

# DÉTECTION DE RUPTURES MULTIPLES À NOYAUX

Damien GARREAU <sup>1</sup> & Sylvain ARLOT <sup>2</sup>

<sup>1</sup> *Sierra project-team*

*Laboratoire d'informatique de l'École Normale Supérieure*

*(CNRS / ENS / INRIA UMR 8548)*

*45 rue d'Ulm, 75005 Paris, France*

*damien.garreau@ens.fr*

<sup>2</sup> *Laboratoire de Mathématiques d'Orsay,*

*Univ. Paris-Sud, CNRS, Université Paris-Saclay*

*91405 Orsay, France*

*sylvain.arlot@math.u-psud.fr*

**Résumé.** La détection de ruptures multiples a pour objet la détection d'un nombre inconnu de changements abrupts dans la distribution des observations d'une série temporelle. Dans ce travail, nous étudions la méthode de détection de ruptures multiples à noyaux introduite par Arlot, Celisse et Harchaoui (2012). À la différence des méthodes actuelles, qui sont calibrées pour détecter des changements dans la moyenne ou la variance d'un signal vectoriel, la méthode que nous étudions peut à la fois détecter d'autres types de changements dans la distribution des observations, et traiter des données structurées. Nous montrons, sous des hypothèses raisonnables, que cette méthode retrouve avec grande probabilité le vrai nombre de ruptures dans un signal i.i.d., et qu'elle localise les ruptures avec une précision que l'on peut quantifier en fonction des paramètres du problème.

**Mots-clés.** détection de ruptures, méthodes à noyaux, inégalités de concentration.

**Abstract.** Multiple change-point detection's goal is to detect an unknown number of steep changes in the law of the observations of a time series. In this work, we study the kernel change-point detection procedure introduced by Arlot, Celisse et Harchaoui (2012). Unlike existing methods, which are designed to find changes in the mean or the variance of a vector-valued signal, the focus of this study can both detect other kinds of changes in the distribution of the observations and deal with structured data. Under reasonable assumptions, we show that kernel change-point detection retrieves the true number of change-points in an i.i.d. signal, and that these change-points can be localized with a precision depending on the problem's parameters.

**Keywords.** change-point detection, kernel methods, concentration inequalities.

## 1 Introduction

La détection de ruptures multiples a pour but la localisation des changements brusques dans une série temporelle, leur nombre n'étant pas supposé connu a priori. La plupart

des méthodes existantes s'intéressent à des changements dans la moyenne ou la variance du signal. Mais il peut arriver qu'on ait à détecter d'autres types de ruptures, comme par exemple un changement dans le mode de la distribution des observations. Il est alors nécessaire de trouver une représentation des données adaptée à cette situation. Une méthode de détection de ruptures multiples à noyaux (que l'on notera désormais KCP) a été proposée par Arlot, Celisse et Harchaoui (2012). Implicitement, KCP représente chaque observation par un élément d'un espace de Hilbert à noyau reproduisant, via une transformation non linéaire.

La méthode KCP repose sur la sélection de modèles par pénalisation. Plus précisément, les modèles correspondent à des segmentations de  $\{1, \dots, n\}$ , où  $n$  est le nombre d'observations. On sélectionne alors le modèle qui minimise la somme du risque quadratique et d'une pénalité. Dans la suite, nous considérons KCP avec une pénalité de la forme

$$\text{pen}(\tau) := \frac{CD_\tau M^2}{n} \left( 1 + \sqrt{c_1 + \log \frac{n}{D_\tau}} \right)^2, \quad (1)$$

où  $D_\tau$  correspond au nombre de segments du modèle, et  $C, M, c_1$ , sont des constantes positives. La pénalité (1) a été proposée par Birgé et Massart (2001), et utilisée notamment par Lebarbier (2005) dans le cadre de la détection de ruptures de moyenne pour des données univariées. Une inégalité oracle a été démontrée pour KCP (Arlot, Celisse et Harchaoui, 2012). Nous allons ici plus loin dans l'analyse, en démontrant que KCP identifie avec grande probabilité le nombre de ruptures et leurs positions.

