# Симулирование случайных процессов с использованием сигнатурных методов

Мащенко Кирилл

Московский государственный университет им. М.В. Ломоносова,
механико-математический факультет,
кафедра теории вероятностей

Научный руководитель: к.ф.-м.н., Житлухин М.В.

Москва
2022 год

# Определения

## Путь

*Путь в* $\mathbb{R}^d$ – это непрерывное отображение $X$ из некоторого интервала $[a, b]$ в $\mathbb{R}^d$. Чтобы подчеркнуть зависимость от времени будем использовать обозначение $X_t = X(t) : [a, b] \mapsto \mathbb{R}^d$.

## Повторный интеграл от пути $X$

Для любого $k \geq 1$ и набора индексов $i_1, \ldots, i_k \in \{1, \ldots, d\}$ определим

$$S(X)_{a,t}^{i_1,\ldots,i_k} = \int\limits_{a<s<t} S(X)_{a,s}^{i_1,\ldots,i_{k-1}} dX_s^{i_k} = \int\limits_{a<t_k<t} \ldots \int\limits_{a<t_1<t_2} dX_{t_1}^{i_1} \ldots dX_{t_k}^{i_k}.$$

Величина $S(X)_{a,t}^{i_1,\ldots,i_k}$ называется $k$-кратным повторным интегралом от пути $X$ по индексам $i_1, \ldots, i_k$.

# Определения

## Сигнатура пути

*Сигнатурой* пути $X : [a, b] \mapsto \mathbb{R}^d$ называется бесконечный набор $S(X)_{a,b}$ всех повторных интегралов от $X$:

$$S(X)_{a,b} = (1,\ S(X)^1_{a,b},\ \dots,\ S(X)^d_{a,b},\ S(X)^{1,1}_{a,b},\ S(X)^{1,2}_{a,b},\ \dots),$$

где первый элемент сигнатуры (соответствующий пустому индексу) по определению считается равным 1, а верхние индексы остальных элементов пробегают набор всевозможных мульти-индексов

$$W = \{(i_1, ..., i_k) \mid k \geq 1, i_1, ..., i_k \in \{1, ..., d\}\}.$$

# Теоремы о сигнатурах

## Теорема о шафл-произведении

Для любого пути $X : [a, b] \mapsto \mathbb{R}^d$ и мульти-индексов $I = (i_1, \dots, i_k)$ и $J = (j_1, \dots, j_m)$ , $i_1, \dots, i_k, j_1, \dots, j_m \in \{1, \dots, d\}$ верно равенство

$$S(X)^I_{a,b} S(X)^J_{a,b} = \sum_{K \in I ш J} S(X)^K_{a,b}.$$

## Тождество Чена

Пусть $a < b < c$ и $X : [a, c] \mapsto \mathbb{R}^d$ . Тогда для любых $i_1, \dots, i_k \in W$ выполнено равенство

$$S(X)^{i_1,\dots,i_k}_{a,c} = \sum_{m=0}^{k} S(X)^{i_1,\dots,i_m}_{a,b} S(X)^{i_{m+1},\dots,i_k}_{b,c}.$$

## Независимость от начальной точки

Рассмотрим путь $X : [a, b] \mapsto \mathbb{R}^d$ и $h \in \mathbb{R}^d$. Пусть путь $Y : [x, y] \mapsto \mathbb{R}^d$ имеет вид $Y_t = X_t + h$. Тогда

$$S(X)_{a,b} = S(Y)_{a,b}.$$

## Независимость от репараметризации времени

Рассмотрим путь $X : [a, b] \mapsto \mathbb{R}^d$ и биективную непрерывную неубывающую функцию $\psi : [x, y] \mapsto [a, b]$. Пусть путь $Y : [x, y] \mapsto \mathbb{R}^d$ имеет вид $Y_t = X_{\psi_t}$. Тогда

$$S(X)_{a,b} = S(Y)_{x,y}.$$

## Однозначное определение пути по его сигнатуре

Для непересекающихся путей по сигнатуре можно полностью определить все точки, через которые пройдёт путь, и порядок их обхода.

