THE ROYAL STATISTICAL SOCIETY

2002 EXAMINATIONS – SOLUTIONS

GRADUATE DIPLOMA

PAPER I – STATISTICAL THEORY & METHODS

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.

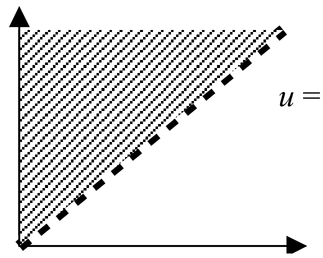
The solutions should NOT be seen as "model answers". Rather, they have been written out in considerable detail and are intended as learning aids.

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.

Graduate Diploma, Statistical Theory & Methods, Paper I, 2002. Question 1



(i) $S(t) = \int_t^\infty f(u)\,du \qquad t \ge 0$ and hence

$$\int_0^\infty S(t)\,dt = \int_0^\infty \int_t^\infty f(u)\,du\,dt$$

$$= \int_{u=0}^\infty f(u)\left\{\int_0^u dt\right\} du \text{, integrating over the shaded region,}$$

$$= \int_0^\infty u f(u)\,du = E[T].$$

(ii) $S(x) = \begin{cases} 1 & 0 \le x < 1 \\ x e^{-(x-1)} & x \ge 1 \end{cases}$

$$E[X] = \int_0^\infty S(x)\,dx = \int_0^1 1\,dx + \int_1^\infty x e^{-(x-1)}\,dx$$

$= 1 + \int_0^\infty (u+1) e^{-u}\,du$ putting $u = x - 1$; now use $\Gamma(m)$ result quoted in the question

$= 1 + \Gamma(2) + \Gamma(1) = 1 + 1 + 1 = 3.$

(iii)

$F_Y(y) = \begin{cases} 0 & \text{for } y \le 1 \quad \text{(by definitions of } X \text{ and of } Y\text{)} \\ P(Y \le y) = F_X\left(\sqrt{y}\right) = 1 - \sqrt{y} e^{-\left(\sqrt{y}-1\right)} & \text{for } y > 1 \end{cases}$

$\therefore\ S_Y(y) = \begin{cases} 1 & \text{for } 0 \le y \le 1 \\ \sqrt{y} e^{-\left(\sqrt{y}-1\right)} & \text{for } y \ge 1 \end{cases}$

From (i), $E[Y] = \int_0^1 1\,dy + \int_1^\infty \sqrt{y}\, e^{-\left(\sqrt{y}-1\right)}\,dy$

$= 1 + 2\int_0^\infty (u+1)^2 e^{-u}\,du$ putting $u = \sqrt{y} - 1$

$= 1 + 2\Gamma(3) + 4\Gamma(2) + 2\Gamma(1)$

$= 1 + (2\times 2) + (4\times 1) + (2\times 1) = 11 = E\left[X^2\right].$

Therefore $\text{Var}(X) = E\left[X^2\right] - \{E[X]\}^2 = 11 - 3^2 = 2.$

Graduate Diploma, Statistical Theory & Methods, Paper I, 2002. Question 2

[Note that $\int_0^\infty u^{m-1}(1-u)^{n-1}\,du = \dfrac{(m-1)!(n-1)!}{(m+n-1)!}$ for all positive integers m, n.]

(a) $$E[U] = \frac{(m+n-1)!}{(m-1)!(n-1)!}\int_0^1 u.u^{m-1}(1-u)^{n-1}\,du$$
$$= \frac{(m+n-1)!}{(m-1)!(n-1)!}\cdot\frac{m!(n-1)!}{(m+n)!} = \frac{m}{m+n}.$$

Similarly, $$E\left[U^2\right] = \frac{(m+n-1)!}{(m-1)!(n-1)!}\cdot\frac{(m+1)!(n-1)!}{(m+n+1)!} = \frac{m(m+1)}{(m+n)(m+n+1)}.$$

$$\therefore \operatorname{Var}(U) = \frac{m(m+1)}{(m+n)(m+n+1)} - \left(\frac{m}{m+n}\right)^2 = \frac{\left(m^2+m\right)(m+n) - m^2(m+n+1)}{(m+n)^2(m+n+1)}$$
$$= \frac{m^3+m^2+m^2n+mn-m^3-m^2n-m^2}{(m+n)^2(m+n+1)} = \frac{mn}{(m+n)^2(m+n+1)}.$$

(b) $$f_X(x) = \int_{y=x}^1 12x^2\,dy = \left[12x^2y\right]_{y=x}^1 = 12x^2(1-x) \qquad \text{(for } 0 \le x \le 1\text{)}.$$
$$f_Y(y) = \int_{x=0}^y 12x^2\,dx = \left[4x^3\right]_{x=0}^y = 4y^3 \qquad \text{(for } 0 \le y \le 1\text{)}.$$

Thus X has beta distribution with m = 3 and n = 2 ["B(3,2)"] and so has mean $\frac{3}{5}$ and variance $\frac{1}{25}$.

