Motivation: Unlocking the Mysteries of the Brain

- The human brain is composed of $10^{11}$ neurons
- **Question**: How do neurons work together to perceive the world, make decisions, and perform other higher-level tasks?
- We will primarily focus on **spike train data**

Sources: Allen Institute for Brain Science (left), Paul De Konnick lab (right)
Spike Train Data
Spike Train Data
Spike Train Data
Spike Train Data

- Spike Train: times at which a neuron “spikes” (transmits a signal)
Functional Connectivity Among Neurons
Functional Connectivity Among Neurons
Functional Connectivity Among Neurons
Functional Connectivity Among Neurons
Functional Connectivity Among Neurons
Functional Connectivity Among Neurons
Learning Functional Connectivity Networks
Learning Functional Connectivity Networks
Learning Functional Connectivity Networks
Learning Functional Connectivity Networks
Learning Functional Connectivity Networks
Challenges in Estimating Functional Connectivity

- May observe thousands of neurons
- Limited theoretical justification
- Short duration of stationary period
Challenges in Estimating Functional Connectivity

- May observe thousands of neurons
- Limited theoretical justification
- Short duration of stationary period
Challenges in Estimating Functional Connectivity

- May observe thousands of neurons
- Limited theoretical justification
- Short duration of stationary period
Challenges in Estimating Functional Connectivity

- May observe thousands of neurons
- Limited theoretical justification
- Short duration of stationary period
Challenges in Estimating Functional Connectivity

- May observe thousands of neurons
- Limited theoretical justification
- Short duration of stationary period
Challenges in Estimating Functional Connectivity

- May observe thousands of neurons
- Limited theoretical justification
- Short duration of stationary period
Hawkes Process

- Introduced by Hawkes (1971)

Spectra of some self-exciting and mutually exciting point processes

By ALAN G. HAWKES

- First applied to spike train data by Brillinger et al.

Identification of Synaptic Interactions*

David R. Brillinger
Department of Statistics, University of California, Berkeley, California

Hugh L. Bryant, Jr., and José P. Segundo
Department of Anatomy and the Brain Research Institute, University of California, Los Angeles, California USA

Received: December 2, 1975
A Linear Hawkes Process

\[ \lambda_j(t) = \mu_j + \sum_{k=1}^{p} \sum_{i: t_{k,i} \leq t} \omega_{j,k}(t-t_{k,i}) \]

- \( \lambda_j(t) dt \equiv P(\text{d}N_j(t) = 1 \mid \mathcal{H}_t) \): intensity process
- \( \text{d}N_j(t) \equiv N_j([t, t + \text{d}t]) \): point process
- \( \mu_j \in \mathbb{R}^+ \): spontaneous rate
- \( \omega_{j,k}(\cdot) : \mathbb{R}^+ \rightarrow \mathbb{R} \): transfer function from k to j
- \( t_{k,i} \in \mathbb{R}^+ \): time when the kth neuron has the ith spike
A Linear Hawkes Process

\[ \lambda_j(t) = \mu_j + \sum_{k=1}^{p} \sum_{i: t_{k,i} \leq t} \omega_{j,k}(t - t_{k,i}) \]

- \( \lambda_j(t) \mathrm{d}t \equiv \mathbb{P}(\mathrm{d}N_j(t) = 1 \mid \mathcal{H}_t) \): intensity process
- \( \mathrm{d}N_j(t) \equiv N_j([t, t + \mathrm{d}t]) \): point process
- \( \mu_j \in \mathbb{R}^+ \): spontaneous rate
- \( \omega_{j,k}(\cdot) : \mathbb{R}^+ \mapsto \mathbb{R} \): transfer function from k to j
- \( t_{k,i} \in \mathbb{R}^+ \): time when the kth neuron has the ith spike

Functional connectivity: there’s an edge from k to j if \( \omega_{j,k}(\cdot) \neq 0 \)
A Simple Hawkes Process

Intensity of Train 1

Time

0 5 10 15 20

0 0.5 1 1.5 2

Train 1

1

2

Intensity of Train 1
A Simple Hawkes Process
A Simple Hawkes Process

Intensity of Train 1
A Simple Hawkes Process
A Simple Hawkes Process

Intensity of Train 1
A Simple Hawkes Process

Intensity of Train 1

Time

Intensity of Train 1

Train 1

1

2
A Simple Hawkes Process
A Simple Hawkes Process
A Simple Hawkes Process
A Simple Hawkes Process

Intensity of Train 2

Time

IntENSITY of TRAIN 2

0 0.5 1 1.5 2

Train 2

Train 1
A Simple Hawkes Process

Intensity of Train 2

Time

Intensity of Train 2
A Simple Hawkes Process
Penalized Regression for Hawkes Processes

Joint work with Shizhe Chen, Eric Shea-Brown, and Daniela Witten

- The multivariate Hawkes process in high dimensions: Beyond mutual excitation (
  \(arXiv:1707.04928\)); invited revision to \textit{Annals of Statistics}
- Nearly assumptionless screening for the mutually-exciting multivariate Hawkes process
  (2017) \textit{Electronic Journal of Statistics}
Penalized Regression for Hawkes Processes

