STAT743 FOUNDATIONS OF STATISTICS (PART II)

Fall 2019

Assignment 2 Solutions

Q. 1 **a)** Let $Y = |X|$ then we have

$$
\begin{aligned}
F_Y(y) &= \mathrm{P}(Y \leqslant y) \\
&= \mathrm{P}(|X| \leqslant y) \\
&= \begin{cases} \mathrm{P}(-y \leqslant X \leqslant y) & \text{if } y > 0 \\ 0 & \text{if } y \leqslant 0 \end{cases} \\
&= \begin{cases} \Phi(y) - \Phi(-y) & \text{if } y > 0 \\ 0 & \text{if } y \leqslant 0 \end{cases}
\end{aligned}
$$

where $\Phi(x)$ is the standard normal cumulative distribution function.

To get the pdf of y we can take derivatives of the cdf

$$
\begin{aligned}
f_Y(y) &= \frac{d}{dy} F_Y(y) \\
&= \begin{cases} \frac{d}{dy}\Phi(y) - \frac{d}{dy}\Phi(-y) & \text{if } y > 0 \\ 0 & \text{if } y \leqslant 0 \end{cases} \\
&= \begin{cases} \phi(y) + \phi(-y) & \text{if } y > 0 \\ 0 & \text{if } y \leqslant 0 \end{cases} \\
&= \begin{cases} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-y^2/2} + \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-(-y)^2/2} & \text{if } y > 0 \\ 0 & \text{if } y \leqslant 0 \end{cases} \\
&= \begin{cases} \sqrt{\frac{2}{\pi}}\mathrm{e}^{-y^2/2} & \text{if } y > 0 \\ 0 & \text{if } y \leqslant 0 \end{cases}
\end{aligned}
$$

[7 marks]

From the pdf given above we get

$$\begin{aligned} \mathrm{E}(Y) &= \int_0^\infty \sqrt{\frac{2}{\pi}} y \mathrm{e}^{-y^2/2}\, dy \\ &= \int_0^\infty \sqrt{\frac{2}{\pi}} \mathrm{e}^{-u}\, du \qquad (\text{where } u = y^2/2) \\ &= -\sqrt{\frac{2}{\pi}} \mathrm{e}^{-u} \bigg|_{u=0}^{\infty} \\ &= \sqrt{\frac{2}{\pi}} \end{aligned}$$

[2 marks]

We can get $E(Y^2)$ using integration by parts but it is easier to note that

$$E(Y^2) = \mathrm{E}(X^2) = \mathrm{Var}(X) = 1$$

Hence we get

$$\mathrm{Var}(Y) = 1 - \frac{2}{\pi}$$

[2 marks]

b) Casella and Berger 4.19

(i) First let $Y = (X_1 - X_2)/\sqrt{2}$. Then we can use Theorems 2.9 and 4.5 in my notes (Theorems 2.3.15 and 4.2.12 in the textbook) along with the moment generating function of the normal on Page 625 of the text book to see that

$$M_Y(t) = M_{X_1}\left(\frac{t}{\sqrt{2}}\right) M_{X_1}\left(\frac{t}{\sqrt{2}}\right) = \exp\left\{\frac{t^2}{4}\right\} \exp\left\{\frac{t^2}{4}\right\} = \exp\left\{\frac{t^2}{2}\right\}$$

and so $Y \sim \mathrm{Normal}(0, 1)$.
Then we can use the result of Example 2.1.9 in the textbook to note that

$$Z = \frac{(X_1 - X_2)^2}{2} = Y^2 \sim \chi_1^2$$

and so the pdf of Z is

$$f_Z(z) = \begin{cases} \dfrac{1}{\sqrt{2\pi}} \dfrac{1}{\sqrt{z}} \mathrm{e}^{-z/2} & \text{for } z > 0 \\ 0 & \text{for } z \leqslant 0 \end{cases}$$

[5 marks]

(ii) We cannot transform to the two random variables described in the question as that is not one-to-one. Instead let us consider the one-to-one transformation

$$\begin{aligned} y_1 &= \frac{x_1}{x_1+x_2} \\ y_2 &= x_1+x_2 \end{aligned} \quad \Longrightarrow \quad \begin{aligned} x_1 &= y_1 y_2 \\ x_2 &= (1-y_1)y_2 \end{aligned}$$

