Федеральное государственное бюджетное образовательное учреждение высшего образования «Сибирский государственный университет науки и технологий имени академика М.Ф. Решетнёва»

На правах рукописи

Браништи Владислав Владимирович

МЕТОДЫ И АЛГОРИТМЫ НАСТРОЙКИ ПРОЕКЦИОННОЙ ОЦЕНКИ ПЛОТНОСТИ ВЕРОЯТНОСТИ СЛУЧАЙНОГО ВЕКТОРА В УСЛОВИЯХ МАЛЫХ ВЫБОРОК

Специальность 05.13.17 – Теоретические основы информатики

Диссертация на соискание учёной степени
кандидата физико-математических наук

Научный руководитель:
доктор физико-математических наук, доцент
Сафонов Константин Владимирович

Красноярск 2018

Оглавление

Введение

Актуальность темы и степень её разработанности. Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных в условиях неопределённости практически всегда предполагает оценивание функции распределения либо плотности вероятности соответствующих величин. В частности, задача оценивания плотности вероятности случайного вектора возникает при разработке методов распознавания образов, фильтрации, распознавания и синтеза изображений [61, 65].

Имеющиеся в настоящее время методы оценивания функции плотности вероятности можно разделить на параметрические и непараметрические. Параметрические методы используются в случае, когда известна структура закона распределения с точностью до параметров, и задача сводится к построению статистических оценок этих параметров, удовлетворяющих заданным условиям (состоятельность, несмещённость и др.). К числу наиболее разработанных параметрических методов относятся метод моментов, метод максимального правдоподобия, метод минимума χ^2 [46, 86]. Однако часто в практических задачах возникают ситуации, когда структура закона распределения неизвестна, т.е. ситуации *непараметрической неопределённости* [131]. При этом априорная информация о функции плотности вероятности $f(x)$ носит более общий характер, например, $f(x)$ может предполагаться непрерывной на данном отрезке, имеющей n-ю производную, имеющей суммируемый квадрат и т.п. Использование параметрических методов при фактическом несовпадении структуры закона распределения приводит к неудовлетворительным результатам. В этом случае используются методы, получившие название непараметрических.

Исторически первой непараметрической оценкой функции плотности вероятности является гистограмма, исследованная К. Пирсоном в 1895 г. Во второй половине 20-го века интерес к непараметрическим методам значительно возрос, о чём свидетельствует ряд работ, посвящённых следующим оценкам: полиграмма [131], оценка k ближайших соседей [124], оценка Розенблатта – Парзена [25, 20], проекционная оценка [144].

При использовании непараметрических методов представляет интерес исследование сходимости получаемых оценок к истинной функции плотности вероятности по заданной метрике, а также оценка скорости сходимости. В связи с этим возникает задача оптимальной настройки оценок функции плотности вероятности. Так, одной из первых формул для расчёта числа интервалов группирования одинаковой длины при построении гистограммы является формула Стэрджеса [31]. В случае использования полиграммы или оценки k ближайших соседей подлежит настройке численный параметр, определяющий степень сглаженности полученной оценки.

При использовании проекционной оценки плотности вероятности случайного вектора $\mathbf{x} = (x_1, \ldots, x_k)$:

$$\hat{f}(\mathbf{x}) = \sum_{j=0}^{l} a_j \psi_j(\mathbf{x})$$

настройке подлежат как численные параметры l, a_1, $\ldots$, a_l, так и ортогональная система функций $\{\psi_j\}$. При этом оптимального набора функций ψ_j для всех плотностей не существует, так как очевидно, что для данной функции плотности вероятности $f(\mathbf{x})$ оптимальным будет любая система, в которой $\psi_0(\mathbf{x}) \equiv \frac{1}{\|f\|} f(\mathbf{x})$. Тогда $l = 0$ и $a_0 = \|f\|$.

Аналогично, при использовании оценки Розенблатта – Парзена:

$$\hat{f}(x_1, \ldots, x_k) = \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{k} \frac{1}{h_j} \Phi_j \left(\frac{x_j - x_{ij}}{h_j} \right)$$

настройке подлежат как параметры h_1, $\ldots$, h_k, так и «ядерные» функции Φ_1, $\ldots$, Φ_k. Как и в случае проекционной оценки, несложно подобрать оптимальные параметры для данного закона распределения.

Задача настройки непараметрических оценок значительно усложняется при отсутствии информации о законе распределения. В большинстве работ, посвящённых этой проблеме, исследования преимущественно выполняются в предположении, что объём выборки $n \to \infty$. Так, асимптотические свойства проекционной оценки исследуются в работах [144, 118, 38]. Для оценки Розенблатта – Парзена в работе [75] для этого случая получено решение для формы ядра Φ в классе усечённых функций, дифференцируемых в заданном интервале. Однако в анализе данных задачу оценивания плотности часто приходится решать при малых n, например, при обработке биомедицинских данных и данных, касающихся производства и эксплуатации дорогостоящих технических систем. Исследования показали, что результаты, полученные при $n \to \infty$, могут оказаться неоптимальными для малых n.

В этих условиях представляет интерес исследование проекционной оценки, так как, в отличие от других видов непараметрических оценок, например, оценки Розенблатта – Парзена или оценки k ближайших соседей, проекционная оценка не содержит в себе всей выборки и допускает компактное аналитическое выражение. Это оказывается более удобным при теоретическом анализе, в приложениях, а также повышает быстродействие алгоритмов классификации и восстановления зависимостей.

Целью диссертационной работы является разработка эффективных методов и алгоритмов настройки непараметрических оценок в условиях малых выборок.

Поставленная цель достигается путём решения следующих **задач**:

а) провести сравнительный анализ известных методов настройки непараметрических оценок;

б) осуществить расширение области применимости проекционной оценки;

в) исследовать возможность применения метода моментов для настройки проекционной оценки и выполнить его обобщение;

г) разработать алгоритмы настройки коэффициентов и длины ряда проекционной оценки функции плотности вероятности случайного вектора, ориентированные на решение задач восстановления зависимостей, классификации и оценивания количества информации;

д) сравнить разработанные методы и алгоритмы с известными алгоритмами настройки проекционной оценки на малых выборках.

Соответствие диссертации паспорту специальности. Диссертационная работа соответствует области исследований специальности 05.13.17 – Теоретические основы информатики по п. 5 «Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечениях, разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений» и п. 7 «Разработка методов распознавания образов, фильтрации, распознавания и синтеза изображений, решающих правил. Моделирование формирования эмпирического знания».

Методы исследования. Основные результаты получены на основе методов теории вероятностей, математической статистики, функционального анализа и теории меры, а также матричного анализа. При численных расчётах функционалов качества получаемых оценок использован метод статистических испытаний.

Научная новизна:

1. Впервые использовано весовое расширение пространства L_2 при построении проекционной оценки для любых функций плотности вероятности, в том числе, с несуммируемым квадратом. Тем самым расширена область применения проекционных оценок, в частности, при решении задач обнаружения закономерностей в данных и распознавания образов.

2. Разработан новый метод настройки коэффициентов проекционной оценки функции плотности вероятности случайного вектора, являющийся обобщением метода моментов. Метод позволяет повысить эффективность проекционной оценки в условиях малых выборок.

3. Предложен новый метод оценивания длины ряда проекционной оценки, в которой коэффициенты настраиваются методом моментов или его обобщением.

4. Разработаны алгоритмы настройки коэффициентов и длины ряда проекционной оценки функции плотности вероятности случайного вектора, которые ориентированы на решение задач восстановления зависимостей, классификации и оценивания количества информации. Предложенные алгоритмы являются более результативными для проекционной оценки в условиях малых выборок, чем алгоритмы, реализующие традиционный подход.

Теоретическая и практическая значимость работы. Работа носит теоретический характер. Результаты могут быть использованы при решении задач восстановления зависимостей, классификации и оценивания количества информации для построения проекционных оценок функции плотности вероятности.

Положения, выносимые на защиту диссертационной работы.

1. Доказательство сходимости проекционной оценки в весовом пространстве $L_{2,w}(\mathbb{R}^k)$ к функции плотности вероятности для любого закона распределения непрерывного случайного вектора при подходящей весовой функции w.

2. Метод настройки коэффициентов проекционной оценки функции плотности вероятности случайного вектора, представляющий собой обобщение метода моментов.

3. Метод оценивания длины ряда проекционной оценки функции плотности вероятности случайного вектора.

4. Алгоритмы настройки коэффициентов и длины ряда проекционной оценки функции плотности вероятности случайного вектора, предназначенные для решения прикладных задач на малых выборках.

Достоверность результатов работы подтверждается математическими доказательствами основных положений, а также численными экспериментами.

Апробация результатов работы. Результаты диссертационной работы докладывались автором на следующих конференциях: Всероссийской конференции «Наука. Технологии. Инновации» (Новосибирск, 2007 г.); Всероссийской конференции «Молодёжь и наука» (Красноярск, 2007, 2014 гг.); Всероссийской конференции «Актуальные проблемы авиации и космонавтики» (Красноярск, 2014, 2015, 2017 гг.); Всероссийской конференции «Наука и АСУ – 2014» (Москва, 2014 г.); Международной конференции «Решетнёвские чтения» (Красноярск, 2014, 2016 гг.).

Результаты работы обсуждались на научно-исследовательских семинарах в Сибирском федеральном университете и Сибирском государственном университете науки и технологий имени академика М. Ф. Решетнёва.

Публикации. По результатам диссертационного исследования опубликовано 12 работ, из которых 4 изданы в журналах, рекомендованных ВАК, 7 в тезисах и трудах конференций и 1 свидетельство о регистрации программы, зарегистрированное в Реестре программ для ЭВМ.

Структура работы. Диссертационная работа состоит из введения, трёх глав, заключения и списка литературы.

В **главе 1** даётся обзор и сравнительный анализ основных методов оценивания функции плотности вероятности (решается задача а) диссертационного исследования). Приводятся необходимые сведения из теории меры и функционального анализа. Определяется пространство $L_p(\Omega, \Sigma, \mu)$ μ-интегрируемых в p-й степени функций, заданных на множестве Ω, где μ – мера, определённая на системе подмножеств Σ множества Ω. Указываются достаточные условия на μ для того, чтобы пространство $L_2(\Omega, \Sigma, \mu)$ было гильбертовым. Приводится постановка задачи оптимизации оценки функции плотности вероятности.

Рассмотрим некоторый класс $\mathcal{F} = \{\hat{f}_\alpha(\mathbf{x}) \mid \alpha \in \mathrm{A}\}$ оценок функции плотности вероятности $f(\mathbf{x})$, где α – набор параметров, A – некоторое множество. Критерием близости оценки $\hat{f}_\alpha(\mathbf{x})$ к истинной плотности является следующий функционал:

$$Q_p\{\hat{f}\} = M\left\{\left\|\hat{f} - f\right\|^p_{L^k_p}\right\}. \tag{1.3}$$

Тогда α выбирается, исходя из условия

$$Q_p\{\hat{f}_\alpha\} \to \min_\alpha.$$

Если $p = 2$, то функционал (1.3) допускает следующее преобразование:

$$Q_2\{\hat{f}_\alpha\} = M\left\{\left\|\hat{f}_\alpha\right\|^2 - 2\left(\hat{f}_\alpha, f\right)\right\} + \|f\|^2.$$

Слагаемое $\|f\|^2$, независимое от α, при минимизации обычно опускается. Тогда приходим к следующей задаче:

$$W\{\hat{f}_\alpha\} = M\left\{\left\|\hat{f}_\alpha\right\|^2 - 2\left(\hat{f}_\alpha, f\right)\right\} \to \min_\alpha,$$

эквивалентной задаче минимизации функционала (1.3). В ряде работ [91, 118] этот подход используется при настройке непараметрических оценок функции плотности вероятности.

Проекционная оценка функции плотности вероятности случайной величины определяется следующим образом [144]:

$$\hat{f}(x) = \sum_{j=0}^{l} a_j \varphi_j(x). \tag{1}$$

Указаны известные [144, 117] методы настройки длины ряда l, коэффициентов a_j, а также основные используемые ортонормальные системы функций $\varphi_j(x)$. Указаны случаи употребления той или иной системы.

Упоминается также рассматриваемое в ряде работ [38, 36] обобщение проекционной оценки в виде

$$\hat{f}(x) = \sum_{j=0}^{l} \lambda_j a_j \varphi_j(x).$$

Оценка Розенблатта – Парзена определяется формулой [20, 25]:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} \Phi\left(\frac{x - x_i}{h}\right).$$

Указаны распространённые методы [75, 48, 91, 88] настройки параметра размытости h, а также основные используемые типы ядерных функций $\Phi(z)$.

Также приводятся некоторые обобщения [88] оценки Розенблатта – Парзена: интегральная и регрессионная оценки функции плотности вероятности.

Кроме того, рассмотрены работы, посвящённые исследованию других непараметрических оценок: гистограммы [31, 14, 128], оценки k ближайших соседей [124], полиграммы k-го порядка [131].

Глава 2 посвящена исследованию проекционной оценки. В ней решаются задачи б), в) диссертационного исследования. Основные результаты второй главы опубликованы в работах [149, 150, 151, 152, 153, 159, 160].

Построение проекционной оценки является мощным непараметрическим методом восстановления функции плотности вероятности [7]. В отличие от оценки Розенблатта – Парзена и других непараметрических оценок проекционная оценка не содержит в себе всей исследуемой выборки и допускает лаконичное математическое выражение. В § 2.1 решается проблема применимости проекционной оценки для оценивания функций плотности вероятности с несуммируемым квадратом за счёт введения весового пространства $L_{2,w}(\Omega)$. Оказалось, что при выполнении следующих условий:

1) $w(\mathbf{x})$ μ-измерима;

2) $w(\mathbf{x})$ положительная почти всюду на Ω;

3) $\operatorname*{ess\,sup}_{\mathbf{x}\in\Omega} w(\mathbf{x}) < +\infty$

данное пространство является гильбертовым. Следовательно в этом пространстве определена проекционная оценка. Показано, что для любой функции плотности вероятности $f(\mathbf{x})$ существует соответствующая весовая функция $w(\mathbf{x})$, при которой $f \in L_{2,w}(\Omega)$, следовательно, проекционная оценка сходится к функции плотности вероятности в этом пространстве.

Рассматривается проблема выбора весовой функции $w(\mathbf{x})$. Доказано необходимое и достаточное условие на весовые функции w_1 и w_2 для того, чтобы пространство $L_{p,w_1}(\Omega)$ было шире пространства $L_{p,w_2}(\Omega)$.

Следствие 2.3.3. Пространство $L_{p,w_1}(\Omega)$ является расширением пространства $L_{p,w_2}(\Omega)$, т.е. $L^k_{p,w_2}(\Omega) \subset L^k_{p,w_1}(\Omega)$, тогда и только тогда, когда выполняются следующие два условия:

1) $\operatorname*{ess\,inf}_{\mathbf{x}\in\Omega} \dfrac{w_1(\mathbf{x})}{w_2(\mathbf{x})} = 0;$

2) $\operatorname*{ess\,sup}_{\mathbf{x}\in\Omega} \dfrac{w_1(\mathbf{x})}{w_2(\mathbf{x})} < +\infty.$

Показано, что при определённых условиях пространство $L_{2,w}(\Omega)$ можно расширить следующим образом:

$$w_0(\mathbf{x}) = \begin{cases} w(\mathbf{x}) \left(\sum_{j=1}^{k} |x_j - a_j|^{\alpha_j}\right)^p, & x \in U_\delta(\mathbf{a}) \\ w(\mathbf{x}), & x \in \Omega \setminus U_\delta(\mathbf{a}) \end{cases},$$

где $\delta > 0$, $U_\delta(\mathbf{a})$ - δ-окрестность точки $\mathbf{a}$, а показатели α_j удовлетворяют следующему неравенству:

$$\frac{1}{\alpha_1} + ... + \frac{1}{\alpha_k} \leq 1.$$

Также рассмотрена возможность настройки коэффициентов проекционной оценки с помощью метода моментов и некоторого его обобщения. Показано, что частным случаем предлагаемого обобщения является традиционный метод настройки коэффициентов. Сравнение эффективности рассматриваемых подходов показало, что предлагаемый метод даёт преимущество перед традиционным, которое оказывается более выраженным при малых объёмах выборки.

Для настройки длины ряда была построена несмещённая оценка $\hat{W}_l$ функционала $W\{\hat{f}_l\}$ от проекционной оценки $\hat{f}_l(\mathbf{x})$, в которой коэффициенты настраиваются обобщённым методом моментов. Данный подход сравнивается с известным подходом, предложенным в работе [117]. Сравнение показало, что проекционные оценки, основанные на методе моментов, показывают аналогичные или лучшие результаты при различных восстанавливаемых распределениях и используемых ортонормальных системах.

При настройке проекционной оценки в многомерном случае рассматривается вопрос о построении базиса в пространстве $L_{2,w}(\Omega)$, где $\Omega \subseteq \mathbb{R}^k$. Оказалось, что если выполняется условие

$$w(x_1, \ldots, x_k) = w_1(x_1) \ldots w_k(x_k). \tag{2.27}$$

причём $w_i : \Omega_i \to \mathbb{R}$, $\Omega_1 \times \cdots \times \Omega_k \supseteq \Omega$, то мера μ_w, индуцированная весовой функцией w является прямым произведением мер μ_{w_i}:

$$\mu_w = \mu_{w_1} \otimes \cdots \otimes \mu_{w_k}, \quad \mu_{w_i}(X) = \int_X w_i(x) dx,$$

и, следовательно, базисом в пространстве $L_{2,w}(\Omega)$ является система функций

$$\psi_{j_1,\ldots,j_k}(x_1, \ldots, x_k) = \varphi_{1,j_1}(x_1) \cdot \ldots \cdot \varphi_{k,j_k}(x_k), \quad j_1, \ldots, j_k = 0, 1, \ldots,$$

где $\{\varphi_{1,j}\}_{j=0}^{\infty}, \ldots, \{\varphi_{k,j}\}_{j=0}^{\infty}$ – базисы в пространствах $L_{2,w_1}(\Omega_1), \ldots, L_{2,w_k}(\Omega_k)$ соответственно. Преимуществом данного подхода является упрощение выкладок и уменьшение вычислительной сложности при построении проекционной оценки. Недостатком является ограничение множества восстанавливаемых функций плотности вероятности. Так, доказана следующая теорема.

Теорема 2.6. При $\Omega \subseteq \mathbb{R}^k$, $k \geq 2$, существуют функции $f \in L_1(\Omega)$, не принадлежащие никакому пространству $L_{p,w}(\Omega)$, в котором весовая функция имеет вид (2.27).

Примером такой функции плотности вероятности при $k = 2$ является

$$f(x_1, x_2) = \frac{3}{8\sqrt{|x_1 - x_2|}}, \quad x_1, x_2 \in [0; 1], \quad x_1 \neq x_2.$$

В § 2.2 исследуется проекционная оценка плотности вероятности, в которой параметры настраиваются методом моментов, а также предлагается некоторое обобщение последнего.

Применение метода моментов для оценивания коэффициентов a_j проекционной оценки сводится к решению системы линейных уравнений, которая в матричном виде записывается следующим образом:

$$\begin{pmatrix} (1,\varphi_0)_w & \dots & (1,\varphi_l)_w \\ \vdots & \ddots & \vdots \\ (x^l,\varphi_0)_w & \dots & (x^l,\varphi_l)_w \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ \vdots \\ a_l \end{pmatrix} = \begin{pmatrix} \hat{\nu}_0 \\ \vdots \\ \hat{\nu}_l \end{pmatrix}, \tag{2}$$

где $(x, y)_w$ – скалярное произведение в весовом пространстве $L_{2,w}$, $\hat{\nu}_j = \frac{1}{n}\sum_{i=1}^{n} x_i^j$ – j-й выборочный начальный момент исследуемой случайной величины, причём $\hat{\nu}_0 \equiv 1$. Если основная матрица системы (2) не вырождена, то она имеет единственное решение, которое берётся в качестве искомых оценок a_j.

Основная идея обобщения метода моментов состоит в том, что для оценивания того же количества параметров a_j используется большее количество выборочных начальных моментов $\hat{\nu}_j$. Пусть требуется оценить $(l+1)$ коэффициентов a_0, …, a_l. Выберем произвольное натуральное $l' > l$ и запишем для него соответствующую систему (2):

$$\begin{pmatrix} (1,\varphi_0)_w & \dots & (1,\varphi_{l'})_w \\ \vdots & \ddots & \vdots \\ (x^{l'},\varphi_0)_w & \dots & (x^{l'},\varphi_{l'})_w \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ \vdots \\ a_{l'} \end{pmatrix} = \begin{pmatrix} \hat{\nu}_0 \\ \vdots \\ \hat{\nu}_{l'} \end{pmatrix}. \tag{3}$$

Если основная матрица системы (3) не вырождена, то имеется единственное решение $(a_0, a_1, \dots, a_{l'})^T$, из которого выбирается подматрица $(a_0, a_1, \dots, a_l)^T$, которая берётся в качестве набора искомых коэффициентов для оценки (1). Данные значения обозначаются через $a_j^{(l')}$. Об оценках $a_j^{(l')}$ в работе доказываются ряд свойств:

– при использовании ортонормированной системы Лежандра выполняется равенство $a_j^{(l')} = a_j^{(l)}$ при любом $l' > l$;

– в общем случае $a_j^{(l')} \neq a_j^{(l)}$, но при определённых условиях на базис при неограниченном увеличении l' и фиксированных n и l оценка $a_j^{(l')}$ сходится к традиционной оценке $a_j = \frac{1}{n}\sum_{i=1}^{n}\varphi_j(x_i)$ почти наверное;

– в общем случае оценки a_j не являются несмещёнными;

– оценка плотности, в которой коэффициенты рассчитываются обобщённым методом моментов при подходящем выборе параметра l' , ближе в среднем квадратичном к истинной плотности при любой длине ряда l.

В § 2.3 рассматривается задача оценивания параметров l и l'. Идея метода взята из работы [117]. Оптимальные значения l^* и $(l')^*$ определяются из условия

$$Q\left\{\hat{f}_l^{(l')}\right\} = M\left\{\left\|\hat{f}_l^{(l')} - f\right\|^2\right\} \to \min_{l,l'}.$$

Вводится функционал

$$W\left\{\hat{f}_l^{(l')}\right\} = M\left\{\left\|\hat{f}_l^{(l')}\right\|^2 - 2\left(\hat{f}_l^{(l')}, f\right)\right\},$$

минимизация которого эквивалентна минимизации функционала Q:

$$\arg\min_{l,l'} Q\left\{\hat{f}_l^{(l')}\right\} = \arg\min_{l,l'} W\left\{\hat{f}_l^{(l')}\right\}.$$

В ходе исследования удалось построить несмещённую оценку $\hat{W}_{l,l'}$ (2.24) функционала $W\left\{\hat{f}_l^{(l')}\right\}$:

$$M\left\{\hat{W}_{l,l'}\right\} = W\left\{\hat{f}_l^{(l')}\right\}.$$

Тогда оценки $\hat{l}$ и $\hat{l}'$ находятся путём минимизации $\hat{W}_{l,l'}$:

$$\left(\hat{l}, \hat{l}'\right) = \arg\min_{l,l'} \hat{W}_{l,l'}.$$

Было проведено сравнение предложенных методов настройки проекционной оценки с традиционными методами, которое показало, что независимо от используемого базиса и восстанавливаемого распределения обобщённый метод моментов даёт лучшие результаты.

В § 2.4 предложенный подход распространяется на многомерный случай. Была получена несмещённая оценка функционала W, путём минимизации которой находятся оценки $\hat{l}_1$, …, $\hat{l}_k$. Сравнение с традиционным подходом на тестовых распределениях показало, что предложенный подход даёт лучшие результаты.

В **главе 3** представлены алгоритмы настройки коэффициентов и длины ряда проекционной оценки функции плотности вероятности случайного вектора, ориентированные на решение задач восстановления зависимостей, классификации и оценивания количества информации, и выполнен сравнительный анализ разработанных методов и алгоритмов с известными алгоритмами настройки проекционной оценки на малых выборках (решаются задачи г) и д) диссертационного исследования). Сравниваемые алгоритмы реализованы на языке Wolfram Language [148].

В § 3.1 приводится постановка задачи восстановления функции (многомерной) регрессии. Для решения данной задачи разработано 3 алгоритма. Первый алгоритм основан на проекционной оценке функции плотности вероятности, в которой параметры настраиваются при помощи метода моментов; два других – на оценке Розенблатта – Парзена, с разными способами настройки параметра размытости. Сравнение разработанных алгоритмов на тестовых задачах показало, что алгоритмы 1.2 и 1.3, основанные на оценке Розенблатта – Парзена эффективнее чем, алгоритм 1.1, основанный на проекционной оценке, причём способ настройки параметра размытости оказался несущественным.

В § 3.2 приводится постановка задачи классификации. Для решения данной задачи разработано 3 алгоритма. Сравнение разработанных алгоритмов на тестовых задачах показало, что улучшение достигается при использовании алгоритма 2.2, основанного на оценке Розенблатта – Парзена.

В § 3.3 рассматривается задача оценивания количества информации. Для решения данной задачи было разработано 4 алгоритма, для которых было выполнено сравнение эффективности на тестовых задачах. Сравнительный анализ показал, что наибольшая точность достигается при использовании алгоритма 3.4, основанного на оценке Розенблатта – Парзена.

В **заключении** сформулированы основные результаты и выводы диссертационной работы.

Глава 1. Обзор методов оценивания функции плотности вероятности

В главе 1 приводится обзор известных методов оценивания функции плотности вероятности: проекционной оценки, оценки Розенблатта – Парзена, гистограммы, полиграммы, оценки k ближайших соседей, метод стохастической регуляризации. Приводятся основные результаты, относящиеся к настройке рассмотренных методов. Сформулированы задачи, требующие дальнейшего исследования.

Также приводятся основные определения и обозначения, используемые в настоящей работе.

§ 1.1. Основные определения и обозначения

В настоящей работе используются стандартные обозначения, применяемые в математическом анализе, функциональном анализе и теории вероятностей [74, 82, 83, 130, 134 и др.]. Иногда в формулировках утверждений и в доказательствах используются логические обозначения.

Пусть A, B, C – произвольные множества. Тогда обозначение $A \sqcup B$ означает объединение непересекающихся множеств (*дизъюнктное объединение*):

$$C = A \sqcup B \;\Leftrightarrow\; C = A \cup B \;\wedge\; A \cap B = \varnothing.$$

Рассмотрим меру μ, заданную на σ-алгебре Σ с единицей Ω.

Мера μ называется *σ-аддитивной*, если

$$\mu \bigsqcup_{n=1}^{\infty} A_n = \sum_{n=1}^{\infty} \mu A_n.$$

Мера μ называется *конечной*, если $\mu(\Omega) < +\infty$. Если множество Ω можно представить в виде

$$\Omega = \bigcup_{n=1}^{\infty} A_n,$$

так что

$$\forall n \in \mathbb{N} \quad \mu A_n < +\infty,$$

то мера μ называется *σ-конечной*.

Мера μ называется *полной* если из $\mu B = 0$ и $A \subset B$ следует $\mu A = 0$.

Последовательность $\{A_1, A_2, ...\}$ множеств из Σ называется *счётным базисом* меры μ, если

$$\forall M \in \Sigma \quad \forall \varepsilon > 0 \quad \exists n \in \mathbb{N} \quad \mu(M \triangle A_n) < \varepsilon.$$

Если μ – полная σ-конечная мера, то упорядоченная тройка $\langle \Omega, \Sigma, \mu \rangle$ называется *пространством с мерой.*

Пусть $\langle \Omega, \Sigma, \mu \rangle$ – произвольное пространство с мерой. Через $L_p(\Omega, \Sigma, \mu)$, где $p \geq 1$ обозначим множество μ-измеримых функций $f : \Omega \to \mathbb{R}$, для которых

$$\int_\Omega |f(x)|^p d\mu(x) < +\infty.$$

В пространстве $L_p(\Omega, \Sigma, \mu)$ вводится норма

$$\|f\|_{L_p(\Omega,\Sigma,\mu)} = \left(\int_\Omega |f(x)|^p d\mu(x) \right)^{1/p}.$$

Полезными являются следующие две известные теоремы [78].

Теорема 1.1. *Если μ – σ-конечная мера, то пространство $L_p(\Omega, \Sigma, \mu)$ банахово.*

Теорема 1.2. *Если мера μ имеет счётный базис, то пространство $L_p(\Omega, \Sigma, \mu)$ сепарабельно.*

При $p = 2$ пространство $L_p(\Omega, \Sigma, \mu)$ является евклидовым, в котором скалярное произведение определяется следующим образом:

$$(f, g) = \int_\Omega f(x)g(x)d\mu(x).$$

Сепарабельное бесконечномерное евклидово пространство, полное относительно своей нормы, называется *гильбертовым пространством.*

Таким образом, если мера μ σ-конечна и имеет счётный базис, то пространство $L_2(\Omega, \Sigma, \mu)$ является гильбертовым.

Если μ – мера Лебега, то пространство $L_p(\Omega, \Sigma, \mu)$ будем обозначать короче: $L_p(\Omega)$. Если $\Omega = \mathbb{R}^k$, то $L_p(\Omega)$ будем обозначать как L_p^k.

Пусть $f,\ f_1,\ f_2,\ \ldots$– последовательность функций в пространстве $L_p(\Omega)$. Если имеет место сходимость но норме

$$\lim_{n\to\infty} \|f_n - f\|_{L_p(\Omega)} = 0,$$

то этот факт мы будем обозначать следующим образом:

$$f_n \xrightarrow[n\to\infty]{L_p(\Omega)} f.$$

Если в пространстве с мерой $\langle\Omega, \Sigma, \mu\rangle$ выполняется $\mu(\Omega) = 1$, то мера μ называется *вероятностью* и обозначается P. Пространство $\langle\Omega, \Sigma, P\rangle$ при этом называется *вероятностным пространством*, а элементы множества Σ называются *случайными событиями*. Вероятность случайного события A будем обозначать как $P\{A\}$.

Рассмотрим вероятностное пространство $\langle\mathbb{R}, \Sigma, P\rangle$. Любое измеримое отображение $\boldsymbol{\xi} : \Sigma \to \mathbb{R}^k$ называется *случайным вектором.* В случае $k = 1$ случайный вектор будем называть *случайной величиной.*

Функция распределения случайного вектора $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_k)$ определяется как

$$F(\mathbf{x}) = F(x_1, \ldots, x_n) = P\{\xi_1 < x_1 \wedge \cdots \wedge \xi_k < x_k\} =$$
$$= P\{\omega \in \Omega \mid \xi_1(\omega) < x_1 \wedge \cdots \wedge x_k(\omega) < x_k.\}$$

Математическое ожидание случайного вектора определяется как вектор математических ожиданий соответствующих компонент:

$$M\{\boldsymbol{\xi}\} = (M\{\xi_1\}, \ldots, M\{\xi_k\}),$$
$$M\{\xi_j\} = \int_{\mathbb{R}^k} x_j \, dF(\mathbf{x}), \quad j = 1, ..., k.$$

Случайный вектор $\boldsymbol{\xi}$ называется *непрерывным*, если его функция распределения $F(\mathbf{x})$ абсолютно непрерывна, т.е. может быть представлена в виде:

$$F(\mathbf{x}) = \int_{D(\mathbf{x})} f(\mathbf{t}) d\mu(\mathbf{t}), \tag{1.1}$$

где μ – мера Лебега,

$$D(\mathbf{x}) = \{\mathbf{t} \in \mathbb{R}^k \mid t_1 < x_1 \wedge \cdots \wedge t_k < x_k\}.$$

При этом любая функция $f(\mathbf{t})$, удовлетворяющая равенству (1.1), называется *функцией плотности вероятности* случайного вектора $\boldsymbol{\xi}$. Очевидно, функция плотности вероятности определена с точностью до эквивалентности.

Множество действительных функций k переменных, являющихся функциями плотности вероятности, будем обозначать символом $\mathcal{D}^k$. Таким образом, $f \in \mathcal{D}^k$ тогда и только тогда, когда выполняются следующие два условия:

$$\begin{aligned} &1)\ f(\mathbf{x}) \geq 0 \text{ почти всюду на } \mathbb{R}^k; \\ &2)\ \int_{\mathbb{R}^k} f(\mathbf{x}) d\mu(\mathbf{x}) = 1. \end{aligned} \tag{1.2}$$

Множества $L_1^k \setminus \{0\}$ и $\mathcal{D}^k$ связаны очевидным сюръективным отображением $\varphi : L_1^k \setminus \{0\} \to \mathcal{D}^k$:

$$\varphi : f \mapsto \frac{1}{\|f\|_{L_1^k}} |f|.$$

Пусть a – некоторая числовая характеристика случайного вектора $\boldsymbol{\xi}$. Под обозначением $\hat{a}$ будем понимать статистическую оценку a. Если $f(x)$ – некоторая функциональная характеристика, например, функция плотности вероятности, то вводится аналогичное обозначение: $\hat{f}(x)$.

Далее вводятся определения, характеризующие близость оценки $\hat{f}(\mathbf{x})$ к оцениваемой функции плотности вероятности $f(\mathbf{x})$.

Выражение

$$M\left\{\left(\hat{f}(\mathbf{x}) - f(\mathbf{x})\right)^2\right\}$$

называется *усреднённой квадратичной ошибкой аппроксимации* в точке $\mathbf{x} \in \Omega$ [75]. Аналогично, назовём усреднённой глобальной ошибкой аппроксимации степени $p \geq 1$ в точке $\mathbf{x}$ выражение

$$M\left\{\left|\hat{f}(\mathbf{x}) - f(\mathbf{x})\right|^p\right\}.$$

Усреднённой глобальной ошибкой аппроксимации (степени p) назовём число

$$Q_p\{\hat{f}\} = M\left\{\left\|\hat{f} - f\right\|_{L_p^k}^p\right\}. \tag{1.3}$$

При этом

$$M\left\{\left\|\hat{f} - f\right\|_{L_2^k}^p\right\} = M\left\{\int_{\mathbb{R}^k} \left|\hat{f}(\mathbf{x}) - f(\mathbf{x})\right|^p d\mu(\mathbf{x})\right\} =$$

$$= \int_{\mathbb{R}^k} M\left\{\left|\hat{f}(\mathbf{x}) - f(\mathbf{x})\right|^p\right\} d\mu(\mathbf{x}).$$

Перестановка операторов интегрирования и математического ожидания возможна в случае существования интеграла под математическим ожиданием [83, с. 337].

Часто рассматривается случай, когда $p = 2$:

$$Q_2\{\hat{f}\} = M\left\{\left\|\hat{f} - f\right\|^2_{L_2^k}\right\} = \int_{\mathbb{R}^k} M\left\{\left(\hat{f}(\mathbf{x}) - f(\mathbf{x})\right)^2\right\} d\mu(\mathbf{x}). \qquad (1.4)$$

Кроме того, определяется *относительная глобальная ошибка аппроксимации* [75]:

$$u^2 = \frac{1}{\|f\|^2_{L_2^k}} \int_{\mathbb{R}^k} M\left\{\left(\hat{f}(\mathbf{x}) - f(\mathbf{x})\right)^2\right\} d\mu(\mathbf{x}) = \frac{Q_2\{\hat{f}\}}{\|f\|^2_{L_2^k}}.$$

Рассмотрим некоторый класс $\mathcal{F} = \{\hat{f}_\alpha(\mathbf{x}) \mid \alpha \in \mathrm{A}\}$ оценок функции плотности вероятности $f(\mathbf{x})$ и поставим задачу минимизации функционала Q_2 по α:

$$Q_2\{\hat{f}_\alpha\} \to \min_{\alpha \in \mathrm{A}}. \qquad (1.5)$$

Функционал Q_2 можно преобразовать следующим образом:

$$Q_2\{\hat{f}_\alpha\} = M\left\{\left\|\hat{f}_\alpha\right\|^2 - 2\left(\hat{f}_\alpha, f\right) + \|f\|^2\right\} = M\left\{\left\|\hat{f}_\alpha\right\|^2 - 2\left(\hat{f}_\alpha, f\right)\right\} + \|f\|^2.$$

Слагаемое $\|f\|^2$ не зависит от α, по этому при минимизации оно может быть отброшено. Тогда получим функционал

$$W\{\hat{f}_\alpha\} = M\left\{\left\|\hat{f}_\alpha\right\|^2 - 2\left(\hat{f}_\alpha, f\right)\right\}, \qquad (1.6)$$

минимизация которого эквивалентна минимизации функционала Q_2:

$$\arg\min_\alpha W\{\hat{f}_\alpha\} = \arg\min_\alpha Q_2\{\hat{f}_\alpha\}.$$

Таким образом, мы получаем задачу оптимизации

$$W\{\hat{f}_\alpha\} \to \min_{\alpha \in \mathrm{A}},$$

эквивалентную задаче (1.5), которая обычно решается при настройке параметров оценки функции плотности вероятности в том или ином виде.

