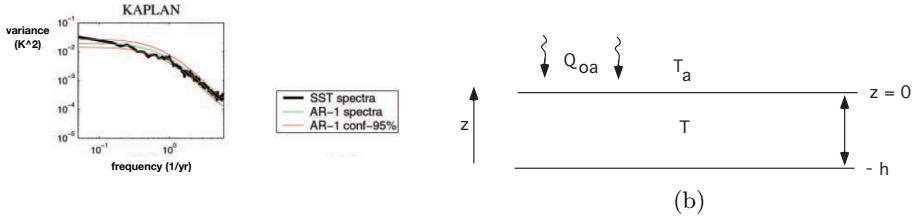


FIGURE 3. *(a) Spectra of midlatitude SST anomalies (Dommenget & Latif 2002) with frequency (1/yr) plotted on the x-axis. (b) Sketch of an upper ocean mixed layer forced by a heat flux Q_{oa}; the mixed-layer depth is indicated by h.*

2. The null-hypothesis of climate variability

From observations, a spectrum of midlatitude sea surface temperature (SST) anomalies (Dommenget & Latif 2002) is plotted in Fig. 3a. To understand this variability, a very idealized model was proposed by Hasselmann (1976). An oceanic mixed layer, with temperature T, is forced by an atmospheric heat flux Q_{oa} Fig. 3a. If we assume that the mixed layer depth h is fixed, the governing equation for T is simply

$$\rho C_p h \frac{dT}{dt} = Q_{oa}, \tag{2.1}$$

where ρ and C_p are the constant density and heat capacity of the ocean water, respectively (Fig. 3b).

The heat flux can be approximated as

$$Q_{oa} = K_Q(T_a - T), \tag{2.2}$$

where K_Q ($\mathrm{Wm^{-2}K^{-1}}$) is a heat exchange coefficient and T_a the atmospheric temperature just above the ocean surface. Combining (2.1) and (2.2), we obtain (with $\xi = \gamma T_a$)

$$\frac{dT}{dt} = -\gamma T + \xi \ ; \ \gamma = \frac{K_Q}{\rho C_p h}. \tag{2.3}$$

In general, the forcing part of this equation (here represented by ξ) will be very irregular (have energy in a wide range of frequencies) and hence can be represented by a random process. In this section, we will discuss how this is formulated mathematically.

2.1. *Random processes*

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.

This is easily generalized to a random vector $\mathbf{X} = (X_1, ..., X_n)$ where each X_i is a random variable. The distribution function $F_{\mathbf{X}}(\mathbf{x})$ is

$$F_{\mathbf{X}}(\mathbf{x}) = P(X_1 \leqslant x_1, ..., X_n \leqslant x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_{\mathbf{X}}(y_1, ..., y_n) dy_1 \dots dy_n, \tag{2.4}$$

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

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}(\det \Sigma)^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu_{\mathbf{X}})\Sigma^{-1}(\mathbf{x}-\mu_{\mathbf{X}})^T}, \tag{2.5}$$

where Σ is the covariance matrix with elements

$$\Sigma_{i,j} = cov[X_i, X_j] = E[(X_i - \mu_{X_i})(X_j - \mu_{X_j})]. \tag{2.6}$$

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

$$F_{X_1,X_2}(x_1, x_2) = F_{X_1}(x_1)F_{X_2}(x_2). \tag{2.7}$$

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

A stochastic process X_t is defined as a collection of random variables

$$(X_t, t \in T) = (X_t(\omega), t \in T, \omega \in \Omega), \tag{2.8}$$

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 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

$$c_X(t, s) = cov(X_t, X_s) = E[(X_t - \mu_X(t))(X_s - \mu_X(s))]. \tag{2.9}$$

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

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

$$(X_{t_1}, ..., X_{t_n}) \stackrel{d}{=} (X_{t_1+h}, ..., X_{t_n+h}), \tag{2.10}$$

where $\stackrel{d}{=}$ indicates equality in distribution sense.

