Probability Theory and Random Processes (MA 225)

Class Notes

September – November, 2020

Instructor
Ayon Ganguly
Department of Mathematics
IIT Guwahati

Contents

Chapter 3

Jointly Distributed Random Variables

In the previous chapter, we studied RV and associated concepts. One of the mains use of the random variable is to model numerical characteristic of a natural phenomena. For example, we may assume that the RV X denote the income of a household. Now, assume that Y denotes the spending of the household. Then $Z = X - Y$ is a RV and it is the savings of the household. Clearly, if we know the values of any of two RVs among X, Y, and Z, the value of other one is known. In such situation, we may want to study the relationship between X and Z. Clearly, the concepts of the previous chapter are not sufficient. Using the tools of the previous chapter, we can study X and Z separately, but not jointly. For example we cannot answer the question: Is savings increases with income? To answer it, we need to know the probability of joint occurrence of events $X \leq x$ and $Z \leq z$ for different values of x and z. In this chapter, we will study the concepts relating to the joint occurrence of multiple RVs. There are plenty of examples, where we need to consider multiple RVs. These examples include height and weight of a person, pollution level and blood pressure, lifetime of a product and its cause of failure, etc.

3.1 Random Vector

Definition 3.1 (Random Vector)**.** *A function* $\boldsymbol{X} : \mathcal{S} \to \mathbb{R}^n$ *is called a random vector.*

Clearly, random vector is a generalization of RV. Note that as $\boldsymbol{X}$ is a function from $\mathcal{S}$ to $\mathbb{R}^n$, $\boldsymbol{X}(\omega)$ can be written as $(X_1(\omega), X_2(\omega), \ldots, X_n(\omega))$, where $X_i : \mathcal{S} \to \mathbb{R}$ for all $i = 1, 2, \ldots, n$. Thus, X_i is a RV for all $i = 1, 2, \ldots, n$. Therefore, each component of a random vector is a RV and we will write $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$.

Definition 3.2 (Joint CDF)**.** *For any random vector* $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$*, the joint cumulative distribution function (JCDF) is defined by*

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = P\left(X_1 \leq x_1, \ldots, X_n \leq x_n\right),$$

for all $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$*. Here*

$$\{X_1 \leq x_1, \ldots, X_n \leq x_n\} = \{X_1 \leq x_1\} \cap \{X_2 \leq x_2\} \cap \ldots \cap \{X_n \leq x_n\}.$$

Now, onward, most of the definitions, theorems, results will be presented for $n = 2$. Thus, we will use $\boldsymbol{X} = (X_1, X_2)$ or $\boldsymbol{X} = (X, Y)$. This is for simplicity of the expressions. However, most of the definitions, theorems, results can be extended for any general value of n. For $n = 2$, the JCDF at the point (x, y) is the probability that the random vector (X, Y) belongs to the shaded region of the Figure 3.1.

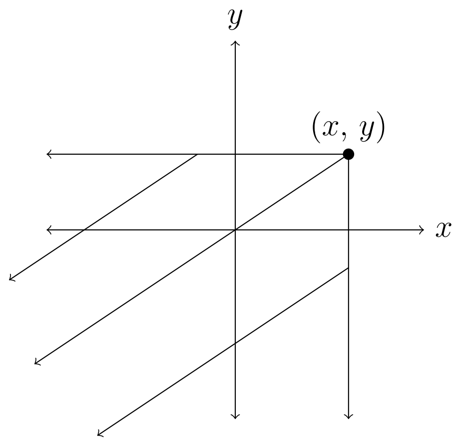


Figure 3.1: CDF is the probability that $(X,\, Y)$ in the marked region

Theorem 3.1. *Let $\boldsymbol{X} = (X,\, Y)$ be a random vector with JCDF $F_{X,Y}(\cdot,\, \cdot)$. Then the CDF of X is given by $F_X(x) = \lim\limits_{y \to \infty} F_{X,Y}(x,\, y)$ for all $x \in \mathbb{R}$. Similarly, the CDF of Y is given by $F_Y(y) = \lim\limits_{x \to \infty} F_{X,Y}(x,\, y)$ for all $y \in \mathbb{R}$.*

Proof: Fix $x \in \mathbb{R}$. Let $\{y_n\}_{n \geq 1}$ be an increasing sequence of real numbers such that $y_n \to \infty$ as $n \to \infty$. Let us define

$$A_n = \{\omega \in \mathcal{S} : X(\omega) \leq x,\, Y(\omega) \leq y_n\}$$

for $n = 1,\, 2,\, 3,\, \ldots$. Clearly, $\{A_n\}_{n \geq 1}$ is an increasing sequence of events, and hence,

$$A = \lim_{n \to \infty} A_n = \cup_{n=1}^{\infty} A_n = \{\omega \in \mathcal{S} : X(\omega) \leq x\}.$$

Now,

$$\lim_{n \to \infty} F_{X,Y}(x,\, y_n) = \lim_{n \to \infty} P(A_n) = P\left(\lim_{n \to \infty} A_n\right) = P(A) = F_X(x).$$

This shows that for any increasing sequence of real numbers $\{y_n\}_{n \geq 1}$ with $y_n \to \infty$ as $n \to \infty$, $\lim_{n \to \infty} F_{X,Y}(x,\, y_n) = F_X(x)$. Thus, $\lim_{y \to \infty} F_{X,Y}(x,\, y) = F_X(x)$ for each fixed $x \in \mathbb{R}$.

Similarly, one can prove that $\lim_{x \to \infty} F_{X,Y}(x,\, y) = F_Y(y)$ for each fixed $y \in \mathbb{R}$. □

Remark 3.1. The previous theorem can be extended for more than two RVs. For example, let $\boldsymbol{X} = (X,\, Y,\, Z)$. In this case we can find CDFs of X, Y, and Z using the following formulas.

$$\begin{aligned}
F_X(x) &= \lim_{y \to \infty} \lim_{z \to \infty} F_{X,Y,Z}(x,\, y,\, z) = \lim_{z \to \infty} \lim_{y \to \infty} F_{X,Y,Z}(x,\, y,\, z) \quad \text{for all } x \in \mathbb{R}, \\
F_Y(y) &= \lim_{x \to \infty} \lim_{z \to \infty} F_{X,Y,Z}(x,\, y,\, z) = \lim_{z \to \infty} \lim_{x \to \infty} F_{X,Y,Z}(x,\, y,\, z) \quad \text{for all } y \in \mathbb{R}, \\
F_Z(z) &= \lim_{x \to \infty} \lim_{y \to \infty} F_{X,Y,Z}(x,\, y,\, z) = \lim_{y \to \infty} \lim_{x \to \infty} F_{X,Y,Z}(x,\, y,\, z) \quad \text{for all } z \in \mathbb{R}.
\end{aligned}$$

We can find the JCDFs of $(X,\, Y)$, $(X,\, Z)$, and $(Y,\, Z)$ as

$$\begin{aligned}
F_{X,Y}(x,\, y) &= \lim_{z \to \infty} F_{X,Y,Z}(x,\, y,\, z) \quad \text{for all } (x,\, y) \in \mathbb{R}^2, \\
F_{X,Z}(x,\, z) &= \lim_{y \to \infty} F_{X,Y,Z}(x,\, y,\, z) \quad \text{for all } (x,\, z) \in \mathbb{R}^2, \\
F_{Y,Z}(y,\, z) &= \lim_{x \to \infty} F_{X,Y,Z}(x,\, y,\, z) \quad \text{for all } (y,\, z) \in \mathbb{R}^2.
\end{aligned}$$

Let $\boldsymbol{X} = (X_1, \ldots, X_n)$. Let $A = \{i_1, i_2, \ldots, i_k\} \subset \{1, 2, \ldots, n\}$. If we want to find the JCDF of $(X_{i_1}, X_{i_2}, \ldots, X_{i_k})$, we need to take limit (tends to infinity) with respect to all the components that are not present in A. †

In the context of random vector, the JCDF of a subset is called marginal CDF. Thus, if $\boldsymbol{X} = (X, Y, Z)$, the CDF of X is called marginal CDF of X. Similarly the JCDF of (X, Y) is called marginal CDF of (X, Y).

Theorem 3.2 (Properties of JCDF). *Let $\boldsymbol{X} = (X, Y)$ be a random vector with JCDF $F_{X,Y}(\cdot, \cdot)$. Then*

1. $\lim\limits_{x\to\infty} \lim\limits_{y\to\infty} F_{X,Y}(x, y) = 1$.
2. $\lim\limits_{x\to-\infty} F_{X,Y}(x, y) = 0$ *for all* $y \in \mathbb{R}$.
3. $\lim\limits_{y\to-\infty} F_{X,Y}(x, y) = 0$ *for all* $x \in \mathbb{R}$.
4. $F_{X,Y}(\cdot, \cdot)$ *is right continuous in each argument keeping other fixed.*
5. *For* $-\infty < a_1 < b_1 < \infty$ *and* $-\infty < a_2 < b_2 < \infty$,

$$F_{X,Y}(b_1, b_2) - F_{X,Y}(b_1, a_2) - F_{X,Y}(a_1, b_2) + F_{X,Y}(a_1, a_2) \geq 0.$$

Proof: The proof of this theorem is similar to that of Theorem 2.1 with standard modification for 2-dimensional functions. Therefore, the proof is skipped here. □

Though we are skipping the proof the previous theorem, let us make a comparison between the properties (that are presented in the Theorem 2.1) of CDF of a RV and that of JCDF of a 2-dimensional random vector. The Property 1 in the Theorem 2.1 states that $F_X(\cdot)$ is non-decreasing. This can be alternatively written as $F_X(x_2) - F_X(x_1) = P(x_1 < X \leq x_2) \geq 0$ for all $x_1 < x_2$. Thus, the non-decreasing property is a consequence of the fact that the probability that a RV is in an interval must be non-negative. Now, a natural extension of an interval in one-dimension is a rectangle in two-dimension. Thus, the equivalent property should be based on the fact that the probability that a 2-dimensional random vector in a rectangle must be non-negative. Let $a_1 < b_1$ and $a_2 < b_2$ be four real numbers. Then the points (a_1, a_2), (a_1, b_2), (b_1, a_2), and (b_1, b_2) forms vertices of the rectangle $(a_1, b_1] \times (a_2, b_2]$. Now,

$$\begin{aligned} P((X, Y) \in (a_1, b_1] \times (a_2, b_2]) &= P(a_1 < X \leq b_1, a_2 < Y \leq b_2) \\ &= F_{X,Y}(b_1, b_2) - F_{X,Y}(b_1, a_2) - F_{X,Y}(a_1, b_2) + F_{X,Y}(a_1, a_2). \end{aligned}$$

Thus, the fact that the probability that a random vector in a rectangle is non-negative gives the Property 5 of Theorem 3.2.

Property 2 of Theorem 2.1 states that $\lim_{x\to\infty} F_X(x) = 1$. This is intuitively tells that if we cover the whole $\mathbb{R}$, then the probability is one. Similarly, for a 2-dimensional random vector if the whole $\mathbb{R}^2$ is covered, the probability is one and we have Property 1 of Theorem 3.2.

Property 3 of Theorem 2.1 states that $\lim_{x\to-\infty} F_X(x) = 0$. Loosely speaking, $\{X \leq x\}$ becomes $\emptyset$ for x tends to $-\infty$. For 2-dimensional random vector, if one of the components tends to $-\infty$, the set becomes empty. If $x \to -\infty$, then $\{X \leq x\}$ becomes $\emptyset$, and hence, $\{X \leq x, Y \leq y\}$ becomes $\emptyset$. Similarly, as $y \to -\infty$, $\{X \leq x, Y \leq y\}$ becomes $\emptyset$. Thus, we have Properties 2 and 3 of Theorem 3.2. Note that for the first property of Theorem 3.2, both the components need to tend to ∞. However, for the Properties 2 and 3, if any of the components tends to $-\infty$ keeping other fixed, then JCDF tends to zero. Property 4 is a straight forward extension of Property 4 of the Theorem 2.1.

Theorem 3.3. *Let $G : \mathbb{R}^2 \to \mathbb{R}$ be a function satisfying conditions 1–5 of the Theorem 3.2. Then G is a JCDF of some 2-dimensional random vector.*

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

This theorem can be used to check if a function is a JCDF or not. Theorems 3.2 and 3.3 can be extended for random vector having more than two components. However, writing the property (5) involves complicated expressions.

3.2 Discrete Random Vector

Definition 3.3 (Discrete Random Vector). *A random vector (X, Y) is said to have a discrete distribution if there exists an atmost countable set $S_{X,Y} \subset \mathbb{R}^2$ such that $P((X, Y) = (x, y)) = P(X = x, Y = y) > 0$ for all $(x, y) \in S_{X,Y}$ and $P((X, Y) \in S_{X,Y}) = 1$. $S_{X,Y}$ is called the support of (X, Y).*

Definition 3.4 (Joint PMF). *Let (X, Y) be a discrete random vector with support $S_{X,Y}$. Define a function $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ by*

$$f_{X,Y}(x, y) = \begin{cases} P(X = x, Y = y) & \text{if } (x, y) \in S_{X,Y} \\ 0 & \text{otherwise.} \end{cases}$$

The function $f_{X,Y}$ is called joint probability mass function (JPMF) of the discrete random vector (X, Y).

Note that discrete random vector is a straight forward extension of DRV. In case of DRV, we need to find an atmost countable set S_X in $\mathbb{R}$ such that $P(X \in S_X) = 1$. For a 2-dimensional discrete random vector $S_{X,Y}$ is atmost countable and a subset of $\mathbb{R}^2$ such that $P((X, Y) \in S_{X,Y}) = 1$. Similarly, the definition of JPMF is also a natural extension of PMF of a DRV. These definitions can be easily extended for more than two dimensional random vectors.

Theorem 3.4 (Properties of JPMF). *Let (X, Y) be a discrete random vector with JPMF $f_{X,Y}(\cdot, \cdot)$ and support $S_{X,Y}$. Then*

1. *$f_{X,Y}(x, y) \geq 0$ for $(x, y) \in \mathbb{R}^2$.*
2. *$\displaystyle\sum_{(x,y) \in S_{X,Y}} f_{X,Y}(x, y) = 1$.*

Proof: The proof of the theorem is straight forward form the definitions of discrete random vector and JPMF. □

Theorem 3.5. *If a function $g : \mathbb{R}^2 \to \mathbb{R}$ satisfy Properties 1 and 2 above for the atmost countable set $D = \{(x, y) \in \mathbb{R}^2 : g(x, y) > 0\}$ in place of $S_{X,Y}$, then g is JPMF of some 2-dimensional discrete random vector.*

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

Theorem 3.5 can be used to check if a function is JPMF or not. Again, Theorems 3.4 and 3.5 can be extended for more than 2-dimensional discrete random vector.

Theorem 3.6 (Marginal PMF from JPMF). *Let (X, Y) be a discrete random vector with JPMF $f_{X,Y}(\cdot, \cdot)$ and support $S_{X,Y}$. Then X and Y are DRVs. The PMF of X is*

$$f_X(x) = \sum_{(x, y) \in S_{X,Y}} f_{X,Y}(x, y) \text{ for all fixed } x \in \mathbb{R}. \tag{3.1}$$

The PMF of Y is given by

$$f_Y(y) = \sum_{(x, y) \in S_{X,Y}} f_{X,Y}(x, y) \text{ for all fixed } y \in \mathbb{R}. \tag{3.2}$$

In this context, $f_X(\cdot)$ and $f_Y(\cdot)$ are called marginal PMF of X and marginal PMF of Y, respectively.

