HMMA237

Détection de ruptures

Cours: Joseph Salmon *Scribes: Luana Timofte et Anas El Benna*

1 Introduction

1.1 Généralités et motivations

Une partie des éléments peuvent se retrouver dans [Har06].

La détection de ruptures (e.g., : Change Point Detection) est un problème classique en traitement de signal et analyse statistique, où l'on cherche à localiser des changements brutes et (presque) instantanés de la distribution d'une série chronologique.

Ce phénomène peut affecter plusieurs grandeurs telles que la moyenne, la variance, la famille de la distribution des données, les composantes spectrales, etc. Le but est donc de pouvoir analyser notre série temporelle par morceaux, et de détecter automatiquement ces changements.

Exemple. *Contrôles de qualité, finance (détection d'anomalies en bourse, krach), optimiser la consommation électrique, biologie (découpage en blocs de gènes), etc.*

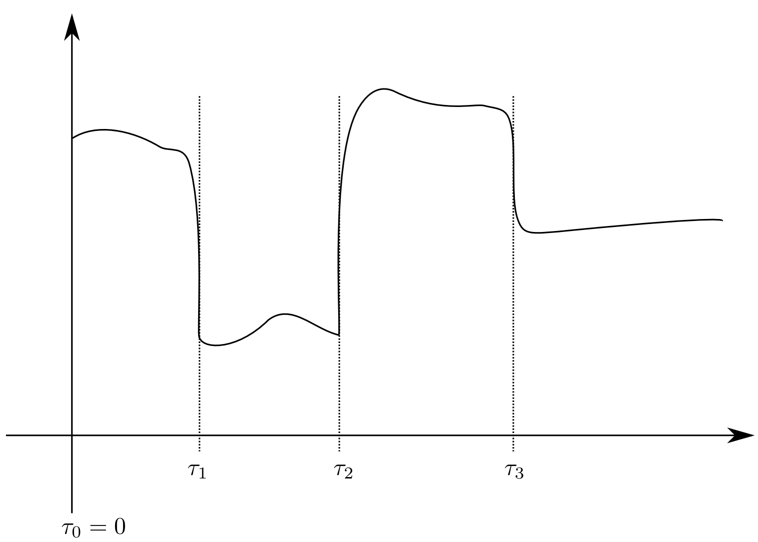


FIGURE 1 – Exemple de série temporelle avec rupture

Remarque. *La figure ci-dessus montre un exemple de série temporelle comportant $R = 3$ ruptures (extrémités exclus) délimitant $R + 1 = 4$ segments.*

2 Modèle mathématique

Definition (Rupture). *Soient un échantillon* $\boldsymbol{X} = (X_1, \ldots, X_n)$ *de variables aléatoires et* $\theta \in \Theta$ *un paramètre représentant les grandeurs affectées par la rupture.*

Pour tout $i \in [\![1, n]\!]$*, on associe à chaque variable* X_i *un paramètre* θ_i*.*

$$\begin{aligned} X_i \text{ est une rupture si } \theta_i \neq \theta_{i+1} \\ X_i \text{ n'est pas une rupture si } \theta_i = \theta_{i+1} \end{aligned} \tag{1}$$

Par convention, X_1 *et* X_n *sont considérées comme des ruptures.*

On suppose que nous avons une série temporelle constante par morceaux, un bruit blanc gaussien, et K ruptures (K connu ici).

On considère le modèle $X_i = f(t_i) + \varepsilon_i$, $\varepsilon_i \underset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ avec :

$$f(t) = \begin{cases} \mu_1, & \text{si } t \in [\tau_0, \tau_1] \\ \vdots \\ \mu_k, & \text{si } t \in [\tau_{K-1}, \tau_K] \end{cases} \tag{2}$$

On prend comme paramètre Θ, la concatenation de toutes les ruptures et des moyennes par blocs :

$$\Theta = (\mu_1, \ldots, \mu_K, \tau_0, \ldots, \tau_K) \in \mathbb{R}^{K+(K+1)} \tag{3}$$

Propriétés :

