Geometric and Negative Binomial Distributions Cheat Sheet

Edexcel FS1

The geometric distribution

If you are carrying out a fixed number of independent trials each with a constant probability of success, p, then you can model the number of trials needed to achieve one success using a geometric distribution. We would write that:

- $X \sim Geo(p)$, where X represents the number of trials needed to achieve one success and p is the probability of success for each trial.

In order for a geometric distribution to be a suitable model:

- trials must be independent of each other,
- the probability of success should be the same for each trial.

These conditions also apply to the binomial distribution.

You can use the following functions to calculate probabilities:

- $P(X = x) = p(1-p)^{x-1}$
- $P(X \le x) = 1-(1-p)^x$
- $P(X \ge x) = (1-p)^{x-1}$

Example 1: Bernhard has to pass an examination to get into law school. He can take the examination as many times as he likes and his probability of passing on any one attempt is 0.3.
(a) Find the probability that:
(i) he passes on his third attempt
(ii) he takes at least four attempts to pass
(b) State two assumptions that you have used in your model

ai) Use $P(X = x)$ with $p = 0.3, x = 3$.	$P(X = 3) = 0.3(1-0.3)^{3-1} = 0.3(0.7)^2 = 0.147$
ii) Use $P(X \ge x)$ with $p = 0.3, x = 4$.	$P(X \ge 4) = (1-0.3)^{4-1} = 0.7^3 = 0.343$
b) State the assumptions for a geometric distribution.	Trials are independent of each other and have the same probability of success (0.3).

Example 2: $X \sim Geo(0.032)$.
a) Given that $P(X = x) = 0.0203\ (4\ d.p.)$, find the value of x.
b) Find the largest value of x such that $P(X \le x) < 0.1$.

a) Use $P(X = x)$ with $p = 0.032, x = x$.	$P(X = x) = 0.032(1-0.032)^{x-1} = 0.0203$
Simplify.	$0.968^{x-1} = 0.6344$
Take logs of both sides.	$\ln 0.968^{x-1} = \ln 0.6344$
Use the power rule for logs.	$(x-1)\ln 0.968 = \ln 0.6344$
Solve for x.	$x - 1 = \frac{\ln 0.6344}{\ln 0.968} = 13.994\ldots$ $x = 1 + 13.994 = 14.994$
State the final answer.	So $x = 15$.
b) Set $P(X \le x)$ less than 0.1	$P(X \le x) = 1-(1-p)^x = 1 - 0.968^x < 0.1$
Simplify.	$0.9 < 0.968^x$
Take logs of both sides:	$\ln 0.9 < \ln 0.968^x$
Use the power rule for logs:	$\ln 0.9 < x \ln 0.968$
Divide by $\ln 0.968$: Since $\ln 0.968 < 0$, the inequality flips.	$x < \frac{\ln 0.9}{\ln 0.968} = 3.24$
Round down to the nearest integer:	So $x_{max} = 3$

Mean and variance for a geometric distribution

If $X \sim Geo(p)$, then

- Mean of $X = E(X) = \frac{1}{p}$
- Variance of $X = Var(X) = \frac{1-p}{p^2}$

Example 3: Uma has a bag of marbles, 15% of which are blue. She puts her hand in the bag and pulls out a marble at random. If the marble is not blue, she puts it back in the bag and tries again.
(a) Calculate:
(i) the mean
(ii) the variance of the number of marbles she pulls out, up to and including the first blue one.
Calculate the probability that:
(b) she pulls out 4 marbles
(c) she pulls out at least 8 marbles

ai) We are told $p = 0.15$.	$E(X) = \frac{1}{0.15} = 6.67\ (3\ s.f.)$
ii) Use $Var(X) = \frac{1-p}{p^2}$ with $p = 0.15$:	$Var(X) = \frac{1-0.15}{0.15^2} = 37.8\ (3\ s.f.)$
b) The question is simply asking to find $P(X = 4)$.	$P(X = 4) = 0.15(1-0.15)^3 = 0.15(0.85)^3 = 0.09$
c) This question is simply asking to find $P(X \ge 8)$.	$P(X \ge 8) = (1-0.15)^{8-1} = 0.321$

Example 4: Wilma is a charity collector and goes door-to-door trying to raise money. Given that the probability of her getting at donation at each house is p, that each house call is independent and the variance is 380, find:
(a) p
(b) the expected number of house calls Wilma must make before getting a donation.

a) We are told $Var(X) = 380$.	$Var(X) = \frac{1-p}{p^2} = 380$
Rearrange the equation.	$380p^2 = 1 - p$ $380p^2 + p - 1 = 0$
Solve the quadratic by quadratic formula.	$p = 0.05$ or $p = -0.05$
p is a probability and so must be non-negative.	$p \ge 0$ so $p = 0.05$.
b) This question is really just asking us to find the expected value of X. Using $E(X) = \frac{1}{p}$:	$E(X) = \frac{1}{0.05} = 20$

The negative binomial distribution

If you are again carrying out a fixed number of independent trials each with a constant probability of success p, then you can model the number of trials needed to achieve r successes using a negative binomial distribution. We would write that:

- $X \sim NB(r, p)$, where X represents the number of trials needed to achieve r successes.

