Probability

Jan-April 2014

Chennai Mathematical Institute

Professor Nandini Kannan

Chapter 6

Multivariate Distributions

6.1 Introduction

In the previous chapters, we have discussed models for a single random variable. We need to expand our univariate results to higher dimensions.

Motivation:

1. Sampling: In any experiment, data is collected on a large group of items/individuals.
2. Multivariate Data: several measurements are taken on the units.

We need models that describe the behaviour of more than one random variable at a time.

Definition: An $n-$dimensional random vector $\boldsymbol{X}$ is a function from the sample space $\mathcal{S}$ into $\mathcal{R}^n$, $n-$dimensional Euclidean space.

Definition: A vector each of whose components is a random variable is called a **random vector**.

If we let $n = 2$, then every outcome in the sample space is associated with an ordered pair $(x, y) \in \mathcal{R}^2$. This defines a two-dimensional or bivariate random vector (X, Y).

Definition: In 2-dimensional Euclidean space $\mathcal{R}^2$ or the plane, the Borel field $\mathcal{B}^2$ is generated by rectangles of the form

$$\{(x, y); a < x \leq b, c < y \leq d\}.$$

It is also generated by product sets of the form

$$B_1 \times B_2 = \{(x, y); x \in B_1, y \in B_2\}$$

where $B_1, B_2 \in \mathcal{B}$.

Definition: Let X and Y be random variables on $(\mathcal{S}, \mathcal{F}, P)$. The random vector (X, Y) induces a probability measure ν on $\mathcal{B}^2$ as follows:

$$\forall \;\; A \in \mathcal{B}^2, \qquad \begin{aligned} \nu(A) &= P\{(X,Y) \in A\} \\ &= P\left(\{s; (X(s), Y(s)) \in A\}\right). \end{aligned}$$

Here ν is the **probability distribution** of (X,Y).

Definition: The joint cumulative distribution function of (X,Y) is the function $F(x,y)$ defined by

$$F(x,y) = P[X \leq x, Y \leq y] \tag{6.1}$$

for all $(x,y) \in \mathcal{R}^2$.

Definition: A two-dimensional or bivariate random vector (X,Y) is **discrete** if it assumes a countable number of values. The function $p(x,y) : \mathcal{R}^2 \to [0,1]$ defined by

$$p(x,y) = P(X = x \text{ and } Y = y) \tag{6.2}$$

is called the joint pmf of (X,Y).

For any set $A \in \mathcal{B}^2$, we have

$$P[(X,Y) \in A] = \sum_{(x,y)\in A} p(x,y).$$

Theorem 6.1.1. *A function $p(x,y)$ is a joint pmf iff*

(a) $p(x,y) \geq 0 \qquad \forall (x,y)$*;*

(b) $\sum_x \sum_y p(x,y) = 1$.

■

The joint pmf completely defines the probability distribution of the random vector (X,Y).

Once we have specified the joint probability model, we may want to find probabilities involving one of the random variables, say X. The pmf of X is called its **marginal distribution**.

Theorem 6.1.2. *Let (X,Y) be a discrete bivariate random vector with joint pmf $p(x,y)$. Then the marginal pmfs of X and Y, denoted by $p_X(.)$ and $p_Y(.)$ respectively, are given by*

$$p_X(x) = \sum_{y\in\mathcal{R}} p(x,y)$$

$$p_Y(y) = \sum_{x\in\mathcal{R}} p(x,y).$$

Proof:

$$
\begin{aligned}
p_X(x) &= P(X = x) = P(X = x, -\infty < Y < \infty) \\
&= \sum_{y \in \mathcal{R}} p(x, y).
\end{aligned}
$$

■

Remarks:

1. The joint pmf completely specifies the marginals.
2. The converse is not true.

Example: Consider the experiment of tossing two tetrahedra (4-sided dice). Let X be the number observed on the lower face of the first tetrahedron. Let Y be the larger of the two lower faces. Find the joint pmf and the marginals.
We have the joint pmf:

$$
\begin{aligned}
&p(1,1) = p(1,2) = p(1,3) = p(1,4) = p(2,3) = p(2,4) = p(3,4) = \frac{1}{16} \\
&p(2,2) = \frac{2}{16}; p(3,3) = \frac{3}{16}; p(4,4) = \frac{4}{16} \\
&p(2,1) = p(3,1) = p(3,2) = p(4,1) = p(4,2) = p(4,3) = 0.
\end{aligned}
$$

From the joint pmf, we have the two marginal pmfs

$$
\begin{aligned}
&p_X(1) = p_X(2) = p_X(3) = p_X(4) = \frac{4}{16} \\
&p_Y(1) = \frac{1}{16}; p_Y(2) = \frac{3}{16}; p_Y(1) = \frac{5}{16}; p_Y(1) = \frac{7}{16}.
\end{aligned}
$$

Suppose we were given the marginals. Is the joint pmf uniquely determined?
Consider the joint pmf $p^*(x, y)$ with

$$
p^*(1,3) = p^*(2,4) = \frac{1}{16} - \epsilon; p^*(1,4) = p^*(2,3) = \frac{1}{16} + \epsilon
$$

$(\epsilon > 0)$, and

$$
p^*(x, y) = p(x, y)
$$

for all the other pairs. Clearly. $p^*(.,.)$ is a valid joint pmf.
However, the two pmfs generate the same marginals.
■

When X and Y are continuous random variables, we may define the joint probability density function $f(x, y)$.

Definition: A function $f(x,y) : \mathcal{R}^2 \to \mathcal{R}$ is called a **joint probability density function** of the continuous bivariate random vector (X, Y) if

$$P[(X,Y) \in A] = \iint_A f(x,y)dxdy, \tag{6.3}$$

for any set $A \in \mathcal{B}^2$.

Theorem 6.1.3. *A function $f(x,y)$ is a joint pdf iff*

(a) $f(x,y) \geq 0 \qquad \forall (x,y) \in \mathcal{R}^2$;

(b) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1.$

■

Example: Assume that for a certain type of washer, both the thickness and the hole diameter vary from item to item. Let X denote the thickness, and let Y denote the diameter. Assume that the joint pdf of X and Y is given by

$$f(x,y) = \begin{cases} \frac{1}{6}(x+y) & 1 \leq x \leq 2, 4 \leq y \leq 5 \\ o & o.w. \end{cases}$$

Find the probability that a randomly chosen washer has a thickness between 1 and 1.5 mm and a diameter between 4.5 and 5 mm.

Definition: Let (X, Y) be a continuous bivariate random vector with joint pdf $f(x,y)$. Then the marginal pdfs of X and Y, denoted by $f_X(.)$ and $f_Y(.)$ respectively, are given by

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

Definition: Let (X, Y) be a continuous bivariate random vector with joint pdf $f(x,y)$. Then the joint cdf is given by

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s,t) ds dt.$$

We have

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} \tag{6.4}$$

for all (x,y) that are continuity points of $f(x,y)$.

Example: A bank operates both a drive-up facility and a walk-up window. On a randomly selected day, let X represent the proportion of time the drive-up facility is in use, and Y represent the proportion of time the walk-up window is in use. The joint pdf is given by

$$f(x,y) = \begin{cases} \frac{6}{5}(x+y^2), & 0 < x < 1, 0 < y < 1; \\ 0, & \text{ow.} \end{cases}$$

Find the probability that neither facility is busy more than one-quarter of the time. We need to find

$$P\left[0 \le X \le \frac{1}{4}, 0 \le Y \le \frac{1}{4}\right] =$$
$$\int_0^{.25} \int_0^{.25} \frac{6}{5}(x+y^2) dx dy = \frac{7}{640}.$$

Find the marginal pdf's of X and Y.
■

Example: ABC cans of deluxe mixed nuts contain almonds, cashewnuts, and peanuts. The net weight of each can is exactly 1 kilo, but the weight contribution of each type of nut is random. Let X be the weight of almonds, and Y the weight of cashews. The joint pdf is given by

$$f(x,y) = 24xy \qquad 0 < x < 1, 0 < y < 1, 0 < x + y < 1.$$

For a fixed x, $f(x,y)$ increases with y. The pdf should be large near the upper boundary and small near the origin.
Compute the probability that the two types of nuts together make up at least 50% of the can.
Find the marginal pdf's.

6.2 Expectation

The concept of expectation extends to random vectors in a straightforward manner. Let (X,Y) be a random vector, and $U = g(X,Y)$ be some real valued function. Then U is a random variable and we can compute its expectation by finding its distribution. We can also find the expectation by using the joint distribution of (X,Y).

