Stochastic Processes

Will Penny

19th May 2011

Introduction

We will

- Show the relation between stochastic differential equations, Gaussian processes and Fokker-Planck methods
- This gives us a formal way of deriving equations for the activity of a population of neurons. These are used to study neural coding and can form generative models of brain imaging data.
- Stochastic processes also provides models of decision making in the brain. These can be fitted to behavioural data and used as regressors in computational fMRI
- This material is essential for understanding the next lecture on Hierarchical Dynamic Models

Stochastic Differential Equations

We consider stochastic differential equations (SDEs)

$$dx(t) = a[x(t), t]dt + b[x(t), t]dw(t)$$

where $dw(t)$ is a Wiener process and a and b are, most generally, time varying functions of the state variable x.

An SDE can be written in integral form

$$x(t) = x(t_0) + \int_{t_0}^{t} a[(x(t'), t']dt' + \int_{t_0}^{t} b[x(t'), t']dw(t')$$

Wiener process

A Wiener process

$$dw_t = w(t + dt) - w(t)$$

is a stochastic process with independent increments

$$w(t + \delta t) - w(t) \sim N(0, \delta t)$$

and is independent of the history of the process up to time t. $N(\mu, \sigma^2)$ denotes a Gaussian density with mean μ and variance σ^2.

Sample Paths

Given the SDE

$$dx(t) = a[x(t), t]dt + b[x(t), t]dw(t)$$

a sample path can be generated by the Euler-Maruyama (EM) method (Higham, 2001)

$$x_{i+1} = x_i + a(x_i, t_i)\Delta t_i + b(x_i, t_i)\Delta w_i$$

where

$$\begin{aligned} x_i &= x(t_i) \\ \Delta t_i &= t_{i+1} - t_i \end{aligned}$$

and

$$\begin{aligned} \Delta w_i &= w(t_{i+1}) - w(t_i) \\ &\sim \mathrm{N}(0, \Delta t_i^2) \end{aligned}$$

Wiener Process

Now consider the SDE

$$dx_t = \mu dt + \sigma dw_t$$

With initial condition $x_0 = 0$, the above equation describes the evolution of a Gaussian density with mean μt and variance $\sigma^2 t$ (to be shown later - see Expectations).

That is, the solution is a Gaussian process

$$p(x_t) = N(\mu t, \sigma^2 t)$$

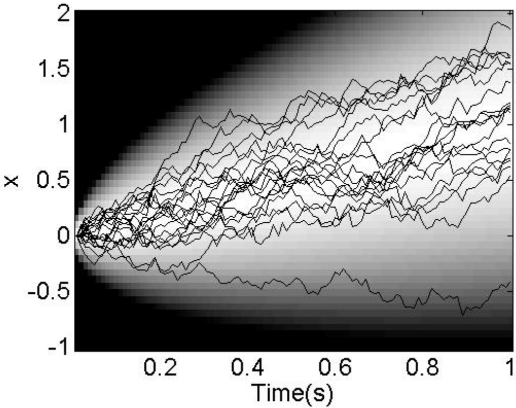
For $\mu = 0$ and $\sigma = 1$ this reverts to the standard Wiener process

$$dx_t = dw_t$$

Wiener Process

$$p(x_t) = N(\mu t, \sigma^2 t)$$

For $\mu = 1$ and $\sigma = 0.05$ we have



The grey scale indicates probability density and the trajectories indicate 20 sample paths.

Gaussian Process

If the SDE

$$dx(t) = a[x(t), t]dt + b[x(t), t]dw(t)$$

has a solution $p(x, t)$ that can be described by a Gaussian we have a Gaussian process.

This is the case for $a[x(t), t]$ and $b[x(t), t]$ being linear functions of $x(t)$.

In the next lecture we will derive expressions for the mean and covariance functions, for the general multivariate case.

Ornstein-Uhlenbeck

An Ornstein-Uhlenbeck (OU) process is given by

$$dx(t) = a[x(t), t]dt + b[x(t), t]dw(t)$$

where

$$\begin{aligned} a[x(t), t] &= ax(t) \\ b[x(t), t] &= \sigma \end{aligned}$$

For a Wiener process we had

$$\begin{aligned} a[x(t), t] &= \mu \\ b[x(t), t] &= \sigma \end{aligned}$$

Some sources also describe $a[x(t), t] = c + ax(t)$ as an OU process. But most (eg Gardiner, 1983) do not.

