EECS 126 Probability and Random Processes — UC Berkeley, Spring 2022
Kannan Ramchandran — February 22, 2022

Midterm 1

Last Name	First Name	SID

Left Neighbor First and Last Name	**Right Neighbor** First and Last Name

***Rules*.**

- **Unless otherwise stated, all your answers need to be justified and your work must be shown. Answers without sufficient justification will get no credit.**
- You have 80 minutes to complete the exam. (DSP students with $X\%$ time accommodation should spend $80 \cdot X\%$ time on the exam).
- This exam is not open book. You may reference one double-sided handwritten sheet of paper. No calculator or phones allowed.
- Collaboration with others is strictly prohibited. If you are caught cheating, you may fail the course and face disciplinary consequences.
- Write in your SID on every page to receive 1 point.

Problem	points earned	out of
SID		1
Problem 1		49
Problem 2		14
Problem 3		26
Problem 4		21
Total		111

1 Potpourri

(a) **Foraging [7 points]**
You find 2 different plants. A plant is green with probability $\frac{2}{3}$ and edible with probability $\frac{2}{5}$. Each plant's properties are independent of each other, and a plant's color is independent of its edibility. You are told that at least one of them is both edible and green. What is the probability that both plants are edible?

Let's define the following events:

- Let $L :=$ at least one of them is edible and green
- Let $E :=$ both plants are edible

Then we find the following:

$$\begin{aligned} P(E \mid L) &= \frac{P(L \mid E)\,P(E)}{P(L)} \\ &= \frac{(1-(1/3)^2)\cdot(2/5)^2}{2/5\cdot 2/3\cdot 2-(2/5\cdot 2/3)^2} \\ &= \frac{4}{13}. \end{aligned}$$

(b) **Comparing Gaussians [7 points]**
Let $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(2,3)$, where X and Y are independent. What is $P(X > Y)$? You may leave your answer in terms of Φ, the CDF of the standard normal distribution.

$$\begin{aligned} P(X > Y) &= P(X - Y > 0) \\ &= P(\mathcal{N}(-2,4) > 0) \\ &= P\left(\mathcal{N}(0,1) > \frac{2}{\sqrt{4}}\right) \\ &= 1 - \Phi(1) \end{aligned}$$

(c) **Uncorrelatedness [7 points]**
If X and Y are uncorrelated, are X^2 and Y also uncorrelated? Provide complete justification for full credit. Additionally, provide a counterexample if the answer is false.

The answer is false.

Consider the following counterexample. Let $X \sim \text{Uniform}[-1, 1]$ and $Y = X^2$. Then $\mathrm{E}[X] = 0$ and $\mathrm{E}[XY] = \mathrm{E}[X^3] = 0$ (since X^3 is an odd function), so $\mathrm{cov}(X, Y) = \mathrm{E}[X^3] - \mathrm{E}[X]\,\mathrm{E}[X^2] = 0$.

However, $\mathrm{cov}(X^2, Y) = \mathrm{E}[X^4] - \mathrm{E}[X^2]^2 = \mathrm{var}(Y) > 0$, since Y is not a constant.

(d) **Important Moments [7 points]**
Given that $\mathrm{E}[X^i] = i!\beta^i$ for $i = 0, 1, 2, \ldots$ what is the distribution of X? Provide complete justification for full credit.

$$
\begin{aligned}
\mathrm{E}[e^{tX}] &= \mathrm{E}\left[\sum_i \frac{(tX)^i}{i!}\right] \\
&= \sum_i \frac{t^i\,\mathrm{E}[X^i]}{i!} \\
&= \sum_i \frac{t^i \cdot i!\beta^i}{i!} \\
&= \sum_i (t\beta)^i \\
&= \frac{1}{1-\beta t}
\end{aligned}
$$

This corresponds to the mgf of Exponential$(1/\beta)$.

(e) **Exponential Sampling [7 points]**
Suppose you are able to generate values from $U \sim \text{Uniform}[0, 2]$. How can you simulate and sample values from an exponential distribution using U and $F_X(x)$, where $X \sim \text{Exponential}(\lambda)$? Provide complete justification for full credit.

