OVERVIEW

First set of lectures: Basics of Brownian motion, Stochastic Differential Equations, Ito calculus, connection with jump processes, Fokker-Planck equations and related PDE's

References:
Oksendal, Stochastic Differential Equations: An Introduction with Applications
Schuss, Introduction to Stochastic Differential Equations
Gardiner, Handbook of Stochatics Methods for Physics and Chemistry

Second set of lecturers: Asymptotic approximations, including large number of events, different time scales, small noise, and stochastic averaging

References:
Stuart and Pavliotis, Multiscale methods: Homogenization and Averaging
Bender and Orszag, Advanced Mathematical Methods for Scientists and Engineers
(Also references with specific examples)

1 From Einstein to Ito

Motivating Examples

1.1 Basic stochastic models: continous time and state space

Characteristics of the noise:

Example 1: Motion of a Brownian particle:

$$mx_{tt} = -\eta x_t + X \tag{1}$$

This is just $F = ma$. x = particle position $x_t = \frac{\partial x}{\partial t}$
X = random fluctuations

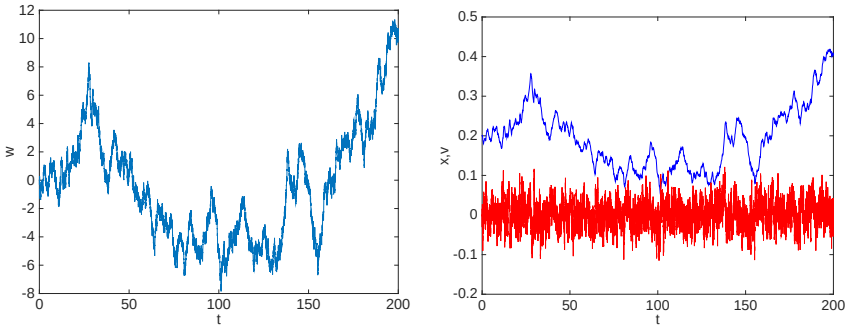


Figure 1: Realization for noise process, particle position, and velocity

Langevin gave an alternative (heuristic) derivation to that of Einstein
Here we include some physical motivation, leading to mathematical definition:
m=mass
η = viscous drag
X= irregular fluctuations: $\langle X \rangle = 0$.

From statistical mechanics: $\frac{1}{2}m\left\langle v^2\right\rangle = \frac{1}{2}m\left\langle x_t^2\right\rangle = \frac{kT}{2}$ as $t \to \infty$
$T =$ temperature
(k is Boltzmann's constant) Then averaging the equation or x gives: $x \to x_0$ as $t \to \infty$.

Note: we are averaging over realizations, $\langle f(x)\rangle = \int f(x)p(x)dx$ for $p(x)$ the (stationary) density of x. The variation about initial condition has mean zero $E[x - x_0] = \langle (x - x_0)\rangle$
Now let's look at the variance $E[x^2]$: (take $x_0 = 0$ for simplicity)

Multiply equation of motion by x, integrate:

$$\Rightarrow mxx_{tt} = -\eta x x_t + X \cdot x \tag{2}$$

Here we have assumed X is mathematically "nice"

$$\Rightarrow \frac{m}{2}\left((x^2)_{tt} - 2(x_t)^2\right) = -\frac{\eta}{2}(x^2)_t + X \cdot x \tag{3}$$

Note: Need to determine features of X, and determine $\left\langle x^2\right\rangle = E[x^2]$.
Averaging over realizations and using the result for $\frac{1}{2}m\left\langle v^2\right\rangle$ gives

$$\frac{m}{2}\left\langle x^2\right\rangle_{tt} = -\frac{\eta}{2}\left\langle x^2\right\rangle_t + kT + \cancelto{0}{\langle X\rangle}\langle x\rangle \tag{4}$$

assuming independence of X and x

Then, we integrate the equation for $\left\langle x^2\right\rangle_t$ to get

$$\left\langle x^2\right\rangle_t = \frac{2kT}{\eta} + Ce^{-\frac{\eta t}{m}} \tag{5}$$

$$\Rightarrow \quad \text{as} \quad t \to \infty \quad \left\langle x^2\right\rangle \to \frac{2kT}{\eta}t \tag{6}$$

Then the variance of the particle position $\langle x^2 \rangle$ behaves as a constant times t

Einstein had derived a similar result, looking for probability density corresponding to diffusion (later).

Note that Langevin's derivation assumes that the usual rules of calculus apply. We will see that is a big assumption, which in general is not the case.

Highlights for the behaviour of X

We consider the equation for the velocity $v = x_t$, to look more closely at the implications for the behavior of the noisy term X What are these fluctuations?

Write the equations of motion of a Brownian particle:

$$mv_t = -\eta v + X \tag{7}$$

Formal solution for v: (assuming X is a "nice" function)

$$v = v_0 e^{-\frac{\eta}{m}t} + \frac{1}{m}\int_0^t e^{-\frac{\eta}{m}(t-s)} X(s) ds \tag{8}$$

so we can consider the variance of v about its mean:

$$\left\langle \left(v - v_0 e^{-\frac{\eta}{m}t} \right)^2 \right\rangle = \left\langle \frac{1}{m^2} \left(\int_0^t e^{-\frac{\eta}{m}(t-s)} X(s) ds \right)^2 \right\rangle \tag{9}$$

Let's now compare the average of this term $(\int \dots ds)^2$ to the result $\frac{1}{2} m \left\langle v^2 \right\rangle = \frac{kT}{2}$

Thinking of the integral as the limit of a sum, we write

$$\int_0^t e^{-\frac{\eta}{m}(t-s)} X(s) ds = \lim_{N\to\infty} \sum_{j=1}^{N} e^{-\frac{\eta}{m}(t-j\Delta s)} \Delta b_j, \tag{10}$$

taking $X(s) \approx \frac{\Delta b}{\Delta s}$

Here b is used, as we are heading toward identifying Brownian motion
That is, we view X as the derivative of random fluctuations (does not exist as $\Delta s \to 0$).

Assuming the Δb_j are independent (independent increments in the random process) and $\left\langle \Delta b_j^2 \right\rangle = q\Delta s$ yields

$$\left\langle \left(\sum_{j=1}^{N} e^{-\frac{\eta}{m}(t-j\Delta s)} \Delta b_j \right)^2 \right\rangle = \sum_{j=1}^{N} e^{-\frac{2\eta}{m}(t-j\Delta s)} q\Delta s + \sum_{j \neq k} () \langle \Delta b_j \Delta b_k \rangle \tag{11}$$

For independent increments, the last term vanishes, yielding, as $\Delta s \to 0$, an integral

$$\int_0^t e^{-\frac{2\eta}{m}(t-s)} q ds$$

which can be easily evaluated as $\text{Const} \cdot (1 - e^{-\frac{2\eta}{m}t})$.

Then, $\frac{m}{2} \left\langle v^2 \right\rangle = const$, as $t \to \infty$. We can then choose q appropriately so that

$$m \left\langle v^2 \right\rangle = kT.$$

So we have $X(s)$ as a "derivative" of the random fluctuations, such that $\left\langle (X(s)\Delta s)^2 \right\rangle = \Delta s$ and, furthermore, have assumed x is independent of X ($\langle Xx \rangle = \langle X \rangle \langle x \rangle$ previously). Essentially this follows from the assumption of independent increments of the random fluctuations.

Note: We have a definition of X (random part) that is consistent with the physical behaviour of v
These calculations illustrate several key mathematical properties of Brownian motion = standard Wiener process $w(t)$:

- The increments $w(t+s)-w(t)$ is independent of t and independent of $w(t)-w(t-u)$ $(u \geq 0, s \geq 0)$
- Paths of $w(t)$ are continuous ($X(t)$ undefined, but integrated, gives continuous paths for w)
- Joint probability distribution of $(w(t_1), w(t_2), ..., w(t_n))$ is mean zero Gaussian, and $E[w^2(t)] = t$ (normalized)

Then $w(t)$ is distributed as $N(0, \sqrt{t})$, with $\mathrm{Var}(w) = t$, and covariance:

$$E[w(s)w(t)] = E([w(t) - w(s)]w(s)) + E[w^2(s)] = 0 + s = \min(t, s) \quad (\text{for } s < t) \tag{12}$$

Simulation of SDE's and Langevin-type stochastic models:
Langevin-type models

$$dx = f(x)dt + qdw \tag{13}$$

$f(x)$ is known as the drift
q is the diffusion coefficient

We write this in differential form, rather than

$$\frac{dx}{dt} = f(x) + q\frac{dw}{dt} \tag{14}$$

since w is continuous but not differentiable.

General Stochastic Differential Equations (SDE):

$$dx = A(x, t)dt + B(x, t)dw \tag{15}$$

The behavior of the increment dw suggests a simple approach for numerical simulation:

Simulate at discrete time steps: $t_j = j\Delta t$

Discrete approximation for the derivative:

$$\frac{dx}{dt} \approx \frac{x(t+\Delta t) - x(t)}{\Delta t}$$

Behavior of dw:

$$dw \approx \Delta w \sim N(0, \sqrt{\Delta t})$$

Then,

$$x(t+\Delta t) = x(t) + A(x(t),t)\,\Delta t + B(x(t),t)\,\sqrt{\Delta t} Z$$

where $Z \sim N(0,1)$.

