Methods of stochastic simulation

Computational Systems Biology II

lectures by Antti Saarinen

22.4.2008

Outline

- What is stochasticity?
- Stochastic phenomena in biology
- Stochastic simulation methods
  - Gillespie method for simulating coupled chemical reactions
  - Stochastic differential equations (SDEs) and Brownian motion

Stochasticity

- from Greek *stokhastikos* = capable of guessing
- "the quality of lacking any predictable order or plan" (randomness, noise)
- "having a probability distribution, usually with finite variance" (statistical)
- "involving a random variable the successive values of which are not independent" (statistical)

Stochastic phenomena in biology

- Pollen in water ($\longrightarrow$ Brownian motion)
- Chemical reactions
- Diffusion
- Bacterial motion
- Behaviour of ion channels on cell membrane
- Electroresponsiveness of a neuron
- Radioactive decay
- etc..

ODEs

- The time evolution of spatially homogeneous mixture of chemically reacting molecules is usually calculated by solving a set of ordinary differential equations.
- N chemical species $\rightarrow$ N differential equations.
- Each equation expresses the rate-of-change of the molecular concentration of one chemical species as a function of the molecular concentrations of all the species.

$$\frac{d[X_i]}{dt} = f_i([X_1], \ldots, [X_N], t)$$

ODEs

- Traditional method based on a *deterministic formulation* of chemical kinetics.
- Reaction constants are viewed as "reaction rates".

$$\frac{d[X_3]}{dt} = k_1[X_1][X_2] - k_2[X_3]$$

- Concentrations are represented by continuous, single-valued functions of time.
- Although adequate in most cases, there are important situations, for which underlying physical assumptions are unrealistic and consequent predictions are unreliable.

Gillespie's stochastic method

- Reaction constants are viewed not as "reaction rates", but as "reaction probabilities per unit time".
- Temporal behaviour of a chemically reacting system takes the form of a Markovian random walk in the N-dimensional space of molecular populations.
- The time evolution is described by a single equation for a grand probability function in which time and the N populations appear as independent variables (the master equation).

Gillespie's stochastic method

- From a *physical* point of view, the stochastic formulation is superior to the deterministic formulation.
- The stochastic approach is always valid whenever the deterministic approach is valid, and is sometimes valid when the deterministic approach is not.
- Gillespie presents a feasible method for numerically calulating the stochastic time evolution of a chemical system.
- Set of deterministic reaction-rate equations for a given chemical system is much easier to solve than the stochastic master equation for the same system.

Gillespie's stochastic method

- The general problem: given a volume V which contains molecules of N chemically active species S_i, determine the time evolution of such a system.
- $X_i :=$ current number of molecules of chemical species S_i in V.
- Chemical species can participate in M reactions R_m, each characterised by a reaction parameter c_m.
- Parameter c_m is called the *reaction propensity*.

Gillespie's stochastic method

- (For definiteness) Reaction R_m is one of the following type
  - * $\longrightarrow$ reaction products,
  - $S_j \longrightarrow$ reaction products,
  - $S_j + S_k \longrightarrow$ reaction products,
  - $2S_j \longrightarrow$ reaction products,
  - $S_j + S_k + S_l \longrightarrow$ reaction products,
  - $S_j + 2S_k \longrightarrow$ reaction products,
  - $3S_j \longrightarrow$ reaction products.
- Each reaction is *unidirectional*, so any reversible reaction must be considered as two separate reactions.

Gillespie's stochastic method

- Fundamental hypothesis: $c_m \Delta t :=$ average probability that a particular combination of R_m reactant molecules will react accordingly in the next time interval Δt.
- The relationship between the propensity c_m and the "reaction rate constant" k_m which is used in the deterministic formulation will be examined later.
- Simulate the time evolution of N quantities $\{X_i\}$, knowing only their initial values, the forms of the M reactions $\{R_m\}$ and the values of the assosiated reaction parameters $\{c_m\}$.

Finding the propensity c_m

- Let's look at reaction $R_m : S_1 + S_2 \rightarrow 2S_3$.
- Molecules are hard spheres with masses m_i and diameters d_i.
- *1–2 collision* will occur if the centre-to-centre distance between an S_1 and S_2 molecule decreases to $d_{12} := (d_1 + d_2)/2$.
- Let v_{12} denote the relative speed of the molecules.
- In the vanishingly small time interval Δt molecule 1 sweeps out relative to molecule 2 a "collision volume" $\Delta V_{coll} = \pi d_{12}^2 v_{12} \Delta t$.
- If the center of molecule 2 lies in ΔV_{coll} then molecules 1 and 2 collide in time Δt.

Finding the propensity c_m

- Problems in deterministic case (spatially homogeneous) when $\Delta t \to 0$, because the number of molecules in ΔV_{coll} will either be 0 or 1.
- Averaging leads to more trouble (e.g. the average number of molecular pairs does not equal to the product of average numers of molecules).
- Herein lies the source of the inexact nature of the deterministic reaction rate equations.
- All these difficulties can be avoided if we consider uniformly (randomly) distributed molecules throughout V.
- Probability that the centre of one S_2 molecule will lie inside ΔV_{coll} is exactly $\Delta V_{coll}/V$.

