Packet switching. Let $N$ be the number of packets per unit time arriving at a network switch. Each packet is routed to output port 1 with probability $p$ and to output port 2 with probability $1 - p$, independent of $N$ and of other packets. Let $X$ be the number of packets per unit time routed to output port 1. Thus

$$X = \begin{cases} 0 & N = 0 \\ \sum_{i=1}^{N} Z_i & N > 0 \end{cases}$$

where $Z_i = \begin{cases} 1 & \text{packet } i \text{ routed to Port 1} \\ 0 & \text{packet } i \text{ routed to Port 2} \end{cases}$, and $Z_1, Z_2, \ldots, Z_N$ are conditionally independent given $N$. Suppose that $N \sim \text{Poisson}(\lambda)$, i.e., has Poisson pmf with parameter $\lambda$.

(a) Find the mean and variance of $X$.

(b) Find the pmf of $X$ and the pmf of $N - X$.

Solution:

(a) Note that given $N = n$, $X \sim \text{Bin}(n, p)$, for $n \neq 0$, and $X = 0$, for $n = 0$.

Thus in either case, $E[X|N = n] = np$, and $\text{Var}(X|N = n) = np(1 - p)$.

Hence, $E[X|N] = np$, and $\text{Var}(X|N) = np(1 - p)$. Thus,

$$E[X] = E[E[X|N]] = E[Np] = \lambda p, \text{ and}$$

$$\text{Var}(X) = E[\text{Var}(X|N)] + \text{Var}(E[X|N]) = E[Np(1 - p)] + \text{Var}(Np) = p(1 - p)E[N] + p^2 \text{Var}(N) = \lambda p(1 - p) + \lambda p^2 = \lambda p.$$

(b) We have, for $n \neq 0$ and $x = 0, 1, 2, \ldots, n$,

$$P(X = x|N = n) = \binom{n}{x} p^x (1 - p)^{n-x}.$$
Thus, for \( x = 1, 2, \ldots \), we have

\[
P(X = x, N \neq 0) = \sum_{n=1}^{\infty} P(X = x | N = n)P(N = n)
\]

\[
= \sum_{n=x}^{\infty} \left( \begin{array}{c} n \\ x \end{array} \right) p^x (1 - p)^{n-x} e^{-\lambda} \frac{\lambda^n}{n!}
\]

\[
= \sum_{n=x}^{\infty} \frac{n!}{x!(n-x)!} p^x (1 - p)^{n-x} e^{-\lambda} \frac{\lambda^n}{n!}
\]

\[
= e^{-\lambda p^x} \sum_{n=x}^{\infty} \frac{(1 - p)^{n-x} \lambda^n}{(n-x)!}
\]

\[
= e^{-\lambda (\lambda p)^x} \sum_{n=x}^{\infty} \frac{(1 - p)^{n-x} \lambda^{n-x}}{(n-x)!}
\]

\[
= e^{-\lambda (\lambda p)^x} e^{\lambda (1-p)}
\]

\[
= e^{-\lambda p (\lambda p)^x}.
\]

We also have, for \( x \neq 0 \), \( P(X = x, N = 0) = 0 \) and thus, for \( x = 1, 2, \ldots \), \( P_X(x) = e^{-\lambda p (\lambda p)^x} \). Hence,

\[
P_X(0) = 1 - \sum_{x=1}^{\infty} P_X(x)
\]

\[
= 1 - \sum_{x=1}^{\infty} e^{-\lambda p (\lambda p)^x}
\]

\[
= e^{-\lambda p}.
\]

Thus, \( X \sim \text{Poisson}(\lambda p) \).

Let \( Y = N - X \). Then, given \( N = n \neq 0 \), \( Y \sim \text{Bin}(n, 1 - p) \), and given \( N = 0 \), \( Y = 0 \). Thus, the whole previous analysis goes through, with \( p \) replaced by \( 1 - p \).

Hence, \( Y \sim \text{Poisson}(\lambda(1-p)) \), and

\[
p_Y(y) = e^{-\lambda(1-p)} \frac{(\lambda(1-p))^y}{y!}, \, y \in \mathbb{N} \cup \{0\}.
\]
We also have, for \( x, y \in \{0, 1, 2, \ldots \}, \)
\[
P(N - X = y|X = x) = \frac{P(N = x + y|X = x)}{P(X = x)}
\]
\[
= \frac{(x+y)p^{x}(1-p)^{y}e^{-\lambda} }{(x+y)!} \frac{\lambda^{x+y} }{(x+y)!}
\]
\[
e^{-\lambda p^x x!} \frac{\lambda y(1-p)^y y!}{y!}
\]
\[= e^{-\lambda(1-p)} \frac{\lambda^y (1-p)^y}{x! y!},
\]
Thus, \( X \) and \( Y = N - X \) are independent, and thus their joint pmf is given by
\[
p_{X,Y}(x,y) = p_{X}(x)p_{Y}(y) = e^{-\lambda} \frac{\lambda^x p^x (1-p)^y}{x! y!},
\]
\( x, y \in \mathbb{N} \cup \{0\}. \)

2. Markov chain. Assume that the continuous random variables \( X_1 \) and \( X_3 \) are independent given \( X_2 \). Show that \( f(x_1, x_2, x_3) = f(x_1)f(x_2|x_1)f(x_3|x_2) = f(x_3)f(x_2|x_3)f(x_1|x_2). \)

Solution: We have, for every \( x_1, x_2, x_3, \)
\[
f_{X_1,X_3|X_2}(x_1, x_3|x_2) = f_{X_1|X_2}(x_1|x_2)f_{X_3|X_2}(x_3|x_2). \tag{1}
\]
Thus, we have
\[
f_{X_1,X_2,X_3}(x_1, x_2, x_3) = f_{X_1,X_3|X_2}(x_1, x_3|x_2)f_{X_2}(x_2)
\]
\[
= f_{X_1|X_2}(x_1|x_2)f_{X_2}(x_2)f_{X_3|X_2}(x_3|x_2) \text{ (from (1))}
\]
\[
= f_{X_1|X_2}(x_1|x_2)f_{X_3|X_2}(x_3|x_2)
\]
\[= f_{X_1}(x_1)f_{X_2}(x_2) f_{X_3|X_2}(x_3|x_2).
\]
Similarly,
\[
f_{X_1,X_2,X_3}(x_1, x_2, x_3) = f_{X_1,X_3|X_2}(x_1, x_3|x_2)f_{X_2}(x_2)
\]
\[
= f_{X_1|X_2}(x_1|x_2)f_{X_2}(x_2)f_{X_3|X_2}(x_3|x_2) \text{ (from (1))}
\]
\[
= f_{X_3|X_2}(x_3|x_2)f_{X_1|X_2}(x_1|x_2)
\]
\[= f_{X_3}(x_3)f_{X_2}(x_2) f_{X_1|X_2}(x_1|x_2).
\]

3. The correlation matrix \( C \) for a random vector \( \mathbf{X} \) is the matrix whose entries are \( c_{ij} = \mathbb{E}(X_i X_j) \).
Show that it has the same properties as the covariance matrix, i.e., that it is real, symmetric, and positive semidefinite definite.

