THE WONG-ZAKAI THEOREM

LIAM HODGKINSON

A one-dimensional SDE can be thought of as the limit of stochastic processes satisfying

$$X_{t+h} = X_t + h\mu(t, X_t) + \sqrt{h} \cdot \sigma(t, X_t) \cdot Z_t,$$

with $Z_t \sim \mathcal{N}(0,1)$ as $h \to 0$. In this case, the standard differential notation for this process in the Ito form is

$$dX_t = \mu(t, X_t)\, dt + \sigma(t, X_t)\, dW_t.$$

The key approach to solving SDEs analytically is Ito's lemma, which is a form of chain rule for SDEs. For this process X_t, there is

$$df(t, X_t) = \left(f_t + \mu f_x + \frac{1}{2}\sigma^2 f_{xx}\right) dt + \sigma f_x dW_t.$$

What makes Ito's lemma different from a regular chain rule is the correction term $\frac{1}{2}\sigma^2 f_{xx} dt$. As an example, by change of variables $Y_t = \log X_t$, it is found that for X_t satisfying geometric Brownian motion

$$dX_t = \mu X_t dt + \sigma X_t dW_t,$$

with μ, σ constant, we have that

$$X_t = X_0 \exp\left[\left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W_t\right],$$

where the correction term is

$$C(t) = \exp\left(-\frac{1}{2}\sigma^2 t\right).$$

Usually, it is quite difficult to solve an SDE analytically, so we attempt to simulate one instead. The standard (simplest) scheme is Euler's scheme where we just use the discretisation above:

$$X_{t+h} = X_t + h\mu(t, X_t) + \sqrt{h} \cdot \sigma(t, X_t) \cdot Z_t,$$

for some fixed step size h. This is an analogue of Euler's scheme for ODEs, which is generally quite terrible. However, Maruyama did show that if μ, σ are smooth, then

$$\mathbb{E}\left[\left\|X_t - X_t^h\right\|^2\right] \le Ch,$$

for some $C > 0$ where X_t^h is the Euler approximation of X_t. We note that since this scheme relies on previous terms, the error compounds exponentially, so a smaller error term (or error term of smaller order) is essential here. An alternative is the implicit Euler scheme given by

$$X_{t+h} = X_t + h\mu(t, X_{t+h}) + \sqrt{h} \cdot \sigma(t, X_t) \cdot Z_t,$$

which works better for stiff equations, but generally seems difficult since the inversion assumes some degree of regularity. The most popular improvement to this scheme is Milstein's method.

Proposition (Milstein Method). *The diffusion process X_t can be discretised by*

$$X_{t+h} = X_t + h\mu(t, X_t) + \sigma(t, X_t)\sqrt{h} \cdot Z_t + \sigma_x(t, X_t)\sigma(t, X_t)\left(Z_t^2 - 1\right) \cdot \frac{h}{2},$$

for a fixed step size h.

Proof. The goal here is to use Ito's lemma in such a way that we transform the SDE into

$$dY_t = A(t, X_t)\,dt + dW_t,$$

so that we don't have to approximate the Ito integral. This is accomplished when $Y_t = F(X_t)$ with $F_x = \sigma(t, x)^{-1}$. By Ito's lemma,

$$dY_t = \left(\frac{\mu(t, X_t)}{\sigma(t, X_t)} - \frac{1}{2}\sigma_x(t, X_t)\right) dt + dW_t,$$

Using a Taylor series expansion

$$\begin{aligned} dX_t &= F^{-1}(Y_t + dY_t) - F^{-1}(Y_t) \\ &= \sigma(t, X_t)\,dY_t + \frac{1}{2}\sigma_x(t, X_t)\sigma(t, X_t)(dY_t)^2 + o\left[(dY_t)^2\right], \end{aligned}$$

and so

$$\begin{aligned} dX_t = &\left[\mu(t, X_t) - \frac{1}{2}\sigma_x(t, X_t)\sigma(t, X_t)\right] dt + \sigma(t, X_t)\,dW_t \\ &+ \frac{1}{2}\sigma_x(t, X_t)\sigma(t, X_t)(dY_t)^2. \end{aligned}$$

