Non-linear, history-dependent response functions

Subhaneil Lahiri

Harvard University

May 25, 2011

Abstract

We look at methods of characterising non-linear, history-dependent responses, particularly the rates of inhomogeneous Poisson processes.

1 Introduction

sec:intro

In this note we will look at responses, $r(t)$, that are *functionals* of some stimulus, $s(t)$,

$$r(t) = r_t[s]. \tag{1}$$

eq:response

We will look at types of series expansion: the Volterra series in §2 that is analogous to the Taylor expansion for function; and the Wiener series in §4 that is analogous to the expansion in Hermite polynomials for functions.

When constructing a set of orthogonal polynomials, one has to choose a weighting function for the integration measure:

$$f \cdot g = \int f(t)g(t)\, w(t)\mathrm{d}t. \tag{2}$$

eq:functionp

For Hermite polynomials, one chooses $w(t) = \mathrm{e}^{-t^2}$, for Legendre polynomials, one chooses $w(t) = \theta(1-t)\theta(1+t)$ and for Laguerre polynomials, one chooses $w(t) = \theta(t)\mathrm{e}^{-t}$.

For orthogonal functionals, one has to choose a measure for the space of functions:

$$F \cdot G = \int F[s]G[s]\, W[s]\mathcal{D}s. \tag{3}$$

eq:functiona

In analogy with the Hermite polynomials, we will choose

$$W[s] \propto \exp\left(-\int \frac{s(t)^2}{2\sigma^2}\, \mathrm{d}t\right). \tag{4}$$

eq:whitenois

There is a statistical way of looking at this that is helpful when thinking about how to measure these quantities. In the case of functions, one can normalise $w(t)$ and think of

1

it as a probability density, then we have $f \cdot g = \langle f(t)g(t) \rangle$, where t is a random variable whose probability distribution is given by $w(t)$.

For functionals, we should think of the input, $s(t)$, as being a *stochastic process*. The stochastic process corresponding to (4) is called zero-mean, Gaussian white noise. We will discuss this in more detail in §3.

When the response function, $r(t)$, is the rate of an inhomogeneous Poisson process, it can be difficult to measure. One has to repeat the stimulus many times and count the events in various time bins. If one wishes to use small time bins, one has to repeat the stimulus more times to get enough data in each bin.

Instead, there is a simpler method where one compute averages of the *stimulus* at various time shifts before each event. With this method, one never needs to determine $r(t)$. We will discuss this further in §5.

2 Volterrra series

Suppose we have a functional that is also a function of the independent variable, e.g. if we have a system that responds to some time varying input, $x(t)$, in a history dependent way (i.e. the system has some memory), then the response at time t is a function of t as well as a functional of $x(t')$. We will assume that the response is invariant under time translations, i.e. $F_t[x(t')] = F_{t+T}[x(t' + T)]$.

The Volterra series is a way of expanding such a functional in powers of the input. It is analogous to the Taylor series expansion of a function.

Definition 1: Volterra functional
A Volterra functional of degree p is one of the form:

$$V_t^{(p)}[x] = \prod_{i=1}^{p} \int \mathrm{d}t_i \, v(t_1, \ldots, t_p) \prod_{i=1}^{p} x(t - t_i). \tag{5}$$

eq:volfunc

Without loss of generality, we may assume that the function $v(t_1, \ldots, t_p)$ is symmetric under interchange of any of its arguments. It is causal if v vanishes whenever any of its arguments is negative.

Definition 2: Volterra series
A Volterra series of degree n is a sum of Volterra functionals of degree less than or equal to n:

$$F_t[x] = \sum_{p=0}^{n} \prod_{i=1}^{p} \int \mathrm{d}t_i \, f_p(t_1, \ldots, t_p) \prod_{i=1}^{p} x(t - t_i). \tag{6}$$

eq:volser

The term "Voltera series", without specification of degree, shall be taken to mean a Volterra series of infinite degree.

Definition 3: Volterra kernel
The pth Volterra kernel is the function $f_p(t_1, \ldots, t_p)$ that appears in (6).

2

Theorem 1 (Determination of Volterra kernels):
Suppose we are given some functional, $F_t[x]$, and we want to express it as a Volterra sreies.
The Volterra kernels that we would use are given by

$$f_p(t_1, \ldots, t_p) = \frac{1}{p!} \frac{\delta^p F_t[0]}{\delta x(t - t_1) \ldots \delta x(t - t_p)}. \tag{7} \quad \boxed{\texttt{eq:volkern}}$$

Proof. Substitute (6) into (7) and see what you get. □

The Volterra expansion is only valid for inputs close to zero if all the functional derivatives in (7) exist, similar to how a Taylor series can only be used for analytic functions. In addition, the Taylor series is only valid inside the radius of convergence, which is equal to the distance to the nearest non-analyticity. This is one of the reasons why orthogonal function expansions are often better than Taylor series: in return for demanding a weaker form of convergence, the class of functions that can be represented is larger.