<u>Definition:</u> Let (X,Y) be a bivariate random vector. The expected value of the random variable $U = g(X,Y)$ is given by

$$E[g(X,Y)] = \begin{cases} \sum_x \sum_y g(x,y)p(x,y), & (X,Y) \text{ discrete;} \\ \int\int g(x,y)f(x,y)dxdy, & (X,Y) \text{ continuous} \end{cases} \tag{6.5}$$

provided the sum or integral exist. If $E|g(X,Y)| = \infty$, we say that $E(g(X,Y))$ **does not exist**. All the properties of expectation hold.

If $g(X,Y) = X$, then

$$E[g(X,Y)] = E(X) = \int\int xf(x,y)dxdy = \int xf_X(x)dx.$$

<u>Definition:</u> Let (X,Y) be a bivariate random vector. The **joint mgf** is given by

$$M_{X,Y}(t_1,t_2) = E[e^{t_1X+t_2Y}], \tag{6.6}$$

provided the expectation exists in a nonempty rectangle containing $(0,0)$.
Clearly

$$M_{X,Y}(t,0) = M_X(t); \qquad M_{X,Y}(0,t) = M_Y(t).$$

The partial derivatives of the joint mgf yield joint moments of the random vector (X,Y).
The $r-$th moment of $X(Y)$ may be obtained from $M_{X,Y}(t_1,t_2)$ by differentiating the function r times with respect to $t_1(t_2)$ and then evaluating the derivative at $t_1 = t_2 = 0$. Product moments such as $E[X^rY^s]$ can be obtained by differentiating the joint mgf r times with respect to t_1 and s times with respect to t_2 and then evaluating the derivative at $t_1 = t_2 = 0$.

<u>Definition:</u> The expected value of a random vector $\boldsymbol{X}$ is the vector $E(\boldsymbol{X})$, whose $i-th$ component is the expected value of the $i-th$ component of $\boldsymbol{X}$. The expected value of $\boldsymbol{X}$ exists if the expected values of the components exist.

6.3 Conditional Distributions

<u>Definition:</u> Let (X,Y) be a discrete bivariate random vector with joint pmf $p(x,y)$, and marginal pmfs $p_X(x)$ and $p_Y(y)$. Let x be a point in the support of X, i.e. $p_X(x) > 0$. The **conditional**

pmf of Y given that $X = x$ is the function of y denoted by $p_{Y|X}(y|x)$ defined by

$$p_{Y|X}(y|x) = P(Y = y|X = x) = \frac{p(x,y)}{p_X(x)}. \tag{6.7}$$

Let y be a point in the support of Y, i.e. $p_Y(y) > 0$. The **conditional pmf** of X given that $Y = y$ is the function of x denoted by $p_{X|Y}(x|y)$ defined by

$$p_{X|Y}(x|y) = P(X = x|Y = y) = \frac{p(x,y)}{p_Y(y)}. \tag{6.8}$$

Clearly

$$p_{Y|X}(y|x) > 0; \qquad p_{X|Y}(x|y) > 0.$$

We also have

$$\sum_y p_{Y|X}(y|x) = \sum_y \frac{p(x,y)}{p_X(x)} = \frac{p_X(x)}{p_X(x)} = 1.$$

So the functions defined in equations 6.7 and 6.8 are valid pmf's.

Example: Consider the experiment of tossing two tetrahedra (4-sided dice). Let X be the number observed on the lower face of the first tetrahedron. Let Y be the larger of the two lower faces. Find the conditional pmfs.

When X and Y are continuous, we have $P(X = 0) = 0$ for all x. So the definition given above needs to be modified.

Definition: Let (X, Y) be a continuous bivariate random vector with joint pdf $f(x, y)$, and marginal pdfs $f_X(x)$ and $f_Y(y)$. For any x such that $f_X(x) > 0$, the **conditional pdf** of Y given that $X = x$ is the function of y denoted by $f_{Y|X}(y|x)$ defined by

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}. \tag{6.9}$$

For any y such that $f_Y(y) > 0$, the **conditional pdf** of X given that $Y = y$ is the function of x denoted by $f_{X|Y}(x|y)$ defined by

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}. \tag{6.10}$$

Clearly

$$f_{Y|X}(y|x) > 0; \qquad f_{X|Y}(x|y) > 0.$$

We also have

$$\int_{-\infty}^{\infty} f_{Y|X}(y|x)dy = \int_{-\infty}^{\infty} \frac{f(x,y)}{f_X(x)}dy = 1.$$

So the function defined in equations 6.9 and 6.10 are valid pdf's.

<u>**Example:**</u> X and Y are continuous random variables with joint pdf given by

$$f(x) = \begin{cases} k(x+y^2) & 0 \le x \le 1, 0 \le y \le 1 \\ o & o.w. \end{cases}$$

(a) Find k.

(b) Find the marginal densities of X and Y.

(c) Find the conditional density of X given $Y = y$.

6.3.1 Conditional Expectation

<u>**Definition:**</u> The **conditional expectation** of $g(Y)$ given that $X = x$ is

$$E[g(Y)|x] = \begin{cases} \sum_y g(y)p_{Y|X}(y|x), & (X,Y) \text{ discrete;} \\ \int g(y)f_{Y|X}(y|x)dy, & (X,Y) \text{ continuous} \end{cases} \tag{6.11}$$

$E(Y|x)$ is the expected value of the conditional distribution $f(y|x)$. The variance of the conditional distribution $f(y|x)$ is called the **conditional variance** of Y given $X = x$.

We have

$$Var(Y|x) = E(Y^2|x) - [E(Y|x)]^2.$$

Theorem 6.3.1.

$$E(Y) = E[E(Y|X)]. \tag{6.12}$$

Proof: $E(Y|X)$ is a random variable whose value depends on the value of X. If $X = x$, the value of the random variable $E(Y|X)$ is $E(Y|x)$. We will illustrate the proof for the case of continuous random variables. The proof for the discrete case is similar with integrals replaced by summations. We have

$$\begin{aligned} E(Y) &= \int\int yf(x,y)dxdy \\ &= \int\left[\int yf(y|x)dy\right] f_X(x)dx \\ &= \int E(Y|x)f_X(x)dx \\ &= E[E(Y|X)]. \end{aligned}$$

■

Theorem 6.3.2.

$$Var(Y) = E[Var(Y|X)] + Var[E(Y|X)]. \tag{6.13}$$

Proof: Consider the RHS of (6.13).

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

■

Example: A certain mouse is placed in the center of a maze, surrounded by three paths that open with varying widths. The first path returns him to the center after two minutes; the second path returns him to the center after four minutes; and the third path leads him out of the maze after one minute. Due to the differing widths, the mouse chooses the first path 50% of the time, the second path 30% of the time, and the third path 20% of the time. Determine the expected number of minutes it will take for the mouse to escape.

Solution:
We have

$$E(T|P=1) = 2 + E(T); \qquad E(T|P=2) = 4 + E(T); \qquad E(T|P=3) = 1.$$

Therefore

$$E(T) = 0.5 \times [2 + E(T)] + 0.3 \times [4 + E(T)] + .2 \times 1 \Rightarrow E(T) = 12.$$

■

Example: Let

$$f(x,y) = \begin{cases} 2, & 0 < x < y < 1; \\ 0, & \text{ow.} \end{cases}$$

Find the marginal and conditional pdfs.

Find $P(Y \geq .5|x = .5)$ and $P(X \geq 1/3|y = 2/3)$.

Find $E(Y|X = x)$.

$$E(Y|x) = \int_x^1 \frac{y}{1-x} dy = \frac{1+x}{2}.$$

■

6.4 Independent Random Variables

Definition: Let (X, Y) be a bivariate random vector with joint pdf $f(x, y)$ (or joint pmf $p(x, y)$) and marginal pdfs (pmfs) $f_X(x)(p_X(x))$ and $f_Y(y)(p_Y(y))$ respectively. The random variables X and Y are said to be **independent** if

$$f(x,y) = f_X(x) f_Y(y) \qquad \text{for all } (x,y) \in \mathcal{R}^2 \tag{6.14}$$

$$(p(x,y) = p_X(x) p_Y(y)). \tag{6.15}$$

If (6.14) does not hold, we say X and Y are **dependent**.
If X and Y are independent, then the conditional pdf of Y given $X = x$ does not depend on the value of x, i.e.

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{f_X(x) f_Y(y)}{f_X(x)} = f_Y(y). \tag{6.16}$$

The knowledge that $X = x$ does not give us any additional information about Y.

