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

Lecture 17: Conditional Expectation, Variance and Covariance

Instructor: Dr. Kuldeep Kumar Kataria *Scribe:*

Example 17.1. *Let* $\underline{X} = (X_1, X_2, X_3)$ *have the joint p.d.f.*

$$f_{\underline{X}}(x_1, x_2, x_3) = \begin{cases} \dfrac{1}{x_1 x_2}, \ 0 < x_3 < x_2 < x_1 < 1, \\ 0, \ \textit{otherwise}. \end{cases}$$

(a) Show that $f_{\underline{X}}(\cdot)$ *is a proper p.d.f.*

(b) Find the marginal p.d.f. of (X_2, X_3).

(c) Find the marginal p.d.f. of X_1.

(d) Find the conditional p.d.f. of X_1 *given* $(X_2, X_3) = (x_2, x_3)$ *where* $0 < x_3 < x_2 < 1$.

(e) Are X_1, X_2 *and* X_3 *independent.*

(f) Find the conditional p.d.f. of (X_1, X_3) *given* $X_2 = x_2$*, where* $0 < x_2 < 1$.

(g) Are X_1 *and* X_3 *independent given* $X_2 = x_2$*, where* $0 < x_2 < 1$.

Solution: (a) Clearly, $f_{\underline{X}}(\underline{x}) \geq 0 \ \forall \ \underline{x} \in \mathbb{R}^3$. Also,

$$\int_{\mathbb{R}^3} f_{\underline{X}}(\underline{x}) \mathrm{d}\underline{x} = \int_0^1 \int_0^{x_1} \int_0^{x_2} \frac{1}{x_1 x_2} \mathrm{d}x_3 \mathrm{d}x_2 \mathrm{d}x_1 = 1.$$

So, $f_{\underline{X}}(\underline{x})$ is a p.d.f.

(b) The marginal p.d.f. of (X_2, X_3) is obtained as,

$$f_{X_2,X_3}(x_2, x_3) = \int_{-\infty}^{\infty} f_{\underline{X}}(\underline{x}) \mathrm{d}x_1 = \int_{x_2}^{1} \frac{1}{x_1 x_2} \mathrm{d}x_1 = -\frac{\ln x_2}{x_2}, \quad 0 < x_3 < x_2 < 1.$$

So,

$$f_{X_2,X_3}(x, y) = \begin{cases} -\dfrac{\ln x}{x}, \ 0 < y < x < 1, \\ 0, \ \text{otherwise}. \end{cases}$$

(c) For $X_1 \in \mathbb{R}$, the marginal of X_1 is

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{\underline{X}}(\underline{x}) \mathrm{d}x_2 \mathrm{d}x_3.$$

Now $0 < x_1 < 1$,

$$f_{X_1}(x_1) = \int_0^{x_1} \int_0^{x_2} \frac{1}{x_1 x_2} \mathrm{d}x_3 \mathrm{d}x_2 = 1.$$

Thus,

$$f_{X_1}(x) = \begin{cases} 1, \ 0 < x < 1, \\ 0, \ \text{otherwise.} \end{cases}$$

(d) The conditional distribution of X_1 given $(X_2, X_3) = (x_2, x_3)$ is

$$f_{X_1|(X_2,X_3)}(x_1|x_2,x_3) = \frac{f_{\underline{X}}(x_1,x_2,x_3)}{f_{X_2,X_3}(x_2,x_3)} = \frac{(1/x_1x_2)}{(-\ln(x_2)/x_2)} = -\frac{1}{x_1 \ln x_2}, \quad x_2 < x_1 < 1.$$

So, the conditional distribution of X_1 given $(X_2, X_3) = (x_2, x_3)$ is

$$f_{X_1|(X_2,X_3)}(x_1|x_2,x_3) = \begin{cases} -\dfrac{1}{x_1 \ln x_2}, \ x_2 < x_1 < 1, \\ 0, \ \text{otherwise.} \end{cases}$$

(e) We have $S_{\underline{X}} = \{\underline{x} \in \mathbb{R}^3 : 0 \le x_3 \le x_2 \le x_1 \le 1\} \ne S_{X_1} \times S_{X_2} \times S_{X_3} = [0,1] \times [0,1] \times [0,1]$. So, X_1, X_2 and X_3 are not independent.

(f) For fixed $x_2 \in \mathbb{R}$, $f_{X_1,X_3|X_2}(x_1,x_3|x_2) \propto f_{X_1,X_2,X_3}(x_1,x_2,x_3)$. For fixed $0 < x_2 < 1$,

$$f_{X_1,X_3|X_2}(x_1,x_3|x_2) = \begin{cases} \dfrac{c(x_2)}{x_1}, \ 0 < x_3 < x_2, \ x_2 < x_1 < 1, \\ 0, \ \text{otherwise.} \end{cases}$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X_1,X_3|X_2}(x_1,x_3|x_2)\mathrm{d}x_1\mathrm{d}x_3 = 1 \implies c(x_2) = -\frac{1}{x_2 \ln x_2}.$$

Thus, for fixed $0 < x_2 < 1$, $f_{X_1,X_3|X_2}(x_1,x_3|x_2) = g_{x_2}(x_1)h_{x_2}(x_3)$, $(x_1,x_3) \in \mathbb{R}^3$ where for fixed $x_2 \in (0,1)$

$$g_{x_2}(x) = \begin{cases} -\dfrac{1}{x x_2 \ln x_2}, \ x_2 < x < 1, \\ 0, \ \text{otherwise.} \end{cases} \quad ; \quad h_{x_2}(y) = \begin{cases} 1, \ 0 < y < x_2, \\ 0, \ \text{otherwise.} \end{cases}$$

$\implies$ given $X_2 = x_2$ $(0 < x_2 < 1)$ X_1 and X_3 are independently distributed.

