Humboldt-Universität zu Berlin
Institut für Physik

Winter Semester 2010/2011

Dr. U. Husemann, V. Ferrara

Statistical Methods of Data Analysis
Problem Set #4

Due Date: Thursday, November 25, 2010, during the lecture

Problem 1: (15 points)

A bivariate Gaussian distribution is given by

$$G_2(x_1,x_2) = \frac{1}{2\pi}\frac{1}{\sigma_1\sigma_2\sqrt{1-\varrho^2}} \exp\left\{-\frac{1}{2(1-\varrho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}-2\varrho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right]\right\},$$

with the usual parameters μ_1, μ_2 for the mean values, σ_1, σ_2 for the widths and ϱ for the correlation coefficient.

i) Work out the marginal PDF $f_{u_1}(u_1)$ with $u_1 = (x_1 - \mu_1)/\sigma_1$. Show that $f_{u_1}(u_1)$ is independent of the correlation coefficient ϱ.

ii) Show that the ellipse that is formed by curves of constant G_2 is rotated by an angle φ according to

$$\tan(2\varphi) = \frac{2\varrho\,\sigma_1\sigma_2}{\sigma_2^2-\sigma_1^2}.$$

iii) Use the ROOT class `TEllipse` to plot the 1σ ellipses of G_2 for the following two sets of parameter values:

Set #	μ_1	μ_2	σ_1	σ_2	ϱ
1	1	1	2	3	0
2	0	0	2	1	-0.9

Tip: You can obtain the lengths of the ellipse's semiaxes from the eigenvalues of the covariance matrix.

Problem 2: (8 points)

Statistics can be employed to describe the thermodynamics of ideal gases:

i) Consider a three-dimensional vector $\vec{q}$ whose components are independent normal random variables with mean equal to zero and variance σ^2. Find the statistical distribution law describing the modulus of $\vec{q}$.

ii) The components (v_x, v_y, v_z) of the velocity $\vec{v}$ of the molecules of an ideal gas are random variables satisfying the central limit theorem conditions. Starting from the distribution of the modulus of $\vec{v}$, derive an expression for the kinetic energy density $f(E)\,\mathrm{d}E$ of the molecules of an ideal gas at the absolute temperature T knowing that the relationship between the gas temperature and the variance of the velocity of the molecules is expressed by $m\sigma^2 = k_BT$, where k_B is the Boltzmann constant.

(see reverse side)

Problem 3: (12 points)

The `TRandom` class of the ROOT package can be used to generate random numbers according to various PDFs.

i) Generate uniform random numbers x_i in the interval (0,1]. Create histograms with the sum of 10, 100 and 1000 random numbers (1000 entries each). Which distribution do they follow? Fit the distribution in ROOT to extract the parameters of the underlying PDF.

ii) Which distribution would the product of uniform random numbers follow? Why is it not practical to histogram the product of random variables? What could be histogrammed instead?

iii) Generate random numbers x_i according to a Gaussian distribution with $\mu = 1$ and $\sigma = 1$. Examine the function

$$\hat{z} = \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma_i^2}.$$

by generating histograms with 1000 entries each of the distribution of $\hat{z}$ for $n = 4$, $n = 12$ and $n = 60$. Which distribution do they follow? Determine expectation value and variance of the distribution both from the statistics box of the histogram and from a fit of the appropriate PDF to the histogram.

Tip: You can either implement the PDF you need for the fit as a custom fit function or look for it in the namespace `ROOT::Math` as defined in ROOT.