About diffusions with jumps, their simulation, and some links with partial integro-differential equations

École d'été : EDP et Probabilités pour les sciences du vivant

Carl Graham

CMAP, École polytechnique, CNRS, Université Paris-Saclay

July 5, 2016

Itô stochastic differential equation

Consider $d \geq 1$ and $r \geq 1$, and continuous functions

$$b := (b_i)_{1 \leq i \leq d} : \mathbb{R}^d \to \mathbb{R}^d , \quad \sigma := (\sigma_{ij})_{\substack{1 \leq i \leq d \\ 1 \leq j \leq r}} : \mathbb{R}^d \to \mathbb{R}^{d \otimes r} . \quad (1)$$

The Itô stochastic differential equation (SDE) with coefficients (1), driven by a $\mathbb{R}^r$-valued Brownian motion $W := (W_t)_{t \in \mathbb{R}_+}$, with solution a $\mathbb{R}^d$-valued continuous process $X := (X_t)_{t \in \mathbb{R}_+}$, is

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t\,, \quad (2a)$$

i.e., since X and b and W are column vectors, for $1 \leq i \leq d$,

$$\mathrm{d}X_t^i = b_i(X_t)\,\mathrm{d}t + \sum_{j=1}^{r} \sigma_{ij}(X_t)\,\mathrm{d}W_t^j\,. \quad (2b)$$

Remarks

If $\sigma \equiv 0$ then this reduces to an

ordinary differential equation (ODE) with vector field b

and we are trying to define the flow of a dynamical system.

What about jumps ?!

Jumps will be modeled and added in later. We are defining first the diffusion between the jumps (if any).
If $\sigma \equiv 0$ this will encompass what some people call PDMP, which were studied far before this acronym was invented (kinetic equations such as neutron transport equations, etc.).

What about PDE ?!

PDE and PIDE will be introduced later, when the time is ripe.

Heuristic infinitesimal interpretation

The Brownian motion W is a continuous process with stationary independent increments, and $W_t \sim \mathcal{N}(0, tI_r)$, *i.e.*, is Gaussian,

$$\mathbb{E}[\,W_t\,] = 0\,, \quad \mathbb{E}[\,W_t W_t^*\,] = tI_r\,, \quad \mathbb{E}[\,W_t^i W_t^j\,] = \begin{cases} t\,, & i = j\,, \\ 0\,, & i \neq j\,. \end{cases}$$

The drift vector b indicates that, for an "infinitesimal time" $\mathrm{d}t > 0$,

$$\mathbb{E}[\,X_{t+\mathrm{d}t} - X_t \mid (X_s, W_s)_{s\le t}\,] \approx b(X_t)\,\mathrm{d}t\,.$$

The dispersion matrix σ yields the diffusion matrix

$$a := (a_{ij})_{1\le i,j\le d} \triangleq \sigma\sigma^*\,, \qquad i.e.\,, \qquad a_{ij} \triangleq \sum_{k=1}^{r} \sigma_{ik}\sigma_{jk}\,,$$

which is such that, for an "infinitesimal time" $\mathrm{d}t > 0$,

$$\mathbb{E}[\,(X_{t+\mathrm{d}t}-X_t-b(X_t)\,\mathrm{d}t)(X_{t+\mathrm{d}t}-X_t-b(X_t)\,\mathrm{d}t)^* \mid (X_s, W_s)_{s\le t}\,] \approx a(X_t)\,\mathrm{d}t$$

of the same order as the expectation, and cannot be neglected.

Modeling perspective

Thus, it should be derived from the model:

- The fact that the evolution is continuous.
  Else jumps will be added in adequately later,
  and the SDE will model what happens in-between jumps.
- The drift vector field b describing the mean evolution, or trend.
- The diffusion matrix a describing the covariance of the noise.
  Since a is symmetric and non-negative, it is always possible for $r \geq d$ to find σ such that $a = \sigma\sigma^*$.
  Seldom can a dispersion matrix σ be derived from the model.
- An initial condition under the form of an initial law π_0
  which may be given or implicit.

Probabilistic modeling, weak solution

We need to make precise the meaning of the SDE (2).

- The coefficients b and σ and an initial law π_0 are given.
- We seek a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$ and
  - a $(\mathscr{F}_t)_{t\geq 0}$-Brownian motion W,
  - a $(\mathscr{F}_t)_{t\geq 0}$-adapted continuous process X,

  such that X_0 has law π_0 and

$$X_t = X_0 + \int_0^t b(X_s)\,\mathrm{d}s + \int_0^t \sigma(X_s)\,\mathrm{d}W_s\,, \tag{3a}$$

i.e., for $1 \leq i \leq d$,

$$X_t^i = X_0^i + \int_0^t b_i(X_s)\,\mathrm{d}s + \sum_{j=1}^r \int_0^t \sigma_{ij}(X_s)\,\mathrm{d}W_s^j\,. \tag{3b}$$

This is the notion of weak solution.

Weak existence and weak uniqueness notions

Definition (Weak existence)

Weak existence is said to hold for SDE (2) if:
For **each** initial law π_0, there exists **some** $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P}, W, X)$ satisfying the above. Then:

- $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P}, W, X)$ is called a weak solution of SDE (2). By abuse of notation, X and (W, X) are also called weak solutions, with an implicit probabilistic set-up.
- The law P of X on the path-space $C(\mathbb{R}_+, \mathbb{R}^d)$ is the actual subject of study, also called a weak solution, or a solution in law.

Definition (Weak uniqueness)

Weak uniqueness, or uniqueness in law, is said to hold for SDE (2) if:
For **each** initial law π_0, there exists **at most one** law P on the path-space $C(\mathbb{R}_+, \mathbb{R}^d)$ which is a weak solution as above.

Diffusion operator

Let the 2nd order differential operator $\mathscr{D}$ act on $f \in C_b^2(\mathbb{R}^d, \mathbb{R})$ as

$$\begin{aligned}\mathscr{D}f(x) &= \sum_{i=1}^{d} b_i(x)\,\partial_i f(x) + \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(x)\,\partial_{ij}^2 f(x)\\ &= b(x)\cdot\nabla f(x) + \frac{1}{2}\,\mathrm{tr}[a(x)\nabla\nabla^* f(x)]\,, \qquad x \in \mathbb{R}^d\,. \qquad (4)\end{aligned}$$

This can be written in divergence form as

$$\mathscr{D}f(x) = b(x)\cdot\nabla f(x) - \frac{1}{2}[\nabla^* a(x)]\nabla f(x) + \frac{1}{2}\nabla\cdot a(x)\nabla f(x)$$

with an adequate corrective term.

