Change-point detection in a Poisson process

E. Lebarbier[(a)]
joint ongoing work with C. Dion-Blanc[(b)] and S. Robin[(b)]

(a) Modal'X, Université Paris Nanterre
(b) Sorbonne Université

Journées MAS 2022, Rouen, 30 Août 2022

Example

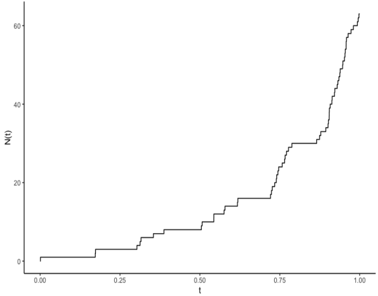
- Studying the activity of a volcano

Example

▶ Studying the activity of a volcano

▶ Counting the number of eruptions (events) that occurs in a time interval

Kilauea eruptions (from 1750 to 1984) *[1]*



Poisson process

Event times:

$$0 < T_1 < \ldots < T_i < \ldots < T_n < 1$$

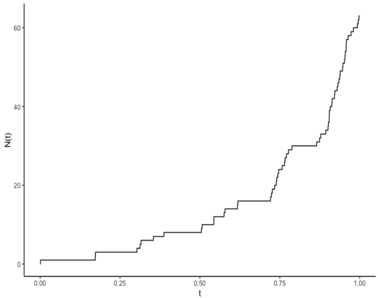
Counting process:

$$N(t) = \sum_{i=1}^{n} \mathbb{1}_{\{T_i \leq t\}}$$

Poisson process:

$$\{N(t)\}_{0 \leq t \leq 1} \sim PP(\lambda(t))$$

Kilauea eruptions



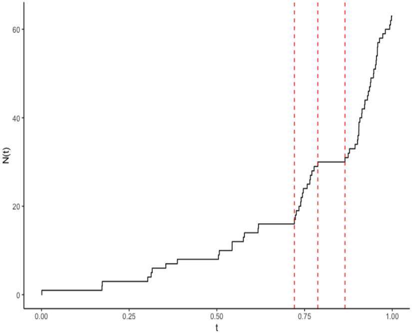
where the intensity function $\lambda(t)$ is

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(N(t+\Delta t) - N(t) = 1)}{\Delta t}, \qquad E[N(s)] - E[N(t)] = \int_t^s \lambda(u)\mathrm{d}u$$

Change-point detection

Question: find time-intervals in which events occurs more (or less) frequently

Kilauea eruptions



Change-point detection

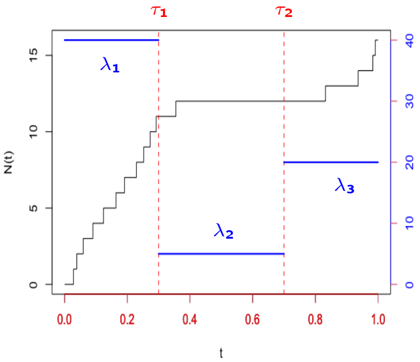
Piecewise constant intensity function.

⋆ Change-points:

$$(\tau_0 =)0 < \tau_1 < \ldots < \tau_{K-1} < 1(= \tau_K)$$

⋆ For $t \in I_k =]\tau_{k-1}, \tau_k]$:

$$\lambda(t) = \lambda_k$$



Change-point detection

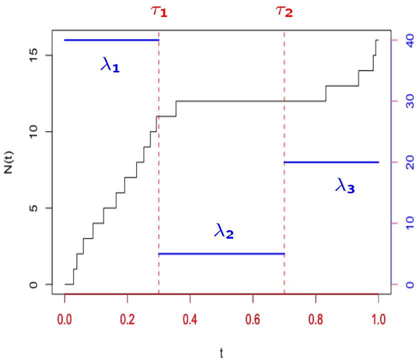
Piecewise constant intensity function.

⋆ Change-points:

$$(\tau_0 =)0 < \tau_1 < \ldots < \tau_{K-1} < 1(= \tau_K)$$

⋆ For $t \in I_k =]\tau_{k-1}, \tau_k]$:

$$\lambda(t) = \lambda_k$$



Aim and challenge

▶ Segmentation with K segments in a exact manner and reasonnably fast → extension of the idea of *[5, 2]* to multiple change-points

▶ Model selection: choose K

Maximum-likelihood segmentation

Property 1. Independence of disjoint time-intervals.

