SCHOOL OF FINANCE AND APPLIED STATISTICS

STATISTICAL INFERENCE

(STAT3013/STAT8027)

TUTORIAL 0 - REVISION EXERCISES - SOLUTIONS

1.(a) Clearly the sample space for U is $S_U = \{0, 1, 2, 3, 4\}$, and we can easily calculate:

$$p_U(0) = Pr(U = 0) = Pr(X = 0, Y = 0) = 0.1$$

$$p_U(1) = Pr(U = 1) = Pr\{(X = 0, Y = 1) \text{ or } (X = 1, Y = 0)\} = 0.1 + 0.25 = 0.35$$

$$p_U(2) = Pr(U = 2) = Pr\{(X = 0, Y = 2) \text{ or } (X = 1, Y = 1) \text{ or } (X = 2, Y = 0)\}$$
$$= 0.2 + 0 + 0.05 = 0.25$$

$$p_U(3) = Pr(U = 3) = Pr\{(X = 1, Y = 2) \text{ or } (X = 2, Y = 1)\} = 0.2 + 0.05 = 0.25$$

$$p_U(4) = Pr(U = 4) = Pr(X = 2, Y = 2) = 0.05$$

(b) The marginal *pmf* of X is:

$$p_X(0) = Pr(X = 0) = Pr(X = 0, 0 \le Y \le 2) = 0.1 + 0.1 + 0.2 = 0.4$$
$$p_X(1) = Pr(X = 1) = Pr(X = 1, 0 \le Y \le 2) = 0.25 + 0 + 0.2 = 0.45$$
$$p_X(2) = Pr(X = 2) = Pr(X = 2, 0 \le Y \le 2) = 0.05 + 0.05 + 0.05 = 0.15$$

and similarly the *pmf* of Y is:

$$p_Y(0) = Pr(Y = 0) = Pr(Y = 0, 0 \le X \le 2) = 0.1 + 0.25 + 0.05 = 0.4$$
$$p_Y(1) = Pr(Y = 1) = Pr(Y = 1, 0 \le X \le 2) = 0.1 + 0 + 0.05 = 0.15$$
$$p_Y(2) = Pr(Y = 2) = Pr(Y = 2, 0 \le X \le 2) = 0.2 + 0.2 + 0.05 = 0.45$$

Now, clearly X and Y are not independent since, for example, $Pr(X = 0, Y = 0) = 0.1 \neq 0.16 = Pr(X = 0)Pr(Y = 0)$.

(c) Using the multiplication rule for independent random variables, we have:

Probability of all possible (x_1, y_1) pairs:

Values of X_1:	Values of Y_1: 0	1	2
0	$0.4 \times 0.4 = 0.16$	$0.4 \times 0.15 = 0.06$	$0.4 \times 0.45 = 0.18$
1	$0.45 \times 0.4 = 0.18$	$0.45 \times 0.15 = 0.0675$	$0.45 \times 0.45 = 0.2025$
2	$0.15 \times 0.4 = 0.06$	$0.15 \times 0.15 = 0.0225$	$0.15 \times 0.45 = 0.0675$

(d) Similar to part *(a)*, the *pmf* of U_1 is calculated as:

$$p_{U_1}(0) = Pr(U_1 = 0) = Pr(X_1 = 0, Y_1 = 0) = 0.16$$

$$p_{U_1}(1) = Pr(U_1 = 1) = Pr\{(X_1 = 0, Y_1 = 1) \text{ or } (X_1 = 1, Y_1 = 0)\} = 0.06 + 0.18 = 0.24$$

$$p_{U_1}(2) = Pr(U_1 = 2) = Pr\{(X_1 = 0, Y_1 = 2) \text{ or } (X_1 = 1, Y_1 = 1) \text{ or } (X_1 = 2, Y_1 = 0)\}$$
$$= 0.18 + 0.0675 + 0.06 = 0.3075$$

$$p_{U_1}(3) = Pr(U_1 = 3) = Pr\{(X_1 = 1, Y_1 = 2) \text{ or } (X_1 = 2, Y_1 = 1)\} = 0.2025 + 0.0225$$
$$= 0.225$$

$$p_{U_1}(4) = Pr(U_1 = 4) = Pr(X_1 = 2, Y_1 = 2) = 0.0675$$

This is different from the *pmf* of U, despite the equality of the marginal distributions of the components of U_1 and U. Thus, the joint distributions is important in determining the distribution of the sum (or any multi-variable function) of random variables.