In general, writing $x_j = x(j\Delta t)$ we have an iterative procedure to get x_j for all j, starting with an initial condition $x = x_0$ and computing to a time $T = K\Delta t$.

$$\begin{aligned} &x(1) = x_0; \\ &\text{for } j = 1 : K-1 \\ &x_{j+1} = x_j + A(x_j, t_j)\,\Delta t + B(x_j, t_j)\,\sqrt{\Delta t} Z_j \\ &\text{end} \end{aligned}$$

Z_j are $K-1$ independent $N(0,1)$ random variables
This is the Euler-Maruyama method. It is generally associated with the Ito interpretation of SDE's (later). This has implications for the dynamics via the interpretation of the noise increments.

Higher order methods: Reference: Kloeden and Platen, Numerical Solution of SDE's, 1992. Both Weak (in distribution) and Strong (pathwise) methods.

For higher dimensional SDE:

$$d\mathbf{x} = \mathbf{A}(\mathbf{x}, t)dt + \mathbf{B}(\mathbf{x}, t)d\mathbf{w} \tag{16}$$

$\mathbf{x}, \mathbf{A}$ are d-dimensional, $\mathbf{w}$ is n-dimensional, B is $d \times n$ dimensional.

Example:

$$mx_{tt} = -\eta x_t + qX \tag{17}$$

For X as above, we write

$$\begin{aligned} dx &= vdt \\ dv &= -\frac{\eta}{m} vdt + \frac{q}{m} dw \end{aligned}$$

Note: the drift is linear (in v), diffusion coefficient is constant
v is an Ornstein-Uhlenbeck (O-U) process - linear drift and constant diffusion coefficient

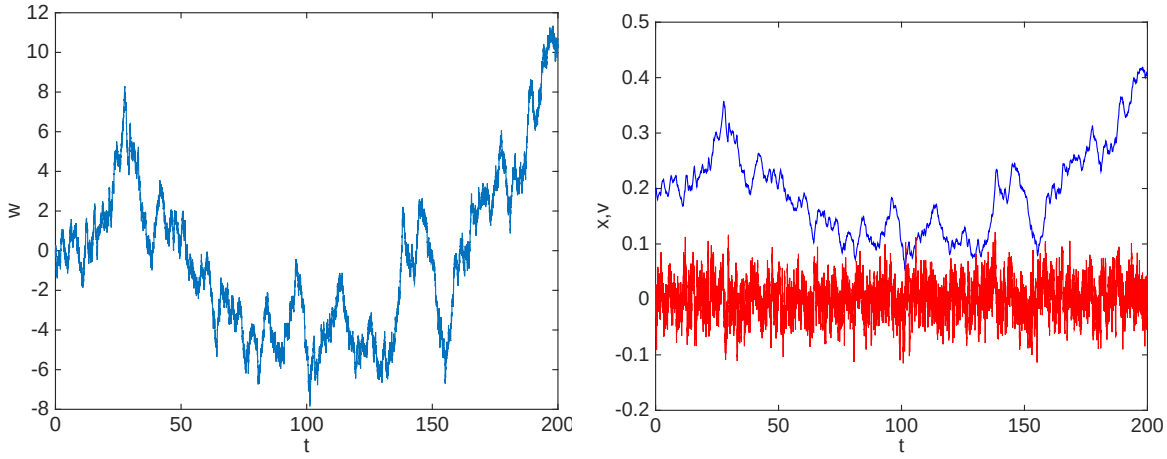


Figure 2: Realization for noise process, particle position, and velocity

$$
\begin{aligned}
&x(1) = x_0; v(1) = v_0 \\
&\text{for } j = 1 : K - 1 \\
&x_{j+1} = v_j \Delta t \\
&v_{j+1} = v_j - \frac{\eta}{m} v_j \Delta t + \frac{q}{m} \sqrt{\Delta t} Z_j \\
&\text{end}
\end{aligned}
$$

Note: v is an Ornstein-Uhlenbeck process (OU process). It has a linear drift term and a constant coefficient for the diffusion term. OU processes have a Gaussian stationary density with a constant mean and variance, i.e. normally distributed. That is, as $t \to \infty$, $v \sim N(0, q/\sqrt{2\eta})$.

Dynamics driven by Jump Processes - "Shot Noise"

$$\frac{dI}{dt} = -\alpha I + q\mu(t) \tag{18}$$

$$\mu(t) = \sum \delta(t - t_k) \tag{19}$$

I=current, t_k=arrival time of electron. Note that $\mu(t)$ is not mean zero.

Define: $N(t)$=sum of Poisson arrivals,

$$\frac{dN}{dt} = \mu \tag{20}$$

formally, but N is a step function. If we want to write the current equation in terms of a mean zero process (analogous to X in previous example) then consider $E[dN] = \lambda dt$, $Var[dN] = \lambda dt$, where λ is the mean arrival rate.

Note: For a Poisson process N with arrival rate λ, then the average number of arrivals in an interval of length t is λt.

Then

$$d\eta = dN - \lambda dt$$

is a centered Poisson increment, with mean zero, and variance λdt

Formally

$$\frac{dI}{dt} = -\alpha I + \lambda q + q\frac{d\eta}{dt} \tag{21}$$

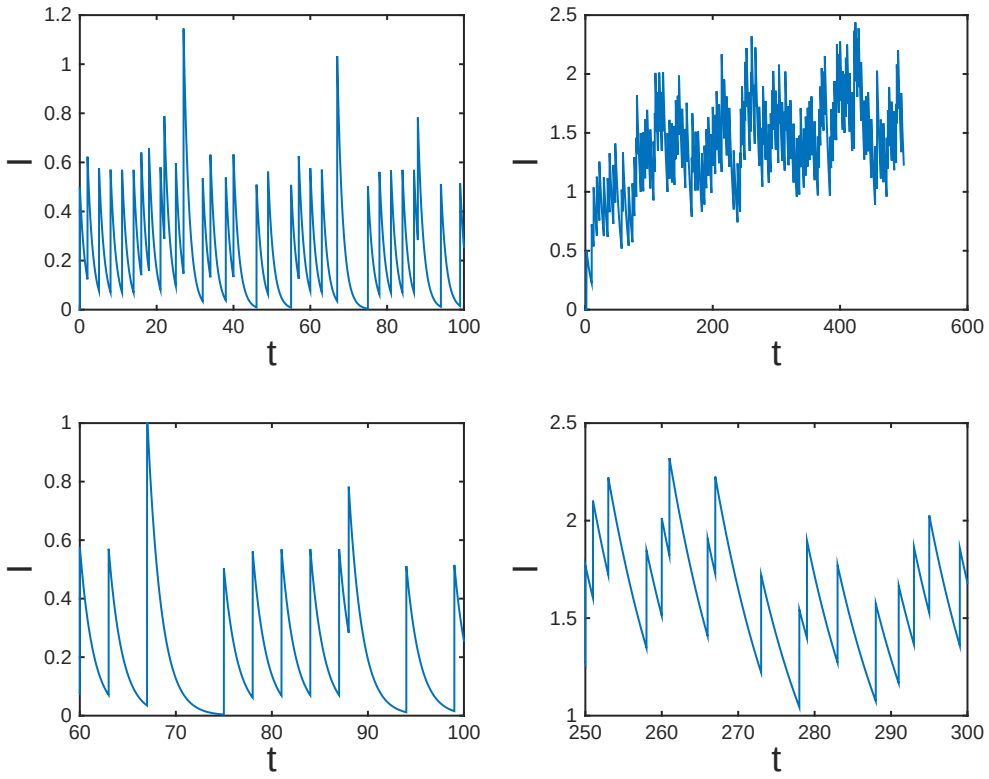


Figure 3: Realization for I. On the left, the rate of jumps is smaller compared with the decay rate of I, on the right the jumps are more frequent. The bottom row figures are zooms of the realizations in the top row.

Note on Numerical simulation

Here η is not a continuous process, rather, events occur at times that are exponentially distributed. Then some time interval increments of length Δt will have an event, and some will not.

Averaging yields

$$\frac{d\langle I\rangle}{dt} = (-\alpha\langle I\rangle + \lambda q), \quad \langle I\rangle \to \frac{\lambda q}{\alpha} \quad \text{as} \quad t \to \infty \tag{22}$$

What about $\langle I^2\rangle$?

Suppose we consider $d(I^2)$ and compare with equation above. We can write $d(I^2)$ in a couple of different ways:

$$d(I^2) = (I + dI)^2 - I^2 = 2IdI + (dI)^2 \tag{23}$$

and

$$(dI)^2 = ((-\alpha I + \lambda q)dt)^2 + 2(-\alpha I + \lambda q)dtd\eta + q^2 d\eta^2 \tag{24}$$

Compare with heuristic as in previous example: multiply equation (21) for I by I and integrate and average:

$$\frac{1}{2}\frac{d\left\langle I^2\right\rangle}{dt} = -\alpha\left\langle I^2\right\rangle + \lambda q\langle I\rangle + q\left\langle I\frac{d\eta}{dt}\right\rangle \tag{25}$$

What about $\left\langle I\frac{d\eta}{dt}\right\rangle$? Is it $= 0$? If so, then we can solve for $\left\langle I^2\right\rangle$ as $t \to \infty$ (left as an exercise).

$$\begin{aligned} \left\langle I^2\right\rangle &= \left(\frac{\lambda q}{\alpha}\right)^2 = \langle I\langle^2 \\ \Rightarrow Var(I) &= \langle I^2\rangle - \langle I\rangle^2 = 0 \end{aligned}$$

! No effect of noise - not correct!

