Chapter 1
Introduction to Stochastic Models in Biology

Susanne Ditlevsen and Adeline Samson

1.1 Introduction

This chapter is concerned with continuous time processes, which are often modeled as a system of ordinary differential equations (ODEs). These models assume that the observed dynamics are driven exclusively by internal, deterministic mechanisms. However, real biological systems will always be exposed to influences that are not completely understood or not feasible to model explicitly. Ignoring these phenomena in the modeling may affect the analysis of the studied biological systems. Therefore there is an increasing need to extend the deterministic models to models that embrace more complex variations in the dynamics. A way of modeling these elements is by including stochastic influences or noise. A natural extension of a deterministic differential equations model is a system of stochastic differential equations (SDEs), where relevant parameters are modeled as suitable stochastic processes, or stochastic processes are added to the driving system equations. This approach assumes that the dynamics are partly driven by noise.

All biological dynamical systems evolve under stochastic forces, if we define stochasticity as the parts of the dynamics that we either cannot predict or understand or that we choose not to include in the explicit modeling. To be realistic, models of biological systems should include random influences, since they are concerned with subsystems of the real world that cannot be sufficiently isolated from effects external to the model. The physiological justification to include erratic behaviors

S. Ditlevsen (✉)
Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark
e-mail: susanne@math.ku.dk

A. Samson
CNRS UMR8145, Laboratoire MAP5, Université Paris Descartes, 45 rue des Saints Pères, 75006 Paris, France
e-mail: adeline.samson@parisdescartes.fr

M. Bachar et al. (eds.), *Stochastic Biomathematical Models*, Lecture Notes in Mathematics 2058, DOI 10.1007/978-3-642-32157-3_1,

in a model can be found in the many factors that cannot be controlled, such as hormonal oscillations, blood pressure variations, respiration, variable neural control of muscle activity, enzymatic processes, energy requirements, cellular metabolism, sympathetic nerve activity, or individual characteristics like body mass index, genes, smoking, stress impacts, etc. Also to be considered are external influences, such as small differences in the experimental procedure, temperature, differences in preparation and administration of drugs (if this is included in the experiment). In addition, experimental runs may be conducted by different experimentalists who inevitably will exhibit small differences in procedures within the protocols. Different sources of errors will require different modeling of the noise, and these factors should be considered as carefully as the modeling of the deterministic part, in order to make the model predictions and parameter values possible to interpret.

It is therefore essential to understand and investigate the influence of noise in the dynamics. In many cases the noise simply blurs the underlying dynamics without qualitatively affecting it, as is the case with measurement noise or in many linear systems. However, in nonlinear dynamical systems with system noise, the noise will often drastically change the corresponding deterministic dynamics. In general, stochastic effects influence the dynamics, and may enhance, diminish or even completely change the dynamic behavior of the system.

1.2 Markov Chains and Discrete-Time Processes

A sequence of stochastic variables $\{X_n, n = 0, 1, \ldots\}$ is called a *stochastic process*. It could for example be measurements every 5 min of the level of blood glucose for a diabetic patient. The simplest type of stochastic process is one where the random variables are assumed independent, but this is often too simple to capture important features of the data. For example if the blood glucose is high, we would also expect it to be high 5 min later. The simplest type of stochastic process incorporating dependence between observations is a Markov process.

Definition 1.1 (Markov chain). A stochastic process $\{X_n, n = 0, 1, \ldots\}$ which can take values in the state space I is called a *discrete-time Markov chain* if for each $n = 0, 1, \ldots,$

$$\mathbb{P}(X_{n+1} = i_{n+1} \,|\, X_0 = i_0, \ldots, X_n = i_n) = \mathbb{P}(X_{n+1} = i_{n+1} \,|\, X_n = i_n)$$

for all possible values of $i_0, \ldots, i_{n+1} \in I$, whenever both sides are well-defined.

This means that conditionally on the present state of the system, its future and past are independent.

A classical example of a stochastic process in discrete time is a random walk. Consider the random migration of a molecule or a small particle arising from motion due to thermal energy. The particle starts at the origin at time 0. At each time unit

the particle moves one distance unit up with probability p or one distance unit down with probability $1-p$, independent of past movements. The random variable X_n then denotes the position of the particle at time n: $X_n = X_{n-1} \pm 1$. This random process $\{X_n\}_{n\in\mathbb{N}_0}$ is a discrete-time Markov chain which has state space the integers. Now let $p = 1/2$ and assume that we *accelerate* the process, so that displacements occur every δ units of time. At the same time, displacements decreases to ϵ units of distance. What happens in the limit of continuous time and space, i.e. when $\delta \to 0$ and $\epsilon \to 0$? Denote $X(t)$ the position of the particle at time t, and assume $X(0) = 0$. Let K denote the number of upward jumps made after a total of n jumps. Then the position of the particle after $n\delta$ units of time is given by

$$X(n\delta) = (K \cdot 1 + (n-K)\cdot(-1))\,\epsilon \;=\; (2K-n)\,\epsilon.$$

Since displacements occur independent of one another, the random variable K has a *binomial distribution* with parameters n and $1/2$. Thus,

$$\begin{aligned}\mathbb{E}\,(X(n\delta)) &= (2\mathbb{E}(K)-n)\,\epsilon = (2n/2-n)\epsilon = 0,\\ \mathrm{Var}\,(X(n\delta)) &= 4\epsilon^2\mathrm{Var}\,(K) = 4\epsilon^2\frac{1}{2}\left(1-\frac{1}{2}\right)n = \epsilon^2 n.\end{aligned}$$

Now let $\delta \to 0$ to obtain a continuous time process. Then

$$\mathrm{Var}\,(X(t))_{t=n\delta} = \epsilon^2 n = \epsilon^2 t/\delta.$$

We see that unless δ and ϵ go to 0 while keeping ϵ^2 proportional to δ, then the variance will be either 0 or infinite—both cases rather uninteresting! Thus, we put $\epsilon^2 = \sigma^2\delta$ for some constant $\sigma > 0$, and obtain a continuous time and space process with $\mathbb{E}\,(X(t)) = 0$ and $\mathrm{Var}\,(X(t)) = \sigma^2 t$ for all $t \geq 0$. With a little extra work and evoking the central limit theorem, one can show that the limiting process has a Gaussian distribution with zero mean and variance $\sigma^2 t$. This process is called *the Wiener process* (Fig. 1.1).

1.3 The Wiener Process (or Brownian Motion)

The most important stochastic process in continuous time is the Wiener process, also called Brownian Motion. It is used as a building block in more elaborate models. In 1828 the Scottish botanist Robert Brown observed that pollen grains suspended in water moved in an apparently random way, changing direction continuously. In 1905, Einstein explained this by the pollen grains being bombarded by water molecules, and Brown only contributed to the theory with his name. The precise mathematical formulation of this phenomenon was given by Norbert Wiener in 1923.

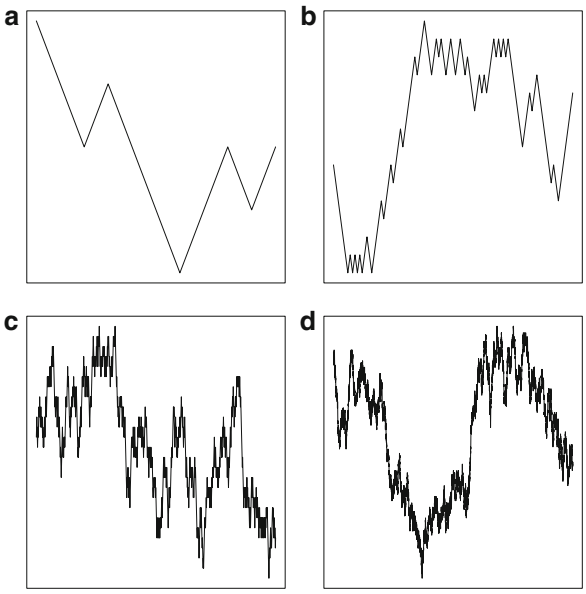

Fig. 1.1 Random walks over the time interval $[0, 1]$ with decreasing time steps δ and jump sizes $\epsilon = \sqrt{\delta}$. (**a**): $\delta = 0.1$. (**b**): $\delta = 0.01$. (**c**): $\delta = 0.001$. (**d**): $\delta = 0.0001$. The random walk approaches a Wiener process for decreasing step size

The Wiener process can be seen as the limit of a random walk when the time steps and the jump sizes go to 0 in a suitable way (see Sect. 1.2) and can formally be defined as follows.

Definition 1.2 (Wiener process). A stochastic process $\{W(t)\}_{t\geq 0}$ is called a *Wiener process* or a *Brownian motion* if

1. $W(0) = 0$.
2. $\{W(t)\}_{t\geq 0}$ has independent increments, i.e.

$$W_{t_1}, W_{t_2} - W_{t_1}, \ldots, W_{t_k} - W_{t_{k-1}}$$

   are independent random variables for all $0 \leq t_1 < t_2 < \cdots < t_k$.
3. $W(t + s) - W(s) \sim \mathcal{N}(0, t)$ for all $t > 0$.

Here, $\mathcal{N}(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2. Thus, the Wiener process is a Gaussian process: a stochastic process X is called a *Gaussian process* if for any finite set of indices $t_1, \ldots, t_k$ the vector of random variables $(X(t_1), \ldots, X(t_k))$ follows a k-dimensional normal distribution. In fact, it can be shown that any continuous time stochastic process with independent increments and finite second moments: $E(X^2(t)) < \infty$ for all t, is a Gaussian process provided that $X(t_0)$ is Gaussian for some t_0. The Wiener process is continuous with mean zero and variance proportional to the elapsed time: $\mathbb{E}(W(t)) = 0$ and $\mathrm{Var}(W(t)) = t$. If $\{X(t)\}_{t\geq 0}$ is a *stationary* stochastic process, then $\{X(t)\}_{t\geq 0}$ has the same distribution as $\{X(t+h)\}_{t\geq 0}$ for all $h > 0$. Thus, the Wiener process cannot

be stationary since the variance increases with t. The autocovariance function is given by $\mathrm{Cov}(W_t, W_s) = \min(s,t)$. The sample paths of a Wiener process behave "wildly" in that they are *nowhere differentiable*. To see what that means define the *total variation* of a real-valued function f on an interval $[a,b] \subset \mathbb{R}$ by the quantity

$$V_a^b(f) = \sup \sum_{k=1}^{n} |f(t_k) - f(t_{k-1})|$$

where the supremum is taken over all finite partitions $a \leq t_0 < \cdots < t_n \leq b$ of $[a,b]$. When $V_a^b(f) < \infty$ and f is right-continuous we say that f is of *bounded variation* on $[a,b]$. Functions that behave sufficiently "nice" are of bounded variation, if for example f is differentiable it is of bounded variation. It turns out that the Wiener process is everywhere of unbounded variation. This happens because the increments $W(t+\Delta t) - W(t)$ is on the order of $\sqrt{\Delta t}$ instead of Δt since the variance is Δt. Heuristically,

$$\begin{aligned} V_a^b(W) &= \sup \sum_{k=1}^{n} |W(t_k) - W(t_{k-1})| \\ &\geq \lim_{n\to\infty} \sum_{k=1}^{n} \left| W\left(a + \frac{k}{n}(b-a)\right) - W\left(a + \frac{(k-1)}{n}(b-a)\right) \right| \\ &\approx \lim_{n\to\infty} \sum_{k=1}^{n} \sqrt{\frac{1}{n}(b-a)} = \lim_{n\to\infty} \sqrt{n(b-a)} = \infty \end{aligned}$$

for *any* interval $[a,b]$. Trying to differentiate we see how this affects the limit

$$\lim_{\Delta t\to 0} \frac{|W(t+\Delta t) - W(t)|}{\Delta t} \approx \lim_{\Delta t\to 0} \frac{|\sqrt{\Delta t}|}{\Delta t} = \infty.$$

Now define the *quadratic variation* of a real-valued function f on $[a,b] \subset \mathbb{R}$ by

$$[f]_a^b = \sup \sum_{k=1}^{n} (f(t_k) - f(t_{k-1}))^2$$

where the supremum is taken as before. For continuous functions of bounded variation the quadratic variation is always 0, and thus, if $[f]_a^b > 0$ then $V_a^b(f) = \infty$. The quadratic variation of a Wiener process over an interval $[s,t]$ equals $t-s$, and in the limit we therefore expect

$$\lim_{\Delta t\to 0} (W(t+\Delta t) - W(t))^2 \approx \Delta t. \tag{1.1}$$

1.4 Stochastic Differential Equations

Assume that the ODE

$$\frac{dx}{dt} = a(x,t) \tag{1.2}$$

describes a one-dimensional dynamical system. Assume that $a(\cdot)$ fulfills conditions such that a unique solution exists, thus $x(t) = x(t; x_0, t_0)$ is a solution satisfying the initial condition $x(t_0) = x_0$. Given the initial condition, we know how the system behaves at all times t, even if we cannot find a solution analytically. We can always solve it numerically up to any desired precision. In many biological systems this is not realistic, and a more realistic model can be obtained if we allow for some randomness in the description.

