Problem 1  (17 points) A random variable $X$ is known to be the sum of $K$ independent and identically distributed exponential random variables, each with an expected value equal to $(K\lambda)^{-1}$. We have only two hypotheses for the value of parameter $K$; these are $H_0 (K = 64)$ and $H_1 (K = 400)$. Before we obtain any experimental data, our a priori guess is that these two hypotheses are equally likely.

The statistic for our hypothesis test is to be the sum of four independent experimental values of $X$. We estimate that false acceptance of $H_0$ will cost us $100$, false rejection of $H_0$ will cost us $200$, and any correct outcome is worth $500$ to us.

Determine approximately the rejection region for $H_0$ which maximizes the expected value of this hypothesis test.

Hint: The sum of $n$ independent and identically distributed exponential random variables is called an Erlang order-$n$ distribution. Your textbook discusses the Erlang family of distributions.

Solution  Let $X_i$ denote the $i$th experimental value of $X$. Let $Z = X_1 + X_2 + X_3 + X_4$, so that $Z$ is the random variable whose value we observe. For notation, let $f_{Z|H_0}(z|H_0)$ denote the conditional PDF of $Z$ given that $H_0$ is true. When $H_0$ is true, then each $X_i$ is a sum of 64 independent and identically distributed exponential random variables with parameter $64\lambda$. Since we are given that $Z$ is a sum of four independent experimental values of $X$, then $X_1, X_2, X_3, X_4$ are all independent. Thus, we can see that when $H_0$ is true, then $Z$ can be seen as the sum of $4(64) = 256$ independent and identically distributed exponential random variables with parameter $64\lambda$. Thus by definition, $Z$ is an Erlang order-256 random variable with the following PDF:

$$f_{Z|H_0}(z|H_0) = \frac{(64\lambda)^{256}z^{255}e^{-64\lambda z}}{255!},$$

where $\lambda$ is the parameter for each of the exponential random variables.

Similarly, let $f_{Z|H_1}(z|H_1)$ denote the conditional PDF of $Z$ given that $H_1$ is true. When $H_1$ is true, then $Z$ can be seen as the sum of $4(400) = 1600$ independent and identically distributed exponential random variables with parameter $400\lambda$. Thus by definition, $Z$ is an Erlang order-1600 random variable with the following PDF:

$$f_{Z|H_1}(z|H_1) = \frac{(400\lambda)^{1600}z^{1599}e^{-400\lambda z}}{1599!},$$

Now let $C_{ij}$ denote the cost of choosing hypothesis $i$ when the real hypothesis is $j$, for $i, j \in \{0, 1\}$. We are given that $C_{00} = C_{11} = -500$, $C_{10} = 200$, and $C_{01} = 100$. In
this case, given an observation \( z \), where \( z > 0 \) since exponential random variables are nonnegative, we have the following:

\[
E[\text{cost of choosing } i|\text{observation } z] = \frac{C_{i0}P(H_0)f_{Z|H_0}(z|H_0)}{f_Z(z)} + \frac{C_{i1}P(H_1)f_{Z|H_1}(z|H_1)}{f_Z(z)}
\]

We choose \( H_0 \) only if \( E[\text{cost of choosing } 0|\text{observation } z] \) is less than \( E[\text{cost of choosing } 1|\text{observation } z] \). By substituting for costs above and \( P(H_0) = P(H_1) = \frac{1}{2} \) as given, then the condition for choosing \( H_0 \) becomes:

\[
\begin{align*}
-500 & \frac{1}{2} \frac{(64\lambda)^{256}z^{255}e^{-64\lambda z}}{255!} + 100 \frac{1}{2} \frac{(400\lambda)^{1600}z^{1599}e^{-400\lambda z}}{1599!} \\
200 & \frac{1}{2} \frac{(64\lambda)^{256}z^{255}e^{-64\lambda z}}{255!} - 500 \frac{1}{2} \frac{(400\lambda)^{1600}z^{1599}e^{-400\lambda z}}{1599!} \\
\Rightarrow & 600 \frac{(400\lambda)^{1600}z^{1599}e^{-400\lambda z}}{1599!} \leq 700 \frac{(64\lambda)^{256}z^{255}e^{-64\lambda z}}{255!}
\end{align*}
\]

By rearranging and combining terms, we get the following condition for choosing hypothesis \( H_0 \):

\[
z^{1344}e^{-336\lambda z} \leq \left( \frac{7 \cdot 1599! \cdot (64\lambda)^{256}}{6 \cdot 255! \cdot (400\lambda)^{1600}} \right) \approx 0,
\]

The above results tell us that we should pick \( H_0 \) only if \( z^{1344}e^{-336\lambda z} \). Note that \( z^{1344}e^{-336\lambda z} \) equals 0 when \( z = 0 \). Thus, the rejection region for \( H_0 \) is all \( z > 0 \).

