Segmentation *and classification* of a *Hawkes* process

S. Robin

joint work with C. Dion-Blanc, E. Lebarbier
and A. Bonnet

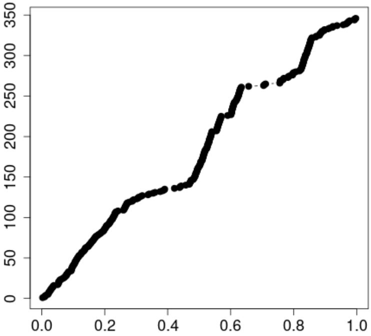
Sorbonne université

Rencontres Mathématiques de Rouen, Jun. 2024

Problem

Counting process

Overnight recording of bat cries in continuous time

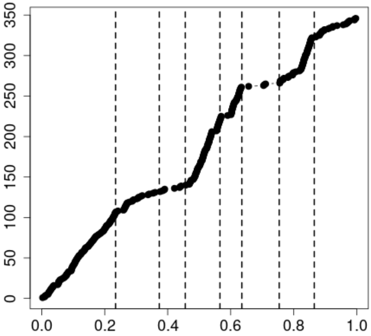


Problem

Counting process

Overnight recording of bat cries in continuous time

- Can we detect changes in the occurrence of events?

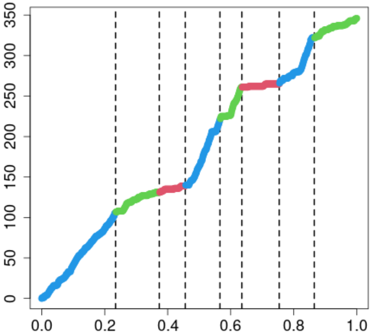


Problem

Counting process

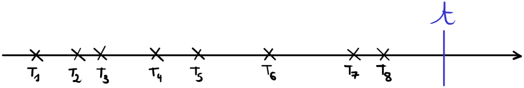
Overnight recording of bat cries in continuous time

- Can we detect changes in the occurrence of events?
- Can we associate each time period with some underlying behavior?



Point process

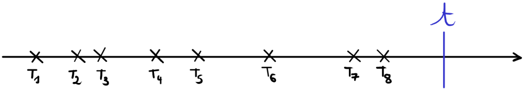
Reminder.



- $(T_k)_{k \geqslant 1}$ a random collection of points
- Count process $N(t) = \sum_{k \geqslant 1} \mathbb{I}\{T_k \leqslant t\}$
- Intensity function $\lambda(t)$: immediate probability of observing an event at time t

Point process

Reminder.



- $(T_k)_{k \geqslant 1}$ a random collection of points
- Count process $N(t) = \sum_{k \geqslant 1} \mathbb{I}\{T_k \leqslant t\}$
- Intensity function $\lambda(t)$: immediate probability of observing an event at time t

Examples

- Homogeneous Poisson process: $\lambda(t) \equiv \lambda$
- Heterogeneous Poisson process: $\lambda(t) =$ deterministic function
- Hawkes process: $\lambda(t) =$ random function of the past

Segmentation (& classification) of a point process

Aim

1. Propose a set of reasonably realistic models;
2. Design an (efficient) algorithm to get the parameter estimates;
3. Choose among the models.

Segmentation (& classification) of a point process

Aim

1. Propose a set of reasonably realistic models;
2. Design an (efficient) algorithm to get the parameter estimates;
3. Choose among the models.

Example Segmentation of a Poisson process [DBLR23]:

1. Model = Poisson process with piece-wise constant intensity function;
2. Algorithm = dynamic programming in (less than) $\mathcal{O}(N(T)^2)$;
3. Model selection = cross validation (using thining)

Outline

Segmentation of a Poisson process (1/3)

Model.

Change-points

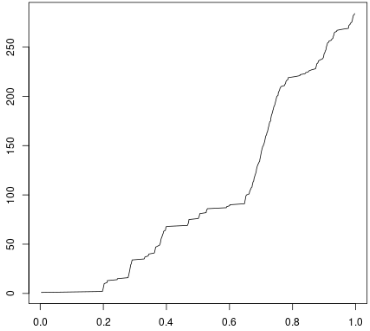
$$(\tau_0 =)0 < \tau_1 \cdots < \tau_{K-1} < 1(= \tau_K)$$

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

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

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

Bat cries[a]



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

Segmentation of a Poisson process (1/3)

Model.

