Advanced Topics In Introductory Probability

A first Course in Probability Theory – Volume III

Nicholas N.N. Nsowah-Nuamah

bookboon
The eBook company

NICHOLAS N.N. NSOWAH-NUAMAH

ADVANCED TOPICS IN INTRODUCTORY PROBABILITY

A FIRST COURSE IN PROBABILITY THEORY – VOLUME III

Download free eBooks at bookboon.com

Advanced Topics In Introductory Probability: A first Course in Probability Theory – Volume III
2nd edition
© 2018 Nicholas N.N. Nsowah-Nuamah & bookboon.com
ISBN 978-87-403-2238-5

CONTENTS

Click on the ad to read more

Download free eBooks at bookboon.com

Download free eBooks at bookboon.com

PART 1

BIVARIATE PROBABILITY DISTRIBUTIONS

I salute the discovery of a single even insignificant truth more highly than all the argumentation on the highest questions which fails to reach a truth
GALILEO (1564–1642)

Download free eBooks at bookboon.com

Chapter 1

PROBABILITY AND DISTRIBUTION FUNCTIONS OF BIVARIATE DISTRIBUTIONS

1.1 INTRODUCTION

So far, all discussions in the two volumes of my book on probability (Nsowah-Nuamah, 2017 and 2018) have been associated with a single random variable X (that is, a one-dimensional or univariate random variable). Frequently, we may be concerned with multivariate situations that simultaneously involve two or more random variables. For instance, if we wanted to study the relationship between weight and height of individual students we might consider weight and height to be two random variables X and Y, respectively, whose values are determined by measuring the weights and heights of the students in the school. Such study will produce the ordered pair (X, Y).

1.2 CONCEPT OF BIVARIATE RANDOM VARIABLES

1.2.1 Definition of Bivariate Random Variables

Many of the concepts discussed for the one-dimensional random variables also hold for higher-dimensional case. Here in most cases we shall limit ourselves to the two-dimensional (bivariate) case; more complex multivariate situations are straightforward generalisations.

Definition 1.1 BIVARIATE RANDOM VARIABLE
If $X = X(s)$ and $Y = Y(s)$ are two real-valued functions on the sample space S, then the pair (X, Y) that assigns a point in the real (x, y) plane to each point $s \in S$ is called a bivariate random variable

Synonyms of a bivariate random variable are a *two-dimensional random variable/vector* Fig. 1.1 is an illustration of a bivariate random variable.

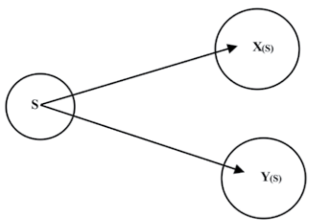


Fig. 1.1 Bivariate Random Variables

Download free eBooks at bookboon.com

1.2.2 Types of Bivariate Random Variables

Multivariate situations, similar to univariate cases, may involve discrete, as well as continuous random variables.

Definition 1.2 DISCRETE BIVARIATE RANDOM VARIABLE
(X, Y) is a discrete bivariate random variable, if each of the random variables X and Y is discrete

Definition 1.3 CONTINUOUS BIVARIATE RANDOM VARIABLE
(X, Y) is a continuous bivariate random variable if each of the random variables is continuous

There are cases where one variable is discrete and the other continuous but this will not be considered here.

1.3 JOINT PROBABILITY DISTRIBUTIONS

A joint distribution is a distribution having two or more random variables, with each random variable still having its own probability distribution, expected value and variance. In addition, for ordered pair values of the random variables, probabilities will exist and the strength of any relationship between the two variables can be measured.

In the multivariate case as in the univariate case we often associate a probability (mass) function with discrete random variables and a probability density function with continuous random variables. We shall take up the discrete case first since it is the easier one to understand.

1.3.1 Joint Probability Distribution of Discrete Random Variables

Suppose that X and Y are discrete random variables, and X takes values $i = 0, 1, 2, \cdots, n$, and Y takes values $j = 1, 2, \cdots, m$. Most often, such a joint distribution is given in table form. Table 1.1 is an n-by-m array which displays the number of occurrences of the various combinations of values of X and Y. We may observe that each row represents values of X and each column represents values of Y. The row and column totals are called *marginal totals*. Such a table is called the *joint frequency distribution*.

Download free eBooks at bookboon.com

Table 1.1 Joint Frequency Distribution of X and Y

X	Y				*Row Totals*
	y_1	y_2	$\cdots$	y_m	
x_1	(x_1, y_1)	(x_1, y_2)	$\cdots$	(x_1, y_m)	$\sum_y x_1$
x_2	(x_2, y_1)	(x_2, y_2)	$\cdots$	(x_2, y_m)	$\sum_y x_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_n	(x_n, y_1)	(x_n, y_2)	$\cdots$	(x_n, y_m)	$\sum_y x_n$
Column Totals	$\sum_x y_1$	$\sum_x y_2$	$\cdots$	$\sum_x y_m$	$\sum_x \sum_y x_i y_j = N$

For example, suppose X and Y are discrete random variables, and X takes values $0, 1, 2, 3$, and Y takes values $1, 2, 3$. Each of the nm row-column intersections in Table 1.2 represents the frequency that belongs to the ordered pair (X, Y).

Table 1.2 Joint Frequency Distribution of X and Y

Values of X	Values of Y			Row Totals
	1	2	3	
0	1	0	0	1
1	0	2	1	3
2	0	2	1	3
3	1	0	0	1
Column Totals	2	4	2	8

Definition 1.4 JOINT PROBABILITY DISTRIBUTION
Let X and Y be discrete random variables with possible values x_i, $i = 1, 2, ..., n$ and $y_j, j = 1, 2, 3, ..., m$, respectively. The joint (or bivariate) probability distribution for X and Y is given by

$$p(x_i, y_j) = P(\{X = x_i\} \cap \{Y = y_j\})$$

defined for all $(x_i,\, y_j)$

The function $p(x_i, y_j)$ is sometimes referred to as the *joint probability mass function* (p.m.f.) or the *joint probability function* (p.f.) of X and Y. This function gives the probability that X will assume a particular value x while at the same time Y will assume a particular value y.

Download free eBooks at bookboon.com

Note

(a) The notation $p(x, y)$ for all (x, y) is the same as writing $p(x_i, y_j)$ for $i = 1, 2, ..., n$ and $j = 1, 2, 3, ..., m$. Sometimes when there is no ambiguity we shall use simply $p(x, y)$.

(b) The joint probability $p(x_i, y_\text{j})$ is sometimes denoted as

$$P(X = x, Y = y),$$

where the comma stands for 'and' or '$\cap$'.

Definition 1.5
If X and Y are discrete random variables with joint probability mass function $p(x_i, y_j)$, then

(a) $p(x_\text{i}, y_\text{j}) \geq 0, \qquad$ for all i and j

(b) $\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) = 1$

Once the joint probability mass function is determined for discrete random variables X and Y, calculation of joint probabilities involving X and Y is straightforward.

Let the value that the random variables X and Y jointly take be denoted by the ordered pair (x_i, y_j). The joint probability $p(x_i, y_\text{j})$ is obtained by counting the number of occurrences of that combination of values X and Y and dividing the count by the total number of all the sample points. Thus,

$$\begin{aligned} P(\{X = x_i\} \cap \{Y = y_j\}) &= \frac{\#(\{X = x_i\} \cap \{Y = y_j\})}{\sum_{i=1}^{n}\sum_{j=1}^{m} \#(\{X = x_i\} \cap \{Y = y_j\})} \\ &= \frac{\#(x_i, y_j)}{\sum_{i=1}^{n}\sum_{j=1}^{m} \#(x_i, y_j)} \end{aligned}$$

where

$\#(x_i, y_j)$ is the number of occurrences in the cell of the ordered pair (x_i, y_j);

$\sum_{i=1}^{n}\sum_{j=1}^{m} \#(x_i, y_j)$ is the total number of all sample points (cells) of the ordered pairs (x_i, y_j), denoted by N.

Download free eBooks at bookboon.com

Joint Probability Distribution of Bivariate Random Variables in Tabular Form

The joint probability distribution may be given in the form of a table of n rows and m columns (See Table 1.3). The upper margins of the table indicate the possible distinct values of X and Y. The numbers in the body of the table are the probabilities for the joint occurrences of the two events corresponding to $X = x_i$ $(1 \leq i \leq n)$ and $Y = y_j$ $(1 \leq i \leq m)$. The row and column totals are the probabilities for the individual random variables and are called ***marginal probabilities*** because they appear on the margins of the table. Such a table is also called the *joint relative frequency distribution*.

Table 1.3 Joint Probability Distribution of X and Y

X	Y				*Row Totals*
	y_1	y_2	$\cdots$	y_m	
x_1	$p(x_1, y_1)$	$p(x_1, y_2)$	$\cdots$	$p(x_1, y_m)$	$p(x_1)$
x_2	$p(x_2, y_1)$	$p(x_2, y_2)$	$\cdots$	$p(x_2, y_m)$	$p(x_2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_n	$p(x_n, y_1)$	$p(x_n, y_2)$	$\cdots$	$p(x_n, y_m)$	$p(x_n)$
Column Totals	$p(y_1)$	$p(y_2)$	$\cdots$	$p(y_m)$	$\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) = 1$

Deloitte.

© Deloitte & Touche LLP and affiliated entities.

Discover the truth at www.deloitte.ca/careers

Download free eBooks at bookboon.com

Click on the ad to read more

Note
The marginal probabilities for X are simply the simple probabilities that $X = x_i$ for values of y_j, where j assumes a value from 1 to m. Similarly, the marginal probabilities for Y are the simple probabilities that $Y = y_j$, where i assumes a value from 1 and n.

It is important to note that the distribution satisfies a joint probability function, namely,

(a) $p(x_i, y_j) \geq 0, \quad \text{for all} \quad i = 1, 2, \cdots, n; \; j = 1, 2, \cdots, m.$

(b) $\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) = 1$

Example 1.1

(a) For the data in Table 1.2, calculate the joint probabilities of X and Y.

(b) Does this distribution satisfy the properties of a joint probability function?

Solution

(a) From Table 1.2, the cell $(\{X = 0\} \cap \{Y = 1\}) = (0, 1)$ contains one element;
Total number of elements in all cells is 8.
Hence

$$\begin{aligned} P(\{X = 0\} \cap \{Y = 1\}) &= p(0, 1) \\ &= \frac{\#(\{X = 0\} \cap \{Y = 1\})}{\sum_{i}^{n}\sum_{j}^{m} \#(\{X = x_i\} \cap \{Y = y_j\})} = \frac{1}{8} \end{aligned}$$

Similarly,

$$\begin{aligned} P(\{X = 0\} \cap \{Y = 2\}) &= p(0, 2) = \frac{0}{8} = 0 \\ P(\{X = 0\} \cap \{Y = 3\}) &= p(0, 3) = \frac{0}{8} = 0 \\ P(\{X = 1\} \cap \{Y = 1\}) &= p(1, 1) = \frac{0}{8} = 0 \\ P(\{X = 1\} \cap \{Y = 2\}) &= p(1, 2) = \frac{2}{8} = \frac{1}{4} \\ P(\{X = 1\} \cap \{Y = 3\}) &= p(1, 3) = \frac{1}{8} = \frac{1}{8} \end{aligned}$$

Download free eBooks at bookboon.com

When probabilities of all possible joint events, $P(X = x_i, Y = y_j)$, have been determined in this fashion, we have a joint probability distribution of X and Y and these results may be presented in a two-way table as shown in the table below:

X	Y		
	1	2	3
0	1/8	0	0
1	0	2/8	1/8
2	0	2/8	1/8
3	1/8	0	0

(b) From the table above,

(i) $p(x_i, y_j) \geq 0, \quad$ for all $\quad i = 0, 1, 2, 3;\ j = 1, 2, 3.$

(ii) $\displaystyle\sum_{i=0}^{3}\sum_{j=1}^{3} p(x_i, y_j) = \frac{1}{8} + \frac{1}{8} + \frac{2}{8} + \frac{2}{8} + \frac{1}{8} + \frac{1}{8} = 1$

Hence this distribution is a joint probability function.

Joint Probability Distribution of Bivariate Random Variables in Expression Form

Sometimes the joint probability distribution of discrete random variables X and Y is given by a formula.

Example 1.2
Given the function

$$p(x, y) = k(3x + 2y), \quad x = 0, 1; \quad y = 0, 1, 2$$

(a) Find the constant $k > 0$ such that the $p(x, y)$ is a joint probability mass function.

(b) Present it in a tabular form for the probabilities associated with the sample points (x, y). Obtain the row and column totals.

Solution
(a) (i) $\quad p(x, y) \geq 0$

(ii)
$$\begin{aligned}\sum_{x=0}^{1}\sum_{y=0}^{2} k(3x + 2y) &= k\sum_{x=0}^{1}[(3x + 0) + (3x + 2) + (3x + 4)]\\ &= k\sum_{x=0}^{1}(9x + 6)\\ &= k[(0 + 6) + (9 + 6)]\\ &= 21k\end{aligned}$$

For $p(x, y)$ to be a joint probability function we must have $21k = 1$ from which

$$k = \frac{1}{21}$$

Download free eBooks at bookboon.com

(b) For the sample point $\{X = 0, Y = 0\} = (0, 0)$

$$p(0,0) = \frac{1}{21}[3(0) + 2(0)] = 0$$

Similarly,

$$\begin{aligned}
p(0,1) &= \frac{1}{21}[3(0) + 2(1)] = \frac{2}{21} \\
p(0,2) &= \frac{1}{21}[3(0) + 2(2)] = \frac{4}{21} \\
p(1,0) &= \frac{1}{21}[3(1) + 2(0)] = \frac{3}{21} \\
p(1,1) &= \frac{1}{21}[3(1) + 2(1)] = \frac{5}{21} \\
p(1,2) &= \frac{1}{21}[3(1) + 2(2)] = \frac{7}{21}
\end{aligned}$$

These results are presented in the following table. Recollect that the row and column totals are the marginal probabilities.

X	Y			Row Totals
	0	1	2	
0	0	2/21	4/21	6/21
1	3/21	5/21	7/21	15/21
Column Totals	3/21	7/21	11/21	1

1.3.2 Joint Distribution of Continuous Random Variables

Definition 1.6 JOINT PROBABILITY DENSITY FUNCTION
Let X and Y be two continuous random variables. The joint (or bivariate) probability density function $f(x, y)$ of X and Y is given by

$$P(\{x_1 \leq X \leq x_2\} \cap \{y_1 \leq Y \leq y_2\}) = \int_{y_1}^{y_2} \int_{x_1}^{x_2} f(x, y)\, dx dy$$

for two pairs (x_1, x_2) and (y_1, y_2) with $x_2 \geq x_1$, $y_2 \geq y_1$

Fig 1.2 depicts the case of the continuous bivariate random variables (X, Y) which assume all the values in the rectangle $x_1 \leq X \leq x_2$ and $y_1 \leq Y \leq y_2$.

Download free eBooks at bookboon.com

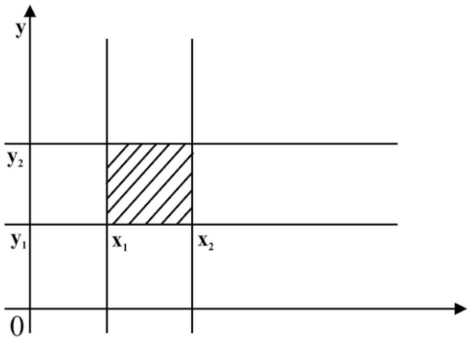


Fig. 1.2 Region assumed by (X, Y) is shaded

Thus, the probability that the event

$$\{x_1 \leq X \leq x_2\} \cap \{y_1 \leq Y \leq y_2\}$$

will fall in the shaded rectangle is

$$P(\{x_1 \leq X \leq x_2\} \cap \{y_1 \leq Y \leq y_2\})$$

This probability can be found by subtracting from the probability that the event will fall in the (semi-infinite) rectangle having the upper-right corner (x_2, y_2) the probabilities that it will fall in the semi-infinite rectangle having the upper-right corner (x_1, y_2) and (x_2, y_1) respectively, and then adding back the probability that it will fall in the semi-infinite rectangle with the upper-right corner at (x_1, y_1). That is,

Download free eBooks at bookboon.com

Click on the ad to read more

$$\begin{aligned} P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) &= P(X \leq x_2, Y \leq y_2) - P(X \leq x_2, Y \leq y_1) \\ &\quad - P(X \leq x_1, Y \leq y_2) + P(X \leq x_1, Y \leq y_1) \end{aligned}$$

For computational purposes, we shall adopt Definition 1.7.

Definition 1.7
Let (X, Y) be a continuous bivariate random variable assuming all values in the region R. The joint probability density function f is a function satisfying the following properties:

Property 1 $\quad f(x, y) \geq 0, \quad$ for all $(x, y) \in \mathcal{R}$

Property 2 $\quad \displaystyle\iint_{\mathcal{R}} f(x, y)\, dx dy = 1$

Property 2 states that the total volume bounded by the surface given by equation $z = f(x, y)$ and the region $\mathcal{R}$ on the xy-plane is equal to 1.

Example 1.3
Given the following function of a two-dimensional continuous random variable (X, Y):

$$f(x, y) = \begin{cases} x^2 + \dfrac{xy}{k}, & 0 \leq x \leq 1 \quad 0 \leq y \leq 2 \\ 0, & \text{elsewhere} \end{cases}$$

where k is a constant.

(a) Find the value of $k > 0$ such that $f(x, y)$ is a probability density function.

(b) Find $P(0 < X < 1,\ 1 < Y < 2)$.

Solution

(a) For $f(x, y)$ to be a p.d.f, it should satisfy the two conditions of Theorem 1.2. Obviously,

$$f(x, y) \geq 0$$

since $x \geq 0$, $y \geq 0$, and $k > 0$.
Also

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)\, dx\, dy = 1$$

Now,

$$\begin{aligned} \int_{y=0}^{2} \int_{x=0}^{1} \left(x^2 + \frac{xy}{k} \right) dx\, dy &= \int_{y=0}^{2} \left[\int_{x=0}^{1} \left(x^2 + \frac{xy}{k} \right) dx \right] dy \\ &= \int_{y=0}^{2} \left[\frac{x^3}{3} + \frac{x^2 y}{2k} \right]_{x=0}^{1} dy \\ &= \int_{y=0}^{2} \left(\frac{1}{3} + \frac{y}{2k} \right) dy \end{aligned}$$

$$\begin{aligned} &= \left[\frac{1}{3}y + \frac{y^2}{4k}\right]_0^2 \\ &= \frac{2}{3} + \frac{1}{k} = 1 \end{aligned}$$

giving $k = 3$.

(b) Using the value of k found in (a) we have

$$\begin{aligned} P(0 < X < 1,\, 1 < Y < 2) &= \int_{x=0}^{1}\int_{y=1}^{2}\left(x^2 + \frac{xy}{3}\right)dx\,dy \\ &= \int_{y=1}^{2}\left[\int_{x=0}^{1}\left(x^2 + \frac{xy}{3}\right)dx\right]dy \\ &= \int_{y=1}^{2}\left[\frac{x^3}{3} + \frac{x^2 y}{6}\right]_{x=0}^{1} dy \\ &= \int_{y=1}^{2}\left(\frac{1}{3} + \frac{y}{6}\right)dy \\ &= \left[\frac{y}{3} + \frac{y^2}{12}\right]_{y=1}^{2} = \frac{7}{12} \end{aligned}$$

1.4 JOINT CUMULATIVE DISTRIBUTION FUNCTIONS

1.4.1 Definition of Joint Bivariate Distribution Function

The joint behaviour of two random variables, X and Y is determined by the joint cumulative distribution function, also called the bivariate cumulative distribution function, or simply the *joint or bivariate distribution function* of the two random variables, X and Y. The definition given in Definition 1.8 is applicable whether X and Y are discrete or continuous.

Definition 1.8 JOINT DISTRIBUTION FUNCTION
For any random variables X and Y, the joint (bivariate) cumulative distribution function $F(x, y)$, is given by

$$F(x, y) = P(\{X \le x\} \cap \{Y \le y\}) = P(X \le x,\, Y \le y)$$

1.4.2 Joint Distribution Function of Discrete Random Variables

The joint cumulative distribution function of random variables X and Y gives the probability that X takes on a value less that or equal to x_i, $i = 1, 2, \cdots, n$ and that the Y takes on a value less than or equal to y_j, $j = 1, 2, \cdots, m$.

Download free eBooks at bookboon.com

Definition 1.9 JOINT DISTRIBUTION FUNCTION **(Discrete Case)**

The joint distribution function of two discrete random variables X and Y is

$$F(x, y) = \sum_{x_i \leq x} \sum_{y_i \leq y} p(x_i, y_j)$$

Example 1.4

Refer to the table of Example 1.1. Calculate

(a) the joint probability $P(2 \leq X \leq 3,\ 1 \leq Y \leq 2)$;

(b) the joint cumulative probability $P(X \leq 1,\ Y \leq 2)$.

Solution

(a) The joint probability $P(2 \leq X \leq 3,\ 1 \leq Y \leq 2)$ is obtained as follows:

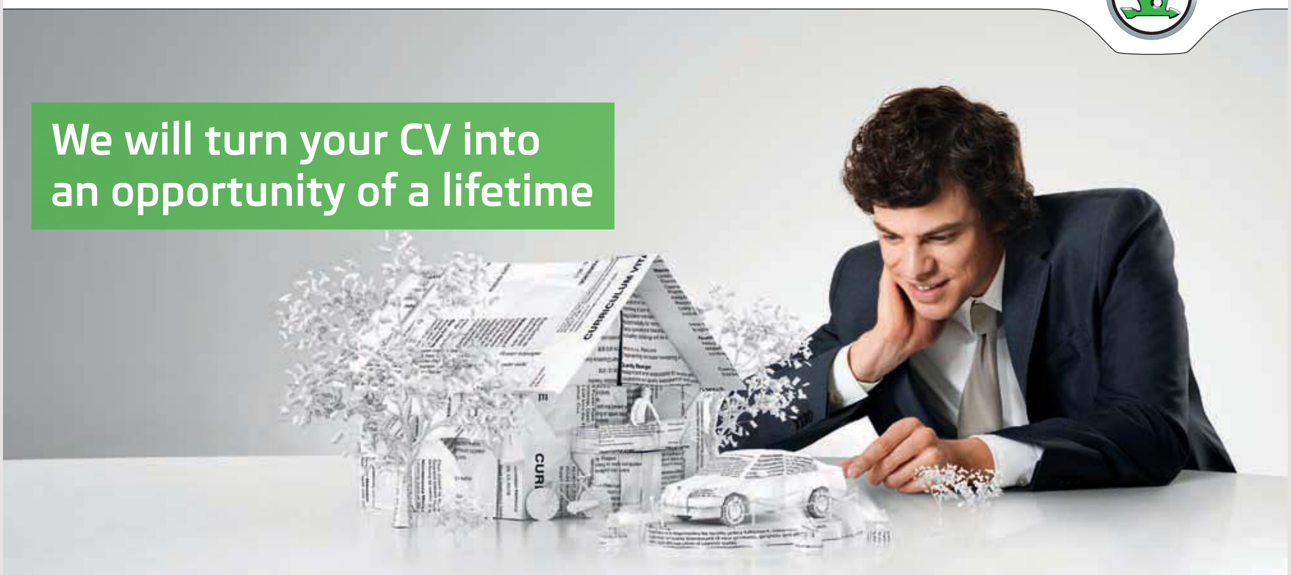
$$\begin{aligned} P(2 \leq X \leq 3,\ 1 \leq Y \leq 2) &= p(2,1) + p(2,2) + p(3,1) + p(3,2) \\ &= 0 + \frac{2}{8} + \frac{1}{8} + 0 = = \frac{3}{8} \end{aligned}$$

(b) The joint probability $P(X \leq 1,\ Y \leq 2)$ is as follows:

$$\begin{aligned} P(X \leq 1,\ Y \leq 2) &= F(1,2), \\ &= p(0,1) + p(0,2) + p(1,1) + p(1,2) \\ &= \frac{1}{8} + 0 + 0 + \frac{2}{8} = \frac{3}{8} \end{aligned}$$

SIMPLY CLEVER

ŠKODA



Do you like cars? Would you like to be a part of a successful brand?
We will appreciate and reward both your enthusiasm and talent.
Send us your CV. You will be surprised where it can take you.

Send us your CV on
www.employerforlife.com

Click on the ad to read more

Example 1.5
Refer to Example 1.2. Calculate

(a) the joint probability distribution $P(0 \leq X \leq 1,\ 1 \leq Y \leq 2)$;

(b) the joint cumulative distribution function $P(X \leq 1,\ Y \leq 1)$;

(c) the joint cumulative distribution function $P(X \leq 1,\ Y \leq 2)$.

Solution

$$\begin{aligned}
\text{(a)} \quad P(0 \leq X \leq 1,\ 1 \leq Y \leq 2) &= \sum_{x=0}^{1}\sum_{y=1}^{2}\frac{1}{21}(3x+2y)\\
&= \frac{1}{21}\sum_{x=0}^{1}[(3x+2)+(3x+4)]\\
&= \frac{1}{21}\sum_{x=0}^{1}(6x+6)\\
&= \frac{1}{21}[(0+6)+(6+6)] = \frac{6}{7}\\
\text{(b)} \qquad P(X \leq 1,\ Y \leq 1) &= F(1,1)\\
&= \sum_{x=0}^{1}\sum_{y=0}^{1}\frac{1}{21}(3x+2y)\\
&= \frac{1}{21}\sum_{x=0}^{1}[(3x+0)+(3x+2)]\\
&= \frac{1}{21}\sum_{x=0}^{1}(6x+2)\\
&= \frac{1}{21}[(0+2)+(6+2)] = \frac{10}{21}
\end{aligned}$$

The reader is asked in Exercise 1.5 to solve part (c) of this example.

1.4.3 Joint Distribution Function of Continuous Random Variables

Definition 1.10 JOINT DISTRIBUTION FUNCTION **(Continuous Case)**
The cumulative distribution function F of the two-dimensional random variable (X, Y) is defined as

$$F(x,y) = P(\{X \leq x\} \cap \{Y \leq y\}) = \int_{-\infty}^{y}\int_{-\infty}^{x} f(s,t)\,ds\,dt$$

where $f(s,t)$ is the value of the joint probability density of X and Y at (s,t)

Download free eBooks at bookboon.com

The joint distribution function gives the probability that the point X, Y belongs to a semi-infinite rectangle in the plane, as shown in Fig. 1.3.

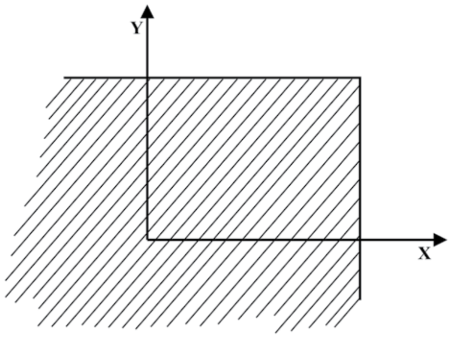


Fig. 1.3 Region of $\{X \leq x, Y \leq y\}$ is shaded

Example 1.6
Refer to Example 1.3. Calculate $P(X \leq 1,\, Y < 1)$.

Solution

$$\begin{aligned}
P(X \leq 1,\, Y < 1) &= \int_{y=0}^{1}\int_{x=0}^{1}\left(x^2 + \frac{xy}{3}\right) dx\, dy \\
&= \int_{y=0}^{1}\left[\int_{x=0}^{1}\left(x^2 + \frac{xy}{3}\right) dx\right] dy \\
&= \int_{y=0}^{1}\left[\frac{x^3}{3} + \frac{x^2 y}{6}\right]_{x=0}^{1} dy \\
&= \int_{y=0}^{1}\left(\frac{1}{3} + \frac{y}{6}\right) dy \\
&= \left[\frac{1}{3}y + \frac{y^2}{12}\right]_{y=0}^{1} \\
&= \frac{5}{12}
\end{aligned}$$

Theorem 1.1
If F is the cumulative distribution function of a two-dimensional random variable with joint probability density function $f(x, y)$ then

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

wherever F is differentiable

Download free eBooks at bookboon.com

Example 1.7
Let

$$F(x,y) = (1 - e^{-x})(1 - e^{-y}), \quad x \geq 0, \quad y \geq 0$$

Find the joint probability density function $f(x,y)$.

Solution

$$\begin{aligned}
\frac{\partial F(x,y)}{\partial x} &= e^{-x}(1 - e^{-y}) \\
\frac{\partial^2 F(x,y)}{\partial x \partial y} &= e^{-x}e^{-y} \\
&= e^{-(x+y)}, \quad x \geq 0, \quad y \geq 0.
\end{aligned}$$

Hence

$$f(x,y) = e^{-(x+y)}, \quad x \geq 0, \ y \geq 0\}$$

Note

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

Click on the ad to read more

1.4.4 Properties of Bivariate Cumulative Distribution Function

The joint c.d.f. of a bivariate random variable has properties which are analogous to those of the univariate random variable.

Property 1
The function $F(x, y)$ is a probability, hence

$$0 \leq F(x, y) \leq 1$$

Property 2
The bivariate distribution function $F(x, y)$ is monotonic increasing, in a wider sense for both variables, that is,
if $x_1 \leq x_2$, then

$$F(x_1, y) \leq F(x_2, y), \quad \text{for } y \text{ fixed}$$

if $y_1 \leq y_2$, then

$$F(x, y_1) \leq F(x, y_2), \quad \text{for } x \text{ fixed}$$

Property 3
The following relations are also true:

(a) $F(-\infty, y) = 0$

(b) $F(x, -\infty) = 0$

(c) $F(+\infty, +\infty) = 1$

(d) $F(-\infty, -\infty) = 0$

(e) $F(x, +\infty) = P(X \leq x) = F_1(x)$, where $F_1(x)$ is the c.d.f of X

(g) $P(a < X < b, \, Y \leq y) = F(b, y) - F(a, y)$

(h) $P(X \leq x, \, c < Y < d) = F(x, d) - F(x, c)$

(f) $P(a < X < b, \, c < Y < d) = F(b, d) - F(a, d) - F(b, c) + F(a, c)$

Property 4
At points of continuity of $f(x, y)$ is

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

1.5 MARGINAL DISTRIBUTION OF BIVARIATE RANDOM VARIABLES

1.5.1 Marginal Distribution of Discrete Bivariate Random Variables

The row totals of Table 1.3 provide us with the probability distribution of X. Similarly, the column totals provide the probability distribution of Y. These are typically called *marginal* probability mass functions because they are found on the margins of tables.

Download free eBooks at bookboon.com

Definition 1.11 MARGINAL BIVARIATE DISTRIBUTIONS OF X AND Y

Let X and Y be discrete random variables with joint probability function $p(x_i, y_j)$. Then the marginal distributions of X and Y respectively are given by

$$g(x_i) = P(X = x_i) = \sum_{j=1}^{m} p(x_i, y_j), \quad i = 1, 2, 3, ..., n$$

$$h(y_j) = P(Y = y_j) = \sum_{i=1}^{n} p(x_i, y_j), \quad j = 1, 2, 3, ..., m$$

Example 1.8

For the data in Table 1.2, find the marginal probability distribution for (a) X and (b) Y.

Solution

To calculate the marginal probabilities the joint probabilities are required. The joint probabilities for this problem have been calculated in Example 1.1.

(a) From the table obtained in Example 1.1. we shall calculate the marginal probabilities for each x_i by fixing i and summing all the joint probabilities across j. Thus:

$$\begin{aligned} P(X=0) &= P(X=0, Y=1) + P(X=0, Y=2) \\ &\quad + P(X=0, Y=3) \\ &= p(0,1) + p(0,2) + p(0,3) = \frac{1}{8} + 0 + 0 = \frac{1}{8} \end{aligned}$$

$$\begin{aligned} P(X=1) &= P(X=1, Y=1) + P(X=1, Y=2) \\ &\quad + P(X=1, Y=3) \\ &= p(1,1) + p(1,2) + p(1,3) = 0 + \frac{2}{8} + \frac{1}{8} = \frac{3}{8} \end{aligned}$$

$$\begin{aligned} P(X=2) &= P(X=2, Y=1) + P(X=2, Y=2) \\ &\quad + P(X=2, Y=3) \\ &= p(2,1) + p(2,2) + p(2,3) = 0 + \frac{2}{8} + \frac{1}{8} = \frac{3}{8} \end{aligned}$$

$$\begin{aligned} P(X=3) &= P(X=3, Y=1) + P(X=3, Y=2) \\ &\quad + P(X=3, Y=3) \\ &= p(3,1) + p(3,2) + p(3,3) = \frac{1}{8} + 0 + 0 = \frac{1}{8} \end{aligned}$$

The results are summarised in the table below:

x_i	0	1	2	3
$g(x_i)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Download free eBooks at bookboon.com

(b) Similar to (a) we fix j and sum all the joint probabilities across i. Hence,

$$\begin{aligned}
P(Y=1) &= P(Y=1,\ X=0)+P(Y=1,\ X=1) \\
&\quad +P(Y=1,\ X=2)+P(Y=1,\ X=3) \\
&= p(1,0)+p(1,1)+p(1,2)+p(1,3) \\
&= \frac{1}{8}+0+0+0=\frac{1}{8}
\end{aligned}$$

$$\begin{aligned}
P(Y=2) &= P(Y=2,\ X=0)+P(Y=2,\ X=1)+ \\
&\quad P(Y=2,X=2)+P(Y=2,X=3)+ \\
&= p(2,0)+p(2,1)+p(2,2)+p(2,3) \\
&= 0+\frac{2}{8}+\frac{2}{8}+0=\frac{4}{8}
\end{aligned}$$

$$\begin{aligned}
P(Y=3) &= P(Y=3,\ X=0)+P(Y=3,\ X=1)+ \\
&\quad P(Y=3,X=2)+P(Y=3,X=3) \\
&= p(3,0)+p(3,1)+p(3,2)+p(3,3) \\
&= 0+\frac{1}{8}+\frac{1}{8}+0=\frac{2}{8}
\end{aligned}$$

Click on the ad to read more

The results are summarised in the table below.

y_j	1	2	3
$g(y_j)$	$\frac{2}{8}$	$\frac{4}{8}$	$\frac{2}{8}$

The results of Examples 1.1 and 1.8 (that is, the joint and marginal probabilities) are usually presented in a single table such as in Table 1.4.

Note

(a) The marginal distributions of X and Y are the ordinary probability distribution functions of X and Y but when derived from the joint distribution function the adjective "marginal" is added.

(b) In marginal distribution, the probability of different values of a random variable in a subset of random variables is determined without reference to any possible values of the other variables.

(c) From the table,

$$\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) = \sum_{j=1}^{m} g(x_i, y_j) = \sum_{i=1}^{n} h(x_i, y_j) = 1$$

that is, the total probability is 1.

Table 1.4 Joint and Marginal Probabilities for Table 1.2

X	Y			Row Totals
	1	2	3	
0	1/8	0	0	1/8
1	0	2/8	1/8	3/8
2	0	2/8	1/8	3/8
3	1/8	0	0	1/8
Column Totals	2/8	4/8	2/8	1

Computation of Marginal Probabilities of Discrete Bivariate Random Variables from Formula

We may also compute marginal distributions from a formula.

Example 1.9
Refer to Example 1.2. Find the marginal distributions of
(a) X, (b) Y.

Download free eBooks at bookboon.com

Solution

(a) The marginal distribution of X is given by:

$$\begin{aligned} g(x) &= \sum_{y=0}^{2} \frac{1}{21}(3x+2y) \\ &= \frac{1}{21}\left[\sum_{y=0}^{2}(3x+2y)\right] \\ &= \frac{1}{21}[(3x+0)+(3x+2)+(3x+4)] \\ &= \frac{1}{7}(3x+2) \end{aligned}$$

(b) The marginal distribution of Y is given by:

$$\begin{aligned} h(y) &= \sum_{x=0}^{1} \frac{1}{21}(3x+2y) \\ &= \frac{1}{21}\sum_{x=0}^{1}(3x+2y) \\ &= \frac{1}{21}[(0+2y)+(3+2y)] \\ &= \frac{1}{21}(3+4y) \end{aligned}$$

1.5.2 Marginal Distribution of Continuous Bivariate Random Variables

Definition 1.12
Suppose $f(x,y)$ be the joint probability density function of the continuous two-dimensional random variable (X,Y). We define $g(x)$ and $h(y)$, the marginal probability density function of X and Y, respectively by

$$g(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$

and

$$h(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$

Example 1.10
Refer to Example 1.3. Find the marginal probability distribution of X and Y.

Solution
Marginal probability distribution of X:

$$\begin{aligned} g(x) &= \int_{-\infty}^{\infty} f(x,y)\,dy \\ &= \int_{0}^{2}\left(x^2+\frac{xy}{3}\right)dy, \qquad 0<x<1 \end{aligned}$$

Download free eBooks at bookboon.com

$$
\begin{aligned}
&= \left[x^2 y + \frac{xy^2}{6}\right]_0^2 \\
&= 2x^2 + \frac{4x}{6}
\end{aligned}
$$

That is,

$$
g(x) = \begin{cases} 2x^2 + \frac{2}{3}x, & 0 < x < 1 \\ 0, & \text{elsewhere} \end{cases}
$$

Marginal probability distribution of Y:

$$
\begin{aligned}
h(y) &= \int_{-\infty}^{\infty} f(x,y)\,dx \\
&= \int_0^1 \left(x^2 + \frac{xy}{3}\right) dx, \quad 0 < y < 2 \\
&= \left[\frac{x^3}{3} + \frac{x^2 y}{6}\right]_0^1 \\
&= \frac{1}{3} + \frac{y}{6}
\end{aligned}
$$

That is

$$
h(y) = \begin{cases} \frac{1}{3} + \frac{y}{6}, & 0 < y < 2 \\ 0, & \text{elsewhere} \end{cases}
$$

Download free eBooks at bookboon.com
Click on the ad to read more

Definition 1.13 MARGINAL CUMULATIVE DISTRIBUTION
The marginal probability distribution function of X, denoted by $F_X(x)$ is

$$F_X(x) = P(X \leq x)$$

and the marginal probability distribution function of Y, denoted by $F_Y(y)$ is

$$F_Y(y) = P(Y \leq y)$$

Thus,

$$\begin{aligned} F(x, y) &= P(X \leq x, Y \leq y) \\ &= \lim_{y \to \infty} F(x, y) \\ &= F_X(x) \\ &= \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(u, y)\, dy\, du \end{aligned}$$

Similarly,

$$\begin{aligned} F(x, y) &= \lim_{x \to \infty} F(x, y) \\ &= F_Y(y) \\ &= \int_{-\infty}^{y} \int_{-\infty}^{\infty} f(x, v)\, dx\, dv \end{aligned}$$

From this, it follows that the probability density function of X alone, known as the *marginal density* of X, is

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

Similarly

$$\begin{aligned} h(y) &= f_Y(y) \\ &= F'_Y(y) \\ &= \int_{-\infty}^{\infty} f(x, y)\, dx \end{aligned}$$

Note
The marginal probability density functions $g(x)$ and $h(y)$ can easily be determined from the knowledge of the joint density function $f(x, y)$. However, the knowledge of the marginal probability density functions does not, in general, uniquely determine the joint density function. The exception occurs when the two random variables are independent.

Download free eBooks at bookboon.com

1.6 CONDITIONAL DISTRIBUTION OF BIVARIATE RANDOM VARIABLES

1.6.1 Conditional Distribution of Discrete Random Variables

The conditional probability distribution of a random variable is analogous to the concept of conditional probability of events. For two events $\mathcal{A}$ and $\mathcal{B}$, the multiplication law gives the probability of the intersection $\mathcal{A} \cap \mathcal{B}$ as

$$P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A})P(\mathcal{B}|\mathcal{A})$$

where $P(\mathcal{A})$ is the unconditional probability of $\mathcal{A}$ and $P(\mathcal{B}|\mathcal{A})$ is the conditional probability of $\mathcal{B}$ given that $\mathcal{A}$ has occurred.

Now, consider $(\{X = x\}) \cap (\{Y = y\})$, represented by the bivariate event (x, y). Then it follows directly from the multiplication law of probability that the bivariate probability for the intersection (x_i, y_j) is

$$p(x_i, y_j) = g(x_i)p(y_j|x_i)$$

or

$$p(x_i, y_j) = h(y_j)p(x_i|y_j)$$

The probabilities $g(x_i)$ and $h(y_j)$ are those associated with the marginal probability distributions for X and Y, respectively. The probability $p(x|y)$ is the probability that the random variable X takes a specific value x given that Y takes on the value y written in full as $P(X = x|Y = y)$. Thus, $P(X = 2|Y = 1)$ is the conditional probability that $X = 2$ given that $Y = 1$. A similar interpretation can be attached to the conditional probability $p(y|x)$.

Note
Conditional distribution is the opposite of marginal distribution, in which the probability of a value of a random variable is determined without reference to the possible values of the other variables.

Definition 1.14 BIVARIATE CONDITIONAL PROBABILITY DISTRIBUTION **(Discrete Case)**
Suppose X and Y are discrete random variables with joint probability distribution $p(x, y)$ then the conditional discrete probability function of X given Y is

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}, \quad P(Y = y) > 0$$

Download free eBooks at bookboon.com

That is, given the joint probability distribution $p(x, y)$ and marginal probability functions $g(x)$ and $h(y)$, respectively, the conditional discrete probability function of X given Y is

$$p(x|y) = \frac{p(x, y)}{h(y)}, \quad h(y) > 0$$

and similarly, the conditional discrete probability function of Y given X is

$$p(y|x) = \frac{p(x, y)}{g(x)}, \quad g(x) > 0$$

This definition shows that if we have the joint probability function of two random variables and desire the conditional distribution for one of them when the other is held fixed, it is merely necessary to divide the joint probability function by the probability function of the fixed variable.

Example 1.11
Refer to the table in Example 1.1. Find

(a) $P(Y = 2|X = 1)$

(b) $P(X = 1|Y = 2)$

(c) $P(X = 1|Y = 1)$

Click on the ad to read more

Solution

(a) $$P(Y=2|X=1) = \frac{P(X=1,\,Y=2)}{P(X=1)} = \frac{p(1,2)}{g(1)} = \frac{\frac{2}{8}}{\frac{3}{8}} = \frac{2}{3}$$

(b) $$P(X=1|Y=2) = \frac{P(X=1,\,Y=2)}{P(Y=2)} = \frac{p(1,2)}{h(2)} = \frac{\frac{2}{8}}{\frac{4}{8}} = \frac{2}{4}$$

(c) $$P(X=1|Y=1) = \frac{P(X=1,\,Y=1)}{P(Y=1)} = \frac{p(1,1)}{h(1)} = \frac{0}{\frac{2}{8}} = 0$$

Example 1.12
Refer to Example 1.2. Find the conditional probability function
(a) $p(x|y)$, (b) $p(y|x)$

Solution

$$p(x|y) = \frac{p(x,y)}{h(y)}$$

(a) From Example 1.2,

$$p(x,y) = \frac{1}{21}(3x+2y)$$

From Example 1.9,

$$h(y) = \frac{1}{21}(3+4y)$$

Hence

$$\begin{aligned} p(x|y) &= \frac{\frac{1}{21}(3x+2y)}{\frac{1}{21}(3+4y)} \\ &= \frac{3x+2y}{(3+4y)} \end{aligned}$$

(b)

$$p(y|x) = \frac{p(x,y)}{g(x)}$$

From Example 1.9,

$$g(x) = \frac{1}{7}(3x+2)$$

Hence

$$\begin{aligned} p(y|x) &= \frac{\frac{1}{21}(3x+2y)}{\frac{1}{7}(3x+2)} \\ &= \frac{7(3x+2y)}{21(3x+2)} \\ &= \frac{3x+2y}{3(3x+2)} \end{aligned}$$

Download free eBooks at bookboon.com

1.6.2 Conditional Distribution of Continuous Variables

Definition 1.15
Let (X, Y) be a continuous two-dimensional random variable with joint probability density function $f(x, y)$. Let g and h be the marginal probability density functions of X and Y respectively. The conditional probability of X for given $Y = y$ is defined by

$$f(x|y) = \frac{f(x,y)}{h(y)}, \qquad h(y) > 0$$

and the conditional probability of Y for given $X = x$ is defined by

$$f(y|x) = \frac{f(x,y)}{g(x)}, \qquad g(x) > 0$$

Example 1.13
Refer to Example 1.3. Find (a) $f(x|y)$, (b) $f(y|x)$.

Solution
From Example 1.10

$$h(y) = \frac{1}{3} + \frac{y}{6}$$

(a)

$$f(x|y) = \frac{f(x,y)}{h(y)}$$

Therefore

$$\begin{aligned} f(x|y) &= \frac{x^2 + \frac{xy}{3}}{\frac{1}{3} + \frac{y}{6}} \\ &= \frac{6x^2 + 2xy}{2 + y}, \quad 0 \le x \le 1; \quad 0 \le y \le 2 \end{aligned}$$

(b) The conditional probability of Y for given $X = x$ is

$$f(y|x) = \frac{f(x,y)}{g(x)}.$$

From Example 1.10,

$$g(x) = 2x^2 + \frac{2}{3}\,x$$

Therefore

$$\begin{aligned} f(y|x) &= \frac{x^2 + \frac{xy}{3}}{2x^2 + \frac{2}{3}\,x} \\ &= \frac{3x^2 + xy}{6x^2 + 2x} \\ &= \frac{3x + y}{6x + 2}, \quad 0 \le y \le 2; \quad 0 \le x \le 1 \end{aligned}$$

Download free eBooks at bookboon.com

Example 1.14
Consider the joint probability density function

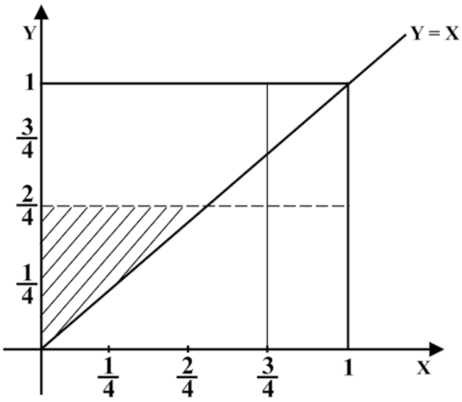
$$f(x,y) = \begin{cases} 6(1-y), & 0 \le x \le y \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

Find $P\left(Y < \frac{1}{2} \middle| X < \frac{3}{4}\right)$.

Solution

$$P\left(Y < \frac{1}{2} \middle| X < \frac{3}{4}\right) = \frac{P\left(Y < \frac{1}{2}, X < \frac{3}{4}\right)}{P\left(X < \frac{3}{4}\right)}$$

We first evaluate the numerator. The region of integration is sketched below.



Click on the ad to read more

Now,

$$\begin{aligned} P\left(Y < \frac{1}{2}, X < \frac{3}{4}\right) &= \int_{y=0}^{\frac{1}{2}} \int_{x=0}^{y} 6(1-y)\,dx\,dy \\ &= \left[3\,y^2 - 2\,y_3\right]_{y=0}^{\frac{1}{2}} \\ &= \frac{1}{2} \end{aligned}$$

The denominator is evaluated as follows. First we find the marginal p.d.f of X.

$$g(x) = \int_{y=x}^{1} 6(1-y)\,dy = 3(1-x)^2 \quad 0 \le x \le 1$$

so that

$$\begin{aligned} P\left(X < \frac{3}{4}\right) &= \int_{x=0}^{\frac{3}{4}} 3(1-x)^2\,dx \\ &= \frac{63}{64} \end{aligned}$$

Hence,

$$P\left(Y < \frac{1}{2} \middle| X < \frac{3}{4}\right) = \frac{1/2}{63/64} = \frac{32}{63}$$

1.7 INDEPENDENCE OF BIVARIATE RANDOM VARIABLES

1.7.1 Definition of Independence of Bivariate Random Variables

We recall from the introductory probability course that, in terms of event probabilities, two events $\mathcal{A}$ and $\mathcal{B}$ are independent if the realization of $\mathcal{A}$ is not affected by the occurrence of $\mathcal{B}$. That is, two events $\mathcal{A}$ and $\mathcal{B}$ are independent if

$$P(\mathcal{A}|\mathcal{B}) = P(\mathcal{A})$$

This condition can be carried over to two random variables. That is, two random variables are independent if the realization of one does not affect the probability distribution of the other.

We also recall that the definition of independence of two events are usually given as

$$P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A})P(\mathcal{B})$$

Similarly, we may carry this condition of independence over to two random variables as in Definition 1.16.

Download free eBooks at bookboon.com

Definition 1.16 INDEPENDENCE OF BIVARIATE RANDOM VARIABLES
Let X have distribution function $F_1(x), Y$ have distribution function $F_2(y)$ with X and Y having joint distribution function $F(x, y)$. Then X and Y are said to be independent if and only if

$$F(x, y) = F_1(x)F_2(y)$$

for every pair of real numbers (x, y)

Thus, the independence of the random variables X and Y implies that their joint distribution function factors into the products of their marginal distribution functions. This definition applies whether the random variables are discrete or continuous.

If X and Y are not independent, they are said to be ***dependent***. It is usually more convenient to verify independence or otherwise with the help of the p.m.f. (in the discrete case) or p.d.f. (in the continuous case).

1.7.2 Independence of Bivariate Discrete Random Variables

Definition 1.17
If X and Y are discrete random variables with joint probability function $p(x, y)$ and marginal probability function $g(x)$ and $h(y)$ respectively, then X and Y are independent if and only if

$$p(x, y) = g(x)h(y)$$

for all pairs of real numbers (x, y)

Example 1.15
Refer to the table in Example 1.1, verify whether or not X and Y are independent.

Solution
Consider the ordered pair $(0, 1)$.

From Example 1.1

$$P(X = 0, Y = 1) = \frac{1}{8}$$

But from the marginal distributions,

$$\begin{aligned} P(X = 0) &= \frac{1}{8} \\ P(Y = 1) &= \frac{2}{8} \\ P(X = 0)P(Y = 1) &= \left(\frac{1}{8}\right)\left(\frac{2}{8}\right) \end{aligned}$$

which does not equal $\frac{1}{8}$. Therefore X and Y are not independent.

Download free eBooks at bookboon.com

Example 1.16
Refer to Example 1.2. Are X and Y independent?

Solution
From Example 1.2

$$p(x,y) = \frac{1}{21}(3x+2y)$$

From Example 1.9,

$$\begin{aligned} g(x) &= \frac{1}{7}(3x+2) \\ h(y) &= \frac{1}{21}(3+4y) \end{aligned}$$

Now

$$\begin{aligned} g(x)h(y) &= \frac{1}{7}(3x+2)\frac{1}{21}(3+4y) \\ &= \frac{1}{147}(9x+12xy+8y+6) \end{aligned}$$

which is not equal to $p(x,y)$; hence X and Y are not independent.

Click on the ad to read more

1.7.3 Independence of Continuous Random Variables

Like the discrete case, we usually verify for independence or lack of it of continuous random variables using the result contained in Definition 1.16.

Definition 1.18
If X and Y are are continuous random variables with a joint density function of $f(x,y)$ and marginal density functions of $g(x)$ and $h(y)$, respectively, then X and Y are independent if and only if

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

for all pairs of real numbers (x,y)

Example 1.17
Refer to Example 1.3. Verify whether X and Y are independent.

Solution
From Example 1.10,

$$\begin{aligned} g(x) &= 2x^2 + \frac{2}{3}\,x \\ h(y) &= \frac{1}{3} + \frac{y}{6} \end{aligned}$$

Now

$$\begin{aligned} g(x)h(y) &= \left(2x^2 + \frac{2}{3}\,x\right)\left(\frac{1}{3} + \frac{y}{6}\right) \\ &= \frac{2}{3}x^2 + \frac{1}{3}x^2 y + \frac{2}{9}x + \frac{1}{9}\,xy \end{aligned}$$

which is not equal to

$$f(x,y) = x^2 + \frac{xy}{3}$$

Hence X and Y are not independent.

Example 1.18
Suppose the joint p.d.f. of X and Y is given by

$$f(x,y) = \begin{cases} 4xy, & 0 < x < 1,\ 0 < y < 1 \\ 0, & \text{elsewhere} \end{cases}$$

Verify whether or not X and Y are independent.

Solution
We have

$$\begin{aligned} g(x) &= \int_0^1 f(x,y)\,dy \\ &= \int_0^1 4xy\,dy \\ &= \left[4x\frac{y^2}{2}\right]_0^1 \\ &= 2x, \quad 0 < x < 1 \end{aligned}$$

Download free eBooks at bookboon.com

Similarly

$$\begin{aligned} h(y) &= \int_0^1 f(x,y)dx \\ &= 2y, \quad 0 < y < 1 \end{aligned}$$

Hence,

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

for all real numbers (x, y), and therefore X and Y are independent.

To conclude this chapter we shall make two important observations.

(a) *Multivariate Distributions*

Our discussion of the bivariate case can be readily extended to multivariate distributions. Apart from added computational labour, joint distributions of three or more variables do not pose new analytical problems. Thus, we simply state here that the joint probability mass function of k discrete random variables $p(x_1, x_2, ..., x_k)$ is specified to be nonnegative and its sum over the k dimensional space is one. There are marginal density functions of various dimensions in that case. Suppose X, Y, and Z are jointly continuous random variables with density function $f(x, y, z)$, The one dimensional marginal distribution of X is

$$f_X(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y,z)\,dy\,dz$$

and the two-dimensional marginal distribution of X and Y is

$$f_{XY}(x,y) = \int_{-\infty}^{\infty} f(x,y,z)\,dz$$

(b) *Algebra of Random Variables*

The four basic operations of algebra, namely, addition, subtraction, multiplication and division can be performed on random variables to form new random variables. That is, if X and Y are random variables and c is a real number, then we can form new variables $X + Y,\ X - Y,\ XY,\ \dfrac{X}{Y},\ c + X,\ cX$. We can also form the random variable $f(X)$, where f is any function on $\mathcal{R}$. That is, if X has range $\{x_1,\ x_2,\ \cdots,\ x_n\}$, then $f(X)$ has range $\{f(x_1),\ f(x_2),\ \cdots,\ f(x_n)\}$. For example, if X is a random variable taking the values $\{1,\ 2,\ 3,\ 4\}$ with probabilities $\frac{1}{2}, \frac{1}{8}, \frac{1}{4}$, and $\frac{1}{8}$, respectively, then X^2 takes the values $\{1,\ 4,\ 9,\ 16\}$ with probabilities $\frac{1}{2}, \frac{1}{8}, \frac{1}{4}$, and $\frac{1}{8}$, respectively.

In general, the probability distributions of cX (that is, adding a random variable to itself c times) and $f(X)$ (if f is one-to-one) such as multiplying a random variable by itself, or finding the root or cosine of a random variable are the same as that of X. However, the determination of the probability distributions of $X \pm Y, XY$, and $\dfrac{X}{Y}$ $(Y \neq 0)$ is not straightforward and this is the subject of the next chapter.

Download free eBooks at bookboon.com

EXERCISES

1.1 Given the following joint frequency distribution of X and Y

X	Y		
	1	2	3
1	3	5	1
2	2	4	6
3	1	2	0
4	4	1	5

Calculate the joint probability of X and Y.

1.2 Refer to Exercise 1.1.

(a) Find the marginal probability distribution of
(i) X, (ii) Y

(b) Verify the independence of X and Y for the data of Exercise 1.1.

