IMPERIAL

Learning with Expected Signatures

Theory and Applications

Lorenzo Lucchese*
lorenzo.lucchese17@imperial.ac.uk

Almut E. D. Veraart
a.veraart@imperial.ac.uk

Mikko S. Pakkanen
m.pakkanen@imperial.ac.uk

*This research has been supported by the EPSRC Centre for Doctoral Training in Mathematics of Random Systems: Analysis, Modelling and Simulation (EP/S023925/1).

The signature

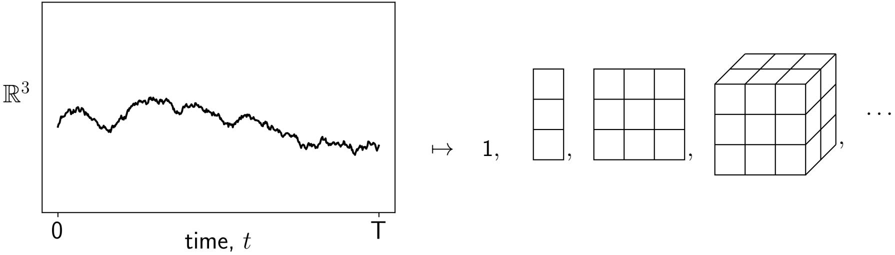
The signature transform of a path $\mathbb{X} = \{\mathbf{X}_t,\ t \in [0, T]\}$ is an embedding

$$C([0, T]; \mathbb{R}^d) \to T((\mathbb{R}^d)) = \mathbb{R} \oplus \mathbb{R}^d \oplus \mathbb{R}^{d \times d} \oplus \cdots,$$

mapping

$$\mathbb{X} \mapsto S(\mathbb{X})_{[0,T]} = \left(1, \int_{0 \le t \le T} \mathrm{d}\mathbf{X}_t, \iint_{0 \le t_1 \le t_2 \le T} \mathrm{d}\mathbf{X}_{t_1} \otimes \mathrm{d}\mathbf{X}_{t_2}, \dots \right),$$

where integration is defined in the *geometric rough* sense.



The expected signature

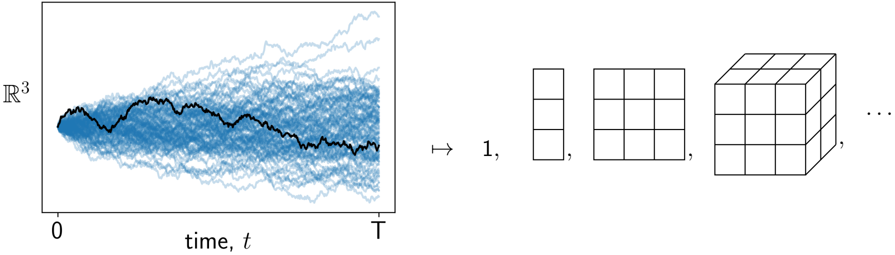
When $\mathbb{X}$ is a stochastic process the signature induces the map

$$\mathbb{P} \in \mathcal{P}(C([0,T];\mathbb{R}^d)) \mapsto \phi(T) := \mathbb{E}[S(\mathbb{X})_{[0,T]}] \in T((\mathbb{R}^d)).$$

The expected signature is characteristic for the law of $\mathbb{X}$

Under suitable conditions, ϕ is injective. [Chevyrev and Lyons, 2016]

This property has been leveraged to develop many ML algorithms.

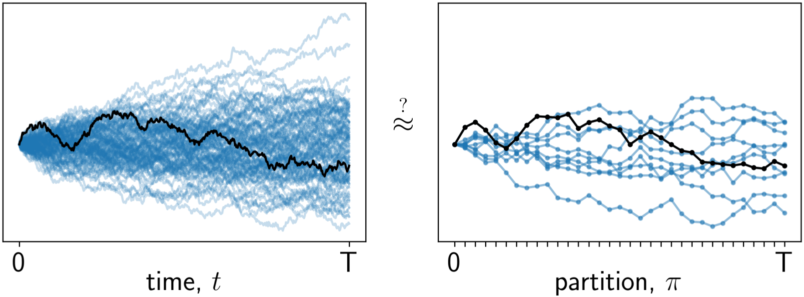


Estimating the expected signature

In practice:

- observe the path over partition π, i.e. the linear interpolation $\mathbb{X}^{\pi}$,
- observe finitely many samples $\mathbb{X}^{1,\pi_1}, \ldots, \mathbb{X}^{N,\pi_N}$.

$$\phi(T) := \mathbb{E}[S(\mathbb{X})_{[0,T]}] \qquad \hat{\phi}^{\Pi(N)}(T) := \frac{1}{N}\sum_{n=1}^{N} S(\mathbb{X}^{n,\pi_n})_{[0,T]}$$



Estimating the expected signature

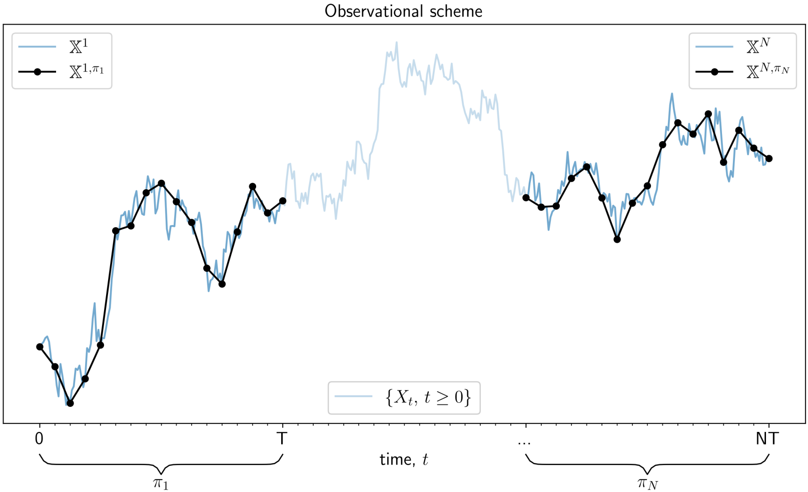


Figure 1: $\{\mathbf{X}_t,\ t \geq 0\}$ over $\Pi(N) := \pi_1 \cup \cdots \cup ((N-1)T + \pi_N) \subseteq [0, NT]$.

Estimating the expected signature

Fixing $k \in \mathbb{N}$ (signature level) decompose

$$\hat{\phi}_k^{\Pi(N)}(T) - \phi_k(T) =$$

$$\frac{1}{N}\sum_{n=1}^{N} S^k(\mathbb{X}^{n,\pi_n})_{[0,T]} - \mathbb{E}[S^k(\mathbb{X})_{[0,T]}] =$$

$$\frac{1}{N}\sum_{n=1}^{N} \underbrace{S^k(\mathbb{X}^{n,\pi_n})_{[0,T]} - S^k(\mathbb{X}^n)_{[0,T]}}_{\text{(in-fill)}} + \underbrace{\frac{1}{N}\sum_{n=1}^{N} S^k(\mathbb{X}^n)_{[0,T]} - \mathbb{E}[S^k(\mathbb{X})_{[0,T]}]}_{\text{(long-span)}}.$$

$S^k(\mathbb{X}^\pi)_{[0,T]} \xrightarrow{L^2} S^k(\mathbb{X})_{[0,T]},\ |\pi| \to 0$

a (dependent) CLT as $N \to \infty$

In-fill asymptotics

Assumption (Continuity of $\mathbb{X}$)

For $0 \leq s < t \leq T$ assume

(Aα) $\|\mathbf{X}_{s,t}\|_{L^p} \lesssim |t-s|^{\alpha}$,

(Aδ) $\|\mathbb{E}_s[\mathbf{X}_{s,t}]\|_{L^p} \lesssim |t-s|^{\delta}$.

