CHAPTER 3   PROBABILITY DISTRIBUTIONS

Objectives: After working through this chapter, you should be able to:

(i)   understand basic concepts of probability distributions, such as random variables and mathematical expectations;

(ii)   show how the Normal probability density function may be used to represent certain types of continuous phenomena;

(iii)   demonstrate how certain types of discrete data can be represented by particular kinds of mathematical models, for instance, the Binomial and Poisson probability distributions.

3.1   Introduction to Probability Distributions

3.1.1   Random Variables

A <u>random variable</u> (R.V.) is a variable that takes on different numerical values determined by the outcome of a random experiment.

Example 1

An experiment of tossing a coin 4 times.

Notation :   Capital letter, X - Random variable
              Lowercase, x   - a possible value of X

A random variable is <u>discrete</u> if it can take on only a limited number of values.

A random variable is <u>continuous</u> if it can take any value in an interval.

The <u>probability distribution</u> of a random variable is a representation of the probabilities for all the possible outcomes. This representation might be algebraic, graphical or tabular.

A table or a formula listing all possible values that a discrete variable can take on, together with the associated probability is called a discrete probability distribution.

Example 2

The probability distribution of the number of heads when a coin is tossed 4 times.

x	0	1	2	3	4
Pr(X = x)	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

i.e. $\qquad \Pr(X = x) = \dfrac{\dbinom{4}{x}}{16}$, $\qquad$ x = 0, 1, 2, 3, 4

In graphic form :

1.   Total area of rectangle = 1
2.   Pr(X = 1) = shaded area

Example 3

An experiment of tossing two fair dice.

Let random variable X be the sum of two dice.

<div align="center">The probability distribution of X</div>

Sum, x	2	3	4	5	6	7	8	9	10	11	12
P(X = x)	$\dfrac{1}{36}$	$\dfrac{2}{36}$	$\dfrac{3}{36}$	$\dfrac{4}{36}$	$\dfrac{5}{36}$	$\dfrac{6}{36}$	$\dfrac{5}{36}$	$\dfrac{4}{36}$	$\dfrac{3}{36}$	$\dfrac{2}{36}$	$\dfrac{1}{36}$

The <u>probability function</u>, f(x), of a discrete random variable X expresses the probability that X takes the value x, as a function of x. That is

$$f(x) = \Pr(X = x)$$

where the function is evaluated at all possible values of x.

Properties of probability function $\Pr(X = x)$:-

1. $\Pr(X = x) \geq 0$ for any value x.
2.   The individual probabilities sum to 1; that is

$$\sum_{x} \Pr(X = x) = 1.$$

Example 4

Find the probability function of the number of boys on a committee of 3 selected at random from 4 boys and 3 girls.

Continuous Probability Distribution

1. The total area under this curve bounded by the x axis is equal to one.
2. The area under the curve between lines x = a and x = b gives the probability that X lies between a and b, which can be denoted by $\Pr(a \leq X \leq b)$.
3. We call f(x) a "probability density function", i.e. p.d.f.

3.1.2 Mathematical Expectations

Expectations for Discrete Random variables

The expected value is the mean of a random variable.

Example 5

A review of textbooks in a segment of the business area found that 81% of all pages of text were error-free, 17% of all pages contained one error, while the remaining 2% contained two errors. Find the expected number of errors per page.

Let random variable X be the number of errors in a page.

x	$\Pr(X = x)$
0	0.81
1	0.17
2	0.02

Expected number of errors per page
$$= 0 \times 0.81 + 1 \times 0.17 + 2 \times 0.02$$
$$= 0.21$$

The <u>expected value</u>, $E[X]$, of a discrete random variable X is defined as

$$E[X] \text{ or } \mu_X = \sum_x x \Pr(X = x)$$

<u>Definition</u> :

Let X be a random variable. The expectation of the squared discrepancy about the mean, $E\left[(X - \mu_X)^2\right]$, is called the variance, denoted σ_X^2, and given by

$$
\begin{aligned}
Var(X) \text{ or } \sigma_X^2 &= E\left[(X - \mu_X)^2\right] \\
&= \sum_x (x - \mu_X)^2 \Pr(X = x) \\
&= \sum_x x^2 \Pr(X = x) - \mu_X^2
\end{aligned}
$$

<u>Properties of a random variable</u>

Let X be a random variable with mean μ_X and variance σ_X^2 and a, b are constants.

