Solution to PS #3

Q1.
Let $Y$ and $X$ be continuous random variables.$^1$

$$E[E[Y|X]] = \int_{-\infty}^{\infty} E[Y|X] f_X(x) dx$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yf_{Y|X}(y) dy f_X(x) dx$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yf_{Y|X}(y) f_X(x) dy dx$$
$$= \int_{-\infty}^{\infty} E[Y] dx$$
$$= E[Y]$$

This is called Law of Iterated Expectation (LIE). For the second one, one way is given in Hogg et. al, p.98. Another way to prove is as follows.

$$\text{Var}(Y) - \text{Var}(E[Y|X]) = \left\{ E\left[ Y^2 \right] - (E[Y])^2 \right\} - \left\{ E\left[ (E[Y|X])^2 \right] - (E[E[Y|X]])^2 \right\}$$
$$= E\left[ Y^2 \right] - E\left[ (E[Y|X])^2 \right] \quad (\because \text{LIE})$$
$$= E\left[ E[Y^2|X] - (E[Y|X])^2 \right] \quad (\because \text{LIE})$$
$$\geq 0$$

where the last inequality comes from that for any $x$,

$$E\left[ Y^2|X = x \right] - (E[Y|X = x])^2 = \text{Var}(Y|X = x) \geq 0$$

and thus the expectation of the LHS of the above equation should be nonnegative. This proof, in fact, leads to the following theorem.$^2$

$$\text{Var}(Y) = \text{Var}(E[Y|X]) + E[\text{Var}(Y|X)]$$

Q2. (a)
Note that two matrices are the same if every coordinate of one matrix coincides with that of the other. Take any $i$ and $j$. The $i,j$-th coordinate of $\text{Var}(X)$ is

$$E\left[ (X_i - \mu_i)(X_j - \mu_j) \right] = E[X_i X_j - X_i \mu_j - \mu_i X_j + \mu_i \mu_j] = E[X_i X_j] - \mu_i \mu_j$$

which is equal to the $i,j$-th coordinate of $E[XX'] - \mu \mu'$.

---

$^1$The proof for discrete random variables is similar.

$^2$The formal proof of the theorem is given in Mike Powell:
(b) Let \( a_{ij} \) be the \( i, j \)-th element of the \( m \times n \) matrix \( A \).

\[
E(AX) = E \begin{pmatrix}
a_{11}X_1 + \cdots + a_{1n}X_n \\
\vdots \\
am_{11}X_1 + \cdots + a_{mn}X_n
\end{pmatrix}
= \begin{pmatrix}
a_{11}E[X_1] + \cdots + a_{1n}E[X_n] \\
\vdots \\
am_{11}E[X_1] + \cdots + a_{mn}E[X_n]
\end{pmatrix}
= AE[X] = A\mu
\]

Therefore,\(^3\)

\[
Var(AX) = E \left[ (AX - A\mu)(AX - A\mu)' \right]
= E \left[ A(X - \mu)(X - \mu)'A' \right]
= AE \left[ (X - \mu)(X - \mu)' \right] A'
= AVar(X)A'
\]

(c) Note that a matrix is symmetric if its \( i, j \)-th element is the same with its \( j, i \)-th element for all \( i \) and \( j \). Take any \( i \) and \( j \). The \( i, j \)-th element of \( Var(X) \) is

\[
E[(X_i - \mu_i)(X_j - \mu_j)] = E[(X_j - \mu_j)(X_i - \mu_i)]
\]

which is the \( j, i \)-th element of \( Var(X) \). So \( Var(X) \) is symmetric. Let \( a \) be any nonstochastic \( n \times 1 \) vector. Then by the result of (b),

\[
a'Var(X)a = Var(a'X) \geq 0
\]

Therefore, \( Var(X) \) is positive semidefinite.

**HMC 2.1.3.**

Note first that

\[
F(b,d) = Pr(X \leq b, Y \leq d) = Pr(X \leq b, c < Y \leq d) + Pr(X \leq b, Y \leq c)
\]

since two sets in the RHS is a partition of the set in LHS. In the same way,

\[
Pr(X \leq b, c < Y \leq d) = Pr(a < X \leq b, c < Y \leq d) + Pr(X \leq a, c < Y \leq d)
= Pr(a < X \leq b, c < Y \leq d) + \left[ Pr(X \leq a, Y \leq d) - Pr(X \leq a, Y \leq c) \right]
= Pr(a < X \leq b, c < Y \leq d) + \left[ F(a, d) - F(a, c) \right]
\]

\(^3\)We can extend this result to \( E[AXB] = AE[X]B \) for any random matrix \( X \) and nonstochastic matrices \( A \) and \( B \).
Plugging back into the first equation,
\[ F(b, d) = \Pr(a < X \leq b, c < Y \leq d) + \left[ F(a, d) - F(a, c) \right] + F(b, c) \]
Rearranging yields
\[ \Pr(a < X \leq b, c < Y \leq d) = F(b, d) - F(b, c) - F(a, d) + F(a, c) \]
which completes the proof.\(^4\)

**HMC 2.1.4.**
Pick \(a = c = 0\) and \(b = d = 1\). Then,
\[ \Pr(0 < X \leq 1, 0 < Y \leq 1) = F(1, 1) - F(0, 1) - F(1, 0) + F(0, 0) = 1 - 1 - 1 + 0 = -1 < 0 \]
which is not possible for distribution function \(F\). So such \(F\) cannot be a distribution function.

