Projects in time series analysis MAP565

- **Programming :** Any software can bused: Scilab, R, octave, python,... The scripts must be archived in a single file (zip, tar or tgz) and should not be printed in the report. Figures and numerical results should be included in the report.
- **Report :** it must be typed (preferably using LaTex) and exported in pdf format. It can be written in French or English.
- **Deadline :** The two files (an archive file containing the scripts + a pdf file for the report) will be emailed to the teacher associated to the project by **midnight on Friday, March 6, 2015**. Emails sent after this date will not be considered.

Contents

1 **Interpolating missing values in a time series** **3**

2 **Smoothing trajectories using the Kalman filter** **5**

3 **Assessing model fit in the spectral domain** **7**

4 **Assessing the significance of a linear trend** **9**

5 **Global warming** **11**

6 **Estimation of the spectral density** **13**

7 **Glacial varve series** **15**

8 **Maximum likelihood estimation for a GARCH process** **16**

1 Interpolating missing values in a time series

cappe@telecom-paristech.fr

Goals

The objective of this project is to optimally interpolate $\ell_2 - \ell_1 + 1$ consecutive values $X_{\mathcal{M}} = (X_{\ell_1}, X_{\ell_1+1}, \ldots, X_{\ell_2-1}, X_{\ell_2})$ of a real valued time series $(X_t)_{t\in\mathbb{Z}}$ given observed values before (X_t, for $t < \ell_1$) and after (X_t, for $t > \ell_2$) the missing section.

The interpolation method is based on L^2 projection ideas using the assumption that $(X_t)_{t\in\mathbb{Z}}$ is an AR(p) process with known parameters. The resulting method can, for instance, be applied to conceal localized defects (clicks, pops, short interruptions) in musical signals.

The AR(1) case

To start with, let us assume that $(X_t)_{t\in\mathbb{Z}}$ is a centered AR(1) process with canonical parameter ϕ (with $|\phi| < 1$) and innovation variance σ^2. Further assume that the missing section consists of a single sample $X_{\mathcal{M}} = X_\ell$ and that we have also observed $X_{\mathcal{P}} = (X_t)_{1\le t\le \ell-2}$, $X_{\ell-1}$, $X_{\ell+1}$ and $X_{\mathcal{F}} = (X_t)_{\ell+2\le t\le n}$.

Question 1 Show that the best linear predictor of $X_{\ell+i}$ (with $i \ge 0$) given $(X_t)_{1\le t\le \ell-1}$ is equal to $\phi^{i+1}X_{\ell-1}$ and that the ratio between the variance of the corresponding prediction error $X_{\ell+i} - \phi^{i+1}X_{\ell-1}$ and that of the original process X_t is equal to $1 - \phi^{2(i+1)}$. What can be said for the backward prediction of $X_{\ell-i}$ (with $i \ge 0$) given $(X_t)_{\ell+1\le t\le n}$?

Question 2 Deduce from the above, that any $Z \in \mathrm{span}(X_{\mathcal{P}}, X_{\ell-1})$ may be written as

$$\mathrm{proj}\left(Z\,|\mathrm{span}(X_{\ell-1})\right) + W\,,$$

where $\mathrm{proj}\left(Z\,|\mathrm{span}(X_{\ell-1})\right)$ denotes the L^2 projection of Z onto the linear subspace $\mathrm{span}(X_{\ell-1})$ and W is orthogonal (in L^2) to $X_{\ell-1}$, X_ℓ and $X_{\ell+1}$.

Similarly, show that elements Z of $\mathrm{span}(X_{\ell+1}, X_{\mathcal{F}})$ can be written as

$$\mathrm{proj}\left(Z\,|\mathrm{span}(X_{\ell+1})\right) + W\,,$$

where W is orthogonal to $X_{\ell+1}$, X_ℓ and $X_{\ell-1}$.

Question 3 Show that this implies that

$$\mathrm{proj}\left(X_\ell\,|\mathrm{span}(X_{\mathcal{P}}, X_{\ell-1}, X_{\ell+1}, X_{\mathcal{F}})\right) = \mathrm{proj}\left(X_\ell\,|\mathrm{span}(X_{\ell-1}, X_{\ell+1})\right)\,,$$

or, in other words, that the best linear predictor of the missing sample X_ℓ is a function of $X_{\ell-1}$ and $X_{\ell+1}$ only.

Question 4 Show that

$$\mathrm{proj}\left(X_\ell\,|\mathrm{span}(X_{\ell-1}, X_{\ell+1})\right) = \frac{\phi}{1+\phi^2}\left(X_{\ell-1} + X_{\ell+1}\right)\,,$$

with the ratio between the variance of the prediction error and that of X_t being equal to $(1-\phi^2)/(1+\phi^2)$; hint: Replace $\mathrm{span}(X_{\ell-1}, X_{\ell+1})$ by $\mathrm{span}\left(X_{\ell-1}, X_{\ell+1} - \mathrm{proj}\left(X_{\ell+1}\,|\mathrm{span}(X_{\ell-1})\right)\right)$.

