BACHELOR OF SCIENCE IN INFORMATION TECHNOLOGY

KIBABII UNIVERSITY

STA 205

Probability and Statistics

Lecture notes

@2021

PROBABILITY DISTRIBUTIONS

SOME SPECIAL RANDOM VARIABLES

These are random variables which occur frequently in applied statistics or in practice. Some of them are discrete while others are continuous.

Discrete random variables

These include uniform, bernoulli, binomial, Poisson, negative binomial, geometric, hypergeometric, etc.

Continuous random variables

These include uniform (rectangular), beta, exponential, normal, gamma, chi-square etc.
We will study some of these distributions.

At the end of it all, you should be able to:

•) state the properties of the random variables described by these distributions.

•) define the distributions by stating their probability functions, their cumulative density functions for continuous distributions, their mean, variances and moment generating functions if they exist.

•) solve probability questions involving these distributions.

The Bernoulli distribution

Bernoulli trail

A Bernoulli trial is a trail which has:

✓ Two outcomes technically termed 'success' and 'failure' and takes values 0 and 1.

✓ Probability of success denoted p is constant in each trail. This implies the probability of failure denoted $q = (1 - p)$.

For example when you toss a coin, you only have two outcomes; Head (H) or Tail (T). The process of tossing a coin i.e. getting a Head or a Tail is called a Bernoulli trial.

Definition

A random variable X is said to have a Bernoulli distribution if its probability mass function is given by

$$p(x) = p(X = x) = \begin{cases} p^x q^{1-x}, & x = 0, 1 \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

Is Equation (1) a pmf?

$p(x) > 0$ and $\sum_{x=0}^{1} p^x q^{1-x} = q + p = 1$. So yes, it is a pmf.

Mean and Variance

$$\begin{aligned} \mu = E(X) &= \sum_{x=0}^{1} x p^x q^{1-x} \\ &= p \end{aligned}$$

$$\begin{aligned} \sigma^2 = Var(X) &= E(X^2) - [E(X)]^2 \\ &= \sum_{x=0}^{1} x^2 p^x q^{1-x} - p^2 \\ &= p - p^2 = p(1-p) \\ &= pq \end{aligned}$$

The Binomial distribution

Using factorial notation to find the number of arrangements of some objects.

Example 1

Find all possible arrangements of

a) 3 objects where one is red, one is blue and one is green.

b) 4 objects where 2 are red and 2 are blue.

Solution

a) The first object can be chosen in 3 ways. The second object can be chosen in 2 ways. The third object can be chosen in 1 way. Hence there are $3 \times 2 \times 1 = 6$ possible ways.

b) If the red objects are labeled R_1 and R_2 and the blue objects B_1 and B_2, then you can treat the objects as 4 different ones and there are 4! ways of arranging them. However, arrangements with R_1R_2 are identical to the arrangements with R_2R_1 and so the total number needs to be divided by 2. A similar argument applies to the Bs and so the number of arrangements is $\dfrac{4!}{2! \times 2!} = 6$

So that

- ✓ n different objects can be arranged in $n! = n \times (n-1) \times \ldots \times 3 \times 2 \times 1$ ways.
- ✓ n objects with a of one type and $(n-a)$ of another can be arranged in $\binom{n}{a} = \frac{n!}{a!(n-a)!}$ ways.

Consider an experiment where we have:

- ✓ A fixed number of trials, n.
- ✓ Each trial has two possible outcomes – Success or Failure.
- ✓ The probability of success, p, in each trial is constant.
- ✓ The trials are independent.

If these four conditions are satisfied we say the random variable X which denotes the number of successes in a fixed number of trials has a binomial distribution and we write $X \sim B(n, p)$. A binomial experiment is a random experiment involving a sequence of independent and identical Bernoulli trials.

