**AW3b Probability Distributions and Approximations**

Under certain circumstances it is possible to approximate one probability distribution with another probability distribution. The benefit of this is to simplify the calculation process. For example, if you have a binomial distribution with n large and you require to calculate the probability that the value of X is greater than a large number then the calculation process by hand would be quite complex. Fortunately, under certain circumstances we can approximate the binomial distribution with either a Poisson or normal distribution where the solution process is not quite as complex.

Please note that with the use of computer software this issue is not of a major concern and this workbook will explore the use of Excel to illustrate how well the approximations are to the binomial or Poisson probability distributions. Finally, the workbook will provide information on two other probability distributions which have business applications: hypergeometric and exponential probability distributions.

The concept of probability is an important aspect of the study of statistics and within this chapter we will introduce the reader to some of the concepts that are relevant to probability. However, the main emphasis of Chapter 4 in the textbook is to focus on the concepts of continuous and discrete probability distributions and not on the fundamentals of probability theory. In the textbook, we will explore the issue of continuous probability distributions (uniform, normal, Student’s t, and F) and then introduce the concept of discrete probability distributions (binomial, and Poisson). Table 1 summarises the probability distributions that are applicable to whether the data variables are discrete/continuous and whether the distributions are symmetric/skewed.

|  |  |
| --- | --- |
|  | Variable type |
| Measured characteristic | Discrete distributions | Continuous distributions |
| Shape | Symmetric | Skewed | Symmetric | Skewed |
| Distribution | UniformBinomial | PoissonHypergeometric | UniformStudent’s tNormal | FExponential |

Table 1

Under certain circumstances it is possible to use other probability distributions that may be used to approximate other probability distributions. The approximations we will explore in this workbook, include:

1. Poisson approximation to the binomial distribution.
2. Normal approximation to the binomial distribution.
3. Normal approximation to the Poisson distribution.

**Poisson approximation to the binomial distribution**

When the number of trials in a binomial situation is very large and when p is small then it can be shown that the binomial probability function can be approximated by the Poisson probability function with λ = np. The larger the n and the smaller the p, the better is the approximation. Equation (1) represents the **Poisson approximation to the binomial distribution** is used to approximate the true (binomial) result:

 (1)

The Poisson random variable theoretically ranges from 0 → ∞. However, when used as an approximation to the binomial distribution, the Poisson random variable - the number of successes out of n observations - cannot be greater than the sample size n.

With large n and small p, equation (1) implies that the probability of observing many successes becomes small and approaches zero quite rapidly. For small values of p (< 0.1), and large values of n, the Poisson distribution will approximate the binomial distribution with  = np. For the binomial distribution with p small (< 0.1) the mean (or expected) value = np and the variance = npq = np(1-p) ≈ np. This implies that for small p the expected and variance for the binomial distribution is approximated by the mean and variance of the Poisson distribution ( = np, VAR(X) = np).

**Example 1**

In a large consignment of apples 3% are rotten. What is the probability that a carton of 60 apples will contain less than 2 rotten apples? We have here a binomial experiment and therefore could easily apply the binomial distribution with p = 0.03, q = 0.97 and n = 60.

**Excel Solution**

The Excel solution is illustrated in Figure 1.



Figure 1

**Excel solution**

n = Cell C3 Value

p = Cell C4 Value

np = Cell C5 Formula: =C3\*C4

Binomial: P( X < 2) = Cell D7 Formula: =BINOM.DIST(1,C3,C4,TRUE)

Poisson: P( X < 2) = Cell D8 Formula: =POISSON.DIST(1,C5,TRUE)

**>>**

We can see from Excel that the Binomial and Poisson distributions provide approximately equal results, 45.92% and 46.28% respectively.

The degree of agreement between the binomial and Poisson probability distributions for this problem can be observed in Figure 2.



