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 1
Hand-in: Friday, 28.04.2023, 22:00 Uhr via CMS

Exercise 1 (5 points). Show that

$$\int_{\mathbb{R}} \exp(-t^2)\,\mathrm{d}t = \sqrt{\pi}.$$

Hint: start by showing that

$$\left(\int_{\mathbb{R}} \exp(-t^2)\,\mathrm{d}t\right)^2 = \int_{\mathbb{R}}\int_{\mathbb{R}} \exp(-t^2 - s^2)\,\mathrm{d}t\,\mathrm{d}s$$

and compute the double integral using polar coordinates.

Definition. A real random variable x is a *Gaussian random variable* with *mean* $\mu \in \mathbb{R}$ and *variance* $\sigma^2 \in (0,\infty)$, denoted by $x \sim N(\mu,\sigma^2)$, if its probability density function ψ is given by

$$\psi : \mathbb{R} \to \mathbb{R}, \quad t \mapsto \psi(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right).$$

If $\mu = 0$ and $\sigma = 1$, then x is also called a *standard* Gaussian.

For a function $f : \mathbb{R} \to \mathbb{R}$, the *expectation* of $f(x)$ is

$$E[f(x)] = \int_{\mathbb{R}} f(t)\psi(t)\,\mathrm{d}t;$$

the *n-th moment* of x is given by

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

Exercise 2 $(3+4+3+3+2$ points$)$**.** Let $x \sim N(\mu, \sigma^2)$.

(a) Use Exercise 1 to compute $E[x^0]$ and $E[x^1]$. Explain the results.

(b) Show that x satisfies the moment recursion

$$E[x^n] = \mu E[x^{n-1}] + (n-1)\sigma^2 E[x^{n-2}] \quad \text{for all integers} \quad n \geq 2.$$

(c) Find the higher moments $E[x^2]$, $E[x^3]$, and $E[x^4]$.

(d) Give an explicit formula for the moments of x in the case $\mu = 0$.

(e) Calculate for the standard Gaussian $x \sim N(0,1)$ the first central moment

$$E\left[|x|\right] = \int_{\mathbb{R}} |t| \psi(t)\,\mathrm{d}t.$$

Exercise 3 $(3+3+4$ points$)$**.** We know from class that

$$\begin{aligned} P\left\{(t_1,\dots,t_p) \in B_p : |t_p| \geq \varepsilon\right\} &= \frac{2\int_\varepsilon^1 \mathrm{vol}[B_{p-1}(\sqrt{1-t^2})]\,\mathrm{d}t}{\mathrm{vol}[B_p]} \\ &= 2\frac{\mathrm{vol}[B_{p-1}]}{\mathrm{vol}[B_p]} \int_\varepsilon^1 (1-t^2)^{\frac{p-1}{2}}\,\mathrm{d}t. \end{aligned}$$

Note that this includes also in particular for $\varepsilon = 0$ a formula for the ratio of the unit balls of consecutive dimensions:

$$1 = 2\frac{\mathrm{vol}[B_{p-1}]}{\mathrm{vol}[B_p]} \int_0^1 (1-t^2)^{\frac{p-1}{2}}\,\mathrm{d}t.$$

By estimating the integrals we want to show from this an estimate for

$$P\left\{(t_1,\dots,t_p) \in B_p : |t_p| \geq \varepsilon\right\}.$$

(a) Prove for $y \geq 0$ the estimate

$$\int_y^\infty \exp(-t^2)\,\mathrm{d}t \leq \frac{\sqrt{\pi}}{2}\exp(-y^2).$$

Hint: treat the cases $y \leq 1$ and $y > 1$ separately.

(b) Let $p \geq 3$. Prove that

$$\int_0^1 (1-t^2)^{\frac{p-1}{2}}\,\mathrm{d}t \geq \int_0^{\frac{1}{\sqrt{p-1}}} (1-t^2)^{\frac{p-1}{2}}\,\mathrm{d}t \geq \frac{1}{2\sqrt{p-1}}.$$

Hint: Bernoulli's inequality states that $(1+a)^b \geq 1 + ab$ for all real numbers $b \geq 1$ and $a \geq -1$.

(c) Let $p \geq 3$. Show that

$$P\{(t_1, \ldots, t_p) \in B_p : |t_p| \geq \varepsilon\} \leq \sqrt{2\pi} \exp\left(-\varepsilon^2 \frac{p-1}{2}\right),$$

and thus

$$P\{(t_1, \ldots, t_p) \in B_p : |t_p| \leq \varepsilon\} \geq 1 - \sqrt{2\pi} \exp\left(-\varepsilon^2 \frac{p-1}{2}\right).$$

Hint: use Lemma 1.4: for $p \geq 1$ und $0 < \varepsilon \leq 1$ we have $(1-\varepsilon)^p \leq \exp(-\varepsilon p)$.