Since $(dY_t)^2 = (dW_t)^2 = dt$, we get our original SDE back. If we take an Euler approximation at this point, replacing dW_t with $\sqrt{h} \cdot Z$, we get Milstein's method. Alternatively, if we take an Euler approximation of only the first order terms, we get the Stratanovich-Euler scheme. □

Kloedon and Platen have shown that

$$\mathbb{E}\left[\left\|X_t - X_t^h\right\|^2\right] \le Ch^2,$$

which is of higher order than Euler's scheme. However, while the stochastic aspect is better approximated, the deterministic term is still problematic. It would be nice if we could take all of the already established and powerful methods from numerical ODE theory and apply them to solving SDEs. Enter Wong and Zakai, who proved the following result:

Theorem (The Wong-Zakai Theorem). *Let $(\Omega, \mathcal{E}, \mathbb{P})$ be a probability space and suppose that:*

(1) *μ, σ, σ_x, σ_t are continuous for $x \in \mathbb{R}$ and $t \in [a, b]$.*
(2) *μ, σ and σ_x are Lipschitz continuous with Lipschitz constant $k > 0$.*
(3) *$|\sigma(x, t)| \ge \beta > 0$ for some $\beta > 0$ and $|\sigma_t(x, t)| \le k\sigma^2(x, t)$.*

Let $\{w_n(t)\}_{n=1}^{\infty}$ be a sequence of regular approximates to a Wiener process W_t satisfying (for $\omega \in \Omega$):

(4) *For each n, $w_n(t, \omega)$ has bounded variation, is continuous and piecewise differentiable in t.*

(5) *There exists some random variables* n_0, k *such that* $w_n(t,\omega) \leq k(\omega)$ *for almost every* $\omega \in \Omega$ *and all* $t \in [a,b]$ *when* $n > n_0(\omega)$.

(6) $w_n(t)$ *converges to* W_t *almost surely.*

If for each $n \in \mathbb{N}$, $x_n(t)$ *is the solution to the ordinary differential equation*

$$\frac{dx_n}{dt} = \mu(t,x_n) - \frac{1}{2}\sigma(t,x_n)\sigma_x(t,x_n) + \sigma(t,x_n)\cdot\frac{dw_n}{dt}$$
$$x_n(a) = x_a$$

on $[a,b]$, *then* $x_n(t)$ *converges almost surely for* $t \in [a,b]$ *to a stochastic process* X_t *satisfying the stochastic differential equation*

$$dX_t = \mu(t,X_t)\,dt + \sigma(t,X_t)\,dW_t$$
$$X_a = x_a$$

as $n \to \infty$.

Theorem (The Wong-Zakai Theorem (Abridged)). *Let* $\mu, \sigma \in C^1$ *be sufficiently regular and let* $\{w_n(t)\}_{n=1}^{\infty}$ *be a sequence of bounded, continuous and piecewise differentiable approximations which converge uniformly almost surely to a Wiener process* W_t. *If for each* $n \in \mathbb{N}$, $x_n(t)$ *is the solution to the ODE*

$$\frac{dx_n}{dt} = \mu(t,x_n) - \frac{1}{2}\sigma(t,x_n)\sigma_x(t,x_n) + \sigma(t,x_n)\cdot\frac{dw_n}{dt}$$
$$x_n(a) = x_a$$

almost everywhere on $[a,b]$, *then* $x_n(t)$ *converges almost surely uniformly in* $t \in [a,b]$ *to a stochastic process* X_t *satisfying the SDE*

$$dX_t = \mu(t,X_t)\,dt + \sigma(t,X_t)\,dW_t$$
$$X_a = x_a$$

as $n \to \infty$.

The proof is actually not very difficult, but we don't have enough time to prove the full result. Instead, we will show a similar (simpler) result, which should hopefully demonstrate why the Stratonivich term is necessary a la the Milstein method.

