

# 1 The Itô & Stratonovich integrals

Recently, I've been reading about stochastic calculus again. Something I found quite confusing was the existence of two formulations of the stochastic calculus; Itô and Stratonovich. When I first read about it, it seemed like there were two mathematical treatments of the same physical process that give different answers. As we will hopefully see, this isn't actually a very accurate description, but it's the impression you can still get from a lot of textbooks and writing on the subject. SDEs require some fairly heavy duty mathematics, and most of the work I was able to find on them wasn't massively accessible, and tends to focus on technical details over intuition. This article is designed to be as concise and simple a statement as possible of the origins of the two formulations, how they are related, and whether the difference really matters<sup>1</sup>. I've tried to avoid going into the measure-theoretic treatment and more technical proofs in favour of a high-level, intuitive overview that is still morally correct. There is a list of references if you want to get deeper into the details of the ideas I'm going over here.

This is a summary and overview, not original work. Most derivations etc. are based on those in the references; any mistakes are my own.

A pdf version of this page is available here if you prefer.

## 2 The problem

The basic problem addressed by SDEs is how to deal with noise in a physical process. That is, we have a system described by a standard differential equation

$$\frac{dx}{dt} = f(x, t) \quad (1)$$

and we want to extend this model to deal with the presence of intrinsic noise. Note that we are not talking about observation noise; the noise effects the evolution of the state  $x$ . Lets call the noise term (a random variable)  $\xi_t$  and we can imagine an equation that looks something like this;

$$\frac{dX}{dt} = f(X, t) + \text{"}\sigma(X, t)\xi_t\text{"} \quad (2)$$

where  $\sigma(x, t)$  is a deterministic function measuring our sensitivity to some standard noise. The reason for the quotation marks will become clear.

Of course, the deterministic equation has the solution

$$x(T) = x(0) + \int_{t=0}^T f(x, t)dt \quad (3)$$

and so we expect it's stochastic counterpart will have solution

$$X(T) = X(0) + \int_{t=0}^T f(X, t)dt + \text{"}\int_{t=0}^T \sigma(X, t)\xi_t dt\text{"} \quad (4)$$

The question, then, is how we assign meaning to the terms in quotation marks. We might want to start by associating our noise  $\xi_t$  to some random process. We may expect it has the following properties, approximately like that of normally distributed noise for discrete observations;

1. The noise is zero mean:  $\mathbb{E}[\xi_t] = 0$  for all  $t$ .
2. The noise is uncorrelated: for any  $t, t'$  such that  $t \neq t'$ ,  $\mathbb{E}[\xi_t \xi_{t'}] = 0$
3. The noise is stationary: for any set of times  $t_1, t_2, \dots, t_n$ , the joint distribution  $P(\xi_{t_1}, \xi_{t_2}, \dots)$  is the same as  $P(\xi_{t+t_1}, \xi_{t+t_2}, \dots)$  for any  $t$ .

---

<sup>1</sup>Spoiler; you should probably understand it, but not lose sleep over it

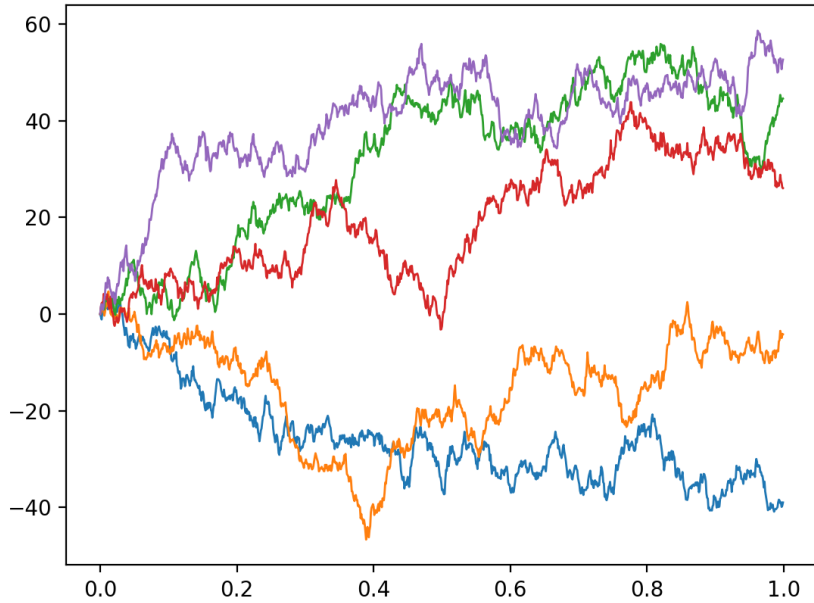


Figure 1: Sampled paths of Brownian motion starting at 0.

The reason we want the first condition is pretty obvious; if the noise had a non zero mean, then it would produce a drift over time, and we would rather treat the drift as part of the deterministic function  $f$ . Similarly, if the noise was non-stationary, we could absorb that into the function  $\sigma$ ; we want a ‘standard’ noise term. These conditions describe white noise, a random signal with no characteristic frequency.

This seems reasonable enough, and fairly well motivated by the noise one encounters in real systems (for example, TV static). However, continuous time white noise is actually quite a bizarre and poorly behaved mathematical object. It turns out to be impossible to construct a white noise process  $\xi_t$  which also has continuous paths, which seems like an important feature for defining well behaved differential equations. Indeed, a true white noise signal would have a completely flat Fourier spectrum, which would imply that it had infinite energy<sup>2</sup>. So we cannot give the equation a sensible meaning by setting  $\xi_t$  to a random process directly.

