Machine Learning for Stochastics
A short intro to stochastic processes

Thorsten Schmidt

Abteilung für Mathematische Stochastik

www.stochastik.uni-freiburg.de
thorsten.schmidt@stochastik.uni-freiburg.de

SS 2025

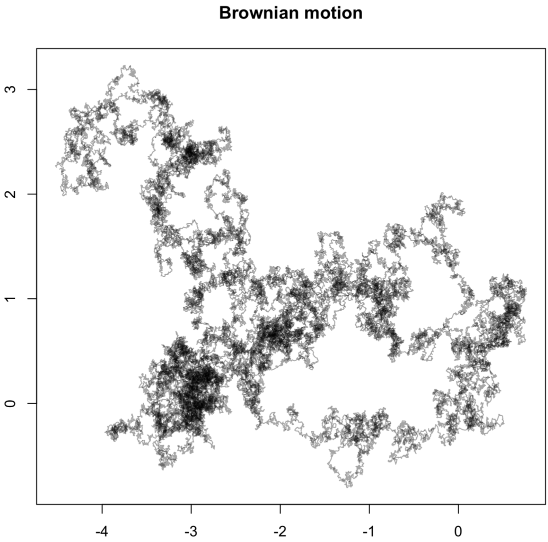
Week 6

In Week 7 we cover the following topics:

- Stochastic processes in discrete / continuous time
- Brownian motion, Poisson process, Lévy process affine processes
- What are semimartingales?
- Stochastic differential equations
- Neural SDEs
- Signatures
- Path-dependency using signatures

Motivation

The botanist Robert Brown observed in 1827 the movement of a particle on water.



- Our goal is to make a precise mathematical framework for understanding and describing such phenomena.
- This is content of a full course (Stochastic Processes), so we can only scratch the surface. Many good books are out, for example: I. Karatzas and S. E. Shreve (1988). **Brownian Motion and Stochastic Calculus**. Springer Verlag. Berlin Heidelberg New York, J. Jacod and A.N. Shiryaev (2003). **Limit Theorems for Stochastic Processes**. 2nd. Berlin: Springer Verlag, Philip Protter (2004). **Stochastic Integration and Differential Equations**. 2nd. Springer Verlag. Berlin Heidelberg New York, D. Revuz and Marc Yor (2005). **Continuous Martingales and Brownian Motion**. 3rd ed. p. cm. Springer Verlag. Berlin Heidelberg New York. Also lecture notes of my course are available.
- Lets start in discrete time.

Discrete time

- Discrete time is much simpler. A stochastic process (on a Polish space E) is a sequence of random variables, i.e.

$$S = (S_t)_{t=0,1,\dots}.$$

- Examples include

$$S_t = \sum_{i=1}^{t} X_i,$$

where X_i are i.i.d., for example $X_1 \sim \mathscr{N}(0,1)$. We can also have other distributions ! (essentially any ...)

- These processes have independent and stationary increments and are Markovian.
- The Polish space guarantees that we can compute conditional expectations for example. But also more general spaces are possible.

Continuous time

- In continuous time, a stochastic process (say on $\mathbb{R}^d$ for simplicity) is a family of random variables,

$$S = (S_t)_{t \geq 0}.$$

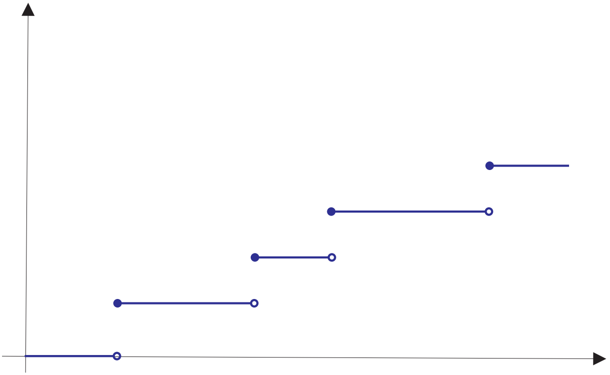
- We say that S has independent increments if $S_t - S_{t'}$ is independent from $S_s - S_{s'}$ whenever $s' \leq s \leq t' \leq t$.
- We say that the increments are stationary if $S_{t+h} - S_t$ has the same distribution as $S_h - S_0$ for all $t \geq 0$ and all $h \geq 0$.
- A process with independent and stationary increments is called **Lévy process**.
- We have a number of examples: the Brownian motion has Gaussian increments, i.e.

$$W_t - W_s \sim \mathcal{N}(0, t-s)$$

- The **Poisson process** takes values in $\mathbb{N}$ and

$$N_t - N_s \sim \text{Poisson}(\lambda(t-s)).$$

Poisson process



Many interesting observations / extensions of the Poisson process are possible: Compound Poisson, time-inhomogeneous Poisson, doubly stochastic Poisson, Semi-Markov processes, Shot-Noise processes, ...

