Neural Coding

(Incomplete notes)

Maneesh Sahani
Gatsby Computational Neuroscience Unit

October 22, 2010

Chapter 1

Introduction

- The brain manipulates information by combining and generating action potentials (or spikes).
- It seems natural to ask how information (about sensory variables; inferences about the world; action plans; cognitive states ...) is represented in spike trains.
- Experimental evidence comes largely from sensory settings — can repeat the same stimulus (although this does not actually guarantee that all information represented is identical, but some is likely to be shared across trials).
  - 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

Chapter 2

The Statistics of Spike Trains

In this lecture we will cover:

- descriptions of spike trains
- point processes
- the homogeneous Poisson process
- the inhomogeneous Poisson process
- general point processes

2.1 Spike Trains

The timecourse of every action potential (AP) in a cell measured at the soma might not be identical, due to differences in the open channel configuration. However, axons tend to contain only the Na^+ and K^+ channels needed for AP propagation, and therefore exhibit little or no timecourse variation. Further, there is no experimental evidence I am aware of to indicate that AP shape affects vesicle release.

Thus, from the point of view of inter-neuron communication the only thing that matters about an AP or spike is its time of occurance.

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

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

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

$$s(t) = \sum_{i=1}^{N} \delta(t - t_i)$$

(D&A call this $\rho(t)$) or using a counting function,

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

(where $\rightarrow t$ in the limit indicates that t is not included in the integral: this might seem a little counter-intuitive, but will match a later definition),

or as a vector by discretizing with time step Δt

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

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

2.2 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 arises in more than one way.

- **Noise**. Perhaps due to vesicle release; or thermal noise in conductances.
- **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 two contributions.

Note that everything about the spike train can be variable, even the spike count on the ith repetition (or "trial") $N_i = \int_0^T d\xi \; s_i(\xi)$

Variability in N_i is on order of the mean.

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

All this requires that we be able to treat spike trains statistically.

2.3 Point Processes

A probabilistic process that produces events of the type

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

is called a **point process**. Clearly this is the statistical object best suited for the description of spike trains. Every point process is associated with a dual **counting process** which produces events of the type

$$\begin{aligned} N(t) \text{ such that } & N(t) \geq 0 \\ & N(t + \Delta t) \geq N(t) \\ & N(t) - N(s) = N[s, t) \in \mathbb{Z} \end{aligned}$$

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

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

$$\left[\begin{array}{l}\text{Recall that the Poisson distribution is a distribution on an integer random variable} \\ n \geq 0. \text{ If } n \sim \mathsf{Poiss}[\mu] \text{ then } \mathsf{P}[n] = \frac{\mu^n e^{-\mu}}{n!}. \text{ The parameter } \mu \text{ is the mean of the} \\ \text{distribution.}\end{array}\right]$$

The most basic point process is called the homogeneous Poisson process. It is parameterised by a single scalar λ which gives the mean rate with which events arrive. Each event is completely independent of all others. Formally, we can define it by way of the asociated counting process, $N_\lambda(t)$, by imposing two conditions:

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

There are two ways to write the second condition. If we assume that $\lim_{ds\to 0} N_\lambda[s,s+ds) \in \{0,1\}$ (technically called conditional orderliness – at most one event occurs at one time) then it is sufficient to assume that

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

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

Consider dividing the interval $[s,t)$ into M bins of length Δ (i.e. $M = (t-s)/\Delta$). If each bin is small enough (we will take the limit $\Delta \to 0$ later) conditional orderliness tells us that the count of spikes in each bin is binary. For a binary random variable, the expected value is the same as the probability of the variable taking the value 1, so we can assume that the expectation $\lambda\Delta$ gives the *probability* of a spike in each interval. Then the distribution of the number of spikes in the whole interval is given by the binomial distribution

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

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 $\Delta \to 0$ or, equivalently, $M \to \infty$

$$\begin{aligned}
&= \frac{\mu^n}{n!}1^n 1^n e^{-\mu} \\
&= \frac{\mu^n e^{-\mu}}{n!}
\end{aligned}$$

So the spike count in any interval is Poisson distributed. This is where the name of the process comes from. 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.** $N_\lambda[s,t) \sim \mathsf{Poiss}[(t-s)\lambda]$.

We now derive a number of properties of the homogeneous Poisson process that will be important.

First, the variance of the count distribution (this is really a property of the Poisson distribution). We can write

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

Thus we have the third property of the homogeneous Poisson process:

3. **Fano factor**[1]**.** $\dfrac{\mathcal{V}ar\left[N_\lambda[s,t)\right]}{\mathcal{E}\left[N_\lambda[s,t)\right]} = 1.$

The next few properties relate to the inter-event (or, for neurons, inter-spike) interval (ISI) statistics. First, it is fairly straightforward to see that, since the counting processes before and after an event t_i are independent, the times to the previous and following spikes are independent from one another.