Why (in what situations) is this solution preferable to the naive linear interpolation estimate $(X_{\ell+1} + X_{\ell+1})/2$?

Question 5 Let $(\epsilon_t)_{t\in\mathbb{Z}}$ denote the innovation process associated to $(X_t)_{t\in\mathbb{Z}}$. Find the matrix $\boldsymbol{\Phi}$ such that $\boldsymbol{\epsilon} = \boldsymbol{\Phi}\boldsymbol{X}$ where

$$\boldsymbol{\epsilon} = \begin{pmatrix}\epsilon_\ell\\ \epsilon_{\ell+1}\end{pmatrix} \quad\text{and}\quad \boldsymbol{X} = \begin{pmatrix}X_{\ell-1}\\ X_\ell\\ X_{\ell+1}\end{pmatrix}.$$

Show that the expression of the projector $\text{proj}\left(X_\ell \,|\text{span}(X_{\ell-1}, X_{\ell+1}\right))$ may also be found as the solution of the program

$$\underset{\text{w.r.t. } X_\ell}{\text{argmin}} \|\boldsymbol{\Phi X}\|^2$$

The general case

For general AR(p) processes and larger gaps to be interpolated, one can show that the two main conclusions observed above still holds:

(i) The optimal linear estimate of the missing data vector

$$\boldsymbol{X}_{\mathcal{M}} = (X_{\ell_1}, X_{\ell_1+1}, \dots, X_{\ell_2-1}, X_{\ell_2})^T$$

only depends on $\boldsymbol{X}_{\mathcal{B}} = (X_{\ell_1-p}, \dots, X_{\ell_1-2}, X_{\ell_1-1})^T$ and

$$\boldsymbol{X}_{\mathcal{A}} = (X_{\ell_2+1}, X_{\ell_2+2}, \dots, X_{\ell_2+p})^T.$$

(ii) The expression of the estimate may be found by writing $\boldsymbol{\epsilon} = \boldsymbol{\Phi X}$, where

$$\boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_{\ell_1} \\ \epsilon_{\ell_1+1} \\ \vdots \\ \epsilon_{\ell_2} \\ \vdots \\ \epsilon_{\ell_2+p} \end{pmatrix} \quad , \quad \boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_{\mathcal{B}} \\ \cdots \\ \boldsymbol{X}_{\mathcal{M}} \\ \cdots \\ \boldsymbol{X}_{\mathcal{A}} \end{bmatrix} ,$$

$\boldsymbol{\Phi}$ is a $(\ell_2 - \ell_1 + p + 1) \times (\ell_2 - \ell_1 + 2p + 1)$ matrix formed from the AR coefficients $\phi_1, \dots, \phi_p$ and solving the program

$$\underset{\text{w.r.t. } \boldsymbol{X}_{\mathcal{M}}}{\text{argmin}} \ \|\boldsymbol{\Phi X}\|^2. \tag{1}$$

Question 6 Solve (1) to show that the best linear predictor of $\boldsymbol{X}_{\mathcal{M}}$ in terms of $\boldsymbol{X}_{\mathcal{B}}$ and $\boldsymbol{X}_{\mathcal{A}}$ is given by

$$-\left(\boldsymbol{\Phi}_{\mathcal{M}}{}^T \boldsymbol{\Phi}_{\mathcal{M}}\right)^{-1} \boldsymbol{\Phi}_{\mathcal{M}}{}^T \left[\boldsymbol{\Phi}_{\mathcal{B}} \ \vdots \ \boldsymbol{\Phi}_{\mathcal{A}}\right] \begin{bmatrix} \boldsymbol{X}_{\mathcal{B}} \\ \cdots \\ \boldsymbol{X}_{\mathcal{A}} \end{bmatrix}$$

where $\boldsymbol{\Phi}_{\mathcal{B}}$, $\boldsymbol{\Phi}_{\mathcal{M}}$ and $\boldsymbol{\Phi}_{\mathcal{A}}$ are matrices of size $(\ell_2 - \ell_1 + p + 1) \times p$, $(\ell_2 - \ell_1 + p + 1) \times (\ell_2 - \ell_1 + 1)$ and $(\ell_2 - \ell_1 + p + 1) \times p$ respectively such that

$$\boldsymbol{\Phi} = \left[\boldsymbol{\Phi}_{\mathcal{B}} \ \vdots \ \boldsymbol{\Phi}_{\mathcal{M}} \ \vdots \ \boldsymbol{\Phi}_{\mathcal{A}}\right]$$

Question 7 Apply the above method to sections of various length[1] of the audio signal downloadable from http://perso.telecom-paristech.fr/~cappe/fr/Enseignement/data/mext. You will previously need to fit an AR model (decide on what value of p you find suitable) to the whole signal. How does the accuracy of the method evolves when interpolating larger sections of the signal?