Suppose the outcome of n bernoulli trials is x successes and $(n-x)$ failures. The probability of this occurrence is

$$\begin{aligned}
\mathsf{SSFSFFSS} \cdots \mathsf{FFS} &= \mathsf{SSS} \ldots \mathsf{SFFF} \ldots \mathsf{F} \\
&= p(S)p(S) \cdots p(S)p(F)p(F) \cdots p(F) \\
&= p^x q^{n-x}
\end{aligned}$$

Yet, the number of ways can we get x successes from n trials is $\binom{n}{x}$ ways. Hence the probability of getting x successes in n trials is

$$\binom{n}{x} p^x q^{n-x}$$

Definition

A random variable X is said to have a binomial distribution if its probability mass function is given by

$$p(x) = p(X = x) = \begin{cases} \binom{n}{x} p^x q^{n-x}, & x = 0, 1, ..., n \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

Is Equation (7) a pmf?

$p(x) > 0$ and $\sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = \left(\binom{n}{0} p^0 q^n + \binom{n}{1} p^1 q^{n-1} + \cdots + \binom{n}{n} p^n q^0\right) = (p+q)^n = 1.$

So yes, it is a pmf.

Before we compute the mean and variance of the Binomial distribution, I want you to confirm that the following two results are true.

Result 1

$$\begin{aligned}\sum_{x=0}^{n}\binom{n}{x}p^x q^{(n-x)} &= \sum_{x=1}^{n}\binom{n-1}{x-1}p^{(x-1)}q^{(n-1-(x-1))},\\ &= \sum_{x=2}^{n}\binom{n-2}{x-2}p^{(x-2)}q^{(n-2-(x-2))},\\ &= 1.\end{aligned}$$

Result 2

$$\begin{aligned}\text{Recall}\quad Var(X) &= E(X^2) - [E(X)]^2,\\ \text{yet}\quad E[X(X-1)] &= E(X^2) - E(X),\\ E(X^2) &= E[X(X-1)] + E(X),\\ \text{hence}\quad Var(X) &= E[X(X-1)] + E(X) - [E(X)]^2.\end{aligned}$$

Mean and Variance

$$\begin{aligned}\mu = E(X) &= \sum_{x=0}^{n} x\binom{n}{x}p^x q^{(n-x)}\\ &= \sum_{x=0}^{n} x\frac{n!}{x!(n-x)!}p^x q^{(n-x)}\\ &= np\sum_{x=1}^{n}\binom{n-1}{x-1}p^{(x-1)}q^{(n-1-(x-1))}\\ &= np\end{aligned}$$

$$\begin{aligned}\sigma^2 = Var(X) &= E[X(X-1)] + E(X) - [E(X)]^2\\ &= \sum_{x=0}^{n} x(x-1)\binom{n}{x}p^x q^{(n-x)} + np - n^2p^2\\ &= n(n-1)p^2\sum_{x=2}^{n}\binom{n-2}{x-2}p^{x-2}q^{(n-2-(x-2))} + np - n^2p^2\\ &= n(n-1)p^2 + np - n^2p^2\\ &= npq\end{aligned}$$

Exercise

1. Let $X \sim B(n, p)$ with $E(X) = 2$ and $Var(X) = \frac{4}{3}$. Find $p(X > 4)$.

2. Assume that on average one telephone number out of 15 called between 2 p.m. and 3 p.m. on weekdays is busy. What is the probability that if 6 randomly selected telephone numbers are called

i) Not more than 3

ii) At least 3 of them are busy?

3. Consider an urn containing M balls k of which are red and the rest black. Let n balls be picked randomly with replacement. Let X be the number of red balls in a sample of size n. Show that $X \sim B(n, p)$.

4. Learn how to use the Cumulative Binomial tables to work out probabilities.

The Poisson distribution

Recall: The function e^x can be defined as a series

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \tag{3}$$

Let $x = \lambda$ in Equation (3) we get

$$e^\lambda = 1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots \tag{4}$$

dividing both sides of Equation (4) by e^λ gives

$$1 = \frac{\lambda^0 e^{-\lambda}}{0!} + \frac{\lambda^1 e^{-\lambda}}{1!} + \frac{\lambda^2 e^{-\lambda}}{2!} + \frac{\lambda^3 e^{-\lambda}}{3!} + \cdots = \sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} \tag{5}$$

Notice that the sum of the infinite series on the right hand side of Equation (5) equals 1 and you could use these values as probabilities to define a probability distribution.

