3 Assessing model fit in the spectral domain

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 variances 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 $\lambda_1, \dots, \lambda_k$, the joint distribution of $I_n(\lambda_1), \dots, I_n(\lambda_k)$ converges to a distribution with independent marginals.

Consider now an ARMA model

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

where (Z_t) is a white noise of variance σ^2 and 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?

[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.

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?