Markov-switching (discrete-time) Hawkes process

S. Robin

joint work with A. Bonnet

LPSM, Sorbonne université

Markov, Semi-Markov Models and Associated Fields , Paris, Jul. 2025

'Motivation'

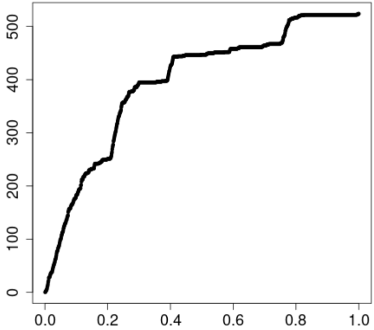
Counting process

Overnight recording of bat cries in continuous time

'Motivation'

Counting process

Overnight recording of bat cries in continuous time

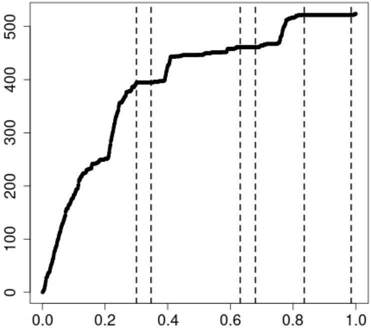


'Motivation'

Counting process

Overnight recording of bat cries in continuous time

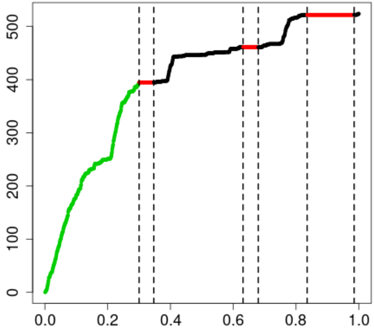
- Can we detect changes in the distribution of events?



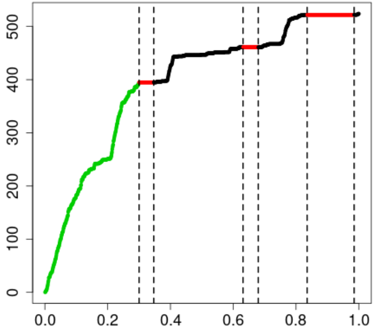
'Motivation'

Counting process
Overnight recording of bat cries in continuous time

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



'Motivation'



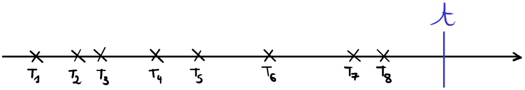
Counting process
Overnight recording of bat cries in continuous time

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

Modelling. Point process with (latent) Markov switching regime

Point process

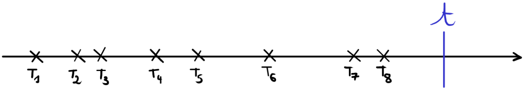
Reminder.



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

Point process

Reminder.



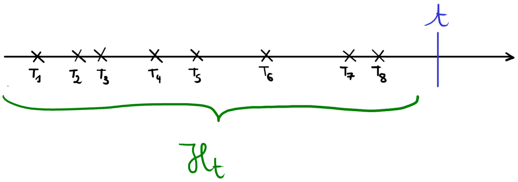
- $(T_k)_{k \geqslant 1}$ a random collection of points
- Count process $H(t) = \sum_{k \geqslant 1} \mathbb{I}\{T_k \leqslant t\}$
- Intensity function $\lambda(t)$: immediate probability of observing an event at time t

Examples

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

Outline

Univariate Hawkes process



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

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

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

Self-exciting exponential Hawkes process

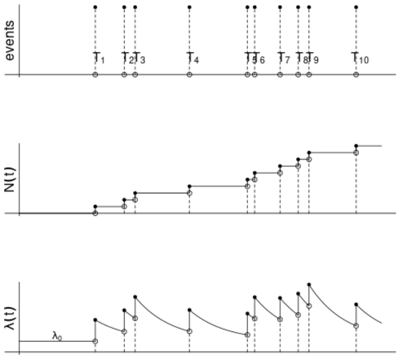
$$\lambda(t) = \lambda_0 + a \sum_{T_k < t} e^{-b(t - T_k)}$$

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

Self-exciting exponential Hawkes process

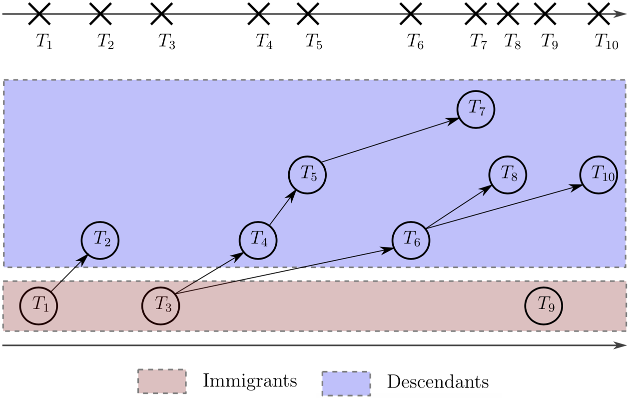
$$\lambda(t) = \lambda_0 + a \sum_{T_k < t} e^{-b(t - T_k)}$$

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, vulcanology, neurosciences, ecology, ...

Cluster representation [HO74]



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

Outline

Discrete-time Hawkes process

Continuous time exponential Hawkes process

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

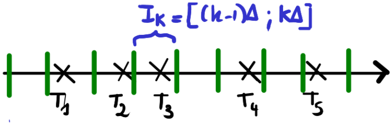
Discrete-time Hawkes process

Continuous time exponential Hawkes process

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

Discretization [Seo15,Kir16,Kir17]

