Probability Theory and Random Processes (MA 225)

Class Notes

September – November, 2020

Instructor
Ayon Ganguly
Department of Mathematics
IIT Guwahati

Contents

1 Probability **2**
1.1 Probability 2
1.1.1 Classical Probability 2
1.1.2 Countable and Uncountable Sets 3
1.1.3 Axiomatic Probability 4
1.1.4 Continuity of Probability 9
1.2 Conditional Probability 11
1.3 Independence 13

2 Random Variable **17**
2.1 Random Variable 17
2.2 Discrete Random Variable 21
2.3 Continuous Random Variable 25
2.4 Expectation of Random Variable 31
2.5 Transformation of Random Variable 35
2.5.1 Technique 1 35
2.5.2 Technique 2 37
2.5.3 Expectation of Function of RV 40
2.5.4 Technique 3 44
2.6 Moment Inequality 45
2.A Gamma Integral 46
2.B Beta Integral 48

Chapter 2

Random Variable

2.1 Random Variable

In most of the practical situations we are interested in numerical characteristic of a random experiment. For example, we may be interested in number of heads out of 10 tosses of a coin, number of bug reported for a newly developed software, value of total yield of a crop in different months of a year in Assam, the level of water in Brahmaputra river each day at a particular site, etc. Hence, it is helpful to use a function which maps a sample space to $\mathbb{R}$. Such a function is called a random variable. Moreover, we have rich mathematical tools on the set of real numbers. These mathematical tools can be used to analysis several properties of probability of the quantity of interest if we can transform any arbitrary sample space to $\mathbb{R}$ or a subset of $\mathbb{R}$.

Definition 2.1 (Random Variable)**.** *Let $\mathcal{S}$ be a sample space of a random experiment. A function $X : \mathcal{S} \to \mathbb{R}$ is called a random variable (RV).*

Example 2.1. Assume that a coin is tossed n times and the tosses are independent. Let $\mathcal{S}$ be the sample space for this random experiment. Let $X : \mathcal{S} \to \mathbb{R}$ be defined by the number of tails out of n tosses. Here, X is a RV, which can take values from the set $\{0, 1, \ldots, n\}$. We can compute probabilities that the RV takes several values from its' range, when we know the probability space $(\mathcal{S}, \mathcal{F}, P)$. For simplicity, suppose that $n = 2$. Then X takes value zero if and only if both the tosses result in heads. X takes value two if and only if both the tosses result in tails. X takes value one if and only if one of the tosses results in head and another in tail. Hence,

$$P(X = 0) = P(HH) = \frac{1}{4}.$$
$$P(X = 2) = P(TT) = \frac{1}{4}.$$
$$P(X = 1) = P(HT, TH) = \frac{1}{2}.$$

Note the technique of the computation of probabilities that the RV takes several values. To compute $P(X = x)$, we first find the inverse image of $X = x$ and then compute the required probability. ||

Example 2.2. Consider the random experiment of rolling a fair die twice. Assume that the throws are independent. Let $X : \mathcal{S} \to \mathbb{R}$ be defined by the sum of the outcomes of two rolls. Clearly, X is a RV. In this case, using the technique of the previous example, $P(X=2)=1/36, P(X=3)=2/36, P(X=4)=3/36, P(X=5)=4/36, P(X=6)=5/36, P(X=7)=6/36,$ $P(X=8)=5/36, P(X=9)=4/36, P(X=10)=3/36,$ $P(X=11)=2/36, P(X=12)=1/36.$ ||

Example 2.3. Suppose we are testing the reliability of a battery. In a reliability testing, an experimenter wants to know different characteristic of lifetime of a product. For example, one may want to know the average lifetime of batteries manufactured by a company, or proportion of batteries that can work beyond two years of use. In a typical reliability experiment, certain number of items are put on a life testing experiment and the failure times of the items are recorded. The outcomes are the lifetime of the product. Thus, the sample space can be taken as $\mathcal{S}=(0,\,\infty)$. Let us define $X:\mathcal{S}\to\mathbb{R}$ by $X_1(\omega)=\omega$. Clearly, X_1 denote the lifetime of the battery. Now, suppose we are mainly interested in whether the battery would last more than 2 years or not. Then we can take a RV $X_2:\mathcal{S}\to\mathbb{R}$ defined by $X_2(\omega)=I_{(2,\infty)}(\omega)$, where I_A denotes the indicator function of the set A and is defined by

$$I_A(x)=\begin{cases}1 & \text{if } x\in A\\ 0 & \text{if } x\notin A.\end{cases} \tag{2.1}$$

Clearly, X_2 indicates if a battery has lifetime more than two years or not. If the lifetime of a battery is less than or equal to two years, then the value of X_2 is zero. On the other hand, the value of X_2 is one if the lifetime is more than two years.

Now, for some interval $I\subset(0,\,\infty)$, $P(I)=\int_I e^{-t}dt$ defines a probability on the Borel σ-field on the positive part of real line. This Borel σ-field is denoted by $\mathcal{B}(0,\infty)$. The exact definition of $\mathcal{B}(0,\,\infty)$ is not possible to provide in this course. However, all kind of intervals that are subset of $(0,\,\infty)$, single-ton sets that are subsets of $(0,\,\infty)$ belong to $\mathcal{B}(0,\,\infty)$. Though it is very difficult to construct, but there exist subsets of $(0,\,\infty)$ which does not belong to $\mathcal{B}(0,\,\infty)$. Thus, $\mathcal{B}(0,\,\infty)$ is not the power set of $(0,\,\infty)$. The Borel σ-field is a very standard and meaningful σ-field, which is used for all most all practical situations. Also, as the exact definition of Borel σ-algebra is not possible to give here, it is also not possible to prove that $P(I)$ defines a probability on $\mathcal{B}(0,\,\infty)$. Please take it as an information and proceed. We can, now, calculate the probability that $X_1\le x$ as

$$P\left(X_1\le x\right)=\int_0^x e^{-t}dt=1-e^{-x}$$

for $x>0$. Also, we can calculate the probability of $X_2=0$ and $X_2=1$ as follows.

$$P(X_2=0)=P\left(X_1\le 2\right)=1-e^{-2} \quad\text{and}\quad P(X_2=1)=P\left(X_1>2\right)=1-P\left(X_1\le 2\right)=e^{-2}.$$

||

Definition 2.2 (Cumulative Distribution Function). *The cumulative distribution function (CDF) of a RV X is a function $F_X:\mathbb{R}\to[0,\infty)$ defined by*

$$F_X(x)=P(X\le x).$$

Note that CDF is defined for all real numbers. Though in the definition $[0,\infty)$ is written as co-domain for CDF, it clear that CDF lies in the interval $[0,1]$ for all real numbers.

Example 2.4 (Continuation of Example 2.1). As the random variable cannot take any negative value, $F_X(x) = P(X \leq x) = 0$ for all $x < 0$. Now, if we try to compute CDF at 0.5, we need to compute $P(X \leq 0.5) = P(X = 0) = 1/4$. The same argument will hold for all $x \in [0, 1)$, and hence, $F_X(x) = 1/4$ for all $x \in [0, 1)$.

Now, suppose we want to compute the CDF at 1.4. To compute it, we need to find $F_X(1.4) = P(X \leq 1.4) = P(X = 0 \text{ or } 1) = P(HH, HT, TH) = 1/4 + 1/2 = 3/4$. Just like the previous case, we can see that $F_X(x) = 3/4$ for all $x \in [1, 2)$. Proceeding in this way, we obtain the CDF of X as

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0, \\ \frac{1}{4} & \text{if } 0 \leq x < 1, \\ \frac{3}{4} & \text{if } 1 \leq x < 2, \\ 1 & \text{if } x \geq 2. \end{cases}$$

||

Example 2.5 (Continuation of Example 2.2). Following the arguments of the last example, one can find the CDF of X and it is given by

$$F_X(x) = \begin{cases} 0 & \text{if } x < 2 \\ 1/36 & \text{if } 2 \leq x < 3 \\ 3/36 & \text{if } 3 \leq x < 4 \\ 6/36 & \text{if } 4 \leq x < 5 \\ 10/36 & \text{if } 5 \leq x < 6 \\ 15/36 & \text{if } 6 \leq x < 7 \\ 21/36 & \text{if } 7 \leq x < 8 \\ 26/36 & \text{if } 8 \leq x < 9 \\ 30/36 & \text{if } 9 \leq x < 10 \\ 33/36 & \text{if } 10 \leq x < 11 \\ 35/36 & \text{if } 11 \leq x < 12 \\ 1 & \text{if } x \geq 12. \end{cases}$$

||

Example 2.6 (Continuation of Example 2.3). As lifetime cannot be negative, $P(X_1 \leq x) = 0$ for all $x < 0$. For $x > 0$, $P(X_1 \leq x) = P(0 < X_1 \leq x) = \int_0^x e^{-t} dt = 1 - e^{-x}$. Here, the first equality holds as $P(X < 0) = 0$. Hence, the CDF of X_1 is given by

$$F_{X_1}(x) = \begin{cases} 0 & \text{if} \quad x < 0, \\ 1 - e^{-x} & \text{if} \quad x \geq 0\,. \end{cases}$$

To find the CDF of X_2, note that X_2 can takes two values, *viz.*, 0 and 1. Hence, the CDF of X_2 is given by

$$F_{X_2}(x) = \begin{cases} 0 & \text{if} \quad x < 0, \\ 1 - e^{-2} & \text{if} \quad 0 \leq x < 1, \\ 1 & \text{if} \quad x \geq 1\,. \end{cases}$$

||

Theorem 2.1 (Properties of CDF). *The CDF of a RV has the following properties:*

1. $F_X(\cdot)$ *is non-decreasing.*
2. $\lim_{x \uparrow \infty} F_X(x) = 1$.
3. $\lim_{x \downarrow -\infty} F_X(x) = 0$.
4. $\lim_{h \downarrow 0} F_X(x+h) = F_X(x)$ *for all* $x \in \mathbb{R}$. *Hence, CDF is right continuous.*
5. $\lim_{h \downarrow 0} F_X(x-h) = F_X(x) - P(X = x)$ *for all* $x \in \mathbb{R}$.

Proof:

1. To show CDF is a non-decreasing function, we need to show that for $x_1 < x_2$, $F_X(x_1) \leq F_X(x_2)$. Now, note that $\{X \leq x_1\} \subset \{X \leq x_2\}$, which implies $P(X \leq x_1) \leq P(X \leq x_2)$. It proves the statement.

2. Let $\{x_n\}_{n\geq 1}$ be an increasing sequence of real numbers such that $x_n \to \infty$ as $n \to \infty$. Let us define a sequence of events $A_n = \{\omega \in \mathcal{S} : X(\omega) \in (-\infty,\, x_n]\}$ for all $n \geq 1$. As $\{x_n\}_{n\geq 1}$ is increasing, $\{A_n\}_{n\geq 1}$ is also increasing sequence of events. Thus,
$$\lim_{n\to\infty} A_n = \cup_{n=1}^{\infty} A_n = \mathcal{S}.$$
Now,
$$\lim_{n\to\infty} F_X(x_n) = \lim_{n\to\infty} P(A_n) = P\left(\lim_{n\to\infty} A_n\right) = P(\mathcal{S}) = 1.$$
This shows that for any increasing sequence $\{x_n\}_{n\geq 1}$ of real numbers with $x_n \to \infty$ as $n \to \infty$, $\lim_{n\to\infty} F_X(x_n) = 1$. Thus, $\lim_{x\to\infty} F_X(x) = 1$.

3. Let $\{x_n\}_{n\geq 1}$ be a decreasing sequence of real numbers such that $x_n \to -\infty$ as $n \to \infty$. Take $B_n = \{\omega \in \mathcal{S} : X(\omega) \leq x_n\}$ for all $n = 1, 2, \ldots$. Then $\{B_n\}_{n\geq 1}$ is a decreasing sequence events and $\lim_{n\to\infty} B_n = \emptyset$. Hence,
$$\lim_{n\to\infty} F_X(x_n) = \lim_{n\to\infty} P(B_n) = P\left(\lim_{n\to\infty} B_n\right) = 0.$$
This shows that for any decreasing sequence $\{x_n\}_{n\geq 1}$ of real numbers such that $x_n \to -\infty$ as $n \to \infty$, $\lim_{n\to\infty} F_X(x_n) = 0$. Hence, $\lim_{x\to-\infty} F_X(x) = 0$.

