

# Describing Spike Trains

Maneesh Sahani

Gatsby Computational Neuroscience Unit  
University College London

February 2024

## Spikes and Neural Coding

- ▶ Most information exchange between neurons (particularly in vertebrates) appears to happen using action potentials or “spikes”.

## Spikes and Neural Coding

- ▶ Most information exchange between neurons (particularly in vertebrates) appears to happen using action potentials or “spikes”.
- ▶ In seeking to understand neural computation, it makes sense to ask how information (about sensory variables; inferences about the world; action plans; cognitive states . . . ) is represented by those spikes.

## Spikes and Neural Coding

- ▶ Most information exchange between neurons (particularly in vertebrates) appears to happen using action potentials or “spikes”.
- ▶ In seeking to understand neural computation, it makes sense to ask how information (about sensory variables; inferences about the world; action plans; cognitive states . . . ) is represented by those spikes.
- ▶ Experimental evidence comes from sensory, and increasingly cognitive, settings.

## Spikes and Neural Coding

- ▶ Most information exchange between neurons (particularly in vertebrates) appears to happen using action potentials or “spikes”.
- ▶ In seeking to understand neural computation, it makes sense to ask how information (about sensory variables; inferences about the world; action plans; cognitive states . . . ) is represented by those spikes.
- ▶ Experimental evidence comes from sensory, and increasingly cognitive, settings.
  - ▶ needs ability to identify and repeat the same content

## Spikes and Neural Coding

- ▶ Most information exchange between neurons (particularly in vertebrates) appears to happen using action potentials or “spikes”.
- ▶ In seeking to understand neural computation, it makes sense to ask how information (about sensory variables; inferences about the world; action plans; cognitive states . . . ) is represented by those spikes.
- ▶ Experimental evidence comes from sensory, and increasingly cognitive, settings.
  - ▶ needs ability to identify and repeat the same content
  - ▶ computational methods are needed to characterise and quantify these results.

## Spikes and Neural Coding

- ▶ Most information exchange between neurons (particularly in vertebrates) appears to happen using action potentials or “spikes”.
- ▶ In seeking to understand neural computation, it makes sense to ask how information (about sensory variables; inferences about the world; action plans; cognitive states . . . ) is represented by those spikes.
- ▶ Experimental evidence comes from sensory, and increasingly cognitive, settings.
  - ▶ needs ability to identify and repeat the same content
  - ▶ computational methods are needed to characterise and quantify these results.
- ▶ Theory can tell us what representations *should* look like.

## Spikes and Neural Coding

- ▶ Most information exchange between neurons (particularly in vertebrates) appears to happen using action potentials or “spikes”.
- ▶ In seeking to understand neural computation, it makes sense to ask how information (about sensory variables; inferences about the world; action plans; cognitive states . . . ) is represented by those spikes.
- ▶ Experimental evidence comes from sensory, and increasingly cognitive, settings.
  - ▶ needs ability to identify and repeat the same content
  - ▶ computational methods are needed to characterise and quantify these results.
- ▶ Theory can tell us what representations *should* look like.
- ▶ Theory also suggests what internal variables might need to be represented:



## Spikes and Neural Coding

- ▶ Most information exchange between neurons (particularly in vertebrates) appears to happen using action potentials or “spikes”.
- ▶ In seeking to understand neural computation, it makes sense to ask how information (about sensory variables; inferences about the world; action plans; cognitive states . . . ) is represented by those spikes.
- ▶ Experimental evidence comes from sensory, and increasingly cognitive, settings.
  - ▶ needs ability to identify and repeat the same content
  - ▶ computational methods are needed to characterise and quantify these results.
- ▶ Theory can tell us what representations *should* look like.
- ▶ Theory also suggests what internal variables might need to be represented:
  - ▶ categorical variables

## Spikes and Neural Coding

- ▶ Most information exchange between neurons (particularly in vertebrates) appears to happen using action potentials or “spikes”.
- ▶ In seeking to understand neural computation, it makes sense to ask how information (about sensory variables; inferences about the world; action plans; cognitive states . . . ) is represented by those spikes.
- ▶ Experimental evidence comes from sensory, and increasingly cognitive, settings.
  - ▶ needs ability to identify and repeat the same content
  - ▶ computational methods are needed to characterise and quantify these results.
- ▶ Theory can tell us what representations *should* look like.
- ▶ Theory also suggests what internal variables might need to be represented:
  - ▶ categorical variables
  - ▶ uncertainty

## Spikes and Neural Coding

- ▶ Most information exchange between neurons (particularly in vertebrates) appears to happen using action potentials or “spikes”.
- ▶ In seeking to understand neural computation, it makes sense to ask how information (about sensory variables; inferences about the world; action plans; cognitive states . . . ) is represented by those spikes.
- ▶ Experimental evidence comes from sensory, and increasingly cognitive, settings.
  - ▶ needs ability to identify and repeat the same content
  - ▶ computational methods are needed to characterise and quantify these results.
- ▶ Theory can tell us what representations *should* look like.
- ▶ Theory also suggests what internal variables might need to be represented:
  - ▶ categorical variables
  - ▶ uncertainty
  - ▶ reward predictions and errors

## Notation for spike trains

A **spike train** is the sequence of times at which a cell spikes:

$$\mathcal{S} = \{t_1, t_2, \dots, t_N\}.$$

It is often useful to write this as a function in time using the Dirac-delta form,

$$s(t) = \sum_{i=1}^N \delta(t - t_i) \quad (\text{D\&A call this } \rho(t))$$

or using a (cumulative) counting function for the number of spikes to time  $t$ ,

$$N(t) = \int_0^{\rightarrow t} d\xi s(\xi)$$

( $\rightarrow t$  means that  $t$  is not included in the integral)

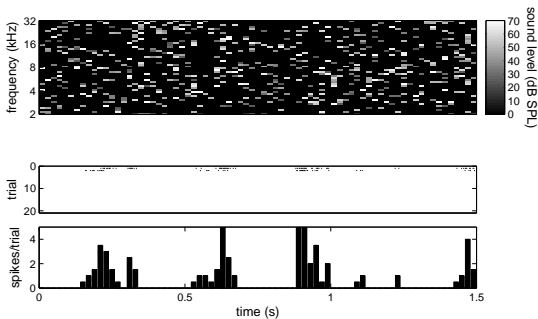
or as a vector by discretizing with time step  $\Delta t$

$$\mathbf{s} = (s_1 \dots s_{T/\Delta t}); \quad s_t = \int_{t-\Delta t}^{\rightarrow t} d\xi s(\xi)$$

Note that the neural refractory period means that for  $\Delta t \approx 1 \text{ ms}$ ,  $s_t$  is binary.

