THE ROYAL STATISTICAL SOCIETY

2006 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.

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. $\log_{10}$.

Graduate Diploma, Statistical Theory & Methods, Paper I, 2006. Question 1

(i) The law of total probability for a partition $\{E_i\}$ of S is

$$P(A) = \sum_{i=1}^{n} P(A|E_i)P(E_i).$$

Using $P(A \cap E_j) = P(E_j|A)P(A) = P(A|E_j)P(E_j)$, we have

$$P(E_j|A) = \frac{P(A|E_j)P(E_j)}{P(A)} = \frac{P(A|E_j)P(E_j)}{\sum_{i=1}^{n} P(A|E_i)P(E_i)}.$$

(ii) Let E_i be the event "i is transmitted", for i = 0 or 1, and let A be the event "there is an error at the receiver".

(a)
$$P(A) = P(A|E_0)P(E_0) + P(A|E_1)P(E_1)$$
$$= P\left(X \le 0 \middle| X \sim \mathrm{N}(1, \sigma^2)\right).\tfrac{1}{2} + P\left(X > 0 \middle| X \sim \mathrm{N}(-1, \sigma^2)\right).\tfrac{1}{2}$$
$$= \frac{1}{2}P\left(Z \le -\frac{1}{\sigma}\right) + \frac{1}{2}P\left(Z > \frac{1}{\sigma}\right) \quad \text{where } Z \sim \mathrm{N}(0,1)$$
$$= \Phi\left(-\frac{1}{\sigma}\right) \quad \text{where } \Phi \text{ is the standard Normal distribution function.}$$

When $\sigma = ½$, $P(A) = \Phi(-2) = 0.0228$.

(b) Let U be the random variable denoting the number of voltage values at the receiver that are greater than 0 (out of 3). The receiver decides that 0 was sent if the value of U is 2 or 3, and that 1 was sent if the value of U is 0 or 1.

So we now have

$$P(A) = P(A|E_0)P(E_0) + P(A|E_1)P(E_1)$$
$$= P(U = 0 \text{ or } 1|E_0).\tfrac{1}{2} + P(U = 2 \text{ or } 3|E_1).\tfrac{1}{2}.$$

If E_0 applies, i.e. 0 was sent, we have (see (ii)(a)) that P(voltage value at receiver > 0) = $P(\mathrm{N}(1, \sigma^2) > 0) = 0.9772$. So $U \sim \mathrm{B}(3, 0.9772)$, and $P(U = 0 \text{ or } 1 \mid E_0) = (0.0228)^3 + 3(0.9772)(0.0228)^2 = 0.00154$.

Similarly, if E_1 applies, i.e. 1 was sent, we have P(voltage value at receiver > 0) = $P(\mathrm{N}(-1, \sigma^2) > 0) = 0.0228$. So $U \sim \mathrm{B}(3, 0.0228)$, and $P(U = 2 \text{ or } 3 \mid E_1) = 3(0.0228)^2(0.9772) + (0.0228)^3 = 0.00154$.

$$\therefore P(A) = 0.00154 \times \tfrac{1}{2} + 0.00154 \times \tfrac{1}{2} = 0.00154.$$

Graduate Diploma, Statistical Theory & Methods, Paper I, 2006. Question 2

(i) (a) $F_W(w) = P(W \le w) = P(-w \le U \le w) = F_U(w) - F_U(-w)$

$= F_U(w) - \{1 - F_U(w)\}$ by symmetry

$= 2F_U(w) - 1$ for $w \ge 0$.

$f_W(w) = \frac{d}{dw} F_W(w) = 2 f_U(w)$ for $w \ge 0$.