4. Fix $x \in \mathbb{R}$. Let $\{x_n\}_{n\geq 1}$ be a decreasing sequence of real numbers such that $x_n \to x$ as $n \to \infty$. Assume that $C_n = \{\omega \in \mathcal{S} : X(\omega) \in (-\infty,\, x_n]\}$ for all $n \geq 1$. Clearly, $\{C_n\}_{n\geq 1}$ is decreasing sequence of events. Hence,
$$\lim_{n\to\infty} C_n = \cap_{n=1}^{\infty} C_n = \{\omega \in \mathcal{S} : X(\omega) \in (-\infty,\, x]\}.$$
Now,
$$\lim_{n\to\infty} F_X(x_n) = \lim_{n\to\infty} P(C_n) = P\left(\lim_{n\to\infty} C_n\right) = F_X(x).$$
This completes the proof.

5. Fix $x \in \mathbb{R}$. Taking an increasing sequence $\{x_n\}_{n\geq 1}$ of real numbers such that $x_n \to x$ as $n \to \infty$, we can prove this part like the previous parts. *(Complete it.)*

$\square$

Theorem 2.2. *Let $G : \mathbb{R} \to \mathbb{R}$ be a function satisfying properties 1–4 of the Theorem 2.1. Then $G(\cdot)$ is a CDF of a RV.*

Proof: The proof of the theorem is out of the scope of this course. □

Though the proof is out of scope, it is an important theorem. This theorem can be use to check if a given function is a CDF or not. We need to check if Properties 1–4 of the Theorem 2.1 are satisfied by the given function or not. If the function satisfy the all the four properties, then it is a CDF. Otherwise, it is not a CDF.

By distribution (or probability distribution) of a RV, we mean how the probability is distributed over the real line for the RV. One of the ways to see the distribution of a RV is through CDF of the RV. Note that a random variable is just a function defined on the sample space and does not depend on probability that is defined on the σ-field. However, the distribution of the RV depends on the probability. Hence, keeping the function same if we change the probability then the RV will remain same but its distribution will change. Consider the Example 2.1 with changed probabilities $P(HH) = 9/16, P(TT) = 1/16, P(HT) = P(TH) = 3/16$. The CDF of X in this case is given by

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{9}{16} & \text{if } 0 \le x < 1 \\ \frac{15}{16} & \text{if } 1 \le x < 2 \\ 1 & \text{if } x \ge 2. \end{cases}$$

Definition 2.3 (Atom)**.** *If $x \in \mathbb{R}$ is such that $P(X = x) > 0$, then x is said to be an atom of the CDF of X or simple atom of X.*

The first property of a CDF says that a CDF is always non-decreasing. That means that a CDF can have only jump discontinuities. From the third property of a CDF, we know that a CDF is always right continuous. However, it can have a discontinuity from the left due to the fourth property. Thus, if a CDF has a discontinuity at a point x, then x is an atom of the CDF. On the other hand if the distribution function of a RV has no atoms, then the CDF of the RV is a continuous function.

We can write probabilities of different events relating to a random variables in terms of CDF of the random variable. Consider the following cases in this regard. By definition of the CDF, $P(X \le a) = F_X(a)$. From the fourth property of CDF (Theorem 2.1), we have $P(X < a) = \lim_{x \uparrow a} F_X(x) = F_X(a-)$ and $P\left(X = a\right) = F_X(a) - F_X(a-)$. Moreover,

$$\begin{aligned}
P(a < X \le b) &= P(X \le b) - P\left(X \le a\right) = F_X(b) - F_X(a). \\
P(a \le X \le b) &= P\left(X \le b\right) = P\left(X < a\right) = F_X(b) - F_X(a-). \\
P(a < X < b) &= P\left(X < b\right) - P\left(X \le a\right) = F_X(b-) - F_X(a). \\
P(a \le X < b) &= P\left(X < b\right) - P\left(X < a\right) = F_X(b-) - F_X(a-).
\end{aligned}$$

2.2 Discrete Random Variable

Definition 2.4 (Discrete Random Variable)**.** *A RV is said to have discrete distribution if there exists an atmost countable set $S_X \subset \mathbb{R}$ such that $P(X = x) > 0$ for all $x \in S_X$ and $\sum_{x \in S_X} P(X = x) = 1$. S_X is called the support of X. A RV having discrete distribution is called a DRV.*

Definition 2.5 (Probability Mass Function)**.** *Let X be a RV having discrete distribution with support S_X. Define a function $f_X : \mathbb{R} \to [0, 1]$ by*

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

The function f_X is called the probability mass function (PMF) of X.

Example 2.7 (Continuation of Example 2.4)**.** In Example 2.4, we have seen the CDF of the RV X is given by

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{4} & \text{if } 0 \le x < 1 \\ \frac{3}{4} & \text{if } 1 \le x < 2 \\ 1 & \text{if } x > 2. \end{cases}$$

Let us check if X is DRV. Let us consider the set $D = \{0,\ 1,\ 2\}$, which is a finite set, and hence, it is atmost countable. Now,

$$P\left(X = 0\right) = F_X(0) - F_X(0-) = \frac{1}{4} > 0.$$
$$P\left(X = 1\right) = F_X(1) - F_X(1-) = \frac{1}{2} > 0.$$
$$P\left(X = 2\right) = F_X(2) - F_X(2-) = \frac{1}{4} > 0.$$

Also, $P(X \in D) = P(X = 0) + P(X = 1) + P(X = 2) = 1$. In this case, we have an atmost countable set $S_X = D$, such that $P(X = x) > 0$ for all $x \in S_X$ and $\sum_{x \in S_X} P(X = x) = 1$. Hence, X is a DRV. ||

Note that $P(X = x)$ is strictly greater than zero if and only if there is an atom at x. In other words, $P(X = x) > 0$ if and only if x is a point of discontinuity of the CDF of the RV. Hence, to check if a RV is DRV or not, we should start with the set of all points of discontinuities of the CDF of the RV. If D denote the set of all points of discontinuities of $F_X(\cdot)$, then we should check if $P(X \in D)$ equal one or not. If $P(X \in D) = 1$, then the corresponding random variable is DRV. If not, then it is not a DRV.

Example 2.8 (Continuation of Example 2.5)**.** The RV X in Example 2.5 can be shown a DRV by taking $S_X = \{2,\ 3, \ldots,\ 12\}$. I leave it as a practice problem and please complete it. ||

Example 2.9 (Continuation of Example 2.6)**.** The CDF of the RVs X_1 is a continuous function, hence, there does not exists any atmost countable set S_X, such that $P\left(X \in S_X\right) = 1$. Therefore, X_1 is not a DRV. You can easily show that X_2 is a DRV. ||

Note that if we are given with a CDF, in principle we should be able to say whether the RV is discrete or not. If the RV is discrete, we should also be able to find the PMF of the RV using the formula

$$f_X(x) = F_X(x) - F_X(x-).$$

On the other hand, if a PMF of a DRV is given, the CDF of the RV can be computed as

$$F_X(x) = \sum_{\substack{y \in S_X \\ y \le x}} f_X(y).$$

Thus, for a DRV, CDF and PMF have a one-one correspondence in the sense that if one of them is given, then other one can be found uniquely.

Theorem 2.3 (Properties of PMF)**.** *Let X be a DRV with PMF $f_X(\cdot)$ and support S_X. Then*

1. $f_X(x) \ge 0$ *for all* $x \in \mathbb{R}$.
2. $\displaystyle\sum_{x \in S_X} f_X(x) = 1$.

Proof: Straight forward using the definition of PMF. □

Theorem 2.4. *Suppose a real valued function $h : \mathbb{R} \to \mathbb{R}$ satisfies the following two conditions:*

1. $h(x) \ge 0$ *for all* $x \in \mathbb{R}$ *and* $D = \{x : h(x) > 0\}$ *is atmost countable.*
2. $\sum_{x \in D} h(x) = 1$.

Then $h(\cdot)$ is a probability mass function of some DRV.

Proof: The proof of this theorem is out of scope of this course. □

Example 2.10 (Bernoulli Distribution)**.** Consider a random experiment which has two possible outcomes. Such a random experiment is called Bernoulli experiment or Bernoulli trial. Examples of Bernoulli experiment include tossing a coin, test if a person is infected with novel corona virus, checking if a mobile phone works more than three years, etc. It is customary to name one of the outcome as success and other as failure. Thus, in a coin toss experiment, we may say that getting a tail is a success and head is failure. In a COVID19 testing experiment, we may say having the virus is a success and not having it a failure. Suppose that the probability of success is $p \in [0, 1]$ and that of failure is $1 - p$. In this case, the sample space is $\mathcal{S} = \{S, F\}$, where S and F denote a success and a failure, respectively. Let us define a RV

$$X = \begin{cases} 1 & \text{if a success occurs} \\ 0 & \text{if a failure occurs.} \end{cases}$$

Then it is clear that $P(X = 1) = p$ and $P(X = 0) = 1 - p$. The corresponding CDF is given by

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - p & \text{if } 0 \le x < 1 \\ 1 & \text{if } x \ge 1. \end{cases}$$

It is easy to show that X is a DRV with PMF

$$f_X(x) = \begin{cases} 1-p & \text{if } x = 0 \\ p & \text{if } x = 1. \end{cases} \tag{2.2}$$

The distribution of a DRV having the PMF given in (2.2) is called a Bernoulli distribution with parameter $p \in [0,\, 1]$ and the RV is called a Bernoulli RV. We will use $X \sim Bernoulli(p)$ to denote that the RV X follows a Bernoulli distribution with parameter p. ||

Example 2.11 (Binomial Distribution)**.** Consider a Bernoulli experiment with probability of success p. The experiment is repeated n times independently. The sample space is given by

$$\mathcal{S} = \left\{(\omega_1,\, \omega_2,\, \ldots,\, \omega_n) : \omega_i \in \{S,\, F\} \text{ for all } i = 1,\, 2,\, \ldots,\, n\right\}.$$

Let the RV X denote the number of successes that occur out of n trials of the Bernoulli experiment. Clearly, X takes values in the set $D = \{0,\, 1,\, \ldots,\, n\}$. Let us try to calculate $P\left(X = k\right)$ for $k \in D$. The event $X = k$ means that there are exactly k successes and $n-k$ failures. Now, the probability of getting a particular arrangement of k successes and $n-k$ failures is $p^k(1-p)^{n-k}$. The event $X = k$ will occurs if any one of $\binom{n}{k}$ arrangements of k successes and $n-k$ failures occurs. Hence,

$$P\left(X = k\right) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for } k \in D.$$

Using binomial expansion, $\sum_{k \in D} P\left(X = k\right) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = 1$. Hence, X is a DRV with PMF

$$f_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{if } x = 0,\, 1,\, \ldots,\, n \\ 0 & \text{otherwise.} \end{cases} \tag{2.3}$$

The distribution of a DRV having the PMF given in (2.3) is called a binomial distribution with parameters $n \in \mathbb{N}$ and $p \in [0,\, 1]$ and the RV is called a binomial RV. We will use $X \sim Bin(n,\, p)$ to denote that the RV X follows a binomial distribution with parameters n and p. ||

Example 2.12 (Geometric Distribution)**.** Let a Bernoulli experiment with success probability p be repeated again and again, independently. The sample space is

$$\mathcal{S} = \left\{(\omega_1,\, \omega_2,\, \ldots) : \omega_i \in \{S,\, F\} \text{ for all } i = 1,\, 2,\, \ldots\right\}.$$

Let X be a function that denotes the number failures before the first success. Clearly, the range of X is $D = \{0,\, 1,\, \ldots\}$. Then for $k \in D$, $P(X = k) = p(1-p)^k$. To see it, notice that the event $X = k$ occurs if and only if there are k failures occur in first k trials, and then a success occurs at the $(k+1)$st trial. As the trials are independent, $P(X = k) = p(1-p)^k$. Using the geometric series, it is easy to see that

$$\sum_{k \in D} P(X = k) = \sum_{k=0}^{n} p(1-p)^k = 1.$$

Hence, X is a DRV with PMF

$$f_X(x) = \begin{cases} p(1-p)^x & \text{if } x = 0,\, 1,\, 2,\, \ldots \\ 0 & \text{otherwise.} \end{cases} \tag{2.4}$$

The distribution of a DRV having the PMF given in (2.4) is called a geometric distribution with parameter $p \in [0, 1]$ and the RV is called a geometric RV. We will use $X \sim Geo(p)$ to denote that the RV X follows a geometric distribution with parameter p. ||

Example 2.13 (Poisson Distribution). Consider the function

$$f_X(x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} & \text{if } x = 0, 1, 2, \dots \\ 0 & \text{otherwise,} \end{cases} \tag{2.5}$$

where $\lambda > 0$. As $f_X(x) \geq 0$ for all $x \in \mathbb{R}$, $D = \{x \in \mathbb{R} : f_X(x) > 0\}$ is countable, and $\sum_{x=0}^{\infty} f_X(x) = 1$, $f_X(\cdot)$ is a PMF. The distribution of a DRV having the PMF given in (2.5) is called a Poisson distribution with parameter $\lambda > 0$ and the RV is called a Poisson RV. We will use $X \sim Poi(\lambda)$ to denote that the RV X follows a Poisson distribution with parameter λ.

