Lesson - Probability and the binomial distribution

Year: Year 12

Probability measures how likely an event is, on a scale from 0 (impossible) to 1 (certain). This chapter extends GCSE probability to Venn diagrams, conditional probability and the binomial distribution, which models repeated trials with two outcomes.

Starter activity

Roll two dice and record the total 36 times. Which total came up most often? Do the results match what you expect from listing the 36 equally likely outcomes? Now count how many outcomes give a total of 7, and compare with the number for a total of 12. Explain the difference. If a friend claims that a total of 7 is "due" after several rolls without a 7, what would you say to explain why each roll is independent of the previous ones?

Probability rules

Definition: mutually exclusive and independent events

Events are **mutually exclusive** if they cannot happen together. Events are **independent** if one does not affect the other.

Rule: laws of probability

- $P(A') = 1 - P(A)$.
- Addition law: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. For mutually exclusive events $P(A \cap B) = 0$.
- Independent events: $P(A \cap B) = P(A) \times P(B)$. To test independence, check this equation.
- Conditional probability: $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$, the probability of A given that B has happened.

Worked example 1

$P(A) = 0.5$, $P(B) = 0.35$, $P(A \cup B) = 0.7$. **Are A and B independent?**

$P(A \cap B) = 0.5 + 0.35 - 0.7 = 0.15$. Independence would need $0.5 \times 0.35 = 0.175$, which is not 0.15, so A and B are **not** independent.

Watch out

"Independent" is not the same as "mutually exclusive". Do not add probabilities for independent events when you need to multiply, and do not use $P(A) + P(B)$ when the events overlap.

Venn diagrams

Method: filling in a Venn diagram

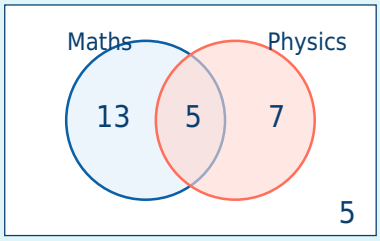
1. Start with the overlap (the "both" region).
2. Subtract it from each set total to get the "only" regions.
3. Subtract all regions from the total to get "neither".

Worked example 2

Of 30 students, 18 study maths, 12 study physics and 5 study both. Find $P(\text{physics} \mid \text{maths})$.

$18 - 5 = 13$ maths only, $12 - 5 = 7$ physics only, $30 - (13 + 5 + 7) = 5$ neither.

$P(\text{maths only}) = \dfrac{13}{30}$, and $P(\text{physics} \mid \text{maths}) = \dfrac{5}{18}$, because we restrict to the 18 who study maths.



Worked example 3

Of 40 pupils, 22 play football, 15 play hockey and 6 play neither. Find $P(\text{hockey} \mid \text{football})$.

34 play at least one sport, so $22 + 15 - x = 34$ and $x = 3$ play both. Then $P(\text{hockey} \mid \text{football}) = \dfrac{3}{22}$.

Tree diagrams

Rule: tree diagrams

Multiply along branches and add across branches. Without replacement the second-stage probabilities change.

Worked example 4

A bag has 5 red and 3 blue counters. Two are taken without replacement. Find $P(\text{both red})$ and $P(\text{one of each})$.

$P(\text{both red}) = \frac{5}{8} \times \frac{4}{7} = \frac{5}{14}$. $P(\text{one of each}) = 2 \times \frac{5}{8} \times \frac{3}{7} = \frac{15}{28}$.

The binomial distribution

Definition: binomial distribution

Use $X \sim B(n, p)$ when there is a fixed number n of trials, each with two outcomes (success or failure), the probability of success p is constant, and the trials are independent.

Rule: binomial probabilities, mean and variance

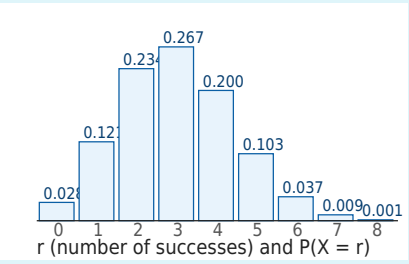
$$P(X = r) = \binom{n}{r} p^r (1-p)^{n-r}, \quad r = 0, 1, \ldots, n.$$

The mean is np and the variance is $np(1-p)$.

Worked example 5

$X \sim B(10, 0.3)$. Find $P(X = 3)$, the mean and the variance.

