ОБ ОЦЕНКЕ ПЛОТНОСТИ РАСПРЕДЕЛЕНИЯ С ПОМОЩЬЮ РЯДА ФУРЬЕ[1]

Беломестный Д. В.[2], Иосипой Л. С.[3]
(*Национальный исследовательский университет «Высшая школа экономики», Москва*)

Рассматривается классическая статистическая задача оценки плотности распределения по выборке из этого распределения. Данная задача возникает в различных прикладных областях при попытке изучить вероятностную структуру некоторого случайного процесса. Например, с помощью оценки плотности можно идентифицировать некоторую структуру в сложной системе, а затем сделать выводы о неизвестных параметрах этой системы. В данной работе предлагается новый способ оценивания плотности распределения, основанный на аппроксимации логарифма плотности рядом Фурье, коэффициенты которого вычисляются с помощью решения некоторой системы линейных уравнений. Анализ теоретических свойств такой оценки является основной задачей данной работы. Основными результатами данной работы являются оценка отклонения в супремум-норме и оценка расстояния Кульбака – Лейблера между аппроксимацией плотности и истинным значением. Полученные оценки являются параметрическими и имеют порядок сходимости «с большой верятностью» $O(1/\sqrt{N})$, что является стандартными порядками в задачах параметрического оценивания. Константы в порядках получены с точностью до некоторого абсолютного множителя, т.е. исследована ее зависимость от всех параметров задачи. В качестве численного примера рассматривается оценка плотности распределения Коши.

Ключевые слова: оценка плотности, ряд Фурье, расстояние Кульбака – Лейблера.

1. Введение

В данной статье рассматривается задача оценки плотности распределения вероятности по выборке из этого распределения. Данная задача возникает в различных прикладных областях при попытке изучить вероятностную структуру некоторого случайного процесса. Эта задача является критически важной, по-

---

[1] *Результаты данной статьи получены при поддержке гранта РНФ №19-71-30020.*

[2] *Денис Витальевич Беломестный, к.ф.-м.н., профессор* (dbelomestny@hse.ru).
[3] *Леонид Сергеевич Иосипой, м.н.с.* (liosipoi@hse.ru).

скольку с помощью плотности распределения можно описывать систему, из которой были получены данные. Таким образом, оценка плотности может быть использована для того, чтобы идентифицировать некоторую структуру в сложной системе, а затем сделать выводы о неизвестных параметрах этой системы.

Оценка плотности распределения является классической статистической задачей. Для ее решения было разработано множество различных подходов, которые можно разделить на параметрические и непараметрические. В параметрическом подходе обычно предполагается, что плотность известна с точностью до конечного числа параметров, в то время как в непараметрическом подходе обычно предполагается, что плотность принадлежит некоторому бесконечному набору кривых (например, классу непрерывно дифференцируемых функций). Литература на данную тему достаточно обширна, см. [7, 4, 2, 1, 14, 6] и список литературы в данных работах. Хорошим введением в данную задачу являются [5, глава 22] и [14, главы 1, 2].

Обозначим через $p: \mathbb{R} \to \mathbb{R}$ неизвестную плотность распределения, которую необходимо оценить по выборке $X_1, \dots, X_N$ (элементы выборки получены независимо). В данной работе предлагается новая параметрическая оценка $p(x)$, которая заключается в приближении функции $\log p(x)$ с помощью ряда Фурье. Коэффициенты Фурье данной аппроксимации вычисляются через коэффициенты Фурье плотности $p(x)$, которые, в свою очередь, оцениваются с помощью выборки $X_1, \dots, X_N$. Анализ теоретических свойств такой оценки является основной задачей данной работы. Обратим внимание, что плотность любого распределения является функцией неотрицательной, что усложняет применение известных методов из классической теории аппроксимации функций. Например, при использовании обычных или тригонометрических полиномов оценка плотности может начать принимать отрицательные значения. Этим обосновывается подход, в котором сначала оценивается $\log p(x)$, а затем от данной оценки берется экспонента.

Обозначения. В данной работе евклидова векторная норма обозначается как $\|\cdot\|$. Так же обозначается и подчиненная ей матричная норма, которую мы будет называть операторной. Для подмножества действительных чисел $T \subset \mathbb{R}$ через $C^s(T)$ мы будем обозначать множество всех непрерывно дифференцируемых действительнозначных функций на T. Для действительнозначной функции $h: T \to \mathbb{R}$ мы положим $\| h \|_\infty = \sup_{x \in T} |h(x)|$ (супремум-норма) и $\| h \|_{L^k(T)} = (\int_T |h(x)|^k dx)^{1/k}$ (L^k-норма), где $k \in \mathbb{N}$. Множество всех функций, для которых $\| h \|_{L^k(T)} < \infty$, мы будем обозначать через $L^k(T)$. Для плотности $p(x)$ некоторого абсолютно непрерывного распределения на $\mathbb{R}$ мы положим $\| h \|_{L^k(T,p)} = (\int_T |h(x)|^k p(x) dx)^{1/k}$. Расстояние Кульбака – Лейблера между плотностями распределения вероятностей p_1 и p_2 будет обозначаться через $KL(p_1, p_2)$. Вероятность и математическое ожидание будут обозначаться соответственно через $\mathbb{P}$ и $\mathbb{E}$.