[1]Of course in this simulated example, there are no missing data and one can compare the interpolation results to the ground truth provided by the actual signal.

2 Smoothing trajectories using the Kalman filter

cappe@telecom-paristech.fr

Goals

The objective of this project is to use the Kalman filter to denoise observations related to the movement of a target in the two dimensional plane. The project illustrates in particular the construction of the state-space model as well as the influence of its parameters.

The model

We consider that the object to be tracked is moving in the two dimensional plane and that its position is represented by the Cartesian coordinates $P_{k,1}$ and $P_{k,2}$. The state equation traditionally used in this problem corresponds to a prior on the target motion that favors regular trajectories, and in particular trajectories with quasi constant velocities (the velocity components $A_{k,1}$ and $A_{k,2}$ represent the derivative of the position $P_{k,1}$ and $P_{k,2}$, respectively). As the measurements are obtained at discrete times, we need however a time discretization argument to define the state space model.

Question 1 Consider a continuous time object moving along a one dimensional axis with position $P(t)$, velocity (first derivative of the position) $V(t)$ and a acceleration (second derivative of the position) $A(t)$. Denote by $t_k = \delta k$, for $k \in \mathbb{Z}$ discretization instants and by $P_k = P(t_k)$, $A_k = A(t_k)$ and $V_k = V(t_k)$ the discretized quantities. Assuming that $A(t) \approx A(t_{k-1})$ for $t \in (t_{k-1}, t_k)$ and that the discretization step δ is small show that the following approximate discretized system of equations holds

$$\begin{aligned} V_k &\approx V_{k-1} + \delta A_{k-1} \,, \\ P_k &\approx P_{k-1} + \delta V_{k-1} + \delta^2/2 A_{k-1} \,. \end{aligned}$$

Question 2 In the two-dimensional case, defining the state vector as $\boldsymbol{X}_k = (P_{k,1}, P_{k,2}, V_{k,1}, V_{k,2})^T$, deduce from what precedes that the state transition can be modelled as $\boldsymbol{X}_k = \boldsymbol{\Phi}\boldsymbol{X}_{k-1} + \boldsymbol{\Pi}\boldsymbol{A}_{k-1}$, where

$$\boldsymbol{\Phi} = \begin{pmatrix} 1 & 0 & \delta & 0 \\ 0 & 1 & 0 & \delta \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} , \quad \boldsymbol{\Pi} = \begin{pmatrix} \delta^2/2 & 0 \\ 0 & \delta^2/2 \\ \delta & 0 \\ 0 & \delta \end{pmatrix} \text{ and } \boldsymbol{A}_{k-1} = \begin{pmatrix} A_{k-1,1} \\ A_{k-1,2} \end{pmatrix} . \tag{2}$$

Question 3 Assuming that $(A_{k,1})_{k\in\mathbb{Z}}$ and $(A_{k,2})_{k\in\mathbb{Z}}$ are two uncorrelated white noise processes with common variance σ^2 convert Equation (2) to the standard form used in the lecture notes (see section 4.2), $\boldsymbol{X}_k = \boldsymbol{\Phi}\boldsymbol{X}_{k-1} + \boldsymbol{W}_k$, and compute the covariance matrix Q of $\boldsymbol{W}_k$.

For the observation equation, we will consider that noisy observations of the positions are available such that

$$\boldsymbol{Y}_k = \boldsymbol{\Psi}\boldsymbol{X}_k + \boldsymbol{V}_k \,,$$

where

$$\boldsymbol{\Psi} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

and $\boldsymbol{V}_k$ is a two-dimensional white noise sequence with covariance matrix proportional to the identity: $R = \rho^2 \mathbf{Id}_2$.

Question 4 Implement the Kalman filtering algorithm as specified in the lectures notes (Algorithm 5).

Question 5 Test it on the sample trajectories provided in
http://perso.telecom-paristech.fr/~cappe/fr/Enseignement/data/traj
taking the sampling period δ equal to 1. These two trajectories correspond to independent noisy observations of a single reference trajectory. How to exploit this information to estimate the parameter ρ^2? Show that using $\boldsymbol{\mu}_0 = 0$ and Σ_0 very large (eg. $\kappa^2 \mathbf{Id}_4$ with a large value of κ) correspond to a sensible initialization such that in particular $\boldsymbol{\Psi} \boldsymbol{X}_{1|1} \approx \boldsymbol{Y}_1$. What is the influence of the parameter σ^2? Is it possible to use both trajectories to improve the filtering results and how?