$P(X = 3) = \binom{10}{3}(0.3)^3(0.7)^7 = 0.267$, mean $= 3$, variance $= 2.1$.



Method: cumulative probabilities

1. Define X and its distribution in words: "Let X be the number of ..., so $X \sim B(n, p)$".
2. Use the calculator's cumulative function for $P(X \leq r)$.
3. Convert: $P(X \geq 4) = 1 - P(X \leq 3)$. "At least one" is $1 - P(\text{none})$.

Worked example 6

90% of seeds germinate. Find $P(\text{all 12 germinate})$, and $P(\text{at least one})$ in 4 trials with $p = 0.2$.

$0.9^{12} = 0.282$. $P(\text{at least one}) = 1 - 0.8^4 = 0.590$.

Worked example 7

How many rolls of a fair die make the chance of at least one six exceed 0.9?

$1 - \left(\frac{5}{6}\right)^n > 0.9$, so $\left(\frac{5}{6}\right)^n$ and $n > \dfrac{\ln 0.1}{\ln(5/6)} = 12.6$. The smallest n is 13.

Worked example 8

A machine makes 5% defective items. In a sample of 20 find $P(X \le 1)$ and the expected number of defectives.

$X \sim B(20, 0.05)$. $P(X \le 1) = 0.95^{20} + 20(0.05)(0.95)^{19} = 0.3585 + 0.3774 = 0.736$. The expected number is $np = 1$.

Note: criticising the model

Comment on the assumptions: are the trials independent (people in one family are not), is p constant, and is the number of trials fixed? State the assumption in the context of the question.

Worked example 9

X takes the values 1, 2, 3, 4 with $P(X = x) = kx$. Find k and $E(X)$.

$k(1 + 2 + 3 + 4) = 1$ so $k = \frac{1}{10}$. $E(X) = \sum xP(X = x) = \dfrac{1 + 4 + 9 + 16}{10} = 3$.

Watch out

Do not forget the $\binom{n}{r}$ coefficient. $P(X$, and $P(X \ge 4) = 1 - P(X \le 3)$, not $1 - P(X \le 4)$.

Worked example 10: Conditional probability from a two-way table

In a class of 30, 18 study French, 12 study Spanish and 5 study both. Find the probability that a student studies French given that they study Spanish.

$P(F \cap S) = \dfrac{5}{30}$ and $P(S) = \dfrac{12}{30}$. So $P(F|S) = \dfrac{5/30}{12/30} = \dfrac{5}{12} = 0.417$. Notice that you can also read it straight from the table: of the 12 Spanish students, 5 also study French. The conditional probability restricts attention to the group named after the vertical bar. Compare with $P(F) = \dfrac{18}{30} = 0.6$: the events are not independent.

Worked example 11: At least one success

The probability that a train is late is 0.15. Find the probability that at least one of 5 trains is late and how many trains are needed to make this probability exceed 0.9.

$P(\text{at least one late}) = 1 - 0.85^5 = 1 - 0.4437 = 0.5563$. For more than 0.9 we need 0.85^n, so $n > \dfrac{\ln 0.1}{\ln 0.85} = 14.2$, so $n = 15$ trains. The trick "1 minus the probability of none" is the quickest method for "at least one" problems because it replaces a long sum with a single power.

Worked example 12: Cumulative binomial probabilities with subtraction

$X \sim B(20, 0.3)$. Given that $P(X \leq 7) = 0.7723$ and $P(X \leq 4) = 0.2375$, find $P(5 \leq X \leq 7)$ and $P(X > 7)$.

$P(5 \leq X \leq 7) = P(X \leq 7) - P(X \leq 4) = 0.7723 - 0.2375 = 0.5348$. Note that you subtract the cumulative probability up to one below the lower limit.
$P(X > 7) = 1 - P(X \leq 7) = 1 - 0.7723 = 0.2277$. The key is to translate every inequality into a form $P(X \leq r)$.

Worked example 13: A tree diagram for three events

The probability of a player winning a point is 0.6 independently. Find the probability that she wins exactly 2 of the next 3 points and compare with a tree diagram.

Binomial: $\binom{3}{2}(0.6)^2(0.4) = 3 \times 0.36 \times 0.4 = 0.432$. Tree diagram: there are three paths with two wins and one loss (WWL, WLW, LWW), each with probability $0.6 \times 0.6 \times 0.4 = 0.144$. The sum is $3 \times 0.144 = 0.432$ ✓. The binomial coefficient counts the paths in the tree that give the same outcome.