OU process

The OU process

$$dx_t = \frac{1}{\tau} x_t dt + \sigma dw_t$$

has solution

$$\begin{aligned} p(x_t) &= N(\mu_t, \sigma_t^2) \\ \sigma_t^2 &= \frac{\sigma^2}{2\tau} (1 - \exp[-2t/\tau]) \\ \mu_t &= x_0 \exp[-t/\tau] \end{aligned}$$

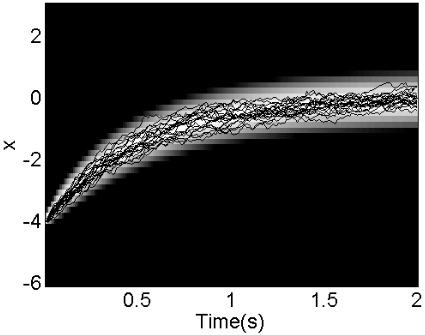
The solution can be derived as shown later (see Expectations).

OU process

OU process with $x_0 = -4$, $\sigma = 0.05$ and $\tau = 0.5$.

$$\sigma_t^2 = \frac{\sigma^2}{2\tau}(1 - \exp[-2t/\tau])$$
$$\mu_t = x_0 \exp[-t/\tau]$$

The grey scale indicates probability density and the trajectories indicate 20 sample paths.



Mean-reverting process

The mean-reverting process

$$dx_t = \frac{1}{\tau}(\mu - x_t)dt + \sigma dw_t$$

The solution of the above equation is a Gaussian density

$$\begin{aligned} p(x_t) &= N(\mu_t, \sigma_t^2) \\ \sigma_t^2 &= \frac{\sigma^2}{2\tau}(1 - \exp[-2t/\tau]) \\ \mu_t &= x_0 \exp[-t/\tau] + \mu(1 - \exp[-t/\tau]) \end{aligned}$$

These expressions can be derived using the stochastic chain rule, and taking expectations (see later).

Mean-reverting process

$$
\begin{aligned}
p(x_t) &= N(\mu_t, \sigma_t^2) \\
\sigma_t^2 &= \frac{\sigma^2}{2\tau}(1 - \exp[-2t/\tau]) \\
\mu_t &= x_0 \exp[-t/\tau] + \mu(1 - \exp[-t/\tau])
\end{aligned}
$$

The density at the steady-state ie. after reverting to the mean is given by a Gaussian with mean μ and variance $\sigma^2/2\tau$.

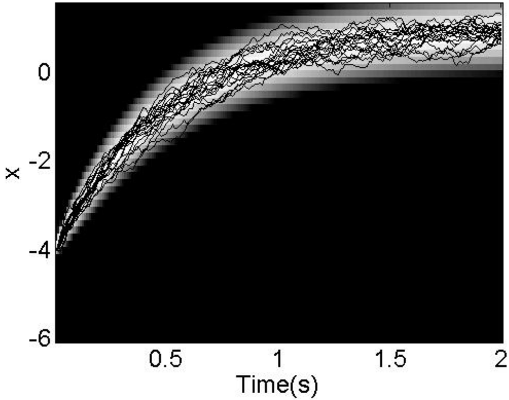
The steady-state density is also known as the Sojourn density.

$$
dx_t = \frac{1}{\tau}(\mu - x_t)dt + \sigma dw_t
$$

The parameter τ therefore determines the time scale at which the Sojourn density is reached.

Mean-reverting process

Mean-reverting process with $x_0 = -4$, $\mu = 1$, $\sigma = 0.05$ and $\tau = 0.5$. The grey scale indicates probability density and the trajectories indicate 20 sample paths.



