Solutions to Homework Problems

1. We will show that
$$(\cup_{n=1}^{\infty} A_n)^c = \cap_{n=1}^{\infty}(A_n^c) \tag{1}$$
by proving that
$$(\cup_{n=1}^{\infty} A_n)^c \subset \cap_{n=1}^{\infty}(A_n^c) \tag{2}$$
and
$$\cap_{n=1}^{\infty}(A_n^c) \subset (\cup_{n=1}^{\infty} A_n)^c. \tag{3}$$
We will first prove (2). Note that
$$\begin{aligned} & x \in (\cup_{n=1}^{\infty} A_n)^c \\ & \Rightarrow x \notin \cup_{n=1}^{\infty} A_n \\ & \Rightarrow x \in A_n^c \text{ for each } n \in \{1, 2, \ldots\} \\ & \Rightarrow x \in \cap_{n=1}^{\infty}(A_n^c), \end{aligned}$$
so (2) holds.

   To prove (3), note that
$$\begin{aligned} & x \in \cap_{n=1}^{\infty}(A_n^c) \\ & \Rightarrow x \in A_n^c \text{ for each } n \in \{1, 2, \ldots\} \\ & \Rightarrow x \notin \cup_{n=1}^{\infty} A_n \\ & \Rightarrow x \in (\cup_{n=1}^{\infty} A_n)^c, \end{aligned}$$
so (3) holds.

   Since both (2) and (3) hold, we have (1).

2. Since $\sigma(\mathcal{C})$ is a σ-field, $\sigma(\mathcal{C})$ is closed under taking completement/countable union/countable intersection. Therefore, we have the following results.

   - The set $\{3\} = A \cap B$ is in $\sigma(\mathcal{C})$ since A and B are in $\sigma(\mathcal{C})$.
   - The set $\{3\}^c$ is in $\sigma(\mathcal{C})$ since $\{3\}$ is in $\sigma(\mathcal{C})$.
   - The set $\{1, 2\} = A \cap \{3\}^c$ is in $\sigma(\mathcal{C})$ since both A and $\{3\}^c$ are in $\sigma(\mathcal{C})$.
   - The set $\{4, 5\} = B \cap \{3\}^c$ is in $\sigma(\mathcal{C})$ since both A and $\{3\}^c$ are in $\sigma(\mathcal{C})$.

   Let $C_1 = \{3\}$, $C_2 = \{1, 2\}$ and $C_3 = \{4, 5\}$, then $\emptyset$, C_1, C_2, C_3 are in $\sigma(\mathcal{C})$. $\sigma(\mathcal{C})$ should also include sets of the form: $D_1 \cup D_2 \cup D_3$, where D_i is $\emptyset$ or C_i for i = 1,2,3. Therefore, $\sigma(\mathcal{C})$ should include the follwing sets:

   - $\emptyset \cup \emptyset \cup \emptyset = \emptyset$,
   - $C_1 \cup \emptyset \cup \emptyset = \{3\}$,
   - $\emptyset \cup C_2 \cup \emptyset = \{1, 2\}$,
   - $\emptyset \cup \emptyset \cup C_3 = \{4, 5\}$,
   - $C_1 \cup C_2 \cup \emptyset = \{3, 1, 2\}$,
   - $C_1 \cup \emptyset \cup C_3 = \{3, 4, 5\}$,
   - $\emptyset \cup C_2 \cup C_3 = \{1, 2, 4, 5\}$,
   - $C_1 \cup C_2 \cup C_3 = \{1, 2, 3, 4, 5\} = \Omega$.

Indeed, $\sigma(\mathcal{C})$ is the collection of the above 8 sets.

3. (a) To find c, note that $\sum_{x:p_X(x)>0} p_X(x) = 1$, so

$$\begin{aligned} 1 &= 0.2 + 0.4 + \sum_{x=1}^{\infty} c \cdot (0.5)^x \\ &= 0.6 + \frac{0.5c}{1-0.5}, \end{aligned}$$

which gives $c = 1 - 0.6 = 0.4$.

(b)

$$\begin{aligned} P(X > 25) &= \sum_{x=26}^{\infty} p_X(x) \\ &= \sum_{x=26}^{\infty} 0.4 \cdot (0.5)^x \\ &= \frac{0.4 \cdot (0.5)^{26}}{1-0.5} = 0.8 \cdot (0.5)^{26}. \end{aligned}$$

4. (a)

$$\begin{aligned} E(XY) &= 1 \cdot 2 \cdot P((X,Y) = (1,2)) + 3 \cdot 2 \cdot P((X,Y) = (3,2)) \\ &\quad + 3 \cdot 6 \cdot P((X,Y) = (3,6)) + +3 \cdot 7 \cdot P((X,Y) = (3,7)) \\ &= 2 \cdot 0.5 + 6 \cdot 0.1 + 18 \cdot 0.3 + 21 \cdot 0.1 \\ &= 9.1 \end{aligned}$$

(b) The possible values of XY are $1 \cdot 2 = 2$, $3 \cdot 2 = 6$, $3 \cdot 6 = 18$ and $3 \cdot 7 = 21$. Let p_{XY} be the PMF of XY, then $p_{XY}(\{2\}) = P((X,Y) = (1,2)) = 0.5$, $p_{XY}(\{6\}) = P((X,Y) = (3,2)) = 0.1$, $p_{XY}(\{18\}) = P((X,Y) = (3,6)) = 0.3$ and $p_{XY}(\{21\}) = P((X,Y) = (3,7)) = 0.1$. Thus

$$\begin{aligned} E(XY) &= 2 \cdot p_{XY}(\{2\}) + 6 \cdot p_{XY}(\{6\}) + 18 \cdot p_{XY}(\{18\}) + 21 \cdot p_{XY}(\{21\}) \\ &= 2 \cdot 0.5 + 6 \cdot 0.1 + 18 \cdot 0.3 + 21 \cdot 0.1 = 9.1. \end{aligned}$$

5. (a)

$$\begin{aligned}
E(X) &= \sum_{x:p_X(x)>0} x p_X(x) \\
&= \sum_{k=0}^{\infty} k\left(e^{-\lambda}\lambda^k/k!\right) \\
&= \sum_{k=1}^{\infty} k\left(e^{-\lambda}\lambda^k/k!\right) \\
&= \sum_{k=1}^{\infty} \left(e^{-\lambda}\lambda^k/(k-1)!\right) \\
(m=k-1) \quad &= \sum_{m=0}^{\infty} \left(e^{-\lambda}\lambda^{m+1}/m!\right) \\
&= \lambda \underbrace{\sum_{m=0}^{\infty} \left(e^{-\lambda}\lambda^{m}/m!\right)}_{=\sum_{m:p_X(m)>0} p_X(m)=1} \\
&= \lambda.
\end{aligned}$$

$$\begin{aligned}
E(X(X-1)) &= \sum_{x:p_X(x)>0} x(x-1) p_X(x) \\
&= \sum_{k=0}^{\infty} k(k-1)\left(e^{-\lambda}\lambda^k/k!\right) \\
&= \sum_{k=2}^{\infty} k(k-1)\left(e^{-\lambda}\lambda^k/k!\right) \\
&= \sum_{k=2}^{\infty} \left(e^{-\lambda}\lambda^k/(k-2)!\right) \\
(m=k-2) \quad &= \sum_{m=0}^{\infty} \left(e^{-\lambda}\lambda^{m+2}/m!\right) \\
&= \lambda^2 \underbrace{\sum_{m=0}^{\infty} \left(e^{-\lambda}\lambda^{m}/m!\right)}_{=\sum_{m:p_X(m)>0} p_X(m)=1} \\
&= \lambda^2.
\end{aligned}$$

(b) From Part (a), $E(X^2-X)=\lambda^2$ and $E(X)=\lambda$, so $E(X^2)=E(X^2-X)+E(X)=\lambda^2+\lambda$, and

$$Var(X)=E(X^2)-(E(X))^2=\lambda^2+\lambda-\lambda^2=\lambda.$$

6. (a)

$$\begin{aligned}
P(X>0.6) &= \int_{0.6}^{\infty} f_{0,1}(x)dx \\
&= \int_{0.6}^{1} 1dx = 0.4.
\end{aligned}$$

(b)

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2 f_{0,1}(x)dx \\ &= \int_0^1 x^2 dx = 1/3. \end{aligned}$$

7. (a) Note that $\{5\}$, $\{1,4\}$, $\{2\}$ are disjoint and

$$\begin{cases} A &= \{1,4\} \cup \{2\} \cup \{5\}; \\ B &= \{1,4\} \cup \{2\}; \\ C &= \{1,4\} \cup \{5\}, \end{cases}$$

so we can take D_1, D_2 and D_3 so that $\{D_1, D_2, D_3\}$ is $\{\{5\}, \{1,4\}$ $\{2\}\}$.

(b) By the additivity of P, we have

$$\begin{cases} P(A) &= P(\{1,4\}) + P(\{2\}) + P(\{5\}); \\ P(B) &= P(\{1,4\}) + P(\{2\}); \\ P(C) &= P(\{1,4\}) + P(\{5\}). \end{cases}$$

In the above equations, treat $P(A)$, $P(B)$ and $P(C)$ as known and solve for $P(\{5\})$, $P(\{1,4\})$ and $P(\{2\})$, then we have

$$\begin{cases} P(\{5\}) &= P(A) - P(B) \\ P(\{1,4\}) &= P(C) - P(A) + P(B) \\ P(\{2\}) &= P(B) - P(C) + P(A) - P(B) \end{cases} \tag{4}$$

and

$$P(\{3\}) = 1 - (P(\{1,4\}) + P(\{2\}) + P(\{5\})) = P(A).$$

If

$$(P(A), P(B), P(C)) = (0.5, 0.3, 0.1), \tag{5}$$

then (4) gives

$$P(\{1,4\}) = P(C) - P(A) + P(B) = 0.1 - 0.5 + 0.3 < 0,$$

which is impossible. Therefore, we cannot have (5).

8. To prove

$$\lim_{n\to\infty} P(A_n) = P\left(\cap_{n=1}^{\infty} A_n\right), \tag{6}$$

note that $\{A_n^c\}_{n=1}^{\infty}$ is an increasing sequence of events in $\mathcal{F}$, so by continuity of P for the increasing case (Fact 1 given in the problem), we have

$$\lim_{n\to\infty} P(A_n^c) = P\left(\cup_{n=1}^{\infty}(A_n^c)\right). \tag{7}$$

Thus

$$\begin{aligned} \lim_{n\to\infty} P(A_n) &= \lim_{n\to\infty}(1 - P(A_n^c)) \\ &\overset{(7)}{=} 1 - P\left(\cup_{n=1}^{\infty} A_n^c\right) \\ &= P\left((\cup_{n=1}^{\infty} A_n^c)^c\right) \\ \text{De Morgan's law} \quad &= P\left(\cap_{n=1}^{\infty} (A_n^c)^c\right) \\ &= P\left(\cap_{n=1}^{\infty} A_n\right) \end{aligned}$$

and (6) holds. Note that the last second equality follows from the result of Problem 1, which is known as a part of the De Morgan's laws.

9. Let
$$A_n = \left(-\frac{1}{n}, \frac{1}{n}\right)$$
for $n = 1,2, \ldots$, then $\{A_n\}_{n=1}^{\infty}$ is a decreasing sequence and $\cap_{n=1}^{\infty} A_n = \{0\}$. Therefore,
$$P(\{0\}) = \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} \left(0.4 + \frac{0.6}{n}\right) = 0.4.$$

10. Let
$$A_n = \{X \le -a_n\} = \{\omega \in \Omega : X(\omega) \in (-\infty, -a_n]\}$$
for $n = 1, 2, \ldots$, then the sequence $\{A_n\}_{n=1}^{\infty}$ is decreasing since $\{a_n\}_{n=1}^{\infty}$ is increasing. By the continuity of P for the decreasing case (the result in problem 8), we have
$$P(\cap_{n=1}^{\infty} A_n) = \lim_{n\to\infty} P(A_n). \tag{8}$$
Since $P(A_n) = F(-a_n)$, (8) becomes
$$\lim_{n\to\infty} F(-a_n) = P(\cap_{n=1}^{\infty} A_n). \tag{9}$$
In addition, the set $\cap_{n=1}^{\infty} A_n = \emptyset$, so
$$P(\cap_{n=1}^{\infty} A_n) = P(\emptyset) = 0,$$
which, together with (9), gives
$$\lim_{n\to\infty} F(-a_n) = 0.$$

Note. You do not have to prove $\cap_{n=1}^{\infty} A_n = \emptyset$, but be sure you understand why the result holds. The explanation is given below.

- To see that $\cap_{n=1}^{\infty} A_n = \emptyset$, note that if there exists $\omega \in \cap_{n=1}^{\infty} A_n$, we must have $X(\omega) \le -a_n$ for all n. Since $\lim_{n\to\infty} -a_n = -\infty$, $X(\omega) = -\infty$, which is impossible since X takes values in R. Therefore, there is no point in $\cap_{n=1}^{\infty} A_n$. That is, $\cap_{n=1}^{\infty} A_n = \emptyset$.

11. We will show that Q is a probability function defined on $\mathcal{F}$ by verifying the following:

    (a) $Q(A) \ge 0$ for all $A \in \mathcal{F}$.

    (b) $Q(\Omega) = 1$.

    (c) Suppose that $\{A_n\}_{n=1}^{\infty}$ is a sequence of disjoint events in $\mathcal{F}$, then
$$Q(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} Q(A_n). \tag{10}$$

- For (a), note that P is a probability function on $\mathcal{F}$, so $Q(A) = P(A|B) = P(A \cap B)/P(B) \ge 0$ since $P(B) > 0$ and $P(A \cap B) \ge 0$.
- For (b), $Q(\Omega) = P(\Omega|B) = P(\Omega \cap B)/P(B) = P(B)/P(B) = 1$.

- For (c), suppose that $\{A_n\}_{n=1}^{\infty}$ is a sequence of disjoint events in $\mathcal{F}$, then $\{A_n \cap B\}_{n=1}^{\infty}$ is also a sequence of disjoint events in $\mathcal{F}$, so

$$P(\cup_{n=1}^{\infty}(A_n \cap B)) = \sum_{n=1}^{\infty} P(A_n \cap B). \tag{11}$$

Compute $Q(\cup_{n=1}^{\infty} A_n)$ using the definition of Q and we have

$$\begin{aligned}
Q(\cup_{n=1}^{\infty} A_n) &= P(\cup_{n=1}^{\infty} A_n | B) \\
&= P(B \cap (\cup_{n=1}^{\infty} A_n))/P(B) \\
&= P(\cup_{n=1}^{\infty}(A_n \cap B))/P(B) \\
&\overset{(11)}{=} \frac{1}{P(B)} \left(\sum_{n=1}^{\infty} P(A_n \cap B) \right) \\
&= \sum_{n=1}^{\infty} \frac{P(A_n \cap B)}{P(B)} \\
&= \sum_{n=1}^{\infty} P(A_n | B) = \sum_{n=1}^{\infty} Q(A_n),
\end{aligned}$$

so (10) holds.

12. (a) $F(0) = \lim_{x \to 0^+} F(x) = \lim_{x \to 0^+} (0.5 + 0.5x) = 0.5$.

(b) $F(1) = \lim_{x \to 1^+} F(x) = \lim_{x \to 1^+} 1 = 1$.

(c) From the definition of F, it is clear that F is continuous at every point that is not in $\{0, 1\}$. Therefore, for $a \in R$, $P(X = a) = 0$ for $a \notin \{0, 1\}$. It remains to compute $P(X = 0)$ and $P(X = 1)$. Direct calculation gives

$$\begin{aligned}
P(X = 0) &= F(0) - \lim_{x \to 0^-} F(x) \\
&= 0.5 - \lim_{x \to 0^-} 0 = 0.5 - 0 = 0.5,
\end{aligned}$$

and

$$\begin{aligned}
P(X = 1) &= F(1) - \lim_{x \to 1^-} F(x) \\
&= 1 - \lim_{x \to 1^-} (0.5 + 0.5x) = 1 - 1 = 0.
\end{aligned}$$

(d)

$$\begin{aligned}
P(0 \le X \le 1) &= P(0 < X \le 1) + P(X = 0) \\
&= F(1) - F(0) + 0.5 \\
&= 1 - 0.5 + 0.5 = 1.
\end{aligned}$$

Here we have used the result that $P(X = 0) = 0.5$ (from Part (c)).

(e) From Part (c), we have that $P(X = a) = 0$ for $a \neq 0$ and $P(X = 0) = 0.5$. If X is discrete, then 0 is the only possbile value of X and we must have $P(X = 0) = 1$, which contradicts with the fact that $P(X = 0) = 0.5$. Therefore, X is not discrete.