- t_i : instants
- μ : les moyennes
- τ_k : les instants de rupture ($\tau_0 = 0$ sans perte de généralité)
- K : le nombre de segments, ou le nombre de ruptures
- T : vecteur de segmentation au niveau des ruptures

Remarque. $\mathbb{E}(X_i) = \mu_k$ *si* $t_i \in [\tau_{k-1}, \tau_k]$

2.1 Contexte

La fonction de vraisemblance est fondamentale pour la détection des ruptures. Elle exprime la densité de probabilité du vecteur des paramètres Θ en vue du vecteur des observations $\mathbf{X}$:

On pose $d_j = |\{\underbrace{i \in [\![1, n]\!] : t_i \in [\tau_{j-1}, \tau_j]}_{S_j}\}|$, et l'on a alors :

$$\mathcal{L}(\Theta; X_1, \ldots, X_n) = \prod_{j=1}^{K} \frac{1}{(2\pi\sigma^2)^{\frac{d_j}{2}}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i\in S_j} (\mu_j - X_i)^2\right) \quad (4)$$

$$= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\sum_{j=1}^{K}\sum_{i\in S_j} \frac{(\mu_j - X_i)^2}{2\sigma^2}\right) \quad (5)$$

On passe maintenant à la log-vraisemblance :

$$-\log \mathcal{L} = c^{te} + \underbrace{\sum_{j=1}^{K}\sum_{i\in S_j} \frac{(\mu_i - X_i)^2}{2\sigma^2}}_{J(\mu;\tau)} \quad (6)$$

Alors on estime :

$$\hat{\mu}_j(\tau_{j-1}, \tau_j) = \frac{1}{d_j} \sum_{i\in S_j} X_i \quad (7)$$

2.2 Problème

On aimerait minimiser J ; $\min_{(K,\mu,\tau)} J(\mu;\tau)$. On suppose que σ^2 et K sont connus.

Alors :

$$\min_{(K,\mu,\tau)} \sum_{j=1}^{K}\sum_{i\in S_j} (\mu_i - X_i)^2 \Leftrightarrow \min_{(K,\tau)} \sum_{j=1}^{K}\sum_{i\in S_j} (\hat{\mu}(\tau_{j-1};\tau_j) - X_i)^2 \quad (8)$$

$$\Leftrightarrow \min_{(K,\tau)} \sum_{j=1}^{K}\sum_{i\in S_j} (\frac{1}{d_j}\sum_{i\in S_j} X_i - X_i)^2 \quad (9)$$

2.2.1 Cas d'une seule rupture (K =2)

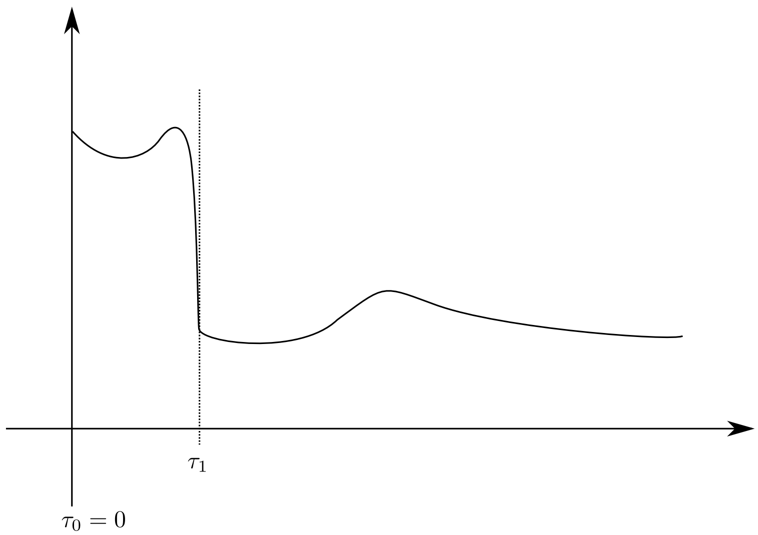
$$f(t) = \begin{cases} \mu_1, & \text{si } t \leqslant \tau_1, \\ \mu_2, & \text{si } t > \tau_1. \end{cases} \quad (10)$$

