STEP Support Programme

Statistics STEP Questions

This is a selection of old STEP 1 and STEP 2 questions which are based on content from the new (2019 onwards) STEP 1 specification.

2010 S1 Q12

This question is about the expectation of a discrete random variable, which used to be on most A-level specifications. The definition of expectation is given in question **1(i)**.

1 Preparation

(i) The expectation of a discrete random variable is defined by:

$$\mathrm{E}(X) = \sum n \times \mathrm{P}(X = n)$$

where the sum is taken over all possible values of n.

The discrete random variable X is defined by:

$$\mathrm{P}(X = x) = \begin{cases} kx & \text{for } x = 1, 2, 3, 4, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the value of k.

(b) Find $\mathrm{P}(X \geqslant 3)$.

(c) Find $\mathrm{E}(X)$.

(ii) I roll a normal six-sided die[1] until I roll a 6; then I stop.

(a) Find the probability that I roll a 6 on the third roll.

(b) Find that probability that I need more than five rolls to get a 6.

You could do $1 - \big(\mathrm{P}(X = 1) + \mathrm{P}(X = 2) + \cdots + \mathrm{P}(X = 5)\big)$ but it is easier to think about what must have happened on the first five rolls if you need more than five rolls to get a 6.

(iii) Find the value of the infinite sum $2 + 0.2 + 0.02 + \cdots$.

(iv) Find the maximum value of $y = x(4 - x)$.

[1] Apparently, the use of 'die' as the singular form of 'dice' is normal in America and in maths books, and certainly on STEP papers. But, at least in conversation, the British never say die.

2 The Probability question

A discrete random variable X takes only positive integer values. Define $\mathrm{E}(X)$ for this case, and show that

$$\mathrm{E}(X) = \sum_{n=1}^{\infty} \mathrm{P}\,(X \geqslant n)\,.$$

I am collecting toy penguins from cereal boxes. Each box contains either one daddy penguin or one mummy penguin. The probability that a given box contains a daddy penguin is p and the probability that a given box contains a mummy penguin is q, where $p \neq 0$, $q \neq 0$ and $p + q = 1$.

Let X be the number of boxes that I need to open to get at least one of each kind of penguin. Show that $\mathrm{P}(X \geqslant 4) = p^3 + q^3$, and that

$$\mathrm{E}(X) = \frac{1}{pq} - 1.$$

Hence show that $\mathrm{E}(X) \geqslant 3$.

Discussion

Even if you cannot do the first part of the question, you can still do the rest. It may be helpful to note that, for example, $3 \times \mathrm{P}\,(X = 3) = \mathrm{P}\,(X = 3) + \mathrm{P}\,(X = 3) + \mathrm{P}\,(X = 3)$.

For $\mathrm{P}\,(X \geqslant 4)$ you could consider $1 - \mathrm{P}\,(X \leqslant 3)$ but this is probably not the easiest way.

For $\mathrm{E}(X)$ be careful with the $n = 1$ case, it does not fit the same pattern as the other values of n.

2009 S1 Q13

Arrangements and combinatorial arguments are not part of A-level maths specifications, but are on the STEP 1 specification. Often they are used in probability questions but you might also need them in other types of questions.

3 **Preparation**

This question involves *arrangements*, and you may like to look back to Assignment 6 before you start. You can leave big numbers in terms of factorials.

(i) 6 women (A, B, C, D, E and F) stand in a line. In how many ways can they arrange themselves in this line?

There are 6 different positions for A, then 5 for B etc.

(ii) The 6 women are joined by 4 men (G, H, I and J). In how many ways can the 10 people arrange themselves?

(iii) The 4 men wish to stand together. We can imagine them being together inside a rope ring.

(a) In how many ways can the 4 men arrange themselves in a line inside their rope?

(b) Imagine now that the rope is one "Person" so that there are 7 "people" i.e. the 6 women and the 1 roped-off collection of men. How many ways can we arrange these 7 "people"?

