Détection d'une rupture en ligne - méthode FOCuS

Synthèse et présentation des résultats du projet d'Algorithmique

Sarugan SRIHARAN

M2 Data Science : Santé, Assurance, Finance

Jeudi 09 Février 2023

Table of Contents

Introduction

- Détection de rupture : sujet de recherche important dans de nombreux secteurs qui nécessitent la surveillance en temps réel des données.
- But : Déterminer automatiquement si un comportement change dans les données temporelles.

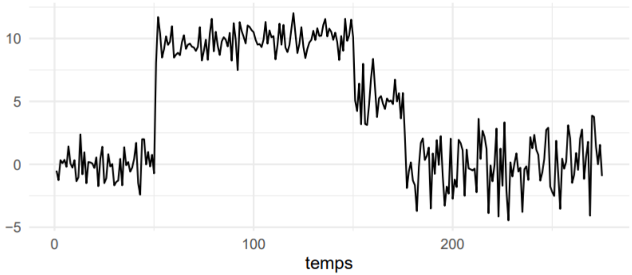

Figure: Données générées aléatoirement

Objectifs

Implémenter l'algorithme FOCuS, proposé par Romano et al. 2022, en R et C++ et de comparer ses performances (algorithmiques et statistiques) à d'autres algorithmes basés sur des approches statistiques

Principe

effectue un test statistique sur chaque point de la série temporelle en divisant la série en deux parties à gauche et à droite du point considéré

- Test de Kolmogorov-Smirnov de comparaison
- Test de la somme des rangs (Wilcoxon)

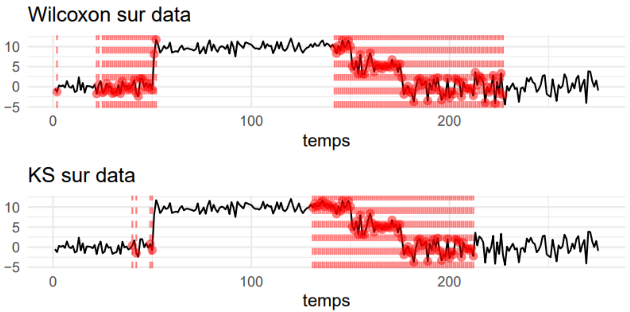

Figure: Test de comparaison sur les données

- **Compléxité :**
  - KS : $O(T^2 log \, \frac{T}{2})$
  - Wilcoxon : $O(T^2)$

Algorithmes naïfs

Test de comparaison connaissant le nombre de ruptures

Principe

On suppose à présent connaître le nombre de ruptures K dans la série temporelle. On cherche alors astucieusement ces points (avec KS et Wilcoxon) en segmentant les intervalles de recherche après chaque itération

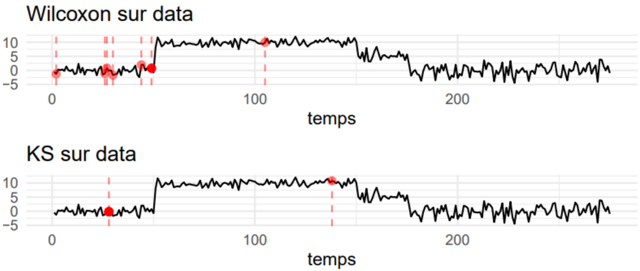

Figure: Test de comparaison sur les données

- **Complexité :**
  - KS : $O(K \times T^2 log\ \frac{T}{2})$
  - Wilcoxon : $O(K \times T^2)$

