Stochastic Differential Equations

MATH574 Final Project - Lecture Notes

By Louis Meunier

We'll be studying differential equations with some notion of "randomness" included. In particular, we'll aim to study equations of the form

$$\frac{\mathrm{d}N}{\mathrm{d}t} = b(t, N) + \sigma(t, N)W_t,$$

where N a function of time, and W_t "noise", time-dependent randomness. We'll talk about:

- What kind of object is W_t?
- What does a solution look like to such an equation?
- How do these equations differ or relate to analogous deterministic (non-random) equations?

1 Probability

What is "W_t"? How do we mathematically capture randomness? Here we'll briefly review some probability theory.

Definition 1 (Probability Space): Let Ω be a nonempty set (our set of possible outcomes; we call it a *sample space*), and $\mathcal{F}$ a collection of subsets of Ω (this is our collection of events; we call it an *event space*), with the properties that

- $\Omega \in \mathcal{F}$;
- $E \in \mathcal{F} \Rightarrow E^c \in \mathcal{F}$; and
- $\{E_n : n \geq 1\} \subseteq \mathcal{F} \Rightarrow \bigcup_n E_n \in \mathcal{F}$. Namely, $\mathcal{F}$ a σ-algebra of subsets of Ω. A function

  $$\mathbb{P} : \mathcal{F} \to [0, 1],$$

  is called a *probability measure* on $\mathcal{F}$ if
- $\mathbb{P}(\Omega) = 1$;
- if $\{E_n : n \geq 1\} \subseteq \mathcal{F}$ a disjoint collection, then $\mathbb{P}\left(\bigcup_n E_n\right) = \sum_n \mathbb{P}(E_n)$.

Then, the triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *probability space*.

This probability space is where we will be mainly working. You can think about $\mathcal{F}$ as a set of possible events following an "experiment", and $\mathbb{P}$ representing the likelihood of a certain event happening.

We'd like a way to "measure" qualities of the experiment:

Definition 2 (Random Variable): A *random variable* is neither random nor a variable; it is a (Borel-measurable) function

$$X : \mathcal{F} \to \mathbb{R}.$$

We are interested in the distribution of X; namely, given any set $B \in \mathfrak{B}_{\mathbb{R}}$, we write

$$\mathbb{P}(X \in B) := \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in B\});$$

this induces a probability measure on $\mathfrak{B}_{\mathbb{R}}$ given by $Q(B) := \mathbb{P}(X \in B)$ (we usually never explicitly write Q, and prefer the second).

How do we specify a random variable? There are many ways, but in any case we must, for any Borel set B, specify the probability that $X \in B$. Suppose X takes values in a finite or countable set S. Then, let

$$\mathbb{P}(X = k) = f(k), \qquad k \in S,$$

where f is a function $S \to [0, 1]$ with the property that $\sum_{k \in S} f(k) = 1$. Then, given any subset $S' \subseteq S$,

$$\mathbb{P}(X \in S') = \sum_{k \in S'} f(k).$$

The function f is called the *point mass function* (pmf) of X; it completely specifies the distribution of X. We define the *cumulative distribution function* (cdf) of X to be, for any $x \in \mathbb{R}$,

$$F(x) := \mathbb{P}(X \leq x) = \sum_{k \in S : k \leq x} f(k).$$

Suppose next X takes values in all of $\mathbb{R}$. We can define the cdf in the same manner; assuming F is (absolutely) continuous, then (*by a strong version of the Fundamental Theorem of Calculus*) there exists a function f such that

$$F(x) = \int_{-\infty}^{x} f(y) \, \mathrm{d}y,$$

for every $x \in \mathbb{R}$. f is called a *probability density function* (pdf).

Note the similarities between the pmf and pdf; the first is summed over, and the second is integrated over, but each represent, in a way, the "point-mass probability" about a point; hence our identical notations.

Example 1:

Discrete	**Continuous**
$X = 1$ with prob. $p, 0$ with prob. $1 - p$, where $p \in (0,1)$; we say X *Bernoulli*.	Let $[a,b] \subseteq \mathbb{R}$ and $f(x) = \frac{1}{b-a} \cdot \mathbb{1}[a,b]$; then, $F(x) = \frac{x-a}{b-a}$ if $x \in [a,b]$ and zero otherwise; we say X *uniform* in $[a,b]$.
Let $f(k) = \frac{e^{-\lambda}\lambda^k}{k!}$ for $k \in \mathbb{N}$ and $\lambda > 0$; we say X *Poisson* with parameter λ.	**Let $\mu, \sigma^2 \in \mathbb{R}$. Define $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$. Then, if the pdf of X is given by f, we say X *normal*, and write $X \sim \mathcal{N}(\mu, \sigma^2)$.**

Definition 3 (Expected Value): The *mean/expected value* of a random variable X is defined as it's "average value" (if one exists). We denote

$$\mathbb{E}[X] = \begin{cases} \int_S x f(x)\, \mathrm{d}x & \text{if } X \text{ continuous} \\ \sum_{x \in S} x f(x) & \text{if } X \text{ discrete} \end{cases},$$

where $S = \{x \in \mathbb{R} \mid f(x) \neq 0\}$ the *support* of X. More generally, for $k \in \mathbb{N}$, the k-th moment of X is defined as $\mathbb{E}[X^k]$, where finite.

X need not have a finite mean. Consider the discrete random variable with $X = k$ with probability $\frac{6}{\pi^2}\frac{1}{k^2}$ for $k \in \mathbb{N}$.

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we denote the space

$$L^2(\mathbb{P}) := \{X : \mathcal{F} \to \mathbb{R} \mid \mathbb{E}[X^2] < \infty\}.$$

This is a Hilbert space equipped with the inner product $\mathbb{E}[XY]$ (namely, a complete inner product space), which we won't prove here but will use extensively in the remaining parts.

