MODULE 6

RANDOM VECTOR AND ITS JOINT DISTRIBUTION

LECTURE 31

Topics

6.7 PROPERTIES OF RANDOM VECTORS HAVING THE SAME DISTRIBUTION

6.7.1 Uniqueness Theorem

6.8 MULTINOMIAL DISTRIBUTION

6.8.1 Multinomial Distribution

6.9 BIVARIATE NORMAL DISTRIBUTION

6.7 PROPERTIES OF RANDOM VECTORS HAVING THE SAME DISTRIBUTION

Definition 7.1

Let $\underline{X}$ and $\underline{Y}$ be two p-dimensional random vectors, defined on the same probability space $(\Omega, \mathcal{F}, P)$. Then $\underline{X}$ and $\underline{Y}$ are *said to have the same distribution* (written as $\underline{X} \stackrel{d}{=} \underline{Y}$) if $F_{\underline{X}}(\underline{x}) = F_{\underline{Y}}(\underline{x})$, $\forall \underline{x} \in \mathbb{R}^p$ (i.e., if they have the same distribution function). ■

The following results are multivariate analogs of theorems stated in Section 4 of Module 3. The proofs of these theorems, being similar to their univariate counterparts, are omitted.

Theorem 7.1

(i) Let $\underline{X}$ and $\underline{Y}$ be p-dimensional random vectors of discrete type with joint p.m.f. $f_{\underline{X}}(\cdot)$ and $f_{\underline{Y}}(\cdot)$, respectively. Then $\underline{X} \stackrel{d}{=} \underline{Y}$ if, and only if, $f_{\underline{X}}(\underline{x}) = f_{\underline{Y}}(\underline{x})$, $\forall \underline{x} \in \mathbb{R}^p$.

(ii) Let $\underline{X}$ and $\underline{Y}$ be p-dimensional random vectors having distribution functions $F_{\underline{X}}(\cdot)$ and $F_{\underline{Y}}(\cdot)$, respectively. Suppose that

$$\frac{\partial^p F_{\underline{X}}(\underline{x})}{\partial x_1 \cdots \partial x_p} \quad \text{and} \quad \frac{\partial^p F_{\underline{Y}}(\underline{x})}{\partial x_1 \cdots \partial x_p}$$

exist everywhere except, possibly, on a set C comprising of countable number of curves. Further suppose that

$$\int_{\mathbb{R}^p} \frac{\partial^p F_{\underline{X}}(\underline{x})}{\partial x_1 \cdots \partial x_p} I_{C^c}(\underline{x}) d\underline{x} = \int_{\mathbb{R}^p} \frac{\partial^p F_{\underline{Y}}(\underline{x})}{\partial x_1 \cdots \partial x_p} I_{C^c}(\underline{x}) d\underline{x} = 1.$$

Then both of them are of absolutely continuous type. Moreover, $\underline{X} \stackrel{d}{=} \underline{Y}$ if and only if, there exist versions of p.d.f.s $f_{\underline{X}}(\cdot)$ and $f_{\underline{Y}}(\cdot)$ of $\underline{X}$ and $\underline{Y}$, respectively, such that $f_{\underline{X}}(\underline{x}) = f_{\underline{Y}}(\underline{x})$, $\forall \underline{x} \in \mathbb{R}^p$. ■

Theorem 7.2

Let $\underline{X}$ and $\underline{Y}$ be p-dimensional random vectors of either discrete type or of absolutely continuous type with $\underline{X} \stackrel{d}{=} \underline{Y}$. Then

(i) For any Borel function $h: \mathbb{R}^p \to \mathbb{R}$, $E\left(h(\underline{X})\right) = E\left(h(\underline{Y})\right)$, provided the expectations are finite;

(ii) For any Borel function $\psi: \mathbb{R}^p \to \mathbb{R}$, $\psi(\underline{X}) \stackrel{d}{=} \psi(\underline{Y})$. ■

6.7.1 Uniqueness Theorem

Theorem 7.3

Let $\underline{X}$ and $\underline{Y}$ be two random vectors of either discrete type or of absolutely continuous type having m.g.f.s $M_{\underline{X}}(\cdot)$ and $M_{\underline{Y}}(\cdot)$, respectively, that are finite on a rectangle $(-\underline{a}, \underline{a})$ for some $\underline{a} = (a_1, a_2, \dots, a_p) \in \mathbb{R}^p$; here $-\underline{a} = (-a_1, -a_2, \dots, -a_p)$ and $(-\underline{a}, \underline{a}) = \{\underline{t} \in \mathbb{R}^p: -a_i < t_i < a_i,\ i = 1, \dots, p.\}$. Suppose that

$$M_{\underline{X}}(\underline{t}) = M_{\underline{Y}}(\underline{t}), \quad \forall \underline{t} \in (-\underline{a}, \underline{a}).$$

