STATS220: Machine Learning Methods for Neural Data Analysis

Poisson processes

Scott Linderman

April 28, 2025

Outline

▶ From Poisson GLMs to Poisson processes

▶ Four ways to sample a Poisson process

▶ Beyond Poisson: Hawkes processes

From Poisson GLMs to Poisson processes

▶ Consider a Poisson GLM with a log link function,

$$\lambda_t = f(\mathbf{x}^\top \boldsymbol{\beta}) \cdot \Delta t \tag{1}$$
$$y_t \sim \text{Po}(\lambda_t) \tag{2}$$
$$\tag{3}$$

where $f : \mathbb{R} \mapsto \mathbb{R}_+$ is the mean function and Δt is the length of the time interval.

▶ This model treats the **number of events** in the interval as a Poisson random variable with mean λ_t, but it says nothing about the **timing** of the events within intervals.

▶ As we take the limit $\Delta t \to 0$ (roughly speaking), we have higher temporal resolution, but infinitely many bins. In this limit, we obtain a **Poisson process**.

▶ The intensity function specifies the expected number of points in each interval of time or space.
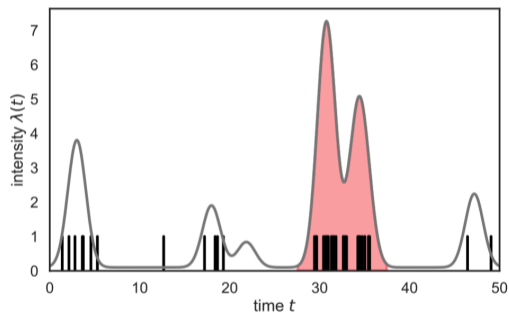
Poisson process

▶ Poisson processes are **stochastic processes** that generate **random sets of points** $\{t_n\}_{n=1}^{N} \subset [0, T]$.

▶ Poisson processes are governed by an **intensity function**, $\lambda(t) : [0, T] \rightarrow \mathbb{R}_+$.

▶ **Property #1:** The number of points in any interval is a Poisson random variable,

$$N(\mathscr{A}) \sim \mathrm{Po}\left(\int_{\mathscr{A}} \lambda(t)\, dt \right) \tag{4}$$

▶ **Property #2:** Disjoint intervals are independent,

$$N(\mathscr{A}) \perp N(\mathscr{B}) \iff \mathscr{A} \cap \mathscr{B} = \varnothing \tag{5}$$
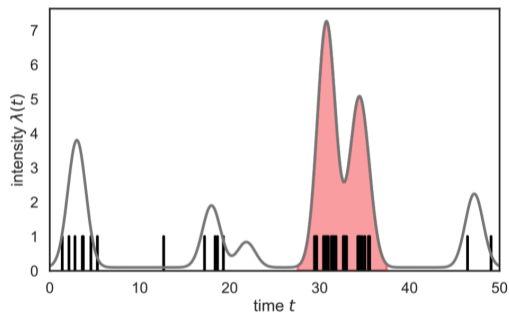
Poisson process

▶ Poisson processes are **stochastic processes** that generate **random sets of points** $\{t_n\}_{n=1}^{N} \subset [0, T]$.

▶ Poisson processes are governed by an **intensity function**, $\lambda(t) : [0, T] \to \mathbb{R}_{+}$.

▶ **Property #1:** The number of points in any interval is a Poisson random variable,

$$N(\mathscr{A}) \sim \mathrm{Po}\left(\int_{\mathscr{A}} \lambda(t)\, \mathrm{d}t \right) \tag{4}$$

▶ **Property #2:** Disjoint intervals are independent,

$$N(\mathscr{A}) \perp\!\!\!\perp N(\mathscr{B}) \iff \mathscr{A} \cap \mathscr{B} = \varnothing \tag{5}$$
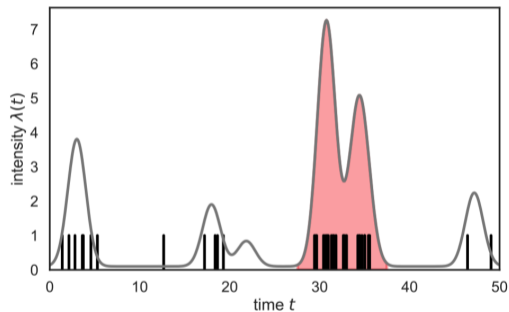
Poisson process

► Poisson processes are **stochastic processes** that generate **random sets of points** $\{t_n\}_{n=1}^{N} \subset [0, T]$.

► Poisson processes are governed by an **intensity function**, $\lambda(t) : [0, T] \to \mathbb{R}_+$.

► **Property #1:** The number of points in any interval is a Poisson random variable,

$$N(\mathscr{A}) \sim \mathrm{Po}\left(\int_{\mathscr{A}} \lambda(t)\, \mathrm{d}t \right) \tag{4}$$

► **Property #2:** Disjoint intervals are independent,

$$N(\mathscr{A}) \perp\!\!\!\perp N(\mathscr{B}) \iff \mathscr{A} \cap \mathscr{B} = \varnothing \tag{5}$$

Example applications of Poisson processes

- ▶ Modeling neural firing rates (of course!)
- ▶ Locations of trees in a forest
- ▶ Locations of stars in astronomical surveys
- ▶ Arrival times of customers in a queue (or HTTP requests to a server)
- ▶ Locations of bombs in London during World War II
- ▶ Times of photon detections on a light sensor
- ▶ Others?