1.3 Refer to Exercise 1.1. Calculate

(a) $P(X < 2, Y \leq 3)$, (b) $P(2 \leq X < 3, 1 \leq Y \leq 2)$

(c) $P(1 < X < 2, Y < 3)$, (d) $P(X < 2, 0 \leq Y < 2)$

Click on the ad to read more

1.4 Refer to Exercise 1.1. Find

(a) $P(Y = 3|X = 1)$, (b) $P(Y = 2|X = 3)$,
(c) $P(X = 2|Y = 2)$, (d) $P(X = 1|Y = 2)$,
(e) $P(X = 3|Y = 1)$, (f) $P(X = 3|Y = 2)$

1.5 Refer to Example 1.5. Solve part (c).

1.6 The joint probability mass function of X and Y is given by

$$p(1,1) = \frac{1}{8}, \quad p(1,2) = \frac{1}{4}, \quad p(2,1) = \frac{1}{8}, \quad p(2,2) = \frac{1}{2}$$

(a) Compute the conditional probability mass function of X, given $Y = i,\ i = 1, 2$.

(b) Are X and Y independent?

1.7 Consider the bivariate discrete random variables X and Y with probability function

$$p(x,y) = \frac{1}{27}(2x + y) \qquad x = 0, 1, 2; \quad y = 0, 1, 2.$$

(a) Verify that $p(x, y)$ is a legitimate probability mass function.

(b) Find the joint probability

$$P(1 \le X \le 2,\, 0 \le Y \le 1)$$

(c) Find the marginal distributions of X and Y.

(d) Find the conditional probability of

(i) Y for given $X = x$

(ii) X for given $Y = y$

(e) Verify whether X and Y are independent or not.

1.8 A bivariate discrete random variable (X, Y) is such that

$$P(X = x,\, Y = y) = \theta^x(1 - \theta)^y,$$

where $0 < \theta < 1$ and $x, y = 1, 2, 3, ...$

(a) Show that the above probability mass function is legitimate

(b) Find the marginal probability distribution of X and Y

(c) Find the conditional probability of
(i) X given $Y = 2$, (ii) Y given $X = 0$

(d) Verify whether X and Y are independent or not.

Download free eBooks at bookboon.com

1.9 A bivariate discrete random variable (X, Y) has the following probability mass function.

$$p(x, y) = \frac{\binom{x+y}{x} p^x (1-p)^y e^{-\lambda} \lambda^{x+y}}{(x+y)!}, \qquad x, y = 0, 1, 2, ...$$

(a) Show that the marginal probability function of X is

$$g(x) = \frac{e^{-\lambda p}(\lambda p)^x}{x!}; \qquad x = 0, 1, 2, ...$$

(b) Find the marginal probability mass function of Y.

(c) Examine whether X and Y are independent.

1.10 Consider the following function of two discrete random variables

$$f(x, y) = \frac{1}{20}(x^2 - y), \qquad for \;\; x = 2, 3; \quad y = 1, 2$$

(a) Find the value of k such that the function is a legitimate probability mass function.

(b) Construct the joint probability table.

(c) Find the marginal probability distributions of X and Y.

(d) Find the conditional probability of

(i) Y for given $X = x$

(ii) X for given $Y = y$

(e) Verify whether or not X and Y are independent.

1.11 Suppose that (X, Y) is a two-dimensional continuous random variable with joint probability density function.

$$f(x, y) = \begin{cases} k(x + y - 2xy), & 0 \le x \le 1, 0 \le y \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

where k is a constant.

(a) Find the value of k.

(b) Find the marginal probability distribution of X and Y.

(c) Find the conditional probability

(i) X given $Y = y$

(ii) Y given $X = x$

(d) Verify whether X and Y are independent or not.

(e) $P(X > Y)$

Download free eBooks at bookboon.com

1.12 A bivariate continuous random variable (X, Y) has joint probability density function

$$f(x,y) = \begin{cases} k(x+2y), & 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0, & \text{elsewhere} \end{cases}$$

where k is a constant.

(a) Find the value of k.

(b) Find the marginal probability distribution of X and Y.

(c) Find the conditional probability of

(i) Y given $X = x$

(ii) X given $Y = y$

(d) Verify whether or not X and Y are independent

1.13 Suppose that a bivariate function is given by

$$f(x,y) = \begin{cases} f(x,y) = a(x^2+xy), & 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0, & \text{elsewhere} \end{cases}$$

Find

(a) the constant a such that $f(x,y)$ is a p.d.f.

(b) $P(X > Y)$;

(c) the marginal p.d.f. of X;

(d) the marginal p.d.f. of Y.

1.14 The joint probability density function of X and Y is given by

$$f(x,y) = \begin{cases} \frac{6}{7}(x^2 + \frac{xy}{2}), & 0 < x < 1, \quad 0 < y < 2 \\ 0, & \text{elsewhere} \end{cases}$$

(a) Verify that this is indeed a joint p.d.f.

(b) Compute the distribution function of X.

(c) Find $P(X > Y)$

(d) Find $P\left(Y < \frac{1}{2} \middle| X > \frac{1}{2}\right)$.

(e) Verify whether X and Y are independent.

1.15 Consider the following joint density function

$$f(x,y) = \lambda^2 e^{\lambda y}, \qquad 0 \leq x \leq y$$

Find

(a) the marginal p.d.f. of X

(b) the marginal p.d.f. of Y;

(c) the conditional p.d.f. of Y given $X = x$;

(d) the conditional p.d.f. of X given $Y = y$

Download free eBooks at bookboon.com

1.16 The number of people that enter a church in a given hour is a Poisson random variable with $\lambda = 10$. Compute the conditional probability that at most 2 men entered the church in a given hour, given that 10 women entered in that hour. What assumptions have you made?

1.17 The joint probability density function of X and Y is given by

$$f(x,y) = \begin{cases} xe^{-(x+y)}, & x > 0, \quad y > 0 \\ 0, & \text{elsewhere} \end{cases}$$

(a) Find the marginal p.d.f of (i) X (ii) Y

(b) Find the conditional p.d.f of

(i) X, given $Y = y$;

(ii) Y given $X = x$.

(c) Verify whether X and Y are independent.

1.18 The joint probability density function of X and Y is given by

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

(a) Find the marginal p.d.f of (i) X, (ii) Y.

(b) Find the conditional p.d.f of
(i) X, given $Y = y$, (ii) Y given $X = x$

1.19 The joint probability density function of X and Y is given by

$$f(x,y) = \lambda_1 \lambda_2 \, exp\{-(\lambda_1 x + \lambda_2 y)\}, \qquad 0 < x < \infty, \; 0 < y < \infty$$

(a) Find the marginal p.d.f. of (i) X (ii) Y.

(b) Verify whether X and Y are independent.

1.20 Show that the conditional densities $f(x|y)$ and $f(y|x)$ are legitimate density functions.

Download free eBooks at bookboon.com

Chapter 2

SUMS, DIFFERENCES, PRODUCTS AND QUOTIENTS OF BIVARIATE DISTRIBUTIONS

2.1 INTRODUCTION

In the previous chapter, we discussed the joint probability distributions, cumulative distributions and marginal distributions of random variables X and Y. We also discussed the independence of bivariate random variables.

In this chapter, we shall take up the distribution of sums, differences, products and quotients of X and Y. The numerical characterisation of the joint distribution of X and Y will be discussed in the next chapter. As has been the practice in this book, discussions will be done separately for discrete and continuous cases.

Download free eBooks at bookboon.com
Click on the ad to read more

2.2 SUMS OF BIVARIATE RANDOM VARIABLES

2.2.1 Sums of Discrete Bivariate Random Variables

We shall introduce the concept of the sum of discrete random variables with an example.

Example 2.1
Consider a sequence of independent experiments in each of which a fixed event is of interest. Suppose $X_1, X_2, \cdots, X_n$ are discrete random variables defined by

$$X_i = \begin{cases} 1, & \text{if the fixed event happens in the } i^{th} \text{ experiment;} \\ 0, & \text{if the fixed event does not happen in the } i^{th} \text{ experiment.} \end{cases}$$

Let

$$Z_n = X_1 + X_2 + \cdots + X_n$$

The random variable Z_n is a sum of discrete random variables X_i, $i = 1, 2, ..., n$, and denotes the number of times the fixed event happens in the n trials.

For instance, the President of Regent University receives visitors coming every day from the two surrounding communities. The number of visitors on a particular day from the communities are X_1 and X_2, respectively. Let $Z_2 = X_1 + X_2$. The random variable Z_2 is a sum of two random variables and gives the total number of visitors coming to the President of Regent University on a particular day.

Distribution of Sums of Discrete Bivariate Random Variables in Tabular Form

Case 1
Suppose that the joint distribution of X and Y is in the form of the table as presented in Table 1.3. Such a table can be referred to as the ***parent joint probability distribution*** table. Then the various sums of the discrete random variables X and Y, denoted as $X + Y$, and their corresponding probabilities may be presented as in Table 2.1. This can be referred to as the ***derived joint probability distribution*** table.

Table 2.1 Various Sums $(X + Y)$ and their Corresponding Probabilities

$x_i + y_j$	$x_1 + y_1$	$x_1 + y_2$	$\cdots$	$x_1 + y_m$	$x_2 + y_1$	$x_2 + y_2$	$\cdots$	$x_2 + y_m$	$x_n + y_1$	$x_n + y_2$	$\cdots$	$x_n + y_m$
$p(x_i, y_j)$	$p(x_1, y_1)$	$p(x_1, y_2)$	$\cdots$	$p(x_1, y_m)$	$p(x_2, y_1)$	$p(x_2, y_2)$	$\cdots$	$p(x_2, y_m)$	$p(x_n, y_1)$	$p(x_n, y_2)$	$\cdots$	$p(x_n, y_m)$

To find the distribution of the sum, we apply the principle of the *probabilities of equivalent events*.[1] By this principle, the probabilities of equivalent events are equal. That is, if $\mathcal{A} \subset \mathcal{S}$ and $\mathcal{B} \subset R_X$ are equivalent events, then we define the probability of the event $\mathcal{B}$, $P(\mathcal{B})$, to be equal to $P(\mathcal{A})$.

This principle is illustrated in some examples below after the following theorem.

Theorem 2.1
The distribution (probability) of the sum of $X = x_i$ and $Y = y_j$ being equal to k is the sum of the probabilities that correspond to all indices i and j that sum to k

Example 2.2
Given the table below, find the distribution of the sum X+Y.

	Y			
X	-1	0	1	Row Totals
-1	0.10	0.20	0.11	0.41
0	0.08	0.02	0.26	0.36
2	0.03	0.17	0.03	0.23
Column Totals	0.21	0.39	0.40	1.00

Solution
We shall list the values of $x + y$ in each cell within parenthesis as in the table below:

	Y			
X	-1	0	1	Row Totals
-1	0.10 (-2)	0.20 (-1)	0.11 (0)	0.41
0	0.08 (-1)	0.02 (0)	0.26 (1)	0.36
2	0.03 (1)	0.17 (2)	0.03 (3)	0.23
Column Totals	0.21	0.39	0.40	1.00

[1]The events $\mathcal{A} \subset \mathcal{S}$ and $\mathcal{B} \subset R_X$ are called equivalent events if

$$\mathcal{A} = \{s \in \mathcal{S} | X(s) \in \mathcal{B}\}$$

where R_X is the range space of the random variable X.

Download free eBooks at bookboon.com

The values for $x+y$, also known as the ***support of*** $x+y$, are $\{-2,-1,0,1,2,3\}$ and the probability of each value[2] $P[(X+Y)=x+y]$ is the sum of all the cell probabilities where that value occurs. For example, there are two (mutually exclusive) ways that the sum $X+Y$ can equal 0, and that is, if $X=0$ and $Y=0$ or if $X=-1$ and $Y=1$, so that the probability of the sum $X+Y=0$ is $P(X=0,\, Y=0)$ or $P(X=-1,\, Y=1)$. That is,

$$\begin{aligned} P[(X+Y)=0] &= P(X=0,\, Y=0)+P(X=-1\, Y=1) \\ &= 0.11+0.02 \\ &= 0.13=p(0) \end{aligned}$$

Again, the only way $X+Y$ can equal 2 is when $X=2$ and $Y=0$ so that the probability of the sum $X+Y=2$ is

$$\begin{aligned} P[(X+Y)=2] &= P(X=2,\, Y=0) \\ &= 0.17=p(2) \end{aligned}$$

Thus, we can summarise the table as follows:

$(x+y)$	-2	-1	0	1	2	3
$p(x+y)$	0.1	0.28	0.13	0.29	0.17	0.03

Aliter
We can obtain this distribution directly from the joint distribution table by going through this example step by step.

[2]Take note that

$$p(x+y)=P[(X+Y)=x+y]$$

Click on the ad to read more

Step 1
Indicate the various summands with their corresponding joint probabilities

$x_i + y_j$	$-1+(-1)$	$-1+0$	$-1+1$	$0+(-1)$	$0+0$	$0+1$
$p(x_i, y_j)$	0.10	0.20	0.11	0.08	0.02	0.26

$2+(-1)$	$2+0$	$2+1$
0.03	0.17	0.03

Step 2
Sum the various values for X and Y

$x+y$	-2	-1	0	-1	0	1	1	2	3
$p(x,y)$	0.1	0.2	0.11	0.08	0.02	0.26	0.03	0.17	0.03

By the principle of equality of the probabilities of equivalent events, we shall perform the operation in Step 3.

Step 3
Sum the various probabilities corresponding to a particular value of the sum $x+y$, denoted by $p(x+y)$ (by the addition rule of probability)[3]. For example, the probability of the sum $x+y=1$, is given by

$$\begin{aligned} P[(X+Y)=1] &= p(1) \\ &= 0.26+0.03=0.29 \end{aligned}$$

Thus, we summarise the table in Step 3 in the table below:

$x+y$	-2	-1	1	0	2	3
$p(x+y)$	0.1	$0.2+0.08$	$0.26+0.03$	$0.11+0.02$	0.17	0.03

Step 4
Present the final result as in the table below:

$(x+y)$	-2	-1	0	1	2	3
$p(x+y)$	0.1	0.28	0.13	0.29	0.17	0.03

We can verify that this distribution of the sum is a probability distribution. That is,

$$0 \leq p(x+y) \leq 1$$

and

$$\sum p(x+y) = 1$$

Case 2
Suppose that the joint distribution of X and Y is in the form of Table 1.3. Suppose also that X and Y are independent so that $p(x_i, y_j) = g(x_i)h(y_j)$. Then the various sums $X+Y$ and their corresponding probabilities may be presented in the form of the following table.

[3]That is if a particular value in the sum appears more than once, their probabilities are added together for the purpose of constructing the probability distribution.

Download free eBooks at bookboon.com

$x_i + y_j$	$x_1 + y_1$	$x_2 + y_2$	$\cdots$	$x_n + y_m$
$p(x_i, y_j)$	$p(x_1)p(y_1)$	$p(x_2)p(y_2)$	$\cdots$	$p(x_n)p(y_m)$

Theorem 2.2
Suppose that X and Y are independent, discrete random variables with marginal probability distributions $g(x)$ and $h(y)$ respectively, then $P(X + Y = k)$ is the sum of the product of $g(x)$ and $h(y)$ for the corresponding value $k = x + y$

Example 2.3
Suppose the joint probabilities of two random variables X and Y are given in the table below.

X	Y: 3	Y: 6	Row Total
-2	0.28	0.12	0.4
4	0.42	0.18	0.6
Column Total	0.7	0.3	1.0

(a) Verify that X and Y are independent random variables.

(b) Find the distribution of the sum of X and Y.

Solution

(a) The marginal probability distributions of the X is

x	-2	4
$g(x)$	0.4	0.6

and the marginal probability distributions of the Y is

y	3	6
$h(y)$	0.7	0.3

so that we have

$$\begin{aligned}
p(-2,3) &= 0.28 = g(-2)h(3) = 0.4(0.7) = 0.28 \\
p(-2,6) &= 0.12 = g(-2)h(6) = 0.4(0.3) = 0.12 \\
p(4,3) &= 0.42 = g(4)h(3) \quad = 0.6(0.7) = 0.42 \\
p(4,6) &= 0.18 = g(4)h(6) \quad = 0.6(0.3) = 0.18
\end{aligned}$$

Hence, X and Y are independent.

Download free eBooks at bookboon.com

(b) It has been shown that X and Y are independent. Hence, we obtain the following table:

$x_i + y_j$	$-2+3$	$-2+6$	$4+3$	$4+6$
$p(x_i + y_j)$	$0.4(0.7)$	$0.4(0.3)$	$0.6(0.7)$	$0.6(0.3)$

Note
Even though the values of the random variables of X and Y are added together, their corresponding probabilities $g(x_i)$ and $h(y_j)$ are multiplied, if X and Y are independent.

By the principle of equality of probabilities of equivalent events, the distribution of the sum of X and Y is presented in the table below:

$x_i + y_j$	1	4	7	10
$p(x_i + y_j)$	0.28	0.12	0.42	0.18

Distribution of Sums of Discrete Bivariate Random Variable in Expression Form

In most cases the distribution of X and Y may not be presented in a table. Suppose that all that we know are the values of $X+Y$ and their corresponding probabilities. In such a case, to obtain the distribution of $X+Y$ we shall use Theorem 2.3.

Need help with your dissertation?

Get in-depth feedback & advice from experts in your topic area. Find out what you can do to improve the quality of your dissertation!

Get Help Now

Go to www.helpmyassignment.co.uk for more info

Helpmyassignment

Download free eBooks at bookboon.com

Click on the ad to read more

Theorem 2.3
Suppose X and Y are nonnegative discrete random variables with joint probability function $p(x, y)$. Let $Z = X + Y$. Then the probability distribution of Z is

$$P(Z = z) = \sum_{x=0}^{z} p(x, z - x)$$

or equivalently,

$$P(Z = z) = \sum_{y=0}^{z} p(z - y, y)$$

Proof

$$\begin{aligned} P(Z = z) &= P(X + Y = z) \\ &= \sum_{x=0}^{z} P(X = x,\, X + Y = z) \\ &= \sum_{x=0}^{z} P(X = x,\, Y = z - x) \qquad (i) \end{aligned}$$

from which the result follows.

Similarly,

$$P(Z = z) = \sum_{y=0}^{z} p(z - y, y)$$

Note
It is important to note the limits of summation. Beyond these limits, one of the two component mass functions is zero. When dealing with densities that are nonzero only on some subset of values, we must always be careful. In case X and Y are allowed to take negative values as well, the lower index of summation is changed from 0 to $-\infty$.

Example 2.4
Refer to Example 1.2, find

(a) the distribution of $Z = X + Y$;

  (i) from the first principle (that is, using the joint distribution of X and Y directly;)
  (ii) using Theorem 2.3.

(b) $P(Z = 2)$,

  (i) using the distribution obtained in (a) (i);
  (ii) using the distribution obtained in (a) (ii).

Download free eBooks at bookboon.com

Solution
Recall from the solution of Example 1.2 that

$$p(x,y) = \frac{1}{21}(3x+2y), \quad \text{for } x = 0,1; \quad y = 0,1,2$$

The various pairs of X and Y that we require are

$$(0,\ 0),\ (0,\ 1),\ (0,\ 2),\ (1,\ 0),\ (1,\ 1),\ (1,\ 2)$$

By the principle of equality of probabilities of equivalent events, the events $\{(0,\ 1)\}$ and $\{1,\ 0\}$ are equivalent to the event $\{Z = 1\}$. Similarly, the events $\{(0,\ 2)\}$ and $\{(1,\ 1)\}$ are equivalent to the event $\{Z = 2\}$. Hence, the possible values of Z are $z = 0,1,2,3$.

(a) (i) Writing the sum of the joint probability mass function as

$$P(X+Y = x+y) = \frac{1}{21}\sum_{x=0}^{1}\sum_{y=0}^{2}(3x+2y)$$

we calculate the distribution of $Z = X+Y$ as follows.

$$\begin{aligned} P(Z=0) &= p(0,0) \\ &= \frac{1}{21}[3(0)+2(0)] \\ &= 0 \end{aligned}$$

$$\begin{aligned} P(Z=1) &= p(0,1)+p(1,0) \\ &= \frac{1}{21}[3(0)+2(1)] + \frac{1}{21}[3(1)+2(0)] \\ &= \frac{5}{21} \end{aligned}$$

$$\begin{aligned} P(Z=2) &= p(0,2)+p(1,1) \\ &= \frac{1}{21}[3(0)+2(2)] + \frac{1}{21}[3(1)+2(1)] \\ &= \frac{9}{21} \end{aligned}$$

$$\begin{aligned} P(Z=3) &= p(1,2) \\ & \frac{1}{21}[3(1)+2(2)] \\ &= \frac{7}{21} \end{aligned}$$

Hence the distribution of $Z = X+Y$ is

$x+y$	0	1	2	3
$p(x+y)$	0	$\frac{5}{21}$	$\frac{9}{21}$	$\frac{7}{21}$

Download free eBooks at bookboon.com

(ii) The distrbution of $Z = X + Y$ by Theorem 2.3 is

$$P(Z = z) = \begin{cases} \sum\limits_{x=0}^{z} p(0,\ z-0), & z = 0, 1, 2, 3 \\[2ex] \sum\limits_{x=1}^{z} p(1,\ z-1), & z = 1, 2, 3 \end{cases}$$

Aliter

$$P(Z = z) = \begin{cases} \sum\limits_{y=0}^{z} p(0,\ z-0), & z = 0, 1, 2, 3 \\[2ex] \sum\limits_{y=1}^{z} p(1,\ z-1), & z = 1, 2, 3 \\[2ex] \sum\limits_{y=2}^{z} p(2,\ z-2), & z = 2, 3 \end{cases}$$

(b) (i) From the table obtained in (a), that is, calculating the probability from first principle, we have

$$P(Z = 2) = \frac{3}{7}$$

Download free eBooks at bookboon.com

Click on the ad to read more

(ii) By Theorem 2.3 (that is, using the results in (a)(ii)) we have

$$\begin{aligned} P(Z=2) &= \sum_{x=0}^{1} p(x, 2-x) \\ &= \frac{1}{21}\sum_{x=0}^{1}[3x+2(2-x)] \\ &= \frac{1}{21}\sum_{x=0}^{1}(x+4) \\ &= \frac{3}{7} \end{aligned}$$

Aliter
Using the alternative formula in Theorem 2.3,

$$\begin{aligned} P(Z=2) &= \sum_{y=0}^{2} p(2-y, y) \\ &= \sum_{y=1}^{2} p(2-y, y), \quad \text{since} \quad p(2,0)=0 \\ &= \frac{1}{21}\sum_{y=1}^{2}[3(2-y)+2y] \\ &= \frac{1}{21}\sum_{y=1}^{2}(6-y) \\ &= \frac{3}{7} \end{aligned}$$

Theorem 2.4 CONVOLUTION THEOREM
(Discrete Case)
Suppose that X and Y are independent nonnegative discrete random variables with marginal probability distributions $g(x)$ and $h(y)$ respectively. Let Z be the sum of X and Y. Then the convolution of g and h is the distribution of the sum Z given by

$$P(Z=z) = \sum_{x=0}^{z} g(x)h(z-x) \qquad z \geq 0$$

or equivalently,

$$P(Z=z) = \sum_{y=0}^{z} g(z-y)h(y) \qquad z \geq 0$$

called the convolution of g and h

Download free eBooks at bookboon.com

Proof
Since X and Y are independent, the proof follows by writing

$$P(X = x, Y = z - x)$$

as

$$P(X = x)P(Y = z - x)$$

in Theorem 2.3.

Theorem 2.4 suggests a method of obtaining probability mass function when $X + Y$ are independent.

When we have two one-dimensional probability mass functions g and h, the convolution formula

$$P(Z = z) = \sum_{x=0}^{z} g(x)h(z - x), \qquad z \geq 0$$

is a one-dimensional probability mass function. Thus the probability function of the sum of two independent random variables is the convolution[4] of the individual probability functions.

Note

(a) Theorem 2.3 also expresses convolution;

(b) The convolution of g and h is denoted as $g * h$.

Since

$$X + Y = Y + X$$

it is easy to verify that

$$\begin{aligned} P(Z = z) &= \sum_{x}^{z} g(x)h(z - x) \\ &= \sum_{y}^{z} g(z - y)h(y) \end{aligned}$$

that is, $g * h = h * g$.

Example 2.5
Suppose that X and Y are independent discrete random variables having probability distributions

$$p_X(x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad \text{and} \quad p_Y(y) = \frac{e^{-\theta}\theta^y}{y!}$$

respectively, where $x = 0, 1, 2, \ldots$ and $y = 0, 1, 2, \ldots$ Find

(a) the distribution of the sum $Z = X + Y$.

(b) $P(X + Y = 4)$, if $\lambda = 1$ and $\theta = 2$.

[4] Some writers prefer to use the French word *Composition* or even the German equivalent *Faltung*

Download free eBooks at bookboon.com

Solution
The joint probability distribution of X and Y is given by

$$p(x,y) = \frac{e^{-\lambda}\lambda^x}{x!} \cdot \frac{e^{-\theta}\theta^y}{y!}$$

(a) By Theorem 2.4

$$P(Z=z) = \sum_{x=0}^{z} \frac{e^{-\lambda}\lambda^x}{x!} \cdot \frac{e^{-\theta}\theta^{z-x}}{(z-x)!}$$

(Note that the summation goes from $x=0$ to $x=z$, since x cannot exceed $z = x+y$).

$$\begin{aligned} P(Z=z) &= \sum_{x=0}^{z} \frac{e^{-(\lambda+\theta)}\lambda^x\theta^{z-x}}{x!(z-x)!} \\ &= e^{-(\lambda+\theta)} \sum_{x=0}^{z} \frac{\lambda^x\theta^{z-x}}{x!(z-x)!} \\ &= \frac{e^{-(\lambda+\theta)}}{z!} \sum_{x=0}^{z} \frac{z!}{x!(z-x)!}\lambda^x\theta^{z-x} \end{aligned}$$

Identifying the sum as the binomial expansion of $(\lambda+\theta)^z$ we obtain

$$P(Z=z) = \frac{e^{-(\lambda+\theta)}(\lambda+\theta)^z}{z!}, \qquad z = 0, 1, 2, \ldots$$

Click on the ad to read more

(b) When $\lambda = 1$ and $\theta = 2$

$$\begin{aligned} P(Z = 4) &= \frac{e^{-(1+2)}(1+2)^4}{4!} \\ &= \frac{e^{-3}(3)^4}{4!} \end{aligned}$$

Example 2.5 shows that, if X and Y are independent random variables having Poisson distribution with parameters λ and θ respectively then their sum has a Poisson distribution with parameter $\lambda + \theta$. That is, the sum of the two independent Poisson random variables is again a Poisson random variable. The ideas here could be used to prove more generally by induction that the sum of a finite Poisson random variables also has a Poisson distribution, a result we proved earlier using the moment generating function in my book "Introductory Probability Theory" (Nsowah-Nuamah, 2017).

Note
In general, if two independent random variables follow the same type of distribution, it is not necessarily true that their sum follows the same type of distribution (see Example 2.7 in the sequel.)

2.2.2 Sums of Continuous Bivariate Random Variables

The formulas for finding sums of continuous random variables are similar to those in the discrete case by replacing the sum by the integral. However, the actual computation is more complex with the continuous case.

Theorem 2.5
Suppose that X and Y are continuous random variables having joint density function $f(x, y)$. Let $Z = X + Y$ and denote the density function of Z by $s(z)$. Then

$$s(z) = \int_{-\infty}^{\infty} f(x, z - x)\, dx$$

or equivalently,

$$s(z) = \int_{-\infty}^{\infty} f(z - y, y)\, dy$$

Proof
We shall use the cumulative distribution function technique. Let S denote the c.d.f. of the random variable Z. Then

$$S(z) = P(Z \le z) = P(X + Y \le z) = \int\!\!\int_{\mathcal{R}} f(x, y)\, dx\, dy$$

where $\mathcal{R}$ is the part of the region over which $x + y < z$, shown shaded in the Figure 2.1.

Download free eBooks at bookboon.com

To represent this integral as an iterated integral, we fix x and then integrate with respect to y from $z-x$ to $-\infty$. Next integrate with respect to x from $-\infty$ to ∞. Thus

$$S(z) = \int_{x=-\infty}^{\infty} \left[\int_{y=-\infty}^{z-x} f(x,y)\,dy\right] dx$$

Let $u = x+y$; for fixed x, $du = dy$. If x is fixed then when $y = z-x$, $u = z$. Thus

$$S(z) = \int_{x=-\infty}^{\infty} \left[\int_{u=-\infty}^{z} f(x, u-x)\,du\right] dx$$

We assume that $f(x,y)$ has continuous partial derivative. Therefore we exchange the order of integration and obtain

$$S(z) = \int_{u=-\infty}^{z} \left[\int_{x=-\infty}^{\infty} f(x, u-x)\,dx\right] du$$

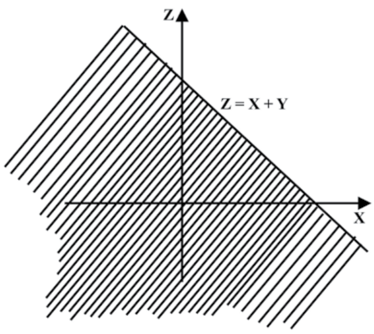


Fig. 2.1 Region $x+y<z$ is shaded

Differentiating $S(z)$ with respect to z using the fundamental theorem of calculus will yield the density function of the random variable Z. Hence,

$$\begin{aligned} \frac{d\,S(z)}{d\,z} &= \frac{d}{d\,z}\left\{\int_{u=-\infty}^{z}\left[\int_{x=-\infty}^{\infty} f(x,u-x)\,dx\right] du\right\} \\ s(z) &= \int_{-\infty}^{\infty} f(x, z-x)\,dx \end{aligned}$$

Reversing the roles played by X and Y, we obtain,

$$s(z) = \int_{-\infty}^{\infty} f(z-y, y)\,dy$$

Example 2.6
Refer to Example 1.3.

(a) Find the distribution of $Z = X+Y$;

(b) Calculate $P(Z \leq 2)$ using

  (i) the distribution of Z;

  (ii) the joint distribution of (X, Y), that is, from the first principle.

Download free eBooks at bookboon.com

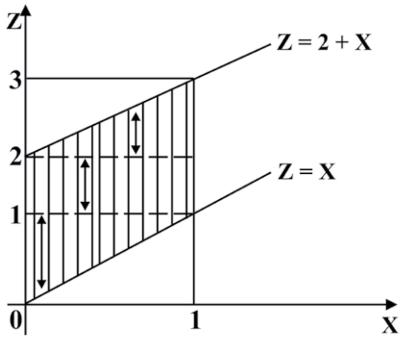
Solution

(a) $Z = X + Y \Longrightarrow Y = Z - X$

Therefore

$$0 < Y < 2 \Longrightarrow 0 \le Z - X \le 2 \Longrightarrow X \le Z \le X + 2$$

To determine the region of integration we sketch below the preceding inequality and $0 < X < 1$.



It can be seen from the sketch that the region of integration is partitioned into three, namely,

$$\begin{array}{rcl} 0 \le X \le Z & \text{when} & 0 < Z \le 1 \\ 0 \le X \le 1 & \text{when} & 1 < Z \le 2 \\ Z - 2 \le X \le 1 & \text{when} & 2 < Z \le 3 \end{array}$$

Click on the ad to read more
Download free eBooks at bookboon.com

For $0 < Z \le 1$,

$$s(z) = \int_{x=0}^{z} \left(x^2 + \frac{x(z-x)}{3}\right) dx = \frac{7}{18} z^3$$

For $1 < Z \le 2$,

$$s(z) = \int_{x=0}^{1} \left(x^2 + \frac{x(z-x)}{3}\right) dx = \frac{2}{9} + \frac{z}{6}$$

Finally, for $2 < Z \le 3$,

$$s(z) = \int_{x=z-2}^{1} \left(x^2 + \frac{x(z-x)}{3}\right) dx = \frac{1}{18}\left(-7z^3 + 36z^2 - 57z + 36\right)$$

Therefore, the distribution of Z is

$$s(z) = \begin{cases} \dfrac{7}{18} z^3, & 0 < z \le 1 \\[2ex] \dfrac{2}{9} + \dfrac{z}{6}, & 1 < z \le 2 \\[2ex] \dfrac{1}{18}\left(-7z^3 + 36z^2 - 57z + 36\right), & 2 < z \le 3 \\ 0, & \text{otherwise} \end{cases}$$

(b) (i) Using the distribution of Z we calculate the required probability as follows:

$$\begin{aligned} P(X \le 2) &= P(0 < z \le 1) + P(1 < z < 2) \\ &= \int_{x=0}^{1} \frac{7}{18} x\, dx + \int_{x=1}^{2} \left(\frac{2}{9} + \frac{x}{6}\right) dx \\ &= \left[\frac{7}{18}\frac{z^4}{4}\right]_0^1 + \left[\frac{1}{3}\left(\frac{2}{3} z + \frac{z^2}{4}\right)\right]_1^2 \\ &= \frac{41}{72} \end{aligned}$$

(ii) Now, using the joint distribution of (X, Y)

$$\begin{aligned} P(Z \le 2) &= \iint\limits_{z<2} f(x,y)\, dy\, dx \\ &= \int_{x=0}^{1} \int_{y=0}^{2-x} \left(x^2 + \frac{xy}{3}\right) dy\, dx \\ &= \int_{x=0}^{1} \left[x^2 y + \frac{xy^2}{6}\right]_{y=0}^{2-x} dx \\ &= \int_{x=0}^{1} \left[x^2(2-x) + \frac{x}{6}(2-x)^2 - \left(x^2 + \frac{x}{6}\right)\right] dx \\ &= \int_{x=0}^{1} \left[(2x^2 - x^3) + \frac{1}{6}(4x - 4x^2 + x^3) - (x^2 + \frac{x}{6})\right] dx \\ &= \left[\frac{2x^2}{3} - \frac{4x^4}{4}\right]_0^1 + \frac{1}{6}\left[2x^2 - \frac{4x^2}{3} + \frac{x^4}{4}\right]_0^1 - \left[\frac{x^3}{3} + \frac{x^2}{12}\right]_0^1 \\ &= \frac{41}{72} \end{aligned}$$

Download free eBooks at bookboon.com

Theorem 2.6 CONVOLUTION THEOREM
(Continuous Case)

Suppose that X and Y are *independent* continuous random variables having joint density function $f(x, y)$ with marginal probability distributions $g(x)$ and $h(y)$ respectively. Let $Z = X + Y$ and denote the p.d.f. of Z by $s(z)$. Then

$$s(z) = \int_{-\infty}^{\infty} g(x)h(z - x)\,dx$$

or equivalently,

$$s(z) = \int_{-\infty}^{\infty} g(z - y)h(y)\,dy$$

The theorem follows from Theorem 2.5 by noting that since X and Y are independent:

$$\begin{aligned} f(x,\, z - x) &= g(x)h(z - x) \\ f(z - y,\, y) &= g(z - y)h(y) \end{aligned}$$

The function s is called the *convolution* of the functions g and h, and the expression in the Theorem is called the *convolution formula.*

If X and Y are nonnegative random variables, then the convolution formula reduces to the following

$$s(z) = \int_0^z g(x)h(z - x)\,dx$$

This is because each of g and h is equal to 0 for negative argument.

Example 2.7

Suppose that X and Y are independent and identically distributed random variables having p.d.f.'s

$$f_{\mathrm{X}}(x) = \lambda\, e^{-\lambda\, x}, \quad x \geq 0$$

and

$$f_{\mathrm{Y}}(y) = \lambda\, e^{-\lambda\, y}, \quad y \geq 0$$

respectively, find the distribution of the sum $Z = X + Y$.

Solution

$$\begin{aligned} f_{\mathrm{Z}}(z) &= \int_0^z \lambda\, e^{-\lambda\, t}\, \lambda\, e^{-\lambda(z-t)}\,dt \\ &= \lambda^2 \int_0^z e^{-\lambda\, z}\,dt \\ &= \lambda^2\, z\, e^{-\lambda\, z}, \qquad z \geq 0 \end{aligned}$$

Note

(a) See note after Theorem 2.3.

(b) This distribution is a gamma distribution with parameters 2 and λ.

2.3 DIFFERENCES OF RANDOM VARIABLES

2.3.1 Differences of Discrete Bivariate Random Variables

The probability distribution of the difference of two independent discrete random variables X and Y may be considered as the probability distribution of the sum of X and $(-Y)$. This is demonstrated in Table 2.2.

Table 2.2 Various Differences $(X - Y)$ and their Corresponding Probabilities

$x_i - y_j$	$x_1 - y_1$	$x_1 - y_2$	$\cdots$	$x_1 - y_m$	$x_2 - y_1$	$x_2 - y_2$	$\cdots$	$\cdots$	$x_2 - y_m$	$x_n - y_1$	$x_n - y_2$	$\cdots$	$x_n - y_m$
$p(x_i, y_j)$	$p(x_1, y_1)$	$p(x_1, y_2)$	$\cdots$	$p(x_1, y_m)$	$p(x_2, y_1)$	$p(x_2, y_2)$	$\cdots$	$\cdots$	$p(x_2, y_m)$	$p(x_n, y_1)$	$p(x_n, y_2)$	$\cdots$	$p(x_n, y_m)$

Download free eBooks at bookboon.com
Click on the ad to read more

To find the distribution of the difference, we similarly apply the principle of the *probabilities of equivalent events.*

Similar to sums of random variables, we shall present the two cases for the difference, namely, when X and Y are not independent and when they are.

Example 2.8
For the data in Example 2.2, find the distribution of the difference of the random variables X and Y.

Solution
Indicate the various difference with their corresponding joint probabilities

$x_i - y_j$	$-1-(-1)$	$-1-0$	$-1-1$	$0-(-1)$	$0-0$	$0-1$
$p(x_i, y_j)$	0.1	0.2	0.11	0.08	0.08	0.26

$2-(-1)$	$2-0$	$2-1$
0.03	0.17	0.03

Step 2
Subtract the various values for Y from the various values for X:

$x-y$	0	-1	-2	1	0	-1	3	2	1
$p(x,y)$	0.1	0.2	0.11	0.08	0.02	0.26	0.03	0.17	0.03

Step 3
By the principle of equality of the probabilities of equivalent events we shall obtain the table below:

$x-y$	-2	-1	0	1	2	3
$p(x+y)$	0.11	$0.2+0.26$	$0.1+0.02$	$0.08+0.03$	0.17	0.03

Step 4
Present the final result as in the table below:

$x-y$	-2	-1	0	1	2	3
$p(x-y)$	0.11	0.46	0.12	0.11	0.17	0.03

which can be presented as follows:

$x-y$	-2	-1	0	1	2	3
$p(x-y)$	0.11	0.46	0.12	0.11	0.17	0.03

We can verify that this distribution of the sum is a probability distribution. That is,

$$0 \leq p(x-y) \leq 1$$

Download free eBooks at bookboon.com

and

$$\sum p(x-y) = 1$$

Example 2.9
For the data in Example 2.3, find the distribution of the difference of the random variables X and Y.

Solution
It has been shown in Example 2.3 that X and Y are independent. Therefor, the various values of the sum $\{x+(-y)\}$ and their corresponding probabilities are given in the following table:

$x_i - y_j$	$-2-3$	$-2-6$	$4-3$	$4-6$
$p(x_i, y_j)$	0.4(0.7)	0.4(0.3)	0.6(0.7)	0.6(0.3)

By the principle of equality of probabilities of equivalent events, we obtain the distribution of the difference of X and Y, presented in the table below:

$x_i - y_j$	-5	-8	1	-2
$p(x_i - y_j)$	0.28	0.12	0.42	0.18

2.3.2 Differences of Continuous Bivariate Random Variables

Theorem 2.7
Suppose that X and Y are continuous random variables having joint density function $f(x, y)$ and let $U = X - Y$. Then

$$f_u(u) = \int_{-\infty}^{\infty} f(x,\, x-u)\, dx$$

or equivalently,

$$f_u(u) = \int_{-\infty}^{\infty} f(u+y,\, y)\, dy$$

Corollary 2.1
Suppose that X and Y are independent nonnegative continuous random variables having joint density function $f(x, y)$ with marginal probability distributions $g(x)$ and $h(y)$ respectively, and let $U = X - Y$. Then

$$f_u(u) = \int_{0}^{\infty} g(x)\, h(x-u)\, dx$$

or equivalently,

$$f_u(u) = \int_{0}^{\infty} g(u+y)\, h(y)\, dy$$

Download free eBooks at bookboon.com

Example 2.10
Refer to Example 1.3.

(a) Find the distribution of $U = X - Y$;

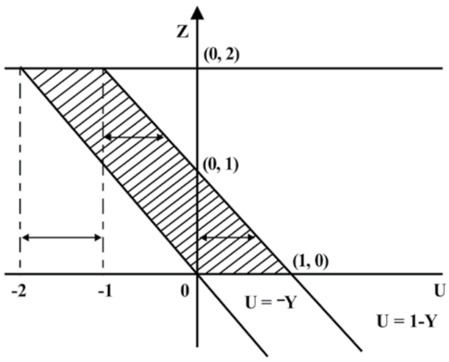
(b) Using the distribution in (a) calculate (i) $P(U \leq 1)$ (ii) $P\left(U < \dfrac{1}{2}\right)$.

Solution

(a) $$X - Y = U \implies X = U + Y.$$
Therefore

$$0 < X < 1 \implies 0 \leq U + Y \leq 1 \implies -Y \leq U \leq 1 - Y$$

The region defined by the preceding inequality and $0 < Y < 2$ is sketched in the diagram below.



It can be seen from the sketch that the region of integration is partitioned into three, namely,

$$\begin{aligned} -U \leq Y \leq 2 &\quad \text{when} \quad -2 \leq U < -1 \\ -U \leq Y \leq 1 - U &\quad \text{when} \quad -1 \leq U < 0 \\ 0 \leq Y \leq 1 - U &\quad \text{when} \quad 0 \leq U \leq 1 \end{aligned}$$

For $-2 \leq U < -1$

$$\begin{aligned} f_u(u) &= \int_{-u}^{2} \left(u^2 + \frac{7}{3}\,u\,y + \frac{4}{3}\,y^3\right)\,dy \\ &= \frac{5}{18}\,u^3 + 2\,u^2 + \frac{14}{3}\,u + \frac{32}{9} \end{aligned}$$

For $-1 \leq U < 0$

$$\begin{aligned} f_u(u) &= \int_{-u}^{1-u} \left(u^2 + \frac{7}{3}u\,y + \frac{4}{3}\,y^2\right)\,dy \\ &= \frac{4}{9} - \frac{u}{6} \end{aligned}$$

For $0 \leq U \leq 1$

$$\begin{aligned} f_u(u) &= \int_{0}^{1-u} \left(u^2 + \frac{7}{3}\,u\,y + \frac{4}{3}\,y^2\right)\,dy \\ &= \frac{4}{9} - \frac{u}{6} - \frac{5}{18}\,u^3 \end{aligned}$$

Download free eBooks at bookboon.com

Therefore

$$f_u(u) = \begin{cases} \frac{5}{18}\,u^3 + 2\,u^2 + \frac{14}{3}\,u + \frac{32}{9}, & -2 \le u < -1 \\ \frac{4}{9} - \frac{u}{6}, & -1 \le u < 0 \\ \frac{4}{9} - \frac{u}{6} - \frac{5}{18}\,u^3, & 0 \le u \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

(b)

$$\begin{aligned} \text{(i)} \quad P(U \le 1) &= P(-2 \le U \le 1) \\ &= P(-2 \le U < 1) + P(-1 \le U < 0) + P(0 \le U \le 1) \\ &= \int_{-2}^{-1} \left(\frac{5}{18}\,u^3 + 2u^2 + \frac{14}{3}\,u + \frac{32}{9} \right) du \\ &\quad + \int_{-1}^{0} \left(\frac{4}{9} - \frac{u}{6} \right) du + \int_{0}^{1} \left(\frac{4}{9} - \frac{u}{6} - \frac{5}{18}\,u^3 \right) du \\ &= \frac{13}{72} + \frac{19}{36} + \frac{21}{72} \\ &= 1 \end{aligned}$$

Download free eBooks at bookboon.com

Click on the ad to read more

$$
\begin{aligned}
\text{(ii)}\quad P\left(U \le \frac{1}{2}\right) &= P(-2 \le U \le 1) + P(-1 \le U < 0) \\
&\quad + P\left(0 \le U \le \frac{1}{2}\right) \\
&= \int_{-2}^{-1} \left(\frac{5}{18}u^3 + 2u^2 + \frac{14}{3}u + \frac{32}{9}\right) du \\
&\quad + \int_{-1}^{0} \left(\frac{4}{9} - \frac{u}{6}\right) du + \int_{0}^{\frac{1}{2}} \left(\frac{4}{9} - \frac{u}{6} - \frac{5}{18}u^3\right) du \\
&= \frac{13}{72} + \frac{19}{36} + \frac{227}{1152} \\
&= \frac{1043}{1152}
\end{aligned}
$$

2.4 PRODUCTS OF BIVARIATE RANDOM VARIABLES

2.4.1 Product of Discrete Bivariate Random Variables

The probability distribution of the product of two discrete random variables X and Y whose joint probability functions are in tabular form can be calculated using the principle of the equality of the probabilities of two equivalent events. This is demonstrated in Table 2.3.

Table 2.3 Various Products XY and their Corresponding Probabilities

x_iy_j	x_1y_1	x_1y_2	$\cdots$	x_1y_m	x_2y_1	x_2y_2	$\cdots$	x_2y_m	x_ny_1	x_ny_2	$\cdots$	x_ny_m
$p(x_i, y_j)$	$p(x_1, y_1)$	$p(x_1, y_2)$	$\cdots$	$p(x_1, y_m)$	$p(x_2, y_1)$	$p(x_2, y_2)$	$\cdots$	$p(x_2, y_m)$	$p(x_n, y_1)$	$p(x_n, y_2)$	$\cdots$	$p(x_n, y_m)$

Example 2.11
For the data in Example 2.2, find the distribution of the product of the random variables X and Y.

Solution
Step 1
Indicate the various products with their corresponding joint probabilities

x_iy_j	$-1(-1)$	$-1(0)$	$-1(1)$	$0(-1)$	$0(0)$	$0(1)$	$2(-1)$	$2(0)$	$2(1)$
$p(x_i, y_j)$	0.1	0.2	0.11	0.08	0.08	0.26	0.03	0.17	0.03

Download free eBooks at bookboon.com

Step 2
Multiply the various values for X by the various values for Y:

xy	1	0	-1	0	0	0	-2	0	2
$p(x,y)$	0.1	0.2	0.11	0.08	0.02	0.26	0.03	0.17	0.03

Step 3
By the principle of equality of the probabilities of equivalent events we shall obtain the table below:

xy	-2	-1	0	1	2
$p(xy)$	0.03	0.11	$0.2+0.08+0.02+0.26+0.17$	0.1	0.03

Step 4
Present the final result as in the table below:

xy	-2	-1	0	1	2
$p(xy)$	0.03	0.11	0.73	0.1	0.03

We can verify that this distribution of the sum is a probability distribution. That is,

$$0 \leq p(xy) \leq 1$$

and

$$\sum p(xy) = 1$$

Example 2.12
For the data in Example 2.3, find the distribution of the product of X and Y.

Solution
It has been shown in Example 2.3 that X and Y are independent. Therefor, the various values of the product xy and their corresponding probabilities are given in the following table:

x_iy_j	$-2(3)$	$-2(6)$	$4(3)$	$4(6)$
$p(x_i,y_j)$	$0.4(0.7)$	$0.4(0.3)$	$0.6(0.7)$	$0.6(0.3)$

By the principle of the equality of probabilities of equivalent events we obtain the result summarised in the following table.

x_iy_j	-6	-12	12	24
$p(x_iy_j)$	0.28	0.12	0.42	0.18

Example 2.13
Suppose there are two independent distributions:

x	-2	4
$g(x)$	0.4	0.6

Download free eBooks at bookboon.com

y	-2	4
$h(y)$	0.4	0.6

Find the distribution of (a) X^2 (b) XY.

Solution

(a)

x^2	4	16
$p(x^2)$	0.4	0.6

(b)

xy	$-2(-2)$	$-2(4)$	$4(-2)$	$4(4)$
$p(x,y)$	$0.4(0.4)$	$0.4(0.6)$	$0.6(0.4)$	$0.6(0.6)$

xy	8	-8	-8	16
$p(x,y)$	0.16	0.24	0.24	0.36

xy	4	-8	16
$p(xy)$	0.16	0.48	0.36

2.4.2 Product of Continuous Bivariate Random Variables

Theorem 2.8
Suppose that X and Y are continuous random variables having joint density function $f(x,y)$ and let $V = XY$. Then

$$f_v(v) = \int_{-\infty}^{\infty} \left(\frac{1}{|x|}\right) f\left(x, \frac{v}{x}\right) dx$$

or equivalently,

$$f_v(v) = \int_{-\infty}^{\infty} \left(\frac{1}{|y|}\right) f\left(\frac{v}{y}, y\right) dy$$

Corollary 2.2
Suppose that X and Y are *independent*, nonnegative continuous random variables having joint density function $f(x,y)$ with marginal probability distributions $g(x)$ and $h(y)$ respectively and let $V = XY$. Then

$$f_v(v) = \int_0^{\infty} \left(\frac{1}{x}\right) h\left(\frac{v}{x}\right) g(x)\, dx, \quad y > 0$$

or equivalently,

$$f_v(v) = \int_0^{\infty} \left(\frac{1}{y}\right) g\left(\frac{v}{y}\right) h(y)\, dy, \quad x > 0$$

We can calculate the probability $P(XY \le t)$ for the independent, nonnegative continuous random variables X and Y thus

$$\begin{aligned} P(XY \le t) &= \iint\limits_{0<xy\le t} g(x)h(y)\,dx\,dy \\ &= \int_0^t \left[\int_{-\infty}^{\infty} \left(\frac{1}{y}\right) g\left(\frac{v}{y}\right) h(y)\,dy\right] dv \\ &= \int_0^t f_v(v)\,dv \end{aligned}$$

The integration is restricted to the first quadrant, since $g(x)h(y)$, the joint density function of X and Y, are zero elsewhere.

Example 2.14
Refer to Example 1.3.

(a) Find the distribution of $V = XY$;

(b) Using the distribution in (a), calculate $P\left(\frac{1}{2} \le U \le 1\right)$.

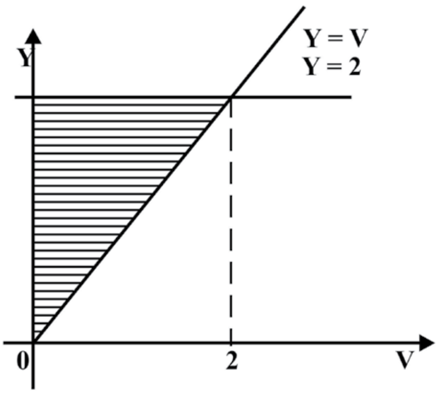
Solution

(a) $\qquad XY = V \implies X = \dfrac{V}{Y}$.

Therefore

$$0 < X < 1 \implies 0 \le \frac{V}{Y} \le 1 \implies 0 \le V \le Y$$

The region defined by the preceding inequality and $0 < Y < 2$ is sketched in the diagram below.



It can be seen from the sketch that the region of integration is

$$V \le Y \le 2 \text{ when } 0 < V \le 2$$

Therefore

$$h(u) = \int_{y=v}^{2} \left(\frac{v^2}{y^3} + \frac{1}{3}\frac{v}{y}\right) dy$$

Download free eBooks at bookboon.com

$$
\begin{aligned}
&= \left[\frac{v}{3}\ln y - \frac{1}{2\,y^2}\right]_v^2 \\
&= \frac{v}{3}\ln\left(\frac{2}{v}\right) + \frac{1}{2} - \frac{v^2}{8}, \qquad\qquad 0 < v \leq 2
\end{aligned}
$$

(b) Now,

$$
\begin{aligned}
P\left(\frac{1}{2} \leq V \leq 1\right) &= \int_{\frac{1}{2}}^{1}\left[\frac{v}{3}\ln\left(\frac{2}{v}\right) + \frac{1}{2} - \frac{v^2}{8}\right] dv \\
&= \left[\frac{v}{2} - \frac{v^3}{24} + \frac{v^2}{6}\left(2 + \ln\frac{2}{v}\right)\right]_{\frac{1}{2}}^{1} \\
&= 0.3338
\end{aligned}
$$

2.5 QUOTIENTS OF BIVARIATE RANDOM VARIABLES

2.5.1 Quotient of Discrete Bivariate Random Variables

The probability distribution of the quotient of two discrete random variables X and Y $\left(\frac{X}{Y}\right)$ whose joint probability functions are in tabular form can be calculated using the principle of the equality of the probabilities of two equivalent events. This is demonstrated in Table 2.4.

Deloitte.

Discover the truth at www.deloitte.ca/careers

© Deloitte & Touche LLP and affiliated entities.

Download free eBooks at bookboon.com

Click on the ad to read more

Table 2.4 Various Quotient $\dfrac{X}{Y}$ and their Corresponding Probabilities

$\frac{x_i}{y_j}$	$\frac{x_1}{y_1}$	$\frac{x_1}{y_2}$	$\cdots$	$\frac{x_1}{y_m}$	$\frac{x_2}{y_1}$	$\frac{x_2}{y_2}$	$\cdots$
$p(x_i, y_j)$	$p(x_1, y_1)$	$p(x_1, y_2)$	$\cdots$	$p(x_1, y_m)$	$p(x_2, y_1)$	$p(x_2, y_2)$	$\cdots$

$\cdots$	$\frac{x_2}{y_m}$	$\frac{x_n}{y_1}$	$\frac{x_n}{y_2}$	$\cdots$	$\frac{x_n}{y_m}$
$\cdots$	$p(x_2, y_m)$	$p(x_n, y_1)$	$p(x_n, y_2)$	$\cdots$	$p(x_n, y_m)$

Suppose X and Y are two independent variables. Then the distribution of $\dfrac{X}{Y},\ Y \neq 0$ can, as before, be calculated using the principle of equality of probabilities of equivalent events.

Example 2.15
Consider the distributions of two independent random variables X and Y.

x	2	5
$g(x)$	0.2	0.8

y	4	8
$h(y)$	0.7	0.3

Find the distribution of $\dfrac{X}{Y}$ and $\dfrac{Y}{X}$.

Solution
Proceeding as before:

x/y	$2/4$	$2/8$	$5/4$	$5/8$
$p(x, y)$	0.2(0.7)	0.2(0.3)	0.8(0.7)	0.8(0.3)

The distribution of $\dfrac{X}{Y}$ is thus

x/y	$1/2$	$1/4$	$5/4$	$5/8$
$p(x/y)$	0.14	0.06	0.56	0.24

The distribution of $\dfrac{Y}{X}$ follows similarly and is given in the table below:

y/x	2	4	$4/5$	$8/5$
$p(y/x)$	0.14	0.06	0.56	0.24

Download free eBooks at bookboon.com

2.5.2 Quotient of Continuous Bivariate Random Variables

Theorem 2.9
Suppose that X and Y are continuous random variables having joint density function $f(x, y)$ and let $W = \dfrac{X}{Y}$. Then

$$f_w(w) = \int_{-\infty}^{\infty} |y|\, f(wy,\, y)\, dy$$

Corollary 2.3
Suppose that X and Y are independent, nonnegative continuous random variables having joint density function $f(x, y)$ with marginal probability distributions $g(x)$ and $h(y)$ respectively, and let $W = \dfrac{X}{Y}$. Then

$$f_w(w) = \int_{0}^{\infty} y\, f(wy)\, h(y)\, dy$$

We can calculate the cumulative distribution function of $\dfrac{X}{Y}$ for the independent, nonnegative random variables X and Y by

$$\begin{aligned}
P\left(\frac{X}{Y} \le t\right) &= \iint\limits_{\substack{0\ <\frac{x}{y}\le t \\ 0<y<\infty}} y\, f(wy) h(y)\, dy \\
&= \int_0^t \left[\int_0^\infty y\, f(wy)\, h(y)\, dy\right] dw \\
&= \int_0^t f_w(w)\, dw
\end{aligned}$$

Example 2.16
Refer to Example 1.3.