A natural extension of a deterministic ODE model is given by an SDE model, where relevant parameters are randomized or modeled as random processes of some suitable form, or simply by adding a noise term to the driving equations of the system. This approach assumes that some degree of noise is present in the dynamics of the process. Here we will use the Wiener process. It leads to a mixed system with both a deterministic and a stochastic part in the following way [21,24]:

$$dX_t = \mu(X_t,t)\,dt + \sigma(X_t,t)\,dW_t \tag{1.3}$$

where $\{X_t = X(t)\}_{t\geq 0}$ is a stochastic process, not a deterministic function like in (1.2). This is indicated by the capital letter. Here $\{W_t = W(t)\}_{t\geq 0}$ is a Wiener process and since it is nowhere differentiable, we need to define what the differential means. It turns out that it is useful to write $dW_t = \xi_t dt$, where $\{\xi_t\}_{t\geq 0}$ is a white noise process, defined as being normally distributed for any fixed t and uncorrelated: $E(\xi_t \xi_s) = 0$ if $s \neq t$. Strictly speaking, the white noise process $\{\xi_t\}_{t\geq 0}$ does not exist as a conventional function of t, but could be interpreted as the generalized derivative of a Wiener process.

The functions $\mu(\cdot)$ and $\sigma(\cdot)$ can be nonlinear, $\mu(\cdot)$ is called the drift coefficient or the deterministic component, and $\sigma(\cdot)$ is called the diffusion coefficient or the stochastic component (system noise), that may depend on the state of the system, X_t. If $\mu(\cdot)$ and $\sigma(\cdot)$ do not depend on t the process is called time-homogeneous. Equation (1.3) should be interpreted in the following way:

$$X_t = X_0 + \int_{t_0}^{t} \mu(X_s,s)\,ds + \int_{t_0}^{t} \sigma(X_s,s)\,dW_s \tag{1.4}$$

where X_0 is a random variable independent of the Wiener process. It could simply be a constant. The first integral on the right hand side can be interpreted as an ordinary integral, but what is the second integral? The Wiener process is nowhere differentiable, so how do we give meaning to this differential?

Let us try the usual tricks from ordinary calculus, where we define the integral for a simple class of functions, and then extend by some approximation procedure to a larger class of functions. We want to define

$$\int_{t_0}^{t} f(s)\, dW_s. \tag{1.5}$$

If $f(t) \equiv \sigma$ is constant we would expect the integral (1.5) to equal $\sigma(W_t - W_{t_0})$. Note that this is a random variable with expectation 0 since the increments of a Wiener process has expectation 0. Assume that $f(t)$ is a non-random step function of the form $f(s) = \sigma_j$ on $t_j \le s < t_{j+1}$ for $j = 1, 2, \ldots, n$ where $t_0 = t_1 < t_2 < \cdots < t_{n+1} = t$. Then we define

$$\int_{t_0}^{t} f(s)\, dW_s = \sum_{j=1}^{n} \sigma_j \left(W_{t_{j+1}} - W_{t_j}\right).$$

It is natural to approximate a given function $f(t)$ by a step function. Now $f(t)$ can be random, but we will only consider functions that are *measurable* with respect to the σ-algebra generated by the random variables $\{W_s\}_{s \le t}$. The concepts of σ-algebra and measurable space will be defined in Chap. 2, for now they are not needed. Intuitively it means that the value of $f(t)$ can be determined from the values of W_s for $s \le t$. For example, we could take $f(t) = W_t$, but not $f(t) = W_{2t}$. We cannot look into the future! Moreover, we require that $\mathbb{E}[\int_{t_0}^{t} f(s)^2 ds] < \infty$. For the rest of this chapter we will always assume these conditions on the integrands.

Define a partition Π_n of the interval $[t_0, t]$ by $t_0 = t_1 < t_2 < \cdots < t_{n+1} = t$ where $|\Pi_n| = \max\{|t_{j+1} - t_j| : j = 1, \ldots, n\}$ is the norm of the partition, and approximate

$$f(t) \approx f(t_j^*) \quad \text{for} \quad t_j \le t < t_{j+1}$$

where the point t_j^* belongs to the interval $[t_j, t_{j+1}]$. Then we define

$$\int_{t_0}^{t} f(s)\, dW_s = \lim_{|\Pi_n| \to 0} \sum_{j=1}^{n} f(t_j^*) \left(W_{t_{j+1}} - W_{t_j}\right).$$

When $f(t)$ is stochastic it turns out that—unlike ordinary integrals—it makes a difference how t_j^* is chosen! To see this consider $f(t) = W_t$ and define two approximations: $t_j^* = t_j$, the left end point, and $t_j^* = t_{j+1}$, the right end point. Taking expectations we see that the two choices yield different results:

$$\begin{aligned}
\mathbb{E}\left[\sum_{j=1}^{n} W_{t_j} \left(W_{t_{j+1}} - W_{t_j}\right)\right] &= \sum_{j=1}^{n} \mathbb{E}\left[W_{t_j} \left(W_{t_{j+1}} - W_{t_j}\right)\right] \\
&= \sum_{j=1}^{n} \mathbb{E}\left[W_{t_j}\right] \mathbb{E}\left[W_{t_{j+1}} - W_{t_j}\right] = 0
\end{aligned}$$

because the Wiener process has independent increments with mean 0. On the other hand,

$$\mathbb{E}\left[\sum_{j=1}^{n} W_{t_{j+1}}\left(W_{t_{j+1}} - W_{t_j}\right)\right] = \sum_{j=1}^{n} \mathbb{E}\left[\left(W_{t_{j+1}} - W_{t_j}\right)^2\right]$$
$$= \sum_{j=1}^{n} \left(t_{j+1} - t_j\right) = t - t_0$$

where we have subtracted $\mathbb{E}\left[\sum_{j=1}^{n} W_{t_j}\left(W_{t_{j+1}} - W_{t_j}\right)\right] = 0$ and rearranged in the first equality sign, and the second equality sign is the variance of the Wiener process. Two useful and common choices are the following:

- *The Itô integral:* $t_j^* = t_j$, the left end point.
- *The Stratonovich integral:* $t_j^* = (t_j + t_{j+1})/2$, the mid point.

There are arguments for using either one or the other, most of them rather technical and we will not enter in this discussion here. Fortunately, though, the difference between the two is an ordinary integral and it is possible to calculate one integral from the other. Here we only use the Itô integral, and we call a process given by an equation of the form (1.3) an Itô process.

Properties of the Itô integral The usual linearity properties are also valid for Itô integrals,

$$\int_{t_0}^{t} f(s)\, dW_s = \int_{t_0}^{t^*} f(s)\, dW_s + \int_{t^*}^{t} f(s)\, dW_s$$

for $t_0 < t^* < t$, and

$$\int_{t_0}^{t} \left(af(s) + bg(s)\right)\, dW_s = a\int_{t_0}^{t} f(s)\, dW_s + b\int_{t_0}^{t} g(s)\, dW_s$$

where a and b are constants. Note that the terms are random variables. Moreover,

$$\mathbb{E}\left[\int_{t_0}^{t} f(s)\, dW_s\right] = 0.$$

Finally, we have *Itô's isometry*: Given the properties of the Wiener process, the variance of the stochastic process $\{\int_{t_0}^{t} f(s)\, dW_s\}_{t\geq 0}$ is equal to

$$\mathrm{Var}\left(\int_{t_0}^{t} f(s)\, dW_s\right) = \int_{t_0}^{t} \mathbb{E}\left[f^2(s)\right] ds.$$

A very important property is that solutions to (1.3) are Markov processes.

Some important examples of Itô processes are the following.

Wiener process with drift Imagine a particle suspended in water which is being bombarded by water molecules. The temperature of the water will influence the force of the bombardment, and thus we need a parameter σ to characterize this. Moreover, there is a water current which drives the particle in a certain direction, and we will assume a parameter μ to characterize the drift. To describe the displacements of the particle, the Wiener process can be generalized to the process

$$dX_t = \mu\, dt + \sigma\, dW_t$$

which has solution

$$X_t = x_0 + \mu\, t + \sigma\, W_t$$

for $X_0 = x_0$. It is thus normally distributed with mean $x_0 + \mu t$ and variance $\sigma^2 t$, as follows from the properties of the standard Wiener process. This process has been proposed as a simplified model for the membrane potential evolution in a neuron, see Chap. 5, Sect. 5.3.2.

Geometric Brownian motion Imagine a drug is supplied as a bolus to the blood stream and that the average metabolic process of the drug can be described by an exponential decay through the deterministic equation $x' = -ax$, where x is the concentration of the drug in plasma and a is the decay rate. The prime $'$ denotes derivative with respect to time. Assume now that the decay rate fluctuates randomly due to the complex working of the enzymatic machinery involved in the breakdown of the drug. That could be described by letting a vary randomly as $a = \mu + \sigma\xi_t$, where $\{\xi_t\}_{t\geq 0}$ is a Gaussian white noise process. Then $\xi_t dt$ can be written as the differential of a Wiener process, dW_t. This leads to the model

$$dX_t = \mu X_t\, dt + \sigma X_t\, dW_t.$$

It is shown below (Example 1.2, Sect. 1.6) that the explicit solution is

$$X_t = X_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W_t\right).$$

The process only takes positive values and X_t conditional on X_0 follows a log-normal distribution with parameters $\log(X_0) + (\mu - \sigma^2/2)t$ and $\sigma^2 t$.