Compound Poisson process

- A typical example of a process with jumps is the **compound Poisson process**.
- Consider a Poisson process N and independent i.i.d. random variables $\xi_1, \xi_2, \ldots$ and let
$$J_t = \sum_{i=1}^{N_t} \xi_i, \qquad t \geq 0.$$
- Then J has independent increments, is of finite variation and $W + J$ is a prototype of a semimartingale.

The Itô theory

- We work on a filtered probability space $(\Omega, \mathscr{F}, \mathbb{F}, P)$, satisfying the usual conditions. A filtration $\mathbb{F}$ is a family of increasing sub-σ-fields.
- For a wide class of processes, i.e. **semimartingales** one can construct a stochastic integral. As usually, this is done by starting with simple processes.
- A simple process is given by

$$H = Y\,\mathbb{1}_{[\![S,T[\![},$$

  where Y is $\mathscr{F}_S$-measurable and $S \leq T$ are finite stopping times.
- A simple predictable process is given by

$$H = Y\,\mathbb{1}_{]\!]S,T]\!]}.$$

  Note that if we have jumps, the integrand needs to be predictable, otherwise adapted is fine.
- For a semimartingale (a càdlàg process given as a sum of a finite variation process and a local martingale), we define

$$(H \cdot X)_t := \int_0^t H_s dX_s = \xi\,(X_{t\wedge T} - X_{t\wedge S}),$$

  where we note that if X takes values in $\mathbb{R}^d$, we view $\xi \in \mathbb{R}^d$ as an element of the dual space $(\mathbb{R}^d)^* = \mathbb{R}^d$. The stochastic integral in this form is then real-valued. Vector-valued stochastic integration is a bit more general.

Theorem

Let X be a semimartingale. The mapping $H \mapsto H \cdot X$ has an extension from the simple processes to the space of locally bounded, predictable processes, such that

(i) *$H \cdot X$ is adapted and càdlàg.*

(ii) *$H \mapsto H \cdot X$ is linear.*

(iii) *If predictable (H^n) converge pointwise to n H, and $|H^n| \leq K$ with a locally bounded, predictable process K, then*

$$(H^n \cdot X)_t \xrightarrow{P} (H \cdot X)_t \qquad \forall\, t > 0.$$

So the limits of simple integrands build a well-defined theory for stochastic integrals, which is very powerful. For a proof we refer to Jacod and Shiryaev (2003), op. cit.

The continuous case

If X is moreover continuous, we can extend the class even to locally square-integrable processes.

- We say a property of X holds locally, if there exists a sequence of stopping times $(T_n) \to \infty$ such that the property holds for all X^{T_n}. The stopped process is defined by

$$X_t^{T_n} = X_{T_n \wedge t}, \qquad t \geq 0.$$

One of the most powerful consequences is the following extension of the chain rule. We have to introduce the quadratic variation of the semimartingale X, which is given as the limit of square increments, i.e.

$$\langle X \rangle_t = \sum_{\Delta \to 0} (\Delta X_s)^2.$$

We know for example, that

$$\langle W \rangle_t = \langle N \rangle_t = t.$$

Theorem (Itô-formula)

Let $X = (X^1, \ldots, X^d)$ be a d-dimensional semimartingale and $f \in \mathscr{C}^2$. Then $f(X)$ is again a semimartingale and

$$\begin{aligned} f(X) = f(X_0) &+ \sum_{i \leq d} D_i f(X_-) \cdot X^i \\ &+ \frac{1}{2} \sum_{i,j \leq d} D_{ij} f(X_-) \cdot \langle X^{i,c}, X^{j,c} \rangle \\ &+ \sum_{0 \leq s \leq \cdot} \Big(f(X_s) - f(X_{s-}) - \sum_{i \leq d} D_i f(X_{s-}) \Delta X_s^i \Big). \end{aligned} \tag{1}$$

Lets look at simpler special cases. For example, if $X = W$, then

$$\begin{aligned} f(X) &= f(X_0) + f'(X) \cdot X + \frac{1}{2} f''(X) \cdot \langle X \rangle \\ &= f(X_0) + \int_0^{\cdot} f'(X_s) dX_s + \frac{1}{2} \int_0^{\cdot} f''(X_s) ds \end{aligned} \tag{2}$$

SDE

- The Itô formula also opens the door to stochastic differential equations, our extension of ODEs.
- As an example think of

$$X_t = e^{W_t}, \qquad t \geq 0$$

  where W is a Brownian motion.
- Now we can apply the Itô formula and obtain ...

