## 9. Week 9

*Remark* 9.1 (Conditional Distribution for discrete random vectors). Let $X = (X_1, X_2, \cdots, X_{p+q})$ be a discrete random vector with support $S_X$ and joint p.m.f. $f_X$. Let $Y = (X_1, X_2, \cdots, X_p)$ and $Z = (X_{p+1}, X_{p+2}, \cdots, X_{p+q})$. Then $Y$ and $Z$ both are discrete random vectors. Let $f_Y$ and $S_Y$ denote the joint p.m.f. and support of $Y$, respectively. Let $f_Z$ and $S_Z$ denote the joint p.m.f. and support of $Z$, respectively. For $z \in S_Z$, consider the set

$$T_z := \{y \in \mathbb{R}^p : (y, z) \in S_X\}.$$

The conditional p.m.f. of $Y$ given $Z = z \in S_Z$ is defined by

$$f_{Y|Z}(y \mid z) := \mathbb{P}(Y = y \mid Z = z) = \frac{\mathbb{P}(Y = y, Z = z)}{\mathbb{P}(Z = z)} = \begin{cases} \frac{f_X(y,z)}{f_Z(z)}, & \text{if } y \in T_z \\ 0, & \text{otherwise.} \end{cases}$$

By definition, $f_{Y|Z}(y \mid z) \geq 0, \forall y \in \mathbb{R}^p$ and $\sum_{y \in \mathbb{R}^p} f_{Y|Z}(y \mid z) = \sum_{y \in T_z} f_{Y|Z}(y \mid z) = 1$. Therefore, for every $z \in S_Z$, the function $y \in \mathbb{R}^p \mapsto f_{Y|Z}(y \mid z)$ is a joint p.m.f. with support $T_z$. We refer to the probability law/distribution described by this p.m.f. as the conditional distribution of $Y$ given $Z = z \in S_Z$. The conditional DF of $Y$ given $Z = z \in S_Z$ is given by

$$F_{Y|Z}(y \mid z) := \mathbb{P}(Y \leq y \mid Z = z) = \frac{\mathbb{P}(Y \leq y, Z = z)}{\mathbb{P}(Z = z)} = \sum_{\substack{t \leq y \\ t \in T_z}} \frac{f_X(t, z)}{f_Z(z)} = \sum_{\substack{t \leq y \\ t \in T_z}} f_{Y|Z}(t \mid z),$$

where $t \leq y$ refers to component-wise inequalities $t_j \leq y_j$ for all $j = 1, 2, \cdots, p$.

**Note 9.2.** For notational convenience, we have discussed the conditional distribution of first $p$ component RVs with respect to the final $q$ component RVs. However, as long as the $(p+q)$-dimensional joint distribution is known, we can discuss the conditional distribution of any of the $k$-component RVs with respect to the other $(p+q-k)$-component RVs.

**Note 9.3.** When values for some of the components RVs are given, the conditional distribution provides an updated probability distribution for the rest of the component RVs.

**Note 9.4.** Let $(X, Y)$ be a 2-dimensional discrete random vector such that $X$ and $Y$ are independent. Then $f_{X,Y}(x, y) = f_X(x) f_Y(y), \forall x, y \in \mathbb{R}$. Then

$$f_{Y|X}(y \mid x) = f_Y(y), \forall x \in S_X, y \in S_Y.$$

This statement can be generalized to higher dimensions with appropriate changes in the notation.

**Example 9.5.** In Example 8.51, we have, for fixed $x \in \{1, 2, 3, 4\}$,

$$f_{Y|X}(y \mid x) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_X(x)}, & \text{if } y \in \{1,2,3,4\} \\ 0, & \text{otherwise.} \end{cases} = \begin{cases} \frac{x+y}{2(2x+5)}, & \text{if } y \in \{1,2,3,4\} \\ 0, & \text{otherwise.} \end{cases}$$

**Definition 9.6** (Continuous Random Vector and its Joint Probability Density Function (Joint p.d.f.))**.** A random vector $X = (X_1, X_2, \cdots, X_p)$ is said to be a continuous random vector if there exists an integrable function $f : \mathbb{R}^p \to [0, \infty)$ such that

$$\begin{aligned} F_X(x) &= \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \cdots, X_p \leq x_p) \\ &= \int_{t_1=-\infty}^{x_1} \int_{t_2=-\infty}^{x_2} \cdots \int_{t_p=-\infty}^{x_p} f(t_1, t_2, \cdots, t_p)\, dt_p dt_{p-1} \cdots dt_2 dt_1, \forall x = (x_1, x_2, \cdots, x_p) \in \mathbb{R}^p. \end{aligned}$$

The function $f$ is called the joint probability density function (joint p.d.f.) of $X$.

*Remark* 9.7. Let $X$ be a continuous random vector with joint DF $F_X$ and joint p.d.f. $f_X$. Then we have the following observations.

(a) $F_X$ is jointly continuous in all co-ordinates.

(b) $\mathbb{P}(X = x) = 0, \forall x \in \mathbb{R}^p$. More generally, if $A \subset \mathbb{R}^p$ is finite or countably infinite, then by the finite/countable additivity of $\mathbb{P}_X$, we have

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

(c) By definition, we have $f_X(x) \geq 0, \forall x \in \mathbb{R}^p$ and

$$\begin{aligned} 1 &= \lim_{x_j \to \infty \, \forall j} F_X(x_1, x_2, \cdots, x_p) \\ &= \lim_{x_j \to \infty \, \forall j} \int_{t_1=-\infty}^{x_1} \int_{t_2=-\infty}^{x_2} \cdots \int_{t_p=-\infty}^{x_p} f(t_1, t_2, \cdots, t_p)\, dt_p dt_{p-1} \cdots dt_2 dt_1 \end{aligned}$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(t_1, t_2, \cdots, t_p)\, dt_p dt_{p-1} \cdots dt_2 dt_1.$$

(d) Suppose that the joint p.d.f. $f_X$ of a $p$-dimensional random vector $X$ is piecewise continuous. Then by the Fundamental Theorem of Calculus (specifically multivariable Calculus), we have

$$f_X(x_1, x_2, \cdots, x_p) = \frac{\partial^p}{\partial x_1 \partial x_2 \cdots \partial x_p} F_X(x_1, x_2, \cdots, x_p),$$

wherever the partial derivative on the right hand side exists.