First, scale any value drawn from U by $\frac{1}{2}$, so $U' = \frac{1}{2}U \sim \text{Uniform}[0, 1]$. This is because $\text{P}(U' \leq x) = \text{P}(\frac{1}{2}U \leq x) = \text{P}(U \leq 2x) = x$, showing $\frac{1}{2}U \sim \text{Uniform}[0, 1]$.

Then the inverse CDF of X is $F_X^{-1}(u) = -\frac{1}{\lambda}\ln(1-u)$. This means that if we draw some u_o from $U' \sim \text{Uniform}[0, 1]$ and compute $x = F_X^{-1}(u_o) = -\frac{1}{\lambda}\ln(1-u_o)$, it will be as if we drew x from an exponential distribution.

This can be shown with the following proof: Let $Y = F^{-1}(U')$. The CDF of Y is $G(y) = \text{P}(Y \leq y) = \text{P}(U' \leq F(y)) = F(y)$. The last equality follows from the CDF of a uniform random variable. Hence, $F^{-1}(U')$ has CDF F. This method is known as inverse transform sampling.

(f) **Matrix Sketching [7 points]**

In lab, we studied Gaussian Sketch and Count Sketch; let's consider the following sketching matrix: *Random Sign Sketch*. Suppose that the entries of $S \in \mathbb{R}^{d \times n}$ are generated such that

$$S_{ij} \sim_{\text{iid}} \begin{cases} +1 & \text{w.p. } 1/2 \\ -1 & \text{w.p. } 1/2 \end{cases}$$

Compute $\text{E}[(S^\top S)_{ij}]$ and $\text{var}[(S^\top S)_{ij}]$. Your answer may depend on d and/or n.

Denote the ith column of S by s_i. We observe the following:

- Notice that the entries of S are i.i.d. samples from a zero-mean distribution. It follows that for $i \neq j$, $\text{E}[(S^\top S)_{ij}] = \text{E}[s_i \cdot s_j] = \text{E}[s_i] \cdot \text{E}[s_j] = 0$ using linearity of expectation and $\text{var}[(S^\top S)_{ij}] = d\,\text{var}[s_{i1}s_{j1}] = d\,\text{E}[s_{11}^2]^2$.
- Similarly, when $i = j$ we can write $\text{E}[\|s_i\|^2] = d\,\text{E}[s_{i1}^2]$ and $\text{var}[\|s_i\|^2] = d\,\text{var}[s_{i1}^2]$.

From our observations above and noting $s_{ij}^2 = 1$ with probability 1, we can immediately deduce $\text{E}[(S^\top S)_{ij}] = d\mathbf{1}_{\{i=j\}}$ and $\text{var}[(S^\top S)_{ij}] = d\mathbf{1}_{\{i \neq j\}}$.

(g) **Polling for IQ [7 points]**
You and Clark discover a new planet called `621 SCEE`! You poll 1 million creatures on this planet for their IQ score (can be negative), but accidentally lose everyone's individual scores. However, you remember that the sample mean is $\frac{1}{2}$ and the sample variance is $\frac{3}{16}$.

With no further assumptions on the scores' possible distribution, choose from Markov's and Chebyshev's inequalities. What is the tightest upper bound on the number of scores ≥ 1?

Since this random variable may be negative, we can't apply Markov's Inequality. Therefore, we need to use Chebyshev's inequality (even though it ends up giving us a worse bound than an incorrect application of Markov).

$$\mathrm{P}(X \geq 1) = \mathrm{P}((X - 1/2)^2 \geq (1 - 1/2)^2) = \mathrm{P}((X - 1/2)^2 \geq 1/4) \leq \frac{\mathrm{var}(X)}{1/4} = 3/4.$$

Then $n \leq 10^6 \cdot \frac{3}{4} = 750000$.

Note: For this question, we awarded full credit to any solution that correctly applies an upper bound on the number of scores ≥ 1.

2 Minecraft!

Merrick is mining for diamonds in Minecraft! His pickaxe has rating $R \sim \text{Uniform}\{1,2,3,4,5\}$ and mining speed $S \sim \mathcal{N}(R, R/2)$ (note that it is possible to have negative mining speed). After a certain time T, Merrick stops mining. For any pickaxe speed S, $\mathrm{P}(T \le t) = 1 - e^{-t/S^2}$.