Change-points

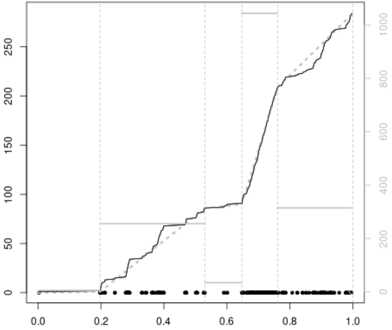
$$(\tau_0 =)0 < \tau_1 \cdots < \tau_{K-1} < 1(= \tau_K)$$

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

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

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

Bat cries[a]



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

- Segmentation algorithm: find the 'optimal' (τ, λ) in a reasonnable time
- Model selection: choose K

Segmentation of a Poisson process (2/3)

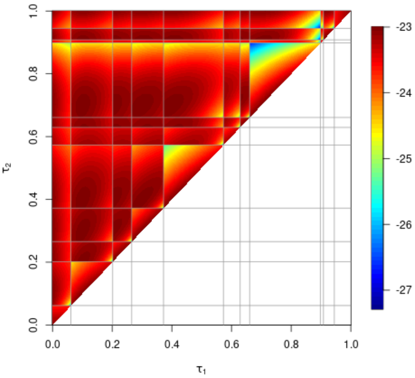
Classical contrasts (negative log-likelihood, least-square) are

- additive wrt the segments and
- concave wrt the length of each segment,

Segmentation of a Poisson process (2/3)

Classical contrasts (negative log-likelihood, least-square) are

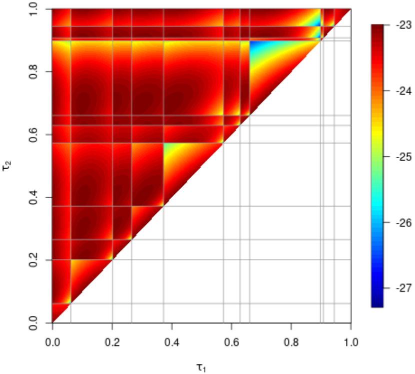
- additive wrt the segments and
- concave wrt the length of each segment,



Segmentation of a Poisson process (2/3)

Classical contrasts (negative log-likelihood, least-square) are

- additive wrt the segments and
- concave wrt the length of each segment,

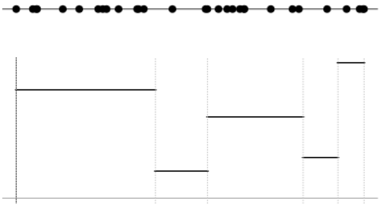


→ The set of optimal change points in included in

$$\{T_1^-, T_1, T_2^-, T_2, \dots T_i^-, T_i, \dots, T_n^-, T_n\}$$

→ The continuous optimization problem turns into a discrete optimization problem
→ Dynamic programming algorithm $= \mathcal{O}(n^2)$.

Segmentation of a Poisson process (3/3)

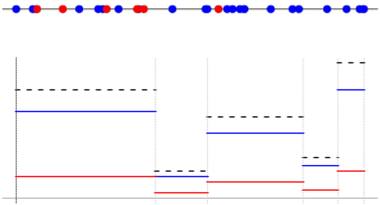


Lazy model selection. Thining property:

- independent processes with proportional internsities and common change point;
- cross-validation procedure to choose K.

[1] Modèles statistiques pour des données dépendantes et applications

Segmentation of a Poisson process (3/3)



Lazy model selection. Thining property:

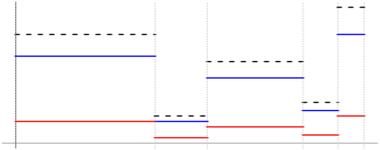
- independent processes with proportional internsities and common change point;
- cross-validation procedure to choose K.

[1] Modèles statistiques pour des données dépendantes et applications

Segmentation of a Poisson process (3/3)

Lazy model selection. Thining property:

- independent processes with proportional internsities and common change point;
- cross-validation procedure to choose K.



