THE ROYAL STATISTICAL SOCIETY

2001 EXAMINATIONS – SOLUTIONS

GRADUATE DIPLOMA

STATISTICAL THEORY AND METHODS

PAPER I

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, 2001. Question 1

(i) For $z = 0, 1, 2, \ldots$

$$P(X+Y=z)=\sum_{x=0}^{z}P(X=x\cap Y=z-x)=\sum_{x=0}^{z}\frac{e^{-\theta}\theta^{x}}{x!}.\frac{e^{-\lambda}\lambda^{z-x}}{(z-x)!} \text{ by independence}$$

$$=\frac{e^{-(\theta+\lambda)}}{z!}\sum_{x=0}^{z}\theta^{x}\lambda^{z-x}\binom{z}{x}=\frac{e^{-(\theta+\lambda)}(\theta+\lambda)^{z}}{z!} \text{ by the binomial theorem.}$$

This is the probability mass function of Poisson, mean $(\theta+\lambda)$.

Hence $X+Y$ has a Poisson distribution, mean $(\theta+\lambda)$.

(ii) $$P(X=x\,|\,X+Y=z)=\frac{P(X=x\cap X+Y=z)}{P(X+Y=z)}=\frac{P(X=x\cap Y=z-x)}{P(X+Y=z)}$$

$$=\frac{\frac{e^{-\theta}\theta^{x}}{x!}.\frac{e^{-\lambda}\lambda^{z-x}}{(z-x)!}}{\frac{e^{-(\lambda+\theta)}(\lambda+\theta)^{z}}{z!}}=\binom{z}{x}\frac{\theta^{x}\lambda^{z-x}}{(\lambda+\theta)^{z}}=\binom{z}{x}\left(\frac{\theta}{\lambda+\theta}\right)^{x}\left(1-\frac{\theta}{\lambda+\theta}\right)^{z-x}.$$

Therefore, given $X+Y=z$, X is binomial $\left(z,\frac{\theta}{\lambda+\theta}\right)$.

(iii) Lecturer A makes $X_1, \ldots, X_6$ mistakes which will follow Poisson (mean = 1.5) independently. Thus $X = X_1 + \ldots + X_6$ is Poisson with mean = 9.

Similarly $Y = Y_1 + \ldots + Y_{12}$, for B, is Poisson with mean 6.

X and Y are independent. Given that $X + Y = 14$, X is binomial $\left(14, \frac{9}{15}\right)$, i.e. $\mathrm{B}\left(14, \frac{3}{5}\right)$.

Therefore $P(X\geq 10\,|\,X+Y=14)=P(X\geq 10\,|\,X\sim \mathrm{B}(14,0.6))$ which is the same as $P(X\leq 4\,|\,X\sim \mathrm{B}(14,0.4))$ and from tables this is 0.2793.

Graduate Diploma, Statistical Theory & Methods, Paper I, 2001. Question 2

(a) Let E_1, E_2, … be a set of mutually exclusive events which exhaust the sample space S (and all $P(E_i)$ are > 0).

Then, for $j = 1, 2, \ldots$, $$P(E_j \mid A) = \frac{P(A \mid E_j)P(E_j)}{\sum_i P(A \mid E_i)P(E_i)}$$

where A is any event in S and $P(A) > 0$.

(b) (i) P(knows answer) = θ, and we assume this leads to the correct answer being written.

P(does not know answer) = $1 - \theta$, in which case the probability of writing the correct answer is 1/5; so the total probability of a correct answer is

$$\theta + \frac{1}{5}(1-\theta) = \frac{1}{5}(1+4\theta).$$

(ii) If X = mark for question,

$$E[X] = 1.\frac{1}{5}(1+4\theta) - \frac{1}{n}\left\{1 - \frac{1+4\theta}{5}\right\}$$
$$= \frac{1}{5}(1+4\theta) + \frac{1}{5n}(4\theta - 4).$$

$E[X] = \theta$ when $5n\theta = n(1+4\theta) + (4\theta - 4)$

or $n\theta - n = 4\theta - 4$ i.e. $n(\theta - 1) = 4(\theta - 1))$

so that $n = 4$.

(iii) Let Y = number of correct answers out of 50, so that

$$Y \sim \text{binomial}\left(50, \frac{1}{5}\{1+4\theta\}\right).$$

If 34 correct answers are given, 16 will be wrong and there will be a deduction of 16/4 = 4 marks, leaving 30.

