UML2014
Exercise set 5
Solutions to be presented in the 11.4.2014 session

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

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

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 > 0 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 mean. You can use the fact that the normal distribution has skewness of zero, i.e. $E(u^3) = 0$. (*Hint:* You might want to use that $E(u^2) = V(u) + E(u)^2$. Also: $(a + b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$.)

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 $\mathrm{kurt}(u) = 0$?

Exercise 2:

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

$$\mathrm{skew}(x) = \mathbb{E}(x^3). \tag{9}$$

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

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

under the constraint that $\|\mathbf{w}\| = 1$.

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

2.2 What is the gradient-ascent optimization iteration, considering the constraint $\|\mathbf{w}\|$?

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 and follows the model $\mathbf{z} = A\mathbf{s}$, where $\mathbf{z} \in \mathbb{R}^N$ is white random vector, A is orthonormal, i.e. $A^T A = I$, and the s_n are independent random variables.

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

3.2 Show that the log-likelihood does not depend anymore on the matrix A if the distribution of 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 $\tilde{p}_i(s_i)$ fulfills

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

for all i, where $g_i = \tilde{p}_i'/\tilde{p}_i$, maximization of the likelihood will lead to the right solution for the mixing matrix B (see Theorem 1 on page 66).

4.1 Assume that s_i is Gaussian (zero mean, unit variance). Is the condition in Eq (11) fulfilled? (Advise: 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 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.