STEP Support Programme

STEP 3 Statistics: Solutions

1 This is a **large** pond, so the frog cannot jump over it (no matter how close he is to the edge).

(i) This question is asking you to find the probability that it takes the frog two jumps to get into the pond if it starts $n - \frac{1}{2} = 2 - \frac{1}{2} = 1\frac{1}{2}$ m away.

If the first jump is 2 m then the frog lands in the pond on this first jump. Hence we need the first jump to be a 1 m jump, which has probability p. After this the frog will be $\frac{1}{2}$ m away from the edge, hence will land in the pond on its next jump (no matter whether it is 1 m or 2 m). Hence $p_2(2) = p$.

(ii) u_1 is the expected number of jumps if the frog starts $1 - \frac{1}{2} = \frac{1}{2}$ m away. The frog will land in the pond on the first jump so $u_1 = 1$.

u_2 is the expected number of jumps if the frog starts $1\frac{1}{2}$ m away. If the first jump is 2 m then it will take 1 jump (probability $p_2(1) = q$) and if the first jump is 1 m then it will take 2 jumps (probability $p_2(2) = p$).

Hence we have:

$$u_2 = q + 2p\,.$$

This can also be written as $1 + p$ or (more usefully for the last part) as $2 - q$.

u_3 is the expected number of jumps if the frog starts $2\frac{1}{2}$ m away. The frog will need either 2 or 3 jumps to get into the pond. If the first jump is 2 m then it will land in the pond on the second jump, and if the first jump is 1 m followed by a 2 m jump it will also land in the pond on the second jump. Hence $p_3(2) = q + pq$. If the first two jumps are 1 m jumps then the frog will land in the pond on the third jump (no matter what size the third jump is), so $p_3(3) = p^2$[1]. We therefore have:

$$\begin{aligned} u_3 &= 2 \times (q + pq) + 3 \times p^2 \\ &= 2\left(q + (1-q)q\right) + 3(1-q)^2 \\ &= 2q + 2q - 2q^2 + 3 - 6q + 3q^2 \\ &= 3 - 2q + q^2 \qquad\qquad \text{as required.} \end{aligned}$$

(iii) Using $u_1 = 1$, $u_2 = 2 - q$ or $u_3 = 3 - 2q + q^2$ we have:

$$1 = A + B + C \tag{1}$$
$$2 - q = -Aq + B + 2C \tag{2}$$
$$3 - 2q + q^2 = Aq^2 + B + 3C \tag{3}$$

Evaluating $(2) - (1)$ gives:

$$1 - q = C - A(1+q) \tag{4}$$

[1] As a sanity check, you can make sure that all the (non-zero) probabilities sum to 1. Here we have $q + pq + p^2 = q + p(q+p) = q + p = 1$

Evaluating (3) − (2) gives:

$$1 - q + q^2 = Aq(1+q) + C \tag{5}$$

Then (5) − (4) gives:

$$\begin{aligned}\left(1 - q + q^2\right) - (1 - q) &= Aq(1+q) + A(1+q) \\ q^2 &= A(1+q)^2 \\ \implies A &= \left(\frac{q}{1+q}\right)^2\end{aligned}$$

Substituting into (4) gives:

$$\begin{aligned}C &= 1 - q + A(1+q) \\ &= 1 - q + \frac{q^2}{1+q} \\ &= \frac{(1-q)(1+q) + q^2}{1+q} \\ &= \frac{1}{1+q}\end{aligned}$$

and substituting into (1) gives:

$$\begin{aligned}B &= 1 - A - C \\ &= 1 - \left(\frac{q}{1+q}\right)^2 - \frac{1}{1+q} \\ &= \frac{(1+q)^2 - q^2 - (1+q)}{1+q} \\ &= \frac{q}{(1+q)^2}\end{aligned}$$

This gives:

$$u_n = \left(\frac{q}{1+q}\right)^2 (-q)^{n-1} + \frac{q}{(1+q)^2} + \frac{n}{1+q}\,.$$

We have $0 < q < 1$, so as $n \to \infty$, $(-q)^{n-1} \to 0$. We also have $\frac{q}{1+q} < 1 \ll n$[2] and so $\frac{q}{(1+q)^2} \ll \frac{n}{1+q}$ and we can ignore the $\frac{q}{(1+q)^2}$ term as being negligible compared to the third term. Hence for large n we have:

$$u_n \approx \frac{n}{1+q} = \frac{n}{p+2q}\,.$$

The jumps are each of integer length, so to cover $\left(n - \frac{1}{2}\right)$ m the frog needs to cover n m. The expected distance covered by the frog in one jump is $p + 2q$ m, so "on average" the frog jumps $p + 2q$ m each time. Therefore the expected number of jumps needed is $\frac{n}{p+2q}$.