A stochastic process $(X_t, t \in T)$ has stationary increments if

$$X_t - X_s \stackrel{d}{=} X_{t+h} - X_{s+h}, \tag{2.11}$$

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

In the following, we associate the indices $1, \ldots, n$ with times $t_1, \ldots, t_n$ and write

$$f_{\mathbf{X}}(x_1, ..., x_n) = p(x_1, t_1; \ldots; x_n, t_n). \tag{2.12}$$

as the joint distribution function. The interpretation of $p(x_1, t_1; x_2; t_2)$ is shown in Fig. 4, where $p(x_1, t_1; x_2; t_2)dx_1dx_2$ is the probability that the process X_t passes through windows of size dx_1 and dx_2 at times t_1 and t_2, respectively.

This can be easily generalized to multi-point joint probability density functions, i.e. $p(x_1, t_1; \ldots; x_n, t_n)$.

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

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \tag{2.13}$$

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

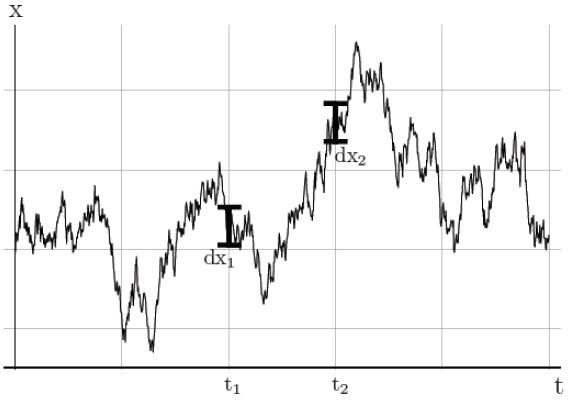


FIGURE 4. *Sketch of the meaning of the joint probability function* $p(x_1, t_1; x_2; t_2)$.

this concept to define a Markov-process for which the following property, the Markov property, holds

$$p(x_n, t_n; \ldots; x_1, t_1 | y_n, \tau_n; \ldots; y_1, \tau_1) = p(x_n, t_n; \ldots; x_1, t_1 | y_n, \tau_n) \tag{2.14}$$

for $t_n \geqslant \ldots \geqslant t_1 \geqslant \tau_n \geqslant \ldots \geqslant \tau_1$. 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 which 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 me 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 2.13)

$$\begin{aligned} & p\,(x_n, t_n; \ldots; x_1, t_1) = p(x_n, t_n; \ldots; x_2, t_2 | x_1, t_1) p(x_1, t_1) = \\ & = p\,(x_n, t_n; \ldots; x_3, t_3 | x_2, t_2; x_1, t_1) p(x_2, t_2 | x_1, t_n 1) p(x_1, t_1) = \\ & = p\,(x_n, t_n; \ldots; x_3, t_3 | x_2, t_2) p(x_2, t_2 | x_1, t_1) p(x_1, t_1) = \\ & = p\,(x_n, t_n | x_{n-1}, t_{n-1}) \ldots p(x_2, t_2 | x_1, t_1) p(x_1, t_1). \end{aligned}$$

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.

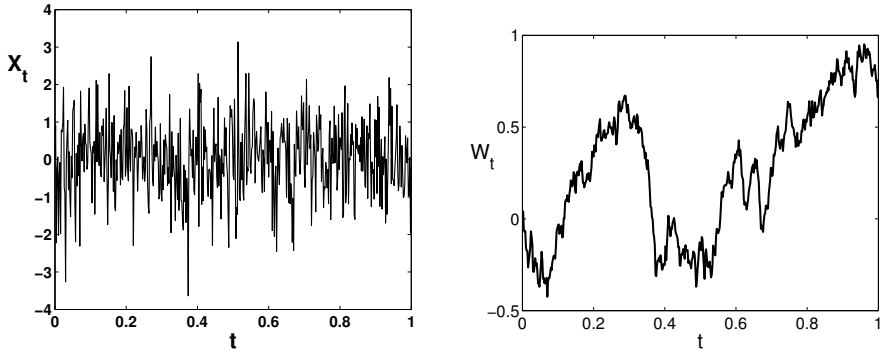
2.2. *Examples of random processes*

A Gaussian process (Fig. 5a) is defined over the interval $T = [0, 1]$ with $0 \leqslant t_1 \leqslant ... \leqslant t_n \leqslant 1$ such that all $X_{t_1}, ..., X_{t_n}$ are independent and standard normally distributed, i.e., 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