Regress each spike train onto others

- Neighbourhood selection
- Estimate incoming edges

Joint work with Shizhe Chen, Eric Shea-Brown, and Daniela Witten

- The multivariate Hawkes process in high dimensions: Beyond mutual excitation (arXiv:1707.04928); invited revision to *Annals of Statistics*
Parameter Estimation via Penalized Regression

- **Model**

\[
\lambda_j(t) = \mu_j + \sum_{k=1}^{p} \sum_{t_k, i \leq t} \omega_{j,k}(t - t_{k,i})
\]
Parameter Estimation via Penalized Regression

- Model
  \[ \lambda_j(t) = \mu_j + \sum_{k=1}^{p} \sum_{t_{k,i} \leq t} \omega_{j,k}(t - t_{k,i}) \]

- Finite-dimensional basis expansion
  \[ \omega_{j,k}(t) \approx \psi(t) \cdot \beta_{j,k} \]
Parameter Estimation via Penalized Regression

- **Model**
  \[
  \lambda_j(t) = \mu_j + \sum_{k=1}^{p} \sum_{k_i \leq t} \omega_{j,k}(t - t_{k,i})
  \]

- **Finite-dimensional basis expansion**
  \[
  \omega_{j,k}(t) \approx \psi(t) \cdot \beta_{j,k}
  \]

- **Squared error loss**
  \[
  \frac{1}{2T} \int_0^T \left\{ \lambda_j(t) \, dt - dN_j(t) \right\}^2
  \]
Parameter Estimation via Penalized Regression

- **Model**

\[ \lambda_j(t) = \mu_j + \sum_{k=1}^{p} \sum_{t_{k,1} \leq t}^{p} \omega_{j,k}(t - t_{k,i}) \]

- **Finite-dimensional basis expansion**

\[ \omega_{j,k}(t) \approx \psi(t) \cdot \beta_{j,k} \]

- **Squared error loss**

\[ \frac{1}{2T} \int_0^T \left\{ \lambda_j(t) \, dt - dN_j(t) \right\}^2 \]

- **Regression**

\[
\begin{aligned}
\text{minimize} & \quad \mu_j, \beta_{j,k} \\
& \quad - \frac{1}{2T} \int_0^T \left\{ \left[ \mu_j + \sum_{k=1}^{p} (\psi(t) \cdot dN_k)(t) \cdot \beta_{j,k} \right]^2 dt - 2 \left[ \mu_j + \sum_{k=1}^{p} (\psi(t) \cdot dN_k)(t) \cdot \beta_{j,k} \right] dN_j(t) \right\} \\
& \quad + \eta_j \sum_{k=1}^{p} \left\{ \frac{1}{T} \int_0^T [(\psi(t) \cdot dN_k)(t) \cdot \beta_{j,k}]^2 dt \right\}^{1/2}
\end{aligned}
\]

(standardized) group lasso penalty
Parameter Estimation via Penalized Regression

- **Model**
  \[
  \lambda_j(t) = \mu_j + \sum_{k=1}^{p} \sum_{t_{k,i} \leq t} \omega_{j,k}(t - t_{k,i})
  \]

- **Finite-dimensional basis expansion**
  \[
  \omega_{j,k}(t) \approx \psi(t) \cdot \beta_{j,k}
  \]

- **Squared error loss**
  \[
  \frac{1}{2T} \int_0^T \left\{ \lambda_j(t) \, dt - dN_j(t) \right\}^2
  \]

- **Regression**
  \[
  \begin{aligned}
  &\text{minimize} \quad \frac{1}{2n} X\beta^T X\beta \\
  &\quad \frac{1}{n} y^T x\beta \\
  &\quad \frac{1}{2T} \int_0^T \left[ \mu_j + \sum_{k=1}^{p} (\psi(t) \ast dN_k)(t) \cdot \beta_{j,k} \right]^2 \, dt - \frac{1}{T} \int_0^T \left[ \mu_j + \sum_{k=1}^{p} (\psi(t) \ast dN_k)(t) \cdot \beta_{j,k} \right] \, dN_j(t) \\
  &\quad + \eta_j \sum_{k=1}^{p} \left\{ \frac{1}{T} \int_0^T [(\psi(t) \ast dN_k)(t) \cdot \beta_{j,k}]^2 \, dt \right\}^{1/2} \\
  &\quad \|X_k \beta_{j,k}\|_2
  \end{aligned}
  \]
Parameter Estimation via Penalized Regression

- **Model**
  \[ \lambda_j(t) = \mu_j + \sum_{k=1}^{p} \sum_{t_k,i \leq t} \omega_{j,k}(t - t_{k,i}) \]

- **Finite-dimensional basis expansion**
  \[ \omega_{j,k}(t) \approx \psi(t) \cdot \beta_{j,k} \]

- **Squared error loss**
  \[ \frac{1}{2T} \int_0^T \left\{ \lambda_j(t) \, dt - dN_j(t) \right\}^2 \]

