NATIONAL UNIVERSITY OF SINGAPORE
MATHEMATICS SOCIETY

PAST YEAR PAPER SOLUTIONS
with credits to Luo Xuan, Ebenezer Lee

MA2216/ST2131 Probability
AY 2010/2011 Sem 1

Question 1

(a) Since $f(x,y)$ is a joint p.d.f.,

$$\iint_{\mathbb{R}^2} f(x,y)\,\mathrm{d}x\mathrm{d}y = 1.$$

i.e.,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Ky^{-1}e^{-y}e^{-\frac{(x-y)^2}{y}}\,\mathrm{d}x\mathrm{d}y = 1$$

The LHS is equal to

$$\begin{aligned}
&K\int_0^{\infty} y^{-1}e^{-y}\int_{-\infty}^{\infty} e^{-\frac{(x-y)^2}{y}}\,\mathrm{d}x\mathrm{d}y\\
&= K\int_0^{\infty} y^{-1}e^{-y}\sqrt{\frac{y}{2}}\sqrt{2\pi}\int_{-\infty}^{\infty}\frac{1}{\sqrt{\frac{y}{2}}\sqrt{2\pi}}\,e^{-\frac{1}{2}\frac{(x-y)^2}{\left(\sqrt{y/2}\right)^2}}\,\mathrm{d}x\mathrm{d}y\\
&= K\sqrt{\pi}\int_0^{\infty} y^{\frac{1}{2}-1}e^{-1y}(1)\,\mathrm{d}y\\
&= K\sqrt{\pi}\,\frac{\Gamma\left(\frac{1}{2}\right)}{(1)^{\frac{1}{2}}}\\
&= K\sqrt{\pi}\sqrt{\pi} = K\pi \qquad \text{which must be equal to the RHS, which is 1.}
\end{aligned}$$

Hence, $K = \frac{1}{\pi}$. $\square$

(b)

$$\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x\\
&= \int_{-\infty}^{\infty} Ky^{-1}e^{-y}e^{-\frac{(x-y)^2}{y}}\,\mathrm{d}x\\
&= Ky^{-1}e^{-y}\int_{-\infty}^{\infty} e^{-\frac{(x-y)^2}{y}}\,\mathrm{d}x\\
&= Ky^{-1}e^{-y}\sqrt{\frac{y}{2}}\sqrt{2\pi} \qquad \text{from part (i)}\\
&= \frac{1}{\sqrt{\pi}}y^{-\frac{1}{2}}e^{-y} \qquad \text{for } y > 0. \quad \square
\end{aligned}$$

Notice that $f_Y(y) = \frac{(1)^{(1/2)}}{\Gamma(\frac{1}{2})}y^{(1/2)-1}e^{-(1)y},\ y > 0$. Thus, $Y \sim \Gamma(\frac{1}{2}, 1)$, from which we also have

$$\mathbb{E}(Y) = \frac{1/2}{1} = \frac{1}{2} \qquad \text{and} \qquad \mathrm{Var}(Y) = \frac{1/2}{1^2} = \frac{1}{2}.$$

(c)

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f(x,y)}{f_Y(y)} \\ &= \frac{\frac{1}{\pi} y^{-1} e^{-y} e^{-\frac{(x-y)^2}{y}}}{\frac{1}{\sqrt{\pi y}} e^{-y}} \\ &= \frac{1}{\sqrt{\pi y}} e^{-\frac{(x-y)^2}{y}} \qquad \text{for } -\infty < x < \infty, \text{ given } Y = y > 0. \end{aligned}$$

Notice that

$$f_{X|Y}(x|y) = \frac{1}{\sqrt{y/2}\sqrt{2\pi}} e^{-\frac{(x-y)^2}{2(\sqrt{y/2})^2}}.$$

Hence, the conditional distribution of X, given $Y = y > 0$, is normal with mean y and variance $\frac{y}{2}$. $\square$

(d) It follows from the conditional distribution in part (iii) that $\mathbb{E}(X|Y=y) = y$. Then, using $\mathbb{E}(X) = \mathbb{E}[\mathbb{E}(X|Y)]$, we have $\mathbb{E}(X) = \mathbb{E}(Y) = \frac{1}{2}$. $\square$

