THE ROYAL STATISTICAL SOCIETY

2010 EXAMINATIONS – SOLUTIONS

GRADUATE DIPLOMA

MODULE 1

PROBILITY DISTRIBUTIONS

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. $\log_{10}$.

© RSS 2010

Graduate Diploma, Module 1, 2010. Question 1

$$P(X = n) = \frac{K}{n(n+2)}, \qquad n = 1, 2, 3, \ldots .$$

(i) For $N \geq 2$, we have $\sum_{n=1}^{N} \frac{K}{n(n+2)} = \frac{K}{2}\sum_{n=1}^{N}\left(\frac{1}{n} - \frac{1}{n+2}\right)$ which, on expanding and then cancelling terms in pairs, collapses to $\frac{K}{2}\left(1 + \frac{1}{2} - \frac{1}{N+1} - \frac{1}{N+2}\right)$.

Thus as $N \to \infty$ the sum converges to $3K/4$, and this immediately implies that $K = 4/3$ for X to have a probability distribution.

(ii) $P(X$ is an odd number$)$

$$= \frac{K}{2}\sum_{n \text{ odd}}\left(\frac{1}{n} - \frac{1}{n+2}\right) = \frac{K}{2}\left\{\left(1 - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{5}\right) + \left(\frac{1}{5} - \frac{1}{7}\right) + \ldots\right\} = \frac{K}{2}.1,$$

i.e. $P(X$ is an odd number$) = 2/3$.

(iii) (a) It follows from part (ii) that $P(X$ is an even number$) = 1/3$.

$X + Y$ is even if X and Y are both even or both odd. By independence, the probability of the former is $(1/3)^2$ and of the latter is $(2/3)^2$. So the total probability that $X + Y$ is even is 5/9.

(b) $$P(X \text{ even} \mid X + Y \text{ even}) = \frac{P(X \text{ even } \cap \ X + Y \text{ even})}{P(X + Y \text{ even})}.$$

The numerator is $P(X$ even and Y even$)$ which, by independence, is $(1/3)(1/3)$. The denominator has been found to be 5/9 in part (iii)(a). So the required probability is $\frac{1/9}{5/9} = \frac{1}{5}$.

Graduate Diploma, Module 1, 2010. Question 2

(i) Let $h(z)$ denote the pgf of S_N, so that $h(z) = E\left(z^{S_N}\right)$.

For a *fixed* value n of N ($n \geq 1$), the pgf of S_n is simply $\{f(z)\}^n$ using the standard result for the pgf of a sum of independent random variables.

We also have $S_0 = 0$ (defined in the question), so the pgf of S_0 is simply 1.

$$\therefore h(z) = E\left[E\left(z^{S_N} \mid N\right)\right] = E\left[\left(f(z)\right)^N\right] = g\left(f(z)\right).$$

(ii) $N \sim$ Poisson(λ). The pgf of this is

$$g(z) = E\left(z^N\right) = \sum_{n=0}^{\infty} z^n \frac{e^{-\lambda}\lambda^n}{n!} = e^{-\lambda}\sum_{n=0}^{\infty}\frac{(\lambda z)^n}{n!} = e^{-\lambda}e^{\lambda z} = e^{-\lambda+\lambda z}.$$

The pgf of each X_i is simply $f(z) = z^0(1-p) + z^1 p = 1 - p + pz$.

Therefore, by the result in part (i), the pgf of S_N is

$$h(z) = \exp\left(-\lambda + \lambda(1 - p + pz)\right) = \exp\left(-p\lambda + p\lambda z\right).$$

This is the pgf of Poisson($p\lambda$) so, by the uniqueness of pgfs, the distribution of S_N is Poisson($p\lambda$).

(iii) Now $f(z)$ denotes the pgf of the number of casualties in any road accident,

$$f(z) = z^0 p_0 + z^1 p_1 + z^2 p_2 + \ldots = p_0 + p_1 z + p_2 z^2 + \ldots\ .$$

Also, $g(z)$ is the pgf of Poisson(1), i.e. $g(z) = e^{z-1}$.

