

## 1 Introduction

Stéphane Robin nous présente une méthode qu'ils ont développé avec Emilie Lebarbier et Charlotte Dion-Blanc. Cette méthode est présentée dans l'article [DION-BLANC et al., 2023](#) sur HAL.

La méthode considère des données de comptages au cours du temps. Un exemple de telles données est celui du nombre de cris de chauve-souris au cours de la nuit. Ces données sont présentées dans sur la figure 1. Ou encore les données d'éruption du volcan Kilauea présentée sur la figure 2.

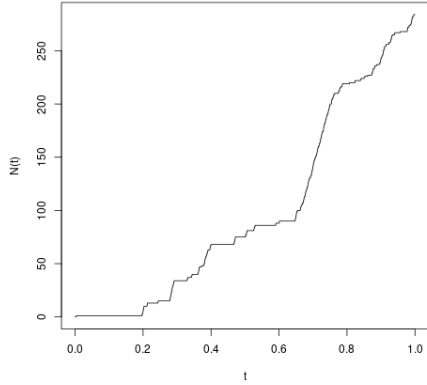


FIGURE 1 – Comptage de cris de chauve-souris (nuit du 17 juillet 2019)

L'intervalle de temps est normalisé,  $t \in [0, 1]$  et les instants d'évènements sont les  $0 < T_1 < \dots T_i < \dots T_n < 1$ . Étant donné qu'il s'agit d'un comptage aléatoire, le processus *naturel* est le processus de comptage,  $N(t) = \sum_{i=1}^n \mathbb{1}_{T_i \leq t}$  et parmi les processus de comptage, le processus de Poisson défini par sa fonction d'intensité  $\lambda(t)$ .

## 2 Méthode

La méthode fait l'hypothèse que la fonction d'intensité est constante par morceaux et qu'il existe des *points de ruptures* les  $(\tau_k)_{0 \leq k \leq K}$ . Et alors pour  $t \in I_k = ]\tau_{k-1}; \tau_k]$ ,  $\lambda(t) = \lambda_k$ . Ainsi l'objectif de la méthode est d'estimer les paramètres  $\theta = ((\tau_k)_{0 \leq k \leq K}, (\lambda_k)_{0 \leq k \leq K})$  et de réaliser une *sélection de modèle* pour obtenir le nombre de segments  $K$ .

### 2.1 Segmentation

Un rappel sur la segmentation en temps discret, nous montre que la programmation dynamique permet ainsi de résoudre le problème d'estimation des paramètres dans ce cas qui semblait apparemment computationnellement complexe.

Dans le cas de la méthode, le problème d'optimisation est

$$(\hat{\tau}, \hat{\lambda}) = \operatorname{argmin}_{\tau \in \mathcal{T}_K, \lambda \in (\mathbb{R}^+)^K} \gamma(\tau, \lambda)$$

L'additivité du contraste :  $\gamma(\tau, \lambda) = \sum_{k=1}^K C(\Delta N_k, \Delta \tau_k, \lambda_k)$  en tant que somme sur les segments aide à la résolution du problème d'optimisation. En effet le  $\lambda = (\lambda_1, \dots, \lambda_K)$  optimal peut être obtenu grâce à la propriété d'additivité en résolvant  $\hat{\lambda}_k = \lambda_k(\tau) = \operatorname{argmin}_{\lambda_k \in \mathbb{R}^+} C(\Delta N_k, \Delta \tau_k)$ . Et si la fonction de contraste est la log-vraisemblance négative, on a :  $\hat{\lambda}_k = \Delta N_k / \Delta \tau_k$ .

Mais le problème difficile est celui de trouver le  $\tau = (\tau_0, \dots, \tau_K)$  optimal, car le problème d'optimisation est alors  $\hat{\tau} = \operatorname{argmin}_{\tau \in \mathcal{T}_K} \hat{\gamma}(\tau)$ ,  $\hat{\gamma}(\tau) = \gamma(\tau, \hat{\lambda}(\tau))$  où  $\mathcal{T}_K$  est l'espace de segmentation **continu**,

$$\mathcal{T}_K = \{\tau \in [0, 1]^{K+1} : 0 = \tau_0 < \tau_1 < \dots < \tau_{K-1} < \tau_K = 1\}$$

Car le contraste n'est ni convexe ni continu par rapport à  $\tau$ .

