STEP Support Programme

STEP 2 Probability and Statistics Questions: Solutions

There are some steps missing in the following solutions so you do need to work through the question yourself and fill these in.

1 The sum of the probabilities must be 1, so we have:

$$\begin{aligned}\sum_{k=1}^{\infty} A\frac{\lambda^k \mathrm{e}^{-\lambda}}{k!} &= A\mathrm{e}^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^k}{k!}\\ &= A\mathrm{e}^{-\lambda}\left(\lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots\right)\\ &= A\mathrm{e}^{-\lambda}\left(\mathrm{e}^{\lambda} - 1\right) = 1\end{aligned}$$

and hence $A = \left(1 - \mathrm{e}^{-\lambda}\right)^{-1}$.

For the mean we have:

$$\begin{aligned}\mathrm{E}(X) &= \sum_{k=1}^{\infty} k \times \frac{A\lambda^k \mathrm{e}^{-\lambda}}{k!}\\ &= A\mathrm{e}^{-\lambda}\sum_{k=1}^{\infty}\frac{k\lambda^k}{k!}\\ &= A\mathrm{e}^{-\lambda}\left(\lambda + \frac{2\lambda^2}{2!} + \frac{3\lambda^3}{3!} + \cdots\right)\\ &= A\mathrm{e}^{-\lambda}\left(\lambda\left[1 + \lambda + \frac{\lambda^2}{2!} + \cdots\right]\right)\\ &= A\mathrm{e}^{-\lambda} \times \lambda \times \mathrm{e}^{\lambda}\,.\end{aligned}$$

Hence we have $\mu = A\lambda = \dfrac{\lambda}{1 - \mathrm{e}^{-\lambda}}$.

$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mu^2$. We have:

$$\begin{aligned}
\mathrm{E}(X^2) &= \sum_{k=1}^{\infty} k^2 \times \frac{A\lambda^k \mathrm{e}^{-\lambda}}{k!} \\
&= A\mathrm{e}^{-\lambda} \sum_{k=1}^{\infty} \frac{k^2\lambda^k}{k!} \\
&= A\mathrm{e}^{-\lambda} \left(\lambda + \frac{2^2\lambda^2}{2!} + \frac{3^2\lambda^3}{3!} + \frac{4^2\lambda^4}{4!} + \cdots \right) \\
&= A\mathrm{e}^{-\lambda} \left(\lambda \left[1 + 2\lambda + \frac{3\lambda^2}{2!} + \frac{4\lambda^3}{3!} + \cdots \right] \right) \\
&= A\mathrm{e}^{-\lambda}\lambda \times \left[\left(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots \right) + \left(\lambda + \frac{2\lambda^2}{2!} + \frac{3\lambda^3}{3!} + \cdots \right) \right] \\
&= A\mathrm{e}^{-\lambda}\lambda \left[\mathrm{e}^{\lambda} + \lambda\mathrm{e}^{\lambda} \right] \\
&= A\lambda(1+\lambda)\,.
\end{aligned}$$

We then have $\mathrm{Var}(X) = \mu(1+\lambda) - \mu^2$ (using $A\lambda = \mu$).

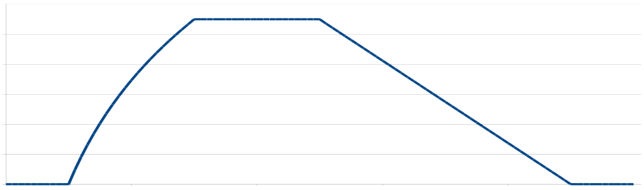
Since λ is a positive constant, we have $0 < \mathrm{e}^{-\lambda} < 1$ and so $\mu = \dfrac{\lambda}{1-\mathrm{e}^{-\lambda}} > \lambda$. We also know that $\mathrm{Var}(X) > 0$ hence $\mu(1-\mu+\lambda) > 0$ and so we have $1-\mu+\lambda > 0$ (since $\mu > 0$).

For the last part, first note that $1 - \mathrm{e}^{-100} \approx 1$, so $A \approx 1$ and $\mathrm{P}(X = 100) \approx \mathrm{P}(Y = 100)$ where Y is a Poisson distribution with mean 100.

