Signature-based validation of real-world economic scenarios

Hervé Andrès, Alexandre Boumezoued, Benjamin Jourdain

Signature-based validation of real-world economic scenarios

Hervé Andrès[1,2], Alexandre Boumezoued[1], and Benjamin Jourdain[2]

[1]Milliman R&D, Paris, France
[2]CERMICS, École des Ponts, INRIA, Marne-la-Vallée, France.

July 29, 2022

Abstract

We propose a new approach for the validation of real-world economic scenario motivated by insurance applications. This approach is based on the statistical test developed by Chevyrev and Oberhauser [6] and relies on the notions of signature and maximum mean distance. This test allows to check whether two samples of stochastic processes paths come from the same distribution. Our contribution is to apply this test to two stochastic processes, namely the fractional Brownian motion and the Black-Scholes dynamics. We analyze its statistical power based on numerical experiments under two constraints:

1. we work in an asymetric setting in which we compare a large sample that represents simulated real-world scenarios and a small sample that mimics information from historical data, both with a monthly time step as often considered in practice and
2. we make the two samples identical from the perspective of validation methods used in practice, i.e. we impose that the marginal distributions of the two samples are the same at a given one-year horizon.

By performing specific transformations of the signature, we obtain high statistical powers and demonstrate the potential of this validation approach for real-world economic scenarios. We also discuss several challenges related to the numerical implementation of this approach, and highlight its domain of validity in terms of distance between models and the volume of data at hand.

Keywords: real-world economic scenarios, economic scenarios validation, insurance, signature, maximum mean distance.

1. Introduction

Real-world economic scenarios provide stochastic forecasts of economic variables like interest rates, equity stocks or indices, inflation, etc. and are widely used in the insurance sector for a

variety of applications including asset and liability management (ALM) studies, strategic asset allocation, computing the Solvency Capital Requirement (SCR) within an Internal Model or pricing assets or liabilities including a risk premium. Unlike risk-neutral economic scenarios that aim at capturing market expectations about future evolutions at some point in time, real-world economic scenarios aim at being realistic in view of the historical data and/or expert expectations about future outcomes. The generation of real-world scenarios raises two essential questions:

1. Which models should be used?
2. Given a set of real-world economic scenarios, how does one assess their reasonableness?

In this article, we will focus on the second question which is often referred to as the scenarios validation. In the risk-neutral framework, the validation step consists for example in verifying the martingality of the discounted values along each scenario. In the real-world framework, the most widespread practice is to perform a so-called point-in-time validation. It consists in analyzing the distribution of some variables derived from the generated scenarios (for example annual log-returns for equity stocks or relative variation for an inflation index) at some specific dates like one year which is the horizon considered in the Solvency II directive. Generally, this analysis only focuses on the first moments of the one year distribution as real-world models are often calibrated by a moment-matching approach. The main drawback of this approach is that it only allows to capture properties of the simulated scenarios at some point in time. In particular, the paths between $t = 0$ and $t = 1$ year leading to the analyzed variables are not studied so that properties like clustering, smoothness, high-order autocorrelation, etc. are not captured. Capturing these properties has its importance as their presence or absence in the economic scenarios can have an impact for the above-mentioned applications, for example on a strategic asset allocation having a monthly rebalancing frequency or on the pricing and hedging of options embedded in life insurance contracts with path-dependent payoff. In this paper, we propose to address this drawback by comparing the distribution of the stochastic process underlying the simulated paths to the distribution of the stochastic process underlying the historical paths. This can be done using a distance between probability measures, called the Maximum Mean Distance (MMD), and a mathematical object, called the signature, allowing to encode a continuous path in an efficient and parsimonious way. Based on these tools, Chevyrev and Oberhauser [6] design a statistical test allowing to accept or reject the hypothesis that the distributions of two samples of paths are equal. Our contribution is to apply this test to the validation of real-world stochastic scenarios. To this end, we specify the validation problem in insurance as an hypothesis testing problem and we study this test from a numerical point of view. Two constraints are considered in the numerical experiments. The first one is to impose that the distribution of the annual increments is the same in the two compared samples, which implies that the current validation procedures cannot distinguish the two samples. Secondly, in order to mimic the operational process of real world scenarios validation in insurance, we consider samples of different sizes: the first sample consisting of the historical paths is of small size (typically below 50) while the second sample consisting of the simulated scenarios is of greater size (typically above 1000). Our aim is to demonstrate the high statistical power of the test under these constraints. Numerical results are presented for two risk drivers, namely the price and the volatility of an equity stock. For the price, the two samples of paths are generated using two specifications of the volatility in the widespread Black-Scholes dynamics, while for the volatility, two samples of paths are generated using different Hurst parameter of a fractional Brownian motion. This model for the volatility is inspired by the work of Gatheral et al [13] who show that the fractional Brownian motion is highly consistent with historical volatility.

In the literature, variety of models have been developed for real-world applications in insurance. Those applications relate to (i) pricing and valuation of insurance products, (ii) hedging strategies for annuity portfolios and (iii) risk calculation for economic capital assessment. On item (i), we can mention the work of Boudreault et al [4] who study the impact on Conditional Tail Expectation provision of GARCH and regime-switching models calibrated on historical data, and the work of Graf et al [15] who perform simulations under the real-world probability measure to estime the risk-return profile of various old-age provision products. On item (ii), Zhu et al [30] measure the hedging error of several dynamic hedging strategies along real-world scenarios for cash balance pension plans while Lin and Yang [23] calculate the value of a large variable annuity portfolio and its hedge using nested simulations (real-world scenarios for the outer simulations and risk-neutral scenarios for the inner simulations). Finally, on item (iii), Hardy et al [20] compare several real-world models for the equity return process in terms of fitting quality and resulting capital requirements and discuss the problem of the validation of real-world scenarios. Similarly, Otero et al [25] measure the impact on the Solvency II capital requirements (SCR) of the use of a regime-switching model in comparison to lognormal, GARCH and E-GARCH models. Floryszcak et al [11] introduce a simple heuristic (but not realistic) model for equity returns allowing to avoid over-assessment of the SCR specifically after market disruptions. On the other hand, Asadi and Al Janadi [1] propose a more complex model for stocks based on ARMA and GARCH processes that results in a higher SCR than in the Solvency II standard model. This literature shows the importance of real-world economic scenarios in various applications in insurance. In particular, we observe that the question of the consistency of the generated real-world scenarios is barely discussed or only from a specific angle like the model likelihood or the ability of the model to reproduce the 1 in 200 worst shock observed on the market.

This article is organized as follows: as a preliminary, we introduce the Maximum Mean Distance and the signature before describing the statistical test proposed by Chevyrev and Oberhauser [6], which is based on these two notions and allows to test if two stochastic processes have the same law using finite numbers of their sample paths. Then, we study this test from a numerical point of view, with a focus on its power in settings that are realistic in view of insurance applications. A final section is dedicated to a more thorough presentation of the signature and its properties.

2. From the MMD and the signature to a two-sample test for stochastic processes

In this section, we start by introducing the Maximum Mean Distance (MMD), which allows to measure how similar two probability measures are. Secondly, Reproducing Kernel Hilbert Spaces (RKHS) are presented as they are key to obtain a simple formula for the MMD. Then, we briefly introduce the signature object and we show how it allows to construct a RKHS that we can use to make the MMD a metric able to discriminate two probability measures defined on the bounded variation paths quotiented by some equivalence relation. Finally, the statistical test underlying the signature-based validation is introduced. In what follows, $\mathcal{X}$ is a topological space.

2.1. The Maximum Mean Distance

Definition 1 (Maximum Mean Distance)**.** *Let $\mathcal{G}$ be a class of functions $f : \mathcal{X} \to \mathbb{R}$ and μ, ν two Borel probability measures defined on $\mathcal{X}$. The Maximum Mean Distance (MMD) is defined as:*

$$MMD_{\mathcal{G}}(\mu, \nu) = \sup_{f \in \mathcal{G}} \left| \int_{\mathcal{X}} f(x)\mu(dx) - \int_{\mathcal{X}} f(x)\nu(dx) \right|. \tag{1}$$

Depending on $\mathcal{G}$, the MMD is not necessarily a metric[1], i.e. we could have $MMD_{\mathcal{G}}(\mu, \nu) = 0$ for some $\mu \neq \nu$ if the class of functions $\mathcal{G}$ is not rich enough. A sufficiently rich class of functions that makes $MMD_{\mathcal{G}}$ a metric is for example the space of bounded continuous functions on $\mathcal{X}$ equipped with a metric d (Lemma 9.3.2 of [8]). A sufficient condition on $\mathcal{G}$ for $MMD_{\mathcal{G}}$ to be a metric will be given in Theorem 11.

As presented in Definition 1, the MMD appears more as a theoretical tool than a practical one since computing this distance seems impossible in practice due to the supremum over a class of functions. However, if this class of function is the unit ball in a reproducing kernel Hilbert space (RKHS), the MMD is much simpler to estimate. Before setting out this result precisely, let us make a quick reminder about Mercer kernels and RKHSs.

Definition 2 (Mercer kernel)**.** *A mapping $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called a Mercer kernel if it is continuous, symmetric and positive semi-definite[2] i.e. for all finite sets $\{x_1, \dots, x_k\} \subset \mathcal{X}$ and for all $(\alpha_1, \dots, \alpha_k) \in \mathbb{R}^k$, the kernel K satisfies:*

$$\sum_{i=1}^{k} \sum_{j=1}^{k} \alpha_i \alpha_j K(x_i, x_j) \geq 0. \tag{2}$$

We set:

$$\mathcal{H}_0 = \text{span}\{K_x := K(x, \cdot) \mid x \in \mathcal{X}\}. \tag{3}$$

With these notations, we have the following theorem due to Moore-Aronszajn (see Theorem 2 of Chapter III in [7]):

Theorem 1 (Moore-Aronszajn)**.** *Let K be a Mercer kernel. Then, there exists a unique Hilbert space $\mathcal{H} \subset \mathbb{R}^{\mathcal{X}}$ with scalar product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ satisfying the following conditions:*

(i) $\mathcal{H}_0$ is dense in $\mathcal{H}$

(ii) For all $f \in \mathcal{H}$, $f(x) = \langle K_x, f \rangle_{\mathcal{H}}$ (reproducing property).

Remark 1. *The obtained Hilbert space $\mathcal{H}$ is said to be a reproducing kernel Hilbert space (RKHS) whose reproducing kernel is K.*

---

[1] It is actually a pseudo-metric, that is a non-negative function $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ such that for all $x, y, z \in \mathcal{X}$,

(i) $d(x, x) = 0$

(ii) $d(x, y) = d(y, x)$ (symmetry)

(iii) $d(x, z) \leq d(x, y) + d(y, z)$ (triangle inequality).

The missing property for d to be a metric is the ability of d to distinguish elements of $\mathcal{X}$, i.e. the property that $d(x, y) = 0$ implies $x = y$ for all $x, y \in \mathcal{X}$.

[2] In the kernel learning litterature, it is common to use the terminology "positive definite" instead of "semi-positive definite" but we prefer the latter one in order to be consistent with the linear algebra standard terminology.

We may now state the main theorem about the MMD.

Theorem 2. *Let $(\mathcal{H}, K)$ a reproducing kernel Hilbert space and let $\mathcal{G} := \{f \in \mathcal{H} \mid \|f\|_{\mathcal{H}} \leq 1\}$. If $\int_{\mathcal{X}} \sqrt{K(x,x)}\mu(dx) < \infty$ and $\int_{\mathcal{X}} \sqrt{K(x,x)}\nu(dx) < \infty$, then for X,X' independent random variables distributed according to μ and Y,Y' independent random variables distributed according to ν and independent from X and X', $K(X,X')$, $K(Y,Y')$ and $K(X,Y)$ are integrable and:*

$$MMD^2_{\mathcal{G}}(\mu,\nu) = \mathbb{E}[K(X,X')] + \mathbb{E}[K(Y,Y')] - 2\mathbb{E}[K(X,Y)]. \tag{4}$$

The proof of this theorem is provided in Appendix A. A natural question at this stage is that of the choice of the Mercer kernel in order to obtain a metric on the space of probability measures defined on the space of continuous mappings or at least on a non-trivial subspace of the space of continuous mappings. In [6], Chevyrev and Oberhauser constructed such Mercer kernel using the signature, which we will know define.

2.2. The signature

We call "path" any continuous mapping from some time interval $[0,T]$ to a finite dimensional vector space E which we equip with a norm $\|\cdot\|_E$. We denote by $\otimes$ the tensor product defined on $E \times E$ and by $E^{\otimes n}$ the tensor space obtained by taking the tensor product of E with itself n times:

$$E^{\otimes n} = \underbrace{E \otimes \cdots \otimes E}_{n \text{ times}}. \tag{5}$$

The space in which the signature takes its values is called the space of formal series of tensors. It can be defined as:

$$T(E) = \left\{(\mathbf{t}^n)_{n\geq 0} \mid \forall n \geq 0, \mathbf{t}^n \in E^{\otimes n}\right\} \tag{6}$$

with the convention $E^{\otimes 0} = \mathbb{R}$.

In a nutshell, the signature of a path X is the collection of all iterated integrals of X against itself. In order to be able to define these iterated integrals of X, one needs to make some assumptions about the regularity of X. A simple assumption is that X is of bounded variation.

Definition 3 (Bounded variation path)**.** *We say that a path $X : [0,T] \to E$ is of bounded variation on $[0,T]$ if its total variation*

$$\|X\|_{1,[0,T]} := \sup_{D=\{t_0,\ldots,t_r\}\in\mathcal{D}} \sum_{i=0}^{r-1} \|X_{t_{i+1}} - X_{t_i}\|_E \tag{7}$$

is finite with $\mathcal{D} = \{D = \{t_0,\ldots,t_r\} \mid r \in \mathbb{N}^*, t_0 = 0 < t_1 < \cdots < t_r = T\}$ the set of all subdivisions of $[0,T]$.

Notation 1. *We denote by $\mathcal{C}^1([0,T],E)$ the set of bounded variation paths from $[0,T]$ to E.*

Remark 2. *Intuitively, a bounded variation path on $[0,T]$ is a path whose graph vertical arc length is finite. In fact, if X is a real-valued continuously differentiable path on $[0,T]$, then*

$$\|X\|_{1,[0,T]} = \int_0^T |X'_t| dt. \tag{8}$$

We can now define the signature of a bounded variation path.

Definition 4 (Signature)**.** *Let $X : [0,T] \to E$ be a bounded variation path. The signature of X on $[0,T]$ is defined as $S_{[0,T]}(X) = (\mathbf{X}^n)_{n\geq 0}$ where by convention $\mathbf{X}^0 = 1$ and*

$$\mathbf{X}^n = \int_{0\leq u_1<u_2<\cdots<u_n\leq T} dX_{u_1} \otimes \cdots \otimes dX_{u_n} \in E^{\otimes n}. \tag{9}$$

where the integrals must be understood in the sense of Riemann-Stieljes. We call $\mathbf{X}^n$ the term of order n of the signature and $S^N_{[0,T]}(X) = (X^n)_{0\leq n\leq N}$ the truncated signature at order N. Note that when the time interval is clear from the context, we will omit it in the subscript of S.

Example 1. *If X is a one-dimensional bounded variation path, then its signature over $[0,T]$ is very simple as it reduces to the powers of the increment $X_T - X_0$, i.e. for any $n \geq 0$:*

$$\mathbf{X}^n = \frac{1}{n!}(X_T - X_0)^n. \tag{10}$$

The above definition could be extended to less regular paths, namely to paths of finite p-variation with $p < 2$. In this case, the integrals can be defined in the sense of Young [29]. However, if $p \geq 2$, it is no longer possible to define the iterated integrals. Still, it is possible to give a sense to the signature but the definition is much more involved and relies on the rough path theory so we refer the interested reader to [24].

For bounded variation paths, the signature takes its values in the space of finite formal series (as a consequence of Proposition 2.2 of [24]):

$$T^*(E) := \left\{ \mathbf{t} \in T(E) \mid \|\mathbf{t}\| := \sqrt{\sum_{n\geq 0} \|\mathbf{t}^n\|^2_{E^{\otimes n}}} < \infty \right\} \tag{11}$$

where $\|\cdot\|_{E^{\otimes n}}$ is the norm induced by the scalar product $\langle\cdot,\cdot\rangle_{E^{\otimes n}}$ defined on $E^{\otimes n}$ by:

$$\langle x, y\rangle_{E^{\otimes n}} = \sqrt{\sum_{I=(i_1,\ldots,i_n)\in\{1,\ldots,d\}^n} x_I y_I} \quad \text{for } x, y \in E^{\otimes n} \tag{12}$$

with d the dimension of E and x_I (resp. y_I) the coefficient at position I of x (resp. y).

The signature is a powerful tool allowing to encode a path in a hierarchical and efficient manner. In fact, two bounded variation paths having the same signature are equal up to an equivalence relation (the so-called tree-like equivalence, denoted by $\sim_t$, and defined in Section 4.2). In other words, the signature is one-to-one on the space $\mathcal{P}^1([0,T],E) := \mathcal{C}^1([0,T],E)/\sim_t$ of bounded variation paths quotiented by the tree-like equivalence relation. This is presented in a more comprehensive manner in Section 4. Now, we would like to characterize the law of stochastic processes with bounded variation sample paths using the expected signature, that is the expectation of the signature taken component-wise. In a way, the expected signature is to stochastic processes what the sequence of moments is to $\mathbb{R}^d$-valued random variables. Thus, in the same way that the sequence of moments characterizes the law of random variables only if the moments do not grow too fast, we need that the terms of the expected signature do not grow too fast in order to be able to characterize the law of stochastic processes. In order to avoid to have to restrict the study to laws with compact support (as assumed by Fawcett [9]), Chevyrev and Oberhauser [6] propose to apply a normalization mapping to the signature ensuring that the norm of the normalized signature is bounded. This property allows them to prove the characterization of the law of a stochastic process by its expected normalized signature

(Theorem 9 in Section 4). One of the consequences of this result is the following theorem, which makes the connection between the MMD and the signature and represents the main theoretical result underlying the signature-based validation. Its proof can be found in Appendix C.

Theorem 3. *Let E be a Hilbert space and $\langle \cdot, \cdot \rangle$ the scalar product on $T^*(E)$ defined by:*

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{n \geq 0} \langle \mathbf{x}^n, \mathbf{y}^n \rangle_{E^{\otimes n}} \tag{13}$$

for all $\mathbf{x}$ and $\mathbf{y}$ in $T^*(E)$. Then the signature kernel defined on $\mathcal{P}^1([0,T],E)$ by:

$$K^{sig}(x,y) = \langle \Phi(x), \Phi(y) \rangle, \tag{14}$$

where Φ is the normalized signature (see Theorem 9 in Section 4), is a Mercer kernel and we denote by $\mathcal{H}^{sig}$ the associated RKHS. Moreover, $MMD_{\mathcal{G}}$ where $\mathcal{G}$ is the unit ball of $\mathcal{H}^{sig}$ is a metric on the space $\mathcal{M}$ defined as:

$$\mathcal{M} = \left\{ \mu \text{ Borel probability measure defined on } \mathcal{P}^1([0,T],E) \middle| \int_{\mathcal{P}^1} \sqrt{K^{sig}(x,x)}\mu(dx) < \infty \right\} \tag{15}$$

and we have:

$$MMD_{\mathcal{G}}(\mu,\nu) = \mathbb{E}[K^{sig}(X,X')] + \mathbb{E}[K^{sig}(Y,Y')] - 2\mathbb{E}[K^{sig}(X,Y)] \tag{16}$$

where X,X' are independent random variables distributed according to μ and Y,Y' are independent random variables distributed according to ν and independent from X and X'.