2. Основные результаты

Чтобы избежать некоторых технических деталей, мы будем предполагать, что плотность $p(x)$ имеет конечный носитель, т.е. что существует такой отрезок $T = [-\tau, \tau]$, $\tau > 0$, для которого $p(x) > 0$ для $x \in T$ и $p(x) = 0$ для $x \notin T$, а также $p(\tau) = p(-\tau)$. Но рассмотренный ниже метод оценки можно применять и в общем случае (необходимо рассмотреть функцию $p_0(x)$, которая для некоторого большого $\tau > 0$ удовлетворяет $p_0(\tau) = p_0(-\tau)$ и которая близка к $p(x)$ на $T = [-\tau, \tau]$, и оценить ошибку аппроксимации $p(x)$ с помощью $p_0(x)$).

Рассмотрим идею, предложенную в [6], которая связывает коэффициенты Фурье функций $p(x)$ и $g(x) = \log p(x)$. Предположим, что обе функции могут быть представлены на T с помощью своих рядов Фурье:

$$p(x) = \sum_{k=-\infty}^{\infty} b_k e^{ik\frac{\pi}{\tau}x}, \quad \text{где} \quad b_k = \frac{1}{2\tau} \int_{-\tau}^{\tau} p(x) e^{-ik\frac{\pi}{\tau}x} dx,$$

$$g(x) = \sum_{k=-\infty}^{\infty} c_k e^{ik\frac{\pi}{\tau}x}, \quad \text{где} \quad c_k = \frac{1}{2\tau} \int_{-\tau}^{\tau} g(x) e^{-ik\frac{\pi}{\tau}x} dx.$$

Тогда коэффициенты b_k и c_k удовлетворяют следующему отношению: для любого $k \in \mathbb{Z}$,

$$2\tau k b_k = k \int_T p(x) e^{-ik\frac{\pi}{\tau}x} dx =$$

$$= i\frac{\tau}{\pi}\left(p(x)e^{-ik\frac{\pi}{\tau}x}|_{-\tau}^{\tau} - \int_T p'(x)e^{-ik\frac{\pi}{\tau}x}dx\right) =$$
$$= -i\frac{\tau}{\pi}\int_T p(x)g'(x)e^{-ik\frac{\pi}{\tau}x}dx =$$
$$= \sum_{j=-\infty}^{\infty} jc_j \int_T p(x)e^{-i(k-j)\frac{\pi}{\tau}x}dx =$$
$$= 2\tau\sum_{j=-\infty}^{\infty} jc_j b_{k-j}.$$

То есть $\sum_{j=-\infty}^{\infty} jb_{k-j}c_j = kb_k$, что можно рассматривать как бесконечную систему линейных уравнений относительно c_j.

В [6] было предложено «обрезать» данную систему уравнений, что приводило к системе $Ax = b$, где

$$(1)\quad A = \begin{pmatrix} b_0 & b_{-1} & \dots & b_{-2n} \\ b_1 & b_0 & \dots & b_{-2n+1} \\ \vdots & \vdots & \ddots & \vdots \\ b_{2n} & b_{2n-1} & \dots & b_0 \end{pmatrix},$$

$$x = \begin{pmatrix} -nc_{-n} \\ -(n-1)c_{-(n-1)} \\ \vdots \\ nc_n \end{pmatrix},$$

$$b = \begin{pmatrix} -nb_{-n} \\ -(n-1)b_{-(n-1)} \\ \vdots \\ nb_n \end{pmatrix}.$$

Решив данную систему, мы сможем найти все $c_{-n}, \dots, c_n$, кроме c_0, так как центральный элемент вектора x равен 0, $c_0 = 0$. Данное обстоятельство не приводит к проблемам, так как благодаря теплицевой структуре системы и наличию некоторой симметрии (подробнее см. [9, с. 317]) решение x тоже будет содержать 0 на данном месте. Обозначим полученную аппроксимацию через

$$g_n(x) := \sum_{k=-n}^{n} c_k e^{ik\frac{\pi}{\tau}x} = c_0 + \sum_{1\le|k|\le n} c_k e^{ik\frac{\pi}{\tau}x},$$

где оставшийся неизвестный коэффициент c_0 найдем из условия

$$\int_T e^{g_n(x)}dx = \int_T p(x)dx = 1,$$

что эквивалентно

$$c_0 = -\log\left(\int_T e^{\left(\sum_{1\le|k|\le n} c_k e^{ik\frac{\pi}{\tau}x}\right)}dx\right).$$

Как уже упоминалось, в данной работе мы изучаем свойства подобной аппроксимации в случае, когда коэффициенты b_k неизвестны и оцениваются с помощью

$$\tilde{b}_k = \frac{1}{2\tau N}\sum_{j=1}^{N} e^{ik\frac{\pi}{\tau}X_j},$$

где, напомним, $X_1, \ldots, X_N$ – это выборка. Данные оценки являются несмещенными и состоятельными. Определим новую «возмущенную» систему уравнений

$$(2)\quad \tilde{A} = \begin{pmatrix} \tilde{b}_0 & \tilde{b}_{-1} & \ldots & \tilde{b}_{-2n} \\ \tilde{b}_1 & \tilde{b}_0 & \ldots & \tilde{b}_{-2n+1} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{b}_{2n} & \tilde{b}_{2n-1} & \ldots & \tilde{b}_0 \end{pmatrix},$$

$$\tilde{x} = \begin{pmatrix} -n\tilde{c}_{-n} \\ -(n-1)\tilde{c}_{-(n-1)} \\ \vdots \\ n\tilde{c}_n \end{pmatrix},$$

$$\tilde{b} = \begin{pmatrix} -n\tilde{b}_{-n} \\ -(n-1)\tilde{b}_{-(n-1)} \\ \vdots \\ n\tilde{b}_n \end{pmatrix}.$$