Finding the propensity c_m

- *Average* probability that a particular 1–2 molecular pair will collide in the next Δt is

$$\left\langle \frac{\Delta V_{coll}}{V} \right\rangle = \frac{\pi d_{12}^2 \langle v_{12} \rangle \Delta t}{V}.$$

- Average relative velocity can be calculated using Maxwell-Bolzmann distributions.
- The above expression corresponds exactly to the quantity $c_m \Delta t$.

Relation between c_m and k_m

- Mathematical relationship between c_m and k_m is always rather simple (e.g. $k_m = V c_m$ or $k_m = V c_m/2$), but from a *physical* standpoint c_m appears to be on much firmer ground.
- The stochastic formulation of chemical kinetics for spatially homogenous systems does indeed take proper account of correlations and fluctuations which are ignored in the deterministic formulation.
- For most systems the difference between the stochastic and the deterministic formulation is academic. However, near chemical instabilities in certain nonlinear systems, fluctuations and correlation can produce dramatic effets.

Reaction probability density

- $P(X_1, \ldots, X_N, t) =$ the probability that there will be X_i molecules of S_i in V at time t.
- The so-called master equation is just the time evolution equation for the function $P(X_1, \ldots, X_N, t)$, and it can be rigorously derived by using simple probability calculus.
- The master equation is usually intractable, both analytically and numerically.
- Problem can be solved using *reaction probability density function* $P(\tau, m)$.

Reaction probability density

- $P(\tau, m)d\tau =$ probability at time t that the next reaction in V will occur in the differential time interval $(t+\tau, t+\tau+d\tau)$ and will be an R_m reaction.
- After some calculations we get

$$P(\tau, m) = h_m c_m \exp\left(-\sum_{i=1}^{M} h_i c_i \tau\right),$$

where h_i is the number of distinct molecular reactant combinations for reaction R_m found to be present in V at time t.

Simulation algorithm

1) Set $t = 0$. Specify initial values $X_1, \ldots, X_N$ and $c_1, \ldots, c_m$. Calculate $h_1 c_1, \ldots, h_M c_M$ which determine $P(\tau, m)$.
2) Generate random pair (τ, m) according to $P(\tau, m)$.
3) Using numbers τ and m andvance t by τ, and change X_i values of those species involved in reaction R_m to reflect the occurrence of one R_m reaction. Then, recalculate $h_i c_i$ for the new $P(\tau, m)$ and go to step 2.

Simulation algorithm

- By carrying out the above procedure one obtains *one possible* realisation of the stochastic process.
- In order to get statistically complete picture of the temporal evolution of the system, one must actually carry out *several independend* simulations with same initial conditions.
- Expected number of S_i molecules, variance or standard deviation for describing the fluctuations which may reasonably be expected.

Summary of Gillespie method

- Relatively simple procedure for calculating the time evolution of any spatially homogeneous chemical system.
- Commonly used in biological stochastic simulations.
- Implemented in many softwares.
- Improved versions also available ("direct" method, "first reaction" method, τ-leap . . .).