Itô formula

The Itô formula applied to a weak solution X of SDE (2) writes

$$
\begin{aligned}
f(X_t) &= f(X_0) + \int_0^t \mathscr{D} f(X_s)\,\mathrm{d}s + \sum_{i=1}^{d}\sum_{j=1}^{r}\int_0^t \sigma_{ij}(X_s)\,\partial_i f(X_s)\,\mathrm{d}W_s^j \\
&= f(X_0) + \int_0^t \mathscr{D} f(X_s)\,\mathrm{d}s + \int_0^t \mathbf{1}_r \cdot \sigma^*(X_s)\,\nabla f(X_s)\,\mathrm{d}W_s^j\,.
\end{aligned}
$$

The last term is a local martingale,
and under suitable integrability controls, is a martingale
with null conditional expectation of increments, w.r.t. past.

Decomposition

The evolution of $f(X)$ is thus decomposed into:

- a "predictable" trend described using the operator $\mathscr{D}$,
- an unpredictable part described by this local martingale.

This is the semi-martingale decomposition of the process.

The Doob-Meyer bracket of the local martingale is given by

$$\sum_{i,j=1}^{d} \int_0^t a_{ij}(X_s)\,\partial_i f(X_s)\,\partial_j f(X_s)\,\mathrm{d}t = \int_0^t \nabla f(X_s) \cdot a(X_s) \nabla f(X_s)\,\mathrm{d}t$$

$$= \int_0^t [\mathscr{D} f^2(X_s) - 2 f(X_s) \mathscr{D} f(X_s)]\,\mathrm{d}t$$

and describes its "covariance" using the « *opérateur carré du champ* ».

Martingale problems

This leads to the notion of martingale problem (MPbm) brilliantly initiated by Stroock and Varadhan.[1]

One could mimic the statements for the SDE using some

$$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}, X)$$

but this would be quite cumbersome.

[1] Daniel W. Stroock and S. R. Srinivasa Varadhan (2006). *Multidimensional diffusion processes*. Reprint of 1997 Ed., 1st Ed. 1979.

Canonical space, filtration, and process

The proper probabilistic set-up and notation is provided by the canonical space

$$\widehat{\Omega} = C(\mathbb{R}_+, \mathbb{R}^d), \qquad \widehat{\mathcal{F}} = \mathcal{B}(\widehat{\Omega}),$$

with the canonical process $\widehat{X} = (\widehat{X}_t)_{t\geq 0}$ given by the projections

$$\widehat{X}_t : \omega = (\omega_t)_{t\geq 0} \in \widehat{\Omega} \mapsto \widehat{X}_t(\omega) = \omega_t\,,$$

and the canonical filtration $(\widehat{\mathcal{F}}_t)_{t\geq 0}$ given by

$$\widehat{\mathcal{F}}_t = \sigma(\widehat{X}_s : s \leq t)\,.$$

Martingale problem on canonical space

Definition (Martingale problem)

A law P on $(\widehat{\Omega}, \widehat{\mathcal{F}}, (\widehat{\mathcal{F}}_t)_{t\geq 0})$ is a solution to the martingale problem for the operator $\mathscr{D}$ in (4) with initial law π_0 if:

- The law under P of $\widehat{X}_0$ is π_0.
- For any f in $C_b^2(\mathbb{R}^d, \mathbb{R})$,

$$f(\widehat{X}_t) - f(\widehat{X}_0) - \int_0^t \mathscr{D} f(\widehat{X}_s)\, \mathrm{d}s\,, \qquad t \geq 0\,, \tag{5}$$

  is a local martingale under P for $(\widehat{\mathcal{F}}_t)_{t\geq 0}$.

Toward the SDE

Note that the Doob-Meyer bracket of this local martingale for any f can be readily computed by applying the martingale problem to f^2 and using the « *opérateur carré du champ* ».

Actually, using the Itô formula it is enough to consider functions which are adequate C_b^2 truncations of

$$x \in \mathbb{R}^d \mapsto x_i \in \mathbb{R}\,, \quad x \in \mathbb{R}^d \mapsto x_i x_j \in \mathbb{R}\,, \qquad 1 \le i, j \le d\,.$$

Weak SDE and martingale problem

Theorem (Equivalence SDE-MPbm)

Assume that b and σ as in (1) and $a : \mathbb{R}^d \to \mathbb{R}^{d\times d}_{sym+}$ are such that

$$a = \sigma\sigma^* .$$

Then, for a law P on $\widehat{\Omega} = C(\mathbb{R}_+, \mathbb{R}^d)$, the two following statements are equivalent:

1. *The law P is a weak solution of the SDE (2) with coeffs b and σ.*
2. *The law P is a solution of the martingale problem (5) for the operator $\mathcal{D}$ defined with coeffs b and a.*

Proof.

$\Rightarrow$: Apply the Itô formula; has already been seen.

$\Leftarrow$: Use a martingale representation result. $\square$

Pathwise constructions imply weak existence …

A modern probabilistic perspective tends to favor pathwise constructions. It is obvious how this may translate into weak existence results.

Pathwise constructions hold mainly under Cauchy-Lipschitz type assumptions and, in dimension 1, under Hölder-1/2 assumptions on σ, yielding results for the Feller and Fischer-Wright diffusions

$$\begin{aligned} dZ_t &= \alpha Z_t \, dt + \sqrt{2\beta Z_t}\, dW_t\,, && \text{on } \mathbb{R}_+\,, \\ dX_t &= \alpha_0(1 - X_t) - \alpha_1 X_t \, dt + \sqrt{2\beta X_t(1 - X_t)}\, dW_t\,, && \text{on } [0, 1]\,, \end{aligned}$$

which are so important in biology, see Yamada and Watanabe.[2]

[2] Toshio Yamada and Shinzo Watanabe (1971). "On the uniqueness of solutions of stochastic differential equations." In: *J. Math. Kyoto Univ.*

From a modeling perspective, the main point is weak uniqueness.