(b) If $U \sim \mathrm{N}(0, \tau^2)$, which is symmetric about 0, then the result in part (a) gives that $W = |U|$ has pdf $f_w(w) = 2 \times \frac{1}{\tau\sqrt{2\pi}} \exp\left(-\frac{w^2}{2\tau^2}\right)$ for $w \ge 0$.

$$E(W) = \int_0^\infty \sqrt{\frac{2}{\pi\tau^2}}\, w e^{-w^2/2\tau^2}\, dw$$

$$= \sqrt{\frac{2}{\pi\tau^2}} \left[e^{-w^2/2\tau^2} \left(-\tau^2\right) \right]_{w=0}^{\infty} = -\sqrt{\frac{2\tau^2}{\pi}} [0 - 1] = \sqrt{\frac{2\tau^2}{\pi}} .$$

$$E(W^2) = \int_0^\infty \sqrt{\frac{2}{\pi\tau^2}}\, w^2 e^{-w^2/2\tau^2}\, dw \qquad \text{(by parts)}$$

$$= \sqrt{\frac{2}{\pi\tau^2}} \left\{ w.\left[-\tau^2 e^{-w^2/2\tau^2} \right]_0^\infty + \int_0^\infty \tau^2 e^{-w^2/2\tau^2} dw \right\}$$

consider pdf of $\mathrm{N}(0, \tau^2)$

$$= \sqrt{\frac{2}{\pi\tau^2}} \left\{ 0 + \tau^2 . \tau\sqrt{2\pi} . \tfrac{1}{2} \right\} = \tau^2 .$$

Note. An alternative approach is to obtain a general expression for $E(W^m)$ for any integer $m > 0$ using gamma functions:

$$E(W^m) = \frac{2^{m/2}\tau^m}{\sqrt{\pi}} \Gamma\left(\frac{m+1}{2}\right).$$

$$\therefore \mathrm{Var}(W) = \tau^2 - \frac{2\tau^2}{\pi} = \left(1 - \frac{2}{\pi}\right)\tau^2 .$$

(ii) For X, Y independent $\mathrm{N}(\mu, \sigma^2)$ random variables, $U = X - Y \sim \mathrm{N}(0, 2\sigma^2)$. Using (i)(b) with $\tau^2 = 2\sigma^2$ gives $E[|U|] = \frac{2\sigma}{\sqrt{\pi}}$. This is the Gini statistic of $\mathrm{N}(\mu, \sigma^2)$.

Graduate Diploma, Statistical Theory & Methods, Paper I, 2006. Question 3

(i) $P(X=x, Y=y) = \dfrac{20!}{x!\,y!(20-x-y)!}\left(\dfrac{1}{4}\right)^{20-y}\left(\dfrac{1}{2}\right)^{y}$ for x and y from 0 to 20.

$$\therefore P(X=5,\ Y=10) = \frac{20!}{5!\,10!\,5!}\left(\frac{1}{4}\right)^{10}\left(\frac{1}{2}\right)^{10} = 0.04336 \ .$$

(ii) Each plant, independently of all the others, has probability ¼ of having red flowers. The number of plants is fixed (20). These are the conditions for a binomial distribution, so $X \sim$ B(20, ¼).

(iii) As in (ii), the number of plants, W, with white flowers is also B(20, ¼).

$$P(W \le 1) = \left(\frac{3}{4}\right)^{20} + 20\left(\frac{3}{4}\right)^{19}\left(\frac{1}{4}\right) = 0.0243 \ .$$

(iv) Given $Y = y$, there are exactly $20 - y$ plants that are not pink, and they are equally likely to be red or white. Independently, each of these $20 - y$ therefore has probability ½ of being red. Hence the required conditional distribution is B($20 - y$, ½).

(v) Let X be the number of the remaining 15 plants having red flowers. As in part (ii), the distribution of X is binomial, now with $n = 15$: $X \sim$ B(15, ¼).

$$\therefore P(X \ge 3) = 1 - P(X \le 2)$$

$= 1 - 0.2361$ (from tables)

$= 0.7369$.

<u>Graduate Diploma, Statistical Theory & Methods, Paper I, 2006. Question 4</u>

(i) As X and Y are independent, their joint pdf is

$$f_{XY}(x,y) = f_X(x) f_Y(y) = \frac{1}{2\pi\sqrt{xy}} e^{-\frac{1}{2}(x+y)}$$

(for $x > 0$, $y > 0$).