Random variables are almost what we were looking for in our initial "noise", W_t; however, we need it to have some dependence on time as well. We address this with the following definition.

Definition 4 (Stochastic Process): A *stochastic process* is a parametrized collection of random variables $\{X_t : t \in T\}$, where T some index set.

For our purposes, we'll usually have the stochastic process varying in "time", and take $T = [0, \infty)$.

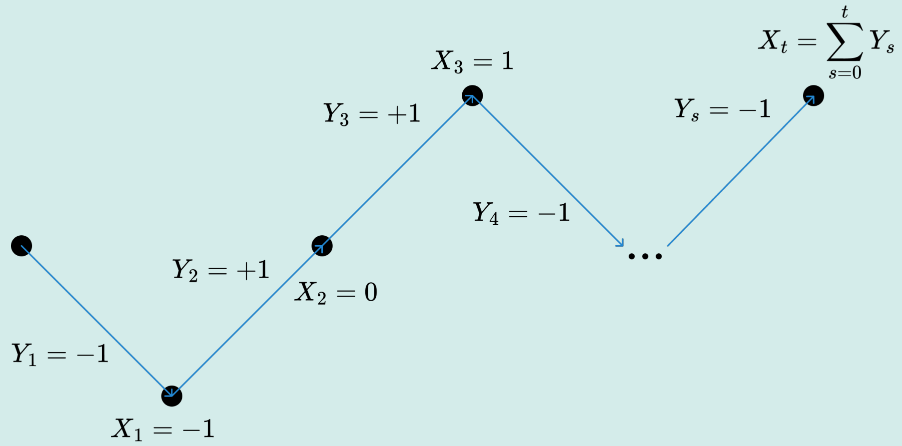
Example 2 (Random Walk): Let $T = \mathbb{N}$ and

$$Y_t = \begin{cases} 1 & \text{with probability } \frac{1}{2} \\ -1 & \text{with probability } \frac{1}{2} \end{cases}$$

and

$$X_t := \sum_{s=0}^{t} Y_s.$$

X_t is called a "random walk"; each Y_s represents a movement left/right along a discrete lattice, and so X_t represents the final "location".



Example 3 (Poisson Process): Let $T = [0, \infty)$, $\lambda > 0$ and for each $t \in T$, X_t be Poisson distributed with parameter λt. Then, $\{X_t : t \in T\}$ is called a *Poisson point process*.

The process we'll focus on is the following:

Definition 5 (Wiener Process/Brownian Motion): A *Wiener process*, $\{B_t : t \geq 0\}$, is a stochastic process with the properties

1. $B_0 = 0$;
2. the increments are independent, i.e. $B_{t+s} - B_t$, $B_{t'+s'} - B_{t'}$ are independent, for every t, t' and $s, s' \geq 0$;
3. the increments are normal, i.e. $B_{t+s} - B_t \sim \mathcal{N}(0, s)$ for every $s \geq 0$;
4. the map $t \mapsto B_t$ is continuous a.s.

This technical definition is fairly useless for any kind of application. We'll briefly look at alternative ways to characterize Brownian motion. Fix $x \in \mathbb{R}^n$. For $y \in \mathbb{R}^n$ and $t > 0$, define

$$\rho(t,x,y) := (2\pi t)^{-n/2} \exp\left(-\frac{|x-y|^2}{2t}\right),$$

where $|\cdot|$ the usual Euclidean norm of $\mathbb{R}^n$. Then, for any set of times $0 \le t_1 \le t_2 \le \cdots \le t_k$ and subsets $F_1, \ldots, F_k \subseteq \mathbb{R}^n$, define the measure

$$\nu_{t_1,\ldots,t_k}(F_1 \times \cdots \times F_k) := \int_{F_1 \times \cdots \times F_k} \rho(t_1, x, x_k)\rho(t_2 - t_1, x_1, x_2) \cdots \rho(t_k - t_{k-1}, x_{k-1}, x_k) \, \mathrm{d}x_1 \cdots \mathrm{d}x_k.$$

By *Kolmogorov's Extension Theorem*, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P}_x)$ and process $\{B_t\}$ such that

$$\mathbb{P}_x\left(B_{t_1} \in F_1, \ldots, B_{t_k} \in F_k\right) = \nu_{t_1,\ldots,t_k}(F_1 \times \cdots \times F_k).$$

Such a process is called the "Brownian motion starting at x" (since, in particular, $\mathbb{P}_x(B_0 = x) = 1$). This space is not unique by any means. However, we can always pick one such that the paths $t \mapsto B_t(\omega)$ are continuous, and hence we may realize Brownian motion as a random distribution on the space $C([0,\infty); \mathbb{R}^n)$. In short, B_t can be thought of as a random n-dimensional curve.

A more constructive, less probability-theoretic construction (and in particular, one that is helpful in numerical simulations) can be given on a closed interval. Let $\{h_n\}$ be an orthonormal basis for $L^2([0,1])$ and $\{\alpha_n\}$ an independent sequence of normal random variables with mean 0 and variance 1, i.e. $\alpha_n \sim \mathcal{N}(0,1)$. Then, define for $t \ge 0$,

$$B_t := \sum_n \alpha_n \cdot \int_0^t h_n(s) \, \mathrm{d}s.$$

This is well-defined since $\mathbb{E}[\alpha_n] = 0$. *To be more specific, such a construction is a so-called "Brownian Bridge", where the endpoints are fixed.*