Many other techniques (PDE techniques, Girsanov tranforms, etc.) and sets of assumptions ensure

weak existence and/or weak uniqueness.

The beautiful Yamada-Watanabe Theorem[3] gives the whole picture.

Ikeda and Watanabe[4] and Karatzas and Shreeve[5] are good reference books for all these notions. Never underestimate Stroock and Varadhan,[6] though.

---

[3] Toshio Yamada and Shinzo Watanabe (1971). "On the uniqueness of solutions of stochastic differential equations." In: *J. Math. Kyoto Univ.*

[4] Nobuyuki Ikeda and Shinzo Watanabe (1989). *Stochastic differential equations and diffusion processes*. 2nd Ed., 1st Ed. 1981.

[5] Ioannis Karatzas and Steven E. Shreve (1991). *Brownian motion and stochastic calculus*. 2nd Ed., 1st Ed. 1988.

[6] Daniel W. Stroock and S. R. Srinivasa Varadhan (2006). *Multidimensional diffusion processes*. Reprint of 1997 Ed., 1st Ed. 1979.

Pathwise uniqueness implies weak uniqueness !

Definition (Pathwise uniqueness)

The SDE (2) with coefficients b and σ has the pathwise uniqueness property if:
For all $x \in \mathbb{R}^d$,
if and X and X' are weak solutions with initial law $\pi_0 = \delta_x$ constructed on the same $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P}, W)$, then $X = X'$, $\mathbb{P}$-a.s.

This actually implies that this holds true for arbitrary initial laws π_0.

Theorem (Yamada-Watanabe)

Pathwise uniqueness implies weak uniqueness.

There is an additional statement about strong solutions, but these mainly come into play in the proof of the Yamada-Watanabe Thm.

Marginal laws

A law P on $\widehat{\Omega} = C(\mathbb{R}_+, \mathbb{R}^d)$ induces by projection a flow of marginals $(\pi_t)_{t\geq 0}$ on $\mathbb{R}^d$. Then π_t is the law of $\widehat{X}_t$ under P.

For the duality bracket between $(\mathcal{M}(\mathbb{R}^d), \|\cdot\|_{\mathrm{TV}})$ and $(\mathcal{B}_b(\mathbb{R}^d), \|\cdot\|_\infty)$ given by

$$\langle \mu , f \rangle = \int_{\mathbb{R}^d} f \, \mathrm{d}\mu \,,$$

it holds that

$$\langle \pi_t , f \rangle = \mathbb{E}^P[\, f(\widehat{X}_t) \,] := \int_{C(\mathbb{R}_+, \mathbb{R}^d)} f(\omega_t) \, P(\mathrm{d}\omega) \,, \qquad f \in \mathcal{B}_b(\mathbb{R}^d) \,.$$

Probabilistic information

Then $P \in \mathcal{M}_+^1(C(\mathbb{R}_+, \mathbb{R}^d))$ implies that $(\pi_t)_{t\geq 0} \in C(\mathbb{R}_+, \mathcal{M}_+^1(\mathbb{R}^d))$.

- The law P contains much more information than $(\pi_t)_{t\geq 0}$! (Don't get a probabilist started on this subject.)
- Nevertheless the information in $(\pi_t)_{t\geq 0}$ can be very relevant.
- The law P of a Markov process can be reconstructed from its marginals $(\pi_t)_{t\geq 0}$, since its transition kernel $(P_t(x, dy))_{x\in\mathbb{R}^d}$ is given by

$$P_t(x, dy) = \pi_t(dy) \quad \text{for} \quad \pi_0 = \delta_x \, .$$

Existence and uniqueness, flow of operators

If the martingale problem has both the existence and the uniqueness property, it is said to be well-posed.

If so, let $(P_t)_{t\geq 0}$ be the family of operators

$$P_t : f \in \mathscr{B}_b(\mathbb{R}^d) \mapsto P_t f \in \mathscr{B}_b(\mathbb{R}^d)$$

given by

$$P_t f(x) \triangleq \mathbb{E}^{P^x}[\,f(\widehat{X}_t)\,] \triangleq \mathbb{E}_x[\,f(\widehat{X}_t)\,] = \langle P_t(x,\bullet), f\rangle\,.$$

With abuse of notation, one often speaks of a solution P on $C(\mathbb{R}_+,\mathbb{R}^d)$ without specifying π_0, and writes that, under P,

$$P_t f(x) = \mathbb{E}_x[\,f(X_t)\,] = \mathbb{E}[\,f(X_t) \mid X_0 = x\,]\,.$$

Markov property, semi-group, generator

Theorem (Markov characterization by MPbm)

Assume that the martingale problem is well-posed.
Then the solutions P on $(\widehat{\Omega}, \widehat{\mathcal{F}}, (\widehat{\mathcal{F}}_t)_{t\geq 0})$ correspond to a Markov process, and $(P_t)_{t\geq 0}$ is its semi-group:

Under P, for $t \geq 0$ and $s \geq 0$,

$$\mathbb{E}[\, f(\widehat{X}_{t+s}) \mid \widehat{\mathcal{F}}_t \,] = P_s f(\widehat{X}_t)\,, \qquad f \in \mathcal{B}_b(\mathbb{R}^d)\,,$$

and hence $P_{t+s} = P_t P_s$ (semi-group property). The generator of the Markov process or of its semi-group is $\mathcal{D}$ so that, for $f \in C_b^2(\mathbb{R}^d)$,

$$\mathcal{D}f(x) = \lim_{\varepsilon\to 0^+} \frac{\mathbb{E}_x[\, f(\widehat{X}_\varepsilon)\,] - f(x)}{\varepsilon} = \lim_{\varepsilon\to 0^+} \frac{P_\varepsilon f(x) - f(x)}{\varepsilon}\,.$$

Relations with partial differential equations

We are now ready to introduce some of the relations between SDE and MPbm on the one hand and PDE on the other.

Good reference books on the subject and on its practical applications are Kushner[7] and Robert Dautray, Pierre-Louis Lions, Étienne Pardoux, *et al.*[8].