Four ways to sample a Poisson process

1. The top-down approach

2. The interval approach

3. The time-rescaling approach

4. The thinning approach

Top-down sampling of a Poisson process

Given $\lambda(t)$ (and a domain $\mathscr{X}$):

1. Sample the **total number of points**

$$N \sim \text{Po}\left(\int_0^T \lambda(t)\,dt\right) \tag{6}$$

2. Sample the **locations** of the points

$$t_n \stackrel{\text{iid}}{\sim} \frac{\lambda(t)}{\int_0^T \lambda(t')\,dt'} \tag{7}$$

for $n = 1, \ldots, N$.

Question: what assumptions are necessary for this procedure to be tractable?

Deriving the Poisson process likelihood

Exercise: from the top-down sampling process, derive the Poisson process likelihood,

$$p\left(\{t_n\}_{n=1}^{N} \mid \lambda(t)\right) = \tag{8}$$

Intervals of a homogeneous Poisson process

▶ A Poisson process is **homogeneous** if its intensity is constant, $\lambda(t) \equiv \lambda$.

▶ **Property #3:** A homogeneous Poisson process on $[0, T] \subset \mathbb{R}$ (e.g. where points correspond to arrival times) has **independent, exponentially distributed intervals**,

$$\Delta_n = t_n - t_{n-1} \overset{\text{iid}}{\sim} \mathrm{Exp}(\lambda) \tag{9}$$

▶ **Property #4:** A homogeneous Poisson process is **memoryless** — the amount of time until the next point arrives is independent of the time elapsed since the previous point arrived.

Intervals of a homogeneous Poisson process

▶ A Poisson process is **homogeneous** if its intensity is constant, $\lambda(t) \equiv \lambda$.

▶ **Property #3:** A homogeneous Poisson process on $[0, T] \subset \mathbb{R}$ (e.g. where points correspond to arrival times) has **independent, exponentially distributed intervals**,

$$\Delta_n = t_n - t_{n-1} \overset{\text{iid}}{\sim} \text{Exp}(\lambda) \tag{9}$$

▶ **Property #4:** A homogeneous Poisson process is **memoryless** — the amount of time until the next point arrives is independent of the time elapsed since the previous point arrived.

Intervals of a homogeneous Poisson process

▶ A Poisson process is **homogeneous** if its intensity is constant, $\lambda(t) \equiv \lambda$.

▶ **Property #3:** A homogeneous Poisson process on $[0, T] \subset \mathbb{R}$ (e.g. where points correspond to arrival times) has **independent, exponentially distributed intervals**,

$$\Delta_n = t_n - t_{n-1} \overset{\text{iid}}{\sim} \mathrm{Exp}(\lambda) \qquad (9)$$

▶ **Property #4:** A homogeneous Poisson process is **memoryless** — the amount of time until the next point arrives is independent of the time elapsed since the previous point arrived.

Sampling a homogeneous Poisson process by simulating intervals

We can sample a homogeneous Poisson process on $[0, T]$ by simulating intervals as follows:

1. Initialize $\mathscr{T} = \varnothing$ and $t_0 = 0$

2. For $n = 1, 2, \ldots$:

   ▶ Sample $\Delta_n \sim \mathrm{Exp}(\lambda)$.

   ▶ Set $t_n = t_{n-1} + \Delta_n$.

   ▶ If $t_n > T$, break and return $\mathscr{T}$,

   ▶ Else, set $\mathscr{T} \leftarrow \mathscr{T} \cup \{t_n\}$.

Deriving the likelihood of a homogeneous Poisson process

Exercise: from the interval sampling process, derive the likelihood of a homogeneous Poisson process. Show that it is the same as what you derived from the top-down sampling process.

Sampling an inhomogeneous Poisson process by time-rescaling

▶ Now consider an **inhomogeneous** Poisson process on $[0, T]$; i.e. one with a non-constant intensity.

▶ Apply the change of variables,

$$t \mapsto \int_0^t \lambda(\tau) \, d\tau \triangleq \Lambda(t) \tag{10}$$

Note that this is an **invertible transformation** when $\lambda(t) > 0$.

▶ Sample a homogeneous Poisson process with unit rate on $[0, \Lambda(T)]$ to get points $\boldsymbol{U} = \{u_n\}_{n=1}^N$. Then set,

$$\mathcal{T} = \{\Lambda^{-1}(u_n) : u_n \in \boldsymbol{U}\}. \tag{11}$$

▶ **Sanity check:** what is the expected value of N?

Sampling an inhomogeneous Poisson process by time-rescaling

► Now consider an **inhomogeneous** Poisson process on $[0, T]$; i.e. one with a non-constant intensity.

► Apply the change of variables,

$$t \mapsto \int_0^t \lambda(\tau)\,\mathrm{d}\tau \triangleq \Lambda(t) \tag{10}$$

Note that this is an **invertible transformation** when $\lambda(t) > 0$.

► Sample a homogeneous Poisson process with unit rate on $[0, \Lambda(T)]$ to get points $\boldsymbol{U} = \{u_n\}_{n=1}^N$. Then set,

$$\mathscr{T} = \{\Lambda^{-1}(u_n) : u_n \in \boldsymbol{U}\}. \tag{11}$$

► **Sanity check:** what is the expected value of N?