13. Let $S_X = \{x : f_{a,b}(x) > 0\} = (a, b)$, then

$$\{(x-a)/(b-a) : x \in (a,b)\} = (0,1).$$

Solving $y = (x-a)/(b-a)$ for x gives $x = a + (b-a)y$. For $y \notin (0,1)$, let $f_Y(y) = 0$, and for $y \in (0,1)$, let

$$f_Y(y) = f_{a,b}\left(a + (b-a)y\right) \left| \frac{d}{dy}\left(a + (b-a)y\right) \right|,$$

then f_Y is a PDF of $(X-a)/(b-a)$. The expression of f_Y can be further simplified: for $y \in R$,

$$\begin{aligned} f_Y(y) &= I_{(0,1)}(y) \cdot f_{a,b}\left(a + (b-a)y\right) \left| \frac{d}{dy}\left(a + (b-a)y\right) \right| \\ &= I_{(0,1)}(y) \cdot \frac{1}{b-a} \cdot I_{(a,b)}(a + (b-a)y) \cdot (b-a) \\ &= I_{(0,1)}(y) \cdot I_{(a,b)}(a + (b-a)y) \\ &= I_{(0,1)}(y). \end{aligned}$$

14. Let $S_X = \{x : f_X(x) > 0\}$, then $S_X = (0, \infty)$ and

$$\{\sqrt{x} : x \in (0,\infty)\} = (0,\infty).$$

Solving $y = \sqrt{x}$ for x gives $x = y^2$. For $y \in (0,\infty)$, let

$$f_Y(y) = f_X(y^2) \left| \frac{d}{dy}\left(y^2\right) \right|,$$

and for $y \notin (0,\infty)$, let $f_Y(y) = 0$, then f_Y is a PDF of $Y = \sqrt{X}$. The expression of f_Y can be further simplified:

$$\begin{aligned} f_Y(y) &= I_{(0,\infty)}(y) \cdot f_X(y^2) \left| \frac{d}{dy}\left(y^2\right) \right| \\ &= I_{(0,\infty)}(y) \cdot 2y^2 e^{-(y^2)^2} I_{(0,\infty)}(y^2) \cdot |2y| \\ &= I_{(0,\infty)}(y) \cdot 4y^3 e^{-y^4} \end{aligned}$$

for $y \in R$.

15. Let F_Y be the CDF of $Y = X^2$. We will find F_Y first. For $t < 0$, $F_Y(t) = P(X^2 \le t) = 0$. For $t = 0$,

$$F_Y(0) = P(X^2 \le 0) = P(X = 0) = 0$$

since X has a PDF. For $t > 0$,

$$\begin{aligned} F_Y(t) &= P(X^2 \le t) \\ &= P(-\sqrt{t} \le X \le \sqrt{t}) \\ &= \int_{-\sqrt{t}}^{0} (-x)dx + \int_0^{\sqrt{t}} 0.5e^{-x}dx \\ (u = -x) \quad &= \int_0^{\sqrt{t}} udu + \int_0^{\sqrt{t}} 0.5e^{-x}dx \\ (y = u^2; y = x^2) \quad &= \int_0^t 0.5dy + \int_0^t \frac{0.5e^{-\sqrt{y}}}{2\sqrt{y}}dy. \qquad (12) \end{aligned}$$

Take

$$f(y) = \left(\frac{1}{2} + \frac{1}{4\sqrt{y}} e^{-\sqrt{y}}\right) I_{(0,\infty)}(y) \tag{13}$$

for $y \in R$, then for $t > 0$,

$$\int_{-\infty}^{t} f(y)dy = \int_{0}^{t} \left(\frac{1}{2} + \frac{e^{-\sqrt{y}}}{4\sqrt{y}}\right) dy \overset{(12)}{=} F_Y(t),$$

and for $t \leq 0$,

$$\int_{-\infty}^{t} f(y)dy = 0 = F_Y(t).$$

We have verify that

$$\int_{-\infty}^{t} f(y)dy = F_Y(t)$$

for all $t \in R$, so the f given in (13) is a PDF of Y.

16. (a)

$$\begin{aligned} P(X > 2) &= 1 - P(X \leq 2) \\ &= 1 - F(2) = 1 - (1 - e^{-4}) = e^{-4}. \end{aligned}$$

(b) Note that

$$F'(x) = \frac{d}{dx}(1 - e^{-2x}) = 2e^{-2x}$$

for $x > 0$ and $F'(x) = 0$ for $x < 0$. Let $f(x) = 2e^{-2x} I_{(0,\infty)}(x)$ for $x \in R$. We will show that f is a PDF of X by verifying

$$F(t) = \int_{-\infty}^{t} f(x)dx \tag{14}$$

for $t \in R$. Note that for $t \leq 0$, $\int_{-\infty}^{t} f(x)dx = 0$ since $f(x) = 0$ for $x \leq 0$, and $F(t) = 0$ for $t \leq 0$. Thus (14) holds clearly for $t \leq 0$. For $t > 0$,

$$\begin{aligned} \int_{-\infty}^{t} f(x)dx &= \int_{0}^{t} (2e^{-2x})dx \\ &= \left(-e^{-2x}\right)\Big|_{x=0}^{t} \\ &= 1 - e^{-2t} = F(t), \end{aligned}$$

so (14) holds for $t > 0$ as well. Since (14) holds for all $t \in R$, f is a PDF of X.

17. (a) Let F be the CDF of X, then for $t \in R$,

$$\begin{aligned} F(t) &= P(X \leq t) \\ &= \int_{-\infty}^{t} 2xe^{-x^2} I_{(0,\infty)}(x)dx \\ &= I_{(0,\infty)}(t) \cdot \int_{0}^{t} 2xe^{-x^2}dx \\ &= I_{(0,\infty)}(t) \cdot (1 - e^{-t^2}). \end{aligned}$$

(b) For $a \in (0,1)$, solving $F(t) = a$ for t gives $e^{-t^2} = 1 - a$ and $t = \sqrt{-\ln(1-a)}$, so the median of the distribution of X is

$$\sqrt{-\ln(1-0.5)} = \sqrt{\ln(2)},$$

and the IQR (interquartile range) of the distribution of X is

$$\sqrt{-\ln(1-0.75)} - \sqrt{-\ln(1-0.25)} = \sqrt{\ln(4)} - \sqrt{\ln(4/3)}.$$

18. (a) Let F be the CDF of $Y = g(X)$. It is clear that $g(x) \leq 0.5$ for every $x \in R$, so $F(0.5) = P(g(X) \leq 0.5) = 1$ and $F(t) = 1$ for $t \geq 0.5$. For $t < 0.5$,

$$\begin{aligned}
F(t) &= P(Y \leq t) \\
&= P(g(X) \leq t \text{ and } X \leq 0.5) + P(g(X) \leq t \text{ and } X > 0.5) \\
&= P(X \leq t \text{ and } X \leq 0.5) + P(0.5 \leq t \text{ and } X > 0.5) \\
&= P(X \leq t) + I_{[0.5,\infty)}(t)P(X > 0.5) \\
&= \int_{-\infty}^{t} e^{-x} \cdot I_{(0,\infty)}(x)dx \\
&= I_{(0,\infty)}(t)\int_{0}^{t} e^{-x}dx \\
&\overset{(t<0.5)}{=} (1-e^{-t})I_{(0,0.5)}(t).
\end{aligned}$$

From the above calculation, the CDF of Y is F, which is given by

$$F(t) = (1-e^{-t})I_{(0,0.5)}(t) + I_{[0.5,\infty)}(t) \tag{15}$$

for $t \in R$.

(b) Y cannot have a PDF since

$$P(Y = 0.5) = P(X \geq 0.5) = \int_{0.5}^{\infty} e^{-x}dx > 0.$$

(c) The CDF of Y is the F given in (15). From (15), $F(t) = 1 > 1-e^{-0.5}$ for $t \geq 0.5$ and for $t < 0.5$, $F(t) = 1 - e^{-t} < 1 - e^{-0.5}$. Therefore, 0.5 is the quantile of order $(1 - e^{-0.5})$ of the distribution of Y.

19. (a) Let $\mu = E(X)$, then by definition,

$$\begin{aligned}
Var(X) &= E(X-\mu)^2 \\
&= E(X^2 - 2\mu X + \mu^2) \\
&= E(X^2) + E(-2\mu X) + E(\mu^2) \\
&= E(X^2) - 2\mu \underbrace{E(X)}_{=\mu} + \mu^2 \\
&= E(X^2) - \mu^2 \\
&= E(X^2) - (E(X))^2.
\end{aligned}$$

(b) Let $Y = X - E(X)$, then $cX - E(cX) = cX - cE(X) = cY$, so

$$\begin{aligned}
Var(cX) &= E(cX - E(cX))^2 \\
&= E\left(c^2Y^2\right) \\
&= c^2E(Y^2) = c^2Var(X).
\end{aligned}$$

20. (b) Note that

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} f_X(x)dx \\
&= \int_{-\infty}^{-1} x\left(\frac{1}{2x^2}\right)dx + \int_{1}^{\infty} x\left(\frac{1}{2x^2}\right)dx,
\end{aligned}
$$

where

$$
\begin{aligned}
&\int_{1}^{\infty} x\left(\frac{1}{2x^2}\right)dx \\
&= \lim_{b\to\infty}\int_{1}^{b}\frac{1}{2x}dx \\
&= \lim_{b\to\infty}\frac{\ln(b)}{2} = \infty
\end{aligned}
\tag{16}
$$

and

$$
\begin{aligned}
&\int_{-\infty}^{-1} x\left(\frac{1}{2x^2}\right)dx \\
&= \lim_{b\to\infty}\int_{-b}^{-1}\frac{1}{2x}dx \\
&\overset{y=-x}{=} \lim_{b\to\infty} -\int_{1}^{b}\frac{1}{2y}dy = -\infty,
\end{aligned}
$$

so $E(X)$ cannot be defined.

(a)

$$
\begin{aligned}
E(X\cdot I_{(0,\infty)}(X)) &= \int_{-\infty}^{\infty} x\cdot I_{(0,\infty)}(x)f_X(x)dx \\
&= \int_{1}^{\infty} x\left(\frac{1}{2x^2}\right)dx \overset{(16)}{=} \infty,
\end{aligned}
$$

so $X\cdot I_{(0,\infty)}(X)$ is not integrable.

21. We will first state and prove the result in Fact 1 to simplify $\lim_{h\to 0^+} P((X,Y)\in (x_0,x_0+h)\times(y_0,y_0+h))/h^2$ and $\lim_{h\to 0^+} P((U,V)\in R_h)/h^2$.

Fact 1 Suppose that f is a real-valued function that is continuous on an open set O in R^2 and $(x_0,y_0)\in O$. Suppose that $\{A_h : h>0\}$ is a collection of regions in R^2 such that for $\varepsilon_1>0$, there exists $\delta_1>0$ such that

$$h\in(0,\delta_1)\Rightarrow (x_0,y_0)\in A_h\subset B((x_0,y_0),\varepsilon_1), \tag{17}$$

where

$$B((x_0,y_0),\varepsilon_1) = \{(x,y)\in R^2 : \|(x,y)-(x_0,y_0)\| < \varepsilon_1\},$$

then

$$\lim_{h\to 0^+}\frac{\int_{A_h} f(x,y)d(x,y)}{\int_{A_h} 1d(x,y)} = f(x_0,y_0). \tag{18}$$

Proof of Fact 1. Since f is continuous at (x_0, y_0), for $\varepsilon_2 > 0$, there exists $\delta_2 > 0$ such that

$$\|(x, y) - (x_0, y_0)\| < \delta_2 \Rightarrow |f(x, y) - f(x_0, y_0)| < \varepsilon_2/2. \tag{19}$$

By assumption, there exists $\delta_1 > 0$ such that (17) holds with ε_1 replaced by δ_2. Thus

$$\begin{aligned} h \in (0, \delta_1) \quad &\Rightarrow \quad (x_0, y_0) \in A_h \subset B((x_0, y_0), \delta_2) \\ &\overset{(19)}{\Rightarrow} \quad |f(x, y) - f(x_0, y_0)| < \varepsilon_2/2 \\ &\Rightarrow \end{aligned} \tag{20}$$

$$\begin{aligned} &\left| \frac{\int_{A_h} f(x, y) d(x, y)}{\int_{A_h} 1 d(x, y)} - f(x_0, y_0) \right| \\ = &\left| \frac{\int_{A_h} f(x, y) - f(x_0, y_0) d(x, y)}{\int_{A_h} 1 d(x, y)} \right| \\ \leq &\frac{\int_{A_h} |f(x, y) - f(x_0, y_0)| d(x, y)}{\int_{A_h} 1 d(x, y)} \\ \overset{(20)}{\leq} &\ \varepsilon_2/2 < \varepsilon_2. \end{aligned}$$

In summary, we have shown that for $\varepsilon_2 > 0$, there exists $\delta_1 > 0$ such that

$$h \in (0, \delta_1) \Rightarrow \left| \frac{\int_{A_h} f(x, y) d(x, y)}{\int_{A_h} 1 d(x, y)} - f(x_0, y_0) \right| < \varepsilon_2,$$

so (18) holds and the proof of Fact 1 is complete.

Next, we simplify $\lim_{h \to 0^+} P((X, Y) \in (x_0, x_0 + h) \times (y_0, y_0 + h))/h^2$. Apply Fact 1 with $f = f_{X,Y}$, $(x_0, y_0) = (x_0, y_0)$ and $A_h = (x_0, x_0 + h) \times (y_0, y_0 + h)$, we have

$$\lim_{h \to 0^+} \frac{\int_{(x_0, x_0+h) \times (y_0, y_0+h)} f_{X,Y}(x, y) d(x, y)}{\int_{(x_0, x_0+h) \times (y_0, y_0+h)} 1 d(x, y)} = f_{X,Y}(x_0, y_0). \tag{21}$$

Since

$$\begin{aligned} &\lim_{h \to 0^+} \frac{P((X, Y) \in (x_0, x_0 + h) \times (y_0, y_0 + h))}{h^2} \\ = &\lim_{h \to 0^+} \frac{\int_{(x_0, x_0+h) \times (y_0, y_0+h)} f_{X,Y}(x, y) d(x, y)}{h^2} \\ = &\underbrace{\lim_{h \to 0^+} \frac{\int_{(x_0, x_0+h) \times (y_0, y_0+h)} f_{X,Y}(x, y) d(x, y)}{\int_{(x_0, x_0+h) \times (y_0, y_0+h)} 1 d(x, y)}}_{= f_{X,Y}(x_0, y_0) \text{ by } (21)} \cdot \underbrace{\lim_{h \to 0^+} \frac{\int_{(x_0, x_0+h) \times (y_0, y_0+h)} 1 d(x, y)}{h^2}}_{=1}, \end{aligned}$$

we have

$$\lim_{h \to 0^+} \frac{P((X, Y) \in (x_0, x_0 + h) \times (y_0, y_0 + h))}{h^2} = f_{X,Y}(x_0, y_0). \tag{22}$$

Next, we simplify $\lim_{h\to 0^+} P((U,V)\in R_h)/h^2$. Apply Fact 1 with $f = f_{U,V}$, $(x_0,y_0)=(u_0,v_0)$ and $A_h = R_h$, we have

$$\lim_{h\to 0^+}\frac{\int_{R_h} f_{U,V}(u,v)d(u,v)}{\int_{R_h} 1d(x,y)} = f_{U,V}(u_0,v_0). \tag{23}$$

Since

$$\begin{aligned}
&\lim_{h\to 0^+}\frac{P((U,V)\in R_h)}{h^2}\\
&= \lim_{h\to 0^+}\frac{\int_{R_h} f_{U,V}(u,v)d(u,v)}{h^2}\\
&= \underbrace{\lim_{h\to 0^+}\frac{\int_{R_h} f_{U,V}(u,v)d(u,v)}{\int_{R_h} 1d(u,v)}}_{=f_{U,V}(u_0,v_0)\text{ by (23)}}\cdot\lim_{h\to 0^+}\frac{\int_{R_h} 1d(u,v)}{h^2}
\end{aligned}$$

and

$$\begin{aligned}
&\int_{R_h} 1d(u,v)\\
=\ & \text{the area of the parallelogram } ABCD\\
=\ & \|\overrightarrow{AB}\|\cdot\|\overrightarrow{AD}\|\cdot\sqrt{1-\left(\frac{\overrightarrow{AB}\cdot\overrightarrow{AD}}{\|\overrightarrow{AB}\|\cdot\|\overrightarrow{AD}\|}\right)^2}\\
=\ & \sqrt{\|\overrightarrow{AB}\|^2\|\overrightarrow{AD}\|^2-(\overrightarrow{AB}\cdot\overrightarrow{AD})^2},
\end{aligned}$$

where

$$\begin{aligned}
\overrightarrow{AB} &= (u_0+u_x(x_0,y_0)h, v_0+v_x(x_0,y_0)h)-(u_0,v_0)\\
&= h\cdot(u_x(x_0,y_0),v_x(x_0,y_0)),
\end{aligned}$$

$$\begin{aligned}
\overrightarrow{AD} &= (u_0+u_y(x_0,y_0)h, v_0+v_y(x_0,y_0)h)-(u_0,v_0)\\
&= h\cdot(u_y(x_0,y_0),v_y(x_0,y_0)),
\end{aligned}$$

and

$$\begin{aligned}
&\|\overrightarrow{AB}\|^2\|\overrightarrow{AD}\|^2-(\overrightarrow{AB}\cdot\overrightarrow{AD})^2\\
=\ & h^2[(u_x(x_0,y_0))^2+(v_x(x_0,y_0)^2)]\cdot h^2[(u_y(x_0,y_0))^2+(v_y(x_0,y_0)^2)]\\
& -[h^2(u_x(x_0,y_0)u_y(x_0,y_0)+v_x(x_0,y_0)v_y(x_0,y_0))]^2\\
=\ & h^4|u_x(x_0,y_0)v_y(x_0,y_0)-v_x(x_0,y_0)u_y(x_0,y_0)|^2,
\end{aligned}$$

we have

$$\begin{aligned}
&\lim_{h\to 0^+}\frac{P((U,V)\in R_h}{h^2}\\
=\ & f_{U,V}(u_0,v_0)\lim_{h\to 0^+}\frac{\int_{R_h} 1d(u,v)}{h^2}\\
=\ & f_{U,V}(u_0,v_0)\lim_{h\to 0^+}\frac{\sqrt{h^4|u_x(x_0,y_0)v_y(x_0,y_0)-v_x(x_0,y_0)u_y(x_0,y_0)|^2}}{h^2}\\
=\ & f_{U,V}(u_0,v_0)|u_x(x_0,y_0)v_y(x_0,y_0)-v_x(x_0,y_0)u_y(x_0,y_0)|. \qquad (24)
\end{aligned}$$