We need a different way, then, to interpret the noise term, if we are going to end up with a sensible mathematical model for noisy dynamics. One way to arrive at a more meaningful result is to split  $T$  into  $N$  discrete intervals and consider the discrete equation

$$X_T = X_0 + \sum_{t=1}^N f(t, X_t) \Delta t + \sum_{t=1}^N \sigma(t, X_t) \Delta W_t \quad (5)$$

where  $\Delta W_t = W_{t+1} - W_t$  are the increments of some random process  $W_t$ . Now, our assumptions imply that we would like the process to have stationary, independent increments with zero mean. It turns out there is only one such process with continuous paths: Brownian motion, also known as the Wiener process, which is a continuous time generalisation of a random walk.

The Wiener process can be thought of as the integral of white noise. It is differentiable nowhere; its derivative would be the white noise process discussed above. Brownian motion  $B_t$  is an example of a Gaussian process; for any finite collection of times  $t_1, t_2, \dots, t_n$ , the joint distribution of  $B_{t_1}, B_{t_2}, \dots$  is a multivariate Gaussian, with mean zero (for Brownian motion starting at zero) and a covariance matrix  $\Sigma$  whose entries are given by

<sup>2</sup>For a signal  $x(t)$ , the energy of the signal is proportional to  $\int_{-\infty}^{\infty} |x(t)|^2 dt$ , which is proportional to  $\int_{-\infty}^{\infty} |\hat{x}(f)|^2 df$  where  $\hat{x}$  is the Fourier transform of  $x$ . White noise would have a flat frequency spectrum, i.e  $\hat{x}(f) = \text{const.}$ . It’s easy to see this leads to the energy being infinite.

$$\Sigma_{ij} = \mathbb{E}[B_{t_i} B_{t_j}] = \min(t_i, t_j) \quad (6)$$

For brevity here, I'm going to mostly take the above as a sensible definition without going much into physical motivations, but Brownian motion has many physical applications, with the original one being the motion of a particle suspended in a fluid. The basic physical intuition behind the idea that  $B_t$  is an integral of white noise is that in an interval  $\Delta t$ , our particle gets hit by a large number of uncorrelated kicks in random directions with bounded variance, which will lead to a Gaussian change of momentum in that interval.

It follows directly from the above definition that Brownian motion has independent increments; if  $t_1 < t_2 < t_3 \dots$ , then

$$\mathbb{E}[(B_{t_{i+1}} - B_{t_i})(B_{t_{j+1}} - B_{t_j})] = \min(t_{i+1}, t_{j+1}) - \min(t_{i+1}, t_j) - \min(t_i, t_{j+1}) + \min(t_i, t_j) \quad (7)$$

$$= t_{i+1} - t_{i+1} - t_i + t_i \quad \text{when } t_i < t_j \quad (8)$$

$$= 0 \quad (9)$$

and the variance of increments is

$$\mathbb{E}[(B_t - B_{t'})^2] = \mathbb{E}[B_t^2] + \mathbb{E}[B_{t'}^2] - 2\mathbb{E}[B_t B_{t'}] \quad (10)$$

$$= t + t' - 2 \min(t, t') \quad (11)$$

$$= |t - t'| \quad (12)$$

Brownian motion is the only process with continuous paths that satisfies our list of criteria, so we will use as our discrete model

$$X_{k+1} - X_k = f(t, X_t)\Delta t + \sigma(t, X_t)\Delta B_t \quad (13)$$

$$X_T = X_0 + \sum_{t=1}^N f(t, X_t)\Delta t + \sum_{t=1}^N \sigma(t, X_t)\Delta B_t \quad (14)$$

in integral or derivative form. We can now try to go back to continuous time by taking the limit as we divide the interval into smaller and smaller pieces ( $N \rightarrow \infty$ ).

### 3 The stochastic integral

Let's quickly recall how the standard (Riemann) integral can be defined. We have some function  $f(t)$  we want to integrate over the range  $0, T$ . First, we partition the interval into  $t_0, t_1, t_2, t_3, \dots$ , and consider

$$\int_0^T f(t)dt = \lim_{N \rightarrow \infty} \sum_{k=0}^{N-1} f(\hat{t}_k)\Delta t_k, \quad \Delta t_k = t_{k+1} - t_k \quad (15)$$

where the width of each interval is  $T/N$ , and for each interval  $\Delta t_k$ , we choose a value  $\hat{t}_k \in [t_k, t_{k+1}]$  at which we evaluate the function. Intuitively: we replace the function with a series of rectangles, whose heights are  $f(\hat{t}_k)$  and whose widths are  $\Delta t_k$ , and then take the limit of the sum of the area of the rectangles as we make this approximation tighter and tighter.

For ordinary, smooth functions, of course, this integral converges to the same value no matter where we choose our rectangle heights  $\hat{t}_k$  in the interval. However, Brownian motion is less well behaved.

To be slightly more precise, the rectangular construction can be thought of as approximating  $f$  as a 'simple function', where simple functions are defined in the following way. Let  $1_A$  be the characteristic function of a set  $A$ ; that is  $1_A(x) = 1$  if  $x \in A$  and 0 otherwise. Then, the rectangular construction is the same as approximating  $f$  with a function of the form

