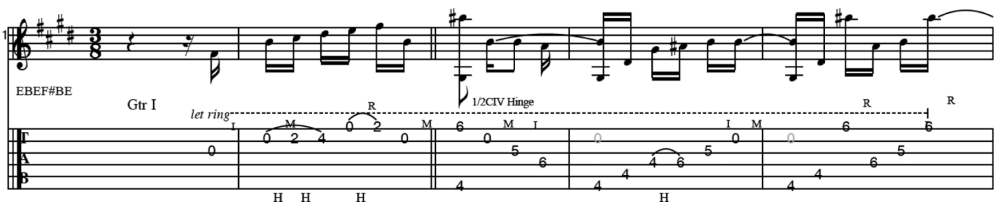
3

Introduction to Stochastic Calculus



Ordering random elements.
EBEF♯BE, Causeway, Alex de Grassi

In this chapter, an introduction to the theory of stochastic calculus is given. Although there are many books on this topic (Van Kampen, 1981; Kloeden et al., 1994; Oksendal, 1995; Mikosch, 2000), many of these are quite mathematically oriented and hence less accessible for climate scientists. Here an attempt is made to present this material from a 'user' point of view (Higham, 2001), while being as rigorous as possible.

3.1 Random variables

The starting point is the concept of a random variable. We are all familiar with examples where random variables play a role, such as throwing a dice and tossing a coin. In this example, we have the only outcomes 'head' or 'tail'. If we attribute 0 to the outcome $\omega =$ 'head' and 1 to the outcome $\omega =$ 'tail', then we can define a random variable $X : \Omega \to \{0, 1\}$, where Ω is the outcome space. In general, a random variable $X(\omega)$ is a real-valued function defined on Ω.

To tackle questions related to the mean and spread of values of the random variable X, we first have to define an algebraic structure on the outcome space Ω. More specifically, a σ-algebra $\mathcal{F}$ is defined on Ω, which at least contains the empty set

$\emptyset$ and its complement Ω. Moreover, if $A \in \mathcal{F}$, then so is its complement A^c, and if $A, B \in \mathcal{F}$, then so are $A \cap B$, $A \cup B$, $A^c \cap B$, $A \cap B^c$ $A^c \cup B$ and $A \cup B^c$. The intuitive meaning of $\mathcal{F}$ is that if one applies the operations $\cap$, $\cup$ and ^{c} to the elements of $\mathcal{F}$ – the 'events' – the result is still an element of $\mathcal{F}$.

On a σ-algebra $\mathcal{F}$, a probability measure $P : \mathcal{F} \to [0, 1]$ is defined as (for $A, B \in \mathcal{F}$)

$$P(\emptyset) = 0\,;\; P(\Omega) = 1\,;\; P(A^c) = 1 - P(A), \tag{3.1a}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{3.1b}$$

For example, if a coin is 'fair', we assign a probability $1/2$ to both events 'head' and 'tail' and therefore $P(\{\omega : X(\omega) = 0\}) = P(\{\omega : X(\omega) = 1\}) = 1/2$.

The distribution function $F_X(x)$ of the random variable X provides a measure of the collection of the probabilities $P(X \leq x) = P(\{\omega : X(\omega) \leq x\})$ for certain $x \in \mathbb{R}$. A distribution function is either continuous or discrete. In the discrete case, with a discrete random variable X, the distribution function is given by

$$F_X(x) = \sum_{k:x_k \leq x} p_k, \tag{3.2}$$

where $p_k = P(X = x_k)$, $0 \leq p_k \leq 1$ for all k and $\sum_k p_k = 1$. Important discrete distributions are the binomial distribution $Bin(n, p)$ for $n \in \mathbb{N}$ and $p \in (0, 1)$ with

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \tag{3.3}$$

for $k \in \mathbb{N}$ and the Poisson distribution $Poi(\lambda)$ containing a parameter $\lambda > 0$ with

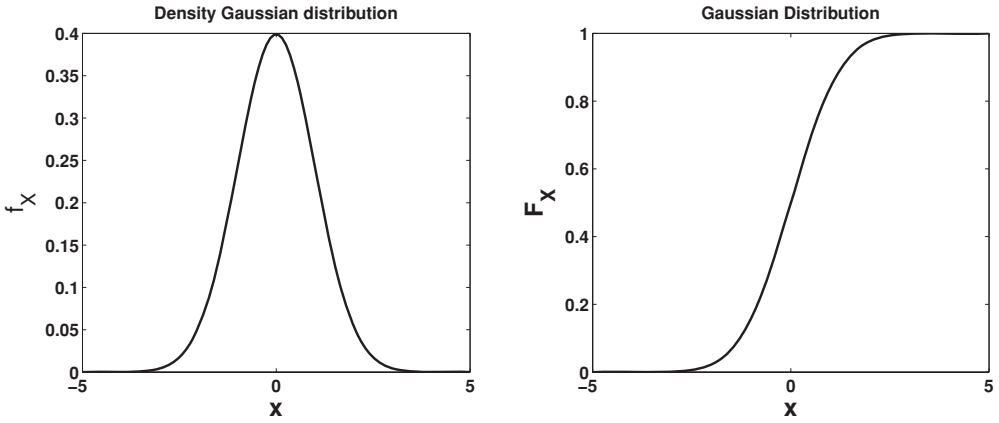
$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}. \tag{3.4}$$

The Poisson distribution has a large applicability. The number of cars that pass a certain point over a given time period, the number of stars in a given volume of space, the spelling mistakes on a page of text all satisfy a Poisson distribution. Note that the Poisson distribution is the limit $n \to \infty$ of the binomial distribution with $p = \lambda/n$ because

$$\lim_{n\to\infty} \binom{n}{k} p^k (1 - p)^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!}. \tag{3.5}$$

In contrast, a continuous random variable has a continuous distribution function F_X, which, in most cases, has a probability density function $f_X \geq 0$ such that

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\, dy;\; \int_{-\infty}^{\infty} f_X(y)\, dy = 1. \tag{3.6}$$



(a) (b)

Figure 3.1 (a) Plot of the Gaussian probability density function (3.8) for $\sigma = 1$ and $\mu = 0$. (b) Plot of the Gaussian distribution function (3.6) for $\sigma = 1$ and $\mu = 0$.

An important example is the uniform distribution $U(a, b)$ on the interval $[a, b] \in \mathbb{R}$ with a probability density function

$$f_X(x) = \frac{1}{b-a};\ x \in (a, b) \tag{3.7}$$

and $f_X(x) = 0$ elsewhere. Another important example is the Gaussian (or normal) distribution $N(\mu, \sigma^2)$ with a density

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{3.8}$$

with $\mu \in \mathbb{R}$ referred to as the mean and $\sigma \in \mathbb{R}$ as the standard deviation. As a special case, the distribution $N(0, 1)$ is called the standard normal distribution, and plots of f_X and F_X of this distribution are given in Fig. 3.1. One can easily show that when X has a distribution $N(\mu, \sigma^2)$, it follows that $Y = (X - \mu)/\sigma$ has a distribution $N(0, 1)$. About 68% of values drawn from a Gaussian distribution are within one standard deviation σ away from the mean, about 95% of the values lie within two standard deviations and about 99.7% are within three standard deviations.

To determine the statistical properties of continuous distributions, the first two moments (mean and variance) are defined as

$$\mu_X = E[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx, \tag{3.9a}$$

$$\sigma_X^2 = Var[X] = E[(X - \mu_X)^2] = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)\, dx. \tag{3.9b}$$

The quantity $E[X]$ is also called the expectation value. A relation exists between σ_X^2 and $E[X^2]$ through

$$\sigma_X^2 = E[(X-\mu_X)^2] = E[X^2] - 2\mu_X E[X] + (E[X])^2 = E[X^2] - \mu_X^2. \tag{3.10}$$

For the uniform distribution, it is easily calculated that

$$\mu_X = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2}, \tag{3.11a}$$

$$\sigma_X^2 = \int_a^b \frac{(x-\mu_X)^2}{b-a} dx = \frac{(b-a)^2}{12}, \tag{3.11b}$$

and for the Gaussian distribution (3.8), we find $\mu_X = \mu$ and $\sigma_X^2 = \sigma^2$.