Solution: Since \( \mathbf{X} \) is real, \( X_i X_j \) is real for every \( i \) and \( j \), hence \( c_{ij} = \mathbb{E}[X_i X_j] \) is real.
We have, for any pair of indices $i$ and $j$,

$$c_{ji} = E[X_jX_i] = E[X_iX_j] = c_{ij},$$
and thus $C$ is symmetric.

For any constant vector $a$ of the same size as $X$, we have

$$a^T Ca = E[a^T XX^T a] = E[(a^T X)^2] \geq 0,$$

since $(a^T X)^2$ is a nonnegative random variable.

4. Let $X \sim \mathcal{N}(0, \Sigma)$, where

$$\Sigma = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 9 \end{bmatrix}.$$

Find the best MSE estimate of $X_1$ given $X_2$ and $X_3$. What is the MSE?

**Solution:** Let us partition the covariance matrix $\Sigma$ as follows:

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

where $\Sigma_{11} = 1$, $\Sigma_{12} = \begin{bmatrix} 2 & 1 \end{bmatrix}$, $\Sigma_{21} = \Sigma_{12}^T$ and $\Sigma_{22} = \begin{bmatrix} 5 & 2 \\ 2 & 9 \end{bmatrix}$. Then,

$$X_1 | \{X_2, X_3\} \sim \mathcal{N}\left(\Sigma_{12}\Sigma_{22}^{-1} \begin{bmatrix} X_2 \\ X_3 \end{bmatrix}, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right).$$

Thus, the best MSE estimate of $X_1$ given $X_2$ and $X_3$ is given by

$$E[X_1 | X_2, X_3] = \Sigma_{12}\Sigma_{22}^{-1} \begin{bmatrix} X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \begin{bmatrix} -\frac{9}{41} & -\frac{2}{41} \\ \frac{9}{41} & \frac{2}{41} \end{bmatrix} \begin{bmatrix} X_2 \\ X_3 \end{bmatrix} = \frac{16X_2 + X_3}{41}.$$
The MSE is given by
\[
E \left[ \text{Var}(X_1|X_2, X_3) \right] = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}
\]
\[
= 1 - \left( \begin{bmatrix} 2 & 1 \end{bmatrix} \begin{bmatrix} \frac{9}{\pi^2} & -\frac{2}{\pi^2} \\ -\frac{2}{\pi^2} & \frac{9}{\pi^2} \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right)
\]
\[
= 1 - \left( \begin{bmatrix} \frac{16}{\pi^2} & \frac{1}{\pi^2} \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right)
\]
\[
= \frac{8}{41}.
\]

5. **Gaussian random vector.** Given a Gaussian random vector \( X \sim \mathcal{N}(\mu, \Sigma) \), where \( \mu = (1 \ 5 \ 2)^T \) and
\[
\Sigma = \begin{bmatrix}
1 & 1 & 0 \\
1 & 4 & 0 \\
0 & 0 & 9
\end{bmatrix},
\]
(a) Find the pdfs of
i. \( X_1 \),
ii. \( X_2 + X_3 \),
iii. \( 2X_1 + X_2 + X_3 \),
iv. \( X_3 \) given \( X_1, X_2 \), and
v. \( (X_2, X_3) \) given \( X_1 \).
(b) What is \( P\{2X_1 + X_2 - X_3 < 0\} \)? Express your answer using the \( Q \) function.
(c) Find the joint pdf on \( Y = AX \), where
\[
A = \begin{bmatrix}
2 & 1 & 1 \\
1 & -1 & 1
\end{bmatrix}.
\]

**Solution:**
(a) i. The marginal pdfs of a jointly Gaussian pdf are Gaussian. Therefore \( X_1 \sim \mathcal{N}(1, 1) \).
ii. Since \( X_2 \) and \( X_3 \) are independent \( (\sigma_{23} = 0) \), the variance of the sum is the sum of the variances. Also the sum of two jointly Gaussian random variables is also Gaussian. Therefore \( X_2 + X_3 \sim \mathcal{N}(7, 13) \).
iii. Since \( 2X_1 + X_2 + X_3 \) is a linear transformation of a Gaussian random vector,
\[
2X_1 + X_2 + X_3 = \begin{bmatrix} 2 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix},
\]

it is a Gaussian random vector with mean and variance
\[
\mu = \begin{bmatrix} 2 & 1 & 1 \\ 5 & 2 \end{bmatrix} = 9 \quad \text{and} \quad \sigma^2 = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 4 & 0 \\ 0 & 0 & 9 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 21.
\]
Thus \( 2X_1 + X_2 + X_3 \sim \mathcal{N}(9, 21) \).
iv. Since \( \sigma_{13} = 0 \), \( X_3 \) and \( X_1 \) are uncorrelated and hence independent since they are jointly Gaussian; similarly, since \( \sigma_{23} = 0 \), \( X_3 \) and \( X_2 \) are independent. Therefore the conditional pdf of \( X_3 \) given \((X_1, X_2)\) is the same as the pdf of \( X_3 \), which is \( \mathcal{N}(2, 9) \).

v. We use the general formula for the conditional Gaussian pdf:

\[
X_2 | \{X_1 = x_1\} \sim \mathcal{N} \left( \Sigma_{12} \Sigma_{11}^{-1} (x - \mu_1) + \mu_2, \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \right)
\]

In the case of \((X_2, X_3) | X_1\),

\[
\Sigma_{11} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \Sigma_{21} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \Sigma_{22} = \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix}.
\]

Therefore the mean and variance of \((X_2, X_3)\) given \( X_1 = x_1 \) are

\[
\mu_{(X_2,X_3)|X_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} [1]^{-1} [x_1 - 1] + \begin{bmatrix} 5 \\ 2 \end{bmatrix} = \begin{bmatrix} x_1 + 4 \\ 2 \end{bmatrix},
\]

\[
\Sigma_{(X_2,X_3)|X_1} = \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 9 \end{bmatrix}.
\]

Thus \( X_2 \) and \( X_3 \) are conditionally independent given \( X_1 \). The conditional densities are \( X_2 | \{X_1 = x_1\} \sim \mathcal{N}(x_1 + 4, 3) \) and \( X_3 | \{X_1 = x\} \sim \mathcal{N}(2, 9) \).

