Санкт-Петербургский государственный университет

НОВОЖИЛОВ Сергей Алексеевич

Выпускная квалификационная работа

Функциональная предельная теорема в методе RFF

Уровень образования:
Направление 01.03.01 «Математика»
Основная образовательная программа СВ.5000.2017 «Математика»

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

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

Санкт-Петербург
2021

Содержание

1 Введение и постановка задачи

Метод Random Fourier Features (RFF) был предложен в [1] и используется для конечно-ранговой аппроксимации коварицонных функций больших размерностей. Так как гауссовские процессы в пространствах высокой размерности используются в некоторых моделях машинного обучения, метод RFF позволяет увеличить скорость работы с такими моделями (например, скорость обучения) без заметных потерь в точности. Опишем метод RFF более детально.

Рассмотрим гауссовский стационарный процесс, задаваемый ядром

$$k(x, y) = k(x - y) \quad x, y \in \mathbb{R}^d.$$

Метод заключается в построении такого отображения $\Phi : \mathbb{R}^d \to \mathbb{R}^M$, что

$$k(x, y) \approx \langle \Phi(x), \Phi(y) \rangle, \quad x, y \in \mathbb{R}^d,$$

где $\langle \cdot, \cdot \rangle$ обозначает скалярное произведение.

Заметим что, имея такую функцию Φ, легко построить случайный процесс с ковариационной функцией $k(\cdot)$. Для этого рассмотрим $M-$мерный вектор ξ, у которого координаты являются независимыми случайными величинами с единичной дисперсией. Проверим, что $k(x, y)$ является ковариационной функцией случайного процесса

$$G(x) := \langle \Phi(x), \xi \rangle .$$

Действительно,

$$\begin{aligned}
\operatorname{cov}(G(x), G(y)) &= \operatorname{cov}(\langle G(x), \xi \rangle, \langle G(y), \xi \rangle) \\
&= \sum_{i,j=1}^{M} \operatorname{cov}(\Phi_i(x)\xi_i, \Phi_j(y)\xi_j) \\
&= \sum_{i=1}^{M} \Phi(x)\Phi(y)\operatorname{cov}(\xi_i, \xi_i) \\
&= \langle \Phi(x), \Phi(y) \rangle \approx k(x, y).
\end{aligned}$$

Для простоты будем считать, что $k(0) = 1$. Тогда по теореме Бохнера k является обратным преобразованием Фурье вероятностной симметричной меры $\mathcal{P}$ на $\mathbb{R}^d$ (симметричность следует из вещественности $k(\cdot)$):

$$k(x) = \int\limits_{\mathbb{R}^d} e^{i\langle s, x \rangle} \mathrm{d}\mathcal{P}(s) \quad x \in \mathbb{R}^d.$$

Рассмотрим семейство $\{S_r\}_{r=1}^M$ независимых случайных величин, $S_r \in \mathbb{R}^d$, которые распределены в соответствии с мерой $\mathcal{P}$. Ещё понадобится семейство $\{\tau_r\}_{r=1}^M$ независимых величин, распределенных равномерно на $[0, 2\pi]$.

В методе RFF отображение Φ строится вероятностно. Покажем, что реализация случайного вектора

$$\Phi(x) = \left(\frac{\sqrt{2}}{\sqrt{M}} \cos(\langle S_1, x \rangle + \tau_1), \ldots, \frac{\sqrt{2}}{\sqrt{M}} \cos(\langle S_M, x \rangle + \tau_M) \right)^\top, \quad x \in \mathbb{R}^d \tag{1}$$

обладает свойством $\langle \Phi(x), \Phi(y) \rangle \approx k(x - y)$.

Для доказательства этого факта нам потребуется несколько технических утверждений.

Лемма 1.
$$k(x) = \mathbb{E}\,[\cos(\langle S_r, x\rangle)].$$

Доказательство. Преобразуем представление ядра k:

$$k(x) = \int\limits_{\mathbb{R}^d} e^{i\langle s,x\rangle} \mathrm{d}\mathcal{P}(s) = \mathbb{E}\,[e^{i\langle S_r,x\rangle}] = \mathbb{E}\,[\cos(\langle S_r, x\rangle) + i\sin(\langle S_r, x\rangle)].$$

Из симметричности меры $\mathcal{P}$ в совокупности с нечетностью функции sin следует:

$$\mathbb{E}\,[i\sin(\langle S_r, x\rangle)] = \int\limits_{\mathbb{R}^d} i\sin(\langle s, x\rangle)\mathrm{d}\mathcal{P}(s) = 0.$$

То есть
$$k(x) = \mathbb{E}\,[\cos(\langle S_r, x\rangle)].$$

$\square$

Для $1 \le r \le M$ определим случайные поля

$$\phi_r(x) := \sqrt{2}\cos(\langle S_r, x\rangle + \tau_r).$$

Лемма 2.
$$\mathbb{E}\,[\phi_r(x)\phi_r(y)] = k(x-y).$$

Доказательство.