## 1.1 Notations

Considérons  $\mathcal{X}$  un espace mesurable dans lequel nous observons  $X_1, \dots, X_n \in \mathcal{X}$  des variables aléatoires i.i.d. Pour tout  $i \in \{1, \dots, n\}$ , on note  $P_{X_i}$  la loi de  $X_i$ . Le problème de détection de ruptures peut se formuler ainsi : étant donné  $(X_i)_{1 \leq i \leq n}$ , notre objectif est de retrouver les changements brusques dans la suite  $P_{X_1}, \dots, P_{X_n}$ .

Une situation importante qu'on peut garder en tête est celle où les  $X_i$  correspondent aux observations aux instants  $t_i = i/n$  d'un processus aléatoire à temps continu sur  $[0, 1]$ , stationnaire sur les  $[t_\ell^*, t_{\ell+1}^*)$  pour tout  $\ell \in \{0, \dots, D^* - 1\}$  avec  $0 = t_0^* < t_1^* < \dots < t_{D^*}^* = 1$  fixés à l'avance. L'objectif est alors de retrouver  $D^*$  et d'estimer  $t_1^*, \dots, t_{D^*-1}^*$  aussi précisément que possible. En particulier, on souhaite estimer ces quantités de manière consistante lorsque  $n$  tend vers  $+\infty$ , situation que l'on nommera « cadre asymptotique » dans ce qui suit.

Pour chaque  $D \in \{1, \dots, n + 1\}$ , nous définissons

$$\mathcal{T}_n^D := \{(\tau_0, \dots, \tau_D) \in \mathbb{N}^{D+1}, 0 = \tau_0 < \tau_1 < \dots < \tau_D = n\},$$

l'ensemble des *segmentations* de  $\{1, \dots, n\}$  en  $D_\tau = D$  segments. Enfin, nous appelons  $\tau_1, \dots, \tau_{D-1}$  les *points de rupture*.

Soit  $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  un noyau semi-défini positif, c'est-à-dire une fonction mesurable telle que pour tout  $x_1, \dots, x_n \in \mathcal{X}$ , la matrice  $(k(x_i, x_j))_{1 \leq i, j \leq n}$  est positive semi-définie. Parmi de nombreux exemples, mentionnons :

- le noyau linéaire  $k^{\text{lin}}(x, y) = \langle x, y \rangle_{\mathbb{R}^d}$ ,
- le noyau gaussien  $k^{\text{G}}(x, y) = \exp(-\|x - y\|^2 / (2h^2))$ , où  $h > 0$  est un paramètre d'échelle.

## 1.2 Critère des moindres carrés à noyaux

Soit  $\ell \in \{1, \dots, D\}$  et  $\tau \in \mathcal{T}_n^D$ , nous définissons la moyenne des observations sur le  $\ell$ -ième segment de  $\tau$  par  $\bar{X}_{\tau_{\ell-1}+1, \tau_\ell}$ . Une mesure naturelle de la qualité d'une segmentation, lorsque  $\mathcal{X} = \mathbb{R}$  est le critère des moindres carrés,

$$\hat{\mathcal{R}}_n(\tau) := \frac{1}{n} \sum_{\ell=1}^D \sum_{i=\tau_{\ell-1}+1}^{\tau_\ell} (X_i - \bar{X}_{\tau_{\ell-1}+1, \tau_\ell})^2. \quad (2)$$

L'algorithme KCP utilise une version « kernelisée » de ce critère,

$$\hat{\mathcal{R}}_n(\tau) := \frac{1}{n} \sum_{i=1}^n k(X_i, X_i) - \frac{1}{n} \sum_{\ell=1}^D \left[ \frac{1}{\tau_\ell - \tau_{\ell-1}} \sum_{i=\tau_{\ell-1}+1}^{\tau_\ell} \sum_{j=\tau_{\ell-1}+1}^{\tau_\ell} k(X_i, X_j) \right]. \quad (3)$$

Il est aisé de vérifier que les définitions (2) et (3) coïncident lorsque  $\mathcal{X} = \mathbb{R}$  et  $k = k^{\text{lin}}$ .