3.2 Stochastic processes

In the previous section, we considered only a random variable X. This is easily generalised to a random vector $\mathbf{X} = (X_1, \ldots, X_n)$, where each X_i is a random variable. The distribution function $F_{\mathbf{X}}(\mathbf{x})$ is generalized as

$$F_{\mathbf{X}}(\mathbf{x}) = P(X_1 \le x_1, \ldots, X_n \le x_n) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} f_{\mathbf{X}}(y_1, \ldots, y_n) dy_1 \ldots dy_n, \tag{3.12}$$

where again $f_{\mathbf{X}}$ is the corresponding probability density function. As an example, the multidimensional Gaussian distribution has a probability density function

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}(\det \Sigma)^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu_{\mathbf{X}})\Sigma^{-1}(\mathbf{x}-\mu_{\mathbf{X}})^T}, \tag{3.13}$$

where Σ is the covariance matrix with elements

$$\Sigma_{i,j} = \mathrm{Cov}[X_i, X_j] = E[(X_i - \mu_{X_i})(X_j - \mu_{X_j})]. \tag{3.14}$$

Two random variables X_1 and X_2 are independent if and only if

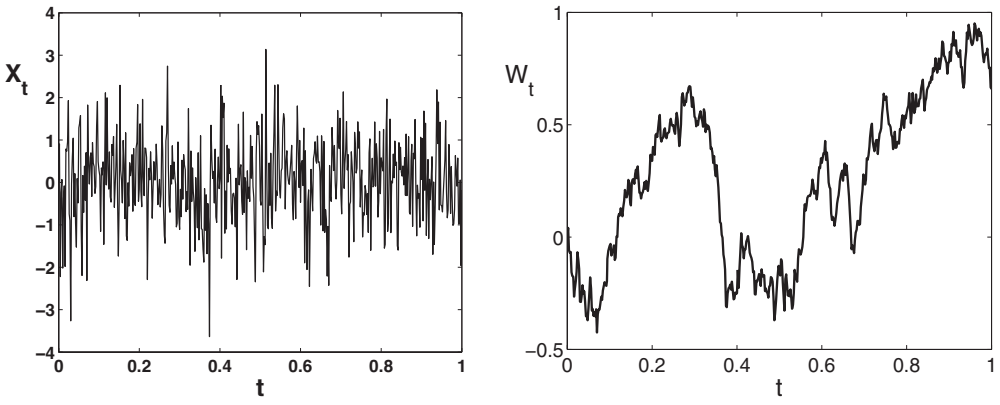
$$F_{X_1,X_2}(x_1, x_2) = F_{X_1}(x_1) F_{X_2}(x_2). \tag{3.15}$$

When two random variables X_i and X_j are independent, then $\mathrm{Cov}[X_i, X_j] = E[(X_i - \mu_{X_i})(X_j - \mu_{X_j})] = E[(X_i - \mu_{X_i})]E[(X_j - \mu_{X_j})] = 0$. If all components in a random vector are independent, then Σ is a diagonal matrix.

We are now ready to define a stochastic process X_t as a time series of random variables

$$(X_t, t \in T) = (X_t(\omega), t \in T, \omega \in \Omega), \tag{3.16}$$

where T denotes the time interval and Ω the outcome space. When t is fixed, then $X_t(\omega)$ is just a random variable. When ω is fixed, then $X_t(\omega)$ is a function of time,



(a) (b)

Figure 3.2 (a) A realization of a Gaussian process. (b) A realization of a Wiener process.

which is called a trajectory, a realization or a sample path. The expectation function of X_t is defined by $\mu_X(t) = E[X_t]$, and the covariance function $c_X(t, s)$ is defined as

$$c_X(t, s) = \text{Cov}(X_t, X_s) = E[(X_t - \mu_X(t))(X_s - \mu_X(s))]. \tag{3.17}$$

In particular, the variance function is given by $\sigma_X^2(t) = c_X(t, t)$.

A Gaussian process (Fig. 3.2a) is defined over the interval $T = [0, 1]$ with $0 \leq t_1 \leq \ldots \leq t_n \leq 1$ such that all $X_{t_1}, \ldots, X_{t_n}$ are independent and standard normally distributed, that is, each X_{t_i} has a distribution function

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

The multidimensional distribution function is then given by

$$F_{\mathbf{X}}(\mathbf{x}) = \Phi(x_1) \ldots \Phi(x_n), \tag{3.18}$$

and the expectation function and (co-)variance functions are given by

$$\mu_X(t) = 0 \; ; \; t \neq s : c_X(t, s) = 0 \; ; \; \sigma_X^2(t) = 1. \tag{3.19}$$

A stochastic process $(X_t, t \in T)$ is strictly stationary if for all choices of $t_1, \ldots, t_n$ and h such that $t_i + h \in T$ for all i, the finite dimensional distributions satisfy

$$(X_{t_1}, \ldots, X_{t_n}) \stackrel{d}{=} (X_{t_1+h}, \ldots, X_{t_n+h}), \tag{3.20}$$

where $\stackrel{d}{=}$ indicates equality in distribution sense. A stochastic process $(X_t, t \in T)$ has stationary increments if

$$X_t - X_s \stackrel{d}{=} X_{t+h} - X_{s+h}, \tag{3.21}$$

for each t, s. A stochastic process $(X_t, t \in T)$ has independent increments if the random variables $X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent for every $t_1 < \ldots < t_n$.

The Wiener process $W_t, t \in [0, \infty)$ is a stochastic process (Fig. 3.2b) with the following properties:

(i) $W_0 = 0$,
(ii) W_t has stationary, independent increments,
(iii) $\forall t > 0 : W_t$ has a $N(0, t)$ distribution, and
(iv) W_t has continuous sample paths.

From these properties, it follows immediately that with $0 \leq s < t \leq T$

$$W_t - W_s \stackrel{d}{=} N(0, t - s), \tag{3.22a}$$

$$\mu_W(t) = 0, \tag{3.22b}$$

$$c_W(t, s) = s. \tag{3.22c}$$

To obtain the result (3.22a), we note that from property (i) and (ii), it follows that $W_t - W_s \stackrel{d}{=} W_{t-s} - W_0 \stackrel{d}{=} W_{t-s}$, and (3.22a) follows then directly from property (iii). Equation (3.22b) follows directly from property (iii), and (3.22c) follows from $c_W(t, s) = E[W_t W_s] = E[(W_t - W_s + W_s)W_s] = E[(W_t - W_s)(W_s - W_0)] + E[W_s^2] = E[W_s^2] = s$, where the last two equalities follow from properties (ii) and (iii).

3.3 Stochastic calculus

In Subsections 3.1 and 3.2, the basic material on stochastic processes was introduced. In this subsection, we introduce basic material of stochastic calculus, in particular the stochastic integral and the Itô lemma.

3.3.1 The stochastic integral

Consider a smooth function $h : [0, T] \to \mathbb{R}$ for which the derivative h' is bounded on $[0, T]$. To define the Riemann integral of h, the interval $[0, T]$ is partitioned into subintervals $0 = t_0 < t_1 < \ldots < t_{N-1} < t_N = T$. The Riemann integral of h is then

given by

$$\int_0^T h(t)dt = \lim_{N\to\infty} \sum_{j=0}^{N-1} h(t_j)(t_{j+1} - t_j). \tag{3.23}$$

The stochastic integral can be defined in a similar way, and two forms exist, the Itô form and the Stratonovich form. The Itô integral of h is

$$\int_0^T h(t)dW_t = \lim_{N\to\infty} \sum_{j=0}^{N-1} h(t_j)(W(t_{j+1}) - W(t_j)), \tag{3.24}$$

where $W(t_j)$ indicates the value of the Wiener process W_t at $t = t_j$. The Stratonovich integral of h is

$$\int_0^T h(t) \circ dW_t = \lim_{N\to\infty} \sum_{j=0}^{N-1} h\left(\frac{t_j + t_{j+1}}{2}\right)(W(t_{j+1}) - W(t_j)). \tag{3.25}$$

Note that the difference between the two forms of the stochastic integral is the time values of h considered with respect to the Wiener process values W. In the Itô integral, only h values at the left endpoint are considered, just as in the Riemann integral. In the Stratonovich integral, values of h at the midpoint of the interval are considered. The Itô and Stratonovich integral in general lead to different outcomes, but a relation exists between these results, and hence both definitions have their use.

Example 3.1 Itô integral of a Wiener process Consider $h(t) = W(t)$, where the function $W(t)$ indicates a Wiener process W_t. For this case, the Itô integral can be evaluated analytically because

$$\begin{aligned}
&\sum_{j=0}^{N-1} W(t_j)(W(t_{j+1}) - W(t_j)) \\
&\quad = \sum_{j=0}^{N-1} \frac{1}{2}\left[W^2(t_{j+1}) - W^2(t_j) - (W(t_{j+1}) - W(t_j))^2\right] \\
&\quad = \frac{1}{2}(W_T^2 - W_0^2) - \frac{1}{2}\sum_{j=0}^{N-1}(W(t_{j+1}) - W(t_j))^2.
\end{aligned}$$

With $dW_j = W(t_{j+1}) - W(t_j)$ and (from (3.22)) $E[(dW_j)^2] = t_{j+1} - t_j = dt$ we obtain

$$E\left[\sum_{j=0}^{N-1}(W(t_{j+1}) - W(t_j))^2\right] = \sum_{j=0}^{N-1} dt = T,$$

and, finally, we find

$$\int_0^T W(t)dW_t = \frac{1}{2}(W_T^2 - T), \tag{3.27}$$

where the equality is interpreted in the mean-square sense. ■

3.3.2 The Itô lemmas

We proceed with the Itô integral and consider the stochastic version of the main theorem of integral calculus,

$$f(b) - f(a) = \int_a^b f'(t)dt, \tag{3.28}$$

for a smooth function f on the interval $[a, b]$. To proceed, we use the notation $dW_t = W_{t+dt} - W_t$ and consider the Taylor-series expansion

$$f(W_x + dW_x) - f(W_x) = f'(W_x)dW_x + \frac{1}{2}f''(W_x)(dW_x)^2 + \cdots. \tag{3.29}$$

With $E[(dW_x)^2] = dx$, we obtain the first Itô lemma by integration of (3.29) over the interval $[s, t]$, that is,

$$f(W_t) - f(W_s) = \int_s^t f'(W_x)dW_x + \int_s^t \frac{1}{2}f''(W_x)dx, \tag{3.30}$$

again with equality in the mean-square sense.

We see that (3.30) is the generalization of (3.28) to the stochastic case. In addition to the first term on the right-hand side, there is now an additional Riemann integral involving the second derivative of f.