$$\begin{aligned}
\mathbb{E}\,[\phi_r(x)\phi_r(y)] &= \frac{1}{2\pi}\int\limits_0^{2\pi}\int\limits_{\mathbb{R}^d} 2\cos(\langle s, x\rangle + \tau)\cos(\langle s, y\rangle + \tau)\mathrm{d}\mathcal{P}(x)\mathrm{d}\tau \\
&= \frac{1}{2\pi}\int\limits_0^{2\pi}\int\limits_{\mathbb{R}^d} \cos(\langle s, x+y\rangle + 2\tau) + \cos(\langle s, x-y\rangle)\mathrm{d}\mathcal{P}(x)\mathrm{d}\tau \\
&= \mathbb{E}\,[\cos(\langle S_r, x-y\rangle)] = k(x-y),
\end{aligned}$$

где последний переход выполнен с использованием леммы 1. $\square$

Теперь, пользуясь законом больших чисел, получаем требуемое утверждение:

$$\begin{aligned}
k(x-y) &= \mathbb{E}\,[\phi_1(x), \phi_1(y)] \\
&\approx \frac{1}{M}\sum_{r=1}^{M} 2\cos(\langle S_r, x\rangle + \tau_m)\cos(\langle S_r, y\rangle + \tau_m) \\
&= \langle \Phi(x), \Phi(y)\rangle\,.
\end{aligned}$$

В дальнейшем будем обозначать

$$k_M(x,y) := \langle \Phi(x), \Phi(y)\rangle \quad x, y \in \mathbb{R}^d. \tag{2}$$

Именно это поле является конечно-ранговой аппроксимацией $k(\cdot)$ в методе RFF. В работе [1] устанавливается экспоненциально быстрая сходимость $k_M(x,y)$ к $k(x-y)$, в том смысле, что для любого $\varepsilon > 0$ верно:

$$\mathbb{P}\{|k_M(x,y) - k(x-y)| > \varepsilon\} \le e^{-cM},$$

где $c = c(\varepsilon)$.

Объектом изучения данной работы будет случайное поле:

$$W_M(x,y) := \sqrt{M}(k_M(x,y) - k(x-y)), \quad x, y \in \mathbb{R}^d. \tag{3}$$

Нормировка $\sqrt{M}$ естественна, так как при такой нормировке конечномерные распределения $W_M(x,y)$ подчиняются центральной предельной теореме.

Покажем центрированность поля $W_M(x,y)$.

Лемма 3.

$$\mathbb{E}\left[k_M(x,y)\right] = k(x-y),$$

А значит и

$$\mathbb{E}\left[W_M(x,y)\right] = 0.$$

Доказательство. По определению

$$k_M(x,y) = \langle \Phi(x), \Phi(y) \rangle = \frac{1}{M}\sum_{r=1}^{M} \phi_r(x)\phi_r(y),$$

по лемме 2:

$$\mathbb{E}\left[k_M(x,y)\right] = \frac{1}{M} M k(x-y) = k(x-y).$$

Утверждение $\mathbb{E}\left[W_M(x,y)\right] = 0$ следует немедленно из вышесказанного и формулы (3). $\square$

Случайные поля, аналогичные полю W_M, возникают в другой, более изученной задаче, а именно в анализе эмпирических характеристических функций. Напомним некоторые определения. Пусть X — случайный вектор в $\mathbb{R}^d$. Пусть $c(t)$ — его характеристическая функция, равная

$$c(t) = \mathbb{E}\left[e^{i\langle X,t \rangle}\right]. \tag{4}$$

Также определим эмпирическую характеристическую функцию:

$$c_M(t) := \frac{1}{M}\sum_{r=1}^{M} e^{i\langle X_r, t\rangle}, \quad t \in \mathbb{R}^d, \tag{5}$$

где X_r независимые равнораспределенные копии X, а также рассмотрим случайное поле

$$C_M(t) = \sqrt{M}(c_M(t) - c(t)). \tag{6}$$

Вопросы сходимости последовательности случайных полей $C_M(t)$ подробно изучены в классических работах [5], [7], [3], [4], [6]. Сходимость понимается в следующем смысле: рассмотрим банахово пространство непрерывных комплекснозначных функций

$$\mathcal{C}([-1/2, 1/2]^d, \|\cdot\|_\infty).$$

Случайные поля $C_M(t)$ можно рассматривать как меры на пространстве функций $\mathcal{C}$, в перечисленных работах вопрос сходимости изучается с точки зрения слабой сходимости мер, соответствующих $C_M(t)$. Подробный обзор результатов, связанных с эмпирическими характеристическими функциями будет дан в разделе 5.

Сформулируем основной результат этой работы:

Теорема. *Существование предела последовательности $C_M(t)$, соответствующих величине, распределенной в соответствии с мерой $\mathcal{P}$, в описанном выше смысле, влечет существование предела последовательности процессов $W_M(x,y)$, соответствующих той же мере $\mathcal{P}$, в смысле слабой сходимости мер в пространстве непрерывных комплекснозначных функций $\mathcal{C}([-1/2,1/2]^{2d}, \|\cdot\|_\infty)$.*

Эта теорема будет доказана в разделе 7.