Lemma 6.4.1. *Let (X, Y) be a bivariate random vector with joint pdf $f(x, y)$ (or joint pmf $p(x, y)$).*

Then X and Y are independent random variables iff there exist functions $g(x)$ and $h(y)$ such that

$$f(x, y) = g(x)h(y) \qquad \text{for all } x \in \mathcal{R} \text{ and } y \in \mathcal{R}. \tag{6.17}$$

<u>Proof:</u> If X and Y are independent, we may define

$$g(x) = f_X(x) \qquad h(y) = f_Y(y),$$

and use (6.14). To prove the "if" part, suppose that

$$f(x, y) = g(x)h(y).$$

Let

$$\int_{-\infty}^{\infty} g(x)dx = c \qquad \int_{-\infty}^{\infty} h(y)dy = d.$$

Then

$$\begin{aligned} c \times d &= \left(\int_{-\infty}^{\infty} g(x)dx\right)\left(\int_{-\infty}^{\infty} h(y)dy\right) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)dxdy \\ &= 1. \end{aligned}$$

We have

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x, y)dy \\ &= \int_{-\infty}^{\infty} g(x)h(y)dy = g(x)d, \end{aligned}$$

and

$$f_Y(y) = h(y)c.$$

These results imply

$$f(x, y) = g(x)h(y) = g(x)h(y)cd = f_X(x)f_Y(y),$$

showing that X and Y are independent.
Replacing integrals with sums proves the result in the discrete case.
■

Theorem 6.4.2. *Let (X, Y) be a bivariate random vector with joint cdf $F(x, y)$. Let X and Y have the marginal cdfs $F_X(x)$ and $F_Y(y)$ respectively. Then X and Y are independent random variables iff*

$$F(x, y) = F_X(x)F_Y(y) \qquad \text{for all } (x, y) \in \mathcal{R}^2. \tag{6.18}$$

■

Theorem 6.4.3. *Let (X, Y) be a bivariate random vector with joint mgf $M_{X,Y}(t_1, t_2)$. Then X and Y are independent random variables iff*

$$M_{X,Y}(t_1, t_2) = M(t_1, 0)M(0, t_2) = M_X(t_1)M_Y(t_2), \tag{6.19}$$

i.e. the joint mgf factors into the product of the marginal mgfs.

■

Theorem 6.4.4. *Let X and Y be independent random variables. Then*

(a) For any $A \subset \mathcal{R}$ and $B \subset \mathcal{R}$

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

(b)

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)].$$

■

Theorem 6.4.5. *Let X and Y be independent random variables with moment generating functions $M_X(t)$ and $M_Y(t)$ respectively. Let $Z = X + Y$. Then*

$$M_Z(t) = M_X(t)M_Y(t). \tag{6.20}$$

■

<u>Example:</u> Let X and Y be independent Poisson random variables with parameters λ_1 and λ_2 respectively. Find the distribution of $X + Y$.

6.5 Bivariate Transformations

Let (X, Y) be a bivariate random vector with a known joint distribution. Consider the bivariate random vector (U, V) where $U = g_1(X, Y)$ and $V = g_2(X, Y)$. Here $g_1(.,.)$ and $g_2(.,.)$ are specified functions. We want to find the joint distribution of (U, V) in terms of the joint distribution of (X, Y).
For any $B \in \mathcal{R}^2$, we have

$$P[(U, V) \in B] = P[(X, Y) \in A],$$

where $A = \{(x, y) : (g_1(x, y), g_2(x, y)) \in B\}$.

If (X, Y) is a discrete random vector with joint pmf $p(x, y)$, let $\mathcal{A}$ be the set of (x, y) for which $p(x, y) > 0$, i.e the support of (X, Y). Then

$$\mathcal{C} = \{(u, v) : u = g_1(x, y), v = g_2(x, y) \text{ for some } (x, y) \in \mathcal{A}\}$$

is the support of (U, V). For any $(u, v) \in \mathcal{C}$, define

$$A_{uv} = \{(x, y) \in \mathcal{A} : g_1(x, y) = u, g_2(x, y) = v\}.$$

Then

$$p_{U,V}(u, v) = P(U = u, V = v) = \sum_{(x,y) \in A_{uv}} p_{X,Y}(x, y).$$

Example: Let X and Y be independent geometric random variables with pmf's

$$\begin{aligned} p_X(x) &= pq^{x-1} \qquad x = 1, 2, \ldots \\ p_Y(y) &= pq^{y-1} \qquad y = 1, 2, \ldots \end{aligned}$$

respectively.
Let $U = \min(X, Y), V = X - Y$.
For $v > 0$, we have $X > Y$. Therefore

$$\begin{aligned} p_{U,V}(u, v) &= P(U = u, V = v) \\ &= P(Y = u, X = u + v) \\ &= pq^{u-1}pq^{u+v-1} \\ &= p^2 q^{2u+v-2}. \end{aligned}$$

For $v < 0$,

$$\begin{aligned} p_{U,V}(u, v) &= P(U = u, V = v) \\ &= P(X = u, Y = u - v) \\ &= pq^{u-1}pq^{u-v-1} \\ &= p^2 q^{2u-v-2}. \end{aligned}$$

For $v = 0$, we have $X = Y$ and

$$\begin{aligned} p_{U,V}(u,v) &= P(U = u, V = 0) \\ &= P(X = Y = u) \\ &= pq^{u-1}pq^{u-1} \\ &= p^2 q^{2u-2}. \end{aligned}$$

■

<u>**Example:**</u> Let X and Y be independent Poisson random variables with pmf's

$$\begin{aligned} p_X(x) &= \frac{e^{-\alpha}\alpha^x}{x!} \qquad x = 0, 1, 2, \ldots \\ p_Y(y) &= \frac{e^{-\beta}\beta^y}{y!} \qquad y = 0, 1, 2, \ldots \end{aligned}$$

respectively.
Let $U = X + Y, V = Y$. We have

$$\begin{aligned} p_{U,V}(u,v) &= P(U = u, V = v) \\ &= P(X = u - v, Y = v) \\ &= \frac{e^{-\alpha}\alpha^{u-v}}{(u-v)!}\frac{e^{-\beta}\beta^v}{v!} \\ &= \frac{e^{-(\alpha+\beta)}\alpha^{u-v}\beta^v}{(u-v)!v!}, \quad v = 0, 1, \ldots; u = v, v+1, \ldots. \end{aligned}$$

The marginal pmf of U is

$$\begin{aligned} p_U(u) &= \sum_{v=0}^{u} \frac{e^{-(\alpha+\beta)}\alpha^{u-v}\beta^v}{(u-v)!v!} \\ &= \frac{e^{-(\alpha+\beta)}}{u!}\sum_{v=0}^{u}\frac{u!}{(u-v)!v!}\alpha^{u-v}\beta^v \\ &= \frac{e^{-(\alpha+\beta)}}{u!}\sum_{v=0}^{u}\binom{u}{v}\alpha^{u-v}\beta^v \\ &= \frac{e^{-(\alpha+\beta)}(\alpha+\beta)^u}{u!}, \; u = 0, 1, \ldots, \end{aligned}$$

which is the pmf of a $P(\alpha + \beta)$ random variable.
The pmf of U can also be obtained using Theorem 6.4.5.
■

Theorem 6.5.1. *Let $X \sim P(\alpha), Y \sim P(\beta)$ be independent random variables. Then $X + Y \sim P(\alpha + \beta)$.*

■

Theorem 6.5.2. *Let $X \sim f_X(x)$ and $Y = g(X)$, where g is a monotone function. Suppose $f_X(x)$ is continuous on $\mathcal{X}$ and $g^{-1}(y)$ has a continuous derivative on $\mathcal{Y}$. Then*

$$f_Y(y) = \begin{cases} f_X[g^{-1}(y)] \left| \frac{d}{dy} g^{-1}(y) \right|, & y \in \mathcal{Y}; \\ 0, & o.w. \end{cases}$$

■

In many applications, the function g may be neither increasing nor decreasing. However, the function may be monotone over certain intervals. It may be possible to divide $\mathcal{X}$ into sets $A_1, \dots, A_k$ such that $g(.)$ is monotone on each set. We can then modify the previous theorem:

Theorem 6.5.3. *Let $X \sim f_X(x)$ and $Y = g(X)$. Suppose there exists a partition $A_0, A_1, \dots, A_k$ of $\mathcal{X}$ such that $P[X \in A_0] = 0$ and $f_X(x)$ is continuous on each A_i. Suppose there exist functions $g_1(x), \dots, g_k(x)$ defined on $A_1, A_2, \dots, A_k$ respectively satisfying*

(i) $g(x) = g_i(x), x \in A_i$;

(ii) $g_i(x)$ *is monotone on* A_i;

(iii) The set $\mathcal{Y} = \{y : y = g_i(x) \ \text{for some } x \in A_i\}$ *is the same for each* $i = 1, \dots, k$; *and*

(iv) $g_i^{-1}(y)$ *has a continuous derivative on* $\mathcal{Y}$ *for each* $i = 1, 2, \dots, k$.

