1. True or False (20 points)

(a) $X_1$ is a Gaussian random variable, and so is $X_2$. Then $X_1$ and $X_2$ are independent if they are uncorrelated.

No.

Unless $X_1$ and $X_2$ are jointly Gaussian, this property can not hold. For a counter example, see handout “Gaussian random vectors” pp. 18.

(b) If $A$ is independent of $B_1$ and $B_2$ respectively, and $B_1 \subseteq B_2$, then $A$ is also independent of $(B_2 - B_1)$.

Yes.

Since $P(A(B_2 - B_1)) = P(AB_2 - AB_1) = P(AB_2) - P(AB_1)$ by using the fact that $AB_1 \subseteq AB_2$, now with $P(AB_2) = P(A)P(B_2)$ and $P(AB_1) = P(A)P(B_1)$, one has that $P(A(B_2 - B_1)) = P(A)P(B_2 - B_1)$ by noting that $B_1 \subseteq B_2$.

(c) Given two random variables $X$ and $Y$, if $E[f(X)g(Y)] = E[f(X)]E[g(Y)]$ for all continuous functions $f(\cdot), g(\cdot)$, then $X$ and $Y$ are independent.

Yes.

Let $f(\cdot), g(\cdot)$ be characteristic functions of $X, Y$, then

$$
\phi_{X,Y}(\omega, \theta) = E[f(X)g(Y)] = E[f(X)]E[g(Y)] = \phi_X(\omega)\phi_Y(\theta)
$$

which gives the independence result.

Or, consider the corresponding probability triple $(\Omega, \mathcal{F}, P)$ with respect to $X, Y$. Let $\forall A \times B \in \mathcal{F}$ and $f(\cdot) = 1_A(\cdot)$, $g(\cdot) = 1_B(\cdot)$, where 1 denotes the indication function ($1_A(\omega) = 1, \omega \in A$, or $1_A(\omega) = 0, \omega \notin A$), then

$$
E[f(X)g(Y)] = E[1_A(X)1_B(Y)] = P(X \in A, Y \in B)
$$

and

$$
E[f(X)] = E[1_A(X)] = P(X \in A), E[g(Y)] = E[1_B(Y)] = P(Y \in B).
$$

Therefore, $X,Y$ are independent.

(d) If $X, Z$ are two independent random variables, then $E[X|Y, Z] = E[X|Y]$ for any random variable $Y$.

No.

Consider a counter-example: with $Y = Z + X$, then $E[X|Y, Z] = Y - Z = X$ which completely recovers $X$, whereas $E[X|Y] = E[X|Z + X]$ only gives a conditional expectation. Or, one could use density function to see that in general \( \frac{f_{X,Y,Z}(x,y,z)}{f_{Y,Z}(y,z)} \neq \frac{f_{X,Y}(x,y)}{f_Y(y)} \) even though $X, Z$ are independent.

2. Conditional distribution (20 points)

Let $X, Y, Z$ be independent and identically distributed non-negative rv’s with common density $f(x) = \exp(-x)$ for $x \geq 0$.

(a) the pdf of $X$, conditional the event that $X \leq 1$. 

(b) the pdf of $X$, conditional on the event $X \leq 1$ AND $X + Y \geq 1$.
(c) the joint pdf of $X$ and $Y$, conditional on the event $X + Y \leq 1$ AND $X + Y + Z \geq 1$.
(d) Show that the pdf of $X$ conditional on the event $X + Y \leq 1$ AND $X + Y + Z \geq 1$ is the same as the pdf of the rv $W = \min\{U_1, U_2\}$, where $U_1$ and $U_2$ are two independent and identically distributed rv’s uniformly distributed in $[0, 1]$.

