MSSC 6010 / Comp. Probability

Instructor: Mehdi Maadooliat

Chapter 5

Department of Mathematics, Statistics and Computer Science

Special thanks to Prof. Ana Militino for providing the original slides of the book.

MARQUETTE UNIVERSITY
Be The Difference.

Chapter 5

Multivariate Probability Distributions

5.1 Joint Distribution of Two Random Variables

5.1.1 Joint pdf for Two Discrete Random Variables

If X and Y are discrete random variables, the function given by

$$p_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y) \tag{5.1}$$

for each pair of values (x, y) within the domain of X and Y is called the joint **pdf** of X and Y. Any function $p_{X,Y}(x,y)$ can be used as a joint **pdf** provided the following properties are satisfied:

1

(i) $p_{X,Y}(x, y) \geq 0$ for all x and y

(ii) $\sum_x \sum_y p_{X,Y}(x, y) = 1$

(iii) $\mathbb{P}[(X, Y) \in A] = \sum\sum_{(x,y)\in A} p_{X,Y}(x, y)$

Property (iii) states that when A is composed of pairs of (x, y) values, the probability $\mathbb{P}[(X, Y) \in A]$ is obtained by summing the joint **pdf** over pairs in A.

2

Example 5.1 ▷ ***Joint Distribution: Mathematics Grades*** ◁ To graduate with a bachelor of science (B.S.) degree in mathematics, all majors must pass Calculus III and Linear Algebra with a grade of B or better. The population of B.S. graduates in mathematics earned grades as given in Table 5.2 on page 6.

Table 5.1: B.S. graduate grades in Linear Algebra and Calculus III

			Linear Algebra (Y)		
			A	B	C
(X)	Calculus III	A	2	13	6
		B	5	85	40
		C	7	33	9

(a) What is the probability of getting a B or better in Linear Algebra?

(b) What is the probability of getting a B or better in Calculus III?

(c) What is the probability of getting a B or better in both Calculus III and Linear Algebra?

3

Table 5.1: B.S. graduate grades in Linear Algebra and Calculus III

			Linear Algebra (Y)		
			A	B	C
(X)	Calculus III	A	2	13	6
		B	5	85	40
		C	7	33	9

Solution: The answers are:

(a) Let the random variables X and Y represent the grades in Calculus III and Linear Algebra respectively. If A_1 represents the pairs of Calculus III and Linear Algebra values such that the grade in Linear Algebra is a B or better, then the probability of getting a B or better in Linear Algebra is written

$$\mathbb{P}\left[(X,Y)\in A_1\right] = \sum_{(x,y)\in A_1}\sum p_{X,Y}(x,y) = \frac{2+5+7+13+85+33}{200}$$

$$= \frac{145}{200}$$

4

Table 5.1: B.S. graduate grades in Linear Algebra and Calculus III

			Linear Algebra (Y)		
			A	B	C
(X)	Calculus III	A	2	13	6
		B	5	85	40
		C	7	33	9

(b) Let the random variables X and Y represent the grades in Calculus III and Linear Algebra respectively. If A_2 represents the pairs of Calculus III and Linear Algebra values such that the grade in Calculus III is a B or better, then the probability of getting a B or better in Calculus III is written

$$\mathbb{P}\left[(X,Y)\in A_2\right] = \sum_{(x,y)\in A_2}\sum p_{X,Y}(x,y) = \frac{2+13+6+5+85+40}{200}$$

$$= \frac{151}{200}$$

5

Table 5.1: B.S. graduate grades in Linear Algebra and Calculus III

			Linear Algebra (Y)		
			A	B	C
(X)	Calculus III	A	2	13	6
		B	5	85	40
		C	7	33	9

(c) Let the random variables X and Y represent the grades in Calculus III and Linear Algebra respectively. If A_3 represents the pairs of Calculus III and Linear Algebra values such that the grade in both Calculus III and Linear Algebra is a B or better, then the probability of getting a B or better in both Calculus III and Linear Algebra is written

$$\mathbb{P}\left[(X,Y) \in A_3\right] = \sum_{(x,y)\in A_3}\sum p_{X,Y}(x,y) = \frac{2+5+13+85}{200} = \frac{105}{200}. \blacksquare$$

For any random variables X and Y, the joint **cdf** is defined in (5.2), while the marginal **pdf**s of X and Y, denoted $p_X(x)$, and $p_Y(y)$

Table 5.2: B.S. graduate grades in Linear Algebra and Calculus III

			Linear Algebra (Y)			
			A	B	C	Total(X)
(X)	Calculus III	A	2	13	6	21
		B	5	85	40	130
		C	7	33	9	49
	Total(Y)		14	131	55	200

respectively are defined in Equations (5.3) and (5.4).

$$F_{X,Y}(x,y) = \mathbb{P}(X \le x, Y \le y), \quad -\infty < x < \infty, \quad -\infty < y < \infty \tag{5.2}$$

$$p_X(x) = \sum_y p_{X,Y}(x,y) \tag{5.3}$$

$$p_Y(y) = \sum_x p_{X,Y}(x,y) \tag{5.4}$$

In (a) of Example 5.1 on page 3, the problem requests the probability of getting a B or better in Linear Algebra. Another way to compute

Table 5.2: B.S. graduate grades in Linear Algebra and Calculus III

			Linear Algebra (Y)			
			A	B	C	Total(X)
(X)	Calculus III	A	2	13	6	21
		B	5	85	40	130
		C	7	33	9	49
Total(Y)			14	131	55	200

the answer is by adding the two marginals $p_Y(A) + p_Y(B) = \frac{14}{200} + \frac{131}{200} = \frac{145}{200}$. Likewise, (b) of Example 5.1 on page 3 can also be solved with the marginal distribution for X: $p_X(A) + p_X(B) = \frac{21}{200} + \frac{130}{200} = \frac{151}{200}$.

8

5.1.2 Joint pdf for Two Continuous Random Variables

The joint **pdf** of two continuous random variables is any integrable function $f_{X,Y}(x,y)$ with the following properties:

(1) $f_{X,Y}(x,y) \geq 0$ for all x and y

(2) $\int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx\, dy = 1$

(3) $\mathbb{P}[(X,Y) \in A] = \iint\limits_{(x,y)\in A} f_{X,Y}(x,y)\, dx\, dy$

Property (3) implies that $\mathbb{P}[(X,Y) \in A]$ is the volume of a solid over the region A bounded by the surface $f_{X,Y}(x,y)$.

9

For any random variables X and Y, the joint **cdf** is defined in (5.5), while the marginal **pdf**s of X and Y, denoted $f_X(x)$, and $f_Y(y)$ respectively are defined in Equations (5.6) and (5.7).

$$F_{X,Y}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(r,s)\,ds\,dr, \quad -\infty < x < \infty, \; -\infty < y < \infty \tag{5.5}$$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy, \quad -\infty < x < \infty \tag{5.6}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx, \quad -\infty < y < \infty \tag{5.7}$$

10

Example 5.2 Given the joint continuous **pdf**

$$f_{X,Y}(x,y) = \begin{cases} 1 & \text{if} \quad 0 \le x \le 1, \quad 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find $F_{X,Y}(x = 0.6, y = 0.8)$.

(b) Find $\mathbb{P}(0.25 \le X \le 0.75, 0.1 \le Y \le 0.9)$.