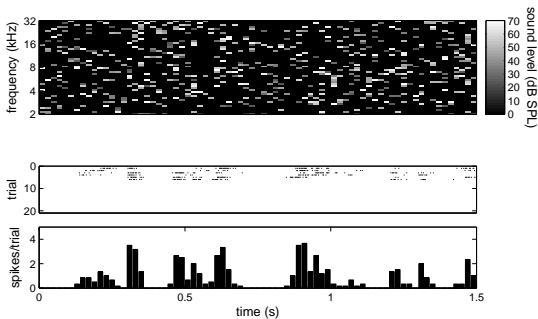
## Variability

Empirically, spike train responses to a repeated stimulus are (very) variable. This is particularly true in the cortex, but might be less so at earlier stages.



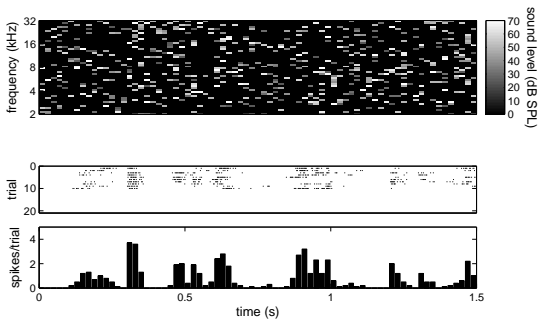
## Variability

Empirically, spike train responses to a repeated stimulus are (very) variable. This is particularly true in the cortex, but might be less so at earlier stages.



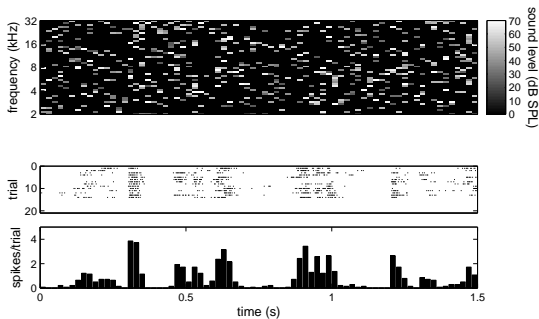
## Variability

Empirically, spike train responses to a repeated stimulus are (very) variable. This is particularly true in the cortex, but might be less so at earlier stages.



## Variability

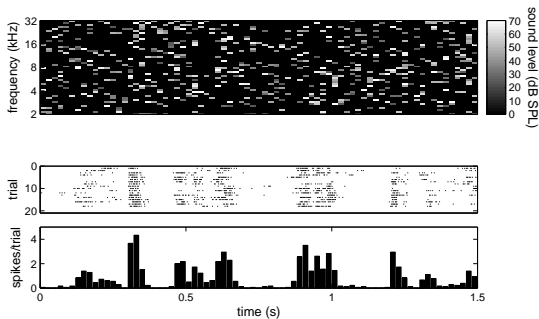
Empirically, spike train responses to a repeated stimulus are (very) variable. This is particularly true in the cortex, but might be less so at earlier stages.





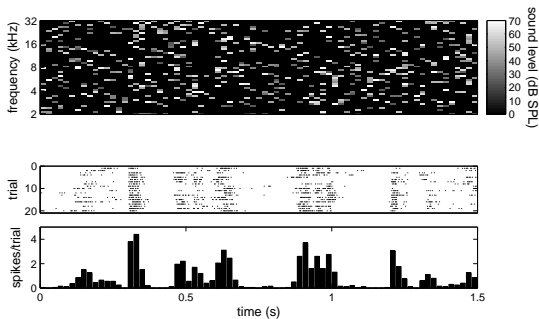
## Variability

Empirically, spike train responses to a repeated stimulus are (very) variable. This is particularly true in the cortex, but might be less so at earlier stages.



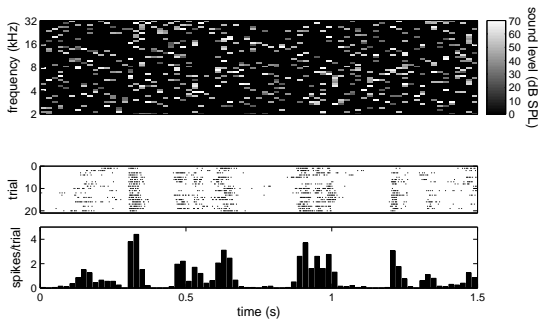
## Variability

Empirically, spike train responses to a repeated stimulus are (very) variable. This is particularly true in the cortex, but might be less so at earlier stages.



## Variability

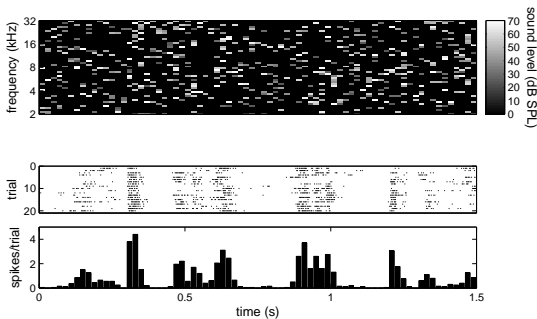
Empirically, spike train responses to a repeated stimulus are (very) variable. This is particularly true in the cortex, but might be less so at earlier stages.



This variability probably arises in more than one way.

## Variability

Empirically, spike train responses to a repeated stimulus are (very) variable. This is particularly true in the cortex, but might be less so at earlier stages.

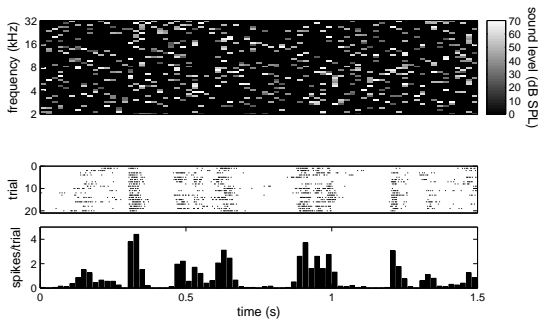


This variability probably arises in more than one way.

- ▶ **Stochasticity.** Perhaps due to vesicle release; or thermal noise in conductances. May be “noise” and/or part of the computation (e.g. sampling).

## Variability

Empirically, spike train responses to a repeated stimulus are (very) variable. This is particularly true in the cortex, but might be less so at earlier stages.

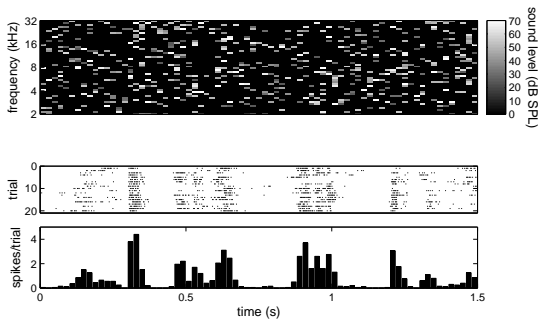


This variability probably arises in more than one way.

