Homework 4
High-Dimensional Probability for Data Science, Fall 2023

Hints are in the back of the homework set.

As before, $C, C_1, C_2, \dots$ and $c, c_1, c_2, \dots$ denote *positive absolute constants of your choice*. See more explanation in Homework 3.

Laplace distribution is often skipped from standard probability courses. But it has some properties that the normal distribution lacks, such as property (c) below. Laplace distribution it a distribution of choice in some areas of data science, particularly in applications for privacy ("Laplace mechanism").

1. Laplace distribution

A random variable X has Laplacian distribution with mean μ and unit scale, denoted $X \sim \mathrm{Lap}(\mu)$, if the probability density function of X is

$$f_\mu(x) = \frac{1}{2} e^{-|x-\mu|}, \quad x \in \mathbb{R}.$$

(a) Check that the mean of X is indeed μ.

(b) Check that variance of X equals 2.

(c) Show that if $X \sim \mathrm{Lap}(\mu)$ and $Y \sim \mathrm{Lap}(\nu)$, then

$$\frac{\mathbb{P}\{X \in B\}}{\mathbb{P}\{Y \in B\}} \le e^{|\mu-\nu|}.$$

(d) Let us say that a random vector $X = (X_1, \dots, X_N)$ has Laplacian distribution $\mathrm{Lap}(\mu)^n$ if all X_i are independent $\mathrm{Lap}(\mu)$ random variables. Check that if $X \sim \mathrm{Lap}(\mu)^N$ and $Y \sim \mathrm{Lap}(\nu)^N$, then

$$\frac{\mathbb{P}\{X \in B\}}{\mathbb{P}\{Y \in B\}} \le e^{|\mu-\nu|N}.$$

In Lecture 5, we studied mean estimation. We introduced the "median-of-means estimator" (MoM), which features tighter confidence intervals than the usual sample mean. However, the guarantee for MoM proved in Lecture 5 (p.4) only works for $t = O(\sqrt{N})$. Here we prove that this restriction is optimal. As we will show now, there does not exist *any* estimator of the mean that works for all t.

2. There is no fully subgaussian estimator of the mean

Disprove the following claim:

> "There exists an absolute constant $c > 0$ such that the following holds. For any $N \in \mathbb{N}$ and every $t \geq 0$, one can find a function $\widehat{\mu} : \mathbb{R}^N \to \mathbb{R}$ that satisfies the inequality
> $$\mathbb{P}\left\{ |\widehat{\mu}(X) - \mu| \geq \frac{t\sigma}{\sqrt{N}} \right\} \leq 2\exp(-ct^2)$$
> whenever the coordinates X_i of the random vector $X = (X_1, \dots, X_N)$ are independent and identically distributed random variables with mean μ and variance σ^2."

We will disprove the claim by working out a version of the *Le Cam's method*, a popular tool for obtaining impossibility results in statistics:

Argue by contradiction. Assuming the claim holds, apply it for two distributions: $X = (X_1, \dots, X_N) \sim \mathrm{Lap}(\mu)^N$ and $Y = (Y_1, \dots, Y_N) \sim \mathrm{Lap}(0)^N$, where μ is a large absolute constant and t is chosen so that $t\sigma/\sqrt{N} = \mu/2$. Next, replace Y with X using Exercise 1, and arrive at a contradiction.

In the next three problems we prove the general Hoeffding's inequality, which we stated in Lecture 5(p.2); we restate it in Problem 5 below. It is a remarkably sharp result with optimal absolute constants. Achieving such precision makes our task a little technical, but it is divided it into small doable steps below.

3. Two-point Hoeffding's lemma

Let X be a mean zero random variable that takes values only two values a and b. Following steps (a)–(d) below, prove that

$$\mathbb{E}\, e^{\lambda X} \leq \exp\left(\frac{\lambda^2(b-a)^2}{8}\right) \quad \text{for any } \lambda \in \mathbb{R}. \tag{1}$$

(a) Assume first that $b - a = 1$, and keep this assumption for steps (a)–(c). Find the probabilities that X takes value a and value b. Then compute the so-called *cumulant* $K(\lambda) := \log \mathbb{E}\, e^{\lambda X}$ for all $\lambda \in \mathbb{R}$.

(b) Check that $K(0) = 0$, $K'(0) = 0$, and $K''(\lambda) \leq 1/4$ for any $\lambda \in \mathbb{R}$.

(c) Conclude that $K(\lambda) \leq \lambda^2/8$ for any $\lambda \in \mathbb{R}$.

(d) Using rescaling, show how to drop the assumption $b - a = 1$.

(e) Demonstrate that the bound (1) is optimal. To this end, show that the constant 8 can not be replaced in general by any strictly smaller constant.

