Review Exercises #2 and #3 (with solutions)

1. **Joint cdf or not.** Consider the function

\[ G(x, y) = \begin{cases} 
1 & \text{if } x + y \geq 0 \\
0 & \text{otherwise.} 
\end{cases} \]

Can \( G \) be a joint cdf for a pair of random variables? Justify your answer.

**Solution**

No. Note that for every \( x \),

\[ \lim_{y \to \infty} G(x, y) = 1. \]

But for any genuine marginal cdf,

\[ \lim_{x \to -\infty} F_X(x) = 0 \neq 1. \]

Therefore \( G(x, y) \) is not a cdf. Alternatively, assume that \( G(x, y) \) is a joint cdf for \( X \) and \( Y \), then

\[
\Pr\{-1 < X \leq 2, -1 < Y \leq 2\} = G(2, 2) - G(-1, 2) - G(2, -1) + G(-1, -1)
\]

\[ = 1 - 1 - 1 + 0 = -1. \]

But this violates the property that the probability of any event must be nonnegative.

2. **Max to Min ratio.** Let \( X_1 \) and \( X_2 \) be two independent random variables, each uniformly distributed between 0 and 1, i.e., \( X_i \sim \text{U}[0, 1] \). Find and sketch the cdf of

\[ Y = \max(X_1, X_2)/\min(X_1, X_2). \]

**Solution**

Since \( X_1 \sim \text{U}[0, 1] \) and \( X_2 \sim \text{U}[0, 1] \) are independent, their joint pdf is uniform over the square \( 0 \leq x_1, x_2 \leq 1 \). The cdf of \( Y \) can be found graphically by calculating the area of the region where \( Y \leq y \), as shown in Figure 1 on page 2.
To see this more clearly, consider $F_Y(y)$ for $y \geq 1$:

$$F_Y(y) = P \left\{ \frac{\max\{X_1, X_2\}}{\min\{X_1, X_2\}} \leq y \right\}$$

$$= P \left\{ \min\{X_1, X_2\} \geq \frac{\max\{X_1, X_2\}}{y} \right\}$$

$$= P \left\{ X_2 \geq \frac{X_1}{y}, \quad X_2 \leq X_1 \right\} + P \left\{ X_1 \geq \frac{X_2}{y}, \quad X_1 < X_2 \right\}$$

$$= P \left\{ \frac{X_1}{y} \leq X_2 \leq X_1 \right\} + P \left\{ \frac{X_2}{y} \leq X_1 < X_2 \right\}$$

$$= 2P \left\{ \frac{X_1}{y} \leq X_2 \leq X_1 \right\} \quad \text{(by symmetry)}$$

$$= 2 \cdot \frac{1}{2} \cdot 1 \cdot \left(1 - \frac{1}{y}\right) = \frac{y - 1}{y}.$$ 

Clearly $F_Y(y) = 0$ for $y < 1$. Note that $f_Y(y) = 1/y^2$ for $y \geq 1$.

3. **Conditional Independence v.s. Independence** Give an example of random variables $X, Y,$ and $Z$ where $f_{X,Z}(x, z) = f_X(x)f_Z(z)$ but $f_{X,Z|Y}(x, z|y) \neq f_{X|Y}(x|y)f_{Z|Y}(z|y)$ i.e. independence does not imply conditional independence.

**Solution**

One solution is let $X, Z \overset{iid}{\sim}$ Bern$(1/2)$ and $Y = X + Z$.

4. **Sum and difference.** Let $X$ and $Y$ be two random variables, and define $U = X - Y$ and $V = X + Y$. Find the minimum MSE linear estimate of $V$ given $U$ as a function of the random variables and $E(X)$, $E(Y)$, $\sigma_X$, $\sigma_Y$, $\rho_{X,Y}$, where $\sigma_X = \sqrt{\text{Var}(X)}$, $\rho_{X,Y} = \text{corr}(X, Y)$.
Solution

First we calculate the first and second moments of $U$ and $V$.