2 Подсчет ковариационной функции случайного поля W_M

Оказывается, что ковариационная функция поля W_M не зависит от M. В этом пункте мы выразим её явно.

Обозначим

$$k_W(v_1, v_2) := \operatorname{cov}(W_M(x_1, y_1), W_M(x_2, y_2)),$$

где $v_i = (x_i, y_i) \in \mathbb{R}^{2d}$. Заметим, что из формул (2), (1), (3) следует, что:

$$W_M(x, y) = \sqrt{M} \left(\frac{1}{M} \sum_{r=1}^{M} (\phi_r(x)\phi_r(y) - k(x - y)) \right). \tag{7}$$

Заметим, что разные слагаемые независимы друг от друга и равнораспределены, поэтому

$$\begin{aligned}\operatorname{cov}(W_M(x_1, y_1), W_M(x_2, y_2)) &= \operatorname{cov}(\phi_1(x_1)\phi_1(y_1) - k(x_1 - y_1), \phi_1(x_2)\phi_1(y_2) - k(x_2 - y_2)) \\ &= \operatorname{cov}(\phi_1(x_1)\phi_1(y_1), \phi_1(x_2)\phi_1(y_2)).\end{aligned}$$

Заметим, что ковариационная функция не зависит от параметра M. Далее заметим, что в силу леммы 2:

$$\mathbb{E}\, \phi(x)\phi(y) = k(x - y),$$

из чего следует:

$$\operatorname{cov}(\phi_1(x_1)\phi_1(y_1), \phi_1(x_2)\phi_1(y_2)) = \mathbb{E}\left[\phi_1(x_1)\phi_1(y_1)\phi_1(x_2)\phi_1(y_2)\right] - k(x_1 - y_1)k(x_2 - y_2).$$

Чтобы посчитать $\mathbb{E}\left[\phi(x_1)\phi(y_1)\phi(x_2)\phi(y_2)\right]$, мы будем пользоваться двумя элементарными фактами.

Факт 1: для любых $a, b, c, d \in \mathbb{C}$ верно

$$\begin{aligned}\cos(a)\cos(b)\cos(c)\cos(d) = \frac{1}{8}(&\cos(a - b - c - d) + \cos(a + b - c - d) + \cos(a - b + c - d) + \\ &+ \cos(a + b + c - d) + \cos(a - b - c + d) + \cos(a + b - c + d) + \\ &+ \cos(a - b + c + d) + \cos(a + b + c + d)),\end{aligned}$$

что является результатом двукратного применения формулы

$$\cos(a)\cos(b) = \frac{1}{2}(\cos(a + b) + \cos(a - b)).$$

Факт 2: для любого $k \in \mathbb{Z} \setminus \{0\}$ верно:

$$\int\limits_0^{2\pi} \cos(a + k\tau)\mathrm{d}\tau = \frac{1}{k} \int\limits_a^{a+2\pi k} \cos(u)\mathrm{d}u = 0.$$

Применяя эти два факта, посчитаем:

$$\begin{aligned}&\mathbb{E}\left[\phi(x_1)\phi(y_1)\phi(x_2)\phi(y_2)\right] = \frac{1}{2\pi} \int\limits_0^{2\pi} \int\limits_{\mathbb{R}^d} 4\cos(\langle s, x_1 \rangle + \tau) \times \\ &\times \cos(\langle s, x_2 \rangle + \tau)\cos(\langle s, y_1 \rangle + \tau)\cos(\langle s, y_2 \rangle + \tau)\mathrm{d}\mathcal{P}(s)\mathrm{d}\tau \\ &= \frac{1}{2\pi} \int\limits_0^{2\pi} \int\limits_{\mathbb{R}^d} \frac{1}{2}(\cos(\langle s, x_1 + x_2 - y_1 - y_2 \rangle) + \cos(\langle s, x_1 - x_2 + y_1 - y_2 \rangle) + \\ &+ \cos(\langle s, x_1 - x_2 - y_1 + y_2 \rangle))\mathrm{d}\mathcal{P}(s)\mathrm{d}\tau \\ &= \frac{1}{2}(k(x_1 + x_2 - y_1 - y_2) + k(x_1 + y_1 - x_2 - y_2) + k(x_1 + y_2 - x_2 - y_1)).\end{aligned}$$

То есть мы проверили, что искомая ковариационная функция имеет вид:

$$k_W(v_1, v_2) = \mathrm{cov}(W_M(x_1, y_1), W_M(x_2, y_2)) = \mathrm{cov}(\phi(x_1)\phi(y_1), \phi(x_2)\phi(y_2))$$
$$= \frac{1}{2}(k(x_1 + x_2 - y_1 - y_2) + k(x_1 + y_1 - x_2 - y_2) + k(x_1 + y_2 - x_2 - y_1)) - k(x_1 - y_1)k(x_2 - y_2).$$