(c) Find $f_X(x)$.

Solution: The answers are:

(a)

$$F_{X,Y}(x = 0.6, y = 0.8) = \int_{0}^{0.6}\int_{0}^{0.8} f_{X,Y}(r,s)\,ds\,dr$$

$$= \int_{0}^{0.6}\int_{0}^{0.8} 1\,ds\,dr = \int_{0}^{0.6} 0.8\,dr = 0.48$$

11

Example 5.2 Given the joint continuous pdf

$$f_{X,Y}(x,y) = \begin{cases} 1 & \text{if} \quad 0 \le x \le 1, \quad 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(b) Find $\mathbb{P}(0.25 \le X \le 0.75, 0.1 \le Y \le 0.9)$.

(c) Find $f_X(x)$.

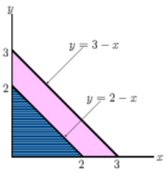
(b)

$$\mathbb{P}(0.25 \le x \le 0.75, 0.1 \le y \le 0.9) = \int_{0.25}^{0.75}\int_{0.1}^{0.9} f_{X,Y}(r,s)\,ds\,dr = \int_{0.25}^{0.75}\int_{0.1}^{0.9} 1\,ds\,dr = \int_{0.25}^{0.75} 0.8\,dr = 0.40$$

(c)

$$f_X(x) = \int_0^1 f_{X,Y}(x,y)\,dy = 1, \quad 0 \le x \le 1 \quad ■$$

Example 5.3 ▷ ***Joint*** **PDF** ◁ Find the value c to make $f_{X,Y}(x,y) = cx$ a valid joint pdf for $x > 0$, $y > 0$, and $2 < x+y < 3$.



Solution:

$$V1 = \int_0^3\int_0^{3-x} cx\,dy\,dx = c\int_0^3 (3x - x^2)\,dx = c\left[\frac{3x^2}{2} - \frac{x^3}{3}\Big|_0^3\right] = \frac{27c}{6}$$

$$V2 = \int_0^2\int_0^{2-x} cx\,dy\,dx = c\int_0^2 (2x - x^2)\,dx = c\left[x^2 - \frac{x^3}{3}\Big|_0^2\right] = \frac{8c}{6}$$

$$V1 - V2 = \frac{27c}{6} - \frac{8c}{6} \overset{\text{set}}{=} 1 \Rightarrow c = \frac{6}{19}$$

MARQUETTE UNIVERSITY
Be The Difference.

5.2 Independent Random Variables

Two random variables are independent if for every pair of x and y values, $p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)$, when X and Y are discrete, or $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$ when X and Y are continuous.

Example 5.4 Use Table 5.2 on page 6 to decide if the random variables X, grade in Calculus III, and Y, grade in Linear Algebra, are dependent.

Table 5.2: B.S. graduate grades in Linear Algebra and Calculus III

		Linear Algebra (Y)			
		A	B	C	Total(X)
(X) Calculus III	A	2	13	6	21
	B	5	85	40	130
	C	7	33	9	49
Total(Y)		14	131	55	200

14

MARQUETTE UNIVERSITY
Be The Difference.

Solution: The random variables X and Y are dependent if $p_{X,Y}(x,y) \neq p_X(x) \cdot p_Y(y)$ for any (x,y). Consider the pair $(x,y) = (A,A)$, that is an A in both Calculus III and in Linear Algebra.

Table 5.2: B.S. graduate grades in Linear Algebra and Calculus III

		Linear Algebra (Y)			
		A	B	C	Total(X)
(X) Calculus III	A	2	13	6	21
	B	5	85	40	130
	C	7	33	9	49
Total(Y)		14	131	55	200

$$p_{X,Y}(A,A) \stackrel{?}{=} p_X(A) \cdot p_Y(A)$$
$$\frac{2}{200} \stackrel{?}{=} \frac{21}{200} \cdot \frac{14}{200}$$
$$\frac{2}{200} \neq \frac{21 \times 14}{40,000}$$
$$0.01 \neq 0.00735$$

Since $0.01 \neq 0.00735$, the random variables X and Y, the grades in Calculus III and Linear Algebra respectively, are dependent. It is important to note that the definition of independence requires all the joint probabilities to be equal to the product of the corresponding row and column marginal probabilities. Consequently, if the joint probability of a single entry is not equal to the product of the corresponding row and column marginal probabilities, the random variables in question are said to be dependent. ∎

15

MARQUETTE UNIVERSITY
Be The Difference.

Example 5.5 Are the random variables X and Y in Example 5.2 independent? Recall that the **pdf** for Example 5.2 was defined as

$$f_{X,Y}(x,y) = \begin{cases} 1 & \text{if} \quad 0 \le x \le 1, \quad 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Solution: Since the marginal **pdf** for X, $f_X(x) = 1$, and the marginal **pdf** for Y, $f_Y(y) = 1$, it follows that X and Y are independent since $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$ for all x and y. ■

16

MARQUETTE UNIVERSITY
Be The Difference.

5.3 Several Random Variables This section examines the joint **pdf** of several random variables by extending the material presented for the joint **pdf** of two discrete random variables and two continuous random variables covered in Section 5. The joint **pdf** of $X_1, X_2, \ldots, X_n$ discrete random variables is any function $p_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$ provided the following properties are satisfied:

(a) $p_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) \ge 0$ for all $x_1, x_2, \ldots, x_n$.

(b) $\sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} p_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = 1$

(c) $\mathbb{P}[(X_1, X_2, \ldots, X_n) \in A] = \sum \sum_{(x_1,x_2,\ldots,x_n) \in A} \cdots \sum p_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)$

17

SEVERAL CONTINUOUS RANDOM VARIABLES

The joint **pdf** of $X_1, X_2, \ldots, X_n$, continuous random variables, is any integrable function $f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)$ such that following properties are satisfied:

(a) $f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) \geq 0$ for all $x_1, x_2, \ldots, x_n$.

(b) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)dx_1\,dx_2\cdots dx_n = 1$

(c)

$$\mathbb{P}\left[(X_1, X_2, \ldots, X_n) \in A\right] = \iint\limits_{(x_1,x_2,\ldots,x_n)\in A}\cdots\int f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)$$

INDEPENDENT RANDOM VARIABLES

Independence for several random variables is simply a generalization of the notion for the independence between two random variables. $X_1, X_2, \ldots, X_n$ are independent if for every subset of the random variables, the joint **pdf** of the subset is equal to the product of the marginal **pdf**s. Further, if $X_1, X_2, \ldots, X_n$ are independent random

Further, if $X_1, X_2, \ldots, X_n$ are independent random variables with respective moment generating functions $M_{X_1}(t), M_{X_2}(t), \ldots, M_{X_n}(t)$ then the moment generating function of $Y = \sum_{i=1}^{n} c_i X_i$ is

$$M_Y(t) = M_{X_1}(c_1 t) \times M_{X_2}(c_2 t) \times \cdots \times M_{X_n}(c_n t). \qquad (5.8)$$

In the case where $X_1, X_2, \ldots, X_n$ are independent normal random variables, a theorem for the distribution of $Y = a_1X_1 + \cdots + a_nX_n$, where $a_1, a_2, \ldots, a_n$ are constants, is stated.

MARQUETTE UNIVERSITY
Be The Difference.