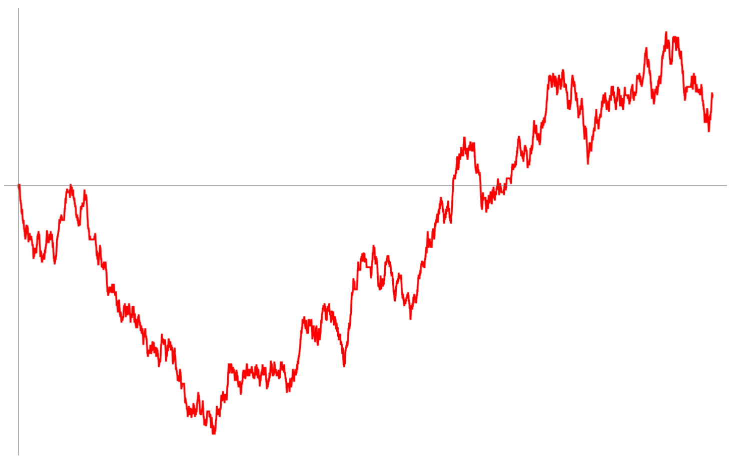


Figure 1: A (simulated) instance of Brownian motion on the line using the construction below.

Yet another construction is as the limit of a random walk (this is my preferred way of thinking about it). Let $\{Y_k : k \geq 1\}$ be an independent and identically distributed (iid) sequence of random variables with $\mathbb{P}(Y_k = \pm 1) = \frac{1}{2}$, and define for $N \geq 1$,

$$X_N := \sum_{k=1}^{N} Y_k.$$

By the *Central Limit Theorem*, $\frac{1}{\sqrt{N}} \cdot X_N \xrightarrow{d} \mathcal{N}(0,1)$. Define

$$B_t^N := \frac{X_{\lfloor Nt \rfloor}}{\sqrt{N}}, \qquad B_t := \lim_{N \to \infty} B_t^N.$$

Remark that in particular,

$$B_t^N = \frac{X_{\lfloor Nt \rfloor}}{\sqrt{N}} = \frac{\sqrt{Nt}}{\sqrt{N}} \cdot \frac{X_{\lfloor Nt \rfloor}}{\sqrt{Nt}} = \sqrt{t} \cdot \frac{X_{\lfloor Nt \rfloor}}{\sqrt{Nt}} \xrightarrow{d} \sqrt{t} \cdot \mathcal{N}(0,1) = \mathcal{N}(0,t).$$

With some difficult, one can prove the remaining properties of Brownian motion.

2 Itô Calculus

Using Brownian motion, we can start to develop (a version of) stochastic calculus. Consider the “differential equation”

$$\frac{\mathrm{d}N}{\mathrm{d}t} = (r(t) + \text{"noise"})N.$$

If we treated this like an ODE, then, we would arrive at an integral equation of the form

$$\int \frac{1}{N} \,\mathrm{d}N = \int r(t) \,\mathrm{d}t + \int \text{"noise"} \,\mathrm{d}t.$$

If we formalize “noise” with a stochastic process of some kind, then we need to a way of “integrating” noise against time. In particular, let’s write W_t in place of “noise”. Then, the more general form of stochastic differential equation (SDE) we’d like to analyze is the following,

$$\frac{\mathrm{d}X}{\mathrm{d}t} = b(t, X_t) + \sigma(t, X_t) W_t. \qquad \dagger$$

In general, we would like the following properties:

- W_{t_1}, W_{t_2} are independent;
- W_t is “stationary”, that is, the joint distribution of $W_{t_1+t}, ..., W_{t_k+t}$ is independent of t;
- $\mathbb{E}[W_t] = 0$;
- W_t has continuous paths.

However, it turns out that these conditions are impossible to simultaneously satisfy! We can, however, realize W_t as a "generalized" process (called the "white noise process"); namely, rather than being a stochastic process on the space of real-valued functions on $[0, \infty)$, we can realize it as a process on the space of tempered distributions on $[0, \infty)$. This is hard to do, so we'll take a different approach.

To motivate this, consider a discretized version of †. For fixed time t, consider a discrete partition $0 = t_0 < t_1 < \cdots < t_m = t$ up to time t. Then, † looks like

$$X(t_{k+1}) - X(t_k) = b(t_k, X(t_k))\Delta t_k + \sigma(t_k, X(t_k))W(t_k)\Delta t_k,$$

where $\Delta t_k := t_{k+1} - t_k$. Put $X_k := X(t_k), W_k := W(t_k)$, and let $\{V_t : t \geq 0\}$ be a stochastic process such that $\Delta V_k = V_{t_{k+1}} - V_{t_k} = W_k \Delta t_k$. Then, † simplifies

$$X_{k+1} - X_k = b(t_k, X_k)\Delta t_k + \sigma(t_k, X_k)\Delta V_k. \qquad ‡$$

In this "discrete setting", it turns out that the only stochastic process V_t that has the properties such as to satisfy those of W_t above is Brownian motion!

Applying ‡ repeatedly, we find

$$X_k = X_0 + \sum_{j=0}^{k-1} b\big(t_j, X_j\big)\Delta t_j + \sum_{j=0}^{k-1} \sigma\big(t_j, X_j\big)\Delta B_j.$$

Letting $\Delta t \to 0$ in some appropriate way and $k = m$, the first summation should converge to $\int_0^t b(s, X_s)\, \mathrm{d}s$, so heuristically the second should converge to "$\int_0^t \sigma(s, X_s)\, \mathrm{d}B_s$". With this as motivation, we now define $\int_0^t f\, \mathrm{d}B_t$ as the limit of this summation behaviour in a proper sense, for a class of functions f.

We'll do this in a manner very similar to the typical construction of the Lebesgue integral. We begin by defining the integral for so-called "simple functions", then for continuous functions by approximation with simple function, then for bounded functions via approximation with continuous functions. Fix times $T > S$. We'll define

$$I[f](\omega) := \int_S^T f(t, \omega)\, \mathrm{d}B_t(\omega).$$

Note that this is a random variable in its own right!