When $\theta = 0.75$, Y is distributed as B(50, 0.8). Using a Normal approximation,

$P(Y \geq 34) = P\left(Z \geq \frac{33.5 - 40}{\sqrt{8}}\right)$ where $Z \sim$ N(0,1).

$$P\left(Z \geq \frac{-6.5}{\sqrt{8}}\right) = P(Z \geq -2.298) = 0.989.$$

(i) $$E\left[U^m\right] = \int_0^\infty \frac{\theta^\alpha u^{m+\alpha-1} e^{-\theta u}}{\Gamma(\alpha)} du = \frac{\theta^\alpha}{\Gamma(\alpha)} \int_0^\infty u^{m+\alpha-1} e^{-\theta u}\, du$$

$$= \frac{\theta^\alpha}{\Gamma(\alpha)} \int_0^\infty \frac{t^{m+\alpha-1} e^{-t}}{\theta^{m+\alpha}} dt \qquad \text{putting } t = \theta u \text{, so } dt = \theta\, du$$

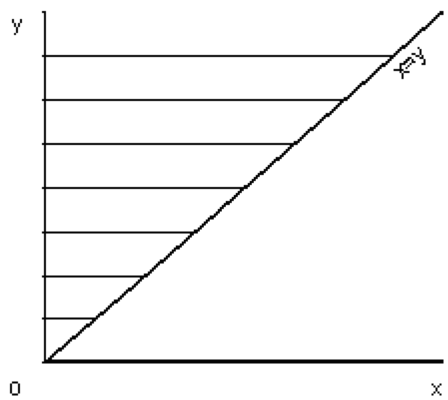
$$= \frac{\theta^\alpha}{\Gamma(\alpha)\theta^{m+\alpha}} \int_0^\infty t^{m+\alpha-1} e^{-t}\, dt = \frac{\theta^\alpha}{\Gamma(\alpha)\theta^{m+\alpha}} . \Gamma(m+\alpha)$$

Hence $E[U] = \dfrac{\theta^\alpha \Gamma(\alpha+1)}{\Gamma(\alpha)\theta^{\alpha+1}} = \dfrac{\alpha}{\theta} = \text{mean}$.

Also, $E\left[U^2\right] = \dfrac{\theta^\alpha \Gamma(\alpha+2)}{\Gamma(\alpha)\theta^{2+\alpha}} = \dfrac{(\alpha+1)\alpha}{\theta^2}$, so variance = $\dfrac{\alpha(\alpha+1)}{\theta^2} - \dfrac{\alpha^2}{\theta^2} = \dfrac{\alpha}{\theta^2}$.

(ii) For fixed x, y lies in (x, ∞).

Therefore the function is defined in the shaded region:



$$f_X(x) = \int_{y=x}^\infty \theta^2 e^{-\theta y} dy = \theta^2 \left[-\frac{1}{\theta} e^{-\theta y} \right]_x^\infty = \theta e^{-\theta x} \qquad (x > 0).$$

Putting $\alpha = 1$ in (i), we see its mean is $\dfrac{1}{\theta}$ and variance $\dfrac{1}{\theta^2}$.

$$f_Y(y) = \int_{x=0}^y \theta^2 e^{-\theta y} dx = \theta^2 e^{-\theta y} [x]_0^y = \theta^2 y e^{-\theta y} \qquad (y > 0).$$

Putting $\alpha = 2$ in (i), we see that mean $= \dfrac{2}{\theta}$ and variance $= \dfrac{2}{\theta^2}$.

$$E[XY] = \int_{y=0}^{\infty} \left\{ \int_{x=0}^{y} \theta^2 xy e^{-\theta y} dx \right\} dy = \int_0^{\infty} \theta^2 y e^{-\theta y} \left[\int_0^{y} x\, dx \right] dy$$

$$= \int_0^{\infty} \frac{1}{2} \theta^2 y^3 e^{-\theta y} dy$$

$$= \frac{1}{2} \theta^2 . \frac{\Gamma(4)}{\theta^4} = \frac{1}{2} \theta^2 . \frac{6}{\theta^4} = \frac{3}{\theta^2} .$$

$$\rho_{XY} = \frac{E[XY] - E[X]E[Y]}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\frac{3}{\theta^2} - \frac{2}{\theta^2}}{\sqrt{\frac{1}{\theta^2} . \frac{2}{\theta^2}}} = \frac{\frac{1}{\theta^2}}{\sqrt{\frac{2}{\theta^4}}} = \frac{1}{\sqrt{2}} = 0.707 .$$