Далее рассматриваются некоторые распространённые методы оценивания функции плотности вероятности.

§ 1.2. Оценки проекционного типа

Пусть восстанавливаемая функция плотности вероятности $f(x)$ непрерывной случайной величины принадлежит некоторому гильбертову пространству $H(\Omega)$ функций, определённых на замкнутом множестве $\Omega \subseteq \mathbb{R}$, для которого

$$\int_{\Omega} f(x) d\mu(x) = 1, \tag{1.7}$$

$\{\varphi_j\}_{j=0}^{\infty}$ – полная ортонормальная система функций (*базис*) пространства $H(\Omega)$. *Проекционная оценка* [144] функции $f(x)$ имеет вид:

$$\hat{f}(x) = \sum_{j=0}^{l} a_j \varphi_j(x), \tag{1.8}$$

где l, a_0, ..., a_l – параметры проекционной оценки, подлежащие оцениванию.

При фиксированной длине ряда l оптимальными коэффициентами a_0, ..., a_l в смысле критерия (1.4), очевидно, являются коэффициенты Фурье:

$$\alpha_j = (f, \varphi_j). \tag{1.9}$$

Если в качестве коэффициентов a_j используются коэффициенты Фурье, то для проекционной оценки вводится следующее обозначение:

$$f_l(x) = \sum_{j=0}^{l} \alpha_j \varphi_j(x). \tag{1.10}$$

При неограниченном увеличении длины ряда l имеет место сходимость по норме пространства $H(\Omega)$:

$$\lim_{l \to \infty} \|f_l - f\|_{H(\Omega)} = 0.$$

Скорость сходимости оценки $f_l(x)$ к истинной плотности $f(x)$ зависит от выбранного базиса $\{\varphi_j\}$. Однако очевидно, что оптимального базиса для всех законов распределения не существует. Действительно, для данной функции $f(x)$ оптимальным в смысле скорости сходимости будет любой базис, в котором $\varphi_0(x) \equiv \frac{1}{\|f\|} f(x)$. В этом случае уже при $l = 0$ оценка (1.10) совпадает с $f(x)$:

$$f_l(x) = f_0(x) = \alpha_0 \varphi_0(x) = (f, \varphi_0)\varphi_0(x) = \frac{1}{\|f\|^2}(f, f) f(x) = f(x).$$

Множество Ω оценивания функции плотности вероятности целесообразно выбирать как можно более узким. В зависимости от имеющейся априорной информации о законе распределения множество Ω может представлять собой всю числовую прямую, луч или отрезок. При некоторых частных значениях Ω широко распространены следующие виды ортонормальных систем:

1) система Лежандра, $\Omega = [-1; 1]$:

$$\varphi_n(x) = \sqrt{\frac{2n+1}{2}} \frac{d^n}{dx^n}(x^2-1)^n, \quad n = 0, 1, \ldots;$$

2) система Чебышёва, $\Omega = [-1; 1]$:

$$\sqrt{\pi}\varphi_n(x) = \begin{cases} \dfrac{1}{\sqrt[4]{1-x^2}}, & n = 0 \\ \dfrac{\cos(n \arccos x)}{\sqrt[4]{1-x^2}}, & n > 0 \end{cases};$$

3) тригонометрическая система, $\Omega = [-\pi; \pi]$:

$$\varphi_n(x) = \begin{cases} \dfrac{1}{\sqrt{x}}, & n = 0 \\ \sin \frac{n+1}{2}\pi x, & n = 1, 3, 5, \ldots \\ \cos \frac{n}{2}\pi x, & n = 2, 4, 6, \ldots \end{cases}$$

4) система Хаара, $\Omega = [0; 1]$:

$$\varphi_0(x) = 1,$$

$$\varphi_{n,i}(x) = \begin{cases} 2^{n/2}, & x \in \left(\dfrac{i-1}{2^n}; \dfrac{i-1}{2^n} + \dfrac{1}{2^{n+1}}\right) \\ -2^{n/2}, & x \in \left(\dfrac{i-1}{2^n} + \dfrac{1}{2^{n+1}}; \dfrac{i-1}{2^n}\right) \\ 0, & \text{иначе} \end{cases},$$

где $n = 1, 2, \ldots, i = 1, \ldots, 2^n$.

Эти же системы используются в случае, если Ω представляет собой произвольный отрезок $[a; b]$. Действительно, линейным преобразованием над ξ можно построить случайную величину ξ', у которой носитель Ω' не выходит за пределы требуемого отрезка.

При $\Omega = [0; +\infty)$ используется

5) система Лагерра:

$$\varphi_n(x) = \frac{e^{x/2}}{n!}\frac{d^n}{dx^n}(x^n e^{-x}), \quad n = 0, 1, \dots$$

Эту же систему можно использовать в случаях, когда Ω представляет собой произвольный луч $[a; +\infty)$ или $(-\infty; b]$.

Наконец, если вовсе нет априорной информации об Ω, то принимается $\Omega = \mathbb{R}$, и используется

6) система Эрмита:

$$\varphi_n(x) = (-1)^n \sqrt{\frac{e^{x^2}}{\sqrt{\pi}2^n n!}}\frac{d^n}{dx^n}e^{-x^2}, \quad n = 0, 1, \dots$$

В [26] указано, что если $f(x)$ непрерывна, суммируема с квадратом и имеет ограниченное изменение, то при использовании базиса Эрмита оценка (1.10) сходится к f равномерно в любом интервале $(a; b)$.

Так как формула (1.9) для расчёта оптимальных коэффициентов a_j проекционной оценки использует вид истинной плотности вероятности $f(x)$, то она не применима, если последняя неизвестна. В этом случае коэффициенты a_j оцениваются по имеющейся выборке $x_1, \dots, x_n$. В [144] предложен метод оценивания a_j, получивший большое распространение:

$$a_j = \frac{1}{n}\sum_{i=0}^{n} \varphi_j(x_i), \quad j = 0, \dots, l. \tag{1.11}$$

Оценку (1.8), в которой коэффициенты a_j рассчитаны по формуле (1.11), будем называть *оценкой Ченцова.*

В работе [117] показано, что при фиксированном объёме выборки n оценка Ченцова с возрастанием l не сходится к истинной плотности $f(x)$. Более того, показано, что при любом конечном n

$$\lim_{l\to\infty} Q_2\{\hat{f}_l\} = +\infty.$$

Отсюда следует, что для данного n существует некоторое оптимальное значение l^* длины ряда (1.8), минимизирующее функционал Q_2. В этой же работе приводится метод оценки значения l^* по выборке. Строится оценка функционала (1.6) в виде

$$\hat{W}_l = \sum_{j=0}^{l} \left(\frac{k+2}{n(n-1)} \sum_{i=1}^{n} \varphi_j^2(x_i) - \frac{n+k+1}{n-1} a_j^2 \right), \tag{1.12}$$

где

$$k = \frac{\sum_{j=1}^{l} \left(\frac{n+3}{n} a_j^2 \sum_{i=1}^{n} \varphi_j^2(x_i) - \frac{2}{n^2} \left(\sum_{i=1}^{n} \varphi_j^2(x_i) \right)^2 - (n+1) a_j^4 \right)}{\sum_{j=1}^{l} \left(\frac{1}{n} \sum_{i=1}^{n} \varphi_j^2(x_i) - a_j^2 \right)^2}.$$

Оценка $\hat{l}$ длины ряда находится путём минимизации $\hat{W}_l$:

$$\hat{l} = \arg\min_l \hat{W}_l.$$

Ряд исследований посвящён обобщению оценки (1.8) в виде

$$\hat{f}(x) = \sum_{j=0}^{l} \lambda_j a_j \varphi_j(x),$$

Так, в [38] приводятся оптимальные значения λ_j, минимизирующие функционал (1.4):

$$\lambda_j = \frac{\alpha_j^2}{\alpha_j^2 + \frac{1}{n} D\{\varphi_j(\xi)\}}, \quad j = 0, \ldots, l.$$

Отсюда следует, что для каждого j

$$\lim_{n \to \infty} \lambda_j = 1.$$

Там же приводятся оценки значений λ_j при конечных n:

$$\lambda_j = \begin{cases} 1, & 0 \le j \le M(n) \\ 0, & j > M(n) \end{cases},$$

где $M(n)$ – некоторая функция объёма выборки.

В работе [36] рассматривается аппроксимация функции плотности вероятности на отрезке $[0; 1]$ тригонометрическим рядов в комплексной форме:

$$\hat{f}(x) = 1 + \sum_{\substack{k=-l/2, \\ k \neq 0}}^{l/2} \frac{b_k}{b_k + \frac{1}{n}} a_k e^{2\pi i k x},$$

где

$$b_k = b_{-k} = \frac{1}{\lambda(2\pi k)^m}, \quad \lambda \geq 0,\ m > 1.$$

Также в работах В. Н. Вапника проекционная оценка рассматривается как частный случай оценки, получаемой с помощью метода стохастической регуляризации. Известно [59], что функция плотности вероятности $f(x)$ и функция распределения $F(x)$ случайной величины ξ связаны соотношением

$$\int_{-\infty}^{+\infty} \eta(x-t) f(t) dt = F(x), \tag{1.13}$$

где $\eta(z)$ – функция Хевисайда [63]:

$$\eta(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}.$$

Данное уравнение представляет собой интегральное уравнение Фредгольма 1-го рода относительно неизвестной плотности $f(x)$. Функция распределения $F(x)$ неизвестна, однако по имеющейся выборке $x_1, \ldots, x_n$ исследуемой случайной величины можно построить эмпирическую функцию распределения $F_n(x)$. По теореме Гливенко [64] с увеличением n эмпирическая функция распределения сходится к $F(x)$ равномерно по x с вероятностью 1:

$$P\left\{\lim_{n\to\infty} \sup_x |F_n(x) - F(x)| = 0\right\} = 1.$$

Однако даже при достаточно больших n точное решение уравнения (1.13) возможно только в классе обобщённых функций и является суммой δ-функций.

В монографии [58] развивается статистическая теория регуляризации, являющаяся обобщением теории регуляризации [132] для случая статистических зависимостей. В соответствии с методами этой теории вводится стабилизирующий функционал $\Omega\{f\}$ и решается задача минимизации

$$\int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} \eta(x-t) f(t) dt - F_n(x) \right)^2 dx + \gamma_n \Omega\{f\} \to \min_f, \tag{1.14}$$

где $\gamma_n \xrightarrow[n\to\infty]{} 0$.

В [48] показано, что если $f \in L_2[-\pi;\pi]$ и стабилизирующий функционал взят в виде $\Omega\{f\} = \|f\|_{L_2}^2$, то решение задачи (1.14) в классе тригонометрических многочленов достигается при

$$\hat{f}(x) = \hat{f}_1(x) + \hat{f}_2(x), \tag{1.15}$$

где

$$\hat{f}_1(x) = \frac{1}{2\pi} + \sum_{k=1}^{\infty} \frac{\left(\frac{1}{n}\sum_{i=1}^{n} \sin kx_i\right)\sin kx + \left(\frac{1}{n}\sum_{i=1}^{n}\cos kx_i\right)\cos kx}{\pi(1+\gamma_n k^2)},$$

$$\hat{f}_2(x) = \frac{2\left(\frac{1}{n}\sum_{i=1}^{n} x_i - \int_{-\pi}^{\pi} x\hat{f}_1(x)dx\right)}{\pi(1+2\Gamma(\gamma_n))}\sum_{k=1}^{\infty}(-1)^k \frac{k\sin kx}{1+\gamma_n k^2},$$

$$\Gamma(\gamma_n) = \sum_{k=1}^{\infty}\frac{1}{1+\gamma_n k^2}.$$

Константу регуляризации γ_n предлагается настраивать, используя статистику ω^2 Колмогорова–Смирнова:

$$\omega_n^2 = n\int_{-\infty}^{+\infty}(F(x) - F_n(x))^2 dF(x).$$

Тогда γ_n для данного объёма выборки n выбирается таким образом, чтобы

$$\int_{-\infty}^{+\infty}(\hat{F} - F_n(x))^2 d\hat{F}(x) = \frac{\omega_n^2}{n},$$

где

$$\hat{F}(x) = \int_{-\infty}^{+\infty}\eta(x-t)\hat{f}(t)dt = \int_{-\infty}^{x}\hat{f}(t)dt,$$

и $\hat{f}(x)$ есть оценка (1.15).

Также γ_n может быть настроена с использованием других статистик [48], [60].

Отметим, что в большом количестве работ, посвящённых исследованию проекционной оценки, предполагается, что оцениваемая функция плотности вероятности $f(x)$ принадлежит пространству L_2. В общем случае для функций плотности вероятности, не принадлежащих L_2, сходимость проекционной оценки не гарантируется. Подробнее этот вопрос обсуждается в главе 2.

§ 1.3. Ядерные оценки

Ядерная оценка функции плотности вероятности была впервые предложена М. Розенблаттом [25], а затем обобщена Э. Парзеном [20]. Ядерная оценка Розенблатта – Парзена имеет вид:

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x-x_i}{h}\right), \tag{1.16}$$

где $h > 0$ – *параметр размытости*, а функция $K(x)$, называемая *ядерной функцией* или *ядром*, удовлетворяет условиям:

а) измерима по Борелю;

б) $\sup\limits_{x}|K(x)| < +\infty$;

в) $\displaystyle\int_{-\infty}^{+\infty}|K(x)|dx = 1$; (1.17)

г) $\lim\limits_{x\to\pm\infty}|xK(x)| = 0$.

Распространённые примеры ядер приведены в табл. 1.1.

Таблица 1.1. Примеры ядерных функций

№	$K(x)$	**Название**
1	$\begin{cases}\frac{1}{2}, & \lvert x\rvert \le 1\\ 0, & \text{иначе}\end{cases}$	прямоугольное
2	$\begin{cases}1-\lvert x\rvert, & \lvert x\rvert \le 1\\ 0, & \text{иначе}\end{cases}$	треугольное
3	$\begin{cases}\frac{3}{4}(1-x^2), & \lvert x\rvert \le 1\\ 0, & \text{иначе}\end{cases}$	параболическое
4	$\begin{cases}(1+2\lvert x\rvert)(1-\lvert x\rvert)^2, & \lvert x\rvert \le 1\\ 0, & \text{иначе}\end{cases}$	кубическое
5	$\begin{cases}\frac{1}{\pi}\sqrt{1-x^2}, & \lvert x\rvert \le 1\\ 0, & \text{иначе}\end{cases}$	эллиптическое
6	$\dfrac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$	гауссово

№	$K(x)$	**Название**
7	$\frac{1}{2}e^{-\lvert x\rvert}$	экспоненциальное
8	$\frac{1}{\pi}\frac{1}{1+x^2}$	Коши
9	$\frac{1}{2\pi}\left(\frac{\sin x/2}{x/2}\right)^2$	–

Заметим, что любое ядро можно преобразовать следующим образом:

$$\bar{K}(x) = \frac{1}{c}K\left(\frac{x}{c}\right), \quad c > 0.$$

Данное преобразование представляет собой сжатие вдоль оси абсцисс и растяжение вдоль оси ординат в c раз. Изменение ядра $K(x)$ на $\bar{K}(x)$, очевидно, не изменяет вида оценки (1.16), так как это приводит лишь к умножению параметра размытости на c. Использование ядра $\bar{K}(x)$ с определённым выбором коэффициента c иногда бывает удобным при теоретических исследованиях.

Например [75], можно выбрать c таким образом, чтобы

$$\int_{-\infty}^{+\infty} x^2\bar{K}(x)dx = 1. \tag{1.18}$$

Искомое значение c равно

$$c = \frac{1}{\int_{-\infty}^{+\infty} x^2K(x)dx}.$$

В работе [75] оценка (1.16) обобщена на многомерный случай:

$$\hat{f}(\mathbf{x}) = \hat{f}(x_1, \ldots, x_k) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{k}\frac{1}{h_j}K_j\left(\frac{x_j - x_{ji}}{h_j}\right), \tag{1.19}$$

функции $K_j(x)$ подчинены более жёстким по сравнению с (1.17) условиям:

$$\begin{aligned}
&\text{а)}\quad 0 \le K_j(x) < C < +\infty;\\
&\text{б)}\quad K_j(x) = K_j(-x);\\
&\text{в)}\quad \int_{-\infty}^{+\infty} K_j(x)dx = 1;\\
&\text{г)}\quad \int_{-\infty}^{+\infty} K_j(x)x^2dx = 1;\\
&\text{д)}\quad \int_{-\infty}^{+\infty} K_j(x)x^m dx < +\infty \text{ при } m \in [0; +\infty).
\end{aligned} \tag{1.20}$$

При выполнении условий (1.20) на ядерные функции, а также условий

$$
\begin{aligned}
&\text{а)} \quad \lim_{n\to\infty} h_j(n) = 0; \\
&\text{б)} \quad \lim_{n\to\infty} n \prod_{j=1}^{k} h_j(n) = +\infty
\end{aligned}
\tag{1.21}
$$

на параметры размытости, оценка (1.19) состоятельна.

Также в работе [75] проводится оптимизация оценки (1.19) по форме ядра в смысле критерия (1.4). Показано, что оптимальная ядерная функция при достаточно большом объёме выборки n не зависит от вида истинной плотности $f(x)$, размерности k и является масштабированной параболической функцией с коэффициентом масштабирования, рассчитанным по формуле (1.18):

$$
K^*(x) = \begin{cases} \frac{3}{4\sqrt{5}}\left(1 - \frac{y^2}{5}\right), & |y| < \sqrt{5} \\ 0, & \text{иначе} \end{cases}.
$$

Также приводится асимптотически оптимальное значение параметра размытости

$$
h_j(n) = h(n) \sim \left(\frac{kL^k}{nM}\right)^{\frac{1}{k+4}}, \tag{1.22}
$$

где

$$
L = \int_{-\infty}^{+\infty} K^2(x)dx,
$$

$$
M = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left(\sum_{i=1}^{k} \frac{\partial^2 f(x_1, ..., x_k)}{\partial x_j^2}\right)^2 dx_1 \dots dx_k.
$$

Выражение (1.22) использует вторые частные производные истинной плотности $f(x)$. Более грубая оценка параметра размытости, полученная в [75], не зависящая от вида истинной плотности вероятности, имеет вид:

$$
h(n) \sim \left(\frac{4}{n(k+2)}\right)^{\frac{1}{k+4}}.
$$

В работе [48] оценка Розенблатта – Парзена рассматривается как частный случай оценки, получаемой методом стохастической регуляризации. Полученная в данной работе оценка имеет вид

$$
\hat{f}(x) = \frac{1}{2n\sqrt{\gamma_n}} \sum_{i=1}^{n} e^{-\frac{|x-x_n|}{\sqrt{\gamma_n}}}.
$$

Данная оценка является оценкой Розенблатта–Парзена с экспоненциальным ядром, в которой параметр размытости является функцией от константы регуляризации γ_n.

Способ оценки параметра размытости в условиях ограниченного объёма выборки n предложен в работе [91]. Функционал (1.6) в случае оценки (1.16) принимает вид:

$$W\{\hat{f}\} = M\left\{\frac{1}{h^2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\int_{\mathbb{R}} K\left(\frac{x-x_i}{h}\right)K\left(\frac{x-x_j}{h}\right)dx - \right.$$
$$\left. - \frac{2}{hn}\sum_{i=1}^{n} M\left\{K\left(\frac{\xi-x_i}{h}\right)\right\}\right\}.$$

В первом интеграле полученного выражения сделаем подстановку

$$x = zh + x_i.$$

Тогда

$$\frac{1}{h}\int_{\mathbb{R}} K\left(\frac{x-x_i}{h}\right)K\left(\frac{x-x_j}{h}\right)dx = \int_{\mathbb{R}} K(z)K\left(z+\frac{x_i-x_j}{h}\right)dz.$$

Правая часть полученного равенства зависит только от вида ядра $K(z)$ и значения $\frac{x_i-x_j}{h}$. Обозначим её через $\tau\left(\frac{x_i-x_j}{h}\right)$. Тогда

$$\int_{\mathbb{R}} K\left(\frac{x-x_i}{h}\right)K\left(\frac{x-x_j}{h}\right)dx = h\tau\left(\frac{x_i-x_j}{h}\right);$$

$$W\{\hat{f}\} = M\left\{\frac{1}{hn^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\tau\left(\frac{x_i-x_j}{h}\right) - \frac{2}{hn}\sum_{i=1}^{n} M\left\{K\left(\frac{\xi-x_i}{h}\right)\right\}\right\}.$$

Для построения несмещённой оценки функционала W, очевидно, достаточно построить несмещённую оценку его второго слагаемого. Но

$$\sum_{i=1}^{n} M\left\{K\left(\frac{\xi-x_i}{h}\right)\right\} = M\left\{\frac{1}{n-1}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} K\left(\frac{x_j-x_i}{h}\right)\right\},$$

откуда получаем

$$W\{\hat{f}\} = M\left\{\frac{1}{hn^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\tau\left(\frac{x_i-x_j}{h}\right) - \frac{2}{hn(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} K\left(\frac{x_j-x_i}{h}\right)\right\}.$$

Таким образом, в качестве несмещённой оценки $\hat{W}$ функционала W берётся выражение

$$\hat{W}(h) = \frac{1}{hn^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \tau \left(\frac{x_i - x_j}{h} \right) - \frac{2}{hn(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} K \left(\frac{x_j - x_i}{h} \right).$$

В случае оценки (2.35) k-мерного случайного вектора аналогично получаем

$$\hat{W}(h_1, \ldots, h_k) = \frac{1}{n^2 h_1 \ldots h_k} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \prod_{j=1}^{k} \tau \left(\frac{x_{i_1,j} - x_{i_2,j}}{h_j} \right) - \\ - \frac{2}{n(n-1) h_1 \ldots h_k} \sum_{i_1=1}^{n} \sum_{\substack{i_2=1 \\ i_2 \neq i_1}}^{n} \prod_{j=1}^{k} K \left(\frac{x_{i_1,j} - x_{i_2,j}}{h_j} \right). \quad (1.23)$$

Оценки $\hat{h}_1, \ldots, \hat{h}_k$ параметров размытости находятся путём минимизации оценки (1.23):

$$(\hat{h}_1, \ldots, \hat{h}_k) = \arg \min_{h_1 \ldots h_k} \hat{W}(h_1, \ldots, h_k). \quad (1.24)$$

Другой способ оценивания параметра размытости основан на применении метода максимального правдоподобия [88]. Вводится функция правдоподобия в виде произведения значений оценки плотности вероятности, построенной по всем точкам x_j, кроме данной:

$$L(h) = \prod_{i=1}^{n} \frac{1}{(n-1)h} \sum_{\substack{j=1 \\ j \neq i}}^{n} K \left(\frac{x_i - x_j}{h} \right).$$

Оценка параметра размытости получается максимизацией этой функции:

$$\hat{h} = \arg \max_{h} L(h). \quad (1.25)$$

На практике обычно максимизируют не саму функцию правдоподобия, а её логарифм:

$$\hat{h} = \arg \max_{h} \ln L(h),$$

где

$$\ln L(h) = \sum_{i=1}^{n} \ln \sum_{\substack{j=1 \\ j \neq i}}^{n} K \left(\frac{x_i - x_j}{h} \right) - n \ln((n-1)h).$$

При этом иногда применяются методы оптимизации, основанные на использовании 1-й производной:

$$\frac{d}{dh}\ln L(h) = -\frac{1}{h^2}\sum_{i=1}^{n}\frac{\sum\limits_{\substack{j=1\\ j\neq i}}^{n} K'\left(\dfrac{x_i-x_j}{h}\right)}{\sum\limits_{\substack{j=1\\ j\neq i}}^{n} K\left(\dfrac{x_i-x_j}{h}\right)} - \frac{n}{h}.$$

В многомерном случае получаем:

$$\ln L(h_1,\ldots,h_k) = \sum_{i_1=1}^{n}\ln\sum_{\substack{i_2=1\\ i_2\neq i_1}}^{n}\prod_{j=1}^{k} K\left(\frac{x_{i_1,j}-x_{i_2,j}}{h}\right) - n\ln((n-1)h_1\ldots h_k); \quad (1.26)$$

$$\frac{\partial}{\partial h_m}\ln L(h_1,\ldots,h_k) =$$

$$= -\frac{1}{h_m^2}\sum_{i_1=1}^{n}\frac{\sum\limits_{\substack{i_2=1\\ i_2\neq i_1}}^{n}\prod\limits_{\substack{j=1\\ j\neq m}}^{k} K\left(\dfrac{x_{i_1,j}-x_{i_2,j}}{h_j}\right)\cdot K'\left(\dfrac{x_{i_1,m}-x_{i_2,m}}{h_m}\right)(x_{i_1,m}-x_{i_2,m})}{\sum\limits_{\substack{i_2=1\\ i_2\neq i_1}}^{n}\prod\limits_{j=1}^{k} K\left(\dfrac{x_{i_1,j}-x_{i_2,j}}{h_j}\right)} - \frac{n}{h_m},$$

$$m = 1,\ldots,k.$$

Кроме того для оценивания параметра размытости примени́м метод k ближайших соседей [88]. В этом случае параметр размытости h настраивается отдельно для каждого выборочного значения x_i таким образом, чтобы интервал $(x_i - h; x_i + h)$ содержал ровно k точек из выборки, где $1 \leq k \leq n-1$. Тогда оценка Розенблатта–Парзена принимает вид

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i(k)}K\left(\frac{x-x_i}{h_i(k)}\right).$$

Параметр k предлагается настраивать методом максимального правдоподобия.

В монографии [88] предложены некоторые модификации оценки Розенблатта – Парзена, которые мы также отнесём к типу ядерных оценок. Пусть

$K(x)$ – произвольная ядерная функция (см. табл. 1.1) и $\beta > 0$. Тогда *интегральная оценка* функции плотности вероятности имеет вид:

$$\hat{f}(x) = \frac{1}{2\beta hn} \sum_{i=1}^{n} \int_{x-\beta}^{x+\beta} K\left(\frac{z - x_i}{h}\right) dz.$$

При $\beta \to +0$ интегральная оценка переходит в оценку Розенблатта–Парзена:

$$\lim_{\beta \to +0} \frac{1}{2\beta hn} \sum_{i=1}^{n} \int_{x-\beta}^{x+\beta} K\left(\frac{z - x_i}{h}\right) dz = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right).$$

Настройка параметров h и β проводится аналогично случаю оценки Розенблатта–Парзена.

Пусть теперь исследуемая случайная величина имеет ограниченный носитель $[a; b]$. Разобьём отрезок $[a; b]$ на N промежутков одинаковой длины. Обозначим через z_j середину j-го промежутка, а через n_j – количество выборочных значений, попавших в j-й промежуток. Тогда *регрессионная оценка* функции плотности вероятности имеет вид:

$$\hat{f}(x) = \frac{1}{h} \sum_{j=1}^{N} \frac{n_j}{n} K\left(\frac{x - z_j}{h}\right).$$

В этом случае настройке подлежат параметры h и N.

§ 1.4. Другие виды оценок

Наиболее распространённым методом непараметрического оценивания закона распределения случайной величины является эмпирическая плотность вероятности или гистограмма [49]. Гистограмма может использоваться при исследовании как непрерывных, так и дискретных случайных величин.

Гистограмму можно рассмотреть как кусочно постоянную функцию $H_n(x)$, определённую на некотором отрезке $[a; b]$. Если задано разбиение (в общем случае, неравномерное) этого отрезка точками

$$a = d_0 < d_1 < \cdots < d_N = b,$$

то $H_n(x)$ можно определить следующим образом:

$$H_n(x) = \sum_{j=1}^{N} \frac{n_j}{n(d_j - d_{j-1})},$$

где n – объём выборки, n_j – число значений случайной величины, попавших в интервал $(d_{j-1}; d_j)$.

Часто встречается случай, когда задано разбиение отрезка $[a; b]$ на интервалы равной длины. При этом

$$H_n(x) = \frac{N}{b-a} \sum_{j=1}^{N} \frac{n_j}{n}.$$

Для оценивания подходящего значения N имеется несколько подходов. Приведём наиболее известные:

1) формула Стерджеса [31]:
$$N = 1 + \log_2 n;$$

2) формула Манна–Вальда [14]:
$$N = 4 \left(\frac{3}{4}(n-1)^2 \right)^{1/5};$$

3) формула Брукса–Каррузера [147]:
$$N = 5 \lg n;$$

4) формула, предложенная в работе [52]:
$$N = \frac{4}{\varkappa} \lg \frac{n}{10},$$
где $\varkappa = \frac{1}{\sqrt{\mu_4/\sigma^4}}$ – коэффициент контрэксцесса;

5) формула, предложенная в работе [146]:
$$N = \left(\frac{n \max |f''(x)|}{4 \max f(x)} \right)^{1/5};$$

6) формула Скотта [27]:
$$N = \left(\frac{n}{6} \int_{-\infty}^{+\infty} (f'(x))^2 dx \right)^{1/3}. \qquad (1.27)$$

В работе [89] предлагается длину интервала выбирать равной удвоенному значению оптимального параметра размытости для оценки Розенблатта–Парзена:

$$h^* = \left(\frac{\int_{-\infty}^{+\infty} K^2(x) dx}{n \int_{-\infty}^{+\infty} (f''(x))^2 dx} \right)^{1/5},$$

где $K(x)$ – параболическое ядро, откуда

$$N = \frac{b-a}{2h^*}.$$

В работе [48] оценка плотности вероятности в виде гистограммы получена методом стохастической регуляризации путём минимизации функционала

$$R(f, F_n) = \int_a^b (F(x) - F_n(x))^2 (d + |f'(x)|) dx + \gamma_n \int_a^b f^2(x) dx,$$

где $F(x)$ и $F_n(x)$ – соответственно теоретическая и эмпирическая функции распределения, d и γ_n – параметры. Показано, что если истинная плотность абсолютно непрерывна, то её оценка, полученная данным способом сходится к ней в метрике пространства $L_2[a; b]$.

В работе [128] дана оценка скорости сходимости гистограммы с увеличением объёма выборки:

$$Q_2\{f_n\} \sim n^{-1/3}.$$

Если оцениваемая плотность дважды дифференцируема, то, как показано в [27],

$$Q_2\{f_n\} = \frac{1}{nh} + \frac{1}{12} h^2 \|f'\|^2 + O\left(\frac{1}{n} + h^2\right).$$

Отсюда, с учётом (1.27), получаем

$$Q_2\{f_n\} \sim n^{-2/3}.$$

В монографии [124] описывается оценка k *ближайших соседей* в виде

$$\hat{f}(x) = \frac{k-1}{2n\rho(x)},$$

где n – объём выборки, $\rho(x)$ – наименьший радиус интервала с центром в точке x, охватывающего k выборочных значений («ближайших соседей»).

Для многомерного случая оценка имеет вид:

$$\hat{f}(\mathbf{x}) = \frac{k-1}{nV(\mathbf{x})},$$

где $\mathbf{x} = (x_1, \ldots, x_m)$, $V(\mathbf{x})$ – объём наименьшего шара с центром в точке $\mathbf{x}$, охватывающего k выборочных значений.

Параметр $k = k_n$ выбирается в зависимости от объёма выборки n. Если выполнены условия

$$\lim_{n\to\infty} k_n = +\infty, \quad \lim_{n\to\infty} \frac{k_n}{n} = 0, \tag{1.28}$$

то оценка является асимптотически несмещённой и состоятельной в каждой точке непрерывности оцениваемой плотности $f(\mathbf{x})$.

Оценка плотности вероятности в виде

$$\hat{f}(x) = \frac{k}{(n-1)(x_{ik} - x_{i(k-1)})}, \quad x \in (x_{i(k-1)}; x_{ik}),$$

где $k \in \left\{1, 2, \ldots, \frac{n}{2}\right\}$ – настраиваемый параметр, называется полиграммой k-го порядка [131]. Как и в случае оценки k ближайших соседей, при выполнении условий (1.28) полиграмма является асимптотически несмещённой и состоятельной оценкой в каждой точке непрерывности плотности вероятности $f(x)$.

Выводы

В данной главе были рассмотрены основные типы непараметрических оценок функции плотности вероятности, а также известные методы их настройки. Сформулированы вопросы, требующие дополнительного исследования. В частности, проблема применимости проекционных оценок при восстановлении плотностей с немуммируемым квадратом, а также проблема сравнения эффективности различных непараметрических оценок при восстановлении плотности вероятности и при решении других практических задач.

Глава 2. Оптимизация проекционной оценки плотности вероятности

В данной главе рассматривается вопрос о применимости проекционной оценки при восстановлении функций плотности вероятности с несуммируемым квадратом. Предлагается положительное решение данного вопроса путём введения весовых функциональных пространств $L_{2,w}(\Omega)$. Также исследуется применение метода моментов при настройке коэффициентов проекционной оценки. В ходе исследования было получено некоторое обобщение метода моментов. Предлагается новый метод настройки длины ряда для проекционной оценки, в которой коэффициенты настраиваются обобщённым методом моментов. Производится сравнение предложенных методов с известными методами настройки проекционной оценки. В § 2.4 даётся обобщение разработанных методов на многомерный случай.

§ 2.1. Обоснование применимости проекционной оценки

Как уже упоминалось, необходимым условием применимости проекционной оценки плотности вероятности является требование, чтобы оцениваемая плотность $f(x)$ принадлежала пространству L_2. Однако данное требование не выполняется уже для некоторых модельных законов распределения. Рассмотрим несколько примеров.

1. Распределение Вейбулла:

$$f(x) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}, & x \geq 0 \\ 0, & \text{иначе} \end{cases},$$

где параметры $k, \lambda > 0$.

Найдём интеграл квадрата функции плотности:

$$\int_{-\infty}^{+\infty} f^2(x)dx = \frac{k^2}{\lambda^{2k}} \int_0^{+\infty} x^{2(k-1)} e^{-2\left(\frac{x}{\lambda}\right)^k} dx \geq \frac{k^2}{\lambda^{2k}} \int_0^{A} x^{2(k-1)} e^{-2\left(\frac{x}{\lambda}\right)^k} dx,$$

где $0 < A < +\infty$. При $k < 1$ последний интеграл является несобственным, так как

$$\lim_{x \to +0} x^{2(k-1)} e^{-2\left(\frac{x}{\lambda}\right)^k} = +\infty.$$

При этом подынтегральная функция является бесконечно большой порядка $(2-2k)$. Если теперь

$$2-2k \geq 1,$$

то интеграл расходится. Окончательно получаем, что при $0 < k \leq \frac{1}{2}$

$$\int_{-\infty}^{+\infty} f^2(x)dx \geq \frac{k^2}{\lambda^{2k}} \int_0^A x^{2(k-1)} e^{-2\left(\frac{x}{\lambda}\right)^k} dx = +\infty,$$

т.е. плотность распределения Вейбулла $f \notin L_2$.

2. Гамма-распределение:

$$f(x) = \begin{cases} \frac{x^{k-1}}{\lambda^k \Gamma(k)} e^{-\frac{x}{\lambda}}, & x \geq 0 \\ 0, & \text{иначе} \end{cases},$$

где параметры $k, \lambda > 0$.

Аналогично распределению Вейбулла получаем:

$$\int_{-\infty}^{+\infty} f^2(x)dx = \frac{1}{\lambda^{2k}\Gamma^2(k)} \int_0^{+\infty} x^{2k-2} e^{-\frac{2x}{\lambda}} dx \geq \frac{1}{\lambda^{2k}\Gamma^2(k)} \int_0^{A} x^{2k-2} e^{-\frac{2x}{\lambda}} dx.$$

Отсюда при $0 < k \leq \frac{1}{2}$ получаем $f \notin L_2$.

В частности, плотность распределения χ^2 с числом степеней свободы $k = 1$ (т.е. гамма-распределения с параметрами $k = \frac{1}{2}$, $\lambda = 2$):

$$f(x) = \frac{1}{2^{1/2}\Gamma\left(\frac{1}{2}\right)} x^{-\frac{1}{2}} e^{-\frac{x}{2}}, \quad x \geq 0$$

не принадлежит L_2.