17.0.1. Expectation and Moments

Let $\underline{X} = (X_1, X_2, \ldots, X_p)$ be a p-dimensional random vector with p.m.f. / p.d.f. $f(\cdot)$ and support S. Let $g : \mathbb{R}^p \to \mathbb{R}$ be a function.

Definition 17.2. *We say that the expected value of $g(\underline{X})$ (denoted by $E(g(\underline{X}))$) is finite and equals*

$$E(g(\underline{X})) = \begin{cases} \displaystyle\sum_{\underline{x} \in S} g(\underline{x})f(\underline{x}), \ \textit{if } \underline{X} \textit{ is discrete,} \\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(\underline{x})f(\underline{x})\mathrm{d}\underline{x}, \ \textit{if } \underline{X} \textit{ is continuous,} \end{cases}$$

provided $\displaystyle\sum_{\underline{x} \in S} |g(\underline{x})|f(\underline{x}) < \infty \left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} |g(\underline{x})|f(\underline{x})\mathrm{d}\underline{x} < \infty\right)$.

Theorem 17.3. *Let $Y = g(\underline{X})$. Then Y has finite expectation iff* $\sum_{y \in S_Y} |y| f_Y(y) < \infty$ $\left(\text{or } \int_{-\infty}^{\infty} |y| f_Y(y) \mathrm{d}y < \infty\right)$ *and in that case*

$$E(g(\underline{X})) = \sum_{y \in S_Y} y f_Y(y) \quad \left(\int_{-\infty}^{\infty} y f_Y(y) \mathrm{d}y\right).$$

Here, S_Y denotes the support of Y and $f_Y(\cdot)$ denotes the p.m.f. / p.d.f. Y.

Some Special Expectations:

(a) For non-negative integers $k_1, k_2, \dots, k_p$, $\mu'_{k_1, k_2, \dots, k_p} = E\left(X_1^{k_1} X_2^{k_2} \dots X_p^{k_p}\right)$, provided it is finite, is called a joint moment of order $k_1 + k_2 + \cdots + k_p$ of $\underline{X}$.

(b) For non-negative integers $k_1, k_2, \dots, k_p$,

$$\mu_{k_1, k_2, \dots, k_p} = E\left((X_1 - E(X_1))^{k_1} (X_2 - E(X_2))^{k_2} \dots (X_p - E(X_p))^{k_p}\right),$$

provided it is finite, is called a joint central moment of order $k_1 + k_2 + \cdots + k_p$ of $\underline{X}$.

(c) The quantity $\mathrm{Cov}(X_1, X_2)$ [redacted]$((X_1 - E(X_1))(X_2 - E(X_2)))$, provided it is finite, is called covariance between X_1 and X_2.

Remark 17.4. *(a)*

$$\begin{aligned}
\mathrm{Cov}(X_1, X_2) &= E\left((X_1 - E(X_1))(X_2 - E(X_2))\right) \\
&= E\left((X_1 - \mu_1)(X_2 - \mu_2)\right) \\
&= E(X_1 X_2 - X_1 \mu_2 - \mu_2 X_1 + \mu_1 \mu_2) = E(X_1 X_2) - E(X_1) E(X_2).
\end{aligned}$$

(b) $\mathrm{Cov}(X_1, X_1) = \mathrm{Var}(X_1)$.

(c) $\mathrm{Cov}(X_1, X_2) = \mathrm{Cov}(X_2, X_1)$

Theorem 17.5. *Let a_i, $i = 1, 2, \dots, p$ and b_j, $j = 1, 2, \dots, r$ are real constants and let X_i, $i = 1, 2, \dots, p$, Y_j, $j = 1, 2, \dots, r$ be random variables. Then*

(a) $E\left(\sum_{i=1}^{p} a_i X_i\right) = \sum_{i=1}^{p} a_i E(X_i)$, *provided the involved expectations are finite.*

(b) $\mathrm{Cov}\left(\sum_{i=1}^{p} a_i X_i, \sum_{j=1}^{r} b_j Y_j\right) = \sum_{i=1}^{p} \sum_{j=1}^{r} a_i b_j \,\mathrm{Cov}(X_i, Y_j)$, *provided the involved expectations are finite.*

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

Proof. (a) (*We will prove for continuous case*).

$$E\left(\sum_{i=1}^{p} a_i X_i\right) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\sum_{i=1}^{p} a_i x_i\right) f_{\underline{X}}(\underline{x}) \mathrm{d}\underline{x} = \sum_{i=1}^{p} a_i \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_i f_{\underline{X}}(\underline{x}) \mathrm{d}\underline{x} = \sum_{i=1}^{p} a_i E(X_i).$$

