RAN Signal M2 MN Orsay-Gif (2024-2025)
Enseignant : Mohammed Nabil EL-KORSO
Contact : `mohammed.el-korso@universite-paris-saclay.fr`

Processus aléatoires

(bruit blanc, prédiction, corrélation, densité spectrale de puissance)

Application : traitements de signaux EEG

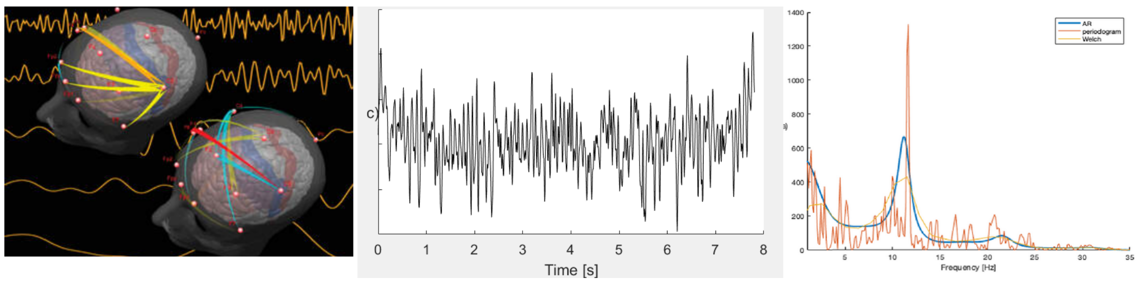

Figure 1: Courtesy of *Practical Biomedical Signal Analysis Using MATLAB By K. J. Blinowska, J. Zygierewicz 2022*

Objectif du TP

The aim of this practical course is to put into practice the theoretical notions learned in class. In particular, stationarity, correlation/variance calculation, signal modeling via AR (auto-regressive) processes, spectral density calculation, etc.

- The first two parts of the tutorial will involve applying these notions to academic toy examples in order to validate the concept.
- The last part will apply these notions in a real data context. Specifically, we will process EEG data to detect a particular rhythm.
- At the end of this tutorial, you'll be able to process an EEG signal and reproduce the Fig. 1

A few basic notions about the targeted application:

1. *"Electroencephalography (EEG) is a method of cerebral exploration that measures the brain's electrical activity using electrodes placed on the scalp, often represented in the form of a trace called an electroencephalogram. Comparable to the electrocardiogram, which studies the functioning of the heart, the EEG is a painless, non-invasive examination that provides information on the neurophysiological activity of the brain over time, and of the cerebral cortex*

in particular, either for diagnostic purposes in neurology, or for research in cognitive neuroscience. The electrical signal at the origin of the EEG is the result of the summation of synchronous postsynaptic potentials from a large number of neurons. On an EEG trace, it is possible to identify rhythmic cerebral electrical activity. These cerebral rhythms are classified according to their frequency, making it possible, for example, to identify or characterize psychological states in fundamental neuroscience, or pathological states in clinical neurology." Wiki

2. In this tutorial, we will focus on the alpha rhythm. "*The alpha rhythm is a brain rhythm, i.e. an EEG oscillation resulting from the brain's electrical activity, with a frequency of between 7.5 and 12.5 Hz. The alpha rhythm appears when the person being recorded closes their eyes and relaxes. Because of their high amplitude, alpha waves were the first signals identified by the inventor of the EEG, Hans Berger.*" Wiki. This type of signal is therefore a good exercise to start processing EEG signals before moving on to more complex phenomena characterizing psychological or pathological states.

Part 1: Prediction, correlation, white noise

Aim of part 1:

The primary aim of this section is to give you the tools you need to model and analyze your signals and observations in the time domain. Fine modeling will enable you to extract information and/or parameters of interest from observations (detection of certain phenomena, estimation of certain parameters, prediction of future values, etc.).

1 Theoretical study and examples

In order to carry out simulations on real data, theoretical preparation is required. A time series (or time signal) denoted x_t is modeled by an autoregressive process of order p, denoted AR(p), if

$$x_t = -\sum_{i=1}^{p} a_i x_{t-i} + \epsilon_t \tag{1}$$

with $epsilon_t$ white noise. Thus, we can see that an AR process is a regression model for time series in which the series is explained by its past values. In the following, we'll assume that ϵ_t has zero mean, as does x_t.