(b) Let \( Y = 2X_1 + X_2 - X_3 \). Similarly as part (a)iii., \( 2X_1 + X_2 - X_3 \) is a linear transformation of a Gaussian random vector,

\[
2X_1 + X_2 - X_3 = \begin{bmatrix} 2 & 1 & -1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix},
\]

it is a Gaussian random vector with mean and variance

\[
\mu = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} \begin{bmatrix} 1 \\ 5 \\ 2 \end{bmatrix} = 5 \quad \text{and} \quad \sigma^2 = \begin{bmatrix} 2 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 4 & 0 \\ 0 & 0 & 9 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} = 21.
\]

Thus \( 2X_1 + X_2 - X_3 \sim \mathcal{N}(5, 21) \), i.e., \( Y \sim \mathcal{N}(5, 21) \). Thus

\[
P\{Y < 0\} = P \left\{ \frac{(Y - 5)}{\sqrt{21}} < \frac{(0 - 5)}{\sqrt{21}} \right\} = Q \left( \frac{5}{\sqrt{21}} \right).
\]

(c) In general, \( AX \sim \mathcal{N}(A\mu_X, A\Sigma_X A^T) \). For this problem,

\[
\mu_Y = A\mu_X = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} \begin{bmatrix} 1 \\ 5 \\ 2 \end{bmatrix} = \begin{bmatrix} 9 \\ -2 \end{bmatrix},
\]

\[
\Sigma_Y = A\Sigma_X A^T = \begin{bmatrix} 2 & 1 & -1 \\ 1 & 4 & 0 \\ 0 & 0 & 9 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 21 & 6 \\ 6 & 12 \end{bmatrix}.
\]

Thus \( Y \sim \mathcal{N} \left( \begin{bmatrix} 9 \\ -2 \end{bmatrix}, \begin{bmatrix} 21 & 6 \\ 6 & 12 \end{bmatrix} \right) \).
6. Additive nonwhite Gaussian noise channel.

Let \( Y_i = X + Z_i \) for \( i = 1, 2, \ldots, n \) be \( n \) observations of a signal \( X \sim N(0, P) \). The additive noise random variables \( Z_1, Z_2, \ldots, Z_n \) are zero mean jointly Gaussian random variables that are independent of \( X \) and have correlation \( \mathbb{E}(Z_i Z_j) = N \cdot 2^{-|i-j|} \) for \( 1 \leq i, j \leq n \).

(a) Find the best MSE estimate of \( X \) given \( Y_1, Y_2, \ldots, Y_n \).

(b) Find the MSE of the estimate in part (a).

Hint: the coefficients for the best estimate are of the form \( h^T = [a \ b \ b \ \cdots \ b \ b \ a] \).

Solution:

(a) Since \( X, Y_1, Y_2, \ldots, Y_n \) are jointly Gaussian and \( \mathbb{E}[X] = 0 \) and \( \mathbb{E}[Y] = 0 \), the best estimate of \( X \) is of the form

\[
\hat{X} = \sum_{i=1}^{n} h_i Y_i.
\]

The vector \( h \) is determined by the equation

\[
\mathbb{E}[YX] = \mathbb{Cov}(Y)h. \tag{2}
\]

We have, for \( 1 \leq i, j \leq n \), \( \mathbb{Cov}(Y_i, Y_j) = \mathbb{E}[Y_i Y_j] = \mathbb{E}[X^2] + \mathbb{E}[Z_i Z_j] = P + N \cdot 2^{-|i-j|} \).

Also, \( \mathbb{E}[Y_i X] = \mathbb{E}[(Z_i + X)X] = \mathbb{E}[X^2] = P \).

Thus, (2) reduces to \( n \) equations with \( n \) unknowns:

\[
\begin{bmatrix}
P & \cdots & P & P + N/2 & \cdots & P + N/2^{n-2} & P + N/2^{n-1} \\
\vdots & & \vdots & \vdots & \ddots & \vdots & \vdots \\
P & \cdots & P + N/2^{n-3} & P + N & \cdots & P + N/2 & P + N/2 \\
P & \cdots & P + N/2^{n-2} & P + N/2^{n-1} & \cdots & P + N/2 & P + N
\end{bmatrix}
\begin{bmatrix}
h_1 \\
h_2 \\
\vdots \\
h_n-1 \\
h_n
\end{bmatrix}
= \begin{bmatrix}
P \\
1 \\
\vdots \\
1 \\
2
\end{bmatrix}.
\]

By the hint, there are only 2 degrees of freedom given, \( a \) and \( b \). Solving this equation using the first 2 rows of the matrix, we obtain

\[
\begin{bmatrix}
h_1 \\
h_2 \\
\vdots \\
h_{n-1} \\
h_n
\end{bmatrix} = \frac{P}{3N + (n + 2)P} \begin{bmatrix}
2 \\
1 \\
\vdots \\
1 \\
2
\end{bmatrix}.
\]
(b) The minimum mean square error is
\[
\text{MSE} = E[(X - \hat{X})^2] \\
= E\left[\left(X - \sum_{i=1}^{n} h_i Y_i\right)^2\right] \\
= E[X^2] + h^T E[YY^T]h - 2\sum_{i=1}^{n} h_i E[XY_i] \\
= P + h^T E[XY] - 2h^T E[YX] \\
= P - P\sum_{i=1}^{n} h_i \\
= P - P\cdot \frac{(n + 2)P}{3N + (n + 2)P} \\
= \frac{3PN}{3N + (n + 2)P}.
\]

7. Let \(X\) and \(Y\) be two random variables. Let \(Z = X + Y\) and let \(W = X - Y\). Find the best linear estimate of \(W\) given \(Z\) as a function of \(E(X), E(Y), \sigma_X, \sigma_Y, \rho_{XY}\) and \(Z\).