Similarly, Y is B(4, 1) and so has mean $\frac{4}{5}$ and variance $\frac{2}{75}$.

$$E[XY] = \int_{y=0}^1\int_{x=0}^y xy.12x^2\,dx\,dy = \int_0^1\left\{\int_0^y 12x^3y\,dx\right\}dy$$
$$= \int_0^1 3y^5\,dy = \left[\frac{1}{2}y^6\right]_0^1 = \frac{1}{2}.$$

$$\therefore \operatorname{Cov}(X,Y) = E[XY] - E[X]E[Y] = \frac{1}{2} - \frac{3}{5}.\frac{4}{5} = \frac{1}{2} - \frac{12}{25} = \frac{1}{50}.$$

$$\therefore \rho_{XY} = \frac{1}{50}\Big/\sqrt{\frac{1}{25}\times\frac{2}{75}} = \frac{1}{50}\Big/\left(\frac{1}{25}\sqrt{\frac{2}{3}}\right) = \frac{1}{2}\sqrt{\frac{3}{2}} = 0.6124.$$

Graduate Diploma, Statistical Theory & Methods, Paper I, 2002. Question 3

(i) $U^2+V^2=(-2\ln X)(\sin^2 2\pi Y+\cos^2 2\pi Y)=-2\ln X$

$\therefore\ -\frac{1}{2}(U^2+V^2) = \ln X$ so that $X=\exp\left[-\frac{1}{2}(U^2+V^2)\right]$

$\frac{U}{V}=\frac{\sin 2\pi Y}{\cos 2\pi Y}=\tan 2\pi Y$, so $Y=\frac{1}{2\pi}\tan^{-1}\left(\frac{U}{V}\right)$.

(ii) Since X and Y are independent $\mathrm{U}(0,1)$, $f(X,Y)=1$ (for $0\le x\le 1$, $0\le y\le 1$).

The jacobian of the transformation from X, Y to U, V is

$$J=\begin{vmatrix}\frac{\partial X}{\partial U} & \frac{\partial X}{\partial V}\\ \frac{\partial Y}{\partial U} & \frac{\partial Y}{\partial V}\end{vmatrix}=\begin{vmatrix}-u\exp\left(-\frac{1}{2}\{u^2+v^2\}\right) & -v\exp\left(-\frac{1}{2}\{u^2+v^2\}\right)\\ \frac{1}{2\pi}.\frac{1}{v}.\frac{1}{1+(u/v)^2} & \frac{1}{2\pi}.\left(-\frac{u}{v^2}\right).\frac{1}{1+(u/v)^2}\end{vmatrix}$$

$$=\frac{1}{2\pi}\exp\left(-\frac{1}{2}\{u^2+v^2\}\right)\left(+\frac{u^2}{v^2}+1\right)\left(\frac{1}{1+(u^2/v^2)}\right) = \frac{1}{2\pi}\exp\left(-\frac{1}{2}\{u^2+v^2\}\right).$$

So $f(u,v) = |J|f(x,y) = \frac{1}{2\pi}\exp\left[-\frac{1}{2}(u^2+v^2)\right]$ (for $-\infty<u<\infty$, $-\infty<v<\infty$).

(iii) $f(u,v)$ can be written as the product $\frac{1}{2\pi}g(u)h(v)$, where $g(u)$, $h(v)$ are respectively $\exp\left(-\frac{1}{2}u^2\right)$, $\exp\left(-\frac{1}{2}v^2\right)$. Over $(-\infty,\infty)$, these will integrate to 1 if they have the factor $\frac{1}{\sqrt{2\pi}}$. Hence U and V are independent and both are $\mathrm{N}(0,1)$: $f(u)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}$ and $f(v)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}v^2}$, defined over $(-\infty,\infty)$.

(iv) Generate a pair of uniform random variates x, y in [0, 1], by any suitable process to produce independent variates.