Based on this theorem, Chevyrev and Oberhauser propose a two-sample statistical test that allows to test whether two samples of paths come from the same distribution and that we now introduce.

2.3. A two-sample test for stochastic processses

Assume that we are given a sample $(X_1, \ldots, X_m)$ consisting of m independent realizations of a stochastic process of unknown law $\mathbb{P}_X$ and an independent sample $(Y_1, \ldots, Y_n)$ consisting of n independent realizations of a stochastic process of unknown law $\mathbb{P}_Y$. We assume that both processes are in $\mathcal{X} = \mathcal{P}^1([0,T],E)$ almost surely. A natural question is whether $\mathbb{P}_X = \mathbb{P}_Y$. Let us consider the following null and alternative hypotheses:

$$\begin{array}{ll} H_0: & \mathbb{P}_X = \mathbb{P}_Y \\ H_1: & \mathbb{P}_X \neq \mathbb{P}_Y. \end{array} \tag{17}$$

According to Theorem 3, we have $MMD_{\mathcal{G}}(\mathbb{P}_X, \mathbb{P}_Y) \neq 0$ under H_1 while $MMD_{\mathcal{G}}(\mathbb{P}_X, \mathbb{P}_Y) = 0$ under H_0 when $\mathcal{G}$ is the unit ball of the reproducing Hilbert space associated to the signature kernel[3]. Moreover,

$$MMD^2_{\mathcal{G}}(\mathbb{P}_X, \mathbb{P}_Y) = \mathbb{E}[K(X,X')] + \mathbb{E}[K(Y,Y')] - 2\mathbb{E}[K(X,Y)] \tag{18}$$

[3]Note that we use the notation K instead of K^{sig} for the signature kernel in this section as there is no ambiguity.

where X,X' are two random variables of law $\mathbb{P}_X$ and Y, Y' are two random variables of law $\mathbb{P}_Y$ with X, X', Y, Y' independent. This suggests to consider the following test statistic:

$$MMD^2_{m,n}(X_1,\dots,X_m,Y_1,\dots,Y_n) := \frac{1}{m(m-1)}\sum_{1\leq i\neq j\leq m} K(X_i,X_j) + \frac{1}{n(n-1)}\sum_{1\leq i\neq j\leq n} K(Y_i,Y_j) - \frac{2}{mn}\sum_{\substack{1\leq i\leq m\\ 1\leq j\leq n}} K(X_i,Y_j) \tag{19}$$

as it is an unbiased estimator of $MMD^2_{\mathcal{G}}(\mathbb{P}_X,\mathbb{P}_Y)$. This estimator is even strongly consistent as stated by the following theorem which is an application of the strong law of large numbers for two-sample U-statistics ([26]).

Theorem 4. *Assuming that*

(i) $\mathbb{E}\left[\sqrt{K(X,X)}\right] < \infty$ *with* X *distributed according to* $\mathbb{P}_X$ *and* $\mathbb{E}\left[\sqrt{K(Y,Y)}\right] < \infty$ *with* Y *distributed according to* $\mathbb{P}_Y$

(ii) $\mathbb{E}[|h|\log^+|h|] < \infty$ *where* $\log^+$ *is the positive part of the logarithm and*

$$h = K(X,X') + K(Y,Y') - \frac{1}{2}\left(K(X,Y) + K(X,Y') + K(X',Y) + K(X',Y')\right) \tag{20}$$

with X, X' *distributed according to* $\mathbb{P}_X$, Y, Y' *distributed according to* $\mathbb{P}_Y$ *and* X, X', Y, Y' *independent*

then[4]*:*

$$MMD^2_{m,n} \underset{m,n\to+\infty}{\overset{a.s.}{\to}} MMD^2_{\mathcal{G}}(\mathbb{P}_X,\mathbb{P}_Y). \tag{21}$$

Under H_1, $MMD^2_{\mathcal{G}}(\mathbb{P}_X,\mathbb{P}_Y) > 0$ so we reject the null hypothesis at level α if $MMD^2_{m,n}$ is greater than some threshold c_α. In [6], Chevyrev and Oberhauser consider a threshold based on Hoeffding's concentration inequality. However, given that this concentration inequality is very general and doesn't use at all the specific form of the MMD estimator, the obtained threshold is very conservative. Consequently, it is very difficult to reject the null hypothesis as reported by Chevyrev and Oberhauser in their numerical experiments. That is why we will present here a different threshold based on the asymptotic distribution of $MMD^2_{m,n}$ under H_0 which is due to Gretton et al (Theorem 12 in [16]).

Theorem 5. *Let us define the kernel* $\tilde{K}$ *by:*

$$\tilde{K}(x,y) = K(x,y) - \mathbb{E}[K(x,X)] - \mathbb{E}[K(X,y)] - \mathbb{E}[K(X,X')] \tag{22}$$

where X *and* X' *are i.i.d. samples drawn from* $\mathbb{P}_X$*. Assume that:*

(i) $\mathbb{E}[\tilde{K}(X,X')^2] < +\infty$

(ii) $m/N \to \rho \in (0,1)$ *as* $N = m + n \to +\infty$.

Under these assumptions, we have:

[4]We omit the dependency on $X_1,\dots,X_m$ and $Y_1,\dots,Y_n$ for notations simplicity.

1. *under H_0:*

$$N \times MMD^2_{m,n} \underset{N\to+\infty}{\overset{\mathcal{L}}{\to}} \frac{1}{\rho(1-\rho)} \sum_{\ell=1}^{+\infty} \lambda_\ell \left(G_\ell^2 - 1\right) \tag{23}$$

where $(G_\ell)_{\ell\geq 1}$ is an infinite sequence of independent standard normal random variables and the λ_ℓ's are the eigenvalues of the operator $S_{\tilde{K}}$ defined as:

$$\begin{array}{rccc} S_{\tilde{K}}: & L_2(\mathbb{P}_X) & \to & L_2(\mathbb{P}_X) \\ & g & \mapsto & \mathbb{E}[\tilde{K}(\cdot, X)g(X)] \end{array} \tag{24}$$

with $L_2(\mathbb{P}_X) := \{g \in \mathcal{X}^{\mathbb{R}} \mid \mathbb{E}[g(X)^2] < \infty\}$.

2. *under H_1:*

$$N \times MMD^2_{m,n} \underset{m,n\to+\infty}{\to} +\infty \tag{25}$$

This theorem indicates that if one wants to have a test with level α, one should take the $1-\alpha$ quantile of the above asymptotic distribution as rejection threshold. In order to approximate this quantile, Gretton et al [17] suggest four approaches:

1. Approximate the asymptotic distribution using a Gamma distribution
2. Approximate the asymptotic distribution using a Pearson distribution
3. Estimate the eigenvalues using the empirical Gram matrix spectrum
4. Use a resampling/bootstrap procedure

In our numerical experiments, we only investigated the third approach which relies on the following theorem (Theorem 1 of [17]).

Theorem 6. *Let $(\lambda_\ell)_{\ell\geq 1}$ be the eigenvalues defined in Theorem 5 and $(G_\ell)_{\ell\geq 1}$ be a sequence of i.i.d. standard normal variables. Define the centered Gram matrix $\hat{A}$ as:*

$$\hat{A} = HAH \tag{26}$$

where $A = (K(Z_i, Z_j))_{1\leq i,j\leq m+n}$ ($Z_i = X_i$ if $i \leq m$ and $Z_i = Y_{i-m}$ if $i > m$) and $H = I_{m+n} - \frac{1}{m+n}\mathbf{1}\mathbf{1}^T$. If $m/N \to \rho \in (0,1)$ as $N = m+n \to +\infty$, then

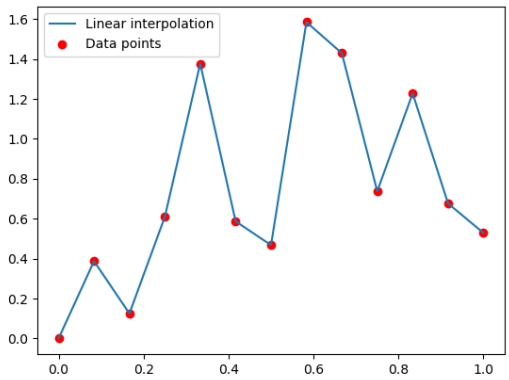
$$\frac{1}{\hat{\rho}(1-\hat{\rho})} \sum_{\ell=1}^{+\infty} \frac{\nu_\ell}{m+n} \left(G_\ell^2 - 1\right) \underset{N\to+\infty}{\overset{\mathcal{L}}{\to}} \frac{1}{\rho(1-\rho)} \sum_{\ell=1}^{+\infty} \lambda_\ell \left(G_\ell^2 - 1\right) \tag{27}$$

where $\hat{\rho} = m/N$ and the ν_ℓ's are the eigenvalues of $\hat{A}$.

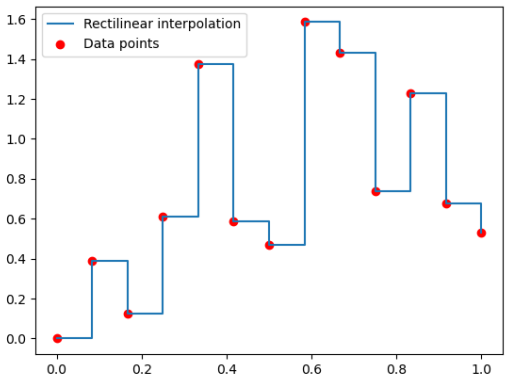
Therefore, we can approximate the asymptotic distribution in Theorem 5 by:

$$\frac{1}{\hat{\rho}(1-\hat{\rho})} \sum_{\ell=1}^{R} \frac{\nu_\ell}{m+n} \left(G_l^2 - 1\right) \tag{28}$$

with R the truncation order and $\nu_1 > \nu_2 > \cdots > \nu_R$ are the R first eigenvalues of $\hat{A}$ in decreasing order. A rejection threshold is then obtained by simulating several realizations of the above random variable and then computing their empirical quantile of level $1-\alpha$.



(a) Linear interpolation illustration



(b) Rectilinear interpolation illustration

Figure 1: Common embeddings

3. Implementation and numerical results

The objective of this section is to show the practical interest of the two-sample test described in the previous section for the validation of real-world economic scenarios. In the sequel, we refer to the two-sample test as the signature-based validation test. As a preliminary, we discuss the challenges implied by the practical implementation of the signature-based validation test.

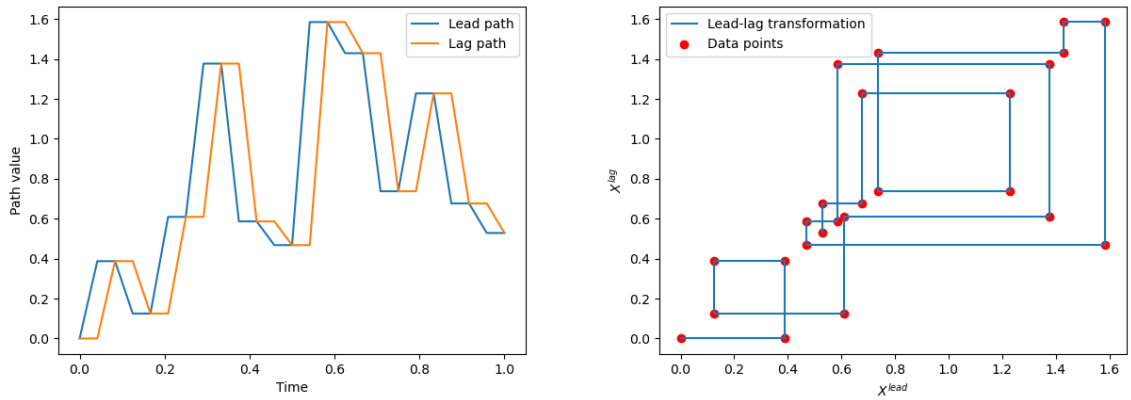
3.1. Practical implementation of the signature-based validation test

3.1.1. Signature of a finite number of observations

In practice, only a finite number of observations of the stochastic processes under study are available and one has to embed these observations into a continuous path in order to be able to compute the signature and *a fortiori* the MMD. The two most popular embeddings in the literature are the linear and the rectilinear embeddings. The former one consists in a plain linear interpolation of the observations, while the latter consists in an interpolation using only parallel shifts with respect to the x and y-axis as illustrated in Figure 1b. In the sequel, we will only use the linear embedding as this choice doesn't seem to have a material impact on the information contained in the obtained signature as shown by the comparative study led by Fermanian (section 4.2 of [10]).

3.1.2. Extracting information of a one-dimensional process

Remember that if X is a one-dimensional bounded variation path, then its signature over $[0, T]$ is equal to the powers of the increment $X_T - X_0$. As a consequence, finer information than the global increment about the evolution of X on $[0, T]$ is lost. In our applications, X represents an economic quantity like the level of an equity index so in most cases it is a one-dimensional stochastic process. In order to nonetheless be able to capture finer information about the evolution of X on $[0, T]$, one can apply a transformation to X to recover a multi-dimensional path. The two most widely used transformations are the time transformation and the lead-



(a) Lead and lag paths

(b) Lead-lag transformation

Figure 2: Lead-lag illustrations

lag transformation. The time transformation consists in considering the two-dimensional path $\hat{X}_t : t \mapsto (t, X_t)$ instead of $t \mapsto X_t$. The lead-lag transformation has been introduced by Gyurkó et al [18] in order to capture the quadratic variation of a path in the signature. Let $t_0 = 0 < t_1 < \cdots < t_N = T$ be a partition of $[0, T]$ and $(X_{t_i})_{i=0,\ldots,N}$ be a finite set of observations of a real-valued process X on this partition. We consider a set of dates $(t_{i+1/2})_{i=0,\ldots,N-1}$ such that $t_i < t_{i+1/2} < t_{i+1}$. The lead-lag transformation of $(X_{t_i})_{i=0,\ldots,N}$ is the two-dimensional path $t \mapsto (X_t^{lead}, X_t^{lag})$ defined on $[0, T]$ where:

1. the lead process $t \mapsto X_t^{lead}$ is the linear interpolation of the points $(X_{t_{i/2}}^{lead})_{i=0,\ldots,2N}$ with:

$$X_{t_{i/2}}^{lead} = \begin{cases} X_{t_j} & \text{if } i = 2j \\ X_{t_{j+1}} & \text{if } i = 2j+1 \end{cases} \tag{29}$$

2. the lag process $t \mapsto X_t^{lag}$ is the linear interpolation of the points $(X_{t_{i/2}}^{lag})_{i=0,\ldots,2N}$ with:

$$X_{t_{i/2}}^{lag} = \begin{cases} X_{t_j} & \text{if } i = 2j \\ X_{t_j} & \text{if } i = 2j+1 \end{cases} \tag{30}$$

Illustrations of the lead and lag paths as well as the lead-lag transformation are provided in Figure 2.

Remark 3. *The choice of the dates $(t_{i+1/2})_{i=0,\ldots,N-1}$ such that $t_i < t_{i+1/2} < t_{i+1}$ can be arbitrary since the signature is invariant by reparametrization (see Proposition 3 in Section 4.2).*

Finally, a third transformation can be constructed from the time and the lead-lag transformations. Indeed, given a finite set of observations $(X_{t_i})_{i=0,\ldots,N}$, one can consider the three-dimensional path $t \mapsto (t, X_t^{lead}, X_t^{lag})$. We call this transformation the time lead-lag transformation.

3.1.3. Numerical computation of the signature and the MMD

The numerical computation of the signature is performed using the `esig` Python package[5]. Because the signature is an infinite object, we compute in practice only the truncated signature up to some specified order R. The choice of the truncation order will be discussed in section 3.2.

[5] `https://esig.readthedocs.io/en/latest/`

3.2. Analysis of the statistical power

In this subsection, we apply the signature-based validation on simulated data, i.e. the two samples of stochastic processes realizations are numerically simulated. Keeping in mind insurance applications, the two-sample test is structured as follows:

- Each path is obtained by a linear interpolation from a set of 13 equally-spaced observations of the stochastic process under study. The first observation (i.e. the initial value of each path) is the same across all paths. These 13 observations represent monthly observations over a period of one year. In insurance practice, computational time constraints around the asset and liability models generally limit the simulation frequency to a monthly time step. The period of one year is justified by the fact that one needs to split the historical path under study into several shorter paths to get a test sample of size greater than 1. Because the number of historical data points is limited (about 30 years of data for the major equity indices), a split frequency of one year appears reasonable given the monthly observation frequency.

- The two samples are assumed to be of different sizes (i.e. $m \neq n$ with the notations of section 2.3). Several sizes m of the first sample will be tested while the size n of the second sample is always set to 1000. The first sample representing historical paths, we will mainly consider small values of m as m will be in practice equal to the number of years of available data (considering a split of the historical path in 1-year length paths as discussed above). The second sample representing simulated paths (for example by an Economic Scenario Generator), we take 1000 simulated paths as it corresponds to a lower bound of the number of scenarios used by insurers.

- As we aim to explore the capability of the two-sample test to capture properties of the paths that cannot be captured by looking at their marginal distribution at some dates, we impose that the distributions of the increment over $[0, T]$ of the two compared stochastic processes are the same. In other words, we only compare stochastic processes $(X_t)_{0\leq t\leq T}$ and $(Y_t)_{0\leq t\leq T}$ satisfying $X_T - X_0 \overset{\mathcal{L}}{=} Y_T - Y_0$. This constraint is motivated by the fact that two models that do not have the same marginal one-year distribution are already discriminated by the current validation methods. Moreover, it is a common practice in the insurance industry to calibrate the real-world models by minimizing the distance between model and historical moments so that the model marginal distribution is often close to the historical marginal distribution. Because of this constraint, we will remove the first order term of the signature in our estimation of the MMD because it does not provide useful statistical information as far as it is equal to the global increment $X_T - X_0$.

In order to measure the ability of the signature-based validation to distinguish two different samples of paths, we compute the statistical power of the underlying test, which is the probability to correctly reject the null hypothesis under H_1, by simulating 1000 times two samples of sizes m and n respectively and counting the number of times that the null hypothesis (the stochastic processes underlying the two samples are the same) is rejected. The rejection threshold is obtained using the empirical Gram matrix spectrum as described in Section 2.3. First, we generate a sample of size m and a sample of size n under H_0 to compute the eigenvalues of the matrix $\hat{A}$ in Theorem 6. Then, we keep the 20 first eigenvalues in decreasing order and we perform 10000 simulations of the random variable in equation (28) whose distribution approximates the MMD asymptotic distribution under H_0. The rejection threshold is obtained as the empirical quantile of level 99% of these realizations.

We will now present numerical results for two stochastic processes: the fractional Brownian motion and the Black-Scholes dynamics.

3.2.1. The fractional Brownian motion.

The fractional Brownian motion (fBm) is a generalization of the standard Brownian motion that, outside this standard case, is neither a semimartingale nor a Markov process and whose sample paths can be more or less regular than those of the standard Brownian motion. More precisely, it is the unique centered Gaussian process $(B_t^H)_{t\geq 0}$ whose covariance function is given by:

$$\mathbb{E}[B_s^H B_t^H] = \frac{1}{2}\left(s^{2H} + t^{2H} - (s-t)^{2H}\right) \quad \forall s,t \geq 0 \tag{31}$$

where $H \in]0,1[$ is called the Hurst parameter. Taking $H = 1/2$, we recover the standard Brownian motion. The fBm exhibits two interesting pathwise properties:

1. the fBm sample paths are $H-\epsilon$ Hölder for all $\epsilon > 0$, that is

$$\mathbb{P}\left(\sup_{s\neq t} \frac{\left|B_t^H - B_s^H\right|}{|t-s|^{H-\epsilon}} < \infty\right) = 1 \quad \forall \epsilon > 0. \tag{32}$$

   Thus, when $H < 1/2$, the fractional Brownian motion sample paths are rougher than those of the standard Brownian motion and when $H > 1/2$, they are smoother.