$$f(t) \simeq \phi(t) = \sum_k \hat{f}_k 1_{[t_k, t_{k+1})}(t) \quad (16)$$

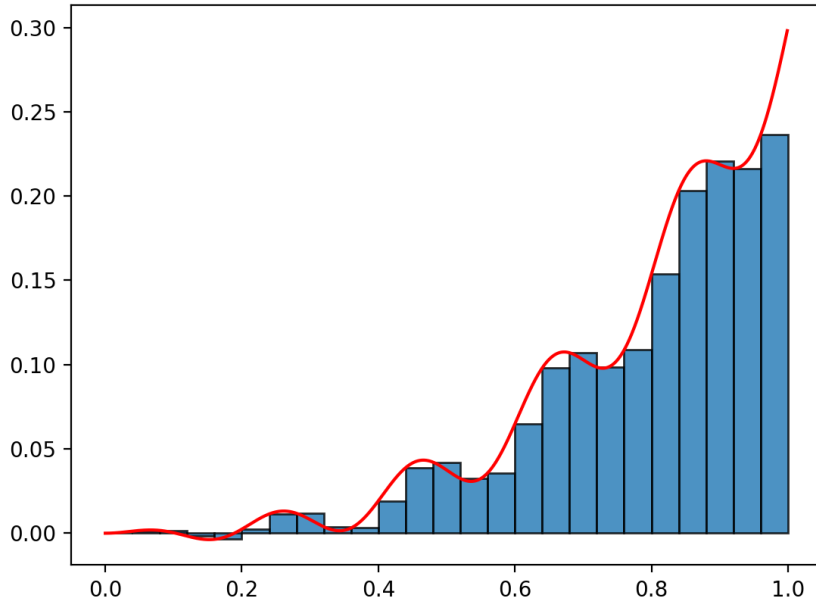


Figure 2: The Riemann integral construction; divide the function up into rectangles. Here, I've evaluated the function at the left edge of each rectangle.

where the  $\hat{f}$  are constants, chosen at some point in the interval as above. For these functions, the area under the curve can be computed very easily as a sum of rectangles. As we take the limit of the intervals becoming infinitesimally wide, the approximation to the function becomes exact and the sum converges to the value of the integral <sup>3</sup>

The obvious thing to do next is to try to extend the same idea to random processes. That is, we can consider functions of the form

$$\phi(t) = \sum_k \hat{e}_k \mathbf{1}_{[t_k, t_{k+1})}(t) \quad (17)$$

for which the integral can be defined in terms of discrete intervals in a straightforward way as

$$\int_0^T \phi(t) dB_t = \sum_k \hat{e}_k \Delta B_k \quad (18)$$

Note that here, the coefficients  $\hat{e}_k$  are not constants, as in the construction for deterministic functions, but random variables, as we are allowed to integrate functions  $\sigma$  which depend on  $X_t$ , which is also a random variable.<sup>4</sup> It doesn't have the same geometrically intuitive nature as the Riemann integral, but there is an analogy here; instead of a finite sum of simple functions which are rectangles, we are now dealing with a finite sum of 'simple random variables' multiplied by increments of a Brownian walk, which are Gaussian.

And we can try to proceed in the same way; having defined a class of functions whose integral is obvious, analogous to the rectangles of the Riemann construction, we can then consider approximating

<sup>3</sup>At least for smooth, well-behaved functions. There exist functions for which this simple picture is actually problematic; for example, consider the function  $\mathbf{1}_{\mathbb{Q}}$ , the indicator function of the rational numbers, which cannot be approximated well as a sum of rectangles. As mentioned, I am trying to avoid these kinds of details here.

<sup>4</sup>Again, I'm assuming here that random processes can be approximated by simple functions in a way that will converge to the true process. This is a non-trivial claim, and in a formal treatment the class of functions for which this is true would have to be defined rigorously. For example, the stochastic integral of the process  $X_t = B_{2t}$  cannot be defined in a sensible way, because this would mean that the current value of the integral depends on the value of the noise in the future. The full details are beyond the scope of this post, and can be found in the references, particularly Øksendal, chapter 3

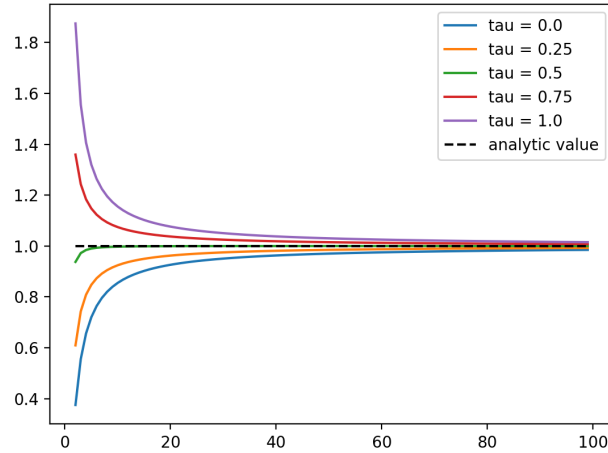


Figure 3: The convergence of the Riemann integral of  $\int_0^1 3x^2$  with  $N$ , the number of subdivisions of the interval, for various choices of evaluation point  $\tau$ . For the Riemann integral, this choice doesn't matter; all values of  $\tau$  converge quickly to the true value of the integral. This is not always the case for stochastic integrals.

the functions we wish to integrate as these simple functions, then letting the size of the interval tend to zero.

However, we will run into difficulties immediately. Unlike for continuous functions, whose sum converges wherever we choose the midpoints  $\hat{f}_k$ , we will see that we can get very different results depending on where we evaluate the integrated function in the random case. Let the evaluation point be  $\tau_k \in [t_k, t_{k+1})$ , and consider integrating  $B_t$  in this way, so  $\hat{e}_k = B_{\tau_k}$ . Then, if we consider the expectation, we have

$$\mathbb{E} \left[ \int_0^T B_t dB_t \right] = \mathbb{E} \left[ \sum_k B_{\tau_k} \Delta B_k \right] \quad (19)$$

$$= \sum_k \mathbb{E} [B_{\tau_k} (B_{t_{k+1}} - B_k)] \quad (20)$$

$$= \sum_k \min(t_{k+1}, \tau_k) - \min(t_k, \tau_k) \quad (21)$$

$$= \sum_k (\tau_k - t_k) \quad (22)$$

but this can be anything between 0 (for  $\tau_k = t_k$ ) and  $T$  (for  $\tau_k = t_{k+1}$ )! So without a choice of  $\tau$ , we are still unable to define the integral of a fairly simple stochastic process without ambiguity. In the above, we didn't bother writing that we were considering a limit, but it clearly doesn't affect the argument, as the different answers have no dependence on  $N$ .