(b) Note that

$$\begin{aligned}
\operatorname{Cov}\left(\sum_{i=1}^{p} a_i X_i, \sum_{j=1}^{r} b_j Y_j\right) &= E\left[\left(\sum_{i=1}^{p} a_i X_i - E\left(\sum_{i=1}^{p} a_i X_i\right)\right)\left(\sum_{j=1}^{r} b_j Y_j - E\left(\sum_{j=1}^{r} b_j Y_j\right)\right)\right] \\
&= E\left[\left(\sum_{i=1}^{p} a_i X_i - \sum_{i=1}^{p} a_i E(X_i)\right)\left(\sum_{j=1}^{r} b_j Y_j - \sum_{j=1}^{r} b_j E(Y_j)\right)\right] \\
&= E\left[\left(\sum_{i=1}^{p} a_i (X_i - E(X_i))\right)\left(\sum_{j=1}^{r} b_j (Y_j - E(Y_j))\right)\right] \\
&= E\left[\left(\sum_{i=1}^{p}\sum_{j=1}^{r} a_i b_j (X_i - E(X_i))(Y_j - E(Y_j))\right)\right] \\
&= \sum_{i=1}^{p}\sum_{j=1}^{r} a_i b_j E\left[(X_i - E(X_i))(Y_j - E(Y_j))\right] = \sum_{i=1}^{p}\sum_{j=1}^{r} a_i b_j \operatorname{Cov}(X_i, Y_j).
\end{aligned}$$

(c) Note that

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

This completes the proof. □

Theorem 17.6. *Let $X_1, X_2, \ldots, X_p$ be independent random variables, let $\psi_i : \mathbb{R} \to \mathbb{R}$, $i = 1, 2, \ldots, p$ be given functions. Then,*

(a) $E\left(\prod_{i=1}^{p} \psi_i(X_i)\right) = \prod_{i=1}^{p} E(\psi_i(X_i))$*, provided the involved expectations are finite,*

(b) for any $A_1, A_2, \ldots, A_p \in \mathscr{B}_p$,

$$\Pr(X_1 \in A_1, X_2 \in A_2, \ldots, X_p \in A_p) = \prod_{i=1}^{p} \Pr(X_i \in A_i),$$

(c) $\psi_1(X_1), \psi_2(X_2), \ldots, \psi_p(X_p)$ are independent random variables.

Proof. (*We will prove for $p = 2$ in continuous case*).

(a)

$$\begin{aligned}E(\psi_1(X_1)\psi_2(X_2)) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi_1(X_1)\psi_2(X_2)f_{X_1,X_2}(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi_1(X_1)\psi_2(X_2)f_{X_1}(x_1)f_{X_2}(x_2)\mathrm{d}x_1\mathrm{d}x_2 \;\; (X_1 \perp\!\!\!\perp X_2)\\ &= \left(\int_{-\infty}^{\infty}\psi_1(X_1)f_{X_1}(x_1)\mathrm{d}x_1\right)\left(\int_{-\infty}^{\infty}\psi_2(X_2)f_{X_2}(x_2)\mathrm{d}x_2\right) = E(\psi_1(X_1))E(\psi_2(X_2)).\end{aligned}$$

(b) Take $\psi_i(X_i) = \begin{cases}1, & \text{if } X_i \in A_i,\\ 0, & \text{otherwise,}\end{cases}$ in (a). Note that $\psi_1(X_1)\psi_2(X_2) = \begin{cases}1, & \text{if } X_i \in A_i,\\ 0, & \text{otherwise.}\end{cases}$

$E(\psi_i(X_i)) = \Pr(X_i \in A_i)$, $i = 1,2$ and $E(\psi_1(X_1)\psi_2(X_2)) = \Pr(X_1 \in A_1, X_2 \in A_2)$. Now the result follows from (a).

(c) Let $Y_i = \psi_i(X_i)$, $i = 1,2$. For fixed $\underline{y} = (y_1, y_2) \in \mathbb{R}^2$, define

$$g_i(X_i) = \begin{cases}1, & \text{if } Y_i = \psi_i(X_i) \le y_i, \; i = 1,2,\\ 0, & \text{otherwise.}\end{cases}$$

Then by (a) $E(g_1(X_1)g_2(X_2)) = E(g_1(X_1))E(g_2(X_2))$. Also,

$$\begin{aligned}g_1(X_1)g_2(X_2) &= \begin{cases}1, & \text{if } \psi_1(X_1) \le y_1, \; \psi_2(X_2) \le y_2,\\ 0, & \text{otherwise,}\end{cases}\\ &= \begin{cases}1, & \text{if } Y_1 \le y_1, \; Y_2 \le y_2,\\ 0, & \text{otherwise.}\end{cases}\end{aligned}$$

So, $E(g_1(X_1)g_2(X_2)) = \Pr(Y_1 \le y_1, Y_2 \le y_2)$ and $E(g_i(X_i)) = \Pr(Y_i \le y_i)$, $i = 1,2$. Consequently, $\Pr(Y_1 \le y_1, Y_2 \le y_2) = \Pr(Y_1 \le y_1)\Pr(Y_2 \le y_2) \;\forall\; (y_1, y_2) \in \mathbb{R}^2 \implies Y_1 = \psi_1(X_1)$ and $Y_2 = \psi_2(X_2)$ are independent random variables. □

Corollary 17.7. *Let $X_1, X_2, \ldots, X_p$ are independent random variables. Then*

(a) $\mathrm{Cov}(X_i, X_j) = 0 \;\forall\; i \neq j$.