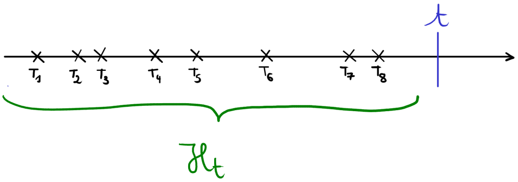
But segmenting a Poisson process

1. Does not provide any classification (although doable);
2. Does not account for the the self exciting (or inhibiting) nature of some processes;
3. Does not fit the scope of RMR2024[1].

[1] Modèles statistiques pour des données dépendantes et applications

Outline

Univariate Hawkes process



(Conditional) intensity function for the Hawkes process [Haw71]:

$$\lambda(t \mid \mathcal{H}_t) = \lambda(t) = \lambda_0 + \sum_{T_k < t} h(t - T_k)$$

- λ_0 = baseline
- h = kernel = influence of past events

Self-exciting exponential Hawkes process

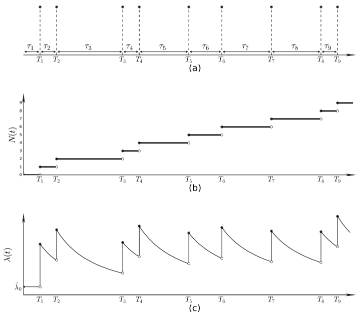
$$\lambda(t) = \lambda_0 + \sum_{T_k < t} a e^{-b(t - T_k)}$$

Self exciting: Each event increases the probability of observing another event

Self-exciting exponential Hawkes process

$$\lambda(t) = \lambda_0 + \sum_{T_k < t} a e^{-b(t - T_k)}$$

Self exciting: Each event increases the probability of observing another event



- Exponential kernel function $h(t) = ae^{-bt}$
- $a \geqslant 0$ to ensure that λ is non negative
- $a/b < 1$ to ensure stationarity
- Applications: sismology, epidemiology, neuroscience, ecology, ...

Discrete time Hawkes process

Continuous time exponential Hawkes process

$$\lambda(t) = \lambda_0 + \sum_{T_k < t} a e^{-b(t-T_k)}$$

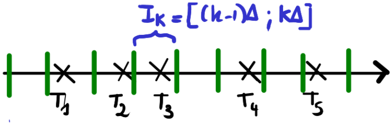
Discrete time Hawkes process

Continuous time exponential Hawkes process

$$\lambda(t) = \lambda_0 + \sum_{T_k < t} a e^{-b(t - T_k)}$$

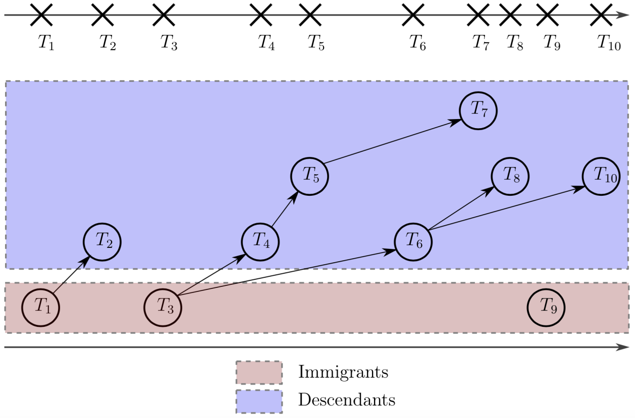
Discretization [Seo15]

- $I_k = [\tau_{k-1}; \tau_k]$ with $\tau_k = k\Delta$
- $N_k = N(I_k)$ the number of events on I_k



- Distribution of $(N_k)_{k \geqslant 1}$?

Cluster representation



- Immigrants arrive at rate λ_0
- Each immigrant or descendant produces new individuals at rate $h(t - T)$

Discrete time Hawkes process

$N_k = N(I_k)$ the number of events on $I_k = [\tau_{k-1}; \tau_k]$

Discrete time Hawkes process

$N_k = N(I_k)$ the number of events on $I_k = [\tau_{k-1}; \tau_k]$

Count distribution

$$N_k \stackrel{\Delta}{=} B_k + \sum_{\ell \leqslant k-1} \sum_{T \in I_\ell} M_T(I_k) + R_k$$

