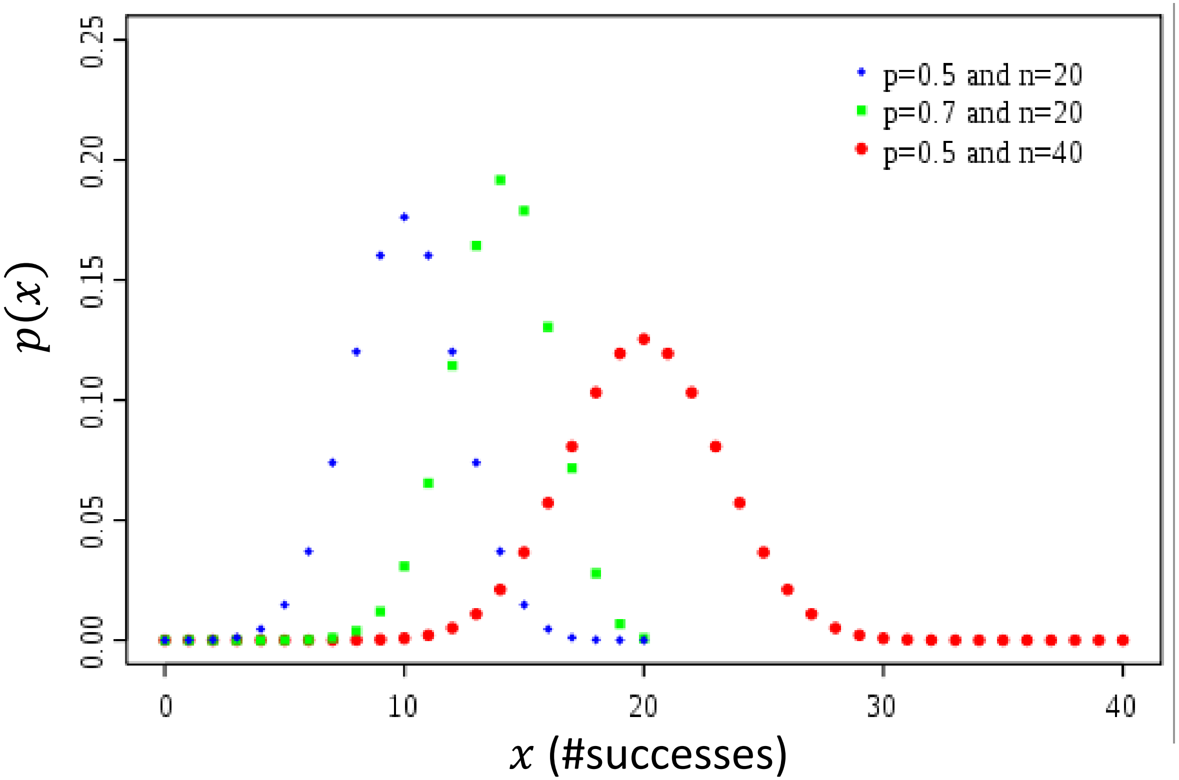
Stats Yr2 Chapter 3: Distribution-N

Binomial Approximation

Approximating a Binomial Distribution



The graph shows the probability function for different Binomial Distributions. Which one resembles another distribution and what distribution does it resemble?

When p is close to 0.5, and n is fairly large, it resembles a normal distribution. The $p = 0.5$ results in the distribution being symmetrical. e.g. For a fair coin toss with 10 throws, we're just as likely to get 1 Head out fo 10 as we are 1 Tail.

Approximating a Binomial Distribution

If we're going to use a normal distribution to approximate a Binomial distribution, it makes sense that we set the mean and standard deviation of the normal distribution to match that of the original binomial distribution:

$$\mu = np$$
$$\sigma = \sqrt{np(1-p)}$$

✐ If n is large and p close to 0.5, then the binomial distribution $X \sim B(n, p)$ can be approximated by the normal distribution $N(\mu, \sigma^2)$ where

$$\mu = np$$
$$\sigma = \sqrt{np(1-p)}$$

Quickfire Questions:

$X \sim B(10,0.2) \quad \rightarrow \quad Y \sim N(2, 1.6)$

$X \sim B(20,0.5) \quad \rightarrow \quad Y \sim N(10, 5)$

$X \sim B(6, 0.3) \quad \rightarrow \quad Y \sim N(1.8, 1.26)$

We tend to use the letter Y to represent the normal distribution approximation of the distribution X.

Why use a normal approximation?

