Computational Statistics

Exercise sheet 6

To be handed in for marking until tutorial of 31/05/2017

Nelson Lima and Luca Amendola, ITP, Heidelberg

31/05/2017

www.thphys.uni-heidelberg.de/~amendola/teaching.html

Problem 1 - Correlated variables [7.5 points]

1. Find the mean, the covariance and correlation matrices of the dataset "bivariate-measurements.txt" available in the Dropbox folder "Sheet 6/Data". You are not supposed to use the built-in functions to compute the covariance (and correlation) matrix but to write your own function from scratch. This will be checked in your code.

2. Plot the dataset. Using the covariance matrix previously estimated, draw the major axis of the ellipse formed by the points. Also plot the 68% and 95% confidence contours of your dataset [Note: for this particular step, the function we have used in the tutorials required, simultaneously, the usage of the packages 'MASS' and 'car'. If you are having problems with the installation of the package 'car', there are other ways to plot the confidence contours that do not require the package 'car'. If the problems persist, let me know.]

3. Consider the ellipse of point (2), of which you have found the eigenvectors. The eigenvectors are normalized to unity and the eigenvalues tell you how long is the corresponding axis. If you stretch this basic ellipse by the factor 1.5152, you obtain the 68.27% confidence-level contour. Your assignment is then to show that the number of elements of "bivariate-measurements.txt" within this ellipse are indeed approximately 68.27% of the total.

Problem 2 - Non-uniform distributions [7.5 points]

The following distribution is often used to model the angle θ at which photons scatter. The photons might be passing through the atmosphere in a climate modeling problem or trough human tissue in some medical application.

Let $0 < \theta < \pi$ and $0 < \phi < 2\pi$. Then the probability distribution function of θ is

$$p(\theta; g) = \frac{1}{4\pi}\frac{1-g^2}{(1+g^2-2g\cos(\theta))^{3/2}}. \tag{1}$$

1. Verify that $p(\theta; g)$ is normalized over the solid angle;

2. Plot $2\pi p(\theta; g)\sin(\theta)$ for $g = 0.1,\ 0.3,\ 0.5$;

3. Now, lets write $p(\theta; g)$ as a function of $\eta = \cos(\theta)$ such that $-1 < \eta < 1$,

$$p(\eta; g) = \frac{1}{2} \frac{1 - g^2}{(1 + g^2 - 2g\eta)^{3/2}} \tag{2}$$

   Develop a way to sample this distribution distribution for g $=$ 0.5, returning the sampled values of η and θ and checking if they agree with their respective probability distribution functions. Turn in a mathematical derivation of your method along with your code.

4. Estimate the mean of θ and its variance.

Problem 3 - Higher order moments of multivariate Gaussian [5 points]

Following what you have learned in the lectures, compute the following higher-order moments of the multivariate gaussian distribution of two variables x_1 and x_2 with covariance matrix C_{ij} such that $\sigma_{\mathrm{ii}} \equiv \sigma_{\mathrm{i}}^2$ and $\sigma_{ij} \neq 0$:

- $E[x_1^4 x_2^2]$
- $E[x_1^3 x_2]$