[2] The symbol "$\ll$" means "much less than".

2 We have:

$$\begin{aligned} S - rS &= 1 + (1+d)r + (1+2d)r^2 + (1+3d)r^3 + \cdots + (1+nd)r^n + \cdots \\ &\qquad - r - (1+d)r^2 - (1+2d)r^3 - \cdots - (1+(n-1)d)r^n - \cdots \\ &= 1 + dr + dr^2 + dr^3 + \cdots + dr^n + \cdots \end{aligned}$$

So we have:

$$\begin{aligned} S(1-r) &= 1 + dr(1 + r + r^2 + r^3 + \cdots) \\ &= 1 + \frac{dr}{1-r} \\ \implies S &= \frac{1}{1-r} + \frac{dr}{(1-r)^2} \qquad (*) \end{aligned}$$

The probability that Arthur hits the target on the first shot is a, the probability that he misses on the first shot then hits on the second is $(1-a)a$, and the probability that he misses the first two then hits on the third is $(1-a)^2a$ etc. The expected number of shots it takes Arthur to hit the target is therefore:

$$\begin{aligned} \mathrm{E}(A) &= a + 2(1-a)a + 3(1-a)^2a + 4(1-a)^3a + \cdots \\ &= a + 2aa' + 3aa'^2 + 4aa'^3 + \cdots \qquad \text{where } a' = 1-a \\ &= a(1 + 2a' + 3a'^2 + 4a'^3 + \cdots) \\ &= a \times \left(\frac{1}{1-a'} + \frac{a'}{(1-a')^2}\right) \qquad \text{using } (*) \text{ with } d=1 \text{ and } r = a' \\ &= a \times \left(\frac{1}{a} + \frac{1-a}{a^2}\right) \\ &= 1 + \frac{1-a}{a} \\ &= \frac{a+1-a}{a} \\ &= \frac{1}{a} \end{aligned}$$

Note that this is an example of a *Geometric distribution*, but no prior knowledge of the Geometric distribution is needed to answer the question. You may already know that the expectation of a Geometric distribution is $\frac{1}{p}$, but as the question tells you to prove it there is no real advantage in having this prior knowledge.

If Arthur is to win the contest then either Arthur hits the target with his first shot, or he misses, then Boadicea misses, then Arthur hits, or the first 4 attempts miss then Arthur hits etc. We have:

$$\begin{aligned} \alpha &= a + a'b'a + a'b'a'b'a + a'b'a'b'a'b'a + \cdots \\ &= a(1 + a'b' + (a'b')^2 + (a'b')^3 + \cdots) \\ &= a \times \frac{1}{1-a'b'} \\ &= \frac{a}{1-a'b'} \end{aligned}$$

Similarly we have:

$$\begin{aligned}\beta &= a'b + a'b'a'b + a'b'a'b'a'b + \cdots \\ &= a'b(1 + a'b' + (a'b')^2 + \cdots) \\ &= \frac{a'b}{1-a'b'}\end{aligned}$$

The expected number of shots is given by:

$$\begin{aligned}\mathrm{E}(X) &= (1\times a) + (2\times a'b) + (3\times a'b'a) + (4\times a'b'a'b) + (5\times a'b'a'b'a) + \cdots \\ &= a\left(1 + 3a'b' + 5(a'b')^2 + \cdots\right) + 2a'b\left(1 + 2a'b' + 3(a'b')^2 + \cdots\right) \\ &= a\times\left(\frac{1}{1-a'b'} + \frac{2a'b'}{(1-a'b')^2}\right) + 2a'b\times\left(\frac{1}{1-a'b'} + \frac{a'b'}{(1-a'b')^2}\right) \qquad \text{using } (*) \\ &= \frac{1}{1-a'b'}\left(a + \frac{2aa'b'}{1-a'b'} + 2a'b + \frac{2a'ba'b'}{1-a'b'}\right) \\ &= \frac{1}{1-a'b'}\left(\frac{a(1-a'b') + 2aa'b' + 2a'b(1-a'b') + 2a'ba'b'}{1-a'b'}\right) \\ &= \frac{1}{1-a'b'}\left(\frac{a + aa'b' + 2a'b}{1-a'b'}\right) \\ &= \frac{1}{1-a'b'}\left(\frac{(1-a') + (1-a')a'b' + 2a'(1-b')}{1-a'b'}\right) \\ &= \frac{1}{1-a'b'}\left(\frac{1 - a' + a'b' - a'a'b' + 2a' - 2a'b')}{1-a'b'}\right) \\ &= \frac{1}{1-a'b'}\left(\frac{1 - a'b' + a' - (a')^2b'}{1-a'b'}\right) \\ &= \frac{1}{1-a'b'}\left(\frac{(1-a'b') + a'(1-a'b')}{1-a'b'}\right) \\ &= \frac{1}{1-a'b'} + \frac{a'}{1-a'b'} \\ &= \frac{\alpha}{a} + \frac{\beta}{b}\end{aligned}$$

It took me a while to get the final answer in the required form. In the end I worked backwards from $\frac{\alpha}{a} + \frac{\beta}{b}$ to work out what I could do to get to the required result. The above solution is probably not the neatest way!

3 It is always a good idea to read the "stem" of the question a couple of times to make sure that you are not asked to do anything in it. In this case the "stem" only has information setting up the question in it.

(i) X_1 will equal 1 if the first letter in the row is A. There are a letter A's out of a total of n letters, so the probability that the first letter is A is $\frac{a}{n}$. Therefore the expectation of X_1 is given by $\mathrm{E}(X_1) = \left(1 \times \frac{a}{n}\right) + \left(0 \times \frac{b}{n}\right) = \frac{a}{n}$.

For X_k we want to find the probability that the $(k-1)$th letter is B and that the kth letter is A. The total number of (distinct) ways of arranging the n letters is $\frac{n!}{a!b!}$, and if we fix the $(k-1)$th and kth letter to be B and A respectively then the number of ways of arranging the $(n-2)$ other letters is $\frac{(n-2)!}{(a-1)!(b-1)!}$. The probability that $X_k = 1$ is therefore $\frac{(n-2)!}{(a-1)!(b-1)!} \times \frac{a!b!}{n!} = \frac{ab}{n(n-1)}$. Therefore $\mathrm{E}(X_k) = \frac{ab}{n(n-1)}$. Note that this does not depend on k (as long as $2 \leqslant k \leqslant n$), which makes sense if you think about the problem a bit.

We have:

$$\begin{aligned}
\mathrm{E}(S) &= \mathrm{E}(X_1 + X_2 + \cdots + X_n) \\
&= \mathrm{E}(X_1) + \mathrm{E}(X_2) + \cdots + \mathrm{E}(X_n) \\
&= \frac{a}{n} + \left(\cancel{(n-1)} \times \frac{ab}{n\cancel{(n-1)}}\right) \\
&= \frac{a+ab}{n} \\
&= \frac{a(b+1)}{n}
\end{aligned}$$

(ii) **(a)** We have:

$$X_1X_j = \begin{cases} 1 & \text{if } X_1 = 1 \text{ and } X_j = 1 \\ 0 & \text{otherwise} \end{cases}$$

Note that X_1 and X_j are not independent, so we cannot multiply the two probabilities found earlier. If $X_1 = X_j = 1$ then the first letter must be A, the $(j-1)$th must be B and the jth must be A. The number of ways of arranging the $n-3$ letters left is $\frac{(n-3)!}{(a-2)!(b-1)!}$. We therefore have:

$$\begin{aligned}
\mathrm{E}(X_1X_j) &= \mathrm{P}(X_1X_j = 1) \\
&= \frac{(n-3)!}{(a-2)!(b-1)!} \times \frac{a!b!}{n!} \\
&= \frac{a(a-1)b}{n(n-1)(n-2)}
\end{aligned}$$

(b) $X_iX_j = 1$ means that X_i and X_j are both equal to 1, so the $(i-1)$th letter is a B, the ith is an A, the $(j-1)$th letter is a B and the jth is an A. Note that the inner sum starts with $j = i+1$, so there is no overlap between the $(i-1)$th, ith, $(j-1)$th and jth letters.