- Tables for the binomial distribution only goes up to $n = 50$ and your calculator will reject large values of n.
- The formula for $P(X = x)$ makes use a factorials. Factorials of large numbers cannot be computed efficiently. Type in 65! for example; your calculator will hesitate! Now imagine how many factorials would be required if you wanted to find $P(X \leq 65)$. ☺

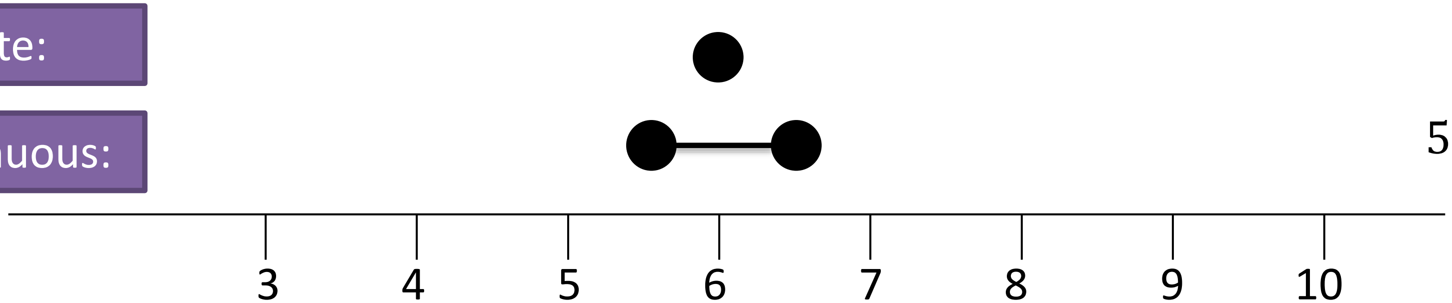
Continuity Corrections

One problem is that the outcomes of a binomial distribution (i.e. number of successes) are **discrete** whereas the Normal distribution is **continuous**.
We apply something called a **continuity correction** to approximate a discrete distribution using a continuous one.

The random variable X represents the time to finish a race in hours. We're interested in knowing the probability Alice took 6 hours to the nearest hour. How would you represent this time on a number line given hours is discrete? And what about if hours was now considered to be continuous (as Y)?

Discrete: $X = 6$

Continuous: $5.5 < Y < 6.5$



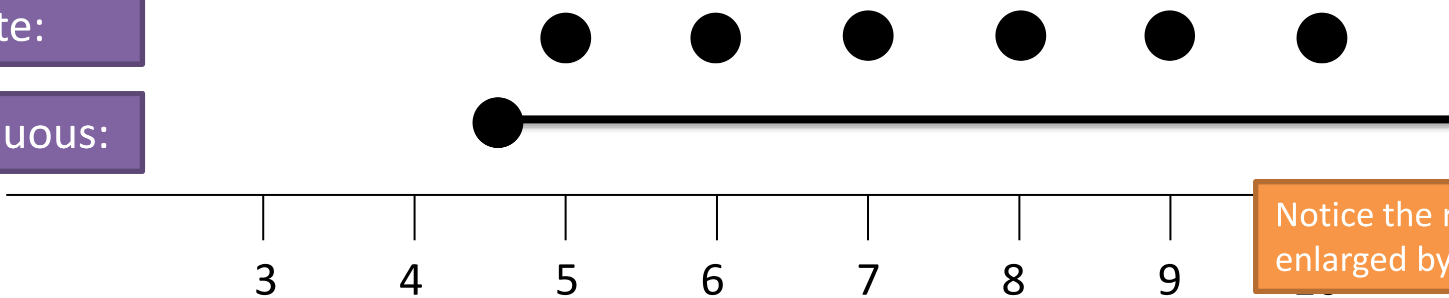
We can't just find $P(Y = 6)$ when Y is continuous, because the probability is effectively 0. But $P(5.5 < Y < 6.5)$ would seem a sensible interval to use because any time between 5.5 and 6.5 would have rounded to 6 hours were it discrete.

Continuity Corrections

If X is a discrete variable, and Y is its continuous equivalent, how would you represent $P(X \geq 5)$ for Y?

Discrete:

Continuous:



$X \geq 5$

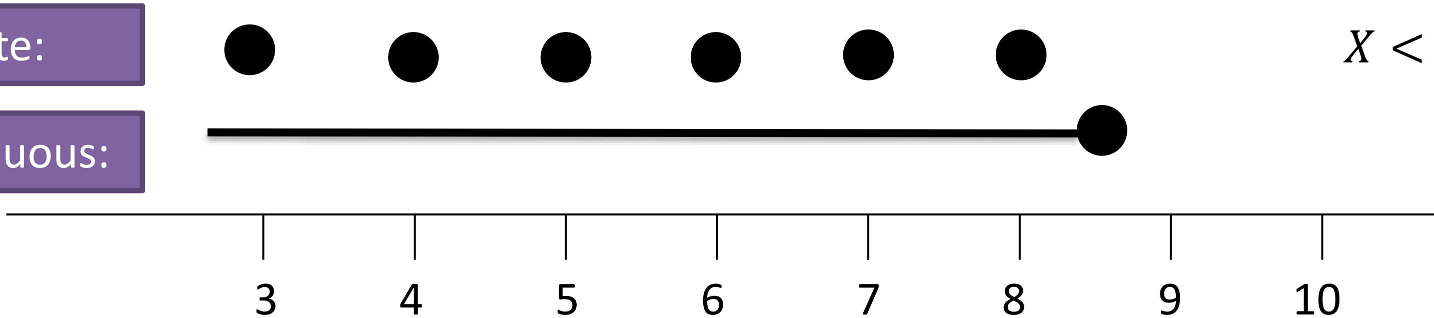
$Y \geq 4.5$

Notice the range has been enlarged by an extra 0.5.

How would represent $P(X < 9)$ for Y?

Discrete:

Continuous:



$X < 9 \rightarrow X \leq 8$

$Y \leq 8.5$

✎ A continuity correction is approximating a discrete range using a continuous one.

1. If > or <, convert to $\geq, \leq$ first.
2. Enlarge the range by 0.5.

Examples

Discrete	Continuous
$P(X \leq 7)$	?
$P(X < 10)$	?
$P(X > 9)$	?
$P(1 \leq X \leq 10)$	?
$P(3 < X < 6)$	?
$P(3 \leq X < 6)$	?
$P(3 < X \leq 6)$	?
$P(X = 3)$	?

✎ A continuity correction is approximating a discrete range using a continuous one.

1. If > or <, convert to ≥, ≤ first.
2. Enlarge the range (at each end) by 0.5.

Examples

Discrete	Continuous
$P(X \leq 7)$	$\approx P(Y \leq 7.5)$
$P(X < 10)$	$= \mathrm{P}(\mathrm{X} \leq 9) \approx P(Y \leq 9.5)$
$P(X > 9)$	$= P(X \geq 10) \approx P(Y \geq 9.5)$
$P(1 \leq X \leq 10)$	$\approx P(0.5 \leq Y \leq 10.5)$
$P(3 < X < 6)$	$= P(4 \leq X \leq 5) \approx P(3.5 \leq Y \leq 5.5)$
$P(3 \leq X < 6)$	$\approx P(2.5 \leq Y \leq 5.5)$
$P(3 < X \leq 6)$	$\approx P(3.5 \leq X \leq 6.5)$
$P(X = 3)$	$\approx P(2.5 \leq X \leq 3.5)$

✐ A continuity correction is approximating a discrete range using a continuous one.

1. If > or <, convert to ≥, ≤ first.
2. Enlarge the range (at each end) by 0.5.

Full Example

[Textbook - Edited] For a particular type of flower bulbs, 55% will produce yellow flowers. A random sample of 80 bulbs is planted.

(a) Calculate the actual probability that there are exactly 50 flowers.
(b) Use a normal approximation to find a estimate that there are exactly 50 flowers.
(c) Hence determine the percentage error of the normal approximation for 50 flowers.

a

?

b

?

c

?

Full Example

[Textbook - Edited] For a particular type of flower bulbs, 55% will produce yellow flowers. A random sample of 80 bulbs is planted.
(a) Calculate the actual probability that there are exactly 50 flowers.
(b) Use a normal approximation to find a estimate that there are exactly 50 flowers.
(c) Hence determine the percentage error of the normal approximation for 50 flowers.

a

$$X \sim B(80, 0.55)$$
$$P(X = 50) = \binom{80}{50} \times 0.55^{50} \times 0.45^{30} = 0.0365$$

b

$$Y \sim N(44, 19.8)$$
$$P(X = 50) \approx P(49.5 < Y < 50.5)$$
$$= 0.9280 - 0.8918 = 0.0362\ (4dp)$$

c

Percentage error:
$$\frac{0.0365 - 0.0362}{0.0365} \times 100 = 0.82\%$$

Test Your Understanding

Edexcel S2 Jan 2004 Q3

The discrete random variable X is distributed $B(n, p)$.

(*a*) Write down the value of p that will give the most accurate estimate when approximating the binomial distribution by a normal distribution.
(1)

(*b*) Give a reason to support your value. **(1)**

(*c*) Given that $n = 200$ and $p = 0.48$, find $P(90 \le X < 105)$. **(7)**

(a) ?

(b) ?

(c) ?

Test Your Understanding

Edexcel S2 Jan 2004 Q3

The discrete random variable X is distributed $B(n, p)$.