(e) Using the formula $\text{Var}(T) = \mathbb{E}\left(T^2\right) - [\mathbb{E}(T)]^2$, we have

$$\mathbb{E}\left(X^2|Y=y\right) = \text{Var}(X|Y) + [\mathbb{E}(X|Y)]^2 = \frac{y}{2} + y^2$$

With this, we find

$$\begin{aligned} \mathbb{E}\left(X^2\right) &= \mathbb{E}\left[\mathbb{E}\left(X^2|Y\right)\right] \\ &= \mathbb{E}\left(\frac{Y}{2} + Y^2\right) \\ &= \frac{1}{2}\mathbb{E}(Y) + \mathbb{E}\left(Y^2\right) \qquad \text{by linearity} \\ &= \frac{1}{2}\left(\frac{1}{2}\right) + \text{Var}(Y) + [\mathbb{E}(Y)]^2 \\ &= \frac{1}{4} + \frac{1}{2} + \left(\frac{1}{2}\right)^2 \\ &= 1 \end{aligned}$$

Finally,

$$\text{Var}(X) = \mathbb{E}\left(X^2\right) - [\mathbb{E}(X)]^2 = 1 - \left(\frac{1}{2}\right)^2 = \frac{3}{4}. \quad \square$$

Question 2

(a) (Step 1.) First, the transformation is given by

$$\begin{cases} x = \frac{1}{2}(u-v) & (1) \\ y = \frac{1}{2}(u+v) & (2) \end{cases}$$

(Step 2.) The inverse transformation can be obtained by separately taking $(1)+(2)$ and $(1)-(2)$. These give

$$\begin{cases} u = y + x & (3) \\ v = y - x & (4) \end{cases}$$

(Step 3.) We find the domain of x and y. The domain of u and v is $\{(u, v) : u > 0, v > 0\}$. These, with (1) and (2) respectively, give

$$\begin{cases} -\infty < x < \infty & \text{and } y > 0 \\ \quad y > -x & \text{and } y > x \end{cases}$$

(Step 4.) The Jacobian

$$J(u, v) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{vmatrix} = \frac{1}{2}$$

The necessary condition of the Jacobian being nonzero for all points in the domain is satisfied.

(Step 5.) Finally, the joint p.d.f. of X and Y is

$$f_{(X,Y)}(x, y) = \frac{1}{|J(u, v)|} f_{(U,V)}(u, v)$$

and, noting that $\Gamma(2) = (2 - 1)! = 1$,

$$\begin{aligned} f_{(U,V)}(u, v) &= f_U(u) f_V(v) && \text{(since } U \text{ and } V \text{ are independent)} \\ &= \frac{(1/2)^2}{\Gamma(2)} u^{2-1} e^{-\frac{1}{2}u} \cdot \frac{(1/2)^2}{\Gamma(2)} v^{2-1} e^{-\frac{1}{2}v} \\ &= \frac{1}{16} uve^{-\frac{1}{2}(u+v)} \end{aligned}$$

Therefore,

$$\begin{aligned} f_{(X,Y)}(x, y) &= \frac{1}{|J(u, v)|} f_{(U,V)}(u, v) \\ &= \frac{1}{|1/2|} \frac{1}{16} uve^{-\frac{1}{2}(u+v)} \\ &= \frac{1}{8}(y + x)(y - x)e^{-y} && \text{from (2), (3) and (4)} \\ &= \frac{1}{8}\left(y^2 - x^2\right) e^{-y} && \text{for } -\infty < x < \infty,\ y > |x|. \quad \square \end{aligned}$$

(b)

$$\begin{aligned} \text{Cov}(X, Y) &= \text{Cov}\left(\frac{1}{2}U - \frac{1}{2}V\ ,\ \frac{1}{2}U + \frac{1}{2}V\right) \\ &= \frac{1}{2} \cdot \frac{1}{2}\ \text{Cov}(U - V\ ,\ U + V) && \text{(by bilinearity)} \\ &= \frac{1}{4}\left[\text{Cov}(U, U) + \text{Cov}(U, V) - \text{Cov}(V, U) - \text{Cov}(V, V)\right] && \text{(by bilinearity)} \\ &= \frac{1}{4}\left[\text{Var}(U) + 0 - 0 - \text{Var}(V)\right] && \text{(since } U \text{ and } V \text{ are independent)} \\ &= 0 \end{aligned}$$

($\text{Var}(U) = \text{Var}(V)$ since U and V are identically distributed.) $\square$