Proof: Define the set

$$D = \{x \in \mathbb{R} : (x, y) \in S_{X,Y} \text{ for some } y \in \mathbb{R}\}.$$

As, $S_{X,Y}$ is atmost countable, D is also atmost countable. Fix $x_0 \in D$. Consider $(x_0, y) \in S_{X,Y}$ and $(x_0, y') \in S_{X,Y}$ such that $y \neq y'$. Then the events

$$\{X = x_0, Y = y\} \text{ and } \{X = x_0, Y = y'\}$$

are disjoint. Now, using theorem of total probability (Theorem 1.16), $P(X = x_0)$ can be found by taking sum over all the points in $S_{X,Y}$ whose first component is x_0. Thus,

$$P(X = x_0) = \sum_{(x_0, y) \in S_{X,Y}} P(X = x, Y = y) = \sum_{(x_0, y) \in S_{X,Y}} f_{X,Y}(x, y),$$

and

$$\sum_{x \in D} P(X = x) = \sum_{x \in D} \sum_{(x, y) \in S_{X,Y}} f_{X,Y}(x, y) = \sum_{(x, y) \in S_{X,Y}} f_{X,Y}(x, y) = 1.$$

Hence, X is a DRV with PMF given by (3.1). Similarly we can prove that Y is also a DRV with PMF given by (3.2) □

Example 3.1. Let (X, Y) be a discrete random vector with JPMF

$$f(x, y) = \begin{cases} cy & \text{if } x = 1, 2, \ldots, n; \, y = 1, 2, \ldots, n \\ 0 & \text{otherwise,} \end{cases}$$

where c is a constant. We can find the value of c based on the properties of JPMF. If $f(\cdot,\, \cdot)$ have to be a JPMF, then $f(x,\, y) \geq 0$ for all $(x,\, y) \in \mathbb{R}^2$, which implies that $c \geq 0$. Also,

$$\sum_{x=1}^{n} \sum_{y=1}^{n} f(x,\, y) = 1 \implies c = \frac{2}{n^2(n+1)}.$$

Thus, the JPMF of $(X,\, Y)$ is given by

$$f(x,\, y) = \begin{cases} \frac{2y}{n^2(n+1)} & \text{if } x = 1,\, 2,\, \ldots,\, n;\; y = 1,\, 2,\, \ldots,\, n \\ 0 & \text{otherwise.} \end{cases}$$

We can also find the marginal PMF of X as follows: Fix $x \in \{1,\, 2,\, \ldots,\, n\}$. Then

$$P\left(X = x\right) = \sum_{(x,\, y) \in S_{X,\, Y}} f(x,\, y) = \sum_{y=1}^{n} cy = \frac{1}{n}.$$

Thus, the marginal PMF of X is given by

$$f_X(x) = \begin{cases} \frac{1}{n} & \text{if } x = 1,\, 2,\, \ldots,\, n \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, we can find the marginal PMF of Y. I leave it as an exercise. ||

Example 3.2. Let $(X,\, Y)$ be a discrete random vector with JPMF

$$f(x,\, y) = \begin{cases} cy & \text{if } x = 1,\, 2,\, \ldots,\, n;\; y = 1,\, 2,\, \ldots,\, n;\; x \leq y \\ 0 & \text{otherwise,} \end{cases}$$

where c is a constant. Note that the function $f(\cdot,\, \cdot)$ is almost similar to that of the previous example. Only difference is in the sets where the functions are strictly positive. In the previous example, f was positive on $\{1,\, 2,\, \ldots,\, n\} \times \{1,\, 2,\, \ldots,\, n\}$. In the current example the set is $\{(x,\, y) \in \mathbb{R}^2 : x = 1,\, 2,\, \ldots,\, n;\; y = 1,\, 2,\, \ldots,\, n;\; x \leq y\}$. However, this changes the probability distribution completely. We will see that the marginal PMFs are also different. Hence, support is an important issue.

The constant c is positive and can be found as follows:

$$\sum_{(x,\, y) \in S_{X,\, Y}} f(x,\, y) = 1 \implies \sum_{y=1}^{n} \sum_{x=1}^{y} cy = 1 \implies c = \frac{6}{n(n+1)(2n+1)}.$$

Please note the range of the summations. Thus, the JPMF of $(X,\, Y)$ is given by

$$f(x,\, y) = \begin{cases} \frac{6y}{n(n+1)(2n+1)} & \text{if } x = 1,\, 2,\, \ldots,\, n;\; y = 1,\, 2,\, \ldots,\, n;\; x \leq y \\ 0 & \text{otherwise} \end{cases}$$

The marginal PMF of X can be found as follows: For $x \in \{1,\, 2,\, \ldots,\, n\}$,

$$P\left(X = x\right) = \sum_{y=x}^{n} cy = \frac{3(n+x)(n-x+1)}{n(n+1)(2n+1)}.$$

Please note the range of the summation above. Thus, the marginal PMF of X is given by

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

The marginal PMF of Y can also be found similarly and is given by

$$f_Y(y) = \begin{cases} \frac{6y^2}{n(n+1)(2n+1)} & \text{if } y = 1,\, 2,\, \ldots,\, n \\ 0 & \text{otherwise.} \end{cases}$$

||

3.3 Continuous Random Vector

Definition 3.5 (Continuous Random Vector)**.** *A random vector $(X,\, Y)$ is said to have a continuous distribution if there exists a non-negative integrable function $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ such that*

$$F_{X,Y}(x,\, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,\, s)dsdt$$

for all $(x,\, y) \in \mathbb{R}^2$. The function $f_{X,Y}$ is called the joint probability density function (JPDF) of $(X,\, Y)$. The set $S_{X,Y} = \{(x,\, y) \in \mathbb{R}^2 : f_{X,Y}(x,\, y) > 0\}$ is called the support of $(X,\, Y)$.

Again, the continuous random vector is a natural extension of CRV. The JPMF exists only for discrete random vector and JPDF exits for continuous random vector.

Theorem 3.7 (Properties of JPDF)**.** *Let $(X,\, Y)$ be a continuous random vector with JPDF $f_{X,Y}(\cdot,\, \cdot)$. Then*

1. $f_{X,Y}(x,\, y) \geq 0$ *for* $(x,\, y) \in \mathbb{R}^2$.
2. $\displaystyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,\, y)dxdy = 1.$

Proof: The proof of the Property 1 is straight forward from the definition of continuous random vector. For the Property 2,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,\, y)dxdy = \lim_{A\to\infty} \lim_{B\to\infty} \int_{-\infty}^{A} \int_{-\infty}^{B} f_{X,Y}(x,\, y)dxdy = \lim_{A\to\infty} \lim_{B\to\infty} F_{X,Y}(A,\, B) = 1.$$

□

Theorem 3.8. *If a function $g : \mathbb{R}^2 \to \mathbb{R}$ satisfy Properties 1 and 2 of the Theorems 3.7, then $g\,(\cdot,\, \cdot)$ is JPDF of some 2-dimensional continuous random vector.*

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

Theorem 3.9 (Marginal PDF from JPDF)**.** *Let $(X,\, Y)$ be a continuous random vector with JPDF $f_{X,Y}\,(\cdot,\, \cdot)$. Then X and Y are CRVs. The PDF of X is given by*

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,\, y)dy \quad \text{for all fixed } x \in \mathbb{R}. \tag{3.3}$$

The PDF of Y is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,\, y)dx \quad \textit{for all fixed } y \in \mathbb{R}. \tag{3.4}$$

In the context of continuous random vector, $f_X(\cdot)$ and $f_Y(\cdot)$ are called marginal PDF of X and marginal PDF of Y, respectively.

Proof: For $x \in \mathbb{R}$,

$$\begin{aligned}
F_X(x) &= \lim_{y\to\infty} F_{X,Y}(x,\, y) \\
&= \lim_{y\to\infty} \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(s,t)dtds \\
&= \int_{-\infty}^{x} \left\{ \lim_{y\to\infty} \int_{-\infty}^{y} f_{X,Y}(s,\, t)dt \right\} ds \\
&= \int_{-\infty}^{x} \left\{ \int_{-\infty}^{\infty} f_{X,Y}(s,\, t)dt \right\} ds \\
&= \int_{-\infty}^{x} g(s)ds,
\end{aligned}$$

where $g(s) = \int_{-\infty}^{\infty} f_{X,Y}(s,\, t)dt$. The third equality holds true as $f_{X,Y}(x,\, y) \geq 0$ for all $(x,\, y) \in \mathbb{R}^2$. Thus, X is a CRV with PDF as given in (3.3). Similarly, we can prove that Y is also CRV with PDF given in (3.4). $\square$

Example 3.3. Let $(X,\, Y)$ be a CRV with JPDF

$$f(x,\, y) = \begin{cases} ce^{-(2x+3y)} & \text{if } 0 < x < y < \infty \\ 0 & \text{otherwise,} \end{cases}$$

where c is a constant. Clearly, $c > 0$ as $f_{X,Y}(x,\, y) \geq 0$ for all $(x,\, y) \in \mathbb{R}^2$. The value of c can be found as follows:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,\, y)dydx = 1 \implies c\int_{0}^{\infty}\int_{x}^{\infty} e^{-(2x+3y)}dydx = 1 \implies c = 15.$$

Note the range of integration. We can find the marginal PDF of X as follows: For $x \leq 0$,

$$f_X(x) = \int_{-\infty}^{\infty} f(x,\, y)dy = 0,$$

as the integrand is zero for all $x \leq 0$. For $x > 0$,

$$f_X(x) = \int_{-\infty}^{x} f(x,\, y)dy + \int_{x}^{\infty} f(x,\, y)dy = \int_{-\infty}^{\infty} f(x,\, y)dy = 15\int_{x}^{\infty} e^{-(2x+3y)}dy = 5e^{-5x},$$

as $f(x,\, y) = 0$ for $y < x$. Thus, the marginal PDF of X is given by

$$f_X(x) = \begin{cases} 5e^{-5x} & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Hence, $X \sim Exp(5)$. Similarly, the marginal PDF of Y can be calculated and is given by

$$f_Y(y) = \begin{cases} \frac{15}{2}e^{-3y}\left(1 - e^{-2y}\right) & \text{if } y > 0 \\ 0 & \text{otherwise.} \end{cases}$$

||

3.4 Expectation of Function of Random Vector

Definition 3.6 (Expectation of Function of Discrete Random Vector)**.** *Let* (X, Y) *be a discrete random vector with JPMF* $f_{X,Y}(\cdot,\, \cdot)$ *and support* $S_{X,Y}$*. Let* $h : \mathbb{R}^2 \to \mathbb{R}$*. Then the expectation of* $h(X, Y)$ *is defined by*

$$E\left(h(X,\, Y)\right) = \sum_{(x,\, y) \in S_{X,\, Y}} h(x,\, y) f_{X,Y}(x,\, y),$$

provided $\displaystyle\sum_{(x,\, y) \in S_{X,\, Y}} |h(x,\, y)| f_{X,Y}(x,\, y) < \infty$*.*

Definition 3.7 (Expectation of Function of Continuous Random Vector)**.** *Let* (X, Y) *be a continuous random vector with JPDF* $f_{X,Y}(\cdot,\, \cdot)$*. Let* $h : \mathbb{R}^2 \to \mathbb{R}$*. Then the expectation of* $h(X, Y)$ *is defined by*

$$E\left(h(X,\, Y)\right) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,\, y) f_{X,Y}(x,\, y) dx dy,$$

provided $\displaystyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |h(x,\, y)| f_{X,Y}(x,\, y) dx dy < \infty$*.*

Theorem 3.10 (Linearity Property of Expectation)**.** *Let* $\boldsymbol{X} = (X_1,\, X_2,\, \ldots,\, X_n)$ *be either a discrete random vector or a continuous random vector. Then*

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$$

where $a_i \in \mathbb{R}$ *is a constant for all* $i = 1,\, 2,\, \ldots,\, n$*. Here we assume that all the expectations exist.*

Proof: We will prove the theorem for continuous random vector $\boldsymbol{X} = (X, Y)$. The proof for general value of n is similar. Also, the proof for discrete random vector can be written easily by replacing integration sign by summation sign. Let f be the JPDF of $\boldsymbol{X}$. Then

$$\begin{aligned}
E\left(a_1 X + a_2 Y\right) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(a_1 x + a_2 y\right) f_{X,Y}(x,\, y) dx dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} a_1 x f_{X,Y}(x,\, y) dx dy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} a_2 y f_{X,Y}(x,\, y) dx dy \\
&= a_1 \int_{-\infty}^{\infty} x \left\{\int_{-\infty}^{\infty} f_{X,Y}(x,\, y) dy\right\} dx + a_2 \int_{-\infty}^{\infty} y \left\{\int_{-\infty}^{\infty} f_{X,Y}(x,\, y) dx\right\} dy \\
&= a_1 \int_{-\infty}^{\infty} x f_X(x) dx + a_2 \int_{-\infty}^{\infty} y f_Y(y) dy \\
&= a_1 E\left(X\right) + a_2 E\left(Y\right).
\end{aligned}$$

$\square$

The previous theorem tells us that we can compute expectation of a linear combination of a random vectors by computing the expectations of individual components of the random vector and then taking the linear combination. Note that the left hand side involves the joint distribution and an n-dimensional integration (for continuous random vector) or an n-dimensional summation (for discrete random vector). However, the right hand side involves n one-dimensional integrations or n one-dimensional summations. Also, we need the marginal distributions (not joint distribution) to compute the right hand side. Sometimes it is much easier to compute n one-dimensional integrations compared to a single n-dimensional integration. Same is true for summations. Also, in many problems, it is easier to obtain the marginal distributions than to obtain joint distribution. The following example illustrate this.

Example 3.4. At a party n men throw their hats into the center of a room. The hats are mixed up and each man randomly selects one. Suppose that we want to calculate the expected number of men who selects their own hat. Let X denote the number of men who selects their own hat. We are interested to find $E(X)$. To compute $E(X)$ directly, we need the distribution of X. It is clear that X should be a DRV with support $S_X = \{0,\, 1,\, \ldots,\, n-2,\, n\}$. Note that $n-1 \notin S_X$ *(why?)*. It is quite easy to find the values $P(X = k)$ for $k = 0$ and n. However, it is quite difficult to find $P\,(X = k)$ for other values of $k \in S_X$ and $k = 0,\, n$. Thus, it becomes a difficult problem if we try to compute $E(X)$ directly.