[7] Harold J. Kushner (1977). *Probability methods for approximations in stochastic control and for elliptic equations*. Mathematics in Science and Engineering, Vol. 129.

[8] Robert Dautray et al. (1989). *Méthodes probabilistes pour les équations de la physique*.

Taking expectations

Assume that the law P on $\widehat{\Omega} = C(\mathbb{R}_+, \mathbb{R}^d)$ and $f \in C_b^2(\mathbb{R}^d, \mathbb{R})$ satisfy that

$$f(\widehat{X}_t) - f(\widehat{X}_0) - \int_0^t \mathscr{D}f(\widehat{X}_s)\,\mathrm{d}s\,, \qquad t \geq 0\,,$$

is an actual martingale, and take expectations. Then

$$\mathbb{E}[\,f(\widehat{X}_t)\,] = \mathbb{E}[\,f(\widehat{X}_0)\,] + \int_0^t \mathbb{E}[\,\mathscr{D}f(\widehat{X}_s)\,]\,\mathrm{d}s\,, \qquad t \geq 0\,. \qquad (6)$$

Note that some probabilistic information is lost.

Forward Kolmogorov eqn, Fokker-Planck eqn

Using the duality bracket $\mathbb{E}[\,g(\widehat{X}_s)\,] = \langle \pi_s\,,\,g \rangle$, eq. (6) writes

$$\begin{aligned} \langle \pi_t\,,\,f \rangle &= \langle \pi_0\,,\,f \rangle + \int_0^t \langle \pi_s\,,\,\mathscr{D} f \rangle\,\mathrm{d}s \\ &= \langle \pi_0\,,\,f \rangle + \int_0^t \langle \mathscr{D}^* \pi_s\,,\,f \rangle\,\mathrm{d}s\,. \end{aligned} \tag{7a}$$

Thus, in weak (distributional) sense $(\pi_t)_{t\geq 0}$ is a solution of the forward Kolmogorov (or Fokker-Planck) equation, in $\mathscr{M}(\mathbb{R}^d)$,

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}\mu_t = \mathscr{D}^* \mu_t\,, \\ \mu_0 = \pi_0\,. \end{cases} \tag{7b}$$

If ever $\mu_t(dx) = m(t,x)\,dx$, we obtain the functional PDE

$$\partial_t m = \mathscr{D}^* m\,.$$

Such eqns are derived by balance or conservation considerations in many applications, and may thus be given a Markov representation.

Computing the adjoint

In distributional sense, and by integration by parts when OK,

$$\mathscr{D}^*\mu(dx) = -\sum_{i=1}^{d} \partial_i[b_i(x)\,\mu(dx)] + \frac{1}{2}\sum_{i,j=1}^{d} \partial_{ij}^2[a_{ij}(x)\,\mu(dx)] \quad (8a)$$

$$= -\Big(\nabla\cdot b(x) - \frac{1}{2}\,\mathrm{tr}[\nabla\nabla^* a]\Big)\mu(dx) - b(x)\cdot\nabla\mu(dx)$$

$$+ \frac{1}{2}[\nabla^* a(x)]\nabla\mu(dx) + \frac{1}{2}\nabla\cdot[a(x)\nabla\mu(dx)] \quad (8b)$$

$$= -\Big(\nabla\cdot b(x) - \frac{1}{2}\,\mathrm{tr}[\nabla\nabla^* a]\Big)\mu(dx) - b(x)\cdot\nabla\mu(dx)$$

$$+ [\nabla^* a(x)]\nabla\mu(dx) + \frac{1}{2}\,\mathrm{tr}[a(x)\nabla\nabla^*\mu(dx)] \quad (8c)$$

where (8b) is in divergence form.

The leading 2nd order term is the same as that of $\mathscr{D}$, see (8c).

Backward Kolmogorov equation

Another perspective on (6): using $\mathbb{E}_x[\,g(\widehat{X}_s)\,] = P_s g(x)$ and

$$\frac{P_{t+\varepsilon} - P_t}{\varepsilon} = P_t \frac{P_\varepsilon - I_d}{\varepsilon} = \frac{P_\varepsilon - I_d}{\varepsilon} P_t \xrightarrow[\varepsilon \to 0^+]{} P_t \mathscr{D} = \mathscr{D} P_t$$

then (6) writes – if all goes well –

$$\begin{aligned} P_t f(x) &= f(x) + \int_0^t P_s \mathscr{D} f(x)\,\mathrm{d}s \\ &= f(x) + \int_0^t \mathscr{D} P_s f(x)\,\mathrm{d}s\,. \end{aligned} \tag{9a}$$

Thus

$$P_t f : x \in \mathbb{R}^d \mapsto P_t f(x) = \mathbb{E}_x[\,f(\widehat{X}_t)\,]$$

is a solution of the backward Kolmogorov equation, in $C_b^2(\mathbb{R}^d)$,

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} u_t = \mathscr{D} u_t\,, \\ u_0 = f\,. \end{cases} \tag{9b}$$

Feynman-Kac formula

The derivation of the backward Kolmogorov equation involves a differentiation backwards in time. This leads us to revert time.

Thus, the backward parabolic PDE

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} u_t + \mathscr{D} u_t = 0\,, & 0 \le t \le T\,, \\ u_T = f\,, \end{cases} \tag{10}$$

has a probab. representation of solution $(w_t)_{0 \le t \le T} = (P_{T-t} f)_{0 \le t \le T}$:

$$w_t(x) = \mathbb{E}_x[\, f(\widehat{X}_{T-t})\,] = \mathbb{E}[\, f(X_T) \mid X_t = x\,]\,, \quad 0 \le t \le T\,,\ x \in \mathbb{R}^d\,.$$

This is a special case of the Feynman-Kac formula.