Log-likelihood. Denoting $\Delta N_k = N(\tau_k) - N(\tau_{k-1}) = N(I_k)$, $\Delta\tau_k = \tau_k - \tau_{k-1}$,

$$\log p_P(N;\tau,\lambda) = \sum_{k=1}^{K} (\Delta N_k \log \lambda_k - \Delta\tau_k \; \lambda_k)$$

Contrast: segment-additive (consequence of Property 1)

$$\gamma(N;\tau,\lambda) = -\log p_P(N;\tau,\lambda) = \sum_{k=1}^{K} C(\Delta N_k, \Delta\tau_k, \lambda_k)$$

Optimization problem.

$$(\widehat{\tau}, \widehat{\lambda}) = \underset{\tau\in\mathcal{M}^K}{\text{argmin}} \; \min_{\lambda} \; \gamma(N;\tau,\lambda)$$

where

$$\mathcal{M}^K := \{\tau = (\tau_1, \dots, \tau_{K-1}) \in (0,1)^{K-1}; 0 = \tau_0 < \tau_1 < \dots < \tau_K = 1\}$$

is the continuous set of all possible partitions of $(0,1)$ with K segments.

Shape of the contrast

Global optimization.

$$(\widehat{\tau}, \widehat{\lambda}) = \underset{\tau \in \mathcal{M}^K}{\operatorname{argmin}} \min_{\lambda} \sum_{k=1}^{K} C(\Delta N_k, \Delta \tau_k, \lambda_k) = \underset{\tau \in \mathcal{M}^K}{\operatorname{argmin}} \sum_{k=1}^{K} \min_{\lambda_k} C(\Delta N_k, \Delta \tau_k, \lambda_k)$$

Shape of the contrast

Global optimization.

$$(\widehat{\tau}, \widehat{\lambda}) = \underset{\tau \in \mathcal{M}^K}{\text{argmin}} \min_{\lambda} \sum_{k=1}^{K} C(\Delta N_k, \Delta \tau_k, \lambda_k) = \underset{\tau \in \mathcal{M}^K}{\text{argmin}} \sum_{k=1}^{K} \min_{\lambda_k} C(\Delta N_k, \Delta \tau_k, \lambda_k)$$

1. Optimization wrt λ.

$$\widehat{\lambda}_k = \frac{\Delta N_k}{\Delta \tau_k} \quad \text{and} \quad \widehat{\gamma}(N; \tau) = \gamma(N; \tau, \widehat{\lambda}) = \sum_{k=1}^{K} \underbrace{C(\Delta N_k, \Delta \tau_k, \widehat{\lambda}_k)}_{\widehat{C}(\Delta N_k, \Delta \tau_k)}$$

Shape of the contrast

Global optimization.

$$(\widehat{\tau}, \widehat{\lambda}) = \underset{\tau \in \mathcal{M}^K}{\operatorname{argmin}} \min_{\lambda} \sum_{k=1}^{K} C(\Delta N_k, \Delta\tau_k, \lambda_k) = \underset{\tau \in \mathcal{M}^K}{\operatorname{argmin}} \sum_{k=1}^{K} \min_{\lambda_k} C(\Delta N_k, \Delta\tau_k, \lambda_k)$$

1. Optimization wrt λ.

$$\widehat{\lambda}_k = \frac{\Delta N_k}{\Delta \tau_k} \quad \text{and} \quad \widehat{\gamma}(N; \tau) = \gamma(N; \tau, \widehat{\lambda}) = \sum_{k=1}^{K} \underbrace{C(\Delta N_k, \Delta\tau_k, \widehat{\lambda}_k)}_{\widehat{C}(\Delta N_k, \Delta\tau_k)}$$

2. Optimization wrt τ.

$$\widehat{\tau} = \underset{\tau \in \mathcal{M}^K}{\operatorname{argmin}} \widehat{\gamma}(N; \tau) = \underset{\tau \in \mathcal{M}^K}{\operatorname{argmin}} \sum_{k=1}^{K} \widehat{C}(\Delta N_k, \Delta\tau_k)$$

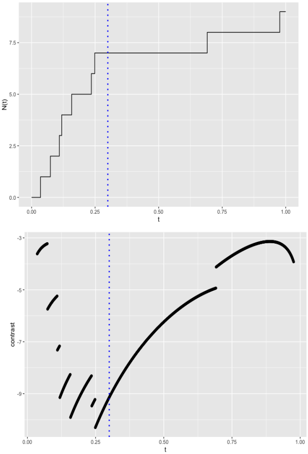
$\rightarrow$ $\widehat{\gamma}(N; \tau)$ is not convex nor even continuous wrt τ.