- ▶ **Stochasticity.** Perhaps due to vesicle release; or thermal noise in conductances. May be “noise” and/or part of the computation (e.g. sampling).
- ▶ **Ongoing processes.** The brain doesn’t just react to sensory input. Ongoing processing is likely to affect firing, particularly in cortex; and there is experimental evidence for this. This might lead to variability on a slower time-scale than noise.

## Variability

Empirically, spike train responses to a repeated stimulus are (very) variable. This is particularly true in the cortex, but might be less so at earlier stages.



This variability probably arises in more than one way.

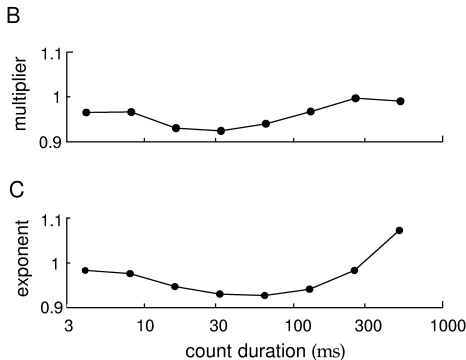
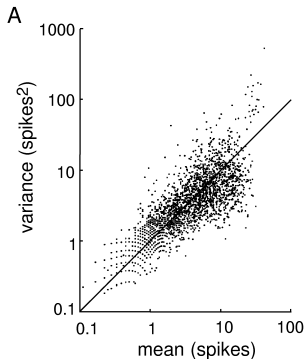
- ▶ **Stochasticity.** Perhaps due to vesicle release; or thermal noise in conductances. May be “noise” and/or part of the computation (e.g. sampling).
- ▶ **Ongoing processes.** The brain doesn’t just react to sensory input. Ongoing processing is likely to affect firing, particularly in cortex; and there is experimental evidence for this. This might lead to variability on a slower time-scale than noise.

We do not know the relative sizes of these contributions.

## Count variability

The variability extends to the total number of spikes (say in response to the  $i$ th stimulus repetition)  $N_i = \int_0^T d\xi s_i(\xi)$

Variability in  $N_i$  is on order of its mean.



Fits of the form  $\text{Var}[N_i] = A \cdot \mathbb{E}[N_i]^B$  yield values of  $A$  and  $B$  between about 1 and 1.5.

## Point Processes

A probabilistic process that produces events of the type

$$S = \{t_1, \dots, t_N\} \subset \mathcal{T}$$

is called a **point process**.

This is the statistical object best suited for the description of spike trains. We take  $\mathcal{T} = [0, T]$  to be an interval of time.

Every point process (on an ordered set) is associated with a dual **counting process** which produces events of the type

$$N(t) \text{ such that } N(t) \geq 0$$

$$N(t') \geq N(t) \text{ if } t' > t$$

$$N(t) - N(s) = N[s, t) \in \mathbb{Z}$$

$N(t)$  gives the number of events with  $t_i < t$ .



## Homogeneous Poisson Process: $N_\lambda(t)$

In the simplest point process, events are all **independent** and occur at a fixed **rate**  $\lambda$ .

Independence is defined formally:

1. **Independence.** For all disjoint intervals  $[s, t)$  and  $[s', t')$ ,  $N_\lambda[s, t) \perp N_\lambda[s', t')$ .

Knowing the number (or times) of one or more events tells us nothing about other possible events.

## Homogeneous Poisson Process: $N_\lambda(t)$

The rate condition can be defined in two ways.

If we assume that  $\lim_{ds \rightarrow 0} N_\lambda[s, s + ds) \in \{0, 1\}$  (technically **conditional orderliness** – at most one event occurs at one time) then it is sufficient to assume that

2. **Mean event rate.**  $\mathbb{E}[N_\lambda[s, t)] = (t - s)\lambda.$

Without assuming conditional orderliness, we could instead define the process by giving the whole distribution of  $N_\lambda[s, t)$ . Instead, we will use the more restrictive defining assumption to derive the distribution.

## Homogeneous Poisson Process: $N_\lambda(t)$

Divide  $[s, t)$  into  $M$  bins of length  $\Delta$  (i.e.  $M = (t - s)/\Delta$ ). If  $\Delta \ll 1/\lambda$  conditional orderliness implies that spike count per bin is binary.

For a binary random variable, the expectation is the same as the probability of event, so  $\lambda\Delta \approx P(N[t, t + \Delta) = 1)$ .

Thus, distribution of  $N[s, t)$  binomial:

$$\begin{aligned} P[N_\lambda[s, t) = n] &= \binom{M}{n} (\lambda\Delta)^n (1 - \lambda\Delta)^{M-n} \\ &= \frac{M!}{n!(M-n)!} \left(\frac{\lambda(t-s)}{M}\right)^n \left(1 - \frac{\lambda(t-s)}{M}\right)^{M-n} \end{aligned}$$

write  $\mu = \lambda(t - s)$

$$= \frac{\mu^n}{n!} \frac{M(M-1)\cdots(M-n+1)}{M^n} \left(1 - \frac{\mu}{M}\right)^{-n} \left(1 - \frac{\mu}{M}\right)^M$$

now take the limit  $\Delta \rightarrow 0$  or, equivalently,  $M \rightarrow \infty$

$$= \frac{\mu^n}{n!} 1^n 1^n e^{-\mu} = \frac{\mu^n e^{-\mu}}{n!}$$

So the spike count in any interval is Poisson distributed.

## Homogeneous Poisson Process: $N_\lambda(t)$

So a Poisson **process** produces event counts which follow the Poisson **distribution**. As we mentioned above, we could instead have dispensed with the conditional orderliness assumption and instead made this a defining property of the process:

2'. **Count distribution.**  $N_\lambda[s, t] \sim \text{Pois}[(t - s)\lambda]$ .

We will now derive a number of properties of the homogeneous Poisson process. These are good to know. We will also employ some tricks in the derivations that can be applied more generally.

## Count Variance

$$\begin{aligned}\mathbb{V}[N_\lambda[s, t]] &= \langle (n - \mu)^2 \rangle \\ &= \langle n^2 \rangle - \mu^2 \\ &= \langle n(n-1) + n \rangle - \mu^2 \\ &= \sum_{n=0}^{\infty} n(n-1) \frac{e^{-\mu} \mu^n}{n!} + \mu - \mu^2 \\ &= \sum_{n=0}^{\infty} \frac{e^{-\mu} \mu^{n-2}}{(n-2)!} \mu^2 + \mu - \mu^2 \\ &= 0 + 0 + \sum_{(n-2)=0}^{\infty} \frac{e^{-\mu} \mu^{n-2}}{(n-2)!} \mu^2 + \mu - \mu^2 \\ &= \mu^2 + \mu - \mu^2 = \mu\end{aligned}$$

Thus:

3. **Fano factor**<sup>1</sup>.  $\frac{\mathbb{V}[N_\lambda[s, t]]}{\mathbb{E}[N_\lambda[s, t]]} = 1.$

---

<sup>1</sup>Note that this ratio (unlike the CV that we will encounter later) is only dimensionless for counting processes, or other dimensionless random variables.

