Санкт-Петербургский государственный университет

Карагодин Никита Алексеевич

Выпускная квалификационная работа

Энергетически эффективные аппроксимации случайных полей

Уровень образования: магистратура
Направление: 01.04.01 «Математика»
Основная образовательная программа: ВМ.5832.2021 «Современная математика»

Научный руководитель:
д.ф.-м.н., профессор
Лифшиц Михаил Анатольевич

Рецензент:
д.ф.-м.н., профессор
Белопольская Яна Исаевна

Санкт-Петербург
2023

Аннотация

Изучено поведение энергетически эффективных аппроксимаций случайных функций. Первая часть посвящена доказательству закона больших чисел о сходимости ошибки для энергетически эффективной аппроксимации броуновского листа в среднем квадратическом и почти наверное. Вторая часть посвящена изучению энергетически эффективных аппроксимаций случайных процессов со стационарными приращениями.

Содержание

1 Введение

Для случайного процесса $B(t), t \in [0, T]$ энергетически эффективной аппроксимацией называют функцию $f(t)$, которая в некотором смысле близка к $B(t)$, но при этом обладает низкой энергией. В общем виде, такую задачу можно описать следующим образом

$$f_{opt,Q,\mathcal{E}}(t) = \arg\min_{f} \int_0^T \left(\mathcal{E}[f](t) + Q(f(t) - B(t))\right) dt, \tag{1}$$

где Q – некоторая функция штрафа, отвечающая за близость f и B, а $\mathcal{E}$ – функционал энергии, контролирующий качество самой аппроксимирующей функции f. Интегральное выражение будем в дальнейшем называть *ошибкой аппроксимации*. Стоит отметить, что на ответ влияет выбор пространства, в котором ищется f. Обычно оно выбирается так, чтобы определение было корректно. Например, когда функционал энергии $\mathcal{E}[f]$ является квадратом некоторой линейной комбинации производных f, можно рассмотреть f из соответствующего пространства Соболева $f \in \mathbb{W}_2^M[0, T]$, где M – максимальный порядок производных в $\mathcal{E}[f]$. Тем не менее, у выбора пространства тоже есть важный смысл – он характеризует то, какие аппроксимации мы бы действительно хотели получить. Этот аспект нужно учитывать при выборе постановки задачи, и мы ещё к нему вернемся при обсуждении результатов данной работы.

Самыми простыми для изучения являются функция штрафа $Q(x) = x^2$ и функционал энергии $\mathcal{E}[f](t) = |f'(t)|^2$, поэтому они используются в классической постановке вопроса. В такой задаче сочетаются расстояние до $B(t)$ в среднем квадратическом и соболевская норма(энергия), то есть

$$f_{opt}(t) = \arg\min_{f \in \mathbb{W}_2^1[0,T]} \int_0^T \left(f'(t)^2 + (f(t) - B(t))^2\right) dt.$$

В таком случае методами вариационного исчисления можно найти явное выражение для функции f_{opt}, которое, конечно, зависит от B. Однако, нас интересует не только сама функция, но и итоговая ошибка аппроксимации, а точнее асимптотическое поведение этой ошибки при больших T.

Помимо штрафной функции и энергетического функционала, важной характеристикой задачи является то, к какому классу принадлежит процесс $B(t)$. В работе [4] изучены энергетически эффективные аппроксимации процессов со стационарными приращениями в широком смысле. Рассматривается классическая постановка задачи, приведённая выше. Показано, что с пренебрежимой разницей в качестве можно взять оптимальную аппроксимацию, обладающую стационарными приращениями в широком смысле, то есть сохранить структурное свойство изначального процесса. К тому же, доказано, что асимптотика средней ошибки аппроксимации при $T \to \infty$ имеет вид $\mathcal{C}T$ для некоторой константы $\mathcal{C}$, которую можно явно найти при помощи спектральной плотности процесса. Для гауссовских процессов и процессов Леви результат дополнен сходимостью нормированной ошибки почти наверное и в L^1 к явно найденному пределу.

В работе [7] изучаются энергетически оптимальные аппроксимации $h(t)$ винеровского процесса $W(t)$, для которых $\mathcal{E}[h] = |h'|^2$ и $Q(x) = 0, x \le r, Q(x) = +\infty, x > r$. Иными словами, на аппроксимацию задаётся ограничение, что она принадлежит некоторому окну ширины r вокруг W, то есть $W(t) - r \le h(t) \le W(t) + r, 0 \le t \le T$, и при этом энергия $\int_0^T h'(t)^2\,dt$ минимальна. Ключевой особенностью здесь является негладкое ограничение на аппроксимацию, что сильно усложняет задачу. Получаемая функция хорошо известна в вариационном анализе и статистике и называется натянутой струной. В работе показано, что асимптотически, при $T \to \infty$, натянутая струна тратит $r^{-2}\mathcal{C}^2$ энергии на единицу времени.

В работе [6] продолжено исследование работы [7] и изучена энергия натянутой струны, сопровождающей случайное блуждание. В обеих работах ширина окна $r = r_T$ постоянна для всего рассматриваемого отрезка, но меняется с ростом T.

В работе [11] результаты работ [6] и [7] обобщены на случай полосы переменной ширины $r = r_t$.

В работах [4] и [13] для стационарного процесса вопрос аппроксимации с оптимальной энергией рассматривается как обобщение задачи прогнозирования, где кроме качества предсказания также играют роль свойства прогнозирующей функции. В таком случае нас уже интересует не просто аппроксимация, но *адаптивная аппроксимация*, которой доступны лишь данные о процессе из прошлого, но не из будущего. В данных работах изначально ставится вопрос оптимизации асимптотики ошибки

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \left(|f(t) - B(t)|^2 + \left| \sum_{m=0}^{M} \ell_m f^{(m)}(t) \right|^2 \right) dt,$$

где $f^{(m)}$ – производная m-го порядка, а ℓ_m – фиксированные комплексные коэффициенты. Однако, поскольку в стационарном случае зачастую применима эргодическая теорема, авторы сразу переходят к изучению задачи в форме

$$\mathbb{E}\,|f(0) - B(0)|^2 + \mathbb{E} \left| \sum_{m=0}^{M} \ell_m f^{(m)}(0) \right|^2 \searrow \min_f .$$

2 Аппроксимации броуновского листа

2.1 Постановка задачи

В задачах аппроксимации большинство результатов посвящено исследованиям различного рода случайных процессов и разным определениям качества аппроксимации. В данной работе мы переходим к изучению аппроксимаций случайных полей. Поскольку винеровский процесс является одним из самых известных и хорошо изученных случайных процессов, мы начнём с рассмотрения аппроксимаций одного из его многопараметрических обобщений – броуновского листа. Одним из ключевых вопросов является выбор определения качества аппроксимации, который непосредственно связан с тем, из какого пространства мы хотим видеть аппроксимации. Здесь мы используем популярный подход из машинного обучения.

Для гауссовской случайной функции с ковариацией $k(x,y)$ её воспроизводящим ядром называют минимальное гильбертово пространство H функций, такое, что при всех y функция $t \to k(t,y)$ лежит в H, со скалярным произведением задаваемым правилом $(k(t,x), k(t,y))_H = k(x,y)$.

Классической задачей машинного обучения является задача поиска целевой функции f в модели $y = f(x) + \xi$, где ξ является шумом, с данным набором данных (x_i, y_i) и метод KRR (от Kernel Ridge Regression) (см. [1] и [8]) ищет функцию в целевом пространстве H, которая оптимизирует функционал риска с регуляризацией

$$f_{opt} = \arg\min_{f \in H} \left(\frac{1}{n} \sum_{i=1}^{n} (f(x_i) - y_i)^2 + \lambda^2 \|f\|_H^2 \right).$$

Здесь первая часть представляет из себя функцию потерь, выражающую близость выбранной функции к реальной, а вторая является регуляризацией, необходимой, чтобы избежать переобучения. Одним из важнейших свойств нормы функции в воспроизводящем ядре является то, что она в некотором смысле контролирует гладкость функции – чем меньше $\|f\|_H$, тем более гладкой является функция. Поэтому константа λ^2 отвечает за гладкость предсказательной функции. Метод KRR не вероятностный, однако он очень тесно связан с байесовским машинным обучением, основанном на гауссовских процессах (см. [5]).