$$F_{\mathbf{X}}(\mathbf{x}) = \Phi(x_1) ... \Phi(x_n), \tag{2.15}$$



(a) (b)

FIGURE 5. *(a) A realization of a Gaussian process. (b) A realization of a Wiener process.*

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

$$\mu_X(t) = 0 \ ; \ t \neq s : c_X(t,s) = 0 \ ; \ \sigma_X^2(t) = 1. \tag{2.16}$$

The Wiener process $W_t, t \in [0,\infty)$ is a stochastic process (Fig. 5b) 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 $0 \leqslant s < t \leqslant T$

$$W_t - W_s \stackrel{d}{=} N(0, t-s) \tag{2.17a}$$

$$\mu_W(t) = 0 \tag{2.17b}$$

$$c_W(t,s) = s \tag{2.17c}$$

To obtain the result (2.17a), 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 (2.17a) follows then directly from property (iii). Equation (2.17b) follows directly from property (iii) and (2.17c) 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).

The Wiener process has transition probability

$$p(x,t|y,s) = \frac{1}{\sqrt{2\pi(t-s)}} e^{-\frac{1}{2}\frac{(x-y)^2}{t-s}} \tag{2.18}$$

for $t > s$ and $p(x,0) = \delta(x)$, and probability density function

$$p(x,t) = \int_{-\infty}^{\infty} p(x,t|y,0)p(y,0)dy = \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2}\frac{x^2}{t}} \tag{2.19}$$

From the probability density functions, it can also be easily shown that

$$E[W_t^{2n}] = \int_{-\infty}^{\infty} w^{2n} p(w,t)dw = \frac{(2n)!}{2^n n!} t^n \tag{2.20}$$

and that odd moments are zero.

To simulate the Wiener process W_t on a computer, we discretize time from t_0 to t_n and start with $W_0 = 0$. Give the current value W_{t_i} at $t = t_i$, we repeatedly draw the

next random value $W_{t_{i+1}}$ using the probability distribution $p(W_{t_{i+1}}, t_{i+1}|W_{t_i}, t_i)$. The increment $\Delta W = W_{t_{i+1}} - W_{t_i}$, it follows from (2.20) that $E[\Delta W^2] = \Delta t$ and hence this suggests that $\Delta W = \mathcal{O}(\sqrt{\Delta t})$. Although $\Delta W/\Delta t = \mathcal{O}(1/\sqrt{\Delta t})$ becomes unbounded when $\Delta t \to \infty$, one can always compute this quantity; this is what is called 'white noise' and is the quantity appearing as ξ in the mixed layer ocean model (2.3). Hence, more formally, $E[\xi(t)] = 0$, $E[\xi(t)\xi(s)] = \delta(t-s)$, and $\xi(t) = dW_t/dt$ (the latter has to be interpreted in a distribution sence).

2.3. *Stochastic calculus*

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 $0 = t_0 < t_1 < ... < t_{N-1} < t_N = T$. The Riemann integral of h is then given by

$$\int_0^T h(t)dt = \lim_{N\to\infty} \sum_{j=0}^{N-1} h(t_j)(t_{j+1} - t_j). \tag{2.21}$$

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

$$\int_0^T h(t)dW_t = \lim_{N\to\infty} \sum_{j=0}^{N-1} h(t_j)(W(t_{j+1}) - W(t_j)), \tag{2.22}$$

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

$$\int_0^T h(t) \circ dW_t = \lim_{N\to\infty} \sum_{j=0}^{N-1} h(\frac{t_j + t_{j+1}}{2})(W(t_{j+1}) - W(t_j)). \tag{2.23}$$

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 exist between these results and hence both definitions have their use.