Then $\underline{X} \stackrel{d}{=} \underline{Y}$. ■

Remark 7.1

If $X_1, X_2, \dots, X_p$ are independent and identically distributed $\left(\text{i.e.}; \ X_i \stackrel{d}{=} X_1, i = 1, \dots, p\right)$, $Y = \sum_{i=1}^{p} X_i$ and $\bar{X} = \frac{1}{p} \sum_{i=1}^{p} X_i$, then

$$M_{\underline{X}}(\underline{t}) = \prod_{i=1}^{p} M_{X_1}(t_i), \quad \underline{t} \in \mathbb{R}^p$$

$$M_Y(t) = \left[M_{X_1}(t)\right]^p, \quad t \in \mathbb{R}$$

and

$$M_{\bar{X}}(t) = \left[M_{X_1}\left(\frac{t}{p}\right)\right]^p, \quad t \in \mathbb{R},$$

provided the expectations are finite. ■

Example 7.1

Let $X_1, X_2, \dots, X_p$ be independent random variable such that $X_i \sim N(\mu_i, \sigma_i^2)$, $-\infty < \mu_i < \infty$, $\sigma_i > 0, i = 1, \dots, p$. If $a_1, \dots, a_p$ are real constants, such that not all of them are zero, then show that

$$\sum_{i=1}^{p} a_i X_i \sim N\left(\sum_{i=1}^{p} a_i \mu_i \,, \sum_{i=1}^{p} a_i^2 {\sigma_i}^2\right).$$

Solution. Let $Y = \sum_{i=1}^{p} a_i X_i$. Then

$$M_Y(t) = E\left(e^{t \sum_{i=1}^{p} a_i X_i}\right)$$

$$= E\left(\prod_{i=1}^{p} e^{t a_i X_i}\right)$$

$$= \prod_{i=1}^{p} E(e^{t a_i X_i}) \qquad (X_1, X_2, \cdots, X_p \text{ are independent})$$

$$= \prod_{i=1}^{p} M_{X_i}(t a_i)$$

$$= \prod_{i=1}^{p} e^{t a_i \mu_i + \frac{{a_i}^2 {\sigma_i}^2 t^2}{2}}, \quad t \in \mathbb{R}$$

$$= e^{t\left(\sum_{i=1}^{p} a_i \mu_i\right) + \frac{\left(\sum_{i=1}^{p} {a_i}^2 {\sigma_i}^2\right) t^2}{2}}, \quad t \in \mathbb{R}\,,$$

which is the m.g.f. of $N\left(\sum_{i=1}^{p} a_i\mu_i\,, \sum_{i=1}^{p} a_i^2{\sigma_i}^2\right)$ distribution. Using Theorem 7.3 it follows that

$$Y \sim N\left(\sum_{i=1}^{p} a_i\mu_i\,, \sum_{i=1}^{p} a_i^2{\sigma_i}^2\right). \blacksquare$$

Example 7.2

Let $X_1, X_2, \ldots, X_p$ be independent random variable such that $X_i \sim \text{Bin}(n_i, \theta),\ 0 < \theta < 1, n_i \in \{1, 2, \ldots\},\ i = 1, \ldots, p$. Show that

$$\sum_{i=1}^{p} X_i \sim \text{Bin}\left(\sum_{i=1}^{p} n_i\,, \theta\right).$$

Solution. Let $Y = \sum_{i=1}^{p} X_i$. Then

$$M_Y(t) = E\left(e^{t\sum_{i=1}^{p} X_i}\right)$$

$$= E\left(\prod_{i=1}^{p} e^{tX_i}\right)$$

$$= \prod_{i=1}^{p} E(e^{tX_i})$$

$$= \prod_{i=1}^{p} M_{X_i}(t)$$

$$= \prod_{i=1}^{p} (1-\theta+\theta e^t)^{n_i},\quad t \in \mathbb{R}$$

$$= (1-\theta+\theta e^t)^{\sum_{i=1}^{p} n_i},\quad t \in \mathbb{R},$$

which is the m.g.f. of $\text{Bin}\left(\sum_{i=1}^{p} n_i\,, \theta\right)$ distribution. Using Theorem 7.3 it follows that

$$Y = \sum_{i=1}^{p} X_i \sim \text{Bin}\left(\sum_{i=1}^{p} n_i\,, \theta\right). \blacksquare$$

Example 7.3

Let $X_1, X_2, \ldots, X_p$ be independent random variables such that $X_i \sim \text{NB}(r_i, \theta)\,, 0 < \theta < 1$, $r_i \in \{1, 2, \ldots\},\ i = 1,2, \ldots, p$. Then show that

$$Y = \sum_{i=1}^{p} X_i \sim \text{NB}\left(\sum_{i=1}^{p} r_i\,, \theta\right).$$

