Problem 1
The rvs $Y$ and $X$ are related as

$$Y = 10 + 20X + \epsilon, \quad \epsilon \sim N(0, 4^2).$$

a) Find the conditional expectation of $Y$ given $X = 3$.

$$(Y|X = x) \sim N(10 + 20x, 4^2)$$

b) Find the conditional distribution of $Y$ given $X = x$.

$$E(Y|X = 3) = 70$$

c) Find the conditional variance of $Y$ given $X = 0$.

$$Var(Y|X = 0) = 4^2$$

Problem 2
Let $(X, Y)$ be a bivariate Normal random variable such that $Y \sim N(4, 3^2)$, $X \sim N(5, 2^2)$, and $\rho(X, Y) = 0.8$.

a) Find the conditional expectations $E(Y|Y = y)$, $E(X|X = x)$, $E(Y|X = x)$ and $E(X|Y = y)$.

$$E(Y|Y = y) = y$$

$$E(X|X = x) = x$$

$$E(Y|X = x) = E(Y) + \rho(X, Y) \frac{\sigma_Y}{\sigma_X} (x - E(x)) = 4 + 0.8 \times \frac{3}{2} \times (x - 5) = 4 + 1.2(x - 5)$$

$$E(X|Y = y) = E(X) + \rho(X, Y) \frac{\sigma_X}{\sigma_Y} (y - E(Y)) = 5 + 0.8 \times \frac{2}{3} \times (y - 4) = 5 + \frac{1.6}{3} (y - 4)$$

b) Find the best mean-square constant forecast of $Y$, $\hat{Y} = c$.

Minimize $E[(Y - c)^2] = E[Y^2 - 2cY + c^2] = E[Y^2] - 2cE[Y] + c^2$. Taking the derivative with respect to $c$ and setting this equal to 0 we have: $\frac{d}{dc} = -2E[Y] + 2c = 0 \rightarrow c = E(Y) = 4$.

c) Find the best mean-square forecast of $Y$ by a function of $X$, $\hat{Y} = f(X)$.

Using the theorem that was presented in lecture, the best forecast of a r.v. $Y$ using r.v. $X$ is the conditional expectation of $Y$ given $X : E(Y|X)$.
To derive the conditional expectation of $Y$ given $X = x$ we first find the conditional distribution of $Y$ given $X = x$. Recall that if $X$ and $Y$ have a joint probability density function $f(x, y)$, then the conditional probability density function of $Y$ given $X = x$ is defined for all values of $x$ such that

$$f_x(x) > 0 \text{ by } f_{Y|X}(y|x) = \frac{f(x, y)}{f_x(x)}.$$  

We are given that $X$ and $Y$ have a bivariate Normal distribution and that $X$ and $Y$ are individually normally distributed. For ease of notation, we will let $E[Y] = \mu_Y$, $E[X] = \mu_X$, $\text{Cov}(XY) = \sigma_{XY}^2$ and $\rho(XY) = \rho$.

The joint probability density function of $X$ and $Y$ is given as:

$$f(x, y) = \frac{1}{2\pi \sqrt{\sigma_X^2 \sigma_Y^2 (1 - \rho^2)}} \exp\left\{-\frac{1}{2(1 - \rho^2)} \left[\frac{(x - \mu_X)^2}{\sigma_X^2} + \frac{(y - \mu_Y)^2}{\sigma_Y^2} - 2\rho \frac{(x - \mu_X)(y - \mu_Y)}{\sigma_X \sigma_Y}\right]\right\}$$

and $X$ has a probability density function given as: $f(x) = \frac{1}{\sqrt{2\pi \sigma_X^2}} \exp\left\{-\frac{1}{2} \left(\frac{x - \mu_X}{\sigma_X}\right)^2\right\}$.

