A Gentle Introduction to Stochastic Differential Equations

We saw how Brownian motion paths can be constructed by using cumulative sums of infinitesimal stochastic jumps in each infinitesimal time step. At first glance, this might seem vaguely similar to what we do when we solve differential equations, except that the jumps, while infinitesimal, are stochastic.

Before talking about stochastic differential equations, it is helpful to review some ordinary differential equation (ODE) concepts. If we have a function of a single variable t, and we are presented with an equation to solve like

$$y'(t) = f(t, y(t)) \tag{1}$$

with initial condition $y(0)$ specified. For example

$$y'(t) = y(t), \quad y(0) = 1,$$

has solution $y(t) = e^t$.

A standard emphnumerical approach to solving an equation (1) with an initial condition is to express it in differential form

$$dy(t) = f(t, y(t))dt, \tag{2}$$

and think of the solution as an integral

$$y(t) = y(0) + \int_{s=0}^{t} f(s, y(s))ds. \tag{3}$$

We can find approximate the integral by employing an *Euler scheme* based on the idea that

$$y(t + \Delta) - y(t) \approx f(t, y(t))\Delta$$

or equivalently

$$y(t + \Delta) \approx y(t) + f(t, y(t))\Delta$$

for a small time increment Δ. So we *numerically integrate* by taking

$$\begin{aligned}
y(\Delta) &= y(0) + f(0, y(0))\Delta \\
y(2\Delta) &= y(\Delta) + f(\Delta, y(\Delta))\Delta \\
y(3\Delta) &= y(2\Delta) + f(2\Delta, y(2\Delta)))\Delta \\
&\vdots
\end{aligned}$$

Brownian Motion

In order to understand what happens on the stochastic side, with stochastic differential equations, we start with the intuitive *construction* to an approximation for *standard* Brownian motion based on a random walk that has been re-scaled in time and space. Start with a tiny time increment Δ and define a (random) function by linearly interpolating a function defined at time points $0, \Delta, 2\Delta, \ldots$ as follows. Let $\delta_1, \delta_2, \ldots,$ be an iid sequence of random variables with δ_i taking value ± 1 each with probability $1/2$. [1] For a standard symmetric random walk starting at position 0 at time 0 we would take $X_0 = 0$ and

$$X_n = X_{n-1} + \delta_n \text{ for } n = 1, 2, \ldots.$$

For Brownian motion, we re-scale in time and space by defining $B_0 = 0$ and

$$B_{n\Delta} = B_{(n-1)\Delta} + \sqrt{\Delta}\ \delta_n, \text{ for } n = 1, 2, \ldots.$$

When we do this, we get functions that look something like shown in the following figure.