(a) Find the distribution of $W = \dfrac{X}{Y}$;

(b) Using the distribution in (a), calculate $P\left(W \le \dfrac{1}{4}\right)$.

Solution

(a) Distribution of $W = \dfrac{X}{Y}$:

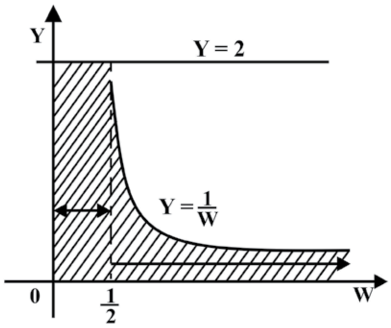
$$\frac{X}{Y} = W \implies X = YW$$

Therefore

$$0 < X < 1 \implies 0 \le YW \le 1 \implies 0 \le W \le \frac{1}{Y}$$

Download free eBooks at bookboon.com

The region defined by the preceding inequality and $0 < Y < 2$ is sketched in the diagram below.



It can be seen from the sketch that the region of integration is partitioned into two, namely,

$$0 < Y \leq 2 \quad \text{when} \quad 0 < W \leq \frac{1}{2}$$
$$0 < Y \leq \frac{1}{W} \quad \text{when} \quad \frac{1}{2} \leq W < \infty$$

$$\begin{aligned} h(w) &= \left(w^2 + \frac{w}{3}\right) \int_0^2 y^3\, dy, \qquad 0 \leq w < \frac{1}{2} \\ &= \left(w^2 + \frac{w}{3}\right) \int_0^{\frac{1}{w}} y^3\, dy, \qquad \frac{1}{2} \leq w < \infty \end{aligned}$$

Download free eBooks at bookboon.com
Click on the ad to read more

Therefore

$$h(w) = \begin{cases} 4(w^2 + \dfrac{w}{3}), & 0 \le w < \frac{1}{2} \\ \\ \dfrac{1}{4}\dfrac{1}{w^3}(w + \dfrac{w}{3}), & \frac{1}{2} \le w < \infty \end{cases}$$

(b) Calculation of $P\left(W \le \dfrac{1}{4}\right)$.

$$\begin{aligned} P\left(\frac{X}{Y} \le \frac{1}{4}\right) &= P\left(W \le \frac{1}{4}\right) \\ &= 4\int_0^{\frac{1}{4}} \left(w^2 + \frac{w}{3}\right) dw \\ &= \frac{1}{16} \end{aligned}$$

EXERCISES

2.1 Refer to Exercise 1.1. Find the distribution of

$(a)\ X+Y$	$(b)\ X-Y$	$(c)\ XY$	$(d)\ \dfrac{X}{Y}$
$(e)\ 2X+3Y$	$(f)\ 5X-Y$	$(g)\ 3X(4Y)$	$(h)\ \dfrac{3X}{7Y}$
$(i)\ X^2+5Y$	$(j)\ 3X$	$(k)\ 3(X+Y)$	$(l)\ Y^3$
$(m)\ Y^2-2Y+3X$			

2.2 Refer to Example 2.14. Find the distribution of

(a) the sum $Z = X + Y$;

(b) $W = 2X$.

2.3 Refer to Exercise 1.6. Find

(a) the distribution of the sum $Z = X + Y$;

(b) $P(Z = 2)$ (i) using the distribution of Z, (ii) from the first principle.

2.4 Refer to Exercise 1.7. Find

(a) the distribution of the sum $Z = X + Y$;

(b) $P(Z = 2)$ (i) using the distribution of Z, (ii) from the first principle.

2.5 Refer to Exercise 1.9. Find

(a) the distribution of the sum $Z = X + Y$;

(b) $P(Z = 4)$ (i) using the distribution of Z, (ii) from the first principle.

(c) Find $P(X = x | X + Y = x + y)$

Download free eBooks at bookboon.com

2.6 Refer to Exercise 1.10. Find

(a) the distribution of the sum $Z = X + Y$;

(b) $P(Z = 4)$ (i) using the distribution of Z, (ii) from the first principle.

2.7 Refer to Exercise 1.11, where $k = 2$. Find

(a) the distribution of the sum $Z = X + Y$;

(b) $P(Z \leq 1)$ (i) using the distribution of Z, (ii) from the first principle.

2.8 Refer to Exercise 1.12, where $k = \dfrac{2}{3}$. Find

(a) the distribution of the sum $Z = X + Y$;

(b) $P(Z =\leq 1)$ (i) using the distribution of Z, (ii) from the first principle.

2.9 Refer to Exercise 1.14. Find

(a) the distribution of the sum $Z = X + Y$;

(b) $P(Z \leq 1)$ (i) using the distribution of Z, (ii) from the first principle.

2.10 Refer to Exercise 1.13, where $a = \dfrac{12}{7}$. Find

(a) the distribution of the sum $Z = X + Y$;

(b) $P(Z \leq 1)$ (i) using the distribution of Z, (ii) from the first principle.

2.11 If X and Y are independent Binomial random variables and X is $b(n, p)$ and Y is $b(m, p)$, use the convolution theorem to show that $X + Y$ is $b(n + m, p)$.

2.12 If X and Y are independent Geometric random variables each with parameter p. Use the convolution theorem to show that $X + Y$ is a negative binomial $NB(2,\, p)$.

2.13 If X and Y are independent Negative Binomial random variables and X is $b^{-}(k_1,\, p)$ and Y is $b^{-}(k_2,\, p)$, use the convolution theorem to show that $X + Y$ is a negative binomial $b^{-}(k_1 + k_2,\, p)$.

2.14 If X and Y are independent Gamma random variables and X is $\Gamma(s_1,\, \lambda)$ and Y is $\Gamma(s_2,\, \lambda)$, use the convolution theorem to show that $X + Y$ is $\Gamma(s_1 + s_2,\, \lambda)$.

2.15 Suppose X and Y are independent Normal random variables and X is $N(\mu_1,\, \sigma_1^2)$ and Y is $N(\mu_2,\, \sigma_2^2)$, use the convolution theorem to show that $X + Y$ is a $N(\mu_1 + \mu_2,\, \sigma_1^2 + \sigma_2^2)$.

2.16 Refer to Example 1.15. Compute $P(X + Y > 2)$.

Download free eBooks at bookboon.com

2.17 The moment generating function of X is given by

$$M_X(t) = exp\left(2e^t - 2\right)$$

and that of Y by

$$M_Y(t) = \left(\frac{1}{4}\right)^{10}\left(3e^t + 1\right)^{10}$$

If X and Y are independent, find $P(X+Y=2)$.

2.18 Refer to Exercise 1.10. Find

(a) the distribution of the difference $U = X - Y$;

(b) (i) $P(U=0)$; (ii) $P(U=1)$; (iii) $P(U=2)$.

2.19 Refer to Exercise 1.11. Find

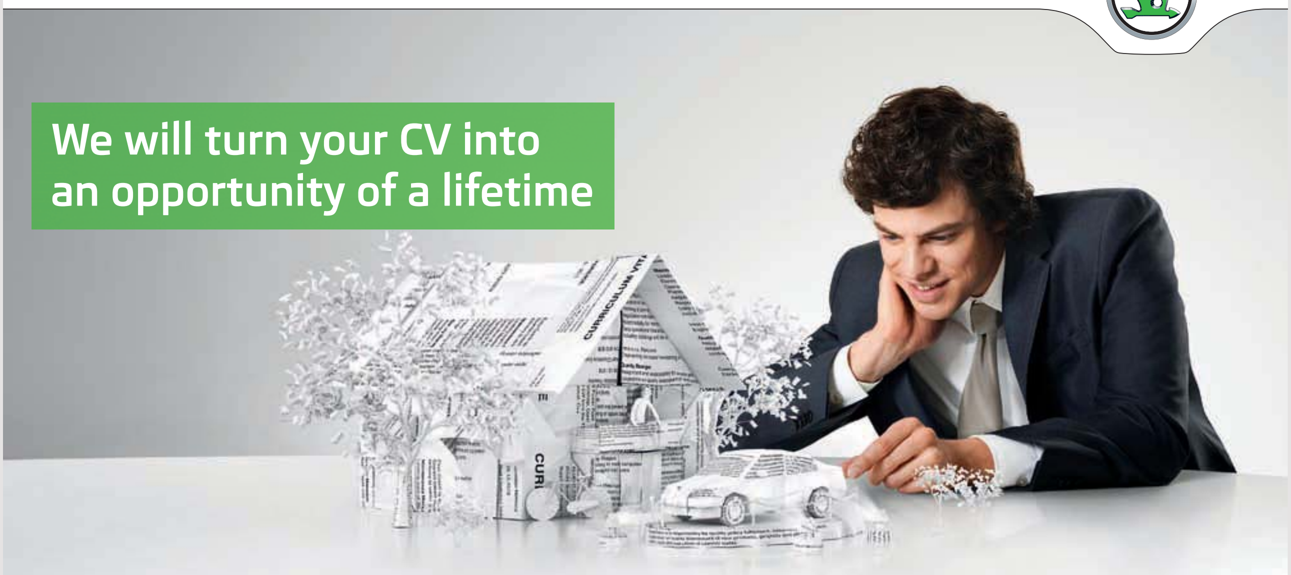
(a) the distribution of the difference $U = X - Y$;

(b) (i) $P(U \leq \frac{1}{2})$, (ii) $P(-\frac{1}{2} \leq U \leq \frac{1}{2})$.

2.20 Refer to Exercise 1.10. Find

(a) the distribution of the product $V = XY$;

(b) (i) $P(V=2)$, (ii) $P(V=3)$ (iii) $P(V=4)$, (iv) $P(V=6)$.

SIMPLY CLEVER
ŠKODA


Do you like cars? Would you like to be a part of a successful brand?
We will appreciate and reward both your enthusiasm and talent.
Send us your CV. You will be surprised where it can take you.

Send us your CV on
www.employerforlife.com

Click on the ad to read more

2.21 Refer to Exercise 1.11. Find

(a) the distribution of the product $U = XY$;
(b) (i) $P(U > \frac{3}{4})$, (ii) $P(\frac{1}{2} < U < 1)$, using the distribution of U,

2.22 Refer to Exercise 2.17. Find $P(XY = 0)$.

2.23 Refer to Exercise 1.10. Find

(a) the distribution of the quotient $W = \frac{X}{Y}$;
(b) (i) $P(\frac{X}{Y} = 1)$; (ii) $P(W = 2)$; (iii) $P(\frac{X}{Y} = 3)$.

2.24 Refer to Exercise 1.11. Find

(a) the distribution of the quotient $W = \frac{X}{Y}$;
(b) $P(W = 5)$.

2.25 Refer to Exercise 1.12, where $k = \frac{2}{3}$. Find

(a) the distribution of $P(X - Y)$;
(b) $P(-\frac{1}{2} \leq X - Y \leq 1)$.

2.26 Refer to Exercise 1.12, where $k = \frac{2}{3}$. Find

(a) the distribution of XY;
(b) $P(XY \leq \frac{4}{5})$.

2.27 Refer to Exercise 1.12, where $k = \frac{2}{3}$. Find

(a) the distribution of $P(\frac{X}{Y})$;
(b) $P(\frac{X}{Y} < 10)$.

2.28 Refer to Exercise 1.13. Find

(a) the distribution of $P(X - Y)$;
(b) $P(-\frac{1}{2} \leq X - Y \leq \frac{1}{2})$.

2.29 Refer to Exercise 1.14. Find

(a) the distribution of XY;
(b) $P(XY \leq 1)$.

2.30 Refer to Exercise 1.14. Find

(a) the distribution of $P\left(\frac{X}{Y}\right)$;
(b) $P\left(\frac{X}{Y} \leq 2\right)$.

Download free eBooks at bookboon.com

Chapter 3

EXPECTATION AND VARIANCE OF BIVARIATE DISTRIBUTIONS

3.1 INTRODUCTION

In the earlier text[5], we discussed the numerical characterisation of a single random variable. Specifically, we discussed in Chapter 6 the concepts of expectation and variance. In this chapter, we shall extend the concept of expectation and variance to the case of bivariate distribution and discuss a few more properties. We shall also realise that in this case, there arises the concept of covariance.

3.2 EXPECTATION OF BIVARIATE RANDOM VARIABLES

To refresh our minds, we recapitulate that the expectation of a discrete random variable X is

$$E(X) = \sum_{i=1}^{n} x_i \, p(x_i)$$

Example 3.1
For the data in Example 2.3, find $E(X)$ and $E(Y)$.

Solution
For convenience we present the data here.

x_i	-2	4
$p(x_i)$	0.4	0.6

$$E(X) = -2(0.4) + 4(0.6) = 1.6$$

y_j	3	6
$p(y_j)$	0.7	0.3

$$E(Y) = 3(0.7) + 6(0.3) = 3.9$$

[5] Nsowah-Nuamah, 2017

Download free eBooks at bookboon.com

3.2.1 Expectation of Functions of Discrete Bivariate Random Variables

The expectation of any function of discrete bivariate random variables, $H(X, Y)$, can be computed either the

(a) parent joint probability distribution; or

(b) derived joint probability distribution.

Expectation from Parent Joint Probability Distribution

The next theorem shows that we can find $E(X)$, $E(Y)$ and $E(XY)$ of a bivariate distribution (X, Y) directly from the parent joint probability distribution without first obtaining the derived joint probability distribution or the marginal distribution.

Definition 3.1
Let (X, Y) be a two-dimensional discrete random variable and let $Z = H(X, Y)$. Then

$$E(Z) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} H(x_i, y_j) p(x_i, y_j)$$

For the finite case,

$$E(Z) = \sum_{i=1}^{n} \sum_{j=1}^{m} H(x_i, y_j)\, p(x_i, y_j)$$

That is,

$$\begin{aligned} E[H(X,Y)] &= H(x_1,y_1)\,p(x_1,y_1) + H(x_1,y_2)\,p(x_1,y_2) + \cdots \\ &\quad + H(x_1,y_m)\,p(x_1,y_m) + H(x_2,y_1)\,p(x_2,y_1) \\ &\quad + H(x_2,y_2)\,p(x_2,y_2) + \cdots + H(x_2,y_m)\,p(x_2,y_m) \\ &\quad + \cdots + H(x_n,y_1)\,p(x_n,y_1) + H(x_n,y_2)\,p(x_n,y_2) \\ &\quad + \cdots + H(x_n,y_m)\,p(x_n,y_m) \end{aligned}$$

Example 3.2
For the data in Example 2.3, find the expectation of the function

$$H(X, Y) = X^2 Y$$

Solution

$$\begin{aligned} E(X^2Y) &= \{(x_1)^2(y_1)\}\,p(x_1,y_1) + \{(x_1)^2(y_2)\}p(x_1,y_2) \\ &\quad + \{(x_2)^2(y_1)\}p(x_2,y_1) + \{(x_2)^2(y_2)\}p(x_2,y_2) \\ &= (-2)^2(3)(0.28) + (-2)^2(6)(0.12) + (4)^2(3)(0.42) \\ &\quad\quad + (4)^2(6)(0.18) \\ &= 43.68 \end{aligned}$$

Download free eBooks at bookboon.com

Theorem 3.1
Let (X, Y) be two-dimensional discrete random variable and let $H(X, Y) = XY$. Then

$$E(XY) = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i y_j \, p(x_i, y_j)$$

Example 3.3
Refer to Example 2.2. Find $E(XY)$ from the parent joint probability distribution.

Solution
The parent distribution of XY is given in the table below.

X	Y			Row Totals
	-1	0	1	
-1	0.10 (1)	0.20 (0)	0.11 (-1)	0.41
0	0.08 (0)	0.02 (0)	0.26 (0)	0.36
2	0.03 (-2)	0.17 (0)	0.03 (2)	0.23
Column Totals	0.21	0.39	0.40	1.00

Download free eBooks at bookboon.com

The values in parentheses are for xy. Multiplying these values by their corresponding probabilities, we obtain the following table:

X	Y		
	-1	0	1
-1	0.10	0.00	-0.11
0	0.00	0.00	0.00
2	-0.06	0.00	0.06

Hence,

$$\begin{aligned} E(XY) &= \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j\, p(x_i, y_j) \\ &= 0.10 + 0.00 + (-0.11) + 0.00 + \cdots + 0.00 + 0.06 \\ &= -0.01 \end{aligned}$$

Corollary 3.1
Let (X, Y) be two-dimensional discrete random variable and let $H(X, Y) = X$. Then

$$E(X) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i\, p(x_i, y_j)$$

and

$$E(Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} y_j\, p(x_i, y_j)$$

Example 3.4
For the data in Example 2.2, find the expectation of the functions

(a) $H(X, Y) = X$

(b) $H(X, Y) = Y$.

Solution
(a) From the table in Example 3.3, we calculate the values of the cells with the x_i values. Thus, for $x_1 = -1$

$$\begin{aligned} x_1\, p(x_1, y_1) &= -1(0.10) = -0.1 \\ x_1\, p(x_1, y_2) &= -1(0.2) = -0.2 \\ x_1\, p(x_1, y_3) &= -1(0.11) = -0.11 \end{aligned}$$

so that the sum for that row is:

$$\sum_{j=1}^{3} x_1\, p(x_1, y_j) = -0.1 + (-0.2) + (-0.11) = -0.41$$

Similarly, for $x_2 = 0$

$$\sum_{j=1}^{3} x_2\, p(x_2, y_j) = 0 + 0 + 0 = 0$$

Download free eBooks at bookboon.com

and for $x_3 = 2$

$$\sum_{j=1}^{3} x_3\, p(x_3, y_j) = 0.06 + 0.34 + 0.06 = 0.46$$

The results are presented in the table below:

X	Y -1	0	1	Row Totals
-1	-0.1	-0.2	-0.11	-0.41
0	0	0	0	0
2	0.06	0.34	0.06	0.46

Hence

$$\begin{aligned} E(X) &= \sum_{i=1}^{3}\sum_{j=1}^{3} x_i\, p(x_i, y_j) \\ &= -0.41 + 0 + 0.46 \\ &= 0.05 \end{aligned}$$

(b) We compute $E(Y)$ as follows:
For $y_1 = -1$

$$\sum_{i=1}^{3} y_1\, p(x_i, y_1) = -0.1 + (-0.08) + (-0.03) = -0.21$$

Similarly, for $y_2 = 0$

$$\sum_{i=1}^{3} y_2\, p(x_i, y_2) = 0 + 0 + 0 = 0$$

and for $y_3 = 1$

$$\sum_{i=1}^{3} y_3\, p(x_i, y_3) = 0.11 + 0.26 + 0.03 = 0.40$$

The results are presented in the table below:

X	Y -1	0	1
-1	-0.1	0	0.11
0	-0.08	0	0.26
2	-0.03	0	0.03
Total	-0.21	0	0.40

Hence

$$\begin{aligned} E(Y) &= \sum_{j=1}^{3}\sum_{i=1}^{3} yj\, p(x_i, y_j) \\ &= -0.21 + 0 + 0.40 \\ &= 0.19 \end{aligned}$$

Download free eBooks at bookboon.com

Expectation from Derived Joint Probability Distribution

The following theorem shows that we can find $E(XY)$ of a bivariate distribution (X, Y) from the derived joint probability distribution without the knowledge of the parent joint probability distribution or the marginal distribution in the case of independence of X and Y.

Theorem 3.2
Let (X, Y) be two-dimensional discrete random variable and let $H(X, Y) = XY$. Then

$$E(XY) = \sum_{xy} xy\, p(xy)$$

Example 3.5
Refer to Example 2.2. Find $E(XY)$ from the derived joint probability distribution.

Solution
From Example 2.11, we have the following:

xy	-2	-1	0	1	2
$p(xy)$	0.03	0.11	0.73	0.1	0.03
$xy\,p(xy)$	-0.06	-0.11	0	0.1	0.06

Download free eBooks at bookboon.com

Click on the ad to read more

Hence

$$\begin{aligned} E(XY) &= \sum_{xy} xy\, p(xy) \\ &= -0.06 + (-0.11) + 0 + 0.1 + 0.06 \\ &= -0.01 \end{aligned}$$

the same result as in Example 3.3.

Expression of Expectation through Marginal Distribution

The following Theorem shows that the expectation of X and Y in a bivariate distribution can be obtained through the marginal distribution.

Theorem 3.3
Let (X, Y) be a two-dimensional random variable with probability mass function $p(x, y)$ and let $H(X, Y) = X$.
Then the function

$$E(X) = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i\, p(x_i, y_j)$$

simplifies to

$$E(X) = \sum_{i=1}^{n} x_i\, p(x_i)$$

Proof

$$\begin{aligned} E(X) &= \sum_{i=1}^{n} \sum_{i=1}^{m} x_i\, p(x_i, y_j) \\ &= \sum_{i=1}^{n} x_i \sum_{i=1}^{m} p(x_i, y_j) \end{aligned}$$

(since the x values do not depend on $y$and can therefore be taken from the summation of y)

$$= \sum_{i=1}^{n} x_i\, p(x_i)$$

(from the definition of the marginal probability mass function of X in Definition 1.11)

Similarly, if $H(X, Y) = Y$. Then

$$E(Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} y_i\, p(x_i, y_j)$$

which also reduces to

$$E(Y) = \sum_{j=1}^{m} y_j\, p(y_j)$$

Download free eBooks at bookboon.com

Example 3.6
Refer to Example 2.2. Using the marginal distributions, find
(a) $E(X)$ (b) $E(Y)$.

Solution
(a) From Example 2.11, the marginal distributions of X is:

x	-1	0	2
$p(x)$	0.41	0.36	0.23
$x\,p(x)$	-0.41	0	0.46

Hence

$$\begin{aligned} E(X) &= \sum_{i=1}^{n} x_i\, p(x_i) \\ &= -41 + 0 + 0.46 \\ &= 0.05 \end{aligned}$$

(b) From Example 3.3, the marginal distributions of Y is:

y	-1	0	1
$p(y)$	0.21	0.39	0.4
$y\,p(y)$	-0.21	0	0.4

Hence

$$\begin{aligned} E(Y) &= E(Y) = \sum_{j=1}^{m} y_j\, p(y_j) \\ &= -21 + 0 + 0.4 \\ &= 0.19 \end{aligned}$$

3.2.2 Expectation of Functions of Continuous Bivariate Random Variables

Definition 3.2
Let (X, Y) be a two-dimensional random variable with probability density function $f(x, y)$ and let $Z = H(X, Y)$.
Then

$$E(Z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H(x, y)\, f(x, y)\, dx\, dy$$

Download free eBooks at bookboon.com

Theorem 3.4
Let (X, Y) be a two-dimensional random variable with probability density function $f(x, y)$ and let $H(X, Y) = XY$. Then expectation of the product of X and Y is

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy\, f(x, y)\, dx\, dy$$

Example 3.7
Refer to Example 1.3. Find $E(XY)$.

Solution

$$\begin{aligned}
E(XY) &= \int_0^2 \int_0^1 xy \left(x^2 + \frac{xy}{3}\right) dx\, dy \\
&= \int_0^2 \left[\int_0^1 \left(x^3 y + \frac{x^2 y^2}{3}\right) dx\right] dy \\
&= \int_0^2 \left(\frac{y}{4} + \frac{y^2}{9}\right) dy \\
&= \frac{4}{8} + \frac{8}{27} = \frac{43}{54}
\end{aligned}$$

Click on the ad to read more

Definition 3.3
Let (X, Y) be a two-dimensional continuous random variable with probability density function $f(x, y)$. Then the expectation of the marginal distributions of X and Y are given by

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x\, g(x)\, dx \\ E(Y) &= \int_{-\infty}^{\infty} y\, h(y)\, dy \end{aligned}$$

where $g(x)$ and $h(y)$ are the marginal probability density function of X and Y respectively

Example 3.8
Refer to Example 1.3. Find (a) $E(X)$, (b) $E(Y)$

Solution
To find the expectation of a random variable from a bivariate distribution of X and Y we require the marginal probability distributions of X and Y.

The marginal probability distribution of X for Example 1.3 was found in Example 1.10 to be

$$g(x) = \begin{cases} 2\,x^2 + \dfrac{2}{3}\,x, & 0 < x < 1 \\ 0, & \text{elsewhere} \end{cases}$$

and that of Y to be

$$h(y) = \begin{cases} \dfrac{1}{3} + \dfrac{y}{6}, & 0 < x < 2 \\ 0, & \text{elsewhere} \end{cases}$$

Hence, by Definition 3.1

$$\begin{aligned} \text{(a)} \qquad E(X) &= \int_{-\infty}^{\infty} x\, g(x)\, dx \\ &= \int_0^1 x \left(2x^2 + \frac{2}{3}x\right) dx \\ &= \int_0^1 \left(2x^3 + \frac{2}{3}\,x^2\right) dx \\ &= \frac{13}{18} \\ \text{(b)} \qquad E(Y) &= \int_0^2 y \left(\frac{1}{3} + \frac{y}{6}\right) dy \\ &= \int_0^2 \left(\frac{1}{3}y + \frac{y^2}{6}\right) dy \\ &= \frac{10}{9} \end{aligned}$$

Download free eBooks at bookboon.com

3.2.3 Properties of Expectation of Bivariate Random Variables

Similar to the univariate case, we shall assume in our discussions of the properties of the expectation of bivariate random variables that both variables have finite expectations.

Property 1 **Expectation of Identical Distributions**

Theorem 3.5
If $X_1, X_2, ..., X_n$ have the same distribution, then they possess a common expectation $E(X_i) = \mu$

Note
The expression "have the same distribution" is equivalent to the expression "are identically distributed". Identical distribution means that the probability density function or the probability mass function of the random variable remains the same from trial to trial.

For example, X and Y are identically distributed if their probability distribution functions are

$$\begin{aligned} g(x) &= e^{-x}, \qquad x \geq 0 \\ h(y) &= e^{-y}, \qquad y \geq 0 \end{aligned}$$

Property 2 **Addition Law of Expectation**

Theorem 3.6
Let X and Y be any two random variables with expectation $E(X)$ and $E(Y)$, respectively. The expectation of their sum is the sum of their expectations. That is,

$$E(X + Y) = E(X) + E(Y)$$

Proof
We shall first prove for discrete case.

$$\begin{aligned} E(X+Y) &= \sum_{i=1}^{n}\sum_{j=1}^{m}(x_i + y_j)\,p(x_i, y_j) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m} x_i\,p(x_i, y_j) + \sum_{i=1}^{n}\sum_{j=1}^{m} y_j\,p(x_i, y_j) \\ &= \sum_{i=1}^{n} x_i \sum_{j=1}^{m} p(x_i, y_j) + \sum_{j=1}^{m} y_j \sum_{i=1}^{n} p(x_i, y_j) \end{aligned}$$

Download free eBooks at bookboon.com

$$
\begin{aligned}
&= \sum_{i=1}^{n} x_i\, p(x_i) + \sum_{j=1}^{m} y_j\, p(x_i) \\
&\qquad \text{(following from the logic of the proof of Theorem 3.3)} \\
&= E(X) + E(Y)
\end{aligned}
$$

We shall now prove for continuous case.

$$
\begin{aligned}
E(X+Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)f(x,y)\,dx\,dy \\
&= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f(x,y)\,dy\,dx + \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f(x,y)\,dx\,dy \\
&= \int_{-\infty}^{\infty} x\, f_X(x)\,dx + \int_{-\infty}^{\infty} y\, f_Y(y)\,dy
\end{aligned}
$$

By the definition of the marginal probability densities, the result follows.

Example 3.9
For the table in Example 2.2, verify that

$$E(X+Y) = E(X) + E(Y)$$

(a) using the parent joint probability distribution table;

(b) using the derived joint probability distribution table.

Download free eBooks at bookboon.com
Click on the ad to read more

Solution

(a) The distribution of $X+Y$ from the parent joint probability distribution table is reproduced from the solution for Example 2.2, taking note that the values for $x+y$ are those in parenthesis:

X	Y -1	0	1	Row Totals
-1	0.10 (-2)	0.20 (-1)	0.11 (0)	0.41
0	0.08 (-1)	0.02 (0)	0.26 (1)	0.36
2	0.03 (1)	0.17 (2)	0.03 (3)	0.23
Column Totals	0.21	0.39	0.40	1.00

To obtain $E(X+Y)$, we multiply the values for $x+y$ by their corresponding probabilities and obtain the following table:

X	Y -1	0	1
-1	-0.20	-0.20	0.00
0	-0.08	0.00	0.26
2	0.03	0.34	0.09

Hence,

$$\begin{aligned}
E(X+Y) &= \sum_{i=1}^{n}\sum_{j=1}^{m}(x_i+y_j)\,p(x_i,y_j) \\
&= -0.20+(-0.20)+0.00+(-0.08)\cdots+0.34+0.09 \\
&= 0.24
\end{aligned}$$

which is the same as

$$E(X)+E(Y) = 0.05+0.19 = 0.24$$

from Example 3.4.

(b) The distribution of $X+Y$ from the derived joint probability distribution table is taken from Example 2.2. Hence, to obtain $E(X+Y)$, we have the following table:

$(x+y)$	-2	-1	0	1	2	3
$p(x+y)$	0.1	0.28	0.13	0.29	0.17	0.03
$(x+y)\,p(x+y)$	-0.2	-0.28	0	0.29	0.34	0.09

Download free eBooks at bookboon.com

Hence,

$$\begin{aligned} E(X+Y) &= \sum (x_i + y_j)\, p(x_i + y_j) \\ &= -0.2 + (-0.28) + 0 + 0.29 + 0.34 + 0.09 \\ &= 0.24 \end{aligned}$$

Corollary 3.2
Let $X_1, X_2, ..., X_n$ be n random variables, where n is finite. Then the expectation of the sum $S_n = X_1 + X_2 + ... + X_n$ is the sum of the expectations of the individual random variables. Thus,

$$\begin{aligned} E(S_n) &= E(X_1 + X_2 + ... + X_n) \\ &= E(X_1) + E(X_2) + ... + E(X_n) \\ &= \sum_{i=1}^{n} E(X_i) \end{aligned}$$

Corollary 3.3
If $X_1, X_2, ..., X_n$ have the same distribution with $E(X_i) = \mu$ for all $1 \leq i \leq n$, then the expectation of their sum $S_n = X_1 + X_2 + ... + X_n$ is

$$E(S_n) = E(X_1 + X_2 + ... + X_n) = n\mu$$

Proof
Since $X_1, X_2, ..., X_n$ have the same distribution, from Theorem 3.3, they have a common expectation:

$$E(X_1) = E(X_2) = ... = E(X_n) = E(X) = \mu$$

From Corollary 3.2

$$E(X_1 + ... + X_n) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \mu = n\mu$$

Example 3.10
Refer to Example 2.13. Find $E(X+Y)$.

Solution

$$\begin{aligned} E(X+Y) &= E(X) + E(Y) \\ &= 1.6 + 1.6 \\ &= 2(1.6) = 3.2 \\ &= n\,\mu \end{aligned}$$

Download free eBooks at bookboon.com

Property 3 **Difference Law of Expectation**

Theorem 3.7
Let X and Y be any two random variables with expectations $E(X)$ and $E(Y)$ respectively. The expectation of their difference is the difference of their expectations. That is,

$$E(X - Y) = E(X) - E(Y)$$

Example 3.11
For the table in Example 2.2, verify that

$$E(X - Y) = E(X) - E(Y)$$

Solution
The distribution of $X - Y$ is given in Example 2.8. Hence, to obtain $E(X + Y)$, we have the following table:

$x - y$	-2	-1	0	1	2	3
$p(x - y)$	0.11	0.46	0.12	0.11	0.17	0.03
$(x - y)\, p(x - y)$	-0.22	-0.46	0	0.11	0.34	0.09

Click on the ad to read more
Download free eBooks at bookboon.com

Hence,

$$\begin{aligned} E(X-Y) &= -0.22+(-0.46)+0+0.11+0.34+0.09 \\ &= -0.14 \end{aligned}$$

which is the same as

$$E(X)-E(Y)=0.05-0.19=-0.14$$

from Example 3.4.

Property 4

Theorem 3.8
If a and b are constants, then for any random variables X and Y

$$E(aX+bY)=aE(X)+bE(Y)$$

Corollary 3.4
Let $X_1, X_2, ..., X_n$ be n random variables. Then

$$E(c_1X_1+c_2X_2+...+c_nX_n)=\sum_{i=1}^{n} c_i\,E(X_i)$$

where c_i's are constants.

If $c_i = c$ for all $1 \leq i \leq n$, then Corollary 3.4 becomes

$$E(c_1X_1+c_2X_2+...+c_nX_n)=c\sum_{i=1}^{n} E(X_i)$$

The proofs of Theorem 3.8 and its corollary are similar to the proofs of Theorem 3.6 and its corollaries.

Note
Theorems 3.6, 3.7 and 3.8 and their corollaries hold whether or not the random variables involved are independent.

Example 3.12
For the data of Example 3.1, if $a = 3$ and $b = 4$, verify whether Property 4 is valid.

Solution
From Example 3.1,

$$\begin{aligned} 3E(X)+4E(Y) &= 3(1.6)+4(3.9) \\ &= 20.4 \end{aligned}$$

Download free eBooks at bookboon.com

Now,

$3x+4y$	$3(-2)+4(3)$	$3(-2)+4(6)$	$3(4)+4(3)$	$3(4)+4(6)$
$p(x,y)$	$0.4(0.7)$	$0.4(0.3)$	$0.6(0.7)$	$0.6(0.3)$

or

$3x+4y$	6	18	24	36
$p(x+y)$	0.28	0.12	0.42	0.18

so that

$$\begin{aligned} E(3X+4Y) &= 6(0.28)+18(0.12)+24(0.42)+36(0.18) \\ &= 1.68+2.16+10.08+6.48 \\ &= 20.4 \end{aligned}$$

Thus, Property 4 is valid.

Property 5 **Expectation of Product of X and Y**

Theorem 3.9
Let X and Y be two random variables with expectations $E(X)$ and $E(Y)$, respectively. Then the expectation of their product is

$$E(XY) = E(X)E(Y) + E\{[X-E(X)][Y-E(Y)]\}$$

The last term appears sufficiently often to deserve a separate name and treatment. It is called the covariance of X and Y and denoted as $\text{Cov}(X,Y)$. The concept of covariance is discussed in detail in the next chapter.

Example 3.13
For the table in Example 2.2, obtain $E\{[X-E(X)][Y-E(Y)]\}$.

Solution
It has been proved in Theorem 4.1 in the sequel that

$$E\{[X-E(X)][Y-E(Y)]\} = E(XY) - E(X)E(Y)$$

From Examples 3.3,

$$E(XY) = -0.01$$

From Examples 3.4,

$$E(X) = 0.05 \quad \text{and} \quad E(Y) = 0.19$$

Hence,

$$\begin{aligned} E\{[X-E(X)][Y-E(Y)]\} &= E(XY) - E(X)E(Y) \\ &= -0.01 + 0.05(0.19) \\ &= -0.0005 \end{aligned}$$

Download free eBooks at bookboon.com

Property 6 **Expectation of Product of Independent Random Variables**

Theorem 3.10
Let X and Y be two independent random variables with expectations $E(X)$ and $E(Y)$ respectively. Then the expectation of their product is equal to the product of their expectations:

$$E(XY) = E(X)E(Y)$$

Proof (for the discrete bivariate case)
The random variable XY can take all values $x_i y_j$, where $i = 1, 2, ..., n$ and $j = 1, 2, ..., m$ with probabilities $g(x_i)h(y_j)$, as long as X and Y are independent random variables. Hence

$$\begin{aligned} E(XY) &= \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j\, g(x_i)\, h(y_j) \\ &= \left(\sum_{i=1}^{n} x_i\, g(x_i)\right)\left(\sum_{j=1}^{m} y_j\, h(y_j)\right) \\ &= E(X)E(Y) \end{aligned}$$

The proof of the bivariate continuous case is similar.

Click on the ad to read more

Note
The converse of Theorem 3.10 does not hold, that is, the random variables X and Y may satisfy the relation $E(XY) = E(X)E(Y)$ without being independent.

Example 3.14
For the data in Example 2.3, verify whether

$$E(XY) = E(X)E(Y)$$

Solution
The distribution of XY is given in Example 2.12. Hence, to obtain $E(XY)$, we have the following table:

xy	-6	-12	12	24
$p(xy)$	0.28	0.12	0.42	0.18
$xy\,p(xy)$	-1.68	-1.44	5.08	4.32

Hence,

$$\begin{aligned} E(XY) &= -1.68 + (-1.44) + 5.08 + 4.32 \\ &= 6.24 \end{aligned}$$

From Example 3.1,

$$E(X) = 1.6 \text{ and } E(Y) = 3.9$$

It has been shown in Example 2.3 that X and Y are independent. Hence,

$$E(X)E(Y) = 6.24$$

which is the same as $E(XY)$ above.

Corollary 3.5
Let $X_1, X_2, \cdots, X_n$ be independent random variables. Then

$$E(X_1 X_2 \cdots X_n) = E(X_1)E(X_2)...E(X_n)$$

This result may be written as

$$E\left(\prod_{i=1}^{n}(X_i)\right) = \prod_{i=1}^{n} E(X_i)$$

where $\prod$ is the "product of".

Proof
We shall demonstrate for the case of a three random variables X, Y and Z.

The product of three random variables $X\,Y\,Z$ may be written as $(XY)Z$. Using Theorem 3.10, we obtain:

$$E(XYZ) = E[(XY)Z] = E(XY)E(Z)$$

Applying again the same theorem to the random variable XY which is a product of random variables X and Y we obtain

$$E(XYZ) = E(X)E(Y)E(Z)$$

Download free eBooks at bookboon.com

This corollary can be proved by mathematical induction. It is easy to prove Theorem 3.10 for any finite number of independent random variables.

Property 7 **Expectation of Quotient of X and Y**

Theorem 3.11
Let X and Y be two random variables with expectations $E(X)$ and $E(Y)$ respectively. Then the expectation of their quotient is

$$E\left(\frac{X}{Y}\right) \approx \frac{E(X)}{E(Y)} - \frac{1}{E(Y)^2}\operatorname{Cov}(X,Y) + \frac{E(X)}{E(Y)^3}\operatorname{Var}(Y)$$

$E(X) \neq 0$, $E(Y) \neq 0$, where $\operatorname{Cov}(X,Y)$ is called the *covariance* of X and Y (see next chapter)

Note
The expectation of the quotient $\dfrac{X}{Y}$ may not exist even though the moments of X and Y exist.

Property 8 **Monotonicity of Expectation**

Theorem 3.12
Suppose X and Y are two random variables with expectations $E(X)$ and $E(Y)$ respectively. If $X \leq Y$, then

$$E(X) \leq E(Y)$$

Proof
We write

$$Y = X + (Y - X)$$

where $Y - X \geq 0$. By Theorem 3.5, we have

$$E(Y) = E(X) + E(Y - X) \geq E(X)$$

Property 9 **Hölder's Inequality of Expectation**

Theorem 3.13
Suppose X and Y are bivariate random variables with finite p^{th} and q^{th} order moments respectively, where p and q are positive numbers satisfying $\dfrac{1}{p} + \dfrac{1}{q} = 1$. Then XY has a finite first moment and

$$E(|XY|) \leq E(|X|^p)^{\frac{1}{p}} E(|Y|^q)^{\frac{1}{q}}$$

Download free eBooks at bookboon.com

The equality holds if and only if $E(|Y|^q) = c\,E(|X|^p)$ for some constant c or $X = 0$.

Property 10 **Cauchy-Schwartz' Inequality of Expectation**

Theorem 3.14
Suppose X and Y are two random variables with second moments $E(X^2)$ and $E(Y^2)$ respectively. Then $E(XY)$ exists and

$$[E(XY)]^2 \leq E(X^2)E(Y^2)$$

Proof
Case 1
If $E(X^2)$ or $E(Y^2)$ is infinite, the theorem is valid.

Case 2
If $E(X^2) = 0$ then we must have $P(X = 0) = 1$ so that $P(XY = 0) = 1$ and $E(XY) = 0$, hence again the theorem is valid. The same argument applies if $E(Y^2) = 0$, so that we may assume that $0 < E(X^2) < \infty$ and $0 < E(Y^2) < \infty$.

Click on the ad to read more

Case 3
Suppose we have a random variable

$$Z_t = tX + Y$$

Let us consider the real valued function

$$E(Z_t)^2 = E(tX+Y)^2$$

For every t we have

$$E(Z_t)^2 = E(tX+Y)^2 \geq 0$$

since

$$(tX+Y)^2 \geq 0$$

so that

$$E(t^2X^2 + 2\,tXY + Y^2) \geq 0 \qquad (i)$$

The term on the left hand side of (i) is a quadratic function in t, which is always nonnegative for all t. Hence, it does not have more than one real root. Its discriminant must, therefore, be less than or equal to zero. The discriminant "$b^2 - 4ac$" for the quadratic equation (i) in t is

$$[2\,E(XY)]^2 - 4\,E(X^2)E(Y^2) \leq 0$$

or

$$4\,[E(XY)]^2 \leq 4\,E(X^2)E(Y^2)$$

from which the result follows.

Note

(a) Equality of Theorem 3.14 holds if and only if one of the random variables equals a constant multiple of the other, say, $Y = t\,X$ for some constant t or, at least one of them is non-zero, say, $X \neq 0$;

(b) This inequality is sometimes simply called *Schwartz' inequality*;

(c) Cauchy-Schwartz' inequality is a special case of Höder's inequality when $p = q = 2$.

Property 11 **Minkowski's Inequality of Expectation**

Theorem 3.15
Suppose X and Y are two random variables with finite p^{th} order moments. Then for $p \geq 1$

$$[E(|X+Y|^p)]^{\frac{1}{p}} \leq E(|X|^p)^{\frac{1}{p}} + E(|Y|^p)^{\frac{1}{p}}$$

Corollary 3.6
Suppose X and Y are two random variables with expectations $E(X)$ and $E(Y)$ respectively. Then

$$E(|X+Y|) \leq E(|X|) + E(|Y|)$$

The proof follows immediately from Theorem 3.14 for the case when $p = 1$.

Download free eBooks at bookboon.com

Property 12 **Expectation of Frequency of Success**

Theorem 3.16
Suppose the random variable M is the frequency of occurrence of the event $\mathcal{A}$ (number of successes) in n independent and identical trials. Then

$$E(M) = np$$

Proof
Let $X_1, X_2, \cdots, X_n$ be n independent random variables which are identically distributed. Suppose

$$X_i = \begin{cases} 1, & \text{if event } \mathcal{A} \text{ occurs in the } i^{th} \text{ trial,} \\ 0, & \text{otherwise} \end{cases}$$

Then X_i are Bernoulli random variables and

$$E(X_i) = p_i = p$$

Letting

$$M = X_1 + X_2 + \cdots + X_n$$

we have

$$\begin{aligned} E(M) &= E(X_1) + E(X_2) + \cdots + E(X_n) \\ &= np \end{aligned}$$

Example 3.15
A box contains 20 black, 20 red and 10 green balls. From the box, 25 balls are randomly selected with replacement.

Find the expectation of the frequency of occurrence of the red balls.

Solution
Let $\mathcal{A} = \{$the occurrence of red balls$\}$. Then

$$p = P(\mathcal{A}) = \frac{20}{50} = \frac{2}{5}$$

Hence the expectation of the frequency of occurrence of the red balls is

$$\begin{aligned} E(M) &= np \\ &= 25\left(\frac{2}{5}\right) = 10 \end{aligned}$$

That is, if we select 25 balls from a box containing 20 black, 20 red and 10 green balls at random with replacement, we are likely to have the red ball appearing 10 times.

Download free eBooks at bookboon.com

Property 13 Expectation of Relative Frequency of Success

Theorem 3.17
Let the random variable $\dfrac{M}{n}$ be the relative frequency of the event $\mathcal{A}$ (or proportion of success) among the n independent and identical trials of experiment $\mathcal{E}$. Then the expectation of the relative frequency equals the probability of the event:

$$E\left(\frac{M}{n}\right) = p$$

Proof
Thus, using Property 4 of expectation (Theorem 3.7)

$$\begin{aligned} E\left(\frac{M}{n}\right) &= \frac{1}{n}E(M) \\ &= \frac{1}{n}(n\,p) \quad \text{(from Theorem 3.15)} \\ &= p \end{aligned}$$

The above theorem says that the expected relative frequency of the event $\mathcal{A}$ is p, where $p = P(\mathcal{A})$. This is intuitively clear and it establishes a connection between the relative frequency of an event and the probability of that event.

Click on the ad to read more

Example 3.16
Refer to Example 3.15. Find the expectation of the relative frequency of the occurrence of the red balls.

Solution
From Example 3.15,

$$p = \frac{2}{5}$$

so that

$$q = P(\overline{\mathcal{A}}) = \frac{3}{5}$$

Hence, the expectation of the relative frequency of occurrence of the red balls is

$$E\left(\frac{M}{n}\right) = p = \frac{2}{5}$$

3.3 VARIANCE OF BIVARIATE RANDOM VARIABLES

3.3.1 Variance of Functions of Bivariate Random Variables

The variance of a univariate random variable has been discussed in the book on "Introductory Probability Theory" (Nsowah-Nuamah, 2017). We extend the discussion to bivariate distributions in this section. A few more properties will also be discussed.

The variance of a bivariate random variables X and Y can be expressed through the ***joint probability functions***.

Definition 3.4
Let (X, Y) be a two-dimensional discrete random variable with probability mass function $p(x_i, y_j)$.
If $H(X, Y) = (X - \mu_X)^2$, then

$$\text{Var}(X) = \sum_{i=1}^{n} \sum_{j=1}^{m} (X - \mu_X)^2 \, p(x_i, y_j)$$

If $H(X, Y) = (Y - \mu_Y)^2$, then

$$\text{Var}(Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} (Y - \mu_Y)^2 \, p(x_i, y_j)$$

where $\mu_X = E(X)$ and $\mu_Y = E(Y)$

Download free eBooks at bookboon.com

Example 3.17
For the data in Example 2.3, find the variance of the (a) X (b) Y.
Solution
(a) From Example 3.1, $E(X) = 1.6$. Hence,

$$\begin{aligned}\text{Var}(X) &= (x_1-\mu_X)^2\, p(x_1,y_1) + (x_1-\mu_X)^2 p(x_1,y_2)\\ &\quad +(x_2-\mu_X)^2 p(x_2,y_1) + (x_2-\mu_X)^2 p(x_2,y_2)\\ &= (-2-1.6)^2(0.28) + (-2-1.6)^2(0.12) + (4-1.6)^2(0.42)\\ &\quad +(4-1.6)^2(0.18)\\ &= 8.64\end{aligned}$$

(b) From Example 3.1, $E(Y) = 3.9$. Hence,

$$\begin{aligned}\text{Var}(Y) &= (y_1-\mu_Y)^2\, p(x_1,y_1) + (y_1-\mu_Y)^2 p(x_1,y_2)\\ &\quad +(y_2-\mu_Y)^2 p(x_2,y_1) + (y_2-\mu_Y)^2 p(x_2,y_2)\\ &= (3-3.9)^2(0.7) + (6-3.9)^2(0.12) + (3-3.9)^2(0.42)\\ &\quad +(6-3.9)^2(0.18)\\ &= 2.23\end{aligned}$$

Definition 3.5
Let (X, Y) be a two-dimensional continuous random variables with probability density function $f(x, y)$.
If $H(X, Y) = (X - \mu_X)^2$, then

$$\text{Var}(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (X-\mu_X)^2\, p(x_i, y_j)$$

If $H(X, Y) = (Y - \mu_Y)^2$, then

$$\text{Var}(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (Y-\mu_Y)^2\, p(x_i, y_j)$$

We may also express the variance of X and Y in a bivariate distribution through the ***marginal probability functions***.

Definition 3.6
Let (X, Y) be a two-dimensional discrete random variable with probability mass function $p(x, y)$. Then the variance of the marginal distributions of X and Y are given by

$$\begin{aligned}\text{Var}(X) &= \sum_{\text{all } x} (x-\mu_X)^2 p(x)\\ \text{Var}(Y) &= \sum_{\text{all } y} (y-\mu_Y)^2 p(y)\end{aligned}$$

where $p(x)$ and $p(y)$ are the marginal probability mass function of X and Y, respectively

Download free eBooks at bookboon.com

Example 3.18
For the data in Example 2.3, find (a) $\text{Var}(X)$ (b) $\text{Var}(Y)$.

Solution

$$\begin{aligned} \text{Var}(X) &= \sum_{\text{all } x} (x - \mu_X)^2 p(x) \\ &= (-2 - 1.6)^2(0.4) + (4(-2 - 1.6)^2(0.6) \\ &= 8.64 \end{aligned}$$

The reader is asked to calculate (b).

Definition 3.7
Let (X, Y) be a two-dimensional continuous random variable with probability density function $f(x, y)$. Then the variance of the marginal distributions of X and Y are given by

$$\begin{aligned} \text{Var}(X) &= \int_{-\infty}^{\infty} (x - \mu_X)^2\, g(x)\, dx \\ \text{Var}(Y) &= \int_{-\infty}^{\infty} (y - \mu_Y)^2 h(y)\, dy \end{aligned}$$

where $g(x)$ and $h(y)$ are the marginal probability density function of X and Y respectively

Click on the ad to read more

Download free eBooks at bookboon.com

Like in the case of the univariate random variable, for computational purposes, we use the following formulas:

$$\begin{aligned}\text{Var}(X) &= E(X^2) - [E(X)]^2 \\ &= \sum_{\text{all } x} x^2 p(x) - (E(X))^2 \quad \text{(Discrete case)} \\ &= \int_{-\infty}^{\infty} x^2 g(x)\,dx - (E(X))^2 \quad \text{(Continuous case)}\end{aligned}$$

$$\begin{aligned}\text{Var}(Y) &= E(Y^2) - [E(Y)]^2 \\ &= \sum_{\text{all } y} y^2 p(y) - (E(Y))^2 \quad \text{(Discrete case)} \\ &= \int_{-\infty}^{\infty} y^2 h(y)\,dy - (E(Y))^2 \quad \text{(Continuous case)}\end{aligned}$$

Example 3.19
Refer to Example 1.3. Find (a) Var(X) (b) Var(Y).

Solution
(a) From Example 3.6, $E(X) = \dfrac{13}{18}$.
Now

$$\begin{aligned}E(X^2) &= \int_0^1 x^2\left(2\,x^2 + \frac{2}{3}\,x\right)dx \\ &= \int_0^1 \left(2\,x^4 + \frac{2}{3}\,x^3\right)dx \\ &= \left[\frac{2\,x^5}{5} + \frac{2\,x^4}{12}\right]_0^1 \\ &= \frac{2}{5} + \frac{2}{12} \\ &= \frac{17}{30}\end{aligned}$$

Therefore

$$\text{Var}(X) = \frac{17}{30} - \left(\frac{13}{18}\right)^2 = 0.04506$$

(b) From Example 3.5, $E(Y) = \dfrac{10}{9}$.
Now

$$\begin{aligned}E(Y^2) &= \int_0^2 y^2\left(\frac{1}{3} + \frac{y}{6}\right)dy \\ &= \int_0^2 \left(\frac{1}{3}y^2 + \frac{y^3}{6}\right)dy \\ &= \left[\frac{y^3}{9} + \frac{y^4}{24}\right]_0^2\end{aligned}$$

Download free eBooks at bookboon.com

$$\begin{aligned} &= \frac{8}{9} + \frac{16}{24} \\ &= \frac{14}{9} \end{aligned}$$

Hence

$$\text{Var}(Y) = \frac{14}{9} - \left(\frac{10}{9}\right)^2 = 0.32099$$

3.3.2 Properties of Variance of Bivariate Random Variables

In discussing the properties of variance, it is assumed that the variances exist.

Property 1 **Variance of Identical and Independent Random Variables**

Theorem 3.18
If the random variables $X_1, X_2, ..., X_n$ have the same distribution, then they all have the same variance σ^2, that is,

$$\text{Var}(X_i) = \sigma^2, \qquad 1 \le i \le n$$

Property 2 **Variance of Sum and Difference of X and Y**

Theorem 3.19
If (X,Y) are two dimensional random variables, then

$$\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y) \pm 2\,\text{Cov}(X, Y)$$

where $\text{Cov}(X, Y)$, is the *covariance* of X and Y

Proof
We shall give the proof for the case $X + Y$. The proof for the case $X - Y$ should be tried by the reader (see Exercise 3.13).

$$\begin{aligned} \text{Var}(X+Y) &= E\{[X+Y-E(X+Y)]^2\} \\ &= E\{[X+Y-E(X)-E(Y)]^2\} \\ &= E\{([X-E(X)]+[Y-E(Y)])^2\} \\ &= E\{(X-E(X))^2+(Y-E(Y))^2 \\ &\qquad +2(X-E(X))(Y-E(Y))\} \\ &= E\{X-E(X)\}^2+E\{Y-E(Y)\}^2 \\ &\qquad +2\{E(X-E(X))(Y-E(Y))\} \\ &= \text{Var}(X)+\text{Var}(Y)+2\,\text{Cov}(X,Y) \end{aligned}$$

Download free eBooks at bookboon.com

Example 3.20
If $\text{Var}(X) = 8$, $\text{Var}(Y) = 5$ and $\text{Cov}(X, Y) = 3$ find
(a) $\text{Var}(X + Y)$, (b) $\text{Var}(X - Y)$

Solution

(a) $\text{Var}(X + Y) = 8 + 5 + 2(3) = 19$

(b) $\text{Var}(X - Y) = 8 + 5 - 2(3) = 7$

The variance is not additive in general, as the expected value. However, with the additional assumption of independence, we obtain Theorem 3.20.

Corollary 3.7
If $X_1, X_2, ..., X_n$ are n dimensional random variables, then

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i) + 2\sum_{\substack{i=1\\i<j}}^{n}\sum_{j=1}^{n} \text{Cov}(X_i, X_j),$$

or equivalently,

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i) + \sum_{\substack{i=1\\i\neq j}}^{n}\sum_{j=1}^{n} \text{Cov}(X_i, X_j),$$

or equivalently,

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \text{Cov}(X_i, X_j)$$

Need help with your dissertation?

Get in-depth feedback & advice from experts in your topic area. Find out what you can do to improve the quality of your dissertation!

Get Help Now

Go to www.helpmyassignment.co.uk for more info

Helpmyassignment

Download free eBooks at bookboon.com

Click on the ad to read more

Property 3 **Sum and Difference of Independent Random Variables**

Theorem 3.20
If (X, Y) are two dimensional random variables and if X and Y are independent then,

$$\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y)$$

Proof
The theorem follows from Theorem 3.19 by noting that $\text{Cov}(X, Y) = 0$ for the case when X and Y are independent (see Theorem 4.7 in the sequel.)

Corollary 3.8 **Bienaymé Equality**
Let X_i $(i = 1, 2, ..., n)$ be independent random variables. Then

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i)$$

This corollary follows from Corollary 3.7.

Note
It is important to observe that

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$$

whether or not the X_i's are independent but it is generally *not* the case that

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i)$$

always.

Example 3.21
For the data in Example 2.3, verify that

(a) $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$;

(b) $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y)$.

Solution
It has been shown in Example 2.3 that X and Y are independent, hence:

(a) $\text{Var}(X + Y) = E(X + Y)^2 - [E(X + Y)]^2$.