(c) How many ways are there of arranging the 10 people so that the men stand together?

(d) Show that the probability that all 4 men stand together is 1/30.

(iv) What is the probability that all the women (but not necessarily the men) stand together?

(v) What is the probability that all the men stand together and all the women stand together?

Note that this is **not** the previous two answers multiplied together as the events are not independent.

(vi) Find the probability that no men stand together by following these steps.

(a) First arrange the 6 women. How many ways can we do this?

(b) There are now 7 gaps between the women which we can slot the first man into, like this:

$$\downarrow W_1 \downarrow W_2 \downarrow W_3 \downarrow W_4 \downarrow W_5 \downarrow W_6 \downarrow$$

where the arrows represent the gaps. Once a man fills up a gap then it is no longer a gap and there are only 6 gaps for the second man. Find the number of ways you can arrange the 4 men in the 7 gaps.

(c) Find hence find the probability that no two men are together.

(vii) Find the probability that there will be a woman at each end of the line, by following these steps.

(a) How many different ways can we choose the two women to be at the ends? (Consider how many choices you have for the first end and then how many for the second end).

(b) How many ways can you arrange the 8 people left?

(c) Use your two previous answers to find the probability that there is a woman at each end of the line.

4 **The Probability question**

I seat n boys and 3 girls in a line at random, so that each order of the $n+3$ children is as likely to occur as any other. Let K be the maximum number of consecutive girls in the line so, for example, $K = 1$ if there is at least one boy between each pair of girls.

(i) Find $\mathrm{P}(K = 3)$.

(ii) Show that

$$\mathrm{P}(K = 1) = \frac{n(n-1)}{(n+2)(n+3)}\,.$$

(iii) Find $\mathrm{E}(K)$.

Discussion

You might like to simplify your final answer, though the question doesn't ask you to do so. More overleaf ...

If you found the last question interesting, you might like to try this STEP question as well:

5 **1995 S1 Q12** A school has n pupils, of whom r play hockey, where $n \geqslant r \geqslant 2$. All n pupils are arranged in a row at random.

(i) What is the probability that there is a hockey player at each end of the row?

(ii) What is the probability that all the hockey players are standing together?

(iii) By considering the gaps between the non-hockey-players, find the probability that no two hockey players are standing together, distinguishing between cases when the probability is zero and when it is non-zero.

Discussion

The last phrase "distinguishing between cases when the probability is zero and when it is non-zero" is a rather poorly phrased attempt to point out that for some values of n and r the probability is zero, i.e. it is not possible for all the hockey players to be separated. A little bit of thinking should give a relationship between n and r for when the hockey players cannot possibly be separated.

The number of ways of choosing r objects from a total of n different objects is $\dfrac{n!}{r!(n-r)!}$, which is written as ${}^nC_r$ or $\binom{n}{r}$. You may find this notation useful.

1999 S2 Q12

6 Preparation

(i) Sketch the graph $y = \dfrac{3x}{5-2x}$.

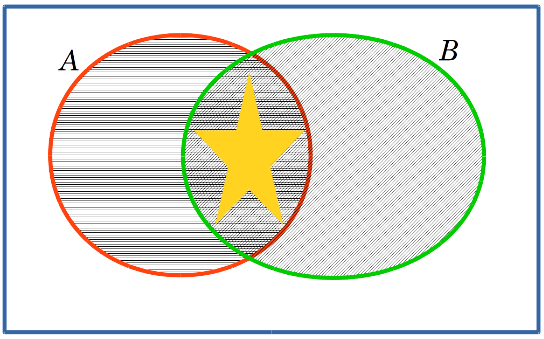
Find the maximum and minimum values of y if x lies in the interval $-1 \leqslant x \leqslant 2$.

To find the horizontal asymptote it may be helpful to divide both the numerator and denominator by x.

(ii) Sketch the graph $y = \dfrac{1-x}{5-2x}$.