Illustration: one change-point *[5, 2]*

- $\widehat{\gamma}(N;\tau)$ is concave in τ in each inter-event interval $[T_i, T_{i+1}[$.

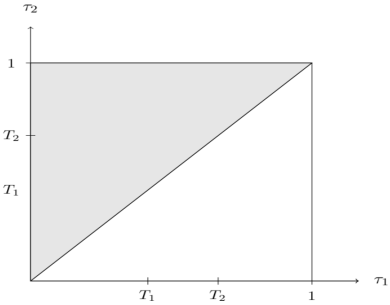
- The optimal change-point $\widehat{\tau}$ in $[T_i, T_{i+1}[$ is one of T_i, T_{i+1}^-

- $\widehat{\tau} \subset \{T_1^-, T_1, T_2^-, T_2, \ldots, T_n^-, T_n\}$

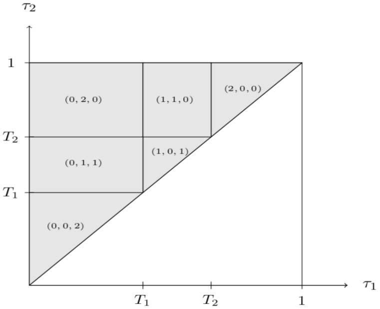


Extension to multiple change-points?

Example: $n = 2$, $K = 3$ (2 change-points)



$\mathcal{M}^K$



Each block corresponds to a specific vector

$$\Delta N = (\Delta N_{\mathbf{1}}, \Delta N_{\mathbf{2}}, \Delta N_{\mathbf{3}})$$

Partitioning the segmentation space

Partitioning the number of events. Define

$$\mathcal{N}^K := \{\nu \in \mathbb{N}^K : \sum_{k=1}^{K} \nu_k = n\},$$

the set of all possible repartitions of the n events into the K segments.

Partitioning the segmentation space. For $\nu \in \mathcal{N}^K$, define $\mathcal{M}_\nu^K(N) := \{\tau \in \mathcal{M}^K : \Delta N_k = \nu_k\}$ and

$$\mathcal{M}^K = \bigcup_{\nu \in \mathcal{N}^K} \mathcal{M}_\nu^K(N)$$

Partitioning the segmentation space

Partitioning the number of events. Define

$$\mathcal{N}^K := \{\nu \in \mathbb{N}^K : \sum_{k=1}^{K} \nu_k = n\},$$

the set of all possible repartitions of the n events into the K segments.

Partitioning the segmentation space. For $\nu \in \mathcal{N}^K$, define $\mathcal{M}^K_\nu(N) := \{\tau \in \mathcal{M}^K : \Delta N_k = \nu_k\}$ and

$$\mathcal{M}^K = \bigcup_{\nu \in \mathcal{N}^K} \mathcal{M}^K_\nu(N)$$

Optimization problem: $\min_{\tau \in \mathcal{M}^K} \widehat{\gamma}(N; \tau) = \min_{\nu \in \mathcal{N}^K} \min_{\tau \in \mathcal{M}^K_\nu(N)} \widehat{\gamma}(N; \tau)$

Partitioning the segmentation space

Partitioning the number of events. Define

$$\mathcal{N}^K := \{\nu \in \mathbb{N}^K : \sum_{k=1}^{K} \nu_k = n\},$$

the set of all possible repartitions of the n events into the K segments.

Partitioning the segmentation space. For $\nu \in \mathcal{N}^K$, define $\mathcal{M}^K_\nu(N) := \{\tau \in \mathcal{M}^K : \Delta N_k = \nu_k\}$ and

$$\mathcal{M}^K = \bigcup_{\nu \in \mathcal{N}^K} \mathcal{M}^K_\nu(N)$$

Optimization problem: $\min_{\tau \in \mathcal{M}^K} \widehat{\gamma}(N;\tau) = \min_{\nu \in \mathcal{N}^K} \min_{\tau \in \mathcal{M}^K_\nu(N)} \widehat{\gamma}(N;\tau)$