1.1 Parameter estimation

The aim of this section is to find a way of estimating the parameters describing an AR(p) process from x_t measurements.

1. **[QT]** *Calculating autocorrelation :* For an AR(p) given by (1) with ϵ_t a white noise of mean zero and variance σ_ϵ^2, show that the autocorrelation, defined by $R_m = E\left\{x_t x_{t-m}\right\}$, is expressed as follows

$$R_m = -\sum_{i=1}^{p} a_i R_{m-i} + \sigma_\epsilon^2 \delta_m \tag{2}$$

   with $\delta_m = 1$ if $m = 0$ and $\delta_m = 0$ otherwise.

2. **[QT]** *AR(p) coefficient expressions:*

   (a) Show that for $m > 0$

$$\boldsymbol{R}\boldsymbol{a} = -\boldsymbol{r} \tag{3}$$

   with $\boldsymbol{a} = [a_1, a_2 \dots a_p]^T$, $\boldsymbol{r} = [R_1, R_2 \dots R_p]^T$ et

$$\boldsymbol{R} = \begin{bmatrix} R_0 & R_{-1} & \dots & R_{1-p} \\ R_1 & R_0 & R_{-1} & \\ \vdots & & & \\ R_{p-1} & \dots & \dots & R_0 \end{bmatrix}$$

(b) Show that

$$\sigma_\epsilon^2 = R_0 + \sum_i^p a_i R_{-i} \tag{4}$$

3. **[QT]** *Estimation of AR(p) coefficients:* To estimate $\boldsymbol{a}$ using (3), we need to know the autocorrelations R_m. However, R_m is unknown but a consistent estimator of the autocorrelation $R_m = E\left\{x_t x_{t-m}\right\}$ is given byfootnoteOptional question: Show that $\hat{R}_m$ is a consistent estimator of R_m

$$\hat{R}_m = \frac{1}{T}\sum_{t=1}^{T} x_t x_{t-m}.$$

Deduce a way to estimate[1] $\boldsymbol{a}$ as

$$\hat{\boldsymbol{a}} = -\hat{\boldsymbol{M}}\hat{\boldsymbol{r}} \tag{5}$$

Given the expression of $\hat{\boldsymbol{M}}$ and $\hat{\boldsymbol{r}}$.

1.2 Numerical simulations

1. Matlab offers various toolboxes for estimating the parameters of an AR(p) model. These include[2]

   - `aryule` based on (5), namely, Yule-Walker/
   - `arburg` based on Burg estimator.
   - `arcov` based on matrix covariance estimation.
   - `armcov` which is based on the modified covariance method.

   Manipulation:

   - Generate several realizations of an AR(5) with $\boldsymbol{a} = [0.2, -0.5, -0.3, 0.1, 0.2]^T$, $\sigma_\epsilon^2 = 10^2$ and $T = 200$. Plot them on the same figure. Note that each realization is different, but they all share the common characteristic of oscillation structures with a similar frequency.
   - Estimate $\boldsymbol{a}$ with `aryule`[3].

2. *Model order estimation:* In reality, the order of the model, p, is not known in practice. It must therefore be estimated. Given the model (1), we might consider choosing a large p in order to fit our data as closely as possible, making the residual error small. That said, increasing the order of the p model will increase the number of parameters to be estimated (as the size of the $\boldsymbol{a}$ vector increases in turn). This has the effect of increasing the variance of our estimates (and therefore low confidence in our estimates). Various criteria exist to find such a compromise. For example, the AIC criterion (Akaike information criterion) penalizes the increase in order p and the residual error as follows

$$AIC(p) = \frac{2p}{T} + \log \sigma_\epsilon \tag{6}$$

[1] Question facultative : Discuter de la consistance de votre estimateur

[2] Question facultative : Expliquez le fondement théorique de chaque méthode. Vous pouvez vous servir du `Help` de Matlab.