We have:

$$\begin{aligned}\mathrm{E}(X_iX_j) &= \mathrm{P}(X_iX_j = 1)\\ &= \frac{(n-4)!}{(a-2)!(b-2)!} \times \frac{a!b!}{n!}\\ &= \frac{a(a-1)b(b-1)}{n(n-1)(n-2)(n-3)}\end{aligned}$$

Noting that this doesn't depend on i or j, we have:

$$\sum_{j=i+2}^{n} \mathrm{E}(X_iX_j) = (n-i-1) \times \frac{a(a-1)b(b-1)}{n(n-1)(n-2)(n-3)}$$ [3]

We then have:

$$\begin{aligned}\sum_{i=2}^{n-2}\left(\sum_{j=i+2}^{n} \mathrm{E}(X_iX_j)\right) &= \sum_{i=2}^{n-2}\Bigg(\cancel{(n-1)} \times \frac{a(a-1)b(b-1)}{n\cancel{(n-1)}(n-2)(n-3)}\\ &\qquad\qquad - i \times \frac{a(a-1)b(b-1)}{n(n-1)(n-2)(n-3)}\Bigg)\\ &= \cancel{(n-3)} \times \frac{a(a-1)b(b-1)}{n(n-2)\cancel{(n-3)}}\\ &\qquad\qquad - \frac{a(a-1)b(b-1)}{n(n-1)(n-2)(n-3)} \times \sum_{i=2}^{n-2} i\\ &= \frac{a(a-1)b(b-1)}{n(n-2)} - \frac{a(a-1)b(b-1)}{\cancel{n}(n-1)(n-2)\cancel{(n-3)}} \times \frac{\cancel{n(n-3)}}{2}\\ &= a(a-1)b(b-1)\left(\frac{2(n-1)}{2n(n-1)(n-2)} - \frac{n}{2n(n-1)(n-2)}\right)\\ &= a(a-1)b(b-1) \times \frac{n-2}{2n(n-1)(n-2)}\\ &= \frac{a(a-1)b(b-1)}{2n(n-1)}\end{aligned}$$

Note that there are several ways you can find $\displaystyle\sum_{i=2}^{n-2} i$. The way I used was to write out $2 + 3 + \cdots + (n-2)$ and then reverse and add, but there are other methods.

[3] First time I did this question I had the number of terms as $n-i-2$ rather than $n-i-1$. Be careful!

(c) We have $\text{Var}(S) = \text{E}(S^2) - (\text{E}(S))^2$. We already have found $\text{E}(S)$, so we need to find $\text{E}(S^2)$. Since $S = \sum_{i=1}^{n} X_i$, when we multiply S by itself we get $S^2 = \sum_{i=1}^{n} X_i^2 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} X_iX_j$.

It is a good idea to sketch out a table showing what happens if you calculate $(X_1+X_2+\cdots+X_n)\times(X_1+X_2+\cdots+X_n)$. The leading diagonal of this gives $X_1^2+X_2^2+\cdots+X_n^2$, and then the other terms are given by $2X_1(X_2+X_3+\cdots+X_n)+2X_2(X_3+\cdots+X_n)+\cdots+2X_{n-1}X_n$. The cross terms can then be written as $2X_1\sum_{j=2}^{n} X_j + 2X_2\sum_{j=3}^{n} X_j + \cdots + 2X_{n-1}\sum_{j=n}^{n} X_j$, which can then by expressed as a nested sum.

