THE ROYAL STATISTICAL SOCIETY

2003 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, 2003. Question 1

(i) $$P(X=x)=\sum_{y=0}^{n-x}P(X=x,Y=y)$$
$$=\sum_{y=0}^{n-x}\frac{n!}{x!y!(n-x-y)!}\theta_1^x\theta_2^y(1-\theta_1-\theta_2)^{n-x-y}$$
$$=\frac{n!}{x!(n-x)!}\theta_1^x\sum_{y=0}^{n-x}\binom{n-x}{y}\theta_2^y(1-\theta_1-\theta_2)^{n-x-y}$$
$$=\binom{n}{x}\theta_1^x\{\theta_2+(1-\theta_1-\theta_2)\}^{n-x}=\binom{n}{x}\theta_1^x(1-\theta_1)^{n-x}$$

which is Binomial(n, θ_1).

It follows by symmetry that Y is Binomial(n, θ_2).

(ii) For $x = 0, 1, \ldots, n-y$,

$$P(X=x|Y=y)=\frac{P(X=x,Y=y)}{P(Y=y)}$$
$$=\frac{n!\theta_1^x\theta_2^y(1-\theta_1-\theta_2)^{n-x-y}}{x!y!(n-x-y)!}\frac{y!(n-y)!}{n!\theta_2^y(1-\theta_2)^{n-y}}$$
$$=\binom{n-y}{x}\left(\frac{\theta_1}{1-\theta_2}\right)^x\left(1-\frac{\theta_1}{1-\theta_2}\right)^{n-y-x},$$

so that, conditional on $Y=y$, X is Binomial$\left(n-y,\ \frac{\theta_1}{1-\theta_2}\right)$.

(iii) P(double six) = $(1/6)^2$ = 1/36. P(no six) = $(5/6)^2$ = 25/36.

The joint distribution of X and Y as defined is given by the multinomial with $\theta_1 = 1/36$, $\theta_2 = 25/36$.

Therefore by (i), $E(X)$ = 10/36 = 5/18, since X is Binomial(10, 1/36).

By (ii), in the case Y = 0, $E(X|\,Y = 0)$ = 10/11, since X will be Binomial(10, 1/11). (There are 11 ways out of 36 of having at least one six.)

(i) (a) The law of total probability is $P(A) = \sum_{i=1}^{n} P(A|E_i)P(E_i)$.

Bayes' Theorem states that $P(E_j|A) = \dfrac{P(A|E_j)P(E_j)}{\sum_{i=1}^{n} P(A|E_i)P(E_i)}$.

(b) Since $\{F_j\}$ partitions S, E_i can be written as the disjoint union of events $\{E_i$ and $F_j\}$. S is the disjoint union of $\{E_i\}$, so S is also the disjoint union of events $\{E_i$ and $F_j\}$. Hence $\{E_i$ and $F_j\}$ partitions S.

Now $P(A) = \sum_{i=1}^{n}\sum_{j=1}^{m} P(A|E_i \text{ and } F_j)P(E_i \text{ and } F_j)$

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

(ii) (a) P(no son haemophiliac)

= P(no son haemophiliac | woman carrier).P(woman carrier)
+ P(no son haemophiliac | woman not carrier).P(woman not carrier)

$$= \left\{\left(\frac{1}{2}\right)^3 \times \left(\frac{1}{2}\right)\right\} + \left\{1 \times \frac{1}{2}\right\} = \frac{9}{16}.$$

P(woman carrier | no son haemophiliac) $= \dfrac{1/16}{9/16} = \dfrac{1}{9}$.

P(daughter carrier)

= P(daughter carrier | woman carrier).P(woman carrier) $= \dfrac{1}{2} \times \dfrac{1}{9} = \dfrac{1}{18}$.

(b) P(at least one girl carrier)

= P(at least one girl carrier | daughter carrier).P(daughter carrier)

$$= \left\{1 - \left(\frac{1}{2}\right)^2\right\} \times \frac{1}{18}$$

$$= \frac{1}{24}.$$

Graduate Diploma, Statistical Theory & Methods, Paper I, 2003. Question 3

(i) The space where X and Y exist jointly is not a rectangular region. It is possible to find points (x, y), e.g. (½, ¾), where both $f(x)$ and $f(y)$ are >0 but $f(x, y) = 0$; thus $f(x, y) \neq f(x).f(y)$.

