

Segmentation de signaux sous contrainte de longueur minimale

Olivier DE CAMBRY

Groupe ESIEE

2,Bd Blaise Pascal, BP 99, 93162 Noisy-Le-Grand Cedex, France

decambro@esiee.fr

Résumé – On considère des signaux aléatoires dans un cadre paramétrique en vue de les segmenter. On utilise un modèle statistique bayésien construit à l'aide de la fonction de vraisemblance. Les paramètres inconnus sont les paramètres du modèle statistique sur chaque segment, le nombre de segments et les instants de changement. On trouve ces paramètres par une méthode hors ligne en résolvant un problème d'optimisation du type M.A.P (*Maximum a posteriori*). Une contrainte sur la longueur des segments peut être ajoutée pour que ce problème ait une solution. Plusieurs algorithmes ont été essayés sur des simulations.

Abstract – This paper presents a method for segmentation of random signals in parametric statistical model. We construct a maximum likelihood criterion with a bayesian model and we use the M.A.P method (*Maximum a posteriori*). The unknown parameters are the numbers of segments, the instants of change and statistical model parameters between changes. A constraint about the segments length can be introduced to obtain a solution to the optimisation problem. Several algorithms have been experimented with simulations to estimate *off-line* parameters.

1 Modèle statistique

1.1 Introduction

On considère un signal aléatoire $X = (X_t)_{t=1,\dots,T}$, réel stationnaire par morceaux. Cela signifie qu'il existe des instants dont le nombre, ν , et les valeurs, $(t(1), \dots, t(\nu))$ sont inconnus auxquels le processus change de loi : on dit que ce sont des instants de changement ou de rupture.

On suppose que entre les instants $t(k)$ et $t(k+1)$ la restriction du signal $X(k) = (X_t)_{t=t(k),\dots,t(k+1)-1}$, est stationnaire pour tout $k = 1, \dots, \nu$ (on notera, par convention, $t(\nu+1) = T+1$).

Le problème est la décomposition de ce signal :

1. Trouver le nombre et la valeur des instants de changements de modèle.
2. Identifier les lois des restrictions à chaque segment du signal.

La méthode proposée n'est pas séquentielle, mais globale, hors ligne. Elle est fondée sur un modèle de processus aléatoires cachés dont les états sous jacents sont au nombre de 2. Elle s'appuie sur le travail de Marc Lavielle dans [3] en apportant des éléments nouveaux notamment la possibilité d'imposer une contrainte sur la longueur d'un segment.

1.2 Modèle statistique

On note, $R = (R_t)_{t=1,\dots,T}$, le processus aléatoire à 2 états :

$$\begin{aligned} R_t &= 0 && \text{si il n'y a pas de changement en } t \\ R_t &= 1 && \text{si il y a changement à l'instant } t \end{aligned} \quad (1)$$

Lorsqu'il y a changement en t le nouveau segment démarre à la valeur t .

Par cohérence on fixera donc, $R_1 = 1$.

D'autre part dans ce modèle les instants de changement deviennent aléatoires. Ils sont notés $(T(1), \dots, T(\nu))$ et ils sont définis par :

$$\begin{aligned} T(1) &= 1 \\ &\text{et} \\ T(k) &= \inf \{T(k-1) < t \leq T, R(t) = 1\} \\ &\text{pour } k = 2, \dots, \nu \end{aligned} \quad (2)$$

Nous supposons dans un premier temps que R est une suite de variable aléatoires indépendantes.

La méthode utilisée est une méthode bayésienne.

On notera pour $t=2,\dots,T$:

$$\lambda = P(R_t = 1) \quad (3)$$

Dans ces conditions la vraisemblance de R est donnée par :

$$L(1, r_2, \dots, r_T; \lambda) = \lambda^{\nu-1} (1-\lambda)^{T-\nu} \quad (4)$$

On notera β la fonction :

$$\beta(r, \lambda) = -\ln L(1, r_2, \dots, r_T; \lambda) \quad (5)$$

Nous supposons que pour $k = 1, \dots, \nu$, la loi de $X(k)$ conditionnellement à R est paramétrique de paramètre $\theta(k)$.

On note $L(x_{t(k-1)+1}, \dots, x_{t(k)}; \theta(k))$, la vraisemblance de la loi de $X(k)$ et $\theta = (\theta(1), \dots, \theta(\nu))$, le vecteur des paramètres.

