STOCHASTIC POPULATION MODELS (SPRING 2015)

STEFAN GERITZ & FENGYING WEI

5. Stochastic differential equations

5.1. **The normal distribution.** The *normal distribution* (or Gaussian distribution) with mean μ and variance σ^2 is the probability distribution on all the real numbers with the probability density

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

To denote that a random variable X is normally distributed with mean μ and variance σ^2, we write $X \sim \mathcal{N}(\mu, \sigma^2)$. One readily verifies that

$$\mathcal{E}\{X\} := \int_{-\infty}^{+\infty} x\, p_{\mu,\sigma^2}(x) dx = \mu \tag{2}$$

and

$$\mathcal{E}\{(X-\mu)^2\} := \int_{-\infty}^{+\infty} (x-\mu)^2\, p_{\mu,\sigma^2}(x) dx = \sigma^2 \tag{3}$$

are indeed the mean and the variance of the distribution. If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$\frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1) \tag{4}$$

which is called the *standard normal distribution* with the probability density

$$p_{0,1}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \tag{5}$$

We have seen that there exists a one-to-one relationship between a function and its Fourier transform. The Fourier transform of a probability density is called the *characteristic function* of the distribution. The characteristic function of the standard normal distribution is

$$\tilde{p}_{0,1}(\omega) = e^{-\frac{1}{2}\omega^2} \tag{6}$$

Below we shall see why statisticians and probability theorists are so obsessed with the normal distribution.

5.2. **The Central Limit Theorem.** The Central Limit Theorem states that if $X_1, X_2, X_3, \dots$ are independent and identically distributed ramdom variables with finite mean μ and finite variance $\sigma^2 > 0$, then

$$\lim_{n\to\infty} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma\sqrt{n}} \sim \mathcal{N}(0,1) \tag{7}$$

The amazing thing is that this conclusion does not depend on what the probability distribution of the X_i actually looks like.

We prove the Central Limit Theorem as follows. First define

$$Z_n := \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma\sqrt{n}} \tag{8}$$

and show that the probability density of Z_n converges to the the probability density of the standard normal distribution as $n \to \infty$, i.e., we prove *convergence in distribution*. It is easier, however, to show that the characteristic function of Z_n converges to the characteristic function of the standard normal distribution, which is the same thing.

To calculate the characteristic function of Z_n, first define

$$Y_i := \frac{X_i - \mu}{\sigma} \quad (i = 1, 2, \dots) \tag{9}$$

The Y_i then are independently and identically distributed random variables with zero mean and unit variance and with a probability density $p_Y(y)$ (which can be calculated from the density of the X_i, but we will not need that). The characteristic function of the Y_i is

$$\begin{aligned} \tilde{p}_Y(\omega) &= \textstyle\int_{-\infty}^{+\infty} p_Y(y)\, e^{-i\omega y} dy \\ &= \textstyle\int_{-\infty}^{+\infty} p_Y(y)\big[1 - i\omega y - \frac{1}{2}\omega^2 y^2 + \mathrm{O}(\omega^3)\big] dy \\ &= 1 - i\omega\mathcal{E}\{Y\} - \tfrac{1}{2}\omega^2\mathcal{E}\{Y^2\} + \mathrm{O}(\omega^3) \\ &= 1 - \tfrac{1}{2}\omega^2 + \mathrm{O}(\omega^3) \end{aligned} \tag{10}$$

Consequently, the characteristic function of the $Y_i/\sqrt{n}$ is

$$\tilde{p}_{\frac{Y}{\sqrt{n}}}(\omega) = \tilde{p}_Y\left(\frac{\omega}{\sqrt{n}}\right) = 1 - \frac{\omega^2}{2n} + \mathrm{O}\left(\frac{\omega^3}{n^{3/2}}\right) \tag{11}$$

Since the probability density of the sum of independent random variables is the convolution of the respective probability densities, the characteristic function of the sum of independent random variables is product of the respective characteristic functions. Applying this to

$$Z_n = \sum_{i=1}^{n} \frac{Y_i}{\sqrt{n}} \tag{12}$$

we find that the characteristic function of Z_n is

$$\tilde{p}_{Z_n}(\omega) = \left(1 - \frac{\omega^2}{2n} + \mathrm{O}\left(\frac{\omega^3}{n^{3/2}}\right)\right)^n \tag{13}$$

and hence

$$\lim_{n\to\infty} \tilde{p}_{Z_n}(\omega) = e^{-\frac{1}{2}\omega^2} \tag{14}$$

which is the characteristic function of the standard normal distribution (see Section 5.1). So, the characteristic function of Z_n converges to that of the standard normal distribution, which completes the proof of the Central Limit Theorem.