We next consider the stochastic version of the main theorem of integral calculus,

$$f(b) - f(a) = \int_a^b f'(t)dt, \tag{2.24}$$

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

$$f(W_x + dW_x) - f(W_x) = f'(W_x)dW_x + \frac{1}{2}f''(W_x)(dW_x)^2 + \dots. \tag{2.25}$$

With $E[(dW_x)^2] = dx$, we obtain the first Itô Lemma by integration of (2.25) over the interval $[s,t]$, i.e.,

$$f(W_t) - f(W_s) = \int_s^t f'(W_x)dW_x + \int_s^t \frac{1}{2}f''(W_x)dx. \tag{2.26}$$

We see that (2.26) is the generalization of (2.24) to the stochastic case. In addition to the

first term in the right hand side, there is now an additional Riemann integral involving the second derivative of f.

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

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

where $f_1 = \partial f/\partial t$, $f_2 = \partial f/\partial y$, $f_{11} = \partial^2 f/\partial t^2$, etc. 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

$$f(t, W_t) - f(s, W_s) = \int_s^t \left[f_1(x, W_x) + \frac{1}{2}f_{22}(x, W_x)\right] dx + \int_s^t f_2(x, W_x)dW_x. \quad (2.27)$$

If $f = f(W_t)$, then $f_1 = 0, f_2 = f', f_{22} = f''$ and the second Itô Lemma (2.27) reduces to the first Itô Lemma (2.26).

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

$$X_t = X_0 + \int_0^t A^{(1)}(s, X_s)ds + \int_0^t A^{(2)}(s, X_s)dW_s. \quad (2.28)$$

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

$$f(t, X_t) - f(s, X_s) = \int_s^t \left[f_1(x, X_x) + A_x^{(1)}f_2(x, X_x) + \frac{1}{2}(A_x^{(2)})^2 f_{22}(x, X_x)\right] dx + \int_s^t A_x^{(2)}f_2(x, X_x)dW_x, \quad (2.29)$$

where $A_x^{(i)} = A^{(i)}(x, X_x)$.

2.4. *Stochastic differential equations*

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

$$\dot{x} = \frac{dx}{dt} = f(t, x) \to dx = f(t, x)dt. \quad (2.30)$$

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

$$x(t) = \int_0^t f(s, x)ds + x_0. \quad (2.31)$$

A general stochastic differential equation is written as

$$dX_t = a(t, X_t)dt + b(t, X_t)dW_t, \quad (2.32)$$

for smooth functions a, b and with initial condition X_0. The formal solution of (2.32) is given

$$X_t - X_0 = \int_0^t a(s, X_s)ds + \int_0^t b(s, X_s)dW_s. \quad (2.33)$$

Either form (2.32) or (2.33) is referred to as the Itô Stochastic Differential Equation (SDE). It can be deduced, see section 3.2.3 of Mikosch (2000), that for each Itô SDE

(2.32), there is an equivalent Stratonovich SDE of the form

$$dX_t = (a(t, X_t) - \frac{1}{2}b(t, X_t)\frac{\partial b}{\partial x}(t, X_t))dt + b(t, X_t) \circ dW_t. \tag{2.34}$$

As an example of the use of the Itô Lemma's, consider the Itô SDE

$$X_t - X_0 = \int_0^t (c_1(s)X_s + c_2(s))ds + \int_0^t \sigma_2(s)dW_s. \tag{2.35}$$

To solve this equation, we use 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 (2.29) is applied to Y_t to give, (with $f_1 = \alpha' x$, $f_2 = \alpha$ and $f_{22} = 0$)

$$\alpha(t)X_t - \alpha(0)X_0 = \int_0^t [\alpha'(x)X_x + (c_1(x)X_x + c_2(x))\alpha(x)]\, dx + \\ + \int_0^t \alpha(x)\sigma_2(x)dW_x.$$

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

$$X_t = \frac{1}{\alpha(t)}\left[X_0 + \int_0^t \alpha(x)c_2(x)dx + \int_0^t \alpha(x)\sigma_2(x)dW_x\right]. \tag{2.36}$$

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 (2.36) becomes

$$X_t = e^{-\gamma t}\left[X_0 + \sigma \int_0^t e^{\gamma x}\, dW_x\right]. \tag{2.37}$$

The stochastic process associated with this solution is called the Ornstein-Uhlenbeck process (Fig. 6a).