The Jacobian of this transformation is

$$|J| = \begin{vmatrix} y_2 & -y_2 \\ y_1 & 1-y_1 \end{vmatrix} = |y_2|$$

Next we need to find the supports for the transformed random variables

$$\begin{aligned} 0 &< x_1 < \infty \\ 0 &< x_2 < \infty \end{aligned} \quad \Longrightarrow \quad \begin{aligned} 0 &< y_1 y_2 < \infty \\ 0 &< (1-y_1)y_2 < \infty \end{aligned} \quad \Longrightarrow \quad \begin{aligned} 0 &< y_1 < 1 \\ 0 &< y_2 < \infty \end{aligned}$$

Thus we get the joint pdf of (Y_1, Y_2) to be

$$\begin{aligned} &f_{Y_1,Y_2}(y_1,y_2) \\ &= f_{X_1,X_2}\big(y_1y_2,(1-y_1)y_2\big)|y_2| \\ &= f_{X_1}(y_1y_2)f_{X_2}((1-y_1)y_2)|y_2| \\ &= \begin{cases} \left(\frac{1}{\Gamma(\alpha)}(y_1y_2)^{\alpha-1}\mathrm{e}^{-y_1y_2}\right)\left(\frac{1}{\Gamma(\beta)}((1-y_1)y_2)^{\beta-1}\mathrm{e}^{-(1-y_1)y_2}\right)y_2 & \begin{array}{l} 0<y_1<1 \\ 0<y_2<\infty \end{array} \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} \frac{1}{\Gamma(\alpha)\Gamma(\beta)}y_1^{\alpha-1}(1-y_1)^{\beta-1}y_2^{\alpha+\beta-1}\mathrm{e}^{-y_2} & 0<y_1<1, 0<y_2<\infty \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}y_1^{\alpha-1}(1-y_1)^{\beta-1}\right)\left(\frac{1}{\Gamma(\alpha+\beta)}y_2^{\alpha+\beta-1}\mathrm{e}^{-y_2}\right) & 0<y_1<1, 0<y_2<\infty \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Since this joint density factors into a product of two marginal densities we can say that Y_1 and Y_2 are independent and that

$$Y_1 = \frac{X_1}{X_1+X_2} \sim \text{Beta}(\alpha,\beta)$$

Interchanging X_1 and X_2 also shows that

$$\frac{X_2}{X_1+X_2} \sim \text{Beta}(\beta,\alpha)$$

We also note that $X_1+X_2 \sim \text{Gamma}(\alpha+\beta,1)$ independently of either of these two Beta random variables. **[9 marks]**

Q. 2 **a) Casella and Berger 4.26**

(i) This is a case where one variable, W, is discrete and the other, Z, is continuous. We will first find $\mathrm{P}(Z \leqslant z, W = i)$ for $i = 0, 1$ and $z > 0$.

$$
\begin{aligned}
\mathrm{P}(Z \leqslant z, W = 0) &= \mathrm{P}(\min(X, Y) \leqslant z, Y \leqslant X) \\
&= \mathrm{P}(Y \leqslant z, Y \leqslant X) \\
&= \int_0^z \int_y^\infty \frac{1}{\lambda} \mathrm{e}^{-x/\lambda} \frac{1}{\mu} \mathrm{e}^{-y/\mu} \, dx \, dy \\
&= \int_0^z \frac{1}{\mu} \mathrm{e}^{-y/\mu} \int_y^\infty \frac{1}{\lambda} \mathrm{e}^{-x/\lambda} \, dx \, dy \\
&= \int_0^z \frac{1}{\mu} \mathrm{e}^{-y/\mu} \mathrm{e}^{-y/\lambda} \, dy \\
&= \int_0^z \frac{1}{\mu} \exp\left\{ -\left(\frac{1}{\mu} + \frac{1}{\lambda} \right) y \right\} \, dy \\
&= \frac{1}{\mu} \left(\frac{1}{\mu} + \frac{1}{\lambda} \right)^{-1} \left(1 - \exp\left\{ -\left(\frac{1}{\lambda} + \frac{1}{\mu} \right) z \right\} \right) \\
&= \frac{\lambda}{\lambda + \mu} \left(1 - \exp\left\{ -\left(\frac{1}{\lambda} + \frac{1}{\mu} \right) z \right\} \right) \\
\mathrm{P}(Z \leqslant z, W = 1) &= \mathrm{P}(X \leqslant z, X \leqslant Y) \\
&= \int_0^z \int_x^\infty \frac{1}{\lambda} \mathrm{e}^{-x/\lambda} \frac{1}{\mu} \mathrm{e}^{-y/\mu} \, dy \, dx \\
&= \frac{\mu}{\lambda + \mu} \left(1 - \exp\left\{ -\left(\frac{1}{\lambda} + \frac{1}{\mu} \right) z \right\} \right)
\end{aligned}
$$