There exist more general versions of the Central Limit Theorem that allow for the X_i to have different distributions.

Since many real-world quantities (such as the landing place of a seed on the ground, the position of a dust particle after a given amount time, etc.) result from the *additive* effect of many unobserved random events, the Central Limit Theorem provides an explanation of the prevalence of the normal distribution in real life.

Likewise, many other real-world quantities (such as the life span of an individual) results from the *multiplicative* effect of a large number of unobserved random events. The *logarithm* of those quantities, therefore, will be approximately normally distributed.

5.3. **The Wiener process.** The Wiener process $\{W(t)\}_{t\geq 0}$ is a continuous-time stochastic process on the real numbers and is characterized by its increments $W(t) - W(s)$ for $t > s$ such that increments over non-overlapping intervals of the same length are independently and identically distributed with zero mean and finite variance. The consequences of this characterization are studied below.

For every regular partition $s = t_0 < t_1 < \cdots < t_n = t$ of the interval (s, t) we have

$$W(t) - W(s) = \sum_{i=1}^{n} \big(W(t_i) - W(t_{i-1})\big) \tag{15}$$

which is a sum of n independently and identically distributed random variables. The partition may be arbitrarily fine by choosing n sufficiently large, and so it follows from the Central Limit Theorem that $W(t) - W(s)$ is normally distributed.

Moreover, as the variance of the sum of independent random variables is equal to the sum of the variances, it follows that the variance of the increment $W(t)-W(s)$ must be proportional to the length of the time interval (s,t). By an appropriate scaling of time, we can take the constant of proportionality to be equal to one, and so the variance of $W(t)-W(s)$ is $t-s$. Thus we conclude that

$$W(t)-W(s)\sim\mathcal{N}(0,t-s) \tag{16}$$

for every $t>s\geq 0$. In particular, if we fix the initial condition $W(0)=0$ and take $s=0$, then

$$W(t)\sim\mathcal{N}(0,t) \tag{17}$$

with

$$\mathcal{E}\{W(t)\}=0 \tag{18}$$

$$\mathcal{E}\{W(t)^2\}=t \tag{19}$$

$$\mathcal{E}\{W(t)W(s)\}=\min\{t,s\} \tag{20}$$

for the mean, variance and auto-covariance of the process.

Realizations of the Wiener process are continuous. This can be understood from

$$W(t+\Delta t)-W(t)\sim\mathcal{N}(0,\Delta t) \tag{21}$$

which for $\Delta t\to 0$ converges to the Dirac delta distribution with all probability mass concentrated at zero, and hence

$$\operatorname{Prob}\left\{\lim_{\Delta t\to 0}|W(t+\Delta t)-W(t)|=0\right\}=1 \tag{22}$$

On the other hand, realizations of the Wiener process are nowhere differentiable, at least not in a similar sense as the process is continuous, because

$$\frac{W(t+\Delta t)-W(t)}{\Delta t}\sim\mathcal{N}\left(0,\frac{1}{\Delta t}\right) \tag{23}$$

which has a divergent variance in the limit $\Delta t\to 0$.

5.4. **Linear stochastic differential equations (SDE).** Consider a stochastic process $\{X(t)\}_{t\geq t_0}$ with $X(t_0)=x_0$ (a.s.) satisfying

$$\mathrm{d}X(t)=-a\,X(t)\mathrm{d}t+b\,\mathrm{d}W(t)\quad(a>0,\ t\geq t_0) \tag{24}$$

where $\mathrm{d}X(t):=X(t+\mathrm{d}t)-X(t)$ and $\mathrm{d}W(t):=W(t+\mathrm{d}t)-W(t)$ for an infinitesimally small time step $\mathrm{d}t$. The $\mathrm{d}W(t)$ is called the *infinitesimal Wiener increment*, and the equation is called a *stochastic differential equation* (SDE); in particular a *linear* stochastic differential equation, because the right hand side is linear in both X and W. We now define what the equation actually means.

