401-4944-20L Mathematics of Data Science: Problem Set 1

(Spring 2020)

Afonso S. Bandeira
bandeira@math.ethz.ch
https://people.math.ethz.ch/~abandeira/

March 31, 2020

This homework is optional and it won't be graded. If you want to discuss a solution (to make sure it is correct) or want to ask questions about a problem stop by office hours or write a TA or myself an email and we can schedule a time to talk. Date is of last update (e.g. correction of typos)

Try not to look up the answers, you'll learn much more if you try to think about the problems without looking up the solutions. If you need hints, feel free to email me.

1.1 Linear Algebra

Problem 1.1 *Show the resut we used in class: If $M \in \mathbb{R}^{n\times n}$ is a symmetric matrix and $d \leq n$ then*

$$\max_{\substack{U\in\mathbb{R}^{n\times d}\\ U^TU=\mathrm{I}_{d\times d}}} \mathrm{Tr}\left(U^TMU\right) = \sum_{k=1}^{d} \lambda_k^{(+)}(M),$$

where $\lambda_k^{(+)}$ is the largest k-th eigenvalue of M.

1.2 Estimators

Problem 1.2 *Given $x_1, \cdots, x_n$ i.i.d. samples from a distribution X with mean μ and covariance Σ, show that*

$$\mu_n = \frac{1}{n}\sum_{k=1}^{n} x_k, \quad \textit{and} \quad \Sigma_n = \frac{1}{n-1}\sum_{k=1}^{n}\left(x_k-\mu_n\right)\left(x_k-\mu_n\right)^T,$$

are unbiased estimators for μ and Σ, i.e., show that $\mathbb{E}[\mu_n] = \mu$ and $\mathbb{E}[\Sigma_n] = \Sigma$.

1.3 Random Matrices

Recall the definition of a standard gaussian Wigner Matrix W: a symmetric random matrix $W \in \mathbb{R}^{n\times n}$ whose diagonal and upper-diagonal entries are independent $W_{ii} \sim \mathcal{N}(0,2)$ and, for $i<j$, $W_{ij} \sim \mathcal{N}(0,1)$. This random matrix emsemble is invariant under orthogonal conjugation: $U^TWU \sim W$ for any $U \in O(n)$. Also, the distribution of the eigenvalues of $\frac{1}{\sqrt{n}}W$ converges to the so-called semicircular law with support $[-2,2]$

$$d\mathrm{SC}(x) = \frac{1}{2\pi}\sqrt{4-x^2}\,1_{[-2,2]}(x).$$

(try it in Matlab, draw an histogram of the distribution of the eigenvalues of $\frac{1}{\sqrt{n}}W$ for, say $n = 500$.)

In the next problem, you will show that the largest eigenvalue of $\frac{1}{\sqrt{n}}W$ has expected value at most 2.[1] For that, we will make use of Slepian's Comparison Lemma.

Slepian's Comparison Lemma is a crucial tool to compare Gaussian Processes. A Gaussian process is a family of gaussian random variables indexed by some set T, more precisely is a family of gaussian random variables $\{X_t\}_{t\in T}$ (if T is finite this is simply a gaussian vector). Given a gaussian process X_t, a particular quantity of interest is $\mathbb{E}[\max_{t\in T} X_t]$. Intuitively, if we have two Gaussian processes X_t and Y_t with mean zero $\mathbb{E}[X_t] = \mathbb{E}[Y_t] = 0$, for all $t \in T$ and same variances $\mathbb{E}\left[X_t^2\right] = \mathbb{E}\left[Y_t^2\right]$ then the process that has the "least correlations" should have a larger maximum (think the maximum entry of vector with i.i.d. gaussian entries versus one always with the same gaussian entry). A simple version of Slepian's Lemma makes this intuition precise:[2]

In the conditions above, if for all $t_1, t_2 \in T$

$$\mathbb{E}[X_{t_1}X_{t_2}] \le \mathbb{E}[Y_{t_1}Y_{t_2}],$$

then

$$\mathbb{E}\left[\max_{t\in T} X_t\right] \ge \mathbb{E}\left[\max_{t\in T} Y_t\right].$$

[1] Note that, a priori, there could be a very large eigenvalue and it would still not contradict the semicircular law, since it does not predict what happens to a vanishing fraction of the eigenvalues.

[2] Although intuitive in some sense, this is a delicate statement about Gaussian random variables, it turns out not to hold for other distributions.

A slightly more general version of it asks that the two Gaussian processes X_t and Y_t have mean zero $\mathbb{E}[X_t] = \mathbb{E}[Y_t] = 0$, for all $t \in T$ but not necessarily the same variances. In that case it says that: If or all $t_1, t_2 \in T$

$$\mathbb{E}[X_{t_1} - X_{t_2}]^2 \geq \mathbb{E}[Y_{t_1} - Y_{t_2}]^2, \tag{1}$$

then

$$\mathbb{E}\left[\max_{t \in T} X_t\right] \geq \mathbb{E}\left[\max_{t \in T} Y_t\right].$$

Problem 1.3 *We will use Slepian's Comparison Lemma to show that*

$$\mathbb{E}\lambda_{\max}(W) \leq 2\sqrt{n}.$$

1. *Note that*

$$\lambda_{\max}(W) = \max_{v:\ \|v\|_2 = 1} v^T W v,$$

which means that, if we take for unit-norm v, $Y_v := v^T W v$ we have that

$$\mathbb{E}\lambda_{\max}(W) = \mathbb{E}\left[\max_{v \in \mathbb{S}^{n-1}} Y_v\right],$$

2. *Use Slepian to compare Y_v with $2X_v$ defined as*

$$X_v = v^T g,$$

where $g \sim \mathcal{N}(0, \mathrm{I}_{n \times n})$

3. *Use Jensen's inequality to upperbound $\mathbb{E}[\max_{v \in \mathbb{S}^{n-1}} X_v]$.*

Problem 1.4 *In this problem you'll derive the limit of the largest eigenvalue of a rank 1 perturbation of a Wigner matrix.*

For this problem, you don't have to justify all of the steps rigorously. You can use the same level of rigor that was used in class to derive the analogue result for sample covariance matrices. Deriving this phenomena rigorously would take considerably more work and is outside of the scope of this homework.

Consider the matrix $M = \frac{1}{\sqrt{n}}W + \beta vv^T$ for $\|v\|_2 = 1$ and W a standard Gaussian Wigner matrix. The purpose of this homework problem is to understand the behavior of $\lambda_{\max}(M)$. Because W is invariant to orthogonal conjugation we can focus on understanding

$$\lambda_{\max}\left(\frac{1}{\sqrt{n}}W + \beta e_1 e_1^T\right).$$

Use the same techniques as used in class to derive the behavior of this quantity.

(Hint: at some point, you'll probably have to integrate $\int_{-2}^{2} \frac{\sqrt{4-x^2}}{y-x} dx$*. You can use the fact that, for* $y > 2$*,* $\int_{-2}^{2} \frac{\sqrt{4-x^2}}{y-x} dx = \pi \left(y - \sqrt{y^2 - 4} \right)$ *(you can also use an integrator software, such as Mathematica, for this).*