FPSAC 2020

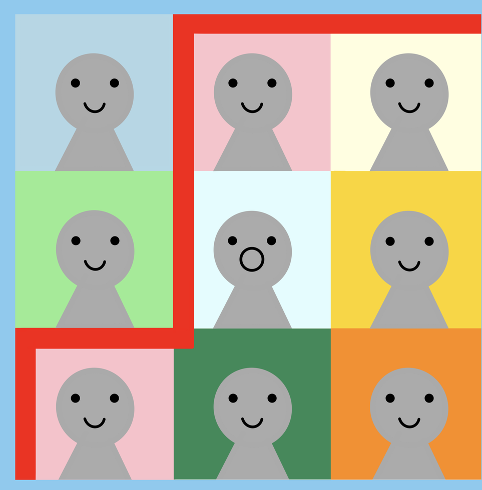
Iterated-sums signature, quasysymmetric functions and time series analysis

Joscha Diehl[1], Kurusch Ebrahimi-Fard[2], Nikolas Tapia[3,4]

[1]Universität Greifswald
[2]Norwegian University of Science and Technology
[3]Weierstrass Institute
[4]TU Berlin

6

We survey results on a recently defined character on the quasi-shuffle algebra, termed *iterated-sums signature*, in the context of time series analysis and dynamic time warping. Algebraically, it relates to quasi-symmetric functions as well as quasi-shuffle algebras.



Quasi-shuffle algebra

Let $A = \{1, \ldots, d\}$. On $H := T(S(A))$ define

$$ua \star vb = (u \star vb)a + (ua \star v)b + (u \star v)[ab].$$

Example:

$$1 \star 2 = 12 + 21 + [12]$$
$$1 \star 23 = 123 + 213 + 231 + [12]3 + 2[13]$$

Quasisymmetric functions

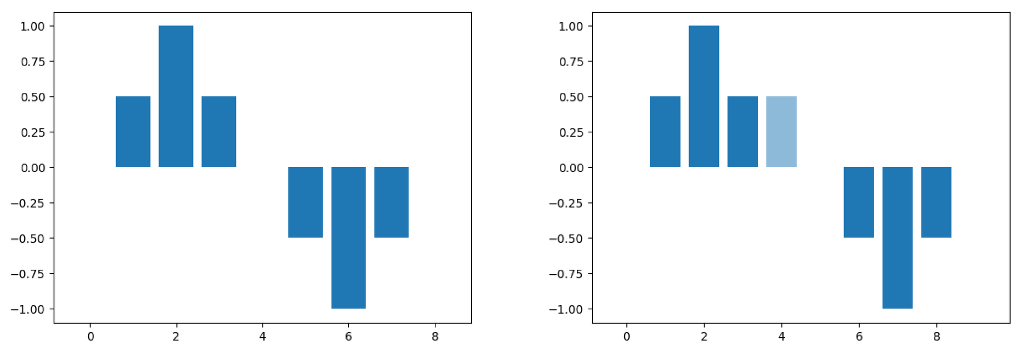
A formal power series $P \in \mathbb{R}\langle X_1, X_2, \ldots \rangle$ is quasisymmetric if the coefficients of the monomials

$$X_{i_1}^{\alpha_1} \cdots X_{i_n}^{\alpha_n} \text{ and } X_{j_1}^{\alpha_1} \cdots X_{j_n}^{\alpha_n}$$

are equal, whenever $i_1 < \cdots < i_n$ and $j_1 < \ldots < i_n$. Examples include

$$M_{(1,2)} = \sum_{i_1 < i_2} X_{i_1} X_{i_2}^2, \quad P_3 = \sum_i X_i^3.$$

Time warping invariance



The iterated-sums signature

We consider time series $\mathbf{x} = (\mathbf{x}_0, \ldots, \mathbf{x}_N) \in (\mathbb{R}^d)^N$. Define the increments $\delta\mathbf{x}_k := \mathbf{x}_{k+1} - \mathbf{x}_k$ for $k = 0, \ldots, N-1$.

Theorem
Let $F : (\mathbb{R}^d)^N \to \mathbb{R}$ be a polynomial map, invariant to time warping and space translations. Then F is realized as a quasisymmetric function on the increments of $\mathbf{x}$.

