ФГБОУ ВО МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В.Ломоносова
МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ

На правах рукописи
УДК 519.2

Ракитько Александр Сергеевич

Идентификация значимых факторов с помощью функционала ошибки

Специальность 1.1.4 —
«Теория вероятностей и математическая статистика»

Диссертация на соискание ученой степени
кандидата физико-математических наук

Научный руководитель:
д.ф.-м.н., профессор
Булинский Александр Вадимович

Москва — 2023

Оглавление

Стр.

Стр.

Введение

Настоящая диссертация подготовлена на кафедре теории вероятностей механико-математического факультета Московского государственного университета им. М.В.Ломоносова и посвящена исследованию математических методов выявления факторов, влияющих на изучаемую случайную функцию отклика.

Актуальность и история вопроса. В последние годы благодаря развитию информационных технологий наблюдается значительный рост объема данных, доступных для анализа. В связи с этим, огромное внимание уделяется современным исследовательским областям, связанным с анализом больших массивов данных, которые в англоязычной литературе носят названия Data Science, Data mining, Big Data, Machine Learning, Deep Learning [3],[53]. Это объясняется тем, что увеличение количества анализируемых данных предоставило возможность обнаруживать более сложные зависимости между переменными, нежели, например, линейные. Ярким примером является возрастающая популярность алгоритмов, в основе которых лежит архитектура нейронных сетей [61]. Подобные методы широко применяются в задачах распознавания речи [55], компьютерном зрении [78], организации сетей связи [1] и других.

Одной из областей, в которых исследователи неизбежно сталкиваются с необходимостью анализа данных высоких размерностей (больших данных), является биоинформатика. Прорыв в данной области был обусловлен прогрессом технологий расшифровки генома человека (Next-Generation Sequencing – секвенирование следующего поколения [4]). В начале 2000-ых годов впервые был расшифрован геном человека [69]. С тех пор стоимость расшифровки одного генома снизилась на несколько порядков, что позволило проводить данный анализ в масштабах целых стран. Развивающиеся информационные технологии, в том числе и квантовые вычисления, в перспективе позволят еще снизить стоимость и ускорить анализ генетических данных [91].

Условно все заболевания можно разделить на два типа: наследственные моногенные и мультифакторные [51]. К мультифакторным относятся многие сердечно-сосудистые заболевания (ишемическая болезнь сердца, гипертония, инсульт), онкологические заболевания (рак груди, рак простаты), болезнь Альцгеймера и многие другие. Мультифакторные болезни имеют более сложный механизм возникновения, нежели моногенные. В риск развития таких заболе-

ваний вносят вклад сразу много факторов (как генетические, так и факторы внешней среды – привычка курить, ожирение, малая подвижность и другие). При этом болезнь может быть не ассоциирована с некоторыми факторами по отдельности, но провоцироваться их совместной комбинацией.

Задача выявления факторов, ассоциированных с риском возникновения некоторого заболевания, является одной из наиболее частых в современной биостатистике. В некоторых исследованиях с помощью статистических тестов проверяется гипотеза о зависимости между генами-кандидатами, предположительно связанными с болезнью, и наличием заболевания у пациента [51],[90], [100]. В других исследованиях осуществляется полногеномный поиск ассоциаций, известный в англоязычной литературе как GWAS (Genome-wide association studies) [51], [93—96]. В таких исследованиях рассматривается зависимость между некоторой случайной функцией Y, обозначающей фенотип (наличие или отсутствие болезни, биохимический показатель крови, способность к обучению и т.д.) и генетическими факторами $X_1, \ldots, X_n$, $n \in \mathbb{N}$, находящихся почти во всех генах и межгенных пространствах. В современных исследованиях количество факторов n может достигать нескольких миллионов. В последние годы активно разрабатывается математический аппарат анализа данных высоких размерностей, например, генетических, когда число факторов n, описывающих одного индивидуума, соизмеримо с размером выборки N (см., [36]). При этом, по-настоящему, ассоциированными с откликом являются лишь некоторые из факторов $X_1, \ldots, X_n$, $n \in \mathbb{N}$. Для практических медицинских задач крайне важно знать, какие именно факторы влияют на функцию отклика. Это позволяет строить прогностические модели для предсказания риска развития заболевания [51], [92], [99], понимать патогенез заболеваний на молекулярно-генетическом уровне и находить мишени для лекарств. Возможно, в будущем в практику войдут и технологии редактирования генома.

Во многих случаях применяется двухэтапная процедура поиска значимых факторов. На первом этапе проводится однофактортный анализ, например, тест хи-квадрат Пирсона. По результатам теста отбираются факторы, показавшие наибольшую зависимость с изучаемой характеристикой. На втором этапе применяется многофакторный анализ понижения размерности данных наблюдений. В последние десятилетия активно разрабатывались такие методы как логистическая регрессия [47], случайные леса [58], LASSO [28], байесовские методы [79],

условная энтропия Шеннона [22; 23; 49],комбинации упомянутых методов [39] и другие.

Как уже отмечалось выше, для некоторых болезней характерны ситуации, когда по отдельности факторы могут давать незначительный вклад в развитие заболевания. Однако, их определенные комбинации могут приводить к существенному увеличению риска болезни. С практической точки зрения это означает необходимость применения нелинейных моделей, которые уже используются в различных областях: от биостатистики [51], [92], до усвоения данных (оценка состояния системы на основании текущих наблюдений, исторических наблюдений и модельных предположений) в гидрометеорологии [97; 98]. С целью выявления комбинаций факторов, влияющих на риск болезни, был предложен метод MDR (multifactor dimensionality reduction) [60]. Сейчас этот алгоритм активно используется в практических исследованиях с целью выявления эффекта взаимодействия генов для различных заболеваний [77]. Впоследствии были разработаны различные модификации данного метода, см., например, обзор [37]. В последние годы продолжают появляться работы, посвященные улучшениям и модификациям MDR алгоритма. В [8] исследуются три варианта MDR метода, позволяющие учитывать популяционную стратификацию индивидуумов. В [56] рассматривается MDR метод для многомерного фенотипа.

В данной диссертации мы продолжаем изучать и развивать метод MDR-EFE (Multifactor Dimensionality Reduction with Error Function Estimation). Впервые алгоритм был предложен в [19] для исследования бинарного отклика, и получил дальнейшее развитие в работах [20],[21],[80],[81],[82],[85]. Метод основан на статистической оценке функционала ошибки вида $Err(f) = |Y - f(X)|\psi(Y)$, где Y – изучаемый случайный отклик, X – вектор факторов, $f(\cdot)$ – предсказательная функция, а $\psi(\cdot)$ – штрафная функция. Оценка функционала ошибки строится по набору независимых одинаково распределенных векторов с помощью кросс-валидации для большей устойчивости алгоритма.

Цель работы. Целью работы является разработка новых методов идентификации значимых факторов с помощью функционала ошибки. В частности, ставится задачи по построению модификаций предложенного ранее MDR-EFE метода на случаи небинарной функции отклика. Также рассматривается модель с объясняющими факторами, имеющими абсолютно непрерывное распределение относительно меры Лебега в пространстве $\mathbb{R}^n$. Предлагается вариант

MDR-EFE метода с последовательным отбором значимых переменных. Развивается теория перестановочных случайных величин. Одной из основных целей работы является изучение асимптотических свойств используемых оценок в предложенных модификациях MDR-EFE метода. Проводится компьютерное моделирование, для иллюстрации работы MDR-EFE метода.

