Change-point detection in a Poisson process

S. Robin

joint ongoing work with E. Lebarbier, C. Dion-Blanc

Sorbonne université

Stats au sommet, Rochebrune, Mar. 2022

Example

Bat cries (night of the 17 jul. 2019)

Example

Point process on $t \in [0, 1]$.

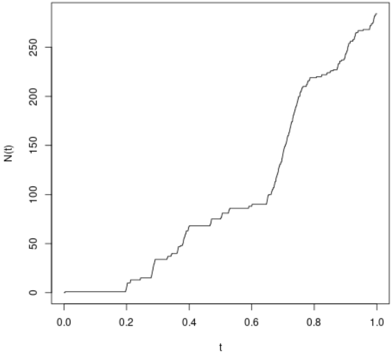
Event times:

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

Counting process:

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

Bat cries (night of the 17 jul. 2019)[a]



[a] source: Vigie-Chiro program, Y. Bas, CESCO-MNHN

Example

Point process on $t \in [0, 1]$.

Event times:

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

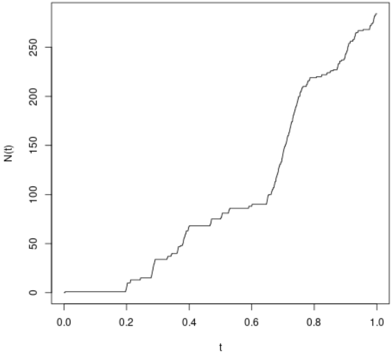
Counting process:

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

Poisson Process.

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

Bat cries (night of the 17 jul. 2019)[a]



[a] source: Vigie-Chiro program, Y. Bas, CESCO-MNHN

Intensity function $\lambda(t)$:

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

Change-point detection

Piecewise constant intensity function.

Change-points

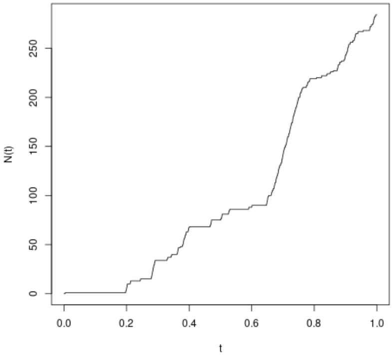
$$(\tau_0 =)0 < \tau_1 \cdots < \tau_{K-1} < 1(= \tau_K)$$

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

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

$\rightarrow$ Continuous piece-wise linear cumulated intensity function

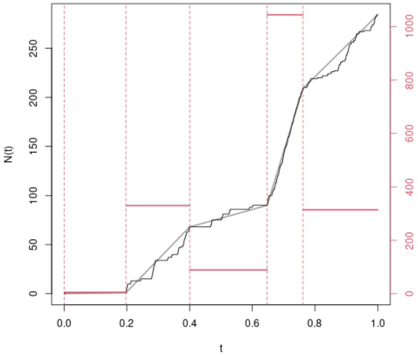
Bat cries (night of the 17 jul. 2019)[a]



[a] source: Vigie-Chiro program, Y. Bas, CESCO-MNHN

Change-point detection

Bat cries (night of the 17 jul. 2019)[a]



[a]source: Vigie-Chiro program, Y. Bas, CESCO-MNHN

Piecewise constant intensity function.

Change-points

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

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

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

$\rightarrow$ Continuous piece-wise linear cumulated intensity function

Aim.

- Segmentation: estimate (τ, λ) reasonnably fast
- Model selection: choose K

Change-point detection

Kilauea eruptions

Piecewise constant intensity function.

Change-points

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

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

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

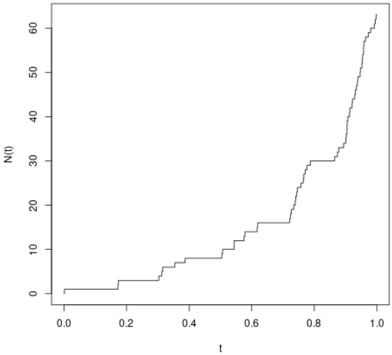
$\rightarrow$ Continuous piece-wise linear cumulated intensity function

Aim.

- Segmentation: estimate (τ, λ) reasonnably fast
- Model selection: choose K

Change-point detection

Kilauea eruptions (from 1750 to 1984)[a]



[a]source: [HB17]

Piecewise constant intensity function.