Definition

A random variable X is said to have a Poisson distribution if its probability mass function is given by

$$p(x) = p(X = x) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!}, & x = 0, 1, \ldots \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

From Equation (5), Equation (6) is a probability mass function. We say that X has a Poisson distribution with parameter λ and denote this as $X \sim P_o(\lambda)$.

Exercise

The random variable $X \sim P_o(1.2)$; find $p(X = 3)$; $p(X \geqslant 1)$; $p(3 < X \leqslant 5)$.

Mean and Variance

$$\begin{aligned}
E(X) &= \sum_{\forall x} x p(x) \\
&= \sum_{x=0}^{\infty} x \frac{\lambda^x e^{-\lambda}}{x!} \\
&= \sum_{x=1}^{\infty} x \frac{\lambda^1 \lambda^{x-1} e^{-\lambda}}{x(x-1)!} \\
&= \lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1} e^{-\lambda}}{(x-1)!} \\
&= \lambda.
\end{aligned}$$

$$\begin{aligned}
Var(X) &= E[X(X-1)] + E[X] - [E[X]]^2 \\
&= \sum_{x=0}^{\infty} x(x-1) \frac{\lambda^x e^{-\lambda}}{x!} + \lambda - \lambda^2 \\
&= \sum_{x=2}^{\infty} x(x-1) \frac{\lambda^2 \lambda^{x-2} e^{-\lambda}}{x(x-1)(x-2)!} + \lambda - \lambda^2 \\
&= \lambda^2 \sum_{x=2}^{\infty} \frac{\lambda^{x-2} e^{-\lambda}}{(x-2)!} + \lambda - \lambda^2 \\
&= \lambda^2 + \lambda - \lambda^2 \\
&= \lambda.
\end{aligned}$$

The fact that the mean equals the variance is an important property of a Poisson distribution. The presence or absence of this property can be a useful indicator of whether or not a Poisson distribution is a suitable model for a particular situation.

Deciding whether or not a Poisson distribution is a suitable model

The random variable X represents the number of events that occur in an interval. The interval may be a fixed length in space or time.

If X is to have a Poisson distribution then the events must occur

- ✓ **singly** in space or time
- ✓ **independently** of each other
- ✓ at a **constant rate** in the sense that the mean number of occurrences in the interval is proportional to the length of the interval.

Such events are said to occur **randomly**. You can use the constant rate idea to adjust the parameter of a Poisson distribution.

The Poisson approximation to Binomial

In some cases you can approximate a binomial distribution with a Poisson distribution. Evaluating binomial probabilities when n is large can be quite difficult and in such circumstances it is sometimes simpler to use an approximation.

If $X \sim B(n, p)$ and

- n is large

- p is small

then X can be approximated by $P_o(np = \lambda)$.
If $np = \lambda$ then $p = \dfrac{\lambda}{n}$

$$\begin{aligned}
p(x) &= \binom{n}{x} p^x q^{n-x}, x = 0, 1, ... \\
&= \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}, x = 0, 1, ... \\
&= \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\
&= \frac{n(n-1)(n-2) \times ... \times (n-(x+1))(n-x)!}{x!(n-x)!} \times \frac{\lambda^x}{n^x} \times \left(1 - \frac{\lambda}{n}\right)^n \times \frac{1}{(1 - \frac{\lambda}{n})^x} \\
&= \frac{n(n-1)(n-2) \times ... \times (n-(x+1))}{n^x} \times \frac{\lambda^x}{x!} \times \left(1 - \frac{\lambda}{n}\right)^n \times \frac{1}{(1 - \frac{\lambda}{n})^x} \\
&= \frac{n}{n} \times \frac{n-1}{n} \times \frac{n-2}{n} \times ... \times \frac{n-(x+1)}{n} \times \frac{\lambda^x}{x!} \times \left(1 - \frac{\lambda}{n}\right)^n \times \frac{1}{(1 - \frac{\lambda}{n})^x} \\
&= 1 \times \left(1 - \frac{1}{n}\right) \times \left(1 - \frac{2}{n}\right) \times ... \times \left(1 - \frac{(x+1)}{n}\right) \times \frac{\lambda^x}{x!} \times \left(1 - \frac{\lambda}{n}\right)^n \times \frac{1}{(1 - \frac{\lambda}{n})^x} \\
\lim_{n\to\infty} p(x) &= \frac{\lambda^x}{x!} \lim_{n\to\infty} 1 \times \left(1 - \frac{1}{n}\right) \times \left(1 - \frac{2}{n}\right) \times ... \times \left(1 - \frac{(x+1)}{n}\right) \times \left(1 - \frac{\lambda}{n}\right)^n \times \frac{1}{(1 - \frac{\lambda}{n})^x} \\
p(x) &= \frac{\lambda^x}{x!} \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n \\
p(x) &= \frac{\lambda^x e^{-\lambda}}{x!}, x = 0, 1, 2, ...
\end{aligned}$$