Hence the joint density/mass function of the random vector (Z, W) is found by taking derivatives with respect to z to get

$$
f_{Z,W}(z, w) = \begin{cases} \dfrac{1}{\mu} \exp\left\{ -\left(\dfrac{1}{\lambda} + \dfrac{1}{\mu} \right) z \right\} & z > 0, w = 0 \\ \dfrac{1}{\lambda} \exp\left\{ -\left(\dfrac{1}{\lambda} + \dfrac{1}{\mu} \right) z \right\} & z > 0, w = 1 \\ 0 & \text{otherwise} \end{cases}
$$

[9 marks]

(ii) To show independence it suffices to show that

$$\mathrm{P}(Z \leqslant z, W = i) = \mathrm{P}(Z \leqslant z)\mathrm{P}(W = i) \quad i = 0, 1, \; z > 0$$

$$\begin{aligned}
\mathrm{P}(W = 0) &= \mathrm{P}(Y \leqslant X) \\
&= \int_0^\infty \int_y^\infty \frac{1}{\lambda} \mathrm{e}^{-x/\lambda} \frac{1}{\mu} \mathrm{e}^{-y/\mu}\, dx\, dy \\
&= \int_0^\infty \frac{1}{\mu} \exp\left\{ - \left(\frac{1}{\lambda} + \frac{1}{\mu} \right) y \right\} dy \\
&= \frac{1}{\mu} \left(\frac{1}{\lambda} + \frac{1}{\mu} \right)^{-1} \\
&= \frac{\lambda}{\lambda + \mu} \\
\mathrm{P}(W = 1) &= 1 - \mathrm{P}(W = 0) = \frac{\mu}{\mu + \lambda}
\end{aligned}$$

$$\begin{aligned}
\mathrm{P}(Z \leqslant z) &= \mathrm{P}(Z \leqslant z, W = 0) + \mathrm{P}(Z \leqslant z, W = 1) \\
&= 1 - \exp\left\{ - \left(\frac{1}{\lambda} + \frac{1}{\mu} \right) z \right\}
\end{aligned}$$

Thus Z and W are independent.

[5 marks]

b) Casella & Berger 4.30

(i) We can use the results from Theorem 4.3 of my notes (Theorems 4.4.3 and 4.4.7 in the textbook). We note that

$$X \sim \mathrm{Uniform}(0, 1) \;\Rightarrow\; \mathrm{E}(X) = \frac{1}{2},\; \mathrm{E}(X^2) = \frac{1}{3},\; \mathrm{Var}(X) = \frac{1}{12}$$

from the properties on Page 626 of the textbook with $a = 0, b = 1$. Hence we have

$$\mathrm{E}(Y) \;=\; \mathrm{E}\big(\mathrm{E}(Y \mid X)\big) \;=\; \mathrm{E}(X) \;=\; \frac{1}{2}.$$

[2 marks]

$$\mathrm{Var}(Y) \;=\; \mathrm{E}\big(\mathrm{Var}(Y \mid X)\big) + \mathrm{Var}\big(\mathrm{E}(Y \mid X)\big) \;=\; \mathrm{E}\left(X^2\right) + \mathrm{Var}(X) \;=\; \frac{1}{3} + \frac{1}{12} \;=\; \frac{5}{12}$$

[2 marks]

$$\mathrm{E}(XY) \;=\; \mathrm{E}\big(\mathrm{E}(XY \mid X)\big) \;=\; \mathrm{E}\big(X\mathrm{E}(Y \mid X)\big) \;=\; \mathrm{E}\left(X^2\right) \;=\; \frac{1}{3}$$

and so

$$\mathrm{Cov}(X,Y) \;=\; \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) \;=\; \frac{1}{3} - \frac{1}{2} \times \frac{1}{2} \;=\; \frac{1}{12}$$

[2 marks]