We would like to be able to use the results already shown, but our expression for S^2 doesn't quite allow this. Going back a step, we have:

$$\begin{aligned} S^2 &= \sum_{i=1}^{n} X_i^2 + 2X_1(X_2+\cdots+X_n) + 2X_2(X_3+\cdots+X_n) + \cdots \\ &= \sum_{i=1}^{n} X_i^2 + 2\sum_{j=2}^{n} X_1X_j + 2\sum_{i=2}^{n-1}\sum_{j=i+1}^{n} X_iX_j \end{aligned}$$

This still doesn't seem to help us use our results. However, it is helpful to note that we cannot have both $X_1 = 1$ and $X_2 = 1$ as this would require the first letter to be both A and B simultaneously. Similarly we have $X_kX_{k+1} = 0$ for all $1 \leqslant k \leqslant n-1$. This means we can re-write S^2 as:

$$S^2 = \sum_{i=1}^{n} X_i^2 + 2\sum_{j=3}^{n} X_1X_j + 2\sum_{i=2}^{n-2}\sum_{j=i+2}^{n} X_iX_j$$

We now have:

$$\begin{aligned} \text{E}(S^2) &= \text{E}\left(\sum_{i=1}^{n} X_i^2\right) + 2\sum_{j=3}^{n} \text{E}(X_1X_j) + 2\sum_{i=2}^{n-2}\sum_{j=i+2}^{n} \text{E}(X_iX_j) \\ &= \frac{a(b+1)}{n} + 2\cancel{(n-2)} \times \frac{a(a-1)b}{n(n-1)\cancel{(n-2)}} + \cancel{2} \times \frac{a(a-1)b(b-1)}{\cancel{2}n(n-1)} \\ &= \frac{a(b+1)}{n} + \frac{a(a-1)b(2+b-1)}{n(n-1)} \\ &= \frac{a(b+1)}{n} + \frac{a(a-1)b(b+1)}{n(n-1)} \end{aligned}$$

Note that $\text{E}(X_i^2) = \text{E}(X_i)$.

We then have:

$$\begin{aligned}
\text{Var}(S) &= \text{E}(S^2) - (\text{E}(S))^2 \\
&= \frac{a(b+1)}{n} + \frac{a(a-1)b(b+1)}{n(n-1)} - \left(\frac{a(b+1)}{n}\right)^2 \\
&= \frac{a(b+1)}{n} \times \left(1 + \frac{(a-1)b}{n-1} - \frac{a(b+1)}{n}\right) \\
&= \frac{a(b+1)}{n} \times \left(\frac{n(n-1) + n(a-1)b - (n-1)a(b+1)}{n(n-1)}\right) \\
&= \frac{a(b+1)}{n^2(n-1)}\big(n^2 - n + nab - nb - a(nb - b + n - 1)\big) \\
&= \frac{a(b+1)}{n^2(n-1)}\big(n^2 - n + \cancel{nab} - nb - \cancel{anb} + ab - an + a)\big) \\
&= \frac{a(b+1)}{n^2(n-1)}\big(n^2 - n(a+b) - n + ab + a)\big) \\
&= \frac{a(b+1)}{n^2(n-1)}\big(\cancel{n^2} - \cancel{n(n)} - (a+b) + ab + a)\big) \quad \text{since } n = a + b \\
&= \frac{a(b+1)}{n^2(n-1)}(ab - b) \\
&= \frac{a(b+1)}{n^2(n-1)}b(a-1) \quad \text{as required}
\end{aligned}$$

This is one of the questions where you have to keep ploughing through the algebra and hope that things simplify. I was a bit worried about the n^2 in the numerator until I remembered that $n = a + b$.

4 The integral of $\mathrm{f}(x)$ from 0 to infinity must be equal to 1, hence:

$$\begin{aligned} 1 &= \int_0^\infty \frac{Ck^{a+1}x^a}{(x+k)^{2a+2}}\,\mathrm{d}x \\ &= Ck^{a+1}\int_0^\infty \frac{x^a}{(x+k)^{2a+2}}\,\mathrm{d}x \end{aligned}$$

Using the result given at the start of the question with $m = a$ and $n = 2a$ gives:

$$\begin{aligned} 1 &= C\cancel{k^{a+1}} \times \frac{a! \times a!}{(2a+1)!\cancel{k^{a+1}}} \\ \implies \quad C &= \frac{(2a+1)!}{a!a!} \end{aligned}$$