(b) For real constants $a_1, a_2, \ldots, a_p$, we have

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

Proof. (a) For $i \neq j$, $\mathrm{Cov}(X_i, X_j) = E(X_iX_j) - E(X_i)E(X_j) = E(X_i)E(X_j) - E(X_i)E(X_j) = 0$.

(b) $\mathrm{Var}\left(\sum_{i=1}^{p} a_iX_i\right) = \sum_{i=1}^{p} a^2\,\mathrm{Var}(X_i) + \sum_{i=1}^{p}\sum_{\substack{j=1\\ j\neq i}}^{p} a_ia_j\,\mathrm{Cov}\,(X_i, X_j) = \sum_{i=1}^{p} a^2\,\mathrm{Var}(X_i)$, (using (a)). □

Definition 17.8. *(a) The correlation between random variables X_1 and X_2 is defined by*

$$\rho(X_1, X_2) = \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)\,\mathrm{Var}(X_2)}},$$

provided $0 < \mathrm{Var}(X_i) < \infty$, $i = 1, 2$.

(b) Random variables X_1 and X_2 are said to be uncorrelated if $\rho(X_1, X_2) = 0$ (or equivalently $\mathrm{Cov}(X_1, X_2) = 0$).

Remark 17.9. *If X_1 and X_2 are independent random variables $\implies$ X_1 and X_2 are uncorrelated. converse may not be true.*

Example 17.10 (Uncorrelated random variables may not be independent)**.** *Let (X, Y) have joint p.m.f.*

$$f(x,y) = \begin{cases} \frac{1}{2}, \text{ if } (x,y) = (0,0), \\ \frac{1}{4}, \text{ if } (x,y) = (1,-1), (1,1), \\ 0, \text{ otherwise.} \end{cases}$$

Then

$$f_X(x) = \begin{cases} \frac{1}{2}, \text{ if } x = 0, 1, \\ 0, \text{ otherwise,} \end{cases} \quad f_Y(y) = \begin{cases} \frac{1}{4}, \text{ if } y = -1, 1, \\ \frac{1}{2}, \text{ if } y = 0, \\ 0, \text{ otherwise.} \end{cases}$$

Clearly, there exists $(x,y) \in \mathbb{R}^2$ such that $f_{X,Y}(x,y) \neq f_X(x) f_Y(y) \implies X$ and Y are not independent (in fact $\Pr(X = Y^2) = 1$).

However, $E(XY) = E(Y) = 0$ and $E(X) = \frac{1}{2} \implies \operatorname{Cov}(X,Y) = 0 \implies \rho(X,Y) = 0$.

Theorem 17.11 (Cauchy-Schwarz Inequality)**.** *For random variables X and Y*

$$(E(XY))^2 \leq E(X^2)E(Y^2) \tag{17.1}$$

provided involved expectations are finite. The equality is attained iff $\Pr(Y = cX) = 1$ or $\Pr(X = cY) = 1$, for some real constant c.

Proof. Case I: $E(X^2) = 0$. In this case $\Pr(X = 0) = 1$. Therefore $\Pr(XY = 0) = 1$ and $E(XY) = 0$. We have inequality in (17.1).

Case II: $E(X^2) > 0$. Then

$$\begin{aligned} & E((Y - cX)^2) \geq 0 \,\forall\, c \in \mathbb{R} \\ \implies & c^2 E(X^2) - 2cE(XY) + E(Y^2) \geq 0 \,\forall\, c \in \mathbb{R} \\ \implies & \text{Discriminant} \leq 0 \implies (2E(XY))^2 - 4(E(X^2))E(Y^2) \leq 0 \implies (E(XY))^2 \leq E(X^2)E(Y^2). \end{aligned}$$

Clearly, equality is attained iff $E((Y - cX)^2) = 0$ for some $c \in \mathbb{R} \implies \Pr(Y = cX) = 1$ for some $c \in \mathbb{R}$. By symmetry $\Pr(X = cY) = 1$ for some $c \in \mathbb{R}$. □

Corollary 17.12. *Let X_1 and X_2 be random variables with $E(X_i) = \mu_i \in (-\infty, \infty)$ and $\operatorname{Var}(X_i) = \sigma_i^2 \in (0, \infty)$, $i = 1, 2$. Then*

(a) $|\rho(X_1, X_2)| \leq 1$.

(b) $|\rho(X_1, X_2)| = 1$ iff $\Pr\left(\frac{X_1 - \mu_1}{\sigma_1} = c\frac{X_2 - \mu_2}{\sigma_2}\right) = 1$ or $\Pr\left(\frac{X_2 - \mu_2}{\sigma_2} = c\frac{X_1 - \mu_1}{\sigma_1}\right) = 1$, for some real constant c.