(a) What is $\text{var}(S)$?

(b) What is $\mathrm{E}[T]$?

(a) Using the law of total variance,

$$\begin{aligned}\mathbb{V}ar(S) &= \mathrm{E}[\mathbb{V}ar(S \mid R)] + \mathbb{V}ar[\mathrm{E}(S \mid R)] \\ &= \mathrm{E}[R/2] + \mathbb{V}ar(R) \\ &= 1.5 + \mathrm{E}[R^2] - (\mathrm{E}[R])^2 \\ &= 3.5.\end{aligned}$$

(b) We notice that $T \sim \text{Exponential}(\frac{1}{S^2})$. Using the tower rule, we get

$$\begin{aligned}\mathrm{E}[T] &= \mathrm{E}[\mathrm{E}[T \mid S]] = \mathrm{E}[S^2] \\ &= \mathrm{E}[\mathrm{E}[S^2 \mid R]] = \mathrm{E}[R^2 + R/2] = 12.5.\end{aligned}$$

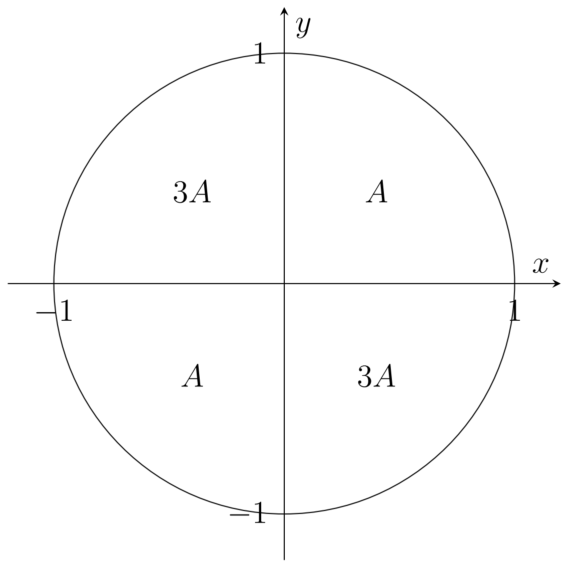
Alternatively, we could calculate $\mathrm{E}[S^2]$ as follows:

$$\begin{aligned}\mathrm{E}[S^2] &= \mathbb{V}ar[S] + (\mathrm{E}[S])^2 \\ &= 3.5 + 9 = 12.5.\end{aligned}$$

 Student ID: ______________________

3 Graphical Density

Let X and Y have joint probability density function as below.



Note that the circle of radius 1 is described by the equation $x^2 + y^2 = 1$.

(a) Find the value of A.

(b) Find the marginal PDFs $f_X(x)$ and $f_Y(y)$.

(c) Are X and Y independent?

(d) Find $\text{cov}(X, Y)$. Are X and Y correlated?

a. The joint PDF $f_{X,Y}$ should have total probability density 1, i.e.

$$A\left(\frac{\pi}{4}\right) + 3A\left(\frac{\pi}{4}\right) + A\left(\frac{\pi}{4}\right) + 3A\left(\frac{\pi}{4}\right) = 2\pi A = 1.$$

Thus $\boxed{A = \frac{1}{2\pi}}$.

b. Let us first consider the case when $x \geq 0$:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy = \int_{-\sqrt{1-x^2}}^{0} 3A\, dy + \int_{0}^{\sqrt{1-x^2}} A\, dy = \boxed{\frac{2}{\pi}\sqrt{1-x^2}}.$$

By reflection about the x-axis, we see that $f_X(x)$ is the same for $x < 0$. Now we observe that $X \overset{d}{=} Y$ by rotational symmetry, so $\boxed{f_Y(y) = \frac{2}{\pi}\sqrt{1-y^2}}$ as well.