This heuristic neglects the term $(dI)^2$ - it is not in the equation for $\left\langle I^2\right\rangle$, obtained from (23)-(24). So we need to examine that equation

$$(dI)^2 = (\quad)(dt)^2 + (\quad)dt\langle d\eta\rangle + q^2\left\langle d\eta^2\right\rangle \tag{26}$$

using the properties of the centered Poisson increment.

$$\langle d\eta \rangle = 0 \qquad \left\langle (d\eta)^2 \right\rangle = \lambda dt$$

Note that the centered Poisson increment $d\eta$ has the same variance as the original Poisson increment dN.

So now, combining results from the expression above, we can get the correct equation for I^2:

$$\begin{aligned} d\left\langle (I)^2 \right\rangle &= 2\langle I dI \rangle + \left\langle (dI)^2 \right\rangle && (27) \\ &= \langle (2I(-\alpha I + \lambda q))dt \rangle + \cancelto{0}{\langle 2qI d\eta \rangle} + q^2\lambda dt + o(dt) && (28) \\ \Rightarrow\ Var\left\langle I^2 \right\rangle &= \frac{q^2\lambda}{2\alpha} \qquad t \to \infty && (29) \end{aligned}$$

Recall that for the Wiener process $\langle w(t) \rangle = 0$ and $\left\langle w^2 \right\rangle = t$. Later we see that the if Poisson events are very frequent, then one can replace Poisson random variables with appropriate "centering" and Brownian motion (normal random variables).

Other types of increments: e.g. Levy Processes

Levy Processes can be characterized by a combination of Brownian motion and jumps.

A simple class of Levy Processes are alpha-stable processes. Instead of having a density with exponentially decaying tails, like the Gaussian distribution, they have "fat" tails; that is, the tails of the density have the behavior

$$p(x) = |x|^{\alpha+1} \ \text{ as } x \to \infty \ \text{ for } \alpha < 2$$

To simulate an SDE with alpha-stable noise increments, e.g.

$$dx = H(x)dt + \kappa dL^{\alpha,\beta} \tag{30}$$

the iterative steps of the Euler-type method take the form

$$x_{n+1} = x_n + H(x_n)\Delta t + \kappa\delta L_n^{\alpha,\beta} \qquad \Delta L_n^{\alpha,\beta} \sim \mathcal{S}_\alpha(\beta, (\Delta t)^{1/\alpha}) \tag{31}$$

The alpha-stable distribution is given in terms of its characteristic function, rather than the density, so $X \sim \mathcal{S}_\alpha^{\beta,\sigma}$ has a characteristic function of the form

$$\begin{aligned} \psi(k) &= \mathcal{F}[p(x)] = \int e^{ikx} p(x) dx = \exp(-\sigma^\alpha |k|^\alpha (1 - i\beta \mathrm{sgn}(k)\phi(k))) \\ \phi(k) &= = \begin{cases} \tan\left(\frac{\pi\alpha}{2}\right) & \alpha \neq 1 \\ -\frac{2}{\pi}\log|k| & \alpha = 1 \end{cases} \end{aligned} \tag{32}$$

Note: for $\alpha = 2$, we have a Gaussian distribution.

Additional details: Dybiec, Gudowska-Nowak, Hanggi, Phys. Rev E, 2006; Asmussen and Glynn, 2007.

2 Ito's Formula

Now we ask:

What are the rules for functions of $w(t)$: $f(w(t))$?

In particular, we are interested in how to treat $df(w(t))$, write stochastic differential equations including $w(t)$

Ito's formula: find equation for $df(x)$ given $dx = a\,dt + b\,dw$

Review of Chain Rule for deterministic functions: if $x'(t) = g(x,t)$, what is df for $f(x,t)$?
Simple example: $y = w^2$

$$\int_s^t dy = w^2(t) - w^2(s) \tag{33}$$

Can we write $dy = a\,dt + b\,dw$? Or,

$$\int dy = \int a dt + \int b dw \tag{34}$$

where a,b may be functions of y,t?

We need to define bdw: how do we evaluate or interpret $b\,dw$?

Ito interpretation:

$$\int_s^t b(w) dw = \sum_j b(w(t_j))(w(t_{j+1}) - w(t_j)) \tag{35}$$

Note that this is the discretization used in the Euler-Maruyama numerical method.

Stratonovich interpretation:

$$\int_s^t b(w)dw = \sum_j b\left(\frac{w(t_j)+w(t_{j+1})}{2}\right)(w(t_{j+1})-w(t_j)) \tag{36}$$

where b is evaluated at the average of $w(t_j)$ and $w(t_{j+1})$

Back to the equation for $y = w^2$: We compare $\int dy$ with $\int w\,dw$ - we might expect these two expressions to be related from the usual rules of calculus. (Why?)

If we use the Ito interpretation, then we rewrite it in a convenient way

$$\begin{aligned}\int_s^t w\,dw &= \frac{1}{2}\left[\sum_j w^2(t_{j+1}) - w^2(t_j) - (w(t_{j+1})-w(t_j))^2\right] &(37)\\ &= \frac{1}{2}(w^2(t)-w^2(s)) - \frac{1}{2}\sum_j (w(t_{j+1})-w(t_j))^2 &(38)\end{aligned}$$

This looks like an integrated quantity evaluated at the endpoints t and s plus a sum of dw_j^2

Considering the expected value of the sum:

$$\begin{aligned}E\left[\sum_{j=1}^N (w(t_{j+1})-w(t_j))^2\right] &= \sum_{j=1}^N E\left[w^2(t_{j+1})\right] + E\left[w^2(t_j)\right] - 2\min(t_{j+1},t_j)\\ &= \sum_{j=1}^N t_{j+1}+t_j-2t_j = \sum_{j=1}^N \Delta t = t-s\end{aligned}$$

where we have used the expression for the covariance of w (12). Furthermore, one can show that

$$Var\left[\sum_{j=1}^N (w(t_{j+1})-w(t_j))^2\right] \sim \frac{1}{N} \tag{39}$$

(reference: Schuss, Introduction to SDE's)

So

$$2\int_s^t w dw = w^2(t) - w^2(s) - t - s = y(t) - y(s) - t - s \tag{40}$$

in probability, as $N \to \infty$. Then for $y = w^2$,

$$dy = \underbrace{2w}_{b} dw + \underbrace{1}_{a=1} dt \tag{41}$$

Note: The expression is not simply $dy = 2w\,dw$, but there is also an additional term dt.

Next, we compute the rule for products

Consider $dx = adt + bdw$ and dw

For a,b constant, $x = at + bw$, for $x(0) = 0$ (using the Ito interpretation)

$$\begin{aligned} xw &= atw + (bw^2) && (42)\\ d(xw) &= atdw + awdt + b(2wdw) + bdt && (43)\\ &= xdw + wdx + bdt && (44) \end{aligned}$$

And we can use this, together with the result for $y = w^2$ to find by induction (exercise)

$$dw^m(t) = m(w(t))^{m-1}dw + \frac{m(m-1)}{2}w^{m-2}dt \quad \text{for} \quad m \geq 2 \tag{45}$$

Then, for any *Polynomial* $P(w)$

$$dP(w) = P'(w)dw + \frac{1}{2}P''(w)dt \tag{46}$$

From there, we can write general "noise" functions as $f = g(t)\phi(w)$, so

$$df = \phi g'(t)dt + gd\phi \tag{47}$$

$$\Rightarrow df = \left[\phi g'(t) + \frac{1}{2}g\phi''(w)\right]dt + g\phi'(w)dw \tag{48}$$

$$= \left[f_t + \frac{1}{2}f_{ww}\right]dt + (f_w)dw \tag{49}$$

Applying this term by term to a general expression,

$$f(w,t) = \sum g(t)\phi(w), \tag{50}$$

yields the same result for $f(w,t)$
Finally, for $f(x,t)$, $x = at + bw$, a,b, constant

$$\frac{\partial f}{\partial t} = \frac{\partial f}{\partial t} + a\frac{\partial f}{\partial x} \tag{51}$$

$$\frac{\partial f}{\partial w} = b\frac{\partial f}{\partial x} \qquad \frac{\partial^2 f}{\partial w^2} = b^2\frac{\partial^2 f}{\partial x^2}$$

Then

$$df = \left[\frac{\partial f}{\partial t} + a\frac{\partial f}{\partial x} + \frac{1}{2}b^2\frac{\partial^2 f}{\partial x^2}\right]dt + b\frac{\partial f}{\partial x}dw \tag{52}$$

This is Ito's formula, relating the equation for f to the equation for x, $dx = adt + bdw$. This can be viewed as the appropriate chain rule for determining the differential of $f(x,t)$ when $dx = adt + bdw$ is interpreted in the Ito sense.

It can be generalized for $a(x,t)$ and $b(x,t)$.

Note: Throughout this we use the Ito interpretation for $()dw$.

Is there a Stratonovich Formula?

(S) $\int f(w)\,dw$ follows the usual rules of calculus

Using $(w_i+w_{i+1})/2 = w_i+\Delta w_i/2$ (S) $\int f(w)\,dw = \sum f\left(\frac{w_i+w_{i+1}}{2}, t_i\right)\Delta w_i \sim \sum f(w_i)\Delta w_i+\frac{(\Delta w_i)^2}{2}f'(w_i)$
so we get the Ito interpretation + a correction
Simple example:

$$(S)\ w\,dw = w\,dw + \frac{(\Delta w)^2}{2} \sim w\,dw + \frac{\Delta t}{2} = \frac{1}{2}d(w^2) \tag{53}$$

The general relationship between Ito and Stratonovich interpretations: $dx = adt + bdw_{(S)} = adt + b\circ dw = (a + \frac{1}{2}bb_x)dt + bdw$

Note: for the simple example above in (53) $2x = w^2$, so $b(x) = w = \sqrt{2x}$, and $bb_x = 1$

How are Stratonovich and Ito intergrals related for $b =$ const?