и аппроксимацию

$$(3)\quad \tilde{g}_n(x) = \sum_{k=-n}^{n} \tilde{c}_k e^{ik\frac{\pi}{\tau}x},$$

где коэффициент $\tilde{c}_0$ определяется из аналогичного условия,

$$\tilde{c}_0 = c_0 = -\log\left(\int_T e^{\left(\sum_{1\le|k|\le n} c_k e^{ik\frac{\pi}{\tau}x}\right)} dx\right).$$

Основными результатами данной работы являются оценка максимального отклонения аппроксимации $\tilde{p}_n = e^{\tilde{g}_n}$ от истинной плотности $p = e^g$, а также оценка расстояния Кульбака – Лейблера между ними. Данные результаты сформулированы ниже.

Теорема 1 (Супремум-норма). Пусть $p \in C^s(T)$, $s \geqslant 2$ и $|p^{(s)}(x)| < \infty$. Пусть также $m = \operatorname{ess\,inf}_{x\in T} p(x)$, $M = \operatorname{ess\,sup}_{x\in T} p(x)$ и $0 < m \le M < \infty$. Если $N \geqslant \left(6n\log(\frac{10n}{\delta})\right)/(\tau^2 m^2)$, тогда с вероятностью не менее $1-\delta$

$$\| g - \tilde{g}_n \|_\infty \lesssim \frac{n^2}{\tau m^2} \sqrt{\frac{\log\left(\frac{n}{\delta}\right)}{N}} + \sqrt{\frac{M\tau}{m}} \frac{1}{n^s},$$

где $\lesssim$ означает неравенство с точностью до некоторой абсолютной константы.

Подобный теоретический результат уже был получен в работе [9], но в детерминированной постановке, когда коэффициенты Фурье $p(x)$ считались известными и их не нужно было оценивать. Обобщение, рассмотренное в данной статье, требует получения равномерной концентрации оценок коэффициентов Фурье $p(x)$, а также исследования устойчивости решения системы линейных уравнений, с помощью которой вычисляются коэффициенты Фурье $\log p(x)$. Обратим также внимание на условие $N \geqslant (6n\log(\frac{10n}{\delta}))/(\tau^2 m^2)$, которое отсутствует в [9] и которое необходимо, чтобы гарантировать близость решений систем уравнений (1) и (2).

С помощью теоремы 1 можно получить следующие следствия.

***Следствие 1* (*Расстояние в супремум-норме*).** В условиях теоремы 1

$$\| \tilde{p}_n - p \|_\infty \lesssim M \left(\frac{n^2}{\tau m^2} \sqrt{\frac{\log\left(\frac{n}{\delta}\right)}{N}} + \sqrt{\frac{M\tau}{m}} \frac{1}{n^s} \right).$$

***Следствие 2* (*Расстояние Кульбака – Лейблера*).** В условиях теоремы 1

$$\mathrm{KL}(p, \tilde{p}_n) \lesssim 2\tau M \left(\frac{n^2}{\tau m^2} \sqrt{\frac{\log(\frac{n}{\delta})}{N}} + \sqrt{\frac{M\tau}{m}} \frac{1}{n^s} \right).$$

Построенная оценка $\tilde{p}_n$ является параметрической оценкой плотности p, и порядки сходимости $O(1/\sqrt{N})$, полученные в следствиях 1 и 2, являются стандартными в задачах параметрического оценивания. В случае непараметрического оценивания порядки сходимости получаются хуже. Так, например, в работе [4, теорема 4] была получена следующая минимаксная оценка (т.е. для наилучшей оценки $\tilde{p}_n$ наихудшей плотности p из некоторого класса): $\mathbb{E} \| \tilde{p}_n - p \|_\infty \lesssim n^{-\frac{s}{2s+1}}$. Здесь оценивается ожидаемое уклонение $\| \tilde{p}_n - p \|_\infty$, а не его значение «с большой вероятностью», как в результатах выше. В последнее время интерес представляют именно оценки «с большой вероятностью» как более информативные. Упомянутый результат из [4] был обобщен в работе [10] на оценку «с большой вероятностью», которая также равна $O(n^{-\frac{s}{2s+1}})$.

Пример. В качестве примера, рассмотрим задачу оценивания плотности Коши $p(x) = \frac{1}{\pi(1+x^2)}$. Зафиксируем параметры $\tau = 3\pi$ и $n = 5$. Оценивать $p(x)$ будем с помощью выборки размера $N = 100$ и $N = 100000$. Для этого, согласно предложенному алгоритму, сначала оценим выборочные коэффициенты Фурье $\tilde{b}_k$ и решим систему (2). После этого найдем $\tilde{c}_0$, вычислим $\tilde{g}_n$ по (3) и положим $\tilde{p}_n = e^{\tilde{g}_n}$. Вне отрезка $T = [-3\pi, 3\pi]$, по определению, $\tilde{p}_n(x) = 0$. Результаты приведены на рис. 1 и 2.