We'll tacitly assume some further technical assumptions on the functions we'll define $I[f]$ *for; namely measurability of* f *in the product space* $\mathfrak{B}_{[0,\infty)} \times \mathcal{F}$*, the "adaptability" of* f *to the filtration* $\mathcal{F}_t := \sigma\Big(\Big\{B_i(s) : \begin{smallmatrix} 1 \leq i \leq n \\ 0 \leq s \leq t \end{smallmatrix}\Big\}\Big)$*, and the* $L^2([S, T])$*-norm of* f *is finite on average.*

Step 1: We call φ a *simple function* if it is of the form

$$\varphi(t, \omega) = \sum_j e_j(\omega)\mathbb{1}\big[t_j, t_{j+1}\big)(t),$$

with $e_j(\omega)$ a function of only ω, $\mathbb{1}[t_j, t_{j+1})$ the indicator function of the interval $[t_j, t_{j+1})$, and the sum is over a finite partition $\{t_j\}$ of $[S, T]$. Then, simply define

$$I(\varphi)(\omega) := \sum_j e_j(\omega)\left(B_{t_{j+1}} - B_{t_j}\right)(\omega).$$

Note that we choose to take the value $e_j(\omega)$ on the "interval" $B_{t_{j+1}} - B_{t_j}$, i.e. the left-hand value. This is somewhat arbitrary, but turns out to be the right choice in ultimately constructing the "Itô integral". Another option (common in physics) is to take the midpoint over the interval, $\frac{e_{j+1}+e_j}{2}$. This leads to what is called the "Stratanovich integral". This choice fundamentally changes the properties that we will develop to follow.

We promptly have the following important identity that will motivate how to proceed, and is perhaps the most important theorem in the theory of Itô Calculus:

Theorem 1 (Itô Isometry):

$$\boxed{\mathbb{E}\left[\left(\int_S^T \varphi(t,\omega)\,\mathrm{d}B_t\right)^2\right] = \mathbb{E}\left[\int_S^T \varphi^2(t,\omega)\,\mathrm{d}t\right]}$$

That is to say, the $L^2(\mathbb{P})$ norm of the Itô integral of φ is equal to the expected value of $L^2([S,T])$ norm of φ.

Proof: Put $\Delta B_j = B_{t_{j+1}} - B_{t_j}$. Then, by properties of Brownian motion, ΔB_i and ΔB_j are independent, and thus

$$\mathbb{E}\left[e_i e_j \Delta B_i \Delta B_j\right] = \begin{cases} \mathbb{E}[e_i^2]\mathbb{E}\left[(\Delta B_i)^2\right] & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} = \begin{cases} \mathbb{E}[e_i^2](t_{i+1} - t_i) & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}.$$

So,

$$\begin{aligned}
\mathbb{E}\left[\left(\int \varphi\,\mathrm{d}B\right)^2\right] &= \mathbb{E}\left[\left(\sum_j e_j \Delta B_j\right)^2\right] \\
&= \mathbb{E}\left[\sum_{i,j} e_i e_j \Delta B_i \Delta B_j\right] \\
&= \sum_{i,j} \mathbb{E}\left[e_i e_j \Delta B_i \Delta B_j\right] \\
&= \sum_i \mathbb{E}[e_i^2(t_{i+1} - t_i)] \\
&= \mathbb{E}\left[\sum_i e_i^2(t_{i+1} - t_i)\right] = \mathbb{E}\left[\int \varphi^2\,\mathrm{d}t\right].
\end{aligned}$$

$\square$

Step 2: Define $I[g]$ for continuous and bounded functions.

Theorem 2: Let g be bounded and continuous. Then, there exists a sequence of step functions φ_n such that $\int_S^T (g - \varphi_n)^2 \, \mathrm{d}t \to 0$ as $n \to \infty$.

Proof: For each $n \geq 1$, let $\{[t_j, t_{j+1})\}$ be a partition of $[S, T]$ with the property that $|t_{j+1} - t_j| < \frac{1}{n}$ for each j. Then define

$$\varphi_n(\omega) := \sum_j g(t_j, \omega) \cdot \mathbb{1}[t_j, t_{j+1})(t);$$

in particular, φ_n takes the value of g on the left endpoint of each interval. Then, the continuity of g immediately gives $\int_S^T (g(t, \omega) - \varphi_n(t, \omega))^2 \, \mathrm{d}t \to 0$ for every ω so in particular

$$\mathbb{E}\left[\int_S^T (g(t, \omega) - \varphi_n(t, \omega))^2 \, \mathrm{d}t\right] \to 0, n \to \infty.$$

$\square$

Step 3: Approximate bounded functions with continuous functions; this is identical to the previous step.

For any bounded function f, then, we can approximate f with step functions, say $\{\varphi_n\}$, such that $\mathbb{E}\left[\int_S^T (f - \varphi_n)^2 \, \mathrm{d}t\right] \to 0$. With the Itô isometry as motivation, then, define the Itô integral of f as

$$\boxed{I[f](\omega) := \lim_{n \to \infty} \int_S^T \varphi_n(t, \omega) \, \mathrm{d}B_t}$$

with the limit taken in $L^2(\mathbb{P})$. We immediately see two possible problems with this definition:

1. Does this limit even exist?
2. Is it unique?

To answer 1., note that by Itô Isometry (for simple functions), for any $n, m \geq 1$, we have

$$\mathbb{E}\left[\left(\int_S^T \varphi_n \,\mathrm{d}W_t - \int_S^T \varphi_m \,\mathrm{d}W_t\right)^2\right] = \mathbb{E}\left[\int_S^T (\varphi_n - \varphi_m)^2 \,\mathrm{d}t\right]$$
$$= \mathbb{E}\Big[\|\varphi_n - \varphi_m\|_{L^2([S,T])}\Big]$$
$$\leq \mathbb{E}\Big[\|\varphi_n - f\|_{L^2([S,T])}\Big] + \mathbb{E}\Big[\|f - \varphi_m\|_{L^2([S,T])}\Big].$$

By construction, $\varphi_n \to f$ in $L^2([S,T])$, so each of these two expectation terms can be made arbitrary small for large n, m. Then, in particular, $\left\{\int_S^T \varphi_n \,\mathrm{d}W_t\right\}$ forms a *Cauchy sequence* in $L^2(\mathbb{P})$. This space is complete, and thus the limit indeed exists in $L^2(\mathbb{P})$.