Formal integration of equation (24) (i.e., acting as if the equation is an ordinary differential equation) leads to

$$X(t) = x_0 e^{-a(t-t_0)} + b\, e^{-a(t-t_0)} \int_{t_0}^{t} e^{a\tau} dW \tag{25}$$

What to make of the integral on the right?

Let $t_0 < t_1 < \cdots < t_n = t$ be a *regular partition* of the interval $(0, t)$. Then we define for any integrable function $f : \mathbb{R}_+ \to \mathbb{R}$

$$\int_{t_0}^{t} f(\tau) \mathrm{d}W := \lim_{n\to\infty} \sum_{i=1}^{n} f(\tau_i)(W(t_i) - W(t_{i-1})) \tag{26}$$

where each $\tau_i \in (t_{i-1}, t_i)$. The sum on the right is a linear combination of n independent $\mathcal{N}(0, t_i - t_{i-1})$ distributed random variables and therefore is itself normally distributed with zero mean and variance

$$\sum_{i=1}^{n} f(\tau_i)^2 (t_i - t_{i-1}) \longrightarrow \int_{t_0}^{t} f(\tau)^2 \mathrm{d}\tau \quad \text{as } n \to \infty. \tag{27}$$

Hence, we get

$$\int_{t_0}^{t} f(\tau) \mathrm{d}W \sim \mathcal{N}\left(0, \int_{t_0}^{t} f(\tau)^2 \mathrm{d}\tau\right) \tag{28}$$

In particular, in equation (25),

$$\int_{t_0}^{t} e^{a\tau} dW \sim \mathcal{N}\left(0, \frac{e^{2a(t-t_0)} - 1}{2a}\right) \tag{29}$$

and so

$$X(t) \sim \mathcal{N}\left(x_0 e^{-a(t-t_0)}, \frac{b^2}{2a}\left(1 - e^{-2a(t-t_0)}\right)\right) \tag{30}$$

Since $a > 0$, it follows that, asymptotically,

$$\lim_{t\to\infty} X(t) \sim \mathcal{N}\left(0, \frac{b^2}{2a}\right) \tag{31}$$

The stochastic differential equation (24) with initial condition $X(t_0) = x_0$ (a.s.) uniquely defines the stochastic process $\{X(t)\}_{t\geq t_0}$. This process is called the *Ornstein-Uhlenbeck process.* Note, however, that while the process is uniquely defined, there are infinitely many different realizations of the process, called *sample paths* or *stochastic orbits.*

5.5. **White noise.** Consider the linear equation (24) for $b \to 1$ and $a \to 0$, which then becomes

$$\mathrm{d}X(t) = \mathrm{d}W(t) \tag{32}$$

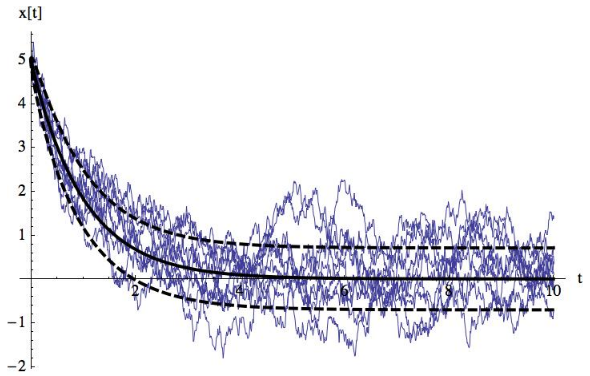

FIGURE 1. Ten sample paths of the Ornstein-Uhlenbeck process (25) with $X(0) = 5$ and $a = b = 1$. Solid line indicates the mean and the dashed lines the mean plus/minus the standard deviation of the distribution of $X(t)$.

and given the initial condition $X(t_0) = x_0$ (a.s.), the solution (30) becomes

$$X(t) \sim \mathcal{N}(x_0, t - t_0) \tag{33}$$

For arbitrary $t > s \geq t_0$ we thus have

$$X(t) - X(s) \sim \mathcal{N}(0, t - s) \tag{34}$$

independently of the initial condition at time t_0. We conclude that any two increments $X(t_1) - X(s_1)$ and $X(t_2) - X(s_2)$ over non-overlapping bounded intervals of equal length are identically and independently distributed with zero mean and finite variance. As these are the defining characteristics of the Wiener process (see section 5.3), it follows that $\{X(t)\}$ in fact *is* a Wiener process.