Unlike the previous cases, we do not discuss the motivation of Poisson distribution here. We will comeback to the issue later. However, this is a very useful distribution, which has a wide range of applications in a diverse number of areas. Applications of Poisson distribution include modelling of number of accidents in a particular spot of a city, number of misprints in a each page of a book having large number of pages, number of dinners in a certain restaurant, number of customers on a server at a given time, etc. ||

Example 2.14. Suppose that an airplane engine will fail, when in flight, with probability $1 - p$ independently from engine to engine. The airplane will make a successful flight if at least 50 percent of its engines remain operating. As each engine is assumed to fail or function independently of what happens with the other engines, it follows that the number of engines remaining operative is a binomial RV with parameter p. Hence, the probability that a four-engine plane makes a successful flight is

$$\binom{4}{2}p^2(1-p)^2 + \binom{4}{3}p^3(1-p) + \binom{4}{4}p^4 = 6p^2(1-p)^2 + 4p^3(1-p) + p^4,$$

whereas the corresponding probability for a two-engine plane is

$$\binom{2}{1}p(1-p) + \binom{2}{2}p^2 = 2p(1-p) + p^2.$$

Hence, the four-engine plane is safer if

$$6p^2(1-p)^2 + 4p^3(1-p) + p^4 \geq 2p(1-p) + p^2 \implies p \geq \frac{2}{3}.$$

Therefore, the four-engine plane is safer when the engine success probability is at least as large as $\frac{2}{3}$, where the two-engine plane is safer if this probability falls below $\frac{2}{3}$. ||

2.3 Continuous Random Variable

Definition 2.6 (Continuous Random Variable and Probability Density Function). *A RV is said to have a continuous distribution if there exists a non-negative integrable function $f_X : \mathbb{R} \to [0, \infty)$ such that*

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

for all $x \in \mathbb{R}$. A RV having a continuous distribution is called a continuous RV (CRV). The function f_X is called the probability density function (PDF). The set $S_X = \{x \in \mathbb{R} : f_X(x) > 0\}$ is called support of X.

Following the definition of CRV, there exists a PDF and $P(X \leq x)$ can be written as the area under the PDF from $-\infty$ to x. Note that for DRV we have PMF and PMF is only defined for DRV, where the PDF is only defined for CRV.

For a continuous RV X, $P(X = a) = 0$ for all $a \in \mathbb{R}$. To see it notice that $P(X = a)$ can be interpreted as the integration of PDF over a single-ton set $\{a\}$. As we know that the integration over a single-ton set is zero, $P(X = a) = 0$ for all $a \in \mathbb{R}$. That means the CDF of a CRV does not have any atom, and hence, the CDF of a CRV is continuous.

Let $f_X(\cdot)$ be a PDF of a CRV. Let a and $b \geq 0$ be a real numbers such that $b \neq f_X(a)$. Define a new function as

$$g(x) = \begin{cases} f_X(x) & \text{if } x \neq a \\ b & \text{if } x = a. \end{cases}$$

It is easy to see that $g(x) \geq 0$ and $F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \int_{-\infty}^{x} g(t)dt$. Hence, $g(\cdot)$ is a PDF corresponding to the CDF $F_X(\cdot)$ and $g(x) \neq f_X(x)$. Thus, PDF is not unique. As a consequence, support of a CRV is also not unique. Note that PMF and support of a DRV are unique. Also, note that $f_X(x)$ is not $P(X = x)$ for $P(X = x)$ is always zero, but $f_X(x)$ can be greater than zero.

For a CRV with PDF $f_X(x)$,

$$P(a < X \leq b) = F_X(b) - F_X(a) = \int_{-\infty}^{b} f_X(t)dt - \int_{-\infty}^{a} f_X(t)dt = \int_{a}^{b} f_X(t)dt.$$

Also, note that $P(a \leq X \leq b) = F_X(b) - F_X(a-) = F_X(b) - F_X(a)$, as the CDF of a CRV is continuous. Therefore, for a CRV X,

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

Example 2.15 (Uniform Distribution)**.** Let the CDF of a RV X is given by

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

where $-\infty < a < b < \infty$. It is very easy to see that for all $x \in \mathbb{R}$, $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$, where

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0 & \text{otherwise.} \end{cases} \tag{2.6}$$

It is also clear that $f_X(x) \geq 0$ for all $x \in \mathbb{R}$. Hence, X is a CRV with PDF (2.6). The distribution of a CRV having the PDF given in (2.6) is called a uniform distribution with parameters a and b $(-\infty < a < b < \infty)$ and the RV is called a uniform RV. We will use $X \sim U(a, b)$ to denote that the RV X follows a uniform distribution with parameters a and b.

Note that $f_X(x)$ is constant in the interval (a, b) and X is just as likely to be near any value on (a, b) as any other value. To check this, note that for any $a < \alpha < \beta < b$,

$$P(\alpha < X < \beta) = \int_{\alpha}^{\beta} f_X(x)dx = \frac{\beta - \alpha}{b - a}.$$

In other words, the probability that X is in any particular subinterval of (a, b) is proportional to the length of that subinterval. ||

Example 2.16 (Exponential Distribution). Consider that X be a RV having CDF given by

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{otherwise,} \end{cases} \tag{2.7}$$

where $\lambda > 0$. Let us define

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{2.8}$$

Clearly, for all $x \in \mathbb{R}$, $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$ and $f_X(x) \geq 0$. Hence, X is a CRV with PDF given in (2.8). The distribution of a CRV having the PDF given in (2.8) is called an exponential distribution with parameter $\lambda > 0$ and the RV is called an exponential RV. We will use $X \sim Exp(\lambda)$ to denote that the RV X follows an exponential distribution with parameter λ. Plot of PDFs and CDFs are given in Figure 2.1 for different values of λ.

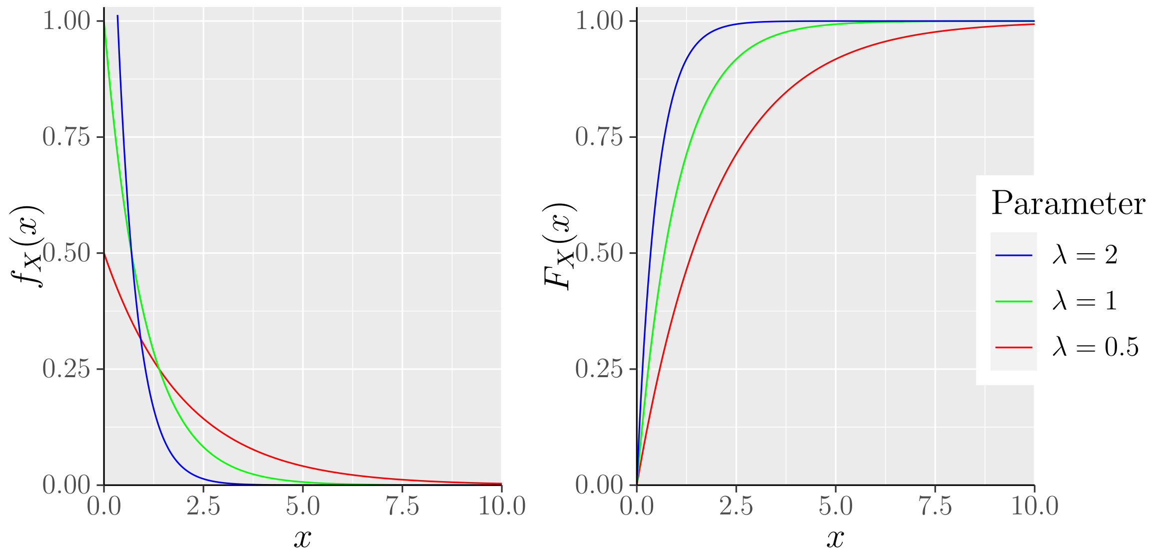

Figure 2.1: Plot of PDFs and CDFs of $Exp(\lambda)$ for several values of λ.

Clearly, PDFs are bounded and decreasing on the positive part of real line. The CDF rushes towards one more quickly for smaller values of λ compared to larger values of the parameter. This means that as λ increases, $P(X \leq x)$ decreases for each fixed $x > 0$. ||

Example 2.17 (Normal Distribution). Let X be a RV with CDF

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \text{ if } -\infty < x < \infty,$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$. It is clear that the corresponding PDF is given by

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \text{ if } -\infty < x < \infty, \tag{2.9}$$

The distribution of a CRV having the PDF given in (2.9) is called a normal distribution with parameters $\mu \in \mathbb{R}$ and $\sigma^2\,(\sigma > 0)$ and the RV is called a normal RV. We will use $X \sim N(\mu,\,\sigma^2)$ to denote that the RV X follows a normal distribution with parameters μ and σ^2. The normal distribution with $\mu = 0$ and $\sigma = 1$ is called a standard normal distribution. If $X \sim N(0,\,1)$, then X is called standard normal RV. The CDF and PDF of a standard normal distribution are denoted by $\Phi(\cdot)$ and $\phi(\cdot)$, respectively. Plots of PDFs and CDFs

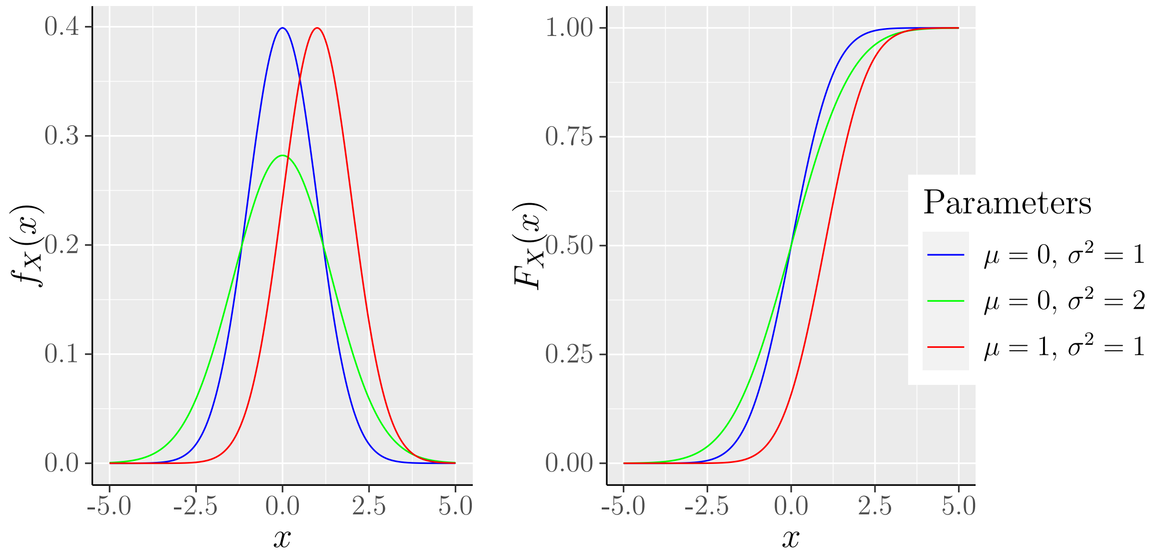

Figure 2.2: Plot of PDFs and CDFs of $N(\mu,\,\sigma^2)$ for several values of parameters.

are given in Figure 2.2 for different values of the parameters. The following points are quite relevant for the plot. The location of the PDF changes with change in μ keeping σ^2 fixed. On the other hand, keeping μ fixed, if we increase σ^2, the PDF becomes flatter and less peaked. Also, the PDF is symmetric with respect to the point μ. ||

Theorem 2.5 (Properties of PDF). *Let X be a CRV with PDF $f_X(\cdot)$. Then*

1. $f_X(x) \geq 0$ *for all* $x \in \mathbb{R}$.
2. $\int_{-\infty}^{\infty} f_X(x) = 1$.