(e) If $X$ is a $p$-dimensional random vector such that its joint DF $F_X$ is continuous on $\mathbb{R}^p$ and such that the partial derivative $\frac{\partial^p}{\partial x_1 \partial x_2 \cdots \partial x_p} F_X$ exists everywhere except possibly on a countable number of curves on $\mathbb{R}^p$. Let $A \subset \mathbb{R}^p$ denote the set of all points on such curves. Then $X$ is a continuous random vector with the joint p.d.f.

$$f_X(x) = \begin{cases} \frac{\partial^p}{\partial x_1 \partial x_2 \cdots \partial x_p} F_X(x), \text{ if } x = (x_1, x_2, \cdots, x_p) \in A^c, \\ 0, \text{ if } x = (x_1, x_2, \cdots, x_p) \in A. \end{cases}$$

(f) The joint p.d.f. of a continuous random vector is not unique. As in the case of continuous RVs, the joint p.d.f. is determined uniquely upto sets of 'volume 0'. Here, we also get versions of the joint p.d.f..

(g) For $A \subset \mathbb{R}^p$, we have

$$\begin{aligned} \mathbb{P}(X \in A) &= \iiint_A f_X(t_1, t_2, \cdots, t_p)\, dt_p dt_{p-1} \cdots dt_2 dt_1 \\ &= \iiint_{\mathbb{R}^p} f_X(t_1, t_2, \cdots, t_p) 1_A(t_1, t_2, \cdots, t_p)\, dt_p dt_{p-1} \cdots dt_2 dt_1, \end{aligned}$$

provided the integral can be defined. We do not prove this statement in this course.

(h) For any $j \in \{1, 2, \cdots, p\}$, for $x_j \in \mathbb{R}$

$$\begin{aligned} F_{X_j}(x_j) &= \mathbb{P}(X_j \in (-\infty, x_j]) \\ &= \mathbb{P}(X_1 \in \mathbb{R}, \cdots, X_{j-1} \in \mathbb{R}, X_j \in (-\infty, x_j], X_{j+1} \in \mathbb{R}, \cdots, X_p \in \mathbb{R}) \\ &= \mathbb{P}(X \in \mathbb{R} \times \cdots \times \mathbb{R} \times (-\infty, x_j] \times \mathbb{R} \times \cdots \times \mathbb{R}) \end{aligned}$$

$$= \int_{t_1=-\infty}^{\infty} \cdots \int_{t_{j-1}=-\infty}^{\infty} \int_{t_j=-\infty}^{x_j} \int_{t_{j+1}=-\infty}^{\infty} \cdots \int_{t_p=-\infty}^{\infty} f_X(t_1, t_2, \cdots, t_p)\, dt_p dt_{p-1} \cdots dt_2 dt_1.$$

Consider $g_j : \mathbb{R} \to \mathbb{R}$ defined by

$$g_j(t_j) := \int_{t_1=-\infty}^{\infty} \cdots \int_{t_{j-1}=-\infty}^{\infty} \int_{t_{j+1}=-\infty}^{\infty} \cdots \int_{t_p=-\infty}^{\infty} f_X(t_1, t_2, \cdots, t_p)\, dt_p \cdots dt_{j-1} dt_{j+1} \cdots dt_2 dt_1.$$

It is immediate that $g_j$ satisfies the properties of a p.d.f. and $F_{X_j}(x_j) = \int_{t_j=-\infty}^{x_j} g_j(t_j)\, dt_j$. Therefore, $X_j$ is a continuous RV with p.d.f. $g_j$. More generally, all marginal distributions of $X$ are also continuous and can be obtained by integrating out the unnecessary co-ordinates. The function $g_j$ is usually referred to as the marginal p.d.f. of $X_j$.

*Remark* 9.8. Let $f : \mathbb{R}^p \to [0, \infty)$ be an integrable function with

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(t_1, t_2, \cdots, t_p)\, dt_p dt_{p-1} \cdots dt_2 dt_1 = 1.$$

Then $f$ is the joint p.d.f. of some $p$-dimensional continuous random vector $X$. We are not going to discuss the proof of this statement in this course.

We can identify the independence of the component RVs for a continuous random vector via the joint p.d.f.. The proof is similar to Theorem 8.48 and is skipped for brevity.

**Theorem 9.9.** *Let $X = (X_1, X_2, \cdots, X_p)$ be a continuous random vector with joint DF $F_X$, joint p.d.f. $f_X$. Let $f_{X_j}$ denote the marginal p.d.f. of $X_j$. Then $X_1, X_2, \cdots, X_p$ are independent if and only if*

$$f_{X_1, X_2, \cdots, X_p}(x_1, x_2, \cdots, x_p) = \prod_{j=1}^{p} f_{X_j}(x_j), \forall x_1, x_2, \cdots, x_p \in \mathbb{R}.$$

**Example 9.10.** Given p.d.f.s $f_1, f_2, \cdots, f_p : \mathbb{R} \to [0, \infty)$, consider the function $f : \mathbb{R}^p \to [0, \infty)$ defined by

$$f(x) := \prod_{j=1}^{p} f_j(x_j), \forall x = (x_1, x_2, \cdots, x_p) \in \mathbb{R}^p.$$

Then

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(t_1, t_2, \cdots, t_p)\, dt_p dt_{p-1} \cdots dt_2 dt_1 = 1.$$

By Remark 9.8, we have that $f$ is the joint p.d.f. of a $p$-dimensional continuous random vector such that the component RVs are independent, by Theorem 9.9. Using this method, we can construct many examples of continuous random vectors.

*Remark* 9.11. Let $X = (X_1, X_2, \cdots, X_p)$ be a continuous random vector with joint p.d.f. $f_X$. Then $X_1, X_2, \cdots, X_p$ are independent if and only if

$$f_{X_1,X_2,\cdots,X_p}(x_1, x_2, \cdots, x_p) = \prod_{j=1}^{p} g_j(x_j), \forall x_1, x_2, \cdots, x_p \in \mathbb{R}$$

for some integrable functions $g_1, g_2, \cdots, g_p : \mathbb{R} \to [0, \infty)$. In this case, the marginal p.d.fs $f_{X_j}$ have the form $c_j g_j$, where the number $c_j$ can be determined from the relation $c_j = \left(\int_{-\infty}^{\infty} g_j(x)\, dx\right)^{-1}$.