(ii) Let us consider the bivariate transformation

$$\begin{array}{rcl} u & = & \dfrac{y}{x} \\ v = x & & \end{array} \quad\Longrightarrow\quad \begin{array}{rcl} x & = & v \\ y = uv & & \end{array}$$

which has Jacobian

$$|J| \;=\; \begin{vmatrix} 0 & 1 \\ v & u \end{vmatrix} \;=\; |v|$$

The support of the transformed variables is found as

$$\begin{array}{c} 0 < x < 1 \\ -\infty < y < \infty \end{array} \quad\Longrightarrow\quad \begin{array}{c} 0 < v < 1 \\ -\infty < uv < \infty \end{array} \quad\Longrightarrow\quad \begin{array}{c} -\infty < u < \infty \\ 0 < v < 1 \end{array}$$

Hence the joint density of (U,V) is

$$\begin{aligned} f_{U,V}(u,v) &= f_{X,Y}(v,uv)|v| \\ &= f_{Y|X}(uv \mid v) f_X(v)|v| \\ &= \left(\frac{1}{\sqrt{2\pi v^2}} \exp\left\{-\frac{(uv-v)^2}{2v^2}\right\}\right) \times 1 \times v \quad \text{for } u \in \mathbb{R}, 0 < v < 1 \\ &= \left(\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(u-1)^2}{2}\right\}\right) \times 1 \quad \text{for } u \in \mathbb{R}, 0 < v < 1 \end{aligned}$$

Since this joint density factors for all $(u,v) \in \mathbb{R}^2$ we have that U and V are independent and that the marginal distribution of U is the normal$(1,1)$ distribution.

[5 marks]

Q. 3 **a)** We will show that the cdf of Y is the same as that of X and so Y is standard normal. Let y be an arbitrary real number.

$$\begin{aligned}
F_Y(y) &= \mathrm{P}(Y \leqslant y) \\
&= \mathrm{P}(Y \leqslant y \mid U < 0.5)\mathrm{P}(U < 0.5) + \mathrm{P}(Y \leqslant y \mid U \geqslant 0.5)\mathrm{P}(U \geqslant 0.5) \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(Law of Total Probability)} \\
&= \frac{1}{2}\mathrm{P}(X < y) + \frac{1}{2}\mathrm{P}(-X < y) \\
&= \frac{1}{2}\mathrm{P}(X < y) + \frac{1}{2}\mathrm{P}(X > -y) \\
&= \frac{1}{2}\Phi(y) + \frac{1}{2}\big(1 - \Phi(-y)\big) \\
&= \frac{1}{2}\Phi(y) + \frac{1}{2}\Phi(-y) \qquad \text{(symmetry of } \phi(x) \text{ about 0)} \\
&= \Phi(y)
\end{aligned}$$

[6 marks]

b) Define the Bernoulli random variable $Z = I(U < 0.5)$ then from Theorem 4.3 in my notes (Theorems 4.4.3 and 4.4.7 in the textbook) we have

$$\begin{aligned}
\mathrm{E}(XY) &= \mathrm{E}\big(\mathrm{E}(XY \mid Z)\big) \\
&= \mathrm{E}(XY \mid Z = 1)\mathrm{P}(Z = 1) + \mathrm{E}(XY \mid Z = 0)\mathrm{P}(Z = 0) \\
&= \frac{1}{2}\mathrm{E}(X^2) + \frac{1}{2}\mathrm{E}(-X^2) \\
&= \frac{1}{2}\mathrm{E}(X^2) - \frac{1}{2}\mathrm{E}(-X^2) \\
&= 0
\end{aligned}$$

And, since X and Y are both standard normal we have $\mathrm{E}(X) = \mathrm{E}(Y) = 0$ so

$$\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) = 0 - 0 \times 0 = 0$$

[4 marks]

There are a number of ways to show that X and Y are not independent. One way is to show that for at least one pair of sets $A \subset \mathbb{R}$, and $B \subset \mathbb{R}$ we have

$$\mathrm{P}(X \in A, Y \in B) \neq \mathrm{P}(X \in A)\mathrm{P}(Y \in B)$$

I will consider $A = B = (-\infty, a)$ for some constant $a < 0$ then we have

$$\mathrm{P}(X \in A)\mathrm{P}(Y \in B) = \mathrm{P}(X < a)\mathrm{P}(Y < a) = \Phi(a) \times \Phi(a)$$