## ISI distribution

The next few properties relate to the inter-spike interval (ISI) statistics.

First, it is fairly straightforward that, since the counting processes before and after event  $t_i$  are independent, the times to the previous and following spikes are independent.

**4. ISI independence.**  $\forall i > 1, \quad t_i - t_{i-1} \perp t_{i+1} - t_i.$

The full distribution of ISIs can be found from the count distribution:

$$\begin{aligned} P[t_{i+1} - t_i \in [\tau, \tau + d\tau)] &= P[N_\lambda[t_i, t_i + \tau) = 0] \times P[N_\lambda[t_i + \tau, t_i + \tau + d\tau) = 1] \\ &= e^{-\lambda\tau} \lambda d\tau e^{-\lambda d\tau} \end{aligned}$$

taking  $d\tau \rightarrow 0$

$$= \lambda e^{-\lambda\tau} d\tau$$

**5. ISI distribution.**  $\forall i \geq 1, \quad t_{i+1} - t_i \sim \text{iid Exponential}[\lambda^{-1}].$

## ISI distribution

5. **ISI distribution.**  $\forall i \geq 1, \quad t_{i+1} - t_i \sim \text{iid Exponential}[\lambda^{-1}]$ .

From this it follows that

6. **Mean ISI.**  $\mathbb{E}[t_{i+1} - t_i] = \lambda^{-1}$

7. **Variance ISI.**  $\mathbb{V}[t_{i+1} - t_i] = \lambda^{-2}$

These two properties imply that the **coefficient of variation** (CV), defined as the ratio of the standard deviation to the mean, of the ISIs generated by an homogeneous Poisson process is 1.

## Joint density

Finally, consider the probability density of observing spike train  $\{t_1 \dots t_N\}$  in interval  $\mathcal{T}$ .

Spike times are independent of one another and arrive at a uniform rate. So:

$$\rho(t_1 \dots t_N) dt_1 \dots dt_N = P[N \text{ spikes in } \mathcal{T}] \times P[i\text{th spike} \in [t_i, t_i + dt_i)] \times$$

[# of equivalent spike orderings]

The first term is given by the Poisson distribution, the second by the uniform distribution of spike times conditioned on  $N$ , and the third is  $N!$ , giving us

$$\begin{aligned} \rho(t_1 \dots t_N) dt_1 \dots dt_N &= \frac{(\lambda T)^N e^{-\lambda T}}{N!} \cdot \frac{dt_1}{T} \dots \frac{dt_N}{T} \cdot N! \\ &= \lambda^N e^{-\lambda T} dt_1 \dots dt_N \end{aligned}$$

We will see another way to write down this same expression while considering the inhomogeneous Poisson process below.



## Inhomogeneous Poisson Process: $N_{\lambda(t)}(t)$

The inhomogeneous Poisson process generalizes the constant event-arrival rate  $\lambda$  to a time-dependent one,  $\lambda(t)$ , while preserving the assumption of independent spike arrival times.

We can quickly summarize the properties of the inhomogeneous process by reference to the homogeneous one.

## Inhomogeneous Poisson Process: $N_{\lambda(t)}(t)$

The inhomogeneous Poisson process generalizes the constant event-arrival rate  $\lambda$  to a time-dependent one,  $\lambda(t)$ , while preserving the assumption of independent spike arrival times.

We can quickly summarize the properties of the inhomogeneous process by reference to the homogeneous one.

To begin, the two defining properties (this time we just state the Poisson distribution property directly.)

1. **Independence.** For all disjoint intervals  $[s, t)$  and  $[s', t')$ ,  $N_{\lambda(t)}[s, t) \perp N_{\lambda(t)}[s', t')$ .
2. **Count distribution.**  $N_{\lambda(t)}[s, t) \sim \text{Pois}[\int_s^t d\xi \lambda(\xi)]$ .

## Inhomogeneous Poisson Process: $N_{\lambda(t)}(t)$

The inhomogeneous Poisson process generalizes the constant event-arrival rate  $\lambda$  to a time-dependent one,  $\lambda(t)$ , while preserving the assumption of independent spike arrival times.

We can quickly summarize the properties of the inhomogeneous process by reference to the homogeneous one.

To begin, the two defining properties (this time we just state the Poisson distribution property directly.)

1. **Independence.** For all disjoint intervals  $[s, t)$  and  $[s', t')$ ,  $N_{\lambda(t)}[s, t) \perp N_{\lambda(t)}[s', t')$ .
2. **Count distribution.**  $N_{\lambda(t)}[s, t) \sim \text{Pois}[\int_s^t d\xi \lambda(\xi)]$ .

The variance in the counts is simply a consequence of the Poisson counting distribution, and so the next property follows directly.

3. **Fano factor.**  $\frac{\mathbb{V}[N_{\lambda(t)}[s, t)]}{\mathbb{E}[N_{\lambda(t)}[s, t)]} = 1$ .

## ISI distribution

Independence of counting in disjoint intervals means that ISIs remain independent.

4. **ISI independence.**  $\forall i > 1, \quad t_i - t_{i-1} \perp t_{i+1} - t_i.$

## ISI distribution

Independence of counting in disjoint intervals means that ISIs remain independent.

4. **ISI independence.**  $\forall i > 1, \quad t_i - t_{i-1} \perp t_{i+1} - t_i.$

The full distribution of ISIs is found in a similar manner to that of the homogeneous process distribution:

$$\begin{aligned} P[t_{i+1} - t_i \in [\tau, \tau + d\tau]] &= P[N_{\lambda(t)}(t_i, t_i + \tau) = 0] P[N_{\lambda(t)}[t_i + \tau, t_i + \tau + d\tau] = 1] \\ &= e^{-\int_{t_i}^{t_i+\tau} \lambda(\xi) d\xi} e^{-\int_{t_i+\tau}^{t_i+\tau+d\tau} \lambda(\xi) d\xi} \int_{t_i+\tau}^{t_i+\tau+d\tau} \lambda(\xi) d\xi \end{aligned}$$

taking  $d\tau \rightarrow 0$

$$\begin{aligned} &= e^{-\int_{t_i}^{t_i+\tau} \lambda(\xi) d\xi} e^{-\lambda(t_i+\tau)d\tau} \lambda(t_i + \tau) d\tau \\ &= e^{-\int_{t_i}^{t_i+\tau} \lambda(\xi) d\xi} \lambda(t_i + \tau) d\tau \end{aligned}$$

5. **ISI distribution.**  $\forall i \geq 1, \quad p(t_{i+1} - t_i) = e^{-\int_{t_i}^{t_{i+1}} \lambda(\xi) d\xi} \lambda(t_{i+1}).$

As the ISI distribution is not *iid* it is not very useful to consider its mean or variance.