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



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

Decomposition of the count

H_k = number of events on $I_k = [\tau_{k-1}; \tau_k]$

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

where

Decomposition of the count

H_k = number of events on $I_k = [\tau_{k-1}; \tau_k]$

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

where

- B_k = number of immigrants within I_k:

$$B_k \sim \mathcal{P}(\mu)$$

  with $\mu = \lambda_0 \Delta$,

Decomposition of the count

H_k = number of events on $I_k = [\tau_{k-1}; \tau_k]$

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

where

- B_k = number of immigrants within I_k:

$$B_k \sim \mathcal{P}(\mu)$$

  with $\mu = \lambda_0 \Delta$,

- $M_T(I_k)$ = number of descendants of $T < \tau_k$ within I_k:

$$M_T(I_k) \sim \mathcal{P}\left(\int_{I_k} a e^{-b(t-T)} \, \mathrm{d}t)\right) = \mathcal{P}\left(\alpha e^{-b(\tau_k - T)}\right)$$

  with $\alpha = a(e^{b\Delta} - 1)/b$,

Decomposition of the count

H_k = number of events on $I_k = [\tau_{k-1}; \tau_k]$

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

where

- B_k = number of immigrants within I_k:

$$B_k \sim \mathcal{P}(\mu)$$

  with $\mu = \lambda_0 \Delta$,

- $M_T(I_k)$ = number of descendants of $T < \tau_k$ within I_k:

$$M_T(I_k) \sim \mathcal{P}\left(\int_{I_k} a e^{-b(t-T)} \, \mathrm{d}t\right) = \mathcal{P}\left(\alpha e^{-b(\tau_k - T)}\right)$$

  with $\alpha = a(e^{b\Delta} - 1)/b$,

- R_k = number of descendants of points $T \in I_k$ within I_k

Discrete time Hawkes process

When Δ is small:

- $R_k \simeq 0$
- For $T \in I_\ell$: $e^{-b(\tau_k - T)} \simeq e^{-b(\tau_k - \tau_\ell)} = \beta^{k-\ell}$ with $\beta = e^{-b\Delta}$, so

$$\sum_{\ell \leqslant k-1} \sum_{T \in I_\ell} M_T(I_k) \overset{\Delta}{\simeq} \sum_{\ell \leqslant k-1} \sum_{T \in I_\ell} \mathcal{P}\left(\alpha \beta^{k-\ell}\right) \overset{\Delta}{=} \mathcal{P}\left(\alpha \sum_{\ell \leqslant k-1} H_{k-\ell} \beta^{\ell-1}\right)$$

Discrete time Hawkes process

When Δ is small:

- $R_k \simeq 0$
- For $T \in I_\ell$: $e^{-b(\tau_k - T)} \simeq e^{-b(\tau_k - \tau_\ell)} = \beta^{k-\ell}$ with $\beta = e^{-b\Delta}$, so

$$\sum_{\ell \leqslant k-1} \sum_{T \in I_\ell} M_T(I_k) \overset{\Delta}{\simeq} \sum_{\ell \leqslant k-1} \sum_{T \in I_\ell} \mathcal{P}\left(\alpha \beta^{k-\ell}\right) \overset{\Delta}{=} \mathcal{P}\left(\alpha \sum_{\ell \leqslant k-1} H_{k-\ell} \beta^{\ell-1}\right)$$

Discrete-time Hawkes process $Y = \{Y_k\}_{k \leqslant 1}$.

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

[Kir16]: convergence toward a continuous-time Hawkes process.

Outline

Markovian representation

Discrete-time Hawkes process $Y = \{Y_k\}_{k \leqslant 1}$.

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

$(Y_k)_{k \geqslant 1}$ is not a Markov chain (because of infinite memory).

Markovian representation

Discrete-time Hawkes process $Y = \{Y_k\}_{k \leqslant 1}$.

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

$(Y_k)_{k \geqslant 1}$ is not a Markov chain (because of infinite memory).

Markovian representation.

- Define

$$U_1 = 0, \qquad U_k = \alpha \sum_{\ell=1}^{k} \beta^{\ell-1} Y_{k-\ell},$$

Markovian representation

Discrete-time Hawkes process $Y = \{Y_k\}_{k \leqslant 1}$.

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

$(Y_k)_{k \geqslant 1}$ is not a Markov chain (because of infinite memory).

Markovian representation.

- Define

$$U_1 = 0, \qquad\qquad U_k = \alpha \sum_{\ell=1}^{k} \beta^{\ell-1} Y_{k-\ell},$$

- then, for $k \geqslant 1$ (with $U_0 = Y_0 = 0$)

$$U_k = \alpha Y_{k-1} + \beta U_{k-1}, \qquad\qquad Y_k \mid U_k \sim \mathcal{P}(\mu + U_k).$$

Markovian representation

Discrete-time Hawkes process $Y = \{Y_k\}_{k \leqslant 1}$.

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

$(Y_k)_{k \geqslant 1}$ is not a Markov chain (because of infinite memory).

Markovian representation.

- Define

$$U_1 = 0, \qquad U_k = \alpha \sum_{\ell=1}^{k} \beta^{\ell-1} Y_{k-\ell},$$

- then, for $k \geqslant 1$ (with $U_0 = Y_0 = 0$)

$$U_k = \alpha Y_{k-1} + \beta U_{k-1}, \qquad Y_k \mid U_k \sim \mathcal{P}(\mu + U_k).$$

so $((Y_k, U_k))_{k \geqslant 1}$ forms a Markov Chain.

Graphical model

Discrete time Hawkes process.

$$(Y_k)_{k \geqslant 1} \sim \textit{Discrete Hawkes}(\mu, \alpha, \beta)$$

$$U_1 = 0, \qquad U_k = \alpha Y_{k-1} + \beta U_{k-1}, \qquad Y_k \sim \mathcal{P}(\mu + U_k)$$

Graphical model

Discrete time Hawkes process.

$$(Y_k)_{k \geqslant 1} \sim \textit{Discrete Hawkes}(\mu, \alpha, \beta)$$

$$U_1 = 0, \qquad U_k = \alpha Y_{k-1} + \beta U_{k-1}, \qquad Y_k \sim \mathcal{P}(\mu + U_k)$$

Graphical model:

$$\begin{array}{ccccccccc}
 & U_{k-1} & \longrightarrow & U_k & \longrightarrow & U_{k+1} & \longrightarrow & U_{k+2} & \dashrightarrow \\
\nearrow & \searrow & \nearrow & \searrow & \nearrow & \searrow & \nearrow & & \\
 & & Y_{k-1} & & Y_k & & Y_{k+1} & &
\end{array}$$

$(U_k)_{k \geqslant 1} =$ memory, $\quad (Y_k)_{k \geqslant 1} =$ observed process.

$$\begin{aligned}
p\left(U_k, Y_k \mid (U_\ell, Y_\ell)_{\ell \leqslant k-1}\right) &= p\left(U_k, Y_k \mid U_{k-1}, Y_{k-1}\right) \\
&= p\left(U_k \mid U_{k-1}, Y_{k-1}\right) p\left(Y_k \mid U_k\right)
\end{aligned}$$