Theorem (Approximating the Ito Integral). *Let* $\sigma(t,x) \in L^1$ *be continuous and differentiable in* x *and let* $\{w_n(t)\}_{n=1}^{\infty}$ *be a sequence of regular approximations which converges almost surely to a Wiener process* W_t. *Then, almost surely*

$$\lim_{n\to\infty}\int_a^b \sigma(t,w_n(t))\,dw_n(t) = \frac{1}{2}\int_a^b \sigma_x(t,W_t)\,dt + \int_a^b \sigma(t,W_t)\,dW_t.$$

Proof. Let $F(t,x) = \int_a^x \sigma(t,s)\,ds$ and observe that by chain rule and FTC

$$\frac{d}{dt}F(t,x(t)) = F_t(t,x(t)) + x'(t)\,F_x(t,x(t))$$
$$= F_t + \sigma x',$$

and so (for $dw_n(t) = w_n'(t)\,dt$)

$$F(b, w_n(b)) - F(a, w_n(a)) = \int_a^b \sigma(t, w_n(t))\,dw_n(t) + \int_a^b F_t(t, w_n(t))\,dt.$$

From almost sure convergence of w_n to W it follows that

$$\lim_{n\to\infty} \int_a^b \sigma(t, w_n(t))\,dw_n(t) = F(b, W_b) - F(a, W_a) - \int_a^b F_t(t, W_t)\,dt,$$

almost surely. But from Ito's lemma

$$F(b, W_b) - F(a, W_a) = \int_a^b \left[F_t(t, W_t) + \frac{1}{2}\sigma_x(t, W_t)\right] dt + \int_a^b \sigma(t, W_t)\,dW_t.$$

□

The trick for the SDE form is similar, although it is more involved (since convergence in the solutions requires a variation of the Gronwall inequality). In that case, we use $F(t,x) = \int_a^x \sigma(t,s)^{-1}\,ds$ and equate coefficients between chain rule and Ito's lemma, just like in the Milstein method proof. It is worth noting that the Wong-Zakai theorem is equivalent to Ito's lemma under these regularity conditions, in the sense that one can be proved from the other.
It can be verified that the coefficients for geometric Brownian motion satisfy these conditions. The natural choice for w_n is a linear spline between points generated by Algorithm 5.15 in the Handbook of Monte Carlo Methods:

Algorithm 1 Generating the Wiener process

(1) Let $0 = t_0 < t_1 < t_2 < \cdots < t_n$ be the set of distinct times for which simulation of the process is desired.
(2) Generate $Z_1, \ldots, Z_n \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ and output:

$$W_{t_k} = W_{t_{k-1}} + \sqrt{t_k - t_{k-1}} \cdot Z_k$$

for $k = 1, \ldots, n$ (there's a typo in the Handbook here by the way; $\sqrt{t_k - t_{k-1}}$ should be $\sqrt{t_i - t_{i-1}}$ instead).

The ODE solver would then run on the points $t_0, t_1, \ldots, t_{n-1}$ with

$$\frac{dw_n}{dt}(t_i) = \frac{Z_i}{\sqrt{t_{i+1} - t_i}},$$

where $Z_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$. It is worth noting that Wong and Zakai stated in their original paper that this result holds for weaker conditions if w_n are polygonal approximations to W_t (like linear splines), so this seems to be the ideal choice. Additionally, the proof of the Wong-Zakai theorem seems to imply that the solution $x_n(t)$ to the ODE is exact at points $t_0, \ldots, t_n$ if w_n is exact at points $t_0, \ldots, t_n$. This leads to the following algorithm:
These ideas have been used in conjunction with the fifth and eighth-order Dormand Prince methods to develop the MATLAB functions `sde45` and `sde853`. We can estimate the error in the processes by computing $\left\|X_t - X_t^h\right\|_{L^2([0,T]\times\Omega)}$ which is given by

$$\left\|X_t - X_t^h\right\|^2_{L^2([0,T]\times\Omega)} = \mathbb{E}\left[\int_0^T \left(X_t - X_t^h\right)^2 dt\right]^{\frac{1}{2}} \approx \mathbb{E}\left[\frac{T}{n}\sum_{i=1}^n \left(X_{t_i} - X_{t_i}^h\right)^2\right].$$