1.3 Construction du critère

Le principe du maximum a posteriori (M.A.P) nous conduit à estimer la suite des états sous-jacent, r , le vecteur des paramètres θ , en maximisant $P(R = r/X = x)$.

De manière classique [2] on obtient alors :

$$P(R = r/X = x) = \frac{1}{Z} \exp -U(\lambda, r, \theta) \quad (6)$$

La fonction U est une fonction d'énergie définie par les équations suivantes :

$$\begin{aligned} \nu(r) &= \sum_{t=1}^T r_t \\ \beta(r, \lambda) &= -(\nu - 1) \ln \lambda - (T - \nu) \ln (1 - \lambda) \end{aligned} \quad (7)$$

$$V(x; r, \theta) = - \sum_{k=1}^{\nu} \ln L(x_{t(k)}, \dots, x_{t(k+1)-1}; \theta(k))$$

$$U(x; \lambda, r, \theta) = V(x; r, \theta) + \beta(r, \lambda)$$

La fonction U est donc la somme de 2 fonctions :

1. La fonction V qui mesure la distance entre l'observation et le modèle
2. La fonction β qui mesure la qualité du modèle

La solution du problème est donc :

$$(\hat{\lambda}, \hat{r}(\hat{\lambda}), \hat{\theta}(\hat{r}, \hat{\lambda})) = \arg \min_{\lambda, r, \theta} U(x; \lambda, r, \theta) \quad (8)$$

On notera que λ n'apparaît que dans la fonction β , alors que θ n'apparaît que dans la fonction V . Ces paramètres peuvent donc être calculés préalablement.

La minimisation en λ se fait simplement. On obtient :

$$\begin{aligned} \hat{\lambda} &= \frac{\nu - 1}{T - 1} \\ \beta^*(\nu) &= \beta(r, \hat{\lambda}) \\ &= (T - 1) \ln (T - 1) - (\nu - 1) \ln (\nu - 1) \\ &\quad - (T - \nu) \ln (T - \nu) \end{aligned} \quad (9)$$

Dans de nombreux cas il existe une formule explicite pour calculer l'estimateur $\hat{\theta}$ de θ à valeur de r fixée. On note alors :

$$V^*(x; r) = V(x; \hat{\theta}(r)) \quad (10)$$

Lorsque ce calcul est possible, la fonction d'énergie U devient :

$$\begin{aligned} U^*(x; r) &= U(x; \hat{\lambda}, \hat{\theta}(r)) \\ &= V^*(x; r) + \beta^*(\nu(r)) \end{aligned} \quad (11)$$

Nous devons alors résoudre le problème de minimisation suivant :

$$\hat{r} = \arg \min_{\lambda, r, \theta} U^*(x; r) \quad (12)$$

2 Exemple

2.1 Changements de moyenne

Considérons l'exemple d'un signal aléatoire gaussien avec des sauts de moyennes.

Le modèle est le suivant :

On note $\epsilon = (\epsilon_t)_{t=1, \dots, T}$ un bruit gaussien centré, de variance σ^2 et $\mu = (\mu_t)_{t=1, \dots, T}$ une fonction constante par morceaux.

On suppose que μ prend ν valeurs successives.

On notera m le vecteur ainsi défini :

$$m = (\mu(1), \dots, \mu(\nu)) \quad (13)$$

Le signal x est alors défini pour $t = 1, \dots, T$ par :

$$\begin{aligned} x_t &= \mu_t + \epsilon_t \\ &= \mu(k) + \epsilon_t \\ &\text{pour } t = t(k) \dots, t(k+1) - 1 \text{ et } k = 1, \dots, \nu \end{aligned} \quad (14)$$

On suppose que la variance σ^2 est connue.

La fonction V est alors définie par :

$$V(x; r, m) = \frac{1}{2 \sigma^2} \sum_{k=1}^{\nu} \sum_{t=t(k)}^{t(k+1)-1} (x_t - \mu(k))^2 \quad (15)$$

La moyenne sur un segment est alors estimée par la moyenne empirique. On note pour $k = 1, \dots, \nu$

$$\hat{\mu}(k) = \frac{1}{t(k+1) - t(k)} \sum_{t=t(k)}^{t(k+1)-1} x_t \quad (16)$$

$$\hat{m} = (\hat{\mu}(1), \dots, \hat{\mu}(\nu))$$

$$V^*(x; r) = V(x; r, \hat{m})$$

Le modèle que nous avons obtenu peut aussi être utilisé pour des variables non gaussiennes. La fonction V est alors simplement le contraste des moindres carrés.