Остальная часть данной статьи посвящена доказательству данных результатов.

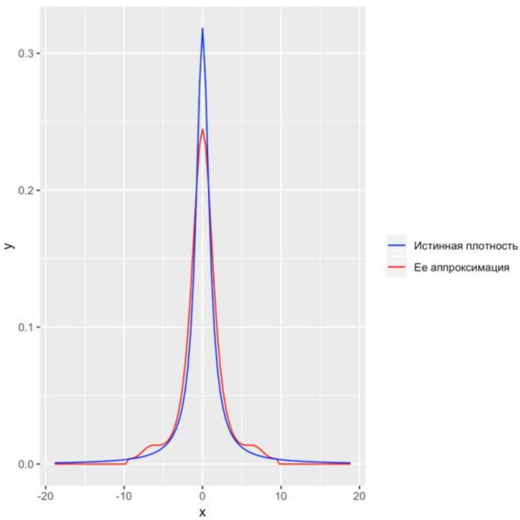

Рис. 1. Оценивание плотности Коши по $N = 100$ наблюдениям

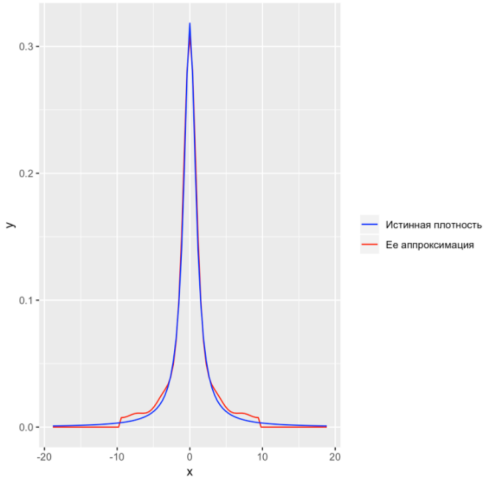

Рис. 2. Оценивание плотности Коши по $N = 100000$ *наблюдениям*

3. Доказательства основных результатов

Доказательство теоремы 1, следствия 1 и следствия 2 разделено на следующие этапы. В разделе 3.1 мы получим оценки на близость всех коэффициентов b_k и $\tilde{b}_k$, которые участвуют в системах уравнений (1) и (2). Затем в разделе 3.2 мы проведем общий анализ стабильности системы линейных уравнений при возмущении. В разделе 3.3 мы оценим евклидово расстояние между решениями x и $\tilde{x}$ систем (1) и (2) соответственно. И, наконец, в разделе 3.4 мы докажем теорему 1, а в разделе 3.5 и разделе 3.6 – следствие 1 и следствие 2 соответственно.

3.1. КОНЦЕНТРАЦИЯ ЭМПИРИЧЕСКИХ КОЭФФИЦИЕНТОВ ФУРЬЕ

В этом разделе мы получим результат, который описывает, насколько хорошо эмпирические коэффициенты Фурье

(4) $\tilde{b}_k = \frac{1}{2\tau N} \sum_{j=1}^{N} e^{ik\frac{\pi}{\tau}X_j}$,

посчитанные по выборке $X_1, \dots, X_n$, близки к истинным коэффициентам Фурье b_k для k $= -2n, \dots, 2n$.

Лемма 1. Пусть b_k являются коэффициентами Фурье функции $p(x)$, а $\tilde{b}_k$ – их эмпирические оценки (4), посчитанные по выборке $X_1, \ldots, X_N$. Тогда для всех b_k из диапазона $k = -2n, \ldots, 2n$ выполнено

$$\left|\tilde{b}_k - b_k\right| \lesssim \sqrt{\frac{\log\left(\frac{10n}{\delta}\right)}{2\tau^2 N}} \quad \text{с вероятностью не менее } 1 - \delta.$$

Доказательство. Обозначим $Y_{k,j} = (2\tau)^{-1} e^{ik\frac{\pi}{\tau}X_j}$ для $j = 1, \ldots, N$. Заметим, что $\tilde{b}_k = \frac{1}{N}\sum_{j=1}^{N} Y_{k,j}$ и $b_k = \mathbb{E}Y_{k,j}$ для всех $j = 1, \ldots, N$. По неравенству Хеффдинга (см. [11]) для любого $k = -2n, \ldots, 2n$,

$$\mathbb{P}\left(\left|\frac{1}{N}\sum_{j=1}^{N} (Y_{k,j} - \mathbb{E}Y_{k,j})\right| \geq t\right) \leq 2e^{-2\tau^2 t^2 N}.$$

Суммируя вероятности для всех $k = -2n, \ldots, 2n$, мы получим

$$\mathbb{P}\left(\left|\tilde{b}_k - b_k\right| \geq t \text{ для всех } k = -2n, \ldots, 2n\right) \leq$$
$$\leq 2(4n+1)e^{-2\tau^2 t^2 N} \leq 10ne^{-2\tau^2 t^2 N}.$$

Следовательно,

$$\mathbb{P}\left(\left|\tilde{b}_k - b_k\right| \geq \sqrt{\frac{\log\left(\frac{10n}{\delta}\right)}{2\tau^2 N}} \text{ для всех } k = -2n, \ldots, 2n\right) \leq \delta,$$

что может быть переписано как

$$\left|\tilde{b}_k - b_k\right| \lesssim \sqrt{\frac{\log\left(\frac{10n}{\delta}\right)}{2\tau^2 N}}$$

для всех $k = -2n, \ldots, 2n$ с вероятностью не менее $1 - \delta$. Лемма доказана.