**Problem 2** (14 points)

(1) You observe a sequence of \( n \) independent Bernoulli trials each with success probability \( p \) but you don’t know \( p \). Compute the maximum likelihood estimate of \( p \).

**Solution** Define the Bernoulli trials as \( \{X_i, 1, 2, \cdots, n\} \), and we observe a sequence of it to be \( x_1, x_2, \cdots, x_n \). For notation, let \( m = \sum_{i=1}^{n} x_i \). Also, let \( H_p \) denote the event: it is true that the Bernoulli trials have success probability \( p \). Then the ML estimate of \( p \), denoted as \( \hat{p} \), is the value of \( p \) which achieves the following maximum:

\[
\max_{p \in [0,1]} P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n|H_p) = \max_{p \in [0,1]} \{p^m(1-p)^{n-m}\}.
\]

Now to maximize the above, we differentiate the above term with respect to \( p \) and set the derivative equal to 0, to obtain:

---
Eighth Problem Assignment

\[
\frac{d}{dp} \left( p^m (1-p)^{n-m} \right) = mp^{m-1}(1-p)^{n-m} - p^m(n-m)(1-p)^{n-m-1} = 0
\]
\[\implies p^{m-1}(1-p)^{n-m-1} [m(1-p) - p(n-m)] = 0
\]
\[\implies m(1-p) - p(n-m) = 0 \implies p = \frac{m}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i
\]

Note that to be complete, one also has to take the second derivative of \( p^m (1-p)^{n-m} \) to indeed show that the maximum value is achieved at \( p = \frac{m}{n} \).

Thus, the ML estimate \( \hat{p} \) is given by:

\[
\hat{p} = \frac{1}{n} \sum_{i=1}^{n} x_i
\]

(2) You observe \( n \) independent and identically distributed exponential random variables all with mean \( \mu_1 \) or \( \mu_2 \), but you don’t know which. Find the maximum likelihood estimate based on your observations.

**Solution** This is similar to part (1) but now the parameter to estimate can take only discrete values, while the observations are continuous. Define the i.i.d exponential random variable as \( Y_i \) for \( i = 1, 2, \cdots, n \) with pdf:

\[
f_Y(y) = \mu e^{-\mu y}, \quad y \geq 0;
\]

we observe a sequence of it to be \( y_1, y_2, \cdots, y_n \) in order to detect whether \( \mu = 1/\mu_1 \) or \( \mu = 1/\mu_2 \). Then this problem is nothing but a hypothesis testing problem with \( \mu \) equally likely to be \( 1/\mu_1 \) or \( 1/\mu_2 \) (where \( \mu_1 \neq \mu_2 \), and the corresponding ML detection rule is:

\[
f_{Y_1,\cdots,Y_n|\mu} \left( y_1, y_2, \cdots, y_n \left| \mu = \frac{1}{\mu_1} \right. \right) \geq f_{Y_1,\cdots,Y_n|\mu} \left( y_1, y_2, \cdots, y_n \left| \mu = \frac{1}{\mu_2} \right. \right)
\]

where the decision is \( \hat{\mu} = 1/\mu_1 \) if the lefthand side is greater, \( \hat{\mu} = 1/\mu_2 \) if the righthand side is greater. This decision rule can be further simplified to be the following:

\[
\frac{1}{n} \sum_{i=1}^{n} y_i \geq \frac{\mu_1\mu_2}{\mu_1 - \mu_2} \ln \frac{\mu_1}{\mu_2}
\]

where again the decision is \( \hat{\mu} = 1/\mu_1 \) if the lefthand side is greater, \( \hat{\mu} = 1/\mu_2 \) if the righthand side is greater.
**Problem 3** (20 points) An airline sells *refundable* tickets for $d$ each. Suppose each airplane holds $n$ seats and each passenger who purchases a ticket shows up for the flight with probability $p$, independently of any other passenger. If a plane is overbooked a passenger is refunded the ticket price and awarded $r$ for the inconvenience. How many tickets per flight should the airline sell to maximize its expected revenue if $n = 150$, $p = 0.9$, $d = 300$, $r = 100$? What if the tickets were *nonrefundable*?