## Joint density

The probability density of the event  $\{t_1 \dots t_N\}$  can be derived by setting the count in intervals **between** spikes to 0, and the count in an interval **around**  $t_i$  to 1. This gives

$$\begin{aligned} p(t_1 \dots t_N) dt_1 \dots dt_N &= P[N[0, t_1) = 0] \times P[N[t_1, t_1 + dt_1) = 1] \times \dots \times P[N(t_N, T) = 0] \\ &= e^{\int_0^{t_1} \lambda(\xi) d\xi} \times \lambda(t_1) dt_1 \times \dots \times e^{\int_{t_N}^T \lambda(\xi) d\xi} \\ &= e^{-\int_0^T \lambda(\xi) d\xi} \prod_{i=1}^N \lambda(t_i) dt_1 \dots dt_N \end{aligned}$$

Setting  $\lambda(t) = \lambda$  gives us the result for the homogeneous process.

## Time rescaling

Finally, we derive an additional important property of the inhomogeneous process. Let us rewrite the density above, by changing variables from  $t$  to  $u$  according to

$$u(t) = \int_0^t \lambda(\xi) d\xi \quad \text{i.e.} \quad u_i = \int_0^{t_i} \lambda(\xi) d\xi$$

Then

$$\begin{aligned} p(u_1 \dots u_n) &= p(t_1 \dots t_n) / \prod_i \frac{du_i}{dt_i} \\ &= e^{-u(T)} \prod_{i=1}^N \lambda(t_i) / \prod_{i=1}^N \lambda(t_i) \\ &= e^{-u(T)} \end{aligned}$$

Comparing this to the density for a homogeneous Poisson process shows that the variables  $u_i$  are distributed according to a homogeneous Poisson process with mean rate  $\lambda = 1$ .

This result is called **time rescaling**, and is central to the study of point processes in time.

## Self-exciting point processes

A self-exciting process has an intensity function which is conditioned on past events

$$\lambda(t) \rightarrow \lambda(t|N(t), t_1 \dots t_{N(t)})$$

It will be helpful to define the notation  $H(t)$  to represent the event *history* at time  $t$ —representing both  $N(t)$  and the times of the corresponding events. Then the self-exciting intensity function can be written  $\lambda(t|H(t))$ .



## Self-exciting point processes

A self-exciting process has an intensity function which is conditioned on past events

$$\lambda(t) \rightarrow \lambda(t|N(t), t_1 \dots t_{N(t)})$$

It will be helpful to define the notation  $H(t)$  to represent the event *history* at time  $t$ —representing both  $N(t)$  and the times of the corresponding events. Then the self-exciting intensity function can be written  $\lambda(t|H(t))$ .

This is actually the most general form of a point process – we can re-express any (conditionally orderly) point process in this form. To see this, consider the point process to be the limit as  $\Delta \rightarrow 0$  of a binary time series  $\{b_1, b_2, \dots, b_{T/\Delta}\}$  and note that

$$P(b_1, b_2, \dots, b_{T/\Delta}) = \prod_i P(b_i | b_{i-1} \dots b_1)$$

## Renewal processes

If the intensity of a self-exciting process depends only on the time since the last spike, i.e.

$$\lambda(t|H(t)) = \lambda(t - t_{N(t)})$$

then the process is called a **renewal** process. ISIs from a renewal process are iid and so we could equivalently have defined the process by its ISI density. This gives an (almost) easy way to write the probability of observing the event  $\{t_1 \dots t_N\}$  in  $T$ . Suppose, for simplicity, that there was an event at  $t_0 = 0$ . Then, if the ISI density is  $\rho(\tau)$ :

$$\rho(t_1 \dots t_N) dt_1 \dots dt_N = \prod_{i=1}^N \rho(t_i - t_{i-1}) \left(1 - \int_0^{T-t_N} d\tau \rho(\tau)\right)$$

The last term gives the probability that no more spikes are observed after  $t_N$ . If had not assumed that there was a spike at 0 we would have needed a similar term at the front.

The conditional intensity (sometimes called **hazard function**) for the renewal process defined by its ISI density  $\rho(\tau)$  is

$$\lambda(t|t_{N(t)}) dt = \frac{\rho(t - t_{N(t)})}{1 - \int_0^{t-t_{N(t)}} d\tau \rho(\tau)} dt$$

which is indeed a function only of  $t - t_{N(t)}$ .

## Gamma-interval process

The specific choice of the **gamma-interval process** with

$$t_{i+1} - t_i \stackrel{\text{iid}}{\sim} \text{Gamma}[\alpha, \beta]$$

where

$$\tau \sim \text{Gamma}[\alpha, \beta] \Rightarrow p(\tau) = \frac{\beta^\alpha}{\Gamma(\alpha)} \tau^{\alpha-1} e^{-\beta\tau}$$

is an important renewal process in theoretical neuroscience, because the ISI distribution has a refractory-like component.

A homogeneous Poisson process is a gamma-interval process (and therefore a renewal process) with  $\alpha = 1$ . The parameter  $\alpha$  is sometimes called the order or the shape-parameter of the gamma-interval process. Larger values of  $\alpha$  shape the polynomial rising part of the Gamma density, thus implementing a relative refractory period. The long-time behaviour is dominated by the exponential decay with coefficient  $\beta$ .

You might wish to show that a gamma-interval process of integral order  $\alpha$  can be constructed by selecting every  $\alpha$ th event from a homogeneous Poisson process.

## Inhomogeneous renewal processes

In an inhomogeneous renewal processes, the rate depends both on time since the last spike and on the current time.

$$\lambda(t) \rightarrow \lambda(t, t - t_{N(t)})$$

Called “inhomogeneous Markov interval” processes by Kass and Ventura (2000).

Two popular ways to construct an inhomogeneous renewal process:

1. **Time-rescaling.** Given unit-mean ISI density  $\rho(\tau)$ , and time-varying intensity  $\rho(t)$ , define:

$$\rho(t_1 \dots t_N) dt_1 \dots dt_N = \prod_{i=1}^N \rho\left(\int_{t_{i-1}}^{t_i} \rho(\xi) d\xi\right) \left(1 - \int_0^{\int_{t_N}^T \rho(\xi) d\xi} d\tau \rho(\tau)\right)$$

2. **Spike-response.**

$$\lambda(t, t - t_{N(t)}) = f(\rho(t), h(t - t_{N(t)}))$$

for a simple  $f$ . Often,  $f$  just multiplies the two functions (or, equivalently, adds log intensities).

The term “spike-response” comes from Gerstner, who uses such spike-triggered currents to create a potentially more tractable approximation to an integrate-and-fire neuron.

These definitions differ in how ISI density depends on  $\rho$ . **Rescaling:** higher rates make time pass faster, so ISI interactions are rescaled. **Spike-response:** a refractory  $h$  may not suppress spikes as well at higher rates, but the duration of influence does not change.

## General Spike-Response processes

This category of processes has come to be used with increasing frequency recently, particularly in a generalised-linear form.

