Homework 1
High-Dimensional Probability for Data Science, Fall 2023

The first two problems reveal a variational characterization of expectation and median of a random variable X. Suppose X represents the income of a randomly chosen person in a town. Can you think of a single, non-random number that describes the town's income in the best way? One candidate is the average income, or the expected value $\mathbb{E}\, X$. By "best approximates" we mean that the expected value $t := \mathbb{E}\, X$ minimizes the MSE (mean squared error) $f(t) = \mathbb{E}(X - t)^2$ over all $t \in \mathbb{R}$.

Another candidate is the median $M(X)$, defined as a number M such that 50% of the people earn less than M and the other 50% earn more than M. In other words, the median is a number that satisfies $\mathbb{P}\left\{X \le M\right\} = \mathbb{P}\left\{X \ge M\right\} = 1/2$. The median also "best approximates" X, but with respect to a slightly different way of measuring the error. While the expected value minimizes the MSE, the median minimizes the MAE (mean absolute error) $g(t) = \mathbb{E}|X - t|$ over all $t \in \mathbb{R}$.

Let us now prove both these claims.

Problem 1 (Expectation minimizes the squared deviation)

Let X be a random variable with finite expectation. Show that the function $f(t) = \mathbb{E}(X - t)^2$ is minimized at $t = \mathbb{E}\, X$.

Hint: Check the identity $\mathbb{E}(X - t)^2 - \mathbb{E}(X - \mu)^2 = (t - \mu)^2$ *where* $\mu = \mathbb{E}\, X$.

Problem 2 (Median minimizes the absolute deviation)

Let X be a random variable with continuous distribution and finite expectation. Show that the function $g(t) = \mathbb{E}|X - t|$ is minimized at $t = M(X)$.

Hint: $\mathbb{E}|X - t|$ *is the expectation of a function of the random variable* X*. Express it as an integral that involves the probability density function of* X*. Break this integral into two integrals, over* $(\infty, t]$ *and over* $[t, \infty)$*. Take the derivative with respect to the variable* t *using Leibnitz integral rule, and set it to zero.*

Problem 3 (Variance of a bounded random variable)

(a) Consider a random variable X that takes values in an interval $[a, b]$. Show that

$$\mathrm{Var}(X) \leq \frac{(b-a)^2}{4}.$$

(b) For any interval $[a, b]$, find a random variable X such that

$$\mathrm{Var}(X) = \frac{(b-a)^2}{4}.$$

Hint: For part (a), use the result of Problem 1 to check that $\mathrm{Var}(X) \leq \mathbb{E}(X-t)^2$ *for any* $t \in \mathbb{R}$*. Now choose a convenient value of* t *to get to the conclusion.*

Next, we will prove a cute identity that was first noted by Charles Stein and is now called Stein's lemma. It has numerous applications in probability, including the modern and powerful Stein's method of proving central limit theorems. In Problem 5, we use Stein's lemma to compute all moments of the standard normal distribution.

Problem 4 (Gaussian integration by parts)

Let $Z \sim N(0, 1)$, i.e. Z is a standard normal random variable. Let $g : \mathbb{R} \to \mathbb{R}$ be a function. Show that

$$\mathbb{E}\, g'(Z) = \mathbb{E}\left[Z\, g(Z)\right].$$

Feel free to make any reasonable regularity assumptions on the function g, such as continuous differentiability, integrability, behavior at $\pm\infty$, etc.

Hint: Express the expectation as an integral that involves the pdf of Z*. Integrate by parts.*

Problem 5 (Gaussian moments)

Let $Z \sim N(0, 1)$ and $n \in \mathbb{N}$.

(a) Check that

$$\mathbb{E}[Z^{n+1}] = n\, \mathbb{E}[Z^{n-1}].$$

(b) Find a formula for $\mathbb{E}[Z^n]$ that involves only n.

Hint: in part (a), use Gaussian integration by parts. Part (b) follows by induction.