Ornstein–Uhlenbeck process Imagine a process subject to a restoring force, i.e. the process is attracted to some constant level but is continuously perturbed by noise. An example is given by the membrane potential of a neuron that is constantly being perturbed by electrical impulses from the surrounding network, and at the same time is attracted to an equilibrium value depending on the resting potentials for various ions present at different concentrations inside the cell and in the interstitium, see Chap. 5, Sect. 5.3.5. This leads to the model

$$dX_t = -\left(\frac{X_t - \alpha}{\tau}\right) dt + \sigma\, dW_t, \tag{1.6}$$

with $\tau, \sigma > 0$. Here τ has units time, and is the typical time constant of the system. The autocorrelation is given by $\text{corr}(X_t, X_{t+s}) = e^{-s/\tau}$, and thus the autocorrelation has decreased with a factor of $1/e$ after τ units of time. It has the explicit solution (due to (1.9) below)

$$X_t = X_0 e^{-t/\tau} + \alpha(1 - e^{-t/\tau}) + e^{-t/\tau} \int_0^t e^{s/\tau} \sigma\, dW_s \tag{1.7}$$

and X_t conditional on X_0 is normally distributed with mean $E(X_t) = X_0 e^{-t/\tau} + \alpha(1-e^{-t/\tau})$ and variance $\text{Var}(X_t) = \sigma^2\tau(1-e^{-2t/\tau})/2$. If X_0 is normally distributed with mean α and variance $\sigma^2\tau/2$, then so is X_t for all t. Thus, contrary to the processes above, the Ornstein–Uhlenbeck process has a stationary solution.

Square-root process In many applications an unrestricted state space is unrealistic, and the variance is often observed to decrease with decreasing distance to some lower level. For example, the hyper-polarization caused by inhibitory reversal potentials in neuron membranes is smaller if the membrane potential is closer to the inhibitory reversal potential, see Chap. 5, Sect. 5.3.6. For simplicity we assume this lower limit in the state space equal to 0. This leads to the model

$$dX_t = -\left(\frac{X_t - \alpha}{\tau}\right) dt + \sigma \sqrt{X_t}\, dW_t. \tag{1.8}$$

The process is also called the Cox-Ingersoll-Ross process in the financial literature [6], or the Feller process in the neuronal literature, because [16] proposed it as a model for population growth. If $2\alpha/(\tau\sigma^2) \geq 1$ the process stays positive, (see Chap. 2, Example 2.3), and admits a stationary distribution. The transition density is a non-central chi-square distribution with conditional mean and variance

$$\begin{aligned}
\mathbb{E}(X_t|X_0) &= \alpha + (X_0 - \alpha)e^{-t/\tau} \\
\text{Var}(X_t|X_0) &= \alpha\frac{\tau\sigma^2}{2}(1 - e^{-t/\tau})^2 + X_0\tau\sigma^2(1 - e^{-t/\tau})e^{-t/\tau}
\end{aligned}$$

The asymptotic stationary distribution is a gamma distribution with shape parameter $2\alpha/(\tau\sigma^2)$ and scale parameter $\tau\sigma^2/2$.

When the diffusion term does not depend on the state variable X_t as in the Wiener process with drift and the Ornstein–Uhlenbeck process, we say that it has *additive noise*. In this case the Itô and the Stratonovich integrals yield the same process, so it does not matter which calculus we choose. In the case of Geometric Brownian motion or the square root process we say that it has *multiplicative noise*. The four processes are simulated in Fig. 1.2.

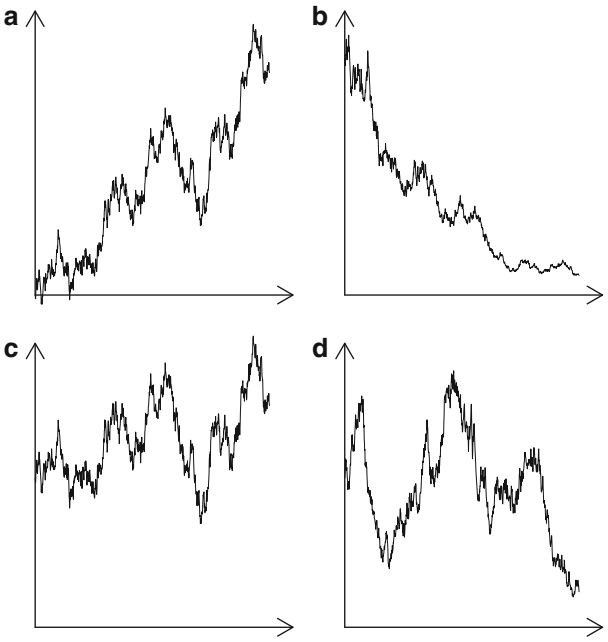

Fig. 1.2 Sample paths from (**a**): a Wiener process with drift, (**b**): a Geometric Brownian motion, (**c**): an Ornstein–Uhlenbeck process, and (**d**): a square-root process. Note how the amplitude of the noise does not change over time for the Wiener and the Ornstein–Uhlenbeck process (**a** and **c**), whereas for Geometric Brownian motion and the square-root process (**b** and **d**), the amplitude of the noise depends on the state variable

1.5 Existence and Uniqueness

To ensure the existence of a solution to (1.3) for $0 \leq t \leq T$ where T is fixed, the following is sufficient:

$$|\mu(t,x)| + |\sigma(t,x)| \leq C(1+|x|)$$

for some constant C [22, 24]. This ensures that $\{X_t\}_{t\geq 0}$ does not explode, i.e that $\{|X_t|\}_{t\geq 0}$ does not tend to ∞ in finite time. To ensure uniqueness of a solution the Lipschitz condition is sufficient:

$$|\mu(t,x) - \mu(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq D|x-y|$$

for some constant D. Note that only *sufficient* conditions are stated, and in many biological applications these are too strict, and weaker conditions can be found. We will not treat these here, though. In Chap. 2 conditions on the functions μ and σ to ensure that the process stays away from the boundaries without assuming the

Lipschitz condition are discussed in detail. Note also that the above conditions are fulfilled for three of the processes described above. The square root process does not fulfill the Lipschitz condition at 0 and is treated in Chap. 2, Example 2.3. In general, many realistic biological models do not fulfill the Lipschitz condition, and the more advanced tools of Chap. 2 are necessary to check if the model is well behaved.

1.6 Itô's Formula

Stochastic differentials do not obey the ordinary chain rule as we know it from classical calculus [24, 34]. An additional term appears because $(dW_t)^2$ behaves like dt, see (1.1). We have

Theorem 1.1 (Itô's formula). *Let $\{X_t\}_{t\geq 0}$ be an Itô process given by*

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)\, dW_t$$

and let $f(t, x)$ be a twice continuously differentiable function in x and once continuously differentiable function in t. Then

$$Y_t = f(t, X_t)$$

is also an Itô process, and

$$dY_t = \frac{\partial f}{\partial t}(t, X_t)dt + \frac{\partial f}{\partial x}(t, X_t)dX_t + \frac{1}{2}\sigma^2(t, X_t)\frac{\partial^2 f}{\partial x^2}(t, X_t)dt. \tag{1.9}$$

The first two terms on the right hand side correspond to the chain rule we know from classical calculus, but an extra term appears in stochastic calculus because the Wiener process is of unbounded variation, and thus the quadratic variation comes into play.

Example 1.1. Let us calculate the integral

$$\int_0^t W_s dW_s.$$

From classical calculus we expect a term like $\frac{1}{2}W_t^2$ in the solution. Thus, we choose $f(t, x) = \frac{1}{2}x^2$ and $X_t = W_t$ and apply Itô's formula to

$$Y_t = f(t, W_t) = \frac{1}{2}W_t^2.$$

We obtain

$$\begin{aligned} dY_t &= \frac{\partial f}{\partial t}(t, W_t)dt + \frac{\partial f}{\partial x}(t, W_t)dW_t + \frac{1}{2}\sigma^2(t, W_t)\frac{\partial^2 f}{\partial x^2}(t, W_t)dt \\ &= 0 + W_t dW_t + \frac{1}{2}dt \end{aligned}$$

because $\sigma^2(t, W_t) = 1$. Hence

$$Y_t = \frac{1}{2}W_t^2 = \int_0^t W_s dW_s + \frac{1}{2}\int_0^t ds = \int_0^t W_s dW_s + \frac{1}{2}t$$

and finally

$$\int_0^t W_s dW_s = \frac{1}{2}W_t^2 - \frac{1}{2}t.$$

Example 1.2. Let us find the solution $\{X_t\}_{t\geq 0}$ to the Geometric Brownian motion

$$dX_t = \mu X_t\, dt + \sigma X_t\, dW_t.$$

Rewrite the equation as

$$\frac{dX_t}{X_t} = \mu\, dt + \sigma\, dW_t.$$

Thus, we have

$$\int_0^t \frac{dX_s}{X_s} = \mu\, t + \sigma\, W_t, \tag{1.10}$$

which suggests to apply Itô's formula on $f(t, x) = \log x$. We obtain

$$\begin{aligned} dY_t = d(\log X_t) &= \frac{\partial f}{\partial t}(t, X_t)dt + \frac{\partial f}{\partial x}(t, X_t)dX_t + \frac{1}{2}\sigma^2(t, X_t)\frac{\partial^2 f}{\partial x^2}(t, X_t)dt \\ &= 0 + \frac{1}{X_t}dX_t + \frac{1}{2}\sigma^2 X_t^2\left(-\frac{1}{X_t^2}\right)dt \;=\; \frac{dX_t}{X_t} - \frac{1}{2}\sigma^2 dt \end{aligned}$$

and thus

$$\frac{dX_t}{X_t} = d(\log X_t) + \frac{1}{2}\sigma^2 dt. \tag{1.11}$$

Integrating (1.11) and using (1.10) we finally obtain

$$\log\frac{X_t}{X_0} = \int_0^t \frac{dX_s}{X_s} - \frac{1}{2}\sigma^2 t \;=\; \mu\, t + \sigma\, W_t - \frac{1}{2}\sigma^2 t$$

and so

$$X_t = X_0 \exp\left\{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma\, W_t\right\}.$$

Note that it is simply the exponential of a Wiener process with drift.

The solution (1.7) of the Ornstein–Uhlenbeck process can be found by multiplying both sides of (1.6) with $e^{-t/\tau}$ and then apply Itô's formula to $e^{-t/\tau}X_t$. We will not do that here.

1.7 Monte Carlo Simulations

The solution of an Itô process is rarely explicit. When no explicit solution is available we can approximate different characteristics of the process by simulation, such as sample paths, moments, qualitative behavior etc. Usually such simulation methods are based on discrete approximations of the continuous solution to an SDE [19, 22]. Different schemes are available depending on how good we want the approximation to be, which comes at a price of computer time. Assume we want to approximate a solution to (1.3) in the time interval $[0, T]$. Consider the time discretization

$$0 = t_0 < t_1 < \cdots < t_j < \cdots < t_N = T$$

and denote the time steps by $\Delta_j = t_{j+1} - t_j$ and the increments of the Wiener process by $\Delta W_j = W_{t_{j+1}} - W_{t_j}$. Then $\Delta W_j \sim \mathcal{N}(0, \Delta_j)$, which we can use to construct approximations by drawing normally distributed numbers from a random number generator. For simplicity assume that the process is time-homogenous.

1.7.1 The Euler–Maruyama Scheme

The simplest scheme, referred to as the Euler–Maruyama scheme, is the stochastic analogue of the deterministic Euler scheme. Approximate the process X_t at the discrete time-points $t_j, 1 \leq j \leq N$ by the recursion

$$Y_{t_{j+1}} = Y_{t_j} + \mu(Y_{t_j})\Delta_j + \sigma(Y_{t_j})\Delta W_j \;\; ; \;\; Y_{t_0} = x_0 \tag{1.12}$$

where $\Delta W_j = \sqrt{\Delta_j} \cdot Z_j$, with Z_j being standard normal variables with mean 0 and variance 1 for all j. This procedure approximates the drift and diffusion functions by constants between time steps, so obviously the approximation improves for smaller time steps. To evaluate the convergence things are more complicated for stochastic processes, and we operate with two criteria of optimality: the *strong* and the *weak* orders of convergence [2, 3, 19, 22].

