-3-
Distributions of discrete random variables

3.1 Basic concepts

Consider selecting 2 persons from a plant site out of 8 chemists (C) and 4 engineers (E).

The probability of choosing 2 chemists = $P(2C) = \frac{C_8^2}{C_{12}^2} = \frac{14}{33}$

The probability of choosing 2 engineers = $P(2E) = \frac{C_4^2}{C_{12}^2} = \frac{1}{11}$

The probability of choosing 1chemist and 1engineer = $P(2C \cap E) = \frac{8\times4}{C_{12}^2} = \frac{16}{33}$

If we define a variable x: "Number of engineers chosen", then we obtain the following results:

$P(x = 0) = \frac{14}{33}$ $P(x = 1) = \frac{16}{33}$ and $P(x = 2) = \frac{1}{11}$

Usually, $P(x = x_i)$ is simply written $p(x_i)$. So that the previous results can be written in a tabulated from as follows:

x_i	0	1	2
$p(x_i)$	$\frac{14}{33}$	$\frac{16}{33}$	$\frac{1}{11}$

In this table, *x* is called a **discrete random variable**. To each event belonging to S, such a number can be assigned. The shown table expresses the **probability distribution (or probability law)** of the discrete random variable x.
It can be readily seen that:

$$\sum_{i=1}^{N} p(x_i) = 1 \tag{3.1}$$

The mean value of a discrete random variable of a population is given by:

$$\mu = \sum_{i=1}^{N} x_i . p(x_i) \tag{3.2}$$

While its standard deviation in the population can be calculated from:

$$\sigma = \sqrt{\sum_{i=1}^{N} x_i^2 . p(x_i) - \mu^2} \tag{3.3}$$

Details of such calculations for the previous example can be tabulated as follows:

x_i	0	1	2	Total
$p(x_i)$	14/33	16/33	1/11	**1**
$x_i . p(x_i)$	0	16/33	2/11	**2/3**
$x_i^2 . p(x_i)$	0	16/33	4/11	**28/33**

So that: $\mu = \frac{2}{3}$

And $\sigma^2 = \frac{28}{33} - \left(\frac{2}{3}\right)^2 = 0.404$

From which: $\boldsymbol{\sigma} = \mathbf{0.636}$

3.2 The uniform discrete distribution

3.2.1 Law of probability of a uniform distribution

A uniform distribution is a distribution where all N values constituting its population have equal probabilities = p.

Following equation (3.1), we get:

$$p = \frac{1}{N} \tag{3.4}$$

This is the case when throwing a uniform dice where the probabilities of obtaining any outcome are equal = $\frac{1}{6}$

3.2.2 Characteristics of a uniform distribution

Mean value

Following equation (3.2), the mean value is:

$$\mu = \sum_{i=1}^{N} p.i = \frac{1}{N}\sum_{i=1}^{N} i = \frac{1}{N}\times(1 + 2 + 3 + \ldots + N) = \frac{N.(N+1)}{2N}$$

$$\mu = \frac{(N+1)}{2} \tag{3.5}$$

Standard deviation

The standard deviation is calculated from equation (3.3):

$$\sigma^2 = \sum_{i=1}^{N} p.i^2 - \mu^2 = \frac{1}{N}\times(1^2 + 2^2 + 3^2 + \ldots + N^2) - \left(\frac{(N+1)}{2}\right)^2$$

$$\sigma^2 = \frac{N(N+1)(2N+1)}{6N}\left(\frac{(N+1)}{2}\right)^2 = \frac{2N^2 + 3N + 1}{6} - \frac{N^2 + 2N + 1}{4} = \frac{N^2 - 1}{12}$$

$$\sigma = \sqrt{\frac{N^2 - 1}{12}} \tag{3.6}$$

This distribution is often used in several practical operations like, for example, assessing the uniformity of the presence of contaminants in waste water or the uniformity of dimensions of products from a factory.

3.3 The binomial distribution

3.3.1 Law of probability of a binomial distribution

Consider the following situation: There are 3 independent reactors in a plant. The probability that at any time, any of them would be operating is 0.85. Therefore if we denote by *x*: "The number of operating reactors", we can calculate the probability distribution of *x* as follows:

p(0) = The probability that none is operating. Since the three reactors are independent then: $p(0) = (1 – 0.85)^3$ = **0.003375**

The probability that one particular reactor is operating is the product $0.85×(1 – 0.85)^2$, but since we have four reactors, then the probability that only one of them be operating will be

$3×0.85×(1 – 0.85)^2$ = **0.057375**

Now, to calculate the probability that 2 of them will be operating we first need to state that these two can be chosen out of three by C_2^3 = 3 methods. Hence *f* (2) = $3×(0.85)^2×(1 – 0.85)$ = **0.325125**

Finally, the probability that all three are operating is $(0.85)^3$ = **0.614125**