Change of variables

Given the deterministic dynamical system

$$\frac{dx(t)}{dt} = a[x(t), t]$$

For a new variable

$$y = f[x(t)]$$

We have from the chain rule

$$\begin{aligned} \frac{dy}{dt} &= \frac{dy}{dx}\frac{dx}{dt} \\ &= f'[x]a[x(t), t] \end{aligned}$$

where $f'[x]$ is the derivative with respect to x. Hence

$$df[x] = f'[x]a[x(t), t]dt$$

Change of variables

For the univariate SDE

$$dx(t) = a[x(t), t]dt + b[x(t), t]dw(t)$$

the dynamical equation for a new variable

$$y = f[x(t)]$$

can be written as follows. First, we note that expanding f in a Taylor series to second order gives

$$f[x(t) + dx(t)] = f[x(t)] + f'[x(t)]dx(t) + \frac{1}{2}f''[x(t)]dx(t)^2$$

Hence

$$df[x(t)] = f'[x(t)]dx(t) + \frac{1}{2}f''[x(t)]dx(t)^2$$

Ito's formula

Hence

$$df[x(t)] = f'[x(t)]dx(t) + \frac{1}{2}f''[x(t)]dx(t)^2$$

Substituting $dx(t)$ and only keeping terms linear in dt gives

$$\begin{aligned} df[x(t)] &= f'[x(t)]\left(a[x(t),t]dt + b[x(t),t]dw(t)\right) \\ &+ \frac{1}{2}f''[x(t)]b[x(t),t]^2[dw(t)]^2 \end{aligned}$$

Now use $[dw(t)]^2 = dt$ (see later) to obtain

$$\begin{aligned} df[x(t)] &= \left(a[x(t),t]f'[x(t)] + \frac{1}{2}b[x(t),t]^2 f''[x(t)]\right)dt \\ &+ b[x(t),t]f'[x(t)]dw(t) \end{aligned}$$

This is known as Ito's formula or the stochastic chain rule (Higham, 2001).

Stochastic versus deterministic chain rule

For DEs we have

$$df[x] = f'[x]a[x(t), t]dt$$

For SDEs we have

$$\begin{aligned} df[x(t)] &= \left(a[x(t), t]f'[x(t)] + \frac{1}{2}b[x(t), t]^2 f''[x(t)] \right) dt \\ &+ b[x(t), t]f'[x(t)]dw(t) \end{aligned}$$

For linear flows the curvature f'' is zero.

Time-varying functions

Given the univariate SDE

$$dx(t) = a[x(t), t]dt + b[x(t), t]dw(t)$$

For a new variable which is a time-varying function of the state

$$y = f[x(t), t]$$

Ito's rule has an extra term

$$\begin{aligned} df[x(t)] &= \left(a[x(t), t]f'[x(t)] + \frac{df}{dt} + \frac{1}{2} b[x(t), t]^2 f''[x(t)] \right) dt \\ &+ b[x(t), t]f'[x(t)]dw(t) \end{aligned}$$

Multivariate SDE

For the multivariate SDE

$$dx = A(x,t)dt + B(x,t)dw(t)$$

the stochastic chain rule is

$$\begin{aligned} df(x) &= \left(\sum_i A_i[x,t] j_i(x) + \frac{1}{2} \sum_{i,j} [B(x,t)B(x,t)^T]_{ij} H_{ij}(x) \right) dt \\ &+ \sum_{i,j} B_{ij}(x,t) j_i(x) dw_j(t) \end{aligned}$$

where

$$\begin{aligned} j_i(x) &= \frac{df_i(x)}{dx} \\ H_{ij}(x) &= \frac{d^2 f_i(x)}{dx_j} \end{aligned}$$

are the gradient and curvature. These formula are useful, for example, for computing moments.

Wiener Process

A Wiener process is defined by the SDE

$$dx_t = \mu dt + \sigma dw_t$$

with initial condition x_0. The integral form is

$$x_t = x_0 + \int_0^t \mu dt + \int_0^t \sigma dw_t$$

Hence

$$x_t = x_0 + \mu t + \sigma [w_t - w_0]$$

Wiener Process

The solution is

$$x_t = x_0 + \mu t + \sigma [w_t - w_0]$$

where

$$\begin{aligned} E[w_t - w_0] &= 0 \\ Var[w_t - w_0] &= t \end{aligned}$$

The mean and variance of x_t are therefore

$$\begin{aligned} E[x_t] &= x_0 + \mu t \\ Var[x_t] &= \sigma^2 t \end{aligned}$$

