

ECE438 - Laboratory 7: Discrete-Time Random Processes (Week 2)

October 6, 2010

1 Bivariate Distributions

In this section, we will study the concept of a bivariate distribution. We will see that bivariate distributions characterize how two random variables are related to each other. We will also see that correlation and covariance are two simple measures of the dependencies between random variables, which can be very useful for analyzing both random variables and random processes.

1.1 Background on Bivariate Distributions

Sometimes we need to account for not just one random variable, but several. In this section, we will examine the case of two random variables—the so called *bivariate* case—but the theory is easily generalized to accommodate more than two.

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

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y). \quad (1)$$

If the joint CDF is sufficiently “smooth”, we can define a joint probability density function,

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

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

$$F_{X,Y}(x, y) = \int_{-\infty}^y \int_{-\infty}^x f_{X,Y}(s, t) ds dt. \quad (3)$$

Questions or comments concerning this laboratory should be directed to Prof. Charles A. Bouman, School of Electrical and Computer Engineering, Purdue University, West Lafayette IN 47907; (765) 494-0340; bouman@ecn.purdue.edu

The random variables X and Y are said to be *independent* if and only if their joint CDF (or PDF) is a separable function, which means

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) . \quad (4)$$

Informally, independence between random variables means that one random variable does not tell you anything about the other. As a consequence of the definition, if X and Y are independent, then the product of their expectations is the expectation of their product.

$$E[XY] = E[X]E[Y] \quad (5)$$

While the joint distribution contains all the information about X and Y , it can be very complex and 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*.

- Correlation

$$E[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X,Y}(x, y) dx dy \quad (6)$$

- Covariance

$$E[(X - \mu_X)(Y - \mu_Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f_{X,Y}(x, y) dx dy \quad (7)$$

- Correlation coefficient

$$\rho_{XY} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} = \frac{E[XY] - \mu_X \mu_Y}{\sigma_X \sigma_Y} \quad (8)$$

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 Samples of Two Random Variables

In the following experiment, we will examine the relationship between the scatter plots for pairs of random samples (X_i, Z_i) and their correlation coefficient. We will see that the correlation coefficient determines the shape of the scatter plot.

Let X and Y be independent Gaussian random variables, each with mean 0 and variance 1. We will consider the correlation between X and Z , where Z is equal to the following:

1. $Z = Y$
2. $Z = (X + Y)/2$
3. $Z = (4 * X + Y)/5$
4. $Z = (99 * X + Y)/100$

Notice that since Z is a linear combination of two Gaussian random variables, Z will also be Gaussian.

Use Matlab to generate 1000 i.i.d. samples of X , denoted as $X_1, X_2, \dots, X_{1000}$. Next, generate 1000 i.i.d. samples of Y , denoted as $Y_1, Y_2, \dots, Y_{1000}$. For each of the four choices of Z , perform the following tasks:

1. Use equation (8) to analytically calculate the correlation coefficient ρ_{XZ} between X and Z . Show all of your work. Remember that independence between X and Y implies that $E[XY] = E[X]E[Y]$. Also remember that X and Y are zero-mean and unit variance.
2. Create samples of Z using your generated samples of X and Y .
3. Generate a scatter plot of the ordered pair of samples (X_i, Z_i) . Do this by plotting points $(X_1, Z_1), (X_2, Z_2), \dots, (X_{1000}, Z_{1000})$. In order to plot points without connecting them with lines, use the `plot` command with the `'.'` format.

`plot(X,Z, '.')`

Use the command `subplot(2,2,n)` ($n=1,2,3,4$) to plot the four cases for Z in the same figure. Be sure to label each plot using the `title` command.

4. Empirically compute an estimate of the correlation coefficient using your samples X_i and Z_i and the following formula.

$$\hat{\rho}_{XZ} = \frac{\sum_{i=1}^N (X_i - \hat{\mu}_X)(Z_i - \hat{\mu}_Z)}{\sqrt{\sum_{i=1}^N (X_i - \hat{\mu}_X)^2 \sum_{i=1}^N (Z_i - \hat{\mu}_Z)^2}}$$

INLAB REPORT:

1. Hand in your derivations of the correlation coefficient ρ_{XZ} along with your numerical estimates of the correlation coefficient $\hat{\rho}_{XZ}$.
2. Why are ρ_{XZ} and $\hat{\rho}_{XZ}$ not exactly equal?
3. Hand in your scatter plots of (X_i, Z_i) for the four cases. Note the theoretical correlation coefficient ρ_{XZ} on each plot.
4. Explain how the scatter plots are related to ρ_{XZ} .

