1. Joint Gaussian Probability
Let $X \sim \mathcal{N}(1, 1)$ and $Y \sim \mathcal{N}(0, 1)$ be jointly Gaussian with covariance $\rho$. What is $P(X > Y)$?

**Solution:**
Note that

$$P(X > Y) = P(X - Y > 0)$$

Any linear combination of jointly gaussian random variables is itself gaussian, so to solve this problem, we try finding the mean and variance of $X - Y$.

$$\text{var}(X - Y) = \text{cov}(X, X - Y)$$
$$= \text{var}(X) - 2 \text{cov}(X, Y) + \text{var}(Y)$$
$$= 2(1 - \rho)$$

and by linearity of expectation, $E[X - Y] = 1$.

Thus,

$$P(X > Y) = P(X - Y > 0)$$
$$= P(\mathcal{N}(1, 2(1 - \rho)) > 0)$$
$$= P\left(\mathcal{N}(0, 1) > \frac{-1}{\sqrt{2(1 - \rho)}}\right)$$
$$= 1 - \Phi\left(\frac{-1}{\sqrt{2(1 - \rho)}}\right)$$
$$= \Phi\left(\frac{1}{\sqrt{2(1 - \rho)}}\right).$$

2. MMSE for Jointly Gaussian Random Variables
Provide justification for each of the following steps (1 - 5) to prove that the LLSE is equal to the MMSE estimator for jointly Gaussian random variables $X$ and $Y$.

Let $g(X) = L[Y \mid X]$.

$$E[(Y - g(X))X] = 0 \quad (1)$$
$$\Rightarrow \text{cov}(Y - g(X), X) = 0 \quad (2)$$
$$\Rightarrow Y - g(X) \text{ is independent of } X \quad (3)$$
$$\Rightarrow E[(Y - g(X))f(X)] = 0 \forall f(\cdot) \quad (4)$$
$$\Rightarrow g(X) = E[Y \mid X] \quad (5)$$

**Solution:**
1. Since \( g(X) \) is the LLSE, \( Y - g(X) \) is orthogonal to all linear functions of \( X \).

2. Since \( Y - g(X) \) has 0 mean, \( E[(Y - g(X))X] = E[(Y - g(X))X] - E[Y - g(X)]E[X] = \text{cov}(Y - g(X), X) \).

3. Since \( X \) and \( Y \) are JG, so are all linear combinations of them, i.e. \( Y - g(X) \) and \( X \) are JG. For JG random variables, uncorrelated implies independent.

4. Since \( Y - g(X) \) and \( X \) are independent, so are any function of \( Y - g(X) \) and any function of \( X \). Therefore \( E[(Y - g(X))f(X)] = E[Y - g(X)]E[f(X)] = 0 \).

5. \( E[Y \mid X] \) is the one and only function \( g(X) \) that satisfies \( E[(Y - g(X))f(X)] = 0 \) for any function \( f(X) \) of \( X \). Since \( g(X) \) satisfies this property, it must be \( E[Y \mid X] \).

### 3. Overlapping Normals

As you will see in the lab, a big part of the Kalman filter is “overlapping” two normal distributions. In particular, suppose at the current time step, you have a state \( X \sim \mathcal{N}(\mu, \sigma_1^2) \) and an observation \( Y \sim X + \mathcal{N}(0, \sigma_2^2) \). The two noises are independent.

(a) Not knowing \( Y \), what is your best guess of \( X \)?

(b) \( X \) and \( Y \) are jointly Gaussian. Write them in \( X = AZ + \mu \) format.

(c) What is \( E[X \mid Y] \)?

(d) Given \( Y = y \), the conditional distribution \( X \mid Y = y \) turns out to be normal, which is the reason why we can continue using the same process for future time steps. In fact, it turns out \( \text{var}(X \mid Y) \) is constant, so it’s always equal to its expectation \( E[\text{var}(X \mid Y)] = E[(X - E[X \mid Y])^2] \). WLOG assume \( \mu = 0 \) and show that this is equal to \( \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \).

**Solution:**

(a) It is to guess \( E[X] = \mu \).

(b) Let \( Z_1, Z_2 \) i.i.d. \( \mathcal{N}(0, 1) \). \( X = \sigma_1 Z_1 + \mu \), and \( Y = X + \sigma_2 Z_2 = \sigma_1 Z_1 + \sigma_2 Z_2 \).

\[
\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} \sigma_1 & 0 \\ \sigma_1 & \sigma_2 \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} + \begin{bmatrix} \mu \\ \mu \end{bmatrix}
\]

(c) Since \( X \) and \( Y \) are jointly Gaussian, \( E[X \mid Y] = L[X \mid Y] \).

\[
L[X \mid Y] = E[X] + \frac{\text{cov}(X, Y)}{\text{var}(Y)}(Y - E[Y])
\]

\( E[X] = E[Y] = \mu \). For the covariance / variance, we can use the covariance matrix of \( X \) and \( Y \), which is found through

\[
AA^T = \begin{bmatrix} \sigma_1 & 0 \\ \sigma_1 & \sigma_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & \sigma_1 \\ 0 & \sigma_2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 + \sigma_2^2 \end{bmatrix}
\]

So

\[
L[X \mid Y] = \mu + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} (Y - \mu)
\]

Note that though it doesn’t look exactly the same, \( \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \) is the idea behind the Kalman gain term in the Kalman filter.
(d) We want to find

\[ \mathbb{E}[(X - E[X \mid Y])^2] = \mathbb{E}[(X - \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} Y)^2] \]

We know that \( Y = X + \mathcal{N}(0, \sigma_2^2) \), but let’s call the noise part \( W \). So \( Y = X + W \), where \( X \) and \( W \) are independent.

\[ \mathbb{E}[(X - E[X \mid Y])^2] = \mathbb{E}[(X - \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} (X + W))^2] \]

\[ = \mathbb{E}[(\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} X - \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} W)^2] \]

Since \( X \) and \( W \) are orthogonal, the cross term disappears.

\[ = \mathbb{E}[(\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} X)^2] + \mathbb{E}[(\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} W)^2] \]

\[ = \left[ \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \right]^2 \mathbb{E}[X^2] + \left[ \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \right]^2 \mathbb{E}[W^2] \]

\[ = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \sigma_1^2 + \left[ \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \right]^2 \sigma_2^2 \]

\[ = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \left[ \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \right] \]

\[ = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \]

Letting \( k = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \), it turns out this is also equal to \( \sigma_1^2 (1 - k) \), which may be familiar from the Kalman filter equation.