Algorithm 2 Solving SDEs using a Runge-Kutta method

(1) Initialise $i = 1$, t_0, x_0 and Brownian motion W on $\{t_0, \ldots, t_n\}$.
(2) Let $h_i = t_i - t_{i-1}$ and $Z_i = W_i - W_{i-1}$.
(3) Compute the intermediary coefficients:

$$x_{i,j} = x_{i-1} + \sum_{l=1}^{j-1} a_{j,l} k_{i,l}$$

$$t_{i,j} = t_{i-1} + h_i \cdot c_j$$

$$k_{i,j} = h_i \left[\mu \left(t_{i,j}, x_{i,j}\right) - \frac{1}{2} \sigma \left(t_{i,j}, x_{i,j}\right) \sigma_x \left(t_{i,j}, x_{i,j}\right) \right] + Z_i \sigma \left(t_{i,j}, x_{i,j}\right)$$

(4) Compute the next step:

$$x_i = x_{i-1} + \sum_j b_j k_{i,j}$$

(5) If $i = n$, return $\boldsymbol{x}$. Otherwise, set $i = i + 1$ and repeat from (2).

TABLE 1. Tests for geometric Brownian motion $dX_t = 2X_t dt + \frac{1}{2} X_t dW_t$ on $t \in [0, 6]$ with 100 sample paths

N	Euler		Milstein		Heun	
	ϵ	Time (s)	ϵ	Time (s)	ϵ	Time (s)
2^2	10026.5	1×10^{-5}	10026.4	1×10^{-5}	9965.57	1×10^{-5}
2^4	7343.74	1.5×10^{-5}	7342.71	1.5×10^{-5}	5322.06	2×10^{-5}
2^6	4064.33	5.8×10^{-5}	4019.45	6.8×10^{-5}	1079.07	7.8×10^{-5}
2^8	1647.39	2.3×10^{-4}	1614.31	2.7×10^{-4}	242.06	3.1×10^{-4}
2^{10}	402.835	8.9×10^{-4}	374.298	1.07×10^{-3}	44.392	1.3×10^{-3}

N	sde45		sde853		sde853-Euler
	ϵ	Time (s)	ϵ	Time (s)	ϵ-to-t ratio
2^2	3828.48	2.7×10^{-4}	237.87	4.1×10^{-4}	1.028
2^4	1.296	9×10^{-4}	0.019	1.5×10^{-3}	1.28×10^4
2^6	0.094	3.9×10^{-3}	8.33×10^{-6}	6.4×10^{-3}	5.34×10^7
2^8	2.464×10^{-3}	0.015	9.31×10^{-9}	0.025	8.49×10^8
2^{10}	7.764×10^{-5}	0.06	1.76×10^{-11}	0.10	7.78×10^{11}

The ϵ-to-t ratio is given by

$$\frac{\epsilon_{\text{Euler}}}{\epsilon_{\texttt{sde853}}} \cdot \frac{t_{\text{Euler}}}{t_{\texttt{sde853}}}$$

for which bigger is better.

Alternatively, if we want to use adaptive methods (like `ode45`), we need w_n to be smooth. The Karhunen-Loeve theorem says that the best (in terms of mean square error) smooth approximation of W_t for $t \in [0, b]$ with only n terms on an orthonormal basis is

$$w_n(t) = \sum_{k=1}^{n} Z_k \cdot \frac{2\sqrt{2b}}{(2k-1)\pi} \cdot \sin\left(\frac{(2k-1)\pi t}{2b}\right)$$

with $Z_k \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$, which has derivative

$$w_n'(t) = \sum_{k=1}^{n} Z_k \cdot \sqrt{\frac{2}{b}} \cdot \cos\left(\frac{(2k-1)\pi t}{2b}\right).$$

The following picture demonstrates the error in some of the methods for the GBM problem with $N = 2^6$.