The probability distribution is therefore:

x_i	0	1	2	3
$p(x_i)$	0.003375	0.057375	0.325125	0.614125

The average value, as calculated from equation (3.2) = **2.55**

While the standard deviation, as calculated from equation (3.3) = **0.618**

The above situation is an example of a distribution that is encountered in many engineering applications, known as **the binomial distribution.**

This distribution is characterized by three main features:

(1) The number of trials must be fixed beforehand

(2) Consecutive elementary trials must be independent

(3) There exist exactly two outcomes for each trial:

Success of probability *p* and failure of probability *q* = 1 – *p*

The law of probability of the binomial distribution is given by:

$$p(x_r) = C_r^n . p^r . q^{n-r} \qquad (3.7)$$

This is the probability of getting (*r*) successes out of (*n*) trials. ($0 \le r \le n$)

We note that:

$$\sum_{i=0}^{n} p(x_i) = \sum_{i=0}^{n} C_r^n . p^r . q^{n-r} = q^n + C_1^n . q^{n-1} . p + C_2^n . q^{n-2} . p^2 + \ldots + p^n = (p+q)^n = 1$$

3.3.2 Characteristics of a binomial distribution

Mean value

From equation (3.2), the mean value is obtained from:

$$\mu = \sum_{i=0}^{n} x_i . p(x_i) = 0.p(0) + 1.p(1) + 2.p(2) + ..n.p(n)$$

$$= 0.q^n + 1.C_1^n .q^{n-1}.p + 2.C_2^n .q^{n-2}.p^2 + ..n.p^n$$

$$= n.p.(q^{n-1} + 2.\frac{(n-1)}{2}.q^{n-2}.p + 3.\frac{(n-1).(n-2)}{3.2}.q^{n-3}.p^3 + ...p^{n-1})$$

$$= n.p.(q^{n-1} + C_1^{n-1}.q^{n-2}.p + C_2^{n-1}.q^{n-3}.p^3 + ...p^{n-1}) = n.p.(n+p)^{n-1} = n.p$$

Hence

$$\mu = n.p \qquad (3.8)$$

Standard deviation

The value of standard deviation is obtained from the following equation, given here without proof.

$$\sigma = \sqrt{n.p.q} \qquad (3.9)$$

If we apply the above equations to the example given in section (3.1):
n = 3, p = 0.85 and q = 0.15
Hence, μ = 3×0.85 = **0.255**
And, σ **=** $\sqrt{3 \times 0.85 \times 0.15}$= **0.618**
These are the same results obtained under (3.3.1)

3.3.3 Calculations of binomial probabilities using EXCEL

Although equation (3.7) can be used to predict the probability of r occurrences out of n, it is more practical to use the EXCEL function BINOM.DIST as follows:

= BINOM.DIST (number_s, trials, probability_s, cumulative)

The value of (r) is put in the first cell, followed by the value of (n) in the second, then (p) in the third. The fourth cell consists of a logical variable:
If $p(r)$ is required, then write FALSE.

If $\sum_{i=0}^{k} p(i)$ is required, then write TRUE. This will calculate $P(x \leq k)$

Example 3.1

There are 5 pumps in a pumping house. At any time, the probability that any of them would not be operating is 0.2. If a random variable is defined as "x = Number of non-operating pumps", find the mean value of x, its standard deviation, and its law of probability. Then deduce the median and the mode of this distribution.

Solution:

Here, p = 0.2, q = 0.8 and n = 5
Hence, from equations (3.5) and (3.6), we get:

μ = 5×0.2 = **1** and $\sigma = \sqrt{5 \times 0.2 \times 0.85}$= **0.894**

This means that, on the average, there will be **one** non-operating pump

As for the law of probability, the BINOM.DIST function is used with logical variable = FALSE.

p (0) = 0.32768
p (1) = 0.4096
p (2) = 0.2048
p (3) = 0.0512
p (4) = 0.0064
p (5) = 0.00032

r	0	1	2	3	4	5
$p(r)$	0.3277	0.4096	0.2048	0.0512	0.0064	0.00032

To get the median, we use the cumulative probability $P(r)$ by setting the logical variable TRUE as follows:

P(0) = 0.32768
P (1) = 0.73728
P (2) = 0.94208
P (3) = 0.99328
P (4) = 0.99968
P (5) = 1

r	0	1	2	3	4	5
$P(r)$	0.3277	0.7373	0.9421	0.9933	0.9997	1

A plot of P against r reveals that one gets P = 0.5 at a value ≈ **0.4**

As for the mode, we take the value corresponding to the maximum probability, that is, p(1) = 0.4096, which would simply correspond to **1**.