Definition. Let $x = (t_1, \ldots, t_p) \in \mathbb{R}^p$. We define the following norms:

- $\|x\|_2 := \sqrt{\sum_{k=1}^{p} t_k^2}$ (Euclidean norm, length, 2-norm)
- $\|x\|_1 := \sum_{k=1}^{p} |t_k|$ (ℓ_1 norm, Manhattan norm, 1-norm)
- $\|x\|_\infty := \max\{|t_k| : 1 \leq k \leq p\}$ (maximum norm, infinity norm)

Exercise 4 $(4+3+3$ points$)$**.** In this exercise, you are tasked with performing some numerical experiments and presenting the results as a histogram similar to the ones shown in the slides of the first lecture. You are free to choose your tools to do this, for example, you can use computer algebra systems with integrated plotting like MATLAB, Maple, or Mathematica, or use a programming language of your choice to compute the values and combine it with some visualization tool to plot the histogram.

As the slides in class, this exercise should give you a feeling for the concentration phenomena. We consider in the following Gaussian random vectors $x \in \mathbb{R}^p$ with independent standard Gaussians as components; i.e., every component of the vector is a Gaussian random variable with mean zero and variance 1 and the components are independent from each other. Such vectors show concentration.

The concentration property says roughly that for our high-dimensional vector $x = (t_1, \ldots, t_p) \in \mathbb{R}^p$ any function $f(x) = f(t_1, \ldots, t_p)$ that depends (in a 'smooth' way) on the components (but not too much on any of them) is essentially constant, and thus close to the average value $E[f(x)]$ of the function. (Later in the course the parentheticals will be made more precise via the notion of Lipschitz functions.) In part (a) we consider the relatively simple situation where the function f is essentially a sum of independent components. In that case the expectation is also quite easy to determine. In part (b), the function f is much more non-linear, and its expectation is not directly clear. In part (c), we arrange our vectors in a matrix form and take as function f the largest eigenvalue of those matrices – these are very non-linear (and not very concrete) functions of the matrix entries, but still 'smooth enough', so that we also have concentration of the eigenvalues.

(a) For f we take here the 1-norm $f(x) = \|x\|_1$ and the 2-norm $f(x) = \|x\|_2$. For each of the two cases plot a histogram of $f(x)$ for $1,000$ realizations of the vector $x \in \mathbb{R}^p$. Do this for $p = 1$, $p = 100$, and $p = 10,000$. You should recognize in those plots the dependence of $E[f(x)]$ on p. Can you explain those values? (For the case of the 1-norm, Exercise 2(e) should be relevant.)

(b) For f we take now the maximum norm $f(x) = \|x\|_\infty$. Plot a histogram of $f(x)$ for 1,000 realizations of the vector $x \in \mathbb{R}^p$. Do this for $p = 1$, $p = 10$, $p = 10,000$, and $p = 100,000$.

The value of $E[f(x)]$ will probably not become clear from the plots. Instead, we can look at some estimates for the concentration: let M be the median of the $f(x_j)$, then for all $\varepsilon > 0$ we have

$$P\left(f(x) > (1+\varepsilon)M\right) \leq \sqrt{\frac{2}{\pi}} \frac{p}{(1+\varepsilon)M} \exp\left(-\frac{1}{2}(1+\varepsilon)^2 M^2\right).$$

Check for some reasonable values for ε whether this is compatible with your data.

(c) We consider now a sample covariance matrix

$$\hat{\Sigma} = \frac{1}{n} \sum_{k=1}^{n} x_k x_k^T = \frac{1}{n} X X^T,$$

where $x_1, \ldots, x_n \in \mathbb{R}^p$ are n independent copies of our p-dimensional random vectors x as above, and $X = [x_1 x_2 \ldots x_n] \in \mathbb{R}^{p \times n}$ is the corresponding data matrix. (Such random matrices $\hat{\Sigma}$ are called *Wishart matrices.*) We take as our function f now the largest eigenvalue of $\hat{\Sigma}$ (which is the same as the square of the largest singular value of the matrix $X/\sqrt{n}$.) This $f(X) = f(x_1, \ldots, x_n)$ is a very non-linear (and not explicit) function of the $p \times n$ independent standard Gaussian entries of the data matrix X. Plot a histogram of $f(X)$ for $1,000$ realizations of the data matrix $X = [x_1 \ldots x_n]$. Do this for $p = n = 1$, $p = n = 10$, $p = n = 50$, $p = n = 100$.

In this case concentration estimates are quite complicated and not very explicit, so let us just quote the following simple rules of thumb (according to the paper "On the distribution of the largest eigenvalue in principal component analysis" by Iain Johnstone): define

$$\mu := \frac{1}{n}\left(\sqrt{n-1} + \sqrt{p}\right)^2 \quad \text{and} \quad \sigma := \frac{1}{n}\left(\sqrt{n-1} + \sqrt{p}\right)\left(\frac{1}{\sqrt{n-1}} + \frac{1}{\sqrt{p}}\right)^{\frac{1}{3}}.$$

Then, about 83% of the distribution is less than μ, about 95% lies below $\mu + \sigma$, and about 99% lies below $\mu + 2\sigma$.

Check whether this is compatible with your data.

Further experimentation is encouraged.