1. $E[aX + b] = a\mu_X + b$
2. $Var(aX + b) = a^2 \sigma_X^2$

<u>Sums and Differences of random variables</u>

Let X and Y be a pair of random variables with means μ_X and μ_Y and variances σ_X^2 and σ_Y^2, and a, b are constants.

1. $E[aX + bY] = a\mu_X + b\mu_Y$
2. $E[aX - bY] = a\mu_X - b\mu_Y$
3.   If X and Y are independent random variables, then

$$Var(aX + bY) = a^2 \sigma_X^2 + b^2 \sigma_Y^2$$

$$Var(aX - bY) = a^2 \sigma_X^2 + b^2 \sigma_Y^2$$

Measurement of risk : Standard Deviation

Example 6

	PROJECT A			PROJECT B	
Profit(x)	Pr(X=x)	x·Pr(X=x)	Profit(x)	Pr(X=x)	x·Pr(X=x)
150	0.3	45	(400)	0.2	(80)
200	0.3	60	300	0.6	180
250	0.4	100	400	0.1	40
			800	0.1	80

Expected value = 205   Expected value = 220
        ===   ===

Without considering risk, choose B.

But: Variance $(X) = \sum (x - \mu)^2 \Pr(X = x)$

$\therefore$ Variance (A) $= (150 - 205)^2(0.3) + (200 - 205)^2(0.3) + (250 - 205)^2(0.4)$
$= 1,725$
SD(A) $= 41.53$

Variance (B) $= (-400 - 220)^2(0.2) + (300 - 220)^2(0.6) + (400 - 220)^2(0.1)$
$+ (800 - 220)^2(0.1)$
$= 117,600$
SD(B) $= 342.93$

$\therefore$ Risk averse management might prefer A.

Coefficient of Variation (C.V.)

Risk can be compared more satisfactorily by taking the ratio of the standard deviation to the mean of profit. That is :

$$\text{C.V.} = \frac{\text{Standard deviation}}{\text{Mean}} \times 100\%$$

$\therefore$ C.V. of project A $= \dfrac{41.53}{205} \times 100\%$
$= 20.3\%$

$$\text{C.V. of project B} = \frac{342.93}{220} \times 100\%$$
$$= 155.9\%$$

As a result, B is more risky.

3.2   The Normal Distribution

Definition:

A continuous random variable X is defined to be a normal random variable if its probability function is given by

$$f(x) = \frac{1}{\sigma\sqrt{(2\pi)}} \exp[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2] \qquad \text{for } -\infty < x < +\infty$$

where μ = the mean of X

σ = the standard deviation of X

$\pi \approx 3.14159$

Example 7

The following figure shows three normal probability distributions, each of which has the same mean but a different standard deviation. Even though these curves differ in appearance, all three are "normal curves".

Notation: $X \sim N(\mu, \sigma^2)$

Properties of the normal distribution:-

1.   It is a continuous distribution.

2.   The curve is symmetric and bell-shaped about a vertical axis through the mean μ, i.e. mean = mode = median = μ.

3.   The total area under the curve and above the horizontal axis is equal to 1.

4.   Area under the normal curve:
     Approximately 68% of the values in a normally distributed population within 1 standard deviation from the mean.
     Approximately 95.5% of the values in a normally distributed population within 2 standard deviation from the mean.
     Approximately 99.7% of the values in a normally distributed population within 3 standard deviation from the mean.

