Mathematics of Data Science, Fall 2025 - Homework 1
Due at 11:49PM on Friday Sep/12 (Gradescope)

Your submitted solutions to assignments should be your own work. While discussing homework problems with peers is permitted, the final work and implementation of any discussed ideas must be executed solely by you. Acknowledge any source you consult. Do not use any type of Large Language Model, e.g., ChatGPT, to blindly answer this assignment. If you do, you might be subject to an oral examination to defend your assignment.

Problem 1 - Useful lemmata

Prove the following facts.

(a) Suppose that X is a zero-mean random variable satisfying that there exists $\sigma > 0$ such that $\mathbb{E}\exp(\lambda X) \leq \exp\left(\frac{\lambda^2\sigma^2}{2}\right)$ for all $\lambda \in \mathbf{R}$. Then, $\mathbb{E}X = 0$ and $\mathrm{Var}(X) \leq \sigma^2$.

(b) Let Z be a random vector with expectation $\mu \in \mathbf{R}^n$ and $\mathbb{E}\|Z\|^2 < \infty$. Then,

$$\mathbb{E}\|Z - \mu\|^2 = \inf_{\omega \in \mathbf{R}^n} \mathbb{E}\|Z - \omega\|^2.$$

(c) Further suppose that Z' has the same distribution as Z and is independent of it. Then,

$$\mathbb{E}\|Z - \mu\|^2 = \frac{1}{2}\mathbb{E}\|Z - Z'\|^2.$$

(d) Suppose that X is a random variable. Then, for any increasing and differentiable function $\varphi\colon \mathbf{R} \to \mathbf{R}$ with $\mathbb{E}\varphi(|X|) < \infty$, we have that

$$\mathbb{E}\varphi(|X|) = \int_0^\infty \varphi'(t)\mathbb{P}(|X| \geq t)dt.$$

Problem 2 - Boosting randomized algorithms

Imagine we have an algorithm for solving some decision problem, like checking if p is prime. Suppose the algorithm makes decisions randomly, giving the correct result with probability $\frac{1}{2} + \varepsilon$ for some $\varepsilon > 0$, only slightly better than pure guessing. To improve the performance, run the algorithm N times and take the majority vote. Show that, for any $\delta \in (0, 1)$, the answer is correct with probability at least $1 - \delta$, as long as

$$N \geq \frac{1}{2\varepsilon^2} \ln\left(\frac{1}{\delta}\right).$$

Problem 3 - Maxima of finitely many sub-Gaussians

Let $X_1, \ldots, X_N$ be mean-zero, σ^2-sub-Gaussian random variables (not necessarily independent) with $N \geq 2$.

(a) Prove that

$$\mathbb{E}\left[\max_{i\le N} X_i\right] \le \sqrt{2\sigma^2 \ln N} \quad \text{and} \quad \mathbb{E}\left[\max_{i\le N} |X_i|\right] \le \sqrt{2\sigma^2 \ln(2N)}.$$

(b) Show that

$$\mathbb{P}\left[\max_{i\le n}\{X_i\} \ge (1+\varepsilon)\,\sigma\sqrt{2\log n}\right] \xrightarrow[n\to\infty]{0} \quad \text{for all } \varepsilon > 0.$$

Hint: use the union bound

$$\mathbb{P}[X \vee Y \ge t] = \mathbb{P}[X \ge t \text{ or } Y \ge t] \le \mathbb{P}[X \ge t] + \Pr[Y \ge t].$$

Problem 4 - Chernoff against moments

Show that for $t \ge 0$

$$\mathbb{P}\big[X - \mathbb{E}X \ge t\big] \le \inf_{p\ge 0} \frac{\mathbb{E}\big[(X-\mathbb{E}X)_+^p\big]}{t^p} \le \inf_{\lambda \ge 0} e^{-\lambda t}\,\mathbb{E}\Big[e^{\lambda(X-\mathbb{E}X)}\Big].$$

Thus the moment bound is at least as good as the Chernoff bound. However, the former is much harder to use than the latter. Explain in words why that is the case. **Hint:** use $\mathbb{E}\left[e^{\lambda(X-\mathbb{E}X)}\right] \ge \mathbb{E}\left[1_{\{X-\mathbb{E}X>0\}}\, e^{\lambda(X-\mathbb{E}X)}\right]$ and expand in a power series.