**Solution:** The best linear estimate of \(W\) given \(Z\) is
\[
\hat{W}(Z) = E[W] + h^*(Z - E[Z]) \\
\]
where
\[
h^* = \text{Var}(Z)^{-1} \text{Cov}(Z, W) \\
= \frac{\text{Cov}(X + Y, X - Y)}{\text{Var}(X + Y)} \\
= \frac{\text{Var}(X) - \text{Cov}(X, Y) + \text{Cov}(Y, X) - \text{Var}(Y)}{\text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)} \\
= \frac{\sigma_X^2 - \sigma_Y^2}{\sigma_X^2 + \sigma_Y^2 + 2\rho_{XY}\sigma_X\sigma_Y}.
\]

Thus, \(\hat{W}(Z) = E[X] - E[Y] + \left(\frac{\sigma_X^2 - \sigma_Y^2}{\sigma_X^2 + \sigma_Y^2 + 2\rho_{XY}\sigma_X\sigma_Y}\right)(Z - E[X] - E[Y]),\) which gives, on simplification,
\[
\hat{W}(Z) = \left(\frac{\sigma_X^2 - \sigma_Y^2}{\sigma_X^2 + \sigma_Y^2 + 2\rho_{XY}\sigma_X\sigma_Y}\right)Z + \frac{2E[X]\left(\sigma_Y^2 + \rho_{XY}\sigma_X\sigma_Y\right) - 2E[Y]\left(\sigma_X^2 + \rho_{XY}\sigma_X\sigma_Y\right)}{\sigma_X^2 + \sigma_Y^2 + 2\rho_{XY}\sigma_X\sigma_Y}.
\]

8. Let \(X\) and \(Y\) be two random variables with joint pdf
\[
f(x, y) = \begin{cases} 
  x + y, & 0 \leq x \leq 1, 0 \leq y \leq 1, \\
  0, & \text{otherwise}.
\end{cases}
\]
(a) Find the MMSE estimator of $X$ given $Y$.
(b) Find the corresponding MSE.
(c) Find the pdf of $Z = \mathbb{E}(X|Y)$.
(d) Find the linear MMSE estimator of $X$ given $Y$.
(e) Find the corresponding MSE.

Solution:

(a) We have $f_Y(y) = \int_0^1 f(x, y) dx = y + 1/2$.

Thus, $f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{x + y}{1/2 + y}$, $(x, y) \in [0, 1]^2$.

Hence,

$$\mathbb{E}[X|Y = y] = \int_0^1 xf_{X|Y}(x|y) dx = \int_0^1 \frac{x(x + y)}{1/2 + y} dx = \frac{2 + 3y}{3(1 + 2y)}.$$ 

Thus, MMSE estimator of $X$ given $Y$ is $\mathbb{E}[X|Y] = \frac{2 + 3Y}{3(1 + 2Y)}$.

(b) We have

$$\mathbb{E}[X^2|Y = y] = \int_0^1 x^2 f_{X|Y}(x|y) dx = \int_0^1 \frac{x^2(x + y)}{1/2 + y} dx = \frac{3 + 4y}{6(1 + 2y)}.$$ 

Thus,

$$\text{Var}(X|Y = y) = \mathbb{E}[X^2|Y = y] - \left(\mathbb{E}[X|Y = y]\right)^2 = \frac{3 + 4y}{6(1 + 2y)} - \left(\frac{2 + 3y}{3(1 + 2y)}\right)^2 = \frac{6y^2 + 6y + 1}{18(1 + 2y)^2}.$$
Thus, the MSE of the MMSE estimator of $X$ given $Y$ is

$$E[\text{Var}(X|Y)] = E\left[\frac{6Y^2 + 6Y + 1}{18(1 + 2Y)^2}\right]$$

$$= \int_0^1 \frac{1}{18(1 + 2y)^2} (6y^2 + 6y + 1) \, dy$$

$$= \int_0^1 \frac{1}{36(1 + 2y)} \, dy$$

$$= \int_0^1 \frac{3}{2} \frac{(1 + 2y)^2 - \frac{1}{2}}{36(1 + 2y)} \, dy$$

$$= \frac{1}{24} \int_0^1 (1 + 2y) \, dy - \frac{1}{144} \int_0^1 \frac{dy}{y + 1/2}$$

$$= \frac{1}{12} - \frac{\ln 3}{144}.$$

(c) We have $Z = E[X|Y] = \frac{2 + 3Y}{3(1 + 2Y)} = \frac{1}{2} + \frac{1}{6(1 + 2Y)}$.

The function $g : [0, 1] \rightarrow [5/9, 2/3]$ defined by $g(t) = \frac{1}{2} + \frac{1}{6(1 + 2t)}$ is continuous and strictly decreasing, hence it is invertible and has an inverse $h : [5/9, 2/3] \rightarrow [0, 1]$ given by

$$h(t) = \frac{1}{6(2t - 1) - \frac{1}{2}}.$$

Thus,

$$f_Z(z) = f_Y(h(z))|h'(z)|$$

$$= \left(\frac{1}{6(2z - 1)}\right) \left(\frac{1}{3(2z - 1)^2}\right)$$

$$= \frac{1}{18(2z - 1)^3}, \quad z \in [5/9, 2/3].$$

(d) We have

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$$

$$= \int_0^1 \int_0^1 xy(x + y) \, dx \, dy - \left(\int_0^1 x(x + 1/2) \, dx\right)^2$$

$$= \frac{1}{3} - \left(\frac{7}{12}\right)^2$$

$$= -\frac{1}{144}, \quad \text{and}$$
\[
\text{Var}(Y) = \text{E}[Y^2] - \left(\text{E}[Y]\right)^2
\]
\[
= \int_0^1 y^2(y+1/2)dy - \left(\int_0^1 y(y+1/2)dy\right)^2
\]
\[
= \frac{5}{12} - \left(\frac{7}{12}\right)^2
\]
\[
= \frac{11}{144}.
\]

Thus, the linear MMSE estimator of \(X\) given \(Y\) is given by
\[
\text{E}[X] + \text{Var}(Y)^{-1} \text{Cov}(X,Y)(Y - \text{E}[Y]) = \frac{7}{12} - \frac{1}{11}(Y - \frac{7}{12})
\]
\[
= \frac{7 - Y}{11}.
\]

(e) The MSE of the linear MMSE estimator of \(X\) given \(Y\) is
\[
\text{Var}(X) - \frac{\text{Cov}(X,Y)^2}{\text{Var}(Y)} = \frac{11}{144} - \left(\frac{1}{144}\right)^2
\]
\[
= \frac{5}{66},
\]
which is very close to the optimal value found in part (b).

9. **Linear estimator.** Consider a channel with the observation \(Y = XZ\), where the signal \(X\) and the noise \(Z\) are uncorrelated Gaussian random Variables. Let \(E[X] = 1\), \(E[Z] = 2\), \(\sigma_X^2 = 5\), and \(\sigma_Z^2 = 8\).

(a) Find the MMSE linear estimate of \(X\) given \(Y\).

(b) Suppose your friend from Caltech tells you that he was able to derive an estimator with a lower MSE. Your friend from MIT disagrees, saying that this is not possible because the signal and the noise are Gaussian, and hence the MMSE linear estimator will also be the MMSE estimator. Could your MIT friend be wrong?