SDE

- The Itô formula also opens the door to stochastic differential equations, our extension of ODEs.
- As an example think of

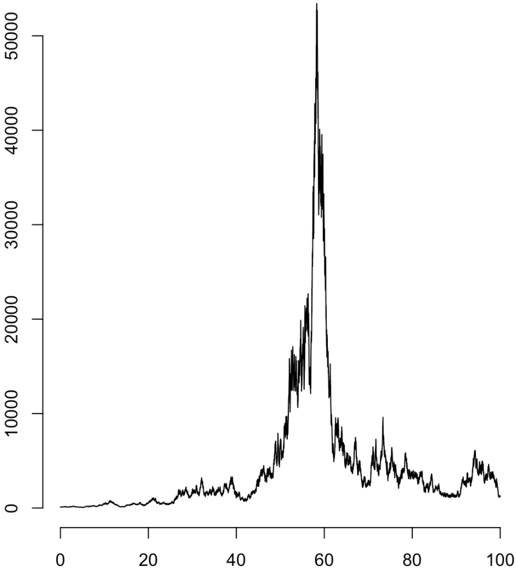
$$X_t = e^{W_t}, \qquad t \geq 0$$

  where W is a Brownian motion.
- Now we can apply the Itô formula and obtain ...
- $$X_t = X_0 + \int_0^t X_t dW_t + \frac{1}{2} \int_0^t X_t dt,$$

  which could be read as

$$dX_t = X_t dW_t + \frac{1}{2} X_t dt.$$

We give an example of $e^{\sigma W}$ with $\sigma = 0.4$



The martingale property

- We can already guess that this may cause problems. Note that W itself is a martingale, i.e. a process where

$$E[M_t| \mathscr{F}_s] = M_s, \qquad 0 \leq s \leq t$$

- but we would expect that this fails for e^W.
- Indeed, we know that $(H \cdot X)$ is a (local) martingale if X is a (local martingale) and as such we need to compensate e^W for the upward drift.
- Indeed, we can show with the Itô formula that

$$e^{W_t - t/2}, \qquad t \geq 0$$

  is a local martingale and we can calculate directly that it is indeed a martignale.
- However, it converges to 0 almost surely, since it is not closeable (i.e. there exist no M_∞ which conserves the martingale property.

We are interested in processes X which satisfy

$$\begin{aligned} dX_t &= b(t, X_t)dt + \sigma(t, X_t)dW_t, \\ X_0 &= \xi \end{aligned} \tag{3}$$

which we always understand as an abbreviation of

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s, \qquad t \geq 0.$$

Definition

A **strong solution** of (3) is a continuous and adapted process X, such that

(i) X is $\mathbb{F}$-adapted,

(ii) $P(X_0 = \xi) = 1$,

(iii) for all $0 \leq t < \infty$, $1 \leq i \leq d$, $1 \leq j \leq r$, it hold that

$$\int_0^t \big(|b_i(s, X_s)| + \sigma_{ij}^2(s, X_s)\big) ds < \infty$$

a.s. and

(iv) the first equation in (3) holds a.s.

Condition (iii) ensures that the integral is well-defined. The first property ensures that the ds-integral always exists, and the second guarantees local square integrability (which suffices for the stochastic integral when X is continuous).

Uniqueness and Existence

Theorem

Are the coefficients b and σ locally Lipschitz-continuous, then strong uniqueness holds for the SDE (3).

We say the global Lipschitz property holds, if

$$\|b(t,x) - b(t,y)\| + \|\sigma(t,x) - \sigma(t,y)\| \leq K\|x - y\|.$$

for all $t \geq 0$ and $x, y \in \mathbb{R}^d$.

Theorem

Assume the global Lipschitz property holds together with

$$\|b(t,x)\|^2 + \|\sigma(t,x)\|^2 \leq K^2\left(1 + \|x\|^2\right) \tag{4}$$

and $E\left[\|\xi\|^2\right] < \infty$. Then there exists a strong solution.

