UML2015
Exercise set 5
Solutions to be presented in the 24.4.2015 session
(There was a mistake in the date)

Exercise 1:
Kurtosis of a zero mean random variable X is defined as

$$\mathrm{kurt}(X) = \mathbb{E}(X^4) - 3(\mathbb{E}(X^2))^2 \tag{1}$$

Kurtosis is a measure of the "peakedness" of the probability distribution of X. Calculate the kurtosis for the

1.1 Uniform distribution $p(x)$,

$$p(x) = \begin{cases} \frac{1}{2\sqrt{3}} & |x| \leq \sqrt{3} \\ 0 & \text{else.} \end{cases} \tag{2}$$

1.2 Laplacian distribution $p(x)$,

$$p(x) = \frac{1}{\sqrt{2}} \exp(-\sqrt{2}|x|). \tag{3}$$

1.3 Gaussian distribution $p(x)$ with mean zero and variance σ^2,

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{x^2}{2\sigma^2}\Big). \tag{4}$$

1.4 Calculate the kurtosis for the following mixture of Gaussians (called a Gaussian scale mixture)

$$p(x) = \frac{1}{2}(p_1(x) + p_2(x)), \tag{5}$$

where

$$p_i(x) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\Big(-\frac{x^2}{2\sigma_i^2}\Big). \tag{6}$$

Show that the kurtosis is always positive if $\sigma_1 \neq \sigma_2$.

1.5 Consider now the following mixture of Gaussians of the same variance but different means:

$$p(y) = \frac{1}{3}\big(p_\mu(y) + p_0(y) + p_{-\mu}(y)\big) \tag{7}$$

where

$$p_a(y) = \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{(y-a)^2}{2}\Big). \tag{8}$$

Calculate the kurtosis and show that it is always negative for nonzero μ.

1.6 Let $U = X + \alpha Y$, where X follows the distribution in Equation (5), and Y has the distribution in Equation (7). Furthermore, X and Y are independent. How can you choose $\alpha \in \mathbb{R}$ so that $\text{kurt}(U) = 0$?

Advice: You'll need the integration by parts technique and the change of integration variable. By planning the calculations well, you can avoid unnecessary work by reusing some integration by parts results multiple times. Also the formula $(a+b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$ can turn out useful.

Exercise 2:

For a zero-mean random variable, skewness of a distribution is defined to be its third moment, i.e.

$$\text{skew}(X) = \mathbb{E}(X^3). \tag{9}$$

It measures the asymmetry of a distribution. If the independent variables $\mathbf{s}$ have a highly asymmetric distribution, skewness can be used to perform ICA.

Suppose Z is $N \times K$ data matrix, and denote as $\mathbf{z}_k$ the columns of Z with each fixed $1 \leq k \leq K$. We would like to maximize

$$J(\mathbf{w}) = \frac{1}{K} \sum_{k=1}^{K} (\mathbf{w} \cdot \mathbf{z}_k)^3 \tag{10}$$

with respect to a vector $\mathbf{w} \in \mathbb{R}^N$ under the constraint $\|\mathbf{w}\| = 1$.

2.1 Find the gradient $\nabla J(\mathbf{w})$.

2.2 What is the gradient-ascent optimization iteration, taking into account the constraint?

2.3 Take the limit of large stepsizes, i.e. $\mu \to \infty$. What is the optimization iteration now?

Exercise 3:

Assume the data $\mathbf{z}_1, \ldots, \mathbf{z}_K$ is iid sample generated by the model $\mathbf{Z} = A\mathbf{S}$, where $\mathbf{Z}$ and $\mathbf{S}$ are random vectors, A is orthogonal $N \times N$ matrix, and the components of $\mathbf{S}$, S_n when $1 \leq n \leq N$, are independent from each other.

3.1 Write down the log-likelihood $\ell(A|\mathbf{z}_1, \ldots, \mathbf{z}_K)$ of A in terms of the distributions $p_{S_n}(s_n)$ which may be arbitrary.

3.2 Show that the log-likelihood does not depend anymore on the matrix A if the S_n are Gaussian.

Exercise 4:

In the maximum likelihood estimation of the ICA model, we may not know the densities of the independent variables $\mathbf{S}$. Therefore, they must be approximated in one way or the other.

We have seen in the lecture that as long the approximation $\tilde{p}_{S_i}(s_i)$ fulfills

$$\mathbb{E}\big(S_i g_i(S_i) - g_i'(S_i)\big) > 0 \tag{11}$$

for all i, where $g_i = \tilde{p}_i'/\tilde{p}_i$ and where the expectation $\mathbb{E}$ has been defined with the true distribution $p_{S_i}(s_i)$ as weight, maximization of the likelihood will lead to the right solution for the mixing matrix B (see Theorem 1 on page 61).

4.1 Assume that S_i is Gaussian (zero mean, unit variance). Is the condition in Eq (11) fulfilled? (Advice: Examine the quantities $\mathbb{E}(S_i g_i(S_i))$ and $\mathbb{E}(g_i'(S_i))$ with integration by parts. Assume that $g_i(s_i)$ grow slower than $\exp(s_i^2/2)$.)

4.2 Suppose you make the choice $g_i(s_i) = s_i^3$. To what does the condition in Eq (11) correspond to? (Assume that S_i is zero mean and normalized to unit variance.)

4.3 Show that making the choice $g_i(s_i) = -s_i$ corresponds to $\tilde{p}_i$ being a Gaussian distribution.