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:

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

taking $d\tau \to 0$

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

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

From this it follows that

6. **Mean ISI.** $\mathcal{E}\left[t_{i+1} - t_i\right] = \lambda^{-1}$
7. **Variance ISI.** $\mathcal{V}ar\left[t_{i+1} - t_i\right] = \lambda^{-2}$

[1]The term Fano Factor comes from semiconductor physics, where it actually means something slightly different. This use is standard in neuroscience. Note that this ratio (unlike the CV that we will encounter later) is only dimensionless for counting processes, or other dimensionless random variables.

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.

Finally, we write down the probability density of observing a spike train $\{t_1 \dots t_N\}$ in some interval T, from an homogeneous Poisson process. Recall that spike times are independent of one another and arrive at a uniform rate. This makes it possible to write down the relevant probability as a product of three terms:

$$p(t_1 \dots t_N)dt_1 \dots dt_N = \mathsf{P}\left[N \text{ spikes in } T\right] \cdot \mathsf{P}\left[i\text{th spike} \in [t_i, t_i + dt_i)\right] \cdot [\#\text{ of equivalent spike orderings}]$$

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

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

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

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

The inhomogeneous Poisson process generalizes the constant event-arrival rate λ to a time-dependent one, $\lambda(t)$, while preserving the assumption of independent spike arrival times. We will quickly summarize the properties of the inhomogeneous process by reference to the homogeneous one.

We begin with the two defining properties, although in this case we will 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.** $N_{\lambda(t)}[s, t) \sim \mathsf{Poiss}[\int_s^t d\xi\ \lambda(\xi)]$.

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

3. **Fano factor.** $\dfrac{\mathcal{V}ar\left[N_{\lambda(t)}[s, t)\right]}{\mathcal{E}\left[N_{\lambda(t)}[s, t)\right]} = 1.$

Also, independence of counting in disjoint intervals means that ISIs remain independent.

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

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

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

taking $d\tau \to 0$

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

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. Instead we pass directly to the probability density of the event $\{t_1 \dots t_N\}$ which 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

$$p(t_1 \dots t_N)dt_1 \dots dt_N = e^{-\int_0^T \lambda(\xi)d\xi} \prod_{i=1}^{N} \lambda(t_i)dt_1 \dots dt_N$$

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

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

Then

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

Comparison with 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.

2.6 Other Point Processes

2.6.1 Self-exciting point processes

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

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

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 $\Delta \to 0$ of a binary time series $\{b_1, b_2, \dots b_{T/\Delta}\}$ and note that

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

2.6.2 Renewal processes

If the intensity of a self-exciting process depends only 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 $p(\tau)$:

$$p(t_1 \dots t_N)\, dt_1 \dots dt_N = \prod_{i=1}^{N} p(t_i - t_{i-1}) \Big(1 - \int_0^{T-t_N} d\tau\; p(\tau)\Big)$$

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 $p(\tau)$ is

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

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

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

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

where

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

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 α is sometimes called the order or the shape-parameter of the gamma-interval process. Larger values of α 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 β.

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

2.6.3 Inhomogeneous renewal processes

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

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

These have been called "inhomogeneous Markov interval" processes by Kass and Ventura (2000).

There are two popular ways to construct an inhomogeneous renewal process, and the two constructions have different properties.

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

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

2. **Spike-response**. Choose

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

for a simple f. Most often, f simply multiplies the two functions (or, equivalently, acts additively on the logarithms). 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.

The substantial difference between these two approaches is in how the rate interacts with the ISI density. In the rescaling approach, higher rates mean that time passes faster. Thus any refractory-like element in the renewal ISI density becomes shorter at higher rates. By contrast, in the spike-response model, a refractory h may be slightly less effective at suppressing spikes at higher rates, but the duration of its 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 the spike-response renewal process can conveniently be written in exponential form:

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

and can by generalised to include influence from all (or a fixed number) of past spikes:

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

We often wish to estimate the parameters of a point process model from spike data. This is made easier by assuming a generalised linear form. In this case, the history term can be expressed that way by writing h in terms of a basis of fixed functions $h_i(\tau)$:

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

If $\rho(t)$ is written as a linear function of some 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).

2.6.4 Birth process

The intensity of a birth process depends only on the *number* of events so far:

$$\lambda(t) \to \lambda(t|N(t))$$

2.6.5 Doubly stochastic Poisson (or Cox) process

In the doubly stochastic process, or Cox process, $\lambda(t)$ is itself a random variable; or depends on another random process $x(t)$. An example is the randomly scaled IHPP:

$$\lambda(t) = s \cdot \rho(t) \qquad \text{with } \rho(t) \text{ fixed and } s \sim \mathsf{Gamma}(\alpha, \beta)$$

These models have been the subject of some recent attention, as a way to model a stimulus-dependent response $\rho(t)$ which is modulated by cortical excitability. The counting process for such a DSPP has a Fano factor > 1. DSPP models also provide a useful way to introduce dependencies between two or more point processes, through correlations in the intensity functions.

