UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Homework 04
Fall 2023

1. **Basic Properties of Jointly Gaussian Random Variables**

   Let $(X_1, \ldots, X_n)$ be a collection of jointly Gaussian random variables with mean vector μ and covariance matrix Σ. Their joint density is given by, for $x \in \mathbb{R}^n$,

   $$f(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\left\{-\frac{1}{2}(x-\mu)^{\mathsf{T}}\Sigma^{-1}(x-\mu)\right\}.$$

   a. Show that $X_1, \ldots, X_n$ are independent if and only if they are pairwise uncorrelated.

   b. Show that any linear combination of $X_1, \ldots, X_n$ will also be a Gaussian random variable. *Hint*: Consider using moment-generating functions.

2. **Gaussian Sine**

   Let X, Y, Z be jointly Gaussian random variables with covariance matrix

$$\begin{bmatrix} 4 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 4 \end{bmatrix}$$

   and mean vector $[0, 2, 0]$. Compute $\mathbb{E}[(\sin X)Y(\sin Z)]$. *Hint:* Condition on (X, Z).

3. **Lognormal Distribution and the Moment Problem**

(Optional) This question seeks to answer the following question: if two distributions have the same moments of all orders, are they necessarily the same? An equivalent way to phrase the problem is: if the moments exist, do they completely determine the distribution?

a. Suppose that Z is a standard Gaussian and let $X = e^Z$. Calculate the density of X. (This is known as the **lognormal** distribution.)

b. Let $f_X(x)$ denote the density of the lognormal density. Define

$$f_a(x) = f_X(x)(1 + a\sin(2\pi \log x)), \qquad x > 0, \quad -1 \le a \le 1.$$

Argue that $f_a(x)$ is a valid density function and show that $f_X(x)$ and $f_a(x)$ have the same moments of all orders by showing that

$$\int_0^\infty x^k f_X(x) \sin(2\pi \log x)\, \mathrm{d}x = 0, \qquad k \in \mathbb{N}.$$

c. Explicitly calculate the moments of the lognormal distribution.

d. Now, let Y_b $(b > 0)$ be a discrete random variable with distribution

$$\Pr(Y_b = be^n) = cb^{-n}e^{-n^2/2}, \qquad n \in \mathbb{Z},$$

where c is chosen to normalize the distribution:

$$\sum_{n=-\infty}^{\infty} cb^{-n}e^{-n^2/2} = 1.$$

Show that Y_b has the same moments as X. This provides a discrete counterexample to the moment problem.

4. **Revisiting Proofs Using Transforms**

a. Let $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$ be independent. Calculate the MGF of $X + Y$, and use this to show that $X + Y \sim \text{Poisson}(\lambda + \mu)$.

b. Calculate the MGF of $X \sim \text{Exponential}(\lambda)$, and use this to find all of the moments of X.

c. Repeat the above part, but for $X \sim \mathcal{N}(0, 1)$.

5. **Coupon Collector Bounds**

Recall the coupon collector's problem, in which there are n different types of coupons. Every box contains a single coupon, and we let the random variable X be the number of boxes bought until one of every type of coupon is obtained. The expected value of X is nH_n, where $H_n := \sum_{i=1}^{n} \frac{1}{i}$ is the *harmonic number* of order n, which satisfies the inequality

$$\ln n \leq H_n \leq \ln n + 1.$$

a. Use Markov's inequality in order to show that

$$\mathbb{P}(X > 2nH_n) \leq \frac{1}{2}.$$

b. Use Chebyshev's inequality in order to show that

$$\mathbb{P}(X > 2nH_n) \leq \frac{\pi^2}{6(\ln n)^2}.$$

Note: You can use Euler's solution to the Basel problem, the identity $\sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}$.

c. Define appropriate events and use the union bound in order to show that

$$\mathbb{P}(X > 2nH_n) \leq \frac{1}{n}.$$

Note: $a_n = (1 - \frac{1}{n})^n$ is a strictly increasing sequence with limit e^{-1}.

6. **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

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

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

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 3×8 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.