A general backward parabolic PDE

Consider the backward parabolic PDE, with terminal value,

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}u_t + \mathscr{D}_t u_t + c_t u_t + d_t = 0\,, & 0 \le t \le T\,, \\ u_T = f\,, \end{cases} \tag{11}$$

with time-dependent coefficients

$$a\colon (t,x) \in \mathbb{R}_+ \times \mathbb{R}^d \mapsto a_t(x) \in \mathbb{R}^{d\otimes d}_{\mathrm{sym}+}\,,$$

$$b\colon (t,x) \in \mathbb{R}_+ \times \mathbb{R}^d \mapsto b_t(x) \in \mathbb{R}^d\,,$$

$$c\colon (t,x) \in \mathbb{R}_+ \times \mathbb{R}^d \mapsto c_t(x) \in \mathbb{R}^d\,,$$

$$d\colon (t,x) \in \mathbb{R}_+ \times \mathbb{R}^d \mapsto d_t(x) \in \mathbb{R}^d\,,$$

and operators $\mathscr{D}_t$ acting on $g \in C_b^2(\mathbb{R}^d)$ as

$$\mathscr{D}_t g(x) = b_t(x) \cdot \nabla g(x) + \frac{1}{2}\,\mathrm{tr}[a_t(x)\nabla\nabla^* g(x)]\,, \qquad x \in \mathbb{R}^d\,.$$

Feynman-Kac formula, general version

Theorem (Feynman-Kac formula)

Assume that this backward parabolic PDE has a nice solution $(w_t)_{t\geq 0}$ *(see Dautray et al.*[a]*, e.g.). Let* $(X_t)_{t\geq 0}$ *be a solution of the time-dependent SDE for* $\mathscr{D}_t$*. Then, for* $0 \leq t \leq T$ *and* $x \in \mathbb{R}^d$*,*

$$w_t(x) = \mathbb{E}\left[\mathrm{e}^{\int_t^T c_s(X_s)\,\mathrm{d}s} f(X_T) + \int_t^T \mathrm{e}^{\int_t^r c_s(X_s)\,\mathrm{d}s} d_r(X_r)\,\mathrm{d}r \,\middle|\, X_t = x \right].$$

This is the Feynman-Kac probabilistic representation formula.

---

[a] Robert Dautray et al. (1989). *Méthodes probabilistes pour les équations de la physique*.

An existence result for the SDE yields a uniqueness result for the PDE (typical of duality), as well as a Monte-Carlo method using the simulation of its solutions starting at time t at x for a duration $T - t$.

Feynman-Kac formula, proof

For $0 \le t \le u \le T$, the Itô formula yields that

$$
\begin{aligned}
Y_{t,u} &\triangleq e^{\int_t^u c_s(X_s)\,\mathrm{d}s} w_u(X_u) + \int_t^u e^{\int_t^r c_s(X_s)\,\mathrm{d}s} d_r(X_r)\,\mathrm{d}r \\
&= w_t(X_t) + M_u - M_t \\
&\quad + \int_t^u e^{\int_t^r c_s(X_s)\,\mathrm{d}s} \underbrace{\left[\tfrac{\mathrm{d}}{\mathrm{d}r} w_r(X_r) + \mathscr{D}_r w_r(X_r) + c_r(X_r) w_r(X_r) + d_r(X_r)\right]}_{=0 \;\text{ using } (11)} \mathrm{d}r
\end{aligned}
$$

and hence

$$
\mathbb{E}[\, Y_{t,T} \mid X_t = x \,] = w_t(x)
$$

so that

$$
w_t(x) = \mathbb{E}\left[e^{\int_t^T c_s(X_s)\,\mathrm{d}s} f(X_T) + \int_t^T e^{\int_t^r c_s(X_s)\,\mathrm{d}s} d_r(X_r)\,\mathrm{d}r \,\middle|\, X_t = x \right]. \quad \square
$$

Simulation of Itô SDE

Very seldom does an Itô SDE have an explicit solution.

In certain situations (typically one-dimensional) it is possible to perform the so-called exact simulation of the Itô SDE.
The known methods are not very practical and can seldom be extended to higher dimensions. This is an active field of research.

Thus, discretization methods are typically used in order to simulate approximate solutions of the SDE.

Euler method

The grand father of such methods is the explicit Euler scheme.

We introduce a discretization step $\varepsilon > 0$, and compute the values of an approximate simulation on the grid $0, \varepsilon, 2\varepsilon, \cdots$, by freezing the arguments of the coefficients in-between the grid-points.

This yields a sequence $(X_t^\varepsilon)_{t=0,\varepsilon,2\varepsilon,\dots}$ as follows:

Explicit Euler scheme:

- Draw X_0^ε according to π_0.
- For $n \in \mathbb{N}$ draw $W_{(n+1)\varepsilon} - W_{n\varepsilon}$ according to $\mathcal{N}(0, \varepsilon I_r)$ and set

$$X_{(n+1)\varepsilon}^\varepsilon = X_{n\varepsilon}^\varepsilon + b(X_{n\varepsilon}^\varepsilon)\,\varepsilon + \sigma(X_{n\varepsilon}^\varepsilon)\,(W_{(n+1)\varepsilon} - W_{n\varepsilon})\,.$$

The Brownian motion $(W_t)_{t\in\mathbb{R}}$ is mathematic fiction.

Three interpolations

Here are 3 ways to interpolate $(X_t^\varepsilon)_{t=0,\varepsilon,2\varepsilon\cdots}$ to obtain $(X_t^\varepsilon)_{t\in\mathbb{R}_+}$:

1. The step-process interpolation
$$X_t^\varepsilon = X_{\lfloor t/\varepsilon\rfloor\varepsilon}^\varepsilon .$$
Obvious, adapted, but discontinuous, and Skorohod comes in.
2. The linear interpolation
$$X_t^\varepsilon = X_{\lfloor t/\varepsilon\rfloor\varepsilon}^\varepsilon + (t/\varepsilon - \lfloor t/\varepsilon\rfloor)\,(X_{(\lfloor t/\varepsilon\rfloor+1)\varepsilon}^\varepsilon - X_{\lfloor t/\varepsilon\rfloor\varepsilon}^\varepsilon) .$$
Natural, computable, continuous, but not adapted.
3. The Brownian interpolation
$$X_t^\varepsilon = X_{\lfloor t/\varepsilon\rfloor\varepsilon}^\varepsilon + b(X_{\lfloor t/\varepsilon\rfloor\varepsilon}^\varepsilon)\,(t - \lfloor t/\varepsilon\rfloor\varepsilon) + \sigma(X_{\lfloor t/\varepsilon\rfloor\varepsilon}^\varepsilon)\,(W_t - W_{\lfloor t/\varepsilon\rfloor\varepsilon})$$
$$= X_0^\varepsilon + \int_0^t b(X_{\lfloor s/\varepsilon\rfloor\varepsilon}^\varepsilon)\,\mathrm{d}s + \int_0^t \sigma(X_{\lfloor s/\varepsilon\rfloor\varepsilon}^\varepsilon)\,\mathrm{d}W_s .$$
Natural, continuous, adapted, but not directly computable.