Outline

Discrete time Hawkes HMM

Model: Q hidden states

- Hidden path: $(Z_k)_{k \geqslant 1}$ homogeneous Markov chain with Q states, transition matrix π and initial distribution ν:

$$(Z_k)_{k \geqslant 1} \sim MC_Q(\nu, \pi)$$

Discrete time Hawkes HMM

Model: Q hidden states

- Hidden path: $(Z_k)_{k\geqslant 1}$ homogeneous Markov chain with Q states, transition matrix π and initial distribution ν:

$$(Z_k)_{k\geqslant 1} \sim MC_Q(\nu, \pi)$$

- Observed counts: for $k \geqslant 1$ and

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

or, for $k \geqslant 1$ (with $U_0 = Y_0 = 0$)

$$U_k = \alpha Y_{k-1} + \beta U_{k-1}, \qquad\qquad Y_k \mid U_k \sim \mathcal{P}\left(\mu_{Z_k} + U_k\right)$$

Discrete time Hawkes HMM

Model: Q hidden states

- Hidden path: $(Z_k)_{k \geqslant 1}$ homogeneous Markov chain with Q states, transition matrix π and initial distribution ν:

$$(Z_k)_{k \geqslant 1} \sim MC_Q(\nu, \pi)$$

- Observed counts: for $k \geqslant 1$ and

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

or, for $k \geqslant 1$ (with $U_0 = Y_0 = 0$)

$$U_k = \alpha Y_{k-1} + \beta U_{k-1}, \qquad Y_k \mid U_k \sim \mathcal{P}\left(\mu_{Z_k} + U_k\right)$$

Assumptions:

- The immigration rate μ varies with the hidden state
- The distribution of the number of offspring (α, β) does not vary with the hidden state

Discrete time Hawkes HMM

Graphical model:

$$\begin{tikzcd}
{} \arrow[rr, dashed] & & Z_{k-1} \arrow[rr] \arrow[dd] & & Z_k \arrow[rr] \arrow[dd] & & Z_{k+1} \arrow[rr, dashed] \arrow[dd] & & {} \\
& U_{k-1} \arrow[rr] \arrow[dr] & & U_k \arrow[rr] \arrow[dr] & & U_{k+1} \arrow[rr] \arrow[dr] & & U_{k+2} \arrow[r, dashed] & {} \\
{} \arrow[ur, dashed] & & Y_{k-1} \arrow[ur] & & Y_k \arrow[ur] & & Y_{k+1} \arrow[ur] & &
\end{tikzcd}$$

$(Z_k)_{k\geqslant 1}$ = hidden path, $\quad (U_k)_{k\geqslant 1}$ = memory, $\quad (Y_k)_{k\geqslant 1}$ = observed process.

Discrete time Hawkes HMM

Graphical model:

$$
\begin{tikzcd}
 & {} \arrow[r, dashed] & Z_{k-1} \arrow[rr] \arrow[dd] & & Z_k \arrow[rr] \arrow[dd] & & Z_{k+1} \arrow[r, dashed] \arrow[dd] & {} & \\
 & U_{k-1} \arrow[rr] \arrow[dr] & & U_k \arrow[rr] \arrow[dr] & & U_{k+1} \arrow[rr] \arrow[dr] & & U_{k+2} \arrow[r, dashed] & {} \\
{} \arrow[ur, dashed] & & Y_{k-1} \arrow[ur] & & Y_k \arrow[ur] & & Y_{k+1} \arrow[ur] & &
\end{tikzcd}
$$

$(Z_k)_{k \geqslant 1}$ = hidden path, $(U_k)_{k \geqslant 1}$ = memory, $(Y_k)_{k \geqslant 1}$ = observed process.

Remarks:

- The memory of the past is 'stored' in the variable U_k, which can still be computed recursively ($U_k = \alpha Y_{k-1} + \beta U_{k-1}$)
- The Markovian property still holds if the influence of the past varies with the hidden state ($\alpha \to \alpha_q, \beta \to \beta_q$).

Outline

Identifiability

Proposition: The model parameter $\theta = (\nu, \pi, (\mu_q)_{1 \leqslant q \leqslant Q}, \alpha, \beta)$ is identifiable from the joint distribution $p_\theta^{Y_1, Y_2, Y_3}$:

$$\theta' \neq \theta \qquad \Rightarrow \qquad p_{\theta'}^{Y_1, Y_2, Y_3} \neq p_\theta^{Y_1, Y_2, Y_3}.$$

[1] The generic technique from [AMR09] does not apply here.

Identifiability

Proposition: The model parameter $\theta = (\nu, \pi, (\mu_q)_{1 \leqslant q \leqslant Q}, \alpha, \beta)$ is identifiable from the joint distribution $p_\theta^{Y_1, Y_2, Y_3}$:

$$\theta' \neq \theta \qquad \Rightarrow \qquad p_{\theta'}^{Y_1, Y_2, Y_3} \neq p_\theta^{Y_1, Y_2, Y_3}.$$

Sketch of proof[1]. Because

$$p_\theta^{Y_1, Y_2, Y_3}(x, y, z) = \sum_{1 \leqslant q, \ell, m \leqslant Q} \nu_q \pi_{q\ell} \pi_{\ell m} \mathcal{P}(x; \mu_q) \mathcal{P}(y; \mu_\ell + \alpha x) \mathcal{P}(z; \mu_m + \alpha\beta x + \alpha y),$$

and because finite Poisson mixtures are identifiable [Tei61], one may identify

[1] The generic technique from [AMR09] does not apply here.

Identifiability

Proposition: The model parameter $\theta = (\nu, \pi, (\mu_q)_{1 \leqslant q \leqslant Q}, \alpha, \beta)$ is identifiable from the joint distribution $p_\theta^{Y_1, Y_2, Y_3}$:

$$\theta' \neq \theta \qquad \Rightarrow \qquad p_{\theta'}^{Y_1, Y_2, Y_3} \neq p_\theta^{Y_1, Y_2, Y_3}.$$

Sketch of proof[1]. Because

$$p_\theta^{Y_1, Y_2, Y_3}(x, y, z) = \sum_{1 \leqslant q, \ell, m \leqslant Q} \nu_q \pi_{q\ell} \pi_{\ell m} \mathcal{P}(x; \mu_q) \mathcal{P}(y; \mu_\ell + \alpha x) \mathcal{P}(z; \mu_m + \alpha\beta x + \alpha y),$$

and because finite Poisson mixtures are identifiable [Tei61], one may identify

1. ν and μ from $p_\theta(Y_1)$, [sum over y and z]

[1] The generic technique from [AMR09] does not apply here.