3 Общий случай

Откажемся от условия $k(0) = 1$. Рассмотрим ядро $k(x, y) = k(x - y)$ стационарного гауссовского процесса. По нему построим нормированное ядро

$$k^0(x) = \frac{k(x)}{k(0)}.$$

По нормированному ядру построим его приближение (применим предыдущий пункт):

$$k^0(x, y) \approx k^0_M(x, y) = \langle \Phi(x), \Phi(y) \rangle .$$

Тогда

$$k(x, y) = k(0)k^0(x, y) \approx k(0)k^0_M(x, y) = \left\langle \sqrt{k(0)}\Phi(x), \sqrt{k(0)}\Phi(y) \right\rangle =: k_M(x, y).$$

Аналогично с нормированным случаем обозначим

$$W_M(x, y) := \sqrt{M}(k_M(x, y) - k(x - y)) = k(0)\sqrt{M}(k^0_M(x, y) - k^0(x - y)),$$

то есть

$$W_M(x, y) = k(0)W^0_M(x, y).$$

Посчитаем ковариационную функцию:

$$\mathrm{cov}(W_M(x_1, y_1), W_M(x_2, y_2)) = \mathrm{cov}(k(0)W^0_M(x_1, y_1), k(0)W^0_M(x_2, y_2))$$
$$= k^2(0)\mathrm{cov}(W^0_M(x_1, y_1), W^0_M(x_2, y_2))$$
$$= \frac{1}{2}k(0)(k(x_1 + x_2 - y_1 - y_2) + k(x_1 + y_1 - x_2 - y_2) + k(x_1 + y_2 - x_2 - y_1)) - k(x_1 - y_1)k(x_2 - y_2).$$

4 Сходимость конечномерных распределений

Опять будем считать, что $k(0) = 1$. Рассмотрим набор из s векторов $\{v_1, v_2, v_3, \dots, v_s\}$, где $v_i = (x_i, y_i) \in \mathbb{R}^{2d}$ и соответствующую матрицу ковариаций

$$\mathbf{\Sigma}_{ij} \; = k_W(v_i, v_j).$$

Пусть $1 \leq r \leq M$. Рассмотрим случайные векторы

$$X_r = \begin{pmatrix} \phi_r(x_1)\phi_r(y_1) \\ \phi_r(x_2)\phi_r(y_2) \\ \vdots \\ \phi_r(x_s)\phi_r(y_s) \end{pmatrix},$$

они независимы и одинаково распределены. Также рассмотрим детерминированный вектор

$$\mu = \begin{pmatrix} k(x_1, y_1) \\ k(x_2, y_2) \\ \vdots \\ k(x_s, y_s) \end{pmatrix}.$$

Как мы выяснили ранее, матрица ковариаций вектора X_r совпадает с $\mathbf{\Sigma}$. Заметим, что

$$\begin{pmatrix} W_M(v_1) \\ W_M(v_2) \\ \vdots \\ W_M(v_s) \end{pmatrix} = \sqrt{M} \left(\frac{1}{M} \sum_{r=1}^{M} X_r - \mu \right).$$

Тогда по многомерной центральной предельной теореме

$$\sqrt{M} \left(\frac{1}{M} \sum_{r=1}^{M} X_r - \mu \right) \xrightarrow[M\to\infty]{D} \mathcal{N}(0, \mathbf{\Sigma})$$

в $\mathbb{R}^s$. Таким образом, мы показали сходимость конечномерных распределений W_M к гауссовскому распределению с нулевым математическим ожиданием и матрицей ковариаций $\mathbf{\Sigma}_{ij} = k_W(v_i, v_j)$.

5 Эмпирические функции распределения

В этом разделе будет перечислены известные результаты, относящиеся к сходимости эмпирической характеристической функции случайной величины.

5.1 Одномерная эмпирическая характеристическая функция

Пусть X случайная величина со значениями в $\mathbb{R}$, пусть $F(x) = \mathbb{P}[X \leq x]$ её функция распределения, а $c(t) = \mathbb{E}\left[e^{iXt}\right]$ её характеристическая функция.

Рассмотрим эмпирическую функцию распределения:

$$F_M(x) := \frac{1}{M} \sum_{r=1}^{M} \mathbb{1}_{[X_r, +\infty)},$$

где X_r независимы и равнораспределены с X. Также определим эмпирическую характеристическую функцию:

$$c_M(t) := \frac{1}{M} \sum_{r=1}^{M} e^{iX_r t}.$$

Рассмотрим:

$$C_M(t) := \sqrt{M}(c_M(t) - c(t)), \quad -\frac{1}{2} \leq t \leq \frac{1}{2}.$$

Рассмотрим банахово пространство комплекснозначных непрерывных функций на отрезке с супремум-нормой:

$$\mathcal{C} = (C[-1/2, 1/2], \ \|\cdot\|_\infty).$$

Из непрерывности характеристической функции $c(t)$ следует, что C_M случайный элемент $\mathcal{C}$. Поэтому C_M задает меру на банаховом пространстве $\mathcal{C}$. Далее будут приведены результаты о слабой сходимости мер, задаваемых C_M.