**Solution** Suppose the airline sells $m$ tickets. It is clear that to maximize its profit, the airline should be selling at least $n$ tickets. We will consider $m > n$. Suppose $k$ passengers turn up on the day of the flight. The revenue of the airline is

$$R_m = \sum_{k=0}^{n} k \times d \cdot \binom{m}{k} p^k (1-p)^{m-k} + \sum_{k=n+1}^{m} \left( n \times d - (k-n) \times r \right) \binom{m}{k} p^k (1-p)^{m-k}$$

The revenue is maximum for the case $m = 169$ (see MATLAB code at end).

If the tickets are *non-refundable* then revenue when $k$ passengers turn up is

$$R_m = \sum_{k=0}^{n} k \times d \cdot \binom{m}{k} p^k (1-p)^{m-k} + \sum_{k=n+1}^{m} \left( m \times d - (k-n) \times (r+d) \right) \binom{m}{k} p^k (1-p)^{m-k}$$

The revenue is maximum for the case $m = 177$ (see MATLAB code at end).
n = 150;
p = 0.9;
d = 300 ;
r = 100 ;
m_total = n:(n+40) ;
warning off;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% Refundable Tickets %%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

cost = zeros(length(m_total),1) ;
for m = m_total
    for k = 0:n
        cost(m-n+1) = cost(m-n+1) + k*d*nchoosek(m,k)*pˆk*(1-p)ˆ(m-k);
    end;
    for k = (n+1):m
        cost(m-n+1) = cost(m-n+1) + (m*d-(k-n)*r)*nchoosek(m,k)*pˆk*(1-p)ˆ(m-k);
    end;
end;
[value,count] = max(cost) ;
disp(sprintf('The maximum occurs at %d',count+n-1)) ;
figure(1);
plot(m_total,cost) ;
ref_cost = cost;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% Nonrefundable Tickets %%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Problem 4 (28 points) We want to transmit one bit of information through a wireless link, by sending a signal which is either \( X = a \) or \( X = -a \) with equal probability. The receiver has two antennas, and at antenna \( i \), we observe:

\[
Y_i = X + Z_i, \quad i = 1, 2,
\]

where \( Z_1, Z_2 \) are independent Gaussian noise with zero mean and variance \( \sigma^2 \).

(1) Derive the MAP rule for detecting the transmitted information based on the observations at the two antennas. What is the error probability?

Solution Define random variable \( \hat{X} \) as the decoded signal; denote \( \sigma_1^2 \) and \( \sigma_2^2 \) as the variance of \( Y_1 \) and \( Y_2 \) respectively. Note that because \( X \) is equally likely to be \(+a\) or \(-a\), then the MAP rule here is to compare two posterior probabilities:

\[
P(X = a | Y_1 = y_1, Y_2 = y_2) \overset{\geq}{\propto} P(X = -a | Y_1 = y_1, Y_2 = y_2) \quad (1)
\]

\[
\implies f_{Y_1,Y_2|X}(y_1,y_2|a) \overset{\geq}{\propto} f_{Y_1,Y_2|X}(y_1,y_2|-a),
\]

where we choose \( \hat{X} = a \) if the lefthandside of the above equations is larger, or we choose \( \hat{X} = -a \) if the righthandside is larger.