Theorem (in-fill)

Let $k \in \mathbb{N}$, $m \geq 2$ and $p = mk$. Assume $\mathbb{X}$ satisfies either

1. (Aα) *for $\alpha > 1/2$, or*
2. (Aα), (Aδ) *for $\alpha = 1/2$, $\delta \geq 1$,*

and π is refining with $|\pi| \downarrow 0$ "fast" then

$$S^k(\mathbb{X}^\pi)_{[0,T]} \xrightarrow{L^m} S^k(\mathbb{X})_{[0,T]}, \ |\pi| \downarrow 0, \quad \textit{with explicit rate.}$$

Examples: Itô processes/diffusions, GPs, fBm with π dyadic.

Long-span asymptotics

	$\{\mathbf{X}_t,\ t \geq 0\}$		$\{\Pi(N),\ N \geq 1\}$	
	$\mathbb{X}^1$	$\{\mathbb{X}^n,\ n \geq 1\}$	π_n	
(a)	❶ or ❷ with $m > 2$	stationary & ergodic	refining	$\lvert\Pi(N)\rvert \to 0$ "fast"
(b)		strong mixing	$\lvert\Pi(N)\rvert \to 0$ "faster"	

Theorem (long-span)

- *(a)* $\implies \hat{\phi}_k^{\Pi(N)}(T) \xrightarrow{L^2} \phi_k(T),\ N \to \infty$
- *(a)* + *(b)* $\implies \sqrt{N}\left(\hat{\phi}_k^{\Pi(N)}(T) - \phi_k(T)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_k),\ N \to \infty$

Examples: Itô diffusions, GPs with $\Pi(N)$ expanding dyadic.

Variance Reduction via Martingale Correction

When $\mathbb{X}$ is a (semi)martingale, the signature is

$$S^k(\mathbb{X})_{[0,T]} = \int_0^T S^{k-1}(\mathbb{X})_{[0,t]} \otimes \circ\, \mathrm{d}\mathbf{X}_t. \qquad \text{(Strat.)}$$

Given N i.i.d. observations of a martingale $\mathbb{X}$, the estimator

$$\hat{\phi}_k^{N,c}(T) := \frac{1}{N}\sum_{n=1}^{N}\Big(S^k(\mathbb{X}^n)_{[0,T]} - cS_c^k(\mathbb{X}^n)_{[0,T]}\Big),$$

$$\text{where} \quad S_c^k(\mathbb{X})_{[0,T]} = \int_0^T S^{k-1}(\mathbb{X})_{[0,t]} \otimes \mathrm{d}\mathbf{X}_t, \qquad \text{(Itô)}$$

improves $\hat{\phi}_k^N(T)$ since for any entry I of the k-th level signature

$$\mathbb{E}[\hat{\phi}_I^{N,c^*}(T)] = \mathbb{E}[\hat{\phi}_I^N(T)], \quad \mathrm{Var}(\hat{\phi}_I^{N,c^*}(T)) = (1-\rho_I^2)\mathrm{Var}(\hat{\phi}_I^N(T)),$$

where $\rho_I := \mathrm{Corr}(S^I(\mathbb{X})_{[0,T]}, S_c^I(\mathbb{X})_{[0,T]})$.

In practice: observe over π and estimate c_π^* from $\mathbb{X}^{1,\pi}, \ldots, \mathbb{X}^{N,\pi}$.

Path-dependent Option Pricing [Lyons et al., 2021]

For a large class of path-dependent payoffs $F = F(\mathbb{X}) \approx \langle f, S(\hat{\mathbb{X}}_{ll})_{[0,T]} \rangle$.
Under pricing measure $\mathbb{Q}$ and deterministic discounting Z_T,

$$\text{price}(F) = \mathbb{E}^{\mathbb{Q}}[Z_T F] \approx \langle f, Z_T \mathbb{E}^{\mathbb{Q}}[S(\hat{\mathbb{X}}_{ll})_{[0,T]}] \rangle = \langle f, Z_T \phi(T) \rangle.$$