Change-points

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

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

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

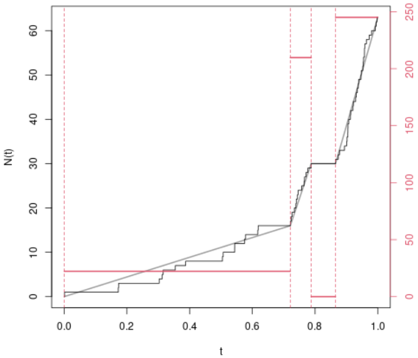
→ Continuous piece-wise linear cumulated intensity function

Aim.

- Segmentation: estimate (τ, λ) reasonnably fast
- Model selection: choose K

Change-point detection

Kilauea eruptions (from 1750 to 1984)[a]



[a] source: [HB17]

Piecewise constant intensity function.

Change-points

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

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

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

$\rightarrow$ Continuous piece-wise linear cumulated intensity function

Aim.

- Segmentation: estimate (τ, λ) reasonnably fast
- Model selection: choose K

Outline

Maximum-likelihood segmentation

Property 1: Independence of disjoint intervals.

Maximum-likelihood segmentation

Property 1: Independence of disjoint intervals.

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

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

Maximum-likelihood segmentation

Property 1: Independence of disjoint intervals.

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

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

Additive contrast. Sum over the segments

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

Maximum-likelihood segmentation

Property 1: Independence of disjoint intervals.

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

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

Additive contrast. Sum over the segments

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

Optimization problem.

$$(\widehat{\tau}, \widehat{\lambda}) = \arg\min_{\tau \in \mathcal{T}^K} \min_{\lambda} \gamma(\tau, \lambda)$$

where $\mathcal{T}^K$ is a *continuous set* and $\gamma(\tau, \lambda)$ is *not convex* nor even *continuous*.

Shape of the contrast fonction

Optimization wrt λ:

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

e.g.: $\widehat{\lambda}_k = \Delta N_k / \Delta\tau_k$

Shape of the contrast fonction

Optimization wrt λ:

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

e.g.: $\widehat{\lambda}_k = \Delta N_k / \Delta\tau_k$

Optimization wrt τ:

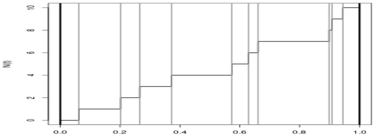
$$\widehat{\tau} = \arg\min_{\tau} \widehat{\gamma}(\tau) = \sum_k \widehat{C}(\Delta N_k, \Delta\tau_k)$$

Shape of the contrast fonction

Optimization wrt λ:

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

e.g.: $\widehat{\lambda}_k = \Delta N_k / \Delta\tau_k$

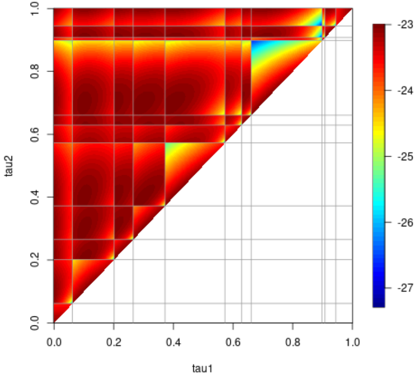


Optimization wrt τ:

$$\widehat{\tau} = \arg\min_{\tau} \widehat{\gamma}(\tau) = \sum_{k} \widehat{C}(\Delta N_k, \Delta\tau_k)$$

Example: $n = 10$, $K = 3$.
Each block corresponds to a specific vector

$$\Delta N = (\Delta N_1, \Delta N_2, \Delta N_3)$$



Shape of the contrast fonction

Optimization wrt λ:

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

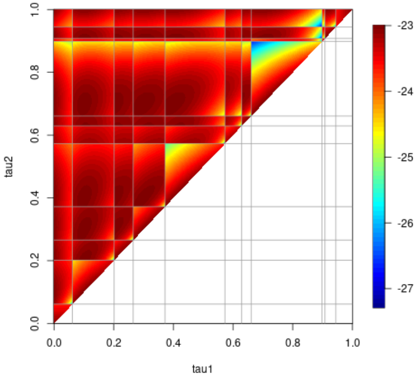
e.g.: $\widehat{\lambda}_k = \Delta N_k / \Delta\tau_k$

Optimization wrt τ:

$$\widehat{\tau} = \arg\min_{\tau} \widehat{\gamma}(\tau) = \sum_{k} \widehat{C}(\Delta N_k, \Delta\tau_k)$$

Example: $n = 10$, $K = 3$.
Each block corresponds to a specific vector

$$\Delta N = (\Delta N_1, \Delta N_2, \Delta N_3)$$



Partitioning the segmentation space

Partitioning the number of events. Define $\mathcal{N}^K = \left\{\nu \in \mathbb{N}^K : \sum_{k=1}^K \nu_k = n\right\}$.