But given \( X = a \), then \( Y_1 = a + Z_1 \) follows \( N(a, \sigma_1^2) \) and \( Y_2 = a + Z_2 \) follows \( N(a, \sigma_2^2) \), where \( Y_1 \) and \( Y_2 \) are independent since \( Z_1 \) and \( Z_2 \) are independent. Also, we let \( \sigma_1^2 = \sigma_2^2 = \sigma^2 \). Thus we have:
\[ f_{Y_1, Y_2|X}(y_1, y_2|a) = f_{Y_1|X}(y_1|a)f_{Y_2|X}(y_2|a) = \frac{1}{2\pi \sigma_1 \sigma_2} e^{-\frac{(y_1-a)^2}{2\sigma_1^2}} e^{-\frac{(y_2-a)^2}{2\sigma_2^2}} \]  

(2)

similarly, given \( X = -a \) we have:

\[ f_{Y_1, Y_2|X}(y_1, y_2|-a) = f_{Y_1|X}(y_1|-a)f_{Y_2|X}(y_2|-a) = \frac{1}{2\pi \sigma_1 \sigma_2} e^{-\frac{(y_1+a)^2}{2\sigma_1^2}} e^{-\frac{(y_2+a)^2}{2\sigma_2^2}} \]  

(3)

Plugging (2) into (3) and (1), we get the simplified MAP rule to be:

\[
\frac{2a}{\sigma_1^2} y_1 + \frac{2a}{\sigma_2^2} y_2 \geq 0 \\
\implies y_1 + y_2 \geq 0 \quad \text{(because } \sigma_1 = \sigma_2 = \sigma) ,
\]

where we choose \( \hat{X} = a \) if the lefthandside of the above equations is larger, or we choose \( \hat{X} = -a \) if the righthandside is larger.

Now to find the error probability, first define vector \( Y = [Y_1, Y_2] \) and a likelihood ratio (a rv) as:

\[
\text{LLR}(Y) = \frac{2a}{\sigma_1^2} y_1 + \frac{2a}{\sigma_2^2} y_2 ,
\]

then we see that if \( X = a \), then \( \text{LLR}(Y) \) is a normal random variable described by \( N \left( \frac{2a^2}{\sigma_1^2} + \frac{4a^2}{\sigma_1^2}, \frac{4a^2}{\sigma_1^2} + \frac{4a^2}{\sigma_2^2} \right) \).

Similarly, if \( X = -a \), then \( \text{LLR}(Y) \) is a normal random variable described by \( N \left( \frac{-2a^2}{\sigma_1^2} - \frac{4a^2}{\sigma_1^2}, \frac{4a^2}{\sigma_1^2} + \frac{4a^2}{\sigma_2^2} \right) \).

Thus, given \( X = a \), the probability of error is:

\[
\Pr(\text{error}|X = a) = P(\text{LLR}(Y) < 0|X = a) = 1 - \Phi \left( \sqrt{\frac{a^2}{\sigma_1^2} + \frac{a^2}{\sigma_2^2}} \right) .
\]

Similarly, we have:

\[
\Pr(\text{error}|X = -a) = P(\text{LLR}(Y) > 0|X = a) = 1 - \Phi \left( \sqrt{\frac{a^2}{\sigma_1^2} + \frac{a^2}{\sigma_2^2}} \right) .
\]

Thus finally by the total probability law, and using \( \sigma_1 = \sigma_2 = \sigma \), we have:

\[
\Pr(\text{error}) = 1 - \Phi \left( \sqrt{\frac{2a^2}{\sigma^2}} \right) = \Phi \left( -\sqrt{\frac{2a^2}{\sigma^2}} \right)
\]
We would like to compare the performance with a receiver having only one antenna. Communications engineers measure the performance benefit in terms of the transmit power saving (on the dB scale) in the dual-antenna system to achieve the same error probability as the single-antenna system. In other words, if \( b \) is the amplitude of \( X \) in the single-antenna system you will be studying:

\[
10 \log_{10} \frac{b^2}{a^2} \text{ (dB)}.
\]

In these terms, what is the performance gain from having two antennas?

**Solution**

We will show that by using one antenna, the probability of error is:

\[
\Pr(\text{error}) = \Phi \left( -\sqrt{\frac{b^2}{\sigma^2}} \right),
\]

where \( b \) is the amplitude of \( X \).

To describe the single-antenna system, suppose at the one antenna we observe: \( Y = X + Z \), where \( Z \) is Gaussian noise with zero mean and variance \( \sigma^2 \). Also, \( X \) is equally likely to be \( b \) or \( -b \). Let \( \hat{X} \) denote the decoded signal. Then the MAP rule here is to compare the two posterior probabilities:

\[
P(\text{X} = b | Y = y) \overset{>}{\underset{<}{\Rightarrow}} P(\text{X} = -b | Y = y)
\]

\[
\Rightarrow f_{Y|X}(y | b) \overset{>}{\underset{<}{\Rightarrow}} f_{Y|X}(y | -b),
\]

where we choose \( \hat{X} = b \) if the lefthandside of the above equations is larger, or we choose \( \hat{X} = -b \) if the righthandside is larger.