The idea of the proof is to use a modification of the Picard-Lindelöf iteration.

Linear equations

If we consider the linear equation

$$\begin{aligned} dX_t &= \Big(A(t)X_t + a(t)\Big)dt + \sigma(t)dW_t, \qquad 0 \le t < \infty, \\ X_0 &= \xi \end{aligned} \tag{5}$$

we expect that everything remains Gaussian and we con obtain a nice theory. As in the deterministic case,

$$X_t = \Phi(t)\left[X_0 + \int_0^t \Phi^{-1}(s)a(s)ds + \int_0^t \Phi^{-1}(s)\sigma(s)dW_s\right], \quad 0 \le t < \infty. \tag{6}$$

where

$$\dot{\Phi}(t) = A(t)\Phi(t) \tag{7}$$

is a fundamental solution of the homogeneous equation (this is a Matrix-differential equations, which has an easy solution for $d = 1$).

The Ornstein-Uhlenbeck process

Here we look for a solution of

$$dX_t = -\alpha X_t dt + \sigma dW_t,$$

with $\alpha, \sigma > 0$. This equation was already studied 1930 by the dutch physicists Leonard Ornstein and George Uhlenbeck.
With our solution method

$$X_t = X_0 + \int_0^t e^{-\alpha(t-s)} dW_s$$

is a solution (which is easily verified by the Itô-formula).

The Brownian bridge

For the Brownian bridge $B_t = W_t - {}^t/_T W_T$ one obtains an adapted representation via the SDE

$$dX_t = \frac{b - X_t}{T - t} dt + dW_t, \qquad 0 \le t < T, \; X_0 = a,$$

with $a, b \in \mathbb{R}$ and $T > 0$. As solution we obtain

$$X_t = a\left(1 - \frac{t}{T}\right) + \frac{b}{T} t + (T - t) \int_0^t \frac{dW_s}{T - s}.$$

We have the covariance function for $a = b = 0$

$$\rho(s, t) = (s \wedge t) - \frac{st}{T},$$

which characterizes the standard Brownian bridge (and coincides with the covariance function of B).

The one-dimensional case

The one-dimensional case can be solved completely: consider

$$dX_t = \big(A(t)X_t + a(t)\big)dt + \big(\Sigma(t)^\top X_t + \sigma(t)^\top\big)dW_t \tag{8}$$

with an r-dimensional Brownian motion W. We only assume that A, a and Σ, σ are $\mathbb{F}$-adapted, measurable and locally bounded.
Set

$$\zeta_t = \int_0^t \Sigma(s)^\top dW_s - \frac{1}{2}\int_0^t \Sigma(s)^\top \Sigma(s)ds,$$
$$Z_t = \exp\Big(\int_0^t A(s)ds + \zeta_t\Big).$$

Satz

The unique strong solution of (8) is given by

$$X_t = Z_t\Big[X_0 + \int_0^t \frac{1}{Z_s}(a(s) - \Sigma(s)^\top \sigma(s))ds + \int_0^t \frac{1}{Z_s}\Sigma(s)^\top dW_s\Big].$$

Affine processes

A much more flexible class is the class of affine processes. They are well described in many textbooks, see for example Damir Filipović (2009). **Term Structure Models: A Graduate Course**. Springer Verlag. Berlin Heidelberg New York. Simulations are studied in Aurélien Alfonsi (2015). **Affine diffusions and related processes: simulation, theory and applications**. Springer. The general d-dimensional semimartingale case was studied in Martin Keller-Ressel, Thorsten Schmidt, Robert Wardenga, et al. (2019). „Affine processes beyond stochastic continuity". In: **The Annals of Applied Probability** 29.6, pp. 3387–3437.

Lets keep it simple and consider the strong solution of the SDE

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t,$$

with initial condition $X_0 = x$. We set $a = \sigma\sigma^\top$.
We call X affine if

$$E\left[e^{iu^\top X_T}|\mathscr{F}_t\right] = \exp\left(\phi(T-t,u) + \psi(T-t,u)^\top X_t\right)$$

for all $u \in \mathbb{R}^d$ and all $0 \leq t \leq T$.
We assume that

$$\phi\colon \mathbb{C}^d \times \mathbb{R}_{\geq 0} \to \mathbb{C}$$
$$\psi\colon \mathbb{C}^d \times \mathbb{R}_{\geq 0} \to \mathbb{C}^d$$

are continuously differentiable functions.