Структура и объем работы. Диссертация, объемом 110 страниц, состоит из введения, трех глав, заключения и списка литературы, насчитывающего 100 наименований. В заключении к диссертации сформулированы возможные направления дальнейшей деятельности.

В первой главе дается описание MDR-EFE метода идентификации значимых факторов с помощью функционала ошибки. Затем предлагается его модификация на случай небинарной функции отклика. Устанавливается критерий сильной состоятельности введенных оценок функционала ошибки в случае небинарной функции отклика. Доказывается теорема, которая обосновывает стратегию выбора набора значимых факторов. Доказывается теорема о сильной состоятельности функционала ошибки в случае объясняющих факторов, имеющих абсолютно непрерывное распределение относительно меры Лебега в пространстве $\mathbb{R}^n$.

Вторая глава посвящена изучению асимптотические свойства оценок функционала ошибки, построенных с помощью процедуры кросс-валидации. Доказывается центральная предельная теорема (ЦПТ) для регуляризованных оценок функционала ошибки в случае небинарной функции отклика. С целью получения дальнейших асимптотических результатов развивается теория перестановочных случайных величин. Доказывается аналог теоремы Эрдеша и Каца для перестановочных случайных величин. Устанавливается новый вариант ЦПТ для перестановочных случайных величин, с помощью которого доказывается новый вариант ЦПТ для оценок функционала ошибок. Полученные результаты о скорости сходимости построенных оценок к предельному распределению используются с целью получения асимптотических доверительных интервалов.

В третьей главе разрабатывается новая версия MDR-EFE метода с последовательным отбором значимых переменных. Для модели наивного байесовского классификатора устанавливаются оценки снизу для вероятности выбора значимого набора факторов MDR-EFE методом с последовательным

отбором переменных. MDR-EFE реализуется в виде программного кода, а его работа иллюстрируется на данных компьютерного моделирования.

Научная новизна работы. Все результаты, представленные в диссертации, являются новыми.

Положения, выносимые на защиту:

1. Критерий сильной состоятельности оценки функционала ошибки в MDR-EFE методе для случая небинарной функции отклика.

2. Теорема, обосновывающая стратегию выбора набора значимых факторов.

3. Достаточные условия сильной состоятельности оценок в случае объясняющих факторов, имеющих абсолютно-непрерывное распределение относительно меры Лебега в пространстве $\mathbb{R}^n$.

4. ЦПТ для регуляризованных оценок функционала ошибки в случае небинарной функции отклика.

5. Новый вариант ЦПТ для серий перестановочных случайных величин. Новый вариант ЦПТ для оценок функционала ошибок.

6. Аналог теоремы Эрдеша и Каца для перестановочных случайных величин.

7. Оценки снизу для вероятности выбора значимого набора факторов MDR-EFE методом с последовательным отбором переменных в случае модели наивного байесовского классификатора получены

Методы исследования. В работе используются классические методы теории вероятностей, вероятностные неравенства, асимптотические результаты для массивов случайных величин, анализ распределений случайных векторов. При доказательстве ЦПТ применяется техника перестановочных случайных величин. Часть теорем доказана с помощью результатов, справедливых для мартингалов.

Практическая и теоретическая значимость работы. Результаты диссертации носят теоретический характер. При этом они допускают и приложения. Разрабатываемый MDR-EFE метод и его модификации могут быть применимы в биостатистических задачах, требующих выявления факторов, оказывающих влияние на изучаемый отклик.

Апробация диссертации. Результаты диссертации докладывались на следующих **конференциях**.

1. International workshop «Probability, Analysis and Geometry», Ульм, Германия, 2013.

2. Международная научная конференция «Современные проблемы математики и механики», посвященная 75-летию академика РАН В.А. Садовничего, Москва, Россия, 2014.

3. XXI Международная научная конференция студентов, аспирантов и молодых ученых «Ломоносов», Москва, Россия, 2014.

4. XXXII International Seminar on Stability Problems for Stochastic Models, Трондхейм, Норвегия, 2014.

5. International Conference on Bioinformatics Models, Methods and Algorithms, Лиссабон, Португалия, 2015.

6. 6th Annual Canadian Human and Statistical Genetics Meeting, Квебек, Канада, 2017.

7. XXIV Международная научная конференция студентов, аспирантов и молодых ученых «Ломоносов», Москва, Россия, 2017.

8. V Международная конференция «Постгеном-2018», Казань, Россия, 2018.

9. The 5th International Conference on Stochastic Methods, Москва, Россия, 2020.

10. International conference «Limit Theorems of Probability Theory and Mathematical Statistics», Ташкент, Узбекистан, 2022.

Результаты диссертации неоднократно докладывались автором на следующих **научно-исследовательских семинарах**.

1. Большой семинар кафедры теории вероятностей под руководством академика РАН, профессора А.Н. Ширяева, механико-математический факультет, Московский государственный университет им. М.В. Ломоносова.

2. «Асимптотический анализ случайных процессов и полей» под руководством доктора физико-математических наук, профессора А.В. Булинского, механико-математический факультет, Московский государственный университет им. М.В. Ломоносова.

3. Аспирантский коллоквиум по теории вероятностей, математической статистике, теории случайных процессов под руководством академика РАН, профессора А.Н. Ширяева, механико-математический факультет, Московский государственный университет им. М.В. Ломоносва.

4. «Forschungsseminar Stochastische Geometrie und raumliche Statistik» под руководством Prof. E.Spodarev (Institut fur Stochastik, Ulm University, Germany, 2014 г.).

Публикации. Основные результаты диссертации изложены в 10 публикациях автора. Из них 4 статьи опубликованы в рецензируемых научных журналах, входящих в базы SCOPUS, Web of Science, RSCI. 2 статьи без соавторов опубликованы в трудах научных конференций. В материалах международных конференций представлены 4 публикации.

Личный вклад автора. Диссертантом совместно с научным руководителем проводился выбор темы, а также осуществлялось планирование всей работы. Профессору А.В.Булинскому принадлежит постановка задач и общий подход к их решению, им также доказаны леммы 2, 5, теоремы 8, 13 и следствия 1, 4. Предложение 2 и следствие 3 доказаны П. Алонсо-Руиз. Автору диссертации принадлежит доказательство остальных лемм, предложений, теорем, следствий, проведение компьютерного моделирования. В начале каждой главы диссертации также приводится список соответствующих публикаций с долей участия авторов.

Благодарность. В заключение автор выражает признательность научному руководителю профессору А.В. Булинскому за большую помощь в работе.

Глава 1. Состоятельность MDR-EFE метода в случаях небинарной функции отклика. Анализ моделей с дискретными и непрерывных объясняющими факторами

При подготовке данной главы диссертации использован материал публикаций [80; 81; 85]. Работа [81] выполнена автором в соавторстве с профессором А.В. Булинским. В публикации А.В. Булинскому принадлежит постановка задач и общий подход к их решению, им также доказана лемма 2 (лемма 2 в диссертации), следствие 1 (следствие 1 в диссертации) и следствие 4 (теорема 8 в диссертации), все остальные результаты доказаны автором диссертации. Работа [80] выполнена автором в соавторстве с профессором А.В. Булинским. В публикации А.В. Булинскому принадлежит постановка задач и общий подход к их решению, им также доказана лемма 1, все остальные результаты доказаны автором диссертации. Публикация [85] выполнена автором без соавторов.