3. Бета-распределение:

$$f(x) = \begin{cases} \frac{1}{\mathrm{B}(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, & x \in [0;1] \\ 0, & \text{иначе} \end{cases},$$

где параметры $\alpha, \beta > 0$, B – бета-функция:

$$\mathrm{B}(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

При $\alpha \in \left(0; \frac{1}{2}\right]$ или $\beta \in \left(0; \frac{1}{2}\right]$ плотность бета-распределения $f(x)$ не принадлежит L_2. Действительно,

$$\int_{-\infty}^{+\infty} f^2(x)dx = \frac{1}{\mathrm{B}^2(\alpha,\beta)} \int_0^1 \frac{1}{x^{2(1-\alpha)}} \frac{1}{(1-x)^{2(1-\beta)}} dx.$$

Теперь при $\alpha \in \left(0; \frac{1}{2}\right]$ последний интеграл расходится в окрестности точки $x = 0$, а при $\beta \in \left(0; \frac{1}{2}\right]$ – в окрестности точки $x = 1$.

В частности, распределение арксинуса имеет плотность

$$f(x) = \begin{cases} \frac{1}{\pi} x^{1/2}(1-x)^{1/2}, & x \in [0;1] \\ 0, & \text{иначе} \end{cases},$$

не принадлежащую L_2.

4. Распределение Фишера:

$$f(x) = \frac{1}{\mathrm{B}\left(\frac{k_1}{2}, \frac{k_2}{2}\right)} \left(\frac{k_1}{k_2}\right)^{\frac{k_1}{2}} \frac{x^{\frac{k_1}{2}-1}}{\left(1 + \frac{k_1}{k_2}x\right)^{\frac{k_1+k_2}{2}}},$$

где параметры $k_1, k_2 \in \mathbb{N}$.

При $k_1 = 1$ (и любых k_2) плотность распределения Фишера $f(x)$ не принадлежит L_2. Действительно,

$$\int_{-\infty}^{+\infty} f^2(x)dx = \frac{1}{k_2 \mathrm{B}^2\left(\frac{1}{2}, \frac{k_2}{2}\right)} \int_0^{+\infty} \frac{1}{x\left(1 + \frac{x}{k_2}\right)^{k_2}} dx.$$

Последний интеграл расходится в окрестности точки $x = 0$.

В частности, плотность распределения отношения двух независимых случайных величин, подчинённых закону распределения χ^2, не принадлежит L_2.

Список примеров функций плотности с несуммируемым квадратом легко можно продолжить. Например, возьмём

$$f(x) = \begin{cases} \frac{1}{2\sqrt{x}}, & x \in [0;1] \\ 0, & \text{иначе} \end{cases}.$$

Тогда

$$\int_{-\infty}^{+\infty} f^2(x)dx = \frac{1}{4} \int_0^1 \frac{dx}{x} = +\infty.$$

Приведённые примеры показывают, что требование принадлежности оцениваемой плотности $f(x)$ пространству L_2 является существенным ограничением.

Данное ограничение частично удаётся снять, если рассматривать их сходимость в пространствах $L_p(\Omega)$, где p принимает некоторое известное значение из промежутка $[1; 2)$. Как показано ниже, в этом случае пространство $L_p(\Omega)$ содержит более широкое множество функций плотности вероятности, чем $L_2(\Omega)$.

Предложение 2.1. *Пусть* $1 \leq p_1 \leq p_2$. *Тогда*

$$L_{p_2}(\Omega) \cap L_1(\Omega) \subseteq L_{p_1}(\Omega) \cap L_1(\Omega).$$

Доказательство. Возьмём произвольную функцию $f \in L_{p_2}(\Omega) \cap L_1(\Omega)$. Тогда для неё выполняется система неравенств

$$\begin{cases} \displaystyle\int_\Omega |f(x)| dx < +\infty \\ \displaystyle\int_\Omega |f(x)|^{p_2} dx < +\infty \end{cases}.$$

Разобьём множество Ω на два непересекающихся подмножества

$$\Omega_1 = \{x \in \Omega \mid |f(x)| < 1\},$$

$$\Omega_2 = \Omega \setminus \Omega_1 = \{x \in \Omega \mid |f(x)| \geq 1\}.$$

Оценим интеграл

$$\int_\Omega |f(x)|^{p_1} dx = \int_{\Omega_1} |f(x)|^{p_1} dx + \int_{\Omega_2} |f(x)|^{p_1} dx.$$

Первое слагаемое конечно:

$$\int_{\Omega_1} |f(x)|^{p_1} dx \leq \int_{\Omega_1} |f(x)| dx \leq \int_\Omega |f(x)| dx < +\infty.$$

Второе слагаемое также конечно:

$$\int_{\Omega_2} |f(x)|^{p_1} dx \leq \int_{\Omega_2} |f(x)|^{p_2} dx \leq \int_\Omega |f(x)|^{p_2} dx < +\infty.$$

Тогда конечна и вся сумма:

$$\int_{\Omega_1} |f(x)|^{p_1} dx + \int_{\Omega_2} |f(x)|^{p_1} dx < +\infty.$$

Получаем, что для функции $f(x)$ выполняется система неравенств

$$\begin{cases} \int_\Omega |f(x)| dx < +\infty \\ \int_\Omega |f(x)|^{p_1} dx < +\infty \end{cases},$$

из чего следует $f \in L_{p_1}(\Omega) \cap L_1(\Omega)$. □

Из предложения 2.1 следует, что если $p \in [1; 2)$, то

$$L_2 \cap L_1 \subseteq L_p \cap L_1.$$

Таким образом, *множество функций плотности вероятности с интегрируемой p-степенью* ($p \in [1; 2)$) *является более широким по сравнению с множеством плотностей с интегрируемым квадратом.*

Ниже приводятся несколько известных теорем о сходимости рядов Фурье в пространствах $L_p(\Omega)$, также следствия из них, используемые в построении проекционных оценок.

Теорема 2.1. *Пусть* $f \in L_p[0; 2\pi]$, *где* $p \in (1; +\infty)$, S_n *– частичная сумма тригонометрического ряда Фурье:*

$$S_n = \frac{a_0}{2} + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx),$$

$$a_k = \int_0^{2\pi} f(x) \cos kx \, dx, \quad b_k = \int_0^{2\pi} f(x) \sin kx \, dx.$$

Тогда

$$\lim_{n \to \infty} \|f - S_n\|_{L_p} = 0.$$

Доказательство этой теоремы приводится, например, в [55, с. 594].

Следствие 2.1.1. *Пусть* $f \in \mathcal{D} \cap L_p[0; 2\pi]$, *где* $p \in (1; +\infty)$. *Тогда проекционная оценка* $f_l(x)$ *с оптимальными коэффициентами, построенная по тригонометрической системе, сходится к* f *в пространстве* $L_p[0; 2\pi]$*:*

$$\lim_{l \to \infty} \|f - f_l\|_{L_p[0;2\pi]} = 0.$$

Теорема 2.2. *Система Хаара является базисом в пространстве* $L_p[0; 1]$ *при* $p \in [1; +\infty)$.

Доказательство приводится в [79, с. 75].

Следствие 2.2.1. *Пусть* $f \in \mathcal{D} \cap L_1[0; 1]$. *Тогда проекционная оценка* f_l *с оптимальными коэффициентами, сходится к* f *в пространстве* $L_1[0; 1]$.

Таким образом, с помощью базиса Хаара можно строить сходящиеся проекционные оценки для плотностей вероятности любых непрерывных случайных величин с ограниченными носителями.

Далее приводятся результаты численных экспериментов, подтверждающих полученные теоретические результаты, а также в ситуациях, когда ответ на вопрос о сходимости неизвестен.

Пример 1. Функция

$$f(x) = \begin{cases} \frac{1}{2\sqrt{x}}, & x \in [0;1] \\ 0, & \text{иначе} \end{cases},$$

рассмотренная выше, не принадлежит пространству $L_2[0;1]$. Однако эта функция принадлежит пространству $L_p[0;1]$, например, при $p = \frac{3}{2}$:

$$\int_{-\infty}^{+\infty} |f(x)|^{3/2} dx = \sqrt{2} < +\infty.$$

В соответствии с теоремами 2.1 и 2.2, ортогональные ряды для этой функции, построенные с использованием тригонометрической системы и системы Хаара, сходятся к $f(x)$ в пространстве $L_{3/2}(\Omega)$.

Найдём коэффициенты разложения функции $f(x)$ по тригонометрической системе:

$$\alpha_0 = \int_{-\pi}^{\pi} f(x)\varphi_0(x)dx = \frac{1}{\sqrt{2\pi}},$$

$$\alpha_{2k-1} = \int_{-\pi}^{\pi} f(x)\varphi_{2k-1}(x)dx = \frac{1}{\sqrt{\pi k}}\mathrm{S}(\sqrt{k}), \quad k = 1, 2, \ldots$$

$$\alpha_{2k} = \int_{-\pi}^{\pi} f(x)\varphi_{2k}(x)dx = \frac{1}{\sqrt{\pi k}}\mathrm{C}(\sqrt{k}), \quad k = 1, 2, \ldots$$

где $\mathrm{S}(x)$ и $\mathrm{C}(x)$ – интегралы Френеля [44]:

$$\mathrm{S}(x) = \sqrt{\frac{2}{\pi}} \int_0^x \sin t^2 dt, \quad \mathrm{C}(x) = \sqrt{\frac{2}{\pi}} \int_0^x \cos t^2 dt.$$

В табл. 2.1 приведены значения расстояния оценки $f_l(x)$ от истинной плотности $f(x)$ в пространстве $L_1[-\pi;\pi]$ при различных значениях длины ряда l.

Таблица 2.1. Сходимость проекционной оценки с тригонометрическим базисом

l	10	20	30	40	50
$\|f - f_l\|_{L_1}$	0.801	0.674	0.611	0.564	0.509

Найдём коэффициенты разложения функции $f(x)$ по системе Хаара:

$$\alpha_0 = \int_0^1 f(x)\varphi_0(x)dx = 1,$$

$$\alpha_{n,k} = \int_0^1 f(x)\varphi_{n,k}(x)dx = \sqrt{4k-2} - \sqrt{k} - \sqrt{k-1}.$$

В табл. 2.2 приведены значения расстояния оценки $f_l(x)$ от истинной плотности $f(x)$ в пространстве $L_1[0;1]$ при различных значениях длины ряда l. В силу специфики системы Хаара длина ряда является степенью числа 2.

Таблица 2.2. Сходимость проекционной оценки с базисом Хаара

l	8	16	32	64	128
$\|f - f_l\|_{L_1}$	0.205	0.148	0.106	0.076	0.054

Далее приводятся результаты разложения функции исследуемой функции плотности вероятности по системе Лежандра. Сходимость ортогонального ряда в этом случае не гарантируется. Коэффициенты разложения имеют вид:

$$\alpha_k = \begin{cases} \sqrt{\dfrac{2}{2k+1}}, & k = 0, 1, 4, 5, \ldots \\ -\sqrt{\dfrac{2}{2k+1}}, & k = 2, 3, 6, 7, \ldots \end{cases}$$

В табл. 2.3 приведены значения расстояния оценки $f_l(x)$ от истинной плотности $f(x)$ в пространстве $L_1[-1;1]$ при различных значениях длины ряда l.

Таблица 2.3. Сходимость проекционной оценки с базисом Лежандра

l	10	20	30	40	50
$\|f - f_l\|_{L_1}$	0.525	0.430	0.378	0.345	0.320

Как показали численные расчёты, в этом случае система Лежандра показывает удовлетворительные результаты.

Пример 2. Рассмотренная выше функция плотности вероятности распределения χ^2 с одной степенью свободы имеет несуммируемый квадрат, однако, она принадлежит пространству $L_{3/2}[0;+\infty)$:

$$\int_0^{+\infty} |f(x)|^{3/2} dx = \frac{\Gamma\left(\frac{1}{4}\right) - \Gamma\left(\frac{1}{4}, \frac{3}{4}\right)}{\sqrt[4]{6\pi^3}} < +\infty,$$

где $\Gamma(\alpha, x)$ – неполная гамма-функция [129]:

$$\Gamma(\alpha, x) = \int_x^{+\infty} e^{-t} t^{\alpha-1} dt.$$

Коэффициенты разложения функции $f(x)$ по системе Лагерра имеют вид:

$$\alpha_j = \int_0^{+\infty} f(x)\varphi_j(x)dx = \frac{(2j-1)!!}{2^{j+\frac{1}{2}} j!}.$$

В табл. 2.4 приведены значения расстояния оценки $f_l(x)$ от истинной плотности $f(x)$ в пространстве $L_1[0;+\infty)$ при различных значениях длины ряда l.

Таблица 2.4. Сходимость проекционной оценки с базисом Лагерра

l	10	20	30	40	50
$\|f - f_l\|_{L_1}$	19.22	132.45	677.76	2852.27	10443.46

Как показали численные расчёты, в этом случае система Лагерра показывает неудовлетворительные результаты.

Таким образом, при построении проекционной оценки функции плотности вероятности в пространствах $L_p(\Omega)$ сходимость не гарантируется в случае неограниченных множеств Ω. В работах [150] определяются весовые функциональные пространства вида $L_{p,w}(\Omega)$. Показано, что для любой функции плотности вероятности имеется хотя бы одно такое пространство, которому она принадлежит.

Пусть $k \in \mathbb{N}$, μ – мера Лебега в $\mathbb{R}^k$, Ω – μ-измеримое подмножество в $\mathbb{R}^k$, $p \geq 1$, $w : \Omega \to \mathbb{R}$, причём

1) $w(\mathbf{x})$ μ-измерима;

2) $w(\mathbf{x})$ положительная почти всюду на Ω;

3) $\operatorname*{ess\,sup}_{\mathbf{x}\in\Omega} w(\mathbf{x}) < +\infty$.

Определение 2.1. ***Весовое функциональное пространство*** $L_{p,w}(\Omega)$ *определим как множество* μ*-измеримых функций* $f : \Omega \to \mathbb{R}$*, удовлетворяющих неравенству*

$$\left(\int_\Omega |f(\mathbf{x})|^p w(\mathbf{x}) d\mu\right)^{1/p} < +\infty.$$

Функцию $w(\mathbf{x})$ *при этом будем называть* ***весовой функцией***.

Левая часть неравенства в определении 2.1 представляет собой норму функции $f(\mathbf{x})$ в пространстве $L_{p,w}(\Omega)$, которую мы обозначим как $\|f\|_{p,w}$.

Отметим важнейшие свойства пространства $L_{p,w}(\Omega)$.

Предложение 2.2. *Всякое пространство* $L_{p,w}(\Omega)$ *является полным по норме* $\|\cdot\|_{p,w}$.

Доказательство. Весовая функция $w(\mathbf{x})$ индуцирует меру на Ω:

$$\mu_w X = \int_X w(\mathbf{x})\, d\mu.$$

При этом множество X является μ_w-измеримым тогда и только тогда, когда оно μ-измеримо, т.е. область определения для μ_w есть система измеримых по Лебегу подмножеств множества Ω. Из определения весовой функции $w(\mathbf{x})$ непосредственно следует, что полученная мера μ_w является σ-конечной, откуда следует полнота пространства $L_{p,w}(\Omega)$. $\square$

Предложение 2.3. *Всякое пространство $L_{p,w}(\Omega)$ является сепарабельным.*

Доказательство. Покажем, что мера μ_w, индуцированная весовой функцией $w(\mathbf{x})$, имеет счётный базис. Так как мера Лебега μ на $\mathbb{R}^k$ имеет счётный базис, то ограничение этой меры на измеримое подмножество $\Omega \subseteq \mathbb{R}$ также имеет счётный базис. Обозначим его через $\{A_n\}$. В силу вышесказанного, каждое A_n является μ_w-измеримым. Покажем, что множество $\{A_n\}$ является базисом для μ_w.

Возьмём произвольное μ_w-измеримое множество $X \subseteq \Omega$ и произвольное $\varepsilon > 0$. В силу абсолютной непрерывности интеграла Лебега существует такое $\delta > 0$, что для любого множества Y из $\mu Y < \delta$ следует

$$\int_Y w(x) d\mu < \varepsilon.$$

Возьмём теперь такое A_n, чтобы выполнялось $\mu(A_n \bigtriangleup X) < \delta$. Тогда

$$\int_{A_n \bigtriangleup X} w(x) d\mu < \varepsilon,$$

что означает

$$\mu_w(A_n \bigtriangleup X) < \varepsilon.$$

Отсюда заключаем, что система множеств $\{A_n\}$ является счётным базисом меры μ_w. $\square$

При $p = 2$ пространство $L_{p,w}(\Omega)$ является евклидовым, так как в этом случае для любых $f, g \in L_{2,w}(\Omega)$ определено скалярное произведение:

$$(f, g)_w = \int_\Omega f(\mathbf{x}) g(\mathbf{x}) w(\mathbf{x}) d\mu.$$

Таким образом, пространство $L_{2,w}(\Omega)$ является (сепарабельным) гильбертовым пространством. Если $f \in L_{2,w}(\Omega)$ и выполняется условие (1.7), то для неё определена проекционная оценка (1.8).

Покажем, что пространств вида $L_{2,w}(\Omega)$ достаточно для построения проекционной оценки любой функции плотности вероятности. Сформулируем более общее предложение для всех пространств вида $L_{p,w}(\Omega)$.

Предложение 2.4. *Пусть* $\boldsymbol{\xi}$ *– непрерывный случайный вектор,* $f(\mathbf{x})$ *– его функция плотности вероятности, множество* $\Omega \subseteq \mathbb{R}^k$ *удовлетворяет условию* (1.7), $p \geq 1$. *Тогда существует такая весовая функция* $w(\mathbf{x})$, *что* $f \in L_{p,w}(\Omega)$.

Доказательство. Определим функцию $w(\mathbf{x})$ на Ω следующим образом:

$$w(\mathbf{x}) = \begin{cases} 1, & f(\mathbf{x}) \leq 1 \\ f^{1-p}(\mathbf{x}), & f(\mathbf{x}) > 1 \end{cases}$$

Тогда $w(\mathbf{x})$ является измеримой, положительной на Ω, и

$$\operatorname*{ess\,sup}_{\mathbf{x} \in \Omega} w(\mathbf{x}) = 1.$$

Таким образом, $w(\mathbf{x})$ является весовой функцией, и пространство $L_{p,w}(\Omega)$ определено.

Покажем, что $f \in L_{p,w}(\Omega)$. Оценим интеграл

$$\int_{\Omega} |f(\mathbf{x})|^p w(\mathbf{x}) d\mu. \tag{2.1}$$

Разобьём множество Ω на два подмножества

$$\Omega_1 = \{\mathbf{x} \in \Omega \mid f(\mathbf{x}) \leq 1\}, \qquad \Omega_2 = \{\mathbf{x} \in \Omega \mid f(\mathbf{x}) > 1\}.$$

Очевидно, что $\Omega = \Omega_1 \sqcup \Omega_2$ (здесь символ $\sqcup$ обозначает дизъюнктное объединение множеств [74, с. 8]). Тогда

$$\int_{\Omega} |f(\mathbf{x})|^p w(\mathbf{x}) d\mu = \int_{\Omega_1} |f(\mathbf{x})|^p w(\mathbf{x}) d\mu + \int_{\Omega_2} |f(\mathbf{x})|^p w(\mathbf{x}) d\mu.$$

Оценим каждый из полученных интегралов, используя то, что функция $f(\mathbf{x})$ неотрицательна почти всюду и суммируема на Ω:

$$\int_{\Omega_1} |f(\mathbf{x})|^p w(\mathbf{x}) d\mu = \int_{\Omega_1} f^p(\mathbf{x}) d\mu \leq \int_{\Omega_1} f(\mathbf{x}) d\mu \leq \int_{\Omega} f(\mathbf{x}) d\mu = 1;$$

$$\int_{\Omega_2} |f(\mathbf{x})|^p w(\mathbf{x}) d\mu = \int_{\Omega_2} f^p(\mathbf{x}) f^{1-p}(\mathbf{x}) d\mu = \int_{\Omega_2} f(\mathbf{x}) d\mu \leq \int_{\Omega} f(\mathbf{x}) d\mu = 1.$$

Из конечности обоих интегралов следует конечность интеграла (2.1). □

В частности, для любой функции плотности вероятности случайного вектора, существует соответствующее гильбертово пространство $L_{2,w}(\Omega)$.

В одномерном случае, т.е. при $k = 1$, при построении пространства $L_{2,w}(\Omega)$ весовая функция $w(x)$ может удовлетворять более слабым условиям [150].

В доказательстве предложения 2.4 при нахождении соответствующей весовой функции используется вид истинной плотности $f(\mathbf{x})$. Важным вопросом построения проекционных оценок является выбор весовой функции без использования вида оцениваемой плотности. Прежде всего выясним, при каких весовых функциях $w_1(\mathbf{x})$ и $w_2(\mathbf{x})$ пространства $L_{p,w_1}(\Omega)$ и $L_{p,w_2}(\Omega)$ являются тождественными.

Теорема 2.3. *Пусть $w(\mathbf{x})$ – измерима и положительна почти всюду на Ω. Тогда измеримая функция $g(\mathbf{x})$, удовлетворяющая условиям*

1) $\displaystyle\int_{\Omega} |g(\mathbf{x})| d\mu = +\infty;$

2) $\displaystyle\int_{\Omega} |g(\mathbf{x})| w(\mathbf{x}) d\mu < +\infty$

существует тогда и только тогда, когда

$$\operatorname*{ess\,inf}_{\mathbf{x}\in\Omega} w(\mathbf{x}) = 0. \tag{2.2}$$

Доказательство. Необходимость. Пусть, напротив, равенство (2.2) не выполняется:

$$\operatorname*{ess\,inf}_{\mathbf{x}\in\Omega} w(\mathbf{x}) = m > 0.$$

Тогда почти для всех $\mathbf{x} \in \Omega$

$$w(\mathbf{x}) \geq m.$$

Если $m = +\infty$, то почти всюду на Ω получаем $w(\mathbf{x}) = +\infty$. Из условия 2) получаем, что $g(x) = 0$ почти всюду на Ω, и условие 1) нарушается.

Пусть теперь $0 < m < +\infty$. Тогда

$$\int_{\Omega} |g(\mathbf{x})| w(\mathbf{x}) d\mu \geq m \int_{\Omega} |g(\mathbf{x})| d\mu = +\infty,$$

и условия 1) и 2) несовместны.

Достаточность. Пусть равенство (2.2) выполнено. Тогда

$$\forall\varepsilon > 0\ \mu\{\mathbf{x} \in \Omega \mid w(\mathbf{x}) < \varepsilon\} > 0.$$

Выберем последовательность $\varepsilon_n \downarrow 0$, например, $\varepsilon_n = \frac{1}{n}$, и обозначим

$$\Omega_n = \{\mathbf{x} \in \Omega \mid w(\mathbf{x}) < \varepsilon_n\}, \quad \Omega_0 = \Omega.$$

Тогда, очевидно,

$$\Omega_0 \supseteq \Omega_1 \supseteq \cdots \supseteq \Omega_n \cdots;$$

$$\mu\Omega_0 \geq \mu\Omega_1 \geq \ldots \geq \mu\Omega_n \ldots$$

В силу непрерывности меры Лебега получаем

$$\lim_{n\to\infty} \mu\Omega_n = \mu \bigcap_{n=1}^{\infty} \Omega_n = \mu\{\mathbf{x} \in \Omega \mid w(\mathbf{x}) = 0\} = 0.$$

Разобьём множество Ω на попарно не пересекающиеся подмножества следующим образом:

$$\Omega = (\Omega_0 \setminus \Omega_1) \sqcup (\Omega_1 \setminus \Omega_2) \sqcup \ldots \sqcup (\Omega_{n-1} \setminus \Omega_n) \sqcup \ldots$$

При этом

$$\lim_{n\to\infty} \mu(\Omega_{n-1} \setminus \Omega_n) \leq \lim_{n\to\infty} \mu\Omega_{n-1} = 0.$$

Кроме того, последовательность $\mu(\Omega_{n-1} \setminus \Omega_n)$ содержит бесконечное количество положительных членов.

Перенумеруем последовательность $\mu(\Omega_{n-1} \setminus \Omega_n)$, отбросив все нулевые члены (сходимость последовательности ε_n при этом не изменится), и будем искать функцию $g(\mathbf{x})$ в виде:

$$g(\mathbf{x}) = g_n \geq 0, \quad \mathbf{x} \in \Omega_{n-1} \setminus \Omega_n, \quad n \in \mathbb{N}.$$

Тогда интегралы в условиях 1) и 2) преобразуются следующим образом:

$$\int_\Omega |g(\mathbf{x})| d\mu = \sum_{n=1}^{\infty} g_n \mu(\Omega_{n-1} \setminus \Omega_n);$$

$$\int_\Omega |g(\mathbf{x})| w(\mathbf{x}) d\mu \leq \sum_{n=1}^{\infty} g_n \varepsilon_{n-1} \mu(\Omega_{n-1} \setminus \Omega_n).$$

При этом полагаем

$$\varepsilon_0 = \operatorname*{ess\,sup}_{\mathbf{x}\in\Omega} w(\mathbf{x}).$$

Подберём теперь последовательность g_n так, чтобы выполнялись условия

$$\sum_{n=1}^{\infty} g_n \mu(\Omega_{n-1} \setminus \Omega_n) = +\infty, \quad \sum_{n=1}^{\infty} g_n \varepsilon_{n-1} \mu(\Omega_{n-1} \setminus \Omega_n) < +\infty.$$

Для этого возьмём

$$g_n = \frac{1}{\mu(\Omega_{n-1} \setminus \Omega_n)} \left(1 - \frac{\varepsilon_n}{\varepsilon_{n-1}}\right)$$

Тогда

$$\sum_{n=1}^{\infty} g_n \varepsilon_{n-1} \mu(\Omega_{n-1} \setminus \Omega_n) = \sum_{n=1}^{\infty} (\varepsilon_{n-1} - \varepsilon_n) = \varepsilon_0 < +\infty;$$

$$\sum_{n=1}^{\infty} g_n \mu(\Omega_{n-1} \setminus \Omega_n) = \sum_{n=1}^{\infty} \left(1 - \frac{\varepsilon_n}{\varepsilon_{n-1}}\right) = \sum_{n=1}^{\infty} \left(1 - \frac{1/\varepsilon_{n-1}}{1/\varepsilon_n}\right).$$

Так как последовательность $\frac{1}{\varepsilon_n}$ монотонно возрастает и не ограничена, то по признаку Сапогова [134, с. 319], последний ряд расходится. Таким образом, искомая функция $g(\mathbf{x})$ определена почти всюду на Ω, а именно за исключением отброшенных элементов последовательности $\Omega_{n-1} \setminus \Omega_n$, имеющих меру 0. В случае необходимости функцию $g(\mathbf{x})$ можно доопределить в остальных точках произвольным образом. □

Следствие 2.3.1. *Для того, чтобы $L_{p,w_1}(\Omega) \setminus L_{p,w_2}(\Omega) \neq \varnothing$, необходимо и достаточно, чтобы*

$$\operatorname*{ess\,inf}_{\mathbf{x}\in\Omega} \frac{w_1(\mathbf{x})}{w_2(\mathbf{x})} = 0. \tag{2.3}$$

Доказательство. Действительно, функция $\frac{w_1(\mathbf{x})}{w_2(\mathbf{x})}$ удовлетворяет условию теоремы 2.3. Тогда условие (2.3) эквивалентно существованию такой функция $g(\mathbf{x})$, что одновременно

$$\int_{\Omega} |g(\mathbf{x})| d\mu = +\infty$$

и

$$\int_{\Omega} |g(\mathbf{x})| \frac{w_1(\mathbf{x})}{w_2(\mathbf{x})} d\mu < +\infty.$$

Определим функцию $f(\mathbf{x})$ следующим образом:

$$f(\mathbf{x}) = \left(\frac{g(\mathbf{x})}{w_2(\mathbf{x})}\right)^{1/p}.$$

Тогда

$$\int_{\Omega} |f(\mathbf{x})|^p w_2(\mathbf{x}) d\mu = \int_{\Omega} |g(\mathbf{x})| d\mu = +\infty;$$

$$\int_{\Omega} |f(\mathbf{x})|^p w_1(\mathbf{x}) d\mu = \int_{\Omega} |g(\mathbf{x})| \frac{w_1(\mathbf{x})}{w_2(\mathbf{x})} d\mu < +\infty.$$

□

Следствие 2.3.2. *Если существуют такие положительные константы m и M, что почти всюду на Ω выполняется*

$$0 < m \leq \frac{w_1(\mathbf{x})}{w_2(\mathbf{x})} \leq M < +\infty,$$

то пространства $L_{p,w_1}(\Omega)$ и $L_{p,w_2}(\Omega)$ совпадают:

$$L_{p,w_1}(\Omega) = L_{p,w_2}(\Omega).$$

Для построения пространства, включающего оцениваемую функцию плотности вероятности предлагается расширить «основное» пространство $L_2(\Omega)$, т.е. построить весовое пространство $L_{2,w}(\Omega)$, содержащее данную функцию плотности и пространство $L_2(\Omega)$ как собственное подмножество. Теорема 2.3 позволяет сформулировать критерий, при котором одно пространство является расширением другого.

Следствие 2.3.3. *Пространство $L_{p,w_1}(\Omega)$ является расширением пространства $L_{p,w_2}(\Omega)$, т.е. $L_{p,w_2}(\Omega) \subset L_{p,w_1}(\Omega)$, тогда и только тогда, когда выполняются следующие два условия:*

1) $\operatorname{ess\,inf}\limits_{\mathbf{x}\in\Omega} \dfrac{w_1(\mathbf{x})}{w_2(\mathbf{x})} = 0;$

2) $\operatorname{ess\,sup}\limits_{\mathbf{x}\in\Omega} \dfrac{w_1(\mathbf{x})}{w_2(\mathbf{x})} < +\infty.$

Таким образом, чтобы построить весовое расширение $L_{2,w}$ пространства L_2, необходимо и достаточно взять весовую функцию, удовлетворяющую условиям:

1) $\operatorname{ess\,inf}\limits_{\mathbf{x}\in\Omega} w(\mathbf{x}) = 0;$

2) $\operatorname{ess\,sup}\limits_{\mathbf{x}\in\Omega} w(\mathbf{x}) < +\infty.$

Рассмотрение условий на весовые функции, при которых одно весовое функциональное пространство является расширением другого, представляет теоретический интерес. Более прикладной задачей, однако, является получение условий, при которых одно весовое функциональное пространство содержит в себе более обширное множество функций плотности вероятности, чем другое. С этой целью доказывается следующая теорема.

Теорема 2.4. *Пусть весовая функция $w(\mathbf{x})$ удовлетворяет равенству (2.2), $p > 1$. Тогда найдётся измеримая на Ω функция $g(\mathbf{x})$, принадлежащая множеству $(L_{p,w}(\Omega) \setminus L_p(\Omega)) \cap L_1(\Omega)$.*

Доказательству теоремы 2.4 предпошлём следующую лемму.

Лемма 2.1. *Пусть числовые последовательности a_n и b_n удовлетворяют условиям:*

1) $a_n, b_n \geq 0$;
2) $\sum\limits_{n=1}^{\infty} a_n = +\infty, \sum\limits_{n=1}^{\infty} b_n = +\infty$;
3) $\varlimsup\limits_{n\to\infty} b_n < +\infty$;
4) $\lim\limits_{n\to\infty} \dfrac{a_n}{b_n} = 0$.

Тогда существует такая последовательность номеров n_k, что

$$\sum_{k=1}^{\infty} a_{n_k} < +\infty, \quad \sum_{k=1}^{\infty} b_{n_k} = +\infty.$$

Доказательство. Обозначим верхний предел последовательности b_n через b. Не умаляя общности, можно считать, что $b_n < b + 1$. В противном случае в рядах $\sum\limits_{n=1}^{\infty} a_n$ и $\sum\limits_{n=1}^{\infty} b_n$ можно отбросить конечное число первых членов. Теперь в силу условия 4) получаем, что для любого натурального m существует такой номер N', что из $n > N'$ следует неравенство

$$\frac{a_n}{b_n} < \frac{1}{2(b+1)m^2},$$

т.е.

$$a_n < \frac{1}{2(b+1)m^2} b_n.$$

Пусть значению $m = 1$ соответствует N_1'. Возьмём $N_1 = N_1'$ и подберём $M_1 \geq 2$ таким образом, чтобы

$$\sum_{n=N_1+1}^{N_1+M_1} b_n > b+1 \quad \text{и} \quad \sum_{n=N_1+1}^{N_1+M_1-1} b_n \leq b+1.$$

Это можно сделать в силу расходимости любого остатка расходящегося ряда. Тогда

$$\sum_{n=N_1+1}^{N_1+M_1} b_n = \sum_{n=N_1+1}^{N_1+M_1-1} b_n + b_{N_1+M_1} \leq 2(b+1);$$

$$\sum_{n=N_1+1}^{N_1+M_1} a_n < \frac{1}{2(b+1)m^2} \sum_{n=N_1+1}^{N_1+M_1} b_n \leq \frac{1}{m^2}.$$

Пусть теперь значению $m = 2$ соответствует N_2'. Возьмём $N_2 = \max\{N_2', N_1 + M_1\}$. Для него аналогично найдём значение M_2. И т.д. Получим последовательность пар (N_m, M_m), где $m \in \mathbb{N}$, каждой из которых соответствует конечный ряд номеров $\{N_m + 1, ..., N_m + M_m\}$. Тогда искомое множество номеров n_k возьмём в виде

$$\{n_k\} = \bigcup_{m=1}^{\infty} \{N_m + 1, ..., N_m + M_m\}.$$

Действительно,

$$\sum_{k=1}^{\infty} a_{n_k} = \sum_{m=1}^{\infty} (a_{N_m+1} + \cdots + a_{N_m+M_m}) \leq \sum_{m=1}^{\infty} \frac{1}{m^2} < +\infty;$$

$$\sum_{k=1}^{\infty} b_{n_k} = \sum_{m=1}^{\infty} (b_{N_m+1} + \cdots + b_{N_m+M_m}) \geq \sum_{m=1}^{\infty} (b+1) = +\infty.$$

$\square$

Доказательство теоремы 2.4. Другими словами, требуется доказать существование функции $g(\mathbf{x})$, удовлетворяющей условиям:

1) $\displaystyle\int_\Omega |g(\mathbf{x})|^p w(\mathbf{x}) d\mu < +\infty;$

2) $\displaystyle\int_\Omega |g(\mathbf{x})|^p d\mu = +\infty;$

3) $\displaystyle\int_\Omega |g(\mathbf{x})| d\mu < +\infty.$

Из равенства (2.2) по теореме 2.3 получаем, что существует такая функция $\tilde{g}(\mathbf{x})$, удовлетворяющая условиям 1) и 2). Если при этом $\tilde{g}(\mathbf{x})$ удовлетворяет условию 3), то можно взять $g(\mathbf{x}) = \tilde{g}(\mathbf{x})$.

Пусть $\tilde{g}(\mathbf{x})$ не удовлетворяет условию 3):

$$\int_\Omega |\tilde{g}(\mathbf{x})| d\mu = +\infty. \tag{2.4}$$

По построению функции $\tilde{g}(\mathbf{x})$ получаем, что

$$\tilde{g}^p(\mathbf{x}) = \tilde{g}^p_n = \frac{1}{\mu(\Omega_{n-1} \setminus \Omega_n)} \left(1 - \frac{\varepsilon_n}{\varepsilon_{n-1}}\right), \quad \mathbf{x} \in \Omega_{n-1} \setminus \Omega_n,$$

где ε_n – некоторая подпоследовательность исходной монотонно убывающей последовательности, сходящейся к 0. Пусть в качестве исходной взята последовательность $\frac{1}{2^n}$. Покажем, что в этом случае

$$\lim_{n\to\infty} \tilde{g}^p_n = +\infty. \tag{2.5}$$

Действительно, пусть $\frac{1}{2^{n_k}}$ – произвольная подпоследовательность последовательности $\frac{1}{2^n}$. Тогда

$$n_k - n_{k-1} \geq 1;$$

$$\frac{\varepsilon_{n_k}}{\varepsilon_{n_{k-1}}} = \frac{1}{2^{n_k - n_{k-1}}} \leq \frac{1}{2};$$

$$\lim_{k\to\infty} \tilde{g}^p_{n_k} = \lim_{k\to\infty} \frac{1}{\mu(\Omega_{n_{k-1}} \setminus \Omega_{n_k})} \left(1 - \frac{\varepsilon_{n_k}}{\varepsilon_{n_{k-1}}}\right) \geq \lim_{k\to\infty} \frac{1}{\mu(\Omega_{n_{k-1}} \setminus \Omega_{n_k})} \cdot \frac{1}{2} = +\infty.$$

Таким образом, равенство (2.5) доказано.

