March 31, 2017

MASTER IN ASTROPHYSICS

Meteor Exoplanet Detection
Exam on the theoretical part (2h, no document)

Notations :
In the following we consider regularly sampled signals at time instants $t = 1, 2, \ldots, n$, with n an even integer.

Bolded symbols refer to vectors (for instance $\mathbf{y} := [y_1\ y_2 \ldots y_n]^\top$) and underlined letters refer to matrices (for instance $\underline{\underline{Y}} := [\mathbf{y}_1 \ldots \mathbf{y}_n]$).

The (angular) Fourier frequencies are denoted by $\lambda_j := \frac{2\pi j}{n}, j = 0, 1, \ldots, n$. The set of frequencies $\{\lambda_j\}, j = 1, \ldots n/2 - 1$, is denoted by $\bar{\Omega}_n$.

Let $\mathbf{f}(\lambda_j) := [\mathrm{e}^{\mathrm{i}\lambda_j}\ \mathrm{e}^{2\mathrm{i}\lambda_j} \ldots \mathrm{e}^{(n-1)\mathrm{i}\lambda_j}]^\top$ and $\underline{\underline{F}} := [\mathbf{f}(\lambda_0)\ \mathbf{f}(\lambda_1)\ \mathbf{f}(\lambda_{n-1})]$.

The Discrete Fourier Transform at frequency λ_j of a vector $\mathbf{y}$ is defined as the weighted scalar product $\widetilde{\mathbf{y}}(\lambda_j) := \frac{1}{n}\langle \mathbf{f}(\lambda_j), \mathbf{y}\rangle$.

I. Distribution of the periodogram (10 points)

1. Assume that the signal $\mathbf{y}$ is a sampled white Gaussian noise, with components $Y_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$.
   (a) Show that the periodogram at frequency λ_j, $P(\lambda_j)$, can be written as the sum $X_r^2 + X_i^2$, and define X_r and X_i.
   (b) Characterize the distribution of X_r and X_i.
   (c) Compute $\mathbb{E}X_r$ and $\mathbb{E}X_i$.
   (d) Compute $\mathrm{var}X_r$ and $\mathrm{var}X_i$ for $j \neq 0, n/2$.
   (e) Deduce from the previous question the distribution of $P(\lambda_j)$, $j \neq 0, n/2$.

2. Assume now that the signal $\mathbf{y}$ is a white Gaussian noise plus a sine at some Fourier frequency $\lambda_k \in \bar{\Omega}_n$. So each component $Y_t = B_t + A\sin(\frac{2\pi k t}{n})$, with $B_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$ and $A \in \mathbb{R}^+$ the sine amplitude.
   (a) Show that for Fourier frequencies $\lambda_j \neq \lambda_k$, the distribution of $P(\lambda_j)$ is the same as in 1(e).
   (b) Derive the distribution of $P(\lambda_k)$, the periodogram at signal frequency.
   (c) Deduce from the previous expression the parameters that will influence the detectability of the signal.

3. Assume now that the signal $\mathbf{y}$ is a sampled colored noise with PSD $S(f)$ (with no signal). Are the results above on the distribution of the periodogram modified with respect to the white Gaussian noise case ? How ?

4. Assume now that the signal $\mathbf{y}$ is a white Gaussian noise plus a sine at some frequency ω_k, and that we zero-pad $\mathbf{y}$, resulting in $n' > n$ samples (so we add $n' - n$ zero samples). Make a picture explaining the structure and distribution of the periodogram when $\omega_k \in \bar{\Omega}_n$ (ω_k is on the Fourier grid) and when $\omega_k \notin \bar{\Omega}_n$. (You do not have to compute precisely the non centrality parameters when they exist.)

II. Hypothesis tests based on the periodogram (10 points)

We consider here a subset Ω_m of m frequencies taken in the set $\{\lambda_1, \lambda_2, \dots \lambda_{n/2-1}\}$.
We denote the order statistics of the periodogram by

$$\max_j(P(\lambda_j)) := P_{(1)} > P_{(2)} > \dots > P_{(m)} := \min_j P(\lambda_j).$$

Under the null hypothesis, $\mathbf{y}$ is composed from n i.i.d. samples of a white noise with variance σ^2.

1. Consider the test
$$\frac{P_{(1)}}{\sigma^2} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma. \tag{1}$$
   (a) Compute the probability of false alarm (P_{FA}) of this test.
   (b) Assume that under the alternative hypothesis, $\mathbf{y}$ is as in question I.2, where the Fourier frequency $\lambda_j \in \Omega_m$ of the sine is unknown. Compute the probability of detection ($\mathrm{P}_{\mathrm{DET}}$) of this test.
2. Explain what problem naturally arises for the previous test when a large number of frequencies are involved in Ω_m.
3. Fisher test.
   (a) Write the Fisher test.
   (b) Justify this test statistic.
   (c) Explain its advantage with respect to the test in expression (1) above.
4. Consider the test
$$\frac{P_{(k)}}{\sigma^2} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma. \tag{2}$$
   (a) Compute the P_{FA} of this test.
   (b) Under which kind of alternatives is this test more powerful than test (1) ? Why ?