Theorem 5.1 If $X_1, X_2, \ldots, X_n$ are independent normal random variables, with means μ_i and standard deviations σ_i for $i = 1, 2, \ldots, n$, the distribution of $Y = a_1X_1 + a_2X_2 + \cdots + a_nX_n$, where $a_1, a_2, \ldots, a_n$ are constants is normal with mean $E[Y] = a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n$,

and variance $\text{var}[Y] = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2$. In other words,

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

20

MARQUETTE UNIVERSITY
Be The Difference.

Proof: Since $X_i \sim N(\mu_i, \sigma_i)$, the **mgf** for X_i is $M_{X_i}(t) = e^{\mu_i t + \frac{\sigma_i^2 t^2}{2}}$ using the **mgf** from Box (4.14) Further, since the $X_1, X_2, \ldots, X_n$ are independent,

$$\begin{aligned} M_Y(t) &= M_{X_1}(ta_1) \times M_{X_2}(ta_2) \times \cdots \times M_{X_n}(ta_n) \\ &= e^{t\sum_{i=1}^{n} a_i\mu_i + t^2 \sum_{i=1}^{n} \frac{a_i^2\sigma_i^2}{2}}, \end{aligned}$$

which is the moment generating function for a normal random variable with mean $\sum_{i=1}^{n} a_i\mu_i$ and variance $\sum_{i=1}^{n} a_i^2\sigma_i^2$.

21

MARQUETTE UNIVERSITY Be The Difference.

5.4 Conditional Distributions Suppose X and Y represent the respective lifetimes (in years) for the male and the female in married couples. If $X = 72$, what is the probability that $Y \geq 75$? In other words, if the male partner of a marriage dies at age 72, how likely is it that the surviving female will live to an age of 75 or more? Questions of this type are answered with conditional distributions. Given two discrete random variables, X and Y, define the conditional **pdf** of X given that $Y = y$ provided that $p_Y(y) > 0$ as

$$p_{X|Y}(x|y) = \mathbb{P}(X = x|Y = y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}. \tag{5.9}$$

22

MARQUETTE UNIVERSITY Be The Difference.

CONDITIONAL DISTRIBUTION FOR TWO CONTINUOUS RANDOM VARIABLES

If the random variables are continuous, the conditional **pdf** of X given that $Y = y$ provided that $f_Y(y) > 0$ is defined as

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}. \tag{5.10}$$

In addition, if X and Y are jointly continuous over an interval A,

$$\mathbb{P}(X \in A|Y = y) = \int_A f_{X|Y}(x|y)\,dx.$$

23

MARQUETTE UNIVERSITY
Be The Difference.

Example 5.7 Let the random variables X and Y have joint **pdf**:

$$f_{X,Y}(x,y) = \begin{cases} \frac{12}{5}x(2-x-y) & \text{for } 0 < x < 1,\ 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the **pdf** of X given $Y = y$, for $0 < y < 1$.

Solution: Using the definition for the conditional **pdf** of X given $Y = y$ from (5.10), write

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_{X,Y}(x,y)}{\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx} = \frac{x(2-x-y)}{\int_0^1 x(2-x-y)\,dx}$$

$$= \frac{x(2-x-y)}{2/3 - y/2} = \frac{6x(2-x-y)}{4-3y} \quad \text{for} \quad 0 < x < 1,\ 0 < y < 1$$

■

24

MARQUETTE UNIVERSITY
Be The Difference.

Example 5.8 ▷ ***Joint Distribution: Radiators*** ◁ A local radiator manufacturer subjects his radiators to two tests. The function that describes the percentage of radiators that pass the two tests is

$$f_{X,Y}(x,y) = 8xy, \quad 0 \le y \le x \le 1 \qquad (5.11)$$

The random variable X represents the percentage of radiators that pass test A, and Y represents the percentage of radiators that pass test B.

(a) Is the function given in (5.11) a **pdf**?

(b) Determine the marginal and conditional **pdf**s for X and Y.

(c) Are X and Y independent?

(d) Compute the probability that less than $\frac{1}{8}$ of the radiators will pass test B given that $\frac{1}{2}$ have passed test A.

(e) Compute the quantities: $E[X]$, $E\left[X^2\right]$, var(X), $E[Y]$, $E\left[Y^2\right]$, and var(Y).

(f) Use S to represent graphically (5.11).

25

$$f_{X,Y}(x,y) = 8xy, \quad 0 \le y \le x \le 1$$

Solution: The answers are:

(a) The function (5.11) is a **pdf** since $f_{X,Y}(x,y)$ is nonnegative and

$$\int_0^1 \int_0^x 8xy\,dy\,dx = 8\int_0^1 \left[x \int_0^x y\,dy \right] dx = 8\int_0^1 \frac{x^3}{2} dx = 1$$

MARQUETTE UNIVERSITY Be The Difference.

$$f_{X,Y}(x,y) = 8xy, \quad 0 \le y \le x \le 1$$

(b) The marginal and conditional **pdf**s are:

$$f_X(x) = \int f(x,y)dy = \int_0^x 8xy\,dy = 4x^3, \quad 0 \le x \le 1$$

$$f_Y(y) = \int f(x,y)dx = \int_y^1 8xy\,dx = 4y(1-y^2), \quad 0 \le y \le 1$$

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{8xy}{4y(1-y^2)} = \frac{2x}{1-y^2}, \quad y \le x \le 1$$

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{8xy}{4x^3} = \frac{2y}{x^2}, \quad 0 \le y \le x$$

MARQUETTE UNIVERSITY
Be The Difference.

$$f_{X,Y}(x,y) = 8xy, \quad 0 \le y \le x \le 1$$

$$f_X(x) = \int f(x,y)dy = \int_0^x 8xy\,dy = 4x^3, \quad 0 \le x \le 1$$
$$f_Y(y) = \int f(x,y)dx = \int_y^1 8xy\,dx = 4y(1-y^2), \quad 0 \le y \le 1$$

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{8xy}{4x^3} = \frac{2y}{x^2}, \quad 0 \le y \le x$$

(c) The random variables X and Y are dependent since $f_{X,Y}(x,y) = 8xy \neq f_X(x) \cdot f_Y(y) = 16x^3y - 16x^3y^3$.

(d) The probability that $\mathbb{P}\left(Y < \frac{1}{8} \middle| X = \frac{1}{2}\right)$ is computed as:

$$\mathbb{P}\left(Y < \frac{1}{8} \middle| X = \frac{1}{2}\right) = \int_0^{\frac{1}{8}} f_{Y|X}(y|\frac{1}{2})\,dy = \int_0^{\frac{1}{8}} \frac{2y}{\frac{1}{4}}\,dy = 4y^2|_0^{\frac{1}{8}} = \frac{1}{16}$$

28

MARQUETTE UNIVERSITY
Be The Difference.