Проверим выполнение условий леммы 2.1 для последовательностей

$$a_n = \tilde{g}_n \mu(\Omega_{n-1} \setminus \Omega_n) \quad \text{и} \quad b_n = \tilde{g}^p_n \mu(\Omega_{n-1} \setminus \Omega_n).$$

Выполнение условия 1) очевидно. По определению функции $\tilde{g}(\mathbf{x})$ имеем:

$$\int_\Omega |\tilde{g}(\mathbf{x})| d\mu = \sum_{n=1}^{\infty} \tilde{g}_n \mu(\Omega_{n-1} \setminus \Omega_n) = \sum_{n=1}^{\infty} a_n = +\infty;$$

$$\int_\Omega |\tilde{g}(\mathbf{x})|^p d\mu = \sum_{n=1}^{\infty} \tilde{g}^p_n \mu(\Omega_{n-1} \setminus \Omega_n) = \sum_{n=1}^{\infty} b_n = +\infty.$$

Далее,

$$\overline{\lim_{n\to\infty}}\, b_n = \overline{\lim_{n\to\infty}}\, \tilde{g}^p_n \mu(\Omega_{n-1} \setminus \Omega_n) = \overline{\lim_{n\to\infty}} \left(1 - \frac{\varepsilon_n}{\varepsilon_{n-1}}\right) \leq 1 < +\infty.$$

Наконец, из (2.5) следует, что

$$\lim_{n\to\infty}\frac{a_n}{b_n}=\lim_{n\to\infty}\frac{\tilde g_n\mu(\Omega_{n-1}\setminus\Omega_n)}{\tilde g_n^p\mu(\Omega_{n-1}\setminus\Omega_n)}=\lim_{n\to\infty}\frac{1}{\tilde g_n^{p-1}}=0.$$

Таким образом, получаем, что последовательности

$$\tilde g_n\mu(\Omega_{n-1}\setminus\Omega_n)\quad \text{и}\quad \tilde g_n^p\mu(\Omega_{n-1}\setminus\Omega_n)$$

удовлетворяют условиям леммы 2.1. Следовательно, имеется последовательность номеров n_k, такая что

$$\sum_{k=1}^{\infty}\tilde g_{n_k}\mu(\Omega_{n_k-1}\setminus\Omega_{n_k})<+\infty,\quad \sum_{k=1}^{\infty}\tilde g_{n_k}^p\mu(\Omega_{n_k-1}\setminus\Omega_{n_k})=+\infty.$$

Тогда искомую функцию $g(\mathbf{x})$ строим следующим образом. Полагаем

$$g(\mathbf{x})=\tilde g_{n_k},\quad \mathbf{x}\in\Omega_{n_k-1}\setminus\Omega_{n_k},\quad k\in\mathbb{N},$$

и $g(\mathbf{x})=0$ на всех остальных множествах $\Omega_{n-1}\setminus\Omega_n$. Так как $0\le g(\mathbf{x})\le\tilde g(\mathbf{x})$, то условие 1) не нарушается:

$$\int_\Omega|g(\mathbf{x})|^pw(\mathbf{x})d\mu\le\int_\Omega|\tilde g(\mathbf{x})|^pw(\mathbf{x})d\mu<+\infty.$$

Выполнение условий 2) и 3) проверяется непосредственно:

$$\int_\Omega|g(\mathbf{x})|^pd\mu=\sum_{n=1}^{\infty}\int_{\Omega_{n-1}\setminus\Omega_n}|g(\mathbf{x})|^pd\mu=\sum_{k=1}^{\infty}\int_{\Omega_{n_k-1}\setminus\Omega_{n_k}}\tilde g_{n_k}^pd\mu=$$
$$=\sum_{k=1}^{\infty}\tilde g_{n_k}^p\mu(\Omega_{n_k-1}\setminus\Omega_{n_k})=+\infty;$$

$$\int_\Omega|g(\mathbf{x})|d\mu=\sum_{n=1}^{\infty}\int_{\Omega_{n-1}\setminus\Omega_n}|g(\mathbf{x})|d\mu=\sum_{k=1}^{\infty}\int_{\Omega_{n_k-1}\setminus\Omega_{n_k}}\tilde g_{n_k}d\mu=$$
$$=\sum_{k=1}^{\infty}\tilde g_{n_k}\mu(\Omega_{n_k-1}\setminus\Omega_{n_k})<+\infty.$$

□

Следствие 2.4.1. *Пусть весовая функция $w(\mathbf{x})$ удовлетворяет условию (2.2) и $p>1$. Тогда существует функция плотности вероятности $f(\mathbf{x})$, принадлежащая множеству $L_{p,w}(\Omega)\setminus L_p(\Omega)$.*

Доказательство. Действительно, по доказанной теореме существует функция

$$g \in (L_{p,w}(\Omega) \setminus L_p(\Omega)) \cap L_1(\Omega).$$

Тогда искомую функцию определим следующим образом:

$$f(\mathbf{x}) = \frac{|g(\mathbf{x})|}{\|g(\mathbf{x})\|_{L_1}}.$$

Тогда $f(\mathbf{x})$ является функцией плотности вероятности:

$$f(\mathbf{x}) \geq 0;$$

$$\int_\Omega |f(\mathbf{x})| d\mu = \frac{1}{\|g(\mathbf{x})\|_1} \int_\Omega |g(\mathbf{x})| d\mu = 1.$$

И, кроме того,

$$\int_\Omega |f(\mathbf{x})|^p w(\mathbf{x}) d\mu = \frac{1}{\|g(\mathbf{x})\|_1^p} \int_\Omega |g(\mathbf{x})|^p w(\mathbf{x}) d\mu < +\infty;$$

$$\int_\Omega |f(\mathbf{x})|^p d\mu = \frac{1}{\|g(\mathbf{x})\|_1^p} \int_\Omega |g(\mathbf{x})|^p d\mu = +\infty.$$

□

Более удобный способ построения весового расширения функционального пространства вытекает из теоремы 2.4 с использованием понятия существенного предела измеримой функции.

Следствие 2.4.2. *Пусть $w_1(\mathbf{x})$ и $w_2(\mathbf{x})$ – весовые функции, удовлетворяющие условию*

$$\operatorname{ess}\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{w_1(\mathbf{x})}{w_2(\mathbf{x})} = 0, \tag{2.6}$$

где $\mathbf{x}_0 \in \Omega$. Тогда

$$L_{p,w_1}(\Omega) \setminus L_{p,w_2}(\Omega) \neq \varnothing.$$

Доказательство. Так как из условия (2.6) следует равенство (2.3), то заключение данного предложения сразу следует из следствия 2.4.1. □

Таким образом, для построения расширения $L_{2,w}^k$ пространства L_2^k достаточно взять весовую функцию $w(\mathbf{x})$, для которой

1) $\exists\, \mathbf{a} \in \Omega \quad \operatorname{ess}\lim\limits_{\mathbf{x}\to\mathbf{a}} w(\mathbf{x}) = 0;$

2) $\operatorname*{ess\,sup}_{\mathbf{x}\in\Omega} w(\mathbf{x}) < +\infty$.

При этом за $\mathbf{a}$ следует выбирать точку, для которой

$$\forall\delta > 0 \quad \int_{U_\delta(\mathbf{a})} f^p(\mathbf{x})w(\mathbf{x}) = +\infty, \tag{2.7}$$

где $U_\delta(\mathbf{a})$ – δ-окрестность точки $\mathbf{a}$.

Предположим, что в области определения Ω точка $\mathbf{a}$, обладающая свойством (2.7) единственна, а также восстанавливаемая функция плотности вероятности $f(\mathbf{x})$ такова, что существует конечный существенный предел

$$\operatorname*{ess\,lim}_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) \sum_{j=1}^{k} |x_j - a_j|^{\alpha_j} \in (0; +\infty), \tag{2.8}$$

где $\alpha_1, \ldots, \alpha_k > 0$ – известные числа. Тогда при любом $\delta > 0$ интеграл (2.7) расходится если и только если

$$\int_{U_\delta(\mathbf{a})} \frac{d\mu}{\sum_{j=1}^{k} |x_j - a_j|^{\alpha_j}} = +\infty.$$

При этом последнее равенство эквивалентно условию

$$\frac{1}{\alpha_1} + ... + \frac{1}{\alpha_k} \le 1.$$

В этом случае можно, выбрав некоторую окрестность $U_\delta(\mathbf{a})$ точки $\mathbf{a}$, построить расширение $L_{p,w_0}(\Omega)$ пространства $L_{p,w}(\Omega)$ следующим образом:

$$w_0(\mathbf{x}) = \begin{cases} w(\mathbf{x}) \left(\sum_{j=1}^{k} |x_j - a_j|^{\alpha_j}\right)^p, & x \in U_\delta(\mathbf{a}) \\ w(\mathbf{x}), & x \in \Omega \setminus U_\delta(\mathbf{a}) \end{cases}, \quad \delta > 0. \tag{2.9}$$

Действительно, так как

$$\operatorname*{ess\,lim}_{\mathbf{x}\to\mathbf{a}} \frac{w_0(\mathbf{x})}{w(\mathbf{x})} = \operatorname*{ess\,lim}_{\mathbf{x}\to\mathbf{a}} \left(\sum_{j=1}^{k} |x_j - a_j|^{\alpha_j}\right)^p = 0$$

и

$$\operatorname*{ess\,sup} \frac{w_0(\mathbf{x})}{w(\mathbf{x})} < +\infty,$$

то пространство $L_{p,w_0}(\Omega)$ является расширением пространства $L_{p,w}(\Omega)$. Кроме того, оно содержит $f(\mathbf{x})$:

$$\int_\Omega |f(\mathbf{x})|^p w_0(\mathbf{x}) d\mu =$$

$$= \int_{U_\delta(\mathbf{a})} |f(\mathbf{x})|^p w(\mathbf{x}) \left(\sum_{j=1}^{k} |x_j - a_j|^{\alpha_j} \right)^p d\mu + \int_{\Omega \setminus U_\delta(\mathbf{a})} |f(\mathbf{x})|^p w(\mathbf{x}) d\mu =$$

$$= \int_{U_\delta(\mathbf{a})} w(\mathbf{x}) d\mu + \int_{\Omega \setminus U_\delta(\mathbf{a})} |f(\mathbf{x})|^p w(\mathbf{x}) d\mu < +\infty.$$

Если восстанавливаемая плотность $f(\mathbf{x})$ имеет конечное число особых точек $\mathbf{a}_1, ..., \mathbf{a}_N$, каждая из которых удовлетворяет условию (2.8), то весовая функция строится в виде

$$w_0(\mathbf{x}) = w(\mathbf{x}) \tau^p_{\delta_1}(\mathbf{x}, \mathbf{a}_1) ... \tau^p_{\delta_N}(\mathbf{x}, \mathbf{a}_N),$$

где

$$\tau_\delta(\mathbf{x}, \mathbf{a}) = \begin{cases} \sum_{j=1}^k |x_j - a_j|^{\alpha_j(\mathbf{a})}, & x \in U_\delta(\mathbf{a}) \\ 1, & x \in \Omega \setminus U_\delta(\mathbf{a}) \end{cases},$$

$\alpha_1(\mathbf{a}), ..., \alpha_N(\mathbf{a})$ – показатели, соответствующие данной точке $\mathbf{a}$.

Случай, когда множество особых точек представляет собой линию, поверхность или гиперповерхность в $\mathbb{R}^k$, может быть рассмотрен в том же порядке идей.

Предложенный способ выборка весовой функции $w(\mathbf{x})$ не использует вид истинной плотности вероятности $f(\mathbf{x})$. Требуемая информация о точках $\mathbf{a} \in \Omega$, в которых выполняется условие (2.7), и о показателях α_j из (2.8) может быть получена из выборки, а также из априорных соображений.

§ 2.2. Методы настройки коэффициентов

При настройке коэффициентов проекционной оценки предполагается, что длина ряда уже рассчитана. Проекционную оценку, в которой длина ряда равна l, будем обозначать через $\hat{f}_l(x)$. Методы оценивания длины ряда рассматриваются в следующем параграфе.

Как уже отмечалось (см. § 1.2), большинство работ, посвящённых проекционной оценке плотности вероятности, используют для настройки коэффициентов a_j формулу (1.11), которая в случае весового пространства $L_{2,w}(\Omega)$ принимает вид:

$$a_j = \frac{1}{n} \sum_{i=1}^{n} \eta_j(x_i), \quad j = 0, \ldots, l, \tag{2.10}$$

где

$$\eta_j(x) = \varphi_j(x) w(x).$$

В этом случае оценки a_j также являются несмещёнными оценками оптимальных коэффициентов, т.е. коэффициентов Фурье α_j разложения функции $f(x)$ по базису $\{\varphi_j\}_{j=0}^{\infty}$ в пространстве $L_{2,w}(\Omega)$:

$$M\{a_j\} = M\{\varphi_j(\xi) w(\xi)\} = \int_{\Omega} f(x) \varphi_j(x) w(x) = \alpha_j.$$

Помимо формул (1.11) и (2.10) для настройки коэффициентов a_j можно использовать другие статистические методы.

Для построения выборочных оценок параметров распределения существует ряд общих методов (см., например, [46, гл. 8]). Однако методы, основанные на функции правдоподобия (в том числе, метод максимального правдоподобия, байесовский подход [46, с. 294] и др.) в данном случае оказываются неприменимыми, так как для построения функции правдоподобия

$$L(a_0, \ldots, a_l) = \prod_{i=1}^{n} \hat{f}_l(x_i)$$

необходимо, чтобы при каждом значении параметров $a_0, \ldots, a_l$ функция $\hat{f}_l(x)$ являлась функцией плотности вероятности, т.е. чтобы выполнялись условия (1.2), что приводит к неоправданному усложнению области допустимых значений аргументов a_0, $\ldots$, a_l. С другой стороны, даже при невыполнении для некоторых x условий (1.2) возможно существование оценок $\hat{f}_l$, близких к истинной плотности f в смысле интегрального критерия (1.3).

В настоящей работе для получения выборочных оценок a_i параметров α_i рассматривается метод моментов, а также предлагается его обобщение, частным случаем которого является метод (2.10).

Пусть оцениваемая функция $f(x)$ плотности вероятности непрерывной случайной величины ξ принадлежит некоторому пространству $L_{2,w}(\Omega) \supseteq L_2(\Omega)$. Тогда для любой такой функции определены числа

$$\nu_j = (x^j, f)_w = \int_\Omega x^j f(x) w(x) dx, \quad j = 0, \ldots, l,$$

которые мы будем называть *взвешенными начальными моментами* случайной величины ξ порядка j относительно веса $w(x)$. При $w(x) \equiv 1$ понятие взвешенного начального момента совпадает с понятием начального момента в обычном смысле.

Определим также *выборочные взвешенные начальные моменты* $\hat{\nu}_j$ следующим образом:

$$\hat{\nu}_j = \frac{1}{n} \sum_{i=1}^{n} x_i^j w(x_i), \quad j = 0, \ldots, l.$$

Очевидно, выборочный взвешенный начальный момент $\hat{\nu}_j$ является несмещённой, а при условии конечности дисперсии случайной величины ξ также и состоятельной, оценкой соответствующего теоретического взвешенного начального момента ν_j.

Согласно методу моментов, приравняем выборочные взвешенные моменты к соответствующим теоретическим, найденным с использованием оценки $\hat{f}_l$ в качестве плотности вероятности:

$$(x^j, \hat{f}_l)_w = \hat{\nu}_j, \quad j = 0, \ldots, l.$$

Используя дистрибутивность скалярного произведения, получаем систему линейных уравнений относительно параметров $a_0, \ldots, a_l$:

$$\begin{cases} a_0(1, \varphi_0)_w + \cdots + a_l(1, \varphi_l)_w = 1 \\ a_0(x, \varphi_0)_w + \cdots + a_l(x, \varphi_l)_w = \hat{\nu}_1 \\ \vdots \\ a_0(x^l, \varphi_0)_w + \cdots + a_l(x^l, \varphi_l)_w = \hat{\nu}_l \end{cases}.$$

Введя матричные обозначения:

$$\mathbf{B} = \begin{pmatrix} (1, \varphi_0)_w & \ldots & (1, \varphi_l)_w \\ \vdots & \ddots & \vdots \\ (x^l, \varphi_0)_w & \ldots & (x^l, \varphi_l)_w \end{pmatrix}, \quad \mathbf{a} = \begin{pmatrix} a_0 \\ \vdots \\ a_l \end{pmatrix}, \quad \boldsymbol{\hat{\nu}} = \begin{pmatrix} \hat{\nu}_0 \\ \vdots \\ \hat{\nu}_l \end{pmatrix},$$

получим

$$\mathbf{B}\mathbf{a} = \hat{\boldsymbol{\nu}}.$$

Если определитель $|\mathbf{B}| \neq 0$, то существует единственное решение

$$\mathbf{a} = \mathbf{B}^{-1}\hat{\boldsymbol{\nu}}. \tag{2.11}$$

Нетрудно показать, что если в пространстве $L_2[-1;1]$ в качестве базиса взята система Лежандра, то $|\mathbf{B}| \neq 0$ при любом l.

Предложение 2.5. *Пусть* $\{\varphi\}_{i=0}^{\infty}$ *– ортонормированная система Лежандра. Тогда*

$$\begin{vmatrix} (1,\varphi_0) & \dots & (1,\varphi_l) \\ \vdots & \ddots & \vdots \\ (x^l,\varphi_0) & \dots & (x^l,\varphi_l) \end{vmatrix} \neq 0 \tag{2.12}$$

при любом $l \geq 0$.

Доказательство. Заметим, что матрица

$$\mathbf{B} = \begin{pmatrix} (1,\varphi_0) & \dots & (1,\varphi_l) \\ \vdots & \ddots & \vdots \\ (x^l,\varphi_0) & \dots & (x^l,\varphi_l) \end{pmatrix} \neq 0$$

является нижнетреугольной, т.е. при $j > i$

$$(x^i, \varphi_j) = 0.$$

Действительно, по построению ортонормированной системы Лежандра,

$$x^i = c_0\varphi_0(x) + \dots + c_i\varphi_i(x),$$

где $c_k = (x^i, \varphi_k)$, $k = 0, \dots, i$. Тогда при $j > i$

$$(x^i, \varphi_j) = (c_0\varphi_0 + \dots + c_i\varphi_i, \varphi_j) = c_0(\varphi_0, \varphi_j) + \dots + c_i(\varphi_i, \varphi_j) = 0.$$

Поэтому определитель (2.12) равен произведению элементов главной диагонали:

$$\begin{vmatrix} (1,\varphi_0) & \dots & (1,\varphi_l) \\ \vdots & \ddots & \vdots \\ (x^l,\varphi_0) & \dots & (x^l,\varphi_l) \end{vmatrix} = (1,\varphi_0) \cdot \dots \cdot (x^l, \varphi_l).$$

Докажем, что каждый из множителей в правой части последнего равенства отличен от 0. Пусть, напротив, существует $k \in 0, \ldots, l$, для которого $(x^k, \varphi_k) = 0$. Разложение функции x^k имеет вид:

$$x^k = c_0\varphi_0(x) + \cdots + c_k\varphi_k(x).$$

Но, по предположению, $c_k = (x^k, \varphi_k) = 0$, поэтому

$$x^k = c_0\varphi_0(x) + \cdots + c_{k-1}\varphi_{k-1}(x).$$

Теперь слева от равенства стоит многочлен степени k, а справа – многочлен степени $k-1$. Полученное противоречие доказывает предложение. $\square$

Данный подход к оцениванию коэффициентов a_j допускает следующее обобщение. Возьмём произвольное $l' \geq l$ и рассчитаем матрицу

$$\mathbf{B}_{l'} = \left\{(x^i, \varphi_j)_w\right\}_{i,j=0}^{l'} = \begin{pmatrix} (1, \varphi_0)_w & \ldots & (1, \varphi_{l'})_w \\ \vdots & \ddots & \vdots \\ (x^{l'}, \varphi_0)_w & \ldots & (x^{l'}, \varphi_{l'})_w \end{pmatrix}$$

размерности $(l'+1) \times (l'+1)$. Обратную к ней матрицу обозначим через $\mathbf{B}_{l'}^{-1}$. Выделив в ней подматрицу $\mathbf{B}_{l,l'}^{-1}$, содержащую первые $(l+1)$ её строк, и умножив её на вектор из $(l'+1)$ выборочных взвешенных начальных моментов $\hat{\boldsymbol{\nu}}_{l'}$, получим вектор

$$\mathbf{a}^{(l')} = \mathbf{B}_{l,l'}^{-1}\hat{\boldsymbol{\nu}}_{l'}, \tag{2.13}$$

значения которого наряду со значениями (2.10) и (2.11) могут использоваться в качестве оценок коэффициентов α_j для (1.8). Соответствующую проекционную оценку функции плотности вероятности обозначим следующим образом:

$$\hat{f}_l^{(l')} = a_0^{(l')}\varphi_0(x) + \cdots + a_l^{(l')}\varphi_l(x). \tag{2.14}$$

Очевидно, что формула (2.11) является частным случаем формулы (2.13) при $l' = l$. Следовательно, оценка плотности вероятности, построенная с помощью коэффициентов (2.11) является частным случаем оценки (2.14). С другой стороны, справедлива следующая теорема.

Теорема 2.5. *Пусть ξ – непрерывная случайная величина, $f(x)$ – её функция плотности вероятности, $f \in L_{2,w}(\Omega)$, где $\operatorname{supp}\xi \subseteq \Omega$, $\{\varphi_j\}_{j=0}^{\infty}$ – базис про-*

странства $L_{2,w}(\Omega)$, удовлетворяющий для любых $x \in \Omega$ следующему условию:

$$\lim_{l\to\infty}\left(\|\mathbf{B}_l^{-1}\|_\infty \max_{0\le k\le l}\left|x^k - \sum_{j=0}^{l}(x^k,\varphi_j)_w\varphi_j(x)\right|\right) = 0, \tag{2.15}$$

где $\|\cdot\|_\infty$ – матричная норма, индуцированная векторной l_∞-нормой [141]*:*

$$\|A\|_\infty = \max_i \sum_j |a_{ij}|.$$

Тогда при неограниченном увеличении l' и фиксированном l оценка (2.14) *сходится к оценке Ченцова по норме пространства $L_{2,w}(\Omega)$ почти наверное:*

$$P\left\{\lim_{l'\to\infty}\left\|\hat{f}_l^{(l')} - \hat{f}_l\right\|_w = 0\right\} = 1.$$

Для доказательства этой теоремы докажем следующую лемму.

Лемма 2.2. *Из условия теоремы* 2.5 *следует, что последовательность случайных величин*

$$\|\mathbf{B}_l^{-1}\hat{\boldsymbol{\nu}} - \hat{\boldsymbol{\nu}}_\eta\|_\infty, \|\mathbf{B}_{l+1}^{-1}\hat{\boldsymbol{\nu}} - \hat{\boldsymbol{\nu}}_\eta\|_\infty, \ldots$$

где

$$\hat{\boldsymbol{\nu}}_\eta = \left(\frac{1}{n}\sum_{i=1}^{n}\eta_0(x_i) \quad \ldots \frac{1}{n}\sum_{i=1}^{n}\eta_l(x_i)\right)^T,$$

сходится к 0 почти наверное:

$$P\left\{\lim_{l'\to\infty}\|\mathbf{B}_{l'}^{-1}\hat{\boldsymbol{\nu}} - \hat{\boldsymbol{\nu}}_\eta\|_\infty = 0\right\} = 1.$$

Доказательство. Заметим, что

$$\left\|\mathbf{B}_{l'}^{-1}\hat{\boldsymbol{\nu}} - \hat{\boldsymbol{\nu}}_\eta\right\|_\infty = \left\|\mathbf{B}_{l'}^{-1}(\hat{\boldsymbol{\nu}} - \mathbf{B}_{l'}\hat{\boldsymbol{\nu}}_\eta)\right\|_\infty \le \left\|\mathbf{B}_{l'}^{-1}\right\|_\infty \|\hat{\boldsymbol{\nu}} - \mathbf{B}_{l'}\hat{\boldsymbol{\nu}}_\eta\|_\infty.$$

В свою очередь,

$$\hat{\boldsymbol{\nu}} - \mathbf{B}_{l'}\hat{\boldsymbol{\nu}}_\eta = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} x_i^0 \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} x_i^{l'} \end{pmatrix} - \begin{pmatrix} (1,\varphi_0)_w & \ldots & (1,\varphi_{l'})_w \\ \vdots & \ddots & \vdots \\ (x^{l'},\varphi_0)_w & \ldots & (x^{l'},\varphi_{l'})_w \end{pmatrix} \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n}\varphi_0(x_i)w(x_i) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n}\varphi_{l'}(x_i)w(x_i) \end{pmatrix} =$$

$$= \begin{pmatrix} \dfrac{1}{n}\sum\limits_{i=1}^{n} w(x_i)\left(1 - \sum\limits_{j=0}^{l'}(1,\varphi_j)_w\varphi_j(x_i)\right) \\ \dfrac{1}{n}\sum\limits_{i=1}^{n} w(x_i)\left(x_i - \sum\limits_{j=0}^{l'}(x,\varphi_j)_w\varphi_j(x_i)\right) \\ \vdots \\ \dfrac{1}{n}\sum\limits_{i=1}^{n} w(x_i)\left(x_i^{l'} - \sum\limits_{j=0}^{l'}(x^{l'},\varphi_j)_w\varphi_j(x_i)\right) \end{pmatrix};$$

$$\|\hat{\boldsymbol{\nu}} - \mathbf{B}_{l'}\hat{\boldsymbol{\nu}}_\eta\|_\infty = \max_{0\le k\le l'}\left|\frac{1}{n}\sum_{i=1}^{n} w(x_i)\left(x_i^k - \sum_{j=0}^{l'}(x^k,\varphi_j)_w\varphi_j(x_i)\right)\right| \le$$

$$\le \frac{1}{n}\sum_{i=1}^{n}\left(|w(x_i)|\max_{0\le k\le l'}\left|x_i^k - \sum_{j=0}^{l'}(x^k,\varphi_j)_w\varphi_j(x_i)\right|\right) \le$$

$$\le \max_{1\le i\le n}\left(|w(x_i)|\max_{0\le k\le l'}\left|x_i^k - \sum_{j=0}^{l'}(x^k,\varphi_j)_w\varphi_j(x_i)\right|\right) \le$$

$$\le \max_{1\le i\le n}|w(x_i)| \cdot \max_{1\le i\le n}\max_{0\le k\le l'}\left|x_i^k - \sum_{j=0}^{l'}(x^k,\varphi_j)_w\varphi_j(x_i)\right|.$$

Тогда

$$\|\mathbf{B}_{l'}^{-1}\hat{\boldsymbol{\nu}} - \hat{\boldsymbol{\nu}}_\eta\|_\infty \le \|\mathbf{B}_{l'}^{-1}\|_\infty \max_{1\le i\le n}|w(x_i)| \cdot \max_{1\le i\le n}\max_{0\le k\le l'}\left|x_i^k - \sum_{j=0}^{l'}(x^k,\varphi_j)_w\varphi_j(x_i)\right|.$$

Переходя к пределу при $l' \to \infty$, получим

$$\lim_{l'\to\infty}\|\mathbf{B}_{l'}^{-1}\hat{\boldsymbol{\nu}} - \hat{\boldsymbol{\nu}}_\eta\|_\infty \le$$

$$\le \max_{1\le i\le n}|w(x_i)|\lim_{l'\to\infty}\left(\|\mathbf{B}_{l'}^{-1}\|_\infty \max_{1\le i\le n}\max_{0\le k\le l'}\left|x_i^k - \sum_{j=0}^{l'}(x^k,\varphi_j)_w\varphi_j(x_i)\right|\right).$$

Таким образом, равенство

$$\max_{1\le i\le n}|w(x_i)|\lim_{l'\to\infty}\left(\|\mathbf{B}_{l'}^{-1}\|_\infty \max_{1\le i\le n}\max_{0\le k\le l'}\left|x_i^k - \sum_{j=0}^{l'}(x^k,\varphi_j)_w\varphi_j(x_i)\right|\right) = 0 \qquad (2.16)$$

влечёт равенство

$$\lim_{l'\to\infty}\|\mathbf{B}_{l'}^{-1}\hat{\boldsymbol{\nu}} - \hat{\boldsymbol{\nu}}_\eta\|_\infty = 0,$$

откуда

$$P\left\{\lim_{l'\to\infty}\|\mathbf{B}_{l'}^{-1}\hat{\boldsymbol{\nu}}-\hat{\boldsymbol{\nu}}_\eta\|_\infty=0\right\}\geq$$

$$\geq P\left\{\max_{1\leq i\leq n}|w(x_i)|\lim_{l'\to\infty}\left(\|\mathbf{B}_{l'}^{-1}\|_\infty\max_{1\leq i\leq n}\max_{0\leq k\leq l'}\left|x_i^k-\sum_{j=0}^{l'}(x^k,\varphi_j)_w\varphi_j(x_i)\right|\right)=0\right\}$$

Определим случайные события A и B следующим образом:

$$A\Leftrightarrow\max_{1\leq i\leq n}|w(x_i)|=0;$$

$$B\Leftrightarrow\lim_{l'\to\infty}\left(\|\mathbf{B}_{l'}^{-1}\|_\infty\max_{1\leq i\leq n}\max_{0\leq k\leq l'}\left|x_i^k-\sum_{j=0}^{l'}(x^k,\varphi_j)_w\right|\right)=0.$$

Тогда случайное событие (2.16) представляет собой сумму случайных событий A и B. Так как весовая функция $w(x)$ по определению положительна почти всюду, то $P\{A\}=0$ и

$$P\{A+B\}=P\{A\}+P\{B\}-P\{AB\}=P\{B\}.$$

Тогда

$$P\left\{\max_{1\leq i\leq n}|w(x_i)|\lim_{l'\to\infty}\left(\|\mathbf{B}_{l'}^{-1}\|_\infty\max_{1\leq i\leq n}\max_{0\leq k\leq l'}\left|x_i^k-\sum_{j=0}^{l'}(x^k,\varphi_j)_w\varphi_j(x_i)\right|\right)=0\right\}=$$

$$=P\left\{\lim_{l'\to\infty}\left(\|\mathbf{B}_{l'}^{-1}\|_\infty\max_{1\leq i\leq n}\max_{0\leq k\leq l'}\left|x_i^k-\sum_{j=0}^{l'}(x^k,\varphi_j)_w\varphi_j(x_i)\right|\right)=0\right\}.$$

Так как для любых $i\in\{1,\ldots,n\}$ выполняются неравенства

$$0\leq\max_{0\leq k\leq l'}\left|x_i^k-\sum_{j=0}^{l'}(x^k,\varphi_j)_w\varphi_j(x_i)\right|\leq\max_{1\leq i\leq n}\max_{0\leq k\leq l'}\left|x_i^k-\sum_{j=0}^{l'}(x^k,\varphi_j)_w\varphi_j(x_i)\right|,$$

то из

$$\lim_{l'\to\infty}\left(\|\mathbf{B}_{l'}^{-1}\|_\infty\max_{1\leq i\leq n}\max_{0\leq k\leq l'}\left|x_i^k-\sum_{j=0}^{l'}(x^k,\varphi_j)_w\right|\right)=0$$

следует

$$\forall i\in\{1,\ldots,n\}\quad\lim_{l'\to\infty}\left(\|\mathbf{B}_{l'}^{-1}\|_\infty\max_{0\leq k\leq l'}\left|x_i^k-\sum_{j=0}^{l'}(x^k,\varphi_j)_w\varphi_j(x_i)\right|\right)=0.$$

Тогда

$$P\left\{\lim_{l'\to\infty}\left(\|\mathbf{B}_{l'}^{-1}\|_\infty \max_{1\le i\le n}\max_{0\le k\le l'}\left|x_i^k-\sum_{j=0}^{l'}(x^k,\varphi_j)_w\varphi_j(x_i)\right|\right)=0\right\}\ge$$

$$\ge P\left\{\bigwedge_{i=1}^{n}\lim_{l'\to\infty}\left(\|\mathbf{B}_{l'}^{-1}\|_\infty \max_{0\le k\le l'}\left|x_i^k-\sum_{j=0}^{l'}(x^k,\varphi_j)_w\varphi_j(x_i)\right|\right)=0\right\}$$

Так как $x_1,\dots,x_n$ – независимая выборка случайной величины ξ, то случайные события под знаком вероятности в правой части полученного неравенства также независимы, и

$$P\left\{\bigwedge_{i=1}^{n}\lim_{l'\to\infty}\left(\|\mathbf{B}_{l'}^{-1}\|_\infty \max_{0\le k\le l'}\left|x_i^k-\sum_{j=0}^{l'}(x^k,\varphi_j)_w\varphi_j(x_i)\right|\right)=0\right\}=$$

$$=P^n\left\{\lim_{l'\to\infty}\left(\|\mathbf{B}_{l'}^{-1}\|_\infty \max_{0\le k\le l'}\left|\xi^k-\sum_{j=0}^{l'}(x^k,\varphi_j)_w\varphi_j(\xi)\right|\right)=0\right\}.$$

Тогда получаем, что

$$P\left\{\lim_{l'\to\infty}\|\mathbf{B}_{l'}^{-1}\hat{\boldsymbol{\nu}}-\hat{\boldsymbol{\nu}}_\eta\|_\infty=0\right\}\ge$$

$$\ge P^n\left\{\lim_{l'\to\infty}\left(\|\mathbf{B}_{l'}^{-1}\|_\infty \max_{0\le k\le l'}\left|\xi^k-\sum_{j=0}^{l'}(x^k,\varphi_j)_w\varphi_j(\xi)\right|\right)=0\right\}.$$

Из условия (2.15) следует, что вероятность в правой части полученного неравенства равна 1. Отсюда получаем окончательно

$$P\left\{\lim_{l'\to\infty}\|\mathbf{B}_{l'}^{-1}\hat{\boldsymbol{\nu}}-\hat{\boldsymbol{\nu}}_\eta\|_\infty=0\right\}\ge 1,$$

т.е.

$$P\left\{\lim_{l'\to\infty}\|\mathbf{B}_{l'}^{-1}\hat{\boldsymbol{\nu}}-\hat{\boldsymbol{\nu}}_\eta\|_\infty=0\right\}=1.$$

□

Доказательство теоремы 2.5. Оценим величину $\left\|\hat f_l^{(l')}-\hat f_l\right\|_w$:

$$\left\|\hat f_l^{(l')}-\hat f_l\right\|_w=\left\|\sum_{j=0}^{l}\left(a_j^{(l')}-a_j\right)\varphi_j\right\|_w\le\sum_{j=0}^{l}\left|a_j^{(l')}-a_j\right|\le(l+1)\left\|a^{(l')}-a\right\|_\infty$$

и перейдём к пределу:

$$\lim_{l'\to\infty}\left\|\hat{f}_l^{(l')}-\hat{f}_l\right\|_w \le (l+1)\lim_{l'\to\infty}\left\|a^{(l')}-a\right\|_\infty .$$

Тогда

$$P\left\{\lim_{l'\to\infty}\left\|\hat{f}_l^{(l')}-\hat{f}_l\right\|_w=0\right\}\ge P\left\{\lim_{l'\to\infty}\left\|a^{(l')}-a\right\|_\infty=0\right\}.$$

Так как $a^{(l')}$ и a – столбцы, содержащие l первых элементов столбцов $\mathbf{B}_{l'}^{-1}\hat{\boldsymbol{\nu}}$ и $\hat{\boldsymbol{\nu}}_\eta$ соответственно, то

$$\left\|a^{(l')}-a\right\|_\infty \le \left\|\mathbf{B}_{l'}^{-1}\hat{\boldsymbol{\nu}}-\hat{\boldsymbol{\nu}}_\eta\right\|_\infty;$$

$$P\left\{\lim_{l'\to\infty}\left\|\hat{f}_l^{(l')}-\hat{f}_l\right\|_w=0\right\}\ge P\left\{\left\|\mathbf{B}_{l'}^{-1}\hat{\boldsymbol{\nu}}-\hat{\boldsymbol{\nu}}_\eta\right\|_\infty=0.\right\}$$

По лемме 2.2 получаем

$$P\left\{\lim_{l'\to\infty}\left\|\hat{f}_l^{(l')}-\hat{f}_l\right\|_w=0\right\}=1.$$

□

Из доказанного следует, что при выполнении условия (2.15) оценка Ченцова является частным случаем оценки (2.14) при $l'\to\infty$. В связи с этим для оценки Ченцова вводится следующее обозначение: $\hat{f}_l^{(\infty)}(x)$. Оценка, полученная с помощью классического метода моментов, является другим предельным частным случаем оценки (2.14) при $l'=l$. Ниже (§ 2.3) будет показано, что обобщённый метод моментов при $l<l'<\infty$ иногда оказывается эффективнее обоих указанных предельных случаев.