Proof: 1. Straight forward form the definition of PDF.

2.
$$\int_{-\infty}^{\infty} f_X(x)dx = \lim_{A\to\infty} \int_{-\infty}^{A} f_X(x)dx = \lim_{A\to\infty} F_X(A) = 1.$$

$\square$

Theorem 2.6. *Suppose a real valued function $g : \mathbb{R} \to \mathbb{R}$ satisfies the following conditions:*

1. $g(x) \geq 0$ *for all* $x \in \mathbb{R}$.

2. $\int_{-\infty}^{\infty} g(x)dx = 1$.

Then $g(\cdot)$ is a PDF of some CRV.

Proof: The proof of this theorem is out of scope of the course. $\square$

Example 2.18 (Gamma Distribution). It can be shown that the improper integral

$$\int_0^\infty t^{\alpha-1}e^{-t}dt$$

converges for all $\alpha > 0$ and diverges for all $\alpha \leq 0$. For the proof of the statement along with some properties of the integral, please see Appendix 2.A at the end of this chapter. For $\alpha > 0$ the integral is denoted by $\Gamma(\alpha)$ and is called gamma integral. Thus,

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}dt \quad \text{for } \alpha > 0.$$

Gamma integral gives a probability distribution of a CRV, which is described below. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}e^{-\beta x} & \text{if } x > 0 \\ 0 & \text{otherwise,} \end{cases}$$

where $\alpha > 0$ and $\beta > 0$. It is easy to see that the function $f(\cdot)$ satisfies both the conditions of the previous theorem. Hence, $f(\cdot)$ is a PDF of a CRV. Thus, we can define Gamma distribution as follows. A RV X is said to have a gamma distribution if the PDF of the RV is given by

$$f_X(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}e^{-\beta x} & \text{if } x > 0 \\ 0 & \text{otherwise,} \end{cases}$$

where $\alpha > 0$ and $\beta > 0$. In this case, the RV X is called a gamma RV. We will use the notation $X \sim Gamma(\alpha, \beta)$ to denote the fact that X follows a gamma distribution with parameters $\alpha > 0$ and $\beta > 0$. The plot of PDFs and CDFs of $Gamma(\alpha, \beta)$ are given in Figure 2.3. The following points are quite visible form the plot. On the positive part of real line, for $\alpha \leq 1$, the PDF is strictly decreasing, where the PDF has a unique maxima for $\alpha > 1$. The PDF is unbounded at the point $x = 0$ for $\alpha < 1$. As the shape of the PDF changes with a change in α keeping β fixed, α is called a shape parameter. If we change β keeping α fixed, the flatness of the PDF changes. Hence, β is called a scale parameter. ||

Example 2.19 (Beta Distribution). The improper integral

$$\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$$

converges if and only if $\alpha > 0$ and $\beta > 0$. For the proof of this statement and some main properties of the integral, please see the Appendix 2.B. For $\alpha > 0$ and $\beta > 0$, this improper integral is denoted by $B(\alpha, \beta)$ and is called beta integral. Thus,

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx \quad \text{for } \alpha > 0 \text{ and } \beta > 0.$$

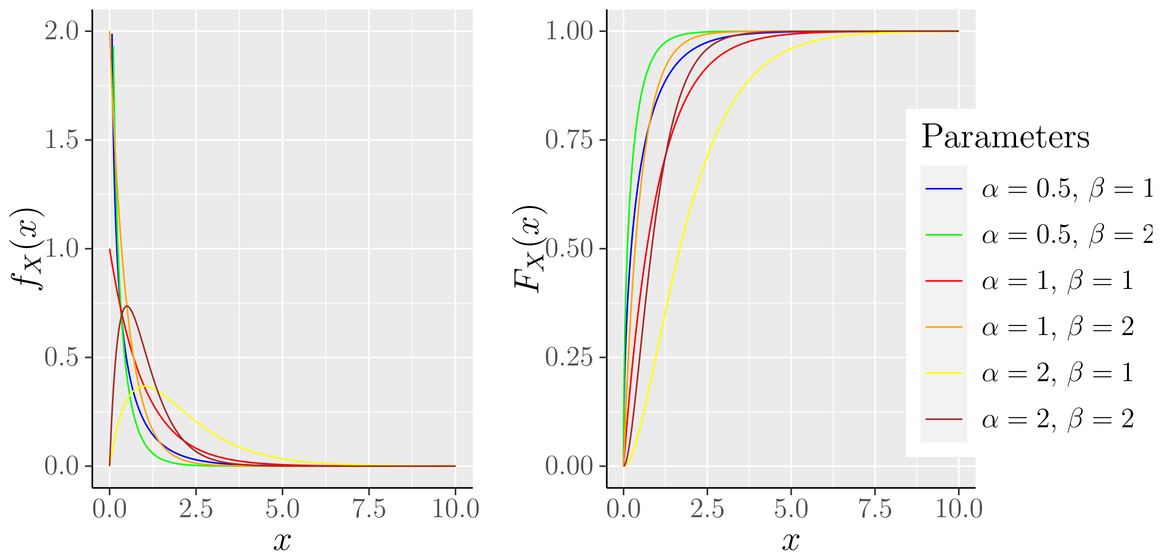

Figure 2.3: Plot of PDFs and CDFs of $Gamma(\alpha,\ \beta)$ for several values of parameters.

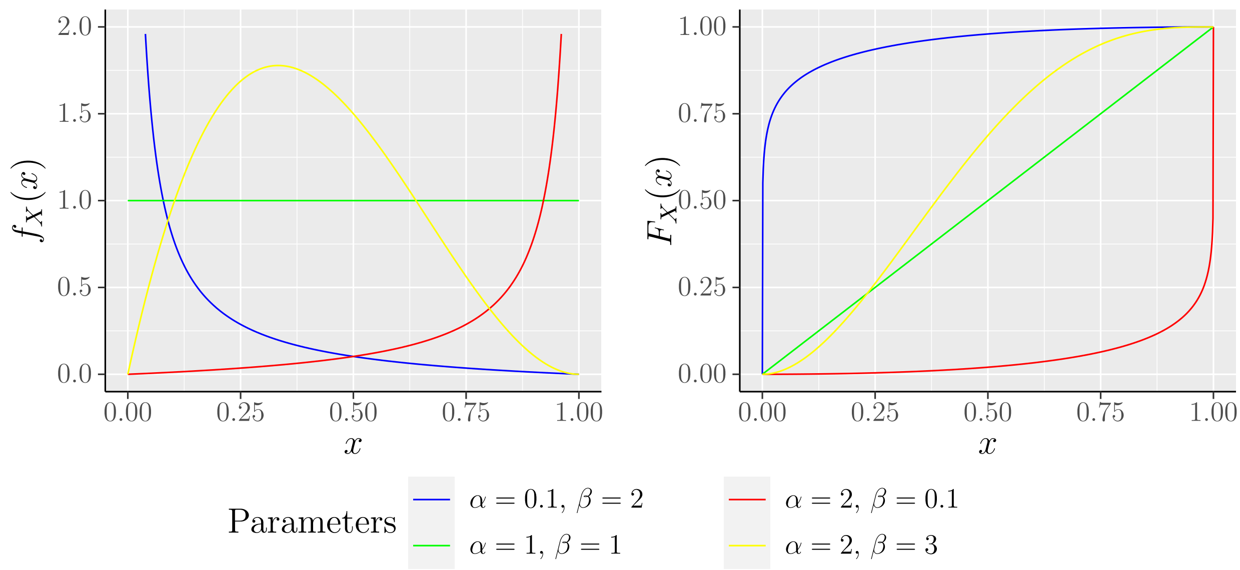

Figure 2.4: Plot of PDFs and CDFs of $Beta(\alpha,\ \beta)$ for several values of parameters.

Like gamma integral, beta integral also gives a PDF, and hence, a distribution of a CRV. Consider the function

$$f(x) = \begin{cases} \frac{1}{B(\alpha,\beta)}\, x^{\alpha-1}(1-x)^{\beta-1} & \text{if } 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to see that $f(\cdot)$ is a PDF. Thus, we have the following definition of beta distribution. A RV X is said to have a beta distribution if the PDF of the RV is given by

$$f_X(x) = \begin{cases} \frac{1}{B(\alpha,\beta)}\, x^{\alpha-1}(1-x)^{\beta-1} & \text{if } 0 < x < 1 \\ 0 & \text{otherwise,} \end{cases}$$

where $\alpha > 0$ and $\beta > 0$. In this case, the RV X is called beta RV with parameters α and β. We will use the notation $X \sim Beta(\alpha,\, \beta)$ to denote the fact that the distribution of X is beta with parameters $\alpha > 0$ and $\beta > 0$. The plot of PDFs and CDFs of $Beta(\alpha,\, \beta)$ is provided in Figure 2.4. It is clear that if $\alpha = \beta = 1$, beta distribution coincides with $U(0,\, 1)$ distribution. The PDF of $Beta(\alpha,\, \beta)$ is unbounded at $x = 0$ and $x = 1$ if $\alpha < 1$ and $\beta < 1$, respectively. For $\alpha \geq 1$ and $\beta \geq 1$, the PDF of beta distribution is bounded. ||

Example 2.20 (RV which is neither discrete nor continuous)**.** Consider the RV X having CDF

$$F_X(x) = \begin{cases} 0 & \text{if} \quad x < -1 \\ x+1 & \text{if} \quad -1 \leq x < -1/2 \\ 1 & \text{if} \quad x \geq -1/2. \end{cases}$$

Note that the CDF is discontinuous only at 1/2. Now, $P(X = 1/2) = 1/2 \neq 1$. Hence, it is not a DRV. On the other hand it cannot be a CRV, as the CDF has discontinuity. Therefore, X is neither DRV nor CRV. Thus, there exists RV, which is not a DRV nor a CRV. Observe that $F_X(x) = \frac{1}{2}F_1(x) + \frac{1}{2}F_2(x)$ for all $x \in \mathbb{R}$, where $F_1(\cdot)$ and $F_2(\cdot)$ are distribution functions and given by

$$F_1(x) = \begin{cases} 0 & \text{if} \quad x < -1 \\ 2(x+1) & \text{if} \quad -1 \leq x < -1/2 \\ 1 & \text{if} \quad x \geq -1/2 \end{cases}$$

and

$$F_2(x) = \begin{cases} 0 & \text{if} \quad x < -1/2 \\ 1 & \text{if} \quad x \geq -1/2. \end{cases}$$

||

2.4 Expectation of Random Variable

Definition 2.7 (Expectation of DRV)**.** *Let X be a discrete RV with PMF $f_X(\cdot)$ and support S_X. The expectation or mean of X is defined by*

$$E(X) = \sum_{x \in S_X} x f_X(x) \quad \text{provided} \quad \sum_{x \in S_X} |x| f_X(x) < \infty.$$

If $\displaystyle\sum_{x \in S_X} |x| f_X(x) = \infty$ then we say that expectation does not exist.

Let us first try to understand the physical meaning of the expectation. This is an intuitive discussion and not mathematically flawless. However, it gives a good intuition of use of expectation. Let X be a discrete random variable that takes values $x_1,\, x_2,\, \ldots,\, x_k$ for some fixed value of k. Let we perform the corresponding random experiment N number of times. Let f_i^N denote frequency of x_i out of N trials. For example let a coin is tossed N times. There are two outcomes. Let X take value 1 if a head appear and 0 otherwise. Let out of

N tosses, f_1^N times head comes and f_2^N times tails comes. In this case, the frequency of 1 is f_1^N and 0 is f_2^N.