Therefore the pgf of the number of casualties next Sunday afternoon is

$$h(z) = g\left(f(z)\right) = \exp\left(f(z) - 1\right) = \exp\left(p_0 - 1 + p_1 z + p_2 z^2 + \ldots\right)$$

$$= \exp\left(p_0 - 1\right).\exp\left(p_1 z + p_2 z^2 + \ldots\right)$$

$$= \exp\left(p_0 - 1\right).\left\{1 + \left(p_1 z + p_2 z^2 + \ldots\right) + \frac{\left(p_1 z + p_2 z^2 + \ldots\right)^2}{2!} + \ldots\right\}.$$

This expands as a power series in z of the form $h(z) = a_0 + a_1 z + a_2 z^2 + \ldots$, and the required probability is just

$$a_0 + a_1 + a_2 = \left(1 + p_1 + p_2 + \frac{p_1^2}{2}\right)\exp\left(p_0 - 1\right).$$

Graduate Diploma, Module 1, 2010. Question 3

The pdf of $W = X + Y$ is $h(w) = \int_{-\infty}^{\infty} f(x)\, f(w-x)\, dx$.

(i) The range of W is given by $0 < w < 4$. We have $f(x) = x/2$ for $0 < x < 2$.

The pdf of W is

$$h(w) = \int_0^2 f(x)\, f(w-x)\, dx \qquad \text{for all } w \ (0 < w < 4)$$

$$= \begin{cases} (\text{for } 0 < w < 2) & \displaystyle\int_0^w \frac{x}{2}\frac{w-x}{2}\, dx \\ (\text{for } 2 < w < 4) & \displaystyle\int_{w-2}^2 \frac{x}{2}\frac{w-x}{2}\, dx \end{cases}$$

$$= \begin{cases} (0 < w < 2) & = \dfrac{1}{4}\left[\dfrac{wx^2}{2} - \dfrac{x^3}{3}\right]_{x=0}^{x=w} = \dfrac{1}{4}\left[\dfrac{w^3}{2} - \dfrac{w^3}{3}\right] = \dfrac{w^3}{24}. \\ (2 < w < 4) & = \dfrac{1}{4}\left[\dfrac{wx^2}{2} - \dfrac{x^3}{3}\right]_{x=w-2}^{x=2} = \dfrac{1}{4}\left[2w - \dfrac{8}{3} - \dfrac{w(w-2)^2}{2} + \dfrac{(w-2)^3}{3}\right] \\ & = \dfrac{1}{24}\left(-w^3 + 24w - 32\right). \end{cases}$$

(ii) Let the pdf and cdf of $V = (X-1)^2$ be $g(v)$ and $G(v)$. The range of V is given by $0 < v < 1$. For this range, we have

$$G(v) = P\left((X-1)^2 \le v\right) = P\left(1 - \sqrt{v} \le X \le 1 + \sqrt{v}\right)$$

$$= F\left(1+\sqrt{v}\right) - F\left(1-\sqrt{v}\right) \quad \text{where } F \text{ denotes the cdf of } X.$$

Hence, differentiating,

$$g(v) = \frac{f\left(1+\sqrt{v}\right)}{2\sqrt{v}} + \frac{f\left(1-\sqrt{v}\right)}{2\sqrt{v}} = \frac{\left(1+\sqrt{v}\right)+\left(1-\sqrt{v}\right)}{4\sqrt{v}} = \frac{1}{2\sqrt{v}}$$

(for $0 < v < 1$).

