# Stochastic Processes, Session 6 — Group Work
## WSS Processes, the Autocorrelation Function and its Estimation

Carles Navarro and Troels Pedersen, Aalborg University, September, 2017

## Exercise Session 6

Go through the exercises below. Allow yourself the time to reflect over your results and discuss them with other students! Use the book for inspiration and for further information.

### 6.1 Estimation of the Mean in WSS processes

One of the properties of WSS processes is that they have a mean function which is constant, i.e. it does not depend on the time index. This suggests the idea that, if only one realization of a WSS process is available, one may attempt to estimate the mean of the process by averaging the different temporal samples of the process. For instance, consider the WSS process:

$$X(n) = U(n) - \frac{1}{2}U(n-1), \quad n = 1, 2, \ldots \tag{1}$$

with $U(n) \overset{i.i.d.}{\sim} \mathcal{N}(3, 9)$. Verify that the process is WSS by computing its mean and ACF.

If we can only observe $N$ samples of a single realization of the process, one may try to estimate its mean by:

$$\hat{\mu}_X = \frac{1}{N}\sum_{n=1}^{N} X(n). \tag{2}$$

a) Generate a realization of the process in (1) with a large number of samples $M$.

b) Next, use $N$ consecutive samples of the process to estimate its mean, with $N = 1, 2, \ldots, M$. Plot the estimates of the mean as a function of the number of samples used. Plot, on top of that, the true mean value for comparison.

c) How does the estimation error behave as a function of the number of used samples? How many sample do you need to use to get *reasonably accurate* results?

Next consider a second random process, defined as

$$Y(n) = A + W(n), \quad n = 1, 2, \ldots \tag{3}$$

with $W(n) \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$, and $A$ being a random variable distributed as $A \sim \mathcal{U}(0, 3)$.

d) Compute the mean function and ACF of $Y(n)$. Is it also WSS?

e) Next, repeat the same steps as for the process in (1). What do you observe?

f) Now, generate multiple realizations of $Y(n)$ and estimate its mean function by averaging over the different realizations, and not over the time index. Do the results make more sense now?

The process $X(n)$ in (1) is a process that is *ergodic in the mean.* This means that the sample mean estimator in (2) tends to the true mean of the process as the number of temporal samples $N$ grows large, i.e.

$$\lim_{N\to\infty} \hat{\mu}_X = \mu_X,$$

which is satisfied when the variance of the sample mean estimator tends to zero as $N$ approaches infinity:

$$\lim_{N\to\infty} \mathbb{V}\mathrm{ar}(\hat{\mu}_X) = 0.$$

When this is the case, one may use *temporal averaging* to estimate the mean of the process and be confident that, with enough temporal samples, the estimator will converge to the correct result.

Unfortunately, not all WSS processes are ergodic in the mean, as you have experienced with the process $Y(n)$ in (3). When processes are not ergodic, one needs to average over different realizations of the process –also known as *ensemble averaging*– to safely estimate the mean value.

## 6.2 Checking Wide-Sense Stationarity

a) Write a program to generate and plot multiple realizations of each of the processes:

   1) Moving average process:

   $$X(n) = U(n) - \frac{1}{2}U(n-1) + \frac{1}{4}U(n-2), \qquad U(n) \overset{i.i.d.}{\sim} \mathcal{N}(0,1), \quad n \in \mathbb{Z}.$$

   2) Randomly phased cosine:

   $$Y(t) = \cos(t + \Theta), \qquad \Theta \sim \mathcal{U}(-\pi, \pi) \quad t \in \mathbb{R}.$$

   3) Random walk:

   $$Z(n) = \sum_{i=1}^{n} B(i), \quad n = 1, 2, \ldots$$

   with i.i.d. samples $B(i)$, $i = 1, 2, \ldots$ drawn according to the pmf

   $$B(i) \overset{\text{iid}}{\sim} p_B(b) = \begin{cases} \frac{1}{2}, & b = +1 \\ \frac{1}{2}, & b = -1 \end{cases}$$

b) Use the realizations generated in your program to estimate the mean and variance functions of the above three processes. Based on the estimates obtained, what can you say about the stationarity / non-stationarity of the processes?

c) Next, derive the necessary results to check whether each of the processes is or is not WSS. Compare your calculations with the estimates obtained from your program.

## 6.3 Estimation of the Autocorrelation Function

Next we will attempt to estimate the autocorrelation function (ACF) of a WSS process. Assume that we can observe the samples $n = 1, \ldots, N$ of a realization of WSS process $X(n)$. Then, the ACF of the process can be estimated as:

$$\hat{R}_X(k) = \frac{1}{N-|k|} \sum_{n=1}^{N-|k|} X(n)X(n+k), \quad |k| \leq N-1. \tag{4}$$

The above estimator is called the *unbiased autocorrelation estimator*, and provides estimates of the ACF for lags $k$ between $-N+1$ and $N-1$. We will learn more about this and other ACF estimators in the next lecture of the course.

a) Write a program that estimates the ACF of a WSS process from a realization consisting of $N$ samples.

b) Apply the estimator to an $N$-samples long realization of those processes from the previous exercise which are WSS. Compare the estimates obtained with the true ACF that you derived before.

c) For which values of $k$ are the estimates of the ACF most accurate? How does the estimates' accuracy change when the length $N$ of the realizations is increased or decreased? Can you, intuitively, guess why?

To finish up the exercise, have a look at the documentation of the MATLAB command `xcorr`. How should the command be applied — meaning with which options or parameters – so that it produces the same result as the estimator in (4)? Check that, indeed, you get the same results.

## 6.4 Autoregressive Random Process

A random process defined as

$$X(n) = aX(n-1) + U(n), \quad \infty < n < \infty \tag{5}$$

with $|a| < 1$ and $U(n) \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$ is called an autoregressive (AR) process of order 1. The name *autoregressive* is due to the regression of $X(n)$ onto $X(n-1)$. AR processes are WSS processes, as you will show next.

a) Compute the mean function of the process.[1]

b) Compute the ACF of the process[2] and plot it for different values of $a$. How do you expect the realizations of the process to look like, depending on the value chosen for $a$?

Next, implement a script that can generate multiple realizations of $\{X(n)\}$ for times $n = 1, 2, \dots$ using (5). For that, assume that $X(0) = 0$.

c) Plot realizations of the process for different values of $a$. Does the behavior fit what you were expecting from the ACF?

d) Estimate the mean and variance functions of the process using multiple realizations. Do your results match what you would expect from a WSS process?

e) How should the process be initialized –i.e., how should $X(0)$ be distributed– so that the process is indeed stationary? Generate realizations with the correct initialization and check that they now have the right mean and variance functions.

f) Finally, estimate the ACF from realizations generated with different values of $a$. Do the estimates match the ACF you derived earlier?

[1] For this, it may be useful to express the process as $X(n) = \sum_{m=0}^{\infty} a^m U(n-m)$

[2] You may find the following result useful: $\sum_{m=0}^{\infty} cr^m = \frac{c}{1-r}$, for $|r| < 1$.