Sampling an inhomogeneous Poisson process by time-rescaling

► Now consider an **inhomogeneous** Poisson process on $[0, T]$; i.e. one with a non-constant intensity.

► Apply the change of variables,

$$t \mapsto \int_0^t \lambda(\tau) \, d\tau \triangleq \Lambda(t) \tag{10}$$

Note that this is an **invertible transformation** when $\lambda(t) > 0$.

► Sample a homogeneous Poisson process with unit rate on $[0, \Lambda(T)]$ to get points $\boldsymbol{U} = \{u_n\}_{n=1}^N$. Then set,

$$\mathcal{T} = \{\Lambda^{-1}(u_n) : u_n \in \boldsymbol{U}\}. \tag{11}$$

► **Sanity check:** what is the expected value of N?

Sampling an inhomogeneous Poisson process by time-rescaling

▶ Now consider an **inhomogeneous** Poisson process on $[0, T]$; i.e. one with a non-constant intensity.

▶ Apply the change of variables,

$$t \mapsto \int_0^t \lambda(\tau) \, d\tau \triangleq \Lambda(t) \tag{10}$$

Note that this is an **invertible transformation** when $\lambda(t) > 0$.

▶ Sample a homogeneous Poisson process with unit rate on $[0, \Lambda(T)]$ to get points $\boldsymbol{U} = \{u_n\}_{n=1}^N$. Then set,

$$\mathcal{T} = \{\Lambda^{-1}(u_n) : u_n \in \boldsymbol{U}\}. \tag{11}$$

▶ **Sanity check:** what is the expected value of N?

Sampling an inhomogeneous Poisson process by time-rescaling, in pictures

Note: this is the analog of **inverse-CDF** sampling.

A Goodness of fit test for inhomogeneous Poisson processes

▶ Brown et al. [2002] used the time-rescaling sampling procedure to develop a goodness-of-fit test for inhomogeneous Poisson processes.

▶ Suppose you observe a set of points $\{t_n\}_{n=1}^N \subset [0, T]$ and you want to test whether they are well-modeled by an inhomogeneous Poisson process with rate $\lambda(t)$.

▶ Let $\Delta_n = \Lambda(t_n) - \Lambda(t_{n-1})$ with $\Lambda(t_0) = 0$. If the model is a good fit, then $\Delta_n \overset{\text{iid}}{\sim} \text{Exp}(1)$.

▶ Perform a further transformation $z_n = 1 - e^{-\Delta_n}$. Then $z_n \overset{\text{iid}}{\sim} \text{Unif}([0, 1])$.

▶ Now sort the z_n's in increasing order into $(z_{(1)}, \ldots, z_{(N)})$, so $z_{(1)}$ is the smallest value.

▶ Intuitively, the points $\left(\frac{n-1/2}{N}, z_{(n)}\right)$ should like along a 45° line.

A Goodness of fit test for inhomogeneous Poisson processes

▶ Brown et al. [2002] used the time-rescaling sampling procedure to develop a goodness-of-fit test for inhomogeneous Poisson processes.

▶ Suppose you observe a set of points $\{t_n\}_{n=1}^N \subset [0, T]$ and you want to test whether they are well-modeled by an inhomogeneous Poisson process with rate $\lambda(t)$.

▶ Let $\Delta_n = \Lambda(t_n) - \Lambda(t_{n-1})$ with $\Lambda(t_0) = 0$. If the model is a good fit, then $\Delta_n \overset{\text{iid}}{\sim} \text{Exp}(1)$.

▶ Perform a further transformation $z_n = 1 - e^{-\Delta_n}$. Then $z_n \overset{\text{iid}}{\sim} \text{Unif}([0, 1])$.

▶ Now sort the z_n's in increasing order into $(z_{(1)}, \ldots, z_{(N)})$, so $z_{(1)}$ is the smallest value.

▶ Intuitively, the points $\left(\frac{n-1/2}{N}, z_{(n)}\right)$ should like along a 45° line.

A Goodness of fit test for inhomogeneous Poisson processes

▶ Brown et al. [2002] used the time-rescaling sampling procedure to develop a goodness-of-fit test for inhomogeneous Poisson processes.

▶ Suppose you observe a set of points $\{t_n\}_{n=1}^{N} \subset [0, T]$ and you want to test whether they are well-modeled by an inhomogeneous Poisson process with rate $\lambda(t)$.

▶ Let $\Delta_n = \Lambda(t_n) - \Lambda(t_{n-1})$ with $\Lambda(t_0) = 0$. If the model is a good fit, then $\Delta_n \overset{\text{iid}}{\sim} \text{Exp}(1)$.

▶ Perform a further transformation $z_n = 1 - e^{-\Delta_n}$. Then $z_n \overset{\text{iid}}{\sim} \text{Unif}([0, 1])$.

▶ Now sort the z_n's in increasing order into $(z_{(1)}, \ldots, z_{(N)})$, so $z_{(1)}$ is the smallest value.

▶ Intuitively, the points $\left(\frac{n - 1/2}{N}, z_{(n)}\right)$ should like along a 45° line.

A Goodness of fit test for inhomogeneous Poisson processes

► Brown et al. [2002] used the time-rescaling sampling procedure to develop a goodness-of-fit test for inhomogeneous Poisson processes.

► Suppose you observe a set of points $\{t_n\}_{n=1}^N \subset [0, T]$ and you want to test whether they are well-modeled by an inhomogeneous Poisson process with rate $\lambda(t)$.