2. the increments are correlated:

$$\mathbb{E}[(B_t^H - B_s^H)(B_v^H - B_u^H)] = \frac{1}{2}\left(|s-v|^{2H} + |t-u|^{2H} - |t-v|^{2H} - |s-u|^{2H}\right) \quad \forall s,t,u,v \geq 0. \tag{33}$$

   In particular, if $s < t < u < v$, then $\mathbb{E}[(B_t^H - B_s^H)(B_v^H - B_u^H)]$ is positive if $H > 1/2$ and negative if $H < 1/2$ since $x \mapsto x^{2H}$ is convex if $H > 1/2$ and concave otherwise.

One of the main motivations for studying this process is the work of Gatheral et al [13] which shows that the historical volatility of many financial indices essentially behaves as a fBm with a Hurst parameter around 10%.

In the following numerical experiments, we will compare samples from a fBm with Hurst parameter H and samples from a fBm with a different Hurst parameter H'. One can easily check that B_1^H has the same distribution than $B_1^{H'}$ since B_1^H and $B_1^{H'}$ are both standard normal variables. Thus, the constraint that both samples have the same one year marginal distribution (see the introduction of Section 3.2) is satisfied[6]. We start with a comparison of fBm paths having a Hurst parameter $H = 0.1$ with fBm paths having a Hurst parameter $H = 0.2$ using the lead-lag transformation. In Figure 3, we plot the statistical power as a function of the truncation order R for different values of the first sample size m (we recall that the size of the second sample is fixed to 1000). We observe that even with small sample sizes, we already obtain a power close to 1 at order 2. Note that the power does not increase with the order but decreases at odd orders when the sample size is smaller than 50. This can be explained by the fact that the odd-order terms of the signature of the lead-lagged fBm are linear combinations of monomials in $B_{t_1}^H, \ldots, B_{t_N}^H$ that are of odd degree. Since (B_t^H) is a centered Gaussian process, the expectation of these terms are zero no matter the value of H. As a consequence, the contribution of odd-order terms of the

[6]Note that a variance rescaling should be performed if one considers a horizon that is different from 1 year.

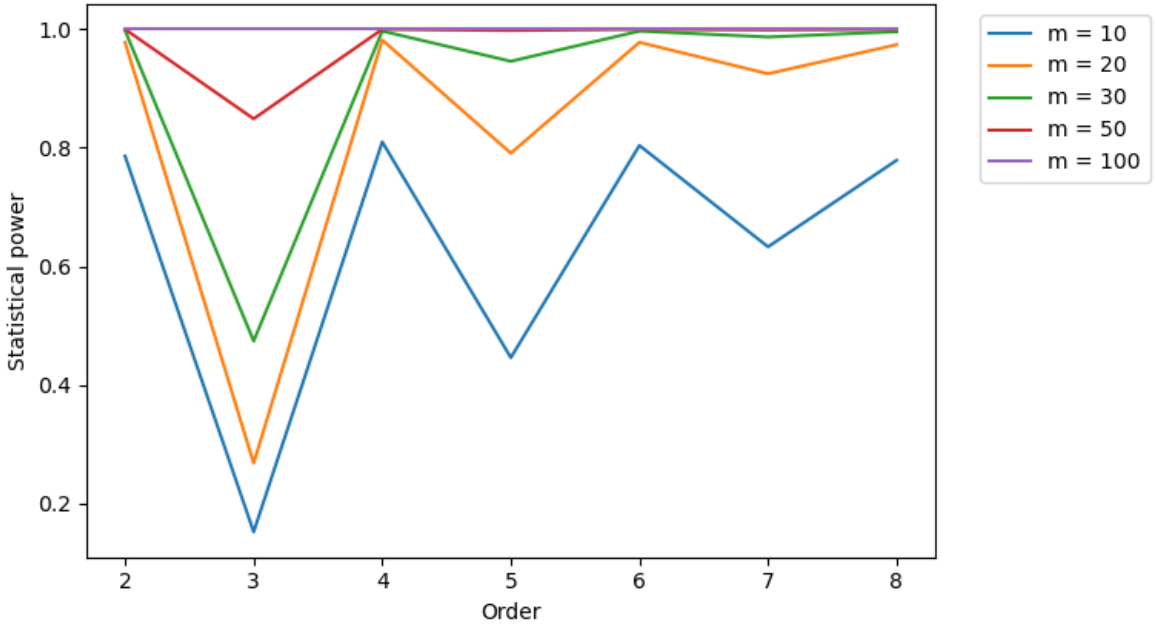


Figure 3: Statistical power of the signature-based validation test ($H = 0.1$ versus $H' = 0.2$) as a function of the truncation order

signature to the MMD is the same under H_0 and under H_1. This is formalized in Proposition 7 of Appendix F. Moreover, if we compute $MMD_{m,n}$ by keeping only one specific order of the signature and we estimate the power of the associated test, we obtain Table 1 which is consistent with the above explanation.

If we conduct the same experiment for $H = 0.1$ versus $H' = 0.5$ (corresponding to the standard Brownian motion), we obtain cumulated powers greater than 99% for all tested orders and sample sizes (even $m = 10$) even if the power of the odd orders is small (below 45%). This is very promising as it shows that the signature-based validation allows to distinguish very accurately rough fBm paths (with a Hurst parameter in the range of those estimated by Gatheral et al [13]) from standard Brownian motion paths even with small sample sizes.

m \Order	2	3	4	5	6	7	8
10	79.4%	9.9%	81.5%	15.3%	80.4%	23.0%	78.1%
20	97.7%	9.8%	98.1%	18.7%	98.0%	28.8%	97.5%
30	99.7%	10.7%	99.7%	20.8%	99.7%	29.8%	99.6%
50	100.0%	9.7%	100.0%	20.7%	100.0%	33.4%	100.0%
100	100.0%	8.4%	100.0%	20.6%	100.0%	34.7%	100.0%
150	100.0%	9.6%	100.0%	21.8%	100.0%	33.4%	100.0%

Table 1: Statistical power of the signature-based validation test ($H = 0.1$ versus $H' = 0.2$) restricted to one order.

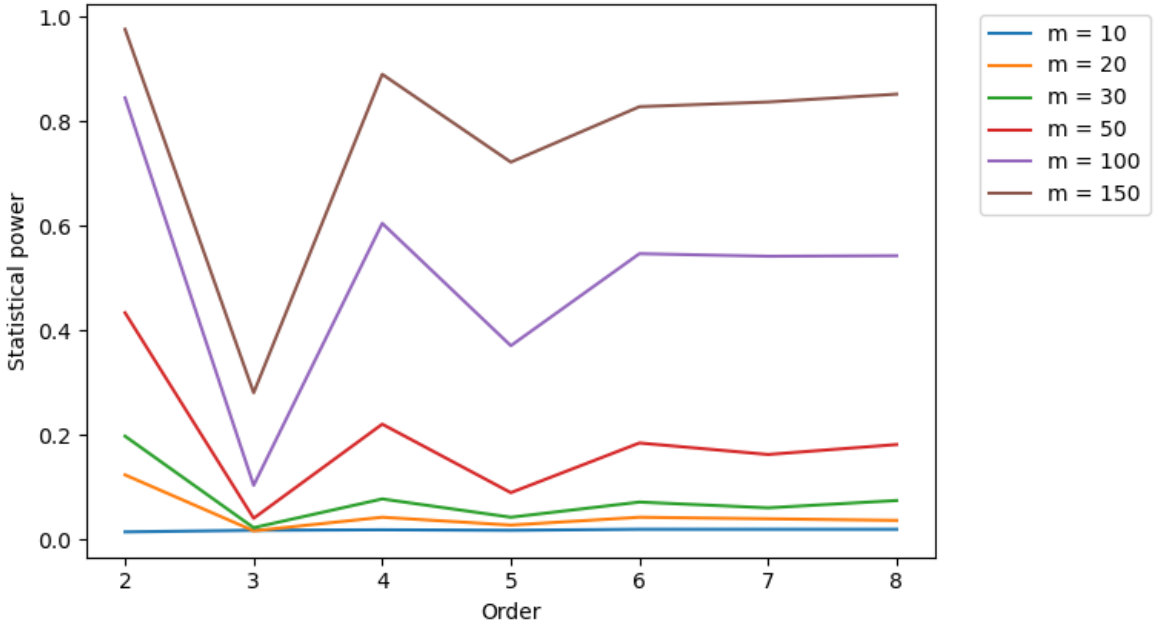


Figure 4: Statistical power of the normalized signature-based validation test ($H = 0.1$ versus $H' = 0.2$) as a function of the truncation order

Note that in these numerical experiments, we have not used any tensor normalization while it is a key ingredient in Theorem 3. This is motivated by the fact that the power is much worse when we use Chevyrev and Oberhauser's normalization (described in Appendix E.1) as one can see on Figure 4. These lower powers can be understood as a consequence of the fact that the normalization is specific to each path. So the normalization can bring the distribution of $S(\hat{B}^{H'})$ closer to the one of $S(\hat{B}^{H})$ than without normalization so that it is harder to distinguish them at fixed sample size. Moreover, if the normalization constant λ is smaller than 1 (which we observe numerically), the high-order terms of the signature become close to zero and their contribution to the MMD is not material. For $H = 0.1$ versus $H' = 0.5$, we observed that the powers remain very close to 100%.

Also note that the lead-lag transformation is key for this model as replacing it by the time transformation results in much lower statistical powers, see Figure 5. This observation is consistent with the previous study from Fermanian [10] which concluded that the lead-lag transformation is the best choice in a learning context.

Before moving to the Black-Scholes dynamics, we present results of the test when the signature is replaced by the log-signature. The log-signature is a more parsimonious - though equivalent - representation of paths than the signature as it contains more zeros. A formal definition of the log-signature and more insights can be found in Section 4. Although no information is lost by the log-signature, it is not clear whether the MMD is still a metric when the signature is replaced by the log-signature in the kernel (see Remark 7 in Section 4). Numerically, the log-signature shows satisfying powers for $H = 0.1$ versus $H' = 0.2$, especially when the truncation order is 2 (see Figure 6). One can remark in particular that the power decreases with the

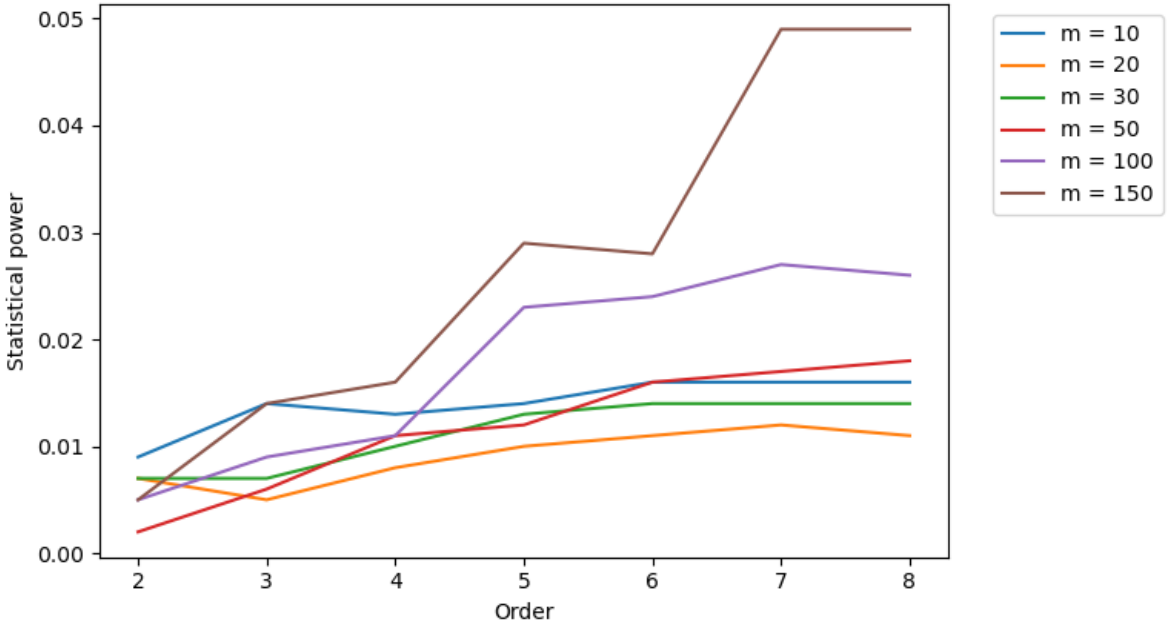


Figure 5: Statistical power of the signature-based validation test ($H = 0.1$ versus $H' = 0.2$) as a function of the truncation order when using the time transformation

order. This observation likely results from the $1/n$ factor appearing in the log-signature formula (equation (59) in Section 4) which makes high-order terms of the signature small so that the even-order terms no longer compensate the odd-order terms. Note that in terms of CPU time, the log-signature is approximately 2.5 times faster to compute than the signature (0.35 ms to compute the log-signature of one path and 0.9 ms for the signature[7]) because of the fact that they are less coefficients to compute in the log-signature than in the signature. Therefore, the log-signature could be preferred to the signature if the CPU time is a constraint in practical applications.

3.2.2. The Black-Scholes dynamics

In the well-known Black-Scholes model [2], the evolution of the stock price $(S_t)_{t\geq 0}$ is modelled using the following dynamics:

$$dS_t = \mu S_t dt + \sigma_t S_t dW_t, \quad S_0 = s_0 \tag{34}$$

where $\mu \in \mathbb{R}$, σ is a deterministic function of the time and $(W_t)_{t\geq 0}$ is a standard Brownian motion. Because of its simplicity, this model is still widespread in the insurance industry, in particular for the modelling of equity and real estate indices.

As for the fractional Brownian motion, we want to compare two parametrizations of this model that share the same one year marginal distribution. For this purpose, we consider a Black-Scholes dynamics (BSd) with drift μ and constant volatility σ and a BSd with the same

[7]The numerical experiments have been conducted on a standard laptop with a 1.8 GHz processor.

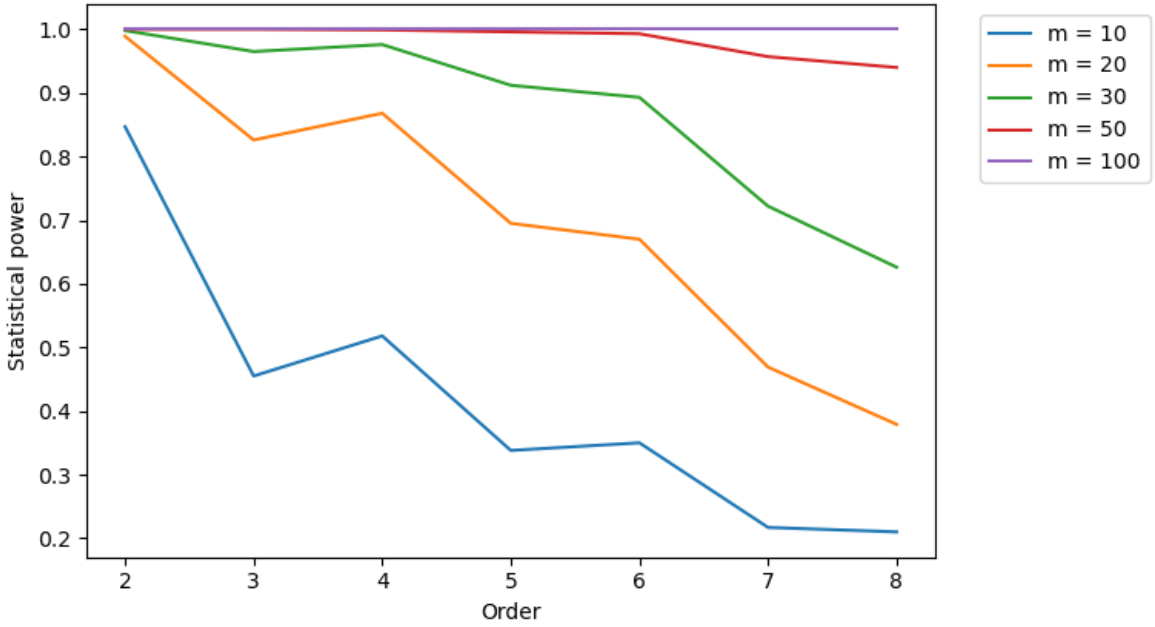


Figure 6: Statistical power of the log-signature-based validation test ($H = 0.1$ versus $H' = 0.2$) as a function of the truncation order

drift μ but with a deterministic volatility $\gamma(t)$ satisfying $\int_0^1 \gamma^2(s)ds = \sigma^2$ which guarantees that the one-year marginal distribution constraint is met. For the sake of simplicity, we take a piecewise constant volatility with $\gamma(t) = \gamma_1$ if $t \in [0, 1/2)$ and $\gamma(t) = \gamma_2$ if $t \in [1/2, 1]$. In this setting, the objective of the test is no longer to distinguish two stochastic processes with different regularity but two stochastic processes with different volatility which is *a priori* more difficult since the volatility is not directly observable in practice. When $\mu = 0$, the first two terms of the signature of the lead-lag transformation have the same distribution in the two parametrizations as the time step converges to 0. This is explained in Example 4 in Section 4.1. We conjecture that this result extends to the full signature so that the two models cannot be distinguished using the signature.

We start by comparing BSd paths with $\mu = 0.05$ and $\sigma = 0.2$ and BSd paths with $\mu = 0.05$, $\gamma_1 = 0$ and $\gamma_2 = \sqrt{2}\sigma$ using the lead-lag transformation. We consider a zero volatility on half of the time interval in order to obtain very different paths in the two samples. Despite this extreme parametrization, we obtain very low powers as shown in Figure 7, even for larger sample sizes. It seems that the constraint of same quadratic variation in both samples makes the signatures from sample 1 too close from those of sample 2. In order to improve the power of the test, we can consider another data transformation which allows to capture information about the initial one-dimensional path in a different manner. We observed that the time lead-lag transformation (see section 3.1.2) allowed to better distinguish the signatures from the two parametrizations given above. However, the order of magnitude of the differentiating coefficients of the signature was significantly smaller from non-differentiating coefficients so that the former were hidden by the latter when computing $MMD_{m,n}$. To address this issue, we applied a normalization to all coefficients of the signature to make sure they are all of the same order of magnitude. Concretely, given two samples $\mathcal{S}_1 = \{X_1, \dots, X_m\}$ and $\mathcal{S}_2 = \{Y_1, \dots, Y_n\}$ of d-dimensional

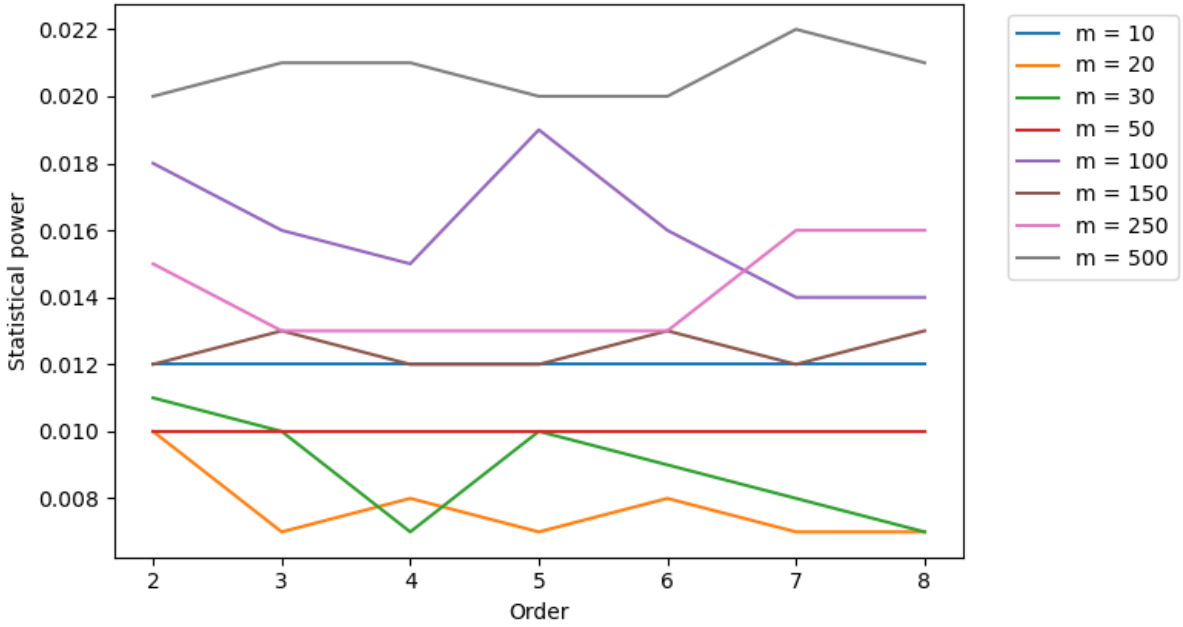


Figure 7: Statistical power of the signature-based validation test (constant volatility BSd versus piecewise constant volatility BSd) as a function of the truncation order.

paths, the normalization is performed as follows:

1. for all $i \in \{1, \dots, m\}$ and for all $j \in \{1, \dots, n\}$, compute $S(X_i)$ and $S(Y_j)$
2. for all $\ell \in \{1, \dots, R\}$ and for all $I = (i_1, \dots, i_\ell) \in \{1, \dots, d\}^\ell$, compute

$$M_I^\ell = \max\left(\max_{i=1,\dots,m} \left| \mathbf{X}_{i,I}^\ell \right|, \max_{j=1,\dots,n} \left| \mathbf{Y}_{j,I}^\ell \right| \right) \tag{35}$$

   where $\mathbf{X}_{i,I}^\ell$ (resp. $\mathbf{Y}_{j,I}^\ell$) is the coefficient at position I of the ℓ-th term of the signature of X_i (resp. Y_j).