Theorem

Assume that X is affine. Then b and a are affine:

$$b(x) = \beta_0 + \sum_{i=1}^{d} \beta_i x_i$$

$$c(x) = \gamma_0 + \sum_{i=1}^{d} \gamma_i x_i$$

for $\beta_0, \dots, \beta_d \in \mathbb{R}^d$ and $\gamma_0, \dots, \gamma_d \in \mathbb{R}^{d \times d}$.
Moreover, ϕ and ψ solve the following Riccati-Equations :

$$\begin{aligned}
\partial_t \phi(u,t) &= \beta_0^\top \psi(u,t) + \frac{1}{2}\psi(u,t)^\top \gamma_0 \psi(u,t) \\
\phi(u,0) &= 0 \\
\partial_t \psi_i(u,t) &= \beta_i^\top \psi(u,t) + \frac{1}{2}\psi(u,t)^\top \gamma_i \psi(u,t) \\
\psi(u,0) &= u
\end{aligned}$$

Riccati equations are equations of the type $f' = af^2$.

We also have the converse direction

Theorem

Assume that b and a are affine and the Riccati equations have solutions such that

$$\operatorname{real}\left(\phi(u,t) + \psi(u,t)^\top x\right) \leq 0 \qquad\qquad \forall u \in i\mathbb{R}^d$$

with $t \geq 0$, $x \in E$, then X is affine.

Neural SDEs

- We now come to the ML application to SDEs.
- Starting point is Ricky TQ Chen et al. (2018). „Neural ordinary differential equations“. In: **Advances in neural information processing systems** 31, where neural ODEs have been introduced. Quite a number of authors generalize this to neural SDEs, for example Belinda Tzen and Maxim Raginsky (2019). „Neural stochastic differential equations: Deep latent gaussian models in the diffusion limit“. In: **arXiv preprint arXiv:1905.09883**; Junteng Jia and Austin R Benson (2019). „Neural jump stochastic differential equations“. In: **Advances in Neural Information Processing Systems** 32.

- We consider a driving r-dimensional semimartingale Z and study the solution to the SDE

$$dX_t = b_\theta(t, X_t)dt + \sigma_\theta(t, X_{t-})dZ_t,$$

  $X_0 = x$ for some $x \in \mathbb{R}^d$.
- This is called a **neural SDE**, if b_θ and σ_θ are given by (deep) neural nets and trained on some given data.
- Typical examples involve that Z is a Brownian motion (classical neural SDEs) or Z is a pure-jump process (neural jump SDEs) or other variants.
- An first application to finance can be found in Samuel N Cohen, Christoph Reisinger, and Sheng Wang (2023). „Arbitrage-free neural-SDE market models". In: **Applied Mathematical Finance** 30.1, pp. 1–46, however in the diffusion setting only.

Path-dependence

- We note that there is in principle no difficulty to extend the setting to the path-dependent case, where
- $$dX_t = b_\theta(t, X_{[0,t]})dt + \sigma_\theta(t, X_{[0,t)})dZ_t,$$

  and $X_{[0,t]} = \{X_s \colon s \in [0,t]\}$ denotes the path of X from 0 to t. Now the functions b and σ live on the path space, here the space of càdlàg functions over $[0,t]$.
- One question is how to find a nice representation of these functions, which can be (in an explainable way) be done by signatures.
- Existence and uniqueness is guaranteed by classical restrictions on the functions.
- The universal approximation theorem (for Banach spaces) shows that all functions on $D([0,T])$ can be arbitrarily well approximated by a neural network.
- However - the training is more complicated. A classical question is of course when you observe only **one** path, how can you learn the parameters of X in a good way. This is typically done via maximum likelihood (which can also be done for stochastic processes through the Girsanov theorem).

Stochastic control

- Of course, one can also consider stochastic control in the following form

$$dX_t = b_\theta(t, X_t, a_t)dt + \sigma_\theta(t, X_{t-}, a_{t-})dZ_t,$$

  with a stochastic control a. One can derive neural HJB equations or solve using methods from reinforcement learning (X here is a continuous-time Markov decision process - if Z is in discrete time, we recover our setting from the first part of our lectures).