- **Regression**
  \[ \minimize_{\mu_j, \beta_{j,k}} \frac{1}{2n} X \beta^T X \beta \]
  \[ + \eta_j \sum_{k=1}^{p} \left\{ \frac{1}{T} \int_0^T \left[ (\psi(t) \ast dN_k(t)) \cdot \beta_{j,k} \right]^2 \, dt \right\}^{1/2} \]
  \[ \| X_k \beta_{j,k} \|_2 \]

- **Estimation via block coordinate descent**
Properties of Penalized Estimation Procedures

- Existing theory relies on the cluster process representation
  - Assumes non-negative transfer functions \((\omega_{j,k} \geq 0)\)
  - Only holds for linear Hawkes processes

A CLUSTER PROCESS REPRESENTATION OF A SELF-EXCITING PROCESS

ALAN G. HAWKES*, University of Durham
DAVID OAKES**, Imperial College London
Gap in Existing Theory: Neurons Excite and Inhibit

Neocortical excitation/inhibition balance in information processing and social dysfunction

A New Concentration Inequality for Hawkes Process

- New theoretical framework that allows inhibition
  - Use the thinning process representation
A New Concentration Inequality for Hawkes Process

- New theoretical framework that allows inhibition
  - Use the thinning process representation
- For any $j, k$, consider

$$\bar{y}_{j,k} \equiv \frac{1}{T} \int_0^T \int_0^T f(t - t') \, dN_k(t') \, dN_j(t)$$

- Here $f(\cdot)$ can be any continuous and integrable function
- $\bar{y}_{j,k}$ covers a wide range of second-order statistics of the Hawkes process, including the cross-covariance
A New Concentration Inequality for Hawkes Process

- New theoretical framework that allows inhibition
  - Use the thinning process representation
- For any $j, k$, consider
  \[
  \bar{y}_{j,k} = \frac{1}{T} \int_0^T \int_0^T f(t - t') \, dN_k(t') \, dN_j(t)
  \]
  - Here $f(\cdot)$ can be any continuous and integrable function
  - $\bar{y}_{j,k}$ covers a wide range of second-order statistics of the Hawkes process, including the cross-covariance
- We have
  \[
  \mathbb{P} \left( |\bar{y}_{j,k} - \mathbb{E}\bar{y}_{j,k}| \geq c_1 T^{-1/5} \right) \leq c_2 T \exp \left( -c_3 T^{1/5} \right)
  \]
An Application of the New Concentration Inequality

• Neighbourhood selection recovers the graph with high probability

\[ \mathbb{P}(\hat{\mathcal{E}} = \mathcal{E}) \geq 1 - o(1) \]

where \( \mathcal{E} \) and \( \hat{\mathcal{E}} \) are true and estimated edges

• Key assumptions
  - \( p^2 T \exp\left(-c_3 T^{1/5}\right) = o(1) \), i.e., we can handle \( p \gg T \)
  - Stationarity
  - Other regularity conditions for lasso-type estimators
A Computational Shortcut

Penalized regression becomes computationally (and statistically) challenging with **many neurons**

- Can we reduce the number of potential edges?
A Computational Shortcut

- Let $V_{j,k}$ be the **cross-covariance** between the $j$th & $k$th neurons
- Consider the graph defined by **marginal screening**

\[
\tilde{\mathcal{E}} = \left\{ (j, k) : \| \hat{V}_{j,k} \|_2 > \eta \right\}
\]

- This ‘correlation graph’ is often used by neuroscientists
- It is computationally (and statistically) efficient
Cross-Correlation Graph

Edge: (1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (1, 7), (1, 8), (1, 9)

Cross Cov.: 1.29
Cross-Correlation Graph

<table>
<thead>
<tr>
<th>Edge</th>
<th>Cross Cov.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 1)</td>
<td>1.29</td>
</tr>
<tr>
<td>(1, 2)</td>
<td>4.35</td>
</tr>
<tr>
<td>(1, 3)</td>
<td></td>
</tr>
<tr>
<td>(1, 4)</td>
<td></td>
</tr>
<tr>
<td>(1, 5)</td>
<td></td>
</tr>
<tr>
<td>(1, 6)</td>
<td></td>
</tr>
<tr>
<td>(1, 7)</td>
<td></td>
</tr>
<tr>
<td>(1, 8)</td>
<td></td>
</tr>
<tr>
<td>(1, 9)</td>
<td></td>
</tr>
</tbody>
</table>
Cross-Correlation Graph

<table>
<thead>
<tr>
<th>Edge</th>
<th>Cross Cov.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 1)</td>
<td>1.29</td>
</tr>
<tr>
<td>(1, 2)</td>
<td>4.35</td>
</tr>
<tr>
<td>(1, 3)</td>
<td>2.47</td>
</tr>
<tr>
<td>(1, 4)</td>
<td></td>
</tr>
<tr>
<td>(1, 5)</td>
<td></td>
</tr>
<tr>
<td>(1, 6)</td>
<td></td>
</tr>
<tr>
<td>(1, 7)</td>
<td></td>
</tr>
<tr>
<td>(1, 8)</td>
<td></td>
</tr>
<tr>
<td>(1, 9)</td>
<td></td>
</tr>
</tbody>
</table>
Cross-Correlation Graph