There is no clear rule as to what constitutes **large** n or **small** p but usually the value of np will be $\leqslant 10$ so that the Poisson table can be used. Generally the larger the value of n and the smaller the value of p the better the approximation will be.

Exercise

1. A car hire firm has 2 cars which it hires out day by day. The number of demands for a car each day is distributed as a Poisson variable with mean 1.5. Calculate the proportion of days on which

    a) neither car is used.

    b) some demand is refused.

2. Look up the following distributions and study them like we have done for the above distributions

a) Hypergeometric distribution

b) Geometric distribution

c) Negative-binomial distribution

THE NORMAL DISTRIBUTION

Introduction

Mass-produced items should conform to a specification. Usually, a mean is aimed for but due to random errors in the production process a tolerance is set on deviations from the mean. For example if we produce piston rings which have a target mean internal diameter of 45 mm then realistically we expect the diameter to deviate only slightly from this value. The deviations from the mean value are often modeled very well by the normal distribution.

Definition

A random variable X is said to have a normal distribution with mean (μ) and variance (σ^2) if its probability density function is given by

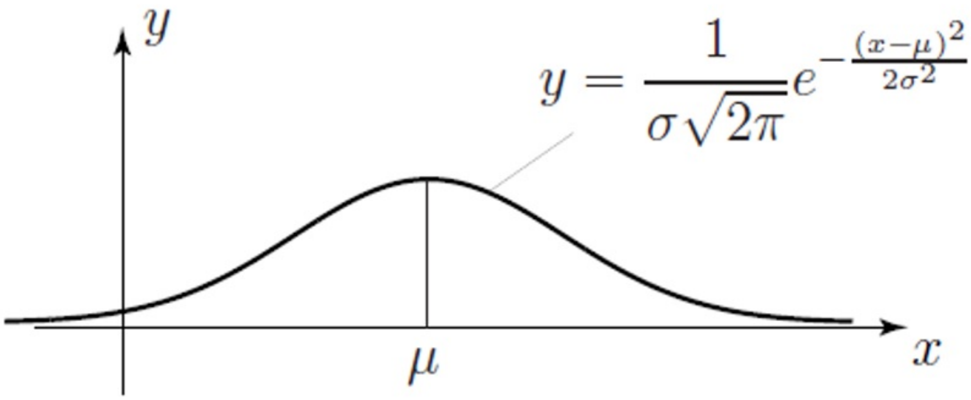
$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, & -\infty < x < \infty, -\infty < \mu < \infty, \sigma^2 > 0 \\ 0, & \text{elsewhere} \end{cases} \tag{7}$$

Many of the techniques in applied statistics are based on the normal distribution.

Characteristics of the Normal Distribution

A normal distribution is completely defined by specifying its mean (the value of μ) and its variance (the value of σ^2). The normal distribution with mean μ and variance σ^2 is written $N(\mu, \sigma^2)$. Hence the distribution $N(20, 25)$ has a mean of 20 and a standard deviation of 5; remember that the second 'parameter' is the variance which is the square of the standard deviation.