And we also have

$$
\begin{aligned}
&\mathrm{P}(X \in A, Y \in B) \\
&\quad = \quad Pr(X < a, Y < a) \\
&\quad = \quad Pr(X < a, Y < a \mid U < 0.5)\mathrm{P}(U < 0.5) + Pr(X < a, Y < a \mid U \geqslant 0.5)\mathrm{P}(U \geqslant 0.5) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(Law of Total Probability)} \\
&\quad = \quad \frac{1}{2}\mathrm{P}(X < a, X < a) + \frac{1}{2}\mathrm{P}(X < a, -X < a) \\
&\quad = \quad \frac{1}{2}\mathrm{P}(X < a) + \frac{1}{2}\mathrm{P}(X < a, X > -a) \\
&\quad = \quad \frac{1}{2}\mathrm{P}(X < a) + \frac{1}{2} \times 0 \qquad \text{(since } a < 0\text{)} \\
&\quad = \quad \frac{1}{2}\Phi(a)
\end{aligned}
$$

And since $a < 0$ we have $\Phi(a) < 0.5$ and so

$$Pr(X < a, Y < a) \;>\; \mathrm{P}(X < a)\mathrm{P}(Y < a)$$

Hence X and Y cannot be independent. **[5 marks]**

c) The conditional distribution of Y given $X = x$ can be found by finding the conditional cdf.

$$
\begin{aligned}
F_{Y|X}(y \mid x) &= \mathrm{P}(Y \leqslant y \mid X = x) \\
&= \mathrm{P}(Y \leqslant y \mid X = x, U < 0.5)\mathrm{P}(U < 0.5) + \mathrm{P}(Y \leqslant y \mid X = x, U \geqslant 0.5)\mathrm{P}(U \geqslant 0.5) \\
&= \frac{1}{2}\mathrm{P}(X \leqslant y \mid X = x) + \frac{1}{2}\mathrm{P}(X \geqslant -y \mid X = x)
\end{aligned}
$$

Now if $y < -|x|$ then $\mathrm{P}(X \leqslant y \mid X = x) = \mathrm{P}(X \geqslant -y \mid X = x) = 0$ and so $F_{Y|X}(y \mid x) = 0$.

If $y \geqslant |x|$ then $\mathrm{P}(X \leqslant y \mid X = x) = \mathrm{P}(X \geqslant -y \mid X = x) = 1$ and so $F_{Y|X}(y \mid x) = 1$.

If $x < 0$ and $x \leqslant y < -x$ then $\mathrm{P}(X \leqslant y \mid X = x) = 1$ and $\mathrm{P}(X \geqslant -y \mid X = x) = 0$ and if $x > 0$ and $-x \leqslant y > x$ then $\mathrm{P}(X \leqslant y \mid X = x) = 0$ and $\mathrm{P}(X \geqslant -y \mid X = x) = 1$. In either of these situations we get $F_{Y|X}(y \mid x) = 0.5$.

Hence the conditional cdf of Y given $X = x$ is

$$F_{Y|X}(y \mid x) \;=\; \begin{cases} 0 & y \;<\; -|x| \\ 0.5 & -|x| \;\leqslant\; y \;<\; |x| \\ 1 & y \;\geqslant\; |x| \end{cases}$$

and this is the cdf corresponding to the discrete random variable with probability mass function

$$f_{Y|X}(y \mid x) \;=\; \begin{cases} 0.5 & y \;=\; -|x| \\ 0.5 & y \;=\; |x| \\ 0 & \text{otherwise} \end{cases}$$

[6 marks]

Since this is conditional pmf of $Y \mid X = x$ we get the joint distribution of (X, Y) to be

$$f_{X,Y}(x, y) \;=\; f_X(x) f_{Y|X}(y \mid x) \;=\; \begin{cases} \dfrac{1}{2\sqrt{2\pi}} \mathrm{e}^{-x^2/2} & x \in \mathbb{R} \text{ and } y \in \{-|x|, |x|\} \\ 0 & \text{otherwise} \end{cases}$$

which is clearly not of the form given Definition 4-8 of my notes (Definition 4.5.10 in the textbook). **[4 marks]**

Q. 4 Note that, contrary to what the book says, the geometric is only properly defined if $0 < \theta < 1$ so I will assume that in this question.

a) The random vector (X, Y) is discrete so it is necessary to express the probability mass of the new random vector in terms of the old. That is we need to express $\mathrm{P}(U = u, V = v)$ in terms of $f_{X,Y}$ as described on Page 4-23 of my notes (Page 156-157 in the text).