We extend the coordinate map $i \mapsto x^i$ to $S(A)$ as an algebra morphism, that is,

$$\mathbf{x}^{[i_1 \cdots i_k]} := \mathbf{x}^{i_1} \cdots \mathbf{x}^{i_k}.$$

Definition
Let $\mathbf{x}$ be a time series and $0 \leq n < m \leq N$. We define a map $\mathrm{ISS}(\mathbf{x})_{n,m} : H \to \mathbb{R}$ by

$$\langle \mathrm{ISS}(\mathbf{x})_{n,m}, u_1 \cdots u_k \rangle = \sum_{n \leq j_1 < \cdots < j_k < m} \delta\mathbf{x}_{j_1}^{u_1} \cdots \delta\mathbf{x}_{j_k}^{u_k}$$

Theorem
The iterated-sums signature map satisfies

1. *Chen's relation: for all $0 \leq n < r < m \leq N$,*

$$\langle \mathrm{ISS}(\mathbf{x})_{n,m}, i_1 \cdots i_k \rangle = \sum_{j=0}^{k} \langle \mathrm{ISS}(\mathbf{x})_{n,r}, i_1 \cdots i_j \rangle \langle \mathrm{ISS}(\mathbf{x})_{r,m}, i_{j+1} \cdots i_k \rangle.$$

2. *the quasi-shuffle relations:*

$$\langle \mathrm{ISS}(\mathbf{x})_{n,m}, u \star v \rangle = \langle \mathrm{ISS}(\mathbf{x})_{n,m}, u \rangle \langle \mathrm{ISS}(\mathbf{x})_{n,m}, v \rangle$$

Quasi-shuffle morphisms

For a formal diffeomorphism $f \in t\mathbb{R}[[t]]$, $f = \sum_n c_n t^n$ define a linear map $\Psi_f : H \to H$ by

$$\Psi_f(u_1 \cdots u_k) := \sum_{I \in C(k)} c_{i_1} \cdots c_{i_p} I[u_1 \cdots u_k]$$

where $I = (i_1, \ldots, i_p) \in C(k)$ is a composition of k of length p and

$$I[u_1 \cdots u_k] := [u_1 \cdots u_{i_1}][u_{i_1+1} \cdots u_{i_1+i_2}] \cdots [u_{i_1+\cdots+i_{p-1}+1} \cdots u_k].$$

Definition
Let $\theta \in \mathbb{R}$ and consider $f_\theta(t) := \frac{1}{\theta}(e^{\theta t} - 1)$. Define the map $\mathrm{ISS}(\mathbf{x})^\theta_{n,m} : H \to \mathbb{R}$ by

$$\mathrm{ISS}(\mathbf{x})^\theta_{n,m} := \mathrm{ISS}(\mathbf{x})_{n,m} \circ \Psi_{f_\theta}.$$

Relation to stochastic integration

Stochastic integrals are defined as the limit in probability of Riemann sums,

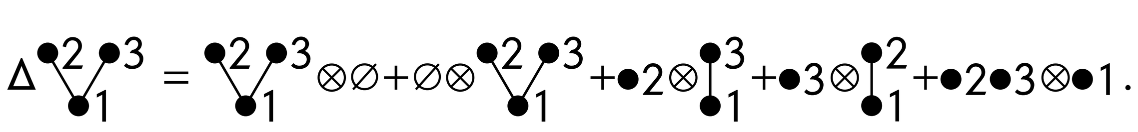
$$\int_0^1 X_t \, \mathrm{d}Y_t \approx \sum_{j=0}^{n} X_{t_j}(Y_{t_{j+1}} - Y_{t_j}).$$

The ISS contains these sums and also provides an alternative description of Itô calculus at the discrete level. Indeed, the quasi-shuffle relations recover Itô's formula

$$\begin{aligned}(X_1 - X_0)(Y_1 - Y_0) &= \left(\sum_{i=0}^{n} \delta\mathbf{x}_i\right)\left(\sum_{i=0}^{n} \delta\mathbf{y}_i\right) \\ &= \sum_{j_1 < j_2} \delta\mathbf{x}_{j_1}\delta\mathbf{y}_{j_2} + \sum_{j_2 < j_1} \delta\mathbf{y}_{j_2}\delta\mathbf{x}_{j_1} + \sum_j \delta\mathbf{x}_j\delta\mathbf{y}_j \\ &\approx \int_0^1 (X_t - X_0)\, \mathrm{d}Y_t + \int_0^1 (Y_t - Y_0)\, \mathrm{d}X_t + \langle X, Y \rangle_1\end{aligned}$$

The Connes-Kreimer Hopf algebra

Here it is denoted by $(H_{\mathrm{CK}}, \cdot, \Delta)$. It is linearly spanned by trees and forests. Its product is the disjoint union of forests. The coproduct is given in terms of *admissible cuts*.