► Let $\Delta_n = \Lambda(t_n) - \Lambda(t_{n-1})$ with $\Lambda(t_0) = 0$. If the model is a good fit, then $\Delta_n \overset{\text{iid}}{\sim} \text{Exp}(1)$.

► Perform a further transformation $z_n = 1 - e^{-\Delta_n}$. Then $z_n \overset{\text{iid}}{\sim} \text{Unif}([0, 1])$.

► Now sort the z_n's in increasing order into $(z_{(1)}, \ldots, z_{(N)})$, so $z_{(1)}$ is the smallest value.

► Intuitively, the points $\left(\frac{n-1/2}{N}, z_{(n)}\right)$ should like along a 45° line.

A Goodness of fit test for inhomogeneous Poisson processes

▶ Brown et al. [2002] used the time-rescaling sampling procedure to develop a goodness-of-fit test for inhomogeneous Poisson processes.

▶ Suppose you observe a set of points $\{t_n\}_{n=1}^N \subset [0, T]$ and you want to test whether they are well-modeled by an inhomogeneous Poisson process with rate $\lambda(t)$.

▶ Let $\Delta_n = \Lambda(t_n) - \Lambda(t_{n-1})$ with $\Lambda(t_0) = 0$. If the model is a good fit, then $\Delta_n \overset{\text{iid}}{\sim} \text{Exp}(1)$.

▶ Perform a further transformation $z_n = 1 - e^{-\Delta_n}$. Then $z_n \overset{\text{iid}}{\sim} \text{Unif}([0, 1])$.

▶ Now sort the z_n's in increasing order into $(z_{(1)}, \ldots, z_{(N)})$, so $z_{(1)}$ is the smallest value.

▶ Intuitively, the points $\left(\frac{n-1/2}{N}, z_{(n)}\right)$ should like along a 45° line.

A Goodness of fit test for inhomogeneous Poisson processes

▶ Brown et al. [2002] used the time-rescaling sampling procedure to develop a goodness-of-fit test for inhomogeneous Poisson processes.

▶ Suppose you observe a set of points $\{t_n\}_{n=1}^N \subset [0, T]$ and you want to test whether they are well-modeled by an inhomogeneous Poisson process with rate $\lambda(t)$.

▶ Let $\Delta_n = \Lambda(t_n) - \Lambda(t_{n-1})$ with $\Lambda(t_0) = 0$. If the model is a good fit, then $\Delta_n \overset{\text{iid}}{\sim} \text{Exp}(1)$.

▶ Perform a further transformation $z_n = 1 - e^{-\Delta_n}$. Then $z_n \overset{\text{iid}}{\sim} \text{Unif}([0, 1])$.

▶ Now sort the z_n's in increasing order into $(z_{(1)}, \ldots, z_{(N)})$, so $z_{(1)}$ is the smallest value.

▶ Intuitively, the points $\left(\frac{n-1/2}{N}, z_{(n)}\right)$ should like along a 45° line.

A Goodness of fit test for inhomogeneous Poisson processes II

► We can check for significant departures from the 45° line using a simple visual test.

► The order statistics $z_{(n)}$ are marginally beta distributed,

$$z_{(n)} \sim \text{Beta}(n, N - n + 1) \tag{12}$$

The mean is $\frac{n}{N+1}$ and its mode is $\frac{n-1}{N-1}$.

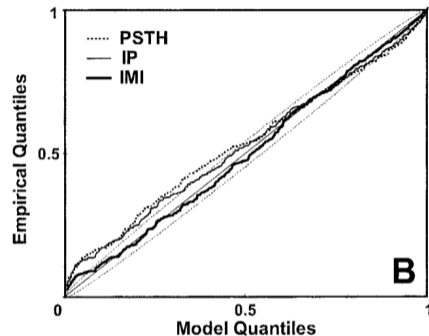► Then, use the 2.5% and 97.5% quantiles of the beta distribution to obtain confidence intervals around the 45° line.

Figure: Figure 1 from Brown et al. [2002].

The Poisson Superposition Principle

► **Property #5:** The union (a.k.a. superposition) of independent Poisson processes is also a Poisson process.

► Suppose we have two independent Poisson processes on the same domain $\mathscr{X}$,

$$\{t_n\}_{n=1}^{N} \sim \mathrm{PP}(\lambda_1(t)) \tag{13}$$
$$\{t'_m\}_{m=1}^{M} \sim \mathrm{PP}(\lambda_2(t)) \tag{14}$$

Then

$$\{t_n\}_{n=1}^{N} \cup \{t'_m\}_{m=1}^{M} \sim \mathrm{PP}(\lambda_1(t) + \lambda_2(t)) \tag{15}$$

► This is the analog of the fact that the sum of independent Poisson random variables is Poisson.

The Poisson Superposition Principle

▶ **Property #5:** The union (a.k.a. superposition) of independent Poisson processes is also a Poisson process.

▶ Suppose we have two independent Poisson processes on the same domain $\mathscr{X}$,

$$\{t_n\}_{n=1}^N \sim \mathrm{PP}(\lambda_1(t)) \tag{13}$$

$$\{t'_m\}_{m=1}^M \sim \mathrm{PP}(\lambda_2(t)) \tag{14}$$

Then

$$\{t_n\}_{n=1}^N \cup \{t'_m\}_{m=1}^M \sim \mathrm{PP}(\lambda_1(t) + \lambda_2(t)) \tag{15}$$

▶ This is the analog of the fact that the sum of independent Poisson random variables is Poisson.