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{8xy}{4y(1-y^2)} = \frac{2x}{1-y^2}, \quad y \le x \le 1$$
$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{8xy}{4x^3} = \frac{2y}{x^2}, \quad 0 \le y \le x$$

Likewise

$$F_{X|Y}(x|y) = \int_{-\infty}^x f(x|y)\,dx = \int_y^x \frac{2x}{1-y^2}\,dx = \frac{x^2-y^2}{1-y^2} \quad y \le x < 1$$

$$F_{Y|X}(y|x) = \int_{-\infty}^y f(y|x)dy = \int_0^y \frac{2y}{x^2}dy = \frac{y^2}{x^2} \quad 0 < y \le x$$

$$P(Y < 1/8|X = 1/2) = F_{Y|X}(y = \tfrac{1}{8}|x = \tfrac{1}{2}) = \frac{(1/8)^2}{(1/2)^2} = \frac{1}{16}$$

29

(e) The quantities $E[X]$, $E\left[X^2\right]$, $\mathrm{var}(X)$, $E[Y]$, $E\left[Y^2\right]$, and $\mathrm{var}(Y)$ are

$$E[X] = \int_0^1 x \cdot 4x^3\,dx = 4\int_0^1 x^4\,dx = \frac{4}{5}$$

$$E\left[X^2\right] = \int_0^1 x^2 \cdot 4x^3\,dx = 4\int_0^1 x^5\,dx = \frac{2}{3}$$

$$\mathrm{var}(X) = E\left[X^2\right] - [E[X]]^2 = \frac{2}{3} - \frac{16}{25} = \frac{2}{75}$$

$$E[Y] = \int_0^1 y \cdot 4y(1-y^2)\,dy = 4\int_0^1 (y^2 - y^4)\,dy = \frac{8}{15}$$

$$E\left[Y^2\right] = \int_0^1 y^2 \cdot 4y(1-y^2)\,dy = 4\int_0^1 (y^3 - y^5)\,dy = \frac{1}{3}$$

$$\mathrm{var}(Y) = E\left[Y^2\right] - [E[Y]]^2 = \frac{1}{3} - \frac{64}{225} = \frac{11}{225}$$

30

(f) The following code can be used to create Figure 5.2

```
➢ function.draw <- function(f, low=-1, hi=1, n=30) {
      r<-seq(low, hi, length=n)
      z <-outer (r,r,f)
      persp(r,r,z,xlab="X",ylab="Y",zlab="Z",theta=-70)
  }
➢ f3 <- function(x,y) {ifelse(x >= y, 8*x*y, 0)}
➢ function.draw(f3,0,1,25)
```

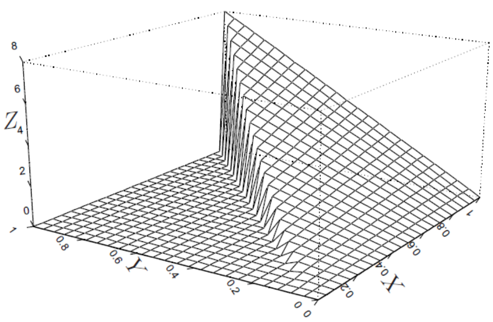


Figure 5.2: Graphical representation of $f_{X,Y}(x,y) = 8xy, \quad 0 \le y \le x \le 1$

31

Be careful not to assume the variance of the sum of two random variables is the sum of the variances of each random variable. Only if X and Y are independent is it true that $\text{var}[X+Y] = \text{var}[X]+\text{var}[Y]$. A simple example to show why this is not true in general is computing $\text{var}[X+X] \neq \text{var}[X]+\text{var}[X]$ since $\text{var}[X+X] = \text{var}[2X] = 4\text{var}[X]$.

However, if $X_1, X_2, \ldots X_n$ are n independent random variables with means $\mu_1, \mu_2, \ldots, \mu_n$, and variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$ respectively, then the mean and variance of $Y = \sum_{i=1}^n c_i X_i$ where the c_is are real valued constants are $\mu_Y = \sum_{i=1}^n c_i \mu_i$ and $\sigma_Y^2 = \sum_{i=1}^n c_i^2 \sigma_i^2$. The proofs of the last two statements are left as exercises for the reader.

32

Example 5.9 Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean μ and standard deviation σ. Find the mean and variance of $Y = \frac{X_1+X_2+\cdots+X_n}{n}$.

Solution: In the expression $Y = \frac{X_1+X_2+\cdots+X_n}{n}$, the c_i values are all $\frac{1}{n}$. Consequently, $\mu_Y = \sum_{i=1}^n \frac{1}{n} \cdot \mu = \mu$ and $\sigma_Y^2 = \sum_{i=1}^n \left(\frac{1}{n}\right)^2 \cdot \sigma^2 = \frac{\sigma^2}{n}$. ■

- **In general:**

- $\text{var}[aX \pm bY] = a^2\text{var}[X] + b^2\text{var}[Y] \pm 2ab\,\text{cov}[X,Y]$,
- **where** $\text{cov}[X,Y] = E\{(X - E[X])(Y - E[Y])\}$

- $\text{var}[\sum_{i=1}^n a_i X_i] = \sum_{i,j=1}^n a_i a_j \text{cov}[X_i, X_j]$
- $= \sum_{i=1}^n a_i^2 \text{var}[X_i] + \sum_{i \neq j} a_i a_j \text{cov}[X_i, X_j]$

33

MARQUETTE UNIVERSITY Be The Difference.

5.5 Expected Values, Covariance, and Correlation

The expected value of $g(X, Y)$ is

$$E[g(X,Y)] = \begin{cases} \sum_x \sum_y g(x,y) \cdot p_{X,Y}(x,y) & \text{if } X \text{ and } Y \text{ are discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) \cdot f_{X,Y}(x,y)\, dx\, dy & \text{if } X \text{ and } Y \text{ are continuous} \end{cases} \tag{5.12}$$

The conditional expectation of X given a value y of Y is written

$$E[X|Y] = \begin{cases} \sum_x x \cdot p_{X|Y}(x|y) & \text{if } X \text{ and } Y \text{ are discrete} \\ \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y)\, dx & \text{if } X \text{ and } Y \text{ are continuous} \end{cases} \tag{5.13}$$

MARQUETTE UNIVERSITY Be The Difference.

Example 5.10 Let the random variables X and Y have joint **pdf**

$$f_{X,Y}(x,y) = \frac{e^{-y/x}e^{-x}}{x} \quad x > 0, \quad y > 0$$

Compute $E[Y|X = x]$.

Solution: First, compute the conditional **pdf** $f_{Y|X}(y|x)$.

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{\frac{e^{-y/x}e^{-x}}{x}}{\int_0^\infty \frac{e^{-y/x}e^{-x}}{x} dy} = \frac{\frac{e^{-y/x}}{x}}{\int_0^\infty \frac{e^{-y/x}}{x} dy}$$

$$= \frac{e^{-y/x}}{x}, \quad x > 0, \quad y > 0$$

Using (5.13) for continuous random variables, write

$$E[Y|X = x] = \int_0^\infty y \cdot \frac{e^{-y/x}}{x} dy$$

Using (5.13) for continuous random variables, write

$$E[Y|X=x] = \int_0^\infty y \cdot \frac{e^{-y/x}}{x}\, dy$$

Integrating by parts with $u = y$, and $dv = \frac{e^{-y/x}}{x}$, obtain

$$E[Y|X=x] = -ye^{-y/x}\Big|_0^\infty + \int_0^\infty e^{-y/x} dy = 0 + -xe^{-y/x}\Big|_0^\infty = x, \quad x > 0.$$

■

USEFUL RESULT

When two random variables, say X and Y, are independent, recall that $f(x,y) = f_X(x) \cdot f_Y(y)$ for the continuous case and $p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)$ for the discrete case. Further, $E[XY] = E[X] \cdot E[Y]$. The last statement is true for both continuous and discrete X and Y. A proof for the discrete case is provided. Note that the proof in the continuous case would simply consist of exchanging the summation signs for integral signs.