Signatures

- A further class of models are the so-called **signature SDEs**. These are path-dependent (neural) SDEs, where the path dependence is explicitly described via signatures.
- In a certain way, this corresponds to an **explainable** representation of the path-dependence.
- For signatures, we however, need to introduce a bit of notation. Signatures go back to Kuo-Tsai Chen (1957). „Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula“. In: **Annals of Mathematics** 65.1, pp. 163–178; Kuo-Tsai Chen (1977). „Iterated path integrals“. In: **Bulletin of the American Mathematical Society** 83.5, pp. 831–879 and play a particular important role in the context of rough path theory initiated by Terry J Lyons (1998). „Differential equations driven by rough signals“. In: **Revista Matemática Iberoamericana** 14.2, pp. 215–310.

Basic notions

- For $n \in \mathbb{N}_0$ consider the n-fold tensor product $(\mathbb{R}^d)^{\otimes n}$ i.e.

$$(x \otimes y \otimes z)_{ijk} = x_i y_j z_k, \quad i, j, k = 1, \dots, d.$$

- The extended tensor algebra is defined as sequences

$$T((\mathbb{R}^d)) := \{\mathbf{a} := (a_0, \dots, a_n, \dots) : a_n \in (\mathbb{R}^d)^{\otimes n}\}.$$

- and the truncated tensor algebra by

$$T^{(N)}(\mathbb{R}^d) := \{\mathbf{a} \in T((\mathbb{R}^d)) : a_n = 0, \forall n > N\},$$

- and the **tensor algebra**

$$T(\mathbb{R}^d) := \bigcup_{N \in \mathbb{N}} T^{(N)}(\mathbb{R}^d).$$

- Note that $T^{(N)}(\mathbb{R}^d)$ has dimension $\sum_{i=0}^{N} d^i = (d^{N+1} - 1)/(d - 1)$.

Basic operations

- Addition, and scalar multiplication are defined in a canonical way
- For each $\mathbf{a}, \mathbf{b} \in T((\mathbb{R}^d))$ and $\lambda \in \mathbb{R}$ we set

$$\begin{aligned} \mathbf{a} + \mathbf{b} &:= (a_0 + b_0, \dots, a_n + b_n, \dots), \\ \lambda \mathbf{a} &:= (\lambda a_0, \dots, \lambda a_n, \dots), \\ \mathbf{a} \otimes \mathbf{b} &:= (c_0, \dots, c_n, \dots), \end{aligned}$$

  where $c_n := \sum_{k=0}^{n} a_k \otimes b_{n-k}$.
- $(T((\mathbb{R}^d)), +, \cdot, \otimes)$ is a real non-commutative algebra with neutral element $\mathbf{1} = (1, 0, \dots, 0, \dots)$.

Multi-indices

- For a multi-index $I := (i_1, \dots, i_n)$ we set $|I| := n$.
- If $n \geq 1$ or $n \geq 2$ we set $I' := (i_1, \dots, i_{n-1})$, and $I'' := (i_1, \dots, i_{n-2})$
- We also use the notation

$$\{I \colon |I| = n\} := \{1, \dots, d\}^n,$$

- Observe that multi-indices can be identified with words[1]
- For a multi-index I we set

$$e_I := e_{i_1} \otimes \cdots \otimes e_{i_n},$$

  where (e_i) is the basis in $\mathbb{R}^d$.
- Then $\{e_I \colon |I| = N\}$ is an orthonormal basis of $(\mathbb{R}^d)^{\otimes N}$.

[1] Given an alphabet Σ, a word is a finite string

$$w = a_1 a_2 \cdots a_n,$$

$a_i \in \Sigma$ and $|w| = n$.

- Moreover $e_\emptyset = (e_1, \dots, e_d)$. Then each $\mathbf{a} \in T((\mathbb{R}^d))$ can be written as

$$\mathbf{a} = \sum_{|I| \geq 0} a_I e_I,$$

with $a_I \in \mathbb{R}$.

- Finally, for each $\mathbf{a} \in T(\mathbb{R}^d)$ and each $\mathbf{b} \in T((\mathbb{R}^d))$ we set

$$\langle \mathbf{a}, \mathbf{b} \rangle := \sum_{|I| \geq 0} \langle a_I, b_I \rangle.$$

Observe in particular that $b_I = \langle e_I, \mathbf{b} \rangle$.