We can then approximate Y with a normal approximation, $Q \sim N(100, 100)$. Applying a continuity correction gives $\mathrm{P}(Y = 100) \approx \mathrm{P}(99.5 < Q < 100.5)$. Normalising this gives $\mathrm{P}\left(\frac{-0.5}{10} < Z < \frac{0.5}{10}\right)$ which is equal to $2 \times \mathrm{P}(0 < Z < 0.05)$. Using normal tables gives us 2×0.0199 and so the answer is 0.04 to 2 d.p.

Before 2019 the STEP formulae book had lots of statistical tables in it (It was a standard A-level formulae book). Very few STEP questions require the use of the statistical tables, and in a case like this what would probably happen is a section of the relevant table would be included in the question.

2 **(i)** The sketch of $\mathrm{f}(x)$ looks something like this:



Note that since the graph is continuous the ends of the "sections" must meet up.

(ii) The fact that the graph is continuous gives us the following conditions:

$$x = 2k \implies \ln k = a - 2kb$$
$$x = 4k \implies a - 4kb = 0$$

This gives $b = \dfrac{\ln k}{2k}$ and $a = 2\ln k$. The total area under $\mathrm{f}(x)$ must be equal to 1. This means we have:

$$\int_1^k \ln x \mathrm{d}x + \int_k^{2k} \ln k \mathrm{d}x + \int_{2k}^{4k} a - bx \, \mathrm{d}x = 1\,.$$

The middle part is just a rectangle with area $k \times \ln k$, and the last integral gives $2ak - 6bk^2$, which simplifies to $k \ln k$ when you substitute for a and b. The first integral requires you to integrate $\ln x$ which can be done by writing this as $1 \times \ln x$ and integrating by parts. The result of the three integrals gives:

$$(k\ln k - k + 1) + (k\ln k) + (k\ln k) = 1 \text{ i.e.}$$
$$3k\ln k = k$$

Since $k > 0$, you can divide by k. Answer: $k = \mathrm{e}^{\frac{1}{3}}$, hence $a = \frac{2}{3}$ and $b = \frac{1}{6}\mathrm{e}^{-\frac{1}{3}}$.

(iii) You are required to find m so that $\int_1^m \mathrm{f}(x)\mathrm{d}x = \frac{1}{2}$, and in order to do this you need to identify which section m lies in.

The first section has an area of:

$$\int_1^k \ln x \mathrm{d}x = 1 - \tfrac{2}{3}\mathrm{e}^{\frac{1}{3}}\,.$$

Since $\mathrm{e} > 1$ this area is less than $1 - \frac{2}{3} = \frac{1}{3}$ and so m does not lie in this section. The other two sections have equal areas, as one is a rectangle height $\ln k$, width k and the other is a triangle height $\ln k$, width $2k$. Hence m lies in the middle section and satisfies:

$$1 - \tfrac{2}{3}\mathrm{e}^{\frac{1}{3}} + \int_k^m \ln k \mathrm{d}x = \tfrac{1}{2} \qquad \text{which gives}$$
$$1 - \tfrac{2}{3}\mathrm{e}^{\frac{1}{3}} + \big[x \ln k\big]_k^m = \tfrac{1}{2} \qquad \text{but } \ln k = \tfrac{1}{3}$$
$$1 - \tfrac{2}{3}\mathrm{e}^{\frac{1}{3}} + \tfrac{1}{3}m - \tfrac{1}{3}\mathrm{e}^{\frac{1}{3}} = \tfrac{1}{2} \qquad \text{i.e.}$$
$$m = 3\left(\mathrm{e}^{\frac{1}{3}} - \tfrac{1}{2}\right)\,.$$

3 **(i)** The probability that Younis wins a particular game is:

$$\mathrm{P}(Y, Y, *) + \mathrm{P}(X, Y, Y) = (1-p)p + p(1-p)p = (1-p^2)p\,.$$

The probability that a particular game is drawn is:

$$\mathrm{P}(X, Y, X) + \mathrm{P}(Y, X, Y) = p(1-p)(1-p) + (1-p)^3 = (1-p)^2\,.$$

If Younis wins the match, he could win it on the first game, or the first game could be drawn then Younis wins the second game, or the first two could be drawn and Younis wins the third etc. This gives us:

$$w = (1-p^2)p \times \Big[1 + (1-p)^2 + \big((1-p)^2\big)^2 + \cdots\Big]$$

which is a geometric progression. The sum is:

$$w = \frac{(1-p^2)p}{1-(1-p)^2} = \frac{(1-p^2)p}{2p-p^2}$$

which simplifies to the required result.

