Saarland University
Faculty of Mathematics and Computer Science
Department of Mathematics
Prof. Dr. Roland Speicher
Dr. Johannes Hoffmann

Exercises for the lecture
High Dimensional Analysis: Random Matrices and Machine Learning
Summer term 2023
Sheet 2
Hand-in: Friday, 12.05.2023, 22:00 Uhr via CMS

(Parts of) exercises that start with the indicator "Bonus" consider advanced or more philosophical questions; they come with extra points, but you cannot get more than 40 points per exercise sheet.

Definition. A random vector $x \in \mathbb{R}^p$ is a *Gaussian random vector* with mean vector $\mu \in \mathbb{R}^p$ and covariance matrix $\Sigma \in \mathbb{R}^{p\times p}$, denoted $x \sim N(\mu, \Sigma)$, if its probability density function ψ is given by

$$\psi(x) = \frac{1}{(2\pi)^{p/2}\det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2}\langle x-\mu, \Sigma^{-1}(x-\mu)\rangle\right).$$

The mean μ can be an arbitrary vector in $\mathbb{R}^p$, but the covariance matrix $\Sigma \in \mathbb{R}^{p\times p}$ has to be positive definite.

If $\mu = 0$ and $\Sigma = I_p$, then x is also called a *standard Gaussian random vector.*

Exercise 1 (6 points). Consider n independent copies $x_1, \dots, x_n \in \mathbb{R}^p$ of Gaussian random vectors with mean zero, where the components of each x_k are independent and half of them has variance 1 and the other half has variance 2. Plot a histogram of the p eigenvalues of the sample covariance matrix

$$\hat{\Sigma} := \frac{1}{n}\sum_{k=1}^{n} x_k x_k^T \in \mathbb{R}^{p\times p}$$

for the following parameters:

(i) $p = 100$, $n = 400$

(ii) $p = 100$, $n = 4000$

(iii) $p = 100$, $n = 40000$

(iv) $p = 500$, $n = 2000$

(v) $p = 1000$, $n = 4000$

in the domain $[0, 4]$. Choose $\frac{1}{10}$ as the width of the bars (or *bins*) in the histogram.

Further experimentation is encouraged.

Exercise 2 ($3 + 3 + 3^* + 3$ points). In this exercise, let $p = 1{,}000$.

(a) Consider n independent copies $x_1, \ldots, x_n \in \mathbb{R}^p$ of standard Gaussian random vectors, i.e., $x_i \sim N(0, I_p)$. As in Exercise 1, plot the histogram for the p eigenvalues of the sample covariance matrix and compare this with the Marchenko-Pastur distribution, which is given by the density

$$\psi(t) = \frac{1}{2\pi} \frac{\sqrt{(\gamma_+ - t)(t - \gamma_-)}}{\gamma t} \qquad \text{on the interval } [\gamma_-, \gamma_+],$$

where

$$\gamma = \frac{p}{n}, \qquad \gamma_- = (1 - \sqrt{\gamma})^2, \qquad \gamma_+ = (1 + \sqrt{\gamma})^2.$$

Do this for $\gamma = \frac{1}{4}$, $\gamma = \frac{1}{2}$ and $\gamma = 1$.

Hint: functions that draw histograms often can also automatically rescale the data to mimic a probability density function, which allows to draw actual densities like Marchenko-Pastur on top for easier comparison.

(b) The above is for $\gamma \leq 1$. How does the formula change for $\gamma > 1$? Plot the cases $\gamma = 2$ and $\gamma = 4$ like above.

(c) Bonus: what is the relation between the case γ and the case $\frac{1}{\gamma}$?
Hint: how are the eigenvalues of XX^T and X^TX for a rectangular matrix X related?