Consider the expectation of the absolute error at the final time instant T of the Euler–Maruyama scheme. It can be shown that there exist constants $K > 0$ and $\delta_0 > 0$ such that

$$\mathbb{E}(|X_T - Y_{t_N}|) \leq K\delta^{0.5}$$

for any time discretization with maximum step size $\delta \in (0, \delta_0)$. We say that the approximating process Y *converges in the strong sense* with order 0.5. This is similar to how approximations are evaluated in deterministic systems, only here we take expectations, since X_T and Y_{t_N} are random variables. Compare with the Euler scheme for an ODE which has order of convergence 1. Sometimes we do not need a close *pathwise* approximation, but only some function of the value at a given

final time T, e.g. $\mathbb{E}(X_T)$, $\mathbb{E}(X_T^2)$ or generally $\mathbb{E}(g(X_T))$. In this case we have that there exist constants $K > 0$ and $\delta_0 > 0$ such that for any polynomial g

$$|\mathbb{E}(g(X_T) - \mathbb{E}(g(Y_{t_N})))| \leq K\delta$$

for any time discretization with maximum step size $\delta \in (0, \delta_0)$. We say that the approximating process Y *converges in the weak sense* with order 1.

1.7.2 The Milstein Scheme

To improve the accuracy of the approximation we add a second-order term that appears from Itô's formula. Approximate X_t by

$$Y_{t_{j+1}} = \underbrace{\underbrace{Y_{t_j} + \mu(Y_{t_j})\Delta_j + \sigma(Y_{t_j})\Delta W_j}_{\text{Euler–Maruyama}} + \frac{1}{2}\sigma(Y_{t_j})\sigma'(Y_{t_j})((\Delta W_j)^2 - \Delta_j)}_{\text{Milstein}} \quad (1.13)$$

where the prime $'$ denotes derivative. It is not obvious exactly how this term appears, but it can be derived through *stochastic Taylor expansions*. The Milstein scheme converges in the strong sense with order 1, and could thus be regarded as the proper generalization of the deterministic Euler-scheme.

If $\sigma(X_t)$ does not depend on $\{X_t\}_{t\geq 0}$ the Euler–Maruyama and the Milstein schemes coincide.

1.8 Inference

Estimation of parameters of an Itô process is a key issue in applications. Estimation in continuously observed Itô processes has been widely studied [23, 27]. However, biological processes are normally observed at discrete times. Parametric inference for discretely observed Itô processes can be complicated depending on the model considered and on the presence of measurement noise. The transition densities are only known for a few specific Itô processes (Wiener process with drift, Ornstein–Uhlenbeck process, Square-root process). Likelihood functions of data sampled from these processes have then an explicit form, and the maximum likelihood estimator (MLE) of the parameters can thus be computed. These estimators have nice statistical properties like consistency and asymptotic normality [7]. Consistency means that when the number of observations go to infinity, the estimator will converge in probability to the true value. Asymptotic normality means that the large sample distribution of the estimator will be close to normal, which is useful for constructing confidence intervals.

When the transition densities are not available, the likelihood function cannot be directly computed. Several estimation methods have been proposed to circumvent this difficulty: methods based on an approximation of the transition density by Hermite expansions [1], simulation based methods, also called Monte Carlo methods [13, 25], martingale estimating functions [4], see also [19, 23, 27] and references therein.

Estimation is more difficult when the process is observed with measurement noise. The likelihood function is explicit in some few specific cases for which filtering techniques can be applied [15]. Otherwise, methods based on simulations or on the Expectation-Maximization algorithm have been developed [10]. The Bayesian approach, which is an alternative to the maximum likelihood approach, can be applied to a large variety of problems when it is combined with sample path simulations or Euler–Maruyama approximations [11].

We present below some of these methods.

1.8.1 Maximum Likelihood

Observations of an Itô process without measurement noise Consider discrete observations $x_0, \ldots, x_N$ at time points $0 = t_0 < t_1 < \ldots < t_j < \ldots < t_N = T$ of an Itô process X which depends on an unknown parameter vector θ,

$$dX_t = \mu(t, X_t; \theta)dt + \sigma(t, X_t; \theta)\, dW_t. \tag{1.14}$$

The vector of observations is denoted $x_{0:N} = (x_0, \ldots, x_N)$. Bayes' rule combined with the Markovian nature of the process X, which the discrete data inherit, imply that the likelihood function of θ is simply the product of transition densities,

$$L(\theta, x_{0:N}) = p(x_0; \theta) \prod_{j=1}^{N} p(x_j | x_{j-1}; \theta), \tag{1.15}$$

where $p(x_0; \theta)$ is the density of the initial variable X_0 and $p(x_t | x_s; \theta), s < t$ is the transition density of X, i.e. the conditional density of X_t at time t, given that it was at $X_s = x_s$ at an earlier time s. We will normally ignore the asymptotically unimportant distribution of X_0 by setting $p(x_0; \theta) = 1$. The vector of partial derivatives of the log-likelihood function with respect to the coordinates of θ is called the *score function*,

$$U(\theta) = \frac{\partial}{\partial \theta} \log L(\theta, x_{0:N}) = \sum_{i=1}^{N} \frac{\partial}{\partial \theta} \log p(x_j | x_{j-1}; \theta), \tag{1.16}$$

which under mild regularity conditions is a *martingale* under $\mathbb{P}_\theta$.

Definition 1.3 (Martingale). A stochastic process $\{X_n, n = 1, 2, \ldots\}$ is called a *martingale* if for all $n = 1, 2, \ldots$,

$$\mathbb{E}(|X_n|) < \infty$$
$$\mathbb{E}(X_{n+1} \mid X_1, \ldots, X_n) = X_n$$

i.e., the conditional expected value of the next observation, given all the past observations, is equal to the last observation.

The MLE usually solves the estimating equation $U(\theta) = 0$. Under mild regularity conditions it is consistent and asymptotically normally distributed [7].

Example 1.3. Let us calculate the likelihood function of an Ornstein–Uhlenbeck process (1.6). The unknown parameter to be estimated is $\theta = (\tau, \alpha, \sigma)$. Denote the length of the observation time intervals by $\Delta_j = t_j - t_{j-1}$, for $j = 1, \ldots, N$. Equation (1.7) provides an explicit expression of X_{t_j} as a function of $X_{t_{j-1}}$ and θ:

$$X_{t_j} = X_{t_{j-1}} e^{-\Delta_j/\tau} + \alpha(1 - e^{-\Delta_j/\tau}) + \eta_i, \quad \eta_i \sim \mathcal{N}(0, \frac{\sigma^2\tau}{2}(1 - e^{-2\Delta_j/\tau})). \quad (1.17)$$

The likelihood (1.15) is thus explicit and equal to

$$L(\theta, x_{0:N}) = p(x_0; \theta) \prod_{j=1}^{N} \varphi\left(x_j; x_{j-1} e^{-\frac{\Delta_j}{\tau}} + \alpha\left(1 - e^{-\frac{\Delta_j}{\tau}}\right), \frac{\sigma^2\tau}{2}\left(1 - e^{-\frac{2\Delta_j}{\tau}}\right)\right),$$

where $\varphi(x; \mu, \sigma^2)$ denotes the density of a Gaussian variable with mean μ and variance σ^2. The unique maximum of the likelihood function provides the MLE $\hat{\theta} = (\hat{\tau}, \hat{\alpha}, \hat{\sigma}^2)$. When $\Delta_j = \Delta$ is constant the MLE is given by the equations

$$\hat{\alpha} = \frac{\sum_{j=1}^{n}(X_j - X_{j-1}e^{-\Delta/\hat{\tau}})}{n(1 - e^{-\Delta/\hat{\tau}})},$$
$$e^{-\Delta/\hat{\tau}} = \frac{\sum_{j=1}^{n}(X_j - \hat{\alpha})(X_{j-1} - \hat{\alpha})}{\sum_{j=1}^{n}(X_{j-1} - \hat{\alpha})^2},$$
$$\hat{\sigma}^2 = \frac{2\sum_{j=1}^{n}(X_j - \hat{\alpha} - (X_{j-1} - \hat{\alpha})e^{-\Delta/\hat{\tau}})^2}{n(1 - e^{-2\Delta/\hat{\tau}})\hat{\tau}}.$$

It requires that $\sum_{j=1}^{n}(X_j - \hat{\alpha})(X_{j-1} - \hat{\alpha}) > 0$. Otherwise there is no solution.

When the transition density function $p(\cdot)$ is unknown, the likelihood is not explicit. A simple approximation to the likelihood function is obtained by approximating the transition density by a Gaussian density with the correct first and second conditional moments,

$$p(x|y; \theta) \approx q(x|y; \theta) = \frac{1}{\sqrt{2\pi\phi(\Delta, x; \theta)}} \exp\left[-\frac{(y - F(\Delta, x; \theta))^2}{2\phi(\Delta, x; \theta)}\right]$$

where $F(\Delta, x; \theta) = \mathbb{E}_\theta(X_\Delta | X_0 = x)$ and $\phi(\Delta, x; \theta) = \mathrm{Var}_\theta(X_\Delta | X_0 = x)$. In this way we obtain the *quasi-likelihood*

$$L(\theta) \approx QL(\theta) = \prod_{j=1}^{N} q(X_{j-1} | X_j; \theta).$$

By differentiation with respect to the parameter, we obtain the quasi-score function

$$\frac{\partial}{\partial \theta} \log QL(\theta) = \sum_{j=1}^{N} \left\{ \frac{\frac{\partial}{\partial \theta} F(\Delta, X_{j-1}; \theta)}{\phi(\Delta, X_{j-1}; \theta)} [X_j - F(\Delta, X_{j-1}; \theta)] \right. \tag{1.18}$$

$$\left. + \frac{\frac{\partial}{\partial \theta} \phi(\Delta, X_{j-1}; \theta)}{2\phi(\Delta, X_{j-1}; \theta)^2} \left[(X_j - F(\Delta, X_{j-1}; \theta))^2 - \phi(\Delta, X_{j-1}; \theta) \right] \right\},$$

which is clearly a martingale under $\mathbb{P}_\theta$. It is a particular case of a quadratic martingale estimating function considered by [4, 5].