3.2. АНАЛИЗ УСТОЙЧИВОСТИ СЛАУ

В данном разделе мы проведем анализ устойчивости решения системы линейных уравнений $Ax = b$ при возмущении матрицы коэффициентов A и столбца свободных членов b. Данный анализ является общим и верен для любой векторной нормы $\|\cdot\|$ и подчиненной ей матричной норме.

Итак, пусть x является решением системы $Ax = b$, а $\tilde{x}$ – решением возмущенной системы $\tilde{A}\tilde{x} = \tilde{b}$. Определим

$$\Delta A = A - \tilde{A} \quad \text{и} \quad \Delta b = b - \tilde{b}.$$

Мы покажем, что при малых значениях $\| \Delta A \|$ и $\| \Delta b \|$ значение $\| x - \tilde{x} \|$ также будет мало. Рассмотрим сначала вспомогательные леммы.

Лемма 2. Пусть I будет единичной матрицей и ΔI будет любой матрицей с $\| \Delta I \| < 1$. Тогда матрица $I - \Delta I$ является невырожденной и

$$\| (I - \Delta I)^{-1} \| \leq \frac{1}{1 - \| \Delta I \|}.$$

Более того,

$$\| I - (I - \Delta I)^{-1} \| \leq \frac{\| \Delta I \|}{1 - \| \Delta I \|}.$$

Доказательство. Так как $\| x \| \leq \| x - \Delta I x \| + \| \Delta I x \|$, мы имеем

$$\| (I - \Delta I)x \| \geq \| x \| - \| \Delta I x \| \geq \| x \| - \| \Delta I \| \| x \| =$$
$$= (1 - \| \Delta I \|) \| x \| > 0$$

для всех $x \neq 0$. Следовательно, матрица $I - \Delta I$ является невырожденной. Рассмотрим равенство $(I - \Delta I)(I - \Delta I)^{-1} = I$, с помощью которого мы получим

(5) $(I - \Delta I)^{-1} = I + \Delta I (I - \Delta I)^{-1}$.

Следовательно,

$$\| (I - \Delta I)^{-1} \| \leq \| I \| + \| \Delta I \| \| (I - \Delta I)^{-1} \|$$
$$= 1 + \| \Delta I \| \| (I - \Delta I)^{-1} \|.$$

Разрешая это неравенство относительно $\| (I - \Delta I)^{-1} \|$, мы докажем первое утверждение леммы. Далее, равенство (5) также влечет

$$I - (I - \Delta I)^{-1} = -\Delta I (I - \Delta I)^{-1}.$$

Используя первое утверждение леммы, мы имеем

$$\| I - (I - \Delta I)^{-1} \| \leq \| \Delta I \| \| (I - \Delta I)^{-1} \| \leq \frac{\| \Delta I \|}{1 - \| \Delta I \|}.$$

Лемма доказана.

Лемма 3. Пусть A является невырожденной матрицей, а ΔA – любой матрицей с $\| A^{-1} \Delta A \| < 1$. Тогда $A + \Delta A$ является невырожденной и

$$\| (A - \Delta A)^{-1} \| \leq \frac{\| A^{-1} \|}{1 - \| A^{-1} \Delta A \|}.$$

Более того,

$$\| A^{-1} - (A - \Delta A)^{-1} \| \leq \frac{\| A^{-1} \|}{1 - \| A^{-1} \Delta A \|} \| A^{-1} \Delta A \|.$$

Доказательство. Рассмотрим равенство $A + \Delta A = A(I + A^{-1}\Delta A)$. Так как $\| A^{-1}\Delta A \| < 1$, по лемме 2 матрица $I + A^{-1}\Delta A$ является невырожденной и, следовательно, $A + \Delta A$ тоже невырождена. Рассмотрим далее равенство

(6) $(A + \Delta A)^{-1} = (I + A^{-1}\Delta A)^{-1}A^{-1}$.

Используя опять лемму 2, мы получим

$$\| (A + \Delta A)^{-1} \| = \| (I + A^{-1}\Delta A)^{-1}A^{-1} \| \leq$$
$$\leq \| (I + A^{-1}\Delta A)^{-1} \| \| A^{-1} \| \leq \frac{\| A^{-1} \|}{1 - \| A^{-1}\Delta A \|},$$

что доказывает первое утверждение леммы. Чтобы оценить норму разности матриц A^{-1} и $(A + \Delta A)^{-1}$, мы используем равенство (6) и лемму 2 еще раз.

$$\| A^{-1} - (A + \Delta A)^{-1} \| \leq \| I - (I + A^{-1}\Delta A)^{-1} \| \| A^{-1} \| \leq$$
$$\leq \frac{\| A^{-1} \|}{1 - \| A^{-1}\Delta A \|} \| A^{-1}\Delta A \|,$$

что доказывает второе утверждение леммы.