From (22), (24) and the result that

$$\lim_{h\to 0^+} \frac{P((X,Y)\in(x_0,x_0+h)\times(y_0,y_0+h))}{h^2} = \lim_{h\to 0^+} \frac{P((U,V)\in R_h)}{h^2},$$

we have

$$f_{X,Y}(x_0,y_0) = f_{U,V}(u_0,v_0)|u_x(x_0,y_0)v_y(x_0,y_0) - v_x(x_0,y_0)u_y(x_0,y_0)|.$$

Since

$$u_x(x_0,y_0)v_y(x_0,y_0) - v_x(x_0,y_0)u_y(x_0,y_0)$$

is the determinant of $J(x_0,y_0)$, we have

$$f_{X,Y}(x_0,y_0) = f_{U,V}(u_0,v_0) \cdot |\text{ determinant of } J(x_0,y_0)|\,.$$

22.

$$\begin{aligned} E(Y) &= E\left(\frac{X-\mu}{\sigma}\right) \\ &= \frac{1}{\sigma}E(X-\mu) \\ &= \frac{1}{\sigma}\left(E(X)+E(-\mu)\right) = \frac{1}{\sigma}\left(\mu+(-\mu)\right) = 0. \end{aligned}$$

$$\begin{aligned} Var(Y) &= Var\left(\frac{X-\mu}{\sigma}\right) \\ &= \frac{1}{\sigma^2}Var(X-\mu) \\ &= \frac{Var(X)}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1. \end{aligned}$$

23. Note that for a positive integer k,

$$\frac{d^k}{dt^k}e^{2t} = 2^k \cdot e^{2t}$$

and

$$\frac{d^k}{dt^k}0.6 = 0,$$

so for $k \in \{1,2,3,4\}$,

$$\frac{d^k}{dt^k}M_X(t) = 0.4 \cdot 2^k \cdot e^{2t}$$

and

$$E(X^k) = \left.\frac{d^k}{dt^k}M_X(t)\right|_{t=0} = 0.4 \cdot 2^k, \tag{25}$$

which gives $E(X) = 0.8$, $E(X^2) = 1.6$, $E(X^3) = 3.2$, and $E(X^4) = 6.4$. Below is another approach for finding

$$\left.\frac{d^k}{dt^k}M_X(t)\right|_{t=0}.$$

Apply the result that

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

for $x \in R$, we have

$$M_X(t) = 0.6 + 0.4\left(1 + \sum_{k=1}^{\infty} \frac{(2t)^k}{k!}\right)$$

for $t \in R$. Therefore, for $k \geq 1$,

$$\frac{d^k}{dt^k} M_X(t)\bigg|_{t=0} = k! \cdot 0.4 \cdot \frac{2^k}{k!} = 0.4 \cdot 2^k,$$

so (25) still holds and the $E(X^k)$ values remain the same as those computed above.

24. Let M_Y be the MGF of Y, then

$$\begin{aligned} M_Y(t) &= E(e^{tY}) \\ &= e^{t\cdot 0} P(Y=0) + e^{t\cdot 2} P(Y=2) \\ &= 1 \cdot 0.6 + e^{2t} \cdot 0.4 = M_X(t) \end{aligned}$$

for $t \in (-\infty, \infty)$, where M_X is the MGF of the random variable X in Problem 23. Since Y and X have the same MGF, they have the same distribution.

25. (a) The random variable X's distribution is the Poisson distribution with mean λ, so X has PMF p_X, where

$$p_X(k) = \begin{cases} e^{-\lambda}\lambda^k/k! & \text{if } k \in \{0, 1, 2, \ldots\}; \\ 0 & \text{otherwise.} \end{cases}$$

Let M_X be the MGF of X, then

$$\begin{aligned} M_X(t) &= E(e^{tX}) \\ &= \sum_{k=0}^{\infty} e^{tk} e^{-\lambda} \cdot \frac{\lambda^k}{k!} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} \\ &= e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)} \end{aligned}$$

for $t \in (-\infty, \infty)$.

(b) From Part (a), $M_X(t) = e^{\lambda(e^t-1)}$ for $t \in (-\infty, \infty)$. Since

$$\frac{d}{dt} M_X(t) = \frac{d}{dt} e^{\lambda(e^t-1)} = e^{\lambda(e^t-1)} \frac{d}{dt} \lambda(e^t - 1) = \lambda e^t M_X(t)$$

and

$$\begin{aligned} \frac{d^2}{dt^2} M_X(t) &= \frac{d}{dt} \lambda e^t M_X(t) \\ &= M_X(t) \frac{d}{dt}(\lambda e^t) + (\lambda e^t) \frac{d}{dt} M_X(t) \\ &= (\lambda e^t)(M_X(t) + M_X'(t)) \end{aligned}$$

for $t \in (-\infty, \infty)$, we have

$$E(X) = M'_X(0) = \lambda e^0 M_X(0) = \lambda$$

and

$$E(X^2) = M''_X(0) = (\lambda e^0)(M_X(0) + M'_X(0)) = \lambda(1+\lambda).$$

Therefore,

$$Var(X) = E(X^2) - (E(X))^2 = \lambda(1+\lambda) - \lambda^2 = \lambda.$$

26.

$$\begin{aligned} M_Y(t) &= E(e^{t(a+bX)}) \\ &= E(e^{ta} \cdot e^{tbX}) \\ &= e^{ta} E(e^{tbX}) \\ &= e^{ta} M_X(tb) \end{aligned}$$

for $|tb| < h$. If $b \neq 0$, then $|tb| < h \Leftrightarrow t \in (-h/|b|, h/|b|)$. Thus if $b \neq 0$,

$$M_Y(t) = e^{ta} M_X(tb) < \infty$$

for $t \in (-h/|b|, h/|b|)$.

27. (a) For $t \in R$,

$$\begin{aligned} M_Z(t) &= E(e^{tZ}) \\ &= \int_{-\infty}^{\infty} e^{tz} f_{0,1}(z) dz \\ &= \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(z-t)^2/2} \cdot e^{t^2/2} dz \\ &= e^{0.5t^2} \int_{-\infty}^{\infty} f_{t,1}(z) dz. \end{aligned}$$

Since $f_{t,1}$ is a PDF, the integral $\int_{-\infty}^{\infty} f_{t,1}(z)dz = 1$ and we have

$$M_Z(t) = \int_{-\infty}^{\infty} e^{tz} f_{0,1}(z) dz = e^{0.5t^2} \tag{26}$$

for $t \in R$.

(b) Note that

$$M_Z(t) = e^{0.5t^2} = \sum_{k=0}^{\infty} \frac{(0.5t^2)^k}{k!}$$

for $t \in R$, so

$$E(Z^6) = M_Z^{(6)}(0) = 6! \frac{(0.5)^3}{3!} = 15.$$

(c) We will show that Y and Z have the same distribution by verifying that they have the same MGF. Let M_Y be the MGF of Y, then

$$
\begin{aligned}
M_Y(t) &= E(e^{tY}) \\
&= E(e^{t(X-\mu)/\sigma}) \\
&= \int_{-\infty}^{\infty} e^{t(x-\mu)/\sigma} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)} dx \\
&\overset{z=(x-\mu)/\sigma}{=} \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-z^2/2} \cdot \sigma dz \\
&= \int_{-\infty}^{\infty} e^{tz} \underbrace{\frac{1}{\sqrt{2\pi}} e^{-z^2/2}}_{=f_{0,1}(z)} dz \\
&\overset{(26)}{=} e^{0.5t^2} = M_Z(t)
\end{aligned}
$$

for $t \in R$. Since Y and Z have the same MGF, Y and Z have the same distribution.

(d) Since $Y = (X-\mu)/\sigma$, we have $X = \mu + \sigma Y$. Let M_X be the MGF of X, then

$$
\begin{aligned}
M_X(t) &= E(e^{tX}) \\
&= E(e^{t(\mu+\sigma Y)}) \\
&\overset{\text{Problem 26}}{=} e^{t\mu} M_Y(t\sigma) \\
&\overset{\text{Part (a)}}{=} e^{t\mu} e^{0.5(t\sigma)^2} \\
&= e^{\mu t + 0.5\sigma^2 t^2}
\end{aligned}
$$

for $t \in R$.

To compute $E(X)$ and $Var(X)$, note that

$$
\begin{aligned}
M_X'(t) &= \frac{d}{dt} e^{\mu t + 0.5\sigma^2 t^2} \\
&= \underbrace{e^{\mu t + 0.5\sigma^2 t^2}}_{=M_X(t)} \cdot \frac{d}{dt}(\mu t + 0.5\sigma^2 t^2) \\
&= M_X(t)(\mu + \sigma^2 t)
\end{aligned}
$$

and

$$
\begin{aligned}
M_X''(t) &= \frac{d}{dt}\left(M_X(t)(\mu + \sigma^2 t)\right) \\
&= M_X'(t)(\mu + \sigma^2 t) + M_X(t) \cdot \sigma^2,
\end{aligned}
$$

so

$$
E(X) = M_X'(0) = \underbrace{M_X(0)}_{=1} \cdot (\mu + \sigma^2 t)\big|_{t=0} \mu = \mu \tag{27}
$$

and

$$
E(X^2) = M_X''(0) = \underbrace{M_X'(0)}_{\overset{(27)}{=}\mu} \cdot (\mu + \sigma^2 t)\big|_{t=0} + \underbrace{M_X(0)}_{=1} \cdot \sigma^2 = \mu^2 + \sigma^2,
$$

which gives

$$
Var(X) = E(X^2) - [E(X)]^2 = \mu^2 + \sigma^2 - \mu^2 = \sigma^2.
$$

28. (a) Let $H(x) = (1 - e^{-x})I_{[0,\infty)}(x)$, then

$$F_{X,Y}(x, y) = 0.5G(x)G(y) + 0.5H(x)H(y)$$

for $(x, y) \in R^2$ and

$$\begin{aligned}
& P(0 < X \le 1 \text{ and } 1 < Y \le 2) \\
& = F_{X,Y}(1,2) - F_{X,Y}(0,2) - F_{X,Y}(1,1) + F_{X,Y}(0,1) \\
& = 0.5G(1)G(2) + 0.5H(1)H(2) - [0.5G(0)G(2) + 0.5H(0)H(2)] \\
& \quad -[0.5G(1)G(1) + 0.5H(1)H(1)] + 0.5G(0)G(1) + 0.5H(0)H(1) \\
& = 0.5(G(2) - G(1))(G(1) - G(0)) + 0.5(H(2) - H(1))(H(1) - H(0)) \\
& = 0.5(1-1)(1-0) + 0.5(1 - e^{-2} - (1 - e^{-1}))(1 - e^{-1} - (1-1)) \\
= \; & 0.5e^{-1} - e^{-2} + 0.5e^{-3}.
\end{aligned}$$

(b) Let F_X be the CDF of X, then

$$\begin{aligned}
F_X(x) &= \lim_{y\to\infty} F_{X,Y}(x, y) \\
&= \lim_{y\to\infty} (0.5G(x)G(y) + 0.5H(x)H(y))
\end{aligned}$$

for $x \in R$. Since

$$\lim_{y\to\infty} G(y) = \lim_{x\to\infty} (xI_{(0,1)}(x) + I_{[1,\infty)}(x)) = 1$$

and

$$\lim_{y\to\infty} H(y) = \lim_{x\to\infty} (1 - e^{-x})I_{[0,\infty)}(x) = 1,$$

we have

$$\begin{aligned}
F_X(x) &= \lim_{y\to\infty} (0.5G(x)G(y) + 0.5H(x)H(y)) \\
&= 0.5G(x) + 0.5H(x) \\
&= 0.5(xI_{(0,1)}(x) + I_{[1,\infty)}(x)) + 0.5(1 - e^{-x})I_{[0,\infty)}(x) \\
&= (0.5x + 0.5 - 0.5e^{-x}))I_{(0,1)}(x) + (1 - 0.5e^{-x})I_{[1,\infty)}(x)
\end{aligned}$$

for $x \in R$.

29. Since

$$\begin{aligned}
& P((X, Y, Z) \in (a, b] \times (c, d] \times (e, f]) \\
= \; & P((X, Y) \in (a, b] \times (c, d] \text{ and } Z \le f) \\
& -P((X, Y) \in (a, b] \times (c, d] \text{ and } Z \le e) \qquad (28)
\end{aligned}$$

We will first express $P((X, Y) \in (a, b] \times (c, d] \text{ and } Z \le z)$ using the joint CDF of (X, Y, Z). Note that for $z \in (-\infty, \infty)$,

$$\begin{aligned}
& P((X, Y) \in (a, b] \times (c, d] \text{ and } Z \le z) \\
= \; & P(X \in (a, b] \text{ and } Y \le d \text{ and } Z \le z) \\
& -P(X \in (a, b] \text{ and } Y \le c \text{ and } Z \le z) \qquad (29)
\end{aligned}$$

and for $y \in \{c, d\}$,

$$\begin{aligned}
& P(X \in (a, b] \text{ and } Y \le y \text{ and } Z \le z) \\
= \; & P(X \le b \text{ and } Y \le y \text{ and } Z \le z) \\
& -P(X \le a \text{ and } Y \le y \text{ and } Z \le z) \\
& = F(b, y, z) - F(a, y, z),
\end{aligned}$$

so (29) becomes

$$\begin{aligned}
&P((X,Y) \in (a,b] \times (c,d] \text{ and } Z \le z) \\
&= F(b,d,z) - F(b,c,z) - (F(a,d,z) - F(a,c,z)) \qquad (30)
\end{aligned}$$

for $z \in (-\infty, \infty)$. In (28), replace each probability of the form $P((X,Y) \in (a,b] \times (c,d] \text{ and } Z \le z)$ ($z = e,\ f$) with the CDF expression in (30), then we have

$$\begin{aligned}
& P((X,Y,Z) \in (a,b] \times (c,d] \times (e,f]) \\
= \;& F(b,d,f) - F(b,c,f) - (F(a,d,f) - F(a,c,f)) \\
& -(F(b,d,e) - F(b,c,e) - (F(a,d,e) - F(a,c,e))) \\
= \;& F(b,d,f) - F(b,c,f) - F(a,d,f) + F(a,c,f) \\
& -F(b,d,e) + F(b,c,e) + F(a,d,e) - F(a,c,e).
\end{aligned}$$

30. (a)

$$\begin{aligned}
1 &= \int_{R^2} f_{X,Y}(x,y)d(x,y) \\
&= \int_0^1 \int_0^1 cxdydx \\
&= \int_0^1 cx \int_0^1 1dydx \\
&= \int_0^1 cxdx = c/2,
\end{aligned}$$

so $c = 2$.

(b)

$$\begin{aligned}
P((X+2Y) \le 1) &= \int_{\{(x,y):x+2y\le 1\}} f_{X,Y}(x,y)d(x,y) \\
&= \int_{\{(x,y):x+2y\le 1\}\cap(0,1)\times(0,1)} 2xd(x,y) \\
&= \int_0^{1/2} \int_0^{(1-2y)} 2xdxdy \\
&= \int_0^{1/2} (1-2y)^2 dy = \frac{1}{6}.
\end{aligned}$$

(c) Let

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$$

for $y \in (-\infty, \infty)$, then f_Y is a PDF of Y. Note that for $y \in R$ and $y \notin (0,1)$, $f_{X,Y}(x,y) = 0$, so

$$\int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = 0$$

for $y \notin (0,1)$. In addition, for $y \in (0,1)$,

$$\int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \int_0^1 2xdx = 1.$$

Therefore,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = I_{(0,1)}(y)$$

for $y \in R$.

31. Let

$$c_1 = \int_{-\infty}^{\infty} g(x)dx \qquad \text{and} \qquad c_2 = \int_{-\infty}^{\infty} h(x)dx,$$

then

$$1 = \int_{R^2} f_{X,Y}(x,y)d(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)dxdy = c_1c_2. \tag{31}$$

Let

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \int_{-\infty}^{\infty} g(x)h(y)dy = c_2 g(x)$$

for $x \in R$ and

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \int_{-\infty}^{\infty} g(x)h(y)dx = c_1 h(y)$$

for $y \in R$, then f_X and f_Y are PDF's of X and Y, respectly, and

$$f_X(x)f_Y(y) = c_2 g(x) c_1 h(y) \stackrel{(31)}{=} g(x)h(y) = f_{X,Y}(x,y)$$

for $(x,y) \in R^2$.