- ✓ The curve is bell-shaped with the centre of the bell located at the value of μ.
- ✓ The height of the bell is controlled by the value of σ.
- ✓ As with all normal distribution curves it is symmetrical about the centre and decays exponentially as $x \longrightarrow \pm\infty$; it is asymptotic to the horizontal axis.



- ✓ As with any probability density function the area under the curve is equal to 1.
- ✓ The Empirical rule gives the proportion of the distribution within standard deviations of the mean so that
  - ✓ $\mu \pm \sigma$ includes 68.2%.
  - ✓ $\mu \pm 2\sigma$ includes 95.4%.
  - ✓ $\mu \pm 3\sigma$ includes 99.97%.

CDF of a Normal distribution

$$F(x) = p(X \leqslant x) = \int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx$$

The cdf of a normal random variable is denoted $\Phi(x)$.

Standard Normal distribution

A standard normal random variable is denoted by Z. It is a special normal random variable with mean $\mu = 0$ and variance $\sigma^2 = 1$ i.e. $Z \sim N(0, 1)$. Hence

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, -\infty \leqslant z \leqslant \infty$$

Areas under the standard normal distribution curve are provided in standard normal tables. Every other normal distribution is converted to a standard normal; we say standardizing a normal distribution by

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

Class Exercise

Use of standard normal table

Example

Let X be a normal random variable with mean $\mu = 100$ and standard deviation $\sigma = 10$. Find $p(90 \leqslant X \leqslant 100)$.

Solution

$$
\begin{aligned}
p(90 \leqslant X \leqslant 100) &= \int_{x=90}^{x=100} N(100, 10^2)\,dx \\
&= \int_{z=\frac{90-100}{10}}^{z=\frac{100-100}{10}} N(0,1)\,dz \\
&= \int_{z=-1}^{z=0} N(0,1)\,dz \\
&= \Phi(0) - \Phi(-1) = \Phi(1) - \Phi(0) \quad \text{due to symmetry of the normal distribution} \\
&= (0.5 - 0.1587) = (0.8413 - 0.5) \\
&= 0.3413
\end{aligned}
$$

NORMAL APPROXIMATIONS

At the end of this Topic you should know how to

1. approximate a binomial distribution by a Poisson distribution
2. approximate a binomial distribution by a normal distribution
3. approximate a Poisson distribution by a normal distribution
4. apply a continuity correction

Introduction

Sometimes the calculations using a binomial or Poisson distribution can be quite cumbersome because of the numbers involved. In many cases the normal distribution provides a simple and accurate approximation to the probability required.

Using a continuity correction

The binomial and Poisson distributions are both discrete distributions but the Normal distribution is a continuous one. Recall that, for a discrete distribution $p(X \leqslant 5) \neq p(X < 5)$ while $p(X < 5) = p(X \leqslant 4)$. However for a continuous distribution such as the normal, $p(X \leqslant 5) = p(X < 5)$ since $p(X = 5) = 0$.

If you are approximating a discrete distribution, X, by a continuous distribution Y, you need to consider how to treat the decimal values of Y between the discrete values of X.
Recall that if $X = 5$ to the nearest integer, then $4.5 \leqslant X < 5.5$; we use this idea when approximating discrete distributions by normal (or any continuous) distribution and we call it a continuity correction.

Example 1

If X is a discrete distribution, apply a continuity correction to the following probabilities: $p(X \leqslant 6)$; $p(X > 10)$; $p(2 \leqslant X < 5)$.
Solution: $p(X \leqslant 6) \approx p(Y < 6.5)$; $p(X > 10) \approx p(Y \geqslant 10.5)$; $p(2 \leqslant X < 5) \approx p(1.5 \leqslant Y < 4.5)$.
The continuity correction can be summarized in the following simple rules

- ✓ First write your probability using $\leqslant$ or $\geqslant$.
- ✓ For p$(X \leqslant n)$ you simply approximate by $p(Y < n + 0.5)$.
- ✓ For $p(X \geqslant n)$ you simply approximate by $p(Y \geqslant n - 0.5)$.

You may find these rules helpful but the required continuity correction can always be found using a simple diagram.