For the inequality it will help to consider $w - \frac{1}{2}$ instead, which gives

$$w - \tfrac{1}{2} = \frac{p^2(1-2p)}{2p(2-p)}\,.$$

The only part which can be negative is the $(1-2p)$ bit, which means that if $p < \frac{1}{2}$ we have $w - \frac{1}{2} > 0$, i.e. $w > \frac{1}{2}$, and if $p > \frac{1}{2}$ we have $w - \frac{1}{2} < 0$, i.e. $w < \frac{1}{2}$.

For the last bit you are expected to justify your answer (no marks given for a lucky guess). You could find a counter example (i.e. two values that show w decreasing as p decreases), but considering the derivative of w with respect to p is probably a better approach.

$$\frac{\mathrm{d}w}{\mathrm{d}p} = \frac{-2p(2-p) - (-1)(1-p^2)}{(2-p)^2} = \frac{p^2 - 4p + 1}{(2-p)^2}\,.$$

Completing the square on the numerator gives $\dfrac{\mathrm{d}w}{\mathrm{d}p} = \dfrac{(2-p)^2 - 3}{(2-p)^2}$. Hence w increases as p increases when $(2-p)^2 - 3 > 0$ i.e. $0 < p < 2 - \sqrt{3}$.

(ii) The probability that after n games no-one has won is $(1-p)^{2n}$, which tends to 0 as $n \to \infty$ (as long as $p > 0$). This means that the probability that Xavier wins is $1 - w$. When $p = \frac{2}{3}$, $w = \frac{5}{12}$ and the probability that Xavier wins is $\frac{7}{12}$.

For the game to be fair, Younis' expected winnings should be 0. This gives:

$$k \times \frac{5}{12} - 1 \times \frac{7}{12} = 0$$

and so $k = \frac{7}{5}$.

(iii) When $p = 0$ the probability that a game is drawn is 1, so no-one ever wins. All the games result in (Y, X, Y).

4 If X is the number of supermarkets in a circle of radius y then $X \sim \text{Po}(k\pi y^2)$ and $\text{P}(X = 0) = \text{e}^{-k\pi y^2}$.

If we have $Y < y$ then there is at least one supermarket within a circle of radius y of the chosen point. The required probability is therefore $1 - \text{P}$(there are no supermarkets within a circle radius y) and we have $\text{P}(Y < y) = 1 - \text{e}^{-k\pi y^2}$. To find the probability density function we need to differentiate this which gives $\text{f}(y) = 2\pi y k \text{e}^{-k\pi y^2}$ as required.

The expectation is given by:

$$\text{E}(Y) = \int_0^\infty 2k\pi y^2 \text{e}^{-k\pi y^2} \, \text{d}y\,.$$

Using integration by parts, and noting that $\dfrac{\text{d}}{\text{d}y}\left(\text{e}^{-k\pi y^2}\right) = 2k\pi y \text{e}^{-k\pi y^2}$ we have:

$$\text{E}(Y) = \left[y\left(\text{e}^{-k\pi y^2}\right)\right]_0^\infty + \int_0^\infty \text{e}^{-k\pi y^2}\,\text{d}y = 0 + \int_0^\infty \text{e}^{-k\pi y^2}\,\text{d}y\,.$$

The last bit looks a little like the result given at the start of the question, and we can use a substitution of $x = y\sqrt{2\pi k}$ to get:

$$\text{E}(Y) = \frac{1}{\sqrt{2k\pi}} \int_0^\infty \text{e}^{-\frac{1}{2}x^2}\text{d}x = \frac{1}{2\sqrt{k}}\,.$$

Variance is given by $\text{E}(Y^2) - [\text{E}(Y)]^2$ and we have:

$$\begin{aligned}
\text{E}(Y^2) &= \int_0^\infty 2k\pi y^3 \text{e}^{-k\pi y^2}\,\text{d}y \\
&= \left[y^2\left(\text{e}^{-k\pi y^2}\right)\right]_0^\infty + \int_0^\infty 2y\text{e}^{-k\pi y^2}\,\text{d}y \\
&= 0 + \frac{1}{k\pi}\int_0^\infty 2k\pi y \text{e}^{-k\pi y^2}\,\text{d}y \\
&= \frac{-1}{k\pi}\left[\text{e}^{-k\pi y^2}\right]_0^\infty = \frac{1}{k\pi}\,.
\end{aligned}$$

Then $\text{Var}(Y) = \dfrac{1}{k\pi} - \dfrac{1}{4k}$ which gives the required result.