32. Let

$$A_n = \{w \in \Omega : X(w) \leq x \text{ and } Y(w) \leq n\}$$

for $n \geq 1$, then it is clear that

$$A_n \subset A_{n+1} \text{ for all } n \geq 1 \tag{32}$$

and it can be shown that

$$\cup_{n=1}^{\infty} A_n = \{w \in \Omega : X(w) \leq x\}, \tag{33}$$

so

$$\begin{aligned}
\lim_{y\to\infty} P(X \leq x \text{ and } Y \leq y) &= \lim_{n\to\infty} P(X \leq x \text{ and } Y \leq n) \\
&= \lim_{n\to\infty} P(A_n) \\
&\stackrel{(32)}{=} P(\cup_{n=1}^{\infty} A_n) \\
&\stackrel{(33)}{=} P(\{w \in \Omega : X(w) \leq x\}) = P(X \leq x).
\end{aligned}$$

Below we will prove (33). Note that it is clear that

$$\cup_{n=1}^{\infty} A_n \subset \{w \in \Omega : X(w) \leq x\} \tag{34}$$

since for every $n \geq 1$,

$$A_n = \{w \in \Omega : X(w) \leq x\} \cap \{w \in \Omega : Y(w) \leq n\} \subset \{w \in \Omega : X(w) \leq x\}.$$

Therefore, to verify (33), it remains to show that

$$\{w \in \Omega : X(w) \leq x\} \subset \cup_{n=1}^{\infty} A_n. \tag{35}$$

To see that (35) holds, note that for $w \in \{w \in \Omega : X(w) \leq x\}$, there must exist a positive integer m such that

$$Y(w) \leq m,$$

otherwise we cannot have $Y(w) \in R$. Thus

$$\begin{aligned}
& w \in \{w \in \Omega : X(w) \leq x\} \\
& \Rightarrow w \in \Omega, X(w) \leq x \text{ and } Y(w) \leq m \text{ for some positive integer } m \\
& \Rightarrow w \in A_m \text{ for some positive integer } m \\
& \Rightarrow w \in \cup_{n=1}^{\infty} A_n.
\end{aligned}$$

Therefore, we have verified (35). Since both (34) and (35) hold, we have

$$\{w \in \Omega : X(w) \leq x\} = \cup_{n=1}^{\infty} A_n$$

and (33) holds.

33. (a) The set $\{(x, y) : f_{X,Y}(x, y) > 0\}$ is $(0, \infty) \times (0, \infty)$. Let

$$S_{U,V} = \{(\sqrt{x^2 + y^2}, \tan^{-1}(y/x)) : (x, y) \in (0, \infty) \times (0, \infty)\}.$$

We will first show that

$$S_{U,V} = (0, \infty) \times (0, \pi/2) \tag{36}$$

by verifying that

$$S_{U,V} \subset (0, \infty) \times (0, \pi/2) \tag{37}$$

and

$$(0, \infty) \times (0, \pi/2) \subset S_{U,V}. \tag{38}$$

To verify (37), suppose that $(u, v) \in S_{U,V}$. Then by the definition of $S_{U,V}$, there exists $(x, y) \in (0, \infty) \times (0, \infty)$ such that

$$\begin{cases} u = \sqrt{x^2 + y^2}; \\ v = \tan^{-1}(y/x). \end{cases} \tag{39}$$

Since (39) holds and $(x, y) \in (0, \infty) \times (0, \infty)$, we have $u = \sqrt{x^2 + y^2} > 0$ and $v = \tan^{-1}(y/x) \in (0, \pi/2)$, so $(u, v) \in (0, \infty) \times (0, \pi/2)$. The verification of (37) is complete.

To verify (38), suppose that $(u, v) \in (0, \infty) \times (0, \pi/2)$. In such case, solving (39) for $(x, y) \in (0, \infty) \times (0, \infty)$ gives

$$\begin{cases} x = u \cos(v); \\ y = u \sin(v). \end{cases} \tag{40}$$

Since $(u, v) \in (0, \infty) \times (0, \pi/2)$, the (x, y) in (40) satisfies $(x, y) \in (0, \infty) \times (0, \infty)$ and (39) holds for the (x, y) in (40). Therefore, $(u, v) \in S_{U,V}$. The verification of (38) is complete.

From (37) and (38), we have (36).

For $(u,v) \in (0,\infty)\times(0,\pi/2)$, let $J(u,v)$ be the determinant of the Jacobian matrix of x and y as functions of (u,v) given in (40), then

$$\begin{aligned}
J(u,v) &= \text{determinant of } \begin{pmatrix} \frac{\partial}{\partial u}u\cos(v) & \frac{\partial}{\partial v}u\cos(v) \\ \frac{\partial}{\partial u}u\sin(v) & \frac{\partial}{\partial v}u\sin(v) \end{pmatrix} \\
&= \text{determinant of } \begin{pmatrix} \cos(v) & -u\sin(v) \\ \sin(v) & u\cos(v) \end{pmatrix} \\
&= u\cos^2(v) - (-u\sin^2(v)) = u.
\end{aligned}$$

For $(u,v) \in R^2$, let

$$f_{U,V}(u,v) = \begin{cases} f_{X,Y}(u\cos(v), u\sin(v))|J(u,v)| & \text{if } (u,v) \in S_{U,V}; \\ 0 & \text{otherwise,} \end{cases}$$

$$= \begin{cases} f_{X,Y}(u\cos(v), u\sin(v))|u| & \text{if } (u,v) \in (0,\infty)\times(0,\pi/2); \\ 0 & \text{otherwise,} \end{cases}$$

then $f_{U,V}$ is a PDF of (U,V). Note that

$$\begin{aligned}
& (u,v)(0,\infty)\times(0,\pi/2) \\
\Rightarrow\ & (u\cos(v), u\sin(v)) \in (0,\infty)\times(0,\infty) \\
\Rightarrow\ & f_{X,Y}(u\cos(v), u\sin(v))|u| \\
& = ce^{-((u\cos(v))^2+(u\sin(v))^2)/2}u \\
& = ce^{-u^2/2}u,
\end{aligned}$$

so the expression of $f_{U,V}$ can be simplified as follows:

$$f_{U,V}(u,v) = \begin{cases} ce^{-u^2/2}u & \text{if } (u,v) \in (0,\infty)\times(0,\pi/2); \\ 0 & \text{otherwise.} \end{cases}$$

(b) For $v \in (-\infty,\infty)$, let

$$\begin{aligned}
f_V(v) &= \int_{-\infty}^{\infty} f_{U,V}(u,v)du \\
&= \begin{cases} \int_0^\infty ce^{-u^2/2}udu = c & \text{if } v \in (0,\pi/2); \\ 0 & \text{otherwise,} \end{cases} \\
&= cI_{(0,\pi/2)}(v)
\end{aligned}$$

then f_V is a PDF of V.

(c) Since $1 = \int_{-\infty}^{\infty} f_V(v)dv = \int_0^{\pi/2} cdv = c\pi/2$, we have $c = 2/\pi$.

34. (a) Let $M_{X,Y}$ be the joint MGF of (X,Y), then for $s<1$, $t<1$,

$$\begin{aligned}
M_{X,Y}(s,t) &= E(e^{sX+tY}) \\
&= \int_{R^2} e^{sx+ty} f_{X,Y}(x,y)d(x,y) \\
&= \int_{(0,\infty)\times(0,\infty)} e^{sx+ty}\left(\frac{1}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}y^{\beta-1}e^{-x-y}\right)d(x,y) \\
&= \frac{1}{\Gamma(\alpha)\Gamma(\beta)}\int_0^\infty\int_0^\infty e^{-(1-s)x}x^{\alpha-1}e^{-(1-t)y}y^{\beta-1}dxdy.
\end{aligned}$$

Note that for $t_0 < 1$ and $a > 0$,

$$
\begin{aligned}
& \int_0^\infty e^{-(1-t_0)x} x^{a-1} dx \\
(\text{ let } z = (1-t_0)x) \quad &= \int_0^\infty e^{-z} \left(\frac{z}{1-t_0}\right)^{a-1} \frac{1}{1-t_0} dz \\
&= (1-t_0)^{-a} \int_0^\infty e^{-z} z^{a-1} dz = (1-t_0)^{-a}\Gamma(a).
\end{aligned}
$$

Apply the above reslut with $(a, t_0) = (\alpha, s)$, (β, t), then we have

$$\int_0^\infty e^{-(1-s)x} x^{\alpha-1} dx = (1-s)^{-\alpha}\Gamma(\alpha) \text{ for } s < 1$$

and

$$\int_0^\infty e^{-(1-t)y} y^{\beta-1} dy = (1-t)^{-\beta}\Gamma(\beta) \text{ for } t < 1.$$

Thus the expression of $M_{X,Y}(s,t)$ can be simplified as follows: for $s < 1$, $t < 1$,

$$
\begin{aligned}
M_{X,Y}(s,t) &= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \int_0^\infty \int_0^\infty e^{-(1-s)x} x^{\alpha-1} e^{-(1-t)y} y^{\beta-1} dx dy \\
&= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \int_0^\infty e^{-(1-t)y} y^{\beta-1} (1-s)^{-\alpha}\Gamma(\alpha) dy \\
&= \frac{1}{\Gamma(\beta)} (1-s)^{-\alpha} \int_0^\infty e^{-(1-t)y} y^{\beta-1} dy \\
&= \frac{1}{\Gamma(\beta)} (1-t)^{-\alpha} (1-t)^{-\beta}\Gamma(\beta) \\
&= (1-s)^{-\alpha}(1-t)^{-\beta}.
\end{aligned}
$$

It is clear that the function M given in the problem is the same as $M_{X,Y}$, which is the MGF of (X,Y).

(b) Let M_X be the MGF of X, then $M_X(s) = M_{X,Y}(s,0) = (1-s)^{-\alpha}$ for $s < 1$.

(c) To find $E(XY)$, we will first compute $\frac{\partial^2}{\partial t \partial s} M_{X,Y}(s,t)$ since

$$E(XY) = \frac{\partial^2}{\partial s \partial t} M_{X,Y}(s,t)\bigg|_{(s,t)=(0,0)}.$$

Note that

$$
\begin{aligned}
\frac{\partial}{\partial s} M_{X,Y}(s,t) &= \frac{\partial}{\partial s} (1-s)^{-\alpha}(1-t)^{-\beta} \\
&= (1-t)^{-\beta} \frac{d}{ds} (1-s)^{-\alpha} \\
&= (1-t)^{-\beta} \alpha (1-s)^{-\alpha-1},
\end{aligned}
$$

so

$$
\begin{aligned}
\frac{\partial^2}{\partial t \partial s} M_{X,Y}(s,t) &= \frac{\partial}{\partial t} (1-t)^{-\beta} \alpha (1-s)^{-\alpha-1} \\
&= \alpha (1-s)^{-\alpha-1} \frac{d}{dt} (1-t)^{-\beta} \\
&= \alpha (1-s)^{-\alpha-1} \beta (1-t)^{-\beta-1}.
\end{aligned}
$$

Therefore,

$$E(XY) = \alpha(1-s)^{-\alpha-1}\beta(1-t)^{-\beta-1}\big|_{(s,t)=(0,0)} = \alpha\beta$$

and

$$E(X) = \frac{\partial}{\partial s} M_{X,Y}(s,t)\bigg|_{(s,t)=(0,0)} = (1-t)^{-\beta}\alpha(1-s)^{-\alpha-1}\big|_{(s,t)=(0,0)} = \alpha.$$

Remark. To find $E(X)$, we can also compute M_X': the derivative of the MGF of X. From the solution to Part (b), we have $M_X(s) = (1-s)^{-\alpha}$, so

$$M_X'(s) = \frac{d}{ds}(1-s)^{-\alpha} = \alpha(1-s)^{-\alpha-1}$$

and $E(X) = M_X'(0) = \alpha$.

35. From the solution to Problem 34 (a), the joint MGF of (X, Y) is $M_{X,Y}$, where

$$M_{X,Y}(s,t) = (1-s)^{-\alpha}(1-t)^{-\beta} \text{ for } s < 1, t < 1. \tag{41}$$

For $a > 0$, let M_a be the MGF of a random variable whose distribution is $\Gamma(a, 1)$, then from the solution to Problem 34 (b), we have for every $a > 0$,

$$M_a(t) = (1-t)^{-a} \text{ for } t < 1. \tag{42}$$

Let M_{X+Y} be the MGF of $X + Y$, then for $t < 1$,

$$\begin{aligned} M_{X+Y}(t) &= E(e^{t(X+Y)}) \\ &= M_{X,Y}(t,t) \\ &\overset{(41)}{=} (1-t)^{-\alpha}(1-t)^{-\beta} \\ &= (1-t)^{-(\alpha+\beta)}. \end{aligned}$$

Since M_{X+Y} is the same as the MGF M_a in (42) with $a = \alpha + \beta$ on $(-\infty, 1)$, the distribution of $(X + Y)$ is $\Gamma(\alpha + \beta, 1)$.

36. (a) Let $A = \{(x, y) : (x + 2y) > 0\}$, then

$$\begin{aligned} P((X+2Y)>0) &= \int_A f_{X,Y}(x,y)d(x,y) \\ &= \int_{A\cap S} c\,d(x,y) \\ &= \int_{\{(x,y):0<(x+2y)<2 \text{ and } -2<(x-2y)<2\}} c\,d(x,y) \\ &\overset{u=x+2y,v=x-2y}{=} \int_{(0,2)\times(-2,2)} c|J(u,v)|d(u,v), \end{aligned}$$

where

$$\begin{aligned} J(u,v) &= \text{deteminant of } \begin{pmatrix} \frac{\partial}{\partial u}\frac{u+v}{2} & \frac{\partial}{\partial v}\frac{u+v}{2} \\ \frac{\partial}{\partial u}\frac{u-v}{4} & \frac{\partial}{\partial v}\frac{u-v}{4} \end{pmatrix} \\ &= \text{deteminant of } \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & -\frac{1}{4} \end{pmatrix} \\ &= \frac{1}{2}\left(-\frac{1}{4}\right) - \frac{1}{2}\cdot\frac{1}{4} \\ &= -\frac{1}{4}, \end{aligned} \tag{43}$$

so

$$\begin{aligned} P((X+2Y)>0) &= \int_{(0,2)\times(-2,2)} \frac{c}{4} d(u,v) \\ &= \int_0^2 \int_{-2}^2 \frac{c}{4} dv du \\ &= 2c. \end{aligned}$$

(b) Since

$$1 = \int_{R^2} f_{X,Y}(x,y) d(x,y) = \int c I_S(x,y) d(x,y),$$

we have

$$\begin{aligned} \frac{1}{c} &= \int_S 1 d(x,y) \\ &= \int_{\{(x,y): -2<(x+2y)<2 \text{ and } -2<(x-2y)<2\}} 1 d(x,y) \\ &\overset{u=x+2y, v=x-2y}{=} \int_{(-2,2)\times(-2,2)} |J(u,v)| d(u,v) \\ &\overset{(43)}{=} \int_{(-2,2)\times(-2,2)} \frac{1}{4} d(u,v) \\ &= \int_{-2}^2 \int_{-2}^2 \frac{1}{4} dv du = 4, \end{aligned}$$

so $c = 1/4$.

37. (a) To find $E(X|Y)$, we will first find a version of the conditional PDF of X given Y. For $y \in (-\infty, \infty)$, let

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx,$$

$R_Y = \{y : f_Y(y) > 0\}$, and

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \tag{44}$$

for $y \in R_Y$, then $\{f_{X|Y=y} : y \in R_Y\}$ is a version of the conditional PDF of X given Y.

To compute f_Y, note that the region S is the interior of the parrallelogram with vertices $(-2,0)$, $(0,1)$, $(2,0)$ and $(0,-1)$, so for $(x,y) \in S$, we have $y \in (-1,1)$ and $f_Y(y) = 0$ for $y \notin (-1,1)$. In addition, given $y \in (-1,1)$,

$$\begin{aligned} & (x,y) \in S \\ & \Leftrightarrow x \in \begin{cases} (-2-2y, 2+2y) & \text{if } -1 < y \le 0; \\ (-2+2y, 2-2y) & \text{if } 0 < y < 1, \end{cases} \end{aligned} \tag{45}$$

so

$$f_Y(y) = \begin{cases} 0 & \text{if } y \ge 1 \text{ or } y \le -1; \\ \int_{-2-2y}^{2+2y} c dx = 4c(1+y) & \text{if } -1 < y \le 0; \\ \int_{-2+2y}^{2-2y} c dx = 4c(1-y) & \text{if } 0 < y < 1. \end{cases} \tag{46}$$

From (46), it is clear that $R_Y = \{y : f_Y(y) > 0\} = (-1, 1)$.
Next, we compute $f_{X|Y=y}$ using (44) for $y \in (-1, 1)$. From (45) and (46), we have that for $y \in (-1, 0]$,

$$f_{X|Y=y}(x) = \begin{cases} \frac{c}{4c(1+y)} = \frac{1}{4(1+y)} & \text{if } x \in (-2-2y, 2+2y); \\ 0 & \text{otherwise,} \end{cases}$$

and for $y \in (0, 1)$,

$$f_{X|Y=y}(x) = \begin{cases} \frac{c}{4c(1-y)} = \frac{1}{4(1-y)} & \text{if } x \in (-2+2y, 2-2y); \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, for $y \in (-1, 1)$,

$$\begin{aligned} E(X|Y=y) &= \int_{-\infty}^{\infty} x f_{X|Y=y}(x)dx \\ &= \begin{cases} \int_{-2-2y}^{2+2y} \frac{x}{4(1+y)}dx = 0 & \text{if } -1 < y \leq 0; \\ \int_{-2+2y}^{2-2y} \frac{x}{4(1-y)}dx = 0 & \text{if } 0 < y < 1, \end{cases} \\ &= 0, \end{aligned}$$

which implies that $E(X|Y) = 0$.