Solution. Similar to solution of Example 7.2 on noting that if $X \sim \text{NB}(r, \theta)$ then

$$M_X(t) = \left(\frac{\theta}{1-(1-\theta)e^t}\right)^r, \quad t < -\ln(1-\theta). \blacksquare$$

Example 7.4

Let $X_1, X_2, \dots, X_p$ be independent random variables such that $X_i \sim P(\lambda_i)$, $\lambda_i > 0, i = 1, \dots, p$. Then show that

$$\sum_{i=1}^{p} X_i \sim P\left(\sum_{i=1}^{p} \lambda_i\right).$$

Solution. Similar to solution of Example 7.2 on noting that if $X \sim P(\lambda)$, $\lambda > 0$, then

$$M_X(t) = e^{\lambda(e^t - 1)}, \ \ t \in \mathbb{R}\,. \blacksquare$$

Example 7.5

Let $X_1, X_2, \dots, X_p$ be independent random variable such that $X_i \sim G(\alpha_i, \theta)$, $\theta > 0$, $\alpha_i > 0$, $i = 1, \dots, p$. Show that

$$\sum_{i=1}^{p} X_i \sim G\left(\sum_{i=1}^{p} \alpha_i\,, \theta\right).$$

Solution. Similar to solution of Example 7.2 on noting that if $X \sim G(\alpha, \theta)$, $\alpha > 0$, $\theta > 0$, then

$$M_X(t) = (1 - t\theta)^{-\alpha}, \quad t < \frac{1}{\theta}. \blacksquare$$

Example 7.6

(i) Let $X_1, X_2, \dots, X_p$ be independent random variables such that $X_i \sim \chi^2_{n_i}$, $n_i \in \{1, 2, \dots\}, i = 1, \dots, p$. Then show that

$$\sum_{i=1}^{p} X_i \sim \chi^2_{\sum_{i=1}^{p} n_i}.$$

(ii) Let $Y_1, Y_2, \ldots, Y_p$ be independent random variables such that $Y_i \sim N(\mu, \sigma^2)$, $i = 1, \ldots, p$, $-\infty < \mu < \infty, \sigma > 0$. Then

$$\sum_{i=1}^{p} \left(\frac{Y_i - \mu}{\sigma}\right)^2 \sim \chi_p^2.$$

Solution.

(i) Note that $X_i \sim \chi_{n_i}^2 = G\left(\frac{n_i}{2}, 2\right), i = 1, \ldots, p$. Now the assertion follows from Example 7.5.

(ii) Follows on using Theorem 4.19 (i)-(ii) of Module 5 and (i) above. ■

We state the following theorem without providing its proof.

Theorem 7.4

Let $\underline{X}$ be a p-dimensional random vector and let $\underline{X} = (\underline{X}_1, \ldots, \underline{X}_k)$, where $\underline{X}_i$ is p_i-dimensional, $i = 1, \ldots, k$, $\sum_{i=1}^{k} p_i = p$. Suppose that there exist $\underline{a}_i \in \mathbb{R}^{p_i}$, $\underline{a}_i \neq \underline{0}$, $i = 1, \ldots, k$, such that $M_{\underline{X}}(\cdot)$ is finite on $(-\underline{a}, \underline{a})$and $M_{\underline{X}_i}(\cdot)$ is finite on $(-\underline{a}_i, \underline{a}_i)$, $i = 1, \ldots, k$, where $\underline{a} = (\underline{a}_1, \ldots, \underline{a}_k)$, and $-\underline{a} = (-\underline{a}_1, \ldots, -\underline{a}_k)$. Then $\underline{X}_1, \ldots, \underline{X}_k$ are independent iff

$$M_{\underline{X}}(\underline{t}_1, \ldots, \underline{t}_k) = \prod_{i=1}^{k} M_{\underline{X}_i}(\underline{t}_i), \ \ \forall \underline{t}_i \in (-\underline{a}_i, \underline{a}_i), i = 1, \ldots, k. \blacksquare$$

6.8 MULTINOMIAL DISTRIBUTION

First let us introduce the notion of multinomial coefficients, which is a generalization of notion of binomial coefficients.

Let $k, n_1, \ldots, n_{k-1}$ and n be non-negative integers such that $k \geq 2$, $\sum_{i=1}^{k-1} n_i \leq n$. Consider a collection of n items comprising of

n_1 identical items of type 1
n_2 identical items of type 2
$\vdots$
n_{k-1} identical items of type $k-1$
$n_k = n - \sum_{i=1}^{k-1} n_i$ identical items of type k.

The number of visually distinguishable ways in which these n items can be arranged in a row is

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\cdots\binom{n-\sum_{i=1}^{k-2}n_i}{n_{k-1}} = \frac{n!}{n_1!\,n_2!\cdots n_{k-1}!\,(n-\sum_{i=1}^{k-1}n_i)!}.$$

The coefficients

$$\binom{n}{n_1 n_2 \cdots n_{k-1}} = \frac{n!}{n_1!\,n_2!\cdots n_{k-1}!\,(n-\sum_{i=1}^{k-1}n_i)!}\ , n_i \geq 0, i=1,\ldots,k-1, \sum_{i=1}^{k-1} n_i \leq n \ (8.1)$$

are called multinomial coefficients.