- $B_k \sim \mathcal{P}(\lambda_0 \Delta)$ discrete immigrant process
- $M_T(I_k) \sim \mathcal{P}\left(c(a, b, \Delta) e^{-b(\tau_{k-1} - T)}\right)$ descendants of $T < \tau_{k-1}$
- R_k number of descendants of points $T \in I_k$ within I_k

Discrete time Hawkes process

$N_k = N(I_k)$ the number of events on $I_k = [\tau_{k-1}; \tau_k]$

Count distribution

$$N_k \triangleq B_k + \sum_{\ell \leqslant k-1} \sum_{T \in I_\ell} M_T(I_k) + R_k$$

- $B_k \sim \mathcal{P}(\lambda_0 \Delta)$ discrete immigrant process
- $M_T(I_k) \sim \mathcal{P}\left(c(a, b, \Delta) e^{-b(\tau_{k-1} - T)}\right)$ descendants of $T < \tau_{k-1}$
- R_k number of descendants of points $T \in I_k$ within I_k

Approximation when Δ is small

- $M_T(I_k) \simeq \mathcal{P}\left(c(a, b, \Delta) e^{-b(\tau_{k-1} - \tau_{\ell-1})}\right)$ for $T \in I_\ell = [\tau_{\ell-1}; \tau_\ell]$
- $R_k \simeq 0$

Markovian reformulation

Approximation of N_k

$$Y_k \mid \{Y_\ell\}_{\ell \leqslant k-1} \sim \mathcal{P}\left(\mu + \sum_{\ell=1}^{\infty} \alpha \beta^\ell Y_{k-\ell}\right),$$

with $\mu = \lambda_0 \Delta$ and α, β depending on a, b, Δ.

$\{Y_k\}_{k \geqslant 1}$ is not a Markov chain but

Markovian reformulation

Approximation of N_k

$$Y_k \mid \{Y_\ell\}_{\ell \leqslant k-1} \sim \mathcal{P}\left(\mu + \sum_{\ell=1}^{\infty} \alpha\beta^\ell Y_{k-\ell}\right),$$

with $\mu = \lambda_0\Delta$ and α, β depending on a, b, Δ.

$\{Y_k\}_{k\geqslant 1}$ is not a Markov chain but defining $\{U_k\}_{k\geqslant 1}$

$$U_1 = 0, \qquad U_k = \alpha Y_{k-1} + \beta U_{k-1},$$

so that

$$Y_k|(U_{k-1}, Y_{k-1}) \sim \mathcal{P}\left(\alpha Y_{k-1} + \beta U_{k-1}\right),$$

then we have that

$$\{(Y_k, U_k)\}_{k\geqslant 1}$$

is a Markov Chain.

Outline

Discrete time Hawkes HMM

Model: Q hidden states

- Hidden path: $\{Z_k\}_{k \geqslant 1}$ homogeneous Markov chain with transition matrix π
- Observed counts: for $k \geqslant 1$, set $U_1 = 0$ and

$$Y_k \mid \{Y_\ell\}_{\ell \leqslant k-1} \sim \mathcal{P}\left(\mu_{Z_k} + \sum_{\ell=1}^{\infty} \alpha \beta^\ell Y_{k-\ell}\right)$$

Discrete time Hawkes HMM

Model: Q hidden states

- Hidden path: $\{Z_k\}_{k \geqslant 1}$ homogeneous Markov chain with transition matrix π
- Observed counts: for $k \geqslant 1$, set $U_1 = 0$ and

$$Y_k \mid \{Y_\ell\}_{\ell \leqslant k-1} \sim \mathcal{P}\left(\mu_{Z_k} + \sum_{\ell=1}^{\infty} \alpha \beta^\ell Y_{k-\ell}\right)$$

Assumptions:

- The immigration rate varies with the hidden path
- The number of offspring does not vary with the hidden path

Discrete time Hawkes HMM

Model: Q hidden states

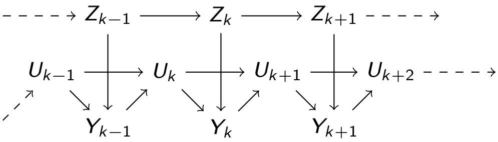
- Hidden path: $\{Z_k\}_{k \geqslant 1}$ homogeneous Markov chain with transition matrix π
- Observed counts: for $k \geqslant 1$, set $U_1 = 0$ and

$$Y_k \mid \{Y_\ell\}_{\ell \leqslant k-1} \sim \mathcal{P}\left(\mu_{Z_k} + \sum_{\ell=1}^{\infty} \alpha \beta^\ell Y_{k-\ell}\right)$$