Find the maximum and minimum values of y if x lies in the interval $-1 \leqslant x \leqslant 2$.

(iii) The diagram below is called a *Venn diagram*. In a Venn diagram, probabilities are represented by areas, and the blue rectangle (representing the whole sample space) has area 1.



The area in the red circle represents P(A) and the area in the green circle. represents P(B). The probability of A **and** B both happening, which is written as $\text{P}(\text{A} \cap \text{B})$, is represented by the area which lies in both the red and the green circles (the bit with a yellow star in it).

The probability that either A or B, or both, occur is written as $\text{P}(\text{A} \cup \text{B})$ and is represented by the whole area covered by the red and green circles. If A and B were *mutually exclusive* then the two circles do not overlap and $\text{P}(\text{A} \cup \text{B}) = \text{P}(\text{A}) + \text{P}(\text{B})$. If A and B are not mutually exclusive, then we have to ensure that we do not count the intersection twice, so:

$$\text{P}(\text{A} \cup \text{B}) = \text{P}(\text{A}) + \text{P}(\text{B}) - \text{P}(\text{A} \cap \text{B})\,.$$

The same diagram can be used to find conditional probabilities. In order to find the conditional probability that A happens given that B happens, we look at the fraction of the area of the B circle that is also covered by the A circle. This gives the conditional probability of A happening given that B happens as:

$$\text{P}(\text{A}|\text{B}) = \frac{\text{P}(\text{A} \cap \text{B})}{\text{P}(\text{B})}\,.$$

In order to calculate conditional probabilities when events A and B are not independent, we use:

$$\text{P}(\text{A}|\text{B}) = \frac{\text{P}(\text{A} \cap \text{B})}{\text{P}(\text{B})} \quad \text{and} \quad \text{P}(\text{B}|\text{A}) = \frac{\text{P}(\text{A} \cap \text{B})}{\text{P}(\text{A})}$$

to give:

$$\text{P}(\text{A}|\text{B})\text{P}(\text{B}) = \text{P}(\text{B}|\text{A})\text{P}(\text{A})$$

and hence:

$$\text{P}(\text{A}|\text{B}) = \frac{\text{P}(\text{B}|\text{A})\text{P}(\text{A})}{\text{P}(\text{B})}$$

This last result is known as **Bayes' Theorem**.

(a) Given that $\text{P}(\text{A}) = 0.3$, $\text{P}(\text{B}) = 0.5$, that the events A and B are independent, find $\text{P}(\text{A} \cup \text{B})$. Find also $\text{P}(\text{A}|\text{B})$ and $\text{P}(\text{B}|\text{A})$.
If the events are *independent* then $\text{P}(\text{A} \cap \text{B}) = \text{P}(\text{A}) \times \text{P}(\text{B})$.

(b) Events A and B (not necessarily independent) satisfy:

$$\text{P}(\text{A}) = 0.6, \quad \text{P}(\text{B}) = 0.5, \quad \text{P}(\text{B}|\text{A}) = 0.4\,.$$

Find: $\text{P}(\text{A}|\text{B})$ and $\text{P}(\text{A} \cup \text{B})$.

(c) 1% of Martians at age forty who participate in routine screening for tentacle rot actually have tentacle rot. It is known that 80% of Martians with tentacle rot will get a positive result in the screening. 10% of Martians without tentacle rot will also get a positive result. A Martian of age forty had a positive result in a routine screening. What is the probability that she actually has tentacle rot?
You can do this is various ways.

- You could just consider a population of 1000 Martians.
- You could draw a Venn diagram.
- You could use Bayes' theorem, in which case you might find the result $\text{P}(\text{B}) = \text{P}(\text{B}|\text{A})\text{P}(\text{A}) + \text{P}(\text{B}|\text{A}')\text{P}(\text{A}')$ useful. (Here A' is the *complement* of A, i.e. the event "A does not happen".