Now

$(x_i + y_j)^2$	$(-2 + 3)^2$	$(-2 + 6)^2$	$(4 + 3)^2$	$(4 + 6)^2$
$p(x_i)p(y_j)$	0.4(0.7)	0.4(0.3)	0.6(0.7)	0.6(0.3)

Download free eBooks at bookboon.com

Therefore, the distribution of $(X+Y)^2$ is

$(x_i+y_j)^2$	1	16	49	100
$p(x_i)p(y_j)$	0.28	0.12	0.42	0.18

$$\begin{aligned} E\left[(X+Y)^2\right] &= \sum (x+y)^2\, p(x)p(y) \\ &= 1(0.28)+16(0.12)+49(0.42)+100(0.18) \\ &= 0.28+1.92+20.58+18 \\ &= 40.78 \end{aligned}$$

From Example 3.7, $E(X+Y)=5.50$.

Therefore,

$$\begin{aligned} \text{Var}(X+Y) &= E\left[(X+Y)^2\right]-[E(X+Y)]^2 \\ &= 40.78-(5.5)^2 \\ &= 10.53 \end{aligned}$$

Now

$$\text{Var}(X)=E(X^2)-[E(X)]^2$$

From Example 3.1, $E(X)=1.6$.

From Example 2.3, we shall find $E(X^2)$. Now,

x_i^2	$(-2)^2$	$(4)^2$
$p(x_i)$	0.4	0.6

$$\begin{aligned} E(X^2) &= \sum_{i=1}^{n} x^2\, p(x) \\ &= (-2)^2(0.4)+(4)^2(0.6) \\ &= 1.6+9.6=11.2 \end{aligned}$$

Therefore

$$\text{Var}(X)=11.2-(1.6)^2=8.64$$

Again

$$\text{Var}(Y)=E(Y^2)-[E(Y)]^2$$

From Example 3.1, $E(Y)=3.9$.

Similarly,

y_j^2	$(3)^2$	$(6)^2$
$p(y_j)$	0.7	0.3

$$\begin{aligned} E(Y^2) &= \sum y^2\, h(y) \\ &= (3)^2(0.7)+(6)^2(0.3) \\ &= 6.3+10.8=17.1 \end{aligned}$$

Download free eBooks at bookboon.com

Therefore

$$\text{Var}(Y) = 17.1 - (3.9)^2 = 1.89$$

Hence,

$$\text{Var}(X+Y) = \text{Var(X)} + \text{Var(Y)}$$

(b) $\quad \text{Var}\,[(X-Y)] = E(X-Y)^2 - [E(X-Y)]^2$

Now

$(x_i - y_j)^2$	$(-2-3)^2$	$(-2-6)^2$	$(4-3)^2$	$(4-6)^2$
$p(x_i)p(y_j)$	$0.4(0.7)$	$0.4(0.3)$	$0.6(0.7)$	$0.6(0.3)$

$(x_i - y_j)^2$	25	64	1	4
$p(x_i)p(y_j)$	0.28	0.12	0.42	0.18

$$\begin{aligned} E(X-Y)^2 &= \sum (x-y)^2\, g(x)\, h(y) \\ &= 25(0.28) + 64(0.12) + 1(0.42) + 4(0.18) \\ &= 7 + 7.68 + 0.42 + 0.72 \\ &= 3.82 \end{aligned}$$

$x_i - y_j$	$-2-3$	$-2-6$	$4-3$	$4-6$
$p(x_i)p(y_j)$	$0.4(0.7)$	$0.4(0.3)$	$0.6(0.7)$	$0.6(0.3)$

$x_i - y_j$	-5	-8	1	-2
$p(x_i)p(y_j)$	0.28	0.12	0.42	0.18

Download free eBooks at bookboon.com

Click on the ad to read more

$$\begin{aligned}E\left[(X-Y)\right] &= \sum (x-y)\,g(x)\,h(y)\\ &= -5(0.28)+(-8)(0.12)+1(0.42)+-2(0.18)\\ &= -1.4-0.96+0.42-0.36\\ &= -2.3\end{aligned}$$

Therefore,

$$[E(X-Y)]^2 = (-2.30)^2 = 5.29$$

Hence,

$$\text{Var}(X-Y) = 15.82 - 5.29 = 10.53$$

But

$$\text{Var}(X) + \text{Var}(Y) = 8.64 + 1.89 = 10.53$$

Hence,

$$\text{Var}(X-Y) = \text{Var}(X) + \text{Var}(Y)$$

The next property is of much importance in both theoretical and applied statistics.

Property 4

Theorem 3.21
If the random variables $X_i\,(i = 1, 2, ..., n)$ are independent and have the same variance, then

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = n\,\sigma^2$$

$\text{Var}(X_i) = \sigma^2$

This theorem follows immediately from Corollary 3.8.

Property 5

Theorem 3.22
If (X, Y) is a two-dimensional random variable, then

$$\text{Var}(aX \pm bY) = a^2\,\text{Var}(X) \pm b^2\,\text{Var}(Y) \pm 2\,a\,b\,\text{Cov}(X, Y)$$

where a and b are constant

Download free eBooks at bookboon.com

Proof
We shall prove for the sum. The proof for the difference is similar.

$$\begin{aligned}\text{Var}(aX+bY) &= E\{[(aX+bY)-E(aX+bY)]^2\}\\ &= E\{a[X-E(X)]+b[Y-E(Y)]^2\}\\ &= a^2E[X-E(X)]^2+b^2E[Y-E(Y)]^2\\ &\quad +2\,a\,b\,E\{[X-E(X)][Y-E(Y)]\}\\ &= a^2\text{Var}(X)+b^2\text{Var}(Y)+2\,a\,b\,\text{Cov}(X,Y)\end{aligned}$$

More generally, we have the following results for the variance of a linear combination of random variables.

Corollary 3.9
If $X_i\,(i=1,2,...,n)$ are n random variables and c_i a constant associated with i^{th} random variable, then

$$\text{Var}\left(\sum_{i=1}^{n} c_iX_i\right) = \sum_{i=1}^{n} c_i^2\,Var(X_i) + 2\sum_{\substack{i=1\\i<j}}^{n}\sum_{j=1}^{n} c_i\,c_j\,\text{Cov}(X_i,X_j)$$

or equivalently,

$$\text{Var}\left(\sum_{i=1}^{n} c_iX_i\right) = \sum_{i=1}^{n} c_i^2\text{Var}(X_i) + \sum_{\substack{i=1\\i\neq j}}^{n}\sum_{j=1}^{n} c_i\,c_j\,\text{Cov}(X_i,X_j)$$

or equivalently,

$$\text{Var}\left(\sum_{i=1}^{n} c_iX_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i\,c_j\text{Cov}(X_i,X_j)$$

Example 3.22
If $\text{Var}(X)=5$, $\text{Var}(Y)=3$ and $\text{Cov}(X,Y)=2$, find

(a) $\text{Var}(4X+6Y)$;

(b) $\text{Var}(4X-6Y)$.

Solution (a)

$$\begin{aligned}(a)\quad \text{Var}(4X+6Y) &= 4^2\,\text{Var}(X)+6^2\,\text{Var}(Y)+2(4)(6)\,\text{Cov}(X,Y)\\ &= 16(5)+36(3)+48(2)\\ &= 284\\ (b)\quad \text{Var}(4X-6Y) &= 4^2\,\text{Var}(X)+6^2\,\text{Var}(Y)-2(4)(6)\,\text{Cov}(X,Y)\\ &= 16(5)+36(3)-48(2)\\ &= 92\end{aligned}$$

Property 6

Theorem 3.23
Let X and Y be two independent random variables. Then

$$\text{Var}(aX\pm bY) = a^2\,\text{Var}(X)+b^2\,\text{Var}(Y)$$

where a and b are constants

Download free eBooks at bookboon.com

Corollary 3.10
Let X_i $(i = 1, 2, ..., n)$ be independent random variables and c_i a constant associated with i^{th} random variable, then

$$\mathrm{Var}\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} c_i^2\, \mathrm{Var}(X_i)$$

From Corollary 3.10, if $c_i = c$ then

$$\mathrm{Var}\left(\sum_{i=1}^{n} c\, X_i\right) = c^2 \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

Example 3.23
If $\mathrm{Var}(X) = 10$ and $\mathrm{Var}(Y) = 6$, find

(a) $\mathrm{Var}(3X + 2Y)$, (b) $\mathrm{Var}(2X + 2Y)$,

(c) $\mathrm{Var}(3X - 2Y)$, (d) $\mathrm{Var}(2X - 2Y)$.

Solution

$$\begin{aligned} \text{(a)} \quad \mathrm{Var}(3X + 2Y) &= 3^2\mathrm{Var}(X) + 2^2\mathrm{Var}(Y) \\ &= 9(10) + 4(6) \\ &= 112 \end{aligned}$$

Download free eBooks at bookboon.com
Click on the ad to read more

$$
\begin{aligned}
\text{(b)} \quad \mathrm{Var}(2X+2Y) &= \mathrm{Var}[2(X+Y)] \\
&= 2^2\,\mathrm{Var}(X+Y) \\
&= 2^2[\mathrm{Var}(X)+\mathrm{Var}(Y)] \\
&= 4(10+6) \\
&= 64
\end{aligned}
$$

The reader should solve (c) and (d).

Property 7 **Variance of Frequency of Success**

Theorem 3.24
Let the random variable M be the frequency of success in n independent trials, then

$$\mathrm{Var}(M) = npq$$

Proof
We found in the proof of Theorem 3.16 that $E(X_i) = p$. To find the variance, we need to find also $E(X_i^2)$:

x_i^2	0	1
$p(x_i)$	q	p

$$E(X_i^2) = 1(p) + 0(q) = p$$

so that

$$
\begin{aligned}
\mathrm{Var}(X_i) &= E(X_i^2) - [E(X_i)]^2 \\
&= p - p^2 = pq
\end{aligned}
$$

Hence

$$
\begin{aligned}
\mathrm{Var}(M) &= \mathrm{Var}(X_1 + X_2 + ... + X_n) \\
&= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n) \\
&= npq
\end{aligned}
$$

Example 3.24
Refer to example 3.15, find $\mathrm{Var}(M)$.

Solution
$n = 25, \quad p = \dfrac{2}{5}, \quad q = \dfrac{3}{5}$

$$
\begin{aligned}
\mathrm{Var}(M) &= npq \\
&= 25\left(\frac{2}{5}\right)\left(\frac{3}{5}\right) \\
&= 6
\end{aligned}
$$

Download free eBooks at bookboon.com

Property 8 **Variance of Relative Frequency of Success**

Theorem 3.25
Let the random variable $\frac{M}{n}$ be the relative frequency of success in n independent trials, then

$$\text{Var}\left(\frac{M}{n}\right) = \frac{pq}{n}$$

Proof

$$\begin{aligned}\text{Var}\left(\frac{M}{n}\right) &= \frac{1}{n^2}\text{Var}(M) \\ &= \frac{npq}{n^2} \quad \text{(by Theorem 3.24)} \\ &= \frac{pq}{n}\end{aligned}$$

Example 3.25
Refer to example 3.15, find $\text{Var}\left(\frac{M}{n}\right)$.

Solution
$n = 25, \quad p = \frac{2}{5}, \quad q = \frac{3}{5}$

$$\begin{aligned}\text{Var}\left(\frac{M}{n}\right) &= \frac{pq}{n} \\ &= \frac{\left(\frac{2}{5}\right)\left(\frac{3}{5}\right)}{25} \\ &= 0.0096\end{aligned}$$

In this chapter we have discussed some moments of bivariate random variables, namely, the expectation and variance. In the next chapter, we shall continue with the discussion of moments and how that can be used to measure the relationship between two random variables.

EXERCISES

3.1 Refer to the data of Exercise 1.1, find

$$\begin{array}{ll} (a)\ E(X+Y) & (b)\ E(X-Y) \\ (c)\ E(XY) & (d)\ E\left(\frac{3X}{7Y}\right) \\ (e)\ E(2X+3Y) & (f)\ E(5X-Y) \\ (g)\ E[(3X)(4Y)] & (h)\ E\left(\frac{3X}{7Y}\right) \\ (i)\ E(X^2+5Y) & (j)\ E[3(X+Y)] \\ (k)\ E(Y^3) & (l)\ E(Y^2-2Y+3X) \end{array}$$

Download free eBooks at bookboon.com

3.2 Refer to Exercise 1.1. Suppose X and Y are not assumed to be independence, find

(a) $\text{Var}(X+Y)$ (b) $\text{Var}(X-Y)$
(c) $\text{Var}(XY)$ (d) $\text{Var}\left(\dfrac{X}{Y}\right)$
(e) $\text{Var}(2X+3Y)$ (f) $\text{Var}(5X-Y)$
(g) $\text{Var}[(3X)(4Y)]$
(h) $\text{Var}\left(\dfrac{3X}{7Y}\right)$
(i) $\text{Var}(X^2+5Y)$
(j) $\text{Var}(3X)$ (k) $\dfrac{1}{7}\text{Var}(3X-2Y)$
(l) $\text{Var}(Y^2)$ (m) $\text{Var}(Y^2-2Y+3X)$
(n) $\text{Var}[3(X+Y)]$

3.3 Refer to Exercise 3.2 and rework, assuming that X and Y are independent.

3.4 A box contains 10 green balls and 15 red balls. 5 balls are randomly picked from the box with replacement. Find the expectation and variance of the frequency and relative frequency of the occurrence of green balls.

3.5 Given a pair of continuous random variables having the joint density

$$f(x,y) = \begin{cases} 24\,x\,y, & x>0,\ y>0 \text{ and } x+y\le 1 \\ 0, & \text{elsewhere} \end{cases}$$

Find

(a) $E(X)$	(b) $E(Y)$	(c) $E(XY)$
(d) $\text{Var}(X)$	(e) $\text{Var}(Y)$	(f) $E(X+Y)$
(g) $E(X-Y)$	(h) $\text{Var}(X+Y)$	(i) $\text{Var}(X-Y)$

assuming independence.

3.6 Given the joint density

$$f(x,y) = \begin{cases} 2, & 0<x<1, \quad 0<y<1, x+y<1 \\ 0, & \text{elsewhere} \end{cases}$$

find

(a) E(X)	(b) E(Y)	(c) E(XY)
(d) Var(X)	(e) Var(Y)	(f) $E(X-Y)$
(g) $E(X+Y)$	(h) $\text{Var}(X+Y)$	(i) $\text{Var}(X-Y)$

if independence is assumed.

3.7 Refer to Exercise 1.7. Find

(a) E(X)	(b) E(Y)	(c) E(XY)
(d) Var(X)	(e) Var(Y)	(f) $E(X-Y)$
(g) $E(X+Y)$	(h) $\text{Var}(X+Y)$	(i) $\text{Var}(X-Y)$

Download free eBooks at bookboon.com

3.8 Refer to Exercise 1.8. Find

(a) E(X)	(b) E(Y)	(c) E(XY)
(d) Var(X)	(e) Var(Y)	(f) $E(X-Y)$
(g) $E(X+Y)$	(h) $\text{Var}(X+Y)$	(i) $\text{Var}(X-Y)$

3.9 Refer to Exercise 1.12. Find

(a) E(X)	(b) E(Y)	(c) E(XY)
(d) Var(X)	(e) Var(Y)	(f) $E(X-Y)$
(g) $E(X+Y)$	(h) $\text{Var}(X+Y)$	(i) $\text{Var}(X-Y)$

3.10 Refer to Exercise 1.14. Find

(a) E(X)	(b) E(Y)	(c) E(XY)
(d) Var(X)	(e) Var(Y)	(f) $E(X-Y)$
(g) $E(X+Y)$	(h) $\text{Var}(X+Y)$	(i) $\text{Var}(X-Y)$

3.11 Refer to Exercise 1.3. Find

(a) E(X)	(b) E(Y)	(c) E(XY)
(d) Var(X)	(e) Var(Y)	(f) $E(X-Y)$
(g) $E(X+Y)$	(h) $\text{Var}(X+Y)$	(i) $\text{Var}(X-Y)$

3.12 Refer to Exercise 1.16. Find

(a) E(X)	(b) E(Y)	(c) E(XY)
(d) Var(X)	(e) Var(Y)	(f) $E(X-Y)$
(g) $E(X+Y)$	(h) $\text{Var}(X+Y)$	(i) $\text{Var}(X-Y)$

3.13 Prove Theorem 3.13.

3.14 Suppose X and Y are independent Normal variables $N(\mu, \sigma^2)$ and $N(\mu, 2\sigma^2)$, respectively. Determine μ if

(a) $\sigma = 3$ and $P(X + 2Y \le 10) = 0.3$,

(b) $\sigma = 2$ and $P(2X + Y < 5) = 0.4$,

(c) $\sigma = 3$ and $P(X + Y \ge 8) = 0.2$.

3.15 Refer to Exercise 3.14. Determine μ and σ if
$P(|2X - Y| > 10) = 0.05$ and $P(Y \le 10) = 0.9$

Download free eBooks at bookboon.com

Chapter 4

MEASURES OF RELATIONSHIP OF BIVARIATE DISTRIBUTIONS

4.1 INTRODUCTION

We recall that the expectation and variance of a univariate random variable are special moments about the origin and the mean respectively, the expectation being the first moment about the origin and the variance, the second moment about the mean. When we have two or more random variables, moments are similarly calculated. In Chapter 3, we calculated the mean and variance of bivariate random variables and also discussed their properties. However, we shall realise in this chapter that in the multivariate case, other types of moments can be calculated.

Click on the ad to read more

4.2 PRODUCT MOMENT

Let $g(X_1, ..., X_n)$ be any function of the random variables $X_1, ..., X_n$ whose density function is $f(x_1, ..., x_n)$. Then the k^{th} moment of $g(X_1, X_2, ..., X_n)$ is defined by

$$E[g^k(X_1, X_2..., X_n)]$$
$$= \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} g^k\,(x_1,\, x_2,\, \cdots, x_n)\, f(x_1, ..., x_n)\, dx_1\, dx_2...dx_n$$

In the case of a bivariate distribution a special type of moment that has been found very useful is the *product moment.* We shall define product moment for both discrete and continuous random variables.

4.2.1 Definition of Product Moment about Origin

Definition 4.1 PRODUCT MOMENT ABOUT ORIGIN
(Discrete Case)
Suppose X and Y are two discrete random variables. Then the product moment of orders r and s about the origin of X and Y respectively is given by

$$M'_{rs} = E\,(X^r Y^s) = \sum_{x=1} \sum_{y=1} x^r\, y^s\, p(x, y)$$

where r and s are nonnegative integers

If $r = s = 1$ this formula reduces to that of $E(XY)$ in Theorem 3.1.

Definition 4.2 PRODUCT MOMENT ABOUT ORIGIN
(Continuous Case)
Suppose X and Y are two continuous random variables. Then the product moment of orders p and q about the origin of X and Y respectively, is given by

$$M'_{pq} = E\,(X^p Y^q) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p\, y^q\, f(x, y)\, dy\, dx$$

where p and q are nonnegative integers

If $p = q = 1$ this formula reduces to that of $E(XY)$ in Theorem 3.4.

The moments described in Definitions 4.1 and 4.2 for the case when both powers are equal to unity are known as the *first product moment about the origin.*

Download free eBooks at bookboon.com

4.2.2 Definition of Product Moment about Mean

Definition 4.3 PRODUCT MOMENT ABOUT MEAN
(Discrete Case)
Suppose X and Y are two discrete random variables with means μ_X and μ_Y respectively. Then the product moment of orders r and s about the mean of X and Y respectively is defined as

$$\begin{aligned} M^*_{rs} &= E[(X-\mu_X)^r(Y-\mu_Y)^s] \\ &= \sum_x \sum_y (x-\mu_X)^r\,(y-\mu_Y)^s\,p(x,y) \end{aligned}$$

If $r = s = 1$, the formula in Definition 4.3 reduces to[6]

$$M^*_{11} = E[(X-\mu_X)(Y-\mu_Y)] = \text{Cov}(X,Y)$$

Definition 4.4 PRODUCT MOMENT ABOUT MEAN
(Continuous Case)
Suppose X and Y are two continuous random variables. Then the product moment of orders p and q about the mean of X and Y respectively is defined as

$$\begin{aligned} M^*_{pq} &= E[(X-\mu_X)^p(Y-\mu_Y)^q] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)^p\,(y-\mu_Y)^q\,f(x,y)\,dy\,dx \end{aligned}$$

If $p = q = 1$ this formula reduces to

$$M^*_{11} = E\,[(X-\mu_X)(Y-\mu_Y)] = \text{Cov}(X,Y)$$

The moments described in Definitions 4.3 and 4.4 for the case when both powers are equal to unity shall conveniently be referred to as the first product moment about the mean or the *first central product moment.*

Example 4.1
Refer to Example 3.8. Calculate the first product moment about the mean.

Solution
The first product moment about the mean, (that is, $p = q = 1$) for which we are required to calculate is given by

$$\mu_{11} = E[(X-\mu_X)(Y-\mu_Y)] = \int_0^1 \int_0^2 (x-\mu_X)(y-\mu_Y)\,f(x,y)\,dy\,dx$$

[6]See detailed discussion of this in Section 4.3.

Download free eBooks at bookboon.com

From Example 3.5

$$\begin{aligned}
\mu_X &= E(X) = \frac{13}{18} \\
\mu_Y &= E(Y) = \frac{10}{9}
\end{aligned}$$

Hence

$$\begin{aligned}
&= \int_0^1 \int_0^2 \left(x - \frac{13}{18}\right)\left(y - \frac{10}{9}\right)\left(x^2 + \frac{xy}{3}\right) dy\, dx \\
&= \frac{1}{486} \int_0^1 (18x - 13)\left[\int_0^2 (27x^2y + 9y^2x - 30x^2 - 10xy)\, dy\right] dx \\
&= \frac{1}{486} \int_0^1 (18x - 13)(54x^2 + 24x - 60x^2 - 20x)\, dx \\
&= \frac{1}{243} \int_0^1 (18x - 13)(2x - 3x^2)\, dx \\
&= \frac{1}{243} \int_0^1 (75x^2 - 54x^3 - 26x)\, dx \\
&= -0.00617
\end{aligned}$$

Click on the ad to read more
Download free eBooks at bookboon.com

4.3 COVARIANCE OF RANDOM VARIABLES

4.3.1 Definition of Covariance

The variance of a random variable is a measure of its variability, and the covariance of two random variables is a measure of their joint variability, or their degree of association.

Covariance was first mentioned and defined in Chapter 3 during the discussion of the properties of variance. In this section, we shall introduce it formally and discuss its properties.

Definition 4.5 COVARIANCE OF X and Y
If (X, Y) is a bivariate random variable with $E(X) = \mu_X$ and $E(Y) = \mu_Y$, then the covariance of X and Y, denoted by $\text{Cov}(X, Y)$, is defined as the expected value of $(X - \mu_X)(Y - \mu_Y)$

$$\text{Cov}(X, Y) = E(X - \mu_X)(Y - \mu_Y)$$

Note
$M_{11}^* = \text{Cov}(X, Y)$ and hence the covariance is sometimes referred to as the first product moment about the mean or simply the product moment.

Theorem 4.1
Let X and Y be random variables with a joint distribution function, then

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

Proof

$$\begin{aligned}
\text{Cov}(X, Y) &= E\{[X - E(X)][Y - E(Y)]\} \\
&= E[XY - YE(X) - XE(Y) + E(X)E(Y)] \\
&= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) \\
&= E(XY) - E(X)E(Y)
\end{aligned}$$

From Theorem 3.1, the covariance may also be defined as

$$\text{Cov}(X, Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i y_j \, p(x_i, y_j) - E(X)E(Y)$$

Download free eBooks at bookboon.com

Note

(a) $\text{Cov}(X, X) = \text{Var}(X) \geq 0$

(b) The covariance need not be finite, or even exist. However, it is finite if both random variables have finite variances.

Example 4.2
For the data in Example 2.3, find the covariance between X and Y.

Solution
It has been shown in Example 3.14,

$$\begin{aligned} E(XY) &= 6.24 \\ E(X)E(Y) &= 6.24 \end{aligned}$$

Hence

$$\begin{aligned} \text{Cov}(X, Y) &= E(XY) - E(X)E(Y) \\ &= 6.24 - 6.24 \\ &= 0 \end{aligned}$$

Example 4.3
Refer to Example 1.3. Find $\text{Cov}(X, Y)$.

Solution
From Examples 3.7 and 3.8,

$$E(X) = \frac{13}{18}; \quad E(Y) = \frac{10}{9}; \quad E(XY) = \frac{43}{54}.$$

Hence

$$\begin{aligned} \text{Cov}(X, Y) &= E(XY) - E(X)E(Y) \\ &= \frac{43}{54} - \left(\frac{13}{18}\right)\left(\frac{10}{9}\right) \\ &= -0.00617 \end{aligned}$$

which is the same as that obtained in Example 4.1.

4.3.2 Properties of Covariance

Property 1 **Symmetry**

Theorem 4.2
Suppose X and Y are two random variables, then

$$\text{Cov}(X, Y) = \text{Cov}(Y, X)$$

Download free eBooks at bookboon.com

Property 2

> **Theorem 4.3**
> Suppose X and Y are two random variables and a is a constant, then
>
> $$\text{Cov}(a + X, Y) = \text{Cov}(X, Y)$$

Proof

$$\begin{aligned}
\text{Cov}(a + X, Y) &= E\{[a + X - E(a + X)][Y - E(Y)]\} \\
&= E\{[X - E(X)][Y - E(Y)]\} \\
&= \text{Cov}(X, Y)
\end{aligned}$$

Corollary 4.1
Suppose X and Y are two random variables and a and b are constants. Then

$$\text{Cov}(a + X, b + Y) = \text{Cov}(X, Y)$$

Property 3

> **Theorem 4.4**
> Suppose X and Y are two random variables, then
>
> $$\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$$

Download free eBooks at bookboon.com
Click on the ad to read more

Proof

$$\begin{aligned}
\operatorname{Cov}(X, Y+Z) &= E([X-E(X)]\{[Y+Z-E(Y+Z)]\}) \\
&= E\{[X-E(X)][Y-E(Y)]+[X-E(X)][Z-E(Z)]\} \\
&= E\{[X-E(X)][Y-E(Y)]\}+E\{[X-E(X)][Z-E(Z)]\} \\
&= \operatorname{Cov}(X,Y)+\operatorname{Cov}(X,Z)
\end{aligned}$$

Property 4

Theorem 4.5
Suppose X and Y are two random variables and a and b are constants, then

$$\operatorname{Cov}(aX, bY) = a\,b\operatorname{Cov}(X,Y)$$

Proof

$$\begin{aligned}
\operatorname{Cov}(aX, bY) &= E\{[aX-aE(X)]\,[bY-bE(Y)]\} \\
&= E\{[a(X-E(X))]\,[b(Y-bE(Y))]\} \\
&= a\,b\,E\{[X-E(X)]\,[Y-E(Y)]\} \\
&= a\,b\operatorname{Cov}(X,Y)
\end{aligned}$$

Combining Corollary 4.1 and Theorem 4.5 , we have another corollary.

Corollary 4.2
Suppose X and Y are two random variables and a, b, c, and d are constants, then

$$\operatorname{Cov}(a+cX, b+dY) = c\,d\operatorname{Cov}(X,Y)$$

In general, the same kind of argument gives the important linear property of covariance.

Property 5 **Bilinear Property of Covariance**

Theorem 4.6
Let $U = a + \sum_{i=1}^{n} c_i X_i$ and $V = b + \sum_{j=1}^{m} d_j\,Y_j$. Then

$$\operatorname{Cov}(U,V) = \sum_{i=1}^{n}\sum_{j=1}^{m} c_i\,d_j\operatorname{Cov}(X_i,\,Y_j)$$

Download free eBooks at bookboon.com

This theorem has many applications. In particular, since

$$\begin{aligned}\operatorname{Cov}(X+Y, X+Y) &= \operatorname{Var}(X+Y) \\ &= \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X,Y)\end{aligned}$$

Property 6 **Independence of Random Variables**

Theorem 4.7
If X and Y are independent random variables then $\operatorname{Cov}(X,Y) = 0$

Proof
From Theorem 4.1

$$\operatorname{Cov}(X,Y) = E(XY) - E(X)E(Y)$$

But if X and Y are independent, then from Theorem 3.10

$$E(XY) = E(X)E(Y)$$

The desired result follows immediately.

Note
If $\operatorname{Cov}(X,Y) = 0$, X and Y need not be independent.

Property 7

Theorem 4.8
Suppose X and Y are random variables having second moments then $\operatorname{Cov}(X,Y)$ is a well-defined finite number and

$$[\operatorname{Cov}(X,Y)]^2 \le \operatorname{Var}(X)\operatorname{Var}(Y)$$

By replacing X by $X - E(X)$ and Y by $Y - E(Y)$ in Theorem 3.14, Theorem 4.8 follows immediately.

Note
The equality holds if and only if Y and X have perfect linear relation.

4.3.3 Uses of Covariance

$\operatorname{Cov}(X,Y)$ is often used as a measure of the linear dependence of X and Y, and the reason for this is that $\operatorname{Cov}(X,Y)$ is a single number (rather than a complicated object such as a joint density function) which contains some useful information about the joint behaviour of X and Y:

Download free eBooks at bookboon.com

(a) Positive values indicate that X increases as Y increases;

(b) Negative values indicate that X decreases as Y increases or vice versa;

(c) A zero value of the covariance would indicate no linear dependence between X and Y.

4.3.4 Limitations of Covariance

The covariance is a bad measure of dependence because:

(a) It is not 'scale-invariant', that is, the numerical value of the covariance is in the product of the units in which the two variables were measured. This makes it difficult to determine whether a particular covariance is large at first glance.

(b) It is unbounded. It can take any value. This means we cannot determine the maximum value of a covariance at which we can say whether it is too large or not.

To deal with these problems, we 're-scale' covariance to obtain the correlation coefficient of random variables X and Y.

Deloitte.
Discover the truth at www.deloitte.ca/careers
© Deloitte & Touche LLP and affiliated entities.
Click on the ad to read more

4.4 CORRELATION COEFFICIENT OF RANDOM VARIABLES

4.4.1 Definition of Correlation Coefficient

The problems associated with covariance can be eliminated by standardising its value by the standard deviations σ_X and σ_Y of X and Y, respectively. This gives the simple coefficient of linear correlation or simply the *correlation coefficient*, denoted as $\rho(X, Y)$ or simply as ρ.

Definition 4.6 CORRELATION COEFFICIENT
If the covariance of X and Y is defined then the correlation coefficient is defined as

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

In this definition, $\dfrac{0}{0} = 0$. Correlation coefficient is also referred to as *Pearson's correlation coefficient* or *product-moment correlation coefficient.* The product-moment correlation coefficient ρ can also be expressed as

$$\rho = \frac{\overline{XY} - \overline{X}\,\overline{Y}}{\sigma_X \sigma_Y} \qquad (i)$$

The reader will be asked in Exercise 4.21 to show that the numerator of expression (i) above is defined by $\text{Cov}(X, Y)$. This expression may also be used to establish a very important property of the correlation coefficient (see Theorem 4.10 in the sequel.)

4.4.2 Properties of Correlation Coefficient

Property 1 **Scale-invariance**

Theorem 4.9
For any $a, b, c, d \in R$ such that $ac \neq 0$,

$$\rho(aX + b, cY + d) = I_{XY}\, \rho(X, Y)$$

where

$$I_{XY} = \begin{cases} +1, & \text{if } ac > 0 \\ -1, & \text{if } ac < 0 \end{cases}$$

Download free eBooks at bookboon.com

Proof

$$\rho(aX+b,\, cY+d) = \frac{\mathrm{Cov}(ax+b,\, cY+d)}{\sqrt{\mathrm{Var}(aX+b)\,\mathrm{Var}(cY+d)}}$$

By Corollary 4.2

$$\mathrm{Cov}(aX+b,\, cY+d) = a\,c\,\mathrm{Cov}(X,Y)$$

and by Theorem 3.22

$$\mathrm{Var}(aX+b) = a^2\,\mathrm{Var}(X) \quad \text{and} \quad \mathrm{Var}(cY+d) = c^2\mathrm{Var}(Y)$$

Hence

$$\begin{aligned}\rho(aX+b,\, cY+d) &= \frac{a\,c\,\mathrm{Cov}(X,Y)}{\sqrt{a^2\,\mathrm{Var}(X)\; c^2\,\mathrm{Var}(Y)}}\\ &= \frac{a\,c}{|a\,c|}\cdot\rho(X,Y)\end{aligned}$$

from which the result follows.

Property 2 **Bounds of Correlation Coefficient**

Theorem 4.10
The correlation coefficient lies between -1 and $+1$, that is,

$$-1 \le \rho \le 1$$

By dividing both sides of the formula in Theorem 4.8 by $\mathrm{Var}(X)\,\mathrm{Var}(Y)$ and finding the square root, Theorem 4.10 follws immediately.

4.4.3 Interpretation of Correlation Coefficient

For the value of ρ to be meaningful, both variables involved, in theory, must be distributed normally. Positive values of ρ show that Y tends to increase with increasing X; negative values of ρ indicate that Y tends to decrease with increasing values of X.

There are specific interpretations of the correlation coefficient in terms of the joint behaviour of X and Y.

(a) ***As a measure of linearity***

The correlation coefficient measures the degree of linearity between X and Y.
When $\rho = \pm 1$, there is perfect linearity between X and Y.

(i) If $\rho(X,Y) = 1$, then Y is an increasing perfect linear function of X.

(ii) If $\rho(X,Y) = -1$, Y is a decreasing perfect linear function of X.

When $\rho = 0$, there is no *linear* relationship between X and Y (see the note below). When the value of ρ is near $+1$ or -1, that is an indication of a high degree of linearity and when ρ is very small (near 0) it indicates a lack of linearity.

Download free eBooks at bookboon.com

Note
When $\rho = 0$ or near 0 it does not indicate the absence of relationship between X and Y. It only indicates no *linear* relationship and it does not preclude the possibility of some nonlinear relationship.

(b) ***As a measure of independence***

When X and Y are independent then $\rho = 0$. However, the converse is not true. That is:

(i) If $\rho = 0$, it does not mean that X and Y are independent.

(ii) If $\rho \neq 0$ it does not mean that X and Y are dependent. The association may be spurious.

In general, correlation and dependence are not equivalent. The existence of a statistical association, even if it is very strong, does *not* establish that an increase in X causes an increase in Y, or that an increase in Y causes an increase in X. A fundamental weakness of observational studies is that they can demonstrate association but not causation.

Example 4.4
For the data in Example 2.3, calculate the correlation coefficient of X and Y and comment on the result.

Solution
From Example 4.2

$$\text{Cov}(X, Y) = 0$$

From Example 3.21

$$\begin{aligned} \sigma_X &= \text{Var}(X) = \sqrt{8.64} = 2.9394 \\ \sigma_Y &= \text{Var}(Y) = \sqrt{1.89} = 1.3748 \end{aligned}$$

Hence

$$\rho = \frac{0}{\sqrt{(1.3748)(2.9394)}} = 0$$

There is no *linear* correlation between X and Y.

Example 4.5
Refer to Example 1.3. Calculate the correlation coefficient.

Solution
From Example 3.7 and 3.8,

$$\begin{aligned} E(XY) &= \frac{43}{54} = 0.79630 \\ E(X) &= \frac{13}{18} = 0.72222 \\ E(Y) &= \frac{10}{9} = 1.11111 \end{aligned}$$

From Example 3.19

$$\begin{aligned} \text{Var}(X) &= 0.04506 \\ \text{Var}(Y) &= 0.32099 \end{aligned}$$

Download free eBooks at bookboon.com

Hence

$$\begin{aligned}\rho &= \frac{E(XY) - E(X)E(Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \\ &= \frac{0.79630 - 0.72222(1.11111)}{\sqrt{(0.04506)(0.32099)}} \\ &= -0.05127\end{aligned}$$

4.4.4 Variance of Sum of Random Variables with Common Variance and Common Correlation Coefficient

We now establish the variance of sum of random variables with correlation coefficient.

Theorem 4.11
Suppose $X_1, X_2, ..., X_n$ are random variables with common variance σ^2 and common correlation $\mathrm{Corr}(X_i, X_j) = \rho, \quad i \neq j \quad (i, j = 1, 2, ..., n)$. Let

$$S_n = X_1 + X_2 + ... + X_n$$

Then

(a) $\mathrm{Var}(S_n) = n\sigma^2 [1 + (n-1)\rho]$

(b) $\mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{\sigma^2}{n} [1 + (n-1)\rho]$

where $\frac{S_n}{n} = \overline{X}_n =$ sample mean

Proof
(a) From Corollary 3.9,

$$\begin{aligned}\mathrm{Var}(S_n) &= \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\,\mathrm{Cov} \sum_{\substack{i \\ i<j}}^{n} \sum_{j}^{n} \mathrm{Cov}(X_i, X_j) \\ &= \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{\substack{i \\ i<j}}^{n} \sum_{j}^{n} \sqrt{\mathrm{Var}(X_i)\mathrm{Var}(X_j)}\end{aligned}$$

Proof
(a) From Corollary 3.9,

$$\begin{aligned}\mathrm{Var}(S_n) &= \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\,\mathrm{Cov} \sum_{\substack{i \\ i<j}}^{n} \sum_{j}^{n} \mathrm{Cov}(X_i, X_j) \\ &= \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{\substack{i \\ i<j}}^{n} \sum_{j}^{n} \sqrt{\mathrm{Var}(X_i)\mathrm{Var}(X_j)} \\ &\quad \times \left(\frac{\mathrm{Cov}\,(X_i, X_j)}{\sqrt{\mathrm{Var}(X_i)\mathrm{Var}(X_j)}}\right)\end{aligned}$$

Download free eBooks at bookboon.com

$$
\begin{aligned}
&= \sum_{i=1}^{n} \sigma^2 + 2 \sum_{\substack{i \\ i<j}}^{n} \sum_{j}^{n} \sqrt{\sigma^2 \sigma^2}\, \rho \\
&= n\sigma^2 + 2 \sum_{\substack{i \\ i<j}}^{n} \sum_{j}^{n} \rho\, \sigma^2 \\
&= n\sigma^2 + 2\, \rho\, \sigma^2 \left(\frac{n^2 - n}{2} \right) \\
&= n\sigma^2 + \rho\, \sigma^2 \left(n^2 - n \right) \\
&= n\sigma^2 + \rho\, \sigma^2\, n\ (n-1) \\
&= n\sigma^2 \left[1 + (n-1)\ \rho \right]
\end{aligned}
$$

$$
\begin{aligned}
\text{(b)} \quad \operatorname{Var}\left(\frac{S_n}{n} \right) &= \frac{1}{n^2} \operatorname{Var}(S_n) \\
&= \frac{\sigma^2}{n} \left[1 + (n-1)\ \rho \right]
\end{aligned}
$$

When X's are independent, $\rho = 0$ and

$$
\operatorname{Var}\left(\frac{S_n}{n} \right) = \operatorname{Var}(\overline{X}) = \frac{\sigma^2}{n}
$$

as will be seen in Theorem 7.2.

Click on the ad to read more

4.5 CONDITIONAL EXPECTATIONS

We encountered in my book "Introductory Probability Theory" (Nsowah-Nuamah, 2017) the concept of conditional probability. Again, in Chapter 1 of this book, we dealt with conditional distributions. In this section, we discuss conditional expectation[7] which is simply the expected value of a variable, given that a set of prior conditions has taken place. As we shall observe later, it is obtained simply by summation or integration with respect to conditional distributions. A precise definition will be given and their properties, some of which are analogous to properties obtained in Chapter 3 for (unconditional) expectations, will be discussed.

4.5.1 Definition of Conditional Expectation

Consider the random variable $(Y|X)$ which may be interpreted as the random variable Y given random variable X. Let $E(Y|X)$ be the conditional expectation of Y for values of X. It is important to remember that $E(Y|X)$ itself is a random variable. Let $E(Y|X = x)$ or simply $E(Y|x)$ represent the expectation of Y conditioned on the event $\{X = x\}$. Since $E(Y|x)$ is a function of x it is a random variable though, strictly speaking, it is one value of the random variable $E(Y|X)$.

Definition 4.7 CONDITIONAL EXPECTATION
(Discrete Case)
If (X, Y) is a two-dimensional discrete random variable, we define the conditional expectation of Y for given $X = x$ as

$$E(Y|X = x) = \sum_y y\, h(y|x)$$

where $h(y|x) = p_{Y|X}(y|x)$ denotes the conditional p.m.f of Y given X

Definition 4.8 CONDITIONAL EXPECTATION
(Continuous Case)
If (X, Y) is a two-dimensional continuous random variable, we define the conditional expectation of Y for given $X = x$ as

$$E(Y|X = x) = \int_{-\infty}^{\infty} y\, h(y|x)\, dy$$

where $h(y|x) = f_{Y|X}(y|x)$ denotes the conditional p.d.f of Y given X

[7] It is sometimes referred to as conditional expected value or conditional mean of a random variable

Download free eBooks at bookboon.com

From the foregoing two definitions we notice that the definition of conditional expectation is almost the same as the definition of expectation, except that instead of a probability (marginal) distribution it uses a conditional probability (marginal) distribution.

Example 4.6
Refer to Example 1.3. Determine: (a) $E(Y|x)$, (b) $E(X|y)$

Solution

(a) $$E(Y|x) = \int_{-\infty}^{\infty} y\, f(y|x)\, dy$$

From Example 1.3,

$$f(y|x) = \frac{3x+y}{6x+2}, \qquad 0 < y < 2, \quad 0 < x < 1$$

Hence

$$\begin{aligned} E(Y|x) &= \int_0^2 y\left(\frac{3x+y}{6x+2}\right) dy \\ &= \int_0^2 \left(\frac{3xy+y^2}{6x+2}\right) dy \\ &= \frac{1}{3x+1}\left(3x+\frac{4}{3}\right) \\ &= \frac{9x+4}{3(3x+1)} \end{aligned}$$

(b) $$E(X|y) = \int_{-\infty}^{\infty} x\, f(x|y)\, dx$$

From Example 1.13,

$$f(x|y) = \frac{6x^2+2xy}{2+y}, \qquad 0 < x < 1, \quad 0 < y < 2$$

Hence

$$\begin{aligned} E(X|y) &= \int_0^1 x\left(\frac{6x^2+2xy}{2+y}\right) dx \\ &= \int_0^1 \left(\frac{6x^3+2x^2y}{2+y}\right) dx \\ &= \frac{1}{2+y}\left(\frac{6x^4}{4}+\frac{2x^3y}{3}\right)\Bigg|_0^1 \\ &= \frac{1}{2+y}\left(\frac{6}{4}+\frac{2y}{3}\right) \\ &= \frac{18+8y}{12(2+y)} \end{aligned}$$

Download free eBooks at bookboon.com

4.5.2 Properties of Conditional Expectation

Property 1

Theorem 4.12
Suppose X and Y are random variables. If Y has a finite expectation and $Y \geq 0$, then

$$E(Y|X) \geq 0$$

Property 2

Theorem 4.13
Suppose X and Y are random variables having finite expectations and a_i are constants, $1 \leq i \leq n$, then

$$E\left(\sum_{i=1}^{n} a_i Y_i | X\right) = \sum_{i=1}^{n} E a_i E(Y_i|X)$$

In particular if $a_i = a$ then

$$E\left(\sum_{i=1}^{n} a_i Y_i | X\right) = a \sum_{i=1}^{n} E(Y_i|X)$$

SIMPLY CLEVER
ŠKODA


Do you like cars? Would you like to be a part of a successful brand?
We will appreciate and reward both your enthusiasm and talent.
Send us your CV. You will be surprised where it can take you.
Send us your CV on
www.employerforlife.com
Download free eBooks at bookboon.com
Click on the ad to read more

Property 3

Theorem 4.14
Suppose Y_1 and Y_2 are random variables with finite expectations. If $Y_1 \leq Y_2$ then

$$E(Y_1|X) \leq E(Y_2|X)$$

Property 4

Theorem 4.15
Suppose X and Y are two independent random variables. If Y has a finite expectation, then

$$E(Y|X) = E(Y)$$

Property 5

Theorem 4.16
Suppose X and Y are two random variables. If Y has a moment of order $r \geq 1$, then

$$\left|E(Y|X)\right|^r \leq [E(|Y|\big|X)]^r \leq E(|Y|^r|X)$$

As pointed out earlier, $E(Y|X)$ is a random variable and hence we may find its expectation.

Property 6

Theorem 4.17
Suppose that X and Y are independent random variables. Then the expectation of conditional expectation is given by

$$E[E(Y|X)] = E(Y)$$

Download free eBooks at bookboon.com

Proof
We will prove this for the continuous case. The discrete case is proved similarly on replacing integrals by summations.

By definition,

$$\begin{aligned} E(Y|x) &= \int_{-\infty}^{+\infty} y\, h(y|x)\, dy \\ &= \int_{-\infty}^{\infty} y\, \frac{f(x,y)}{g(x)}\, dy \end{aligned}$$

where $f(x,y)$ is the joint probability density function of (X,Y) and $g(x)$ is the marginal probability density function of X.

Multiply both sides of the equation by $g(x)$:

$$g(x)E[Y|x] = \int_{-\infty}^{\infty} \left\{ y\, \frac{f(x,y)}{g(x)}\, dy \right\} g(x)$$

Taking the integral of both sides with respect to x, gives an expectation of $E(Y|x)$, namely, $E[E(Y|X)]$. That is,

$$\begin{aligned} E[E(Y|X)] &= \int_{-\infty}^{\infty} E(Y|x)\, g(x)\, dx \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} y \frac{f(x,y)}{g(x)}\, dy \right] g(x)\, dx \end{aligned}$$

If all the expectations exist, it is permissible to write the above iterated integral with the order of integration reversed. Thus

$$\begin{aligned} E[E(Y|X)] &= \int_{-\infty}^{\infty} y \left[\int_{-\infty}^{\infty} f(x,y)\, dx \right] dy \\ &= \int_{-\infty}^{\infty} y\, h(y)\, dy \\ &= E(Y) \end{aligned}$$

Theorem 4.17 gives what might be called the *law of total expectation:* the expectation of a random variable Y can be calculated by weighting the conditional expectations appropriately and summing or integrating.

Property 7

Theorem 4.18
Let $\widehat{Y} = E(Y|x)$ be the conditional expectation of Y given X. Then

$$E[(Y - \widehat{Y})^2] \leq E[(Y - \pi)^2]$$

where $\pi = \pi(x)$ is any other function of X

Proof

$$\begin{aligned} E\{(Y-\pi)^2\} &= E[\{(Y-\widehat{Y}) + (\widehat{Y} - \pi)\}^2] \\ &= E[(Y-\widehat{Y})^2] + 2E[(Y-\widehat{Y})(\widehat{Y}-\pi)] + E[(\widehat{Y}-\pi)^2] \\ &= E[(Y-\widehat{Y})^2] + E[(\widehat{Y}-\pi)^2] \qquad (i) \end{aligned}$$

Download free eBooks at bookboon.com

because we can show that the second term is zero. Thus,

$$\begin{aligned} E(Y-\widehat{Y})(\widehat{Y}-\pi) &= \int_x \int_y (y-\widehat{y})(\widehat{y}-\pi)f(x,y)\partial y \partial x \\ &= \int_x \left[\int_y (y-\widehat{y})f(y|x)\partial y\right](\widehat{y}-\pi)f(x)\partial x \end{aligned}$$

since $f(x,y)$ can be expressed as the product of the conditional density function of Y given X and the marginal density function of Y, that is, $f(x,y)$ has been factored as $f(x,y) = f(y|x)f(x)$. And also

$$\begin{aligned} \int_y (y-\widehat{y})f(y|x)\partial y &= \int_y yf(y|x)\partial y - \int_y \widehat{y}f(y|x)\partial y \\ &= E(Y|x) - E(Y|x) \\ &= 0 \end{aligned}$$

Now, from (i),

$$\begin{aligned} E\{(Y-\pi)^2\} &= E[(Y-\widehat{Y})^2] + E[(\widehat{Y}-\pi)^2] \\ &\geq E[(Y-\widehat{Y})^2] \end{aligned}$$

from which the result follows.

Theorem 4.19
If the joint distribution of Y and X is a Normal distribution, then the function $E(Y|x)$ may be expressed as

$$E(Y|x) = \alpha + \beta x$$

Theorem 4.19 indicates that the conditional expectation of Y given X is a linear function of X. As we shall discuss in the sequel, the function is described as a linear regression function.

Property 8

Theorem 4.20
Suppose that X and Y are independent random variables. Then the variance of conditional expectation Y is given by

$$\text{Var}[E(Y|X)] = \text{Var}(Y) - E[\text{Var}(Y|X)]$$

Proof
Just as $E(Y|X)$, $\text{Var}(Y|X)$ is also a random variable and has expectation $E[\text{Var}(Y|X)]$. Now,

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

Download free eBooks at bookboon.com

$$
\begin{aligned}
& -E\{[E(Y|X)]^2\} \\
= \; & E(Y^2) - [E(Y)]^2 - \mathrm{Var}[E(Y|X)] \quad \text{(from Theorem 4.17)} \\
= \; & \mathrm{Var}(Y) - \mathrm{Var}[E(Y|X)
\end{aligned}
$$

from which the result follows.

4.6 CONDITIONAL VARIANCES

In this section we discuss conditional variances and prove a useful formula relating conditional and unconditional variances.

4.6.1 Definition of Conditional Variance

Definition 4.9 CONDITIONAL VARIANCE
If (X, Y) is a two-dimensional discrete random variable, we define the conditional variance of Y given X as

$$\mathrm{Var}(Y|X) = E(Y^2|X) - [E(Y|X)]^2$$

provided that $E(Y|X)$ exists

Click on the ad to read more

Thus for the discrete case, if $p_X(x) > 0$ and if Y has a second moment, then the conditional variance of Y given $X = x$ is given by

$$\operatorname{Var}(Y|X = x) = \sum_y y^2 p_{Y|X}(y|x) - \left[\sum_y y\, p_{Y|X}(y|x)\right]^2$$

The conditional variance of Y given $X = x$ for the continuous case is defined similarly by replacing the summations with integrals.

Another way of definition the conditional variance is given in Definition 4.9a.

Definition 4.9a CONDITIONAL VARIANCE
If (X, Y) is a two-dimensional discrete random variable, then the conditional variance of Y given X is

$$\operatorname{Var}(Y|X) = E\{[Y - E(Y|X)]^2 \big| X\}$$

provided that $E(Y|X)$ exists

4.6.2 Properties of Conditional Variance

Property 1

Theorem 4.21
Suppose that X and Y are independent random variables. Then the expectation of conditional variance

$$E[\operatorname{Var}(Y|X)] = E[E(Y^2|X)] - E\{[E(Y|X)]^2\}$$

Property 2

Theorem 4.22
Suppose that X and Y are independent random variables. Then the expectation of conditional variance is given by

$$E[\operatorname{Var}(Y|X)] = \operatorname{Var}(Y) - \operatorname{Var}[E(Y|X)]$$

Proof
This follows from the proof of Theorem 4.20.

We shall notice in Theorem 4.23 in the sequel derived from Theorem 4.20 that the *unconditional* variance $\operatorname{Var}(Y)$ is not just the expected value of the conditional variance but also the variance of conditional expectation.

Download free eBooks at bookboon.com

Theorem 4.23
Suppose that X and Y are independent random variables. Then

$$\mathrm{Var}(Y) = \mathrm{Var}[E(Y|X)] + E[\mathrm{Var}(Y|X)]$$

Proof
This follows from the proof of Theorem 4.20.

Aliter

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

since the cross-product term vanishes. Conditioning on X the two terms of the expression on the right side, we obtain for the first term:

$$E\{E[Y - E(Y|X)]^2|X\} = E[\mathrm{Var}(Y|X)] \quad \text{(by Definition 4.9a)}$$

and the second term:

$$E[E[Y|X) - E(Y|X)]^2 = \mathrm{Var}[E(Y|X)]$$

Hence

$$\mathrm{Var}(Y) = E[\mathrm{Var}(Y|X)] + \mathrm{Var}[E(Y|X)]$$

4.7 REGRESSION CURVES

The graph of $E(Y|x)$ as a function of x, is called the regression curve (of the mean) of Y on X. Similarly, the graph of $E(X|y)$ as a function of y is known as the regression curve (of the mean) of X on Y.

4.7.1 Definition of Linear Regression

The linear regression function was encountered in Theorem 4.19 as the expected value of Y given X. In this section, we shall discuss in detail its special characteristics.

Definition 4.10 LINEAR REGRESSION
The two-dimensional random variable (X, Y) is said to have a linear regression of Y on X if

$$E(Y|x) = \alpha + \beta x$$

where α and β are constants

Download free eBooks at bookboon.com

That is, the regression curve of Y on X is a straight line, called the regression line of Y on X. The constants α and β are the *parameters* of the linear regression equation. The constant α is called the *intercept*, which is the point where the regression line cuts the y-axis. The constant β is called the *slope of the regression equation* or the *regression coefficient* of Y on X which measures a change in Y per unit change in X.

The constants α and β are unknown parameters and have to be estimated. We discuss here two estimation methods, namely,

(a) method of moments;

(b) method of least suares.

4.7.2 Method of Moments for Estimating Linear Regression Function

The method of moments for estimating the linear regression function attempts to find expressions for α and β that are in terms of the first-and second-order moments of the joint distribution, namely, $E(X)$, $E(Y)$, $\mathrm{Var}(X)$ and $\mathrm{Cov}(X, Y)$.

Download free eBooks at bookboon.com
Click on the ad to read more

Theorem 4.24
If (X, Y) has linear regression of Y on X, then

$$\begin{aligned}\alpha &= E(Y) - \beta\, E(X)\\ \beta &= \frac{\mathrm{Cov}(XY)}{\mathrm{Var}(X)}\end{aligned}$$

Proof
Proof for Estimating α:
From Definition 4.10, the regression function is given by

$$E(Y|x) = \alpha + \beta x \qquad (i)$$

If we multiply (i) by $f(x)$ and integrate it with respect to x, that is,

$$\int_x E(Y|x)f(x)\,dx = \int_x (\alpha + \beta x)f(x)\,dx$$

we obtain

$$E(Y) = \alpha + \beta E(X) \qquad (ii)$$

so that

$$\alpha = E(Y) - \beta E(X) \qquad (iii)$$

Proof for Estimating β:
Let us substitute (iii) into (i) to obtain

$$E(Y|x) = E(Y) + \beta(X - E(X)) \qquad (iv)$$

Multiplying (i) by x and $f(x)$ and then integrating with respect to x, we have

$$\begin{aligned}\int_x xE(Y|x)f(x)\,dx &= \int_x x(\alpha + \beta x)f(x)\,dx\\ E(XY) &= \alpha E(X) + \beta E(X^2) \qquad (v)\end{aligned}$$

Multiplying (ii) by $E(X)$, we get

$$E(X)E(Y) = \alpha E(X) + \beta[E(X)]^2 \qquad (vi)$$

Subtracting (vi) from (v) gives

$$E(XY) - E(X)E(Y) = \beta\{E(X^2) - E[(X)]^2\} \qquad (vi)$$

so that

$$\begin{aligned}\beta &= \frac{E(XY) - E(X)E(Y)}{E(X^2) - E[(X)]^2}\\ &= \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)} \quad \text{(from Theorem 4.1)}\end{aligned}$$

or

$$\beta = \frac{E[X - E(X)][(Y - E(Y)]}{E\{[(X - E[(X)]^2\}}$$

Download free eBooks at bookboon.com

4.7.3 Least Squares Method of Estimating Linear Regression Function

Let the values $(x_1, y_1), (x_2, y_2), \cdots (x_n, y_n)$, be plotted as points in the x, y plane. Then the problem of estimating a linear regression function can be treated as a problem of fitting a straight line to this set of points. The best known method is the least squares approach.