Another approach is to compute an approximation to $p(\cdot)$ based on the Euler–Maruyama (1.12) or the Milstein (1.13) schemes. In general, this approximation will converge to $p(\cdot)$ as $\Delta \to 0$. More precisely, the Euler–Maruyama approximation of (1.15) consists in replacing $p(\cdot)$ by the Gaussian density of the Euler–Maruyama scheme:

$$L(\theta, x_{0:N}) \approx p(x_0, \theta) \prod_{j=1}^{N} \varphi(x_j; x_{j-1} + \Delta_j \mu(t_{j-1}, x_{j-1}, \theta), \Delta_j \sigma^2(t_{j-1}, x_{j-1}, \theta)).$$

When the interval lengths (Δ_j) are fixed and large, the Euler–Maruyama scheme provides a poor approximation to the diffusion. An alternative is to approximate the transition density via simulation of finer sample paths. A set of auxiliary latent data points are introduced between each pair of observations. Along these auxiliary latent data points, the process can be finely sampled and the likelihood function is then approximated via numerical integration (also called Monte Carlo method) [13, 25]. We detail the approximation of the transition density $p(x_j | x_{j-1}; \theta)$ on the interval $[t_{j-1}, t_j]$ for a fixed $j \in \{1, \ldots, N\}$. The time interval $[t_{j-1}, t_j]$ is discretized in K sub-intervals $t_{j-1} = \tau_0^{(j)} < \tau_1^{(j)} < \ldots < \tau_k^{(j)} < \ldots < \tau_K^{(j)} = t_j$. The transition density $p(x_j | x_{j-1}; \theta)$ can be written as

$$p\left(x_j | x_{j-1}; \theta\right) = p\left(x_{\tau_K^{(j)}} | x_{\tau_0^{(j)}}; \theta\right) = \int p\left(x_{\tau_K^{(j)}} | X_{\tau_{K-1}^{(j)}}, \ldots, X_{\tau_1^{(j)}}, x_{\tau_0^{(j)}}; \theta\right)$$

$$\times p\left(X_{\tau_{K-1}^{(j)}}, \ldots, X_{\tau_1^{(j)}} | x_{\tau_0^{(j)}}; \theta\right) dX_{\tau_{K-1}^{(j)}} \ldots dX_{\tau_1^{(j)}}$$

$$= \mathbb{E}\left[p(x_{\tau_K^{(j)}} | X_{\tau_{K-1}^{(j)}}; \theta) \right],$$

where the expectation is taken under the distribution $p(X_{\tau_{K-1}^{(j)}}, \ldots, X_{\tau_1^{(j)}} \mid x_{\tau_0^{(j)}}; \theta)$. By simulating M independent sample paths $(x^m_{\tau_{K-1}^{(j)}}, \ldots, x^m_{\tau_1^{(j)}})_{m=1,\ldots,M}$ under this distribution, the transition density $p(x_j|x_{j-1};\theta)$ can be approximated by

$$p^{(M)}(x_j|x_{j-1};\theta) = \frac{1}{M}\sum_{m=1}^{M} p\left(x_j|x^m_{\tau_{K-1}^{(j)}}, \ldots, x^m_{\tau_1^{(j)}}, x_{j-1};\theta\right) \qquad (1.19)$$
$$= \frac{1}{M}\sum_{m=1}^{M} p\left(x_j|x^m_{\tau_{K-1}^{(j)}};\theta\right).$$

By the law of large numbers, the approximating transition density $p^{(M)}(x_j|x_{j-1};\theta)$ converges to $p(x_j|x_{j-1};\theta)$. For a given j, the simulation of the sample paths $(x^m_{\tau_{K-1}^{(j)}}, \ldots, x^m_{\tau_1^{(j)}})_{m=1,\ldots,M}$ can be performed using the Euler–Maruyama or Milstein schemes with the initial condition x_{j-1}. The densities of the right side of (1.19) are then explicit Gaussian densities. The Euler–Maruyama approximation gives

$$p^{(M)}_{EM}(x_j|x_{j-1};\theta)$$
$$= \frac{1}{M}\sum_{m=1}^{M}\left[\varphi\left(x_j; x^m_{\tau_{K-1}^{(j)}} + \Delta_K^{(j)}\mu(\tau_{K-1}^{(j)}, x^m_{\tau_{K-1}^{(j)}}, \theta), \Delta_K^{(j)}\sigma^2(\tau_{K-1}^{(j)}, x^m_{\tau_{K-1}^{(j)}}, \theta)\right)\right.$$

with $\Delta_k^{(j)} = \tau_k^{(j)} - \tau_{k-1}^{(j)}$.

However, this approach can have poor convergence properties as the simulations are based on unconditional distributions, especially the variance resulting from the Monte Carlo integration can be large. A more appropriate strategy to reduce the variance consists in importance sampling: instead of simulating the sample paths using the Euler–Maruyama or the Milstein schemes, the trajectories $(x^m_{\tau_1^{(j)}}, \ldots, x^m_{\tau_{K-1}^{(j)}})_{m=1,\ldots,M}$ are generated using Brownian bridges, conditioning the proposed bridge on the events x_{j-1} and x_j [12]. More precisely, for $k = 1, \ldots, (K-1)$, $x^m_{\tau_k^{(j)}}$ is simulated with:

$$x^m_{\tau_k^{(j)}} = x_{t_{j-1}} + \frac{x_{t_j} - x_{t_{j-1}}}{t_j - t_{j-1}}(\tau_k^{(j)} - t_{j-1}) + \overline{B}^m_{\tau_k^{(j)}}, \qquad (1.20)$$

where $\overline{B}$ is a standard Brownian bridge on $[t_{j-1}, t_j]$ equal to zero for $t = t_{j-1}$ and $t = t_j$, which can be easily simulated.

Observations of an Itô process with measurement noise Consider that the Itô process is discretely observed with measurement noise. Let $y_{0:N} = (y_0, \ldots, y_N)$ denote the vector of noisy observations:

$$y_j = X_{t_j} + \gamma \varepsilon_j, \tag{1.21}$$

where X is defined by (1.14), the ε_j's are the measurement error random variables assumed to be independent, identically distributed with a centered normal distribution with unit variance and independent of $\{X_t\}_{t \geq 0}$, and γ is the measurement noise level. The observed process is no longer Markov. The likelihood function of the data $y_{0:N}$ can be computed by recursive conditioning:

$$L(\theta, y_{0:N}) = p(y_0; \theta) \prod_{j=1}^{N} p(y_j | y_{0:j-1}; \theta),$$

where $y_{0:j} = (y_0, \ldots, y_j)$ is the vector of observations until time t_j. It is thus sufficient to compute the distribution of y_j given $y_{0:j-1}$ which can be written

$$p(y_j | y_{0:j-1}; \theta) = \int p(y_j | X_{t_j}; \theta) p(X_{t_j} | y_{0:j-1}; \theta) dX_{t_j}.$$

This conditional distribution is rarely explicit, though for the Ornstein–Uhlenbeck process it is. Since the innovation noise η_j of the discretization of the Ornstein–Uhlenbeck process (1.17) and the observation noise ε_j are Gaussian variables, the law of y_j given $y_{0:j-1}$ can be obtained by elementary computations on Gaussian laws if we know the mean and covariance of the conditional distribution of X_{t_j} given $y_{0:j-1}$. This conditional distribution can be exactly computed using Kalman recursions as proposed by [15, 26]. The Kalman filter is an iterative procedure which computes recursively the following conditional quantities: $\hat{X}_j^-(\theta) = \mathbb{E}(X_{t_j} | y_{0:j-1}; \theta)$, $V_j^-(\theta) = \mathbb{E}((X_{t_j} - \hat{X}_j^-)^2; \theta)$, $\hat{X}_j(\theta) = \mathbb{E}(X_{t_j} | y_{0:j}; \theta)$, $V_j(\theta) = \mathbb{E}((X_{t_j} - \hat{X}_j)^2; \theta)$. The exact likelihood of $y_{0:N}$ is then equal to

$$L(\theta, y_{0:N}) = \prod_{j=0}^{N} \frac{1}{\sqrt{2\pi (V_j^-(\theta) + \sigma^2)}} \exp\left(-\frac{1}{2} \frac{(y_j - \hat{X}_j^-(\theta))^2}{(V_j^-(\theta) + \sigma^2)} \right). \tag{1.22}$$

When the unobserved diffusion is not an Ornstein–Uhlenbeck process, Monte Carlo methods can be used similarly to the case without measurement noise.

1.8.2 Bayesian Approach

Bayesian estimation is an alternative to the MLE, which takes advantage of prior knowledge of parameter values. For example, biologists may know that the decay rate of a drug elimination is most probably close to some pre-specified value. This is

incorporated into the model by assuming a prior distribution for the parameters. The Bayesian approach consists in estimating the posterior distribution of the parameter θ given the observations and the prior distribution.

Denote $p(\theta)$ the prior distribution (θ is thus a random variable). When the Itô process is observed without measurement noise, the posterior distribution given the observations $x_{0:N}$ is

$$p(\theta|x_{0:N}) = \frac{p(\theta, x_{0:N})}{p(x_{0:N})} = \frac{p(x_{0:N}|\theta)p(\theta)}{p(x_{0:N})},$$

where $p(x_{0:N}|\theta)$ is the likelihood function, and $p(x_{0:N}) = \int p(\theta, x_{0:N})d\theta$ is the marginal distribution of the data $x_{0:N}$. In general, the posterior distribution has no closed form because $p(x_{0:N})$ is not explicit. Classical Bayesian estimators propose to approximate the posterior distribution via simulation of samples $(\theta^m)_{1\leq m\leq M}$ using Markov Chain Monte Carlo (MCMC) techniques. The aim is to simulate a Markov Chain with the target distribution $p(\theta|x_{0:N})$ as stationary distribution. Usual MCMC techniques are the Metropolis–Hastings and the Gibbs algorithms [28]. The Metropolis–Hastings algorithm, an accept–reject algorithm, requires an explicit expression of $p(x_{0:N}|\theta)$ for the computation of the acceptance probability. This is rarely the case for Itô processes and approaches similar to the MLE framework can be used: $p(\theta, x_{0:N})$ can be approximated via the Euler–Maruyama scheme by a Gaussian density [14], and Brownian bridges can be used to reduce the variance of the MCMC integration [14, 29].

When the diffusion is observed with measurement noise (1.21), the posterior distribution given the observations $y_{0:N}$ is

$$\begin{aligned} p(\theta|y_{0:N}) &= \int p(\theta, X_{t_0}, \ldots, X_{t_N}|y_{0:N})dX_{t_0}\ldots dX_{t_N} \\ &= \int \frac{p(y_{0:N}|\theta, X_{t_0}, \ldots, X_{t_N})p(\theta, X_{t_0}, \ldots, X_{t_N})}{p(y_{0:N})}dX_{t_0}\ldots dX_{t_N}. \end{aligned}$$

Simulations of $(\theta, X_{t_0}, \ldots, X_{t_N})$ under $p(\theta, X_{t_0}, \ldots, X_{t_N}|y_{0:N})$ provide samples of θ under the posterior distribution. Similarly to the case without measurement noise, the MCMC approach combined with Gaussian approximations are used to simulate samples under this target distribution.

1.8.3 Martingale Estimating Functions

The score function (1.16) can be approximated by means of martingales of a similar form. Suppose we have a collection of real valued functions $h_j(x, y, ; \theta)$, $j = 1, \ldots, N$ satisfying

$$\int h_j(x, y; \theta)p(y \mid x; \theta)dy = 0 \tag{1.23}$$

for all x and θ. Consider estimating functions of the form

$$G_n(\theta) = \sum_{i=1}^{n} a(X_{i-1}, \theta) h(X_{i-1}, X_i; \theta), \tag{1.24}$$

where $h = (h_1, \ldots, h_N)^T$, and the $p \times N$ weight matrix $a(x, \theta)$ is a function of x such that (1.24) is $\mathbb{P}_\theta$-integrable. It follows from (1.23) that $G_n(\theta)$ is a martingale under $\mathbb{P}_\theta$ for all θ. An estimating function with this property is called a *martingale estimating function*. The matrix a determines how much weight is given to each of the h_j's in the estimation procedure. This weight matrix can be chosen in an optimal way using the theory of optimal estimating functions. We will not treat this here, see [4,5,31,32] for details.