Предлагается в качестве энергии аппроксимации положить её норму в ядре исследуемого поля. Такой подход к определению аппроксимации можно рассматривать как экстремальный случай алгоритма KRR, когда в тренировочных данных известны все значения (x_i, y_i) и они имеют вид $(x_i, B(x_i))$, то есть настоящая предсказываемая функция является траекторией броуновского листа. Тогда оптимизируемый функционал превращается в

$$f_{opt} = \arg\min_{f \in H} \left(\int_0^T (f(x) - B(x))^2 \, dx + \lambda^2 \|f\|_H^2 \right).$$

В частности, исследуемая ошибка будет минимальной возможной ошибкой, которая может получиться при использовании KRR с воспроизводящим ядром броуновского листа.

Дадим строгое описание исследумой задачи. Пусть $B(t_1, \ldots, t_d)$ – броуновский лист в $\mathbb{R}_+^d$, то есть центрированное гауссовское случайное поле с ковариационной функцией

$$\mathrm{Cov}(B(t_1, \ldots, t_d), B(s_1, \ldots, s_d)) = \min(t_1, s_1) \cdot \ldots \cdot \min(t_d, s_d).$$

Обозначим $\bar{t} = (t_1, \ldots, t_d)$ и $[0, \bar{t}] = [0, t_1] \times \ldots \times [0, t_d]$. В новых обозначениях

$$\mathrm{Cov}(B(\bar{t}), B(\bar{s})) = \lambda^d([0, \bar{t}] \cap [0, \bar{s}]).$$

Хорошо известно, что для броуновского листа, суженного на $[0,\overline{T}]$, ядром является пространство функций

$$H([0,\overline{T}]) = \left\{ f(\bar{t}), \bar{t} \in [0,\overline{T}] \, : \, f(\bar{t}) = \int_{[0,\bar{t}]} \ell(\bar{u}) d\bar{u}, \; \ell(\bar{u}) \in L^2(\mathbb{R}_+^d) \right\}.$$

Иными словами, это функции, чья обобщённая смешанная производная $\frac{\partial^d f}{\partial t_1 \ldots \partial t_d}$ квадратично интегрируема. Скалярное произведение в ядре определяется как

$$(f_1(\bar{t}), f_2(\bar{t}))_{H([0,\overline{T}])} = \int_{[0,\overline{T}]} \frac{\partial f_1}{\partial t_1 \ldots \partial t_d}(\bar{t}) \frac{\partial f_2}{\partial t_1 \ldots \partial t_d}(\bar{t}) \, d\bar{t}.$$

Нас интересует аппроксимация броуновского листа функцией из ядра, которая при этом имеет относительно малую норму (энергию). Объединим эти два требования в один общий функционал, контролируя важность одной из компонент при помощи параметра λ. Ошибкой аппроксимации назовём

$$\mathcal{E}_{\overline{T}}(f, B) = \min_{f \in H([0,\overline{T}])} \int_{[0,\overline{T}]} \left(|f(\bar{t}) - B(\bar{t})|^2 + \lambda^2 \left(\frac{\partial^d f}{\partial t_1 \ldots \partial t_d} \right)^2 (\bar{t}) \right) d\bar{t}.$$

Отметим, что решение такой задачи с общим параметром λ позволяет разделить компоненты при помощи метода множителей Лагранжа и решить две дополнительные задачи – оптимизацию близости аппроксимации при данном ограничении на энергию и минимизацию энергии при данном ограничении на близость.

2.2 Основной результат

Теорема 1 *С ростом объёма параметрического множества $T = T_1 \ldots T_d$ нормированная ошибка аппроксимации сходится в L^2*

$$\frac{\mathcal{E}_{\overline{T}}(f, B)}{T(\ln T)^{d-1}} \xrightarrow{L^2} \frac{\lambda}{2\pi^{d-1}(d-1)!} \qquad \text{при } T \to \infty.$$

Если при этом существует такое $c > 0$, что $\min(T_1, \ldots, T_d) > c$, то имеет место сходимость почти наверное

$$\frac{\mathcal{E}_{\overline{T}}(f, B)}{T(\ln T)^{d-1}} \xrightarrow{\text{п.н.}} \frac{\lambda}{2\pi^{d-1}(d-1)!} \qquad \text{при } T \to \infty.$$

Доказательство:

Сначала оценим асимптотику среднего и дисперсии ошибки. Воспользуемся самоподобием броуновского листа

$$B(s_1, \ldots, s_d) \overset{\mathcal{D}}{=} (T_1 \ldots T_d)^{-1/2} B(s_1 T_1, \ldots, s_d T_d),$$

чтобы сделать замену переменной в интеграле $(t_1,\dots,t_d)=(s_1T_1,\dots,s_dT_d)$. Обозначим $T=T_1\dots T_d$. Тогда

$$\begin{aligned}\mathcal{E}_{\overline{T}}(f,B) &\overset{\mathcal{D}}{=} \mathcal{E}^*_{\overline{T}}(g,B)\\ &= \min_{g\in H([0,1]^d)}\int_{[0,1]^d}\left(T^2|g(\bar s)-B(\bar s)|^2+\lambda^2\left(\frac{\partial^d g}{\partial s_1\dots\partial s_d}\right)^2(\bar s)\right)d\bar s.\end{aligned}$$

Функции $e_n(s)=\sqrt{2}\sin((n-1/2)\pi s)$ образуют ортогональную систему в $L^2([0,1])$ из собственных функций ковариационного оператора стандартного броуновского движения, $\gamma_n=((n-1/2)\pi)^{-2}$ – соответствующие e_n собственные числа. Тогда $e_{\bar n}(\bar s)=\prod_i e_{n_i}(s_i)$ образуют ортогоналную систему в $L^2([0,1]^d)$ из собственных функций ковариационного оператора броуновского листа, а $\gamma_{\bar n}=\prod_i\gamma_{n_i}$ являются соответствующими $e_{\bar n}$ собственными числами.

Воспользуемся разложением Кархунена – Лоэва поля B в $[0,1]^d$. Для последовательности гауссовских случайных величин $\omega_{\bar n}$ с $\mathbb{E}\,\omega_{\bar n}=0$ и $\operatorname{Var}\omega_{\bar n}=\gamma_{\bar n}$ имеет место представление

$$B(\bar s)=\sum_{\bar n\in\mathbb{N}^d}\omega_{\bar n}e_{\bar n}(\bar s),\ \ \bar s\in[0,1]^d.$$

В то же время, набор функций $(e_{\bar n})_{\bar n\in\mathbb{N}^d}$ является ортогональным базисом пространства H, причём $\|e_{\bar n}\|_H=\gamma_{\bar n}^{-1/2}$. Поэтому можно разложить функцию g в ряд в пространстве H

$$g(\bar s)=\sum_{\bar n\in\mathbb{N}^d}g_{\bar n}e_{\bar n}(\bar s),$$

и представить норму в виде

$$\|g\|_H^2=\sum_{\bar n\in\mathbb{N}^d}\gamma_{\bar n}^{-1}g_{\bar n}^2.$$

Пространство H непрерывно вложено в $L^2([0,1]^d)$, поэтому то же разложение для g имеет место и в $L^2([0,1]^d)$, откуда

$$\int_0^1\dots\int_0^1 T^2|g(\bar s)-B(\bar s)|^2\,ds_1\dots ds_d=T^2\sum_{\bar n\in\mathbb{N}^d}(g_{\bar n}-\omega_{\bar n})^2.$$

Получается, что общая ошибка записывается в виде

$$\mathcal{E}^*_{\overline{T}}(g,B)=\sum_{\bar n\in\mathbb{N}^d}\left(T^2(g_{\bar n}-\omega_{\bar n})^2+\lambda^2\gamma_{\bar n}^{-1}g_{\bar n}^2\right).$$