**Example 9.12.** Let $Z = (X, Y)$ be a 2-dimensional continuous random vector with the joint p.d.f. of the form

$$f_Z(x, y) = \begin{cases} \alpha xy, & \text{if } 0 < x < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

for some constant $\alpha \in \mathbb{R}$. For $f_Z$ to take non-negative values, we must have $\alpha > 0$. Now,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_Z(x, y)\, dxdy = \int_{y=0}^{1}\int_{x=0}^{y} \alpha xy\, dxdy = \int_{y=0}^{1} \alpha \frac{y^3}{2}\, dy = \frac{\alpha}{8}.$$

For $f_Z$ to be a joint p.d.f., we need $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_Z(x, y)\, dxdy = 1$ and hence $\alpha = 8 > 0$. Also note that for this value of $\alpha$, $f_Z$ takes non-negative values. The marginal p.d.f. $f_X$ of $X$ can now be computed as follows.

$$f_X(x) = \int_{-\infty}^{\infty} f_Z(x, y)\, dy = \begin{cases} \int_{y=x}^{1} 8xy\, dy, & \text{if } x \in (0, 1) \\ 0, & \text{otherwise} \end{cases} = \begin{cases} 4x[1 - x^2], & \text{if } x \in (0, 1) \\ 0, & \text{otherwise} \end{cases}.$$

The marginal p.d.f. $f_Y$ of $Y$ follows by a similar computation.

$$f_Y(y) = \int_{-\infty}^{\infty} f_Z(x, y)\, dx = \begin{cases} \int_{x=0}^{y} 8xy\, dx, & \text{if } y \in (0, 1) \\ 0, & \text{otherwise} \end{cases} = \begin{cases} 4y^3, & \text{if } y \in (0, 1) \\ 0, & \text{otherwise} \end{cases}.$$

Observe that $f_Z(\frac{1}{2}, \frac{1}{2}) = 0$ and $f_X(\frac{1}{2})f_Y(\frac{1}{2}) = \frac{3}{2} \times \frac{1}{2} = \frac{3}{4}$. Hence $X$ and $Y$ are not independent.

**Example 9.13.** Let $U = (X, Y, Z)$ be a 3-dimensional continuous random vector with the joint p.d.f. of the form

$$f_U(x, y, z) = \begin{cases} \alpha xyz, & \text{if } x, y, z \in (0, 1) \\ 0, & \text{otherwise} \end{cases}$$

for some constant $\alpha \in \mathbb{R}$. For $f_Z$ to take non-negative values, we must have $\alpha > 0$. Now,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_U(x, y, z)\, dxdydz = \int_{x=0}^{1}\int_{y=0}^{1}\int_{z=0}^{1} \alpha xyz\, dxdydz = \frac{\alpha}{8}.$$

For $f_U$ to be a joint p.d.f., we need $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_U(x, y, z)\, dxdydz = 1$ and hence $\alpha = 8 > 0$. Also note that for this value of $\alpha$, $f_U$ takes non-negative values. The marginal p.d.f. $f_X$ of $X$ can now be computed as follows.

$$f_X(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_U(x, y, z)\, dydz = \begin{cases} \int_{z=0}^{1}\int_{y=0}^{1} 8xyz\, dydz, & \text{if } x \in (0, 1) \\ 0, & \text{otherwise} \end{cases} = \begin{cases} 2x, & \text{if } x \in (0, 1) \\ 0, & \text{otherwise} \end{cases}.$$

By the symmetry of $f_U(x, y, z)$ in the variables $x, y$ and $z$, we conclude that $X \stackrel{d}{=} Y \stackrel{d}{=} Z$. Observe that $f_{X,Y,Z}(x, y, z) = f_X(x)f_Y(y)f_Z(z), \forall x, y, z$ and hence the RVs $X, Y, Z$ are independent.

**Note 9.14.** There are random vectors which are neither discrete nor continuous. We do not discuss such examples in this course.

*Remark* 9.15 (Conditional Distribution for continuous random vectors). We now discuss an analogue of conditional distributions as discussed in Remark 9.1 for discrete random vectors. To avoid notational complexity, we work in dimension 2. Let $(X, Y)$ be a 2-dimensional continuous random vector with joint DF $F_{X,Y}$ and joint p.d.f. $f_{X,Y}$. Let $f_X$ and $f_Y$ denote the marginal p.d.fs of $X$ and $Y$ respectively. Since $\mathbb{P}(X = x) = 0, \forall x \in \mathbb{R}$, expressions of the form $\mathbb{P}(Y \in A \mid X = x)$ are not defined for $A \subset \mathbb{R}$. We consider $x \in \mathbb{R}$ such that $f_X(x) > 0$ and look at the following computation. For $y \in \mathbb{R}$,

$$\lim_{h \downarrow 0} \mathbb{P}(Y \leq y \mid x - h < X \leq x) = \lim_{h \downarrow 0} \frac{\mathbb{P}(Y \leq y, x - h < X \leq x)}{\mathbb{P}(x - h < X \leq x)}$$

$$= \lim_{h \downarrow 0} \frac{\int_{x-h}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s)\, dsdt}{\int_{x-h}^{x} f_X(t)\, dt}$$
$$= \lim_{h \downarrow 0} \frac{\frac{1}{h}\int_{x-h}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s)\, dsdt}{\frac{1}{h}\int_{x-h}^{x} f_X(t)\, dt}$$
$$= \frac{\int_{-\infty}^{y} f_{X,Y}(x,s)\, ds}{f_X(x)}$$
$$= \int_{-\infty}^{y} \frac{f_{X,Y}(x,s)}{f_X(x)}\, ds$$

Here, we have assumed continuity of the p.d.fs. Motivated by the above computation, we define the conditional DF of $Y$ given $X = x$ (provided $f_X(x) > 0$) by

$$F_{Y|X}(y \mid x) := \lim_{h \downarrow 0} \mathbb{P}(Y \le y \mid x - h < X \le x),\ y \in \mathbb{R}$$

and the conditional p.d.f. of $Y$ given $X = x$ (provided $f_X(x) > 0$) by

$$f_{Y|X}(y \mid x) := \frac{f_{X,Y}(x,y)}{f_X(x)},\ y \in \mathbb{R}.$$

These calculations generalizes to the higher dimensions as follows. Let $X = (X_1, X_2, \cdots, X_{p+q})$ be a continuous random vector with joint p.d.f. $f_X$. Let $Y = (X_1, X_2, \cdots, X_p)$ and $Z = (X_{p+1}, X_{p+2}, \cdots, X_{p+q})$. If $z \in \mathbb{R}^q$ be such that $f_Z(z) > 0$, then we define the conditional DF of $Y$ given $Z = z$ by

$$F_{Y|Z}(y \mid z) := \lim_{\substack{h_j \downarrow 0 \\ j = p+1, p+2, \cdots, p+q}} \mathbb{P}(X_1 \le y_1, \cdots, X_p \le y_p \mid x_j - h_j < X_j \le x_j, \forall j),\ y \in \mathbb{R}^p$$

and the conditional p.d.f. of $Y$ given $Z = z$ by

$$f_{Y|Z}(y \mid z) := \frac{f_{Y,Z}(y,z)}{f_Z(z)},\ y \in \mathbb{R}^p.$$

**Note 9.16.** For notational convenience, we have discussed the conditional distribution of first $p$ component RVs with respect to the final $q$ component RVs. However, as long as the $(p+q)$-dimensional joint distribution is known, we can discuss the conditional distribution of any of the $k$-component RVs with respect to the other $(p+q-k)$-component RVs.