$U = \dfrac{X}{Y}$, $V = Y$; hence $X = UV$ and $Y = V$.

The Jacobian of the transformation is $\begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = v$.

Thus the joint pdf of U and V is

$$f_{UV}(u,v) = f_{XY}(x,y)v$$

$$= v\frac{1}{2\pi\sqrt{uv^2}} e^{-(uv+v)/2} = \frac{1}{2\pi\sqrt{u}} e^{-(u+1)v/2} \qquad \text{for } u > 0,\ v > 0.$$

The marginal distribution of U is

$$f_U(u) = \int_{v=0}^{\infty} \frac{1}{2\pi\sqrt{u}} e^{-(u+1)v/2} dv = \frac{1}{2\pi\sqrt{u}} \left[-\frac{2}{(u+1)} e^{-(u+1)v/2} \right]_{v=0}^{\infty}$$

$$= \frac{1}{\pi(u+1)\sqrt{u}}$$, as required. [Note: this is the pdf of $F_{1,1}$.]

(ii) If W_1, W_2 are independent N(0, σ^2), then $\dfrac{W_1^2}{\sigma^2}$ and $\dfrac{W_2^2}{\sigma^2}$ are independent χ_1^2 random variables.

Hence $U = \left(\dfrac{W_1}{W_2}\right)^2 \sim F_{1,1}$.

Now let $T = \sqrt{U}$. Using the formula for the pdf in a monotonic transformation, the pdf of T can be written down as

$$f_T(t) = f_U(t^2).\frac{du}{dt} = \frac{1}{\pi(1+t^2)t} 2t = \frac{2}{\pi(1+t^2)} \qquad \text{(for } t > 0\text{)}.$$

Graduate Diploma, Statistical Theory & Methods, Paper I, 2006. Question 5

$$f(x)=\frac{\theta^{\alpha}x^{\alpha-1}e^{-\theta x}}{\Gamma(\alpha)} \quad \text{for } x>0, \text{ where } \Gamma(\alpha)=\int_0^{\infty}x^{\alpha-1}e^{-x}dx$$

(i) $$M_X(t)=E\left(e^{tX}\right)=\int_0^{\infty}e^{tx}\frac{\theta^{\alpha}x^{\alpha-1}e^{-\theta x}}{\Gamma(\alpha)}=\frac{\theta^{\alpha}}{\Gamma(\alpha)}\int_0^{\infty}x^{\alpha-1}e^{-(\theta-t)x}dx\ .$$

If $t<\theta$ this integral converges to give (by substituting $(\theta-t)x=u$)

$$M_X(t)=\frac{\theta^{\alpha}}{\Gamma(\alpha)}\frac{1}{(\theta-t)^{\alpha}}\Gamma(\alpha)=\left(\frac{1}{1-(t/\theta)}\right)^{\alpha} \quad \text{(for } t<\theta\text{)}.$$

$E(X)=M_X'(0)$. We have $M_X'(t)=\dfrac{d}{dt}\left(\dfrac{\theta^{\alpha}}{(\theta-t)^{\alpha}}\right)=\dfrac{\alpha\theta^{\alpha}}{(\theta-t)^{\alpha+1}}$, so $E(X)=\dfrac{\alpha\theta^{\alpha}}{\theta^{\alpha+1}}=\dfrac{\alpha}{\theta}$

$E\left(X^2\right)=M_X''(0)$. We have

$$M_X''(t)=\frac{d}{dt}\left(\frac{\alpha\theta^{\alpha}}{(\theta-t)^{\alpha+1}}\right)=\frac{\alpha(\alpha+1)\theta^{\alpha}}{(\theta-t)^{\alpha+2}}, \quad \text{so } E(X^2)=\frac{\alpha(\alpha+1)\theta^{\alpha}}{\theta^{\alpha+2}}=\frac{\alpha(\alpha+1)}{\theta^2}\ .$$

$$\therefore \text{Var}(X)=E\left(X^2\right)-\left[E(X)\right]^2=\frac{\alpha(\alpha+1)}{\theta^2}-\frac{\alpha^2}{\theta^2}=\frac{\alpha}{\theta^2}\ .$$

