Mock Exam of Computational Statistics - Summer 2017

Name:
Matriculation n.:

Note that this exam is composed of n questions worth a total of m points. You have two hours and thirty minutes to complete it. Please **do not forget** to turn over the sheet!

A. *Descriptive statistics*

1. Please compute the moments $\langle x^0 \rangle$, $\langle x^2 \rangle$ and $\langle x^4 \rangle$ of the distribution $p(x)dx \propto \exp(-|\mathrm{x}|)$ for the range $-\infty < x < +\infty$.
2. Why are the odd moments $\langle x^{2n+1} \rangle$ equal to 0?
3. Please compute the moment generating function and use it to show $\langle x^0 \rangle = 0$.
4. Imagine the transformation to a new variable $y = x^2$. What is the transformation of the probability density?

B. *Uniform random variables*

1. Compute the mean $E[X]$ and variance $Var[X]$ of a uniformly distributed variable X between a and b. Simply stating these values is not enough, and an explicit calculation is needed.
2. Now, consider you have two random variables X and Y distribute between a and b. Compute the mean of the variable $Z = X + Y$ using the moment generating function of Z.

C. [Pseudo-code] *Non-uniform distributions*

The Cauchy distribution is given by

$$p(x) = N \frac{1}{1+x^2}, \ -\infty < x < \infty. \tag{1}$$

1. Assuming you are writing in the programming language R, compute the normalization constant N.
2. Analytically derive the cumulative density function CDF of the Cauchy distribution. You can use the result $\int 1/(1+x^2) = \mathrm{arctan(x)} + \mathrm{const.}$
3. Using the previous result, write a code to sample from the Cauchy distribution using the inversion sampling method.
4. Once you have generated your sample, write a code to compute its mean and variance without using R's built-in functions.

D. *Fit of a linear model*

1. Please derive the fit of a polynomial $q(x) = ax^3 + bx^2 + cx + d$ of second order to n data points (x_i, y_i) with Gaussian error σ_i. Start by writing down a suitable likelihood.
2. Please isolate an expression for the respective χ^2 function.
3. Please formulate a linear system of equations applying the principle of maximum likelihood.
4. Please write down the formal solution of the previous system of equations.

E. t-*test*

Suppose we have 49 data points with sample mean $\overline{x} = 6.25$ and sample variance $s = 36$. We want to test the following hypotheses:

1. H_0: the data is drawn from $N(4, \sigma^2)$, where σ is unknown. H_A: the data is drawn from $N(\mu, \sigma^2)$ where $\mu \neq 4$. Test for significance at the $\alpha = 0.05$ level. Use the t-table to find the p value.
2. Draw a picture showing the null pdf, the rejection region and the area used to compute the p-value for the previous part.