Worked example 14: Using the mean and variance to test a model

In 200 trials, the number of successes in blocks of 10 had mean 3.1 and variance 2.2. Is a binomial model $B(10, p)$ reasonable?

For a binomial distribution $p = \dfrac{3.1}{10} = 0.31$, so the variance would be $npq = 10 \times 0.31 \times 0.69 = 2.14$. The observed variance 2.2 is close to 2.14, which supports the binomial model. If the observed variance had been much larger (say 5) the trials would not be independent or the probability would vary, and the model would be questioned.

Worked example 15: Bayes-style reasoning with a table of numbers

In a town of 10 000 people, 1% have a condition. A screening test detects 90% of cases and wrongly flags 5% of healthy people. Of those who test

positive, what proportion really have the condition?

Imagine the 10 000 people. 1% is 100 people with the condition: 90 of them test positive. The other 9 900 are healthy: 5% is 495 false positives. Total positives $= 90 + 495 = 585$. The proportion who truly have the condition is $\frac{90}{585} = 0.154$, about 15%. Counting people instead of multiplying probabilities often makes conditional probability easier to understand, and it shows why screening for rare conditions produces many false alarms.

Worked example 16: Independent events and the multiplication rule

The probability that Ali passes his driving test first time is 0.6 and that Ben passes is 0.7, independently. Find the probability that at least one passes and that exactly one passes.

$P(\text{neither passes}) = 0.4 \times 0.3 = 0.12$, so $P(\text{at least one}) = 1 - 0.12 = 0.88$.
$P(\text{exactly one}) = 0.6 \times 0.3 + 0.4 \times 0.7 = 0.18 + 0.28 = 0.46$. Check: both pass has probability $0.6 \times 0.7 = 0.42$, and $0.12 + 0.46 + 0.42 = 1$ ✓. Using the complement saves work in "at least one" questions.

Worked example 17: Finding an unknown probability

The events A and B are independent with $P(A) = 0.4$ and $P(A \cup B) = 0.7$. Find $P(B)$.

For independent events
$P(A \cup B) = P(A) + P(B) - P(A)P(B) = 0.4 + P(B) - 0.4P(B) = 0.4 + 0.6P(B)$.
Setting this equal to 0.7: $0.6P(B) = 0.3$, so $P(B) = 0.5$. Check:
$P(A \cap B) = 0.4 \times 0.5 = 0.2$ and $P(A \cup B) = 0.4 + 0.5 - 0.2 = 0.7$ ✓.

Method: a conditional probability question

1. Identify the event after the vertical bar: this is the reduced sample space.
2. Find $P(A \cap B)$ using a Venn diagram, a table or a tree.
3. Divide by $P(B)$ to get $P(A|B) = \frac{P(A \cap B)}{P(B)}$.
4. Check against a direct count in the reduced sample space.
5. Test independence by comparing $P(A \cap B)$ with $P(A)P(B)$.

Watch out: common probability errors

- Adding probabilities of events that are not mutually exclusive without subtracting the overlap.
- Multiplying probabilities along a tree for events that are not independent without updating the second-stage probabilities.
- Writing $P(X \geq 3) = 1 - P(X \leq 3)$ instead of $1 - P(X \leq 2)$.

- Using the binomial distribution without checking the conditions (fixed n, independence, constant p).
- Confusing $P(A|B)$ with $P(B|A)$.

Note: the language of probability

An **event** is a set of outcomes. The **complement** A' contains all outcomes not in A. **Mutually exclusive** events cannot happen together ($P(A \cap B) = 0$) while **independent** events do not influence each other ($P(A \cap B) = P(A)P(B)$). A **random variable** X takes numerical values with a probability distribution; for the binomial distribution $B(n, p)$ the values are $0, 1, \ldots, n$. When you write an answer, always state which event you are finding, define the variable and give the answer as a fraction or a decimal to at least 3 decimal places.

Did you know?

Probability theory began in 1654 when the gambler Antoine Gombaud asked Blaise Pascal about a problem in a game of dice, and Pascal and Pierre de Fermat exchanged letters that founded the subject. The binomial distribution was first worked out by Jacob Bernoulli, whose book *Ars Conjectandi* was published in 1713. Today the same mathematics is used to decide how many bulbs to test in a factory, how reliable a medical test is, and whether a new medicine works better than a placebo.

Otto's tip

"At least one" is $1-$"none". It turns a long sum into one short calculation.