Example 1.4. The martingale estimating function (1.18) is of the type (1.24) with $N = 2$, $h_1(x, y; \theta) = y - F(\Delta, x; \theta)$ and $h_2(x, y; \theta) = (y - F(\Delta, x; \theta))^2 - \phi(\Delta, x, \theta)$. The weight matrix is

$$\left(\frac{\partial_\theta F(\Delta, x; \theta)}{\phi(\Delta, x; \theta)}, \frac{\partial_\theta \phi(\Delta, x; \theta)}{2\phi^2(\Delta, x; \theta)\Delta} \right).$$

Example 1.5. A generally applicable quadratic martingale estimating function for model (1.14) is

$$\begin{aligned} G_n(\theta) = \sum_{i=1}^{n} \Bigg\{ & \frac{\partial_\theta \mu(X_{i-1}; \theta)}{\sigma^2(X_{i-1}; \theta)} [X_i - F(X_{i-1}; \theta)] \\ & + \frac{\partial_\theta \sigma^2(X_{i-1}; \theta)}{2\sigma^4(X_{i-1}; \theta)\Delta} \left[(X_i - F(X_{i-1}; \theta))^2 - \phi(X_{i-1}; \theta) \right] \Bigg\}. \end{aligned} \tag{1.25}$$

For the square-root process (1.8) the quadratic martingale estimating function (1.25) is

$$G_n(\theta) = \begin{pmatrix} \displaystyle\sum_{i=1}^{n} \frac{1}{X_{i-1}} \left[X_i - X_{i-1} e^{-\Delta/\tau} - \alpha(1 - e^{-\Delta/\tau}) \right] \\ \displaystyle\sum_{i=1}^{n} \left[X_i - X_{i-1} e^{-\Delta/\tau} - \alpha(1 - e^{-\Delta/\tau}) \right] \\ \displaystyle\sum_{i=1}^{n} \frac{1}{X_{i-1}} \Big[\left(X_i - X_{i-1} e^{-\Delta/\tau} - \alpha(1 - e^{-\Delta/\tau}) \right)^2 \\ \qquad -\sigma^2 \tau \left\{ (\alpha/2 - X_{i-1}) e^{-2\Delta/\tau} - (\alpha - X_{i-1}) e^{-\Delta/\tau} + \alpha/2 \right\} \Big] \end{pmatrix}$$

1.9 Biological Applications

To end this chapter, we will give a few examples of the use of stochastic models in biology. Examples of applications in neuroscience can be found in Chaps. 5–8.

1.9.1 Oncology

This work has been realized by Benjamin Favetto, Adeline Samson, Daniel Balvay, Isabelle Thomassin, Valentine Genon-Catalot, Charles-André Cuenod and Yves Rozenholc.

In anti-cancer therapy, it is of importance to assess tumor aggressiveness as well as to follow and monitor the in vivo effects of treatments [17, 30]. This can be performed via dynamic contrast enhanced imaging (DCEI) by studying the tissue microvascularisation and angiogenesis. This facilitates a better treatment monitoring by optimizing *in vivo* the therapeutic strategy.

The DCEI experiment consists in injecting a contrast agent to the patient and recording a medical images sequence, which measures the evolution of the contrast agent concentration along time. The pharmacokinetics of the contrast agent is modeled by a bidimensional differential system. In this pharmacokinetic model, the contrast agent within a voxel of tissue is assumed to be either in the plasma compartment or inside the interstitial compartment. We assume that exchanges inside a voxel are (1) from the arteries (input) into the blood plasma; (2) from the blood plasma into the veins (output) and (3) between blood plasma and interstitial space. The quantities of contrast agent in a single unit voxel at time t are denoted $AIF(t)$, $Q_P(t)$ and $Q_I(t)$ for artery, plasma and interstitial compartments, respectively. The biological parameters and constraints are as follows: $F_T \geq 0$ is the tissue blood perfusion flow per unit volume of tissue (in $\mathrm{ml{\cdot}min^{-1}{\cdot}100\,ml^{-1}}$), $V_b \geq 0$ is the part of whole blood volume (in %), $V_e \geq 0$ is the part of extravascular extracellular space fractional volume (in %), and $PS \geq 0$ is the permeability surface area product per unit volume of tissue (in $\mathrm{ml{\cdot}min^{-1}{\cdot}100\,ml^{-1}}$). We have that $V_b + V_e < 100$. The hematocrit is the proportion of blood volume consisting of red blood cells and assumed to be $h = 0.4$. The delay with which the contrast agent arrives from the arteries to the plasma is denoted δ. Both t and δ are measured in seconds.

The contrast agent kinetics can be modeled by the following ODE model:

$$\begin{aligned}
\frac{dQ_P(t)}{dt} &= \frac{F_T}{1-h}\,AIF(t-\delta) - \frac{PS}{V_b(1-h)}Q_P(t) + \frac{PS}{V_e}Q_I(t) - \frac{F_T}{V_b(1-h)}Q_P(t) \\
\frac{dQ_I(t)}{dt} &= \frac{PS}{V_b(1-h)}Q_P(t) - \frac{PS}{V_e}Q_I(t).
\end{aligned} \tag{1.26}$$

We assume that no contrast agent exists inside the body before the acquisition and hence the initial conditions are $Q_P(t_0) = Q_I(t_0) = AIF(t_0) = 0$. Note that $AIF(t)$ is a given function for all t, controlled by the experimentalist.

However, this deterministic model is unable to capture the random fluctuations observed along time. For example, it fails to capture the contrast agent dosing and sampling errors, the measurement errors in the arterial input function, or the random fluctuations along time in the plasma/interstitial permeability. These variations are unpredictable. Our main hypothesis is that a more realistic model can be obtained by a stochastic approach. We introduce an SDE model by adding random components:

$$\begin{aligned} dQ_P(t) &= \left(\frac{F_T}{1-h}\,AIF(t-\delta) - \frac{PS}{V_b(1-h)}Q_P(t) + \frac{PS}{V_e}Q_I(t) - \frac{F_T}{V_b(1-h)}Q_P(t)\right)dt \\ &\quad + \sigma_1 dW_t^1 \qquad (1.27) \\ dQ_I(t) &= \left(\frac{PS}{V_b(1-h)}Q_P(t) - \frac{PS}{V_e}Q_I(t)\right)dt + \sigma_2 dW_t^2 \end{aligned}$$

where W_t^1 and W_t^2 are two independent Wiener processes, and σ_1, σ_2 are the standard deviations of the random perturbations. The initial conditions are the same as above. This Itô process is a bidimensional Ornstein Uhlenbeck process.

In our biological context, only the sum $S(t) = Q_P(t) + Q_I(t)$ can be measured. Noisy and discrete measurements $(y_i, i = 0, \ldots, N)$ of $S(t)$ are performed at times $t_0 = 0 < t_1 < \ldots < t_N = T$. The observation model is thus:

$$y_i = S(t_i) + \gamma\varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0,1)$$

where $(\varepsilon_i)_{i=0,\ldots,N}$ are assumed to be independent, and γ is the unknown standard deviation of the Gaussian noise. The model parameters are denoted $\theta^{\text{ODE}} = (F_T, V_b, PS, V_e, \delta, \gamma^2)$ and $\theta^{\text{SDE}} = (F_T, V_b, PS, V_e, \delta, \sigma_1, \sigma_2, \gamma^2)$ for the ODE and SDE models, respectively.

MLEs $\hat{\theta}$ of the model parameters are obtained by applying the standard least squares method for the ODE model and the Kalman filter approach for the SDE model. Predictions of both models are computed as the solution of the differential system (1.26) computed in $\hat{\theta}^{\text{ODE}}$ for the ODE model and as the conditional expectation of Q_P and Q_I given the whole data $(y_{0:N})$ for the SDE model.

The ODE and SDE models were applied to two signals to estimate the parameters $\hat{\theta}^{\text{ODE}}$ and $\hat{\theta}^{\text{SDE}}$, their standard deviations and the associated predictions $\hat{Q}_P$, $\hat{Q}_I$ and $\hat{S}$. The ODE and SDE residuals were computed as the difference between the observations $y_{0:N}$ and the predictions $\hat{S}$ of the corresponding model. Signal 1 results are summarized in Table 1.1 and Fig. 1.3. For this signal, the ODE and SDE estimates and the predictions of the quantity of contrast agent are identical.

For signal 2, the ODE and SDE estimates were different. The SDE predicted quantity of contrast agent in the interstitial compartment $\hat{Q}_I^{\text{SDE}}(t)$ was always null ($\hat{Q}_I^{\text{SDE}}(t) = 0 \quad \forall t$) while the ODE prediction $\hat{Q}_I^{\text{ODE}}(t)$ was not (Fig. 1.4). The ODE

Table 1.1 Estimated parameters for oncology signal 1 data, using the ODE and the SDE models

Parameters	ODE model	SDE model
F_T	48.7	48.7
V_b	40.5	40.5
PS	13.3	13.3
V_e	29.4	29.4
δ	6.0	6.0
γ	8.02	7.86
σ_1	–	$< 10^{-3}$
σ_2	–	$< 10^{-3}$

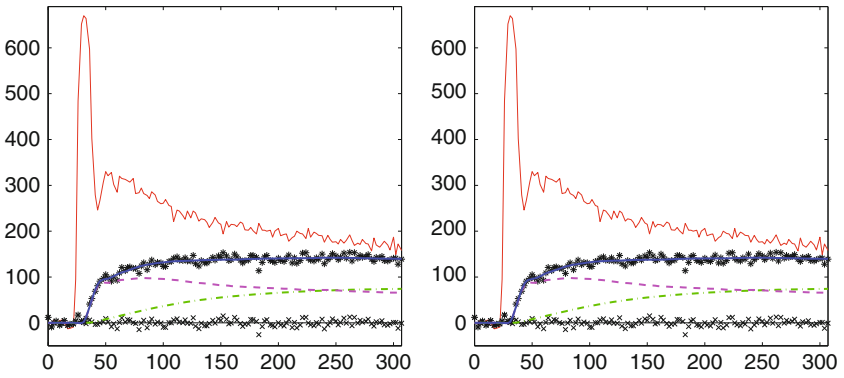

Fig. 1.3 Predictions for signal 1 data, obtained with the ODE model (*left*) and the SDE model (*right*): *black stars* (∗) are the tissue observations (y_i), the AIF observations are represented by the *red line*, *crosses* (×) are the residuals. The *plain blue*, *dashed pink* and *dash-dotted green lines* are the predictions for $S(t)$, $Q_P(t)$ and $Q_I(t)$, respectively

model detected exchanges inside the voxel between the two compartments. The ODE residuals were correlated, especially between times $t = 40$ and $t = 75$, contrary to the SDE residuals. Parameter estimates obtained by the ODE and the SDE models are different (Table 1.2). The SDE estimated blood volume ($\hat{V}_b^{\text{SDE}} = 53.5$) is larger than the ODE estimate ($\hat{V}_b^{\text{ODE}} = 41.3$). The SDE estimated permeability surface ($\widehat{PS}^{\text{SDE}} = 0.81$) is much less than the ODE estimate ($\widehat{PS}^{\text{ODE}} = 2.96$). As $\hat{V}_b^{\text{ODE}} + \hat{V}_e^{\text{ODE}} = 100$, the ODE estimation has stopped at a boundary of the optimization domain. This suggests a more careful look. We removed the 2 (and then the 5) last times of observations. While the SDE estimation remained stable when removing observations (up to changes in the last digits), the ODE estimation changed totally showing its poor stability. This variability induces even an inversion in the prediction of the quantity of the contrast agent in the two compartments. The results with the 2 or 5 last observations removed are added in Table 1.2, only for the ODE estimation. Figure 1.4 illustrates these results by zooming in on the predictions for each estimation.