The fix is basically to just make a choice, and simply define the integral to be the limit as evaluated for a choice of  $\tau$ . The Itô rule says to evaluate the integrand at the left endpoint, corresponding to  $\tau_k = t_k$ , while the Stratonovich rule says to evaluate at the midpoint, corresponding to  $\tau_k = \frac{1}{2}(t_{k+1} + t_k)$ . We could, in principle, choose  $\tau_k = \lambda t_k + (1 - \lambda)t_{k+1}$  for any value  $\lambda \in [0, 1]$ , but these two choices turn out to be the most natural and convenient ones.

## 4 Itô or Stratonovich?

The obvious question is which interpretation is the right one?. After all, we just saw that the equations give different answers. If we let

$$\int_0^T b(x, t) dB_t \quad (23)$$

denote the Itô integral and

$$\int_0^T b(x, t) \circ dB_t \quad (24)$$

denote Stratonovich's, then we have two options to interpret our original prototype of a stochastic differential equation. Itô says

$$X(T) = X(0) + \int_{t=0}^T f(X, t) dt + \int_{t=0}^T \sigma(X, t) dB_t \quad (25)$$

and Stratonovich says

$$X(T) = X(0) + \int_{t=0}^T f(X, t) dt + \int_{t=0}^T \sigma(X, t) \circ dB_t \quad (26)$$

and these describe different processes, as we have just seen. Surely one must be “better?”. Lets briefly look at the properties of the two integrals before we return to this question.

#### 4.1 Itô

The biggest advantage of Itô's formulation is that the evaluations of the function are totally uncorrelated with the increment, by construction. That is to say, that because we evaluate the function  $f(x, B_t, t)$  at the left point of the midpoint, our limiting sum is

$$\int f(x, B_t, t) dB_t = \lim_{N \rightarrow \infty} \sum_k f(x_{t_k}, B_{t_k}, t_k) (B_{t_{k+1}} - B_{t_k}) \quad (27)$$

But because the increment  $(B_{t_{k+1}} - B_{t_k})$  is uncorrelated with  $B_{t_k}$ , it's also uncorrelated with any function of  $B_{t_k}$ , and so this always has mean zero.

This leads to the nice property that an Itô integral

$$\int_S^T f(x, t) dB_t \quad (28)$$

always has mean zero, and is a martingale. A stochastic process  $M_t$  is a martingale if its mean is bounded, and the conditional expectation given access to it's history up to time  $t'$  is equal to the value at that time; that is

$$\mathbb{E}[|M_t|] < \infty \quad (29)$$

$$\mathbb{E}[M_t | \{M_s, s \leq t'\}] = M_{t'}, \quad \text{for all } t' \leq t \quad (30)$$

In less formal language, this means that the “best guess” for the future value of a martingale is always the most recent observed value; knowing  $M_{t'}$  is just as good as knowing the entire history of  $M_t$  up to  $t'$ . The origins of the term “martingale” are historical; they have been extensively mathematically analysed in the context of gambling<sup>5</sup>. If  $W_t$  is your wealth at the  $t^{\text{th}}$  round of a fair game, like betting on a coin flip, then  $W_t$  is a martingale; on average you end up with the same amount of money you started with.

It is slightly difficult to give a non-technical overview, but this property turns out to be very convenient mathematically, due to some nice theoretical properties of martingales. For instance, one can

<sup>5</sup>According to Wikipedia, martingale originally referred to an entertainingly poorly conceived class of gambling strategies, and later the name became attached to the games these strategies were designed for, with an important result in martingale theory being the impossibility of a winning strategy in a fair game of chance, given only bounded wealth and time.

make use of the martingale property when computing the conditional expectation of an Itô process. In general, it is easier to analyse and prove theorems about the Itô integral for this reason.

The choice of the function evaluation “before” the observation of noise also seems natural if you think of a stochastic process as being the limit of a discrete one as the time-per-round becomes infinitesimal, as we used above to derive an SDE. Obviously, in a discrete game, you have to make your decision before you observe new information, and the Itô integral makes the current function value independent of the current noise increment.<sup>6</sup>

On the other hand, the Itô integral leads to a modification of the chain rule of calculus for changing variables. If we have an Itô process

$$dX_t = f(X_t, t)dt + \sigma(X_t, t)dB_t \quad (31)$$

and we want the equation of a new random process  $Y_t$ , where  $Y_t = g(t, X_t)$ , and  $g$  is a twice differentiable function, the Itô SDE for  $Y_t$  is given by

$$dY_t = \frac{\partial g}{\partial t}(t, X_t)dt + \frac{\partial g}{\partial x}(t, X_t)dX_t + \frac{1}{2} \frac{\partial^2 g}{\partial x^2}(t, X_t)\sigma(X_t, t)^2 dt \quad (32)$$

the famous Itô’s lemma. Here, we have an additional second order term in the noise added to the standard chain rule. I won’t go through the derivation here, but I will provide a brief, hand-wavy explanation. From the properties of Brownian motion, we can see that