Edge | Cross Cov.
-----|-----------
(1, 1) | 1.29
(1, 2) | 4.35
(1, 3) | 2.47
(1, 4) | 3.21
(1, 5) |
(1, 6) |
(1, 7) |
(1, 8) |
(1, 9) |
Cross-Correlation Graph

<table>
<thead>
<tr>
<th>Edge</th>
<th>Cross Cov.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 1)</td>
<td>1.29</td>
</tr>
<tr>
<td>(1, 2)</td>
<td>4.35</td>
</tr>
<tr>
<td>(1, 3)</td>
<td>2.47</td>
</tr>
<tr>
<td>(1, 4)</td>
<td>3.21</td>
</tr>
<tr>
<td>(1, 5)</td>
<td>1.70</td>
</tr>
<tr>
<td>(1, 6)</td>
<td>1.91</td>
</tr>
<tr>
<td>(1, 7)</td>
<td>1.47</td>
</tr>
<tr>
<td>(1, 8)</td>
<td>0.93</td>
</tr>
<tr>
<td>(1, 9)</td>
<td>2.23</td>
</tr>
</tbody>
</table>
Cross-Correlation Graph

The true graph

The screened graph

<table>
<thead>
<tr>
<th>Edge</th>
<th>Cross Cov.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 1)</td>
<td>1.29</td>
</tr>
<tr>
<td>(1, 2)</td>
<td>4.35</td>
</tr>
<tr>
<td>(1, 3)</td>
<td>2.47</td>
</tr>
<tr>
<td>(1, 4)</td>
<td>2.41</td>
</tr>
<tr>
<td>(1, 5)</td>
<td>1.70</td>
</tr>
<tr>
<td>(1, 6)</td>
<td>1.91</td>
</tr>
<tr>
<td>(1, 7)</td>
<td>1.47</td>
</tr>
<tr>
<td>(1, 8)</td>
<td>0.93</td>
</tr>
<tr>
<td>(1, 9)</td>
<td>2.23</td>
</tr>
</tbody>
</table>
Cross-Correlation Graph

The true graph

The screened graph

Edge | Cross Cov.
--- | ------
(1, 1) | 1.29
(1, 2) | 4.35
(1, 3) | 2.47
(1, 4) | 2.41
(1, 5) | 1.70
(1, 6) | 1.91
(1, 7) | 1.47
(1, 8) | 0.93
(1, 9) | 2.23
Properties of Screening

Recall

\[ \tilde{\mathcal{E}} = \left\{ (j, k) : \|\tilde{V}_{j,k}\|_2 > \eta \right\} \]

\[ \mathcal{E} = \left\{ (j, k) : \omega_{j,k} \neq 0 \right\} \]

Q: How does \( \tilde{\mathcal{E}} \) relate to the functional connectivity network, \( \mathcal{E} \)?
Properties of Screening

- If the process is mutually exciting,
Properties of Screening

- If the process is mutually exciting,
  - $\mathcal{E} \subseteq \tilde{\mathcal{E}}$
  - $\text{card}(\tilde{\mathcal{E}}) \approx \text{card}(\mathcal{E})$
Properties of Screening

- If the process is mutually exciting,
  - $\mathcal{E} \subseteq \tilde{\mathcal{E}}$
  - $\text{card}(\tilde{\mathcal{E}}) \approx \text{card}(\mathcal{E})$

- These results can be shown using our new theoretical framework

- Unlike existing approaches, they do not require extra assumption
Properties of Screening

- If the process is mutually exciting,
  - $\mathcal{E} \subseteq \tilde{\mathcal{E}}$
  - $\text{card}(\tilde{\mathcal{E}}) \approx \text{card}(\mathcal{E})$

- These results can be shown using our new theoretical framework
- Unlike existing approaches, they do not require extra assumption
Properties of Screening

● What if there are negative edges?
Properties of Screening

- What if there are negative edges?
Properties of Screening

- What if there are negative edges?
  - Even with negative edges, screening detects connected components of the graph.
Properties of Screening

- What if there are negative edges?
  - Even with negative edges, screening detects connected components of the graph
Properties of Screening

- What if there are negative edges?
  - Even with negative edges, screening detects connected components of the graph
Properties of Screening

- What if there are negative edges?
  - Even with negative edges, screening detects connected components of the graph
Neurons in Cat Visual Cortex
Addressing Non-Stationarity: Piecewise Stationary VARs

Motivation: Analyzing EEG Data
Addressing Non-Stationarity: Piecewise Stationary VARs

Motivation: Analyzing EEG Data
Addressing Non-Stationarity: Piecewise Stationary VARs

Motivation: Analyzing EEG Data
Addressing Non-Stationarity: Piecewise Stationary VARs

Motivation: Analyzing EEG Data

Brain connectivities expected to change after seizure

- Goal: To locate the seizure and estimate before/after networks
Addressing Non-Stationarity: Piecewise Stationary VARs

**Our proposal:** A 3-step procedure based on **total variation penalty**
Addressing Non-Stationarity: Piecewise Stationary VARs