2.2 Changements de moyenne et de variance

Dans ce cas, la variance change aussi de valeur aux instants de changements. On introduit

$$s^2 = (\sigma^2(1), \dots, \sigma^2(\nu)) \quad (17)$$

On a alors

$$\hat{\mu}(k) = \frac{1}{t(k+1) - t(k)} \sum_{t=t(k)}^{t(k+1)-1} x_t \quad (18)$$

$$\hat{\sigma}^2(k) = \frac{1}{t(k+1) - t(k)} \sum_{k=1}^{\nu} \sum_{t=t(k)}^{t(k+1)-1} (x_t - \hat{\mu}(k))^2 \quad (19)$$

$$V(x; r, m, s^2) = T + \frac{1}{2} \sum_{k=1}^{\nu} (t(k+1) - t(k)) \ln \hat{\sigma}^2(k) \quad (20)$$

3 Contrainte sur le modèle

3.1 Nécessité d'une contrainte

Dans les paragraphes précédents on a supposé que le processus des états sous-jacent R , est constitué de variables aléatoires indépendantes. En fait lorsqu'il n'y a pas beaucoup d'observations, l'algorithme a tendance à tronçonner le signal en de nombreux segments. Cela est possible car l'hypothèse d'indépendance entraîne la possibilité d'avoir des ruptures très rapprochées. Par exemple dans le cas du changement de moyenne, la fonction d'énergie est minimum (et vaut 0), lorsque chaque point représente un segment.

Plusieurs types de contraintes peuvent être envisagés :

1. Fixer le nombre de segments :

On peut alors utiliser les mêmes algorithmes que ceux que nous décrivons dans la suite en les arrêtant à la valeur choisie. Mais nous souhaitons ici conserver le nombre de changement comme une inconnue.

2. Modifier la fonction d'énergie

C'est la solution préconisée dans [3].

La fonction $\beta(r)$ est remplacée par :

$$\beta_0(r) = b\nu \quad (21)$$

La fonction β_0 ne dépend donc que du nombre de changement et d'un paramètre de réglage b dont la valeur est fixé par l'expérimentateur en fonction de ses attentes. L'inconvénient de cette approche est que le paramètre b est un paramètre abstrait peu accessible dans la pratique.

3. Ajouter une contrainte sur la taille des segments

On ajoute une contrainte supplémentaire que l'on peut choisir en fonction de la connaissance préalable que l'on a du signal que l'on étudie, qui est d'imposer une longueur minimale pour chaque segment notée τ . La solution que l'on cherche n'est alors pas le minimum global de la fonction d'énergie mais le premier minimum local en partant de 0.

3.2 Contrainte de longueur minimale des segments

Le processus R devient alors un processus aléatoire semi-markovien.

En fait lorsque $\tau = 0$ le processus est composé de variables indépendantes (comme en 1.2).

Si $\tau = 1$ le processus est une chaîne de Markov à 2 états, 0 et 1, de matrice de transition :

$$P = \begin{bmatrix} 1 - \lambda & \lambda \\ 1 & 0 \end{bmatrix} \quad (22)$$

De plus on a toujours $R_1 = 1$;

Pour $\tau > 1$, le processus R est un processus semi-markovien, à temps discret que l'on supposera à 2 états 0 et 1, dont la transition est définie de la manière suivante. On note $(U_j)_{j=1,2,\dots}$, la suite des instants de changement d'états du processus R et $(\rho_j)_{j=1,2,\dots}$ la suite des états

successifs.

On a alors

$$\rho_j = R_{u_j} \quad (23)$$

De plus par convention $U_1 = 1$ et $\rho_1 = 1$.

On peut alors définir les probabilités de transition du processus :

On note

$$A_j = (u_1, \dots, u_{j-1}; r_1, \dots, r_{j-1}) \quad (24)$$

la suite des états successifs jusqu'à l'instant $j - 1$ et

$$\pi(u_j, r_j; u_{j-1}, r_{j-1}) = P(U_j = u_j, \rho_j = r_j / A_j) \quad (25)$$

la probabilité de transition du processus.