(e) We calculate: $E(X) = \sum_{i=0}^{2} ip_X(i) = (0 \times 0.4) + (1 \times 0.45) + (2 \times 0.15) = 0.75$, and

$$\begin{aligned}
E(Y) &= \sum_{i=0}^{2} ip_Y(i) = (0 \times 0.4) + (1 \times 0.15) + (2 \times 0.45) = 1.05 \\
E(Y^2) &= \sum_{i=0}^{2} i^2 p_Y(i) = (0^2 \times 0.4) + (1^2 \times 0.15) + (2^2 \times 0.45) = 1.95 \\
Var(Y) &= E(Y^2) - \{E(Y)\}^2 = 1.95 - (1.05)^2 = 0.8475
\end{aligned}$$

(f) We note that $E(X|Y = y) = \sum_{i=0}^{2} ip_{X|Y}(i|y)$ where $p_{X|Y}(i|y)$ is the conditional *pmf* of X given $Y = y$ which we can calculate for all possible pairs (x, y) as:

$$\begin{aligned}
p_{X|Y}(0|0) &= Pr(X = 0|Y = 0) = \frac{Pr(X = 0, Y = 0)}{Pr(Y = 0)} = \frac{0.1}{0.4} = 0.25 \\
p_{X|Y}(1|0) &= Pr(X = 1|Y = 0) = \frac{Pr(X = 1, Y = 0)}{Pr(Y = 0)} = \frac{0.25}{0.4} = 0.625 \\
p_{X|Y}(2|0) &= Pr(X = 2|Y = 0) = \frac{Pr(X = 2, Y = 0)}{Pr(Y = 0)} = \frac{0.05}{0.4} = 0.125 \\
p_{X|Y}(0|1) &= Pr(X = 0|Y = 1) = \frac{Pr(X = 0, Y = 1)}{Pr(Y = 1)} = \frac{0.1}{0.15} = 0.667 \\
p_{X|Y}(1|1) &= Pr(X = \overset{1}{\cancel{0}}|Y = 1) = \frac{Pr(X = 1, Y = 1)}{Pr(Y = 1)} = \frac{0}{0.15} = 0 \\
p_{X|Y}(2|1) &= Pr(X = \overset{2}{\cancel{0}}|Y = 1) = \frac{Pr(X = 2, Y = 1)}{Pr(Y = 1)} = \frac{0.05}{0.15} = 0.333 \\
p_{X|Y}(0|2) &= Pr(X = 0|Y = 2) = \frac{Pr(X = 0, Y = 2)}{Pr(Y = 2)} = \frac{0.2}{0.45} = 0.444 \\
p_{X|Y}(1|2) &= Pr(X = \overset{1}{\cancel{0}}|Y = 2) = \frac{Pr(X = 1, Y = 2)}{Pr(Y = 2)} = \frac{0.2}{0.45} = 0.444 \\
p_{X|Y}(2|2) &= Pr(X = \overset{2}{\cancel{0}}|Y = 2) = \frac{Pr(X = 2, Y = 2)}{Pr(Y = 2)} = \frac{0.05}{0.45} = 0.111
\end{aligned}$$

Thus, we can calculate

$$\begin{aligned}
E(X|Y = 0) &= (0 \times 0.25) + (1 \times 0.625) + (2 \times 0.125) = 0.875 \\
E(X|Y = 1) &= (0 \times 0.667) + (1 \times 0) + (2 \times 0.333) = 0.667 \\
E(X|Y = 2) &= (0 \times 0.444) + (1 \times 0.444) + (2 \times 0.111) = 0.667
\end{aligned}$$

Finally, then, we see that

$$E\{E(X|Y)\} = \sum_{i=0}^{2} E(X|Y = i)p_Y(i) = (0.875 \times 0.4) + (0.667 \times 0.15) + (0.667 \times 0.45) = 0.75,$$

which is the same as $E(X)$ which we calculated in part *(e)*.