Можно показать, что для систем Лежандра, Чебышёва и тригонометрической условие (2.15) выполняется для всех $x\in(-1;1)$. Причём для системы Лежандра выражение под пределом равно 0 при любом $l\ge 0$.

Действительно, если $\{\varphi_j\}_{j=0}^\infty$ – ортонормальная система Лежандра, то функция $\varphi_j(x)$ является многочленом степени j, и

$$x^k=\sum_{j=0}^k (x^k,\varphi_j)\varphi_j(x),\quad k=0,1,2,\dots;$$

$$\|\mathbf{B}_l^{-1}\|_\infty \max_{0\le k\le l}\left|x^k - \sum_{j=0}^{l}(x^k,\varphi_j)\varphi_j(x)\right| =$$

$$= \|\mathbf{B}_l^{-1}\|_\infty \max_{0\le k\le l}\left|x^k - \sum_{j=0}^{k}(x^k,\varphi_j)\varphi_j(x)\right| = \|\mathbf{B}_l^{-1}\|_\infty \max_{0\le k\le l}\left|x^k - x^k\right| = 0.$$

Более того, так как для системы Лежандра матрица $\mathbf{B}_{l'}$ является верхнетреугольной, то её можно представить в виде

$$\mathbf{B}_{l'} = \begin{pmatrix} \mathbf{B}_l & \mathbf{0} \\ \mathbf{G} & \mathbf{H} \end{pmatrix}.$$

Тогда

$$\mathbf{B}_{l'}^{-1} = \begin{pmatrix} \mathbf{B}_l^{-1} & \mathbf{0} \\ -\mathbf{H}^{-1}\mathbf{G}\mathbf{B}_l^{-1} & \mathbf{H}^{-1} \end{pmatrix}.$$

Действительно,

$$\begin{pmatrix} \mathbf{B}_l^{-1} & \mathbf{0} \\ -\mathbf{H}^{-1}\mathbf{G}\mathbf{B}_l^{-1} & \mathbf{H}^{-1} \end{pmatrix}\begin{pmatrix} \mathbf{B}_l & \mathbf{0} \\ \mathbf{G} & \mathbf{H} \end{pmatrix} = \begin{pmatrix} \mathbf{E} & \mathbf{0} \\ \mathbf{0} & \mathbf{E} \end{pmatrix} = \mathbf{E}.$$

Отсюда следует, что

$$\mathbf{B}_{l,l'} = \begin{pmatrix} \mathbf{B}_l^{-1} & \mathbf{0} \end{pmatrix};$$

$$\mathbf{a}^{(l')} = \mathbf{B}_{l,l'}^{-1}\hat{\boldsymbol{\nu}}_{l'} = \mathbf{B}_l^{-1}\hat{\boldsymbol{\nu}}_l = \mathbf{a}^{(l)}.$$

С другой стороны, из доказательства предложения 2.5 следует, что

$$\mathbf{B}_l\mathbf{a}^{(\infty)} = \begin{pmatrix} (x^0,\varphi_0) & 0 & \dots & 0 \\ (x^1,\varphi_0) & (x^1,\varphi_1) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ (x^l,\varphi_0) & (x^l,\varphi_1) & \dots & (x^l,\varphi_l) \end{pmatrix}\begin{pmatrix} \dfrac{1}{n}\sum\limits_{i=1}^{n}\varphi_0(x_i) \\ \dfrac{1}{n}\sum\limits_{i=1}^{n}\varphi_1(x_i) \\ \vdots \\ \dfrac{1}{n}\sum\limits_{i=1}^{n}\varphi_l(x_i) \end{pmatrix} =$$

$$
= \begin{pmatrix} \dfrac{1}{n}\sum\limits_{i=1}^{n}(x^0,\varphi_0)\varphi_0(x_i) \\ \sum\limits_{j=0}^{1}\left(\dfrac{1}{n}\sum\limits_{i=1}^{n}(x^1,\varphi_j)\varphi_j(x_i)\right) \\ \vdots \\ \sum\limits_{j=0}^{l}\left(\dfrac{1}{n}\sum\limits_{i=1}^{n}(x^l,\varphi_j)\varphi_j(x_i)\right) \end{pmatrix} =
$$

$$
= \begin{pmatrix} \dfrac{1}{n}\sum\limits_{i=1}^{n}(x^0,\varphi_0)\varphi_0(x_i) \\ \dfrac{1}{n}\sum\limits_{i=1}^{n}\sum\limits_{j=0}^{1}(x^1,\varphi_j)\varphi_j(x_i) \\ \vdots \\ \dfrac{1}{n}\sum\limits_{i=1}^{n}\sum\limits_{j=0}^{l}(x^l,\varphi_j)\varphi_j(x_i) \end{pmatrix} = \begin{pmatrix} \dfrac{1}{n}\sum\limits_{i=1}^{n}x_i^0 \\ \dfrac{1}{n}\sum\limits_{i=1}^{n}x_i^1 \\ \vdots \\ \dfrac{1}{n}\sum\limits_{i=1}^{n}x_i^l \end{pmatrix} = \hat{\boldsymbol{\nu}}_l.
$$

Тогда

$$
\mathbf{a}^{(\infty)} = \mathbf{B}_l^{-1}\hat{\boldsymbol{\nu}}_l = \mathbf{a}^{(l)} = \mathbf{a}^{(l')}. \tag{2.17}
$$

Таким образом, *при использовании системы Лежандра для данной длины ряда l оценка Ченцова и оценка, полученная обобщённым методом моментов, тождественны*:

$$
\hat{f}_l^{(l)}(x) \equiv \hat{f}_l^{(l')}(x) \equiv \hat{f}_l^{(\infty)}(x).
$$

Рассмотрим некоторые свойства оценок (2.13). Найдём математические ожидания:

$$
M\left\{\mathbf{a}^{(l')}\right\} = M\left\{\mathbf{B}_{l,l'}^{-1}\hat{\boldsymbol{\nu}}\right\} = \mathbf{B}_{l,l'}^{-1}M\left\{\hat{\boldsymbol{\nu}}\right\},
$$

причём

$$
M\left\{\hat{\boldsymbol{\nu}}\right\} = \begin{pmatrix} M\left\{\hat{\boldsymbol{\nu}}_0\right\} \\ \vdots \\ M\left\{\hat{\boldsymbol{\nu}}_{l'}\right\} \end{pmatrix} = \begin{pmatrix} M\left\{\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i^0 w(x_i)\right\} \\ \vdots \\ M\left\{\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i^{l'} w(x_i)\right\} \end{pmatrix} =
$$

$$= \begin{pmatrix} M\left\{\xi^0 w(\xi)\right\} \\ \vdots \\ M\left\{\xi^{l'} w(\xi)\right\} \end{pmatrix} = \begin{pmatrix} \nu_0 \\ \vdots \\ \nu_l \end{pmatrix} = \boldsymbol{\nu}$$

Тогда

$$M\left\{a^{(l')}\right\} = \mathbf{B}^{-1}_{l,l'}\boldsymbol{\nu}, \quad l' < \infty. \tag{2.18}$$

При $l' = \infty$, оценки, как известно, не смещены:

$$M\left\{a_j^{(\infty)}\right\} = M\left\{\frac{1}{n}\sum_{i=1}^{n}\varphi_j(x_i)w(x_i)\right\} = (\varphi_j, f)_w = \alpha_j, \quad j = 0, \ldots, l.$$

В общем случае оценка (2.13) смещена, что легко показать следующим примером. Пусть ξ равномерно распределена на отрезке $[0;1]$:

$$f(x) = \begin{cases} 1, & x \in [0;1] \\ 0, & \text{иначе} \end{cases}.$$

Тогда уже при $l' = l = 2$ для ортонормальной системы Чебышёва:

$$\boldsymbol{\alpha} \approx \begin{pmatrix} 0.676 \\ 0.532 \\ -0.191 \end{pmatrix}, \quad \mathbf{B} \approx \begin{pmatrix} 1.352 & 0 & -0.382 \\ 0 & 0.765 & 0 \\ 0.541 & 0 & 0.255 \end{pmatrix};$$

для тригонометрической системы:

$$\boldsymbol{\alpha} \approx \begin{pmatrix} 0.399 \\ 0.259 \\ 0.474 \end{pmatrix}, \quad \mathbf{B} \approx \begin{pmatrix} 2.507 & 0 & 0 \\ 0 & 3.545 & 0 \\ 8.246 & 0 & -7.09 \end{pmatrix}.$$

В обоих случаях равенство

$$M\{\mathbf{a}\} = \boldsymbol{\alpha}$$

не выполняется:

$$M\{\mathbf{a}_{\text{Чеб}}\} \approx \begin{pmatrix} 1.352 & 0 & -0.382 \\ 0 & 0.765 & 0 \\ 0.541 & 0 & 0.255 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 1/2 \\ 1/3 \end{pmatrix} \approx \begin{pmatrix} 0.693 \\ 0.654 \\ -0.163 \end{pmatrix}$$

и

$$M\{\mathbf{a}_{\text{триг}}\} \approx \begin{pmatrix} 2.507 & 0 & 0 \\ 0 & 3.545 & 0 \\ 8.246 & 0 & -7.09 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 1/2 \\ 1/3 \end{pmatrix} \approx \begin{pmatrix} 0.399 \\ 0.141 \\ 0.417 \end{pmatrix}$$

соответственно.

Рассчитаем ковариационную матрицу

$$\Sigma\left\{\mathbf{a}^{(l')}\right\} = \begin{pmatrix} \operatorname{cov}\left\{a_0^{(l')}, a_0^{(l')}\right\} & \dots & \operatorname{cov}\left\{a_0^{(l')}, a_l^{(l')}\right\} \\ \vdots & \ddots & \vdots \\ \operatorname{cov}\left\{a_l^{(l')}, a_0^{(l')}\right\} & \dots & \operatorname{cov}\left\{a_l^{(l')}, a_l^{(l')}\right\} \end{pmatrix}$$

оценок (2.13):

$$\Sigma\left\{\mathbf{a}^{(l')}\right\} = M\left\{\mathbf{a}\cdot\left(\mathbf{a}^{(l')}\right)^T\right\} - M\{\mathbf{a}\}M\left\{\left(\mathbf{a}^{(l')}\right)^T\right\} =$$
$$= M\left\{\mathbf{B}_{l,l'}^{-1}\hat{\boldsymbol{\nu}}\hat{\boldsymbol{\nu}}^T\left(\mathbf{B}_{l,l'}^{-1}\right)^T\right\} - M\left\{\mathbf{B}_{l,l'}^{-1}\hat{\boldsymbol{\nu}}\right\}M\left\{\hat{\boldsymbol{\nu}}^T\left(\mathbf{B}_{l,l'}^{-1}\right)^T\right\} =$$
$$= \mathbf{B}_{l,l'}^{-1}M\left\{\hat{\boldsymbol{\nu}}\hat{\boldsymbol{\nu}}^T\right\}\left(\mathbf{B}_{l,l'}^{-1}\right)^T - \mathbf{B}_{l,l'}^{-1}M\left\{\hat{\boldsymbol{\nu}}\right\}M\left\{\hat{\boldsymbol{\nu}}^T\right\}\left(\mathbf{B}_{l,l'}^{-1}\right)^T =$$
$$= \mathbf{B}_{l,l'}^{-1}\left(M\left\{\hat{\boldsymbol{\nu}}\hat{\boldsymbol{\nu}}^T\right\} - M\left\{\hat{\boldsymbol{\nu}}\right\}M\left\{\hat{\boldsymbol{\nu}}^T\right\}\right)\left(\mathbf{B}_{l,l'}^{-1}\right)^T = \mathbf{B}_{l,l'}^{-1}\Sigma\left\{\hat{\boldsymbol{\nu}}\right\}\left(\mathbf{B}_{l,l'}^{-1}\right)^T.$$

При этом

$$\Sigma\left\{\hat{\boldsymbol{\nu}}\right\} = \begin{pmatrix} \operatorname{cov}\left\{\hat{\nu}_0, \hat{\nu}_0\right\} & \dots & \operatorname{cov}\left\{\hat{\nu}_0, \hat{\nu}_{l'}\right\} \\ \vdots & \ddots & \vdots \\ \operatorname{cov}\left\{\hat{\nu}_{l'}, \hat{\nu}_0\right\} & \dots & \operatorname{cov}\left\{\hat{\nu}_{l'}, \hat{\nu}_{l'}\right\} \end{pmatrix};$$

$$\operatorname{cov}\left\{\hat{\nu}_{l'}, \hat{\nu}_{l'}\right\} = \frac{1}{n}\operatorname{cov}\left\{\xi^{j_1}w(\xi), \xi^{j_2}w(\xi)\right\}, \quad j_1, j_2 = 0, \dots, l'.$$

Теперь, введя обозначение

$$\boldsymbol{\Xi} = \begin{pmatrix} \xi^0 w(\xi) & \xi^1 w(\xi) & \dots & \xi^{l'} w(\xi) \end{pmatrix}^T,$$

получим

$$\Sigma\left\{\hat{\boldsymbol{\nu}}\right\} = \frac{1}{n}\Sigma\{\boldsymbol{\Xi}\};$$

$$\Sigma\left\{\mathbf{a}^{(l')}\right\} = \frac{1}{n}\mathbf{B}_{l,l'}^{-1}\Sigma\{\boldsymbol{\Xi}\}\left(\mathbf{B}_{l,l'}^{-1}\right)^T. \tag{2.19}$$

В случае оценки Ченцова получаем:

$$\Sigma\left\{\mathbf{a}^{(\infty)}\right\} = \frac{1}{n}\Sigma\{\mathbf{H}\},$$

где

$$\mathbf{H} = \begin{pmatrix} \eta_0(\xi) & \eta_1(\xi) & \dots & \eta_l(\xi) \end{pmatrix}^T.$$

Отсюда видно, что если случайная величина ξ имеет ограниченные начальные моменты всех порядков до $2l'$ включительно, то при неограниченном увеличении объёма выборки n дисперсии оценок (2.13) стремятся к 0.

Сравнение метода Ченцова и обобщённого метода моментов показало, что в некоторых случаях проекционная оценка, в которой коэффициенты рассчитываются по формуле (2.13), оказывается ближе к истинной плотности $f(x)$, чем оценка Ченцова.

Для сравнения была взята мера близости в виде функционала

$$Q_2^2\{\hat{f}\} = M\left\{\left\|\hat{f} - f\right\|_{2,w}^2\right\}. \tag{2.20}$$

Для удобства вычислений преобразуем функционал (2.20) следующим образом:

$$M\left\{\left\|\hat{f} - f\right\|_{2,w}^2\right\} = \sum_{j=0}^{l} M\left\{a_j^2\right\} - 2\sum_{j=0}^{l} \alpha_j M\left\{a_j\right\} + \|f\|_{2,w}^2.$$

Если $\hat{f}_l(x) \equiv \hat{f}_l^{(\infty)}(x)$ – оценка Ченцова, то

$$M\left\{a_j^{(\infty)}\right\} = \alpha_j, \quad M\left\{\left(a_j^{(\infty)}\right)^2\right\} = \frac{1}{n} M\left\{\eta_j^2(\xi)\right\} + \frac{n-1}{n}\alpha_j^2,$$

и

$$Q\left\{\hat{f}_l^{(\infty)}\right\} = \frac{1}{n}\sum_{j=0}^{l} M\left\{\eta_j^2(\xi)\right\} - \frac{n+1}{n}\sum_{j=0}^{l}\alpha_j^2 + \|f\|_{2,w}^2.$$

Если $\hat{f}_l(x) \equiv \hat{f}_l^{(l')}(x)$ – оценка, построенная по обобщённому методу моментов, то из формул (2.18) и (2.19) получаем:

$$\sum_{j=0}^{l} \alpha_j M\left\{a_j^{(l')}\right\} = \boldsymbol{\alpha}^T \mathbf{B}_{l,l'}^{-1} \boldsymbol{\nu},$$

$$\sum_{j=0}^{l} M\left\{\left(a_j^{(l')}\right)^2\right\} = \operatorname{tr} M\left\{\mathbf{a}^{(l')}\left(\mathbf{a}^{(l')}\right)^T\right\} =$$
$$= \operatorname{tr} \mathbf{B}_{l,l'}^{-1}\left(\frac{1}{n} M\{\boldsymbol{\Xi}\boldsymbol{\Xi}^T\} + \frac{n-1}{n}\boldsymbol{\nu}\boldsymbol{\nu}^T\right)(\mathbf{B}_{l,l'}^{-1})^T,$$

и

$$M\left\{\left\|\hat{f}-f\right\|_{2,w}^{2}\right\}=$$
$$=\operatorname{tr}\mathbf{B}_{l,l'}^{-1}\left(\frac{1}{n}M\{\boldsymbol{\Xi}\boldsymbol{\Xi}^T\}+\frac{n-1}{n}\boldsymbol{\nu}\boldsymbol{\nu}^T\right)(\mathbf{B}_{l,l'}^{-1})^T-2\boldsymbol{\alpha}^T\mathbf{B}_{l,l'}^{-1}\boldsymbol{\nu}+\|f\|_{2,w}^2.$$

Рассмотрим пример. Пусть случайная величина ξ распределена равномерно на отрезке $[0;1]$:

$$f(x)=\begin{cases}1, & x\in[0;1]\\ 0, & \text{иначе}\end{cases}.$$

Построим проекционную оценку функции плотности вероятности этой случайной величины в пространстве $L_2[0;\pi]$, используя в качестве базиса систему

$$\varphi_j(x)=\sqrt{\frac{2}{\pi}}\sin jx,\quad j=1,2,\ldots$$

В этом случае для оценки Ченцова функционал (2.20) имеет вид:

$$Q\left\{\hat{f}_l^{(\infty)}\right\}=\frac{1}{\pi n}\sum_{j=1}^{l}\frac{j-\sin j\cos j}{j}-\frac{n+1}{n}\cdot\frac{2}{\pi}\sum_{k=1}^{l}\left(\frac{1-\cos j}{j}\right)^2+1.$$

Значения функционала (2.20) для оценки Ченцова $\hat{f}_l^{(\infty)}$ и наилучшей из оценок $\hat{f}_l^{(l')}(x)$, полученных по обобщённому моменту моментов при объёме выборки $n=10$ и параметрах $l=1,\ldots,10$, $l'=l,\ldots,40$, занесены в табл. 2.5.

Таблица 2.5. Сравнение методов настройки коэффициентов

l	1	2	3	4	5
$Q_2^2\left\{\hat{f}_l^{(\infty)}\right\}$	0.8694	0.5561	0.2813	0.1895	0.2087
$\min\limits_{l\le l'\le 40} Q_2^2\left\{\hat{f}_l^{(l')}\right\}$	0.8693	0.5553	0.2813	0.1869	0.2087
l	6	7	8	9	10
$Q_2^2\left\{\hat{f}_l^{(\infty)}\right\}$	0.2419	0.2706	0.2887	0.2903	0.2970
$\min\limits_{l\le l'\le 40} Q_2^2\left\{\hat{f}_l^{(l')}\right\}$	0.2419	0.2706	0.2887	0.2902	0.2970

Как показали расчёты, оценка плотности вероятности, в которой коэффициенты рассчитаны по формуле (2.13) оказывается близкой по качеству или лучшей, чем оценка Ченцова.

§ 2.3. Методы настройки длины ряда

Как показывают исследования, значения l и l' существенно влияют на качество восстановления функции плотности вероятности. Для определения оптимальных значений l^* и $(l')^*$ приходится производить перебор значений $Q\left\{\hat{f}_l^{(l')}\right\}$, для чего необходимо знать вид истинной плотности $f(x)$. Важным вопросом является оценивание параметров l и l' по выборке.

Для построения оценок $\hat{l}$ и $\hat{l}'$ будем максимизировать оценку функционала (1.6). В случае проекционной оценки (1.8) в пространстве $L_{2,w}(\Omega)$ функционал (1.6) преобразуется следующим образом:

$$W\{\hat{f}_l\} = M\left\{\left\|\hat{f}_l\right\|_{2,w}^2 - 2\left(\hat{f}_l, f\right)_w\right\} = \sum_{j=0}^{l}\left(M\{a_j^2\} - 2\alpha_j M\{a_j\}\right),$$

где

$$\alpha_j = (f,\varphi_j)_w = \int_\Omega f(x)\varphi_j(x)w(x)dx = M\{\varphi_j(\xi)w(\xi)\} = M\{\eta_j(\xi)\}.$$

Для оценки Ченцова получаем:

$$M\{a_j\} = \alpha_j;$$

$$W\{\hat{f}_l\} = \sum_{j=0}^{l}\left(M\{a_j^2\} - 2M^2\{a_j\}\right) = \sum_{j=0}^{l}\left(M\{a_j^2\} - 2\left(M\{a_j^2\} - D\{a_j\}\right)\right) =$$

$$= \sum_{j=0}^{l}\left(2D\{a_j\} - M\{a_j^2\}\right) = \sum_{j=0}^{l}\left(\frac{2}{n}D\{\eta_j(\xi)\} - M\{a_j^2\}\right) =$$

$$= M\left\{\sum_{j=0}^{l}\left(\frac{2}{n}s_{\eta_j}^2 - a_j^2\right)\right\} = M\left\{\sum_{j=0}^{l}\left(\frac{2}{n}s_{\eta_j}^2 - m_{\eta_j}^2\right)\right\},$$

где m и s обозначают выборочное среднее и исправленную выборочную дисперсию соответственно:

$$m_\xi = \frac{1}{n}\sum_{i=1}^{n}x_i, \quad s_\xi = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - m_\xi)^2}.$$

Следовательно, значение

$$\hat{W}_l = \sum_{j=0}^{l}\left(\frac{2}{n}s_{\eta_j}^2 - m_{\eta_j}^2\right) \tag{2.21}$$

является несмещённой оценкой функционала $W\{\hat{f}_l\}$. Тогда оценку $\hat{l}$ можно получить минимизацией величины $\hat{W}_l$:

$$\hat{l} = \arg\min_l \hat{W}_l. \tag{2.22}$$

Этот метод наряду с методом оценивания путём максимизации значения (1.12) может быть использован при построении проекционной оценки плотности вероятности. Причём, как показано в [159], метод, основанный на максимизации значения (2.21), часто оказывается близким или лучшим метода (1.12).

Данный подход можно применить также и при настройке параметров l и l' для проекционной оценки, в которой коэффициенты оцениваются обобщёнными методом моментов. В этом случае функционал (1.6) с использованием матричных обозначений преобразуется следующим образом:

$$\begin{aligned}
W\left\{\hat{f}_l^{(l')}\right\} &= \sum_{j=0}^{l}\left(M\left\{\left(a_j^{(l')}\right)^2\right\} - 2\alpha_j M\left\{a_j^{(l')}\right\}\right) = \\
&= M\left\{\left(\mathbf{a}^{(l')}\right)^T \mathbf{a}^{(l')}\right\} - 2\boldsymbol{\alpha}^T M\left\{\mathbf{a}^{(l')}\right\} = \\
&= \operatorname{tr} M\left\{\mathbf{a}^{(l')}\left(\mathbf{a}^{(l')}\right)^T\right\} - 2\operatorname{tr}\left(M\left\{\mathbf{a}^{(l')}\right\}\boldsymbol{\alpha}^T\right) = \\
&= \operatorname{tr} M\left\{\mathbf{B}_{l,l'}^{-1}\hat{\boldsymbol{\nu}}\hat{\boldsymbol{\nu}}^T\left(\mathbf{B}_{l,l'}^{-1}\right)^T\right\} - 2\operatorname{tr}\left(M\left\{\mathbf{B}_{l,l'}^{-1}\hat{\boldsymbol{\nu}}\right\}\boldsymbol{\alpha}^T\right) = \\
&= \operatorname{tr} M\left\{\mathbf{B}_{l,l'}^{-1}\hat{\boldsymbol{\nu}}\hat{\boldsymbol{\nu}}^T\left(\mathbf{B}_{l,l'}^{-1}\right)^T\right\} - 2\operatorname{tr}\left(\mathbf{B}_{l,l'}^{-1}\boldsymbol{\nu}\boldsymbol{\alpha}^T\right).
\end{aligned} \tag{2.23}$$

Для построения несмещённой оценки функционала (2.23) построим несмещённые оценки элементов матрицы

$$\mathbf{G} = \boldsymbol{\nu}\boldsymbol{\alpha}^T = \begin{pmatrix} \nu_0\alpha_0 & \dots & \nu_0\alpha_l \\ \vdots & \ddots & \vdots \\ \nu_{l'}\alpha_0 & \dots & \nu_{l'}\alpha_l \end{pmatrix} = \begin{pmatrix} M\{\xi^0\}M\{\eta_0(\xi)\} & \dots & M\{\xi^0\}M\{\eta_l(\xi)\} \\ \vdots & \ddots & \vdots \\ M\{\xi^{l'}\}M\{\eta_0(\xi)\} & \dots & M\{\xi^{l'}\}M\{\eta_l(\xi)\} \end{pmatrix},$$

для чего воспользуемся несмещённой выборочной оценкой ковариационного момента:

$$\operatorname{cov}\{X, Y\} = M\left\{\frac{1}{n-1}\sum_{i=1}^{n} x_i y_i - \frac{1}{n(n-1)}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i\right\},$$

откуда

$$M\{X\}M\{Y\} = M\{XY\} - \operatorname{cov}\{X,Y\} = \\ = M\left\{\frac{1}{n(n-1)}\left(\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i y_i\right)\right\}.$$

Получаем

$$M\{\xi^{j_1}\}M\{\eta_{j_2}(\xi)\} = M\left\{\frac{1}{n(n-1)}\left(\sum_{i=1}^{n} x_i^{j_1} \sum_{i=1}^{n} \eta_{j_2}(x_i) - \sum_{i=1}^{n} x_i^{j_1}\eta_{j_2}(x_i)\right)\right\},$$

где $j_1 = 0, \ldots, l'$, $j_2 = 0, \ldots, l$.

Тогда несмещённой оценкой матрицы $\mathbf{G}$ является матрица

$$\hat{\mathbf{G}} = \|g_{j_1,j_2}\|_{\substack{j_1=0,\ldots,l',\\ j_2=0,\ldots,l}}$$

в которой

$$g_{j_1,j_2} = \frac{1}{n(n-1)}\left(\sum_{i=1}^{n} x_i^{j_1} \sum_{i=1}^{n} \eta_{j_2}(x_i) - \sum_{i=1}^{n} x_i^{j_1}\eta_{j_2}(x_i)\right),$$

а для несмещённой оценки функционала (2.23) получаем выражение:

$$\hat{W}_{l,l'} = \operatorname{tr} \mathbf{B}_{l,l'}^{-1} \hat{\boldsymbol{\nu}} \hat{\boldsymbol{\nu}}^T \left(\mathbf{B}_{l,l'}^{-1}\right)^T - 2 \operatorname{tr} \mathbf{B}_{l,l'}^{-1} \hat{\mathbf{G}}$$

или

$$\hat{W}_{l,l'} = \operatorname{tr} \mathbf{B}_{l,l'}^{-1} \left(\hat{\boldsymbol{\nu}} \hat{\boldsymbol{\nu}}^T \left(\mathbf{B}_{l,l'}^{-1}\right)^T - 2\hat{\mathbf{G}}\right). \tag{2.24}$$

Теперь оценки $\hat{l}$ и $\hat{l}'$ находятся минимизацией значения $\hat{W}_{l,l'}$:

$$\left(\hat{l}, \hat{l}'\right) = \arg\min_{l,l'} \hat{W}_{l,l'}. \tag{2.25}$$

При этом если используется классический метод моментов, то оценивается только один параметр l:

$$\hat{l} = \arg\min_{l} \hat{W}_{l,l}. \tag{2.26}$$

Введём следующие обозначения для рассматриваемых оценок функции плотности вероятности: см. табл. 2.6.

Таблица 2.6. Обозначения оценок плотности вероятности

Обозначение	Метод оценивания a_j	Метод оценивания l и l'
$\hat{f}_1(x)$	$a_j = a_j^{(\infty)} = \frac{1}{n}\sum_{i=1}^{n}\eta_j(x_i)$	максимизация оценки (1.12)
$\hat{f}_2(x)$		максимизация оценки (2.21)
$\hat{f}_3(x)$	$\mathbf{a} = \mathbf{a}^{(l)} = \mathbf{B}_l^{-1}\hat{\boldsymbol{\nu}}_l$	формула (2.26)
$\hat{f}_4(x)$	$\mathbf{a} = \mathbf{a}^{(l')} = \mathbf{B}_{l,l'}^{-1}\hat{\boldsymbol{\nu}}_{l'}$	формула (2.25)

В оценках $\hat{f}_1(x)$ и $\hat{f}_2(x)$ для настройки параметров используется метод Ченцова (2.10). В оценках $\hat{f}_3(x)$ и $\hat{f}_4(x)$ используется подход, предлагаемый в [151], основанный на применении метода моментов (формулы (2.11) и (2.13) соответственно). Оценки $\hat{f}_1(x)$ и $\hat{f}_2(x)$ различаются между собой способом оценивания длины ряда l: в первом случае применяется оценка (1.12) функционала качества $W\{\hat{f}\}$, во втором случае – несмещённая оценка (2.21) этого функционала.

При этом если функция плотности вероятности восстанавливается в пространстве $L_2[-1;1]$ и в качестве базиса используется система Лежандра, то можно показать, что оценки $\hat{f}_2(x)$, $\hat{f}_3(x)$ и $\hat{f}_4(x)$ тождественны.

В самом деле, в §1.3 было показано, что при данной длине ряда l коэффициенты данных оценок совпадают (2.17). Также нетрудно показать, что в этом случае для любых $l \in \mathbb{N}$ и $l' \geq l$ выполняется тождество

$$\hat{W}_{l,l'} = \hat{W}_{l,l}.$$

Отсюда следует, что оценки $\hat{f}_3(x)$ и $\hat{f}_4(x)$ тождественны.

Покажем теперь, что тождественны оценки $\hat{f}_2(x)$ и $\hat{f}_3(x)$. Для этого достаточно показать, что совпадают оценки (2.21) и (2.24) функционала (1.6), по которым оценивается длина ряда l, т.е. что равенство

$$\operatorname{tr} \mathbf{B}_{l,l}^{-1} \hat{\boldsymbol{\nu}} \hat{\boldsymbol{\nu}}^T \left(\mathbf{B}_{l,l}^{-1}\right)^T - 2 \operatorname{tr} \mathbf{B}_{l,l}^{-1} \hat{\mathbf{G}} = \sum_{j=0}^{l} \left(\frac{2}{n} s_{\varphi_j}^2 - m_{\varphi_j}^2 \right)$$

выполняется при любых значениях l и n, а также $x_1, \ldots, x_n$.

Действительно, так как $\mathbf{B}_{l,l}^{-1} \hat{\boldsymbol{\nu}} = \mathbf{a}^{(\infty)}$, то

$$\operatorname{tr} \mathbf{B}_{l,l}^{-1} \hat{\boldsymbol{\nu}} \hat{\boldsymbol{\nu}}^T \left(\mathbf{B}_{l,l}^{-1}\right)^T = \operatorname{tr} \mathbf{a}^{(\infty)} \left(\mathbf{a}^{(\infty)}\right)^T = \sum_{j=0}^{l} m_{\varphi_j}^2.$$

Преобразуем теперь слагаемое $\operatorname{tr} \mathbf{B}_{l,l} \hat{\mathbf{G}}$. Так как

$$\mathbf{B}_l \cdot \left\| \sum_{i=1}^{n} \varphi_{j_1}(x_i) \sum_{i=1}^{n} \varphi_{j_2}(x_i) - \sum_{i=1}^{n} \varphi_{j_1}(x_i) \varphi_{j_2}(x_i) \right\|_{\substack{j_1=0,\ldots,l \\ j_2=0,\ldots,l}} =$$

$$= \left\| \sum_{k=0}^{l} (x^{j_1}, \varphi_k) \left(\sum_{i=1}^{n} \varphi_k(x_i) \sum_{i=1}^{n} \varphi_{j_2}(x_i) - \sum_{i=1}^{n} \varphi_k(x_i) \varphi_{j_2}(x_i) \right) \right\|_{\substack{j_1=0,\ldots,l \\ j_2=0,\ldots,l}} =$$

$$= \left\| \sum_{i=1}^{n} \sum_{k=0}^{l} (x^{j_1}, \varphi_k)\varphi_k(x_i) \cdot \sum_{i=1}^{n} \varphi_{j_2}(x_i) - \sum_{i=1}^{n} \sum_{j=0}^{l} (x^{j_1}, \varphi_k)\varphi_k(x_i)\varphi_{j_2}(x_i) \right\|_{\substack{j_1=0,\dots,l \\ j_2=0,\dots,l}} =$$

$$= \left\| \sum_{i=1}^{n} x_i^{j_1} \sum_{i=1}^{n} \varphi_{j_2}(x_i) - \sum_{i=1}^{n} \sum_{j=0}^{l} x_i^{j_1} \varphi_{j_2}(x_i) \right\|_{\substack{j_1=0,\dots,l \\ j_2=0,\dots,l}},$$

то

$$\mathbf{B}_l \cdot \frac{1}{n(n-1)} \left\| \sum_{i=1}^{n} \varphi_{j_1}(x_i) \sum_{i=1}^{n} \varphi_{j_2}(x_i) - \sum_{i=1}^{n} \varphi_{j_1}(x_i)\varphi_{j_2}(x_i) \right\|_{\substack{j_1=0,\dots,l \\ j_2=0,\dots,l}} = \hat{\mathbf{G}};$$

$$\mathbf{B}_l^{-1}\hat{\mathbf{G}} = \frac{1}{n(n-1)} \left\| \sum_{i=1}^{n} \varphi_{j_1}(x_i) \sum_{i=1}^{n} \varphi_{j_2}(x_i) - \sum_{i=1}^{n} \varphi_{j_1}(x_i)\varphi_{j_2}(x_i) \right\|_{\substack{j_1=0,\dots,l \\ j_2=0,\dots,l}};$$

$$\operatorname{tr} \mathbf{B}_l^{-1}\hat{\mathbf{G}} = \frac{1}{n(n-1)} \sum_{j=0}^{l} \left(\left(\sum_{i=1}^{n} \varphi_j(x_i) \right)^2 - \sum_{i=1}^{n} \varphi_j^2(x_i) \right).$$

Окончательно получаем:

$$\operatorname{tr} \mathbf{B}_{l,l}^{-1} \hat{\boldsymbol{\nu}} \hat{\boldsymbol{\nu}}^T \left(\mathbf{B}_{l,l}^{-1} \right)^T - 2 \operatorname{tr} \mathbf{B}_{l,l}^{-1} \hat{\mathbf{G}} =$$

$$= \sum_{j=0}^{l} m_{\varphi_j}^2 - \frac{2}{n(n-1)} \sum_{j=0}^{l} \left(\left(\sum_{i=1}^{n} \varphi_j(x_i) \right)^2 - \sum_{i=1}^{n} \varphi_j^2(x_i) \right) =$$

$$= \sum_{j=0}^{l} \left(\frac{2}{n} \frac{n}{n-1} \left(m_{\varphi_j^2} - m_{\varphi_j}^2 \right) - m_{\varphi_j}^2 \right) = \sum_{j=0}^{l} \left(\frac{2}{n} s_{\varphi_j}^2 - m_{\varphi_j}^2 \right).$$

Всё вышесказанное сформулируем в виде следующего предложения.