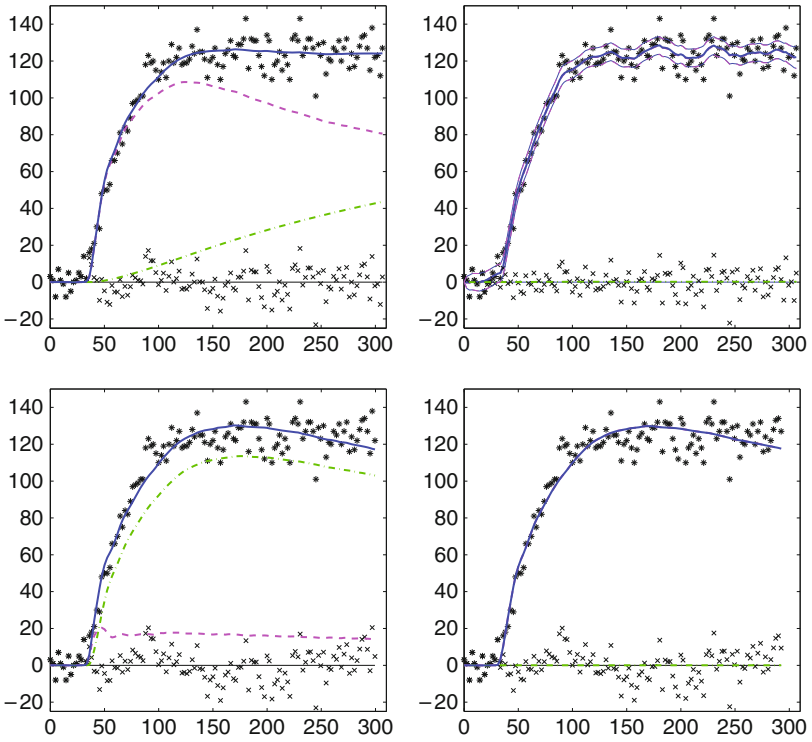

Fig. 1.4 *Top figures*: predictions for oncology signal 2 data, obtained with the ODE model (*left*) and the SDE model (*right*): *black stars* (∗) are the tissue observations (y_i), *crosses* (×) are the residuals. The *plain blue, dashed pink and dash-dotted green lines* are respectively the predictions for $S(t)$, $Q_P(t)$ and $Q_I(t)$. For the SDE model, each prediction curve is surrounded by its 95% confidence intervals. *Bottom figures*: predictions obtained with the ODE model removing the last 2 observations (*left*) and the last 5 observations (*right*)

Table 1.2 Estimated parameters for oncology signal 2 data, using the ODE model, the SDE model and using the ODE model after removing the last 2 and the last 5 observations

Parameters	ODE model	SDE model[a]	ODE without 3 last times	ODE without 5 last times
F_T	24.6	20.0	32.4	20.3
V_b	41.3	53.5	6.6	52.9
PS	2.96	0.81	43.2	0.04
V_e	58.7	0.04	27.9	0.002
δ	10.5	9.68	9.5	7.49
γ	7.55	6.51	8.4	8.19
σ_1		1.22		
σ_2		0.02		

[a]The results were exactly the same after dropping the last 2 or 5 observations

In conclusion, the use of a stochastic version of the two-compartment model avoids the instability sometimes observed with the classical two-compartment model. The SDE approach provides a more robust parameter estimation, adding reliability to the two-compartment models.

1.9.2 Agronomy

This work has been realized by Sophie Donnet, Jean-Louis Foulley and Adeline Samson [11].

Growth curve data consist of repeated measurements of a growth process over time among a population of individuals. In agronomy, growth data allow differentiating animal or vegetal phenotypes by characterizing the dynamics of the underlying biological process. In gynaecology or pediatrics, height and weight of children are regularly recorded to control their development. The parametric statistical approach used to analyze these data is a parametric growth function, such as the Gompertz, logistic, Richards or Weibull functions [35], which prescribe monotone increasing growth, whatever the parameter values. These models have proved their efficiency in animal genetics [18,20] and in pediatrics [33]. However, as pointed out by [8], the used function may not capture the exact process, as responses for some individuals may display some local fluctuations such as weight decreases or growth slow-down. These phenomena are not due to measurement errors but are induced by an underlying biological process that is still unknown today. In animal genetics, a wrong modeling of these curves could affect the genetic analysis. In fetal growth, the detection of growth slow-down is a crucial indicator of fetal development problems. Thus, we propose to model these variations in growth curves by an Itô process. The parameter estimation is based on a Bayesian approach.

We focus on the modeling of chicken growth. Data y are noisy weight measurements of chickens at weeks $t = 0, 4, 6, 8, 12, 16, 20, 24, 28, 32, 36, 40$ after birth (Fig. 1.5). These data are classically analyzed with a Gompertz function:

$$x(t) = Ae^{-Be^{-Ct}}, \tag{1.28}$$

which depends on the three parameters A, B, C and verifies the following ODE

$$x'(t) = BCe^{-Ct}x(t), \quad x(0) = Ae^{-B}. \tag{1.29}$$

A heteroscedastic error model is usually required to obtain satisfactory results. For simplicity we model the logarithm of the data y with the logarithm of the Gompertz function (1.28) and add a measurement error with constant variance:

$$\log y_j = \log A - Be^{-Ct_j} + \gamma\varepsilon_j,\ \varepsilon_j \sim_{i.i.d.} \mathcal{N}(0,1),\ \forall j = 0, \ldots, n_i. \tag{1.30}$$

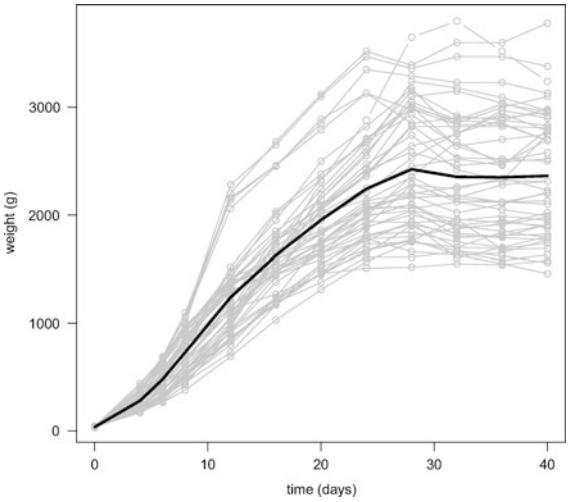

Fig. 1.5 Growth curves of the 50 chickens and mean growth curve in *dashed bold line*

The log-parametrization for A and C was used to ensure that parameters are positive. We estimate the posterior distribution of the parameters $(\log A, B, \log C, \gamma^2)$ of this ODE model.

The SDE model is deduced from the Gompertz equation (1.29):

$$dX_t = BCe^{-Ct}X_t dt + \sigma X_t dW_t, \quad X_0 = Ae^{-B}, \tag{1.31}$$

where the diffusion coefficient is set equal to σX_t given the heteroscedasticity of the process. This means that the standard error of the random perturbations of the growth rate is proportional to the weight. This Itô process belongs to the family of Geometric Brownian motions with time inhomogeneous drift. The Itô process (1.31) has an explicit solution. Indeed, set $Z_t = \log(X_t)$. By Itô's formula (1.9), the conditional distribution of Z_{t+h} given (Z_s), $s \le t, h > 0$ is:

$$Z_{t+h}|(Z_s)_{s\le t} \sim \mathcal{N}(Z_t - Be^{-Ct}(e^{-Ch} - 1) - \frac{1}{2}\sigma^2 h, \sigma^2 h), \quad Z_0 = \log(A) - B.$$

Thus, $X_t = Ae^{-Be^{-Ct}}e^{-\frac{1}{2}\sigma^2 t + \sigma W_t}$ and $X_0 = Ae^{-B}$. As a consequence, X_t is a multiplicative random perturbation of the solution of the Gompertz model. Due to the assumption of the non-negativity of A, X_t is almost surely non-negative, which is a natural constraint to model weight records.

We then discretize the SDE:

$$Z_{t_j}|Z_{t_{j-1}} \sim \mathcal{N}\left(Z_{t_{j-1}} - Be^{-Ct_{j-1}}(e^{-C(t_j - t_{j-1})} - 1) - \frac{1}{2}\sigma^2(t_j - t_{j-1}), \sigma^2(t_j - t_{j-1})\right).$$

Table 1.3 Posterior distributions for the ODE and SDE models on chicken growth data: average of estimated parameters and their 95% credibility intervals (95% CI)

	ODE		SDE	
	Average	95% CI	Average	95% CI
$\log A$	7.77	[7.70; 7.84]	7.75	[7.67; 7.83]
B	4.17	[4.11; 4.23]	4.15	[4.08; 4.22]
$\log C$	2.75	[2.70; 2.81]	2.78	[2.71; 2.84]
γ^{-2}	225.5	[197.4; 255.5]	630.2	[463.8; 797.9]
σ^2			0.09	[0.07; 0.12]

The SDE model on the logarithm of data is thus defined as:

$$\begin{aligned}
(\log y_0, \log y_1, \ldots, \log y_N)^T &= (\log(A) - B, Z_{t_1}, \ldots, Z_{t_N})^T + \gamma\varepsilon, \\
&\quad \varepsilon \sim_{i.i.d.} \mathcal{N}(0, \mathbf{I}_{N+1}) \\
(Z_{t_1}, \ldots, Z_{t_N})^T &= \left(\log(A) - Be^{-Ct_1}, \ldots, \log(A) - Be^{-Ct_N}\right)^T \\
&\quad -\sigma^2 (t_1, \ldots, t_N)^T + \eta, \\
&\quad \eta \sim_{i.i.d} \mathcal{N}\left(0_J, \sigma^2 t\right) \\
t &= (\min(t_j, t_{j'}))_{1 \le j, j' \le n_i}, \qquad (1.32)
\end{aligned}$$

where ^{T} denotes transposition. By a Bayesian approach we estimate the posterior distribution of the parameters $(\log A, B, \log C, \sigma^2, \gamma^2)$. We consider Gaussian prior distributions for $(\log A, B, \log C)$, an inverse Gamma prior distribution for σ^2 as suggested by [9] for hierarchical models, and an inverse Gamma prior distribution for γ^2. The posterior distribution is approximated with an MCMC algorithm.