**Solution:**

(a) We know that the best linear estimate is given by the formula
\[
\hat{X} = \frac{\text{Cov}(X,Y)}{\sigma_Y^2}(Y - \text{E}(Y)) + \text{E}(X).
\]

Note that \(X\) and \(Z\) Gaussian and uncorrelated implies they are independent. Therefore,
\[
E(Y) = E(XZ) = E(X)E(Z) = 2,
\]
\[
E(XY) = E(X^2Z) = E(X^2)E(Z) = (\sigma_X^2 + E^2(X))E(Z) = 12,
\]
\[
E(Y^2) = E(X^2Z^2) = E(X^2)E(Z^2) = (\sigma_X^2 + E^2(X))(\sigma_Z^2 + E^2(Z)) = 72,
\]
\[
\sigma_Y^2 = E(Y^2) - E^2(Y) = 68,
\]
\[
\text{Cov}(X,Y) = E(XY) - E(X)E(Y) = \frac{5}{34}.
\]
Using all of the above, we get
\[ \hat{X} = \frac{5}{34} Y + \frac{12}{17}. \]

(b) The fact that the best linear estimate equals the best MMSE estimate when input and noise are independent Gaussians is only known to be true for additive channels. For multiplicative channels this need not be the case in general. In the following, we prove \( Y \) is not Gaussian by contradiction.

Suppose \( Y \) is Gaussian, then \( Y \sim N(2, 68) \). We have
\[ f_Y(y) = \frac{1}{\sqrt{2\pi \times 68}} e^{-\frac{(y-2)^2}{2 \times 68}}. \]

On the other hand, as a function of two random Variables, \( Y \) has pdf
\[ f_Y(y) = \int_{-\infty}^{\infty} f_X(x)f_Z\left(\frac{y}{x}\right)\,dx. \]

But these two expressions are not consistent, because
\[
\begin{align*}
f_Y(0) &= \int_{-\infty}^{\infty} f_X(x)f_Z\left(\frac{0}{x}\right)\,dx = f_Z(0) \int_{-\infty}^{\infty} f_X(x)\,dx = f_Z(0) \\
&= \frac{1}{\sqrt{2\pi \times 8}} e^{-\frac{(0-2)^2}{2 \times 8}} \\
&\neq \frac{1}{\sqrt{2\pi \times 68}} e^{-\frac{(0-2)^2}{2 \times 68}} = f_Y(0),
\end{align*}
\]
which is a contradiction. Hence, \( X \) and \( Y \) are not joint Gaussian, and we might be able to derive an estimator with a lower MSE.

10. Additive-noise channel with path gain. Consider the additive noise channel shown in the figure below, where \( X \) and \( Z \) are zero mean and uncorrelated, and \( a \) and \( b \) are constants.

Find the MMSE linear estimate of \( X \) given \( Y \) and its MSE in terms only of \( \sigma_X, \sigma_Z, a, \) and \( b \).

Solution: By the theorem of MMSE linear estimate, we have
\[ \hat{X} = \frac{\text{Cov}(X, Y)}{\sigma_Y^2} (Y - \text{E}(Y)) + \text{E}(X). \]
Since $X$ and $Z$ are zero mean and uncorrelated, we have

\[
E(X) = 0,
\]
\[
E(Y) = b(aE(X) + E(Z)) = 0,
\]
\[
\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = E(Xb(aX + Z)) = ab\sigma_X^2,
\]
\[
\sigma_Y^2 = E(Y^2) - (E(Y))^2 = E(b^2(aX + Z)^2) = b^2a^2\sigma_X^2 + b^2\sigma_Z^2.
\]

Hence, the best linear MSE estimate of $X$ given $Y$ is given by

\[
\hat{X} = \frac{a\sigma_X^2}{b^2a^2\sigma_X^2 + b\sigma_Z^2}Y.
\]

11. \textit{Worst noise distribution.} Consider an additive noise channel $Y = X + Z$, where the signal $X \sim \mathcal{N}(0, P)$ and the noise $Z$ has zero mean and variance $N$. Assume $X$ and $Z$ are independent.

Find a distribution of $Z$ that maximizes the minimum MSE of estimating $X$ given $Y$, i.e., the distribution of the worst noise $Z$ that has the given mean and variance. You need to justify your answer.

\textbf{Solution:} Let us calculate the MSE of the linear MMSE estimator of $X$ given $Y$, which will serve as an upper bound on the minimum MSE of estimating $X$ given $Y$.

We have, since $X$ and $Z$ are independent,

\[
\text{Var}(Y) = \text{Var}(X + Z) = \text{Var}(X) + \text{Var}(Z) = P + N, \text{ and}
\]
\[
\text{Cov}(X, Y) = \text{Cov}(X, X + Z) = \text{Var}(X) + \text{Cov}(X, Z) = \text{Var}(X) = P.
\]

Thus, the MSE of the linear MMSE estimator of $X$ given $Y$ is

\[
\text{Var}(X) - \frac{\left(\text{Cov}(X, Y)\right)^2}{\text{Var}(Y)} = P - \frac{P^2}{P + N} = \frac{PN}{P + N}.
\]

Thus, the minimum MSE of estimating $X$ given $Y$ is upper-bounded by $\frac{PN}{P + N}$, and we also know that the linear MMSE estimator is the same as the MMSE estimator when $X$ and $Y$ are jointly Gaussian.

$X$ and $Y$ are jointly Gaussian if $Z \sim \mathcal{N}(0, N)$, and this is thus the noise distribution that
makes the MSE of the linear MMSE estimator the same as the overall minimum MSE.

Thus, the distribution $Z \sim \mathcal{N}(0, N)$ maximizes the minimum MSE of estimating $X$ given $Y$.

12. **Image processing.** A pixel signal $X \sim U[-k, k]$ is digitized to obtain

$\tilde{X} = i + \frac{1}{2}$, if $i < X \leq i + 1$, $i = -k, -k + 1, \ldots, k - 2, k - 1$.

To improve the visual appearance, the digitized value $\tilde{X}$ is dithered by adding an independent noise $Z$ with mean $E(Z) = 0$ and variance $\text{Var}(Z) = N$ to obtain $Y = \tilde{X} + Z$.

(a) Find the correlation of $X$ and $Y$.

(b) Find the MMSE linear estimate of $X$ given $Y$. Your answer should be in terms only of $k$, $N$, and $Y$.