Then

$$f_Y(y) = \begin{cases} \sum_{i=1}^{k} f_X[g_i^{-1}(y)] \left| \frac{d}{dy} g_i^{-1}(y) \right|, & y \in \mathcal{Y}; \\ 0, & o.w. \end{cases}$$

Here A_0 represents the exceptional set.

■

Let (X, Y) be a bivariate continuous random vector with joint pdf $f_{X,Y}(x, y)$. We can write the joint pdf of (U, V) in terms of $f_{X,Y}(x, y)$. Let $\mathcal{A}$ denote the support of (X, Y) and $\mathcal{C}$ denote the support of (U, V).
Let us assume the transformations $U = g_1(X, Y)$ and $V = g_2(X, Y)$ are one-to-one from $\mathcal{A}$ onto $\mathcal{C}$. Let

$$x = h_1(u, v) \qquad y = h_2(u, v)$$

be the inverse transformations. The Jacobian of the transformation is the determinant of the matrix of partial derivatives:

$$|J| = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}.$$

The joint pdf of (U, V) is given by

$$f_{U,V}(u, v) = f_{X,Y}[h_1(u, v), h_2(u, v)]|J|.$$

If the transformations are not one-to-one, we can partition the space $\mathcal{A}$ as before.

Example: Let X and Y be independent random variables with $X \sim G(\alpha_1, \beta), Y \sim G(\alpha_2, \beta)$. Let

$$U = \frac{X}{X+Y} \qquad V = X + Y.$$

Find the joint pdf of (U, V) and the marginals.
Solution: We have $\mathcal{A}$ as the positive quadrant of $\mathcal{R}^2$. Then $\mathcal{C} = \{(u,v); 0 < u < 1, v > 0\}$. The transformations defined above are one-to-one from $\mathcal{A}$ onto $\mathcal{C}$. We have

$$x = uv \qquad y = v(1-u),$$

and the Jacobian is v. We have

$$|J| = \left| \begin{array}{cc} v & u \\ -v & (1-u) \end{array} \right| = |v(1-u) + vu| = v.$$

We have

$$\begin{aligned} f_{U,V}(u,v) &= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}}(uv)^{\alpha_1-1}[v(1-u)]^{\alpha_2-1}\exp\left[-\frac{uv}{\beta}\right] \\ &\times \exp\left[-\frac{v(1-u)}{\beta}\right]v \\ &= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}}u^{\alpha_1-1}(1-u)^{\alpha_2-1}v^{\alpha_1+\alpha_2-1}e^{-v/\beta}, \end{aligned}$$

for $0 < u < 1, v > 0$. The joint pdf factors as the product of two functions, which implies U and V are independent. We have $U \sim Beta(\alpha_1, \alpha_2)$ and $V \sim G(\alpha_1 + \alpha_2, \beta)$.

■

Example: Let X and Y be independent standard normal random variables. Let

$$U = \frac{X}{Y} \qquad V = |Y|.$$

Find the joint pdf of (U, V) and the marginals.
We have $\mathcal{A} = \mathcal{R}^2$. Therefore $\mathcal{C} = \mathcal{R} \times R^+$.
If $Y = 0$, we can define U and V to be any value since $P(Y = 0) = 0$. The transformation is not one-to-one, since (x, y) and $(-x, -y)$ are both mapped to the same (u, v). We can partition the support $\mathcal{A}$ into three regions

$$\mathcal{A}_1 = \{(x, y), y > 0\},\ \mathcal{A}_2 = \{(x, y), y < 0\},\ \mathcal{A}_0 = \{(x, y), y = 0\}.$$

We have $P[(X, Y) \in \mathcal{A}_0] = 0$. The image of $\mathcal{A}_1$ under the transformation is $\{(u, v), v > 0\}$. The image of $\mathcal{A}_2$ under the transformation is also $\{(u, v), v > 0\}$. The inverse transformations on $\mathcal{A}_1$ are

$$x = uv \qquad y = v$$

and the inverse transformations on $\mathcal{A}_2$ are

$$x = -uv \qquad y = -v.$$

The Jacobians are both v. We have

$$\begin{aligned}
f_{U,V}(u, v) &= \frac{1}{2\pi} \exp\left[-\frac{(uv)^2}{2}\right] \exp\left[-\frac{v^2}{2}\right] |v| \\
&+ \frac{1}{2\pi} \exp\left[-\frac{(-uv)^2}{2}\right] \exp\left[-\frac{(-v)^2}{2}\right] |v| \\
&= \frac{v}{\pi} \exp\left[-\frac{(u^2+1)v^2}{2}\right], \qquad v > 0; u \in \mathcal{R}.
\end{aligned}$$

The marginal pdf of U is

$$\begin{aligned}
f_U(u) &= \int_0^\infty \frac{v}{\pi} \exp\left[-\frac{(u^2+1)v^2}{2}\right] dv \\
&= \frac{1}{\pi} \frac{1}{1+u^2}
\end{aligned}$$

which is the pdf of a Cauchy random variable.

The marginal pdf of V is

$$\begin{aligned} f_V(v) &= \int_{-\infty}^{\infty} \frac{v}{\pi} \exp\left[-\frac{(u^2+1)v^2}{2}\right] du \\ &= \frac{v}{\pi} \exp[-v^2/2] \sqrt{2\pi} \frac{1}{v} \\ &= \sqrt{\frac{2}{\pi}} \exp[-v^2/2], \end{aligned}$$

which is the Folded Normal distribution.
■

Example: Let X and Y be independent $U(0,1)$ random variables. Let $U = X+Y, V = X-Y$. Find the joint pdf of (U,V) and the marginals.
We have $\mathcal{A}$ as the unit square. The space $\mathcal{C}$ is given by the figure below.
The transformation is clearly one-to-one. We have

$$X = \frac{U+V}{2} \qquad Y = \frac{U-V}{2}.$$

The Jacobian is 1/2.
We must have

$$0 < \frac{u+v}{2} < 1 \qquad 0 < \frac{u-v}{2} < 1,$$

which leads to

$$u+v > 0, u-v > 0, u+v < 2, u-v < 2.$$

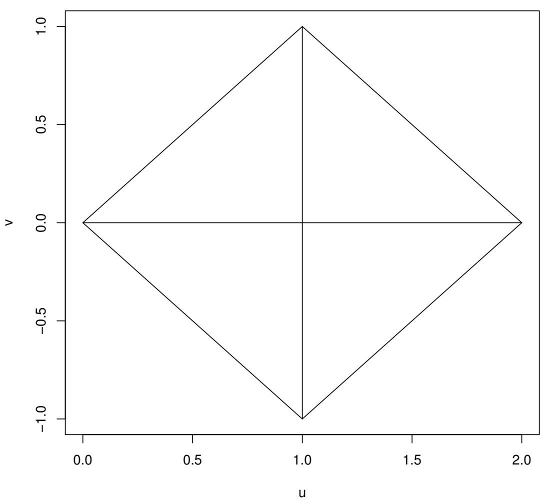
Therefore,

$$f_{U,V}(u,v) = \frac{1}{2}$$

over the region $\mathcal{C}$.
The marginal pdf of U is

$$\begin{aligned} f_U(u) &= \int_{-u}^{u} \frac{1}{2} dv = u \qquad 0 < u < 1, \\ f_U(u) &= \int_{u-2}^{2-u} \frac{1}{2} dv = 2-u \qquad 1 < u < 2. \end{aligned}$$

■



6.6 Covariance and Correlation

In the previous section, we have discussed the notion of independence of random variables. If two random variables are dependent, can we measure the strength of their relationship?
In order to quantify the relationship, we introduce the ideas of covariance and correlation.
Definition: Let X and Y be two random variables with finite variances. The **covariance** of X and Y is defined as

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E(XY) - \mu_X\mu_Y. \tag{6.21}$$

$Cov(X,Y)$ will be positive when $X-\mu_X$ and $Y-\mu_Y$ tend to have the same sign with high probability. $Cov)X,Y)$ will be negative when $X-\mu_X$ and $Y-\mu_Y$ tend to have opposite signs with high probability. Therefore, the sign of the covariance provides information about the relationship between X and Y. However, the magnitude of the covariance does not in itself provide information on the strength of the relationship since it depends on the variability.

Remark: Equation (6.21) defines an inner product on the linear space spanned by X and Y. We have

$$< X, X >= ||X||^2 \qquad Cov(X,X) = Var(X).$$

Definition: Let X and Y be two random variables with finite variances. The **correlation** of X and Y is the number defined by

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X\sigma_Y}. \tag{6.22}$$

The value $\rho(X,Y)$ is called the **correlation coefficient**.

Example: Let

$$f(x,y) = \begin{cases} 2, & 0 < x < y < 1; \\ 0, & \text{ow.} \end{cases}$$

Find $\rho(X,Y)$.

