Discrete distributions, including the discrete uniform and the binomial

S4.1 · Statistics · revision notes

Discrete random variables given by a table or a formula, the discrete uniform distribution, and the binomial distribution $B(n, p)$ as a model: its conditions, individual and cumulative probabilities from your calculator, and criticising the model in context.

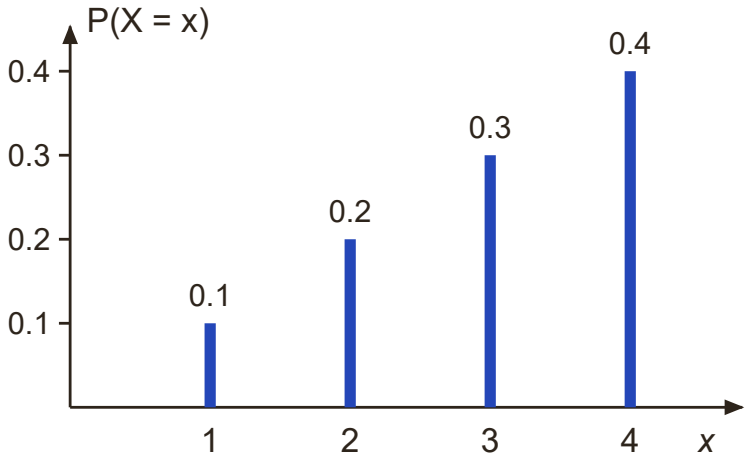
Key facts

- $\Sigma P(X = x) = 1$
- $X \sim B(n, p)$: $P(X = r) = \binom{n}{r} p^r (1-p)^{n-r}$
- $P(X \geq r) = 1 - P(X \leq r - 1)$
- Binomial mean np, variance $np(1-p)$
- Discrete uniform over n values: $P(X = x) = \frac{1}{n}$
- Binomial conditions: fixed n, two outcomes, independent trials, constant p
- $P(X > r) = 1 - P(X \leq r)$

Notes

Discrete random variables

- A **discrete random variable** X takes separate values (often counts), each with a probability.
- Its probability distribution can be a table or a formula, e.g. $P(X = x) = kx$ for $x = 1, 2, 3, 4$. The probabilities add up to 1: $\Sigma P(X = x) = 1$



P(X = x) = kx with k = 1/10: the probabilities add up to 1

- The cumulative probability $P(X \leq x)$ is the sum of the probabilities up to and including x
- A **discrete uniform distribution** has n equally likely values, each with probability $\frac{1}{n}$, e.g. the score on a fair die.
- In A level Mathematics you are not asked to calculate the mean or variance of a general discrete random variable.

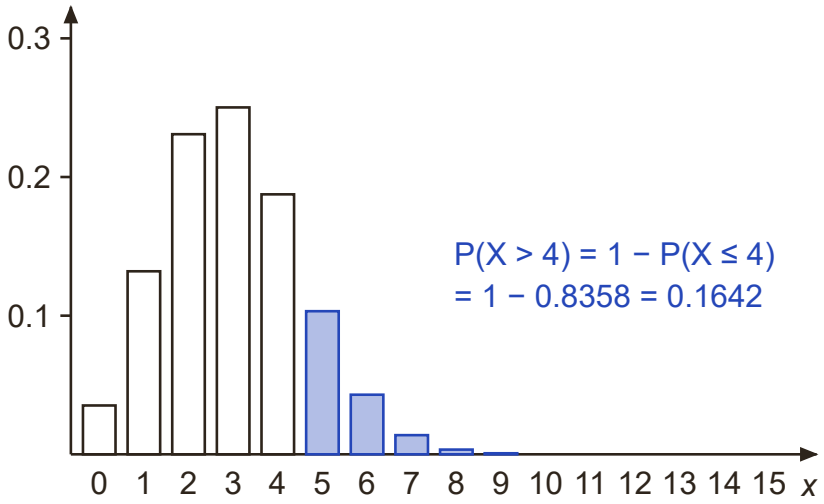
The binomial distribution

- $X \sim B(n, p)$ models the number of successes in n trials when: there is a fixed number of trials; each trial has two outcomes (success or failure); the trials are independent; the probability of success, p, is the same for every trial.
- $P(X = r) = \binom{n}{r} p^r (1-p)^{n-r}$ for $r = 0, 1, \ldots, n$. Your calculator gives $P(X = r)$ (binomial PD) and $P(X \leq r)$ (binomial CD).
- The mean (expected number of successes) is np and the variance is $np(1-p)$

Cumulative probabilities

$$P(X < r) = P(X \leq r - 1)$$

- $P(X > r) = 1 - P(X \leq r)$ and $P(X \geq r) = 1 - P(X \leq r - 1)$



X ~ B(15, 0.2): P(X > 4) is everything from 5 upwards

- $P(a < X \leq b) = P(X \leq b) - P(X \leq a)$ and
  $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a - 1)$

Criticising a binomial model

- Independence can fail when trials influence each other, e.g. illness spreading within a class.
- p may not be constant, e.g. when conditions change, or when sampling without replacement from a small population. When sampling from a large population, p can be treated as constant.