Proof:

$$\begin{aligned} E[XY] &= \sum_x \sum_y xy\, p_{X,Y}(x,y) = \sum_x \sum_y xy\, p_X(x)\, p_Y(y) \\ &= \sum_y y\, p_Y(y) \sum_x x\, p_X(x) \\ &= E[Y]E[X] \end{aligned}$$

MARQUETTE UNIVERSITY Be The Difference.

Example 5.11 Use the joint **pdf** provided in Example 5.8 on page 33 and compute $E[XY]$.

$$f_{X,Y}(x,y) = 8xy, \quad 0 \le y \le x \le 1$$

$$f_X(x) = \int f(x,y)dy = \int_0^x 8xy\,dy = 4x^3, \quad 0 \le x \le 1$$

$$f_Y(y) = \int f(x,y)dx = \int_y^1 8xy\,dx = 4y(1-y^2), \quad 0 \le y \le 1$$

$$E[X] = \int_0^1 x \cdot 4x^3\,dx = 4\int_0^1 x^4\,dx = \frac{4}{5}$$

$$E[Y] = \int_0^1 y \cdot 4y(1-y^2)\,dy = 4\int_0^1 (y^2 - y^4)\,dy = \frac{8}{15}$$

Solution:

$$E[XY] = \int_0^1 \int_0^x xy \cdot 8xy\,dy\,dx = 8\int_0^1 \left[x^2 \int_0^x y^2\,dy\right] dx = 8\int_0^1 \frac{x^5}{3}\,dx = \frac{4}{9}$$

Since the random variables X and Y were found to be dependent in part (c) of Example 5.8 on page 33, note that

$$E[XY] = \frac{4}{9} \ne E[X] \cdot E[Y] = \frac{4}{5} \cdot \frac{8}{15} = \frac{32}{75}$$

MARQUETTE UNIVERSITY Be The Difference.

5.5.2 Covariance When two variables, X and Y, are not independent or when it is noted that $E[XY] \ne E[X] \cdot E[Y]$, one is naturally interested in some measure of their dependency. The covariance of X and Y written $Cov[X,Y]$, provides one measure of the degree to which X and Y tend to move linearly in either the same or opposite directions. The covariance of two random variables X and Y is defined as

$$Cov[X,Y] = E\big[(X-\mu_X)(Y-\mu_Y)\big] = \begin{cases} \sum_x \sum_y (x-\mu_X)(y-\mu_Y)p_{X,Y}(x,y) & X, Y \text{ discrete} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y)f(x,y)\,dx\,dy & X, Y \text{ continuo} \end{cases} \tag{5.14}$$

At times, it will be easier to work with the shortcut formula $Cov[X,Y] = E[XY] - \mu_X \cdot \mu_Y$ instead of using the definition in (5.14).

Example 5.13 Compute the covariance between X and Y for Example 5.8 on page 33. In part (e) of Example 5.8, $E[X]$ and $E[Y]$ were computed to be $\frac{4}{5}$ and $\frac{8}{15}$ respectively, and in Example 5.11 it was found that $E[XY] = \frac{4}{9}$.

$$f_{X,Y}(x,y) = 8xy, \quad 0 \le y \le x \le 1$$

$$E[X] = \int_0^1 x \cdot 4x^3\,dx = 4\int_0^1 x^4\,dx = \frac{4}{5}$$

$$E[Y] = \int_0^1 y \cdot 4y(1-y^2)\,dy = 4\int_0^1 (y^2 - y^4)\,dy = \frac{8}{15}$$

$$E[XY] = \int_0^1\int_0^x xy \cdot 8xy\,dy\,dx = 8\int_0^1 \left[x^2 \int_0^x y^2\,dy\right] dx = 8\int_0^1 \frac{x^5}{3}\,dx = \frac{4}{9}$$

Solution:

$$Cov[X,Y] = E[XY] - \mu_X\mu_Y = \frac{4}{9} - \frac{4}{5}\cdot\frac{8}{15} = \frac{4}{225}$$

Example 5.12 Compute the covariance between X_1 and Y_1 for the values provided in Table 5.3 on the next page given that $p_{X,Y}(x,y) = \frac{1}{10}$ for each (x,y) pair.

Table 5.3: Values Used to Compute Covariance for Figure 5.3

X_1	Y_1	X_2	Y_2	X_3	Y_3
58	80	58	1200	25.5	30.0
72	80	72	1200	27.0	33.0
72	90	72	1100	30.0	34.5
86	90	86	1100	33.0	33.0
86	100	86	1000	34.5	30.0
100	100	100	1000	33.0	27.0
100	110	100	900	30.0	25.5
114	110	114	900	27.0	27.0
114	120	114	800	28.8	30.0
128	120	128	800	31.2	30.0

MARQUETTE UNIVERSITY
Be The Difference.

Solution:

$$p_{X_1}(x) = \sum_y p_{X_1,Y_1}(x,y)$$

$$\mu_{X_1} = \sum_x x \cdot p_{X_1}(x) = \frac{58+72+\cdots+128}{10} = 93$$

$$\mu_{Y_1} = \sum_y y \cdot p_{Y_1}(y) = \frac{80+80+\cdots+120}{10} = 100$$

$$Cov[X_1,Y_1] = \sum_x \sum_y (x-\mu_{X_1})(y-\mu_{Y_1})p_{X_1,Y_1}(x,y)$$

$$= (58-93)\cdot(80-100)\cdot\frac{1}{10} + (72-93)\cdot(80-100)\cdot\frac{1}{10} +$$

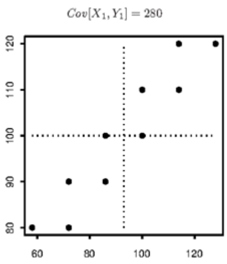
$$+ (128-120)\cdot(120-100)\cdot\frac{1}{10}$$

$$= 280$$

42

MARQUETTE UNIVERSITY
Be The Difference.

- A $Cov[X,Y] > 0$ indicates that generally, as X increases, so does Y (that is, X and Y move in the same direction);
- To gain an intuitive understanding of covariance, see Figure 5.3 on page which has both horizontal and vertical dotted lines to indicate μ_{X_i} and μ_{Y_i}

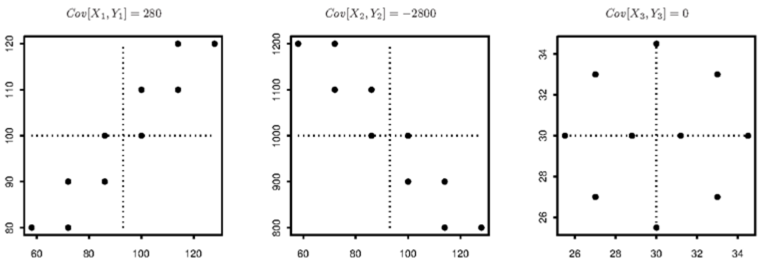


- The first plot in Figure 5.3 exhibits a strong positive relationship.
- By this it is meant that large values of X tend to occur with large values of Y and small values of X tend to occur with small values of Y.
- Consequently, $(x-\mu_X)$ will tend to have the same sign as $(y-\mu_Y)$, so their product will be positive.

