BINOMIAL DISTRIBUTION

Bernoulli trials can help you to understand the most important of the discrete probability distributions, the binomial distribution.

Suppose you are conducting an experiment, consisting of n trials, where:

- n is determined before the experiment begins
- all n trials are identical Bernoulli trials, with probability of success p and probability of failure $q = 1 - p$
- all the trials are independent, so that the outcome from any one trial has no effect on the outcome of any other trial.

In these cases there is a **binomial random variable**, which in turn has a **binomial probability distribution**.

There is a shorthand notation to indicate a binomial distribution:

$X \sim B(n, p)$ indicates a random variable X that has a binomial distribution with n identical trials and a probability of success of p.

n and p are called the *parameters* of the distribution.

$X \sim B(n, p)$ is read as 'X is distributed as a binomial variable with parameters n and p'.

Consider the experiment of drawing three cards, one at a time with replacement, from a standard pack of 52 playing cards. If interested in the number X of hearts cards selected, then X is a binomial variable. Consider drawing a heart card to be a success. In terms of notation, $n = 3$ (the number of trials) and $p = \frac{1}{4}$ (the probability of success on any particular trial, as $P(\text{heart}) = \frac{1}{4}$.

So, $X \sim B\left(3, \frac{1}{4}\right)$.

If you were drawing three cards from a pack without replacement, then X would not be a binomial variable because the probability of success p would change with each draw.

Even if an experiment has only two possible outcomes, it is not automatically binomial.

Example 2

Find the probability of obtaining exactly two hearts in a selection of three cards if the card is replaced after each selection.

Solution

There are three ways to obtain exactly two hearts: HHN, HNH, NHH, where H represents 'getting a heart' and N represents 'not getting a heart'.

In each case, the probability is equal to $\frac{1}{4}\times\frac{1}{4}\times\frac{3}{4}$.

$$P(\text{exactly two hearts}) = 3\times\left(\frac{1}{4}\right)^2\times\frac{3}{4} = \frac{9}{64}$$

Binomial theorem

In Chapter 6 Permutations and combinations, you developed Pascal's triangle and the general expansion of $(a + b)^n$ using the binomial theorem.

This is written as:

$$(a+b)^n = {}^nC_0a^n + {}^nC_1a^{n-1}b + {}^nC_2a^{n-2}b^2 + \dots + {}^nC_ra^{n-r}b^r + \dots + {}^nC_nb^n$$

where ${}^nC_r = \frac{n!}{(n-r)!r!}$

Example 3

Calculate, using technology or by hand, the coefficients in the expansion of $(a+b)^6$.

Solution

${}^6C_0 = 1, {}^6C_1 = 6, {}^6C_2 = 15, {}^6C_3 = 20, {}^6C_4 = 15, {}^6C_5 = 6, {}^6C_6 = 1$

Thus $(a+b)^6 = a^6 + 6a^5b + 15a^4b^2 + 20a^3b^3 + 15a^2b^4 + 6ab^5 + b^6$.

It is usual to replace ${}^nC_r$ by $\begin{pmatrix} n \\ r \end{pmatrix}$ when working with binomial probabilities. Do not mix up this notation with the column vector!

It is easy to work out simple examples like Example **2**, but when finding the probability of drawing 13 hearts from 30 draws, it is not as easy to list all the possible outcomes. Instead you can use ${}^{30}C_{13}$ to find the total number of possible outcomes. With the binomial distribution, use the alternative notation $\begin{pmatrix} n \\ r \end{pmatrix}$ or in this case, $\begin{pmatrix} 30 \\ 13 \end{pmatrix}$.

Consider the probability of obtaining any one of these outcomes, such as 13H followed by 17N. The probability of this is $\left(\frac{1}{4}\right)^{13} \times \left(\frac{3}{4}\right)^{17}$.