The least squares method states that the sum of the squares of the difference between the observed value of Y and the corresponding fitted curve value of Y, which we denote by $\hat{Y}$, must be minimum. The values of the parameters obtained by this minimisation determine what is known as the best fitting curve in the sense of least squares.

Theorem 4.25
If (X, Y) has linear regression of Y on X, then

$$\alpha = \mu_Y - \beta\,\mu_X$$

$$\beta = \frac{\sigma_{XY}}{\sigma_X^2}$$

where $\mu_Y = E(X)$, $\sigma_X^2 = \mathrm{Var}(X)$ and $\sigma_{XY} = \mathrm{Cov}(X, Y)$

Proof
Let us find the best function of the form $h(x) = \alpha + \beta\, x$. This merely requires optimising over the two parameters α and β. Now, we can write[8]

$$\begin{aligned}
\mathrm{Var}(Y - \alpha - \beta\, X) &= E[(Y - \alpha - \beta\, X)^2] - [E(Y - \alpha - \beta\, X)]^2 \\
E[(Y - \alpha - \beta\, X)^2] &= \mathrm{Var}(Y - \alpha - \beta\, X) + [E(Y - \alpha - \beta\, X)]^2 \\
&= (\sigma_Y^2 + \beta^2\sigma_X^2 - 2\beta\,\sigma_{XY}) + [E(Y - \alpha - \beta\, X)]^2 \\
&\quad \text{(by Theorem 3.22; and also Var}(\alpha) = 0\text{, a property} \\
&\quad \text{of variance)}
\end{aligned}$$

The second term of the expression on the right hand side of the equation does not depend on α so α can be chosen so as to minimize the second term.

Recall that to minimize a function we set the derivative of the function to zero. Thus

$$\left\{[E(Y - \alpha - \beta\, X)]^2\right\}' = 2[E(Y - \alpha - \beta\, X)] = 0$$

giving

$$\begin{aligned}
\mu_Y - \alpha - \beta\,\mu_X &= 0 \\
\alpha &= \mu_Y - \beta\,\mu_X
\end{aligned}$$

[8] Recollect that

$$E(X) = E(X^2) - [E(X)]^2$$

Download free eBooks at bookboon.com

Now to minimize the first term we set the derivative with respect to β equal to zero, to obtain

$$\left(\sigma_Y^2 + \beta^2\sigma_X^2 - 2\,\beta\,\sigma_{XY}\right)' = 0$$

giving

$$\begin{aligned} 2\,\beta\,\sigma_X^2 - 2\,\sigma_{XY} &= 0 \\ \beta &= \frac{\sigma_{XY}}{\sigma_X^2} \end{aligned}$$

Corollary 4.3
If β is a coefficient of a linear regression function of X on Y, then

$$\beta = \rho\,\frac{\sigma_Y}{\sigma_X}$$

Proof
It is sufficient to show that

$$\frac{\sigma_{XY}}{\sigma_X^2} = \rho\,\frac{\sigma_Y}{\sigma_X}$$

Thus

$$\beta = \frac{\sigma_{XY}}{\sigma_X^2}$$

Download free eBooks at bookboon.com
Click on the ad to read more

$$\begin{aligned} &= \frac{\sigma_{XY}}{\sigma_X \sigma_X} \cdot \frac{\sigma_Y}{\sigma_Y} \\ &= \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \cdot \frac{\sigma_Y}{\sigma_X} \\ &= \rho \, \frac{\sigma_Y}{\sigma_X} \end{aligned}$$

Similarly, (X, Y) is said to have linear regression of X on Y if

$$E(X|y) = \tau + \theta \, y \qquad (i)$$

with τ and θ as constants. That is, the regression curve of X on Y is a straight line, which is called the regression line of X on Y. The number θ is called the regression coefficient of X on Y and is defined as

$$\theta = \frac{\sigma_{XY}}{\sigma_Y^2} = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(Y)} \qquad (ii)$$

Theorem 4.26
Let us define $\mathrm{Var}(\hat{Y}) = \mathrm{Var}(Y - \beta\, X)$ as the mean squared prediction error. Then

$$\mathrm{Var}(\hat{Y}) = \sigma_Y^2(1 - \rho^2)$$

Proof

$$\begin{aligned} \mathrm{Var}(Y - \beta\, X) &= \mathrm{Var}(Y) + \mathrm{Var}(\beta\, X) - 2\mathrm{Cov}(Y, \beta\, X) \\ &= \sigma_Y^2 + \beta^2\, \sigma_X^2 - 2\, \beta\, \sigma_{XY} \\ &= \sigma_Y^2 + \frac{\sigma_{XY}^2}{\sigma_X^4}\, \sigma_X^2 - 2\, \frac{\sigma_{XY}}{\sigma_X^2}\, \sigma_{XY} \\ &= \sigma_Y^2 - \frac{\sigma_{XY}^2}{\sigma_X^2} \\ &= \sigma_Y^2 - \left(\frac{\sigma_{XY}}{\sigma_X} \cdot \frac{\sigma_Y}{\sigma_Y} \right)^2 \\ &= \sigma_Y^2 - \rho^2\, \sigma_Y^2 \\ &= \sigma_Y^2 \left(1 - \rho^2\right) \end{aligned}$$

Example 4.7
Refer to Example 1.3.

(a) Find the linear regression of Y on X;

(b) Calculate the mean squared prediction error.

Solution
(a) From Example 3.19

$$\begin{aligned} \sigma_X^2 &= \mathrm{Var}X = 0.04506 \\ \sigma_Y^2 &= \mathrm{Var}Y = 0.32099 \\ \sigma_{XY} &= \mathrm{Cov(X,Y)} = -0.00617 \end{aligned}$$

Download free eBooks at bookboon.com

Hence

$$\beta = \frac{\sigma_{XY}}{\sigma_X^2} = -\left(\frac{0.00617}{0.04506}\right) = -0.13693$$

From Example 3.8,

$$\begin{aligned}\mu_X &= E(X) = \frac{13}{18} = 0.72222 \\ \mu_Y &= E(Y) = \frac{10}{9} = 1.11111\end{aligned}$$

Hence

$$\begin{aligned}\alpha &= \mu_Y - \beta\,\mu_X \\ &= 1.11111 - (-0.13693)(0.72222) = 0.90111\end{aligned}$$

Finally,

$$E(Y|x) = 0.90111 - 0.13693\,x$$

(b) Calculation of the mean squared prediction error

$$\begin{aligned}\text{Var}(\hat{Y}) &= \sigma_Y^2(1-\rho^2) \\ &= 0.32099[1-(-0.05127)] = 0.33745\end{aligned}$$

Since the value for β is negative, it means as X increases Y decreases. That is, a unit increase in X corresponds to about 1.14 decrease in Y (where the 1.14 is in the same unit of measurement as Y.)

Theorem 4.27
Let (X, Y) be a two-dimensional random variable with

$$E(X) = \mu_X, \quad E(Y) = \mu_Y, \quad V(X) = \sigma_X^2 \quad \text{and} \quad V(Y) = \sigma_Y^2$$

If the regression of Y on X is linear, then

$$E(Y|x) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X)$$

where ρ is the correlation coefficient between Y and X

Proof
Substituting the expressions for α and β from Theorem 4.25 into Definition 4.10, we get

$$\begin{aligned}E(Y|x) &= (\mu_Y - \beta\,\mu_X) + \frac{\sigma_{XY}}{\sigma_X^2}\,x \\ &= \mu_X - \frac{\sigma_{XY}}{\sigma_X^2}\,\mu_X + \frac{\sigma_{XY}}{\sigma_X^2}\,x \\ &= \mu_Y - \frac{\sigma_{XY}}{\sigma_X^2}(X-\mu_X) \\ &= \mu_Y - \rho\frac{\sigma_Y}{\sigma_X}(X-\mu_X) \quad (from Corollary 4.3)\end{aligned}$$

Download free eBooks at bookboon.com

Similarly, it can be proved that if the regression of X on Y is linear, then

$$E(X|y) = \mu_X + \theta\frac{\sigma_X}{\sigma_Y}(X - \mu_Y)$$

It follows that if a regression is linear and $\rho = 0$, then $E(Y|x)$ does not depend on X and $E(X|y)$ does not depend on Y.

Example 4.8
Refer to Example 1.3. Write the regression equation of Y on X.

Solution
From Example 3.8

$$\begin{aligned} E(X) &= \frac{13}{18} = 0.72222 \\ E(Y) &= \frac{10}{9} = 1.11111 \quad \text{(to 5 decimal places)} \end{aligned}$$

From Example 3.19,

$$\begin{aligned} \sigma_X &= \sqrt{0.04506} = 0.21227 \\ \sigma_Y &= \sqrt{0.32099} = 0.56656 \end{aligned}$$

From Example 4.5,

$$\rho = -0.05127.$$

Click on the ad to read more

Hence

$$\begin{aligned} E(Y|x) &= 1.11111 - 0.05127 \frac{0.56656}{0.21227} (x - 0.72222) \\ &= 1.11111 - 1.13684(x - 0.72222) \end{aligned}$$

Theorem 4.28
The product of the regression coefficients β and θ in the linear regressions $E(Y|x)$ and $E(X|y)$, respectively, is equal to the square of the correlation coefficient of X and Y:

$$\rho^2 = \beta\theta$$

Proof
In Corollary 4.3 and (ii) above,

$$\beta = \rho \frac{\sigma_Y}{\sigma_X}$$

and

$$\theta = \rho \frac{\sigma_X}{\sigma_Y}$$

Hence

$$\beta\theta = \rho \frac{\sigma_Y}{\sigma_X} \cdot \rho \frac{\sigma_X}{\sigma_Y}$$

from which the result follows.

Note

(a) The sign of the regression coefficient is determined by ρ, since $\sigma_X > 0$ and $\sigma_Y > 0$.

(b) The two linear regression equations, $E(Y|x)$ and $E(X|y)$, have the same sign of slope.

(c) The regression line passes through the point $(EXY) = [(E(X), E(Y)]$, which is the expected value of the joint distribution.

(d) The point of the intersection of the linear regression curves $E(Y|x)$ and $E(X|y)$ is $[(E(X), E(Y)]$.

This chapter concludes discussions on bivariate distributions which we started in Chapter 1. By induction we can easily extent most of the concepts discussed so far to multivariate distributions in general.

EXERCISES

4.1 For the data of Exercise 1.1, find the covariance between X and Y.

4.2 For the data of Exercise 1.1, find the correlation coefficient of X and Y.

Download free eBooks at bookboon.com

4.3 If $\text{Var}(W) = 5$, $\text{Var}(X) = 3$, $\text{Var}(Y) = 4$ and $\text{Var}(Z) = 1$, find

(a) $\text{Cov}(X, W)$, (b) $\text{Cov}(X, Y)$, (c) $\text{Cov}(Y, W)$,

(d) $\text{Cov}(X, Z)$, (e) $\text{Cov}(Y, Z)$, (f) $\text{Cov}(W, Z)$

given that the correlation coefficient between any pair is 0.6.

4.4 Refer to Exercise 4.3. If $a = 4$, $b = 2$, $c = 2$ and $d = 3$, find

(a) $\text{Cov}(c + X, Y)$,	(b) Cov(c+X, a+Y),
(c) $\text{Cov}(dX + bY)$,	(d) $\text{Cov}(a + dY, b + cZ)$,
(e) $\text{Cov}(aX, bZ + cW)$,	(f) $\text{Cov}(aW + bX, cY + dZ)$,
(g) $\text{Cov}(c + aW, c + aW)$,	(h) $\text{Cov}(bY + cW, X)$

4.5 A die is rolled twice. Let X be the sum of the outcomes, and Y, the first outcome minus the second. Compute $\text{Cov}(X, Y)$

4.6 Prove Theorem 4.5.

4.7 Prove Theorem 4.7.

4.8 Prove Theorem 4.8.

4.9 Refer to Exercise 1.7, find the correlation coefficient of X and Y.

4.10 Refer to Exercise 1.10, find the correlation coefficient of X and Y.

4.11 Suppose the random variable X and Y have the joint p.d.f.

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

Compute the correlation coefficient of X and Y.

4.12 Suppose the joint p.d.f. of X and Y is given by

$$f(x, y) = \begin{cases} \frac{1}{y} e^{-y}, & 0 < x < y, \quad 0 < y < \infty \\ 0, & \text{elsewhere} \end{cases}$$

Compute (a) $E(X|y)$, (b) $E(X^2|y)$

4.13 Refer to Exercise 1.7.

(a) Use the least squares method to obtain the linear regression equation of (i) Y on X; (ii) X on Y.

(b) (i) Find the product of the regression coefficients in a(i) and a(ii);

(ii) Take the square root of your results in b(i);

(iii) Compare the result in b(ii) with that in Exercise 4.9.

4.14 Refer to Exercise 1.10. Find
(a) $\text{Var}(Y|X)$ (b) $\text{Var}(X|Y)$.

Download free eBooks at bookboon.com

4.15 Refer to Exercise 1.7. Find
(a) $\text{Var}(Y|X)$ (b) $\text{Var}(X|Y)$.

4.16 Refer to Exercise 1.11. Find
(a) $\text{Var}(Y|X)$ (b) $\text{Var}(X|Y)$.

4.17 Verify Theorem 4.16 using Table 1.1.

4.18 Refer to Exercise 4.10.

(a) Find the regression of X on Y.
(b) Calculate the mean squared prediction error.

4.19 Refer to Exercise 4.16.

(a) Find the regression of X on Y.
(b) Calculate the mean squared prediction error.

4.20 Show that the point of intersection of $Y = E(Y|x)$ and $X = E(X|y)$ is (μ_X, μ_Y)

4.21 Show that $\rho(X, Y)$ can be defined as

$$Cov(X, Y) = \overline{XY} - \overline{XY}$$

Hence show that the correlation coeficient is scale invariant.

4.22 Prove Theorem 4.10 for the discrete case.

Click on the ad to read more

PART 2

STATISTICAL INEQUALITIES, LIMIT LAWS AND SAMPLING DISTRIBUTIONS

Life is the art of drawing sufficient conclusions from insufficient premises
SAMUEL BUTLER

Download free eBooks at bookboon.com

Chapter 5

STATISTICAL INEQUALITIES AND LIMIT LAWS

5.1 INTRODUCTION

5.1.1 Statistical Inequalities

There are two well-known inequalities due to two Russian mathematicians, Andrey Markov and and his teacher, Pafnuty Chebyshev[9], which play an important role in statistical work.

If we know the probability distribution of a random variable X, we may compute $E(X)$ and $\mathrm{Var}(X)$, if these exist. However, the converse is not true. That is, from a knowledge of $E(X)$ and $\mathrm{Var}(X)$, we cannot reconstruct the probability distribution of X. This is because to describe a probability distribution completely, we need to know the probability function of the random variable. However, Markov's and Chebyshev's inequalities enable us to derive lower (or upper) bound on such probabilities when only the mean, or both the mean and the variance of the distribution are known.

[9] These names appear in other publications as Markoff and Tchebysheff respectively. In this book we have used the spellings Markov and Chebyshev, the official American transliterations from Russian for consistency.

Click on the ad to read more
Download free eBooks at bookboon.com

5.1.2 Limit Laws

Perhaps, the two most important laws of probability theory and statistics are the "Law of Large Numbers" and the "Central Limit Theorem". Both are related to the notion of the "sum of random variables". Recall from Chapter 3 that if $X_i \quad (i = 1, 2, ..., n)$ are random variables, then their sum $S_n = X_1 + X_2 + \cdots + X_n$ is also a random variable with expectation $E(S_n)$ which is the sum of the expectations of the individual random variables if n is finite (see Corollary 3.3). Furthermore, if the X_i's are also independent, then the additivity property of variance also holds for the variance of S_n; that is, the variance of the sum, $\text{Var}(S_n)$, is the sum of the individual variances (see Corollary 3.9). Finally, it follows that if the X_i's have the same distribution, then they possess a common mean μ (Theorem 3.3), a common variance σ^2 (Theorem 3.16), and $E(S_n) = n\mu$ (Corollary 3.4). Theorem 3.19 states that if X_i's are independent and identically distributed, then $\text{Var}(S_n) = n\sigma^2$.

Except for a few exceptional cases such as some special probability distributions discussed in the previous chapters the distributions involving sums may be very complicated. Computations of probabilities with direct evaluation are extremely difficult. Fortunately, under the 'Law of Large Numbers' and the 'Central Limit Theorem' (see Sections 5.4 and 5.5, respectively), the evaluation of probabilities involving sums of random variables turn out to be fairly simple.

In the subsequent sections we shall discuss statistical inequalities and limit theorems in detail.

5.2 MARKOV'S INEQUALITY

Markov's inequality relates a probability to an expectation and provides an upper bound for the probability that a non-negative function of a random variable is greater than or equal to some positive constant.

Theorem 5.1 MARKOV'S INEQUALITY
Let X be a nonnegative random variable having a finite expectation $E(X)$. Let ϵ be any positive number, then the probability that X is no less than ϵ is no greater than the expectation of X divided by ϵ:

$$P(X \geq \epsilon) \leq \frac{E(X)}{\epsilon}$$

Proof
This theorem is valid for both continuous and discrete cases. We shall first prove for the discrete case.

Download free eBooks at bookboon.com

(a) Discrete Case
Suppose that the probability distribution of X has the form in the following table:

X	x_1	x_2	$\cdots$	x_k	$\cdots$	x_n
$P(X=x)$	$p(x_1)$	$p(x_2)$	$\cdots$	$p(x_k)$	$\cdots$	$p(x_n)$

Suppose also that the values of the random variable are arranged in ascending order

$$0 \leq x_1 < x_2 < \cdots < x_n$$

Note that $E(X) \geq 0$, since $X \geq 0$. We shall consider three cases.

Case 1
If X takes only zero values, then $E(X) = 0$ and for any constant $\epsilon > 0$

$$P(X \geq \epsilon) = 0 \leq \frac{0}{\epsilon}$$

That is, the theorem is valid.

Suppose now that not all x_i equal zero. Clearly,

$$E(X) > 0$$

Take an arbitrary constant $\epsilon > 0$.

Case 2
If $\epsilon > x_n$, then $\{X < \epsilon\}$ is a 'certain event' and

$$P(X < \epsilon) = 1 \leq \frac{E(X)}{\epsilon}$$

that is, the theorem is again valid in this case.

Case 3
Let $\epsilon \leq x_n$ and $x_k,\ x_{k+1},\ \cdots,\ x_n$ be all values of X greater than ϵ (if in a special case $\epsilon \leq x_1,\quad$ then $\quad k = 1$)

By definition,

$$\begin{aligned} E(X) &= x_1p_1 + x_2p_2 + \cdots + x_kp_k + \cdots + x_np_n \\ &\geq x_kp_k + \cdots + x_np_n \\ &\geq \epsilon(p_k + \cdots + p_n) \qquad\qquad (i) \end{aligned}$$

By addition rule of probability

$$p_k + p_{k+1} + \cdots + p_n = P(X \ \geq \epsilon)$$

Then from (i)

$$E(X) \geq \epsilon P(X \geq \epsilon)$$

or

$$P(X \geq \epsilon) \leq \frac{E(X)}{\epsilon} \qquad\qquad (ii)$$

Download free eBooks at bookboon.com

Since

$$P(X < \epsilon) = 1 - P(X \geq \epsilon)$$

then from (ii) it follows that

$$P(X < \epsilon) \geq 1 - \frac{E(X)}{\epsilon} \qquad (iii)$$

(b) Continuous case

$$\begin{aligned}
E(X) &= \int_0^\infty x f(x)\,dx \\
&= \int_0^\epsilon x f(x)\,dx + \int_\epsilon^\infty x f(x)\,dx \\
&\geq \int_\epsilon^\infty x f(x)\,dx \qquad \text{since} \quad \int_0^\epsilon x f(x)\,dx \geq 0 \\
&\geq \int_\epsilon^\infty \epsilon f(x) dx \quad \text{since} \quad \epsilon > 0 \\
&= \epsilon \int_\epsilon^\infty f(x)\,dx \\
&= \epsilon P(X \geq \epsilon)
\end{aligned}$$

Therefore

$$P(X \geq \epsilon) \leq \frac{E(X)}{\epsilon},$$

Download free eBooks at bookboon.com

and hence

$$P(X < \epsilon) \geq 1 - \frac{E(X)}{\epsilon}$$

Aliter

Let

$$Y = \begin{cases} 0, & \text{if } X < \epsilon \\ \epsilon, & \text{if } X \geq \epsilon \end{cases}$$

Then

$$P(Y = 0) = P(X < \epsilon)$$

and

$$P(Y = \epsilon) = P(X \geq \epsilon)$$

Hence

$$\begin{aligned} E(Y) &= 0 \cdot P(Y = 0) + \epsilon \cdot P(Y = \epsilon) \\ &= \epsilon\, P(X \geq \epsilon) \end{aligned}$$

Clearly,

$$X \geq Y$$

Hence

$$E(X) \geq E(Y) = \epsilon\, P(X \geq \epsilon)$$

Therefore

$$P(X \geq \epsilon) \leq \frac{E(X)}{\epsilon}$$

Note

(1) The Markov's inequality may be equivalently stated as in Theorem 5.1(b).

> **Theorem 5.1(b)** MARKOV'S INEQUALITY
> Let X be a nonnegative random variable having a finite expectation $E(X)$. Let ϵ be any positive number, then
>
> $$P(X < \epsilon) \geq 1 - \frac{E(X)}{\epsilon}$$

Proof

Since $X \geq \epsilon$ and $X < \epsilon$ are complementary events, Theorem 5.1(b) follows.

(2) The Markov's inequality had appeared earlier in the work of Pafnuty Chebyshev, and for this reason it is sometimes referred to in other books as the first Chebyshev's inequality. Such books refer to the Chebyshev's inequality, discussed in the sequel, as the second Chebyshev inequality.

Download free eBooks at bookboon.com

Example 5.1
A textile factory produces on the average 150 bales of suiting material a month. Suppose the number of bales of a suiting material produced each month is a random variable. Find the bounds for the probability that a particular month's production will be

(a) at least 160 bales;

(b) less than 200 bales.

Solution
Let X be the number of bales of the suiting material produced in a month.
$E(X) = 150$

(a) We are required to find $P(X \geq 160)$.
It is better to use the first formulation of the Markov's inequality in Theorem 5.1 with $\epsilon = 160$. Hence

$$P(X \geq 160) \leq \frac{150}{160} = \frac{15}{16}$$

(b) We are required to find $P(X < 200)$. Using the second formulation we have

$$\begin{aligned} P(X < 200) &\geq 1 - \frac{150}{200} \\ &= \frac{1}{4} \end{aligned}$$

5.3 CHEBYSHEV'S INEQUALITY

Chebyshev's inequality was first formulated by Irnée-Jules Bienaymé in 1853 but without proof and it was not until 1867 that Chebyshev provided the proof. For this reason, the inequality is sometimes called Bienaymé-Chebyshev inequality.

5.3.1 Basic Chebyshev's Inequality

Theorem 5.2(a) CHEBYSHEV'S INEQUALITY
Let X be a random variable with finite expectation $E(X) = \mu$ and finite variance $\mathrm{Var}(X)$. Then for any real number $\epsilon > 0$,

$$P\left(|X - \mu| \geq \epsilon\right) \leq \frac{\mathrm{Var}(X)}{\epsilon^2}$$

Download free eBooks at bookboon.com

Proof
The inequality

$$|X - \mu| \geq \epsilon$$

is equivalent to the inequality

$$(X - \mu)^2 \geq \epsilon^2$$

The random variable $(X - \mu)^2$ takes only nonnegative values.
Applying to it the Markov's inequality, we have:

$$\begin{aligned} P(|X - \mu| \geq \epsilon) &= P[(X - \mu)^2 \geq \epsilon^2] \\ &\leq \frac{E\left[(X - \mu)^2\right]}{\epsilon^2} \\ &\leq \frac{\mathrm{Var}(X)}{\epsilon^2} \end{aligned}$$

Note

(1) The Chebyshev's inequality may be equivalently stated as in Theorem 5.2(b).

> **Theorem 5.2(b)** CHEBYSHEV'S INEQUALITY
> Let X be a random variable with finite expectation $E(X) = \mu$ and finite variance $\mathrm{Var}(X)$. Then for any real number $\epsilon > 0$,
>
> $$P\left(|X - \mu| < \epsilon\right) \geq 1 - \frac{\mathrm{Var}(X)}{\epsilon^2}$$

Proof
Since

$$\{|X - \mu| \geq \epsilon\}$$

and

$$\{|X - \mu| < \epsilon\}$$

are complementary events, Theorem 5.2(b) follows.

(2) As indicated earlier, the Theorem 5.2 is referred to in some books as the second Chebyshev's inequality.

Example 5.2
An electric station services an area with 12,000 bulbs. The probability of switching on each of these bulbs every evening is 0.9. What are the bounds for the probability that the number of bulbs switched on in the area in one particular evening is different from its expected value in absolute terms by (a) less than 100? (b) at least 120?

Download free eBooks at bookboon.com

Solution
Let

$$\begin{aligned} X &= \text{the number of bulbs switched on in the area in that evening} \\ \mu &= E(X) = np = 12000(0.9) = 10800 \\ \mathrm{Var}(X) &= npq = 12000(0.9)(0.1) = 1080 \end{aligned}$$

(a) We are required to calculate $P(|X-\mu| < 100)$.
We shall use the second Chebyshev's Inequality

$$P(|X-\mu| < \epsilon) \geq 1 - \frac{\mathrm{Var}(X)}{\epsilon^2}$$

with $\epsilon = 100$.
Then

$$P(|X-10800| < 100) \geq 1 - \frac{1080}{(100)^2} = 0.892$$

(b) We are required to calculate $P(|X-\mu| \geq 120)$.
We shall have to use the first Chebyshev's inequality

$$P(|X-10800| \geq 120) \leq \frac{1080}{(120)^2} = 0.075$$

Click on the ad to read more

Example 5.3
Suppose that a random variable has mean 5 and standard deviation 1.5. Use the Chebyshev's inequality to estimate the probability that an outcome lies between 3 and 7.

Solution
$\mu = 5, \quad \sigma = 1.5$
Since we wish to estimate the probability of an outcome lying between 3 and 7, we have

$$P(3 - 5 < X - \mu < 7 - 5) = P(-2 < X - \mu < 2) = P(|X - \mu| < 2)$$

That is $\epsilon = 2$. By Chebyshev's inequality

$$P(|X - \mu| < 2) \geq 1 - \frac{\text{Var}(X)}{\epsilon^2} = 1 - \frac{1.5}{2^2} = 0.625$$

The desired probability is at least 0.625. That is, if the experiment is repeated a large number of times, we expect, at least, 62.5% of the outcome to be between 3 and 7.

The following two theorems are other forms by which the Chebyshev's inequality are expressed.

Theorem 5.3(a)
Let X be a random variable with finite expectation $E(X) = \mu$ and finite variance $\text{Var}(X)$. Then for any positive number ϵ,

$$P(|X - \mu| \geq c\sigma) \leq \frac{1}{c^2}$$

where $\quad \epsilon = c\sigma \quad$ and $\quad \sigma = \sqrt{\text{Var}(X)}$

In words, the above theorem states that the probability that X assumes a value outside the interval from $\mu - c\sigma$ to $\mu + c\sigma$ is never more than $\dfrac{1}{c^2}$.

Theorem 5.3(b)
Let X be a random variable with finite expectation $E(X) = \mu$ and finite variance $\text{Var}(X)$. Then for any real number $\epsilon > 0$,

$$P(|X - \mu| < c\sigma) \geq 1 - \frac{1}{c^2}$$

where $\quad \epsilon = c\sigma \quad$ and $\quad \sigma = \sqrt{\text{Var}(X)}$

This states that the probability of the event that X takes on a value x which is within c standard deviations of its expectation is at least $1 - \dfrac{1}{c^2}$, no matter what c happens to be. That is, the probability that X assumes a value within the interval from $\mu - c\,\sigma$ to $\mu + c\,\sigma$ is never less than $1 - \dfrac{1}{c^2}$.

Download free eBooks at bookboon.com

Note

Theorems 5.3(a) and 5.3(b) are used only when $c > 1$. Now, if $c < 1$, then $1 - \frac{1}{c^2} < 0$ or $\frac{1}{c^2} > 1$, but we know that the probability of any event ranges from zero to one. Thus, Chebyshev's inequality of Theorems 5.3(a) and 5.3(b) is trivially true when $c < 1$.

Example 5.4

Suppose a random variable X has an expectation $\mu = 4.6$ and a variance $\sigma^2 = 2.25$. Find the bounds for the following probabilities:

(a) $P(|X - \mu| < 2\sigma)$ (b) $P(|X - \mu| < 3\sigma)$

Solution

$\mu = 4.6, \quad \sigma = \sqrt{2.25} = 1.5$

(a) $P[|X - 4.6| < 2(1.5)] \geq 1 - \frac{1}{2^2} = \frac{3}{4} = 0.75$

Thus,

$$P[|X - 4.6| < 3] \geq 0.75$$

(b) $P[|X - 4.6| \leq 3(1.5)] \geq 1 - \frac{1}{3^2} = \frac{8}{9} = 0.8889$

Thus,

$$P(|X - 4.6| < 4.5) \geq 0.8889$$

From Example 5.4, we can say that the probability that the random variable X will take on a value within two standard deviations from the mean is at least $\frac{3}{4}$, and the probability that X will take on a value within three standard deviations from the mean is at least $\frac{8}{9}$. It is in this sense that the standard deviation σ controls the spread or dispersion of the distribution of a random variable.

5.3.2 Special Cases of Chebyshev's Inequality

The following special cases of Chebyshev's inequality are very helpful in sampling theory. They can be used to determine the minimum sample sizes (number of trials).

Theorem 5.4

Let $\overline{X_n}$ be the sample mean based on a random sample of size n on a random variable X with expectation μ and finite variance σ^2. Then for any real number ϵ,

$$P(|\overline{X}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$$

or equivalently,

$$P(|\overline{X}_n - \mu| < \epsilon) \geq 1 - \frac{\sigma^2}{n\epsilon^2}$$

where n is the sample size

Download free eBooks at bookboon.com

Example 5.5
Let $\overline{X}$ denote the mean of random variables X_i with expectation $\mu = 100$ and variance $\sigma^2 = 2.5$. Find the bound for the probability that in 120 trials the mean differs from the expected value in absolute terms by less than 0.8.

Solution
$\epsilon = 0.8, \quad n = 120, \quad \sigma^2 = 2.5$

$$P(|\overline{X}_n - \mu| < 0.8) \geq 1 - \frac{2.5}{120(0.8)^2} = 0.974$$

Example 5.6
Referring to the $\overline{X}$ of Example 5.5, determine the size of n such that

$$P(|\overline{X}_n - \mu| < 0.8) \geq 0.99$$

Solution

$$P(|\overline{X}_n - 100| < 0.8) \geq 0.99$$

Click on the ad to read more

Comparing this with Theorem 5.4 we have:

$$\begin{aligned} 1-\frac{\sigma^2}{n\epsilon^2} &= 0.99 \\ \Rightarrow \qquad \frac{\sigma^2}{n\epsilon^2} &= 0.01 \end{aligned}$$

Rearranging we obtain

$$n \geq \frac{\sigma^2}{0.01\epsilon^2} = \frac{2.5}{0.01(0.8)^2} = 390.6$$

That is, we need at least 391 trials in order that the probability will be at least 0.99 that the sample mean $\overline{X}_n$ will lie within 0.8 of the expectation.

Going through the solution of Example 5.5, we may write the formula for determining the sample size n by Chebyshev's inequality as

$$n = \frac{\sigma^2}{q\epsilon^2}$$

where $q = 1 - p$.

Another form of Chebyshev's inequality which is of great importance is the Bernoulli forms which are the applications of the De Moivre-Laplace Integral Theorem.

Theorem 5.5 FIRST BERNOULLI THEOREM
Let $\mathcal{E}$ be an experiment and $\mathcal{A}$ an event associated with $\mathcal{E}$. Consider n independent trials of $\mathcal{E}$. Let the random variable M_n be the number of times event $\mathcal{A}$ occurs among the n trials. Then by Chebyshev's inequality

$$P(|M_n - np| \geq \epsilon) \leq \frac{npq}{\epsilon^2}$$

or equivalently,

$$P(|M_n - np| < \epsilon) \geq 1 - \frac{npq}{\epsilon^2}$$

Proof
The Theorem follows from Chebyshev's inequality (Theorem 5.2) by remembering that $E(M_n) = np$ and $\mathrm{Var}(M_n) = npq$.

The reader should attempt Exercise 5.26 (a). It is a typical example of Theorem 5.5.

Theorem 5.5 is true for any value we give to the $\epsilon > 0$. In particular, if we replace ϵ by $n\,\delta$, where δ is thought of as "small", then we have the following theorem.

Download free eBooks at bookboon.com

Theorem 5.6 SECOND BERNOULLI THEOREM
Let $\mathcal{E}$ be an experiment and $\mathcal{A}$ be an event associated with $\mathcal{E}$. Consider n independent trials of $\mathcal{E}$. Let $\dfrac{M_n}{n}$ be the relative frequency of the occurrence of event $\mathcal{A}$ in n trials. Then

$$P\left(\left|\frac{M_n}{n} - p\right| \geq n\,\delta\right) \leq \frac{pq}{n\delta^2}$$

or equivalently,

$$P\left(\left|\frac{M_n}{n} - p\right| < n\,\delta\right) \geq 1 - \frac{pq}{n\delta^2}$$

For any value of δ, the right side of the second part of Theorem 5.6 converges to 1 as $n \to \infty$. What it means is that for any value of δ, *no matter how small,* the probability that the proportion of successes in n trials differs from the theoretical probability p by less than δ tends to 1 as the number of trials increases without bound. That is, it is guaranteed that the *observed* relative frequency of successes will converge to the *theoretical* relative frequency (as measured by p) as the number of trials tends to infinity (see Section 5.4).

Example 5.7
The probability of the occurrence of an event $\mathcal{A}$ in each trial of an experiment is $\dfrac{2}{3}$. Using Chebyshev's inequality, find a lower bound for the probability that in 10,000 trials the deviation of the relative frequency of the event $\mathcal{A}$ from the true probability of $\mathcal{A}$ will be less than 0.01.

Solution
The relative frequency in n independent trials is a random variable. Its expectation equals $p = \dfrac{2}{3}$.
We are required to calculate

$$P\left(\left|\frac{M_n}{10,000} - \frac{2}{3}\right| < 0.01\right)$$

Then,

$$\begin{aligned} P\left(\left|\frac{M_n}{10,000} - \frac{2}{3}\right| < 0.01\right) &\geq 1 - \frac{\left(\frac{2}{3}\right)\left(\frac{1}{3}\right)}{(10,000)(0.01)^2} \\ &= 1 - \frac{2}{9} = \frac{7}{9} \approx 0.778 \end{aligned}$$

Thus, with probability not less than 0.778 we may expect that in 10,000 trials the relative frequency of event $\mathcal{A}$ will deviate from its probability for less than 0.01.

Download free eBooks at bookboon.com

Example 5.8
How many times should a fair die be tossed in order to be at least 95% sure that the relative frequency of having a four come up is within 0.01 of the theoretical probability $\frac{1}{6}$.

Solution
$p = \frac{1}{6}, \quad 1 - p = \frac{5}{6} \quad \text{and} \quad \epsilon = 0.01$

Let $\frac{M_n}{n}$ be the relative frequency. Then

$$P\left(\left|\frac{M_n}{n} - p\right| < 0.01\right) \geq 0.95$$

By the second Bernoulli Theorem,

$$P\left(\left|\frac{S_n}{n} - p\right| < \epsilon\right) \geq 1 - \frac{pq}{n\epsilon^2}$$

Thus

$$\begin{aligned} 1 - \frac{pq}{n\epsilon^2} &= 0.95 \\ \Rightarrow \quad \frac{pq}{n\epsilon^2} &= 0.05 \\ \Rightarrow \quad n &= \frac{pq}{0.05(\epsilon^2)} \\ &= \frac{\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)}{0.05(0.01)^2} \\ &= 27777.778 \\ &\approx 27,778 \end{aligned}$$

The Chebyshev's inequality provides only crude and general results. This limitation arises because of its complete universality; it is valid under any distribution provided both the expectation and the variance are finite. But this condition will always be satisfied when the number of values of the random variable X is finite. In the process of achieving such a general result, however, this inequality is not particularly tight in terms of the bound achieved on particular distributions. For most distributions that arise in practice, there are far sharper bounds for $P(|X - \mu| < \epsilon)$ than that given by Chebyshev's inequality. For instance, if the random variable X is normally distributed, then the following formula derived in Chapter 11 of Volume I is more exact:

$$P(|X - \mu| < \epsilon) = 2\Phi\left(\frac{\epsilon}{\sigma}\right) - 1$$

Example 5.9
Referring to Example 5.4, use the Normal distribution to find the probabilities and compare the results.

Solution
$\mu = 4.6, \quad \sigma = \sqrt{2.25} = 1.5$

Download free eBooks at bookboon.com

(a) $2\sigma = 2(1.5) = 3$

$$\begin{aligned} P(|X-\mu| < 2\sigma) &= P\left(|X-\mu| < 3\right) \\ &= 2\Phi\left(\frac{3}{1.5}\right) - 1 \\ &= 2\Phi(2) - 1 \\ &= 2(0.9972) - 1 \\ &= 0.9544 \end{aligned}$$

(b) $3\sigma = 3(1.5) = 4.4$

$$\begin{aligned} P\left(|X-\mu| < 3\sigma\right) &= P\left(|X-\mu| < 4.5\right) \\ &= 2\Phi\left(\frac{4.5}{1.5}\right) - 1 \\ &= 2\Phi(3) - 1 \\ &= 2(0.9987) - 1 \\ &= 0.9974 \end{aligned}$$

In both cases the probabilities are greater than those obtained by Chebyshev's inequality in Example 5.4.

Need help with your dissertation?

Get in-depth feedback & advice from experts in your topic area. Find out what you can do to improve the quality of your dissertation!

Get Help Now

Go to www.helpmyassignment.co.uk for more info

Helpmyassignment

Click on the ad to read more

Note that Chebyshev's inequality provides nontrivial estimate of the probability of the event $\{|X-\mu|<\epsilon\}$ only in the case when the variance of the random variable X is sufficiently small, namely, less than ϵ^2.

Clearly, the left side of the inequality in Theorem 5.2, expressing the probability of an event is nonnegative. It gives the estimate of the probability of the event $\{|X-\mu|<\epsilon\}$, if the right side is positive.
Thus,

$$1-\frac{\mathrm{Var}(X)}{\epsilon^2}>0, \quad \text{or} \quad \frac{\mathrm{Var}(X)}{\epsilon^2}<1, \quad \text{or} \quad \mathrm{Var}(X)<\epsilon^2$$

If, on the other hand, $\mathrm{Var}(X)>\epsilon^2$, then the right side of the inequality of Theorem 5.2 will become negative and the Chebyshev inequality will give

$$P(|X-\mu\leq\epsilon)<-a$$

where $a>0$.

Thus, Chebyshev's inequality is trivial for the case when $\mathrm{Var}(X)>\epsilon^2$. This of course reduces the role of Chebyshev's inequality in its application to practical problems; however, its theoretical importance is great. It is the starting point for several theoretical development. It provides us with a convenient interpretation of the concept of variance (or standard deviation). It can also be used to provide a simple proof for the law of large numbers in the next section.

Note
The inequality that helps us derive bounds for sums of independent random variables is the Komogorov's inequality.

Suppose that $X_1, X_2, \cdots, X_n$ are independent random variables with mean zero and finite variances and let $S_k=X_1+X_2+\cdots+X_n$. Then for $\epsilon>0$

$$P\left(\max_{1\leq k\leq n}|S_k|\geq\epsilon\right)\leq\frac{\mathrm{Var}(S_n)}{\epsilon^2}$$

For the proof see page 248 of Billinsley, P. (1979) listed in the bibliography.

For $n=1$, Komogorov's inequality reduces to Chebyshev's inequality. For $n>1$, Chebyshev's inequality gives the same bound for the single relation $|S_k|\geq\epsilon$, so that Komogorov's inequality is considerably stronger.

5.4 LAW OF LARGE NUMBERS

5.4.1 Definition of Law of Large Numbers

The Law of Large Numbers is an important consequence of Chebyshev's inequality. It is commonly believed that if a fair coin is tossed many times the number of heads is about the same as the number of tails. The law of large numbers is a mathematical formulation of this belief.

Download free eBooks at bookboon.com

Definition 5.1 LAW OF LARGE NUMBERS
The law of large numbers states that the *average* of a number of independent and identically distributed random variables *converge* to the expected value of the underlying distribution of X as the number of random variables increases

Whatever the "law of averages" may be, it is certainly not reasonable in a hundred tosses of a fair coin to expect exactly 50 tails or a million tosses to expect exactly 500,000 heads. To be more reasonable, perhaps the best we can get is that usually the proportion of tails in n tosses is *close* to $\frac{1}{2}$. J.E. Kerrich who, while interned in Denmark as a British subject during the Second World War, tossed a coin 10,000 times and recorded 5067 heads.

Before we give the various forms of the law of large numbers we shall discuss briefly the term "converge" used in the definition. Suppose we have a sequence $X_1, X_2, \cdots$ of random variables defined on the same sample space. In what follows we shall distinguish various possible modes in which the sequence $\{X_n\}$ may converge. Convergence generally comes in two categories which we call *strong* and *weak*.

Theorem 5.7 CONVERGENCE IN PROBABILITY
Suppose $\{X_n\}$ is a sequence of random variables ($n = 1, 2, \cdots$). Then if for every $\epsilon > 0$

$$P(|X_n - \theta| > \epsilon)$$

approaches zero as n approaches infinity, then $\{X_n\}$ is said to converge in probability to θ

Symbolically, we may write convergence in probability as

$$\lim_{n\to\infty} P(|X_n - \theta| > \epsilon) \to 0$$

or

$$X_n \xrightarrow{p} \theta$$

Synonyms for convergence in probability are *stochastic convergence, convergence in measure, or weak convergence.*

Theorem 5.8 CONVERGENCE WITH PROBABILITY
Suppose $\{X_n\}$ is a sequence of random variables ($n = 1, 2, \cdots$). Then if

$$P(|X_n = X|)$$

is 1 as n approaches infinity, then $\{X_n\}$ is said to converge with probability one (w.p. 1) to X

Download free eBooks at bookboon.com

That is, X_n actually converges to X except in a negligible number of cases. This form of convergence is also called *almost sure convergence (a.s.)*, or *almost everywhere convergence (a.e.)*, or *strong convergence.*

Symbolically, we may write convergence in probability as

$$\lim_{n\to\infty} P(|X_n = X|) = 1$$

or

$$X_n \xrightarrow{a.s.} X$$

These two types of convergence[10] are the basis of the two laws of large numbers, namely, the weak and the strong laws of large numbers. For our purposes, which are statistical, the weak law of large numbers is the central concept and when the "law of large numbers" is referred to without qualification, this one is implied. We shall discuss it in much detail later but now we state without proof, the strong law of large numbers.

[10] The other type of convergence that plays an important role in Probability theory is convergence in distribution, also called *complete convergence or weak convergence.* Suppose $\{X_n\}$ is a sequence of random variables, $(n = 1, 2, \cdots)$ and let $F_n(t)$ be their c.d.f.'s. If for every t at which $F_0(t)$ is continuous,

$$\lim_{n\to\infty} F_n(t) = F_0(t)$$

then X_n is said to converge in distribution to X_0, denoted by $F_n(t) \xrightarrow{d} F_0(t)$.

Click on the ad to read more
Download free eBooks at bookboon.com

5.4.2 Strong Law of Large Numbers

Even though the strong law of large numbers may not be realistic, what might not be realised in the real-world situation may sometimes be achieved in a purely theoretical sense. Such a possibility was unravelled by Borel in 1909, who established the *strong law of large numbers* in the case of independent Bernoulli trials.

Theorem 5.9 STRONG LAW OF LARGE NUMBERS **(Bernoulli Form)**

Let M_n be a random variable of the number of successes in n Bernoulli trials so that $\frac{M_n}{n}$ is the proportion of successes. Then

$$P\left(\lim_{n\to\infty}\frac{M_n}{n}=p\right)=1$$

where p is the probability of success

Theorem 5.10 STRONG LAW OF LARGE NUMBERS **(Khinchin Form)**

Let $X_1, X_2, \cdots, X_n$ be a sequence of independent random variables having a common distribution and let $E(X)=\mu$. Then there is a limit number μ, such that

$$P\left(\lim_{n\to\infty}\overline{X}_n=\mu\right)=1$$

where $\overline{X}_n=\sum_{i=1}^{n}\frac{X_i}{n}$

The proof of this theorem goes beyond this book. To see it refer to page 250 of Billingsley, P. (1979), listed in the bibliography.

The strong law of large numbers makes better sense than the weak law and it is indispensable for certain theoretical investigations. It is indeed the foundation of a mathematical theory of probability based on the concept of relative frequency. It is also known that the sample mean, $\overline{X}_n$, converges with probability 1 to the mean μ, provided that the latter exists. The strong law of large numbers is usually not given much attention in most statistical textbooks, including this one partly because it is not realistic to assert this almost sure convergence, for in any experiment we can neither be 100 percent sure nor 100 percent accurate, otherwise the phenomenon will not be a random one. Secondly, in general, it is easier to prove the weak law of large numbers than the strong law of large numbers.

Download free eBooks at bookboon.com

5.4.3 Weak Law of Large Numbers

The weak law of large numbers is one of the earliest and most famous of the limit laws of probability. We shall consider two of its forms, namely, Bernoulli law (which relates to proportions) and Khinchin law (which relates to the means).

Bernoulli Law of Large Numbers *(Discrete Case)*

The first formulation of the law of large numbers, known as the Bernoulli law of large numbers, was given and proved by Jakob Bernoulli and published posthumously in his book "Ars Conjectandi" in 1713 as a crowning achievement. It states that if S_n represents the number of successes in n identical Bernoulli trials, with probability of success p in each trial, then the relative frequency $\frac{M_n}{n}$ is very likely to be close to p when n is a sufficiently large and fixed integer. The law in a sense justifies the use of the frequency definition of probability discussed in Chapter 3 of Volume I and it brings the theory of probability into contact with practice.

Mathematically, the Bernoulli law of large numbers may be expressed in the following theorem.

Theorem 5.11 BERNOULLI LAW OF LARGE NUMBERS
Let $X_1, X_2, \cdots, X_n$, be a sequence of independent Bernoulli random variables, each taking on the value 1 (with the occurrence of an event $\mathcal{A}$) or 0 (with non-occurrence of an event $\mathcal{A}$), and let $p = P(X_i = 1); i = 1, 2, ..., n$. Then for sums M_n, the number of successes in n trials, where

$$M_n = \sum_{i=1}^{n} X_i,$$

$$\lim_{n\to\infty} P\left(\left|\frac{M_n}{n} - p\right| < \epsilon\right) = 1$$

for all $\epsilon > 0$ or equivalently,

$$\lim_{n\to\infty} P\left(\left|\frac{M_n}{n} - p\right| \geq \epsilon\right) = 0$$

for all $\epsilon > 0$

That is, for n large, $\frac{M_n}{n}$ is very close to p. In other words the law of large numbers may be stated as follows: In n trials the probability that the relative number of successes $\frac{M_n}{n}$, deviates numerically from the true probability p by not more than $\epsilon\,(\epsilon > 0)$, approaches 1 as n approaches infinity.

Or equivalently, in n trials the probability that the relative number of successes $\frac{M_n}{n}$, deviates numerically from the true probability p by more than $\epsilon\,(\epsilon > 0)$, approaches 0 as n approaches infinity.

Download free eBooks at bookboon.com

Khinchin Theorem of Large Numbers *(Continuous Case)*

The modern discussion of a mathematical version of the weak law of large numbers was first published by the Russian mathematician, Khinchin in 1929. The Khinchin Theorem states "If X_i $(i = 1, 2, ..., n)$ are independent and identically distributed random variables and if $E(X_i) = \mu$ exists, then when n becomes a very large integer, the probability that the random variable $\overline{X}$ will differ from the common expectation μ of X_i by not more than any arbitrarily prescribed small difference ϵ is very close to one." Mathematically, this theorem can be expressed in the form as in the following theorem.

Theorem 5.12(a) KHINCHIN LAW OF LARGE NUMBERS I
Let $X_1, \cdots, X_n$ be a sequence of independent and identically distributed continuous random variables with finite common mean μ and finite common variance σ^2. Then for any real number $\epsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\overline{X}_n - \mu\right| < \epsilon\right) = 1$$

or equivalently,

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| < \epsilon\right) = 1$$

where $S_n = X_1 + X_2 + \cdots + X_n$

Download free eBooks at bookboon.com

Click on the ad to read more

Proof
By Theorem 5.4,

$$P\left(\left|\overline{X}_n - \mu\right| < \epsilon\right) = 1 - \frac{\sigma^2}{n\epsilon^2}$$

so that

$$\begin{aligned} 1 - \frac{\sigma^2}{n\epsilon^2} \; &\le \; P(|\overline{X}_n - \mu| < \epsilon) \; \le \; 1 \\ \lim_{n\to\infty}\left(1 - \frac{\sigma^2}{n\epsilon^2}\right) \; &\le \; \lim_{n\to\infty} P(|\overline{X}_n - \mu| < \epsilon) \; \le \; 1 \\ 1 \; &\le \; \lim_{n\to\infty} P(|\overline{X}_n - \mu| < \epsilon) \; \le \; 1 \end{aligned}$$

By Sandwich rule of limits

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| < \epsilon) = 1$$

It is in this sense that the arithmetic mean "converges" to $E(X) = \mu$. That is, by averaging an increasingly large number of observations of the value of a quantity, we can obtain increasingly more accurate measures of the expectation of that quantity.

Aliter
We first find $E(\overline{X}_n)$ and $\text{Var}(\overline{X}_n)$:

$$E(\overline{X}_n) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu \qquad \text{(see Chapter 7).}$$

Since the X_i are independent and identically distributed,

$$\text{Var}(\overline{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n} \text{Var}(X_i) = \frac{\sigma^2}{n} \qquad \text{(see Chapter 7).}$$

This desired result now follows immediately from Chebyshev's inequality as follows

$$P\left(\left|\overline{X}_n - \mu\right| < \epsilon\right) \ge 1 - \frac{\text{Var}(\overline{X}_n)}{\epsilon^2} = 1 - \frac{\sigma^2}{n\epsilon^2} \to 1 \text{ as } n \to \infty$$

Theorem 5.12(b) KHINCHIN LAW OF LARGE NUMBERS II
The law of large numbers may equivalently be stated as

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| \ge \epsilon) = 0$$

or equivalently,

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| \ge \epsilon\right) = 0$$

where $S_n = X_1 + X_2 + \cdots + X_n$

Download free eBooks at bookboon.com

The reader is asked in Exercise 5.27 to prove this theorem.

Note
To apply the weak law of large numbers, the random variables do not have to be identically distributed; they just need to have the same mean and the same finite variance.

5.5 CENTRAL LIMIT THEOREM

5.5.1 Definition of Central Limit Theorem

Among all probability laws, the *Central Limit Theorem* (CLT) is considered the most remarkable theoretical formulation from both practical and theoretical viewpoints. It provides some reassurance when we are not certain whether observations are normally distributed. Whatever the nature of the random variables X_1, X_2, $\cdots$ – whether they are discrete or continuous, Bernoulli or gamma – as n gets larger and larger the distribution of the sum (after standardisation) always gets closer and closer to that of the Standard Normal distribution. In fact, it is the most important reason for the central role of the Normal distribution in Statistics.

Generally, the name "Central Limit Theorem" is used to describe any convergence theorem in which the Normal distribution appears as the limit. More particularly, it applies to sums of large numbers or random variables having approximately the Standard Normal distribution after standardisation. Though the name "Central Limit Theorem" was given by G. Polya in 1920, the concept was first introduced by De Moivre early in the eighteenth century. De Moivre proved special cases in 1733 and Laplace later in his book of 1812 gave the first (incomplete) statement of the general form of the CLT so the CLT of Theorem 5.16 in the sequel (the Bernoulli form) is often referred to as the De Moivre–Laplace CLT. The modern CLT (Theorem 5.15 in the sequel) began to emerge from the work of the Russian school of Probabilists–notably Chebyshev and his students Markov and Lyapunov towards the close of the nineteenth century. In fact Lyapunov proved the first general version of CLT in about 1901, though the details of his proof were complicated.

The central limit theorem has been expressed in many forms, differing in various degrees of abstraction and generality. We shall present some of the simplest versions but first we state its basic theorems.

Theorem 5.13 POLYA'S THEOREM
Let $X_1, X_2, \cdots, X_n$ be a sequence of independent and identically distributed random variables with finite mean $E(X) = \mu$ and fine variance $\text{Var}(X) = \sigma^2 > 0$ and let

$$Z_n^* = \frac{\sqrt{n}(\overline{X_n} - \mu)}{\sigma}$$

Then

$$P(Z_n^* \leq z^*) \xrightarrow{d} \Phi(z^*) \quad \text{as } n \to \infty$$

where Φ is the c.d.f. of the Standard Normal distribution

Download free eBooks at bookboon.com

We shall observe in Theorem 5.17 in the sequel that Polya's Theorem is the central limit theorem of means.

> **Theorem 5.14** CENTRAL LIMIT THEOREM **(Lindberg-Levy Theorem)**
> Suppose $Z_1, Z_2, \cdots, Z_n$ are independent and identically distributed Standard Normal random variables and let
>
> $$S_n^* = \frac{Z_1 + Z_2 + \cdots + Z_n}{\sqrt{n}}$$
>
> Then
>
> $$P(S_n^* \leq s^*) \to \Phi(s^*) \quad \text{as } n \to \infty$$
>
> where Φ is the c.d.f. of the Standard Normal distribution

For proof see page 316 of Dudewicz and Mishra, 1988, listed among the bibliography.

Click on the ad to read more

5.5.2 Central Limit Theorem of Sums

The Central Limit Theorem is concerned with the distribution of the "sum of random variables". If $X_1, X_2, ...,$ is a sequence of independent random variables with expectation $E(X)$ and variance Var(X) and if

$$S_n = \sum_{i=1}^{n} X_i$$

we know from the law of large numbers that $\frac{S_n}{n}$ converges to $E(X)$ in probability. The Central Limit Theorem is concerned not with the fact that the ratio $\frac{S_n}{n}$ converges to $E(X)$ but with how it fluctuates around $E(X)$. To analyse these fluctuations, we standardise the sum

$$T_n = \frac{S_n - E(S_n)}{\sqrt{\mathrm{Var}(S_n)}} \qquad (i)$$

It is easily verified that T_n has mean 0 and variance 1.

As usual we shall discuss two forms of CLT, namely, the Khinchin form (continuous case) and the Bernoulli form (discrete case.)

Continuous Case

Theorem 5.15 CENTRAL LIMIT THEOREM **(Khinchin Form)**
Let $X_1, \cdots, X_n$ be a sequence of independently and identically distributed random variables, each having mean μ and variance σ^2. Let

$$S_n = \sum_{i=1}^{n} X_i.$$

Then the distribution of

$$T_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

tends to the Standard Normal distribution when n becomes infinite

Proof
Since X_n's are independent and identically distributed, then from Corollary 3.4 and Theorem 3.19

$$E(S_n) = n\mu, \quad \text{and} \quad \mathrm{Var}(S_n) = n\sigma^2$$

Substituting these expressions in (i) above the result follows immediately.