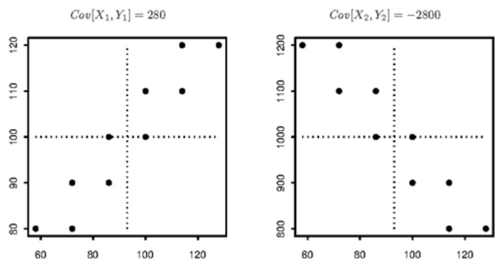
43

- whereas, a $Cov[X,Y] < 0$ indicates that generally, as X increases Y decreases (that is, X and Y move in opposite directions).

- In the center plot of Figure 5.3, the relationship between the two variables is negative, and note that $(x-\mu_{X_2})$ and $(y-\mu_{Y_2})$ tend to have opposite signs, which makes most of their products negative.



When one examines the first two plots in Figure 5.3 on page 58, the dependency in the left plot seems to be about as strong as the dependency in the center plot, just in the opposite direction. However, the $Cov[X,Y] = 280$ in the left plot and $Cov[X,Y] = -2800$ in the center plot. It turns out that the dependencies are the same (just in opposite directions), but the units of measurement for the Y variable in the center plot are a factor of 10 times larger than those in the left plot. So, it turns out that covariance is unit dependent. To eliminate this unit dependency, scale the covariance.



5.5.3 Correlation

- The **correlation coefficient** between X and Y denoted $\rho_{X,Y}$, or simply ρ is a scale independent measure of linear dependency between two random variables.
- The independence in scale is achieved by dividing the covariance by $\sigma_X \sigma_Y$.
- Specifically, define the correlation between X and Y as

$$\rho_{X,Y} = \frac{Cov[X,Y]}{\sigma_X \sigma_Y} \tag{5.15}$$

- The correlation coefficient measures the degree of linear dependency between two random variables and is bounded by -1 and $+1$.
- The values $\rho = -1$ and $\rho = +1$ indicate perfect negative and positive relationships between two random variables. When $\rho = 0$ there is an absence of linear dependency between X and Y.

- If X and Y are independent, it is also true that $\rho = 0$; however, $\rho = 0$ does not imply independence.
- A similar statement is true for the $Cov[X,Y]$. That is, if X and Y are independent, $Cov[X,Y] = 0$; however, $Cov[X,Y] = 0$ does not imply independence.

Example 5.14 Compute $\rho_{X,Y}$ for Example 5.8 on page 33. Recall that $Cov[X,Y] = \frac{4}{225}$ was computed in Example 5.13 on page 64, and $\text{var}[X] = \frac{2}{75}$ and $\text{var}[Y] = \frac{11}{225}$ in part (e) of Example 5.8 on page 33.

Solution:

$$\rho_{X,Y} = \frac{Cov[X,Y]}{\sigma_X \sigma_Y} = \frac{\frac{4}{225}}{\sqrt{\frac{2}{75} \cdot \frac{11}{225}}} = 0.4924$$

■

MARQUETTE UNIVERSITY Be The Difference.

Example 5.15 Given the random variables X and Y with their joint probability distribution provided in Table 5.4, verify that although $Cov[X,Y]=0$, X and Y are dependent.

Table 5.4: Joint probability distribution for X and Y

		Y		
		-1	0	1
X	-1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
	0	$\frac{1}{8}$	0	$\frac{1}{8}$
	1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

48

MARQUETTE UNIVERSITY Be The Difference.

		Y		
		-1	0	1
X	-1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
	0	$\frac{1}{8}$	0	$\frac{1}{8}$
	1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

Solution: Start by computing the quantities $E[XY]$, $E[X]$, and $E[Y]$ to use in the shortcut formula for the covariance.

$$E[X] = (-1)\cdot\frac{3}{8} + (0)\cdot\frac{2}{8} + (1)\cdot\frac{3}{8} = 0$$
$$E[Y] = (-1)\cdot\frac{3}{8} + (0)\cdot\frac{2}{8} + (1)\cdot\frac{3}{8} = 0$$
$$E[XY] = (-1\cdot -1)\cdot\frac{1}{8} + \cdots + (1\cdot 1)\cdot\frac{1}{8} = 0$$
$$Cov[X,Y] = E[XY] - E[X]\cdot E[Y] = 0$$

The covariance for this problem is 0. However, the random variables are dependent since

$$\mathbb{P}(X=-1, Y=-1) = \frac{1}{8} \neq \mathbb{P}(X=-1)\cdot\mathbb{P}(Y=-1) = \frac{3}{8}\cdot\frac{3}{8} = \frac{9}{64}.$$

This example reinforces the idea that a covariance or correlation coefficient of 0 does not imply independence. ■

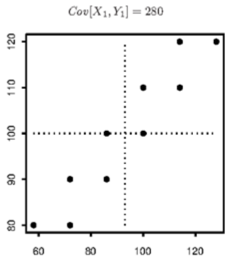
49

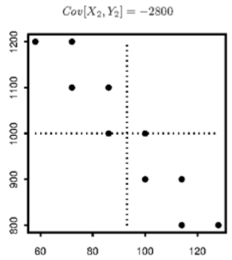
MARQUETTE UNIVERSITY Be The Difference.

Example 5.16 Compute ρ_{X_1,Y_1} for Example 5.12 on page 59. Recall that $\mu_{X_1} = 93$, $\mu_{Y_1} = 100$ and $Cov[X_1, Y_1] = 280$.

X_1	Y_1
58	80
72	80
72	90
86	90
86	100
100	100
100	110
114	110
114	120
128	120

X_2	Y_2
58	1200
72	1200
72	1100
86	1100
86	1000
100	1000
100	900
114	900
114	800
128	800





50

MARQUETTE UNIVERSITY Be The Difference.

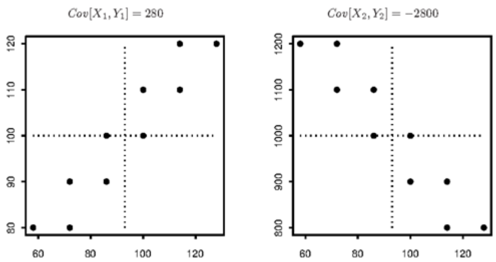
Solution: Start by computing the quantities $E\left[X_1^2\right]$, $E\left[Y_1^2\right]$, σ_{X_1} and σ_{X_2}.

$$E\left[X_1^2\right] = \sum_x x^2 p_{X_1}(x)$$
$$= 58^2 \cdot \frac{1}{10} + 72^2 \cdot \frac{1}{10} + \cdots + 128^2 \cdot \frac{1}{10} = 9090$$
$$E\left[Y_1^2\right] = \sum_y y^2 p_{Y_1}(y)$$
$$= 80^2 \cdot \frac{1}{10} + 80^2 \cdot \frac{1}{10} + \cdots + 120^2 \cdot \frac{1}{10} = 10200$$
$$\text{var}[X_1] = E\left[X_1^2\right] - \left(E[X_1]\right)^2 = 9090 - 93^2 = 441$$
$$\sigma_{X_1} = \sqrt{\text{var}[X_1]} = \sqrt{441} = 21$$
$$\text{var}[Y_1] = E\left[Y_1^2\right] - \left(E[Y_1]\right)^2 = 10200 - 100^2 = 200$$
$$\sigma_{Y_1} = \sqrt{\text{var}[Y_1]} = \sqrt{200} = 14.14214$$
$$\rho_{X_1,Y_1} = \frac{Cov[X_1, Y_1]}{\sigma_{X_1}\sigma_{Y_1}} = \frac{280}{21 \times 14.14214} = 0.9428087$$

51

```
> cor(X1,Y1)
[1] 0.942809
```

It is worthwhile to note that $\rho_{X_1,Y_1} = 0.9428087$ and $\rho_{X_2,Y_2} = -0.9428087$ for the left and center plots respectively in Figure 5.3
In other words, the correlations have the same absolute magnitude for both plots, even though the absolute values of the covariances differ by a factor of ten.