Remark (impulse response). The experimental analogue of (7) involves measuring the response to a series of impulses. Let

$$x(t) = \sum_{i=1}^{p} \epsilon_i \delta(t - t_i), \tag{8} \quad \boxed{\texttt{eq:imulses}}$$

$$F(t; \vec{t}, \vec{\epsilon}) = F_t[x],$$

where $\vec{t}$ represents $(t_1, \ldots, t_p)$, etc. Then one can determine the Volterra kernels via

$$f_p(t - t_1, \ldots, t - t_p) = \frac{1}{p!} \frac{\partial^p F(t; \vec{t}, \vec{0})}{\partial \epsilon_1 \ldots \partial \epsilon_p}. \tag{9} \quad \boxed{\texttt{eq:impkern}}$$

However, one has to perform this experiment several times, varying $\vec{\epsilon}$ and $\vec{t}$, in order to determine one of the Volterra kernels.

3 Gaussian white noise

Definition 4: Stationary process
A stochastic process, $x(t)$ is said to be stationary if

$$\langle x(t_1) x(t_2) \ldots x(t_n) \rangle = \langle x(t_1 + \tau) x(t_2 + \tau) \ldots x(t_n + \tau) \rangle \qquad \forall n, t_i, \tau \tag{10} \quad \boxed{\texttt{eq:statproc}}$$

Definition 5: Gaussian process
A stochastic process is said to be Gaussian if

$$\left\langle e^{i \int dt\, [J(t) x(t)]} \right\rangle = e^{i \int dt\, [J(t) \mu(t)] - \frac{1}{2} \int dt dt'\, [J(t) J(t') \phi(t, t')]}. \tag{11} \quad \boxed{\texttt{eq:gaussproc}}$$

3

This means that all correlation functions can be expressed in terms of the connected one and two-point functions, e.g.

$$
\begin{aligned}
\langle x(t_1)\rangle &= \mu(t_1),\\
\langle x(t_1)x(t_2)\rangle &= \mu(t_1)\mu(t_2) + \phi(t_1,t_2),\\
\langle x(t_1)x(t_2)x(t_3)\rangle &= \mu(t_1)\mu(t_2)\mu(t_3) + \phi(t_1,t_2)\mu(t_3) + \phi(t_1,t_3)\mu(t_2) + \phi(t_2,t_3)\mu(t_1).
\end{aligned}
\tag{12}
$$

`eq:gaussmom`

If a process is Gaussian and stationary,

$$
\begin{aligned}
\mu(t) &= \mu,\\
\phi(t,t') &= \phi(t-t').
\end{aligned}
\tag{13}
$$

`eq:statgauss`

The Fourier transform of $\phi(t)$, $\tilde{\phi}(\omega)$, is called the **power density spectrum** of $x(t)$.

Definition 6: Gaussian white noise

Gaussian white noise of power σ^2 is a stationary Gaussian process with

$$
\phi(t-t') = \sigma^2\delta(t-t'), \qquad \tilde{\phi}(\omega) = \sigma^2.
\tag{14}
$$

`eq:gaussianw`

i.e. the power density spectrum is the same for all frequencies. In practice, one can only have white noise up to some high frequency cut-off.

Most of the time, we will be dealing with zero-mean gaussian white noise, i.e.

$$
\mu = 0, \qquad \left\langle e^{i\int dt\,[J(t)x(t)]}\right\rangle = e^{-\frac{\sigma^2}{2}\int dt\,[J(t)]^2}.
\tag{15}
$$

`eq:zeromeanw`

4   Wiener series

`sec:wiener`

The Wiener series is an alternative series expansion of a response functional (the type described at the beginning of §2) that is analogous to an expansion of a function as a series of orthogonal functions. For this, we need to define an inner product for functionals:

$$
F \cdot G = \langle F_t[x]G_t[x]\rangle \propto \int \mathcal{D}x \left\{ F_t[x]G_t[x]\, e^{-\int dt\left[\frac{x(t)^2}{2\sigma^2}\right]}\right\},
\tag{16}
$$

`eq:innerfunc`

i.e. $x(t)$ is zero mean Gaussian white noise with power σ^2.

Definition 7: Wiener functional

A Wiener functional of degree p is a Volterra series of degree p that is orthogonal to all Volterra functionals of degree less than p.

Once the pth Volterra kernel of a degree p Wiener functional has been specified, all the lower order Volterra kernels are determined. We will use the notation $G_t^{(p)}[k;x]$ for a degree p Wiener functional whose p^{th} Volterra kernel is k, at time t with input x. e.g. [1,

4

ch.12]:

$$G_t^{(0)}[k;x] = k,$$

$$G_t^{(1)}[k;x] = \int dt_1 \, k(t_1) \, x(t - t_1),$$

$$G_t^{(2)}[k;x] = \int dt_1 dt_2 \, k(t_1, t_2) \, x(t - t_1) \, x(t - t_2) - \sigma^2 \int dt_1 \, k(t_1, t_1), \qquad (17) \quad \boxed{\texttt{eq:wienerex}}$$

$$G_t^{(3)}[k;x] = \int dt_1 dt_2 dt_3 \, k(t_1, t_2, t_3) \, x(t - t_1) \, x(t - t_2) \, x(t - t_3)$$

$$- 3\sigma^2 \int dt_1 dt_2 \, k(t_1, t_2, t_2) \, x(t - t_1).$$

Note that the Wiener functionals also depend on the choice of σ^2.

Definition 8: Wiener series

A Wiener series of degree n is a sum of Wiener functionals of degree less than or equal to n:

$$F_t[x] = \sum_{p=0}^{n} G_t^{(p)}[k_p; x]. \qquad (18) \quad \boxed{\texttt{eq:wienerser}}$$

The term "Wiener series", without specification of degree, shall be taken to mean a Wiener series of infinite degree.

