STOCHASTIC POPULATION MODELS
(SPRING 2011)

STEFAN GERITZ
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF HELSINKI

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 $X_1, X_2, X_3, \ldots$ 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

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

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

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

On the other hand, realizations of the Wiener process are nowhere differentiable, at least not in the sense 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. **The Gaussian white noise.** The Gaussian white noise is a stochastic process $\{\xi(t)\}_{t \geq 0}$ where all the $\xi(t)$ are independently and identically distributed with

$$\mathcal{E}\{\xi(t)\} = 0 \tag{24}$$

and

$$\mathcal{E}\{\xi(t)\xi(s)\} = \delta(t - s) \tag{25}$$

for the mean and the auto-covariance (where δ is the Dirac delta distribution). In particular, the white noise has infinite variance.

The white noise obviously is a weird process. To make sense of it, define

$$X(t) := \int_0^t \xi(\tau) d\tau \tag{26}$$

Explicit calculation shows that $\{X(t)\}_{t \geq 0}$ has all the properties of the Wiener process and therefore is the Wiener process. In particular, we have

$$\begin{aligned} \mathcal{E}\left\{(X(t) - X(s))^2\right\} &= \textstyle\int_s^t \int_s^t \mathcal{E}\{\xi(\tau')\xi(\tau'')\} d\tau'' d\tau' \\ &= \textstyle\int_s^t \int_s^t \delta(\tau' - \tau'') d\tau'' d\tau' \\ &= \textstyle\int_s^t 1 d\tau' \\ &= t - s \end{aligned} \tag{27}$$

Differentiation of expression (26) gives the stochastic differential equation (SDE)

$$\begin{cases} dX(t) = \xi(t)dt \\ \\ X(0) = 0 \text{ a.s.} \end{cases} \tag{28}$$

The solution of this equation (as we have just seen) is the Wiener process:

$$X(t) = W(t) \tag{29}$$

The white noise thus can be seen (or meaningfully defined) as the derivative of the Wiener process. We usually write dW instead of ξdt, called the *infinitesimal Wiener increment*, i.e., the Wiener increment $W(t+dt) - W(t)$ over an interval of infinitesimal dt length.

5.5. **Linear stochastic differential equations.** Consider the linear SDE

$$\begin{cases} dX &= -aX\,dt + b\,dW \\ X(0) &= x_0 \end{cases} \tag{30}$$

which defines the *Ornstein-Uhlenbeck process*. Formal integration leads to

$$X(t) = x_0 e^{-at} + be^{-at} \int_0^t e^{a\tau} dW \tag{31}$$

What to make of the integral on the right?

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

$$\int_0^t f(\tau)dW = \lim_{n\to\infty} \sum_{i=1}^n f(\tau_i)(W(t_i) - W(t_{i-1})) \tag{32}$$

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

$$\sum_{i=1}^n f(\tau)^2 (t_i - t_{i-1}) \longrightarrow \int_0^t f(\tau)^2 d\tau \tag{33}$$

as $n \to \infty$. Hence we get

$$\int_0^t f(\tau)dW \sim \mathcal{N}\left(0, \int_0^t f(\tau)^2 d\tau\right) \tag{34}$$

In particular, in equation (31),

$$\int_0^t e^{a\tau} dW \sim \mathcal{N}\left(0, \frac{e^{2at} - 1}{2a}\right) \tag{35}$$

and so

$$X(t) \sim \mathcal{N}\left(x_0 e^{-at}, \frac{b^2}{2a}\left(1 - e^{-2at}\right)\right) \tag{36}$$

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

$$\lim_{t\to\infty} X(t) \sim \mathcal{N}\left(0, \frac{b^2}{2a}\right) \tag{37}$$