You may also see written $X \sim Negative\ B(r, p)$

In order for a negative binomial distribution to be a suitable model:

- trials must be independent of each other,
- the probability of success should be the same for each trial.

You can use the probability mass function to calculate probabilities:

- $P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$

Questions that involve the use of the negative binomial distribution will often include parts where you also need to use the regular binomial distribution. You need to know when to use each distribution. To find probabilities of the form $P(X \le x)$ or $P(X \ge x)$, you will need to instead consider a binomial distribution (see example 5).

Example 5: The random variable X has negative binomial distribution, Negative B(5, 0.7). Find:
(a) $P(X = 10)$
(b) $P(X \le 6)$
(c) $P(X > 12)$

a) Use $P(X = x)$ with $p = 0.7, x = 10, r = 5$.	$P(X = 10) = \binom{9}{4} 0.7^5 (0.3)^5 = 0.051$
b) $P(X \le 6)$ just means find the probability that there are 5 successes in 6 trials or less. So, we use a binomial distribution with $n = 6, p = 0.7$ and find $P(X \ge 5)$.	Let Y represent the number of successes in 6 trials. Then $Y \sim B(6, 0.7)$. $P(required) = P(Y \ge 5) = 1 - P(Y \le 4) = 1 - 0.580 = 0.420$
c) $P(X > 12)$ means find the probability it takes more than 12 trials to achieve 5 successes. In other words, find the probability there are less than 5 successes in 12 trials. This is easily found using a binomial distribution.	Let X' represent the number of successes in 12 trials. Then $X' \sim B(12, 0.7)$. $P(required) = P(X' \le 4) = 0.009$

Example 6: Denise takes part in a multiple-choice quiz where she picks the answers at random. Given that her probability of picking any correct answer is 0.25, find the probability that:
(a) She gets exactly two correct answers in the first ten questions.
(b) She picks her fourth correct answer on her seventh question.
(c) Her third correct answer occurs on or before the tenth question.

a) We need to use a negative binomial distribution here. We use $r = 2$ since there are 2 successes in 10 trials.	Let X represent the number of questions needed to achieve 2 correct answers. Then $X \sim NB(2, 0.25)$.
Use the probability mass function for the NB distribution ($r = 2, p = 0.25$).	$P(required) = P(X = 10) = \binom{9}{1} 0.25^2 (0.75)^8 = 0.056$
b) We do not use a NB distribution here because we are looking for the fourth success to occur specifically on the seventh question.	Let Y be the number of correct answers in 6 questions. Then $Y \sim B(6, 0.25)$ $P(required) = P(3\ correct\ in\ first\ 6) \times 0.25$
Use the binomial distribution for the first 6 questions to find $P(3\ correct\ in\ 6\ questions)$	$P(Y = 3) = \binom{6}{3} 0.25^3 (0.75)^3 = 0.132$
Then multiply by 0.25 (the probability of success in the next trial).	$P(required) = 0.25 \times 0.185 = 0.033$
c) Again, we need to use a binomial distribution since we aren't simply looking for the probability of r successes in x trials.	Let Z be the number of correct answers in the first 10 questions. $Z \sim B(10, 0.25)$
We need to find the probability there are at least 3 correct answers in the first 10 questions.	$P(required) = P(Z \ge 3) = 1 - P(Z \le 2) = 1 - 0.526 = 0.474$

Mean and variance for a negative binomial distribution

If $X \sim NB(r, p)$ then

- Mean of $X = E(X) = \frac{r}{p}$
- Variance of $X = Var(X) = \frac{r(1-p)}{p^2}$

These are identical to the mean/variance for the geometric distribution but multiplied by r.

Example 7: Kelly and her classmates are taking part in a competition where students take turns attempting to solve a puzzle. The probability that each student solves the puzzle is 0.7. The random variable X represents the number of students who need to attempt the puzzle before five have solved it. Using a negative binomial model, find the mean and variance of X.

Use $P(X = x)$ with $p = 0.3, x = 3$.	Let X be the number of trials needed before 5 students solve it. Then $X \sim NB(5, 0.7)$
Use $E(X) = \frac{r}{p}$	$E(X) = \frac{5}{0.7} = 7.14$
Use $Var(X) = \frac{r(1-p)}{p^2}$	$Var(X) = \frac{5(1-0.7)}{0.7^2} = 3.06$

Example 8: The random variable X has negative binomial distribution with mean 6 and variance 3. Find:
(a) the value of p and the value of r.
(b) $P(X = 7)$

a) Set the mean equal to 6 and variance equal to 3.	$\frac{r}{p} = 6 \therefore r = 6p$ [1] $\frac{r(1-p)}{p^2} = 3$ [2]
Now we need to solve [1] and [2] simultaneously. Substitute [1] into [2].	$\frac{6p(1-p)}{p^2} = 3$
Simplify.	$6(1-p) = 3p$ $9p = 6$ $\therefore p = \frac{2}{3}, r = 6p = 4$
b) Use $P(X = x)$ with $p = \frac{2}{3}, x = 7,\ r = 4$.	$P(X = 4) = \binom{6}{3} \frac{2}{3}^4 \left(\frac{1}{3}\right)^3 = 0.146$