Несложно убедиться, в предположении, что слабый предел существует, то это мера определяется центрированным гауссовским процессом Y с ковариационной функцией

$$k_Y(s,t) = c(t-s) - c(t)c(-s).$$

В статье М.Б. Маркуса [5] приведено необходимое и достаточное условие сходимости C_M к гауссовскому процессу Y.

Теорема (М.Б. Маркус, [5])**.** *Пусть*

$$\sigma^2(t) = 4 \int\limits_{\mathbb{R}} \sin^2 \frac{xt}{2} dF(x) = 2(1 - \operatorname{Re}\, c(t)), \quad -1 \le t \le 1.$$

Пусть

$$m_\sigma(x) = \lambda\{u \in [-1,1] : \ \sigma(u) < x\},$$

где λ — мера Лебега. Определим неубывающую перестановку $\bar{\sigma}$ функции σ как

$$\bar{\sigma}(u) = \sup\{y : \ m_\sigma(y) < u\}.$$

Пусть

$$I(\sigma) = \int\limits_0^2 \frac{\bar{\sigma}(s)}{s \log \sqrt{\frac{16}{s}}} ds.$$

Тогда если $I(\sigma) < \infty$, то $C_M(t)$ слабо сходится в пространстве $\mathcal{C}$ к процессу Y.

Если же $I(\sigma) = \infty$, то у гауссовского процесса Y не существует непрерывных модификаций, поэтому $C_M(t)$ не имеет слабого предела в $\mathcal{C}$.

Дж. Юкич в статье [7] предложил необходимое и достаточное условие, использующее метрическую энтропию (при некоторых дополнительных предположениях).

Начнем с определения так называемой скобочной метрической энтропии (metric entropy with bracketing). Пусть $(\Omega, \mathcal{F})$ пространство с заданной на нем сигма-алгеброй. Обозначим $\mathcal{L}^0(\Omega, \mathcal{F})$ множество вещественно значных $\mathcal{F}$-измеримых функций. Для $f, g \in \mathcal{L}^0(\Omega, \mathcal{F})$ определим

$$[f, g] := \{h \in \mathcal{L}^0(\Omega, \mathcal{F}) : \ f \le h \le g\}.$$

Пусть $(\Omega, \mathcal{F}, P)$ вероятностное пространство. Для произвольного класса функций $\mathcal{A} \subset \mathcal{L}^1(\Omega, \mathcal{F}, P)$ и произвольного $\varepsilon > 0$ определим скобочную метрическую энтропию $N^B(\varepsilon, \mathcal{A}, P)$ как минимальное m, при котором существует набор функций

$$\{f_1, f_2, \ldots, f_m\} \subset \mathcal{L}^1(\Omega, \mathcal{F}, P)$$

такой, что

$$\mathcal{A} \subset \bigcup_{i,j} \{[\, f_i, f_j] : \int |f_i - f_j| dP < \varepsilon\}.$$

Теорема (Дж Юкич, [7])**.** *Пусть $\Omega = [-1/2, 1/2]$. Пусть у случайной величины X существует плотность p. К тому же, пусть существует $x_0 > 0$, что $p(x) + p(-x)$ убывает при $x > x_0$. Рассмотрим класс функций $\mathcal{A} := \{x \mapsto e^{itx} : \ |t| < 1/2\}$. Тогда C_M слабо сходится к гауссовскому процессу тогда и только тогда, когда*

$$\int\limits_0^1 \sqrt{N^B(\varepsilon^2, \mathcal{A}, P)} d\varepsilon < \infty.$$

5.2 Многомерная эмпирическая характеристическая функция

Пусть X — случайный вектор в $\mathbb{R}^d$. Пусть $c(t), c_M(t), C_M(t)$ определены как в (4), (5), (6).

Рассмотрим банахово пространство непрерывных комплекснозначных функций $\mathcal{C}([-1/2, 1/2]^d, \|\cdot\|_\infty)$, и рассмотрим вопрос слабой сходимости мер, задаваемых C_M на $\mathcal{C}$.

В своей статье Ш. Чёргё [3] приводит аналогичное одномерному случаю необходимое и достаточное условие существования слабого предела.

