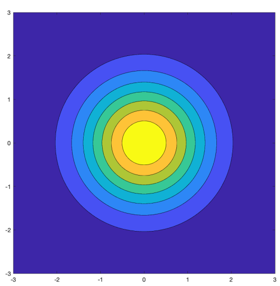
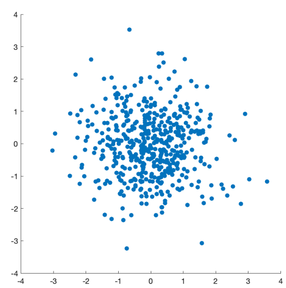
EK381 Probability, Statistics, and Data Science for Engineers | Boston University
Prof Carruthers | Summer 2025

Homework 7: finish by 6/16

Reading: Notes: Chapter 4, 5. **Videos:** 5.1 - 5.6

Problem 7.1 (**Video 5.1, 5.2, 5.3**) In the videos, we introduced the contour plot as a useful way to visualize the joint PDF. The *scatter plot* is a related method for visualizing two-dimensional datasets, which places a marker (e.g., a dot) for each point in the dataset. For example, below are the contour plot and scatter plot (for 500 data points) for a pair of independent Gaussian$(0, 1)$ random variables.





We have provided several scenarios describing pairs of (continuous) random variables via contour plots, scatter plots, and equations. For each of the following parts, list the scenarios that satisfy the specified criteria. For the equations, this will require some quick calculations, but for the contour and scatter plots you should be able to reason visually without calculations.

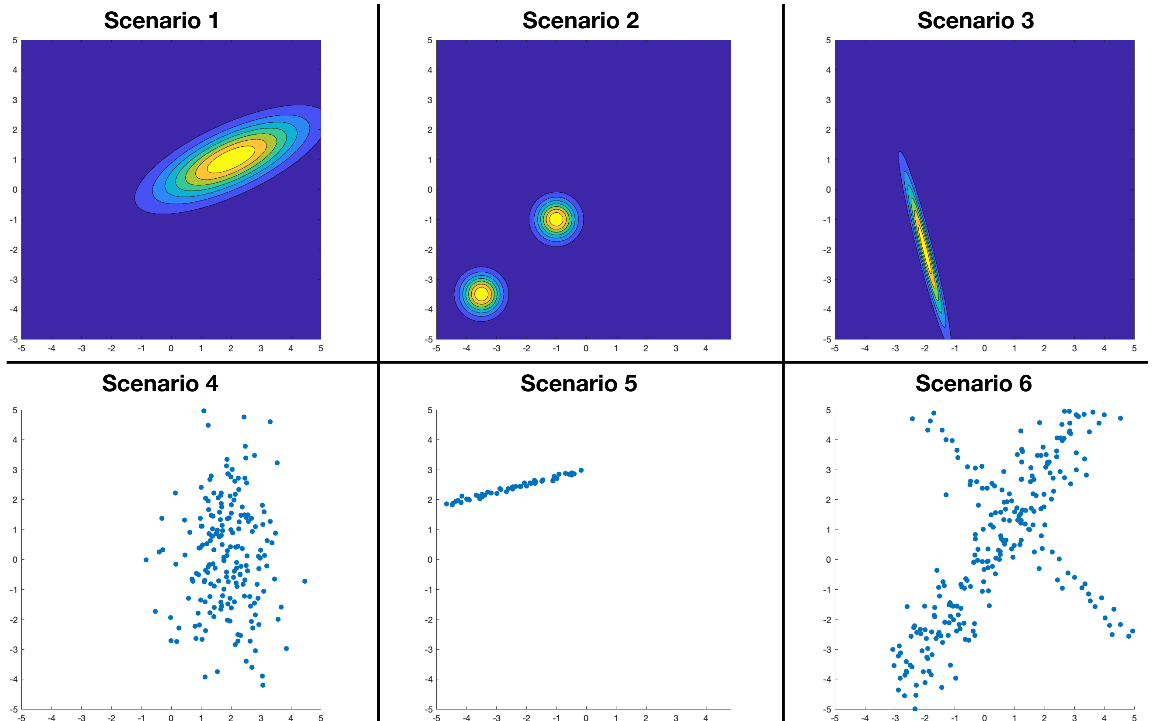
(a) $\mathbb{E}[X]$ noticeably more than 0

(b) $\mathbb{E}[Y]$ noticeably less than 0

(c) $\mathsf{Var}[X]$ noticeably larger than $\mathsf{Var}[Y]$

(d) $\mathsf{Cov}[X, Y]$ noticeably more than 0

(e) $\left|\rho_{X,Y}\right|$ close to 1



Scenario 7

$X = U - V, \quad Y = U + V$

$\mathbb{E}[U] = 1, \ \mathbb{E}[V] = 2,$

$\mathsf{Var}[U] = \mathsf{Var}[V] = 1,$

$\mathsf{Cov}[U, V] = -\frac{1}{2}$

Scenario 8

$Y = X + Z$

X is Gaussian$(0, 1)$

Z is Gaussian$(0, 3)$

X, Z independent

Scenario 9

$Y = -3X + 5$

$\mathbb{E}[X] = 2, \ \mathsf{Var}[X] = 1$

Problem 7.2 (Video 5.1 - 5.6, Quick Calculations)

For each of the scenarios below, determine the requested quantities. (You should be able to do this without any long calculations or integration.)