Definition 9: Wiener kernel

The pth Wiener kernel is the function $k_p(t_1, \ldots, t_p)$ that appears in (18).

Theorem 2 (Inner product of Wiener functionals):

The inner product of two Wiener functionals of arbitrary degree with kernels f and g is:

$$\left\langle G_t^{(p)}[f; x] G_t^{(q)}[g; x] \right\rangle = \delta_{p,q} \, p! \sigma^{2p} \int dt_1 \ldots dt_p \, f(t_1, \ldots, t_p) g(t_1, \ldots, t_p). \qquad (19) \quad \boxed{\texttt{eq:wienerinn}}$$

Proof. see [1, §14.2]   □

A particularly useful set of Wiener functionals are:

$$D_{t,\vec{\tau}}^{(p)}[x] = G_t^{(p)}[\delta_{\vec{\tau}}^{(p)}; x],$$

$$\text{where} \quad \delta_{\vec{\tau}}^{(p)}(t_1, \ldots, t_p) = \frac{1}{p!} \sum_{\pi \in S_p} \prod_{i=1}^{p} \delta(t_i - \tau_{\pi(i)}), \qquad (20) \quad \boxed{\texttt{eq:wienerdel}}$$

5

where S_p is the group of permutations of p objects. e.g.

$$
\begin{aligned}
D_{t,\vec{\tau}}^{(0)} &= 1, \\
D_{t,\vec{\tau}}^{(1)} &= x(t-\tau_1), \\
D_{t,\vec{\tau}}^{(2)} &= x(t-\tau_1)\,x(t-\tau_2) - \sigma^2\delta(\tau_1-\tau_2), \\
D_{t,\vec{\tau}}^{(3)} &= x(t-\tau_1)\,x(t-\tau_2)\,x(t-\tau_3) \\
&\quad - \sigma^2[\delta(\tau_1-\tau_2)\,x(t-\tau_3) + \delta(\tau_1-\tau_3)\,x(t-\tau_2) + \delta(\tau_2-\tau_3)\,x(t-\tau_1)]. \\
D_{t,\vec{\tau}}^{(4)} &= x(t-\tau_1)\,x(t-\tau_2)\,x(t-\tau_3)\,x(t-\tau_4) \\
&\quad - \sigma^2[\delta(\tau_1-\tau_2)\,x(t-\tau_3)\,x(t-\tau_4) + \delta(\tau_1-\tau_3)\,x(t-\tau_2)\,x(t-\tau_4) \\
&\qquad + \delta(\tau_2-\tau_3)\,x(t-\tau_1)\,x(t-\tau_4) + \delta(\tau_1-\tau_4)\,x(t-\tau_2)\,x(t-\tau_3) \\
&\qquad + \delta(\tau_2-\tau_4)\,x(t-\tau_1)\,x(t-\tau_3) + \delta(\tau_3-\tau_4)\,x(t-\tau_1)\,x(t-\tau_2)] \\
&\quad + \sigma^4[\delta(\tau_1-\tau_2)\,\delta(\tau_3-\tau_4) + \delta(\tau_1-\tau_3)\,\delta(\tau_2-\tau_4) + \delta(\tau_1-\tau_4)\,\delta(\tau_2-\tau_3)].
\end{aligned}
\tag{21}
$$

In general:

$$
D_{t,\vec{\tau}}^{(p)}[x] = \exp\left(-\frac{\sigma^2}{2}\int du\,\frac{\delta^2}{\delta x(u)^2}\right)\prod_{i=1}^{p} x(t-\tau_i).
\tag{22}
$$

Proof. It's not difficult to see that the highest degree terms in (20) and (22) agree. Therefore we only need to show that it is orthogonal to all lower degree Volterra functionals. In fact, it suffices to show that it is orthogonal to $\prod_{i=1}^{q} x(t-s_i)$ for $q < p$, as these form a basis for the Volterra functionals.

$$
\left\langle \left[e^{-\frac{\sigma^2}{2}\int du\,\frac{\delta^2}{\delta x(u)^2}}\prod_{i=1}^{p} x(t-\tau_i)\right]\prod_{k=1}^{q} x(t-s_k)\right\rangle = 0 \quad \forall q < p.
\tag{23}
$$

We can construct a generating function for the left hand side:

$$
Z[J,j] = \left[e^{-\frac{\sigma^2}{2}\int du\,\frac{\delta^2}{\delta x(u)^2}}e^{i\int dv\,[J(v)x(t-v)]}\right]e^{i\int ds\,[j(s)x(t-s)]},
$$

$$
\frac{(-i)^{p+q}\delta^{p+q}Z[0,0]}{\delta J(\tau_1)\ldots\delta J(\tau_p)\delta j(s_1)\ldots\delta j(s_q)} = \left[e^{-\frac{\sigma^2}{2}\int du\,\frac{\delta^2}{\delta x(u)^2}}\prod_{i=1}^{p} x(t-\tau_i)\right]\prod_{k=1}^{q} x(t-s_k).
\tag{24}
$$