**Solution:**

(a) From the definition of $\tilde{X}$, we know $P\{\tilde{X} = i + \frac{1}{2}\} = P\{i < X \leq i + 1\} = \frac{1}{2k}$. By the law of total expectation, we have

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = E(X(\tilde{X} + Z)) = E(X\tilde{X})$$

$$= \sum_{i=-k}^{k-1} E[X\tilde{X} | i < X \leq i + 1]P(i < X \leq i + 1)$$

$$= \sum_{i=-k}^{k-1} \int_{i}^{i+1} x(i + \frac{1}{2}) \frac{1}{2k} dx = \frac{1}{8k} \sum_{i=-k}^{k-1} (2i + 1)^2 = \frac{1}{4k} \sum_{i=1}^{k} (2i - 1)^2$$

$$= \frac{4k^2 - 1}{12}.$$

Since, $\sum_{i=1}^{k} i^2 = k(k + 1)(2k + 1)/6$.

(b) We have

$E(X) = 0$,

$E(Y) = E(\tilde{X}) + E(Z) = 0$,

$\sigma^2_Y = \text{Var}\tilde{X} + \text{Var}Z = \sum_{i=-k}^{k-1} (i + \frac{1}{2})^2 \frac{1}{2k} + N = \frac{1}{4k} \sum_{i=0}^{k-1} (2i + 1)^2 + N = \frac{4k^2 - 1}{12} + N$.

Then, the best linear MMSE estimate of $X$ given $Y$ is given by

$$\hat{X} = \frac{\text{Cov}(X, Y)}{\sigma^2_Y} (Y - E(Y)) + E(X) = \frac{\frac{4k^2 - 1}{12}}{\frac{4k^2 - 1}{12} + N} Y$$

$$= \frac{4k^2 - 1}{4k^2 - 1 + 12N} Y.$$
13. **Noise cancellation.** A classical problem in statistical signal processing involves estimating a weak signal (e.g., the heartbeat of a fetus) in the presence of a strong interference (the heartbeat of its mother) by making two observations; one with the weak signal present and one without (by placing one microphone on the mother’s belly and another close to her heart). The observations can then be combined to estimate the weak signal by “canceling out” the interference. The following is a simple version of this application.

Let the weak signal $X$ be a random variable with mean $\mu$ and variance $P$, and the observations be $Y_1 = X + Z_1$ ($Z_1$ being the strong interference), and $Y_2 = Z_1 + Z_2$ ($Z_2$ is a measurement noise), where $Z_1$ and $Z_2$ are zero mean with variances $N_1$ and $N_2$, respectively. Assume that $X, Z_1$ and $Z_2$ are uncorrelated. Find the MMSE linear estimate of $X$ given $Y_1$ and $Y_2$ and its MSE. Interpret the results.

**Solution:** This is a vector linear MSE problem. Since $Z_1$ and $Z_2$ are zero mean, $\mu_X = \mu_{Y_1} = \mu$ and $\mu_{Y_2} = 0$. We first normalize the random variables by subtracting off their means to get $X' = X - \mu$, and

$$Y' = \begin{bmatrix} Y_1 - \mu \\ Y_2 \end{bmatrix}.$$ 

Now using the orthogonality principle we can find the best linear MSE estimate $\hat{X}'$ of $X'$. To do so we first find

$$\Sigma_Y = \begin{bmatrix} P + N_1 & N_1 \\ N_1 & N_1 + N_2 \end{bmatrix} \quad \text{and} \quad \Sigma_{YX} = \begin{bmatrix} P \\ 0 \end{bmatrix}.$$ 

Thus,

$$\hat{X}' = \Sigma_{YX} \Sigma_Y^{-1} Y' = \begin{bmatrix} P \\ 0 \end{bmatrix} \frac{1}{P(N_1 + N_2) + N_1 N_2} \begin{bmatrix} N_1 + N_2 \\ -N_1 \\ -P + N_1 \end{bmatrix} Y' = \frac{P}{P(N_1 + N_2) + N_1 N_2} \begin{bmatrix} (N_1 + N_2) \\ -N_1 \end{bmatrix} Y'.$$

The best linear MSE estimate is $\hat{X} = \hat{X}' + \mu$. Thus,

$$\hat{X} = \frac{P}{P(N_1 + N_2) + N_1 N_2} ((N_1 + N_2)(Y_1 - \mu) - N_1 Y_2) + \mu$$

$$= \frac{1}{P(N_1 + N_2) + N_1 N_2} (P((N_1 + N_2)Y_1 - N_1 Y_2)) + N_1 N_2 \mu).$$

The MSE can be calculated by

$$\text{MSE} = \sigma_X^2 - \Sigma_{YX}^T \Sigma_Y^{-1} \Sigma_{YX}$$

$$= P - \frac{P}{P(N_1 + N_2) + N_1 N_2} \begin{bmatrix} (N_1 + N_2) \\ -N_1 \end{bmatrix} \begin{bmatrix} P \\ 0 \end{bmatrix}$$

$$= P - \frac{P^2(N_1 + N_2)}{P(N_1 + N_2) + N_1 N_2} \frac{N_1 N_2}{P N_1 N_2}$$

$$= \frac{P(N_1 + N_2) + N_1 N_2}{P(N_1 + N_2) + N_1 N_2}.$$
The equation for the MSE makes perfect sense. First, note that if \( N_1 \) and \( N_2 \) are held constant but \( P \) goes to infinity, the MSE tends to \( \frac{N_1 N_2}{N_1 + N_2} \). Next, note that if both \( N_1 \) and \( N_2 \) go to infinity, the MSE goes to \( \sigma_X^2 \), i.e., the estimate becomes worthless. Finally, note that if either \( N_1 \) or \( N_2 \) goes to 0, the MSE also goes to 0. This is because the estimator will then use the measurement with zero noise Variance and perfectly determine the signal \( X \).

14. **Nonlinear estimator.** Consider a channel with the observation \( Y = XZ \), where the signal \( X \) and the noise \( Z \) are uncorrelated Gaussian random Variables. Let \( E[X] = 1 \), \( E[Z] = 2 \), \( \sigma_X^2 = 5 \), and \( \sigma_Z^2 = 8 \).

(a) Using the fact that
\[ E(W^3) = \mu + 3\mu\sigma^2 \quad \text{and} \quad E(W^4) = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 \]
for \( W \sim \mathcal{N}(\mu, \sigma^2) \), find the mean and CoVariance matrix of \( [X \ Y \ Y^2]^T \).

(b) Find the MMSE linear estimate of \( X \) given \( Y \) and the corresponding MSE.

(c) Find the MMSE linear estimate of \( X \) given \( Y^2 \) and the corresponding MSE.

(d) Find the MMSE linear estimate of \( X \) given \( Y \) and \( Y^2 \) and the corresponding MSE.

(e) Compare your answers in parts (b) through (d). Is the MMSE estimate of \( X \) given \( Y \) (namely, \( E(X|Y) \)) linear? This would answer Problem 1 in Homework Set #6.