The arithmetic average (or mean) of the observed outcomes is given by

$$Ave(N) = \frac{1}{N} \sum_{i=1}^{k} x_i f_i^N.$$

Let us see what will happen to $Ave(N)$ as $N \to \infty$. Note that

$$\lim_{N\to\infty} Ave(N) = \lim_{N\to\infty} \sum_{i=1}^{k} x_i \frac{f_i^N}{N} = \sum_{i=1}^{k} x_i \lim_{N\to\infty} \frac{f_i^N}{N} = \sum_{i=1}^{N} x_i p_i.$$

Consider the quantity $\lim_{N\to\infty} \frac{f_i^N}{N}$. This quantity can be interpreted as the proportion of times x_i observed out of N trials, when N is very large. Therefore, $\frac{f_i^N}{N}$ should, intuitively, converge to probability of $X = x_i$, which is denoted by p_i. This discussion shows that $E(X)$ can be interpreted as the long run mean (or average or weighted average) of the values that are assumed by a DRV.

The condition $\sum_{x\in S_X} |x| f_X(x) < \infty$ is imposed to ensure that the series $\sum_{x\in S_X} x f_X(x)$ converges absolutely. Note that a series $\sum_{n=1}^{\infty} x_n$ is called to converge absolutely if $\sum_{n=1}^{\infty} |x_n| < \infty$. Also, we know that if the series $\sum_{n=1}^{\infty} |x_n| < \infty$, then $\sum_{n=1}^{\infty} x_n$ converges. For expectation, $\sum_{x\in S_X} x f_X(x)$ absolutely converges if $\sum_{x\in S_X} |x f_X(x)| < \infty \implies \sum_{x\in S_X} |x| f_X(x) < \infty$, as $f_X(x) > 0$ for all $x \in S_X$.

If S_X is finite, $\sum_{x\in S_X} |x| f_X(x)$ is always finite. Therefore, $E(X)$ exists when S_X is finite. When S_X is countably infinite, we need to check if $\sum_{x\in S_X} |x| f_X(x)$ is finite or not.

Example 2.21. Let a fair die is rolled and let X denote the outcome of the roll. It is easy to see that X is a DRV with PMF

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

Now, $\sum_{x\in S_X} |x| f_X(s) = \sum_{x=1}^{6} \frac{x}{6} < \infty$ as it is a finite sum. Hence, $E(X)$ exists and is given by

$$E(X) = \sum_{x=1}^{6} \frac{x}{6} = 3.5.$$

Note that the value of $E(X)$ does not belongs to the support of X. In other words, in this example $P\left(X = E(X)\right) = 0$. ||

Example 2.22. Let $X \sim Bin(n,\, p)$. In this case, the support is $S_X = \{0,\, 1,\, \ldots,\, n\}$, which is finite. Hence, $E(X)$ exists. The expectation can be calculated as follows.

$$\begin{aligned}
E(X) &= \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} x \times \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\
&= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!\,((n-1)-(x-1))!} p^{x-1} (1-p)^{(n-1)-(x-1)} \\
&= np \times (p+1-p)^{n-1} \\
&= np.
\end{aligned}$$

For $Bin(n,\, p)$ distribution, the success probability is p and we are performing n independent trials. If we interpret the probability as long term proportion, then out of n trials there will be approximately np successes. Hence, the value of the expectation is quite intuitive. ||

Example 2.23. Let $X \sim Geo(p)$. In this case, the support is $S_X = \{0,\, 1,\, \ldots\}$, which is a countably infinite set. Hence, first we need to check if the condition of the definition of expectation hold or not. To check it, we can proceed as follows:

$$\sum_{x=0}^{\infty} |x| f_X(x) = \sum_{x=1}^{\infty} x p (1-p)^x = p \sum_{x=1}^{\infty} x q^x,$$

where $q = 1-p$. Now, consider the partial sum

$$S_n = \sum_{x=1}^{n} x q^x = \frac{q(1-q^n)}{p^2} - \frac{n q^{n+1}}{p} \to \frac{q}{p^2}$$

as $n \to \infty$. Hence, $\sum_{x=0}^{\infty} |x| f_X(x) < \infty$ and the $E(X)$ exists. The expectation is given by

$$E(X) = \sum_{x=0}^{\infty} x f_X(x) = \frac{q}{p}.$$

||

Example 2.24. $X \sim Poi(\lambda)$. It can be shown that the expectation exists and $E(X) = \lambda$. Technique of showing it is similar to that of the above example. Therefore, I leave it for you to complete. ||

Example 2.25. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} \frac{c}{x^2}, & x \in \mathbb{N}, \\ 0 & \text{otherwise}, \end{cases}$$

where $c = \left(\sum_{n=1}^{\infty} \frac{1}{n^2} \right)^{-1}$. Clearly, $f(x) \geq 0$. The function $f(\cdot)$ is strictly greater than zero only on the set of natural numbers $\mathbb{N}$ and $\sum_{x=1}^{n} f(x) = 1$. Hence, $f(\cdot)$ is a PMF of a DRV, say X, with support $\mathbb{N}$, which is countably infinite. To compute $E(X)$, we first need to check if $\sum_{x=1}^{n} |x|f(x) < \infty$. Now,

$$\sum_{x=0}^{\infty} |x|f(x) = c\sum_{n=1}^{\infty} \frac{1}{n},$$

which is not finite. Hence, $E(X)$ does not exist in this case. $||$

Definition 2.8 (Expectation of CRV)**.** *Let X be a CRV with PDF $f_X(\cdot)$. The expectation of X is defined by*

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)dx \quad \text{provided} \quad \int_{-\infty}^{\infty} |x| f_X(x)dx < \infty.$$

Example 2.26. Let $X \sim U(a, b)$. First we need to check if $\displaystyle\int_a^b |x| \times \frac{1}{b-a}dx < \infty$.

For $a < b < 0$, $\displaystyle\int_a^b \frac{|x|}{b-a}dx = \int_a^b \frac{-x}{b-a}dx = -\frac{a+b}{2} < \infty$.

For $a < 0 < b$, $\displaystyle\int_a^b \frac{|x|}{b-a}dx = \int_a^0 \frac{-x}{b-a}dx + \int_0^b \frac{x}{b-a}dx = \frac{a^2+b^2}{2(b-a)} < \infty$.

For $0 < a < b$, $\displaystyle\int_a^b \frac{|x|}{b-a}dx = \int_a^b \frac{x}{b-a}dx = \frac{a+b}{2} < \infty$.

Hence, $E(X)$ exists and $E(X) = \displaystyle\int_a^b \frac{x}{b-a}dx = \frac{a+b}{2}$. Loosely speaking, for a uniform distribution all points in $(a,\, b)$ are equally likely, and hence, it is expected that the mean should be the middle point of the interval $(a,\, b)$. $||$

Example 2.27. Let $X \sim N(\mu, \sigma^2)$. Here, $E(X)$ exists for the following argument.

$$\begin{aligned}
\int_{-\infty}^{\infty} |x| f_X(x)dx &= \int_{-\infty}^{\infty} |x| \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |\mu + \sigma z| e^{-z^2/2}dz, \quad \text{taking } z = \frac{x-\mu}{\sigma} \\
&\leq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(|\mu| + \sigma|z|\right) e^{-z^2/2}dz \\
&= \frac{\sqrt{2}}{\sqrt{\pi}}|\mu| \int_0^{\infty} e^{-z^2/2}dz + \frac{\sqrt{2}\sigma}{\sqrt{\pi}} \int_0^{\infty} z e^{-z^2/2}dz.
\end{aligned}$$

Now, for a $A > 0$, $\displaystyle\int_0^A z e^{-z^2/2}dz = 1 - e^{-A^2/2}$, which implies that $\displaystyle\int_0^{\infty} z e^{-z^2/2}dz < \infty$. Also, notice that

$$\int_0^{\infty} e^{-z^2/2}dz = \int_0^1 e^{-z^2/2}dz + \int_1^{\infty} e^{-z^2/2}dz \leq \int_0^1 e^{-z^2/2}dz + \int_1^{\infty} z e^{-z^2/2}dz < \infty.$$

Here, the first integration is a proper definite integration, where the integrand is bounded and continuous, and the range of integration is bounded. The second integration is finite by the argument used in the previous case. The expectation of X is given by

$$\begin{aligned}
\int_{-\infty}^{\infty} x f_X(x)dx &= \int_{-\infty}^{\infty} x\frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\mu+\sigma z)e^{-z^2/2}dz, \quad \text{taking } z = \frac{x-\mu}{\sigma} \\
&= \mu\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz + \frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty} ze^{-z^2/2/}dz \\
&= \mu.
\end{aligned}$$

Here, the value of the integration in the first term is one, as the integrand is the PDF of a $N(0,\, 1)$ distribution. The value of the integration in the second term is zero, as the integrand is an odd function. ||

Example 2.28. Let X be a CRV having PDF $f_X(x) = \frac{1}{\pi(1+x^2)}$ for all $x \in \mathbb{R}$. The distribution of a CRV having this PDF is called Cauchy distribution. To check if $E(X)$ exist, we can proceed as follows.

$$\int_{-\infty}^{\infty} \frac{|x|}{\pi(1+x^2)}dx = \int_{-\infty}^{0} \frac{-x}{\pi(1+x^2)}dx + \int_{0}^{\infty} \frac{x}{\pi(1+x^2)}dx.$$

Consider the second term. Notice that

$$\int_0^A \frac{x}{1+x^2}dx = \frac{1}{2}\ln(1+A^2) \to \infty \text{ as } A \to \infty.$$

Thus, $\displaystyle\int_{-\infty}^{\infty} \frac{|x|}{\pi(1+x^2)}dx$ is not finite, hence, $E(X)$ does not exist. Note that $\displaystyle\int_{-A}^{A} \frac{x}{1+x^2}dx =$ 0 as the integrand is an odd function. Thus, $\displaystyle\int_{-\infty}^{\infty} \frac{x}{1+x^2}dx$ is conditionally integrable, but $\displaystyle\int_{-\infty}^{\infty} \frac{x}{1+x^2}dx$ is not integrable. ||

2.5 Transformation of Random Variable

Let $g : \mathbb{R} \to \mathbb{R}$ be a function and X be a RV. Then $Y = g(X)$ is also a RV. To see it, notice that a RV is a function from sample space to $\mathbb{R}$. Clearly, for $\omega \in \mathcal{S}$, $Y(\omega) = g(X(\omega)) \in \mathbb{R}$. In this section, our main aim is to find the distribution (CDF/PMF/PDF) of $Y = g(X)$, when the distribution of X is known. We will discuss mainly three techniques. In this section, we will also discuss about the expectation of a function of a RV.

2.5.1 Technique 1

In this technique, we will try to find the CDF of $Y = g(X)$ using the definition of CDF. That means that we will try to find $F_Y(y) = P\left(Y \le y\right) = P\left(g(X) \le y\right)$ for all $y \in \mathbb{R}$. Note that CDF exists for all type of RVs. Therefore, this technique can be used for any type of RV. This technique is best understood by examples.