Assumptions:

- The immigration rate varies with the hidden path
- The number of offspring does not vary with the hidden path

Graphical model:



Inference

Aim: Infer the parameter $\theta = ((\mu_q)_{1 \leqslant q \leqslant Q}, \pi)$

$$\widehat{\theta} = \log p_\theta(Y)$$

Inference

Aim: Infer the parameter $\theta = ((\mu_q)_{1 \leqslant q \leqslant Q}, \pi)$

$$\widehat{\theta} = \log p_\theta(Y)$$

EM algorithm: [DLR77]

$$\theta^{(h+1)} = \underbrace{\arg\max_\theta}_{\text{M step}} \underbrace{\mathbb{E}_{\theta^{(h)}}}_{\text{E step}} [\log p_\theta(Y, Z) \mid Y]$$

- E step: Evaluate $\ell^{(h)}(\theta) = \mathbb{E}_{\theta^{(h)}}[\log p_\theta(Y, Z) \mid Y]$ (forward-backward recursion)
- M step: Gradient descent, computing $\nabla_\theta \ell^{(h)}(\theta)$ by recursion

Inference

Aim: Infer the parameter $\theta = ((\mu_q)_{1 \leqslant q \leqslant Q}, \pi)$

$$\widehat{\theta} = \log p_\theta(Y)$$

EM algorithm: [DLR77]

$$\theta^{(h+1)} = \underbrace{\arg\max_\theta}_{\text{M step}} \underbrace{\mathbb{E}_{\theta^{(h)}}}_{\text{E step}} [\log p_\theta(Y, Z) \mid Y]$$

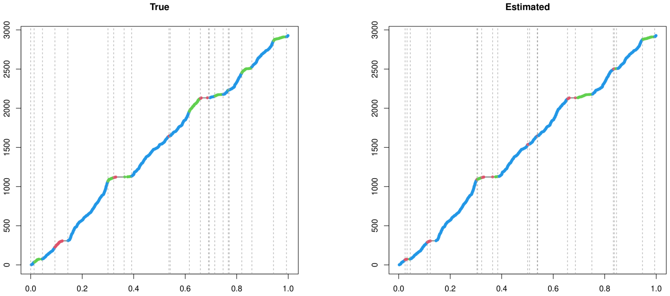
- E step: Evaluate $\ell^{(h)}(\theta) = \mathbb{E}_{\theta^{(h)}}[\log p_\theta(Y, Z) \mid Y]$ (forward-backward recursion)
- M step: Gradient descent, computing $\nabla_\theta \ell^{(h)}(\theta)$ by recursion

By-product: Classification

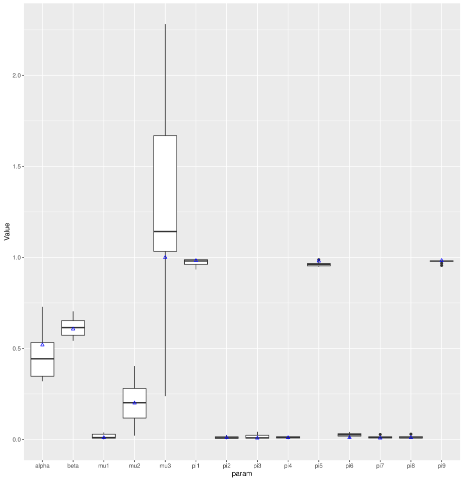
$$\widehat{Z}_k = \arg\max_q P_{\widehat{\theta}}\{Z_k = q \mid Y\}, \qquad \widehat{Z} = \arg\max_z p_{\widehat{\theta}}(Y, Z = z)$$

Synthetic data: classification

	States	1	2	3
Hawkes HMM	1	273.7	28.7	14.2
	2	37	166.3	96.9
	3	4.7	24.7	353.8
Poisson HMM	1	181	122.8	12.8
	2	136	111.1	53.1
	3	45.4	115.2	222.6