$$\mathbb{E}[\Delta B_k^2] = \Delta t_k \quad (33)$$

which implies that, as we take the limit

$$(dB_t)^2 = dt \quad (34)$$

The infinitesimal of  $dY_t$  is given by it’s Taylor expansion;

$$dY_t = \frac{\partial g}{\partial t}dt + \frac{\partial g}{\partial x}dX_t + \frac{1}{2} \frac{\partial^2 g}{\partial x^2}(dX_t)^2 + \dots \quad (35)$$

$$(36)$$

where the second order term in  $X_t$  is relevant because

$$(dX_t)^2 = f^2(dt)^2 + 2f\sigma dt dB_t + \sigma^2(dB_t)^2 \quad (37)$$

In the infinitesimal limit,  $dt^2$  and  $dt dB_t$  become negligible faster than  $dt$  and  $dB_t$ , but as we saw,  $(dB_t)^2$  is of the same order as  $dt$  and tends to it in probability. Since the Itô choice means that  $\Delta B_k$  is independent of  $f_k$  and  $\sigma_k$ , this leads to the rule  $dt^2 = dt dB_t = 0$ ,  $dB_t^2 = dt$ , from which we get the last term in Itô’s lemma.

## 4.2 Stratonovich

Itô’s lemma is a neat piece of mathematics, but it’s also quite awkward to work with. Stratonovich’s choice avoids this; his choice of evaluation point preserves the standard chain rule, but in the process necessarily gives up the useful martingale properties of the Itô integral.

I mentioned that Itô has a certain intuitive appeal if you think of a stochastic equation as the limit of a discrete time process. Similarly, the Stratonovich integral has a nice physical motivation. Recall,

---

<sup>6</sup>This occasionally leads to the (slightly misleading) claim that the Itô integral “does not look into the future”. This is of course true, but it somewhat implies that the Stratonovich formulation does, which is not the case. Both formulations lead to SDEs which depend only on the past. The non-anticipatory choice in Itô’s formulation does mean that the noise term is uncorrelated with the present value of  $f$ , which is what gives the Itô process it’s nice probabilistic properties. But one does not need the gift of prophecy to use Stratonovich’s formulation, as can be seen later from the fact that Stratonovich processes have equivalent Itô formulations and vice versa, which would be impossible if one class of equations was able to “look into the future” and the other was not.

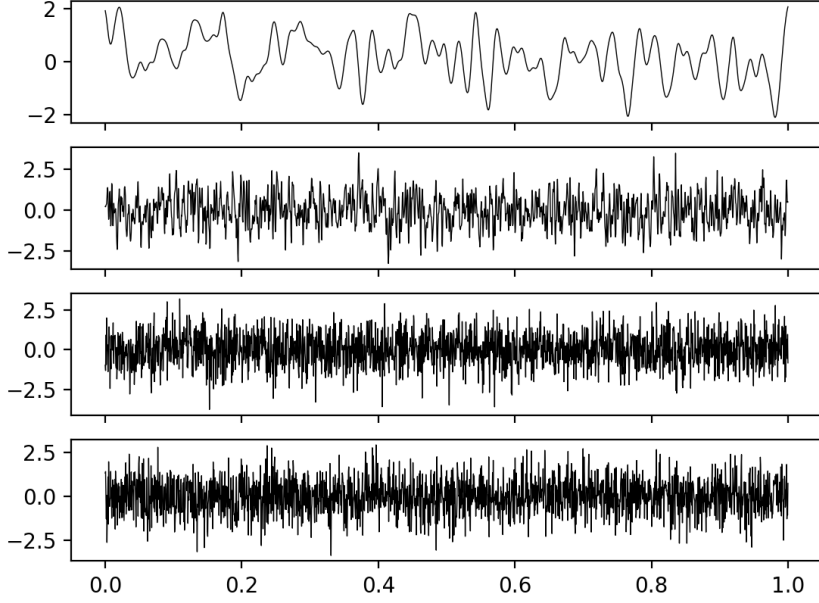


Figure 4: The first three figures are draws from a Gaussian process with a finite correlation length, with this length being decreased gradually. These have smooth, continuous paths. The bottom panel shows a draw from true white noise. As the length scale decreases, the finite correlation noise becomes increasingly similar to white noise, so we can think of white noise as an idealised limit of a series of smooth, continuous processes with increasingly small correlation time.

originally, one of the desired properties of white noise  $\xi$  that we wanted was total independence w.r.t time

$$\mathbb{E}[\xi_t \xi_{t'}] = 0, \quad t \neq t' \quad (38)$$

which prevents you from constructing a noise process with continuous paths, and leads to the absurd requirement that we have a signal with a flat power spectrum for all frequencies. The white noise condition above says that for any  $\epsilon$ ,  $\xi_t$  and  $\xi_{t+\epsilon}$  are totally statistically independent. A more realistic version of this condition is that the characteristic time scale of correlation is finite, but very small, at least compared to the time scale on which we are measuring the system. For example, consider the noise induced in a small object suspended in a fluid, the canonical example of a physical random walk. It's obvious that while the kicks from particle collisions happen on a much smaller timescale than the macro motion of the particle, they will have some finite, though tiny, correlation time since the particles in the fluid move at finite speed.