The product form of spike-response renewal process can be written:

$$\lambda(t, t - t_{N(t)}) = e^{\rho(t) + h(t - t_{N(t)})}$$

and then generalised to include influence from all (or  $> 1$ ) past spikes:

$$\lambda(t|H(t)) = e^{\rho(t) + \sum_j h(t - t_{N(t)-j})}$$

Often, we wish to estimate the parameters of a point-process model from spike data. Assuming a generalised linear form makes this easier. Write history influence  $h$  in terms of a basis of fixed functions  $h_i(\tau)$ :

$$\lambda(t|H(t)) = e^{\rho(t) + \sum_{ij} \alpha_j h_i(t - t_{N(t)-j})}$$

If  $\rho(t)$  is also written as linear function of external covariates, then the complete model can be fit by the standard methods used for generalised linear models (GLMs: note, a different use of this abbreviation to the commonly used models for fMRI data).

## Doubly stochastic point processes (DSPP)

Some variability may originate outside the spiking process itself. Modelled, by having intensity depend on a random variable  $\sigma$  (or random process  $\sigma(t)$ ).

(Inhomogeneous) Poisson conditioned on  $\sigma \Rightarrow$  Cox process.

- ▶ Random scale:

$$\sigma \sim \text{Gamma}(\alpha, \beta); \quad \lambda(t; \sigma) = \sigma \rho(t)$$

Used to model evoked response  $\rho(t)$  modulated by cortical excitability.

- ▶ Log-normal Cox process:

$$\sigma(t) \sim \mathcal{GP}(\mu(t), K(t, t')); \quad \lambda(t; \sigma) = e^{\sigma(t)} + \rho(t)$$

Used to capture slow fluctuations in spontaneous activity [ $\rho(t) = 0$ ] or in noise around evoked response. ( $\rho$  may appear in exponent instead  $\Rightarrow$  fluctuating scale.)

- ▶ Counting process for Cox process events has Fano factor  $> 1$ . For Gamma-scaled Poisson distribution is negative binomial.
- ▶ DSPP models also provide a useful way to introduce dependencies between two or more point processes, through correlations in the intensity functions (e.g. GPFA using Cox processes).

## Doubly stochastic or self exciting?

- ▶ Recall that *any* point process can be written in self-exciting form.

## Doubly stochastic or self exciting?

- ▶ Recall that *any* point process can be written in self-exciting form.
- ⇒ in principle, dependence on  $\sigma(t)$  can be captured by dependence on  $H(t)$  instead.



## Doubly stochastic or self exciting?

- ▶ Recall that *any* point process can be written in self-exciting form.
- ⇒ in principle, dependence on  $\sigma(t)$  can be captured by dependence on  $H(t)$  instead.
- ▶ Without separate access to  $\sigma$  we could choose either form.

## Doubly stochastic or self exciting?

- ▶ Recall that *any* point process can be written in self-exciting form.
- ⇒ in principle, dependence on  $\sigma(t)$  can be captured by dependence on  $H(t)$  instead.
- ▶ Without separate access to  $\sigma$  we could choose either form.
- ▶ The doubly stochastic view is sometimes preferred:

## Doubly stochastic or self exciting?

- ▶ Recall that *any* point process can be written in self-exciting form.
- ⇒ in principle, dependence on  $\sigma(t)$  can be captured by dependence on  $H(t)$  instead.
- ▶ Without separate access to  $\sigma$  we could choose either form.
- ▶ The doubly stochastic view is sometimes preferred:
  - ▶ May provide a cleaner representation of the structure ( $\rho(\sigma)$  and  $\lambda(t; \sigma)$  may be simple, while  $\lambda(t|H(t))$  is not.), thus simplifying analysis or calculations.

## Doubly stochastic or self exciting?

- ▶ Recall that *any* point process can be written in self-exciting form.
- ⇒ in principle, dependence on  $\sigma(t)$  can be captured by dependence on  $H(t)$  instead.
- ▶ Without separate access to  $\sigma$  we could choose either form.
- ▶ The doubly stochastic view is sometimes preferred:
  - ▶ May provide a cleaner representation of the structure ( $p(\sigma)$  and  $\lambda(t; \sigma)$  may be simple, while  $\lambda(t|H(t))$  is not.), thus simplifying analysis or calculations.
  - ▶ May provide an more interpretable model: separates spike-mechanism effects (refractoriness, bursting . . . ) from variable input.

## Doubly stochastic or self exciting?

- ▶ Recall that *any* point process can be written in self-exciting form.
- ⇒ in principle, dependence on  $\sigma(t)$  can be captured by dependence on  $H(t)$  instead.
- ▶ Without separate access to  $\sigma$  we could choose either form.
- ▶ The doubly stochastic view is sometimes preferred:
  - ▶ May provide a cleaner representation of the structure ( $\rho(\sigma)$  and  $\lambda(t; \sigma)$  may be simple, while  $\lambda(t|H(t))$  is not.), thus simplifying analysis or calculations.
  - ▶ May provide an more interpretable model: separates spike-mechanism effects (refractoriness, bursting . . . ) from variable input.
- ▶ Indeed, may be helpful for modelling to consider doubly-stochastic renewal, IMI or other non-Poisson processes, even though these are theoretically redundant.

## Doubly stochastic or self exciting?

- ▶ Recall that *any* point process can be written in self-exciting form.
- ⇒ in principle, dependence on  $\sigma(t)$  can be captured by dependence on  $H(t)$  instead.
- ▶ Without separate access to  $\sigma$  we could choose either form.
- ▶ The doubly stochastic view is sometimes preferred:
  - ▶ May provide a cleaner representation of the structure ( $\rho(\sigma)$  and  $\lambda(t; \sigma)$  may be simple, while  $\lambda(t|H(t))$  is not.), thus simplifying analysis or calculations.
  - ▶ May provide an more interpretable model: separates spike-mechanism effects (refractoriness, bursting . . . ) from variable input.
- ▶ Indeed, may be helpful for modelling to consider doubly-stochastic renewal, IMI or other non-Poisson processes, even though these are theoretically redundant.
- ▶ Exact choice of model separation between self-excitation and external drive can strongly influence scientific interpretation (cf. later discussion of correlation functions).