Let us try to solve it by converting the problem into a multidimensional problem. For $i = 1,\, 2,\, \ldots,\, n$, let us define the RV

$$X_i = \begin{cases} 1 & \text{if } i\text{th person takes his own hat} \\ 0 & \text{otherwise.} \end{cases}$$

Clearly, $X = X_1 + X_2 + \ldots + X_n$. Thus, using the previous theorem, $E(X) = E(X_1) + \ldots + E(X_n)$. Now, we need to compute $E(X_i)$ for all $i = 1,\, 2,\, \ldots,\, n$. Note that $P\,(X_i = 1) = \frac{1}{n}$ for all $i = 1,\, 2,\, \ldots,\, n$. Hence, $E(X_i) = \frac{1}{n}$ for all $i = 1,\, 2,\, \ldots,\, n$ implies $E(X) = 1$. Thus, on an average only one person takes his own hat, and it does not depend on n, the number of persons present in the game. ||

3.5 Some Useful Remarks

Remark 3.2. In Theorem 3.6, we have seen that if $(X,\, Y)$ is a discrete random vector then X and Y are DRVs. On the other hand, suppose that X and Y are DRVs with supports S_X and S_Y, respectively. Suppose that $S = S_X \times S_Y$. Clearly, S is atmost countable. Then

$$\begin{aligned} P\left((X,\, Y) \in S\right) &= \sum_{(x,\, y) \in S} P\left(X = x,\, Y = y\right) \\ &= \sum_{x \in S_X} \left\{ \sum_{y \in S_Y} P\left(X = x,\, Y = y\right) \right\} \\ &= \sum_{x \in S_X} P\left(X = x\right), \quad \text{using theorem of total probability} \\ &= 1. \end{aligned}$$

Let $T = \{(x,\, y) \in \mathbb{R}^2 : P\left(X = x,\, Y = y\right) > 0\}$. Then $T \subseteq S$, and hence, T is atmost countable. Also, $P\left((X,\, Y) \in T\right) = 1$. Thus, $(X,\, Y)$ is discrete random vector. This discussion shows that $(X,\, Y)$ is discrete random vector if and only if X and Y are DRVs. †

Remark 3.3. If (X, Y) is continuous random vector, then

$$P\left((X, Y) \in A\right)\right) = \int\int_{(x,\, y) \in A} f_{X,Y}(x,\, y)dxdy,$$

for all $A \subseteq \mathbb{R}^2$ such that the integration is possible. This statement can be seen as an extension of the fact that if X is a CRV with PDF $f(\cdot)$, then $P\left(X \in B\right) = \int_B f(x)dx$. However, the mathematical proof is out of the scope of this course. †

Remark 3.4. In Theorem 3.9, we have seen that if (X, Y) is continuous random vector, then X and Y are continuous random variables. However, the converse, in general, is not true. Thus, (X, Y) may not be a continuous random vector even if X and Y are CRVs. Consider the following example in this regards. Let X be a CRV. Suppose that $Y = X$. Then X and Y are CRVs. It is clear that $P\left(X = Y\right) = 1$. Now, if possible, assume that (X, Y) is a continuous random vector. Thus, (X, Y) has a JPDF, say $f(\cdot, \cdot)$. Then

$$P\left(X = Y\right) = \int\int_{x=y} f(x, y)dxdy = 0.$$

The last equality is true as a double integral $\int\int_B g(x,\, y)dxdy$ can be interpreted as the volume under the function $g(\cdot, \cdot)$ over the set B. As the area of the set $\{(x,\, y) \in \mathbb{R}^2 : x = y\}$ is zero, the volume is also zero. This is a contradiction to the fact that $P\left(X = Y\right) = 1$. Thus, our assumption is wrong and (X, Y) is not a continuous random vector.

In general, if there exists a set $A \subset \mathbb{R}^2$ whose area is zero and $P\left((X, Y) \in A\right) > 0$, then (X, Y) does not have a JPDF, and hence, (X, Y) is not a continuous random vector. †

Remark 3.5. In Theorems 3.6 and 3.9, we have seen that the marginal distributions can be recovered form the joint distribution. However, the converse, in general, is not true. Let us illustrate it using the following example. †

Example 3.5. Let $f(\cdot)$ and $g(\cdot)$ be two PDFs and $F(\cdot)$ and $G(\cdot)$ be the corresponding CDFs, respectively. Define, for $-1 < \alpha < 1$,

$$h(x,\, y) = f(x)g(y)\left\{1 + \alpha(1 - 2F(x))(1 - 2G(y))\right\}.$$

First, we will show that $h(\cdot, \cdot)$ is a JPDF of a two-dimensional random vector. As $0 \leq F(x) \leq 1$, $-1 \leq 1 - 2F(x) \leq 1$. Similarly $-1 \leq 1 - 2G(y) \leq 1$. Hence, for all $(x,\, y) \in \mathbb{R}^2$,

$$-|\alpha| \leq \alpha\left(1 - 2F(x)\right)\left(1 - 2G(y)\right) \leq |\alpha| \implies 1 + \alpha\left(1 - 2F(x)\right)\left(1 - 2G(y)\right) \geq 0.$$

As $f(x) \geq 0$ and $g(y) \geq 0$, $h(x,\, y) \geq 0$ for all $(x,\, y) \in \mathbb{R}^2$. Also,

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,\, y)dxdy &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)g(y)\left\{1 + \alpha(1 - 2F(x))(1 - 2G(y))\right\}dxdy \\
&= 1 + \alpha\left(\int_{\infty}^{\infty} f(x)\left(1 - 2F(x)\right)dx\right)\left(\int_{-\infty}^{\infty} g(y)\left(1 - 2G(y)\right)dy\right) \\
&= 1 + \alpha\left(\int_{\infty}^{\infty}\left(1 - 2F(x)\right)dF(x)\right)\left(\int_{-\infty}^{\infty}\left(1 - 2G(y)\right)dG(y)\right) \\
&= 1.
\end{aligned}$$

Thus, $h(\cdot, \cdot)$ is a JPDF of a 2-dimensional continuous random vector, say (X, Y). Let us try to find the marginal PDFs of X and Y. The marginal PDF of X is

$$\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} h(x, y)dy \\
&= f(x)\int_{-\infty}^{\infty} g(y)dy + \alpha f(x)\left(1 - 2F(x)\right)\int_{-\infty}^{\infty} g(y)\left(1 - 2G(y)\right)dy \\
&= f(x).
\end{aligned}$$

Similarly the marginal PDF of Y is $g(\cdot)$. Thus, the marginal PDFs of X and Y does not depend on α. However, the JPDF depends on the value of α. For different values of α, we have different JPDF, but the marginals remain same. Hence, given the marginal distributions, in general, we cannot construct the joint distribution. ||

3.6 Independent Random Variables

Definition 3.8 (Independent RVs). *The random variables $X_1, X_2, \ldots, X_n$ are said to be independent if*

$$F_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i),$$

for all $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, where $F_{X_i}(\cdot)$ is the marginal CDF of X_i.

Thus, two RVs X and Y are independent if and only if the events $E_x = \{X \leq x\}$ and $F_y = \{Y \leq y\}$ are independent for all $(x, y) \in \mathbb{R}^2$. For discrete random vector, an equivalent definition of independence can be given in terms of JPMF and marginal PMFs. Similarly, for continuous random vector, an equivalent definition of independence can be given in terms of JPDF and marginal PDFs. Next, we will give these two alternative definitions, however we will not prove the equivalence of the respective definitions. Nonetheless, these alternative definitions are very handy in many applications.

Definition 3.9 (Alternative Definition for Discrete Random Vector). *Let $(X_1, X_2, \ldots, X_n)$ be a discrete random vector with JPMF $f_{X_1, X_2, \ldots, X_n}(\cdot, \ldots, \cdot)$. Then $X_1, X_2, \ldots, X_n$ are said to be independent if*

$$f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

for all $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, where $f_{X_i}(\cdot)$ is the marginal PMF of X_i, $i = 1, 2, \ldots, n$.

Definition 3.10 (Alternative Definition for Continuous Random Vector). *Let $(X_1, X_2, \ldots, X_n)$ be a continuous random vector with JPDF $f_{X_1, X_2, \ldots, X_n}(\cdot, \ldots, \cdot)$. Then $X_1, X_2, \ldots, X_n$ are said to be independent if*

$$f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

for all $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, where $f_{X_i}(\cdot)$ is the marginal PDF of X_i, $i = 1, 2, \ldots, n$.

In the last section, we have pointed out that, in general, the joint distribution cannot be recovered from the marginal distributions. However, if $X_1, X_2, \ldots, X_n$ are independent RVs and if we know the marginal distributions of X_i for all $i = 1, 2, \ldots, n$, then we can write $F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i)$ for all $(x_1, \ldots, x_n) \in \mathbb{R}^n$. Thus, we can recover the joint distribution from the marginal distributions if the RVs are known to be independent. Moreover, if $X_1, X_2, \ldots, X_n$ are independent CRVs, then $(X_1, \ldots, X_n)$ is a continuous random vector.

Theorem 3.11. *If X and Y are independent, then*

$$E\left(g(X)h(Y)\right) = E\left(g(X)\right) E\left(h(Y)\right),$$

provided all the expectations exist.

Proof: We will prove it for continuous random vector (X, Y). For discrete random vector, it can be proved by replacing the integration sign by summation sign.

$$\begin{aligned} E\left(g(X)h(Y)\right) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_{X,Y}(x, y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)dxdy, \quad \text{as } X \text{ and } Y \text{ are independent} \\ &= \left(\int_{-\infty}^{\infty} g(x)f_X(x)dx\right)\left(\int_{-\infty}^{\infty} h(y)f_Y(y)dy\right) \\ &= E\left(g(X)\right) E\left(h(Y)\right). \end{aligned}$$

□

3.7 Covariance and Correlation Coefficient

Definition 3.11 (Covariance)**.** *The covariance of two random variables X and Y is defined by*

$$Cov(X, Y) = E\left[(X - E(X))(Y - E(Y))\right] = E(XY) - E(X)E(Y).$$

Definition 3.12 (Correlation Coefficient)**.** *The correlation coefficient of X and Y is defined by*

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}.$$

Remark 3.6. Let X and Y be independent, then

$$Cov(X, Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0.$$

We get the second equality using the previous theorem. Thus, $\rho(X, Y) = 0$. However, the converse is not true in general. That means that there exists dependent RVs X and Y such that $Cov(X, Y) = 0$. Let us consider the following example in this regards. †

Example 3.6. Let $X \sim N(0, 1)$ and $Y = X^2$. Then

$$Cov(X, Y) = E\left(XY\right) - E(X)E(Y) = E\left(X^3\right) - E(X)E(Y).$$

It is easy to see that $E(X) = 0$ and $E(X^3) = 0$. Hence, $Cov(X, Y) = 0$. Now, $P\left(X \le -5\right) = \Phi(-5) \ne 0$ and $P\left(Y \le 1\right) = P\left(-1 \le X \le 1\right) = 2\Phi(1) - 1 \ne 0$. However, $P\left(X \le -5, Y \le 1\right) = 0$. Thus, X and Y are not independent. ||

Theorem 3.12. *Let X, Y, Z, X_1, $\dots$, X_n, Y_1, $\dots$, Y_m be RVs such that all the necessary expectations for the followings exist. Then*

1. $Cov(X, X) = Var(X)$.
2. $Cov(X, Y) = Cov(Y, X)$.
3. $Cov(aX, Y) = a\,Cov(X, Y)$ *for a real constant* a.
4. $Cov(X + Z, Y) = Cov(X, Y) + Cov(Z, Y)$.
5. $Cov\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j Cov(X_i, Y_j)$ *for real constants* a_1, a_2, $\dots$, a_n *and* b_1, b_2, $\dots$, b_m.
6. $Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i<j} Cov(X_i, Y_j)$.
7. *If* X_i*'s are independent, then* $Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)$.

Proof:

1. Straight forward form the definition.
2. Straight forward form the definition.
3. $$\begin{aligned} Cov\,(aX, Y) &= E\left(\left(aX - E\,(aX)\right)\left(Y - E(Y)\right)\right) \\ &= aE\left(\left(X - E(X)\right)\left(Y - E(Y)\right)\right) \\ &= a\,Cov(X, Y). \end{aligned}$$
4. Straight forward from the definition.
5. Combining 2, 3, and 4, we can prove it.
6. Combining 1 and 5, this proof is trivial.
7. Using Remark 3.6, it can be readily obtained from 5.

□

Theorem 3.13. $|\rho(X, Y)| \le 1$ *provided it exists.*

Proof: Note that for any $\lambda \in \mathbb{R}$,

$$Var(X + \lambda Y) \ge 0 \implies \lambda^2 Var(Y) + 2\lambda Cov(X, Y) + Var(X) \ge 0.$$

That means that the quadratic equation $\lambda^2 Var(Y) + 2\lambda Cov(X, Y) + Var(X) = 0$ either has one solution or no solution. Hence,

$$4\left(Cov\,(X, Y)\right)^2 - 4Var(X)Var(Y) \le 0 \implies |\rho(X, Y)| \le 1.$$

□

3.8 Transformation Techniques

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ be a random vector and $g : \mathbb{R}^n \to \mathbb{R}^m$. Clearly, $\boldsymbol{Y} = g(\boldsymbol{X})$ is a m-dimensional random vector. In this section, we will discuss different methods to find the distribution of the random vector $\boldsymbol{Y} = g(\boldsymbol{X})$. Like the previous chapter, there are mainly three techniques to obtain the distribution of $\boldsymbol{Y} = g(\boldsymbol{X})$.

3.8.1 Technique 1

In Technique 1, we try to find the JCDF of $\boldsymbol{Y} = g(\boldsymbol{X})$ given the distribution of $\boldsymbol{X}$. As before, we will discuss this technique using examples.

Example 3.7. Let X_1 and X_2 be identically and independently distributed (*i.i.d.*) $U(0, 1)$ random variables. Suppose we want to find the CDF of $Y = X_1 + X_2$. Now,

$$F_Y(y) = P(Y \leq y) = P(X_1 + X_2 \leq y) = \int\int_{x_1+x_2\leq y} f_{X_1, X_2}(x_1, x_2) dx_1 dx_2. \tag{3.5}$$

As $X_1 \sim U(0, 1)$, $X_2 \sim U(0, 1)$ and X_1 and X_2 are independent RVs, the JPDF of (X_1, X_2) is given by

$$f_{X_1, X_2}(x_1, x_2) = \begin{cases} 1 & \text{if } 0 < x_1 < 1,\ 0 < x_2 < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Thus, the JPDF of (X_1, X_2) is positive only on the unit square $(0, 1) \times (0, 1)$, which is indicated by gray shade in Figure 3.2. Now, to compute the integration in (3.5), we need to consider the following cases.