2.6.6 Joint Models

Placeholder section.

- 2D point process is not correct model
- superimposed processes
- infinitely divisible Poisson process
- multivariate self-exciting process
- log-linear spike-response models with interaction terms
- doubly stochastic processes
  - common input to log-linear spike-response models
  - Gaussian process factor analysis

Chapter 3

Measuring point processes

We now turn from the probabilistic theory of point processes to the question of how best to characterise a set measured of events. Suppose we repeatedly measure spike trains, $s(t)$, elicited from a single neuron under, as far as possible, constant experimental conditions. Let the kth measured spike train be

$$s^{(k)}(t) = \sum_{i=1}^{N^{(k)}} \delta(t - t_i^{(k)}).$$

We might take one of a number of approaches to characterising $s^{(k)}(t)$, and it's relationship to the experimental stimulus (or task):

- Construct a parametric model for the intensity of the point process, possibly dependent on the stimulus $a(t)$:
$$s^{(k)}(t) \sim \lambda\left(t, a[0,t), N^{(k)}(t), t_1^{(k)}, \ldots, t_{N^{(k)}(t)}^{(k)}, \theta\right),$$
thus characterising the stimulus-response function of the neuron. This is the **encoding** approach, to be discussed in a subsequent lecture. Here, we just note that the time-rescaling result discussed above provides a way to evaluate the goodness-of-fit of a point-process encoding model.
- Construct an algorithm to estimate $a(t)$ from $s^{(k)}(t)$:
$$\widehat{a}(t) = F[s^{(k)}[0,t)]$$
as accurately as possible. Not always causal. This **decoding** approach (also discussed in a later lecture) may be interpreted as asking what the neuron tells the animal about the outside world.
- Estimate nonparametric features (usually moments) of the distribution of $s^{(k)}(t)$. This is what we discuss below.

3.1 Mean response functions and the PSTH

The simplest non-parametric characterisation of the process that generated $s^{(k)}(t)$ is its **mean intensity**:

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

Note that this is *not* the intensity function for the point process, unless that process is Poisson. Instead it is the *marginal* intensity, obtained by integrating over all random variables besides time:

$$\overline{\lambda}(t, a(\cdot)) \stackrel{\text{def}}{=} \int dN(t) \int dt_1 \ \ldots dt_{N(t)} \ p\left(t_1 \ldots t_{N(t)}\right) \lambda\left(t, a(\cdot), N(t), t_1, \ldots, t_{N(t)}\right)$$

For finite K, estimating $\overline{\lambda}$ by summing δ-functions yields spiky results. Instead, we usually construct a histogram

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

This is called the Post- (or Peri-) Stimulus-Time Histogram or PSTH.

If we expect $\overline{\lambda}(t)$ to be a smooth function of time, we might instead use a kernel $\phi(\tau)$ to construct the estimate:

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

There has been some work (which we won't discuss here) on choosing the width of ϕ adaptively, possibly even in a way that depends on the local density of spikes. [Note: the literature contains many examples where a histogram is constructed first and then smoothed with a kernel. There is little theoretical justification for this practice, rather than sampling the smoothed kernel estimate given above.]

Alternatively, it is possible to use a smooth prior (e.g., derived from a suitable Gaussian Process) on a time-varying element of the intensity. This could be the intensity function itself for an inhomogeneous Poisson process, or else a time-varying element that combines with fixed spike-dependent elements to form the intensity (c.f. the function ρ in our discussion of inhomogeneous renewal processes).

3.2 Correlation functions and Correlograms

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 the angle brackets indicate expectation with respect to random draws of $s(t)$ from the process. This is the time-averaged local second moment of the joint distribution on $s(t)$; by contrast, $\overline{\lambda}(t)$ was the (non-time-averaged) first moment. Note that, since $s(t)$ is a sum of δ functions, $R_{ss}(0) = \infty$ for this definition.

An alternative definition for R_{ss} is in terms of a time-averaged conditional first moment. It is the mean intensity at time $t+\tau$, conditioned on an event having occured at time t, and averaged over t. That is,

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

where the conditioning means that $t_i = t$ for some i, and the angle brackets represent expectation with respect to $N(T)$ and the times of all but the ith event. In this case, $R_{ss}^{alt}(0)$ gives the average probability of two events occuring at the same time, which is 0, by definition, for a conditionally orderly process. In what follows, we will stick to the first (i.e., second moment) definition.