3. for all $i \in \{1, \dots, m\}$, for all $j \in \{1, \dots, n\}$, for all $\ell \in \{1, \dots, R\}$ and for all $I = (i_1, \dots, i_\ell) \in \{1, \dots, d\}^\ell$, compute the normalized signature as:

$$\hat{\mathbf{X}}_{i,I}^\ell = \frac{\mathbf{X}_{i,I}^\ell}{M_I^\ell} \quad \text{and} \quad \hat{\mathbf{Y}}_{j,I}^\ell = \frac{\mathbf{Y}_{j,I}^\ell}{M_I^\ell}. \tag{36}$$

This procedure guarantees that all coefficients of the signature lie within $[-1, 1]$. Using this normalization for the time lead-lag normalization, the power of the test is significantly better than with the plain lead-lag transformation, as shown in Figure 8. In Figure 9, we show that the power can be further improved by considering the log-signature instead of the signature.

We also considered a slight variation of the BSd that has autocorrelation. Let $(t_i)_{0 \le i \le N}$ be an equally-spaced partition of $[0, 1]$ with $\Delta t = t_{i+1} - t_i = 1/N$. The autocorrelated discretized

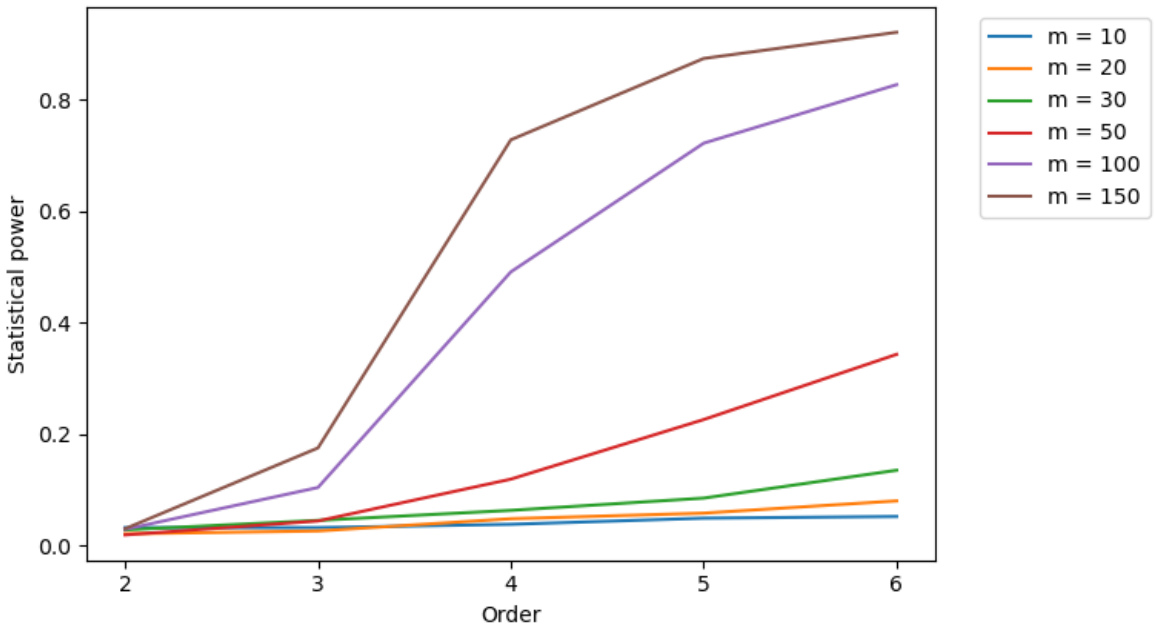


Figure 8: Statistical power of the signature-based validation test (constant volatility BSd versus piecewise constant volatility BSd) as a function of the truncation order when using the time lead-lag transformation with the normalization procedure.

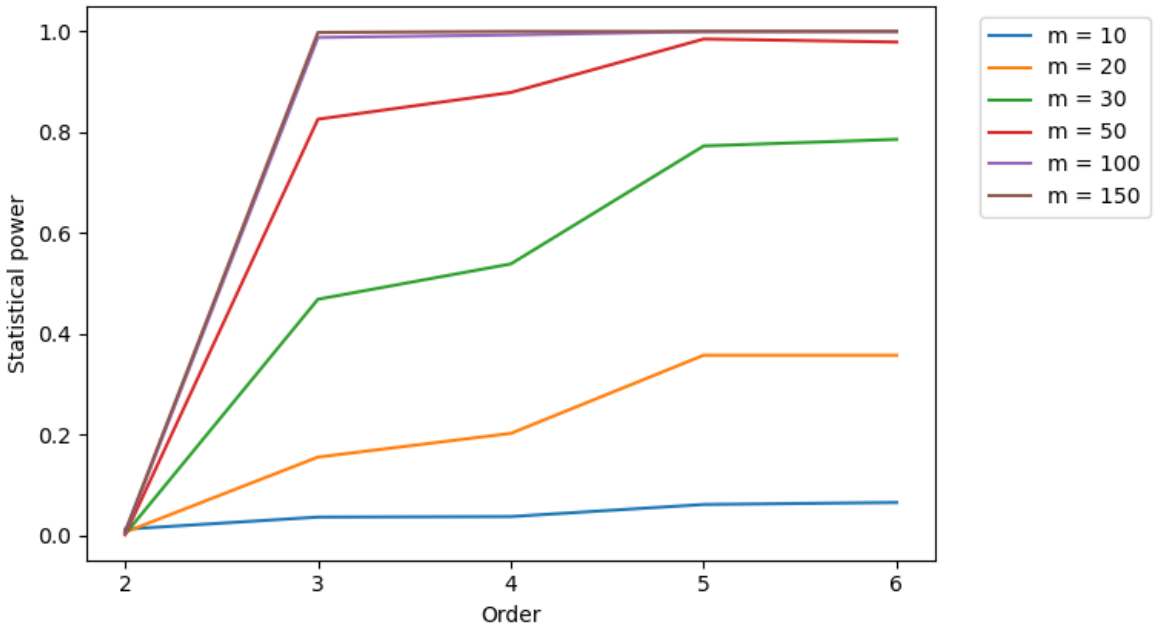


Figure 9: Statistical power of the log-signature-based validation test (constant volatility BSd versus piecewise constant volatility BSd) as a function of the truncation order when using the time lead-lag transformation with the normalization procedure.

BSd $(S^C_{t_i})_{0\leq i\leq N}$ is defined as follows:

$$\left\{\begin{array}{rcl} S^C_{t_{i+1}} & = & S^C_{t_i}\exp\left((\mu^C-\frac{\gamma^2_{t_i}}{2})\Delta t+\gamma_{t_i}\sqrt{\Delta t}G_{i+1}\right), \quad i\in\{0,\dots,N-1\} \\ S^C_{t_0} & = & s_0 \end{array}\right. \tag{37}$$

where $(G_i)_{1\leq i\leq N}$ is a sequence of standard normal random variables satisfying:

$$Cov(G_i,G_j)=\left\{\begin{array}{ll} 1 & \text{if } i=j \\ \rho & \text{if } i+1=j \\ 0 & \text{otherwise.}\end{array}\right. \tag{38}$$

The covariance matrix of $(G_1,\dots,G_N)$ is positive definite when $\rho\in\left(-\frac{1}{2\cos\left(\frac{\pi}{N+1}\right)},\frac{1}{2\cos\left(\frac{\pi}{N+1}\right)}\right)$. Indeed, the covariance matrix is a tridiagonal Toepliz matrix so its eigenvalues are given by (page 59 in [27]):

$$\lambda_k=1+2\rho\cos\left(\frac{k\pi}{N+1}\right), \quad k=1,\dots,N. \tag{39}$$

In our framework, we have $N=12$ and we can check that $[-0.5,0.5]\subset\left(-\frac{1}{2\cos\left(\frac{\pi}{N+1}\right)},\frac{1}{2\cos\left(\frac{\pi}{N+1}\right)}\right)$. In Figure 10, we compare BSd paths with $\mu=0.05$ and $\sigma=0.2$ and autocorrelated BSd paths with correlation $\rho\in\{-0.5,-0.4,\dots,0.5\}$ and a piecewise constant volatility as in the previous setting but with $\gamma_1=\sigma/\sqrt{2}$ and (γ_2,μ^C) chosen such that S^C_1 has the same distribution as $s_0e^{\mu-\sigma^2/2+\sigma W_1}$. Here, the first sample size is fixed to $m=30$. While it was not possible to distinguish BSd paths with different volatility functions using the lead-lag transformation, we observe that the introduction of autocorrelation makes it again possible even with a small sample size. More precisely, except if $\rho\in\{-0.1,0,0.1,0.2\}$, we obtain a power greater than 90% at order 2. We note however a decrease of the power with the truncation order as it appears that apart from the term of order 2, all the other terms of the signature are very close between the two samples.

4. Examples and properties of the signature

The signature being already defined in Section 2.2, the purpose of this section is to provide more insights on the signature thanks to examples and a brief overview of its main properties.

4.1. Some examples

First, we present several examples that allow to better understand the signature and the log-signature.

Example 2. *If $X:[0,T]\to E$ is a linear path , i.e. $X_t=X_0+(X_T-X_0)\frac{t}{T}$, then for any $n\geq 0$:*

$$\mathbf{X}^n=\frac{1}{n!}(X_T-X_0)^{\otimes n}. \tag{40}$$

Example 3. *If E is a vector space of dimension 2, the second order term of the signature is given by:*

$$\mathbf{X}^2=\int_0^T\int_0^t dX_s\otimes dX_t=\begin{pmatrix}\int_0^T\int_0^t dX^{(1)}_s dX^{(1)}_t & \int_0^T\int_0^t dX^{(1)}_s dX^{(2)}_t \\ \int_0^T\int_0^t dX^{(2)}_s dX^{(1)}_t & \int_0^T\int_0^t dX^{(2)}_s dX^{(2)}_t\end{pmatrix}. \tag{41}$$

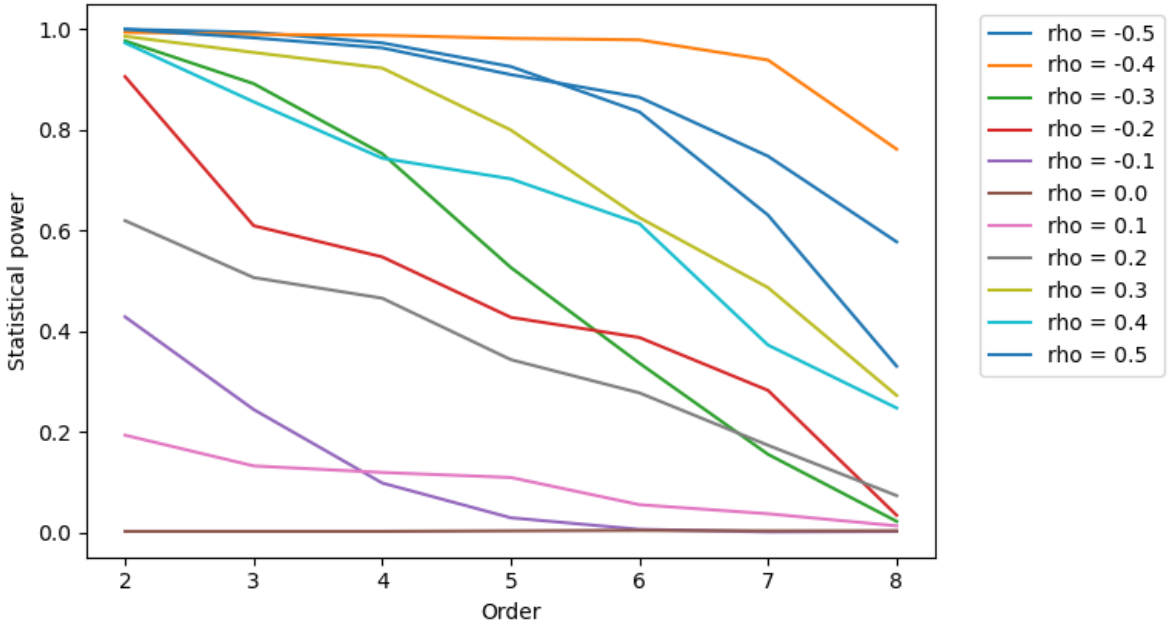


Figure 10: Statistical power of the log-signature-based validation test (constant volatility BSd versus autocorrelated BSd) as a function of the truncation order when using the lead-lag transformation with the normalization procedure.

Note that the difference of the anti-diagonal coefficients of $\mathbf{X}^2$ *corresponds, up to a factor* $1/2$*, to the Lévy area of the curve* $t \mapsto (X_t^1, X_t^2)$ *which is defined as:*

$$\mathcal{A}^{Levy} = \frac{1}{2}\left(\int_0^T (X_t^1 - X_0^1)dX_t^2 - \int_0^T (X_t^2 - X_0^2)dX_t^1\right). \tag{42}$$

It is the signed area between the curve and the chord connecting the two endpoints (see Figure 11).

In Section 3.1.2, we mentioned that the lead-lag transformation allows to capture the quadratic variation of a path in the signature. More precisely, the Levy area of the lead-lag transformation is the quadratic variation up to a factor $1/2$ as stated by the following proposition.

Proposition 1. *Let* $t_0 = 0 < t_1 < \cdots < t_N = T$ *be a partition of* $[0, T]$ *and* $(X_{t_i})_{i=0,\dots,N}$ *be a finite set of observations of a real-valued process* X *on this partition. The Levy area of the lead-lag transformation of* $(X_{t_i})_{i=0,\dots,N}$ *is equal to the quadratic variation of* X *on the partition* $(t_i)_{i=0,\dots,N}$ *up to a factor* $1/2$*, i.e.*

$$\frac{1}{2}\left(\int_0^T (X_t^{lead} - X_0^{lead})dX_t^{lag} - \int_0^T (X_t^{lag} - X_0^{lag})dX_t^{lead}\right) = \frac{1}{2}\sum_{i=0}^{N-1}(X_{t_{i+1}} - X_{t_i})^2. \tag{43}$$

Proof. Using the notations introduced above, we have:

$$X_t^{lead} = \begin{cases} X_{t_i} + \frac{X_{t_{i+1}} - X_{t_i}}{t_{i+1/2} - t_i}(t - t_i) & \text{if } t \in [t_i, t_{i+1/2}] \\ X_{t_{i+1}} & \text{if } t \in [t_{i+1/2}, t_{i+1}] \end{cases} \tag{44}$$

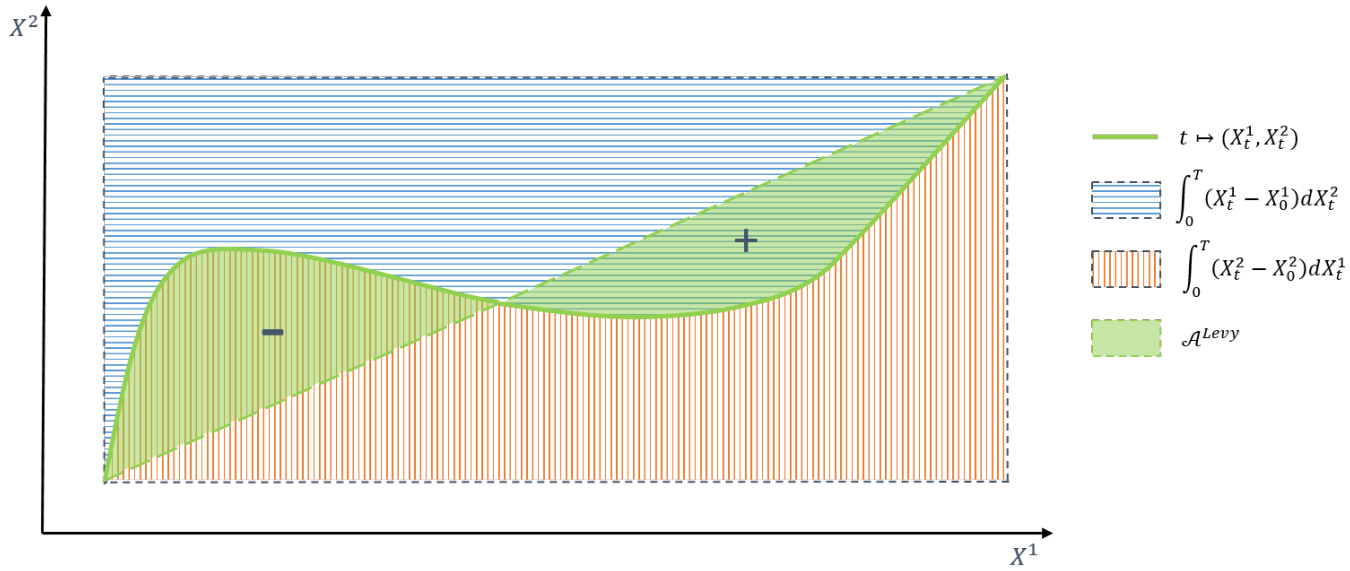


Figure 11: Illustration of the Lévy area. The blue dashed area corresponds to the integral $\int_0^T (X_t^1 - X_0^1) dX_t^2$ while the red dashed area corresponds to the integral $\int_0^T (X_t^2 - X_0^2) dX_t^1$. Taking the difference between these two areas yields the Levy area (represented in green transparent) up to a factor 2 because of a double counting. The '+' (resp. '−') sign indicates that the surrounding area is counted positively (resp. negatively).

and

$$X_t^{lag} = \begin{cases} X_{t_i} & \text{if } t \in [t_i, t_{i+1/2}] \\ X_{t_i} + \frac{X_{t_{i+1}} - X_{t_i}}{t_{i+1} - t_{i+1/2}} (t - t_{i+1/2}) & \text{if } t \in [t_{i+1/2}, t_{i+1}]. \end{cases} \tag{45}$$

We deduce that,

$$\begin{aligned} \int_0^T (X_t^{lead} - X_0^{lead}) dX_t^{lag} &= \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1/2}} (X_t^{lead} - X_0^{lead}) dX_t^{lag} + \int_{t_{i+1/2}}^{t_{i+1}} (X_t^{lead} - X_0^{lead}) dX_t^{lag} \\ &= \sum_{i=0}^{N-1} \int_{t_{i+1/2}}^{t_{i+1}} (X_{t_{i+1}} - X_0) \frac{X_{t_{i+1}} - X_{t_i}}{t_{i+1} - t_{i+1/2}} dt \\ &= \sum_{i=0}^{N-1} (X_{t_{i+1}} - X_0)(X_{t_{i+1}} - X_{t_i}) \\ &= \sum_{i=0}^{N-1} \left[(X_{t_{i+1}} - X_{t_i})^2 + (X_{t_i} - X_0)(X_{t_{i+1}} - X_{t_i}) \right]. \end{aligned}$$

A similar calculation yields

$$\int_0^T (X_t^{lag} - X_0^{lag}) dX_t^{lead} = \sum_{i=0}^{N-1} (X_{t_i} - X_0)(X_{t_{i+1}} - X_{t_i}). \tag{46}$$

Hence,

$$\int_0^T (X_t^{lead} - X_0^{lead}) dX_t^{lag} - \int_0^T (X_t^{lag} - X_0^{lag}) dX_t^{lead} = \sum_{i=0}^{N-1} (X_{t_{i+1}} - X_{t_i})^2. \tag{47}$$

□

We have seen in our numerical experiments in Section 3.2.2 that the lead-lag transformation is not always sufficient to distinguish two close models. In the following example, we show that, as the time step converges to 0, the first two terms of the signature of the lead-lag transformation of a driftless Black-Scholes dynamics with constant volatility have the same distributions as the first two terms of the signature of the lead-lag transformation of a driftless Black-Scholes dynamics with a time-dependent deterministic volatility if the total variances at time T of both models are the same.