Partitioning the segmentation space

Partitioning the number of events. Define $\mathcal{N}^K = \left\{\nu \in \mathbb{N}^K : \sum_{k=1}^K \nu_k = n\right\}$.

Partitioning the segmentation space. For $\nu \in \mathcal{N}_K$, define $\mathcal{T}_\nu^K = \left\{\tau \in \mathcal{T}^K : \Delta N = \nu\right\}$ so that

$$\min_{\tau \in \mathcal{T}^K} \widehat{\gamma}(\tau) = \min_{\nu \in \mathcal{N}^K} \min_{\tau \in \mathcal{T}_\nu^K} \widehat{\gamma}(\tau,).$$

Partitioning the segmentation space

Partitioning the number of events. Define $\mathcal{N}^K = \left\{\nu \in \mathbb{N}^K : \sum_{k=1}^K \nu_k = n\right\}$.

Partitioning the segmentation space. For $\nu \in \mathcal{N}_K$, define $\mathcal{T}_\nu^K = \left\{\tau \in \mathcal{T}^K : \Delta N = \nu\right\}$ so that

$$\min_{\tau \in \mathcal{T}^K} \widehat{\gamma}(\tau) = \min_{\nu \in \mathcal{N}^K} \min_{\tau \in \mathcal{T}_\nu^K} \widehat{\gamma}(\tau,).$$

Property. If $K \leqslant n$ and, for each $\nu \in \mathcal{N}^K$, $\widehat{\gamma}(\tau)$ is strictly concave wrt $\tau \in \mathcal{T}_\nu^K$, then

$$\widehat{\tau} = \arg\min_{\tau \in \mathcal{T}^K} \widehat{\gamma}(\tau) \subset \{T_1^-, T_1, T_2^-, T_2^-, \dots T_n^-, T_n\}.$$

Partitioning the segmentation space

Partitioning the number of events. Define $\mathcal{N}^K = \left\{\nu \in \mathbb{N}^K : \sum_{k=1}^K \nu_k = n\right\}$.

Partitioning the segmentation space. For $\nu \in \mathcal{N}_K$, define $\mathcal{T}_\nu^K = \left\{\tau \in \mathcal{T}^K : \Delta N = \nu\right\}$ so that

$$\min_{\tau \in \mathcal{T}^K} \widehat{\gamma}(\tau) = \min_{\nu \in \mathcal{N}^K} \min_{\tau \in \mathcal{T}_\nu^K} \widehat{\gamma}(\tau,).$$

Property. If $K \leqslant n$ and, for each $\nu \in \mathcal{N}^K$, $\widehat{\gamma}(\tau)$ is strictly concave wrt $\tau \in \mathcal{T}_\nu^K$, then

$$\widehat{\tau} = \arg\min_{\tau \in \mathcal{T}^K} \widehat{\gamma}(\tau) \subset \{T_1^-, T_1, T_2^-, T_2^-, \dots T_n^-, T_n\}.$$

Consequence. $\widehat{\tau}$ can be obtained by dynamic programming over the $2n + 2$ possible change-points

$$\mathcal{S} = \{0, T_1^-, T_1, T_2^-, T_2, \dots T_n^-, T_n, 1\}.$$

Alternative constrast

Remark.

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

Alternative constrast

Remark.

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

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

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

which gives:

$$\begin{aligned} C(\Delta N_k, \Delta \tau_k) &= -\log p_{a,b}(\{N(t)\}_{t \in I_k}) \\ &= \text{cst} - \log \Gamma(a + \Delta N_k) + (a + \Delta N_k) \log(b + \Delta \tau_k) \end{aligned}$$

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

Outline

Model selection

Property 2: Thining.