Based on the usual decomposition of second moments ($\langle x^2 \rangle = \langle (x-\mu)^2 \rangle + \mu^2$) we can decompose the autocorrelation functions thus:

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

where $\overline{\Lambda}$ is the time-averaged mean rate, and $Q_{ss}(\tau)$ 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".

For an (inhomogeneous) Poisson process $Q_{ss}(\tau) = \delta(\tau)$, by independence. For a general self-exciting process, it indicates (to second order) dependence on nearby spike times. Thus, it is often used to look for oscillatory structure in spike trains (where spikes tend to repeat around with fixed intervals, but at random phase with respect to the stimulus) or similar spike-timing relationships. Note, however, that since any (conditionally orderly) point process is a self-exciting process, *any* non-Poisson process will have a non-δ autocovariance function, even if nearby spike timing relationships are not the most natural (or causal) way to think about the process. In particular, think about the effects of random (but slow) variations in a non-constant $\lambda(t)$, as in a DSPP.

Correlation functions are typically estimated by constructing **correlograms**, which are histograms of the time differences between (not necessarily adjacent) spikes. The covariance function is then estimated by subtracting an estimate of the correlation of the mean intensity; this estimate is perhaps best constructed from the PSTH, but, in practice is often obtained by constructing a "shifted" or "shuffled" correlogram where time differences are taken between spikes from two different trials.

3.3 Power spectra and coherences

Another way to describe the second order statistics of a process is in the frequency domain, through power-spectra, spectrograms and (for multiple processes) coherence. These are increasingly used in neuroscience, but are are beyond our present scope.

3.4 Multiple spike trains

Thus far we have restricted ourselves to spike trains from a single cell. Often, we may actually 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 we can express dependences between the processes generally by dependence on all previous events in all cells:

$$\lambda^{(c)}(t) \to \lambda^{(c)}\left(t|N^{(c)}(t), t_1^{(c)}, \ldots, t_{N^{(c)}(t)}^{(c)}, \{N^{(c')}(t), t_1^{(c')}, \ldots, t_{N^{(c')}(t)}^{(c')}\}\right)$$

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.

The techniques for measuring relationships between cells are analogous to those described for single processes: namely, by cross-correlogram estimates of the cross-correlation function:

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

shift- or shuffle-corrected correlogram estimates of the cross-covariance function:

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

or by cross-spectra or empirical coherences.

Note that, as for autocovariograms, structure in a cross-covariogram needn't imply that dependencies between individual spike times are the most natural way to think about the interaction between the processes – DSPPs with shared latents may also give significant cross-covariance structure.

Parametric models for multiple spike trains have also recently appeared in the experimental and neural-data-modelling literature. Further discussion of these is beyond our present scope.

Chapter 4

Encoding time-varying stimuli

4.1 Introduction

Now suppose $x(t)$ varies (continuously) in time.

- Real world is not constant
- If $\lambda()$ or $\overline{\lambda}()$ depends on many aspects of stimulus can be laborious to explore each sequentially.
- Interactions between stimuli at different times might be non-linear and significant.

General form:

$$\lambda(t|x[0,t), H(t))$$

This cannot be done non-parametrically: $x[0,t)$ is too large, even if we assume a finite encoding window $\lambda(t|x[t-k,t), H(t))$.

Thus we need to build parametric models.

Two approaches:

- Model PTSH $\overline{\lambda}(t|x[0,t))$. Obviates need for history dependence. If "meaning" of spikes is independent, then this may be quantity of interest.

  Models for the PSTH are often optimised by minimising squared error, which is equivalent to assuming Gaussian noise. These methods are closely related to techniques of *system identification* in the engineering literature.
- Model spike trains using $\lambda(t|x[0,t), H(t))$. Ignores history at peril.

4.2 Linear Models

4.2.1 PSTH models

History integrated away. The basic linear model is a filter:

$$\overline{\lambda}(t|x[t-k,t)) = \int_0^k d\xi x(t-\xi)w(\xi)$$

In practice, $x(t)$ may be multidimensional: an image or the instantaneous spectrum of a sound. In that case, so is $w(\xi)$, and the product is a an inner product.

The filter w is most often found by linear regression, minimising the squared error between the PSTH and the filtered stimulus. This would be maximum likelihood if the noise process $\overline{\lambda}(t) - r(t)$ were stationary white Gaussian noise.

Discretise time. Lag-matrix form.

$$\mathbf{N}^\mathsf{T} = \mathsf{X}^\mathsf{T}\mathbf{w}$$

Solution given by pseudo inverse

$$\mathbf{w} = (\mathsf{X}\mathsf{X}^\mathsf{T})^{-1}\mathsf{X}\mathbf{N}^\mathsf{T}$$

4.2.2 Regularisation

In practice, dimensionality of $\mathbf{w}$ can be very high, and the linear model can overfit.