Example 2.29. Let the RV X has the following PMF:

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

Consider $Y = X^2$. Clearly, for $y < 0$, $F_Y(y) = P\left(X^2 \leq y\right) = P\left(X \in \emptyset\right) = 0$. For $y \geq 0$,

$$F_Y(y) = P\left(X^2 \leq y\right) = P\left(-\sqrt{y} \leq X \leq \sqrt{y}\right).$$

Now, for $0 \leq y < 1$,

$$F_Y(y) = P(X = 0) = \frac{1}{7}.$$

For $1 \leq y < 4$,

$$F_Y(y) = P(X = 0 \text{ or } 1 \text{ or } -1) = \frac{3}{7}.$$

For $4 \leq y < 9$,

$$F_Y(y) = P(X = 0 \text{ or } 1 \text{ or } -1 \text{ or } 2 \text{ or } -2) = \frac{11}{14}.$$

For $y \geq 9$,

$$F_Y(y) = P(X = 0 \text{ or } 1 \text{ or } -1 \text{ or } 2 \text{ or } -2 \text{ or } 3) = 1.$$

Hence, the CDF of Y is

$$F_Y(y) = \begin{cases} 0 & \text{if } y < 0 \\ \frac{1}{7} & \text{if } 0 \leq y < 1 \\ \frac{3}{7} & \text{if } 1 \leq y < 4 \\ \frac{11}{14} & \text{if } 4 \leq y < 9 \\ 1 & \text{if } y \geq 9. \end{cases}$$

In this case, Y is a DRV *(Why? Also, find the PMF of Y.).* ||

Example 2.30. Let the RV X has the following PDF:

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

Again consider the RV $Y = X^2$. For $y < 0$, $F_Y(y) = 0$. Like the previous example, for $y \geq 0$, $F_Y(y) = P\left(-\sqrt{y} \leq X \leq \sqrt{y}\right)$. Now, for $0 \leq y < 1$,

$$F_Y(y) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{|x|}{2} dx = \frac{y}{2}.$$

For $1 \leq y < 4$,

$$F_Y(y) = \int_{-1}^{1} \frac{|x|}{2} dx + \int_{1}^{\sqrt{y}} \frac{x}{3} dx = \frac{1}{6}\left(2 + y\right).$$

For $y \geq 4$,

$$F_Y(y) = \int_{-1}^{2} f(x) = 1.$$

It is clear that Y is a CRV *(Why? Also, find the PDF of Y.).* ||

Example 2.31. Let the RV X has the following PDF:

$$f(x) = \begin{cases} e^{-x} & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Suppose that we want to find the distribution of $Y = [X]$. Here, $[x]$ denotes the largest integer not exceeding x. First notice that $F_Y(y) = P(Y \leq y) = P([X] \leq y) = 0$ for all $y < 0$. For $0 \leq y < 1$,

$$F_Y(y) = P([X] \leq y) = P(X < 1) = \int_{-\infty}^{1} f(x)dx = \int_0^1 e^{-x}dx = 1 - e^{-1}.$$

For $1 \leq y < 2$,

$$F_Y(y) = P([X] \leq y) = P(X < 2) = \int_{-\infty}^{2} f(x)dx = \int_0^2 e^{-x}dx = 1 - e^{-2}.$$

In general, for $i \leq y < i + 1$, where $i = 0,\, 1,\, 2, \ldots$,

$$F_Y(y) = P([X] \leq y) = P(X < i + 1) = \int_{-\infty}^{i+1} f(x)dx = \int_0^{i+1} e^{-x}dx = 1 - e^{-(i+1)}.$$

Thus, the CDF of Y is given by

$$F_Y(y) = \begin{cases} 0 & \text{if } y < 0 \\ 1 - e^{-(i+1)} & \text{if } i \leq y < i + 1, i = 0,\, 1,\, 2,\, \ldots. \end{cases}$$

Now, you can easily check that Y is a DRV by finding $P(Y = y)$ for all $y = 0,\, 1,\, 2,\, \ldots$ and then showing that $\sum_{y=0}^{\infty} P(Y = y) = 1$. Please complete. ||

The basic idea here is to write the event $Y \leq y$ as $X \in A_y$ for appropriate set A_y. In the previous example, we have written $Y \leq y$ as $X \in (-\infty, i + 1)$ for $y \in [i,\, i + 1)$. Then using the distribution of X, one needs to find the probability of the event $X \in A_y$.

2.5.2 Technique 2

In Technique 2, we try to find PMF (if Y is DRV) or PDF (if Y is CRV) of Y directly without finding its' CDF. Obviously, first we need to understand whether Y is DRV or CRV. This technique is mainly based on two theorems. The first theorem consider the case when X is DRV. We will see that if X is DRV, then Y is also a DRV. The second theorem addresses the case when X is CRV. We will see the under some condition, Y is a CRV if X is CRV. With examples, we will illustrate that if the conditions do not hold, then Y can be DRV as well as CRV. Hence, those conditions are important. Let us start with an example.

Example 2.32. Let the RV X has the following PMF:

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

Consider $Y = X^2$ and suppose that we want to find PMF or PDF, whatever applicable, of Y. Note that the support of X is $S_X = \{-2, -1, 0, 1, 2, 3\}$. Intuition says that Y should takes value from the set $D = \{0, 1, 4, 9\}$ with positive probabilities. Based on this intuition, we will try to find $P(Y = y)$ for all $y \in D$ and then check if $\sum_{y \in D} P(Y = y)$ equal one or not.

$$\begin{aligned}
P(Y = 0) &= P(X = 0) = \frac{1}{7}. \\
P(Y = 1) &= P(X = 1 \text{ or } -1) = \frac{2}{7}. \\
P(Y = 4) &= P(X = 2 \text{ or } -2) = \frac{5}{14}. \\
P(Y = 9) &= P(X = 3 \text{ or } -3) = \frac{3}{14}.
\end{aligned}$$

Note that again to compute $P(Y = y)$, we first find the inverse image of $Y = y$ as $X \in A_y$ and then used the distribution of X. Thus, $A_y = \{x \in \mathbb{R} : x^2 = y\}$. In the last case, $P(Y = 9)$, suggests that even we do not need to consider all the elements x such that $x^2 = 9$. We need to only consider those x, which are in S_X and $x^2 = y$. Thus, we can take $A_y = \{x \in S_X : x^2 = y\}$. It is clear that $\sum_{y \in D} P(Y = y) = 1$. Hence, Y is a DRV with support D and PMF

$$f(y) = \begin{cases} \frac{1}{7} & \text{if } y = 0 \\ \frac{2}{7} & \text{if } y = 1 \\ \frac{5}{14} & \text{if } y = 4 \\ \frac{4}{14} & \text{if } y = 9 \\ 0 & \text{otherwise.} \end{cases}$$

||

Theorem 2.7. *Let X be a DRV with PMF $f_X(\cdot)$ and support S_X. Let $g : \mathbb{R} \to \mathbb{R}$ and $Y = g(X)$. Then Y is a DRV with support $S_Y = \{g(x) : x \in S_X\}$ and PMF*

$$f_Y(y) = \begin{cases} \sum\limits_{x \in A_y} f_X(x) & \text{if } y \in S_Y \\ 0 & \text{otherwise,} \end{cases} \tag{2.10}$$

where $A_y = \{x \in S_X : g(x) = y\}$.

Proof: For $y \in g(S_X)$, $P(Y = y) = P(X \in A_y) = \sum_{x \in A_y} f_X(x)$. Also, $g(S_X)$ is atmost countable as S_X is countable. Now, we will try to show that $\sum_{y \in g(S_X)} P(Y = y) = 1$ or equivalently $\sum_{y \in g(S_X)} \sum_{x \in A_y} f_X(x) = 1$. Notice that

$$\bigcup_{y \in g(S_X)} A_y = S_X.$$

To see it first assume that $x \in S_X$ which implies that $y = g(x) \in g(S_X)$. Hence, for some $y \in g(S_X)$, $x \in A_y$. On the other hand, if $x \in \bigcup_{y \in g(S_X)} A_y$, then for some $y \in g(S_X)$, $x \in A_y \subset S_X$. Hence, $x \in S_X$.

Next notice that

$$A_{y_1} \bigcap A_{y_2} = \emptyset \quad \text{for } y_1 \neq y_2 \in g(S_X).$$

We can prove this claim by contradiction. Suppose that $A_{y_1} \bigcap A_{y_2} \neq \emptyset$. That means there exists at least one $x \in A_{y_1}$ and $x \in A_{y_2}$. Hence, $g(x) = y_1 = y_2$, which is a contradiction to the fact that $y_1 \neq y_2$. Thus,

$$\sum_{y \in g(S_X)} \sum_{x \in A_y} f_X(x) = \sum_{x \in S_X} f_X(x) = 1.$$

Hence, Y is a DRV with support $S_y = g(S_X)$ and PMF as given in (2.10) $\square$

Example 2.33. Let $X \sim Bin(n, p)$. Suppose that we are interested to find the distribution of $Y = n - X$. As X is a DRV, using the above theorem, Y is also DRV. Here, $S_X = \{0, 1, \ldots, n\} = S_Y$. For any $y \in S_Y$, $A_y = \{n - y\}$. Hence, the PMF of Y is

$$\begin{aligned}
f_Y(y) &= \begin{cases} f_X(n-y) & \text{if } y = 0, 1, \ldots, n \\ 0 & \text{otherwise} \end{cases} \\
&= \begin{cases} \binom{n}{n-y} p^{n-y} (1-p)^{n-n+y} & \text{if } y = 0, 1, \ldots, n \\ 0 & \text{otherwise} \end{cases} \\
&= \begin{cases} \binom{n}{y} (1-p)^y p^{n-y} & \text{if } y = 0, 1, \ldots, n \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}$$

Hence, $Y \sim Bin(n, 1-p)$. Note that $Y = n - X$ is the number of failures out of n trials. Therefore, this result is well justified. ||

Theorem 2.8. *Let X be a CRV with PDF $f_X(\cdot)$ and support S_X, which is an interval. Let $g : S_X \to \mathbb{R}$ be a differentiable function and either $g'(x) < 0$ for all $x \in S_X$ or $g'(x) > 0$ for all $x \in S_X$. Then the RV $Y = g(X)$ is a CRV with PDF*

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| & \textit{for } y \in g(S_X) \\ 0 & \textit{otherwise.} \end{cases}$$

Proof: The proof of the theorem can be done using the results of transformation of variable technique in integration. However, it is skipped here. $\square$

Example 2.34. Let $X \sim U(0, 1)$. Suppose that $g(x) = -\ln x$ for $x \in (0, 1)$. Also, the support of X is $S_X = (0, 1)$, which is an interval. Clearly, $g'(x) < 0$ for all $x \in (0, 1)$. The inverse of $g(\cdot)$ is $g^{-1}(y) = e^{-y}$ for all $y \in g(S_X) = (0, \infty)$. Hence, $Y = -\ln X$ is a CRV with PDF

$$\begin{aligned}
f_Y(y) &= \begin{cases} f_X\left(g^{-1}(y)\right) \left| \frac{d}{dy} g^{-1}(y) \right| & \text{if } y > 0 \\ 0 & \text{otherwise} \end{cases} \\
&= \begin{cases} e^{-y} & \text{if } y > 0 \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}$$

Therefore, $Y = -\ln X \sim Exp(1)$. ||

Example 2.35. Let $X \sim Exp(1)$. Suppose we are interested to find the distribution of $Y = X^2$. Here, $g(x) = x^2$ for $x \in S_X = (0, \infty)$. Also, $g'(x) = 2x > 0$ for all $x > 0$. Hence, $Y = X^2$ is a CRV. Note that $g^{-1}(y) = \sqrt{y}$. Thus, the PDF of Y is

$$f_Y(y) = \begin{cases} e^{-\sqrt{y}} \times \frac{1}{2\sqrt{y}} & \text{if } y > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Note that in this case the function $g(x) = x^2$ defined on $\mathbb{R}$ is not strictly monotone. However, we need to check only on the support of X and $g(\cdot)$ is strictly monotone on $(0, \infty)$. ||

Example 2.36. Let $X \sim N(0, 1)$. Suppose that we want to find the distribution of $Y = X^2$. Note that the support of X is $\mathbb{R}$ and $g'(x) = 2x$ does not take only positive or negative values on $\mathbb{R}$. Hence, we cannot use Theorem 2.8. However, we can use technique 1 to obtain the CDF of Y and then check the type of the RV Y. The CDF of Y is given by

$$F_Y(y) = \begin{cases} 0 & \text{if } y < 0 \\ 2\Phi\left(\sqrt{y}\right) - 1 & \text{if } y \geq 0. \end{cases}$$

It is easy to see that $F_Y(y) = \displaystyle\int_{-\infty}^{y} f_Y(t)dt$, where

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{y}}\phi\left(\sqrt{y}\right) & \text{if } y > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Thus, Y is a CRV. This example shows that even if some of the conditions of the Theorem 2.8 do not hold true, the RV Y could be CRV. On the other hand, consider the Example 2.31, where $g(x) = [x]$. This function also does not satisfy the strictly monotone condition of the Theorem 2.8 and we have seen that Y is a DRV. Thus, the conditions in the Theorem 2.8 are important and they are sufficient conditions, but not necessary. ||

2.5.3 Expectation of Function of RV

In this subsection, we will consider the expectation of a function of a RV. Let X be a RV and $g : S_X \to \mathbb{R}$ be a function. Then we have seen that $Y = g(X)$ is a RV. Naturally, we may want talk about the expectation of Y, if it exist. Note that one of the way to check if $E(Y)$ exists and then to compute it is to find the PMF or PDF of Y and then use the definition of expectation as we have done earlier. Let us consider an example.