First we note that the support of this joint distribution is $\{(u, v) : u = 1, 2, \ldots,\ v = 0, \pm 1, \pm 2, \ldots\}$. It is easiest to take the three cases $v < 0$, $v > 0$ and $v = 0$ separately and then we shall put them all together in a single formula if possible.

To have $v < 0$ we require $X < Y$ and so $U = X$. Thus

$$\begin{aligned}
\mathrm{P}(U = u, V = v) &= \mathrm{P}(X = u, X - Y = v) \\
&= \mathrm{P}(X = u, Y = u - v) \\
&= \theta(1-\theta)^{u-1}\theta(1-\theta)^{u-v-1} \\
&= \theta^2(1-\theta)^{2u-v-2} \quad \text{for} \quad u = 1, 2, \ldots, \\
&\qquad\qquad\qquad\qquad\qquad\ \ v = -1, -2, \ldots \\
&= \theta^2(1-\theta)^{2u+|v|-2} \quad \text{for} \quad u = 1, 2, \ldots, \\
&\qquad\qquad\qquad\qquad\qquad\ \ v = -1, -2, \ldots
\end{aligned}$$

To have $v > 0$ we require $X > Y$ and so $U = Y$. Thus

$$\begin{aligned}
\mathrm{P}(U = u, V = v) &= \mathrm{P}(Y = u, X - Y = v) \\
&= \mathrm{P}(X = u + v, Y = u) \\
&= \theta^2(1-\theta)^{2u+v-2} \quad \text{for} \quad u = 1, 2, \ldots, \\
&\qquad\qquad\qquad\qquad\qquad\ \ v = 1, 2, \ldots \\
&= \theta^2(1-\theta)^{2u+|v|-2} \quad \text{for} \quad u = 1, 2, \ldots, \\
&\qquad\qquad\qquad\qquad\qquad\ \ v = 1, 2, \ldots
\end{aligned}$$

Finally we can only have $v = 0$ if $X = Y$ and so

$$\begin{aligned}
\mathrm{P}(U = u, V = 0) &= \mathrm{P}(X = u, Y = u) \\
&= \theta^2(1-\theta)^{2u-2} \quad \text{for } u = 1, 2, \ldots
\end{aligned}$$

Putting all of these results together we get

$$f_{U,V}(u, v \mid \theta) = \theta^2(1-\theta)^{2u+|v|-2} = \left[(1-\theta)^{2u}\right]\theta^2(1-\theta)^{|v|-2}$$

for $u = 1, 2, \ldots,\ v = 0, \pm 1, \pm 2, \ldots$. **[7 marks]**

Since this joint pmf factors into a function of u times a function of v, for all possible u and v, we have that the random variables U and V are independent from Lemma 4.1 of my notes (Lemma 4.2.7 in the textbook). **[1 mark]**

To get the probability mass function of U we note that $f_U(u) \propto (1-\theta)^{2u}$ so we only need to get the constant of proportionality

$$\sum_{u=1}^{\infty}(1-\theta)^{2u} = \sum_{u=1}^{\infty}\left((1-\theta)^2\right)^u = \frac{(1-\theta)^2}{1-(1-\theta)^2}$$

using standard results for summing geometric series with ratio $r=(1-\theta)^2<1$.

Hence the pmf for U is

$$f_U(u \mid \theta) = \frac{1-(1-\theta)^2}{(1-\theta)^2}\left((1-\theta)^2\right)^u = \left(1-(1-\theta)^2\right)\left((1-\theta)^2\right)^{u-1} \quad \text{for } u=1,2,\ldots$$

That is, $U \sim \text{geometric}(p=\theta(2-\theta))$. **[3 marks]**

To get the marginal for V it is simplest to note that, since U and V are independent,

$$f_V(v \mid \theta) = \frac{f_{U,V}(u,v)}{f_U(u)} = \frac{\theta^2(1-\theta)^{2u+|v|-2}}{\theta(2-\theta)(1-\theta)^{2(u-1)}} = \frac{\theta}{2-\theta}(1-\theta)^{|v|} \quad \text{for } v=0,\pm1,\pm2,\ldots$$