Note that, for $k=2$ (so that $0 \leq n_1 \leq n$), multinomial coefficients (8.1) reduce to binomial coefficients

$$\binom{n}{n_1} = \frac{n!}{n_1!\,(n-n_1)!}\ , \quad n_1 \in \{0,1,\ldots,n\}.$$

Also note that, for real numbers $x_1,\ldots,x_k$,

$$(x_1+x_2\cdots+x_k)^n = \underbrace{(x_1+x_2+\cdots+x_k)(x_1+x_2+\cdots+x_k)\cdots(x_1+x_2+\cdots+x_k)}_{\text{Product of } n \text{ quantities}}.$$

A typical term in expansion of above product is an arrangement of $n_1\ x_1^{'}s,\ n_2\ x_2^{'}s, \ldots$, $n_{k-1}\ x_{k-1}^{'}s$ and $n_k = (n-\sum_{i=1}^{k-1}n_i)\ x_k^{'}s,\ n_i \in \{0,1,\ldots\},\ n_1+n_2+\cdots+n_{k-1} \leq n$ (such as $x_1, x_3, x_4, x_2, x_1, x_2 \ldots x_{k-2}x_8$). Each such term equals ${x_1}^{n_1}{x_2}^{n_2}\cdots {x_k}^{n_k}$ and total number of visually distinguishable ways of arranging $n_1\ x_1^{'}s, n_2\ x_2^{'}s, \ldots, n_{k-1}\ x_{k-1}^{'}s$ $\left(n-\sum_{i=1}^{k-1}n_i\right) x_k^{'}s$ is $\binom{n}{n_1 n_2 \cdots n_{k-1}}$.

Thus, we have

$$(x_1+x_2+\cdots+x_k)^n = \sum_{\substack{n_1=0 \\ n_1+n_{2+\cdots}}}^{n} \cdots \sum_{\substack{n_{k-1}=0 \\ +n_{k-1}\leq n}}^{n} \binom{n}{n_1 n_2 \cdots n_{k-1}} x_1^{n_1}x_2^{n_2}\cdots x_k^{n_k}.$$

6.8.1 Multinomial Distribution

Example 8.1

Consider a random experiment that can result in one of $p+1$ $(p \geq 1)$ possible outcomes $A_1, A_2, \dots, A_{p+1}$, where $A_i \cap A_j = \phi, i \neq j$ and $\bigcup_{i=1}^{p+1} A_i = \Omega$. Let $P(A_i) = \theta_i \in (0,1), i = 1, \dots, p$, and $\sum_{i=1}^{p} \theta_i < 1$ so that $P(A_{p+1}) = 1 - \sum_{i=1}^{p} \theta_i \in (0,1)$. Suppose that the random experiment is repeated n times independently.

Define

$$X_i = \text{number of times event } A_i \text{ occurs in } n \text{ trials}, i = 1, \dots, p+1.$$

Then one may be interested in the joint probability distribution of $\underline{X} = (X_1, X_2, \dots, X_{p+1})$. Note that

$$X_{p+1} = n - \sum_{i=1}^{p} X_i = \text{number of times } A_{p+1} \text{ occurs}$$

is completely determined by $\underline{X} = (X_1, X_2, \dots, X_p)$ and therefore only distribution of $\underline{X} = (X_1, \dots, X_p)$ may be of interest. Let $S_{\underline{X}} = \{\underline{x} = (x_1, \dots, x_p): x_i \in \{0, 1, \dots, n\}, i = 1, \dots, p, \ \sum_{i=1}^{p} x_i \leq n\}$. Then

$$f_{\underline{X}}(x_1, \dots, x_p) = P(\{X_1 = x_1, \dots, X_p = x_p\})$$

$$= \begin{cases} \dfrac{n!}{x_1! \cdots x_p! \, (n - \sum_{i=1}^{p} x_i)!} \theta_1^{x_1} \cdots \theta_p^{x_p} \left(1 - \sum_{i=1}^{p} \theta_i\right)^{(n - \sum_{i=1}^{p} x_i)} & \text{if } \underline{x} \in S_{\underline{X}} \\ 0, & \text{otherwise} \end{cases} \quad (8.2) \blacksquare$$

Definition 8.1

The probability distribution given by (8.2) is called a multinomial distribution with n trials and cell probabilities $\theta_1, \dots, \theta_p$ $\left(\text{denoted by Mult}(\text{n}, \theta_1, \dots, \theta_p)\right)$. $\blacksquare$

Note that, for $p = 1$, $\text{Mult}(n, \theta_1)$ distribution is nothing but the $\text{Bin}(n, \theta_1)$ distribution.