Gaussian prior: $\mathbf{w} \sim \mathcal{N}\left(\mathbf{0}, \mathsf{A}^{-1}\right)$

$$\mathbf{w} = (\mathsf{X}\mathsf{X}^\mathsf{T} + \sigma_n^2\mathsf{A})^{-1}\mathsf{X}\mathbf{N}^\mathsf{T}$$

Ridge regression.

A can also enforce smoothness.

Bayesian evidence optimisation – can be used to degree of smoothness and sparsity given data.

Alternative sparsity by L1 prior.

4.2.3 Spike models

A similar framework could be used for spikes:

$$\lambda(t|x[t-k,t), H(t)) = \int_0^k d\xi x(t-\xi)w(\xi)$$

No explicit history dependence.

If we go ahead and use linear regression here (although squared error/Gaussian noise is not obviously a good idea), we get:

$$\mathbf{w} = (\mathsf{X}\mathsf{X}^\mathsf{T})^{-1}\mathsf{X}\mathbf{s}^\mathsf{T}$$

Now, $\mathbf{s}$ is a vector of zeros and ones. So $\mathsf{X}\mathbf{s}^\mathsf{T}$ is the sum of all stimuli that precede a spike. If $x(t)$ is uncorrelated on average then $\mathsf{X}\mathsf{X}^\mathsf{T} = (T-k)\mathsf{I}$ and

$$\mathbf{w} = \frac{1}{T-k}\sum_i \mathbf{x}_t s_t = \frac{1}{T-k}\sum_i \mathbf{x}_{t_i} = \text{spike-triggered average}$$

Geometrical view

4.2.4 But

Spike rate can't be negative! Neurons are rarely perfectly linear in stimulus intensity.

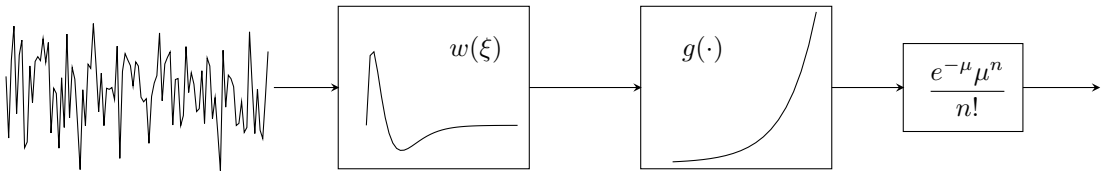
4.3 LNP cascade models

The simplest parametric extension to the linear model is to introduce a "static" nonlinearity in a cascade following the linear filter. If the nonlinearity has non-negative range, its output is a valid intensity function. Thus, we can use a point process likelihood model (or often, in practice, a discretised approximation thereto).

In equations, we can write

$$s(t) \sim \lambda(t|\{x(\cdot)\}_{t-k}^{t})$$
$$\lambda(t|\{x(\cdot)\}_{t-k}^{t}) = g\Big(\int_0^k d\xi\, x(t-\xi)w(\xi)\Big)$$

where the first equation indicates that $s(t)$ is drawn from a conditional intensity function without history dependence (the P), and the second gives the (LN) form of that conditional intensity.



How can we estimate $w(\xi)$ in this model?

4.3.1 Bussgang's Theorem

One helpful result dates back to the 1950s systems identification literature, and is due to Bussgang. It says that if:

1. we study an LN cascade (also known as Wiener cascade) given by

$$y = g(\mathbf{x}^{\mathsf{T}}\mathbf{w}) + \text{noise},$$

2. and we use input (lag) vectors $\mathbf{x}$ that are circularly symmetrically distributed, so that $p(\mathbf{x})$ depends only on $\|\mathbf{x}\|$ (or, in other words, equal probability lines are circles),

3. and the expected value of the correlation between input and output is non-zero

$$\langle y\mathbf{x}\rangle \neq 0,$$

then the correlation is an unbiased estimator of the *direction* of $\mathbf{w}$.

Geometric interpretation. We can use our geometrical view of the STA to see that this holds. Essentially, if all that matters to the probability of spiking at a point in stimulus space is the projection of that point onto a line, and stimuli are distributed with equal density on each side of the line, then the perpendicular contributions to the STA cancel (on average), and the expected STA points along the line.

You will be asked to prove the result more formally as an exercise.

Reconstructing the non-linearity. Once we know the STA, we can recover an estimate of the non-linearity. Consider the discretised case, and write $x_t = \mathbf{x}_t^\mathsf{T}\mathbf{w}$ and $s_t \in \{0, 1\}$ to indicate whether a spike occurred or not. We have two choices:

- Kernel regression, where we think of s as a function of x and smooth:

$$g(x)\,\delta t = \frac{\sum_t \phi(x - x_t)s_t}{\sum_t \phi(x - x_t)}.$$

  Here $\phi()$ is a fixed smoothing kernel function – often Gaussian in shape.