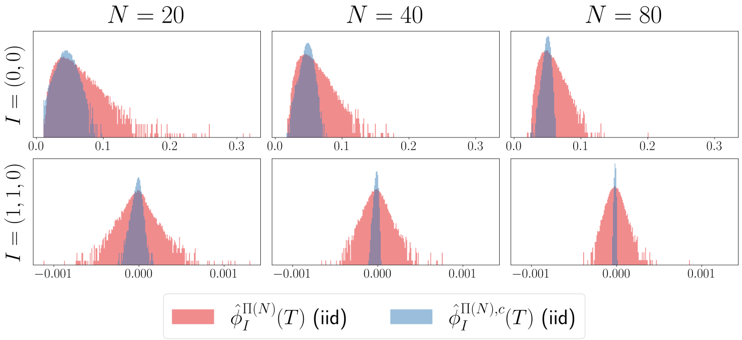
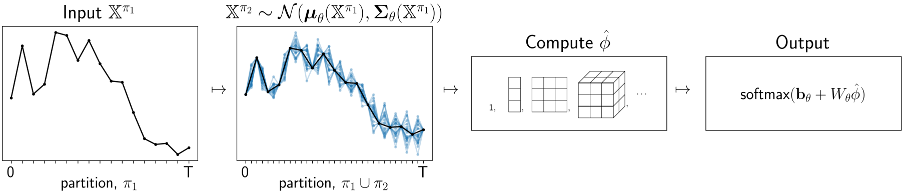


Figure 2: Distributions of $\hat{\phi}^{\Pi(N),\hat{c}^*}(T)$ where $\mathbb{Q}$ defines a Heston process.

GP augmented ES classifier [Triggiano and Romito, 2024]

Time series classification model



trained via gradient descent with categorical cross-entropy loss.

	Predictive Accuracy [%]		
	FBM	OU	Bidim
GPES	95.62 (0.18)	62.20 (0.70)	79.33 (0.46)
GPES-MC	95.26 (0.70)	88.26 (0.31)	88.97 (0.44)
t-stat	1.49	−101.92	−45.52
p-value	0.15	0.00	0.00

Signature of Expected Signature [Lemercier et al., 2021]

$\forall$ continuous $f : \mathcal{P}(C([0,T];\mathbb{R}^d)) \mapsto \mathbb{R}$, $\exists \beta \in T((\mathbb{R}^d))^*$ s.t.

$$f(\mathbb{P}) \overset{\phi \text{ characteristic}}{\approx} F(\phi(T)) \overset{\text{universality } S}{\approx} \langle \beta, S(\Phi)_{[0,T]} \rangle.$$

Model-free distributional regression:

$$y_i = \langle \beta, S(\hat{\Phi}^c(\mathbb{X}_i^{1,\pi_1}, \ldots, \mathbb{X}_i^{N,\pi_N}))_{[0,T]} \rangle + \epsilon_i.$$

	Predictive MSE [$\times 10^{-3}$]				
	Ideal Gas		Rough Volatility		
	small r	large r	$N = 20$	$N = 50$	$N = 100$
SES	12.7 (2.3)	0.9 (0.3)	1.49 (0.39)	0.33 (0.13)	0.20 (0.08)
SES-MC	13.1 (4.5)	0.7 (0.2)	1.26 (0.48)	0.31 (0.09)	0.19 (0.05)
t-stat	-0.29	1.41	0.87	0.63	0.29
p-value	0.79	0.23	0.43	0.56	0.79

Bibliography I

Chevyrev, I. and Lyons, T. (2016).
Characteristic functions of measures on geometric rough paths.
The Annals of Probability, 44(6):4049–4082.

Lemercier, M., Salvi, C., Damoulas, T., Bonilla, E. V., and Lyons, T. (2021).
Distribution regression for sequential data.
In *24th International Conference on Artificial Intelligence and Statistics (AISTATS 2021)*, Proceedings of Machine Learning Research, pages 3754–3762. Journal of Machine Learning Research.

Lyons, T., Nejad, S., and Pérez Arribas, I. (2021).
Non-parametric pricing and hedging of exotic derivatives.
Applied Mathematical Finance, 27(6):457–494.

Triggiano, F. and Romito, M. (2024).
Gaussian Processes Based Data Augmentation and Expected Signature for Time Series Classification.
IEEE Access, 12:80884–80895.