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 1.325 (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 1.326. 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 1.190. We skip the proof for brevity.

Proposition 1.327. *(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,*

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

(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,

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

Note 1.328. 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 1.329 (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, $\mathbb{E}[(X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))]$ 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

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

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 $(0, 0, \cdots, 0) \in A$.

Remark 1.330*.* 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 $a_1, \dots, a_p$ 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 $i, j = 1, \dots, p$.

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

(d) For all $i, j = 1, \dots, p$, 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 $X_1, X_2, \dots, X_p, Y_1, Y_2, \dots, Y_q$ be RVs, and let $a_1, \dots, a_p, b_1, \dots, b_q$ 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 1.307 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,

$$\begin{aligned}
\mathbb{E}\left(\prod_{i=1}^{2} h_i(X_i)\right) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_1(x_1)h_2(x_2)f_X(x_1,x_2)\,dx_1dx_2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_1(x_1)h_2(x_2)f_{X_1}(x_1)f_{X_2}(x_2)\,dx_1dx_2 \\
&= \left(\int_{-\infty}^{\infty} h_1(x_1)f_{X_1}(x_1)\,dx_1\right)\left(\int_{-\infty}^{\infty} h_2(x_2)f_{X_2}(x_2)\,dx_2\right) \\
&= \prod_{i=1}^{2} \mathbb{E}h_i(X_i).
\end{aligned}$$

(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 $\mathbb{E}h_i(X_i) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} 1_{A_i}(x_i)f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_p}(x_p)\,dx_1dx_2\cdots dx_p = \int_{-\infty}^{\infty} 1_{A_i}(x_i)f_{X_i}(x_i)\,dx_i = \mathbb{P}(X_i \in A_i)$, 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

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

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

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

$$= \prod_{i=1}^{p} F_{h(X_i)}(y_i).$$

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 $\mathbb{E}(X_1X_2) = (\mathbb{E}X_1)(\mathbb{E}X_2)$ 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

$Cov\left(X_i, X_j\right)$

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

where $\Psi_X(t) := \ln M_X(t), t \in A$. Compare this with the one-dimensional case in Proposition 1.215.

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

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

(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.

Proposition 1.331 (Cauchy-Schwarz Inequality)**.** *Let X and Y be RVs defined on the same probability space. Then,*

$$
(\mathbb{E}(XY))^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2),
$$

provided the expectations exist. The equality occurs if and only if $\mathbb{P}(Y = cX) = 1$ or $\mathbb{P}(X = cY) = 1$ for some $c \in \mathbb{R}$.

Proof. First we consider the case when $\mathbb{E}X^2 = 0$. Then $\mathbb{P}(X = 0) = 1$ and consequently $\mathbb{P}(XY = 0) = 1$ and $\mathbb{E}(XY) = 0$. The equality holds.

Now, assume that $\mathbb{E}X^2 > 0$. Now, for all $c \in \mathbb{R}$, we have $\mathbb{E}(Y - cX)^2 = c^2\mathbb{E}X^2 - 2c\mathbb{E}(XY) + \mathbb{E}Y^2 \geq 0$. Hence, the discriminant $(2\mathbb{E}(XY))^2 - 4\mathbb{E}X^2\mathbb{E}Y^2$ must be non-positive, which proves the statement.

If the equality holds for some $\mathbb{E}(Y - cX)^2 = 0$ for some c, then we have $\mathbb{P}(Y = cX) = 1$. If $\mathbb{P}(Y = cX) = 1$ for some c, then $\mathbb{E}(Y - cX)^2 = 0$. Interchanging the roles of X and Y, we can discuss the case involving $\mathbb{E}(X - cY)^2$ and $\mathbb{P}(X = cY)$. □

Corollary 1.332. *Let X and Y be RVs defined on the same probability space. Then,*

$$(Cov(X, Y))^2 \leq Var(X)\, Var(Y),$$

provided the covariance and the variances exist.

Proof. Take $U = X - \mathbb{E}X$ and $V = Y - \mathbb{E}Y$. Applying the Cauchy-Schwarz inequality to U and V, the result follows. □

Definition 1.333 (Correlation between RVs)**.** Let X and Y be RVs defined on the same probability space. If $0 < Var(X) < \infty, 0 < Var(Y) < \infty$, then we call

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

as the Correlation between X and Y. We say X and Y are uncorrelated if $\rho(X, Y) = 0$ or equivalently $Cov(X, Y) = 0$.

Note 1.334. By Corollary 1.332, $|\rho(X, Y)| \leq 1$ for any two RVs X and Y defined on the same probability space.