If we allow our noise process to have some finite length-scale  $\tau$ , then we can draw a process with continuous paths. To be concrete, consider the squared-exponential process, a Gaussian process with covariance

$$\Sigma_{ij} = \exp\left[-\frac{|t_i - t_j|^2}{\tau^2}\right] \quad (39)$$

Clearly, as  $\tau \rightarrow 0$ , this approaches the idealised white noise process  $\xi$ , as can be seen in the figure above. We can therefore consider having some process  $B_t^\tau$ , such that as  $B_t^\tau \rightarrow B_t$  as  $\tau \rightarrow 0$ , and consider what the process

$$\frac{dX_t^\tau}{dt} = f + \sigma \frac{dB_t^\tau}{dt} \quad (40)$$



converges towards. Making this statement more rigorous requires going back to measure theory, and considering the equation that is conditioned on an element of the event space  $\Omega$ , that is  $B_t^\tau(\omega)$ . This equation is then a standard ODE for all  $\tau$ , since  $B_t^\tau(\omega)$  are deterministic, continuous functions, and one can show that the  $X_t^\tau(\omega)$  converges to  $X_t(\omega)$  for almost all  $\omega$ . More intuitively, we can imagine that the unknown ODE input  $B_t^\tau$  is picked ahead of time, meaning that the system is just a normal ODE with an unobservable state component. These are random ODE's in the sense that they have an unknown input, but their mathematical behaviour is the same as standard ODEs. It turns out that the solution to these equations converges to the Stratonovich equation

$$dX_t = f dt + \sigma \circ dB_t \quad (41)$$

in the limit of white noise ( $\tau \rightarrow 0$ ). So the physical interpretation of the Stratonovich process is as the limit of a continuous time process with coloured noise. Intuitively, this is why Stratonovich preserves the normal chain rule; because we define it as the limit of a series of ODEs, which obviously obey the standard chain rule. However, Stratonovich integrals are not Martingales - with the evaluation of  $f$  at the midpoint of the interval,  $f$  and  $\Delta B_t$  are no longer independent by construction.

## 5 Does it matter?

I've tried to give an idea of the "philosophical" justification for the approaches taken by the two formulations. To an extent, the popularity of the two reflects this; the Itô integral is more popular in mathematics and finance, where the interpretation as the limit of a discrete game is somewhat appealing, and (more importantly) the martingale property is convenient. Stratonovich's rule is more popular in physics, where the limit of smooth noise argument is more compelling. However, it's easy to exaggerate the importance of this. In particular, the two are mathematically equivalent in the following sense; any Stratonovich process

$$dX_t = f(X_t, t)dt + \sigma(X_t, t) \circ dB_t \quad (42)$$

has an equivalent Itô process with identical solutions, which is given by

$$dX_t = f(X_t, t)dt + \sigma(X_t, t)dB_t + \frac{1}{2} \frac{\partial \sigma}{\partial x}(X_t, t)\sigma(X_t, t)dt \quad (43)$$

This formula holds in both directions. So if we already have a well defined SDE, either in the Itô or Stratonovich sense, then we can convert between the two conventions arbitrarily, depending on which properties we feel are more convenient for the problem at hand. Notice that the conversion formula basically comes down to a modification of the drift ( $dt$ ) term.

The only time 'interpretation' comes into it is if we start with the "pre equation" model

$$\frac{dX}{dt} = f(X, t)dt + \sigma(X, t)\xi_t \quad (44)$$

where we just know our system should be "noisy". As we have seen, this equation as written has no real rigorous meaning. If we know  $f$  and  $\sigma$ , however, we could reasonably choose to interpret it either as the Itô SDE

$$dX_t = f(X_t, t)dt + \sigma(X_t, t)dB_t \quad (45)$$

or as

$$dX_t = f(X_t, t)dt + \sigma(X_t, t) \circ dB_t \quad (46)$$

which are not equivalent, in general, as we have seen.

The only time this is really relevant, though, is if we have some prior knowledge of the “noise-free”  $f$ , and we would like the  $f$  of the stochastic equation to match, in which case choosing the “correct” interpretation is important. An example would be in electronic engineering, say, where we may have a strong understanding of the noise-free case, and where the Stratonovich interpretation of the noise is much more physically compelling. However, an Itô formulation of the noise process is still possible. In this case, though, the deterministic component of the SDE may not match the  $f$  we would have expected from the noise free analysis (because of the modification to the drift term in the conversion formula above).

However, the situation where we somehow “know”  $f$  and  $\sigma$  a priori, and have to make a deep philosophical decision as to whether interpreting the noise as the limit of a discrete time process or as the limit of coloured noise, is slightly unrealistic, occurring mostly in SDE textbooks. In reality, we can choose either formulation, then we will fit it to data; in other words, we have  $X_t$  as given, not  $f, \sigma$ , and we need to choose the  $f, \sigma$  that will give us the answers we observe. Our choice of formulation will not affect the behaviour of the system, but it may change the interpretation of the drift and diffusion coefficients in our model. To make this concrete, let’s go through an example.

## 6 An example - a noisy population growth equation

An illustrative example is the simplest form of the noisy population growth equation

$$\frac{dN}{dt} = N(r + \gamma\xi_t) \quad (47)$$

where  $r$  is the (constant) average growth rate for a population of size  $N$ ,  $\gamma$  is a constant, and  $\xi_t$  is white noise as before. This can be generalised to let  $r, \gamma$  depend on  $t$  or  $N$ , but the simplest case will be sufficient for us here. In standard form, the corresponding SDE is

$$dN_t = N_t r dt + \gamma N_t dW_t \quad \text{where } dW_t = dB_t \text{ or } \circ dB_t \quad (48)$$

depending on whether we choose the Itô or Stratonovich interpretation of the noise (remember that without a specification, the equation involving  $\xi_t$  doesn’t really have any meaning).