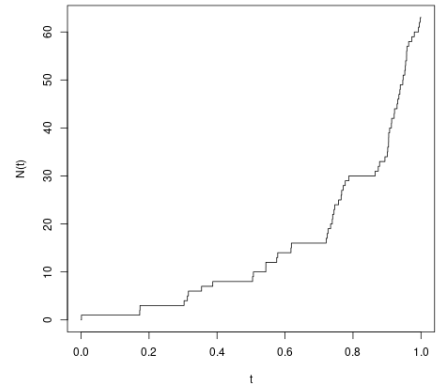


FIGURE 2 – Données d'éruption du Kilauea, 1750 - 1984

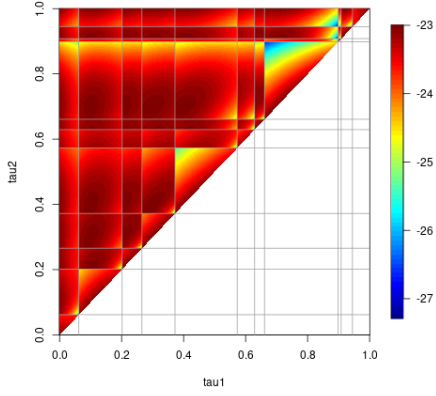


FIGURE 3 – Fonction de contraste pour  $n = 10$ ,  $K = 3$  et  $\tau = (\tau_1, \tau_2)$

La figure 3 présente les valeurs de la fonction de contraste pour les paramètres donnés. Sur la figure, chaque "bloc" correspond à une partition des nombre d'évènements ( $\Delta N = (\Delta N_1, \Delta N_2, \Delta N_3)$ ).

L'idée est alors de partitionner le nombre d'évènements  $\mathcal{N} = \{\nu \in \mathbb{N}^K : \sum_{k=1}^K \nu_k = n\}$  où  $\nu_k$  est le nombre d'évènements dans le segment  $k$ . Puis à partir de cette partition, on peut partitionner l'espace de segmentation  $\mathcal{T}(\nu) = \{\tau \in \mathcal{T}_K : \Delta N = \nu\}$  qui satisfait la partition  $\nu$ . Ce qui permet de réécrire le problème de minimisation :

$$\min_{\tau \in \mathcal{T}_K} \hat{\gamma}(\tau) = \min_{\nu \in \mathcal{N}^K} \min_{\tau \in \mathcal{T}(\nu)} \hat{\gamma}(\tau)$$

Et alors grâce au fait que

### 3 Apport personnel

### Bibliographie

### Références

DION-BLANC, C., LEBARBIER, E., & ROBIN, S. S. (2023). Détection de Ruptures Multiples Pour Les Processus de Poisson. *54ème Journées De Statistique De Société Française De Statistique*. Récupérée 31 janvier 2024, à partir de [https://hal.science/hal-](https://hal.science/hal-04403138)

04403138

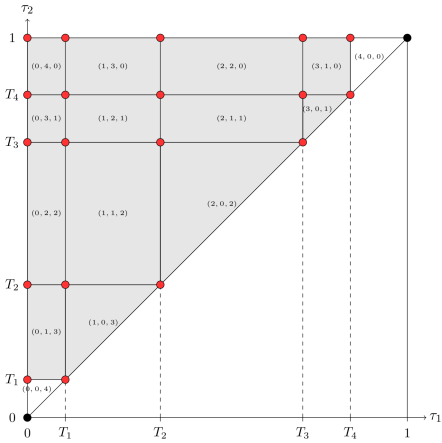


FIGURE 4 – Partitionnement de l'espace de segmentation pour  $K = 3$  et  $\tau = (\tau_1, \tau_2)$