(*a*) Write down the value of p that will give the most accurate estimate when approximating the binomial distribution by a normal distribution.
(1)

(*b*) Give a reason to support your value. **(1)**

(*c*) Given that $n = 200$ and $p = 0.48$, find $P(90 \le X < 105)$. **(7)**

(a)	$p = \frac{1}{2}$		**B1**	
				(1)
(b)	Binomial distribution is symmetrical		**B1**	
				(1)
(c)	Since n is large and $p \approx 0.5$ then use normal approximation,	Can be implied below	**M1**	
	$np = 96$ and $npq = 49.92$		**A1A1**	
	$P(90 \le X < 105) \approx P(89.5 \le Y \le 104.5)$ where $Y \sim N(96, 49.92)$	± 0.5 cc on both	**M1,**	
	$\approx P\left(\frac{89.5-96}{\sqrt{49.92}} \le Z \le \frac{104.5-96}{\sqrt{49.92}}\right)$	Standardisation of both	**M1**	
	$\approx P(-0.92 \le Z \le 1.20)$	awrt -0.92 & 1.20	**A1**	
	$\approx 0.7055\text{-}0.7070$	4dp in range	**A1**	
				(7)
				(Total 9 Marks)

Exercise 3.6

Pearson Stats/Mechanics Year 2
Page 27-28

Homework Exercise

1 For each of the following binomial random variables, X:

i state, with reasons, whether X can be approximated by a normal distribution.

ii if appropriate, write down the normal approximation to X in the form $N(\mu, \sigma^2)$, giving the values of μ and σ.

a $X \sim B(120, 0.6)$ **b** $X \sim B(20, 0.5)$ **c** $X \sim B(250, 0.52)$

d $X \sim B(300, 0.85)$ **e** $X \sim B(400, 0.48)$ **f** $X \sim B(1000, 0.58)$

2 The random variable $X \sim B(150, 0.45)$. Use a suitable approximation to estimate:

a $P(X \leqslant 60)$ **b** $P(X > 75)$ **c** $P(65 \leqslant X \leqslant 80)$

3 The random variable $X \sim B(200, 0.53)$. Use a suitable approximation to estimate:

a $P(X < 90)$ **b** $P(100 \leqslant X < 110)$ **c** $P(X = 105)$

4 The random variable $X \sim B(100, 0.6)$. Use a suitable approximation to estimate:

a $P(X > 58)$ **b** $P(60 < X \leqslant 72)$ **c** $P(X = 70)$

5 A fair coin is tossed 70 times. Use a suitable approximation to estimate the probability of obtaining more than 45 heads.

6 The probability of a roulette ball landing on red when the wheel is spun is $\frac{50}{101}$.
On one day in a casino, the wheel is spun 1200 times.
Estimate the probability that the ball lands on red in at least half of these spins.

Homework Exercise

7 **a** Write down two conditions under which the normal distribution may be used as an approximation to the binomial distribution. **(2 marks)**

A company sells orchids of which 45% produce pink flowers.

A random sample of 20 orchids is taken and X produce pink flowers.

b Find $P(X = 10)$. **(1 mark)**

A second random sample of 240 orchids is taken.

c Using a suitable approximation, find the probability that fewer than 110 orchids produce pink flowers. **(3 marks)**

d The probability that at least q orchids produce pink flowers is 0.2. Find q. **(3 marks)**

8 A drill bit manufacturer claims that 52% of its bits last longer than 40 hours.

A random sample of 30 bits is taken and X last longer than 40 hours.

a Find $P(X < 17)$. **(1 mark)**

A second random sample of 600 drill bits is taken.

b Using a suitable approximation, find the probability that between 300 and 350 bits last longer than 40 hours. **(3 marks)**

9 A particular breakfast cereal has prizes in 56% of the boxes. A random sample of 100 boxes is taken.

a Find the exact value of the probability that exactly 55 boxes contain a prize. **(1 mark)**

b Find the percentage error when using a normal approximation to calculate the probability that exactly 55 boxes contain prizes. **(4 marks)**

Homework Answers

For Chapter 3, student answers may differ slightly from those shown here when calculators are used rather than table values.

1 **a** **i** Yes, n is large (> 50) and p is close to 0.5.
ii $X \sim N(72, 5.37^2)$
b **i** No, n is not large enough (< 50).
c **i** Yes, n is large (> 50) and p is close to 0.5.
ii $X \sim N(130, 7.90^2)$
d **i** No, p is too far from 0.5.
e **i** Yes, n is large (> 50) and p is close to 0.5.
ii $X \sim N(192, 9.99^2)$
f **i** Yes, n is large (> 50) and p is close to 0.5.
ii $X \sim N(580, 15.6^2)$

2 **a** 0.1253 **b** 0.0946 **c** 0.6723
3 **a** 0.0097 **b** 0.5596 **c** 0.0559
4 **a** 0.6203 **b** 0.4540 **c** 0.0102
5 0.006
6 0.3767
7 **a** n large, p close to 0.5. **b** 0.1593
c 0.5772 **d** 115
8 **a** 0.6277 **b** 0.8457
9 **a** 0.0786 **b** 0.26%