Sample Midterm Problems Solutions

1. a. Independence does not generally imply conditional independence. We can also see it from
\[ E[X_1, X_2|X_3] = \text{Cov}(X_1, X_2|X_3) + E[X_1|X_3] E[X_2|X_3], \]
where the covariance term can be either positive (e.g., by letting \( X_3 = X_1 - X_2 \)), or negative (e.g., by letting \( X_3 = X_1 + X_2 \)).

b. \( \leq \). By the law of conditional variances (and conditioning both sides on \( Y \)), it follows that
\[ E[\text{Var}(X|Y)] = E[\text{Var}(X|Y, Z)] + E[\text{Var}(E(X|Y, Z)|Y)]. \]
Thus \( E[\text{Var}(X|Y)] \geq E[\text{Var}(X|Y, Z)] \). This makes sense because with more observations \( Y, Z \), the MSE of the best estimate of the signal \( X \) should be less than or equal to that observing only \( Y \).

c. \( \leq \). From the previous result, it follows that
\[ E[\text{Var}(X|Y, g(Y))] \leq E[\text{Var}(X|g(Y))]. \]
But \( g(Y) \) is completely determined by \( Y \), thus \( E[\text{Var}(X|Y, g(Y))] = E[\text{Var}(X|Y)]. \) This result makes sense because in general \( Y \) provides better information about the signal \( X \) than any function of it.

d. \( \geq \). First note that \( E(X^2|Z) \geq \text{Var}(X|Z) \) and \( E(Y^2|Z) \geq \text{Var}(Y|Z) \). Thus
\[ E(X^2|Z) E(Y^2|Z) \geq \text{Var}(X|Z) \text{Var}(Y|Z). \]
Now, using Schwarz inequality, we obtain
\[ \text{Var}(X|Z) \text{Var}(Y|Z) \geq (\text{Cov}(X, Y|Z))^2. \]
Taking expectations of both sides, we obtain
\[ E[\text{Var}(X|Z) \text{Var}(Y|Z)] \geq E[(\text{Cov}(X, Y|Z))^2]. \]
But \( E[(\text{Cov}(X, Y|Z))^2] \geq [E(\text{Cov}(X, Y|Z))]^2. \)

e. \( \leq \). We use Jensen’s inequality twice and the fact that \( E(X) \leq 1 \)
\[ E \left( \log_2(1 + \sqrt{X}) \right) \leq \log_2 \left( 1 + E \left( \sqrt{X} \right) \right) \]
\[ \leq \log_2 \left( 1 + \sqrt{E(X)} \right) \]
\[ \leq \log_2(1 + 1) \]
\[ \leq 1. \]
f. Consider
\[
P\{(XY)^2 > 16\} \leq \frac{E[(XY)^2]}{16}, \quad \text{by Markov inequality}
\]
\[
\leq \frac{\sqrt{E(X^4)E(Y^4)}}{16}, \quad \text{by Schwarz inequality}
\]
\[
= \frac{1}{8}.
\]

2. a. First, we compute the cdf \(F_L(l)\) as follows. For \(l \in [0, 1]\),
\[
F_L(l) = P\{L \leq l\}
\]
\[
= P\{L \leq l, X \geq Y\} + P\{L \leq l, X < Y\}
\]
\[
= P\{X \leq l, X \geq Y\} + P\{1 - X \leq l, X < Y\}
\]
\[
= \frac{l^2}{2} + \frac{l^2}{2}
\]
\[
= l^2,
\]
where we have used the fact that \(X\) and \(Y\) are independent and have pdf \(U[0, 1]\). Their joint pdf is thus constant over the square \([0, 1] \times [0, 1]\). From the cdf, we get the pdf by taking the derivative,
\[
f_L(l) = 2l, \quad \text{for } l \in [0, 1].
\]
b. For \(Y = y \in (0, 1)\), we take expectation with respect to \(X\), since \(X\) and \(Y\) are independent. Thus
\[
E(L | Y = y) = \int_0^y (1 - x) \, dx + \int_y^1 x \, dx
\]
\[
= y - \frac{y^2}{2} + \frac{1}{2} - \frac{y^2}{2} = y + \frac{1}{2} - y^2.
\]
Therefore, \(E(L | Y) = \frac{1}{2} + Y - Y^2\).

3. **Wireless Channel.** To find the best linear MSE estimate of \((X_1 + X_2)\) given \(Y\), we need to compute the first and second order statistics. Consider
\[
E(X_1 + X_2) = 0,
\]
\[
E(Y) = E(H_1 X_1 + H_2 X_2 + Z) = E(H_1) E(X_1) + E(H_2) E(X_2) + E(Z) = 0,
\]
\[
\text{Var}(Y) = E[(H_1 X_1 + H_2 X_2 + Z)^2]
\]
\[
\overset{1}{=} E[(H_1 X_1)^2] + E[(H_2 X_2)^2] + 2 E(H_1 H_2 X_1 X_2) + E(Z^2)
\]
\[
= E(H_1^2) E(X_1^2) + E(H_2^2) E(X_2^2) + 2 E(H_1) E(H_2) E(X_1 X_2) + N
\]
\[
= 2P + 2P + 2\rho P + N = 2P(2 + \rho) + N,
\]
\[
\text{Cov}(X_1 + X_2, Y) = E[(X_1 + X_2)(H_1 X_1 + H_2 X_2 + Z)]
\]
\[
= E(H_1 X_1 X_2) + E(H_2 X_1 X_2) + E(H_1 X_1^2) + E(H_2 X_2^2)
\]
\[
= (E(H_1) + E(H_2)) E(X_1 X_2) + E(H_1) E(X_1^2) + E(H_2) E(X_2^2)
\]
\[
= 2P(1 + \rho),
\]
where (1) follows because \( Z \) is independent of the other random variables and thus the cross terms are equal to 0.