5.6 Multinomial Distribution

- The multinomial distribution is a generalization of the binomial distribution. Recall that each trial in a binomial experiment results in only one of two mutually exclusive outcomes.

- Experiments where each trial can result in any one of k possible mutually exclusive outcomes $A_1, \ldots, A_k$ with probabilities $\mathbb{P}(A_i) = \pi_i$, $0 < \pi_i < 1$, for $i = 1, \ldots, k$ such that $\sum_{i=1}^{k} \pi_i = 1$ can be modelled with the **multinomial distribution**.

- Specifically, the multinomial distribution computes the probability that A_1 occurs x_1 times, A_2 occurs x_2 times, $\ldots, A_k$ occurs x_k times in n independent trials where $x_1 + x_2 + \cdots + x_k = n$.

- To derive the probability distribution function, reason in a fashion similar to that done with the binomial.

MARQUETTE UNIVERSITY Be The Difference.

- Since the trials are independent, any specified ordering yielding x_1 outcomes for A_1, x_2 outcomes for $A_2, \dots,$ and x_k outcomes for A_k will occur with probability $\pi_1^{x_1}\pi_2^{x_2}\cdots\pi_k^{x_k}$.
- The total number of orderings yielding x_1 outcomes for A_1, x_2 outcomes for $A_2, \dots,$ and x_k outcomes for A_k is $\frac{n!}{x_1!x_2!\cdots x_k!}$.
- With these two facts in mind, the probability distribution, mean, variance, and **mgf** of a multinomial distribution can be derived. All are found in Box (5.16).

Multinomial Distribution $\mathbf{X} \sim MN(n, \pi_1, \dots, \pi_k)$

$$\mathbb{P}(\mathbf{X} = (x_1, \dots, x_k)|n, \pi_1, \dots, \pi_k) = \frac{n!}{x_1!x_2!\cdots x_k!}\pi_1^{x_1}\pi_2^{x_2}\cdots\pi_k^{x_k}$$

$$E[X_i] = n\pi_i$$

$$\text{var}[X_i] = n\pi_i(1-\pi_i)$$

given that each $X_i \sim Bin(n, \pi_i)$

$$M_{\mathbf{X}}(t) = (\pi_1 e^{t_1} + \pi_2 e^{t_2} + \cdots + \pi_{k-1}e^{t_{k-1}} + \pi_k e^{t_k})^n \quad (5.16)$$

MARQUETTE UNIVERSITY Be The Difference.

Example 5.17 The probability a particular type of light bulb lasts less than 500 hours is 0.5 and the probability the same type light bulb lasts more than 800 hours is 0.2. In a random sample of ten light bulbs, what is the probability of obtaining exactly four light bulbs that last less than 500 hours and two light bulbs that last more than 800 hours?

Solution: Let the random variables $X_1, X_2,$ and X_3 denote the number of light bulbs that last less than 500 hours, the number of light bulbs that last between 500 and 800 hours, and the number of light bulbs that last more than 800 hours respectively. Since $\pi_1 = 0.5$, $\pi_2 = 0.3$ and $\pi_3 = 0.2$, use the first equation in Box (5.16) and compute $\mathbb{P}(X_1 = 4, X_2 = 4, X_3 = 2)$ as

$$\begin{aligned}\mathbb{P}(X_1 = 4, X_2 = 4, X_3 = 2|10, 0.5, 0.3, 0.2) &= \frac{10!}{4!4!2!}(0.5)^4(0.3)^4(0.2)^2 \\ &= 0.0638.\end{aligned}$$

5.7 Bivariate Normal Distribution

The joint distribution of the random variables X and Y is said to have a **bivariate normal** distribution when its joint density takes the form

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}, \quad (5.17)$$

for $-\infty < x, y < +\infty$, where $\mu_X = E[X]$, $\mu_Y = E[Y]$, $\sigma_X^2 = \text{var}[X]$, $\sigma_Y^2 = \text{var}[Y]$ and ρ is the correlation coefficient between X and Y.

An equivalent representation of (5.17) is given in (5.18) where $\mathbf{X} = (X,Y)^T$ is a vector of random variables where T represents the transpose, $\boldsymbol{\mu} = (\mu_X, \mu_Y)^T$, is a vector of constants, and $\boldsymbol{\Sigma}$ a 2×2 nonsingular matrix such that its inverse $\boldsymbol{\Sigma}^{-1}$ exists and the determinant $|\boldsymbol{\Sigma}| \neq 0$ where

$$\boldsymbol{\Sigma} = \begin{pmatrix} \text{var}[X] & Cov[X,Y] \\ Cov[Y,X] & \text{var}[Y] \end{pmatrix}.$$

$$f(\mathbf{x}) = \frac{1}{(\sqrt{2\pi})^2}|\boldsymbol{\Sigma}|^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})\right\} \quad (5.18)$$

MULTIVARIATE NORMAL DISTRIBUTION

The shorthand notation used to denote a multivariate (bivariate being a subset) normal distribution is $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. In general, $\boldsymbol{\Sigma}$ represents what is called the variance covariance matrix. When $\mathbf{X} = (X_1, X_2, \ldots, X_n)$

$$\boldsymbol{\Sigma} = E[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T] = E\left[\begin{pmatrix} X_1-\mu_1 \\ \vdots \\ X_n-\mu_n \end{pmatrix}(X_1-\mu_1, \ldots, X_n-\mu_n)\right]$$

$$= \begin{pmatrix} \sigma^2_{X_1} & \ldots & Cov(X_1, X_n) \\ \vdots & \ddots & \vdots \\ Cov(X_n, X_1) & \ldots & \sigma^2_{X_n} \end{pmatrix}.$$

$$f(\mathbf{x}) = \frac{1}{(\sqrt{2\pi})^n}|\boldsymbol{\Sigma}|^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})\right\}$$