Identifiability

Proposition: The model parameter $\theta = (\nu, \pi, (\mu_q)_{1 \leqslant q \leqslant Q}, \alpha, \beta)$ is identifiable from the joint distribution $p_\theta^{Y_1, Y_2, Y_3}$:

$$\theta' \neq \theta \qquad \Rightarrow \qquad p_{\theta'}^{Y_1, Y_2, Y_3} \neq p_\theta^{Y_1, Y_2, Y_3}.$$

Sketch of proof[1]. Because

$$p_\theta^{Y_1, Y_2, Y_3}(x, y, z) = \sum_{1 \leqslant q, \ell, m \leqslant Q} \nu_q \pi_{q\ell} \pi_{\ell m} \mathcal{P}(x; \mu_q) \mathcal{P}(y; \mu_\ell + \alpha x) \mathcal{P}(z; \mu_m + \alpha\beta x + \alpha y),$$

and because finite Poisson mixtures are identifiable [Tei61], one may identify

1. ν and μ from $p_\theta(Y_1)$, [sum over y and z]
2. then α from $p_\theta(Y_2 \mid Y_1 = 1)$, [fix $x = 1$, sum over z]

[1] The generic technique from [AMR09] does not apply here.

Identifiability

Proposition: The model parameter $\theta = (\nu, \pi, (\mu_q)_{1 \leqslant q \leqslant Q}, \alpha, \beta)$ is identifiable from the joint distribution $p_\theta^{Y_1, Y_2, Y_3}$:

$$\theta' \neq \theta \qquad \Rightarrow \qquad p_{\theta'}^{Y_1, Y_2, Y_3} \neq p_\theta^{Y_1, Y_2, Y_3}.$$

Sketch of proof[1]. Because

$$p_\theta^{Y_1, Y_2, Y_3}(x, y, z) = \sum_{1 \leqslant q, \ell, m \leqslant Q} \nu_q \pi_{q\ell} \pi_{\ell m} \mathcal{P}(x; \mu_q) \mathcal{P}(y; \mu_\ell + \alpha x) \mathcal{P}(z; \mu_m + \alpha\beta x + \alpha y),$$

and because finite Poisson mixtures are identifiable [Tei61], one may identify

1. ν and μ from $p_\theta(Y_1)$, [sum over y and z]
2. then α from $p_\theta(Y_2 \mid Y_1 = 1)$, [fix $x = 1$, sum over z]
3. then β from $p_\theta(Y_3 \mid Y_1 = 1, Y_2 = 0)$, [fix $x = 1$, $y = 0$]

[1] The generic technique from [AMR09] does not apply here.

Identifiability

Proposition: The model parameter $\theta = (\nu, \pi, (\mu_q)_{1 \leqslant q \leqslant Q}, \alpha, \beta)$ is identifiable from the joint distribution $p_\theta^{Y_1, Y_2, Y_3}$:

$$\theta' \neq \theta \qquad \Rightarrow \qquad p_{\theta'}^{Y_1, Y_2, Y_3} \neq p_\theta^{Y_1, Y_2, Y_3}.$$

Sketch of proof[1]. Because

$$p_\theta^{Y_1, Y_2, Y_3}(x, y, z) = \sum_{1 \leqslant q, \ell, m \leqslant Q} \nu_q \pi_{q\ell} \pi_{\ell m} \mathcal{P}(x; \mu_q) \mathcal{P}(y; \mu_\ell + \alpha x) \mathcal{P}(z; \mu_m + \alpha \beta x + \alpha y),$$

and because finite Poisson mixtures are identifiable [Tei61], one may identify

1. ν and μ from $p_\theta(Y_1)$, [sum over y and z]
2. then α from $p_\theta(Y_2 \mid Y_1 = 1)$, [fix $x = 1$, sum over z]
3. then β from $p_\theta(Y_3 \mid Y_1 = 1, Y_2 = 0)$, [fix $x = 1$, $y = 0$]
4. then π from the joint mixture (proven identifiable) [sum over z]

$$p_\theta^{Y_1, Y_2}(x, y) = \sum_{1 \leqslant q, \ell \leqslant Q} \nu_q \pi_{q\ell} \mathcal{P}(x; \mu_q) \mathcal{P}(y; \mu_\ell + \alpha x).$$

[1] The generic technique from [AMR09] does not apply here.

Inference

Aim: Infer the parameter θ

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

Inference

Aim: Infer the parameter θ

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

EM algorithm for HMM: [DLR77,CMR05]

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

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

Inference

Classification:

Marginal: $$\widehat{Z}_k = \arg\max_q P_{\widehat{\theta}}\{Z_k = q \mid Y\},$$

Joint (Viterbi): $$\widehat{Z} = \arg\max_z P_{\widehat{\theta}}\{Z = z \mid Y\}$$

Inference

Classification:

Marginal: $$\widehat{Z}_k = \arg\max_q P_{\widehat{\theta}}\{Z_k = q \mid Y\},$$

Joint (Viterbi): $$\widehat{Z} = \arg\max_z P_{\widehat{\theta}}\{Z = z \mid Y\}$$

Model selection: Penalized likelihood

$$AIC_Q = \log p_{\widehat{\theta}_Q}(Y) - D_Q,$$

$$BIC_Q = \log p_{\widehat{\theta}_Q}(Y) - D_Q \frac{\log(N)}{2}$$

with D_Q = number of parameters $= 2 + Q^2$ and N = number of time bins.