Graduate Diploma, Statistical Theory & Methods, Paper I, 2001. Question 4

(i) By independence, the joint p.d.f. of X and Y is

$$f(x,y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} . \sqrt{\frac{2}{\pi}} e^{-y^2/2} \qquad (-\infty < x < \infty;\ y > 0)$$

$$= \frac{1}{\pi} e^{-(x^2+y^2)/2} \qquad (-\infty < x < \infty;\ y > 0),$$

$X = UV$ and $Y = V$, so $\frac{\partial x}{\partial u} = v$, $\frac{\partial x}{\partial v} = u$, $\frac{\partial y}{\partial u} = 0$ and $\frac{\partial y}{\partial v} = 1$.

Hence the Jacobian J of the transformation from X, Y to U, V is $\begin{vmatrix} v & 0 \\ u & 1 \end{vmatrix}$ i.e. v, so $|J| = v$.

The pdf becomes $f(u,v) = \frac{v}{\pi} e^{-\frac{1}{2}v^2(1+u^2)}$ (for $v > 0$, $-\infty < u < \infty$).

(ii) $$f_u(u) = \int_0^\infty \frac{v}{\pi} e^{-\frac{1}{2}v^2(1+u^2)} dv \qquad (-\infty < u < \infty)$$

$$= \int_0^\infty \frac{1}{\pi} . \frac{1}{1+u^2} . e^{-t}\, dt \quad \text{putting } t = \frac{1}{2}v^2(1+u^2), \text{ so } dt = v(1+u^2)\,dv$$

$$= \frac{1}{(u^2+1)\pi}, \qquad -\infty < u < \infty.$$

(iii) X should be N(0, 1), and W independently χ^2_m.

$U = \frac{X}{Y}$, and X is N(0, 1). Also $Y = \sqrt{\frac{W}{1}}$ where W is χ^2_1, since the square of N(0, 1) is χ^2_1.

[Note that a square root may be positive or negative whereas χ^2 must be positive; hence the definition of Y = |Z|.]

(i) $$M_X(t) = E\left[e^{Xt}\right] = \sum_{x=0}^{n} \binom{n}{x} \theta^x (1-\theta)^{n-x} e^{xt}$$

$$= \sum_{x=0}^{n} \binom{n}{x} \left(\theta e^t\right)^x (1-\theta)^{n-x} = \left(1-\theta+\theta e^t\right)^n$$ by the binomial theorem.

$M'_X(t) = n\left(1-\theta+\theta e^t\right)^{n-1} \theta e^t$, so $M'_X(0) = n\theta = E[X]$.

$M''_X(t) = n(n-1)\left(1-\theta+\theta e^t\right)^{n-2}\left(\theta e^t\right)^2 + n\left(1-\theta+\theta e^t\right)^{n-1}\theta e^t$.

$\therefore\ M''_X(0) = n(n-1)\theta^2 + n\theta = E\left[X^2\right]$.

So $\operatorname{Var}(X) = n(n-1)\theta^2 + n\theta - n^2\theta^2 = n\theta(1-\theta)$.

(ii) $$M_z(t) = E\left[e^{Zt}\right] = \exp\left(\frac{-n\theta t}{\sqrt{n\theta(1-\theta)}}\right) E\left[\exp\left(\frac{Xt}{\sqrt{n\theta(1-\theta)}}\right)\right]$$

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

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

giving $$\ln\{M_z(t)\} = \frac{-n\theta t}{\sqrt{n\theta(1-\theta)}} + n\ln\left(1-\theta+\theta e^{\frac{t}{\sqrt{n\theta(1-\theta)}}}\right).$$

Now,

$$\ln\left\{1+\theta\left(e^{\frac{t}{\sqrt{n\theta(1-\theta)}}}-1\right)\right\} = \ln\left\{1+\theta\left(\frac{t}{\sqrt{n\theta(1-\theta)}} + \frac{t^2}{2n\theta(1-\theta)} + \ldots\right)\right\}$$

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

Hence

$$\ln\left\{M_Z(t)\right\} = \frac{-n\theta t}{\sqrt{n\theta(1-\theta)}} + n\theta\left(\frac{t}{\sqrt{n\theta(1-\theta)}} + \frac{t^2}{2n\theta(1-\theta)} + \ldots\right) - \frac{1}{2}n\theta^2\left(\frac{t^2}{n\theta(1-\theta)} + \ldots\right) + o\left(n^{-1}\right)$$

$$= \frac{t^2}{2(1-\theta)} - \frac{\theta t^2}{2(1-\theta)} + \ldots + o\left(n^{-1}\right)$$

$$= \frac{1}{2}t^2 + \ldots + o\left(n^{-1}\right).$$

Therefore $M_z(t) \rightarrow e^{t^2/2}$, which is the moment generating function of N(0, 1).