The Poisson Superposition Principle

▶ **Property #5:** The union (a.k.a. superposition) of independent Poisson processes is also a Poisson process.

▶ Suppose we have two independent Poisson processes on the same domain $\mathscr{X}$,

$$\{t_n\}_{n=1}^N \sim \mathrm{PP}(\lambda_1(t)) \tag{13}$$

$$\{t'_m\}_{m=1}^M \sim \mathrm{PP}(\lambda_2(t)) \tag{14}$$

Then

$$\{t_n\}_{n=1}^N \cup \{t'_m\}_{m=1}^M \sim \mathrm{PP}(\lambda_1(t) + \lambda_2(t)) \tag{15}$$

▶ This is the analog of the fact that the sum of independent Poisson random variables is Poisson.

Poisson thinning

▶ The opposite of Poisson superposition is **Poisson thinning**.

▶ Suppose we have points $\{t_n\}_{n=1}^N \sim \mathrm{PP}(\lambda(t))$ where $\lambda(t) = \lambda_1(t) + \lambda_2(t)$.

▶ Sample independent binary variables

$$z_n \sim \mathrm{Bern}\left(\frac{\lambda_1(t_n)}{\lambda_1(t_n) + \lambda_2(t_n)}\right). \tag{16}$$

▶ Then $\{t_n : z_n = 1\} \sim \mathrm{PP}(\lambda_1(t))$ and $\{t_n : z_n = 0\} \sim \mathrm{PP}(\lambda_2(t))$.

Sampling a Poisson process by thinning

Exercise: Use Poisson thinning to sample an inhomogeneous Poisson process with a bounded intensity, $\lambda(t) \leq \lambda_{\max}$.

Question: What Monte Carlo sampling method is this akin to?

Outline

▶ Defining properties of a Poisson process

▶ Four ways to sample a Poisson process

▶ **Beyond Poisson: Hawkes Processes**

What's not to love about Poisson processes?

Conditional intensity functions

▶ One way of introducing dependence is via an **autoregressive model**. Consider a point process on a time interval $[0, T]$.

▶ Let $\lambda(t \mid \mathcal{H}_t)$ denote a **conditional intensity function** where $\mathcal{H}_t$ is the **history** of points before time t.

▶ Technically, $\mathcal{H}_t$ is a **filtration** in the language of stochastic processes.

▶ Allowing past points to influence the intensity function enables more complex, non-Poisson models.

Conditional intensity functions

▶ One way of introducing dependence is via an **autoregressive model**. Consider a point process on a time interval $[0, T]$.

▶ Let $\lambda(t \mid \mathscr{H}_t)$ denote a **conditional intensity function** where $\mathscr{H}_t$ is the **history** of points before time t.

▶ Technically, $\mathscr{H}_t$ is a **filtration** in the language of stochastic processes.

▶ Allowing past points to influence the intensity function enables more complex, non-Poisson models.

Conditional intensity functions

▶ One way of introducing dependence is via an **autoregressive model**. Consider a point process on a time interval $[0, T]$.

▶ Let $\lambda(t \mid \mathcal{H}_t)$ denote a **conditional intensity function** where $\mathcal{H}_t$ is the **history** of points before time t.

▶ Technically, $\mathcal{H}_t$ is a **filtration** in the language of stochastic processes.

▶ Allowing past points to influence the intensity function enables more complex, non-Poisson models.

Conditional intensity functions

► One way of introducing dependence is via an **autoregressive model**. Consider a point process on a time interval $[0, T]$.

► Let $\lambda(t \mid \mathcal{H}_t)$ denote a **conditional intensity function** where $\mathcal{H}_t$ is the **history** of points before time t.

► Technically, $\mathcal{H}_t$ is a **filtration** in the language of stochastic processes.

► Allowing past points to influence the intensity function enables more complex, non-Poisson models.

Hawkes processes

▶ Hawkes processes [Hawkes, 1971] are **self-exciting point processes**.

▶ Their conditional intensity function is modeled as,

$$\lambda(t \mid \mathcal{H}_t) = \lambda_0 + \sum_{t_n \in \mathcal{H}_t} h(t - t_n), \tag{17}$$

where $h : \mathbb{R}_+ \mapsto \mathbb{R}_+$ is a positive **impulse response** or **influence function**.

▶ For example, the impulse responses could be modeled as exponential functions,

$$h(\Delta t) = \frac{w}{\tau} e^{-\frac{\Delta t}{\tau}} = w \cdot \mathrm{Exp}(\Delta t; \tau), \tag{18}$$

where $\tau \in \mathbb{R}_+$ is a time-constant governing the rate of decay and $w \in \mathbb{R}_+$ is a scaling parameter such that $\int_0^\infty h(\Delta t) \, d\Delta t = w$.

Hawkes processes

► Hawkes processes [Hawkes, 1971] are **self-exciting point processes**.

► Their conditional intensity function is modeled as,

$$\lambda(t \mid \mathscr{H}_t) = \lambda_0 + \sum_{t_n \in \mathscr{H}_t} h(t - t_n), \tag{17}$$

where $h : \mathbb{R}_+ \mapsto \mathbb{R}_+$ is a positive **impulse response** or **influence function**.