[3] Question facultative : Tracez, en fonction de σ_ϵ^2 l'erreur quadratique moyenne des estimées données par $\left\|E\left\{(\hat{\boldsymbol{a}} - \boldsymbol{a})(\hat{\boldsymbol{a}} - \boldsymbol{a})^T\right\}\right\|^2$. Conclure quant à la consistance de votre estimateur.

(a) **[QT]** Estimate $sigma_\epsilon$ using (4). Note that the Matlab function `aryule` gives you an estimate of σ_ϵ.

(b) Take your code and plot AIC(p) for different values of p. Estimate $\sigma_{epsilon}$ using `aryule`, apply (6) and find the minimum. Does this correspond to the actual order of your model?

Part 2: Power spectral density, transfer function

Aim of Part 2:

The temporal representation of some signals is not sufficient to extract the desired information. Let's take the example of the EEG signal shown in Fig. 1. Cerebral rhythms are classified according to their frequency and can be used, for example, to identify or characterize psychological states in fundamental neuroscience, or pathological states in clinical neurology. The aim of this section is therefore to take advantage of temporal modeling (1) to better describe the spectral representation.

2 Power spectral density: Calculation and representation

1. **[QT]** Using eq (1) and the Fourier transform, show that

$$X(f) = H(f)E(f) \tag{7}$$

   with $X(f)$ and $E(f)$ respectively the Fourier transform of x_t and ϵ_t. Show that the frequency transfer function $H(f)$ is as follows

$$H(f) = \frac{1}{\sum_{k=0}^{p} a_k \exp^{-j2\pi kf}} \tag{8}$$

2. **[QT]** Show that the spectral density $S(f) = X(f)X(f)^*$, is written as follows

$$S(f) = \frac{\text{Const}}{\left|\sum_{k=0}^{p} a_k \exp^{-j2\pi kf}\right|^2} \tag{9}$$

3. *Manipulation:*

   (a) Generate an AR(3) signal with $\boldsymbol{a} = [0.2, 0.5, 0.1]^T$, for a sampling frequency of 1 Hz and $T = 200$.

   (b) Matlab's `pyulear` function gives you an estimate of the power spectral density. Use this function to plot the spectral density.

   (c) Superimpose on this plot the function $S(f)$ calculated in (9).

   (d) Repeat the experiment several times, then conclude with the use of `pyulear`.

Part 3: Application to EEG signal processing

Objectif of Part 3:

After 3 hours of work, you're now ready to use your knowledge for a real-world application of interest: Using AR modeling and spectral representation to detect specific brain rhythms.

1. Create a new M-file script (don't forget to start with `clear all` and `close all`)
2. Download the eeg1.mat data using the `load` function. What does this database contain?
3. Plot the signal observed by the EEG electrode (remember to specify the axes: nature and unit).
4. We're going to use an AR[4]odelYou can use another model, but you'll have to redo parts 1 and 2 ;). Plot the criterion AIC(p) for $p = 1, \ldots, 50$. Then estimate the order of the $\hat{p}$ model and the $\hat{boldsymbola}$ parameters.
5. Use the `pyulear` function to plot the power spectral density.
6. Superimpose the periodogram on this plot using the `periodogram`[5].
7. Compare the two and conclude whether AR modeling is worthwhile.
8. Does the patient have an alpha rhythm?[6]?
9. Now you're ready to explore and characterize brain rhythms by their frequency, in order to identify psychological or pathological states. For example, have fun with the `https://archive.ics.uci.edu/datasets?search=eeg`

[4] m

[5] A periodogram is used to estimate the spectral density of a signal by Fourier transforming its auto-correlation function.. The periodogram is therefore a non-parametric method that does not require signal modelling.

[6] "The alpha rhythm is a brain rhythm, i.e. an EEG oscillation resulting from the brain's electrical activity, with a frequency of between 7.5 and 12.5 Hz. The alpha rhythm appears when the person being recorded closes their eyes and relaxes. Because of their high amplitude, alpha waves were the first signals identified by the inventor of the EEG, Hans Berger." wiki