To answer 2., suppose we take two different sequences φ_n, ψ_n such that $\mathbb{E}\big[\int (f - \varphi_n)^2 \,\mathrm{d}t\big], \mathbb{E}\big[\int (f - \psi_n)^2 \,\mathrm{d}t\big]$ both converge to zero. Then, following identical work as in addressing 1.,

$$\mathbb{E}\left[\left(\int |\varphi_n - \psi_n| \,\mathrm{d}W_t\right)^2\right] = \mathbb{E}\left[\int (\varphi_n - \psi_n)^2 \,\mathrm{d}t\right]$$
$$\leq \mathbb{E}\left[\int (\varphi_n - f)^2 \,\mathrm{d}t\right] + \mathbb{E}\left[\int (\psi_n - f)^2\right] \underset{n\to\infty}{\to} 0.$$

In particular, $\int |\varphi_n - \psi_n| \,\mathrm{d}W_t \to 0$ in $L^2(\mathbb{P})$ from which it follows $\lim_n \int \varphi_n \,\mathrm{d}W_t = \lim_n \int \psi_n \,\mathrm{d}W_t$, so the limit indeed unique.

Thus, we've developed a manner of interpreting integrating a function against a stochastic process. But how do we actually compute this? And more importantly, what's the connection to SDEs?

First, we can rewrite † in "integral form" (by formally multiplying both sides by $\mathrm{d}t$)

$$\mathrm{d}X_t = b \,\mathrm{d}t + \sigma W_t \,\mathrm{d}t.$$

Integrating with respect to $\mathrm{d}t$ from 0 to t, we find

$$X_t = X_0 + \int_0^t b(s,\omega) \,\mathrm{d}s + \text{"} \int_0^t \sigma(s,\omega) W_t \,\mathrm{d}t\text{"}.$$

To properly make sense of the integral on the right, we need to have some manner of integrating against W_t and $\mathrm{d}t$. Recall that in our earlier formulation, we never worked with W_t directly, but rather with "discretized steps" of it, namely we found that in some formal sense,

$$\text{"}\Delta B_k = B_{t_{k+1}} - B_{t_k} = W_k \Delta t_k.\text{"}$$

Dividing both sides by Δt_k, and taking $\Delta t_k \to 0$, we are inspired to formally write that W_t is the derivative with respect to time of B_t; that is,

$$"W_t = \frac{\mathrm{d}B_t}{\mathrm{d}t}", \qquad \text{"White noise is the derivative of Brownian motion."}$$

One should be sure to understand that this is a purely heuristic notion, hence the quotation marks. Indeed, one can show that the sample paths $t \mapsto B_t$ derived from Brownian motion are almost surely nowhere differentiable, so this statement doesn't even make sense in a fixed-time point of view. Nonetheless, it is still a useful mental image if nothing else.

From here, our integral formula becomes

$$X_t = X_0 + \int_0^t b(s,\omega)\,\mathrm{d}s + "\int_0^t \sigma(s,\omega)\frac{\mathrm{d}B_t}{\mathrm{d}t}\cdot \mathrm{d}t".$$

In short, then, we have good reason to accept the right-hand integral as the Itô integral of $\sigma(s,\omega)$ on $[0,t]$! With this as motivation, we make the following formal definition:

Definition 6 (Itô Process): Let B_t be the standard 1-dimensional Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, an *Itô Process* on $(\Omega, \mathcal{F}, \mathbb{P})$ is a stochastic process $\{X_t : t \geq 0\}$ of the form

$$X_t = X_0 + \int_0^t u(s,\omega)\,\mathrm{d}s + \int_0^t v(s,\omega)\,\mathrm{d}B_s,$$

where u, v subject to certain technical conditions (of note: the $L^2([0,t])$ (resp. $L^1([0,t])$)-norm of $v(t,\omega)$ (resp. $u(t,\omega)$) is finite for every $t \geq 0$ almost surely) and $\int_0^t (\cdots)\,\mathrm{d}B_t$ the Itô integral of v. We'll typically denote such a process in the concise (and suggestive) formulation

$$\mathrm{d}X_t = u\,\mathrm{d}t + v\,\mathrm{d}B_t.$$

Thus, we finally have a concrete way of speaking of our SDE †; namely, given such an equation, we'll say that X_t a solution if it is an Itô process with $u = b, v = \sigma$.

All of this has been fairly formal, and we have no real manner of algebraically manipulating such formulas yet. The following lemma, which can be thought of as a sort of "chain rule", will be crucial in this direction:

Theorem 3 (Itô's Lemma): Let $g \in C^2([0,\infty),\mathbb{R})$. Then, if Y_t a stochastic process defined by $Y_t = g(t, X_t)$, then

$$\mathrm{d}Y_t = \frac{\partial g}{\partial t}(t, X_t)\,\mathrm{d}t + \frac{\partial g}{\partial x}(t, X_t)\,\mathrm{d}X_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t, X_t)(\mathrm{d}X_t)^2,$$

where we compute using the rules

$$(\mathrm{d}t)^2 = \mathrm{d}B_t \cdot \mathrm{d}t = \mathrm{d}t \cdot \mathrm{d}B_t = 0, \qquad \mathrm{d}B_t \cdot \mathrm{d}B_t = \mathrm{d}t.$$