$$E(V) = E(X) + E(Y)$$
$$E(U) = E(X) - E(Y)$$

$$\sigma^2_V = \sigma^2_X + \sigma^2_Y + 2\rho_{X,Y}\sigma_X\sigma_Y$$
$$\sigma^2_U = \sigma^2_X + \sigma^2_Y - 2\rho_{X,Y}\sigma_X\sigma_Y$$

$$\text{Cov}(V, U) = \sigma^2_X - \sigma^2_Y.$$ 

The minimum MSE linear estimate of $V$ given $U$ is given by

$$\hat{V} = \frac{\text{Cov}(V, U)}{\sigma^2_U}(U - E(U)) + E(V).$$

Plugging in the moments of $U$ and $V$ gives the answer.

$$\hat{V} = \frac{\sigma^2_X - \sigma^2_Y}{\sigma^2_X + \sigma^2_Y - 2\rho_{X,Y}\sigma_X\sigma_Y}(U - (E(X) - E(Y)) + (E(X) + E(Y))$$

Note that $U$ and $V$ are positively correlated if $\sigma^2_X > \sigma^2_Y$, negatively correlated if $\sigma^2_X < \sigma^2_Y$, and uncorrelated if $\sigma^2_X = \sigma^2_Y$.

5. Covariance matrices. Which of the following matrices can be a covariance matrix? Justify your answer. Either construct a random vector $X$ with the given covariance matrix as a function of the i.i.d. zero mean unit variance random variables $Z_1, Z_2, Z_3$, or establish a contradiction as was done in lecture.

(a) $\begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}$  
(b) $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$  
(c) $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}$  
(d) $\begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 3 \end{bmatrix}$

Solution

a. No: not symmetric.

b. Yes: covariance matrix of $X_1 = Z_1 + Z_2$ and $X_2 = Z_1 + Z_3$.

c. Yes: covariance matrix of $X_1 = Z_1$, $X_2 = Z_1 + Z_2$, and $X_3 = Z_1 + Z_2 + Z_3$.

d. No: several justifications.
   - $\sigma^2_{Z_3} = 9 > \sigma^2_{Z_2}\sigma_{Z_3} = 6$, which contradicts the Schwarz inequality.
   - The matrix is not nonnegative definite since the determinant is $-2$.
   - One of the eigenvalues is negative ($\lambda_1 = -0.8056$).

6. Generating a GRV. Let $[X, Y]^T$ be a jointly Gaussian random vector with parameters:

$$\mu_X = 3, \mu_Y = 2, \sigma_X = 2, \sigma_Y = 1, \rho_{X,Y} = 0.5.$$ 

Using the uniform random number generator of MATLAB (the function `rand`) and an appropriate transformation, generate 5000 realizations of $[X, Y]^T$. In your code, you are not allowed to use the MATLAB function `randn`. Plot the realizations as points in the $x$-$y$-plane.
Solution

Here is a Matlab code snippet. The corresponding plot is shown in Figure 2 on page 4.

```matlab
muX = 3;
muy = 2;
sigmaX = 2;
sigmaY = 1;
rho = 0.5;
nPoints = 5000;

% Covariance matrix of the vector [X,Y]^T
Sigma = [sigmaX^2, rho*sigmaX*sigmaY; rho*sigmaX*sigmaY, sigmaY^2];

% Coloring matrix
SigmaHalf = chol(Sigma, 'lower'); % lower triangular square root

% Generate Z=[X,Y]^T realizations
Z = rand(2, nPoints); % uniform
Z = norminv(Z); % white Gaussian
Z = SigmaHalf*Z; % colored Gaussian
Z = Z + repmat([muX; muY], [1, nPoints]); % non-zero-mean colored Gaussian

% Make plot
figure;
plot(Z(1,:), Z(2,:), 'rx');
xlabel('x');
ylabel('y');
print -depsc2 generatingGRV_figure
```

Figure 2: Figure for the problem *Generating a GRV*.