OU Process

An OU process is defined by the SDE

$$dx_t = -ax_t dt + \sigma dw_t$$

with initial condition x_0. We can transform this equation so that x_t does not appear on the right hand side. This can be achieved with the transformation

$$\begin{aligned} y &= f(x) \\ &= x \exp(at) \end{aligned}$$

where y is a time-varying function of x. We have

$$\begin{aligned} \frac{df}{dx} &= \exp(at) \\ \frac{d^2 f}{dx^2} &= 0 \\ \frac{df}{dt} &= ax \exp(at) \end{aligned}$$

OU Process

We have

$$\frac{df}{dx} = \exp(at)$$
$$\frac{d^2 f}{dx^2} = 0$$
$$\frac{df}{dt} = ax\exp(at)$$

From the stochastic chain rule we have

$$\begin{aligned} dy &= \left(-ax\frac{df}{dx} + \frac{df}{dt} + \frac{1}{2}\sigma^2\frac{d^2 f}{dx^2}\right)dt + \sigma\frac{df}{dx}dw \\ &= \left(-ax\exp(at) + ax\exp(at)\right)dt + \sigma\exp(at)dw_t \end{aligned}$$

Hence

$$dy = \sigma\exp(at)dw_t$$

OU Process

Hence

$$dy = \sigma \exp(at) dw_t$$

Now integrate from 0 to t

$$y_t - y_0 = \sigma \int_0^t \exp(as) dw_s$$
$$x_t \exp(at) = x_0 + \sigma \int_0^t \exp(as) dw_s$$

So

$$x_t = x_0 \exp(-at) + \sigma \int_0^t \exp(a(s-t)) dw_s$$

We then have

$$E[x_t] = x_0 \exp(-at)$$

The variance can be computed similarly (next week we will compute variance and covariance functions for univariate and multivariate linear SDEs).

FN Population

Rodriguez and Tuckwell (1996) consider the nonlinear stochastic spiking neuron model given by the addition of a stochastic noise term to the Fitzhugh-Nagumo equations

$$\begin{aligned} dv &= [f(v) - r + I]dt + \beta dw \\ dr &= b(v - \gamma r)dt \end{aligned}$$

where v is a voltage variable, r is a recovery variable, I is applied current

$$f(v) = kv(v - a)(1 - v)$$

and we use the following parameter values $a = 0.1$, $b = 0.015$, $\gamma = 0.2$, $I = 1.5$ and $\beta = 0.01$.

The above equation defines a nonlinear SDE.

Gaussian approximation

We consider the SDE

$$dx_i = f_i(x)dt + \sigma_i dw_i$$

where $f_i(x)$ defines the flow of the ith variable, and dw_i is a Wiener process. We have $i = 1..N$ variables.

The corresponding noise covariance matrix V has diagonal entries given by $V_{ii} = \sigma_i^2$, and zero off-diagonal entries.

Let

$$\begin{aligned} J_{ij}(x) &= \frac{df_i(x)}{dx_j} \\ H_{ijk}(x) &= \frac{d^2 f_i(x)}{dx_j dx_k} \end{aligned}$$

where J contains gradients (the 'Jacobian' matrix) and H is the curvature.

Gaussian approximation

The density over the state variables is approximated using a Gaussian

$$p(x) \sim N(x; \mu, C)$$

with mean μ and covariance C. Rodriguez and Tuckwell (1996) show using the stochastic chain rule that the dynamical equations for the moments can then be written

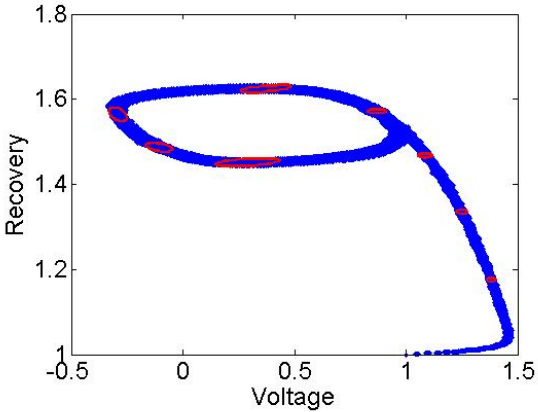
$$\begin{aligned} \dot{\mu}_i &= f_i(\mu) + \frac{1}{2} \sum_{j,k} H_{ijk}(\mu) C_{jk} \\ \dot{C}_{ij} &= \sum_k J_{ik}(\mu) C_{jk} + \sum_k J_{jk}(\mu) C_{ik} + V_{ij} \end{aligned}$$