Example: Let

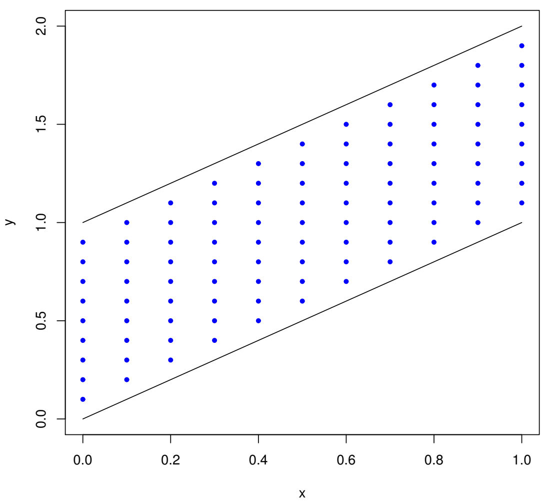
$$f(x,y) = \begin{cases} 1, & 0 < x < 1, x < y < x+1; \\ 0, & \text{ow.} \end{cases}$$

Find $\rho(X,Y)$.

We have $f_X(x) = 1; 0 < x < 1$. This implies $\mu_X = 1/2, \sigma_X^2 = 1/12$. The marginal of Y is

$$f_Y(y) = \begin{cases} y, & 0 < y < 1; \\ 2-y, & 1 \le y < 2. \end{cases}$$

We have $\mu_Y = 1, \sigma_Y^2 = 1/6$. Further $E(XY) = 7/12$. The correlation is $1/\sqrt{2}$.



Theorem 6.6.1. *If X and Y are independent, then $Cov(X,Y) = 0$ and $\rho(X,Y) = 0$.*

Proof: If X and Y are independent,

$$\begin{aligned} Cov(X,Y) &= E(XY) - \mu_X\mu_Y \\ &= E(X)E(Y) - \mu_X\mu_Y \\ &= 0. \end{aligned}$$

If $Cov(X,Y) = 0$, then clearly $\rho(X,Y) = 0$.

■

The converse of the result is not true. If $\rho(X,Y) = 0$, X and Y may still exhibit some form of dependence. Covariance and correlation measure a particular kind of linear relationship between X and Y.

Examples: Consider the random variable X with pmf

$$p(X = -1) = P(X = 0) = P(X = 1) = \frac{1}{3}.$$

Let Y be a random variable defined as follows:

$$Y = \begin{cases} 0, & \text{if } X \neq 0; \\ 1, & \text{if } X = 0. \end{cases}$$

Then $XY = 0$. $E(XY) = 0, E(X) = 0$. This implies

$$Cov(X,Y) = 0.$$

However, X and Y are clearly dependent.

■

Theorem 6.6.2. *If X and Y are two random variables and a, b are constants, then*

$$Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X,Y). \tag{6.23}$$

■

From the theorem we see that if X and Y are positively correlated, the the variation in $X + Y$ is greater than the sum of the variations in X and Y. If the correlation is negative, the variance of $X + Y$ is smaller than the sum of the variances.

Theorem 6.6.3. *For any random variables X and Y,*

(a) $-1 \leq \rho(X,Y) \leq 1$.

(b) $|\rho(X,Y)| = 1$ *if and only if there exist numbers $a \neq 0$ and b such that $P(Y = aX + b) = 1$. If $\rho(X,Y) = 1$, then $a > 0$, and if $\rho(X,Y) = -1$, then $a < 0$.*

Proof: Without loss of generality, let $\mu_X = \mu_Y = 0$. Consider the function $h(t)$ defined by

$$h(t) = E[(tX - Y)^2] = E[t^2X^2 - 2tXY + Y^2] \geq 0.$$

Since the quadratic function $h(t) \geq 0$ for all t, the discriminant must be negative. This implies

$$\begin{aligned}
&4[E(XY)]^2 - 4E(X^2)E(Y^2) \leq 0 \\
&\Leftrightarrow [2Cov(X,Y)]^2 - 4\sigma_X^2\sigma_Y^2 \leq 0 \\
&\Leftrightarrow |\rho(X,Y)| \leq 1.
\end{aligned}$$

This proves part (a) of the theorem.
We also note that $|\rho(X,Y)| = 1$ if and only if the discriminant is equal to 0. This implies the function $h(t)$ has a single root. We see that $h(t) = 0$ if and only if

$$E[(tX - Y)^2] = 0,$$

which implies

$$P[tX - Y = 0] = 1 \Leftrightarrow P[tX = Y] = 1 \Leftrightarrow P[Y = aX] = 1.$$

The root t is $Cov(X,Y)/\sigma_X^2$. Therefore a has the same sign as the correlation coefficient.
■

Remark: If there is a line $y = ax + b$, $a \neq 0$ such that the values of (X,Y) have a high probability of being near this line, then the correlation between X and Y will be near ± 1. Therefore, the correlation measures the strength of the "linear relationship" between X and Y.

Example: Let

$$f(x,y) = \begin{cases} 10, & 0 < x < 1, x < y < x + \frac{1}{10}; \\ 0, & \text{ow.} \end{cases}$$

Find $\rho(X,Y)$.

We have

$$f_X(x) = \int_x^{x+1/10} \frac{1}{10} dy = 1,$$

for $0 < x < 1$, i.e. $X \sim U(0,1)$. This implies $\mu_X = 1/2, Var(X) = 1/12$. The conditional pdf of Y given $X = x$ is

$$f(y|x) = 10 \qquad x < y < x + \frac{1}{10},$$

i.e $Y|x = x \sim U(x, x + 1/10)$. Therefore

$$E(Y|x) = x + \frac{1}{20}.$$

Using Theorem 6.3.1,

$$\begin{aligned}
\mu_Y = E(Y) &= \int_0^1 E(Y|X) f_X(x) dx \\
&= \int_0^1 (x + .05) dx \\
&= \frac{1}{2} + \frac{1}{20} = \frac{11}{20}.
\end{aligned}$$

The conditional variance of Y given $X = x$ is

$$Var(Y|X = x) = \frac{1}{1200},$$

From Theorem 6.3.2, we have

$$\sigma_Y^2 = \frac{1}{1200} + Var\left(X + \frac{1}{20}\right) = \frac{1}{1200} + \frac{1}{12}.$$

We also have

$$\begin{aligned} E(XY) &= \int_0^1 \int_x^{x+1/10} 10xydydx \\ &= \int_0^1 10x \left. \frac{y^2}{2} \right|_x^{x+1/10} \\ &= \int_0^1 \left(x^2 + \frac{x}{20}\right) dx \\ &= \frac{1}{3} + \frac{1}{40} = \frac{43}{120}. \end{aligned}$$

Therefore

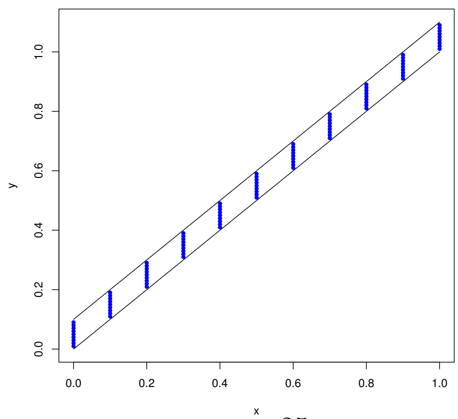
$$Cov(X, Y) = \frac{43}{120} - \frac{1}{2} \times \frac{11}{20} = \frac{1}{12},$$

and

$$\rho(X, Y) = \frac{\frac{1}{12}}{\sqrt{\frac{1}{12}}\sqrt{\frac{1}{1200} + \frac{1}{12}}} = \sqrt{\frac{100}{101}}.$$

Compare this with the previous example. Why are they so different? In both examples, there is a linear relationship between X and Y, However, the relationship is much stronger for this example, because knowing that $X = x$ gives us more information about the value of Y.

■



6.7 Multivariate Distributions

In the previous sections, we have considered bivariate distributions. We will now extend the definitions given earlier to the case of three or more random variables.
Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be an $n-$dimensional random vector. Then $\boldsymbol{X}$ is a function defined from the sample space $\mathcal{S}$ into $\mathcal{R}^n$.
If the sample space of $\boldsymbol{X}$ is countable, then we can define the joint pmf by

$$p(\boldsymbol{x}) = P(X_1 = x_1, \ldots, X_n = x_n)$$

for each $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathcal{R}^n$. For any $\boldsymbol{A} \in \mathcal{B}^n$,

$$P(\boldsymbol{X} \in A) = \sum_{\boldsymbol{x} \in A} p(\boldsymbol{x}.$$

If $\boldsymbol{X}$ is a continuous random vector, then the joint pdf is a function $f(x_1, \ldots, x_n)$ that satisfies

$$P(\boldsymbol{X} \in A) = \int \ldots \int_A f(x_1, \ldots, x_n) dx_1 \ldots dx_n.$$

The joint pmf and pdf satisfy the standard properties: they are non-negative functions and either sum to one or integrate to one.