Figure 2

1. Binomial solution

P(less than 2 rotten apples) = P(X < 2)

P(X < 2) = P(X = 0) + P(X = 1)







P(X < 2) = 0.4592

1. Poisson solution

P(less than 2 rotten) = P(X < 2) = P(X = 0) + P(X = 1)

Since n is large and p is small we can use the Poisson distribution. To check if the Poisson distribution is appropriate calculate the mean and variance: mean = np = 60 \* 0.03 = 1.8, and variance = npq = 60 \* 0.03 \* 0.97 = 1.746. Comparing the two values we see that they are approximately equal and the binomial distribution can be approximated using the Poisson distribution:

.

Table 2 compares the two solutions:

|  |  |  |
| --- | --- | --- |
| Problem | Binomial | Poisson |
| P(X < 2) | 0.4592 | 0.4627 |

Table 2

**Example 2**

Suppose that a manufacturing machine is known to produce 1% defective components is used to produce 40 components. Calculate the probability that two defective items are produced. This problem requires calculating P(X = 2) when X ~ Bin (40, 0.01). Figure 3 illustrates the Excel solution

**Excel solution**



Figure 3

Table 3 compares the two solutions:

|  |  |  |
| --- | --- | --- |
| Problem | Binomial | Poisson |
| P(X = 2) | 0.0532 | 0.0536 |

Table 3

#### Check your understanding

X1

A new telephone directory is to be published. Before publication entries are proof - read for errors and any corrections made. Experience suggests that, on average, 0.1% of the entries require correction and that entries requiring correction are randomly distributed. The directory contains 800 pages with 300 entries per page. Two methods for making corrections are proposed: Method A (costs 50p per page containing one correction and £1.50 per page containing two or more corrections), and Method B (costs £1 per page containing one or more corrections). Which method based on cost should be used?

X2

A manufacturer produces lightbulbs that are packed into boxes of 100. Historically, the company states that 0.5% of the lightbulbs produced are defective. Calculate: (a) probability of no defectives, and (b) the probability of 2 or more defectives.

**Normal approximation to the binomial distribution**

Computing binomial probabilities using the binomial probability distribution can be difficult for large values of n. If we were undertaking the calculation using tables then usually tables are supplied up to a value of n of 50 and for values of the probability of success, p. We have seen that the Poisson distribution can be used to approximate the binomial distribution when n > 20 and p < 0.1.

The binomial mean and variance are given by the terms: mean = np and variance = npq. Substituting these terms into textbook equation (3.6) gives equation (2):

 (2)

The **normal distribution to the binomial distribution** can be used to approximate the binomial probabilities when n is large and p is close to 0.5 and np > 5 (and nq > 5), with mean () and variance ().

**Example 3**

Assume you have a fair coin and wish to know the probability that you would get 8 heads out of 10 flips. The binomial distribution has a mean of µ = np = 10 \* 0.5 = 5 and a variance of σ2 = npq = 10 \* 0.5 \* 0.5 = 2.5. The standard deviation is therefore 1.5811. A total of 8 heads is 1.8973 standard deviations above the mean of the distribution [(8-5)/1.5811].

The question then is, ‘what is the probability of getting a value exactly 1.8973 standard deviations above the mean?’ The answer to this question is to remember that the probability of a particular event for a normal distribution is zero given that a particular event (or value of X) will not have an actual area within the normal distribution.

The problem is that the binomial distribution is a discrete probability distribution whereas the normal distribution is a continuous distribution. The solution is to round off and consider any value from 7.5 to 8.5 to represent an outcome of 8 heads.

Using this approach, we can solve discrete binomial problems with a normal approximation if we transform X = 8 for the binomial to the region 7.5 - 8.5 for the normal distribution.

The area is shaded in Figure 4 illustrates the normal distribution approximation of the binomial distribution probability of obtaining 8 heads. We can see that the binomial probability distribution solution, P(X = 8) Binomial ≈ P (7.5 ≤ X ≤ 8.5) Normal.



Figure 4

**Excel solution**

The Excel solution to this problem is illustrated in Figure 5.



