STA2001 Probability and Statistics I
Computer-based Exercise 8

Pooi Yuen KAM

Problem 1. The Correlation Coefficient

We study here the joint behavior of two correlated random variables as a function of their correlation coefficient ρ. These correlated random variables can be generated as follows for any desired correlation coefficient.

The output sequence of the random number generator *rand* in MATLAB on your computer is a sequence of independent random variables, each of which is uniformly distributed over the interval $[0, 1]$. Similar command in Python is *numpy.random.uniform*.

Using this generator, you can generate independent samples of a zero-mean random variable X that is uniformly distributed over the interval $[-L, +L]$. Determine analytically the value of L so that $\text{var}[X] = 1$.

In the same way, you can generate independent samples of a zero-mean, uniform random variable Y with $\text{var}[Y] = 1$.

Consider now the random variable Z is given by

$$Z = \rho X + \sqrt{1-\rho^2}Y \tag{1}$$

where X and Y are independent. Show analytically that $E[Z] = 0$ and $\text{var}[Z] = 1$, and that $E[XZ] = \rho$.

Now, we compare the analytical results with numerical results. Using the random number generator, generate N independent samples $(x_i, y_i)_{i=1}^N$ of the random variables (X, Y) as described above. Then, we consider N pairs of samples $(x_i, z_i)_{i=1}^N$ for $\rho = 0, \pm 0.5, \pm 1$, and $N = 1,000$ and $5,000$, where each z_i is obtained according to (1). For each value of ρ, plot the samples $(x_i, z_i)_{i=1}^N$ on the xz-plane together with the line $z = \rho x$. Observe how the samples are scattered around the straight line.

For each value of ρ and N, compute $\frac{1}{N}\sum_{i=1}^N x_i z_i$, and compare it with the value of ρ used. This should give you an appreciation of the role of the correlation coefficient.