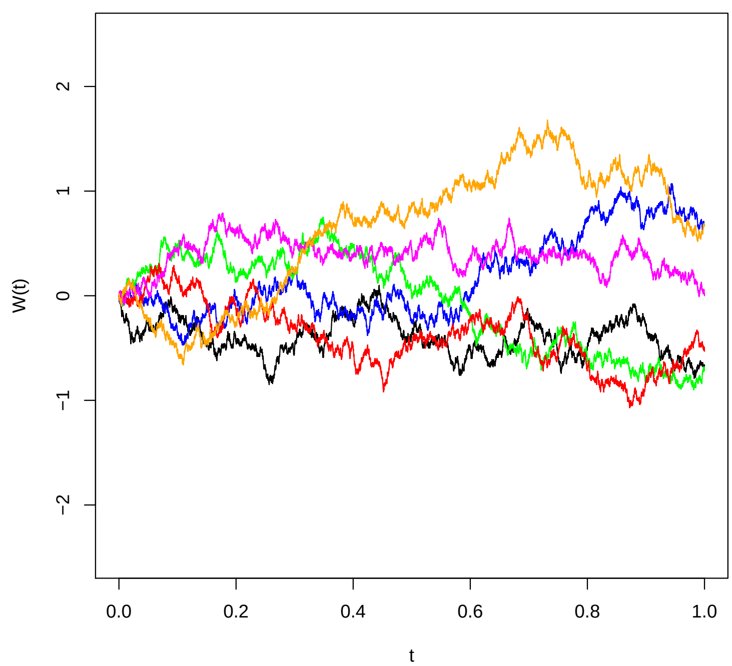

Figure 1: Six realizations of standard Brownian motion based on re-scaled random walk with $\Delta = 1/10,000$. The sample paths appear continuous but they do not appear differentiable.

[1] Actually, any distribution that has mean 0 and variance 1 can be used as the distribution of the δ_i.

To get actual Brownian motion we need to let $\Delta \to 0$ but hopefully, you get a sense of what happens. We a random function (referred to as a *stochastic process* $B_t, t > 0$ satisfying the following properties.

(1) $B_0 = 0$

(2) Independent increments property: for all $0 = t_0 < t_1 < t_2 < \cdots < t_k$ the increments $B_{t_0}, B_{t_1} - B_{t_0}, B_{t_2} - B_{t_1}, \ldots, B_{t_k} - B_{t_{k-1}}$, are independent random variables,

(3) $B_t - B_s \sim N(0, t-s)$ for $t > s$, and

(4) the function $t \to B_t$ is continuous but nowhere differentiable with probability 1.

As a consequence of these properties, for $s < t$ we have

$$0 \overset{(2)}{=} \mathrm{Cov}(B_s, B_t - B_s) = \mathrm{Cov}(B_s, B_t) - \mathrm{Cov}(B_s, B_s) = \mathrm{Cov}(B_s, B_t) - \mathrm{Var}(B_s) \overset{(3)}{=} \mathrm{Cov}(B_s, B_t) - s,$$

so

$$\mathrm{Cov}(B_s, B_t) = s.$$

Generally,

$$\mathrm{Cov}(B_s, B_t) = \min\{s, t\}.$$

Stochastic Differential Equations (SDE's)

Stochastic differential equations generalize ordinary differential equations of the type in (1) by allowing for infinitesimal random *shocks*. They are easier to understand when we think of the version in differential form (3). We take the differential $dB(t)$ of Brownian motion as a primitive object and write equations, for example this equation

$$dS_t = f(t, S_t)dt + g(t, S_t)dB_t, \tag{4}$$

can be interpreted as saying that for small values of Δ we have

$$S_{t+\Delta} - S_t = f(t, S_t)\Delta + g(t, S_t)(B_{t+\Delta} - B_t).$$

In the construction above, we saw that we could take $B_{t+\Delta} - B_t = \sqrt{\Delta}\,\delta$ where $\delta = \pm 1$ each with probability 1/2.

As a consequence, sample paths for an equation like (4) starting with $S_0 = 0$ can be *realized* using the following numerical scheme for a small time increment Δ.

$$\begin{aligned}
S_0 &= 0 \\
S_\Delta &= S_0 + f(0, S_0)\Delta + g(0, S_0)\sqrt{\Delta}\ \delta_1 \\
S_{2\Delta} &= S_\Delta + f(\Delta, S_\Delta)\Delta + g(\Delta, S_\Delta)\sqrt{\Delta}\ \delta_2 \\
S_{3\Delta} &= S_{2\Delta} + f(2\Delta, S_{2\Delta})\Delta + g(2\Delta, S_{2\Delta})\sqrt{\Delta}\ \delta_3 \\
&\vdots
\end{aligned}$$

where, as above $\delta_1, \delta_2, \ldots$ are iid with $\delta_i = \pm 1$ each with probability 1/2.