For $y < 0$, consider the Figure 3.2a. As the integrand in (3.5) is zero over the region $\{(x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 \leq y\}$ for $y < 0$,

$$F_Y(y) = 0.$$

For $0 \leq y < 1$, consider the Figure 3.2b. The integrand is positive only on the shaded region in the set $\{(x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 \leq y\}$. Therefore,

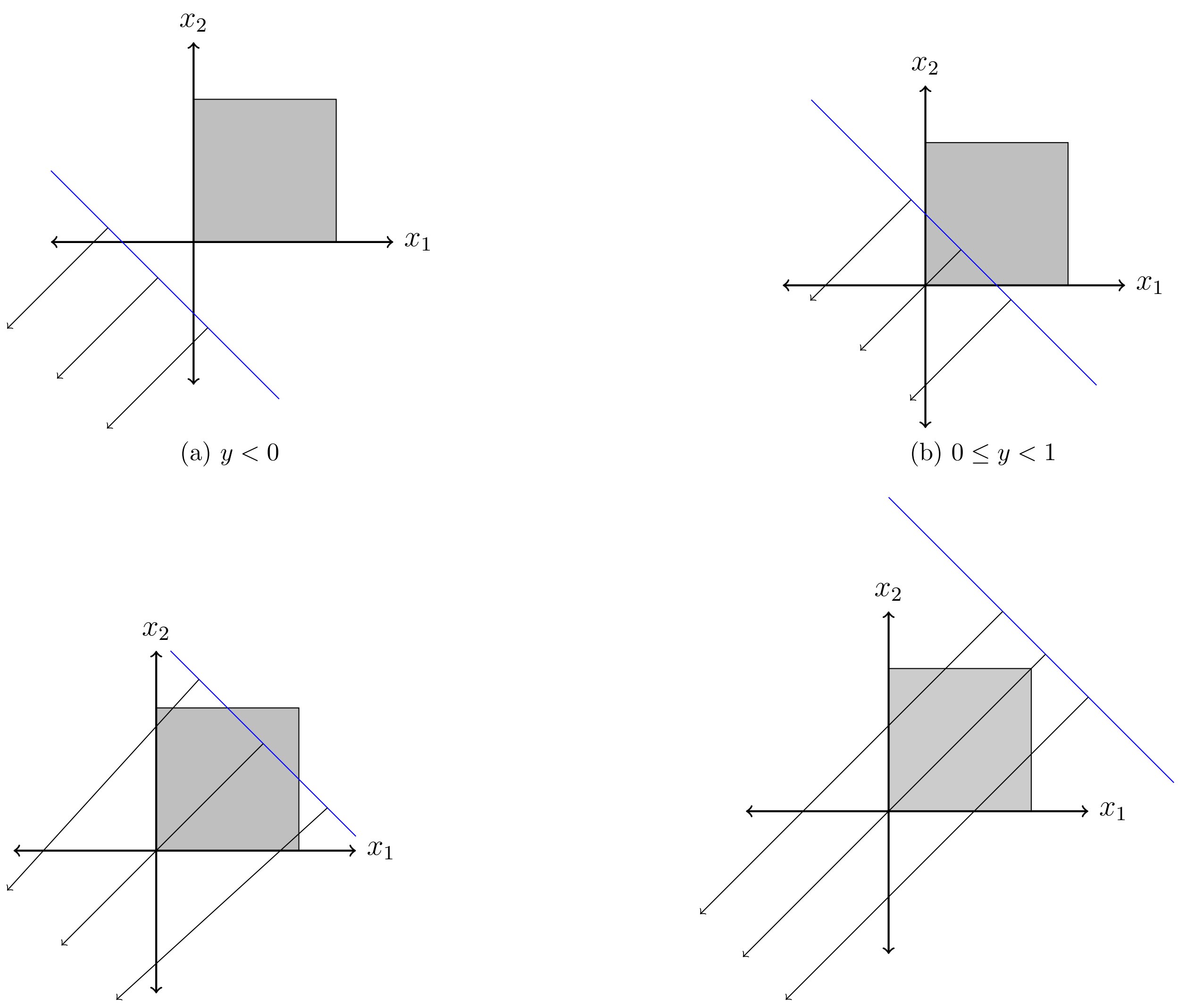
$$F_Y(y) = \int_0^y \int_0^{y-x_2} dx_1 dx_2 = \frac{1}{2} y^2.$$

For $1 \leq y < 2$, consider the Figure 3.2c. The integrand is positive only on the shaded region in the set $\{(x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 \leq y\}$. Therefore,

$$F_Y(y) = 1 - \int_{y-1}^1 \int_{y-x_2}^1 dx_1 dx_2 = 1 - \frac{1}{2}(2 - y)^2.$$

For $y \geq 2$, consider the Figure 3.2d. The integrand is positive on the shaded region in the set $\{(x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 \leq y\}$ and the square $(0, 1) \times (0, 1)$ is completely inside the set $\{(x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 \leq y\}$. Therefore,

$$F_Y(y) = 1.$$



(a) $y < 0$ (b) $0 \le y < 1$

(c) $1 \le y < 2$ (d) $y \ge 2$

Figure 3.2: Plot for Example 3.7

Thus, the CDF of $Y = X_1 + X_2$ is given by

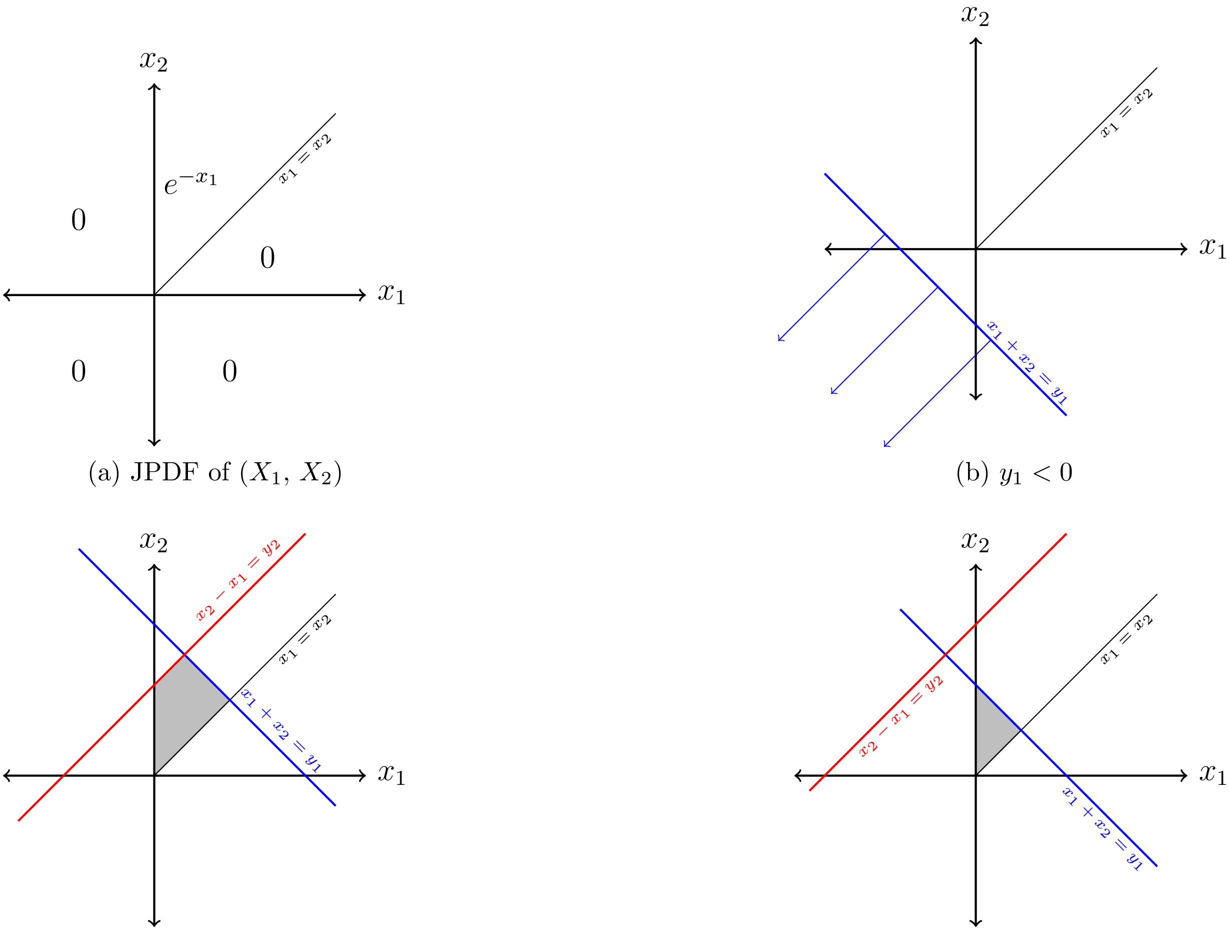
$$F_Y(y) = \begin{cases} 0 & \text{if } y < 0 \\ \frac{1}{2}y^2 & \text{if } 0 \le y < 1 \\ 1 - \frac{1}{2}(2-y)^2 & \text{if } 1 \le y < 2 \\ 1 & \text{if } y \ge 2. \end{cases}$$

It can be shown that Y is a CRV *(why?)*. ||

Example 3.8. Let the JPDF of (X_1, X_2) be given by

$$f_{X_1, X_2}(x_1, x_2) = \begin{cases} e^{-x_1} & \text{if } 0 < x_1 < x_2 < \infty \\ 0 & \text{otherwise.} \end{cases}$$

Suppose that we want to find the JCDF of $Y_1 = X_1 + X_2$ and $Y_2 = X_2 - X_1$. Note that the JPDF of (X_1, X_2) is positive only on the set $S_{X_1, X_2} = \{(x_1, x_2) \in \mathbb{R}^2 : 0 < x_1 < x_2 < \infty\}$.



(a) JPDF of $(X_1,\, X_2)$

(b) $y_1 < 0$

(c) $0 < y_2 < y_1 < \infty$

(d) $0 < y_1 < y_2 < \infty$

Figure 3.3: Plot for Example 3.8

See Figure 3.3a. Now, let $A_{y_1,\,y_2} = \{(x_1,\, x_2) \in \mathbb{R} : x_1 + x_2 \le y_1,\, x_2 - x_1 \le y_2\}$. Then

$$F_{Y_1,\,Y_2}(y_1,\, y_2) = P\left(X_1 + X_2 \le y_1,\, X_2 - X_1 \le y_2\right) = \int\int_{A_{y_1,\,y_2}} f_{X_1,\,X_2}(x_1,\, x_2)dx_2dx_1. \quad (3.6)$$

Suppose that $y_1 < 0$. Then $F_{Y_1}(y_1) = 0$. See the Figure 3.3b. As $F_{Y_1,\,Y_2}(y_1,\, y_2) \le \min\left\{F_{Y_1}(y_1),\, F_{Y_2}(y_2)\right\}$, $F_{Y_1,\,Y_2}(y_1,\, y_2) = 0$ for $y_1 < 0$. Similarly, $F_{Y_1,\,Y_2}(y_1,\, y_2) = 0$ for $y_2 < 0$. For $0 < y_2 < y_1 < \infty$, $A_{y_1,y_2} \cap S_{X_1,\,X_2}$ is the shaded region of the Figure 3.3c. Therefore,

$$\begin{aligned} F_{Y_1,\,Y_2}(y_1,\, y_2) &= \int_0^{\frac{y_1-y_2}{2}} \int_{x_1}^{x_1+y_1} e^{-x_1}dx_2dx_1 + \int_{\frac{y_1-y_2}{2}}^{\frac{y_1}{2}} \int_{x_1}^{y_1-x_1} e^{-x_1}dx_2dx_1 \\ &= y_1 + e^{-\frac{y_1}{2}} - (y_1 - y_2 + 2)e^{-\frac{y_1-y_2}{2}}. \end{aligned}$$

For $0 < y_1 < y_2 < \infty$, $A_{y_1,y_2} \cap S_{X_1,\,X_2}$ is indicated by the shaded region in the Figure 3.3d. Therefore,

$$F_{Y_1,\,Y_2}(y_1,\, y_2) = \int_0^{\frac{y_1}{2}} \int_{x_1}^{y_1-x_1} e^{-x_1}dx_2dx_1 = y_1 + 2e^{-\frac{y_1}{2}} - 2.$$

Thus, the JCDF of $(Y_1, Y_2) = (X_1 + X_2, X_2 - X_1)$ is given by

$$F_{Y_1, Y_2}(y_1, y_2) = \begin{cases} 0 & \text{if } y_1 < 0 \text{ or } y_2 < 0 \\ y_1 + e^{-\frac{y_1}{2}} - (y_1 - y_2 + 2)e^{-\frac{y_1 - y_2}{2}} & \text{if } 0 < y_2 \leq y_1 < \infty \\ y_1 + 2e^{-\frac{y_1}{2}} - 2 & \text{if } 0 < y_1 < y_2 < \infty. \end{cases}$$

It can be shown that (Y_1, Y_2) is a continuous random vector. This is easy and therefore left as an exercise. ||

3.8.2 Technique 2

Like the previous chapter, this technique is based on two theorems, one for discrete random vector and another for continuous random vector.

Theorem 3.14. *Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ be a discrete random vector with JPMF $f_{\boldsymbol{X}}$ and support $S_{\boldsymbol{X}}$. Let $g_i : \mathbb{R}^n \to \mathbb{R}$ for all $i = 1, 2, \ldots, k$. Let $Y_i = g_i(\boldsymbol{X})$ for $i = 1, 2, \ldots, k$. Then $\boldsymbol{Y} = (Y_1, \ldots, Y_k)$ is a discrete random vector with JPMF*

$$f_{\boldsymbol{Y}}(y_1, \ldots, y_k) = \begin{cases} \displaystyle\sum_{\boldsymbol{x} \in A_{\boldsymbol{y}}} f_{\boldsymbol{X}}(\boldsymbol{x}) & \text{if } (y_1, \ldots, y_k) \in S_{\boldsymbol{Y}} \\ 0 & \text{otherwise,} \end{cases}$$

where $A_{\boldsymbol{y}} = \{\boldsymbol{x} \in S_{\boldsymbol{X}} : g_i(\boldsymbol{x}) = y_i, \, i = 1, \ldots, k\}$ and $S_{\boldsymbol{Y}} = \{(g_1(\boldsymbol{x}), \ldots, g_k(\boldsymbol{x})) : \boldsymbol{x} \in S_{\boldsymbol{X}}\}$.