In matrix notation this is

$$\begin{aligned} \dot{\mu} &= f(\mu) + \frac{1}{2} Tr[CH(\mu)] \\ \dot{C} &= J(\mu)C + CJ(\mu)^T + V \end{aligned}$$

FN Population

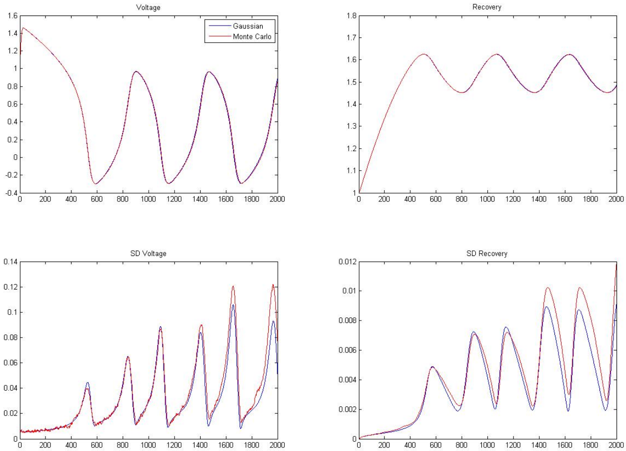
The blue mass indicates the trajectories of 64 EM sample paths. The red lines show probability contours for the Gaussian approximation for every 20 time steps (only the first eight plotted).



Uncertainty increases during periods of greatest flow increase (see effect of Jacobian on $\dot{C}$).

FN Population

The figure compares the evolution of the mean and variance of each of the states as computed using the Gaussian approximation versus EM simulations.



Fokker-Planck

Harrison et al. (2005) consider a stochastic integrate and fire neuron with a single excitatory synapse

$$\begin{aligned} \dot{x} &= f(x) + s(t) \\ s(t) &= h \sum_n \delta(t - t_n) \end{aligned}$$

where x is the membrane potential, t_n is the time of the nth incoming spike, and h is the magnitude of the post synaptic potential.

In a small time interval we can write

$$\begin{aligned} s(t) &= hr(t) \\ r(t) &= \frac{1}{T} \int_0^T \sum_n \delta(\tau - t_n) d\tau \end{aligned}$$

where $r(t)$ is the mean spike rate over that interval.

Ensemble density

We consider an ensemble of cells that receive a common driving current $f(x)$ and whose synapses receive spikes at the same average rate, but differ in the exact arrival time and number of spikes

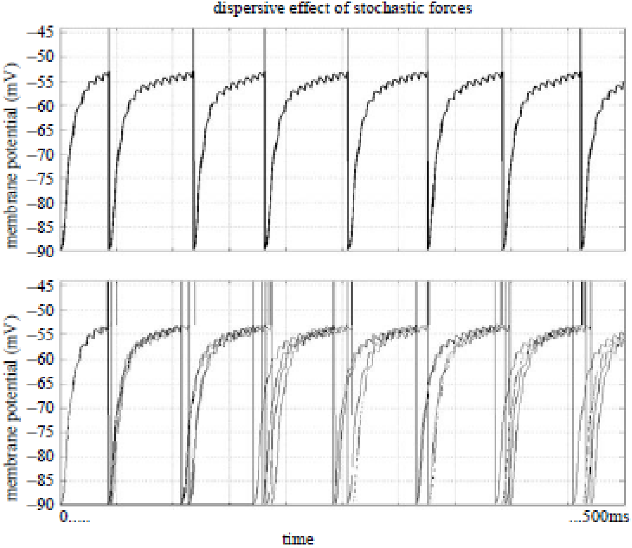
$$
\begin{aligned}
\dot{x} &= f(x) + s(t) \\
s(t) &= hr(t) \\
r(t) &= \frac{1}{T}\int_0^T \sum_n \delta(\tau - t_n) d\tau
\end{aligned}
$$

This induces a probability density, $p(x)$ over membrane potentials.

Additionally, the membrane potential is reset once a threshold is reached.

Evolution of ensemble density

The top plot shows a single sample path.