Предложение 2.6. *Пусть восстанавливаемая функция плотности вероятности $f(x)$ принадлежит пространству $L_2[-1;1]$, φ_j – базис Лежандра в этом пространстве. Тогда оценки $\hat{f}_2(x)$, $\hat{f}_3(x)$ и $\hat{f}_4(x)$ из* табл. 2.6 *тождественны:*

$$\hat{f}_2(x) \equiv \hat{f}_3(x) \equiv \hat{f}_4(x).$$

При использовании других ортонормальных систем все оценки $\hat{f}_1(x)$, $\hat{f}_2(x)$, $\hat{f}_3(x)$ и $\hat{f}_4(x)$ различаются. В работе было проведено сравнение качества указанных оценок для следующих распределений:

1) равномерное распределение на отрезке $\left[0; \frac{1}{2}\right]$:

$$f(x) = \begin{cases} 2, & x \in \left[0; \frac{1}{2}\right] \\ 0, & \text{иначе} \end{cases};$$

2) кубическое распределение на отрезке $\left[-\frac{1}{2}; \frac{1}{2}\right]$:

$$f(x) = \begin{cases} 2(1+4|x|)(1-2|x|)^2, & x \in \left[-\frac{1}{2}; \frac{1}{2}\right] \\ 0, & \text{иначе} \end{cases};$$

3) показательное распределение с параметром $\lambda = 1$:

$$f(x) = \begin{cases} e^{-x}, & x \in [0; +\infty) \\ 0, & \text{иначе} \end{cases};$$

4) нормальное распределение с параметрами $\mu = 1$, $\sigma = 1$:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-1)^2}{2}}.$$

Соответствие между восстанавливаемым распределением, пространством, в котором строится оценка $\hat{f}(x)$, и используемым базисом приведена в табл. 2.7.

Таблица 2.7. Характеристика восстанавливаемых распределений

№	**Распределение**	**Пространство**	**Базис**
1	*равномерное*	$L_2[-1; 1]$	Лежандра
2	*кубическое*	$L_2[-1; 1]$	Лежандра
3	*показательное*	$L_2[0; +\infty)$	Лагерра
4	*нормальное*	$L_2(-\infty; +\infty)$	Эрмита

В качестве меры близости оценки $\hat{f}(x)$ к истинной плотности $f(x)$ взят функционал (2.20). Однако так как теоретический расчёт значений (2.20) для рассматриваемых оценок затруднён, то его значения находятся приближённо методом статистических испытаний.

Метод состоит в следующем. Пусть требуется оценить неизвестное значение z, представляющее собой математическое ожидание случайной величины Z, причём закон распределения случайной величины Z неизвестен. При помощи датчиков псевдослучайных чисел генерируется независимая выборка $z_1, \ldots, z_n$ значений случайной величины Z. Тогда оценку величины z можно взять в виде

$$z \approx \bar{Z} = \frac{1}{n} \sum_{i=1}^{n} z_i.$$

Погрешность данного значения рассчитывается по правилу «трёх сигм»:

$$z \approx \bar{Z} \pm 3\sigma_{\bar{Z}},$$

где $\sigma_{\bar{Z}}$ – среднеквадратичное отклонение выборочного среднего $\bar{Z}$, которое, как известно, связано со среднеквадратичным отклонением случайной величины Z следующим образом:

$$\sigma_{\bar{Z}} = \frac{1}{\sqrt{n}} \sigma_Z.$$

Тогда

$$z \approx \bar{Z} \pm \frac{3}{\sqrt{n}} \sigma_Z.$$

Значение σ_Z, в свою очередь, также оценивается по выборке случайной величины Z:

$$\sigma_Z \approx \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (z_i - \bar{Z})^2}.$$

Если теперь взять $n = 900$, то получаем простую формулу:

$$M\{Z\} \approx \bar{Z} \pm 0.1\sigma_Z.$$

Результаты расчётов значений функционала (2.20) занесены в табл. 2.8.

Таблица 2.8. Сравнение качества оценок

Распределение	$\hat{f}_1(x)$	$\hat{f}_2(x)$	$\hat{f}_3(x)$	$\hat{f}_4(x)$
равномерное	0.466 ± 0.033	$\mathbf{0.433 \pm 0.030}$	$\mathbf{0.433 \pm 0.030}$	$\mathbf{0.433 \pm 0.030}$
кубическое	0.225 ± 0.032	$\mathbf{0.198 \pm 0.028}$	$\mathbf{0.198 \pm 0.028}$	$\mathbf{0.198 \pm 0.028}$
показательное	0.078 ± 0.013	0.071 ± 0.010	$\mathbf{0.049 \pm 0.008}$	$\mathbf{0.047 \pm 0.008}$
нормальное	0.051 ± 0.006	0.047 ± 0.005	0.050 ± 0.005	0.048 ± 0.005

Как видно из таблицы, на базисах, отличных от базиса Лежандра, при имеющейся погрешности оценка $\hat{f}_4(x)$ не даёт существенного преимущества по сравнению с оценкой $\hat{f}_3(x)$. Это объясняется как незначительной разницей значений функционала качества (2.20) при различных значениях длины ряда l, так и ухудшением качества оценивания из-за увеличения количества настраиваемых параметров оценки $\hat{f}_l^{(l')}(x)$ по сравнению с оценкой $\hat{f}_l^{(l)}(x)$.

В целом, *проекционные оценки, основанные на методе моментов, показывают аналогичные или лучшие результаты при различных восстанавливаемых распределениях и используемых ортонормальных системах по сравнению с оценками* $\hat{f}_1(x)$ *и* $\hat{f}_2(x)$.

§ 2.4. Многомерный случай

В настоящем разделе рассматриваются особенности применения рассмотренных подходов к оцениванию коэффициентов и длины ряда проекционной оценки функции плотности вероятности многомерной случайной величины (случайного вектора). Как уже было показано (предложение 2.4), для любого непрерывного случайного вектора $\boldsymbol{\xi}$ существует пространство $L_{2,w}(\Omega)$, содержащее его функцию плотности вероятности. Важный класс пространств этого вида образуют пространства, в которых

$$\Omega = \Omega_1 \times \dots \times \Omega_k, \quad \Omega_i \subseteq \mathbb{R},$$

а мера μ_w является прямым произведением линейных (1-мерных) мер [83, с. 332]:

$$\mu_w = \mu_{w_1} \otimes \dots \otimes \mu_{w_k}, \quad \mu_{w_i}(X) = \int_X w_i(x)dx,$$

где все $w_i(x)$ удовлетворяют условиям 1) – 3) в определении 2.1. В этом случае

$$\mu_w(X_1 \times \dots \times X_k) = \int_{X_1} \cdots \int_{X_k} w_1(x_1)\dots w_k(x_k)dx_1\dots dx_k,$$

откуда

$$w(x_1, \dots, x_k) = w_1(x_1)\dots w_k(x_k). \tag{2.27}$$

Очевидно обратное: если весовая функция $w(x_1, \dots x_k)$ может быть представлена в виде (2.27), то мера

$$\mu_w(x) = \int \cdots \int_X w(x_1, \dots, x_k)dx_1\dots dx_k$$

является прямым произведением мер μ_{w_1}, ..., μ_{w_k}, где

$$\mu_{w_i}(X_i) = \int_{X_i} w_i(x)dx, \quad i = 1, \dots, k,$$

$$X_1 \times \cdots \times X_k \supseteq X.$$

В случае (2.27) легко построить базис пространства $L_{2,w}(\Omega)$. Действительно, если системы функций $\{\varphi_{1,j}\}_{j=0}^{\infty}$, ..., $\{\varphi_{k,j}\}_{j=0}^{\infty}$ являются базисами в пространствах $L_{2,w_1}(\Omega_1)$, ..., $L_{2,w_k}(\Omega_k)$ соответственно, то система функций

$$\psi_{j_1,\ldots,j_k}(x_1,\ldots,x_k) = \varphi_{1,j_1}(x_1) \cdot \ldots \cdot \varphi_{k,j_k}(x_k), \quad j_1,\ldots,j_k = 0,1,\ldots$$

является базисом в пространстве $L_{2,w}(\Omega_1 \times \ldots, \times\Omega_k)$.

Недостатком данного подхода является ограничение множества восстанавливаемых функций плотности вероятности. Другими словами, имеет место следующая теорема.

Теорема 2.6. *При $k \geq 2$ существуют функции $f \in L_1(\Omega)$, не принадлежащие никакому пространству $L_{p,w}(\Omega)$, в котором весовая функция имеет вид (2.27).*

Обозначим через $U_\delta(\mathbf{a})$ δ-окрестность точки $\mathbf{a} \in \Omega \subseteq \mathbb{R}^k$:

$$U_\delta(\mathbf{a}) = \left\{ \mathbf{x} \in \Omega \,\middle|\, \sum_{j=1}^{k} (x_j - a_j)^2 < \delta^2 \right\}.$$

В формулировке следующей леммы используется понятие существенного предела измеримой функции.

Лемма 2.3. *Пусть весовая функция $w(\mathbf{x})$ такова, что для некоторой точки $\mathbf{a} \in \Omega$ выполняется*

$$\forall\delta > 0 \quad L_{p,w}^k(U_\delta(\mathbf{a})) \setminus L_p^k(U_\delta(\mathbf{a})) \neq \varnothing. \tag{2.28}$$

Тогда

$$\operatorname*{ess\,lim}_{\mathbf{x}\to\mathbf{a}} w(\mathbf{x}) = 0.$$

Доказательство. Пусть, напротив,

$$\operatorname*{ess\,lim}_{\mathbf{x}\to\mathbf{a}} w(\mathbf{x}) = A > 0.$$

Тогда для любых $\varepsilon > 0$ найдётся такая окрестность $U_\delta(\mathbf{a})$ точки $\mathbf{a}$, что почти для всех $x \in U_\delta(\mathbf{a})$ выполняется неравенство

$$A - \varepsilon < w(\mathbf{x}) < A + \varepsilon.$$

В частности при $\varepsilon = \frac{A}{2}$ получаем, что почти для всех $x \in U_\delta(\mathbf{a})$ выполняется

$$w(\mathbf{x}) > \frac{A}{2}.$$

Отсюда следует

$$\operatorname{ess}\inf_{x \in U_\delta(\mathbf{a})} w(\mathbf{x}) \geq \frac{A}{2} > 0.$$

По следствию 2.3.1 получаем

$$L^k_{p,w}(U_\delta(\mathbf{a})) \setminus L^k_p(U_\delta(\mathbf{a})) = \varnothing,$$

что противоречит условию (2.28). □

Лемма 2.4. *Пусть* Σ *– некоторая* σ*-алгебра множеств с единицей* Ω, $\mu : \Sigma \to [0; +\infty)$ *– полная* σ*-аддитивная мера со счётным базисом,* $f : \Omega \to [0; +\infty)$ *– измеримая функция,*

$$A = \{x \in \Omega \mid f(x) = 0\},$$

$$B = \{a \in \Omega \mid \operatorname{ess}\lim_{x \to a} f(x) = 0\}.$$

Тогда

$$\mu(A \bigtriangleup B) = 0.$$

Доказательство. Так как

$$A \bigtriangleup B = (A \setminus B) \sqcup (B \setminus A),$$

то

$$\mu(A \bigtriangleup B) = \mu(A \setminus B) + \mu(B \setminus A). \tag{2.29}$$

Найдём $\mu(A \setminus B)$. Для этого заметим, что

$$A \setminus B = A \cap (\Omega \setminus B).$$

Введём обозначения:

$$B^* = \Omega \setminus B = \{a \in \Omega \mid \operatorname{ess}\lim_{x \to a} f(x) > 0\};$$

$$B^*_n = \left\{ a \in \Omega \,\middle|\, \operatorname{ess}\lim_{x \to a} f(x) > \frac{1}{n} \right\}.$$

Тогда
$$B_1^* \subseteq B_2^* \subseteq \ldots \subseteq B_n^* \subseteq \ldots;$$
$$B^* = \bigcup_{n=1}^{\infty} B_n^*;$$
$$A \cap (\Omega \setminus B) = A \cap \bigcup_{n=1}^{\infty} B_n^* = \bigcup_{n=1}^{\infty} (A \cap B_n^*);$$
$$\mu(A \setminus B) = \lim_{n\to\infty} \mu(A \cap B_n^*). \tag{2.30}$$

Условие
$$\operatorname{ess}\lim_{x\to a} f(x) > \frac{1}{n}$$
означает, что существует такая окрестность $U_n(a)$ точки a, что почти для всех $x \in U_n(a)$ выполняется неравенство
$$f(x) > \frac{1}{2n}. \tag{2.31}$$
При этом можно считать, что окрестности $U_n(a)$ выбираются из счётного базиса меры μ.

Окружим каждую точку $a \in B_n^*$ соответствующей окрестностью $U_n(a)$. Очевидно, что
$$B_n^* \subseteq \bigcup_{a\in B_n^*} U_n(a).$$
Так как все $U_n(a)$ выбираются из счётного базиса меры μ, последнее объединение фактически состоит из не более чем счётного множества таких окрестностей. Занумеруем их:
$$\bigcup_{a\in B_n^*} U_n(a) = \bigcup_m U_{n,m}.$$
Тогда
$$A \cap B_n^* \subseteq A \cap \bigcup_m U_{n,m} = \bigcup_m (A \cap U_{n,m});$$
$$\mu(A \cap B_n^*) \le \mu\left(\bigcup_m (A \cap U_{n,m})\right) \le \sum_m \mu(A \cap U_{n,m}).$$
Так как почти для всех $x \in U_{n,m}$ выполняется неравенство (2.31), то при любых n и m
$$\mu(A \cap U_{n,m}) = 0.$$

Тогда

$$\mu(A \cap B_n^*) \leq \sum_m \mu(A \cap U_{n,m}) \leq 0,$$

и из равенства (2.30) получаем

$$\mu(A \setminus B) = 0.$$

Применим аналогичные рассуждения при нахождении $\mu(B \setminus A)$:

$$B \setminus A = B \cap (\Omega \setminus A).$$

Введём обозначения:

$$A^* = \Omega \setminus A = \{a \in \Omega \mid f(x) > 0\};$$

$$A_n^* = \left\{a \in \Omega \,\middle|\, f(a) > \frac{1}{n}\right\}.$$

Тогда

$$A_1^* \subseteq A_2^* \subseteq \ldots \subseteq A_n^* \subseteq \ldots;$$

$$A^* = \bigcup_{n=1}^{\infty} A_n^*;$$

$$B \cap (\Omega \setminus A) = \bigcup_{n=1}^{\infty} (B \cap A_n^*);$$

$$\mu(B \setminus A) = \lim_{n\to\infty} (B \cap A_n^*). \tag{2.32}$$

Условие

$$\operatorname{ess}\lim_{x\to a} f(x) = 0$$

означает что для любых натуральных k существует такая окрестность $U_k(a)$ точки a, что почти для всех $x \in U_k(a)$ выполняется неравенство

$$f(x) < \frac{1}{k}.$$

Тогда множество B можно представить в виде

$$B = \bigcap_{k=1}^{\infty} B_k,$$

где

$$B_k = \left\{a \in \Omega \,\middle|\, \exists U_k(a)\ \mu\left\{x \in U_k(a) \,\middle|\, f(x) \geq \frac{1}{k}\right\} = 0\right\}.$$

При этом также считаем, что окрестности $U_k(a)$ выбраны из счётного базиса меры μ.

Выберем произвольную точку $a \in B_k$ и окружим её соответствующей окрестностью $U_k(a)$. Очевидно, что

$$B_k \subseteq \bigcup_{a \in B_k} U_k(a), \quad k = 1, 2, ...$$

В частности, если взять $k = n$, то получим

$$B \subseteq B_n \subseteq \bigcup_m U_{n,m};$$

$$B \cap A_n^* \subseteq \bigcup_m U_{n,m} \cap A_n^* = \bigcup_m (U_{n,m} \cap A_n^*)$$

$$\mu\left(B \cap A_n^*\right) \leq \sum_m \mu\left(U_{n,m} \cap A_n^*\right).$$

Так как почти для всех $a \in U_{n,m}$ выполняется неравенство $f(a) < \frac{1}{n}$, то для любых $m, n \in \mathbb{N}$

$$\mu\left(U_{n,m} \cap A_n^*\right) = 0;$$

$$\mu\left(B \cap A_n^*\right) = 0.$$

Наконец, из равенства (2.32) получаем

$$\mu(B \setminus A) = 0.$$

Теперь из (2.29) получим окончательно:

$$\mu(A \triangle B) = 0.$$

□

Доказательство теоремы 2.6. Теперь достаточно взять функцию $f(\mathbf{x})$, у которой множество точек $\mathbf{a}$, удовлетворяющих условию (2.28), таково, что её проекции на координатные оси Ox_1, ..., Ox_k имеют положительные меры:

$$\mu_j\, \pi_j \left\{ \mathbf{a} \in \Omega \mid \forall \delta > 0 \; f \in L^k_{p,w}(U_\delta(\mathbf{a})) \setminus L^k_p(U_\delta(\mathbf{a})) \right\} > 0, \quad j = 1, ..., k,$$

где μ_j – мера Лебега на оси Ox_j.

Например, если $\Omega_1 = ... = \Omega_k = [0; 1]$, то таким множеством является, в частности, отрезок, соединяющий точки $(0, ..., 0)$ и $(1, ..., 1)$.

Действительно, если ввести обозначения

$$X_j = \{a \in \Omega_j \mid \operatorname{ess} \lim_{x \to a} w_j(x) = 0\}, \quad j = 1, ..., k,$$

то, согласно лемме 2.3,

$$\pi_j \left\{ \mathbf{a} \in \Omega \,\middle|\, \forall \delta > 0 \ f \in L^k_{p,w}(U_\delta(\mathbf{a})) \setminus L^k_p(U_\delta(\mathbf{a})) \right\} \subseteq X_j,$$

откуда

$$\mu_j X_j > 0.$$

Рассмотрим множество

$$D = \{\mathbf{a} \in \Omega \mid \operatorname{ess} \lim_{\mathbf{x} \to \mathbf{a}} w(\mathbf{x}) = 0\}.$$

Так как из

$$\operatorname{ess} \lim_{x_j \to a_j} w_j(x_j) = 0, \quad j = 1, ..., k,$$

следует

$$\operatorname{ess} \lim_{\mathbf{x} \to \mathbf{a}} w(\mathbf{x}) = 0,$$

то

$$X_1 \times ... \times X_k = \{(a_1, ..., a_k) \in \Omega \mid a_1 \in X_1 \wedge ... \wedge a_k \in X_k\} \subseteq D,$$

откуда

$$\mu D \geq \mu_1 X_1 \cdot \ldots \cdot \mu_k X_k > 0.$$

Наконец, воспользовавшись леммой 2.4, заключаем, что

$$\mu\{\mathbf{a} \in \Omega \mid w(\mathbf{a}) = 0\} = \mu\{\mathbf{a} \in \Omega \mid \operatorname{ess} \lim_{\mathbf{x} \to \mathbf{a}} w(\mathbf{x}) = 0\} > 0,$$

что противоречит определению весовой функции как функции, положительной почти всюду. $\square$

Примером такой функции плотности вероятности для размерности $k = 2$ является функция

$$f(x_1, x_2) = \frac{3}{8\sqrt{|x_1 - x_2|}}, \quad x_1, x_2 \in [0; 1], \quad x_1 \neq x_2.$$

Отметим, что ограниченная почти всюду функция плотности вероятности всегда принадлежит пространству L^k_2. Действительно, пусть почти для всех $\mathbf{x} \in \mathbb{R}^k$ выполняется неравенство $f(\mathbf{x}) \leq M < +\infty$ и

$$\int_{\mathbb{R}^k} f(\mathbf{x}) d\mu = 1.$$

Тогда

$$\int_{\mathbb{R}^k} f^2(\mathbf{x})d\mu \leq M \int_{\mathbb{R}^k} f(\mathbf{x})d\mu = M < +\infty.$$

Следовательно, вводить весовое пространство $L^k_{2,w}$ имеет смысл в случаях, когда не гарантируется ограниченность восстанавливаемой функции плотности вероятности.

В общем случае проекционная оценка плотности вероятности случайного вектора имеет вид:

$$\hat{f}_{l_1,\dots,l_k}(x_1,\dots,x_k) = \sum_{j_1=0}^{l_1} \cdots \sum_{j_k=0}^{l_k} a_{j_1,\dots,j_k} \psi_{j_1,\dots,j_k}(x_1,\dots,x_k),$$

где $a_{j_1,\dots,j_k}$, $l_1, \dots, l_k$ – параметры проекционной оценки, подлежащие настройке.

Для настройки коэффициентов $a_{j_1,\dots,j_k}$ по выборке

$$\mathbf{x}_i = (x_{i1},\dots,x_{ik}), \quad i = 1,\dots,n$$

значений исследуемого случайного вектора $\boldsymbol{\xi}$ могут быть применены методы, рассмотренные в § 2.2. Введём следующие обозначения:

– столбец базисных функций:

$$\boldsymbol{\psi} = (\psi_{0,\dots,0},\dots,\psi_{l_1,\dots,l_k})^T,$$

причём

$$\eta_{j_1,\dots,j_k}(\mathbf{x}) = \psi_{j_1,\dots,j_k}(\mathbf{x}) w(\mathbf{x}),$$

$$\boldsymbol{\eta} = (\eta_{0,\dots,0},\dots,\eta_{l_1,\dots,l_k})^T;$$

– столбец коэффициентов Фурье:

$$\boldsymbol{\alpha} = (\alpha_{0,\dots,0},\dots,\alpha_{l_1,\dots,l_k})^T;$$

– столбец параметров:

$$\mathbf{a} = (a_{0,\dots,0},\dots,a_{l_1,\dots,l_k})^T;$$

– столбец произведения степеней компонент случайного вектора $\boldsymbol{\xi}$ и весовой функции:

$$\boldsymbol{\Xi} = \Big(\xi_1^0 \dots \xi_k^0 w(\xi_1,\dots,\xi_k), \ \ \dots, \ \ \xi_1^{l_1} \dots \xi_k^{l_k} w(\xi_1,\dots,\xi_k)\Big)^T;$$

– столбец *смешанных взвешенных начальных моментов* компонент случайного вектора $\boldsymbol{\xi}$:

$$\boldsymbol{\nu} = M\{\boldsymbol{\Xi}\} = \left\| (x_1^{j_1} \dots x_k^{j_k}, f)_w \right\|_{l_1,\dots,l_k}^T;$$

– столбец *выборочных смешанных взвешенных начальных моментов* компонент случайного вектора $\boldsymbol{\xi}$:

$$\hat{\boldsymbol{\nu}} = (\hat{\nu}_{0,\dots,0}, \dots, \hat{\nu}_{l_1,\dots,l_k})^T,$$

где

$$\hat{\nu}_{j_1,\dots,j_k} = \frac{1}{n}\sum_{i=1}^{n} x_{i1}^{j_1} \dots x_{ik}^{j_k} w(\mathbf{x}_i);$$

– матрица скалярных произведений мономов $x_1^{j_1} \dots x_k^{j_k}$ на базисные функции:

$$\mathbf{B} = \begin{pmatrix} (x_1^0 \dots x_k^0, \psi_{0,\dots,0})_w & \dots & (x_1^0 \dots x_k^0, \psi_{l_1,\dots,l_k})_w \\ \vdots & \ddots & \vdots \\ (x_1^{l_1} \dots x_k^{l_k}, \psi_{0,\dots,0})_w & \dots & (x_1^{l_1} \dots x_k^{l_k}, \psi_{l_1,\dots,l_k})_w \end{pmatrix}.$$

Таким образом, распространяя формулу (2.10) на многомерный случай, используя введённые обозначения, получим:

$$\mathbf{a} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\psi}(\mathbf{x}_i) w(\mathbf{x}_i) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\eta}(\mathbf{x}_i). \tag{2.33}$$

Формула (2.11) сохраняет свой вид:

$$\mathbf{a} = \mathbf{B}^{-1}\hat{\boldsymbol{\nu}}. \tag{2.34}$$

Столбец коэффициентов $\mathbf{a}$, рассчитанный с помощью формулы (2.33) будем по аналогии с обозначением в § 2.2 обозначать через $\mathbf{a}^{(\infty)}$, а рассчитанный с помощью (2.34) – через $\mathbf{a}^{(l_1,\dots,l_k)}$. Получаемые оценки плотности вероятности обозначим через $\hat{f}_{l_1,\dots,l_k}^{(\infty)}(x_1,\dots,x_k)$ и $\hat{f}_{l_1,\dots,l_k}^{(l_1,\dots,l_k)}(x_1,\dots,x_k)$ соответственно.

На многомерный случай также можно распространить подход к оцениванию коэффициентов $a_{j_1,\dots,j_k}$ с помощью обобщения метода моментов, рассмотренного в § 2.2, однако из-за громоздкости последнего и незначительного различия качества восстановления плотности вероятности по сравнению с классическим методом моментов это не представляется целесообразным.

Для настройки параметров $l_1, \ldots, l_k$ построим оценку функционала (1.6), который в многомерном случае имеет вид:

$$W\left\{\hat{f}_{l_1,\ldots,l_k}\right\} = M\{\mathbf{a}^T\mathbf{a}\} - 2\boldsymbol{\alpha}^T M\{\mathbf{a}\}. \tag{2.35}$$

Если $\mathbf{a} = \mathbf{a}^{(\infty)}$, то

$$W\left\{\hat{f}_{l_1,\ldots,l_k}\right\} = W\left\{\hat{f}^{(\infty)}_{l_1,\ldots,l_k}\right\} = \frac{1}{n}M\{\boldsymbol{\eta}^T(\boldsymbol{\xi})\boldsymbol{\eta}(\boldsymbol{\xi})\} - \frac{n+1}{n}\boldsymbol{\alpha}^T\boldsymbol{\alpha} =$$

$$= \frac{1}{n}\operatorname{tr}\Sigma\left\{\boldsymbol{\eta}(\boldsymbol{\xi})\right\} - \boldsymbol{\alpha}^T\boldsymbol{\alpha}.$$

Если $\mathbf{a} = \mathbf{a}^{(l_1,\ldots,l_k)}$, то

$$W\left\{\hat{f}_{l_1,\ldots,l_k}\right\} = W\left\{\hat{f}^{(l_1,\ldots,l_k)}_{l_1,\ldots,l_k}\right\} =$$

$$= \frac{1}{n}\operatorname{tr}\mathbf{B}^{-1}\operatorname{Cov}\{\boldsymbol{\Xi}\}(\mathbf{B}^{-1})^T - \left(2\boldsymbol{\alpha}^T - \boldsymbol{\nu}^T(\mathbf{B}^{-1})^T\right)\mathbf{B}^{-1}\boldsymbol{\nu}.$$

Для настройки параметров l_1, …, l_k рассматривается несколько подходов. Во-первых, может быть применён метод, аналогичный методу, рассмотренному в § 2.3, основанный на построении несмещённых оценок функционала (2.35). Так, для оценок $\hat{f}^{(\infty)}_{l_1,\ldots,l_k}(x_1,\ldots,x_k)$ и $\hat{f}^{(l_1,\ldots,l_k)}_{l_1,\ldots,l_k}(x_1,\ldots,x_k)$ получаем

$$\hat{W}_{l_1,\ldots,l_k} = \hat{W}\left\{\hat{f}^{(\infty)}_{l_1,\ldots,l_k}\right\} = \frac{2}{n-1}m_{\boldsymbol{\psi}^T\boldsymbol{\psi}} - \frac{n+1}{n-1}m^T_{\boldsymbol{\psi}}m_{\boldsymbol{\psi}}, \tag{2.36}$$

$$\hat{W}_{l_1,\ldots,l_k} = \hat{W}\left\{\hat{f}^{(l_1,\ldots,l_k)}_{l_1,\ldots,l_k}\right\} = \hat{\boldsymbol{\nu}}^T(\mathbf{B}^{-1})^T\mathbf{B}^{-1}\hat{\boldsymbol{\nu}} - 2\operatorname{tr}\mathbf{B}^{-1}\hat{\mathbf{G}}. \tag{2.37}$$

Тогда оценки $\hat{l}_1$, …, $\hat{l}_k$ могут быть найдены путём максимизации соответствующих оценок функционала (2.35) по значениям l_1, …, l_k. Полученные оценки функции плотности вероятности обозначим через $\hat{f}_1(\mathbf{x})$ и $\hat{f}_2(\mathbf{x})$.

Второй подход состоит в уменьшении числа настраиваемых параметров. Пусть $l_1 = \cdots = l_k = l$. Тогда максимизация значений оценок (2.36) и (2.37) будет проводиться под одному значению l. Полученные в этом случае оценки функции плотности вероятности обозначим через $\hat{f}_3(\mathbf{x})$ и $\hat{f}_4(\mathbf{x})$.

Сравнение качества полученных оценок проводилось на следующих законах распределения:

1) двумерное равномерное распределение в круге:

$$f(x_1,x_2) = \frac{1}{\pi}, \quad x_1^2 + x_2^2 \le 1;$$

2) трёхмерное равномерное распределение в шаре:

$$f(x_1, x_2, x_3) = \frac{3}{4\pi}, \quad x_1^2 + x_2^2 + x_3^2 \leq 1;$$

3) двумерное нормальное распределение:

$$f(x_1, x_2) = \frac{1}{2\pi} e^{-\frac{x_1^2 + x_2^2}{2}};$$

4) трёхмерное нормальное распределение:

$$f(x_1, x_2, x_3) = \frac{1}{(2\pi)^{3/2}} e^{-\frac{x_1^2 + x_2^2 + x_3^2}{2}};$$

5) «распределение №5»:

$$f(x_1, x_2) = \frac{3}{(x_1 + 3)^3} \left((x_1 + 3)^2 - (4x_2)^2 \right),$$

$$|x_1| \leq 1, \ |x_2| \leq 1, \ 4|x_2| \leq x_1 + 3.$$

Результаты численных расчётов значений функционала (2.20) для рассматриваемых оценок и законов распределения представлены в табл. 2.9.

Таблица 2.9. Сравнение качества оценок многомерной плотности

Распред.	$Q\{\hat{f}_1\}$	$Q\{\hat{f}_2\}$	$Q\{\hat{f}_3\}$	$Q\{\hat{f}_4\}$
1	0.0518 ± 0.0024	0.0518 ± 0.0024	$\mathbf{0.0403 \pm 0.0015}$	$\mathbf{0.0403 \pm 0.0015}$
2	0.4713 ± 0.0020	0.4713 ± 0.0020	$\mathbf{0.0476 \pm 0.0011}$	$\mathbf{0.0476 \pm 0.0011}$
3	0.00202 ± 0.00030	0.00263 ± 0.00035	$\mathbf{0.00111 \pm 0.00027}$	0.00219 ± 0.00041
4	0.000580 ± 0.000068	$\mathbf{0.000122 \pm 0.000017}$	0.000288 ± 0.000072	$\mathbf{0.000120 \pm 0.000017}$
5	$\mathbf{0.4638 \pm 0.0043}$	$\mathbf{0.4638 \pm 0.0043}$	0.4962 ± 0.0042	0.4962 ± 0.0042

Как видно из таблицы, подход, при котором $l_1 = \cdots = l_k = l$, не всегда приводит к улучшению качества аппроксимации. Например, в случае асимметричного распределения №5 с помощью него были получены существенно худшие результаты. В случае остальных распределений (симметричных) данный подход улучшает качество аппроксимации.

Выводы

В данной главе решена задача расширения области применимости проекционной оценки для плотностей с несуммируемым квадратом. Определено гильбертово пространство $L_{2,w}(\Omega)$. Доказано (предл. 2.4), что для любой функции плотности вероятности $f(\mathbf{x})$ проекционная оценка сходится к $f(\mathbf{x})$ в пространстве $L_{2,w}(\Omega)$ при подходящем выборе весовой функции $w(\mathbf{x})$. Рассмотрены способы построения весовой функции, заданным образом расширяющие «обычное» гильбертово пространство $L_2(\Omega)$.

Также решена задача исследования возможности применения метода моментов при настройке проекционной оценки. В ходе решения данной задачи было предложено обобщение метода моментов, которое позволило повысить эффективность проекционной оценки. Предложенный подход распространяется на многомерный случай.

Глава 3. Применение оценок плотности вероятности

В данной главе рассматриваются приложения методов восстановления функции плотности вероятности при решении прикладных задач: оценивание функции регрессии, классификация, оценивание количества информации. Производится сравнение методов, основанных на проекционной оценке, с методами, основанными на оценке Розенблатта – Парзена, при различных способах настройки параметров. Сравнение происходит на тестовых задачах при помощи функционалов качества, численные значения которых находятся методом статистических испытаний.

§ 3.1. Оценивание функции регрессии

Одним из основных приложений алгоритмов восстановления функции плотности вероятности является задача оценивания функции регрессии. Задача оценивания функции регрессии относится к классу задач аппроксимации стохастических зависимостей и возникает при построении непараметрических моделей статических объектов [88].

Функцией (множественной) регрессии φ случайной величины Y на случайные величины $X_1, \ldots, X_k$ называется условное математическое ожидание:

$$\varphi(x_1, \ldots, x_k) = M\{Y \mid X_1 = x_1, \ldots, X_k = x_k\}.$$

Пусть имеется выборка $(\mathbf{x}_i, y_i) = (x_{i1}, \ldots, x_{ik}, y_i)$, $i = 1, \ldots, n$, независимых наблюдений непрерывного случайного вектора $(X_1, \ldots, X_k, Y)$, имеющего неизвестную плотность вероятности $f(x_1, \ldots, x_k, y)$. Требуется построить оценку $\hat{\varphi}(x_1, \ldots, x_k)$ функции регрессии Y на $X_1, \ldots, X_k$.

По определению функции регрессии имеем:

$$\varphi(x_1, \ldots, x_k) = \frac{\displaystyle\int_{\mathbb{R}} y f(x_1, \ldots, x_k, y) dy}{\displaystyle\int_{\mathbb{R}} f(x_1, \ldots, x_k, y) dy}.$$

Оценку функции регрессии получим, взяв вместо истинной плотности f

её непараметрическую оценку $\hat{f}$:

$$\hat{\varphi}(x_1, \dots, x_k) = \frac{\displaystyle\int_{\mathbb{R}} y\hat{f}(x_1, \dots, x_k, y)dy}{\displaystyle\int_{\mathbb{R}} \hat{f}(x_1, \dots, x_k, y)dy}.$$

В случае использования проекционной оценки

$$\hat{f}(x_1, \dots, x_k, y) = \sum_{j_1=0}^{l_1} \dots \sum_{j_k=0}^{l_k} \sum_{j_{k+1}=0}^{l_{k+1}} a_{j_1,\dots,j_k,j_{k+1}} \psi_{j_1,\dots,j_k,j_{k+1}}(x_1, \dots, x_k, y)$$

получаем следующее выражение для оценки функции регрессии:

$$\hat{\varphi}(x_1, \dots, x_k) = \frac{\displaystyle\sum_{j_1=0}^{l_1} \dots \sum_{j_k=0}^{l_k} \sum_{j_{k+1}=0}^{l_{k+1}} a_{j_1,\dots,j_k,j_{k+1}} \int_{\mathbb{R}} y\psi_{j_1,\dots,j_k,j_{k+1}}(x_1, \dots, x_k, y)dy}{\displaystyle\sum_{j_1=0}^{l_1} \dots \sum_{j_k=0}^{l_k} \sum_{j_{k+1}=0}^{l_{k+1}} a_{j_1,\dots,j_k,j_{k+1}} \int_{\mathbb{R}} \psi_{j_1,\dots,j_k,j_{k+1}}(x_1, \dots, x_k, y)dy}.$$

При этом если базисные функции $\psi_{j_1,\dots,j_k,j_{k+1}}$ можно представить в виде (см. § 2.4)

$$\psi_{j_1,\dots,j_k,j_{k+1}}(x_1, \dots, x_k, y) = \varphi_{j_1}(x_1) \dots \varphi_{j_k}(x_k)\varphi_{j_{k+1}}(y),$$

то

$$\hat{\varphi}(x_1, \dots, x_k) = \frac{\displaystyle\sum_{j_1=0}^{l_1} \dots \sum_{j_k=0}^{l_k} \sum_{j_{k+1}=0}^{l_{k+1}} a_{j_1,\dots,j_k,j_{k+1}} \varphi_{j_1}(x_1) \dots \varphi_{j_k}(x_k) \int_{\mathbb{R}} y\varphi_{j_{k+1}}(y)dy}{\displaystyle\sum_{j_1=0}^{l_1} \dots \sum_{j_k=0}^{l_k} \sum_{j_{k+1}=0}^{l_{k+1}} a_{j_1,\dots,j_k,j_{k+1}} \varphi_{j_1}(x_1) \dots \varphi_{j_k}(x_k) \int_{\mathbb{R}} \varphi_{j_{k+1}}(y)dy}.$$

Для настройки параметров $l_1, \dots, l_{k+1}, a_{0,\dots,0}, \dots, a_{l_1,\dots,a_{l_{k+1}}}$ могут быть применены методы, рассмотренные в § 2.4.