Notice that the terms $(\frac{y - \mu_Y}{\sigma_Y})^2 - 2\rho (\frac{x - \mu_X}{\sigma_X})(\frac{y - \mu_Y}{\sigma_Y})$ in the exponent can be written as:

$$\frac{1}{\sigma_Y^2} [y - \mu_Y - \frac{\sigma_Y}{\sigma_X} (x - \mu_X)]^2 - \frac{\rho^2}{\sigma_X^2} (x - \mu_X)^2.$$  

Recognizing that $\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$ or that $\frac{\sigma_Y}{\sigma_X} = \frac{\sigma_{XY}}{\sigma_X^2}$, we can write the complete exponent as:

$$- \frac{1}{2(1 - \rho^2)} \left[\frac{(y - \mu_Y)^2}{\sigma_Y^2} - 2\rho \frac{(x - \mu_X)(y - \mu_Y)}{\sigma_X \sigma_Y} + \left(\frac{x - \mu_X}{\sigma_X}\right)^2\right] =$$

$$- \frac{1}{2\sigma_Y^2(1 - \rho^2)} (y - \mu_Y - \frac{\sigma_Y}{\sigma_X} (x - \mu_X))^2 - \frac{1}{2(1 - \rho^2)} \left(\frac{\sigma_Y^2}{\sigma_X^2} - \frac{\rho^2}{\sigma_X^2}\right)(x - \mu_X)^2 =$$

$$- \frac{1}{2\sigma_Y^2(1 - \rho^2)} (y - \mu_Y - \frac{\sigma_Y}{\sigma_X} (x - \mu_X))^2 - \frac{1}{2} \left(\frac{1 - \mu_Y}{\sigma_Y}\right)^2.$$  

The constant term $2\pi \sqrt{\sigma_X^2 \sigma_Y^2 (1 - \rho^2)} = \sqrt{2\pi} \sqrt{\sigma_X^2} \sqrt{2\pi} \sqrt{\sigma_Y^2} \sqrt{(1 - \rho^2)}$. We can now divide the density $f(x, y)$ by the marginal density $f(x)$ to obtain the conditional density:

$$f(y|x) = \frac{1}{\sqrt{2\pi \sigma_Y^2 (1 - \rho^2)}} \exp\left\{-\frac{1}{2\sigma_Y^2(1 - \rho^2)} [y - \mu_Y - \left(\frac{\sigma_{XY}}{\sigma_X^2}\right)(x - \mu_X)]^2\right\}. $$

Notice that the conditional distribution of $Y$ given $X = x$ is Normal with mean of $\mu_Y + \left(\frac{\sigma_{XY}}{\sigma_X^2}\right)(x - \mu_X)$ and variance $\sigma_Y^2(1 - \rho^2)$. Using the fact that $\frac{\sigma_Y}{\sigma_X} = \frac{\sigma_{XY}}{\sigma_X^2}$ we conclude that $E[Y|X = x] = \mu_Y + \frac{\sigma_{XY}}{\sigma_X} (x - \mu_X)$.

\(\text{d) Find the best mean-square forecast of } X \text{ by a function of } Y, \hat{X} = f(Y).\)

By the same argument as in part (c) the best forecast of a r.v. $X$ using r.v. $Y$ is the conditional expectation of $X$ given $Y : E(X|Y)$. Switching $X$ and $Y$ in the computations of part (c) we have:

$$E[X|Y = y] = \mu_X + \frac{\rho \sigma_X}{\sigma_Y} (y - \mu_Y).$$
Problem 3
Consider rvs $Y$ and $X$ with finite variances. We notice that when $\rho(X, Y) = \pm 1$, the rvs $Y$ and $X$ are (deterministically) linearly related, and the slopes of the lines $Y = \beta_0 + \beta_1 X$ and $X = \alpha_0 + \alpha_1 Y$ are reciprocal to each other:

$$\alpha_1 \beta_1 = 1.$$  

**True or False:** If $\rho(X, Y) \neq \pm 1$, the slopes of the best mean-square forecast line $\hat{Y} = \beta_0 + \beta_1 X$ and $\hat{X} = \alpha_0 + \alpha_1 Y$ are reciprocal to each other? False.

First consider the slope of the best mean-square forecast line $\hat{Y} = \beta_0 + \beta_1 X$ which is given as: $\beta_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$. Given $\rho(X, Y) = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2} \sqrt{\sum(Y_i - \bar{Y})^2}}$ we can write

$$\beta_1 = \rho(X, Y) \frac{\sqrt{\sum(Y_i - \bar{Y})^2}}{\sqrt{\sum(X_i - \bar{X})^2}}.$$  

Now consider the slope of the best mean-square forecast line $\hat{X} = \alpha_0 + \alpha_1 Y$ which is given as: $\alpha_1 = \frac{\sum(Y_i - \bar{Y})(X_i - \bar{X})}{\sum(Y_i - \bar{Y})^2}$. Using the expression for $\rho(X, Y)$ we can write $\alpha_1 = \rho(X, Y) \frac{\sqrt{\sum(X_i - \bar{X})^2}}{\sqrt{\sum(Y_i - \bar{Y})^2}}$.