(b) For $y \in (-1, 1)$,

$$\begin{aligned} Var(X|Y=y) &= E(X^2|Y=y) - (E(X|Y=y))^2 \\ &= E(X^2|Y=y) - 0^2 \\ &= \int_{-\infty}^{\infty} x^2 f_{X|Y=y}(x)dx \\ &= \begin{cases} \int_{-2-2y}^{2+2y} \frac{x^2}{4(1+y)}dx = \frac{4(1+y)^2}{3} & \text{if } -1 < y \leq 0; \\ \int_{-2+2y}^{2-2y} \frac{x^2}{4(1-y)}dx = \frac{4(1-y)^2}{3} & \text{if } 0 < y < 1; \end{cases} \end{aligned}$$

so

$$Var(X|Y) = \begin{cases} 4(1+Y)^2/3 & \text{if } -1 < Y \leq 0; \\ 4(1-Y)^2/3 & \text{if } 0 < Y < 1. \end{cases}$$

38. (a) We first compute the conditional probabilities $P(X = x|Y = -3)$ to find $E(X|Y = -3)$ and $Var(X|Y = -3)$. Since

$$\begin{aligned} P(Y=-3) &= P((X,Y) = (0,-3)) + P((X,Y) = (1,-3)) \\ &= 0.4 + 0.1 = 0.5, \end{aligned}$$

$$P(X = 0|Y = -3) = \frac{P((X,Y) = (0,-3))}{P(Y=-3)} = \frac{0.4}{0.5} = 0.8,$$

and

$$P(X = 1|Y = -3) = \frac{P((X,Y) = (1,-3))}{P(Y=-3)} = \frac{0.1}{0.5} = 0.2,$$

we have

$$E(X|Y=-3) = 0 \times P(X=0|Y=-3) + 1 \times P(X=1|Y=-3) = 0 \cdot 0.8 + 1 \cdot 0.2 = 0.2$$

and

$$\begin{aligned} Var(X|Y=-3) &= E(X^2|Y=-3) - (E(X|Y=-3))^2 \\ &= 0^2 \times P(X=0|Y=-3) + 1^2 \times P(X=1|Y=-3) - (0.2)^2 \\ &= 0 \cdot 0.8 + 1 \cdot 0.2 - 0.04 = 0.16. \end{aligned}$$

Next, we will find $E(X|Y=2)$ and $Var(X|Y=2)$. Since

$$P(Y=2) = P((X,Y)=(1,2)) = 0.5,$$

we have

$$P(X=1|Y=2) = \frac{P((X,Y)=(1,2))}{P(Y=2)} = \frac{0.5}{0.5} = 1,$$

and

$$P(X=x|Y=2) = \frac{P((X,Y)=(x,2))}{P(Y=2)} = 0$$

for $x \neq 1$, we have

$$E(X|Y=2) = 1 \times P(X=1|Y=2) = 1$$

and

$$\begin{aligned} Var(X|Y=2) &= E(X^2|Y=2) - (E(X|Y=2))^2 \\ &= 1^2 \times \underbrace{P(X=0|Y=2)}_{=1} - (1)^2 = 0. \end{aligned}$$

In summary, we have

$$E(X|Y) = 0.2I_{\{-3\}}(Y) + I_{\{2\}}(Y)$$

and

$$Var(X|Y) = 0.16I_{\{-3\}}(Y).$$

(b) To find $E(Var(X|Y))$, note that $Var(X|Y)$ can be 0.16 or 0 with probabilities $P(Y=-3)$ and $P(Y=2)$ respectively, so

$$\begin{aligned} E(Var(X|Y)) &= 0.16 \times P(Y=-3) + 0 \times P(Y=2) \\ &= 0.16 \times 0.5 = 0.08. \end{aligned}$$

To find $Var(E(X|Y))$, note that $E(X|Y)$ can be 0.2 or 1 with probabilities $P(Y=-3)$ and $P(Y=2)$ respectively, so

$$\begin{aligned} E(E(X|Y))^2 &= (0.2)^2 \times P(Y=-3) + 1^2 \times P(Y=2) \\ &= 0.04 \times 0.5 + 1 \times 0.5 = 0.52 \end{aligned}$$

and

$$\begin{aligned} E(E(X|Y)) &= (0.2) \times P(Y=-3) + 1 \times P(Y=2) \\ &= 0.2 \times 0.5 + 1 \times 0.5 = 0.6, \end{aligned}$$

which gives

$$\begin{aligned} Var(E(X|Y)) &= E(E(X|Y))^2 - (E(E(X|Y)))^2 \\ &= 0.52 - (0.6)^2 = 0.16. \end{aligned}$$

To find $Var(X)$ using the joint PMF of (X, Y), note that

$$E(X^2) = 1^2 \times 0.5 + 0^2 \times 0.4 + 1^2 \times 0.1 = 0.6$$

and

$$E(X) = 1 \times 0.5 + 0 \times 0.4 + 1 \times 0.1 = 0.6,$$

so $Var(X) = E(X^2) - (E(X))^2 = 0.6 - (0.6)^2 = 0.24$.
From the above calculation, we have $Var(X) = 0.24$ and

$$E(Var(X|Y)) + Var(E(X|Y)) = 0.08 + 0.16 = 0.24,$$

so the equality

$$Var(X) = E(Var(X|Y)) + Var(E(X|Y))$$

holds for the (X, Y) in this problem.

39. Let

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dx$$

for $y \in R$. Let $R_Y = \{y : f_Y(y) > 0\}$ and

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

for $y \in R_Y$, then for $y \in R_Y$, we have

$$E(u(X, Y)|Y = y) = \int_{-\infty}^{\infty} u(x, y)f_{X|Y=y}(x)dx. \tag{47}$$

(a) For $y \in R_Y$,

$$\begin{aligned} E(g(X, Y)h(Y)|Y = y) &\overset{(47)}{=} \int_{-\infty}^{\infty} g(x, y)f_{X|Y=y}(x)dx \\ &= h(y)\int_{-\infty}^{\infty} g(x, y)f_{X|Y=y}(x)dx \\ &\overset{(47)}{=} h(y)E(g(X, Y)|Y = y), \end{aligned}$$

so $E(g(X, Y)h(Y)|Y) = h(Y)E(g(X, Y)|Y)$.

(b) For $y \in R_Y$,

$$\begin{aligned} &E(g_1(X, Y) + g_2(X, Y)|Y = y) \\ \overset{(47)}{=}\ &\int_{-\infty}^{\infty} (g_1(x, y) + g_2(x, y))f_{X|Y=y}(x)dx \\ =\ &\int_{-\infty}^{\infty} g_1(x, y)f_{X|Y=y}(x)dx + \int_{-\infty}^{\infty} g_2(x, y)f_{X|Y=y}(x)dx \\ \overset{(47)}{=}\ &E(g_1(X, Y)|Y = y) + E(g_2(X, Y)|Y = y), \end{aligned}$$

so

$$E(g_1(X, Y) + g_2(X, Y)|Y) = E(g_1(X, Y)|Y) + E(g_2(X, Y)|Y).$$

40. (a) For $(x, y) \in R^2$, let

$$f_{X,Y}(x, y) = f_{0,1}(y) f_{1+2y,1}(x),$$

then $f_{X,Y}$ is a PDF of (X, Y), and

$$\begin{aligned}
f_{X,Y}(x, y) &= \frac{1}{\sqrt{2\pi}} e^{-0.5y^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-0.5(x-(1+2y))^2} \\
&= \frac{1}{2\pi} e^{-0.5((2y-(x-1))^2+y^2)} \\
&= \frac{1}{2\pi} e^{-2.5(y-0.4(x-1))^2 - 0.1(x-1)^2} \\
&= \frac{1}{\sqrt{2\pi/5}} e^{-2.5(y-0.4(x-1))^2} \frac{1}{\sqrt{2\pi \cdot 5}} e^{-0.1(x-1)^2} \\
&= f_{1,\sqrt{5}}(x) f_{0.4(x-1),\sqrt{1/5}}(y)
\end{aligned}$$

for $(x, y) \in R^2$. Since

$$\begin{aligned}
\int_{-\infty}^{\infty} f_{X,Y}(x, y) dy &= \int_{-\infty}^{\infty} f_{1,\sqrt{5}}(x) f_{0.4(x-1),\sqrt{1/5}}(y) dy \\
&= f_{1,\sqrt{5}}(x)
\end{aligned}$$

for $x \in R$, $f_{1,\sqrt{5}}$ is a PDF of X.

(b) From Part (a), we have for $x \in R$,

$$\int_{-\infty}^{\infty} f_{X,Y}(x, y) dy = f_{1,\sqrt{5}}(x)$$

and

$$\frac{f_{X,Y}(x, y)}{f_{1,\sqrt{5}}(x)} = \frac{f_{1,\sqrt{5}}(x) f_{0.4(x-1),\sqrt{1/5}}(y)}{f_{1,\sqrt{5}}(x)} = f_{0.4(x-1),\sqrt{1/5}}(y)$$

for $y \in R$. Thus $\{f_{0.4(x-1),\sqrt{1/5}} : x \in R\}$ is a version of the conditional PDF of Y given X and

$$E(Y|X) = \int_{-\infty}^{\infty} y f_{0.4(x-1),\sqrt{1/5}}(y) dy \Big|_{x=X} = 0.4(X - 1).$$

Here we have used the result that for $\sigma > 0$ and $\mu \in R$,

$$\int_{-\infty}^{\infty} y f_{\mu,\sigma}(y) dy = \mu$$

since $f_{\mu,\sigma}$ is a PDF of $N(\mu, \sigma^2)$.

(c) X and Y are not independent. To see this, note that if X and Y are independent, then $E(Y|X) = E(Y)$ is a constant, which implies that $Var(E(Y|X)) = 0$. From Part (b), we have $E(Y|X) = 0.4(X - 1)$, which is a random variable of variance $(0.4)^2 Var(X) = (0.4)^2 \cdot (\sqrt{5})^2 > 0$, so X and Y cannot be independent.

41. Let p_X and p_Y be PMF's of X and Y respectively, and let

$$p_{X,Y}(x, y) = p_X(x) p_Y(y)$$

for $(x, y) \in R^2$. Since X and Y are independent, $p_{X,Y}$ is a joint PMF of (X, Y). Let $R_X = \{x : p_X(x) > 0\}$ and $R_Y = \{y : p_Y(y) > 0\}$, then

$$\begin{aligned}
E(u(X)v(Y) &= \sum_{y \in R_Y} \sum_{x \in R_X} u(x)v(y)p_{X,Y}(x, y) \\
&= \sum_{y \in R_Y} \sum_{x \in R_X} u(x)v(y)p_X(x)p_Y(y) \\
&= \sum_{y \in R_Y} v(y)p_Y(y) \underbrace{\left(\sum_{x \in R_X} u(x)p_X(x) \right)}_{E(u(X))} \\
&= E(u(X)) \underbrace{\sum_{y \in R_Y} v(y)p_Y(y)}_{E(v(Y))} \\
&= E(u(X))E(v(Y)).
\end{aligned}$$

42. Let $M_{X+Y,X-Y}$ be the joint MGF of $(X + Y, X - Y)$, then

$$\begin{aligned}
& M_{X+Y,X-Y}(t_1, t_2) \\
= \; & E(e^{t_1(X+Y)+t_2(X-Y)}) \\
= \; & E(e^{(t_1+t_2)X+(t_1-t_2)Y}) \\
= \; & M_{X,Y}(t_1 + t_2, t_1 - t_2) \\
\overset{(41)}{=} \; & (1 - (t_1 + t_2))^{-\alpha}(1 - (t_1 - t_2))^{-\beta} \qquad (48)
\end{aligned}$$

for (t_1, t_2) such that $t_1 + t_2 < 1$ and $t_1 - t_2 < 1$.

Let M_{X-Y} and M_{X+Y} be the marginal MGFs of $X - Y$ and $X + Y$ respectively, then from the joint MGF $M_{X+Y,X-Y}$ given in (48),

$$M_{X-Y}(t_2) = M_{X+Y,X-Y}(0, t_2) = (1 - t_2)^{-\alpha}(1 + t_2)^{-\beta}$$

for $t_2 \in (-1, 1)$ and

$$M_{X+Y}(t_1) = M_{X+Y,X-Y}(t_1, 0) = (1 - t_1)^{-(\alpha+\beta)}$$

for $t_1 \in (-\infty, 1)$. Thus

$$M_{X+Y}(t_1)M_{X-Y}(t_2) = (1 - t_1)^{-(\alpha+\beta)}(1 - t_2)^{-\alpha}(1 + t_2)^{-\beta} \qquad (49)$$

for $(t_1, t_2) \in (-\infty, 1) \times (-1, 1)$. It is clear that from (48) and (49), $M_{X+Y,X-Y}(t_1, t_2)$ and $M_{X+Y}(t_1)M_{X-Y}(t_2)$ are not the same for all (t_1, t_2) in $\{(t_1, t_2) : t_1 + t_2 < 1 \text{ and } t_1 - t_2 < 1\} \cap (-\infty, 1) \times (-1, 1)$. Therefore, $X + Y$ and $X - Y$ are not independent.

43. X and Y are not independent. To see this, note that for discrete random variables X and Y that are independent of each other, we have that for y such that $P(Y = y) > 0$,

$$\begin{aligned}
P(X = x|Y = y) &= \frac{P((X, Y) = (x, y))}{P(Y = y)} \\
&= \frac{P(X = x)P(Y = y)}{P(Y = y)} = P(X = x)
\end{aligned}$$

for all x. In Problem 38, if X and Y are independent, then we must have

$$P(X=0|Y=2)=P(X=0)=P(X=0|Y=-3).$$

However, from the calculation in the solution to Problem 40, we have $P(X=0|Y=2)=0$ and $P(X=0|Y=-3)=0.8$. Since $P(X=0|Y=2)\neq P(X=0|Y=-3)$, X and Y cannot be independent.

44. For $x\in\{x_1,\dots,x_m\}$, let

$$p(x)=P(X=x|Y=y_1),$$

then by assumption, $P(X=x|Y=y_j)=p(x)$ for $j\in\{1,\dots,n\}$. Thus for $x\in\{x_1,\dots,x_m\}$,

$$\begin{aligned} P(X=x) &= \sum_{j=1}^{n} P(X=x|Y=y_j)P(Y=y_j) \\ &= \sum_{j=1}^{n} p(x)P(Y=y_j) \\ &= p(x)\underbrace{\sum_{j=1}^{n} P(Y=y_j)}_{1}=p(x), \end{aligned}$$

and for $x\in\{x_1,\dots,x_m\}$, $y\in\{y_1,\dots,y_n\}$,

$$P(X=x|Y=y)=p(x)=P(X=x),$$

which implies that $P(X=x \text{ and } Y=y)/P(Y=y)=P(X=x)$ and

$$P(X=x \text{ and } Y=y)=P(X=x)P(Y=y)$$

for all $x\in\{x_1,\dots,x_m\}$, $y\in\{y_1,\dots,y_n\}$. Therefore, X and Y are independent.

45. (a) Let $Z=(Z_1,Z_2,Z_3)^T$ and

$$f_Z(z)=\left(\frac{1}{\sqrt{2\pi}}\right)^3 e^{-\frac{1}{2}z^Tz}$$

for $z=(z_1,z_2,z_3)^T\in R^3$, then f_Z is a PDF of Z since Z_1, Z_2, Z_3 are IID $N(0,1)$ random variables. Let $Y=(Y_1,Y_2,Y_3)^T$ and

$$A=\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}^{-1}=\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix},$$

then $Z=AY$. For $y=(y_1,y_2,y_3)^T\in R^3$, let

$$\begin{aligned} f_Y(y) &= f_Z(Ay)|\det(A)| \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^3 e^{-\frac{1}{2}(Ay)^TAy}|\det(A)|, \end{aligned}$$

where $\det(A)$ is the determinant of A, then f_Y is a PDF of $Y = (Y_1, Y_2, Y_3)^T$. Since $\det(A) = 1$, we have

$$f_Y(y) = \left(\frac{1}{\sqrt{2\pi}}\right)^3 e^{-\frac{1}{2}y^T B y}$$

for $y \in R^3$, where

$$B = A^T A = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}.$$

(b) Since

$$Cov(Y_1, Y_3) = Cov(Z_1 + Z_2 + Z_3, Z_3) = Var(Z_3) = 1 \neq 0,$$

Y_1 and Y_3 are not independent. Thus Y_1 and (Y_2, Y_3) are not independent.

46. We will show that $\bar{X}$ and Y are independent using the joint MGF of $\bar{X}$ and Y. Let $M_{\bar{X},Y}$ be the joint MGF of $\bar{X}$ and Y and let $M_{X_1,\ldots,X_n}$ be the joint MGF of $(X_1, \ldots, X_n)$, then for $(s_1, \ldots, s_n) \in R^n$,

$$\begin{aligned} M_{X_1,\ldots,X_n}(s_1, \ldots, s_n) &= E(e^{s_1 X_1}) \cdots E(e^{s_n X_n}) \\ &= \prod_{i=1}^{n} e^{\mu s_i + 0.5\sigma^2 s_i^2} \\ &= e^{\mu(\sum_{i=1}^n s_i)} e^{0.5\sigma^2(\sum_{i=1}^n s_i^2)}, \end{aligned}$$

and for $s \in (-\infty, \infty)$ and $\boldsymbol{t} = (t_1, \ldots, t_n)^T \in R^n$,

$$\begin{aligned} M_{\bar{X},Y}(s, \boldsymbol{t}) &= Ee^{s\bar{X} + \sum_{i=1}^n t_i(X_i - \bar{X})} \\ (\bar{t} = \sum_{i=1}^{n} t_i/n) &= Ee^{(s - n\bar{t})\bar{X} + \sum_{i=1}^n t_i X_i} \\ &= M_{X_1,\ldots,X_n}(t_1 + (s - n\bar{t})/n, \ldots, t_n + (s - n\bar{t})/n) \\ &= e^{\mu(\sum_{i=1}^n (t_i + (s - n\bar{t})/n))} e^{0.5\sigma^2(\sum_{i=1}^n (t_i + (s - n\bar{t})/n))^2)} \\ &= e^{s\mu} e^{0.5\sigma^2((\sum_{i=1}^n t_i^2) + (s^2/n) - n\bar{t}^2)} \\ &= e^{s\mu} e^{0.5\sigma^2(s^2/n)} e^{0.5\sigma^2((\sum_{i=1}^n t_i^2) - n\bar{t}^2)}. \end{aligned} \tag{50}$$

Let $M_{\bar{X}}$ and M_Y be the MGFs of $\bar{X}$ and Y respectively, then for $s \in (-\infty, \infty)$,

$$M_{\bar{X}}(s) = M_{\bar{X},Y}(s, \boldsymbol{t})|_{\boldsymbol{t}=(0,\ldots,0)^T} = e^{s\mu} e^{0.5\sigma^2(s^2/n)}$$

and for $\boldsymbol{t} = (t_1, \ldots, t_n)^T \in R^n$,

$$M_Y(\boldsymbol{t}) = M_{\bar{X},Y}(0, \boldsymbol{t}) = e^{0.5\sigma^2((\sum_{i=1}^n t_i^2) - n\bar{t}^2)}.$$

Therefore, for $s \in (-\infty, \infty)$ and $\boldsymbol{t} = (t_1, \ldots, t_n)^T \in R^n$,

$$\begin{aligned} M_{\bar{X}}(s) M_Y(\boldsymbol{t}) &= e^{s\mu} e^{0.5\sigma^2(s^2/n)} e^{0.5\sigma^2((\sum_{i=1}^n t_i^2) - n\bar{t}^2)} \\ &\overset{(50)}{=} M_{\bar{X},Y}(s, \boldsymbol{t}), \end{aligned}$$

so $\bar{X}$ and Y are independent.