Graduate Diploma, Module 1, 2010. Question 4

For $n = 1, 2, 3, \ldots$, the event $X = n$ occurs when the first $n - 1$ trials are failures and the nth trial is a success. By independence, the probability of this is simply

$$P(X = n) = pq^{n-1} \quad (n = 1, 2, 3, \ldots) \quad \text{where } q = 1 - p \text{ as usual.}$$

$$\therefore E(X) = \sum_{n=1}^{\infty} npq^{n-1} = p + 2pq + 3pq^2 + \ldots = p(1-q)^{-2} = \frac{1}{p}.$$

[There are various methods for summing this series; for example, it can be considered as a "GP of GPs",

$p + pq + pq^2 + \ldots$
$\quad + pq + pq^2 + \ldots$
$\quad\quad + pq^2 + \ldots$
$\quad\quad\quad + \ldots$]

To find Var(X), consider first

$$E\left(X(X-1)\right) = \sum_{n=1}^{\infty} n(n-1)\, pq^{n-1} = 2pq\sum_{n=2}^{\infty}\binom{n}{2} q^{n-2} = 2pq(1-q)^{-3} = \frac{2q}{p^2}.$$

$$\therefore \text{Var}(X) = E\left(X(X-1)\right) + E(X) - \left\{E(X)\right\}^2 = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{q}{p^2}.$$

To find $E(Y)$, we use $E(Y) = E(E(Y \mid X))$. Given $X = n$, the distribution of Y is B(n, p), so $E(Y \mid X) = np$. $\therefore E(Y) = E(Xp) = (1/p)p = 1$.

To find Var(Y), we use $\text{Var}(Y) = E(\text{Var}(Y \mid X)) + \text{Var}(E(Y \mid X))$.

We have $\text{Var}(Y \mid X = n) = npq$, so the first term is $E(Xpq) = (1/p)pq = q$. Also, the second term is $\text{Var}(Xp) = p^2\text{Var}(X) = p^2(q/p^2) = q$. $\therefore \text{Var}(Y) = 2q$.

The result $E(Y) = 1$ should cause no surprise because the number of trials in the original experiment that produced the value of X was constructed so as to have exactly one success.

As the probability of success, p, increases, we would anticipate that fewer trials would be needed to produce the first success. The variance of the number of successes in a given number, n, of trials is proportional to n. Thus we can anticipate that an increase in p will result in a decrease in Var(Y) – as indeed happens, since $\text{Var}(Y) = 2(1 - p)$.

Graduate Diploma, Module 1, 2010. Question 5

(i) A_n occurs when X_n is the largest value among $\{X_1, X_2, \ldots, X_n\}$. By symmetry, any of these values is equally likely to be the largest, so $P(A_n) = 1/n$.

(ii) Without loss of generality, suppose that $m < n$. Now suppose that X_n is the largest of all the observations, so that A_n has occurred. There are $(n-1)!$ ways of arranging $X_1, X_2, \ldots, X_{n-1}$, and all these orderings are equally likely. Consider any of these $(n-1)!$ arrangements. By symmetry, the probability that the mth is the largest of the first m is $1/m$, i.e. X_m is the largest of the first m in a fraction $1/m$ of all the $(n-1)!$ arrangements. Thus $P(A_m \mid A_n) = \{(n-1)!/m\}/(n-1)! = 1/m = P(A_m)$. Thus A_m and A_n are independent.

(iii) Define $I_r = \begin{cases} 1 & \text{if } A_r \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$, so that $S_n = \sum_{r=1}^{n} I_r$. We have $P(I_r = 1) = 1/r$.

Therefore $E(I_r) = 1/r$. Also, $E(I_r^2) = 1/r$, and thus $\text{Var}(I_r) = \frac{1}{r} - \frac{1}{r^2}$.

$$\therefore E(S_n) = \sum_{r=1}^{n} \frac{1}{r} \quad \text{and} \quad \text{Var}(S_n) = \sum_{r=1}^{n} \text{Var}(I_r) = \sum_{r=1}^{n} \left(\frac{1}{r} - \frac{1}{r^2} \right).$$

(iv) When $n = 100$, the results in part (iii) and the approximations quoted in the question give $E(S_{100}) \approx \log 100 = 4.605$ and $\text{Var}(S_{100}) \approx \log 100 - (\pi^2/6) = 1.72^2$.