Proof. Let $X = \frac{X_1 - \mu_1}{\sigma_1}$ and $Y = \frac{X_2 - \mu_2}{\sigma_2}$. Using Cauchy-Schwarz inequality $(E(XY))^2 \leq E(X^2)E(Y^2)$ but

$$E(X^2) = \frac{E(X_1 - \mu_1)^2}{\sigma_1^2} = 1 \text{ and } E(Y^2) = \frac{E(X_2 - \mu_2)^2}{\sigma_2^2} = 1.$$

Thus

$$\left(\frac{E((X_1 - \mu_1)(X_2 - \mu_2))}{\sigma_1 \sigma_2}\right)^2 \leq 1 \implies \rho^2(X_1, X_2) \leq 1 \implies |\rho(X_1, X_2)| \leq 1$$

and equality is attained iff $\Pr(X = cY) = 1$, for some real constants $c \implies \Pr\left(\frac{X_1 - \mu_1}{\sigma_1} = c\frac{X_2 - \mu_2}{\sigma_2}\right) = 1$ for some real constants c. □

17.0.2. Conditional Expectation, Conditional Variance and Conditional Covariance

Definition 17.13. *(a) Let $\underline{X}$ be a p-dimensional random vector and $\underline{Y}$ be a q-dimensional random vector. Let $\underline{y} \in \mathbb{R}^q$ be such that $f_{\underline{Y}}(\underline{y}) > 0$ and let $\psi : \mathbb{R}^p \to \mathbb{R}$ be a given function. Here $f_{\underline{Y}}(\cdot)$ is the p.d.f. / p.m.f. of random vector $\underline{Y}$. Then*

(i) The conditional expectation of $\psi(\underline{X})$ given $\underline{Y} = \underline{y}$ (denoted by $E(\psi(\underline{X})|\underline{Y} = \underline{y})$) is the expectation of $\psi(\underline{X})$ under the conditional distribution of $\underline{X}$ given $\underline{Y} = \underline{y}$.

(ii) The conditional variance of $\psi(\underline{X})$ given $\underline{Y} = \underline{y}$ (denoted by $\mathrm{Var}(\psi(\underline{X})|\underline{Y} = \underline{y})$) is the variance of $\psi(\underline{X})$ under the conditional distribution of $\underline{X}$ given $\underline{Y} = \underline{y}$.

(b) Let X_1 and X_2 be two random variables and $\underline{Y}$ be a q-dimensional random vector. Then the conditional covariance between X_1 and X_2 given $\underline{Y} = \underline{y}$, (denoted by $\mathrm{Cov}(X_1, X_2|\underline{Y} = \underline{y})$) is the covariance between X_1 and X_2 under the conditional distribution of (X_1, X_2) given $\underline{Y} = \underline{y}$.

Notation Let for $\underline{y} \in \{\underline{t} \in \mathbb{R}^q : f_{\underline{Y}}(\underline{t}) > 0\}$, $\psi_1(\underline{y}) = E(\psi(\underline{X})|\underline{Y} = \underline{y})$ and $\psi_2(\underline{y}) = \mathrm{Var}(\psi(\underline{X})|\underline{Y} = \underline{y})$ and $\psi_3(\underline{y}) = \mathrm{Cov}(X_1, X_2|\underline{Y} = \underline{y})$. We denote $\psi_1(\underline{Y}) = E(\psi(\underline{X})|\underline{Y})$ and $\psi_2(\underline{Y}) = \mathrm{Var}(\psi(\underline{X})|\underline{Y})$ and $\psi_3(\underline{Y}) = \mathrm{Cov}(X_1, X_2|\underline{Y})$.

Theorem 17.14. *Under the above notation*

(a) $E(\psi(\underline{X})) = E(E(\psi(\underline{X})|\underline{Y}))$,

(b) $\mathrm{Var}(\psi(\underline{X})) = \mathrm{Var}(E(\psi(\underline{X})|\underline{Y})) + E(\mathrm{Var}(\psi(\underline{X})|\underline{Y}))$,

(c) $\mathrm{Cov}(X_1, X_2) = \mathrm{Cov}(E(X_1|\underline{Y}), E(X_2|\underline{Y})) + E(\mathrm{Cov}(X_1, X_2|\underline{Y}))$.

Proof. (*We will prove for $p = q = 1$ continuous case*).
(a)

$$\begin{aligned} E(E(\psi(\underline{X})|\underline{Y})) &= \int_{-\infty}^{\infty} E(\psi(\underline{X})|\underline{Y} = \underline{y}) f_{\underline{Y}}(\underline{y}) \mathrm{d}\underline{y} \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \psi(\underline{x}) f_{X|Y}(x|y) \mathrm{d}x\right] f_{\underline{Y}}(\underline{y}) \mathrm{d}\underline{y} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi(x) f_{X|Y}(x|y) f_Y(y) \mathrm{d}x\mathrm{d}y = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi(x) f_{X,Y}(x,y) \mathrm{d}x\mathrm{d}y = E(\psi(X)). \end{aligned}$$

(b) Follows from (c).

