SI 140A: Probability and Statistics for EECS
Spring 2025, Instructor: Xavier Lagorce, TA: Junye Wang, Chuyan Zhou

Homework 7

Due: 23:59 (CST), May 9, 2025
Full Mark: 50 points, 6 problems including 1 bonus.

Read all the instructions below carefully before you start working on the assignment, and before you make a submission.

- You are required to write down all the major steps towards making your conclusions; otherwise you may obtain limited points of the problem.
- Write your homework in English; otherwise you will get no points of this homework.
- Any form of plagiarism will lead to 0 point of this homework.
- If there are any descriptions unclear, we recommend you to ask on Piazza.

Problem 1

(Independence of Joint Normals *2'+2'+2'***)**

Let X, Y, Z be r.v.s such that $X \sim \mathcal{N}(0,1)$, and conditional on $X = x$, Y and Z are i.i.d. $\mathcal{N}(x,1)$.

(a) Find the joint PDF of X, Y, Z.

(b) By definition, Y and Z are conditionally independent given X. Discuss intuitively whether or not Y and Z are also unconditionally independent.

(c) Find the joint PDF of Y and Z. You can leave your answer as an integral, though the integral can be done with some algebra (such as completing the square) and facts about the Normal distribution.

Problem 2

(Play with Min & Max *5'+5'***)**

Let U_1, U_2, U_3 be i.i.d. Unif$(0,1)$, and let $L = \min(U_1, U_2, U_3)$, $M = \max(U_1, U_2, U_3)$.

(a) Find the marginal CDF and marginal PDF of M, and the joint CDF and joint PDF of L, M.

Hint: For the latter, start by considering $P(L \geq l, M \leq m)$.

(b) Find the conditional PDF of M given L.

Problem 3

(Multivariate Normal *4'+4'+4'***)**

Let X and Y be i.i.d. $\mathcal{N}(0,1)$, and let S be a random sign 1 or -1 , with equal probabilities independent of (X, Y).

(a) Determine whether or not $(X, Y, X+Y)$ is Multivariate Normal.

(b) Determine whether or not $(X, Y, SX+SY)$ is Multivariate Normal.

(c) Determine whether or not (SX, SY) is Multivariate Normal.

Problem 4

(Bivariate Normal & Correlations *3'+3'+5'***)**

Let Z_1, Z_2 be two i.i.d. random variables satisfying standard normal distributions, i.e., $Z_1, Z_2 \sim \mathcal{N}(0,1)$. Define

$$X = \sigma_X Z_1 + \mu_X$$
$$Y = \sigma_Y \left(\rho Z_1 + \sqrt{1-\rho^2} Z_2\right) + \mu_Y$$

where $\sigma_X > 0, \sigma_Y > 0, -1 < \rho < 1$.

(a) Show that X and Y are bivariate normal.

(b) Find the correlation coefficient between X and Y, i.e., $\mathrm{Corr}(X,Y)$.

(c) Find the joint PDF of X and Y.

Problem 5

(Interpreting Covariance *3'+5'+3'***)**

This problem explores a visual interpretation of covariance. Data are collected for $n \geq 2$ individuals, where for each individual two variables are measured (e.g., height and weight). Assume independence across individuals (e.g., person 1's variables gives no information about the other people), but not within individuals (e.g., a person's height and weight may be correlated).

Let $(x_1, y_1), \ldots, (x_n, y_n)$ be the n data points. The data are considered here as fixed, known numbers-they are the observed values after performing an experiment. Imagine plotting all the points (x_i, y_i) in the plane, and drawing the rectangle determined by each pair of points. For example, the points $(1,3)$ and $(4,6)$ determine the rectangle with vertices $(1,3), (1,6), (4,6), (4,3)$.