## Joint Models

- ▶ 2D point process is not correct model

## Joint Models

- ▶ 2D point process is not correct model
- ▶ superimposed processes



## Joint Models

- ▶ 2D point process is not correct model
- ▶ superimposed processes
- ▶ infinitely divisible Poisson process

## Joint Models

- ▶ 2D point process is not correct model
- ▶ superimposed processes
- ▶ infinitely divisible Poisson process
- ▶ multivariate self-exciting process

## Joint Models

- ▶ 2D point process is not correct model
- ▶ superimposed processes
- ▶ infinitely divisible Poisson process
- ▶ multivariate self-exciting process
  - ▶ linear dependence – Hawkes process

## Joint Models

- ▶ 2D point process is not correct model
- ▶ superimposed processes
- ▶ infinitely divisible Poisson process
- ▶ multivariate self-exciting process
  - ▶ linear dependence – Hawkes process
  - ▶ log-linear dependence – GLM (or generalised Hawkes process)

## Joint Models

- ▶ 2D point process is not correct model
- ▶ superimposed processes
- ▶ infinitely divisible Poisson process
- ▶ multivariate self-exciting process
  - ▶ linear dependence – Hawkes process
  - ▶ log-linear dependence – GLM (or generalised Hawkes process)
- ▶ doubly stochastic processes

## Joint Models

- ▶ 2D point process is not correct model
- ▶ superimposed processes
- ▶ infinitely divisible Poisson process
- ▶ multivariate self-exciting process
  - ▶ linear dependence – Hawkes process
  - ▶ log-linear dependence – GLM (or generalised Hawkes process)
- ▶ doubly stochastic processes
  - ▶ Common input to log-linear spike-response models

## Joint Models

- ▶ 2D point process is not correct model
- ▶ superimposed processes
- ▶ infinitely divisible Poisson process
- ▶ multivariate self-exciting process
  - ▶ linear dependence – Hawkes process
  - ▶ log-linear dependence – GLM (or generalised Hawkes process)
- ▶ doubly stochastic processes
  - ▶ Common input to log-linear spike-response models
  - ▶ Gaussian process factor analysis

## Measuring point processes

Event data  $\Rightarrow$  point process generative models. What about measurement?

Consider spike trains from repeated experiments under (as far as possible) constant experimental conditions.

$$s^{(k)}(t) = \sum_{i=1}^{N^{(k)}} \delta(t - t_i^{(k)}) \quad (\text{trials } k = 1 \dots K)$$

How to characterise  $s^{(k)}(t)$ , and relationship to stimulus (or task)?

- ▶ Parametric point-process model, possibly dependent on stimulus  $a(t)$ :

$$s^{(k)}(t) \sim \lambda \left( t, a[0, t], N^{(k)}(t), t_1^{(k)}, \dots, t_{N^{(k)}(t)}^{(k)}, \theta \right).$$

**Encoding:** stimulus-response function. Discussed later.

- ▶ Construct an algorithm to estimate  $a(t)$  from  $s^{(k)}(t)$ :

$$\hat{a}(t) = F[s^{(k)}[0, t]].$$

**Decoding:** what does the neuron say about the world? (Not always causal). Also discussed later.

- ▶ Estimate nonparametric features (usually **moments**) of the distribution of  $s^{(k)}(t)$ .



## Mean intensity and PSTH

Simplest non-parametric characterisation of spike process is with **mean intensity**:

$$\bar{\lambda}(t) = \langle s(t) \rangle = \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{k=1}^K s^{(k)}(t)$$

*Not* the intensity function for the point process (unless Poisson) – *marginalised* over history.

$$\bar{\lambda}(t, a(\cdot)) \equiv \int dN(t) \int dt_1 \dots dt_{N(t)} p(t_1 \dots t_{N(t)}) \lambda(t, a(\cdot), N(t), t_1, \dots, t_{N(t)})$$

For finite  $K$ , estimating  $\bar{\lambda}$  by summing  $\delta$ -functions yields spiky results. Instead, histogram:

$$\widehat{\bar{N}}[t, t + \Delta t] = \frac{1}{K} \sum_{k=1}^K N^{(k)}[t, t + \Delta t]$$

This is called the Post- (or Peri-) Stimulus-Time Histogram  $\Rightarrow$  **PSTH**.

## Smoothing the PSTH

If we expect  $\bar{\lambda}(t)$  to be smooth, could use kernel  $\phi(\tau)$ :

$$\widehat{\bar{\lambda}}(t) = \frac{1}{K} \sum_{k=1}^K \int d\tau \phi(\tau) s^{(k)}(t - \tau)$$

Resembles kernel density estimation (without normalisation).

Width of  $\phi$  could be chosen adaptively, depending on local density of spikes. Note: sampling a smoothed function makes *much* more sense than smoothing a binned histogram!

Alternatively, can impose a smooth prior (e.g. GP) on time-varying aspect of intensity:  $\lambda(t)$  for inhomog Poisson, or (say)  $\rho(t)$  for inhomog gamma-interval of order  $\gamma$ :

$$\rho \sim \mathcal{N}(\mu \mathbf{1}, K_{\theta})$$

$$\rho(t_1 \dots t_N | \rho) = \prod_{i=1}^N \left[ \frac{\gamma x_{t_i}}{\Gamma(\gamma)} \left( \gamma \sum_{j=t_{i-1}}^{t_i-1} \rho_j \Delta \right)^{\gamma-1} \exp \left\{ -\gamma \sum_{j=t_{i-1}}^{t_i-1} \rho_j \Delta \right\} \right]$$

Posterior on  $\rho(t)$  can be found by approximate inference (Laplace/EP).

## Autocorrelation

The autocorrelation function for a process that generates spike trains  $s(t)$  is:

$$R_{ss}(\tau) = \left\langle \frac{1}{T} \int dt s(t)s(t-\tau) \right\rangle$$

where  $\langle \cdot \rangle$  is expectation wrt to random draws of  $s(t)$  from the process.

**Time-averaged** local second moment of joint on  $s(t)$ ;  $\bar{\lambda}(t)$  was the (non-time-averaged) first moment.

Note that, since  $s(t)$  is a sum  $\delta$  functions,  $R_{ss}(0) = \infty$  under this definition.

Alternatively, could define  $R_{ss}$  as time-averaged conditional first moment  $\Rightarrow$  mean intensity at  $t + \tau$ , conditioned on event at  $t$ , averaged over  $t$ .

$$R_{ss}^{alt}(\tau) = \frac{1}{T} \int dt \langle \lambda(t+\tau | \exists i : t_i = t) \rangle,$$

where  $\langle \cdot \rangle$  is expectation with respect to  $N(T)$  and  $t_{j \neq i}$ . Now  $R_{ss}^{alt}(0) = 0$ .

We will stick to the first (i.e., second moment) definition.

## Autocovariance

Using the identity  $\langle x^2 \rangle = \langle (x - \mu)^2 \rangle + \mu^2$ , we can decompose the autocorrelation function:

$$R_{ss}(\tau) = \bar{\lambda}^2 + \frac{1}{T} \int dt (\bar{\lambda}(t) - \bar{\lambda})(\bar{\lambda}(t - \tau) - \bar{\lambda}) + \underbrace{\left\langle \frac{1}{T} \int dt (s(t) - \bar{\lambda}(t))(s(t - \tau) - \bar{\lambda}(t - \tau)) \right\rangle}_{Q_{ss}(\tau)}$$

where  $\bar{\lambda}$  is the time-averaged mean rate.  $Q_{ss}(\tau)$  is called the **autocovariance** function.

[D&A call  $Q_{ss}$  the autocorrelation function; in the experimental literature, estimates of  $Q_{ss}$  are usually called “shift-” or “shuffle-corrected autocorrelograms” (see next slide).]

