1. Given the following distribution function:

\[ f_{XY}(x, y) = \begin{cases} 
\frac{1}{\pi} e^{-\frac{x^2+y^2}{2}} & \text{if } xy \geq 0 \\
0 & \text{otherwise}
\end{cases} \]

(a) Show that the distribution has normal marginal distributions for \( X \) and \( Y \) but is not jointly normal. (Hint: it may help to sketch it first.)

(b) Are \( X \) and \( Y \) independent?

2. For a typical QPSK digital modulation the bit error probability is given by \( p_b = Q(\sqrt{2S}) \) where \( S \) is the signal to noise ratio (SNR). The previous homework described a method of sending a block of 4 bits uncoded versus sending the 4 bits with 3 extra bits added (7 bits total) so that one error could be corrected.

(a) For the uncoded case, compute \( S_4(p_c) \), the SNR required so that the probability that the 4 bits is corrupted is \( p_c \). Compute it for \( p_c \in \{10^{-2}, 10^{-4}, 10^{-6}\} \).

(b) For the coded case, compute \( S_7(p_c) \), the SNR required so that the probability that the 7 bits is corrupted is \( p_c \). Compute it for \( p_c \in \{10^{-2}, 10^{-4}, 10^{-6}\} \). Note that each bit gets only 4/7 of the total energy per block (i.e. \( p_b = Q(\sqrt{8S/7}) \)).

(c) For each \( p_c \), compute the ratio of \( S_4(p_c)/S_7(p_c) \) in dB. Naive users complain that a code like the 7 bit code requires extra effort to send the data (in this case 7/4 more effort if \( S \) is fixed). Is this statement justified?

3. Let \( Z = X + Y \) where \( X \sim N(\mu_X, \sigma_X^2) \) and \( Y \sim N(\mu_Y, \sigma_Y^2) \). Using the techniques described in class, show \( Z \sim N(\mu_Z, \sigma_Z^2) \) and compute \( \mu_Z \) and \( \sigma_Z^2 \) in terms of the parameters for \( X \) and \( Y \).

4. Let \( Y = 10^{X/10} \) (i.e. \( X \) is \( Y \) expressed in dB). Let \( X \sim N(0, \sigma^2) \), where \( \sigma \) has units of dB.

(a) Compute the density of \( Y \), \( f_Y \).

(b) Let \( Y_1 \) and \( Y_2 \) be two independent draws from the distribution \( Y \). Let \( Z = Y_1/Y_2 \). Compute the density of \( Z \), \( f_Z \). How does \( f_Z \) compare to \( f_Y \)?
5. A wood processing machine classifies each piece of lumber into wood species with an optical sensor that measures the brightness $X$ of the wood. The two types of wood are cherry (C) and maple (M). The brightness of cherry is distributed as $N(9, 9)$. The brightness of maple is distributed as $N(13, 4)$. Twice as much maple is processed as cherry and it is fed randomly into the machine.

(a) Given the measured brightness is $x$, compute the probability that the wood is maple or cherry (i.e. compute $P[M|X = x]$ and $P[C|X = x]$).

(b) Plot $P[M|X = x]$ and $P[C|X = x]$ on the same graph vs. $x$. Show for which values of $x$ maple is more likely and for which cherry is more likely. This defines the classifier rule.

(c) The probability of error given the wood is cherry ($P[E|C]$) is the total probability of all $x$ where the classifier decides the wood is maple given it is cherry. $P[E|M]$ is similarly defined. Compute $P[E|C]$ and $P[E|M]$.

(d) Compute the total error probability, $P[E]$. Convince me that no other classification rule on $X$ would yield a lower error probability.

6. In the previous problem, an alternative formulation is as follows. A wood processing machine identifies the cherry wood from the other wood (in this case maple). We can get four outcomes:

- True Positive: Classify as cherry when it is cherry.
- False Positive: Classify as cherry when it is maple.
- True Negative: Classify as maple when it is maple.
- False Negative: Classify as maple when it is cherry.

The goal in this formulation might be to minimize false positives or to maximize some utility function on the probability of getting any one of these outcomes. Suppose that the classifier rule is to define a threshold $x_c$ and to classify the wood as cherry if $x \leq x_c$, otherwise classify it as maple.

(a) Plot the true positive rate ($P[E^C|C] = 1 - P[E|C]$) vs. the false positive rate ($P[E|M]$) as $x_c$ is varied. This defines the *receiver operating characteristic* (ROC) for this simple classifier. Indicate the point of minimum error probability.

(b) On the same plot, indicate the true positive rate and false positive rate performance point of the minimum error classifier from Problem 5. Plot the ROC of an ideal perfect classifier that makes no mistakes.

(c) Suppose cherry is a premium wood and customers become upset when non-cherry wood is sold as cherry. Customers will reject an entire shipment if more than 1% is non-cherry wood. At what point would you operate on the ROC? At this point, what percentage of cherry wood are you misclassifying as non-premium wood?

Extra credit: For Problem 5 how would you define the optimal ROC? Compute the optimal ROC for this problem.