47. For $t < 1$,

$$\begin{aligned} E(e^{tZ_1^2/2}) &= \int_{-\infty}^{\infty} e^{0.5tz^2} \frac{1}{\sqrt{2\pi}} e^{-0.5z^2} dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-0.5(1-t)z^2} dz \\ &= \frac{\sqrt{2\pi(1/1-t)}}{\sqrt{2\pi}} = (1-t)^{-1/2}, \end{aligned}$$

so

$$\begin{aligned} E(e^{t(U/2)}) &= E(e^{t(Z_1^2+\cdots+Z_m^2)/2}) \\ &= [E(e^{tZ_1^2/2})]^m \qquad \text{(since } Z_1, \ldots, Z_m \text{ are IID)} \\ &= ((1-t)^{-1/2})^m = (1-t)^{-m/2} \end{aligned}$$

for $t < 1$, and the MGF of $U/2$ is the same as the MGF of $\Gamma(a,1)$ given in (42) in the solution to Problem 35 with $a = m/2$, so $U/2 \sim \Gamma(m/2, 1)$.

48. For $i \in \{1, \ldots, n\}$, $j \in \{1, \ldots, k\}$, let

$$(W_{i,1}, \ldots, W_{i,m})$$

be the i-th row of W and let

$$(b_{1,j}, \ldots, b_{m,j})^T$$

be the j-th column of B. Then, for $i \in \{1, \ldots, n\}$, $j \in \{1, \ldots, k\}$, the (i,j)-th element of WB is

$$W_{i,1}b_{1,j} + \cdots + W_{i,m}b_{m,j},$$

so the (i,j)-th element of $E(WB)$

$$\begin{aligned} &E(W_{i,1}b_{1,j} + \cdots + W_{i,m}b_{m,j}) \\ &= E(W_{i,1})b_{1,j} + \cdots + E(W_{i,m})b_{m,j}. \end{aligned}$$

Moreover, the (i,j)-th element of $E(W)B$ is also

$$E(W_{i,1})b_{1,j} + \cdots + E(W_{i,m})b_{m,j}$$

since

$$(E(W_{i,1}), \ldots, E(W_{i,m}))$$

is the i-th row of $E(W)$ and

$$(b_{1,j}, \ldots, b_{m,j})^T$$

is the j-th column of B. We have verified that the (i,j)-th element of $E(WB)$ and the (i,j)-th element of $E(W)B$ are both equal to

$$E(W_{i,1})b_{1,j} + \cdots + E(W_{i,m})b_{m,j}$$

for $i \in \{1, \ldots, n\}$, $j \in \{1, \ldots, k\}$. Thus $E(WB) = E(W)B$.

49. Since

$$AX - E(AX) = AX - AE(X) = A[X - E(X)],$$

the covariance matrix of AX is

$$\begin{aligned} E([AX - E(AX)][AX - E(AX)]^T) &= E(A[X - E(X)][X - E(X)]^T A^T) \\ &= AE([X - E(X)][X - E(X)]^T A^T) \\ &= A \underbrace{E([X - E(X)][X - E(X)]^T)}_{\text{covariance matrix of } X} A^T \\ &= A\Sigma A^T. \end{aligned}$$

50. Let $X = (X_1, \ldots, X_m)^T$ and $Y = (Y_1, \ldots, Y_n)^T$. Let $\mu_X = E(X)$, $\mu_Y = E(Y)$, and let Σ_X and Σ_Y be the covariance matrices of X and Y respectively, then the mean vector of $(X^T, Y^T)^T$ is $(\mu_X^T, \mu_Y^T)^T$ and the covariance matrix of $(X^T, Y^T)^T$ is

$$\begin{pmatrix} \Sigma_X & O_{m\times n} \\ O_{n\times m} & \Sigma_Y \end{pmatrix},$$

where $O_{a\times b}$ denotes the $a \times b$ matrix of 0's for positive integers a and b.

Let $M_{X,Y}$, M_X and M_Y be the MGFs of $(X^T, Y^T)^T$, X and Y respectively. Then for $s = (s_1, \ldots, s_m)^T \in R^m$ and $t = (t_1, \ldots, t_n)^T \in R^n$,

$$\begin{aligned} \ln M_{X,Y}(s,t) &= \begin{pmatrix} s^T & t^T \end{pmatrix} \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} + \frac{1}{2} \begin{pmatrix} s^T & t^T \end{pmatrix} \begin{pmatrix} \Sigma_X & O_{m\times n} \\ O_{n\times m} & \Sigma_Y \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix} \\ &= s^T \mu_X + t^T \mu_Y + \frac{1}{2} \begin{pmatrix} \underbrace{s^T \Sigma_X}_{1\times m} & \underbrace{t^T \Sigma_Y}_{1\times n} \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix} \\ &= s^T \mu_X + t^T \mu_Y + \frac{1}{2} \left(s^T \Sigma_X s + t^T \Sigma_Y t \right) \\ &= \ln M_{X,Y}(s,t)\big|_{t=(0,\ldots 0)^T} + \ln M_{X,Y}(s,t)\big|_{s=(0,\ldots 0)^T} \\ &= \ln M_X(s) + \ln M_Y(t), \end{aligned}$$

so

$$M_{X,Y}(s,t) = M_X(s) M_Y(t)$$

for all $s = (s_1, \ldots, s_m)^T \in R^m$ and $t = (t_1, \ldots, t_n)^T \in R^n$, which implies that X and Y are independent.

51. (a) Since the distribution of $(X, Y, Z)^T$ is a multivariate normal distribution and for a constant $b \in R$,

$$\begin{pmatrix} X \\ Y - bX \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -b & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix},$$

the distribution of $(X, Y - bX)^T$ is a multivariate normal distribution. In such case,

$$\begin{aligned} &Y - bX \text{ and } X \text{ are independent} \\ &\Leftrightarrow Cov(X, Y - bX) = 0 \\ &\Leftrightarrow Cov(X, Y) - bCov(X, X) = 0. \end{aligned}$$

Take

$$b = \frac{Cov(X,Y)}{var(X)} = \frac{2}{10} = \frac{1}{5},$$

then $Y - bX$ and X are independent.

(b) Since the distribution of $(X, Y, Z)^T$ is a multivariate normal distribution and for constants c, $d \in R$,

$$\begin{pmatrix} X \\ Y \\ Z - cY - dX \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -d & -c & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix},$$

the distribution of $(X, Y, Z - cY - dX)^T$ is a multivariate normal distribution. In such case,

$$\begin{aligned} & Z - cY - dX \text{ and } (X,Y) \text{ are independent} \\ & \Leftrightarrow Cov(X, Z - cY - dX) = 0 \text{ and } Cov(Y, Z - cY - dX) = 0 \\ & \Leftrightarrow \begin{pmatrix} Cov(X,Z) \\ Cov(Y,Z) \end{pmatrix} = \begin{pmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{pmatrix} \begin{pmatrix} d \\ c \end{pmatrix}. \end{aligned}$$

Take

$$\begin{aligned} \begin{pmatrix} d \\ c \end{pmatrix} &= \begin{pmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{pmatrix}^{-1} \begin{pmatrix} Cov(X,Z) \\ Cov(Y,Z) \end{pmatrix} \\ &= \begin{pmatrix} 10 & 2 \\ 2 & 16 \end{pmatrix}^{-1} \begin{pmatrix} 5 \\ 3 \end{pmatrix} \\ &= \frac{1}{10 \cdot 16 - 2 \cdot 2} \begin{pmatrix} 16 & -2 \\ -2 & 10 \end{pmatrix} \begin{pmatrix} 5 \\ 3 \end{pmatrix} \\ &= \begin{pmatrix} 74/156 \\ 20/156 \end{pmatrix} = \begin{pmatrix} 37/78 \\ 5/39 \end{pmatrix}, \end{aligned}$$

then $Z - cY - dX$ and (X, Y) are independent.

52. For $\boldsymbol{z} = (z_1, \ldots, z_n) \in R^n$, let $f_{Z_1,\ldots,Z_n}(\boldsymbol{z}) = f_{0,1}(z_1) \cdots f_{0,1}(z_n)$ for $(z_1, \ldots, z_n) \in R^n$, where $f_{0,1}$ is the $N(\mu, \sigma^2)$ PDF $f_{\mu,\sigma}$ defined in Problem 27 with $\mu = 0$ and $\sigma = 1$, then $f_{Z_1,\ldots,Z_n}$ is a PDF of $(Z_1, \ldots, Z_n)$. Let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^T$. For $\boldsymbol{y} = (y_1, \ldots, y_n)^T \in R^n$, let

$$(z_1(\boldsymbol{y}), \ldots, z_n(\boldsymbol{y}))^T = A^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) \tag{51}$$

and define

$$f_{Y_1,\ldots,Y_n}(\boldsymbol{y}) = f_{Z_1,\ldots,Z_n}(A^{-1}(\boldsymbol{y} - \boldsymbol{\mu}))|J|,$$

where

$$J = \det \begin{pmatrix} \frac{\partial}{\partial y_1} z_1(\boldsymbol{y}) & \cdots & \frac{\partial}{\partial y_n} z_1(\boldsymbol{y}) \\ \frac{\partial}{\partial y_1} z_2(\boldsymbol{y}) & \cdots & \frac{\partial}{\partial y_n} z_2(\boldsymbol{y}) \\ \vdots & \vdots & \vdots \\ \frac{\partial}{\partial y_1} z_n(\boldsymbol{y}) & \cdots & \frac{\partial}{\partial y_n} z_n(\boldsymbol{y}) \end{pmatrix} \overset{(51)}{=} \det(A^{-1}),$$

then $f_{Y_1,\ldots,Y_n}$ is a PDF of $(Y_1, \ldots, Y_n)$. Note that

$$\det(A)\det(A^{-1}) = \det(AA^{-1}) = 1$$

and

$$\det(AA^T) = \det(A)\det(A^T) = (\det(A))^2,$$

so

$$|\det(A^{-1})| = |1/\det(A)| = (\det(AA^T))^{-1/2},$$

and

$$\begin{aligned} f_{Y_1,\ldots,Y_n}(\boldsymbol{y}) &= f_{Z_1,\ldots,Z_n}(A^{-1}(\boldsymbol{y}-\boldsymbol{\mu}))|J| \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-0.5(\boldsymbol{y}-\boldsymbol{\mu})^T(A^{-1})^T A^{-1}(\boldsymbol{y}-\boldsymbol{\mu})}|\det(A^{-1})| \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-0.5(\boldsymbol{y}-\boldsymbol{\mu})^T(AA^T)^{-1}(\boldsymbol{y}-\boldsymbol{\mu})}(\det(AA^T))^{-1/2} \end{aligned}$$

for $\boldsymbol{y} \in R^n$. Thus $f_{Y_1,\ldots,Y_n}$ is a PDF of $(Y_1,\ldots,Y_n)$ that is determined by AA^T and $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_n)^T$.

53. Since

$$\begin{aligned} \frac{d}{db}S(b) &= \frac{d}{db}E(U-bX)^2 \\ &= \frac{d}{db}E(U^2 - 2bUX + b^2X^2) \\ &= \frac{d}{db}\left[E(U^2) - 2bE(UX) + b^2E(X^2)\right] \\ &= -2E(UX) + 2bE(X^2) \end{aligned}$$

and

$$\begin{aligned} E\left[\frac{d}{db}(U-bX)^2\right] &= E\left[2(U-bX)\frac{d}{db}(U-bX)\right] \\ &= E[2(U-bX)(-X)] \\ &= E[-2UX + 2bX^2] \\ &= -2E(UX) + 2bE(X^2), \end{aligned}$$

we have

$$\frac{d}{db}S(b) = E\left[\frac{d}{db}(U-bX)^2\right].$$

54. Let $\boldsymbol{a} = (1,1,1,0)^T$, then

$$\boldsymbol{a}^T\boldsymbol{X} = (X_1 + X_2 + X_3)$$

and

$$
\begin{aligned}
(Var(X_1+X_2+X_3)) &= (Var(\boldsymbol{a}^T\boldsymbol{X})) \\
&= \boldsymbol{a}^T\Sigma\boldsymbol{a} \\
&= \begin{pmatrix} 1 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} \\
&= (4).
\end{aligned}
$$

55. (a) For $(x,y) \in R^2$,

$$
\begin{aligned}
f_{X,Y}(x,y) &= \int_{-\infty}^{\infty} f_{X,Y,Z}(x,y,z)dz \\
&= c\int_0^{\infty} e^{-(x^2+4xy+5y^2)}ze^{-z}dz \\
&= ce^{-(x^2+4xy+5y^2)}\int_0^{\infty} ze^{-z}dz \\
&= ce^{-(x^2+4xy+5y^2)}.
\end{aligned}
$$

Here the last equality follows from

$$
\int_0^{\infty} ze^{-z}dz = \lim_{b\to\infty}\left((-ze^{-z})|_0^b + \int_0^b e^{-z}dz\right) = 1. \qquad (52)
$$

For $(x,y) \in R^2$, let

$$
\begin{aligned}
f_{Z|(X,Y)=(x,y)}(z) &= \frac{f_{X,Y,Z}(x,y,z)}{f_{X,Y}(x,y)} \\
&= \frac{ce^{-(x^2+4xy+5y^2)}ze^{-z}I_{(0,\infty)}(z)}{ce^{-(x^2+4xy+5y^2)}} \\
&= ze^{-z}I_{(0,\infty)}(z)
\end{aligned}
$$

for $z \in R$, then $\{f_{Z|(X,Y)=(x,y)} : (x,y) \in R^2\}$ is a version of the conditional PDF of Z given (X,Y).

(b) From the solution to Part (a), $f_{X,Y}$ is a PDF of (X,Y), where

$$
f_{X,Y}(x,y) = ce^{-(x^2+4xy+5y^2)}
$$

for $(x, y) \in R^2$. For $y \in R$, let

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx \\ &= \int_{-\infty}^{\infty} ce^{-(x^2+4xy+5y^2)}dx \\ &= \int_{-\infty}^{\infty} ce^{-(x+2y)^2-y^2}dx \\ &= ce^{-y^2}\sqrt{2\pi\cdot 0.5}\underbrace{\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\cdot 0.5}}e^{-(x+2y)^2}dx}_{\text{integral of } N(-2y,0.5) \text{ PDF}} \\ &= ce^{-y^2}\sqrt{\pi}, \end{aligned}$$

then f_Y is a PDF of Y, so

$$\int_{-\infty}^{\infty} ce^{-y^2}\sqrt{\pi}dy = 1$$

and

$$\begin{aligned} c &= \frac{1}{\sqrt{\pi}}\left(\int_{-\infty}^{\infty} e^{-y^2}dy\right)^{-1} \\ &= \frac{1}{\sqrt{\pi}}\left(\sqrt{2\pi\cdot 0.5}\underbrace{\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\cdot 0.5}}e^{-y^2}dy}_{\text{integral of } N(0,0.5) \text{ PDF}}\right)^{-1} \\ &= \frac{1}{\sqrt{\pi}}\cdot\frac{1}{\sqrt{2\pi\cdot 0.5}} = \frac{1}{\pi}. \end{aligned}$$

For $y \in R = \{y : f_Y(y) > 0\}$, let

$$\begin{aligned} f_{X|Y=y}(x) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} \\ &= \frac{e^{-(x+2y)^2-y^2}/\pi}{e^{-y^2}/\sqrt{\pi}} \\ &= \frac{1}{\sqrt{\pi}}e^{-(x+2y)^2} \end{aligned}$$

for $x \in (-\infty, \infty)$, then $\{f_{X|Y=y} : y \in R\}$ is a version of the conditional PDF of X given Y.

(c) For $y \in R$, $f_{X|Y=y}$ is a PDF of $N(-2y, 0.5)$. Let $U \sim N(-2y, 0.5)$, then $E(U) = -2y$, $Var(U) = 0.5$,

$$E(U^2) = Var(U) + [E(U)]^2 = 0.5 + (-2y)^2 = 0.5 + 4y^2,$$

$$\begin{aligned} E(X|Y) &= \left.\int_{-\infty}^{\infty} xf_{X|Y=y}(x)dx\right|_{y=Y} \\ &= E(U)|_{y=Y} \\ &= (-2y)|_{y=Y} = -2Y, \end{aligned}$$

$$
\begin{aligned}
E(X^2|Y) &= \int_{-\infty}^{\infty} x^2 f_{X|Y=y}(x)dx\Big|_{y=Y} \\
&= E(U^2)\big|_{y=Y} \\
&= (0.5+4y^2)\big|_{y=Y} = 0.5+4Y^2,
\end{aligned}
$$

and

$$
Var(X|Y) = E(X^2|Y) - [E(X|Y)]^2 = 0.5 + 4Y^2 - (-2Y)^2 = 0.5.
$$

(d) Since $E(X|Y) = -2Y$ is a linear function of Y, $E(X|Y) = -2Y$ is the best linear predictor of X based on Y.