5 The number of texts that George's phone receives in a given hour is $X \sim \mathrm{Po}(\lambda)$. If he waits between one and two hours for the first text then there were no texts in the first hour followed by at least one in the second hour. The probability is therefore:

$$\begin{aligned} p &= \mathrm{P}(X=0) \times P(X \geqslant 1) \\ p &= \mathrm{P}(X=0) \times (1 - \mathrm{P}(X=0)) \\ p &= \mathrm{e}^{-\lambda}\left(1-\mathrm{e}^{-\lambda}\right) \\ p\mathrm{e}^{\lambda} &= 1-\mathrm{e}^{-\lambda} \\ p\mathrm{e}^{2\lambda} &= \mathrm{e}^{\lambda}-1 \\ p\mathrm{e}^{2\lambda} - \mathrm{e}^{\lambda} + 1 &= 0 \end{aligned}$$

Solving this equation gives $\mathrm{e}^{\lambda} = \dfrac{1 \pm \sqrt{1-4p}}{2p}$. For e^{λ} to be real we need $4p < 1$. We also need to show that both solutions for e^{λ} give positive values of λ, which means we need $\mathrm{e}^{\lambda} > 1$. Consider the two solutions separately.

For the positive square root we need $\dfrac{1+\sqrt{1-4p}}{2p} > 1$ i.e. $1 + \sqrt{1-4p} > 2p$. The LHS of this is bigger than 1, and since $4p < 1$ the RHS is less than $\frac{1}{2}$, hence this is true and this value of e^{λ} gives a positive value of λ.

For the second solution we need:

$$\begin{aligned} \frac{1-\sqrt{1-4p}}{2p} &> 1 \\ 1-\sqrt{1-4p} &> 2p \\ 1-2p &> \sqrt{1-4p} \\ (1-2p)^2 &> 1-4p \\ 1-4p+4p^2 &> 1-4p \\ 4p^2 &> 0 \end{aligned}$$

which is true, and hence the second solution for e^{λ} gives a positive value of λ.

For each of Mildred's two phones the probability that she waits between one and two hours for a text is p, which means that λ_1 and λ_2 are solutions of $p\mathrm{e}^{2\lambda} - \mathrm{e}^{\lambda} + 1 = 0$. We are told that the two means are different, so WLOG let λ_1 be the larger one. We have:

$$\begin{aligned} \mathrm{e}^{\lambda_1} &= \frac{1+\sqrt{1-4p}}{2p} \\ \mathrm{e}^{\lambda_2} &= \frac{1-\sqrt{1-4p}}{2p} \end{aligned}$$

This means that $\mathrm{e}^{\lambda_1+\lambda_2} = \dfrac{1}{p}$, and so $\lambda_1 + \lambda_2 = -\ln p$.

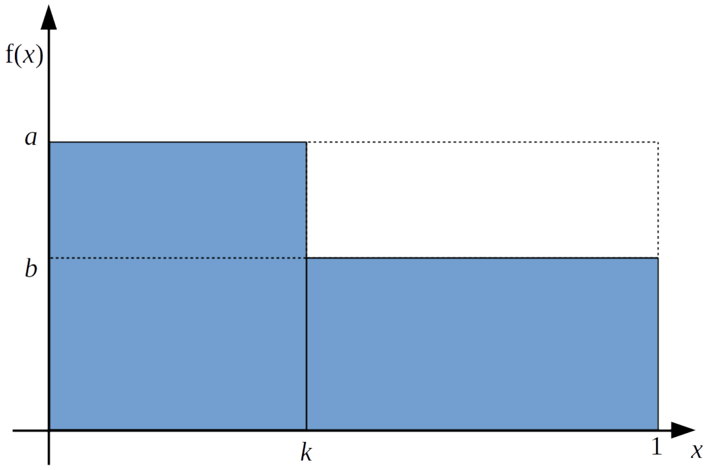
The total number of texts that Mildred receives on her two phones is distributed as $\mathrm{Po}(\lambda_1+\lambda_2)$. The probability that she waits between one and two hours for her first text is:

$$\begin{aligned} \mathrm{e}^{-(\lambda_1+\lambda_2)}\left(1-\mathrm{e}^{-(\lambda_1+\lambda_2)}\right) &= \mathrm{e}^{\ln p}\left(1-\mathrm{e}^{\ln p}\right) \\ &= p(1-p) \end{aligned}$$

6 The Probability question

Don't forget to answer the request in the "stem" of the question!