Поэтому достаточно оптимизировать каждое слагаемое отдельно по $g_{\bar n}$. Слагаемое с $g_{\bar n}$ квадратично, поэтому принимает минимальное значение при

$$g_{\bar n}=\frac{\omega_{\bar n}}{1+\frac{\lambda^2}{T^2\gamma_{\bar n}}}$$

и это значение равно
$$\frac{\omega_{\bar n}^2}{\frac{\gamma_{\bar n}}{\lambda^2}+\frac{1}{T^2}}.$$
Следовательно, оптимальная функция g имеет вид
$$g_{opt}(\bar s)=\sum_{\bar n\in\mathbb{N}^d}\frac{\omega_{\bar n}}{1+\frac{\lambda^2}{T^2\gamma_{\bar n}}}e_{\bar n}(\bar s),$$
а ошибка равна
$$\mathcal{E}^*_{\overline{T}}(g_{opt},B)=\sum_{\bar n\in\mathbb{N}^d}\frac{\omega_{\bar n}^2}{\frac{\gamma_{\bar n}}{\lambda^2}+\frac{1}{T^2}}.$$
Здесь стоит отметить, что все описанные ряды сходятся в L^2 и почти наверное по теореме Колмогорова о двух рядах. Напомним, что $\omega_{\bar n}^2$ – независимые случайные величины с
$$\mathbb{E}\,\omega_{\bar n}^2=\gamma_{\bar n},\qquad \operatorname{Var}\omega_{\bar n}^2=2\gamma_{\bar n}^2.$$
Поскольку $\gamma_{\bar n}^{-1}=\prod_i\pi^2(n_i-1/2)^2$, имеем
$$\begin{aligned}\mathbb{E}\,\mathcal{E}^*_{\overline{T}}(g_{opt},B)&=\lambda^2\sum_{\bar n\in\mathbb{N}^d}\frac{1}{1+\prod_i\frac{\lambda^{2/d}\pi^2(n_i-1/2)^2}{T_i^2}},\\ \operatorname{Var}\mathcal{E}^*_{\overline{T}}(g_{opt},B)&=2\lambda^4\sum_{\bar n\in\mathbb{N}^d}\frac{1}{\left(1+\prod_i\frac{\lambda^{2/d}\pi^2(n_i-1/2)^2}{T_i^2}\right)^2},\end{aligned}\tag{2}$$
Докажем, что при $T\to\infty$ выполнено
$$\begin{aligned}\mathbb{E}\,\mathcal{E}^*_{\overline{T}}(g_{opt},B)&\sim\frac{\lambda}{2\pi^{d-1}(d-1)!}T(\ln T)^{d-1},\\ \operatorname{Var}\mathcal{E}^*_{\overline{T}}(g_{opt},B)&\sim\frac{\lambda^3}{2\pi^{d-1}(d-1)!}T(\ln T)^{d-1}.\end{aligned}$$
Для удобства обозначим $2r_i=\lambda^{1/d}\pi/T_i$. Тогда оба соотношения следует из следующей леммы.

Лемма 2 *При любых $r_i\geq 0, 1\leq i\leq d$, таких, что $\prod_i r_i\to 0$, выполнены соотношения*
$$\begin{aligned}\sum_{\bar n\in\mathbb{N}^d}\frac{1}{1+\prod_i(4r_i^2(n_i-1/2)^2)}&\sim\frac{\pi}{2(d-1)!}\frac{(-\sum_i\ln r_i)^{d-1}}{\prod_i 2r_i},\\ \sum_{\bar n\in\mathbb{N}^d}\frac{1}{(1+\prod_i(4r_i^2(n_i-1/2)^2))^2}&\sim\frac{\pi}{4(d-1)!}\frac{(-\sum_i\ln r_i)^{d-1}}{\prod_i 2r_i}.\end{aligned}$$

Доказательство: Доказательство будем вести индукцией по d.
База индкуции. При $d = 1$ утверждение верно, так как обе суммы практически являются суммами Дарбу сходящихся интегралов. При $r \to 0$ имеем

$$\sum_{n\in\mathbb{N}} \frac{1}{1+4r^2(n-1/2)^2} \sim \frac{1}{2r}\int_0^\infty \frac{1}{1+x^2}\,dx = \frac{\pi}{4r}$$

и

$$\sum_{n\in\mathbb{N}} \frac{1}{(1+4r^2(n-1/2)^2)^2} \sim \frac{1}{2r}\int_0^\infty \frac{1}{(1+x^2)^2}\,dx = \frac{\pi}{8r}.$$

Индукционный переход. Из предположения индукции следует, что сумма по граничным гиперплоскостям асимптотически мала, а именно при $\prod r_i \to 0$ выполнено

$$\begin{aligned}\sum_{n_1=1,\bar{n}_{-1}\in\mathbb{N}^{d-1}} \frac{1}{1+\prod_i(4r_i^2(n_i-1/2)^2)} &= \sum_{\bar{n}_{-1}\in\mathbb{N}^{d-1}} \frac{1}{1+r_1^2\prod_{i\neq 1}4r_i^2(n_i-1/2)^2}\\ &\sim \frac{\pi}{2(d-2)!}\frac{(-\sum_i \ln r_i)^{d-2}}{1/2\prod_i 2r_i},\end{aligned}$$

Из монотонности функции $\bar{x} \to \frac{1}{1+x_1^2\dots x_d^2}$ по каждой координате в положительном октанте $\mathbb{R}^d_+$ следует, что

$$\sum_{\bar{n}\in\mathbb{N}^d} \frac{1}{1+\prod_i(4r_i^2(n_i-1/2)^2)} \geq \frac{1}{\prod_i 2r_i}\int_{r_1}^\infty dx_1\dots\int_{r_d}^\infty dx_d\left(\frac{1}{1+\prod_i x_i^2}\right)$$

и вместе с этим

$$\sum_{\bar{n}\in(\mathbb{N}\setminus\{1\})^d} \frac{1}{1+\prod_i(4r_i^2(n_i-1/2)^2)} \leq \frac{1}{\prod_i 2r_i}\int_{r_1}^\infty dx_1\dots\int_{r_d}^\infty dx_d\left(\frac{1}{1+\prod_i x_i^2}\right).$$

Мы показали, что разность двух приведённых сумм асимптотически мала по сравнению с предполагаемым ответом. Поэтому достаточно доказать, что при $\prod_i r_i \to 0$ выполнено

$$\int_{r_1}^\infty dx_1\dots\int_{r_d}^\infty dx_d\left(\frac{1}{1+x_1^2\dots x_d^2}\right) \sim \frac{\pi(-\sum_i \ln r_i)^{d-1}}{2(d-1)!}.$$

Аналогично, чтобы разобраться с суммой для дисперсии, нужно проверить, что при $\prod_i r_i \to 0$ выполнено

$$\int_{r_1}^\infty dx_1\dots\int_{r_d}^\infty dx_d\left(\frac{1}{1+x_1^2\dots x_d^2}\right)^2 \sim \frac{\pi(-\sum_i \ln r_i)^{d-1}}{4(d-1)!}.$$

Проинтегрируем оба интеграла по x_d и получим

$$\int_{r_1}^\infty\dots\int_{r_{d-1}}^\infty\left(\frac{\pi}{2}-\arctan(x_1\dots x_{d-1}r_d)\right)\frac{dx_{d-1}\dots dx_1}{x_1\dots x_{d-1}}$$

и

$$\int_{r_1}^{\infty}\dots\int_{r_{d-1}}^{\infty}\frac{1}{2}\left(\frac{\pi}{2}-\arctan(x_1\dots x_{d-1}r_d)-\frac{x_1\dots x_{d-1}r_d}{1+x_1^2\dots x_{d-1}^2r_d^2}\right)\frac{dx_{d-1}\dots dx_1}{x_1\dots x_{d-1}}.$$

Сразу разберёмся с лишней частью во втором интеграле. По предположению индукции при $\prod_i r_i \to 0$ верно

$$\int_{r_1}^{\infty}\dots\int_{r_{d-1}}^{\infty}\frac{r_d dx_{d-1}\dots dx_1}{1+x_1^2\dots x_{d-1}^2r_d^2}\sim\frac{\pi(-\sum_i \ln r_i)^{d-2}}{2(d-1)!}=o\left((-\sum_i \ln r_i)^{d-1}\right).$$

Таким образом, нам остаётся доказать следующее утверждение

Лемма 3 *При $r_i \geq 0, \prod_i r_i \to 0$ выполнено*

$$\int_{r_1}^{\infty}\dots\int_{r_{d-1}}^{\infty}\left(\frac{\pi}{2}-\arctan(x_1\dots x_{d-1}r_d)\right)\frac{dx_{d-1}\dots dx_1}{x_1\dots x_{d-1}}\sim\frac{\pi(-\sum_i \ln r_i)^{d-1}}{2(d-1)!}.$$

Для доказательства нам пригодится явное значение следующего интеграла. Проверку этого равенства можно найти в разделе 2.3.