We know given \( X = b \), then \( Y = b + Z \) follows \( N(b, \sigma^2) \). Therefore:

\[
f_{Y|X}(y | b) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(y-b)^2}{2\sigma^2}}
\]

Similarly,

\[
f_{Y|X}(y | -b) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(y+b)^2}{2\sigma^2}}
\]

Plugging these values into the MAP rule above, we get the following:

\[
-(y - b)^2 \overset{>}{\underset{<}{\Rightarrow}} -(y + b)^2 \Rightarrow y \overset{>}{\underset{<}{\Rightarrow}} 0,
\]
where we choose $\hat{X} = b$ if the lefthandside of the above equations is larger, or we choose $\hat{X} = -b$ if the righthandside is larger. Now we can calculate the probability of error as follows:

$$
\Pr(\text{error}|X = b) = P(Y < 0|X = b) = P\left(\frac{Y - b}{\sigma} < \frac{-b}{\sigma}|X = b\right) = \Phi\left(\frac{-b}{\sigma}\right)
$$

$$
\Pr(\text{error}|X = -b) = P(Y > 0|X = -b) = P\left(\frac{Y + b}{\sigma} > \frac{b}{\sigma}|X = -b\right) = \Phi\left(\frac{-b}{\sigma}\right)
$$

Thus we see that $\Pr(\text{error}) = \Phi\left(\frac{-b}{\sigma}\right) = \Phi\left(-\sqrt{\frac{b^2}{\sigma^2}}\right).

Now, to do a fair comparison between one antenna system and two antennas system, we let the probability of errors be the same, i.e. $\Pr(\text{error}) = \Pr(\text{error})_1$, and compare the corresponding $a^2$ and $b^2$:

$$
\Pr(\text{error}) = \Pr(\text{error})_1 \implies b^2 = 2a^2.
$$

Thus the power saving of having two antennas is:

$$
10 \log_{10} \frac{b^2}{2a^2} = 10 \log(2) \approx 3 \text{ (dB)}
$$

(3) Generalize your answers in the previous parts to a system with $n$ receive antennas, with independent and identically distributed $\mathcal{N}(0, \sigma^2)$ noise at each antenna.

**Solution** As we do not apply any technique specific to the number of observations in analysis in part a), the results there can be directly extended to a system with $n$ receive antennas. The MAP rule is:

$$
\sum_{i=1}^{n} \frac{2a}{\sigma_i^2} y_i \geq 0
$$

$$
\implies \sum_{i=1}^{n} y_i \geq 0 \quad (\text{because } \sigma_1 = \sigma_2 = \cdots = \sigma_n = \sigma)
$$

where we choose $\hat{X} = a$ if the lefthandside is greater; we choose $\hat{X} = -a$ if the righthandside is greater. The corresponding probability of error is:

$$
\Pr(\text{error}) = 1 - \Phi\left(\sqrt{\frac{\sum_{i=1}^{n} a^2}{\sigma^2}}\right) = 1 - \Phi\left(\sqrt{\frac{n a^2}{\sigma^2}}\right)
$$
since $\sigma_1 = \sigma_2 = \cdots = \sigma_n = \sigma$. Similarly, having $n$ receive antennas achieves diversity gain of $n$.

(4) Suppose now that some antennas have better reception than other antennas, so that the noise variance at each antenna are different. Derive the MAP detector and its error probability in terms of the noise variances. Explain why your MAP detector makes intuitive sense.

**Solution** In all our previous analysis, we use the fact that $\sigma_1 = \sigma_2 = \cdots = \sigma_n = \sigma$ only in the last simplification step, thus all the analysis and the results up to the last step is just for the general case where noise variances are different, e.g. the MAP detector now is just:

$$\sum_{i=1}^{n} \frac{2a}{\sigma_i^2} y_i \leq 0,$$

where we choose $\hat{X} = a$ if the lefthandside is greater; we choose $\hat{X} = -a$ if the righthandside is greater.