Note on the increments: We approximate standard Brownian using cumulative sums of increments, where in every time interval Δ we can take our jumps to be random variables whose distribution has mean 0 and variance Δ (i.e. standard deviation $\sqrt{\Delta}$. So we can take the jumps to be $\pm\sqrt{\Delta}$ with equal probability of an down or up jump. We can also take our jumps to be $N(0, \Delta)$ distributed. We get processes with the same limiting behavior in the limit as the time increments go to zero. This fact is related to the central limit theorem.

Ito's Lemma.

We are often presented with a process S_t satisfying an SDE and we want to determine the SDE satisfied by e.g. $Y_t \stackrel{def}{=} f(S_t)$ where f is a suitably smooth function. For ODE's, the situation is relatively simple. If we have

$$dy(t) = g(t, y(t))dt$$

then

$$d(t, f(y(t))) = \left\{ \frac{\partial}{\partial x} f(t, y(t))y'(t) + \frac{\partial}{\partial y} f(y(t), t) \right\} dt$$

where the partial derivative with respect to x refers to the first argument of f and the partial derivative with respect to y refers to the second argument of f.

Things get a little bit more interesting in the SDE case. If S_t satisfies

$$dS_t = f(t, S_t)dt + g(t, S_t)dB_t \tag{5}$$

and we define $Y_t = h(t, S_t)$. To get an SDE, for suitably smooth h we can expand in a Taylor series and write

$$h(t+dt, s+ds) \approx h(t,s) + \frac{\partial h}{\partial x}(t,s)dt + \frac{\partial h}{\partial y}(t,s)ds + \frac{1}{2}\frac{\partial^2 h}{\partial x^2}(t,s)dt^2 + \frac{1}{2}\frac{\partial^2 h}{\partial y^2}(t,s)ds^2 + \frac{\partial^2 h}{\partial x \partial y}(t,s)dtds.$$

which suggests that

$$dh(t,S_t) \approx \frac{\partial h}{\partial x}(t,S_t)dt + \frac{\partial h}{\partial y}(t,S_t)dS_t + \frac{1}{2}\frac{\partial^2 h}{\partial x^2}(t,S_t)dt^2 + \frac{1}{2}\frac{\partial^2 h}{\partial y^2}(t,S_t)dS_t^2 + \frac{\partial^2 h}{\partial x \partial y}(t,S_t)dtdS_t.$$

We can substitute the expression for dS_t from the SDE (5) and obtain

$$\begin{gathered} dh(t,S_t) \approx \frac{\partial h}{\partial x}(t,S_t)dt + \frac{\partial h}{\partial y}(t,S_t)\left\{f(t,S_t)dt + g(t,S_t)dB_t\right\} + \\ \frac{1}{2}\frac{\partial^2 h}{\partial x^2}(t,S_t)dt^2 + \frac{1}{2}\frac{\partial^2 h}{\partial y^2}(t,S_t)\left\{f(t,S_t)dt + g(t,S_t)dB_t\right\}^2 + \\ \frac{\partial^2 h}{\partial x \partial y}(t,S_t)dt\left\{f(t,S_t)dt + g(t,S_t)dB_t\right\}. \end{gathered}$$

and collecting like terms yields

$$\begin{gathered} dh(t,S_t) \approx \left(\frac{\partial h}{\partial x}(t,S_t) + \frac{\partial h}{\partial y}(t,S_t)f(t,S_t)\right)dt + \frac{\partial h}{\partial y}(t,S_t)g(t,S_t)dB_t + \\ \frac{1}{2}\left(\frac{\partial^2 h}{\partial x^2}(t,S_t) + \frac{\partial^2 h}{\partial y^2}(t,S_t)f(t,S_t)^2 + 2\frac{\partial^2 h}{\partial x \partial y}(t,S_t)f(t,S_t)\right)dt^2 \\ +\frac{1}{2}\frac{\partial^2 h}{\partial y^2}g(t,S_t)^2 dB_t^2 + \\ \left(\frac{\partial^2 h}{\partial x \partial y}(t,S_t) + \frac{\partial^2 h}{\partial y^2}f(t,S_t)\right)g(t,S_t)\ dt\ dB_t. \end{gathered}$$