Outline

Simulation design ($Q = 3$)

- Baseline continuous parameters: $m^0 = [10,\ 200,\ 1000], \quad a^0 = 40, \quad b = 160$

Simulation design ($Q = 3$)

- Baseline continuous parameters: $m^0 = [10,\ 200,\ 1000], \quad a^0 = 40, \quad b = 160$

- Increasing signal: $\lambda = 0.5,\ 1,\ 1.5,\ 2$

$$a = \lambda\ a^0, \qquad m = \lambda\ m^0.$$

Simulation design ($Q = 3$)

- Baseline continuous parameters: $m^0 = [10,\ 200,\ 1000], \quad a^0 = 40, \quad b = 160$
- Increasing signal: $\lambda = 0.5,\ 1,\ 1.5,\ 2$

$$a = \lambda\ a^0, \qquad m = \lambda\ m^0.$$

- Simulated process:

$$(H_t)_{0 \leqslant t \leqslant 1} \sim \textit{Heterogeneous Continuous Hawkes}(a, b^0, m)$$

Simulation design ($Q = 3$)

- Baseline continuous parameters: $m^0 = [10,\ 200,\ 1000], \quad a^0 = 40, \quad b = 160$

- Increasing signal: $\lambda = 0.5,\ 1,\ 1.5,\ 2$

$$a = \lambda\ a^0, \qquad m = \lambda\ m^0.$$

- Simulated process:

$$(H_t)_{0 \leqslant t \leqslant 1} \sim \textit{Heterogeneous Continuous Hawkes}(a, b^0, m)$$

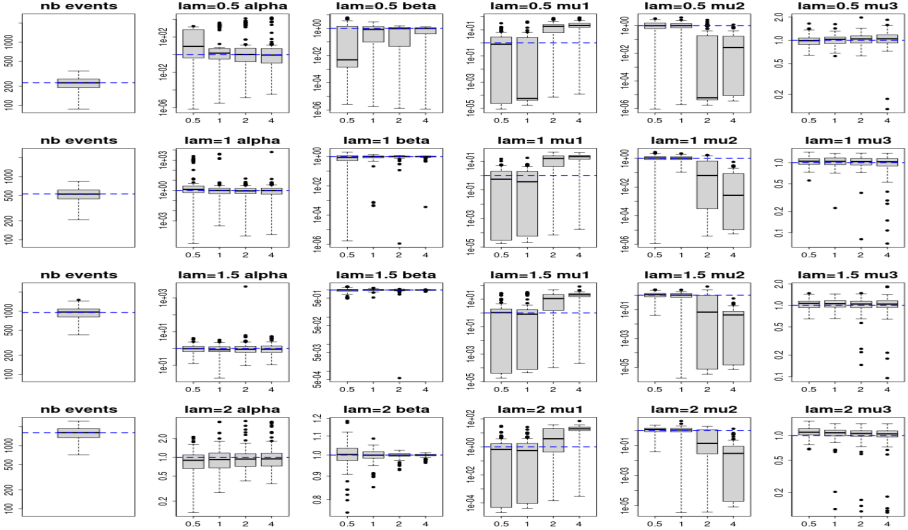
- Discretized process: $n = H(1), \quad N = c\ n, \quad c = 0.5,\ 1,\ 2,\ 4,$

$$Y_k = H\left(\left[\frac{k-1}{N}; \frac{k}{N}\right]\right), \qquad k = 1, \dots N.$$

$\rightarrow$ not a discrete-time Hawkes process as defined earlier

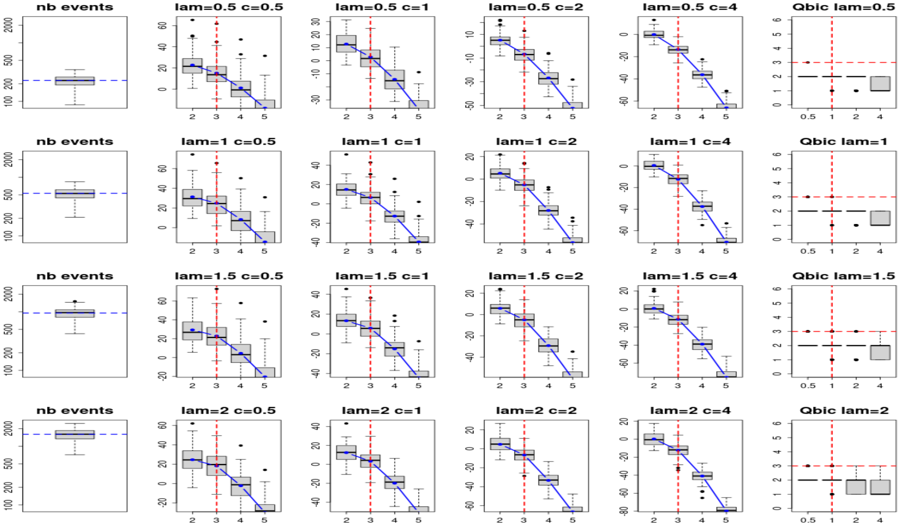
Simulation results ($Q^* = 3$)

Parameter estimates. Distribution of $\widehat{\theta} - \theta^*$ ($Q = Q^*$)



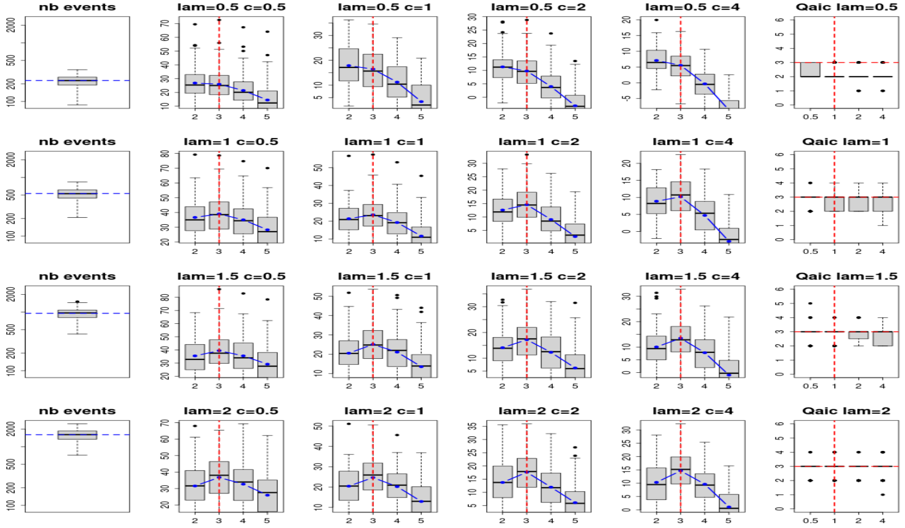
Simulation results ($Q^* = 3$)