Example

The Brusselator, a "limit cycle" chemical oscillator

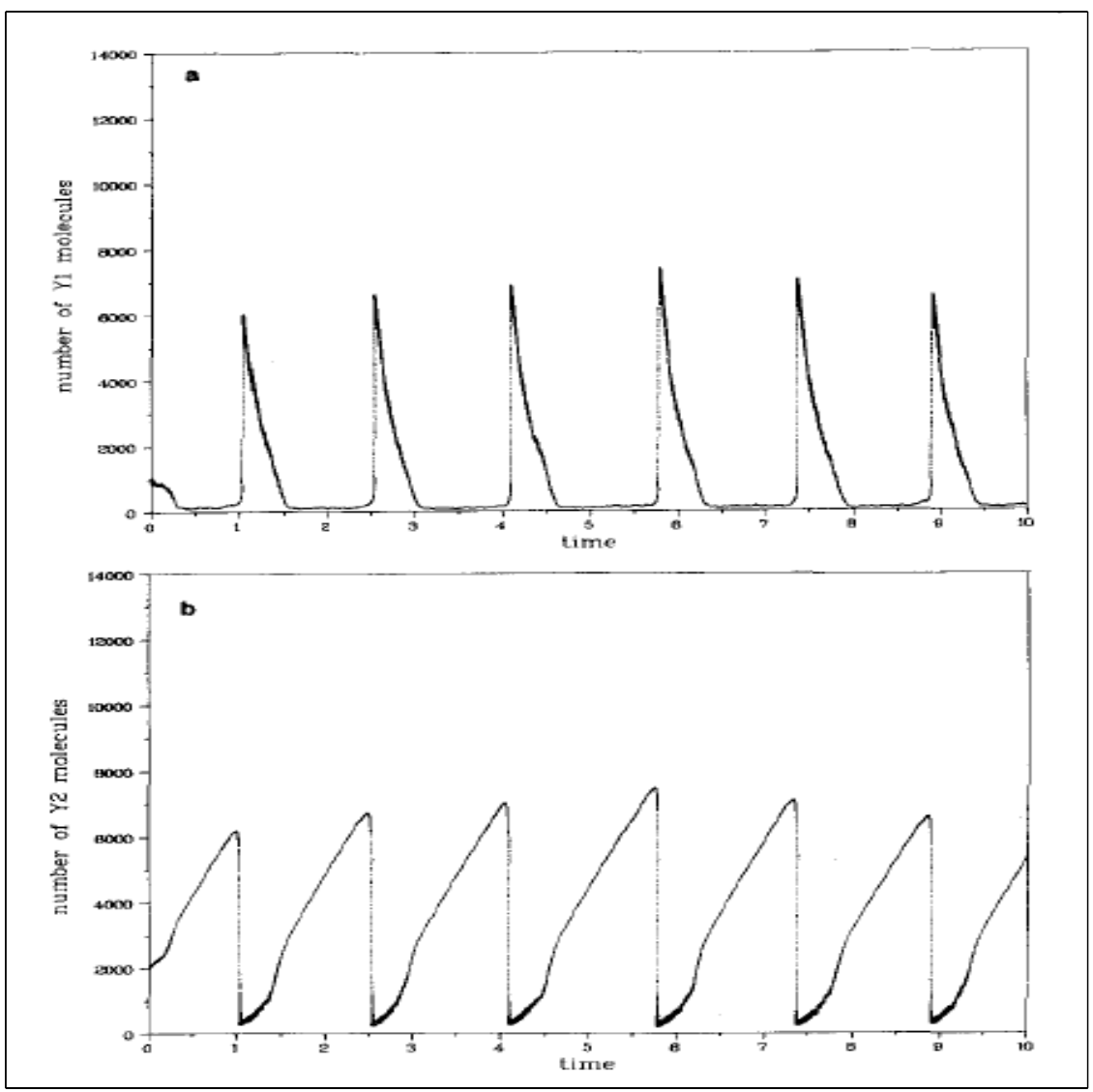
(1) $X_1 \xrightarrow{c_1} Y_1$

(2) $X_2 + Y_1 \xrightarrow{c_2} Y_2 + Z_1$