We suggest that you try all three ways and see which you like best.

7 The Probability question

It is known that there are three manufacturers A, B, C, who can produce micro chip MB666. The probability that a randomly selected MB666 is produced by A is $2p$, and the corresponding probabilities for B and C are p and $1-3p$, respectively, where $0 \leqslant p \leqslant \frac{1}{3}$. It is also known that 70% of MB666 micro chips from A are sound and that the corresponding percentages for B and C are 80% and 90%, respectively.

Find in terms of p, the conditional probability, P(A|S), that if a randomly selected MB666 chip is found to be sound then it came from A, and also the conditional probability, P(C|S), that if it is sound then it came from C.

A quality inspector took a random sample of one MB666 micro chip and found it to be sound. She then traced its place of manufacture to be A, and so estimated p by calculating the value of p that corresponds to the greatest value of P(A|S). A second quality inspector also a took random sample of one MB666 chip and found it to be sound. Later he traced its place of manufacture to be C and so estimated p by applying the procedure of his colleague to P(C|S).

Determine the values of the two estimates and comment briefly on the results obtained.

Discussion

A lot of reading. But you shouldn't be put off by questions that look long on the page: often (and especially for mechanics and probability), they are long because there are quite a few things to say that take up space, but hardly need saying (stuff about no wind resistance, randomness, etc); and sometimes it is because some new concept is being explained — once you have grasped the concept, the question becomes fairly straightforward.

2006 S2 Q13

8 Preparation

(i) A group of 5 children, aged 2, 4, 6, 8 and 10, line up randomly (so that each order in the line is equally likely) for birthday cake.

(a) Find the probability that the youngest child is first in the queue.

(b) Find the probability that the second child in the queue is the oldest given that the first child is the youngest.

(ii) I create a four-digit number by placing the digits $1, 2, 3$ and 4 in random order (each order being equally likely).

(a) How many such numbers can I create?

(b) How many of these numbers are divisible by 4?
There are various ways of doing this — the low-tech approach is to write out all the different options.

(c) What is the probability that I create a number bigger than 2413?

(d) If the first digit of my number is 1, what is the probability that the number is divisible by 4?

9 I know that ice-creams come in n different sizes, but I don't know what the sizes are. I am offered one of each in succession, in random order. I am certainly going to choose one - the bigger the better - but I am not allowed more than one. My strategy is to reject the first ice-cream I am offered and choose the first one thereafter that is bigger than the first one I was offered; if the first ice-cream offered is in fact the biggest one, then I have to put up with the last one, however small.

Let $\mathrm{P}_n(k)$ be the probability that I choose the kth biggest ice-cream, where $k = 1$ is the biggest and $k = n$ is the smallest.

(i) Show that $\mathrm{P}_4(1) = \frac{11}{24}$ and find $\mathrm{P}_4(2)$, $\mathrm{P}_4(3)$ and $\mathrm{P}_4(4)$.

(ii) Find an expression for $\mathrm{P}_n(1)$.

Discussion

The expression in part **(ii)** will be in the form of a sum.

2008 S2 Q13

10 Preparation

(i) I have a row of 5 current buns, one of which has a cherry on top. If I close my eyes and pick two at random, what is the probability that I pick the bun with a cherry on top?

(ii) Now I have n current buns, of which one has a cherry on top. If I pick r buns, what is the probability that I pick the one with a cherry on top?

You are asked to do the same thing below in part **(iv)**, but here you should use a simple argument.

(iii) The number of ways of choosing r objects from a total of n different objects is $\frac{n!}{r!(n-r)!}$, which is written as ${}^nC_r$ or $\binom{n}{r}$. These are also binomial coefficients i.e. the coefficient of x^r in the expansion of $(1+x)^n$.

A team of 5 has to be chosen from a class of 10 swimmers, of whom exactly 2 are male. Find the probability that a team of 5 chosen at random (the probability of each choice being equally likely) will include the 2 males.