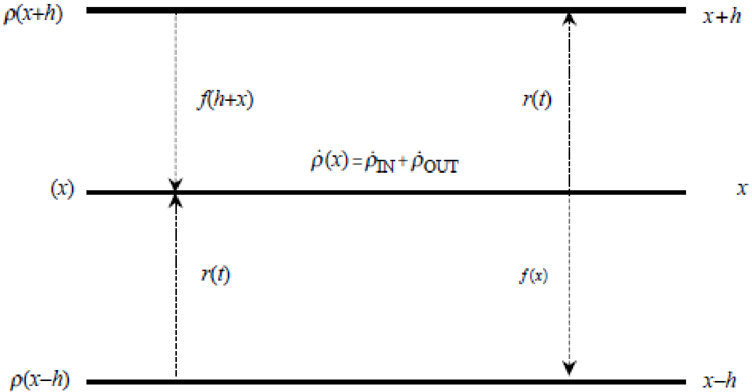


The bottom plot shows five sample paths.

Evolution of ensemble density

If we ignore the reset mechanism

$$\dot{\rho}_{IN} = \rho(x+h)f(x+h) + \rho(x-h)r(t) \qquad \dot{\rho}_{OUT} = -\rho(x)r(t) - \rho(x)f(x)$$



Evolution of ensemble density

$$
\begin{aligned}
\dot{p}_{in} &= -p(x+h)f(x+h) + p(x-h)r(t) \\
\dot{p}_{out} &= p(x)r(t) - p(x)f(x) \\
\dot{p} &= \dot{p}_{in} - \dot{p}_{out}
\end{aligned}
$$

Hence

$$\dot{p} = f(x)p(x) - f(x+h)p(x+h) + r(t)[p(x-h) - p(x)]$$

We can write the first term as

$$f(x)p(x) - f(x+h)p(x+h) = -\frac{d[f(x)p(x)]}{dx}$$

and for the second a Taylor series shows that

$$p(x-h) = p(x) - h\frac{dp(x)}{dx} + \frac{1}{2}h^2\frac{d^2p(x)}{dx^2}$$

Hence

$$\dot{p} = -\frac{d[f(x)p(x)]}{dx} - hr\frac{dp(x)}{dx} + \frac{1}{2}h^2 r\frac{d^2p(x)}{dx^2}$$

Evolution of ensemble density

$$\dot{p} = -\frac{d[f(x)p(x)]}{dx} - hr\frac{dp(x)}{dx} + \frac{1}{2}h^2 r\frac{d^2 p(x)}{dx^2}$$

Letting

$$\begin{aligned} s &= hr \\ \sigma^2 &= h^2 r \end{aligned}$$

we have

$$\dot{p} = -\frac{d[(f(x)+s)p(x)]}{dx} + \frac{1}{2}\sigma^2\frac{d^2 p(x)}{dx^2}$$

Writing the total flow as $g(x) = f(x) + s$ gives

$$\dot{p} = -\frac{d[g(x)p(x)]}{dx} + \frac{1}{2}\sigma^2\frac{d^2 p(x)}{dx^2}$$

This is the Fokker-Planck equation.

Fokker-Planck

The Fokker-Planck equation is

$$\dot{p} = -\frac{d[g(x)p(x)]}{dx} + \frac{1}{2}\sigma^2 \frac{d^2 p(x)}{dx^2}$$

Tuckell (1998) and Gerstner (2002) derive FP equations for populations of integrate and fire cells in a similar manner. They also allow for multiple synapse types and include an additional flux (flow of probability mass) for describing the spike reset mechanism.

Fokker-Planck

The Fokker-Planck equation can also be derived by applying Ito's rule and taking expectations.

We start with the SDE

$$dx = a(x,t)dt + b(x,t)dw$$

Following Ermentrout (p.292)

2AFC tasks

In a two-alternative forced-choice (2AFC) task, in which information is arriving discretely, the optimal decision is implemented with a sequential likelihood ratio test (SLRT).