Proof: The proof of the theorem is similar to that of Theorem 2.7. □

Example 3.9. Let $X_1 \sim Poi(\lambda_1)$ and $X_2 \sim Poi(\lambda_2)$. Also, assume that X_1 and X_2 are independent. Then $Y = X_1 + X_2 \sim Poi(\lambda_1 + \lambda_2)$. To see it, we can apply Theorem 3.14. First note that the JPMF if (X_1, X_2) is given by

$$f_{X_1, X_2}(x_1, x_2) = \begin{cases} \frac{e^{-(\lambda_1 + \lambda_2)} \lambda_1^{x_1} \lambda_2^{x_2}}{x_1! x_2!} & \text{if } x_1 = 0, 1, \ldots; \, x_2 = 0, 1, \ldots \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, $S_{X_1, X_2} = \{0, 1, 2, \ldots\} \times \{0, 1, 2, \ldots\}$, which implies that $S_Y = \{0, 1, 2, \ldots\}$. For $y \in S_Y$, $A_y = \{(x, y - x) : x = 0, 1, \ldots, y\}$. Hence, using the Theorem 3.14, for $y \in S_Y$,

$$f_Y(y) = \sum_{(x_1, x_2) \in A_y} \frac{e^{-(\lambda_1 + \lambda_2)} \lambda_1^{x_1} \lambda_2^{x_2}}{x_1! x_2!} = \frac{e^{-(\lambda_1 + \lambda_2)}}{y!} \sum_{x=0}^{y} \binom{y}{x} \lambda_1^x \lambda_2^{y-x} = \frac{1}{y!} e^{-(\lambda_1 + \lambda_2)} (\lambda_1 + \lambda_2)^y .$$

Thus, the PMF of $Y = X_1 + X_2$ is

$$f_Y(y) = \begin{cases} \frac{1}{y!} e^{-(\lambda_1 + \lambda_2)} (\lambda_1 + \lambda_2)^y & \text{if } y = 0, 1, \ldots \\ 0 & \text{otherwise,} \end{cases}$$

which is PMF of a $P(\lambda_1 + \lambda_2)$. Hence, $X_1 + X_2 \sim P(\lambda_1 + \lambda_2)$. ||

Example 3.10. Let $X_1 \sim Bin(n_1, p)$ and $X_2 \sim Bin(n_2, p)$. We also assume that X_1 and X_2 are independent. Suppose that we want to find the PMF of $Y = X_1 + X_2$. Note that X_1 and X_2 are the numbers of successes out of n_1 and n_2 independent Bernoulli trials, respectively. In both the cases the probability of success is p. Therefore, Y is the number of successes out of $n_1 + n_2$ Bernoulli trials with success probability p. As X_1 and X_2 are independent, these $n_1 + n_2$ Bernoulli trials can be assumed to be independent. Hence, the distribution of Y must be $Bin(n_1 + n_2,\, p)$. Let us now check if we get the same distribution using the Theorem 3.14. The JPMF of X_1 and X_2 is

$$f_{X_1,\,X_2}(x_1,\, x_2) = \begin{cases} \binom{n_1}{x_1}\binom{n_2}{x_2} p^{x_1+x_2}(1-p)^{n_1+n_2-x_1-x_2} & \text{if } x_1 = 0,\, 1,\, \ldots,\, n_1;\ x_2 = 0,\, 1,\, \ldots,\, n_2 \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, $S_{X_1,\,X_2} = \{0,\, 1,\, \ldots,\, n_1\} \times \{0,\, 1,\, \ldots,\, n_2\}$. Without loss of generality, we assume that $n_1 \leq n_2$. If not, exchange the roles of X_1 and X_2. Now, $S_Y = \{0,\, 1,\, \ldots,\, n_1 + n_2\}$. For $y \in S_Y$,

$$\begin{aligned} A_y &= \{(x_1,\, x_2) \in S_{X_1,\,X_2} : x_1 + x_2 = y\} \\ &= \begin{cases} \{(x,\, y-x) : x = 0,\, 1,\, \ldots,\, y\} & \text{if } 0 \leq y \leq n_1 \\ \{(x,\, y-x) : x = 0,\, 1,\, \ldots,\, n_1\} & \text{if } n_1 < y \leq n_2 \\ \{(x,\, y-x) : x = y-n_2,\, \ldots,\, n_1\} & \text{if } n_2 < y \leq n_1 + n_2. \end{cases} \end{aligned}$$

Hence, for $y \in S_Y$ and $y \leq n_1$,

$$f_Y(y) = \sum_{x=0}^{y} \binom{n_1}{x}\binom{n_2}{y-x} p^y (1-p)^{n_1+n_2-y} = \binom{n_1+n_2}{y} p^y (1-p)^{n_1+n_2-y}.$$

The last equality can be proved by collecting the coefficient of x^y from both sides of the following expression:

$$(1+x)^{n_1}(1+x)^{n_2} = \left\{\sum_{i=0}^{n_1} \binom{n_1}{i} x^i\right\} \times \left\{\sum_{i=0}^{n_2} \binom{n_2}{i} x^i\right\}.$$

For $y \in S_Y$ and $n_1 < y \leq n_2$,

$$f_Y(y) = \sum_{x=0}^{n_1} \binom{n_1}{x}\binom{n_2}{y-x} p^y (1-p)^{n_1+n_2-y} = \binom{n_1+n_2}{y} p^y (1-p)^{n_1+n_2-y}.$$

For $y \in S_Y$ and $n_2 < y \leq n_1 + n_2$,

$$f_Y(y) = \sum_{x=y-n_2}^{n_1} \binom{n_1}{x}\binom{n_2}{y-x} p^y (1-p)^{n_1+n_2-y} = \binom{n_1+n_2}{y} p^y (1-p)^{n_1+n_2-y}.$$

Thus, $X_1 + X_2 \sim Bin\,(n_1 + n_2,\, p)$. Note that independence of X_1 and X_2 and same value of probability of success are important for the result. ||

Theorem 3.15. *Let $\boldsymbol{X} = (X_1,\, \ldots,\, X_n)$ be a continuous random vector with JPDF $f_{\boldsymbol{X}}$.*

1. *Let $y_i = g_i(\boldsymbol{x})$, $i = 1,\, 2,\, \ldots,\, n$ be $\mathbb{R}^n \to \mathbb{R}$ functions such that*

$$\boldsymbol{y} = g(\boldsymbol{x}) = (g_1(\boldsymbol{x}),\, \ldots,\, g_n(\boldsymbol{x}))$$

is one-to-one. That means that there exists the inverse transformation $x_i = h_i(\boldsymbol{y})$, $i = 1,\, 2,\, \ldots,\, n$ defined on the range of the transformation.

2. *Assume that both the mapping and its' inverse are continuous.*

3. *Assume that partial derivatives* $\frac{\partial x_i}{\partial y_j}$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n$, *exist and are continuous.*

4. *Assume that the Jacobian of the inverse transformation*

$$J \doteq \det \left(\frac{\partial x_i}{\partial y_j} \right)_{i,j=1,2,\ldots,n} \neq 0$$

on the range of the transformation.

Then $\boldsymbol{Y} = (g_1(\boldsymbol{X}), \ldots, g_n(\boldsymbol{X}))$ *is a continuous random vector with JPDF*

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{X}}(h_1(\boldsymbol{y}), \ldots, h_n(\boldsymbol{y}))|J|.$$

Proof: The proof of this theorem can be done using transformation of variable technique for multiple integration. However, the proof is skipped here. □

Remark 3.7. Note that g is a vector valued function. As g should be one-to-one, the dimension of g should be same as dimension of the argument of g. Though we have written that $g_i : \mathbb{R}^n \to \mathbb{R}$ in the previous theorem, the conclusion of the theorem is valid if we replace $g_i : \mathbb{R}^n \to \mathbb{R}$ by $g_i : S_{\boldsymbol{X}} \to \mathbb{R}$. Moreover, the theorem gives us sufficient conditions for $g(\boldsymbol{X})$ to be a continuous random vector, when $\boldsymbol{X}$ is continuous random vector. Thus, $g(\boldsymbol{X})$ can be a continuous random vector even if the conditions of the previous theorem do not hold true. †

Example 3.11. Let X_1 and X_2 be *i.i.d.* $U(0, 1)$ random variables. We want to find the JPDF of $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$. Clearly,

$$g_1(x_1, x_2) = x_1 + x_2 \quad \text{and} \quad g_2(x_1, x_2) = x_1 - x_2.$$

Thus, $\boldsymbol{y} = (y_1, y_2) = g(x_1, x_2) = (g_1(x_1, x_2), g_2(x_1, x_2)) = (x_1 + x_2, x_1 - x_2)$. Now, if $(x_1, x_2) \neq (\tilde{x}_1, \tilde{x}_2)$, then $g\,(x_1, x_2) \neq g(\tilde{x}_1, \tilde{x}_2)$. If not, then $x_1 + x_2 = \tilde{x}_1 + \tilde{x}_2$ and $x_1 - x_2 = \tilde{x}_1 - \tilde{x}_2$, which implies $x_1 = \tilde{x}_1$ and $x_2 = \tilde{x}_2$. This is a contradiction. Hence, the function $g(\cdot, \cdot)$ is one-to-one. The inverse function is given by $h(y_1, y_2) = (h_1(y_1, y_2), h_2(y_1, y_2))$, where $x_1 = h_1(y_1, y_2) = \frac{1}{2}(y_1 + y_2)$ and $x_2 = h_2(y_1, y_2) = \frac{1}{2}(y_1 - y_2)$. Clearly, both the mapping and inverse mapping are continuous. Now,

$$\frac{\partial x_1}{\partial y_1} = \frac{1}{2}, \; \frac{\partial x_1}{\partial y_2} = \frac{1}{2}, \; \frac{\partial x_2}{\partial y_1} = \frac{1}{2}, \quad \text{and} \quad \frac{\partial x_2}{\partial y_2} = -\frac{1}{2}.$$

All the partial derivatives are continuous. The Jacobian is

$$J = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2} \neq 0.$$

Thus, all the four conditions of the Theorem 3.15 hold, and hence, $\boldsymbol{Y} = (Y_1, Y_2)$ is a continuous random vector with JPDF

$$f_{Y_1, Y_2}(y_1, y_2) = f_{X_1, X_2}\left(\frac{1}{2}(y_1 + y_2), \frac{1}{2}(y_1 - y_2) \right) \left| -\frac{1}{2} \right|$$

$$= \begin{cases} \frac{1}{2} & \text{if } 0 < y_1 + y_2 < 2,\ 0 < y_1 - y_2 < 2 \\ 0 & \text{otherwise.} \end{cases}$$

Note that in Example 3.7, we have found the distribution of $X_1 + X_2$. You may find the marginal distribution of X_1+X_2 from JPDF above and check if you are getting same marginal distribution. ||

Example 3.12. Let X_1 and X_2 be *i.i.d.* $N(0,\,1)$ random variables. We want to find the PDF of $Y_1 = X_1/X_2$. Note that we cannot use Theorem 3.15 directly here as we have a single function $g_1(x_1,\,x_2) = \frac{x_1}{x_2}$. Thus, we need to bring an auxiliary new function $g_2(x_1,\,x_2)$ such that $g(x_1,\,x_2) = (g_1(x_1,\,x_2),\,g_2(x_1,\,x_2))$ satisfies all the conditions of Theorem 3.15. Let us take $g_2(x_1,\,x_2) = x_2$. Clearly, $g(x_1,\,x_2)$ is a one-to-one function. Here, the inverse function is $h(y_1,\,y_2) = (h_1(y_1,\,y_2),\,h_2(y_1,\,y_2))$, where $x_1 = h_1(y_1,\,y_2) = y_1y_2$ and $x_2 = h_2(y_1,\,y_2) = y_2$. It is easy to see that mapping g and its' inverse are continuous. Also,

$$\frac{\partial x_1}{\partial y_1} = y_2,\ \frac{\partial x_1}{\partial y_2} = y_1,\ \frac{\partial x_2}{\partial y_1} = 0, \quad \text{and} \quad \frac{\partial x_2}{\partial y_2} = 1.$$

All the partial derivatives are continuous. Hence, the Jacobian is

$$J = \begin{vmatrix} y_2 & y_1 \\ 0 & 1 \end{vmatrix} = y_2.$$

Thus, all the four conditions of the Theorem 3.15 hold, and hence, $\boldsymbol{Y} = \left(\frac{X_1}{X_2},\,X_2\right)$ is a continuous random vector with JPDF

$$f_{Y_1,Y_2}(y_1,\,y_2) = \frac{1}{2\pi}e^{-\frac{1}{2}(1+y_1^2)y_2^2}\,|y_2| \quad \text{for } (y_1,\,y_2) \in \mathbb{R}^2.$$

Now, we can find the marginal PDF of Y_1 from the JPDF of $(Y_1,\,Y_2)$. The marginal PDF of Y_1 is given by

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} \frac{|y_2|}{2\pi}e^{-\frac{1}{2}(1+y_1^2)y_2^2}dy_2 = \frac{1}{\pi}\int_0^{\infty} y_2e^{-\frac{1}{2}(1+y_1^2)y_2^2}dy_2 = \frac{1}{\pi(1+y_1^2)}$$

for all $y_1 \in \mathbb{R}$. Thus, $Y_1 \sim Cauchy(0,\,1)$. ||

Theorem 3.16. *If X and Y are independent, then $g(X)$ and $h(Y)$ are also independent.*

Proof: The exact proof of the theorem cannot be given in the course. However, intuitively it makes sense. X and Y are independent. That means that there is no effect of one of the RVs on the other. Now, g being a function of X only and h being a function of Y only, there should be no effect of $g(X)$ on $h(Y)$ and vice versa. □

3.8.3 Technique 3

The Technique 3 depends on the MGF. Hence, first we need to define the MGF of a random vector.

Definition 3.13 (Moment Generating Function)**.** *Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ be a random vector. The MGF of $\boldsymbol{X}$ at $\boldsymbol{t} = (t_1, t_2, \ldots, t_n)$ is defined by*

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = E\left(\exp\left(\sum_{i=1}^{n} t_iX_i\right)\right)$$

provided the expectation exists in a neighborhood of origin $\mathbf{0} = (0,\,0,\,\ldots,\,0)$.

Theorem 3.17. $E(X_1^{r_1}X_2^{r_2}\cdots X_n^{r_n}) = \frac{\partial^{r_1+r_2+\ldots+r_n}}{\partial t_1^{r_1}\partial t_2^{r_2}\ldots\partial t_n^{r_n}} M_{\boldsymbol{X}}(\boldsymbol{t})\Big|_{\boldsymbol{t}=0}$.

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

Theorem 3.18. *X and Y are independent iff $M_{X,Y}(t_1,t_2) = M_X(t_1)M_Y(t_2)$ in a neighborhood of the origin.*

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

Note that if X and Y are independent, then using Theorems 3.11, it is straight forward to see that $M_{X,Y}(t,\,s) = M_X(t)M_Y(s)$. Also, note that $E\left(g(X)h(Y)\right) = E\left(g(X)\right)E\left(h(Y)\right)$ for some functions g and h does not imply that X and Y are independent. In particular $E(XY) = E(X)E(Y)$ does not imply X and Y are independent. Please revisit Example 3.6 in this regard.