Exercise: Demonstrate this for a general expression for dx? What does the result look like for products and polynomials, as considered in deriving Ito's formula.

Other Processes:

For Levy Processes, the analogy for the Stratonovich interpretation is known as the Marcus interpretation. Here, written in an SDE with increments of an alpha-stable process

$$dx \;=\; a(x)dt + b(x)\diamond dL^{\alpha,\beta} \tag{54}$$

For $\alpha = 2$ it is equivalent to the Stratonovich interpretation.

To implement this for evaluation/simulation, the Marcus integral takes the form

$$\int_0^t b(x_s)\diamond dL_s^{\alpha,\beta} = \sum_{s\leq t} \theta(1;\Delta L_s, x_s^-) - x_s^-$$
$$\frac{d\theta}{dr} = \Delta L_x b(\theta),\ \theta(0) = z_s^- \tag{55}$$

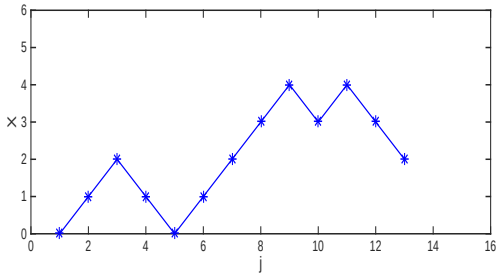
The quantity θ here is known as the Marcus map, and it is a time-like variable that travels infinitely fast along a curve connecting across the jumps. Then, a numerical approximation of $dx = a(x)dt + b(x)\diamond dL^{\alpha,\beta}$ is given by

$$x_{n+1} = x_n + a(x_n)\Delta t + (\theta(1;\Delta L^{\alpha,\beta}, x_n) - x_n] \tag{56}$$

Additional references on implementation:
Grigoriu, Phys. Rev. E, 2009.
Asmussen and Glynn, Stochastic Simulation: Algorithms and Analysis, 2007.

3 Master and Forward equations

Random walk as a Markov process



We define X_n as a random walk, where X_n =position after n steps, with probabilities of taking a step up or down:

P(step up: $X_{n+1} = X_n + 1$) = p

P(step down: $X_{n+1} = X_n - 1$) = 1 - p

Each step is independent of previous steps (Markovian) (depends only on present location), so we can write $P(X_n = k)$ in terms of the conditional probability, conditioning on the location at the previous step:

$$\begin{aligned} P(X_n = k) &= P(X_n = k|X_{n-1} = k-1)P(X_{n-1} = k-1) \\ &\quad + P(X_n = k|X_{n-1} = k+1)P(X_{n-1} = k+1) \\ &= pP(X_{n-1} = k-1) + (1-p)P(X_{n-1} = k+1) \end{aligned}$$

In general,

$$\begin{aligned} P(X_n = k) \;=\; & P(\text{ location after n steps, } m \text{ + steps, } l \text{ - steps,} \\ & \text{with } m - l = k,\, m + l = n) \end{aligned}$$

Now let's look at how we can use this to derive an equation for $P(X_n)$, and to derive a partial differential equation for this probability in the continuum limit - that is the limit as we "zoom out", so steps and time intervals shrink.

Random walk: use the shorthand, $P(k,j) = P(X_j = k)$

$$\begin{aligned} P(k,j) &= P(X_j = k) = pP(X_{j-1} = k-1) + (1-p)P(X_{j-1} = k+1) \\ &= pP(k-1, j-1) + (1-p)P(k+1, j-1) \end{aligned}$$

For $p = \frac{1}{2}$ (symmetric) we rewrite this as:

$$\begin{aligned} P(k,j) - P(k,j-1) &= \frac{1}{2}[P(k-1,j-1) + P(k+1,j-1) - \\ &\qquad 2P(k,j-1)] \end{aligned}$$

It doesn't matter what the grid size is, so, take time step size Δt, space step size ΔX,

$$\begin{aligned} P(X_k,t_j) - P(X_k,t_j - \Delta t) &= \frac{1}{2}[P(X_k - \Delta X, t_j - \Delta t) + \\ &\qquad P(X_k + \Delta X, t_j - \Delta t) - 2P(X_k, t_j - \Delta t)] \end{aligned}$$

Suppose

$$(\Delta X)^2 = C\Delta t$$

Then the equation for $P(X_k, t_j)$ becomes (for all k):

$$\frac{P(X_k,t_j) - P(X_k,t_j - \Delta t)}{C\Delta t} = \frac{1}{2(\Delta X)^2}[P(X_k - \Delta X, t_j - \Delta t) + \tag{57}$$

$$P(X_k + \Delta X, t_j - \Delta t) - 2P(X_k, t_j - \Delta t)] \tag{58}$$

$$\implies P_t = \frac{C}{2}P_{XX} \qquad \text{as } \Delta t,\ (\Delta x)^2 \to 0 \tag{59}$$

Exercise: Show how to obtain (59) via a Taylor series expansion about $\Delta t = 0$ and $\Delta X = 0$, with relationship $\Delta t \propto (\Delta X)^2$.

This PDE is the diffusion equation or the heat equation.

The solution is a mean zero Gaussian with variance proportional to t! This is the density of Brownian motion.

$$P(X,t) = \frac{1}{\sqrt{\pi C t}} e^{X^2/(Ct)} \qquad \text{for } P(X,0) = \delta(X) \tag{60}$$

$$C/2 = \text{ diffusion coefficient} \tag{61}$$

Exercise: Show that (60) is the solution to (59).
So we get Brownian motion under a diffusive scaling for the random walk.

This also illustrates the self-similarity of this diffusion. In particular, if $w(t)$ is a standard Brownian motion, then so is

$$\frac{1}{\sqrt{\alpha}} w(\alpha t), \qquad \langle (\frac{1}{\sqrt{\alpha}} w(\alpha t))^2 \rangle = t$$

Asymptotic approximations
In the next sections we cover different types of asymptotic behavior in SDE's.

Examples:
1. Approximating systems with a large number of discrete events with a continuous process: Related to central limit theorem: normal approximation for a large number N of realizations

2. Small noise asymptotics and behavior in the state space for: boundary layers in state space

3. Systems with multiple time scales: quasi-steady approximations and stochastic averaging

Example: Quasi-steady approximations in deterministic dynamics

Michaelis-Menten Model: (deterministic model)

$$S + E \underset{k_3}{\overset{k_1}{\rightleftarrows}} C \xrightarrow{k_2} P + E$$

substrate + enzyme $\Longrightarrow$ complex $\Longrightarrow$ further reaction to product + enzyme

$$\begin{aligned} S \;+\; E &\xrightarrow{k_1} C \\ \dot{C} &= k_1 SE - k_3 C - k_2 C \\ \dot{S} &= -k_1 SE + k_3 C \\ \dot{E} &= -k_1 SE + k_2 C + k_3 C \\ \dot{P} &= k_2 C \end{aligned} \tag{62}$$

$E + C =$ conserved quantity $= E_0$ Take $(C(0) = 0)$ so we can eliminate E

Here the k_j are the reaction rates.

$$\left.\begin{aligned}\dot{C} &= k_1 S(E_0 - C) - k_3 C - k_2 C \\ \dot{S} &= -k_1 S(E_0 - C) + k_3 C \\ \dot{P} &= k_2 C\end{aligned}\right] \Rightarrow \begin{array}{l}\text{linear stability shows that} \\ C = 0, \quad S = 0 \\ \text{is stable fixed point.}\end{array} \tag{63}$$

Note that this is a nonlinear system of ODE's - in general it is difficult to get a closed-form expression for systems of this type.

Instead, suppose $E_0 \ll 1$ (small amount of enzyme); then also small amount of C.
The notation $\ll$ is interpreted as "is much less than", and is usually related to orders of magnitude - that is the initial amount of enzyme E_0 is small compared with other quantities that are larger (sometimes stated as $O(1)$ - order one, that is, a constant that is not large or small).

$$C = E_0 c \tag{64}$$

$$\Rightarrow E_0 \dot{c} = k_1 S(1-c)E_0 - k_3 E_0 c - k_2 E_0 c \tag{65}$$

$$\dot{S} = -k_1 S(1-c)E_0 + k_3 E_0 c \tag{66}$$

To see the leading order behavior, $T = E_0 t$ (note this is a short time)

$$\Rightarrow E_0 c_T = k_1 S_1 (1-c) - k_3 c - k_2 c \tag{67}$$

$$S_T = -k_1 S(1-c) + k_3 c \tag{68}$$

Leading order: set $E_0 c_T = 0$

$$\left.\Rightarrow c = \frac{k_1 S}{k_1 S + k_2 + k_3}\right] \begin{array}{l}\text{like a steady–state, if} \\ S \text{ was a constant}\end{array} \tag{69}$$

$$S_T = -k_1 S(1-c) + k_3 c \tag{70}$$

Leading order indicates comparing the relative sizes of different terms, under the assumption that $E_0 \ll 1$ while other quantities are $O(1)$.