The corresponding probability of error is just:

$$Pr(\text{error}) = 1 - \Phi \left( \sqrt{\sum_{i=1}^{n} \frac{a^2}{\sigma_i^2}} \right) = \Phi \left( -\sqrt{\sum_{i=1}^{n} \frac{a^2}{\sigma_i^2}} \right).$$

The intuition behind the rule is that the noisier components of observations play a less significant role in the estimation of $X$.

**Problem 5** (21 points) Consider again the wireless communication system in the previous question. Instead of sending one bit of information, we would like to send a continuous random variable $X$ (think of this as an analog signal like voice or video). We model $X$ as normally distributed with mean 0 and variance $a^2$.

**Solution** For the solutions to all parts of this problem, we let the two observations be denoted by:

$$Y_i = X + Z_i \quad i = 1, 2,$$

where $Z_1$, $Z_2$ are independent Gaussian noise with mean zero and variance $\sigma_1^2, \sigma_2^2$, and they are independent with $X$.

(1) Find the minimum mean square estimator of $X$ given the observations at the two antennas.
**Solution**  The Minimum Mean Square Error (MMSE) estimation of $X$, denoted by $\hat{X}$, is nothing but the conditional expectation of $X$ given $Y_1, Y_2$:

$$\hat{X} = \mathbb{E}[X|Y_1 = y_1, Y_2 = y_2].$$

For convenience, define $\text{SNR}_1 = a^2/\sigma_1^2$, and $\text{SNR}_2 = a^2/\sigma_2^2$. When $\sigma_1 = \sigma_2$, then we define $\text{SNR} = \text{SNR}_1 = \text{SNR}_2$.

To compute the expectation above, we first compute the conditional density of $X$ as follows.

$$f_{X|Y_1, Y_2}(x|y_1, y_2) = \frac{f_{Y_1, Y_2}(y_1, y_2|x)f_X(x)}{f_{Y_1, Y_2}(y_1, y_2)} = \frac{f_{Y_1}(y_1|x)f_{Y_2}(y_2|x)f_X(x)}{C_1},$$

where we have set $C_1 = f_{Y_1, Y_2}(y_1, y_2)$ because this is a constant with respect to $x$, and also we have used the fact that $Y_1$ and $Y_2$ are conditionally independent given $X$. Now, we know from earlier steps that given $X = x$, then $Y_1, Y_2$ are independent normal random variables with mean $x$ and variances $\sigma_1^2, \sigma_2^2$, respectively. Thus we have by substituting into the above equation:

$$f_{X|Y_1, Y_2}(x|y_1, y_2) = \frac{1}{C_1} \frac{1}{(2\pi)^{3/2}a\sigma_1\sigma_2} e^{-\frac{(y_1-x)^2}{2\sigma_1^2} - \frac{(y_2-x)^2}{2\sigma_2^2} - \frac{x^2}{2a^2}} \cdot \frac{C_2}{C_1} e^{-(x - \frac{\text{SNR}_1}{1+\text{SNR}_1+\text{SNR}_2} y_1 - \frac{\text{SNR}_2}{1+\text{SNR}_1+\text{SNR}_2} y_2)^2/(1+2\text{SNR}^2)} ,$$

where $C_2$ is a constant that denotes all terms which do not include an $x$ (we have removed these for convenience). (the second step was shown by completing the square in the $e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ term). Comparing this PDF to the PDF for a Gaussian random variable, we see that $X$ given $Y_1, Y_2$ is also Gaussian. There is no need to compute the value of $C_2/C_1$ since a PDF must integrate to 1.

In other words, given $Y_1 = y_1$ and $Y_2 = y_2$, then $X$ is nothing but a Gaussian random variable with mean $\frac{\text{SNR}_1}{1+\text{SNR}_1+\text{SNR}_2} y_1 - \frac{\text{SNR}_2}{1+\text{SNR}_1+\text{SNR}_2} y_2$ and variance $\frac{a^2}{1+2\text{SNR}^2}$. Thus, we get the MMSE estimation of $X$ to be:

$$\hat{X} = \mathbb{E}[X|Y_1 = y_1, Y_2 = y_2] = \frac{\text{SNR}_1}{1+\text{SNR}_1+\text{SNR}_2} y_1 - \frac{\text{SNR}_2}{1+\text{SNR}_1+\text{SNR}_2} y_2,$$

In the case where $\sigma_1 = \sigma_2$, then we get $\text{SNR}_1 = \text{SNR}_2 = \text{SNR}$ and

$$\hat{X} = \frac{\text{SNR}}{1+2\text{SNR}} (y_1 + y_2)$$
We can see again that for Gaussian random variables, the MMSE estimate is nothing but a linear combination of the observations. The noisier observations play a less significant role in the estimation of $X$.