Using (15),

$$
Z[J,j] = e^{\frac{\sigma^2}{2}\int du\,[J(u)]^2}e^{i\int ds\,[J(t-s)+j(t-s)]x(s)},
$$

$$
\langle Z[J,j]\rangle = e^{-\frac{\sigma^2}{2}\int ds\,j(s)[j(s)+2J(s)]},
$$

$$
\frac{(-i)^p\delta^p\langle Z[0,0]\rangle}{\delta J(\tau_1)\ldots\delta J(\tau_p)} = (i\sigma^2)^p\left(\prod_{i=1}^{p} j(\tau_i)\right)e^{-\frac{\sigma^2}{2}\int ds\,[j(s)]^2}.
$$

Therefore, if we take the functional derivative with respect to $j(s_j)$ q times, with $q < p$, we will still have some of the factors of $j(\tau_i)$ left in front. When we set $j(s) = 0$ after that, we will get zero, i.e.

$$\frac{(-\mathrm{i})^{p+q}\delta^{p+q}\langle Z[0,0]\rangle}{\delta J(\tau_1)\ldots\delta J(\tau_p)\delta j(s_1)\ldots\delta j(s_q)} = \left\langle\left[e^{-\frac{\sigma^2}{2}\int\mathrm{d}\tau\frac{\delta^2}{\delta x(\tau)^2}}\prod_{i=1}^{p}x(t-\tau_i)\right]\prod_{k=1}^{q}x(t-s_k)\right\rangle$$
$$= 0 \quad \forall q < p.$$

$\square$

Using (19), we have

$$\left\langle F_t[x]D_{t,\vec{\tau}}^{(p)}[x]\right\rangle = p!\,\sigma^{2p}\,k_p(\tau_1,\ldots,\tau_p), \tag{25}$$

eq:wienerker

where k_p is the pth Wiener kernel of F_t. This formula can be used to determine the kernels to use if we want to represent a functional as a Wiener series. One can show that this choice is optimal, in a least squares sense [1, §15.2]. Note that the kernels depend on the choice of σ^2, which should not be a surprise, as the Wiener functionals themselves depend on it.

If you are worried by the presence of delta functions in (20), note the alternative expression [2]:

$$\left\langle\left(F_t[x] - \sum_{q=0}^{p-1}G_t^{(p)}[k_q;x]\right)\prod_{i=1}^{p}x(t-\tau_i)\right\rangle = p!\,\sigma^{2p}\,k_p(\tau_1,\ldots,\tau_p). \tag{26}$$

eq:wieneralt

This formula agrees with (25) because $\langle x(s)x(t)\rangle = \delta(s-t)$. In a practical situation, delta functions cannot be realised, but we see that one should simply replace the delta functions in (20) with the actual autocorrelation function of the stimulus.

4.1   Estimators

eq:wienerest

There are several sources of error in estimating the response functions in an experiment, e.g.:

- finite sample size,

- non-whiteness of stimulus (perfectly white noise is an idealisation requiring infinite power and temporal resolution),

- non-gaussianity of the stimulus (truly gaussian quantities can go to $\pm\infty$),

- noise in measurements of response and stimulus.

We will concentrate on the first of these for now. A great deal of discussion of the practical issues of using these methods can be found in [3].

Suppose we wish to measure a response function $k_p(\tau_1, \ldots, \tau_n)$ for $\tau_1, \ldots, \tau_n < \tau^*$. Say we present the system with zero-mean Gaussian white noise of power σ^2 at least over a time period $-\tau^* \leq t \leq T$, and we record the response $F_t[x]$ at least for the time period $0 \leq t \leq T$. The obvious choice of estimator is

$$\hat{k}_p(\tau_1, \ldots, \tau_n) = \frac{1}{p!\,\sigma^{2p}\,T} \int_0^T \mathrm{d}t\, F_t[x] D_{t_i,\vec{\tau}}^{(p)}[x]. \tag{27}$$ `eq:wienerest`

It is trivial to show that this estimator is unbiased

$$\hat{k}_p(\tau_1, \ldots, \tau_n) = \frac{1}{p!\,\sigma^{2p}\,T} \int \mathrm{d}t\, \left\langle F_t[x] D_{t_i,\vec{\tau}}^{(p)}[x] \right\rangle = \frac{1}{T} \int \mathrm{d}t\, k_p(\tau_1, \ldots, \tau_p) \tag{28}$$ `eq:wienerest`
$$= k_p(\tau_1, \ldots, \tau_p).$$

It is much harder to show that this is consistent. First we define some generating functionals:

$$\Gamma[J] = \left\langle F_t[x(t')] \mathrm{e}^{\frac{\mathrm{i}}{\sigma^2} \int \mathrm{d}u\, x(t-u)J(u)} \right\rangle,$$

$$W[J] = \left\langle F_t[x(t')] \mathrm{e}^{-\frac{\sigma^2}{2} \int \mathrm{d}v\, \frac{\delta^2}{\delta x(v)^2}} \mathrm{e}^{\frac{\mathrm{i}}{\sigma^2} \int \mathrm{d}u\, x(t-u)J(u)} \right\rangle = \mathrm{e}^{\frac{1}{2\sigma^2} \int \mathrm{d}v\, J(v)^2} \Gamma[J],$$

$$\widehat{\Gamma}[J] = \frac{1}{T} \int \mathrm{d}t\, \left[F_t[x(t')] \mathrm{e}^{\frac{\mathrm{i}}{\sigma^2} \int \mathrm{d}u\, x(t-u)J(u)} \right], \tag{29}$$ `eq:wienergen`