Definition 3.14. *Two n-dimensional random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ are said to have the same distribution, denoted by $\boldsymbol{X} \stackrel{d}{=} \boldsymbol{Y}$, if $F_{\boldsymbol{X}}(\boldsymbol{x}) = F_{\boldsymbol{Y}}(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^n$.*

Theorem 3.19. *Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two n-dimensional random vectors. Let $M_{\boldsymbol{X}}(\boldsymbol{t}) = M_{\boldsymbol{Y}}(\boldsymbol{t})$ for all $\boldsymbol{t}$ in a neighborhood around $\boldsymbol{0}$, then $\boldsymbol{X} \stackrel{d}{=} \boldsymbol{Y}$.*

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

Example 3.13. Let X_i, $i = 1, 2, \ldots, k$ be independent $Bin(n_i, p)$ RVs. Let us try to find the distribution of $Y = \sum_{i=1}^k X_i$. Now, the MGF of Y is

$$M_Y(t) = E\left(e^{tY}\right) = E\left(\exp\left(t\sum_{i=1}^k X_i\right)\right) = E\left(\prod_{i=1}^k e^{tX_i}\right) = \prod_{i=1}^k E(e^{tX_i}) = \prod_{i=1}^k M_{X_i}(t).$$

The fourth equality is true as the RVs $X_1,\,X_2,\,\ldots,\,X_k$ are independent. In Example 2.40, we have seen that the MGF of $X \sim Bin(n,\,p)$ is $M_X(t) = (1-p+pe^t)^n$ for all $t \in \mathbb{R}$. Thus, the MGF of Y is

$$M_Y(t) = \prod_{i=1}^k (1-p+pe^t)^{n_i} = (1-p+pe^t)^{\sum_{i=1}^k n_i}$$

for $t \in \mathbb{R}$. Let $Z \sim Bin\left(\sum_{i=1}^k n_i,\,p\right)$, then $M_Z(t) = M_Y(t)$ for all $t \in \mathbb{R}$. Thus, $Y \stackrel{d}{=} Z \sim Bin\left(\sum_{i=1}^k n_i,\,p\right)$. Note that this example is an extension of Example 3.10. ||

Example 3.14. Let $X_1,\,X_2,\,\ldots,\,X_k \stackrel{i.i.d.}{\sim} Exp(\lambda)$ and $Y = \sum_{i=1}^k X_i$. Then the MGF of Y is

$$M_Y(t) = \prod_{i=1}^k M_{X_i}(t) = \left[M_{X_1}(t)\right]^k = \left(1 - \frac{t}{\lambda}\right)^{-k}$$

for all $t < \lambda$. The second equality is due to the fact that X_i has same distribution for all $i = 1,\,2,\,\ldots,\,k$. The third equality is hold true form Example 2.41. Let $Z \sim Gamma(k, \lambda)$. Then $M_Z(t) = M_Y(t)$ for all $t < \lambda$. Hence, $Y \sim Gamma(k,\,\lambda)$. ||

Example 3.15. Let X_i, $i = 1, 2, \ldots, k$ be independent $N(\mu_i, \sigma_i^2)$ RVs. Then $\sum_{i=1}^k X_i \sim N\left(\sum_{i=1}^k \mu_i, \sum_{i=1}^k \sigma_i^2\right)$. This can be proved following the same technique as the last example. I am leaving it as an exercise. ||

Definition 3.15 (Expectation of a Random Vector)**.** *Expectation of a random vector is given by*

$$E(\boldsymbol{X}) = (EX_1, EX_2, \ldots, EX_n)' = \boldsymbol{\mu}.$$

Definition 3.16 (Variance-Covariance Matrix of a Random Vector)**.** *The variance-covariance matrix of a n-dimensional random vector, denoted by* Σ, *is defined by*

$$\Sigma = [Cov(X_i, X_j)]_{i,j=1}^n = E(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})'.$$

3.9 Conditional Distribution

3.9.1 For Discrete Random Vector

Definition 3.17. *Let* (X, Y) *be a discrete random vector with JPMF* $f_{X,Y}(\cdot, \cdot)$. *Suppose the marginal PMF of* Y *is* $f_Y(\cdot)$. *The conditional PMF of* X, *given* $Y = y$ *is defined by*

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

provided $f_Y(y) > 0$.

Note that $f_{X,Y}(x,\, y) = P\,(X = x,\, Y = y)$ and $f_Y(y) = P(Y = y)$. Thus, the conditional PMF of X given $Y = y$ is $P\,(X = x|Y = y)$. As we know that $P(A|B)$ is defined if $P(B) > 0$, here we need the condition that $f_Y(y) = P(Y = y) > 0$. Hence, $f_{X|Y}(x|y)$ is only defined for $y \in S_Y$.

Example 3.16. Let $X_1 \sim Poi(\lambda_1)$, $X_2 \sim Poi(\lambda_2)$. Also, assume that X_1 and X_2 are independent. In Example 3.9, we have seen that $X_1 + X_2 \sim Poi(\lambda_1 + \lambda_2)$ and the support of $X_1 + X_2$ is $S = \{0,\, 1,\, 2,\, \ldots\}$. Hence, the conditional PMF of X_1 given $X_1 + X_2 = y$ is defined for all $y \in S$, and the conditional PMF of X given $X_1 + X_2 = y$ is given by

$$\begin{aligned}
f_{X_1|X_1+X_2}(x|y) &= \frac{f_{X_1,\, X_1+X_2}(x,\, y)}{f_{X_1+X_2}(y)} \\
&= \frac{P\,(X_1 = x,\, X_1 + X_2 = y)}{P\,(X_1 + X_2 = y)} \\
&= \frac{P\,(X_1 = x,\, X_2 = y - x)}{P\,(X_1 + X_2 = y)} \\
&= \frac{P\,(X_1 = x)\, P\,(X_2 = y - x)}{P\,(X_1 + X_2 = y)}, \quad \text{as } X_1 \text{ and } X_2 \text{ are independent} \\
&= \begin{cases} \frac{e^{-\lambda_1}\frac{\lambda_1^x}{x!} \times e^{-\lambda_2}\frac{\lambda_2^{y-x}}{(y-x)!}}{e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^y}{y!}} & \text{if } x = 0,\, 1,\, 2\, \ldots,\, y \\ 0 & \text{otherwise} \end{cases} \\
&= \begin{cases} \binom{y}{x}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^x \left(1 - \frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{y-x} & \text{if } x = 0,\, 1,\, 2\, \ldots,\, y \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}$$

Thus, $X_1|X_1 + X_2 = y \sim Bin\left(y,\, \frac{\lambda_1}{\lambda_1+\lambda_2}\right)$. Note that the support of the conditional PMF is $\{0,\, 1,\, \ldots,\, y\}$. ||

Definition 3.18. *The conditional CDF of X given $Y = y$ is defined by*

$$F_{X|Y}(x|y) = P(X \leq x|Y = y) = \sum_{\{u \leq x : (u,y) \in S_{X,Y}\}} f_{X|Y}(u|y),$$

provided $f_Y(y) > 0$.

Definition 3.19 (Conditional Expectation for Discrete Random Vector)**.** *The conditional expectation of $h(X)$ given $Y = y$ is defined by*

$$E(h(X)|Y = y) = \sum_{\{x : (x,y) \in S_{X,Y}\}} h(x) f_{X|Y}(x|y),$$

provided it is absolutely summable.

Remark 3.8. Note that for fixed $y \in S_Y$, $f_{X|Y}(\cdot|y)$ is PMF. Thus, conditional expectation is an expectation with respect to the distribution specified by the PMF $f_{X|Y}(\cdot|y)$, and hence, conditional expectation satisfies all the properties of expectation. For example, if $h_1(x) \leq h_2(x)$ for all $x \in \mathbb{R}$, then

$$E\left(h_1(X)|Y = y\right) \leq E\left(h_2(X)|Y = y\right),$$

provided the expectations exist. †

Example 3.17. Let $X \sim Poi(\lambda_1)$, $Y \sim Poi(\lambda_2)$. Let X and Y be independent. In Example 3.16, we have seen that $X|X+Y = n \sim Bin(n,\, \frac{\lambda_1}{\lambda_1+\lambda_2})$ for all $n = 0,\, 1,\, \ldots$. Hence, the conditional expectation of X given $X + Y = n$ is $\frac{n\lambda_1}{\lambda_1+\lambda_2}$. ||

Example 3.18. Suppose that a system has n components. Suppose that on a rainy day, component i functions with probability p_i, $i = 1,\, 2,\, \ldots,\, n$. Also, assume that the components work independently. We want to calculate the conditional expected number of components that will function tomorrow given that it will rain tomorrow. Again we will use the indicator RVs as we used in Example 3.4 to count the number of components that will work tomorrow. Let

$$X_i = \begin{cases} 1 & \text{if component } i \text{ functions tomorrow} \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad Y = \begin{cases} 1 & \text{if it rains tomorrow} \\ 0 & \text{otherwise.} \end{cases}$$

Then the desired expectation can be obtained as follows:

$$E\left[\sum_{i=1}^{n} X_i|Y = 1\right] = \sum_{i=1}^{n} E\left(X_i|Y = 1\right) = \sum_{i=1}^{n} p_i.$$

The last equality is due to the fact that $P\left(X_i|Y = 1\right) = p_i$ for all $i = 1,\, 2,\, \ldots,\, n$. ||

Theorem 3.20. *If X and Y are independent DRVs, then $f_{X|Y}(x|y) = f_X(x)$ for all $x \in \mathbb{R}$ and $y \in S_Y$.*

Proof: The proof is straight forward from the definition of conditional PMF. □

3.9.2 For Continuous Random Vector

Let (X, Y) be a continuous random vector. Note that, in this case, Y is a CRV, and hence, $P(Y = y) = 0$ for all $y \in \mathbb{R}$. As a result, we cannot define the conditional probabilities in the same way as we defined for the discrete random vector in the previous subsection. Like CRV or continuous random vector, we will first define the conditional CDF for a continuous random vector (X, Y), then conditional PDF.

Definition 3.20 (Conditional CDF)**.** *Let (X, Y) be a continuous random vector. The conditional CDF of X given $Y = y$ is defined as*

$$F_{X|Y}(x|y) = \lim_{\epsilon \downarrow 0} P(X \leq x | Y \in (y-\epsilon, y+\epsilon]),$$

provided the limit exists.

Note that CDF of a random variable, X, is defined by $P(X \leq x)$. Ideally, we want to see what is the value of $P(X \leq x|Y = y)$. However, when (X, Y) is continuous random vector, we have a problem to define $P(X \leq x|Y = y)$ as $P(Y = y) = 0$ for all $y \in \mathbb{R}$. One of the way to overcome this difficulty is to proceed as follows. We can replace the event $\{Y = y\}$ by $Y \in (y - \epsilon,\, y + \epsilon]$ and then take limit that ϵ drops to zero. Of course, this makes sense if the limit exists. Motivated by this intuition, we have the previous definition of conditional CDF of X given $Y = y$ for a continuous random vector (X, Y).

Definition 3.21 (Conditional PDF)**.** *Let (X, Y) be a continuous random vector with conditional CDF $F_{X|Y}(\cdot|y)$ of X given $Y = y$. Define the conditional PDF of X given $Y = y$, $f_{X|Y}(x|y)$, as the non-negative integrable function satisfying*

$$F_{X|Y}(x|y) = \int_{-\infty}^{x} f_{X|Y}(t|y)dt \quad \text{for all } x \in \mathbb{R}.$$

Theorem 3.21. *Let $f_{X,Y}$ be the JPDF of (X,Y) and let f_Y be the marginal PDF of Y. If $f_Y(y) > 0$, then the conditional PDF of X given $Y = y$ exists and is given by*

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

Proof: This is not an exact proof, but a overview of it.

$$\begin{aligned}
\lim_{\varepsilon \downarrow 0} P(X \leq x | y - \varepsilon < Y \leq y + \varepsilon) &= \lim_{\varepsilon \downarrow 0} \frac{P(X \leq x,\, y - \varepsilon < Y \leq y + \varepsilon)}{P(y - \varepsilon < Y \leq y + \varepsilon)} \\
&= \lim_{\varepsilon \downarrow 0} \frac{\int_{-\infty}^{x} \int_{y-\varepsilon}^{y+\varepsilon} f_{X,Y}(t,\, s) ds dt}{\int_{y-\varepsilon}^{y+\varepsilon} f_Y(s) ds} \\
&= \int_{-\infty}^{x} \frac{\lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} \int_{y-\varepsilon}^{y+\varepsilon} f_{X,Y}(t,\, s) ds}{\lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} \int_{y-\varepsilon}^{y+\varepsilon} f_Y(s) ds} dt \\
&= \int_{-\infty}^{x} \frac{f_{X,Y}(t,\, y)}{f_Y(y)} dt.
\end{aligned}$$

The last equality is due to fundamental theorems of calculus. Thus, the conditional PDF is given by $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$ for those values of $y \in \mathbb{R}$ for which $f_Y(y) > 0$. $\square$

Definition 3.22 (Conditional Expectation for Continuous Random Vector)**.** *The conditional expectation of $h(X)$ given $Y = y$, is defined for all values of y such that $f_Y(y) > 0$, by*

$$E(h(X)|Y = y) = \int_{-\infty}^{\infty} h(x) f_{X|Y}(x|y) dx,$$

provided it is absolutely integrable.

Remark 3.9. Note that for fixed $y \in S_Y$, $f_{X|Y}(\cdot|y)$ is a PDF. Therefore, conditional expectation is an expectation with respect to the PDF $f_{X|Y}(\cdot|y)$. Thus, $E(X|Y = y)$ satisfies all properties of unconditional expectation. †

Example 3.19. Suppose the JPDF of (X, Y) is given by

$$f_{X,Y}(x, y) = \begin{cases} 6xy(2 - x - y) & 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

The marginal PDF of Y is

$$f_Y(y) = \begin{cases} \frac{1}{6} y(4 - 3y) & \text{if } 0 < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Hence, the conditional PDF of X given $Y = y \in (0,\ 1)$ is given by

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f_{X,Y}(x,\ y)}{f_Y(y)} \\ &= \begin{cases} \frac{6x(2-x-y)}{4-3y} & \text{if } 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

The conditional expectation of X given $Y = y \in (0,\ 1)$ is

$$E\left(X|Y = y\right) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx = \frac{6}{4 - 3y} \int_0^1 x^2 (2 - x - y) dx = \frac{5 - 4y}{2\left(4 - 3y\right)}.$$

Note that for conditional PDF, the ranges of both of x and y are important and need to mention unambiguously. Similarly for computing conditional expectation, we need the appropriate ranges of x and y. ||

Example 3.20. Let the joint PDF of $(X,\ Y)$ be

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{2} y e^{-xy} & \text{if } 0 < x < \infty,\ 0 < y < 2 \\ 0 & \text{otherwise.} \end{cases}$$

The marginal PDF of Y is

$$f_Y(y) = \begin{cases} \frac{1}{2} & \text{if } 0 < y < 2 \\ 0 & \text{otherwise.} \end{cases}$$

For $y \in (0,\ 2)$, the conditional PDF of X given $Y = y$ is

$$f_{X|Y}(x|y) = \begin{cases} y e^{-yx} & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Note that $Y \sim U(0,\, 2)$ and $X|Y = y \sim Exp(y)$ for $y \in (0,\, 2)$. Hence,

$$E\left(e^{\frac{X}{2}}|Y = 1\right) = \int_0^\infty e^{-x+\frac{x}{2}}dx = 2.$$

||

Theorem 3.22. *If $(X,\, Y)$ is a continuous random vector such that X and Y are independent random variables, then $f_{X|Y}(x|y) = f_X(x)$ for all $x \in \mathbb{R}$ and for all $y \in S_Y$.*

Proof: The proof is straight forward from the definition. □

3.9.3 Computing Expectation by Conditioning

Suppose that (X, Y) is either a discrete random vector or a continuous random vector. Then the conditional expectation $E(X|Y = y)$ is a function of y. Let we denote $g(y) = E(X|Y = y)$. Then $g(Y)$ is a function of RV Y. Thus, $g(Y) = E(X|Y)$ is again a random variable. With this understanding, we have the following theorem.

Theorem 3.23. *$E(X) = E(E(X|Y))$.*

Proof: We will prove it for continuous random vector $(X,\, Y)$. For the discrete random vector, the proof can be obtained by replacing the integration sign by summation sign.

$$\begin{aligned}
EE(X|Y) &= \int_{-\infty}^{\infty} E\left(X|Y = y\right) f_Y(y)dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X|Y}(x|y) f_Y(y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,\, y)dxdy \\
&= E(X).
\end{aligned}$$

□

In the previous theorem, the outside expectation is with respect to the distribution of Y, as $E\left(X|Y\right)$ is a function of Y. This theorem can be used to solve many problems. This theorem states that we can compute the average of different parts of a population separately and then take an weighted sum of those average to obtain the overall average. Let there be three columns of seating arrangement (like some of the lectures halls at IITG) in a class and we want to find the average of heights of the students in the class. Let $\overline{x}_i$ denote the average of the height of the students seating in the ith column and n_i be the number of students who is seating in ith column. Then the overall average is

$$\overline{x} = \frac{n_1}{n}\overline{x}_1 + \frac{n_2}{n}\overline{x}_2 + \frac{n_3}{n}\overline{x}_3,$$

where $n = n_1 + n_2 + n_3$. Note that $\frac{n_i}{n}$ can be interpreted as the probability that a student in column i and $\overline{x}_i$ is the conditional expectation of height given that the student is in the column i. Therefore, the overall average is $EE\left(X|Y\right)$, where X denotes the height of a student and Y is an indicator of the column number. We can take $Y = i$ if the student is in the ith column. Thus, the above theorem is a generalization of what we have learned in school.