Therefore, we have $\alpha_1 \beta_1 = [\rho(X, Y) \frac{\sqrt{\sum(X_i - \bar{X})^2}}{\sqrt{\sum(Y_i - \bar{Y})^2}}] * [\rho(X, Y) \frac{\sqrt{\sum(Y_i - \bar{Y})^2}}{\sqrt{\sum(X_i - \bar{X})^2}}] = (\rho(X, Y))^2$.

In any other case except when $\rho(X, Y) = \pm 1$ the quantity $(\rho(X, Y))^2 \neq 1$.

Problem 4
Consider rvs $Y$ and $X$ with finite variances.

a) Formulate the necessary and sufficient conditions for the best mean-square forecast of $Y$ by a linear function of $X$ to be $\hat{Y} = X$.

Let a linear function of $X$ be given as $f(X) = a + bX$. We wish to find $a, b$ such that we minimize:

$$E(Y - f(X))^2 = E(Y - (a + bX))^2.$$  

$$E(Y - (a + bX))^2 = E[Y^2 - 2Y(a + bX) + (a + bX)^2] = E[Y^2 - 2Ya - 2bXY + a^2 + 2abX + b^2]$$

Using the linearity property of the expectations operator and recognizing that $a$ and $b$ are constants we have: $E(Y - (a + bX))^2 = E(Y^2) - 2aE(Y) - 2bE(XY) + a^2 + 2abE(X) + b^2E(X^2)$

We wish to find $a$ and $b$ that minimize $E(Y - (a + bx))^2$ so we take partial derivatives with respect to $a$ and $b$ and set them equal to 0:

$$\frac{\partial}{\partial a} = -2E(Y) + 2a + 2bE(X) = 0$$  

$$\frac{\partial}{\partial b} = -2E(XY) + 2aE(X) + 2bE(X^2) = 0$$

To solve for $a$ and $b$ notice that from $\frac{\partial}{\partial a}$ we can write: $a = E(Y) - bE(X)$ and
Recognizing that \[ E(Y) - E(X)E(Y) \rightarrow b[E(X^2) - [E(X)]^2] = E(XY) - E(X)E(Y) \]

we have: \( b = \frac{\text{Cov}(XY)}{\text{Var}(X)} \) and \( a = E(Y) - \frac{\text{Cov}(XY)}{\text{Var}(X)} \cdot E(X) \).

Given the above, we will have \( a = 0 \) and \( b = 1 \) when \( \text{Cov}(XY) = \text{Var}(X) \) and \( E(Y) = E(X) \). These are the necessary and sufficient conditions for the best mean-square forecast of \( Y \) by a linear function of \( X \) to be \( \hat{Y} = X \).

b) It is known that the best mean-square linear forecast of \( Y \) by a (linear) function of \( X \) is \( \hat{Y} = X \), and \( \text{Var}(Y) = 4 \). Find the range of possible values for the standard deviation of \( X \), \( \sigma_X \).

We know from part (a) that if the best mean-square linear forecast of \( Y \) by a (linear) function of \( X \) is \( \hat{Y} = X \) then \( \text{Cov}(XY) = \text{Var}(X) \). Additionally, we are given that \( \text{Var}(Y) = 4 \). Thus we know that \( \sigma_Y = 2 \).

Given \( \text{Cov}(XY) = \sigma_X^2 \), if we divide both sides of this expression by \( \sigma_X \cdot \sigma_Y \) we have

\[
\frac{\text{Cov}(XY)}{\sigma_X \sigma_Y} = \frac{\sigma_X^2}{\sigma_X \sigma_Y} \text{ which can be written as: } \rho(XY) = \frac{\sigma_X}{\sigma_Y} \text{ or } \sigma_Y \cdot \rho(XY) = \sigma_X. \]

Since \(-1 \leq \rho(XY) \leq 1\) and \( \sigma_X \) must be strictly positive, we have \( 0 < \sigma_X \leq 2 \).

**Problem 5**

Suppose that \( X_t \) is a stationary time series with acf \( \rho(h) \) and variance \( \sigma^2 \).

a) Show that the best mean-square forecast of \( X_{t+h} \) in the form

\[
\hat{X}_{t+h} = a \cdot X_t + b
\]

corresponds to \( a = \rho(h) \), \( b = E(X_0)(1 - \rho(h)) \).