Теорема (Ш. Чёргё [3]). *Пусть*

$$\sigma^2(t-s) = 1 - \mathrm{Re}(C(t-s)).$$

Пусть

$$m_\sigma(y) = \lambda_d\{t:\ \|t\|_\infty < \frac{1}{2}, \sigma(y) < t\}, \quad 0 \le y \le 1.$$

Рассмотрим неубывающую перестановку

$$\bar{\sigma}(u) = \sup\{y:\ m_\sigma(y) < u\}.$$

Тогда C_M слабо сходится к некоторому гауссовскому процессу (который однозначно определяется своими конечномерными распределениями) тогда и только тогда, когда

$$\int\limits_0^2 \frac{\bar{\sigma}(u)}{\sqrt{h \log \frac{1}{h}}} dh < \infty.$$

6 Некоторые соображения о связи задач

Сделаем замену $u = x - y,\ v = x + y$. Воспользуемся формулой произведения косинусов и перепишем наше поле $W_M(x, y)$ в несколько ином виде:

$$\begin{aligned} W_M(x,y) &= \sqrt{M}\left(\frac{1}{M}\sum_{r=1}^{M} 2\cos(\langle S_r, x\rangle + \tau_r)\cos(\langle S_r, y\rangle + \tau_r) - k(x-y)\right) \\ &= \sqrt{M}\left(\frac{1}{M}\sum_{r=1}^{M}\cos(\langle S_r, x-y\rangle) - k(x-y)\right) + \frac{1}{\sqrt{M}}\sum_{r=1}^{M}\cos(\langle S_r, x+y\rangle + 2\tau_r) \\ &=: W_M^I(u) + W_M^{II}(v). \end{aligned}$$

Заметим, что $W_M^I(u) = \mathrm{Re}(C_M(u))$, где C_M эмпирическая характеристическая функция для случайной величины, распределенной в соответствии с мерой $\mathcal{P}$.

Найдем $\mathrm{cov}(W_M^{II}(v_1), W_M^{II}(v_2))$. Для этого заметим, что слагаемые попарно независимы.

$$\begin{aligned} \mathrm{cov}(W_M^{II}(v_1), W_M^{II}(v_2)) &= \mathrm{cov}(\cos(\langle S_1, v_1\rangle + 2\tau_1), \cos(\langle S_1, v_2\rangle + 2\tau_1)) \\ &= \frac{1}{2}\mathbb{E}\left[\cos(\langle S_1, v_1 - v_2\rangle)\right] + \frac{1}{2}\mathbb{E}\left[\cos(\langle S_1, v_1 + v_2\rangle) + 4\tau_1\right] = \\ &= \frac{k(v_1 - v_2)}{2}. \end{aligned}$$

То есть если W_M^{II} к чему-то сходится, то это изначальный гауссовский процесс, умноженный на $\frac{1}{\sqrt{2}}$.

Покажем, что $\mathrm{cov}(W_M^I(u), W_M^{II}(v)) = 0$. Для этого заметим, что слагаемые с разными индексами независимы, поэтому можно искать ковариацию

$$\mathrm{cov}(\cos(\langle S_1, u\rangle) - k(u), \cos(\langle S_1, v\rangle + 2\tau_1)) = \mathbb{E}\left[(\cos(\langle S_1, u\rangle) - k(u))\cos(\langle S_1, v\rangle + 2\tau_1)\right] = 0,$$

где математическое ожидание равно нулю, так как результатом интегрировании по $\mathrm{d}\tau$ по периоду функции cos является ноль.

То есть мы разложили наше поле в сумму двух, причем предельные гауссовские процессы, соответствующие этим полям, некоррелированны.

7 Основной результат

В этом пункте мы докажем следующую теорему:

Теорема. *Описанный выше процесс $W_M(x, y)$ сходится к гауссовскому полю (в смысле слабой сходимости мер в $\mathcal{C}([-1/2, 1/2]^{2d}, \|\cdot\|_\infty)$ топологии), если функция $C_M(t) = \sqrt{M}(c_M(t) - c(t))$ для величины, распределенной в соответствии с мерой $\mathcal{P}$, удовлетворяет центральной предельной теореме.*

Доказательство. В доказательстве мы будем опираться на результаты статьи [3]. В этой статье Ш. Чёргё, в свою очередь, опирается на общий ход второго доказательства, которое приводит М. Маркус в статье [5] (Theorem 1, стр 198-200). Подход М. Маркуса основывается на применении следующей теоремы:

Теорема (Ферник, Theorem 1.3 [8]). *Пусть E банахово пространство, пусть Z случайная величина со значениями в E. Тогда, чтобы Z удовлетворяла (стандартной) центральной теореме в пространстве E, необходимо и достаточно, чтобы для произвольного $\varepsilon > 0$ существовала случайная величина Y со значениями в E, которая бы удовлетворяла центральной предельной теореме в E, и такая, что для всех $n > 0$ было выполнено:*

$$\mathbb{E}\left\|\frac{1}{\sqrt{M}}\sum_{r=1}^{M}(Z_r - Y_r)\right\| < \varepsilon,$$

где Z_r, Y_r независимые копии Z, Y.