The first Itô lemma (3.30) is a powerful tool to compute stochastic integrals explicitly. Consider, for example, $f(t) = t^2$, with $f'(t) = 2t$ and $f''(t) = 2$; with (3.30) we then find

$$W_t^2 - W_s^2 = 2\int_s^t W_x dW_x + \int_s^t dx. \tag{3.31}$$

With $s = 0$ and $t = T$, this gives

$$W_T^2 - W_0^2 = 2\int_0^T W_x dW_x + \int_0^T dx \Rightarrow \int_0^T W_x dW_x = \frac{1}{2}(W_T^2 - T), \tag{3.32}$$

which is the same as (3.27).

There are two extensions of the first Itô lemma. Consider a stochastic process $f(t, W_t)$ for which the function $f(t, y)$ is smooth. Again by Taylor series expansion,

we find

$$f(x+dx, W_{x+dx}) - f(x, W_x) = f_1(x, W_x)dx + f_2(x, W_x)dW_x$$
$$+\frac{1}{2}\left[f_{11}(x, W_x)(dx)^2 + 2f_{12}(x, W_x)dxdW_x + f_{22}(x, W_x)(dW_x)^2\right] + \cdots,$$

where $f_1 = \partial f/\partial t$, $f_2 = \partial f/\partial y$, $f_{11} = \partial^2 f/\partial t^2$, and so forth. We use again that $E[(dW_x)^2] = dx$, neglect higher-order terms dx^2 and $dx\, dW_x$ and integrate over the interval $[s, t]$ to obtain the second Itô lemma,

$$f(t, W_t) - f(s, W_s) = \int_s^t \left[f_1(x, W_x) + \frac{1}{2}f_{22}(x, W_x)\right]dx + \int_s^t f_2(x, W_x)dW_x. \tag{3.33}$$

If $f = f(W_t)$, then $f_1 = 0$, $f_2 = f'$, $f_{22} = f''$, and the second Itô lemma (3.33) reduces to the first Itô lemma (3.30).

Example 3.2 The Itô exponential An application of the Itô lemmas is the determination of the Itô exponential. In the deterministic case, we know that $f(t) = e^t$ has the special property that

$$f(t) - f(s) = \int_s^t f(x)dx.$$

The Itô exponential is a process X_t with the property

$$X_t - X_s = \int_s^t X_x\, dx.$$

To determine this process, we first try $f(t) = e^t$; using the first Itô lemma (3.30) then gives

$$e^{W_t} - e^{W_s} = \int_s^t e^{W_x}dW_x + \frac{1}{2}\int_s^t e^{W_x}dx,$$

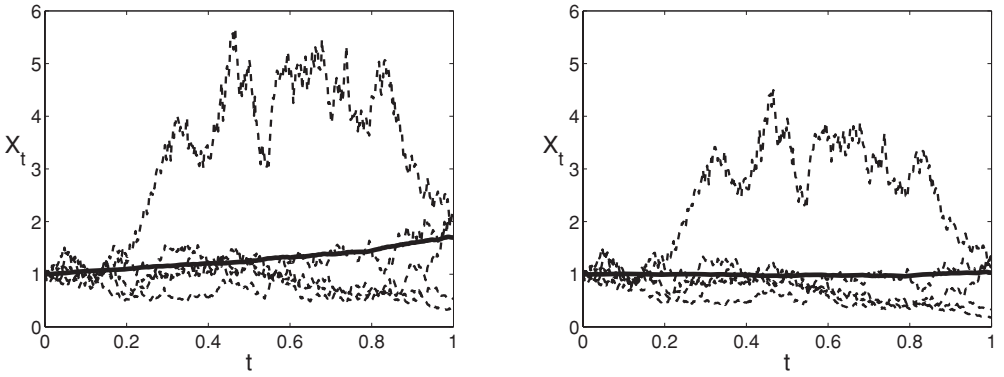
and as the latter integral is always positive, the process $X_t = e^{W_t}$ is not the Itô exponential. The correct Itô exponential follows from the use of the second Itô lemma (3.33) with $f(t, y) = e^{y-t/2}$, such that with $f_1 = -\frac{1}{2}e^{(y-t/2)}$, $f_2 = f_{22} = f$, we get

$$e^{W_t - \frac{t}{2}} - e^{W_s - \frac{s}{2}} = \int_s^t e^{W_x - \frac{x}{2}}dW_x,$$

and hence $X_t = e^{W_t - t/2}$ is the desired Itô exponential. For illustration, both stochastic processes are plotted in Fig. 3.3. ■

A third extension of the main theorem of integral calculus is for stochastic processes of the form $f(t, X_t)$, where X_t is given by

$$X_t = X_0 + \int_0^t A^{(1)}(s, X_s)ds + \int_0^t A^{(2)}(s, X_s)dW_s. \tag{3.34}$$



(a) (b)

Figure 3.3 (a) The process $X_t = e^{W_t}$. (b) The process $X_t = e^{W_t - t/2}$. In each case, five sample paths are plotted together with the mean over 1,000 sample paths.

Here the $A^{(i)}$ are smooth functions of s and X_s. Using the same procedure as for the first two Itô lemmas (Taylor series, neglect higher-order terms and $E[(dW_x)^2] = dx$), leads to the third Itô lemma,

$$f(t, X_t) - f(s, X_s) = \int_s^t \left[f_1(x, X_x) + A_x^{(1)} f_2(x, X_x) + \frac{1}{2}(A_x^{(2)})^2 f_{22}(x, X_x) \right] dx + \int_s^t A_x^{(2)} f_2(x, X_x) dW_x, \tag{3.35}$$

where $A_x^{(i)} = A^{(i)}(x, X_x)$. This is the most general form of the Itô lemmas, and it is used in the study of stochastic differential equations in Section 3.4.

3.3.3 Itô versus Stratonovich

There is a direct relation between the Itô and Stratonovich integrals of a smooth function f, which is given by

$$\int_0^T f(W_t) \circ dW_t = \int_0^T f(W_t) dW_t + \frac{1}{2} \int_0^T f'(W_t) dt, \tag{3.36}$$

where the first integral on the right-hand side is the Itô integral. This can be shown by defining $y_j = (t_{j+1} + t_j)/2$ and then considering the sum

$$I_N = \sum_{j=0}^{N-1} f(W_{y_j})(W_{t_{j+1}} - W_{t_j}).$$

Using the Taylor expansion

$$f(W_{y_j}) = f(W_{t_j}) + f'(W_{t_j})(W_{y_j} - W_{t_j}) + \cdots,$$

gives

$$I_N \cong \sum_{j=0}^{N-1} \left[f(W_{t_j})(W_{t_{j+1}} - W_{t_j}) + f'(W_{t_j})(W_{y_j} - W_{t_j})(W_{t_{j+1}} - W_{t_j}) \right].$$

With $W_{t_{j+1}} - W_{t_j} = W_{t_{j+1}} - W_{y_j} + W_{y_j} - W_{t_j}$ in the second sum we obtain

$$\begin{aligned} I_N \cong & \sum_{j=0}^{N-1} f(W_{t_j})(W_{t_{j+1}} - W_{t_j}) \\ & + \sum_{j=0}^{N-1} f'(W_{t_j})(W_{y_j} - W_{t_j})^2 \\ & + \sum_{j=0}^{N-1} f'(W_{t_j})(W_{y_j} - W_{t_j})(W_{t_{j+1}} - W_{t_j}). \end{aligned}$$

In the limit $N \to \infty$, the first term converges (in the mean-square sense) to the Itô integral of f, the second term converges to the second term on the right-hand side of (3.36) and the third sum converges to zero (this requires some more detailed analysis; see Chapter 2 of Mikosch [2000]).

The Stratonovich calculus is more similar to the deterministic calculus. For example, if we take $f(t) = g'(t)$ in the first Itô lemma (3.30), we find

$$\begin{aligned} & \int_0^T g'(W_x) dW_x + \frac{1}{2} \int_0^T g''(W_x) dx = g(W_T) - g(W_0) \\ \Rightarrow & \int_0^T f(W_x) dW_x + \frac{1}{2} \int_0^T f'(W_x) dx = g(W_T) - g(W_0). \end{aligned}$$

With (3.36), we find

$$\int_0^T f(W_x) \circ dW_x = \int_0^T g'(W_x) \circ dW_x = g(W_T) - g(W_0), \tag{3.37}$$

which is similar to the classical main theorem of integral calculus. The Stratonovich exponential is, therefore, simply $X_t = e^{W_t}$ (Fig. 3.3a).

3.4 Stochastic differential equations

A general scalar ordinary differential equation (ODE) is written as

$$\frac{dx}{dt} = f(t, x) \to dx = f(t, x) dt. \tag{3.38}$$

With an initial condition $x(0) = x_0$, it has a formal solution,

$$x(t) = \int_0^t f(s, x)ds + x_0. \tag{3.39}$$

A general stochastic differential equation is written as

$$dX_t = a(t, X_t)dt + b(t, X_t)dW_t, \tag{3.40}$$

for smooth functions a, b and with initial condition X_0. The formal solution of (3.40) is given,

$$X_t - X_0 = \int_0^t a(s, X_s)ds + \int_0^t b(s, X_s)dW_s. \tag{3.41}$$

Either form (3.40) or (3.41) is referred to as the Itô stochastic differential equation (SDE). It can be deduced from (3.36) (see Section 3.2.3 of Mikosch [2000]) that, for each Itô SDE (3.40), there is an equivalent Stratonovich SDE of the form

$$dX_t = (a(t, X_t) - \frac{1}{2}b(t, X_t)\frac{\partial b}{\partial x}(t, X_t))dt + b(t, X_t) \circ dW_t. \tag{3.42}$$

A strong solution X_t of an Itô SDE (or equivalent Stratonovich SDE) has the following properties:

(i) X_t satisfies (3.41), and at time t, it is a function of W_s for $s \leq t$ and the coefficient functions a and b.
(ii) The integrals in (3.41) are well defined in terms of Riemann or Itô stochastic integrals.