Note. Another way to solve this part of the problem is to obtain the best linear predictor of X based on Y using $E(X)$, $E(Y)$, $Var(X)$, $Var(Y)$ and $Cov(X,Y)$. Below we will compute these quantities first. From Part (b),

$$
f_Y(y) = \frac{1}{\sqrt{\pi}} e^{-y^2}
$$

for $y \in R$, so f_Y is a PDF of $N(0,\sigma^2)$ with $\sigma^2 = 0.5$ and $Y \sim N(0, 0.5)$. Thus

$$
E(Y) = 0 \text{ and } Var(Y) = 0.5. \tag{53}
$$

To find $Cov(X,Y)$, we will first compute $E(X)$, $E(X^2)$, $E(XY)$:

$$
E(X) = E[E(X|Y)] = E(-2Y) = -2E(Y) \overset{(53)}{=} 0, \tag{54}
$$

$$
\begin{aligned}
Var(X) &= E(X^2) \\
&= E[E(X^2|Y)] \\
&= E(0.5+4Y^2) \\
&= 0.5 + 4[Var(Y) + (E(Y))^2] \\
&\overset{(53)}{=} 0.5 + 4 \cdot 0.5 = 2.5,
\end{aligned} \tag{55}
$$

$$
\begin{aligned}
E(XY) &= E[E(XY|Y)] \\
&= E[YE(X|Y)] \\
&= E[Y(-2Y)] \\
&= -2E(Y^2) \\
&= -2[Var(Y) + (E(Y))^2] \\
&\overset{(53)}{=} -2 \cdot 0.5 = -1.
\end{aligned} \tag{56}
$$

From (54), (56) and (53), we have

$$
Cov(X,Y) = E(XY) - E(X)E(Y) = -1 - 0 \cdot 0 = -1. \tag{57}
$$

Let $a + bY$ be the best linear predictor of X based on Y, then

$$
E(X - (a+bY)) = 0 \tag{58}
$$

and

$$
Cov(X - (a+bY), Y) = 0. \tag{59}
$$

From (58),

$$a = E(X) - bE(Y) \overset{(54)(53)}{=} 0 - b \cdot 0 = 0.$$

From (59),

$$b = \frac{Cov(X,Y)}{Var(Y)} \overset{(57)(53)}{=} \frac{-1}{0.5} = -2,$$

so the best linear predictor of X based on Y is $a + bY = -2Y$.

56. Let

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$$

for $y \in (-\infty, \infty)$, then f_Y is a PDF of Y. Let $f_{X,Y}(x,y) = g_y(x) f_Y(y)$ for $(x,y) \in R^2$, then $f_{X,Y}$ is a joint PDF of (X,Y). The expression of $f_{X,Y}$ can be re-written as follows:

$$\begin{aligned} f_{X,Y}(x,y) &= g_y(x) f_Y(y) \\ &= \frac{1}{\sqrt{2\pi}} e^{-0.5(x-y)^2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \\ &= \frac{1}{2\pi} e^{-y^2+xy-0.5x^2} \\ &= \frac{1}{2\pi} e^{-(y-0.5x)^2-0.25x^2} \end{aligned}$$

for $(x,y) \in R^2$. Let

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy \\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-(y-0.5x)^2-0.25x^2} dy \\ &= \frac{1}{2\pi} e^{-0.25x^2} \cdot \sqrt{2\pi \cdot 0.5} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi \cdot 0.5}} e^{-(y-0.5x)^2} dy}_{\text{integral of } N(0.5x,0.5) \text{ PDF}} \\ &= \frac{1}{\sqrt{2\pi \cdot 2}} e^{-0.25x^2} \end{aligned}$$

for $x \in R$ and for $x \in R$, let

$$\begin{aligned} f_{Y|X=x}(y) &= \frac{f_{X,Y}(x,y)}{f_X(x)} \\ &= \frac{e^{-(y-0.5x)^2-0.25x^2}/(2\pi)}{e^{-0.25x^2}/\sqrt{2\pi \cdot 2}} \\ &= \frac{1}{\sqrt{\pi}} e^{-(y-0.5x)^2} \end{aligned}$$

for $y \in R$, then $\{f_{Y|X=x} : x \in R\}$ is a version of the conditional PDF of Y given X.

57. We will introduce some notation first.

- For $\mu \in R$, $\sigma > 0$, let $f_{\mu,\sigma}$ denote the PDF of $N(\mu, \sigma)$ given in Problem 27.

- For $\boldsymbol{\mu} \in R^p$ and Σ: a $p \times p$ covariance matrix, let $g_{\boldsymbol{\mu},\Sigma}$ denote the PDF of $N(\boldsymbol{\mu}, \Sigma)$ given in Problem 52:

$$g_{\boldsymbol{\mu},\Sigma}(\boldsymbol{x}) = \frac{1}{(2\pi)^{p/2}\sqrt{\det(\Sigma)}} e^{-0.5(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}, \boldsymbol{x} \in R^p.$$

(a) We first find the best linear predictor of X_1 based on X_2 and X_3. Let $a_0 + b_0X_2 + c_0X_3$ be the best linear predictor of X_1 based on X_2 and X_3, then we have

$$Cov(X_1 - (a_0 + b_0X_2 + c_0X_3), X_2) = 0,$$

$$Cov(X_1 - (a_0 + b_0X_2 + c_0X_3), X_3) = 0$$

and

$$E(X_1) = E(a_0 + b_0X_2 + c_0X_3).$$

Therefore, (b_0, c_0) can be obtained by solving

$$\begin{pmatrix} Var(X_2) & Cov(X_2, X_3) \\ Cov(X_2, X_3) & Var(X_3) \end{pmatrix} \begin{pmatrix} b_0 \\ c_0 \end{pmatrix} = \begin{pmatrix} Cov(X_1, X_2) \\ Cov(X_1, X_3) \end{pmatrix},$$

which is

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} b_0 \\ c_0 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix},$$

which gives $b_0 = -1/2$ and $c_0 = 0$. In addition, a_0 can be obtained by

$$a_0 = E(X_1) - E(b_0X_2 + c_0X_3) = 0 - ((-1/2) \cdot 0 + 0 \cdot 0) = 0.$$

The best linear predictor of X_1 based on X_2 and X_3 is $-X_2/2$.

Next, we find the best linear predictor of X_4 based on X_2 and X_3. Let $a_0 + b_0X_2 + c_0X_3$ be the best linear predictor of X_4 based on X_2 and X_3, then we have

$$Cov(X_4 - (a_0 + b_0X_2 + c_0X_3), X_2) = 0,$$

$$Cov(X_4 - (a_0 + b_0X_2 + c_0X_3), X_3) = 0$$

and

$$E(X_4) = E(a_0 + b_0X_2 + c_0X_3).$$

Therefore, (b_0, c_0) can be obtained by solving

$$\begin{pmatrix} Var(X_2) & Cov(X_2, X_3) \\ Cov(X_2, X_3) & Var(X_3) \end{pmatrix} \begin{pmatrix} b_0 \\ c_0 \end{pmatrix} = \begin{pmatrix} Cov(X_4, X_2) \\ Cov(X_4, X_3) \end{pmatrix},$$

which is

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} b_0 \\ c_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

which gives $b_0 = 0$ and $c_0 = 0$. In addition, a_0 can be obtained by

$$a_0 = E(X_1) - E(b_0X_2 + c_0X_3) = 0 - 0 = 0.$$

The best linear predictor of X_4 based on X_2 and X_3 is 0.

From the above calculation, the best linear predictor of $(X_1, X_4)^T$ based on X_2 and X_3 is $(-X_2/2, 0)^T$.

(b) From the solution to Part (a), the best linear predictor of X_1 based on X_2 and X_3 is $-X_2/2$. Since $-X_2/2$ is a linear function of X_2 such that

$$Cov(X_1 - (-X_2/2), X_2) = 0$$

and

$$E[X_1 - (-X_2/2)] = 0,$$

$-X_2/2$ is the best linear predictor of X_1 based on X_2 with expected squared prediction error

$$\begin{aligned} E(X_1 - (-X_2/2))^2 &= Var(X_1 + X_2/2) + [E(X_1 - (-X_2/2))]^2 \\ &= Var(X_1) + 2Cov(X_1, X_2/2) + Var(X_2/2) + 0^2 \\ &= Var(X_1) + Cov(X_1, X_2) + Var(X_2)/4 \\ &= 2 + (-1) + 2/4 = 3/2. \end{aligned}$$

Therefore, a version of the conditional PDF of X_1 given X_2 is $\left\{f_{-x_2/2,\sqrt{3/2}} : x_2 \in R\right\}$.

(c) We will first show that the best linear predictor of X_1 based on $(X_2, X_3, X_4)^T$ is $-X_2/2$ by verifying

$$Cov(X_1 - (-X_2/2), X_2) = 0, \tag{60}$$

$$Cov(X_1 - (-X_2/2), X_3) = 0, \tag{61}$$

$$Cov(X_1 - (-X_2/2), X_4) = 0, \tag{62}$$

and

$$E[X_1 - (-X_2/2)] = 0. \tag{63}$$

Note that from the solution to Part (a), $-X_2/2$ is the best linear predictor of X_1 based on X_2 and X_3, so (60), (61) and (63) hold. Moreover, (62) hold true since $Cov(X_1, X_4) = 0 = Cov(X_2, X_4)$, so we have shown that the best linear predictor of X_1 based on X_2, X_3 and X_4 is $-X_2/2$.

The expected squared prediction error for predicting X_1 using $-X_2/2$ is

$$E(X_1 - (-X_2/2))^2 = 3/2,$$

which has been computed in the solution to Part (b). Thus a version of the conditional PDF of X_1 given $(X_2, X_3, X_4)^T$ is $\left\{f_{-x_2/2,\sqrt{3/2}} : (x_2, x_3, x_4)^T \in R^3\right\}$.

(d) Since the distribution of $(X_1, X_2, X_3, X_4)^T$ is a multivariate normal distribution and $Cov(X_i, X_j) = 0$ for $i \in \{1, 2\}$ and $j \in \{3, 4\}$, $(X_1, X_2)^T$ and $(X_3, X_4)^T$ are independent. Let Σ_0 be the covarinace matrix of $(X_1, X_2)^T$, then $g_{(0,0)^T,\Sigma_0}$ is a PDF of $(X_1, X_2)^T$, and $\{g_{(0,0)^T,\Sigma_0} : (x_3, x_4)^T \in R^2\}$ is a version of the conditional PDF of $(X_1, X_2)^T$ give $(X_3, X_4)^T$.

(e) Since $(X_1, X_2)^T$ and $(X_3, X_4)^T$ are independent,

$$\begin{aligned} E(X_1 X_2 | X_3, X_4) &= E(X_1 X_2) \\ &= Cov(X_1, X_2) + E(X_1)E(X_2) \\ &= -1 + 0 = -1. \end{aligned}$$

58. For $\boldsymbol{x} \in R^n$ and $\boldsymbol{u} \in R^m$, let

$$f_{\boldsymbol{X},\boldsymbol{U}}(\boldsymbol{x},\boldsymbol{u}) = f_{\boldsymbol{X}}(\boldsymbol{x}) f_{\boldsymbol{U}}(\boldsymbol{u}),$$

then $f_{\boldsymbol{X},\boldsymbol{U}}$ is a PDF of $(\boldsymbol{X},\boldsymbol{U})$ since $\boldsymbol{X}$ and $\boldsymbol{U}$ are independent. Consider the transform T such that

$$T\left(\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{u} \end{pmatrix}\right) = \begin{pmatrix} \boldsymbol{x} \\ g(\boldsymbol{x}) + \boldsymbol{u} \end{pmatrix},$$

then

$$T\left(\begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{U} \end{pmatrix}\right) = \begin{pmatrix} \boldsymbol{X} \\ g(\boldsymbol{X}) + \boldsymbol{U} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{pmatrix}.$$

Let

$$f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y}) = f_{\boldsymbol{X},\boldsymbol{U}}(\boldsymbol{x},\boldsymbol{y} - g(\boldsymbol{x}))|J|$$

for $\boldsymbol{x} \in R^n$, $\boldsymbol{y} \in R^m$, where

$$\begin{aligned} J &= \det\begin{pmatrix} \frac{\partial}{\partial \boldsymbol{x}}\boldsymbol{x} & \frac{\partial}{\partial \boldsymbol{y}}\boldsymbol{x} \\ \frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{y} - g(\boldsymbol{x})) & \frac{\partial}{\partial \boldsymbol{y}}(\boldsymbol{y} - g(\boldsymbol{x})) \end{pmatrix} \\ &= \det\begin{pmatrix} I_{n\times n} & O_{n\times m} \\ \frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{y} - g(\boldsymbol{x})) & I_{m\times m} \end{pmatrix} = 1. \end{aligned}$$

Here for $\boldsymbol{v}$: a vector value function of a vector $\boldsymbol{w}$, $\frac{\partial}{\partial \boldsymbol{w}}\boldsymbol{v}$ denotes the matrix whose (i,j)-th element is the partial derivative of the i-th component of $\boldsymbol{v}$ with respect to the j-th component of $\boldsymbol{w}$, $I_{n\times n}$ and $I_{m\times m}$ are identity matrices of sizes $n \times n$ and $m \times m$ respectively, and $O_{n\times m}$ is the $n \times m$ matrix of zeros. From the above calculation, $f_{\boldsymbol{X},\boldsymbol{Y}}$ is a joint PDF of $(\boldsymbol{X},\boldsymbol{Y})$ and

$$f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y}) = f_{\boldsymbol{X},\boldsymbol{U}}(\boldsymbol{x},\boldsymbol{y} - g(\boldsymbol{x}))|J| = f_{\boldsymbol{X}}(\boldsymbol{x}) f_{\boldsymbol{U}}(\boldsymbol{y} - g(\boldsymbol{x})) \tag{64}$$

for $\boldsymbol{x} \in R^n$ and $\boldsymbol{y} \in R^m$. It is clear that $f_{\boldsymbol{X},\boldsymbol{Y}} > 0$ since $f_{\boldsymbol{X}} > 0$ and $f_{\boldsymbol{U}} > 0$ by assumption.

Let

$$g_0(\boldsymbol{x}) = \int_{R^m} f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y}) d\boldsymbol{y}$$

for $\boldsymbol{x} \in R^n$, then $g_0(\boldsymbol{x}) > 0$ since $f_{\boldsymbol{X},\boldsymbol{Y}} > 0$. For $\boldsymbol{x} \in R^n = \{\boldsymbol{x} : g_0(\boldsymbol{x}) > 0\}$, let

$$g_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}(\boldsymbol{y}) = \frac{f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y})}{g_0(\boldsymbol{x})}$$

for $\boldsymbol{y} \in R^m$, then a version of the conditional PDF of $\boldsymbol{Y}$ given $\boldsymbol{X}$ is $\{g_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}} : \boldsymbol{x} \in R^n\}$. Next, we will simplify the expressions of $g_0(\boldsymbol{x})$ and $g_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}$. Note that for $\boldsymbol{x} \in R^n$,

$$\begin{aligned} g_0(\boldsymbol{x}) &= \int_{R^m} f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y}) d\boldsymbol{y} \\ &\overset{(64)}{=} \int_{R^m} f_{\boldsymbol{X}}(\boldsymbol{x}) f_{\boldsymbol{U}}(\boldsymbol{y} - g(\boldsymbol{x})) d\boldsymbol{y} \\ &= f_{\boldsymbol{X}}(\boldsymbol{x}) \int_{R^m} f_{\boldsymbol{U}}(\underbrace{\boldsymbol{y} - g(\boldsymbol{x})}_{\boldsymbol{u}}) d\boldsymbol{y} \\ &= f_{\boldsymbol{X}}(\boldsymbol{x}) \underbrace{\int_{R^m} f_{\boldsymbol{U}}(\boldsymbol{u}) d\boldsymbol{u}}_{=1 \text{ since } f_{\boldsymbol{U}} \text{ is a PDF}} \\ &= f_{\boldsymbol{X}}(\boldsymbol{x}), \end{aligned} \tag{65}$$

and for $\boldsymbol{x} \in R^n$,

$$\begin{aligned} g_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}(\boldsymbol{y}) &= \frac{f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y})}{g_0(\boldsymbol{x})} \\ &\overset{(65)}{=} \frac{f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y})}{f_{\boldsymbol{X}}(\boldsymbol{x})} \\ &\overset{(64)}{=} \frac{f_{\boldsymbol{X}}(\boldsymbol{x}) f_{\boldsymbol{U}}(\boldsymbol{u} - g(\boldsymbol{x}))}{f_{\boldsymbol{X}}(\boldsymbol{x})} \\ &= f_{\boldsymbol{U}}(\boldsymbol{y} - g(\boldsymbol{x})) \\ &= f_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}(\boldsymbol{y}) \text{ (by definition)} \end{aligned}$$

for $\boldsymbol{y} \in R^m$. Therefore, a version of the conditional PDF of $\boldsymbol{Y}$ given $\boldsymbol{X}$ is $\{g_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}} : \boldsymbol{x} \in R^n\} = \{f_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}} : \boldsymbol{x} \in R^n\}$.