Remark 1.335 (Correlation and Independence). If X and Y are independent RVs defined on the same probability space, then by Remark 1.330(i), $Cov(X, Y) = 0$ and hence X and Y are uncorrelated. However, the converse is not true. We illustrate this problem with examples.

(a) Let $X = (X_1, X_2)$ be a bivariate discrete random vector, i.e. a 2-dimensional discrete random vector with joint p.m.f. given by

$$f_X(x_1, x_2) = \begin{cases} \frac{1}{2}, & \text{if } (x_1, x_2) = (0, 0), \\ \frac{1}{4}, & \text{if } (x_1, x_2) = (1, 1) \text{ or } (1, -1), \\ 0, & \text{otherwise.} \end{cases}$$

The marginal p.m.fs are

$$f_{X_1}(x_1) = \begin{cases} \frac{1}{2}, \text{ if } x_1 \in \{0,1\} \\ 0, \text{ otherwise} \end{cases} \quad , \quad f_{X_2}(x_2) = \begin{cases} \frac{1}{2}, \text{ if } x_2 = 0 \\ \frac{1}{4}, \text{ if } x_2 \in \{1,-1\} \\ 0, \text{ otherwise} \end{cases}$$

We have $f_{X_1,X_2}(0,0) = \frac{1}{2} \neq \frac{1}{4} = f_{X_1}(0)f_{X_2}(0)$ and hence X_1 and X_2 are not independent. But, $\mathbb{E}X_1 = \frac{1}{2}, \mathbb{E}X_2 = 0, \mathbb{E}(X_1X_2) = 0, Var(X_1) > 0$ and $Var(X_2) > 0$. Therefore $Cov(X_1, X_2) = 0$ and hence X_1 and X_2 are uncorrelated.

(b) Let $X = (X_1, X_2)$ be a bivariate continuous random vector, i.e. a 2-dimensional continuous random vector with joint p.d.f. given by

$$f_X(x_1, x_2) = \begin{cases} 1, & \text{if } 0 < |x_2| \leq x_1 < 1 \\ 0, & \text{otherwise.} \end{cases}$$

Then,

$$\mathbb{E}(X_1X_2) = \int_0^1 \int_{-x_1}^{x_1} x_1x_2 \, \mathrm{d}x_2 \, \mathrm{d}x_1 = 0,$$

and

$$\mathbb{E}(X_1) = \int_0^1 \int_{-x_1}^{x_1} x_1 \, \mathrm{d}x_2 \, \mathrm{d}x_1 = \frac{2}{3}, \quad \mathbb{E}(X_2) = \int_0^1 \int_{-x_1}^{x_1} x_2 \, \mathrm{d}x_2 \, \mathrm{d}x_1 = 0.$$

Hence, $\mathbb{E}(X_1X_2) = (\mathbb{E}X_1)(\mathbb{E}X_2)$, which implies $Cov(X_1, X_2) = 0$. A similar computation shows $Var(X_1)$ and $Var(X_2)$ exists and are non-zero. Hence, X_1 and X_2 are uncorrelated. Now, by computing the marginal p.d.f.s f_{X_1} and f_{X_2}, it is immediate that the equality

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2)$$

does not hold for all $x = (x_1, x_2) \in \mathbb{R}^2$. Here, X_1 and X_2 are not independent. The verification with the marginal p.d.f.s is left as an exercise in practice problem set 8.

We now discuss the concept of equality of distribution for random vectors. As we shall see, the ideas remain the same as in the case of RVs.

Definition 1.336 (Identically distributed random vectors)**.** Let X and Y be two p-dimensional random vectors, possibly defined on different probability spaces. We say that they have the same law/distribution, or equivalently, X and Y are identically distributed or equivalently, X and Y are equal in law/distribution, denoted by $X \stackrel{d}{=} Y$, if $F_X(x) = F_Y(x), \forall x \in \mathbb{R}^p$.

Remark 1.337. As discussed in the case of RVs in Remark 1.219, we can check whether two random vectors are identically distributed or not via other quantities that describe their law/distribution.

(a) Let X and Y be p-dimensional discrete random vectors with joint p.m.f.s f_X and f_Y, respectively. Then X and Y are identically distributed if and only if $f_X(x) = f_Y(x), \forall x \in \mathbb{R}^p$.
(b) Let X and Y be p-dimensional continuous random vectors with joint p.d.f.s f_X and f_Y, respectively. Then X and Y are identically distributed if and only if $f_X(x) = f_Y(x), \forall x \in \mathbb{R}^p$.
(c) Let X and Y be p-dimensional random vectors such that their joint MGFs M_X and M_Y exist and agree on $(-a_1, a_1) \times (-a_2, a_2) \times \cdots (-a_p, a_p)$ for some $a_1, a_2, \cdots, a_p > 0$, then X and Y are identically distributed.
(d) Let X and Y be identically distributed p-dimensional random vectors. Then for any function $h : \mathbb{R}^p \to \mathbb{R}^q$, we have $h(X)$ and $h(Y)$ are identically distributed q-dimensional random vectors.