The signed area contributed by (x_i, y_i) and (x_j, y_j) is the area of the rectangle they determine if the slope of the line between them is positive, and is the negative of the area of the rectangle they determine if the slope of the line between them is negative. (Define the signed area to be 0 if $x_i = x_j$ or $y_i = y_j$, since then the rectangle is degenerate.) So the signed area is positive if a higher x value goes with a higher y value for the pair of points, and negative otherwise. Assume that the x_i are all distinct and the y_i are all distinct.

(a) The sample covariance of the data is defined to be

$$r = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

where

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \text{ and } \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

are the sample means. (There are differing conventions about whether to divide by $n-1$ or n in the definition of sample covariance, but that need not concern us for this problem.)

Let (X, Y) be one of the (x_i, y_i) pairs, chosen uniformly at random. Determine precisely how $\mathrm{Cov}(X, Y)$ is related to the sample covariance.

(b) Let (X, Y) be as in (a), and $(\tilde{X}, \tilde{Y})$ be an independent draw from the same distribution. That is, (X, Y) and $(\tilde{X}, \tilde{Y})$ are randomly chosen from the n points, independently (so it is possible for the same point to be chosen twice).

Express the total signed area of the rectangles as a constant times $E((X - \tilde{X})(Y - \tilde{Y}))$. Then show that the sample covariance of the data is a constant times the total signed area of the rectangles.

Hint: Consider $E((X - \tilde{X})(Y - \tilde{Y}))$ in two ways: as the average signed area of the random rectangle formed by (X, Y) and $(\tilde{X}, \tilde{Y})$, and using properties of expectation to relate it to $\mathrm{Cov}(X, Y)$. For the former, consider the n^2 possibilities for which

Problem 5 continued on next page...

point (X, Y) is and which point $(\tilde{X}, \tilde{Y})$; note that n such choices result in degenerate rectangles.

(c) Based on the interpretation from (b), give intuitive explanations of why for any r.v.s W_1, W_2, W_3 and constants a_1, a_2, covariance has the following properties:

(i) $\mathrm{Cov}(W_1, W_2) = \mathrm{Cov}(W_2, W_1)$;

(ii) $\mathrm{Cov}(a_1 W_1, a_2 W_2) = a_1 a_2 \mathrm{Cov}(W_1, W_2)$;

(iii) $\mathrm{Cov}(W_1 + a_1, W_2 + a_2) = \mathrm{Cov}(W_1, W_2)$;

(iv) $\mathrm{Cov}(W_1, W_2 + W_3) = \mathrm{Cov}(W_1, W_2) + \mathrm{Cov}(W_1, W_3)$.

Problem 6

(Bonus: Conditional Independence *2'+2'+2'+2'+2'***)**

We use the notation $X \perp\!\!\!\perp Y \mid Z$ to represent the statement: random variables X and Y are conditionally independent given random variable Z. Now given any four continuous random variables X, Y, Z, W, show the following properties of conditional independence:

(a) Symmetry:
$$X \perp\!\!\!\perp Y \mid Z \iff Y \perp\!\!\!\perp X \mid Z.$$

(b) Decomposition:
$$X \perp\!\!\!\perp (Y, W) \mid Z \Rightarrow X \perp\!\!\!\perp Y \mid Z.$$

(c) Weak Union:
$$X \perp\!\!\!\perp (Y, W) \mid Z \Rightarrow X \perp\!\!\!\perp (Y, W) \mid (Z, W).$$

(d) Contract:
$$X \perp\!\!\!\perp Y \mid Z \;\&\; X \perp\!\!\!\perp W \mid (Y, Z) \iff X \perp\!\!\!\perp (Y, W) \mid Z.$$

(e) Intersection: For any positive joint PDF of X, Y, Z, W,
$$X \perp\!\!\!\perp Y \mid (Z, W) \;\&\; X \perp\!\!\!\perp Z \mid (Y, W) \iff X \perp\!\!\!\perp (Y, Z) \mid W.$$

In fact, these properties are found by Judea Pearl, who won 2011 Turing Award for fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning. As Judea Pearl commented: "Exploiting conditional independence to generate fast probabilistic computations is one of the main contributions CS has made to probability theory."