Предложение 1. Пусть x является решением системы $Ax = b$, а $\tilde{x}$ является решением возмущенной системы $\tilde{A}\tilde{x} = \tilde{b}$. Обозначим $\Delta A = A - \tilde{A}$ и $\Delta b = b - \tilde{b}$. Если $\| A^{-1} \| \| \Delta A \| < 1/2$, тогда

$$\| \tilde{x} - x \| \leq 2 \| A^{-1} \| (\| A^{-1} \| \| b \| \| \Delta A \| + \| \Delta b \|).$$

Доказательство. Мы имеем

$$\| \tilde{x} - x \| = \| (A + \Delta A)^{-1}(b + \Delta b) - A^{-1}b \| \leq$$
$$\leq \| (A + \Delta A)^{-1}b - A^{-1}b \| + \| (A + \Delta A)^{-1}\Delta b \| \leq$$
$$\leq \| (A + \Delta A)^{-1} - A^{-1} \| \cdot \| b \| + \| (A + \Delta A)^{-1} \| \cdot \| \Delta b \|.$$

По условию $\| A^{-1}\Delta A \| \leq \| A^{-1} \| \| \Delta A \| < 1/2$. Следовательно, применив лемму 3, мы получим

$$\| \tilde{x} - x \| \leq \frac{\| A^{-1} \|}{1 - \| A^{-1}\Delta A \|} (\| A^{-1}\Delta A \| \| b \| + \| \Delta b \|) \leq$$
$$\leq 2 \| A^{-1} \| (\| A^{-1} \| \| b \| \| \Delta A \| + \| \Delta b \|).$$

Предложение доказано.

3.3. ВСПОМОГАТЕЛЬНЫЕ ОЦЕНКИ

Приведем без доказательства вспомогательные утверждения, которые будут использованы позже. Первое утверждение связывает спектр матрицы A, определенной в (1), с множеством значений плотности $p(x)$.

Лемма 4 **([*13, теорема 1.2*]).** Пусть $p \in L^1(T)$ и пусть $m = \operatorname{ess\,inf}_{x\in T} p(x)$, $M = \operatorname{ess\,sup}_{x\in T} p(x)$. Определим для любого целого n матрицу $A = (b_{j-k})_{j,k=1,\dots,n}$, где b_k – коэффициенты Фурье функции $p(x)$. Тогда матрица A является эрмитовой и все ее собственные значения лежат в интервале (m, M).

В следующем утверждении показано, что функция g_n, которая является аппроксимацией функции g, является решением некоторой оптимизационной задачи.

Лемма 5 **([*9, теорема 2*]).** Пусть $p \in C^s(T)$, $s \geqslant 1$ и $|p^{(s)}(x)| < \infty$. Пусть также $m = \operatorname{ess\,inf}_{x\in T} p(x)$, $M = \operatorname{ess\,sup}_{x\in T} p(x)$ и $0 < m \le M < \infty$. Тогда функция g_n, определенная в (3), является единственным решением оптимизационной задачи

$$\begin{cases} \min\limits_{h\in F_n} \| (g)' - (h)' \|^2_{L^2(T,p)}, \\ \int_T e^{g_n(x)} = 1, \end{cases}$$

где F_n – линейная оболочка множества функций $\{1, e^{\pm i\frac{\pi}{\tau}x}, \ \dots, e^{\pm in\frac{\pi}{\tau}x}\}$.

В следующем предложении мы воспользуемся предыдущими двумя леммами, чтобы охарактеризовать евклидову норму разности векторов x и $\tilde{x}$, решениями систем уравнений (1) и (2) соответственно.

Предложение 2. Пусть $p \in C^s(T)$, $s \geqslant 2$ и $|p^{(s)}(x)| < \infty$. Пусть также для $m = \operatorname{ess\,inf}_{x\in T} p(x)$, $M = \operatorname{ess\,sup}_{x\in T} p(x)$ и $0 < m \le M < \infty$. Если $N \geqslant \left(6n \log(\frac{10n}{\delta})\right)/(\tau^2 m^2)$, тогда с вероятностью не менее $1 - \delta$,

$$\| \tilde{x} - x \| \lesssim \frac{1}{m^2\tau}\sqrt{\frac{n^3 \log(\frac{n}{\delta})}{N}},$$

где $\lesssim$ означает неравенство с точностью до некоторой абсолютной константы.

Доказательство. Согласно предложению 1, если
$\| A^{-1} \| \| \Delta A \| < 1/2$, то

(7) $\| \tilde{x} - x \| \le 2 \| A^{-1} \| (\| A^{-1} \| \| b \| \| \Delta A \| + \| \Delta b \|)$.

Оценим каждый множитель в этой формуле отдельно. Напомним, что у гладкой функции $p \in C^s(T)$ с $|p^{(s)}(x)| < \infty$

коэффициенты Фурье убывают как $b_k \le c|k|^{-s}$, где c – это некоторая константа (см., например, [3]).