Предложение 4 *Рассмотрим область $S=\{x\,|\,\forall i\, x_i \geq r_i, x_1\dots x_k \leq A\}$, с параметрами $A \geq \prod r_i$. Тогда имеет место равенство*

$$\int_S \frac{1}{x_1\dots x_k}\,dx_1\dots dx_k=\frac{1}{k!}\left(\ln A-\sum_i \ln r_i\right)^k.$$

Сделаем замену $p = x_1\dots x_{d-1}r_d$. Её якобиан $\frac{dp dx_1\dots dx_{d-2}}{dx_1\dots dx_{d-1}} = x_1\dots x_{d-2}r_d$. Старая область интегрирования $\{\bar{x}\,|\,x_i \geq r_i, 1\leq i\leq d-1\}$ переходит в $\{\bar{x}\,|\,x_i \geq r_i, 1 \leq i \leq d-2, x_1\dots x_{d-2} \leq \frac{p}{r_{d-1}r_d}\}$. Интеграл переписывается в виде

$$\int_{r_1\dots r_d}^{\infty} dp\left(\frac{\pi/2-\arctan(p)}{p}\int_S\frac{dx_1\dots dx_{d-2}}{x_1\dots x_{d-2}}\right),$$

где область S имеет вид $\{\bar{x}\,|\,x_i \geq r_i, x_1\dots x_{d-2} \leq \frac{p}{r_{d-1}r_d}\}$. Согласно предложению 4 внутренний интеграл по области S можно найти явно. Подставляя, получаем

$$\frac{1}{(d-2)!}\int_{r_1\dots r_d}^{\infty}\left(\frac{\pi/2-\arctan(p)}{p}\left(\ln p-\sum_{i=1}^{d}\ln r_i\right)^{d-2}\right)\,dp.$$

Обозначим $r = r_1\dots r_d$. Сделаем замену $p = re^t, dp = pdt$ и получим

$$\frac{1}{(d-2)!}\int_0^{\infty}(\pi/2-\arctan(re^t))t^{d-2}\,dt.$$

Чтобы найти асимптотику при $r \to 0$, обозначим $g(r) = \sqrt{-\ln r}$ и разобьём луч $[0, \infty)$ на три части $[0, -\ln r - g(r)]$, $[-\ln r - g(r), -\ln r + g(r)]$, $[-\ln r + g(r), \infty)$. Интегралы по соответствующим частям обозначим I_1, I_2, I_3.

Поскольку на первом участке $0 \le re^t \le e^{-g(r)} \to 0$, мы можем равномерно по участку записать $\pi/2 - \arctan(re^t) = \pi/2(1 + o(1))$, откуда

$$\begin{aligned} I_1 &= \int_0^{-\ln r - g(r)} (\pi/2 - \arctan(re^t)) t^{d-2}\, dt \\ &= (1 + o(1)) \int_0^{-\ln r - g(r)} \frac{\pi}{2} t^{d-2}\, dt \\ &= (1 + o(1)) \frac{\pi}{2(d-1)} (-\ln r - g(r))^{d-1} \\ &= (1 + o(1)) \frac{\pi}{2(d-1)} (-\ln r)^{d-1}. \end{aligned}$$

На втором участке используем элементарную верхнюю оценку

$$\begin{aligned} I_2 &= \int_{-\ln r - g(r)}^{-\ln r + g(r)} (\pi/2 - \arctan(re^t)) t^{d-2}\, dt \le \pi g(r)(-\ln r + g(r))^{d-2} \\ &= o((-\ln r)^{d-1}). \end{aligned}$$

Поскольку на третьем участке $re^t \ge e^{g(r)} \to \infty$, мы можем равномерно по участку записать $\pi/2 - \arctan(re^t) = (1 + o(1)) \frac{1}{re^t}$, а значит

$$\begin{aligned} I_3 &= \int_{-\ln r + g(r)}^{\infty} (\pi/2 - \arctan re^t) t^{d-2}\, dt \\ &\le \frac{1 + o(1)}{r} \int_{-\ln r + g(r)}^{\infty} e^{-t} t^{d-2}\, dt \\ &= \frac{1 + o(1)}{r} e^{\ln r - g(r)} (-\ln r + g(r))^{d-2} \\ &= o((-\ln r)^{d-1}). \end{aligned}$$

Таким образом, только I_1 вносит значимый вклад в асимптотику. Остаётся подставить $r = r_1 \dots r_d$ и получить требуемое. $\square$

Из доказанной Леммы 2, с учётом равенств (2), получаем

$$\mathbb{E}\, \mathcal{E}_{\overline{T}}(f_{opt}, B) = \mathbb{E}\, \mathcal{E}^*_{\overline{T}}(g_{opt}, B) \sim \frac{\lambda}{2\pi^{d-1}(d-1)!} T (\ln T)^{d-1} \tag{3}$$

и

$$\operatorname{Var} \mathcal{E}_{\overline{T}}(f_{opt}, B) = \operatorname{Var} \mathcal{E}^*_{\overline{T}}(g_{opt}, B) \sim \frac{\lambda^3}{2\pi^{d-1}(d-1)!} T (\ln T)^{d-1}. \tag{4}$$

Докажем теперь сходимость нормированной оптимальной ошибки почти наверное и в L^2.

Сходимость в L^2 следует автоматически, ведь для случайной величины

$$Q = \frac{1}{T(\ln T)^{d-1}}\mathcal{E}_{\overline{T}}(f_{opt}, B) - \frac{\lambda}{2\pi^{d-1}(d-1)!}$$

имеем $\mathbb{E}\, Q = o(1)$ и $\operatorname{Var} Q = o(1)$, откуда

$$\mathbb{E}\, Q^2 = (\mathbb{E}\, Q)^2 + \operatorname{Var} Q = o(1).$$

Теперь докажем сходимость почти наверное.

Сначала докажем сходимость для некоторого дискретного набора прямоугольников, а затем произвольный прямоугольник зажмём между двумя прямоугольниками из этого набора. Зафиксируем произвольное $a > 1$ и $c > 0$. Рассмотрим прямоугольники со сторонами $\bar{R} = (ca^{\alpha_1}, \ldots, ca^{\alpha_d})$, где $\alpha_i \in \mathbb{N}_0$. Покажем, что при стремлении объёма $R = c^d a^{\alpha_1+\ldots+\alpha_d} \to \infty$ имеет место сходимость п.н.

$$\frac{\mathcal{E}_{\overline{R}}(f_{opt}, B)}{R(\ln R)^{d-1}} \to \frac{\lambda}{2\pi^{d-1}(d-1)!}.$$

Рассмотрим центрированную последовательность

$$Q_{\overline{R}} = \frac{\mathcal{E}_{\overline{R}}(f_{opt}, B)}{R(\ln R)^{d-1}} - \mathbb{E}\, \left(\frac{\mathcal{E}_{\overline{R}}(f_{opt}, B)}{R(\ln R)^{d-1}}\right).$$

Поскольку из соотношения (3) следует

$$\mathbb{E}\, \left(\frac{\mathcal{E}_{\overline{R}}(f_{opt}, B)}{R(\ln R)^{d-1}}\right) \to \frac{\lambda}{2\pi^{d-1}(d-1)!},$$

достаточно доказать, что $Q_{\overline{R}} \to 0$ почти наверное. Из соотношения (4) следует, что для некоторой константы C

$$\mathbb{P}(|Q_{\overline{R}}| > R^{-1/3}) < \operatorname{Var} Q_{\overline{R}} R^{2/3} < CR^{-1/3}.$$

Вспомним, что $R = c^d a^{\sum_i \alpha_i}$, где вектор степеней $\bar{\alpha}$ пробегает $\mathbb{N}_0^d$. Тогда

$$\sum_{\bar{\alpha}\in\mathbb{N}_0^d} \mathbb{P}(|Q_{\overline{R}}| > R^{-1/3}) < C \sum_{\bar{\alpha}\in\mathbb{N}_0^d} c^{-d/3} a^{-\sum_i \alpha_i/3} = Cc^{-d/3} \left(\sum_{\alpha\in\mathbb{N}_0} a^{-\alpha/3}\right)^d < \infty.$$

Следовательно, по лемме Бореля – Кантелли, почти наверное выполнено лишь конечное число событий $\{|Q_{\overline{R}}| > R^{-1/3}\}$. Таким образом, при $R \to \infty$ выполнено $Q_{\overline{R}} \to 0$ почти наверное. Теперь вернёмся к изначальной последовательности. Поскольку $\min(T_i)$ отделён от 0, мы можем выбрать $c < \min(T_i)$. Тогда прямоугольник из последовательности со сторонами $\bar{T}$ зажимается между прямоугольником со сторонами $\bar{R}$ и прямоугольником со сторонами $a\bar{R}$. Поскольку с ростом параметрического множества ошибка аппроксимации растёт, получаем

$$\frac{\mathcal{E}_{\overline{R}}(f_{opt}, B)}{a^d R(\ln a^d R)^{d-1}} \le \frac{\mathcal{E}_{\overline{T}}(f_{opt}, B)}{T(\ln T)^{d-1}} \le \frac{\mathcal{E}_{a\overline{R}}(f_{opt}, B)}{R(\ln R)^{d-1}}.$$