Convergence

Theorem

Assume that the martingale problem is well-posed (existence and uniqueness) and that b and a appearing in $\mathscr{D}$ are continuous.
Let P on $\widehat{\Omega} = C(\mathbb{R}_+, \mathbb{R}^d)$ be the solution of the martingale problem with initial law π_0 on $\mathbb{R}^d$.
Let $(X_t^\varepsilon)_{t\in\mathbb{R}_+}$ for $\varepsilon > 0$ be one of these continuous interpolations of the Euler scheme, and P^ε be their laws on the path-space $C(\mathbb{R}_+, \mathbb{R}^d)$.
Then

$$P^\varepsilon \xrightarrow[\varepsilon\to 0]{\text{weak}} P \ \text{ in } \ \mathscr{M}_+^1(C(\mathbb{R}_+, \mathbb{R}^d))\,.$$

Central limit theorem

Under suitable additional assumptions, there is a functional central limit theorem.

Assume that $X := (X_t)_{t\in\mathbb{R}_+}$ is the solution of the SDE
and $X_0 = X_0^\varepsilon$ and $X^\varepsilon := (X_t^\varepsilon)_{t\in\mathbb{R}_+}$ are the Brownian interpolations of the Euler schemes, and that all use the same Brownian motion W.
Then

$$\frac{1}{\sqrt{\varepsilon}}(X - X^\varepsilon) \xrightarrow[\varepsilon\to 0]{\text{in law}} Z\,,$$

where $Z := (Z_t)_{t\geq 0}$ is the unique weak solution of either of the SDE, with initial data 0,

$$dZ_t = b'(X_t)Z_t\,dt + \sqrt{\sigma'(X_t)^2(Z_t)^2 + \frac{1}{2}\sigma(X_t)^2\sigma'(X_t)^2}\,dW_t\,,$$

$$dZ_t = b'(X_t)Z_t\,dt + \sigma'(X_t)Z_t\,dW_t^1 + \frac{1}{\sqrt{2}}\sigma(X_t)\sigma'(X_t)\,dW_t^2\,.$$

Higher-order schemes

Among others, Milstein[9] presented several higher-order schemes.

One of these is known as the Milstein scheme. It applies the Itô formula to $\sigma(X_s) - \sigma(X_{\lfloor s/\varepsilon \rfloor \varepsilon})$ to attain order $O(\varepsilon)$.
Unfortunately its algorithmic complexity is usually much larger than the one for the Euler scheme: it involves the partial derivatives of σ, as well as the stochastic integrals

$$\int_{n\varepsilon}^{(n+1)\varepsilon} (W_s^i - W_{n\varepsilon}^i)\, \mathrm{d}W_s^j$$

which are not easily simulatable when $i \neq j$.

[9] G. N. Milstein (1978). "A method with second order accuracy for the integration of stochastic differential equations". In: *Teor. Verojatnost. i Primenen.* 2; G. N. Milstein (1995). *Numerical integration of stochastic differential equations*. Translated and revised from the 1988 Russian original.

The Milstein scheme

Since

$$\int_{n\varepsilon}^{(n+1)\varepsilon} (W_s^i - W_{n\varepsilon}^i)\,\mathrm{d}W_s^i = \frac{1}{2}\Big[\big(W_{(n+1)\varepsilon}^i - W_{n\varepsilon}^i\big)^2 - \varepsilon\Big]$$

this problem vanishes when $r = 1$, and we may use the following.

Explicit Milstein scheme, 1-dim.:

- Draw X_0^ε according to π_0.
- For $n \in \mathbb{N}$ draw $W_{(n+1)\varepsilon} - W_{n\varepsilon}$ according to $\mathcal{N}(0, \varepsilon I_r)$ and set

$$\begin{aligned} X_{(n+1)\varepsilon}^\varepsilon &= X_{n\varepsilon}^\varepsilon + b(X_{n\varepsilon}^\varepsilon)\,\varepsilon + \sigma(X_{n\varepsilon}^\varepsilon)\,(W_{(n+1)\varepsilon} - W_{n\varepsilon}) \\ &\quad + \frac{1}{2}\,[\nabla\sigma^*(X_{n\varepsilon}^\varepsilon)]\,\sigma(X_{n\varepsilon}^\varepsilon)\,\Big[\big(W_{(n+1)\varepsilon} - W_{n\varepsilon}\big)^2 - \varepsilon\Big]\,. \end{aligned}$$

A similar simplification exists for $r > 1$ under a commutativity hypothesis on σ and $\nabla\sigma^*$.

The Lévy measure : Jumps at last !

The jumps of a process $(X_t)_{t\geq 0}$ with sample paths in $D(\mathbb{R}_+, \mathbb{R}^d)$ can be specified through the Lévy kernel $(L(x,dh))_{x\in\mathbb{R}^d}$ satisfying

$$L(x,dh) \in \mathcal{M}_+(\mathbb{R}^d), \quad \text{loc. bdd in } x, \quad L(x,\{0\}) = 0\,.$$

The non-negative function and probability kernel given resp. by

$$\lambda(x) \triangleq L(x,\mathbb{R}^d), \; l(x,dh) \triangleq \frac{L(x,dh)}{\lambda(x)} \mathbb{1}_{\{\lambda(x)\neq 0\}} + \delta_0(dh)\mathbb{1}_{\{\lambda(x)=0\}}\,,$$

describe the instantaneous intensity of jumps at position x and the law of the (potential) jumps from x to $x+h$ as follows:

$$\mathbb{P}\{\text{ no jumps on } [u,v]\,\} = \mathrm{e}^{-\int_u^v \lambda(X_t)\,\mathrm{d}t}\,, \qquad 0 \leq u \leq v\,,$$

and in case of jump at time t then

$$\text{the law of jumps from } X_{t-} \text{ to } X_t = X_{t-} + h \text{ is } l(X_{t-}, dh)\,.$$