1.1 Выявление значимых факторов с помощью функционала ошибки

В первой главе диссертации рассматривается задача выявления значимых факторов в рамках непараметрических моделей. Подобная задача возникает во многих медико-биологических исследованиях, в которых существует две и более выборки индивидуумов, отличающихся по какому-либо признаку. Например, выборка людей с некоторым заболеванием и выборка здоровых индивидуумов. Начнем с введения основных обозначений, формализации задачи и с описания MDR-EFE метода в первоначальной формулировке, предложенной в [19]. Далее все случайные величины заданы на некотором вероятностном пространстве $(\Omega, \mathcal{F}, \mathbb{P})$. Интеграл Лебега от случайной величины $\xi : \Omega \to \mathbb{R}$ по мере $\mathbb{P}$ будем обозначать $\mathbb{E}(\xi)$. Пусть $X = (X_1, \ldots, X_n)$ – случайный *вектор факторов* с компонентами $X_i : \Omega \to \mathbb{R}$, где $i = 1, \ldots, n$. Можно считать, что каждый индивид j в выборке $(X^1,Y^1), \ldots, (X^N,Y^N)$, имеющей то же распределение, что (X,Y), описывается вектором факторов X^j. В [19] предполагалось, что X_i, $i = 1, \ldots,n$, принимает значения в дискретном множестве $\{0,1,2\}$. Множество

значений случайного вектора X будем обозначать $\mathbb{X}$. Таким образом, в работе [19] рассматривается $\mathbb{X} = \{0,1,2\}^n$. В данной главе мы предложим обобщение MDR-EFE метода на случай, когда компоненты вектора X принимают значения в множестве вещественных чисел $\mathbb{R}$, то есть $\mathbb{X} = \mathbb{R}^n$. Случайная величина Y будет обозначать функцию отклика – значение некоторого признака, по которому мы различаем выборки (например, больных от здоровых). Множество значений Y будем обозначать $\mathbb{Y}$. В [19] рассматривалась бинарная функция отклика $Y : \Omega \to \{-1,1\}$. Однако, часто двух значений функции отклика недостаточно для того, чтобы описать возможное разнообразие исследуемого признака. Поэтому в данной главе также будут приведены результаты для небинарной функции отклика.

В [19] неслучайная функция $f : \mathbb{X} \to \{-1,1\}$ используется как функция предсказаний значений Y для некоторого индивидуума на основании значений вектора X для соответствующего инивидуума. Кроме того, в MDR-EFE методе фигурирует штрафная функция $\psi : \{-1,1\} \to \mathbb{R}_+$ (тривиальный случай $\psi \equiv 0$ исключается из рассмотрения). Качество предсказаний Y значениями $f(X)$ исследуется с помощью функционала ошибки:

$$Err(f) := \mathbb{E}|Y - f(X)|\psi(Y). \tag{1.1}$$

Будем называть *оптимальными* такие функции $f_{opt} : \mathbb{X} \to \{-1,1\}$, которые являются решением задачи $Err(f) \to \inf$, где нижняя грань берется по всем функциям $f : \mathbb{X} \to \{-1,1\}$. Согласно [25] все оптимальные функции имеют вид

$$f_{opt} = \mathbb{I}\{A\} - \mathbb{I}\left\{\overline{A}\right\}, \quad A \in \mathcal{A}, \tag{1.2}$$

$\mathbb{I}\{A\}$ обозначает индикатор множества A ($\mathbb{I}\{\varnothing\} := 0$), и $\mathcal{A}$ состоит из множеств

$$A = \{x \in M : F(x) < 0\} \cup B \cup C.$$

Здесь B – произвольное подмножество множества $\{x \in M : F(x) = 0\}$, где M – множество всех значений вектора X, имеющих отличную от нуля вероятность, то есть $M = \{x \in \mathbb{X} : \mathbb{P}(X = x) > 0\}$, а функция $F(x)$ задается формулой

$$F(x) = \psi(-1)\mathbb{P}(Y = -1|X = x) - \psi(1)\mathbb{P}(Y = 1|X = x), \quad x \in M.$$

C – произвольное подмножество $\overline{M} := \mathbb{X} \setminus M$. Рассмотрим $A^* = \{x \in M : F(x) < 0\}$. В силу того, что $\psi(-1) + \psi(1) \neq 0$, имеем

$$A^* = \{x \in M : \mathbb{P}(Y = 1|X = x) > \gamma(\psi)\}, \tag{1.3}$$

где

$$\gamma(\psi) := \psi(-1)/(\psi(-1)+\psi(1)). \tag{1.4}$$

Несложно преобразовать выражение (1.1) к виду

$$Err(f) = 2\sum_{y\in\{-1,1\}} \psi(y)\mathbb{P}(Y=y, f(X)\neq y). \tag{1.5}$$

Поскольку закон распределения вектора (X,Y) обычно не известен, выводы о качестве приближения Y с помощью $f(X)$ базируются на оценках функционала ошибки $Err(f)$.

Пусть $\xi^1,\xi^2,\ldots$ – последовательность независимых одинаково распределенных (н.о.р.) случайных векторов, имеющих тот же закон распределения, что и вектор (X,Y). Для $N\in\mathbb{N}$ положим $\xi_N = (\xi^1,\ldots,\xi^N)$. Мы будем аппроксимировать $Err(f)$ при $N\to\infty$ с помощью алгоритма предсказания. Он использует функцию $f_{PA} = f_{PA}(x,\xi_N)$ со значениями в множестве $\{-1,1\}$, которая определена для $x\in\mathbb{X}$ и случайной выборки ξ_N. Если $S\subset\{1,\ldots,N\}$ (символ "$\subset$" понимается как нестрогое включение "$\subseteq$"), то положим $\xi_N(S) = \{\xi^j : j\in S\}$ и $\overline{S} := \{1,\ldots,N\}\setminus S$. Для $K\in\mathbb{N}$ $(K>1)$ введем разбиение множества $\{1,\ldots,N\}$ на подмножества

$$S_k(N) = \{(k-1)[N/K]+1,\ldots,k[N/K]\mathbb{I}\{k<K\} + N\mathbb{I}\{k=K\}\}, \tag{1.6}$$

где $k=1,\ldots,K$, $[a]$ – целая часть числа $a\in\mathbb{R}$, $\mathbb{I}\{A\}$ – индикатор множества A. Построим оценку введенного функционала ошибки $Err(f)$, основываясь на выборке ξ_N, алгоритме предсказания с f_{PA} и применяя K-кратную кросс-валидацию, где $K\in\mathbb{N}$, $K>1$. А именно, следуя [20], положим

$$\widehat{Err}_K(f_{PA},\xi_N)$$
$$:= 2\sum_{y\in\{-1,1\}} \frac{1}{K}\sum_{k=1}^{K}\sum_{j\in S_k(N)} \frac{\widehat{\psi}(y,S_k(N))\mathbb{I}\{Y^j=y, f_{PA}(X^j,\xi_N(\overline{S_k(N)}))\neq y\}}{\#S_k(N)}, \tag{1.7}$$

$\#S$ обозначает мощность множества S, и для каждого $k=1,\ldots,K$ случайные величины $\widehat{\psi}(y,S_k(N))$ являются сильно состоятельными (при $N\to\infty$) оценками значений $\psi(y)$, $y\in\{-1,1\}$, построенными по $\{Y^j, j\in S_k(N)\}$.