- Looking at the fraction of stimuli at (or around) a particular value of x that generate spike

$$g(x)\,\delta t = p(\mathsf{spike}|x) = \frac{p(x, \mathsf{spike})}{p(x)}$$

  Here, spike is the *event* of a spike, i.e. $s = 1$, rather than a random variable. This ratio may be obtained by histogramming the stimulus distribution with and without a spike. However, that estimator is then frequently smoothed! An alternative is to use a kernel density estimator for x. Then:

$$g(x)\,\delta t = \frac{p(x, \mathsf{spike})}{p(x)} = \frac{\sum_t \psi(x - x_t)s_t}{\sum_t \psi(x - x_t)}.$$

  where $\psi()$ is a fixed kernel with unit integral (although it doesn't actually matter, because any other non-zero integral will cancel in the ratio). This is the same expression as for kernel regression.

Bias due to non-symmetry The first and third premises of Bussgang's theorem are rather obvious. But the second premise is very important.

A silly counterexample would be if stimuli were confined to a single direction that was not parallel to $\mathbf{w}$. But obviously weakening this confinement would still have an effect.

Can the bias be fixed by whitening? Only if the original distribution was elliptically symmetric. The result really needs *symmetry*. Just decorrelation (corresponding to a spherical covariance matrix) is not enough.

One alternative is to resample (or reweight) inputs to make their distribution (or weighted distribution) more elliptical. This ends up throwing away (or discounting) a considerable amount of data, and may be impractical in high dimensions.

4.3.2 Parametric max-likelihood – GLMs

Another approach is to use a specific parametric form of $g()$, and find ML (or penalised/regularised/MAP) weights under the (discretised) point-process likelihood.

$$\ell(\mathbf{w}) = \sum_t s_t \log[g(\mathbf{x}_t^\mathsf{T}\mathbf{w})\,\delta t] - [g(\mathbf{x}_t^\mathsf{T}\mathbf{w})\,\delta t]$$

(assuming $s_t \in \{0, 1\}$ so $s_t! = 1$). [This is the Poisson counting process discretisation. It's also possible to use a Bernoulli discretisation. Both have the same limit at $\delta t \to 0$.]

This is an example of a standard statistical model called a generalised linear model (GLM) in which a non-linear "link" function of the mean parameter of an exponential-family distribution on the output is given by a linear combination of the inputs. GLMs are very well studied.

With certain restrictions on $g()$, the likelihood of a GLM is concave, and thus easily optimised by gradient or second-order methods. A special second order method called Iteratively Reweighted Least Squares (IRLS) is standard. If the function $g()$ is the *canonical* link function, IRLS is equivalent to a Newton method.

For the Poisson case above it is obvious by inspection that the likelihood will be concave as long as the function $g()$ is convex and log-concave (i.e. $\log g$ is concave). Examples include:

$$g(x) = \Theta(x)\, x^\alpha, \quad \alpha \geq 1$$
$$g(x) = e^{\alpha x}$$
$$g(x) = \frac{1}{|\alpha|} \log(1 + e^{\alpha x})$$

Here $\Theta()$ is the Heaviside function. The last function is often useful. It is a softened threshold-linear function. Exercise: show that it is indeed convex and log-concave.]

Regularisation. The number of parameters in the Poisson-GLM is the same as in the linear model. Thus regularisation is equally important.

- L1-norm regularisation, promoting sparsity within a basis, is easy to implement (usually in a gradient method) and preserves the concavity of the posterior.
- Fixed Gaussian regularisation is also concave. However, optimisation of parameters within the Gaussian "prior" is no longer analytically tractable, and must usually be approximated.

4.3.3 Non-parametric Max-likelihood

What if we don't know or don't want to assume the form of the nonlinearity? We can attempt to learn it.

- One approach is to use a basis function expansion for g:

  $$g(x) = \sum_l a_l g_l(x)$$

  for fixed functions $g_l(x)$.

- Another is to use a "non-parametric" method such as kernel regression, where

  $$g(x)\, \delta t = \alpha \frac{\sum_t \phi(x - x_t) s_t}{\sum_t \phi(x - x_t)}$$

  for a fixed kernel ϕ. As we saw above, this is equivalent to measuring the ratio $p(x, \mathsf{spike})/p(x)$. [We have introduced the scale parameter α here to help reduce one form of sensitivity to the kernel

shape. Write $\gamma(x)$ for the unscaled kernel regressor. Then, at the ML solution for α

$$\frac{\partial \ell(\mathbf{w})}{\partial \alpha} = \frac{\partial}{\partial \alpha} \sum_t s_t \log[g(x_t)\,\delta t] - [g(x_t)\,\delta t] = \sum_t s_t \frac{1}{\alpha} - \gamma(x_t) = 0$$

$$\Rightarrow \alpha = \sum_t s_t \Big/ \sum_t \gamma(x_t)$$

and so $\sum_t g(x_t) = \sum_t s_t$ whatever the kernel.]