Though we have discussed expectations when (X, Y) is either discrete random vector or continuous random vector, the above theorem is still valid if one of them is DRV and other is CRV. We will see some applications where one of them is CRV and other one is DRV. Of course, we will not go for proper definition (which is out of scope of this course) or proof in these cases. If Y is a DRV, then

$$E(X) = EE(X|Y) = \sum_{y \in S_Y} E(X|Y=y)\, f_Y(y).$$

If Y is a CRV, then

$$E(X) = EE(X|Y) = \int_{-\infty}^{\infty} E(X|Y=y)\, f_Y(y) dy.$$

Example 3.21. Virat will read either one chapter of his probability book or one chapter of his history book. If the number of misprints in a chapter of his probability and history book is Poisson with mean 2 and 5 respectively, then assuming that Virat is equally likely to choose either book, we can compute the expected number of misprints that he will come across using the above theorem. Let X denote the number of misprint and

$$Y = \begin{cases} 1 & \text{if Virat read the probability book} \\ 2 & \text{if Virat read the history book.} \end{cases}$$

We need to find $E(X)$. Note that $P(Y=1) = P(Y=2) = \frac{1}{2}$, $E(X|Y=1) = 2$ and $E(X|Y=2) = 5$. Hence,

$$E(X) = EE(X|Y) = P(Y=1)\, E(X|Y=1) + P(Y=2)\, E(X|Y=2) = \frac{1}{2}(2+5) = 3.5.$$

Thus, expected number of misprint that Virat will come across is 3.5. ||

Theorem 3.24. *$E(X - E(X|Y))^2 \leq E(X - f(Y))^2$ for any function f.*

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

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

Now,

$$\begin{aligned} E\left[(X - \mu(Y))(\mu(Y) - f(Y))\right] &= EE\left[(X - \mu(Y))\,(\mu(Y) - f(Y))\,|Y\right] \\ &= E\left[(\mu(Y) - f(Y)) E(X - \mu(Y)|Y)\right] \\ &= E\left[(\mu(Y) - f(Y))\,(\mu(Y) - \mu(Y))\right] \\ &= 0. \end{aligned}$$

The first equality is due to the Theorem 3.23. For the second equality, notice that $\mu(Y) - f(Y)$, being a function of Y only, acts as a constant when Y is given. Hence, $\mu(Y) - f(Y)$ comes out of the conditional expectation. Thus,

$$E(X - f(Y))^2 = E(X - \mu(Y))^2 + E(\mu(Y) - f(Y))^2 \geq E(X - \mu(Y))^2,$$

as $E(\mu(Y) - f(Y))^2 \geq 0$. The equality holds if and only if $E(\mu(Y) - f(Y))^2 = 0$. It can be shown that $E(\mu(Y) - f(Y))^2 = 0$ if and only if $f(Y) = \mu(Y) = E(X|Y)$. $\square$

Recall the Theorem 2.12, which states that if we do not have any extra information, then the "best estimate" of X is $E(X)$. The Theorem 3.24 states that if we have information on the RV Y, then the "best estimate" of X is changes and becomes $E(X|Y)$.

Definition 3.23 (Conditional Variance)**.** *Let (X, Y) be a random vector. Then the conditional variance of X given $Y = y$ is defined by*

$$Var(X|Y = y) = E((X - E(X|Y))^2|Y = y) = E(X^2|Y = y) - (E(X|Y = y))^2.$$

Like expectation, $Var(X|Y) = \sigma^2(Y)$ is a RV, where $\sigma^2(y) = Var(X|Y = y)$. Then we have the following theorem. The following theorem says that the overall variance can be computed by calculating variances and expectations of different parts and then aggregating.

Theorem 3.25. $Var(X) = E(Var(X|Y)) + Var(E(X|Y))$.

Proof: Let $\mu(Y) = E(X|Y)$ and $\mu = E(X) = EE(X|Y)$. Then

$$\begin{aligned} Var(X) &= E(X-\mu)^2 \\ &= E(X-\mu(Y))^2 + E(\mu(Y)-\mu)^2 + 2E[(X-\mu(Y))(\mu(Y)-\mu)]. \end{aligned}$$

Now, using Theorem 3.23,

$$E[(X-\mu(Y))(\mu(Y)-\mu)] = EE[(X-\mu(Y))(\mu(Y)-\mu)|Y] = 0.$$

Thus,

$$\begin{aligned} Var(X) &= E(X-\mu(Y))^2 + E(\mu(Y)-\mu)^2 \\ &= EE\left[(X-\mu(Y))^2|Y\right] + E[\mu(Y) - E(\mu(Y))]^2 \\ &= E(Var(X|Y)) + Var(\mu(Y)) \\ &= E(Var(X|Y)) + Var(E(X|Y)). \end{aligned}$$

Note that it is easy to remember the formula. On the right hand side, one is expectation of conditional variance and another is variance of conditional expectation. □

Example 3.22. Let $X_0, X_1, X_2, \ldots$ be a sequence of *i.i.d.* RVs with mean μ and variance σ^2. Let $N \sim Bin(n, p)$ and is independent of X_i's for all $i = 0, 1, \ldots$. Define $S = \sum_{i=0}^{N} X_i$. Note that the RV S is the sum of random number of RVs. This type of RVs are called compound RVs. Compound random variables are quite important in many practical situations. For example, consider a car insurance company. The number of accidents that a customer meets in a year is a RV. Let N denotes the number of accident in a year. Now, assume that X_i denotes the claim by the customer after the ith accident. Then S is the total claim made by the customer. Now, it is important for the insurance company to have an idea of average and variance of claims made by a customer. Let us try to compute $E(S)$ and $Var(S)$. Note that

$$E(S|N=n) = E\left(\sum_{i=0}^{N} X_i|N=n\right) = E\left(\sum_{i=0}^{n} X_i|N=n\right) = E\left(\sum_{i=0}^{n} X_i\right) = (n+1)\mu.$$

The second equality is true as under the condition $N = n$, the sum has n components. The third equality is true due to the fact that N and X_i's are independent. Note that $\sum_{i=1}^{N} X_i$

and N are not independent. However, when we put a specific value of N in $\sum_{i=1}^{N} X_i$ to get $\sum_{i=1}^{n} X_i$, then the later does not involve N and becomes independent of N. Thus, we have $E(S|N) = (N+1)\mu$. Hence,

$$E(S) = EE(S|N) = E[(N+1)\mu] = (np+1)\mu.$$

Now,

$$Var(S|N=n) = Var\left(\sum_{i=0}^{N} X_i | N = n\right) = Var\left(\sum_{i=0}^{n} X_i | N = n\right) = Var\left(\sum_{i=0}^{n} X_i\right) = (n+1)\sigma^2.$$

Thus, $Var(S|N) = (N+1)\sigma^2$. Hence,

$$\begin{aligned} Var(S) &= E\left(Var(S|N)\right) + Var\left(E(S|N)\right) \\ &= E\left[(N+1)\sigma^2\right] + Var\left[(N+1)\mu\right] \\ &= (np+1)\sigma^2 + np(1-p)\mu^2. \end{aligned}$$

||

Theorem 3.23 can be used to compute probability by conditioning. We have seen that $P(X \in A) = E(I_A(X))$, where I_A is the indicator function of the set A. Also, note that $E(I_A(X)|Y = y) = P(X \in A|Y = y)$. Therefore, we can write

$$\begin{aligned} P(A) = P(X \in A) &= E(I_A(X)) = EE(I_A(X)|Y) \\ &= \begin{cases} \displaystyle\sum_{y \in S_Y} P(A|Y=y)P(Y=y) & \text{for } Y \text{ discrete} \\ \displaystyle\int_{-\infty}^{\infty} P(A|Y=y)f_Y(y)dy & \text{for } Y \text{ continuous.} \end{cases} \end{aligned}$$

When Y is a DRV, $P(E) = \sum_{y \in S_Y} P(E|Y=y)P(Y=y)$ can be concluded from the Theorem 1.16. Of course, the result for CRV cannot be obtained form the Theorem 1.16.

Example 3.23. Let X and Y be independent CRVs having PDFs f_X and f_Y, respectively. Then

$$\begin{aligned} P(X < Y) &= \int_{-\infty}^{\infty} P(X < Y|Y = y) f_Y(y)dy \\ &= \int_{-\infty}^{\infty} P(X < y|Y = y) f_Y(y)dy \\ &= \int_{-\infty}^{\infty} P(X < y) f_Y(y)dy \\ &= \int_{-\infty}^{\infty} F_X(y)f_Y(y)dy, \end{aligned}$$

where $F_X(\cdot)$ is the CDF corresponding to $f_X(\cdot)$. ||

Example 3.24. Let X and Y be *i.i.d.* CRVs having common PDF $f(\cdot)$ and CDF $F(\cdot)$. Then using the last example

$$P(X < Y) = \int_{-\infty}^{\infty} F(y)f(y)dy = \frac{1}{2}.$$

Now, as X and Y are *i.i.d.*, $P(Y < X) = P(X > Y) = \frac{1}{2}$. Thus, $P(X = Y) = 1 - P(X < Y) - P(X > Y) = 0$. ||

Example 3.25. Suppose X and Y are two independent RVs, either discrete or continuous. Let us study the RV $Z = X + Y$ and try to see if this is a CRV or DRV

We know that (X, Y) is a discrete random vector if X and Y are DRVs, and hence, $Z = X + Y$ is a DRV. The PMF of Z, for $z \in \mathbb{R}$, is

$$\begin{aligned}
f_Z(z) &= P(X + Y = z) \\
&= \sum_{y \in S_y} P(X + Y = z | Y = y)\, P(Y = y) \\
&= \sum_{y \in S_Y} P(X + y = z | Y = y)\, f_Y(y) \\
&= \sum_{y \in S_Y} P(X = z - y)\, f_Y(y) \\
&= \sum_{y \in S_Y} f_X(z - y) f_Y(y).
\end{aligned}$$

Now, assume that X and Y are CRVs. Let us first find the CDF of Z and then check what type of RV Z is. The CDF of Z, for $z \in \mathbb{R}$, is

$$\begin{aligned}
F_Z(z) &= P(X + Y \le z) \\
&= \int_{-\infty}^{\infty} P(X + Y \le z | Y = y)\, f_Y(y) dy \\
&= \int_{-\infty}^{\infty} P(X + y \le z | Y = y)\, f_Y(y) dy \\
&= \int_{-\infty}^{\infty} P(X \le z - y)\, f_Y(y) dy, \quad \text{as } X \text{ and } Y \text{ are independent} \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f_X(x) f_Y(y) dx dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{z} f_X(x' - y) f_Y(y) dx' dy, \quad \text{taking } x' = x + y \\
&= \int_{-\infty}^{z} \left\{ \int_{-\infty}^{\infty} f_X(x' - y) f_Y(y) dy \right\} dx' \quad \text{for all } z \in \mathbb{R}.
\end{aligned}$$

Thus, $X + Y$ is a CRV with PDF

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(z - y) f_Y(y) dy \quad \text{for all } z \in \mathbb{R}.$$

Note that changing the role of X and Y, we can write the PDF of Z as

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_Y(z - x) f_X(x) dx \quad \text{for all } z \in \mathbb{R}.$$

Now, assume that X is a CRV and Y is a DRV. Then the CDF of $Z = X + Y$ is

$$\begin{aligned}
F_Z(z) &= P(X + Y \le z) \\
&= \sum_{y \in S_Y} P(X + Y \le z | Y = y)\, f_Y(y)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{y \in S_Y} \int_{-\infty}^{z-y} f_X(x) f_Y(y) dx \\
&= \sum_{y \in S_Y} \int_{-\infty}^{z} f_X(x'-y) f_Y(y) dx', \quad \text{taking } x' = x + y \\
&= \int_{-\infty}^{z} \left\{ \sum_{y \in S_Y} f_X(x'-y) f_Y(y) \right\} dx' \quad \text{for all } z \in \mathbb{R}.
\end{aligned}
$$

Therefore, $X + Y$ is a CRV with PDF

$$
f_{X+Y}(z) = \sum_{y \in S_Y} f_X(z-y) f_Y(y) \quad \text{for all } z \in \mathbb{R}.
$$

To summarize, if X and Y are independent, then the RV $X+Y$ is continuous if at least one of X or Y is CRV. If both of them are DRVs, then $X+Y$ is a DRV. ||

Definition 3.24 (Conditional Expectation for given Event)**.** *Let (X, Y) be a random vector. Then*

$$
E\left(h(X,\,Y) | (X,\,Y) \in A\right) = \frac{E(h(X,\,Y) I_A(X,\,Y))}{P\left((X,\,Y) \in A\right)}.
$$

Example 3.26. Let $X \sim Exp(1)$. Then

$$
E\left(X | X \geq 2\right) = \frac{E\left(X I_{[2,\,\infty)}(X)\right)}{P\left(X \geq 2\right)} = \frac{\displaystyle\int_0^\infty x I_{[2,\,\infty)}(x) e^{-x} dx}{\displaystyle\int_2^\infty e^{-x} dx} = e^2 \int_2^\infty x e^{-x} dx = 3.
$$

||

Example 3.27. $(X,\,Y)$ is uniform on unit square. Then

$$
E\left(X | X+Y > 1\right) = \frac{E\left(X I_{(1,\,\infty)}(X+Y)\right)}{P\left(X+Y > 1\right)} = \frac{\displaystyle\int_0^1 \int_{1-x}^1 x dy dx}{\displaystyle\int_0^1 \int_{1-y}^1 dx dy} = \frac{2}{3}.
$$

||

Example 3.28. A rod of length l is broken into two parts. Then the expected length of the shorter part is

$$
E\left(X | X < \frac{l}{2}\right),
$$

where $X \sim U(0,\,l)$. Thus, the expected length of the shorter part is

$$
E\left(X | X < \frac{l}{2}\right) = \frac{E\left(X I_{(-\infty,\,\frac{l}{2})}(X)\right)}{P\left(X < \frac{l}{2}\right)} = \frac{\displaystyle\frac{1}{l}\int_0^{\frac{l}{2}} x dx}{\displaystyle\frac{1}{l}\int_0^{\frac{l}{2}} dx} = \frac{l}{4}.
$$

An alternative formulation is as follows: The required quantity is $E\left[\min\left\{X,\,l-X\right\}\right]$. Please calculate and check if you are getting the same value. ||

3.10 Bivariate Normal Distribution

Recall that a CRV X is said to have a univariate normal distribution if the PDF of X is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \text{ for all } x \in \mathbb{R},$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$. In this case, $X \sim N(\mu, \sigma^2)$ is used to denote the RV X follows a normal distribution with parameters μ and σ^2. Note that if $X \sim N(\mu, \sigma^2)$, then all moments of X exist. In particular, $E(X)$ and $Var(X)$ exist, and they are given by $E(X) = \mu$ and $Var(X) = \sigma^2$. This means that a normal distribution is completely specified by its mean and variance.