Using the approximating Normal distribution with continuity correction, we therefore have

$$P(S_{100} \geq 10) \approx P\left(\text{N}(4.605, 1.72^2) > 9.5\right)$$

$$= P\left(\text{N}(0,1) > \frac{9.5 - 4.605}{1.72}\right) = P\left(\text{N}(0,1) > 2.846\right) = 0.0023.$$

<u>Graduate Diploma, Module 1, 2010. Question 6</u>

(i) Clearly $E(U) = 0.5$. $\therefore$ $E(W) = E(V) + KE(U - 0.5) = E(V)$.

We have $E(V) = \int \exp(u^2) f(u)\,du$ where $f(u)$ is the pdf of U, i.e.

$$E(V) = \int_0^1 \exp(u^2).1du = \theta \quad \text{using the notation introduced in the question.}$$

(ii) $\text{Cov}(U, V) = E(UV) - E(U)E(V) = E(UV) - \tfrac{1}{2}\theta$.

$$E(UV) = \int_0^1 u\exp(u^2)\,du = \left[\frac{\exp(u^2)}{2}\right]_0^1 = \frac{e-1}{2}.$$

$$\therefore \text{Cov}(U,V) = \frac{e-1}{2} - \frac{\theta}{2} = \frac{e-1-\theta}{2}, \text{ as required.}$$

(iii) Let $\sigma^2 = \text{Var}(V)$. Then $\text{Var}(W) = \sigma^2 + K^2\text{Var}(U) + 2K\text{Cov}(U, V)$. The value of $\text{Var}(U)$ is 1/12 [this may be quoted or is easily found as $E(U^2) - \{E(U)\}^2$] and $\text{Cov}(U, V)$ has been obtained in part (ii). $\text{Var}(W)$ is a quadratic in K with positive coefficient of K^2, so it has a unique minimum which we easily find by differentiation. We have

$$d\text{Var}(W)/dK = 2K\text{Var}(U) + 2\text{Cov}(U, V) = K/6 + 2\text{Cov}(U, V),$$

and setting this equal to zero gives $K = -12\text{Cov}(U, V) = -6(e - 1 - \theta)$.

(iv) Given the preliminary estimate $\theta \approx 1.46$, we take $K = -6(e - 1 - 1.46) = -1.55$. With this value of K, calculate the value of W for each value of U. We expect that W will have small variance, and thus computing the mean of a large number of these values of W is expected to give the correct value of θ to a high degree of precision.

Graduate Diploma, Module 1, 2010. Question 7

(i) For any $t > 0$, we have $P(-2\log U > t) = P(U < \exp(-t/2)) = \exp(-t/2)$. Thus $P(-2\log U < t) = 1 - \exp(-t/2)$ and, by differentiating, the density function of $-2\log U$ is $(1/2)\exp(-t/2)$ for $t > 0$ – the exponential distribution with mean 2.

(ii) The joint density of X and Y is $f(x,y) = \frac{1}{2\pi}\exp\left(-\frac{x^2+y^2}{2}\right)$ for all x and y.

For $x = r\cos\theta$ and $y = r\sin\theta$, we have

$$\frac{\partial x}{\partial r} = \cos\theta, \quad \frac{\partial x}{\partial \theta} = -r\sin\theta, \quad \frac{\partial y}{\partial r} = \sin\theta, \quad \frac{\partial y}{\partial \theta} = r\cos\theta.$$

Hence the Jacobian is $\begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r\left(\cos^2\theta + \sin^2\theta\right) = r$.

Also, we have $x^2 + y^2 = r^2$.

So the joint density of (R, Θ) is

$$g(r,\theta) = \frac{1}{2\pi}\exp\left(-\frac{r^2}{2}\right).r = \frac{r}{2\pi}\exp\left(-\frac{r^2}{2}\right)$$

as required, and this is clearly valid for $r > 0$ and $0 < \theta < 2\pi$.

Thus the density function of R is

$$\int_0^{2\pi} \frac{r}{2\pi}\exp\left(-\frac{r^2}{2}\right)d\theta = r\exp\left(-r^2/2\right), \quad \text{for } r > 0.$$

(iii) *Hence:*

For $v > 0$, we have $P\left(R^2 < v\right) = P\left(R < \sqrt{v}\right) = \int_0^{\sqrt{v}} r\exp(-r^2/2)dr$

$$= \left[-\exp\left(-r^2/2\right)\right]_0^{\sqrt{v}} = 1 - \exp\left(-v/2\right).$$

Differentiating, the density function is $\exp(-v/2)/2$ for $v > 0$ – the exponential distribution with mean 2.