Given uniform prior probabilities $p(r = L) = p(r = R) = 0.5$, a decision based on the posterior probability is identical to one based on the likelihood, or log-likelihood ratio

$$I_t = \log \left(\frac{p(X_t | r = L)}{p(X_t | r = R)} \right)$$

where $X_t = [x_1, x_2, ..., x_t]$ comprises all data points up to time t. This can be accumulated sequentially as

$$I_t = I_{t-1} + \delta I_t$$

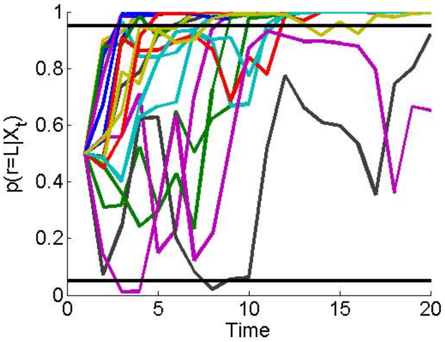
where

$$\delta I_t = \log \left(\frac{p(x_t | r = L)}{p(x_t | r = R)} \right)$$

2AFC tasks

The posterior is a sigmoid function of the accumulated log-likelihood ratio.

$$
\begin{aligned}
p(r = L|X_t) &= \frac{p(X_t|r = L)}{p(X_t|r = L) + p(X_t|r = R)} \\
&= \frac{1}{1 + \exp(-l_t)}
\end{aligned}
$$

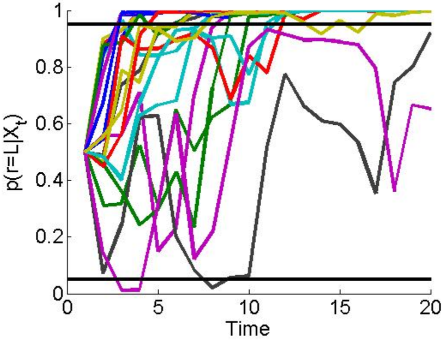


A left decision is made when the posterior exceeds β. A right decision is made if it exceeds $1 - \beta$.

2AFC tasks

Unequal priors can be accomodated using

$$l_0 = \log \frac{p(r = L)}{p(r = R)}$$



Continuous Time

In continuous time this is equivalent to a drift diffusion model (DDM) (Bogacz, 2006).

$$dI = adt + cdw$$

As we have seen this corresponds to a Wiener process.

Additionally we assume that $x_0 = 0$ and that a positive/negative decision (eg left/right button press) is made if x crosses z before/after $-z$.

The decision time (DT) is the time at which the crossing occurs.

Continuous Time

The correct rate, CR, and mean decision time, DT, are then given by analytic expressions (Bogacz et al 2006)

$$
\begin{aligned}
CR &= \frac{1}{1 + \exp(-2az/c^2)} \\
DT &= \frac{z}{a} \tanh\left(\frac{az}{c^2}\right) \\
z &= \log\left(\frac{\beta}{1-\beta}\right)
\end{aligned}
$$

These formulae can be derived using the backward Fokker-Planck equation (Gardiner, 1983; Moehlis et al. 2004).

Continuum Limit

In a small time interval Δt the mean and variance of the Wiener process are

$$\begin{aligned} E[\Delta I] &= a\Delta t \\ Var[\Delta I] &= c^2 \Delta t \end{aligned}$$

In the discrete time model we have

$$\Delta I = \log \frac{p(x|r_L)}{p(x|r_R)}$$

For Gaussian likelihoods with means μ_L for left and μ_R for right, with common variance σ^2

$$\begin{aligned} E[\Delta I] &= \frac{(\mu_L - \mu_R)^2}{2\sigma^2} \\ Var[\Delta I] &= \frac{(\mu_L - \mu_R)^2}{\sigma^2} \end{aligned}$$