Definition:

The distribution of a normal random variable with $\mu = 0$ and $\sigma = 1$ is called a standard normal distribution. Usually a standard normal random variable is denoted by Z.

            Notation: $Z \sim N(0, 1)$

Remark: Usually a table of Z is set up to find the probability $P(Z \geq z)$ for $z \geq 0$.

Example 8

Given $Z \sim N(0, 1)$, find

(a) $P(Z > 1.73)$
(b) $P(0 < Z < 1.73)$
(c) $P(-2.42 < Z < 0.8)$
(d) $P(1.8 < Z < 2.8)$
(e) the value z that has
   (i) 5% of the area below it;
   (ii) 39.44% of the area between 0 and z.

Theorem :

If X is a normal random variable with mean μ and standard deviation σ, then

$$Z = \frac{X - \mu}{\sigma}$$

is a standard normal random variable and hence

$$\Pr(x_1 < X < x_2) = \Pr\left(\frac{x_1 - \mu}{\sigma} < Z < \frac{x_2 - \mu}{\sigma}\right)$$

Example 9

Given $X \sim N(50, 10^2)$, find $P(45 < X < 62)$.

Example 10

The charge account at a certain department store is approximately normally distributed with an average balance of $80 and a standard deviation of $30. What is the probability that a charge account randomly selected has a balance

(a)   over $125;
(b)   between $65 and $95.

Solution:
Let X be the balance of charge account ($)
$X \sim N(80, 30^2)$

Example 11

On an examination the average grade was 74 and the standard deviation was 7. If 12% of the class are given A's, and the grades are curved to follow a normal distribution, what is the lowest possible A and the highest possible B?

Solution:
Let X be the examination grade
$X \sim N(74, 7^2)$

3.3   The Binomial Distribution

A binomial experiment possesses the following properties:
1.   There are n identical observations or trials.
2.   Each trial has two possible outcomes, one called "success" and the other "failure". The outcomes are mutually exclusive and collectively exhaustive for each trial.
3.   The probabilities of success p and of failure 1 − p remain the same for all trials.
4.   The outcomes of trials are independent of each other.

Example 12

1.   In testing 10 items as they come off an assembly line, where each test or trial may indicate a defective or a non-defective item.

2.   Five cards are drawn <u>with replacement</u> from an ordinary deck and each trial is labelled a success or failure depending on whether the card is red or black.

Definition:

In a binomial experiment with a constant probability p of success at each trial, the probability distribution of the binomial random variable X, the number of successes in n independent trials, is called the binomial distribution.

$$\text{Notation:} \qquad X \sim b(n, p)$$

$$P(X = x) = \binom{n}{x} p^x q^{n-x} \qquad x = 0, 1, \ldots, n$$

$$p + q = 1$$

Example 13

Of a large number of mass-produced articles, one-tenth are defective. Find the probabilities that a random sample of 20 will obtain
(a)   exactly two defective articles;
(b)   at least two defective articles.

Solution:
Let X be the number of defective articles in the 20
$X \sim b(20, 0.1)$

Example 14

A test consists of 6 questions, and to pass the test a student has to answer at least 4 questions correctly. Each question has three possible answers, of which only one is correct. If a student guesses on each question, what is the probability that the student will pass the test?

Solution:
Let X be the number of correctly answered questions in the 6
X ~ b(6, 1/3)

Theorem

The mean and variance of the binomial distribution with parameters n and p are $\mu = np$ and $\sigma^2 = npq$ respectively where $p + q = 1$.

Example 15

A packaging machine produces 20 percent defective packages. A random sample of ten packages is selected, what are the mean and standard deviation of the binomial distribution of that process?

Solution:
Let X be the number of defective packages in the 10
X ~ b(10, 0.2)

The Normal Approximation to the Binomial Distribution Theorem:

Given X is a random variable which follows the binomial distribution with parameters n and p, then

$$P(X = x) = P(\frac{(x - 0.5) - np}{\sqrt{(npq)}} < Z < \frac{(x + 0.5) - np}{\sqrt{(npq)}})$$

if n is large and p is not close to 0 or 1 (i.e. $0.1 < p < 0.9$).