We thus find that, after having shown in section 5.3 that the Wiener process is nowhere differentiable, it nonetheless can be described by a differential equation, or more precisely, by a *stochastic* differential equation. The Wiener process is, in some sense, differentiable after all! The derivative of the Wiener process is called *white noise*, which is a stochastic process itself.

Let $\{\xi(t)\}$ be the process we just called white noise. Then (32) can be rewritten as

$$\mathrm{d}W(t) = \xi(t)\mathrm{d}t \tag{35}$$

White noise is a stochastic process where all the $\xi(t)$ are independently and identically distributed with

$$\mathcal{E}\{\xi(t)\} = 0 \tag{36}$$

and

$$\mathcal{E}\{\xi(t)\xi(s)\} = \delta(t - s) \tag{37}$$

for the mean and the auto-covariance, where δ denotes the Dirac delta distribution. In particular, the white noise has infinite variance. White noise obviously is a weird process and begins to make sense only after integration, which gives the Wiener process. A similar situation we find in the case of the Dirac delta distribution, which can be interpreted as the derivative of the unit step function at zero, exactly where the function in the classical sense is non-differentiable.

5.6. **Non-linear stochastic differential equations.** Most of the time we shall be dealing with *linear* SDEs. Still it is important to know something of non-linear SDEs as well, especially those SDEs with *multiplicative noise* (i.e., with a non-linear noise term).

Consider the non-linear SDE with multiplicative noise

$$\mathrm{d}X = W\mathrm{d}W \tag{38}$$

which can formally be solved as

$$X(t) = \int_{t_0}^{t} W\mathrm{d}W \tag{39}$$

But what to make of this integral? What does it mean? Lets proceed as in the previous section, i.e., let $t_0 < t_1 < \cdots < t_n = t$ be a *regular partition* of the interval $(0, t)$, and define

$$\int_{t_0}^{t} W\mathrm{d}W := \lim_{n\to\infty} \sum_{i=1}^{n} W(\tau_i)(W(t_i) - W(t_{i-1})) \tag{40}$$

where (and now we want to be explicit about this)

$$\tau_i = (1-\alpha)t_{i-1} + \alpha t_i \tag{41}$$

for some fixed $\alpha \in [0, 1]$. One wouldn't like the definition of the integral to depend on the particular choice of *alpha*, but unfortunately it will. To see this we take the expectation of the sum on the right side and use the property of the Wiener process that

$$\mathcal{E}\{W(t)W(s)\} = \min\{t, s\} \tag{42}$$

Since $t_{i-1} \le \tau_i \le t_i$, this gives

$$\begin{aligned} &\textstyle\sum_{i=1}^{n} \Big(\mathcal{E}\{W(\tau_i)W(t_i)\} - \mathcal{E}\{W(\tau_i)W(t_{i-1})\}\Big) \\ &= \textstyle\sum_{i=1}^{n} (\tau_i - t_{i-1}) \\ &= \alpha(t - t_0) \end{aligned} \tag{43}$$

Hence we find

$$\mathcal{E}\left\{\int_{t_0}^{t} W\mathrm{d}W\right\} = \alpha t \tag{44}$$

which depends on the particular choice of α. But then also the solution (39) of the SDE (38) will depend on the choice of α.

Now consider the more general non-linear SDE with multiplicative noise

$$\mathrm{d}X = f(X)\mathrm{d}t + g(X)\mathrm{d}W \tag{45}$$

Formally we can define a solution of this equation as any stochastic process $\{X(t)\}_{t\geq 0}$ that satisfies the integral equation

$$X(t) = \int_{t_0}^{t} f(X)\mathrm{d}\tau + \int_{t_0}^{t} g(X)\mathrm{d}W \tag{46}$$

The first integral will not give any problem, but the definition of the second integral as a limit of a sum over a partition requires an extra rule that specifies at what points the integrant is to be sampled. Without such integration rule, the SDE is not well-defined and not interpretable.

