STEP 2016, Paper 1, Q13 – Solution (3 pages; 29/5/18)

Note: This question concerns variables that have the exponential (aka the negative exponential) distribution. Although it is possible to answer it without any knowledge of this distribution (which isn't required by the specification, strictly speaking), the official solution uses results about the distribution - rather unfairly. The examiners' report observes that virtually no one did part (ii).

(i) [This uses a standard device for continuous distributions: work with the CDF, and then convert to the pdf.]

$P(T < t) = 1 - P(T > t) = 1 - [P(X > t)]^n$, where X is the arrival time of a single email

And $P(X > t) = 1 - P(X < t) = 1 - \int_0^t \lambda e^{-\lambda x} dx$

$= 1 - \left[-e^{-\lambda x}\right]_0^t = 1 - \left(1 - e^{-\lambda t}\right) = e^{-\lambda t}$

So $P(T < t) = 1 - e^{-n\lambda t}$

and T has pdf $\frac{d}{dt}\left(1 - e^{-n\lambda t}\right) = n\lambda e^{-n\lambda t}$, as required.

[The official solution allows candidates to just quote $E(T)$ from the formulae booklet "for an easy 6 marks", as T has the exponential distribution with parameter $n\lambda$.]

$E(T) = \int_0^\infty tn\lambda e^{-n\lambda t} dt$

By Parts: $\left[-te^{-n\lambda t}\right]_0^\infty - \int_0^\infty -e^{-n\lambda t}\, dt$

$= 0 + \left[-\frac{1}{n\lambda} e^{-n\lambda t}\right]_0^\infty = \frac{1}{n\lambda}$

[Note: It is acceptable to assume that $e^{n\lambda t}$ grows faster than t, so that $te^{-n\lambda t} \to 0\ as\ t \to \infty$]

(ii) ["Write down" often means that the result can be deduced by some lateral thinking.]

Considering a case by case approach (and using 'concrete' cases relating to the individual emails, rather than something more nebulous, such as the "1st email to arrive"),

P(2nd email arrives after time t) =

P(1st email arrives after time t)

+ P(email #1 arrives before time t)

× P(1st email amongst other n − 1 emails arrives after time t)

+ P(email #2 arrives before time t)

× P(1st email amongst other n − 1 emails arrives after time t)

$+ \cdots$

[we are using the fact that the emails arrive independently of each other; ie no conditional probabilities are needed]

$= e^{-n\lambda t} + n(1 - e^{-\lambda t})e^{-(n-1)\lambda t}$

pdf for 2nd email:

$\frac{d}{dt}(1 - e^{-n\lambda t} - n\left(1 - e^{-\lambda t}\right)e^{-(n-1)\lambda t})$

$= n\lambda e^{-n\lambda t} - n(\lambda e^{-\lambda t})e^{-(n-1)\lambda t}$

$-n\left(1 - e^{-\lambda t}\right)(-1)(n-1)\lambda e^{-(n-1)\lambda t}$

$= n\lambda e^{-n\lambda t} - n\lambda e^{-n\lambda t}$

$+n(n-1)\lambda\{e^{-(n-1)\lambda t} - e^{-n\lambda t}\}$

$= n(n-1)\lambda\{e^{-(n-1)\lambda t} - e^{-n\lambda t}\}$

[In the mark scheme, there doesn't seem to be any mention of the above parts of (ii). The 4 marks available only relate to finding the expected value!]

Then the expected value is

$\int_0^\infty t.n(n-1)\lambda\{e^{-(n-1)\lambda t} - e^{-n\lambda t}\}dt$

This integral can be worked out using Parts, as before, or we can write it as

$n\int_0^\infty t.(n-1)\lambda e^{-(n-1)\lambda t}dt - (n-1)\int_0^\infty t.n\lambda e^{-n\lambda t}dt$

and the two integrals are the expected values of exponential variables with parameters $(n-1)\lambda$ and $n\lambda$,

so that the answer is $n.\frac{1}{(n-1)\lambda} - (n-1).\frac{1}{n\lambda}$

$= \frac{1}{\lambda}(\frac{n}{n-1} - \frac{n-1}{n})$

$= \frac{1}{\lambda}(\frac{1}{n} - 1 + \frac{n}{n-1})$

$= \frac{1}{\lambda}(\frac{1}{n} + \frac{n-(n-1)}{n-1})$

$= \frac{1}{\lambda}(\frac{1}{n} + \frac{1}{n-1})$, as required.

[This agrees with the 'memory-less' property of the exponential distribution (ie the remaining waiting time at any point is independent of the time elapsed). The expected time until the arrival of the 1st email is $\frac{1}{\lambda n}$ (from (i)), and then the additional expected time until the arrival of the 2nd email is $\frac{1}{\lambda(n-1)}$ (as we now have $n-1$ emails that haven't yet arrived).