On a :

$$U(\tau) = \sum_{i:t_i\leqslant\tau} (X_i - \overline{X}_{1:\tau})^2 + \sum_{i:t_i>\tau} (X_i - X_{\tau+1:n})^2 \quad (11)$$

Pour simplifier, on ré-indexe $i \leftarrow t_i$. Et on utilise le principe de la **Programmation dynamique** pour calculer $U(1), \ldots, U(n-1)$:

FIGURE 2 – Exemple de série temporelle avec 1 rupture



et

$$\Sigma_1 = X_1,\ \Sigma_2 = X_1 + X_2, \ldots, \Sigma_n = X_1 + \cdots + X_n \tag{12}$$
$$\Omega_1 = X_n, \Omega_2 = X_n + X_{n-1}, \ldots, \Omega_n = X_n + \cdots + X_1 \ . \tag{13}$$

En pratique on calcule ça avec une relation de récurrence :

$$\Sigma_1 = X_1 \tag{14}$$
$$\Sigma_2 = \Sigma_1 + X_2 \tag{15}$$
$$\vdots \tag{16}$$
$$\Sigma_n = \Sigma_{n-1} + X_n \ . \tag{17}$$

$$\Omega_1 = \Omega_n \tag{18}$$
$$\Omega_2 = \Omega_1 + X_{n-1} \tag{19}$$
$$\vdots \tag{20}$$
$$\Omega_n = \Omega_{n-1} + X_1 \ . \tag{21}$$

Remarque. *On peut remarquer aussi que :* $\Omega_{n-k} = \Sigma_n - \Sigma_k$

Nous calculons ensuite les vecteurs $[\tilde{\Sigma}_1, \ldots, \tilde{\Sigma}_n]$ et $[\tilde{\Omega}_1, \ldots, \tilde{\Omega}_n]$. Avec $\forall k \in [\![1, n]\!]$; $\tilde{\Sigma}_k = \sum_{j=1}^{k} X_j^2$ et $\tilde{\Omega}_k = \sum_{j=n-k+1}^{n} X_j$

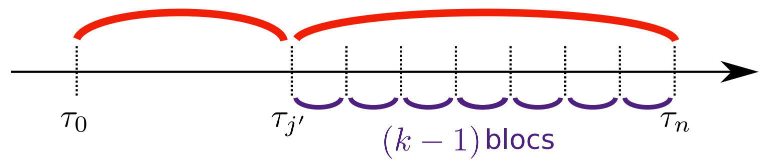


FIGURE 3 – Exemple de programmation dynamique

2.2.2 Algorithme

La programmation dynamique consiste à calculer $U(\tau), \forall\tau$, avec les $\Omega_\tau, \Sigma_\tau, \tilde{\Omega}_\tau, \tilde{\Sigma}_\tau$, et trouver le plus petit.

$$U(\tau_1, \ldots, \tau_n) = \sum_{j=1}^{K} \sum_{i \in S_j} (\hat{\mu}_j - X_i)^2 \tag{22}$$

$$= \sum_{j=1}^{K} u(\tau_{j-1} - \tau_j) \tag{23}$$

Avec :

$$\begin{cases} u(j,n)^{(k)} = & \underset{j'<j<n}{min} [u(j,j') + u(j',n)^{(k-1)}] \\ u(j,n)^{(0)} = & u(j,n) \end{cases} \tag{24}$$

Premier cas :

$$U^{(1)}(j', n) = \min_{j'<j<n} U(j,j') + U(j,n) \tag{25}$$

- **Valeur optimale :** $U^{(k)}(0,n)$

Remarque. ***Perspective :*** *alternative avec des régularisations de type TV (TV-graphe, TV = total variation), sur* $\sum_{i=1}^{n} |z_i - z_{i-1}|$

Différence d'amplitude entre les signaux :

$$\min_z \frac{1}{2} \sum (X_i - z_i)^2 + \lambda \sum |z_i - z_{i-1}| \tag{26}$$

Avec $\lambda > 0$*, un paramètre de régularisation.*

Références

[Har06] Flore Harlé. Détection de ruptures multiples dans des séries temporelles multivariées : application à l'inférence de réseaux de dépendance. PhD thesis, Université de Grenoble Alpes, 2006. 1