In either case, we can substitute this expression into the likelihood (note that $g()$ will only be evaluated at x_t) and differentiate. The parameters are then learnt by (sometimes very tedious) gradient ascent.

4.3.4 Maximising information

A related (actually equivalent) approach goes under the name of *Maximally Informative Dimensions* (MID).

Let $x = \mathbf{x}^\mathsf{T}\mathbf{w}$ be the projected stimulus. Then Sharpee et al. (2003), following Brenner et al. (2000), give an expression for the mutual information between a spike arrival time and the projected stimulus:

$$I(\mathbf{w}) = \int dx\; P_\mathbf{w}(x|\mathsf{spike}) \log \left[\frac{P_\mathbf{w}(x|\mathsf{spike})}{P_\mathbf{w}(x)} \right]$$

where the distributions on x obviously depend on the projection vector $\mathbf{w}$ and spike is the event of a spike occuring (not the random variable indicating whether a spike occurred or not). This is obviously also the KL divergence between the spike-conditioned distribution and the marginal stimulus distribution.

The MID approach is to attempt to maximise this mutual information with respect to $\mathbf{w}$. In practice, of course, we must maximise an *estimate* of this quantity, and it turns out that maximizing the conventional estimator is exactly equivalent to the non-parametric maximum likelihood approach described above. To see this, we note that:

- In the estimator, the integral over x is replaced by a sum over time. Thus, we replace

  $$\int dx\; P_\mathbf{w}(x|\mathsf{spike}) f(x) \quad \text{by} \quad \frac{1}{P(\mathsf{spike})} \sum_t s_t f(x_t)$$

  where $s_t = 1$ when there was a spike at time t and 0 otherwise (and $P(\mathsf{spike})$ is unknown, but does not depend on $\mathbf{w}$).

- By Bayes, the ratio

  $$\frac{P_\mathbf{w}(x|\mathsf{spike})}{P_\mathbf{w}(x)} = \frac{P_\mathbf{w}(\mathsf{spike}|x)}{P(\mathsf{spike})}$$

  (note that $P(\mathsf{spike})$ doesn't depend on $\mathbf{w}$).

[Actually, just the first of these observations, coupled with the discussion of the equivalence of the two non-parametric nonlinearity-reconstruction methods above, would suffice to show the point.]

Then

$$\begin{aligned}
\widehat{I}(\mathbf{w}) &= \frac{1}{P(\mathsf{spike})} \sum_t s_t \log \left[\frac{\widehat{P}_\mathbf{w}(\mathsf{spike}|x_t)}{P(\mathsf{spike})} \right] \\
&= \frac{1}{P(\mathsf{spike})} \sum_t s_t \log \widehat{P}_\mathbf{w}(\mathsf{spike}|x_t) - \frac{1}{P(\mathsf{spike})} \sum_t s_t \log P(\mathsf{spike}) \\
&= \frac{1}{P(\mathsf{spike})} \sum_t s_t \log \widehat{P}_\mathbf{w}(\mathsf{spike}|x_t) - \log P(\mathsf{spike})
\end{aligned}$$

Identifying $\widehat{P}_{\mathbf{w}}(\mathsf{spike}|x)$ with a non-parametric estimate of $g(\mathbf{x}^{\mathsf{T}}\mathbf{w})\delta t$ we get the first term of the Poisson likelihood. The second term in our expression for the information doesn't depend on $\mathbf{w}$. What about the second term of the likelihood?

Provided $g()$ has sufficient freedom in scale (as it does in the non-parametric case above) it will always be true that at the maximum-likelihood value

$$\sum_t g(x_t)\delta t = \sum_t s_t \delta t = P(\mathsf{spike})$$

and so this term does not effect the ML choice of $\mathbf{w}$.

4.3.5 Common ground

This allows us to view the STC and nonparametric/MID methods as varying implementations of a common idea: the best estimate of (the direction of) $\mathbf{w}$ is given by the direction in which the spike-conditioned ensemble of stimuli differs most from the background ensemble. The difference between the methods is the measure of the "difference" between the distributions that they exploit.

- The STA maximises the difference in *means* between the spike-conditioned ensemble and the overall one (which is taken to have 0 mean).
- The non-parametric LNP (or equivalently, the MID) maximises the KL divergence between the two distributions (or equivalently $\langle \log p(x|\mathsf{spike})\rangle$).
- The STC approach (below) maximises the difference in second moment.

