THE ROYAL STATISTICAL SOCIETY

2010 EXAMINATIONS – SOLUTIONS

HIGHER CERTIFICATE

MODULE 2

PROBABILITY MODELS

The Society provides these solutions to assist candidates preparing for the examinations in future years and for the information of any other persons using the examinations.

The solutions should NOT be seen as "model answers". Rather, they have been written out in considerable detail and are intended as learning aids.

Users of the solutions should always be aware that in many cases there are valid alternative methods. Also, in the many cases where discussion is called for, there may be other valid points that could be made.

While every care has been taken with the preparation of these solutions, the Society will not be responsible for any errors or omissions.

The Society will not enter into any correspondence in respect of these solutions.

Note. In accordance with the convention used in the Society's examination papers, the notation log denotes logarithm to base *e*. Logarithms to any other base are explicitly identified, e.g. $\log_{10}$.

Higher Certificate, Module 2, 2010. Question 1

(i) $E(X) = 2p$ $\quad\quad$ $\text{Var}(X) = 2p(1-p)$ $\quad\quad$ $P(X = 2) = p^2$

$$P(X=0 \mid X<2) = \frac{P\left[(X=0)\cap(X<2)\right]}{P(X<2)} = \frac{P(X=0)}{1-p^2} = \frac{(1-p)^2}{1-p^2} = \frac{1-p}{1+p}$$

$$P(X=1 \mid X<2) = \frac{P\left[(X=1)\cap(X<2)\right]}{P(X<2)} = \frac{P(X=1)}{1-p^2} = \frac{2p(1-p)}{1-p^2} = \frac{2p}{1+p}$$

[Alternatively, $P(X = 1 \mid X < 2)$ may be obtained as $1 - P(X = 0 \mid X < 2)$.]

(ii) (a) Each X is the number of successes in two Bernoulli trials with probability p of success. So, noting that the Xs are independent and p is the same for all of them, Y is the number of successes in 200 such trials and we have $Y \sim \text{B}(200, p)$.

(b) With $p = 2/3$, $Y \sim \text{N}\left(\frac{400}{3}, \frac{400}{9}\right)$, approximately.

$$P(Y > 140) \approx 1 - \Phi\left(\frac{140.5 - (400/3)}{20/3}\right) = 1 - \Phi(1.075)$$

(where Φ denotes the standard Normal cdf as usual)

$= 1 - 0.859 = 0.141.$

(c) With $p = 0.02$, $Y \sim \text{Poisson}(4)$, approximately.

$P(Y > 2) = 1 - P(Y \le 2)$

$= 1 - 0.2381$ from tables; or by use of $1 - e^{-4}\left(1 + 4 + \frac{4^2}{2}\right)$

$= 0.762$ approximately.

(d) Now $p = 0.98$.

Consider $200 - Y \sim \text{B}(200, 0.02) \sim \text{Poisson}(4)$ approximately.

$P(Y \le 197) = P(200 - Y \ge 3)$

$\approx 1 - P(\text{Poisson}(4) \le 2) = 0.762$ approximately (as in (c)).

Higher Certificate, Module 2, 2010. Question 2

$W \sim N(24, 1)$

(i) $P(W > 25) = 1 - \Phi\left(\frac{25-24}{1}\right) = 1 - \Phi(1) = 1 - 0.8413 = 0.1587.$

(ii) (a) $P(C = 0) = P(W \le 25) = \Phi(1) = 0.8413.$

$P(C = 5) = P(25 < W \le 26)$

$= \Phi\left(\frac{26-24}{1}\right) - \Phi(1) = \Phi(2) - \Phi(1) = 0.9772 - 0.8413 = 0.1359.$

$P(C = 10) = P(26 < W \le 27)$

$= \Phi\left(\frac{27-24}{1}\right) - \Phi(2) = \Phi(3) - \Phi(2) = 0.9987 - 0.9772 = 0.0215.$

(b) We have the following for C.

c	0	5	10	15
$P(C = c)$	0.8413	0.1359	0.0215	0.0013
$cP(C = c)$	0	0.6795	0.215	0.0195
$c^2P(C = c)$	0	3.3975	2.15	0.2925

So $E(C)$ = row total for the "$cP(C = c)$" row = 0.914.