**Solution:**

(a) Since \( X \) and \( Z \) are uncorrelated Gaussian random Variables, they are independent. We have
\[
E(X^2) = \sigma_X^2 + E^2(X) = 5 + 1 = 6, \\
E(X^3) = 1 + 3 \times 1 \times 5 = 16, \\
E(X^4) = 1 + 6 \times 1 \times 5 + 3 \times 25 = 106.
\]

\[
E(Z^2) = \sigma_Z^2 + E^2(Z) = 8 + 4 = 12, \\
E(Z^3) = 2 + 3 \times 2 \times 8 = 50, \\
E(Z^4) = 2^4 + 6 \times 4 \times 8 + 3 \times 64 = 400.
\]

Since \( X \) and \( Z \) are independent, we have
\[
E(Y) = E(XZ) = E(X)E(Z) = 2, \\
E(Y^2) = E(X^2Z^2) = E(X^2)E(Z^2) = 6 \times 12 = 72, \\
E(Y^3) = E(X^3Z^3) = E(X^3)E(Z^3) = 16 \times 50 = 800, \\
E(Y^4) = E(X^4)E(Z^4) = 106 \times 400 = 42400.
\]

Therefore, the mean of \( [X \ Y \ Y^2]^T \) is \( [1 \ 2 \ 72]^T \).
\[ \text{Var}(Y) = E(Y^2) - E^2(Y) = 72 - 4 = 68, \]
\[ \text{Var}(Y^2) = E(Y^4) - E^2(Y^2) = 37216, \]
\[ \text{Cov}(X, Y) = E(XY) - E(X)E(Y) = E(X^2)E(Z) - E(X)E(Y) = 10, \]
\[ \text{Cov}(X, Y^2) = E(XY^2) - E(X)E(Y^2) = E(X^3)E(Z^2) - E(X)E(Y^2) = 120, \]
\[ \text{Cov}(Y, Y^2) = E(YY^2) - E(Y)E(Y^2) = E(X^3)E(Z^3) - E(Y)E(Y^2) = 656. \]

Therefore, the CoVariance matrix of \([X\ Y\ Y^2]^T\) is
\[
\begin{bmatrix}
5 & 10 & 120 \\
10 & 68 & 656 \\
120 & 656 & 37216
\end{bmatrix}.
\]

(b) The MMSE linear estimate of \(X\) given \(Y\) is
\[
\hat{X} = \frac{\text{Cov}(X, Y)}{\text{Var}(Y)}(Y - E(Y)) + E(X) = \frac{10}{68}(Y - 2) + 1 = \frac{5}{34}Y + \frac{24}{34},
\]
and its MSE is given by
\[
\text{MSE} = \text{Var}(X) - \frac{\text{Cov}^2(X, Y)}{\text{Var}(Y)} = 5 - \frac{100}{68} = 3.5294.
\]

(c) The MMSE linear estimate of \(X\) given \(Y^2\) is
\[
\hat{X} = \frac{\text{Cov}(X, Y^2)}{\text{Var}(Y^2)}(Y^2 - E(Y^2)) + E(X) = \frac{120}{37216}(Y^2 - 72) + 1 = \frac{15}{4652}Y^2 + \frac{893}{1163},
\]
and its MSE is given by
\[
\text{MSE} = \text{Var}(X) - \frac{\text{Cov}^2(X, Y^2)}{\text{Var}(Y^2)} = 5 - \frac{14400}{37216} = 4.6131.
\]

(d) We first normalize the random Variables by subtracting off their means to get
\[
X' = X - E(X) = X - 2, \\
Y' = Y - E(Y) = Y - 2, \\
Y'^2 = Y^2 - E(Y^2) = Y^2 - 72.
\]

Using the covariance matrix in part a, we have
\[
\Sigma_{[Y\ Y^2]^T|X} = \begin{bmatrix} 10 & 120 \end{bmatrix}^T, \\
\Sigma_{[Y\ Y^2]^T} = \begin{bmatrix} 68 & 656 \\ 656 & 37216 \end{bmatrix}.
\]
Therefore,
\[ \hat{X}' = \Sigma_{[Y \ y^2]T} X \Sigma_{[Y \ y^2]T}^{-1} \begin{bmatrix} Y' \\ Y'^2 \end{bmatrix} = 0.1397Y' + 0.0008Y'^2, \]
and hence
\[ \hat{X} = \hat{X}' + X = 0.1397(Y - 2) + 0.0008(Y^2 - 72) + 1 = 0.1397Y + 0.0008Y^2 + 0.663. \]
The corresponding MSE is given by
\[ \text{MSE} = \text{Var}(X) - \Sigma_{[Y \ y^2]T} X \Sigma_{[Y \ y^2]T}^{-1} \Sigma_{[Y \ y^2]T} X = 3.5115. \]

(e) MSE linear estimate of \( X \) given \( Y \) and \( Y^2 \) results in the minimum MSE among the three. Therefore, MSE linear estimate of \( X \) given \( Y \) does not have the minimum MSE and MMSE estimate of \( X \) given \( Y \) is not linear.

15. **Additive shot noise channel.** Consider an additive noise channel \( Y = X + Z \), where the signal \( X \sim \mathcal{N}(0, 1) \), and the noise \( Z|\{X = x\} \sim \mathcal{N}(0, x^2) \), i.e., the noise power of increases linearly with the signal squared.

(a) Find \( E(Z^2) \).
(b) Find the best linear MSE estimate of \( X \) given \( Y \).

**Solution:**

(a) \( E[Z^2|X = x] = x^2 \), and thus
\[
\]

(b) Given \( X = x \), \( Y = x + Z \sim \mathcal{N}(x, x^2) \).

Thus, \( E[Y|X] = X \) and \( \text{Var}(Y|X) = X^2 \). Hence,
\[
\text{Var}(Y) = E[\text{Var}(Y|X)] + \text{Var}(E[Y|X]) = E[X^2] + \text{Var}(X) = 2.
\]
We also have \( E[Y] = E[E[Y|X]] = 0. \)

Moreover, \( f_{Z|X}(z|x) = \frac{1}{\sqrt{2\pi x^2}} \exp \left[ -\frac{z^2}{2x^2} \right] \), and hence
\[
f_{X,Z}(x, z) = f_X(x)f_{Z|X}(z|x) = \frac{1}{2\pi|x|} \exp \left[ -\frac{1}{2} \left( x^2 + \frac{z^2}{x^2} \right) \right].
\]

18
Thus,

$$\text{Cov}(X, Y) = \text{Cov}(X, X + Z) = \text{Var}(X) + \text{Cov}(X, Z) = 1 + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x z \frac{1}{2\pi|x|} \exp \left( -\frac{1}{2} \left( x^2 + \frac{z^2}{x^2} \right) \right) \, dx \, dz = 1 + 0 = 1,$$

where the integral becomes zero since for each $x$, the integrand is an odd function of $z$.

