# The normal approximation to the binomial; choosing a distribution

S4.2, S4.3 · Statistics · revision notes

Approximating $B(n, p)$ by $N(np, np(1-p))$ when $n$ is large and $p$ is close to 0.5, using a continuity correction, and choosing a suitable distribution (binomial, normal or discrete uniform) for a context, including recognising when a model is not appropriate.

## Key facts

- $B(n, p) \approx N(np, np(1-p))$ when $n$ is large and $p$ is close to 0.5
- $P(X \geq r) \approx P(Y > r - 0.5)$
- $P(X = r) \approx P(r - 0.5 < Y < r + 0.5)$
- Binomial: fixed $n$, two outcomes, independent trials, constant $p$
- Discrete uniform: equally likely values
- $P(X \leq r) \approx P(Y < r + 0.5)$
- Normal: continuous, symmetrical, bell-shaped

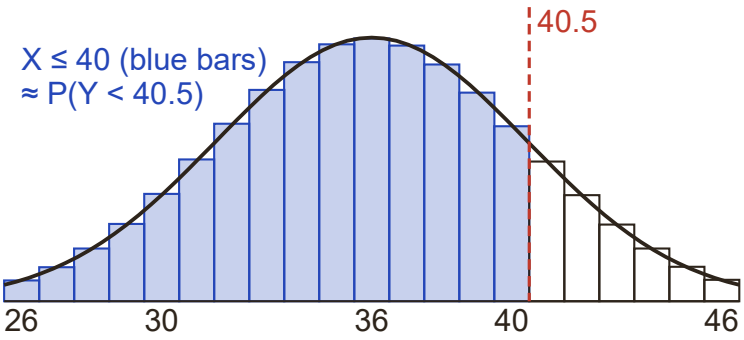
## Notes

### The normal approximation

- If $X \sim B(n, p)$ with $n$ large and $p$ close to 0.5, then $X$ can be approximated by $Y \sim N(np, np(1-p))$
- Under these conditions the binomial distribution is roughly symmetrical and bell-shaped.
- Your calculator can find binomial probabilities exactly, so questions normally tell you when to use the approximation (e.g. ‘using a suitable approximation’).

## Continuity correction

- The binomial distribution is discrete but the normal is continuous, so each integer $r$ is represented by the interval from $r - 0.5$ to $r + 0.5$
- $P(X \leq r) \approx P(Y < r + 0.5)$ and $P(X < r) \approx P(Y < r - 0.5)$




Each bar r covers r − 0.5 to r + 0.5, so the normal area must stop at 40.5

- $P(X \geq r) \approx P(Y > r - 0.5)$ and $P(X > r) \approx P(Y > r + 0.5)$

$P(X = r) \approx P(r - 0.5 < Y < r + 0.5)$

## Choosing a distribution

- **Discrete uniform**: a fixed set of equally likely values.
- **Binomial**: the number of successes in a fixed number of independent trials, each with the same probability of success.
- **Normal**: a continuous variable that is roughly symmetrical, with most values near the mean.
- A binomial model is not suitable if there is no fixed number of trials, the trials are not independent, or the probability changes.
- A normal model is not suitable if the data are clearly skewed, or if it gives a noticeable probability of impossible values (e.g. negative times). Check whether $\mu - 3\sigma$ and $\mu + 3\sigma$ are realistic values.

## How to answer each type of question

M = method mark, A = answer mark, B = a mark on its own.

## Use a normal approximation to a binomial distribution

5 to 7 marks | **Grade B**

1. Check the conditions ($n$ large, $p$ close to 0.5) and write $Y \sim N(np, np(1-p))$
2. Turn strict inequalities into $\leq$ or $\geq$, then apply the continuity correction.
3. Use normal CD with $\sigma = \sqrt{np(1-p)}$

**Example.** In a large town, 45% of adults own a bicycle. A random sample of 80 adults is taken, and $X$ is the number who own a bicycle.
(a) Explain why a normal distribution can be used to approximate the distribution of $X$.
(b) Use a normal approximation to estimate $P(X \leq 40)$ and $P(30 \leq X < 38)$

---

**Model answer.**

(a) $n = 80$ is large and $p = 0.45$ is close to 0.5 (B1)

(b) $Y \sim N(36, 19.8)$ (B1)

$P(X \leq 40) \approx P(Y < 40.5)$ (M1) $= 0.8441$ (A1)

$P(30 \leq X < 38) = P(30 \leq X \leq 37)$

$\approx P(29.5 < Y < 37.5)$ (M1) $= 0.5599$ (A1)

(The exact binomial values are 0.8441 and 0.5619, so the approximation is good.)