(a) Construct u, v as above to give independent N(0,1) variates.

(b) u^2, v^2 are independent χ_1^2 distributed variates. Hence u^2+v^2 is a χ_2^2 variate.

Graduate Diploma, Statistical Theory & Methods, Paper I, 2002. Question 4

(i) $$M_X(t) = E\left[e^{Xt}\right] = \int_0^\infty e^{xt}.\theta e^{-\theta x}dx = \int_0^\infty \theta e^{-(\theta-t)x}dx$$

$$= \theta\left[\frac{-e^{-(\theta-t)x}}{\theta-t}\right]_0^\infty = \frac{\theta}{\theta-t} \qquad \text{(converges for } t < 0\text{).}$$

$$M'_X(t) = \frac{\theta}{(\theta-t)^2}; \qquad M''_X(t) = \frac{2\theta}{(\theta-t)^3}.$$

$$E[X] = M'_X(0) = \frac{1}{\theta}.$$

$$E\left[X^2\right] = M''_X(0) = \frac{2}{\theta^2}, \quad \text{hence } \operatorname{Var}(X) = \frac{2}{\theta^2} - \left(\frac{1}{\theta}\right)^2 = \frac{1}{\theta^2}.$$

(ii) Using the convolution and "linear transformation" results for moment generating functions,

$$M_Z(t) = e^{-t\sqrt{n}}\left\{M_X\left(\frac{\theta t}{\sqrt{n}}\right)\right\}^n = e^{-t\sqrt{n}}\left(1-\frac{t}{\sqrt{n}}\right)^{-n}$$

$$= e^{-t\sqrt{n}}\left\{1+\left(-\frac{t}{\sqrt{n}}\right)\right\}^{-n},$$

so that

$$\ln M_Z(t) = -t\sqrt{n} - n\ln\left\{1+\left(-\frac{t}{\sqrt{n}}\right)\right\}$$

$$= -t\sqrt{n} - n\left(-\frac{t}{\sqrt{n}} - \frac{1}{2}\left(\frac{t}{\sqrt{n}}\right)^2 - \frac{1}{3}\left(\frac{t}{\sqrt{n}}\right)^3 - \ldots\right)$$

$$= -t\sqrt{n} + t\sqrt{n} + \frac{1}{2}t^2 + \frac{1}{3}\frac{t^3}{\sqrt{n}} + \ldots$$

$$\to \frac{1}{2}t^2 \text{ as } n \to \infty$$

so that $M_Z(t) \to e^{-\frac{1}{2}t^2}$ as $n \to \infty$.

This is the mgf of $\mathrm{N}(0,1)$, so $Z \to \mathrm{N}(0,1)$.

Graduate Diploma, Statistical Theory & Methods, Paper I, 2002. Question 5

(i) $F_1\left(u_{(1)}\right) = P\left(U_{(1)} \le u_{(1)}\right) = 1 - P\left(U_{(1)} > u_{(1)}\right) = 1 - \left[1 - F\left(u_{(1)}\right)\right]^n$

$= 1 - \left(1 - u_{(1)}\right)^n$ for U(0,1) (for $0 \le u_{(1)} \le 1$).

Hence $f_1\left(u_{(1)}\right) = n\left(1 - u_{(1)}\right)^{n-1}$ (for $0 \le u_{(1)} \le 1$).

(ii) Using the multinomial expression for one observation at u_1, one at u_2 and $n - 2$ observations greater than u_2,

$$f_{1,2}\left(u_{(1)}, u_{(2)}\right) = \frac{n!}{1!1!(n-2)!} 1.1.\left(1 - F\left(u_{(2)}\right)\right)^{n-2} \qquad \text{(since } f\left(u_{(j)}\right) = 1\text{)}$$

$$= n(n-1)\left(1 - u_{(2)}\right)^{(n-2)} \qquad 0 < u_{(1)},\ u_{(2)} < 1.$$

(iii) Change variables to $W = U_{(2)} - U_{(1)}$, $Z = U_{(1)}$.