The joint cumulative distribution function is defined as

$$F(\boldsymbol{x}) = P(X_1 \leq x_1, \ldots, X_n \leq x_n).$$

If $\boldsymbol{X}$ is continuous, then the joint pdf is obtained as

$$f(\boldsymbol{x}) = \frac{\partial^n}{\partial x_1 \ldots \partial x_n} F(\boldsymbol{x}).$$

Let $g(\boldsymbol{x})$ be a real valued function defined on the sample space of $\boldsymbol{X}$. Then $g(\boldsymbol{X})$ is a random variable and

$$E[g(\boldsymbol{X})] = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} g(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x}; \qquad E[g(\boldsymbol{X})] = \sum_{\boldsymbol{x} \in \mathcal{R}^n} p(\boldsymbol{x}),$$

for the continuous and discrete cases, respectively. The properties of expectation carry over to the case of multiple random variables.

We can define the marginal pmf or pdf of any subset of the random vector. We have

$$f(x_1, \ldots x_k) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n) dx_{k+1} \ldots dx_n$$

as the marginal distribution of $(X_1, \ldots, X_k)$. The discrete case can be similarly defined.

The conditional pdf of a subset of random variables given the values of the remaining variables is defined as

$$f(x_{k+1},\ldots,x_n|x_1,\ldots x_k)=\frac{f(x_1,\ldots,x_n)}{f(x_1,\ldots x_k)},$$

provided the denominator is not zero.

6.7.1 The Multinomial Distribution

Consider the following experiment:

1. The experiment consists of n identical trials.
2. Each trial can result in one of k possible outcomes.
3. The probability of the $i-$th outcome is p_i and remains constant from trial to trial.
4. The trials are independent.

We have

$$\sum_{i=1}^{k} p_i = 1.$$

Let X_i be the random variable that records the total number of times outcome i is observed in the n trials. We have

$$\sum_{i=1}^{k} = n.$$

Then

$$P(X_1=x_1,\ldots,X_k=x_k)=\begin{cases}\left(\begin{array}{c} n \\ x_1,\ x_2,\ldots,x_k\end{array}\right)p_1^{x_1}\ldots p_k^{x_k}, & n=\sum x_i; \\ 0, & \text{ow.}\end{cases} \tag{6.24}$$

where

$$\left(\begin{array}{c} n \\ x_1,\ x_2,\ldots,x_k\end{array}\right)=\frac{n!}{x_1!\ldots x_k!}$$

is called a multinomial coefficient.
$(X_1,\ldots,X_k)$ is said to have a **multinomial distribution** with n trials and probabilities $(p_1,\ldots,p_k)$.

To show the probabilities sum to one, we have an extension of the Binomial Theorem:

Theorem 6.7.1. *Multinomial Theorem.*

$$(p_1+\ldots p_k)^n=\sum_A \frac{n!}{x_1!\ldots x_k!}p_1^{x_1}\ldots p_k^{x_k}=1,$$

where A is the set of vectors $(x_1,\ldots,x_n)$ such that each x_i is non-negative and $\sum x_i = n$.

■

Theorem 6.7.2. *The marginal pmf of X_i is $Bin(n, p_i)$, i=1, ..., k.*

Proof:

$$\begin{aligned}
P(X_i = x) &= \sum_{x_2, x_3, \ldots, x_k} \frac{n!}{x! \ldots x_k!} p_1^x \ldots p_k^{x_k} \\
&= \frac{n!}{x!(n-x)!} p_1^x (1-p_1)^{n-x} \\
&\quad \sum_{x_2, x_3, \ldots, x_k} \frac{(n-x)!}{x_2! \ldots x_k!} \left(\frac{p_2}{1-p_1} \right)^{x_2} \ldots \left(\frac{p_k}{1-p_1} \right)^{x_k} \\
&= \binom{n}{x} p_1^x (1-p_1)^{n-x},
\end{aligned}$$

which is the pmf of the Binomial distribution with parameters n and p_1.
Here we use the fact that $x_2 + \ldots + x_k = n - x$ and $p_2 + \ldots + p_k = 1 - p_1$,
■

Theorem 6.7.3. *The joint mgf of $\boldsymbol{X}$ is given by*

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = E[e^{\boldsymbol{t}' \boldsymbol{X}}] = (p_1 e^{t_1} + \ldots + p_k e^{t_k})^n.$$

Here $\boldsymbol{t}' \boldsymbol{X} = t_1 X_1 + \ldots + t_k X_k$.

Proof: We have

$$\begin{aligned}
M_{\boldsymbol{X}}(\boldsymbol{t}) &= E[e^{t_1 X_1 + \ldots + t_k X_k}] \\
&= \sum_{\boldsymbol{x}} e^{\sum t_i x_i} \binom{n}{x_1,\ x_2, \ldots, x_k} p_1^{x_1} \ldots p_k^{x_k} \\
&= \sum_{\boldsymbol{x}} \binom{n}{x_1,\ x_2, \ldots, x_k} (p_1 e^{t_1})^{x_1} \ldots (p_k e^{t_k})^{x_k} \\
&= (p_1 e^{t_1} + \ldots + p_k e^{t_k})^n,
\end{aligned}$$

for all $(t_1, \ldots, t_k) \in \mathcal{R}^k$. The last step follows from the Multinomial Theorem.
■

Remarks:

1. Letting all the $t_j's$ equal 0, except for t_i, we have

$$\begin{aligned}
M_{X_i}(t) &= M(0, \ldots, 0, t, 0, \ldots, 0) \\
&= (p_1 + p_i e^t + \ldots p_k)^n \\
&= (1 - p_i + p_i e^t)^n = (q_i + p_i e^t)^n,
\end{aligned}$$

which is the mgf of a $Bin(n, p_i)$ random variable.

2.

$$\begin{aligned} M_{X_1,X_2}(t_1,t_2) &= M_{\boldsymbol{X}}(t_1,t_2,\ldots,0) \\ &= [p_1e^{t_1}+p_2e^{t_2}+(1-p_1-p_2)]^n, \end{aligned}$$

which is the mgf of the **Trinomial Distribution.**

3. Using the form of the joint mgf, we have

$$\begin{aligned} \frac{\partial}{\partial t_1}M_{\boldsymbol{X}}(\boldsymbol{t}) &= n(p_1e^{t_1}+\ldots+p_ke^{t_k})^{n-1}p_1e^{t_1} \\ \frac{\partial^2}{\partial t_1t_2}M_{\boldsymbol{X}}(\boldsymbol{t}) &= n(n-1)(p_1e^{t_1}+\ldots+p_ke^{t_k})^{n-2}p_1e^{t_1}p_2e^{t_2}. \end{aligned}$$

Letting $t_1=0$, we have

$$E[X_i = np_i.$$

Letting $t_1=t_2=0$, we have

$$E[X_1X_2]=n(n-1)p_1p_2.$$

4. Using the results in the previous step, we have

$$Cov(X_1,X_2)=n(n-1)p_1p_2-n^2p_1p_2=-np_1p_2,$$

indicating a negative correlation between X_1 and X_2.

<u>**Definition:**</u> Let $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$ be random vectors with joint pdf $f(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)$. Let $f_{\mathbf{X}_i}(\boldsymbol{x}_i)$ be the marginal pdf of $\boldsymbol{X}_i$. Then $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$ are said to be **mutually independent random vectors** if, for every $(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)$,

$$f(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)=\prod_{i=1}^n f_{\mathbf{X}_i}(\boldsymbol{x}_i).$$

If the X_i's are all one-dimensional, then $X_1,\ldots,X_n$ are called **mutually independent random variables**.
The definition for discrete random variables is similar.

Mutual independence implies that any pair of random variables are pairwise independent. It is a much stronger statement than pairwise independence.