► For example, the impulse responses could be modeled as exponential functions,

$$h(\Delta t) = \frac{w}{\tau} e^{-\frac{\Delta t}{\tau}} = w \cdot \mathrm{Exp}(\Delta t; \tau), \tag{18}$$

where $\tau \in \mathbb{R}_+$ is a time-constant governing the rate of decay and $w \in \mathbb{R}_+$ is a scaling parameter such that $\int_0^\infty h(\Delta t) \, d\Delta t = w$.

Hawkes processes

▶ Hawkes processes [Hawkes, 1971] are **self-exciting point processes**.

▶ Their conditional intensity function is modeled as,

$$\lambda(t \mid \mathscr{H}_t) = \lambda_0 + \sum_{t_n \in \mathscr{H}_t} h(t - t_n), \tag{17}$$

where $h : \mathbb{R}_+ \mapsto \mathbb{R}_+$ is a positive **impulse response** or **influence function**.

▶ For example, the impulse responses could be modeled as exponential functions,

$$h(\Delta t) = \tfrac{w}{\tau} e^{-\frac{\Delta t}{\tau}} = w \cdot \mathrm{Exp}(\Delta t; \tau), \tag{18}$$

where $\tau \in \mathbb{R}_+$ is a time-constant governing the rate of decay and $w \in \mathbb{R}_+$ is a scaling parameter such that $\int_0^\infty h(\Delta t) \, \mathrm{d}\Delta t = w$.

Hawkes processes, in pictures

Maximum likelihood estimation for Hawkes processes

▶ Suppose we observe a collection of time points $\{t_n\}_{n=1}^{N} \subset [0, T]$ and want to estimate the parameters $\theta = (\lambda_0, w)$ of a Hawkes process with an exponential impulse response function. (Consider τ to be fixed.)

▶ The Hawkes process log likelihood is just like that of a Poisson process,

$$\log p(\{t_n\}_{n=1}^{N} \mid \theta) = -\int_0^T \lambda_\theta(t \mid \mathcal{H}_t) \, \mathrm{d}t + \sum_{n=1}^{N} \log \lambda_\theta(t_n \mid \mathcal{H}_t) \tag{19}$$

Maximum likelihood estimation for Hawkes processes

▶ Substituting in the form of the conditional intensity, we can simplify the log likelihood to,

$$\log p(\{t_n\}_{n=1}^N \mid \boldsymbol{\theta}) = -\int_0^T \left[\lambda_0 + w \sum_{t_n \in \mathcal{H}_t} \mathrm{Exp}(t - t_n; \tau) \, \mathrm{d}t \right]$$
$$+ \sum_{n=1}^N \log \left(\lambda_0 + w \sum_{t_m \in \mathcal{H}_{t_n}} \mathrm{Exp}(t_n - t_m; \tau) \right) \quad (20)$$

$$\approx -\boldsymbol{\theta}^\top \boldsymbol{\phi}_0 + \sum_{n=1}^N \log \left(\boldsymbol{\theta}^\top \boldsymbol{\phi}_n \right) \quad (21)$$

where $\boldsymbol{\phi}_0 = (T, N)^\top$ and $\boldsymbol{\phi}_n = \left(1, \sum_{t_m \in \mathcal{H}_{t_n}} \mathrm{Exp}(t_n - t_m; \tau) \right)^\top$.

▶ **Questions:** What approximation did we make? How would you maximize the log likelihood as a function of $\boldsymbol{\theta}$?

Nonlinear Hawkes processes

Hawkes processes only allow for excitatory impulse responses ($w \in \mathbb{R}_+$). What if we want to model **self-inhibitory** influences as well? A natural generalization of the Hawkes process is,

$$\lambda(t \mid \mathscr{H}_t) = f(\beta_0 + \beta \sum_{t_n \in \mathscr{H}_t} g(t - t_n))$$

where $f : \mathbb{R} \mapsto \mathbb{R}_+$ is a rectifying nonlinear function and $\beta_0, \beta \in \mathbb{R}$ can be positive or negative.

These models are a bit more challenging because the Poisson process likelihood requires the integrated intesity function, which is generally intractable. Instead, we can approximate the integral with a Riemann sum,

$$\int \lambda(t \mid \mathscr{H}_t) \, dt \approx \sum_{i=0}^{\lfloor T/\Delta \rfloor - 1} \lambda(i\Delta \mid \mathscr{H}_{i\Delta})\Delta$$

which essentially treats the intensity as piecewise constant within each interval of length Δ.

Nonlinear Hawkes processes

In that approximation, the Poisson process likelihood is equivalent, up to a scale factor, to a product of Poisson likelihoods,

$$
\begin{aligned}
p(\{t_n\}_{n=1}^N) &\approx \prod_{i=0}^{\lfloor T/\Delta \rfloor - 1} \text{Po}\big(N_i \mid \lambda(i\Delta \mid \mathcal{H}_{i\Delta})\Delta\big) \\
&= \prod_{i=0}^{\lfloor T/\Delta \rfloor - 1} \text{Po}\left(N_i \mid f\left(\beta_0 + \beta \sum_{t_n \in \mathcal{H}_{i\Delta}} g(t - t_n) \right)\Delta \right)
\end{aligned}
$$

where N_i is the number of events in the i-th interval, $[iD, (i+1)D)$.

This is a Poisson GLM with weights $\beta = (\beta_0, \beta)$! We can fit it using all the techniques we developed earlier in the course.

Marked point processes

▶ Now suppose we observed points from *S* difference **sources**.