The important notion here is that if we change the Wiener process W_t, then also the strong solution X_t changes, but the functional relation between X_t and W_t remains the same. For a weak solution X_t of (3.41), the sample path dependence is not needed; for given X_0, a and b, we just have to find a W_t for which (3.41) holds. There exist Itô SDEs that have only weak solutions!

Assume that the initial condition X_0 has a finite second moment ($E[X_0^2] < \infty$) and that for all $t \in [0, T]$ and $x, y \in \mathbb{R}$, the coefficient functions $a(t, x)$ and $b(t, x)$ are continuous and satisfy a Lipschitz condition in the second variable, that is,

$$|a(t, x) - a(t, y)| + |b(t, x) - b(t, y)| \leq K|x - y|, \tag{3.43}$$

for certain $K > 0$, then the Itô SDE (3.41) has a unique strong solution X_t on $[0, T]$. For the proof, see Oksendal (1995), Theorem 5.5. In most practical applications, such conditions are satisfied, and existence and uniqueness of strong solutions are guaranteed.

As an example, consider the simple case $a = 0$ and $b = 1$, which gives the SDE

$$X_t - X_0 = \int_0^t dW_s. \tag{3.44}$$

The solution is $X_t = W_t$ (just by evaluating the stochastic integral), and hence the Wiener process is itself a strong solution of this SDE.

The general linear scalar Itô SDE is defined as

$$X_t - X_0 = \int_0^t (c_1(s)X_s + c_2(s))ds + \int_0^t (\sigma_1(s)X_s + \sigma_2(s))dW_s, \tag{3.45}$$

with smooth and bounded functions c_1, c_2, σ_1 and σ_2 on the interval $[0, T]$. With $a(t, x) = c_1(t)x + c_2(t)$ and $b(t, x) = \sigma_1(t)x + \sigma_2(t)$, we find from the Lipschitz condition

$$\begin{aligned} |a(t, x) - a(t, y)| + |b(t, x) - b(t, y)| &= |c_1(t)(x - y)| + |\sigma_1(t)(x - y)| \\ &\leq K|(x - y)|, \end{aligned} \tag{3.46}$$

where $K = \max_{t\in[0,T]}(|c_1(t)| + |\sigma_1(t)|)$ and hence (3.45) has a unique strong solution on every interval $[0, T]$. In the following subsections, we consider some specific cases.

3.4.1 Pure additive noise

When $\sigma_1 = 0$, the stochastic integral in (3.45) does not depend on the solution X_t, and this case is referred to as 'pure additive' noise. The Itô SDE becomes

$$X_t - X_0 = \int_0^t (c_1(s)X_s + c_2(s))ds + \int_0^t \sigma_2(s)dW_s. \tag{3.47}$$

To solve this equation, we introduce the process $Y_t = f(t, X_t) = \alpha(t)X_t$ where

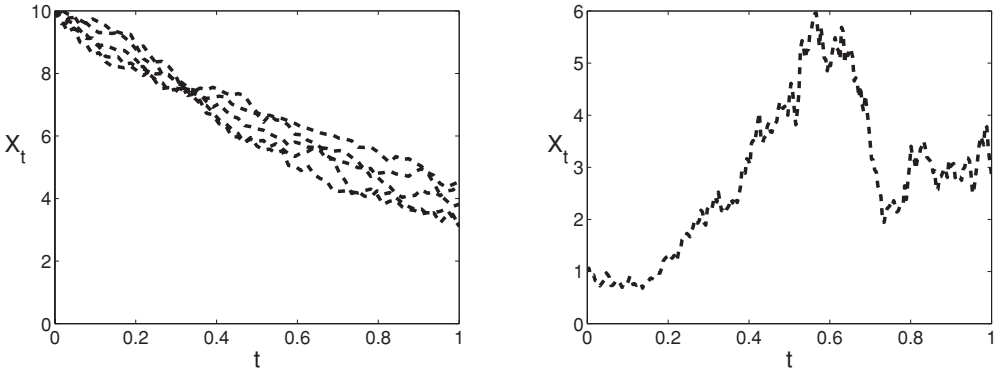
$$\alpha(t) = e^{-\int_0^t c_1(s)ds}.$$

With $A^{(1)} = c_1X + c_2$ and $A^{(2)} = \sigma_2$, the third Itô lemma (3.35) is applied to Y_t to give (with $f_1 = \alpha' x$, $f_2 = \alpha$ and $f_{22} = 0$)

$$\begin{aligned} \alpha(t)X_t - \alpha(0)X_0 = &\int_0^t \left[\alpha'(x)X_x + (c_1(x)X_x + c_2(x))\alpha(x)\right]dx \\ &+ \int_0^t \alpha(x)\sigma_2(x)dW_x. \end{aligned}$$

Because $\alpha' = -c_1\alpha$ and $\alpha(0) = 1$, we find

$$X_t = \frac{1}{\alpha(t)}\left[X_0 + \int_0^t \alpha(x)c_2(x)dx + \int_0^t \alpha(x)\sigma_2(x)dW_x\right]. \tag{3.48}$$



(a) (b)

Figure 3.4 (a) Five sample paths of the Ornstein-Uhlenbeck process (3.49) with $\gamma = 1$, $\sigma = 1$ and $X_0 = 10$. (b) One sample path of the solution (3.53) for $\lambda = 2$, $\mu = 1$ and $X_0 = 1$.

A prominent example is the Langevin equation, for which $c_1(t) = -\gamma$, $c_2 = 0$ and $\sigma_2 = \sigma$, where γ and σ are constants. We find that in this case

$$\alpha(t) = e^{\int_0^t \gamma ds} = e^{\gamma t},$$

and the solution (3.48) becomes

$$X_t = e^{-\gamma t}\left[X_0 + \sigma \int_0^t e^{\gamma x}\, dW_x\right]. \tag{3.49}$$

The stochastic process associated with this solution is called the Ornstein-Uhlenbeck process (Fig. 3.4a).

3.4.2 Pure multiplicative noise

When $\sigma_2 = 0$, the stochastic integral in (3.45) depends only on the solution X_t, and this case is referred to as 'multiplicative' (or state-dependent) noise. With also $c_2 = 0$, the Itô SDE (3.45) becomes

$$X_t - X_0 = \int_0^t c_1(s)X_s ds + \int_0^t \sigma_1(s)X_s dW_s. \tag{3.50}$$

We apply again the third Itô lemma (3.35), but now on the process $Y_t = \ln X_t$. With $f(t, x) = \ln x$, $f_1 = 0$, $f_2 = 1/x$ and $f_{22} = -1/x^2$, we find

$$\ln X_t - \ln X_0 = \int_0^t \left[c_1(x)X_x \frac{1}{X_x} + \frac{1}{2}(\sigma_1(x)X_x)^2(-\frac{1}{X_x^2})\right] dx + \int_0^t \sigma_1(x)X_x \frac{1}{X_x} dW_x, \tag{3.51}$$

and hence the solution is given by

$$X_t = X_0 \exp\left\{\int_0^t \left[c_1(x) - \frac{1}{2}\sigma_1^2(x)\right]dx + \int_0^t \sigma_1(x)dW_x\right\}. \tag{3.52}$$

An example is given by the special case $c_1 = \lambda$ and $\sigma_1 = \mu$, where λ, μ are constants. In this case, the solution (Fig. 3.4b) follows immediately as

$$X_t = X_0 e^{(\lambda - \mu^2/2)t + \mu W(t)}. \tag{3.53}$$

This solution is often used to test the convergence of several numerical solution methods for SDEs (Section 3.6).

3.4.3 Complete solution of the linear scalar Itô SDE

The general linear one-dimensional Itô SDE (3.45), now written as

$$Y_t - Y_0 = \int_0^t (c_1(s)Y_s + c_2(s))ds + \int_0^t (\sigma_1(s)Y_s + \sigma_2(s))dW_s, \tag{3.54}$$

can be completely solved analytically. We refer to Section 3.3.3 of Mikosch (2000) for the derivation and present only the answer here for completeness as

$$Y_t = X_t(Y_0 + \int_0^t [c_2(x) - \sigma_1(x)\sigma_2(x)]\, X_x^{-1}dx + \int_0^t \sigma_2(x)X_x^{-1}dW_x), \tag{3.55}$$

where X_t is the solution (3.52).

3.4.4 Mean and variance

Once we have obtained solutions to SDEs, we want to know their statistical properties. Let $\mu_X(t)$ indicate the mean and $Var(X_t) = \sigma_X^2(t)$ the variance of the stochastic process X. Note that these in general depend on time, as for every fixed t, we have a distribution of X. There are basically two ways to obtain this information: (i) by direct evaluation using the discrete form of the stochastic integral and (ii) through formulation of ODEs for these statistical properties. We will illustrate both approaches with the Ornstein-Uhlenbeck process (3.49).