Definition 3.25 (Bivariate Normal). *A two dimensional random vector* $\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ *is said to have a bivariate normal distribution if* $aX_1 + bX_2$ *is a univariate normal for all* $(a, b) \in \mathbb{R}^2 \setminus (0, 0)$.

Theorem 3.26. *If* $\boldsymbol{X}$ *has bivariate normal distribution, then each of* X_1 *and* X_2 *is univariate normal. Hence,* $E(X_1)$, $E(X_2)$, $Var(X_1)$, $Var(X_2)$, *and* $Cov(X_1, X_2)$ *exist.*

Proof: Taking $a = 1$ and $b = 0$, $aX_1 + bX_2 = X_1$ follows normal distribution. Similarly, X_2 follows normal distribution. As all moments of a normal RV exist, $E(X_1)$, $E(X_2)$, $Var(X_1)$, and $Var(X_2)$ exist. As $|Cov(X_1, X_2)| \leq \sqrt{Var(X_1)Var(X_2)}$ (see the proof of the Theorem 3.13), $Cov(X_1, X_2)$ exists. □

Let us denote $\boldsymbol{\mu} = E(\boldsymbol{X}) = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ and $\Sigma = Var(\boldsymbol{X}) = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$, where $\mu_1 = E(X_1)$, $\mu_2 = E(X_2)$, $\sigma_{11} = Var(X_1)$, $\sigma_{22} = Var(X_2)$, and $\sigma_{12} = \sigma_{21} = Cov(X_1, X_2)$.

Theorem 3.27. *Let* $\boldsymbol{X}$ *be a bivariate normal random vector. If* $\boldsymbol{\mu} = E(\boldsymbol{X})$ *and* $\Sigma = Var(\boldsymbol{X})$, *then for any fixed* $\boldsymbol{u} = (a, b) \in \mathbb{R}^2 \setminus (0, 0)$,

$$\boldsymbol{u}'\boldsymbol{X} \sim N(\boldsymbol{u}'\boldsymbol{\mu}, \boldsymbol{u}'\Sigma\boldsymbol{u}).$$

Proof: As $\boldsymbol{u}'\boldsymbol{X} = aX_1 + bX_2$, $\boldsymbol{u}'\boldsymbol{X}$ follows a univariate normal distribution. Now,

$$E(\boldsymbol{u}'\boldsymbol{X}) = a\mu_1 + b\mu_2 = \boldsymbol{u}'\boldsymbol{\mu}.$$

and

$$Var(\boldsymbol{u}'\boldsymbol{X}) = a^2\sigma_{11} + b^2\sigma_{22} + 2ab\sigma_{12} = \boldsymbol{u}'\Sigma\boldsymbol{u}.$$

Thus, $\boldsymbol{u}'\boldsymbol{X} \sim N(\boldsymbol{u}'\boldsymbol{\mu}, \boldsymbol{u}'\Sigma\boldsymbol{u})$. □

Theorem 3.28. *Let* $\boldsymbol{X}$ *be a bivariate normal random vector with* $\boldsymbol{\mu} = E(\boldsymbol{X})$ *and* $\Sigma = Var(\boldsymbol{X})$, *then the MGF of* $\boldsymbol{X}$ *is given by*

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = e^{\boldsymbol{t}'\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t}}$$

for all $\boldsymbol{t} \in \mathbb{R}^2$.

Proof: The JMGF of $\boldsymbol{X}$ is

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = E\left(e^{\boldsymbol{t}'\boldsymbol{X}}\right) = M_{\boldsymbol{t}'\boldsymbol{X}}(1). \tag{3.7}$$

As $\boldsymbol{X}$ has a bivariate normal distribution, $\boldsymbol{t}'\boldsymbol{X} \sim N\left(\boldsymbol{t}'\boldsymbol{\mu},\, \boldsymbol{t}'\Sigma\boldsymbol{t}\right)$. Now, using Example 2.42, the theorem is immediate. $\square$

The Theorem 3.28 shows that the bivariate normal distribution is completely specified by the mean vector $\boldsymbol{\mu}$ and the variance-covariance matrix Σ. We will use the notation $\boldsymbol{X} \sim N_2(\boldsymbol{\mu},\, \Sigma)$ to denote that the random vector $\boldsymbol{X}$ follows a bivariate normal distribution with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix Σ.

Theorem 3.29. *If $\boldsymbol{X} \sim N_2(\boldsymbol{\mu},\, \Sigma)$, then $X_1 \sim N(\mu_1,\, \sigma_{11})$ and $X_2 \sim N(\mu_2,\, \sigma_{22})$.*

Proof: The proof of the theorem is immediate from Theorem 3.27. $\square$

The converse of the Theorem 3.29 is not true in general. Consider the following example in this regard.

Example 3.29. Let $X \sim N(0,\, 1)$. Let Z be a DRV, which is independent of X and

$$P(Z = 1) = 0.5 = P(Z = -1).$$

Then $Y = ZX \sim N(0,\, 1)$. To see it, notice that for all $y \in \mathbb{R}$,

$$\begin{aligned}
P\left(Y \le y\right) &= P\left(ZX \le y\right) \\
&= P\left(ZX \le y | Z = 1\right) P\left(Z = 1\right) + P\left(ZX \le y | Z = -1\right) P\left(Z = -1\right) \\
&= \frac{1}{2} P\left(X \le y\right) + \frac{1}{2} P\left(X \ge -y\right) \\
&= \Phi(y).
\end{aligned}$$

Thus, $X \sim N(0,\, 1)$ and $Y \sim N(0,\, 1)$. However, $(X,\, Y)$ is not a bivariate normal random vector. To see it, observe that

$$P\left(X + Y = 0\right) = P\left(X + ZX = 0\right) = P(Z = -1) = \frac{1}{2}.$$

That means that $X + Y$ does not follow a univariate normal distribution, and hence, $(X,\, Y)$ is not a bivariate normal random vector, though $X \sim N(0,\, 1)$ and $Y \sim N(0,\, 1)$. ||

Theorem 3.30. *If $\boldsymbol{X} \sim N_2(\boldsymbol{\mu},\, \Sigma)$ and $Cov(X_1, X_2) = 0$, then X_1 and X_2 are independent.*

Proof: In this case, $\Sigma = diag(\sigma_{11},\, \sigma_{22})$. Hence, the JMGF of $(X_1,\, X_2)$ is

$$\begin{aligned}
M_{X_1,\, X_2}(t_1,\, t_2) &= e^{t_1\mu_1 + \frac{1}{2}\sigma_{11}t_1^2} \times e^{t_2\mu_2 + \frac{1}{2}\sigma_{22}t_2^2} \\
&= M_{X_1}(t_1) M_{X_2}(t_2),
\end{aligned}$$

where $M_{X_i}(\cdot)$ is the MGF of X_i, $i = 1,\, 2$. This shows that X_1 and X_2 are independent. $\square$

Note that we have discussed that two random variable can be dependent even if covariance between them zero. The bivariate normal random vector is special in this respect.

Theorem 3.31 (Probability Density Function). *Let $\boldsymbol{X} \sim N_2(\boldsymbol{\mu}, \Sigma)$ be such that Σ is invertible, then, for all $\boldsymbol{x} \in \mathbb{R}^2$, $\boldsymbol{X}$ has a joint PDF given by*

$$f(\boldsymbol{x}) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}$$
$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\, e^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right\}}$$

where $\sigma_1 = \sqrt{\sigma_{11}}$, $\sigma_2 = \sqrt{\sigma_{22}}$, ρ is correlation coefficient between X_1 and X_2.

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

Theorem 3.32 (Conditional Probability Density Function). *Let $\boldsymbol{X} \sim N_2(\boldsymbol{\mu}, \Sigma)$ be such that Σ is invertible, then for all $x_2 \in \mathbb{R}$, the conditional PDF of X_1 given $X_2 = x_2$ is given by*

$$f_{X_1|X_2}(x_1|x_2) = \frac{1}{\sigma_{1|2}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x_1-\mu_{1|2}}{\sigma_{1|2}}\right)^2\right] \quad \textit{for } x_1 \in \mathbb{R},$$

where $\mu_{1|2} = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2-\mu_2)$ and $\sigma^2_{1|2} = \sigma^2_1(1-\rho^2)$. Thus, $X_1|X_2 = x_2 \sim N\left(\mu_{1|2},\, \sigma^2_{1|2}\right)$.

Proof: Easy to see from the fact that

$$f_{X_1|X_2}(x_1|x_2) = \frac{f_{X_1,X_2}(x_1,\, x_2)}{f_{X_2}(x_2)}.$$

Of course, you need to perform some algebra. □

Corollary 3.1. *Under the condition of the Theorem 3.32, $E(X_1|X_2 = x_2) = \mu_{1|2} = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2-\mu_2)$ and $Var(X_1|X_2 = x_2) = \sigma^2_{1|2} = \sigma^2_1(1-\rho^2)$ for all $x_2 \in \mathbb{R}$. Hence, the conditional variance does not depend on x_2.*

Proof: Straight forward from the previous theorem. □

3.11 Some Results on Independent and Identically Distributed Normal RVs

Theorem 3.33. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. $N(0,1)$ random variables. Then*

$$\sum_{i=1}^{n} X_i^2 \sim Gamma(n/2, 1/2) \equiv \chi^2_n.$$

Proof: The MGF of X_1^2 is given by

$$M_{X_1^2}(t) = E\left(e^{tX_1^2}\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-(\frac{1}{2}-t)x^2}dx = (1-2t)^{-\frac{1}{2}},$$

for $t < \frac{1}{2}$. Hence, the MGF of $T = \sum_{i=1}^{n} X_i^2$

$$M_T(t) = \prod_{i=1}^{n} M_{X_i^2}(t) = (1-2t)^{-\frac{n}{2}},$$

where $t < \frac{1}{2}$. Thus, $T = \sum_{i=1}^{n} X_i^2 \sim Gamma(\frac{n}{2},\, \frac{1}{2})$. This distribution is also known as χ^2 distribution with degrees of freedom n. Thus, the sum of squares of n *i.i.d.* $N(0,\, 1)$ has a χ^2 distribution with degrees of freedom n. □

Theorem 3.34. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$ random variables. Let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$. Then $\overline{X}$ and S^2 are independently distributed and*

$$\overline{X} \sim N(\mu, \sigma^2/n) \quad \textit{and} \quad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Proof: Let A be an $n \times n$ orthogonal matrix, whose first row is

$$\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right).$$

Note that such a matrix exists as we can start with the row and construct a basis of $\mathbb{R}^n$. Then Gram-Schmidt orthogonalization will give us the required matrix. As A is orthogonal, its' inverse exists and $A^{-1} = A^T$, the transpose of A. Now consider the transformation of random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)'$ given by

$$\boldsymbol{Y} = A\boldsymbol{X}.$$

First, we shall try to find the distribution of $\boldsymbol{Y}$. Note that the transformation $g(\boldsymbol{x}) = A\boldsymbol{x}$ is a one-to-one transformation as A is invertible. The inverse transformation is given by $\boldsymbol{x} = A'\boldsymbol{y}$. Hence, the Jacobian of the inverse transformation is $J = det(A)$. As A is orthogonal, absolute value of $det(A)$ is one. Now, as $X_1, X_2, \ldots, X_n$ are *i.i.d.* $N(\mu, \sigma^2)$ RVs, the JPDF of $\boldsymbol{X}$, for $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)' \in \mathbb{R}^n$, is

$$\begin{aligned} f_{\boldsymbol{X}}(\boldsymbol{x}) &= \frac{1}{\left(\sigma\sqrt{2\pi}\right)^n} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right] \\ &= \frac{1}{\left(\sigma\sqrt{2\pi}\right)^n} \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{x} - \boldsymbol{\mu})'(\boldsymbol{x} - \boldsymbol{\mu})\right], \end{aligned}$$

where $\boldsymbol{\mu} = (\mu, \mu, \ldots, \mu)'$ is a n component vector. Thus, the JPDF of $\boldsymbol{Y}$, for $\boldsymbol{y} \in \mathbb{R}^n$, is

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}) &= f_{\boldsymbol{X}}(A'\boldsymbol{y}) \\ &= \frac{1}{\left(\sigma\sqrt{2\pi}\right)^n} \exp\left[-\frac{1}{2\sigma^2}(A'\boldsymbol{y} - \boldsymbol{\mu})'(A'\boldsymbol{y} - \boldsymbol{\mu})\right] \\ &= \frac{1}{\left(\sigma\sqrt{2\pi}\right)^n} \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{\eta})'(\boldsymbol{y} - \boldsymbol{\eta})\right], \end{aligned}$$

where $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_n)' = A\boldsymbol{\mu}$. Note that $\eta_1 = \sqrt{n}\mu$. Moreover,

$$\boldsymbol{\eta}'\boldsymbol{\eta} = \boldsymbol{\mu}'\boldsymbol{\mu} \implies \sum_{i=1}^{n}\eta_i^2 = n\mu^2 \implies \sum_{i=2}^{n}\eta_i^2 = n\mu^2 - \eta_1^2 = 0.$$

Thus, $\eta_i = 0$ for $i = 2, 3, \ldots, n$. Hence the JPDF of $\boldsymbol{Y}$ is

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}(y_1 - \sqrt{n}\mu)^2}\left\{\prod_{i=2}^{n}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{y_i^2}{2\sigma^2}}\right\} \quad \text{for } \boldsymbol{y} = (y_1, y_2, \ldots, y_n)' \in \mathbb{R}^n.$$

Therefore, $Y_1, Y_2, \ldots, Y_n$ are independent RVs and $Y_1 \sim N(\sqrt{n}\mu, \sigma^2)$ and $Y_i \sim N(0, \sigma^2)$ for $i = 2, 3, \ldots, n$, where $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)'$. Now,

$$Y_1 = \sqrt{n}\,\overline{X} \implies \sqrt{n}\,\overline{X} \sim N(\sqrt{n}\mu, \sigma^2) \implies \overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

Again,

$$\boldsymbol{Y}'\boldsymbol{Y} = \boldsymbol{X}'\boldsymbol{X} \implies \sum_{i=2}^{n} Y_i^2 = \sum_{i=1}^{n} X_i^2 - Y_1^2 = \sum_{i=1}^{n} X_i^2 - n\overline{X}^2 = (n-1)S^2.$$

For $i = 2, 3, \ldots, n$, $\frac{Y_i}{\sigma}$ are *i.i.d.* $N(0, 1)$ RVs. Thus, using the previous theorem

$$\frac{(n-1)S^2}{\sigma^2} = \sum_{i=2}^{n} \left(\frac{Y_i}{\sigma}\right)^2 \sim \chi^2_{n-1}.$$

Notice that $\overline{X}$ is a function of Y_1 only, and S^2 is a function of $Y_2, Y_3, \ldots, Y_n$. As Y_i's are independent, $\overline{X}$ and S^2 are independent. □