7. Definition of Gaussian random vector. In lecture notes #3 we defined Gaussian random vectors via the joint pdf. There are other equivalent definitions, including the following very revealing definition. A random vector $\mathbf{X}$ with mean $\mu$ and covariance matrix $\Sigma$ is a GRV if and only if $Y = \mathbf{a}^T \mathbf{X}$ is Gaussian for every real vector $\mathbf{a} \neq \mathbf{0}$. In the lecture notes (Property 2) we showed that any linear transformation of a GRV results in a GRV. Thus the definition given in the lecture notes implies this new definition. In this problem you will prove the converse, i.e., that the new definition implies that the joint pdf of $\mathbf{X}$ has the form given in the lecture notes. You will do this using the characteristic function as follows:

a. Write down the definition of the characteristic function for $\mathbf{X}$. 
b. Define $Y = \omega^T X$. Note that the characteristic function of $X$ reduces to the characteristic function of $Y$ evaluated at $\omega = 1$.

c. By the new definition, $Y$ is Gaussian. Use this fact to write the characteristic function of $X$ in terms of the mean and variance of $Y$.

d. Write down the mean and variance of $Y$ in terms of $\omega$ and the mean and covariance matrix of $X$ and substitute in the characteristic function of $X$.

e. Conclude that the joint pdf of $X$ is Gaussian.

**Solution**

a. Just copying the definition from the lecture notes:

$$\Phi_X(\omega) = E\left(e^{i\omega^T X}\right).$$

b. Let $Y = \omega^T X$. Then $\Phi_X(\omega) = E(e^{iY}) = \Phi_Y(1)$.

c. Since $Y$ is Gaussian by the new definition, its characteristic function is

$$\Phi_Y(\omega) = e^{-\frac{1}{2}\omega^2 \sigma_Y^2 + i\omega \mu_Y}.$$

Thus

$$\Phi_X(\omega) = \Phi_Y(1) = e^{-\frac{1}{2}\sigma_Y^2 + i\mu_Y}.$$

d. By linearity (lecture notes #3),

$$\mu_Y = \omega^T \mu \quad \text{and} \quad \sigma_Y^2 = \omega^T \Sigma \omega.$$

e. Combining the previous two steps,

$$\Phi_X(\omega) = e^{-\frac{1}{2}\omega^T \Sigma \omega + i\omega^T \mu},$$

which is the characteristic function of a Gaussian pdf with mean $\mu$ and covariance matrix $\Sigma$. Therefore $X$ is a a Gaussian random vector.

Note: This proof shows the power of the characteristic function. Try to prove this without using the characteristic function!

8. **Jensen’s inequality.** A function $g(x)$ is said to be convex on an interval $(a, b)$ if for every $x_1, x_2$ in $(a, b)$ and for every $\lambda$ satisfying $0 \leq \lambda \leq 1$,

$$g(\lambda x_1 + (1-\lambda)x_2) \leq \lambda g(x_1) + (1-\lambda)g(x_2).$$

Further, $g(x)$ is said to be strictly convex if equality holds only for $\lambda = 0$ and $\lambda = 1$. It can be shown that if $g(x)$ is twice differentiable, then it is convex iff $g''(x) \geq 0$ for all $x$ in $(a, b)$ and strictly convex iff $g''(x) > 0$ for all $x$ in $(a, b)$. See Figure 3.

a. Show that if $g(x)$ is convex on $(a, b)$ and $X \in \mathcal{X} \subset (a, b)$ is a discrete random variable, then

$$E(g(X)) \geq g(E(X)).$$
Hint: Use induction on the number of values with nonzero probability.
Note: Jensen’s inequality holds for continuous random variables as well, which can be shown by discretizing the random variable and using a limiting argument. (You need not prove this.)
For the following parts, find the inequality relationship (≤ or ≥) and justify your answer.

b. \( E(e^{2X}) \) and \( e^{E(2X)} \).
c. \( E(\ln X) \) and \( \ln(E(X)) \) for \( X \geq 0 \).
d. \( (E(X^2))^6 \) and \( E(X^{12}) \).

**Solution (20 points)**

a. Proof by mathematical induction. For two \( x \) points with nonzero \( p_X(x) \), the proof follows by the definition of convexity. Now, suppose that the proof holds for \( n \) \( x \) points with nonzero \( p_X(x) \). To prove that it holds for \( n + 1 \) points, we need to show that

\[
g\left(\sum_{i=1}^{n+1} p_X(x_i)x_i\right) \leq \sum_{i=1}^{n+1} p_X(x_i)g(x_i).
\]

Define \( \lambda \) by

\[
\lambda = \sum_{i=1}^{n} p_X(x_i)x_i \Rightarrow 1 - \lambda = p_X(x_{n+1}).
\]