Synthetic data: parameter estimation



Outline

Model selection

Aim: select the number of hidden states Q

Model selection

Aim: select the number of hidden states Q

Penalized likelihood:

$$\log p_\theta(Y) = \mathbb{E}_\theta[\log p_\theta(Y, Z) \mid Y] + \mathcal{H}(p_\theta(Z \mid Y))$$

→ Standard criterion for discrete time HMM

$$BIC(Q) = \log p_{\widehat{\theta}_Q}(Y) - pen(\widehat{\theta}_Q),$$
$$ICL(Q) = \log p_\theta(Y) - \mathcal{H}(p_{\widehat{\theta}_Q}(Z \mid Y)) - pen(\widehat{\theta}_Q)$$

with

$$pen(\widehat{\theta}_Q) = \frac{1}{2}\log(N)(Q^2 + 2)$$

where N = number of time steps (i.e. discretized intervals) = tuning parameter

Goodness-of-fit

Time-change theorem [DVJ03] A sequence $(T_k)_{k \geqslant 1}$ is a realization of N if and only if $(\Lambda(T_k))_{k \geqslant 1}$ is a realization of a homogeneous Poisson process with unit intensity. where

$$\Lambda(t) = \int_0^t \lambda(u) du \qquad \text{(Compensator)}$$

Goodness-of-fit

Time-change theorem [DVJ03] A sequence $(T_k)_{k \geqslant 1}$ is a realization of N if and only if $(\Lambda(T_k))_{k \geqslant 1}$ is a realization of a homogeneous Poisson process with unit intensity. where

$$\Lambda(t) = \int_0^t \lambda(u) du \qquad \text{(Compensator)}$$

Goodness-of-fit test

- H_0: "$(T_k)_{k \geqslant 1}$ is a realization of a HMM-Hawkes process with parameter θ".
- Kolmogorov-Smirnov test between

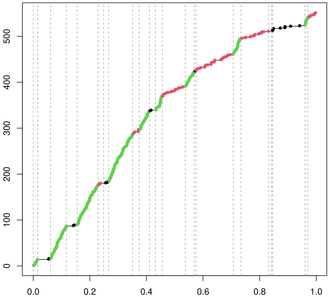
$$\left(\Lambda_{\widehat{\theta}}(T_{k+1}) - \Lambda_{\widehat{\theta}}(T_k)\right)_{k \geqslant 1}$$

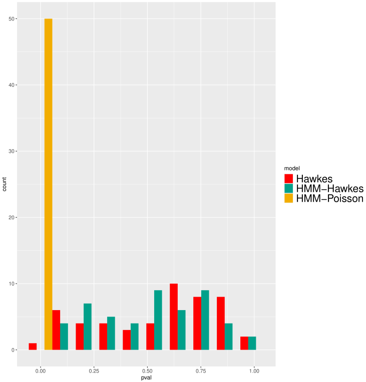
  and an exponential distribution $\mathcal{E}(1)$.

Comments

- Same test for alternative models (Hawkes, HMM-Poisson)
- Train/test samples (resampling procedure, [RBRGTM14])

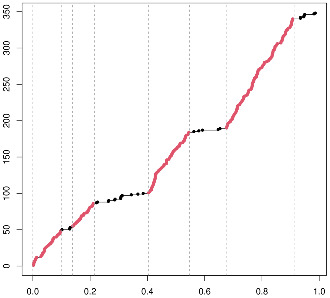
Synthetic data: Goodness-of-fit (1/2)

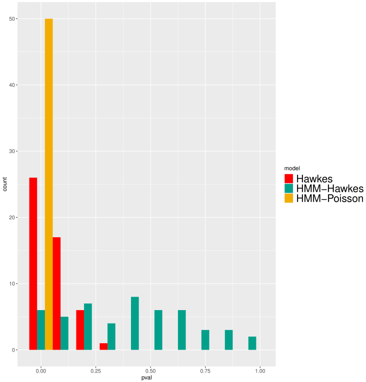




- The test rejects the homogeneous Poisson but does not differentiate the homogeneous Hawkes process from the HMM-Hawkes process.