Мы хотим гарантировать, что сходимость (в определенном смысле, когда $N\to\infty$) $f_{PA}(\cdot,\xi_N)$ к $f(\cdot)$ обеспечивает соотношение

$$\widehat{Err}_K(f_{PA},\xi_N) \to Err(f) \quad \text{п.н., } N\to\infty.$$

Предсказательный алгоритм f_{PA} будем строить следующим образом:

$$f_{PA}(x,\xi_N) = \begin{cases} 1, & \widehat{\mathbb{P}}(Y=1|X=x) > \widehat{\gamma}(\psi), \\ -1, & \text{иначе}, \end{cases} \tag{1.8}$$

где $\widehat{\mathbb{P}}(Y=1|X=x)$ – некоторая оценка условной вероятности $\mathbb{P}(Y=1|X=x)$, а $\widehat{\gamma}(\psi)$ – оценка пороговой функции $\gamma(\psi)$, заданная формулой (1.4).

1.2 Состоятельность MDR-EFE метода в случае небинарной функции отклика

Как отмечалось выше, ранее исследовались случаи бинарной функции отклика Y. В данном разделе мы обобщим результаты для MDR-EFE метода на случай, когда Y принимает некоторое конечное множество значений. Подобный подход важен с практической точки зрения, поскольку позволяет более детально дифференцировать выборки исследуемых [56; 73; 74].

1.2.1 Обозначения и вспомогательные результаты

Пусть $X = (X_1,\ldots,X_n)$ – случайный вектор с компонентами $X_k : \Omega \to \{0,1,\ldots,s\}$, где $k = 1,\ldots,n$ и $s,n \in \mathbb{N}$. Положим $\mathbb{X} = \{0,\ldots,s\}^n$, $\mathbb{Y} = \{-m,\ldots,0,\ldots,m\}$, здесь $m \in \mathbb{N}$. Мы предполагаем, что $Y : \Omega \to \mathbb{Y}$, $f : \mathbb{X} \to \mathbb{Y}$ и штрафная функция $\psi : \mathbb{Y} \to \mathbb{R}_+$. Тривиальный случай $\psi \equiv 0$ исключается из рассмотрения.

Замечание 1. В медицине функция отклика Y, как правило, связана с состоянием здоровья пациента. Для этого используется некоторая шкала значений, отражающая степень развития заболевания. Если, предположим, данная шкала состоит из значений $\{0,1,\ldots,m\}$, то наша модель сводится к данному случаю предположением, что Y принимает значения $\{-m,\ldots,-1\}$ с нулевой вероятностью. Более того, мы можем считать, что Y принимает произвольные рациональные значения $0 \leqslant x_1 \leqslant \ldots \leqslant x_m$, где $x_k = s_k/L$ ($s_k \in \mathbb{N}$, $L \in \mathbb{N}$,

$k = 1,\ldots,m$). Тогда введем соответствие $x_k \mapsto s_k$, $k = 1,\ldots,m$, и рассмотрим $\mathbb{Y} = \{-s_m,\ldots,0,\ldots,\ldots,s_m\}$. Мы используем сильно состоятельные оценки штрафной функции, и если нам известно, что $\mathbb{P}(Y = y) = 0$ для некоторого $y \in \mathbb{Y}$, тогда мы можем положить $\psi(y) = 0$ и $\widehat{\psi}(y) \equiv 0$ для таких $y \in \mathbb{Y}$ ($N \in \mathbb{N}$), и все установленные в дальнейшем результаты останутся справедливы. Напомним также, что мы устанавливаем значимость отклонения $f(X)$ от Y с помощью штрафной функции ψ.

Рассмотрим множества $A_y = \{x \in \mathbb{X} : f(x) = y\}$, где $y \in \mathbb{Y}$, и положим, как и ранее, $M = \{x \in \mathbb{X} : \mathbb{P}(X = x) > 0\}$. Тогда мы можем представить $Err(f)$ в виде

$$Err(f) = \sum_{y,z\in\mathbb{Y}} |y - z|\psi(y)\mathbb{P}(Y = y,f(X) = z) = \sum_{z\in\mathbb{Y}} \sum_{x\in A_z} w^\top(x)q(z). \qquad (1.9)$$

Здесь $q(z)$ – столбец с номером z матрицы Q размерности $(2m+1)\times(2m+1)$ с элементами $q_{y,z} = |y - z|$, $y,z \in \mathbb{Y}$ (элемент $q_{-m,-m}$ находится в левом верхнем углу Q),

$$w(x) = (\psi(-m)\mathbb{P}(Y = -m,X = x),\ldots,\psi(m)\mathbb{P}(Y = m,X = x))^\top$$

и $\top$ обозначает транспонирование. Все векторы рассматриваются как столбцы. Как обычно, $\#A$ обозначает мощность конечного множества A.

Для каждого непустого множества J такого, что $J \subset \mathbb{Y}$, положим

$$B_J = \{x \in \mathbb{X}\colon w^\top(x)q(y) = w^\top(x)q(z),\ y,z \in J;\\ w^\top(x)q(y) < w^\top(x)q(v),\ y \in J, v \in \mathbb{Y} \setminus J\}. \quad (1.10)$$

Если $J = \mathbb{Y}$, то $B_{\mathbb{Y}} = \{x \in \mathbb{X}\colon w^\top(x)q(y) = w^\top(x)q(z),\ y,z \in \mathbb{Y}\}$. Заметим, что $B_J \cap B_I = \varnothing$, если $J \neq I (I,J \subset \mathbb{Y})$. Более того,

$$\cup_{J\subset\mathbb{Y},J\neq\varnothing} B_J = \mathbb{X}. \qquad (1.11)$$

Мы пишем B_y для B_J, когда $J = \{y\}$, $y \in \mathbb{Y}$. Как и прежде, $\mathbb{I}\{A\}$ – индикатор множества A, $\mathbb{I}\{\varnothing\} := 0$.