2. We calculate the *mgf* as follows:

$$
\begin{aligned}
m_X(t) = E(e^{tX}) &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2\sigma^2}(x-\mu)^2 \right\} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2\sigma^2}(x^2 - 2\mu x + \mu^2 - 2\sigma^2 tx) \right\} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}\{x^2 - 2(\mu+\sigma^2 t)x + (\mu+\sigma^2 t)^2 - 2\mu\sigma^2 t - \sigma^4 t^2\} \right] dx \\
&= \exp\left(\mu t + \tfrac{1}{2}\sigma^2 t^2\right) \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}\{x - (\mu+\sigma^2 t)\}^2 \right] dx \\
&= \exp\left(\mu t + \tfrac{1}{2}\sigma^2 t^2\right),
\end{aligned}
$$

where the last equality follows by applying the fact provided in the hint.

3.(a) First, we note that $Y = X^2$ implies that $X = \sqrt{Y}$ and $dX = 0.5Y^{-1/2}dY$. So, using the change of variable formula, we have (for $y > 0$):

$$
\begin{aligned}
f_Y(y) &= f_X\{x(y)\}\left|\frac{dx(y)}{dy}\right| \\
&= \sqrt{\frac{2}{\pi}} \exp\left\{ -\tfrac{1}{2}(\sqrt{y})^2 \right\} 0.5y^{-1/2}dy \\
&= \sqrt{\frac{1}{2\pi y}} \exp\left(-\tfrac{1}{2}y \right)
\end{aligned}
$$

(b) On inspection, and recalling that $\Gamma(1/2) = \sqrt{\pi}$, we recognise this as the density of a chi-squared distribution with 1 degree of freedom. Note that since the square of a standard normal random variable has a chi-squared distribution with 1 degree of freedom, $Z^2 = |Z|^2$, we see that $X = |Z|$. Thus, the name of the distribution of X arises from the fact that it is the absolute value of a standard normal random variable or, more colorfully, X has a normal distribution which is "folded" over at the origin.

4. If we define

$$A = \{(u_1, u_2) : \ g_1(u_1, u_2) \leq y_1, \ g_2(u_1, u_2) \leq y_2\}$$

then we see that

$$
\begin{aligned}
A_h &= [(v_1, v_2) : \ \{h_1(v_1, v_2), h_2(v_1, v_2)\} \in A] \\
&= [(v_1, v_2) : \ g_1\{h_1(v_1, v_2), h_2(v_1, v_2)\} \leq y_1, \ g_2\{h_1(v_1, v_2), h_2(v_1, v_2)\} \leq y_2] \\
&= \{(v_1, v_2) : \ v_1 \leq y_1, \ v_2 \leq y_2\}.
\end{aligned}
$$

Therefore, we know that

$$
\begin{aligned}
\iint\limits_{A} f_{X_1X_2}(u_1, u_2)du_1du_2 &= \iint\limits_{A_h} f_{X_1X_2}\{h_1(v_1, v_2), h_2(v_1, v_2)\}|J(v_1, v_2)|dv_1dv_2 \\
&= \int_{-\infty}^{y_2}\int_{-\infty}^{y_1} f_{X_1X_2}\{h_1(v_1, v_2), h_2(v_1, v_2)\}|J(v_1, v_2)|dv_1dv_2.
\end{aligned}
$$

In addition, we see that the event $\{(Y_1, Y_2) \in A_h\} = \{Y_1 \le y_1, Y_2 \le y_2\}$ is equivalent to the event $[\{h_1(Y_1, Y_2), h_2(Y_1, Y_2)\} \in A] = \{(X_1, X_2) \in A\}$ by the definition of the set A_h. Thus, we have:

$$
\begin{aligned}
\iint_A f_{X_1X_2}(u_1, u_2)du_1du_2 &= Pr\{(X_1, X_2) \in A\} = Pr\{(Y_1, Y_2) \in A_h\} \\
&= Pr(Y_1 \le y_1, Y_2 \le y_2) \\
&= F_{Y_1Y_2}(y_1, y_2) \\
&= \int_{-\infty}^{y_2} \int_{-\infty}^{y_1} f_{Y_1Y_2}(v_1, v_2)dv_1dv_2.
\end{aligned}
$$

$$J(y_1, y_2) = \left\{ \begin{matrix} \frac{\partial}{\partial y_1} h_1(y_1, y_2) & \frac{\partial}{\partial y_1} h_2(y_1, y_2) \\ \frac{\partial}{\partial y_2} h_1(y_1, y_2) & \frac{\partial}{\partial y_2} h_2(y_1, y_2) \end{matrix} \right)$$

So, combining these facts shows that

$$f_{Y_1Y_2}(v_1, v_2) = f_{X_1X_2}\{h_1(v_1, v_2), h_2(v_1, v_2)\}|J(v_1, v_2)|,$$

which is the desired result.