$$\widehat{W}[J] = \frac{1}{T} \int \mathrm{d}t\, \left[F_t[x(t')] \mathrm{e}^{-\frac{\sigma^2}{2} \int \mathrm{d}v\, \frac{\delta^2}{\delta x(v)^2}} \mathrm{e}^{\frac{\mathrm{i}}{\sigma^2} \int \mathrm{d}u\, x(t-u)J(u)} \right] = \mathrm{e}^{\frac{1}{2\sigma^2} \int \mathrm{d}v\, J(v)^2} \Gamma[J],$$

then

$$k_p(\tau_1, \ldots, \tau_p) = \frac{(-\mathrm{i})^p}{p!} \frac{\delta^p W[J]}{\delta J(\tau_1) \ldots \delta J(\tau_p)},$$

$$\hat{k}_p(\tau_1, \ldots, \tau_p) = \frac{(-\mathrm{i})^p}{p!} \frac{\delta^p \widehat{W}[J]}{\delta J(\tau_1) \ldots \delta J(\tau_p)}. \tag{30}$$ `eq:wienergen`

As a consequence of (28), $\left\langle \widehat{W}[J] \right\rangle = W[J]$.

`wienercons` **Theorem 3 (Consistency of estimators (27)):** The covariance of the estimators (27) vanishes as the sample size goes to infinity:

$$\lim_{T \to \infty} \mathrm{Cov}\left(\widehat{W}[J], \widehat{W}[\tilde{J}] \right) = 0. \tag{31}$$ `eq:wienercon`

As we have already showed that they are unbiased, this is sufficient for consistency.

We will need to use an identity from Sidney Coleman:

$$F\left(-\mathrm{i}\frac{\partial}{\partial x} \right) G(x) = G\left(-\mathrm{i}\frac{\partial}{\partial y} \right) F(y)\, \mathrm{e}^{\mathrm{i}xy} \bigg|_{y=0}. \tag{32}$$ `eq:coleman`

It can be proved by expressing both sides in terms of the Fourier transforms of F and G and verifying that the integrands are equal. It holds equally well when we replace functions by functionals and partial derivatives by functional derivatives.

Proof. First let's look at $\widehat{\Gamma}[J]$:

$$
\begin{aligned}
\left\langle \widehat{\Gamma}[J] \right\rangle &= \int \frac{\mathrm{d}t}{T} \left\langle F_t[x(t')] e^{\frac{i}{\sigma^2} \int \mathrm{d}u\, x(u) J(t-u)} \right\rangle \\
&= \int \frac{\mathrm{d}t}{T} F_t \left[-i\sigma^2 \frac{\delta}{\delta J(t-t')} \right] \left\langle e^{\frac{i}{\sigma^2} \int \mathrm{d}u\, x(u) J(t-u)} \right\rangle \\
&= \int \frac{\mathrm{d}t}{T} F_t \left[-i\sigma^2 \frac{\delta}{\delta J(t-t')} \right] e^{-\frac{1}{2\sigma^2} \int \mathrm{d}u\, J(t-u)^2} \\
&= e^{-\frac{\sigma^2}{2} \int \mathrm{d}v\, \frac{\delta^2}{\delta x(v)^2}} \int \frac{\mathrm{d}t}{T} F_t[x(t')]\, e^{\frac{i}{\sigma^2} \int \mathrm{d}u\, x(t-u) J(u)} \Bigg|_{x=0}.
\end{aligned}
$$
(33)   eq:wienergen

Using essentially the same argument:

$$
\left\langle \widehat{\Gamma}[J]\widehat{\Gamma}[\widetilde{J}] \right\rangle = e^{-\frac{\sigma^2}{2} \int \mathrm{d}v \left[\frac{\delta}{\delta x(v)} + \frac{\delta}{\delta y(v)} \right]^2} \int \frac{\mathrm{d}t\,\mathrm{d}s}{T^2} F_t[x(t')] F_s[y(s')] e^{\frac{i}{\sigma^2} \int \mathrm{d}u\, [x(t-u)J(u) + y(s-u)\widetilde{J}(u)]} \Bigg|_{x=y=0}.
$$
(34)   eq:wienergen

Combining (33) and (34) leads to

$$
\begin{aligned}
\mathrm{Cov}\left(\widehat{\Gamma}[J], \widehat{\Gamma}[\widetilde{J}] \right) = &\left(e^{-\sigma^2 \int \mathrm{d}v\, \frac{\delta}{\delta x(v)} \frac{\delta}{\delta y(v)}} - 1 \right) e^{-\frac{\sigma^2}{2} \int \mathrm{d}v \left[\frac{\delta^2}{\delta x(v)^2} + \frac{\delta^2}{\delta y(v)^2} \right]} \\
&\int \frac{\mathrm{d}t\,\mathrm{d}s}{T^2} F_t[x(t')] F_s[y(s')] e^{\frac{i}{\sigma^2} \int \mathrm{d}u\, [x(t-u)J(u) + y(s-u)\widetilde{J}(u)]} \Bigg|_{x=y=0}.
\end{aligned}
$$
(35)   eq:wienergen

Note that

$$
\int \mathrm{d}v\, \frac{\delta}{\delta x(v)} \frac{\delta}{\delta y(v)}\, x(u) y(v) = \frac{1}{2} \int \mathrm{d}v\, \frac{\delta^2}{\delta x(v)^2}\, x(u) x(v) = \delta(u - v).
$$