(c) No. The domain of x and y is $\{(x, y) : -\infty < x < \infty,\ y > |x|\}$. This shows that y depends on x. $\square$

Question 3

(a) We know that for an indicator random variable I_A of an event A, $\mathbb{E}(I_A) = \mathbb{P}(A)$.
Hence, for each i,

$$\mathbb{E}(X_i) = \mathbb{P}\{\text{ball } r_i \text{ is withdrawn}\} = \frac{1 \cdot \binom{29}{11}}{\binom{30}{12}} = 0.4 \quad \square$$

Similarly, for each j,

$$\mathbb{E}(Y_j) = \mathbb{P}\{j\text{th ball is withdrawn}\} = \frac{1 \cdot \binom{29}{11}}{\binom{30}{12}} = 0.4 \quad \square$$

Since X_i and Y_j are independent,

$$\mathbb{E}(X_i Y_j) = \mathbb{E}(X_i)\,\mathbb{E}(Y_j) = (0.4)(0.4) = 0.16. \quad \square$$

$$\begin{aligned}
\operatorname{Cov}(X, Y) &= \operatorname{Cov}\left(\sum_{i=1}^{10} X_i\,, \sum_{j=1}^{8} Y_j\right) \\
&= \sum_{i=1}^{10}\sum_{j=1}^{8} \operatorname{Cov}(X_i, Y_j) \qquad \text{(by bilinearity)} \\
&= \sum_{i=1}^{10}\sum_{j=1}^{8} 0 \qquad\qquad\qquad \text{(since } X_i \text{ and } Y_j \text{ are independent)} \\
&= 0. \quad \square
\end{aligned}$$

(b) Notice that X and Y are indicator random variables.
Let A and B be, respectively, the events that X and Y represent, i.e.,

$$X = \begin{cases} 1, & \text{if } A \text{ occurs} \\ 0, & \text{if } A \text{ does not occur} \end{cases} \qquad Y = \begin{cases} 1, & \text{if } B \text{ occurs} \\ 0, & \text{if } B \text{ does not occur} \end{cases}$$

To show that "X and Y are independent $\iff \operatorname{Cov}(X, Y) = 0$" :
The direction "$\Longrightarrow$" is true for any random variables X and Y. For the direction "$\Longleftarrow$",

$$\begin{aligned}
& \operatorname{Cov}(X, Y) = 0 \\
\Rightarrow\quad & \mathbb{E}(XY) - \mathbb{E}(X)\,\mathbb{E}(Y) = 0 \\
\Rightarrow\quad & \mathbb{P}(AB) - \mathbb{P}(A)\,\mathbb{P}(B) = 0 \\
\Rightarrow\quad & \mathbb{P}(AB) = \mathbb{P}(A)\,\mathbb{P}(B)
\end{aligned}$$

i.e., A and B are independent. Hence, X and Y are independent. $\square$

Question 4

(a) We have independent $X_1,\ X_2, \cdots \sim$ U(0,1) and $N \sim$ Geom(p). Let $q = 1 - p$.
Also, notice that M takes values between 0 and 1. Thus, for $0 \le x < 1$,

$$\begin{aligned}
\mathbb{P}(M \le x) &= \sum_{n=1}^{\infty} \mathbb{P}\{M \le x | N = n\} \cdot \mathbb{P}\{N = n\} \\
&= \sum_{n=1}^{\infty} \mathbb{P}\{\max(X_1, X_2, \cdots, X_n) \le x\} \cdot q^{n-1}p \\
&= \sum_{n=1}^{\infty} \mathbb{P}\{X_1 \le x, X_2 \le x, \cdots, X_n \le x\} \cdot q^{n-1}p \\
&= \sum_{n=1}^{\infty} \mathbb{P}\{X_1 \le x\}\, \mathbb{P}\{X_2 \le x\} \cdots \mathbb{P}\{X_n \le x\} \cdot q^{n-1}p \qquad (\text{since } X_1, X_2, \cdots \text{ are independent}) \\
&= \sum_{n=1}^{\infty} \mathbb{P}\{X_1 \le x\}^n \cdot q^{n-1}p \qquad (\text{since } X_1, X_2, \cdots \text{ are identically distributed}) \\
&= \sum_{n=1}^{\infty} \left(\frac{x-0}{1-0}\right)^n \cdot q^{n-1}p \qquad \text{for } 0 \le x < 1 \\
&= px \sum_{n=1}^{\infty} (qx)^{n-1} \\
&= px \cdot \frac{1}{1-qx} \qquad (|qx| < 1 \text{ since } q, x \in (0,1)) \\
&= \frac{px}{1-(1-p)x} \qquad \text{for } 0 \le x < 1
\end{aligned}$$