3 Assessing model fit in the spectral domain

cappe@telecom-paristech.fr

Goals

In this project we consider a famous time series which has attracted the curiosity of astronomers for more than three centuries. We will use this example to asses the fit of autoregressive models in the frequency domain.

Preliminaries

Recall from the lecture notes that the so-called periodogram

$$I_n(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} X_t \mathrm{e}^{-it\lambda} \right|^2 ,$$

is the Fourier transform of the empirical covariance function $\hat{\gamma}_n$ (see Section 5.2).

We will first assume that X_t is a Gaussian white noise of variance σ^2.

Question 1 Show that the real and imaginary parts of $F_n(\lambda) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t \mathrm{e}^{-it\lambda}$ are jointly Gaussian variables with mean zero and expectation tending to $\sigma^2/2$ and that their covariance tends to zero as $n \to \infty$.

Question 2 Deduce from what precedes that the distribution of $I_n(\lambda)$ converges to an exponential distribution with mean $\sigma^2/2\pi$.

Question 3 Proceeding similarly, show that for distinct frequencies $\lambda_1, \ldots, \lambda_k$, the joint distribution of $I_n(\lambda_1), \ldots, I_n(\lambda_k)$ converges to a distribution with independent marginals.

Consider now an ARMA model

$$X_t = \sum_{k=1}^{p} \phi_k X_{t-k} + \sum_{k=1}^{q} Z_{t-k} + Z_t ,$$

where (Z_t) is a white noise of variance σ^2 an define

$$f(\lambda) = \frac{\left|1 + \sum_{k=1}^{q} \theta_k \mathrm{e}^{-ik\lambda}\right|^2}{\left|1 - \sum_{k=1}^{p} \phi_k \mathrm{e}^{-ik\lambda}\right|^2} .$$

A result known as Bartlett decomposition states that

$$I_n^X(\lambda) = f(\lambda) I_n^Z(\lambda) + R_n(\lambda) ,$$

where the variance of the remainder term $R_n(\lambda)$ is bounded by an $O(n^{-1})$ term and I_n^X and I_n^Z denote the periodograms of (X_t) and (Z_t) respectively.

In other words, if we assume (Z_t) to be a Gaussian process, it means that $I_n^X(\lambda)$ when scaled by $f(\lambda)$ should behave approximately as the periodogram of the Gaussian white noise investigated in the previous question.

We will apply this idea to the sunspot numbers series available from
http://perso.telecom-paristech.fr/˜cappe/fr/Enseignement/data/yearssn
to asses the fit of AR models. We will consider the original series with its empirical mean subtracted.

Question 4 Estimate AR models of various order p (from 1 to 10 for instance) from the centered sunspot series using the Yule Walker approach.

Question 5 Plot both the periodogram of the series and the spectral densities $\hat{f}_{n,p}$ corresponding to the estimated AR models for different values of p. It is recommended to plot the periodogram and the spectral densities on the log scale and to represent only frequencies between 0 and π due to the symmetry of properties of both I_n and $\hat{f}_{n,p}$. Both I_n and $\hat{f}_{n,p}$ can be computed using the Fast Fourier Transform (FFT)[2]. When does the fit appears to be acceptable?

Question 6 For a more objective answer, one can consider Quantile-Quantile (QQ) plots of $I_n(\lambda)/\hat{f}_{n,p}(\lambda)$ against the exponential distribution or using the Kolmogorov–Smirnov test.

Question 7 An issue of interest about this series is whether or not it is indeed stationary. Try replicating the previous experiment (with the p selected in the previous question) but estimating the periodogram from the first half of the data and the AR model from the second half (or vice versa). Does this experiment suggest obvious signs of non stationarity?

[2]Even if, strictly speaking, the result of Question 3 has been shown for fixed frequencies (ie. not depending on n) and not for frequencies of the form $\lambda_k = 2\pi k/n$ as required for computation by the FFT algorithm.

4 Assessing the significance of a linear trend

cappe@telecom-paristech.fr

Goals

The aim of this project is to assess the statistical significance of an observed increasing trend in an historical series of temperatures. We will in particular show that ignoring the correlated nature of the residuals may lead to an incorrect assessment of the level of significance of the statistical test.
Keywords: Least squares regression, test of significance, autoregressive modeling, weighted least squares.

