Problems on *statistics and data analysis (MVComp2)*

lecturer: Björn Malte Schäfer
head tutor: Alessio Spurio Mancini
summer term 2016

Problem sheet 3

To be handed in during the exercise group on 23.May.2016

1. *Gram-Charlier series* *(10 points)*

If a Gaussian distribution with zero mean and variance $\sigma^2 = \kappa_2$ is weakly perturbed by the presence of a non-vanishing third and fourth cumulant κ_3 and κ_4, respectively, the distribution $p(x)dx$ can be approximated with the Gram-Charlier series

$$p(x)\mathrm{d}x = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \times \left[1 + \frac{\kappa_3}{3!\sigma^3} H_3\left(\frac{x}{\sigma}\right) + \frac{\kappa_4}{4!\sigma^4} H_4\left(\frac{x}{\sigma}\right)\right] \mathrm{d}x,$$

with the argument x/σ of the Hermite polynomials $H_n(x/\sigma)$.

(a) The Hermite polynomials $H_n(x/\sigma)$ can be computed by n-fold differentiation of a Gaussian,

$$H_n\left(\frac{x}{\sigma}\right) = (-\sigma)^n \exp\left(\frac{x^2}{2\sigma^2}\right) \frac{\mathrm{d}^n}{\mathrm{d}x^n} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

Check that this is true by calculating $H_1(x/\sigma)$, $H_2(x/\sigma)$, $H_3(x/\sigma)$, $H_4(x/\sigma)$ and verifying that you get

$$H_1\left(\frac{x}{\sigma}\right) = \frac{x}{\sigma}, \quad H_2\left(\frac{x}{\sigma}\right) = \left(\frac{x}{\sigma}\right)^2 - 1, \quad H_3\left(\frac{x}{\sigma}\right) = \left(\frac{x}{\sigma}\right)^3 - 3\frac{x}{\sigma}, \quad H_4\left(\frac{x}{\sigma}\right) = \left(\frac{x}{\sigma}\right)^4 - 6\left(\frac{x}{\sigma}\right)^2 + 3.$$

(b) Compute (by integration by parts) the cumulative distribution $P(x)$ and express it in terms of the cumulative function of the Gaussian distribution, $\Phi(x/\sigma)$, which can be written as

$$\Phi\left(\frac{x}{\sigma}\right) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}\sigma}\right)\right),$$

where $\mathrm{erf}(x/\sigma)$ is the error function.

(c) By using the derivative relation

$$\frac{\mathrm{d}}{\mathrm{d}x} H_n\left(\frac{x}{\sigma}\right) = \frac{n}{\sigma} H_{n-1}\left(\frac{x}{\sigma}\right)$$

of the Hermite polynomials $H_n(x/\sigma)$, verify that the derivative of the Gram-Charlier series takes the form

$$\frac{\mathrm{d}}{\mathrm{d}x} p(x) = -\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \times \left[\frac{x^2}{\sigma^2} + \frac{\kappa_3}{3!\sigma^4} H_4\left(\frac{x}{\sigma}\right) + \frac{\kappa_4}{4!\sigma^5} H_5\left(\frac{x}{\sigma}\right)\right].$$

(d) Show by direct computation of the mean and variance that their values do not depend on the value of κ_3 and κ_4.

2. *Fit of a third-order polynomial* *(10 points)*

Work out explicitly a max-likelihood fit of a third-order polynomial $y(x) = p_0 + p_1 x + p_2 x^2 + p_3 x^3$ by minimising the squared distance between the data points y_i and the model function $y(x)$ at the positions x_i:

(a) compose the matrix of moments,

(b) after stating the condition for the matrix of moments to be invertible, calculate the inverse,

(c) use what you found to fit the set of data contained in the file `third_order_pol.dat`, which you can download from the Moodle webpage of the course. In order to load data from file, you can use `numpy.loadtxt`: in the text file you will find explanations on how to actually import the dataset.

3. *Python exercise* *(10 points)*

Write a Python script to verify the central limit theorem: consider the sum of m uniformly distributed random numbers with zero mean and show that as you increase m the distribution of the sum converges towards a Gaussian. Do this by plotting the distributions for $m = 1, 2, 4, 8$.