- $\| A^{-1} \|$: по лемме 4 $\| A^{-1} \| \le 1/m$.
- $\| b \|$: $\| b \|^2 = \sum_{k=-n}^{n} k^2 b_k^2 \le 2c \sum_{k=1}^{n} k^{-(2s-2)} \le$
  $\le 2c \sum_{k=1}^{\infty} k^{-2} \le c\pi^2/3$.
- $\| \Delta A \|$: используя лемму 1, с вероятностью не менее $1-\delta$
  $$\| \Delta A \| \le \sqrt{2n+1} \max_{1 \le i,j \le 2n+1} |(\Delta A)_{i,j}| \le \sqrt{\frac{(2n+1)\log(\frac{10n}{\delta})}{2\tau^2 N}}.$$
- $\| \Delta b \|$: используя лемму 1, с вероятностью не менее $1-\delta$
  $$\| \Delta b \|^2 = \sum_{k=-n}^{n} k^2 (\tilde{b}_k - b_k)^2 \le \frac{2\log\left(\frac{10n}{\delta}\right)}{N} \sum_{k=-n}^{n} k^2 \le$$
  $$\le \frac{2(2n^3+3n^2+n)\log(\frac{10n}{\delta})}{3N},$$

где в последнем неравенстве мы использовали формулу Фаульхабера.

Заметим, что условие $N \geqslant (6n\log(\frac{10n}{\delta}))/(\tau^2 m^2)$ влечет $\| A^{-1} \| \le 1/m$. Подставив полученные оценки в (7) (пренебрегая константами для простоты), мы получим

$$\| \tilde{x} - x \| \lesssim \frac{1}{m}\left(\frac{1}{m}\sqrt{\frac{n\log(\frac{n}{\delta})}{\tau^2 N}} + \sqrt{\frac{n^3\log(\frac{n}{\delta})}{\tau^2 N}}\right) \lesssim \frac{1}{m^2\tau}\sqrt{\frac{n^3\log(\frac{n}{\delta})}{N}}.$$

Предложение доказано.

3.4. ДОКАЗАТЕЛЬСТВО ТЕОРЕМЫ 1

Заметим, что по построению $\int_T \tilde{p}_n(x)dx = \int_T p(x)dx = 1$. Так как обе функции $\tilde{p}_n(x)$ и $p(x)$ непрерывны, то существует $x_0 \in T$, что $\tilde{p}_n(x_0) = p(x_0)$. Следовательно, $\tilde{g}_n(x_0) = g(x_0)$. Используя этот факт и неравенство Коши – Буняковского, мы получим

$$\| \tilde{g}_n - g \|_\infty = \sup_{x \in T} \left| \int_{x_0}^{x} \left(\tilde{g}'_n(x) - g'(x)\right)dx \right| \le$$
$$\le \int_T |\tilde{g}'_n(x) - g'(x)|dx \le \sqrt{2\tau} \| \tilde{g}'_n - g' \|_{L^2(T)}.$$

По неравенству треугольника

$$\| \tilde{g}'_n - g' \|_{L^2(T)} \le \| \tilde{g}'_n - g'_n \|_{L^2(T)} + \| g'_n - g' \|_{L^2(T)}.$$

Оценим каждое слагаемое отдельно. Для первого слагаемого, с помощью неравенства Коши – Буняковского и предложения 2, мы получим

$$\| \tilde{g}'_n - g'_n \|_{L^2(T)} = \frac{\pi}{\tau}\left(\int_T \left|\sum_{k=-n}^{n} k(\tilde{c}_k - c_k)e^{\frac{ik\pi x}{\tau}}\right|^2 dx\right)^{1/2} \leq$$

$$\leq \frac{\pi}{\tau} \| \tilde{x} - x \| \cdot \sqrt{2\tau} \cdot \sqrt{2n+1} \lesssim \frac{n^2}{\tau^{3/2} m^2}\sqrt{\frac{n^3 \log(\frac{n}{\delta})}{\tau^2 N}},$$

где векторы x и $\tilde{x}$ определены в (1) и (2). Для второго слагаемого мы получим

$$\| g'_n - g' \|_{L^2(T)}^2 \leq \frac{1}{m} \| g'_n - g' \|_{L^2(T,p)}^2 \leq$$

$$\leq \frac{1}{m} \inf_{h \in F_n} \| g' - h' \|_{L^2(T,p)}^2 \leq \frac{M}{m} \inf_{h \in F_n} \| g' - h' \|_{L^2(T)}^2.$$

Согласно неравенству Джексона – Стечкина (см., например, [8]),

$$\inf_{h \in F_n} \| g' - h' \|_{L^2(T)} \leq \frac{c}{n^s},$$

где c – некоторая (зависящая только от s) константа. Собрав все полученные оценки, мы получим

$$\| g - \tilde{g}_n \|_\infty \lesssim \frac{n^2}{\tau m^2}\sqrt{\frac{n^3 \log\left(\frac{n}{\delta}\right)}{N}} + \sqrt{\frac{M\tau}{m}}\frac{1}{n^s},$$

с вероятностью не менее $1 - \delta$.