**Solution:**
(a) We start from finding cumulative density function

$$F_X(x|X \leq 1) = P\{X \leq x|X \leq 1\} = \frac{P\{X \leq x, X \leq 1\}}{P\{X \leq 1\}}$$

$$= \begin{cases} 
\frac{\int_0^x e^{-\alpha} d\alpha}{\int_0^1 e^{-\alpha} d\alpha} = \frac{1 - e^{-x}}{1 - e^{-1}} & \text{when } x \geq 1 \\
\frac{\int_0^1 e^{-\alpha} d\alpha}{\int_0^1 e^{-\alpha} d\alpha} = 1 & \text{when } 0 \leq x \leq 1 
\end{cases}$$

Hence, the probability density function is given by

$$f_X(x|X \leq 1) = \frac{d}{dx}F_X(x|X \leq 1) = \begin{cases} 
0 & x \geq 1 \\
\frac{e^{-x}}{1 - e^{-1}} & 0 \leq x \leq 1 
\end{cases}$$

(b) Similarly, let $0 \leq x \leq 1$, one has

$$F_X(x|X \leq 1, X + Y \leq 1) = \frac{P\{X \leq x, X \leq 1, X + Y \leq 1\}}{P\{X \leq x, X + Y \leq 1\}}$$

$$= \frac{\int_0^x \int_0^{1-x} e^{-u} e^{-y} du \, dy}{\int_0^1 \int_0^{1-x} e^{-u} e^{-y} du \, dy}$$

$$= \frac{1 - e^{-x} - e^{-1}x}{1 - 2e^{-1}}$$

Note that the integrations in the above expression are double integrals. The density function is then given by

$$f_X(x|X \leq 1, X + Y \leq 1) = \begin{cases} 
0 & \text{else} \\
\frac{e^{-x} - e^{-1}}{1 - 2e^{-1}} & 0 \leq x \leq 1 
\end{cases}$$

(c) Similar to part b), we consider $0 \leq x \leq 1, 0 \leq y \leq 1$ with $x + y \leq 1$,

$$F_{X,Y}(x,y|X + Y \leq 1, X + Y + Z \geq 1) = \frac{P\{X \leq x, Y \leq y, X + Y \leq 1, X + Y + Z \geq 1\}}{P\{X \leq x, Y \leq y, X + Y \leq 1\}}$$

$$= \frac{\int_0^x \int_0^{1-x} \int_0^\infty e^{-u} e^{-v} e^{-z} du \, dv \, dz}{\int_0^1 \int_0^{1-x} \int_0^\infty e^{-u} e^{-v} e^{-z} du \, dv \, dz}$$

$$= \frac{1}{2} e^{-1} xy - 2xy$$

2
Hence,
\[
F_{X,Y}(x,y | X+Y \leq 1, X+Y+Z \geq 1) = \begin{cases} 
2xy & 0 \leq x \leq 1, 0 \leq y \leq 1, x+y \leq 1 \\
1 & x \geq 0, y \geq 0, x+y \geq 1 
\end{cases}
\]
\[
f_{X,Y}(x,y | X+Y \leq 1, X+Y+Z \geq 1) = \begin{cases} 
2 & 0 \leq x \leq 1, 0 \leq y \leq 1, x+y \leq 1 \\
0 & x \geq 0, y \geq 0, x+y \geq 1 
\end{cases}
\]

(d) Since the density of \{ \{X,Y\} | X+Y \leq 1, X+Y+Z \geq 1 \} is obtained in part c), the density on X can be obtained by integrating this joint density function:
\[
f_X(x | X+Y \leq 1, X+Y+Z \geq 1) = \int_0^{1-x} f_{X,Y}(x,y | X+Y \leq 1, X+Y+Z \geq 1) dy = 2(1-x), \quad 0 \leq x \leq 1.
\]

Note that conditioned on \{ X+Y \leq 1, X+Y+Z \geq 1 \}, the random variable Y can only take values inside \([0, 1-x]\), which is the upper and lower limits used in the above formula.

Let \( W = \min(U_1, U_2) \) where \( U_1, U_2 \) are i.i.d uniformly distributed random variables in \([0, 1]\). Consulting the lecture notes,
\[
F_W(w) = F_{U_1}(w) + F_{U_2}(w) - F_{U_1}(w)F_{U_2}(w) = 2w - w^2, \quad 0 \leq w \leq 1.
\]

Hence, the density of \( W \) is given by
\[
f_W(w) = \begin{cases} 
2(1-w) & 0 \leq w \leq 1 \\
0 & \text{else}
\end{cases}
\]

This has the same form as the density of \( X \).

