Eötvös Loránd University

# APPLICATION OF SIGNATURES FOR FORECASTING

Modelling project work 1

Bat-erdene Egshiglen

Supervisor: Tikosi Kinga

Budapest, Hungary

2020, I semester

When people attempt to predict the future scenarios, one needs to have a basic idea of the past and present. It can be practically easy if we have a constant or a simple smooth graph. But what about cases when this data we are working with is highly non-linear? The answer is to find a linear structure, which can be considered as a supporter or a base. However, depending on what we consider it to be a underpin for our function, the prediction may be wrong. When we try to solve this we use signatures of paths.

The goal of this project is to understand what this method is and its usages for forecasting a chosen time-series, for example, financial data. In real life, the graph of stock prices will be a non-smooth path, that is why when it comes to a rough path, we consider the signature of it as a fundamental object, instead of models which cannot fit the data or which cannot generalise it.

In the first semester, we focused on the basics, such as what is a signature and how does it work on simple smooth deterministic functions.

**Definition 0.1.** A path in a Euclidean space, which we will denote by $X$, is a continuous mapping from an interval $[a, b]$ to $\mathbb{R}^m$. We will use $X_t$ for a path parameterized by a point $t \in [a, b]$.

$$X_t = \{X_t^1, X_t^2 \ldots, X_t^m\}, \tag{1}$$

we call these coordinate paths.

**Definition 0.2.** Let us assume $X_t : [a, b] \mapsto \mathbb{R}^m$, then the signature of the path $X_t$ is an infinite series of the iterated integrals

$$S(X)_{a,b} = (1, S(X)^1_{a,b}, S(X)^2_{a,b} \ldots, S(X)^m_{a,b}, S(X)^{11}_{a,b}, \ldots). \tag{2}$$

To be precise, we define them as:

$$S(X)^i_{a,t} = \int_{a<s<b} dX^i_s = X^i_t - X^i_0 \tag{3}$$

for any $i \in \{1, \dots, m\}$, $X_t = \{X_t^1, X_t^2 \dots, X_t^m\}$, additionally

$$S(X)_{a,t}^{i,j} = \int_{a<s<b} S(X)_{a,s}^{i} \, dX_s^j = \int_{a<r<s<t} dX_r^i dX_s^j \tag{4}$$

for any $i, j \in \{1, \dots, m\}$.

Furthermore, we will continue recursively, for any positive integer $n$, we consider the collection of indices (multi-indices) $i_1, i_2, \dots, i_n \in \{1, 2 \dots, m\}$ and define the signature as

$$S(X)_{a,t}^{i_1,\dots,i_n} = \int_{a<t_m<t} \dots \int_{a<t_1<t_2} dX_{t_1}^{i_1} \dots dX_{t_n}^{i_n} \tag{5}$$

This time we only consider $\mathbb{R}^2$ and $\mathbb{R}$.

**Remark.** If we have the paths $X : [a, b] \mapsto \mathbb{R}$, $Y : [a, b] \mapsto \mathbb{R}$ and a function $f : \mathbb{R} \mapsto \mathbb{R}$, then

$$\int_a^b f(X_t) \; dX_t = \int_a^b f(X_t)\dot{X}_t \; dt, \tag{6}$$

$$\int_a^b Y_t \; dX_t = \int_a^b Y_t\dot{X}_t \; dt \tag{7}$$

Since we now know how we compute these types of integrals, let us see some examples on computing signatures.

**Example 0.1.** $X_t = \{X_t^1\} = \{t^3\}$, where $t \in [a, b] = [0, 1]$.

$$S(X)_{0,1}^1 = \int_{0<t<1} dX_t^1 = \int_0^1 3t^2 dt = X_1^1 - X_0^1 = 1^3 - 0^3 = 1$$

$$S(X)_{0,1}^{1,1} = \int_{0<t_1<t_2<1} dX_{t_1}^1 dX_{t_2}^1 = \int_0^1 \left( \int_0^{t_2} 3t_1^2 dt_1 \right) dt_2 = 1/4$$

So we will compute for $\{1, 11, 111, \dots\}$.

**Example 0.2.** Let us define $X_t = \{X_t^1, X_t^2\} = \{5 + t^2, (t^2 + 1)^2\}$, where $t \in [a, b] = [1, 5]$. So we know $dX_t = \{dX_t^1, dX_t^2\} = \{2tdt, 4t(t^2 + 1)dt\}$. If we compute the iterated integrals we get:

$$S(X)_{1,5}^1 = \int_{1<t<5} dX_t^1 = \int_1^5 2tdt = X_5^1 - X_1^1 = 30 - 6 = 24$$

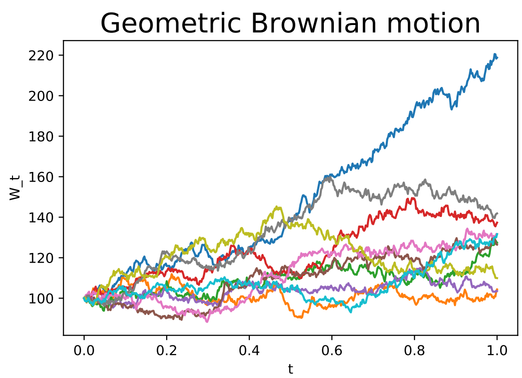
$$S(X)^2_{1,5} = \int_{1<t<5} dX^2_t = \int_1^5 4t(t^2+1)dt = X^2_5 - X^2_1 = 676 - 4 = 672$$

$$S(X)^{1,1}_{1,5} = \int_{1<t_1<t_2<5} dX^1_{t_1} dX^1_{t_2} = \int_1^5 \left( \int_1^{t_2} 2t_1 dt_1 \right) 2t_2 dt_2 = \int_1^5 (\ t_2^2 - 1) 2t_2 dt_2 = 288$$

$$S(X)^{1,2}_{1,5} = \int_{1<t_1<t_2<5} dX^1_{t_1} dX^2_{t_2} = \int_1^5 \left( \int_1^{t_2} 2t_1 dt_1 \right) 4t_2(t_2^2+1) dt_2 = \int_1^5 (\ t_2^2-1) 4t_2(t_2^2+1) dt_2$$

and so on, since the superscripts run along the set of all multi-indexes, we will have an infinite combination$\{1, 2, 11, 12, 21, 22, 111, 112, 121 \ldots\}$.

Eventually, we would like to deal with rough non-smooth functions, like the Geometric Brownian motion (following picture), which is significant when it comes to stock prices.




# References

[1] Adeline Fermanian: Signature and statistical learning,
`https://afermanian.github.io/docs/thesis_fermanian.pdf`

[2] Ilya Chevyreva and Andrey Kormilitzin: A Primer on the Signature Method in Machine Learning,
`https://arxiv.org/pdf/1603.03788.pdf`

[3] Tikosi Kinga: python code for Geometric Brownian motion,