3.5. ДОКАЗАТЕЛЬСТВО СЛЕДСТВИЯ 1

Чтобы оценить расстояние между $\tilde{p}_n(x)$ и $p(x)$ в супремум-норме, заметим, что

$$\| \tilde{p}_n - p \|_\infty \leq \| p \|_\infty \left\|\frac{\tilde{p}_n}{p} - 1\right\|_\infty \leq M\left|e^{\|\tilde{g}_n - g\|_\infty} - 1\right| \lesssim$$

$$\lesssim M \| \tilde{g}_n - g \|_\infty,$$

так как в разложении экспоненты в ряд Тейлора все последующие члены будут иметь больший порядок малости. Подстановка оценки из теоремы 1 завершает доказательство.

3.6. ДОКАЗАТЕЛЬСТВО СЛЕДСТВИЯ 2

Для расстояния Кульбака – Лейблера мы имеем

$$KL(p, \tilde{p}_n) = \int_T p(x) \log\frac{p(x)}{\tilde{p}_n(x)} dx \leq$$

$$\leq M \int_T |g(x) - \tilde{g}_n(x)| dx \leq 2\tau M \| g(x) - \tilde{g}_n(x) \|_\infty.$$

Подстановка оценки из теоремы 1 завершает доказательство.

Литература

1. ГОЛУБЕВ Г.К. *Непараметрическое оценивание гладких плотностей распределения в L_2* // Проблемы передачи информации. – 1992. – №28. – С. 52–62.
2. ЕФРОЙМОВИЧ С.Ю. *Непараметрическое оценивание плотности неизвестной гладкости* // Теория вероятностей и ее применения. – 1985. – №30. – С. 524–534.
3. ЗОРИЧ В.А. *Математический анализ.* – М.: МЦНМО, 2012.
4. ИБРАГИМОВ И.А., ХАСЬМИНСКИЙ Р.З. *Об оценке плотности распределения* // Записки научных семинаров ЛОМИ. – 1986. – №153. – С. 45–59.
5. ЛАГУТИН М.Б. *Наглядная математическая статистика*. – М.: Бином. Лаборатория знаний, 2009.
6. BORWEIN J.M., HUANG W.Z. *A fast heuristic method for polynomial moment problems with Boltzmann-Shannon entropy* // SIAM J.Opt. – 1995. – Vol. 5. – P. 68–99.
7. BRETAGNOLLE J., HUBER C. *Estimation des densites: risque minimax* // Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete – 1979. – Vol. 47. – P. 119–137.
8. DEVORE R.A., LORENTZ G.G. *Constructive Approximation*. – Springer, Berlin, 1993.
9. FASINO D. *Approximation of nonnegative functions by means of exponential trigonometric polynomials* // J. of Computational and Applied Mathematics. – 2002. – Vol. 140. – P. 315–329.
10. JIANG H. *Uniform Convergence Rates for Kernel Density Estimation* // Proc. of the 34th Int. Conference on Machine Learning, PMLR. – 2017. – Vol. 70 – P. 1694–1703.
11. HOEFFDING W. *Probability inequalities for sums of bounded random variables* // J. of the American Statistical Association. – 1963. – Vol. 58, No. 301. – P. 13–30.
12. SERRA S. *On the extreme spectral properties of Toeplitz matrices generated by L^1 functions with several minima (maxima)* // BIT Numerical Mathematics. – 1996. – Vol. 36. – P. 135–142.
13. SERRA S. *On the extreme eigenvalues of Hermitian (block) Toeplitz matrices* // Linear Algebra and its Applications. – 1998. – Vol. 270. – P. 109–129.

14. TSYBAKOV A. *Introduction to Nonparametric Estimation. Springer Series in Statistics* // Springer Series in Statistics. – 2009.
15. TYRTYSHNIKOV E. *A Unifying Approach to Some Old and New Theorems on Distribution and Clustering* // Linear Algebra and its Applications – 1996. – Vol. 232, No. 1. – P. 2–43.

ON DENSITY ESTIMATION VIA FOURIER SERIES

Denis Belomestny, National Research University Higher School of Economics, Moscow, Cand.Sc., professor (dbelomestny@hse.ru).
Leonid Iosipoi, National Research University Higher School of Economics, Moscow, junior research fellow (liosipoi@hse.ru).

Abstract: In this paper, we consider the classical statistical problem of probability density estimation based on a sample from this distribution. This problem naturally arises in many applications when one aims at investigation of a probability structure in a random process. For instance, it is possible to identify some structure in a complex system using density estimation. In this paper, a new approach to estimate a density function is proposed. This approach is based on approximation of a log-density via Fourier series with coefficients obtained by solving a system of linear equations. Analysis of theoretical properties of such estimate is the main purpose of this work. As the main results, we prove bounds on the difference between target density and its approximation in the supremum norm and the Kullback-Leibler divergence. Obtained rates are parametric and have order $O(1/\sqrt{N})$ *with high probability, which is a standard rate in parametric estimation problems. The constants in the rates are obtained up to an absolute factor, which means that we investigated the dependence on all parameters. As a numerical example, we consider a problem of Cauchy density estimation.*

Keywords: density estimation, Fourier series, Kullback–Leibler divergence.

УДК 519.23
ББК 22.172
DOI: 10.25728/ubs.2019.82.2

Статья представлена к публикации членом редакционной коллегии Э.Ю. Калимулиной.

Поступила в редакцию 03.09.2019.
Опубликована 30.11.2019.