Thus, the best linear MSE estimate of $X$ given $Y$ is

$$\text{E}[X] + \frac{\text{Cov}(X, Y)}{\text{Var}(Y)}(Y - \text{E}[Y]) = 0 + \frac{1}{2}(Y - 0) = \frac{Y}{2}.$$

16. Estimation vs. detection. Let the signal

$$X = \begin{cases} +1, & \text{with probability } \frac{1}{2} \\ -1, & \text{with probability } \frac{1}{2}, \end{cases}$$

and the noise $Z \sim \text{Unif}[-2, 2]$ be independent random variables. Their sum $Y = X + Z$ is observed.

(a) Find the best MSE estimate of $X$ given $Y$ and its MSE.

(b) Now suppose we use a decoder to decide whether $X = +1$ or $X = -1$ so that the probability of error is minimized. Find the optimal decoder and its probability of error. Compare the optimal decoder’s MSE to the minimum MSE.

Solution:

(a) We have

$$f_{Y|X}(y | +1) = f_{1+Z}(y) = f_Z(y - 1) = \frac{1}{4} \mathbb{1}_{[-1,3]}(y),$$

and

$$f_{Y|X}(y | -1) = f_{Z-1}(y) = f_Z(y + 1) = \frac{1}{4} \mathbb{1}_{[-3,1]}(y).$$
Thus,

\[
p_{X|Y}(1|y) = \frac{f_{Y|X}(y|1)}{\sum_{x \in \{-1, 1\}} f_{Y|X}(y|x)}
\]

\[
= \frac{1_{[-1,3]}(y)}{1_{[-1,3]}(y) + 1_{[-3,1]}(y)}
\]

\[
= \begin{cases} 
0, & -3 \leq y < -1 \\
1/2, & -1 \leq y < 1 \\
1, & 1 \leq y < 3.
\end{cases}
\]

Similarly,

\[
p_{X|Y}(-1|y) = \begin{cases} 
1, & -3 \leq y < -1 \\
1/2, & -1 \leq y < 1 \\
0, & 1 \leq y < 3.
\end{cases}
\]

Thus, the best MSE estimate of $X$ given $Y = y$ is

\[
E[X|Y = y] = \begin{cases} 
-1, & -3 \leq y < -1 \\
0, & -1 \leq y < 1 \\
1, & 1 \leq y < 3.
\end{cases}
\]

We also have $E[X^2|Y = y] = 1$ and thus,

\[
\text{Var}(X|Y) = 1_{[-1,1]}(Y).
\]

Thus the MSE of the best MSE estimate of $X$ given $Y$ is

\[
E[\text{Var}(X|Y)] = P(Y \in [-1, 1])
\]

\[
= P(Y \in [-1, 1]|X = +1)P(X = +1) + P(Y \in [-1, 1]|X = -1)P(X = -1)
\]

\[
= 1/2.
\]

(b) From the posterior probabilities computed in the previous part, we see that

\[
p_{X|Y}(+1|y) > p_{X|Y}(-1|y) \text{ if } y \in (1, 3], \text{ and } p_{X|Y}(+1|y) < p_{X|Y}(-1|y) \text{ if } y \in [-3, -1).
\]

Thus the decision rule should be

\[
g^*(y) = \begin{cases} 
+1, & y \in (1, 3] \\
-1, & y \in [-3, -1).
\end{cases}
\]

For $y \in [-1, 1]$, the posteriors are equal, and hence, let us arbitrarily choose $g^*(y) = +1$ for $y \in [-1, 1]$. 

20
For this decoding rule, there is no error if $X = +1$, and so, the probability of error is given by

$$P\left(\{\text{error}\}\right) = P\left(Y \notin [-3, -1] \mid X = -1\right)P(X = +1)$$

$$= \frac{1}{2} \times \frac{1}{2}$$

$$= \frac{1}{4}.$$  

Since there is no error if $X = +1$, the decoder’s MSE is given by

$$E[(g^*(Y) - X)^2 \mid X = -1]P(X = -1) = 4P\left(Y \in [-1, 1] \mid X = -1\right)P(X = -1)$$

$$= 4 \times \frac{1}{2} \times \frac{1}{2}$$

$$= 1.$$  

As expected, this is larger than the minimum MSE.

17. **Prediction.** Let $X$ be a random process with zero mean and covariance matrix

$$\Sigma_X = \begin{bmatrix}
1 & \alpha & \alpha^2 & \cdots & \alpha^{n-1} \\
\alpha & 1 & \alpha & \cdots & \\
\alpha^2 & \alpha & 1 & \cdots & \\
\vdots & \vdots & \ddots & \ddots & \\
\alpha^{n-1} & \cdots & & 1
\end{bmatrix}$$

for $\vert \alpha \vert < 1$. $X_1, X_2, \ldots, X_{n-1}$ are observed, find the best linear MSE estimate (predictor) of $X_n$. Compute its MSE.

**Solution:** Let $Y = (X_1, X_2, \ldots X_{n-1})$. Since $E[X_n] = 0$ and $E[Y] = 0$, the best linear MSE estimate of $X_n$ given $Y$ is given by $\hat{X}_n = h^T Y$, where the vector $h$ satisfies the equation

$$E[YX_n] = E[YY^T]h. \quad (3)$$

We see that $E[YY^T]$ is simply $\Sigma_X$ with the last row and last column removed, and $E[YX_n]$ is simply the last column of $\Sigma_X$, with the last element removed.

Thus, $E[YX_n] = \begin{bmatrix}\alpha^{n-1} \\
\alpha^{n-2} \\
\vdots \\
\alpha \end{bmatrix}$, and the last column of $E[YY^T]$ is $\begin{bmatrix}\alpha^{n-2} \\
\alpha^{n-3} \\
\vdots \\
1 \end{bmatrix}$.

Thus, $E[YX_n]$ is simply a constant multiple of the last column of $E[YY^T]$, and thus, (2) is solved by
\[ h = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \alpha \end{bmatrix}. \]

Hence, the best linear MSE estimate of \( X_n \) given \( Y \) is given by \( \hat{X}_n = \alpha X_{n-1} \).

The MSE of this estimate is
\[
E[(X_n - \alpha X_{n-1})^2] = E[X_n^2] + \alpha^2 E[X_{n-1}^2] - 2\alpha E[X_n X_{n-1}]
= 1 - \alpha^2.
\]