Synthetic data: Goodness-of-fit (2/2)

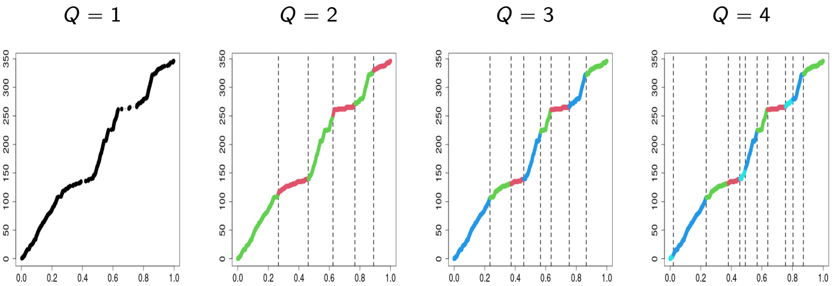




- The test is able to detect that the point process is neither an homogeneous Hawkes nor a HMM-Poisson process

Preliminary results on bat cries

Back to the recording of bat cries over one night:



- $Q = 2, 3$, hidden states? ($\widehat{Q}_{BIC} = 1$ or 2, depending on N)
- States = behavior (transit, foraging), species?

Outline

Some future works (1/2)

Multivariate Hawkes process: Consider p simultaneous processes $(N^{(i)})_{1 \leqslant i \leqslant p}$ (i.e. p neurons, bat species, ...)

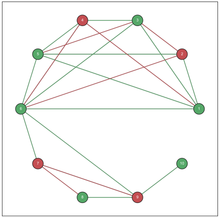
$$\lambda^{(i)}(t) = \lambda_0^i + \sum_{i=1}^{M} \sum_{T_k^j < t} h_{i,j}(t - T_k^j)$$

Exponential version:

$$h_{i,j}(t - T_k^j) = a_{i,j} e^{-b(t - T_k^j)}$$

where sparse interaction matrix $A = [a_{i,j}]_{1 \leqslant i,j \leqslant p}$

- Interaction network between neurons, species, ...



Some future works (2/2)

Modelling inhibition: Non-linear Hawkes process

$$\lambda(t) = \phi\left(\lambda_0 + \sum_{T_k \leqslant t} h(t - T_k)\right)$$

with $h < 0$.

Some future works (2/2)

Modelling inhibition: Non-linear Hawkes process

$$\lambda(t) = \phi\left(\lambda_0 + \sum_{T_k \leqslant t} h(t - T_k)\right)$$

with $h < 0$.

Effect of the hidden state: State-dependent parameters α and/or β

$$Y_k \mid \{Y_\ell\}_{\ell \leqslant k-1} \sim \mathcal{P}\left(\mu_{Z_k} + \sum_{\ell=1}^{\infty} \alpha_{Z_k} (\beta_{Z_k})^\ell Y_{k-\ell}\right)$$

References

Felix Cheysson. *hawkesbow: Estimation of Hawkes Processes from Binned Observations*, 2021. R package version 1.0.2.

C Dion-Blanc, E Lebarbier, and S Robin. Multiple change-point detection for Poisson processes. Technical Report 2302.09103, arXiv, 2023.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 39:1–38, 1977.

D. J Daley and D. Vere-Jones. *An introduction to the theory of point processes: volume I: elementary theory and methods*. Springer, 2003.

A. G Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

P. Reynaud-Bouret, V. Rivoirard, F. Grammont, and C. Tuleau-Malot. Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. *The Journal of Mathematical Neuroscience*, 4:1–41, 2014.

Y. Seol. Limit theorems for discrete hawkes processes. *Statistics & Probability Letters*, 99:223–229, 2015.

Backup: Discrete HMM

Conversion formulas from continuous to discrete Hawkes

$$\alpha = \frac{e^{b\Delta} - 1}{b}, \qquad \beta = e^{-b\Delta}$$

Backup: Discrete HMM

Conversion formulas from continuous to discrete Hawkes

$$\alpha = \frac{e^{b\Delta} - 1}{b}, \qquad \beta = e^{-b\Delta}$$

3-step initialization

- Homogeneous Hawkes for the reproduction parameters α and β (`hawkesbow` R package [Che21])
- Poisson-HMM for the rates $\mu_1, \dots, \mu_Q$
- Correction $\mu_k \to \widetilde{\mu}_k$ to account for reproduction rate

Backup: GoF

Change-time for the recording of bat cries over one night:

$Q = 1$ $Q = 2$ $Q = 3$ $Q = 4$