Remark: If both np and nq are greater than 5, the approximation will be good.

Example 16

A process yields 10% defective items. If 100 items are randomly selected from the process, what is the probability that the number of defective items exceeds 13?

Solution:
Let X be the number of defective items in the 100
X ~ b(100, 0.1)

Example 17

A multiple-choice quiz has 200 questions each with four possible answers of which only one is the correct answer. What is the probability that sheer guesswork yields from 25 to 30 correct answers for 80 of the 200 problems about which the student has no knowledge?

Solution:
Let X be the number of correct answers in the 80
X ~ b(80, ¼)

3.4   The Poisson Distribution

Experiments yielding numerical values of a random variable X, the number of successes (observations) occurring during a given time interval (or in a specified region) are often called Poisson experiments.

A Poisson experiment has the following properties:
1.   The number of successes in any interval is independent of the number of successes in other non-overlapping intervals.
2.   The probability of a single success occurring during a short interval is proportional to the length of the time interval and does not depend on the number of successes occurring outside this time interval.
3.   The probability of more than one success in a very small interval is negligible.

Examples of random variables following Poisson Distribution

1.   The number of customers who arrive during a time period of length t.
2.   The number of telephone calls per hour received by an office.
3.   The number of typing errors per page.
4.   The number of accidents per day at a junction.

Definition:

The probability distribution of the Poisson random variable X is called the Poisson distribution.

Notation: $X \sim Po(\lambda)$

where λ is the average number of successes occuring in the given time interval.

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad x = 0, 1, 2, \dots$$

$$e \approx 2.718282$$

Example 18

The average number of radioactive particles passing through a counter during 1 millisecond in a laboratory experiment is 4. What is the probability that 6 particles enter the counter in a given millisecond?

Solution:
Let X be the number of radioactive particles passing through the counter in 1 millisecond.
$X \sim Po(4)$

Example 19

Ships arrive in a harbour at a mean rate of two per hour. Suppose that this situation can be described by a Poisson distribution. Find the probabilities for a 30-minute period that

(a)   No ships arrive;
(b)   Three ships arrive.

Solution:
Let X be the number of arrivals in 30 minutes
X ~ Po(1)

Theorem:

The mean and variance of the Poisson distribution are both equal to λ.

Poisson approximation to the binomial distribution

If n is large and p is near 0 or near 1.00 in the binomial distribution, then the binomial distribution can be approximated by the Poisson distribution with parameter λ = np.

Example 20

If the probability that an individual suffers a bad reaction from a certain injection is 0.001, determine the probability that out of 2000 individuals, more than 2 individuals will suffer a bad reaction.

According to binomial:
Required probability

$$= 1 - \left\{ \binom{2000}{0}(0.001)^0(0.999)^{2000} + \binom{2000}{1}(0.001)^1(0.999)^{1999} + \binom{2000}{2}(0.001)^2(0.999)^{1998} \right\}$$

Using Poisson distribution:

$$\Pr(0 \text{ suffers}) = \frac{2^0 e^{-2}}{0!} = \frac{1}{e^2} \qquad \because \qquad \lambda = np = 2$$

$$\Pr(1 \text{ suffers}) = \frac{2^1 e^{-2}}{1!} = \frac{2}{e^2}$$

$$\text{Pr(2 suffer)} = \frac{2^2 e^{-2}}{2!} = \frac{2}{e^2}$$

$$\text{Required probability} = 1 - \frac{5}{e^2} = 0.323$$

General speaking, the Poisson distribution will provide a good approximation to binomial when

(i)   n is at least 20 and p is at most 0.05; or

(ii)   n is at least 100, the approximation will generally be excellent provided $p < 0.1$.

Example 21

Two percent of the output of a machine is defective. A lot of 300 pieces will be produced. Determine the probability that exactly four pieces will be defective.