(ii) $Z=-\sqrt{\alpha}+\left(\dfrac{\theta}{\sqrt{\alpha}}\right)X$; hence, using the "linear transformation" result for mgfs,

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

$$\therefore \log M_Z(t)=-t\sqrt{\alpha}-\alpha\log\left\{1-\frac{t}{\sqrt{\alpha}}\right\} = -t\sqrt{\alpha}-\alpha\left\{-\frac{t}{\sqrt{\alpha}}-\frac{1}{2}\frac{t^2}{\alpha}-\frac{1}{3}\frac{t^3}{\alpha\sqrt{\alpha}}-\dots\right\}$$

$$=\frac{1}{2}t^2+\frac{1}{3}\frac{t^3}{\sqrt{\alpha}}+\dots \rightarrow \frac{1}{2}t^2 \quad \text{as } \alpha\rightarrow\infty\ .$$

Thus $M_Z(t)\rightarrow\exp\left(\frac{1}{2}t^2\right)$ which is the mgf of N(0, 1). So the distribution of Z tends to N(0, 1), i.e. Z is approximately N(0, 1) for large α. Hence, by "unstandardising", X is approximately $\text{N}\left(\dfrac{\alpha}{\theta},\ \dfrac{\alpha}{\theta^2}\right)$ for large α.

Graduate Diploma, Statistical Theory & Methods, Paper I, 2006. Question 6

$$f(x)=\frac{1}{\theta}, \quad \text{so } F(x)=\frac{x}{\theta} \qquad \text{(both for } 0 < x < \theta\text{)}$$

(i) The pdf of $X_{(j)}$ is $\dfrac{n!}{(j-1)!1!(n-j)!}\left[F(x)\right]^{j-1}\left[1-F(x)\right]^{n-j} f(x)$

$$=\frac{n!}{(j-1)!(n-j)!}\frac{x^{j-1}(\theta-x)^{n-j}}{\theta^n} \qquad (0 < x < \theta)$$

$$E\left(X_{(j)}\right)=\frac{n!}{(j-1)!(n-j)!}\int_0^\theta \frac{x^j(\theta-x)^{n-j}}{\theta^n}dx$$

$$=\frac{n!}{(j-1)!(n-j)!}\int_0^\theta \left(\frac{x}{\theta}\right)^j\left(1-\frac{x}{\theta}\right)^{n-j}dx \qquad \text{Put } y=\frac{x}{\theta}$$

$$=\frac{n!\theta}{(j-1)!(n-j)!}\int_0^1 y^j(1-y)^{n-j}dy$$

Use the beta function formula or repeated integration by parts

$$=\frac{n!\theta}{(j-1)!(n-j)!}\frac{j!(n-j)!}{(n+1)!} = \frac{j\theta}{n+1}.$$

$$E\left(X_{(j)}^{\ 2}\right)=\frac{n!}{(j-1)!(n-j)!}\int_0^\theta \frac{x^{j+1}(\theta-x)^{n-j}}{\theta^n}dx \qquad \text{Proceed similarly}$$

$$=\frac{n!\theta^2}{(j-1)!(n-j)!}\frac{(j+1)!(n-j)!}{(n+2)!} = \frac{j(j+1)\theta^2}{(n+1)(n+2)}.$$

$$\therefore \text{Var}\left(X_{(j)}\right) = \frac{j(j+1)\theta^2}{(n+1)(n+2)}-\frac{j^2\theta^2}{(n+1)^2} = \frac{j\theta^2}{n+1}\left\{\frac{j+1}{n+2}-\frac{j}{n+1}\right\}$$

$$=\frac{j\theta^2}{n+1}\frac{\left[(j+1)(n+1)-j(n+2)\right]}{(n+1)(n+2)} = \frac{j(n+1-j)\theta^2}{(n+1)^2(n+2)}.$$

From the $E(X_{(j)})$ result above, $E(U) = E\left(X_{(n)}\right)-E\left(X_{(1)}\right) = \dfrac{n\theta}{n+1}-\dfrac{\theta}{n+1} = \dfrac{n-1}{n+1}\theta$.