If we look at a Volterra expansion of F, (6), we see that t only appears in $x(t - u_a)$ and s only appears in $y(s - v_a)$. As we will set $x = y = 0$ at the end, the only s and t dependence will come from the delta functions. Furthermore, in the delta functions resulting from $\frac{\sigma^2}{2} \int \mathrm{d}v \left[\frac{\delta^2}{\delta x(v)^2} + \frac{\delta^2}{\delta y(v)^2} \right]$, the s and t dependence will cancel. Because

$$
e^{-\sigma^2 \int \mathrm{d}v\, \frac{\delta}{\delta x(v)} \frac{\delta}{\delta y(v)}} - 1 = \sum_{n=1}^{\infty} \frac{(-\sigma^2)^n}{n!} \left[\int \mathrm{d}v\, \frac{\delta}{\delta x(v)} \frac{\delta}{\delta y(v)} \right]^n,
$$

every term will have at least one delta function of the form $\delta(s - t + u_a - v_b)$. Terms with multiple delta functions of that for can be processed using

$$
\delta(s-t+u_a-v_b)\delta(s-t+u_c-v_d)\delta(s-t+u_e-v_f)\ldots = \delta(s-t+u_a-v_b)\delta(v_b-u_a+u_c-v_d)\delta(v_b-u_a+u_e-v_f)\ldots.
$$

9

Therefore, the only s and t dependence will come from a single delta function of the form $\delta(s - t + u_a - v_b)$ in each term. This allows us to do the s and t integrals

$$\int \frac{\mathrm{d}t\mathrm{d}s}{T^2} \delta(s - t + u_a - v_b) = \frac{1}{T}.$$

This will be the only T dependent part of the whole thing. In conclusion

$$\mathrm{Cov}\left(\widehat{W}[J], \widehat{W}[\widetilde{J}]\right) = \mathrm{e}^{\frac{1}{2\sigma^2} \int \mathrm{d}v\, [J(v)^2 + \widetilde{J}(v)^2]} \mathrm{Cov}\left(\widehat{\Gamma}[J], \widehat{\Gamma}[\widetilde{J}]\right) \propto \frac{1}{T}, \tag{36}$$

which implies (31). $\qquad\qquad\square$

5   Inhomogeneous Poisson processes

A Poisson process consists of a series of discrete events whose timing is stochastic. The probability of an event occurring at time t is given by a rate function, $r(t)$, (for a homogeneous Poisson process, the rate is constant). This probability is independent of the timing of any other events, but in our model it can depend on the history of some stimulus, $s(t)$.

$$P(\text{event time} \in [t, t + \mathrm{d}t]) = r(t)\mathrm{d}t, \qquad r(t) = r_t[s]. \tag{37}$$

We also assume time translational invariance: $r_t[s(t')] = r_{t+T}[s(t' + T)]$. The probability of a particular event train, $\{t_1, \ldots, t_n\}$, is given by

$$P(t_1, \ldots, t_n)\, \mathrm{d}t_1 \ldots \mathrm{d}t_n = \mathrm{e}^{- \int \mathrm{d}t\, r(t)} \prod_{i=1}^{n} r(t_i)\, \mathrm{d}t_i, \qquad t_1 < t_2 < \ldots < t_n, \tag{38}$$

where the exponential factor comes from demanding that there are no events at times other than $\{t_1, \ldots, t_n\}$. When integrating over all $\{t_1, \ldots, t_n\}$, one can ignore the restriction over the order of the t_i provided one also divides by $n!$. When integrating over all possibilities, one must also sum over all possible event counts, n.

Measuring the rate function can be difficult. If one performs N repetitions with tim bins of width Δt, the uncertainty scales as $1/\sqrt{N\Delta t}$. Using small time bins will require many repetitions. We will now discuss a method of measuring the response to stimuli that circumvents the need to measure $r_t[s]$ if we wanted to use (9) or (25) directly.

We are going to consider responses to stimuli that are stochastic processes themselves. As we will have two stochastic processes, the stimuli and the timing of events, it will be useful to define some notation.

Notation (outcomes). Let $\hat{t}$, $\widehat{\{t_i\}}$ and $\hat{s}$ denote the following outcomes:

$$\begin{aligned} \hat{t} &= \{\text{outcomes} \mid \text{an event occurs at time } t\}, \\ \widehat{\{t_i\}} &= \{\text{outcomes} \mid \text{events occur only at times } \{t_i\}\}, \\ \hat{s} &= \{\text{outcomes} \mid \text{the stimulus was } s(t)\}. \end{aligned} \tag{39}$$

10

Notation (partial expectations). Let the symbols $\langle\ldots\rangle_A$ and $\langle\ldots\rangle_{A|B}$ denote the following partial expectation values

$$\langle\ldots\rangle_A = \int \ldots P(A)\,\mathrm{d}A,$$
$$\langle\ldots\rangle_{A|B} = \int \ldots P(A|B)\,\mathrm{d}A. \tag{40}$$

5.1 Event-triggered correlators

In this section, we will discuss the theoretical concept of an event-triggered correlator, i.e. the expectation values will be integrals over a probability distribution. We will postpone discussion of how to actually measure these quantities to the next section. In neuroscience, the events of interest are when a neuron fires an action potential, hence the quantity usually measured is a spike-triggered average [4, 5].