You will need to find the total number of ways of choosing 5 from the class of 10, and also (to work out the number of these in which there are two males) the number of ways of picking 3 from the remaining 8 once you have picked the 2 males.

(iv) Use the ${}^nC_r$ notation to find the probability that, if I pick r current buns from a total of n, I get the one with a cherry on top.

You need to consider the total number of ways of picking r buns out of the n as well as the number of ways that include the 1 with a cherry on top.

(v) I have n chocolates of which 2 are filled with chilli and the rest are filled with caramel. If I pick r at random, show that the probability that I manage to avoid the chilli ones is $\dfrac{(n-r)(n-r-1)}{n(n-1)}$.

(vi) Show that ${}^nC_r = {}^{n-1}C_r + {}^{n-1}C_{r-1}$.

This can be done either by manipulating factorials or by a clever "choosing" argument. We suggest that you try both ways. Note that this is the relationship that we see in Pascal's triangle: an entry in a given row is formed by adding together the two closest entries in the row above.

(vii) The sequence $a_0, a_1, \ldots, a_n$ has the property that $a_i/a_{i-1} > 1$ for $i \leqslant k$ and $a_i/a_{i-1} < 1$ for $i > k$. What is the largest term of the sequence?

11 The Probability question

Bag P and bag Q each contain n counters, where $n \geqslant 2$. The counters are identical in shape and size, but coloured either black or white. First, k counters $(0 \leqslant k \leqslant n)$ are drawn at random from bag P and placed in bag Q. Then, k counters are drawn at random from bag Q and placed in bag P.

(i) If initially $n-1$ counters in bag P are white and one is black, and all n counters in bag Q are white, find the probability in terms of n and k that the black counter ends up in bag P.

Find the value or values of k for which this probability is maximised.

(ii) If initially $n-1$ counters in bag P are white and one is black, and $n-1$ counters in bag Q are white and one is black, find the probability in terms of n and k that the black counters end up in the same bag.

Find the value or values of k for which this probability is maximised.

2015 S2 Q12

12 Preparation

(i) Suppose that we want to make a choice from 3 options, P, Q and R, but only have a fair coin to do this.

One option is to toss the coin twice and then:

- if we get HH choose P,
- if we get HT choose Q,
- if we get TH choose R,
- and if we get TT then repeat, i.e. toss the coin twice again.

(a) What is the probability that we choose P on the first two tosses of the coin?

(b) What is the probability that we don't make a choice on the first two tosses, but choose P on the second two tosses?

(c) Write down an infinite sum for the probability that we eventually choose P and hence find the probability that we choose P. Comment on your answer.

(ii) This time we keep tossing the coin until we get one of the following sequences:

- if HH appears first, choose P,
- if HT appears first, choose Q,
- and if TH appears first, choose R.

So if we have the sequence "TTTTTH" we would choose R as out of the three sequences, TH appears first.

By considering what happens if the first toss is a Head or a Tail, find the probabilities that you choose P, Q and R.

13 The Probability question

Four players A, B, C and D play a coin-tossing game with a fair coin. Each player chooses a sequence of heads and tails, as follows:

Player A: HHT; Player B: THH; Player C: TTH; Player D: HTT.

The coin is then tossed until one of these sequences occurs, in which case the corresponding player is the winner.

(i) Show that, if only A and B play, then A has a probability of $\frac{1}{4}$ of winning.

(ii) If all four players play together, find the probabilities of each one winning.

(iii) Only B and C play. What is the probability of C winning if the first two tosses are TT?

Let the probabilities of C winning if the first two tosses are HT, TH and HH be p, q and r, respectively. Show that $p = \frac{1}{2} + \frac{1}{2}q$.

Find the probability that C wins.

Discussion

It is hard to write a preparation for this question which does not give too much away. For the first part think carefully about who must win after the first two tosses in the 4 different cases.

For the last part, you are asked to derive a given equation; then you need to find two more equations.