Figure 5

**Excel solution**

Binomial

n = Cell D5 Value

p = Cell D6 Value

mean  = Cell D7 Formula: =D5\*D6

Variance 2 = Cell D8 Formula: =D5\*D6\*(1-D6)

SD  = Cell D9 Formula: =SQRT(D8)

Number of heads X = Cell D10 Value

P(X=8) = Cell D11 Formula: =BINOM.DIST(D10,D5,D6,FALSE)

Normal

Lower X1 = Cell D13 Formula: =D10-0.5

Upper X2 = Cell D14 Formula: =D10+0.5

P(X1 ≤ 7.5) = Cell D15 Formula: =NORM.DIST(D13,D7,D9,TRUE)

P(X2 ≤ 8.5) = Cell D16 Formula: =NORM.DIST(D14,D7,D9,TRUE)

P(7.5 ≤ X ≤ 8.5) = Cell D18 Formula: =D16-D15

We can see from Excel that the two probabilities agree with one another. The binomial probability of obtaining 8 heads from 10 flips is 0.043945 and the normal approximation probability of containing 8 heads is 0.043495. Table 4 compares the two solutions:

|  |  |  |
| --- | --- | --- |
| Problem | Binomial | Normal |
| P(X = 8) | 0.043945 | 0.043495 |

Table 4

**>>**

The probability of obtaining 8 heads from 10 flips of a fair coin is approximately 4.3%.

**Example 4**

Enquiries at a travel agent lead only sometimes to a holiday booking being made. The agent needs to take 35 bookings per week to break even. If during a week there are 100 enquiries and the probability of a booking in each case is 0.4, find the probability that the agent will at least break even in this week?

To solve this problem let X represent the number of bookings per week, p represents the probability that a booking will be made p = 0.4, and n represents the number of possible bookings over the week, n = 100.

The area is shaded in Figure 6 illustrates the normal approximation of the binomial probability of obtaining at least 35 bookings. We can see that the binomial probability distribution solution, P(X ≥ 35) Binomial ≈ 1 - P (X ≤ 34.5) Normal.



Figure 6

**Excel solution**

The Excel solution to this problem is illustrated in Figure 7.



Figure 7

**Excel solution**

n = Cell D3 Value

p = Cell D4 Value

mean  = Cell D5 Formula: =D3\*D4

Variance 2 = Cell D6 Formula: =D3\*D4\*(1-D4)

SD  = Cell D7 Formula: =SQRT(D6)

Binomial

P(X ≥ 35) = 1 - P(X ≤ 34)?

Binomial X = Cell D12 Value

P(X ≥ 35) = 1 - P(X ≤ 34) = Cell D13 Formula:=1-BINOM.DIST(D12,D3,D4,TRUE)

Normal

P(X ≥ 34.5)?

Normal X = Cell D18 Value

P(X ≥ 34.5) = Cell D19 Formula: =1-NORM.DIST(D18,D5,D7,TRUE)

Z = Cell D21 Formula: =(D18-D5)/D7

P(X ≥ 34.5) = Cell D22 Formula: =1-NORM.S.DIST(D21, TRUE)

We can see from Excel that the two probabilities agree with one another. The binomial probability of obtaining at least 35 bookings is 0.86966347 and the normal approximation probability of obtaining at least 35 bookings is 0.86921388. Table 5 compares the two solutions:

|  |  |  |
| --- | --- | --- |
| Problem | Binomial | Normal |
| P(X ≥ 35) | 0.86966347 | 0.86921388 |

Table 5

**>>**

The probability of obtaining at least 35 bookings is 87.0%.

(a) Binomial solution

P(X = 35 or more) = P(X  35) = P(X = 35 or 36 or 37 .........or 100)

This would be quite difficult to solve without the aid of calculator or some other computational device, for example, a spreadsheet. From Excel we find that this probability value is P(X  35) = 0.8697.

(b) Normal approximation solution (n = 100, p = 0.4)

 = np = 0.4\*100 = 40 and  =  = 4.899.

P(X  35 for binomial)  P(X  34.5 for normal)

P(X  35 for binomial)   = P(Z ≥ - 1.12) = 0.8692

Comparing the two answers we can see that good agreement has been reached.