Theorem 8.1

Let $\underline{X} = (X_1, X_2, \dots, X_p) \sim \text{Mult}(n, \theta_1, \dots, \theta_p)$, where $n \in \{1,2, \dots\}, \theta_i \in (0,1), i = 1, \dots, p$ and $\sum_{i=1}^{p} \theta_i < 1$. Then

(i) $X_i \sim \text{Bin}(n, \theta_i), i = 1, \ldots, p$;
(ii) $X_i + X_j \sim \text{Bin}(n, \theta_i + \theta_j), i, j = 1, \ldots, p, i \neq j$;
(iii) $E(X_i) = n\theta_i$ and $\text{Var}(X_i) = n\theta_i(1 - \theta_i),\ i = 1, \ldots, p$;
(iv) $\text{Cov}(X_i, X_j) = -n\theta_i\theta_j,\ i, j = 1, \ldots, p, i \neq j$.

Proof.

(i) Fix $i \in \{1, \ldots, p\}$. In a given trial of the random experiment treat the occurrence of outcome A_i as success and that of any other A_j , $j \neq i$ (i.e., non-occurrence of A_i) as failure. Then we have a sequence of n independent Bernoulli trials with probability of success in each trial as $P(A_i) = \theta_i$. Therefore

$$X_i = \#\text{ of success in } n \text{ independent Bernoulli trials} \sim \text{Bin}(n, \theta_i).$$

(ii) Fix $i, j \in \{1, \ldots, p\}, i \neq j$. In a given trial of the random experiment treat the occurrence of A_i or A_j $\left(\text{i.e., occurrence of } A_i \cup A_j\right)$ as success and its non-occurrence as failure. Then we have a sequence of n independent Bernoulli trials with probability of success in each trial as $P(A_i \cup A_j) = P(A_i) + P(A_j) = \theta_i + \theta_j$ and, therefore,

$$X_i + X_j = \#\text{ of successes in } n \text{ independent Bernoulli trials} \sim \text{Bin}(n, \theta_i + \theta_j).$$

(iii) Follows from (i) on using properties of binomial distribution.

(iv) Fix $i, j \in \{1, \ldots, p\}, i \neq j$. Then

$$X_i + X_j \sim \text{Bin}(n, \theta_i + \theta_j)$$

$$\Longrightarrow \text{Var}(X_i + X_j) = n(\theta_i + \theta_j)(1 - \theta_i - \theta_j)$$

$$\Longrightarrow \text{Var}(X_i) + \text{Var}(X_j) + 2\,\text{Cov}(X_i, X_j) = n(\theta_i + \theta_j)(1 - \theta_i - \theta_j)$$

$$\Longrightarrow n\theta_i(1 - \theta_i) + n\theta_j(1 - \theta_j) + 2\,\text{Cov}(X_i, X_j) = n(\theta_i + \theta_j)(1 - \theta_i - \theta_j)$$

$$\Longrightarrow \text{Cov}(X_i, X_j) = -n\theta_i\theta_j\ , i \neq j. \ \blacksquare$$

The joint m.g.f. of $\underline{X} = (X_1, X_2, \ldots, X_p) \sim \text{Mult}(n, \theta_1, \ldots, \theta_p)$ is given by

$$M_{\underline{X}}(t) = \sum_{\substack{x_1=0 \\ x_1+\cdots}}^{n} \cdots \sum_{\substack{x_p=0 \\ +x_p \leq n}}^{n} e^{t_1x_1+\cdots+t_px_p} \frac{n!}{x_1!\, x_2! \cdots x_p!\, (n - \sum_{i=1}^{p} x_i)!} \theta_1^{x_1} \cdots \theta_p^{x_p} \left(1 - \sum_{i=1}^{p} \theta_i\right)^{n - \sum_{i=1}^{p} x_i}$$

$$= \sum_{\substack{x_1=0 \\ x_1+\cdots}}^{n} \cdots \sum_{\substack{x_p=0 \\ +x_p\leq n}}^{n} \frac{n!}{x_1!\,x_2!\cdots.x_p!\,(n-\sum_{i=1}^{p} x_i)!}(\theta_1 e^{t_1})^{x_1}\cdots(\theta_p e^{t_p})^{x_p}\left(1-\sum_{\mathrm{i=1}}^{\mathrm{p}}\theta_i\right)^{n-\sum_{i=1}^{p} x_1}$$

$$= \left(\theta_1 e^{t_1}+\cdots+\theta_2 e^{t_2}+1-\sum_{i=1}^{p}\theta_i\right)^n, \qquad \underline{t}\in\mathbb{R}^p.$$