The terms dt^2 and $dtdB_t$ are of smaller order than dt and can be ignored, but a new phenomenon emerges in that we can view $dB_t^2 = dt$, which can be understood from the strong law of large numbers. Think about integrating dB_t^2 from t to $t+s$ for some $s > 0$. From our construction, consider jumps $\sqrt{s/n}\delta_j$ in time periods of size s/n where n is large and the δ_j are iid ± 1, with equal probability. Then

$$\int_t^{t+s} dB_t^2 \approx \sum_{i=1}^{n}(B_{t+is/n} - B_{t+(i-1)s/n})^2 = \sum_{i=1}^{n}(\sqrt{s/n}\delta_i)^2 = s\frac{1}{n}\sum_{i=1}^{n}\delta_i^2 \to s,$$

as $n \to \infty$. So we see that integrating dB_t^2 is the same as integrating dt.

This is not meant to be a completely rigorous argument, but hopefully you get a feel for why in Ito's Lemma we get an extra term, and we end up with the conclusion that

Ito's Lemma. Under appropriate smoothness conditions

$$dh(t, S_t) = \left(\frac{\partial h}{\partial x}(t, S_t) + \frac{\partial h}{\partial y}(t, S_t) f(t, S_t) + \frac{1}{2}\frac{\partial^2 h}{\partial y^2} g(t, S_t)^2 \right) dt + \frac{\partial h}{\partial y}(t, S_t) g(t, S_t) dB_t.$$

Example. Geometric Brownian motion

If we start with S_t as Brownian motion, so that $dS_t = dB_t$, (hence $f(x, y) \equiv 0$ and $g(x, y) \equiv 1$ and we define

$$Y_t = \mu t + \sigma B_t = h(t, B_t),$$

where $h(x, y) = \mu x + \sigma y$ then $\frac{\partial h}{\partial x} = \mu$, $\frac{\partial h}{\partial y} = \sigma$ and Ito's lemma tells us Y_t satisfies the SDE

$$dY_t = \mu dt + \sigma dB_t$$

This is referred to as Brownian motion with drift.