An equivalent formulation

We may equivalently use a "jump kernel" $(K(x,dh))_{x\in\mathbb{R}^d}$ satisfying

$$K(x,dy) \in \mathcal{M}_+(\mathbb{R}^d), \quad \text{loc. bdd in } x, \quad K(x,\{x\}) = 0\,.$$

The non-negative function and probability kernel given resp. by

$$\lambda(x) = K(x,\mathbb{R}^d), \;\; k(x,dy) \triangleq \frac{K(x,dy)}{\lambda(x)} \mathbb{1}_{\{\lambda(x)\neq 0\}} + \delta_x(dh)\mathbb{1}_{\{\lambda(x)=0\}}\,,$$

describe the instantaneous intensity of jumps at position x and the law of the (potential) jumps from x to y as follows:

$$\mathbb{P}\{\,\text{no jumps on } [u,v]\,\} = \mathrm{e}^{-\int_u^v \lambda(X_t)\,\mathrm{d}t}\,, \qquad 0 \le u \le v\,,$$

and in case of jump at time t then

$$\text{the law of jumps from } X_{t-} \text{ to } X_t = y \text{ is } k(X_{t-},dy)\,.$$

Image measures

The kernels $(L(x, dh))_{x\in\mathbb{R}^d}$ and $(K(x, dh))_{x\in\mathbb{R}^d}$ are image measures of one another, and satisfy that

$$\int_{\mathbb{R}^d} f(y)\,K(x, dy) = \int_{\mathbb{R}^d} f(x+h)\,L(x, dh)\,, \qquad f \in \mathscr{B}_b(\mathbb{R}^d)\,.$$

Intermediate versions of these two formulations may be used to alleviate notation in modeling.

Diffusion with jumps

Consider a process which evolves according to $\mathscr{D}$ in between jumps, and jumps according to a Lévy kernel L or a jump kernel K as described above.

One can try to specify this through

- either an Itô-Tanaka SDE involving a Poisson point process,
- or a SDE involving time-changed marked Poisson processes,

but this may be awkward.

Using these representations and the Itô formula,
or by direct computation, it will be seen what follows.

Martingale problem

Let $\mathcal{J}$ denote the integral operator acting on f in $\mathcal{B}_b(\mathbb{R}^d, \mathbb{R})$ as

$$\begin{aligned}\mathcal{J} f(x) &= \int_{\mathbb{R}^d} [f(x+h) - f(x)]\, L(x, dh) \\ &= \int_{\mathbb{R}^d} [f(y) - f(x)]\, K(x, dy)\,, \qquad x \in \mathbb{R}^d\,.\end{aligned}$$

This determines L up to mass at 0 and $K(x, \cdot)$ up to mass at x, which were assumed to be null. Define

$$\mathcal{L} = \mathcal{D} + \mathcal{J} \quad \text{on} \quad C_b^2(\mathbb{R}^d, \mathbb{R})\,.$$

Then for any f in a suitable subset of $C_b^2(\mathbb{R}^d, \mathbb{R})$,

$$f(X_t) - f(X_0) - \int_0^t \mathcal{L} f(X_s)\, \mathrm{d}s\,, \qquad t \geq 0\,, \tag{12}$$

should be a local martingale (we need to control large jumps).

A neuroscience caricature

We model a caricature of a neural network as a jump diffusion using a martingale problem.

There are d neurons, each with potential in $\mathbb{R}$. Let $(e_j)_{1\le j\le d}$ denote the the canonical basis of $\mathbb{R}^d$. The generator writes

$$\mathscr{L}f(x) = \sum_{1\le i\le d} \mathscr{L}_i f(x) \quad \text{(where } \mathscr{L}_i \text{ acts and depends only on } x_i\text{)}$$

$$+ \sum_{1\le i\le d} \int_{h\in\mathbb{R}_+} \left[f\left(x - h\,e_i + \sum_{1\le j\ne i\le d} w_{ij}(h, x_j)\,e_j \right) - f(x) \right] L_i(x_i, dh)$$

where $w_{ij}(h, x)$ quantifies the effect of a discharge of amplitude h of neuron i on neuron $j \ne i$ having potential x, through an excitatory synapse if $w_{ij} > 0$ and an inhibitory synapse if $w_{ij} < 0$.

There should be mean-field limits under suitable assumptions.

An individual-based SISR model

We model a system of N individuals with global generic state $x = (x_{kn})_{1 \le k \le K, 1 \le n \le N}$ in $\mathbb{R}^{K \otimes N}$, where $x_{\bullet n} \triangleq (x_{kn})_{1 \le k \le K}$ in $\mathbb{R}^K$ represents the state of the n-th individual as follows:

- if $x_{1n} = 0$, 1, or -1, then it is Susceptible, Infected, or Removed,
- and $x = (x_{kn})_{2 \le k \le K}$ describes its position, anti-body count, vaccination status, viral load, phenotype, age, etc.

For $h \in \mathbb{R}^K$ let $h_{\bullet n} \in \mathbb{R}^{K \times N}$ be s.t. $[h_{\bullet n}]_{\bullet, n} = h$ and $[h_{\bullet n}]_{\bullet, p} = 0$ if $p \neq n$. The generator writes (healing, removal, etc., is in the $\mathscr{L}_n$)

$$\mathscr{L} f(x) = \sum_{1 \le n \le N} \mathscr{L}_n f(x) \quad \text{(where } \mathscr{L}_n \text{ acts and depends only on } x_{\bullet n}\text{)}$$

$$+ \sum_{1 \le m \neq n \le N} \int_{h \in \mathbb{R}^K} [f(x + h_{\bullet n}) - f(x)] \, L_{mn}((x_{\bullet m}, x_{\bullet n}), dh) \, .$$

There should be mean-field limits under suitable assumptions.

Partial integro-differential equations

The results on the forward and backwards Kolmogorov equations and on the Feynman-Kac Formula can be suitably adapted using

$$\mathscr{L} = \mathscr{D} + \mathscr{J}\,, \qquad \mathscr{L}^* = \mathscr{D}^* + \mathscr{J}^*,$$

instead of $\mathscr{D}$ and $\mathscr{D}^*$.

This leads to partial integro-differential equations (PIDE).