In Problem 4 we found that for two random variables, \( Y \) and \( X \), the best mean-square forecast of \( Y \) by a linear function of \( X \) was given by \( \hat{Y} = aX + b \) where the coefficient on \( X \) was given as \( a = \frac{\text{Cov}(XY)}{\text{Var}(X)} \) and the constant term was given as \( b = E(Y) - \frac{\text{Cov}(XY)}{\text{Var}(X)} \cdot E(X) \).

Replacing \( X \) and \( X_t \) and \( Y \) with \( X_{t+h} \) we have: \( a = \frac{\text{Cov}(X_t, X_{t+h})}{\text{Var}(X_t)} = \rho(h) \) and \( b = E(X_{t+h}) - \frac{\text{Cov}(X_t, X_{t+h})}{\text{Var}(X_t)} \cdot E(X_t) = E(X_{t+h}) - \rho(h) \cdot E(X_t) \). By the stationarity of the series we have \( E(X_{t+h}) = E(X_t) = E(X_0) \). Therefore, \( b = E(X_{t+h}) - \rho(h) \cdot E(X_t) = E(X_0)(1 - \rho(h)) \).

b) What is the mean-square error of this forecast?
\[ \hat{X}_{t+h} = \rho X_t + E(X_0)(1 - \rho). \] As in part (a), by the stationarity of the process we have, \( E(X_0) = E(X_t) = E(X_{t+h}) = \mu_X, \) so we have: \( \hat{X}_{t+h} = \rho X_t + \mu_X(1 - \rho). \)

\[
\epsilon = X_{t+h} - \hat{X}_{t+h} = X_{t+h} - (\rho X_t + \mu_X(1 - \rho)) = X_{t+h} - \rho X_t - \mu_X(1 - \rho)
\]

\[
Var(\epsilon) = E[\epsilon^2] = E[X_{t+h} - \rho X_t - \mu_X(1 - \rho)]^2 = E[X_{t+h}^2 - 2\rho X_t X_{t+h} - 2\mu_X X_{t+h}(1 - \rho) + 2\mu_X(1 - \rho)] =
\]

\[
E(X_{t+h}^2) + \rho^2 E(X_t^2) + \mu^2(1 - \rho)^2 - 2\rho E(X_t X_{t+h}) - 2\mu_X(1 - \rho) + 2\mu^2(1 - \rho)
\]

Gathering all of the terms in \( \mu^2 \) we have:

\[
Var(\epsilon) = \mu^2[2\rho - \rho \sigma^2 - 1] + E(X_{t+h}^2) + \rho^2 E(X_t^2) - 2\rho E(X_t X_{t+h}) = \]

\[
-\mu^2(1 - \rho)^2 + E(X_{t+h}^2) + \rho^2 E(X_t^2) - 2\rho E(X_t X_{t+h})
\]

Notice the following:

1. \( Var(X_t) = E(X_t^2) - [E(X_t)]^2 \rightarrow Var(X_t) = E(X_t^2) - \mu^2_X. \) Letting \( \sigma^2_X \) represent \( Var(X_t) \) we have: \( E(X_t^2) = \sigma^2_X + \mu^2_X. \) Also, by stationarity of the process we have \( E(X_{t+h}^2) = E(X_t^2) = \sigma^2_X + \mu^2_X. \) For convenience we will drop the subscript and replace \( E(X_{t+h}^2) \) and \( E(X_t^2) \) with \( \sigma^2 + \mu^2. \)

2. \[
Cov(X_t X_{t+h}) = \gamma(h) = E[(X_t - \mu_X)(X_{t+h} - \mu_X)] =
\]

\[
E(X_t X_{t+h}) - \mu_X E(X_t) - \mu_X E(X_{t+h}) + \mu^2_X =
\]

\[
E(X_t X_{t+h}) - \mu^2_X \rightarrow E(X_t X_{t+h}) = \gamma(h) + \mu^2_X = \rho \sigma^2 + \mu^2
\]

Making these substitutions we have:

\[
Var(\epsilon) = -\mu^2(1 - \rho)^2 + \sigma^2 + \mu^2 + \rho^2(\sigma^2 + \mu^2) - 2\rho(\rho \sigma^2 + \mu^2)
\]

After opening up all of the parentheses and cancelling terms, we are left with: \( Var(\epsilon) = \sigma^2(1 - \rho^2). \)