Following approach (i), we first determine

$$E[\int_0^t e^{\gamma x}dW_x] = E[\lim_{N\to\infty} \sum_{j=0}^{N-1} e^{\gamma t_j}(W(t_{j+1}) - W(t_j))]$$

$$= \lim_{N\to\infty} \sum_{j=0}^{N-1} e^{\gamma t_j} E[W(t_{j+1})] - E[W(t_j))] = 0,$$

and hence

$$\mu_X(t) = E[e^{-\gamma t}\left[X_0 + \int_0^t e^{\gamma x}\sigma\, dW_x\right]] = e^{-\gamma t}E[X_0]. \tag{3.56}$$

Next, we determine (using $Var(cX) = c^2 Var(X)$ and $Var(dW_j) = dt$)

$$\begin{aligned} Var\left(\int_0^t e^{\gamma x} dW_x\right) &= Var\left(\lim_{N\to\infty}\sum_{j=0}^{N-1} e^{\gamma t_j}(W(t_{j+1}) - W(t_j))\right) \\ &= \lim_{N\to\infty}\sum_{j=0}^{N-1} e^{2\gamma t_j} Var(W(t_{j+1}) - W(t_j)) \\ &= \lim_{N\to\infty}\sum_{j=0}^{N-1} e^{2\gamma t_j}(t_{j+1} - t_j) = \int_0^t e^{2\gamma x} dx = \frac{e^{2\gamma t} - 1}{2\gamma}, \end{aligned}$$

and hence for X_t in (3.49),

$$Var(X_t) = e^{-2\gamma t}\left(Var(X_0) + \sigma^2\frac{e^{2\gamma t} - 1}{2\gamma}\right). \tag{3.57}$$

Following approach (ii), we consider the general linear scalar SDE given by (3.45) and apply directly the expectation operator, which gives (using $E[dW_s] = 0$),

$$\mu_X(t) = E[X_t] = E[X_0] + \int_0^t (c_1(s)E[X_s] + c_2(s))ds. \tag{3.58}$$

Differentiation of this relation to t gives

$$\mu_X'(t) = c_1(t)\mu_X(t) + c_2(t), \tag{3.59}$$

with initial conditions $\mu_X(0) = E[X_0]$. For the Ornstein-Uhlenbeck process (with $c_1 = -\gamma$, $c_2 = 0$), we find

$$\mu_X'(t) = -\gamma\mu_X(t) \Rightarrow \mu_X(t) = \mu_X(0)e^{-\gamma t}, \tag{3.60}$$

which is in correspondence with (3.56).

To obtain the variance using approach (ii), we define $q_X(t) = E[X_t^2]$. An expression for X_t^2 can be found using the third Itô lemma (3.35) with $f(t, x) = x^2$ (such that $f_1 = 0$, $f_2 = 2x$, $f_{22} = 2$). This gives for (3.45)

$$\begin{aligned} X_t^2 - X_0^2 = &\int_0^t \left[2X_s(c_1X_s + c_2) + (\sigma_1X_s + \sigma_2)^2\right]ds \\ &+ \int_0^t 2X_s(\sigma_1X_s + \sigma_2)dW_s. \end{aligned}$$

Taking the expectation operator and using the fact that X_s and dW_s are independent then gives

$$E[X_t^2] = E[X_0^2] + \int_0^t \left[2(c_1 E[X_s^2] + c_2 E[X_s]) + (\sigma_1^2 E[X_s^2] + 2\sigma_1\sigma_2 E[X_s] + \sigma_2^2\right] ds$$

or

$$q_X(t) = q_X(0) + \int_0^t \left[(2c_1 + \sigma_1^2) q_X(s) + (2c_2 + 2\sigma_1\sigma_2)\mu_X(s) + \sigma_2^2\right] ds.$$

Differentiating to time, we finally obtain

$$q_X'(t) = (2c_1 + \sigma_1^2) q_X(t) + (2c_2 + 2\sigma_1\sigma_2)\mu_X(t) + \sigma_2^2. \tag{3.61}$$

For the Ornstein-Uhlenbeck process (with $c_1 = -\gamma$, $c_2 = 0$, $\sigma_1 = 0$, $\sigma_2 = \sigma$), this reduces to

$$q_X'(t) = -2\gamma q_X(t) + \sigma^2,$$

with initial conditions $q_X(0) = E[X_0^2]$, having the solution

$$q_X(t) = e^{-2\gamma t}\left[\frac{\sigma^2}{2\gamma}(e^{2\gamma t} - 1) + q_X(0)\right].$$

Finally, the variance of the process follows from

$$\begin{aligned} Var(X_t) = E[X_t^2] - (E[X_t])^2 &= q_X(t) - \mu_X^2(t) \\ &= \frac{\sigma^2}{2\gamma}(1 - e^{-2\gamma t}) + e^{-2\gamma t}(q_X(0) - \mu_X^2(0)) \\ &= e^{-2\gamma t} Var(X_0) + \frac{\sigma^2}{2\gamma}(1 - e^{-2\gamma t}), \end{aligned} \tag{3.62}$$

which is in agreement with (3.57).

3.5 The Fokker-Planck equations

An alternative approach to computing statistical properties of a stochastic process associated with an SDE is the determination of its Fokker-Planck equation (FPE).

3.5.1 Markov processes

Recall from Section 3.2 that the distribution function $F_{\mathbf{X}}(\mathbf{x})$ for a random vector $\mathbf{X}$ was given by (3.12), for convenience here written again as

$$F_{\mathbf{X}}(\mathbf{x}) = P(X_1 \le x_1, \ldots, X_n \le x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\mathbf{X}}(y_1, \ldots, y_n) dy_1 \ldots dy_n, \tag{3.63}$$

where $f_{\mathbf{X}}$ is the (joint) probability density function. In the following, we associate the indices $1, \ldots, n$ with times $t_1, \ldots, t_n$ and write

$$f_{\mathbf{X}}(x_1, \ldots, x_n) = p(x_1, t_1; \ldots; x_n, t_n). \tag{3.64}$$

The conditional probability $P(A|B)$ of two events A and B is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \tag{3.65}$$

where $P(A)$ and $P(B)$ are the probabilities of the events A and B, respectively. Hence the probability $P(A|B)$ concerns events A, which are contained in the set B. We need this concept to define a Markov process for which the following property, the Markov property, holds,

$$p(x_1, t_1; \ldots; x_n, t_n|y_1, \tau_1; \ldots; y_n, \tau_n) = p(x_1, t_1; \ldots; x_n, t_n|y_1, \tau_1) \tag{3.66}$$

for $t_1 \geq \ldots \geq t_n \geq \tau_1 \geq \ldots \geq \tau_n$. Loosely speaking, for a Markov process, one can make future predictions based solely on its present state just as well as one could knowing the process's full history.

A well-known example of a Markov process is the following: suppose that someone is popping many kernels of popcorn, and each kernel will pop at an independent, uniformly random time within the next time interval. Let X_t denote the number of kernels that have popped up to time t. If after some amount of time, one wants to guess how many kernels will pop in the next second, one needs only know how many kernels have popped. It will not help to know when they popped, so knowing X_t for previous times t will not inform the guess any better.

For a Markov process, we find (using 3.65)

$$\begin{aligned}
p(x_1, t_1; \ldots; x_n, t_n) &= p(x_1, t_1; \ldots; x_{n-1}, t_{n-1}|x_n, t_n)p(x_n, t_n) \\
&= p(x_1, t_1; \ldots; x_{n-2}, t_{n-2}|x_{n-1}, t_{n-1}; x_n, t_n)p(x_{n-1}, t_{n-1}|x_n, t_n)p(x_n, t_n) \\
&= p(x_1, t_1; \ldots; x_{n-2}, t_{n-2}|x_{n-1}, t_{n-1})p(x_{n-1}, t_{n-1}|x_n, t_n)p(x_n, t_n) \\
&= p(x_1, t_1|x_2, t_2) \ldots p(x_{n-1}, t_{n-1}|x_n, t_n)p(x_n, t_n).
\end{aligned}$$

Hence only the so-called transition probability $p(x_{i-1}, t_{i-1}|x_i, t_i)$ is needed to describe the joint probability density function of a Markov process.

3.5.2 Forward Fokker-Planck equation

We now return to the general Itô SDE (3.40) given by

$$X_t = X_0 + \int_0^t a(X_s, s)ds + \int_0^t b(X_s, s)dW_s \tag{3.67}$$

for smooth functions a and b. With X_0 given, the future time development is uniquely determined by $W_t, t > 0$. As W_t for $t > 0$ is independent of X_t for $t < 0$, we conclude that X_t for $t > 0$ is independent of X_t for $t < 0$, provided X_0 is known, and hence X_t is a Markov process (for a more extensive discussion, see Gardiner [2002], Section 4.3.2). Let the transition probability be indicated by $p(x, t) = p(x, t|x_0, t_0)$.