Although there are infinitely many possibilities for the α, in practice only two are being used: $\alpha = 0$ and $\alpha = 1/2$. The first leads to the so-called *Ito calculus*, and the second to the *Stratonovitch calculus*. Without proof we state that the following SDEs together with their respective integration rules as indicated in parentheses have the same solutions, i.e., describe the same stochastic process:

$$\mathrm{d}X = f(X)\mathrm{d}t + g(X)\mathrm{d}W \quad \text{(S)} \tag{47}$$

$$\mathrm{d}X = \Big(f(X) + \frac{1}{2}g'(X)g(X)\Big)\mathrm{d}t + g(X)\mathrm{d}W \quad \text{(I)} \tag{48}$$

Each of these two descriptions has its own advantages and disadvantages. We state without proof that in the Stratonovitch calculus we can integrate and differentiate as if the stochastic processes were ordinary functions, whereas in the Ito calculus the calculation of expectations as well as the numerical integration of the SDE are easier. Depending on what particular information we want to get out of a given SDE, we may prefer either the Ito or the Stratonovitch representation. This will be made clear with the following section.

Note from equations (47) and (**??**) that if g is the constant function (i.e., $g'(x) = 0$ for all x), then the Ito interpretation and the Stratonovitch interpretation make no difference at all. That is why in the case of a linear stochastic stochastic differential equation we do not need to specify the integration rule (see section 5.4).

5.7. **Example.** Consider the non-linear SDE

$$\begin{cases} \mathrm{d}X &= X\mathrm{d}W \quad (\mathrm{S}) \\ X(0) &= x_0 \end{cases} \tag{49}$$

As indicated between the parentheses, the equation has to be integrated according to the Stratonovich rule which follows the ordinary rules of calculus. This gives

$$X(t) = x_0 E^{W(t)} \tag{50}$$

In other words, $X(t)$ has a log-normal distribution, i.e., $\log X(t) \sim \mathcal{N}(\log x_0, t)$.

Next, consider the SDE

$$\begin{cases} \mathrm{d}X &= X\mathrm{d}W \quad (\mathrm{I}) \\ X(0) &= x_0 \end{cases} \tag{51}$$

which looks the same but which has to be integrated according to a different rule, namely the Ito rule. Easiest way to solve the equation is first to transform it into a Stratonovitch equation. A comparison with the Ito equation (48) shows that then we should take $g(X) = X$ and $f(X) = -\frac{1}{2}X$. Substituting these functions into the Stratonovitch equation (47) gives

$$\begin{cases} \mathrm{d}X &= -\frac{1}{2}X\mathrm{d}t + X\mathrm{d}W \quad (\mathrm{S}) \\ X(0) &= x_0 \end{cases} \tag{52}$$

Equations (51) and (52) have the same solutions, but the latter can be integrated using the ordinary rules of calculus, which gives

$$X(t) = x_0 e^{-\frac{t}{2} + W(t)} \tag{53}$$

So, $X(t)$ gain has a log-normal distribution, but with different parameters, i.e., $\log X(t) \sim \mathcal{N}(\log x_0 - \frac{1}{2}t, t)$.

It is obvious why one would like to have an SDE in the Stratonovitch form, because then we can use the normal rules of calculus. But what is the Ito form good for? To begin with, the Ito SDE is good to calculate expected values. Remember from the definition (40) that for the Ito integral the integrant is sampled *at the beginning* of each interval and therefore is independent of the Wiener increment over that interval. Thus, taking expectations in the Ito equation (51) and using that the expectation of the product of two independent random variables is equal to the product of their expectations, we get

$$\begin{cases} \mathrm{d}\mathcal{E}\{X\} &= \mathcal{E}\{X\mathrm{d}W\} \\ &= \mathcal{E}\{X\}\mathcal{E}\{\mathrm{d}W\} \\ &= \mathcal{E}\{X\} \cdot 0 \\ &= 0 \\ \mathcal{E}\{X(0)\} &= x_0 \end{cases} \tag{54}$$