Therefore,

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

$$= \left[n\theta_i e^{t_i}\left(\theta_1 e^{t_1}+\cdots+\theta_2 e^{t_p}+1-\sum_{i=1}^{p}\theta_i\right)^{n-1}\right]_{\underline{t}=\underline{0}}$$

$$= n\theta_i\,, \;\; i=1,\dots,p.$$

$$E(X_iX_j) = \left[\frac{\partial^2}{\partial t_i \partial t_j} M_{\underline{X}}(\underline{t})\right]_{\underline{t}=\underline{0}}$$

$$= \left[n(n-1)\theta_i\theta_j e^{t_i+t_j}\left(\theta_1 e^{t_1}+\cdots+\theta_p e^{t_p}+1-\sum_{i=1}^{p}\theta_i\right)^{n-2}\right]_{\underline{t}=\underline{0}}$$

$$= n(n-1)\theta_i\theta_j,\;\; i,j\in\{1,\dots,p\},\;\; i\neq j.$$

$$\mathrm{Cov}(X_i,X_j) = E(X_iX_j)-E(X_i)E(X_j) = -n\theta_i\theta_j,\; i\neq j.$$

$$E(X_i^2) = \left[\frac{\partial^2}{\partial t_i^2} M_{\underline{X}}(\underline{t})\right]_{\underline{t}=\underline{0}}$$

$$= \Bigg[n(n-1)\theta_i^2 e^{2t_i}\left(\theta_1 e^{t_1}+\cdots+\theta_p e^{t_p}+1-\sum_{i=1}^{p}\theta_i\right)^{n-2}$$

$$+\, n\theta_i e^{t_i}\left(\theta_1 e^{t_1}+\cdots+\theta_p e^{t_p}+1-\sum_{i=1}^{p}\theta_i\right)^{n-1}\Bigg]_{\underline{t}=\underline{0}}$$

$$= n(n-1)\theta_i^2+n\theta_i,\;\; i=1,\dots,p. \blacksquare$$

6.9 BIVARIATE NORMAL DISTRIBUTION

Definition 9.1

A bivariate random vector $\underline{X} = (X_1, X_2)$ is said to have a *bivariate normal distribution* $N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ if, for some $-\infty < \mu_i < \infty, i = 1, 2,$ $\sigma_i > 0, i = 1, 2,$ and $-1 < \rho < 1$, the joint p.d.f. of $\underline{X} = (X_1, X_2)$ is given by

$$f_{X_1,X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]}, \ \underline{x} = (x_1, x_2) \in \mathbb{R}^2. \ \blacksquare$$

Note that $f_{X_1,X_2}(\underline{x}) \geq 0,\ \forall \underline{x} \in \mathbb{R}^2$ and on making the transformation $z_1 = \frac{x_1-\mu_1}{\sigma_1}$ and $z_2 = \frac{x_2-\mu_2}{\sigma_2}$ in the integral below, we have

$$I = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2) d\underline{x}$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho^2)}\left(z_1^2 - 2\rho z_1 z_2 + z_2^2\right)} d\underline{z}$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho^2)}\left(z_2^2 - \rho^2 z_2^2\right)} \underbrace{\left\{\int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho^2)}\left(z_1 - \rho z_2\right)^2} dz_1\right\}}_{=\sqrt{1-\rho^2}\sqrt{2\pi}} dz_2$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{{z_2}^2}{2}} dz_2$$

$$= 1$$

Therefore $f_{X_1,X_2}(x_1, x_2)$ is a p.d.f..

Theorem 9.1

Suppose that $\underline{X} = (X_1, X_2) \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho), -\infty < \mu_i < \infty, i = 1, 2, \sigma_i > 0, i = 1, 2$ and $-1 < \rho < 1$. Then,

(i) $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$;

(ii) for a fixed $x_2 \in \mathbb{R}$, the conditional distribution of X_1 given that $X_2 = x_2$ is $N\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), \sigma_1^2(1-\rho^2)\right)$ (written as $X_1|X_2 = x_2 \sim N\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), \sigma_1^2(1-\rho^2)\right)$;

(iii) for a given $x_1 \in \mathbb{R}$, the conditional distribution of X_2 given $X_1 = x_1$ is $N\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1), \sigma_2^2(1-\rho^2)\right)$(written as $X_2|X_1 = x_1 \sim N\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1), \sigma_2^2(1-\rho^2)\right)$;

(iv) the m.g.f. of $\underline{X} = (X_1, X_2)$ is

$$M_{X_1,X_2}(t_1,t_2) = e^{\mu_1 t_1 + \mu_2 t_2 + \frac{\sigma_1^2 t_1^2}{2} + \frac{\sigma_2^2 t_2^2}{2} + \rho\sigma_1\sigma_2 t_1 t_2}, \ \ \underline{t} = (t_1, t_2) \in \mathbb{R}^2;$$

(v) for real constants c_1 and c_2 such that $c_1^2 + c_2^2 > 0$

$$c_1X_1 + c_2X_2 \sim N(c_1\mu_1 + c_2\mu_2, c_1^2\sigma_1^2 + c_2^2\sigma_2^2 + 2\,\rho c_1 c_2\sigma_1\sigma_2);$$

(vi) $\rho(X_1, X_2) = \rho$;

(vii) X_1 and X_2 are independent if, and only if, $\rho = 0$.