Otherwise:

As $R^2 = X^2 + Y^2$, it is the sum of the squares of two independent standard Normal variables and hence has the χ^2 distribution with 2 d.f. – which is the same as the exponential distribution with mean 2.

Graduate Diploma, Module 1, 2010. Question 8

(i) Let the four variables be W_1, W_2, W_3, W_4. Then we have

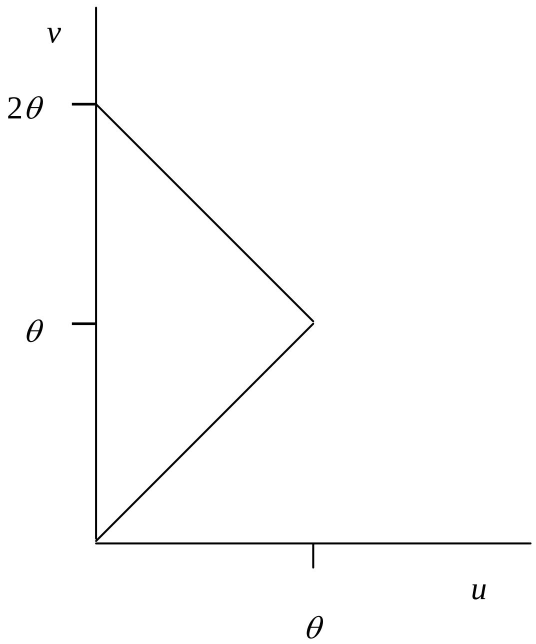
$$P(X > x,\ Y < y) = P(x < W_1, W_2, W_3, W_4 < y)$$

$$= \{P(x < W < y)\}^4 \quad \text{by independence and identical distribution of the } W_i$$

$$= (y - x)^4/\theta^4 \quad = F(x, y) \text{ say.}$$

The joint density is $-\dfrac{\partial^2 F(x,y)}{\partial x \partial y} = \dfrac{12(y-x)^2}{\theta^4}$, for $0 < x < y < \theta$.

(ii) With $U = Y - X$ and $V = Y + X$, we have $Y = (U + V)/2$ and $X = (V - U)/2$. Each relevant partial derivative ($\partial x/\partial u$ etc) is $\pm½$, so the Jacobian is (½)(½) + (½)(½) = ½. So the joint density of U and V is $6u^2/\theta^4$. This is clearly valid for $0 < u < \theta$; also, given a value u of U, the value of V must be between u and $2\theta - u$. So the joint density is non-zero for $0 < u < \theta$ and $u < v < 2\theta - u$. A sketch of this region is shown below.



Solution continued on next page

(iii) The marginal density of U is $\int_u^{2\theta-u} \frac{6u^2}{\theta^4}\,dv = \frac{12u^2(\theta-u)}{\theta^4}$, for $0 < u < \theta$.

For $0 < v < \theta$, the marginal density of V is $\int_0^v \frac{6u^2}{\theta^4}\,du = \frac{2v^3}{\theta^4}$.

For $\theta < v < 2\theta$, the marginal density of V is $\int_0^{2\theta-v} \frac{6u^2}{\theta^4}\,du = \frac{2(2\theta-v)^3}{\theta^4}$.

(iv) From the information given, clearly

$c_1 = 5$, $c_2 = 5/4$, $c_3 = 5/3$, $c_4 = 1$.

The respective variances are simply

$5^2.(2\theta^2/75) = 2\theta^2/3$,

$(5/4)^2.(2\theta^2/75) = \theta^2/24$,

$(5/3)^2.(\theta^2/25) = \theta^2/9$,

$1^2.(\theta^2/15) = \theta^2/15$.

The second of these is the smallest, i.e. c_2Y has the smallest variance.