How to answer each type of question

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

Find an unknown constant and probabilities

2 to 4 marks | Grade E

1. Write out $P(X = x)$ for every value in terms of k
2. Add them and set the total equal to 1.
3. Add the probabilities of the values you need.

Example. The discrete random variable X has probability distribution $P(X = x) = k(x + 1)$ for $x = 1, 2, 3, 4$.
(a) Find the value of k.
(b) Find $P(X \geq 3)$

Model answer.

(a) $k(2 + 3 + 4 + 5) = 1$ (M1), so $14k = 1$ and $k = \frac{1}{14}$ (A1)

(b) $\frac{4 + 5}{14} = \frac{9}{14}$ (B1)

Model with $B(n, p)$ and find probabilities

6 to 8 marks | Grade C

1. Define X and write its distribution, e.g. $X \sim B(15, 0.2)$
2. State assumptions in context: independence and a constant probability.
3. Rewrite each inequality in terms of $P(X = r)$ or $P(X \leq r)$
4. Use your calculator and give answers to 4 decimal places.

Example. 20% of seeds of a certain type fail to germinate. A gardener plants 15 of these seeds, and X is the number of seeds that fail to germinate.
(a) State two assumptions needed to model X by $B(15, 0.2)$.
(b) Find $P(X = 3)$.
(c) Find $P(X > 4)$.
(d) Find $P(2 \leq X \leq 5)$

Model answer.

(a) The seeds germinate or fail independently of each other (B1); each seed has the same probability, 0.2, of failing (B1)

(b) $\binom{15}{3}(0.2)^3(0.8)^{12}$ (M1) $= 0.2501$ (A1)

(c) $1 - P(X \leq 4) = 1 - 0.8358$ (M1) $= 0.1642$ (A1)

(d) $P(X \leq 5) - P(X \leq 1)$
$= 0.9389 - 0.1671$ (M1) $= 0.7718$ (A1)

Criticise a binomial model

1 to 2 marks | **Grade B**

1. Pick the condition most likely to fail in this context.
2. Explain why, referring to the situation.

Example. On a typical day, 10% of the students at a school are absent. A teacher models the number of students absent from a class of 30 on a given day by $B(30, 0.1)$. Give two reasons why this model may not be suitable.

Model answer.

Absences may not be independent: illnesses such as colds spread between students in the same class (B1)

The probability of being absent may not be the same for every student or every day, e.g. there may be more absences in winter (B1)

Find n from a probability condition

3 to 4 marks | **Grade A**

1. Write the condition, often $1 - P(X = 0) >$ a given value.
2. Rearrange to $(1 - p)^n <$ a value.
3. Take logs: the inequality reverses when you divide by the log of a number less than 1.
4. Round up to the next whole number and check.

Example. Each component made by a machine is faulty with probability 0.04, independently of other components. Find the smallest number of components that must be checked for the probability of finding at least one faulty component to be greater than 0.9.

Model answer.

$1 - 0.96^n > 0.9$ (M1), so $0.96^n < 0.1$

$$n > \frac{\log 0.1}{\log 0.96} = 56.4$$ (M1)

$n = 57$ (A1)

Don't lose marks

- Using $P(X \leq 4)$ for 'fewer than 4' (it should be $P(X \leq 3)$).
- Using $1 - P(X \leq 5)$ for $P(X \geq 5)$ (it should be $1 - P(X \leq 4)$).
- Stating the binomial conditions without context.
- Forgetting to reverse the inequality when dividing by the logarithm of a number less than 1.

More tips

Memory tricks

- Words to inequalities: ‘at least 3’ is $X \geq 3$; ‘more than 3’ is $X \geq 4$; ‘fewer than 3’ is $X \leq 2$; ‘at most 3’ is $X \leq 3$
- Many calculators' binomial CD only gives $P(X \leq r)$, so rewrite everything in that form first.
- Swapping to failures: if $X \sim B(n, p)$ counts successes, the number of failures, $n - X$, is $B(n, 1-p)$

Exam technique

- Write the distribution and the probability statement, e.g. $X \sim B(15, 0.2)$ and $P(X > 4) = 1 - P(X \leq 4)$, before the answer: they earn method marks.
- Give probabilities to 4 decimal places (or 3 significant figures or better) unless told otherwise.
- The two assumptions usually wanted are independence and a constant probability, both in context.

What each grade needs

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

Grade E — **Use the fact that probabilities sum to 1**
Solve $\Sigma P(X = x) = 1$ to find an unknown such as k

Grade E — **Recognise a discrete uniform distribution**
Each of the n possible values has probability $\frac{1}{n}$, e.g. the score on a fair die.

Grade D — **Calculate a binomial probability for one value**
Use $P(X = r) = \binom{n}{r} p^r (1-p)^{n-r}$ or your calculator's binomial PD.

Grade C — **Use cumulative binomial probabilities with inequalities**
Rewrite ‘more than’, ‘at least’ and ‘between’ in terms of $P(X \leq r)$

Grade C — **State binomial conditions and assumptions in context**
Fixed number of trials, two outcomes, independent trials, constant probability of success.

Grade B — **Criticise a binomial model in context**
Explain why independence or a constant probability may not hold.

Grade A — **Find n or p from a probability condition**
E.g. solve $1 - (1-p)^n > 0.9$ using logarithms.

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.