Our proposal: A 3-step procedure based on total variation penalty
Addressing Non-Stationarity: Piecewise Stationary VARs

Our proposal: A 3-step procedure based on total variation penalty
Addressing Non-Stationarity: Piecewise Stationary VARs

Our proposal: A 3-step procedure based on total variation penalty

Joint work with Abolfazl Safikhani (Columbia Univ)

Acknowledgment

● Allen Institute for Brain Sciences

● Funding
  ○ NIH: NIGMS & NHLBI
  ○ NSF: DMS & DMS/NIGMS

● References
Acknowledgment

● Allen Institute for Brain Sciences

● Funding
  ○ NIH: NIGMS & NHLBI
  ○ NSF: DMS & DMS/NIGMS

● References

Thank You!
4th Annual Summer Institute in Statistics for Big Data (SISBID)

The Summer Institute for Big Data (SISBID) is designed to introduce biologists, quantitative scientists, and statisticians to modern statistical techniques for the analysis of biological big data.

Key Dates

- Modules: July 11-28
- Registration now open
Appendix I
Theory for Hawkes Process with Inhibitions
Recap: One-Dimensional Linear Hawkes Process

\[ \lambda(t) = \mu + \sum_{i : t_i \leq t} \omega(t - t_i) \]

- \( \lambda(t) dt \equiv \mathbb{P}(dN(t) = 1 \mid \mathcal{H}_t) \): intensity process
- \( dN(t) \equiv N([t, t + dt]) \): point process
- \( \mu \in \mathbb{R}^+ \): spontaneous rate
- \( \omega(\cdot) : \mathbb{R}^+ \to \mathbb{R} \): transfer function
- \( t_i \in \mathbb{R}^+ \): time of the ith spike
Hawkes Process is Temporally Dependent by Definition

- Recall the form of intensity

\[ \lambda(t) = \mu + \sum_{i : t_i \leq t} \omega(t - t_i) \]

- Observation of \( dN \) on \( (t + \Delta, t'] \) might still depend on \( \{dN(s) : s \in [0, t]\} \)

- Question: how much "independent" information is there in \( \{dN(s) : s \in [0, T]\} \)?
Hawkes Process is Temporally Dependent by Definition

- Recall the form of intensity

\[ \lambda(t) = \mu + \sum_{i: t_i \leq t} \omega(t - t_i) \]

- Observation of \( dN \) on \((t + \Delta, t']\) might still depend on \( \{dN(s) : s \in [0, t]\} \)

- Question: how much “independent” information is there in \( \{dN(s) : s \in [0, T]\} \)?

Key to understanding the Hawkes process: quantifying the temporal dependence
Temporal Dependence of a Hawkes Process

- $\tau$-dependence coefficient for the Hawkes process $dN$

$$\tau_N(\Delta) \equiv \sup_{u>0} \tau(dN(u + \Delta), \mathcal{H}_u), \quad \Delta > 0$$

- Here $\tau(dN(u + \Delta), \mathcal{H}_u)$ is a dependence measure between $dN(u + \Delta)$ and the history $\mathcal{H}_u$

- Coupling theorem by Dedecker and Prieur (2004)

$$\tau(dN(u + \Delta), \mathcal{H}_u) \leq \mathbb{E}[dN(u + \Delta) - d\tilde{N}(u + \Delta)]$$

s.t. $dN(u + \Delta) \overset{d}{=} d\tilde{N}(u + \Delta)$ and $d\tilde{N}(u + \Delta) \perp \mathcal{H}_u$
Temporal Dependence of a Hawkes Process

- $\tau$-dependence coefficient for the Hawkes process $dN$

$$\tau_N(\Delta) \equiv \sup_{u > 0} \tau(dN(u + \Delta), \mathcal{H}_u), \quad \Delta > 0$$

- Here $\tau(dN(u + \Delta), \mathcal{H}_u)$ is a dependence measure between $dN(u + \Delta)$ and the history $\mathcal{H}_u$
- Coupling theorem by Dedecker and Prieur (2004)

$$\tau(dN(u + \Delta), \mathcal{H}_u) \leq \mathbb{E}[dN(u + \Delta) - d\tilde{N}(u + \Delta)]$$

s.t. $dN(u + \Delta) \overset{d}{=} d\tilde{N}(u + \Delta)$ and $d\tilde{N}(u + \Delta) \perp \mathcal{H}_u$

How to construct $dN$ and $d\tilde{N}$?
Existing Theory Assumes Non-Negative Transfer Functions

• Construct $dN$ and $d\tilde{N}$ using the cluster process representation (Hawkes and Oakes 1974)

A CLUSTER PROCESS REPRESENTATION OF A SELF-EXCITING PROCESS

ALAN G. HAWKES*, University of Durham
DAVID OAKES**, Imperial College London

• Only applies to mutually-exciting processes ($\omega \geq 0$)
Existing Theory Assumes Non-Negative Transfer Functions

- Construct $dN$ and $d\tilde{N}$ using the cluster process representation (Hawkes and Oakes 1974)

  **A CLUSTER PROCESS REPRESENTATION OF A SELF-EXCITING PROCESS**

  ALAN G. HAWKES*, University of Durham  
  DAVID OAKES**, Imperial College London

- Only applies to mutually-exciting processes ($\omega \geq 0$)