(c) From (a), we have

$$\mathrm{Cov}(X_1, X_2) = E((X_1 - E(X_1))(X_2 - E(X_2))) = E\left[E\left[(X_1 - E(X_1))(X_2 - E(X_2))|\underline{Y}\right]\right].$$

Now,

$$
\begin{aligned}
& E\left[(X_1 - E(X_1))(X_2 - E(X_2))|\underline{Y}\right] \\
& \quad = E\left[(X_1 - E(X_1|\underline{Y}) + E(X_1|\underline{Y}) - E(X_1))(X_2 - E(X_2|Y) + E(X_2|Y) - E(X_2)|\underline{Y})\right] \\
& \quad = E\left[(X_1 - E(X_1|Y))(X_2 - E(X_2|\underline{Y}))|\underline{Y}\right] + (E(X_1|\underline{Y}) - E(X_1))(E(X_2|\underline{Y}) - E(X_2)) \\
& \quad = \mathrm{Cov}(X_1, X_2|\underline{Y}) + (E(X_1|\underline{Y}) - E(X_1))(E(X_2|Y) - E(X_2)). \\
\implies \mathrm{Cov}(X_1, X_2) & = E(\mathrm{Cov}(X_1, X_2|\underline{Y})) + E\left[(E(X_1|\underline{Y}) - E(X_1))(E(X_2|Y) - E(X_2))\right] \\
& = \mathrm{Cov}(E(X_1|\underline{Y}), E(X_2|\underline{Y})) + E(\mathrm{Cov}(X_1, X_2|\underline{Y})).
\end{aligned}
$$

This completes the proof. $\square$

17.0.3. Joint Moment Generating Function

Let $\underline{X} = (X_1, X_2, \ldots, X_p)$ be a p-dimensional random vector with p.d.f. /p.m.f. $f_{\underline{Y}}(\cdot)$. $A = \{\underline{t} = (t_1, t_2, \ldots, t_p) \in \mathbb{R}^p : E\left(e^{\sum_{i=1}^p t_i X_i}\right) < \infty\}$.

Definition 17.15. *(a) The function $M_{\underline{X}} : A \to \mathbb{R}$ defined by*

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

is called the joint moment generating function (m.g.f.) of random vector $\underline{X} = (X_1, X_2, \ldots, X_p)$.

Notation: For $\underline{a} = (a_1, a_2, \ldots, a_p) \in \mathbb{R}^p$, $-\underline{a} = (-a_1, -a_2, \ldots, -a_p)$ and $(-\underline{a}, \underline{a}) = (-a_1, a_1) \times \cdots \times (-a_p, a_p)$, $\underline{a} = (a_1, a_2, \ldots, a_p) > 0 \iff a_i > 0, i = 1, 2, \ldots, p$.

Remark 17.16. *(i) As $M_{\underline{X}}(\underline{0}) = 1$, we have $A \neq \phi$. Moreover $M_{\underline{X}}(\underline{t}) > 0 \;\forall\, \underline{t} \in A$.*

(ii) If $X_1, X_2, \ldots, X_p$ are independent then

$$M_{\underline{X}}(\underline{t}) = E\left(e^{\sum_{i=1}^p t_i X_i}\right) = E\left(\prod_{i=1}^p e^{t_i X_i}\right) = \prod_{i=1}^p E\left(e^{t_i X_i}\right) = \prod_{i=1}^p M_{X_i}(t_i) \;\forall\, \underline{t} \in A.$$

Conversely, suppose that $A \subseteq (-\underline{a}, \underline{a})$ for some $\underline{a} > 0$ and $M_{\underline{X}}(\underline{t}) = \prod_{i=1}^p M_{X_i}(t_i) \;\forall\, \underline{t} \in A$, then it can be shown that $X_1, X_2, \ldots, X_p$ are independent.

(iii) Let $X_1, X_2, \ldots, X_p$ be independent random variables and let $Y = \sum_{i=1}^p X_i$, then

$$M_Y(t) = E\left(e^{t\sum_{i=1}^p X_i}\right) = E\left(\prod_{i=1}^p e^{tX_i}\right) = \prod_{i=1}^p E\left(e^{tX_i}\right) = \prod_{i=1}^p M_{X_i}(t), \; t \in A.$$

In particular, if $X_1, X_2, \ldots, X_p$ are independent and identically distributed (iid) with common m.g.f. $M(t)$, then $M_Y(t) = (M(t))^p$, $t \in A$.

Theorem 17.17. *Suppose that the joint m.g.f. $M_{\underline{X}}(\underline{t})$ is finite on a rectangle $(-\underline{a}, \underline{a}) \in \mathbb{R}^p$, $\underline{a} > 0$. Then $M_{\underline{X}}(\underline{t})$ posseses partial derivatives of all order in $(-\underline{a}, \underline{a})$. Furthermore, for non-negative integers $k_1, k_2, \ldots, k_p$*

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

Proof. (*We give an outline of the proof*).