As far as I know, this distribution is not a named distribution but it is a discrete analogue of the double exponential distribution. **[3 marks]**

b) (i) First let us give the joint distribution of (Y,Λ)

$$\begin{aligned}
f_{Y,\Lambda}(y,\lambda \mid \alpha,\beta) &= f_{Y|\Lambda}(y \mid \lambda) f_\Lambda(\lambda \mid \alpha,\beta) \\
&= \begin{cases} \left(\dfrac{\lambda^y \mathrm{e}^{-\lambda}}{y!}\right)\left(\dfrac{1}{\Gamma(\alpha)\beta^\alpha}\lambda^{\alpha-1}\mathrm{e}^{-\lambda/\beta}\right) & y=0,1,2,\ldots;\lambda>0 \\ 0 & \text{otherwise} \end{cases} \\
&= \begin{cases} \dfrac{1}{y!\Gamma(\alpha)\beta^\alpha}\lambda^{y+\alpha-1}\exp\left\{-\left(1-\dfrac{1}{\beta}\right)\lambda\right\} & y=0,1,2,\ldots;\lambda>0 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

Then for $y=0,1,2,\ldots$ we have

$$\begin{aligned}
f_Y(y \mid \alpha,\beta) &= \int_{-\infty}^{\infty} f_{Y,\Lambda}(y,\lambda \mid \alpha,\beta)\,d\lambda \\
&= \int_0^\infty \frac{1}{y!\Gamma(\alpha)\beta^\alpha}\lambda^{y+\alpha-1}\exp\left\{-\left(1-\frac{1}{\beta}\right)\lambda\right\}d\lambda \\
&= \frac{1}{y!\Gamma(\alpha)\beta^\alpha}\int_0^\infty \lambda^{y+\alpha-1}\exp\left\{-\left(1-\frac{1}{\beta}\right)\lambda\right\}d\lambda \\
&= \frac{1}{y!\Gamma(\alpha)\beta^\alpha} \times \frac{\Gamma(y+\alpha)}{\left(1+\frac{1}{\beta}\right)^{y+\alpha}}
\end{aligned}$$

Thus the probability mass function for Y from this hierarchical model is

$$f_Y(y) = \begin{cases} \dfrac{\Gamma(y+\alpha)}{y!\Gamma(\alpha)} \dfrac{\beta^y}{(\beta+1)^{y+\alpha}} & \text{for } y = 0, 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases}$$

[5 marks]

Now if we suppose that α is a positive integer then $\Gamma(\alpha) = (\alpha - 1)!$ and $\Gamma(y+\alpha) = (y + \alpha - 1)!$ so we can rewrite the pmf for postive integers y as

$$f_Y(y \mid \alpha, \beta) = \frac{(y+\alpha-1)!}{y!(\alpha-1)!} \frac{\beta^y}{(\beta+1)^{y+\alpha}} = \binom{y+r-1}{y} \left(\frac{1}{\beta+1}\right)^r \left(1 - \frac{1}{\beta+1}\right)^y$$

and we recognise this as the probability mass function of a negative binomial with $r = \alpha$ and $p = 1/(\beta + 1)$. **[2 marks]**

(ii) To find the mean and variance of Y, we will use Theorem 4.3 in my notes (Theorems 4.4.3 and 4.4.7 in the textbook) and standard results for the mean and variance of the Poisson and Gamma distributions as given on Pages 622 and 624 of the textbook.

$$\mathrm{E}(Y) = \mathrm{E}\big(\mathrm{E}(Y \mid \Lambda)\big) = \mathrm{E}(\Lambda) = \alpha\beta$$

[1 mark]

$$\mathrm{Var}(Y) = \mathrm{E}\big(\mathrm{Var}(Y \mid \Lambda)\big) + \mathrm{Var}\big(\mathrm{E}(Y \mid \Lambda)\big) = \mathrm{E}(\Lambda) + \mathrm{Var}(\Lambda) = \alpha\beta + \alpha\beta^2 = \alpha\beta(1+\beta)$$

[2 marks]

Now, if we set $\mathrm{E}(Y) = \alpha\beta = \mu$ then we see that

$$\mathrm{Var}(Y) = \alpha\beta + \frac{1}{\alpha}(\alpha\beta)^2 = \mathrm{E}(Y) + \frac{1}{\alpha}\big(\mathrm{E}(Y)\big)^2$$

[1 mark]

For a regular Poisson random variable we have $\mathrm{Var}(Y) = \mathrm{E}(Y)$ and so the use of this mixture distribution gives increased variability relative to the Poisson (since $\alpha > 0$). This is called *overdispersion* in models for count data and is very commonly observed in practice so being able to model it using this mixture is very useful.