Problem Set #1 – Introduction and Concentration inequalities

Notation In this problem set and throughout the course we will use the following notation:

- $a_n \sim b_n$ means $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$.
- $a_n \lesssim b_n$ means that there exists an *absolute* constant $C > 0$ such that $a_n \leq b_n$. By "absolute" it means that this constant does not depend on the problem quantities defining a_n, b_n (e.g., the random variable). We can think of it as a numerical constant, which for simplicity of analysis is not computed exactly.
- $a_n = \Theta(b_n)$ means $\lim_{n\to\infty} \frac{|a_n|}{b_n} < \infty$ *and* $\lim_{n\to\infty} \frac{|a_n|}{b_n} > 0$.

Preliminaries Let X, Y be random variables. Consult [V, Ch. 1] and verify that you are comfortable with the following:

- Moment generating function $\mathbb{E}[e^{sX}]$.
- L_p spaces, their norm $\|X\|_{L_p} := (\mathbb{E}|X|^p)^{1/p}$, for $p \in (0, \infty)$ (*what makes the last sentence inaccurate?*)
  - (specifically) L_2 space and its inner product $\langle X, Y \rangle_{L_2} = \mathbb{E}[XY]$.
- Basic inequalities: Jensen, Minkowski, Cauchy-Schwarz, *Hölder* .
- Basic tail inequalities: Markov, Chebyshev, Chernoff.
- The central limit theorem.

1. (*Gaussian integral*)

   (a) Prove that for $a > 0, b \in \mathbb{R}$:

$$\int e^{-at^2-bt} \cdot \mathrm{d}t = \sqrt{\frac{\pi}{a}} \cdot e^{b^2/(4a)}.$$

(b) Let $S \sim N(0,3/4)$. Compute $\mathbb{E}[e^{S^2/2}]$.

2. *(From tails to expectation)* Let $\{X_i\}_{i=1}^n$ be distributed i.i.d., $X_i \in \mathbb{R}^d$ such that $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[X_i X_i^T] = I_d$ (such X_i's are called *isotropic*). Let $\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^n X_i X_i^T$ be the empirical covariance matrix. We have stated that

$$\mathbb{P}\left[\lambda_{\max}(\hat{\Sigma}) > 1 + C\left(\sqrt{\frac{d}{n}} + t + \sqrt{t}\right)\right] \leq e^{-nt}. \tag{1}$$

   (a) Show that with probability larger than $1-\delta$

$$\lambda_{\max}(\hat{\Sigma}) \leq 1 + C_1\left(\sqrt{\frac{d+\log\left(\frac{1}{\delta}\right)}{n}} + \frac{\log\left(\frac{1}{\delta}\right)}{n}\right).$$

   What is C_1?

   (b) Recall the *integral identity*: $\mathbb{E}[X] = \int_0^\infty \mathbb{P}[X > s]\mathrm{d}s$ if $X > 0$ a.s. . Show that

$$\mathbb{E}\left[\lambda_{\max}(\hat{\Sigma}) - 1\right] \leq C_2 \cdot \sqrt{\frac{d}{n}}.$$

3. *(Probability of exact sample statistics)* Your highly skeptic friend doesn't believe the coin you are holding is fair. To prove him wrong you flip the coin n times, with the hope of getting *exactly* $n/2$ heads and $n/2$ tails. What is the probability that you will be successful? Show that probably *not*:

$$\mathbb{P}\left[\sum_{i=1}^n X_i = n/2\right] = \Theta\left(\frac{1}{\sqrt{n}}\right)$$

for $X_i \sim \text{Bernoulli}(\frac{1}{2})$, i.i.d..
Hint: Stirling's approximation $n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$ *might be useful here.*