For now, choose the Itô convention. In this case we have that,

$$\int_0^t \frac{dN_t}{N} = rt + \gamma B_t \quad (49)$$

if we set  $B_0 = 0$ . Now, we can evaluate this by applying Itô’s lemma to  $g(x, t) = \ln(x)$ , which gives

$$d(\ln N_t) = \frac{dN_t}{N_t} - \frac{1}{2}\gamma^2 dt \quad (50)$$

and so

$$\ln \frac{N_t}{N_0} = (r - \frac{1}{2}\gamma^2)t + \gamma B_t \quad (51)$$

$$N_t = N_0 \exp[(r - \frac{1}{2}\gamma^2)t + \gamma B_t] \quad (52)$$

If we let  $\bar{N}_t$  be the solution to the Stratonovich interpretation, we would have obtained

$$\bar{N}_t = \bar{N}_0 \exp[rt + \gamma B_t] \quad (53)$$

as the  $dt$  term in the log expansion comes from Itô’s lemma, which does not apply in this case. Note these are both processes of the form

$$X_t = X_0 \exp(\alpha t + \beta B_t) \quad (54)$$

and the difference between the two interpretations comes down to a modification of  $\alpha$ .

This can then be used to give some predictions about the long term behaviour of the solution. We can use the law of the iterated logarithm, which states that for a random walk,

$$\limsup_{t \rightarrow \infty} \frac{B_t}{\sqrt{2t \log \log t}} = 1 \quad (55)$$

almost surely<sup>7</sup>. This implies that  $B_t$  will be bounded by  $\sqrt{2t \log \log t}$  as  $t \rightarrow \infty$ . Since this is dominated by  $t$  for large  $t$ , we have that the  $t$  term in the exponential will dominate the long term behaviour unless  $\alpha = 0$ . This implies that (almost surely)

$$\begin{aligned} \alpha > 0 &\Rightarrow N_t \rightarrow \infty \text{ as } t \rightarrow \infty \\ \alpha < 0 &\Rightarrow N_t \rightarrow 0 \text{ as } t \rightarrow \infty \end{aligned}$$

This seems to be a dramatic illustration of the differences between the approaches; Itô says that  $N_t \rightarrow 0$  almost surely (i.e the population will eventually become extinct with certainty) if  $r - \frac{1}{2}\gamma^2 < 0$ . Stratonovich, though, predicts the same as the non-noisy case; the population will tend to extinction only if  $r < 0$ .

Let's look at another difference. What is  $\mathbb{E}[N_t]$ ? Again, let's do the Itô process; we have

$$\mathbb{E}[N_t] = N_0 \exp\left(\left(r - \frac{1}{2}\gamma^2\right)t\right) \mathbb{E}[\exp(\gamma B_t)] \quad (56)$$

By using Itô's formula on  $Y_t = \exp(\gamma B_t)$ , we get

$$dY_t = \gamma \exp(\gamma B_t) dB_t + \frac{1}{2} \gamma^2 \exp(\gamma B_t) dt \quad (57)$$

$$\Rightarrow Y_T = Y_0 + \gamma \int_0^T Y_t dB_t + \frac{1}{2} \gamma^2 \int_0^T Y_t dt \quad (58)$$

Using that Itô integrals have mean 0, we have

$$\mathbb{E}[Y_T] = \mathbb{E}[Y_0] + \frac{1}{2} \mathbb{E}[\gamma^2 \int_0^T Y_t dt] \quad (59)$$

$$\Rightarrow \frac{d}{dt} \mathbb{E}[Y_t] = \frac{1}{2} \gamma^2 \mathbb{E}[Y_t] \quad (60)$$

$$\Rightarrow \mathbb{E}[Y_t] = e^{\frac{1}{2} \gamma^2 t} \quad (61)$$

since  $Y_0 = 1$ . This implies that  $\mathbb{E}[N_t] = N_0 e^{rt}$ , the same as the deterministic case, despite the slightly odd look of the Itô solution. For the Stratonovich formulation, one instead obtains

$$\mathbb{E}[\bar{N}_t] = \exp\left(r + \frac{1}{2}\gamma^2\right)t \quad (62)$$

which is different from the deterministic case. On the other hand, if we look at the expectation of  $\log N_t$ , we find that

$$\mathbb{E}[\log N_t] = \log N_0 + \left(r - \frac{1}{2}\gamma^2\right)t \quad (63)$$

$$\mathbb{E}[\log \bar{N}_t] = \log \bar{N}_0 + rt \quad (64)$$

which makes Itô different from the deterministic case.

So which one is right? It's not at all clear whether population growth is 'really' the limit of a discrete time process or the limit of a coloured noise process, so there doesn't seem to be a strong reason to

---

<sup>7</sup>'Almost surely' means 'with probability 1'

prefer one interpretation or the other. And the two interpretations give very different predictions, even up to whether a species will go extinct? Isn't this a huge problem?

Actually, the confusion<sup>8</sup> here comes, as I mentioned, from a hidden assumption. We have assumed that we have been given  $r$  and  $\gamma$  from on high, and asked to choose which kind of noise to apply. This is not realistic; in reality, we would infer the  $r$  and  $\gamma$  from some empirical data. The system is fixed, and the choice of a formulation will lead to a different interpretation of  $r$  and  $\gamma$ , rather than the other way around!