Model selection: BIC. Distribution of $BIC_Q - BIC_1$



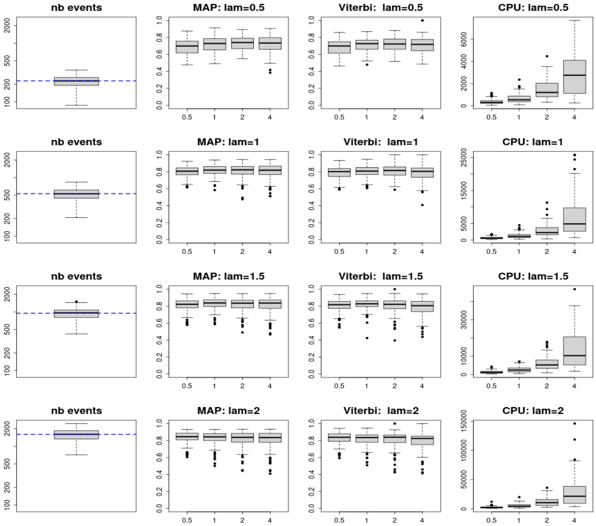
Simulation results ($Q^* = 3$)

Model selection: AIC. Distribution of $AIC_Q - AIC_1$



Simulation results ($Q^* = 3$)

Classification. MAP / Viterbi (+ comput. time)



Simulation conclusions

- Inference easier when more signal (large λ)!!!
- Inference easier with thinner discretization step (large N)
  But at the price of a higher computational cost
- BIC does not capture the right number of states
  Sequences not simulated according to the model
- AIC does, with reasonable signal (λ) and discretization (N)
  Blind to the simulation shift from the model?

Simulation conclusions

- Inference easier when more signal (large λ)!!!
- Inference easier with thinner discretization step (large N)
  But at the price of a higher computational cost
- BIC does not capture the right number of states
  Sequences not simulated according to the model
- AIC does, with reasonable signal (λ) and discretization (N)
  Blind to the simulation shift from the model?

Practical recommendations.

Take $N = 2n$ and use AIC

Outline

Bat cries

Data set. 1555 overnight recordings all over France

Bat cries

Data set. 1555 overnight recordings all over France

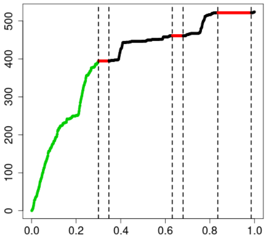
Poisson vs Hawkes / Homogeneous vs HMM. Best model based on AIC

	Poisson	Hawkes	Total
Homogeneous	34	353	387
Hidden Markov	24	1144	1168
Total	58	1497	1555

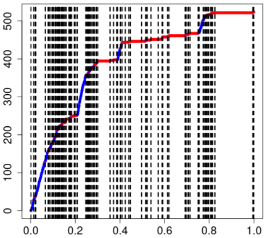
- Memory (95%) and heterogeneity (75%) are present in most sequences
- Hawkes-HMM best fits almost 3 sequences out of 4.

Example

Hawkes HMM ($\widehat{Q} = 3$)



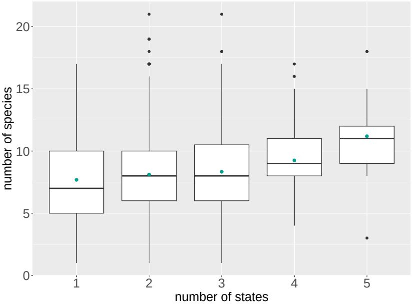
Poisson HMM ($\widehat{Q} = 4$)



- Poisson-HMM needs many state changes to account for self-excitation
- Hawkes-HMM state changes do not correspond to slope changes

States and species

The number of bat species was also recorded



- The number of states does not match the number of species

Outline

Summary

What we did.

- The discretized Hawkes process (with exponential kernel) is a Markov model

  $\rightarrow$ The discretized Markov switching Hawkes process is a hidden Markov model

- The standard EM machinery applies to achieve maximum likelihood inference

- Not shown: initialization based on existing estimation procedures for homogeneous Hawkes ([Che21]) and Poisson HMM

Discussion

What we did not do.

- Goodness-of-fit: 'Poissonisation' (on-going)
- Model selection: derive a proper (BIC?) criterion accounting for the discretization step
- Understand the inferred latent states in terms of animal behavior, biogeography, species, ...

Discussion

What we did not do.

- Goodness-of-fit: 'Poissonisation' (on-going)
- Model selection: derive a proper (BIC?) criterion accounting for the discretization step
- Understand the inferred latent states in terms of animal behavior, biogeography, species, ...

In parallel. With C. Dion-Blanc, D. Hawat and E. Lebarbier

- Efficient change-point detection ('segmentation') in (marked) Poisson & Hawkes processes.
  $\rightarrow$ Dynamic programming applies [DBHLR24].
- Segmentation-classification of Poisson processes

References

E.A. Allman, C. Matias, and J.A. Rhodes. Identifiability of parameters in latent structure models with many observed variables. *The Annals of Statistics*, pages 3099–3132, 2009.

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

O. Cappé, E. Moulines, and T. Rydén. *Inference in Hidden Markov Models*. Springer, 2005.

C. Dion-Blanc, D. Hawat, E Lebarbier, and S Robin. Multiple change-point detection for Poisson processes. Technical Report 2302.09103, arXiv, 2024.

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.

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

A. G Hawkes and D. Oakes. A cluster process representation of a self-exciting process. *Journal of applied probability*, 11(3):493–503, 1974.

M. Kirchner. Hawkes and INAR(∞) processes. *Stochastic Processes and their Applications*, 126(8):2494–2525, 2016.

M. Kirchner. An estimation procedure for the hawkes process. *Quantitative Finance*, 17(4):571–595, 2017.

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

H Teicher. Identifiability of mixtures. *The Annals of Mathematical Statistics*, 32(1):244–248, 1961.

Discrete HMM

Conversion formulas from continuous to discrete Hawkes

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

Discrete HMM

Conversion formulas from continuous to discrete Hawkes

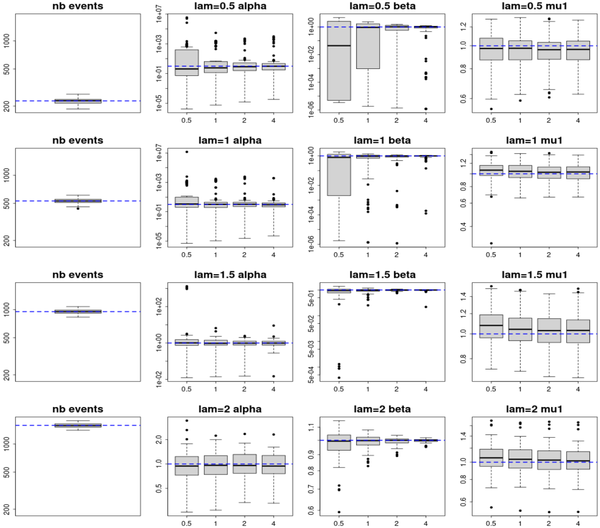
$$\alpha = \frac{a(e^{b\Delta} - 1)}{b}, \qquad \beta = e^{-b\Delta}$$