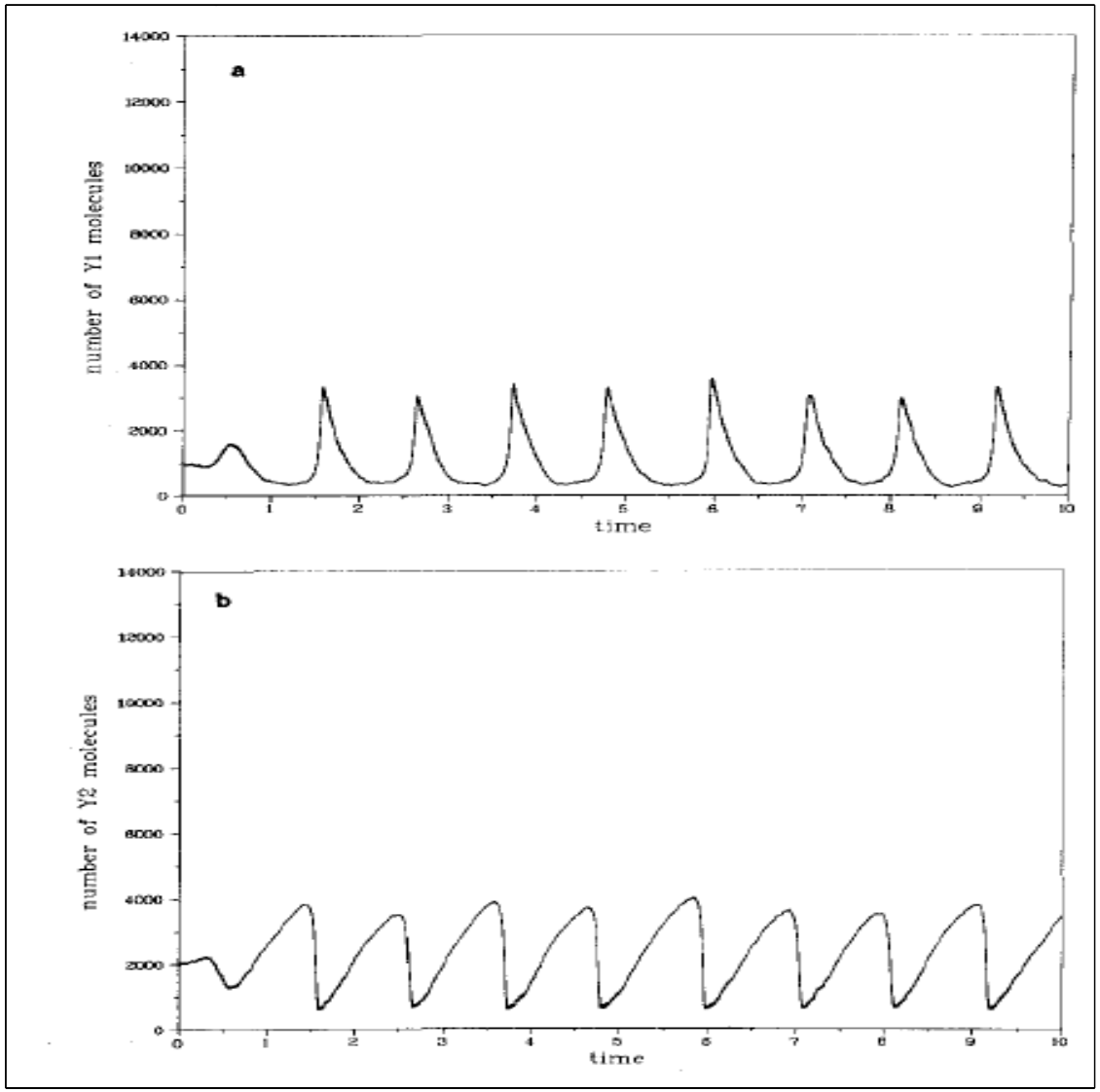
(3) $2Y_1 + Y_2 \xrightarrow{c_3} 3Y_1$

(4) $Y_1 \xrightarrow{c_4} Z_2$

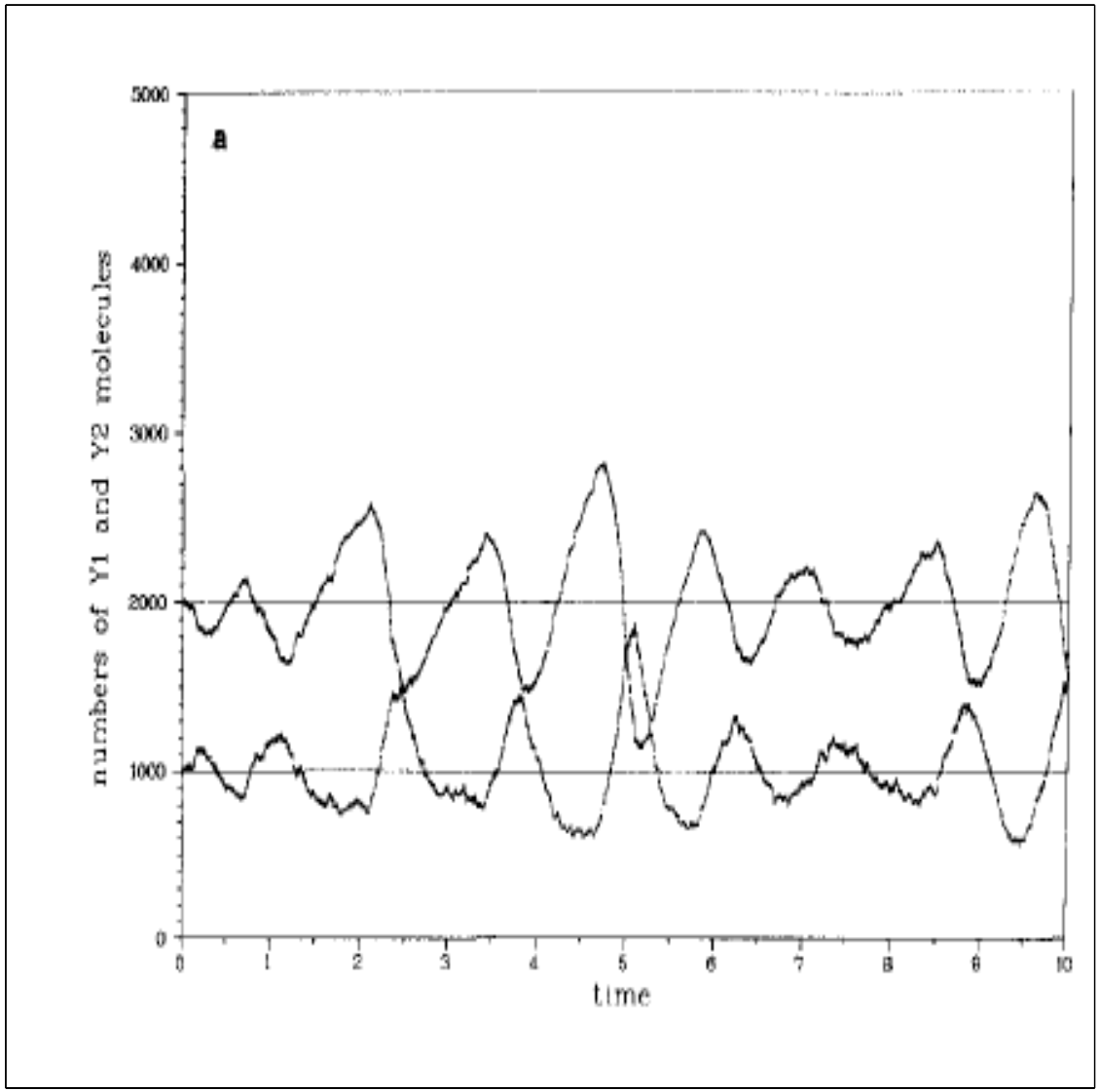