Now by definition of the expectation operator, for any smooth function $f : \mathbb{R} \to \mathbb{R}$, we find

$$E[f(X_t)] = \int_{-\infty}^{\infty} f(x)p(x, t)dx \tag{3.68}$$

and hence

$$\frac{d}{dt}E[f(X_t)] = \int_{-\infty}^{\infty} f(x)\frac{\partial p}{\partial t}(x, t)dx. \tag{3.69}$$

However, when we use the third Itô lemma (3.35) for f, we find

$$\begin{aligned} f(X_t) - f(X_0) = &\int_0^t \left[a(X_s, s)\frac{\partial f}{\partial x}(X_s) + \frac{1}{2}b^2(X_s, s)\frac{\partial^2 f}{\partial x^2}(X_s)\right] ds \\ &+ \int_0^t b(X_s, s)\frac{\partial f}{\partial x}(X_s)dW_s. \end{aligned} \tag{3.70}$$

Taking the expectation operator of (3.70), using (3.68) and differentiating the result to t then gives

$$\frac{d}{dt}E[f(X_t)] = \int_{-\infty}^{\infty} \left[a(x, t)\frac{\partial f}{\partial x}(x) + \frac{1}{2}b^2(x, t)\frac{\partial^2 f}{\partial x^2}(x)\right] p(x, t)dx. \tag{3.71}$$

Combining (3.71) and (3.69), we find

$$\int_{-\infty}^{\infty} \left[(a(x, t)\frac{\partial f}{\partial x}(x) + \frac{1}{2}b^2(x, t)\frac{\partial^2 f}{\partial x^2}(x))p(x, t) - f(x)\frac{\partial p}{\partial t}(x, t)\right] dx = 0.$$

When furthermore it is assumed that $p, \partial p/\partial x \to 0$ for $x \to \pm\infty$, then partial integration of the terms with a and b gives

$$\int_{-\infty}^{\infty} f\left(\frac{\partial p}{\partial t} + \frac{\partial(ap)}{\partial x} - \frac{1}{2}\frac{\partial^2(pb^2)}{\partial x^2}\right) dx = 0, \tag{3.72}$$

and as f is arbitrary, we find for p the forward Fokker-Planck equation

$$\frac{\partial p}{\partial t} + \frac{\partial(ap)}{\partial x} - \frac{1}{2}\frac{\partial^2(pb^2)}{\partial x^2} = 0. \tag{3.73}$$

Once this Fokker-Planck equation is solved, the probability distribution of the stochastic process X_t is totally determined.

Example 3.3 Forward FPE of the Ornstein-Uhlenbeck process For the Ornstein-Uhlenbeck process we have $a(x,t) = -\gamma x$ and $b(x,t) = \sigma$. The Fokker-Planck equation (3.73), then becomes

$$\frac{\partial p}{\partial t} = \frac{\partial(\gamma x p)}{\partial x} + \frac{\sigma^2}{2}\frac{\partial^2 p}{\partial x^2},$$

with $p, \partial p/\partial x \to 0$ for $x \to \pm\infty$. In many cases, only the stationary distribution of p is desired. When putting the time derivative to zero, the resulting equation can be integrated to x to give

$$\gamma x p + \frac{\sigma^2}{2}\frac{\partial p}{\partial x} = C_1,$$

and $C_1 = 0$ through the boundary conditions. Integrating once more, we find

$$p(x) = C_2 e^{-\frac{\gamma}{\sigma^2}x^2}; \quad \int_{-\infty}^{\infty} p(x)dx = 1 \Rightarrow C_2 = \sqrt{\frac{\gamma}{\pi}}\frac{1}{\sigma}.$$

We immediately conclude from this normal distribution that $\mu[X] = 0$ and that $Var[X] = \sigma^2/(2\gamma)$, which is exactly the limit $t \to \infty$ of the expression (3.62). ■

3.5.3 Backward Fokker-Planck equation

In a forward problem, we have information about the state X_t of a particular process at time t (i.e., a probability distribution $p(x,t)$), and we want to know the probability distribution at a later time $s > t$. Hence $p(x,t)$ serves as an initial condition for the forward Fokker-Planck equation. In many problems, however, we want to know for every state at time t the probability of ending up at a future time $s > t$ in a target state X_s. A typical example is a so-called exit time problem, where the target state is outside a specific domain, as discussed in the next subsection. To solve this problem, the probability distribution $p(x,s)$ serves as a final condition, and we have to integrate the adjoint equation of the forward Fokker-Planck equation back in time. This adjoint equation is called the backward Fokker-Planck equation. With the inner product

$$< f, g >= \int_{-\infty}^{\infty} f(x)g(x)dx, \tag{3.74}$$

the adjoint operator $L^\dagger$ is defined as $< Lf, g >=< f, L^\dagger g >$. Writing (3.73) as $\partial p/\partial t = Lp$, we can write

$$< Lf, g >= \int_{-\infty}^{\infty} g\left(-\frac{\partial(af)}{\partial x} + \frac{1}{2}\frac{\partial^2(fb^2)}{\partial x^2}\right)dx. \tag{3.75}$$

Using partial integration (and with vanishing boundary conditions for f and g at $x = \pm\infty$) leads to

$$< f, L^{\dagger} g > \int_{-\infty}^{\infty} \left(a \frac{\partial g}{\partial x} + \frac{b^2}{2} \frac{\partial^2 g}{\partial x^2} \right) f dx \tag{3.76}$$

and hence the backward Fokker-Planck equation is given by (see, e.g., Risken, 1989)

$$-\frac{\partial p}{\partial t} = a \frac{\partial p}{\partial x} + \frac{b^2}{2} \frac{\partial^2 p}{\partial x^2}. \tag{3.77}$$

3.5.4 Exit-time problems

The backward Fokker-Planck equation is used to solve so-called exit-time problems. Let a particle be at a position x at $t = 0$ and assume that its position for later times is described by a stochastic process X_t satisfying the SDE (3.40). The problem is to estimate the average time it will take to leave a certain interval $x \in (x_0, x_1)$. When the particle reaches either x_0 or x_1, it is removed from the system; hence when it is still in (x_0, x_1), it has never left the interval.

To solve this exit-time problem, let the time when the particle leaves the interval be indicated by $T(x)$. For the probability $P(T(x) \geq t)$ we can write

$$P(T(x) \geq t) = \int_{x_1}^{x_0} p(y, t|x, 0) dy \equiv G(x, t), \tag{3.78}$$

because the right-hand side is precisely the probability that the particle is still in the interval (x_0, x_1).

As argued in the previous subsection, the probability $p(y, t|x, 0)$ satisfies a backward Fokker-Planck equation (3.77). Because $p(y, t|x, 0) = p(y, 0|x, -t)$ (time-shift), we find that $G(x, t)$ satisfies

$$\frac{\partial G}{\partial t} = a \frac{\partial G}{\partial x} + \frac{b^2}{2} \frac{\partial^2 G}{\partial x^2}. \tag{3.79}$$

Because $p(y, 0|x, 0) = \delta(x - y)$, the initial conditions are $G(x, 0) = 1$ for $x \in (x_0, x_1)$ and $G(x, 0) = 0$ elsewhere. If $x = x_0$ or $x = x_1$, the particle is absorbed immediately; hence $P(T(x) \geq t) = 0$. The boundary conditions for G hence become

$$G(x_0, t) = G(x_1, t) = 0. \tag{3.80}$$

Because G is the probability that $T(x) \geq t$, the mean first passage time $\overline{T}(x)$ can be written as the expectation value

$$\overline{T}(x) = \int_0^{\infty} t \, d(1 - G(x, t)) = -\int_0^{\infty} t \frac{\partial G}{\partial t} \, dt = \int_0^{\infty} G(x, t) \, dt. \tag{3.81}$$

When (3.79) is integrated in time over the interval $(0, \infty)$ we find

$$G(x, \infty) - G(x, 0) = -1 = a\frac{\partial \overline{T}}{\partial x} + \frac{b^2}{2}\frac{\partial^2 \overline{T}}{\partial x^2}, \tag{3.82}$$

with boundary conditions $\overline{T}(x_0) = \overline{T}(x_1) = 0$; the general solution for $\overline{T}(x)$ can be found in Gardiner (2002), Section 5.2.7.

We now consider the special case that $x_0 \to -\infty$, and the particle is only removed at $x = x_1$. Because at $x = x_0$ the probability will not depend on x anymore, we now require boundary conditions

$$\frac{\partial G}{\partial x}(x_0, t) = G(x_1, t) = 0, \tag{3.83}$$

instead of (3.80) and, consequently, $\overline{T}'(x_0) = \overline{T}(x_1) = 0$. To solve (3.82), we put $S(x) = \overline{T}'(x)$, solve for the homogeneous problem first and then determine a particular solution by variation of constants. The result (with the boundary condition at $x = x_0$ included) is

$$S(x) = \int_{-\infty}^{x} \frac{-2}{b^2(s)}\psi(s)ds, \tag{3.84}$$

where $\psi(x)$ is defined as

$$\psi(x) = e^{\int_{-\infty}^{x} 2\frac{a(s)}{b^2(s)}ds}. \tag{3.85}$$

By integrating $S(x)$, the solution for the mean first passage time is given by

$$\overline{T}(x) = 2\int_{x}^{x_1}\left[\frac{1}{\psi(y)}\int_{-\infty}^{y}\frac{\psi(z)}{b^2(z)}dz\right]dy. \tag{3.86}$$

Example 3.4 Exit times for a double-well potential Consider the Itô stochastic differential equation

$$dX_t = -V'(x)dt + \sigma dW_t,$$

for which the stationary probability density function can be explicitly solved from the forward Fokker-Planck equation (just as in Example 3.3) as

$$p(x) = Ce^{-2\frac{V(x)}{\sigma^2}}, \quad \int_{-\infty}^{\infty} p(x)\,dx = 1.$$

The function $V(x)$ can be viewed as a potential, and let us assume that it has a minimum at $x = x_0$ and a maximum at $x = x_1$. In Fig. 3.5, the potential $V(x) = x^4/4 - x^2/2$ is plotted, for which $x_0 = -1$ and $x_1 = 0$.

