MASTER IN ASTROPHYSICS

Homework 3 : The periodogram

S. Sulis, D. Mary
sophia.sulis@oca.eu

Objective : Study the classical and averaged periodogram properties. Design Monte Carlo simulations. Study the sparsity of the RV Keplerian signature in the Fourier domain.

1 Classical periodogram

Definition [1]:

$$\begin{aligned} P(\nu_k) &= \frac{1}{N}\Big|\mathrm{FT}[X]\Big|^2 \\ &= \frac{1}{N}\Big|\sum_{j=1}^{N} X(t_j)\mathrm{e}^{-i2\pi\nu_k j}\Big|^2 \\ &= \frac{1}{N}\Big|\sum_{j=1}^{N} X(t_j)\cos\left(2\pi\nu_k j\right) + \sum_{j=1}^{N} X(t_j)\sin\left(2\pi\nu_k j\right)\Big|^2 \end{aligned} \tag{1}$$

with $\{\nu_k = \frac{k}{N}\}_{k=0,\ldots,\frac{N}{2}}$ the Fourier frequencies.

(Asymptotic) Distribution:

$$P(\nu_k) \sim \begin{cases} \frac{S(\nu_k)}{2}\chi_2^2, & \forall k \in [1,..,\frac{N}{2}-1], \\ S(\nu_k)\chi_1^2, & \text{for } k = 0, \frac{N}{2}, \end{cases}$$

with $S(\nu_k)$ the noise PSD. In the case of a WGN, $S(\nu_k) = \sigma^2$.

Exercise 1: Classical periodogram properties

Generate $X_t \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma)$ with $\mu = 0$ m/s, $\sigma^2 = 1m^2/s^2$, $N = 1024$ and assume the series is evenly sampled with $\Delta_t = 1$s.

1. What are the units of the periodogram ?
2. Show the orthogonality between the sines and cosines functions in (1).
3. Verify the periodogram distribution at one frequency with MC simulations. Compare with the theoretical χ_2^2 distribution.
4. Derive, in the WGN case first and then for a general colored noise with PSD S, the expressions of the bias and variance of the periodogram as a PSD estimate.
   Note : the bias and variance are defined as :
   $$b(f) := \mathbb{E}P(f) - S(f)$$
   $$\text{var}P(f) := \mathbb{E}(P(f) - \mathbb{E}P(f))^2.$$
5. Verify bias and variance by Monte Carlo simulations in the WGN case.
6. Verify bias and variance for an autoregressive process of order $o = 2$, $\sigma^2 = 1$ and coefficients $c_j = [0.8, -0.5]$.

2 Periodogram of a sinusoidal signal

If $X(t)$ is a sinusoidal signal with frequency f_s, then :

- if $f_s = \nu_k$ for some k, the signal $X(t)$ and the sinusoidal components ($\mathrm{e}^{-i\omega_s t_j}$ and $\mathrm{e}^{i\omega_s t_j}$) of (1) are perfectly in phase, and X_t is orthogonal to the other exponentials.

- if $f_s \notin \Omega_N$ (i.e., not on the Fourier grid), the sinusoidal component is orthogonal to none of the Fourier frequencies.

We define the Dirichlet kernel (i.e., the FT of a sampled rectangular window) as:

$$D_N(\nu) := \frac{1}{N}\sum_{j=1}^{N} \mathrm{e}^{i2\pi\nu j},$$

and the Fejèr kernel as

$$K_N(\nu) := \left|D_N(\nu)\right|^2 = \left(\frac{\sin(N\pi\nu)}{N\sin(\pi\nu)}\right)^2.$$

We note that $K_N(0) = 1$.

Exercise 2: Periodogram of a pure sinusoid

1. Using the Dirichlet kernel and the Fejèr kernel, develop the analytical expression of the classical periodogram in the case of an evenly sampled real sinusoidal signal $X(t) = A\sin(2\pi f_s t + \Phi_s)$.

2. Generate $X(t) = A\sin(2\pi f_s t)$ with $N = 100$, $A = 1$ and explain what happens when

   (a) $f_s = \nu_{60}$,

   (b) $f_s = (\nu_{61} - \nu_{60})/2$.

   What do you anticipate for the periodogram distribution at different frequencies for a time series composed of this pure sine plus a WGN ?

3. *Frequency resolution.* Generate $X(t) = \sum_{i=1}^{2} \sin(2\pi f_i t)$ with $f_1 = 0.102\text{Hz}$, $f_2 = 0.104\text{Hz}$ ($\Delta t = 1s$). Compare the result for $N = 100, 500$ and 1000. Which value of N do you need to be able to resolve the periodogram peaks?

4. Increase N by adding zeros to the end of the simulated time series and apply the classical periodogram. What is the effect of zero padding? Does it affect the ability to resolve close frequencies?

5. Compare the periodogram of Venus' RV signature obtained with $T = [1, 10, 100, 1000]$ years (all other parameters constant) and $\Delta t = 30$ days.

6. Same question but change now the eccentricity, all other parameters equal. Interpret the results.

3 Averaged periodogram

To reduce the variance of the classical periodogram, we can:

- Separate the N samples in L blocks assumed uncorrelated (assume also $N = L \times m$),
- Calculate the periodogram over each block,
- Average the L obtained periodograms to obtain the "averaged periodogram" [2]:

$$\overline{P}_L(\nu'_k) := \frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{m}\Big|\sum_{j=1}^{m} X_\ell(t_j)\mathrm{e}^{-i2\pi\nu j}\Big|^2, \qquad (2)$$

with $\nu'_k := \frac{k}{m}$.

Distribution:

$$\overline{P}_L(\nu_k) \sim \begin{cases} \dfrac{S(\nu_k)}{2L}\chi^2_{2L}, \ \forall k \ \in \ [1,..,\dfrac{m}{2}-1], \\ \dfrac{S(\nu_k)}{L}\chi^2_{L}, \ \text{ for } \ k=0, \dfrac{m}{2}. \end{cases}$$

Exercise 3: Averaged periodogram properties

1. Comment on the difference (resolution, bandwidth) between the classical periodogram (over N Fourier frequencies) and the averaged periodogram (over m frequencies).
2. Generate $X_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$ with $\mu = 0$, $\sigma^2 = 1$, $N = 5024$ and sampling step 1s. Verify the averaged periodogram distribution with $L = 10$ and 10^4 MC simulations.
3. Add a sinusoidal signal ($A = 1\ m/s,\ f_s = 0.2$Hz) to the WGN and test different blocks size $L = [1, 10, 100, 500]$. Which advantage(s) and drawback(s) of this periodogram do you observe ?
4. Compare numerically the bias and variance of this periodogram with the classical one on the AR(2) of the previous Exercise.

References

[1] Schuster, A., Journal of Geophysical Research, On the investigation of hidden periodicities, p.13, vol.3, 1898.

[2] Bartlett, M.S., Biometrica, Periodogram analysis and continuous spectra, p.1-16, vol.37, 1950.

[3] Parsen, E., The Annals of Mathematical Statistics, num.3, p. 1065-1076, The Institute of Mathematical Statistics, On Estimation of a Probability Density Function and Mode, vol.33, 1962,

[4] Brillinger, D.R., Holden Day, San Francisco, Time Series : Data Analysis and Theory, 1981,