Preliminary analysis

The time series (Y_k) we will consider in this project is taken from the book *Introduction to Time Series and Forecasting* by P. J. Brockwell and R. A. Davis and consists of $n = 99$ values that correspond to the average maximum temperature over the month of September for the years 1895-1993 in an area of the USA whose vegetation is characterized as tundra. The data can be downloaded from http://perso.telecom-paristech.fr/~cappe/fr/Enseignement/data/tundra/

Question 1 We first fit a simple linear regression model (with constant offset β_0 and linear slope β_1) to the data. Using the canonical matrix form of the regression model

$$\underbrace{\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}}_{\boldsymbol{Y}} = \underbrace{\begin{pmatrix} 1 & 1 \\ \vdots & \vdots \\ 1 & n \end{pmatrix}}_{\boldsymbol{X}} \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}}_{\boldsymbol{\beta}} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} \tag{3}$$

recall the expressions of the least squares estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ of, respectively, $\boldsymbol{\beta}$ and the variance σ^2 of the noise sequence (ϵ_k).

Question 2 Show that under the Gaussian assumption, $\hat{\boldsymbol{\beta}}/\sigma$ is a centered two-dimensional Gaussian variable with covariance matrix $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$. Use the approximation that $\hat{\boldsymbol{\beta}}/\hat{\sigma}$ follows the same distribution[3] to obtain a confidence interval for β_1. Show that the observations suggest an increasing trend of the temperature, that is however only moderately significant (conclusive at the 95% confidence level but not so at the 99% level).

Refined analysis using weighted least squares

Examination of the series of residual from the previous fit, $R_k = Y_k - (\hat{\beta}_0 + k\hat{\beta}_1)$, however suggests that the assumption that the noise sequence is IID is most likely incorrect; we will see that this fact may impact the significance of the previous test.

Question 3 Fit an AR(1) model to the series (R_k) of residuals. Justify, using methods presented in the MAP565 course that it is useless to consider higher order models.

Question 4 Deduce from what precedes an estimate of the covariance matrix $\boldsymbol{\Gamma}$ of the vector $(R_1, \ldots, R_n)^T$. We will now fit the model of (3) assuming that the noise sequence is multivariate Gaussian with covariance $\boldsymbol{\Gamma}$. Show that the maximum likelihood estimator of $\boldsymbol{\beta}$ is now given by the weighted least squares estimate $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y}$, where $\boldsymbol{W} = \boldsymbol{\Gamma}^{-1}$. Further show that $\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}$.

[3]Note that it is known that $\hat{\boldsymbol{\beta}}/\hat{\sigma}$ indeed follows a multivariate Student-t distribution. Why do we neglect this fact here?

Question 5 Explain why the previous computation can be equivalently implemented by "pre-withening" the series (Y_k) as well as the two columns of the regression matrix by filtering them with the fitted AR polynomial $1 - \hat{\phi}B$ and using regular (unweighted) least squares.

Question 6 Implement the weighted least squares estimator as well the computation of the refined confidence interval for the slope β_1. Show that now the estimate of the slope parameter is unchanged but that it is no more significant at the 95% confidence level. Can you give an intuitive interpretation of this result?

5 Global warming

agathe.guilloux@upmc.fr

Goal

The aim of this project is to study time-series measuring the evolution of global temperature. We consider a particular global temperature series record, that can be found online here:
`http://www.lsta.upmc.fr/guilloux/X/Timeseries/globtemp.dat`
This file contains 142 observations (from the year to 1856 to 1997), corresponding to land-air average temperatures anomalies. We recommend to use the `R` software for this project, but it is not mandatory.

Temperature prediction

Question 1 Load the data, and observe it. Note the apparent upward trend in the series. From now on, we denote by $X = (X_t)_t$ the series containing the temperatures from 1900 up to 1997. Construct this series.

Question 2 Do the series look stationary?

Question 3 Fit a simple linear regression model

$$X_t = \beta_1 + \beta_2 t + \varepsilon_t,$$

where ε_t is a residual. Give the values of the least-squares estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ of β_1 and β_2. Print the series and the linear regression curve together. Assuming that the residuals are i.i.d. and Gaussian, what is the distribution of $(\hat{\beta}_1, \hat{\beta}_2)$?

Question 4 Compute the p-value of the T-test for the assumption $H_0 : \beta_2 = 0$. What is the conclusion of the test?

Question 5 Now, we want to remove the trend of X, to obtain a series that seems stationary. Remove the estimated trend using the previous linear regression fit by computing the series $Y = (Y_t)$ where

$$Y_t = X_t - \hat{\beta}_1 + \hat{\beta}_2 t. \tag{4}$$

Plot the autocorrelation (ACF) and partial autocorrelation (PACF) of X and Y, and comment your result.

Question 6 Another way of detrending a time series is to differentiate it, namely to apply the operator $\nabla = I - B$ one or sereval times. Compute the differentiated series by applying ∇ one time, and plot its ACF and PACF. Explain why this operator helps to remove the trend. Compare both way of detrending by comparing the ACFs.