Approximating a binomial distribution by a normal distribution

The random variable $X \sim B(n, p)$; Recall: $E[X] = np$ and $Var[X] = np(1 - p)$. If n is large calculating the binomial coefficients can be difficult and if n is very large most calculators will be unable to perform the calculation (try $\binom{500}{240}$ on your calculator).
If p is close to 0.5, then the binomial distribution will be fairly symmetric and so it is reasonable to assume that a normal distribution might provide a suitable approximation.
It is clearly sensible to choose a normal distribution which has the same mean and variance as the original binomial distribution and this leads to the following approximation.
If $X \sim B(n, p)$ and

- n is large
- p is close to 0.5

Then X can be approximated by $Y \sim N(np, np(1 - p))$ i.e. $\sigma = \sqrt{(np(1 - p))}$. There is no definitive answer to the question of how large n should be or how close to 0.5 p should be.

The simple answer is that the larger n is and the closer p is to 0.5 the better. However the approximation does work well for relatively small values of n if p is close to 0.5 as the following example illustrates.

Example 2

The random variable $X \sim B(20, 0.4)$

a) Use tables to find $p(X \leqslant 6)$.

b) Use a normal approximation to estimate $p(X \leqslant 6)$.

Solution

a) $p(X \leqslant 6) = 0.2500$.

b) $X \sim B(20, 0.4)$ so $Y \sim N(20 \times 0.4, 20 \times 0.4 \times 0.6) = Y \sim N(8, 4.8)$; hence $p(X \leqslant 6) = p(Y < 6.5) = p\left(Z < \frac{6.5-8}{\sqrt{4.8}}\right) = p(Z < -0.6846) = 0.2483$.

NB: A calculator would give 0.24678... here both of these values are equal to 0.25 to 2 s.f. thus demonstrating that the approximation has worked well.

Approximating a Poisson distribution by a normal distribution

If the mean of a Poisson distribution is large then a normal approximation can be used. Just like for the binomial distribution we choose the normal distribution to have the same mean as the original Poisson distribution mean and the same variance as the original Poisson distribution. Hence if λ is large, $X \sim P_o(\lambda)$ can be approximated by $Y \sim N(\lambda, \lambda)$.

As before there is no simple answer, other than the larger the better, to the question of how large λ should be before a Poisson distribution can be approximated to a normal distribution.

Example 3

The random variable $X \sim P_o(25)$ Use a normal approximation to estimate

a) $p(X > 30)$.

b) $p(18 \leqslant X < 35)$.

Solution

$X \sim P_o(25)$ so $Y \sim N(25, 25)$

a) $p(X > 30) \approx p(Y > 30.5) = p(Z > \frac{30.5-25}{5}) = p(Z > 1.1) = 0.1357$.

b) $p(18 \leqslant X < 35) = p(17.5 \leqslant X < 34.5) = p(\frac{17.5-25}{5} \leqslant Z < \frac{34.5-25}{5}) = p(-1.5 \leqslant Z < 1.9) = 0.9045$.

Class Exercise

A car hire firm has a large fleet of cars for hire by the day and it is found that the fleet suffers breakdowns at the rate of 21 per week. Assuming that the breakdowns occur at a constant rate, randomly in time and independently of one another, use a suitable approximation to estimate the probability that in any one week more than 27 breakdowns occur. (Answer = 0.0778).

Choosing the appropriate approximation

We have seen that a binomial distribution can sometimes be approximated by a Poisson distribution and sometimes it is approximated by a normal distribution.

If you are approximating a binomial distribution by a normal distribution you should always go directly to the normal distribution and not via a Poisson distribution as this involves one not two approximations and should therefore be more accurate.

For a binomial distribution, there are two possible approximations, depending on whether p lies close to 0.5 (in which case a normal distribution is used) or p is small in which case a Poisson distribution is used).

If in doubt over which approximation is appropriate a useful 'rule of thumb' is to calculate the mean np and, if this is less than or equal to 10, you should be able to use the Poisson tables so that approximation can be used. If the mean is more than 10 then a normal approximation is usually suitable.