Notation 1.338 (i.i.d RVs)**.** We say that RVs $X_1, \cdots, X_p$ defined on the same probability space are independent and identically distributed, then we usually use the short hand notation i.i.d. and say that $X_1, \cdots, X_p$ are i.i.d..

Note 1.339. We can generalize the concept of independence of RVs to independence of random vectors. To avoid notational complexity, we do not discuss this in this course.

Definition 1.340. (a) A random sample is a collection of i.i.d. RVs.
(b) A random sample of size n is a collection of n i.i.d. RVs $X_1, X_2, \cdots, X_n$.

(c) Let $X_1, X_2, \cdots, X_n$ be a random sample of size n. If the common DF is F or the common p.m.f./p.d.f. is f, then we call $X_1, X_2, \cdots, X_n$ to be a random sample from a distribution having a DF F or p.m.f./p.d.f. f.

(d) A function of one or more RVs that does not depend on any unknown parameter is called a statistic.

Example 1.341. Suppose that $X_1, \cdots, X_n$ are i.i.d. with the common distribution being $Poisson(\theta)$ or $Exponential(\theta)$ for some unknown $\theta \in (0, \infty)$. Here, θ is a unknown parameter.

(a) $\bar{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$ is a statistic and is usually referred to as the sample mean.

(b) $X_1 - \theta$ is not a statistic.

(c) $S_n^2 := \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is a statistic and is usually referred to as the sample variance. Depending on the situation, we sometimes work with $\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

(d) The value of S_n such that S_n^2 is the sample variance, is referred to as the sample standard deviation.

(e) For $r = 1, \cdots, n$, we denote by $X_{(r:n)}$ the r-th smallest of $X_1, \cdots, X_n$. By definition, $X_{(1:n)} \leq \cdots \leq X_{(n:n)}$ and these are called the order statistics of the random sample. If n is understood, then we simply write $X_{(r)}$ to denote the r-th order statistic.

Note 1.342. Let X_1, X_2 be a random sample of size 2. Then $X_{(1)} = \min\{X_1, X_2\} = \frac{1}{2}(X_1 + X_2) - \frac{1}{2}|X_1 - X_2|$ and $X_{(2)} = \max\{X_1, X_2\} = \frac{1}{2}(X_1 + X_2) + \frac{1}{2}|X_1 - X_2|$ are RVs. Using similar arguments, it follows that the order statistics from any random sample of size n are RVs. The joint distribution of the order statistics is therefore of interest.

Note 1.343. Let $X_1, \cdots, X_n$ be a random sample of continuous RVs with the common p.d.f. f. Then,

$$\mathbb{P}(X_{(1)} < X_{(2)}) < \cdots < X_{(n)}) = 1$$

and hence $X_{(r)}, r = 1, \cdots, n$ are defined uniquely with probability one.

Proposition 1.344. *Let $X_1, \cdots, X_n$ be a random sample of continuous RVs with the common DF F and the common p.d.f. f. The joint p.d.f. of $(X_{(1)}, \cdots, X_{(n)})$ is given by*

$$g(y_1, \cdots, y_n) = \begin{cases} n! \prod_{i=1}^{n} f(y_i), & \textit{if } y_1 < \cdots < y_n, \\ 0, & \textit{otherwise.} \end{cases}$$

Further the marginal p.d.f. of $X_{(r)}$ is given by

$$g_{X_{(r)}}(y) = \frac{n!}{(r-1)!(n-r)!}(F(y))^{r-1}(1-F(y))^{n-r} f(y), \forall y \in \mathbb{R}.$$

Proof. Observe that a sample value $(y_1, \cdots, y_n)$ of $(X_{(1)}, \cdots, X_{(n)})$ is related to a sample $(x_1, \cdots, x_n)$ of $(X_1, \cdots, X_n)$ in the following way

$$(y_1, \cdots, y_n) = (x_{(1)}, \cdots, x_{(n)}),$$

$x_{(r)}$ being the r-th smallest of $x_1, \cdots, x_n$. Note that $y_r = x_{(r)}$.