Then

\[
\sum_{i=1}^{n+1} p_X(x_i)g(x_i) = (1 - \lambda)g(x_{n+1}) + \lambda \sum_{i=1}^{n} \frac{p_X(x_i)}{\lambda} g(x_i)
\]

\[
\geq (1 - \lambda)g(x_{n+1}) + \lambda g\left(\sum_{i=1}^{n} \frac{p_X(x_i)}{\lambda} g(x_i)\right)
\]

\[
\geq g\left((1 - \lambda)x_{n+1} + \lambda \sum_{i=1}^{n} \frac{p_X(x_i)}{\lambda} g(x_i)\right)
\]

\[
= g\left(\sum_{i=1}^{n+1} p_X(x_i)x_i\right).
\]

The first inequality follows from the induction hypothesis and the second follows from the definition of convexity.

Once a function \( g(X) \) has been shown to be convex, applying Jensens inequality \( E(g(X)) \geq g(E(X)) \) gives the inequality relationship.

b. \( e^{2x} \) is convex since

\[
d^2 \frac{e^{2x}}{dx^2} = 4e^{2x} \geq 0
\]

and therefore

\[
E(e^{2X}) \geq e^{2E(X)} = e^{E(2X)}.
\]
c. \(-\ln(x)\) is convex since 
\[
\frac{d^2}{dx^2}(-\ln(x)) = \frac{1}{x^2} > 0
\]
and therefore 
\[E(-\ln(X)) \geq -\ln(E(X)).\]
Consequently, by linearity of expectation 
\[E(\ln(X)) \leq \ln(E(X)).\]
d. Let \(Y = X^2\). Now \(y^6\) is convex since 
\[
\frac{d^2}{dy^2}y^6 = 30y^4 \geq 0
\]
and therefore 
\[E(Y^{12}) = E(Y^6) \geq (E(Y))^6 = (E(X^2))^6.\]

9. \textit{Additive noise channel with path gain.} Consider the additive noise channel shown in Figure 7, 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 \(\sigma_X^2\), \(\sigma_Z^2\), \(a\), and \(b\).

\[\begin{align*}
X &\rightarrow a &\rightarrow b &\rightarrow Y = b(aX + Z)
\end{align*}\]

\textbf{Solution (10 points)}

First we find the mean and variance of \(Y\) and its covariance with \(X\). In the following we use the notation \(\sigma_X^2 = P\) and \(\sigma_Z^2 = N\).

\[
E(Y) = E(abX + bZ) = abE(X) + bE(Z) = 0
\]

\[
\text{Var}(Y) = E(abX + bZ)^2 - (E(abX + bZ))^2
\]
\[
= E(a^2b^2X^2 + 2ab^2XZ + b^2Z^2) - E(Y)^2
\]
\[
= a^2b^2E(X^2) + 2ab^2E(X)E(Z) + b^2E(Z^2) - E(Y)^2
\]
\[
= a^2b^2P + 0 + b^2N - 0 = a^2b^2P + b^2N
\]

\[
\text{Cov}(X,Y) = E[(X - E(X))(Y - E(Y))]
\]
\[
= E(XY)
\]
\[
= E[X(abX + bZ)]
\]
\[
= abE(X^2) + bE(XZ)
\]
\[
= abP + bE(X)E(Z)
\]
\[
= abP
\]
The MMSE linear estimate of $X$ given $Y$ is given by

\[ \hat{X} = \frac{\text{Cov}(X,Y)}{\sigma_Y^2} (Y - \text{E}(Y)) + \text{E}(X) \]

\[ = \frac{a b P}{a^2 b^2 P + b^2 N} (Y - \text{E}(Y)) + \text{E}(X) \]

\[ = \frac{a P}{b(a^2 P + N)} Y \]

The MSE of the linear estimate is the minimum MMSE:

\[ \text{MMSE} = \sigma_X^2 - \frac{\text{Cov}^2(X,Y)}{\sigma_Y^2} \]

\[ = P - \frac{a^2 b^2 P^2}{a^2 b^2 P + b^2 N} \]

\[ = \frac{a^2 P^2 + P N - a^2 P^2}{a^2 P + N} = \frac{P N}{a^2 P + N} \]