**AW3c Other useful probability distributions**

Two other important probability distributions in business will be explored in this section:

* Hypergeometric probability distribution.
* Exponential probability distribution.

**Hypergeometric discrete probability distribution and binomial approximation**

Other types of discrete probability distributions include the **hypergeometric discrete probability distribution** which measures, like the binomial distribution, the number of successes from n observations of the experiment. Unlike the binomial which involves replacement and therefore the probability of success (p) is constant the hypergeometric distribution involves sampling without replacement. In this case the probability of success (p) is dependent upon the outcome of the previous run of the experiment. The **characteristics of a hypergeometric distribution** are as follows:

* Discrete distribution.
* The population is finite and known.
* Sampling is without replacement.
* Each outcome is either success or failure.
* The number of successes is known.

Consider a population of size N known to contain M defective items where we select a sample without replacement of size n which contains r defective items. The probability of obtaining r defective items is given by equation (1) and is known as the hypergeometric distribution.

$P\left(X = r\right)= \frac{ × }{}$ (1)

It can be shown that the mean (or expected value) and variance/standard deviation are given by equations (2) - (5) respectively:

$p= \frac{M}{N}$ (2)

$Mean=E\left(X\right)=np$ (3)

$Variance = VAR\left(X\right)=np \left(1-p\right) \frac{N-M}{N-1}$ (4)

$Standard deviation=SD\left(X\right)= \sqrt{VAR(X)}$ (5)

**Example 1**

A manufacturer is concerned at the quality of the items produced by a new production process. A batch of 10 items are sampled which contains 4 defective items. If we draw samples of size 3 without replacement, from the batch of 10. Calculate (a) the probability that a sample contains 2 defective items, and (b) the mean and standard deviation for the number of defectives from the sample.

From the example data, we have: N = 10, M = 4, n = 3, and r = 2. Substituting these values into equation (1) gives:

$$P\left(X=2\right)= \frac{}{}$$

$$P\left(X=2\right)= \frac{}{}$$

$$P\left(X=2\right)= \frac{6 ×6}{120}$$

$$P\left(X=2\right)= 0.3$$

From equations (2) – (5) gives:

$$p= \frac{M}{N}= \frac{4}{10}=0.4$$

$$Mean=E\left(X\right)= np = 3 \* 0.4 = 1.2$$

$$VAR\left(X\right)=np \left(1-p\right) \frac{N-M}{N-1}=3\*0.4\*\left(1-0.4\right)\* \frac{10-4}{10-1}=0.48$$

$$SD \left(X\right)= \sqrt{VAR(X)}= \sqrt{0.48}= 0.69$$

**>>**

We conclude that we have a 30% chance of selecting 2 defectives from the sample and the mean number of defectives from the sample of 3 is 1.2 with a standard deviation of 0.69.

**Excel solution**

Figure 1 illustrates the Excel solution.



Figure 1

**Excel solution**

N = Cell C3 Value

M = Cell C4 Value

n = Cell C5 Value

r = Cell C6 Value

MCr = Cell C8 Formula:=COMBIN(C4,C6)

N-MCn-r = Cell C9 Formula:=COMBIN(C3-C4,C5-C6)

NCn = Cell C10 Formula:=COMBIN(C3,C5)

P(X = 2) = Cell C12 Formula:=C8\*C9/C10

or P(X = 2) = Cell C13 Formula:=HYPGEOM.DIST(C6,C5,C4,C3,FALSE)

p = Cell C14 Formula:=C4/C3

E(X) = Cell C15 Formula:=C5\*C14

VAR(X) = Cell C16 Formula:=C5\*C14\*(1-C14)\*(C3-C4)/(C3-1)

SD(X) = Cell C17 Formula:=SQRT(C16)

If the sample size is small compared to the population size then the ratio n/N is close to zero and sampling with replacement and sampling without replacement will give similar results. This tells us that the value obtained by the hypergeometric and binomial distributions will give approximately the same results when n/N is small.

$Hyp \left(n, M, N\right) ≈ Bin (n,\frac{M}{N})$ (6)

A rule of thumb to use is that if the sample size is at most 5% - 10% of the population size then the binomial distribution will give a good approximation to the hypergeometric distribution, as given by equation (2). It should be noted that the hypergeometric distribution will give the exact value with the binomial distribution providing an approximation to this exact solution when n small compared to the population size, N.