How to construct $dN$ and $d\tilde{N}$ with inhibition?
Represent Processes by Thinning a Poisson Process

- Any point processes $dN(t)$ of intensity $\lambda(t)$ can be represented by homogeneous Poisson processes via thinning (Kerstain 1964, Brémaud and Massoulié 1996, Daley and Vere-Jones 2003)

$$dN(t) \equiv M([t, t + dt] \times [0, \lambda(t)])$$

"Full" Process

$M$: 2D homogeneous Poisson process of intensity 1 (i.e., lives on $\mathbb{R}^2$)
Represent Processes by Thinning a Poisson Process

- Any point processes $dN(t)$ of intensity $\lambda(t)$ can be represented by homogeneous Poisson processes via thinning (Kerstain 1964, Brémaud and Massoulié 1996, Daley and Vere-Jones 2003)

$$dN(t) \equiv M([t, t + dt) \times [0, \lambda(t)])$$

"Full" Process & "Thinned" Process

$M$: 2D homogeneous Poisson process of intensity 1 (i.e., lives on $\mathbb{R}^2$)

$\lambda(t)$: $\mathcal{M}_t$-predictable process

$\mathcal{M}_t$: $\sigma$-field of $M$ on $(-\infty, t] \times \mathbb{R}$
Represent Processes by Thinning a Poisson Process

- Any point processes $dN(t)$ of intensity $\lambda(t)$ can be represented by homogeneous Poisson processes via thinning (Kerstain 1964, Brémaud and Massoulié 1996, Daley and Vere-Jones 2003)

\[ dN(t) \equiv M\left( [t, t + dt] \times [0, \lambda(t)] \right) \]

“Full” Process & “Thinned” Process

$M$: 2D homogeneous Poisson process of intensity 1 (i.e., lives on $\mathbb{R}^2$)

$\lambda(t)$: $\mathcal{M}_t$-predictable process

$\mathcal{M}_t$: $\sigma$-field of $M$ on $(-\infty, t] \times \mathbb{R}$

This representation applies to any **stationary** Hawkes process!
Bounding the Temporal Dependence Using the Thinning Process Representation

- Represent both $dN$ and $d\tilde{N}$ using the thinning processes

\[
\begin{align*}
\tilde{M} \perp M & \quad \text{on } [0, u] \times \mathbb{R} \\
\tilde{M} \equiv M & \quad \text{on } (u, \infty) \times \mathbb{R}
\end{align*}
\]

\[
\begin{align*}
dN(t) & \equiv M([t, t+dt) \times [0, \lambda(t)]) \\
d\tilde{N}(t) & \equiv \tilde{M}([t, t+dt) \times [0, \tilde{\lambda}(t)])
\end{align*}
\]
Bounding the Temporal Dependence Using the Thinning Process Representation

- Represent both $dN$ and $\tilde{dN}$ using the thinning processes

\[
\begin{align*}
\tilde{M} &\perp M \quad \text{on } [0, u] \times \mathbb{R} \\
\tilde{M} &\equiv M \quad \text{on } (u, \infty) \times \mathbb{R}
\end{align*}
\]

\[
\begin{align*}
dN(t) &\equiv M([t, t+dt) \times [0, \lambda(t)]) \\
d\tilde{N}(t) &\equiv \tilde{M}([t, t+dt) \times [0, \tilde{\lambda}(t)])
\end{align*}
\]

- The construction guarantees that

\[dN(u + \Delta) \overset{d}{=} d\tilde{N}(u + \Delta) \quad \text{and} \]

\[d\tilde{N}(u + \Delta) \perp \mathcal{H}_u \quad \text{(since } d\tilde{N}(u + \Delta) \perp \mathcal{M}_u \& \mathcal{H}_u \subset \mathcal{M}_u)\]

- We can show that

\[\tau(\Delta) \leq \sup_{u>0} \mathbb{E}[dN(u + \Delta) - d\tilde{N}(u + \Delta)] \leq a_1 \exp(-a_2 \Delta)\]
Appendix II
Iterative Construction of
Thinning Process Representation
for Hawkes Process
Recap: One-Dimensional Linear Hawkes Process

\[ \lambda(t) = \mu + \sum_{i: t_i \leq t} \omega(t - t_i) \equiv \mu + (\omega * dN)(t) \]

- \( \lambda(t) dt \equiv P(dN(t) = 1 | \mathcal{H}_t) \): intensity process
- \( dN(t) \equiv N([t, t + dt]) \): point process
- \( \mu \in \mathbb{R}^+ \): spontaneous rate
- \( \omega(\cdot) : \mathbb{R}^+ \mapsto \mathbb{R} \): transfer function
- \( t_i \in \mathbb{R}^+ \): time of the ith spike
- \( (\omega * dN)(t) \equiv \int_0^\infty \omega(\Delta) dN(t - \Delta) \): convolution of \( \omega \) and \( dN \)
Thinning Process Representation of the Hawkes Process