**Note 9.17.** Let $(X, Y)$ be a 2-dimensional continuous random vector such that $X$ and $Y$ are independent. Then $f_{X,Y}(x, y) = f_X(x) f_Y(y), \forall x, y \in \mathbb{R}$. Then

$$f_{Y|X}(y \mid x) = f_Y(y), \forall y \in \mathbb{R},$$

provided $f_X(x) > 0$. This statement can be generalized to higher dimensions with appropriate changes in the notation.

**Example 9.18.** In Example 9.12, we have, for fixed $x \in (0, 1)$,

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

Earlier, we have discussed about the distribution of functions of RVs. We now generalize the same concept for random vectors.

*Remark* 9.19. Let $X = (X_1, \ldots, X_p)$ be a $p$-dimensional discrete/continuous random vector with joint p.m.f./p.d.f. $f_X$. We are interested in the distribution of $Y = h(X)$ for functions $h : \mathbb{R}^p \to \mathbb{R}^q$. Here, $Y = (Y_1, \ldots, Y_q)$ is a $q$-dimensional random vector with $Y_j = h_j(X_1, \ldots, X_p)$, where $h_j : \mathbb{R}^p \to \mathbb{R}, j = 1, 2, \cdots, q$ denotes the component functions of $h$. The distribution of $Y$ is uniquely determined as soon as we are able to compute the joint DF $F_Y$ of $Y$. Note that

$$F_Y(y_1, \cdots, y_q) = \mathbb{P}(Y_1 \le y_1, \cdots, Y_q \le y_q) = \mathbb{P}(h_1(X) \le y_1, \cdots, h_q(X) \le y_q), \forall (y_1, \cdots, y_q) \in \mathbb{R}^q.$$

Once the joint DF $F_Y$ is known, the joint p.m.f./p.d.f. of $Y$ can then be deduced by standard techniques.

**Example 9.20.** Let $X_1 \sim Uniform(0, 1)$ and $X_2 \sim Uniform(0, 1)$ be independent RVs. Suppose we are interested in the distribution of $Y = X_1 + X_2$. By independence of $X_1$ and $X_2$, the joint p.d.f. $(X_1, X_2)$ is given by

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

Consider the function $h : \mathbb{R}^2 \to \mathbb{R}$ defined by $h(x_1, x_2) := x_1 + x_2, \forall (x_1, x_2) \in \mathbb{R}^2$. Then $Y = h(X_1, X_2)$. Now, for $y \in \mathbb{R}$

$$\begin{aligned}
F_Y(y) &= \mathbb{P}(Y \le y) \\
&= \mathbb{P}(h(X_1, X_2) \le y) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 1_{(-\infty, y]}(h(x_1, x_2)) f_{X_1, X_2}(x_1, x_2)\, dx_1 dx_2 \\
&= \int_0^1 \int_0^1 1_{(-\infty, y]}(x_1 + x_2)\, dx_1 dx_2 \\
&= \begin{cases} 0, \text{ if } y < 0, \\ \int_{x_1=0}^{y} \int_{x_2=0}^{y-x_1} dx_2 dx_1, \text{ if } 0 \le y < 1, \\ 1 - \frac{1}{2} \times (2-y) \times (2-y), \text{ if } 1 \le y < 2, \\ 1, \text{ if } y \ge 2 \end{cases} \\
&= \begin{cases} 0, \text{ if } y < 0, \\ \frac{y^2}{2}, \text{ if } 0 \le y < 1, \\ \frac{4y - y^2 - 2}{2}, \text{ if } 1 \le y < 2, \\ 1, \text{ if } y \ge 2 \end{cases}
\end{aligned}$$

Here, $F_Y$ is differentiable everywhere except possibly at the points $0, 1, 2$ and

$$F_Y'(y) = \begin{cases} y, \text{ if } y \in (0, 1), \\ 2 - y, \text{ if } y \in (1, 2), \\ 0, \text{ otherwise.} \end{cases}$$

Observe that $\int_{-\infty}^{\infty} F_Y'(y)\, dy = 1$ and the derivative is non-negative. Hence, $Y$ is a continuous RV with the p.d.f. given by $F_Y'$.

As done in the case of RVs, in the setting of Remark 9.19, we consider the computation of the joint p.m.f./p.d.f. of $Y$ directly, instead of computing the joint DF $F_Y$ first. The next result is a direct generalization of Theorem 5.21 and we skip the proof for brevity.

**Theorem 9.21** (Change of Variables for Discrete random vectors). *Let $X = (X_1, \ldots, X_p)$ be a $p$-dimensional discrete random vector with joint p.m.f. $f_X$ and support $S_X$. Let $h = (h_1, \cdots, h_q) : \mathbb{R}^p \to \mathbb{R}^q$ be a function and let $Y = (Y_1, \cdots, Y_q) = h(X) = (h_1(X), \cdots, h_q(X))$. Then $Y$ is a discrete random vector with support*

$$S_Y = h(S_X) = \{h(x) : x \in S_X\},$$

*joint p.m.f.*

$$f_Y(y) = \begin{cases} \displaystyle\sum_{\substack{x \in S_X \\ h(x)=y}} f_X(x), & \text{if } y \in S_Y, \\ 0, & \text{otherwise} \end{cases}$$

*and joint DF*

$$F_Y(y) = \sum_{\substack{x \in S_X \\ h(x) \leq y}} f_X(x), \forall y \in \mathbb{R}^q.$$