(ii)
$$E\left(X^r Y^s\right) = \int_0^1 6x^{r+1}\left(\int_0^{1-x} y^s\,dy\right)dx$$

$$= 6\int_0^1 x^{r+1}\left[\frac{(1-x)^{s+1}}{s+1}\right]dx$$

$$= \frac{6}{s+1}\frac{(r+1)!(s+1)!}{(r+s+3)!} \qquad \text{using the result quoted in the paper}$$

$$= \frac{6(r+1)!s!}{(r+s+3)!} \qquad \text{for any non-negative integers } r, s.$$

Hence

$$E(X) = \frac{6.\,2!\,0!}{4!} = \frac{6.2.1}{4.3.2.1} = \frac{1}{2},$$

$$E\left(X^2\right) = \frac{6.\,3!\,0!}{5!} = \frac{6.3.2.1.1}{5.4.3.2.1} = \frac{3}{10}, \quad \text{so } \operatorname{Var}(X) = \frac{3}{10} - \left(\frac{1}{2}\right)^2 = \frac{1}{20}.$$

$$E(Y) = \frac{6.\,1!\,1!}{4!} = \frac{1}{4},$$

$$E\left(Y^2\right) = \frac{6.\,1!\,2!}{5!} = \frac{1}{10}, \quad \text{so } \operatorname{Var}(Y) = \frac{1}{10} - \left(\frac{1}{4}\right)^2 = \frac{3}{80}.$$

$$E(X,Y) = \frac{6.\,2!\,1!}{5!} = \frac{1}{10}, \quad \text{so } \operatorname{Cov}(X,Y) = \frac{1}{10} - \frac{1}{2}.\frac{1}{4} = -\frac{1}{40},$$

and so $$\rho_{XY} = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X).\operatorname{Var}(Y)}} = -\frac{1}{40}\sqrt{\frac{20\times 80}{3}} = -\frac{1}{\sqrt{3}}.$$

(iii) For $0 \le w \le 1$, $P(X+Y \le w) = \int_0^w 6x\left(\int_0^{w-x} dy\right)dx$

$$= 6\int_0^w x(w-x)\,dx = 6\left[\frac{1}{2}wx^2 - \frac{1}{3}x^3\right]_0^w = w^3,$$

so that the cumulative distribution function is $F_W(w) = w^3$ (for $0 \le w \le 1$) and the probability density function is $f_W(w) = 3w^2$ (for $0 \le w \le 1$).

Graduate Diploma, Statistical Theory & Methods, Paper I, 2003. Question 4

(i) As X and Y are independent,

$$f(x,y) = (\text{pdf of } X)(\text{pdf of } Y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{1}{2\sigma^2}\left(x^2+y^2\right)\right)$$

(for $-\infty < x < \infty,\ -\infty < y < \infty$).

For $X = R\cos\phi,\ Y = R\sin\phi,$ the Jacobian is

$$|J| = \begin{vmatrix} \frac{\partial X}{\partial R} & \frac{\partial X}{\partial \phi} \\ \frac{\partial Y}{\partial R} & \frac{\partial Y}{\partial \phi} \end{vmatrix} = \begin{vmatrix} \cos\phi & -R\sin\phi \\ \sin\phi & R\cos\phi \end{vmatrix} = R\left(\cos^2\phi + \sin^2\phi\right) = R\,.$$

So the joint pdf of R, ϕ is $g(r,\phi) = \dfrac{r}{2\pi\sigma^2}\exp\left(-\dfrac{r^2}{2\sigma^2}\right)$, for $0 \le r,\ 0 \le \phi \le 2\pi$.

(ii) The pdf of R is $\displaystyle\int_{\phi=0}^{\phi=2\pi} g(r,\phi)\,d\phi = \frac{r}{\sigma^2}\exp\left(-\frac{r^2}{2\sigma^2}\right)$, for $r \ge 0$.