For the subparts below, we note that the marginal PDFs are even, i.e. $f_X(x) = f_X(-x)$.

c. X and Y are independent iff $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$ at every point. Let us consider some (x,y) in the first quadrant. We see that

$$f_X(x) \cdot f_Y(y) = f_X(x) \cdot f_Y(-y).$$

However, $f_{X,Y}(x,y) = A$, while $f_{X,Y}(x,-y) = 3A \neq A$. The product of the marginals cannot be equal to two different values at once. Thus X and Y are $\boxed{\text{dependent}}$.

d. We recall that $\text{cov}(X,Y) = \text{E}(XY) - \text{E}(X) \cdot \text{E}(Y)$. By reflectional symmetry of $f_X(x)$ about the y-axis, $\text{E}(X) = 0$. More formally,

$$\int_{-1}^{0} x f_X(x)\, dx + \int_{0}^{1} x f_X(x)\, dx = \int_{1}^{0} (-x) f_X(-x)\, d(-x) + \int_{0}^{1} x f_X(x)\, dx = 0.$$

By rotational symmetry once more, $\text{E}(Y) = \text{E}(X) = 0$. Now let

$$I := \int_0^1 \int_0^{\sqrt{1-x^2}} xy f_{X,Y}(x,y)\, dy\, dx$$

be the expected value of XY in the first quadrant. By symmetry of the form $xy = (-x)(-y)$ and $-xy = (-x)y$, the overall expectation is $\text{E}(XY) = I - 3I + I - 3I = -4I$, where

$$I = A \int_0^1 \left(x \left[\frac{y^2}{2}\right]_0^{\sqrt{1-x^2}} \right) dx = A \int_0^1 \frac{x - x^3}{2}\, dx = \frac{1}{16\pi}.$$

Therefore $\boxed{\text{cov}(X,Y) = -\frac{1}{4\pi}}$, and $\text{cov}(X,Y) \neq 0$ means that X and Y are $\boxed{\text{correlated}}$.

4 Some Ticking Tocks

Axel has a collection of 101 faulty stopwatches, and starts them all of them at once. Suppose that the amount of time each stopwatch runs before stopping by itself, in minutes, is i.i.d. Exponential(2).

(a) What is the probability that each stopwatch lasts at least half a minute?

(b) Find the PDF of Y, the time it takes until all stopwatches stop.

(c) Find the expected value of the median time it takes for the stopwatches to stop. Feel free to express your answer in terms of $H_n = \frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{n}$, the nth harmonic number.

Let $X_i \sim$ Exponential(2) be how long stopwatch i lasts.

a) Given n exponentially distributed variables $X_i \sim \text{Exponential}(\lambda_i)$, we know that $\min_i X_i \sim \text{Exponential}(\sum_{i=1}^{n} \lambda_i)$. Thus, the minimum amount of time before any stopwatch stops is exponentially distributed with parameter 202. The probability this is greater than 0.5 is $e^{-202 \cdot 0.5} = e^{-101}$ using the CDF.

b) Let Y be the maximum time of all stopwatches. We shall first find the CDF of Y, which is $\mathrm{P}(Y \le y) = \prod_{i=1}^{101} \mathrm{P}(X_i \le y) = (1 - e^{-2y})^{101}$. Differentiating, we derive $f_Y(y) = 202e^{-2y}(1 - e^{-2y})^{100}$.

c) Define Z_k as the kth order statistic, i.e. the kth smallest out of the X_i's. For convenience, define $Z_0 = 0$. We wish to find $\mathrm{E}[Z_{51}]$.

$Z_{i+1} - Z_i$ represents the interval of time when exactly i timers have stopped, or equivalently when $101 - i$ timers are still running. Since all stopwatches stop at exponentially distributed times, the time from when the ith stopwatch stops to when the $i + 1$th stopwatch stops is determined by the next earliest stopwatch out of $101 - i$. Using the property mentioned in the first part, $Z_{i+1} - Z_i \sim \text{Exponential}(2(101 - i))$ and $\mathrm{E}[Z_{i+1} - Z_i] = \frac{1}{2(101-i)}$.

Now we can compute $\mathrm{E}[Z_{51}]$ as

$$\begin{aligned}
\mathrm{E}[Z_{51}] &= \sum_{i=0}^{50} \mathrm{E}[Z_{i+1} - Z_i] \\
&= \sum_{i=0}^{50} \frac{1}{2(101 - i)} \\
&= \frac{1}{2} \sum_{j=51}^{101} \frac{1}{j} \\
&= \frac{1}{2}(H_{101} - H_{50}).
\end{aligned}$$