We want to determine the mean first passage time of a particle in the potential well near $x = x_0$ over the potential barrier at $x = x_1$ and use the notation $T(x_0 \to x_1)$. In the solution (3.86), the function ψ is given by

$$\psi(x) = e^{-\frac{2V(x)}{\sigma^2}},$$

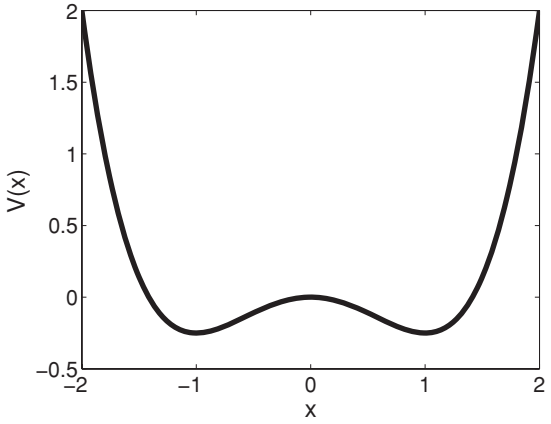


Figure 3.5 Plot of the potential function $V(x) = x^4/4 - x^2/2$ with $x_0 = -1$ (minimum) and $x_1 = 0$ (maximum).

and hence $T(x_0 \to x_1)$ is given by

$$T(x_0 \to x_1) = \frac{2}{\sigma^2} \int_{x_0}^{x_1} \left[e^{\frac{2V(y)}{\sigma^2}} \int_{-\infty}^{y} e^{-\frac{2V(z)}{\sigma^2}} dz \right] dy.$$

When σ is small, the integral over z will only be relevant for values of y near x_1 because, otherwise, the integral values of the first integrand will be small. Hence we can approximate the integral by taking $y = x_1$ in the inner integral, giving

$$T(x_0 \to x_1) \approx \frac{2}{\sigma^2} \int_{x_0}^{x_1} e^{\frac{2V(y)}{\sigma^2}} dy \int_{-\infty}^{x_1} e^{-\frac{2V(z)}{\sigma^2}} dz. \tag{3.87}$$

■

3.5.5 *Variance in fast-slow systems*

The Fokker-Planck approach can also be applied efficiently to study the change in variance in fast-slow systems (as discussed in Section 2.3.3) when a bifurcation point is approached. Consider, for example, the two-dimensional Itô SDE (Kuehn, 2011)

$$dX_t = f(X_t, Y_t)dt + \sigma dW_t, \tag{3.88a}$$

$$dY_t = \epsilon\, dt, \tag{3.88b}$$

where X_t, Y_t are both one-dimensional stochastic variables and σ is constant.

For $\epsilon = 0$, the forward Fokker-Planck for the probability density function $p^y(x, t)$ is

$$\frac{\partial p^y}{\partial t} = -\frac{\partial (f(x, y)p^y)}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2 p^y}{\partial x^2}, \tag{3.89}$$

and the stationary distribution $\bar{p}^y$ is given by

$$\bar{p}^y(x) = \frac{1}{N} e^{\int_a^x \frac{2}{\sigma^2} f(s,y)\, ds}, \tag{3.90}$$

where N is again a normalization factor.

Consider now, for example, the normal form of the saddle node, given by $f_1(x, y) = -y - x^2$ (cf. Example 2.3). Here, the interval (a, b) consists of the points that do not escape to infinity, that is, $a = -\sqrt{-y}$, $b = \infty$. Evaluating the integral in (3.90) then gives

$$\int_{-\sqrt{-y}}^{x} (-y - s^2) ds = -yx - \frac{1}{3}x^3 + \frac{2}{3}(-y)^{3/2},$$

and hence the equilibrium probability density function is given by

$$\bar{p}_1^y(x) = \frac{1}{N_1} e^{\frac{2}{\sigma^2}(-yx - \frac{1}{3}x^3 + \frac{2}{3}(-y)^{3/2})}. \tag{3.91}$$

Similar expressions can be derived for the transcritical bifurcation ($f_2(x, y) = yx - x^2$, $a = y$, $b = \infty$) and the subcritical pitchfork bifurcation ($f_3(x, y) = yx + x^3$, $a = -\sqrt{-y}$, $b = \sqrt{-y}$), giving expressions (Kuehn, 2011)

$$\bar{p}_2^y(x) = \frac{1}{N_2} e^{\frac{2}{\sigma^2}(\frac{1}{2}yx^2 - \frac{1}{3}x^3 - \frac{1}{6}y^3)}, \tag{3.92a}$$

$$\bar{p}_3^y(x) = \frac{1}{N_3} e^{\frac{2}{\sigma^2}(\frac{1}{2}yx^2 + \frac{1}{4}x^4 + \frac{1}{4}y^2)}. \tag{3.92b}$$

The variance of the system can be directly determined from the probability density function and is plotted for the three different bifurcations as a function of y in Fig. 3.6, for $\sigma = 0.1$. For each case, the variance increases sharply when the critical transition at $y = 0$ is approached. The distance of the local maximum to the critical transition decreases with decreasing noise level σ. This suggests that increased variance may be used as indicator of the approach to the critical transition.

3.6 Numerical solutions of SDEs

The numerical solution of SDEs is more involved than the solution of the deterministic counterparts. Consider an Itô SDE of the form

$$X(t) = X(0) + \int_0^t f(X(s))ds + \int_0^t g(X(s))dW(s), \tag{3.93}$$

where the notation of the stochastic integral is slightly changed. Let us define a partition $\tau_j = j\Delta t$, $j = 0, \ldots, n$ on $[0, T]$ with $\Delta t = T/n$ and indicate the numerical solution at τ_j with X_j (which is the reason for changing the notation) and the analytical solution with $X(\tau_j)$.

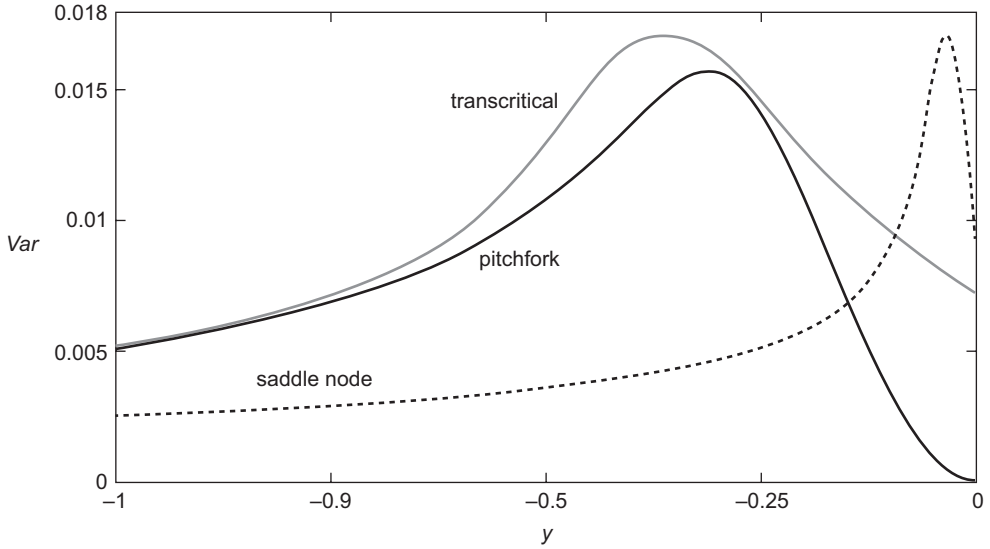


Figure 3.6 Variance for the stochastic system (3.88) for normal forms of three bifurcations and $\sigma = 0.1$ (figure slightly modified from Kuehn [2011]).

The order η of strong convergence for fixed k is such that

$$E[|X_k - X(\tau_k)|] \leq (\Delta t)^\eta. \tag{3.94}$$

Strong convergence therefore implies that the mean of the error converges to zero. On the contrary, weak convergence indicates only convergence of the expectation (error in the mean), and its order η is determined by

$$|E[X_k] - E[X(\tau_k)]| \leq (\Delta t)^\eta. \tag{3.95}$$

In the following subsections, two much-used schemes and their convergence behaviour are presented.