Question 7 A random walk can be used as well to detrend a series. Indeed, instead of using a deterministic linear function, we can use a stochastic model

$$\mu_t = \delta + \mu_{t-1} + w_t \tag{5}$$

of the trend, the series X being modelled now as

$$X_t = \mu_t + \varepsilon_t, \tag{6}$$

where δ is a constant and (w_t) is a white noise independent of (ε_t). Prove that $\nabla(X_t)$ is stationnary in this case. Fit this model on the data (namely, estimate δ). Assuming that w is a general second order stationary process with autocovariance γ_w, compute the autocovariance of X.

Question 8 Fit an ARIMA(p, d, q) model to the series X. Choose the p, d, q parameters by doing back-fitting tests (or, optionally, using a model-selection criterion such as AIC, AICc, BIC). Give a precise description of the methods used for the choice of these parameters.

Question 9 After choosing an appropriate ARIMA model, predict the value of the series for the next 10 years, with standard errors. Plot the series and its prediction with standard errors over the next 10 years.

6 Estimation of the spectral density

agathe.guilloux@upmc.fr

Goal

The aim of this project is to study an estimator of the spectral density, and to compute it on a dataset. We recommend to use the R software for this project, but it is not mandatory.

Spectral density estimation

Let $(X_t)_{t\in\mathbb{Z}}$ be a centered second order stationary process in $\mathbb{R}^{\mathbb{Z}}$ with autocovariance γ and spectral density f. We assume that $\gamma \in \ell_1$. We want to estimate the spectral density f of X based on observations $\{X_1, \dots, X_T\}$.

Question 1 Write f_X in function of γ. We consider the following estimator of the autocovariance:

$$\hat{\gamma}_T(p) = \begin{cases} \dfrac{1}{T} \displaystyle\sum_{t=|p|+1}^{T} X_t X_{t-|p|} & \text{if } |p| < T, \\ 0 \ \text{ if } |p| \geq T, \end{cases}$$

and an estimator of $f(\omega)$ as

$$\hat{f}(\omega; d_T) = \frac{1}{2\pi} \sum_{|d|\leq d_T} \hat{\gamma}_T(d) e^{-id\omega},$$

where $(d_T)_{T\in\mathbb{N}}$ is a sequence of integers going to $+\infty$ and such that $d_T < T$ for any $T \geq 1$.

Question 2 Prove that $\lim_{T\to+\infty} \mathbb{E}(f(\omega, d_T)) = f(\omega)$. What does it mean concerning the estimator $\hat{f}(\omega, d_T)$?

Question 3 From now on, we assume that $\{X_t\}_{t\in\mathbb{Z}}$ is centered IID with moments of order 4. Prove that $\mathrm{cov}(\hat{\gamma}_T(p), \hat{\gamma}_T(q)) = 0$ for $0 < p < q < T$, and that consequently

$$\mathrm{var}(\hat{f}(\omega; d_T)) = \frac{\mathrm{var}(X_0^2)}{4\pi^2 T} + \frac{1}{T\pi^2} \sum_{d=1}^{d_T} \big(1 - \frac{d}{T}\big)(\mathbb{E}(X_0^2))^2 \cos^2(d\omega).$$

Question 4 Recall that the periodogram is given, for any frequency $\omega \in [-\pi, \pi]$, by

$$I_T(\omega) = \frac{1}{2\pi T} \Big|\sum_{t=1}^{T} X_t e^{i\omega t}\Big|^2.$$

Prove that $I_T(\omega) = \hat{f}(\omega, T-1)$. Compute the limit of $I_T(\omega)$ and conclude that it is not a convergent estimator of $f(\omega)$.

Question 5 Give conditions on the sequence $\{d_T\}_{T\in\mathbb{N}}$ that ensure the convergence of the variance of $\hat{f}(\omega; d_T)$ to zero. Deduce that in this case, the estimator is convergent in L^2 (namely the $L^2([-\pi, \pi])$ distance between f and $\hat{f}(\cdot, d_T)$ goes to zero).

Question 6 Load the data from:

`http://www.lsta.upmc.fr/guilloux/X/Timeseries/sunspots.dat`

in the software of your choice. This data contains monthly mean relative sunspot numbers from 1749 to 1983, collected at Swiss Federal Observatory, Zurich until 1960, then Tokyo Astronomical Observatory. Compute the raw periodogram $I_n(\cdot)$ and the estimator $\hat{f}(\cdot, d_T)$

for some choices of d_T, and plot them. Plot also the vertical line at frequency $\omega = 1/120$. Comment.

Now, apply a moving average of order m to this series (namely, if m is even, compute $Y_t = \frac{1}{m+1}(X_{t-m/2} + \cdots + X_{t+m/2})$, and similarly for m odd). Compute the raw periodogram and the estimator $\hat{f}(\cdot, d_T)$, together with vertical lines at the frequencies $\omega \in \{1/120, 1/m, 2/m, \cdots\}$. Comment.