## 1.3 Algorithme

Intuitivement, une bonne segmentation doit minimiser le critère des moindres carrés (3). Cependant, si l'on ne contraint pas le nombre de segments, c'est systématiquement la segmentation  $\tau = (0, 1, \dots, n)$  qui réalisera ce minimum, ce qui n'est pas souhaitable. Une solution est de *pénaliser* le critère (3), par exemple en fonction du nombre de segments de la segmentation. Nous considérons une pénalité de la forme (1). L'algorithme KCP choisit alors la segmentation  $\hat{\tau}$  solution du problème de minimisation

$$\hat{\tau} \in \arg \min_{\tau \in \mathcal{T}_n} \left\{ \hat{\mathcal{R}}_n(\tau) + \text{pen}(\tau) \right\}, \quad (4)$$

où  $c_1$  et  $c_2$  sont des constantes positives à déterminer.

La Figure 1 décrit l'algorithme qui permet de résoudre le problème (4).

**Entrées :**  $X_1, \dots, X_n \in \mathcal{X}$ ,  
 $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ ,  
 $c_1, c_2 > 0$ ,  
 $D_{\max} \in \{1, \dots, n-1\}$ .  
**Étape #1 :**  $\forall D \in \{1, \dots, D_{\max}\}$ , trouver  $\hat{\tau}(D) \in \arg \min_{\tau \in \mathcal{T}_n^D} \left\{ \hat{\mathcal{R}}_n(\tau) \right\}$   
(par programmation dynamique)  
**Étape #2 :** Trouver  
 $\hat{D} \in \arg \min_{1 \leq D \leq D_{\max}} \left\{ \hat{\mathcal{R}}_n(\hat{\tau}(D)) + \frac{CM^2D}{n} \left( 1 + \sqrt{c_1 + \log \frac{n}{D}} \right)^2 \right\}$   
**Sortie :**  $\hat{\tau} = \hat{\tau}(\hat{D})$ .

FIGURE 1 – L’algorithme KCP de détection de ruptures à noyaux.

## 2 Résultats théoriques

### 2.1 Hypothèses

À tout noyau défini positif  $k$ , on peut associer un espace de Hilbert à noyau reproduisant  $\mathcal{H}$  (nous renvoyons à Aronszajn (1950) pour la théorie des noyaux définis positifs). On peut alors voir l’algorithme KCP comme la recherche, par moindres carrés pénalisés, de ruptures dans la « moyenne »  $\mu_i^*$  de  $k(X_i, \cdot) \in \mathcal{H}$ . Formellement,  $\mu_i^*$  peut se définir comme la fonction  $\mathcal{X} \rightarrow \mathbb{R}$  qui à  $x \in \mathcal{X}$  associe  $\mathbb{E}[k(X_i, x)]$ . Cette définition est possible lorsque  $k$  vérifie une hypothèse technique, ce qui est le cas pour tous les noyaux classiques.

Les garanties théoriques que nous apportons pour KCP nécessitent également une hypothèse sur le noyau. Une possibilité est de supposer que

$$\exists M \in (0, +\infty), \quad \forall 1 \leq i \leq n, \quad k(X_i, X_i) \leq M^2. \quad (\text{Db})$$

Nous faisons aussi parfois l’hypothèse suivante, plus faible que (Db),

$$\exists V \in (0, +\infty), \quad \forall 1 \leq i \leq n, \quad \mathbb{E}[k(X_i, X_i)] \leq V. \quad (\text{V})$$

Remarquons que (Db) implique (V) avec  $V = M^2$ , et que le noyau gaussien vérifie (Db) avec  $M = 1$ .

## 2.2 Nombre de ruptures

Soit  $\underline{\Delta} := \inf_{\mu_i^* \neq \mu_{i+1}^*} \|\mu_i^* - \mu_{i+1}^*\|_{\mathcal{H}}$  la taille du plus petit saut de  $\mu^*$  dans  $\mathcal{H}$  et  $\underline{\Delta} := n^{-1} \inf_{\ell} |\tau_{\ell}^* - \tau_{\ell-1}^*|$  la taille (normalisée) du plus petit segment de  $\tau^*$ . Si  $\underline{\Delta}$  est suffisamment grand et si la constante  $C$  devant la pénalité est également suffisamment grande, alors l'algorithme KCP retrouve le vrai nombre de ruptures dans  $\mu^*$ , comme le montre le résultat suivant.