**HMC 2.1.7.**
\[
\begin{align*}
\Pr(Z \leq z) &= \Pr \left( Y \leq \frac{z}{X} \right) \\
&= \Pr \left( Y \leq \frac{z}{X}, 0 \leq X \leq z \right) + \Pr \left( Y \leq \frac{z}{X}, z < X \leq 1 \right) \\
&= \Pr \left( 0 < Y \leq 1, 0 \leq X \leq z \right) + \Pr \left( 0 < Y \leq \frac{z}{X}, z < X \leq 1 \right) \\
&= \int_0^z \int_0^1 f(x, y) dydx + \int_0^1 \int_z^1 f(x, y) dydx \\
&= z - z \log z
\end{align*}
\]
for \(z \in (0, 1)\). So the pdf of \(Z\) is \(f_Z(z) = -\log z\) when \(z \in (0, 1)\) and 0 otherwise.\(^5\)

---

\(^4\)There is another way to prove which uses Lebesgue integral.
\[
\Pr(a < X \leq b, c < Y \leq d) = \int_a^b \int_c^d d^2 F(x, y) \\
= \int_{-\infty}^c \int_{-\infty}^d d^2 F(x, y) - \int_{-\infty}^c \int_{-\infty}^d d^2 F(x, y) \\
= \left[ \int_{-\infty}^b \int_{-\infty}^d d^2 F(x, y) - \int_{-\infty}^b \int_{-\infty}^c d^2 F(x, y) - \int_{-\infty}^a \int_{-\infty}^c d^2 F(x, y) \right] \\
= \left[ F(b, d) - F(b, c) \right] - \left[ F(a, d) + F(a, c) \right]
\]

\(^5\)Another way is
\[
\begin{align*}
\Pr(Z \leq z) &= 1 - \Pr(Z > z) \\
&= 1 - \Pr \left( Y > \frac{z}{X} \right) \\
&= 1 - \int_z^1 \int_0^1 f(x, y) dydx \\
&= z - z \log z
\end{align*}
\]
We can also introduce another variable \(W = X\) and use Tranformation theorem. Note that we obtain joint pdf of \(Z\) and \(W\), so we have to calculate marginal pdf of \(Z\).
HMC 2.2.4.
Note that $X_2 = Y_2$ and $X_1 = Y_1X_2 = Y_1Y_2$. Since $x_1$ and $x_2$ satisfy $0 < x_1 < x_2 < 1$, $y_1$ and $y_2$ would satisfy $0 < y_1y_2 < y_2 < 1$. The support of $Y$ is $S_Y = \{(y_1, y_2)|0 < y_1, y_2 < 1\}$. By Transformation theorem,

$$f_Y(y) = f_X(y_1y_2, y_2) \left| \det \begin{pmatrix} y_2 & y_1 \\ 0 & 1 \end{pmatrix} \right| = 8y_1y_2^3$$

when $y \in S_Y$ and 0 otherwise.

HMC 2.2.5. (a)
Note that $X_2 = Y_2$ and $X_1 = Y_1 - X_2 = Y_1 - Y_2$. The support of $Y$ is $S_Y = R^2$.

$$f_Y(y) = f_X(y_1 - y_2, y_2) \left| \det \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \right| = f_X(y_1 - y_2, y_2)$$

(b) This is direct application of (a).

$$f_Y(y_1) = \int_{-\infty}^{\infty} f_Y(y)dy_2 = \int_{-\infty}^{\infty} f_X(y_1 - y_2, y_2)dy_2$$

HMC 2.3.11. (a)
The most plausible assumption is that

$$f_{X_1}(x_1) = \begin{cases} 1 & \text{if } x_1 \in (0, 1) \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad f_{X_2|X_1=x_1}(x_2) = \begin{cases} \frac{1}{x_1} & \text{if } x_2 \in (0, x_1) \\ 0 & \text{otherwise} \end{cases}$$