- ▶ For an (inhomogeneous) Poisson process  $Q_{ss}(\tau) = \delta(\tau)$ , by independence.
- ▶ For a general self-exciting process,  $Q_{ss}(\tau)$  gives (to second order) dependence on nearby spike times.
- ▶ Often used to look for oscillatory structure in spike trains (where spikes tend to repeat around fixed intervals, but at random phase wrt stimulus) or similar spike-timing relationships.

## Estimating correlation functions

- ▶ Correlation functions are typically estimated by constructing **correlograms**: histograms of time *differences* between (not necessarily adjacent) spikes.
- ▶ Covariance function is estimated by subtracting an estimate of the correlation of the mean intensity:

$$\begin{aligned} & \frac{1}{T} \int dt \widehat{\lambda}(t) \widehat{\lambda}(t - \tau) \\ &= \frac{1}{TK^2} \int dt \sum_k s^{(k)}(t) \sum_{k'} s^{(k')} (t - \tau) = \frac{1}{TK^2} \sum_{kk'} \int dt s^{(k)}(t) s^{(k')} (t - \tau) \end{aligned}$$

'Shift' or 'shuffle' correction.

- ▶ May also be constructed in frequency domain: power spectrum, spectrogram, coherence (for multiple processes). Usually based on FT of binary-binned spike trains.

## Interpreting covariance functions

$$Q_{ss}(\tau) = \left\langle \frac{1}{T} \int dt (s(t) - \bar{\lambda}(t))(s(t - \tau) - \bar{\lambda}(t - \tau)) \right\rangle$$

- ▶ (Auto)covariance captures coordination in spike timing around  $\bar{\lambda}(t)$ . For a self-exciting point process this is related to the history-conditioned intensity  $\lambda(t|H(t))$ .
- ▶ But *any* point process can be written in self-exciting form, so *any* non-Poisson process has non- $\delta$  autocovariance.
- ▶ Consider a doubly-stochastic (non-Poisson) point process, with intensity and conditional mean intensity:

$$\lambda(t|H(t); \sigma); \quad \bar{\lambda}(t; \sigma) = \langle \lambda(t|H(t); \sigma)_{H(t)|\sigma} \rangle$$

dependent on a random variable  $\sigma$ . We have  $\bar{\lambda}(t) = \langle \bar{\lambda}(t; \sigma) \rangle_{\sigma}$ . Now,

$$Q_{ss}(\tau) = \left\langle \frac{1}{T} \int dt (s(t) - \bar{\lambda}(t; \sigma))(s(t - \tau) - \bar{\lambda}(t - \tau; \sigma)) \right\rangle \\ + \left\langle \frac{1}{T} \int dt (\bar{\lambda}(t; \sigma) - \bar{\lambda}(t))(\bar{\lambda}(t - \tau; \sigma) - \bar{\lambda}(t - \tau)) \right\rangle$$

So covariance can arise from local spike time interactions *or* from influence of  $\sigma$ .

- ▶ Covariance in  $\bar{\lambda}(t; \sigma)$  can be localised in time if influence of  $\sigma$  is localised or if it is slow (e.g. neuromodulation / excitability/ latency shifts) but  $\bar{\lambda}(t)$  varies rapidly  $\Rightarrow$  narrow correlogram peaks do not necessarily indicate spike-time interactions.

## Multiple spike trains

Often, we may be interested in simultaneously modelling responses from many cells.

If no two processes can generate events at precisely the same time (a form of conditional orderliness), or if simultaneous spiking events are independent, then dependences between the processes generally by dependence on all previous events in all cells:

$$\lambda^{(c)}(t) \rightarrow \lambda^{(c)}\left(t \mid N^{(c)}(t), t_1^{(c)}, \dots, t_{N^{(c)}(t)}^{(c)}, \{N^{(c')}(t), t_1^{(c')}, \dots, t_{N^{(c')}(t)}^{(c')}\}\right)$$

This is analogous to the self-exciting point process intensity function.

Dependencies can also be expressed by other forms, for example by DSPPs with the latent random process shared (or correlated) between cells. Such representations may often be more natural or causally accurate.

## Cross-correlations

Time averaged second-order statistic of interaction between spike times in different cells:

$$R_{s^{(c)}s^{(c')}}(\tau) = \left\langle \frac{1}{T} \int dt s^{(c)}(t)s^{(c')}(t-\tau) \right\rangle.$$

Much of the discussion of autocorrelations carries over:

- ▶ Separates into cross-correlation of mean intensities and cross-covariance function:

$$Q_{s^{(c)}s^{(c')}}(\tau) = \left\langle \frac{1}{T} \int dt (s^{(c)}(t) - \bar{\lambda}^{(c)}(t))(s^{(c')}(t-\tau) - \bar{\lambda}^{(c')}(t-\tau)) \right\rangle.$$

- ▶ Crosscorrelation estimated by correlogram: histogram of intervals from spikes in one cell to spikes in the other.
- ▶ Shift- or shuffle-correction used to obtain estimate of cross-covariance.
- ▶ May also be obtained in the frequency domain as cross-spectrum or coherence.



## Interpreting cross-covariance functions

Much of the discussion of auto-covariance interpretation carries over.

- ▶  $s^{(c)}(t)$  independent of  $s^{(c')}(t')$  for all  $t, t' \Rightarrow Q_{s^{(c)}s^{(c')}}(\tau) = 0$ .
- ▶ Peaks (offset from  $\tau = 0$ ) may reflect direct effect of spikes in one cell on another (e.g. synaptic input); but such patterns are rare.
- ▶ Peaks centered on  $\tau = 0$  may reflect precise synchronisation; perhaps driven by common pre-synaptic spiking or by population-wide oscillation.
- ▶ But consider DSPP models with  $\lambda^{(c)}(t; \sigma)$  and  $\lambda^{(c')}(t; \sigma)$  depending on the same (or correlated) random variable:

$$Q_{s^{(c)}s^{(c')}}(\tau) = \left\langle \frac{1}{T} \int dt (s^{(c)}(t) - \bar{\lambda}^{(c)}(t; \sigma))(s^{(c')}(t - \tau) - \bar{\lambda}^{(c')}(t - \tau; \sigma)) \right\rangle \\ + \left\langle \frac{1}{T} \int dt (\bar{\lambda}^{(c)}(t; \sigma) - \bar{\lambda}^{(c)}(t))(\bar{\lambda}^{(c')}(t - \tau; \sigma) - \bar{\lambda}^{(c')}(t - \tau)) \right\rangle$$

Covariance can arise from correlations in  $\bar{\lambda}(t; \sigma)$  and can be temporally sharp even if influence of  $\sigma$  is slow: consider common scale modulation of sharp stimulus response; or common shifts in latency.

- ▶ As with autocorrelation, estimating biophysical significance from correlation structure is not straightforward.