- $\{N(t)\} \sim PP(\lambda(t))$
- Sample the event times (prob. v)

Model selection

Property 2: Thining.

- $\{N(t)\} \sim PP(\lambda(t))$
- Sample the event times (prob. v)

$$\{N^L(t)\} \sim PP(v\lambda(t)), \qquad \{N^T(t)\} \sim PP((1-v)\lambda(t)), \qquad \{N^L(t)\} \perp \{N^T(t)\}$$

Model selection

Property 2: Thining.

- $\{N(t)\} \sim PP(\lambda(t))$
- Sample the event times (prob. v)

$$\{N^L(t)\} \sim PP(v\lambda(t)), \qquad \{N^T(t)\} \sim PP((1-v)\lambda(t)), \qquad \{N^L(t)\} \perp \{N^T(t)\}$$

Consequence. If $\lambda(t)$ is piece-wise constant with parms $\tau = (\tau_k)$ and $\lambda = (\lambda_k)$, then

Model selection

Property 2: Thining.

- $\{N(t)\} \sim PP(\lambda(t))$
- Sample the event times (prob. v)

$$\{N^L(t)\} \sim PP(v\lambda(t)), \qquad \{N^T(t)\} \sim PP((1-v)\lambda(t)), \qquad \{N^L(t)\} \perp \{N^T(t)\}$$

Consequence. If $\lambda(t)$ is piece-wise constant with parms $\tau = (\tau_k)$ and $\lambda = (\lambda_k)$, then

- $\lambda^L(t)$ piece-wise constant with change points (τ_k) and intensities $(v\lambda_k)$,
- $\lambda^T(t)$ piece-wise constant with change points (τ_k) and intensities $((1-v)\lambda_k)$,
- $\{N^L(t)\} \perp \{N^T(t)\}$.

Cross validation

Sampling event times provides two independent Poisson processes with same change points.

Cross validation

Sampling event times provides two independent Poisson processes with same change points.

Cross-validation. For $1 \leqslant K \leqslant K_{\max}$,

- Repeat for $1 \leqslant m \leqslant M$:

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

  2 – Estimate $\widehat{\tau}^{L,m}$ and $\widehat{\lambda}^{L,m}$ from $\{N^{L,m}(t)\}$,

  3 – Compute the contrast $\gamma_K^{T,m} = \gamma\left(\{N^T(t)\}; \widehat{\tau}^{L,m}, \dfrac{1-v}{v}\widehat{\lambda}^{L,m}\right)$.

Cross validation

Sampling event times provides two independent Poisson processes with same change points.

Cross-validation. For $1 \leqslant K \leqslant K_{\max}$,

- Repeat for $1 \leqslant m \leqslant M$:

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

  2 – Estimate $\widehat{\tau}^{L,m}$ and $\widehat{\lambda}^{L,m}$ from $\{N^{L,m}(t)\}$,

  3 – Compute the contrast $\gamma_K^{T,m} = \gamma\left(\{N^T(t)\}; \widehat{\tau}^{L,m}, \frac{1-v}{v}\widehat{\lambda}^{L,m}\right)$.

- Compute

$$\overline{\gamma}_K = \frac{1}{M}\sum_{m=1}^{M} \gamma_K^{T,m}$$

Cross validation

Sampling event times provides two independent Poisson processes with same change points.

Cross-validation. For $1 \leqslant K \leqslant K_{\max}$,

- Repeat for $1 \leqslant m \leqslant M$:

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

  2 – Estimate $\widehat{\tau}^{L,m}$ and $\widehat{\lambda}^{L,m}$ from $\{N^{L,m}(t)\}$,

  3 – Compute the contrast $\gamma_K^{T,m} = \gamma\left(\{N^T(t)\}; \widehat{\tau}^{L,m}, \frac{1-v}{v}\widehat{\lambda}^{L,m}\right)$.