Опишем функции $f : \mathbb{X} \to \mathbb{Y}$, которые являются решениями оптимизационной задачи $Err(f) \to \inf$. Другими словами, мы будем искать все такие разбиения, A_y, $y \in \mathbb{Y}$, множества $\mathbb{X}$, что

$$f = \sum_{y\in\mathbb{Y}} y\,\mathbb{I}\{A_y\} \qquad (1.12)$$

соответсвует минимальному значению $Err(f)$. Любую такую функцию f мы будем называть *оптимальной функцией*. Ввиду (1.9) можно утверждать, что $B_y \subset A_y$ для каждого $y \in \mathbb{Y}$. Если $\cup_{y\in\mathbb{Y}} B_y \neq \mathbb{X}$, то для каждого

$$x \in \mathbb{X} \setminus \cup_{y\in\mathbb{Y}} B_y$$

существует такое $J = J(x)$, что $x \in B_J$, где $J \subset \mathbb{Y}$ и $\#J > 1$. В таком случае можно включать x в любое множество A_y с y, принадлежащим J (другими словами, расширить одно из множеств B_y, $y \in J$, значением элемента x). Таким образом мы получим, что $A_y = B_y \cup C_y$, где C_y, $y \in \mathbb{Y}$, образуют разбиение множества $\mathbb{X} \setminus \cup_{y\in\mathbb{Y}} B_y$. Очевидно, такой алгоритм построения ведет к оптимальной функции f с минимальной величиной $Err(f)$. Иной выбор A_y, $y \in \mathbb{Y}$, приведет к f с большими значениям $Err(f)$. Следовательно, мы приходим к следующему утверждению.

Лемма 1. *Любая функция $f : \mathbb{X} \to \mathbb{Y}$, являющаяся решением задачи $Err(f) \to \inf$, имеет вид (1.12) с A_y, $y \in \mathbb{Y}$, заданными выше.*

Замечание 2. Ясно, что мы можем указать уникальный способ построения множеств C_y, $y \in \mathbb{Y}$. Например, если $\mathbb{X} \setminus \cup_{y\in\mathbb{Y}} B_y \neq \varnothing$, тогда для каждого $x \in \mathbb{X} \setminus \cup_{y\in\mathbb{Y}} B_y$ найдется такое B_J, что $x \in B_J$ ($J = J(x)$, $J \subset \mathbb{Y}$, $\sharp\, J > 1$). Если $J = \{y_1, \ldots, y_r\}$, где $y_1 < \ldots < y_r$, то мы включаем x в C_{y_1}. Отметим также, что мы можем рассматривать оптимальную функцию f с $A_y^* = A_y \cap M$ для $y \in \mathbb{Y} \setminus \{0\}$ и $A_0^* = A_0 \cup (\mathbb{X} \setminus M)$.

Кроме того, удобно описать B_y следующим образом:

$$x \in B_y \Longleftrightarrow \begin{cases} w^\top(x)q(-m) < w^\top(x)q(-m+1), & y = -m, \\ w^\top(x)q(y) < w^\top(x)q(z), \quad z = y \pm 1, & y \neq \pm m, \\ w^\top(x)q(m-1) > w^\top(x)q(m), & y = m. \end{cases} \tag{1.13}$$

В частности, B_y может быть пустым множеством. Для того, чтобы показать, что (1.13) выполняется, мы определим для каждого $y \in \mathbb{Y}$, $y > -m$, вектор $\Delta(y) := q(y) - q(y-1)$. Ясно, что

$$\Delta(y) = (\underbrace{1, \ldots, 1}_{m+y}, \underbrace{-1, \ldots, -1}_{m-y+1})^\top. \tag{1.14}$$

Неравенство $w^\top(x)q(y) < w^\top(x)q(y+1)$ перепишем эквивалентным образом:

$$w^\top(x)\Delta(y+1) > 0. \tag{1.15}$$

Для всех $x \in \mathbb{X}$ вектор $w(x)$ имеет неотрицательные компоненты

$$w_y(x) := \psi(y)\mathbb{P}(Y = y, X = x), \quad y \in \mathbb{Y}. \tag{1.16}$$

Поэтому неравенство $w^\top(x)\Delta(y+1) > 0$ и (1.14) влекут $w^\top(x)\Delta(z) > 0$, если $z \geqslant y+1$ ($z \in \mathbb{Y}$). Для $z \geqslant y+1$ ($z \in \mathbb{Y}$) имеем

$$w^\top(x)(q(z) - q(y)) = \sum_{k=y+1}^{z} w^\top(x)\Delta(k). \tag{1.17}$$

Следовательно, $w^\top(x)(q(z)-q(y)) > 0$. Аналогичным образом можно показать, что неравенство $w^\top(x)q(y) < w^\top(x)q(y-1)$ влечет для $k < y$, $k \in \mathbb{Y}$, соотношение $w^\top(x)q(y) < w^\top(x)q(k)$. Таким образом, (1.13) установлено. Используя (1.17), мы получаем, что множества $J = \{y_1, \ldots, y_r\}$, фигурирующее в замечании 2, имеют вид $\{y_1, y_1+1, \ldots, y_1+r-1\}$.

Для $x \in \mathbb{X}$ рассмотрим вектор $L(x)$, с $2m$ компонентами

$$L_y(x) := w^\top(x)\Delta(y) = w_{-m}(x) + \ldots + w_{y-1}(x) - w_y(x) - \ldots - w_m(x), \tag{1.18}$$

здесь $y \in \mathbb{Y}$, $y > -m$. Тогда, в силу (1.13) имеем для каждого $y \in \mathbb{Y}$,

$$x \in B_y \Longleftrightarrow \begin{cases} L_{-m+1}(x) > 0, & y = -m, \\ L_{y+1}(x) > 0, L_y(x) < 0, & y \neq \pm m, \\ L_m(x) < 0, & y = m. \end{cases} \tag{1.19}$$

Далее мы воспользуемся свойством вектор-функции $L(x)$, $x \in \mathbb{X}$, устанавливаемом в следующей лемме.

Лемма 2. *Пусть $L_t(x) = 0$ и $L_z(x) = 0$ для некоторых $x \in \mathbb{X}$, $t, z \in \mathbb{Y}$, $-m < t < z$. Тогда $L_y(x) = 0$ для всех таких $y \in \mathbb{Y}$, что $t \leqslant y \leqslant z$.*

Доказательство. Для каждого $x \in \mathbb{X}$ вектор $w(x)$ имеет неотрицательные компоненты. Формула (1.18) показывает, что для любого $x \in \mathbb{X}$ функция $L_y(x)$ является неубывающей функцией по y ($y \in \mathbb{Y}, y > -m$). Это наблюдение приводит к утверждению леммы $\square$.

Используя замечание 2, удобно сделать следующий выбор *оптимальной функции* f_{opt}. А именно, согласно (1.19) мы можем записать

$$f_{opt}(x) = y \Longleftrightarrow \begin{cases} L_{-m+1}(x) \geqslant 0, & y = -m, \\ L_{y+1}(x) \geqslant 0, L_y(x) < 0, & y \neq \pm m, \\ L_m(x) < 0, & y = m. \end{cases} \tag{1.20}$$

В действительности, согласно замечанию 1, мы имеем $A_m = B_m$, и, следовательно, можем написать в (1.20) строгое неравенство $L_m(x) < 0$ при $y = m$.