Noting that a is greater than b, we can draw a sketch of the probability density function, as below:



The shaded area has to be equal to 1. From the diagram we can see that the shaded area is less than $a \times 1 = a$ (i.e. a rectangle of height a and width 1). This means that we have $a > 1$. Similarly the shaded area is greater than $b \times 1 = b$, and so $b < 1$.

(i) We have:

$$\begin{aligned}\mathrm{E}(X) &= \int_0^1 x \times \mathrm{f}(x)\,\mathrm{d}x \\ &= \int_0^k ax\,\mathrm{d}x + \int_k^1 bx\,\mathrm{d}x \\ &= \left[\tfrac{1}{2}ax^2\right]_0^k + \left[\tfrac{1}{2}bx^2\right]_k^1 \\ &= \tfrac{1}{2}\left(ak^2 + b - bk^2\right)\end{aligned}$$

This doesn't look like the required result as there are some "k"'s involved.

We know that the total area under the graph has to be 1, so we have:

$$ak + b(1-k) = 1 \implies k = \frac{1-b}{a-b}$$

Substituting this into our expression for $\mathrm{E}(X)$ we have:

$$\begin{aligned}
\mathrm{E}(X) &= \tfrac{1}{2}\left((a-b)k^2 + b\right) \\
&= \tfrac{1}{2}\left((a-b) \times \left(\frac{1-b}{a-b}\right)^2 + b\right) \\
&= \tfrac{1}{2}\left(\frac{(1-b)^2}{a-b} + b\right) \\
&= \frac{1}{2(a-b)}\Big((1-b)^2 + b(a-b)\Big) \\
&= \frac{1}{2(a-b)}\Big(1 - 2b + \cancel{b^2} + ab - \cancel{b^2}\Big) \\
&= \frac{1-2b+ab}{2(a-b)} \qquad \text{as required}
\end{aligned}$$

(ii) The median will be less than or equal to k if and only if $ak \geqslant \frac{1}{2}$, which means that:

$$\begin{aligned}
a \times \frac{1-b}{a-b} &\geqslant \frac{1}{2} \\
2a(1-b) &\geqslant a-b \\
2a - 2ab &\geqslant a-b \\
a+b &\geqslant 2ab
\end{aligned}$$

In this case the median is given by $a \times M = \frac{1}{2} \implies M = \frac{1}{2a}$.

If we have $ak \leqslant \frac{1}{2}$ (and so $a+b \leqslant 2ab$) then the median is less than or equal to k. Here we need $(1-M) \times b = \frac{1}{2} \implies M = 1 - \frac{1}{2b}$.

(iii) There are two different cases we need to check. In each case, consider $M - \mathrm{E}(X)$ and try and show that this is negative.

- $M = \dfrac{1}{2a}$

$$\begin{aligned}
M - \mathrm{E}(X) &= \frac{1}{2a} - \frac{1-2b+ab}{2(a-b)} \\
&= \frac{(a-b) - a(1-2b+ab)}{2a(a-b)} \\
&= \frac{\cancel{a} - b - \cancel{a} + 2ba - a^2b}{2a(a-b)} \\
&= \frac{-b(a^2-2a+1)}{2a(a-b)} \\
&= -\frac{b(a-1)^2}{2a(a-b)} < 0 \\
&\quad \text{since } b, a, (a-b), (a-1)^2 \text{ are all positive}
\end{aligned}$$

Hence $M < \mathrm{E}(X)$.

- $M = 1 - \dfrac{1}{2b}$

$$
\begin{aligned}
M - \mathrm{E}(X) &= \left(1 - \frac{1}{2b}\right) - \frac{1 - 2b + ab}{2(a-b)} \\
&= \frac{2b(a-b) - (a-b) - b(1 - 2b + ab)}{2b(a-b)} \\
&= \frac{2ab - \cancel{2b^2} - a + \cancel{b} - \cancel{b} + \cancel{2b^2} - ab^2}{2b(a-b)} \\
&= \frac{2ab - a - ab^2}{2b(a-b)} \\
&= \frac{-a(1 - 2b + b^2)}{2b(a-b)} \\
&= -\frac{a(1-b)^2}{2b(a-b)} < 0 \\
&\qquad \text{since } a, b, (a-b), (1-b)^2 \text{ are all positive}
\end{aligned}
$$

Hence $M < \mathrm{E}(X)$.