Thus Z $\rightarrow$ N(0, 1).

Graduate Diploma, Statistical Theory & Methods, Paper I, 2001. Question 6

(i) The random variable R has probability mass function

$$P(R=r)=\phi(1-\phi)r \qquad (r = 0, 1, 2, \ldots).$$

The pgf is therefore $g_R(t)=\sum_{r=0}^{\infty} t^r\phi(1-\phi)^r = \phi\sum_r \{t(1-\phi)\}^r$

where $|t|<\dfrac{1}{1-\phi}$ for convergence. This gives $\dfrac{\phi}{[1-t(1-\phi)]}$.

$$g_R'(t)=\frac{\phi(1-\phi)}{[1-t(1-\phi)]^2}, \text{ and } E[R]=g_R'(1)=\frac{\phi(1-\phi)}{\{1-(1-\phi)\}^2}=\frac{1-\phi}{\phi}.$$

$$g_R''(t)=\frac{2\phi(1-\phi)^2}{[1-t(1-\phi)]^3}, \quad E[R(R-1)]=g_R''(1)=\frac{2\phi(1-\phi)^2}{\phi^3}=\frac{2(1-\phi)^2}{\phi^2}.$$

So $\mathrm{Var}(R)=E[R(R-1)]+E[\mathrm{R}]-(E[R])^2=\dfrac{2(1-\phi)^2}{\phi^2}+\dfrac{(1-\phi)}{\phi}-\dfrac{(1-\phi)^2}{\phi^2}=\dfrac{1-\phi}{\phi^2}$.

(ii) $P(X=x\mid y)=\dfrac{e^{-y}y^x}{x!} \qquad (x=0, 1, 2, \ldots)$, and so

$$P(X=x)=\int_0^\infty \frac{e^{-y}y^x}{x!}\theta e^{-\theta y}\,dy=\frac{\theta}{x!}.\frac{1}{(1+\theta)^{x+1}}\int_0^\infty t^x e^{-t}\,dt \qquad \text{putting } t=(1+\theta)y$$

$$=\frac{\theta}{(1+\theta)^{x+1}} \quad \text{for } x=0, 1, 2, \ldots .$$

This is the distribution of (i) with $\phi=\dfrac{\theta}{1+\theta}$.

(iii) Therefore from part (i) $E[X] = \frac{1}{\theta}$, $\text{Var}(\text{X}) = \frac{1+\theta}{\theta^2}$.

We have $E\left[E(X \mid Y)\right] = E[Y]$; and Y has mean $\frac{1}{\theta}$. Thus we indeed have $E\left[E(X|Y)\right] = E[X]$ as required.

Also $E\left[\text{Var}(X \mid Y)\right]$ will be $\frac{1}{\theta}$ (note the properties of a Poisson distribution: mean = variance).

Finally, $\text{Var}\left[E(X \mid Y)\right] = \text{Var}(Y) = \frac{1}{\theta^2}$.

So

$$E\left[\text{Var}(X|Y)\right] + \text{Var}\left[E(X|Y)\right] = \frac{1}{\theta} + \frac{1}{\theta^2}$$

which equals $\frac{1+\theta}{\theta^2}$ (= Var(X)) as required.

Graduate Diploma, Statistical Theory & Methods, Paper I, 2001. Question 7

(a) $F_Y(y) = P(Y \le y) = P\left(F^{-1}(U) \le y\right) = P\left(U \le F(y)\right)$.

But U is uniform(0, 1), and so $P(U \le u) = u$ for $0 \le u \le 1$.

Hence $F_Y(y) = P\left(U \le F(y)\right) = F(y)$, the same cdf as X.

$\therefore$ Y has the same distribution as X.