Definition 10: Event-triggered correlator

The event-triggered correlator at delays $(t_1,\ldots,t_n)$ is given by

$$\mathrm{ETC}(t_1,\ldots,t_n) = \langle s(t-t_1)\ldots s(t-t_n)\rangle_{\hat{s}|\hat{t}}. \tag{41}$$

Theorem 4:

The event-triggered correlator and the rate function are related by

$$\mathrm{ETC}(t_1,\ldots,t_n) = \frac{\langle r_t[s]s(t-t_1)\ldots s(t-t_n)\rangle}{\langle r_t[s]\rangle}. \tag{42}$$

Proof. Bayes' theorem states that

$$P(\hat{s}|\hat{t}) = \frac{P(\hat{t}|\hat{s})P(\hat{s})}{P(\hat{t})}. \tag{43}$$

However, $P(\hat{t}|\hat{s}) = r_t[s]$ and $P(\hat{t}) = \int P(\hat{t}|\hat{s})P(\hat{s})\mathcal{D}s = \langle r_t[s]\rangle$. Therefore

$$\begin{aligned}
\mathrm{ETC}(t_1,\ldots,t_n) &= \int s(t-t_1)\ldots s(t-t_n)\,P(\hat{s}|\hat{t})\mathcal{D}s \\
&= \frac{\int s(t-t_1)\ldots s(t-t_n)\,r_t[s]\,P(\hat{s})\mathcal{D}s}{\langle r_t[s]\rangle} \\
&= \frac{\langle r_t[s]s(t-t_1)\ldots s(t-t_n)\rangle}{\langle r_t[s]\rangle}.
\end{aligned} \tag{44}$$

$\square$

As these expectations are linear in the $s(t-t_i)$, we can use (25)in conjunction with (42) to compute the Wiener kernels of the rate:

$$k_p(\tau_1,\ldots,\tau_p) = \frac{\langle r_t[s]\rangle\left\langle D_{t,\vec{\tau}}^{(p)}[s]\right\rangle_{\hat{s}|\hat{t}}}{p!\,\sigma^{2p}}, \tag{45}$$

where $s(t)$ is Gaussian white noise of power σ^2.

5.2   Unbiased estimators

The expression on the right hand side of (45) involves a theoretical expectation value, with respect to the true probability distributions that we will not know in an experimental situation. We have to estimate it from the results of an experiment.

Suppose we present the system with zero-mean Gaussian white noise of power σ^2 over a time period T, and we see that events take place at times $(t_1, \ldots, t_n)$. The obvious choice of estimator is

$$\hat{k}_p(\tau_1, \ldots, \tau_n) = \frac{1}{p!\,\sigma^{2p}} \frac{n}{T} \frac{1}{n} \sum_{i=1}^{n} D_{t_i,\vec{\tau}}^{(p)}[s]$$

$$= \frac{\sum_{i=1}^{n} D_{t_i,\vec{\tau}}^{(p)}[s]}{p!\,\sigma^{2p}\,T}. \tag{46}$$

Theorem 5:
This estimator is unbiased.

Proof. First, let's average over event times:

$$\left\langle \hat{k}_p(\tau_1, \ldots, \tau_n) \right\rangle_{\widehat{\{t_i\}}|\hat{s}} = \frac{\sum_{n=1}^{\infty} \frac{1}{n!} \int dt_1 \ldots dt_n \left(\prod_{i=1}^{n} r_{t_i}[s]\right) e^{-\int dt\, r_t[s]} \sum_{j=1}^{n} D_{t_j,\vec{\tau}}^{(p)}[s]}{p!\,\sigma^{2p}\,T}$$

$$= \frac{e^{-\int dt\, r_t[s]} \sum_{n=1}^{\infty} \frac{1}{n!} \int dt_1 \ldots dt_n \sum_{j=1}^{n} \left(\prod_{i\neq j} r_{t_i}[s]\right) r_{t_j}[s]\, D_{t_j,\vec{\tau}}^{(p)}[s]}{p!\,\sigma^{2p}\,T}$$

$$= \frac{e^{-\int dt\, r_t[s]} \sum_{n=1}^{\infty} \frac{1}{(n-1)!} \left(\int dt\, r_t[s]\right)^{n-1} \int dt\, r_t[s]\, D_{t,\vec{\tau}}^{(p)}[s]}{p!\,\sigma^{2p}\,T}$$

$$= \frac{\int dt\, r_t[s]\, D_{t,\vec{\tau}}^{(p)}[s]}{p!\,\sigma^{2p}\,T}.$$

Now we can also average over stimuli (see (28))

$$\left\langle \hat{k}_p(\tau_1, \ldots, \tau_n) \right\rangle = \left\langle \left\langle \hat{k}_p(\tau_1, \ldots, \tau_n) \right\rangle_{\widehat{\{t_i\}}|\hat{s}} \right\rangle_{\hat{s}} = \left\langle \frac{\int dt\, r_t[s]\, D_{t,\vec{\tau}}^{(p)}[s]}{p!\,\sigma^{2p}\,T} \right\rangle$$

$$= k_p(\tau_1, \ldots, \tau_n).$$

$\square$

5.3   Estimator variance and consistency

In this section we will compute the variances of the estimators defined in the previous section and show that they are consistent, i.e. that the variance vanishes in the limit $T \to \infty$ (we have already shown that the bias vanishes). Again, these will be theoretical variances, with respect to the true probability distributions. We will postpone the definition of estimators for the variance until the next section.