7 Glacial varve series

agathe.guilloux@upmc.fr

Goal

This project is about geolodical data. It concerns paleoclimatic varves of melting glaciers, that deposit yearly layers of sand and silt during the spring melting seasons, which can be reconstructed over a period of deglaciation that began in New England about 12,600 years and ended about 6,000 years ago. Such sedimentary deposits, called varves, can be used as proxies for paleoclimatic temperature, since, in a warm year, more sand and silt are deposited from the receding glacier.

Model fitting and prediction

Question 1 Load the data from

`http://www.lsta.upmc.fr/guilloux/X/Timeseries/varve.dat`

in a series $X = (X_t)$ and plot it. The variation in thicknesses increases in proportion to the amount deposited, so the variance seems to be increasing with time. Propose a simple transformation $Y = (Y_t)$ of X that makes it look like more stationary. Plot histograms of the data before an after transformation. What can you say about the distribution of the transformed data?

Question 2 Plot the ACF and PACF of Y. Comment about the stationarity of Y.

Question 3 Compute the differenciated process $\nabla(Y_t) = Y_t - Y_{t-1}$, and plot its ACF and PACF. In view of these plots, which ARIMA model would you choose to model the series Y?

Question 4 Fit this model to the series. Do a diagnostic of this fit, by plotting the residuals, by computing the ACF and PACF of the residuals, and by applying a statistical test of adequation to a white noise of the residuals. Is your fit satisfactory? If it is not, change the parameters of the ARIMA, and do again the checks on the residuals. Give the value of the estimated parameters with standard errors. Using this fit, predict the series over the next 24 months, and plot the prediction with standard errors. Comment.

Question 5 Consider the univariate state-space model given by state conditions $Z_0 = W_0$, $Z_t = Z_{t-1}+W_t$ and observations $Y_t = Z_t + V_t$, $t = 1, 2, \cdots$, where W_t and V_t are independent, Gaussian, white noise processes with $\text{var}(W_t) = \sigma_w^2$ and $\text{var}(V_t) = \sigma_v^2$. Show that under this assumption, (Y_t) follows an ARIMA(0,1,1) model. Fit this model to the series Y of transformed varve data series, and compare it to the results obtained before.

8 Maximum likelihood estimation for a GARCH process

agathe.guilloux@upmc.fr

Goal

Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models are a common tool in finance, in particular for the analysis and prediction of volatility for low-frequency data. Indeed, these models provide a simple way of reproducing some *stylized facts*, namely, if $X_t = \log(P_t/P_{t-1})$ is the log-returns of an asset P_t, we observe the following: $\{X_t\}$ is uncorrelated but not independent; the volatility $\sigma_t = \sqrt{\mathrm{var}(X_t|X_{t-1}, X_{t-2}, \ldots)}$ changes with t and we observe *clusters* (periods with strong volatility); the distribution of X_t is not Gaussian (the Black-Scholes models is not satisfied) but heavy-tailed; the conditional distributions of $X_t|(X_{t-1} > 0)$ and $X_t|(X_{t-1} < 0)$ are different (Leverage effect).

In this project we will apply maximum likelihood estimation (MLE) to fit a GARCH model on a financial times-series. For this project it is recommended to use the `R` software, with the `MASS` and `tseries` libraries. Don't hesitate to read the documentation of the `tseries` library, since many built-in functions are available, and the numerical questions below can all be easily answered using the appropriate functions.

Questions

Question 1 Load the `tseries` and `MASS` libraries using the `library` command. Load data by typing `data(EuStockMarkets)`. Plot the time-series available in this dataset (using `plot.ts`) and plot the log-returns of these series. Plot the auto-correlation functions (ACF) of the log-returns and ACF of the squared log-returns. Comment.

Question 2 Suppose that we observe $X_1, \ldots, X_n$, where $\{X_t\}$ satisfies a GARCH model

$$\begin{aligned} X_t &= \sigma_t \varepsilon_t \\ \sigma_t^2 &= a_0 + \sum_{j=1}^{p} a_j X_{t-j}^2 + \sum_{j=1}^{q} b_j \sigma_{t-j}^2, \end{aligned}$$

where we assume that $\{\varepsilon_t\}$ is i.i.d and $N(0,1)$. Denote $\theta = (a_0, \ldots, a_p, b_1, \ldots b_q)$ the non-negative parameters of the model. From the data $\{X_1, \ldots, X_n\}$ we want to estimate θ. For this, one typically use MLE, which consist in finding the maximum of $\theta \mapsto \ell(\theta)$ where $\ell(\theta) = \ell(\theta; X_1, \ldots, X_n)$ is the logarithm of the joint distribution of the data (the *log-likelihood*).