[This can also be seen because R, ϕ exist in a "rectangular" space, and the joint pdf can be written as $\left(\dfrac{r}{\sigma^2}e^{-r^2/(2\sigma^2)}\right)\left(\dfrac{1}{2\pi}\right)$ which factorises, so R, ϕ are independent.]

(iii) The cumulative distribution function of R is

$$F(r) = \int_0^r \frac{u}{\sigma^2}\exp\left(-\frac{u^2}{2\sigma^2}\right)du = \int_0^{r^2/(2\sigma^2)} e^{-w}\,dw \qquad \left(\text{putting } w = \frac{u^2}{2\sigma^2}\right)$$

$$= 1 - e^{-r^2/(2\sigma^2)}.$$

$\therefore F(k\sigma) = 1 - e^{-k^2/2}$, which is to be 0.5. This gives 0.5 = $e^{-k^2/2}$, or $k = \sqrt{-2\log 0.5} = 1.18$.

(i) $M_X(t) = E\left[e^{tX}\right] = \sum_{x=0}^{\infty} e^{tx} \cdot \frac{e^{-\mu}\mu^x}{x!} = e^{-\mu} \sum_{x=0}^{\infty} \frac{\left(\mu e^t\right)^x}{x!} = e^{-\mu} \cdot \exp\left(\mu e^t\right)$

$$= \exp\left\{\left(e^t - 1\right)\mu\right\}.$$

We have $E[X] = M_X'(0)$ and $E\left[X^2\right] = M_X''(0)$. Differentiating $M_X(t)$ gives

$$M'(t) = \mu e^t \exp\left\{\mu\left(e^t - 1\right)\right\}, \quad \text{so } M'(0) = \mu, \quad \text{and}$$

$$M''(t) = \mu e^t \exp\left\{\mu\left(e^t - 1\right)\right\} + \mu^2 e^{2t} \exp\left\{\mu\left(e^t - 1\right)\right\}, \quad \text{so } M''(0) = \mu + \mu^2.$$

Hence $\mathrm{Var}(X) = E\left[X^2\right] - \left(E[X]\right)^2 = \mu + \mu^2 - (\mu)^2 = \mu$.

(Note. The results for $E[X]$ and $E\left[X^2\right]$ can also be obtained from the power series expansion of $M_X(t)$.)

(ii) $Z = \frac{X - \mu}{\sqrt{\mu}} = \frac{1}{\sqrt{\mu}} X - \sqrt{\mu}$, so (using the "linear transformation" result for moment generating functions) we have

$$M_Z(t) = e^{-t\sqrt{\mu}} \cdot M_X\left(\frac{t}{\sqrt{\mu}}\right) = e^{-t\sqrt{\mu}} \exp\left\{\mu\left(e^{t/\sqrt{\mu}} - 1\right)\right\}.$$

Taking logarithms (base *e*),

$$\log\left(M_Z(t)\right) = -t\sqrt{\mu} + \mu\left(e^{t/\sqrt{\mu}} - 1\right) = -t\sqrt{\mu} + \mu\left(1 + \frac{t}{\sqrt{\mu}} + \frac{t^2}{2\mu} + \frac{t^3}{6\mu^{3/2}} + \ldots - 1\right)$$

$$= \frac{1}{2}t^2 + \frac{t^3}{6\sqrt{\mu}} + \ldots \quad \to \frac{1}{2}t^2 \text{ as } \mu \to \infty.$$

Hence $M_Z(t) \to \exp\left(t^2/2\right)$ as $\mu \to \infty$, and this is the moment generating function of N(0, 1). Hence the limiting distribution of Z is N(0, 1).

(iii) $W = \sum_{i=1}^{n} Y_i$ and the m.g.f. of Y_i is $M_i(t) = \exp\left\{\left(e^t - 1\right)\mu_i\right\}$. By independence, $M_W(t) = \prod_{i=1}^{n} \exp\left\{\left(e^t - 1\right)\mu_i\right\} = \exp\left\{\left(e^t - 1\right)\sum_{i=1}^{n} \mu_i\right\}$, i.e. the same form as the original Poisson m.g.f. but with parameter $\sum \mu_i$, so the distribution of W is Poisson with parameter $\sum \mu_i$.