5.(a) We note that the defining equations for R and Θ show that:

$$
\begin{aligned}
\tan\Theta = \frac{Y}{X} \quad &\Longrightarrow \quad \frac{\sin\Theta}{\cos\Theta} = \frac{Y}{X} \\
&\Longrightarrow \quad X\sin\Theta = Y\cos\Theta \\
&\Longrightarrow \quad X^2\sin^2\Theta = Y^2\cos^2\Theta \\
&\Longrightarrow \quad X^2(1-\cos^2\Theta) = Y^2\cos^2\Theta \\
&\Longrightarrow \quad X^2 = (X^2+Y^2)\cos^2\Theta \\
&\Longrightarrow \quad X^2 = R^2\cos^2\Theta \\
&\Longrightarrow \quad X = R\cos\Theta.
\end{aligned}
$$

Therefore, we have $Y = X\tan\Theta = R\cos\Theta\tan\Theta = R\sin\Theta$.

(b) We have $X = h_1(R, \Theta) = R\cos\Theta$ and $Y = h_2(R, \Theta) = R\sin\Theta$. So, the Jacobian matrix of the transformation has determinant:

$$|J(r,\theta)| = \begin{vmatrix} \frac{\partial}{\partial r}h_1(r,\theta) & \frac{\partial}{\partial \theta}h_1(r,\theta) \\ \frac{\partial}{\partial r}h_2(r,\theta) & \frac{\partial}{\partial \theta}h_2(r,\theta) \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r\cos^2\theta + r\sin^2\theta = r.$$

Also, the joint density of X and Y can be easily seen to be:

$$
\begin{aligned}
f_{XY}(x,y) = f_X(x)f_Y(y) &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\} \\
&= \frac{1}{2\pi\sigma^2}\exp\left[-\frac{1}{2\sigma^2}\{(x-\mu)^2+(y-\mu)^2\}\right],
\end{aligned}
$$

since we have assumed they are independent and identically normally distributed with mean μ and variance σ^2. Therefore, the joint density function of R and Θ is given by:

$$
\begin{aligned}
f_{R\Theta}(r,\theta) &= |J(r,\theta)|f_{XY}(r\cos\theta, r\sin\theta) \\
&= \frac{r}{2\pi\sigma^2}\exp\left[-\frac{1}{2\sigma^2}\{(r\cos\theta-\mu)^2+(r\sin\theta-\mu)^2\}\right] \\
&= \frac{r}{2\pi\sigma^2}\exp\left\{-\frac{1}{2\sigma^2}(r^2\cos^2\theta - 2r\mu\cos\theta + \mu^2 + r^2\sin^2\theta - 2r\mu\sin\theta + \mu^2)\right\} \\
&= \frac{r}{2\pi\sigma^2}\exp\left[-\frac{1}{2\sigma^2}\{r^2 - 2r\mu(\cos\theta+\sin\theta) + 2\mu^2\}\right],
\end{aligned}
$$

where we have again used the fact that $\sin^2\theta + \cos^2\theta = 1$ and the range of definition of the density is $0 < r < \infty$, $-\pi < \theta < \pi$.

(c) When $\mu = 0$, we see that the joint density is given by:

$$f_{R\Theta}(r,\theta) = \frac{r}{2\pi\sigma^2}\exp\left(-\frac{1}{2\sigma^2}r^2\right),$$

which does not depend on the value θ. Furthermore, it is not difficult to see that:

$$\int_0^\infty \frac{r}{\sigma^2}\exp\left(-\frac{1}{2\sigma^2}r^2\right)dr = -\exp\left(-\frac{1}{2\sigma^2}r^2\right)\Big|_{r=0}^{\infty} = 1,$$

implying that $\frac{r}{\sigma^2}\exp\left(-\frac{1}{2\sigma^2}r^2\right)$ is a density function (indeed, it is the density of the so-called Raleigh distribution). Therefore, we can see the joint density of R and Θ factors as:

$$f_{R\Theta}(r,\theta) = \frac{r}{2\pi\sigma^2}\exp\left(-\frac{1}{2\sigma^2}r^2\right) = \left\{\frac{r}{\sigma^2}\exp\left(-\frac{1}{2\sigma^2}r^2\right)\right\}\left(\frac{1}{2\pi}\right) = f_R(r)f_\Theta(\theta).$$

Clearly, then, when $\mu = 0$, R and Θ are independent. Moreover, the marginal density function for Θ is given by $f_\Theta(\theta) = 2\pi^{-1}$ for $\theta \in (-\pi,\pi)$. In other words, Θ is uniformly distributed on the interval $(-\pi,\pi)$.