Solution continued on next page

(ii) The joint pdf of $X_{(1)}$ and $X_{(n)}$ is

$$g(x_{(1)},x_{(n)}) = n(n-1)\left[F\left(x_{(n)}\right)-F\left(x_{(1)}\right)\right]^{n-2} f\left(x_{(1)}\right)f\left(x_{(n)}\right)$$

$$= \frac{n\left(n-1\right)\left(x_n - x_1\right)^{n-2}}{\theta^n}, \qquad 0 < x_{(1)} < x_{(n)} < \theta .$$

$$\therefore E\left[\left(X_{(n)} - X_{(1)}\right)^2\right] = \frac{n\left(n-1\right)}{\theta^n}\int_{x_{(1)}=0}^{\theta}\int_{x_{(n)}=x_{(1)}}^{\theta}\left(x_{(n)} - x_{(1)}\right)^2\left(x_{(n)} - x_{(1)}\right)^{n-2} dx_{(n)}dx_{(1)}$$

$$= \frac{n\left(n-1\right)}{\theta^n}\int_0^{\theta}\frac{(\theta - x_{(1)})^{n+1}}{n+1}\ dx_{(1)}$$

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

Hence

$$\mathrm{Var}\left(X_{(n)} - X_{(1)}\right) = \frac{n\left(n-1\right)\theta^2}{(n+1)(n+2)} - \left(\frac{n-1}{n+1}\theta\right)^2$$

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

(iii) We have

$$\mathrm{Var}\left(\frac{n+1}{n}X_{(n)}\right) = \left(\frac{n+1}{n}\right)^2\frac{n\theta^2}{(n+1)^2(n+2)} = \frac{\theta^2}{n\left(n+2\right)}$$

$$\mathrm{Var}\left(\frac{n+1}{n-1}U\right) = \left(\frac{n+1}{n-1}\right)^2\frac{2\left(n-1\right)\theta^2}{(n+1)^2(n+2)} = \frac{2\theta^2}{(n-1)(n+2)}$$

Thus the variance of the first of these estimators is smaller, for all n, so use this.

Graduate Diploma, Statistical Theory & Methods, Paper I, 2006. Question 7

(i) (a) For a discrete distribution, first construct the cdf $F(x)$.

x	1	2	3	4	5	6	≥ 7
$P(X=x)$	0.3333	0.2222	0.1481	0.0988	0.0658	0.0439	0.0878
$F(x)$	0.3333	0.5555	0.7036	0.8024	0.8683	0.9122	1

$u_1 = 0.1269$ which is ≤ 0.3333, so x_1 is taken as 1.

$u_2 = 0.2473$ which is ≤ 0.3333, so x_2 is taken as 1.

$u_3 = 0.5107$ which is in the range (0.3333, 0.5555), so x_3 is taken as 2.

$u_4 = 0.9068$ which is in the range (0.8693, 0.9122), so x_4 is taken as 6.

(b) For a continuous distribution, first find the cdf $F(x)$. Here we have

$$F(x) = \int_0^x \frac{dt}{(1+t)^2} = \left[-\frac{1}{1+t}\right]_0^x = -\frac{1}{1+x} + 1 = \frac{x}{1+x} .$$

So a given value u from U(0, 1) gives $u = x/(1 + x)$; so the required random variates are given by $x = u/(1 - u)$.