Looking at the limits suggests that $u = \frac{k^2}{x}$ might be a good choice. This gives:

$$\begin{aligned} \int_0^v \frac{x^a}{(x+k)^{2a+2}}\,\mathrm{d}x &= \int_\infty^{\frac{k^2}{v}} \frac{\frac{k^{2a}}{u^a}}{\left(\frac{k^2}{u}+k\right)^{2a+2}} \times \left(-\frac{x^2}{k^2}\right)\,\mathrm{d}u \\ &= -\int_\infty^{\frac{k^2}{v}} \frac{\frac{k^{2a}}{u^a}}{\left(\frac{k^2}{u}+k\right)^{2a+2}} \times \left(\frac{k^2}{u^2}\right)\,\mathrm{d}u \\ &= \int_{\frac{k^2}{v}}^\infty \frac{k^{2a} \times u^{a+2}}{(k^2+ku)^{2a+2}} \times \left(\frac{k^2}{u^2}\right)\,\mathrm{d}u \\ &= \int_{\frac{k^2}{v}}^\infty \frac{k^{2a+2} \times u^a}{(k^2+ku)^{2a+2}}\,\mathrm{d}u \\ &= \int_{\frac{k^2}{v}}^\infty \frac{k^{2a+2} \times u^a}{k^{2a+2}\,(k+u)^{2a+2}}\,\mathrm{d}u \\ &= \int_{\frac{k^2}{v}}^\infty \frac{u^a}{(k+u)^{2a+2}}\,\mathrm{d}u \qquad \text{as required.} \end{aligned}$$

For the median we need m such that $\int_0^m \mathrm{f}(x)\,\mathrm{d}x = \int_m^\infty \mathrm{f}(x)\,\mathrm{d}x \quad (= \frac{1}{2})$. This means that the median value will be when $v = \frac{k^2}{v} \implies v = k$.

Note that the question said *deduce* i.e. you must use what you have just shown.

The expectation is given by:

$$\begin{aligned} \mathrm{E}(X) &= \int_0^\infty \frac{Ck^{a+1}x^{a+1}}{(x+k)^{2a+2}}\,\mathrm{d}x \\ &= \frac{(2a+1)!}{a!a!}k^{a+1}\int_0^\infty \frac{x^{a+1}}{(x+k)^{2a+2}}\,\mathrm{d}x \\ &= \frac{(2a+1)!}{a!a!}k^{a+1} \times \frac{(a+1)!(a-1)!}{(2a+1)!k^a} \\ &= \frac{k(a+1)}{a} \end{aligned}$$

$T < t$ if and only if $V > \dfrac{s}{t}$ and so we have:

$$\begin{aligned}\mathrm{P}(T < t) &= P\left(V > \frac{s}{t}\right)\\ &= \int_{\frac{s}{t}}^{\infty} \frac{Ck^{a+1}x^a}{(x+k)^{2a+2}}\,\mathrm{d}x\end{aligned}$$

Using the suggested substitution $u = \frac{s}{x}$ gives us:

$$\begin{aligned}\mathrm{P}(T < T) &= \int_{t}^{0} \frac{Ck^{a+1}\left(\frac{s}{u}\right)^a}{(\frac{s}{u}+k)^{2a+2}} \times \left(-\frac{x^2}{s}\right)\mathrm{d}u\\ &= -\int_{t}^{0} \frac{Ck^{a+1}s^a u^{a+2}}{(s+ku)^{2a+2}} \times \left(\frac{s}{u^2}\right)\mathrm{d}u\\ &= \int_{0}^{t} \frac{Ck^{a+1}s^{a+1}u^a}{(s+ku)^{2a+2}}\,\mathrm{d}u\end{aligned}$$

Hence the density function of T is $\dfrac{Ck^{a+1}s^{a+1}u^a}{(s+Ky)^{2a+2}}$.

Dividing top and bottom by k^{2a+2} means that the density function can be written as:

$$\frac{C\left(\frac{s}{k}\right)^{a+1}u^a}{\left(\frac{s}{k}+u\right)^{2a+2}}$$

which has the same form as the density function for V just with k replaced by $\dfrac{s}{k}$. Hence the median value of T is $\dfrac{s}{k}$ and the expected value of T is $\dfrac{s(a+1)}{Kay}$.

Hence:

$$\text{median time} \times \text{median speed} = \frac{s}{k} \times k = s$$

and

$$\mathrm{E}(\text{time}) \times \mathrm{E}(\text{speed}) = \frac{s(a+1)}{\cancel{k}a} \times \frac{\cancel{k}(a+1)}{a} = s \times \left(\frac{a+1}{a}\right)^2$$

which is greater than s.