Теперь рассмотрим случайные векторы φ и χ с соответствующими компонентами

$$\varphi_y = \psi(y)\mathbb{I}\{Y = y\}, \quad \chi_y = \mathbb{I}\{X \in A_y\}, \ y \in \mathbb{Y}.$$

Тогда перепишем (1.9) как

$$Err(f) = \mathsf{E}\varphi^\top Q\,\chi.$$

Заметим, что Q может быть представлена как сумма $2m$ симметрических матриц с элементами 0 или 1.

$$Q = \begin{pmatrix} 0 & 1 & 1 & \dots & 1 & 1 & 1 \\ 1 & 0 & 1 & \dots & 1 & 1 & 1 \\ 1 & 1 & 0 & \dots & 1 & 1 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & \dots & 0 & 1 & 1 \\ 1 & 1 & 1 & \dots & 1 & 0 & 1 \\ 1 & 1 & 1 & \dots & 1 & 1 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 1 & \dots & 1 & 1 & 1 \\ 0 & 0 & 0 & \dots & 1 & 1 & 1 \\ 1 & 0 & 0 & \dots & 1 & 1 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & \dots & 0 & 0 & 1 \\ 1 & 1 & 1 & \dots & 0 & 0 & 0 \\ 1 & 1 & 1 & \dots & 1 & 0 & 0 \end{pmatrix}$$

$$+ \dots + \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 & 0 \end{pmatrix}. \quad (1.21)$$

Другими словами,

$$Q = \sum_{i=0}^{2m-1} Q^{(i)}, \quad (1.22)$$

где матрица $Q^{(i)} = (q^{(i)}_{y,z})_{y,z\in\mathbb{Y}}$ имеет элементы $q^{(i)}_{y,z} = 0$ при $|y - z| \leqslant i$ и $q^{(i)}_{y,z} = 1$ в противном случае. Формула (1.22) позволяет записать $Err(f)$ в виде

$$Err(f) = \sum_{i=0}^{2m-1} \sum_{i-m<|y|\leqslant m} \psi(y)\mathbb{P}(Y = y, |f(X) - y| > i). \quad (1.23)$$

Здесь мы использовали то, что в представлении Q в виде суммы матриц некоторые из этих матриц имеют нулевые строки.

Распределение (X,Y) неизвестно, следовательно, для произвольной $f:\mathbb{X}\to\mathbb{Y}$, мы не можем вычислить $Err(f)$. Таким образом, естественно, что статистические выводы о качестве предсказания отклика Y значениями $f(X)$ основываются на оценках функционала ошибки $Err(f)$.

Как и ранее, пусть $\xi^1,\xi^2,\ldots$ – последовательность н.о.р. случайных векторов, имеющих такое же распределение, что и (X,Y). Для $N\in\mathbb{N}$ положим $\xi_N=(\xi^1,\ldots,\xi^N)$. Мы будем приближать $Err(f)$ используя ξ_N (при $N\to\infty$) и *предсказательный алгоритм* (PA). Данный предсказательный алгоритм использует функцию $f_{PA}=f_{PA}(x,\xi_N)$, заданную для $x\in\mathbb{X}$ и ξ_N и принимающую значения в $\mathbb{Y}$. В (1.8) описана функция f_{PA} для бинарного отклика. Более точно, мы имеем дело с *семейством функций* $f_{PA}(x,\upsilon_m)$ (со значениями в $\mathbb{Y}$), определенными для $x\in\mathbb{X}$ и $\upsilon_m\in(\mathbb{X}\times\mathbb{Y})^m$, где $m\in\mathbb{N}$, $m\leqslant N$. Для упрощения обозначений мы пишем $f_{PA}(x,\upsilon_m)$ вместо $f^m_{PA}(x,\upsilon_m)$. Для $S\subset\{1,\ldots,N\}$ положим $\xi_N(S)=\{\xi^j,j\in S\}$ и $\overline{S}:=\{1,\ldots,N\}\setminus S$. Для $K\in\mathbb{N}$, $(K>1)$ введем разбиение множества $\{1,\ldots,N\}$ согласно (1.6). Следуя [25], мы можем построить оценки $Err(f)$, используя ξ_N, предсказательный алгоритм f_{PA} и K-кросс-валидацию [10]. А именно, положим

$$\widehat{Err}_K(f_{PA},\xi_N):=\sum_{i=0}^{2m-1}\sum_{i-m<|y|\leqslant m}\frac{1}{K}\sum_{k=1}^{K}\sum_{j\in S_k(N)}\widehat{\psi}(y,\xi_N(S_k(N)))\times\frac{\mathbb{I}\{Y^j=y,|f_{PA}(X^j,\xi_N(\overline{S_k(N)}))-y|>i\}}{\sharp S_k(N)}. \quad (1.24)$$

Здесь для каждого $k\in\{1,\ldots,K\}$ оценка $\widehat{\psi}(y,\xi_N(S_k(N))$ является сильно состоятельной оценкой $\psi(y)$ (при $N\to\infty$) для всех $y\in\mathbb{Y}$, т.е.

$$\widehat{\psi}(y,\xi_N(S_k(N)))\to\psi(y)\ \text{ п.н.},\ y\in\mathbb{Y},\ N\to\infty. \quad (1.25)$$

В практических приложениях функция $\psi(\cdot)$ и ее статистическая оценка $\widehat{\psi}(\cdot,\cdot)$ выбираются таким образом, что соотношение (1.25) выполняется.

1.2.2 Критерий сильной состоятельности оценки функционала ошибки

Мы хотим, чтобы сходимость (в определенном смысле) $f_{PA}(\cdot,\xi_N)$ к $f(\cdot)$ при $N \to \infty$ влекла соотношение

$$\widehat{Err}_K(f_{PA},\xi_N) \to Err(f) \quad \text{п.н.}, \quad N \to \infty. \tag{1.26}$$

Здесь и далее сумма по пустому множеству полагается равной нулю.

Теорема 1. *Пусть $\xi^1,\xi^2,\ldots$ – последовательность независимых одинаково распределенных случайных величин с таким же законом распределения, что и (X,Y), ψ – штрафная функция, $f : \mathbb{X} \to \mathbb{Y}$ и f_{PA} задают предсказательный алгоритм. Предположим, что существует такое непустое множество $U \subset \mathbb{X}$, что для каждого $x \in U$ и всех $k = 1,\ldots,K$ имеем*

$$f_{PA}(x,\xi_N(\overline{S_k(N)})) \to f(x) \quad \textit{п.н.}, \quad N \to \infty. \tag{1.27}$$

Тогда (1.26) выполняется в том и только том случае, если

$$\sum_{k=1}^{K} \sum_{x \in \mathbb{X}\setminus U} w^\top(x)\, Q\, \delta(N,x,k) \to 0 \quad \textit{п.н.}, \quad N \to \infty, \tag{1.28}$$

где для $x \in \mathbb{X}$, $N \in \mathbb{N}$ и $k = 1,\ldots,K$ вектор $\delta(N,x,k)$ имеет компоненты

$$\delta_y(N,x,k) = \mathbb{I}\{f_{PA}(x,\xi_N(\overline{S_k(N)})) = y\} - \mathbb{I}\{f(x) = y\}, \ y \in \mathbb{Y}.$$

Доказательство. Соотношение (1.26) эквивалентно следующему:

$$\sum_{i=0}^{2m-1} \sum_{i-m<|y|\leqslant m} \frac{1}{K} \sum_{k=1}^{K} \sum_{j \in S_k(N)} \psi(y) \times \frac{\mathbb{I}\{Y^j = y, |f_{PA}(X^j,\xi_N(\overline{S_k(N)})) - y| > i\}}{\sharp S_k(N)} \to Err(f) \ \text{ п.н.} \tag{1.29}$$

при $N \to \infty$. В самом деле, (1.25) выполняется, и для произвольных $\omega \in \Omega$, $i = 0,\ldots,2m-1$ и $k = 1,\ldots,K$ имеем

$$\frac{1}{\sharp S_k(N)} \sum_{j \in S_k(N)} \mathbb{I}\{Y^j = y, |f_{PA}(X^j,\xi_N(\overline{S_k(N)})) - y| > i\} \leqslant 1.$$

В работе [43] был установлен усиленный закон больших чисел для массивов независимых величин (УЗБЧ).