4. *(Orlicz spaces)* A function $\psi : [0,\infty) \to [0,\infty)$ is called an Orlicz function if it is convex, increasing, and satisfies

$$\psi(0) = 0, \ \lim_{x\to\infty} \psi(x) = \infty.$$

For a random variable X, define

$$\|X\|_\psi := \inf\{t > 0 : \mathbb{E}\psi(|X|/t) \leq 1\} \tag{2}$$

   (a) Prove that $\|\cdot\|_\psi$ defines a *norm* on the *Orlicz space* $L_\psi := \{X : \|X\|_\psi \leq \infty\}$.
   Hint: It holds that $\frac{|X+Y|}{\|X+Y\|_\psi} \leq \alpha\frac{|X|}{\|X\|_\psi} + (1-\alpha)\frac{|Y|}{\|Y\|_\psi}$. *What is* α*?*

   (b) What is L_ψ for:

i. $\psi = x^p$, $p \geq 1$ (what about $p \in [0,1]$?).
ii. $\psi_2 = e^{x^2} - 1$.
iii. $\psi_1 = e^{|x|} - 1$.

5. *(Centering)* We have defined subGaussian and subExponential distributions only for zero mean random variables. Is the zero-mean assumption essential? Let X be a random variable, possibly with $\mathbb{E}X \neq 0$

   (a) Prove that $\|X - \mathbb{E}X\|_{L_2} \leq \|X\|_{L_2}$ (you are probably well-familiar with this property).

   (b) Prove that $\|X - \mathbb{E}X\|_{\psi} \leq C\|X\|_{\psi}$ for some absolute constant C, for both $\psi = \psi_2$ and $\psi = \psi_1$.

   (c) Show that $C > 1$ must hold for ψ_2 (unlike the L_2 case).
   Hint: For a counterexample, it seems easiest to consider X with support $\{\pm 1\}$, and numerically check the condition (2).

6. *(Boosting randomized algorithms)* Assume that a random algorithm with a binary (or "yes"/"no") output returns a correct answer with probability $\frac{1}{2} + \delta$. Note that this is only slightly better than a completely random guess. Being unsatisfied with this poor performance, you decide to run the algorithm n times. Let $A_i = \mathbb{1}\{\text{The algorithm outputs a correct answer in the } i\text{-th attempt}\}$, i.e. $A_i \sim \text{Bernoulli}(\frac{1}{2} + \delta)$. Note that specifically, they are *independent*. We decide on the final answer according to a simple *majority* vote on the $\{A_i\}$. For any fixed $\varepsilon \in (0,1)$, prove that if $n \geq \frac{1}{2\delta^2} \log\left(\frac{1}{\varepsilon}\right)$ then the answer is correct with probability larger than $1 - \varepsilon$.

7. *(Median-of-means estimator for robust estimation)* Suppose you are required to find an estimator $\hat{\mu}$ of the mean $\mu = \mathbb{E}[X]$ such that $\hat{\mu}$ is ε-accurate, i.e., $\hat{\mu} \in (\mu - \varepsilon, \mu + \varepsilon)$ with high probability. For this task, you are given i.i.d. samples $\{X_i\}$ of X. Assume that $\text{Var}X^2 < \infty$.

   (a) Prove that there exists $c_1 > 0$ such that the sample mean $\tilde{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is ε-accurate with probability larger than $\frac{3}{4}$ if $n > \frac{c_1\sigma^2}{\varepsilon^2}$ where $\sigma^2 = \text{Var}X$.

   (b) Consider the following median-of-means estimator: Compute k estimators $\tilde{\mu}_j = \frac{1}{n}\sum_{i=1}^{n} X_{i,j}$, $j \in [k]$, where each is based on different n i.i.d. samples (the total nk samples are all i.i.d.). Let $\mu^* = \mathbb{M}(\tilde{\mu}_1, \tilde{\mu}_2, \ldots, \tilde{\mu}_k)$ be the *median* of these k estimators. Prove that there exists $c_2 > 0$ such that μ^* is ε-accurate with probability larger than $1 - \delta$ for

   $$nk = c_2 \log(\frac{1}{\delta})\frac{\sigma^2}{\varepsilon^2}.$$

   Hint: The $\tilde{\mu}_i$ are random algorithms which are "correct" with probability $\frac{3}{4}$. We are *boosting* these *weak estimators* to obtain a more reliable estimator. The price to be paid for this high probability is only logarithmic in $1/\delta$.