From this process, the covariance function is calculated (just by working out the integrals) as

$$c_{OU}(t, s) = E[X_t X_{t+s}] = \frac{\sigma^2}{2\gamma}(e^{-\gamma s} - e^{-\gamma(2t+s)}) \tag{2.38}$$

and the spectrum follows from the Fourier transform as

$$P(\omega) = \frac{\sigma^2}{\gamma^2 + \omega^2}$$

which is a Lorentz spectrum. These spectra (indicated there by AR(1) in Fig. 3a) are called red-noise spectra (the OU process is the red noise process) and they decay for large ω (compared to γ) as ω^{-2}.

2.5. *The Fokker-Planck equation*

We now return to the general Itô SDE (2.32) given by

$$X_t = X_0 + \int_0^t a(X_s, s)ds + \int_0^t b(X_s, s)dW_s \tag{2.39}$$

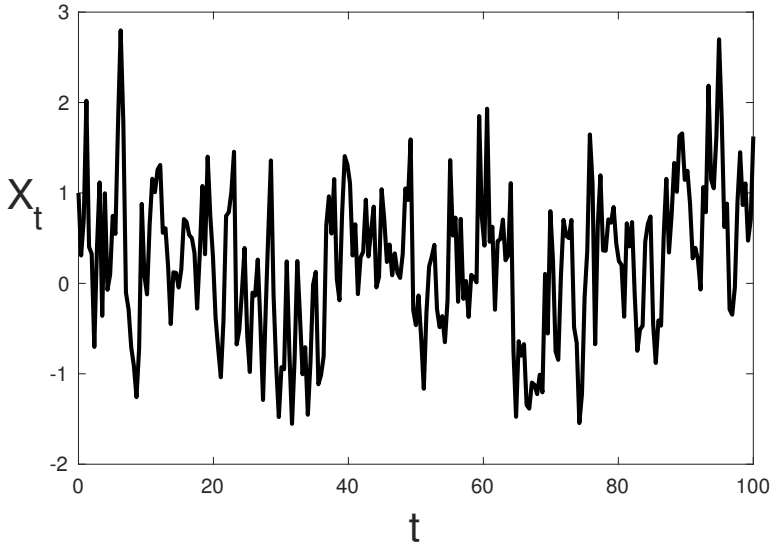


FIGURE 6. *A path of the Ornstein-Uhlenbeck process (2.37) with* $\gamma = 1$, $\sigma = 1$ *and* $X_0 = 1$.

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

$$E[f(X_t)] = \int_{-\infty}^{\infty} f(x)p(x,t)dx, \tag{2.40}$$

and hence

$$\frac{d}{dt}E[f(X_t)] = \int_{-\infty}^{\infty} f(x)\frac{\partial p}{\partial t}(x,t)dx. \tag{2.41}$$

On the other hand, when we use the third Itô Lemma for f, we find

$$\begin{aligned} f(X_t) - f(X_0) = &\int_0^t \left[a(X_s,s)\frac{\partial f}{\partial x}(X_s) + \frac{1}{2}b^2(X_s,s)\frac{\partial^2 f}{\partial^2 x}(X_s)\right] ds \\ &+ \int_0^t b(X_s,s)\frac{\partial f}{\partial x}(X_s)dW_s. \end{aligned} \tag{2.42}$$