3-step initialization

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

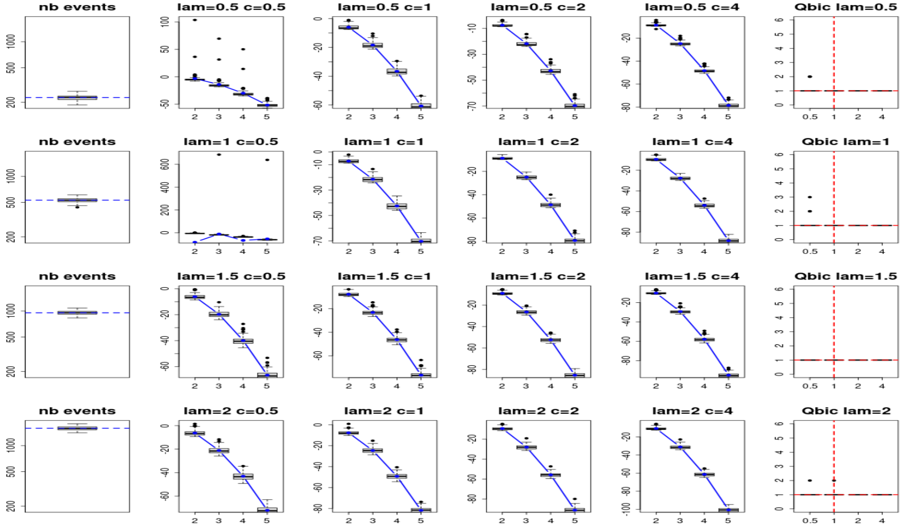
Simulation results ($Q^* = 1$, $N = cn$, $m^0 = 400$)

Parameter estimates. Distribution of $\widehat{\theta} - \theta^*$ ($Q = Q^*$)



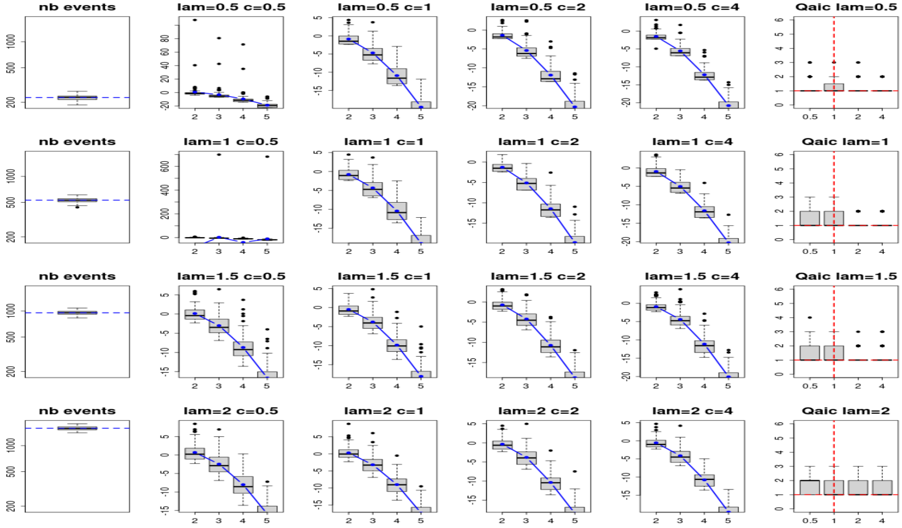
Simulation results ($Q^* = 1$, $N = cn$, $m^0 = 400$)

Model selection: BIC. Distribution of $BIC_Q - BIC_1$



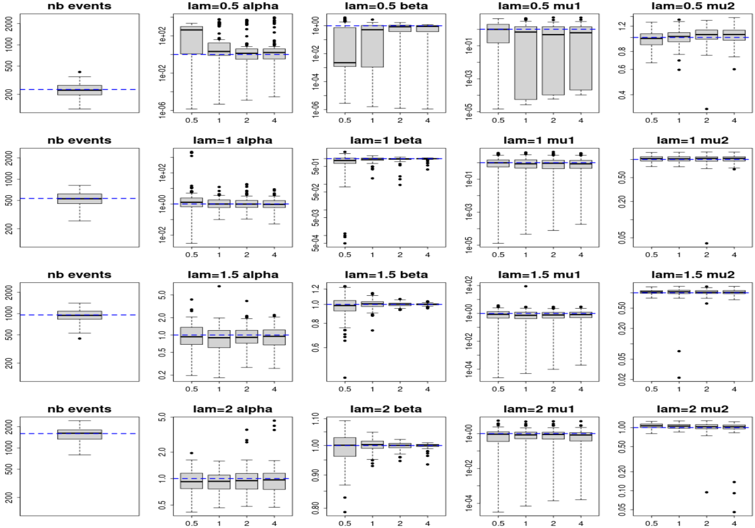
Simulation results ($Q^* = 1$, $N = cn$, $m^0 = 400$)

Model selection: AIC. Distribution of $AIC_Q - AIC_1$



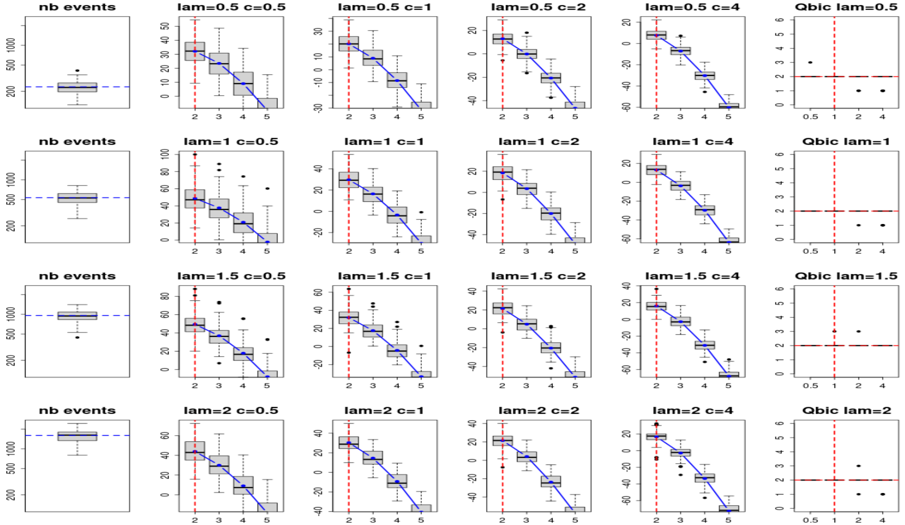
Simulation results ($Q^* = 2,\ N = cn,\ m^0 = [10,\ 800]$)