Proof.

(i) For $x_1 \in \mathbb{R}$

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1,x_2)dx_2$$

$$= \frac{e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}}}{2\pi\,\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho^2)}\left[\frac{x_2-\mu_2}{\sigma_2} - \rho\frac{x_1-\mu_1}{\sigma_1}\right]^2} dx_2$$

$$= \frac{e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}}}{2\pi\,\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma_2^2(1-\rho^2)}\left[x_2 - \left(\mu_2 + \frac{\rho\sigma_2}{\sigma_1}(x_1-\mu_1)\right)\right]^2} dx_2$$

$$= \frac{e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}}}{2\pi\,\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}$$

$$= \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{(x_1-\mu_1)^2}{2{\sigma_1}^2}},$$

which is the p.d.f. of $N(\mu_1, \sigma_1^2)$ distribution. Thus $X_1 \sim N(\mu_1, \sigma_1^2)$. By symmetry $X_2 \sim N(\mu_2, \sigma_2^2)$.

(ii) Fix $x_2 \in \mathbb{R}$. Then

$$\begin{aligned} f_{X_1|X_2}(x_1|x_2) &= c_1(x_2) f_{X_1,X_2}(x_1, x_2) \\ &= c_2(x_2)\, e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1} - \rho\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right)^2\right]} \\ &= c_2(x_2) e^{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left(x_1 - \left(\mu_1 + \frac{\rho\sigma_1}{\sigma_2}(x_2-\mu_2)\right)\right)^2}, \quad x_1 \in \mathbb{R}, \end{aligned}$$

where $c_2(x_2)$ is the normalizing constant, i.e., $c_2(x_2)$ satisfies

$$\int_{-\infty}^{\infty} f_{X_1|X_2}(x_1|x_2)\, dx_1 = 1.$$

Clearly, for a fixed $x_2 \in \mathbb{R}$, $f_{X_1|X_2}(\cdot\,|x_2)$ is the p.d.f. of $N\left(\mu_1 + \frac{\rho\sigma_1}{\sigma_2}(x_2 - \mu_2), \sigma_1^2(1-\rho^2)\right)$ distribution.

(iii) Follows from (ii) on using symmetry.

(iv) For $\underline{t} = (t_1, t_2) \in \mathbb{R}^2$, using Theorem 5.5, we have

$$\begin{aligned} M_{X_1,X_2}(t_1, t_2) &= E(e^{t_1X_1 + t_2X_2}) \\ &= E\left(E(e^{t_1X_1+t_2X_2}|X_2)\right) \\ &= E\left(E^{t_2X_2} E(e^{t_1X_1}|X_2)\right). \end{aligned}$$

For a fixed $x_2 \in \mathbb{R}$, since $X_1|X_2 = x_2 \sim N\left(\mu_1 + \frac{\rho\sigma_1}{\sigma_2}(x_2-\mu_2), \sigma_1^2(1-\rho^2)\right)$, on using Theorem 4.2 (i), Module 5, we get

$$E(e^{t_1X_1}|X_2 = x_2) = e^{\left\{\mu_1 + \frac{\rho\sigma_1}{\sigma_2}(x_2-\mu_2)\right\}t_1 + \frac{\sigma_1^2(1-\rho^2)t_1^2}{2}}, \quad t_1 \in \mathbb{R}.$$

Therefore, for $\underline{t} = (t_1, t_2) \in \mathbb{R}^2$,

$$M_{X_1,X_2}(t_1, t_2) = E\left(e^{t_2X_2} e^{\left\{\mu_1 + \frac{\rho\sigma_1}{\sigma_2}(X_2-\mu_2)\right\}t_1 + \frac{\sigma_1^2(1-\rho^2)t_1^2}{2}}\right)$$

$$= e^{\mu_1 t_1 + \frac{\sigma_1^2(1-\rho^2)t_1^2}{2} - \frac{\rho\sigma_1}{\sigma_2}\mu_2 t_1} E\left(e^{\left(t_2 + \frac{\rho\sigma_1}{\sigma_2}t_1\right)X_2}\right).$$

Since $X_2 \sim N(\mu_2, \sigma_2^2)$, on using Theorem 4.2 (i), Module 5, we get

$$M_{X_1,X_2}(t_1,t_2) = e^{\mu_1 t_1 + \frac{\sigma_1^2(1-\rho^2)t_1^2}{2} - \frac{\rho\sigma_1}{\sigma_2}\mu_2 t_1} e^{\left(t_2 + \frac{\rho\sigma_1}{\sigma_2}t_1\right)\mu_2 + \frac{\sigma_2^2\left(t_2 + \frac{\rho\sigma_1}{\sigma_2}t_1\right)^2}{2}}$$

$$= e^{\mu_1 t_1 + \mu_2 t_1 + \frac{\sigma_1^2 t_1^2}{2} + \frac{\sigma_2^2 t_2^2}{2} + \rho\sigma_1\sigma_2 t_1 t_2}, \quad \underline{t} = (t_1,t_2) \in \mathbb{R}^2.$$