**Example 9.22.** Fix $p \in (0, 1)$ and let $n_1, \cdots, n_q$ be positive integers. Let $X_1, \cdots, X_q$ be independent RVs with $X_i \sim Binomial(n_i, p), i = 1, \cdots, q$. Here, the with p.m.f.s are given by

$$f_{X_i}(x_i) = \begin{cases} \binom{n_i}{x} p^x (1-p)^{n_i - x}, \forall x \in \{0, 1, \cdots, n_i\}, \\ 0, \text{ otherwise} \end{cases}$$

for $i = 1, \cdots, q$. Using independence, the joint p.m.f. is given by

$$f_X(x_1, \cdots, x_q) = \begin{cases} \displaystyle\prod_{i=1}^{q} \binom{n_i}{x_i} p^{\sum_{i=1}^q x_i} (1-p)^{n - \sum_{i=1}^q x_i}, \forall (x_1, \cdots, x_q) \in \prod_{i=1}^{q} \{0, 1, \cdots, n_i\}, \\ 0, \text{ otherwise} \end{cases}$$

where $n = n_1 + \cdots + n_q$. Consider $Y = X_1 + \cdots + X_q$. Now, if $y \notin \{0, 1, \cdots, n\}$, $f_Y(y) = \mathbb{P}(X_1 + \cdots + X_q = y) = 0$ and if $y \in \{0, 1, \cdots, n\}$, then

$$\begin{aligned} f_Y(y) &= \mathbb{P}(X_1 + \cdots + X_q = y) \\ &= \sum_{\substack{(x_1, \cdots, x_q) \in \prod_{i=1}^q \{0,1,\cdots,n_i\} \\ x_1 + \cdots + x_q = y}} f_X(x_1, \cdots, x_q) \end{aligned}$$

$$= p^y(1-p)^{n-y} \sum_{\substack{(x_1,\cdots,x_q)\in\prod_{i=1}^q\{0,1,\cdots,n_i\} \\ x_1+\cdots+x_q=y}} \prod_{i=1}^q \binom{n_i}{x_i}$$

$$= \binom{n}{y} p^y(1-p)^{n-y}.$$

Therefore, $Y = X_1 + \cdots + X_q \sim Binomial(n,p)$ with $n = n_1 + \cdots + n_q$.

*Remark* 9.23. We had earlier mentioned that $Bernoulli(p)$ distribution is the same as $Binomial(1,p)$ distribution. Using the above computation, we can identify a $Binomial(n,p)$ RV as a sum of $n$ independent RVs each having distribution $Bernoulli(p)$. We shall come back to this observation in later lectures.

For continuous random vectors, we have the following generalization of Theorem 6.1. Proof of this result is being skipped.

**Theorem 9.24.** *Let $X = (X_1,\ldots,X_p)$ be a $p$-dimensional continuous random vector with joint p.d.f. $f_X$. Suppose that $\{x \in \mathbb{R}^p : f_X(x) > 0\}$ can be written as a disjoint union $\cup_{i=1}^k S_i$ of open sets in $\mathbb{R}^p$.*

*Let $h^j : \mathbb{R}^p \to \mathbb{R}, j = 1,\cdots,p$ be functions such that $h = (h^1,\cdots,h^p) : S_i \to \mathbb{R}^p$ is one-to-one with inverse $h_i^{-1} = ((h_i^1)^{-1},\cdots,(h_i^p)^{-1})$ for each $i = 1,\cdots,k$. Moreover, assume that $(h_i^j)^{-1}, i = 1,2,\cdots,k; j = 1,\cdots,p$ have continuous partial derivatives and the Jacobian determinant of the transformation*

$$J_i := \begin{vmatrix} \frac{\partial (h_i^1)^{-1}}{\partial y_1}(t) & \cdots & \frac{\partial (h_i^1)^{-1}}{\partial y_p}(y) \\ \vdots & \vdots & \vdots \\ \frac{\partial (h_i^p)^{-1}}{\partial y_1}(y) & \cdots & \frac{\partial (h_i^p)^{-1}}{\partial y_p}(y) \end{vmatrix} \neq 0, \forall i = 1,\cdots,k.$$

*Then the $p$-dimensional random vector $Y = (Y_1,\cdots,Y_p) = h(X) = (h^1(X),\cdots,h^p(X))$ is a continuous with joint p.d.f.*

$$f_Y(y) = \sum_{i=1}^k f_X((h_i^1)^{-1}(y),\cdots,(h_i^p)^{-1}(y))\,|J_i|\,1_{h(S_i)}(y).$$

**Example 9.25.** Fix $\lambda > 0$. Let $X_1 \sim Exponential(\lambda)$ and $X_2 \sim Exponential(\lambda)$ be independent RVs defined on the same probability space. The joint distribution of $(X_1,X_2)$ is given by the joint p.d.f.

$$f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1)f_{X_2}(x_2) = \begin{cases} \frac{1}{\lambda^2}\exp(-\frac{x_1+x_2}{\lambda}), \text{ if } x_1 > 0, x_2 > 0 \\ 0, \text{ otherwise.} \end{cases}$$

Consider the function

$$h(x_1,x_2) = \begin{cases} (x_1+x_2, \frac{x_1}{x_1+x_2}), \forall x_1 > 0, x_2 > 0, \\ 0, \text{otherwise.} \end{cases}$$

Here, $\{(x_1,x_2) \in \mathbb{R}^2 : f_{X_1,X_2}(x_1,x_2) > 0\} = \{(x_1,x_2) \in \mathbb{R}^2 : x_1 > 0, x_2 > 0\}$ and $h : \{(x_1,x_2) \in \mathbb{R}^2 : x_1 > 0, x_2 > 0\} \to \mathbb{R}^2$ is one-to-one with range $(0,\infty) \times (0,1)$. The inverse function is $h^{-1}(y_1,y_2) = (y_1y_2, y_1(1-y_2))$ for $(y_1,y_2) \in (0,\infty) \times (0,1)$ with Jacobian determinant given by

$$J(y_1,y_2) = \begin{vmatrix} y_2 & y_1 \\ 1-y_2 & -y_1 \end{vmatrix} = -y_1.$$

Now, $Y = (Y_1,Y_2) = h(X_1,X_2) = (X_1+X_2, \frac{X_1}{X_1+X_2})$ has the joint p.d.f. given by

$$\begin{aligned} f_{Y_1,Y_2}(y_1,y_2) &= \begin{cases} f_{X_1,X_2}(y_1y_2, y_1(1-y_2))|J(y_1,y_2)|, \text{ if } y_1 > 0, y_2 \in (0,1) \\ 0, \text{ otherwise.} \end{cases} \\ &= \begin{cases} \frac{1}{\lambda^2}y_1\exp\left(-\frac{y_1}{\lambda}\right), \text{ if } y_1 > 0, y_2 \in (0,1) \\ 0, \text{ otherwise.} \end{cases} \end{aligned}$$

Now, we compute the marginal distributions. The marginal p.d.f. $f_{Y_1}$ is given by