(2) What is the associated mean-square error? How much power do you save by using the extra antenna?

**Solution** The mean-square error is calculated as follows, by using $\hat{X} = E[X|Y_1, Y_2]$ and the conditional variance from part (1) to obtain:

$$
E \left[ (X - \hat{X})^2 \right] = E \left[ E \left[ (X - \hat{X})^2 | Y_1, Y_2 \right] \right] \\
= E \left[ E \left[ (X - E[X|Y_1, Y_2])^2 | Y_1, Y_2 \right] \right] \\
= E[\text{var}(X|Y_1, Y_2)] = E \left[ \frac{a^2}{1 + \text{SNR}_1 + \text{SNR}_2} \right] \\
= \frac{a^2}{1 + \text{SNR}_1 + \text{SNR}_2}
$$

Thus, when $\text{SNR}_1 = \text{SNR}_2 = \text{SNR}$, then the mean-square error is $\frac{a^2}{1 + 2\text{SNR}}$.

Consider the case of using one antenna. It can be shown, using similar steps, that the mean-squared error of using one antenna is $\frac{a^2}{1 + \text{SNR}}$. Thus we get the SNR gain to be:

$$
\frac{1}{1 + 2\text{SNR}_2} = \frac{1}{1 + \text{SNR}_1} \quad \Rightarrow \quad \frac{\text{SNR}_1}{\text{SNR}_2} = 2 \approx 3 \text{ (dB)}
$$

(3) Redo the first part if the noise variances are different at the antennas, say $\sigma_1^2$ and $\sigma_2^2$.

Does your result make intuitive sense?

**Solution** This solution was completed in the previous two parts for the case when $\text{SNR}_1 \neq \text{SNR}_2$. In particular, we showed that the minimum mean square estimator of $X$ given the observations at the antennas is:

$$
\hat{X} = E[X|Y_1 = y_1, Y_2 = y_2] = \frac{\text{SNR}_1}{1 + \text{SNR}_1 + \text{SNR}_2}y_1 - \frac{\text{SNR}_2}{1 + \text{SNR}_1 + \text{SNR}_2}y_2
$$

The corresponding mean-square error is:

$$
E \left[ (X - \hat{X})^2 \right] = \frac{a^2}{1 + \text{SNR}_1 + \text{SNR}_2}
$$

**Extra Credit (10 points)** There are 240 students in a literature class. Our model states that $X$, the numerical grade for any individual student, is an independent Gaussian random
variable with a standard deviation equal to $10\sqrt{2}$. Assuming that our model is correct, we wish to perform a significance test on the hypothesis that $E[X]$ is equal to 60.

Determine the highest and lowest class averages which will result in the acceptance of this hypothesis:

(1) At the 0.02 level of significance

**Solution**

Let $X_i$ denote the grade of the $i$th student, where $1 \leq i \leq 240$. We know that $X_i$ is a normal random variable with variance $10\sqrt{2}$. Thus, the class average $Z$ is a random variable determined as follows:

$$Z = \frac{1}{240} \sum_{i=1}^{240} X_i.$$

Suppose our hypothesis is true, so that $E[X_i] = 60$ for all $i$. Then in this case, assuming the hypothesis is true, $Z$ is a normal random variable with mean and variance calculated as follows, with $H$ denoting the event that our hypothesis is true ($E[X_i] = 60$).

$$E[Z|H] = E\left[\frac{1}{240} \sum_{i=1}^{240} X_i \bigg| H\right] = \frac{1}{240} \sum_{i=1}^{240} E[X_i|H] = 60.$$

$$\text{var}(Z|H) = \text{var}\left(\frac{1}{240} \sum_{i=1}^{240} X_i \bigg| H\right) = \frac{1}{(240)^2} \sum_{i=1}^{240} \text{var}(X_i|H) = \frac{10\sqrt{2}}{240} = \frac{\sqrt{2}}{24}.$$

From the problem statement, we would like to find the highest and lowest class averages which will result in the acceptance of this hypothesis. Now we know that the distribution of a normal random variable is symmetric about its mean. Thus, the highest and lowest class averages which result in acceptance of the hypothesis must be some numbers $60 + \delta$ and $60 - \delta$, respectively, for some $\delta > 0$. 