Proof: This is essentially fancy Taylor expanding, and observing which terms go to zero. Fix $t > 0$. We'll assume that u, v are simple functions; for general u, v, we can approximate by simple functions and take limits to arrive to the final formula. Take a partitioning of $[0,t]$ given by $\{t_j\}_{j=1}^N$, and denote $\Delta t_j = t_{j+1} - t_j$, $\Delta X_j = X_{t_{j+1}} - X_{t_j}$. Then, we find

$$\begin{aligned} g(t, X_t) = g(0, X_0) &+ \sum_j \frac{\partial g}{\partial t}\Delta t_j + \sum_j \frac{\partial g}{\partial x}\Delta X_j \qquad (\mathcal{O}(1) \text{ terms}) \\ &+ \frac{1}{2}\sum_j \frac{\partial^2 g}{\partial t^2}(\Delta t_j)^2 + \sum_j \frac{\partial^2 g}{\partial t \partial x}(\Delta t_j)(\Delta X_j) + \frac{1}{2}\sum_j \frac{\partial^2 g}{\partial x^2}(\Delta X_j)^2 \qquad (\mathcal{O}(2) \text{ terms}) \\ &+ \sum_j R_j, \end{aligned}$$

where in each summation the ∂g terms are evaluated at (t_j, X_{t_j}), and the remainder term(s) $R_j = o\left(|\Delta t_j|^2 + |\Delta X_j|^2\right)$, by Taylor's theorem, and will in particular go to zero as $\Delta t_j \to 0$. Moreover, the first two terms will converge to $\int_0^T \frac{\partial g}{\partial t}(s, X_s)\,\mathrm{d}s$, $\int_0^T \frac{\partial g}{\partial x}(s, X_s)\,\mathrm{d}X_s$ respectively. For the $\mathcal{O}(2)$ terms, we know that since u, v simple,

$$(\Delta X_j)^2 = u_j^2(\Delta t_j)^2 + 2u_j v_j(\Delta t_j)(\Delta B_j) + v_j^2(\Delta B_j)^2,$$

where u_j, v_j correspond to u, v evaluated at t_j. Thus, the third $\mathcal{O}(2)$ term becomes

$$\sum_j \partial_x^2 g(\Delta X_j)^2 = \sum_j \partial_x^2 g u_j^2(\Delta t_j)^2 + \sum_j \partial_x^2 g(\Delta t_j)(\Delta B_j) + \sum_j \partial_x^2 g v_j^2(\Delta B_j)^2.$$

Because of the Δt_j^2 term in the first term and the fact that $\mathbb{E}\left[(\Delta B_j)^2\right] = \Delta t_j$, it's not hard to show that the first two terms converge to 0 as $\Delta t_j \to 0$. The last term, perhaps surprisingly, converges to $\int \partial_x^2 g v^2\,\mathrm{d}s$ (namely, it loses all dependence on the Brownian motion)! The crucial computation to show this is as follows. Put $a(t) = \partial_x^2 g(t)v^2(t)$ and so $a_j := a(t_j)$. Then,

$$\begin{aligned}
\mathbb{E}\left[\left(\sum_j a_j(\Delta B_j)^2 - \sum_j a_j \Delta t_j\right)^2\right] &= \sum_{i,j} \mathbb{E}\Big[a_i a_j\big((\Delta B_i)^2 - \Delta t_i\big)\big((\Delta B_j)^2 - \Delta t_j\big)\Big] \\
\text{by independence} \qquad &= \sum_j \mathbb{E}\left[a_j^2\big((\Delta B_j)^2 - \Delta t_j\big)^2\right] \\
&= \sum_j \mathbb{E}\big[a_j^2\big]\mathbb{E}\Big[(\Delta B_j)^4 - 2(\Delta B_j)^2 \Delta t_j + (\Delta t_j)^2\Big] \\
&= 2\sum_j \mathbb{E}\big[a_j^2\big](\Delta t_j)^2 \underset{\Delta t_j \to 0}{\to} 0.
\end{aligned}$$

Since the second term in the original expectation $\to \int a^2 \, \mathrm{d}s$, this completes the proof of the claimed convergence. A similar argument gives that the other two $\mathcal{O}(2)$ terms $\to 0$ as $\Delta t_j \to 0$. So, in all, we've shown that in the limit as $\Delta t_j \to 0$,

$$g(t, X_t) = g(0, X_0) + \int_0^T \partial_t g(s, X_s)\, \mathrm{d}s + \int_0^T \partial_x g(s, X_s)\, \mathrm{d}X_s + \frac{1}{2}\int_0^T \partial_x^2 g(s, X_s)\, \mathrm{d}s$$

□

We should, really, be as little more careful in all of our limit taking. Specifically, we aren't always explicit with what metric we are taking our limits with respect to, because usually its clear from context.

The last result we showed, people like to summarize by writing "$(\mathrm{d}B_t)^2 = \mathrm{d}t$*". This is, again, a nonsensical thing to write, but is a helpful heuristic.*

3 Stochastic Differential Equations

Unfortunately, it's typically hard to actually find explicit solutions to most SDEs of the form we've considered here (much like in the case of ODEs!). However, as in the ODE case, we have a existence and uniqueness theorem, which we present in one-dimension for simplicity:

Theorem 4 (Existence and Uniqueness for Scalar SDEs): Let $T > 0$, and $b, \sigma : [0,T] \times \mathbb{R} \to \mathbb{R}$ be measurable functions such that

$$|b(t,x)| + |\sigma(t,x)| \le C(1 + |x|)$$

for every $x \in \mathbb{R}, t \in [0,T]$ for some constant C, and

$$|b(t,x) - b(t,y)| + |\sigma(t,x) - \sigma(t,y)| \le D|x - y|$$

for every $x, y \in \mathbb{R}, t \in [0,T]$ for some constant D. Let Z be a random variable with distribution independent of the Brownian motion $\{B_s : s \ge 0\}$ and such that $\mathbb{E}[|Z|^2] < \infty$. Then, the SDE

$$\mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}B_t, \qquad t \in [0,T], X_0 = Z$$

admits a unique, time-continuous solution $X_t(\omega)$ with the property that

$$\mathbb{E}\left[\int_0^T |X_t|^2 \,\mathrm{d}t\right] < \infty.$$

Just as in the ODE case, its easy to construct SDEs that violate these conditions and admit multiple solutions. For instance, $\frac{\mathrm{d}X_t}{\mathrm{d}t} = 3X_t^{2/3}, X_0 = 0$ has an infinite number of solutions given by $\begin{cases} 0 & t \le a \\ (t-a)^3 & t > a \end{cases}$ for any $a > 0$.

Proof: We provide just a sketch. For existence, define a sequence of stochastic process by $Y_t^0 = X_0$ and inductively for $k \ge 0$,

$$Y_t^{(k+1)} = X_0 + \int_0^t b\left(s, Y_s^{(k)}\right) \mathrm{d}s + \int_0^t \sigma\left(s, Y_s^{(k)}\right) \mathrm{d}B_s.$$

One can then show

$$\mathbb{E}\left[|Y_t^{(k+1)} - Y_k^{(k)}|^2\right] \le \frac{A^{k+1} t^{k+1}}{(k+1)!},$$

where A a constant only dependent on C, D, T and $\mathbb{E}[|X_0|^2]$, all as given in the statement. Thus, for any $n, m \ge 1$, one applies the triangle inequality to find that

$$\left\|Y_t^{(m)} - Y_t^{(n)}\right\|_{L^2(\lambda \times \mathbb{P})} \le \sum_{k=n}^{m-1} \left(\frac{A^{k+1} T^{k+2}}{(k+2)!}\right)^{1/2},$$

where λ the Lebesgue measure on $[0,T]$ and the $L^2(\lambda \times \mathbb{P})$ norm is given by

$$\|X_t\|_{L^2(\lambda \times \mathbb{P})} := \mathbb{E}\left[\int_0^T |X_t|^2 \,\mathrm{d}t\right]^{1/2}.$$

This summation converges to zero as $n, m \to \infty$ hence $\left\{Y_t^{(m)}\right\}$ forms a Cauchy sequence in $L^2(\lambda \times \mathbb{P})$ and thus by completeness converges to some process X_t. From here, it's not hard to show that X_t actually does satisfy the SDE in question, then ensuring continuity is all that remains (this is a little tricky).

For uniqueness, suppose $X_t, \widehat{X}_t$ are two solutions. Then, one readily finds that

$$\mathbb{E}\left[|X_t - \widehat{X}_t|^2\right] \leq 3(1+t)D^2 \int_0^t \mathbb{E}\left[|X_s - \widehat{X}_s\ |^2\right] \mathrm{d}s,$$

so if we define $v(t) := \mathbb{E}\left[|X_t - \widehat{X}_t|^2\right]$, we find that there exists a constant A independent of time such that

$$v(t) \leq A \int_0^t v(s)\, \mathrm{d}s$$

for all $0 \leq t \leq T$. By the *Gronwall Lemma*, then, $v(t) \leq F \cdot \exp(At)$ where $F = v(0) = \mathbb{E}\left[|X_0 - \widehat{X}_0|^2\right] = 0$; hence, $v(t) = 0$ hence $X_t = \widehat{X}_t$ for a.e. $t \in [0, T]$. □

Let's actually solve an equation. We consider, in Itô process formulation,

$$\mathrm{d}N_t = rN_t\, \mathrm{d}t + \alpha N_t\, \mathrm{d}B_t,$$

where r, α are constants and N_0 is specified. General solution methods are difficult, but let's try to take advantage of the Itô Lemma to guess and verify. Suppose

$$g(t, x) := \ln(x).$$

Then

$$\begin{aligned}
\mathrm{d}(g(t, N_t)) = \mathrm{d}(\ln(N_t)) &= \frac{1}{N_t}\, \mathrm{d}N_t + \frac{1}{2}\left(-\frac{1}{N_t^2}\right)(\mathrm{d}N_t)^2 \\
&= \frac{\mathrm{d}N_t}{N_t} - \frac{1}{2N_t^2}\alpha^2 N_t^2\, \mathrm{d}t \\
&= \frac{\mathrm{d}N_t}{N_t} - \frac{1}{2}\alpha^2\, \mathrm{d}t \\
\Rightarrow \frac{\mathrm{d}N_t}{N_t} &= \mathrm{d}\ln(N_t) + \frac{1}{2}\alpha^2\, \mathrm{d}t.
\end{aligned}$$

So, we find that on the one hand,

$$\int_0^t \frac{\mathrm{d}N_s}{N_s} = rt + \alpha B_t$$

from the solution formula, while also

$$
\begin{aligned}
\int_0^t \frac{\mathrm{d}N_s}{N_s} &= \int_0^t \mathrm{d}\ln(N_s) + \int_0^t \frac{1}{2}\alpha^2\,\mathrm{d}s \\
&= \ln(N_t) - \ln(N_0) + \frac{1}{2}\alpha^2 t \\
&= \ln\left(\frac{N_t}{N_0}\right) + \frac{1}{2}\alpha^2 t,
\end{aligned}
$$

so equating these two we find

$$
\ln\left(\frac{N_t}{N_0}\right) = \left(r - \frac{1}{2}\alpha^2\right)t + \alpha B_t
$$

$$
\boxed{\Rightarrow N_t = N_0 \exp\big((r - \tfrac{1}{2}\alpha^2)t + \alpha B_t\big)}
$$