Also, $E(C^2)$ = row total for the "$c^2P(C = c)$" row = 5.84,

and therefore $\text{Var}(C) = 5.84 - 0.914^2 = 5.005$.

(c) $E(C_T) = 91400$, $\text{Var}(C_T) = 500500$.

We use a Normal approximation to the distribution of C_T.

The upper 95% point of this is $91400 + (1.6449 \times \sqrt{500500}) = 92564$.

(d) Independence may hold for people travelling separately but is most unlikely to hold for families or other groups – they may, for example, try to equalise their loads to minimise the excess cost – and this will affect the variance of C_T.

"100 000" must be a rough figure for the total number of passengers.

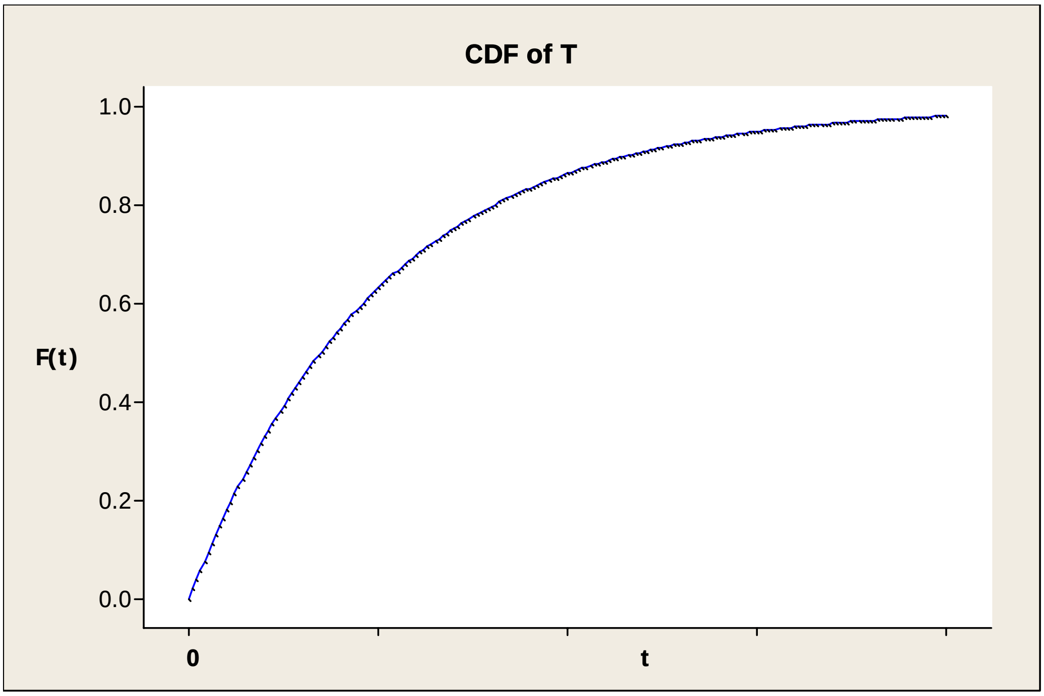
The distribution of C is clearly positively skew, so a large number of passengers is needed to validate the use of a Normal approximation for C_T. The given total of 100 000 (even as a rough figure) is probably large enough.

Higher Certificate, Module 2, 2010. Question 3

$$f_T(t) = \lambda e^{-\lambda t}, \quad t > 0, \quad \lambda > 0$$

(i) The cdf is $F_T(t) = \int_0^t \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_0^t = 1 - e^{-\lambda t}$ for $t > 0$.

[Also $F_T(t) = 0$ for $t \le 0$.]



[Note. The graph should of course be a smooth curve. It may not be shown as such, due to the limits of electronic reproduction.]

(ii) $P(a < T \le b) = F_T(b) - F_T(a) = \left(1 - e^{-\lambda b}\right) - \left(1 - e^{-\lambda a}\right) = e^{-\lambda a} - e^{-\lambda b}$.

(iii) $P(0 < T \le 1) = 1 - e^{-\lambda}$.

$P(1 < T \le 2) = e^{-\lambda} - e^{-2\lambda} = e^{-\lambda}\left(1 - e^{-\lambda}\right)$.

So we have $1 - e^{-\lambda} = 2\left\{e^{-\lambda}\left(1 - e^{-\lambda}\right)\right\}$ which gives $e^{\lambda} = 2$ so $\lambda = \log 2 = 0.693$.