Note
These theorems will also hold when $X_1, \cdots, X_n$ are independent random variables with the same mean and the same finite variance but not necessarily identically distributed.

Download free eBooks at bookboon.com

Discrete Case

The De Moivre-Laplace Theorem discussed in Chapter 12 of Volume I is actually a special case of the Central Limit Theorem, namely, the Bernoulli form of CLT of frequency.

The Bernoulli and other forms of the central limit theorem arise out of Theorem 5.15. For instance, if the random variables X_i's are independent Bernoulli random variables, their sum

$$M_n = \sum_{i=1}^{n} X_i$$

is a binomial random variable. Substituting the expressions

$$E(M_n) = np \quad \text{and} \quad \text{Var}(M_n) = npq$$

in (i) above, the result follows immediately.

Theorem 5.16 CENTRAL LIMIT THEOREM **(Bernoulli Form)**
Let $X_1, X_2, \cdots, X_n$ be independent Bernoulli random variables with parameter p so that $E(X_i) = p$ and $\text{Var}(X_i) = pq$. If

$$M_n = \sum_{i=1}^{n} X_i$$

then

$$Y_n = \frac{M_n - np}{\sqrt{npq}}$$

tends to the standard Normal distribution as $n \to \infty$

Thus, the central limit theorem may mathematically be stated as

$$\lim_{n\to\infty} P(Z_n \leq z) = \Phi(z),$$

where Φ is the cumulative distribution function of the standard Normal distribution.

Significance of the Central Limit Theorem

The computational significance of the CLT is that for large n, we can express the cumulative distribution function of T_n in terms of $N(0, 1)$ as follows.

In the case of frequencies,

$$P(Y_n \leq y) \approx \Phi\left(\frac{y - np}{npq}\right)$$

where y is a particular value of Y_n.
In the case of continuous random variables

$$P(T_n \leq t) \approx \Phi\left(\frac{t - n\mu}{\sigma\sqrt{n}}\right)$$

where t is a particular value of T_n.

Download free eBooks at bookboon.com

The logical question that arises at this point is: How large should the n be to enable us apply the Central Limit Theorem. There is no single answer, since this depends on the closeness of approximation required and the actual distribution forms of X_i's. *If the random variables X_i's are normally distributed, then no matter how small n is, their sum is also normally distributed* and $P(T_n \leq t)$ provides exact probabilities. If nothing is known about the distribution patterns of X_i's, or if the distribution of X_i's differs greatly from normality then n must be large enough to guarantee approximate normality for S_n. One rule of thumb states that, in most practical situations, n equals at least 30 is satisfactory. In general, the approximation of the sum of random variables to normality becomes better and better as the sample size increases.

Example 5.10
A die is thrown ninety times. At a given throw the expected number of points is $\frac{7}{2}$ and the variance is $\frac{32}{12}$. Find

(a) the expectation of the sum of points;

(b) the variance of the sum of points;

(c) the probability that the sum obtained is at most 300.

Download free eBooks at bookboon.com
Click on the ad to read more

Solution
$n = 90, \quad \mu = \dfrac{7}{2}, \quad \sigma^2 = \dfrac{32}{12}$

(a) $E(S_n) = n\mu = 90\left(\dfrac{7}{2}\right) = 315$

(b) $\mathrm{Var}(S_n) = n\sigma^2 = 90\left(\dfrac{32}{12}\right) = 240$

(c) $s = 300$

$$P(S_n \leq s) = P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq \frac{s - n\mu}{\sigma\sqrt{n}}\right) = P\left(T_n \leq \frac{s - n\mu}{\sigma\sqrt{n}}\right)$$

$$\begin{aligned}
\text{Thus,} \qquad P(S_n \leq 300) &= P\left(T_n \leq \frac{300 - 315}{\sqrt{\frac{32}{12}}\sqrt{90}}\right) \\
&= P\left(T_n \leq \frac{-15}{(1.6330)(9.4868)}\right) \\
&= P(T_n \leq -0.968) \\
&= 1 - \Phi(0.97) \\
&= 1 - 0.8340 = 0.166
\end{aligned}$$

5.5.3 Central Limit Theorem of Means

Another important version of the Central Limit Theorem is expressed in terms of the means of independent and identically distributed random variables.

Theorem 5.17 CENTRAL LIMIT THEOREM **(of Means)**
Let $X_1, \cdots, X_n$ be n independent and identically distributed random variables with means equal to μ and variances equal to σ^2. Let $\overline{X_n}$ be the random variable defined by

$$\overline{X_n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

The distribution of

$$Z_n = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$$

tends to the Standard Normal distribution as $n \to \infty$

Note
Z_n is the standardised mean of the random variables $X_1, \cdots, X_n$ and it is equal to T_n in Theorem 5.16.

Download free eBooks at bookboon.com

Proof
The proof is quite sophisticated. An outline of a proof is given in Hoel (1971), and Freund and Walpole (1971) listed among the references.

Example 5.11
Suppose that I.Q. is a random variable with mean $\mu = 100$ and standard deviation $\sigma = 25$. What is the probability that in a class of 40 students

(a) the average I.Q. exceeds 100?

(b) a single person's I.Q. exceeds 100?

Solution

$$n = 40,\ \mu = 110,\ \sigma = 25,\ \sigma_{\overline{X}} = \frac{25}{\sqrt{40}} = 3.9528$$

By the Central Limit Theorem, the mean I.Q. of the 40 students is normally distributed. Hence
(a) The probability that the average I.Q. exceeds 100 is

$$\begin{aligned} P(\overline{X} \geq 100) &= P\left(Z \geq \frac{110 - 100}{3.9528}\right) \\ &= P(Z \geq 2.53) \\ &= 1 - \Phi(2.53) \\ &= 1 - 0.9943 \\ &= 0.0057 \end{aligned}$$

So, approximately, 0.6% of the mean I.Q.'s of a class of 40 students will exceed 100.

(b) The probability that a single person's I.Q. exceeds 100 is

$$\begin{aligned} P(X > 100) &= P\left(Z \geq \frac{110 - 100}{25}\right) \\ &= P(Z \geq 0.4) \\ &= 1 - \Phi(0.4) \\ &= 1 - 0.6554 \\ &= 0.3446 \end{aligned}$$

Theorem 5.18
Suppose that $X_1, X_2, \cdots, X_n$ are independent and identically distributed random variables, each having mean μ and $\mathrm{Var}(X) = \sigma^2$. Suppose also that the average of the measurements, $\overline{X}$, is used as an estimate of μ. Then

$$P(|\overline{X} - \mu| < \epsilon) \approx \Phi\left(\frac{\epsilon\sqrt{n}}{\sigma}\right) - \Phi\left(-\frac{\epsilon\sqrt{n}}{\sigma}\right)$$

Download free eBooks at bookboon.com

Proof

Suppose that we wish to find

$$P(|\overline{X} - \mu| < \epsilon)$$

for some constant $\epsilon > 0$. To use the Central Limit Theorem to approximate this probability, we standardise the mean, using $E(\overline{X}) = \mu$ and $\mathrm{Var}(\overline{X}) = \dfrac{\sigma^2}{n}$;

$$\begin{aligned} P(|\overline{X} - \mu| < \epsilon) &= P(-\epsilon < \overline{X} - \mu < \epsilon) \\ &= P\left(\frac{-\epsilon}{\frac{\sigma}{\sqrt{n}}} < \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < \frac{\epsilon}{\frac{\sigma}{\sqrt{n}}} \right) \\ &\approx \Phi\left(\frac{\epsilon\sqrt{n}}{\sigma} \right) - \Phi\left(-\frac{\epsilon\sqrt{n}}{\sigma} \right) \end{aligned}$$

If we use the Standard Normal Table I in the Appendix, then we may write this formula as

$$P(|\overline{X} - \mu| < \epsilon) \approx 2\Phi\left(\frac{\epsilon\sqrt{n}}{\sigma} \right) - 1$$

On the other hand

$$P(|\overline{X} - \mu| < \epsilon) \approx 2\Psi\left(\frac{\epsilon\sqrt{n}}{\sigma} \right)$$

if we use Table II in the Appendix.

Click on the ad to read more

The Law of Large Numbers tells us that $\overline{X}$ converges to μ in probability; so we can hope that $\overline{X}$ is close to μ if n is large. Chebyshev's inequality allows us to determine the bound of the probability of an error of a given size, but the Central Limit Theorem gives a much sharper approximation to the actual error.

Example 5.12
Suppose that 25 measurements are taken with $\sigma = 1.2$. Find the probability that the sample mean $\overline{X}$ deviates from μ by less than 0.3.

Solution
$n = 25, \quad \sigma = 1.2, \quad e = 0.3$

$$\begin{aligned} P(|\overline{X} - \mu| < 0.3) &\approx \Phi\left(\frac{0.3\sqrt{25}}{1.2}\right) - \Phi\left(-\frac{0.3\sqrt{25}}{1.2}\right) \\ &= 2\Phi\left(\frac{0.3\sqrt{25}}{1.2}\right) - 1 \\ &= 2\Phi(1.25) - 1 \\ &= 0.7888 \end{aligned}$$

This sort of reasoning can be turned around. That is, given ϵ and Y, n can be found such that

$$P(|\overline{X_n} - \mu| < \epsilon) \geq \theta$$

where θ is the probability that corresponds to the z value in the Standard Normal Table (see Tables I and II in the Appendix).

With the Central Limit Theorem we can derive a formula for calculating the value of n such that the probability that the mean deviates from μ by ϵ is θ :

$$n = \left\{\frac{\sigma}{\epsilon}\,\Phi^{-1}\left(\frac{1+\theta}{2}\right)\right\}^2$$

if we use the Full Normal Table (Table I in the Appendix), or equivalently,

$$n = \left\{\frac{\sigma}{\epsilon}\,\Psi^{-1}\left(\frac{\theta}{2}\right)\right\}^2$$

using the Half Normal Table (Table II in the Appendix).

Example 5.13
Refer to Example 5.12. Suppose $\sigma = 1.2$ and $\epsilon = 0.3$, find n such that the probability that the mean deviates from μ by 0.3.

Solution
$\sigma^2 = 2 \quad \epsilon = 0.25 \quad \theta = 0.75$

Therefore

$$n = \left[\frac{2}{0.25}\,\Phi^{-1}\left(\frac{1+0.75}{2}\right)\right]^2 = [8\,\Phi^{-1}(.875)]^2 = 8(1.15)^2 = 84.64 \approx 85$$

Download free eBooks at bookboon.com

5.5.4 Central Limit Theorem of Proportions

Sometimes we may have to work with proportions of successes in n independent trials than with actual number of successes. In this case $\frac{M_n - np}{\sqrt{npq}}$ becomes

$$\frac{\dfrac{M_n}{n} - p}{\sqrt{\dfrac{pq}{n}}}$$

and this leads us to the following theorem.

> **Theorem 5.19** CENTRAL LIMIT THEOREM **(Proportions)**
>
> Suppose in n Bernoulli trials with M_n successes, $\dfrac{M_n}{n}$ is the proportion (relative frequency) of successes and the probability of success in each trial is p. Then
>
> $$Y_n = \frac{\dfrac{M_n}{n} - p}{\sqrt{\dfrac{pq}{n}}}$$
>
> tends to the Standard Normal distribution as $n \to \infty$

We conclude this chapter by summarising the two limit laws as follows. Generally, if a theorem is concerned with stating conditions under which variables converge (in some sense) to the expected average, then it is a *law of large numbers*. If on the other hand, the theorem is concerned with determining conditions under which the sum of a large number of random variables has a probability distribution that is approximately Normal, it is considered a *central limit theorem*.

EXERCISES

5.1 Refer to Example 5.1. Find the bounds for the probability that the month's production will be

(a) at most 300 bales;

(b) at least 180 bales.

5.2 The mean lifetime of a certain electrical device is 4 years. Find the lower bounds for the probability that a randomly selected device from a consignment of such devices will not exceed 20 years.

Download free eBooks at bookboon.com

5.3 The amount of savings in a certain Rural Bank is ten million cedis. Suppose the probability that an amount of at most two hundred thousand cedis drawn by a customer selected at random is 0.8. What can we say about the number of customers.

5.4 Refer to Example 5.2. Find the bounds for the probability that the number of bulbs switched on in the area in that evening is different from its expected value in absolute terms by

(a) at least 80 bulbs;

(b) at most 200 bulbs;

(c) not more than 150.

5.5 The distribution of a random variable X is given by the following table:

x_i	1	2	3	4	5	6
$p(x_i)$	0.05	0.10	0.25	0.30	0.20	0.10

(a) Find $P(|X - E(X)| < 2)$

(b) Use Chebyshev's inequality to estimate the lower bounds of the probability in (a) above. *Hint:* $E(X) = 3.8$ and $\text{Var}(X) = 1.77$.

5.6 Suppose that a random variable has mean 35 and standard deviation 5. Use the Chebyshev's inequality to estimate the probability that an outcome will lie between
(a) 24 and 46 (b) 18 and 52 (c) 31 and 39

5.7 A Pharmaceutical Company supplies boxes of paracetamol tablets to a particular hospital. Owing to the packaging procedure, not every box contains exactly 150 tablets. It has been found out that on the average the number of tablets in a box is indeed 150 and the standard deviation is 4. If the company supplies the hospital $1,000$ boxes, estimate the number of boxes having between 140 and 160 tablets.

5.8 Suppose that a random variable has mean 80 and standard deviation 6. Use the Chebyshev's inequality to find the value of ϵ for which the probability that the outcome lies between $80 - \epsilon$ and $80 + \epsilon$ is at least $\frac{5}{12}$.

5.9 Suppose that a random variable has mean 25 and standard deviation 0.67. Use the Chebyshev's inequality to find the value of ϵ for which the probability that the outcome lies between $25 - \epsilon$ and $25 + \epsilon$ is at most $\frac{10}{13}$.

Download free eBooks at bookboon.com

5.10 Refer to Example 5.4. Find the bounds for the following probabilities:

(a) $P(|X - \mu| \leq \sigma)$
(b) $P(|X - \mu| \geq 3.3\sigma)$
(c) $P(|X - \mu| \leq 2.5\sigma)$
(d) $P(|X - \mu| \geq 1.65\sigma)$
(e) $P(|X - \mu| \geq 2.2\sigma)$

5.11 Suppose that the number of hours a certain type of light bulb will burn before requiring replacement has a mean of $2,000$ hours and standard deviation of 150 hours. If $1,000$ such bulbs are installed in a new house, estimate the number that will require replacement between $1,200$ and $2,800$ hours from the time of installation.

5.12 Refer to Example 5.7. Find an upper bound for the probability that in 5,000 trials the deviation of the relative frequency of the event $\mathcal{A}$ from its probability will not exceed 0.08.

Deloitte.
Discover the truth at www.deloitte.ca/careers
© Deloitte & Touche LLP and affiliated entities.
Click on the ad to read more

5.13 A manufacturing company produces ground measuring poles with a mean length of 2 metres and a standard deviation of 1.5 metre per pole. Assume that the pole lengths, X, are statistically independent. A container of 100 poles is sent to a customer for distribution. Find the probability that the total length of all poles in the container is

(a) not more than 195 metres
(b) at least 220 metres
(c) between 180 and 210

5.14 Refer to Exercise 5.13. How many observations should be included in the sample if we wish $\overline{X}$, the mean length of poles, to be within 0.1 metres of μ with probability 0.95.

5.15 The final scores of the students in Diploma class over a period of four years have a mean 60 and a variance 64. In a particular year there were 100 students with a mean score of 58. Calculate the probability that the sample mean, $\overline{X}$ is at most 58.

5.16 A bolt manufacturer knows that, on average, 5% of his production is defective. He gives a guarantee on a shipment of $10,000$ bolts, by promising to refund the money if more than K bolts are defective. How small can the manufacturer choose K and still be assured that he needs not give a refund of more than 1% of his shipment?

5.17 $\overline{X}_n$ is to be used as an approximation to μ. If σ is known to be 2, calculate the sample size n that will be required to ensure that the error in this approximation will be less than 0.05 with probability of at least 0.90.

5.18 Suppose that the number of students that enrol in the Basic Statistics course in the Faculty of Social Studies at the University of Ghana is a Poisson random variable with mean 500. The Co-ordinator of the course has decided that if the number enrolling is 350 or more he will split the group into two, otherwise they will all be in the same class. What is the probability that the class will be split?

5.19 The mean and standard deviation of the ages of statistics students of the University of Ghana are 20 years and 5.8 years respectively. What is the probability that a random sample of 50 students will have a mean age of between 18 and 23 years?

5.20 The mean age and the standard deviation of Statistics students are 18 years and 1.8 years respectively. What is the probability that a random sample of 50 students will have a mean of 16 and 20 years?

5.21 Use the Chebyshev's inequality to determine how large n should be in order that you are at least 95 per cent certain that the sample mean $\overline{X}_n$ is within $0.9\,\sigma$ of the population mean μ?

5.22 Use the Chebyshev's inequality to determine how large n should be in order that you are at most 95 per cent certain that $\overline{X}_n$ is different from μ by at least $0.2\,\sigma$?

Download free eBooks at bookboon.com

5.23 Let $X_1, X_2, \cdots, \overline{X_n}$ be independent, identically distributed random variables each having p.d.f.

$$f(x) = \begin{cases} \frac{1}{2}e^{-\frac{x}{2}}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

Let $S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Use the Chebyshev's inequality to estimate the minimum possible value of n, given that

$$P\{|S_n - E(S_n)| \ge 0.25\} \le 0.75$$

For this value of n, calculate the upper bound to the given probability obtained by applying Chebyshev's inequality.

5.24 The burning time of a certain type of lamp is an exponential random variable with mean $\mu = 20$ hours and standard deviation 14 hours. What is the probability that 80 of these lamps will provide a total of more than 2000 hours of burning time?

5.25 Twenty numbers are rounded off to the nearest integer and then added. Assume the individual round-off errors are independent and uniformly distributed over $(-\frac{1}{2}, \frac{1}{2})$. Find the probability that the given sum will differ from the sum of the original twenty numbers by 3 or more using
(a) Chebyshev's inequality (b) Central Limit Theorem.

5.26 Suppose a coin is tossed 26 times. Estimate the probability that the number of heads will deviate from 12 by less than 4, using
(a) Chebyshev's inequality (b) Central Limit Theorem.

5.27 Prove Theorems 5.11.

5.28 Prove Theorem 5.12(b).

5.29 A consignment of 1,800 items has been produced by a manufacturing company whose products are 4% defective. Suppose the company's production process is an independent trial process with each item produced having a probability of 0.3 of being defective. Use Chebyshev's inequality to give a bound on the probability that in the batch of 1,800 items the number of defective is between 39 and 69.

5.30 Past experience indicates that wire rods purchased from a certain company have a mean breaking strength of 400 kg and a standard deviation of 20 kg. How many rods should you select so that you would be certain with probability 0.95 that your sample mean would not be in error by more than 3 kg.

Download free eBooks at bookboon.com

Chapter 6

SAMPLING DISTRIBUTIONS I
Basic Concepts

6.1 INTRODUCTION

So far we have gone through the various aspects of probability theory in my earlier books. In the Volume I (Nsowah-Nuamah, 2017), we discussed the basic theorems in probability in where we emphasised conditional probability, Bayes' Theorem and Independence. We considered the Random Variable and it was there that the concept of probability distribution was first introduced. We also discussed the Numerical Characteristics of the random variable and there we encountered the concepts of Mathematical Expectation, Variance, Moments, and Moment Generating Function. In the Volume II (Nsowah-Nuamah, 2018) we considered some special probability distributions. These included the Bernoulli distribution, Binomial distribution, Geometric distribution, Negative binomial distribution, Poisson distribution, Hypergeometric distribution, Multinomial distribution, Uniform distribution, Exponential distribution, Gamma distribution, Beta distribution and Normal distribution.

Download free eBooks at bookboon.com
Click on the ad to read more

In Chapters 1 to 4 of this book, we extended the concept of probability distributions to bivariate distributions. In Chapter 5 we discussed the basic Inequalities in Statistics (Markov's and Chebyshev's Inequalities) and the two main Limit Laws (the Central Limit Theorem and the Law of Large Numbers) which have wider and practical applications in Statistics.

We can now conclude the book with some analytical statistics concerned with inferential procedures. The present chapter and the subsequent ones serve as a bridge between the topics discussed in this book and statistical inference.

6.2 STATISTICAL INFERENCE

Statistical inference involves drawing conclusions on a large data set on the basis of information obtained from a small portion of the large collection. Two main concepts that form the basis of statistical inference are *population* and *sample*.

6.2.1 Population

Definition 6.1 POPULATION
A population is a large data set that contains all the members (elements) of interest

Definition 6.2 FINITE POPULATION
A finite population is a population that has a stated or limited number of members

Examples of a finite population include all students in the University of Ghana, all households in Accra, all the mud houses in Kumasi. It is possible (though not always practical) to count the number of elements in a finite population.

Definition 6.3 INFINITE POPULATION
An infinite population is composed of a limitless number of members

The elements in an infinite population cannot be counted completely no matter how long the counting process is carried on. An example of an infinite population is an experiment of tossing a coin to determine whether or not the coin is biased. Theoretically, this experiment could be carried out an infinite number of times.

Download free eBooks at bookboon.com

Philosophically, no truly infinite population of physical objects exists. After all, given unlimited resources and time, we could enumerate even the grains of sand, on a particular seashore. As a practical matter, then, we will use the term infinite population when we are talking about a population that could not be enumerated in a reasonable period of time.

6.2.2 Sample

It is obvious that in an infinite population complete enumeration cannot be secured. Even in a finite but large population it is often costly, time consuming, or even physically impossible to conduct a census (complete enumeration). Any inferences on the population should therefore be based on a part of the population called the sample.

Definition 6.4 SAMPLE
A sample is a part of the population

An example of a sample includes selecting a few pupils from a particular school in order to study the socio-economic background of pupils in that school.

If a sample of n elements is drawn, the sample is said to be of size n.

6.2.3 Parameter and Statistic

From Chapter 7 through 12, we discussed a number of special distributions but did not specify how they were realised in any particular case. Those probabilities that we were calculating depended on certain parameters describing such distributions.

Population Parameter

Definition 6.5 POPULATION PARAMETER
A population parameter is a measure computed from observations of the population

Examples of population parameter (or simply *parameter*) are the values of p in the case of the binomial distribution, λ in the case of the Poisson distribution or μ and σ in the case of the Normal distribution. Other parameters include the median and the range.

Most often the probability distribution of a population is not known precisely although we may have some idea of or at least be able to make some hypothesis concerning its general behaviour. Thus, for example, we may have some reason to suppose that a particular population is binomially distributed. In such a case we would not know the values p and so we might have to estimate.

In short, we have no direct knowledge of the probability distribution and therefore have to approximate it empirically by a frequency distribution. The number of experiments performed for this purpose, called a sample, is necessarily finite.

Download free eBooks at bookboon.com

Definition 6.6 SAMPLE STATISTIC
A sample statistic (or simply statistic) is a measure calculated from the sample

Examples of statistic include the sample mean, the sample proportion, the sample variance, the sample range.

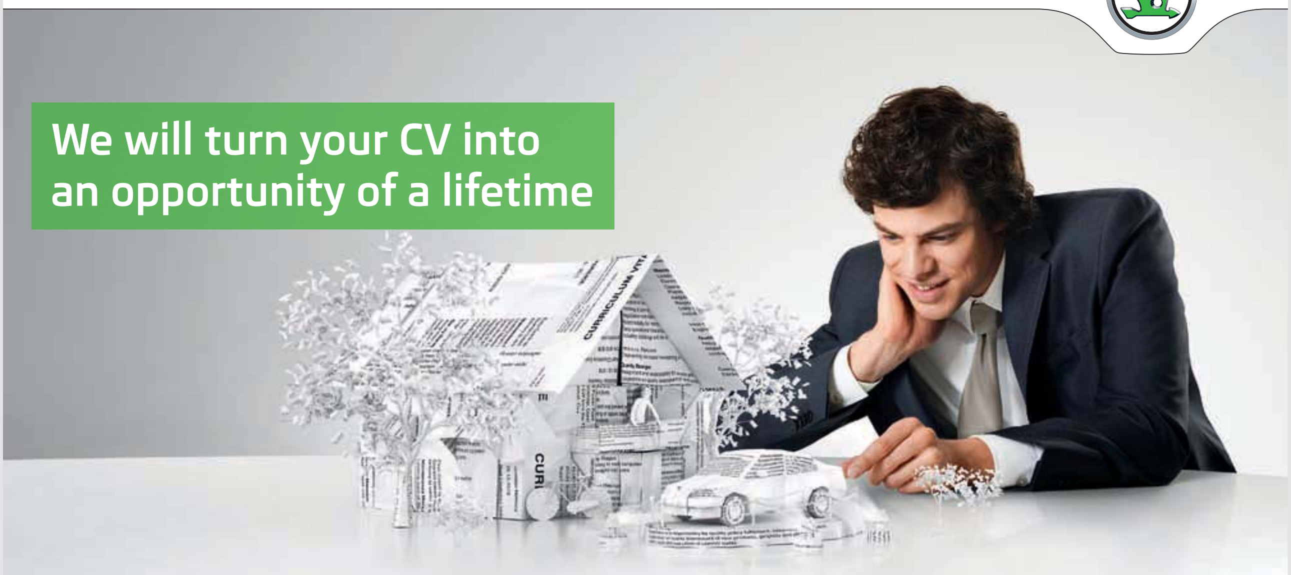
Definition 6.7 SAMPLING
The process of obtaining samples from a population is called sampling

As a rule, small populations, for obvious reasons, are not sampled. Instead, the entire population is examined. A sample that contains all the members in the population is called an *exhaustive sampling*, or a 100 per cent sampling, which of course, are only other names for *census*.

There are basically two types of sampling: *probability sampling* and *non-probability sampling*. In this book we shall consider only probability sampling because it is only for probability sampling that there are sound statistical procedures for drawing conclusions concerning the population of interest based on the sample drawn.

SIMPLY CLEVER

ŠKODA



Do you like cars? Would you like to be a part of a successful brand?
We will appreciate and reward both your enthusiasm and talent.
Send us your CV. You will be surprised where it can take you.

Send us your CV on
www.employerforlife.com

Download free eBooks at bookboon.com

Click on the ad to read more

6.3 PROBABILITY SAMPLING

6.3.1 Definition of Probability Sampling

Probability sampling is the scientific method by which the units of the sample are chosen based on some definite pre-assigned probability. The various types of probability sampling include the cases where:

(a) each member of the population has an equal chance of being selected;

(b) the members of the population have different probabilities of being selected;

(c) the probability of selecting a member is proportional to the sample size.

> **Definition 6.8** PROBABILITY SAMPLE
> A probability sample is a sample drawn from a population in such a way that every member of the population has a known and nonzero probability of being included in the sample

For a probability sample neither the sampler nor the member of the population can decide which member will be included in the sample. The selection is achieved by the operation of chance alone. Synonym for probability sampling is *random sampling*. Random sampling methods include simple random sampling, systematic sampling, and *stratified sampling*. Samples obtained by taking every k^{th} name in a list are called systematic samples. Sometimes a population P is divided into groups, P_1, P_2, $\cdots$, P_r called *strata* and random samples are taken from these strata and combined to get a stratified random sample. This is often done when we want to be sure that we shall have specified numbers of subjects from each stratum. Other sampling techniques include cluster sampling and multistage sampling. For details about sampling techniques see Nuamah (1994).

The simple random sampling is the only one that can be considered a true random sampling and in this book unless otherwise stated, by random sampling we mean simple random sampling.

6.3.2 Simple Random Sampling

Before giving the conventional definition of simple random sampling we shall first consider a mathematical definition.

Let $f(x)$ be the density function of a continuous random variable X for the population being sampled. We are now interested in the values of X taken by the elements of the sample. Suppose that k different samples of size n each are drawn from this population. Let $x_i^{(j)}$ represent the resulting sample values for the jth sample $(j = 1, 2, \cdots, k)$ of size i $\quad (i = 1, 2, \cdots, n)$.

Download free eBooks at bookboon.com

That is,

$$
\begin{array}{lll}
1^{st} & \text{sample:} & x_1^{(1)}, x_2^{(1)}, \cdots, x_i^{(1)}, \cdots, x_n^{(1)} \\
2^{nd} & \text{sample:} & x_1^{(2)}, x_2^{(2)}, \cdots, x_i^{(2)}, \cdots, x_n^{(2)} \\
\vdots & & \vdots \\
j^{th} & \text{sample:} & x_1^{(j)}, x_2^{(j)}, \cdots, x_i^{(j)}, \cdots, x_n^{(j)} \\
\vdots & & \vdots \\
k^{th} & \text{sample:} & x_1^{(k)}, x_2^{(k)}, \cdots, x_i^{(k)}, \cdots, x_n^{(k)}
\end{array}
$$

The values in the i^{th} column may be considered the values of a random variable X_i corresponding to the i^{th} trial of the sample having probability density function $f_i(x_i)$. Thus,

$$
\begin{array}{rcll}
X_1 & = & x_1^{(1)}, x_1^{(2)}, \cdots, x_1^{(j)}, \cdots, x_1^{(k)} & \text{with density function} \quad f_1(x_1) \\
X_2 & = & x_2^{(1)}, x_2^{(2)}, \cdots, x_2^{(j)}, \cdots, x_2^{(k)} & \text{with density function} \quad f_2(x_2) \\
\vdots & & \vdots & \\
X_i & = & x_i^{(1)}, x_i^{(2)}, \cdots, x_i^{(j)}, \cdots, x_i^{(k)} & \text{with density function} \quad f_i(x_i) \\
\vdots & & \vdots & \\
X_n & = & x_n^{(1)}, x_n^{(2)}, \cdots, x_{(n)}^{j}, \cdots, x_n^{(k)} & \text{with density function} \quad f_n(x_n)
\end{array}
$$

The probability density function $f(x_i)$ has to fulfill two conditions in order to describe the process of simple random sampling. These are given in Definition 6.9.

Definition 6.9 SIMPLE RANDOM SAMPLING
(Mathematical Definition)
Simple random sampling is a method of sampling for which

(a) $f(x_1, x_2, ..., x_n) = f_1(x_1) f_2(x_2) ... f_n(x_n)$ and

(b) $f_1(x_1) = f_2(x_2) = ... = f_n(x_n) = f(x)$

where $f(x)$ is the probability density function of the random variable X for the population being sampled

Although the random variable X has been treated as continuous variable, Definition 6.9 applies to both continuous and discrete variables. The definition means that

(a) the successive selection of the element in the population are independent and

(b) the density function of the random variable remains the same from selection to selection as that of the population.

Download free eBooks at bookboon.com

The independence condition can, to a certain extent, be controlled by comparing the empirical distributions for the 1^{st}, 2^{nd}, $\cdots$ elements obtained in a large number of samples. However, it is not at all easy to be sure that the sample is in fact drawn from a population described by the probability density function $f(x)$.

Definition 6.10 SIMPLE RANDOM SAMPLE
(Mathematical Definition)
Suppose $x_1, x_2, ..., x_n$ are values assumed by a corresponding set of random variables $X_1, X_2, ..., X_n$ which are independent and which have the same distribution, then the x_i's are referred to as a simple random sample

Conventionally, we shall define simple random sampling and simple random sample as follows.

Definition 6.11 SIMPLE RANDOM SAMPLING
(Conventional Definition)
Simple random sampling is any technique designed to draw sample members from a population in such a way that each member in the population has an equal chance of being selected

Definition 6.12 SIMPLE RANDOM SAMPLE **(Conventional Definition)**

A simple random sample is a sample drawn in such a way that

(a) every member of the population has the same probability of being included in the sample,

(b) different members of the population are selected independently (that is, selection of one member has no effect on selection of another)

We must emphasise that in actual process of sampling, it is often quite difficult to maintain randomness and conscious effort must be made to ensure that conditions (a) and (b) in Definition 6.12 are fulfilled. There is no general recipe that can be given to ensure that the results from sampling are reliable besides the use of objective procedures. Suppose that there is a *sampling frame* which is the list of all members of the population of interest. Then probably, the best method of simple random sampling is the use of a computer generated table of random numbers.

Random Number Table

A random number table is a table of numbers constructed by a process that ensures that

(a) in any position in the table, each of the numbers 0 through 9 has equal probability of $\frac{1}{10}$ of occurring.

(b) The occurrence of any number in one part of the table is independent of the occurrence of any number in any other part of the table (i.e. knowing the numbers in one part of the table tells us nothing about the numbers in another part of the table).

Table X of Appendix A presents a table of random numbers.

We shall illustrate in the following example how random numbers are used in the selection of a sample.

Example 6.1
Select 15 students from a student population of 250 using a table of random numbers.

Solution

(a) *Assign a number to each member in the population*

Compile the list of the students and assign each of the 250 students a number from 001 to 250. Each student number is called a sampling number.

Download free eBooks at bookboon.com

(b) *Select a page*

If the table of random numbers is more than one page then select a page at random (for example, by writing the page numbers on pieces of paper and picking one at random). Suppose the third page of Table X is selected.

(c) *Pick an arbitrary starting point in the random table*

These random numbers are in groups of five digits. This grouping is arbitrary. Because our population size ($N = 250$) has three digits, we need to pick three-digit numbers from the table.

Suppose we decide to take the first three digits of each five-digit number. To pick the starting point, close your eyes and place your pencil point on the table so that we do not subconsciously look for a starting point that includes a member of interest to us. Assume that your pencil falls in the *fourth* column of numbers in the *eleventh* row.

The number is 46201. The first three digits of this number are 462. This number is thrown out because it is larger than the population size ($N = 250$). There is no student whose number is 462.

(d) *Determine the numbers*

We select numbers for our sample by moving in some (predetermined) direction away from the starting point. We can go up, down or across. Suppose we had decided beforehand to move across the row. Reading the next first three-digit number we get 045 and the student assigned number 045 is the first member of our sample. The next three-digit number going across to the right gives us 425 which we throw out. Moving across we look for the next three digit number which is not larger than the population size of 250. This is 211 so the second person in the sample is the student numbered 211. The next three-digit number is 763 which we throw out. This process of picking numbers continue until 15 students (the sample size n) are selected. The resulting samples are:

045 211 006 194 214 212 029 276 044 165 154 080
076 043 244

Note
If, by chance, the same number occurs more than once, we would ignore it, after it has appeared for the first time.

6.4 SAMPLING WITH AND WITHOUT REPLACEMENT

If we select a member from a population we have the choice of replacing or not replacing the member in the population before a second selection. In the first case a particular member can come up again and again, whereas in the second, it can come up only once. This leads us to Definitions 6.13 and 6.14.

Definition 6.13 RANDOM SAMPLING WITH REPLACEMENT
If n members are selected from a population of N members, sequentially one by one, such that at every stage the probability of selection of any member is $\frac{1}{N}$ irrespective of whether this member has been selected earlier or not, the procedure is known as random sampling with replacement

In general, when sampling is with replacement from a population of N members, the number of possible samples of size n is equal to N^n.

Sampling with replacement is often associated with an infinite population. Thus the expressions "sample from an infinite population", "sampling with replacement", "samples are independent" and "simple random sample" are all equivalent.

A finite population which is sampled with replacement can theoretically be considered infinite since samples of any size can be drawn without exhausting the population. For more practical purposes, sampling from a finite population which is very large can be considered as sampling from an infinite population.

Definition 6.14 RANDOM SAMPLING WITHOUT REPLACEMENT
If at each stage of selection of n members from a population of N members sequentially one by one, a member is selected with equal probability only from among members not selected earlier, sampling is said to be random sampling without replacement

Sampling without replacement is often associated with finite population and the expressions "sample from a finite population", "sampling without replacement" and "samples are not independent" are all equivalent.

When drawing samples of size n from a finite population of size N without replacement, and ignoring the order in which the sample values are drawn, the number of possible samples is given by the combination of N objects taken n at a time. From Chapter 2 this is given by

Download free eBooks at bookboon.com

$$\binom{N}{n} = \frac{N!}{n!\,(N-n)!}$$

The probability for each subset of n of the N objects of the finite population is

$$\frac{1}{\binom{N}{n}}$$

We can also determine the joint probability distributions of the random variables from a random sample of size n from a finite population by

$$f(x_1, x_2, \cdots, x_n) = \frac{1}{N(N-1)\cdots(N-n+1)}$$

To generalise: when a simple random sample of size n is drawn from a finite population with replacement, or from an infinite population with or without replacement, we have n identically distributed random variables $X_1, X_2, \cdots, X_n$ all possessing the same distribution as the parent population. Also, if the population is finite but large then even when sampling is without replacement, the sample observations, x_i, can in practice be treated as if they were identical and independent.

EXERCISES

6.1 Write short notes on the following:

(a) Simple random sampling
(b) Systematic sampling
(c) Stratified sampling
(d) Cluster sampling
(e) multistage sampling

6.2 The management of an establishment wants to estimate the age of twenty workers from the workers population of 300. Use random numbers to draw this sample.

Download free eBooks at bookboon.com

Chapter 7

SAMPLING DISTRIBUTIONS II
Sampling Distribution of Statistics

7.1 INTRODUCTION

7.1.1 Definition of Statistic

Having determined in Chapter 6 what a random sample is, we can now examine the distributions of statistics calculated from random samples.

Though we have already defined what "statistic" is, we shall now present its mathematical definition.

Definition 7.1 STATISTIC
Let $X_1, X_2, ..., X_n$ be a random sample from a random variable and let $x_1, x_2, ..., x_n$ be the values assumed by the sample. Then the real-valued function

$$Y = G(X_1, X_2, ..., X_n)$$

is a statistic which assumes the value $y = G(x_1, x_2, ..., x_n)$

Download free eBooks at bookboon.com
Click on the ad to read more

Examples of a statistic include

(a) the sample mean: $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$

(b) the sample variance[11]: $\hat{s}^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})$

(c) Minimum of the sample: $K = \min(X_1, X_2, ..., X_n)$

(d) Maximum of the sample: $M = \max(X_1, X_2, ..., X_n)$

(e) the sample range: $R = M - K$

That is, a statistic is a formula for obtaining values of the function. However, sometimes the term statistic is used to refer to the values themselves. Thus we may speak of the statistic $y = G(x_1, x_2, ..., x_n)$ when we really should say that y is the value of the statistic $Y = G(X_1, X_2, ..., X_n)$.

7.1.2 Sample Observations as Random Variables

Earlier we have referred to sampling as a process of selecting a sample from a population or the range of a random variable. An observation made from the range of a random variable is, again a random variable. This is because an observed value can assume any one of all possible values of a random variable.

Suppose $X_1, X_2, ..., X_n$ is a simple random sample and $x_1, x_2, ..., x_n$ are the actual observations of a specific sample that has been drawn. Thus, the term "random sample" refers to the observations thought of as random variables but not to a particular set of actual realisations. However, since the X_i are identically distributed, a specific collection of sample observations can be thought of as the particular values of a random variable X. In connection with this, for a random sample of size n, we may treat the x_i as realisation of X : $X = x_1, X = x_2, ..., X = x_n$. Consequently, if the sample is random then sample observations thought of as independent random variables and sample observations thought of as a particular set of sample values are equivalent and can be used interchangeably.

7.1.3 Definition and Construction of Sampling Distribution of Statistics

We can consider all possible samples of size n which can be drawn from the population, and for each sample we compute a statistic of interest. The idea that sample observations are random variables leads us to the conclusion that any quantity computed from sample observations must also be a random

[11] The reason for dividing by $n-1$ rather than the seemingly more logical choice, n, will be explained in Section 7.5.

Download free eBooks at bookboon.com

variable. As a random variable, the computed quantity or statistic has a probability distribution of its own called probability distribution of the statistic or sampling distribution of the statistic which may be established directly from the parent population distribution.

Definition 7.2 SAMPLING DISTRIBUTION OF A STATISTIC

The distribution of all possible values that can be assumed by some statistic, computed from samples of the same size randomly drawn from the same population, is called the sampling distribution of that statistic

For a given sampling distribution we are interested in knowing three things, namely, the mean, the variance and the form of the distribution.

Even though in practice we would not at all attempt to empirically construct the sampling distribution of a statistic, we can conceptualise the manner in which it can be done when sampling is from a discrete, finite population of size N. The procedure is as follows:

(a) Randomly select all possible samples of size n from the finite population of size N.

(b) Compute the statistic of interest for each sample.

(c) List the different distinct observed values of the statistic together with the corresponding frequency of occurrence of each distinct observed value of the statistic.

(d) Compute the relative frequency (probabilities) of the occurrence of the observed statistic.

For infinite or large finite populations, one could approximate sampling distribution of a statistic by drawing large number of independent simple random samples and by proceeding in the manner just described above. In fact, the actual construction of a sampling distribution according to the steps given above is a tedious task. Fortunately, there are theorems that simplify things for us.

In the subsequent sections we shall introduce sampling distributions for the most frequently encountered statistics: the mean, proportion and variance. We shall now introduce an example that will be used most often in this chapter.

Example 7.1
Suppose a population consists of four numbers: $4, 5, 7, 8$.
Find

(a) the population mean;

(b) the population variance and population standard deviation.

Download free eBooks at bookboon.com

Solution

(a) The population mean μ is

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{4+5+7+8}{4} = \frac{24}{4} = 6$$

(b) The population variance σ^2 is

$$\begin{aligned}
\sigma^2 &= \frac{\sum_{i=1}^{N}(X_i-\mu)^2}{N} \\
&= \frac{(4-6)^2+(5-6)^2+(7-6)^2+(8-6)^2}{4} \\
&= \frac{4+1+1+4}{4} \\
&= 2.5
\end{aligned}$$

The population standard deviation is

$$\sigma = \sqrt{2.5} = 1.5811$$

Click on the ad to read more

7.2 SAMPLING DISTRIBUTION OF MEANS

7.2.1 Definition of Sampling Distribution of Means

Let $x_1, x_2, ..., x_n$ be the values assumed by a corresponding set of random variables $X_1, X_2, ..., X_n$. We refer to the x_i's as a random sample of size n. We are accustomed to thinking of the sample mean $\overline{X}$ as a specific number calculated from a specific sample. If we consider only one sample, this is correct. If we consider more than one sample, we need to think of $\overline{X}$ differently. The value of $\overline{X}$ usually changes as the sample changes. When we select different samples from a population, we generally get different values for $\overline{X}$. The mean $\overline{X}$ of such a sample is a value assumed by the random variable

$$\overline{X} = \frac{X_1 + X_2 + ... + X_n}{n}$$

The sample mean $\overline{X}$ therefore is a random variable because $X_1, X_2, ..., X_n$ corresponding to n trials of the sample values $x_1, x_2, ..., x_n$ are random variables. After a particular random sample has been taken, $\overline{X}$ will be a number, but before it has been drawn it will be a random variable capable of assuming any value that the original variable X can assume.

Definition 7.3 SAMPLING DISTRIBUTION OF MEANS
The probability distribution of the sample statistic $\overline{X}$, is called the sampling distribution for the sample mean or the sampling distribution of means

7.2.2 Sampling Distribution of Means from Infinite Population

In the previous section we conceptualised the manner in which the sampling distribution of a statistic can be empirically constructed. In the following example we shall illustrate the procedure with the sampling distribution of means.

Example 7.2
Refer to Example 7.1.

(a) List all possible samples of size two that can be drawn from the population with replacement.

(b) Find

  (i) the mean of the distribution of the means;
  (ii) the variance and standard deviation of the distribution of the means.

Solution

(a) $N = 4, \quad n = 2$
There are $N^n = 4^2 = 16$ possible samples of size two that can be drawn from a population of size four (where sampling is with replacement).

Download free eBooks at bookboon.com

These are

$$\begin{array}{cccccccc} (4,4) & (4,5) & (4,7) & (4,8) & (5,4) & (5,5) & (5,7) & (5,8) \\ (7,4) & (7,5) & (7,7) & (7,8) & (8,4) & (8,5) & (8,7) & (8,8) \end{array}$$

Note
Here the notation (a,b) is an ordered pair. For example, $(4,5)$ denotes "first a 4 and then a 5" and is different from $(5,4)$ which denotes "first a 5 and then a 4."

(b) (i) The corresponding sample means $\overline{X}_i$ are

$$\begin{array}{cccccccc} 4.0 & 4.5 & 5.5 & 6.0 & 4.5 & 5.0 & 6.0 & 6.5 \\ 5.5 & 6.0 & 7.0 & 7.5 & 6.0 & 6.5 & 7.5 & 8.0 \end{array}$$

The sample mean and their associated probabilities (relative frequencies) are as follows:

Possible values of $\overline{X}_i$	Frequency	Relative Frequency (Probability)
4.0	1	0.0625
4.5	2	0.1250
5.0	1	0.0625
5.5	2	0.1250
6.0	4	0.2500
6.5	2	0.1250
7.0	1	0.0625
7.5	2	0.1250
8.0	1	0.0625
Total	16	1.0000

For the distribution of the sample means $\overline{X}_i$, the expected value is

$$\begin{aligned} \mu_{\overline{X}} &= E(\overline{X}_i) = \sum \overline{x}_i \, p(\overline{x}_i) \\ &= 4(0.0625) + 4.5(0.1250) + 5(0.0625) + 5.5(0.1250) \\ &\quad +6(0.2500) + 6.5(0.1250) + 7(0.0625)0 + 7.5(0.1250) \\ &\quad +8(0.0625) \\ &= 6 \end{aligned}$$

which is identical with the population mean obtained in Example 7.1(a). Thus the mean of the sample means is equal to the population mean.

Note
Without calculating the probabilities we could have obtained the mean of the sampling distribution of means from the raw data (the sample means themselves) as:

$$\begin{aligned} \mu_{\overline{X}} &= \frac{\sum \overline{X}}{N^2} = \frac{4.0 + 4.5 + 5.0 + \cdots + 6.5 + 7.0 + 7.5 + 8.0}{4^2} \\ &= \frac{96}{16} \\ &= 6 \end{aligned}$$

Download free eBooks at bookboon.com

(ii) The variance of the distribution of sample means is:

$$\begin{aligned}
\sigma^2_{\overline{X}} &= \frac{\sum(X_i - \mu_X)^2}{N^n} \\
&= \frac{(4-6)^2 + (4.5-6)^2 + \cdots + (8-6)^2}{16} \\
&= \frac{4 + 2.25 + 1 + 0.25 + 0 + 0.25 + 1 + 2.25 + 4}{16} \\
&= 1.25
\end{aligned}$$

The standard deviation of the distribution of the means is

$$\sigma_{\overline{X}} = \sqrt{1.25} = 1.118$$

Note
The variance of the sampling distribution of means is not equal to the population variance.

As has been pointed out earlier, it is burdensome to construct the sampling distribution of means in the way it has been done. In practice, we shall employ the following two theorems to obtain the mean and variance of the sample mean.

Download free eBooks at bookboon.com
Click on the ad to read more

Theorem 7.1
Let X be a random variable with expectation $E(X) = \mu$ and variance $\text{Var}(X) = \sigma^2$. Let $\overline{X}$ be the sample mean of a random sample of size n. Then the mean of the sampling distribution of means is given by

$$\mu_{\overline{X}} = E(\overline{X}) = \mu$$

The theorem above states that the expected value of the sample mean is the population mean.

Proof
Since the sample mean is defined as

$$\begin{aligned}
\overline{X} &= \frac{X_1 + X_2 + \cdots + X_n}{n} \\
E(\overline{X}) &= \frac{1}{n}\Big[E(X_1) + E(X_2) + \cdots + E(X_n)\Big] \\
&= \frac{1}{n}(n\mu) \\
&= \mu
\end{aligned}$$

Theorem 7.2
If a population is infinite or if sampling is with replacement, then the variance of the sampling distribution of means, denoted by $\sigma^2_{\overline{X}}$, is given by

$$\text{Var}(\overline{X}) = \sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$$

That is, the variance of the sampling distribution is equal to the population variance divided by the size of the sample used to obtain the sampling distribution.

Proof

$$\begin{aligned}
\text{Var}(\overline{X}) &= \text{Var}\left(\frac{X_1}{n} + \frac{X_2}{n} + ... + \frac{X_n}{n}\right) \\
&= \frac{1}{n^2}\Big[\text{Var}(X_1) + \text{Var}(X_2) + ... + \text{Var}(X_n)\Big] \\
&= \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}
\end{aligned}$$

Note
Theorems 7.1 and 7.2 do not assume Normality of the "parent" population.

Definition 7.4 STANDARD ERROR OF MEAN
The positive square root of the variance of the mean is referred to as the standard error of the sample mean

Download free eBooks at bookboon.com

The standard error of the mean measures chance variations of sample mean from sample to sample. Denoting the standard error by $\sigma_{\overline{X}}$, it is given by

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

This formula shows that the standard error of the mean decreases when n, the *sample size*, is increased. This means that when n is sufficiently large and we actually have more information, sample means can be expected to be closer to μ, the quantity which $\overline{X}$ is usually supposed to estimate.

Example 7.3
Refer to Example 7.1. Employ Theorems 7.1 and 7.2 to obtain the variance of the sampling distribution of means if a sample of size two is drawn with replacement from the population.

Solution
From Example 7.2, $\quad N = 4, \quad \sigma^2 = 2.5.$
Also, the sample size $n = 2$. Hence

$$\text{Var}(\overline{X}) = \frac{\sigma^2}{n} = \frac{2.5}{2} = 1.25$$

Although Theorems 7.1 and 7.2 give us some characteristics of the sampling distribution, it does not permit us to calculate probabilities because we do not know the form of the sampling distributions. To be able to do this we need to use the Central Limit Theorem, as it can be seen later.

7.2.3 Sampling Distributions of Means from Finite Population

The discussions so far have been based on the assumption that samples are either drawn with replacement or are drawn from infinite populations. In general, we do not sample with replacement, and in most practical situations it is necessary to sample from a finite population; hence, we need to become familiar with the behaviour of the sampling distribution of the sample mean when sampling is drawn without replacement or from finite populations.

Example 7.4
Refer to Example 7.1

(a) List all possible samples of size two that can be drawn from the population without replacement.

(b) Find

  (i) the mean of the distribution of the means;

  (ii) the variance and standard deviation of the distribution of the means.

Download free eBooks at bookboon.com

Solution
$N = 4, \quad n = 2$

(a) There are $\binom{4}{2} = 6$ samples of size two which can be drawn without replacement, namely,

$$(4,5) \quad (4,7) \quad (4,8) \quad (5,7) \quad (5,8) \quad (7,8)$$

Note
The sample $(4,5)$ for example, is considered the same as $(5,4)$.

(b) (i) The means of the corresponding samples, $\overline{X}_i$ in (a) are

$$4.5 \quad 5.5 \quad 6.0 \quad 6.0 \quad 6.5 \quad 7.5$$

Therefore the mean of the sampling distribution of means is

$$\begin{aligned}\mu_{\overline{X}} &= \frac{\sum_{i=1}^{n} \overline{X}}{\binom{4}{2}} \\ &= \frac{4.5+5.5+6.0+6.0+6.5+7.5}{6} \\ &= 6\end{aligned}$$

Click on the ad to read more
Download free eBooks at bookboon.com

Note
Once again the mean of the sampling distribution of means is equal to the population mean.

(ii) The variance of this sampling distribution is

$$\begin{aligned}
\sigma^2_{\overline{X}} &= \frac{\sum(X_i - \mu_{\overline{X}})}{\binom{N}{n}} \\
&= \frac{(4.5-6)^2 + (5.5-6)^2 + \cdots + (7.5-6)^2}{\binom{4}{2}} \\
&= \frac{2.25 + 0.25 + 0 + 0 + 0.25 + 2.25}{6} \\
&= \frac{5}{6} = 0.83333 \\
\sigma_{\overline{X}} &= \sqrt{\frac{5}{6}} = 0.9129
\end{aligned}$$

Note
As in the case of sampling with replacement, the variance of the sampling distribution is not equal to the population variance. Moreover, it is not equal to the population variance divided by the sample size ($\sigma^2_{\overline{X}} = \frac{5}{6} \neq \frac{2.5}{4}$). The formula for obtaining the variance of the sampling distribution of means in the case of sampling without replacement is given in Theorem 7.3.

Theorem 7.3
If $\overline{X}$ is the mean of a random sample of size n from a finite population of size N (or if sampling is without replacement), whose mean is μ and variance is σ^2, then

$$\begin{aligned}
E(\overline{X}) &= \mu \\
\mathrm{Var}(\overline{X}) &= \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}
\end{aligned}$$

The formula for obtaining $\mathrm{Var}(\overline{X})$ in Theorem 7.2, which applies to values assumed by independent random variables (or sampling with replacement from a finite population) and the one in Theorem 7.3, which applies to sampling without replacement from a finite population, differ by the factor $\frac{N-n}{N-1}$. If N, the size of the population, is larger compared to n, the size of the sample, the difference between the two formulas becomes negligible. Indeed, the formula in Theorem 7.2 is frequently used as an approximation for the variance of the distribution of $\overline{X}$ for samples obtained without replacement from sufficiently large finite populations.

Download free eBooks at bookboon.com

Example 7.5
For the data of Example 7.1 if a sample of size two, is drawn without replacement, calculate the variance and hence the standard deviation of the sampling distribution of means.

Solution
Since sampling is without replacement, we use the formula in Theorem 7.3

$$\begin{aligned}
\sigma^2 &= 2.5, \quad N = 4, \quad n = 2 \\
\text{Var}(\overline{X}) &= \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1} \\
&= \frac{2.5}{2} \cdot \frac{4-2}{4-1} \\
&= 0.83333
\end{aligned}$$

Hence the standard error of the mean is

$$\sigma_{\overline{X}} = \sqrt{0.83333} = 0.9129$$

Note
These results are equal to the ones obtained in Example 7.4

7.2.4 Sampling Distribution of Means from Normal Distribution

In sampling distribution of means we are assured of at least an approximately Normal distribution under three conditions. These are when sampling is from

(i) a normally distributed population;

(ii) a non-normally distributed population but the sample size is large;

(iii) a population whose distribution is unknown but the sample size is large.

Theorem 7.4
If the population from which random samples are taken is normally distributed with mean μ and variance σ^2 then the sample mean $\overline{X}$ is normally distributed with mean μ and variance $\dfrac{\sigma^2}{n}$

Proof
From Theorem 6.22 of Volume I it follows that

$$M_{\overline{X}}(t) = M_{\frac{1}{n}(X_1+X_2...+X_n)}(t) = M_{X_1+...+X_n}\left(\frac{t}{n}\right)$$

Since the sampling is random (that is, simple random), the variables X_1, $X_2, .., X_n$ are independent, and therefore Theorem 6.28 of Volume I may be applied to give

$$M_X(t) = M_{X_1}\left(\frac{t}{n}\right) M_{X_2}\left(\frac{t}{n}\right) ...M_{X\,n}\left(\frac{t}{n}\right) \qquad (i)$$

Download free eBooks at bookboon.com

From Definition 7.2 all the random variables $X_1, X_2, ..., X_n$ have the same probability density function, namely that of X, and hence the same moment generating function. Consequently, all the moment generating functions on the right hand side of (i) are the same function, namely the moment generating function of the random variable X. Thus

$$M_X(t) = \left[M_X\left(\frac{t}{n}\right)\right]^n$$

Since the moment generating function of the Normal distribution $N(\mu,\, \sigma^2)$ is given by (see Theorem 11.7 of Volume I)

$$M_X(t) = e^{\mu\, t + \frac{1}{2}\, t^2 \sigma^2},$$

replacing t by $\dfrac{t}{n}$ in the formula will yield

$$\begin{aligned} M_X(t) &= e^{\left[\mu\,\frac{t}{n} + \frac{1}{2}\left(\frac{t^2}{n^2}\right)\sigma^2\right]^n} \\ &= e^{\mu\, t + \frac{1}{2}\, t^2\, \frac{\sigma^2}{n}} \end{aligned}$$

Therefore, by the Uniqueness Property of m.g.f. (Theorem 6.26 of Volume I), $\overline{X}$ has the Normal distribution $N(\mu,\, \dfrac{\sigma^2}{n})$.