Чтобы применить эту теорему, сначала определим

$$V_M^I(u) = \sqrt{M}\left(\frac{1}{M}\sum_{r=1}^{M} e^{i\langle S_r, u\rangle} - k(u)\right),$$

$$V_M^{II}(v) = \frac{1}{\sqrt{M}}\sum_{r=1}^{M} e^{i\langle S_r, v\rangle + 2i\tau_r},$$

$$V_M(u, v) := V_M^I(u) + V_M^{II}(v),$$

(очевидно, что $W_M, W_M^{I,II} = \Re V_M, \Re V_M^{I,II}$). Перепишем представления $V_M^{I,II}$ в виде интегралов по случайным мерам:

$$V_M^I(u) = \int\limits_{\mathbb{R}^d} e^{i\langle x, u\rangle} d\sqrt{M}\left(\frac{1}{M}\sum_{r=1}^{M} \mathbb{1}_{S_r \le x} - \mathcal{P}(x)\right),$$

$$V_M^{II}(v) = \int\limits_{\mathbb{R}^d} e^{i\langle x, v\rangle} d\left(\frac{1}{\sqrt{M}}\sum_{r=1}^{M} e^{2i\tau_r}\mathbb{1}_{S_r \le x}\right).$$

Обозначим множество точек $\{x \in \mathbb{R}^d :\ \|x\|_\infty < R\}$ как Π_R, а его дополнение в $\mathbb{R}^d$ как $\overline{\Pi}_R$. В качестве приближений ($\frac{1}{\sqrt{M}}\sum Y_r$, в терминах теоремы), мы будем использовать следующие интегралы:

$$\dot{V}_M^I(u) = \int\limits_{\Pi_R} e^{i\langle x, u\rangle} d\sqrt{M}\left(\frac{1}{M}\sum_{r=1}^{M} \mathbb{1}_{S_r \le x} - \mathcal{P}(x)\right),$$

$$\dot{V}_M^{II}(v) = \int\limits_{\Pi_R} e^{i\langle x, v\rangle} d\left(\frac{1}{\sqrt{M}}\sum_{r=1}^{M} e^{2i\tau_r}\mathbb{1}_{S_r \le x}\right).$$

Их дополнения будем обозначать как:

$$\mathring{V}_M^I(u) = \int\limits_{\overline{\Pi}_R} e^{i\langle x,u\rangle} d\sqrt{M}\left(\frac{1}{M}\sum_{r=1}^{M} \mathbb{1}_{S_r\le x} - \mathcal{P}(x)\right),$$

$$\mathring{V}_M^{II}(v) = \int\limits_{\overline{\Pi}_R} e^{i\langle x,v\rangle} d\left(\frac{1}{\sqrt{M}}\sum_{r=1}^{M} e^{2i\tau_r}\mathbb{1}_{S_r\le x}\right).$$

Дополнительно будем предполагать, что граница Π_R имеет нулевую меру:

$$\mathcal{P}\left(\partial\Pi_R\right) = 0,$$

что мы можем делать в силу того, что существует континуальное число кубических окрестностей нуля, из которых положительную меру границы может иметь лишь счётное число кубов.
Чтобы воспользоваться теоремой, нужно доказать, что:

1. для любого $\varepsilon > 0$ можно выбрать такое достаточно большое R, что
$$\mathbb{E}\,\|\mathring{V}_M^I(u) + \mathring{V}_M^{II}(v)\|_\infty \le \varepsilon;$$
2. приближающая последовательность $\dot{V}_M^I(u)+\dot{V}_M^{II}(v)$ удовлетворяет условиям центральной предельной теоремы.

Сначала оценим внешнюю часть:

$$\mathbb{E}\,\|\mathring{V}_M^I(u) + \mathring{V}_M^{II}(v)\|_\infty \le \mathbb{E}\,\|\mathring{V}_M^I(u)\|_\infty + \mathbb{E}\,\|\mathring{V}_M^{II}(v)\|_\infty.$$

В работе [3] (Theorem 3.1 стр 209; Theorem 7.1) доказаны следующие оценки:

$$\mathbb{E}\,\|\mathring{V}_M^I\|_\infty, \mathbb{E}\,\|\mathring{V}_M^{II}\|_\infty \le \alpha(R) + C\int\limits_0^{\beta(R)}\left[\log\left(1+\frac{1}{m(h/2)}\right)\right]^{1/2} dh,$$

где $\alpha(R), \beta(R) \to 0$ при $R \to \infty$, а $m(h) = \lambda^d\{\|t\|_\infty \le 1/2 : \sqrt{1-\Re c(t)} < h\}$ ($c(t)$ – характеристическая функция величины распределенной в соответствии с $\mathcal{P}$). В этой же работе устанавливается необходимость и достаточноть конечности интеграла

$$\int\limits_0 \left[\log\left(1+\frac{1}{m(h/2)}\right)\right]^{1/2} dh$$

для справедливости центральной предельной теоремы для $C_M(t) = \sqrt{n}(c_M(t) - c(t))$.

