max planck institut
informatik

You can discuss these problems with other students, but everybody must hand in their own answers. You can use computers etc. to perform the algebraic operations, but you must show the intermediate steps (and "computer said so" is never a valid answer). You can return either computer-typeset solutions by email (but no scanned or photographed solutions are accepted), or legibly hand-written or computer-typeset solutions personally to the lecture. Notice that the DL is strict. *Remember to write your name and matriculation number to every answer sheet! If you want to discuss the solutions with the tutor, the tutorial meeting is the time to do that. If you cannot attend the tutorial meeting but want to discuss with the tutor, you must schedule a meeting with the tutor via email.*

Problem 1 (Correlation matrix). Let $\boldsymbol{x} = (x_i)_{i=1}^n$ be a (column) vector of n zero-centered random variables. The *covariance* $\operatorname{cov}(x_i, x_j)$ is defined as

$$\operatorname{cov}(x_i, x_j) = \mathbb{E}[x_i x_j] \ , \tag{1.1}$$

The *correlation matrix* $\boldsymbol{\Sigma}$ is defined as

$$\boldsymbol{\Sigma} = \mathbb{E}[\boldsymbol{x}\boldsymbol{x}^T] = \big(\operatorname{cov}(x_i, x_j)\big)_{i,j} \ . \tag{1.2}$$

What are the requirements for random variables x_i that ensure that the covariance matrix is an identity matrix? Give the requirements, and prove that if all x_i satisfy them, $\boldsymbol{\Sigma}$ is an identity matrix.

Hint: consider what $\boldsymbol{\Sigma}_{i,i} = \operatorname{cov}(x_i, x_i)$ tells about random variable x_i.

Problem 2 (Kurtosis of a sum). Recall that the *kurtosis* of a random variable X with zero mean is

$$\operatorname{kurt}(X) = \mathbb{E}[X^4] - 3(\mathbb{E}[X^2])^2 \ . \tag{2.1}$$

One way to understand the importance of the factor 3 in (2.1) is to consider a sum of two independent random variables. Let X and Y be two independent random variables with zero mean and unit variance, i.e.

$$\mathbb{E}[X] = 0 \qquad \mathbb{E}[X^2] = 1 \tag{2.2}$$

$$\mathbb{E}[Y] = 0 \qquad \mathbb{E}[Y^2] = 1 \ . \tag{2.3}$$

Show that

$$\operatorname{kurt}(X + Y) = \operatorname{kurt}(X) + \operatorname{kurt}(Y) \ . \tag{2.4}$$

Can you see the importance of factor 3?

Hint: Use binomial formula and linearity of expectation.

Problem 3 (Kurtosis of normal distribution). Another way to see the importance of the factor 3 is to consider the kurtosis of normal distribution. We will prove that if X is normally distributed with 0 mean, then $\operatorname{kurt} X = 0$. To compute the kurtosis, we need the fourth moment $\mathbb{E}[X^4]$. To compute it, we use very powerful and general technique of *moment-generating functions*. The moment-generating function of random variable Y is

$$M_Y(t) = \mathbb{E}[\exp(tX)] = \mathbb{E}[\mathrm{e}^{tX}] \ , \quad t \in \mathbb{R} \ . \tag{3.1}$$

One important feature of moment-generating functions is that if we know M_Y, we can easily compute the nth moment of Y by differentiating M_Y n times and evaluating the derivative at origin. In other words,

$$\frac{d^n M_Y}{dt^n}(0) = \mathbb{E}[Y^n] \ , \tag{3.2}$$

where $\frac{d^n M_Y}{dt^n}(0)$ is the nth derivative of M_Y evaluated at origin. (Here we assume that the derivative exists.)

The moment-generating function for normally distributed X with 0 mean and variance σ^2 is

$$M_X(t) = \exp(\sigma^2 t^2/2) \ . \tag{3.3}$$

Use (3.3) to compute $\mathbb{E}[X^4]$ and conclude that $\mathrm{kurt}(X) = 0$.

Problem 4 (Kurtosis of various distributions). Plot the probability density functions (or probability mass functions) and compute the kurtosis of the following probability distributions (the probability density (or mass) function and moment-generating function are given below the distribution). Are the distributions sub- or super-Gaussian? Notice that all of the distributions have zero mean.