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f_{Y_1,Y_2}(y_1,y_2)\,dy_2 = \begin{cases} \frac{1}{\lambda^2}y_1\exp\left(-\frac{y_1}{\lambda}\right), \text{ if } y_1 > 0 \\ 0, \text{ otherwise} \end{cases}$$

and the marginal p.d.f. $f_{Y_2}$ is given by

$$f_{Y_2}(y_2) = \int_{-\infty}^{\infty} f_{Y_1,Y_2}(y_1,y_2)\,dy_1 = \begin{cases} 1, \text{ if } y_2 \in (0,1) \\ 0, \text{ otherwise} \end{cases}.$$

Therefore $Y_1 = X_1 + X_2 \sim Gamma(2, \lambda)$ and $Y_2 = \frac{X_1}{X_1+X_2} \sim Uniform(0,1)$. Moreover,

$$f_{Y_1,Y_2}(y_1, y_2) = f_{Y_1}(y_1) f_{Y_2}(y_2), \forall (y_1, y_2) \in \mathbb{R}^2$$

and hence $Y_1$ and $Y_2$ are independent.

*Remark* 9.26. We had earlier mentioned that $Exponential(\lambda)$ distribution is the same as $Gamma(1, \lambda)$ distribution. Using the above computation, we can identify a $Gamma(2, \lambda)$ RV as a sum of two independent RVs each having distribution $Gamma(1, \lambda)$. A more general property in this regard is mentioned in practice problem set 8.

We now consider expectations for random vectors and for functions of random vectors. The concepts are same as discussed in the case of RVs.

**Definition 9.27** (Expectation/Mean/Expected Value for functions of Random Vectors)**.** Let $X = (X_1, X_2, \cdots, X_p)$ be a $p$-dimensional discrete/continuous random vector with joint p.m.f./p.d.f. $f_X$. Let $h : \mathbb{R}^p \to \mathbb{R}$ be a function. Then $h(X)$ is an one-dimensional random vectors, i.e. an RV. We say that the expectation of $h(X)$, denoted by $\mathbb{E}h(X)$, is defined as the quantity

$$\mathbb{E}h(X) := \begin{cases} \sum_{x \in S_X} h(x) f_X(x), & \text{if } \sum_{x \in S_X} |h(x)| f_X(x) < \infty \text{ for discrete } X, \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(x) f_X(x)\, dx, & \text{if } \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |h(x)| f_X(x)\, dx < \infty \text{ for continuous } X. \end{cases}$$

In the discrete case, $S_X$ denotes the support of $X$.

*Remark* 9.28. If the sum or the integral above converges absolutely, we say that the expectation $\mathbb{E}h(X)$ exists or equivalently, $\mathbb{E}h(X)$ is finite. Otherwise, we shall say that the expectation $\mathbb{E}h(X)$ does not exist.

The following results is a generalization of Proposition 6.19. We skip the proof for brevity.

**Proposition 9.29.** *(a) Let $X = (X_1, X_2, \cdots, X_p)$ be a discrete random vector with joint p.m.f. $f_X$ and support $S_X$ and let $h : \mathbb{R}^p \to \mathbb{R}$ be a function. Consider the discrete RV $Y := h(X)$ with p.m.f. $f_Y$ and support $S_Y$. Then $\mathbb{E}Y$ exists if and only if $\sum_{y \in S_Y} |y| f_Y(y) < \infty$ and in* *this case,*

$$\mathbb{E}Y = \mathbb{E}h(X) = \sum_{x\in S_X} h(x)f_X(x) = \sum_{y\in S_Y} yf_Y(y).$$

*(b) Let $X = (X_1, X_2, \cdots, X_p)$ be a continuous random vector with joint p.d.f. $f_X$. Let $h : \mathbb{R}^p \to \mathbb{R}$ be a function such that the RV $Y := h(X)$ is continuous with p.d.f. $f_Y$. Then $\mathbb{E}Y$ exists if and only if $\int_{-\infty}^{\infty} |y| f_Y(y)\, dy < \infty$ and in this case,*

$$\mathbb{E}Y = \mathbb{E}h(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} h(x)f_X(x)\, dx = \int_{-\infty}^{\infty} yf_Y(y)\, dy.$$

**Note 9.30.** As considered for the case of RVs, by choosing different functions $h : \mathbb{R}^p \to \mathbb{R}$, we obtain several quantities of interest of the form $\mathbb{E}h(X)$ for a $p$-dimensional random vector $X$.

**Definition 9.31** (Some special expectations for Random Vectors)**.** Let $X = (X_1, X_2, \cdots, X_p)$ be a $p$-dimensional discrete/continuous random vector.

(a) (Joint Moments) For non-negative integers $k_1, \ldots, k_p$, let $h(x) := x_1^{k_1} \cdots x_p^{k_p}, \forall x \in \mathbb{R}^p$. Then,

$$\mu'_{k_1,\ldots,k_p} := \mathbb{E}\left(X_1^{k_1}\cdots X_p^{k_p}\right)$$

is called a joint moment of order $k_1 + \cdots + k_p$ of $X$, provided it exists.

(b) (Joint Central Moments) For non-negative integers $k_1, \ldots, k_p$, let

$$h(x) := (x_1 - \mathbb{E}(X_1))^{k_1}\cdots(x_p - \mathbb{E}(X_p))^{k_p}, \forall x \in \mathbb{R}^p.$$

Then

$$\mu_{k_1,\ldots,k_p} := \mathbb{E}\left((X_1 - \mathbb{E}(X_1))^{k_1}\cdots(X_p - \mathbb{E}(X_p))^{k_p}\right)$$

is called a joint central moment of order $k_1 + \cdots + k_p$ of $X$, provided it exists.

(c) (Covariance) Fix $i, j = 1, \ldots, p$. Let $h(x) := (x_i - \mathbb{E}(X_i))(x_j - \mathbb{E}(X_j)), \forall x = (x_1, x_2, \cdots, x_p) \in \mathbb{R}^p$. Then, $\mathbb{E}\left[(X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))\right]$ is called the covariance between $X_i$ and $X_j$, provided it exists. We shall denote this quantity by $Cov(X_i, X_j)$.

