Homework 1
Probability for Data Science, Fall 2022

Problem 1 (Expectation as the optimal estimator)

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

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

The rest of the problems are about *random vectors*, i.e. vectors $X = (X_1, \ldots, X_n)$ in $\mathbb{R}^n$ whose coordinates X_i are random variables. The coordinates X_i are not necessarily independent. The expected value of a random vector is defined coordinate-wise, that is $\mathbb{E}\,X = (\mathbb{E}\,X_1, \ldots, \mathbb{E}\,X_n)$.

We will use the standard notation of linear algebra, where $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$ denotes the inner product on $\mathbb{R}^n$, and $\|x\|_2 = \sqrt{\langle x, x\rangle} = (\sum_{i=1}^n x_i^2)^{1/2}$ denotes the Euclidean norm on $\mathbb{R}^n$.

Let us check that the standard properties of expectation of random variables, which you studied in the introductory probability course, pass on to random vectors.

Problem 2 (Expectation of random vectors)

Let X and Y be random vectors in $\mathbb{R}^n$.

(a) (Linearity) Prove that for any constants $a, b \in \mathbb{R}$, we have

$$\mathbb{E}(aX + bY) = a\,\mathbb{E}(X) + b\,\mathbb{E}(Y).$$

(b) (Multiplicativity) Prove that if X and Y are independent, then

$$\mathbb{E}\langle X, Y\rangle = \langle \mathbb{E}\,X, \mathbb{E}\,Y\rangle.$$

This generalizes the identity $\mathbb{E}(XY) = (\mathbb{E}\,X)(\mathbb{E}\,Y)$ for independent random variables.

Now you can prove the following generalization of Problem 1 for random vectors.

Problem 3 (Expectation as optimal estimator: random vectors)

Let X be a random vector with finite expectation. Show that the function $f(a) = \mathbb{E}\|X-a\|_2^2$ is minimized at $a = \mathbb{E}\,X$.

Hint: extend the identity mentioned in Problem 1 to random vectors.

As you may recall, the variance of a random variable X is defined as

$$\operatorname{Var}(X) = \mathbb{E}(X - \mathbb{E}\, X)^2.$$

By expanding the square and simplifying, we obrtain an equivalent definition:

$$\operatorname{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}\, X)^2.$$

Let us extend the last identity for random vectors.

Problem 4 (A variance-like identity)

For a random vector in $\mathbb{R}^n$, define

$$V(X) = \mathbb{E}\|X - \mathbb{E}\, X\|_2^2. \tag{1}$$

(a) Prove that

$$V(X) = \mathbb{E}\|X\|_2^2 - \|\mathbb{E}\, X\|_2^2.$$

(b) Prove that

$$V(X) = \frac{1}{2}\, \mathbb{E}\, \|X - X'\|_2^2,$$

where X' is an independent copy of X. (The latter means that the random variables X and X' are independent and identically distributed.)

Recall the additivity property of variance: if random variables $X_1, \ldots, X_k$ are independent, then

$$\operatorname{Var}(X_1 + \cdots + X_k) = \operatorname{Var}(X_1) + \cdots + \operatorname{Var}(X_k).$$

Let us extend this property for random vectors. We will do it in two steps.

Problem 5 (Additivity of variance)

(a) Check that if X and Y are independent random vectors in $\mathbb{R}^n$ with zero means, then

$$\mathbb{E}\|X + Y\|_2^2 = \mathbb{E}\|X\|_2^2 + \mathbb{E}\|Y\|_2^2.$$

(b) Prove that if random vectors $X_1, \ldots, X_k$ in $\mathbb{R}^n$ are independent, then

$$V(X_1 + \cdots + X_k) = V(X_1) + \cdots + V(X_k),$$

where $V(\cdot)$ is defined in (1).

(Hint: if $k = 2$ and all X_i have zero mean, part (b) reduces to part (a). To deduce part (b) in full generality, you may use induction on k, and consider $X_i' = X_i - \mathbb{E}\, X_i$ that have zero mean.)