The system (70) is the quasi-steady approximation: S is treated like a constant in the equation for c. The solution is compared to the solution of the full system (62) in the figure. Note that even though E_0 is not very small, the approximation still does well after an initial transient.

What is the physical interpretation: c changes quickly to adjust to the value of S, i.e. S is used up quickly in the reaction with E. Meanwhile S changes slowly, and looks like a constant relative to c.

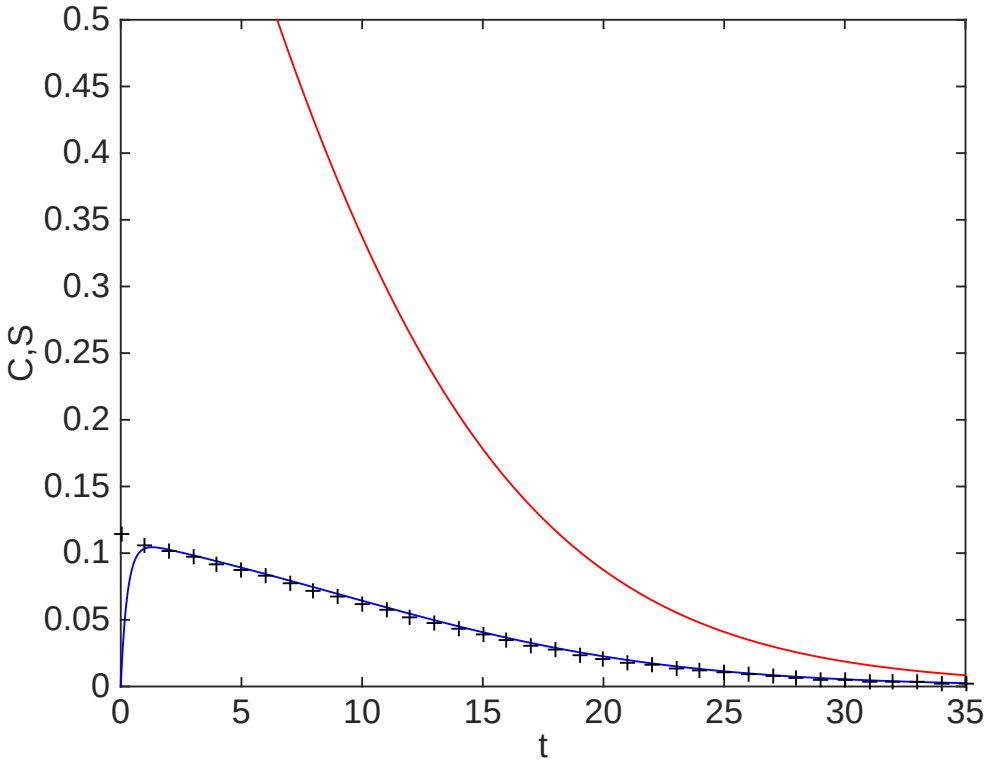


Figure 4: Realization M-M process, with $E_0 \ll 1$. The red line is S, the blue line is C, both obtained from (62). The $+$'s are obtained for c from the quasi-steady approximation (70. Here $E_0 = .2$.

4 Approximations of interacting individuals

The SIR model: Simple epidemiological model
Reference: Kuske, Greenwood, Gordillo, 2007, Journal of Theoretical Biology; Chaffee and Kuske, Bull. Math Bio, 2011:

Models: spread of disease Susceptible population = S: don't have the disease
Infective population = I: has the disease
Population of recovered individuals R: had the disease, can't get it again

For either probabilistic or deterministic models, we think in terms of rates

$$\begin{array}{ll} \text{transition} & \text{rate} \\ S \to S+1 & \mu N \\ S \to S-1 & \beta SI/N + \mu S \\ I \to I+1 & \beta SI/N \\ I \to I-1 & (\gamma+\mu)I \\ R \to R+1 & \gamma I \\ R \to R-1 & \mu R \end{array} \tag{71}$$

N = total population size: we will look at the case for large N, $N \gg 1$
μ = birth/death rate μ
γ = recovery rate
$\beta I/N$ = average number of contacts with infectives per susceptible per unit time.
N = total population size

Stochastic model: the rates (per unit time) are the conditional transition rates of the stochastic (Poisson) process (S, I, R)

Can write this as a Continuous time Markov process: $\{(S_t, I_t, R_t) : t \in [0, \infty)\}$, state space Z_+^3. N = expected population size $N = E[S + I + R]$

We need to break this down into Poisson events - births, deaths, infections, recoveries

For example: $P(S_{t+\Delta t} = s + 1 | S_t = s) = \mu N \Delta t + o(\Delta t)$
A birth process is a Poisson event with probability $\mu N \Delta t$ for a birth in time interval of length Δt

Other events:
$P(S_{t+\Delta t} = s + 1 | S_t = s) = \mu N \Delta t + o(\Delta t)$,
$P(S_{t+\Delta t} = s - 1 | S_t = s) = \mu S \Delta t + o(\Delta t)$ decrease in susceptibles due to death
$P(S_{t+\Delta t} = s - 1 | S_t = s) = \beta SI/N \Delta t + o(\Delta t)$ decrease in susceptibles due to infections
$P(I_{t+\Delta t} = i + 1 | I_t = i) = \beta SI/N \Delta t + o(\Delta t)$, increase in infectives (only by infections)

Approximations

Deterministic (mean field) model

$$\begin{aligned} \frac{dS}{dt} &= \mu(N - S) - \beta \frac{SI}{N}, \\ \frac{dI}{dt} &= \beta \frac{SI}{N} - (\gamma + \mu) I, \end{aligned} \tag{72}$$

Note: R equation is redundant, $S + I + R = N$.

Basic reproductive number, $R_0 = \frac{\beta}{\mu + \gamma}$
If $R_0 > 1$ there is a stable endemic ($I \neq 0$) equilibrium. (ODE exercise)

We express the stochastic equations of the process in a form easily compared with the equations (72) of the deterministic model.

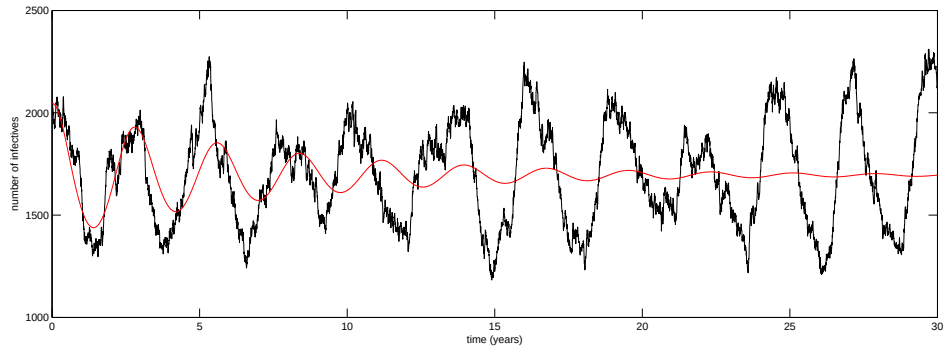


Figure 5: A realization of a stochastic SIR model (black) and its deterministic counterpart (red) for $N = 2000000$, $\mu = 1/55$, $R_0 = 15$ and $\gamma = 20$. Both show the number infected. Note that the stochastic result does not settle near an equlibrium value, but rather has a distinct nearly regular oscillation.

Approximation of the stochastic process with a continuous time, continuous state space: Typically for larger populations, with short time increments. $N \gg 1$

Recall diffusion processes - approximating a random walk with a Brownian motion for $(\Delta x)^2 = \Delta t$.

Ultimately, we will approximate Poisson process increments with Brownian motion increments.

It is useful to recall the centered Poisson increments that were considered in the model for shot noise. Then the centered increments will have mean zero and variance proportional to Δt.
Viewing our Poisson processes as increments: To each increment we add and subtract its conditional expectation, conditioned on the value of the process at the beginning of the time increment of length Δt. Each increment of the process is then the sum of the expected value of the increment of the process and the centered increment,

For a Poisson process U, composed of two independent types of Poisson events with probability P_+, P_- for increasing or decreasing by one, we have

$$P(U_{t+\Delta t} = u+1 | U_t = u) = P_+ \Delta t + o(\Delta t)$$
$$P(U_{t+\Delta t} = u-1 | U_t = u) = P_- \Delta t + o(\Delta t)$$

Then the equation for an increment of U is written as the average change plus the increments:

$$\Delta U = (P_+ - P_-)\Delta t + Z_+ + Z_-,$$
$$Z_{+/-} = U^{+/-} - E[U^{+/-}]$$

Z_+ Z_- are centered Poisson increments, that is $E[Z_{+/-}] = 0$ with variances $P_{+/-}\Delta t$

Note: Can simplify, for two processes independent: $\Delta U = (P_+ - P_-)\Delta t + Z$
Z is a centered Poisson increment, that is $E[Z] = 0$ with variance that is the sum of the two variances $(P_+ + P_-)\Delta t$

Recall: σdW has distribution $N(0, \sigma\sqrt{\Delta t})$, so Z has zero mean and standard deviation proportional to $\sqrt{\Delta t}$, as does the standard Brownian motion.