The Stratonovich integral

- Let X be a continuous semimartingale, and Y an adapted process. The **Stratonovich integral** is defined by

$$\int_0^T Y_t \circ dX_t := \lim_{|\Pi| \to 0} \sum_{i=0}^{n-1} Y_{\frac{t_i + t_{i+1}}{2}} (X_{t_{i+1}} - X_{t_i}),$$

  where $\Pi = \{0 = t_0 < t_1 < \cdots < t_n = T\}$ is a partition of the interval $[0, T]$, and $|\Pi| = \max(t_{i+1} - t_i)$ is the mesh of the partition.
- If Y is also a semimartingale, then

$$\int_0^\cdot Y_t \circ dX_t = \int_0^\cdot Y_t \, dX_t + \frac{1}{2}[X, Y]_t.$$

The signature

We consider a continuous, d-dimensional semimartingale $(X_t)_{t\in[0,T]}$

Definition

The **signature of** X is the $T((\mathbb{R}^d))$-valued process $(s,t)\mapsto \mathbb{X}_{s,t}$ whose components are recursively defined as

$$\langle e_\emptyset, \mathbb{X}_{s,t}\rangle := 1, \qquad \langle e_I, \mathbb{X}_{s,t}\rangle := \int_s^t \langle e_{I'}, \mathbb{X}_{s,r}\rangle \circ \mathrm{d}X_r^{i_n},$$

for each $I=(i_1,\dots,i_n)$, $I'=(i_1,\dots,i_{n-1})$ and $0\le s\le t\le T$, where $\circ$ denotes the Stratonovich integral. Its projection $\mathbb{X}^N$ on $T^{(N)}(\mathbb{R}^d)$ is given by

$$\mathbb{X}_{s,t}^N = \sum_{|I|\le N} \langle e_I, \mathbb{X}_{s,t}\rangle e_I$$

and is called **signature of** X **truncated at level** N. If $s=0$, we use the notation $\mathbb{X}_t$ and $\mathbb{X}_t^N$, respectively.

- Note that the signature of X is equal to the signature of $X - c$, $c \in \mathbb{R}$.
- Moreover,

$$\mathbb{X}_t = \Bigg(1, \int_0^t 1 \circ \underset{\cdot}{\mathsf{X}}^1_s, \dots, \int_0^t 1 \circ \underset{\cdot}{\mathsf{X}}^d_s, \int_0^t \Bigg(\int_0^s 1 \circ \underset{\cdot}{\mathsf{X}}^1_r\Bigg) \circ \underset{\cdot}{\mathsf{X}}^1_s,$$
$$\int_0^t \Bigg(\int_0^s 1 \circ \underset{\cdot}{\mathsf{X}}^1_r\Bigg) \circ \underset{\cdot}{\mathsf{X}}^2_s, \dots, \int_0^t \Bigg(\int_0^s 1 \circ \underset{\cdot}{\mathsf{X}}^d_r\Bigg) \circ \underset{\cdot}{\mathsf{X}}^d_s, \dots \Bigg).$$

- Using Itô integrals,

$$\mathbb{X}_t = \Bigg(1, X^1_t - X^1_0, \dots, X^d_t - X^d_0, \int_0^t (X^1_s - X^1_0) dX^1_s + \frac{1}{2}[X^1]_t,$$
$$\int_0^t (X^1_s - X^1_0) dX^2_s + \frac{1}{2}[X^1, X^2]_t, \dots,$$
$$\int_0^t (X^d_s - X^d_0) dX^d_s + \frac{1}{2}[X^d]_t, \dots \Bigg),$$

Example

Let $(X_t)_{t\in[0,T]}$ be a continuous semimartingale with $X_0 = 0$. Then the (Stratonovich) integration by parts formula yields, for any $i, j \in \{1, \dots, d\}$

$$\begin{aligned}\langle e_i, \mathbb{X}_T\rangle\langle e_j, \mathbb{X}_T\rangle = X_T^i X_T^j &= \int_0^T X_t^i \circ dX_t^j + \int_0^T X_t^j \circ dX_t^i, \\ &= \langle e_i \otimes e_j, \mathbb{X}_T\rangle + \langle e_j \otimes e_i, \mathbb{X}_T\rangle \\ &= \langle e_i \sqcup\!\sqcup\, e_j, \mathbb{X}_T\rangle.\end{aligned}$$

Definition

For multi-indices $I := (i_1, \dots, i_n)$ and $J := (j_1, \dots, j_m)$ the **shuffle product** is

$$e_I \sqcup\!\sqcup\, e_J := (e_{I'} \sqcup\!\sqcup\, e_J) \otimes e_{i_n} + (e_I \sqcup\!\sqcup\, e_{J'}) \otimes e_{j_m},$$

with $e_I \sqcup\!\sqcup\, e_\emptyset := e_\emptyset \sqcup\!\sqcup\, e_I = e_I$. It extends to $\mathbf{a}, \mathbf{b} \in T(\mathbb{R}^d)$ as

$$\mathbf{a} \sqcup\!\sqcup\, \mathbf{b} = \sum_{|I|,|J|\geq 0} a_I b_J (e_I \sqcup\!\sqcup\, e_J).$$

