UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

**Homework 03**
Spring 2023

1. **Matrix Sketching**

Matrix sketching is an important technique in randomized linear algebra for doing large computations efficiently. For example, to compute $\mathbf{A}^T \times \mathbf{B}$ for two large matrices $\mathbf{A}$ and $\mathbf{B}$, we can use a random sketch matrix $\mathbf{S}$ to compute a "sketch" $\mathbf{SA}$ of $\mathbf{A}$, and a sketch $\mathbf{SB}$ of $\mathbf{B}$. Such a sketching matrix has the property that

$$\mathbf{S}^T\mathbf{S} \approx \mathbf{I},$$

so that the approximate multiplication $(\mathbf{SA})^T(\mathbf{SB}) = \mathbf{A}^T\mathbf{S}^T\mathbf{SB}$ is close to $\mathbf{A}^T\mathbf{B}$.

In this problem, we will discuss two popular sketching schemes and understand how they help in approximate computation. Let $\hat{\mathbf{I}} = \mathbf{S}^T\mathbf{S}$, and let the dimension of the sketch matrix $\mathbf{S}$ be $d \times n$ (where typically $d \ll n$).

a. **Gaussian sketch**. Let the sketch matrix be

$$\mathbf{S} = \frac{1}{\sqrt{d}} \begin{bmatrix} S_{1,1} & \cdots & S_{1,n} \\ \vdots & \ddots & \vdots \\ S_{d,1} & \cdots & S_{d,n} \end{bmatrix},$$

where the $S_{i,j}$ are chosen i.i.d. from $\mathcal{N}(0,1)$ for all $i \in [1,d]$ and $j \in [1,n]$. Show that the elementwise mean and variance of the matrix $\hat{\mathbf{I}}$, as functions of $d$, are

$$\mathbb{E}(\hat{I}_{i,j}) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

$$\text{var}(\hat{I}_{i,j}) = \begin{cases} \frac{2}{d} & \text{if } i = j \\ \frac{1}{d} & \text{otherwise.} \end{cases}$$

You can use without proof the fact that $\mathbb{E}(Z^4) = 3$ for $Z \sim \mathcal{N}(0,1)$.

b. **Count sketch**. For each column $j \in [1,n]$ of $\mathbf{S}$, choose a row $i$ uniformly randomly from $[1,d]$. Set

$$S_{i,j} = \begin{cases} 1 & \text{with probability } \frac{1}{2} \\ -1 & \text{with probability } \frac{1}{2}, \end{cases}$$

and assign $S_{k,j} = 0$ for all $k \neq i$. An example of a $3 \times 8$ count sketch matrix is

$$\begin{bmatrix} 0 & -1 & 1 & 0 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix}.$$

Show that the elementwise mean and variance of the matrix $\hat{\mathbf{I}}$ are

$$\mathbb{E}(\hat{I}_{i,j}) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

$$\mathrm{var}(\hat{I}_{i,j}) = \begin{cases} 0 & \text{if } i = j \\ \frac{1}{d} & \text{otherwise.} \end{cases}$$

Note that for sufficiently large $d$, the matrix $\hat{\mathbf{I}}$ is close to the identity matrix in both cases. We use this fact in the lab to do an approximate matrix multiplication.

2. **Properties of the CDF**

   The **cumulative distribution function**, or cdf, of a random variable $X$ is the function $F(x) = \mathbb{P}(X \leq x)$.

   a. Using the properties of a probability measure, show that $F$ is nondecreasing: if $x \leq y$, then $F(x) \leq F(y)$.
   b. Show that $F$ is right-continuous: if $x_1, x_2, \ldots$ is a decreasing sequence converging to $y$, then $F(x_1), F(x_2), \ldots$ converges to $F(y)$.
   c. Show that $F$ is *normalized*: $\lim_{x \to -\infty} F(x) = 0$, and $\lim_{x \to \infty} F(x) = 1$.

   *Hint*: For parts b and c, it may help to revisit question 1b of discussion 01.