2 Autocorrelation for Filtered Random Processes

In this section, we will generate *discrete-time random processes* and then analyze their behavior using the correlation measure introduced in the previous section.

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

$$r_{XX}(m) = E[X_n X_{n+m}] \quad m = \dots, -1, 0, 1, \dots \quad (9)$$

Note that for a WSS random process, the autocorrelation does not vary with n . Also, since $E[X_n X_{n+m}] = E[X_{n+m} X_n]$, the autocorrelation is an even function of the “lag” value m .

Intuitively, the autocorrelation determines how strong a relation there is between samples separated by a lag value of m . For example, if X is a sequence of independent identically distributed (i.i.d.) random variables each with zero mean and variance σ_X^2 , then the autocorrelation is given by

$$\begin{aligned} r_{XX}(m) &= E[X_n X_{n+m}] \\ &= \begin{cases} E[X_n]E[X_{n+m}] & \text{if } m \neq 0 \\ E[X_n^2] & \text{if } m = 0 \end{cases} \\ &= \sigma_X^2 \delta(m) . \end{aligned} \quad (10)$$

We use the term *white* or *white noise* to describe this type of random process. More precisely, a random process is called *white* if its values X_n and X_{n+m} are uncorrelated for every $m \neq 0$.

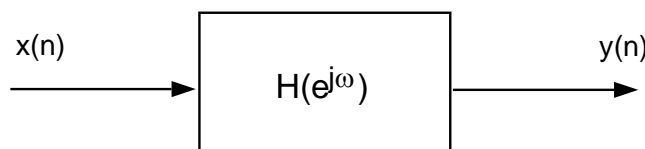


Figure 1: A new LTI system diagram

If we run a white random process X_n through an LTI filter as in figure 1, the output random variables Y_n may become correlated. In fact, it can be shown that the output autocorrelation $r_{YY}(m)$ is related to the input autocorrelation $r_{XX}(m)$ through the filter’s impulse response $h(m)$.

$$r_{YY}(m) = h(m) * h(-m) * r_{XX}(m) \quad (11)$$

2.2 Experiment

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

$$y(n) = x(n) - x(n-1) + x(n-2) \quad (12)$$

Calculate the theoretical autocorrelation of Y_n using (10) and (11). Show all of your work.

Generate 1000 independent samples of a Gaussian random variable X with mean 0 and variance 1. Filter the samples using (12). We will denote the filtered signal Y_i , $i = 1, 2, \dots, 1000$.

Draw 4 scatter plots using the form *subplot(2,2,n)*, ($n = 1, 2, 3, 4$). The first scatter plot should consist of points, (Y_i, Y_{i+1}) , ($i = 1, 2, \dots, 900$). Notice that this correlates samples that are separated by a lag of “1”. The other 3 scatter plots should consist of the points (Y_i, Y_{i+2}) , (Y_i, Y_{i+3}) , (Y_i, Y_{i+4}) , ($i = 1, 2, \dots, 900$), respectively. What can you deduce about the random process from these scatter plots?

For real applications, the theoretical autocorrelation may be unknown. Therefore, $r_{YY}(m)$ may be estimated by the *sample autocorrelation*, $r'_{YY}(m)$ defined by

$$r'_{YY}(m) = \frac{1}{N - |m|} \sum_{n=0}^{N-|m|-1} Y(n)Y(n + |m|) \quad - (N - 1) \leq m \leq N - 1 \quad (13)$$

where N is the number of samples of Y .

Use Matlab to calculate the sample autocorrelation of Y_n using (13). Plot both the theoretical autocorrelation $r_{YY}(m)$, and the sample autocorrelation $r'_{YY}(m)$ versus m for $-20 \leq m \leq 20$. Use *subplot* to place them in the same figure. Does equation (13) produce a reasonable approximation of the true autocorrelation?

INLAB REPORT:

For the filter in equation (12),

1. Show your derivation of the theoretical output autocorrelation, $r_{YY}(m)$.
2. Hand in the four scatter plots. Label each plot with the corresponding theoretical correlation, using $r_{YY}(m)$. What can you conclude about the output random process from these plots?
3. Hand in your plots of $r_{YY}(m)$ and $r'_{YY}(m)$ versus m . Does equation (13) produce a reasonable approximation of the true autocorrelation? For what value of m does $r_{YY}(m)$ reach its maximum? For what value of m does $r'_{YY}(m)$ reach its maximum?
4. Hand in your Matlab code.