Отсюда, применяя доказанную сходимость, получаем что при $T \to \infty$ почти наверное

$$\begin{aligned} \limsup \frac{\mathcal{E}_{\overline{T}}(f_{opt}, B)}{T(\ln T)^{d-1}} &\le \frac{a^d \lambda}{2\pi^{d-1}(d-1)!}, \\ \liminf \frac{\mathcal{E}_{\overline{T}}(f_{opt}, B)}{T(\ln T)^{d-1}} &\ge \frac{a^{-d} \lambda}{2\pi^{d-1}(d-1)!}. \end{aligned}$$

Это утверждение верно для произвольного рационального $a > 1$, из чего следует, что при $T \to \infty$ почти наверное

$$\liminf \frac{\mathcal{E}_{\overline{T}}(f_{opt}, B)}{T(\ln T)^{d-1}} = \limsup \frac{\mathcal{E}_{\overline{T}}(f_{opt}, B)}{T(\ln T)^{d-1}} = \frac{\lambda}{2\pi^{d-1}(d-1)!}.$$

□

Для полноты картины заметим, что от ограничения на отделимость $\min(T_1, \dots, T_d)$ от нуля отказаться нельзя. Без него можно брать каждый следующий прямоугольник гораздо более вытянутым, чем предыдущий. Тогда пересечение любого прямоугольника с предыдущими будет очень мало, поэтому ошибки на них будут практически независимы, из-за чего не будет сходимости почти наверное.

2.3 Доказательство предложения 4

Для непустой области $S = \{\bar{x} \,|\, \forall i\, x_i \ge r_i, x_1 \dots x_k \le A\}$ проверим, что

$$\int_S \frac{1}{x_1 \dots x_k}\, dx_1 \dots dx_k = \frac{1}{k!} \left(\ln A - \sum_i \ln r_i \right)^k.$$

Доказательство: Сначала сделаем замену $y_i = x_i / r_i \ge 1$ и обозначим $B = A / \prod_i r_i$. Тогда область S переходит в $S_1 = \{\bar{y} \,|\, y_i \ge 1, \prod_i y_i \le B\}$. Учитывая, что $dx_1 \dots dx_k = \prod_i r_i dy_1 \dots dy_k$, получим

$$\int_S \frac{1}{x_1 \dots x_k} dx_1 \dots dx_k = \int_{S_1} \frac{1}{y_1 \dots y_k} dy_1 \dots dy_k.$$

Теперь сделаем замену $t_j = \prod_{i \le j} y_i$. Область S_1 переходит в $S_2 = \{\bar{t} \,|\, 1 \le t_1 \le t_2 \le \dots \le t_k \le B\}$, при этом $dt_1 \dots dt_k = t_1 \dots t_{k-1} dy_1 \dots dy_k$. Поэтому

$$\int_{S_1} \frac{1}{y_1 \dots y_k} dy_1 \dots dy_k = \int_{S_2} \frac{1}{t_1 \dots t_k} dt_1 \dots t_k.$$

Наконец, последний интеграл симметричен по переменным $t_1, \dots, t_k$, а значит не зависит от порядка. Поэтому

$$\int_{1 \le t_1 \le \ldots \le t_k \le B} \frac{dt_1 \ldots dt_k}{t_1 \ldots t_k} = \frac{1}{k!} \int_1^B \cdots \int_1^B \frac{dt_1 \ldots dt_k}{t_1 \ldots t_k} = \frac{1}{k!} (\ln B)^k.$$

$\square$

3 Аппроксимации процессов со стационарными приращениями

Вторая часть данной работы посвящена изучению аппроксимаций процессов со стационарными приращениями.

3.1 Основные понятия

Вернёмся к задаче в изначальной постановке, когда нас интересует поведение функции

$$f_{opt,Q,\mathcal{E}}(t) = \arg\min_f \int_0^T \left(\mathcal{E}[f](t) + Q(f(t) - B(t))\right) dt.$$

Ограничимся классическим случаем L^2 нормы расстояния до процесса, соответствующей штрафу $Q(x) = x^2$. Нас интересует случай, когда B обладает стационарными приращениями в широком смысле, то есть он обладает конечными вторыми моментами и среднее и ковариационная функция процессов $\Delta_t(s) = B(t+s) - B(t)$ одинаковы для всех t. Вдобавок, будем считать, что B непрерывен в среднем квадратическом, то есть $B(t) \to B(t_0)$ в $L^2(\Omega, \mathbb{P})$ при $t \to t_0$. В дальнейшем нам понадобится корреляционная теория, начало которой было положено в работе А.Я. Хинчина [14]. Напомним основные моменты.

Для начала рассмотрим комплекснозначный, центрированный, стационарный в широком смысле и непрерывный в среднем квадратическом случайный процесс $B(t), t \in \mathbb{R}$. По теореме Хинчина, его ковариационная функция допускает спектральное представление

$$K(t) := \mathbb{E}\, B(t)\overline{B(0)} = \int_{\mathbb{R}} e^{itu} \mu(du).$$

Конечная мера $\mu(du)$ на $\mathbb{R}$ называется спектральной мерой процесса B. Тогда сам процесс допускает спектральное представление в виде случайного интеграла

$$B(t) = \int_{\mathbb{R}} e^{itu} \mathcal{W}(du),$$

где $\mathcal{W}$ – комплекснозначная центрированная мера с некоррелированными значениями на $\mathbb{R}$, связанная с мерой μ равенством $\mathbb{E}\,|\mathcal{W}(A)|^2 = \mu(A)$. Для любых двух функций f, g в пространстве $L^2(\mathbb{R}, \mu)$ стохастические интегралы

$$\int_{\mathbb{R}} f(u)\mathcal{W}(du), \qquad \int_{\mathbb{R}} g(u)\mathcal{W}(du)$$

корректно определены как центрированные случайные величины с конечным вторым моментом, причём

$$\mathbb{E} \int_{\mathbb{R}} f(u)\mathcal{W}(du)\overline{\int_{\mathbb{R}} g(u)\mathcal{W}(du)} = \int_{\mathbb{R}} f(u)\overline{g(u)}\mu(du) = (f,g)_{L^2(\mathbb{R},\mu)}.$$

Таким образом, поскольку семейство $\{e^{itu}\}_{t\in\mathbb{R}}$ образует базис в $L^2(\mathbb{R}, \mu)$, соответствие между $B(t)$ и e^{itu} продолжается до линейной изометрии между Гильбертовым пространством $H_B = \overline{\mathrm{span}}\{B(t), t \in \mathbb{R}\}$ снабжённым стандартным скалярным произведением $(\xi, \nu) = \mathbb{E}\,\xi\overline{\nu}$, и $L^2(\mathbb{R}, \mu)$. Изометрия задаётся формулой $f \mapsto \int_{\mathbb{R}} f(u)\mathcal{W}(du)$.

Аналогичная теория развита для процессов со стационарными приращениями (см. [15], [10]). Мы будем использовать то, что как и в случае стационарных в широком смысле процессов, $B(t)$ допускает спектральное представление в схожей форме

$$B(t) = \xi_0 + \xi_1 t + \int_{\mathbb{R}} (e^{itu} - 1)\mathcal{W}(du), \tag{5}$$

где ξ_0 и ξ_1 – случайные величины с конечным вторым моментом, $\mathcal{W}(du)$ – комплекснозначная центрированная случайная мера с некоррелированными значениями и мерой контроля $\mu(du)$, некоррелированная с ξ_0, ξ_1.

Поскольку постоянный сдвиг не влияет на задачу поиска аппроксимации, не умаляя общности положим $\xi_0 = B(0) = 0$.

Спектральная мера процесса со стационарными приращениями не обязана быть конечной, но заведомо удовлетворяет условию Леви

$$\int_{\mathbb{R}} \min(u^2, 1)\mu(du) < \infty.$$

Если дополнительно предположить, что случайная мера $\mathcal{W}(du)$ гауссовская, то процесс $B(t)$ тоже будет гауссовским. Самым известным примером гауссовских процессов со стационарными приращениями является дробное броуновское движение $W^{(H)}(t)$, где $0 < H \leq 1$ – фиксированный параметр. Его спектральная мера (как и спектральная мера любого другого процесса с такой ковариацией) имеет вид

$$\mu_H(du) := \frac{\Gamma(2H+1)\sin(\pi H)}{2\pi}\frac{du}{|u|^{2H+1}},$$

а дрифт отстутсвует, то есть $\xi_1 = 0$. Классическое броуновское движение является частным случаем, при $H = 1/2$.