Example 4. *Consider X and Y the solutions of the following SDE's*

$$\begin{array}{ll} dX_t &= \sigma X_t dW_t \quad ,X_0 = 1 \\ dY_t &= \gamma(t) Y_t dW_t \quad ,Y_0 = 1 \end{array} \tag{48}$$

where $(W_t)_{t\geq 0}$ is a Brownian motion and γ is a deterministic function satisfying $\int_0^T \gamma(t)^2 dt = \sigma^2 T$. The explicit formulas of X and Y write:

$$\begin{aligned} X_t &= \exp\left(\sigma W_t - \frac{1}{2}\sigma^2 t\right) \\ Y_t &= \exp\left(\int_0^t \gamma(s) dW_s - \frac{1}{2}\int_0^t \gamma(s)^2 ds\right). \end{aligned} \tag{49}$$

Let us denote by $\hat{X}_N$ (resp. $\hat{Y}_N$) the lead-lag transformation of X (resp. Y) on a partition $(t_i)_{i=0,\dots,N}$ of $[0,T]$ such that $t_i = iT/N$. The constraint $\int_0^T \gamma(t)^2 dt = \sigma^2 T$ implies that $X_T \overset{d}{=} Y_T$ so the first order terms of the signatures of $\hat{X}_N$ and $\hat{Y}_N$ (which reduce to the increments of X and Y over $[0,T]$) have the same distribution for all $N \geq 1$. The second order term of the signature of $\hat{Y}_N$ is given by (see the proof of the above proposition):

$$\hat{\mathbf{Y}}_N^2 = \begin{pmatrix} \frac{1}{2}(Y_T - Y_0)^2 & \sum_{i=0}^{N-1}\left[(Y_{t_{i+1}} - Y_{t_i})^2 + (Y_{t_i} - Y_0)(Y_{t_{i+1}} - Y_{t_i})\right] \\ \sum_{i=0}^{N-1}(Y_{t_i} - Y_0)(Y_{t_{i+1}} - Y_{t_i}) & \frac{1}{2}(Y_T - Y_0)^2 \end{pmatrix} \tag{50}$$

Now, given that Y is a square-integrable continuous martingale, the coefficient at position $(1,2)$ of $\hat{\mathbf{Y}}_N$ converges in probability as $N \to +\infty$ to:

$$\langle Y\rangle_T + \int_0^T (Y_t - Y_0) dY_t = \frac{1}{2}\left[(Y_T - Y_0)^2 + \langle Y\rangle_T\right] \tag{51}$$

where $\langle Y\rangle$ denotes the quadratic variation process of Y and the equality is obtained using the integration by parts formula. Similarly, the coefficient at position $(2,1)$ of $\hat{\mathbf{Y}}_N$ converges in probability as $N \to +\infty$ to:

$$\int_0^T (Y_t - Y_0) dY_t = \frac{1}{2}\left[(Y_T - Y_0)^2 - \langle Y\rangle_T\right]. \tag{52}$$

The same convergences hold for $\hat{\mathbf{X}}_N$. Now remark that the processes $\left(\int_0^t \gamma(s) dW_s\right)_{t\geq 0}$ and $\left(W_{\int_0^t \gamma(s)^2 ds}\right)_{t\geq 0}$ are both Gaussian processes with the same mean and the same covariance function, we deduce that they have the same distribution. Analogously, $(\sigma W_t)_{t\geq 0}$ has the same distribution as $(W_{\sigma^2 t})$. We deduce that:

$$\begin{aligned} (X_t)_{t\geq 0} &\overset{d}{=} \left(\exp\left(W_{\sigma^2 t} - \frac{1}{2}\sigma^2 t\right)\right)_{t\geq 0} \\ (Y_t)_{t\geq 0} &\overset{d}{=} \left(\exp\left(W_{\int_0^t \gamma(s)^2 ds} - \frac{1}{2}\int_0^t \gamma(s)^2 ds\right)\right)_{t\geq 0}. \end{aligned} \tag{53}$$

Setting $\varphi(t) = \frac{1}{\sigma^2}\int_0^t \gamma(s)^2 ds$, *we deduce that* $(Y_t)_{0\leq t\leq T} \overset{d}{=} (X_{\varphi(t)})_{0\leq t\leq T}$. *As a consequence,* $(Y_t, \langle Y\rangle_t)_{0\leq t\leq T}$ *has the same distribution as* $(X_{\varphi(t)}, \langle X\rangle_{\varphi(t)} - \langle X\rangle_{\varphi(0)})_{0\leq t\leq T}$. *Since* $\varphi(0) = 0$ *and* $\varphi(T) = T$, *we conclude that the limit of* $\hat{\mathbf{Y}}_N$ *has the same distribution as the limit of* $\hat{\mathbf{X}}_N$.

Let us now introduce more formally the log-signature. We recall that the space of formal series of tensors is defined as:

$$T(E) = \left\{(\mathbf{t}^n)_{n\geq 0} \mid \forall n \geq 0, \mathbf{t}^n \in E^{\otimes n}\right\} \tag{54}$$

with the convention $E^{\otimes 0} = \mathbb{R}$. This space can be equipped with the following operations: for $\mathbf{t}$, $\mathbf{u} \in T(E)$, $\lambda \in \mathbb{R}$,

$$\begin{array}{rcl} \mathbf{t} + \mathbf{u} & = & (\mathbf{t}^n + \mathbf{u}^n)_{n\geq 0} \\ \lambda\mathbf{t} & = & (\lambda\mathbf{t}^n)_{n\geq 0} \\ \mathbf{t} \otimes \mathbf{u} & = & \left(\mathbf{v}^n = \sum_{k=0}^n \mathbf{t}^k \otimes \mathbf{u}^{n-k}\right)_{n\geq 0}. \end{array} \tag{55}$$

Since by convention the term of order 0 of the signature is set to 1, the signature takes its values in the following affine subspace of $T(E)$:

$$T_1(E) = \left\{\mathbf{t} \in T(E) \mid \mathbf{t}^0 = 1\right\}. \tag{56}$$

A closely related subspace of $T(E)$ is the following:

$$T_0(E) = \left\{\mathbf{t} \in T(E) \mid \mathbf{t}^0 = 0\right\}. \tag{57}$$

In fact, there is a bijection between $T_1(E)$ and $T_0(E)$ (Lemma 2.21 in [24]):

Proposition 2. *Let us define the exponential mapping as:*

$$\begin{array}{rcl} T_0(E) & \to & T_1(E) \\ \mathbf{t} & \mapsto & \exp(\mathbf{t}) := \displaystyle\sum_{n\geq 0} \frac{\mathbf{t}^{\otimes n}}{n!} \end{array} \tag{58}$$

with the convention $t^{\otimes 0} = 1$ *and the logarithm mapping as:*

$$\begin{array}{rcl} T_1(E) & \to & T_0(E) \\ \mathbf{t} & \mapsto & \log(\mathbf{t}) := \displaystyle\sum_{n\geq 1} \frac{(-1)^{n-1}}{n} (\mathbf{t} - \mathbf{1})^{\otimes n} \end{array} \tag{59}$$

where $\mathbf{1} = (1, 0, \ldots, 0, \ldots) \in T_1(E)$. *The exponential mapping is bijective from* $T_0(E)$ *to* $T_1(E)$ *and its inverse is the logarithm mapping.*

Example 3 (continued)**.** *Using the exponential and the logarithm mappings, we can rewrite the signature in Example 2 in the following way:*

$$S(X) = \exp(X_T - X_0) \tag{60}$$

where $X_T - X_0$ *should be interpreted as the element* $(0, X_T - X_0, 0, \ldots, 0, \ldots)$ *of* $T_0(E)$. *Moreover,*

$$\log(S(X)) = X_T - X_0. \tag{61}$$

Using the logarithm, it is therefore possible to define the log-signature of a path X as $\log(S(X))$. Although there is a one-to-one correspondence between the signature and the log-signature, the log-signature is a more parsimonious representation of the path than the signature in the sense that it removes the redundancies. This can be seen in Example 3: the only non-zero term of the log-signature of a linear path is the term of order 1 which contains the increments of the path. In comparison to the signature, all the powers of the increments have disappeared. However, no information is lost. More generally, it can be shown (see for example [22]) that the log-signature has more zeros than the signature. As such, it represents a useful object for applications as it allows to avoid the exponential increase of the size of the truncated signature with the order. Indeed, if E is a vector space of dimension d, the term of order n of the signature has d^n elements.

Example 5. *Let us consider $X \in \mathcal{C}^1([0,T],\mathbb{R}^2)$. Denoting by $\mathbf{lX}^n$ the term of order n of the log-signature, the second order term of the log-signature writes:*

$$\mathbf{lX}^2 = \mathbf{X}^2 - \frac{1}{2}\mathbf{X}^1 \otimes \mathbf{X}^1 \tag{62}$$

where $\mathbf{X}^2$ comes from the first term ($n=1$) of the log series in equation (59) and $\mathbf{X}^1 \otimes \mathbf{X}^1$ comes from the second term ($n=2$). We have:

$$\mathbf{X}^2 = \begin{pmatrix} \frac{(X_T^1 - X_0^1)^2}{2} & \int_0^T (X_t^1 - X_0^1) dX_t^2 \\ \int_0^T (X_t^2 - X_0^2) dX_t^1 & \frac{(X_T^2 - X_0^2)^2}{2} \end{pmatrix} \tag{63}$$

and

$$\mathbf{X}^1 \otimes \mathbf{X}^1 = \begin{pmatrix} (X_T^1 - X_0^1)^2 & (X_T^1 - X_0^1)(X_T^2 - X_0^2) \\ (X_T^1 - X_0^1)(X_T^2 - X_0^2) & (X_T^2 - X_0^2)^2 \end{pmatrix}. \tag{64}$$

Using the integration by part formula, we obtain:

$$\mathbf{lX}^2 = \underbrace{\frac{1}{2}\left(\int_0^T (X_t^1 - X_0^1) dX_t^2 - \int_0^T (X_t^2 - X_0^2) dX_t^1 \right)}_{\text{Lévy area of } X} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{65}$$

Hence, the second order term of the log-signature reduces to the Lévy area.

Remark 4. *Note that only the N first terms of the logarithm series (59) contribute to the N-th term of the log-signature. Indeed, for $n > N$, the contributions to $E^{\otimes N}$ of $(\mathbf{t}-\mathbf{1})^{\otimes n}$ always involve some product by $(\mathbf{t}-\mathbf{1})^0 = 0$.*

4.2. Properties

We have seen in the first subsection that the signature allows to capture some information about the path. A natural question at this stage is how much information about X does the signature of X contain. This subsection aims at answering this question.

Proposition 3 (Invariance under time reparametrization)**.** *Let $X \in \mathcal{C}^1([0,T],E)$ and consider $\varphi : [0,T] \to [0,T]$ a non-decreasing surjection. If we set $\tilde{X}_t = X_{\varphi(t)}$, then:*

$$S(\tilde{X}) = S(X). \tag{66}$$

This first property (see Proposition 7.10 in [12] for a proof) means that the speed at which the path is traversed is not captured by the signature. The signature is also invariant by translation. Indeed, if we define $\bar{X}_t = x + X_t$, then $d\bar{X}_t = dX_t$ and by definition of the signature we have $S(\bar{X}) = S(X)$. The next property we will outline is Chen's identity. Before introducing it, we need the following definition.

Definition 5 (Concatenation)**.** *Let $X \in \mathcal{C}^1([0,t], E)$ and $Y \in \mathcal{C}^1([t,T], E)$. The concatenation of X and Y is the path in $\mathcal{C}^1([0,T], E)$ defined as:*

$$(X * Y)_s = \begin{cases} X_s & \text{if } s \in [0,t] \\ X_t + Y_s - Y_t & \text{if } s \in [t,T]. \end{cases} \tag{67}$$

Theorem 7 (Chen's identity)**.** *Let $X \in \mathcal{C}^1([0,t], E)$ and $Y \in \mathcal{C}^1([t,T], E)$. Then,*

$$S_{[0,T]}(X * Y) = S_{[0,t]}(X) \otimes S_{[t,T]}(Y). \tag{68}$$

A proof can be found in Theorem 2.9 of [24]. A useful application of Chen's identity is the computation of the signature of a piecewise linear path. Let $(t_i)_{0 \le i \le n}$ be a subdivision of $[0,T]$ and $X : [0,T] \to E$ be a path such that for $t \in [t_i, t_{i+1}]$ with $0 \le i \le n-1$,

$$X_t = X_{t_i} + \frac{X_{t_{i+1}} - X_{t_i}}{t_{i+1} - t_i}(t - t_i). \tag{69}$$

Then by Chen's identity,

$$S_{[0,T]}(X) = \bigotimes_{i=0}^{n-1} S_{[t_i, t_{i+1}]}(X). \tag{70}$$

Given that X is linear on each $[t_i, t_{i+1}]$, then $S_{[t_i,t_{i+1}]}(X) = \exp(X_{t_{i+1}} - X_{t_i})$ and

$$S_{[0,T]}(X) = \bigotimes_{i=0}^{n-1} \exp(X_{t_{i+1}} - X_{t_i}). \tag{71}$$

In general, the right hand side cannot be simplified to $\exp(X_T - X_0)$ because the tensor product $\otimes$ is not commutative. Another consequence of Chen's identity is the following proposition (Proposition 2.14 in [24]).

Proposition 4 (Time-reversal)**.** *Let $X \in \mathcal{C}^1([0,T], E)$. Define $\overleftarrow{X}$ as $\overleftarrow{X}_t = X_{2T-t}$ for $t \in [T, 2T]$. Then,*

$$S_{[0,2T]}(X * \overleftarrow{X}) = S_{[0,T]}(X) \otimes S_{[T,2T]}(\overleftarrow{X}) = \mathbf{1} \tag{72}$$

where we recall that $\mathbf{1} = (1, 0, \dots, 0, \dots) \in T_1(E)$.

Because constant paths also have $\mathbf{1}$ as signature, the above proposition implies that $X * \overleftarrow{X}$ has the same signature as constant paths.

Due to the invariance by reparametrisation and by translation and the time-reversal property, it is clear that if two paths have the same signature, then they are not necessarily equal. In other words, the signature mapping is not injective. Fortunately, the presented invariances and the time-reversal property are essentially the only cases when paths can differ but have the same signature. To make this precise, we need the notion of tree-like paths.

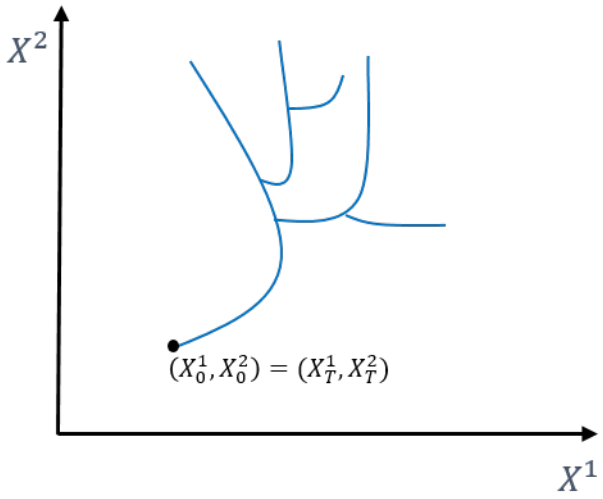


Figure 12: Example of a tree like path.

Definition 6 (Tree-like path)**.** *A path $X : [0,T] \to E$ is tree-like if there exists a continuous function $h : [0,T] \to [0,+\infty[$ such that $h(0) = h(T) = 0$ and for all $s,t \in [0,T]$ with $s \leq t$:*

$$\|X_t - X_s\|_E \leq h(s) + h(t) - 2 \inf_{u \in [s,t]} h(u). \tag{73}$$

This function is called a height function for the path X.

Remark 5. *Note that a tree-like path necessarily satisfies $X_0 = X_T$. Indeed, by Definition 6:*

$$\|X_T - X_0\|_E \leq h(0) + h(T) - 2 \inf_{u \in [0,T]} h(u) = 0 \tag{74}$$

because $h(0) = h(T) = 0$ and h is non-negative. Therefore, one way to turn a tree-like path into a path that is not tree-like is to consider the path $t \mapsto (t, X_t)$ obtained as the time transformation of X.

As suggested by their name, tree-like paths are paths whose graph looks like a tree (see Figure 12), i.e. an acyclic and connected graph in graph theory and the height function h corresponds to the depth of each node of the tree in a depth-first search. Another equivalent way to see tree-like paths is to see them as paths that can be reduced to a constant path by removing pieces of the form $W * \overleftarrow{W}$. For example, if X and Y are non-constant paths, $X * Y * \overleftarrow{Y} * \overleftarrow{X}$ is an example of tree-like path. This notion of tree-like paths is crucial to understand the information that is not captured by the signature as Hambly and Lyons [19] showed that the signature determines the path up to tree-like equivalence, which we will now define.

Definition 7 (Tree-like equivalence)**.** *For X and Y two paths, we say that X and Y are tree-like equivalent if $X * \overleftarrow{Y}$ is a tree-like path. This relation is denoted by $X \sim_t Y$.*

We can now state Hambly and Lyons's theorem.