- Construct the thinning process representation of the Hawkes process $dN: \lambda(t) = \mu + (\omega \ast dN)(t)$
- Brémaud and Massoulié 1996: Hawkes process as the limit of a sequence of thinning processes
  \[ \lambda^{(n)}(t) = \mu + (\omega \ast dN^{(n-1)})(t) \quad (\lambda^{(1)}(t) = \mu) \]
  \[ dN^{(n)}(t) = M([t, t + dt] \times [0, \lambda^{(n)}(t)]) \]
Thinning Process Representation of the Hawkes Process

- Construct the thinning process representation of the Hawkes process \( dN: \lambda(t) = \mu + (\omega \ast dN)(t) \)
- Brémaud and Massoulié 1996: Hawkes process as the limit of a sequence of thinning processes
  \[
  \lambda^{(n)}(t) = \mu + (\omega \ast dN^{(n-1)})(t) \quad (\lambda^{(1)}(t) = \mu)
  \]
  \[
  dN^{(n)}(t) = M([t, t + dt) \times [0, \lambda^{(n)}(t)])
  \]

\[ n = 2 \]

- Intensity in the previous iteration
- Intensity in the current iteration
- Removed spikes
- New spikes
Thinning Process Representation of the Hawkes Process

- Construct the thinning process representation of the Hawkes process $dN$: $\lambda(t) = \mu + (\omega \ast dN)(t)$
- Brémaud and Massoulié 1996: Hawkes process as the limit of a sequence of thinning processes

\[
\lambda^{(n)}(t) = \mu + (\omega \ast dN^{(n-1)})(t) \quad (\lambda^{(1)}(t) = \mu)
\]

\[
dN^{(n)}(t) = M([t, t + dt) \times [0, \lambda^{(n)}(t)])
\]

$n = 3$

- Intensity in the previous iteration
- Intensity in the current iteration
- Removed spikes
- New spikes
Thinning Process Representation of the Hawkes Process

- Construct the thinning process representation of the Hawkes process $dN$: $\lambda(t) = \mu + (\omega \ast dN)(t)$
- Brémaud and Massoulié 1996: Hawkes process as the limit of a sequence of thinning processes
  $$\lambda^{(n)}(t) = \mu + (\omega \ast dN^{(n-1)})(t) \quad (\lambda^{(1)}(t) = \mu)$$
  $$dN^{(n)}(t) = M\left([t, t + dt] \times [0, \lambda^{(n)}(t)]\right)$$

![Diagram](image)
Thinning Process Representation of the Hawkes Process

- Construct the thinning process representation of the Hawkes process $dN$: $\lambda(t) = \mu + (\omega \ast dN)(t)$

- Brémaud and Massoulié 1996: Hawkes process as the limit of a sequence of thinning processes

\[
\lambda^{(n)}(t) = \mu + (\omega \ast dN^{(n-1)})(t) \quad (\lambda^{(1)}(t) = \mu)
\]

\[
dN^{(n)}(t) = M([t, t + dt) \times [0, \lambda^{(n)}(t)])
\]

$n = 5$

- Intensity in the previous iteration
- Intensity in the current iteration
- Removed spikes
- New spikes
Thinning Process Representation of the Hawkes Process

- Construct the thinning process representation of the Hawkes process $dN$: $\lambda(t) = \mu + (\omega \ast dN)(t)$
- Brémaud and Massoulié 1996: Hawkes process as the limit of a sequence of thinning processes
  
  $\lambda^{(n)}(t) = \mu + (\omega \ast dN^{(n-1)})(t)$ \hspace{1cm} (\lambda^{(1)}(t) = \mu)$
  
  $dN^{(n)}(t) = M([t, t + dt) \times [0, \lambda^{(n)}(t)])$

\[ n = 6 \]
Thinning Process Representation of the Hawkes Process

- Construct the thinning process representation of the Hawkes process $dN$: $\lambda(t) = \mu + (\omega \ast dN)(t)$
- Brémaud and Massoulié 1996: Hawkes process as the limit of a sequence of thinning processes

$$\lambda^{(n)}(t) = \mu + (\omega \ast dN^{(n-1)})(t) \quad (\lambda^{(1)}(t) = \mu)$$

$$dN^{(n)}(t) = M([t, t + dt] \times [0, \lambda^{(n)}(t)])$$

$n = 7$

- . . . . . . . Intensity in the previous iteration
- Red Intensity in the current iteration
- Removed spikes
- New spikes
Thinning Process Representation of the Hawkes Process

- Construct the thinning process representation of the Hawkes process \( dN: \lambda(t) = \mu + (\omega * dN)(t) \)
- Brémaud and Massoulié 1996: Hawkes process as the limit of a sequence of thinning processes
  \[
  \lambda^{(n)}(t) = \mu + (\omega * dN^{(n-1)})(t) \quad (\lambda^{(1)}(t) = \mu)
  \]
  \[
  dN^{(n)}(t) = M([t, t+dt] \times [0, \lambda^{(n)}(t)])
  \]

\( n = 8 \)

- Intensity in the previous iteration
- Intensity in the current iteration
- Removed spikes
- New spikes
Thinning Process Representation of the Hawkes Process