Objectif

- Introduire des méthodes statistiques plus appropriées et plus efficaces (complexité en $O(T)$ pour CUSUM et $O(T^2)$ pour Page-CUSUM)
- En supposant que la moyenne de pré-rupture μ_0 est connue : l'idée est de surveiller la valeur absolue des sommes partielles

$$S(s, T) = \sum_{t=s+1}^{T}(x_t - \mu_0)$$

CUSUM $\quad C(T) = \frac{1}{\sqrt{T}}|S(0, T)|$

Page-CUSUM $\quad P(T) = \max_{0 \leq w < T} \frac{1}{\sqrt{w}}|S(T-w, T)|$

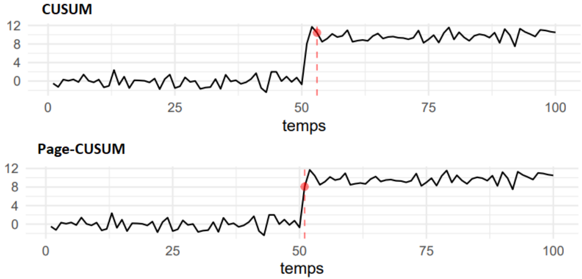

Figure: CUSUM et Page-Cusum sur les 100 premières valeurs

Objectif

- Réappliquer ces procédures lorsqu'on trouve un point de rupture afin de trouver plusieurs points de rupture
- Idée : remplacement des moyennes à priori connues avec les données de la série temporelle.

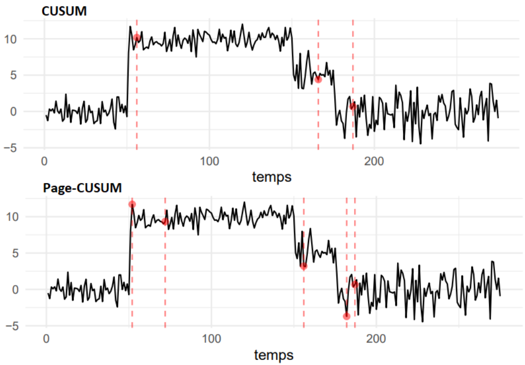

Figure: CUSUM et Page-Cusum hors-ligne sur les données

- **Complexité :** $O(T)$ pour CUSUM et $O(T^2)$ pour Page-CUSUM

Objectif

- On suppose que la moyenne à priori $\mu_0 = 0$, et qu'on connaît aussi la moyenne à posteriori μ_1
- permet de détecter un point de rupture en utilisant des statistiques calculées récursivement à partir des log-vraisemblances

$$LR(x_t, \mu_1) = 2\mu_1(x_t - \frac{\mu_1}{2})$$

Statistique séquentielle de Page :

$$Q_{T,\mu_1} = \max_{0 \leq s \leq T} \sum_{t=s+1}^{T} \frac{1}{2} LR(x_t, \mu_1)$$

$$\text{avec} \begin{cases} Q_{0,\mu_1} = 0 \\ Q_{t,\mu_1} = \max\{0, Q_{t-1,\mu_1} + \frac{1}{2} LR(x_T, \mu_1)\} \end{cases}$$

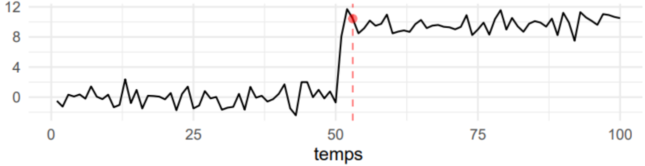

Figure: Page séquentielle sur les 100 premières valeurs

- **Complexité :** $O(T)$

FOCuS en ligne

Objectif

- utilisé sans forcément connaître la moyenne après le changement
- met à jour de manière récursive un modèle de quadratique par morceaux, et le maximum de ce modèle est utilisé comme statistique de test pour détecter un changement.

- **Complexité :** $O(logT)$ par itération

Bibliographie

1. Romano, Gaetano and Eckley, Idris and Fearnhead, Paul and Rigaill, Guillem (2021) Fast Online Changepoint Detection via Functional Pruning CUSUM statistics, arXiv, 10.48550/ARXIV.2110.08205
2. ES Page. A test for a change in a parameter occurring at an unknown point. Biometrika, 42(3/4):523–527, 1955.
3. Claudia Kirch, Silke Weber, et al. Modified sequential change point procedures based on estimating functions. Electronic Journal of Statistics, 12(1):1579–1613, 2018