(d) (Joint Moment Generating Function, or simply, Joint MGF) We define

$$A := \left\{t = (t_1, t_2, \ldots, t_p) \in \mathbb{R}^p : \mathbb{E}\left(e^{\sum_{i=1}^{p} t_i X_i}\right) < \infty\right\},$$

and consider the function $M_X : A \to \mathbb{R}$ defined by

$$M_X(t) = \mathbb{E}\left(e^{\sum_{i=1}^{p} t_i X_i}\right), \quad \forall t = (t_1, t_2, \ldots, t_p) \in A.$$

If $(-a_1, a_1) \times (-a_2, a_2) \times \cdots \times (-a_p, a_p) \subseteq A$ for some $a_1, a_2, \cdots, a_p > 0$, then the function $M_X$ is called the joint moment generating function (joint MGF) of the random vector $X$. Note that $t = (0, 0, \cdots, 0) \in \mathbb{R}^p$ yields $M_X(t) = 1$ and hence $(0, 0, \cdots, 0) \in A$.

(e) (Joint Characteristic Function) Define $\Phi_X : \mathbb{R}^p \to \mathbb{C}$ by

$$\Phi_X(t) = \mathbb{E}\left(e^{\sum_{j=1}^{p} i\, t_j X_j}\right) = \mathbb{E}\exp\left(i\, t.X\right) = \mathbb{E}\cos(t.X) + i\,\mathbb{E}\sin(t.X), \quad \forall t = (t_1, t_2, \ldots, t_p) \in \mathbb{R}^p,$$

where $i$ denotes the complex number $\sqrt{-1}$ and $t.X = \sum_{j=1}^{p} t_j X_j$ is the standard dot product in $\mathbb{R}^p$. This function exists for all $t \in \mathbb{R}^p$.

*Remark* 9.32. We now list some properties of the above quantities. The properties are being stated under the assumption that the expectations involved exist. Let $X = (X_1, X_2, \cdots, X_p)$ be a $p$-dimensional discrete/continuous random vector.

(a) Let $a_1, \ldots, a_p$ be real constants. Then, $\mathbb{E}\left(\sum_{i=1}^{p} a_i X_i\right) = \sum_{i=1}^{p} a_i \mathbb{E} X_i$. To see this for discrete $X$, observe that

$$\mathbb{E}\left(\sum_{i=1}^{p} a_i X_i\right) = \sum_{x \in S_X} \sum_{i=1}^{p} a_i x_i f_X(x) = \sum_{i=1}^{p} \sum_{x \in S_X} a_i x_i f_X(x) = \sum_{i=1}^{p} a_i \mathbb{E} X_i.$$

The interchange of the order of summation is allowed due to absolute convergence of the series involved. The proof for continuous $X$ is similar.

(b) $Cov(X_i, X_j) = Cov(X_j, X_i)$, for all $i, j = 1, \ldots, p$.

(c) $Cov(X_i, X_i) = Var(X_i)$, for all $i = 1, \ldots, p$.

(d) For all $i, j = 1, \ldots, p$, we have

$$\begin{aligned} Cov(X_i, X_j) &= \mathbb{E}\left[X_i X_j - X_i(\mathbb{E} X_j) - X_j(\mathbb{E} X_i) + (\mathbb{E} X_i)(\mathbb{E} X_j)\right] \\ &= \mathbb{E}\left(X_i X_j\right) - \mathbb{E}\left(X_i\right) \mathbb{E}\left(X_j\right) \end{aligned}$$

(e) Let $X_1, X_2, \ldots, X_p, Y_1, Y_2, \ldots, Y_q$ be RVs, and let $a_1, \ldots, a_p, b_1, \ldots, b_q$ be real constants. Then,

$$Cov\left(\sum_{i=1}^{p} a_i X_i, \sum_{j=1}^{q} b_j Y_j\right) = \sum_{i=1}^{p}\sum_{j=1}^{q} a_i b_j Cov\left(X_i, Y_j\right).$$

In particular,

$$\begin{aligned} Var\left(\sum_{i=1}^{p} a_i X_i\right) &= \sum_{i=1}^{p} a_i^2 Var\left(X_i\right) + \sum_{i=1}^{p}\sum_{\substack{j=1\\ j\neq i}}^{p} a_i a_j Cov\left(X_i, X_j\right) \\ &= \sum_{i=1}^{p} a_i^2 Var\left(X_i\right) + 2\sum_{1\leq i<j\leq p} a_i a_j Cov\left(X_i, X_j\right). \end{aligned}$$

(f) Let $X_1, X_2, \cdots, X_p$ be independent and let $h_1, h_2, \cdots, h_p : \mathbb{R} \to \mathbb{R}$ be functions. Then

$$\mathbb{E}\left(\prod_{i=1}^{p} h_i(X_i)\right) = \prod_{i=1}^{p} \mathbb{E}h_i(X_i).$$

For simplicity, we discuss the proof when $p = 2$ and $X = (X_1, X_2)$ is continuous with joint p.d.f. $f_X$. Recall from Theorem 9.9 that $f_X(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2), \forall x_1, x_2, \in \mathbb{R}$. Then,

$$\begin{aligned} \mathbb{E}\left(\prod_{i=1}^{2} h_i(X_i)\right) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_1(x_1)h_2(x_2) f_X(x_1, x_2)\, dx_1 dx_2 \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_1(x_1)h_2(x_2) f_{X_1}(x_1) f_{X_2}(x_2)\, dx_1 dx_2 \\ &= \left(\int_{-\infty}^{\infty} h_1(x_1) f_{X_1}(x_1)\, dx_1\right)\left(\int_{-\infty}^{\infty} h_2(x_2) f_{X_2}(x_2)\, dx_2\right) \\ &= \prod_{i=1}^{2} \mathbb{E}h_i(X_i). \end{aligned}$$

(g) This is a special case of statement (f). Let $A_1, A_2, \cdots, A_p \subseteq \mathbb{R}$. Consider the functions

$$h_i(x_i) := \begin{cases} 1, & \text{if } x \in A_i \\ 0, & \text{otherwise.} \end{cases} = 1_{A_i}(x_i), \forall x_i \in \mathbb{R}, i = 1, 2, \cdots, p.$$

Note that $\mathbb{E}h_i(X_i) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} 1_{A_i}(x_i) f_{X_1}(x_1) f_{X_2}(x_2)\cdots f_{X_p}(x_p)\, dx_1 dx_2 \cdots dx_p = \int_{-\infty}^{\infty} 1_{A_i}(x_i) f_{X_i}(x_i)\, dx_i = \mathbb{P}(X_i \in A_i)$, when $X$ is continuous. The same equality is also true when $X$ is discrete. Now, consider the function $h : \mathbb{R}^p \to \mathbb{R}$ defined by $h(x) = \prod_{i=1}^p h_i(x_i), \forall x \in \mathbb{R}^p$. Using (f), we have

$$\mathbb{P}(X_1 \in A_1, X_2 \in A_2, \cdots, X_p \in A_p) = \prod_{i=1}^p \mathbb{P}(X_i \in A_i).$$