3. MMSE and LLSE (20 points)

Let \( X \) be a scalar random variable with a known distribution that is symmetric about 0 (hence you may assume all moments of \( X \) are known). Let \( Y = X^3 \) be your observation. Express your answers in terms of \( Y \) and the moments of \( X \):

(a) What is the best estimate \( g(Y) \) for \( X \) which minimizes the mean squared error \( E[(g(Y) - X)^2] \)?

Solution: The best estimate \( g(Y) = Y^{\frac{1}{3}} \) which renders the mean squared error zero.

(b) What is the best linear estimate \( aY + b \) for \( X \) which minimizes the squared error \( E[(aY + b - X)^2] \)?

Solution: First, noting that \( E[X] = 0 \) by symmetry, it is obvious \( b = 0 \). Then, setting first order derivative of \( E[(aY - X)^2] \) to zero, one obtains
\[
a = \frac{\text{cov}(X,Y)}{\text{var}(Y)} = \frac{E[Y^{\frac{4}{3}}]}{E[Y^2]} \]
(see problem set #3 solution set for details).

(c) Now suppose you have two more observations: \( Z_1 = X + W \) and \( Z_2 = X - 2W \) where \( W \) is \( N(0,1) \) and independent from \( X \) and \( Y \). What is the best linear estimate for \( X \) given \( Y, Z_1 \) and \( Z_2 \)?

Solution: Noting that \( X \) can be completely recovered from \( Z_1, Z_2 \), the best linear estimate is given by \( X = \frac{1}{3}(2Z_1 + Z_2) + 0 \cdot Y \).

(d) Let \( V = X^2 \), what is the best estimate \( g(V) \) for \( X \) which minimizes the mean squared error \( E[(g(V) - X)^2] \)? Explain.

Solution: With \( V = X^2 \), it follows the estimate of \( X = \pm \sqrt{V} \) with equal probabilities. Then, the best estimate is \( g(V) = 0 \) since otherwise the mean squared error would be larger.
4. Summation and multiplication of random sequences (20 points)

Let \( X_n \rightarrow X \) and \( Y_n \rightarrow Y \) in probability.

(a) Prove that \( aX_n + bY_n \rightarrow aX + bY \) in probability for any constants \( a, b \in \mathbb{R} \).

**Solution:** The conclusion is trivial if \( a, b = 0 \). Hence, we assume that \( a, b \neq 0 \). For \( \forall \epsilon > 0 \), let 

\[ A = \{|a||X_n - X| + |b||Y_n - Y| > \epsilon\} \]

then

\[
P(|(aX_n + bY_n) - (aX + bY)| > \epsilon) \\
\leq P(A) = P(A, |X_n - X| > \frac{\epsilon}{2|a|}) + P(A, |X_n - X| \leq \frac{\epsilon}{2|a|}) \\
\leq P(|X_n - X| > \frac{\epsilon}{2|a|} ) + P(|Y_n - Y| > \frac{\epsilon}{2|b|}, |X_n - X| \leq \frac{\epsilon}{2|a|} ) \\
\leq P(|X_n - X| > \frac{\epsilon}{2|a|}) + P(|Y_n - Y| > \frac{\epsilon}{2|b|}) \\
\rightarrow 0, \ \epsilon \rightarrow 0.
\]

Here, we repeatedly use the simple inequality \( P(A) \leq P(B) \) for sets (events) \( A \subseteq B \).

(b) Prove that \( X_nY_n \rightarrow XY \) in probability.