Solution:
Let X be the number of defective pieces in the 300
X ~ b(300, 0.02)

EXERCISE: PROBABILITY DISTRIBUTIONS

1. If a set of measurements are normally distributed, what percentage of these differ from the mean by

    (a)   more than half the standard deviation,
    (b)   less than three quarters of the standard deviation?

2. If $\bar{x}$ is the mean and s is the standard deviation of a set of measurements which are normally distributed, what percentage of the measurements are

    (a)   within the range $(\bar{x} \pm 2s)$
    (b)   outside the range $(\bar{x} \pm 1.2s)$
    (c)   greater than $(\bar{x} - 1.5s)$?

3. In the preceding problem find the constant a such that the percentage of the cases

    (a)   within the range $(\bar{x} \pm as)$ is 75%
    (b)   less than $(\bar{x} - as)$ is 22%.

4. The mean inside diameter of a sample of 200 washers produced by a machine is 5.02mm and the Std. deviation is 0.05mm. The purpose for which these washers are intended allows a maximum tolerance in the diameter of 4.96 to 5.08mm, otherwise the washers are considered defective. Determine the percentage of defective washers produced by the machine, assuming the diameters are normally distributed.

5. The average monthly earnings of a group of 10,000 unskilled engineering workers employed by firms in northeast China in 1997 was ¥1000 and the standard deviation was ¥200. Assuming that the earnings were normally distributed, find how many workers earned:

    (a)   less than ¥1000
    (b)   more than ¥600 but less than ¥800
    (c)   more than ¥1000 but less than ¥1200
    (d)   above ¥1200.

6. If a set of grades on a statistics examination are approximately normally distributed with a mean of 74 and a standard deviation of 7.9 find:

    (a)   The lowest passing grade if the lowest 10% of the students are give F_s.
    (b)   The highest B if the top 5% of the students are given A_s.

7.   The average life of a certain type of a small motor is 10 years, with a standard deviation of 2 years. The manufacturer replaces free all motors that fail while under guarantee. If he is willing to replace only 3% of the motors that fail, how long a guarantee should he offer? Assume that the lives of the motors follow a normal distribution.

8.   Find the probability that in a family of 4 children there will be (a) at least 1 boy, (b) at least 1 boy and 1 girl. Assume that the probability of a male birth is $\frac{1}{2}$.

9.   A basketball player hits on 75% of his shots from the free-throw line. What is the probability that he makes exactly 2 of his next 4 free shots?

10.   A pheasant hunter brings down 75% of the birds he shoots at. What is the probability that at least 3 of the next 5 pheasants shot at will escape? If X represents the number of pheasants that escape when 5 pheasants are shot at, find the probability distribution of X.

11.   A basketball player hits on 60% of his shots from the floor. What is the probability that he makes less than one half of his next 100 shots?

12.   A fair coin is tossed 400 times. Use the normal-curve approximation to find the probability of obtaining:

    (a)   Between 185 and 210 heads inclusive
    (b)   Exactly 205 heads
    (c)   Less than 176 or more than 227 heads.

13.   Ten percent of the tools produced in a certain manufacturing process turn out to be defective. Find the probability that in a sample of 10 tools chosen at random, exactly two will be defective by using:

    (a)   the binomial distribution,
    (b)   the Poisson approximation to the binomial.

14.   Suppose that on the average 1 person in every 1000 is an alcoholic. Find the probability that a random sample of 8000 people will yield fewer than 7 alcoholics.

15.   Suppose that on the average 1 person in 1000 makes a numerical error in preparing his income tax return. If 10,000 forms are selected at random and examined, find the probability that 6, 7, or 8 of the forms will be in error.

16. A secretary makes 2 errors per page on the average. What is the probability that she makes

   (a)   4 or more errors on the next page,
   (b)   no error?

17. The probability that a person dies from a certain respiratory infection is 0.002. Find the probability that fewer than 5 of the next 2000 so infected will die.