(v) Let c_1 and c_2 be real constants such that $c_1^2 + c_2^2 > 0$ and let $Y = c_1X_1 + c_2X_2$. Then, for $t \in \mathbb{R}$,

$$M_Y(t) = E(e^{tY})$$

$$= E(e^{tc_1X_1 + tc_2X_2})$$

$$= M_{X_1,X_2}(tc_1, tc_2)$$

$$= e^{(c_1\mu_1 + c_2\mu_2)t + \frac{\left(c_1^2\sigma_1^2 + c_2^2\sigma_2^2 + 2\rho c_1c_2\sigma_1\sigma_2\right)t^2}{2}},$$

which is the m.g.f. of $N\,(c_1\mu_1 + c_2\mu_2, c_1^2\sigma_1^2 + c_2^2\sigma_2^2 + 2\rho c_1c_2\sigma_1\sigma_2)$ distribution. Thus, by Theorem 7.3,

$$Y \sim N(c_1\mu_1 + c_2\mu_2, c_1^2\sigma_1^2 + c_2^2\sigma_2^2 + 2\rho c_1c_2\sigma_1\sigma_2).$$

(vi) By (i), $\mathrm{Var}(X_1) = \sigma_1^2$ and $\mathrm{Var}\,(X_2) = \sigma_2^2$. Also, for $\underline{t} = (t_1,t_2) \in \mathbb{R}^2$,

$$\psi_{X_1,X_2}(t_1,t_2) = \ln M_{X_1,X_2}\,(t_1,t_2) = \mu_1 t_1 + \mu_2 t_2 + \frac{\sigma_1^2 t_1^2}{2} + \frac{\sigma_2^2 t_2^2}{2} + \rho\sigma_1\sigma_2 t_1 t_2$$

$$\mathrm{Cov}(X_1,X_2) = \left[\frac{\partial^2}{\partial t_1 \partial t_2}\psi_{X_1,X_2}(t_1,t_2)\right]_{\underline{t}=\underline{0}} = \rho\sigma_1\sigma_2$$

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

(vii) Since independent random variables are uncorrelated it follows from (vi) that if X_1 and X_2 are independent then $\rho = 0$. Conversely suppose that $\rho = 0$. Then, for $\underline{x} = (x_1,x_2) \in \mathbb{R}^2$,

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]}$$

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

Now the assertion follows on using Theorem 4.2 (i). ■

Theorem 9.2

Let $\underline{X} = (X_1, X_2)$ be a bivariate random vector with $E(X_i) = \mu_i \in (-\infty, \infty), \mathrm{Var}(X_i) = \sigma_i^2, i = 1, 2$ and $\mathrm{Cov}(X_1, X_2) = \rho \in (-1, 1)$. Then $\underline{X} \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ if, and only if, for any real constants t_1 and t_2 such that $t_1^2 + \mathrm{t}_2^2 > 0$, $Y = t_1 X_1 + t_2 X_2 \sim N(t_1 \mu_1 + t_2\, \mu_2, t_1^2 \sigma_1^2 + t_2^2 \sigma_2^2 + \ 2\rho t_1 t_2 \sigma_1 \sigma_2)$.

Proof. Clearly the necessary part of the assertion follows from Theorem 9.1(v). Conversely suppose that for all real constants t_1 and t_2 with $t_1^2 + \mathrm{t}_2^2 > 0$,

$$Y = t_1 X_1 + t_2 X_2 \sim N(t_1 \mu_1 + t_2 \mu_2, t_1^2 \sigma_1^2 + \mathrm{t}_2^2 \sigma_2^2 + 2\rho t_1 t_2 \sigma_1 \sigma_2). \qquad (9.1)$$

Then, for $\underline{t} = (t_1, t_2) \in \mathbb{R}^2$,

$$\begin{aligned} M_{X_1, X_2}(t_1, t_2) &= E(e^{t_1 X_1 + t_2 X_2}) \\ &= E(e^Y) \\ &= M_Y(1) \\ &= e^{t_1 \mu_1 + t_2 \mu_2 + \frac{t_1^2 \sigma_1^2 + t_2^2 \sigma_2^2 + 2\rho\, t_1 t_2 \sigma_1 \sigma_2}{2}}, \qquad \text{(using (9.11))} \end{aligned}$$

which is the m.g.f. of $N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ distribution. Now using Theorem 7.3 it follows that $\underline{X} = (X_1, X_2) \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. ■