▶ We can represent the points as a set of tuples, $\{(t_n, s_n)\}_{n=1}^{N}$ where $t_n \in [0, T]$ denotes the time and $s_n \in \{1, \ldots, S\}$ denotes the source of the *n*-th point.

▶ We will model them as a **marked point process**.

▶ Like before, we have a (conditional) intensity function, but this time is defined over time and marks,

$$\lambda(t, s \mid \mathscr{H}_t) : [0, T] \times \{1, \ldots, S\} \mapsto \mathbb{R}_+ \tag{22}$$

▶ When *s* takes on a discrete set of values, we often use the shorthand,

$$\lambda_s(t \mid \mathscr{H}_t) \triangleq \lambda(t, s \mid \mathscr{H}_t) \tag{23}$$

to denote the intensity for the *s*-th source.

Multivariate Hawkes processes

▶ A **multivariate** Hawkes process is a marked point process with **mutually excitatory** interactions.

▶ It is defined by the conditional intensity functions,

$$\lambda_s(t \mid \mathcal{H}_t) = \lambda_{s,0} + \sum_{(t_n,s_n)\in\mathcal{H}_t} h_{s_n,s}(t-t_n). \tag{24}$$

where $h_{s,s'}(\Delta t)$ is a **directed impulse response** from points on source s to the intensity of s'.

▶ Again, it is common to model the impulse responses as weighted probability densities; e.g.,

$$h_{s,s'}(\Delta t) = w_{s,s'} \cdot \text{Exp}(\Delta t; \tau_{s,s'}) \tag{25}$$

where $w_{s,s'}$ is the weight.

▶ Like before, the weights can be estimated using maximum likelihood estimation.

Multivariate Hawkes processes

▶ A **multivariate** Hawkes process is a marked point process with **mutually excitatory** interactions.

▶ It is defined by the conditional intensity functions,

$$\lambda_s(t \mid \mathcal{H}_t) = \lambda_{s,0} + \sum_{(t_n,s_n)\in\mathcal{H}_t} h_{s_n,s}(t - t_n). \tag{24}$$

where $h_{s,s'}(\Delta t)$ is a **directed impulse response** from points on source s to the intensity of s'.

▶ Again, it is common to model the impulse responses as weighted probability densities; e.g.,

$$h_{s,s'}(\Delta t) = w_{s,s'} \cdot \mathrm{Exp}(\Delta t; \tau_{s,s'}) \tag{25}$$

where $w_{s,s'}$ is the weight.

▶ Like before, the weights can be estimated using maximum likelihood estimation.

Multivariate Hawkes processes

▶ A **multivariate** Hawkes process is a marked point process with **mutually excitatory** interactions.

▶ It is defined by the conditional intensity functions,

$$\lambda_s(t \mid \mathcal{H}_t) = \lambda_{s,0} + \sum_{(t_n,s_n) \in \mathcal{H}_t} h_{s_n,s}(t - t_n). \tag{24}$$

where $h_{s,s'}(\Delta t)$ is a **directed impulse response** from points on source s to the intensity of s'.

▶ Again, it is common to model the impulse responses as weighted probability densities; e.g.,

$$h_{s,s'}(\Delta t) = w_{s,s'} \cdot \mathrm{Exp}(\Delta t; \tau_{s,s'}) \tag{25}$$

where $w_{s,s'}$ is the weight.

▶ Like before, the weights can be estimated using maximum likelihood estimation.
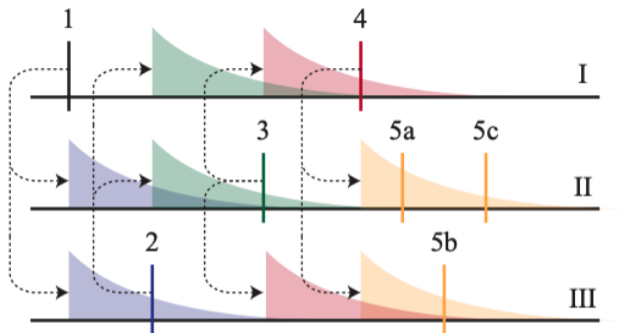
Multivariate Hawkes Processes II

Figure 1: Illustration of a Hawkes process. Events induce impulse responses on connected processes and spawn "child" events. See the main text for a complete description.

From Linderman and Adams [2014].

Discovering latent network structure in point process data

▶ We can think of the weights as defining a **directed network**,

$$\mathbf{W} = \begin{bmatrix} w_{1,1} & \cdots & w_{1,S} \\ \vdots & & \vdots \\ w_{S,1} & \cdots & w_{S,S} \end{bmatrix} \tag{26}$$

where $w_{s,s'} \in \mathbb{R}_+$ is the strength of influence that events (points) on source s induce on the intensity of source s'.

▶ However, we don't directly observe the network. We only observed it indirectly through the point process.

▶ **Question:** when is a multivariate Hawkes process stable, in that the intensity tends to a finite value in the infinite time limit?

Multivariate Hawkes processes as Poisson clustering processes

▶ Note that the conditional intensity in eq. (24) is a sum of a background intensity and a bunch of non-negative impulse responses.

$$\lambda_s(t \mid \mathscr{H}_t) = \lambda_{0,s} + \sum_{(t_n, s_n) \in \mathscr{H}_t} h_{s_n, s}(t - t_n). \tag{27}$$

▶ **Question:** which property of Poisson processes applied to such intensities?