4. General Hoeffding's lemma

(a) Let Z and X be mean zero random variables. Suppose that Z takes values in some interval $[a, b]$, while X takes values in the two-point set $\{a, b\}$. Show that

$$\mathbb{E}\, e^{\lambda Z} \le \mathbb{E}\, e^{\lambda X} \quad \text{for all } \lambda \in \mathbb{R}.$$

(b) Using part (a), quickly extend the two-point Hoeffding's lemma from Exercise 3. Namely, note that the MGF of any mean zero random variable Z that takes values in an interval $[a, b]$ satisfies

$$\mathbb{E}\, e^{\lambda Z} \le \exp\left(\frac{\lambda^2 (b-a)^2}{8}\right) \quad \text{for all } \lambda \in \mathbb{R}.$$

5. General Hoeffding's inequality

Using the result of Exercise 4, prove the following general result known as Hoeffding's inequality. Let $X_1, \ldots, X_N$ be independent random variables with each X_i taking values in $[a_i, b_i]$. Then the sum $S_N = X_1 + \cdots + X_N$ satisfies

$$\mathbb{P}\left\{|S_N - \mathbb{E}\, S_N| \ge t\right\} \le 2\exp\left(-\frac{2t^2}{\sum_{i=1}^N (b_i - a_i)^2}\right) \quad \text{for all } t \ge 0.$$

In Lecture 6 we proved a Discrepancy theorem, which guarantees with high probability, a random sample of points in the square has low discrepancy. A given set of points is said to have "low discrepancy" if the fraction of points that fall into any given rectangle is approximately equal to the area of the rectangle. One of us asked if rectangles are special. Can the same result be proved for circles? Let us prove that it can!

6. Discrepancy for circles

Extend the Discrepancy theorem (Lecture 6) for circles. Namely, prove that a set of N independent points uniformly drawn from the square $S = [0, 1]^2$ satisfies the following with probability at least 0.99:

$$\sup_Q \left| \frac{N_Q}{N} - \operatorname{Area}(Q \cap S) \right| \le C\sqrt{\frac{\log N}{N}}.$$

Here the supremum is over all circles Q whose centers are in S, the number N_Q denotes the number of the points in Q, and C is an absolute constant.

(You do not need to write down the entire proof. It is sufficient to explain how to modify the argument in Lecture 6.)

TURN OVER FOR HINTS

Hints for Problem 1.

(c) Using triangle inequality, check that

$$\frac{f_\mu(x)}{f_\nu(x)} \leq e^{|\mu-\nu|} \quad \text{for all } x \in \mathbb{R}.$$

Express $\mathbb{P}\{X \in B\}$ in terms of the density $f_\mu(x)$, and similarly for $\mathbb{P}\{Y \in B\}$.

(d) Recall how the density of the joint distribution of independent random variables X_i is related to the densities of X_i.

Hints for Problem 2.

Use the claim to get bounds for $p_1 := \mathbb{P}\left\{|\widehat{\mu}(X) - \mu| \geq \mu/2\right\}$ and for $\mathbb{P}\left\{|\widehat{\mu}(Y)| \geq \mu/2\right\}$. Next use Exercise 1 to conclude a bound for $p_2 := \mathbb{P}\left\{|\widehat{\mu}(X)| \geq \mu/2\right\}$. Choosing μ as a sufficiently large constant, you should be able to obtain $p_1 < 1/2$ and $p_2 < 1/2$. This implies that with some positive probability, both $|\widehat{\mu}(X) - \mu| < \mu/2$ and $|\widehat{\mu}(X)| < \mu/2$ hold together (why?) Arrive at a contradiction using triangle inequality.

Hints for Problem 3.

(a) The two probabilities satisfy two linear equations: their sum equals 1, and the expected value of X equals 0. Write down and solve this equations. If you do everything correctly, you can compute the cumulant and the result should be $K(\lambda) = \lambda a + \log\left(b - ae^\lambda\right)$.

(b) Differentiating twice should yield

$$K''(\lambda) = \frac{-ae^\lambda \cdot b}{(-ae^\lambda + b)^2}.$$

Now use the AM-GM inequality $\sqrt{xy} \leq \frac{x+y}{2}$ for $x = -ae^\lambda$ and $y = b$.

(c) Write down a linear approximation of $K(\lambda)$ using Taylor's theorem with remainder in Lagrange form. Use the information from part (a).

(d) Use part (c) for $\lambda_0 = \lambda(b-a)$ and $X_0 = X/(b-a)$.

(e) Consider the symmetric Bernoulli random variable X, i.e. one which takes values -1 and 1 with probabilities 1/2 each. Expand both sides into Taylor series around $\lambda = 0$ with the remainder in the Peano form, and take the limit as $\lambda \to 0$.

Hints for Problem 4.

(a) By rescaling, we can assume without loss of generality that $b - a = 1$. Use the convexity of the exponential function $f(z) = e^{\lambda z}$ to conclude that

$$e^{\lambda z} \leq (b-z)e^{\lambda a} + (z-a)e^{\lambda b} \quad \text{for all } z \in [a,b].$$

(Plot a graph of the exponential function and the right hand side; what do you see?) Substitute $z = Z$, take expectation on both sides use that $\mathbb{E}\, Z = 0$, and get $\mathbb{E}\, e^{\lambda Z} \leq$

$be^{\lambda a} - ae^{\lambda b}$. The right hand side here should be the same as the MGF of X, which you computed in Problem 3(a).

Hints for Problem 5.

Follow the MGF argument developed in Lecture 5. Use Hoeffding's lemma (Exercise 4) to bound the MGF of each X_i.

Hints for Problem 6.

Follow the ε-net argument developed in Lecture 6. In the discretization step, choose a convenient "net of circles" – a finite family $\mathcal{Q}$ of circles such that (a) the cardinality of not too large, and (b) every circle Q can be nicely approximated by some circle from the family $\mathcal{Q}$. To this end, discretize the position of the center using a grid, and discretize the value of the radius similarly.