When is a Poisson random variable well-approximated by a Gaussian random variable? Typically when there are enough events so that the Poisson parameter is large (e.g. > 10)

So we can write:

$$\begin{aligned} \Delta S &= (\mu(N-S) - \beta\tfrac{SI}{N})\,\Delta t + \Delta Z_1 - \Delta Z_2, \\ \Delta I &= (\beta\tfrac{SI}{N} - (\gamma+\mu)I)\,\Delta t + \Delta Z_2 - \Delta Z_3. \end{aligned} \tag{73}$$

The increments ΔZ_1, ΔZ_2, ΔZ_3 are independent, centered Poisson variables with variances $\mu(N+S)\Delta t$, $\beta\frac{SI}{N}\Delta t$ and $(\gamma+\mu)I\Delta t$, respectively.

Note here that S and I have fluctuations that are not independent from each other. ΔZ_2 is the increment of the process that influences both S and I (infection), so it must appear in both equations. Thus the noise in the equations is correlated.

We replace ΔZ_i by increments of Brownian motion, dW_i, with the same standard deviations,

$$\begin{aligned} dS &= \left(\mu(N-S) - \frac{\beta}{N}SI\right)\, dt + G_1 dW_1 - G_2 dW_2, \\ dI &= \left(\frac{\beta}{N}SI - (\gamma+\mu)I\right)\, dt + G_2 dW_2 - G_3 dW_3, \\ G_1 &= \sqrt{\mu(N+S)}, \quad G_2 = \sqrt{\frac{\beta}{N}SI}, \quad G_3 = \sqrt{(\gamma+\mu)I}. \end{aligned} \tag{74}$$

This is the diffusion approximation for the stochastic SIR model
Note - these have the Ito interpretation.
Note - some noise coefficients may be large or small? If the noise is small, why not just use the deterministic equations (72)?

References:

Allen, Allen, Arciniega, Greenwood, Stochastic Analysis and Applications, 2008.

$$\begin{aligned} dX &= f(t,\mathbf{X}) + G(t,\mathbf{X})d\mathbf{W} \\ dY &= f(t,\mathbf{Y}) + B(t,\mathbf{Y})d\mathbf{V} \end{aligned}$$

X, Y are d-dimensional, $\mathbf{W}$ is m-dimensional, $\mathbf{V}$ is d-dimensional, $m > d$. For $GG^T = H$ and $B = H^{1/2}$, for f, G, B satisfying certain continuity conditions, giving pathwise unique solutions. Then X and Y have the same distribution. (weak sense)

So we can express the noise terms in different forms, can choose for our convenience. In the example above, it is easy to derive, easy to see influence of different biological processes in the equations.

Approximating Poisson with Normal

Expect N is relatively large (larger rates for Poisson processes in a time interval)

Are stochastic terms significant? i.e. if N is large, are the fluctuations significant?

Consider rescaled system $S = Nu$, $I = Nv$ - here u is the proportion of the total population that are susceptible, v is the proportion that are infected. Both u, v are between 0 and 1.

$$\begin{aligned} du &= \left(\mu(1-u) - \frac{\beta}{u}v\right)\, dt + g_1 dW_1 - g_2 dW_2, \\ dv &= \left(\frac{\beta}{N}SI - (\gamma+\mu)I\right)\, dt + g_2 dW_2 - g_3 dW_3, \\ g_1 = \sqrt{\mu(1+u)/N}, &\quad g_2 = \sqrt{\frac{\beta}{N}uv}, \quad g_3 = \sqrt{(\gamma+\mu)v/N}. \end{aligned} \tag{75}$$

Note, coefficients g_j of noise term scale as $N^{-1/2}$, vanish as $N \to \infty$. Then we recover the mean field (deterministic) model (72).

What if N is just large, rather than infinite?

We consider the system with the basic reproductive number, $R_0 = \frac{\beta}{\mu+\gamma} > 1$. Then the deterministic system has a unique nontrivial stable equilibrium point (S_{eq}, I_{eq}) at

$$S_{eq} = \frac{N}{R_0}, \quad I_{eq} = \frac{N\mu}{\beta}(R_0 - 1). \tag{76}$$

Consider simulations for different parameter values: Why is the noise sometimes significant?

Let's use different rescaled variables, to reflect this non-trivial stable equilibrium. We introduce the dimensionless variables

$$u = \frac{S - S_{eq}}{S_{eq}}, \quad v = \frac{I - I_{eq}}{I_{eq}},$$

Note S_{eq} and I_{eq} are both proportional to N. u, v are fluctuations around the equilibrum values, normalized by the size of that equilibrium.

$t \to \Omega t$, $\Omega = \sqrt{\frac{\beta\mu}{R_0}(R_0 - 1)}$ (for convenience) to get the equations (exercise)

$$\begin{aligned}
du &= \frac{1}{\Omega}[(-\mu - \frac{\beta I_{eq}}{N})u - \frac{\beta I_{eq}}{N}v - \frac{\beta I_{eq}}{N}uv]dt + \\
&\quad + \sqrt{\frac{\mu}{\Omega S_{eq}^2}(N + S_{eq}(u+1))}dW_1(t) - \sqrt{\frac{\beta I_{eq}}{\Omega N S_{eq}}(v+1)(u+1)}dW_2(t), \\
dv &= \frac{1}{\Omega}[\frac{\beta S_{eq}}{N}(u+v) + \frac{\beta S_{eq}}{N}uv - (\gamma+\mu)v]dt + \\
&\quad + \sqrt{\frac{\beta S_{eq}}{\Omega N I_{eq}}(v+1)(u+1)}dW_2(t) - \sqrt{\frac{\gamma+\mu}{\Omega I_{eq}}(v+1)}dW_3(t).
\end{aligned} \tag{77}$$

Noise coefficients: still $N^{-1/2}$

The Power Spectral density (PSD) is the modulus of the Fourier transform of the process, here graphed as a function of frequency. Large peak indicates most of the energy is in a particular frequency. This indicates a dominant oscillation with that frequency. Fluctuations indicated by the width of the peak.

Consider equations linearized about $u = v = 0$ (near equilibrium). This is a linear approximation about the equilibrium, and this linear approximation describes a multi-dimensional Ornstein-Uhlenbeck process.

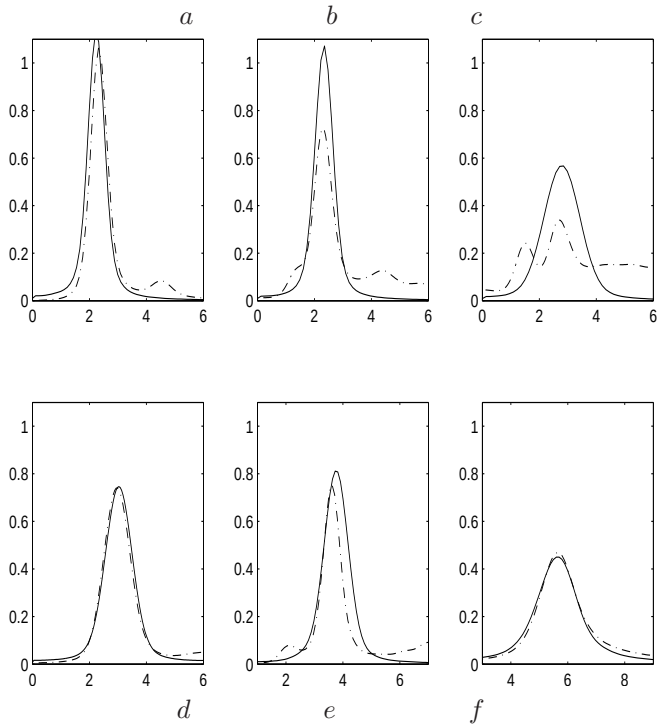


Figure 6: Graphs of the PSD (vertical axis) vs. frequency of the oscillations, a)$R_0 = 15, \gamma = 20, \mu = 1/55, N = 500000$. b)$R_0 = 15, \gamma = 25, \mu = 1/55, N = 500000$. c)$R_0 = 15, \gamma = 30, \mu = 1/55, N = 500000$. d)$R_0 = 15, \gamma = 30, \mu = 1/55, N = 2000000$. e)$R_0 = 7, \gamma = 33, \mu = 1/55, N = 2000000$. f)$R_0 = 10, \gamma = 35, \mu = 1/10, N = 500000$.

$$d\begin{pmatrix} u \\ v \end{pmatrix} = \mathbf{M}\begin{pmatrix} u \\ v \end{pmatrix} dt + \mathbf{G}\begin{pmatrix} dW_1 \\ dW_2 \\ dW_3 \end{pmatrix}, \qquad \mathbf{M} = \begin{pmatrix} -\frac{\mu R_0}{\Omega} & -\mu\frac{R_0-1}{\Omega} \\ \frac{\beta}{\Omega R_0} & 0 \end{pmatrix}, \tag{78}$$

$$\mathbf{G} = \begin{pmatrix} \sqrt{\frac{\mu}{\Omega S_{eq}^2}(N+S_{eq})} & -\sqrt{\frac{\beta I_{eq}}{\Omega N S_{eq}}} & 0 \\ 0 & \sqrt{\frac{\beta S_{eq}}{\Omega N I_{eq}}} & -\sqrt{\frac{\gamma+\mu}{\Omega I_{eq}}} \end{pmatrix} = \begin{pmatrix} g_1 & -b^2 g_2 & 0 \\ 0 & g_2 & -g_2 \end{pmatrix}.$$