Hence $U_{(1)} = Z$ and $U_{(2)} = W + Z$.

$$J = \begin{vmatrix} \dfrac{\partial U_{(1)}}{\partial W} & \dfrac{\partial U_{(1)}}{\partial Z} \\ \dfrac{\partial U_{(2)}}{\partial W} & \dfrac{\partial U_{(2)}}{\partial Z} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 1 & 1 \end{vmatrix} = -1, \text{ so } |J| = 1.$$

$\therefore f(w,z) = n(n-1)\left(1 - \{w+z\}\right)^{n-2} \qquad 0 \le w \le 1,\ 0 \le z \le 1,\ 0 \le w+z \le 1.$

$\therefore f_W(w) = n(n-1)\int_{z=0}^{1-w}(1-w-z)^{n-2}dz$ put $z = y(1 - w)$;
then $1 - w - z = (1 - w)(1 - y)$
and $dz = (1 - w)dy$

$$= n(n-1)\int_0^1 \{(1-w)(1-y)\}^{n-2}(1-w)\,dy$$

$$= n(n-1)(1-w)^{n-1}\int_0^1 (1-y)^{n-2}\,dy$$

$$= n(n-1)(1-w)^{n-1}\left[-\frac{(1-y)^{n-1}}{n-1}\right]_0^1 = n(1-w)^{n-1} \qquad \text{(for } 0 \le w \le 1\text{)},$$

which is the same pdf as that of $U_{(1)}$.

(iv) For $n = 10$, $f_W(w) = 10(1-w)^9 \qquad 0 \le w \le 1$.

$$P(W < 0.1) = \int_0^{0.1} 10(1-w)^9\,dw = \left[-(1-w)^{10}\right]_0^{0.1} = 1 - (0.9)^{10} = 0.6513.$$

(i) (a) P(not found) = P(not in region 1) + P(in 1 but not found)
$= \theta_2 + \theta_3 + \theta_1(1-\alpha) = 1 - \alpha\theta_1$.

(b) Let R_i be the event that the aircraft came down in region I and NF the event that it is not found. By Bayes' theorem,

$$P(R_1 \mid NF) = \frac{P(NF \mid R_1)P(R_1)}{P(NF)} = \frac{(1-\alpha)\theta_1}{1-\alpha\theta_1}.$$

At this stage, $P(NF \mid R_2) = P(NF \mid R_3) = 1$ since R_2, R_3 have not been examined.

Hence $P(R_2 \mid NF) = \dfrac{\theta_2}{1-\alpha\theta_1}$ and $P(R_3 \mid NF) = \dfrac{\theta_3}{1-\alpha\theta_1}$.

(ii) Once all three regions have been searched,

$$P(NF) = P(NF \mid R_1)P(R_1) + P(NF \mid R_2)P(R_2) + P(NF \mid R_3)P(R_3)$$
$$= (1-\alpha)\theta_1 + (1-\alpha)\theta_2 + (1-\alpha)\theta_3 = 1-\alpha.$$

So $P(R_i \mid NF) = \dfrac{P(NF \mid R_i)P(R_i)}{(1-\alpha)} = \dfrac{(1-\alpha)\theta_i}{(1-\alpha)} = \theta_i$.

(iii) Given that the aircraft is actually in region i, then it may only be spotted on sorties numbers $3(k-1)+i$, for k = 1, 2, 3, … . The probability that it is spotted for the first time on sortie number $3(k-1)+i$ is $(1-\alpha)^{k-1}\alpha$, since the previous $(k-1)$ sorties in i were "failures".

Hence $E[X \mid \text{aircraft in region } i] = \sum_{k=1}^{\infty}\{3(k-1)+i\}(1-\alpha)^{k-1}\alpha$

$$= 3\alpha\sum_{k=1}^{\infty}k(1-\alpha)^{k-1} + (i-3)\alpha\sum_{k=1}^{\infty}(1-\alpha)^{k-1}.$$

For a geometric series, we have $1 + y + y^2 + y^3 + \ldots\ldots = \dfrac{1}{1-y}$

and $1 + 2y + 3y^2 + \ldots\ldots = \dfrac{d}{dy}\left(\dfrac{1}{1-y}\right) = \dfrac{1}{(1-y)^2}$.

Hence the above sum is $\left(3\alpha.\dfrac{1}{\alpha^2}\right) + \alpha(i-3).\dfrac{1}{\alpha} = \dfrac{3}{\alpha} + i - 3$.

Therefore $E[X] = \left(\dfrac{3}{\alpha} - 2\right)\theta_1 + \left(\dfrac{3}{\alpha} - 1\right)\theta_2 + \left(\dfrac{3}{\alpha}\right)\theta_3 = \dfrac{3}{\alpha} - 2\theta_1 - \theta_2$.