Example



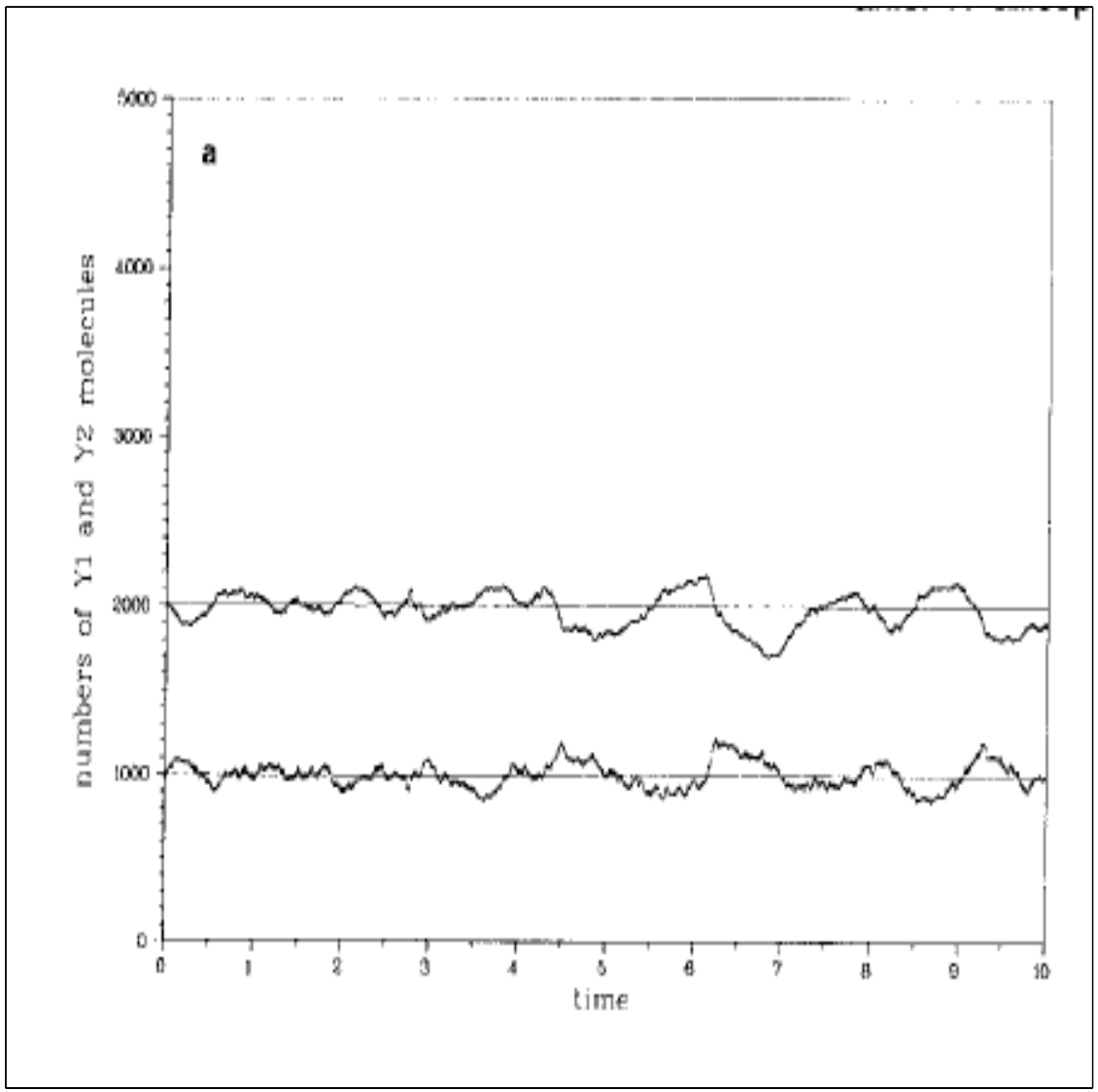
Example



Example


a
numbers of Y1 and Y2 molecules
5000
4000
3000
2000
1000
0
0
1
2
3
4
5
6
7
8
9
10
time