3.4 The Poisson distribution

3.4.1 Law of probability of the Poisson distribution

This distribution is like the binomial distribution in that it is characterized by the same three conditions cited under (3.3.1). However, it is usually used whenever the value of the single probability of success p is very small. The probability of getting (r) successes out of (n) defines the following law of probability:

$$p(r) = \frac{\lambda^r . e^{-\lambda}}{r!} \quad (3.10)$$

Where: $\lambda = n.p$

This distribution is commonly used in engineering applications dealing with events of low probability, such as equipment or power failure in a plant, defective items in a high-quality product, etc. It is generally restricted to the cases where: $\boldsymbol{n} > \mathbf{50}$ ($n \to \infty$) and $\boldsymbol{\lambda = n.p} < \mathbf{6}$, although it is common to use it whenever the probability of occurrence is very low and the number of possible outcomes moderately elevated.

We note that:

$$\sum_{i=0}^{\infty} p(x_i) = p(0) + p(1) + p(2) + \ldots \infty =$$

$$= \frac{\lambda^0.e^{-\lambda}}{0!} + \frac{\lambda^1.e^{-\lambda}}{1!} + \frac{\lambda^2.e^{-\lambda}}{2!} + \frac{\lambda^3.e^{-\lambda}}{3!} + ...\infty = e^{-\lambda}[1 + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + ...\infty]$$

$= e^{-\lambda}.e^{\lambda} = 1$

3.4.2 Characteristics of the Poisson distribution

Mean value

$$\mu = \sum_{i=0}^{\infty} x_i.p(x_i) = 0.p(0) + 1.p(1) + 2.p(2) + ...\infty$$

$$= 0.\frac{\lambda^0.e^{-\lambda}}{0!} + 1.\frac{\lambda^1.e^{-\lambda}}{1!} + 2.\frac{\lambda^2.e^{-\lambda}}{2!} + 3.\frac{\lambda^3.e^{-\lambda}}{3!} + ...\infty$$

$$= e^{-\lambda}[1.\frac{\lambda^1}{1!} + 2.\frac{\lambda^2}{2!} + 3.\frac{\lambda^3}{3!} + ...\infty] = e^{-\lambda}.\lambda[1 + \frac{\lambda}{1!} + .\frac{\lambda 2}{2!} + ...\infty] = e^{-\lambda}.e^{\lambda}.\lambda$$

$$\mu = \lambda = n.p \qquad (3.11)$$

Standard deviation

The following formula gives the value of the standard deviation of a variable following the Poisson distribution:

$$\sigma = \sqrt{\lambda} \qquad (3.12)$$

3.4.3 Calculation of Poisson probabilities using EXCEL

The probability values in a binomial distribution can be readily obtained using the function POISSON.DIST as follows:

POISSON.DIST(*x*, mean, cumulative)

If *p(r)* is required, the probability shall be calculated from:

POISSON.DIST(*r*, λ, false)

If the cumulative probability $P(r) = P(x \le r)$ is required, then the function is:

POISSON.DIST(*r,* λ, true)

Example 3.2

It is known that 3% of a certain batch of detergent packages is defective. Out of a sample of 100 items, calculate the probability that the number of defective items will be: None, 1, 2, 3, 4. Deduce the median value.

Solution:

To apply the Poisson distribution, we have to check the following:

$p = 0.03 \approx 0$

$n = 100 > 50$

$\lambda = n.p = 100 \times 0.03 = 3 < 6$.

Therefore:

p (0) = **0.04978**

p (1) = **0.1493**

p (2) = **0.2240**

p (3) = **0.2240**
p (4) = **0.1680**

We note that the values of probability first increase to reach a maximum **modal value** at $r \approx \lambda$. It then decreases for higher values.

Cumulative values are obtained using the TRUE logical variable and the **median** value was deduced as equal to **2.3** corresponding to a cumulative probability of 0.5 as follows:

*P(*0) = 0.04978
*P(*1) = 0.19908
*P(*2) = 0.42308
*P(*3) = 0.64708…….

Example 3.3
It was found that, on average, a production line is stopped five times a year. Calculate the probability that in a certain year:
(1) Exactly four times
(2) At most four times
(3) At least four times

Solution:
(1) Using EXCEL with mean = 5 and $x = 4$, we get for a FALSE cumulative input: *p* (4) = **0.1755**

(2) Using EXCEL with mean = 5 and $x = 4$, we get for a TRUE cumulative input: *P*(4) = **0.4405**

(3) This means calculating $P(x \geq 4) = 1 - P(x < 4) = 1 - P(x \leq 3) = 1 - 0.2650$ = **0.7350**

3.5 Comparing the binomial and the Poisson distributions

Actually, the Poisson distribution is a special case of the binomial when (*n*) is very big and (*p*) very small. Figure (3.1) shows a comparison between the two distributions for $n = 10$ and $p = 0.1$
Even though $n < 50$, this figure shows that the two distributions are very close.