4.4 Multi-dimensional LNP Models

4.4.1 STC

Another measure of the difference between two distributions is the difference in their variances (or second moments). Let us write:

$$\mathsf{C}_{\mathsf{spike}} = \sum_t s_t \mathbf{x}_t \mathbf{x}_t^{\mathsf{T}} / \sum_t s_t$$

This is the second (zero-centred) moment of the spike-conditioned distribution. We can compare this to the second moment of the overall stimulus distribution

$$\mathsf{C}_{\mathsf{spike}} = \frac{1}{T-k}\sum_t \mathbf{x}_t \mathbf{x}_t^{\mathsf{T}}.$$

In particular, for any direction $\mathbf{u}$ in stimulus space which does not effect the probability of firing, s_t is independent of $\mathbf{u}^{\mathsf{T}}\mathbf{x}$ and $\mathbf{u}^{\mathsf{T}}\mathsf{C}_{\mathsf{spike}}\mathbf{u} = \mathbf{u}^{\mathsf{T}}C\mathbf{u}$ in expectation. Thus, we want to look for directions in which these two quadratic forms differ. [Note that this condition only applies one way — it is not necessarily true that if $\mathbf{u}^{\mathsf{T}}\mathsf{C}_{\mathsf{spike}}\mathbf{u} = \mathbf{u}^{\mathsf{T}}C\mathbf{u}$ then the cell is insensitive to the direction $\mathbf{u}$]. These directions are usually found by looking for significantly non-zero eigenvalues of the difference $\mathsf{C}_{\mathsf{spike}} - \mathsf{C}$.

Significance. To find significantly non-zero eigenvalues we have to have a null distribution for the empirical distribution of eigenvalues whose expectations are actually zero. This is not easy to find, because the distribution of $\mathsf{C}_{\mathsf{spike}}$ is unknown. It is possible to try to construct an estimate by sampling methods; although the most common approach is simply to sort the eigenvalues, identify an apparently smoothly varying set of middle values, and then select the few values that seem to lie outside the trend at the two extremes.

Removing the STA. As $\mathsf{C}_{\mathsf{spike}} - \mathsf{C}$ is symmetric, its eigenvectors are orthogonal. These identify a subspace to which the neuron is sensitive. If the cell has a non-zero STA, it is likely to fall within this space. It is common to want to find a subspace orthogonal to the STA – this could be done by projecting the STA direction out of the subspace once it is identified; but for some reason it is more usual in the literature to project this out first. Let $\widehat{\mathbf{w}}_{\mathsf{STA}}$ be a unit vector parallel to the STA. Then we can write

$$\mathbf{x}_t^{\perp} = \mathbf{x}_t - (\mathbf{x}_t^{\mathsf{T}} \widehat{\mathbf{w}}_{\mathsf{STA}}) \widehat{\mathbf{w}}_{\mathsf{STA}}, \tag{4.1}$$

and

$$C^{\perp} = \frac{1}{T-k} \sum_t \mathbf{x}_t^{\perp} \mathbf{x}_t^{\perp\mathsf{T}} \tag{4.2}$$

$$C_s^{\perp} pike = \frac{1}{\sum_t s_t} \sum_t s_t \mathbf{x}_t^{\perp} \mathbf{x}_t^{\perp\mathsf{T}} \tag{4.3}$$

and then examine the eigenspectrum of $C^{\perp}_{\mathsf{spike}} - C^{\perp}$. In this case the empirical spike-triggered ensemble $\{\mathbf{x}^{\perp}\}$ is zero-mean, so that the outerproduct matrices are true covariance matrices rather than just zero-centred second moments.

Incompleteness. As mentioned above, there is no guarantee that the neuron is actually insensitive to directions outside the eigenspace – all we know is that the conditional variance does not change. Variance-preserving (or approximately variance preserving) transformations will be invisible.

Bias due to non-symmetry We said a few words about bias in STA analysis above. The same concern applies to the STC. Except that now, spherical (or elliptical, and then whitened before the analysis) symmetry alone is insufficient, we actually require Gaussianity. The intuition is straightforward. The multivariate Gaussian is the only distribution that is spherically symmetric, and has the property that the projections of the random vector onto any orthogonal basis are independent. This second result means that if we select vectors based on the value of their projection in one direction, we do not change their distribution (and, in particular, their variance) in any orthogonal direction. Conversely, if the distribution does not have this "any orthogonal components are independent" property — i.e., is non-Gaussian — then selection acting along one direction *can* affect others, and so the STC-derived subspace will be incorrect on average.

4.5 Other topics

4.5.1 NL cascade models

Multilinear formulation

4.5.2 Spike interactions

Spike currents in GLMs

4.5.3 General nonlinear models

- Volterra series
- M-sequences
- Kernel methods