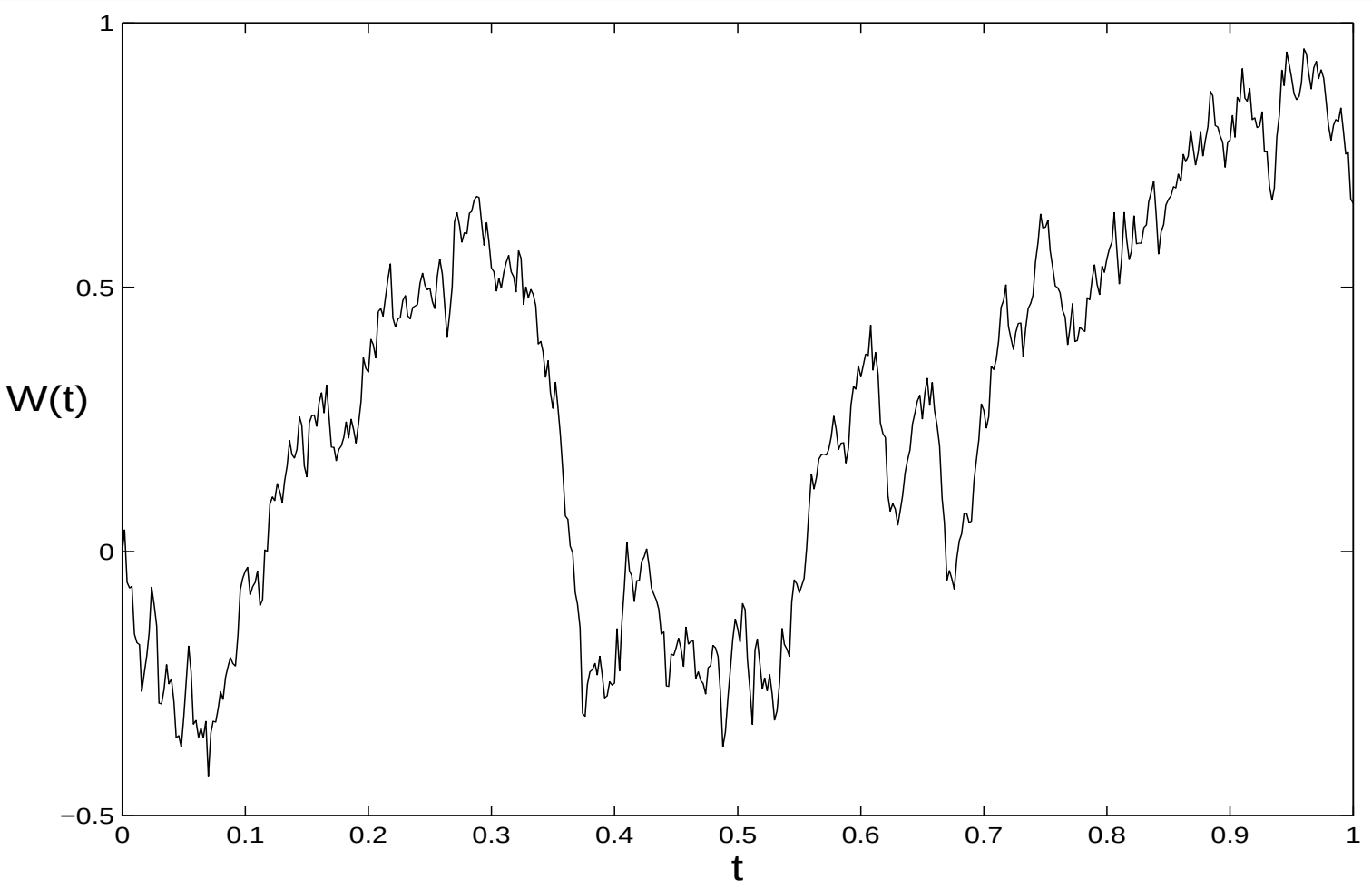
Example


a
numbers of Y1 and Y2 molecules
5000
4000
3000
2000
1000
0
0
1
2
3
4
5
6
7
8
9
10
time

SDEs and Brownian motion

- SDE models play a prominent role in a range of aplication areas, including biology, chemistry, epidemiology, mechanics, microelectronics, economics and finance.
- Complete understanding requires familiarity with advanced probability theory and stochastic processes.
- Simple simulation can be carried out just with background knowledge of Euler's method for ODEs and an intuitive understanding of random variables (..says Higham).

SDEs and Brownian motion



A sample path of one-dimensional Brownian motion.

SDEs and Brownian motion

- Brownian motion incorporates the natural randomness observed in biological phenomena into the biological models.
- A proper stochastic process (unlike commonly used "white noise").
- Brownian motion has analytically desirable properties (e.g. normally distributed, independent increments).