(a) Let X and Y be random variables with $\mathbb{E}[X] = \mathbb{E}[Y] = 1$, $\mathbb{E}[X^2] = \mathbb{E}[Y^2] = 5$, and $\rho_{X,Y} = -\frac{1}{4}$. Calculate $\mathsf{Var}[X]$ and $\mathsf{Cov}[X, Y]$.

(b) Let X be a random variable with $\mathsf{Var}[X] = 5$ and let $Y = 2X - 3$. Calculate $\mathsf{Cov}[X, Y]$ and $\rho_{X,Y}$.

(c) Let X and Y be random variables with $\mathbb{E}[X] = 1$, $\mathbb{E}[Y] = -1$, $\mathsf{Var}[X] = 4$, $\mathsf{Var}[Y] = 2$, and $\mathsf{Cov}[X, Y] = 1$. Let $W = X + Y$ and $Z = 2X - Y$. Calculate $\mathbb{E}[W + Z]$ and $\mathsf{Cov}[W, Z]$.

(d) Let X and Y be jointly Gaussian random variables with $\mathbb{E}[X] = 2$, $\mathbb{E}[Y] = -1$, $\mathsf{Var}[X] = 4$, $\mathsf{Var}[Y] = 1$, and $\mathsf{Cov}[X, Y] = -2$. Calculate $\mathbb{P}[2X - Y \leq 1]$ and $\mathbb{P}[X \leq Y]$. (You may leave your answers in terms of the standard normal CDF $\Phi(z)$.)

(e) Let X and Y be jointly Gaussian with $\mathbb{E}[X] = \mathbb{E}[Y] = 0$, $\mathsf{Var}[X] = \mathsf{Var}[Y] = 2$, and $\mathsf{Cov}[X, Y] = 1$. Calculate $\mathbb{E}[X|Y = 2]$ and $\mathbb{P}[X < 0|Y = 2]$. (You may leave your second answer in terms of the $\Phi(z)$ function.)

(f) Let X_1 and X_2 be jointly Gaussian with $\mathbb{E}[X_1] = \mathbb{E}[X_2] = 0$, $\mathsf{Var}[X_1] = \mathsf{Var}[X_2] = 1$, and $\mathsf{Cov}[X_1, X_2] = \frac{1}{3}$. Define $Y_1 = 2X_1 + 1$ and $Y_2 = X_1 - X_2 - 1$. Define the random vector $\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$. Calculate the mean vector $\underline{\mu}_{\underline{Y}}$ and the covariance matrix $\mathbf{\Sigma}_{\underline{Y}}$.

Problem 7.3 (**Video 5.1, 5.2, 5.3, 5.4**, **Lecture Problem**)

A robot is tasked with moving 4 meters in both the x-direction and y-direction, starting from the origin. However, due to wheel slippage and uneven terrain, its actual movement deviates from the intended path. These deviations are modeled by a bivariate Gaussian distribution, where the errors in both directions are positively correlated. Specifically, (X, Y) is jointly Gaussian with means $\mathbb{E}[X] = \mathbb{E}[Y] = 4$, variances $\mathsf{Var}[X] = \mathsf{Var}[Y] = 0.25$, and $\mathsf{Cov}[X, Y] = 0.1$.

(a) Compute the expected value $\mathbb{E}[W]$ and variance $\mathsf{Var}[W]$ of the robot's movement in the direction $W = \frac{1}{\sqrt{2}}X + \frac{1}{\sqrt{2}}Y$.

(b) Calculate the probability that the robot's movement along this direction exceeds 6 meters.

(c) Suppose that, after the robot has completed its movement, we determine that its exact position along the y-axis is 4.5 meters. Find the conditional distribution of the robot's position X given that Y = 4.5. Don't forget the parameters.

(d) More generally, for a given y position, compute the conditional expected value of its x position, $E[X|Y = y]$.

Problem 7.4 (**Video 5.5, 5.6**, **Lecture Problem**)

Jointly Gaussian random variables play an important role in probability theory, due partly to the fact that linear combinations of Gaussians are themselves Gaussian. This allows us to answer complex questions by only calculating means and variances. Here, we will explore an application of this phenomenon to the expected value. Let $X_1, \ldots, X_n$ be independent Gaussian random variables, with expected values $\mathbb{E}[X_i] = \mu$ and variances $\mathsf{Var}[X_i] = \sigma^2$ for $i = 1, \ldots, n$ and let $Y = \frac{1}{n}\sum_{i=1}^{n} X_i$ be their average. Intuitively, the random variable Y should get "closer" to μ as the number of samples n increases. Below, we will try to make this intuition precise.

(a) Define $\underline{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$. Determine its mean vector $\mathbb{E}[\underline{X}]$ and covariance matrix $\mathbf{\Sigma}_{\underline{X}}$.

(b) Express $Y = \mathbf{A}\underline{X}$ for some matrix $\mathbf{A}$. (Note that row vectors and column vectors are special cases of matrices.)

(c) Determine $\mathbb{E}[Y]$.

(d) Determine $\mathsf{Var}[Y]$.

(e) Calculate the probability that Y is more than $\delta > 0$ away from its mean, $\mathbb{P}\big[|Y - \mathbb{E}[Y]| > \delta\big]$. Express your answer in terms of the standard normal complementary CDF $Q(z) = 1 - \Phi(z)$.

(f) Using the fact that $Q(3.29) = \frac{1}{2000}$, calculate how many samples n are needed, as a function of the variance σ^2 to guarantee that $\mathbb{P}\big[|Y - \mathbb{E}[Y]| > \frac{1}{10}\big]$ is $\frac{1}{1000}$ or smaller.