Theorem 6.7.4. *Let $X_1,\ldots,X_n$ be mutually independent random variables. Let $g_1,\ldots,g_n$ be real valued functions such that $g_i(x_i)$ is a function of only $x_i, i=1,\ldots,n$. Then*

$$E[g_1(X_1)\ldots g_n(X_n)]=E[g_1(X_1)]\ldots E[g_n(X_n)].$$

■

Theorem 6.7.5. *Let $X_1,\ldots,X_n$ be mutually independent random variables with moment generating* *functions $M_{X_1}(t), \dots, M_{X_n}(t)$ respectively. Let $Z = X_1 + \dots + X_n$. Then*

$$M_Z(t) = M_{X_1}(t) \dots M_{X_n}(t). \tag{6.25}$$

If all the n random variables have the same distribution with mgf $M_X(t)$, then

$$M_Z(t) = [M_X(t)]^n. \tag{6.26}$$

■

Theorem 6.7.6. *Let $X_1, \dots, X_n$ be mutually independent normal random variables with $X_i \sim N(\mu_i, \sigma_i^2)$. Then $Y = a_1X_1 + \dots + a_nX_n$ is normally distributed with mean*

$$\mu_Y = a_1\mu_1 + \dots + a_n\mu_n$$

and variance

$$\sigma_Y^2 = a_1^2\sigma_1^2 + \dots + a_n^2\sigma_n^2.$$

■

Example: Let $X_1, \dots, X_n$ be mutually independent gamma random variables with $X_i \sim G(\alpha_i, \beta)$. Then $Y = X_1 + \dots + X_n \sim G(\sum \alpha_i, \beta)$.

■

6.8 The Multivariate Normal

Let $Y_1, \ldots, Y_p$ be p random variables. Let $\boldsymbol{Y}$ be a $p \times 1$ column vector, with the $i-$ th element equal to Y_i. Then $\boldsymbol{Y}$ is a **random vector**.
Let $\boldsymbol{\mu}$ be the $p \times 1$ vector

$$\boldsymbol{\mu} = \begin{bmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_p) \end{bmatrix}.$$

This is called the **mean vector**. The dispersion matrix or variance-covariance matrix is given by

$$\boldsymbol{\Sigma} = E[(\boldsymbol{Y} - \boldsymbol{\mu})(\boldsymbol{Y} - \boldsymbol{\mu})^T] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \ldots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \ldots & \sigma_{pp} \end{bmatrix},$$

where

$$\sigma_{ii} = \sigma_i^2 = Var(Y_i)$$

and

$$\sigma_{ij} = Cov(Y_i, Y_j).$$

The dispersion matrix is a symmetric $p \times p$ matrix, with $\boldsymbol{\Sigma}$ being non-negative definite (nnd). We have

$$\boldsymbol{x}^T \boldsymbol{\Sigma} \boldsymbol{x} = \sum_i \sum_j x_i x_j \sigma_{ij} = Var(x_1 Y_1 + \ldots + x_p Y_p)$$

which is ≥ 0 and $= 0$ if $x_1 Y_1 + \ldots + x_p Y_p = 0$..
The correlation matrix is given by

$$\boldsymbol{P} == \begin{bmatrix} 1 & \rho_{12} & \ldots & \rho_{1p} \\ \rho_{21} & 1 & \ldots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \ldots & 1 \end{bmatrix},$$

where

$$\rho_{ij} = Corr(Y_i, Y_j).$$

Let $Z_1, \ldots, Z_p$ be iid $N(0,1)$ random variables. Let

$$\boldsymbol{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{bmatrix}.$$

The joint pdf of $Z_1, \ldots, Z_p$ is given by

$$\begin{aligned} f(z_1, \ldots, z_p) &= \left(\frac{1}{2\pi}\right)^{p/2} exp\left\{-\frac{1}{2}\sum_{i=1}^{p} z_i^2\right\} \\ &= \left(\frac{1}{2\pi}\right)^{p/2} exp\left\{-\frac{1}{2}\boldsymbol{z}^T\boldsymbol{z}\right\}. \end{aligned}$$

We have

$$E(\boldsymbol{Z}) = \begin{bmatrix} E(Z_1) \\ E(Z_2) \\ \vdots \\ E(Z_p) \end{bmatrix} = \boldsymbol{0},$$

the zero vector. The dispersion matrix is given by

$$\boldsymbol{\Sigma} = E[\boldsymbol{Z}\boldsymbol{Z}^T] = \boldsymbol{I}.$$

This is the pdf of $\boldsymbol{Z}$, the $p-$variate standard normal $N_p(\boldsymbol{0}, \boldsymbol{I})$.

Theorem 6.8.1. *Let $\boldsymbol{Z} \sim N_p(\boldsymbol{0}, \boldsymbol{I})$. The moment generating function of $\boldsymbol{Z}$ is*

$$m_{\boldsymbol{Z}}(\boldsymbol{t}) = E(e^{\boldsymbol{t}'\boldsymbol{Z}}) = exp(\boldsymbol{t}'\boldsymbol{t}/2), \qquad \boldsymbol{t} \in \mathcal{R}^p.$$

Proof:

$$\begin{aligned} m_{\boldsymbol{Z}}(\boldsymbol{t}) &= E(e^{\boldsymbol{t}'\boldsymbol{Z}}) \\ &= E(e^{t_1 Z_1 + \ldots + t_p Z_p}) \\ &= \prod_{i=1}^{p} E(e^{t_i Z_i}) = \prod_{i=1}^{p} e^{t_i^2/2} \\ &= e^{\sum_{i=1}^{p} t_i^2/2} = exp(\boldsymbol{t}'\boldsymbol{t}/2) \end{aligned}$$

for all $\boldsymbol{t} \in \mathcal{R}^p$.

■

Theorem 6.8.2. *Let $\boldsymbol{Z} \sim N_p(\boldsymbol{0}, \boldsymbol{I})$. Let $\boldsymbol{\alpha}$ be any non-zero vector of constants and β a scalar. Then*

$$X = \boldsymbol{\alpha}'\boldsymbol{Z} + \beta \sim N(\beta, \boldsymbol{\alpha}'\boldsymbol{\alpha}).$$

■

Definition: A random vector $\boldsymbol{Y}$ is said to have a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and dispersion matrix $\boldsymbol{\Sigma}$ (nnd) if it has the same distribution as that of $\boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{Z}$, where $\boldsymbol{Z} \sim N_p(\boldsymbol{0}, \boldsymbol{I})$ and $\boldsymbol{A}$ is any nonsingular (or such that $\rho(\boldsymbol{A}) = \rho(\boldsymbol{\Sigma}) = K > 0$) matrix such that $\boldsymbol{A}\boldsymbol{A}' = \boldsymbol{\Sigma}$. [If $\boldsymbol{\Sigma} > 0$, then $\boldsymbol{A}$ is of rank p and square; otherwise, $\boldsymbol{A}$ is of order $p \times K$ and of rank $= K$.) We write $\boldsymbol{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Theorem 6.8.3. *Let $\boldsymbol{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The moment generating function of $\boldsymbol{Y}$ is*

$$m_{\boldsymbol{Y}}(\boldsymbol{t}) = exp\left[\boldsymbol{t}^{'}\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}^{'}\boldsymbol{\Sigma}\boldsymbol{t}\right].$$

<u>**Proof:**</u>

$$\begin{aligned}
m_{\boldsymbol{Y}}(\boldsymbol{t}) &= E(e^{\boldsymbol{t}^{'}\boldsymbol{Y}}) \\
&= E(e^{\boldsymbol{t}^{'}(\boldsymbol{\mu}+\boldsymbol{A}\boldsymbol{Z})}) \\
&= e^{\boldsymbol{t}^{'}\boldsymbol{\mu}}E(e^{(\boldsymbol{t}^{'}\boldsymbol{A})\boldsymbol{Z}}) \\
&= e^{\boldsymbol{t}^{'}\boldsymbol{\mu}}E(e^{(\boldsymbol{A}^{'}\boldsymbol{t})^{'}\boldsymbol{Z}}) = e^{\boldsymbol{t}^{'}\boldsymbol{\mu}}E(e^{\boldsymbol{l}^{'}\boldsymbol{Z}}) \\
&= e^{\boldsymbol{t}^{'}\boldsymbol{\mu}}exp(\boldsymbol{l}^{'}\boldsymbol{l}/2) = e^{\boldsymbol{t}^{'}\boldsymbol{\mu}}exp\{(\boldsymbol{t}^{'}\boldsymbol{A}\boldsymbol{A}^{'}\boldsymbol{t}/2)\} \\
&= exp\left[\boldsymbol{t}^{'}\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}^{'}\boldsymbol{\Sigma}\boldsymbol{t}\right]
\end{aligned}$$

■

To obtain the density of the multivariate normal, we use the theory of transformations. Let $\boldsymbol{A}$ be nonsingular, which implies $\boldsymbol{\Sigma}$ is positive definite. If

$$\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{Z},$$

then $\boldsymbol{Z} = \boldsymbol{A}^{-1}(\boldsymbol{Y} - \boldsymbol{\mu})$. The Jacobian of the transformation is

$$|\boldsymbol{J}| = \left|\frac{\partial \boldsymbol{Z}}{\partial \boldsymbol{Y}}\right| = |\boldsymbol{A}^{-1}| = |\boldsymbol{\Sigma}|^{-1/2}.$$

The pdf of $\boldsymbol{Y}$ is then given by

$$f(\boldsymbol{y}) = \left(\frac{1}{2\pi}\right)^{p/2} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} exp\left\{-\frac{1}{2}(\boldsymbol{Y} - \boldsymbol{\mu})^{'}\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{\mu})\right\}.$$

Theorem 6.8.4. *Distribution of a linear combination. Let $\boldsymbol{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let $\boldsymbol{B}$ be a $p \times q$ matrix of constants and $\boldsymbol{b}$ a $q \times 1$ vector of constants. Then*

$$\boldsymbol{V} = \boldsymbol{B}\boldsymbol{Y} + \boldsymbol{b} \sim N_q(\boldsymbol{B}\boldsymbol{\mu} + \boldsymbol{b}, \boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}^{'}).$$

The proof uses the mgf of $\boldsymbol{Y}$.