3 Correlation of Two Random Processes

3.1 Background

The *cross-correlation* is a function used to describe the correlation between two separate random processes. If X and Y are jointly WSS random processes, the cross-correlation is defined by

$$c_{XY}(m) = E[X_n Y_{n+m}] \quad m = \dots, -1, 0, 1, \dots \quad (14)$$

Similar to the definition of the sample autocorrelation introduced in the previous section, we can define the *sample cross-correlation* for a pair of data sets. The *sample cross-correlation* between two finite random sequences X_n and Y_n is defined as

$$c'_{XY}(m) = \frac{1}{N-m} \sum_{n=0}^{N-m-1} X(n)Y(n+m) \quad 0 \leq m \leq N-1 \quad (15)$$

$$c'_{XY}(m) = \frac{1}{N-|m|} \sum_{n=|m|}^{N-1} X(n)Y(n+m) \quad 1-N \leq m < 0 \quad (16)$$

where N is the number of samples in *each* sequence. Notice that the cross-correlation is not an even function of m . Hence a two-sided definition is required.

Cross-correlation of signals is often used in applications of sonar and radar, for example to estimate the distance to a target. In a basic radar set-up, a zero-mean signal $X(n)$ is transmitted, which then reflects off a target after traveling for $D/2$ seconds. The reflected signal is received, amplified, and then digitized to form $Y(n)$. If we summarize the attenuation and amplification of the received signal by the constant α , then

$$Y(n) = \alpha X(n-D) + W(n) \quad (17)$$

where $W(n)$ is additive noise from the environment and receiver electronics.

In order to compute the distance to the target, we must estimate the delay D . We can do this using the cross-correlation. The cross-correlation c_{XY} can be calculated by substituting (17) into (14).

$$\begin{aligned} c_{XY}(m) &= E[X(n)Y(n+m)] \\ &= E[X(n)(\alpha X(n-D+m) + W(n+m))] \\ &= \alpha E[X(n)X(n-D+m)] + E[X(n)]E[W(n+m)] \\ &= \alpha E[X(n)X(n-D+m)] \end{aligned}$$

Here we have used the assumptions that $X(n)$ and $W(n+m)$ are uncorrelated and zero-mean. By applying the definition of autocorrelation, we see that

$$c_{XY}(m) = \alpha r_{XX}(m-D) \quad (18)$$

Because $r_{XX}(m-D)$ reaches its maximum when $m=D$, we can find the delay D by searching for a peak in the cross correlation $c_{XY}(m)$. Usually the transmitted signal $X(n)$ is designed so that $r_{XX}(m)$ has a large peak at $m=0$.

3.2 Experiment

Down load [radar.mat](#)

Using (15) and (16), write a Matlab function $\mathbf{C}=\text{CorR}(X,Y,m)$ to compute the sample cross-correlation between two discrete-time random processes, X and Y , for a single lag value m .

To test your function, generate two length 1000 sequences of zero-mean Gaussian random variables, denoted as X_n and Z_n . Then compute the new sequence $Y_n = X_n + Z_n$. Use CorR to calculate the sample cross-correlation between X and Y for lags $-10 \leq m \leq 10$. Plot your cross-correlation function.

INLAB REPORT:

1. Submit your plot for the cross-correlation between X and Y . Label the m -axis with the corresponding lag values.
2. Which value of m produces the largest cross-correlation? Why?
3. Is the cross-correlation function an even function of m ? Why or why not?
4. Hand in the code for your CorR function.

Next we will do an experiment to illustrate how cross-correlation can be used to measure time delay in radar applications. Download the MAT file [radar.mat](#) and load it using the command `load radar`. The vectors *trans* and *received* contain two signals corresponding to the transmitted and received signals for a radar system. First compute the autocorrelation of the signal *trans* for the lags $-100 \leq m \leq 100$. (Hint: Use your CorR function.)

Next, compute the sample cross-correlation between the signal *trans* and *received* for the range of lag values $-100 \leq m \leq 100$, using your CorR function. Determine the delay D .

INLAB REPORT:

1. Plot the transmitted signal and the received signal on a single figure using *subplot*. Can you estimate the delay D by a visual inspection of the received signal?
2. Plot the sample autocorrelation of the transmitted signal, $r'_{XX}(m)$ vs. m for $-100 \leq m \leq 100$.
3. Plot the sample cross-correlation of the transmitted signal and the received signal, $c'_{XY}(m)$ vs. m for $-100 \leq m \leq 100$.
4. Determine the delay D from the sample correlation. How did you determine this?

References

- [1] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed., McGraw-Hill, New York, 1991.