Proposition (Shuffle property)

We have that

$$\langle e_I, \mathbb{X} \rangle \langle e_J, \mathbb{X} \rangle = \langle e_I \sqcup\!\sqcup\, e_J, \mathbb{X} \rangle. \tag{9}$$

Lemma (Uniqueness of the signature)

Let $(X_t)_{t\in[0,T]}$ and $(Y_t)_{t\in[0,T]}$ be two continuous $\mathbb{R}^d$-valued semimartingales with $X_0 = Y_0 = 0$. Set $\widehat{X}_t := (t, X_t)$, $\widehat{Y}_t := (t, Y_t)$ and let $\widehat{\mathbb{X}}$ and $\widehat{\mathbb{Y}}$ be the corresponding signature processes. Then $\widehat{\mathbb{X}}_T = \widehat{\mathbb{Y}}_T$ if and only if $X_t = Y_t$ for each $t \in [0, T]$.

Lemma (Chen's identity)

Let $(X_t)_{t\in[0,T]}$ be an $\mathbb{R}^d$-valued semimartingale. Then

$$\mathbb{X}_{s,t} = \mathbb{X}_{s,u} \otimes \mathbb{X}_{u,t}$$

for each $s \le u \le t \le T$.

Universal approximation theorem

- Loosely speaking, it states that every quantity of the form

  $$f\Big((\widehat{\mathbb{X}}_t^2)_{t\in[0,T]}\Big)$$

  for some continuous map f and some $T > 0$ can be approximated arbitrarily well on compact sets by linear functions of the signature of the form $\langle \ell, \widehat{\mathbb{X}}_T \rangle$ where $\ell \in T(\mathbb{R}^d)$. Observe that the latter just involves the **final value** $\widehat{\mathbb{X}}_T$ of $\widehat{\mathbb{X}}$, instead of its whole trajectory.
- The proof is done by an application of the stone Weierstrass-theorem.

For each $N \in \mathbb{N}$ define the set

$$\mathcal{S}^{(N)} := \{(\widehat{\mathbb{X}}_t^N)_{t\in[0,T]}(\omega) \colon \omega \in \Omega\},$$

which, without loss of generality (passing to a subset of Ω of measure 1), corresponds to a set of signature paths of $\widehat{X}$ up to time T.

Theorem (Universal approximation theorem)

Let K be a compact subset of $\mathcal{S}^{(2)}$ and consider a continuous map $f : K \to \mathbb{R}$.[2] *Then for every $\varepsilon > 0$ there exists some $\ell \in T(\mathbb{R}^d)$ such that*

$$\sup_{(\widehat{\mathbb{X}}_t^2)_{t\in[0,T]}\in K} |f((\widehat{\mathbb{X}}_t^2)_{t\in[0,T]}) - \langle \ell, \widehat{\mathbb{X}}_T \rangle| < \varepsilon,$$

almost surely.

[2] Compactness and continuity are defined with respect to $d_{\mathcal{S}^{(2)}}$.

Signature models

We now come to the core of Christa Cuchiero, Guido Gazzani, and Sara Svaluto-Ferro (2023). „Signature-based models: Theory and calibration“. In: **SIAM journal on financial mathematics** 14.3, pp. 910–957. By $\widehat{\mathbb{X}}_t$ we denote the signature of the extension $\hat{X}_t$

Definition

A **signature model** is a stochastic process of the form

$$S_n(\ell)_t := \ell_\emptyset + \sum_{0<|I|\leq n} \ell_I \langle e_I, \widehat{\mathbb{X}}_t \rangle, \tag{10}$$

where $n \in \mathbb{N}$ and $\ell := \{\ell_\emptyset, \ell_I : 0 < |I| \leq n\}$.

Itô-integrals of processes of form (10) with respect to processes of form (10) are again processes of form (10). This includes in particular the signature $\widehat{\mathbb{S}}_n(\ell)$ of $\widehat{S}_n(\ell)_t := (t, S_n(\ell)_t)$ or expressions of the form

$$\int_0^{\cdot} S_n(\ell)_s dX_s^i.$$

In particular, this points to polynomial processes, which have been well studied.