Example 2.37. Let X be a DRV with PMF

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

Let $Y = X^2$. In Example 2.32, we have seen that the PMF of Y is

$$f(y) = \begin{cases} \frac{1}{7} & \text{if } y = 0 \\ \frac{2}{7} & \text{if } y = 1 \\ \frac{5}{14} & \text{if } y = 4 \\ \frac{4}{14} & \text{if } y = 9 \\ 0 & \text{otherwise.} \end{cases}$$

Thus, we can compute the expectation of Y based on this PMF. Note that in this case S_Y is finite, and hence, $E(Y)$ exists. The expectation of Y is given by

$$E(Y) = 0 \times \frac{1}{7} + 1 \times \frac{2}{7} + 4 \times \frac{5}{14} + 9 \times \frac{4}{14} = \frac{30}{7}.$$

However, we can compute the expectation without computing the PMF of Y. Notice that

$$\begin{aligned}
E(Y) &= 0 \times \frac{1}{7} + 1 \times \frac{2}{7} + 4 \times \frac{5}{14} + 9 \times \frac{4}{14} \\
&= 0^2 \times P(X = 0) + 1^2 \times P(X = 1) + (-1)^2 \times P(X = -1) \\
&\qquad + 2^2 \times P(X = 2) + (-2)^2 \times P(X = -2) + 3^2 \times P(X = 3) \\
&= \sum_{x \in S_X} x^2 P(X = x).
\end{aligned}$$

The benefit of this is that we do not need to compute the PMF of Y. We can use the PMF of X to compute the $E(Y) = E(X^2)$. ||

Theorem 2.9. *Let X be a DRV with PMF $f_X(\cdot)$ and support S_X. Let $g : \mathbb{R} \to \mathbb{R}$. Then*

$$E\left[g(X)\right] = \sum_{x \in S_X} g(x) f_X(x) \quad \textit{provided} \quad \sum_{x \in S_X} |g(x)| f_X(x) < \infty.$$

Proof: Using the Theorem 2.7,

$$E(Y) = \sum_{y \in S_Y} y \sum_{x \in A_y} f_X(x) = \sum_{y \in S_Y} \sum_{x \in A_y} g(x) f_X(x) = \sum_{x \in S_X} g(x) f_X(x).$$

□

Theorem 2.10. *Let X be a CRV with PDF $f_X(\cdot)$. Let $g : \mathbb{R} \to \mathbb{R}$. Then*

$$E\left[g(X)\right] = \int_{-\infty}^{\infty} g(x) f_X(x) dx \quad \textit{provided} \quad \int_{-\infty}^{\infty} |g(x)| f_X(x) dx < \infty.$$

Proof: The proof of this theorem is out of the scope of this course. □

As we have pointed out that $Y = g(X)$ may be a DRV or a CRV if X is a CRV. However, the previous theorem tells us that whatever the type of the RV Y is, we can use the theorem if X is a CRV.

Example 2.38 (Continuation of Example 2.32)**.** Let $X \sim U(0, 1)$. Then the expectation of $Y = -\ln X$ exists if $\int_0^1 |-\ln x| dx = -\int_0^1 \ln x dx < \infty$. Now,

$$\int_\varepsilon^1 \ln x dx = -1 - \varepsilon \ln \varepsilon + \varepsilon \to -1 \quad \text{as } \varepsilon \downarrow 0.$$

Hence, $E(Y)$ exists and is given by

$$E(Y) = -\int_0^1 \ln x dx = 1.$$

||

Example 2.39 (Continuation of Example 2.31). Let $X \sim Exp(1)$. The expectation of $Y = [X]$ exists if

$$\int_0^\infty |[x]|\, e^{-x} dx = \sum_{i=0}^\infty \int_i^{i+1} |[x]|\, e^{-x} dx = \sum_{i=0}^\infty i \left(e^{-i} - e^{-(i+1)}\right) < \infty.$$

Now, using the technique that is used in Example 2.23, we can show that $\sum_{i=0}^\infty ie^{-i} < \infty$, which implies that $E(Y)$ exists. The expectation of Y is given by

$$\int_0^\infty [x]e^{-x} dx = \sum_{i=0}^\infty \int_i^{i+1} [x]e^{-x} dx = \sum_{i=0}^\infty i \left(e^{-i} - e^{-(i+1)}\right) = \frac{e^{-1} - e^{-2}}{(1 - e^{-1})^2}.$$

||

Theorem 2.11. *Let X be a RV (either DRV or CRV).*

1. *Let $A \subset \mathbb{R}$. Then $E(I_A(X)) = P(X \in A)$.*
2. *Let $h_1(x) \leq h_2(x)$ for all $x \in \mathbb{R}$ be two real valued functions. Then $E[h_1(X)] \leq E[h_2(X)]$, provided all the expectations exist.*
3. *Let $a < b$ be two fixed real numbers such that $S_X \subset [a, b]$. Then $a \leq E(X) \leq b$, provided the expectation exists.*
4. *$E(a + bX) = a + bE(X)$, where a and b are two fixed real numbers.*
5. *Let $h_1(\cdot), \ldots, h_p(\cdot)$ be real valued functions of real numbers such that $E(h_i(X))$ exists for all $i = 1, 2, \ldots, p$, then*

$$E\left(\sum_{i=1}^p h_i(X)\right) = \sum_{i=1}^p E\left(h_i(X)\right).$$

Proof: The proof this theorem is straight forward and therefore, left as an exercise. □

Definition 2.9 (Raw Moment). *For $r = 1, 2, \ldots$, $\mu_r = E(X^r)$ is called rth raw moment of X, if the expectation exists.*

Definition 2.10 (Central Moment). *$\mu_r' = E\left[(X - E(X))^r\right]$ is called rth central moment of X, if the expectations exist.*

Definition 2.11 (Variance). *$\mu_2' = E\left[(X - E(X))^2\right]$ is called variance of X when it exists and is denoted by $Var(X)$. Note that $Var(X) = E(X^2) - (E(X))^2$.*

Theorem 2.12. *$E\left(X - E(X)\right)^2 \leq E\left(X - a\right)^2$ for all $a \in \mathbb{R}$.*

Proof: Let us denote $\mu = E(X)$. Then

$$\begin{aligned}
E\left(X - a\right)^2 &= E\left(X - \mu + \mu - a\right)^2 \\
&= E\left(X - \mu\right)^2 + E\left(\mu - a\right)^2 + 2E\left[(X - \mu)(\mu - a)\right] \\
&= E\left(X - \mu\right)^2 + \left(\mu - a\right)^2 + 2(\mu - a)E\left(X - \mu\right) \\
&= E\left(X - \mu\right)^2 + \left(\mu - a\right)^2 \\
&\geq E\left(X - \mu\right)^2.
\end{aligned}$$

The third equality is due to the fact that $\mu - a$ is a constant. The fourth equality holds true as $E(X-\mu) = E(X) - \mu = 0$. Finally, the last inequality is due to the fact that $(\mu - a)^2 \geq 0$. Note that $E(X-a)^2 = E(X-\mu)^2$ if and only if $(\mu - a)^2 = 0$, which implies and is implied by $a = \mu$. □

Note that $E(X-a)^2$ can be regarded as the average error if we use a instead of X. Thus, if we do not know the value of X and need to provide a guess for the same, then it is safer to use $E(X)$ as a guess than any other value, as the average error is minimum for $E(X)$. Therefore, $E(X)$ is the "best estimate" of X.

Definition 2.12 (Moment Generating Function)**.** *The moment generating function (MGF) of a RV X is defined by*

$$M_X(t) = E\left(e^{tX}\right)$$

provided there exists a real number $a > 0$ such that the expectation exists for all $t \in (-a,\, a)$.

Example 2.40. Let $X \sim Bin(n,\, p)$. Then

$$E\left(e^{tX}\right) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=0}^{n} \binom{n}{x} \left(pe^t\right)^x (1-p)^{n-x} = (1-p+pe^t)^n$$

for all $t \in \mathbb{R}$. As the sum is a finite sum, the sum converges for all $t \in \mathbb{R}$. As $E\left(e^{tX}\right)$ exists for all $t \in \mathbb{R}$, we can take any value of $a > 0$. However, we write that MGF exists for all $t \in \mathbb{R}$ in such situations. Therefore, the MGF of X is given by

$$M_X(t) = (1-p+pe^t)^n$$

for all $t \in \mathbb{R}$. ||

Example 2.41. Let $X \sim Exp(\lambda)$. Then

$$E\left(e^{tX}\right) = \lambda \int_0^\infty e^{tx} e^{-\lambda x} dx = \lambda \int_0^\infty e^{-(\lambda - t)x} dx.$$

This integration converges if and only if $t < \lambda$. Hence,

$$E\left(e^{tX}\right) = \left(1 - \frac{t}{\lambda}\right)^{-1}$$

for all $t < \lambda$. Clearly, we can take $a = \lambda > 0$ and $E\left(e^{tX}\right)$ exists for all $t \in (-\lambda,\, \lambda)$. Thus, MGF of X exists and is given by

$$M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-1}$$

for all $t < \lambda$. Notice that the range of t, for which $E\left(e^{tX}\right)$ exists, is important and need to specify unambiguously. ||

Example 2.42. Let $X \sim N(\mu,\, \sigma^2)$. Then

$$
\begin{aligned}
E\left(e^{tX}\right) &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2\sigma^2}\left(x^2 - 2\mu x + \mu^2 - 2\sigma^2 tx\right)\right] dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2\sigma^2}\left(x^2 - 2\left(\mu + t\sigma^2\right)x + \left(\mu + t\sigma^2\right)^2 - \left(\mu + t\sigma^2\right)^2 + \mu^2\right)\right] dx \\
&= \exp\left[-\frac{1}{2\sigma^2}\left(\mu^2 - \mu^2 - t^2\sigma^4 - 2\mu\sigma^2 t\right)\right] \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}\left(x - \mu - t\sigma^2\right)^2\right] dx \\
&= \exp\left[\mu t + \frac{1}{2}t^2\sigma^2\right]
\end{aligned}
$$

for all $t \in \mathbb{R}$. Here, the last equality follows from the fact that the integrand (in the last but one line) is the PDF of a $N(\mu + t\sigma^2,\, \sigma^2)$ distribution. Therefore, the MGF of X is

$$M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

for all $t \in \mathbb{R}$. ||

Theorem 2.13. *If the MGF $M_X(t)$ exist for $t \in (-a,\, a)$ for some $a > 0$, the derivatives of all order exist at $t = 0$ and*

$$E\left(X^k\right) = \left.\frac{d^k}{dt^k} M_X(t)\right|_{t=0}$$

for all positive integer k.

Proof: Out of the scope of this course. □

Though the proof of the previous theorem is out of scope, it is quite important. Note that this theorem tells us that if MGF exists then all raw moment, and hence, all central moments exists. Also, we can compute raw moments using the MGF, and hence, the name MGF. Moreover, we need to take rth derivative of MGF at zero to compute rth raw moments. Also, for the main theorem of the next subsection, we need the condition that two MGFs have to be equal in a neighborhood of zero. These are the reasons of keeping the condition that the MGF exists if $E\left(e^{tX}\right)$ exists in a neighborhood of zero in the definition of MGF. Note that if $E\left(e^{tX}\right)$ exists only on an interval not including zero, then $E\left(e^{tX}\right)$ is of no use.