There is an isomorphism between the Hopf subalgebra $\tilde{H}_{\mathrm{CK}}$ formed by ladder trees and the quasi-shuffle Hopf algebra H. We denote this map by $F : H \to \tilde{H}_{\mathrm{CK}}$.

Moments and cumulants

When the time series under considerations is a random sequence $\mathbf{x}$, the ISS is itself a random map on the quasi-shuffle algebra.

Definition
The expectation map of ISS is the linear map $\mu_{\mathbf{x}} : H \to \mathbb{R}$ given by

$$\langle \mu_{\mathbf{x}}, u_1 \cdots u_k \rangle = \mathbb{E}[\langle \mathrm{ISS}(\mathbf{x})_{0,N}, u_1 \cdots u_k \rangle] = \mathbb{E}\left[\sum_{j_1 < \cdots < j_k} \delta\mathbf{x}_{j_1}^{u_1} \cdots \delta\mathbf{x}_{j_k}^{u_k}\right]$$

There exists a unique map $\kappa_{\mathbf{x}} : H \to \mathbb{R}$ such that $\mu_{\mathbf{x}} = \exp_*(\kappa_{\mathbf{x}})$ where the exponential is with respect to the convolution product of linear maps on H, that is

$$\mu_{\mathbf{x}} = \sum_{n=1}^{\infty} \frac{1}{n!} \kappa_{\mathbf{x}}^{*n} \quad \kappa_{\mathbf{x}} = \sum_{n=0}^{\infty} \frac{(-1)^{n-1}}{n} \mu_{\mathbf{x}}^{*n}.$$

We define $\tilde{\mu}_{\mathbf{x}} := \mu_{\mathbf{x}} \circ F^{-1}$, $\tilde{\kappa}_{\mathbf{x}} := \kappa_{\mathbf{x}} \circ F^{-1}$.

Proposition
This sum can be expressed in terms of linearly ordered partitions:

$$\kappa_x^{w_1 \star \cdots \star w_n} = -\sum_{m=1}^{|t_1 \cdots t_n|} \frac{(-1)^m}{m} \sum_{\pi \in OP_m} \tilde{\mu}'_x(t_{\pi_1}) \cdots \tilde{\mu}'_x(t_{\pi_m}).$$

The sum on the right-hand side runs over ordered partition with m blocks. The computation of order m, $\pi := \{\pi_1, \cdots, \pi_m\}$ and its blocks π_i is obtained by partitioning $I \cup J = [n]$ into two subsets, where $I \neq \emptyset$. Then consider the corresponding subsets of trees, $t_I = t_{i_1} \cdots t_{i_p}$ and $t_J = t_{j_{p+1}} \cdots t_{j_n}$. Apply to each tree in t_I a single non-empty cut. This produces a tensor product of forests $t'_I \times t''_I$. Define the set $\pi_1 := \{t'_I\}$. and the forest $t''_I t_J$. Repeat the procedure to define the blocks π_2, π_3, up to π_m for $1 \leq m \leq |t_1 \cdots t_n|$.

Presenter

Nikolas Tapia
E tapia@wias-berlin
Mohrenstr. 39
10117 Berlin
Germany

Funding

NT BMS Excelence Cluster MATH+ EF1
Project EF1-5 On Robustness of Deep Neural Networks
KEF Research Council of Norway
Project 302831 Computational Dynamics and Stochastics on Manifolds (CODYSMA).

References

[1] P. Bonnier and H. Oberhauser, "Signature cumulants, ordered partitions, and independence of stochastic processes". (2019) arXiv: 1908.06496 [math.PR]
[2] J. Diehl, K. Ebrahimi-Fard and N. Tapia, "Time-warping invariants of multidimensional time series", Acta Appl. Math. (2020).
[3] M. E. Hoffman and K. Ihara, "*Quasi-shuffle products revisited*", J. Algebra **481** (2017). pp. 293–326.