Теорема 2 (Теорема 2, [43]). *Пусть* $\{X_{n,k}\}$ *– массив построчно независимых случайных величин, для которых справедливо* $\mathbb{E}X_{n,k} = 0$ *при всех* n *и* k*. Кроме того, пусть* $\{X_{n,k}\}$ *равномерно ограничены случайной величиной* X*, удовлетворяющей неравенству* $\mathbb{E}|X|^{2p} < \infty$ *при некотором* $1 \leqslant p < 2$*. Тогда*

$$\frac{1}{n^{1/p}} \sum_{k=1}^{n} X_{n,k} \to 0$$

вполне при $n \to \infty$*.*

Из теоремы 2 для всех y и i следует:

$$\frac{1}{\sharp S_k(N)} \sum_{j \in S_k(N)} \mathbb{I}\{Y^j = y, |f(X^j) - y| > i\} \to \mathbb{P}(Y = y, |f(X) - y| > i) \quad \text{п.н.}$$

при $N \to \infty$. Действительно, массив случайных величин

$$\Big\{\mathbb{I}\{Y^j = y, |f(X^j) - y| > i\} - \mathbb{P}(Y = y, |f(X) - y| > i)\Big\}_{N,j}$$

состоит из построчно независимых центрированных случайных величин, равномерно ограниченных константой 2. Осталось отметить, что сходимость вполне влечет сходимость почти наверное.

Для всех $k = 1, \ldots, K$ мы имеем

$$\sum_{i=0}^{2m-1} \sum_{i-m<|y|\leqslant m} \sum_{j \in S_k(N)} \frac{\psi(y)\mathbb{I}\{Y^j = y, |f(X^j) - y| > i\}}{\sharp S_k(N)} \to Err(f) \quad \text{п.н.} \quad (1.30)$$

при $N \to \infty$. Для $y \in \mathbb{Y}$, $N \in \mathbb{N}$, $k = 1, \ldots, K$ и $i = 0, \ldots, 2m-1$ введем случайные величины

$$Q^{(i)}_{N,k}(y) = \frac{1}{\sharp S_k(N)} \sum_{j \in S_k(N)} \mathbb{I}\{Y^j = y\} F^{(i)}_{N,k}(X^j, y),$$

здесь

$$F^{(i)}_{N,k}(x,y) := \mathbb{I}\{|f_{PA}(x, \xi_N(\overline{S_k(N)})) - y| > i\} - \mathbb{I}\{|f(x) - y| > i\}. \quad (1.31)$$

Ввиду (1.30) выражение (1.29) эквивалентно следующему:

$$\sum_{k=1}^{K} \sum_{i=0}^{2m-1} \sum_{i-m<|y|\leqslant m} \psi(y) Q^{(i)}_{N,k}(y) \to 0 \ \text{ п.н., } N \to \infty. \quad (1.32)$$

Мы можем записать $Q^{(i)}_{N,k}(y) = Q^{(i),U}_{N,k}(y) + Q^{(i),\mathbb{X}\setminus U}_{N,k}(y)$, $i = 0,\dots,2m-1$, где

$$Q^{(i),V}_{N,k}(y) = \frac{1}{\sharp S_k(N)} \sum_{j\in S_k(N)} \mathbb{I}\{X^j \in V\}\mathbb{I}\{Y^j = y\}F^{(i)}_{N,k}(X^j,y),\ V \subset \mathbb{X}.$$

Для $y \in \mathbb{Y}$, $N \in \mathbb{N}$, $k = 1,\dots,K$ и $i = 0,\dots,2m-1$ приходим к неравенствам

$$|Q^{(i),U}_{N,k}(y)| \leqslant \sum_{x\in U} \left|\mathbb{I}\{|f_{PA}(x,\xi_N(\overline{S_k(N)})) - y| > i\} - \mathbb{I}\{|f(x) - y| > i\}\right|. \quad (1.33)$$

Функции f и f_{PA} принимают значения в $\mathbb{Y}$. Таким образом, (1.27) влечет, что для каждого $x \in U$, $k = 1,\dots,K$ и почти всех $\omega \in \Omega$ можно найти такое достаточно большое целое $N_0(x,k,\omega)$, что $f_{PA}(x,\xi_N(\overline{S_k(N)})) = f(x)$ при $N \geqslant N_0(x,k,\omega)$. Следовательно, $Q^{(i),U}_{N,k}(y) = 0$ для всех $i \in \{0,\dots,2m-1\}$, $y \in \mathbb{Y}$, $k = 1,\dots,K$ и почти всех $\omega \in \Omega$, когда $N \geqslant N_0(\omega) = \max_{x\in U,k=1,\dots,K} N_0(x,k,\omega)$. Очевидно, $N_0 < \infty$ п.н., потому что $\sharp U < \infty$. Таким образом, мы показали, что

$$\sum_{k=1}^{K}\sum_{i=0}^{2m-1}\sum_{i-m<|y|\leqslant m} \psi(y)Q^{(i),U}_{N,k}(y) \to 0 \quad \text{п.н.},\ N \to \infty. \quad (1.34)$$

Если $U = \mathbb{X}$, то тогда $Q^{(i),\mathbb{X}\setminus U}_{N,k}(y) = 0$ для всех возможных i, N, k и y. В этом случае (1.34) эквивалентно (1.32). Следовательно, для $U = \mathbb{X}$ утверждение теоремы доказано. Пусть теперь $U \neq \mathbb{X}$. Положим

$$\tau_N(\mathbb{X}\setminus U) = \sum_{k=1}^{K}\sum_{i=0}^{2m-1}\sum_{i-m<|y|\leqslant m} \psi(y)Q^{(i),\mathbb{X}\setminus U}_{N,k}(y).$$

Очевидно, что

$$\tau_N(\mathbb{X}\setminus U) = \sum_{k=1}^{K}\sum_{x\in\mathbb{X}\setminus U}\sum_{i=0}^{2m-1}\sum_{i-m<|y|\leqslant m} \psi(y) \times \frac{1}{\sharp S_k(N)} \sum_{j\in S_k(N)} \mathbb{I}\{X^j = x\}\mathbb{I}\{Y^j = y\}F^{(i)}_{N,k}(x,y). \quad (1.35)$$

В силу УЗБЧ для массивов (теорема 2), для всех $x \in \mathbb{X}$, $y \in \mathbb{Y}$ и $k = 1,\dots,K$,

$$\frac{1}{\sharp S_k(N)} \sum_{j\in S_k(N)} \mathbb{I}\{X^j = x\}\mathbb{I}\{Y^j = y\} \to \mathbb{P}(X = x, Y = y) \quad \text{п.н.},\ N \to \infty. \quad (1.36)$$