Theorem 8. *Let $X \in \mathcal{C}^1([0,T],E)$. Then $S(X) = \mathbf{1}$ if and only if X is a tree-like path. Moreover, if $Y \in \mathcal{C}^1([0,T],E)$ is another bounded variation path, then $S(X) = S(Y)$ if and only if $X \sim_t Y$.*

This theorem can be understood as follows: two paths will have the same signature if and only if one can be obtained from the second by using translations, by changing the traversal speeds and by removing parts of the form $W * \overleftarrow{W}$. This uniqueness result has then been extended to a more general class of paths (namely weakly geometric rough paths) by Boedihardjo et al [3].

Remark 6. *The conclusion of Theorem 8 still holds if the signature is replaced by the log-signature since the log mapping is a bijection. Note however that the first statement of the theorem should be modified as follows:* $\log(S(X)) = \mathbf{0}$ *if and only if* X *is a tree-like path where* $\mathbf{0} = (0, \dots, 0, \dots) \in T_0(E)$.

We have seen that in dimension 1, the signature only captures the path increment between 0 and T (see Example 1 in Section 2.2) so that the signature will only allow to distinguish paths X and Y such that $X_T - X_0 = Y_T - Y_0$. This result is actually a consequence of the following proposition and Theorem 8.

Proposition 5. *If* E *is a one-dimensional real vector space and* X, Y *are* E*-valued paths such that* $X_T - X_0 = Y_T - Y_0$*, then* X *and* Y *are tree-like equivalent.*

Proof. Since any one-dimensional real vector space is isometrically isomorph to $\mathbb{R}$, we can assume that $E = \mathbb{R}$. Let X and Y be two paths from $[0, T]$ to $\mathbb{R}$ such that $X_T - X_0 = Y_T - Y_0$. Let us set $Z = X * \overleftarrow{Y}$ and $h(t) = |Z_t - Z_0|$ for $t \in [0, 2T]$. Using the definition of concatenation operator and the fact that $X_T - X_0 = Y_T - Y_0$, we have $Z_0 = X_0$ and $Z_{2T} = X_T + Y_0 - Y_T = X_0$ so that $h(0) = h(2T) = 0$. The non-negativity of h results from the non-negativity of the absolute value. Moreover, the continuity of X and Y imply the continuity of Z by definition of the concatenation operator, so h is continuous as well. The only remaining property to show is inequality (73). Let $s, t \in [0, 2T]$ with $s \leq t$. Let us assume that $Z_s \leq Z_t$ (the proof in the case $Z_t \leq Z_s$ is similar) so that $|Z_t - Z_s| = Z_t - Z_s = Z_t - Z_0 - (Z_s - Z_0)$. We distinguish three cases:

- If $Z_0 \leq Z_s \leq Z_t$, then $h(t) = Z_t - Z_0$ and $h(s) = Z_s - Z_0$. Thus,

$$|Z_t - Z_s| = h(t) - h(s) \leq h(t) - \inf_{u \in [s,t]} h(u) \leq h(t) + h(s) - 2 \inf_{u \in [s,t]} h(u). \tag{75}$$

- If $Z_s \leq Z_0 \leq Z_t$, then $h(t) = Z_t - Z_0$ and $h(s) = Z_0 - Z_s$. Thus,

$$|Z_t - Z_s| = h(t) + h(s) = h(t) + h(s) - 2 \inf_{u \in [s,t]} h(u) \tag{76}$$

  because by the intermediate value theorem, there exists $v \in [s, t]$ such that $Z_v = Z_0$ which implies $\inf_{u \in [s,t]} h(u) = 0$.

- If $Z_s \leq Z_t \leq Z_0$, then $h(t) = Z_0 - Z_t$ and $h(s) = Z_0 - Z_s$. Thus,

$$|Z_t - Z_s| = h(s) - h(t) \leq h(s) - \inf_{u \in [s,t]} h(u) \leq h(t) + h(s) - 2 \inf_{u \in [s,t]} h(u). \tag{77}$$

Hence, h is a height function of Z and Z is tree-like. □

4.3. Signature and stochastic processes

In the last two subsections, the signature has been presented in a deterministic setting. However, it is clear that the stated results in the previous subsection remain true for stochastic processes by defining the signature as a random variable. In view of the uniqueness theorem from Hambly and Lyons, a natural question at this stage is whether the signature allows to characterize the law of stochastic processes. A first positive answer has been provided by Chevyrev and Lyons [5]. They succeeded to construct a characteristic function for the signature of stochastic processes and they proved that it characterizes the law of stochastic processes in the same way as the traditional characteristic function does for random variables. However, this construction is quite abstract and as such is not suitable for applications so far. They also gave some technical conditions under which the expected signature (defined as $\mathbb{E}[S(X)]$ where X is a stochastic process) characterizes the law.

These results have then been extended by Chevyrev and Oberhauser [6]. They showed that by considering a normalization of the signature, the expected normalized signature characterizes the law of stochastic processes under mild regularity assumptions. This result is stronger than the one from Chevyrev and Lyons as it requires less assumptions. We now provide a brief description of their main result. More details can be found in Appendix E.

Let us denote by $T_1^*(E)$ the subset of $T^*(E)$ (see equation 11) defined by:

$$T_1^*(E) := \left\{ \mathbf{t} \in T^*(E) \mid \mathbf{t}^0 = 1 \right\}. \tag{78}$$

We define a tensor normalization as follows:

Definition 8 (Tensor normalization)**.** *A tensor normalization is a continuous injective map of the form*

$$\begin{array}{rcl} \Lambda : T_1^*(E) & \to & \{\mathbf{t} \in T_1^*(E) \mid \|\mathbf{t}\| \leq K\} \\ \mathbf{t} & \mapsto & (\mathbf{t}^0, \lambda(\mathbf{t})\mathbf{t}^1, \lambda(\mathbf{t})^2\mathbf{t}^2, \dots, \lambda(\mathbf{t})^n\mathbf{t}^n, \dots). \end{array} \tag{79}$$

where $K > 0$ is a constant and $\lambda : T_1^*(E) \to (0, +\infty)$ is a positive function.

The existence of such object is stated in Appendix E.1. We can now state a simplified version of Chevyrev and Oberhauser's main theorem:

Theorem 9. *Let $X = (X_t)_{t\in[0,T]}$ and $Y = (Y_t)_{t\in[0,T]}$ be two stochastic processes defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that X and Y are in $\mathcal{P}^1([0,T], E)$ almost surely where $\mathcal{P}^1([0,T], E) = \mathcal{C}^1([0,T], E)/\sim_t$ is the space of bounded variation paths quotiented by the tree-like equivalence relation. Let Λ be a tensor normalization and define the normalized signature as $\Phi = \Lambda \circ S$. Then,*

$$\mathbb{E}[\Phi(X)] = \mathbb{E}[\Phi(Y)] \text{ iff } X \overset{d}{=} Y. \tag{80}$$

Of course, stochastic processes of interest for financial applications do not have bounded variation. However, as in practice we only consider the values of stochastic processes over a finite grid and we interpolate them linearly, the bounded variation assumption is verified.

Remark 7. *The proof of this theorem does not work anymore if we replace the signature by the log-signature. Indeed, one of the key ingredients of the proof is the shuffle product identity (stated below and proved in Theorem 2.15 of [24]) which holds for the signature but not for the*

log-signature. If $X : [0,T] \to E$ with E of dimension d, then:

$$\begin{aligned}&\int_{0\leq u_1<u_2\leq\cdots<u_m\leq T} dX_{u_1}^{i_1}\dots dX_{u_m}^{i_m}\int_{0\leq v_1<v_2\leq\cdots<v_n\leq T} dX_{v_1}^{i_1}\dots dX_{v_n}^{i_n}\\&=\sum_{\sigma\in Shuffles(m,n)}\int_{0\leq w_1<w_2\cdots<w_{m+n}\leq T} dX_{w_1}^{k_{\sigma(1)}}\dots dX_{w_{m+n}}^{k_{\sigma(m+n)}}\end{aligned}\tag{81}$$

for $(i_1,\dots,i_m)\in\{1,\dots,d\}^m$, $(j_1,\dots,j_n)\in\{1,\dots,d\}^n$, $(k_1,\dots,k_{m+n})=(i_1,\dots,i_m,j_1,\dots,j_n)$ and

$$Shuffles(m,n)=\{\sigma\in\mathfrak{S}_{m+n}\mid \sigma(1)<\dots<\sigma(m)\ \text{and}\ \sigma(m+1)<\dots<\sigma(m+n)\}\tag{82}$$

with $\mathfrak{S}_{m+n}$ the permutation group of $\{1,\dots,m+n\}$.

5. Concluding remarks

We propose a new approach for the validation of real-world economic scenario motivated by insurance applications. This approach is based on the statistical test developed by Chevyrev and Oberhauser [6] and relies on the notions of signature and maximum mean distance. This test allows to check whether two samples of stochastic processes paths come from the same distribution. Our contribution is to apply this test to two stochastic processes, namely the fractional Brownian motion and the Black-Scholes dynamics. We analyze its statistical power based on numerical experiments under two constraints: first on the asymetry in terms of size of samples and second on the equality of the marginal one-year distributions between the samples. The numerical experiments have highlighted the need to configure the test specifically for each stochastic process to achieve a good statistical power. In particular the truncation order, the path transformation (lead-lag or time lead-lag), the normalization and the signature type (signature or log-signature) are key parameters to be adjusted for each model. For example, the test allows to distinguish fBm paths with Hurst parameter $H = 0.1$ from fBm paths with Hurst parameter $H' = 0.2$ using the lead-lag transformation and the signature. Besides, the test allows to distinguish BSd paths with constant volatility from BSd paths with piecewise constant volatility using the time lead-lag transformation and the log-signature with a proper normalization. These results indicate that this test represents a promising validation tool for real-world scenarios in a practical framework motivated by insurance applications.

Appendix A. Proof of Theorem 2

As a preliminary to the proof, we give some definitions and lemmas.

Definition 9 (Mean embedding). *Let $\mathcal{H} \subset \mathbb{R}^{\mathcal{X}}$ be a RKHS and μ be a probability measure on $\mathcal{X}$. When it exists, the mean embedding of μ is defined as the element $\varphi_\mu \in \mathcal{H}$ satisfying $\int_{\mathcal{X}} f(x)\mu(dx) = \langle f, \varphi_\mu \rangle$ for all $f \in \mathcal{H}$.*

The mean embedding doesn't always exist. The following lemma ensures that under the assumptions of Theorem 2 the mean embedding exists.

Lemma 1. *If $\int_{\mathcal{X}} \sqrt{K(x,x)}\mu(dx) < \infty$, then the mean embedding of μ exists and we have:*

$$\varphi_\mu(x) = \int_{\mathcal{X}} K(x,y)\mu(dy). \tag{83}$$

Proof. The idea is simply to apply Riesz representation theorem. We define the linear form T_μ on $\mathcal{H}$ as $T_\mu(f) = \int_{\mathcal{X}} f(x)\mu(dx)$ for all $f \in \mathcal{H}$ and we show that is is continuous. Let $f \in \mathcal{H}$, then f is integrable with respect to μ since for all $x \in \mathcal{X}$,

$$|f(x)| = |\langle f, K_x \rangle_{\mathcal{H}}| \leq \|f\|_{\mathcal{H}} \|K_x\|_{\mathcal{H}} = \|f\|_{\mathcal{H}} \sqrt{\langle K_x, K_x \rangle_{\mathcal{H}}} = \|f\|_{\mathcal{H}} \sqrt{K(x,x)} \tag{84}$$

where the first and last equalities are obtained using the reproducing property. By hypothesis $\sqrt{K(x,x)}$ is integrable with respect to μ, so

$$|T_\mu(f)| \leq \int_{\mathcal{X}} |f(x)|\mu(dx) \leq \|f\|_{\mathcal{H}} \int_{\mathcal{X}} \sqrt{K(x,x)}\mu(dx). \tag{85}$$

Thus, T_μ is continuous and by Riesz representation theorem, there exists $\varphi_\mu \in \mathcal{H}$ such that $T_\mu(f) = \langle f, \varphi_\mu \rangle$ for all $f \in \mathcal{H}$. In particular, we have for all $x \in \mathcal{X}$:

$$\varphi_\mu(x) = \langle \varphi_\mu, K_x \rangle = T_\mu(K_x) = \int_{\mathcal{X}} K(x,y)\mu(dy). \tag{86}$$

□

The mean embedding allows to rewrite the MMD in a very simple form when $\mathcal{G}$ is the unit ball in a reproducing kernel Hilbert space.

Lemma 2. *Under the assumptions of Lemma 1, for two probability measures μ and ν, we have:*

$$MMD^2_{\mathcal{G}}(\mu,\nu) = \|\varphi_\mu - \varphi_\nu\|^2 \tag{87}$$

when $\mathcal{G} = \{f \in \mathcal{H} \mid \|f\| \leq 1\}$

Proof.

$$\begin{aligned} MMD^2_{\mathcal{G}}(\mu,\nu) &= \left[\sup_{\|f\| \leq 1} \left| \int_{\mathcal{X}} f(x)\mu(dx) - \int_{\mathcal{X}} f(x)\nu(dx) \right| \right]^2 \\ &= \left[\sup_{\|f\| \leq 1} |\langle \varphi_\mu - \varphi_\nu, f \rangle| \right]^2 \\ &= \|\varphi_\mu - \varphi_\nu\|^2. \end{aligned} \tag{88}$$

□

The proof of Theorem 2 is now straightforward:

Proof of Theorem 2. According to Lemma 2, we have:

$$\begin{aligned} MMD^2_{\mathcal{G}}(\mu,\nu) &= \|\varphi_\mu - \varphi_\nu\|^2 \\ &= \langle \varphi_\mu, \varphi_\mu \rangle + \langle \varphi_\nu, \varphi_\nu \rangle - 2\langle \varphi_\mu, \varphi_\nu \rangle \\ &= \int_{\mathcal{X}} \varphi_\mu(x)\mu(dx) + \int_{\mathcal{X}} \varphi_\nu(x)\nu(dx) - 2\int_{\mathcal{X}} \varphi_\nu(x)\mu(dx) \\ &= \int_{\mathcal{X}\times\mathcal{X}} K(x,y)\mu(dx)\mu(dy) + \int_{\mathcal{X}\times\mathcal{X}} K(x,y)\nu(dx)\nu(dy) - 2\int_{\mathcal{X}\times\mathcal{X}} K(x,y)\mu(dx)\nu(dy), \end{aligned} \tag{89}$$

where the third equality comes from the definition of the mean embeddings and the fourth equality comes from Lemma 1. □

Appendix B. Sufficient condition for the MMD to be a metric

In this section, we provide a sufficient condition on the Mercer kernel K for the MMD to be a metric. For this purpose, we introduce the following definition.

Definition 10. *Let K be a Mercer kernel and denote by $\mathcal{H}$ the associated RKHS. Let $\mathcal{F} \subset \mathbb{R}^{\mathcal{X}}$ be a TVS such that $\mathcal{H} \subset \mathcal{F}$. We say that K is:*

- *universal to $\mathcal{F}$ if the RKHS $\mathcal{H}$ of K is dense in $\mathcal{F}$*
- *characteristic to $\mathcal{F}'$ if the map:*

$$\begin{array}{rccc} \kappa: & \mathcal{F}' & \to & \mathcal{H}' \\ & D & \mapsto & D_{|\mathcal{H}} \end{array} \tag{90}$$

  is one-to-one.

We have a similar theorem to Theorem 13 (the proof is similar as well).

Theorem 10. *Suppose that $\mathcal{F}$ is a locally convex topological vector space (TVS). Then a kernel K is universal to $\mathcal{F}$ iff K is characteristic to $\mathcal{F}'$.*

We can now state a sufficient condition on the kernel K for the MMD to be a metric.

Theorem 11. *Let K be a Mercer kernel and $\mathcal{F}$ be the space of bounded continuous functions on $\mathcal{X}$. If K is universal to $\mathcal{F}$, then $MMD_{\mathcal{G}}$ is a metric on the space $\mathcal{M}$ defined as:*

$$\mathcal{M} = \left\{ \mu \text{ Borel probability measure defined on } \mathcal{X} \middle| \int_{\mathcal{X}} \sqrt{K(x,x)}\mu(dx) < \infty \right\} \tag{91}$$

when $\mathcal{G}$ is the unit ball of the RKHS $\mathcal{H}$ associated to K.

Proof. Since the MMD is a pseudo-metric, we only have to prove that $MMD_{\mathcal{G}}(\mu,\nu) = 0$ implies $\mu = \nu$. Assume that $MMD_{\mathcal{G}}(\mu,\nu) = 0$. According to Lemma 2, we therefore have $\|\varphi_\mu - \varphi_\nu\| = 0$, i.e. $\varphi_\mu = \varphi_\nu$. As a consequence, we have for all $f \in \mathcal{H}$:

$$\int_{\mathcal{X}} f(x)\mu(dx) = \langle f, \varphi_\mu \rangle = \langle f, \varphi_\nu \rangle = \int_{\mathcal{X}} f(x)\nu(dx). \tag{92}$$

Thus, μ and ν (seen as linear forms) are equal on $\mathcal{H}$. Since K is universal to $\mathcal{F}$, it is characteristic to $\mathcal{F}'$ which is the space of finite regular Borel measures on $\mathcal{X}$ (see [14] for a proof). The definition of characteristicness allows to conclude that $\mu = \nu$. □

Remark 8. *Since universality is equivalent to characteristicness, the previous theorem can be stated using one or the other assumption. We prefer the assumption of universality as it is easier to prove in practice.*

Appendix C. Proof of Theorem 3

The proof relies on two following lemmas. The first one states the continuity of the signature which is necessary to ensure that the signature kernel is a Mercer kernel. The second lemma, proved by Chevyrev and Oberhauser [6], states the equivalence between the universality of a map Ψ and the universality of the associated kernel $K(x,y) := \langle \Psi(x), \Psi(y) \rangle$ which is key to prove that the MMD is a metric. We first introduce the definition of the total variation norm.

Definition 11 (Total variation norm)**.** *Let E be a vector space endowed with a norm $\|\cdot\|_E$. The total variation norm on the set $\mathcal{C}^1([0,T],E)$ of bounded variation paths is defined as:*

$$\|X\|_{\mathcal{C}^1([0,T],E)} = \|X\|_{1,[0,T]} + \sup_{t\in[0,T]} \|X_t\|_E \tag{93}$$

where we recall that $\|\cdot\|_{1,[0,T]}$ is the total variation (see Definition 3).

Lemma 3. *The signature map defined on bounded variation paths:*

$$\begin{array}{rcl} S : \left(\mathcal{C}^1([0,T],E), \|\cdot\|_{\mathcal{C}^1([0,T],E)}\right) & \to & (T_1^*(E), \|\cdot\|) \\ X & \mapsto & S(X) \end{array} \tag{94}$$

is continuous (we recall that $\|\cdot\|$ is the norm defined in equation (11)).