So, the expectation of $X(t)$ stays constant in time, i.e.,

$$\mathcal{E}\{X(t)\} = x_0 \tag{55}$$

How is this for the Stratonovitch equation (49)? Remember that in the definition of the Stratonovitch integral, the integrant is sampled *in the middle* of each interval of the partition, and so the integrant and the Wiener increment over that interval are not independent. To calculate the expectation of the process defined by the Stratonovitch equation (49) we must first put it in the Ito form in order to be able to exploit that the expectation of $X\mathrm{d}W$ is equal to the product of expectations of X and $\mathrm{d}W$. From equations (47) and (48) we see that the Stratonovitch equation (49) is equivalent to the Ito equation

$$\begin{cases} \mathrm{d}X &= \frac{1}{2}X\mathrm{d}t + X\mathrm{d}W \quad \text{(I)} \\ X(0) &= x_0 \end{cases} \tag{56}$$

Taking expectations we get

$$\begin{cases} \mathrm{d}\mathcal{E}\{X\} &= \frac{1}{2}\mathcal{E}\{X\}\mathrm{d}t + \mathcal{E}\{X\mathrm{d}W\} \\ &= \frac{1}{2}\mathcal{E}\{X\}\mathrm{d}t + \mathcal{E}\{X\}\mathcal{E}\{\mathrm{d}W\} \\ &= \frac{1}{2}\mathcal{E}\{X\}\mathrm{d}t + \mathcal{E}\{X\}\cdot 0 \\ &= \frac{1}{2}\mathcal{E}\{X\}\mathrm{d}t \\ \mathcal{E}\{X(0)\} &= x_0 \end{cases} \tag{57}$$

and so

$$\mathcal{E}\{X(t)\} = x_0 e^{\frac{1}{2}t} \tag{58}$$

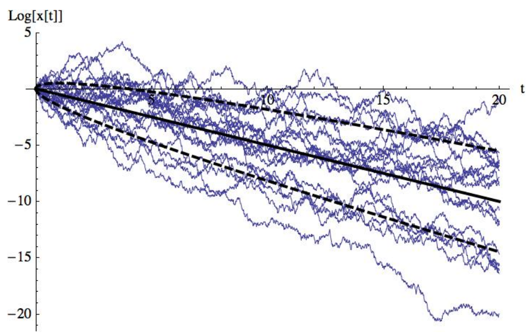

FIGURE 2. Numerical integration of the Ito equation $\mathrm{d}X = X\mathrm{d}W$. Twenty sample paths of $\log X(t)$ versus t with initial condition $\log X(0) = 0$. Solid line indicates the mean and the dashed lines the mean plus/minus the standard deviation of the distribution of $\log X(t)$.

5.8. **Numerical integration of SDEs.** The Stratonovitch equation we can solve using de rules of ordinary calculus. The Ito equation is handy when it comes to calculating expectations, because the integrant and the Wiener increment are probabilistically independent. Another advantage of the Ito form is that it suggest a method for solving the SDE numerically.

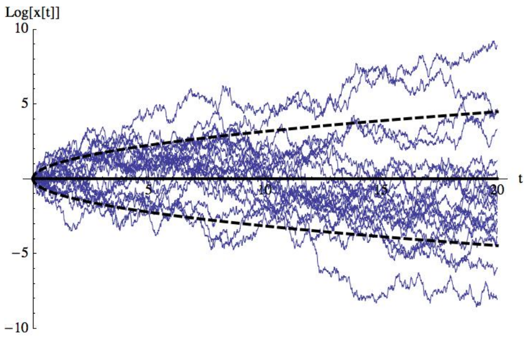

FIGURE 3. Numerical integration of the Stratonovich equation $\mathrm{d}X = X\mathrm{d}W$, which was solved by numerically integrating the equivalent Ito equation $\mathrm{d}X = \frac{1}{2}X\mathrm{d}t + X\mathrm{d}W$. Twenty sample paths of $\log X(t)$ versus t with initial condition $\log X(0) = 0$. Solid line indicates the mean and the dashed lines the mean plus/minus the standard deviation of the distribution of $\log X(t)$.