**Théorème 1.** *Il existe deux constantes numériques  $L_1, L_2 > 0$  telles que l'on a le résultat suivant. Soit  $y \geq 1$ . Supposons que (Db) est vérifiée et que  $\text{pen}$  est définie par (1) avec  $c_1 = 4 + y$  et  $C > L_1 D^*(y + \log D^*)$ . Supposons également que*

$$\frac{\underline{\Delta}^2}{M^2} > \frac{L_2 C D^*}{n \underline{\Delta}} \left( y + \log \frac{n}{D^*} \right). \quad (5)$$

Alors  $\mathbb{P}(\widehat{D} = D^*) \geq 1 - e^{-y}$ .

Remarquons que dans le cadre asymptotique, si l'on prend  $y = 2 \log(n)$ , le terme de droite dans (5) tend vers 0 comme  $\log^2(n)/n$ , tandis que le terme de gauche reste fixe; ainsi,  $\widehat{D}$  tend vers  $D^*$  presque sûrement.

## 2.3 Localisation des ruptures

Dans cette section, nous ne considérons que des segmentations qui appartiennent à

$$\mathcal{T}_n^{D^*}(\lambda) := \left\{ \tau \in \mathcal{T}_n^{D^*}, \inf_{\ell} |\tau_{\ell} - \tau_{\ell-1}| \geq n\lambda \right\},$$

pour un certain  $\lambda > 0$ . En d'autres termes, ce sont des segmentations qui ont le bon nombre de ruptures et telles que la longueur de leur plus petit segment est plus grande qu'un seuil fixé à l'avance. Afin de s'intéresser à la position des ruptures estimées par rapport aux vraies ruptures (celles de  $\mu^*$ ), il faut définir une *distance* entre segmentations. Nous considérons la distance suivante entre segmentations en  $D^*$  segments :

$$\forall \tau^1, \tau^2 \in \mathcal{T}_n^{D^*}, \quad \delta_{\infty}(\tau^1, \tau^2) := \frac{1}{n} \max_{1 \leq i \leq D^*-1} |\tau_i^1 - \tau_i^2|. \quad (6)$$

Notons  $\overline{\Delta}$  la taille du plus grand saut de  $\mu^*$  dans  $\mathcal{H}$ , par analogie à  $\underline{\Delta}$ . Le résultat suivant montre que la minimisation du risque empirique « kernelisé » sur  $\mathcal{T}_n^{D^*}(\lambda)$  permet de retrouver les positions des vraies ruptures.

**Théorème 2.** *Supposons que (V) est vérifiée et que  $\widehat{\tau} \in \mathcal{T}_n^{D^*}(n^{-1/2})$  minimise  $\widehat{\mathcal{R}}_n(\cdot)$ . Alors il existe des constantes  $K_1, K_2 > 0$ , ne dépendant que de  $\underline{\Delta}$  et  $\overline{\Delta}$ , telles que, pour tout  $0 < z/\sqrt{n} < \underline{\Delta}$  et  $n$  suffisamment grand,*

$$\mathbb{P} \left( \delta_{\infty}(\widehat{\tau}, \tau^*) \geq \frac{z}{\sqrt{n}} \right) \leq \frac{K_1}{z} + \frac{K_2}{z^2}.$$

Dans le cadre asymptotique, en prenant  $z = z_n \rightarrow +\infty$ , on obtient que  $\delta_\infty(\hat{\tau}, \tau^*)$  converge en probabilité vers 0 à une vitesse  $n^{-1/2}$ . En combinant les Théorèmes 1 et 2, on obtient que sous l'hypothèse (Db), avec grande probabilité, l'algorithme KCP (légèrement modifié) retrouve le nombre de ruptures exact et localise chacune des vraies ruptures avec une précision de l'ordre de  $n^{-1/2}$ .

## Bibliographie

- [1] Arlot, S., Celisse, A. et Harchaoui, Z. (2012), Kernel change-point detection. <http://arxiv.org/abs/1202.3878>.
- [2] Aronszajn, N. (1950), Theory of reproducing kernels, *Transactions of the American mathematical society*, 337-404.
- [3] L. Birgé and P. Massart, Gaussian model selection. *Journal of the European Mathematical Society*, 3(3) :203–268, 2001.
- [4] É. Lebarbier. Detecting multiple change-points in the mean of a Gaussian process by model selection. *Signal Proces.*, 85 :717–736, 2005.