Multivariate Statistical Analysis, Exercise 1, Fall 2011, Prof. S.C. Jeng

September 30, 2011 TA: H.C. Cheng

(Geometrical Interpretation of Statistics)

From an experiment we obtain a size N random sample $\mathbf{x}_1, \dots \mathbf{x}_N$ from some unknown population $f(\mathbf{x})$, where each random sample $\mathbf{x} = \begin{bmatrix} x_1 & \cdots & x_p \end{bmatrix}$ contains p variables. The random sample can be formulated as the following matrix

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \cdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Np} \end{bmatrix},$$

where in this experiment the number of samples N equals 10000 and the number of variables of each sample data $p = 2$. Please answer the following questions.

(a) Plot the locations of the random sample on the $\mathbb{R}^2$ space. Generate the histogram of the sample data and see its distributions.

(b) Calculate the sample mean $\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i$, sample variance $s_{ii} = \frac{1}{N-1}\sum_{j=1}^{N}(x_{ji} - \bar{x}_i)$, $i = 1, 2$, and the sample correlation coefficient $r_{x_1x_2} = \frac{\sum_{j=1}^{N}(x_{j1}-\bar{x}_1)(x_{j2}-\bar{x}_2)}{(N-1)}$. Are the variables of the population statistically independent? (Assume that the measurement error and the noise contaminating the random sample is sufficiently small such that we can judge the statistical independence based on the statistics of random sample.)

(c) Rotate the coordinate through the angle $\theta = \frac{1}{2}\tan^{-1}\left(\frac{2\rho\sqrt{s_{11}s_{22}}}{s_{11}-s_{22}}\right)$. Plot locations and the histograms of the rotated random sample as in (a).

(d) Calculate the sample correlation coefficient of the rotated random sample. Are the variables of the rotated population statistically independent?