When you consider all possible outcomes:

$P(\text{exactly 13 hearts from 30 draws}) = \begin{pmatrix} 30 \\ 13 \end{pmatrix} \times \left(\frac{1}{4}\right)^{13} \times \left(\frac{3}{4}\right)^{17}$

Using technology to evaluate, the expression is equal to 0.013 414 448 8… which can be rounded to 0.0134.

A general expression for the probability that X takes a particular value x is as follows:

$$P(X = x) = \begin{pmatrix} n \\ x \end{pmatrix} p^x (1-p)^{n-x}$$

Recalling that the probability of failure is sometimes written as q, where $p + q = 1$, this can also be written as:

$$P(X = x) = \begin{pmatrix} n \\ x \end{pmatrix} p^x q^{n-x}$$

Example 4

A variable x follows the distribution $X \sim B(10, 0.6)$.

Find $P(X = 5)$ for this distribution, expressing your answer correct to four decimal places.

Solution

Identify n (the number of trials), x (the number of successes), p (the probability of success), and $1 - p$ (the probability of failure):

$n = 10$ $X = 5$ $p = 0.6$ $1 - p = 0.4$

Substitute into the formula: $P(X = x) = \begin{pmatrix} n \\ x \end{pmatrix} p^x (1-p)^{n-x}$

$P(X = 5) = \begin{pmatrix} 10 \\ 5 \end{pmatrix} (0.6)^5 (0.4)^5$

$P(X = 5) = 0.2007$

There is a clear link between the binomial probability distribution and the binomial theorem, although the order of the terms has been reversed. You can use this link to show that the sum of the binomial probabilities equals one. Remember, it is a condition for a probability distribution that the sum of the probabilities equals one.

Using the expansion of the binomial theorem:

$$\begin{aligned}\sum_{i=0}^{n}\binom{n}{i}p^{n-i}(1-p)^{i} &= (p+1-p)^{n}\\ &= 1^{n}\\ &= 1\end{aligned}$$

You can also think of a binomial distribution in terms of the expansion of $(p+q)^n$, where $p+q=1$, and p is the probability of success.

Example 5

The probability of any particular egg being cracked in a carton containing a dozen eggs is 0.05.

Find the probability that exactly three eggs are cracked, stating your answer correct to three decimal places.

Solution

Binomial probability is appropriate because there are two outcomes, a fixed probability of success and a defined number of trials.

Identify n (the number of trials), x (the number of successes, i.e. cracked eggs), p (the probability of success), and $1-p$ (the probability of failure):

$n=12, x=3, p=0.05, 1-p=0.95$

Substitute into the formula: $P(X=x)=\binom{n}{x}p^{x}(1-p)^{n-x}$

$$P(X=3)=\binom{12}{3}(0.05)^{3}(0.95)^{9}$$

$$P(X=3)=0.017$$

Example 6

A particular medical test correctly identifies whether or not a person has an illness 98% of the time. If 10 people are tested, find the following probabilities, correct to three decimal places:

(a) No people are incorrectly diagnosed.

(b) At least one person is incorrectly diagnosed.

Solution

(a) Identify n, x and p: $n=10, x=10, p=0.98$

Substitute into the formula: $P(X=x)=\binom{n}{x}p^{x}(1-p)^{n-x}$

$$P(X=10)=\binom{10}{10}(0.98)^{10}(0.02)^{0}$$

$$P(X=10)=0.817$$

(b) P(at least one incorrectly diagnosed) $=1-p$ (none incorrectly diagnosed)

(None incorrectly diagnosed is the same as all correctly diagnosed)

$$\begin{aligned}P(X<10) &= 1-P(X=10)\\ &= 1-0.817\\ &= 0.183\end{aligned}$$

Note that $P(X \geq a) = 1 - P(X < a)$.

This can be a useful result in many situations.

Example 7

Accurate Andy is a darts player whose favourite shot is the triple-twenty. Andy is successful with this shot 80% of the time. In each of the following situations, Andy has five shots. In each case state your answer correct to four decimal places where necessary.