Thus the best linear estimate is

\[
\hat{U} = \frac{2P(1 + \rho)}{2P(2 + \rho) + N}Y.
\]

4. a. i. TRUE. \( a \) is the variance of \( X \), and thus non-negative. The strict inequality holds by assumption.

ii. NEITHER. \( b_1 \) is a covariance and can be either positive or negative.

iii. TRUE. Since \( |\Sigma| = a^2 - b_1^2 - b_2^2 > 0 \).

iv. TRUE. Since \( X \) and \( Z \) are uncorrelated and Gaussian.

v. FALSE. The joint distribution of \( X \) and \( Z \) conditioned on \( Y = y \) is Gaussian.

Using property #4 of Gaussian random vectors, the conditional covariance matrix is

\[
\text{Cov}(X, Z|Y = y) = \begin{bmatrix}
a & b_1 \\
0 & a
\end{bmatrix} - \begin{bmatrix}
b_1 \\
b_2
\end{bmatrix} a^{-1} \begin{bmatrix}
b_1 & b_2
\end{bmatrix}
\]

\[
= \begin{bmatrix}
a - b_1^2/a & -b_1 b_2/a \\
-b_1 b_2/a & a - b_2^2/a
\end{bmatrix},
\]

which has non-zero off-diagonal elements. We conclude that \( X \) and \( Z \) are correlated given \( Y \), and thus not independent.

b. The best estimate of \( X \) given \( Y \) is

\[
\hat{X} = \text{E}(X|Y) = \frac{b_1}{a} Y.
\]

Thus we can write \( [\hat{X} \ Z]^T \) as a linear transformation of \( [Y \ Z]^T \) as

\[
[\hat{X} \ Z] = \begin{bmatrix}
b_1/a & 0 \\
0 & 1
\end{bmatrix} [Y \ Z].
\]

The transformed covariance matrix is

\[
\begin{bmatrix}
b_1/a & 0 \\
0 & 1
\end{bmatrix} \begin{bmatrix}
a & b_2 \\
b_2 & a
\end{bmatrix} \begin{bmatrix}
b_1/a & 0 \\
0 & 1
\end{bmatrix} = \begin{bmatrix}
b_1^2/a & b_1 b_2/a \\
b_1 b_2/a & a
\end{bmatrix}.
\]

Thus,

\[
[\hat{X} \ Z] \sim \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix}
b_1^2/a & b_1 b_2/a \\
b_1 b_2/a & a
\end{bmatrix} \right).
\]

5. Two coins.

a. Since \( p_\Theta(1) = p_\Theta(2) = 1/2 \), the maximum likelihood decoder minimizes the probability of error. The conditional pmf of \( (X_1, X_2) \) given \( \Theta = 1 \) is equal to:

\[
P \{X_1 = x_1, X_2 = x_2|\Theta = 1\} = P \{X_1 = x_1|\Theta = 1\} P \{X_2 = x_2|\Theta = 1\} = \frac{1}{4}.
\]
To derive the conditional pmf of \((X_1, X_2)\) given \(\Theta = 2\) we apply the law of total probability and the conditional independence assumption:

\[
P \{ X_1 = x_1, X_2 = x_2 | \Theta = 2 \} = \int_{u=0}^{1} f_P(p) p_{X_2|P}(x_2 | u) \, du
\]

Evaluating this expression, we obtain:

\[
P \{ X_1 = 1, X_2 = 1 | \Theta = 2 \} = \int_{u=0}^{1} u^2 \, du = \frac{1}{3},
\]

\[
P \{ X_1 = 0, X_2 = 0 | \Theta = 2 \} = \int_{u=0}^{1} (1 - u)^2 \, du = \frac{1}{3},
\]

\[
P \{ X_1 = 1, X_2 = 2 | \Theta = 2 \} = \int_{u=0}^{1} u (1 - u) \, du = \frac{1}{6}.
\]

As a result, by the maximum likelihood criterion, the detection rule is to choose \(\Theta = 2\) if the flips are the same (i.e., \(X_1 = X_2\)) and \(\Theta = 1\) if they are different.

b. The event of an error occurring for the detection rule derived in the previous part can be decomposed into two disjoint events: the flips being the same if \(\Theta = 1\) and the flips being different if \(\Theta = 2\). We compute the probability of these events and add them to obtain the probability of error:

\[
P \{ X_1 = X_2, \Theta = 1 \} = p_\Theta (1) P \{ X_1 = X_2 | \Theta = 1 \} = \frac{1}{2} \left( \frac{1}{4} + \frac{1}{4} \right) = \frac{1}{4},
\]

\[
P \{ X_1 \neq X_2, \Theta = 2 \} = p_\Theta (2) P \{ X_1 \neq X_2 | \Theta = 1 \}
\]

\[
= \frac{1}{2} \cdot 2 P \{ X_1 = 1, X_2 = 2 | \Theta = 2 \} = \frac{1}{6}.
\]

Hence,

\[
P_e = \frac{5}{12}.
\]