(d) When $\mu \neq 0$, we have seen that the general expression for the joint density of R and Θ is given by:

$$\begin{aligned} f_{R\Theta}(r,\theta) &= \frac{r}{2\pi\sigma^2}\exp\left[-\frac{1}{2\sigma^2}\{r^2 - 2r\mu(\cos\theta + \sin\theta) + 2\mu^2\}\right] \\ &= \frac{r}{2\pi\sigma^2}\exp\left[-\frac{1}{2\sigma^2}\{r^2 + 2\mu^2\}\right]\exp\left\{\frac{r\mu}{\sigma^2}(\cos\theta + \sin\theta)\right\} \\ &= C(r)\exp\{a(r)(\cos\theta + \sin\theta)\}. \end{aligned}$$

Now, if $\mu > 0$, then for any given value of r we have $C(r), a(r) > 0$, which means that the function $C(r)\exp\{a(r)(\cos\theta + \sin\theta)\}$ is maximised whenever $(\cos\theta + \sin\theta)$ is. Differentiating this function and setting to zero, yields:

$$-\sin\theta + \cos\theta = 0 \quad\Longrightarrow\quad \sin\theta = \cos\theta \quad\Longrightarrow\quad \tan\theta = 1 \quad\Longrightarrow\quad \theta = \frac{\pi}{4}, -\frac{3\pi}{4}.$$

(Alternatively, we could have simply differentiated the joint density function directly, but this would have been a rather messy way to arrive at the same answer). Taking second derivatives yields $-\cos\theta - \sin\theta$ which is positive at $\theta = -3\pi/4$, implying this value is a minimum. Thus, when $\mu > 0$, the maximum occurs at $\theta = \pi/4$. Alternatively, when $\mu < 0$, we have $a(r) < 0$, which means the density is maximised when $(\cos\theta + \sin\theta)$ is minimised. So, the preceding reasoning shows that when $\mu < 0$, the density is maximised when $\theta = -3\pi/4$.

(d) If we think of X and Y as representing random $x-$ and $y-$co-ordinates in the Euclidean plane, then Θ is the random angle associated with the point (X,Y), when expressed in polar co-ordinates. The fact that Θ is uniform when $\mu = 0$ means that under this condition the point (X,Y) is equally likely to lie in any direction from the origin. Alternatively, if

$\mu \neq 0$, we see that the point (X, Y) is more likely to lie along the 45° line. This approach can be generalised to a k-vector of *iid* normals and then leads to a geometric interpretation of the t-test of $H_0 : \mu = 0$ versus $H_A : \mu \neq 0$. However, a full development of these ideas is beyond the scope of this course.

6. We can calculate the *CDF* of Y as:

$$Pr(Y \leq y) = Pr\{F(X) \leq y\} = Pr\{X \leq F^{-1}(y)\} = F\{F^{-1}(y)\} = y,$$

for $0 \leq y \leq 1$ (otherwise $F^{-1}(y)$ is undefined). Of course, this is precisely the *CDF* of a uniform distribution on (0,1). (Alternatively, if we don't recognise the *CDF*. we could differentiate to find the density function $f(y) = 1$ for $0 \leq y \leq 1$, which is the *pdf* of a uniform distribution on the unit interval.) Now, if we can generate a random uniform value, U, then the above discussion shows that $X = F^{-1}(U)$ will be a random value having a distribution with *CDF* $F(x)$ (provided, of course, that the desired it CDF is invertible, which can be guaranteed if the distribution in question is continuous). Therefore, since the exponential distribution is continuous and has $F(x) = 1 - e^{-x/\mu}$, we see that:

$$F(x) = 1 - e^{-x/\mu} \qquad \Longrightarrow \qquad x = -\mu \log\{1 - F(x)\} \qquad \Longrightarrow \qquad F^{-1}(x) = -\mu \log\{1 - x\}$$

which means that the value $X = -\mu \log(1-U)$ will be a random exponentially distributed value with mean μ. This approach to generating random values from uniform values is sometimes referred to as *the probability inverse transform* method.