В случае использования оценки Розенблатта – Парзена

$$\hat{f}(x_1, \dots, x_k, y) = \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{k} \frac{1}{h_j} K\left(\frac{x_j - x_{ij}}{h_j}\right) \frac{1}{h_{k+1}} K\left(\frac{y - y_i}{h_{k+1}}\right)$$

получаем следующее выражение для оценки функции регрессии:

$$\hat{\varphi}(x_1, \dots, x_k) = \frac{\displaystyle\sum_{i=1}^{n} \prod_{j=1}^{k} K\left(\frac{x_j - x_{ij}}{h_j}\right) \int_{\mathbb{R}} yK\left(\frac{y - y_i}{h_{k+1}}\right) dy}{\displaystyle\sum_{i=1}^{n} \prod_{j=1}^{k} K\left(\frac{x_j - x_{ij}}{h_j}\right) \int_{\mathbb{R}} K\left(\frac{y - y_i}{h_{k+1}}\right) dy}.$$

При этом если ядро $K(\cdot)$ удовлетворяет условиям

1) $\displaystyle\int_{\mathbb{R}} K(z)dz = 1$,

2) $\displaystyle\int_{\mathbb{R}} zK(z)dz = 0$,

то

$$\hat{\varphi}(x_1, \ldots, x_k) = \frac{\sum\limits_{i=1}^{n} y_i \prod\limits_{j=1}^{k} K\left(\frac{x_j - x_{ij}}{h_j}\right)}{\sum\limits_{i=1}^{n} \prod\limits_{j=1}^{k} K\left(\frac{x_j - x_{ij}}{h_j}\right)}.$$

Для настройки параметров размытости h_1, …, h_k могут быть применены методы, рассмотренные в главе 1.

В настоящей работе проводится сравнение качества алгоритмов восстановления функции регрессии, основанных на проекционной оценке и оценке Розенблатта – Парзена. Качество восстановления функции регрессии определялось при решении следующей задачи. Имеется случайная величина X, равномерно распределённая на отрезке $[-1; 1]$. Случайная величина Y определена следующим образом:

$$Y = X + \varepsilon,$$

где ε – центрированная нормально распределённая помеха со среднеквадратичным отклонением $\sigma = 0.05$. Истинная функция регрессии Y на X, очевидно, равна

$$\varphi(x) = x.$$

Дана независимая выборка

$$(x_1, y_1), \ldots, (x_n, y_n)$$

объёма $n = 100$ случайного вектора (X, Y). Требуется построить оценку $\hat{\varphi}(x)$ функции регрессии. Качество восстановления функции регрессии определяется при помощи функционала

$$Q\{\hat{\varphi}\} = M\left\{\|\hat{\varphi} - \varphi\|_{L_1[-1;1]}\right\}, \tag{3.1}$$

значение которого находится приближённо методом статистических испытаний (см. § 2.3).

Для решения данной задачи использовались алгоритмы расчёта параметров проекционной оценки плотности вероятности, рассмотренные в главе 2, а также методы настройки параметра размытости оценки Розенблатта – Парзена, рассмотренные в главе 1.

При построении как проекционной оценки, так и оценки Розенблатта – Парзена функции плотности вероятности случайного вектора (X, Y) считается известным априори, что его компоненты имеют одинаковую размерность, носитель не выходит за пределы квадрата $[-1; 1] \times [-1; 1]$:

$$\operatorname{supp}(X, Y) \subseteq [-1; 1] \times [-1; 1],$$

а также что

$$\operatorname*{ess\,sup}_{x,y} f(x, y) < +\infty.$$

Таким образом, мы заключаем, что $f \in L_2^2$, и в качестве базиса $\{\psi_{j_1,j_2}\}$ выбирается система произведений пар многочленов Лежандра:

$$\psi_{j_1,j_2}(x, y) = \varphi_{j_1}(x)\varphi_{j_2}(y), \quad j_1, j_2 = 0, 1, \ldots$$

Так как в этом случае оценка Ченцова и оценка, построенная при помощи метода моментов, тождественны, то при решении данной задачи в качестве проекционной оценки используется оценка первого вида как менее затратная по вычислениям.

Так как для функций системы Лежандра справедливы следующие равенства:

$$\forall j \neq 0 \ \int_{-1}^{1} \varphi_j(x) dx = 0,$$

$$\forall j \neq 1 \ \int_{-1}^{1} x\varphi_j(x) dx = 0,$$

то после соответствующих упрощений получаем следующее выражение для оценки функции регрессии:

$$\hat{\varphi}(x) = \frac{\sum_{j=0}^{\hat{l}} \left(\varphi_j(x) \sum_{i=1}^{n} y_i \varphi_j(x_i) \right)}{\sum_{j=0}^{\hat{l}} \left(\varphi_j(x) \sum_{i=1}^{n} \varphi_j(x_i) \right)}.$$

Значение $\hat{l}$ находится путём максимизации выражения (2.36) при условии $l_1 = l_2 = l$:

$$\hat{l} = \arg\max_{l} \hat{W}_{l,l}.$$

Из сходимости оценки $\hat{f}(x,y)$ функции плотности вероятности в пространстве $L_2(\Omega)$, вообще говоря, ещё не следует сходимость оценки $\hat{\varphi}(x)$ функции регрессии даже в $L_1(\Omega)$. В связи с этим в ходе метода статистических испытаний могут возникать (и возникают) случаи, когда

$$\|\hat{\varphi}(x) - \varphi(x)\|_{L_1[-1;1]} = +\infty,$$

обусловленные наличием вертикальных асимптот у функции $\hat{\varphi}(x)$. Однако, используя имеющуюся априорную информацию о носителе случайной величины Y, этого можно избежать следующим образом. Если значение оценки больше 1, то считаем, что значение оценки равно 1; если значение оценки меньше -1, то считаем, что оценка равна -1. Таким образом, определяется усечённая оценка функции регрессии:

$$\bar{\varphi}(x) = \begin{cases} 1, & \hat{\varphi}(x) > 1 \\ \hat{\varphi}(x), & -1 \le \hat{\varphi}(x) \le 1 \\ -1, & \hat{\varphi}(x) < -1 \end{cases},$$

от которой затем находится значение функционала качества (3.1). Оценку функции регрессии, построенную описанными методом обозначим через $\bar{\varphi}_1(x)$.

На рис. 3.1 приводится пример построения оценки $\bar{\varphi}_1(x)$.

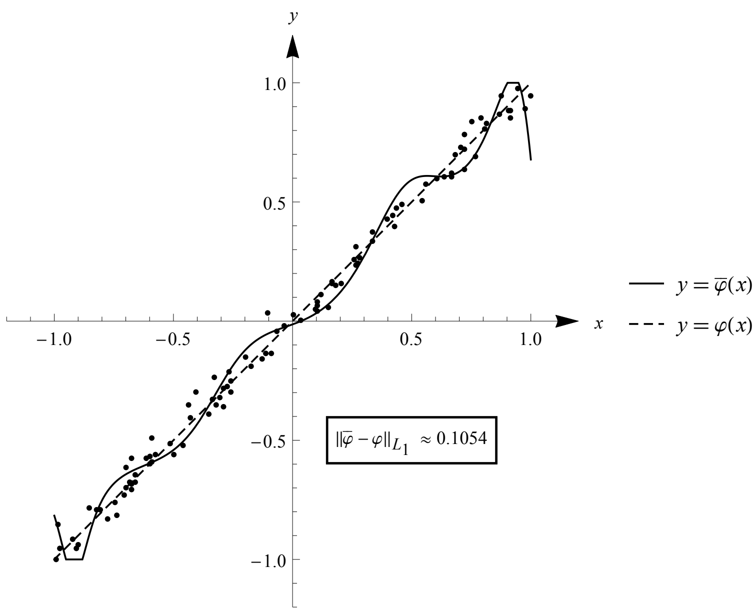

Рис. 3.1. Пример восстановления функции регрессии на основе проекционной оценки функции совместной плотности вероятности случайных величин X и Y.

При решении задачи восстановления функции регрессии с использованием оценки Розенблатта – Парзена возьмём параболическое ядро (см. табл. 1.1), которое является асимптотически оптимальным при объёме выборки $n \to \infty$. Так как из условия считается известным, что случайные величины X и Y имеют одинаковую размерность, то параметры размытости h_1 и h_2 имеют общее значение: $h_1 = h_2 = h$. В этом случае оценка Розенблатта – Парзена совместной плотности вероятности случайных величин X и Y имеет вид:

$$\hat{f}(x, y) = \frac{1}{h^2 n} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) K\left(\frac{y - y_i}{h}\right),$$

оценка функции регрессии:

$$\hat{\varphi}(x) = \frac{\displaystyle\sum_{i=1}^{n} y_i K\left(\frac{x - x_i}{h}\right)}{\displaystyle\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)}.$$

Для настройки параметра размытости h используем два метода: метод минимизации оценки (1.23) функционала $W\{\hat{f}\}$ и метод максимального правдоподобия. Оценка (1.23) в данном случае имеет вид:

$$\hat{W}(h) = \frac{1}{h^2 n^2} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \tau\left(\frac{x - x_i}{h}\right) \tau\left(\frac{y - y_i}{h}\right) -$$

$$- \frac{2}{h^2 n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} K\left(\frac{x - x_i}{h}\right) K\left(\frac{y - y_i}{h}\right),$$

где

$$\tau(z) = \begin{cases} \frac{3}{160}(32 - 40z^2 + 20|z|^3 - |z|^5), & d \in [-2; 2] \\ 0 & \text{иначе} \end{cases}. \qquad (3.2)$$

Для численной минимизации $\hat{W}(h)$ можно использовать различные методы, причём, как показано в [155], наиболее эффективным по скорости вычислений является метод Пауэлла. Оценку функции регрессии, основанную на оценке Розенблатта – Парзена, в которой параметр размытости рассчитывается по формуле (1.24), обозначим через $\bar{\varphi}_2(x)$.

Пример построения оценки $\bar{\varphi}_2(x)$ приводится на рис. 3.2.

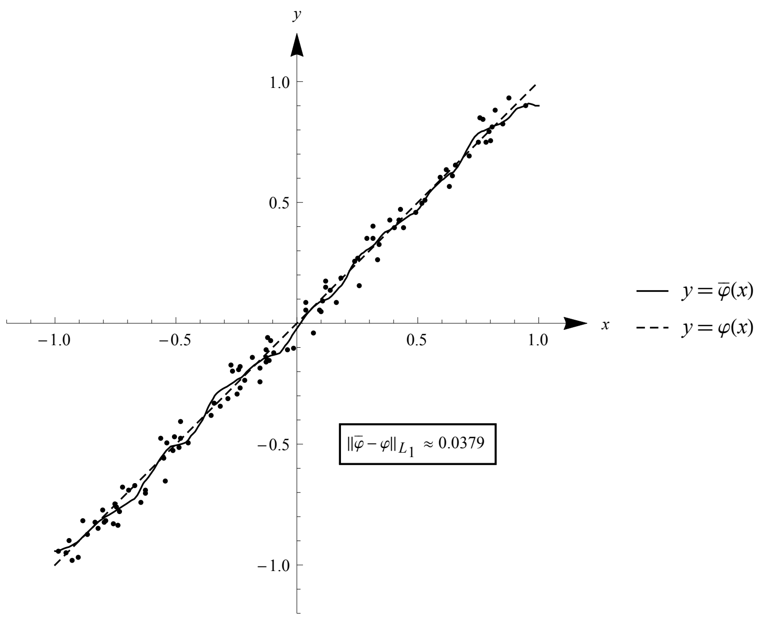

Рис. 3.2. Пример восстановления функции регрессии на основе оценки Розенблатта – Парзена с параметром размытости, рассчитанным по формуле (1.24).

Выражение для логарифма функции правдоподобия в данном случае имеет вид:

$$\ln L(h) = \sum_{i=1}^{n} \ln \sum_{\substack{j=1 \\ j \neq i}}^{n} K\left(\frac{x_i - x_j}{h}\right) K\left(\frac{y_i - y_j}{h}\right) - n \ln((n-1)h^2). \qquad (3.3)$$

При этом область определения функции (3.3) определяется следующим неравенством:

$$h > \max_{1\le i\le n} \min_{\substack{1\le j\le n\\ j\ne i}} \max\{|x_i - x_j|, |y_i - y_j|\}.$$

Оценку функции регрессии, основанную на оценке Розенблатта – Парзена, в которой параметр размытости рассчитывается методом максимального правдоподобия, обозначим через $\bar{\varphi}_3(x)$.

Пример построения оценки $\bar{\varphi}_3(x)$ приводится на рис. 3.3.

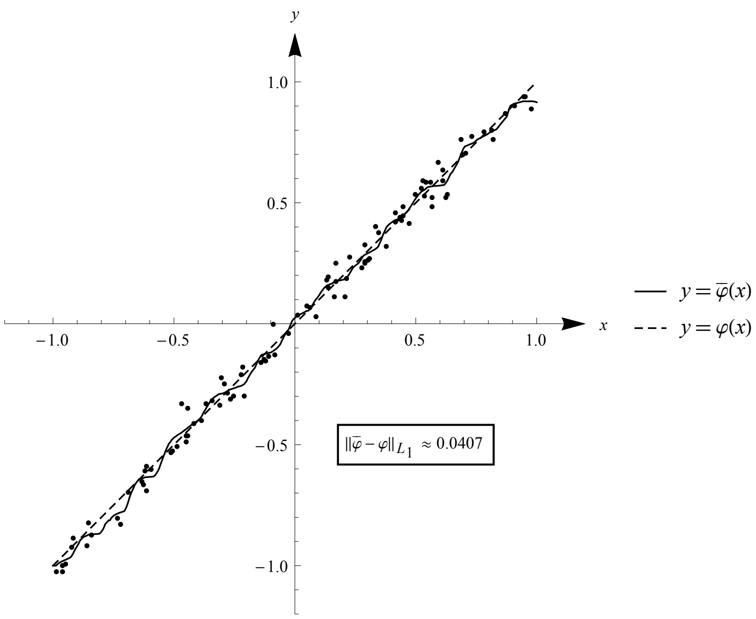

Рис. 3.3. Пример восстановления функции регрессии на основе оценки Розенблатта – Парзена с параметром размытости, рассчитанным методом максимального правдоподобия.

Результаты расчёта значений функционала (3.1) для оценок $\bar{\varphi}_1(x)$, $\bar{\varphi}_2(x)$, $\bar{\varphi}_3(x)$ занесены в табл. 3.1.

Таблица 3.1. Результаты восстановления функции регрессии

Обозначение	**Оценка плотности**	**Оценка параметров**	$Q\{\hat{I}\}$
$\bar{\varphi}_1$	проекционная (1.8), базис Лежандра	формулы (1.11) и (2.22)	0.1101 ± 0.0054
$\bar{\varphi}_2$	Розенблатта – Парзена (1.16), параболическое ядро	формула (1.24)	$\mathbf{0.03896 \pm 0.00074}$
$\bar{\varphi}_3$	Розенблатта – Парзена (1.16), параболическое ядро	формула (1.25)	$\mathbf{0.03906 \pm 0.00078}$

Таким образом, численный расчёт показал, что алгоритмы восстановления функции регрессии на основе оценки Розенблатта – Парзена успешнее справляются с задачей по сравнению с алгоритмами на основе проекционной оценки. При этом выбор метода настройки параметра размытости оказался несущественным.

§ 3.2. Классификация

Другим важным приложением алгоритмов восстановления функции плотности вероятности является задача классификации и распознавания образов. Разработка эффективных систем распознавания объектов различной природы: зрительных образов, речи, явлений, ситуаций и т.п. – представляет собой задачу исключительной важности [65]. Методы распознавания в настоящее время широко применяются в технике, медицине, сельском хозяйстве и других областях знания. Различные методы классификации и распознавания образов рассматриваются в таких работах как [47, 61, 65, 92, 124].

В данном параграфе рассматривается постановка задачи классификации с обучением, взятая из [124].

Пусть имеется $N = 2$ класса объектов из пространства $\Omega \subseteq \mathbb{R}^k$, называемого пространством информативных признаков, и дана обучающая выборка:

$$\mathbf{x}_1 = (x_{11}, \ldots, x_{1k}), \ldots, \mathbf{x}_n = (x_{n1}, \ldots, x_{nk}),$$

такая что для каждого объекта $\mathbf{x}_i$ известно, какому классу о принадлежит. Требуется сформулировать решающее правило, по которому каждый элемента из Ω можно отнести к определённому классу.

Для решения этой задачи в каждом векторе $\mathbf{x}_i$ введём дополнительную переменную y, равную 1, если $\mathbf{x}_i$ принадлежит 1-му классу, и равную -1, если $\mathbf{x}_i$ принадлежит 2-му классу. Решающую функцию представляем в виде функции регрессии случайной величины Y на случайный вектор $(X_1, \ldots, X_k)$:

$$\eta(x_1, \ldots, x_k) = M\{Y \mid X_1 = x_1, \ldots, X_k = x_k\} = \int_{\mathbb{R}} y f_{Y|X}(x_1, \ldots, x_k, y) dy,$$

где $f_{Y|X}(x_1, \ldots, x_k, y)$ – условная плотность вероятности случайной величины Y при условии $(X_1, \ldots, X_n) = (x_1, \ldots, x_n)$. Оценку решающей функции получим, заменив в последнем выражении функцию $f_{Y|X}$ её непараметрической оценкой

$\hat{f}_{Y|X}$, построенной по выборке

$$(x_{i1},\ldots,x_{ik},y_i),\quad i=1,\ldots,n:$$

$$\hat{\eta}(x_1,\ldots,x_k)=\int_{\mathbb{R}} y\hat{f}_{Y|X}(x_1,\ldots,x_k,y)dy.$$

Таким образом, задача классификации сводится к задаче оценивания функции регрессии. Решающее правило при этом имеет вид:

1) если $\hat{\eta}(x_1,\ldots,x_k)>0$, то объект $(x_1,\ldots,x_k)$ относим к 1-му классу;

2) если $\hat{\eta}(x_1,\ldots,x_k)<0$, то объект $(x_1,\ldots,x_k)$ относим ко 2-му классу.

Качество решения задачи классификации определим следующим образом. Пусть

$$S_1=\{\mathbf{x}\in\mathbb{R}^k \mid \eta(\mathbf{x})>0\}\quad \text{и}\quad S_2=\{\mathbf{x}\in\mathbb{R}^k \mid \eta(\mathbf{x})<0\}$$

есть области принятия решения о включении объекта $\mathbf{x}$ в 1-й и 2-й класс соответственно при истинной решающей функции $\eta(\mathbf{x})$, и

$$\hat{S}_1=\{\mathbf{x}\in\mathbb{R}^k \mid \hat{\eta}(\mathbf{x})>0\}\quad \text{и}\quad \hat{S}_2=\{\mathbf{x}\in\mathbb{R}^k \mid \hat{\eta}(\mathbf{x})<0\}$$

есть соответствующие области при полученной оценке решающей функции $\hat{\eta}(\mathbf{x})$. Тогда критерий качества определим в виде функционала

$$Q\{\hat{\eta}\}=M\{\mu(S_1\cap\hat{S}_1)+\mu(S_2\cap\hat{S}_2)\}.$$

Очевидно, чем больше значение Q, тем точнее решение задачи классификации.

Задача классификация в общем случае при числе классов $N>2$ несколько усложняется. Предполагается, что сами классы не отличаются один от другого, т.е. задача не изменяется при переименовании классов. Из этого следует, что если классы соотнести с точками некоторого метрического пространства (по аналогии с точками $y=1$, $y=-1$ в случае $N=2$), то расстояния между любыми двумя классами должны быть одинаковыми. Этого можно добиться, взяв арифметическое пространство размерности $N-1$, а в качестве N точек – вершины правильного $(N-1)$-мерного симплекса. Пусть $A_1(a_{11},\ldots,a_{1,N-1})$,

$\ldots,\ A_N(a_{N1},\ldots,a_{N,N-1})$ – выбранные точки, соответствующие N классам. По обучающей выборке

$$(x_{i1},\ldots,x_{ik},y_{i1},\ldots,y_{ik}),\quad i=1,\ldots,n,$$

где $(x_{i1},\ldots,x_{ik}) \in \Omega$, $(y_{i1},\ldots,y_{ik}) \in \{A_1,...,A_N\}$, построим решающую функцию в виде оценки функции регрессии вектора $(Y_1,\ldots,Y_{N-1})$ на вектор $X_1\ldots X_k$:

$$\hat{\eta}(x_1,\ldots x_k)=\int\limits_{\mathbb{R}^{N-1}}\!\!\cdots\!\int(y_1,\ldots,y_{N-1})\hat{f}_{Y|X}(x_1,\ldots x_k,y_1,\ldots,y_{N-1})dy_1\ldots dy_{N-1}.$$

В этом случае значением решающей функции является $(N-1)$-мерный вектор. Решающее правило при этом формулируется следующим образом: считается, что объект $\mathbf{x}=(x_1,\ldots x_k)$ принадлежит классу j тогда и только тогда, когда расстояние $\|\hat{\eta}(\mathbf{x})-A_j\|$ является наименьшим среди всех расстояний $\|\hat{\eta}(\mathbf{x})-A_i\|$:

$$\|\hat{\eta}(\mathbf{x})-A_j\|=\min_{1\le i\le N}\|\hat{\eta}(\mathbf{x})-A_i\|,\quad j=1,\ldots N. \tag{3.4}$$

Равенства (3.4) определяют N областей $\hat{S}_1,\ \ldots,\ \hat{S}_N$, дающих в объединении множество Ω, причём если $i\neq j$, то $\mu(\hat{S}_i\cap\hat{S}_j)=0$. Через $S_1,\ \ldots,\ S_N$ обозначим соответствующие области при истинной решающей функции. Близость набора классов $\hat{S}=(\hat{S}_1,\ldots,\hat{S}_N)$, оценённых по обучающей выборке, к набору $S=S_1\ldots,S_N$ определим следующим образом:

$$\rho(\hat{S},S)=\sum_{i=1}^{N}\mu(S_i\cap\hat{S}_i).$$

Тогда критерий качества решения задачи определим как среднюю близость классов, получаемых в результате оценивания, к истинным классам:

$$Q\{\hat{\eta}\}=M\left\{\rho(\hat{S},S)\right\}. \tag{3.5}$$

Качество классификации рассчитывалось при решении следующей тестовой задачи. В пространстве $\Omega=[-1;1]\times[-1;1]$ информативных признаков имеется $N=2$ класса объектов. Объекты первого класса генерируются при помощи двумерного нормального распределения с параметрами

$$\mu_1 = 1/2,\ \mu_2 = -1/2,\ \sigma_1 = \sigma_2 = 1/2,\ r = 0,$$

усечённого на квадрат $[-1;1] \times [-1;1]$, объекты второго класса – при помощи аналогичного распределения с параметрами

$$\mu_1 = -1/2,\ \mu_2 = 1/2,\ \sigma_1 = \sigma_2 = 1/2,\ r = 0.$$

Каждое из двух распределений выбирается равновероятно. На рис. 3.4 представлена случайная выборка объёма $n = 50$ объектов 1-го и 2-го классов. Пунктиром показана прямая, разделяющая области S_1 и S_2. В данном случае она имеет уравнение $x_2 = x_1$.

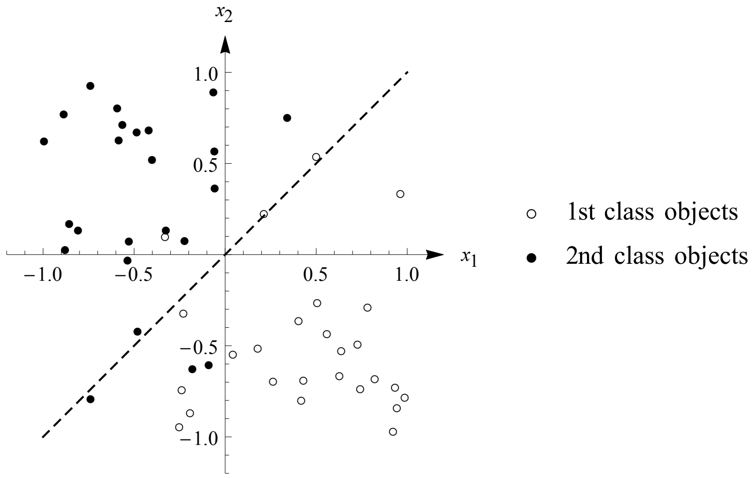

Рис. 3.4. Пример исходных данных для задачи классификации.

Для решения данной задачи построим обучающую выборку

$$(x_{i1}, x_{i2}, y_i), \quad i = 1, \ldots n,$$

где y_i принимает значения 1 или -1 в зависимости от того, 1-му или 2-му классу принадлежит объект $\mathbf{x}_i = (x_{i1}, x_{i2})$. По этой обучающей выборке построим решающую функцию в виде оценки функции регрессии случайной величины Y на случайный вектор (X_1, X_2). Для построения оценки функции регрессии используем проекционную оценку и оценку Розенблатта – Парзена функции плотности вероятности случайного вектора (X_1, X_2, Y). Априори считаем известным, что

1) $\operatorname{supp}(X_1, X_2) \subseteq [-1;1] \times [-1;1]$;

2) величины X_1 и X_2 имеют одинаковую размерность,

3) $\operatorname{supp} Y = \{-1, 1\}$;

4) $\operatorname*{ess\,sup}_{x_1, x_2} f(x_1, x_2) < +\infty$.

При использовании проекционной оценки с базисом Лежандра $\{\varphi_j\}$ решающая функция имеет вид (после упрощений):

$$\hat{\eta}(x_1, x_2) = \frac{\sum_{j_1=0}^{l} \sum_{j_2=0}^{l} \left(\varphi_{j_1}(x_1)\varphi_{j_2}(x_2) \sum_{i=1}^{n} y_i \varphi_{j_1}(x_{i1})\varphi_{j_2}(x_{i2}) \right)}{\sum_{j_1=0}^{l} \sum_{j_2=0}^{l} \left(\varphi_{j_1}(x_1)\varphi_{j_2}(x_2) \sum_{i=1}^{n} \varphi_{j_1}(x_{i1})\varphi_{j_2}(x_{i2}) \right)}.$$

Длина ряда l оценивается путём минимизации оценки функционала $W\{\hat{f}\}$, где $\hat{f} = \hat{f}(x_1, x_2)$ – проекционная оценка плотности вероятности случайного вектора (X_1, X_2), построенная по выборке (x_{11}, x_{12}), $\ldots$, (x_{n1}, x_{n2}). Полученную таким образом оценку решающей функции обозначим через $\hat{\eta}_1$. Пример решения задачи классификации в этом случае приведён на рис. 3.5.

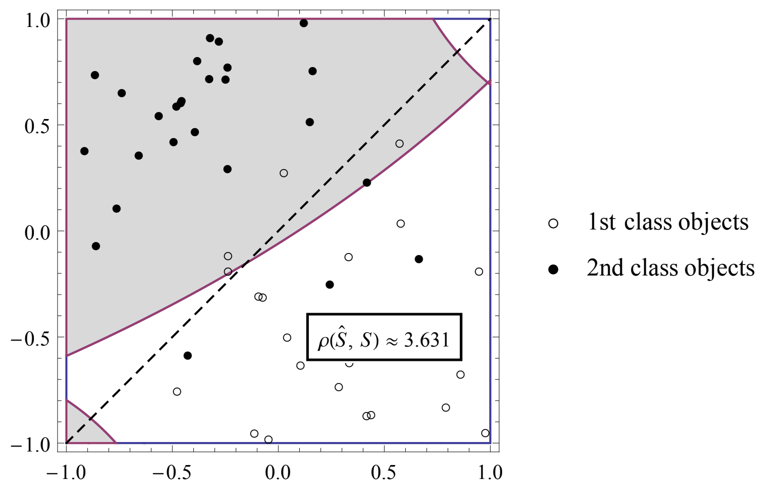

Рис. 3.5. Пример решения задачи классификации с использованием проекционной оценки.

При использовании оценки Розенблатта – Парзена решающая функция имеет вид:

$$\hat{\eta}(x_1, x_2) = \frac{\sum_{i=1}^{n} y_i K\left(\frac{x_1 - x_{i1}}{h}\right) K\left(\frac{x_2 - x_{i2}}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x_1 - x_{i1}}{h}\right) K\left(\frac{x_2 - x_{i2}}{h}\right)},$$

где K – параболическое ядро, параметр размытости h настраивается методом минимизации оценки (1.23), либо методом максимального правдоподобия по выборке (x_{11}, x_{12}), $\ldots$, (x_{n1}, x_{n2}).

Оценка (1.23) в данном случае имеет вид:

$$\hat{W}(h) = \frac{1}{h^3 n^2} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \tau\left(\frac{x_{i_1,1} - x_{i_2,1}}{h}\right) \tau\left(\frac{x_{i_1,2} - x_{i_2,2}}{h}\right) - \\ - \frac{2}{h^3 n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} K\left(\frac{x_{i_1,1} - x_{i_2,1}}{h}\right) K\left(\frac{x_{i_1,2} - x_{i_2,2}}{h}\right), \quad (3.6)$$

где $\tau(z)$ определяется по формуле (3.2). Для нахождения значения h, минимизирующего оценку $\hat{W}(h)$, используется метод золотого сечения. Полученную таким образом оценку решающей функции обозначим через $\hat{\eta}_2$. Пример решения задачи классификации в этом случае приведён на рис. 3.6.

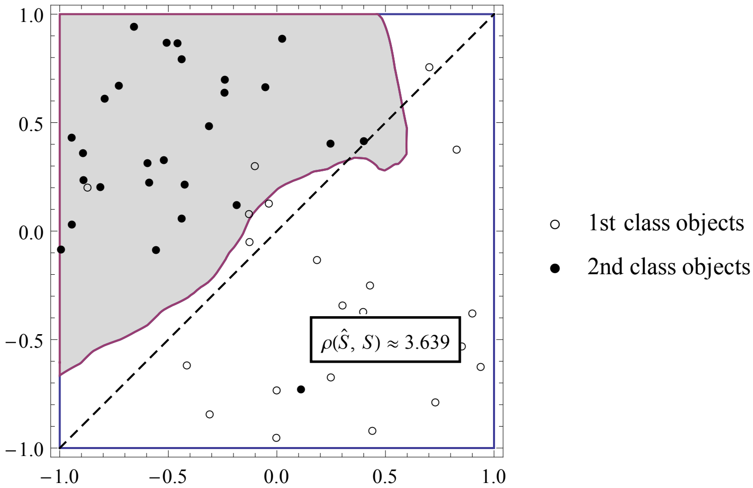

Рис. 3.6. Пример решения задачи классификации с использованием Розенблатта – Парзена, в которой $\hat{h} = \arg\min_h \hat{W}(h)$.

Логарифм функции правдоподобия (1.26) в данном случае имеет вид:

$$\ln L(h) = \\ = \sum_{i_1=1}^{n} \ln \sum_{\substack{i_2=1 \\ i_2 \neq i_1}}^{n} K\left(\frac{x_{i_1,1} - x_{i_2,1}}{h}\right) K\left(\frac{x_{i_1,2} - x_{i_2,2}}{h}\right) - n \ln((n-1)h^2). \quad (3.7)$$

Для нахождения значения h, минимизирующего (3.7), также используется метод золотого сечения. Полученную таким образом оценку решающей функции обозначим через $\hat{\eta}_3$. Пример решения задачи классификации в этом случае приведён на рис. 3.7.

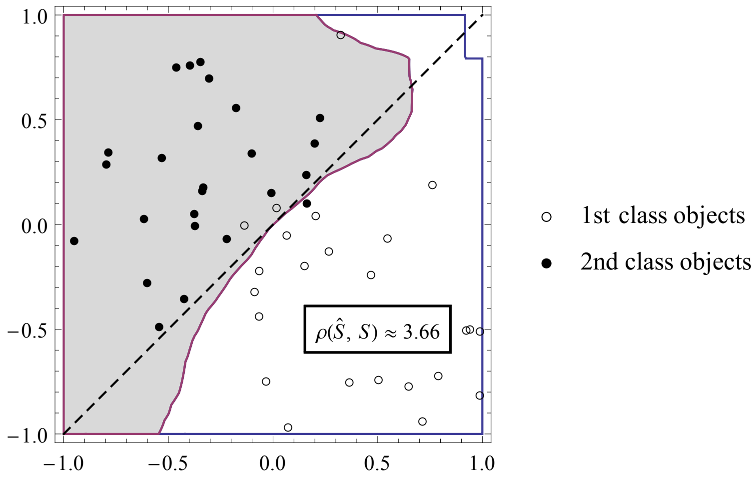

Рис. 3.7. Пример решения задачи классификации с использованием Розенблатта – Парзена, в которой $\hat{h} = \arg\min_h \ln L(h)$.

Результаты расчёта значений функционала (3.5) для оценок $\hat{\eta}_1$, $\hat{\eta}_2$, $\hat{\eta}_3$ методом статистических испытаний занесены в табл. 3.2.

Таблица 3.2. Результаты решения задачи классификации

Обозначение	**Оценка плотности**	**Оценка параметров**	$Q\{\hat{I}\}$
$\hat{\eta}_1$	проекционная (1.8), базис Лежандра	формулы (1.11) и (2.22)	3.55 ± 0.04
$\hat{\eta}_2$	Розенблатта – Парзена (1.16), параболическое ядро	формула (1.24)	$\mathbf{3.56 \pm 0.02}$
$\hat{\eta}_3$	Розенблатта – Парзена (1.16), параболическое ядро	формула (1.25)	3.51 ± 0.02

Результаты расчётов позволяют сделать вывод о том, что небольшое улучшение в качестве решения задачи классификации достигается при использовании оценки Розенблатта – Парзена, в которой параметр размытости настраивается методом минимизации оценки функционала качества.

§ 3.3. Оценивание количества информации

Рассмотрим задачу оценивания количества информации, содержащейся в непрерывном случайном векторе $\mathbf{X} = (X_1, \ldots, X_{k_1})$ о случайном векторе $\mathbf{Y} = (Y_1, \ldots, Y_{k_2})$. Данная задача рассматривается в [124], и является актуальной в теории управления и в теории сигналов. Фундаментальные результаты по непараметрическому оцениванию функционалов от плотностей с использованием ядерных оценок содержатся в монографии [72].

Пусть имеется две случайные величины X и Y, такие что случайный вектор (X, Y) имеет плотность $f(x, y)$. Маргинальные функции плотности вероятности случайных величин X и Y обозначим через $f_X(x)$ и $f_Y(y)$ соответственно.

Дифференциальной энтропией случайной величины называется число, противоположное математическому ожиданию от логарифма функции плотности вероятности этой случайной величины:

$$H\{X\} = -M\{\ln f_X(X)\} = -\int_{\mathbb{R}} f_X(x) \ln f_X(x) dx;$$

$$H\{Y\} = -M\{\ln f_Y(Y)\} = -\int_{\mathbb{R}} f_Y(x) \ln f_Y(y) dy.$$

Аналогично, совместная энтропия случайных величин X и Y определяется как энтропия случайного вектора (X, Y):

$$H\{X, Y\} = -M\{\ln f(X, Y)\} = -\iint_{\mathbb{R}^{\nvDash}} f(x, y) \ln f(x, y) dy.$$

Условной энтропией случайной величины Y при условии $X = x$ называется следующий функционал:

$$H\{Y \mid X = x\} = -\int_{\mathbb{R}} f_{Y|X}(x, y) \ln f_{Y|X}(x, y) dy,$$

где $f_{Y|X}(x, y)$ – условная плотность случайной величины Y:

$$f_{Y|X}(x, y) = \frac{f(x, y)}{f_X(x)}.$$

Условная энтропия представляет собой функцию одной переменной x. Рассмотрим случайную величину ξ, подчинённую тому же закону распределения, что и случайная величина X. Тогда выражение $H\{Y \mid X = \xi\}$ определяет некоторую случайную величину, математическое ожидание которой называется средней условной энтропией случайной величины Y относительно X:

$$H\{Y \mid X\} = M\{H\{Y \mid X = \xi\}\}.$$

Как известно, средняя условная энтропия выражается через безусловную:

$$\begin{aligned} H\{Y \mid X\} &= \int_{\mathbb{R}} H\{Y \mid X = x\} f_X(x) dx = \\ &= -\iint_{\mathbb{R}^2} \ln f_{Y|X}(x, y) f_{Y|X}(x, y) f_X(x) dx dy = -\iint_{\mathbb{R}^2} \ln \frac{f(x, y)}{f_X(x)} f(x, y) dx dy = \\ &= -\iint_{\mathbb{R}^2} \ln f(x, y) f(x, y) dx dy + \iint_{\mathbb{R}^2} \ln f_X(x) f(x, y) dx dy. \end{aligned}$$

Таким образом,

$$H\{Y \mid X\} = H\{X, Y\} - H\{X\}. \tag{3.8}$$

Количеством информации, содержащейся в случайной величине X о случайной величине Y, называется разность между безусловной энтропией случайной величины Y и её условной энтропией относительно случайной величины X:

$$I\{Y \mid X\} = H\{Y\} - H\{Y \mid X\}.$$

Или, с учётом (3.8),

$$I\{Y \mid X\} = H\{X\} + H\{Y\} - H\{X, Y\}.$$

Очевидно, количество информации $I\{Y \mid X\}$ совпадает с количеством информации $I\{X \mid Y\}$.