Consider the Ito equation

$$\mathrm{d}X = h(X)\mathrm{d}t + g(X)\mathrm{d}W \quad \text{(I)} \tag{59}$$

For small $\Delta t > 0$ we can approximate this by

$$\Delta X(t) = h(X(t))\Delta t + g(X(t))\Delta W(t) \tag{60}$$

where

$$\begin{aligned} \Delta X(t) &:= X(t + \Delta t) - X(t) \\ \Delta W(t) &:= W(t + \Delta t) - W(t) \\ \Delta W(t) &\sim \mathcal{N}(0, \Delta t) \end{aligned} \tag{61}$$

A similar discretization of the Stratonovitch equation

$$\mathrm{d}X = h(X)\mathrm{d}t + g(X)\mathrm{d}W \quad \text{(S)} \tag{62}$$

would give

$$\Delta X(t) = h(X(t + \frac{1}{2}\Delta t))\Delta t + g(X(t + \frac{1}{2}\Delta t))\Delta W(t) \tag{63}$$

which, if we know X only up to and including time t (but not further), is not practical.

So, let's go back to the discretization of the Ito equation, which we can rewrite as

$$\begin{aligned} X(t + \Delta t) &= X(t) + h(X(t))\Delta t + g(X(t))\sqrt{\Delta t}\, Z(t) \\ Z(t) &\sim \mathcal{N}(0, 1) \text{ and i.i.d. for all } t \geq 0 \end{aligned} \tag{64}$$

Most program packages contain a random number generator for the standard normal distribution $\mathcal{N}(0,1)$. Numerical iteration of the discretization of the Ito equation n times gives us an approximation of a *sample path* (or *realization*) of the stochastic process $\{X(t)\}_{t\geq 0}$ for the times $t = 0, \Delta t, 2\Delta t, 3\Delta t, \ldots, n\Delta t$.

5.9. **SDEs and the Fokker-Planck equation** There are at least two ways to describe a stochastic process $\{X(t)\}$: (*i*) one can describe the evolution of individual realizations (i.e., sample paths) of the process – this is in essence what is done by a stochastic differential equation. (*ii*) One can also describe how the probability distribution $p(x,t)$ of the random variable $X(t)$ changes over time – this is done by the so-called Fokker-Planck equation. It should not come as a surprise that what the Fokker-Planck equation, too, depends on the sampling oprocedure, i.e., on $\alpha \in [0,1]$:

$$\partial_t p(x,t) = -\partial_x\Big(\big[f(x) + \alpha g'(x)g(x)\big]p(x,t)\Big) + \frac{1}{2}\partial_{xx}\Big(g(x)^2 p(x,t)\Big) \tag{65}$$

This we give without proof. Hence, the Ito differential equation

$$\mathrm{d}X = f(X)\mathrm{d}t + g(X)\mathrm{d}W \quad \text{(I)} \tag{66}$$

with $\alpha = 0$ corresponds to the Fokker-Planck equation

$$\partial_t p(x,t) = -\partial_x\big(f(x)p(x,t)\big) + \frac{1}{2}\partial_{xx}\big(g(x)^2 p(x,t)\big) \tag{67}$$

and the Stratonovitch equation

$$\mathrm{d}X = f(X)\mathrm{d}t + g(X)\mathrm{d}W \quad \text{(S)} \tag{68}$$

with $\alpha = 1/2$ corresponds to the Fokker-Planck equation

$$\partial_t p(x,t) = -\partial_x\big(\big[f(x) + \frac{1}{2}g'(x)g(x)\big]p(x,t)\big) + \frac{1}{2}\partial_{xx}\big(g(x)^2 p(x,t)\big) \tag{69}$$

The Focker-Planck equation contains the same information as the corresponding SDE – they are equivalent representations of the same stochastic process.

In particular, the linear equation

$$\mathrm{d}X = 2D\mathrm{d}W \tag{70}$$

(no need to specify whether it is Ito or Stratonovics; see last paragraph of section 5.6) is equivalent to the so-called *heat equation*

$$\partial_t p(x,t) = D\partial_{xx}p(x,t) \tag{71}$$

with diffusion coefficient $D > 0$.