# Логарифмические сигнатуры

## Связь сигнатур и формальных степенных рядов

Сигнатура может быть “закодирована” как элемент пространства формальных степенных рядов:

$$S(X)_{a,b} = \sum_{k=0}^{\infty} \sum_{i_1,\dots,i_k=1}^{d} S(X)_{a,b}^{i_1,\dots,i_k} e_{i_1}\dots e_{i_k}.$$

## Определение логарифмической сигнатуры

*Логарифмической сигнатурой* пути $X : [a,b] \mapsto \mathbb{R}^d$ называется формальный степенной ряд

$$\log S(X)_{a,b} = \sum_{n=1} \frac{(-1)^n}{n} (\mathbf{1} - S(X)_{a,b})^{\otimes n}.$$

# Сферы применения генеративных моделей

- Анонимизация данных (финансовые и медицинские данные).
- Заведомо маленькое количество данных (ограничение доступа к данным).
- Тестирование стратегий (необходимы другие данные во избежание переобучения модели).

# Работа Buhler, Horvath, Lyons, Arribas, Wood (2020)

- Разработана генеративная модель для финансовых временных рядов, основанная на сигнатурах и стабильно работающая на маленьком количестве данных.
- Показывается различие между классическими подходами и подходом машинного обучения.
- Предложен тест для проверки схожести распределения сгенерированных и исходных путей, основанный на сигнатурных методах.
- Предложен алгоритм обращения логарифмической сигнатуры.

# Алгоритм обращения логарифмической сигнатуры

- Ищем произвольный путь $X : [a, b] \mapsto \mathbb{R}^d$, соответствующий известной логарифмической сигнатуре.
- Требуется найти кусочно-линейный путь $\widehat{X}$, наилучшим образом приближающий путь $X$, с фиксированным шагом $\Delta t$ и значениями приращений каждой координаты из множества $\{-n_i \cdot h_i, \dots, n_i \cdot h_i\}$ с шагом $h_i$.

# Алгоритм обращения логарифмической сигнатуры

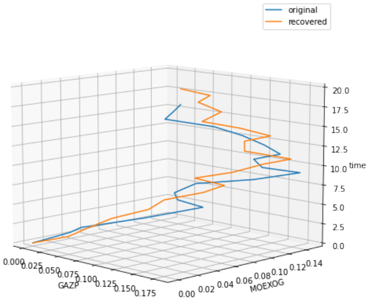
## Теорема

Пусть даны фиксированные числа $h_i \in \mathbb{R}_+, n_i \in \mathbb{N}, \Delta t \in \mathbb{R}_+$. Пусть $X : [a, b] \mapsto \mathbb{R}^d$ - непрерывный и кусочно-дифференцируемый путь, у которого $\forall i : |X_i'(t)| \leq \frac{h_i \cdot n_i}{\Delta t}$. Тогда существует кусочно-линейный путь $\widehat{X} : [a, b] \mapsto \mathbb{R}^d$ с приращениями по каждой координате $i$ со значениями из множества $\{-n_i \cdot h_i, \ldots, 0, \ldots, n_i \cdot h_i\}$ и "изломами" в точках $k \cdot \Delta t,\ k \in \mathbb{Z}$ такой, что $\forall i : \sup\limits_{t \in \mathbb{R}_+} |X_i(t) - \widehat{X}_i(t)| \leq h_i \cdot (2n_i + \frac{1}{2})$.

## Замечание

Если провести достаточное количество итераций генерирования приращений каждой координаты пути $\widehat{X}_i$ из арифметического распределения со значениями из соответствующего множества $\{-n_i \cdot h_i, \ldots, n_i \cdot h_i\}$, то с вероятностью, равной единице, будет получен оптимальный путь $\widehat{X}$

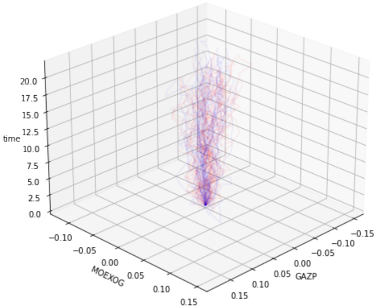
# Алгоритм обращения логарифмической сигнатуры




Сравнение оригинального пути и пути, восстановленного по его логарифмической сигнатуре.

# Генерация рынка

Приведём пример симуляции совместных траекторий процессов цен акций компании ПАО "Газпром" (код GAZP) и отраслевого индекса нефти и газа Московской биржи (код MOEXOG): красным цветом нарисованы сгенерированные пути, синим - исходные.




Оригинальные и сгенерированные совместные траектории процессов цен активов GAZP и MOEXOG.

Спасибо за внимание!