- Compute

$$\overline{\gamma}_K = \frac{1}{M}\sum_{m=1}^{M} \gamma_K^{T,m}$$

- Select

$$\widehat{K} = \arg\min_K \overline{\gamma}_K$$

Illustration

Poisson-Gamma contrast. Set $a^L = a^T = 1$, $b^L = 1/n_L$, $b^T = 1/n_T$, and compute

$$\begin{aligned} - \log p(\{N^T(t)\} \mid \widehat{\tau}^L) = & \sum_{k=1}^{K} (a + \Delta N_k^T) \log(a + \Delta\tau_k) - \log\Gamma(a + \Delta N_k^L) \\ & - K \left(a^T \log b^T - \log\Gamma(a^T)\right) \end{aligned}$$

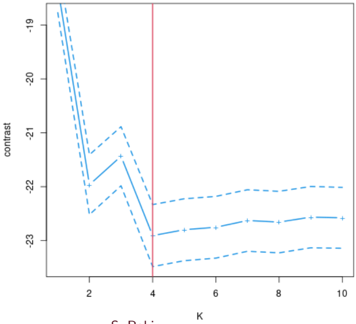
Illustration

Poisson-Gamma contrast. Set $a^L = a^T = 1$, $b^L = 1/n_L$, $b^T = 1/n_T$, and compute

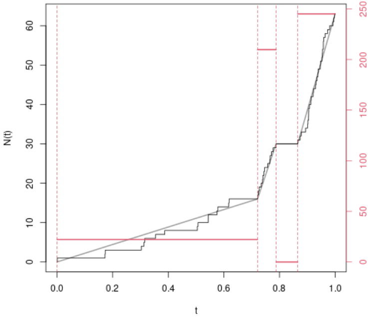
$$-\log p(\{N^T(t)\} \mid \hat{\tau}^L) = \sum_{k=1}^{K} (a + \Delta N_k^T) \log(a + \Delta \tau_k) - \log \Gamma(a + \Delta N_k^L)$$
$$- K \left(a^T \log b^T - \log \Gamma(a^T)\right)$$

Kilauea eruptions

Model selection via CV



Resulting segmentation



Outline

Estimation

Model selection

Extensions

Extensions

Marked Poisson Process.

- $\{Y(t)\}_{0 \leqslant t \leqslant 1} \sim MPP(\lambda(t), \mu(t))$:

$$\{N(t)\}_{0 \leqslant t \leqslant 1} \sim PP(\lambda(t)), \qquad \text{at each } T_i: \quad X_i \sim \mathcal{F}(\mu(T_i))$$

- Works the same way, provided that concavity holds.
- Bat cries: X = bat species or cry duration.

Extensions

Marked Poisson Process.

- $\{Y(t)\}_{0 \leqslant t \leqslant 1} \sim MPP(\lambda(t), \mu(t))$:

$$\{N(t)\}_{0 \leqslant t \leqslant 1} \sim PP(\lambda(t)), \qquad \text{at each } T_i: \quad X_i \sim \mathcal{F}(\mu(T_i))$$

- Works the same way, provided that concavity holds.
- Bat cries: $X =$ bat species or cry duration.

Segmentation-clustering.

- Each segment belongs to a class $1 \leqslant q \leqslant Q$ (with probability π_q and intensity $\lambda_k = \ell_q$),
- Combination of EM and DP algorithms [PRLD07],
- Class = animal behaviour (hunt, transit, ...)

Extensions

Marked Poisson Process.

- $\{Y(t)\}_{0 \leqslant t \leqslant 1} \sim MPP(\lambda(t), \mu(t))$:

$$\{N(t)\}_{0 \leqslant t \leqslant 1} \sim PP(\lambda(t)), \qquad \text{at each } T_i: \quad X_i \sim \mathcal{F}(\mu(T_i))$$

- Works the same way, provided that concavity holds.
- Bat cries: $X =$ bat species or cry duration.

Segmentation-clustering.

- Each segment belongs to a class $1 \leqslant q \leqslant Q$ (with probability π_q and intensity $\lambda_k = \ell_q$),
- Combination of EM and DP algorithms [PRLD07],
- Class = animal behaviour (hunt, transit, ...)

And also.

- Theoretically grounded model selection criterion,
- Other desirable contrasts, ...

References I

C-H Ho and M Bhaduri. A quantitative insight into the dependence dynamics of the Kilauea and Mauna Loa volcanoes, Hawaii. *Mathematical Geosciences*, 49(7):893–911, 2017.

F. Picard, S. Robin, E Lebarbier, and J-J Daudin. A segmentation/clustering model for the analysis of array CGH data. *Biometrics*, 63(3):758–766, 2007.

Appendix

$$|\mathcal{N}_K| = \sum_{h=\lfloor (K-1)/2 \rfloor}^{K} \binom{n-1}{h-1} \binom{h+1}{K-h}$$