Solution continued on next page

(iv) For $t > c > 0$, we have

$$P(T > t \mid T > c) = \frac{P(T > t \text{ and } T > c)}{P(T > c)} = \frac{P(T > t)}{P(T > c)} = \frac{e^{-\lambda t}}{e^{-\lambda c}} = e^{-\lambda(t-c)}.$$

Hence the conditional pdf of T given that $T > c$ is $\frac{d}{dt}\left(1 - e^{-\lambda(t-c)}\right) = \lambda e^{-\lambda(t-c)}$ (for $t > c$).

Arguing similarly for the random variable $T - c$, we first find, for $t > 0$,

$$P(T - c > t \mid T > c) = \frac{P(T - c > t \text{ and } T > c)}{P(T > c)} = \frac{P(T > t + c)}{P(T > c)} = \frac{e^{-\lambda(t+c)}}{e^{-\lambda c}} = e^{-\lambda t}.$$

Hence the conditional pdf of $T - c$ given that $T > c$ is $\lambda e^{-\lambda t}$, for $t > 0$.

Thus the conditional distribution of $T - c$ given that $T > c$ is the same as the unconditional distribution of T, for any constant $c > 0$.

Higher Certificate, Module 2, 2009. Question 4

(i) $X \sim$ Poisson(2).

(a) $P(X = 0) = 0.1353$ from tables (or as e^{-2}).

(b) $P(X > 2) = 1 - P(X \leq 2) = 1 - 0.6767$ from tables, or by use of

$$1 - e^{-2}\left(1 + 2 + \frac{2^2}{2}\right)$$

$= 0.3233.$

(ii) With obvious notation, $X_A \sim$ Poisson(0.2) and $X_B \sim$ Poisson(0.3).

(a) $P(\text{no flaws}) = P(X_A = 0 \text{ and } X_B = 0)$

$= P(X_A = 0).P(X_B = 0)$ by independence

$= e^{-0.2}e^{-0.3} = 0.8187 \times 0.7408 = 0.6065.$

[Alternative method: $X_A + X_B \sim$ Poisson(0.5), so $P(X_A + X_B = 0)$ $= e^{-0.5} = 0.6065$.]

(b) $P(\text{exactly one flaw}) = P\{(X_A = 0 \text{ and } X_B = 1) \text{ or } (X_B = 0 \text{ and } X_A = 1)\}$

$= P(X_A = 0 \text{ and } X_B = 1) + P(X_B = 0 \text{ and } X_A = 1)$

$= P(X_A = 0).P(X_B = 1) + P(X_B = 0).P(X_A = 1)$ by independence

$= e^{-0.2} \times (0.3)e^{-0.3} + e^{-0.3} \times (0.2)e^{-0.2}$

$= 0.3033$ [Or by use of the alterative method as above.]

Solution continued on next page

(iii) (a)

$$P(A|7\text{ flaws}) = \frac{P(7\text{ flaws}|A)P(A)}{P(7\text{ flaws})}$$

$$= \frac{P(7\text{ flaws}|A)P(A)}{P(7\text{ flaws}|A)P(A) + P(7\text{ flaws}|B)P(B)}$$

$$= \frac{\frac{e^{-4}4^7}{7!}\times 0.75}{\left(\frac{e^{-4}4^7}{7!}\times 0.75\right) + \left(\frac{e^{-6}6^7}{7!}\times 0.25\right)} = 0.5647.$$

(b) Repeating this calculation for 8 flaws:

$$P(A|8\text{ flaws}) = \frac{P(8\text{ flaws}|A)P(A)}{P(8\text{ flaws})}$$

$$= \frac{P(8\text{ flaws}|A)P(A)}{P(8\text{ flaws}|A)P(A) + P(8\text{ flaws}|B)P(B)}$$

$$= \frac{\frac{e^{-4}4^8}{8!}\times 0.75}{\left(\frac{e^{-4}4^8}{8!}\times 0.75\right) + \left(\frac{e^{-6}6^8}{8!}\times 0.25\right)} = 0.4638.$$

The rigging contains more rope from company A than from B; but the rope from B is less reliable than that from A. Thus, as we find increasingly many flaws in the rope, the probability that it came from A reduces to less than ½.