(b)

$$\Pr(X_1 + X_2 \geq 1) = \Pr\left(\frac{1}{2} \leq X_1 \leq 1, X_1 + X_2 \geq 1\right)$$

$$= \Pr\left(\frac{1}{2} \leq X_1 \leq 1, 1 - X_1 \leq X_2 \leq X_1\right)$$

$$= \int_{\frac{1}{2}}^{1} \int_{1-x_1}^{x_1} f_{X_2|X_1=x_1}(x_2)dx_2f_{X_1}(x_1)dx_1$$

$$= 1 - \log 2$$

(c)

$$f_X(x) = f_{X_2|X_1=x_1}(x_2)f_{X_1}(x_1) = \frac{1}{x_1}$$

when $0 < x_2 < x_1 < 1$ and 0 otherwise. Since $f_{X_2}(x_2) = \int_{x_2}^{1} f_X(x)dx_1 = -\log x_2$,

$$f_{X_1|X_2=x_2}(x_1) = \frac{f_X(x)}{f_{X_2}(x_2)} = -\frac{1}{x_1 \log x_2}$$

when $x_1 \in (x_2, 1)$ and 0 otherwise.

$$E[X_1|X_2 = x_2] = \int_{x_2}^{1} x_1 f_{X_1|X_2=x_2}(x_1)dx_1 = \left[-\frac{x_1}{\log x_2}\right]_{x_2}^{1} = \frac{x_2 - 1}{\log x_2}$$
HMC 2.4.11.

\[ \text{Var}(X_1 + X_2) = \text{Var}(X_1) + 2\text{Cov}(X_1, X_2) + \text{Var}(X_2) = \sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2 = 2(1 + \rho)\sigma^2 \]

Therefore, by Chebyshev inequality,

\[ \Pr\left( \left| (X_1 + X_2) - (\mu_1 + \mu_2) \right| \geq b \right) \leq \frac{\text{Var}(X_1 + X_2)}{b^2} = \frac{2(1 + \rho)\sigma^2}{b^2} \]

Letting \( b = k\sigma \),

\[ \Pr\left( \left| (X_1 - \mu_1) + (X_2 - \mu_2) \right| \geq k\sigma \right) \leq \frac{2(1 + \rho)}{k^2} \]

HMC 2.5.7.

From the support, \((x_2 + 2)^2 < 1 - (x_1 - 1)^2\), or \(-2 - \sqrt{1 - (x_1 - 1)^2} < x_1 < -2 + \sqrt{1 - (x_1 - 1)^2} \).

\[
\begin{align*}
  f_{X_1}(x_1) &= \int_{-2 - \sqrt{1 - (x_1 - 1)^2}}^{-2 + \sqrt{1 - (x_1 - 1)^2}} f_X(x)dx_2 = \frac{2\sqrt{1 - (x_1 - 1)^2}}{\pi} \quad \text{when } 0 < x_1 < 2 \\
  f_{X_2}(x_1) &= \int_{1 - \sqrt{1 - (x_2 + 2)^2}}^{1 + \sqrt{1 - (x_2 + 2)^2}} f_X(x)dx_1 = \frac{2\sqrt{1 - (x_2 + 2)^2}}{\pi} \quad \text{when } -3 < x_2 < -1 
\end{align*}
\]

They are not independent since the support of \( X_1 \) varies with \( X_2 \). In fact, let \( x_1 = .2 \) and \( x_2 = -1.2 \), then \((x_1 - 1)^2 + (x_2 + 2)^2 = 1.28 > 1\), so \( x \) is not in the support, therefore

\[ f_{X_1}(x_1)f_{X_2}(x_2) = \frac{2\sqrt{0.36}}{\pi} \cdot \frac{2\sqrt{0.36}}{\pi} = \frac{1.44}{\pi^2} \neq 0 = f_X(x) \]

HMC 2.6.6.

\[ E[X_1 - \mu_1 | X_2, X_3] = b_2(X_2 - \mu_2) + b_3(X_3 - \mu_3) \quad (1) \]

Multiply \((X_2 - \mu_2)\) on both sides of (1) and take expectation, then, \( \rho_{12}\sigma_1\sigma_2 = b_2\sigma_2 + b_3\rho_{23}\sigma_2\sigma_3 \), or

\[ \rho_{12}\sigma_1 = b_2\sigma_2 + b_3\rho_{23}\sigma_3 \quad (2) \]

Multiply \((X_3 - \mu_3)\) on both sides of (1) and take expectation, then, \( \rho_{13}\sigma_1\sigma_3 = b_2\rho_{23}\sigma_2\sigma_3 + b_3\sigma_3^2 \), or

\[ \rho_{13}\sigma_1 = b_2\rho_{23}\sigma_2 + b_3\sigma_3 \quad (3) \]

Solving the system equation (2) and (3),

\[
\begin{align*}
  b_2 &= \frac{\sigma_1(\rho_{12} - \rho_{13}\rho_{23})}{\sigma_2(1 - \rho_{23}^2)} \\
  b_3 &= \frac{\sigma_1(\rho_{13} - \rho_{12}\rho_{23})}{\sigma_3(1 - \rho_{23}^2)}
\end{align*}
\]