DEVOIR

DÉTECTER UNE PÉRIODE CACHÉE DANS UN SIGNAL

1 Motivations

On cherche à détecter une période cachée dans un signal bruité. Un signal est vu comme un vecteur dans $\mathbb{C}^n$, avec un pas d'échantillonage constant. Le but de ce devoir est de construire une règle de décision statistique pour détecter la présence d'une période caché dans le signal. Le signal est observé avec du bruit, ce qui rend cette tache éventuellement difficile.

2 Cadre mathématique

Soit $n \in \mathbb{N} - \{0\}$. On considère un signal (un vecteur) $x = (x_1, \ldots, x_n) \in \mathbb{C}^n$. L'espace $\mathbb{C}^n$ est muni du produit scalaire hermitien $\langle x, y \rangle = \sum_{t=1}^n x_t \bar{y}_t$. Pour $j \in \mathbb{Z}$, on considère la fréquence $\omega_j = 2\pi j/n$, et on définit les vecteurs

$$e_j = \frac{1}{\sqrt{n}} \begin{bmatrix} e^{i\omega_j} \\ \vdots \\ e^{in\omega_j} \end{bmatrix}$$

pour $j \in F_n$, où on pose $F_n = \big\{ j \in \mathbb{Z} : -\pi < \omega_j \leq \pi \big\}$.

3 Questions

1. Montrer que le système $\{e_j : j \in F_n\}$ est une base orthonormée de $\mathbb{C}^n$. En déduire que l'on peut écrire
$$x = \sum_{j \in F_n} a_j e_j,$$
où
$$a_j = \frac{1}{\sqrt{n}} \sum_{k=1}^n x_k e^{-ik\omega_j},$$
ce qui signifie que l'on peut décomposer le signal x sur les "harmoniques" e_j. On appelle la suite $\{a_j : j \in F_n\}$ la transformée de Fourier discrète du signal x.

2. On définit le *périodogramme* d'un vecteur $x \in \mathbb{C}^n$
$$I_x(\omega_j) = |\langle x, e_j \rangle|^2 = \frac{1}{n} \Big| \sum_{k=1}^n x_k e^{-ik\omega_j} \Big|^2.$$
Montrer que $\|x\|_2^2 = \sum_{j \in F_n} I_x(\omega_j)$.

3. Supposons à partir de maintenant que le signal x est réel, c'est-à-dire $x \in \mathbb{R}^n$. Montrer que
$$x = a_0 e_0 + \sum_{j=1}^{[(n-1)/2]} (a_j e_j + \bar{a}_j e_{-j}) + a_{n/2} e_{n/2},$$
où l'on pose par convention $e_{n/2} = 0$ si n est impair. Si on pose $a_j = r_j e^{i\theta_j}$ avec $r_j > 0$ et $\theta_j \in]-\pi, \pi]$, montrer que l'on peut décomposer le signal sur la base des cosinus et sinus:
$$x = a_0 e_0 + \sum_{j=1}^{[(n-1)/2]} \sqrt{2} r_j (\cos(\theta_j) c_j - \sin(\theta_j) s_j) + a_{n/2} e_{n/2},$$
où l'on a noté
$$c_j = \sqrt{\frac{2}{n}} \begin{bmatrix} \cos(\omega_j) \\ \vdots \\ \cos(n\omega_j) \end{bmatrix} \text{ and } s_j = \sqrt{\frac{2}{n}} \begin{bmatrix} \sin(\omega_j) \\ \vdots \\ \sin(n\omega_j) \end{bmatrix}$$
et en déduire que $\{e_0, c_1, \ldots, c_{[(n-1)/2]}, s_1, \ldots, s_{[(n-1)/2]}, e_{n/2}\}$ forme une base orthonormale de $\mathbb{R}^n$. En déduire une décomposition de $\|x\|_2^2$. Supposons à partir de maintenant, pour simplifier, que n est impair et posons $q = (n-1)/2$.

4. Montrer que
$$2I_x(\omega_j) = \| \operatorname{proj}_{\operatorname{span}(c_j, s_j)}(x) \|^2$$
pour $j = 1, \ldots, q$, où $\| \operatorname{proj}_{\operatorname{span}(c_j, s_j)}(x) \|^2$ est la norme euclidienne au carré de la projection de x dans l'espace vectoriel engendré par c_j et s_j.