Without the B_t term in the original equation, we would have found solution

$$
N_t = N_0 \exp(rt).
$$

Generally, processes of the form

$$
X_t = X_0 \exp(\mu t + \alpha B_t)
$$

for constants μ, α are called "geometric Brownian motion". A sample path:

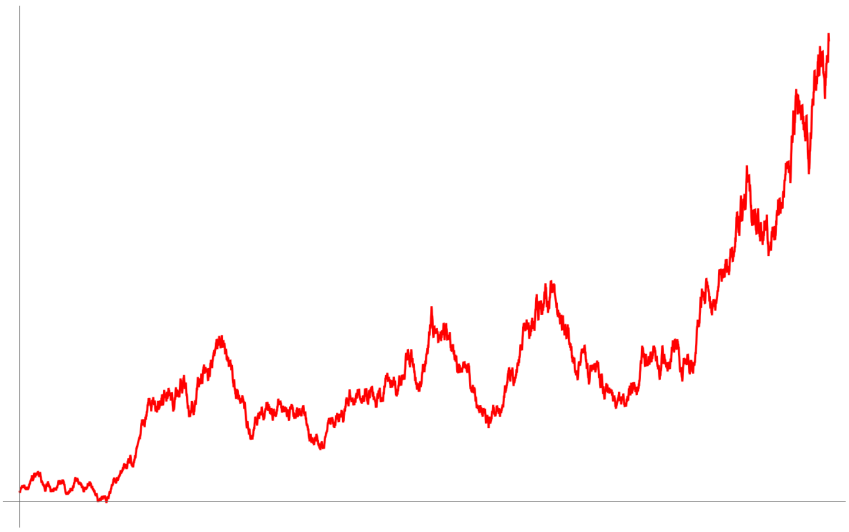


Figure 2: A (simulated) instance of geometric Brownian motion with $\mu = \alpha = 1$.

How much of an "effect" does the noise term have on the solution? Let us compute the "average location" of the solution at a given time t; i.e., the expected value of N_t. Assuming N_0 is independent of B_t for all time, we can write

$$
\begin{aligned}
\mathbb{E}[N_t] &= \mathbb{E}[N_0]\mathbb{E}\left[\exp\left(\left(r-\frac{1}{2}\alpha^2\right)t+\alpha B_t\right)\right] \\
&= \mathbb{E}[N_0]\cdot\underbrace{\exp\left(\left(r-\frac{1}{2}\alpha^2\right)t\right)}_{\text{deterministic}}\mathbb{E}[\exp(\alpha B_t)].
\end{aligned}
$$

From our earlier construction, we know $B_t \sim \mathcal{N}(0,t)$, so we can directly compute this last term using the formula for the pdf of such a distribution:

$$
\begin{aligned}
\mathbb{E}[\exp(\alpha B_t)] &= \frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}}\exp(\alpha x)\exp\left(-\frac{1}{2t}x^2\right)\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}}\exp\left(-\frac{1}{2t}\{x^2-2t\alpha x\}\right)\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}}\exp\left(-\frac{1}{2t}\left\{(x-t\alpha)^2-t^2\alpha^2\right\}\right)\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi t}}\exp\left(\frac{t\alpha^2}{2}\right)\int_{\mathbb{R}}\exp(-u^2)\sqrt{2t}\,\mathrm{d}u \\
&= \frac{1}{\sqrt{2\pi t}}\sqrt{2\pi t}\exp\left(\frac{t\alpha^2}{2}\right) = \exp\left(\frac{t\alpha^2}{2}\right).
\end{aligned}
$$

Hence, bringing everything back together,

$$
\mathbb{E}[N_t] = \mathbb{E}[N_0]\exp(rt).
$$

But this is precisely the formula we had for the deterministic case. In summary, while we have a different solution "pointwise" in the stochastic framework, the "average solution" is identical to the deterministic one.

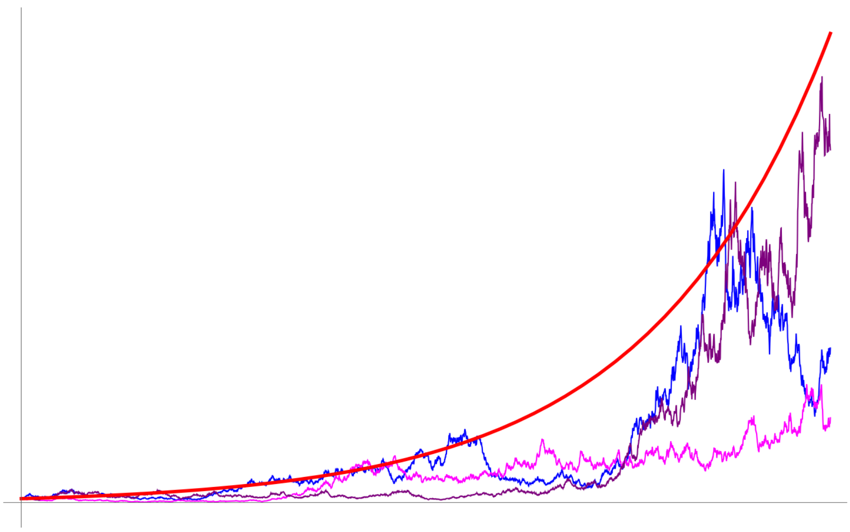


Figure 3: N_t (with $r=1$), deterministic situation (i.e. $\alpha=0$) in red, 3 instances of stochastic solution (with $\alpha=1$) in purple, blue, magenta.

Sources.

[1] B. Øksendal, "Stochastic Differential Equations: An Introduction with Applications." 2013.

[2] T. Hida, "Brownian Motion." 1980.

[3] D. Higham and P. Kloeden, "An Introduction to the Numerical Simulation of Stochastic Differential Equations." 2021.