<u>Graduate Diploma, Statistical Theory & Methods, Paper I, 2002. Question 7</u>

(i) First generate by any available method a pseudo-random number between 0 and 1; call it u.

Now set $F(x) = u$, and solve this equation to find $x = F^{-1}(u)$. This value x is a pseudo-random member of the specified distribution.

If this is to work, F must be easily invertible, either algebraically or numerically.

(ii) (a) $F(x) = 1 - e^{-x}$.

If $u = F(x) = 1 - e^{-x}$, then $x = -\ln(1-u)$.

For the given four numbers, using them as u, we find

$x = 0.183;\ 0.269;\ 1.505;\ 3.442.$

[NOTE: if u is U(0,1), so is $(1 - u)$; so $x = -\ln u$ could be used.]

(b) $F(x) = \int_0^x (4t - 4t^3)\,dt = \left[2t^2 - t^4\right]_0^x = 2x^2 - x^4$ (for $0 \le x \le 1$).

If $u = 2x^2 - x^4$, then we have $x^4 - 2x^2 + u = 0$, i.e. $(x^2 - 1)^2 - 1 + u = 0$, or $x^2 - 1 = -\sqrt{1-u}$ (taking negative square root to obtain $x < 1$), which gives $x = \sqrt{1 - \sqrt{1-u}}$. This gives $x = 0.295;\ 0.355;\ 0.727;\ 0.906$.

(c) For the Poisson distribution, tables can be used to set up the cumulative distribution (e.g. Examination Tables XII) or the c.d.f. can be calculated by hand. When $\lambda = 2$, we have:

		u
$P(X=0) = 0.1353$	so $F(0) = 0.1353$	
$P(X=1) = 0.2707$	so $F(1) = 0.4060$	← 0.167, 0.236
$P(X=2) = 0.2707$	so $F(2) = 0.6767$	
$P(X=3) = 0.1804$	so $F(3) = 0.8571$	← 0.778
$P(X=4) = 0.0902$	so $F(4) = 0.9473$	
$P(X=5) = 0.0361$	so $F(5) = 0.9834$	← 0.968

and so on.

Any value of u up to 0.1352 corresponds to $x = 1$; u from 0.1353 to 0.4059 to $x = 2$; and so on. So we find <u>1, 1, 3, 5</u> as the random sample from the Poisson distribution with mean 2.

F needs to be worked out as far into the tail of the distribution as necessary to use all the given values of u.

Graduate Diploma, Statistical Theory & Methods, Paper I, 2002. Question 8

(i) Markov chain model is given by one-step transition matrix:

	L	D	W
L	0.5	0.4	0.1
D	0.3	0.4	0.3
W	0.2	0.4	0.4

Call this $\mathbf{T}$.

(ii) The two-step matrix is

$$\mathbf{T}^2 = \begin{pmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.4 & 0.4 \end{pmatrix}\begin{pmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.4 & 0.4 \end{pmatrix} = \begin{pmatrix} 0.39 & 0.40 & 0.21 \\ 0.33 & 0.40 & 0.27 \\ 0.30 & 0.40 & 0.30 \end{pmatrix}$$

So having lost game 1, game 3 is won with probability 0.21.

(iii) $\Pi = (\pi_L \ \pi_D \ \pi_W)$, the stationary distribution, is given by

$\Pi = \Pi\mathbf{T}$, i.e.
$$\pi_L = 0.5\pi_L + 0.3\pi_D + 0.2\pi_W$$
$$\pi_D = 0.4\pi_L + 0.4\pi_D + 0.4\pi_W = 0.4 \quad (\text{using } \pi_L + \pi_D + \pi_W = 1)$$
$$\pi_W = 0.1\pi_L + 0.3\pi_D + 0.4\pi_W$$

So, inserting $\pi_D = 0.4$, we have $0.5\pi_L = 0.12 + 0.2\pi_W$
and $0.6\pi_W = 0.12 + 0.1\pi_L$.

$\therefore\ 3.0\pi_W = 0.60 + 0.5\pi_L = 0.60 + 0.12 + 0.2\pi_W$, i.e. $2.8\pi_W = 0.72$.
Hence $\pi_W = 0.2571$ and $\pi_L = 0.24 + 0.4\pi_W = 0.3429$.

The expected number of points per game is $(0\times\pi_L) + (1\times\pi_D) + (3\times\pi_W) =$ 1.1713.