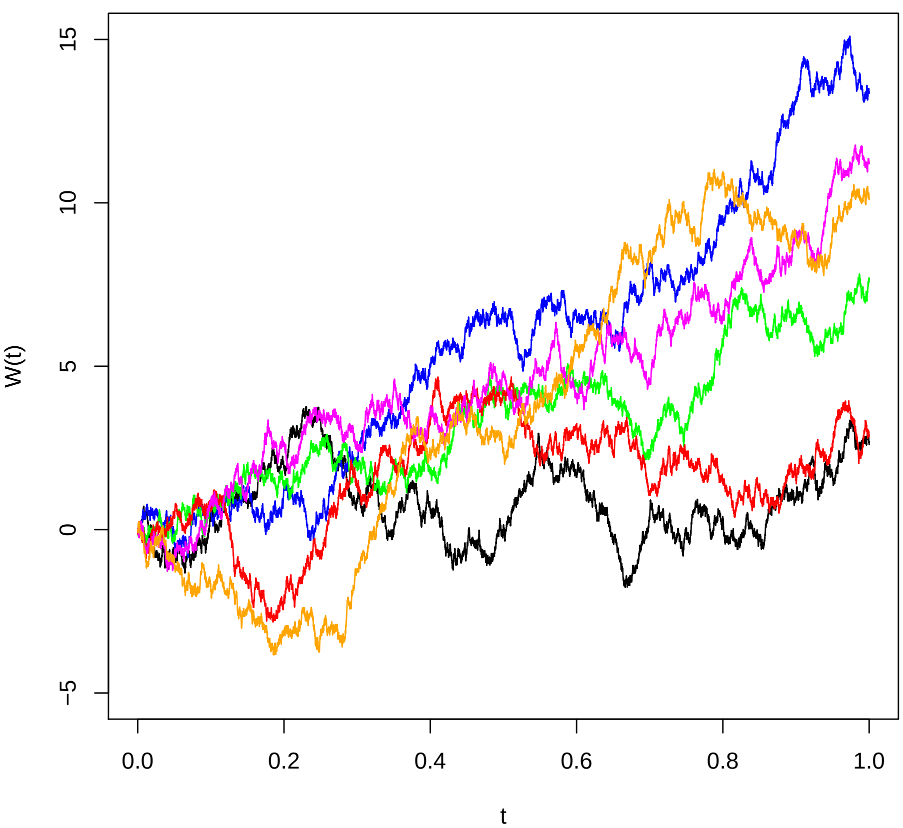

Figure 2: Six realizations of Brownian motion with drift with $\mu = 2$ and $\sigma = 5$.

For a slightly more complicated example, take Y_t as in the last example,

$$dY_t = f(t, Y_t)dt + g(t, Y_t)dB_t$$

where $f \equiv \mu$ and $g \equiv \sigma$, and take

$$X_t = e^{Y_t} = h(t, Y_t),$$

where $h(x, y) = e^y$. Such a process is useful in finance because it remains positive for all times t.

Ito's lemma gives

$$\begin{aligned} dX_t &= (\mu e^{Y_t} + \frac{1}{2}\sigma^2 e^{Y_t})dt + \sigma e^{Y_t} dB_t \\ &= (\mu + \frac{1}{2}\sigma^2)X_t)dt + \sigma X_t dB_t \end{aligned}$$

Exercise. Suppose we have

$$dX_t = \mu X_t dt + \sigma X_t dB_t$$

what stochastic differential equation is satisfied by $\log(X_t)$?

Calibration of the Geometric Brownian Motion Model

Suppose we observe a GBM process starting at G_0 at time 0 at times $t_0 = 0 < t_1 < t_2 < \cdots < t_N$. Define

$$\Delta t_i := t_{i+1} - t_i$$

and

$$\Delta \log G_i := \log G_{t_{i+1}} - \log G_{t_i}$$

for $i = 0, \ldots, N-1$. Then the $\Delta \log G_i$ are independent with $\Delta \log G_i \sim N(\Delta t_i \mu, \Delta t_i \sigma^2)$ and we can write down a likelihood function

$$L(\mu, \sigma) = \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi}\sqrt{\Delta t_i}\sigma} \exp\left\{ -\frac{1}{2} \left(\frac{\Delta \log G_i - \mu \Delta t_i}{\sigma \sqrt{\Delta t_i}} \right)^2 \right\}$$

and the log likelihood takes the form

$$\log L(\mu, \sigma) = -\frac{N}{2}\log(2\pi) - N\log(\sigma) - \frac{1}{2}\sum_{i=0}^{N-1}\log(\Delta t_i) - \frac{1}{2\sigma^2}\sum_{i=0}^{N-1}\left(\frac{\Delta \log G_i - \mu\Delta t_i}{\sqrt{\Delta t_i}}\right)^2$$

This function is quadratic and concave in μ and we can differentiate with respect to μ and equate to zero to give

$$\hat{\mu} = \frac{\sum_{i=0}^{N-1} \Delta \log G_i}{\sum_{i=0}^{N-1} \Delta t_i} = \frac{\log(G_{t_N}) - \log(G_0)}{t_N - t_0} = \frac{\log(G_{t_N}) - \log(G_0)}{t_N} =$$

The distribution of this estimator is $N(\mu, \sigma^2/T_N)$. Subsitution of $\hat{\mu}$ into the log likelihood, differentiating with respect to σ and equating to zero gives

$$\hat{\sigma} = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} \frac{(\Delta \log G_i - \hat{\mu} \Delta t_i)^2}{\Delta t_i}}$$