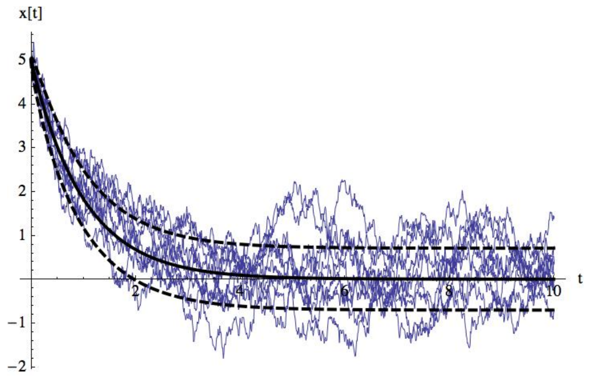

FIGURE 1. Ten sample paths of the Ornstein-Uhlenbeck process (31) with $X(0) = 5$ and $a = b = 1$. Solid line indicates the mean and the dashed lines the mean plus/minus the standard deviation of the distribution of $X(t)$.

5.6. **Ito and Stratonovitch.** 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

$$dX = WdW \tag{38}$$

which can solved as

$$X(t) = \int_0^t WdW \tag{39}$$

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

$$\int_0^t WdW = \lim_{n\to\infty} \sum_{i=1}^{n} W(\tau_i)(W(t_i) - W(t_{i-1})) \tag{40}$$

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

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

for 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 $t_{i-1} \leq \tau_i \leq t_i$, this gives

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

Hence we find

$$\mathcal{E}\left\{\int_0^t W dW\right\} = \alpha t \tag{44}$$

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 general non-linear SDE with multiplicative noise

$$dX = f(X)dt + g(X)dW \tag{45}$$

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

$$X(t) = \int_0^t f(X)d\tau + \int_0^t g(X)dW \tag{46}$$

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. There are infinitely many possibilities, but 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:

$$dX = f(X)dt + g(X)dW \quad \text{(S)} \tag{47}$$

$$dX = \Big(f(X) + \frac{1}{2}g'(X)g(X)\Big)dt + g(X)dW \quad \text{(I)} \tag{48}$$

We also state without proof that in the Stratonovitch calculus we can integrate and differentiate as if the stochastic processes are ordinary functions. 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 example.

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

$$\left\{ \begin{array}{rcl} dX & = & XdW \quad \text{(S)} \\ X(0) & = & x_0 \end{array} \right. \tag{49}$$

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

$$\left\{\begin{array}{rcl} dX & = & XdW \quad \text{(I)} \\ X(0) & = & x_0 \end{array}\right. \tag{51}$$

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

$$\left\{\begin{array}{rcl} dX & = & -\frac{1}{2}Xdt + XdW \quad \text{(S)} \\ X(0) & = & x_0 \end{array}\right. \tag{52}$$

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

$$\left\{\begin{array}{rcl} d\mathcal{E}\{X\} & = & \mathcal{E}\{XdW\} \\ & = & \mathcal{E}\{X\}\mathcal{E}\{dW\} \\ & = & \mathcal{E}\{X\} \cdot 0 \\ & = & 0 \\ \mathcal{E}\{X(0)\} & = & x_0 \end{array}\right. \tag{54}$$

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

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

How is this for the Stratonovitch equation (49)? Remember that inn 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 XdW is equal to the product of expectations of X and dW. From equations (47) and (48) we see that the Stratonovitch equation (49) is equivalent to the

Ito equation

$$\left\{ \begin{array}{rcl} dX &=& \frac{1}{2}X dt + X dW \quad \text{(I)} \\ X(0) &=& x_0 \end{array} \right. \tag{56}$$

Taking expectations we get

$$\left\{ \begin{array}{rcl} d\mathcal{E}\{X\} &=& \frac{1}{2}\mathcal{E}\{X\}dt + \mathcal{E}\{X dW\} \\ &=& \frac{1}{2}\mathcal{E}\{X\}dt + \mathcal{E}\{X\}\mathcal{E}\{dW\} \\ &=& \frac{1}{2}\mathcal{E}\{X\}dt + \mathcal{E}\{X\} \cdot 0 \\ &=& \frac{1}{2}\mathcal{E}\{X\}dt \\ \mathcal{E}\{X(0)\} &=& x_0 \end{array} \right. \tag{57}$$

and so

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

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.