5. Supposons maintenant que x est un vecteur aléatoire de loi $N(0, \sigma^2 I_n)$ et posons
$$V_j = \frac{I_x(\omega_j)}{\sigma^2}$$
pour $j = 1, \ldots, q$. Montrer que les variables $V_1, \ldots, V_q$ sont i.i.d et de loi $\chi^2(2)/2$, montrer que cette loi est également la loi $\mathcal{E}(1)$ (exponentielle de paramètre 1), qui admet comme densité $f(v) = e^{-v} \mathbf{1}_{v \geq 0}$ par rapport à la mesure de Lebesgue.

6. Soient $U_1, \ldots, U_{q-1}$ des variables aléatoires i.i.d de loi uniforme sur $[0, 1]$, et posons $(U_{(1)}, \ldots, U_{(q-1)})$ le vecteur $(U_1, \ldots, U_{(q-1)})$ réordonné de façon croissante, c'est-à-dire tel que
$$U_{(1)} \leq \cdots \leq U_{(q-1)}.$$
Montrer que la densité jointe de $(U_{(1)}, \ldots, U_{(q-1)})$ est donnée par
$$g(u_1, \ldots, u_{q-1}) = (q-1)! \mathbf{1}_{0 \leq u_1 \leq \cdots \leq u_{q-1} \leq 1}.$$

7. Posons $S_j = \sum_{k=1}^{j} V_k$ pour $j = 1, \ldots, q-1$. Montrer que $(S_1, \ldots, S_q)$ a pour densité jointe
$$f_{(S_1, \ldots, S_q)}(s_1, \ldots, s_q) = e^{-s_q} \mathbf{1}_{0 \leq s_1 \leq \cdots \leq s_q}$$

8. Montrer que S_q est de loi $\Gamma(q, 1)$, dont la densité est donnée par
$$f_{S_q}(s) = \frac{1}{(q-1)!} s^{q-1} e^{-s} \mathbf{1}_{s \geq 0}.$$

9. En déduire que la densité conditionnelle de $(S_1, \ldots, S_{q-1}|S_q)$ est donnée par

$$f_{(S_1,\ldots,S_{q-1}|S_q)}(s_1, \ldots, s_{q-1}|s_q) = (q-1)!s_q^{-(q-1)}\mathbf{1}_{0\leq s_1\leq\cdots\leq s_q}$$

pour $s_q \geq 0$.

10. Posons

$$Y_j = \frac{S_j}{S_q}$$

pour $j = 1, \ldots, q-1$. Montrer que la densité jointe de $(Y_1, \ldots, Y_{q-1})$ est égale à g.

11. Posons $Y_0 = 0$, $Y_q = 1$ et

$$M_q = \max_{i=1,\ldots,q}(Y_i - Y_{i-1}) = \frac{\max_{j=1,\ldots,q} I(\omega_j)}{\sum_{k=1}^q I(\omega_k)}.$$

Expliquer intuitivement pourquoi on s'intéresse à la variable aléatoire M_q pour détecter une périodicité dans le signal x.

12. Nous cherchons à expliciter la loi de la variable aléatoire M_q. Montrer que la loi de M_q est la loi de la longueur maximale des intervalles $[U_{(i-1)}, U_{(i)}]$ pour $i = 1, \ldots, q$, où $U_{(0)} = 1$, $U_{(q)} = 1$, et $U_{(1)} \leq \cdots \leq U_{(q-1)}$ où $U_1, \ldots, U_{q-1}$ sont des variables i.i.d de loi uniforme sur $[0, 1]$.

13. Posons $G_q(u) = \mathbb{P}(M_q \leq u)$. Montrer que G_q vérifie la relation de récurrence

$$G_q(u) = (q-1)\int_0^u G_{q-1}(1-y)(1-y)^{q-2}dy,$$

pour $q \geq 3$.

14. En déduire que

$$G_q(u) = \sum_{j=0}^{q}(-1)^j\binom{q}{j}(1-ju)_+^{q-1}.$$

avec $x_+ := \max\{0, x\}$ pour tout $x \in \mathbb{R}$.

15. En utilisant la statistique M_q, construisez un test pour l'hypothèse nulle $H_0 : x \sim N(0, \sigma^2)$ contre l'alternative $H_1 : x \sim N(y, \sigma^2)$ où y est un signal avec une périodicité, et expliquer pourquoi on s'attend à ce que le test ait une bonne puissance. Ce test a été introduit par Fisher (encore lui !) pour détecter des périodicités cachées dans des signaux.