where

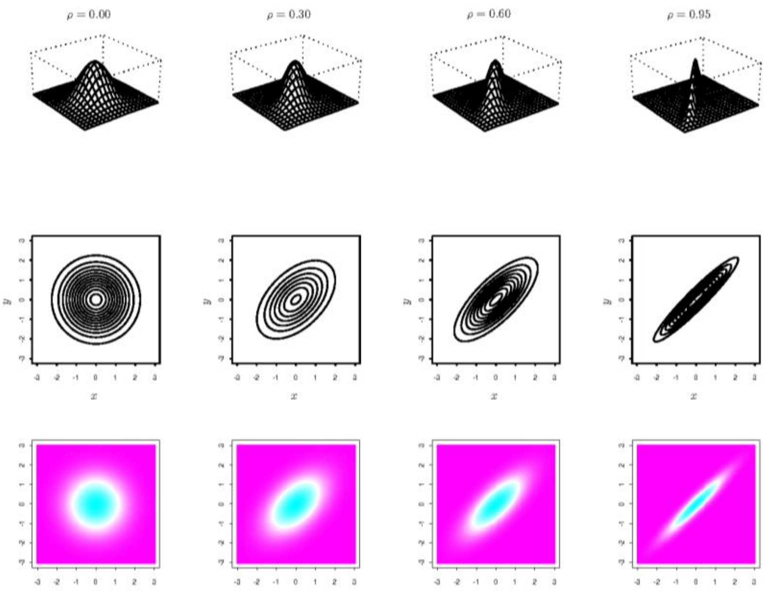
- **Mean vector:** $\boldsymbol{\mu}$
- **Covariance Matrix:** $\boldsymbol{\Sigma}$
- **MGF:** $M_{\mathbf{X}}(\mathbf{t}) = \exp\left(\boldsymbol{\mu}^T\mathbf{t} + \frac{1}{2}\mathbf{t}^T\boldsymbol{\Sigma}\mathbf{t}\right)$

58

Different representations of four bivariate normal distributions all with parameters $\mu_X = \mu_Y = 0$, $\sigma_X = \sigma_Y = 1$, and ρ values of 0, 0.30, 0.60, and 0.95 respectively are provided in Figure 5.4.

```
> function1.draw <- function(f, low=-1, hi=1, n=50){
>     xy <- seq(low, hi, length = n)
>     z <- outer(xy, xy, f)
>     persp(xy, xy, z, axes=FALSE, box=TRUE)
> }
> f1 <- function(x,y) {
>     r <- 0.30
>     exp( (x^2-2*r*x*y+y^2) / (-2*(1-r^2)) )/
>     (2*pi*sqrt(1-r^2))
> }
> par(mfrow=c(1,3), pty="s")
> x <- seq(-3,3,length=100)
> y <- x
> function1.draw(f1,-3,3,20)
> contour(x,y,outer(x,y,f1),nlevels=10)
> image(x,y,outer(x,y,f1),zlim=range(outer(x,y,f1)))
```

59



SOME FACTS ABOUT BIVARIATE NORMAL DISTRIBUTION

- **If** $\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left[\boldsymbol{\mu} = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{YX} & \sigma_X^2 \end{pmatrix}\right]$, **and** $\rho = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}$**:**

The following facts about the bivariate normal distribution are listed without proof.

(a) The marginal distribution of X is $N(\mu_X, \sigma_X)$.

(b) The marginal distribution of Y is $N(\mu_Y, \sigma_Y)$.

(c) If X and Y have a bivariate normal distribution, the conditional density of Y given $X = x$ is a normal distribution with mean $\mu_{Y|x} = E(Y|x) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$ and variance $\sigma_{Y|x}^2 = \sigma_Y^2(1 - \rho^2)$.

(d) Given any two constants a and b, the distribution of $aX + bY$ is

$$N(a\mu_X + b\mu_Y, \sqrt{a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho\sigma_X\sigma_Y})$$

SOME FACTS ABOUT MULTIVARIATE NORMAL DISTRIBUTION

MARQUETTE UNIVERSITY Be The Difference.

- If $\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left[\mu = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{pmatrix}\right]$:

The following facts about the bivariate normal distribution are listed without proof.

a) **The marginal distribution of X is $N(\mu_X, \Sigma_X)$**

b) **The marginal distribution of Y is $N(\mu_Y, \Sigma_Y)$**

c) **The conditional distribution of Y given $X = x$ is a multivariate normal distribution with mean**
   - ✓ $\mu_{Y|X} = \mu_Y + \Sigma_{YX}\Sigma_X^{-1}(x - \mu_X)$

   and variance
   - ✓ $\Sigma_{Y|X} = \Sigma_Y - \Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}$

d) **Given any matrix A and vector b, $AX + b$ is a multivariate normal distribution with mean and variance:**
   - ✓ $\mu_{AX+b} = A\mu_X + b$ $\qquad \Sigma_{AX+b} = A\Sigma_X A^T$

62

MARQUETTE UNIVERSITY Be The Difference.

Example 5.18 ▷ *Bivariate Normal Grades* ◁ Let us assume that the distribution of grades for a particular group of students where X and Y represent the respective grade point averages in high school and the first year of college respectively follow a bivariate normal distribution with parameters $\mu_X = 3.2$, $\mu_Y = 2.4$, $\sigma_X = 0.4$, $\sigma_Y = 0.6$, and $\rho = 0.6$. Find the following:

(a) $\mathbb{P}(Y < 1.8)$

(b) $\mathbb{P}(Y < 1.8 \,|\, X = 2.5)$

(c) $\mathbb{P}(Y > 3.0)$

(d) $\mathbb{P}(Y > 3.0 \,|\, X = 2.5)$

63

Solution: The answers are computed first manually then with S.

(a) Using the parameters given in the problem,

$$\mathbb{P}(Y < 1.8) = \mathbb{P}\left(\frac{Y-2.4}{0.6} < \frac{1.8-2.4}{0.6}\right) = \mathbb{P}(Z < -1) = 0.1586$$

```
> pnorm(1.8,2.4,.6)
[1] 0.1586553
```

(b) First, find the quantities $\mu_{Y|x=2.5}$ and $\sigma_{Y|x=2.5}$.

$$\mu_{Y|x=2.5} = E(Y|x=2.5) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)$$
$$= 2.4 + 0.6 \cdot \frac{0.6}{0.4} \cdot (2.5-3.2) = 1.77$$

$$\sigma^2_{Y|x=2.5} = \sigma^2_Y(1-\rho^2) = 0.6^2 \cdot \left(1-0.6^2\right) = 0.2304 \Rightarrow \sigma_{Y|x=2.5} = 0.48$$

$\mathbb{P}(Y < 1.8|X = 2.5) = \mathbb{P}\left(\frac{Y-1.77}{0.48} < \frac{1.8-1.77}{0.48}\right) = \mathbb{P}(Z < 0.0625) = 0.5249.$

```
> pnorm(1.8,1.77,.48)
[1] 0.5249177
```

(c) Using the parameters given in the problem,

$$\mathbb{P}(Y > 3.0) = 1 - \mathbb{P}(Y \le 3.0) = 1 - \mathbb{P}\left(\frac{Y-2.4}{0.6} \le \frac{3.0-2.4}{0.6}\right)$$
$$= 1 - \mathbb{P}(Z \le 1) = 0.1586$$

```
> 1-pnorm(3,2.4,.6)
[1] 0.1586553
```

(d) Using the quantities $\mu_{Y|x}$ and $\sigma_{Y|x}$ from (b),

$$\begin{aligned}\mathbb{P}(Y > 3.0|X = 2.5) &= 1 - \mathbb{P}(Y \leq 3.0 \,|\, X = 2.5)\\ &= 1 - \mathbb{P}\left(\frac{Y - 1.77}{0.48} \leq \frac{3.0 - 1.77}{0.48}\right)\\ &= 1 - \mathbb{P}(Z \leq 2.5625)\\ &= 0.0052.\end{aligned}$$

```
> 1-pnorm(3,1.77,.48)
[1] 0.005196079
```

■

QUESTIONS?

- **ANY QUESTION?**