3.6.1 The Euler-Maruyama scheme

The Euler-Maruyama scheme for (3.93) is

$$X_j - X_{j-1} = f(X_{j-1})\Delta t + g(X_{j-1})(W(\tau_j) - W(\tau_{j-1})). \tag{3.96}$$

As an example, we consider $f(X) = \lambda X$ and $g(X) = \mu X$ for which we derived the analytical solution (3.53) as

$$X(t) = X(0)e^{(\lambda - \frac{\mu^2}{2})t + \mu W(t)}, \tag{3.97}$$

and we take $\lambda = 2, \mu = 1$ on the domain $[0, 1]$. First, a sample path of a Wiener process with a time step $\delta t = 2^{-8}$ and the exact solution (3.97) is shown as the drawn

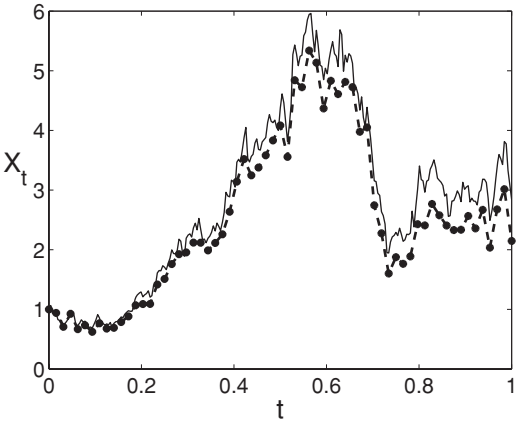


Figure 3.7 Application of the Euler scheme to the Itô SDE (3.50) for $X_0 = 1$, $\lambda = 2$ and $\mu = 1$. The drawn curve is the exact solution, and the labelled curve is the numerical solution with $\Delta t = 2^6$.

curve in Fig. 3.7. The numerical solution is computed with values of $\Delta t = R\delta t$ with $R = 4, 2$ and 1, and the solution for $R = 4$ is plotted in Fig. 3.7. For this case, the endpoint errors for $R = 4, 2, 1$ are 0.6907, 0.1595 and 0.0821, respectively.

To consider the strong and weak convergence of the Euler-Maruyama scheme, we choose $\delta t = 2^{-9}$ and compute 1,000 different sample paths of the Wiener process. For each path, the SDE is integrated with five different step sizes ($\Delta t = 2^{p-1}\delta t$, $p = 1, \ldots, 5$. The endpoint errors are computed for each of these paths, and the sample mean is computed over the 1,000 sample paths for each Δt. The result is shown in Fig. 3.8a, from which we see that the order of strong convergence is indeed near to a $1/2$. A power law fit to the four points gives an exponent $\eta = 0.5384$. The same is done to study weak convergence of the Euler-Maruyama method. In Fig. 3.8b, the error in the expectation value is plotted versus Δt. A power-law fit to the four points gives an exponent $\eta = 0.9858$. These results confirm that the Euler-Maruyama scheme has a strong convergence with order $\eta = 1/2$ and a weak convergence with order $\eta = 1$ (Kloeden and Platen, 1999).

3.6.2 The Milstein scheme

To improve the order of strong convergence, we need higher-order terms to be included into the discretization scheme. One of these schemes is the Milstein scheme, which we present now for the Itô SDE (3.93). We first write the discretization as

$$X_{t_j} - X_{t_{j-1}} = \int_{t_{j-1}}^{t_j} f(X_s)ds + \int_{t_{j-1}}^{t_j} g(X_s)dW_s, \tag{3.98}$$

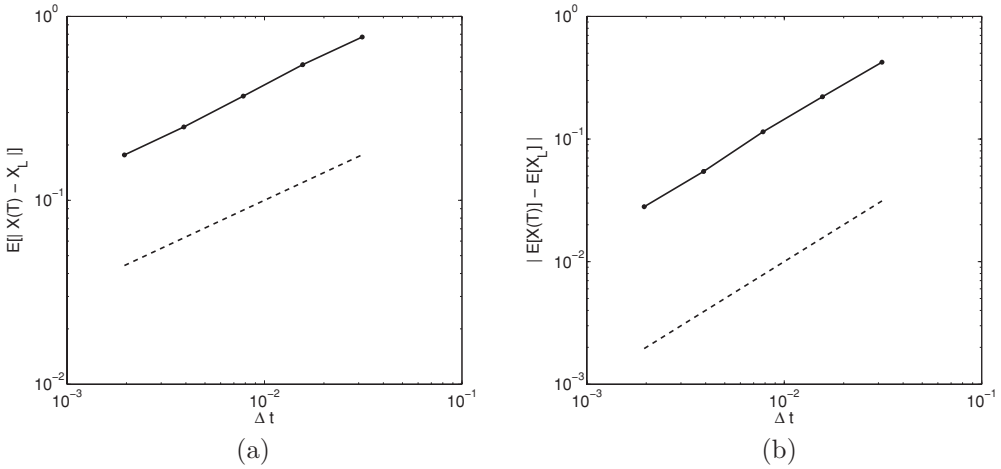


Figure 3.8 (a) Strong convergence of the Euler-Maruyama scheme to the Itô SDE (3.50) with $c_1 = \lambda$ and $\sigma_1 = \mu$ for $X_0 = 1$, $\lambda = 2$ and $\mu = 1$. The dotted curve is a line with slope $1/2$. (b) Weak convergence for the same case. The dotted curve is a line with slope 1.

and recover the Euler-Maruyama scheme, with $\Delta_j = t_j - t_{j-1}$ and $\Delta_j W = W_{t_j} - W_{t_{j-1}}$, as

$$\int_{t_{j-1}}^{t_j} f(X_s)ds = f(X_{t_{j-1}})\Delta_j, \tag{3.99a}$$

$$\int_{t_{j-1}}^{t_j} g(X_s)dW_s = g(X_{t_{j-1}})\Delta_j W. \tag{3.99b}$$

The crucial step in the derivation of higher-order schemes is the application of the third Itô lemma (3.35) for a function $f(x)$, with $f_1 = 0$, $f_2 = f'$, $f_{22} = f''$, whereas $A^{(1)} = f$ and $A^{(2)} = g$, according to (3.98). We then find

$$f(X_s) - f(X_{t_{j-1}}) = \int_{t_{j-1}}^{s} \left[ff' + \frac{1}{2}g^2 f'' \right] dy + \int_{t_{j-1}}^{s} gf' dW_y, \tag{3.100}$$

where the integration argument, y, has been suppressed for clarity. We do the same for the function g to obtain

$$g(X_s) - g(X_{t_{j-1}}) = \int_{t_{j-1}}^{s} \left[fg' + \frac{1}{2}g^2 g'' \right] dy + \int_{t_{j-1}}^{s} gg' dW_y. \tag{3.101}$$

Next we substitute the last two expressions into (3.98) and obtain

$$X_{t_j} - X_{t_{j-1}} = f(X_{t_{j-1}})\Delta_j + g(X_{t_{j-1}})\Delta_j W + R_j^1 + R_j^2, \tag{3.102}$$

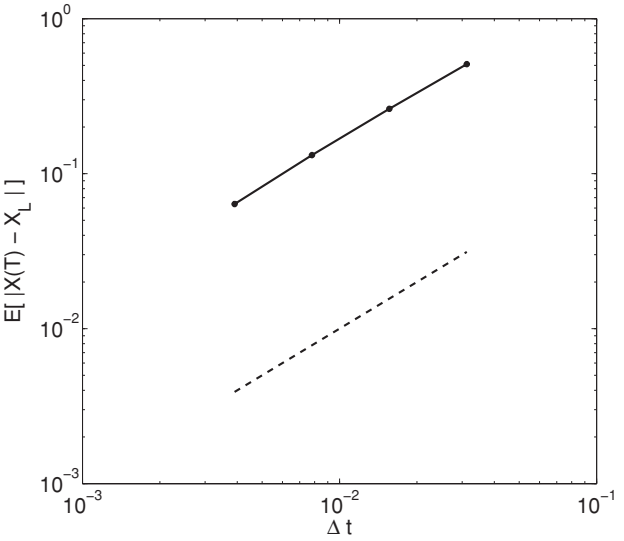


Figure 3.9 Test of strong convergence for the Milstein scheme for the same problem as in Fig 3.8. The dotted line has a slope 1.

with

$$R_j^1 = \int_{t_{j-1}}^{t_j} \int_{t_{j-1}}^{s} gg' dW_y dW_s. \tag{3.103}$$

As R_j^2 can be shown to be of smaller magnitude than R_j^1 (Kloeden and Platen, 1999), what remains is to evaluate R_j^1 as

$$R_j^1 = \frac{1}{2} g(X_{t_{j-1}}) g'(X_{t_{j-1}})((\Delta_j W)^2 - \Delta_j) \tag{3.104}$$

to finally give the Milstein scheme

$$X_{t_j} - X_{t_{j-1}} = f(X_{t_{j-1}})\Delta_j + g(X_{t_{j-1}})\Delta_j W + \frac{1}{2} g(X_{t_{j-1}}) g'(X_{t_{j-1}})((\Delta_j W)^2 - \Delta_j). \tag{3.105}$$

Again, we consider the Itô SDE with $f(X) = \lambda X$ and $g(X) = \mu X$ for which we derived the analytical solution (3.97), and we take $\lambda = 2$, $\mu = 1$ on the domain $[0, 1]$. To consider the strong and weak convergence of the Milstein scheme, we choose $\delta t = 2^{-12}$ and compute 1,000 different sample paths of the Wiener process. For each path, the SDE is integrated with four different step sizes $\Delta t = R\delta t$ with $R = 16, 32, 64$ and 128. The endpoint errors are computed for each of these paths, and the sample mean is computed over the 1,000 sample paths for each Δt. The results are shown in Fig. 3.9, from which we see that the order of convergence is indeed near to 1. A power-law fit to the four points gives an exponent $\eta = 0.9992$, and

hence the Milstein scheme has a faster strong convergence than the Euler-Maruyama scheme.

By repeatedly using the Itô lemma, even higher-order schemes can be used. A summary of these schemes is given in Kloeden and Platen (1999). A Fortran 95 version of some of these schemes is available at http://steck.us/computer.html.