Indeed, I have been slightly disingenuous here in using the same letters  $r$  and  $\gamma$  for both processes  $N_t$  and  $\bar{N}_t$ . Let's call the Itô coefficients  $r_i, \gamma_i$  and the Stratonovich  $r_s, \gamma_s$ . Now, what do the parameters mean? It's clear  $r$  should be some kind of average instantaneous growth rate. If we let the population at a time be  $x$ , and consider the instantaneous average conditioned on  $N_t = x$ , we have that the arithmetic average growth rate is

$$R_a(x) = \lim_{\Delta t \rightarrow 0} \mathbb{E}\left[\frac{N_{t+\Delta t} - N_t}{\Delta t} \mid N_t = x\right] \quad (65)$$

$$= \frac{1}{x} \lim_{\Delta t \rightarrow 0} \frac{\mathbb{E}[N_{t+\Delta t} \mid N_t = x] - x}{\Delta t} \quad (66)$$

Now, again treating the Itô solution, we have that, using the solution we worked out earlier

$$N_{t+\Delta t} \mid (N_t = x) = x \exp\left[\left(r_i - \frac{1}{2}\gamma_i^2\right)\Delta t + \gamma_i(B_{t+\Delta t} - B_t)\right] \quad (67)$$

Because  $B_{t+\Delta t} - B_t$  follows a Gaussian distribution with variance  $\Delta t$ , we see that this follows a log-normal distribution, with parameters  $\mu = \log x + \left(r_i - \frac{1}{2}\gamma_i^2\right)\Delta t$  and  $\sigma^2 = \gamma_i^2\Delta t$ . Then we can use the standard formula for the mean of a log normal distribution

$$\mathbb{E}[N_{t+\Delta t} \mid N_t = x] = \exp\left(\mu + \frac{\sigma^2}{2}\right) \quad (68)$$

$$= x \exp(r_i \Delta t) \quad (69)$$

So we can see that the arithmetic average growth rate is equal to the Itô growth parameter;

$$R_a(x) = \frac{1}{x} \lim_{\Delta t \rightarrow 0} \frac{x \exp(r_i \Delta t) - x}{\Delta t} \quad (70)$$

$$= \frac{1}{x} \lim_{\Delta t \rightarrow 0} \frac{x(1 + r_i \Delta t + \mathcal{O}(\Delta t^2)) - x}{\Delta t} \quad (71)$$

$$= r_i, \quad (72)$$

If we perform a similar calculation for the Stratonovich solution, we obtain that  $r_s = R_a - \frac{1}{2}\gamma_s^2$ . Note that we can also get this simply by using the Itô/Stratonovich conversion formula above.

One can also show that if you consider the geometric average  $R_g$  instead of the arithmetic, which is arguably more natural for geometric growth, then you find that the Stratonovich rate  $r_s = R_g$ , and  $r_i = R_g + \frac{1}{2}\gamma_i^2$ .

The key point here is that the solutions give identical predictions when they are phrased in terms of the actual measurable quantities in the system, the average growth rate for some specified average. Recall that we had above that Itô predicted extinction if  $r_i - \frac{1}{2}\gamma_i^2 < 0$ , while Stratonovich predicted it for  $r_s < 0$ . We see, then, that in fact Itô and Stratonovich completely agree; the population will tend to extinction if the geometric average growth rate is negative, or if the arithmetic average is smaller than half the square of the variance parameter  $\gamma$  (which is the same in both formulations, as can be seen by comparing the  $B_t$  term in the solutions).

<sup>8</sup>The confused are in good company; even Øksendal's book, something of a standard text for SDEs, is a little misleading here, in my opinion. The overview of this example is taken from chapter 5 in his book. He writes 'The two solutions have fundamentally different properties and it is an interesting question which one gives the best description of the situation'.

## 7 Conclusion

We have given an overview of the difference between the two formulations of the stochastic integral. We have seen that the differences between them, though they appear dramatic at first glance, are in fact largely semantic.

Both the Itô and Stratonovich formalisms have a ‘physical’ motivation; Itô as the limit of a discrete random process as the length of the discrete intervals tends towards zero, and Stratonovich as the limit of a process with coloured noise as the frequency spectrum tends to being uniform. It should be clear that these are both idealisations. While this is sometimes relevant if one of these limit processes is clearly more appropriate than the other, this is of secondary importance compared to which choice has more convenient mathematical properties for the task at hand. Often, neither of the limit processes is particularly compelling, as in the population growth example. As we have seen, while having to choose can appear to have a dramatic effect on the solutions, the difference between the solutions is not really a ‘physical’ difference, but only changes how we interpret the coefficients in our models.

The difference between the two is often framed as ‘given  $f$  and  $\sigma$ , we need to choose Itô or Stratonovich, and we will get a different  $X_t$  depending on our choice’. A more accurate way to think about it would be that we have  $X_t$ , the true behaviour of our system, and our choice of formalism determines how we will choose  $f$  and  $\sigma$ . Paradoxical differences can be eliminated under sufficiently careful treatment. We must remember that the tail does not wag the dog; we need to make a choice between the Itô and Stratonovich in order to make our mathematics work; either is fine so long as we are consistent, and the system we are modelling is not likely to be affected by our choice of convention.

## 8 References

The discussion on the population growth example is largely taken from [4], which is an excellent resource, and develops and generalises the brief version presented here.

[2] is a physics influenced exposition of the argument and is quite clear, as well as entertainingly written.

[1] is something of a canonical reference for SDE theory; it treats the controversy briefly, but I draw on it heavily for the the formulation of the Itô integral. Population growth is also used as an example in this book, though he does not mention the resolution in terms of the actual meaning of the parameters under both formulations.

[3] is an older textbook, with more of a physic-ish than maths-ish inclination. I occasionally found this book invaluable as a way to get a more intuitive feel for some of the more abstract stuff.

1. Øksendal, Bernt. "Stochastic differential equations." Springer, Berlin, Heidelberg, 2003
2. Van Kampen, N. G. "Itô versus Stratonovich." *Journal of Statistical Physics* 24.1 (1981): 175-187.
3. Gardiner, Crispin W. "Handbook of stochastic methods for physics, chemistry and the natural sciences", Springer, Berlin, Heidelberg, 1986
4. Braumann, Carlos A. "Itô versus Stratonovich calculus in random population growth." *Mathematical biosciences* 206.1 (2007): 81-107.