**Solution:**

Assume \( X \) is finite a.s., i.e. \( P\{|X| < \infty\} = 1 \), and so does \( Y \). Now, we first prove that if \( X_n \rightarrow X \) i.p., then \( X_nY \rightarrow XY \) i.p.. For \( \forall N \geq 1 \), note the following inequality,

\[
P(|X_nY - XY| > \epsilon) \\
= P(|Y||X_n - X| > \epsilon, |Y| < N) + P(|Y||X_n - X| > \epsilon, |Y| \geq N) \\
\leq P(|X_n - X| > \epsilon N, |Y| < N) + P(|Y| \geq N) \\
\leq P(|X_n - X| > \epsilon N) + P(|Y| \geq N) \\
\rightarrow 0, \ \ N \rightarrow \infty, \ \epsilon \rightarrow 0.
\]

Next, let \( B = \{|X_nY_n - XY_n| + |XY_n - XY| > \epsilon\} \), the convergence of \( X_nY_n \) can be proved using the following relationship,

\[
P(|X_nY_n - XY| > \epsilon) = P(|X_nY_n - XY_n + XY_n - XY| > \epsilon) \\
\leq P(B) = P(B, |XY_n - XY| > \epsilon/2) + P(B, |XY_n - XY| \leq \epsilon/2) \\
\leq P(|XY_n - XY| > \epsilon/2) + P(|X_nY_n - XY_n| > \epsilon/2, |XY_n - XY| \leq \epsilon/2) \\
\leq P(|XY_n - XY| > \epsilon/2) + P(|X_nY_n - XY_n| > \epsilon/2) \\
\rightarrow 0, \ \epsilon \rightarrow 0.
\]

Or, one first prove (using a similar argument) that if \( X_n \rightarrow X \) i.p., then \( X_n^2 \rightarrow X^2 \) i.p. Next, using the fact that \( X_nY_n = [(X_n + Y_n)^2 - (X_n - Y_n)^2]/4, XY = [(X + Y)^2 - (X - Y)^2]/4 \) and the conclusion from part a), it is straightforward to see \( X_nY_n \rightarrow XY \) i.p.

5. Characteristic function (20 points)

Suppose \( X \) and \( Y \) are independent and identically distributed with means 0 and variances 1 and let \( \phi(\omega) \) be the common characteristic function, i.e. \( \phi(\omega) = E[e^{i\omega X}] = E[e^{i\omega Y}] \). Furthermore suppose that \( E[X - Y |X + Y] = 0 \) and \( \text{Var}(X - Y |X + Y) = 2 \).

(a) Deduce that

\[
\phi(\omega)^2 = \phi'(\omega)^2 - \phi(\omega)\phi''(\omega), \tag{1}
\]

**Solution:** Let \( U \equiv X + Y, V \equiv X - Y \). Then, their joint characteristic function is given by

\[
\psi(\omega, \theta) = E[e^{i\omega U + i\theta V}] = E[e^{i\omega (X + Y) + i\theta (X - Y)}] = \phi(\omega + \theta)\phi(\omega - \theta). \tag{2}
\]

Using what is given,

\[
\frac{\partial^2}{\partial \theta^2} \psi(\omega, \theta) \bigg|_{\theta=0} = -E[V^2 e^{i\omega U}] = -E\{e^{i\omega U}E[V^2 |U]| = -E[2e^{i\omega U}] = -2\phi(\omega)^2.
\]
However, by (2),

\[
\frac{\partial^2}{\partial \theta^2} \psi(\omega, \theta) \bigg|_{\theta = 0} = 2\{\phi''(\omega)\phi(\omega) - \phi'(\omega)^2\},
\]

yielding the required differential equation (1).

(b) Notice that the above equation is equivalent to

\[
\frac{d}{d\omega} \left( \frac{\phi'(\omega)}{\phi(\omega)} \right) = -1.
\]

Using this fact to show that \(X, Y\) are in fact \(N(0, 1)\) variables.

**Solution:** Solving the differential equation (3), one has that

\[
\log \phi(\omega) = a + b\omega - \frac{1}{2}\omega^2
\]

for constants \(a, b\), whence \(\phi(\omega) = e^{-\frac{1}{2}\omega^2}\). Therefore, \(X, Y\) are Gaussian \(N(0, 1)\) r.v.'s.