Чтобы доказать, что $\dot{V}_M^I(u)+\dot{V}_M^{II}(v)$ удовлетворяет условию центральной предельной теоремы, будем пользоваться следующим d-мерным обобщением теоремы Биллингсли, Theorem 12.3 [9]:

Теорема (Критерий плотности)**.** *Последовательность X_n является плотной, если последовательность $\{X_n(0)\}$ плотна, а также существуют такие константы C и $\delta > 0$, что равномерно для всех n выполнено:*

$$\mathbb{E}\,|X_n(t_1) - X_n(t_2)|^2 \le C\|t_1 - t_2\|^{1+\delta}.$$

Иными словами, мы хотим получить оценку на

$$\mathbb{E}\,|(\overset{\bullet}{V}{}^{I}_{M}(u_1) - \overset{\bullet}{V}{}^{I}_{M}(u_2)) + (\overset{\bullet}{V}{}^{II}_{M}(v_1) - \overset{\bullet}{V}{}^{II}_{M}(v_2))|^2.$$

В силу общего неравенства

$$\mathbb{E}\,(X+Y)^2 \le 2\mathbb{E}\,X^2 + 2\mathbb{E}\,Y^2,$$

наша задача упрощается, и достаточно доказать независимо неравенства для $(\overset{\bullet}{V}{}^{I}_{M}(u_1) - \overset{\bullet}{V}{}^{I}_{M}(u_2))$ и $(\overset{\bullet}{V}{}^{II}_{M}(v_1) - \overset{\bullet}{V}{}^{II}_{M}(v_2))$. Разберемся с первым:

$$\begin{aligned}\mathbb{E}\,|\overset{\bullet}{V}{}^{I}_{M}(u_1) - \overset{\bullet}{V}{}^{I}_{M}(u_2)|^2 &\le 2\int\limits_{\Pi_R} (1-\cos\langle u_1-u_2, x\rangle)d\mathcal{P}(x)\\ &\le 2\int\limits_{\Pi_R} |\,\langle u_1-u_2, x\rangle\,|^{1+\delta} d\mathcal{P}(x) \le C\|u_1-u_2\|^{1+\delta},\end{aligned}$$

где δ выбрана так, чтобы

$$1-\cos(t) < |t|^{1+\delta}.$$

На самом деле, нам подходит любое $\delta \in [0, 0.1]$, так как $1-\cos(t) = o(t)$ в окрестности нуля.

Заметим также, что $\mathcal{P}$ является мерой контроля меры $de^{i2\tau_r}\mathbb{1}_{S_r\le x}$. Действительно, обозначим эту меру $Z = de^{i2\tau_r}\mathbb{1}_{S_r\le x}$, пусть $A_1, A_2 \subset \Pi_R$, тогда (пользуясь тем, что мера Z центрирована):

$$\begin{aligned}\mathrm{cov}(Z(A_1), Z(A_2)) = \mathbb{E}\,_\tau e^{2i\tau_r}\mathcal{P}(A_1\setminus A_2)\cdot 0 + \mathbb{E}\,_\tau \mathcal{P}(A_1\cap A_2)e^{2i\tau_r}\overline{e^{2i\tau_r}} +\\ + \mathbb{E}\,_\tau 0\cdot \overline{e^{2i\tau_r}\mathcal{P}(A_2\setminus A_1)} = \mathcal{P}(A_2\cap A_2).\end{aligned}$$

Поэтому предыдущее неравенство работает без изменений:

$$\begin{aligned}\mathbb{E}\,|\overset{\bullet}{V}{}^{II}_{M}(v_1) - \overset{\bullet}{V}{}^{II}_{M}(v_2)|^2 &\le 2\int\limits_{\Pi_R} (1-\cos\langle v_1-v_2, x\rangle)d\mathcal{P}(x)\\ &\le 2\int\limits_{\Pi_R} |\,\langle v_1-v_2, x\rangle\,|^{1+\delta} d\mathcal{P}(x) \le C\|v_1-v_2\|^{1+\delta},\end{aligned}$$

что завершает доказательство теоремы. □

Список литературы

[1] A. Rahimi, B. Recht. *Random features for large scale kernel machines.* NIPS. 20. 1177-1184 (2007).

[2] Т.Д. Мосеева, М.А. Лифшиц. *Описание метода Random Fourier Features.* Рабочие материалы.

[3] S. Csörgő. *Multivariate empirical characteristic functions.* Z. Wahrscheinlichkeitstheorie verw Gebiete 55, 203–229 (1981).

[4] M.T. Lacey. *Laws of the iterated logarithm for the empirical characteristic function,* The Annals of Probability, Ann. Probab. 17(1), 292-300.

[5] M.B. Marcus. *Weak convergence of the empirical characteristic function,* The Annals of Probability, Ann. Probab. 9(2), 194-201.

[6] S. Csörgő. *Limit behaviour of the empirical characteristic function,* The Annals of Probability, Ann. Probab. 9(1), 130-144 (1981).

[7] J. Yukich. *Weak convergence of the empirical characteristic function.* Proc. Amer. Math. Soc. 95(3), 470-473 (1985).

[8] X. Fernique. *Continuité et théorème central limite pour les transformées de Fourier des mesures aléatoires du second ordre.* Z. Wahrsch. verw. Geb. 42, 57–66 (1978).

[9] P. Billingsley. *Convergence of probability measures.* Wiley (1968).