$u_1 = 0.1269 \rightarrow x_1 = 0.1269/0.8731 = 0.1453.$

$u_2 = 0.2473 \rightarrow x_2 = 0.2473/0.7527 = 0.3286.$

$u_3 = 0.5107 \rightarrow x_3 = 0.5107/0.4893 = 1.0437.$

$u_4 = 0.9068 \rightarrow x_4 = 0.9068/0.0932 = 9.7296.$

(ii) (a) For the exponential distribution with cdf $F(x) = 1 - e^{-\lambda x}$, the inverse cdf method (as in (i)(b)) gives $x = -\frac{1}{\lambda}\log(1-u)$. For each of the machines A, B and C, we have $\lambda = 0.01$.

Simulated lifetime of machine A is $x_A = -\frac{1}{0.01}\log(1-0.1269) = 13.57$.

Simulated lifetime of machine B is $x_B = -\frac{1}{0.01}\log(1-0.2473) = 28.41$.

Simulated lifetime of machine C is $x_C = -\frac{1}{0.01}\log(1-0.5107) = 71.48$.

The repair time has $\lambda = 0.4$, so $x_R = -\frac{1}{0.4}\log(1-0.9068) = 5.93$.

(b) A fails at time 13.57, and is replaced by C. A returns from repair at 13.57 + 5.93 = 19.50. However, B does not fail until 28.41. Hence the repair is complete before the next failure.

Graduate Diploma, Statistical Theory & Methods, Paper I, 2006. Question 8

(i) $P(Y > k) = \phi\left\{(1-\phi)^k + (1-\phi)^{k+1} + (1-\phi)^{k+2} + \cdots\right\}$

$$= \phi(1-\phi)^k\left\{1 + (1-\phi) + (1-\phi)^2 + \cdots\right\} = \frac{\phi(1-\phi)^k}{1-(1-\phi)}$$

$$= (1-\phi)^k$$

$$\therefore P(Y = k + y \mid Y > k) = \frac{P(Y = k+y)}{P(Y > k)} = \frac{\phi(1-\phi)^{k+y-1}}{(1-\phi)^k} = \phi(1-\phi)^{y-1} = P(Y = y).$$

(ii) The probability that a customer who is being served in time interval t completes service in time interval $(t + 1)$ is always ϕ, by (i), irrespective of how long that customer has been waiting for service previously. Hence we have the Markov property.

The transition probabilities are

$$p_{01} = \theta, \quad p_{00} = 1-\theta$$

$$p_{j\,j-1} = \phi(1-\theta), \quad p_{j\,j+1} = \theta(1-\phi), \quad p_{j\,j} = 1-\theta-\phi+2\theta\phi \ .$$

(iii) For $\theta = ¼$, $\phi = ½$, the transition matrix is

$$\mathbf{P} = \begin{bmatrix} 3/4 & 1/4 & 0 & 0 & 0 & \cdots \\ 3/8 & 1/2 & 1/8 & 0 & 0 & \cdots \\ 0 & 3/8 & 1/2 & 1/8 & 0 & \cdots \\ 0 & 0 & 3/8 & 1/2 & 1/8 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} .$$

Hence for the stationary probabilities we have:

$$\frac{3}{4}\pi_0 + \frac{3}{8}\pi_1 = \pi_0 \quad \text{so that } \pi_0 = \frac{3}{2}\pi_1;$$

$$\frac{1}{4}\pi_0 + \frac{1}{2}\pi_1 + \frac{3}{8}\pi_2 = \pi_1 \quad \text{so that } \pi_1 = \frac{1}{2}\pi_0 + \frac{3}{4}\pi_2;$$

$$\frac{1}{8}\pi_{j-1} + \frac{1}{2}\pi_j + \frac{3}{8}\pi_{j+1} = \pi_j \quad \text{for } j \ge 2, \quad \text{so that } \pi_j = \frac{1}{4}\pi_{j-1} + \frac{3}{4}\pi_{j+1} .$$

The given probabilities ($\pi_0 = ½$, $\pi_j = 1/3^j$ for $j = 1, 2, \ldots$) can be shown to satisfy these equations by substitution.