Sample Midterm Problems Solutions

1. a. None. 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

\[ E[\text{Var}(X | Y)] = E[\text{Var}(X | Y, Z)] + E[\text{Var}(E(X | Y, Z))]. \]

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. \( \leq \). Consider
\[
P\{ (XY)^2 > 16 \} \leq \frac{E[(XY)^2]}{16}, \text{ by Markov inequality} \]
\[
\leq \sqrt{E(X^4)E(Y^4)} \leq \frac{1}{16}, \text{ by Schwarz inequality} \]
\[
= \frac{1}{8}.
\]

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,
\[
f_L(l) = 2l, \text{ for } l \in [0, 1].
\]

b. For \( Y = y \in (0, 1) \), we take expectation with respect to \( X \), since \( X \) and \( Y \) are independent. Thus
\[
E(L \mid Y = y) = \int_0^y (1 - x) \, dx + \int_y^1 x \, dx
= y - \frac{y^2}{2} + \frac{1}{2} - \frac{y^2}{2} = y + \frac{1}{2} - y^2.
\]
Therefore, \( E(L \mid Y) = \frac{1}{2} + Y - Y^2 \).

3. As convergence in probability is implied by both m.s. and a.s. convergence, we have to show that the statement holds for convergence in probability only. By definition we have:
\[
\forall \epsilon > 0 \ P\{ |X_n - X| \geq \epsilon \} \xrightarrow{n \to \infty} 0
\]
\[
\forall \epsilon > 0 \ P\{ |X_n - Y| \geq \epsilon \} \xrightarrow{n \to \infty} 0
\]
Using de Morgan’s law and a union bound we get for any two events \( A, B : P\{ A \cap B \} \geq 1 - P\{ A^c \cup B^c \} \geq 1 - P\{ A^c \} - P\{ B^c \} \), so by choosing the \( A = \{ |X_n - X| < \epsilon \} \), \( B = \{ |X_n - Y| < \epsilon \} \) we get:
\[
P\{ |X_n - X| < \epsilon, |X_n - Y| < \epsilon \} \geq 1 - P\{ |X_n - X| \geq \epsilon \} - P\{ |X_n - Y| \geq \epsilon \}
\]
\[
P\{ |X - Y| < 2\epsilon \} \geq P\{ |X_n - X| < \epsilon, |X_n - Y| < \epsilon \}
\]
Combining the two inequalities we get \( \forall \epsilon > 0 : \)
\[
P\{ |X - Y| < 2\epsilon \} \geq 1 - P\{ |X_n - X| \geq \epsilon \} - P\{ |X_n - Y| \geq \epsilon \} \xrightarrow{n \to \infty} 1.
4. Gaussian multiple access channel.
   a. Define
   \[ \Theta = \begin{cases} 
   \theta_1 & \text{if the first sender is active} \\
   \theta_2 & \text{if the second sender is active,} 
   \end{cases} \]
   where
   \[ \theta_1 = \begin{bmatrix} +\sqrt{P} \\
   -\sqrt{P} \end{bmatrix}, \quad \theta_2 = \begin{bmatrix} -\sqrt{P} \\
   +\sqrt{P} \end{bmatrix}. \]
   We would like to detect \( \Theta \) from the observation \( Y = \Theta + Z \), where \( Z \sim \mathcal{N}(0, NI) \).
   Since the noise is white and the signals are equally likely, the MAP rule reduces to
   the ML decoding rule, which in turn can be written as the minimum distance decoding rule
   \[ \hat{\Theta}(y) = \begin{cases} 
   \theta_1 & \text{if } y^T (\theta_2 - \theta_1) < \frac{1}{2} (\theta_2^T \theta_2 - \theta_1^T \theta_1) \\
   \theta_2 & \text{otherwise.} 
   \end{cases} \]
   Substituting \( \theta_1 \) and \( \theta_2 \) and simplifying, the rule becomes
   \[ \hat{\Theta}(y) = \begin{cases} 
   \theta_1 & \text{if } y_1 > y_2 \\
   \theta_2 & \text{otherwise.} 
   \end{cases} \]
   b. The minimum probability of error is
   \[ P_e = Q \left( \sqrt{\frac{2P}{N}} \right), \]
   where we have substituted the signal power \( \theta_1^T \theta_1 = \theta_2^T \theta_2 = 2P \).

5. 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 \( Y = y \) as Gaussian.
   Using property #4 of Gaussian random vectors, the conditional covariance matrix is
   \[ \text{Cov}(X, Z | Y = y) = \begin{bmatrix} a & 0 \\
   0 & a \end{bmatrix} - \begin{bmatrix} b_1 \\
   b_2 \end{bmatrix} a^{-1} \begin{bmatrix} b_1 & b_2 \end{bmatrix} \]
   \[ = \begin{bmatrix} a - b_1^2/a & -b_1b_2/a \\
   -b_1b_2/a & a - b_2^2/a \end{bmatrix}, \]
   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

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

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

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

The transformed covariance matrix is

\[
\begin{bmatrix} b_1/a & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & b_2 \\ b_2 & a \end{bmatrix} \begin{bmatrix} b_1/a & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} b_1^2/a & b_1b_2/a \\ b_1b_2/a & a \end{bmatrix}.
\]

Thus,

\[
\begin{bmatrix} \hat{X} \\ Z \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} b_1^2/a & b_1b_2/a \\ b_1b_2/a & a \end{bmatrix} \right).
\]