(h) Continue with the assumptions of statement (f). For fixed $y_1, y_2, \cdots, y_p \in \mathbb{R}$, consider the functions $g_1, g_2, \cdots, g_p : \mathbb{R} \to \mathbb{R}$ defined by

$$g_i(x_i) := \begin{cases} 1, & \text{if } h_i(x_i) \leq y_i \\ 0, & \text{otherwise.} \end{cases} \quad \forall x_i \in \mathbb{R}, i = 1, 2, \cdots, p.$$

Note that $\mathbb{E} g_i(X_i) = \mathbb{P}(h_i(X_i) \leq y_i) = F_{h_i(X_i)}(y_i), \forall i$ and

$$\begin{aligned} F_{h_1(X_1), h_2(X_2), \cdots, h_p(X_p)}(y_1, y_2, \cdots, y_p) &= \mathbb{P}(h_1(X_1) \leq y_1, h_2(X_2) \leq y_2, \cdots, h_p(X_p) \leq y_p) \\ &= \prod_{i=1}^p \mathbb{P}(h_i(X_i) \leq y_i) \\ &= \prod_{i=1}^p F_{h(X_i)}(y_i). \end{aligned}$$

Hence, the RVs $h_1(X_1), h_2(X_2), \cdots, h_p(X_p)$ are independent.

(i) Let $X_1, X_2$ be independent RVs. Then $\mathbb{E}(X_1 X_2) = (\mathbb{E} X_1)(\mathbb{E} X_2)$ and hence, using (d),

$$Cov(X_1, X_2) = 0.$$

Further, if $X_1, X_2, \cdots, X_p$ are independent, then using (e),

$$Var\left(\sum_{i=1}^p a_i X_i\right) = \sum_{i=1}^p a_i^2 Var\left(X_i\right)$$

for all real constants $a_1, a_2, \cdots, a_p$.

(j) Recall that $M_X : A \to \mathbb{R}$ is given by

$$M_X(t) = \mathbb{E}\left(e^{\sum_{i=1}^p t_i X_i}\right), \quad \forall t = (t_1, t_2, \ldots, t_p) \in A,$$

with

$$A := \left\{t = (t_1, t_2, \ldots, t_p) \in \mathbb{R}^p : \mathbb{E}\left(e^{\sum_{i=1}^p t_i X_i}\right) < \infty\right\}.$$

Taking $t = (0, 0, \cdots, 0) \in \mathbb{R}^p$ yields $M_X(0, 0, \cdots, 0) = 1$ and hence $(0, 0, \cdots, 0) \in A$. In particular, $A \neq \emptyset$. Also, $M_X(t) > 0, \forall t \in A$.

(k) If $t = (0, \cdots, 0, t_i, 0, \cdots, 0) \in A$, then $M_X(t) = \mathbb{E}\left(e^{\sum_{k=1}^p t_k X_k}\right) = \mathbb{E}\left(e^{t_i X_i}\right) = M_{X_i}(t_i)$. Similarly, if $t = (0, \cdots, 0, t_i, 0, \cdots, 0, t_j, 0, \cdots, 0) \in A$, then $M_X(t) = \mathbb{E}\left(e^{\sum_{k=1}^p t_k X_k}\right) = \mathbb{E}\left(e^{t_i X_i + t_j X_j}\right) = M_{X_i, X_j}(t_i, t_j)$.

(l) This result is being stated without proof. If $(-a_1, a_1) \times (-a_2, a_2) \times \cdots \times (-a_p, a_p) \subseteq A$ for some $a_1, a_2, \cdots, a_p > 0$, then $M_X$ possesses partial derivatives of all orders in $(-a_1, a_1) \times (-a_2, a_2) \times \cdots \times (-a_p, a_p)$. Furthermore, for non-negative integers $k_1, \ldots, k_p$

$$\mathbb{E}\left(X_1^{k_1} X_2^{k_2} \cdots X_p^{k_p}\right) = \left[\frac{\partial^{k_1+k_2+k_3+\cdots+k_p}}{\partial t_1^{k_1} \cdots \partial t_p^{k_p}} M_X(t)\right]_{(t_1, t_2 \ldots t_p) = (0, \ldots, 0)}.$$

For $i \neq j$ with $i, j \in \{1, \ldots, p\}$, we have

$$\begin{aligned}
&Cov\,(X_i, X_j) \\
&= \mathbb{E}\,(X_i X_j) - \mathbb{E}\,(X_i)\,\mathbb{E}\,(X_j) \\
&= \left[\frac{\partial^2}{\partial t_i \partial t_j} M_X(t)\right]_{(t_1, t_2 \ldots t_p) = (0, \ldots, 0)} - \left[\frac{\partial}{\partial t_i} M_X(t)\right]_{(t_1, t_2 \ldots t_p) = (0, \ldots, 0)} \left[\frac{\partial}{\partial t_j} M_X(t)\right]_{(t_1, t_2 \ldots t_p) = (0, \ldots, 0)} \\
&= \left[\frac{\partial^2}{\partial t_i \partial t_j} \Psi_X(t)\right],
\end{aligned}$$

where $\Psi_X(t) := \ln M_X(t), t \in A$. Compare this with the one-dimensional case in Proposition 6.45.

(m) If $X_1, X_2, \cdots, X_p$ are independent, then for all $t \in A$,

$$M_X(t) = \mathbb{E}\left(e^{\sum_{i=1}^p t_i X_i}\right) = \mathbb{E}\left(\prod_{i=1}^p e^{t_i X_i}\right) = \prod_{i=1}^p \mathbb{E}\left(e^{t_i X_i}\right) = \prod_{i=1}^p M_{X_i}(t_i).$$

(n) If $(-a_1, a_1) \times (-a_2, a_2) \times \cdots \times (-a_p, a_p) \subseteq A$ for some $a_1, a_2, \cdots, a_p > 0$ and $M_X(t) = \prod_{i=1}^p M_{X_i}(t_i), \forall t \in A$, then it can be shown that $X_1, X_2, \cdots, X_p$ are independent. We do not discuss the proof of this result in this course.

(o) As discussed for the case of random variables, the joint Characteristic function for random vectors has nice properties similar to the joint MGF. If some joint moment for the random vectors exists, then it can be recovered from the partial derivatives of the joint Characteristic function etc.. We do not discuss these properties in detail in this course.