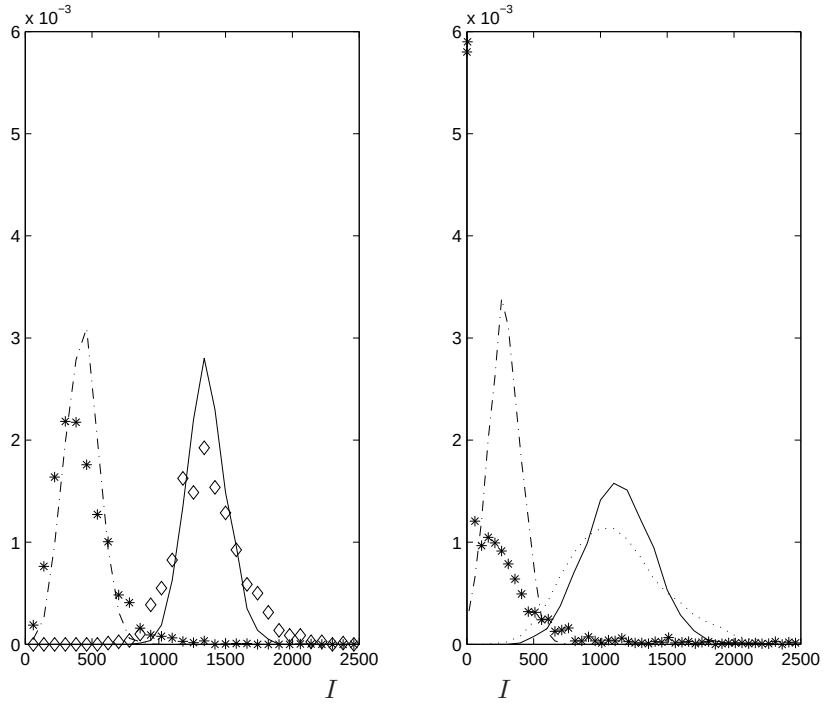


Figure 7: Stationary density $p(I)$ (2000 realizations for a value of $t > 100$) Left: (a) ($*$'s and dash-dotted line) and (f) (diamonds and solid line) Right, (c) ($*$'s and dash-dotted line), (d) (dotted and solid line)

Solutions of the deterministic version of (78) are given in terms of the eigenvalues of $\mathbf{M}$, which are

$$\lambda = -\epsilon^2 \pm \sqrt{\epsilon^4 - 1}\,,$$

where

$$\epsilon^2 = \frac{\mu R_0}{2\Omega}.$$

What does the solution look like for $\epsilon \ll 1$? Under what conditions is $\epsilon \ll 1$?

$$\begin{pmatrix} u \\ v \end{pmatrix} \sim C_1 e^{-\epsilon^2 t} \begin{pmatrix} b\cos t \\ \sin t \end{pmatrix} + C_2 e^{-\epsilon^2 t} \begin{pmatrix} b\sin t \\ -\cos t \end{pmatrix}, \tag{79}$$

dropping $O(\epsilon^2)$ corrections.

For N large and finite - drift and diffusion coefficients both small
Oscillations have weak (slow) decay, noise coefficients may play a role

This is shown in the references, where u and v are sinusoidal with stochastic amplitudes that are O-U processes.

5 Equation(s) for the probability density:

General continuous time/space

We derived an equation for $P(x,t)$ using conditional probability - that is conditioning on starting at x_0 at time t, then taking a step up or down.

In general, this is a useful approach to deriving and equation for $P(x,t)$
Chapman-Kolmogorov equation:

$$P(X(t) = y|X(s) = x) = p(y,t,x,s) = \int p(y,t,z,\tau)p(z,\tau,x,s)dz \quad s < \tau < t \tag{80}$$

This is the probability of going from x to y via z, "summed" over all possible intermediate z. While in general this doesn't look like a simplification of the problem, it may be, given the particular problem and the choice of z.

This general equation is used to derive the PDE for the probability density $p(y,t|x,s)$ (often, $s = 0$) as follows:

We will show that $\frac{\partial p}{\partial t} = L^* p$ where

$$L^* p = \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(b^2(y,t)p(y,t,x,s)\right) - \frac{\partial}{\partial y}\left(a(y,t)p(y,t,x,s)\right) \quad \text{(FPE)} \tag{81}$$

where $d\xi = adt + bdw$, $p(\xi = y,t|\xi = x,s) = p(y,t,x,s)$. This is the Fokker-Planck equation (FPE) or more generally, the Forward Kolmogorov equation.

This is accomplished by showing:

$$\int g(y)\frac{\partial p}{\partial t}\,dy = \int g(y)L^* p(y,t,x,s)\,dy \tag{82}$$

for an appropriate "nice" test function g (weak sense). To do this, we show that

$$\int \left\{ \int p(z, t+h, y, t)[g(z) - g(y)]dz \right\} p(y, t, x, s)\, dy \tag{83}$$

$$= \int [p(y, t+h, x, s) - p(y, t, x, s)]g(y)\, dy \quad \text{(FPI)} \tag{84}$$

The RHS of FPI can be approximated with

$$h \int p_t(y, t, x, s)g(y)dy, \qquad h = \Delta t \tag{85}$$

as $h \to 0$.

The LHS of FPI can be written as:

$$\int \left\{ \int p(z, t+h, y, t)[g'(y)(z-y) + 1/2g''(y)(z-y)^2 ...]\, dz \right\} p(y, t, x, s)\, dy \tag{86}$$

by using a Taylor series expansion about $z = y$.

Borrowing the $\frac{1}{h}$ from (85), the LHS can be written as

$$\int \left(\frac{1}{h} E\left[\xi(t+h) - \xi(t)\right] g'(y) + \frac{1}{2} g''(y) \frac{1}{h} E\left[\left(\xi(t+h) - \xi(t)\right)^2 \right] \right) p(y, t, x, s)\, dy \tag{87}$$

Defining

$$\frac{1}{h} E\left[\xi(t+h) - \xi(t)\right] = a(y, t) \quad \text{as} \quad h \to 0 \tag{88}$$

$$\frac{1}{h} E\left[\left(\xi(t+h) - \xi(t)\right)^2 \right] = b^2(y, t) \quad \text{as} \quad h \to 0 \tag{89}$$

which is essentially the statement that the particle displacement is $a(y, t)h$ with variance $b^2 h$ (thus a diffusion process). Together the RHS and LHS of FPI yields:

$$\int p_t(y,t,x,s)g(y)dy = \int \left[a(y,t)g'(y) + \frac{g''(y)}{2}b^2(y,t)\right] p(y,t,x,s)dy \tag{90}$$

Integrating by parts - which moves the derivatives to the terms ap and b^2p - yields the FPE for $p(y,t,x,s)$ (weak version - integrated with test function $g(y)dy$). Then we get $p_t = L^*p$.

To show FPI: Write RHS:

$$\int p(z,t+h,x,s)g(z)dz - \int p(y,t,x,s)g(y)dy \tag{91}$$

First term

$$\begin{aligned}\int p(z,t+h,x,s)g(z)dz &= \int \left(\int p(z,t+h,y,t)p(y,t,x,s)\,dy\right) g(z)\,dz \\ &= \int p(y,t,x,s)\left(\int p(z,t+h,y,t)g(z)\,dz\right) dy\end{aligned}$$

using Chapman-Kolmogorov equation. Second term:

$$\begin{aligned}\int p(y,t,x,s)\cdot 1\cdot g(y)dy &= \int p(y,t,x,s)\left(\int p(z,t+h,y,t)dz\right) g(y)dy \\ &= \int p(y,t,x,s)\left(\int p(z,t+h,y,t)g(y)dz\right) dy\end{aligned}$$

Subtract first and second to get the LHS of FPI.

So we can find the density of $dy = a\,dt + b\,dw$ by solving a PDE $\frac{\partial p}{\partial t} = L^*p$. Note here that $b\,dw$ is in the sense of Ito.

Computationally, we can find $p(x,t)$ using methods for solving a PDE.

We can also obtain a numerical approximation to $p(x,t)$ via a simulation of the SDE for x.

This is accomplished by simulating the SDE for x to time t using N (independent) realizations of the SDE. Recording the N values of $x(t)$, we construct a histogram from these values, recording the number of realizations that fall into bins of width Δx. Then

$$p(X,t)\Delta x \approx k_i/N \tag{92}$$

where $0 \leq k_i \leq N$ is the number of realizations of $x(t)$ that take values in the interval between X and $X + \Delta x$. We see examples of these approximations in later applications.

5.1 Exit Time

Another key quantity is Expected Exit Time (Expected Transition Time).

Later, we will see that the equation for mean exit time has the form, $u_t + Lu = -1$ and $u(\xi \in \partial\Omega, t) = 0$, where

$$Lu = au_\xi + \frac{1}{2}b^2 u_{\xi\xi} \quad \text{where } d\xi = adt + bdw$$

L is the adjoint of L^*, where $p_t = L^* p$.

6 Mean exit time

Another key quantity is Expected Exit Time (Expected Transition Time).

If we define τ_x=time to reach certain state, $\partial\Omega$, given initial state $\xi(s) = x \in \Omega$,

$E[\tau_x]$=Expected time to exit Ω and reach new state ($\partial\Omega$) given $\xi(s) = x$.