3.6 The use of trial solutions

In many cases, one needs to obtain an unknown value by trial. The easiest way is to use the Goal – Seek module available in EXCEL on pressing the DATA key. Choose the option "What if analysis". The following example illustrates the use of that module.

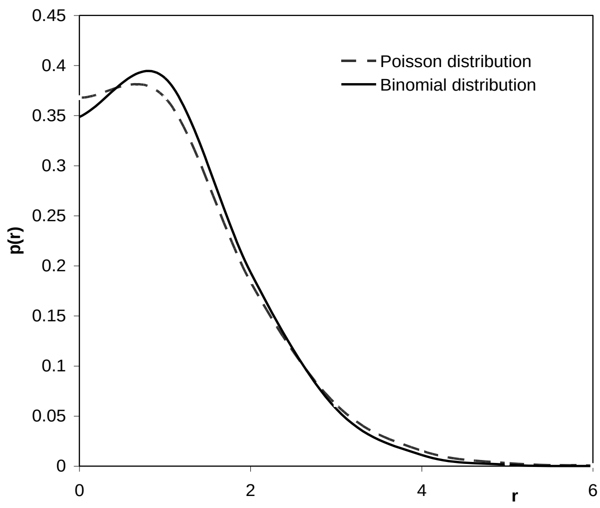


Fig (3.1): Comparison between the two distributions

Example 3.5

It was found from previous practice that there is a 2% probability that 2 accidents take place yearly for company cars. Assuming a Poisson distribution, estimate the mean number of annual accidents.

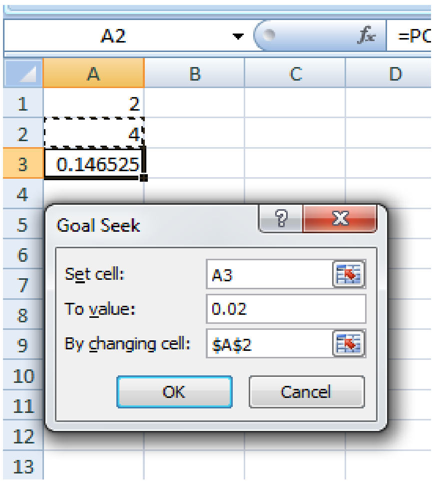
Solution:

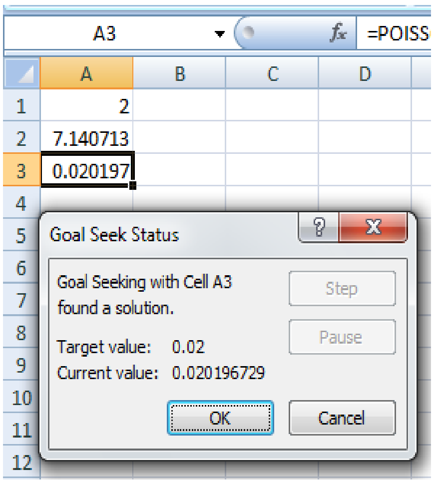
In this example $p(2) = 0.02$ and it is required to calculate the value of λ.

Start by assuming any value of λ (say 4). Write the assumed value of λ (4) in cell A2. Calculate the Poisson probability of $p(2)$ accordingly. You get $p(2) = 0.1465$. Display that results in cell A3.

Then use the goal – seek module as shown:

Now press OK to get the exact value of λ in cell A2 = 7.14





3.7 Exercise problems

(1) X is a random variable that can take the values: $\{1, 2, 3, 4\}$. The probability that $X = r$ is given by:

$P(X = r) = 0.021\ e^{A.X}$

Find the value of A then obtain the expectation and the variance of X.

(2) Given a random variable with the following distribution:

X_i	1.34	2.55	3.67	4.88	5.21	m
$P(X_i)$	n	0.3	0.25	0.1	0.15	0.1

Given that the standard deviation of X is 1.66, find the values of m and n.

(3) The nominal weight of a detergent box is 5 kg. It was found that 25% of boxes have slightly lower weights. In a sample of 50 randomly chosen boxes, find the probability that 10 of them will have weights less than 5 kg.

(4) There are 80 light bulbs in an indoor sector in a chemical plant. On an average day, 3 of them will be defective and will have to be replaced. Use a suitable distribution to calculate the probability that on a given day, 5 bulbs will be out of order.

(5) The proportion of students passing a certain exam is 68%. Out of a group of 120 students, calculate the following probabilities:
(a) Exactly 80 students will pass.
(b) At most 80 students will pass.

(6) In a lot of 1000 items, the probability of finding at least 5 defective items is 0.02, what is the mean number of defective items?

(7) Long term practice has shown that on average, there are 2 fires a year on a certain plant site. What is the probability that in a certain year there will be no fire at all? Three fires? At least three fires?