Theorem

$\widehat{\gamma}(N;\tau)$ is concave wrt $\tau \in \mathcal{M}^K_\nu(N)$, thus

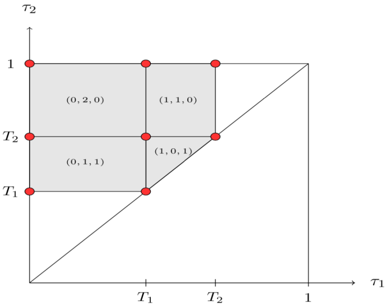
$$\widehat{\tau} = \underset{\tau \in \mathcal{M}^K_\nu(N)}{\text{argmin}} \; \widehat{\gamma}(N;\tau) \in \{T_{\nu_1}, T^-_{\nu_1+1}\} \times \{T_{\nu_1+\nu_2}, T^-_{\nu_1+\nu_2+1}\} \times \ldots \times \{T_{\nu_1+\ldots\nu_K}, T^-_{\nu_1+\ldots\nu_K}\}$$

$\rightarrow$ always true for any contrast s.t. $\widehat{C}(\Delta N_k, \Delta\tau_k)$ is concave wrt $\Delta\tau_k$

Consequence: a discrete optimization problem

⋆ For each block, $\widehat{\tau}$ is necessarly localized on the boudaries

⋆ $\widehat{\tau} \subset \{T_1^-, T_1, T_2^-, T_2, \ldots, T_n^-, T_n\}$



$\widehat{\tau}$ can be obtained by the the usual Dynamic Programming algorithm on the $2n$ possible change-points

$$\mathrm{tp} = \{T_1^-, T_1, , T_2^-, T_2, \ldots, T_n^-, T_n\},$$

with complexity $\mathcal{O}((2n)^2K)$.

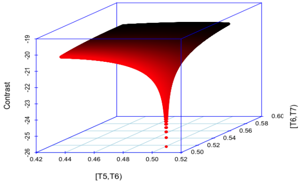
Alternative contrast

Undesirable Poisson likelihood solution:

- tp includes segments with length 0 (e.g. $I_k =]T_i^-, T_i]$, $\Delta N_k = 1$),
- ... which are optimal for the log-likelihood contrast: $\widehat{C}(\Delta N_k = 1, \Delta\tau_k = 0) = -\infty$

For $K = 3$ and $\nu = (5, 1, n-6)$, the optimal change-points in $\mathcal{M}_\nu^K(N)$ are

$$(\widehat{\tau_1}, \widehat{\tau_2}) = (T_6^-, T_6)$$



Alternative contrast

Poisson-Gamma model. For each segment $1 \leq k \leq K$:

$$\lambda_k \sim \mathcal{G}am(a, b), \qquad \{N(t)\}_{t \in I_k} | \lambda_k \sim PP(\lambda_k)$$

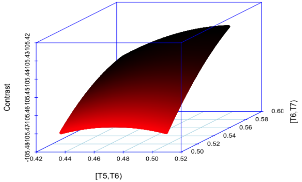
which gives

$$\begin{aligned} \gamma_{(a,b)}(N; \tau) &= -\log p_{PG}(N; \tau, a, b) = -\log \int p(N, \lambda; \tau, a, b) \, \mathrm{d}\lambda \\ &= \text{Cst} + \sum_{k=1}^{K} \left((a + \Delta N_k) \log (b + \Delta \tau_k) - \log \Gamma(a + \Delta N_k) \right) \end{aligned}$$

$\rightarrow$ enjoys the concavity property, but avoids segments with null length.

The optimal change-points in $\mathcal{M}_{\nu}^{K}(N)$ are

$$(\widehat{\tau_1}, \widehat{\tau_2}) = (T_5, T_6)$$



$a = n^{3/2}$, $b = n^{1/2}$

Model selection via cross validation

Property 2. The sum of two independent Poisson processes is a Poisson process.

Model selection via cross validation

Property 2. The sum of two independent Poisson processes is a Poisson process.

Consequence

- $\{N(t)\}_t \sim PP(\lambda(t))$

Model selection via cross validation

Property 2. The sum of two independent Poisson processes is a Poisson process.

Consequence

- $\{N(t)\}_t \sim PP(\lambda(t))$

- Sampling the event times with probability f

$$\{N^L(t)\}_t \sim PP(\underbrace{\lambda^L(t)}_{f\lambda(t)}), \qquad \{N^T(t)\}_t \sim PP(\underbrace{\lambda^T(t)}_{(1-f)\lambda(t)}), \qquad \{N^L(t)\}_t \perp \{N^T(t)\}_t$$

Model selection via cross validation

Property 2. The sum of two independent Poisson processes is a Poisson process.