3.2 Неадаптивная аппроксимация

Мы ограничимся тем, что будем искать аппроксимирующий процесс в виде

$$X(t) = \xi_1 t + \int_{\mathbb{R}} e^{itu}(\hat{g}(u) - 1)\mathcal{W}(du), \tag{6}$$

где $\hat{g} - 1 \in L^2(\mathbb{R}, \mu)$ и $\hat{g}(0) = 1$. Сейчас нас интересует средняя ошибка аппроксимации, записываемая в виде

$$\int_0^T \left(\mathbb{E}\,|X(t) - B(t)|^2 + \sum_{m=1}^{M} \ell_m \mathbb{E}\,|X^{(m)}(t)|^2 \right) dt.$$

И мы будем искать процесс, минимизирующий именно её. Такой выбор не случаен, поскольку для процесса аппроксимации X в форме (6) компоненты $X - B$ и все производные $X^{(m)}$ являются стационарными в широком смысле процессами, а значит нормированная ошибка, при наличии эргодичности, сходится к своему среднему. Поэтому, интересуясь асимптотикой при больших T, сразу ограничимся изучением средней ошибки.

Заданная форма аппроксимации (6) может показаться неестественной на первый взгляд, поэтому давайте приведём эвристическое рассуждение, обосновывающее её выбор. Для начала рассмотрим процесс скользящего среднего, в общем случае определяемый как свёртка процесса B с некоторой весовой функцией, имеющий вид

$$Y(t) = \int_{\mathbb{R}} g(\tau) B(t + \tau) d\tau,$$

где $g(\tau)$ – функция веса. Это стандартный объект задач линейной аппроксимации. Представим его в спектральном виде, подставив формулу (5) для $B(t + \tau)$

$$Y(t) = \int_{\mathbb{R}} \left(g(\tau)\xi_1 t + g(\tau)\xi_1 \tau + g(\tau) \int_{\mathbb{R}} (e^{itu + i\tau u} - 1)\mathcal{W}(du) \right) d\tau$$

Проинтегрируем все составляющие по τ и получим для обратного преобразования Фурье в форме $\hat{g}(u) = \int_{\mathbb{R}} e^{i\tau u} g(\tau) d\tau$ представление в виде

$$Y(t) = \xi_1 \int_{\mathbb{R}} g(\tau)\tau d\tau + \hat{g}(0)\xi_1 t + \int_{\mathbb{R}} (e^{itu}\hat{g}(u) - \hat{g}(0))\mathcal{W}(du).$$

Таким образом, выбор аппроксимации в форме (6) связан с рассмотрением процессов скользящего среднего. Сама по себе такая постановка является вопросом линейной аппроксимации, тесно связанным по смыслу с известными задачами линейной интерполяции и прогнозирования. К тому же, полученный таким образом аппроксимирующий процесс может быть легко построен на основе траектории $B(t)$. Этих причин уже достаточно, чтобы изучать именно линейные аппроксимации, однако, следующее рассуждение позволяет предположить, что такая функция, в дополнение ко всему прочему, даёт ошибку оптимального порядка.

Подойдём к задаче с другой стороны. Пусть функционал энергии сочетает в себе малость производных функции следующим полиномиальным образом

$$\mathcal{E}[f](t) = \sum_{m=1}^{M} \ell_m |f^{(m)}(t)|^2.$$

Такой функционал является некоторым обобщением соболевской нормы, в котором относительная важность производных разных порядков контролируется коэффициентами ℓ_m. Обозначим полином $\ell(x) := \sum_{m=1}^{M} \ell_m x^m$. Учитывая заданные Q и $\mathcal{E}$ задача (1) имеет вид

$$\int_0^T \left(|f(t) - B(t)|^2 + \sum_{m=1}^{M} \ell_m |f^{(m)}(t)|^2 \right) dt \searrow \min_f .$$

Это стандартная задача вариационного исчисления, поэтому, не учитывая граничные значения, приравнивая вариационную производную такого функционала к нулю, мы получаем необходимое условие экстремальности

$$f + \sum_{m=1}^{m} (-1)^m \ell_m f^{(2m)} = B.$$

Это линейное дифференциальное уравнение. Его решением будет свёртка функции B с ядром g, полученным как решение аналогичного обобщённого уравнения

$$f + \sum_{m=1}^{m} (-1)^m \ell_m f^{(2m)} = \delta,$$

где в правой части стоит дельта-функция Дирака. В частности, оптимальное решение представляет из себя процесс скользящего среднего, что объясняет наше решение рассматривать именно такой класс. Более того, пойдём дальше и возьмём у уравнения для ядра свёртки обратное преобразование Фурье в той же форме, что и ранее. Поскольку в заданной форме преобразованием Фурье дельта-функции является константа 1, получится уравнение

$$\hat{g}(\tau)(1 + \ell(\tau^2)) = 1.$$

Тогда $\hat{g}(0) = 1$. Более того, $g(t)$ симметрична, поэтому в смысле главного значения

$$\int_{\mathbb{R}} g(\tau) \tau d\tau = 0.$$

Подставим эти наблюдения в спектральное представление процесса скользящего среднего и получим

$$Y(t) = \xi_1 t + \int_{\mathbb{R}} (e^{itu} \hat{g}(u) - 1) \mathcal{W}(du).$$

Поэтому наложенное ограничение на аппроксимирующий процесс естественно, хотя и не является строго оптимальным. Вопрос о том, почему полученная функция качественно оптимальна (то есть пренебрежимо слабо проигрывает в качестве аппроксимации оптимальной функции) является предметом будущего исследования. Более того, из эвристического рассуждения мы получили форму функции веса $\hat{g}$, а именно $\hat{g}(u) = 1/(1+\ell(u^2))$, которая в действительности является правильным ответом. Получим то же равенство иначе, пользуясь лишь предположением, что аппроксимирующий процесс имеет форму (6). Заметим, что

$$X(t) - B(t) = \int_{\mathbb{R}} e^{itu}(\hat{g}(u) - 1)\mathcal{W}(du)$$

и

$$X'(t) = \xi_1 + \int_{\mathbb{R}} iue^{itu}\hat{g}(u)\mathcal{W}(du), \qquad X^{(m)}(t) = \int_{\mathbb{R}} (iu)^m e^{itu}\hat{g}(u)\mathcal{W}(du).$$

Таким образом, $X - B$ и все производные X стационарны в широком смысле. Из изометрического свойства спектрального представления следует, что

$$\mathbb{E}\,|X(t) - B(t)|^2 = \int_{\mathbb{R}} |\hat{g}(u) - 1|^2 \mu(du)$$

$$\mathbb{E}\,|X'(t)|^2 = \mathbb{E}\,|\xi_1|^2 + \int_{\mathbb{R}} u^2 |\hat{g}(u)|^2 \mu(du), \qquad \mathbb{E}\,|X^{(m)}(t)|^2 = \int_{\mathbb{R}} u^{2m} |\hat{g}(u)|^2 \mu(du).$$

Откуда средняя ошибка аппроксимации равна

$$\int_0^T \left(\mathbb{E}\,|X(t) - B(t)|^2 + \sum_{m=1}^{M} \ell_m \mathbb{E}\,|X_m|^2 \right)$$
$$= \int_0^T \int_{\mathbb{R}} |\hat{g}(u) - 1|^2 \mu(du) + \ell_1 \mathbb{E}\,|\xi_1|^2 + \int_{\mathbb{R}} \ell(u^2) |\hat{g}(u)|^2 \mu(du),$$

что может быть записано как

$$T \left(\ell_1 \mathbb{E}\,|\xi_1|^2 + \int_{\mathbb{R}} |\hat{g}(u) - 1|^2 + \ell(u^2) |\hat{g}(u)|^2 \mu(du) \right).$$

Для любых комплексных значений $\hat{g}(u)$ подинтегральное выражение можно переписать следующим образом

$$|\hat{g}(u) - 1|^2 + \ell(u^2)|\hat{g}(u)|^2 = (\ell(u^2) + 1) \left| \hat{g}(u) - \frac{1}{\ell(u^2) + 1} \right|^2 + \frac{\ell(u^2)}{\ell(u^2) + 1}.$$