(a) Find the probability that Andy makes exactly two triple-twenties.
(b) Find the probability that Andy makes more than three triple-twenties.
(c) Find the probability that Andy makes triple-twenties on his first and fifth attempts only.

Solution

(a) $n = 5, x = 2, p = 0.8$

Substitute into the formula: $P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$

$$P(X = 2) = \binom{5}{2}(0.8)^2(0.2)^3$$

$$P(X = 2) = 0.0512$$

(b) P(more than three successes) = P(four successes) + P(five successes)

$$P(X > 3) = P(X = 4) + P(X = 5)$$

$$= \binom{5}{4}(0.8)^4(0.2)^1 + \binom{5}{5}(0.8)^5(0.2)^0$$

$$= 0.4096 + 0.32768$$

$$\approx 0.7373$$

(c) Certain conditions have been specified, therefore you cannot use a complete term from the binomial distribution. In this case, consider only the sequence SFFFS, not the more general case of two successes anywhere within the five trials.

P(triple-twenty on first and fifth only) $= 0.8 \times 0.2 \times 0.2 \times 0.2 \times 0.8$

$$= 0.00512$$

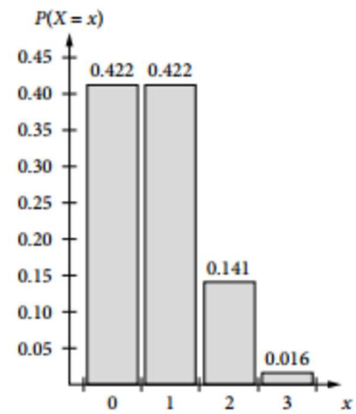
$$\approx 0.0051$$

Graph of the binomial distribution

Consider again the experiment where you draw three cards, with replacement, from a standard pack of 52 playing cards. Remember, you are interested in the number of hearts that occur in the three cards drawn, so $X \sim B(3, 0.25)$. The following table shows the distribution of the random variable:

x	0	1	2	3
$P(X = x)$	0.422	0.422	0.141	0.016

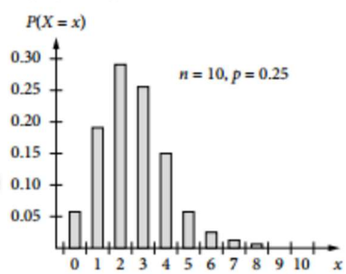
Here, the sum of the probabilities is 1.001. This is due to the rounding of the probabilities and can therefore be regarded as 1. A graph of this distribution is shown:



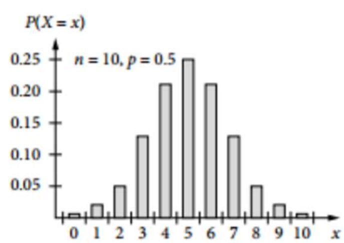
Now, consider the effect that the values n and p have on the graph of the distribution. The following diagrams will help you understand the effect each of these variables has on the graph.

Set $n = 10$.

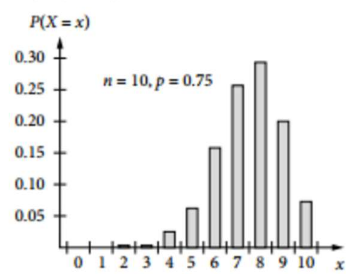
$B(10, 0.25)$



$B(10, 0.5)$



$B(10, 0.75)$



For graphs of the binomial distribution, awhere $X \sim B(n, p)$:

- If $p < 0.5$ the graph is skewed to the right (positively skewed).
- If $p = 0.5$ the graph is symmetric about the mean.
- If $p > 0.5$ the graph is skewed to the left (negatively skewed).
- As n increases, the graph clusters more tightly about the mode but retains the same shape as other distributions with the same value of p.