Taking the expectation operator of (2.42), using (2.40) and differentiating the result to t then gives

$$\frac{d}{dt}E[f(X_t)] = \int_{-\infty}^{\infty} \left[a(x,t)\frac{\partial f}{\partial x}(x) + \frac{1}{2}b^2(x,t)\frac{\partial^2 f}{\partial^2 x}(x)\right] p(x,t;x_0,t_0)dx. \tag{2.43}$$

Combining (2.43) and (2.41), we find

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

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 finally gives

$$\int_{-\infty}^{\infty} f\ (\frac{\partial p}{\partial t} + \frac{\partial (ap)}{\partial x} - \frac{1}{2}\frac{\partial^2 (pb^2)}{\partial^2 x})\ dx = 0, \tag{2.44}$$

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

$$\frac{\partial p}{\partial t} + \frac{\partial (ap)}{\partial x} - \frac{1}{2}\frac{\partial^2 (pb^2)}{\partial^2 x} = 0 \tag{2.45}$$

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

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

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

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

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

We immediately conclude from this normal distribution that $\mu[X] = 0$ and that $Var[X] = \sigma^2/(2\gamma)$. The latter is also the limit for $t \to \infty$ in the variance $c_{OU}(t,0)$ in (2.38).

The multi-dimensional generalization of the Fokker-Planck equation of the SDE

$$d\mathbf{X}_t = \mathbf{f}(\mathbf{X_t}, t)dt + \mathbf{g}(\mathbf{X_t}, t)d\mathbf{W}_t \tag{2.46}$$

where $\mathbf{W}$ is a $d \times n$ vector of Wiener processes and $\mathbf{g}$ is usually an $n \times d$ matrix, is (in tensor notation)

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x_i}(f_i p) + \frac{1}{2}\frac{\partial^2}{\partial x_i \partial x_j}(D_{ij} p) \tag{2.47}$$

where $D_{ij} = g_{ik}g_{kj}$.

2.6. *Backward Kolmogorov equation*

For general stochastic processes, the following relations always holds

$$p(x_3, t_3; x_1, t_1) = \int p(x_3, t_3; x_2, t_2; x_1, t_1)dx_2 = p(x_3, t_3|x_1, t_1)p(x_1, t_1) \tag{2.48}$$

For a Markov process, we can write $p(x_3, t_3; x_1, t_1) = p(x_3, t_3|x_2, t_2)p(x_2, t_2|x_1, t_1)p(x_1, t_1)$ and hence combing this with (2.48) gives the Chapman-Kolmogorov equation

$$p(x_3, t_3|x_1, t_1) = \int p(x_3, t_3|x_2, t_2)p(x_2, t_2|x_1, t_1)dx_2 \tag{2.49}$$

Using $x = x_3, t = t_3, x_1 = y, t_1 = s$ and $x_2 = x', t_2 = t'$ and differentiating (2.49) to t' gives

$$0 = \int_{-\infty}^{\infty} \frac{\partial p(x,t|x',t')}{\partial t'} p(x',t'|y,s) + p(x,t|x',t')\frac{\partial p(x',t'|y,s)}{\partial t'}dt' \tag{2.50}$$

Using the Fokker-Planck equation in the second term gives

$$0 = \int_{-\infty}^{\infty} dx' \left[\frac{\partial p(x,t|x',t')}{\partial t'} p(x',t'|y,s) + p(x,t|x',t') \left(\frac{\partial}{\partial x'}(-f(x') + \frac{1}{2}\frac{\partial}{\partial x'} g^2(x')) p(x',t'|y,s) \right) \right] =$$
$$= \int_{-\infty}^{\infty} dx' p(x',t'|y,s) \left[\frac{\partial p(x,t|x',t')}{\partial t'} + f(x') \frac{\partial p(x,t|x',t')}{\partial x'} + \frac{1}{2} g^2(x') \frac{\partial^2 p(x,t|x',t')}{\partial^2 x'} \right] \tag{2.51}$$

Letting $t' \to s$ gives $p(x',t'|y,s) = \delta(x'-y)$ which gives the Backward-Kolmogorov Equation (BKE)

$$-\frac{\partial p(x,t|y,s)}{\partial s} = f(y) \frac{\partial p(x,t|y,s)}{\partial y} + \frac{1}{2} g^2(y) \frac{\partial^2 p(x,t|y,s)}{\partial^2 s} \tag{2.52}$$

We will use this equation in section 4 to compute exit times for a particular interval.