- Construct the thinning process representation of the Hawkes process $dN$: $\lambda(t) = \mu + (\omega * dN)(t)$
- Brémaud and Massoulié 1996: Hawkes process as the limit of a sequence of thinning processes

$$\lambda^{(n)}(t) = \mu + (\omega * dN^{(n-1)})(t) \quad (\lambda^{(1)}(t) = \mu)$$

$$dN^{(n)}(t) = M([t, t + dt) \times [0, \lambda^{(n)}(t)])$$
Thinning Process Representation of the Hawkes Process

- Construct the thinning process representation of the Hawkes process $dN$:
  \[ \lambda(t) = \mu + (\omega \ast dN)(t) \]

- Brémaud and Massoulié 1996: Hawkes process as the limit of a sequence of thinning processes
  \[
  \lambda^{(n)}(t) = \mu + (\omega \ast dN^{(n-1)})(t) \quad (\lambda^{(1)}(t) = \mu) \\
  dN^{(n)}(t) = M([t, t + dt) \times [0, \lambda^{(n)}(t)])
  \]

\[
\{\lambda^{(n)}\}_n \text{ converges if } \int_0^\infty |\omega(x)|\,dx < 1, \text{ with the limit } \\
\lambda^{(\infty)}(t) = \mu + (\omega \ast dN^{(\infty)})(t)
\]
Appendix III
Cluster Process Representation for
the Hawkes Process
Cluster Process Representation

- Proposed by Hawkes and Oakes (1974)
- Represent a Hawkes process as the summation of processes
- Consider a one-dimensional Hawkes process

\[ dN(t) = dN^{[0]}(t) + dN^{[1]}(t) + dN^{[2]}(t) + \ldots + dN^{[N^0(t)]}(t) \]

Hawkes process \( \lambda(t) = \mu + \sum_{t_i < t} \omega(t - t_i) \)
Cluster Process Representation

- Proposed by Hawkes and Oakes (1974)
- Represent a Hawkes process as the summation of processes
- Consider a one-dimensional Hawkes process

\[ dN(t) = dN^{0}(t) + dN^{1}(t) + dN^{2}(t) + \ldots + dN^{N(t)}(t) \]

Hawkes process

\[ \lambda(t) = \mu + \sum_{t_i < t} \omega(t - t_i) \]

Ancestral process

\[ \lambda^{0}(t) = \mu \]
Cluster Process Representation

- Proposed by Hawkes and Oakes (1974)
- Represent a Hawkes process as the summation of processes
- Consider a one-dimensional Hawkes process

\[ dN(t) = dN^{(0)}(t) + dN^{(1)}(t) + dN^{(2)}(t) + \ldots + dN^{(N^0(t))}(t) \]

Hawkes process \( \lambda(t) = \mu + \sum_{t_i < t} \omega(t - t_i) \)

Descendants \( \lambda^{(k)}(t) = \sum_{t_{k,i} < t} \omega(t - t_{k,i}) \)

Ancestral process \( \lambda^{(0)}(t) = \mu \)
Cluster Process Representation

- Proposed by Hawkes and Oakes (1974)
- Represent a Hawkes process as the summation of processes
- Consider a one-dimensional Hawkes process

\[ dN(t) = dN^{(0)}(t) + dN^{(1)}(t) + dN^{(2)}(t) + \ldots + dN^{(N^0(t))}(t) \]

Hawkes process
\[ \lambda(t) = \mu + \sum_{t_i < t} \omega(t - t_i) \]

Descendants
\[ \lambda^{(k)}(t) = \sum_{t_{k,i} < t} \omega(t - t_{k,i}) \]

Ancestral process
\[ \lambda^{(0)}(t) = \mu \]
Cluster Process Representation

- Proposed by Hawkes and Oakes (1974)
- Represent a Hawkes process as the summation of processes
- Consider a one-dimensional Hawkes process

\[ dN(t) = dN^{[0]}(t) + dN^{[1]}(t) + dN^{[2]}(t) + \ldots + dN^{[N^0(t)]}(t) \]

Hawkes process
\[ \lambda(t) = \mu + \sum_{t_i < t} \omega(t - t_i) \]

Descendants
\[ \lambda^{[k]}(t) = \sum_{t_{k,i} < t} \omega(t - t_{k,i}) \]

Ancestral process
\[ \lambda^{[0]}(t) = \mu \]
Cluster Process Representation

- Proposed by Hawkes and Oakes (1974)
- Represent a Hawkes process as the summation of processes
- Consider a one-dimensional Hawkes process

\[ dN(t) = dN^{\{0\}}(t) + dN^{\{1\}}(t) + dN^{\{2\}}(t) + \ldots + dN^{\{N^0(t)\}}(t) \]

Hawkes process \( \lambda(t) = \mu + \sum_{t_i < t} \omega(t - t_i) \)

Descendants \( \lambda^{\{k\}}(t) = \sum_{t_{k,i} < t} \omega(t - t_{k,i}) \)

Ancestral process \( \lambda^{\{0\}}(t) = \mu \)

Only holds for linear Hawkes processes with \( \omega(\cdot) \geq 0 \)
The End