Click on the ad to read more

Example 7.6
A random sample of 20 students was taken from a normally distributed population with mean age of 17 years and variance of 5.

(a) Find the

  (i) expectation of the sample mean;

  (ii) variance and hence the standard deviation of the sample mean.

(b) What is the probability that the mean of this sample will fall between 16 and 23?

Solution

(a) $n = 20, \quad \mu = 17, \quad \sigma^2 = 5$

  (i) $E(\overline{X}) = \mu = 17$

  (ii) $\text{Var}(\overline{X}) = \dfrac{\sigma^2}{n} = \dfrac{5}{20} = 0.25$

  $$\sigma_{\overline{X}} = \sqrt{\frac{\sigma^2}{n}} = \sqrt{0.025} = 0.5$$

(b) Though $n = 20$ is small, we still apply the Normal distribution because the population from which the sample is drawn is normally distributed.

$$\begin{aligned}
P(16 < \overline{X} < 23) &= P\left(\frac{16-17}{\sqrt{0.5}} < \frac{\overline{X}-17}{\sqrt{0.5}} < \frac{23-17}{\sqrt{0.5}}\right) \\
&= P(-2 < Z < 12) \\
&= \Phi(12) - \Phi(-2) \\
&= 1 - \{1 - \Phi(2)\} \\
&= \Phi(2) \\
&= 0.9872
\end{aligned}$$

7.2.5 Sampling Distribution of Means from Non-Normal Distribution

We are often faced with the problem of sampling from non-normally distributed populations or from populations whose distributions are not known. Under these two conditions, we shall have to take large samples since when the sample size is sufficiently large, by virtue of the Central Limit Theorem (Theorem 5.16), the means of random samples from any distribution will tend to be normally distributed with a mean μ and variance $\dfrac{\sigma^2}{n}$. This permits the use of the Normal distribution approximation as though sampling is from normally distributed populations. Thus inference procedures based on the sample mean can often use the Normal distribution. But we must be careful not to imput normality to the original observations.

Download free eBooks at bookboon.com

Example 7.7
A random sample of size 100 is taken from a population with mean 60 and variance 300.

(a) Find the

(i) expectation of the sample mean,

(ii) variance and hence the standard error of the sample mean.

(b) What is the probability that the sample mean will be less than 56?

Solution

$$n = 100 \quad \mu = 60 \quad \text{Var}(X) = 300$$

(a) (i) $E(X) = \mu = 60$

(ii) $\text{Var}(\overline{X}) = \dfrac{300}{100} = 3$

and hence $\sigma = \sqrt{3} = 1.732$

(b) The sampling distribution of the mean is not known but the sample size is large so by the Central Limit Theorem we approximate it by the Normal distribution.

$$\begin{aligned} P(\overline{X}_n < 56) &= P\left(\frac{\overline{X}_n - 60}{\sqrt{3}} < \frac{56-60}{\sqrt{3}}\right) \\ &= P(Z_n < -2.31) \\ &= \Phi(-2.31) \\ &= 1 - \Phi(2.31) \\ &= 1 - 0.9896 \\ &= 0.0104 \end{aligned}$$

7.3 SAMPLING DISTRIBUTION OF PROPORTIONS

7.3.1 Construction of Sampling Distribution of Proportions

In the previous section we discussed the sampling distribution of a statistic computed from measured variables, namely, the sample mean. In this section we shall deal with the sampling distribution of statistics that result from counts or frequency data, namely, a sample proportion.

Let a population proportion, p, be defined as $p = \dfrac{k}{N}$, where k is the number of members of the population that possess a certain trait and N is the total number of members of the population. Let us also designate the sample proportion by the symbol $\hat{p}$ (read p hat) and is defined as $\dfrac{x}{n}$, where x is the number of elements in the sample that possess a certain trait and n is the sample size. The sample distribution of the sample proportion could be constructed experimentally in exactly the same manner as was suggested in the case of the sample mean. From the population, which we assume to be finite, we would

Download free eBooks at bookboon.com

(a) take all possible samples of a given size,

(b) for each sample compute the sample proportion $\hat{p}$,

(c) prepare a frequency distribution of $\hat{p}$ by listing the different values of $\hat{p}$ along with their frequencies of occurrence.

This frequency distribution (as well as the corresponding relative frequency distribution) would constitute the sampling distribution of $\hat{p}$.

7.3.2 Sampling Distribution of Proportions With Replacement

A sample proportion $\hat{p}$ may be considered as a proportion of success which is obtained by dividing the number of success by the sample size n. Hence if a random sample of size n is obtained with replacement, then the sampling distribution of $\hat{p}$ obeys the binomial probability law.

Theorem 7.5
If sampling is with replacement, then

(a) $E(\hat{p}) = p$

(b) $\text{Var}(\hat{p}) = \dfrac{p(1-p)}{n}$

where $\hat{p}$ is the sample proportion

Click on the ad to read more

Example 7.8
A consignment of bulbs has 20 percent defective. If a random sample of 250 is drawn with replacement from this consignment. What is the

(a) expectation of the sample proportion?

(b) variance and hence the standard error of the sample proportion?

Solution
$p = 0.2, \quad n = 250$

(a) $E(\hat{p}) = p = 0.2$

(b)

$$\begin{aligned} \text{Var}(\hat{p}) &= \frac{p(1-p)}{n} \\ &= \frac{0.2(1-0.2)}{250} = 0.00064 \\ \sigma_{\hat{p}} &= \sqrt{0.00064} = 0.0253 \end{aligned}$$

7.3.3 Sampling Distribution of Proportions Without Replacement

If the sampling is without replacement, then the sampling distribution of $\hat{p}$ obeys the hypergeometric probability law.

Theorem 7.6
If sampling is without replacement, then

(a) $E(\hat{p}) = p$

(b) $\text{Var}(\hat{p}) = \dfrac{p(1-p)}{n} \cdot \dfrac{N-n}{N-1}$

where $\hat{p}$ is the sample proportion

That is, $E(\hat{p})$ is still identical with the population proportion p but the variance is adjusted by a finite population correction factor $\dfrac{N-n}{N-1}$.

Example 7.9
In a class of 50 students, twenty are females and thirty are males. A random sample of twelve students are drawn from this class without replacement. Determine the variance and hence the standard error of the proportion of male students in the sample of twelve.

Download free eBooks at bookboon.com

Solution
Let k represent the number of female students in the class. Hence

$$N = 50, \quad k = 20, \quad n = 12$$

The population proportion of females in the class is therefore

$$p = \frac{k}{N} = \frac{20}{50} = 0.4$$

Since this is sampling without replacement, we have

$$\begin{aligned}
\sigma^2_{\hat{p}} = \text{Var}(\hat{p}) &= \frac{p(1-p)}{n} \cdot \frac{N-n}{N-1} \\
&= \frac{0.4(1-0.4)}{12} \cdot \frac{50-20}{50-1} \\
&= 0.012245 \\
\sigma_{\hat{p}} &= \sqrt{0.012245} = 0.1107
\end{aligned}$$

7.3.4 Sampling Distribution of Proportions from Non-Normal Population

The distribution of $\hat{p}$ for a random sample taken with replacement approaches the Normal distribution when n becomes infinite. For a random sampling without replacement, when the sample size is large, the distribution of sample proportions is approximately normally distributed by virtue of the Central Limit Theorem.

Theorem 7.7
The variable

$$\frac{\hat{p} - p}{\sqrt{\sigma^2_{\hat{p}}}}$$

approaches the Standard Normal distribution when n becomes infinite, where $\hat{p}$ is a sample proportion that occurs in a sample of size n

The question that now arises is how large does the sample size have to be for the use of the Normal approximation to be valid? A widely used criterion is that both np and $n(1-p)$ must be greater than 5.

Example 7.10
Refer to Example 7.8; what is the probability that 22% will be defective?

Solution
$n = 250, \quad p = 0.2$
A specific value of $\hat{p}$, denoted by p_0 is $p_0 = 0.22$. We are required to calculate $P(\hat{p} \le p_0)$.

$$P(\hat{p} \le 0.22) = P\left(Z \le \frac{0.22 - 0.2}{\sqrt{\dfrac{0.2(1-0.2)}{250}}}\right)$$

Download free eBooks at bookboon.com

$$
\begin{aligned}
&= P\left(Z \le \frac{0.02}{0.0253}\right) \\
&= P(Z \le 0.7906) = \Phi(0.7906) \\
&= 0.61141
\end{aligned}
$$

The Normal approximation may be improved by the continuity correction factor, a device that makes an adjustment for the fact that a discrete distribution is being approximated by a continuous distribution. The correction factor is more important if n is small. With the continuity correction factor,

$$
\begin{aligned}
P(\hat{p} \le p_0) &= P\left(Z \le \frac{p_0 + \dfrac{0.5}{n} - p}{\sqrt{\dfrac{p(1-p)}{n}}}\right), \qquad x < np \\
&= \Phi\left(\frac{p_0 + \dfrac{0.5}{n} - p}{\sqrt{\dfrac{p(1-p)}{n}}}\right)
\end{aligned}
$$

or

$$
P(\hat{p} \le p_0) = P\left(Z \le \frac{p_0 - \dfrac{0.5}{n} - p}{\sqrt{\dfrac{p(1-p)}{n}}}\right), \qquad x > np
$$

Download free eBooks at bookboon.com
Click on the ad to read more

$$= \Phi\left(\frac{p_0 - \dfrac{0.5}{n} - p}{\sqrt{\dfrac{p(1-p)}{n}}}\right)$$

Example 7.11
Refer to Example 7.9. Calculate the probability that out of the selected twelve students, five or less of them will be females.

Solution
$n = 12, \quad p = 0.4 \quad p_0 = \dfrac{5}{12} = 0.42$

$$\begin{aligned} P(\hat{p} \le 0.42) &= P\left(Z \le \frac{\left(0.42 - \dfrac{0.5}{12}\right) - 0.4}{\sqrt{\dfrac{0.4(1-0.4)}{12}}}\right) \\ &= P\left(Z \le \frac{-0.002}{0.01414}\right) \\ &= \Phi(-0.1414) = 1 - \Phi(0.01414) \\ &= 1 - 0.0056 \\ &= 0.9944 \end{aligned}$$

7.4 SAMPLING DISTRIBUTION OF DIFFERENCES

7.4.1 Construction of Sampling Distribution of Differences

There are many problems in applied statistics where our interest is in two populations, such as knowing something about the difference between two population means or the difference between two population proportions. In one situation we may wish to know if it is reasonable to conclude that the two population means (or proportions) are different. In another situation we may wish to know the magnitude of the difference between two population means (or proportions). A knowledge of the sampling distribution of the difference between two means (or proportions) is useful in investigations of this type.

To empirically construct the sampling distribution of the difference between two sample means (or proportions) we would adopt the following procedure.

Suppose there are two populations, Population 1 and Population 2. We would draw all possible random samples of size n_1 from Population 1 of size N_1. There would be $\binom{N_1}{n_1}$ such samples. For each set of sample data, the sample statistic (mean or proportion) would be computed. From Population 2 of size N_2 we would draw separately and independently all possible random samples of size $\binom{N_2}{n_2}$ in all. The sample mean (or proportion) of each sample

Download free eBooks at bookboon.com

is computed and the difference between all possible pairs of the means (or proportions) is taken. The sampling distribution of the difference between sample means (or proportions) would consist of all such distinct differences, accompanied by their frequencies or relative frequencies of occurrence.

7.4.2 Sampling Distribution of Difference between Means

Theorem 7.8
Let $X_{11}, X_{12}, ..., X_{1n_1}, X_{21}, X_{22}, ...X_{2n_2}$ be $n_1 + n_2$ independent random variables, the first n_1, having identical distributions with mean μ_1 and variance σ_1^2, and the remaining n_2 having identical distributions with mean μ_2 and variance σ_2^2. Then

$$E(\overline{X}_1 - \overline{X}_2) = \mu_1 - \mu_2$$

and

$$\text{Var}(\overline{X}_1 - \overline{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

where $\overline{X}_1$ and $\overline{X}_2$ are the sample means of Populations 1 and 2 respectively

Example 7.12
Two brands of tyres are being compared by a motor firm. Brand A has a mean life of 24,000 km and a standard deviation of 1,200 km, while Brand

B has a mean life of 22,000 km and a standard deviation of 1,080 km. A random sample of 100 and 90 tyres were taken from Brand A and Brand B respectively. Compute the mean and variance of the sampling distribution of the difference between the two means.

Solution
$n_1 = 100 \quad n_2 = 90$
$\mu_1 = 24,000 \quad \mu_2 = 22,000$
$\sigma_1 = 1,200 \quad \sigma_2 = 1,080$

$$\begin{aligned}
\mu_1 - \mu_2 &= 24,000 - 22,000 = 2,000 \\
\text{Var}(\overline{X}_1 - \overline{X}_2) &= \frac{(1200)^2}{100} + \frac{(1080)^2}{90} \\
&= 14,400 + 12,960 \\
&= 27,360
\end{aligned}$$

Download free eBooks at bookboon.com

When sampling is from non-normally distributed populations or from populations whose distribution are not known, we need to take large samples so that we can apply the Central Limit Theorem, under which the difference between two samples is approximately normally distributed with mean $\mu_1 - \mu_2$ and variance $\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{n}$. To find probabilities associated with specific values of the statistic, our procedure would be the same as that given when sampling is from normally distributed populations. For example, the problem in Example 7.12 is considered to be normally distributed because n_1 and n_2 are both far greater than 30.

7.4.3 Sampling Distribution of Difference between Proportions

Suppose two random samples of different sizes are drawn from binomial populations. If we need to compare the number of successes in the samples we have to work with their proportions.

Theorem 7.9
If independent random samples of different sizes n_1 and n_2 are drawn from two binomial populations with proportions p_1 and p_2 respectively, then the distribution of the difference between the two sample proportions, $\hat{p}_1 - \hat{p}_2$ has mean

$$E(\hat{p}_1 - \hat{p}_2) = p_1 - p_2$$

and variance

$$\mathrm{Var}(\hat{p}_1 - \hat{p}_2) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$$

Example 7.13
In a national election, 60 percent of voters in Community A are in favour of a certain candidate and 50 percent in Community B are in favour of the candidate. If a sample of 210 voters from Community A and 160 voters from Community B are drawn, find

(a) the mean of the difference between the sample proportions;

(b) the variance and standard deviation of the sampling distribution of difference between the two sample proportions.

Solution

$$p_1 = 0.6 \quad p_2 = 0.5 \quad n_1 = 200 \quad n_2 = 160$$

(a) $\quad E(\hat{p}_1 - \hat{p}_2) = p_1 - p_2 = 0.6 - 0.5 = 0.1$

Download free eBooks at bookboon.com

$$
\begin{aligned}
(a) \qquad \operatorname{Var}(\hat{p}_1 - \hat{p}_2) &= \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2} \\
&= \frac{0.6(1-0.6)}{200} + \frac{0.5(1-0.5)}{160} \\
&= \frac{0.6(0.4)}{200} + \frac{0.5(0.5)}{160} \\
&= 0.0012 + 0.00156 \\
&= 0.00276 \\
\sigma_{(\hat{p}_1 - \hat{p}_2)} &= \sqrt{0.00276} = 0.0525
\end{aligned}
$$

If n_1 and n_2 are sufficiently large then the sampling distribution of $\hat{p}_1 - \hat{p}_2$ is approximately Normal. That is, by the Central Limit Theorem

$$
P(\hat{p}_1 - \hat{p}_2 \le p_{01} - p_{02}) \approx \Phi\left(\frac{(p_{01} - p_{02}) - (p_1 - p_2)}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}} \right)
$$

Example 7.14
In Example 7.13 if in the sample of voters from Community A and Community B there were 30 and 20 voters respectively in favour of the candidate what is the probability of obtaining this or a smaller difference in the sample proportions if the belief about the population parameters is correct?

Click on the ad to read more

Solution

$$\begin{aligned}
x_1 &= 30, \quad x_2 = 20 \\
p_{01} &= \frac{x_1}{n_1} = \frac{30}{210} = 0.143 \\
p_{02} &= \frac{x_2}{n_2} = \frac{20}{160} = 0.125 \\
p_{01} - p_{02} &= 0.143 - 0.125 = 0.018 \\
p_1 - p_2 &= 0.6 - 0.5 = 0.1
\end{aligned}$$

The desired probability is

$$\begin{aligned}
P(\hat{p}_1 - \hat{p}_2 \leq 0.018) &\approx \Phi\left(\frac{(p_{01} - p_{02}) - (p_1 - p_2)}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}}\right) \\
&= \Phi\left(\frac{0.018 - 0.1}{\sqrt{\dfrac{0.6(0.4)}{200} + \dfrac{0.5(0.5)}{160}}}\right) \\
&= \Phi\left(\frac{-0.082}{\sqrt{0.0012 + 0.00276}}\right) \\
&= \Phi\left(\frac{-0.082}{0.0629}\right) \\
&= \Phi(-1.30) = 0.0968
\end{aligned}$$

7.5 SAMPLING DISTRIBUTION OF VARIANCE

If $X_1, X_2, ..., X_n$ denote the random variables for a sample j of size n, then the random variable s_j^2 which denotes the *sample variance* is defined by

$$s_j^2 = \frac{(X_{1j} - \overline{X}_j)^2 + (X_{2j} - \overline{X}_j)^2 + ... + (X_{nj} - \overline{X}_j)^2}{n}$$

where $\overline{X}_{n_j}$, for instance, means the n^{th} value of the j^{th} sample and $\overline{X}_j$ is the mean of the j^{th} sample. To find the sampling distribution of variance we need to calculate the mean and variance of all the possible sample values.

7.5.1 Sampling Distribution of Variance with Replacement

Example 7.15
With reference to Examples 7.1, if a sample of size two is drawn with replacement from the population, find

(a) the mean of the sampling distribution of variances;

(b) the variance of the sampling distribution of the variances.

Download free eBooks at bookboon.com

Solution
The sample values and means have been obtained in Example 7.2. For the sake of convenience, we reproduce them here. The sample values i for each j^{th} sample of size two (X_{ij}) are:

$$\begin{array}{cccccccc} (4,4) & (4,5) & (4,7) & (4,8) & (5,4) & (5,5) & (5,7) & (5,8) \\ (7,4) & (7,5) & (7,7) & (7,8) & (8,4) & (8,5) & (8,7) & (8,8) \end{array}$$

The corresponding sample means for each j^{th} sample $(\overline{X_j})$ are:

$$\begin{array}{cccccccc} 4.0 & 4.5 & 5.5 & 6.0 & 4.5 & 5.0 & 6.0 & 6.5 \\ 5.5 & 6.0 & 7.0 & 7.5 & 6.0 & 6.5 & 7.5 & 8.0 \end{array}$$

The corresponding sample variances s_j^2 for the 16 samples are:

$$\begin{array}{cccccccc} 0 & 0.25 & 2.25 & 4.00 & 0.25 & 0 & 1.00 & 2.25 \\ 2.25 & 1.00 & 0 & 0.25 & 4.00 & 2.25 & 0.25 & 0 \end{array}$$

Note that we have defined the sample variance as

$$s_j^2 = \frac{\sum_{i=1}^n (X_{ij} - \overline{X}_j)^2}{n}$$

(a) Mean of sampling distribution of variances

$$\begin{aligned} E(s_j^2) &= \mu_{s_j} = \frac{\sum_{j=1}^k s_j^2}{N^n} \\ &= \frac{0 + 0.25 + 2.25 + ... + 0.25 + 0}{16} \\ &= 1.25 \end{aligned}$$

(b) The variance of the sampling distribution of variances

$$\begin{aligned} \text{Var}(s_j^2) &= \frac{\sum_{j=1}^k (s_j^2 - \mu_{s^2})^2}{N^n} \\ &= \frac{(0-1.25)^2 + (0.25-1.25)^2 + (2.25-1.25)^2 + ...}{16} \\ &\quad \frac{+(0.25-1.25)^2 + (0-1.25)^2}{16} \\ &= \frac{1.5625 + 1.0000 + 1.0000 + ... + 1.0000 + 1.5625}{16} \\ &= \frac{29.5}{16} \\ &= 1.84375 \end{aligned}$$

Download free eBooks at bookboon.com

Theorem 7.10
Let $X_1, ..., X_n$ be a random sample of size n, from a random variable X with expectation μ and variance σ^2. Let s^2 be the sample variance defined as

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(X - \overline{X})^2$$

If sampling is from an infinite population or with replacement from a finite population, then

$$E(s^2) = \frac{n-1}{n}\sigma^2$$

Proof

$$(X_i - \overline{X})^2 = [(X_i - \mu) - (\overline{X} - \mu)]^2$$

Taking the sum of both sides

$$\begin{aligned}
\sum_{i=1}^{n}(X_i - \overline{X})^2 &= \sum_{i=1}^{n}[(X_i - \mu)^2 - 2(\overline{X} - \mu)(X_i - \mu) + (\overline{X} - \mu)^2] \\
&= \sum_{i=1}^{n}[(X_i - \mu)^2 - 2(\overline{X} - \mu)\sum_{i=1}^{n}(X_i - \mu) \\
&\qquad +n(\overline{X} - \mu)^2] \qquad (i)
\end{aligned}$$

Need help with your dissertation?

Get in-depth feedback & advice from experts in your topic area. Find out what you can do to improve the quality of your dissertation!

Get Help Now

Go to www.helpmyassignment.co.uk for more info

Helpmyassignment

Click on the ad to read more

But

$$\begin{aligned}\sum_{i=1}^{n}(X_i-\mu) &= \sum_{i=1}^{n} X_i - \sum \mu\\ &= n\overline{X} - n\,\mu\\ &= n(\overline{X}-\mu) \qquad (ii)\end{aligned}$$

Substituting (ii) into (i) we obtain

$$\begin{aligned}\sum_{i=1}^{n}(X_i-\overline{X})^2 &= \sum_{i=1}^{n}(X-\mu)^2 - 2n\,(\overline{X}-\mu)^2 + n(\overline{X}-\mu)^2\\ &= \sum_{i=1}^{n}(X_i-\mu)^2 - n(\overline{X}-\mu)^2\end{aligned}$$

Taking the expectation of both sides, we obtain:

$$\begin{aligned}E\Big[\sum_{i=1}^{n}(X_i-\overline{X})^2\Big] &= E\Big[\sum_{i=1}^{n}(X_i-\mu)^2\Big] - nE\Big[(\overline{X}-\mu)^2\Big]\\ &= \sum_{i=1}^{n} E(X_i-\mu)^2 - nE[(\overline{X}-\mu)^2]\\ &= \sum_{i=1}^{n} \mathrm{Var}(X_i) - n\,\mathrm{Var}(\overline{X})\\ &= n\sigma^2 - n\left(\frac{\sigma^2}{n}\right)\\ &= (n-1)\sigma^2 \qquad (iii)\end{aligned}$$

Now,

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$$

and

$$\begin{aligned}E(s^2) &= E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2\right]\\ &= \frac{1}{n}\left\{E\left[\sum_{i=1}^{n}(X_i-\overline{X})\right]\right\}\\ &= \frac{n-1}{n}\,\sigma^2 \qquad (iv)\end{aligned}$$

which is very nearly σ^2 for large values of n (say $n \geq 30$). Thus, if we use s^2 as an estimate of σ^2 we shall be underestimating the population variance. To eliminate such a bias we often use what is called the desired unbiased estimator of variance (also called the corrected sample variance) defined by

$$\hat{s}^2 = \frac{n}{n-1}\,s^2$$

From this definition

$$\begin{aligned}E(\hat{s}^2) &= E\left(\frac{n}{n-1}\,s^2\right)\\ &= \frac{n}{n-1}\,E(s^2)\end{aligned}$$

Download free eBooks at bookboon.com

$$\begin{aligned} &= \frac{n}{n-1} \cdot \frac{n-1}{n}\sigma^2 \\ &= \sigma^2 \end{aligned}$$

The computational formula for the corrected sample variance is

$$\begin{aligned} \hat{s}^2 &= \frac{n}{n-1} \cdot \frac{1}{n} \sum (X_i - \overline{X})^2 \\ &= \frac{\sum (X_i - \overline{X})^2}{n-1} \end{aligned}$$

Because of this, some statisticians choose to define the sample variance by simply replacing n by $n-1$ in the denominator of the definition of s^2 given in Example 7.15. Others continue to define the sample variance by maintaining the n in the denominator, because by doing this many later results are simplified. In practice the corrected sample variance is used when $n < 30$. When $n \geq 30$ the sample variance is used. In fact if n is large there is very little difference in the two results.

Example 7.16
Refer to Example 7.1. List all possible samples of size two that can be drawn from the population with replacement and calculate

(a) the variance of the sample values;

(b) the corrected sample variance;

(c) the expectation of the sample variance.

Solution
From Example 7.2, $N = 4$, $\sigma^2 = 2.5$. From Example 7.3, $n = 2$ and $N^n = 16$. The mean of the sample means is 6.

(a) To find the variance we subtract the mean of the sample means from each of the sample means, square it and divide it by the number of the sample means.
Thus,

$$\begin{aligned} s^2 &= \frac{(4-6)^2 + (4.5-6)^2 + \cdots + (7.5-6)^2 + (8-6)^2}{16} \\ &= 1.25 \end{aligned}$$

(b) The corrected sample variance is:

$$\begin{aligned} \hat{s}^2 &= \frac{(4-6)^2 + (4-6)^2 + \cdots + (8-6)^2 + (8-8)^2}{16-1} \\ &= 2.503 \end{aligned}$$

From the results in Example 7.16, we may note the following.

(i) The expectation of the sample variance is

$$E(s^2) = \frac{n-1}{n}\sigma^2 = \frac{2-1}{2}\Big(2.5\Big) = 1.25$$

which is the same as the mean of the sampling distribution of variances obtained in Example 7.15(a).

(ii) As the number of sample values increases, s^2 and $\hat{s}^2$ tend to be the same. Using the definitional formula for the corrected sample variance we obtain

$$\hat{s}^2 = \frac{n}{n-1}\,s^2 = \frac{2}{2-1}(1.25) = 2.5$$

which is the same as the value obtained in part (b) of the example above with a difference of 0.003.

(iii) The expectation of the corrected sample variance is

$$E(\hat{s}^2) = \frac{n}{n-1}\,E(s^2) = \frac{2}{2-1}(1.25) = 2.5 = \sigma^2$$

(see Example 7.1).

7.5.2 Sampling Distribution of Variance Without Replacement

Example 7.17
With reference to Examples 7.1, if a sample of size two is drawn without replacement, find
(a) the mean, (b) the variance
of the sampling distribution of variances.

Solution
The sample values and means have been obtained in Example 7.5. For the sake of convenience, we reproduce them here.

$$(4,5) \quad (4,7) \quad (4,8) \quad (5,7) \quad (5,8) \quad (7,8)$$

Click on the ad to read more

The corresponding sample means are

$$(4.5) \quad (5.5) \quad (6.0) \quad (6.0) \quad (6.5) \quad (7.5)$$

(a) Mean of sampling distribution of variances:

$$\begin{aligned} E(s_j^2) &= \mu_{s_j^2} = \frac{\sum s^2}{\binom{N}{n}} \\ &= \frac{0.25 + 2.25 + 4.00 + 1.00 + 2.25 + 0.25}{6} \\ &= \frac{10}{6} = 1.6667 \end{aligned}$$

(b) The variance of sampling distribution of variances is

$$\begin{aligned} \text{Var}(s_j^2) &= \frac{\sum_{j=1}^{k}(s_j^2 - \mu_{s^2})}{\binom{4}{2}} \\ &= \frac{(0.25 - 1.6667)^2 + (2.25 - 1.6667)^2}{6} \\ &\quad + \frac{(4.00 - 1.6667)^2 + (1 - 1.6667)^2}{6} \\ &\quad + \frac{(2.25 - 1.6667)^2 + (0.25 - 1.6667)^2}{6} \\ &= \frac{2.0070 + 0.3404 + \cdots + 0.3404 + 2.0070}{6} \\ &= \frac{10.5836}{6} \\ &= 1.7639 \end{aligned}$$

Theorem 7.11
Let $X_1, X_2, ..., X_n$ be a random sample of size n, with expectation μ and variance σ^2. Let s^2 be the sample variance defined as

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

If sampling is without replacement from a finite population of size N, then the variance of the sampling distribution of variances is

$$\text{Var}(s^2) = \left(\frac{N}{N-1}\right)\left(\frac{n-1}{n}\right)\sigma^2$$

Download free eBooks at bookboon.com

Note
As $N \to \infty$ this result reduces to that in Theorem 7.10.

Example 7.18
With reference to Examples 7.17 and 7.5, find the variance of the sample variance s^2 if sampling is without replacement.

Solution
$N = 4, \quad n = 2, \quad \sigma^2 = 2.5$ (from Example 7.2). Hence,

$$\begin{aligned} \mathrm{Var}(s^2) &= \left(\frac{4}{4-1}\right)\left(\frac{2-1}{2}\right)(2.5) \\ &= 1.6667 \end{aligned}$$

which is equal to the results obtained in Example 7.17.

To find the sampling distribution of the sample variance, it is more convenient to do so in terms of the related random variable.

Theorem 7.12
If random samples of size n are taken from a population having a Normal distribution, and if

$$s^2 = \frac{\sum_i (X_i - X)^2}{n-1}$$

then

$$\frac{(n-1)s^2}{\sigma^2}$$

has a Chi-square distribution with $n-1$ degrees of freedom

Note
Chi-square distribution is discussed in Chapter 8.

Theorem 7.13
If X is normally distributed with mean μ and variance σ^2 and $X_1, X_2 ..., X_n$ is a random sample of size n of X, then the random variable

$$V = \frac{\sum_i (X_i - \mu)^2}{\sigma^2}$$

will possess a Chi-square distribution with n degrees of freedom

This chapter concludes discussions on basic theory on sampling and sampling distributions. The next chapter continues discussions on sampling distributions which are based on small samples.

Download free eBooks at bookboon.com

EXERCISES

7.1 A population consists of five values $2, 4, 6, 8$ and 14.

(a) Calculate the mean and variance of the population.

(b) List all possible distinct samples of size 2 drawn with replacement from the population and compute the mean of each sample.

(c) Construct the sampling distribution of the sample means and calculate the mean and variance of the distribution.

(d) Comment on your results.

7.2 Rework Exercise 7.1 for the case when a random sample of size 2 is drawn without replacement.

7.3 Referring to Exercise 7.1, calculate

(a) the expectation of the sample mean;

(b) the variance and hence the standard deviation of the sample mean if sampling is made; (i) with replacement; (ii) without replacement.

Download free eBooks at bookboon.com
Click on the ad to read more

7.4 A random sample of 25 adults is taken from a Normal population with mean height of 1.65 metres and standard deviation of 0.8 metres. Find

(a) (i) the expectation of the sample mean;
(ii) the variance and standard deviation of the sample mean;

(b) the probability that the height of an adult selected at random will be
(i) less than 1.50 metres;

(ii) not less than 1.70 metres; (iii) between 1.55 and 1.68.

7.5 The weights of the products of a certain factory have a mean of 140 kilograms and a standard deviation of 20 kilograms. If 180 of the products are selected at random, find

(a) (i) the expectation of the sample mean;
(ii) the variance and standard deviation of the sample mean;

(b) the probability that a product selected at random will weigh
(i) at most 150 kilograms; (ii) at least 120 kilograms;
(iii) between 110 and 155 kilograms.

7.6 The time that a cashier spends in processing each person's order is an independent random variable with mean of 45 minutes and standard deviation of 30 minutes. What is the approximate probability that the orders of 100 persons can be processed in less than 4000 minutes?

7.7 Suppose a population consists of 10 identical balls in a box, four of which are white and six green. A random sample of five is drawn without replacement. Find

(a) the expectation of the sample proportion of white balls;

(b) the variance and hence the standard error of the sample proportion.

7.8 Steel wires produced by a certain factory have a mean tensile strength of $1,000$ kilograms and a variance of 900 kilograms. If a random sample of 250 wires is drawn from the production line during a certain month with a total output of $100,000$ units, find

(a) (i) the expectation and (ii) the variance of the sample mean tensile strength;

(b) the probability that the sample mean will
(i) be more than $1,010$ kilograms;
(ii) be less than $1,003$ kilograms;
(iii) be between 995 and $1,006$ kilograms;
(iv) differ from $1,000$ by 5 kilograms or more;
(v) differ from $1,000$ by at most 7 kilograms.

7.9 Repeat Exercise 7.7 for the case when sample is with replacement.

Download free eBooks at bookboon.com

7.10 A group of children consists of twenty males and twenty-five females. A random sample of twenty-three children is drawn from this group with replacement.

(a) Find

(i) the expectation of the proportion of females;

(ii) the variance and hence the standard error of the sample proportion of females.

(b) What is the probability that the sample proportion of females is at least 0.6?

7.11 Rework Exercise 7.10 if sampling was made without replacement.

7.12 Thirty-six percent of University staff are against a strike action. If a sample of 120 of them are drawn at random with replacement, find

(a) (i) the expectation of the sample proportion,

(ii) the variance and hence the standard error of the sample proportion,

of the University staff who are against a strike action.

(b) What is the probability that the proportion of the University staff who are against the strike will be between 0.5 and 0.7.

7.13 Rework Exercise 7.12 if sampling is without replacement and if there are 400 university staff members.

7.14 A box contains 80 black balls and 60 white balls. Two samples of 40 marbles each are randomly selected with replacement from the box and their colours noted. Suppose that the first and second samples contained 30 and 25 black balls respectively. Find

(a) the variance of the difference in proportion of black balls in the two sample;

(b) the probability that the difference between the two samples does not exceed 10 balls.

7.15 Rework Exercise 7.14 if sampling is without replacement.

7.16 A random sample of size 5 is drawn from a population which is normally distributed with mean 49 and variance 9. A second random sample of size 4 is drawn from a different population which is also normally distributed with mean 39 and variance 4. The two samples are independent with means $\overline{X}_1$ and $\overline{X}_2$ respectively. Let $W = \overline{X}_1 - \overline{X}_2$. Calculate (a) the mean and variance of W; (b) $P(W > 8.2)$.

7.17 With reference to Exercise 7.2, find

(a) (i) the mean; (ii) the variance;
of the sampling distribution of variances;

(b) the corrected variance for sampling.

7.18 With reference to Exercise 7.1, find

(a) (i) the mean (ii) the variance
of the sampling distribution of variances;

(b) the corrected variance for sampling.

Download free eBooks at bookboon.com

Chapter 8

DISTRIBUTIONS DERIVED FROM NORMAL DISTRIBUTION

8.1 INTRODUCTION

So far we have encountered distributions which are related to the Normal distribution on the basis of the Central Limit Theorem. Recall that by the Central Limit Theorem (CLT) the probability distribution of a sample statistic from an independent identical non-Normal distribution is approximately Normal if the sample size is sufficiently large. When the sample size is small we cannot apply the CLT or assume that sampling distributions are Normal. The sampling distribution in this case differs from one case to another. Besides the binomial and Poisson distributions there are three probability distributions, which are byproducts of the Normal distribution such can be described by one of them under certain conditions. These are the *Chi-square* (χ^2) distribution, the *Student's* t (briefly t) distribution, and the *Fisher's variance ratio* (briefly F) distribution.

Click on the ad to read more
Download free eBooks at bookboon.com

8.2 χ^2 DISTRIBUTION

Perhaps, the Chi-square distribution is, of all statistical distributions, the most widely used one in statistical applications. We have already encountered the Chi-square distribution in Chapter 19 where we considered sampling distribution of the sample variance.

8.2.1 Definition of Chi-square Distribution

Recall from Volume II (Nsowah-Nuamah, 2018) that the Chi-square distribution is a special type of the Gamma distribution with $\alpha = \frac{v}{2}$ and $\beta = 2$.

Definition 8.1 CHI-SQUARE DISTRIBUTION
A continuous random variable X is said to have a Chi-square distribution with parameter $v > 0$ if its probability density function is given by

$$f(x) = \frac{1}{2^{\frac{v}{2}}\Gamma\left(\frac{v}{2}\right)} x^{\frac{v}{2}-1} e^{-\frac{x}{2}}, \qquad x > 0$$

The Chi-square distribution is completely defined by its parameter v which is called the number of degrees of freedom (df). If a random variable X follows a Chi-square distribution with v degrees of freedom then we write $X \sim \chi^2_{(v)}$. Fig. 8.1 depicts the density curves of the Chi-square distributions for a few selected values of v.

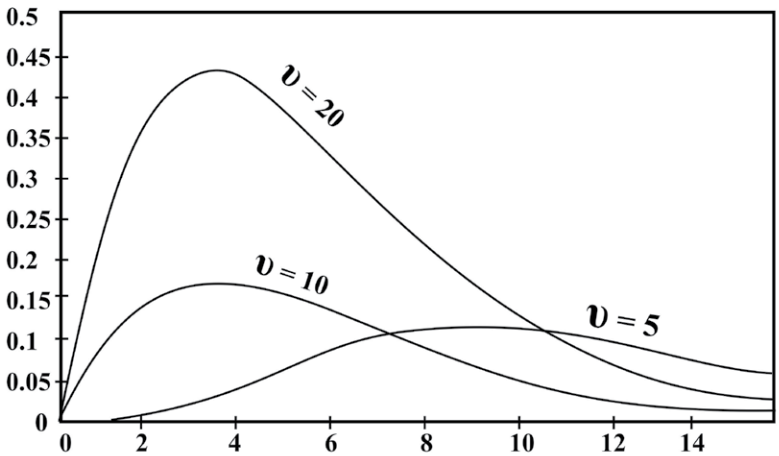


Fig. 8.1 χ^2 Distribution with various Degrees of Freedom

Download free eBooks at bookboon.com

8.2.2 Properties of Chi-square Distribution

Property 1

Theorem 8.1
Suppose X has a χ^2 with parameter v distribution. Then the moment-generating function of X is

$$M_X(t) = (1-2t)^{-\frac{v}{2}}, \qquad t < \frac{1}{2}$$

The proof of this theorem is left as an exercise for the reader (see Exercise 8.1)

Example 8.1
Suppose X has a moment generating function (m.g.f)

$$M_X(t) = (1-2t)^{-10}, \qquad \text{for } t < \frac{1}{2}$$

then this is a χ^2 distribution. What is its degree of freedom.

Solution
The Chi-square distribution has v degrees of freedom. The exponent of the m.g.f. equals -10. Hence $-10 = -\frac{v}{2}$ which gives $v = 20$.

Example 8.2
Suppose X has a Chi-square distribution with 5 degrees of freedom. Find its m.g.f.

Solution
From Theorem 8.1 with $v = 5$, the moment generating function of X is

$$M_X(t) = (1-2t)^{-\frac{v}{2}} = (1-2t)^{-\frac{5}{2}}$$

Property 2

Theorem 8.2
Suppose X has a χ^2 distribution with parameter v. Then

$$\text{(a) } E(X) = v \quad \text{(b) } \mathrm{Var}(X) = 2\,v$$

Download free eBooks at bookboon.com

Proof
Since the Chi-square distribution is a special case of the Gamma distribution when $\alpha = \frac{v}{2}$ and $\beta = 2$, its expectation and variance is obtained simply by substituting these values in the formulas of Theorem 10.18. Thus,

Example 8.3
Refer to Example 8.2. Find the (a) mean (b) variance.

Solution
(a) $E(X) = v = 5$ (b) $\mathrm{Var}(X) = 2\,v = 10$

Property 3

Theorem 8.3
The cumulative distribution function of the Chi-square distribution is given by

$$2^{\frac{v}{2}}\Gamma\left(\frac{v}{2}\right)\int_0^x t^{\frac{v}{2}-1}\,e^{-\frac{x}{2}}\,dt$$

Download free eBooks at bookboon.com
Click on the ad to read more

Property 4

Theorem 8.4
If $X_1, X_2, ..., X_n$ are independent random variables having Chi-square distributions with $v_1,\ v_2,\ ...\ ,\ v_n$ degrees of freedom respectively, then the distribution of $Y = X_1 + X_2 + ... + X_n$ will possess a Chi-square distribution with $v_1 + v_2 + ... + v_n$ degrees of freedom

Corollary 8.1
If $X_1,\ X_2,\ ...,\ X_n$ are independent random variables having Chi-square distributions with 1 degree of freedom each, then the distribution of $Y = X_1 + X_2 + ... + X_n$ will possess a Chi-square distribution with n degrees of freedom.

Corollary 8.2
If X and Y are independent and $X \sim \chi^2_v$ and $Y \sim \chi^2_u$, then $Z = X + Y \sim \chi^2_{v+u}$.

Property 5

Theorem 8.5
If X_1, and X_2 are independent random variables, X_1 has a Chi-square distribution with v_1 degrees of freedom and $X_1 + X_2$ has a Chi-square distribution with $v\,(> v_1)$ degrees of freedom, then X_2 has a Chi-square distribution with $v - v_1$ degrees of freedom.

The Chi-square distribution is positively skewed (Fig. 8.1). It has been extensively tabulated. Table I in Appendix A contains the values of $\chi^2_{\alpha,v}$ for

$$P\left(\chi^2 \le \chi^2_{\alpha,v}\right) = \int_0^{\chi^2_{\alpha,v}} f(x)\,dx = 1 - \alpha$$

where

$$\alpha = 0.005,\ 0.01,\ 0.025,\ 0.05,\ 0.95,\ 0.975,\ 0.99,\ 0.995 \,\text{and}\ v = 1, 2, ..., 30$$

It is a convention in statistics to use the same symbol χ^2 for both the random variable and a value of that random variable. Thus percentile values of the Chi-square distribution with v degrees of freedom is denoted by $\chi^2_{\alpha,v}$ or simply χ^2_α if n is understood, where the suffix α is used from now on and in applications of probability and Statistics, to denote "lower percentile" ($100\alpha\%$ point).

Example 8.4
Find from Table I the following:
(a) $\chi^2_{0.025,\,13}$ (b) $\chi^2_{0.99,\,5}$ (c) $\chi^2_{0.99,\,28}$ (d) $\chi^2_{0,95,\,3}$ (e) $\chi^2_{0.975,\,10}$

Download free eBooks at bookboon.com

Solution

(a) This is a Chi-square for probability 0.025 and degrees of freedom of 13. Now, referring to Table I, we proceed downward under column labelled d.f. until we reach entry 12. Then we proceed right to meet column headed $\chi^2_{0.025}$. The value at that meeting point, which is 24.736, is the required value of $\chi^2_{0.025,12}$. Therefore the probability that a χ^2-distributed random variable with 12 degrees of freedom does not exceed 24.736 is 0.025.

(b) This is a Chi-square for probability 0.99 and degree of freedom of 5. Proceeding as in (a), that is, we go to Table I and proceed downward under column v until we reach the entry 4. Then proceed right to meet column headed $\chi^2_{0.99}$. The value at the meeting point 0.554 is the required value of $\chi^2_{0.99,5}$. Therefore the probability that a χ^2-distributed random variable with 5 degrees of freedom does not exceed 0.554 is 0.99.

The reader should complete this example.

Example 8.5
Suppose a certain Chi-square distribution value for a probability of 0.975 is 30.2. What is the degree of freedom.

Solution
Refer to Table I and proceed downward under column 0.975 until you locate the value 30.2 in the body of the table. Then proceed left to meet column headed v. The value of $v = 17$ is the required degrees of freedom.

8.2.3 Relationship Between Chi-Square and Normal Distributions

Chi-square Distribution appears in the theory associated with random variables which are normally distributed.

> **Theorem 8.6**
> Let Z be a Standard Normal variable. Then $U = Z^2$ is a Chi-square distribution with 1 degree of freedom

The Chi-square distribution with 1 degree of freedom is denoted by $\chi^2_{(1)}$.

Proof
In Theorem 10.28, we proved that

$$P(U \leq u) = \frac{1}{\sqrt{2\pi}} \int_0^u t^{-\frac{1}{2}} e^{-\frac{t}{2}} dt \qquad (i)$$

If we put $v = 1$ in the Chi-square cumulative distribution function (Theorem 8.3), that is, if we consider Chi-square distribution with 1 degree of freedom, and recalling also that $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, we will obtain expression (i).

Download free eBooks at bookboon.com

Note
If $X \sim N(\mu, \sigma)$, then

$$\left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi_1^2$$

Theorem 8.7
Let $Z_1, Z_2, \ldots, Z_v$ be n independent Standard Normal random variables. Then the sum of the squares of these variables is a Chi-square variable, χ^2, with v degrees of freedom. That is,

$$\begin{aligned} \chi^2_{(v)} &= Z_1^2 + Z_2^2 + \ldots + Z_v^2 \\ &= \sum_{i=1}^{v} Z_i^2 \quad \text{where } Z_i \text{ are independent} \end{aligned}$$

Example 8.6
Suppose $Z_1, Z_2, \ldots, Z_v$ is a random sample from the Standard Normal distribution. Find the number c such that

$$P\left(\sum_{i=1}^{22} Z_i^2 > c\right) = 0.10$$

Click on the ad to read more
Download free eBooks at bookboon.com

Solution

By Theorem 8.7, $\sum_{i=1}^{22} Z_i$ has a χ^2 distribution with 22 degrees of freedom.

Reading from Table I in Appendix A the row labelled 22 d.f. and the column headed by an upper-tail area of 0.10, we get the number 30.813. Therefore

$$P\left(\sum_{i=1}^{22} Z_i^2 > 30.813\right) = 0.10$$

which may equivalently be stated as

$$P\left(\sum_{i=1}^{22} Z_i^2 \leq 30.813\right) = 0.90$$

Thus $c = 30.813$

Theorem 8.8
The Chi-square distribution approaches the Normal distribution $N(v, 2v)$ as the degrees of freedom v gets large

In practice when $v > 30$, Chi-square probabilities can be computed by employing Normal functions in the usual fashion.

8.2.4 Relationship Between Chi-Square and Multinomial Distributions

Theorem 8.9
Suppose $X_1, X_2, \ldots, X_s$ has a joint multinomial distribution with parameters n, $p_1, p_2, \cdots, p_s$. Then the sum

$$\sum_{i=1}^{s} \frac{(X_i - np_i)^2}{np_i}$$

tends to the Chi-square distribution as $n \to \infty$

Theorem 8.9 is very important in applied statistics. It is used for goodness-of-fit test and contingency-table tests (tests of independence and homogeneity). The details of these tests may be found in Statistics books.

8.3 t DISTRIBUTION

The t-distribution was first introduced by W. Gosset who published his work under the name of "Student".

Download free eBooks at bookboon.com

8.3.1 Definition of t Distribution

Definition 8.2 t DISTRIBUTION
A continuous random variable X is said to have the Student's t distribution with parameter v if its probability density function is given by

$$f(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\,\pi}\,\Gamma\left(\frac{v}{2}\right)}\left(1+\frac{x^2}{v}\right)^{-\frac{v+1}{2}} \qquad -\infty < x < \infty$$

The parameter v is called the degree of freedom. The t distribution with v degrees of freedom is usually denoted by $t_{\alpha,v}$ or simply t_v. Fig. 8.2 shows the curve of the density of the t distribution for various degrees of freedom.

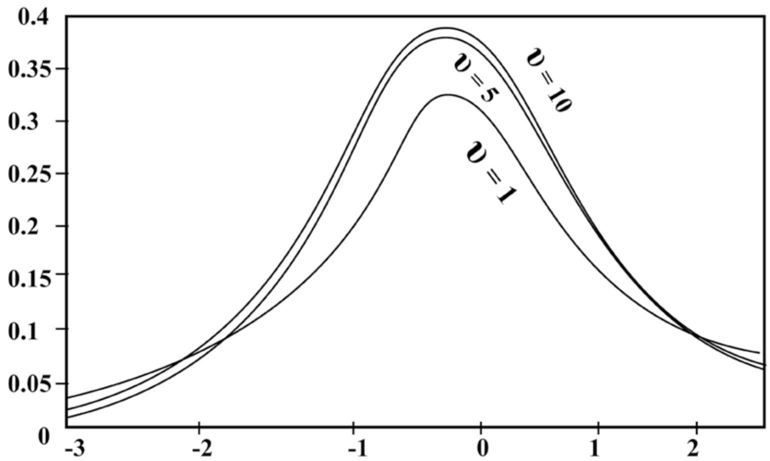


Fig. 8.2 t Distribution with various Degrees of Freedom

8.3.2 Properties of t Distribution

Property 1
The t distribution is symmetric and for $\alpha = 0.50$, $t_{(1-\alpha),v} = -t_{\alpha,v}$

Property 2

Theorem 8.10
Suppose a random variable X has the Student's t distribution with parameter v. Then

(a) $E(X) = 0$, for $v > 1$

(a) $\text{Var}(X) = \dfrac{v}{v-2}, \quad \text{for } v > 2$

Thus a t-distribution possesses no mean when $v = 1$ and the variance does not exist for $v \leq 2$.

Property 2

Theorem 8.11
If X_1, $X_2 \cdots$, X_n is a random sample from the Normal population having the mean μ and variance σ^2, and let $\overline{X}$ be the sample mean, then

$$T = \frac{\overline{X} - \mu}{\hat{s}/\sqrt{n}}$$

has the Student's t distribution with $v = n - 1$ degrees of freedom where

$$\hat{s}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

The t-distribution has been tabulated. Table II in Appendix A contains values of $t_{(1-\alpha),v}$ for

$$\alpha = 0.10, 0.05, 0.025, 0.01, 0.005 \text{ and } v = 1, 2, 3, 4, ..., 30, 40, 60, 120, \infty,$$

where $t_{\alpha,v}$ is such that the area to the left of $t_{\alpha,v}$ under the curve of the t distribution with v degrees of freedom is equal to $1 - \alpha$. That is, $t_{1-\alpha,v}$ is such that

$$P\left(t \leq t_{\alpha,v}\right) = \int_{-\infty}^{t_{\alpha,v}} f(x)\, dx = 1 - \alpha$$

To say that $P\left(t \leq t_{\alpha,v}\right) = 1 - \alpha$ is the same as saying that

$$P\left(t > t_{-\alpha,v}\right) = 1 - P(t > t_{\alpha;\, v})$$

It is therefore uncommon to find texts reporting the α instead of $1 - \alpha$ at the top of the t distribution table.

Example 8.7
If t distribution has 25 degrees of freedom, find
(a) $t_{0.975,25}$ (b) $t_{0.05,25}$

Solution

(a) From Table II in Appendix A with $v =$ d.f. $= 25$ and $1 - \alpha = 0.975$ we proceed downward under column d.f. until we reach 25. Then we proceed right to meet column headed 0.975. The value at that meeting point, which is 2.0595 is the required value.

Therefore the probability that a random variable with t distribution with 25 degrees of freedom does not exceed 3.345 is 0.975.

Download free eBooks at bookboon.com

(b) The value of t at $\alpha = 0.05$ can be obtained from the value of t at $1 - \alpha = 0.95$ in Table II in Appendix A. Using the same procedure as in (a), the reader will see that $t_{0.05,25} = 1.708$.

Therefore the probability that a t- distributed random variable with 25 degrees of freedom does not exceed 1.708 is 0.90.

8.3.3 Relationship Between t, χ^2 and Normal Distributions

Theorem 8.12
If Z is a Standard Normal variable and χ^2 is a Chi-square distribution with v degrees of freedom and Z and χ^2 are independently distributed then the distribution

$$T = \frac{Z}{\sqrt{\frac{\chi^2}{v}}}$$

has a Student's t distribution with v degrees of freedom

Proof
We simply express the given ratio in a different form as follows:

$$\begin{aligned}\frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}} &= \frac{\left(\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}\right)}{\sqrt{\frac{s^2}{\sigma^2}}} \\ &= \frac{Z}{\sqrt{\frac{(n-1)s^2}{(n-1)\sigma^2}}} \\ &= \frac{Z}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}}\end{aligned}$$

since $\frac{(n-1)s^2}{\sigma^2} = \chi^2_{n-1}$ by Theorem 7.12

By Definition 8.11, this is a t distribution with $n-1$ degrees of freedom.

Note
This quantity is usually called 'Student's t and the corresponding distribution is called 'Student's t distribution.

Theorem 8.13
The t curve approaches a Normal curve as n approaches infinity

The t distribution, however, has a greater dispersion than the Standard Normal distribution. In practice, we can treat t_n as $N(0, 1)$ when $n > 30$.

Download free eBooks at bookboon.com

The t distribution, however, has a greater dispersion than the Standard Normal distribution. In practice, we can treat t_n as $N(0,1)$ when $n > 30$.

8.4 F DISTRIBUTION

The F distribution was named after Sir R.A. Fisher.

8.4.1 Definition of F Distribution

Definition 8.3 F DISTRIBUTION
A continuous random variable X is said to have the distribution with parameter v if its probability density function is given by

$$f(x) = \frac{\Gamma\left(\dfrac{v_1+v_2}{2}\right)}{\Gamma\left(\dfrac{v_1}{2}\right)\Gamma\left(\dfrac{v_2}{2}\right)} \frac{\left[\left(\dfrac{v_1}{v_2}\right)x\right]^{\frac{v_1}{2}-1}}{\left[1+\left(\dfrac{v_1}{v_2}\right)x\right]^{\dfrac{v_1+v_2}{2}}}$$

where $x > 0$

Deloitte.
Discover the truth at www.deloitte.ca/careers
© Deloitte & Touche LLP and affiliated entities.
Download free eBooks at bookboon.com
Click on the ad to read more

The F distribution with v_1 and v_2 degrees of freedom is denoted by $F_{\alpha;v_1,v_2}$. Note that the order v_1, v_2 is important, that is, in general,

$$F_{\alpha;v_1;v_2} \neq F_{\alpha;v_2;v_1}$$

Density curves for a few selected degrees of freedom of an F distribution are shown in Fig 8.3.

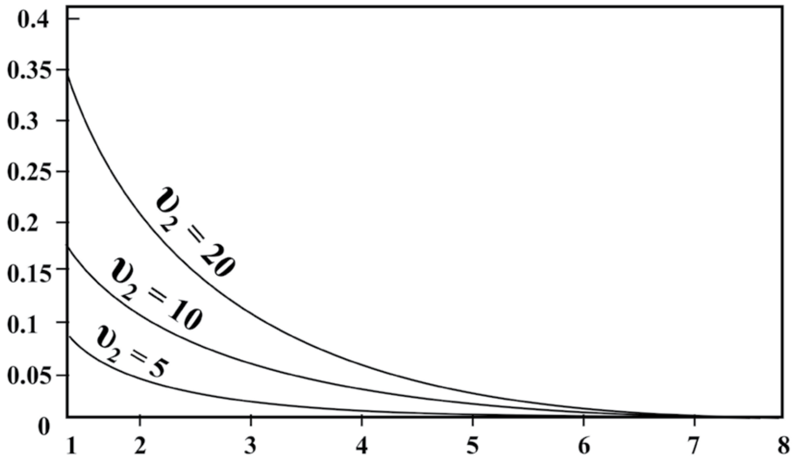


Fig. 8.3 (a) Typical Curves of F Distribution
($v_1 = 2$ with various values of v_2)

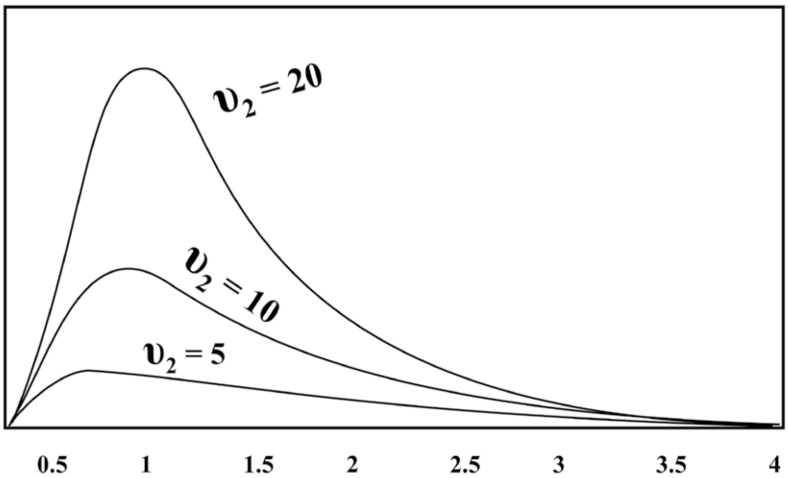


Fig. 8.3 (b) Typical Curves of F Distribution
($v_1 = 20$ with various values of v_2)

Download free eBooks at bookboon.com

8.4.2 Properties of F distribution

Property 1
As a ratio of two non-negative values ($\chi_v^2 \geq 0$), an F random variable ranges in value from 0 to $+\infty$.