Due on March 16, 2007
Now we determine $\delta$. By definition, $\delta$ needs to be the smallest number satisfying:

$$P( |Z - 60| > \delta | H) = 0.02$$

$$\implies P( Z - 60 > \delta | H) + P( Z - 60 < -\delta | H) = 0.02$$

$$\implies 2P( Z - 60 > \delta | H) = 0.02 \implies P( Z < 60 - \delta | H) = 0.01$$

where we have used the fact that the distribution of a normal random variable is symmetric about its mean, so $P( Z - 60 > \delta | H) = P( Z - 60 < -\delta | H)$. Thus we have the following:

$$P( Z < 60 - \delta | H) = 0.01 \implies P \left( \frac{(Z - 60)24}{\sqrt{2}} < \frac{(60 - \delta - 60)(24)}{\sqrt{2}} \bigg| H \right) = 0.01$$

$$\implies P \left( \frac{(Z - 60)24}{\sqrt{2}} < -\frac{\delta(24)}{\sqrt{2}} \bigg| H \right) = 0.01$$

Now we obtain the value for $\frac{-\delta(24)}{\sqrt{2}}$. Since $\frac{(Z - 60)(24)}{\sqrt{2}}$ is a standard normal random variable (mean 0 and variance 1), then we can use the chart on page 155 of the textbook to find that:

$$P \left( \frac{(Z - 60)(24)}{\sqrt{2}} \leq 2.39 \bigg| H \right) \approx 0.99 \implies P \left( \frac{(Z - 60)(24)}{\sqrt{2}} > 2.39 \bigg| H \right) \approx 0.01$$

$$\implies P \left( \frac{(Z - 60)(24)}{\sqrt{2}} < -2.39 \bigg| H \right) \approx 0.01 ,$$

where again we have used the property that the PDF of a normal random variable is symmetric about its mean. Thus, we have shown that $\frac{-\delta(24)}{\sqrt{2}} = -2.39$. Solving for $\delta$ gives us:

$$\delta = \frac{2.39\sqrt{2}}{24} \approx 0.1408$$

Thus the highest and lowest class averages for acceptance at the 0.02 level of significance are given by $60 + \delta = 60.1408$ and $60 - \delta = 59.8592$, respectively.
(2) At the 0.50 level of significance

**Solution** Proceeding similarly to the previous part of the problem, we can apply the same steps to show that we need to find $\delta$ satisfying the following:

$$P \left( \frac{Z - 60}{24} \frac{24}{\sqrt{2}} < -\frac{\delta(24)}{\sqrt{2}} \middle| H \right) = 0.25 ,$$

because if the hypothesis is true, then the distribution of $Z$ is not dependent on the level of significance that we assign to our test. Thus, we need to find the value of $-\frac{\delta(24)}{\sqrt{2}}$ above. We know the following from page 155 and because $\frac{(Z-60)(24)}{\sqrt{2}}$ is a standard normal random variable (conditioned on $H$ being true):

$$P \left( \frac{(Z - 60)(24)}{\sqrt{2}} \leq 0.675 \middle| H \right) \approx 0.75 \implies P \left( \frac{(Z - 60)(24)}{\sqrt{2}} > 0.675 \middle| H \right) \approx 0.25$$

$$\implies P \left( \frac{(Z - 60)(24)}{\sqrt{2}} < -0.675 \middle| H \right) \approx 0.25 ,$$

Thus we have shown that $-\frac{\delta(24)}{\sqrt{2}} = -0.675$. Solving for $\delta$ gives us:

$$\delta = \frac{(0.675)\sqrt{2}}{24} \approx 0.0398 \approx 0.04$$

Thus the highest and lowest class averages for acceptance at the 0.50 level of significance are given by $60 + \delta = 60.04$ and $60 - \delta = 59.96$, respectively.