1. Continuous uniform from $a = -1$ to $b = 1$
   $f(x; a, b) = 1/(b-a)$ for $x \in [a, b]$ and $M(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}$ for $t \neq 0$ and $M(t) = 1$ for $t = 0$.
2. Centered binomial with $n = 100$ and $p = 1/2$ (i.e. $Y \sim X - np$ where $X \sim \mathrm{Binomial}(n, p)$)
   $f(k; n, p) = \binom{n}{k+np} p^{k+np}(1-p)^{n-(k+np)}$ for $k \in \{-np, 1-np, 2-np, \ldots, n-1-np, n-np\}$ and $M(t) = (1-p)^n e^{-npt}\left(\frac{-e^t p + p - 1}{p-1}\right)^n$
3. Laplace distribution for $\mu = 0$ and $b = 1$
   $f(x; \mu, b) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right)$ and $M(t) = \exp(\mu t)/(1 - b^2 t^2)$ for $|t| < 1/b$

Your solution must contain printed plots (even if your solutions are otherwise hand-written) and the values of the second and fourth moment and the kurtosis. You can use computers to evaluate the second and fourth moment and you do not need to report, e.g. the fourth derivative of the moment-generating function (only the value of it when evaluated at zero). You must return the final outcomes, that is, the plots and the values of the moments, not only scripts that would plot the plots or evaluate the moments.

Your solution is acceptable if you solve it for two out of the three distributions. You can gain an extra point if you return solutions for all three distributions.

Problem 5 (Whitening). Most textbooks (and Wikipedia) explain the whitening process as follows: Given data matrix $\boldsymbol{A}$ (where rows are observations and columns variables), compute the correlation matrix $\boldsymbol{C} = \boldsymbol{A}^T\boldsymbol{A}$. Then, compute the *eigendecomposition* of $\boldsymbol{C}$, $\boldsymbol{C} = \boldsymbol{Q}\boldsymbol{\Delta}\boldsymbol{Q}^T$, where $\boldsymbol{Q}$ is an orthogonal matrix and $\boldsymbol{\Delta}$ is diagonal matrix with non-negative entries. To whiten $\boldsymbol{A}$, we multiply $\boldsymbol{A}$ from right with $\boldsymbol{Q}\boldsymbol{\Delta}^{-1/2}$, where $(\boldsymbol{\Delta}^{-1/2})_{ii} = 1/\sqrt{(\boldsymbol{\Delta})_{ii}}$ if $(\boldsymbol{\Delta})_{ii} \neq 0$ and $(\boldsymbol{\Delta}^{-1/2})_{ii} = 0$ otherwise.

In the lectures it was claimed that if $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ is the SVD of $\boldsymbol{A}$, then the whitened $\boldsymbol{A}$ is $\boldsymbol{U}$. Prove that these two processes yield the same solution, that is

$$\boldsymbol{U} = \boldsymbol{A}\boldsymbol{Q}\boldsymbol{\Delta}^{-1/2} \ . \tag{5.1}$$

Hint: eigendecomposition is unique, that is, if $\boldsymbol{C} = \boldsymbol{Q}\boldsymbol{\Delta}\boldsymbol{Q}^T$ for some orthogonal $\boldsymbol{Q}$ and diagonal $\boldsymbol{\Delta}$ with nonnegative entries, then $\boldsymbol{Q}\boldsymbol{\Delta}\boldsymbol{Q}^T$ is the eigendecomposition of $\boldsymbol{C}$. Use the SVD of $\boldsymbol{A}$ to express $\boldsymbol{C}$ and find a definition of $\boldsymbol{Q}$ and $\boldsymbol{\Delta}$ in terms of SVD of $\boldsymbol{A}$.

Problem 6 ($-\log\cosh$). In the MLE ICA, logarithms of super-Gaussian distribution's probability density functions are estimated with

$$\log p(x) = \alpha - 2\log\cosh(x)\ . \tag{6.1}$$

The gradient algorithm for MLE uses function $g(x)$, defined as the first derivative of $\log p(x)$, denoted $(\log p(x))'$. Show that

$$(\log p(x))' = -2\tanh(x)\ . \tag{6.2}$$