Stochastic Processes, Session 8 — Group Work
Estimation of ACF and PSD: Sample Autocorrelation Function and Periodogram

Carles Navarro Manchón and Troels Pedersen
Aalborg University, September 2017

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.

8.1 ACF and PSD of two Moving Average Processes

Consider the following WSS random processes (both special cases of a MA(1) process):

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

- $$Y(n) = W(n) - W(n-1), \quad n = 1, 2, \ldots$$
  with
  $$W(n) \overset{i.i.d.}{\sim} \mathcal{N}(0,1), \quad n \in \mathbb{Z}.$$

a) Can you guess which of the two processes will behave more "wildly" – larger variations from one sample to the next – and which one will vary in a "smoother" manner? Discuss your intuition about it with your group mates. Write down your conclusions and the reasoning behind them.

b) As we know, a common way of characterizing the behavior of a WSS random process is to compute its autocorrelation function (ACF) and its power spectral density (PSD). Do so for the two processes above.

c) Plot (by hand or computer) ACF and the PSD of each of the processes. Compare the results for both processes, and discuss whether the obtained results confirm your initial guess.

d) Write a script that generates multiple realizations of each of the processes. Make sure the number of realizations and the length of each of them is programmed as a parameter, so that you can easily change them later on.

e) Plot several realizations of the two processes and discuss whether the realizations you observe are in line with your earlier guesses.

8.2 The (Biased and Unbiased) Sample Autocorrelation Function and Periodogram

Write a program to compute the biased and unbiased sample autocorrelation functions [1] of a realization of a process. Use it to estimate the ACF of the two processes from the previous exercise. You may rely on existing algorithms (such as `xcorr` in Matlab), but be sure to read their documentation first. Plot the results of both estimators, along with the ACF you derived before. Discuss the following questions with your group mate(s):

a) Which of the two ACF estimators seems to give the most accurate results? Does this contradict what you would expect, given the name (biased/unbiased sample ACF) of the estimators?

b) Derive the expected value of both ACF estimators, and discuss with your group mates how it affects the estimates of the ACF of these two particular processes.

c) Vary the length of the realizations you are using to estimate the ACF. Does increasing the length of the realization help improve the accuracy of the estimates?

d) Finally, generate several realizations of the same process and plot their respective sample autocorrelation functions. Is there any way in which you can use several realizations of the process in order to improve the accuracy of your ACF estimates?

e) Extend your program to compute the periodogram (**Do not use the MATLAB routine, program it yourselves**). Next, estimate the PSD of the two processes of Exercise 8.1. Try this for several different realizations. How do the periodograms compare to each other?

f) Next, vary the number of samples of the realizations used to compute the periodogram. Does increasing the length of the realization help improve the accuracy of the estimates? How about the frequency resolution?

g) In order to understand the effect observed above, derive the expected value of the periodogram of an N-sample long realization of a WSS process $\{Z(n)\}$ with PSD $S_Z(f)$. As you can check, the result depends on the PSD $S_Z(f)$, which is convolved with the Fourier transform of the Bartlett window: the Fejér kernel[2]. Plot the Fejér kernel for different values of N and discuss the result with your groupmates.

h) Now, we will apply an averaged periodogram[3] to improve the accuracy of the estimates of the PSD. Compute the periodograms of several realizations of the

[1] See Section 6.1.2 in [LN].
[2] See [LN, Section 6.1.3].
[3] See Section 17.7 in [Kay].

same process and average them. Compare again your results to the true PSD of the process. How many periodograms do you need to average to obtain results that visually resemble the PSD of the process?

8.3 Working with Real-Life Data: Wolfer Sunspot Data

Sunspots are temporary phenomena caused by intense magnetic activity which create areas of reduced temperature on the surface of the Sun. Their name comes from the fact that such areas appear as dark spots on the surface of the Sun, which can be observed from Earth with the aid of a telescope. Their occurrence is correlated with the emission of solar flares: the emission of massive clouds of electrons, ions and atoms through the corona of the sun into space. If the solar flares particles reach the earth, they can disrupt all radio communications. They are also a major concern for spacecraft electronics. The observation of the number of sunspots is used to predict the possibility of large solar flares.

In 1848 in Zurich, Rudolf Wolf started computing and recording the number of sunspots present in the surface of the Sun every year. The resulting time-series is usually referred to as the Wolfer sunspot data, and can be loaded in Matlab by using the command: `load sunspot.dat`.[4]

a) Plot the sunspot number as a function of time. What kind of behavior do you observe? How do you expect this effect to manifest itself in the ACF and the PSD of the process?

b) Use your previously programmed functions/scripts to compute the sample ACF and the periodogram of the data. First, though, subtract the mean of the data, so that the resulting set of samples have zero mean.

c) At which frequencies do you observe the peaks of the obtained periodogram? How can you relate this frequencies to the temporal behavior observed in the data set?

d) Next, compute an averaged periodogram. This time, only one realization of the process is available, so you will need to slice the data set into smaller sets, and average the corresponding periodograms. What do you gain and what do you lose by doing so?

8.4 The Blackman-Tukey Estimator

The Blackman-Tuckey estimator[5] of the PSD consists of applying a lag window to the biased sample autocorrelation function before taking its Fourier transform. The estimate of the PSD obtained in this way corresponds to a filtered version of the

[4] A .mat file is available at our Moodle site if needed.

[5] See [LN, Section 6.1.3].

periodogram, which is smoothed by the frequency response of the chosen lag window. Typically, the obtained estimates have lower variance than the periodogram, although spectral resolution is lost in this process. Different windows can be applied, each with its own advantages and pitfalls.

MATLAB provides built-in functions that generate the most commonly used lag windows of a desired length: check the command `window` in the help browser to inspect the different options provided and the usage of the command. You may also use the command `wvtool` to visualize the time and frequency representations of each of the windows.

a) Use the command `wvtool` to inspect the time and frequency responses of several lag windows. Try, for instance, with a Blackman window, a Hamming window, and Gaussian windows with different width. (Use the MATLAB help to learn the usage of the commands!). Discuss advantages or disadvantages of using each of the windows with your group mates.

b) Modify your script to compute the periodogram, so that you add the option of including a lag window to smooth the obtained PSD estimates. Next, apply different types of windows to estimate the PSD of the Wolfer sunspot data.

c) Try now with a Gaussian window. You can increase and decrease the width of the window by changing the corresponding parameter in the `window` command. What do you gain and what do you lose by increasing or decreasing the window width?