Now, the actual values $x_1, \cdots, x_n$ may have been arranged in a different order than $x_{(1)}, \cdots, x_{(n)}$. In fact, the values $x_{(1)}, \cdots, x_{(n)}$ arise from one of the $n!$ permutations of the values $x_1, \cdots, x_n$. But, any such transformation/permutation is obtained by the action of a permutation matrix on the vector $(x_1, \cdots, x_n)$. For example, if $x_1 < x_2 < \cdots < x_{n-2} < x_n < x_{n-1}$, then $x_{(1)} = x_1, \cdots, x_{(n-2)} = x_{n-2}, x_{(n-1)} = x_n, x_{(n)} = x_{n-1}$ which interchanges the $n-1$ and n-th values, i.e. x_{n-1} and x_n.

Hence, the Jacobian matrix for this transformation is the same as the corresponding permutation matrix and the Jacobian determinant is ± 1.

Since $X_1, \cdots, X_n$ are i.i.d., the joint p.d.f. of $(X_1, \cdots, X_n)$ is given by

$$f_{X_1, \cdots, X_n}(x_1, \cdots, x_n) = f(x_1) \times \cdots \times f(x_n), \forall (x_1, \cdots, x_n) \in \mathbb{R}^n.$$

Using Theorem 1.322, we have joint p.d.f. of $(X_{(1)}, \cdots, X_{(n)})$ is given by

$$g(y_1, \cdots, y_n) = \begin{cases} n! \prod_{i=1}^{n} f(y_i), & \text{if } y_1 < \cdots < y_n, \\ 0, & \text{otherwise.} \end{cases}$$

The marginal p.d.f. of $X_{(r)}$ can now be computed for $y \in \mathbb{R}$,

$$g_{X_{(r)}}(y)$$
$$= \int_{y_{r-1}=-\infty}^{y} \int_{y_{r-2}=-\infty}^{y_{r-1}} \cdots \int_{y_1=-\infty}^{y_2} \int_{y_{r+1}=y}^{\infty} \int_{y_{r+2}=y_{r+1}}^{\infty} \cdots \int_{y_n=y_{n-1}}^{\infty} n! \prod_{i=1}^{n} f(y_i)\, dy_n dy_{n-1} \cdots dy_{r+1} dy_1 dy_2 \cdots dy_{r-1}$$

The above integral simplifies to the result stated above. □

Example 1.345. Let X_1, X_2, X_3 be a random sample from $Uniform(0,1)$ distribution. The common p.d.f. here is given by

$$f(x) = \begin{cases} 1, \text{ if } x \in (0,1) \\ 0, \text{ otherwise.} \end{cases}$$

By the above result, the joint p.d.f. of $(X_{(1)}, X_{(2)}, X_{(3)})$ is given by

$$g(y_1, y_2, y_3) = \begin{cases} 6, \text{ if } 0 < y_1 < y_2 < y_3 < 1, \\ 0, \text{ otherwise.} \end{cases}$$

and the marginal p.d.f. of $X_{(1)}$ is

$$g(y_1) = \begin{cases} 3(1-y_1)^2, \text{ if } y_1 \in (0,1) \\ 0, \text{ otherwise.} \end{cases}$$

Remark 1.346. For random samples from discrete distributions, there is no general formula or result which helps in computing the joint distribution of the order statistics. Usually they are done by a case-by-case analysis. Let X_1, X_2, X_3 be a random sample from $Bernoulli(p)$ distribution, for some $p \in (0,1)$. The common p.m.f. here is given by

$$f(x) = \begin{cases} p, \text{ if } x = 1 \\ 1-p, \text{ if } x = 0 \\ 0, \text{ otherwise.} \end{cases}$$

Note that $X_{(1)}$ is also a $\{0,1\}$-valued RV with $X_{(1)} = \min\{X_1, X_2, X_3\} = 1$ if and only if $X_1 = X_2 = X_3 = 1$. Then using independence,

$$\mathbb{P}(X_{(1)} = 1) = \mathbb{P}(X_1 = 1, X_2 = 1, X_3 = 1) = \mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 1)\mathbb{P}(X_3 = 1) = p^3$$

and $\mathbb{P}(X_{(1)} = 0) = 1 - \mathbb{P}(X_{(1)} = 1) = 1 - p^3$. Therefore, $X_{(1)} \sim Bernoulli(p^3)$. Similarly, $X_{(3)} \sim Bernoulli(1 - (1-p)^3)$. The distribution of $X_{(2)}$ is left as an exercise in problem set 8.

Earlier, in Week 7, we have discussed the concept of conditional distributions. In this week, we have also discussed the concept of expectation of a random vector. Combining these two concepts, we are led to the following.