Multivariate Hawkes processes as Poisson clustering processes

▶ Note that the conditional intensity is a sum of a background intensity and a bunch of non-negative impulse responses,

$$\lambda_s(t \mid \mathcal{H}_t) = \lambda_{s,0} + \sum_{(t_n, s_n) \in \mathcal{H}_t} h_{s_n, s}(t - t_n). \tag{28}$$

▶ **Question:** which property of Poisson processes applied to such intensities?

▶ Using the **Poisson superposition principle**, we can partition the points $\mathcal{T}_s = \{t_n : s_n = s\}$ from source s into **clusters** attributed to either the background or to one of the impulse responses.

$$\mathcal{T}_s = \bigcup_{n=0}^{N} \mathcal{T}_{s,n} \tag{29}$$

where

$$\mathcal{T}_{s,0} \sim \mathrm{PP}(\lambda_{s,0}) \qquad \qquad \text{[background points]} \tag{30}$$

$$\mathcal{T}_{s,n} \sim \mathrm{PP}(h_{s_n, s}(t - t_n)) \qquad \qquad \text{[points induced by } (t_n, s_n)] \tag{31}$$

Multivariate Hawkes processes as Poisson clustering processes

▶ Now the weights have an intuitive interpretation. Plugging in the definition of the impulse response,

$$\mathcal{T}_{s,n} \sim \mathrm{PP}\Big(w_{s_n,s} \cdot \mathrm{Exp}(t - t_n; \tau_{s_n,s}) \Big). \tag{32}$$

▶ **Question:** What is the expected number of points induced by this impulse response, i.e. $\mathbb{E}[|\mathcal{T}_{s,n}|]$?

Conjugate Bayesian inference for multivariate Hawkes processes

▶ Let's put a gamma prior on the weights,

$$w_{s,s'} \sim \mathrm{Ga}(\alpha, \beta). \tag{33}$$

▶ **Question:** suppose we know the partition of points; i.e. we knew the clusters $\mathcal{T}_{s,n}$. What is the conditional distribution,

$$p\left(w_{s,s'} \mid \{\{\mathcal{T}_{s,n}\}_{n=0}^{N}\}_{s=1}^{S}\right) = \tag{34}$$

Conjugate Bayesian inference for multivariate Hawkes processes II

▶ We don't know the partition of spikes in general, but we do know its conditional distribution!

▶ Let $z_n \in \{0, \ldots, n-1\}$ denote the cluster to which the n-th spike is assigned, with $z_n = 0$ denoting the background cluster. With this notation,

$$\mathcal{T}_{s,n} = \{(t_{n'}, s_{n'}) : s_{n'} = s \wedge z_{n'} = n\}. \tag{35}$$

▶ **Question:** what is the conditional distribution of the cluster assignment,

$$p(z_n \mid \{(t_n, s_n)\}_{n=1}^{N}; \boldsymbol{\theta}) = \tag{36}$$

▶ Using these two conditional distributions, we can derive a simple Gibbs sampling algorithm that alternates between sampling the weights given the clusters and the clusters given the weights.

Beyond Poisson: Hawkes processes

▶ Hawkes processes are only one way of going beyond Poisson processes.

▶ Whereas Hawkes processes take an autoregressive approach, **doubly stochastic point processes** (a.k.a. **Cox processes**) take a latent variable approach.

▶ In these models, the intensity itself is modeled as a stochastic process,

$$\lambda(t) \sim p(\lambda). \tag{37}$$

▶ For example, consider the model,

$$\lambda(t) = g(f(t)) \quad \text{where} \quad f \sim \mathrm{GP}(\mu(\cdot), K(\cdot, \cdot)). \tag{38}$$

When g is the exponential function, this is called a **log Gaussian Cox process**. When g is the sigmoid function, this is called a **sigmoidal Gaussian Cox process** [Adams et al., 2009].

▶ Aternatively, take λ to be a convolution of a Poisson process with a non-negative kernel; this is called a Neyman-Scott process [Wang et al., 2022, e.g.].

Conclusion

▶ Poisson processes are stochastic processes that generate discrete sets of points.

▶ They are defined by an intensity function $\lambda(t)$, which specifies the expected number of points in each interval of time or space.

▶ We can build in dependencies by conditioning on past points or introducing latent variables.

▶ Poisson process modeling boils down to inferring the intensity. We can take various parametric and nonparametric approaches.

▶ The hardness comes about when the integral in the Poisson process likelihood is intractable.

▶ As we will see next time, Poisson processes are also mathematical building blocks for Bayesian nonparametric models with random measures.

References

Emery N Brown, Riccardo Barbieri, Valérie Ventura, Robert E Kass, and Loren M Frank. The time-rescaling theorem and its application to neural spike train data analysis. *Neural computation*, 14(2):325–346, 2002.

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

Scott Linderman and Ryan Adams. Discovering latent network structure in point process data. In *International Conference on Machine Learning*, pages 1413–1421. PMLR, 2014.

Ryan Prescott Adams, Iain Murray, and David JC MacKay. Tractable nonparametric Bayesian inference in poisson processes with Gaussian process intensities. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 9–16, 2009.

Yixin Wang, Anthony Degleris, Alex H Williams, and Scott W Linderman. Spatiotemporal clustering with Neyman-Scott processes via connections to Bayesian nonparametric mixture models. *arXiv preprint arXiv:2201.05044*, 2022.