## Choose a suitable distribution

1 to 2 marks each | **Grade A**

1. A count of successes in a fixed number of independent trials with a constant probability: binomial.
2. A continuous, roughly symmetrical measurement: normal.
3. A set of equally likely values: discrete uniform.
4. If none fits, say which condition fails.

**Example.** For each variable, suggest a suitable distribution, or explain why neither a binomial nor a normal distribution is suitable.
(a) The number of sixes in 30 rolls of a fair die
(b) The lengths of the leaves on a large tree
(c) The number of goals scored by a team in a football match

---

**Model answer.**

(a) $B(30, \frac{1}{6})$: a fixed number of independent rolls, each with probability $\frac{1}{6}$ of a six (B1)

(b) Normal: length is continuous and likely to be roughly symmetrical about a typical value (B1)

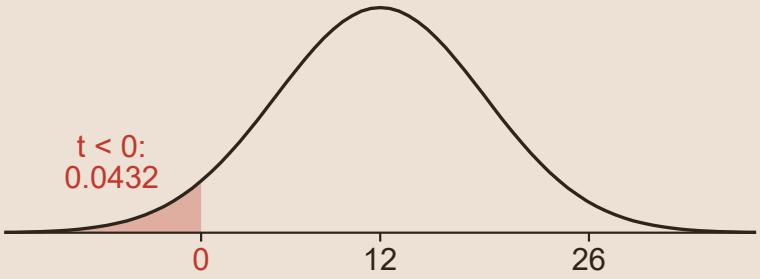
(c) Neither: there is no fixed number of trials, so it is not binomial; it is a discrete count of small values that is likely to be skewed, so it is not normal (B1)

**Show that a normal model is not suitable** 2 to 3 marks **Grade A***

1. Use the model to find the probability of an impossible value (e.g. below 0).
2. Or compare the model with the data: skew, or $\mu \pm 3\sigma$ outside the possible range.
3. Conclude in context.

**Example.** The times, $T$ minutes, that customers spend in a shop have mean 12 and standard deviation 7. A manager suggests modelling $T$ by $N(12, 7^2)$. Explain why this model is not suitable.

**Model answer.**




$P(T < 0) = P(Z < -1.714)$ (M1) $= 0.0432$ (A1), so the model predicts that about 4% of customers spend a negative time in the shop, which is impossible (B1). The times are likely to be positively skewed instead.

## Don't lose marks

- Leaving out the continuity correction, or applying it in the wrong direction.
- Using $np(1-p)$ as the standard deviation.
- Giving the conditions as '$n$ large and $p$ small': for the normal approximation, $p$ must be close to 0.5.
- Choosing a normal model for a discrete count with only a few possible values.

## More tips

### Memory tricks

- Continuity correction: draw a number line, mark the integers you want, and extend half a unit beyond the end ones.
- Change $<$ and $>$ into $\leq$ and $\geq$ first: $P(X < 38) = P(X \leq 37)$, then add 0.5.
- In $N(np, np(1-p))$ the second number is the variance: take its square root for $\sigma$

### Exam technique

- State the approximating distribution with both parameters, e.g. $Y \sim N(36, 19.8)$: this is often a mark on its own.
- When justifying the approximation, give both conditions: $n$ is large and $p$ is close to 0.5.
- When choosing a model, give a reason from the context for each condition you rely on.

## What each grade needs

What you need to be able to do, from the first marks up to the top grade.

**Grade E**

### Recognise when a binomial model is suitable

The variable counts successes in a fixed number of independent trials with a constant probability of success.

**Grade D**

### State the conditions for the normal approximation

$n$ is large and $p$ is close to 0.5.

**Grade C**

### Find the parameters of the approximating normal

$\mu = np$ and $\sigma^2 = np(1-p)$

**Grade C**

### Apply a continuity correction correctly

Extend half a unit beyond the end integers you want, e.g. $P(X \leq 40) \approx P(Y < 40.5)$

**Grade B**

### Approximate binomial probabilities using a normal distribution

Combine the parameters, the continuity correction and normal CD.

**Grade A**

### Choose and justify a distribution for a context

Decide between discrete uniform, binomial and normal, and explain when none of them fits.

**Grade A***

### Judge a model against evidence

E.g. show that a normal model gives impossible values, or compare an approximation with the exact probability.

---

Free notes, Full packs, practice papers and mock exams at chhetriacademy.com. Written for this site; not taken from past papers. Stuck? 1-to-1 tutoring with a free 30-minute trial lesson.