Definition 1.347 (Conditional Expectation, Conditional Variance and Conditional Covariance)**.** Let $X = (X_1, X_2, \cdots, X_{p+q})$ be a $p+q$-dimensional random vector with joint p.m.f./p.d.f. f_X. Let the joint p.m.f./p.d.f. $Y = (X_1, X_2, \cdots, X_p)$ and $Z = (X_{p+1}, X_{p+2}, \cdots, X_{p+q})$ be denoted by f_Y and f_Z, respectively. Let $h : \mathbb{R}^p \to \mathbb{R}$ be a function. Let $z \in \mathbb{R}^q$ be such that $f_Z(z) > 0$.

(a) The conditional expectation of $h(Y)$ given $Z = z$, denoted by $\mathbb{E}(h(Y) \mid Z = z)$, is the expectation of $h(Y)$ under the conditional distribution of Y given $Z = z$.

(b) The conditional variance of $h(Y)$ given $Z = z$, denoted by $Var(h(Y) \mid Z = z)$, is the variance of $h(Y)$ under the conditional distribution of Y given $Z = z$.

(c) Let $1 \leq i \neq j \leq p$. The conditional covariance between X_i and X_j given $Z = z$, denoted by $Cov(X_i, X_j \mid Z = z)$, is the covariance between X_i and X_j under the conditional distribution of (X_i, X_j) given $Z = z$.

Notation 1.348. On $\{z \in \mathbb{R}^q : f_Z(z) > 0\}$, consider the function, $g_1(z) := \mathbb{E}(h(Y) \mid Z = z)$. We denote the RV $g_1(Z)$ by $\mathbb{E}(h(Y) \mid Z)$. Similarly, define the RVs $Var(h(Y) \mid Z)$ and $Cov(X_1, X_2 \mid Z)$

Proposition 1.349. *The following are properties of Conditional Expectation, Conditional Variance and Conditional Covariance. Here, we assume that the relevant expectations exist.*

(a) $\mathbb{E}h(Y) = \mathbb{E}(\mathbb{E}(h(Y) \mid Z)$.

(b) $Var(h(Y)) = Var(\mathbb{E}(h(Y) \mid Z)) + \mathbb{E}Var(h(Y) \mid Z)$.

(c) $Cov(X_1, X_2) = Cov(\mathbb{E}(X_1 \mid Z), \mathbb{E}(X_2 \mid Z)) + \mathbb{E}Cov(X_1, X_2 \mid Z)$.

Proof. We only prove the first statement under a simple assumption. The general case and other statements can be proved using appropriate generalization.

Take $p = q = 1$ and let $X = (Y, Z)$ be a 2-dimensional continuous random vector. Then,

$$\begin{aligned}
\mathbb{E}(\mathbb{E}(h(Y) \mid Z) &= \int_{-\infty}^{\infty} \mathbb{E}(h(Y) \mid Z = z) f_Z(z)\, dz \\
&= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} h(y) f_{Y|Z}(y \mid z)\, dy\right] f_Z(z)\, dz \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(y) f_{Y,Z}(y, z)\, dydz \\
&= \mathbb{E}h(Y).
\end{aligned}$$

□

Example 1.350. We shall see computations for conditional expectations in a later lecture.

We look at examples of discrete RVs in relation with random experiments.

Remark 1.351 (Binomial RVs via random experiments). Recall that in Remark 1.228, we have seen Bernoulli RVs arising from Bernoulli trials. Now, consider the same random experiment with two outcomes 'Success' and 'Failure' with probability of success $p \in (0, 1)$. Now, consider n independent Bernoulli trials of this experiment with the RV X_i being 1 for 'Success' and 0 for 'Failure' in the i-th trial for $i = 1, 2, \cdots, n$. Then, $X_1, X_2, \cdots, X_n$ is a random sample of size n from the *Bernoulli*(p) distribution. Now, the total number X of successes in the n trials is given by $X = X_1 + X_2 + \cdots + X_n$ and hence, by Remark 1.321, $X \sim Binomial(n, p)$. A $Binomial(n, p)$ RV can therefore be interpretated as the number of successes in n trials of a random experiment with two outcomes 'Success' and 'Failure' with probability of success $p \in (0, 1)$. Here, we have kept p fixed over all the trials.

Example 1.352. Suppose that a standard six-sided fair die is rolled at random 4 times independently. We now consider the probability that all the rolls result in a number at least 5. In each roll, obtaining at least 5 has the probability $\frac{2}{6} = \frac{1}{3}$ - we treat this as the probability of success in one trial. Repeating the trial three times independently gives us the number of success as