#### Check your understanding

X1

Use the hypergeometric distribution to calculate the following probabilities:

1. P(X = 4) given N = 16, M = 7, and n = 5.
2. P(X = 0) given N = 156, M = 17, and n = 6.
3. P(X < 3) given N = 12, M = 8, and n = 6.
4. P(X > 4) given N = 34, M = 12, and n = 7.

X2

The manufacturer of car tyres is known to produce 24 defective tyres for every 200 tyres. From these 200 tyres, a sample of 5 is selected for testing. Calculate (a) the probability that the sample contains 2 defective tyres, (b) the expected number of defectives in the sample of size 5, and (c) the standard deviation of the number of defectives in the sample of size 5. Would a binomial distribution approximation be appropriate to calculate part (a)? If yes, calculate this probability.

**Exponential distribution**

Other types of continuous probability distributions include the **exponential probability distribution** which is closely related to the Poisson distribution. The Poisson distribution is discrete and describes random occurrences over some interval (time or distance), whereas the exponential distribution is continuous and describes a probability distribution of the times between random occurrences. The characteristics of an exponential distribution are as follows:

* Continuous distribution.
* Distribution skewed to the right as illustrated in Figure MMMMM.
* The value of x ranges from zero to infinity.
* The value of the distribution when x = 0 is f(x) = average value.

If x is described by an exponential function with mean 1/λ, then the probability density function, f(x), is given by equation (7).

$f\left(x\right)= λ e^{-λx}$ (7)

Where λ (pronounced ‘lambda’) represents the mean number of occurrences per given time interval. Figure 2 represents the relationship between the probability density function, f(x), and the value of x when the value of λ = 0.4.



Figure 2

It can be shown that the mean and standard deviation for a exponential distribution are given by equations (8) and (9) respectively.

$Mean= \frac{1}{λ}$ (8)

$Standard deviation= \frac{1}{λ}$ (9)

The probability P(X < x) is given by equation () and is represented by the shaded region in Figure 3 is the area under the probability density function (PDF) to the left of X = x.

$P\left(X < x\right)=1- e^{- λ x}$ (10)



Figure 3

Furthermore, the probability that X > x is given by equation (11)

$P\left(X > x\right)= e^{- λ x}$ (11)

**Example 2**

Students arrive at a pizza shop according to an approximate Poisson process with a mean rate of 12 customers per hour. (a) what is the mean time between customer arrivals, and (b) what is the probability that the pizza shop waits more than 2 minutes for the next customer to enter the shop? Given we are dealing with the time between Poisson arrival event then we will use the exponential function to solve part (a) and part (b).

(a) Mean time between arrivals (minutes)

Given Poisson distribution with λ = 12/60 = 0.2 arrivals per minute.

Employing the exponential distribution to calculate the mean time between arrivals in minutes

Mean time between arrivals = 1/λ = 1/0.2 = 5 minutes

(b) Probability that more than 2 minutes between arrivals

From equation (), P(X > 2) = e-0.2\*2 = 0.6703

**Excel solution**

Figure 4 illustrates the Excel solution.



Figure 4

**Excel solution**

Poisson distribution

Number of arrivals per hour = Cell C4 Value

Mean number of arrivals per minute, l = Cell C5 Formula:=C4/60

Exponential distribution

Mean time between arrivals = Cell C8 Formula:=1/C5

Probability must wait more than 2 minutes for another customer arrival

x = Cell C11 Value

P(X > x) = Cell C12

 Formula:=1-EXPON.DIST(C11,C5,TRUE)

**>>**

We conclude that (a) the mean time between arrivals is 5 minutes with (b) a 67% chance that the piazza shop will have to wait more than 2 minutes between customer arrivals

#### Check your understanding

X3

A manufacturer reports that the number of hours, X, for which a new dab radio will work before requiring repairs follows an exponential distribution with the probability density function f(x) = 0.0001e-0.0001. This information is considered not appropriate for customers and the manufacturer decides to state the mean in the information supplied to customers. Calculate the average time for repair of the dab radios.

X4

A local hospital keeps detailed records for the arrival times at the accident and emergency department. The records show that at the busiest times patients arrive at an average rate of 0.7 per minute. If the arrivals follow a Poisson distribution, calculate the probability that a patient arrives less than 22 seconds after the previous patient.