Graduate Diploma, Statistical Theory & Methods, Paper I, 2003. Question 6

(i) For the Weibull distribution, $F(w)=\int_0^w \alpha\theta t^{\theta-1}e^{-\alpha t^\theta}dt$; put $u=at^\theta$ to give $F(w)=\int_0^{\alpha w^\theta} e^{-u}du=1-\exp(-\alpha w^\theta)$. Thus, from the formula $h(w)=\dfrac{f(u)}{1-F(u)}$, we have $h(w)=\dfrac{\alpha\theta w^{\theta-1}\exp(-\alpha w^\theta)}{\exp(-\alpha w^\theta)}=\alpha\theta w^{\theta-1}$. This hazard function is constant if $\theta = 1$; it decreases as w increases if $\theta < 1$.

(ii) $G(y)=P(Y\le y)=P(X_1\le y \text{ or } X_2\le y)$

$$=P(X_1\le y)+P(X_2\le y)-P(X_1\le y \text{ and } X_2\le y)$$

$$=P(X_1\le y)+P(X_2\le y)-P(X_1\le y)P(X_2\le y) \quad \text{by independence}$$

$$=F_1(y)+F_2(y)-F_1(y)F_2(y).$$

Hence $g(y)=G'(y)=f_1(y)+f_2(y)-f_1(y)F_2(y)-f_2(y)F_1(y) \quad (\text{for } y\ge 0)$.

$$\therefore h(y)=\frac{g(y)}{1-G(y)}=\frac{g(y)}{1-F_1(y)-F_2(y)+F_1(y)F_2(y)}=\frac{g(y)}{(1-F_1(y))(1-F_2(y))}$$

$$=\frac{f_1(y)\{1-F_2(y)\}+f_2(y)\{1-F_1(y)\}}{\{1-F_1(y)\}\{1-F_2(y)\}}=h_1(y)+h_2(y).$$

If X_i is Weibull(α_i,θ), this gives $h(y)=h_1(y)+h_2(y)=\alpha_1\theta y^{\theta-1}+\alpha_2\theta y^{\theta-1}$ $=(\alpha_1+\alpha_2)\theta y^{\theta-1}$, which is the hazard function of Weibull$(\alpha_1+\alpha_2,\theta)$.

(iii) $G(y) = P(\text{both components fail in time } y) = F_1(y)F_2(y)$ by independence. For identical components, $G(y)=\{F(y)\}^2$, which gives $g(y)=2F(y)f(y)$ and so $h(y)=\dfrac{2F(y)f(y)}{1-\{F(y)\}^2}=\dfrac{2F(y)f(y)}{\{1-F(y)\}\{1+F(y)\}}$. Now, $\dfrac{F(y)}{1+F(y)}\le\dfrac{1}{2}$ (as $0 \le F(y) \le 1$), and therefore $h(y)\le\dfrac{f(y)}{1-F(y)}$, as required.

(i) (a) $\binom{15}{2} = 105$, so $P(0) = \binom{10}{2}/105 = \frac{45}{105}$, and similarly $P(1)$ = 50/105 and $P(2)$ = 10/105. Hence the probability function ($f(x)$) and c.d.f. ($F(x)$) are

x	0	1	2
$f(x)$	0.4286	0.4762	0.0952
$F(x)$	0.4286	0.9048	1.0000

The inverse c.d.f. method produces $x = 0$ if the random number is ≤ 0.4286, $x = 1$ if the random number is between 0.4287 and 0.9048, and $x = 2$ for 0.9049 upwards. Hence we obtain 1, 0, 2, 1.

(b) $F(x) = x^3$ (for $0 \le x \le 1$). The inverse c.d.f. method sets $u = F(x) = x^3$, so $x = u^{1/3}$. So we obtain 0.8142, 0.6960, 0.9962, 0.7894.

(ii) Generating a $N(9, (1/2)^2)$ random variable requires a N(0, 1) z, found as $\Phi^{-1}(u)$, followed by a transformation $x = 9 + \frac{1}{2}z$.