2.5.4 Technique 3

Definition 2.13 (Same in Distribution)**.** *Two RVs X and Y are said to be same in distribution (denoted by $X \stackrel{d}{=} Y$) if $F_X(x) = F_Y(x)$ for all $x \in \mathbb{R}$.*

Theorem 2.14. *Let X and Y be two RVs having MGFs $M_X(\cdot)$ and $M_Y(\cdot)$, respectively. Suppose that there exists a positive real number a such that $M_X(t) = M_Y(t)$ for all $t \in (-a,\, a)$. Then X and Y are same in distribution.*

Proof: The proof of this theorem is out of scope of this course. □

In general, for some function $E(f(X)) = E(f(Y))$ does not imply that $X \stackrel{d}{=} Y$. However, MGF is very special in this respect. Note that this theorem can be use to find the distribution of a function of a RV as illustrated in the following example.

Example 2.43. Let $X \sim N(\mu,\, \sigma^2)$. Suppose we are interested to find the distribution of $Y = a + bX$, which is a linear combination of X. Assume that $b \neq 0$. Otherwise $Y = a$ with probability one. First, let us try to find the MGF of Y. Note that

$$E\left(e^{tY}\right) = E\left(e^{t(a+bX)}\right) = e^{ta}E\left(e^{tbX}\right) = e^{ta}M_X(tb).$$

Now, from the Example 2.42, $M_X(t)$ exists for all $t \in \mathbb{R}$. Hence,

$$E(e^{tY}) = e^{ta}e^{\mu bt + \frac{1}{2}b^2t^2\sigma^2} = e^{(a+b\mu)t + \frac{1}{2}(b\sigma)^2t^2}$$

for all $t \in \mathbb{R}$. Suppose that $Z \sim N(a + b\mu,\, b^2\sigma^2)$. Then the MGF of Z is

$$M_Z(t) = e^{(a+b\mu)t + \frac{1}{2}b^2\sigma^2t^2}$$

for all $t \in \mathbb{R}$. Thus, the MGFs of Y and Z are same for all $t \in \mathbb{R}$. Thus, $Y \stackrel{d}{=} Z \sim N(a + b\mu,\, b^2\sigma^2)$. Note that to use the technique 3, we need to identify the MGF of Y. ||

2.6 Moment Inequality

In this section we will discuss some inequalities involving probability and moment of a random variables. These inequalities give upper bound of probability of the events of the form $|X| \geq c$ in terms of moment of X. Obviously, if the upper bound is greater than or equal to one, then these inequality do not provide any extra information. Therefore, these inequalities are meaningful if the upper bond turn out to be less than one.

Theorem 2.15. *Let X be a RV and $g : [0,\, \infty) \to [0,\, \infty)$ be a non-decreasing function such that $E(g(|X|))$ is finite. Then for any $c > 0$ with $g(c) > 0$,*

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

Proof: Notice that for all $x \in \mathbb{R}$,

$$g(c)I_{(-\infty,\, c] \cup [c,\, \infty)}(x) \leq g(|x|), \tag{2.11}$$

where

$$I_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

To see it, first let $|x| \geq c$. Then the left hand side is $g(c)$ and the right hand side is $g(|x|)$. As $g(\cdot)$ is a non-decreasing function, (2.11) holds true for $|x| \geq c$. For $-c < x < c$, the left hand side is zero and right hand side is $g(|x|)$. As the range of $g(\cdot)$ in non-negative part of real line, (2.11) holds true for $-c < x < c$. Thus, (2.11) holds true for all $x \in \mathbb{R}$. Now, using Theorem 2.11,

$$E\left[g(c)I_{(-\infty,\, c] \cup [c,\, \infty)}(X)\right] \leq E\left(g\left(|X|\right)\right) \implies g(c)P\left(|X| \geq c\right) \leq E\left(g\left(|X|\right)\right).$$

Now, as $g(c) > 0$, proof of the theorem completes. □

Corollary 2.1 (Markov Inequality). *Let X be a RV with $E\left(|X|^r\right) < \infty$ for some $r > 0$. Then for any $c > 0$,*

$$P(|X| \geq c) \leq \frac{E\left(|X|^r\right)}{c^r}.$$

Proof: Take $g(x) = x^r$ in the previous theorem. $\square$

Corollary 2.2 (Chebyshev Inequality). *Let X be a RV with $E(X^2) < \infty$. Let us denote $\mu = E(X)$ and $\sigma^2 = Var(X)$. Then for any $k > 0$,*

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

Proof: Let $Y = X - \mu$. Taking $r = 2$ and $c = k$ in the previous corollary, we get

$$P\left(|Y| \geq k\right) \leq \frac{E\left(|Y|^2\right)}{k^2} \implies P\left(|X - \mu| \leq k\right) \geq \frac{E\left((X-\mu)^2\right)}{k^2} = \frac{\sigma^2}{k^2}.$$

$\square$

Example 2.44. Let X be a DRV with PMF

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

Then $E(X) = 0$ and $E(X^2) = 1/4$, which implies that $Var(X) = 1/4$. Now, using Chebyshev inequality and taking $k = 1$, $P\left(|X| \geq 1\right) \leq \frac{1}{4}$. On the other hand, using PMF of X, $P\left(|X| \geq 1\right) = \frac{1}{4}$. This example shows that in general we do not have a better upper bound of $P\left(|X| \geq k\right)$ than the upper bound provided by Chebyshev inequality. Thus, the Chebyshev inequality is tight. Of course, if we have extra information on the RV X, then it may be possible to have a better upper bound of $P\left(|X| \geq k\right)$ compared to that is given by Chebyshev inequality. ||

2.A Gamma Integral

Lemma 2.1. *The improper integral*

$$\int_0^\infty t^{n-1}e^{-t}dt$$

converges absolutely for all $n \in \mathbb{N}$.

Proof: As $\lim\limits_{t \to \infty} t^{n-1}e^{-t/2} = 0$ for all $n \in \mathbb{N}$, there exists $t_0 > 0$ such that $t^{n-1} \leq e^{t/2}$ for all $t > t_0$. Now,

$$\int_0^\infty t^{n-1}e^{-t}dt = \int_0^{t_0} t^{n-1}e^{-t}dt + \int_{t_0}^\infty t^{n-1}e^{-t}dt,$$

where the first integral is a proper integral and the second is an improper integral. The second integral converges as

$$\int_{t_0}^\infty t^{n-1}e^{-t}dt \leq \int_{t_0}^\infty e^{-t/2}dt = \lim_{A \to \infty} \int_{t_0}^A e^{-t/2}dt = 2e^{-\frac{t_0}{2}} < \infty.$$

$\square$

Lemma 2.2. *The improper integral*

$$\int_0^\infty t^{\alpha-1}e^{-t}dt$$

converges absolutely for all $\alpha \geq 1$.

Proof: For $t > 1$, $t^{\alpha-1}e^{-t} \leq t^{[\alpha]}e^{-t}$, where $[x]$ denotes the largest integer not exceeding $x \geq 0$. Now,

$$\int_0^\infty t^{\alpha-1}e^{-t}dt = \int_0^1 t^{\alpha-1}e^{-t}dt + \int_1^\infty t^{\alpha-1}e^{-t}dt,$$

where the first integral is a proper integral and the second is an improper integral. The second integral converges as

$$\int_1^\infty t^{\alpha-1}e^{-t}dt \leq \int_1^\infty t^{[\alpha]}e^{-t/2}dt < \int_0^\infty t^{[\alpha]}e^{-t}dt,$$

which converges by the Lemma 2.1. □

Lemma 2.3. *The improper integral*

$$\int_0^\infty t^{\alpha-1}e^{-t}dt$$

converges absolutely for all $0 < \alpha < 1$.

Proof: Notice that $t^{\alpha-1}e^{-t} \leq e^{-t}$ for $t > 1$. Now,

$$\int_0^\infty t^{\alpha-1}e^{-t}dt = \int_0^1 t^{\alpha-1}e^{-t}dt + \int_1^\infty t^{\alpha-1}e^{-t}dt,$$

where both the integrals on the right hand side are improper. Consider the first integral. Notice that $e^{1+t} \geq 1 \implies e^{-t} \leq e$ for $t \geq 0$. Hence,

$$\int_0^1 t^{\alpha-1}e^{-t}dt \leq e\int_0^1 t^{\alpha-1} < \infty.$$

Now, consider the second integral.

$$\int_1^\infty t^{\alpha-1}e^{-t}dt \leq \int_1^\infty e^{-t}dt < \infty.$$

Thus, both the integral converges, which proves the Lemma. □

Theorem 2.16. *The improper integral*

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}dt$$

converges absolutely for all $\alpha > 0$.

Proof: Combining Lemmas 2.1, 2.2, and 2.3, the proof of the theorem is immediate. □

Theorem 2.17. *The functional equation* $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ *holds for* $\alpha > 0$.

Proof:

$$\begin{aligned}\Gamma(\alpha+1) &= \int_0^\infty t^\alpha e^{-t} dt \\ &= \left[t^\alpha \int e^{-t} dt + \int \alpha t^{\alpha-1} e^{-t} dt\right]_0^\infty \\ &= \alpha \int_0^\infty t^{\alpha-1} e^{-t} dt \\ &= \alpha\Gamma(\alpha).\end{aligned}$$

□

Theorem 2.18. $\Gamma(n+1) = n!$ *for* $n = 1, 2, \ldots$.

Proof: It is very easy to see (using integration by parts) that the theorem holds for $n = 1$. Now, using the previous theorem, proof of current theorem is immediate. □

2.B Beta Integral

Theorem 2.19. *For* $\alpha > 0$ *and* $\beta > 0$*, the integral*

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx$$

converges.

Proof: Note that for $\alpha \geq 1$ and $\beta \geq 1$, the integral is a proper integral. The integral is improper if $0 < \alpha < 1$ and/or $0 < \beta < 1$. If $\alpha < 1$, $f(x) \to \infty$ as $x \downarrow 0$, where $f(\cdot)$ is the integrand. Similarly, $f(x) \to \infty$ as $x \uparrow 1$ if $\beta < 1$. First notice that

$$\int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx = \int_0^{1/2} x^{\alpha-1}(1-x)^{\beta-1} dx + \int_{1/2}^1 x^{\alpha-1}(1-x)^{\beta-1} dx.$$

The first and second integrals on the right hand side are improper if $\alpha < 1$ and $\beta < 1$, respectively. Let us first prove that the first integral converges. Note that for $0 < x < 1/2$, $(1-x)^{\beta-1} \leq A(\beta)$, where

$$A(\beta) = \begin{cases} 1 & \text{if } \beta \geq 1 \\ \left(\frac{1}{2}\right)^{\beta-1} & \text{if } 0 < \beta < 1. \end{cases}$$

Thus,

$$\int_0^{1/2} x^{\alpha-1}(1-x)^{\beta-1} dx \leq A(\beta) \int_0^{1/2} x^{\alpha-1} dx < \infty.$$

Similarly, we can prove that the second integral converges by deducing an upper bound of $x^{\alpha-1}$ for $1/2 < x < 1$. □

Theorem 2.20. *For* $\alpha > 0$ *and* $\beta > 0$*,*

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

Proof:

$$\begin{aligned}
\Gamma(\alpha+\beta)B(\alpha,\,\beta) &= \left(\int_0^\infty x^{\alpha+\beta-1}e^{-x}dx\right)\left(\int_0^1 y^{\alpha-1}(1-y)^{\beta-1}dy\right)\\
&= \int_0^\infty \int_0^1 x^{\alpha+\beta-1}e^{-x}y^{\alpha-1}(1-y)^{\beta-1}dydx\\
&= \int_0^\infty \int_0^\infty z_1^{\alpha-1}z_2^{\beta-1}e^{-(z_1+z_2)}dz_1dz_2, \quad \text{taking } z_1 = xy \text{ and } z_2 = (1-y)x\\
&= \Gamma(\alpha)\Gamma(\beta).
\end{aligned}$$

□

Corollary 2.3. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

Proof: Using the last theorem,

$$\left(\Gamma\!\left(\frac{1}{2}\right)\right)^2 = B\left(\frac{1}{2},\frac{1}{2}\right) = \int_0^1 \frac{1}{\sqrt{x(1-x)}}dx = \pi.$$

As $\Gamma(1/2)$ is integration of a positive function, $\Gamma(1/2) > 0$. Hence, $\Gamma(1/2) = \sqrt{\pi}$. □