Property 2

Theorem 8.14
Suppose a random variable X has the F distribution with parameters v_1 and v_2. Then

(a) $E(X) = \dfrac{v_2}{v_2 - 2}, \quad v_2 > 2$

(b) $\text{Var}(X) = \dfrac{2v_2^2(v_1 + v_2 - 2)}{v_1(v_2 - 2)^2(v_2 - 4)}, \quad v_2 > 4$

Theorem 8.14 implies that an F variable has no mean when $v_2 \leq 2$ and has no variance when $v_2 \leq 4$.

Property 3 RECIPROCAL PROPERTY

Theorem 8.15
Suppose X has the distribution F_{v_1,v_2} then $Y = \dfrac{1}{X}$ has the distribution F_{v_2,v_1}, that is,

$$F_{(1-\alpha,v_1,v_2)} = \frac{1}{F_{(\alpha,v_1,v_2)}}$$

where both α and $1 - \alpha$ are lower - tail percentage points of an F distribution

Property 4
F is a positively skewed distribution; however, its skewness decreases with increasing v_1 and v_2.

The F distribution has been tabulated for $\alpha = 0.50, 0.90, 0.95, 0.975, 0.99,$ $0.995, 0.999$. Table III in Appendix A gives the percentage points for the right tail of several F distributions at 1, 5 and 10 percent levels. To say that $P(F > F_{v_1,v_2}) = \alpha$ is the same as saying that

$$P(0 \leq F \leq F_{v_1,v_2}) = 1 - \alpha$$

Download free eBooks at bookboon.com

Example 8.8
If the F distribution has degrees of freedom $v_1 = 30$ and $v_2 = 24$, find the table value of
(a) $F_{0.90;30,24}$ (b) $F_{0.95;12,8}$ (c) $F_{0.05;12,8}$

Solution

(a) Table III in Appendix A gives the percentage point of the distribution at 90 percent levels. Read the value of the table at the meeting point of the degrees of freedom of the numerator, $v_1 = 30$ and degrees of freedom of the denominator, $v_2 = 24$. The value is 1.67. That is, the probability that an F-distributed random variable with 30 numerator degrees of freedom and 24 denominator degrees of freedom does not exceed 1.67 is 0.90.

(b) From Table III in Appendix A, the value is 3.28. Hence

$$F_{0.95;12,8} = 3.28$$

(c) By Theorem 8.15

$$F_{0.05;12,5} = \frac{1}{F_{0.95;8,12}} = \frac{1}{3.28} = 0.305$$

Download free eBooks at bookboon.com
Click on the ad to read more

8.4.3 Relationship among χ^2, t and F Distributions

Theorem 8.16
Let $\chi^2_{(v_1)}$ and $\chi^2_{(v_2)}$ be two independent χ^2 random variables with parameters v_1 and v_2 respectively. Then F-distribution is given by

$$F_{(v_1,v_2)} = \frac{\chi^2_{(v_1)}}{v_1} \Big/ \frac{\chi^2_{(v_2)}}{v_2}$$

where v_1 is the degrees of freedom of the numerator and v_2 is the degrees of freedom of the denominator

Theorem 8.17
Let $\chi^2_{\alpha,v}$ be a Chi-square random variable with v degrees of freedom. Then

$$F_{\alpha,v,\infty} = \frac{\chi^2_{\alpha,v}}{v}$$

Theorem 8.18
Let $F_{1,v}$ be an F distribution with degrees of freedom 1 and v. Then

$$F_{1-\alpha;\,1,\,v} = t^2_{1-\frac{\alpha}{2},\,v}$$

Proof
From Theorem 8.12

$$T^2 = \frac{Z^2}{\frac{\chi^2}{v}}$$

But from Theorem 12.6
$Z^2 \sim \chi^2$ with 1 degree of freedom. Hence,

$$t_v^2 = \frac{\frac{Z^2}{1}}{\frac{\chi^2}{v}} = F_{1,\,v}$$

Example 8.9
Verify that $F_{0.95;1,16} = t^2_{.975,16}$

Solution
From Table III in Appendix A,

$$F_{0.95;1,16} = F_{1-0.05;1,16} = 4.49$$

From Table III in Appendix A,

$$t_{1-\frac{\alpha}{2},16} = t_{0.975,16} = 2.12$$

Comparing the F and t values, we observe that

$$4.49 = (2.12)^2$$

Download free eBooks at bookboon.com

The χ^2, t and F distributions have many important applications. The reader will find them in most aspects of both theoretical and applied Statistics especially under Inferential Statistics (estimation and hypothesis testing).

EXERCISES

8.1 Prove Theorem 8.1.

8.2 A random variable X has the χ^2 distribution. Find the degrees of freedom if its moment generating function is

(a) $M_X(t) = (1-2t)^{-18}$,

(b) $M_X(t) = \dfrac{1}{(1-2t)^6}$,

(c) $M_X(t) = (1-2t)^{-32}$,

where $t < \frac{1}{2}$

8.3 If X has Chi-Square distribution

$$f(x) = \frac{2}{9}\,x^5\,e^{-\frac{x}{2}}$$

How many degrees of freedom has $f(x)$.

8.4 A random variable X has the χ^2 distribution. Find the m.g.f. of X if the distribution has degrees of freedom (a) 4, (b) 10, (c) 28.

8.5 Referring to Exercise 8.3. Find the mean and variance of the χ^2 variable.

8.6 Suppose X and Y are independent χ^2 random variables with v_1 and v_2 degrees of freedom, respectively. Show that the m.g.f. of $Z = X + Y$ is

$$M_X(t) = (1-2t)^{-(\frac{v_1+v_2}{2})}$$

8.7 If a random variable X has the χ^2 distribution. Find from Table I in Appendix A the value for
(a) $\chi^2_{0.90,15}$ (b) $\chi^2_{0.99,8}$ (c) $\chi^2_{0.025,21}$

8.8 If X has the χ^2 distribution with degrees of freedom of 12, find x such that

(a) $P(X \le x) = 0.05$ (b) $P(X \ge x) = 0.975$

8.9 If X has the t distribution, find from Table II in Appendix A the value for
(a) $t_{0.95,28}$ (b) $t_{0.975,15}$ (c) $t_{0.90,22}$

Download free eBooks at bookboon.com

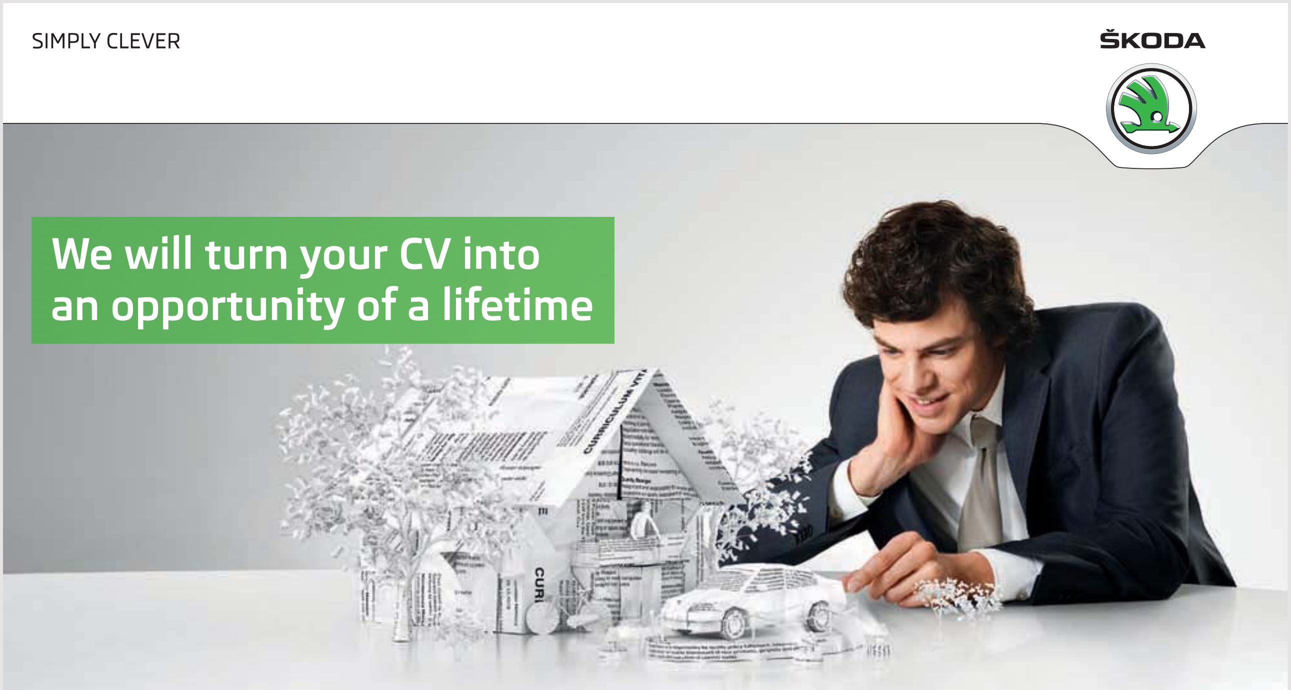
8.10 Suppose X has t distribution and given
(a) $P(X \geq 1.323)$; (b) $P(X \leq 3.078)$
find (i) the degrees of freedom (ii) α

8.11 If X is F_{v_1,v_2}, find from Table III in Appendix A the value for (a) $F_{0.95,20,30}$ (b) $F_{0.90,20,24}$

8.12 Suppose X has the F distribution with $v_1 = 15$ and $v_2 = 8$, find x such that (a) $P(X \geq x) = 0.10$ (b) $P(X \leq x) = 0.10$

8.13 Suppose X is F_{v_1,v_2}. Find the degrees of freedom if
(a) $P(X \geq 2.49) = 0.05$ (b) $P(X \leq 3.34) = 0.99$

8.14 Verify that (a) $F_{0.95;1,20} = t^2_{0.975,20}$ (b) $F_{0.90;1,7} = t^2_{0.95,7}$



Do you like cars? Would you like to be a part of a successful brand?
We will appreciate and reward both your enthusiasm and talent.
Send us your CV. You will be surprised where it can take you.
Send us your CV on
www.employerforlife.com
Click on the ad to read more
Download free eBooks at bookboon.com

STATISTICAL TABLES

Download free eBooks at bookboon.com

Table I

Standard Normal Distribution
(Full Table)

$$\Phi(z) = \frac{1}{\sqrt{2\,\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}}\, dt$$

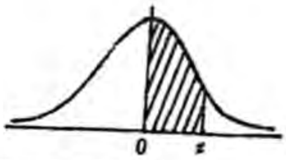
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	z
0.00	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	:5319	.5359	0.00
0.10	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753	0.10
0.20	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6054	.6103	.6141	0.20
0.30	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517	0.30
0.40	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879	0.40
0.50	.6915	.6950	.6985	.7019	.7054.	.7088	.7123	.7157	.7190	.7224	0.50
0.60	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549	0.60
0.70	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852	0.70
0.80	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133	0.80
0.90	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389	0.90
1.00	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621	1.00
1.10	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830	1.10
1.20	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015	1.20
1.30	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177	1.30
1.40	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319	1.40
1.50	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441	1.50
1.60	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545	1.60
1.70	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633	1.70
1.80	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706	1.80
1.90	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767	1.90
2.00	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817	2.00
2.10	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857	2.10
2.20	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890	2.20
2.30	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916	2.30
2.40	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936	2.40
2.50	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952	2.50
2.60	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964	2.60
2.70	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974	2.70
2.80	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981	2.80
2.90	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986	2.90
3.00	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990	3.00
3.10	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993	3.10
3.20	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995	3.20
3.30	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997	3.30
3.40	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998	3.40
3.50	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9993	.9998	.9998	3.50
3.60	.9998	.9998	.9999.	.9999	.9999	.9999	.9999	.9999	.9999	.9999	3.60
3.70	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	3.70
3.80	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	3.80

Table II

Standard Normal Distribution

(Half Table)

$$\Phi(z) = \frac{1}{\sqrt{2\,\pi}} \int_0^z e^{-\frac{t^2}{2}}\, dt$$



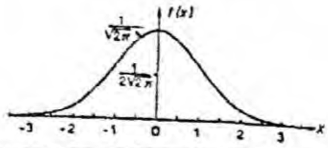
z	00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2703	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998

Download free eBooks at bookboon.com

Table III

Normal Density Function

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.39894	.39892	.39886	.39876	.39862	.39844	.39822	.39797	.39767	.39733
0.1	.39695	.39654	.39608	.39359	.39505	.39448	.39387	.39322	.39253	.39181
0.2	.39104	.39024	.38940	.38853	.38762	.38667	.38568	.38466	.38361	.38251
0.3	.38139	.38023	.37903	.37780	.37654	.37524	.37391	.37255	.37115	.36973
0.4	.36827	.36678	.36526	.36371	.36213	.36053	.35889	.35723	.35553	.35381
0.5	.35207	.35029	.34849	.34667	.34482	.34294	.34105	.33912	.33718	.33521
0.6	.33322	.33121	.32918	.32713	.32506	.32297	.32086	.31874	.31659	.31443
0.7	.31225	.31006	.30785	.30563	.30339	.30114	.29887	.29659	.29431	.29200
0.8	.28969	.28737	.28504	.28269	.28034	.27798	.27562	.27324	.27086	.26848
0;9	.26609	.26369	.26129	.25888	.25647	.25406	.25164	.24923	.24681	.24439
1.0	.24197	.23955	.23713	.23471	.23230	.22988	.22747	.22506	.22265	.22025
1.1	.21785	.21546	.21307	.21069	.20831	.20594	.20357	.20121	.19886	.19652
1.2	.19419	.19186	.18954	.18724	.18494	.18265	.18037	.17810	.17585	.17360
1.3	.17137	.16915	.16694	.16474	.16256	.16038	.15822	.15608	.15395	.15183
1.4	.14973	.14764	.14556	.14350	.14146	.13943	.13742	.13542	.13344	.13147
1.5	.12952	.12758	.12566	.12376	.12188	.12001	.1181[illegible]	.11632	.11450	.11270
1.6	.11092	.10915	.10741	.10567	.10396	.10226	.10059	.09893	.09728	.09566
1.7	.09405	.09246	.09089	.08933	.08780	.08628	.08478	.08329	.08183	.08038
1.8	.07895	.07754	.07614	.07477	.07341	.07206	.07074	.06943	.06814	.06687
1.9	.06562	.06438	.06316	.06195	.06077	.05959	.05844	.05730	.05618	.05508
2.0	.05399	.05292	.05186-	.05082	.04980	.04879	.04780	.04682	.04586	.04491
2.1	.04398	.04307	.04217	.04128	.04041	.03955	.03871	.03788	.03706	.03626
2.2	.03547	.03470	.03394	.03319	.03246	.03174	.03103	.03034	.02965	.02898
2.3	.02833	.02768	.02705	.02643	.02582	.02522	.02463	.62406	.02349	.02294
2.4	.02239	.02186	.02134	.02083	.02033	.01984	.01936	.01888	.01842	.01797
2.5	.01753	.01709	.01667	.01625	.01585	.01545	.01506	.01468	.01431	.01394
2.6	.01358	.01323	.01289	.01256	.01223	.01191	.01160	.01130	.01100	.01071
2.7	.01042	.01014	.00987	.00961	.00935	.00909	.00885	.00861	.00837	.00814
2.8	.00792	.00770	.00748	.00727	.00707	.00687	.00668	.00649	.00631	.00613
2.9	.00595	.00578	.00562	.00545	.00530	.00514	.00499	.00485	.00470	.00457
3.0	.00443	.00430	.00417	.00405	.00393	.00381	.00370	.00358	.00348	.00337
3.1	.00327	.00317	.00307	.00298	.00288	.00279	.00271	.00262	.00254	.00246
3.2	.00238	.00231	.00224	.00216	.00210	.00203	.00196	.00190	.00184	.00178
3.3	.00172	.00167	.00161	.00156	.00151	.00146	.00141	.00136	.00132	.00127
3.4	.00123	.00119	.00115	.00111	.00107	.00104	.00100	.00097	.00094	.00090
3.5	.00087	.00084	.00081	.00079	.00076	.00073	.00071	.00068	.00066	.00063
3.6	.00061	.00059	.00057	.00055	.00053	.00051	.00049	.90047	.00046	.00044
3.7	.00042	.00041	.00039	.00038	.00037	.00035	.00034	.00033	.00031	.00030
3.8	.00029	.00028	.00027	.00026	.00025	.00024	.00023	.00022	.00021	.00021
3.9	.00020	.00019	.00018	.00018	.00017	.00016	.00016	.00015	.00014	.00014
4.0	.00013	.00013	.00012	.00012	.00011	.90011	.00011	.60010	.00010	.00009

Download free eBooks at bookboon.com

Table IV

Percentiles of Chi Square Distribution

$d.f$	.005	.025	.05	.90	.95	.975	.99	.995
1	.0000393	.000982	.00393	2.706	3.841	5.024	6.635	7.879
2	.0100	.0506	.103	4.605	5.991	7.378	9.210	10.597
3	.0717	.216	.352	6.251	7.815	9.348	11.345	12.838
4	.207	.484	.711	7.779	9.488	11.141	13.277	14.860
5	.412	.831	1.145	9.236	11.070	12.8A2	15.086	16.750
6	.676	1.237	1.635	10.645	12.592	14.449	16.812	18.548
7	.989 1	.690	2.167	12.017	14.067	16.013	18.475	20.278
8	1.344	2.180	2.733	13.362	15.507	17.535	20.090	21.955
9	1.735	2.700	3.325	14.684	16.919	19.023	21.666	23.589
10	2.156	3.247	3.940	15.987	18.307	20.483	23.209	25.188
11	2.603	3.816	4.575	17.275	19.675	21.920	24.725	26.757
12	3.074	4.404	5.226	18.549	21.026	23.336	26.217	28.300
13	3.565	5.009	5.892	19.812	22.362	24.736	27.688	29.819
14	4.075	5.629	6.571	21.064	23.685	26.119	29.141	31.319
15	4.601	6.262	7.261	22.307	24.996	27.488	30.578	32.801
16	5.142	6.908	7.962	23.542	26.296	28.845	32.000	34.267
17	5.697	7.564	8.672	24.769	27.587	30.191	33.409	35.718
18	6.265	8.231	9.390	25.989	28.869	31.526	34.805	37.156
19	6.844	8.907	10.117	27.204	30.144	32.852	36.191	38.532
20	7.434	9.591	10.851	28.412	31.410	34.170	37.566	39.997
21	8.034	10.283	11.591	29.615	32.671	35.479	38.932	41.401
22	8.643	10.982	12.338	30.813	33.924	36.781	40.289	42.796
23	9.260	11.688	13.091	32.007	35.172	8.076	41.638	4.181
24	9.886	12.401	13.848	33.196	36.415	39.364	42.980	45.558
25	10.520	13.120	14.611	34.382	37.652	40.646	44.314	46.928
26	11.160	13.844	15.379	35.563	38.885	41.923	45.642	48.290
27	11.808	14.573	16.151	36.741	40.113	43.194	46.963	49.645
28	12.461	15.308	16.928	37.916	41.337	44.461	48.278	50.993
29	13.121	16.047	17.708	39.087	42.557	45.722	49.588	52.336
30	13.787	16.791	18.493	40.256	43.773	46.979	50.892	53.672
35	17.192	26.569	22.465	46.059	49.802	53.203	57.342	60.275
40	20.707	24.433	26.509	51.805	55.758	59.342	63.691	66.766
45	24.311	28.366	30.612	57.505	61.656	65.410	69.957	73.166
50	27.991	32.357	34.764	63.167	67.505	71.420	76.154	79.490
60	35.535	40.482	43.188	74.397	79.082	83.298	88.379	91.952
70	43.275	48.758	51.739	85.527	90.531	95.023	100.425	104.215
80	51.172	57.153	60.391	96.578	101.879	106.629	112.329	116.321
90	59.196	65.647	69.126	107.565	113.145	118.136	124.116	28.299
100	67.328	74.222	77.929	118.498	124.342	129.561	135.807	40.169

Download free eBooks at bookboon.com

Table V

Percentiles of t Distribution

$d.f$	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$	$t_{0.999}$	$t_{0.9995}$
1	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.183	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.500	4.785	5.408
8	1.397	1.860	2.306	2.896	3.335	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.813	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.132	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
35	1.306	1.690	2.030	2.438	2.724	3.340	3.591
40	1.303	1.684	2.021	2.423	2.705	3.307	3.551
45	1.301	1.679	2.014	2.412	2.690	3.282	3.520
50	1.299	1.676	2.009	2.403	2.678	3.261	3.496
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
70	1.294	1.667	1.995	2.381	2.648	3.211	3.435
80	1.292	1.664	1.990	2.374	2.639	3.195	3.416
90	1.291	1.662	1.987	2.368	2.632	3.183	3.402
100	1.290	1.660	1.984	2.364	2.626	3.183	3.402
120	1.289	1.658	1.980	2.358	2.618	3.174	3.391
140	1.288	1.656	1.977	2.353	2.611	3.160	3.373
160	1.287	1.654	1.975	2.350	2.607	3.142	3.352
∞	1.282	1.646	1.962	2.330	2.581	3.098	3.300

Source: This table is adapted from *The Biometrika Tables for Statisticians*, Vol. 1, 3rd ed., 1966 by E.S. Pearson and H.O. Hartley, with permission of the Biometrika Trustees.

Download free eBooks at bookboon.com

Table VI

Percentiles of F Distribution
$F_{0.90}$

v_2	Numerator Degrees of Freedom (v_1) 1	2	3	4	5	6	7	8	9
1	39.86	49.50	53.59	55.83	57.24	58.20	56.91	59.44	59.86
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
4	4.54	4.32	4.19	4.11	4.05	4.01	3.96	3.95	3.94
5	4.06	3.76	3.62	3.52	3.45	3.40	3.37	3.34	3.32
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.67
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63

Download free eBooks at bookboon.com

Percentiles of F Distribution (*continued*)

v_2	Numerator Degrees of Freedom (v_1)									
	10	12	15	20	24	30	40	60	120	∞
1	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
2	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3	5.23	5.22	5.22	5.18	5.18	5.17	5.16	5.15	5.14	5.13
4	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
5	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
6	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
8	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10	2.32	2.28	2.24	2.26	2.18	2.16	2.43	2.11	2.08	2.06
11	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
13	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
15	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
16	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.92
17	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18	1.98	1.93	1.89	1.84	1.81	1.78	1 75	1.72	1.69	1.66
19	1.96	1.91	1.86	1.81	1.79	1.76	1 73	1.70	1.67	1.63
20	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1 59
22	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
24	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
27	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
28	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
29	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
30	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
40	1.76	1.71	1.66	1.61	1.57	1 54	1.51	1.47	1.42	1.38
60	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
120	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00

Source: This table is adapted from *The Biometrika Tables for Statisticians*, Vol. 1, 3rd ed., 1966 by E.S. Pearson and H.O. Hartley, with permission of the Biometrika Trustees.

Percentiles of F Distribution

$F_{0.95}$

v_2	Numerator Degrees of Freedom (v_1) 1	2	3	4	5	6	7	8	9
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
2	18.51	19.00	19.6	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	3.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.69	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

Download free eBooks at bookboon.com

Percentiles of F Distribution (*continued*)

v_2	Numerator Degrees of Freedom (v_1) 10	12	15	20	24	30	40	60	120	∞
1	241.9	243.9	243.9	248.0	249.1	230.1	251.1	252.2	253.3	254.3
2	19.40	19.41	19.43	19.45	19.45	19.46	19.47	9.48	19.49	19.50
3	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.96	5.91	5.86	5.30	5.77	5.75	5.72	5.69	5.66	5.63
5	4.74	4.68	4.56	4.53	5.40	4.46	4.43	4.40	4.36	
6	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67	
7	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.98	2.91	1.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.75	2.69	2.62	2.54	2.59	2.47	2.43	2.38	2.34	2.13
13	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.37	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.25	2.18	2.11	2.03	2.98	1.89	1.89	1.84	1.79	1.73
25	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Source: This table is adapted from *The Biometrika Tables for Statisticians*, Vol. 1, 3rd ed., 1966 by E.S. Pearson and H.O. Hartley, with permission of the Biometrika Trustees.

Download free eBooks at bookboon.com

Percentiles of F Distribution

$F_{0.99}$

v_2	Numerator Degrees of Freedom (v_1)								
	1	2	3	4	5	6	7	8	9
1	4052	4999.5	5403	5625	5764	5859	5928	5981	6022
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
4	21.20	18.00	61.69	15.98	15.52	15.21	14.98	14.80	14.66
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
9	10.56	8.02	6.99	6.42	6.06	5.86	5.61	5.47	5.35
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
12	9.13	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
26	7.72	5.51	4.64	4.14	3.82	3.59	3.42	3.29	3.18
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
30	7.56	5.39	4.51	4.02	3.10	3.47	3.30	3.17	3.07
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41

Download free eBooks at bookboon.com

Percentiles of F Distribution (*continued*)

v_2	Numerator Degrees of Freedom (v_1)									
	10	12	15	20	24	30	40	60	120	∞
1	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	27.23	27.05	26.87	26.69	26.00	26.50	26.41	26.32	26.22	26.13
4	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.59
20	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
31	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
32	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
33	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
34	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

Source: This table is adapted from *The Biometrika Tables for Statisticians*, Vol. 1, 3rd ed., 1966 by E.S. Pearson and H.O. Hartley, with permission of the Biometrika Trustees.

Table VII

Random Numbers

53872	34774	19087	81775	71440	12082	75092	34608	75448	13148
04226	62404	71577	00984	56056	32404	87641	53392	92561	33388
28666	44190	75524	62038	21423	46281	92238	96306	72606	80601
63817	30279	14088	86434	16183	06401	90586	80292	54555	47371
22359	16442	83879	47486	19838	32252	39560	95851	36758	36141
50968	28728	83525	16031	77583	65578	84794	51367	32535	83834
39652	24248	96617	91200	10769	52386	39559	75921	49.375	22847
35493	00529	69632	29684	80284	87828	72418	80950	86311	34016
75687	53919	80439	20534	96185	72345	96391	52625	50866	45132
31509	93521	10681	44124	88345	84969	88768	48819	22311	41235
40389	76282	37506	60661	23295	67357	95419	10864	87833	09152
59244	54664	63424	97899	44153	69251	08781	18604	02312	21658
99876	17075	40934	08912	96196	58503	63613	24486	98092	45672
06457	50072	18060	71023	84349	40984	59487	77782	32107	53770
14297	07687	05517	10362	35783	62236	63764	45542	68889	03862
51661	57130	97441	29590	21634	79772	73801	70122	46467	47152
53455	41788	16117	09698	24409	05079	76603	57563	33461	46791
48086	31512	62819	27689	63744	11023	11184	87679	22218	70139
19108	01602	96950	41536	39974	88287	83546	69187	45539	78263
39001	77727	33095	58785	29179	45421	71416	20418	38558	78700
72346	55617	14714	21930	14851	38209	52202	03979	05970	74483
19094	64359	89829	10942	53101	37758	29583	26792	42840	45872
82247	77127	01652	50774	04970	83300	33760	22172	67516	62135
75968	18386	31874	52249	21015	20365	57475	32756	58268	75739
01963	38095	99960	91307	99654	74279	80145	53303	11870	50485
64828	15817	80923	55226	51893	93362	15757	47430	84855	95822
64347	61578	44160	06266	35118	52558	56436	96155	10293	67506
54746	52337	84826	39012	59118	19851	10156	78167	41473	99025
22241	41501	02993	99340	91044	67268	51088	12751	74008	33773
11906	20043	10415	44425	31712	54831	85591	62237	88797	14382
76637	07609	95378	95580	86909	50609	99008	99042	50364	36664
93896	47120	98926	30636	28136	49458	84145	79205	79517	93446
75292	88232	14360	12455	13656	65736	70428	66917	64412	38502
98792	29828	10577	48184	29433	98278	22543	76155	82107	22066
65751	91049	94127	47558	99880	79667	86254	72797	67117	44699
72064	62102	39155	79462	82975	02638	00302	79476	72656	84003
01227	35821	80607	61734	02600	45564	72344	71034	48370	96826
44768	56504	13993	59701	88238	92483	09497	66058	36651	37927
69838	91226	85736	72247	64099	86305	49877	76215	66980	30228
01800	39313	57730	84410	47637	81369	51830	43530	58937	91901

Download free eBooks at bookboon.com

Random Numbers *(Continued)*

98892	53633	33909	81674	91956	84531	60422	55574	31670	61059
95398	77381	21912	24873	26372	12044	43234	08503	86716	08095
28982	24589	88896	31137	87512	33216	29665	26014	02919	17619
31303	70209	42174	10757	98531	35725	68208	61239	26705	41916
08457	10085	35741	79416	72457	59502	46986	09051	70963	19759
30698	80818	90073	78320	83675	78361	49929	70495	92247	04118
27142	41186	52273	81087	67396	16795	98542	83820	48765	24164
36775	63628	70856	43164	88426	51415	37514	24870	55665	05111
02560	51679	79600	23297	36434	17174	00109	02731	05909	58959
36744	66697	08331	50201	56303	09171	55995	60232	31305	30689
66482	04302	29770	46201	04588	42575	99318	84406	83405	21186
76375	41539	65940	57820	29283	94564	96598	00619	60468	97175
95772	72925	19454	63712	21401	96665	77750	21218	02990	50796
59013	81632	85000	39180	99975	73253	46534	59083	60243	27664
52392	04440	45628	34976	92012	16596	28596	15493	80754	48760
08027	07629	04339	77570	47155	77128	24498	67455	06320	82004
27284	39416	57313	99508	71443	42543	73335	68620	87559	77927
20513	38581	82309	69951	82658	60958	18290	60534	30741	89647
36076	12821	68723	37934	62818	64157	54590	98263	70109	06755
60679	43862	43675	03653	21060	81096	71332	28930	44207	08354
49416	58370	63738	87515	39290	87656	36130	23490	30963	57350
65757	39149	11780	92494	41335	35835	69882	56431	08091	01981
17379	77731	65131	44979	90939	29184	76634	58007	34873	81816
00757	13129	09648	07644	81689	68088	34882	04971	27565	66577
68276	79035	78273	83412	97128	81003	65938	85510	78367	29116
64716	91696	45448	92281	73854	67452	52145	41582	81549	82414
83695	11496	57066	48153	74754	56383	09253	65456	32438	96351
58275	66797	35380	41155	44389	94860	42074	31178	27967	12666
58005	84170	29999	23631	93032	41592	55688	78599	59902	21568
99993	80083	08810	07244	42067	76669	19686	64064	67141	80520
31692	51607	89056	74472	91284	20263	16039	94491	33767	71915
82997	58320	04852	52595	95514	56543	06636	61291	67504	57205
05043	40582	46051	60261	04996	82256	47375	87507	05112	88489
75781	38768	70475	00601	18378	32077	36523	30843	07057	78126
21033	15175	30741	45814	92222	16704	00197	51267	33224	40276
99092	60991	12571	71753	65214	33885	82939	50723	88987	69761
07204	93373	85112	29610	30375	64836	18459	08235	67650	72930
88859	97254	07771	21393	64657	42013	12753	03028	24224	24918
39497	91407	72900	15699	58653	38063	25072	48698	88083	48040
09726	18075	45852	54968	43743	82050	78412	79456	95032	10984

Random Numbers *(Continued)*

11756	45441	59948	57975	92422	70057	50210	30345	55912	31638
39056	86614	53643	62909	27198	04454	33789	86463	66603	48083
88086	93172	68311	39164	42012	10447	45933	28844	36844	57684
12648	27948	76750	19915	66815	34015	43011	27150	94264	89516
16254	87661	66181	68609	58626	58428	75051	27558	49463	66646
69682	19109	94189	94626	09299	10649	55405	54571	57855	54921
61336	86663	13010	40412	50139	30769	13048	61407	41056	60510
65727	66488	12304	70011	93324	58764	87274	43103	96002	06984
55705	34418	99401	32635	42984	40981	91750	27431	05142	77950
95402	51746	98184	38830	97590	00066	82770	42325	28778	83571
79228	94510	57711	64366	89040	43278	69072	22003	89465	61483
48103	56760	82564	33649	35176	32278	51357	05489	47462	55931
70969	27677	99621	63065,	73194	70462	19316	77945	45004	39895
69931	20237	75246	59124	12484	22012	79731	82435	56301	99752
37208	22741	41946	74109	03760	24094	40210	76617	52317	50643
60151	92327	85150	27728	64813	47667	66078	03628	95240	03808
46210	47674	53747	95354	67757	75477	26396	09592	96239	50854
55399	48142	12284	95298	56399	61358	87541	12998	79639	63633
23677	64950	97041	43088	80143	34294	91468	01066	90350	78891
41947	70066	90311	17133	11674	00826	75760	37586	33621	14199
16972	42181	87945	94104	95701	00743	75411	51930	54869	98991
74938	79042	38473	89672	45752	35715	89537	78155	09851	24983
78075	53671	81047	92759	94519	59473	91679	90536	41676	35230
76744	26190	21649	79753	21287	17698	39490	00533	34823	08134
82273	69293	23383	59365	18258	54530	47274	69686	55081	28731
30239	23081	09526	26055	87099	41372	55542	32754	87317	94638
41177	77163	38252	10349	49511	17540	61781	32769	51662	55606
07715	88600	69730	78912	19642	39764	47146	19472	84012	08887
16855	47454	98638	15189	87345	80509	33392	50866	17629	28208
27985	61979	02979	98092	41184	73815	57939	91057	04860	66667
77411	98433	42302	86602	26596	64175	64359	97570	64437	55592
19453	18731	01039	18933	92188	83767	56148	56261	79920	78514
03381	35119	30355	08287	00448	32800	24106	04054	70572	71063
11659	27315	09204	26213	57325	51470	56108	23141	16121	53925
35032	14283	20642	15311	36238	12070	67596	00017	51789	90737
32061	51250	39825	08554	88716	40945	68579	33784	62025	32535
81855	16888	24630	15077	47256	08529	54837	24161	95621	53483
48422	09247	43406	16093	01168	28523	31406	49360	99243	85090
86190	56195	31409	88248	52436	70161	98500	74702	99546	74570
90627	37048	50285	69189	97489	83007	31477	13908	97472	74448

Download free eBooks at bookboon.com

Random Numbers *(Continued)*

60103	76739	57644	56746	63005	08804	47081	65928	65045	58629
09606	69465	16536	94055	86328	56533	16670	57295	26249	18524
62479	29610	03235	51050	15855	66828	08115	16166	32854	74206
40232	52840	02512	99258	09327	55073	86030	29933	00528	67359
10690	55550	81275	78369	33658	47000	89425	60573	81137	25474
73958	38949	99568	72713	22665	03244	17399	83950	66820	08701
56554	57926	41529	00619	51972	09442	60298	81066	28362	11165
35676	20333	77622	93718	57255	09780	26798	60083	58959	15691
01383	85677	96572	16401	31379	88519	41325	33938	36312	03327
29448	88487	05814	82402	42132	85708	49754	57495	57655	78614
56863	94737	68661	43498	33376	81659	07422	58435	24855	15523
20269	34456	48608	11787	86056	88290	17463	66628	03033	80771
06790	99803	86439	94235	48560	62912	82302	43198	97087	97104
73690	79726	06492	77431	49864	69775	46450	02122	09083	92746
76222	20006	98660	88690	01190	05588	76651	03461	11987	80756
18434	21893	80472	19499	80423	58643	27088	66458	78358	56606
20463	75133	40713	84279	56045	79079	20212	91560	60548	95128
27105	77095	72016	23683	01386	40381	74673	11811	36625	62958
47736	56338	07546	36084	73126	33364	78730	47282	76795	95719
60938	13970	90288	79457	50343	92054	12541	93216	58624	37392
02743	59982	92806	62853	39755	42550	31081	38860	35712	78632
74802	59354	91213	26293	18112	93831	01473	10798	18229	18642
06933	78651	45636	77509	28610	34307	68045	15107	62935	34149
40345	80092	50587	18535	19001	82179	12572	77589	33459	35130
70055	98685	10244	11760	21952	73985	68903	66934	42442	07608
34552	76373	40928	93696	97711	15818	31004	03263	05626	07460
45253	86947	42417	28778	14936	94099	90775	42001	86675	62770
71558	21692	84077	17814	33316	49494	31817	90127	39485	92302
95474	76468	12019	04274	01893	23930	88771	31142	65859	28948
34619	91898	28499	00279	35351	87736	83909	43736	19258	95068
44546	75524	68535	77434	18543	15479	58850	73802	01636	82735
22917	96024	04784	05809	52788	83577	02269	68632	23310	46261
33043	31433	47833	75234	74539	38529	57893	45997	71749	28666
99357	54593	21688	64216	85938	51742	12898	09737	61504	18946
01072	31679	80961	34029	56463	09594	11939	51777	64796	52452
90838	50179	42064	62987	13072	84227	24060	59438	05695	38136
35914	39441	90149	67957	16955	39960	26142	45600	75486	74103
87047	77284	12753	45644	47843	55781	06672	57548	84706	25453
93727	46613	48045	49685	28385	37200	98473	56808	86774	07305
37439	50362	44171	18495	57370	77691	28006	55318	39723	25299

Download free eBooks at bookboon.com

Random Numbers *(Continued)*

95330	01985	24128	60514	42539	91907	25694	37097	39566	24043
09760	32388	05601	49923	66126	54146	61213	52234	48381	89442
01534	81967	15337	95831	84643	40792	47562	95494	62087	18064
11234	59350	48368	57195	36287	03046	87136	36057	93913	70080
71056	48762	80221	59683	27504	21121	94711	11807	80882	48359
34208	05374	60104	43178	97247	24875	26259	67622	14657	80354
47132	62839	82198	92445	60650	76219	02772	48651	66449	89213
55685	93302	43019	45861	83008	20544	41665	99439	70606	28974
55045	17219	66737	59080	78489	12626	60661	53733	70062	14289
01923	33647	98442	59293	83318	33425	76412	87062	01295	11083
07202	76476	71888	54845	17468	41964	68691	59662	55905	26898
68825	68242	95750	11033	58634	78411	08523	19313	29327	47526
68525	06496	17446	41378	32368	82019	66101	56733	43308	82641
80819	33515	97373	43064	16221	99697	3795	07947	12935	49391
64209	96929	26044	49283	56545	67200	21325	85056	51345	06309
30156	29121	75874	42399	41121	90643	19585	06364	47203	19679
50467	14282	89098	66717	14753	73356	47781	34165	82842	00121
53764	83212	26675	64184	64455	29023	03181	13674	08838	83829
81727	35572	95469	36825	81882	95083	68323	14965	34166	32351
30807	55558	96026	97398	21723	86560	52617	07771	6188	48234
75104	23682	78756	72728	85940	57290	75507	78715	01426	02310
06180	62724	36835	80288	25075	32609	33312	21348	87710	55457
22098	34834	66117	36252	82717	50585	43639	79999	07414	84003
13173	64783	20984	11929	18849	76211	77375	49561	96747	67007
75273	36108	55265	15653	82270	99216	27805	60088	06056	97377
89849	65756	44454	04602	14292	74458	57777	35934	05160	26359
91108	43562	18883	16569	49599	73871	67101	12054	56492	15981
51843	01542	17881	12954	94913	39583	94969	61146	35907	72184
02644	23564	85464	62941	92571	89377	85004	84654	20465	86212
38608	83374	74032	62183	08740	05279	30455	31032	71512	16476
43164	28909	88624	14992	85359	10193	32491	14769	63694	92640
80933	52950	45646	36636	05085	28053	27596	54873	68476	65823
67690	96766	69250	19344	47855	43489	77479	62418	54079	40069
68579	17014	25362	15114	30982	27250	29052	71115	83369	46776
46353	39733	44677	50133	26623	15979	10651	04263	34087	67005
30039	09532	52215	09164	20930	88230	43403	63230	83525	93550
89200	92772	42195	91634	39272	46462	76835	27755	03151	75692
58118	57942	14807	68214	76093	47484	24468	91764	52907	16675
97230	33027	70166	43232	98802	70715	30216	35586	18909	79658

Source: Reprinted from pages of *A Million Random Digits with 100,000 Normal Deviates* by RAND Corporation (New York: The Free Press, 1955). ©1955 and 1983 by the RAND Corporation. Used by permission.

Download free eBooks at bookboon.com

Answers to Odd-Numbered Exercises

Chapter 1

1.1

X	Y		
	1	2	3
1	$\frac{3}{34}$	$\frac{5}{34}$	$\frac{1}{34}$
2	$\frac{2}{34}$	$\frac{4}{34}$	$\frac{6}{34}$
3	$\frac{1}{34}$	$\frac{2}{34}$	0
4	$\frac{4}{34}$	$\frac{1}{34}$	$\frac{5}{34}$

1.3 (a) $\frac{9}{34}$ (c) $\frac{7}{34}$ 1.5 $\frac{21}{21} = 1$ 1.7 (c) $\frac{1}{9}(2x+1)$ 1.9 (a) $\frac{e^{-\lambda p}(\lambda p)^x}{x!},\ x = 0, 1, 2, \cdots$ 1.11 (c) (i) $2(x+y-2xy),\ 0 \le x \le 1,\ 0 \le y \le 1$ 1.13 $\frac{12}{7}$ (c) $\frac{12}{7}\left(x^2 + \frac{x}{2}\right),\ 0 \le x \le 1$ 1.15 (a) $\lambda e^{-\lambda x},\ x \ge 0$ (c) $\lambda e^{-\lambda(y-x)},\ y \ge x$ 1.17 (a) (i)$x\,e^{-x},\ y \ge 0$

Chapter 2

2.1 (a)

$X+Y$	2	3	4	5	6	7
$P(X+Y=k)$	$\frac{3}{34}$	$\frac{7}{34}$	$\frac{6}{34}$	$\frac{12}{34}$	$\frac{1}{34}$	$\frac{5}{34}$

(c)

XY	1	2	3	4	6	8	9	12
$P(XY=k)$	$\frac{3}{34}$	$\frac{7}{34}$	$\frac{2}{34}$	$\frac{8}{34}$	$\frac{8}{34}$	$\frac{1}{34}$	0	$\frac{5}{34}$

(e)

$2X+3Y$	5	7	8	9	10	11	12	13	14	15	17
$P(2X+3Y=k)$	$\frac{3}{34}$	$\frac{2}{34}$	$\frac{5}{34}$	$\frac{1}{34}$	$\frac{4}{34}$	$\frac{5}{34}$	$\frac{2}{34}$	$\frac{6}{34}$	$\frac{1}{34}$	0	$\frac{5}{34}$

(g)

$3(X)4(Y) = 12XY$	12	24	36	48	72	96	108	144
$P(12XY=k)$	$\frac{3}{34}$	$\frac{7}{34}$	$\frac{2}{34}$	$\frac{8}{34}$	$\frac{8}{34}$	$\frac{1}{34}$	0	$\frac{5}{34}$

(j)

$3X$	3	6	9	12
$P(3X=k)$	$\frac{9}{34}$	$\frac{12}{34}$	$\frac{3}{34}$	$\frac{10}{34}$

2.3

$X+Y$	2	3	4
$P(X+Y)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{1}{2}$

2.5(a) $\frac{\lambda^z e^{-\lambda}}{z!}, \quad z = 0, 1, 2, \cdots$ 2.7(a) $s(z) = \begin{cases} \frac{2}{3}z^2(3-z), & 0 \le z \le 1 \\ \frac{2}{3}z^2(4-3z^2+z^3), & 1 \le z \le 2 \end{cases}$

Download free eBooks at bookboon.com

2.9(a) $s(z) = \begin{cases} \dfrac{5}{14}z^3, & 0 < z \le 1 \\ \dfrac{3}{7}\left(\dfrac{z}{2} + \dfrac{1}{3}\right), & 1 < z \le 2 \\ \dfrac{3}{7}\left(\dfrac{5}{6}z^3 - 4z^2 + \dfrac{11}{2}z + 36\right), & 2 < z \le 3 \end{cases}$

2.17 6.027×10^{-5} 2.19 (a) $h(u) = \begin{cases} \frac{2}{3}(u^3+1), & -1 \le u < 0 \\ \frac{2}{3}(1-u^3), & 0 \le u \le 1 \end{cases}$

2.21 $4[1 - u(1 - \ln u)], \quad 0 \le u \le 1$ 2.23

$\frac{X}{Y}$	1	$\frac{3}{2}$	2	3
$P(\frac{X}{Y} = k)$	$\frac{2}{20}$	$\frac{7}{20}$	$\frac{3}{20}$	$\frac{8}{20}$

2.25 $h(u) = \begin{cases} 1 + \frac{2}{3}u - \frac{u^2}{3}, & -1 \le u < 0 \\ 1 - \frac{4}{3}u + \frac{u^2}{3}, & 0 \le u \le 1 \end{cases}$ 2.27 $h(u) = \begin{cases} \frac{2}{9}(u+2), & 0 \le u < 1 \\ \frac{2}{9u^3}(u+2), & 1 \le u < \infty \end{cases}$

2.29 $\frac{3}{7}(1 - \frac{u^2}{4} + u\ln(\frac{2}{u})), 0 < u < 2$

Chapter 3

3.1(a) 4.47 3.3 (a) 2.01 (c) 10.39 (e) 11.24 (g) 1496.32 (i) 52.80 (k)2.11 (m) 3.21 3.5(a)$\frac{2}{5}$ (c) $\frac{2}{15}$ (e) 0.04 (g) 0 (i) 0.08 3.15 6.46

Chapter 4

4.1 0.064 4.3 (a) 2.3 (c) 2.68 (e) 1.2 4.5 0 4.9 -0.1832 4.11 $-\frac{1}{11}$ 4.13(a)(i) $-0.21x + 1.52$ (ii) $-0.16y + 1.63$ 4.15 $\frac{24x^2+24x+2}{9(2x+1)^2}$ 4.19 $0.05y + 2.68$

Chapter 5

5.1(a) $\frac{1}{2}$ 5.3 $n \le 250$ 5.5(a) 0.85 5.7 0.84 5.9 1.3947 5.11 0.9648 5.13(a) 0.3409 (c) 0.8185 5.15 0.0062 5.17 1600 5.19 0.9927 5.21 $n \approx 25$ 5.23 $n \approx 85,\ \epsilon = 0.43$ 5.25(a) 0.1852 5.29 0.2328

Chapter 7

7.1(a) 3.66.8 (c) $E(X) = 6.8$, Var $= 8.48$ 7.3(a) 6.8 7.5(a)(i) 140 (ii) 1.49 0.85 7.7(a)(i) 1,000 (ii) 7.9 $s.d = 0.657$ 7.11(a)(i) 0.56 (ii) 0.4855 7.15(a) 0.004 7.17(a)(ii) 458.8

Chapter 8

8.3 12 8.5(a) $E(X) = 12$, Var $= 24$ 8.7(a) 22.367 (c) 10.285 8.9(a) 1.701 (c) 1.321 8.11 1.93

Download free eBooks at bookboon.com

Bibliography

Applebaum, D., (1996), *Probability and Information: An Integrated Approach ,* Cambridge University Press, Cambridge.

Bajpal, A.C., Calus, L.M. and Fairley, J.A., (1978), *Statistical Methods for Engineers and Scientists,* John Wiley and Sons, New York.

Barlow, R. (1989), *Statistics, A Guide to the Use of Statistical Methods in the Physical Sciences,* John Wiley and Sons, New York.

Bartoszyński, R and Niewiadomska-Bugay, M (1996), *Probability and Statistical Inference,* John Wiley and Sons, New York.

Beaumont, G.P. (1986), *Probability and Random Variables,* Ellis Norwood Ltd, Chichester.

Berger, M.A. (1993) *An Introduction to Probability and Stochastic Processes,* Springer-Verlag, New York.

Billinsley, P. (1979) *Probability and Measure*, John Wiley and Sons, New York.

Birnbaum, Z.W. (1962), *Introduction to Probability and Mathematical Statistics,* Harper and Brothers Publishers, New York.

Blake, Ian F.B. (1979), *An Introduction to Applied Probability,* John Wiley and Sons, New York.

Bradley, J., (1988), *Introduction to Discrete Mathematics, Addison-Wesley Publishing Company,* Reading.

Breiman L., (1972), *Society for Industrial and Applied Maths,* Siam, Philadelphia.

Brémaud, Pierre (1994), *An Introduction to Probabilistic Modeling,* Springer - Verlag, New York.

Chow, Y.S., Teicher, H. (1988), *Probability Theory*, Springer-Verlag, New York.

David F.N. (1951), *Probability Theory for Statistical Methods*, Cambridge University Press, Cambridge.

Drake, A.W. (1967), *Fundamentals of Applied Probability Theory,* McGraw-Hill, New York.

Dudewicz, E.J. and Mishra, S.N. (1988), *Modern Mathematical Statistics,* John Wiley and Sons, New York.

Feller, W. (1970) *An Introduction to Probability Theory and its Applications* Vol. I, Second Edition, John Wiley and Sons, New York.

Feller, W. (1971) *An Introduction to Probability Theory and its Applications* Vol. II, Second Edition, John Wiley and Sons, New York.

Freund, J.E.and Walpole, R.E. (1971), *Mathematical Statistics,* Prentice - Hall, Inc, Englewood Cliffs, New Jersey.

Fristedt, B. Gray, L. (1997), *Modern Approach to Probability Theory*, Birhhäuser, Boston.

Gnedenko, B.V. and Khinchin, A.Ya. (1962), *An Elementary Introduction to the Theory of Probability,* Dover Publications, Inc., New York.

Goldberg, S. (1960), *Probability, An Introduction*, Prentice-Hall, Inc., Englewood Cliffe, New Jersey.

Grimmeth, G. and Weslsh, D. (1986), *Probability, An Introduction*, Clarendon Press, Oxford.

Guttman, I., Wilks, S.S and Hunter, J.S. (1982), *Introduction Engineering Statistics,* John Wiley and Sons, New York.

Haln, G.J., and Shapiro, S.S. (1967) *Statistical Models in Engineering*, John WIley and Sons.

Download free eBooks at bookboon.com

Hoel, P.G. (1984), *Mathematical Statistics*, Fifth ed., John Wiley and Sons, New York.

Hoel, P.G., Port, S.C, and Stone, J.S. (1971), *Introduction to Probability Theory*, Houghton Mifflin.

Hogg, R.V. and Cragg, A.T. (1978), *Introduction to Mathematical Statistics*, Fourth Edition, Macmillan, New York.

Johnson, N.L. and Kotz, S.(1969) *Distributions in Statistics: Discrete Distributions* , John Wiley and Sons, New York.

Johnson, N.L. and Kotz, S.(1970) *Distributions in Statistics: Continuous Multivariate Distributions*, John Wiley and Sons, New York.

Johnson, N.L. and Kotz, S.(1970) *Distributions in Statistics: Continuous Univariate Distributions 1 & 2*, John Wiley and Sons, New York.

Kendal, M.G. and Stuart, A. (1963) *The Advanced Theory of Statistics,*, Vol. 1, 2nd ed., Charles Griffin & Company Ltd., London.

Kendal, M.G. and Stuart, A. (1976) *The Advanced Theory of Statistics,*, Vol. 3, 4th ed., Charles Griffin & Company Ltd., London.

Kendal, M.G. and Stuart, A. (1979) *The Advanced Theory of Statistics,*, Vol. 2, 4th ed., Charles Griffin & Company Ltd., London.

Kolmogorov, A.N. (1956), *Foundations of the Theory of Probability,* Chelsea Publishing Company, New York.

Lai, Chung Kai (1975), *Elementary Probability: Theory with Stochastic Processes,* Springer - Verlag, New York.

Lamperti, J. (1966), *Probability: A Survey of the Mathematical Theory*, W.A. Benjamin, INC., New York.

Lindgren, B.W. (1968) *Statistical Theory,* McMillan Company, New York.

Lupton R., (1993), *Statistics in Theory and Practice*, Princeton, Princeton University Press.

Mendenhall, W. and Scheaffer, R.L. (1990), *Mathematical Statistics With Applications,* PWS-KENT Publishing Company, Mass.

Meyer, D. (1970), *Probability*, W. A. Benjamin, Inc. New York.

Meyer, P.L. (1970), Introductory Probability and Its Applications, 2nd Edition, Addison Wesley Publishing Company.

Miller, I. and Freund, J.E. (1987), *Probability and Statistics for Engineers,* Third Edition, Prentice-Hall, Englewood Cliffs, New Delhi.

Mood, A.M., Graybill, F.A. and Boes, D.C (1974), *Introduction to the Theorey of Statistics*, Third Edition, McGraw-Hill, New York.

Nsowah-Nuamah, N.N.N., (2017), *A First Course in Probability Theory, Vol. I: Introductory Probability Theory and Random Variables*, Ghana Universities Press, Accra, Ghana.

Nsowah-Nuamah, N.N.N., (2018), *A First Course in Probability Theory, Vol. II: Theoretical Probability Distributions*, Ghana Universities Press, Accra, Ghana.

Page, L.B. (1989), *Probability for Engineering with Applications to Reliability,* Computer Science Press.

Prohorov, Yu.V. and Rozanov, Yu.A. (1969) *Probability Theory,* Springer-Verlag, New York.

Robinson, E.A. (1985), *Probability Theory and Applications* International Human Resources Dev. Corporation.

Ross, S. (1984), *A First Course in Probability*, 2nd Edition, Macmillan Publishing Company.

Download free eBooks at bookboon.com

Spiegel, M.R (1980), *Probability and Statistics,* McGraw-Hill Book Company, New York.

Stirzaker, D.(1994), *Elementaty Probability*, Cambridge University Press, Cambridge.

Stoyanov. J, et. al. (1989), *Exercise Manual in Probability Theory,* Kluwer Academic Publishers, London.

Studies in the History of Statistics and Probability, Vol. 1, edited by Pearson, E.S., and Kendal, M. (1970), Charles Griffin & Company Ltd., London.

Studies in the History of Statistics and Probability, Vol. 1, edited by Pearson, E.S., and Kendal, M. (1977), Charles Griffin & Company Ltd., London.

Trivedi, K.S. (1988), *Probability and Statistics with Reliability, Queuing, and Computer Science Applications*, Prentice-Hall of India, New Delhi.

Tucker, H.G. (1963), *An Introduction to Probability and Mathematical Statistics*, Academic Press, New York.

Wayne, W. D. (1991), *Biostatistics: A Foundation for Analysis in the Health Sciences,* Fifth Edition, John Wiley and Sons, Singapore.

Wilks, S.S. (1962), *Mathematical Statistics,* John Wiley and Sons, New York.

Download free eBooks at bookboon.com