Поэтому итоговая ошибка равна

$$T \left(\ell_1 \mathbb{E}\,|\xi_1|^2 + \int_{\mathbb{R}} (\ell(u^2) + 1) \left| \hat{g}(u) - \frac{1}{\ell(u^2) + 1} \right|^2 + \frac{\ell(u^2)}{\ell(u^2) + 1} \mu(du) \right). \qquad (7)$$

Такая ошибка будет минимальна при $\hat{g}(u) = \frac{1}{\ell(u^2)+1}$. Получается, что ошибка неадаптивной аппроксимации равна

$$T\left(\ell_1 \mathbb{E}\,|\xi_1|^2 + \int_{\mathbb{R}} \frac{\ell(u^2)}{\ell(u^2)+1}\mu(du)\right). \tag{8}$$

Заметим, что в данном случае весовая функция g, определяющая $X(t)$, зависит лишь от функционала энергии, но не от спектральной меры. Получается, что полученное представление траекторное и может быть найдено для любого сэмпла процесса, даже без знания ковариационной структуры самого процесса. Для этого нужно взять скользящее среднее с весом $g(\tau)$ таким, что $\hat{g} = 1/(\ell(u^2)+1)$. Поскольку $\ell(u^2)+1$ – это многочлен, преобразование Фурье такой функции нетрудно найти стандартными способами.

Отметим, что в выводе итоговой формулы для аппроксимации мы не использовали тот факт, что функционал энергии принимает форму многочлена. На самом деле аналогичная формула верна для более общего вида энергии, однако всё ещё при ограничении, что аппроксимация берется вида (6). Тем не менее, для выбранной нами формы энергии проще получить уравнение методами вариационного исчисления, которое не только эвристически обосновывает выбор данного класса, но и является важным шагом для будущего доказательства оптимальности такого сужения. Поэтому, исследования более общих функционалов энергии представляются возможными, но первым шагом является работа с описанным классом.

3.3 Адаптивная аппроксимация

Обсудим, что происходит, когда нас интересует адаптивная аппроксимация, то есть аппроксимация которой доступны лишь значения процесса B до данного момента времени. Такая задача очень схожа по смыслу с задачами прогнозирования, развивавшимися в работах Колмогорова [12] и Винера [9]. В отличие от неадаптивного случая, здесь ответ значительно зависит от спектральной меры процесса. Колмогоровым была исследована форма ошибки в довольно общем случае, а Винером получен вид оптимального предсказания для рациональных плотностей. Сам подход, как и полученные результаты, можно использовать в нашей задаче, когда мы ищем адаптивную энергетически эффективную аппроксимацию, то есть $X(t) \in H_t := \overline{\mathrm{span}}\{B(s), s \le t | L^2(\Omega, \mathbb{P})\}$. Здесь мы обсудим, какую форму принимает поставленная задача и как использовать подход теории прогнозирования для её решения. Опустим возможность наличия тренда у процесса, положив $\xi_1 = 0$, поскольку это компонента конечного ранга, не представляющая большого интереса в задачах прогнозирования. Тогда

$$B(t) = \int_{\mathbb{R}} (e^{itu} - 1)\mathcal{W}(du).$$

Тогда спектральная изометрия $f \to \int_{\mathbb{R}} f(u)\mathcal{W}(du)$ переводит пространство $H := H_{+\infty}$ в $\overline{\mathrm{span}}\{e^{isu} - 1, s \in \mathbb{R} | L^2(\mathbb{R}, \mu)\}$. В то же время, пространство H_t, порождённое значениями процесса до момента времени t, переходит в $\overline{\mathrm{span}}\{e^{isu}-1, s \leq t | L^2(\mathbb{R}, \mu)\}$.

Рассмотрим ту же задачу поиска аппроксимации в виде (6), учитывая что $\xi_1 = 0$,

$$X(t) = \int_{\mathbb{R}} (e^{itu}\hat{g}(u) - 1)\mathcal{W}(du),$$

где $\hat{g}(0) = 1$. Теперь, однако, нас интересует адаптивная аппроксимация, поэтому мы добавляем ограничение $X(t) \in H_t$. За счёт спектральной изометрии это утверждение равносильно $e^{itu}\hat{g}(u) - 1 \in \overline{\mathrm{span}}\{e^{isu}-1, s \leq t | L^2(\mathbb{R}, \mu)\}$, что упрощается до $\hat{g}(u) - 1 \in \overline{\mathrm{span}}\{e^{isu}-1, s \leq 0 | L^2(\mathbb{R}, \mu)\}$. В силу стационарности, средняя ошибка аппроксимации равна

$$T\left(\mathbb{E}\,|X(0) - B(0)|^2 + \sum_{m=1}^{M} \ell_m \mathbb{E}\,|X^{(m)}|^2\right).$$

Как и в неадаптивном случае, ошибка может быть выписана в виде (7), а именно

$$T\left(\int_{\mathbb{R}} (\ell(u^2)+1)\left|\hat{g}(u) - \frac{1}{\ell(u^2)+1}\right|^2 + \frac{\ell(u^2)}{\ell(u^2)+1}\mu(du)\right). \tag{9}$$

Второе слагаемое представляет из себя ошибку неадаптивной аппроксимации, которая от текущего выбора $\hat{g}$ не зависит. Получается, что нам нужно минимизировать первое слагаемое, которое является дополнительной платой за незнание будущего

$$\int_{\mathbb{R}} (\ell(u^2)+1)\left|\hat{g}(u) - \frac{1}{\ell(u^2)+1}\right|^2 \mu(du) \searrow \min$$

по $\hat{g}(u) - 1 \in \overline{\mathrm{span}}\{e^{isu}-1, s \leq 0 | L^2(\mathbb{R}, \mu\}$. Задача в такой постановке очень схожа с классической задачей прогнозирования, поэтому к ней разумно применить технику из той же области. Отдельно отметим, что до этого момента не было ограничений на спектральную меру, кроме условия её существования. К тому же, функционал энергии можно было бы рассматривать в более общей форме, хоть мы и оставили его многочленом для удобства.

Предположим, что спектральная мера удовлетворяет Колмогоровскому условию регулярности, то есть имеет такую плотность $\mu(du) = f(u)du$, что

$$\int_{\mathbb{R}} \frac{|\ln f(u)|}{1+u^2} du < \infty.$$

Тогда существует факторизация

$$f(u) = \gamma_f(u)\overline{\gamma_f(u)}, \tag{10}$$

где γ_f – внешняя функция, см. [2, стр. 38]. В свою очередь, поскольку $\ell(u^2)$ – многочлен с вещественными положительными коэффициентами, $1+\ell(u^2)$ не имеет вещественных корней, поэтому его можно разложить на множители согласно основной теореме алгебры следующим образом

$$1+\ell(u^2)=C\prod_{m=1}^{M}(u-\beta_m)(u-\bar{\beta}_m),$$

где $C>0$ и $\mathrm{Im}(\beta_m)>0$. Тогда рассмотрим

$$\lambda_\ell(u)=C^{1/2}\prod_{m=1}^{M}(u-\beta_m).$$

Мы получили факторизацию

$$\ell(u^2)+1=\lambda_\ell(u)\overline{\lambda_\ell(u)}, \tag{11}$$

При этом,

$$\frac{1}{u-\beta_m}=i\int_0^\infty \exp(-i(u-\beta_m)\tau)d\tau.$$

Поэтому для некоторой комплекснозначной меры ν_ℓ функция $\lambda_\ell(u)$ представима в виде

$$\frac{1}{\lambda_\ell(u)}=\int_0^\infty e^{-i\tau u}\nu_\ell(d\tau). \tag{12}$$

Запишем ошибку при помощи определённых факторизаций (10) и (11). Получим

$$\int_{\mathbb{R}}\left|(\lambda_\ell\gamma_f\hat{g})(u)-\frac{\gamma_f(u)}{\overline{\lambda_\ell(u)}}\right|^2du=\int_{\mathbb{R}}\left|(\lambda_\ell\gamma_f(\hat{g}-1))(u)-\left(\frac{\gamma_f(u)}{\overline{\lambda_\ell(u)}}-\gamma_f(u)\lambda_\ell(u)\right)\right|^2du.$$

Нужно понять, в каком пространстве может лежать функция $\lambda_\ell\gamma_f(\hat{g}-1)$, если $\hat{g}-1$ пробегает всё пространство $\overline{\mathrm{span}}\{e^{isu}-1, s\le 0|L^2(\mathbb{R},\mu)\}$. Тогда оптимальным выбором $\lambda_\ell\gamma_f(\hat{g}-1)$ будет выступать проектор на соответствующее пространство функции $\gamma_f/\overline{\lambda_\ell}-\gamma_f\lambda_\ell$, откуда можно будет найти ответ. Так выглядит общий алгоритм решения, однако в этом подходе имеются некоторые трудности, связанные с бесконечностью меры μ, которые остаются для будущих исследований.