The operator in the right hand side of (2.52) is the formal adjoint of the Fokker-Planck operator and it often referred to as the generator and indicated by $\mathcal{K}$. The multi-dimensional extension of the BKE is in tensor notation given by

$$-\frac{\partial p}{\partial t} = f_i \frac{\partial p}{\partial x_i} + \frac{D_{ij}}{2} \frac{\partial^2 p}{\partial x_i \partial x_j} \tag{2.53}$$

where again $D_{ij} = g_{ik} g_{kj}$.

2.7. *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

$$X(t) = X(0) + \int_0^t f(X(s)) ds + \int_0^t g(X(s)) dW(s), \tag{2.54}$$

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 $X(\tau_j)$.

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

$$E[|X_k - X(\tau_k)|] \leqslant (\Delta t)^\eta. \tag{2.55}$$

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

$$|E[X_k] - E[X(\tau_k)]| \leqslant (\Delta t)^\eta. \tag{2.56}$$

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

The Euler-Maruyama scheme for (2.54) is

$$X_j - X_{j-1} = f(X_{j-1})\Delta t + g(X_{j-1})(W(\tau_j) - W(\tau_{j-1})). \tag{2.57}$$

It turns out that the order of strong convergence is only $\eta = 1/2$ and of weak convergence is $\eta = 1$. 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 (2.54). We first write the discretization as

$$X_{t_j} - X_{t_{j-1}} = \int_{t_{j-1}}^{t_j} f(X_s)ds + \int_{t_{j-1}}^{t_j} g(X_s)dW_s, \tag{2.58}$$

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

$$\int_{t_{j-1}}^{t_j} f(X_s)ds = f(X_{t_{j-1}})\Delta_j t \tag{2.59a}$$

$$\int_{t_{j-1}}^{t_j} g(X_s)dW_s = g(X_{t_{j-1}})\Delta_j W. \tag{2.59b}$$

The crucial step in the derivation of higher order schemes is the application of the third Itô Lemma for a function $f(x)$, with $f_1 = 0, f_2 = f', f_{22} = f''$, while $A^{(1)} = f$ and $A^{(2)} = g$, according to (2.58). We then find

$$f(X_s) - f(X_{t_{j-1}}) = \int_{t_{j-1}}^{s} \left[ff' + \frac{1}{2}g^2 f''\right] dy + \int_{t_{j-1}}^{s} gf'dW_y, \tag{2.60}$$

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

$$g(X_s) - g(X_{t_{j-1}}) = \int_{t_{j-1}}^{s} \left[fg' + \frac{1}{2}g^2 g''\right] dy + \int_{t_{j-1}}^{s} gg'dW_y. \tag{2.61}$$

Next we substitute the last two expressions into (2.58) and obtain

$$X_{t_j} - X_{t_{j-1}} = f(X_{t_{j-1}})\Delta_j t + g(X_{t_{j-1}})\Delta_j W + R_j^1 + R_j^2 \tag{2.62}$$

$$R_j^1 = \int_{t_{j-1}}^{t_j} \int_{t_{j-1}}^{s} gg'dW_y dW_s, \tag{2.63}$$

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

$$R_j^1 = \frac{1}{2}g(X_{t_{j-1}})g'(X_{t_{j-1}})((\Delta_j W)^2 - \Delta_j t), \tag{2.64}$$

to finally give the Milstein scheme

$$X_{t_j} - X_{t_{j-1}} = f(X_{t_{j-1}})\Delta_j t + g(X_{t_{j-1}})\Delta_j W + \frac{1}{2}g(X_{t_{j-1}})g'(X_{t_{j-1}})((\Delta_j W)^2 - \Delta_j t). \tag{2.65}$$