#### Check your understanding

X3

Given X is a discrete binomial random variable with p = 0.3 and n = 20: (a) can we use the normal approximation to estimate the binomial probability, (b) what if n is changed to 15, and (c) if n = 40 and p = 0.1 is the normal approximation appropriate?

X4

A tyre manufacturer estimates that 8% of tyres are defective. If 1600 tyres are sampled. Use a normal approximation to calculate (a) the probability that 150 tyres or less are defective, and (b) calculate the probability that 150 tyres are defective.

**Normal approximation to the Poisson distribution**

The normal distribution can also be used to approximate the Poisson distribution whenever the parameter λ, the expected number of successes, equals or exceeds 5. Since the value of the mean and the variance of a Poisson distribution are the same (µ = λ = σ2 then σ =. Substituting these terms into textbook equation (3.6) gives equation (3).

 (3)

The **normal approximation to the Poisson distribution** improves as the value of the mean (λ) grows larger and at a large enough value we can assume that the Z variable is normally distributed.

**Example 5**

The average number of broken eggs per lorry is known to be 50. What is the probability that there will be more than 70 broken eggs on a particular lorry load?

We may use the normal approximation to the Poisson distribution, where the mean and variance are calculated as follows: mean () and variance (). Require P(X > 70 for Poisson)  P(X > 70.5 for normal).

The area is shaded in Figure 8 illustrates the normal approximation of the Poisson probability of obtaining more than 70 broken eggs. We can see that the Poisson probability distribution solution, P(X > 70) Poisson ≈ P (X ≥ 70.5) normal.



Figure 8

**Excel solution**

The Excel solution to this problem is illustrated in Figure 9.



Figure 9

**Excel Solution**

mean  = Cell D3 Value

Variance 2 = Cell D4 Value

SD  = Cell D5 Formula: =SQRT(D4)

Poisson

P(X >70) = 1 - P(X ≤ 70)?

Poisson X = Cell D10 Value

P(X >70) = 1 - P(X ≤ 70) = Cell D11 Formula: =1-POISSON.DIST(D10,D3,TRUE)

Normal P(X ≥ 70.5)?

Normal X = Cell D16 Value

P(X ≥ 70.5) = Cell D17 Formula: =1-NORM.DIST(D16,D3,D5,TRUE)

Z = Cell D19 Formula: =(D16-D3)/D5

P(X ≥ 70.5) = Cell D20 Formula: =1-NORM.S.DIST(D19, TRUE)

We can see from Excel that the two probabilities closely agree with one another. The Poisson probability of obtaining more than 70 broken eggs is 0.002971 and the normal approximation probability of obtaining more than 70 broken eggs is 0.001871.

Table 6 compares the two solutions:

|  |  |  |
| --- | --- | --- |
| Problem | Poisson | Normal |
| P(X > 70) | 0.002971 | 0.001871 |

Table 6

**>>**

The probability of obtaining more than 70 broken eggs is 0.2%.

(a) Poisson solution

P(X > 70) = P(X = 71 or 72 .........)

This would be quite difficult to solve without the aid of calculator or some other computational device, for example, a spreadsheet. From Excel we find that this Poisson probability value is P(X > 70) = 0.002971.

(b) Normal approximation solution

 = λ = 50 and  =  = 7.071068.

P(X > 70 for Poisson)  P(X  70.5 for normal)

P(X > 70 for Poisson)   = P (Z ≥ 2.899138) = 0.001871.

Comparing the two answers we can see that good agreement has been reached.

#### Check your understanding

X5

A local maternity hospital has an average of 36 births per week. Use this information to calculate the following probabilities: (a) find the probability that there are fewer than thirty births in a given week, (b) find the probability that there will be more than forty births in a given week, and (c) find the probability that there will be between thirty and forty births in a given week?

X6

In a factory the average number of machine stoppages due to malfunctions is 12 per day. Use Excel to calculate the probability of having 15 or less stoppages per day via the Poisson distribution and the normal approximation.