$$M_{\underline{X}}(t_1, t_2, \ldots, t_p) = E\left(e^{\sum_{i=1}^p t_i X_i}\right) = \int_{\mathbb{R}^p} e^{\sum_{i=1}^p t_i x_i} f_{\underline{X}}(\underline{x}) \mathrm{d}\underline{x},$$

$$\frac{\partial^{k_1+\cdots+k_p}}{\partial t_1^{k_1} \cdots \partial t_p^{k_p}} M_{\underline{X}}(\underline{t}) = \int_{\mathbb{R}^p} x_1^{k_1} \cdots x_p^{k_p} e^{\sum_{i=1}^p t_i X_i} f_{\underline{X}}(\underline{x}) \mathrm{d}\underline{x},$$

$$\left[\frac{\partial^{k_1+\cdots+k_p}}{\partial t_1^{k_1} \cdots \partial t_p^{k_p}} M_{\underline{X}}(\underline{t})\right]_{\underline{t}=\underline{0}} = \int_{\mathbb{R}^p} x_1^{k_1} \cdots x_p^{k_p} f_{\underline{X}}(\underline{x}) \mathrm{d}\underline{x} = E\left(X_1^{k_1} X_2^{k_2} \ldots X_p^{k_p}\right).$$

This completes the proof. □

Let $\psi_{\underline{X}}(\underline{t}) = \ln M_{\underline{X}}(\underline{t})$, $\underline{t} \in (-\underline{a}, \underline{a})$. Then

$$E(X_i) = \left[\frac{\partial}{\partial t_i} M_{\underline{X}}(\underline{t})\right]_{\underline{t}=\underline{0}} = \left[\frac{\partial}{\partial t_i} \psi_{\underline{X}}(\underline{t})\right]_{\underline{t}=\underline{0}},$$

$$E(X_i^m) = \left[\frac{\partial^m}{\partial t_i^m} M_{\underline{X}}(\underline{t})\right]_{\underline{t}=\underline{0}}, \quad m = 1, 2, \ldots, \ i = 1, 2, \ldots, p,$$

$$\mathrm{Var}(X_i) = \left[\frac{\partial^2}{\partial t_i^2} M_{\underline{X}}(\underline{t})\right]_{\underline{t}=\underline{0}} - \left(\left[\frac{\partial}{\partial t_i} M_{\underline{X}}(\underline{t})\right]_{\underline{t}=\underline{0}}\right)^2 = \left[\frac{\partial^2}{\partial t_i^2} \psi_{\underline{X}}(\underline{t})\right]_{\underline{t}=\underline{0}}, \ i = 1, 2, \ldots, p,$$

provided $M_{\underline{X}}(\underline{t})$ is finite on $(-\underline{a}, \underline{a})$, for some $\underline{a} > 0$. For $i \neq j$, if $M_{\underline{X}}(\underline{t})$ is finite on $(-\underline{a}, \underline{a})$, for some $\underline{a} > 0$,

$$\begin{aligned}\mathrm{Cov}(X_i, X_j) &= E(X_i X_j) - E(X_i)E(X_j) = E((X_i - E(X_i))(X_j - E(X_j))) \\ &= \left[\frac{\partial^2}{\partial t_i \partial t_j} M_{\underline{X}}(\underline{t})\right]_{\underline{t}=\underline{0}} - \left[\frac{\partial}{\partial t_i} M_{\underline{X}}(\underline{t})\right]_{\underline{t}=\underline{0}} \left[\frac{\partial}{\partial t_j} M_{\underline{X}}(\underline{t})\right]_{\underline{t}=\underline{0}} = \left[\frac{\partial^2}{\partial t_i \partial t_j} \psi_{\underline{X}}(\underline{t})\right]_{\underline{t}=\underline{0}}.\end{aligned}$$

Moreover,

$$M_{\underline{X}}(0, \ldots, 0, t_i, 0, \ldots, 0) = E(e^{t_i X_i}) = M_{X_i}(t_i), \ i = 1, 2, \ldots, p,$$
$$M_{\underline{X}}(0, \ldots, 0, t_i, 0, \ldots, 0, t_j, 0, \ldots, 0) = E(e^{t_i X_i + t_j X_j}) = M_{X_i, X_j}(t_i, t_j),$$

provided the m.g.f. is finite.

17.0.4. Equality in Distribution

Definition 17.18. *Two p-dimensional random vectors $\underline{X}$ and $\underline{Y}$ are said to have the same distribution (written as $\underline{X} \stackrel{d}{=} \underline{Y}$) if $F_{\underline{X}}(\underline{x}) = F_{\underline{Y}}(\underline{x})\ \forall\ \underline{x} \in \mathbb{R}^p$.*

Theorem 17.19. *(a) Let $\underline{X}$ and $\underline{Y}$ be discrete random vectors with p.m.f.s $f_{\underline{X}}(\cdot)$ and $f_{\underline{Y}}(\cdot)$, respectively. Then*

$$\underline{X} \stackrel{d}{=} \underline{Y} \iff f_{\underline{X}}(\underline{x}) = f_{\underline{Y}}(\underline{x})\ \forall\ \underline{x} \in \mathbb{R}^p.$$

(b) Let $\underline{X}$ and $\underline{Y}$ be continuous random vectors. Then

$$\underline{X} \stackrel{d}{=} \underline{Y} \iff f_{\underline{X}}(\underline{x}) = f_{\underline{Y}}(\underline{x})\ \forall\ \underline{x} \in \mathbb{R}^p,$$

for some versions $f_{\underline{X}}(\cdot)$ and $f_{\underline{Y}}(\cdot)$ of p.d.f.s of $\underline{X}$ and $\underline{Y}$, respectively.