Prove that

$$\ell(\theta; x_1, \ldots, x_n) = \ell(\theta; x_1) + \sum_{j=2}^{n} \ell(\theta; x_j | x_1, \ldots, x_{j-1}),$$

where $\ell(x_t|x_1, \ldots, x_{t-1})$ is the logarithm of the density of X_t conditional on $X_1, \ldots, X_{t-1}$ when the parameter is θ.

Question 3 Assuming that $X_0, \ldots, X_{-p+1}$ and $(\sigma_0^2, \ldots, \sigma_{-q+1}^2)$ are observed, prove that

$$\begin{aligned} &\ell(\theta; X_1, \ldots, X_n | X_0, \ldots, X_{1-p}, \sigma_0^2, \ldots, \sigma_{-q+1}^2) \\ &\qquad = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\Big(\log v_t(\theta) + \frac{X_t^2}{v_t(\theta)}\Big), \end{aligned}$$

where

$$v_t(\theta) = \begin{cases} \sigma_t^2 & \text{if } t \le 0 \\ a_0 + \sum_{j=1}^{p} a_j X_{t-j}^2 + \sum_{j=1}^{q} b_j v_{t-j}(\theta) & \text{otherwise,} \end{cases}$$

Question 4 In practice, we don't observe $X_0, \dots, X_{1-p}, \sigma_0^2, \dots, \sigma_{-q+1}^2$, hence we need to use some heuristics to approximate it. Typically, one chooses $X_t = 0$ for $t \leq 0$. For σ_t^2, with $t \geq 0$, we use the following remark: first, prove that $v_t(\theta)$ satisfies

$$Q(B)(v_t(\theta)) = a_0 + P(B)(X_t^2),$$

with B the backshift operator and $P(z) = \sum_{j=1}^p a_j z^j$ and $Q(z) = 1 - \sum_{j=1}^q b_j z^j$. Then, deduce that

$$v_t(\theta) = \frac{a_0}{1 - \sum_{j=1}^q b_j} + \sum_{j \geq 1} \pi_j X_{t-j}^2$$

when $\sum_{j=1}^q b_j < 1$, where you will define the coefficients π_j from P and Q. Explain why the choice

$$v_t(\theta) = \frac{a_0}{1 - \sum_{j=1}^q b_j}$$

for $t \leq 0$ seems reasonable.

Question 5 The MLE estimator of θ is therefore given by

$$\hat{\theta} \in \operatorname*{argmin}_{\theta} \sum_{t=1}^n \Big(\log v_t(\theta) + \frac{X_t^2}{v_t(\theta)} \Big). \tag{7}$$

Fit a GARCH(1, 1) model to one of the series in the `EuStockMarkets` dataset (using the `garch` routine). Plot the corresponding residuals, their ACF, the ACF of the squared residuals, and comment. Plot the estimated conditional variance. Propose statistical tests for the Gaussianity and independence of the residuals, and compute their p-values. Comment.

Question 6 The argmin in (7) can be approximated by a second-order optimization algorithm, such as the Newton algorithm. Use the `optim` routine from `R` to compute this estimator, and compare your results with the built-in function `garch` from the `tseries` library on one of the datasets mentioned above.

Question 7 The higher the log-likelihood $\ell(\hat{\theta})$ of an estimator $\hat{\theta}$ is, the best is the *goodness-of-fit* of $\hat{\theta}$. However, *over-fitting* the model typically leads to poor predictions. A way to deal with this is to choose the parameters p and q using a model-selection criterion, such as the AIC (Akaike's Information Criterion) or the BIC (Bayesian Information Criterion). AIC and BIC suggest to choose the parameters p, q by minimizing a penalized log-likelihood:

$$\begin{aligned} (AIC): \quad & (\hat{p}, \hat{q}) = \operatorname*{argmin}_{p,q} -2\ell(\hat{\theta}_{p,q}) + 2\mathrm{df}(p,q), \\ (BIC): \quad & (\hat{p}, \hat{q}) = \operatorname*{argmin}_{p,q} -2\ell(\hat{\theta}_{p,q}) + \log n \, \mathrm{df}(p,q), \end{aligned}$$

where $\hat{\theta}_{p,q}$ is an estimator of the parameters in the GARCH(p, q) model, n is the number of observations and $\mathrm{df}(p,q)$ is the number of parameters in the GARCH(p, q) model.

Compute the log-likelihood (using `logLik`) of fits of GARCH models with increasing values of p and q. Comment. (**Warning**: the command `logLik(fit)` where `fit` is a GARCH fit returns the value of $-\ell(\hat{\theta})$ instead of $\ell(\hat{\theta})$). Use both AIC and BIC criterions to choose a GARCH models for the series mentioned above. Comment.