SDEs and Brownian motion

- (ODE) $\frac{dX(t)}{dt} = f(X(t), t)$

$$\Rightarrow X(t) = X(0) + \int_0^t f(X(t), t)dt.$$

- (SDE) $dX(t) = f(X(t), t)dt + g(X(t), t)dW$

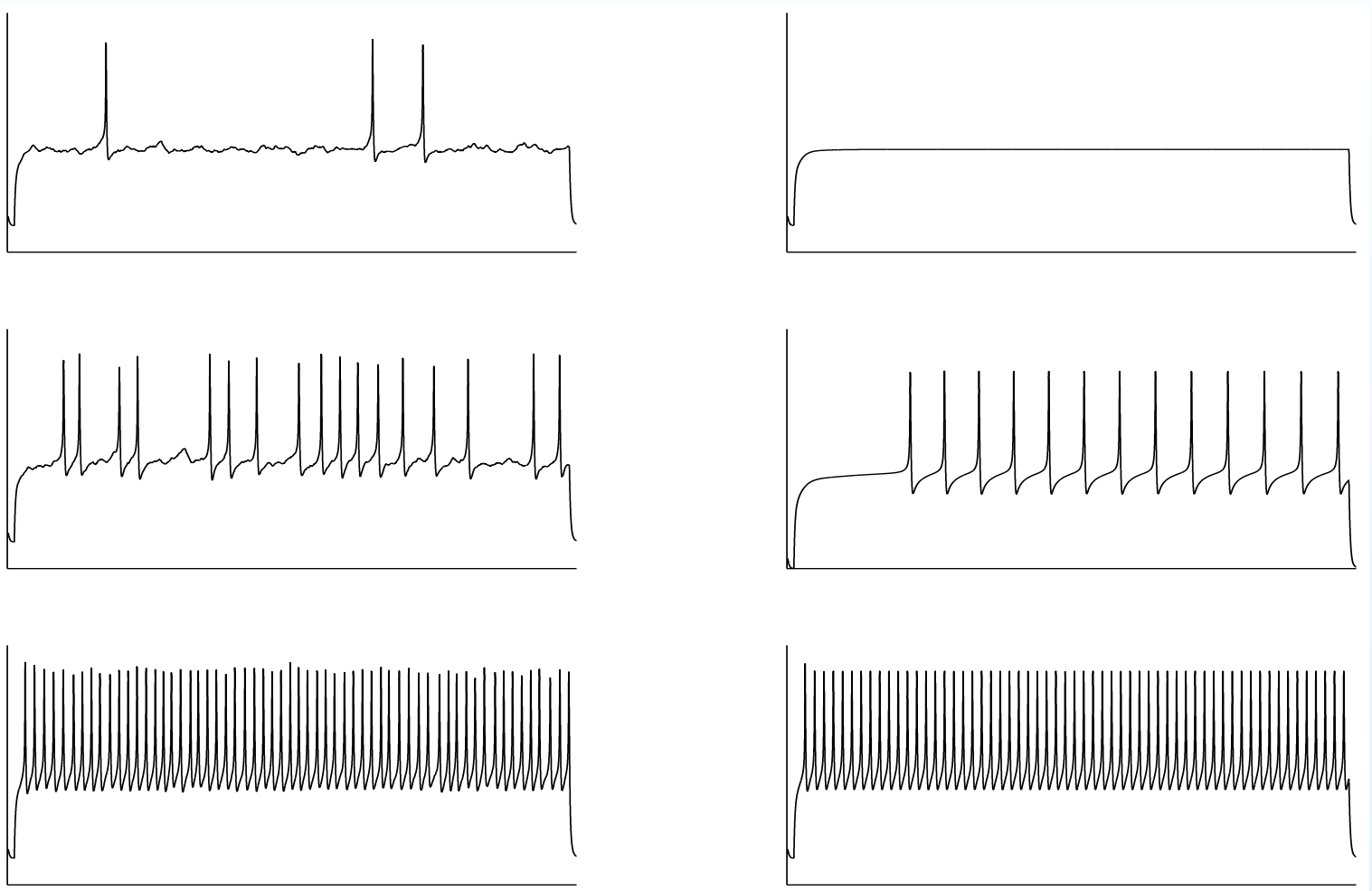
$$\Rightarrow X(t) = X(0) + \int_0^t f(X(t), t)dt$$