Equating moments gives

$$\frac{a}{c^2} = \frac{1}{2}$$

Continuous versus Discrete

Because

$$\frac{a}{c^2} = \frac{1}{2}$$

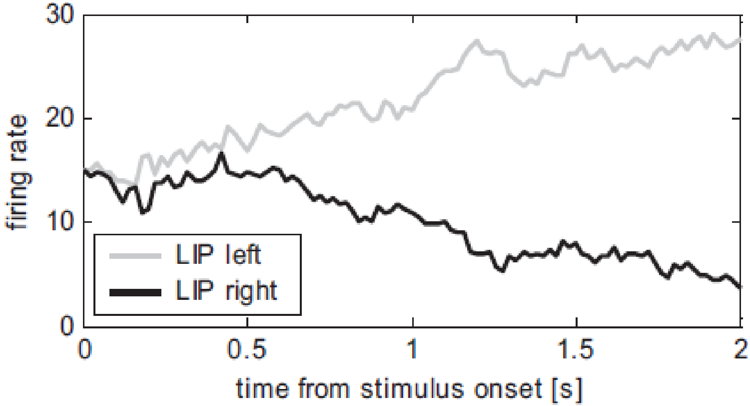
we have

$$\begin{aligned} CR &= \beta \\ DT &= \frac{z}{a}(2\beta - 1) \\ z &= \log\left(\frac{\beta}{1-\beta}\right) \end{aligned}$$

This is intuitively satisfying because the correct rate is simply equal to the probability threshold β.

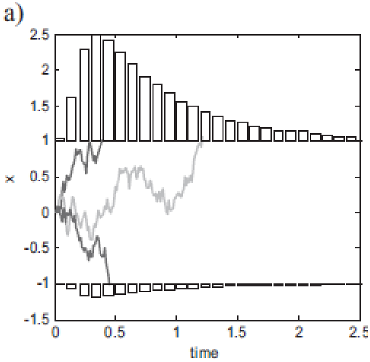
Drift-Diffusion Models

This sort of DDM behaviour has been observed among the firing rates of various cells in 2AFC tasks.



Decision Making Models

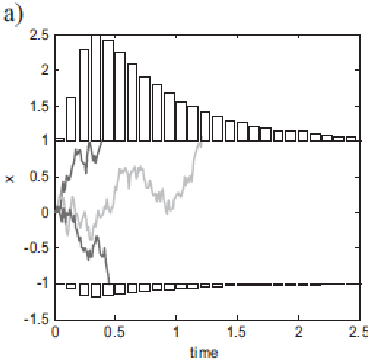
Can get analytic results for whole distribution of RTs (not just mean).



DDMs can be fitted to behavioural data (quantiles of RTs and error rates) to estimate a, σ and z. These can be used as regressors in computational fMRI.

Decision Making Models

Similar results can be derived for OU processes (Moehlis, 2004).



OU models better describe neuronal implementions where evidence for left versus right decisions are accumulated in separate populations which inhibit each other (Bogacz, 2006; Wang 2002).

Integration

Integration of deterministic functions is defined via the Riemann sum

$$\int_0^T h(t)dt = \lim_{N\to\infty} \sum_{j=1}^{N} h(t_j)[t_{j+1} - t_j]$$

for $t_j = jT/N$.

Integration of SDEs is defined by Ito's rule

$$\int_0^T h(t)dw_t = \lim_{N\to\infty} \sum_{j=1}^{N} h(t_j)[w(t_{j+1}) - w(t_j)]$$

where $w(t_j)$ are sample paths. This is a stochastic equivalent of the Riemann sum. Using $h([t_{j+1} + t_j]/2)$ here leads to Stratanovich's rule (Gardiner, 1983).

This definitition is necessary, for example, to compute expectations of stochastic processes.

References

R. Bogacz et al (2006) The physics of optimal decisions making: a formal analysis of models of performance in two-alternative forced-choice tasks. Psychological Review 113, 700-765.

C. Gardiner (1983) Handbook of Stochastic Methods. Springer-Verlag, 1983.

W. Gerstner and W. Kistler (2002) Spiking Neuron Models: Single Neurons, Populations, Plasticity. Cambridge.

L. Harrison et al. (2005) Stochastic models of neuronal dynamics. Phil. Trans. R. Soc. B. 360, 1075-1091.

D. Higham (2001) An algorithmic introduction to numerical simulation of stochastic differential equations. SIAM Review 43, 525-546.

J. Moehlis et al (2004) Optimising reward rate in two alternative forced-choise tasks: mathematical formalism. Technical Report, Princeton University.

R. Rodriguez and H. Tuckwell (1996) Statistical properties of stochastic nonlinear dynamical models of single spiking neurons and neural networks. Phys Rev E 54, 5585-5590.

H. Tuckwell (1998) Introduction to Theoretical Neurobiology Vol2: Nonlinear and Stochastic Theories. Cambdridge.

X. Wang (2002) Probabilistic decision making by slow reverberation in cortical circuits. Neuron 36, 1-20.