3. **Change of Variables**

   Let $X$ be a continuous random variable with cdf $F_X$ and pdf $f_X > 0$ everywhere, and let $Y = g(X)$, where $g$ is a differentiable function.

   a. Suppose that $g$ is also invertible. Find the pdf of $Y$, $f_Y$, in terms of $g$ and $f_X$.

   b. Let $U \sim \text{Uniform}([0, 1])$. Using the conclusion from part a, show that $F_X^{-1}(U)$ has the same distribution as $X$. (This allows us to generate a given random variable given only a uniform random number generator.)

   c. Now suppose that $g(x) = x^2$. Find the pdf of $Y$ in terms of the pdf of $X$. Also find the pdf of $Y$ when $X$ is a standard normal random variable in particular.

      (Note that this $g$ is not invertible, unlike in part a.)

4. **Gaussian Confidence Interval**

A $C\%$ **confidence interval** for a parameter $\theta$ is the interval containing $\theta$ of smallest length, such that $\theta$ falls in the interval with probability at least $C\%$.

Suppose that a given population has Gaussian distribution with unknown mean $\mu$ and variance $\sigma^2$. We draw $n$ independent samples; let the average of the samples be $\bar{\mu}$.

a. Find a 95% confidence interval for $\mu$.

b. Suppose $\sigma^2 = 1$. How many independent samples at minimum do we need to construct a 99% confidence interval for $\mu$ with length at most 1?

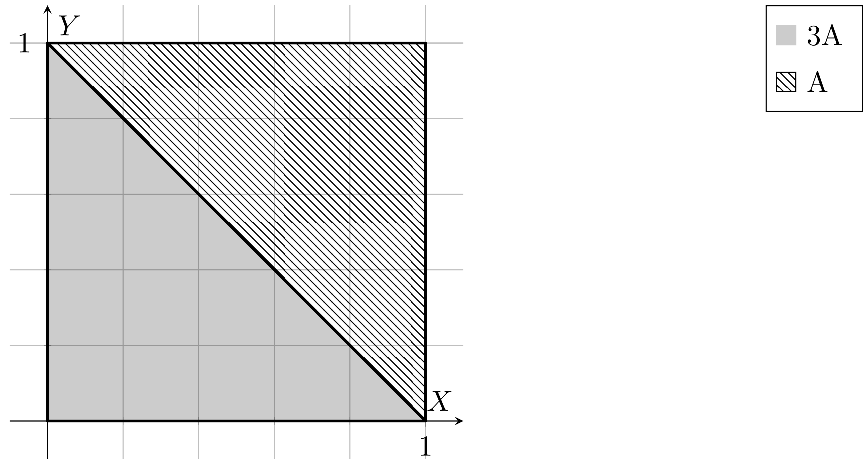
5. **Binomial with Random Parameter**

   Let $U \sim \text{Uniform}([0,1])$, and suppose that $X$ has distribution $\text{Binomial}(n,p)$ given that $U = p$. Find $\mathbb{E}(U^2 X)$ and $\mathbb{E}(U^2 X^2)$.

   *Hint*: Rather than working directly with the definition of expectation, consider the properties of conditional expectation.

6. **Graphical Density**

The following figure depicts the joint density $f_{X,Y}$ of $X$ and $Y$.




a. Are $X$ and $Y$ independent? Remember to justify your answer.

b. What is the value of $A$?

c. Compute $f_X(x)$.

d. Compute $\mathbb{E}(Y \mid X = x)$. You may leave your answer as a fraction of terms containing $x$, but you may not have an integral.

e. What is $\mathbb{E}(X - Y \mid X + Y)$?

7. **Joint Density for Exponential Distribution**

a. If $X \sim \text{Exponential}(\lambda)$ and $Y \sim \text{Exponential}(\mu)$ are independent, compute $\mathbb{P}(X < Y)$.

b. If $X_1, \ldots, X_n$ are independent and Exponentially distributed with parameters $\lambda_1, \ldots, \lambda_n$, show that $\min_{1\le k\le n} X_k \sim \text{Exponential}(\sum_{j=1}^n \lambda_j)$.

c. Deduce that

$$\mathbb{P}\left(X_i = \min_{1\le k\le n} X_k\right) = \frac{\lambda_i}{\sum_{j=1}^n \lambda_j}.$$