For the forward equation, a quick computation shows that

$$\begin{aligned}\mathscr{J}^*\mu(dx) &= \int_{y\in\mathbb{R}^d} [K(y, dx)\,\mu(\mathrm{d}y) - K(x, dy)\,\mu(\mathrm{d}x)] \\ &= \int_{y\in\mathbb{R}^d} K(y, dx)\,\mu(\mathrm{d}y) - \lambda(x)\,\mu(\mathrm{d}x)\end{aligned}$$

which has a natural interpretation
as a balance or conservation equation.

Construction and simulation

The question now is how to construct and simulate such a diffusion with jumps, which is Markov process in case the above martingale problem is well-posed.

In order to construct a Markov process $(X_t)_{t\geq 0}$, the natural idea is

- to construct the process between jump instants according to the (well-posed) martingale problem for $\mathscr{D}$,
- and to determine successively the jump instants, as well as either the jump amplitudes according to the Lévy kernel L, or the jump locations according to the jump kernel K.

This method of construction will succeed if jump instants do not accumulate in finite time, yielding existence as well as uniqueness.

We shall see several ways to do so.

The method of true jumps

True jumps method:

1. Draw X_0 according to π_0. Set $n = 1$ and $T_0 = 0$.
2. When the process $(X_t)_{0 \le t \le T_{n-1}}$ has been constructed:
   - Draw E according to $\mathcal{E}(1)$ (exponential law).
   - Construct $(X_t)_{T_{n-1} \le t < T_n}$ according to $\mathcal{D}$ for

$$T_n = \inf \left\{ u \ge T_{n-1} : \int_{T_{n-1}}^{u} \lambda(X_t)\,\mathrm{d}t \ge E \right\}.$$

   - Draw y according to $k(X_{T_n -}, dy)$ and set $X_{T_n} = y$.
   - Set $n = n + 1$ and go back to 2).

The T_n for $n \ge 1$ which are finite are the true jump instants of the process. None of the draws of the $\mathcal{E}(1)$ is wasted.
This construction corresponds to the SDE
involving time-changed marked Poisson processes.

Advantages and inconvenients

From a theoretical point of view the only problem is that these jump instants may accumulate in finite time, *i.e.*,

$$\mathbb{P}\left\{\lim_{n\to\infty} T_n < \infty\right\} > 0\,.$$

A simple sufficient condition for this not to happen is that

$$\sup_{x\in\mathbb{R}^d} \lambda(x) < \infty\,,$$

but there are many other conditions, for instance based on infinite returns in a set on which λ is bounded.

From a practical point of view, computing the integrals

$$\int_{T_{n-1}}^{u} \lambda(X_t)\,\mathrm{d}t$$

may be very computationally expensive.

The method of fictitious jumps

Assume that you know a bound β such that

$$\sup_{x \in \mathbb{R}^d} \lambda(x) \leq \beta < \infty \,.$$

Fictitious jumps method:

1. Draw X_0 according to π_0. Set $n = 1$ and $T_0 = 0$.
2. When the process $(X_t)_{0 \leq t \leq T_{n-1}}$ has been constructed:
   - Draw E according to $\mathcal{E}(\beta) \sim \mathcal{E}(1)/\beta$.
   - Construct $(X_t)_{T_{n-1} \leq t < T_n}$ according to $\mathcal{D}$ where

     $$T_n = T_{n-1} + E \,.$$

   - With probability $\frac{\lambda(X_{T_n-})}{\beta}$ draw y according to $k(X_{T_n-}, dy)$ and set $X_{T_n} = y$, else set $X_{T_n} = X_{T_n-}$.
   - Set $n = n + 1$ and go back to 2).

Advantages and inconvenients

The $(T_n)_{n\geq 1}$ are the jump instants of a Poisson process of intensity β. This simulation corresponds to the SDE with Poisson point process.

The actual jump instants of $(X_t)_{t\geq 0}$ are a thinning of $(T_n)_{n\geq 1}$ only involving minimal computations.

The main inconvenient of this method is that not only it requires that

$$\sup_{x\in\mathbb{R}^d} \lambda(x) \leq \beta < \infty\,,$$

but if the bound is poor, many draws are lost.

If β is large and λ varies widely over $\mathbb{R}^d$, then the time-step is small and this method corresponds to a costly worst-case scenario.

The method of subdomains

If λ varies widely over $\mathbb{R}^d$, and even if it is unbounded, the method of subdomains can be used. The space $\mathbb{R}^d$ is first partitioned in subdomains $\mathcal{O}_i$ on which λ is bounded by β_i and varies little.

Method of subdomains:

- The fictitious jump method with intensity β_i is used while the process remains in $\mathcal{O}_i$.
- It is necessary to detect when the process crosses over to another subdomain $\mathcal{O}_j$, and stop it at the boundary.
- The simulation must be then restarted using the fictitious jump method with intensity β_j, or perhaps $\beta_i \vee \beta_j$ as long as the simulated process remains close to the boundary.

Refining the partition approximates the true jump method.

Bibliography I

Thank you !

- Dautray, Robert et al. (1989). *Méthodes probabilistes pour les équations de la physique*.
- Ikeda, Nobuyuki and Shinzo Watanabe (1989). *Stochastic differential equations and diffusion processes*. 2nd Ed., 1st Ed. 1981.
- Karatzas, Ioannis and Steven E. Shreve (1991). *Brownian motion and stochastic calculus*. 2nd Ed., 1st Ed. 1988.
- Kushner, Harold J. (1977). *Probability methods for approximations in stochastic control and for elliptic equations*. Mathematics in Science and Engineering, Vol. 129.

Bibliography II

- Milstein, G. N. (1978). “A method with second order accuracy for the integration of stochastic differential equations”. In: *Teor. Verojatnost. i Primenen.* 2.
- – (1995). *Numerical integration of stochastic differential equations*. Translated and revised from the 1988 Russian original.
- Stroock, Daniel W. and S. R. Srinivasa Varadhan (2006). *Multidimensional diffusion processes*. Reprint of 1997 Ed., 1st Ed. 1979.
- Yamada, Toshio and Shinzo Watanabe (1971). “On the uniqueness of solutions of stochastic differential equations.” In: *J. Math. Kyoto Univ.*