(c) Let $\underline{X}$ and $\underline{Y}$ be p-dimensional random vectors and let $\psi : \mathbb{R}^p \to \mathbb{R}^q$ be a given function. Then

$$\underline{X} \stackrel{d}{=} \underline{Y} \iff \psi(\underline{X}) \stackrel{d}{=} \psi(\underline{Y})$$

(d) Let $\underline{X}$ and $\underline{Y}$ be p-dimensional random vectors with finite m.g.f.s $M_{\underline{X}}(\underline{t})$ and $M_{\underline{Y}}(\underline{t})$ on a rectangle $(-\underline{a}, \underline{a})$, for some $\underline{a} > 0$. Then

$$M_{\underline{X}}(\underline{t}) = M_{\underline{Y}}(\underline{t}) \; \forall \, (-\underline{a}, \underline{a}) \implies \underline{X} \stackrel{d}{=} \underline{Y}.$$

17.0.5. Some Generalizations

Let $\underline{X}_i$: a p_i- dimensional random vector, $i = 1, 2, \ldots, m$. $F_{\underline{X}_i}$: d.f. of $\underline{X}_i$, $i = 1, 2, \ldots, m$, $f_{\underline{X}_i}$: p.m.f. / p.d.f. of $\underline{X}_i$, $i = 1, 2, \ldots, m$, $\sum_{i=1}^{p} p_i = p$, $\underline{X} = (\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_m)$: p-dimensional random vector with d.f. $F_{\underline{X}_i}(\cdot)$ and p.m.f. / p.d.f. $f_{\underline{X}_i}(\cdot)$.

Definition 17.20. *The random vectors $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_m$ are said to be independent if for any subcollection $\{\underline{X}_{i_1}, \underline{X}_{i_2}, \ldots, \underline{X}_{i_q}\}$ of $\{\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_m\}$ $(2 \leq q \leq m)$*

$$F_{\underline{X}_{i_1}, \underline{X}_{i_2}, \ldots, \underline{X}_{i_q}}(\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_q) = \prod_{j=1}^{q} F_{\underline{X}_{i_j}}(\underline{x}_j) \; \forall \, \underline{x} = (\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_q) \in \mathbb{R}^{\sum_{j=1}^{q} p_{i_j}}.$$

Remark 17.21. *$\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_m$ are independent $\implies$ random variables in any subset of $\{\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_m\}$ are independent.*

Theorem 17.22. *(a) The following statements are equivalent:*

(i) $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_m$ are independent random vectors.

(ii) $F_{\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_m}(\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m) = \prod_{i=1}^{m} F_{\underline{X}_i}(\underline{x}_i) \; \forall \, \underline{x} = (\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m) \in \mathbb{R}^p$.

(iii) $f_{\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_m}(\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m) = \prod_{i=1}^{m} f_{\underline{X}_i}(\underline{x}_i) \; \forall \, \underline{x} = (\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m) \in \mathbb{R}^p$.

(iv) $f_{\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_m}(\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m) = \prod_{i=1}^{m} g_i(\underline{x}_i) \; \forall \, \underline{x} = (\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m) \in \mathbb{R}^p$ for some non-negative real valued function $g_i : \mathbb{R}^p \to \mathbb{R}$, $i = 1, 2, \ldots, m$.

(v) $\Pr(\underline{X}_i \in A_i, \; i =, 2, \ldots, m) = \prod_{i=1}^{m} \Pr(\underline{X}_i \in A_i) \; \forall \, A_i \in \mathscr{B}_{p_i}$, $i = 1, 2, \ldots, m$.

(b) If $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_m$ are independent random vectors, then

(i) $E\left(\prod_{i=1}^{m} \psi_i(\underline{X}_i)\right) = \prod_{i=1}^{m} E\left(\psi_i(\underline{X}_i)\right)$ for any functions ψ_i, $i = 1, 2, \ldots, m$.

(ii) $\psi_1(\underline{X}_1), \psi_2(\underline{X}_2), \ldots, \psi_m(\underline{X}_m)$ are independent random vectors for any functions $\psi_1, \psi_2, \ldots, \psi_m$.

Definition 17.23. *Let Δ be an arbitrary index set. The random vectors $\{\underline{X}_\lambda : \lambda \in \Delta\}$ are said to be independent if random variables in any finite subcollection of $\{\underline{X}_\lambda : \lambda \in \Delta\}$ are independent.*

Theorem 17.24. *Under the notation of Theorem 17.22, $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_m$ are independent random vectors $\iff$ for some $\underline{a} > 0$ and $\forall \, \underline{t} = (\underline{t}_1, \underline{t}_2, \ldots, \underline{t}_m) \in (-\underline{a}, \underline{a})$,*

$$M_{\underline{X}}(\underline{t}_1, \underline{t}_2, \ldots, \underline{t}_m) = \prod_{i=1}^{m} M_{\underline{X}_i}(\underline{t}_i).$$