UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Homework 5
Spring 2024

1. **Polar coordinates**

   Let X and Y be independent standard normal random variables. The pair (X, Y) can be described in polar coordinates in terms of random variables $R \geq 0$ and $\Theta \in [0, 2\pi]$, so that

   $$X = R \cos \Theta, \quad Y = R \sin \Theta.$$

   (a) Show that the joint CDF of R, Θ is given by $\frac{1}{2\pi} \int_0^\theta \int_0^r e^{-r'^2/2} r' dr' d\theta'$.

   (b) Differentiate the joint CDF to find the pdf of R, Θ. Find immediately the marginal distribution of Θ.

   (c) Show that R^2 has an exponential distribution with parameter $1/2$.

2. **Gaussian practice**

   Let (η, ζ) be joint Gaussian random variables, such that $\mathbb{E}[\xi] = \mathbb{E}[\eta] = 0$, $\text{Var}(\xi) = \text{Var}(\eta) = 1$, and $\text{Cov}(\eta, \xi) = \rho$.

   (a) Are the random variables $\eta - \xi$ and $\eta + \xi$ independent?

   (b) Show that $\mathbb{E}[\min(\xi, \eta)] = -\sqrt{\frac{1-\rho}{\pi}}$.

   Hint: $\int x e^{-\alpha x^2} dx = -\frac{e^{-\alpha x^2}}{2\alpha} + c$.

3. **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 3×8 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

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

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

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.

4. **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}.

5. **Subgaussian Random Variables**

We say that a random variable Y is *subgaussian* with parameter σ^2 (written $Y \in SG(\sigma^2)$) if

$$\mathbb{E}\, e^{\lambda Y} \leq \exp\left(\frac{1}{2}\lambda^2\sigma^2\right), \quad \text{for all } \lambda \in \mathbb{R}.$$

Subgaussian random variables consist a whole class of random variables which have a similar tail with Gaussian random variables, and thus they share similar concentration properties. We are going to examine some properties of subgaussian random variables in this sequel of problems.

Prove the following:

a. If $Y \sim N(0, \sigma^2)$, then $Y \in SG(\sigma^2)$.

b. If $\Pr(Y = 1) = \Pr(Y = -1) = 1/2$, then $Y \in SG(1)$.
Hint: use Taylor expansion for the exponentials and observe what is happening to the odd terms.

c. If $Y \in SG(\sigma^2)$, then $\mathbb{E}[Y] = 0$.
Hint: what is the power series for $\mathbb{E}\, e^{\lambda Y}$? How can you extract $\mathbb{E}\, Y$ from the power series?

d. If $Y \in SG(\sigma^2)$, then $\mathrm{var}(Y) \leq \sigma^2$.

e. If $Y_1, \ldots, Y_n$ are independent random variables with $Y_i \in SG(\sigma_i^2)$ for $i = 1, \ldots, n$, and $a_1, \ldots, a_n$ are scalars then

$$a_1 Y_1 + \ldots + a_n Y_n \in SG(a_1^2\sigma_1^2 + \ldots + a_n^2\sigma_n^2).$$

f. If $Y \in SG(\sigma^2)$, then

$$\Pr(Y \geq t) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right) \quad \text{and} \quad \Pr(-Y \geq t) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right)$$

for every $t \geq 0$.

6. **The Weak Law of Large Numbers**

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with common mean μ and MGF M_X. We assume that $M_X(s)$ is finite when $s \in (-d, d)$ for some $d > 0$. Let

$$\bar{X}_n := \frac{X_1 + \ldots + X_n}{n}.$$

a. Show that the transform (or MGF) associated with $\bar{X}_n$ satisfies

$$M_{\bar{X}_n}(s) = M_X(s/n)^n.$$

b. Suppose that the transform $M_X(s)$ has a first-order Taylor series expansion around $s = 0$ of the form

$$M_X(s) = a + bs + o(s),$$

where $o(s)$ is a function that satisfies $\lim_{s \to 0} o(s)/s = 0$. Find a and b in terms of μ.

c. Show that for all $s \in (-d, d)$,

$$\lim_{n \to \infty} M_{\bar{X}_n}(s) = e^{\mu s}.$$

Hint: If $(a_n)_{n \in \mathbb{N}}$ is a sequence of real numbers converging to a, then $\lim_{n \to \infty} (1 + \frac{a_n}{n})^n = e^a$.

d. Deduce that $\bar{X}_n \xrightarrow{\mathsf{d}} \mu$. Note that the pointwise convergence of MGFs is equivalent to convergence in distribution.