# Lab Com2

# Random Variables and Random Processes

## 1. Random Variables

### 1.1 Background

Two random variables X and Y have cumulative distribution functions (CDFs) $F_X(x)$ and $F_Y(y)$, respecitvely. Since there may be an interaction between X and Y, the marginal statistics may not fully describe their behaviour. Therefore, we define a joint CDF as

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y)$$

If the joint CDF is continuous, we can define a joint probability density function,

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$$

Conversely, the joint probability density function may be used to calculate the joint CDF:

$$F_{X,Y}(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u,v) du\, dv$$

The random variables X and Y are said to be independent if and only if

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

If X and Y are independent, then the product of their expectations is the expectation of their product.

$$E[XY] = E[Y]E[Y]$$

The joint distribution contains all the information about X and Y, but it is often difficult to calculate. In many applications, a simple measure of the dependencies of X and Y can be very useful. Three such measures are the correlation, covariance, and the correlation coefficient.

i. Correlation: $E[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X,Y}(x,y) dx\, dy$

ii. Covariance: $E[(X - m_X)(Y - m_Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - m_X)(y - m_Y) f_{X,Y}(x,y) dx\, dy$

iii. Correlation coefficient: $\rho_{XY} = \frac{E[(X-m_X)(Y-m_Y)]}{\sigma_X \sigma_Y} = \frac{E[XY]-m_X m_Y}{\sigma_X \sigma_Y}$

If the correlation coefficient is 0, then X and Y are said to be uncorrelated. Notice that independence implies uncorrelatedness, however the converse is not true.

### 1.2 Instructions

In this part, we will examine the relationship between the scatter plot and correlation coefficient for two random variables.

i. Let X and Y be independent Gaussian random variables, each with mean 0 and variance 1. Generate 1000 i.i.d. (independent identical distributed) samples of X (denoted as $X_1$, $X_2$, . . . , $X_{1000}$), and 1000 i.i.d. samples of Y (denoted as $Y_1$, $Y_2$, . . . , $Y_{1000}$).

ii. Define another random process W = 4X + Y, create samples of W using your generated samples of X and Y in (i), denoted as $W_1$, $W_2$, . . . , $W_{1000}$.

iii. Define another random process Z = 99X + Y, create samples of Z using your generated samples of X and Y in (i), denoted as $Z_1, Z_2, \dots, Z_{1000}$.

iv. Generate scatter plots of the ordered pair of samples $(X_i, Y_i)$, $(X_i, W_i)$, $(Y_i, W_i)$, $(X_i, Z_i)$, and $(Y_i, Z_i)$.

Note: Use the command subplot to plot these scatter plots.

v. Compute the empirical correlation coefficient using your samples using the following equation:

$$\hat{\rho}_{XY} = \frac{\sum_{i=1}^{N}(X_i-\hat{\mu}_X)(Y_i-\hat{\mu}_Y)}{\sqrt{\sum_{i=1}^{N}(X_i-\hat{\mu}_X)^2 \sum_{i=1}^{N}(Y-\hat{\mu}_Y)^2}}$$

vi. Calculate the theoretical correlation coefficient between X and Y ($\rho_{XY}$), between X and W ($\rho_{XW}$), between Y and W ($\rho_{YW}$), between X and Z ($\rho_{XZ}$), between Y and Z ($\rho_{YZ}$), using the equation in section 1.

### 1.3 Instructions

a) Give Matlab program for steps i to v.

b) Display the scatter plots obtained in step iv.

c) Give value of the empirical correlation computed in step v.

d) Give derivations of the theoretical correlation coefficients in step vi.

e) Explain why are the empirical and theoretical correlation coefficients not exactly equal?

f) Explain how the scatter plots are related to correlation coefficients.

## 2. Random Processes

### 2.1 Background

A discrete-time random process $X_n$ is simply a sequence of random variables. So for each n, $X_n$ is a random variable. The autocorrelation is an important function for characterizing the behavior of random processes. If X is a wide-sense stationary (WSS) random process, the autocorrelation is defined by

$$R_X(m) = E[X_n X_{m+n}]; m = \cdots, -1, 0, 1, \dots$$

Note that for a WSS random process:

i. the autocorrelation does not vary with n.

ii. the autocorrelation is an even function of the "lag" value m.

The autocorrelation determines how strong a relation there is between samples separated by a lag value of m.

If X is a sequence of independent identically distributed (i.i.d.) random variables each with zero mean and variance $\sigma_X{}^2$, then the autocorrelation is given by

$$R_X(m) = \sigma_X{}^2 \delta(m)$$

This type of random process is defined as white or white noise.

If a white random process $X_n$ passes through an linear time-invariant (LTI) filter with a response of h(m), it can be shown that the output autocorrelation $R_Y(m)$ is given by

$$R_Y(m) = h(m) * h(-m) * R_X(m)$$

## 2.2 Instructions

In this part, we will generate discrete-time random processes and then analyze their behavior using the correlation measure.

i. Consider a white Gaussian random process $X_n$ with mean 0 and variance 1 as input to the following filter.

$$y(n) = x(n) - x(n-1) + x(n-2)$$

ii. Generate 1000 independent samples of a Gaussian random variable X with mean 0 and variance 1. Filter the samples using the equation in (i). Denote the filtered signal as $Y_i$, i = 1, 2, . . . , 1000.

iii. Draw the following scatter plots: ($Y_i$, $Y_{i+1}$), ($Y_i$, $Y_{i+2}$), ($Y_i$, $Y_{i+3}$) and ($Y_i$, $Y_{i+4}$) (i = 1, 2, . . . , 900). Use subplot command to show these plots side by side.

iv. Calculate the estimated sample autocorrelation defined by

$$\hat{R}_Y(m) = \frac{1}{N-|m|}\sum_{n=0}^{N-|m|+1} Y(n)Y(n+|m|)\,;\ -(N-1) \le m \le N-1,$$

where N is the number of samples of Y.

v. Derive and calculate the theoretical autocorrelation of $Y_n$ using equations given in section 2.1. (This part is not done using Matlab)

vi. Plot both the theoretical autocorrelation $R_Y(m)$, and the sample autocorrelation $\hat{R}_Y(m)$ versus m for −20 ≤ m ≤ 20.

## 2.3 Instructions

a) Give Matlab program for steps ii to vi, except step v.

b) Display the scatter plots obtained from step iii.

c) Give the estimated sample autocorrelation obtained from step iv.

d) Show the derivation and calculation of the theoretical autocorrelation in step v.

e) Display the scatter plot for step vi.

f) Does the equation given in step iv produce a reasonable approximation of the true autocorrelation? Compare the values of m that give maximum $R_Y(m)$ and $\hat{R}_Y(m)$?

*END*