59. By assumption, X and ε are two random variables such that $X \sim N(0, \sigma_1^2)$, $\varepsilon \sim N(0, \sigma_2^2)$, and X and ε are independent, so $Q(\sigma_1, \sigma_2)$: the distribution of $(X + \varepsilon)$ is

$$N(E(X + \varepsilon), Var(X + \varepsilon)) = N(0, \sigma_1^2 + \sigma_2^2).$$

Note that for $\sigma_1 > 0$, $\sigma_2 > 0$ such that $\sigma_1 \neq \sigma_2$, we have

- both (σ_1, σ_2) and (σ_2, σ_1) are in $(0, \infty) \times (0, \infty)$, and
- $Q_{\sigma_1,\sigma_2} = Q_{\sigma_2,\sigma_1}$ but $(\sigma_1, \sigma_2) \neq (\sigma_2, \sigma_1)$.

Therefore, the family $\mathcal{C} = \{Q_{\sigma_1,\sigma_2} : (\sigma_1, \sigma_2) \in (0, \infty) \times (0, \infty)\}$ is not identifiable.

60. (a)(b)(d) are true. (c) is false.

61. For $\varepsilon > 0$, we have

$$\begin{aligned} 0 &\leq P(|\hat{\theta} - \theta| > \varepsilon) \\ &= P(|\hat{\theta} - E(\hat{\theta})| > \varepsilon) \text{ (since } E(\hat{\theta}) = \theta \text{ by assumption)} \\ &\leq \frac{Var(\hat{\theta})}{\varepsilon^2} \text{ (Chebyshev's inequality).} \end{aligned} \tag{66}$$

Since $\lim_{n\to\infty} Var(\hat{\theta}) = 0$ by assumption, we have

$$\lim_{n\to\infty} \frac{Var(\hat{\theta})}{\varepsilon^2} = \frac{1}{\varepsilon^2} \lim_{n\to\infty} Var(\hat{\theta}) = 0,$$

which, together with (66), implies that

$$\lim_{n\to\infty} P(|\hat{\theta} - \theta| > \varepsilon) = 0$$

for $\varepsilon > 0$, so the estimator $\hat{\theta}$ converges to θ in probability and $\hat{\theta}$ is a consistent estimator of θ.

62. Since

$$\|(X_{1,n}, \ldots, X_{k,n})^T - (Y_1, \ldots, Y_k)^T\|^2 = |X_{1,n} - Y_1|^2 + \cdots + |X_{k,n} - Y_k|^2,$$

for $\varepsilon > 0$,

$$\begin{aligned}
&|(X_{j,n}, -Y_j|^2 \le \frac{\varepsilon^2}{k} \text{ for each } j \in \{1, \ldots, k\} \\
&\Rightarrow \|(X_{1,n}, \ldots, X_{k,n})^T - (Y_1, \ldots, Y_k)^T\|^2 \le \varepsilon^2 \\
&\Rightarrow \|(X_{1,n}, \ldots, X_{k,n})^T - (Y_1, \ldots, Y_k)^T\| \le \varepsilon,
\end{aligned}$$

so

$$\begin{aligned}
0 &\le P\left(\|(X_{1,n}, \ldots, X_{k,n})^T - (Y_1, \ldots, Y_k)^T\| > \varepsilon\right) \\
&\le P\left(\cup_{j=1}^k \left\{|(X_{j,n}, -Y_j|^2 > \frac{\varepsilon^2}{k}\right\}\right) \\
&\le \sum_{j=1}^k P\left(\left\{|X_{j,n}, -Y_j| > \frac{\varepsilon}{\sqrt{k}}\right\}\right). \qquad (67)
\end{aligned}$$

Suppose that $X_{j,n}$ converges to X_j in probability for each $j \in \{1, \ldots, k\}$, then for $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(\left\{|X_{j,n}, -Y_j| > \frac{\varepsilon}{\sqrt{k}}\right\}\right) = 0 \qquad (68)$$

for each $j \in \{1, \ldots, k\}$. From (68) and (67), we have

$$\lim_{n\to\infty} P\left(\|(X_{1,n}, \ldots, X_{k,n})^T - (Y_1, \ldots, Y_k)^T\| > \varepsilon\right) = 0$$

for $\varepsilon > 0$. That is, we have

$$(X_{1,n}, \ldots, X_{k,n})^T \stackrel{\mathcal{P}}{\to} (Y_1, \ldots, Y_k)^T$$

as $n \to \infty$. The proof of Fact 4 is complete.

63. Suppose that X_n and X are random vectors on $(\Omega, \mathcal{F}, P)$, where $\mathcal{F}$ is a σ-field on Ω and P is a probability function defined on $\mathcal{F}$. Let

$$A = \{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}$$

and

$$B = \{\omega \in \Omega : \lim_{n\to\infty} g(X_n(\omega)) = g(X(\omega))\}.$$

Then by the assumption that X_n converges to X almost surely as $n \to \infty$, we have

$$P(A) = 1. \qquad (69)$$

We will show that

$$A \subset B, \qquad (70)$$

then we have

$$0 \le P(B^c) \stackrel{(70)}{\le} P(A^c) \stackrel{(69)}{=} 0$$

$$\Rightarrow P(B^c) = 0 \Rightarrow P(B) = 1,$$

which implies $g(X_n)$ converges to $g(X)$ almost surely as $n \to \infty$.

It remains to prove (70). Note that for $\omega \in A$, we have

$$\lim_{n\to\infty} X_n(\omega) = X(\omega),$$

which, together with the assumption that g is continuous on R^m, implies that

$$\lim_{n\to\infty} g(X_n(\omega)) = g(X(\omega)),$$

which implies that $\omega \in B$. Therefore, $A \subset B$ and (70) holds.

64. Since $X_1, \ldots, X_n$ are IID and

$$E(X_1) = \int_0^\theta x \cdot \frac{1}{\theta} dx = \frac{\theta}{2}$$

is finite, by SLLN, $\bar{X}$ converges to $\theta/2$ almost surely as $n \to \infty$, so $\bar{X}$ is a consistent estimator of $\theta/2$. Let $g(x) = 1/(2x)$ for $x > 0$, then g is continuous at $\theta/2$ for $\theta > 0$. Apply the continuous mapping theorem for convergence in probability, $g(\bar{X}) = 1/(2\bar{X})$ is a consistent estimator of $g(\theta/2) = 1/(2 \cdot \theta/2) = 1/\theta$.

65. (a) Let M_Y be the MGF of Y, where $Y \sim \Gamma(\alpha, 1)$ as given in the problem. Since $X_1 \sim \beta Y$, we will find $E(X_1)$ and $E(X_1^2)$ by finding $E(Y)$ and $E(Y^2)$ first using M_Y.

By the solution to Problem 34(b), $M_Y(s) = (1-s)^{-\alpha}$ for $s < 1$. Since for $\alpha > 0$,

$$\frac{d}{ds} M_Y(s) = \alpha(1-s)^{-\alpha-1}$$

and

$$\frac{d^2}{ds^2} M_Y(s) = \alpha(\alpha+1)(1-s)^{-\alpha-2},$$

we have

$$E(Y) = \alpha(1-s)^{-\alpha-1}\big|_{s=0} = \alpha \tag{71}$$

and

$$E(Y^2) = \alpha(\alpha+1)(1-s)^{-\alpha-2}\big|_{s=0} = \alpha(\alpha+1). \tag{72}$$

Since $X_1 \sim \beta Y$,

$$E(X_1) = E(\beta Y) = \beta E(Y) \overset{(71)}{=} \alpha\beta$$

and

$$E(X_1^2) = E((\beta Y)^2) = \beta^2 E(Y^2) \overset{(72)}{=} \beta^2\alpha(\alpha+1),$$

which gives

$$Var(X_1) = E(X_1^2) - [E(X_1)]^2 = \beta^2\alpha(\alpha+1) - (\alpha\beta)^2 = \alpha\beta^2.$$

(b) Let $\mu = E(X_1)$ and $\mu_2 = E(X_1^2)$, then from Part (a), we have

$$\begin{cases} \mu = \alpha\beta; \\ \mu_2 = \alpha\beta^2 + (\alpha\beta)^2. \end{cases} \tag{73}$$

In (73), we can solve for α and β as functions of μ and μ_2, which gives

$$\beta = \frac{\mu_2 - \mu^2}{\mu} \text{ and } \alpha = \frac{\mu}{\beta} = \frac{\mu}{(\mu_2 - \mu^2)/\mu} = \frac{\mu^2}{\mu_2 - \mu^2}.$$

Let

$$g(s,t) = \frac{s^2}{t-s^2} \text{ and } h(s,t) = \frac{t-s^2}{s}$$

for $(s,t) \in R^2$ such that $s > 0$ and $t > s^2$, then we have

(i) $\alpha = g(\mu, \mu_2)$,

(ii) $\beta = h(\mu, \mu_2)$, and

(iii) g and h are continuous at (μ, μ_2) since by assumption, $\alpha > 0$ and $\beta > 0$, which implies that

$$\mu = \alpha\beta > 0 \text{ and } \mu_2 = \alpha\beta^2 + (\alpha\beta)^2 > \mu^2.$$

Let $\bar{X} = \sum_{i=1}^n X_i/n$ and $\bar{Y} = \sum_{i=1}^n X_i^2/n$, then $(\bar{X}, \bar{Y})^T$ is a consistent estimator of $(\mu, \mu_2)^T$ by SLLN. By the continuous mapping theorem for convergence in probability,

$$(g(\bar{X}, \bar{Y}), h(\bar{X}, \bar{Y}))^T = \left(\frac{(\bar{X})^2}{\bar{Y} - (\bar{X})^2}, \frac{\bar{Y} - (\bar{X})^2}{\bar{X}} \right)^T$$

is a consistent estimator for

$$(g(\mu, \mu_2), h(\mu, \mu_2))^T = \left(\frac{\mu^2}{\mu_2 - \mu^2}, \frac{\mu_2 - \mu^2}{\mu} \right)^T = (\alpha, \beta)^T.$$

66. (a) Let $\boldsymbol{U}_i = (I_{\{a_1\}}(X_i), I_{\{a_2\}}(X_i), I_{\{a_3\}}(X_i))^T$ for $i = 1, \ldots, n$ and $\bar{\boldsymbol{U}} = \sum_{i=1}^n \boldsymbol{U}_i/n$, then $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_n$ are IID random vectors and $E(\boldsymbol{U}_1) = (p_1, p_2, p_3)^T$, so

$$Y_n = \sqrt{n} \left(\begin{pmatrix} \hat{p}_1 \\ \hat{p}_2 \\ \hat{p}_3 \end{pmatrix} - \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} \right) = \sqrt{n} \left(\bar{\boldsymbol{U}} - E(\boldsymbol{U}_1) \right).$$

By C.L.T., Y_n converges to $N(\mathbf{0}, \Sigma_0)$ in distribution, where $\mathbf{0} = (0,0,0)^T$ and

$$\begin{aligned} \Sigma_0 &= \text{covariance matrix of } \boldsymbol{U}_1 \\ &= \begin{pmatrix} Var(I_{\{a_1\}}(X_1)) & Cov(I_{\{a_1\}}(X_1), I_{\{a_2\}}(X_1)) & Cov(I_{\{a_1\}}(X_1), I_{\{a_3\}}(X_1)) \\ Cov(I_{\{a_2\}}(X_1), I_{\{a_1\}}(X_1)) & Var(I_{\{a_2\}}(X_1)) & Cov(I_{\{a_2\}}(X_1), I_{\{a_3\}}(X_1)) \\ Cov(I_{\{a_3\}}(X_1), I_{\{a_1\}}(X_1)) & Cov(I_{\{a_3\}}(X_1), I_{\{a_2\}}(X_1)) & Var(I_{\{a_3\}}(X_1)) \end{pmatrix}. \end{aligned}$$

Note that for $j, k \in \{1,2,3\}$ such that $j \neq k$,

$$\begin{aligned} Cov(I_{\{a_j\}}(X_1), I_{\{a_k\}}(X_1)) &= E[I_{\{a_j\}}(X_1) \cdot I_{\{a_k\}}(X_1)] - E(I_{\{a_j\}}(X_1))E(I_{\{a_k\}}(X_1)) \\ &= 0 - p_j p_k = -p_j p_k, \end{aligned}$$

and for $j \in \{1,2,3\}$,

$$\begin{aligned} Var(I_{\{a_j\}}(X_1)) &= E[(I_{\{a_j\}}(X_1))^2] - [E(I_{\{a_j\}}(X_1))]^2 \\ &= E[(I_{\{a_j\}}(X_1))] - p_j^2 \\ &= p_j - p_j^2. \end{aligned}$$

Thus

$$\Sigma_0 = \begin{pmatrix} p_1 - p_1^2 & -p_1p_2 & -p_1p_3 \\ -p_1p_2 & p_2 - p_2^2 & -p_2p_3 \\ -p_1p_3 & -p_2p_3 & p_3 - p_3^2 \end{pmatrix} \tag{74}$$

and the limiting distribution of Y_n is $N(\mathbf{0}, \Sigma_0)$.

(b) Let Y be a random vector such that $Y \sim N(\mathbf{0}, \Sigma_0)$, where Σ_0 is given in (74), then by Part (a), Y_n converges to Y in distribution as $n \to \infty$. For $\boldsymbol{y} = (y_1, y_2, y_3)^T \in R^3$, define $g(\boldsymbol{y}) = A^{-1}\boldsymbol{y}$, then g is continuous on R^3. By the continuous mapping theorem for convergence in distribution, we have

$$A^{-1}Y_n = g(Y_n) \text{ converges in distribution to } g(Y) = A^{-1}Y$$

as $n \to \infty$, so the limiting distribution of $A^{-1}Y_n$ is the distribution of $A^{-1}Y$. Since $Y \sim N(\mathbf{0}, \Sigma_0)$, the distribution of $A^{-1}Y$ is

$$N(E(A^{-1}Y), A^{-1}\Sigma_0(A^{-1})^T) = N(\mathbf{0}, A^{-1}\Sigma_0 A^{-1})$$

since

$$A^{-1} = \begin{pmatrix} 1/\sqrt{p}_1 & 0 & 0 \\ 0 & 1/\sqrt{p}_2 & 0 \\ 0 & 0 & 1/\sqrt{p}_3 \end{pmatrix} = (A^{-1})^T.$$

Let $\Sigma = A^{-1}\Sigma_0 A^{-1}$, then the limiting distribution of $A^{-1}Y_n$ is $N(\mathbf{0}, \Sigma)$.

To show that $\Sigma^2 = \Sigma$, let $\boldsymbol{p} = (p_1, p_2, p_3)^T$, then $\boldsymbol{pp}^T$ is a 3×3 matrix whose (i, j)-th element is $p_i p_j$ for $i, j \in \{1, 2, 3\}$ and

$$\begin{aligned}
\Sigma_0 &\overset{(74)}{=} \begin{pmatrix} p_1 - p_1^2 & -p_1p_2 & -p_1p_3 \\ -p_1p_2 & p_2 - p_2^2 & -p_2p_3 \\ -p_1p_3 & -p_2p_3 & p_3 - p_3^2 \end{pmatrix} \\
&= \begin{pmatrix} p_1 & 0 & 0 \\ 0 & p_2 & 0 \\ 0 & 0 & p_3 \end{pmatrix} + \begin{pmatrix} -p_1^2 & -p_1p_2 & -p_1p_3 \\ -p_1p_2 & -p_2^2 & -p_2p_3 \\ -p_1p_3 & -p_2p_3 & -p_3^2 \end{pmatrix} \\
&= A^2 - \boldsymbol{pp}^T,
\end{aligned}$$

which implies that

$$\begin{aligned}
\Sigma &= A^{-1}\Sigma_0 A^{-1} \\
&= A^{-1}(A^2 - \boldsymbol{pp}^T)A^{-1} \\
&= A^{-1}A^2A^{-1} - A^{-1}\boldsymbol{pp}^T A^{-1} \\
&= I_{3\times 3} - A^{-1}\boldsymbol{pp}^T A^{-1}, \qquad (75)
\end{aligned}$$

where $I_{3\times 3}$ is the 3×3 identity matrix. Moreover,

$$\boldsymbol{p}^T A^{-1} = (p_1 \quad p_2 \quad p_3) \begin{pmatrix} 1/\sqrt{p}_1 & 0 & 0 \\ 0 & 1/\sqrt{p}_2 & 0 \\ 0 & 0 & 1/\sqrt{p}_3 \end{pmatrix} = (\sqrt{p_1} \quad \sqrt{p_2} \quad \sqrt{p_3}). \qquad (76)$$

Therefore,

$$\begin{aligned}
\Sigma^2 &\overset{(75)}{=} \Sigma\left(I_{3\times 3} - A^{-1}\boldsymbol{pp}^T A^{-1}\right) \\
&= \Sigma - \Sigma A^{-1}\boldsymbol{pp}^T A^{-1} \\
&\overset{(75)}{=} \Sigma - \left(I_{3\times 3} - A^{-1}\boldsymbol{pp}^T A^{-1}\right) A^{-1}\boldsymbol{pp}^T A^{-1} \\
&\overset{(76)}{=} \Sigma - \left(A^{-1}\boldsymbol{pp}^T A^{-1} - A^{-1}\boldsymbol{p} \underbrace{(\sqrt{p_1} \quad \sqrt{p_2} \quad \sqrt{p_3}) \begin{pmatrix} \sqrt{p_1} \\ \sqrt{p_2} \\ \sqrt{p_3} \end{pmatrix}}_{=1} \boldsymbol{p}^T A^{-1} \right) \\
&= \Sigma.
\end{aligned}$$