We can write this as

$$E[\tau_x] = \int_0^\infty \int_\Omega p(y,t,x,s)dydt = \int_0^\infty P(\tau_x > t)dt \tag{93}$$

since $P(\tau_x > t)$ is the probability that $y \in \Omega$ at time t. Can also compute this, noting that for $u(\xi,t)$ with $d\xi = adt + bdw$,

$$du(\xi(t),t) = \left(\frac{\partial u}{\partial t} + \frac{\partial u}{\partial \xi}a + \frac{1}{2}b^2\frac{\partial^2 u}{\partial \xi^2}\right)dt + \frac{\partial u}{\partial \xi}bdw \tag{94}$$

using Ito's formula, so that

$$u(\xi(t),t) = u(\xi(s),s) + \int_0^t [u_t + Lu]\, d\hat{t} + \int_0^t \frac{\partial u}{\partial \xi}bdw \tag{95}$$

(Below we use $s = 0$)
Note $Lu = au_\xi + \frac{1}{2}b^2 u_{\xi\xi}$, is the adjoint operator for $L^*u = -(au)_\xi + \frac{1}{2}(bu)_{\xi\xi}$

Suppose, $u_t + Lu = -1$ and $u(\xi \in \partial\Omega, t) = 0$. Then substitute $t = \tau_x$ above and take expected values,

$$E[u(\xi(\tau_x),\tau_x)] = u(x,s) + E\left[\int_0^{\tau_x}(-1)dt\right] + 0 \tag{96}$$

$$\Rightarrow\ 0 = u(x,s) - E[\tau_x] \qquad (+s \text{ if } s \neq 0) \tag{97}$$

Example: Exit time of a particle from potential $U(x)$.

Simple example: $U(x) = \frac{x^2}{2}$, on $-1 < x < 1$

Particle position: $d\xi = -\xi dt + \sqrt{2\epsilon}dw = -U'(\xi)dt + \sqrt{2\epsilon}dw$

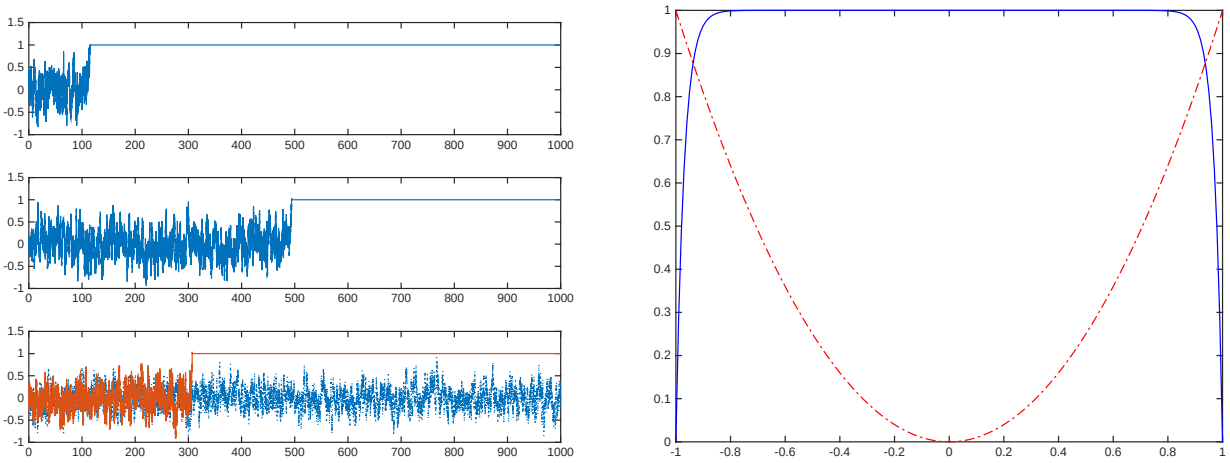


Figure 8: Left: Realizations of ξ vs. t for $\epsilon = .05, .04, .03$ (top to bottom). Right: The uniform approximation for q (blue line), compared to the potential $U(x)$ (red dash-dotted line).

Note: this is a time homogeneous case $\Rightarrow$
$\epsilon u'' - xu' = -1$ for u= mean exit time from interval $[-1, 1]$
$u(-1) = u(1) = 0$

This is an example which we could solve exactly, but we see how the asymptotic solution gives physical insight: Consider for $\epsilon \ll 1$. Then we use boundary layer theory to get an approximation to the solution.

For small ϵ (small noise), expect $u \to \infty$ as $\epsilon \to 0$ (at least for $x \neq \pm 1$)
so take $u = C(\epsilon)q(x)$ for $C(\epsilon) \to \infty$ as $\epsilon \to 0$

Then, substituting into equation for u,

$$\epsilon q'' - xq' \sim 0 \qquad q(-1) = q(1) = 0 \tag{98}$$

For $\epsilon = 0$, we have, to leading order: $\Rightarrow q_0 =$ constant constant $\neq 0$

However, the boundary conditions are not satisfied, so we must find the boundary layer behaviour. The solution q_0 is the "outer" solution (away from the boundaries).

Next we have to look near the boundaries, and zoom in. Then, taking $\epsilon z = 1 - x$, as a local variable near $x = 1$, yields

$$\frac{\epsilon}{\epsilon^2} q_{zz} - \frac{1 - \epsilon z}{\epsilon} q_z = 0 \quad , \quad q \sim q_0 + \epsilon q_1 + ... \tag{99}$$

$$\Rightarrow \quad q_{0zz} - q_{0z} = 0 \quad , \quad q_0(z = 0) = 0 \tag{100}$$

$$q_0 = \left(1 - e^{-z}\right) = \left(1 - e^{-\frac{(1-x)}{\epsilon}}\right) \tag{101}$$

Thus we see the different behavior of q in this local region - q is near 1 when x is not near 1. q approaches zero as x approaches 1. Similarly for x near -1.

Then we can construct the uniform solution:

$$\begin{aligned} q &\sim 1 && \text{away from } x = \pm 1 \\ q &\sim 1 - e^{-\frac{(1-x)}{\epsilon}} && \text{near } x = 1 \\ q &\sim 1 - e^{-\frac{(x+1)}{\epsilon}} && \text{year } x = -1 \end{aligned}$$

Boundary layer approximations $\rightarrow$ constant, as matched to "outer" approximation
Uniform approximation: Add boundary layer expansions + outer approximation - parts matched

$$u = C(\epsilon)\left(1 - e^{-\frac{(1-x)}{\epsilon}} - e^{-\frac{(x+1)}{\epsilon}}\right) \tag{102}$$

Still, we don't know $C(\epsilon)$, but can it determine using our knowledge of $p(x)$ (invariant density). That is, p satisfies equation with the adjoint operator: $L^* p = 0$.

We have the equation for u

$$Lu = -1 \tag{103}$$

and

$$L^* p = 0 = \epsilon p_{xx} + (xp)_x \Rightarrow p(x) = \frac{1}{\sqrt{2\pi\epsilon}} e^{-\frac{x^2}{2\epsilon}} \tag{104}$$

So,

$$\int pLu dx = p\epsilon u'|_{-1}^{1} + \int \cancelto{0}{(L^* p)} u dx = -\int_{-1}^{1} 1 \cdot p dx \tag{105}$$

If we use the uniform expansion for u we can evaluate (exercise), then we get the equation for $C(\epsilon)$ from the calculation of $p(x)$

$$C(\epsilon) \sim K\sqrt{2\pi\epsilon} e^{-\frac{U(1)}{\epsilon}} + \text{ higher order corrections} \tag{106}$$

Additional exercise: Could there be a longer time to escape from the bottom of the potential? Show that there is no internal layer in the construction of the uniform solution. Use local variable $v = \epsilon^\alpha x$, find appropriate α and equation for $q(v)$. A contradiction in matching to the outer solution results, which shows no layer at $x = 0$.

Asymptotic methods for determining $p(x,t)$
Another asymptotic method that is commonly used to determine $p(x,t)$ in the small noise case is the WKB method - another asymptotic method fro solving PDE's.

Let's see how this works for the simple example above. The equation for $p(x,t)$ is

$$p_t = (xp)_x + \epsilon p_{xx} = L^* p \tag{107}$$

The WKB method is based on the observation that is we take $\epsilon \to 0$ in the equation for p, we eliminate the highest order derivative. This doesn't make sense, as it eliminates the diffusion from the equation.

So, instead we should look for a form of the solution that better balances the effect of the diffusion with the other contributions.

$$p(x,t) = K(x,t)\exp -\psi/\epsilon \tag{108}$$

This makes intuitive sense, particularly if we are expecting densities that decay exponentially for large values.

Substituting into the equation for $p(x,t)$, we get:

$$K_t - \frac{1}{\epsilon}K\psi_t = x\left[K_x - \frac{\psi_x}{\epsilon}K\right] + K + \epsilon\left[-\frac{\psi_{xx}}{\epsilon}K + \frac{(\psi_x)^2}{\epsilon^2} - 2K_x\frac{\psi_x}{\epsilon} + K_{xx}\right] \tag{109}$$

Here we have terms with coefficients corresponding to different orders of ϵ: $O(\epsilon^{-1})$ and $O(\epsilon^0)$. For simplicity, we consider the case of the invariant density - no t dependence.

Then, to leading order (terms with coefficient of ϵ^{-1}) we have

$$-x\psi_x + \psi_x^2 = 0 \tag{110}$$

from which we obtain $\psi =$ constant or $\psi = x^2/2$. This is a nonlinear first order PDE for ψ - the eikonal equation. The constant solution is the trivial solution, and is also not normalizable for p a probability density.

Then, the equation for K (terms with coefficient ϵ^0) is

$$(1-\psi_{xx})K + (x-2\psi_x)K_x + K_{xx} = -xK_x + K_{xx} = 0 \tag{111}$$

The solution for K that is normalizable, i.e. $\int K(x)e^{-\psi(x)/\epsilon}dx$ is $K=$ constant. Then we obtain the result given above in (104).