$$+ \int_0^t g(X(t), t)dW.$$

SDEs and Brownian motion

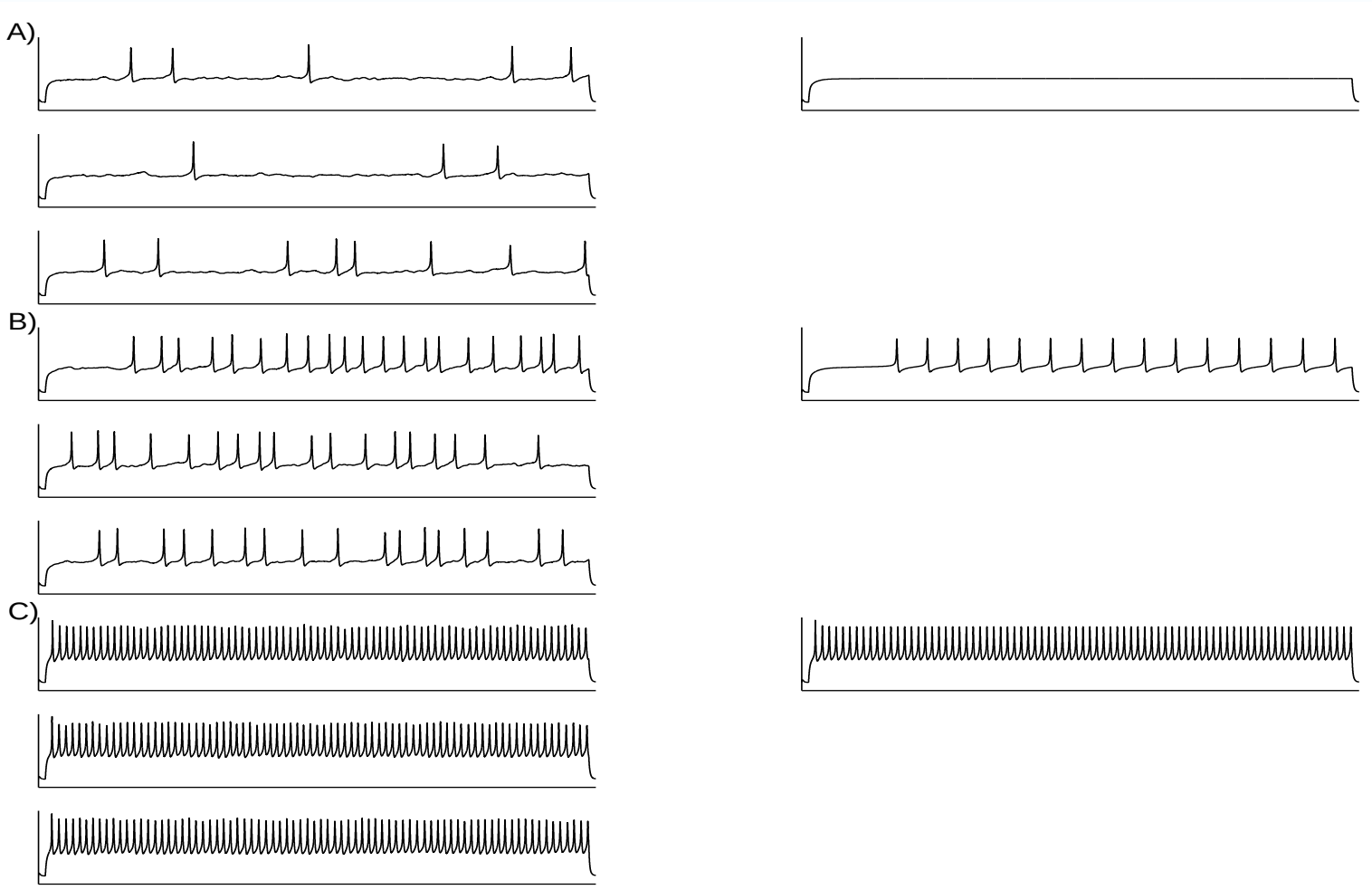
- Handling of stochastic integrals needs new type of calculus.
- Derivation of analytical results is possible, but usually tedious.
- Simulation of SDEs is, on the other hand, rather straightforward.
- Presently not implemented in existing simulation software.
- Own simulation software needed (e.g. MATLAB scripts).

SDEs and Brownian motion

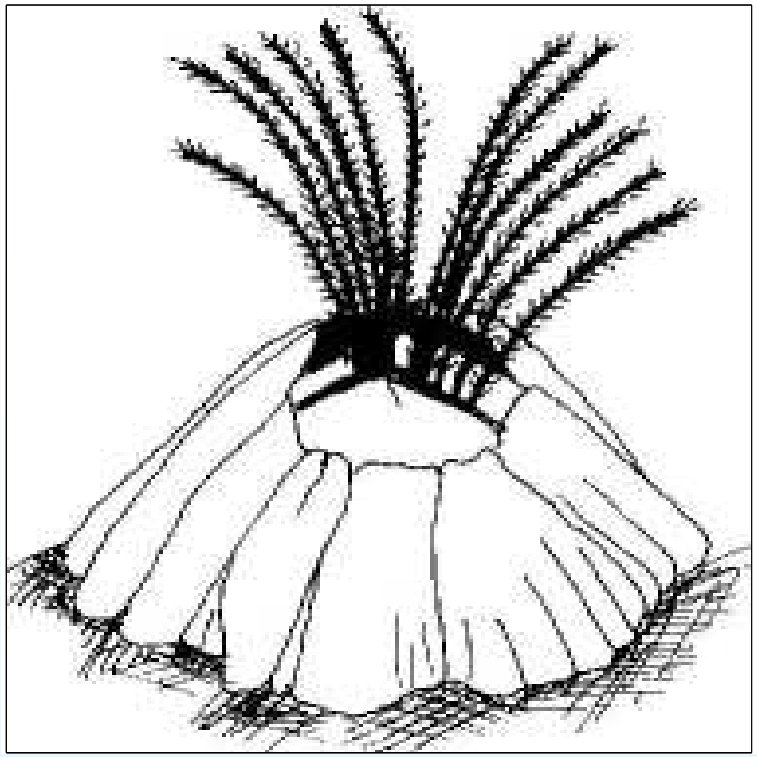


Deterministic and stochastic model for cerebellar granule cell.

SDEs and Brownian motion



Exercises



Simulation of a barnacle muscle fibre.

References

- Daniel T. Gillespie, "A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions," *Journal of Computational Physics*, vol. 22, no. 4, 1976.
- Desmond J. Higham, "An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations," *SIAM Review*, vol. 43, no. 3, 2001.