(d) Now change in $x_i \sim N(0, I_p)$ the covariance matrix from I_p to Σ by replacing the $(1,1)$-entry 1 with $1 + \beta$ and plot again the histograms from above for all combinations of $\gamma \in \left\{\frac{1}{4}, \frac{1}{2}, 1\right\}$ and $\beta \in \{1, 2\}$.

The BBP (Baik, Ben Arous, Péché) transition predicts that (in the limit $n \to \infty$) the eigenvalue $1 + \beta$ of Σ survives as a visible outlier in the eigenvalues of $\hat{\Sigma}$, as long as $\beta \geq \sqrt{\gamma}$, and then sits at the position $(1 + \beta)(1 + \frac{\gamma}{\beta})$. Check whether this is confirmed by your data!

Exercise 3 ($3 + 3$ points). Let $x \in \mathbb{R}^p$ be a random vector with probability density function $\psi : \mathbb{R}^p \to \mathbb{R}$, then the expectation of x is

$$E[x] = \int_{\mathbb{R}^p} x\psi(x) \, \mathrm{d}x \in \mathbb{R}^p.$$

and the covariance of x is

$$\Sigma(x) = E[xx^T] - E[x]E[x]^T \in \mathbb{R}^{p \times p}.$$

Let $A \in \mathbb{R}^{p \times p}$ and $b \in \mathbb{R}^p$.

(a) Show that E is linear in the sense that $E[Ax + b] = AE[x] + b$.

(b) Write $\Sigma(Ax + b)$ in terms of $\Sigma(x)$.

Exercise 4 ($3+3+3+2^*$ points)**.**

(a) Show that for a standard Gaussian random variable $x \sim N(0, I_p)$ we have $E[x] = 0$ and $\Sigma(x) = I_p$.

(b) Let $y = Ax + b$ be an affine transformation of $x \sim N(\mu, \Sigma)$ by an invertible matrix $A \in \mathbb{R}^{p\times p}$ and an arbitrary vector $b \in \mathbb{R}^p$. Find $\tilde{\mu}$ and $\tilde{\Sigma}$ such that $y \sim N(\tilde{\mu}, \tilde{\Sigma})$.

(c) Conclude that for $x \sim N(\mu, \Sigma)$ we have $E[x] = \mu$ and $\Sigma(x) = \Sigma$.

(d) Bonus: the affine transformation $y = Ax + b$ for $x \sim N(0, I_p)$ also makes sense for arbitrary matrices A that are not necessarily invertible. It seems appropriate to also call this a Gaussian random vector. Are there uniform descriptions which support this point of view?

Exercise 5 ($5+5$ points)**.** We will address here concentration estimates for the law of large numbers, and see that control of higher moments allows stronger estimates. Let x_i be a sequence of independent and identically distributed random variables with common mean $\mu = E[x_i]$ and write $X := (x_1, x_2, \dots)$ We put

$$S_n(X) = S_n(x_1, \dots, x_n) := \frac{1}{n}\sum_{i=1}^{n} x_i.$$

(a) Assume that the variance $V[x_i]$ is finite. Prove that we have then the weak law of large numbers, i.e., convergence in probability of S_n to the mean: for any $\varepsilon > 0$

$$P\{(x_1, \dots, x_n) : |S_n(X) - \mu| \geq \varepsilon\} \xrightarrow{n\to\infty} 0.$$

(b) Assume that the fourth moment of the x_i is finite, i.e. $E[x_i^4] < \infty$ (note that this implies that also all moments of smaller order are finite). Show that we then have

$$\sum_{n=1}^{\infty} P\{(x_1, \dots, x_n) : |S_n(X) - \mu| \geq \varepsilon\} < \infty.$$

(Note: by the Borel-Cantelli Lemma, this then implies the strong law of large numbers, i.e., $S_n \to \mu$ almost surely.)

One should also note that our assumptions for the weak and strong law of large numbers are far from optimal. Even the existence of the variance is not needed for them, but for proofs of such general versions one needs other tools than our simple consequences of the Chebyshev/Markov inequalities.