(b) (i)

x	1	2	3	4	…
$P(X = x)$	1/2	1/4	1/8	1/16	
$P(X \le x)$	0.5000	0.7500	0.8750	0.9375	

$u_1 = 0.205$ which is < 0.5, hence $x_1 = 1$.

$u_2 = 0.476$, so also $x_2 = 1$.

$u_3 = 0.879$; $0.8750 < u_3 < 0.9375$, hence $x_3 = 4$.

$u_4 = 0.924$, so also $x_4 = 4$.

Sample is 1, 1, 4, 4.

(ii) For Pareto, $F_X(x) = \int_3^x \frac{18}{t^3}\,dt = \left[-\frac{9}{t^2}\right]_3^x = 1 - \frac{9}{x^2} \quad (x \ge 3)$.

$u = F_X(x) = 1 - \frac{9}{x^2}$ gives $x = \frac{3}{\sqrt{1-u}}$. The random variates therefore are 3.365, 4.144, 8.624, 10.882.

(iii) Using Normal tables, the four values corresponding to u_i are $x_i = \Phi^{-1}(u_i)$, i.e. –0.82, –0.06, 1.17, 1.43.

Graduate Diploma, Statistical Theory & Methods, Paper I, 2001. Question 8

(i) If there are X_i red balls in the first urn at step *i,* write

$$P_{r,s} = P(X_{i+1} = s \mid X_i = r) \quad \text{for } r,\ s = 0,1,...,n\,.$$

Then $P_{0,1} = 1$; also $P_{n,n-1} = 1$.

When $X_i = r$, the first urn contains *r* red and (*n* – *r*) black, and the second (*n* – *r*) red and *r* black.

$$P_{r,r-1} = P(\text{choose red in 1st urn} \mid X_i = r).P(\text{choose black in 2nd urn} \mid X_i = r)$$
$$= \left(\frac{r}{n}\right)^2.$$

$P_{r,r} = 2\left(\frac{r}{n}\right)\left(1-\frac{r}{n}\right)$ by similar arguments; and also

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

Otherwise $P_{r,s} = 0$.

(ii) The probability π_i is $P(i \text{ red balls in 1st urn})$.

$\therefore\ \pi_0 = \left(\frac{1}{n}\right)^2 \pi_1$; and $\pi_n = \left(\frac{1}{n}\right)^2 \pi_{n-1}$. (*)

For *r* = 1, 2, …, *n* – 1,

$$\pi_r = P_{r-1,r}\pi_{r-1} + P_{r,r}\pi_r + P_{r+1,r}\pi_{r+1},$$

i.e. $\pi_r = \left(1-\frac{r-1}{n}\right)^2 \pi_{r-1} + 2\left(\frac{r}{n}\right)\left(1-\frac{r}{n}\right)\pi_r + \left(\frac{r+1}{n}\right)^2 \pi_{r+1}$. (*)

The equations (*) define the process, with $\sum_{r=0}^{n} \pi_r = 1$.

(iii) From (*),

$$\pi_0 = \frac{1}{9}\pi_1$$

$$\pi_3 = \frac{1}{9}\pi_2,$$

and $\pi_0 + \pi_1 + \pi_2 + \pi_3 = 1$.

Also $\pi_1 = (1-0)^2 \pi_0 + 2\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)\pi_1 + \left(\frac{2}{3}\right)^2 \pi_2$; substituting gives

$\frac{1}{9}\pi_1 + \pi_1 + \pi_2 + \frac{1}{9}\pi_2 = 1 = \frac{10}{9}(\pi_1 + \pi_2)$, so $\pi_1 + \pi_2 = \frac{9}{10}$.

Now use $\pi_0 = \frac{1}{9}\pi_1$; $\pi_2 = \frac{9}{10} - \pi_1$; $\pi_3 = \frac{1}{10} - \frac{1}{9}\pi_1$.

Therefore

$$\pi_1 = \pi_0 + \frac{4}{9}\pi_1 + \frac{4}{9}\pi_2$$
$$= \frac{1}{9}\pi_1 + \frac{4}{9}\pi_1 + \frac{4}{9}\left(\frac{9}{10} - \pi_1\right) = \frac{1}{9}\pi_1 + \frac{2}{5}.$$

$\therefore \frac{8}{9}\pi_1 = \frac{2}{5}$, or $\pi_1 = \frac{9}{20}$

$$\pi_0 = \frac{1}{20}$$

$$\pi_2 = \frac{9}{20}$$

$$\pi_3 = \frac{1}{20}.$$