Proof. Consider a sequence of bounded variation paths $(X_n)_{n\geq 0}$ converging to some $X \in \mathcal{C}^1([0,T],E)$ in total variation norm. We have:

$$\begin{aligned} \|S(X_n) - S(X)\| &= \sum_{k\geq 0} \|\mathbf{X}_n^k - \mathbf{X}^k\|_{E^{\otimes k}} \\ &= \sum_{k=0}^{N} \|\mathbf{X}_n^k - \mathbf{X}^k\|_{E^{\otimes k}} + \sum_{k\geq N+1} \|\mathbf{X}_n^k - \mathbf{X}^k\|_{E^{\otimes k}} \end{aligned} \tag{95}$$

for some $N \geq 0$. First, note that we have:

$$\begin{aligned} \|\mathbf{X}_n^k - \mathbf{X}^k\|_{E^{\otimes k}} &\leq \|\mathbf{X}_n^k\|_{E^{\otimes k}} + \|\mathbf{X}^k\|_{E^{\otimes k}} \\ &\leq \frac{1}{k!}\left(\|X_n\|_{1,[0,T]}^k + \|X\|_{1,[0,T]}^k\right) \\ &\leq \frac{1}{k!}\left(\|X_n\|_{\mathcal{C}^1([0,T],E)}^k + \|X\|_{\mathcal{C}^1([0,T],E)}^k\right) \\ &\leq \frac{1}{k!}\left(\|X_n\|_{\mathcal{C}^1([0,T],E)}^k + \|X\|_{\mathcal{C}^1([0,T],E)}^k\right) \\ &\leq \frac{1}{k!}\left(M^k + \|X\|_{\mathcal{C}^1([0,T],E)}^k\right) \end{aligned} \tag{96}$$

where the second inequality results from Proposition 2.2 in [24] and $M = \sup_{n\geq 0} \|X_n\|_{\mathcal{C}^1([0,T],E)}$ which is finite since $(X_n)_{n\geq 0}$ is convergent in total variation norm. We deduce that $\sup_{n\geq 0} \sum_{k\geq N+1} \|\mathbf{X}_n^k -$

$\mathbf{X}^k\|_{E^{\otimes k}} \underset{N\to+\infty}{\to} 0$. Moreover, we have:

$$\begin{aligned}\sum_{k=0}^{N}\|\mathbf{X}_n^k-\mathbf{X}^k\|_{E^{\otimes k}} &= \|S^N_{[0,T]}(X_n)-S^N_{[0,T]}(X)\| \\ &\leq \sup_{t\in[0,T]}\|S^N_{[0,t]}(X_n)-S^N_{[0,t]}(X)\| \\ &\leq \|S^N_{[0,\cdot]}(X_n)-S^N_{[0,\cdot]}(X)\|_{\mathcal{C}^1([0,T],T^N(E))}.\end{aligned} \tag{97}$$

According to Corollary 2.11 in [24], the truncated signature

$$\begin{array}{rccl} S^N_{[0,\cdot]}:\mathcal{C}^1([0,T],E) & \to & \mathcal{C}^1([0,T],T^N(E)) \\ X & \mapsto & t\mapsto S^N_{[0,t]}(X)\end{array} \tag{98}$$

where $T^N(E)$ is the space of truncated formal series of tensors of order N, is continuous in total variation norm. Thus, $\sum_{k=0}^N\|\mathbf{X}_n^k-\mathbf{X}^k\|_{E^{\otimes k}} \underset{n\to+\infty}{\to} 0$. Hence $\|S(X_n)-S(X)\| \underset{n\to+\infty}{\to} 0$. □

Lemma 4. *Consider a map $\Psi:\mathcal{X}\to E$ where E is equipped with a scalar product $\langle\cdot,\cdot\rangle$ and define $K(x,y)=\langle\Psi(x),\Psi(y)\rangle$. If $\mathcal{F}$ is a locally convex TVS such that $\mathcal{H}_0:=\operatorname{span}\{K_x:=K(x,\cdot)\mid x\in\mathcal{X}\}$ continuously embeds into $\mathcal{F}$, then Ψ is universal to $\mathcal{F}$ iff K is universal to $\mathcal{F}$.*

Proof of Theorem 3. We start by showing that K^{sig} is a Mercer kernel. The symmetry of K^{sig} is deduced from the symmetry of $\langle\cdot,\cdot\rangle$. The continuity arises from the continuity of the signature (Lemma 3), which is preserved by the tensor normalization, and of $\langle\cdot,\cdot\rangle$. Finally, for $\alpha_1,\dots,\alpha_n\in\mathbb{R}$ and $x_1,\dots,x_n\in\mathcal{P}^1([0,T],E)$, we have:

$$\sum_{i=1}^n\sum_{j=1}^n\alpha_i\alpha_jK^{sig}(x_i,x_j)=\sum_{i=1}^n\sum_{j=1}^n\alpha_i\alpha_j\left\langle\Phi(x_i),\Phi(x_j)\right\rangle=\left\|\sum_{i=1}^n\alpha_i\Phi(x_i)\right\|^2\geq 0. \tag{99}$$

Thus, K^{sig} is also positive semi-definite and thereby K^{sig} is a Mercer kernel. As such, Moore-Aronszajn's theorem can be applied and we can associate a RKHS $\mathcal{H}^{sig}$ to the kernel K^{sig}. A direct application of Theorem 2 yields equation (16).
The only remaining point to show is that $MMD_{\mathcal{G}}$ is a metric. Let $\mathcal{F}$ be the space of bounded continuous functions on $\mathcal{P}^1([0,T],E)$. According to Theorem 14 in Appendix E.3, Φ is universal to $\mathcal{F}$. Next, we show that $\mathcal{H}_0:=\operatorname{span}\left\{K_x^{sig}:=K^{sig}(x,\cdot)\mid x\in\mathcal{P}^1([0,T],E)\right\}$ continuously embeds into $\mathcal{F}$. Let us equip $\mathcal{H}_0$ with the scalar product $\langle\cdot,\cdot\rangle_{\mathcal{H}_0}$ defined by:

$$\langle f,g\rangle_{\mathcal{H}_0}:=\sum_{i=1}^r\sum_{j=1}^s\alpha_i\beta_jK^{sig}(x_i,y_j)\quad\text{for } f=\sum_{i=1}^r\alpha_iK^{sig}_{x_i}\text{ and } g=\sum_{i=1}^s\beta_iK^{sig}_{y_i}. \tag{100}$$

It is clear that $\mathcal{H}_0$ satisfies the reproducing property. Therefore, if $f\in\mathcal{H}_0$, then for all $x\in\mathcal{P}^1([0,T],E)$,

$$f(x)=\langle f,K_x^{sig}\rangle_{\mathcal{H}_0}\leq\|f\|_{\mathcal{H}_0}\|K_x^{sig}\|_{\mathcal{H}_0} \tag{101}$$

using the Cauchy-Schwarz inequality. By definition of $\|\cdot\|_{\mathcal{H}_0}$, we have

$$\|K_x^{sig}\|_{\mathcal{H}_0}=\sqrt{K^{sig}(x,x)}=\sqrt{\langle\Phi(x),\Phi(x)\rangle_{T(E)}}=\|\Phi(x)\|_{T(E)}\leq K \tag{102}$$

where the last inequality comes from the definition of the tensor normalization. We deduce that:

$$\|f\|_\infty=\sup_{x\in\mathcal{X}}|f(x)|\leq K\|f\|_{\mathcal{H}_0} \tag{103}$$

i.e. $\mathcal{H}_0$ continuously embeds into $\mathcal{F}$ endowed with the uniform norm $\|\cdot\|_\infty$. Thus, according to Lemma 4, K^{sig} is universal to $\mathcal{F}$. Finally, using Theorem 11, we conclude that $MMD_{\mathcal{G}}$ is a metric. □

Remark 9. *A natural question is whether Theorem 3 can be extended to $K^{sig}(x,y) = K(\Phi(x),\Phi(y))$ where K is a Mercer kernel on $T(E)\times T(E)$. It is easy to show that in this case, K^{sig} is still a Mercer kernel. However, we don't know if the conclusion of Lemma 4 still holds with this definition.*

Appendix D. Asymptotic distribution of $MMD_{m,n}$ under H_1

The asymptotic distribution of $MMD_{m,n}$ under H_1 can be obtained by applying a central limit theorem for U-statistics. Before going further, let us define what are U-statistics. In this section, E is an arbitrary set.

Definition 12 (One-sample U-statistic)**.** *Let $X_1,\dots,X_n$ be i.i.d. $\mathbb{R}^d$-valued random variables and $h : E^r \to \mathbb{R}$ a measurable function with $r \leq n$. A one-sample U-statistic with kernel h is defined as:*

$$U_n = \frac{(n-r)!}{n!} \sum_{(i_1,\dots,i_r)\in I_{r,n}} h(X_{i_1},\dots,X_{i_r}) \tag{104}$$

where:

$$I_{r,n} = \{(i_1,\dots,i_r) \in \{1,\dots,n\}^r \mid i_k \neq i_\ell, \forall k \neq \ell\}\,. \tag{105}$$

Remark 10. *If h is symmetric in its arguments, U_n can be rewritten as:*

$$U_n = \binom{n}{r}^{-1} \sum_{1\leq i_1<\dots<i_r\leq n} h(X_{i_1},\dots,X_{i_r}) \tag{106}$$

This definition can be extended to multiple samples:

Definition 13 (k-sample U-statistic)**.** *Let $(X_i^1)_{1\leq i\leq n_1},\dots,(X_i^k)_{1\leq i\leq n_k}$ be k samples of i.i.d. E-valued random variables and $h : E^{r_1}\times\cdots\times E^{r_k}\to\mathbb{R}$ a measurable function. A k-sample U-statistic with kernel h is defined as:*

$$U_{\mathbf{n}} = \prod_{i=1}^{k} \frac{(n_i-r_i)!}{n_i!} \sum_{(\mathbf{i^1},\dots,\mathbf{i^k})\in I_{r_1,n_1}\times\cdots\times I_{r_k,n_k}} h(\mathbf{X^1_{i^1}},\dots,\mathbf{X^k_{i^k}}) \tag{107}$$

where we used the notations $\mathbf{n} = (n_1,\dots,n_k)$, $\mathbf{i}^\ell = (i^\ell_1,\dots,i^\ell_{r_\ell})$ and $\mathbf{X}^\ell_{\mathbf{i}^\ell} = (X^\ell_{i^\ell_1},\dots,X^\ell_{i^\ell_{r_\ell}})$ for $\ell\in\{1,\dots,k\}$.

Remark 11. *$(X_i^1)_{1\leq i\leq n_1},\dots,(X_i^k)_{1\leq i\leq n_k}$ need not to be distributed according to the same probability distribution.*

Let us define h by:

$$h(x,x',y,y') = K^{sig}(x,x')+K^{sig}(y,y')-\frac{1}{2}\left(K^{sig}(x,y)+K^{sig}(x,y')+K^{sig}(x',y)+K^{sig}(x',y')\right). \tag{108}$$

Then, it is straightforward to check that the estimator $MMD^2_{m,n}$ in equation (19) can be rewritten as:

$$MMD^2_{m,n} = \binom{m}{2}^{-1} \binom{n}{2}^{-1} \sum_{\substack{1\leq i<j\leq m \\ 1\leq k<\ell\leq n}} h(X_i, X_j, Y_k, Y_\ell). \tag{109}$$

As such, $MMD^2_{m,n}$ is a two-sample U-statistic. We have the following central limit theorem for two-sample U-statistics (see Theorem 1 of section 3.7.1 in [21] for a proof):

Theorem 12. *Let U_{n_1,n_2} be a two-sample U-statistic:*

$$U_{n_1,n_2} = \binom{n_1}{m_1}^{-1} \binom{n_2}{m_2}^{-1} \sum_{1\leq i_1<\cdots<i_{m_1}\leq n_1} \sum_{1\leq j_1<\ldots j_{m_2}\leq n_2} h(X_{i_1}, \ldots, X_{i_{m_1}}, Y_{j_1}, \ldots, Y_{j_{m_2}}) \tag{110}$$

where:

- *$X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ are i.i.d. samples from two distributions $\mathbb{P}_X$ and $\mathbb{P}_Y$*
- *h is invariant for permutations of $X_1, \ldots, X_{m_1}$ and of $Y_1, \ldots, Y_{m_2}$*

Under the assumptions that

(i) $\mathbb{E}\left[h(X_1, \ldots, X_{m_1}, Y_1, \ldots, Y_{m_2})^2\right] < +\infty$

(ii) $n_1/N \to \rho \in (0,1)$ *as* $N = n_1 + n_2 \to +\infty$

we have:

$$\sqrt{N}(U_{n_1,n_2} - \theta) \underset{N\to+\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}\left(0, \frac{m_1^2}{\rho}\sigma_1^2 + \frac{m_2^2}{1-\rho}\sigma_2^2\right) \tag{111}$$

where $\theta = \mathbb{E}[h(X_1, \ldots, X_{m_1}, Y_1, \ldots, Y_{m_2})]$, $\sigma_1^2 = Var(h_1(X_1))$ and $\sigma_2^2 = Var(h_2(Y_1))$ with $h_1(x) = \mathbb{E}[h(x, X_2, \ldots, X_{m_1}, Y_1, \ldots, Y_{m_2})]$ and $h_2(y) = \mathbb{E}[h(X_1, X_2, \ldots, X_{m_1}, y, Y_2, \ldots, Y_{m_2})]$.

Remark 12. *This theorem can be stated more generally for k-sample U-statistics but we present here only the two-sample case for simplicity.*

Under H_0, we have $\mathbb{P}_X = \mathbb{P}_Y$ which implies that $h_1(x) = h_2(y) = 0$ for all x and y with h given in equation (108). Thus, the asymptotic variance is zero and a factor growing quicker to infinity than $\sqrt{N}$ is needed to obtain a non-vanishing limit. On the contrary, under H_1, we have $\mathbb{P}_X \neq \mathbb{P}_Y$ so that the asymptotic variance is not necessarily zero, which makes Theorem 12 interesting.

Appendix E. Technical details on the characterization of stochastic processes law by the expected normalized signature

Appendix E.1. Normalization existence

In this subsection, we discuss the existence of tensor normalizations. The following proposition is proved in Proposition A.2. of [6].

Proposition 6. *Let $\psi : [1, +\infty) \to [1, +\infty)$ with $\psi(1) = 1$. For $\mathbf{t} \in T_1^*(E)$, let $\lambda(\mathbf{t}) \geq 0$ denote the unique λ such that:*

$$\sum_{n\geq 0} \lambda^{2n} \|\mathbf{t}^n\|^2_{E^{\otimes n}} = \psi(\|\mathbf{t}\|) \tag{112}$$

and define Λ as in Definition 8 with $\lambda(\mathbf{t})$ as specified above. Denote further $\|\psi\|_\infty = \sup_{x\geq 1} \psi(x)$. Under the following assumptions on ψ:

(i) ψ is one-to-one,

(ii) $\|\psi\|_\infty < \infty$,

(iii) $\sup_{x\geq 1} \frac{\psi(x)}{x^2} \leq 1$,

(iv) ψ is K-Lipschitz for some $K > 0$,

then Λ is a tensor normalization.

Remark 13. *Although it is not discussed by Chevyrev and Oberhauser, a solution to equation (112) a priori doesn't always exist. Indeed, the left-hand term can be seen as a power series in λ (with only even terms):*

$$P(\lambda) = \sum_{n\geq 0} \lambda^{2n} \|\mathbf{t}^n\|^2_{E^{\otimes n}}. \tag{113}$$

Its radius of convergence is greater or equal to 1 since by definition of $T^*(E)$, we have $\|\mathbf{t}\| < \infty$. Thus, on $[0, 1]$, P is continous and strictly increasing with $P(0) = 1$ and $P(1) = \|\mathbf{t}\|$. As such, if $\psi(\|t\|)$ is greater than $\|\mathbf{t}\|$, there is a priori no reason for the existence of a solution to (112) without further assumption on the rate of decay of the sequence $(\|\mathbf{t}^n\|_{E^{\otimes n}})_{n\geq 0}$. However, if we think $\mathbf{t}$ as the signature of some path X, we know that (see [24]) $\|\mathbf{t}^n\|_{E^{\otimes n}}$ decays as $\frac{\|X\|^n_{1,[0,T]}}{n!}$ if X is of bounded variation. As a consequence, if $\mathbf{t}$ is the signature of some path, the radius of convergence of P is infinite and equation (112) has indeed a unique solution. Anyway, in practice the power series is truncated and there is always a unique solution belonging to $\mathbb{R}_+$.

In their numerical experiments, Chevyrev and Oberhauser consider the following form for ψ:

$$\psi(x) = \begin{cases} x^2 & \text{if } x \leq \sqrt{M} \\ M + M^{1+a}(M^{-a} - x^{-2a})/a & \text{if } x > \sqrt{M} \end{cases} \tag{114}$$

with $M = 4$ and $a = 1$. This function clearly satisfies the hypotheses of Proposition 6 so Λ constructed from ψ is a tensor normalization.

Appendix E.2. Universality and characteristicness

In this subsection, we define the notions of universality and characteristicness that underlie the main results of Chevyrev and Oberhauser.

Let $\mathcal{X}$ be a topological space and $\mathcal{F} \subset \mathbb{R}^{\mathcal{X}}$ a topological vector space (TVS). The space $\mathcal{X}$ is the input space and $\mathcal{F}$ is the hypothesis space (in learning problems, it corresponds to the space in which we are looking for the function we want to learn). For a TVS E, we call any map:

$$\Phi : \mathcal{X} \to E \tag{115}$$

a feature map for which E is the feature space.

Notation: we will denote by E' the dual of E, that is the space of all continuous linear forms on E, and by E^* the algebraic dual of E, that is the space of all linear forms on E (not necessarily continuous).

Definition 14 (Universality and characteristicness)**.** *Suppose that $\langle l, \Phi(\cdot)\rangle \in \mathcal{F}$ for all $l \in E'$. We say that Φ is*

- *universal to $\mathcal{F}$ if the map*

$$\begin{array}{rccc} \iota: & E' & \to & \mathbb{R}^{\mathcal{X}} \\ & l & \mapsto & \langle l, \Phi(\cdot)\rangle \end{array} \tag{116}$$

  has a dense image in $\mathcal{F}$, i.e. $\overline{\iota(E')} = \mathcal{F}$.

- *characteristic to a subset $\mathcal{P} \subset \mathcal{F}'$ if the map*

$$\begin{array}{rccc} \kappa: & \mathcal{P} & \to & (E')^* \\ & D & \mapsto & [l \mapsto D(\langle l, \Phi(\cdot)\rangle)] \end{array} \tag{117}$$

  is one-to-one.

These two notions are actually equivalent under mild assumptions. We refer to [28] for a definition of a locally convex topological vector space.

Theorem 13. *Suppose that $\mathcal{F}$ is a locally convex TVS. Then, a feature map Φ is universal to $\mathcal{F}$ iff Φ is characteristic to $\mathcal{F}'$.*

Proof. Let us first assume that Φ is universal to $\mathcal{F}$. Let $D_1, D_2 \in \mathcal{F}'$ such that $\kappa(D_1) = \kappa(D_2)$. Then, for all $l \in E'$, we have $\kappa(D_1)(l) = \kappa(D_2)(l)$, that is $D_1(\langle l, \Phi(\cdot)\rangle) = D_2(\langle l, \Phi(\cdot)\rangle)$, which rewrites $D_1(\iota(l)) = D_2(\iota(l))$. Therefore, D_1 and D_2 coincide on $\iota(E')$. Since $\iota(E')$ is dense in $\mathcal{F}$ by hypothesis and D_1 and D_2 are continuous, we have $D_1 = D_2$ on $\mathcal{F}$. Hence, κ is one-to-one and Φ is characteristic to $\mathcal{F}'$.
Now assume that Φ is not universal to $\mathcal{F}$, i.e. $\iota(E')$ is not dense in $\mathcal{F}$. By assumption, we have $\iota(E') \subset \mathcal{F}$, $\overline{\iota(E')} \neq \mathcal{F}$ and $\mathcal{F}$ is a locally convex TVS. According to a corollary of the Hahn-Banach theorem (see corollary 3 of theorem 18.1 in [28]), there exists under these assumptions $f \in \mathcal{F}'$ such that $f \neq 0$ but $f|_{\iota(E')} = 0$. As a consequence, $\kappa(f)(l) = f(\iota(l)) = 0$ for all $l \in E'$, so we have $\kappa(f) = \kappa(0)$ but $f \neq 0$. Hence, κ is not one-to-one and Φ is not characteristic to $\mathcal{F}'$. □

The following lemma combined with Theorem 13 is useful to state that measures are characterized by their action against a feature map.