Consider the Ito equation

$$dX = h(X)dt + g(X)dW \quad \text{(I)} \tag{59}$$

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

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

A similar discretization of the Stratonovitch equation

$$dX = h(X)dt + g(X)dW \quad \text{(S)} \tag{62}$$

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 only know X up to and including time t, is not practical.

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

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

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 $t = 0, \Delta t, 2\Delta t, 3\Delta t, \ldots, n\Delta t$.

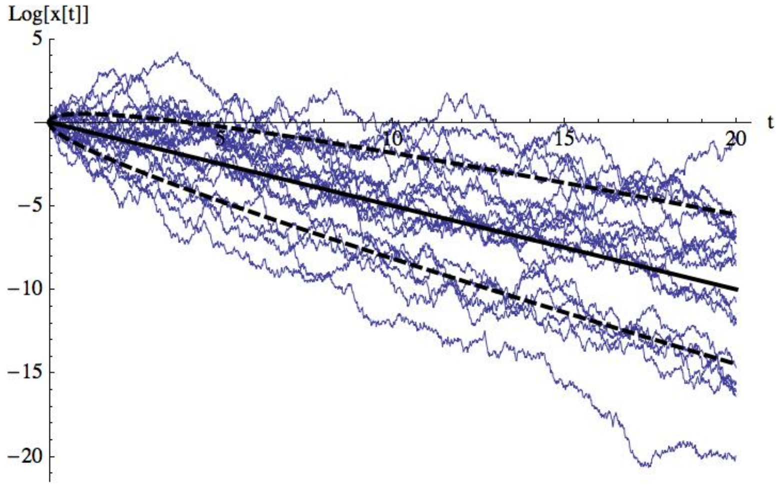

FIGURE 2. Numerical integration of the Ito equation $dX = XdW$. Twenty sample paths of $\log X(t)$ versus t with initial condition $\log X(0) = 0$. Solid line indicates the mean and the dashed lines the mean plus/minus the standard deviation of the distribution of $\log X(t)$.

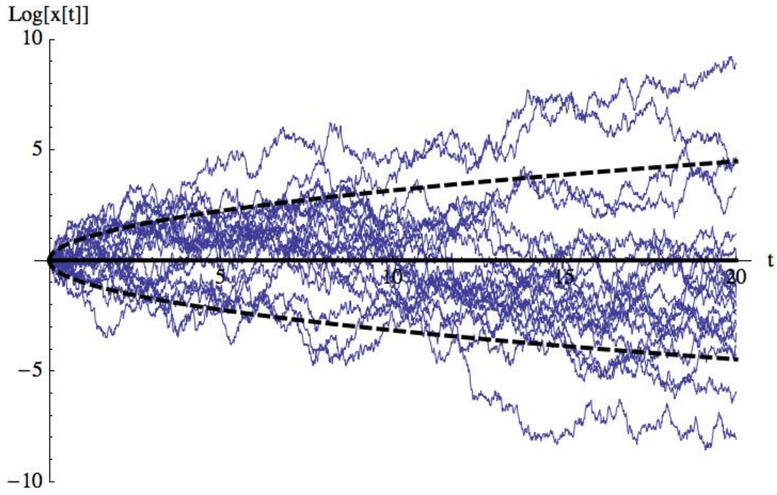

FIGURE 3. Numerical integration of the Stratonovich equation $dX = XdW$, which was solved by numerically integrating the equivalent Ito equation $dX = \frac{1}{2}Xdt + XdW$. Twenty sample paths of $\log X(t)$ versus t with initial condition $\log X(0) = 0$. Solid line indicates the mean and the dashed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