Consequence

- $\{N(t)\}_t \sim PP(\lambda(t))$

- Sampling the event times with probability f

$$\{N^L(t)\}_t \sim PP(\underbrace{\lambda^L(t)}_{f\lambda(t)}), \qquad \{N^T(t)\}_t \sim PP(\underbrace{\lambda^T(t)}_{(1-f)\lambda(t)}), \qquad \{N^L(t)\}_t \perp \{N^T(t)\}_t$$

- If $\lambda(t)$ is piecewise constant with change-points $\tau = (\tau_k)$, then $\lambda^L(t)$ and $\lambda^T(t)$ are piecewise constant with same change-points (τ_k)

Cross validation

Cross validation. For $1 \leq K \leq K_{\mathbf{max}}$,

- Repeat for $1 \leq b \leq B$:

  1. Sample the event times to form $\{N^{L,b}(t)\}_t$ (learn) and $\{N^{T,b}(t)\}_t$ (test)

Poisson contrast.	Poisson-Gamma contrast.
2. Estimate $\tau^{L,b}$ and $\lambda^{L,b}$ using $\gamma(N^{L,b}; \tau, \lambda)$	2. Estimate $\tau^{L,b}$ using $\gamma_{(a^L, b^L)}(N^{L,b}; \tau)$
3. Compute $\gamma_K^{T,b} = \gamma(N^{T,b}; \widehat{\tau}^{L,b}, \frac{1-f}{f}\widehat{\lambda}^{L,b})$	3. Compute $\gamma_K^{T,b} = \gamma_{(a^T, b^T)}(N^{T,b}; \widehat{\tau}^{L,b})$

- Compute:

$$\bar{\gamma}_K = \frac{1}{B}\sum_{b=1}^{B} \gamma_K^{T,b}$$

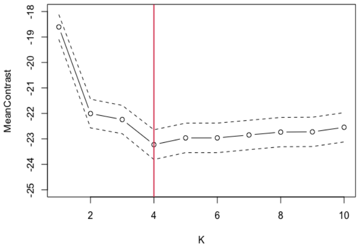
Choose K as

$$\widehat{K} = \underset{K}{\operatorname{argmin}}\ \bar{\gamma}_K$$

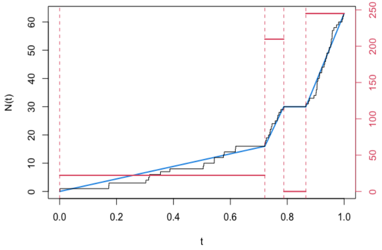
Kilauea eruptions

Poisson-Gamma contrast. Cross validation with $a^L = 1, b^L = 1/n^L$ and $a^T = 1, b^T = 1/n^T$.

Model selection via CV



Resulting segmentation



Ongoing and future works

Ongoing works

- Simulation study
- Marked Poisson Process $MPP(\lambda(t), \mu(t))$.
  - $\{N(t)\}_{\mathbf{0} \leq t \leq \mathbf{1}} \sim PP(\lambda(t))$, and at each T_i: $Y_i \sim \mathcal{F}(\mu(T_i))$
  - Works the same way, provided that concavity holds.

Future works on PP and MPP

- Model selection. Developp a theoretically grounded model selection criterion
- Segmentation/clustering. Adapted for recurrent events on one process
  - Each segment belong to a class $1 \leq q \leq Q$ with probability π_q and intensity $\lambda_k = \lambda_q$
  - Combination of EM and DP algorithms as proposed by *[4]*

Extension to Hawkes process.

Chih-Hsiang Ho and Moinak Bhaduri, *A quantitative insight into the dependence dynamics of the kilauea and mauna loa volcanoes, hawaii*, Mathematical Geosciences 49 (2017), no. 7, 893–911.

Qing Li, *Recurrent-event models for change-points detection*, Ph.D. thesis, Virginia Tech, 2015.

Qing Li, Kehui Yao, and Xinyu Zhang, *A change-point detection and clustering method in the recurrent-event context*, Journal of Statistical Computation and Simulation 90 (2020), no. 6, 1131–1149.

Franck Picard, Stéphane Robin, E Lebarbier, and J-J Daudin, *A segmentation/clustering model for the analysis of array cgh data*, Biometrics 63 (2007), no. 3, 758–766.

Tae Young Yang and Lynn Kuo, *Bayesian binary segmentation procedure for a poisson process with multiple changepoints*, Journal of Computational and Graphical Statistics 10 (2001), no. 4, 772–785.