Stochastic Processes, Session 11 – Group Work
Parametric Estimation of Random Processes

Carles Navarro Manchón and Troels Pedersen
Aalborg University, October, 2016

Session 11

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

11.1 Fitting ARMA Models to Internet Traffic Data

In this exercise we are going to fit MA(q), AR(p) and ARMA(p, q) processes to a data series. The data we consider is the hourly Internet traffic (in bits) measured in the United Kingdom academic network backbone during Christmas 2004.[1] The data can be downloaded from the moodle page of the course in `.mat` format.

1. Plot the data. From the plot, you can observe that the data does not have zero mean. Since ARMA processes have zero mean, it would be convenient for the data to hold the same property. Subtract the sample mean from the data and plot it again.

2. Calculate and plot the sample autocorrelation function and the periodogram of the data. The data has a strong periodicity. Discuss how this shows in the estimated autocorrelation function and the periodogram.

3. Apply differencing to remove periodicity (See [LN, p. 6-18]). Repeat the procedure in item 2 for the differenced data. Compare the results to those you obtained before.

4. Now, we will try to find the parameters of an AR(3) model that fits the data (both the raw data and the differenced data). This can be accomplished by using the Yule-Walker equations (see [LN, pp. 2-7 and 2-8], which for an AR(P) process have the general form

$$R_X(k) = \sum_{i=1}^{P} \phi_i R_X(k-i) + \sigma_Z^2 \delta(k), \quad k \geq 0. \tag{1}$$

   Since we want to find the parameters of an AR(3) process, we need to estimate four parameters: ϕ_1, ϕ_2, ϕ_3, and σ_Z^2.

[1] Source: Time Series Data Library.

- Write up the Yule-Walker equations in (1) for $k = 0, 1, 2, 3$ and $P = 3$. Now you have a system with 4 equations, in which the values of the ACF for $|k| \leq 3$ can be plugged in from the sample autocorrelation function obtained earlier. How many unknown variables do you have in this system of equations?
- Solve the system of equations to find the parameters of the desired AR process. (Hint: Write it up in matrix-vector form and solve it with MATLAB/Python.)
- Plot the PSD of the fitted models and compare the resulting estimated power spectra to the periodograms.

5. Next, we will fit MA and ARMA models to the data. To do so, we will use the MATLAB routine `armax`. Read the MATLAB documentation about `armax` and discuss how to use it to fit ARMA models to the data. If in doubt, ask for the teachers' help. Those of you using Python can find similar routines. Check http://statsmodels.sourceforge.net/stable/tsa.html.

6. Fit an MA(3) and an ARMA(3,3) model using the `armax` routine or equivalent. Plot the resulting PSDs and compare them to the periodograms and the AR(3) PSDs.

7. Increase the order of the ARMA models used to fit the data and compare the estimated models. Read about "Occams Razor" – e.g. in Wikipedia – and discuss with your partners about how this concept relates to the selection of the ARMA model order.

8. Next, use the estimated models to generate synthetic realizations of the data. Remember to add the mean of the data that you subtracted earlier. Compare those realizations with the original data set, and discuss differences and similarities. Can you relate the trends that you observe in the realizations to the estimated PSDs you plotted earlier?