Rough path signature methods for time series data

Teo Deveney

University of Bath

ITT7 Presentation

October 12, 2017

Approximation by signatures

Any 'path' can be written

$$y_t = \left(\sum_0^\infty A^n S_t^n \right) y_0$$

where the signatures S_t are

$$S_t := \sum_0^\infty \int_{u_1 \leq \cdots \leq u_n} \ldots \int d\gamma_{u_1} \otimes \cdots \otimes d\gamma_{u_n} \in \bigoplus_n^\infty E^{\otimes n}.$$

We can show that the signatures decay with n at rate

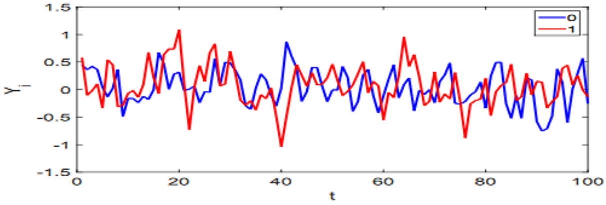
$$S_t^n \leq \frac{|\gamma_t|^n}{n!}$$

So that the uniform error in the truncated sum is

$$\left|\left| y_t - \sum_0^{N-1} A^n \int_{u_1 \leq \cdots \leq u_n} \ldots \int d\gamma_{u_1} \otimes \cdots \otimes d\gamma_{u_n} \right|\right| \leq \left(\sum_N^\infty \frac{||A||^n |\gamma_t|^n}{n!} \right) ||y_0||.$$

Idea:

Suppose we observe some discrete data $Y_1, \ldots, Y_N$. We can convert this into a path via several methods. E.g linear splines, cumulative splines, lead-lag interpolation etc...



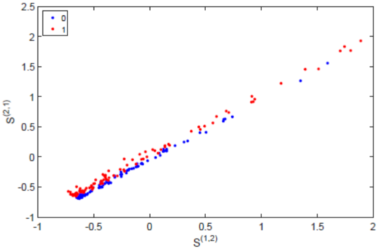
Then the signatures of these paths can be calculated and used for inference.

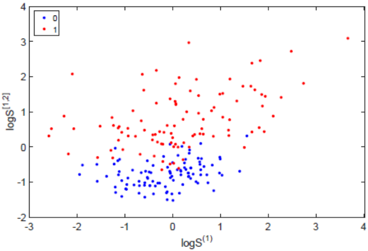
Fast convergence of the signature approximations means that only few terms of the signature are required to be calculated.

Samples of 500 time series are drawn from ARMA(1,1) processes

$$0 : Y_t - 0.4Y_{t-1} = 0.5 + \epsilon_t 0.5\epsilon_{t-1}$$
$$1 : Y_t - 0.8Y_{t-1} = 0.5 + \epsilon_t 0.7\epsilon_{t-1}$$

Transforming into paths and plotting their signatures





A linear classifier seperates these with 90% accuracy using only the second level of the signature!

Many things can be seen as paths. Signature methods were used to win a worldwide competition on Chinese handwriting recognition.