First we define some new generating functions, in addition to those in (29):

$$\widehat{\overline{\Gamma}}[J] = \frac{1}{T}\sum_i e^{\frac{i}{\sigma^2}\int du\, s(t-u)J(u)},$$

$$\widehat{\overline{W}}[J] = \frac{1}{T}\sum_i e^{-\frac{\sigma^2}{2}\int d\tau\, \frac{\delta^2}{\delta x(\tau)^2}} e^{\frac{i}{\sigma^2}\int du\, s(t-u)J(u)} = e^{\frac{1}{2\sigma^2}\int dv\, J(v)^2}\widehat{\overline{\Gamma}}[J], \qquad (47) \quad \boxed{\texttt{eq:ratewiene}}$$

$$\hat{k}_p(\tau_1,\dots,\tau_p) = \frac{(-i)^p}{p!}\frac{\delta^p \widehat{\overline{W}}[J]}{\delta J(\tau_1)\dots\delta J(\tau_p)}.$$

The results of the previous section imply that $\left\langle \widehat{\overline{W}}[J]\right\rangle = W[J]$.

Theorem 6 (Consitency of estimators (46)):
The covariance of the estimators (46) vanishes as the sample size goes to infinity:

$$\lim_{T\to\infty}\text{Cov}\left(\widehat{\overline{W}}[J],\widehat{\overline{W}}[\tilde{J}]\right) = 0. \qquad (48) \quad \boxed{\texttt{eq:ratewiene}}$$

As we have already showed that they are unbiased, this is sufficient for consistency.

Proof. We will suppress the dependence of $r(t) = r_t[s]$ on $s(t)$ and the range of sums and products.

$$\left\langle \widehat{\overline{\Gamma}}[J]\widehat{\overline{\Gamma}}[\tilde{J}]\right\rangle_{\widehat{\{t_i\}|\hat{s}}} = \sum_n \frac{e^{-\int dv\, r(v)}}{n!T^2}\int d^n t\prod_i r(t_i)\sum_{jk} e^{\frac{i}{\sigma^2}\int du\,[s(t_j-u)J(u)+s(t_k-u)\tilde{J}(u)]}$$

$$= \sum_n \frac{e^{-\int dv\, r(v)}}{n!T^2}\int d^n t\left\{\sum_{j\neq k}\left[\left(\prod_{i\neq j,k}r(t_i)\right)r(t_j)r(t_k)e^{\frac{i}{\sigma^2}\int du\,[s(t_j-u)J(u)+s(t_k-u)\tilde{J}(u)]}\right]\right.$$

$$\left. +\sum_{j=k}\left[\left(\prod_{i\neq j}r(t_i)\right)r(t_j)e^{\frac{i}{\sigma^2}\int du\, s(t_j-u)[J(u)+\tilde{J}(u)]}\right]\right\}$$

$$= \sum_n \frac{e^{-\int dv\, r(v)}}{n!T^2}\left\{n(n-1)\left(\int dt\, r(t)\right)^{n-2}\left(\int dt\, r(t)e^{\frac{i}{\sigma^2}\int du\,[s(t_j-u)J(u)]}\right)\right.$$

$$\times\left(\int dt\, r(t)e^{\frac{i}{\sigma^2}\int du\,[s(t_j-u)\tilde{J}(u)]}\right)$$

$$\left. +n\left(\int dt\, r(t)\right)^{n-1}\left(\int dt\, r(t)e^{\frac{i}{\sigma^2}\int du\, s(t_j-u)[J(u)+\tilde{J}(u)]}\right)\right\}$$

$$= \widehat{\overline{\Gamma}}[J]\widehat{\overline{\Gamma}}[\tilde{J}] + \frac{\widehat{\overline{\Gamma}}[J+\tilde{J}]}{T}.$$

If we then average over stimuli, we find

$$\text{Cov}\left(\widehat{\overline{W}}[J],\widehat{\overline{W}}[\tilde{J}]\right) = \text{Cov}\left(\widehat{W}[J]\widehat{W}[\tilde{J}]\right) + \frac{e^{-\int dv\, \frac{J(v)\tilde{J}(v)}{\sigma^2}}}{T}W[J+\tilde{J}]$$

Given that we showed that the first term vanishes in the limit to prove Th.3, (48) is clearly true. $\qquad\square$

5.4   Standard error

In this section we will define estimators for the variances computed in the previous section. These will allow us to put error bars on the estimators in (46).

References

hetzen:1980 [1] M. Schetzen, *The Voltera and Wiener theories of nonlinear systems.* Wiley-Interscience, New York, USA, 1980.

Lee1965 [2] Y. W. Lee and M. Schetzen, "Measurement of the Wiener Kernels of a Non-linear System by Cross-correlation," *International Journal of Control* **2** (1965) no. 3, 237–254.

relis:1978 [3] P. Marmarelis and V. Marmarelis, *Analysis of Physiological Systems: The White-Noise Approach.* Plenum Press, New York, USA, 1978.

Abbot:2001 [4] P. Dayan and L. Abbot, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems.* MIT Press, Cambridge MA, USA, 2001.

Kuyper:1968 [5] E. de Boer and P. Kuyper, "Triggered correlation," *IEEE BIomedical Engineering* **15** (1968) 169–179, PubMed:5667803.