Если $\mathbf{X} = (X_1, \ldots, X_{k_1})$ и $\mathbf{Y} = (Y_1, \ldots, Y_{k_2})$ – непрерывные случайные векторы, такие что случайный вектор

$$(\mathbf{X}, \mathbf{Y}) = (X_1, \ldots, X_{k_1}, Y_1, \ldots, Y_{k_2}),$$

составленный из них, имеет плотность

$$f(\mathbf{x}, \mathbf{y}) = f(x_1, \ldots, x_{k_1}, y_1, \ldots, y_{k_2}),$$

то получаем аналогично:

$$H\{\mathbf{X}\} = -M\{\ln f_{\mathbf{X}}(\mathbf{X})\};\ H\{\mathbf{Y}\} = -M\{\ln f_{\mathbf{Y}}(\mathbf{Y})\};$$

$$H\{\mathbf{Y} \mid \mathbf{X}\} = H\{\mathbf{X}, \mathbf{Y}\} - H\{\mathbf{X}\};$$

$$I\{\mathbf{Y} \mid \mathbf{X}\} = H\{\mathbf{X}\} + H\{\mathbf{Y}\} - H\{\mathbf{X}, \mathbf{Y}\}.$$

Задача оценивания количества информации возникает, например, в теории связи при определении эффективности работы системы связи [113].

Пусть имеется независимая выборка (x_i, y_i), $i = 1, \ldots, n$, значений непрерывного случайного вектора (X, Y). Требуется оценить количество информации $I\{Y \mid X\}$.

Оценку энтропии $\hat{H}\{X\}$ будем искать в виде [124]

$$\hat{H}\{X\} = -\hat{M}\{\ln \hat{f}_X(x)\} = -\frac{1}{n}\sum_{i=1}^{n} \ln \hat{f}_i(x_i),$$

где $\hat{f}_i(x)$ – оценка функции плотности вероятности, построенная по всей выборке $x_1, \ldots, x_n$, за исключением точки x_i.

Оценку количества информации соответственно:

$$\hat{I}\{Y \mid X\} = \hat{H}\{X\} + \hat{H}\{Y\} - \hat{H}\{X, Y\}. \tag{3.9}$$

Качество оценивания количества информации можно определить как среднее отклонение от истинного количества информации:

$$Q\{\hat{I}\} = M\{|\hat{I} - I|\}. \tag{3.10}$$

В качестве тестового закона совместного распределения случайных величин был взят двумерный нормальный закон с параметрами $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$ и произвольным коэффициентом корреляции r. Маргинальные и совместная энтропии компонент X и Y в этом случае:

$$H\{X\} = H\{Y\} = \frac{1}{2}(1 + \ln 2\pi);$$

$$H\{X, Y\} = 1 + \ln 2\pi + \ln\sqrt{1 - r^2}.$$

Следовательно, количество информации

$$I\{Y \mid X\} = I\{X \mid Y\} = -\ln\sqrt{1 - r^2}.$$

Оценку количества информации $\hat{I}\{Y \mid X\}$ будем искать в виде (3.9), где в качестве оценки совместной плотности $\hat{f}(x, y)$ берётся проекционная оценка или оценка Розенблатта – Парзена. В зависимости от метода настройки параметров выбранных оценок плотности вводится четыре вида оценок количества информации (см. табл. 3.3).

Для каждой из этих оценок рассчитывалось значение функционала качества (3.10). Результаты приведены в табл. 3.3.

Таблица 3.3. Результаты оценивания количества информации

Обозначение	**Оценка плотности**	**Оценка параметров**	$Q\{\hat{I}\}$
$\hat{I}_1$	проекционная (1.8), базис Эрмита	формулы (1.11) и (2.22)	0.143 ± 0.01
$\hat{I}_2$	проекционная (1.8), базис Эрмита	формулы (2.11) и (2.26)	0.135 ± 0.01
$\hat{I}_3$	Розенблатта – Парзена (1.16), ядро Гаусса	формула (1.24)	0.10 ± 0.01
$\hat{I}_4$	Розенблатта – Парзена (1.16), ядро Гаусса	формула (1.25)	$\mathbf{0.093 \pm 0.003}$

Как показали расчёты, наилучшая точность при оценивании количества информации достигается при использовании оценки Розенблатта – Парзена (1.16), в которой параметр размытости h настраивается с помощью максимизации функции правдоподобия (1.25).

Выводы

В данной главе решена задача сравнения предложенных методов настройки проекционной оценки с имеющимися, а также с другими видами непараметрических оценок при решении прикладных задач. В ходе исследования было установлено, что оценка Розенблатта – Парзена является более предпочтительной при решении рассмотренных тестовых задач. Вместе с тем, при настройке проекционной оценки предложенный метод оказался более предпочтительным по сравнению с традиционным.

Заключение

В ходе выполнения диссертационной работы были получены следующие результаты:

– показано, что весовое гильбертово пространство $L_{2,w}(\Omega)$ может быть использовано для построения проекционной оценки любой функции плотности вероятности (предл. 2.4);

– найден критерий на весовую функцию $w(\mathbf{x})$ для расширения пространства $L_2(\Omega)$ до пространства $L_{2,w}(\Omega)$, которое содержит более широкое множество функций плотности вероятности (теорема 2.4);

– предложен способ построения весовой функции $w(\mathbf{x})$, при котором соответствующее расширение $L_{2,w}(\Omega)$ пространства $L_{2,w}$ содержит оцениваемую функцию плотности вероятности $f(\mathbf{x})$ (формула (2.9));

– предложен новый метод настройки коэффициентов проекционной оценки функции плотности вероятности случайного вектора, являющийся обобщением метода моментов (формула (2.13));

– доказано, что при определённых условиях частным случаем предложенного обобщения является традиционный метод оценивания коэффициентов (теорема 2.5);

– предложен новый метод оценивания длины ряда проекционной оценки, в которой коэффициенты настраиваются методом моментов или его обобщением (формула (2.25));

– экспериментально установлено, что на малых выборках обобщение метода моментов позволяет повысить эффективность проекционной оценки (табл. 2.5, 2.8);

– экспериментально установлено, что для прикладных задач (восстановление функции регрессии, классификация, оценка количества информации) более предпочтительной является оценка Розенблатта – Парзена.

Также было экспериментально установлено, что условиях малых выборок метод моментов является более предпочтительным при настройке проекционной оценки. В тех случаях, когда нет возможности использовать ядерные оценки (например, ограниченные вычислительные ресурсы), целесообразно использовать проекционную оценку, так как она не содержит всю исследуемую выборку и допускает лаконичное математическое выражение. При этом для настройки длины ряда l рекомендуется использовать предложенный подход.

Используемый метод сравнения алгоритмов восстановления плотности вероятности и полученные численные результаты могут быть также использованы при сравнении эффективности любых непараметрических оценок функции плотности вероятности.

Список литературы

1. Bartlett, M. S. Statistical estimation of density functions // The Indian Journal of Statistics. Series A. – 1963. – Vol. 25, no. 3. – P. 245–254.

2. Bauer, H. Probability Theory and elements of Measure Theory. London: Academic Press, 1981.

3. Bickel, P. J. On some global measures of deviation of density function estimates / P. J. Bickel, M. Rosenblatt // The Annals of Statistics. – 1973. Vol. 1, No. 6. – P. 1071–1095.

4. Davis, Kathryn B. Mean square error properties of density estimates // The Annals of Statistics. – 1975. – Vol. 3, no. 4. – P. 1025–1030.

5. Devroye, L. P. The strong uniform consistency of nearest neighbor density estimates / L. P. Devroye, T. J. Wagner // The Annals of Statistics. – 1977. – Vol. 5, no. 3. – P. 536–540.

6. Dudley, R. Probabilities and Metrics. – Aarhus: Aarhus University, 1976.

7. Efromovich, S. Orthogonal series density estimation // WIREs Computational Statistics. – 2010. – No. 2. – P. 467 – 476.

8. Ghosh, M. Nonparametric sequential Bayes estimation of the distribution function / M. Ghosh, Bh. Mukherjee // Sequential Analysis. – 2005. – Vol. 24. – P. 389–409.

9. Heinhold, J. Ingenieur-Statistik / J. Heinhold, K.-W. Gaede. – München; Wien: Springler Verlag, 1964. – 352 p.

10. Ito K. Introduction to Probability Theory. – Cambridge: Cambridge University Press, 1984.

11. Kingman, J. Introduction to Measure and Probability // J. Kingman, S. Taylor. – Cambridge: Cambridge University Press, 1966.

12. Laha, R. Probability Theory // R. Laha, V. Rogatgi. – New York: John Wiley and Sons, 1979.

13. Loftsgaarden, D. O. A nonparametric estimate of a multivariate density function / D. O. Loftsgaarden, C. P. Quessenberry // The Annals of Mathematical Statistics. – 1965. – Vol. 36, no. 3. – P. 1049–1051.

14. Mann, H. B. On the choice of number of intervals in the application of the

chi-square test / H. B. Mann, A. Wald // The Annals of Mathematical Statistics. – 1942. – Vol. 18. – P. 50–54.

15. Meyer, T. G. Bounds for estimation of density functions and their derivatives // The Annals of Statistics. – 1977. – Vol. 5, no. 1. – P. 136–142.

16. De Montricher, G. F. Nonparametric maximum likelihood estimation of probability densities by penalty function methods / G. F. De Montricher, R. A. Tapia, J. R. Thompson // The Annals of Statistics. – 1975. – Vol. 3, no. 6. – P. 1329–1348.

17. Medvedev, A. V. Nonparametric theory of control system design // Вестник СибГАУ. – 2002. – Выпуск 3. – С. 44–55.

18. Mukhopadhyay, N. An overview of sequential nonparametric density estimation // Nonlinear Analysis. – 1997.

19. Murthy, V. K. Estimation of probability density // The Annals of Mathematical Statistics. – 1965. – Vol. 36, no. 3. – P. 1027–1031.

20. Parzen, E. On estimation of a probability density function and mode // The Annals of Mathematical Statistics. – 1962. – Vol. 35, no. 3. – P. 1065–1076.

21. Pickands, J. Efficient estimation of a probability density // The Annals of Mathematical Statistics. – 1969. – Vol. 40, no. 3. – P. 854–864.

22. Pollard, H. The mean convergence of orthogonal series // Duke Mathematical Journal. – 1949. – Vol. 16, no. 1. – P. 189–191.

23. Reiss, R. D. Consistency of a certain class of empirical density functions // Metrika. – 1975. – Vol. 22, no. 4. – P. 189–203.

24. Rosenblatt, M. A quadratic measure of deviation of two-dimensional density estimates and a test of independence // The Annals of Statistics. – 1975. – Vol. 3, no. 1. – P. 1–14.

25. Rosenblatt, M. Remarks on some nonparametric estimates of a density function // The Annals of Mathematical Statistics. – 1956. – Vol. 27, no. 3. – P. 832–837.

26. Schwartz, S. C. Estimation of probability density by an orthogonal series // The Annals of Mathematical Statistics. – 1967. – Vol. 38, no. 4. – P. 1261–1265.

27. Scott, D. W. On optimal and data-based histograms // Biometrika. –

1979. – Vol. 66. – P. 605–610.

28. Schuster, E. F. Estimation of a probability density function and its derivatives // The Annals of Mathematical Statistics. – 1969. – Vol. 40, no. 4.

29. Singh, R. S. Nonparametric estimation of mixed partial derivatives of a multivariate density // Journal of Multivariate Analysis. – 1976. – Vol. 6, no. 1. – P. 111–122.

30. Srivastava, R. C. Estimation of probability density function based on random number of observations with applications // International Statistical Review. – 1974. – Vol. 41, no. 1. – P. 77–86.

31. Sturges, H. A. The choice of a class interval // Journal of the American Statistical Association. – 1926. – Vol. 21. – P. 65–66.

32. Tarasenko, F. P. On evaluation of an unknown probability density function, the direct estimation of entropy from independent observations of a continuous random variable, and the distribution-free entropy test of goodness-of-fit // Proceedings of the IEEE. – 1968. – Vol. 56, no. 11. – P. 2052–2053.

33. Tarter, M. E. An introduction to the implementation and theory of nonparametric density estimation / M. E. Tarter, R. A. Kronmal // The American Statistician. – 1976. – Vol. 30, no. 3. – P. 105–112.

34. Wagner, T. J. Comments on nonparametric estimates of probability density.

35. Wagner, T. J. Nonparametric estimates of probability densities // IEEE Transactions on Information Theory. – 1975. – Vol. 21. – P. 438–440.

36. Wahba, G. Data-based optimal smoothing of orthogonal series density estimates // The Annals of Statistics. – 1981. – Vol. 9, 1. – P. 146–156.

37. Wahba, G. Optimal convergence properties of variance knot, kernel and orthogonal series methods for density functions // The Annals of Statistics. – 1975. – Vol. 3, no. 1. – P. 15–20.

38. Watson, G. Density estimation by orthogonal series // The Annals of Mathematical Statistics. – 1967. – Vol. 40, no. 4. – P. 1496–1498.

39. Watson, G. On the estimation of the probability density / G. Watson, M. Leadbetter // The Annals of Mathematical Statistics. – 1963. – Vol. 32, no. 2. –

P. 480–491.

40. Winter, B. B. Rate of strong consistency of two nonparametric density estimators // The Annals of Statistics. – 1975. – Vol. 3, no. 3. – P. 759–776.

41. Woodroofe, M. On choosing a delta-sequence // The Annals of Mathematical Statistics. – 1970. – Vol. 41, no. 5. – P. 1665–1671.

42. Zhivoglyadov, V. P. On adaptation algorithms for computer control systems / V. P. Zhivoglyadov, A. V. Medvedev, B. M. Mirkin // 5th IFAC world congress. – Paris, 1972. – P. 107–114.

43. Zhivoglyadov, V. P. Stochastic systems dual control and optimization under nonparametric uncertainty conditions / V. P. Zhivoglyadov, A. V. Medvedev // Proceedings of VI IFAC congress on stochastic control. – Hungary, 1974. – P. 209–215.

44. Абрамовиц, М. Справочник по специальным функциям с формулами, графиками и математическими таблицами / Абрамовиц, М., Стиган И.; пер. с англ. В. А. Диткина и Л. Н. Кармазиной – М.: Наука, 1979. – 831 с.

45. Абусев, Р. А. О непараметрических оценках в групповой классификации // Статистические проблемы управления. – Выпуск 27. – Вильнюс, 1978. – С. 91–97.

46. Айвазян, С. А. Прикладная статистика: Основы моделирования и первичная обработка данных / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин. – М.: Финансы и статистика, 1983. – 471 с.

47. Айвазян, С. А. Прикладная статистика: Классификация и снижение размерности / С. А. Айвазян, В. М. Бухштабер, И. С. Енюков, Л. Д. Мешалкин; под ред. С. А. Айвазяна. – М.: Финансы и статистика, 1989. – 607 с.

48. Айду, Ф. А. Оценивание плотности вероятностей на основе метода стохастической регуляризации / Ф. А. Айду, В. Н. Вапник // Автоматика и телемеханика. – 1989. – №4. – С. 84–97.

49. Акимов, С. С. Методы решения задачи восстановления плотности вероятности по выборке из генеральной совокупности // Естественные и математические науки в современном мире. – 2014. – № 1 (13). – С. 29–35.

50. Алексеев, В. Г. О статистических оценках некоторых функционалов от

плотности вероятности // Теория вероятностей и математическая статистика, 1976. – Выпуск 15. – С. 3–9.

51. Алексеев, В. Г. Об оценке плотности вероятности и ее производных // Математические заметки, 1972. – Том 12, №5. – С. 621–626.

52. Алексеева, И. У. Теоретическое и экспериментальное исследование законов распределения погрешностей, их классификация и методы оценки их параметров: Автореф. дис. на соиск. учен. степени канд. техн. наук. – Л.: Ленинградский политехнический институт, 1975. – 20 с.

53. Андерсон, Т. Введение в многомерный статистический анализ. – М.: Физматгиз, 1963.

54. Бари, Н. К. Биортогональные системы и базисы в гильбертовом пространстве // Учёные записки Московского государственного университета. – Том IV. Математика. – №148. – М.: Изд-во Московского университета, 1951. – С. 69–107.

55. Бари, Н. К. Тригонометрические ряды. – М.: Физматгиз, 1961. – 936 с.

56. Бернштейн, С. Теория вероятностей. – 4-е изд. – М.: Гостехиздат, 1946.

57. Боровков, А. А. Теория вероятностей. – 3-е изд. – М.: Эдиториал УРСС, 1999. – 472 с.

58. Вапник, В. Н. Восстановление зависимостей по эмпирическим данным. – М.: Наука, 1971. – 448 с.

59. Вапник, В. Н. Непараметрические методы восстановления плотности вероятностей / В. Н. Вапник, А. Р. Стефанюк // Автоматика и телемеханика. – 1978. – Выпуск 8. – С. 38–52.

60. Вапник, В. Н. О скорости сходимости в L_2 проекционной оценки плотности вероятности / В. Н. Вапник, Н. М. Маркович, А. Р. Стефанюк // Автоматика и телемеханика. – 1992. – Выпуск 5. – С. 64–74.

61. Вапник, В. Н. Теория распознавания образов / В. Н. Вапник, А. Я. Червоненкис. – М.: Наука, 1974. – 416 с.

62. Васильев, В. А. Непараметрическое оценивание функционалов от распределений стационарных последовательностей / В. А. Васильев, А. В. Добро-

видов, Г. М. Кошкин – М.: Наука, 2004. – 508 с.

63. Волков, И. К. Интегральные преобразования и операционное исчисление: Учебник для вузов / И. К. Волков, А. Н. Канатников – 2-е изд. – М.: Издательство МГТУ им. Н. Э. Баумана, 2002. – 228 с.

64. Гнеденко, Б. В. Курс теории вероятностей : Учебник. – 8-е изд. – М.: УРСС, 2005. – 448 с.

65. Горелик, А. Л. Методы распознавания : Учеб. пособие / А. Л. Горелик, В. А. Скрипкин. – 2-е изд. – М.: Высшая школа, 1984. – 208 с.

66. Деврой, Л. Непараметрическое оценивание плотности. L_1-подход / Л. Деврой, Л. Дьёрфи. – М.: Мир, 1988. – 408 с.

67. Дмитриев, Ю. Г. Использование дополнительной информации при непараметрическом оценивании функционалов плотности / Ю. Г. Дмитриев, Г. М. Кошкин // Автоматика и телемеханика. – 1987. – №10. – С. 47–59.

68. Дмитриев, Ю. Г. К вопросу о статистическом оценивании нелинейных функционалов от плотностей вероятности / Ю. Г. Дмитриев, Ф. П. Тарасенко // Труды Сибирского физико-технического института при ТГУ, 1973. – Выпуск 63. – С. 154–168.

69. Дмитриев, Ю. Г. Об использовании априорной информации при оценивании линейных функционалов от распределений / Ю. Г. Дмитриев, Ф. П. Тарасенко // Математическая статистика и её приложения. – Выпуск 4. – Томск: ТГУ, 1976. – С. 52–62.

70. Дмитриев, Ю. Г. Об одном классе непараметрических оценок нелинейных функционалов плотности / Ю. Г. Дмитриев, Ф. П. Тарасенко // Теория вероятностей и её приложения, 1974. – Том 19, № 2. – С. 404–409.

71. Дмитриев, Ю. Г. Об оценивании функционалов от плотности вероятности и ее производных / Ю. Г. Дмитриев, Ф. П. Тарасенко // Теория вероятностей и её приложения, 1973. – Том 18, №3. – С. 662–668.

72. Добровидов, А. В. Непараметрическое оценивание сигналов / А. В. Добровидов, Г. М. Кошкин. – М.: Наука, 1997. – 329 с.

73. Дуда, Р. Разпознавание образов и анализ сцен / Р. Дуда, П. Харт. –

М.: Мир, 1976. – 509 с.

74. Дьяченко, М. И. Мера и интеграл / М. И. Дьяченко, П. Л. Ульянов. – М.: Факториал, 1998. – 160 с.

75. Епанечников, В. А. Непараметрическая оценка многомерной плотности вероятности // Теория вероятностей и её применения. – 1969. – Том 14, № 1. – С. 156–162.

76. Живоглядов, В. П. Непараметрические алгоритмы адаптации. / В. П. Живоглядов, А. В. Медведев. – Фрунзе: Илим, 1974. – 136 с.

77. Иванилов, А. А.. Непараметрическая оценка производной функции регрессии и ее применение к задаче идентификации / А. А. Иванилов, С. Ф. Ковязин // Адаптивные системы и их приложения. – Новосибирск: Наука, 1978. – С. 109–119.

78. Канторович, Л. В. Функциональный анализ / Л. В. Канторович, Г. П. Акилов. – 3-е изд., перераб. – М.: Наука, 1984. – 752 с.

79. Кашин, Б. С. Ортогональные ряды / Кашин Б. С., Саакян А. А. – Изд 2-е, доп. – М.: Изд-во АФЦ, 1999. – 560 с.

80. Кендалл, М. Теория распределений / М. Кендалл, А. Стюарт. – М.: Наука, 1966. – 588 с.

81. Кобзарь, А. И. Прикладная математическая статистика: Для инженеров и научных работников / А. И. Кобзарь. – М.: Физматлит, 2006. – 816 с.

82. Колмогоров, А. Н. Основные понятия теории вероятностей. – 2-е изд. – М.: Наука, 1974. – 120 с.

83. Колмогоров, А. Н. Элементы теории функций и функционального анализа / А. Н. Колмогоров, С. В. Фомин. – 7-е изд. – М.: Физматлит, 2004. – 572 с.

84. Конаков, В. Д. – Непараметрическая оценка плотности распределения вероятностей // Теория вероятностей и её приложения, 1972. – Том 17, № 2. – С 377–379.

85. Кошкин, Г. М. Рекуррентное оценивание плотности вероятности и линии регрессии по зависимой выборке / Г. М. Кошкин, Ф. П. Тарасенко // Ма-

тематическая статистика и её приложения. – Выпуск 4. – Томск: Изд-во ТГУ, 1976. – С. 122–138.

86. Крамер, Г. Математические методы статистики / Г. Крамер; под ред. А. Н. Колмогорова. – 2-е изд., стер. – М.: Мир, 1975. – 648 с.

87. Ламперти, Дж. Вероятность. – М.: Наука, 1973.

88. Лапко, А. В. Непараметрические модели и алгоритмы обработки информации: учебное пособие / А. В. Лапко, В. А. Лапко. – Красноярск: Изд-во СибГАУ, 2010. – 220 с.

89. Лапко, А. В. Дискретизация интервала измерения значений случайной величины на основе результатов оптимизации непараметрической оценки плотности вероятности / А. В. Лапко, В. А. Лапко // Информатика и системы управления. – 2013. – № 4 (38). – С. 63–69.

90. Лапко, А. В. К анализу непараметрических алгоритмов распознавания образов / А. В. Лапко, А. В. Медведев // Статистические проблемы управления. – Выпуск 14. – Вильнюс, 1976. – С. 105–116.

91. Лапко, А. В. К оптимизации некоторых непараметрических оценок / А. В. Лапко, А. В. Медведев, Е. А. Тишина // Применение вычислительных машин в системах управления непрерывным производством. – Фрунзе: Илим, 1975. – С. 93–107.

92. Лапко, А. В. Непараметрические методы классификации и их применение. – Новосибирск: Наука, 1993. – 152.

93. Лемешко, Б. Ю. О выборе числа интервалов в критериях согласия типа χ^2 / Б. Ю. Лемешко, Е. В. Чимитова // Заводская лаборатория. Диагностика материалов. – 2003. – Том 69. – № 1. – С. 61–67.

94. Ленг, С. Алгебра. – М.: Мир, 1968. – 564 с.

95. Лоэв, М. Теория вероятностей / М. Лоэв; пер. с англ. Б. А. Севастьянова – М.: Издательство иностранной литературы, 1962. – 720 с.

96. Люстерник, Л. А. Краткий курс функционального анализа: Учебное пособие / Л. А. Люстерник, В. И. Соболев. – 2-е изд. – СПб.: Лань, 2009. – 272 с.

97. Мания, Г. М. Статистическое оценивание распределения вероятностей.

– Тбилиси: Изд-во Тбилисского университета, 1974. – 238 с.

98. Медведев, А. В. Адаптивные непараметрические системы: препринт. – Красноярск: ВЦ СО АН СССР, 1979. – 54 с.

99. Медведев, А. В. К непараметрической оценке многомерной плотности вероятности // Исследование и оптимизация стохастических распределённых систем. – Фрунзе: Илим, 1971. – С. 101–107.

100. Медведев, А. В. Непараметрические оценки плотности вероятности и ее производных // Автоматизация промышленного эксперимента. – Фрунзе: Илим, 1973. – С. 22–31.

101. Медведев, А. В. Непараметрические системы адаптации. – Новосибирск: Наука, 1983. – 174 с.

102. Медведев, А. В. Непараметрические системы обучения и адаптации: препринт. – Красноярск: ВЦ СО АН СССР, 1981. – 72 с.

103. Медведев, А. В. О сходимости непараметрических алгоритмов управления // Известия АН Киргизской ССР. – 1975. – Выпуск 1. – С. 27–32.

104. Медведев, А. В. Теория непараметрических систем. Активные процессы – I // Вестник СибГАУ. – 2011. – № 4 (37). – С. 52–57.

105. Медведев, А. В. Теория непараметрических систем. Моделирование // Вестник СибГАУ. – 2010. – № 4 (30). – С. 4–9.

106. Меламед, И. А. Об интегральной среднеквадратичной ошибке некоторых многомерных непараметрических оценок плотности вероятности // Сообщения АН ГССР. – 1973. – Том 71, №2. – С. 293–296.

107. Моторный, В. П. О сходимости в среднем рядов Фурье по многочленам Лежандра // Известия АН СССР. Серия математическая. – 1973. – Том 37, выпуск 1. – С. 135–147.

108. Надарая, Э. А. О непараметрических оценках плотности вероятности и регрессии // Теория вероятностей и её применение. – 1965. – Том 10, выпуск 1. С. 199–203.

109. Надарая, Э. А. Об интегральной среднеквадратичной ошибке некоторых непараметрических оценок плотности распределения // Сообщения АН

ГССР. – 1972. – Том 68, № 1. – С. 33–36.

110. Надарая, Э. А. Оценки плотности двумерного распределения // Сообщения АН ГССР, 1964. – Том 32, №2. – С. 267–268.

111. Натансон, И. П. Теория функций вещественной переменной: Учебное пособие / И. П. Натансон. – 3-е изд. – М.: Наука, 1974. – 480 с.

112. Непараметрическое оценивание функционалов по стационарным выборкам / Ю. Г. Дмитриев [и др.]. – Томск: Изд-во ТГУ, 1974. – 87 с.

113. Нефедов, В. И. Общая теория связи : учебник для бакалавриата и магистратуры / В. И. Нефедов, А. С. Сигов. – М.: Издательство Юрайт, 2018. – 495 с.

114. Никифоров, А. Ф. Специальные функции математической физики // А. Ф. Никифоров, В. Б. Уваров. – М.: Интеллект, 2007. – 344 с.

115. Новицкий, П. В. Оценка погрешностей результатов измерений / П. В. Новицкий, И. А. Зограф. – Л.: Энергоатомиздат, 1991. – 304 с.

116. Новосёлов, А. А. О выборе структуры моделей адаптивных систем // Стохастические системы управления. – Новосибирск: Наука, 1979. – С. 72–77.

117. Новосёлов, А. А. Об оптимальном выборе структуры функции плотности вероятности и регрессии: препринт / А. А. Новосёлов. – Красноярск: ВЦ СО АН СССР, 1979. – 31 с.

118. Новосёлов, А. А. Оптимальная параметризация плотности вероятности и функции регрессии // Адаптация и обучение в системах управления и принятия решений. – Новосибирск: Наука, 1982. – С. 139 – 147.

119. Новосёлов, А. А. Параметризация моделей управляемых систем // Вестник СибГАУ. – 2010. – № 5. – С. 52–56.

120. Серых, А. П. О непараметрических оценках плотности, использующих статистическую эквивалентность выборочных блоков / А. П. Серых, Ф. П. Тарасенко, Н. Г. Черкашин // Математическая статистика и её приложения. – 1976. – Выпуск 4. – С. 173–186.

121. Сираджинов, С. Х. Об оценке плотности вероятности по зависимым выборкам / С. Х. Сираджинов, М. А. Мирзахмедов, А. К. Хосни // УзССР

Фангар Акад. докл. – 1977. – Выпуск 2. – С. 3–6.

122. Обучающие системы обработки информации и принятия решений: непараметрический подход / А. В. Лапко [и др.]. – Новосибирск: Наука. Сибирская издательская фирма РАН, 1996. – 296 с.

123. Рубан, А. И. Идентификация стохастических объектов на основе непараметрического подхода // Автоматика и телемеханика. – 1979. – № 11. – С. 106–117.

124. Рубан, А. И. Методы анализа данных : Учебное пособие / А. И. Рубан. – 2-е изд. – Красноярск: ИПЦ КГТУ, 2004. – 319 с.

125. Самаров, А. М. О минимаксной границе риска непараметрических оценок плотности // Проблемы передачи информации. – 1976. – Том 12, № 3. – С. 108–111.

126. Сергеев, В. Л. Об одном классе непараметрических оценок плотности. (Редколлегия к журналу «Известия высших учебных заведений. Физика»). Томск, 1977. – 26 с. (Рукопись депонирована в ВИНИТИ, 1977, № 58-77 Деп.)

127. Смирнов, В. И. Курс высшей математики. Том 5. – М.: Наука, 1974. – 656 с.

128. Смирнов, Н. В. О построении доверительной области для плотности распределения случайной величины // Доклады АН СССР, 1950. – Т. 74, 2. – С. 189–192.

129. Справочник для студентов: Высшая математика. Физика. Теоретическая механика. Сопротивление материалов / А. Д. Полянин [и др.]. – М.: Астрель, 2000. – 480 стр.

130. Стоянов, Й. Контрпримеры в теории вероятностей : Электронное издание / Й. Стоянов. – М.: МЦНМО, 2014. – 294 с.

131. Тарасенко, Ф. П. Непараметрическая статистика. – Томск, Изд-во ТГУ, 1976. – 294 с.

132. Тихонов, А. Н. Методы решения некорректных задач // А. Н. Тихонов, В. Я. Арсенин. – 2-е изд. – М.: Наука, 1979. – 285 с.

133. Фельдбаум, А. А. Основы теории оптимальных автоматических си-

стем. – М.: Наука, 1966. – 552 с.

134. Фихтенгольц, Г. М. Курс дифференциального и интегрального исчисления: В 3-х т. Т. 2 / Г. М. Фихтенгольц; пред. и прим. А. А. Флоринского. – 8-е изд. – М.: Физматлит, 2003. – 864 с.

135. Халмош, П. Теория меры. – М.: Факториал Пресс, 2003. – 256 с.

136. Хашимов, Ш. А. Замечания об оценках кривой регрессии и плотности распределения // Предельные теоремы и статистика. – Ташкент: Фан, 1976. – С. 163–171.

137. Хашимов, Ш. А. О среднеквадратической ошибке непараметрических оценок функции плотности // Случайные процессы и статистические выводы. – Выпуск 5. – Ташкент: Фан, 1975. С. 175–181.

138. Хашимов, Ш. А. Равномерная среднеквадратическая сходимость оценки плотности вероятности // Случайные процессы и статистические выводы. – Выпуск 4. – Ташкент: Фан, 1974. С. 185–193.

139. Хашимов, Ш. А. Скорость сходимости в оценке плотности вероятности и функции распределения // Сообщения АН ГССР. – Том 75, № 2. – С. 277–280.

140. Хелемский, А. Я. Лекции по функциональному анализу. – М.: МЦНМО, 2004. – 552 с.

141. Хорн, Р. Матричный анализ / Р. Хорн, Ч. Джонсон. – М.: Мир, 1989. – 655 с.

142. Цыпкин, Я. З. Адаптация и обучение в автоматических системах. – М.: Наука, 1968. – 399 с.

143. Цыпкин, Я. З. Основы теории обучающихся систем. – М.: Наука, 1970.

144. Ченцов, Н. Н. Оценка неизвестной плотности распределения по наблюдениям // ДАН СССР. – 1962. – 147, 1. – С. 45–48.

145. Ченцов, Н. Н. Статистические решающие правила и оптимальные выводы / Н. Н. Ченцов. – М.: Физматлит, 1972. – 520 с.

146. Численные методы анализа случайных процессов / М. Е. Лившиц [и др.]. – М.: Наука, 1976. – 128 с.

147. Шторм, Р. Теория вероятностей. Математическая статистика. Стати-

стический контроль качества. – М.: Мир, 1970. – 368 с.

148. Язык Wolfram Language [Электронный ресурс] / Wolfram, 2018. – Режим доступа: https://www.wolfram.com/language/

Публикации основных результатов работы в изданиях, рекомендованных ВАК:

149. Branishti, V. V. On some Properties of Weighted Hilbert Spaces // Journal of Siberian Federal University. Mathematics & Physics. – 2017. – № 10 (4). – С. 410–421.

150. Браништи, В. В. Введение пространства $L_{2,w}$ при построении проекционной оценки плотности вероятности // Вестник СибГАУ. – 2016. – № 1. – С. 19–26.

151. Браништи, В. В. Некоторые обобщения метода моментов при оценивании плотности вероятности в виде ортогонального ряда // Вестник СибГАУ. – 2015. – Том 16, № 3. – С. 566–571.

152. Браништи, В. В. О параметрическом оценивании функции плотности вероятности // Научно-технический вестник Поволжья. – 2014. – № 1. – С. 13–16.

Свидетельства о государственной регистрации программ для ЭВМ:

153. Браништи, В. В. Непараметрическое оценивание плотности распределения вероятности случайной величины / СибГАУ. – Свидетельство о государственной регистрации программы для ЭВМ №2008610808 от 15.02.2008 г.

Публикации основных результатов работы в других изданиях:

154. Браништи, В. В. Один способ расчёта коэффициента размытости для непараметрической оценки Розенблатта – Парзена плотности распределения вероятности случайной величины // Наука. Технологии. Инновации: Материалы всероссийской научной конференции (Новосибирск, 6–9 декабря 2007 г.). Часть 1. – Новосибирск: Изд-во НГТУ, 2007. С. 18–19.

155. Браништи, В. В. Оптимизация алгоритмов настройки коэффициента размытости для непараметрических оценок // Молодежь и наука: сборник материалов всероссийской научно-технической конференции [Электронный ресурс] // Отв. ред. О. А. Краев. – Красноярск: Сиб. федер. ун-т, 2014. – Режим

доступа: http://conf.sfu-kras.ru/sites/mn2014/pdf/d02/s14/s14_002.pdf

156. Браништи, В. В. Построение оценок плотности вероятности в виде суммы дельта-образных функций // Национальная ассоциация ученых. – 2015. – № 4 (9). Часть 7. – С. 10–13.

157. Браништи, В. В. Построение проекционных оценок для плотностей вероятности с неинтегрируемым квадратом // Решетнёвские чтения: материалы международной научно-практической конференции (Красноярск, 9–12 ноября 2016 г.). – Красноярск: СибГАУ, 2016. – С. 96–98.

158. Браништи, В. В. Применение метода моментов при оценивании функции плотности вероятности в виде линейных комбинаций ортогональных функций // Решетнёвские чтения: материалы международной научной конференции (Красноярск, 11–13 ноября 2014 г.). – Красноярск: СибГАУ, 2014. С. 22–24.

159. Браништи, В. В. Сравнение двух алгоритмов настройки длины ряда для проекционной оценки плотности вероятности // Актуальные проблемы гуманитарных и естественных наук. – 2016. – № 9 (92), ч. 1. – С. 10–14.

160. Браништи, В. В. Сравнение проекционных оценок плотности вероятности // Актуальные проблемы авиации и космонавтики: сборник материалов (Красноярск, 14 апреля 2017 г.). – Том 2. – Красноярск: СибГАУ, 2017. – С. 262–264.