Alors

$$\pi(u_{j-1} + 1, 1; u_{j-1}, 1) = 1 \quad (26)$$

$$\pi(u_{j-1} + 1, 1; u_{j-1}, 1) = 1 \quad (27)$$

et

$$\pi(u_{j-1} + t, 1; u_{j-1}, 0) = \begin{cases} \lambda & \text{si } t \geq \tau \\ 0 & \text{si } 1 \leq t < \tau \end{cases} \quad (28)$$

$$\pi(u_{j-1} + t, 0; u_{j-1}, 0) = \begin{cases} 1 - \lambda & \text{si } t \geq \tau \\ 1 & \text{si } 1 \leq t < \tau \end{cases} \quad (29)$$

On peut alors remarquer que pour $j = 2k$, $\rho_{2k} = 0$ et pour $j = 2k + 1$, $\rho_{2k+1} = 1$.

Cela signifie que le processus R est entièrement déterminé par la suite des temps $(u_j)_{j=1,2,\dots,\nu}$, sachant que à l'instant u_{2k} il y a transition de l'état 0 à l'état 1 et à l'instant u_{2k+1} il y a transition de l'état 1 à l'état 0.

La suite $(u_2, \dots, u_{2\nu})$ est la suite des instants de rupture $(t(1), \dots, t(\nu))$. La loi de u_n est déterminée pour $m = 1, \dots, \nu$ par :

$$P(u_{2k} - u_{2k-1} = t) = \begin{cases} \lambda(1 - \lambda)^{t-\tau} & \text{si } t \geq \tau \\ 0 & \text{si } 1 \leq t < \tau \end{cases} \quad (30)$$

et

$$P(u_{2k+1} - u_{2k} = 1) = 1 \quad (31)$$

La vraisemblance de ce processus se calcule aisément.

On note

$$M_\tau = \max(T, T(\nu) + \tau) \quad (32)$$

On a alors :

$$\begin{aligned} \beta(r, \lambda) &= -\ln L(1, r_2, \dots, r_T; \lambda) \\ &= -(\nu - 1) \ln \lambda \\ &\quad - (M_\tau - (\tau + 1)\nu) \ln(1 - \lambda) \end{aligned} \quad (33)$$

On peut alors calculer l'estimateur du maximum de vraisemblance $\hat{\lambda}$ de λ et la fonction β^* associée.

$$\hat{\lambda} = \frac{\nu - 1}{M_\tau - \tau\nu - 1} \quad (34)$$

$$\begin{aligned} \beta^*(\nu, T(\nu), \tau) &= \beta(r, \hat{\lambda}) \\ &= (M_\tau - (\tau + 1)\nu) \ln(M_\tau - (\tau + 1)\nu) \\ &\quad - (\nu - 1) \ln(\nu - 1) \\ &\quad - (M_\tau - \tau\nu - 1) \ln(M_\tau - \tau\nu - 1) \end{aligned} \quad (35)$$

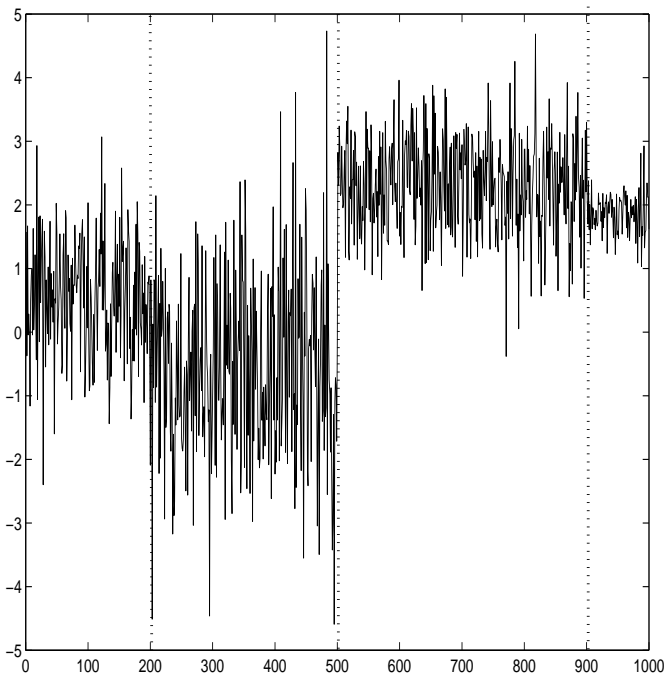


FIG. 1 – Saut de moyenne et variance

4 Simulation numérique

Numériquement on peut calculer la solution de ce problème de 3 manières :