For $u = 0.5398$, we get $z = 0.10$ and hence $x = 9.05$.
For $u = 0.3372$, we get $z = -0.42$ and hence $x = 8.79$.
For $u = 0.9887$, we get $z = 2.28$ and hence $x = 10.14$.
For $u = 0.4920$, we get $z = -0.02$ and hence $x = 8.99$.

Beginning at 11.00 a.m. and working in decimals of a minute, the times taken to reach B, C, D, E will be 9.05, 8.79, 10.14, 8.99 minutes. Notice that this means that the bus will need to "wait time" at B and C. The bus leaves B at 11.10 and C at 11.20. It then leaves D at 30.14 minutes past 11.00, to arrive at E at 39.13 minutes past 11.00. It will have waited 0.95 minutes at B, 1.21 minutes at C, and 0 minutes at D.

A sample of arrival times at E could be generated in this way using a larger simulation, and the sample mean used to estimate the expected arrival time. The number of times in the sample, n say, that E is not reached until after 11.40 a.m. could be used in estimating the probability of a late arrival: $\hat{p} = \frac{n}{\text{number of simulations}}$.

Graduate Diploma, Statistical Theory & Methods, Paper I, 2003. Question 8

(i) The states of the Markov Chain are 0 (not obese) and 1 (obese). If X_i is the state reached at age i (i = 0, 1, 2, … years) and $p_{rs} = P\left(X_{i+1} = s \middle| X_i = r\right)$ for r = 0, 1 and s= 0, 1, the transition matrix is $\mathbf{P} = \left[p_{rs}\right] = \begin{bmatrix} 1-\phi & \phi \\ 1-\theta & \theta \end{bmatrix}$.

(ii) The two-step transition matrix is

$$\mathbf{P}^2 = \begin{bmatrix} 1-\phi & \phi \\ 1-\theta & \theta \end{bmatrix} \begin{bmatrix} 1-\phi & \phi \\ 1-\theta & \theta \end{bmatrix} = \begin{bmatrix} (1-\phi)^2 + \phi(1-\theta) & \phi(1-\phi+\theta) \\ (1-\theta)(1-\phi+\theta) & \theta^2 + \phi(1-\theta) \end{bmatrix}.$$

All children are non-obese (state 0) at 0 years. So the probability that a child is obese (state 1) at 2 years is given by the "top right" element of $\mathbf{P}^2$, i.e. it is $\phi(1-\phi+\theta)$.

(iii) The proportion of children who have never been obese at any stage up to and including 3 years is $(p_{00})^3 = (1-\phi)^3$.

(iv) $p_{i+1} = \theta p_i + \phi(1-p_i) = \phi + (\theta-\phi)p_i$.

Inserting i = 0 in the expression given in the question gives $\dfrac{1-(\theta-\phi)^0}{1-(\theta-\phi)}$ which equals 0 as required (all children are non-obese at age 0). Now supposing the result holds for p_i ($i \geq 0$), we have

$$p_{i+1} = \phi + (\theta-\phi).\frac{1-(\theta-\phi)^i}{1-(\theta-\phi)}\phi = \phi\left\{1 + \frac{\theta-\phi}{1-(\theta-\phi)}\left[1-(\theta-\phi)^i\right]\right\}$$

$$= \phi\frac{1-(\theta-\phi)+(\theta-\phi)-(\theta-\phi)^{i+1}}{1-(\theta-\phi)} = \frac{1-(\theta-\phi)^{i+1}}{1-(\theta-\phi)}\phi.$$

Hence by induction the result is true for all $i \geq 0$.

(v) As i increases, $p_i \to \dfrac{1-0}{1-(\theta-\phi)}\phi$ since $\theta - \phi < 1$, i.e. $p_i \to \dfrac{\phi}{1-(\theta-\phi)}$.

For $\theta = 0.8$ and $\phi = 0.1$, $p_i \to \dfrac{0.1}{1-0.7} = \dfrac{1}{3}$, so we expect approximately one-third of this adult population to be obese.