Продемонстрируем, как такой метод работает для конечной меры μ, используя тот факт, что $1\in L^2(\mu)$. А именно, мы докажем следующее

Теорема 5 *Для конечной меры $\mu=f(u)du$ с условием регулярности Колмогорова оптимальное решение из класса* (6) *для задачи адаптивной аппроксимации, минимизирующее ошибку* (9), *имеет функцию веса*

$$\hat{g}=1+\frac{Q}{\lambda_\ell\gamma_f},$$

где Q – ортогональная проекция

$$\frac{\gamma_f}{\overline{\lambda f}} - \gamma_f \lambda_\ell$$

на пространство $\mathcal{L}' := cl\{\sum \alpha_s e^{isu}, s \leq 0, \sum_s \alpha_s = 0 | L^2(\mathbb{R}, \Lambda)\}$, *а функции* γ_f *и* λ_ℓ *определены согласно факторизации* (10) *и* (11).

Доказательство: В силу конечности меры, $1 \in L^2(\mathbb{R}, \mu)$. Поэтому пространство $\mathcal{L} := \overline{\text{span}}\{e^{isu} - 1, s \leq 0 | L^2(\mathbb{R}, \mu)\}$, которому обязана принадлежать $\hat{g} - 1$, можно представить в виде $\text{cl}\{\sum_s \alpha_s e^{isu}, s \leq 0, \sum_s \alpha_s = 0 | L^2(\mathbb{R}, \mu)\}$, то есть замыкание линейных комбинаций экспонент, которые равны 0 в нуле. Это линейное подпространство $\overline{\text{span}}\{e^{isu}, s \leq 0 | L^2(\mathbb{R}, \mu)\}$ коразмерности 1. Заметим, что если $h \in \mathcal{L}$ и $uh(u) \in L^2(\mathbb{R}, \mu)$, то $uh(u) \in \mathcal{L}$. Чтобы это понять, достаточно приблизить $uh(u)$ функциями

$$\frac{1 - e^{-i\delta u}}{i\delta} h(u) \to uh(u), \text{при } \delta \to 0,$$

где вся последовательность лежит в $\mathcal{L}$. Тогда, поскольку λ_ℓ – многочлен, то при условии, что $\lambda_\ell h$ квадратично интегрируема, а $h \in \mathcal{L}$, получаем $\lambda_\ell h \in \mathcal{L}$.

Возьмём произвольное допустимое $\hat{g} - 1 \in \mathcal{L}$. Тогда $\lambda_\ell(\hat{g} - 1) \in \mathcal{L}$, иначе, если бы эта функция не была бы квадратично интегрируема, ошибка была бы бесконечно большая. Что происходит с пространством $\mathcal{L}$ при домножении на γ_f? Поскольку γ_f является внешней функцией, $\overline{\text{span}}\{\gamma_f e^{isu}, s \leq 0 | L^2(\mathbb{R}, \Lambda)\} = L^2_{\leq 0}$, где $L^2_{\leq 0} := \overline{\text{span}}\{e^{isu}, s \leq 0 | L^2(\mathbb{R}, \Lambda)\}$, Λ – стандартная мера Лебега на прямой. Дополнительное ограничение на то, что сумма коэффициентов в оболочке равна 0, влечёт принадлежность $\gamma_f h$ схожему пространству $\mathcal{L}' = \text{cl}\{\sum \alpha_s e^{isu}, s \leq 0, \sum_s \alpha_s = 0 | L^2(\mathbb{R}, \Lambda)\}$ коразмерности 1. Но образ обладает коразмерностью не больше 1, поэтому он в точности совпадает с $\mathcal{L}'$.

Значит, ошибка не может быть меньше, чем в случае, когда $\lambda_\ell \gamma_f(\hat{g} - 1)$ является ортогональной проекцией $\gamma_f / \overline{\lambda_\ell} - \gamma_f \lambda_\ell$ на пространство $\mathcal{L}'$. Поймём, что такая функция подходит, то есть $\hat{g} - 1 \in \mathcal{L}$. Заметим, что поскольку γ_f – внешняя функция, для нашего выбора функции проекции $\lambda_\ell(\hat{g} - 1) \in \mathcal{L}$. В то же время, за счёт представления (12) получаем $\hat{g} - 1 \in \mathcal{L}$.

□

Смысл этого доказательства заключается в поиске пространства, которое пробегают $\lambda_\ell \gamma_f(\hat{g} - 1)$, когда $\hat{g} - 1$ пробегает $\mathcal{L}$. В данном вопросе значительно помогает понятие внешних функций, а именно использованная характеризация, более подробно с ней можно ознакомиться в [2]. Значительным отличием случая бесконечной меры, который интересен, поскольку охватывает такие процессы, как дробное броуновское движение, является тот факт, что функции e^{itu} перестают быть квадратично интегрируемыми, а потому использование внешних функций таким же образом не представляется возможным. Тем не менее, задача всё ещё сводится к изучению конкретного Гильбертова пространства, поэтому её можно решать отдельно, пользуясь явным представлением функций γ_f и λ_ℓ. Так, пользуясь знанием спектральной меры дробного броуновского движения, можно выписать явное представление описанного пространства и найти явную формулу для оптимальной адаптивной аппроксимации. Такой подход планируется провести в будущей работе. Отметим, что представленный Винером подход [9] для поиска явной формы оптимального прогноза можно перенять для нашей задаче, чтобы найти явный вид оптимальной аппроксимации в представленном случае для хороших спектральных плотностей.

Список литературы

[1] A.Berlinet, C.Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, Kluwer, 2004.

[2] H. Dym, H.P̃. McKean, *Gaussian Processes, Function Theory, and the Inverse Spectral Problem*, Dover Publications, 2008

[3] I.Ibragimov, Z.Kabluchko, M.Lifshits, *Some extensions of linear approximation and prediction problems for stationary processes*, Stoch. Proc. Appl., 2019, 129, 2758–2782.

[4] Z.Kabluchko, M.Lifshits, *Least energy approximations for processes with stationary increments*, J. Theor. Probab., 2017, 30, 1, 268–296.

[5] M.Kanagawa, P.Hennig, D.Sejdinovic, B.K. Sriperumbudur, *Gaussian Processes and Kernel Methods: A Review on Connections and Equivalences*, arXiv:1807.02582

[6] M.Lifshits, A.Siuniaev, *Energy of taut strings accompanying random walk*, Probab. Math. Stat., 2021, 41, 1, 9–23.

[7] M.Lifshits, E.Setterqvist, *Energy of taut string accompanying Wiener process*, Stoch. Proc. Appl., 2015, 125, 401–427.

[8] C.E.Rasmussen, C.K.I.Williams, *Gaussian Processes for Machine Learning*, the MIT Press, 2006.

[9] N. Wiener, *Extrapolation, interpolation and smoothing of stationary time series*, New York, 1949

[10] Yaglom, *A.M. An Introduction to the Theory of Stationary Random Functions. Revised English edition.* Dover, New York, 2004

[11] Д. И. Блинова, М. А. Лифшиц, *Энергия натянутых струн, сопровождающих винеровский процесс и случайное блуждание в полосе переменной ширины*, Зап. научн. сем. ПОМИ, 495, 64–86.

[12] А. Н . Колмогоров, *Интерполирование и экстраполирование стационарных случайных последовательностей*, Изв. АН СССР, серия матем., т. 5, No 1 (1941), 3—14

[13] М. А. Лифшиц, З. А. Каблучко, *Адаптивная энергетически эффективная аппроксимация стационарных процессов*, Известия РАН, сер. Математическая. 2019, 83, 5, 27–52.

[14] А. Я. Хинчин, *Теория корреляции стационарных стохастических процессов*, Успехи матем. наук, вып. V (1938), 42—51 (впервые эта статья была опубликована в Math. Ann. в 1934 г.)

[15] А. М. Яглом, *Корреляционная теория процессов со случайными стационарными n-ми приращениями*, Матем. сб., 1955, том 79, номер 1, 141–196