Posterior expectations of the parameters are presented in Table 1.3. The estimate of σ^2 is strictly positive and its confidence interval is far away from zero. This implies that the dynamical process that most likely represents the growth is an Itô process with non-negligible noise. Diagnostic tools to validate the models are applied to both ODE and SDE models. Figure 1.6 presents the posterior predictive distributions of both models computed for each time point. Centered and symmetrical posterior predictive distributions correspond to a "good" model. There is a clear improvement in the posterior predictive distributions from the ODE to the SDE model for the whole population, both at early and late ages. The predictive abilities of the two models can be compared on the posterior expectation of the squared errors using cross-validation techniques. New data sets denoted y_{-j} are constructed by dropping the jth measurement. The error is then:

$$r_j^k = \mathbb{E}\left[(\log(y_j^{rep,k}) - \log(y_j))^2 | y_{-j}\right], \quad k = 1, 2$$

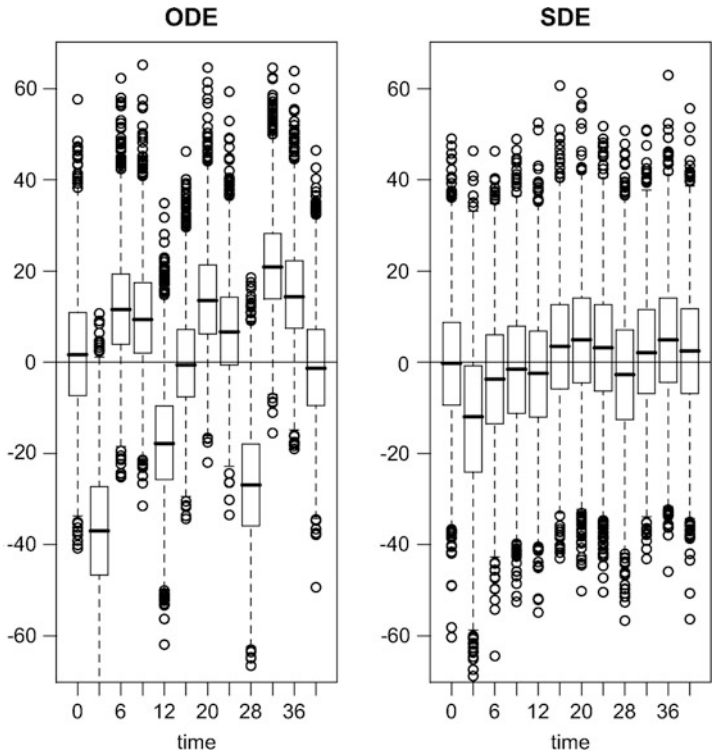

Fig. 1.6 Posterior predictive distributions for the ODE and SDE models on chicken growth data

with $y_j^{rep,k}$ drawn from the predictive distribution $p(y_j^{rep,k}|y_{-j})$. Averaging in r_j^k is with respect to the posterior uncertainty in the parameters of the model. We performed that comparison for the last observation $j = 12$, which is especially critical with respect to the growth pattern studied here. These quantities are $r_{12}^{ode} = 0.56$ and $r_{12}^{sde} = 0.48$, resulting in a reduction of the squared errors of prediction of 14% when using SDE vs ODE. Figure 1.7 reports, for four subjects, the observed weights, the ODE prediction, the empirical mean of the last 1,000 simulated trajectories of the SDE (1.32) generated during the estimation algorithm, their empirical 95% confidence limits (from the 2.5th percentile to the 97.5th percentile) and one simulated trajectory. Subjects 4 and 13 are examples of subjects with no growth slow-down. Both ODE and SDE models satisfactorily fit the observations. Subject 14 has a small observed weight decrease. For subject 1, the weight decrease is more important. For both subjects, the ODE model fails to capture this phenomenon while the SDE model does.

In conclusion, on the presented data set, the introduction of this SDE model leads to a clear validation of the model (Fig. 1.6) which was not the case in the standard model, justifying the introduction of the new stochastic component.

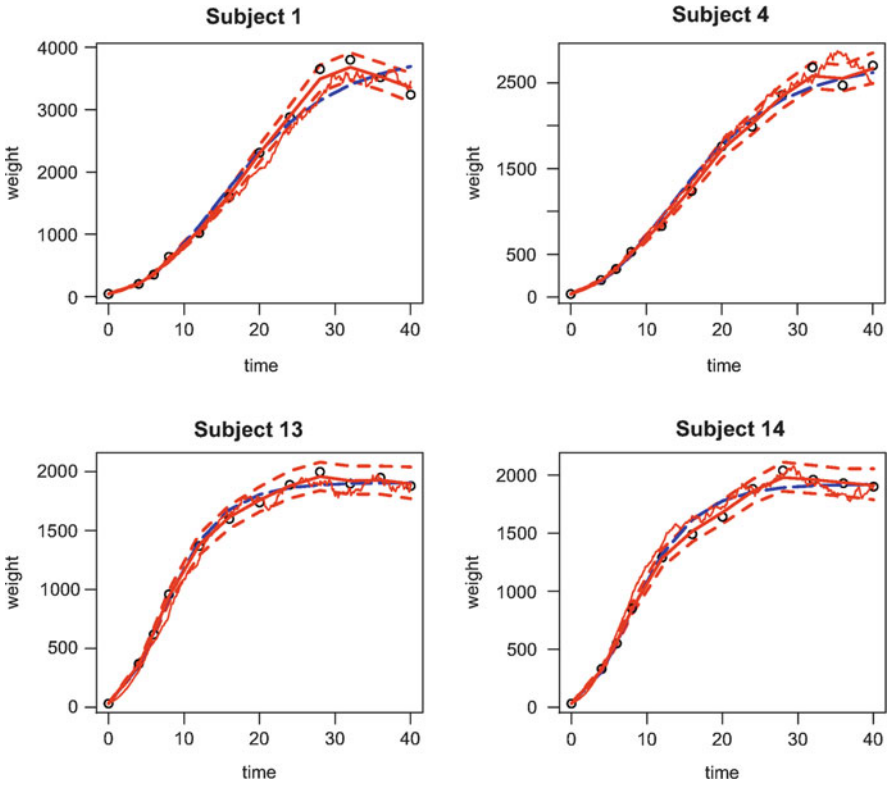

Fig. 1.7 Observations (*circles*), predictions obtained with the ODE model (*long dashed line*), mean SDE prediction (*smooth solid line*), 95% credibility interval obtained with the SDE model (*dotted line*) and one SDE realization (*solid line*), for subjects 1, 4, 13 and 14

References

1. Aït-Sahalia, Y.: Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. Econometrica **70**(1), 223–262 (2002)
2. Bally, V., Talay, D.: The law of the Euler scheme for stochastic differential equations (I): convergence rate of the distribution function. Probab. Theor. Relat. Field **104**(1), 43–60 (1996)
3. Bally, V., Talay, D.: The law of the Euler scheme for stochastic differential equations (II): convergence rate of the density. Monte Carlo Meth. Appl. **2**, 93–128 (1996)
4. Bibby, B.M., Sørensen, M.: Martingale estimation functions for discretely observed diffusion processes. Bernoulli **1**(1/2), 017–039 (1995)
5. Bibby, B.M., Sørensen, M.: On estimation for discretely observed diffusions: a review. Theor. Stoch. Process. **2(18)**, 49–56 (1996)
6. Cox, J.C., Ingersoll, J.E., Ross, S.A.: A theory of the term structure of interest rates. Econometrica **53**, 385–407 (1985)
7. Dacunha-Castelle, D., Florens-Zmirou, D.: Estimation of the coefficients of a diffusion from discrete observations. Stochastics **19**(4), 263–284 (1986)

8. Davidian, M., Giltinan, D.M.: Nonlinear models for repeated measurements: An overview and update. J. Agr. Biol. Environ. Stat. **8**, 387–419 (2003)
9. De la Cruz-Mesia, R., Marshall, G.: Non-linear random effects models with continuous time autoregressive errors: a Bayesian approach. Stat. Med. **25**, 1471–1484 (2006)
10. Donnet, S., Samson, A.: Parametric inference for mixed models defined by stochastic differential equations. ESAIM Probab. Stat. **12**, 196–218 (2008)
11. Donnet, S., Foulley, J.L., Samson, A.: Bayesian analysis of growth curves using mixed models defined by stochastic differential equations. Biometrics **66**(3), 733–741 (2010)
12. Durham, G.B., Gallant, A.R.: Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. J. Bus. Econ. Stat. **20**, 297–338 (2002)
13. Elerian, O., Chib, S., Shephard, N.: Likelihood inference for discretely observed nonlinear diffusions. Econometrica **69**(4), 959–993 (2001)
14. Eraker, B.: MCMC analysis of diffusion models with application to finance. J. Bus. Econ. Stat. **19**(2), 177–191 (2001)
15. Favetto, B., Samson, A.: Parameter estimation for a bidimensional partially observed Ornstein-Uhlenbeck process with biological application. Scand. J. Stat. **37**, 200–220 (2010)
16. Feller, W.: Diffusion processes in genetics. In: Neyman, J. (ed.) Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, pp. 227–246. University of California Press, Berkeley (1951)
17. Fournier, L., Thiam, R., Cuénod, C.-A., Medioni, J., Trinquart, L., Balvay, D., Banu, E., Balcaceres, J., Frija, G., Oudard, S.: Dynamic contrast-enhanced CT (DCE-CT) as an early biomarker of response in metastatic renal cell carcinoma (mRCC) under anti-angiogenic treatment. J. Clin. Oncol. ASCO Annu. Meet. Proc. (Post-Meeting Edition) **25** (2007)
18. Hou, W., Garvan, C.W., Zhao, W., Behnke, M., Eyler, F., Wu, R.: A general model for detecting genetic determinants underlying longitudinal traits with unequally spaced measurements and nonstationary covariance structure. Biostatistics **6**, 420–433 (2005)
19. Iacus, S.M.: Simulation and Inference for Stochastic Differential Equations. With R examples. Springer, New York (2008)
20. Jaffrézic, F., Meza, C., Lavielle, M., Foulley, J.L.: Genetic analysis of growth curves using the SAEM algorithm. Genet. Sel. Evol. **38**, 583–600 (2006)
21. Karlin, S., Taylor, H.M.: A Second Course in Stochastic Processes. Academic, New York (1981)
22. Kloeden, P., Platen, E.: Numerical Solution of Stochastic Differential Equations. Springer, New York (1999)
23. Kutoyants, T.: Parameter Estimation for Stochastic Processes. Helderman Verlag, Berlin (1984)
24. Øksendal, B.: Stochastic Differential Equations. An Introduction with Applications, 6th edn. Universitext. Springer, Berlin (2003)
25. Pedersen, A.: A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. Scand. J. Stat. **22**(1), 55–71 (1995)
26. Pedersen, A.R.: Statistical analysis of gaussian diffusion processes based on incomplete discrete observations. Research Report, Department of Theoretical Statistics, University of Aarhus, 297 (1994)
27. Prakasa Rao, B.: Statistical Inference for Diffusion Type Processes. Arnold, London (1999)
28. Robert, C.P.: Bayesian computational methods. In: Handbook of Computational Statistics, pp. 719–765. Springer, Berlin (2004)
29. Roberts, G.O., Stramer, O.: On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. Biometrika **88**(3), 603–621 (2001)
30. Rosen, M.A., Schnall, M.D.: Dynamic contrast-enhanced magnetic resonance imaging for assessing tumor vascularity and vascular effects of targeted therapies in renal cell carcinoma. Clin. Cancer Res. **13**(2), 770–6 (2007)
31. Sørensen, M.: Parametric inference for discretely sampled stochastic differential equations. In: Andersen, T.G., Davis, R.A., Kreiss, J.P., Mikosch, T. (eds.) Handbook of Financial Time Series, pp. 531–553. Springer, Heidelberg (2009)

32. Sørensen, M.: Estimating functions for diffusion-type processes. In: Kessler, M., Lindner, A., Sørensen, M. (eds.) Statistical Methods for Stochastic Differential Equations. Chapmann & Hall/CRC Monographs on Statistics & Applied Probability, London (2012)
33. Spyrides, M.H., Struchiner, C.J., Barbosa, M.T., Kac, G.: Effect of predominant breastfeeding duration on infant growth: a prospective study using nonlinear mixed effect models. J. Pediatr. **84**, 237–243 (2008)
34. Taylor, H.M., Karlin, S.: An Introduction to Stochastic Modeling, 3rd edn. Academic, San Diego, CA (1998)
35. Zimmerman, D., Núnez-Antón, V.: Parametric modelling of growth curve data: an overview. Test **10**, 1–73 (2001)

http://www.springer.com/978-3-642-32156-6

Stochastic Biomathematical Models
with Applications to Neuronal Modeling
Bachar, M.; Batzel, J.J.; Ditlevsen, S. (Eds.)
2013, XVI, 206 p. 34 illus., 13 illus. in color., Softcover
ISBN: 978-3-642-32156-6