Parameter estimates. Distribution of $\widehat{\theta} - \theta^*$ ($Q = Q^*$)



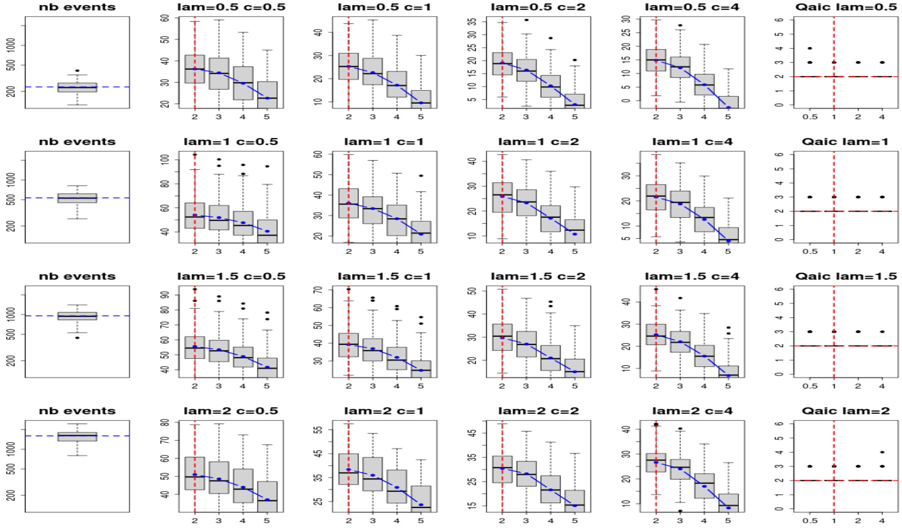
Simulation results ($Q^* = 2$, $N = cn$, $m^0 = [10,\ 800]$)

Model selection: BIC. Distribution of $BIC_Q - BIC_1$



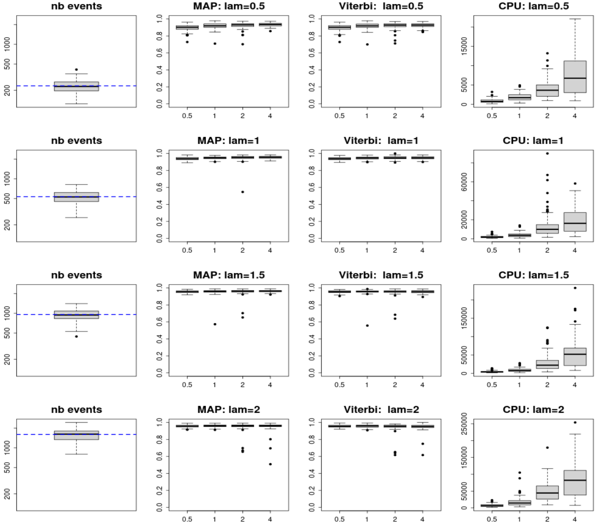
Simulation results ($Q^* = 2,\ N = cn,\ m^0 = [10,\ 800]$)

Model selection: AIC. Distribution of $AIC_Q - AIC_1$



Simulation results ($Q^* = 2$, $N = cn$, $m^0 = [10, 800]$)

Classification. MAP / Viterbi (+ comput. time)

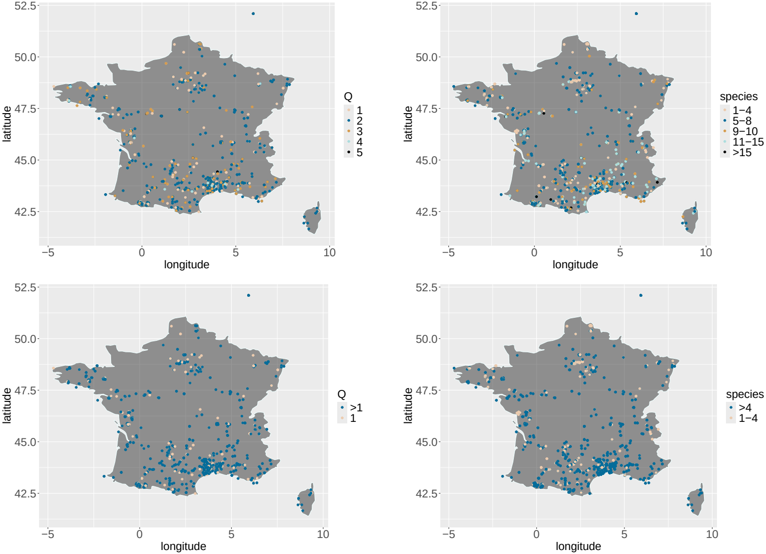


Model comparison for bat cries sequences

Poisson vs Hawkes / Homogeneous vs HMM. Best model based on BIC

	Poisson	Hawkes	Total
Homogeneous	132	775	907
Hidden Markov	21	627	648
Total	153	1402	1555

States and locations



Non exponential kernel function h

Compact support. Suppose that h has no exponential form, but

$$t > L\Delta \qquad \Rightarrow \qquad h(t) = 0.$$

Homogeneous discrete Hawkes process.

$$\left(Y_k \mid Y_{1:(k-1)}\right) \sim \mathcal{P}\left(\mu + \sum_{\ell=1}^{k-1} \alpha_\ell Y_{k-\ell}\right) \qquad \text{with} \quad \alpha_\ell = \int_{(\ell-1)\Delta}^{\ell\Delta} h(t) \, \mathrm{d}t.$$

Markovian representation. Define $U_k = \sum_{\ell \geqslant 1} \alpha_\ell Y_{k-\ell}$, then

$$\left(Y_k \mid Y_{1:(k-1)}\right) = (Y_k \mid U_k) \sim \mathcal{P}\left(\mu + U_k\right)$$

and $((Y_k, U_k))_{k \geqslant 1}$ forms a Markov chain (of order L).