Lemma 5. *Under the assumptions:*

(i) $(E, \langle\cdot,\cdot\rangle_E)$ is a Hilbert space,

(ii) $\mathcal{F}'$ consists of finite measures on $\mathcal{X}$,

(iii) $\|\Phi\|_\infty = \sup_{x\in\mathcal{X}} \|\Phi(x)\|_E < \infty$,

then the characteristicness of the feature map Φ (i.e. the injectivity of κ) implies the injectivity of the mapping $\tilde{\kappa}$ defined as:

$$\begin{array}{rccc} \tilde{\kappa}: & \mathcal{F}' & \to & E \\ & \mu & \mapsto & \int_{\mathcal{X}} \Phi(x)\mu(dx) \end{array} \tag{118}$$

Proof. Let μ_1 and μ_2 in $\mathcal{F}'$ such that $\tilde{\kappa}(\mu_1) = \tilde{\kappa}(\mu_2)$. This can be rewritted:

$$\int_{\mathcal{X}} \Phi(x)\mu_1(dx) = \int_{\mathcal{X}} \Phi(x)\mu_2(dx). \tag{119}$$

Let $l \in E'$, we have:

$$\left\langle l, \int_{\mathcal{X}} \Phi(x)\mu_1(dx) \right\rangle = \left\langle l, \int_{\mathcal{X}} \Phi(x)\mu_2(dx) \right\rangle \tag{120}$$

Identifying l with its Riesz representation, we have:

$$|\langle l, \Phi(x)\rangle_E| \leq \|l\|_E \|\Phi(x)\|_E \leq \|l\|_E \|\Phi\|_\infty \tag{121}$$

Since μ_1 is a finite measure, we deduce that $\int_{\mathcal{X}} |\langle l, \Phi(x)\rangle_E| \, \mu_1(dx) < \infty$. Thus, we can interchange the integral and the scalar product, which yields:

$$\int_{\mathcal{X}} \langle l, \Phi(x)\rangle \mu_1(dx) = \int_{\mathcal{X}} \langle l, \Phi(x)\rangle \mu_2(dx), \tag{122}$$

which is exactly $\kappa(\mu_1)(l) = \kappa(\mu_2)(l)$. Thus $\mu_1 = \mu_2$ by injectivity of κ. □

Appendix E.3. General version of Theorem 9

In this section, we present a more general version of Theorem 9.

We recall that $\mathcal{C}^1([0,T], E)$ is the set of bounded variation paths from $[0,T]$ to E and $\mathcal{P}^1([0,T], E)$ is the set of bounded variation paths quotiented by the tree-like equivalence relation. As there is no ambiguity here, we juste write $\mathcal{C}^1$ and $\mathcal{P}^1$. We denote by $C_b(\mathcal{P}^1, \mathbb{R})$ the space of continuous bounded functions from $\mathcal{P}^1$ to $\mathbb{R}$. This space can be equipped with the strict topology which is for example defined in [14]. With these notations, we have the following theorem due to Chevyrev and Oberhauser (Theorem 6.1. in [6]).

Theorem 14. *Let Λ be a tensor normalization. Then the map*

$$\begin{array}{rcl} \Phi : \mathcal{P}^1 & \to & T_1^*(E) \\ X & \mapsto & \Lambda \circ S(X) \end{array} \tag{123}$$

(i) is a continuous injection from $\mathcal{P}^1$ into a bounded subset of $T_1^*(E)$,

(ii) is universal to $\mathcal{F} := C_b(\mathcal{P}^1, \mathbb{R})$ equipped with the strict topology,

(iii) is characteristic to the space of finite regular Borel measures on $\mathcal{P}^1$.

The same statement holds when $\mathcal{P}^1$ is replaced by $\mathcal{C}^1$ and Φ is replaced by $\bar{\Phi} := \Lambda \circ S(\bar{X})$ where $\bar{X}(t) = (X(t), t)$.

Remark 14. *Theorem 9 follows from (iii) and Lemma 5.*

Appendix F. Zero odd-order restricted MMD between centered Gaussian processes

In this section, we justify theoretically why the MMD restricted to odd orders doesn't allow to distinguish fBm's with different Hurst parameters. We first need the following lemma.

Lemma 6. *Let X be a real-valued stochastic process. We consider a vector $(X_{t_0}, X_{t_1}, \dots, X_{t_N})$ of values of X on a partition $0 = t_0 < t_1 < \cdots < t_N = T$ of $[0,T]$ and we denote by $\hat{X}$ the lead-lag transformation of X on this partition. Then, for all $k \in \mathbb{N}$, for all $I = (i_1, \dots, i_k) \in \{1,2\}^k$ (with the convention $I = 1$ if $k = 0$) and for all $s \in [0,T]$, we have:*

$$\hat{\mathbf{X}}^k_{I,[0,s]} = \sum_{\alpha \in \mathcal{J}_k(\tau(s)+1)} P_\alpha(s) X_{t_0}^{\alpha_0} X_{t_1}^{\alpha_1} \dots X_{t_{\tau(s)+1}}^{\alpha_{\tau(s)+1}} \tag{124}$$

where

- $\hat{\mathbf{X}}^k_{I,[0,s]}$ *is the coefficient at position I of the k-th term of the signature of $\hat{X}$ on $[0,s]$*
- $\tau(s) \in \{0, \dots, N-1\}$ *such that $t_{\tau(s)} \leq s < t_{\tau(s)+1}$*
- $\mathcal{J}_k(l) = \left\{ \alpha = (\alpha_0, \dots, \alpha_l) \in \{0, \dots, k\}^l \mid \sum_{i=0}^l \alpha_i = k \right\}$
- $P_\alpha(s)$ *is a polynomial that depends on α and $t_0, \dots, t_N$ but not on X.*

Proof. We proceed by induction on k. If $k = 0$, then by definition of the signature, $\hat{\mathbf{X}}^0_{1,[0,s]} = 1$ for all $s \in [0,T]$ so (124) is verified. Let us assume that (124) is verified for some $k \geq 0$. Let $I = (i_1, \dots, i_{k+1}) \in \{1,2\}^{k+1}$ and $s \in [0,T]$. We set $I' = (i_1, \dots, i_k)$. By definition of the signature, we have:

$$\begin{aligned}
\hat{\mathbf{X}}^{k+1}_{I,[0,s]} &= \int_{0<u_1<\cdots<u_{k+1}<s} d\hat{X}^{i_1}_{u_1} \times \cdots \times d\hat{X}^{i_{k+1}}_{u_{k+1}} \\
&= \int_0^s \hat{\mathbf{X}}^k_{I',[0,u]} d\hat{X}^{i_k+1}_u \\
&= \sum_{i=0}^{\tau(s)-1} \int_{t_i}^{t_{i+1}} \hat{\mathbf{X}}^k_{I',[0,u]} d\hat{X}^{i_k+1}_u + \int_{\tau(s)}^s \hat{\mathbf{X}}^k_{I',[0,u]} d\hat{X}^{i_k+1}_u \\
&= \sum_{i=0}^{\tau(s)-1} \int_{t_i}^{t_{i+1}} \left(\sum_{\alpha \in \mathcal{J}_k(\tau(u)+1)} P_\alpha(u) X_{t_0}^{\alpha_0} \dots X_{\tau(u)+1}^{\alpha_{\tau(u)+1}} \right) d\hat{X}^{i_k+1}_u \\
&+ \int_{\tau(s)}^s \left(\sum_{\alpha \in \mathcal{J}_k(\tau(u)+1)} P_\alpha(u) X_{t_0}^{\alpha_0} \dots X_{\tau(u)+1}^{\alpha_{\tau(u)+1}} \right) d\hat{X}^{i_k+1}_u \\
&= \sum_{i=0}^{\tau(s)-1} \sum_{\alpha \in \mathcal{J}_k(i+1)} X_{t_0}^{\alpha_0} \dots X_{t_{i+1}}^{\alpha_{i+1}} \int_{t_i}^{t_{i+1}} P_\alpha(u) d\hat{X}^{i_k+1}_u \\
&+ \sum_{\alpha \in \mathcal{J}_k(\tau(s)+1)} X_{t_0}^{\alpha_0} \dots X_{\tau(s)+1}^{\alpha_{\tau(s)+1}} \int_{\tau(s)}^s P_\alpha(u) d\hat{X}^{i_k+1}_u
\end{aligned} \tag{125}$$

Using the definition of the lead-lag transformation, it is easy to see that if $i_{k+1} = 1$ (corresponding to the lead component) and $u \in [t_i, t_{i+1}]$:

$$d\hat{X}^{i_{k+1}}_u = \frac{X_{t_{i+1}} - X_{t_i}}{t_{i+1/2} - t_i} \mathbb{1}_{[t_i, t_{i+1/2}]}(u) du \tag{126}$$

with the notations introduced in subsection 3.1.2. Similarly, if $i_{k+1} = 2$ (corresponding to the lag component) and $u \in [t_i, t_{i+1}]$:

$$d\hat{X}^{i_{k+1}}_u = \frac{X_{t_{i+1}} - X_{t_i}}{t_{i+1} - t_{i+1/2}} \mathbb{1}_{[t_{i+1/2}, t_{i+1}]}(u) du. \tag{127}$$

Since the integral of a polynomial is still a polynomial and for $i \in \{0, \dots, \tau(s)-1\}$, $\mathcal{J}_k(i+1)$ can be seen as a subset of $\mathcal{J}_k(\tau(s)+1)$ by adding zero entries, we finally get

$$\hat{\mathbf{X}}^{k+1}_{I,[0,s]} = \sum_{\alpha \in \mathcal{J}_{k+1}(\tau(s)+1)} \tilde{P}_\alpha(s) X_{t_0}^{\alpha_0} X_{t_1}^{\alpha_1} \dots X_{t_{\tau(s)+1}}^{\alpha_{\tau(s)+1}} \tag{128}$$

□

We recall that the signature kernel is defined by $K^{sig}(X,Y) = \langle S(X), S(Y) \rangle$ where $\langle \cdot, \cdot \rangle$ is the scalar product on $T^*(E)$ defined by:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{n \geq 0} \langle \mathbf{x}^n, \mathbf{y}^n \rangle_{E^{\otimes n}} \tag{129}$$

for $\mathbf{x}$ and $\mathbf{y}$ in $T^*(E)$. We now define the signature kernel restricted at order l as:

$$K^{sig}_\ell(X,Y) = \langle \mathbf{X}^\ell, \mathbf{Y}^\ell \rangle_{E^{\otimes \ell}} \tag{130}$$

as well as the MMD restricted at order ℓ as:

$$MMD^\ell_{\mathcal{G}}(\mathbb{P}_X, \mathbb{P}_Y) = \mathbb{E}[K^{sig}_\ell(X_1, X_2)] + \mathbb{E}[K^{sig}_\ell(Y_1, Y_2)] - 2\mathbb{E}[K^{sig}_\ell(X_1, Y_1)] \tag{131}$$

where $\mathcal{G}$ is the unit ball of the RKHS associated to the signature kernel, X_1,X_2 are two independent random variables of law $\mathbb{P}_X$ and Y_1,Y_2 are two independent random variables of law $\mathbb{P}_Y$.

Proposition 7. *Let X and Y be two independent centered unidimensional Gaussian processes and let us denote by $\hat{X}$ and $\hat{Y}$ the lead-lag transformation of X and Y respectively. Then, if ℓ is odd:*

$$MMD^\ell_{\mathcal{G}}(\mathbb{P}_{\hat{X}}, \mathbb{P}_{\hat{Y}}) = 0 \tag{132}$$

Proof. We have:

$$\begin{aligned} MMD^\ell_{\mathcal{G}}(\mathbb{P}_{\hat{X}}, \mathbb{P}_{\hat{Y}}) &= \mathbb{E}[K^{sig}_\ell(\hat{X}_1, \hat{X}_2)] + \mathbb{E}[K^{sig}_\ell(\hat{Y}_1, \hat{Y}_2)] - 2\mathbb{E}[K^{sig}_\ell(\hat{X}_1, \hat{Y}_1)] \\ &= \mathbb{E}[\langle \hat{\mathbf{X}}^\ell_1, \hat{\mathbf{X}}^\ell_2 \rangle_{E^{\otimes \ell}}] + \mathbb{E}[\langle \hat{\mathbf{Y}}^\ell_1, \hat{\mathbf{Y}}^\ell_2 \rangle_{E^{\otimes \ell}}] - 2\mathbb{E}[\langle \hat{\mathbf{X}}^\ell_1, \hat{\mathbf{Y}}^\ell_1 \rangle_{E^{\otimes \ell}}]. \end{aligned} \tag{133}$$

By definition of the scalar product $\langle \cdot, \cdot \rangle_{E^{\otimes \ell}}$:

$$\langle \hat{\mathbf{X}}^\ell_1, \hat{\mathbf{X}}^\ell_2 \rangle_{E^{\otimes \ell}} = \sum_{I=(i_1,\dots,i_\ell) \in \{1,2\}^\ell} \hat{\mathbf{X}}^\ell_{1,I} \hat{\mathbf{X}}^\ell_{2,I}. \tag{134}$$

Thus, by independence of X_1 and X_2, we get:

$$\mathbb{E}[\langle \hat{\mathbf{X}}^\ell_1, \hat{\mathbf{X}}^\ell_2 \rangle_{E^{\otimes \ell}}] = \sum_{I=(i_1,\dots,i_\ell) \in \{1,2\}^\ell} \mathbb{E}[\hat{\mathbf{X}}^\ell_{1,I}] \mathbb{E}[\hat{\mathbf{X}}^\ell_{2,I}] \tag{135}$$

According to Lemma 6, $\hat{\mathbf{X}}^\ell_{1,I}$ is a linear combination of monomials in $X_{1,t_0}, \dots, X_{1,t_N}$ of degree ℓ. Given that ℓ is odd and X_1 is a centered Gaussian process, we obtain that $\mathbb{E}[\hat{\mathbf{X}}^\ell_{1,I}] = 0$ for all $I \in \{1,2\}^\ell$. So $\mathbb{E}[\langle \hat{\mathbf{X}}^\ell_1, \hat{\mathbf{X}}^\ell_2 \rangle_{E^{\otimes \ell}}] = 0$. The same reasoning based on the independence between Y_1 and Y_2 and between X_1 and Y_1 yields $\mathbb{E}[\langle \hat{\mathbf{Y}}^\ell_1, \hat{\mathbf{Y}}^\ell_2 \rangle_{E^{\otimes \ell}}] = 0$ and $\mathbb{E}[\langle \hat{\mathbf{X}}^\ell_1, \hat{\mathbf{Y}}^\ell_1 \rangle_{E^{\otimes \ell}}] = 0$, so that $MMD^\ell_{\mathcal{G}}(\mathbb{P}_{\hat{X}}, \mathbb{P}_{\hat{Y}}) = 0$. □

References

[1] Saeed Asadi and Mazin A. M. Al Janabi. Measuring market and credit risk under Solvency II: evaluation of the standard technique versus internal models for stock and bond markets. *Eur. Actuar. J.*, 10(2):425–456, 2020.

[2] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *J. Polit. Econ.*, 81(3):637–654, 1973.

[3] Horatio Boedihardjo, Xi Geng, Terry Lyons, and Danyu Yang. The signature of a rough path: uniqueness. *Adv. Math.*, 293:720–737, 2016.

[4] Mathieu Boudreault and Christian-Marc Panneton. Multivariate models of equity returns for investment guarantees valuation. *N. Am. Actuar. J.*, 13(1):36–53, 2009.

[5] Ilya Chevyrev and Terry Lyons. Characteristic functions of measures on geometric rough paths. *Ann. Probab.*, 44(6):4049–4082, 2016.

[6] Ilya Chevyrev and Harald Oberhauser. Signature moments to characterize laws of stochastic processes. *arXiv preprint arXiv:1810.10971*, 2018.

[7] Felipe Cucker and Steve Smale. On the mathematical foundations of learning. *Bull. Amer. Math. Soc. (N.S.)*, 39(1):1–49, 2002.

[8] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.

[9] Thomas Fawcett. *Problems in stochastic analysis: Connections between rough paths and non-commutative harmonic analysis*. PhD thesis, University of Oxford, 2002.

[10] Adeline Fermanian. Embedding and learning with signatures. *Comput. Statist. Data Anal.*, 157:Paper No. 107148, 23, 2021.

[11] A. Floryszczak, J. Lévy Véhel, and M. Majri. A conditional equity risk model for regulatory assessment. *Astin Bull.*, 49(1):217–242, 2019.

[12] Peter K. Friz and Nicolas B. Victoir. *Multidimensional stochastic processes as rough paths*, volume 120 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010. Theory and applications.

[13] Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. Volatility is rough. *Quant. Finance*, 18(6):933–949, 2018.

[14] Robin Giles. A generalization of the strict topology. *Trans. Amer. Math. Soc.*, 161:467–474, 1971.

[15] Stefan Graf, Lena Haertel, Alexander Kling, and Jochen Ruß. The impact of inflation risk on financial planning and risk-return profiles. *Astin Bull.*, 44(2):335–365, 2014.

[16] Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *J. Mach. Learn. Res.*, 13:723–773, 2012.

[17] Arthur Gretton, Kenji Fukumizu, Zaid Harchaoui, and Bharath K. Sriperumbudur. A fast, consistent kernel two-sample test. *Advances in neural information processing systems*, 22, 2009.

[18] Lajos Gergely Gyurkó, Terry Lyons, Mark Kontkowski, and Jonathan Field. Extracting information from the signature of a financial data stream. *arXiv preprint arXiv:1307.7244*, 2013.

[19] Ben Hambly and Terry Lyons. Uniqueness for the signature of a path of bounded variation and the reduced path group. *Ann. of Math. (2)*, 171(1):109–167, 2010.

[20] Mary R. Hardy, R. Keith Freeland, and Matthew C. Till. Validation of long-term equity return models for equity-linked guarantees. *N. Am. Actuar. J.*, 10(4):28–47, 2006.

[21] A. J. Lee. *U-statistics*, volume 110 of *Statistics: Textbooks and Monographs*. Marcel Dekker, Inc., New York, 1990. Theory and practice.

[22] Shujian Liao, Terry Lyons, Weixin Yang, and Hao Ni. Learning stochastic differential equations using rnn with log signature features. *arXiv preprint arXiv:1908.08286*, 2019.

[23] X. Sheldon Lin and Shuai Yang. Efficient dynamic hedging for large variable annuity portfolios with multiple underlying assets. *Astin Bull.*, 50(3):913–957, 2020.

[24] Terry J. Lyons, Michael Caruana, and Thierry Lévy. *Differential equations driven by rough paths*, volume 1908 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. Lectures from the 34th Summer School on Probability Theory held in Saint-Flour, July 6–24, 2004, With an introduction concerning the Summer School by Jean Picard.

[25] L Otero, P Durán, S Fernández, and M Vivel. Estimating insurers capital requirements through markov switching models in the solvency ii framework. *International Research Journal of Finance and Economics*, 86:20–38, 2012.

[26] Pranab Kumar Sen. Almost sure convergence of generalized U-statistics. *Ann. Probability*, 5(2):287–290, 1977.

[27] Gordon D. Smith. *Numerical solution of partial differential equations*. Oxford Applied Mathematics and Computing Science Series. The Clarendon Press, Oxford University Press, New York, third edition, 1985. Finite difference methods.

[28] François Trèves. *Topological vector spaces, distributions and kernels*. Academic Press, New York-London, 1967.

[29] L. C. Young. An inequality of the Hölder type, connected with Stieltjes integration. *Acta Math.*, 67(1):251–282, 1936.

[30] Xiaobai Zhu, Mary R. Hardy, and David Saunders. Dynamic hedging strategies for cash balance pension plans. *Astin Bull.*, 48(3):1245–1275, 2018.