1. L'utilisation de l'algorithme de Fischer basé sur la programmation dynamique [1] : cet algorithme a le mérite de donner une solution exacte et donc une segmentation optimale mais a le défaut d'être très long.
2. L'utilisation d'algorithme de recuit simulé avec différentes stratégies [2],[3] : le résultat est variable et surévalue souvent le nombre de changements.
3. La segmentation par instant de changement successif, consistant à calculer la position de l'instant de changement minimisant la fonction d'énergie, puis répéter l'opération pour trouver l'instant de changement suivant minimisant la fonction d'énergie, sans remettre en cause les instants déjà obtenus. On s'arrête lorsque la fonction d'énergie augmente. Cette méthode est très rapide et donne sur toutes les simulations essayées le résultat optimal c'est à dire celui donné par l'algorithme de Fischer.

La figure 1 représente un signal avec 3 changements de moyenne et de variance aux instants : 200, 500 et 900.

Pour $\tau = 50$, on a représenté la fonction d'énergie minimale en fonction du nombre de segments dans la figure 2.

On a obtenu les résultats suivants :

1. **Algorithme de Fischer**
Changements : $r = [1\ 199\ 409\ 501\ 901\ 982]$
Temps de calcul : $t \simeq 2$ heures
2. **Recuit simulé**
Nombre de changements : $\nu = 15$

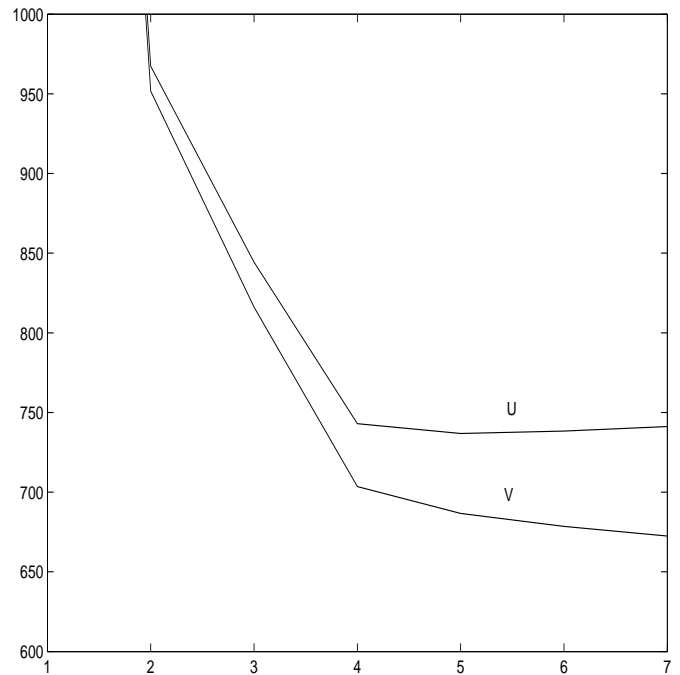


FIG. 2 – Fonction d'énergie

3. Détection successive

Changements : $r = [1\ 199\ 409\ 501\ 901\ 982]$

Temps de calcul : $t = 17s$

On peut remarquer que si on avait mis comme contrainte le nombre de segments on aurait obtenu : $r = [1\ 199\ 501\ 901]$ et que d'autre part si $\tau = 1$, on n'obtient pas de solution, la fonction d'énergie minimale étant décroissante.

Références

- [1] Fisher W.D., *On Grouping for Maximum Homogeneity*, J. of American Statist. Assoc., Vol. 53, p.789-798, 1958.
- [2] Guyon X., *Méthodes numériques par chaînes de Markov*, polycopié du Cours de l'école d'été de Mathématiques, Mérida, Venezuela, Septembre 1999.
- [3] Lavielle M., *Optimal Segmentation of Random Processes*, IEEE Transactions on Signal Processing, Vol. 46, N°5, 1998.
- [4] Lavielle M., *An Application of MCMC Methods for the Multiple Change-points Problem*, Signal Processing, 81, p.39-53, 2001.