■

Theorem 6.8.5. *Let $\boldsymbol{Y} \sim N_p(\boldsymbol{0}, \sigma^2\boldsymbol{I})$. Let $\boldsymbol{V} = \boldsymbol{P}\boldsymbol{Y}$, where $\boldsymbol{P}$ is orthogonal. Then*

$$\boldsymbol{V} \sim N_p(\boldsymbol{0}, \sigma^2\boldsymbol{I}).$$

■

An orthogonal transformation preserves the distribution.

Theorem 6.8.6. *Reproductive Property: Let* $\boldsymbol{Y}_i \sim N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), i = 1, \ldots, n$. *Then for fixed constants* $a_1, \ldots, a_n$, *we have*

$$\boldsymbol{V} = a_1\boldsymbol{Y}_1 + \ldots + a_n\boldsymbol{Y}_n \sim N_p\left(\sum a_i\boldsymbol{\mu}_i, \sum a_i^2\boldsymbol{\Sigma}_i\right).$$

■

Corollary 6.8.7. *Let* $\boldsymbol{Y}_i \sim N_p\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right), i = 1, \ldots, n$ *be iid random vectors. Then*

$$\overline{\boldsymbol{Y}} = \frac{\boldsymbol{Y}_1 + \ldots + \boldsymbol{Y}_n}{n} \sim N_p\left(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma}\right).$$

■

6.8.1 Marginal and Conditional Distributions

Let

$$\boldsymbol{Y}_{p\times 1} = \begin{bmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{bmatrix},$$

where $\boldsymbol{Y}_1$ is a $q \times 1$ vector. Partition $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ appropriately.

Theorem 6.8.8. *If* $\boldsymbol{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, *then*

$$\boldsymbol{Y}_1 \sim N_q(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}).$$

<u>Proof</u>

Choose the $q \times p$ matrix $\boldsymbol{B} = [\boldsymbol{I}_{q\times q} : \mathbf{0}]$. Using the previous result, we have

$$\boldsymbol{BY} = \boldsymbol{Y}_1 \sim N_q(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}).$$

■

Theorem 6.8.9. *If* $\boldsymbol{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, *then* $\boldsymbol{Y}_1$ *and* $\boldsymbol{Y}_2$ *are independent iff*

$$\boldsymbol{\Sigma}_{12} = \mathbf{0}; \qquad \boldsymbol{\Sigma}_{21} = \mathbf{0}.$$

<u>Proof:</u> We have

$$\begin{aligned} m_{\boldsymbol{Y}}(\boldsymbol{t}) &= E(e^{\boldsymbol{t}'\boldsymbol{Y}}) \\ &= exp\left[\boldsymbol{t}_1'\boldsymbol{\mu}_1 + \boldsymbol{t}_2'\boldsymbol{\mu}_2 + \frac{\boldsymbol{t}_1'\boldsymbol{\Sigma}_{11}\boldsymbol{t}_1}{2} + \frac{\boldsymbol{t}_1'\boldsymbol{\Sigma}_{12}\boldsymbol{t}_2}{2} + \frac{\boldsymbol{t}_2'\boldsymbol{\Sigma}_{21}\boldsymbol{t}_1}{2} + \frac{\boldsymbol{t}_2'\boldsymbol{\Sigma}_{22}\boldsymbol{t}_2}{2}\right] \end{aligned}$$

If $\boldsymbol{\Sigma}_{12} = \mathbf{0}$ and $\boldsymbol{\Sigma}_{21} = \mathbf{0}$, then the moment generating function reduces to

$$m_{\boldsymbol{Y}}(\boldsymbol{t}) = exp\left[\boldsymbol{t}_1'\boldsymbol{\mu}_1 + \frac{\boldsymbol{t}_1'\boldsymbol{\Sigma}_{11}\boldsymbol{t}_1}{2}\right] exp\left[\boldsymbol{t}_2'\boldsymbol{\mu}_2 + \frac{\boldsymbol{t}_2'\boldsymbol{\Sigma}_{22}\boldsymbol{t}_2}{2}\right],$$

implying independence.
If they are independent, then the covariances must be 0. ∎

6.8.2 Conditional Distributions

Theorem 6.8.10. *Let $\boldsymbol{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $R(\boldsymbol{\Sigma}) = p$. The conditional distribution of $\boldsymbol{Y}_1$ given $\boldsymbol{Y}_2 = \boldsymbol{C}_2$ is*

$$N_q(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{C}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11.2}),$$

where

$$\boldsymbol{\Sigma}_{11.2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}.$$

Proof: The joint pdf of $\boldsymbol{Y}_1, \boldsymbol{Y}_2$ is the distribution of $\boldsymbol{Y}$:

$$f(\boldsymbol{y}) = \left(\frac{1}{2\pi}\right)^{p/2} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} exp\left\{-\frac{1}{2}(\boldsymbol{Y} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{\mu})\right\}.$$

We have

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} [\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}]^{-1} & -\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22.1}^{-1} \\ -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11.2}^{-1} & \boldsymbol{\Sigma}_{22.1}^{-1} \end{bmatrix}.$$

The quadratic form in the exponent of the pdf may be written as

$$\begin{aligned}
(\boldsymbol{Y} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}) &= (\boldsymbol{Y}_1 - \boldsymbol{\mu}_1)'\boldsymbol{\Sigma}_{11.2}^{-1}(\boldsymbol{Y}_1 - \boldsymbol{\mu}_1) \\
&- (\boldsymbol{Y}_1 - \boldsymbol{\mu}_1)'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22.1}^{-1}(\boldsymbol{C}_2 - \boldsymbol{\mu}_2) \\
&- (\boldsymbol{C}_2 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_{22.1}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{Y}_1 - \boldsymbol{\mu}_1) \\
&+ (\boldsymbol{C}_2 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_{22.1}^{-1}(\boldsymbol{C}_2 - \boldsymbol{\mu}_2) \\
&= Q_1.
\end{aligned}$$

We have

$$f(\boldsymbol{y}_1|\boldsymbol{y}_2 = \boldsymbol{C}_2) = \frac{f(\boldsymbol{y})}{f_{\boldsymbol{Y}_2}(\boldsymbol{C}_2)}.$$

The marginal distribution of $\boldsymbol{Y}_2 \sim N_{p-q}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$. Therefore

$$f(\boldsymbol{y}_1|\boldsymbol{y}_2) = \left(\frac{1}{2\pi}\right)^{q/2} \frac{|\boldsymbol{\Sigma}_{22}|^{1/2}}{|\boldsymbol{\Sigma}|^{1/2}} exp\left\{-\frac{1}{2}Q_1 + (\boldsymbol{C}_2 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{C}_2 - \boldsymbol{\mu}_2)\right\}.$$

We have $|\boldsymbol{\Sigma}| = |\boldsymbol{\Sigma}_{22}|\ |\boldsymbol{\Sigma}_{11.2}|$. The term in the exponent reduces to

$$[(\boldsymbol{Y}_1 - \boldsymbol{\mu}_1) - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{C}_2 - \boldsymbol{\mu}_2)]'\boldsymbol{\Sigma}_{11.2}^{-1}[(\boldsymbol{Y}_1 - \boldsymbol{\mu}_1) - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{C}_2 - \boldsymbol{\mu}_2)].$$

For $\boldsymbol{Y}_2$ fixed at $\boldsymbol{C}_2$, the term

$$\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{C}_2 - \boldsymbol{\mu}_2)$$

is fixed. Therefore

$$\boldsymbol{Y}_1|\boldsymbol{Y}_2 \sim N_q(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{C}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11.2}),$$

The equation

$$E[\boldsymbol{Y}_1|\boldsymbol{Y}_2 = \boldsymbol{C}_2] = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{C}_2 - \boldsymbol{\mu}_2)$$

is called the regression of $\boldsymbol{Y}_1$ on the variables in $\boldsymbol{Y}_2$.

■

Remarks: The mean of the conditional distribution depends on the $\boldsymbol{C}_2$, but the conditional variance is the same for all values of $\boldsymbol{C}_2$.