We can thus deduce that

$$\mathbb{P}(M \le x) = \begin{cases} 0, & \text{for } x < 0 \\ \frac{px}{1-(1-p)x}, & \text{for } 0 \le x < 1 \\ 1, & \text{for } x \ge 1 \end{cases} \qquad \square$$

(b) We first calculate

$$\begin{aligned}
\mathbb{E}(S_n) &= \mathbb{E}(X_1 + X_2 + \cdots + X_n) \\
&= \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n) \qquad (\text{by linearity}) \\
&= n\,\mathbb{E}(X_1) \qquad (\text{since } X_1, X_2, \cdots, X_n \text{ are identically distributed}) \\
&= n\,\mathbb{P}(X_1 = 1) \qquad (\text{since } X_1 \text{ is an indicator random variable}) \\
&= np
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Var}(S_n) &= \mathrm{Var}(X_1 + X_2 + \cdots + X_n) \\
&= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n) \\
&\qquad (\text{since } X_1, X_2, \cdots, X_n \text{ are independent, the covariances vanish}) \\
&= n\,\mathrm{Var}(X_1) \qquad (\text{since } X_1, X_2, \cdots, X_n \text{ are identically distributed}) \\
&= np(1-p)
\end{aligned}$$

Thus, for large n, by CLT, $S_n \sim \mathrm{N}(np, np(1-p))$ approx. Equivalently,

$$\overline{S_n} = \frac{S_n}{n} \sim \mathrm{N}\left(p, \frac{p(1-p)}{n}\right) \text{ approx.}$$

(i.)

$$\begin{aligned}
\mathbb{P}\left\{\left|\overline{S_n} - p\right| \geq c\right\} &= \mathbb{P}\left\{\left|\frac{\overline{S_n} - p}{\sqrt{\frac{p(1-p)}{n}}}\right| \geq \frac{c}{\sqrt{\frac{p(1-p)}{n}}}\right\} \\
&\approx \mathbb{P}\left\{|Z| \geq \frac{c}{\frac{1/2}{\sqrt{900}}}\right\} \qquad \left(\because \sqrt{p(1-p)} \approx 1/2\right) \\
&= \mathbb{P}\left\{|Z| \geq 60c\right\} \\
&= 2\,\mathbb{P}\{Z \geq 60c\} \qquad \text{(by symmetry)}
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathbb{P}\{Z \geq 60c\} &\approx \frac{1}{2}\,\mathbb{P}\left\{\left|\overline{S_n} - p\right| \geq c\right\} = \frac{1}{2}(0.01) = 0.005 \\
60c &\approx z_{0.005} = 2.58 \\
c &\approx 0.043 \quad \square
\end{aligned}$$

(ii)

$$\begin{aligned}
\mathbb{P}\left\{\left|\overline{S_n - p}\right| \geq 0.025\right\} &= \mathbb{P}\left\{\left|\frac{\overline{S_n} - p}{\sqrt{\frac{p(1-p)}{n}}}\right| \geq \frac{0.025}{\sqrt{\frac{p(1-p)}{n}}}\right\} \\
&\approx \mathbb{P}\left\{|Z| \geq \frac{0.025}{\frac{1/2}{\sqrt{n}}}\right\} \\
&= \mathbb{P}\left\{|Z| \geq 0.05\sqrt{n}\right\} \\
&= 2\,\mathbb{P}\left\{Z \geq 0.05\sqrt{n}\right\} \qquad \text{(by symmetry)}
\end{aligned}$$

Thus,

$$\begin{aligned}
2\,\mathbb{P}\{Z \geq 0.05\sqrt{n}\} &\approx \mathbb{P}\left\{\left|\overline{S_n - p}\right| \geq 0.025\right\} = 0.01 \\
\mathbb{P}\{Z \geq 0.05\sqrt{n}\} &\approx 0.005 \\
0.05\sqrt{n} &\approx z_{0.005} = 2.58 \\
n &\approx (51.6)^2 \approx 2663 \quad \square
\end{aligned}$$