Тогда (1.35) и (1.36) показывают, что $\lim_{N\to\infty}\tau_N(\mathbb{X}\setminus U)=0$ п.н. в том и только том случае, когда

$$\nu_N:=\sum_{k=1}^{K}\sum_{x\in\mathbb{X}\setminus U}\sum_{i=0}^{2m-1}\sum_{i-m<|y|\leqslant m}w_y(x)F_{N,k}^{(i)}(x,y)\to 0\quad\text{п.н., } N\to\infty. \tag{1.37}$$

Принимая во внимание, что $\mathbb{I}\{\cup_{j\in J}D_j\}=\sum_{j=1}^{J}\mathbb{I}\{D_j\}$ для попарно непересекающихся множеств $D_1,\ldots,D_J$, мы имеем

$$\nu_N=\sum_{k=1}^{K}\sum_{x\in\mathbb{X}\setminus U}\sum_{i=0}^{2m-1}\sum_{i-m<|y|\leqslant m}\sum_{|r-y|>i}w_y(x)\delta_r(N,x,k)$$

$$=\sum_{k=1}^{K}\sum_{x\in\mathbb{X}\setminus U}\sum_{y\in\mathbb{Y}}w_y(x)\sum_{i=0}^{I(y)}\sum_{|r-y|>i}\delta_r(N,x,k), \tag{1.38}$$

где $I(y)=2m-1$ для $y=\pm m$ и $I(y)=m-1+|y|$ при $|y|<m$ ($y\in\mathbb{Y}$). Заметим, что

$$\sum_{i=0}^{I(y)}\sum_{|r-y|>i}\delta_r(N,x,k)=\sum_{r=-m}^{m}\sum_{i<|r-y|}\delta_r(N,x,k)=\sum_{r=-m}^{m}|y-r|\delta_r(N,x,k)$$

$$=(Q\delta(N,x,k))_y, \tag{1.39}$$

так как $|r-y|-1\leqslant I(y)$ для всех $r,y\in\mathbb{Y}$. Здесь $(Q\delta(N,x,k))_y$, $y\in\mathbb{Y}$, – компоненты вектора $Q\delta(N,x,k)$. Следовательно, (1.38) и (1.39) означают, что условие (1.37) эквивалентно (1.28). Доказательство завершено. □

Замечание 3. В теореме 1 устанавливается условие, которое должно выполняться вне "хорошего множества" U, на котором справедливо (1.27), для того, чтобы гарантировать соотношение (1.26). Далее мы покажем, что возможно эффективно проверять условия (1.27) и (1.28). Заметим, что утверждение Теоремы 1 будет выполняться, если вместо $\xi^1,\ldots,\xi^N$ мы рассмотрим независимые случайные векторы $\xi_{N,1},\ldots,\xi_{N,N}$ такие, что $\xi_{N,j}:=(X^{(N,j)},Y^{(N,j)})$ совпадают по распределению с (X,Y), $j=1,\ldots,N$, $N\in\mathbb{N}$.

Замечание 4. В теореме 1 мы не предполагали, что непустое множество U состоит из всех $x\in\mathbb{X}$, удовлетворяющих условию (1.27). Однако, если соотношение (1.27) справедливо для некоторых $z\in\mathbb{X}\setminus U$, тогда $f_{PA}(z,\xi_N(\overline{S_k(N)}))=f(z)$ п.н. для $k=1,\ldots,K$ при достаточно больших N (например, при $N>N_0(\omega)$).

Следовательно, (1.37) эквивалетно аналогичному условию, в котором суммирование по $x \in \mathbb{X} \setminus U$ заменено суммированием по $x \in \mathbb{X} \setminus (U \cup \{z\})$. Таким образом, мы получим эквивалентную формулировку теоремы 1, если U состоит из всех $x \in \mathbb{X}$, удовлетворяющих (1.27). Более того, если не существует непустого $U \subset \mathbb{X}$, удовлетворяющего (1.27), тогда соотношение (1.26) равносильно (1.37) с $U = \varnothing$. Поэтому в теореме 1 можно предполагать отличие U от пустого множества.

Замечание 5. Пусть выполнено (1.26), и для некоторой константы C_0 имеем

$$\widehat{\psi}(y,\xi_N(S_k(N))) \leqslant C_0 \text{ п.н. при } N \in \mathbb{N},\ k = 1,\ldots,K,\ y \in \mathbb{Y}. \tag{1.40}$$

Тогда $\widehat{Err}_K(f_{PA},\xi_N) \leqslant 2m(2m+1)C_0$, $N \in \mathbb{N}$. Из теоремы Лебега о мажорируемой сходимости следует, что $\widehat{Err}_K(f_{PA},\xi_N)$ асимптотически несмещенная оценка $Err(f)$.

Для $N \in \mathbb{N}$, $x \in \mathbb{X}$, $k \in \{1,\ldots,K\}$ и $t \in \mathbb{Y}$ зададим случайный вектор $I(N,x,k,t)$ с компонентами

$$I_y(N,x,k,t) = \begin{cases} -\mathbb{I}\{f_{PA}(x,\xi_N(\overline{S_k(N)})) < y\}, & -m < y \leqslant t, \\ \mathbb{I}\{f_{PA}(x,\xi_N(\overline{S_k(N)})) \geqslant y\}, & t < y \leqslant m. \end{cases} \tag{1.41}$$

Если $t = -m$, тогда $\{-m < y \leqslant t\} = \varnothing$ и $I_y(N,x,k,-m) = \mathbb{I}\{f_{PA}(x,\xi_N(\overline{S_k(N)})) \geqslant y\}$; если $t = m$, тогда $\{m < y \leqslant m\} = \varnothing$ и $I_y(N,x,k,m) = -\mathbb{I}\{f_{PA}(x,\xi_N(\overline{S_k(N)})) \leqslant y-1\}$, здесь $-m < y \leqslant m$, $y \in \mathbb{Y}$.

Следствие 1. *Условие* (1.28) *Теоремы 1 эквивалентно следующему:*

$$\sum_{k=1}^{K} \sum_{t \in \mathbb{Y}} \sum_{x \in \mathbb{X}(t,U)} L^{\top}(x) I(N,x,k,t) \to 0 \ \text{ п.н.},\ N \to \infty, \tag{1.42}$$

где $L^{\top}(x) := (L_{-m+1}(x),\ldots,L_m(x))$, $L_y(x)$, $y = -(m-1),\ldots,m$, *определены в* (1.18) *и* $\mathbb{X}(t,U) := (\mathbb{X} \setminus U) \cap \{x \in M : f(x) = t\}$.

Доказательство. Условие (1.28) может быть переписано в виде

$$\sum_{k=1}^{K} \sum_{t \in \mathbb{Y}} \sum_{x \in \mathbb{X}(t,U)} w^{\top}(x)\, Q\, \delta(N,x,k) \to 0 \ \text{ п.н.},\ N \to \infty. \tag{1.43}$$

Заметим, что для $x \in \mathbb{X}(t,U)$ выполняется равенство

$$w^{\top}(x)\, Q\, \delta(N,x,k) = w^{\top}(x) \sum_{y \in \mathbb{Y}} \mathbb{I}\{f_{PA}(x,\xi_N(\overline{S_k(N)})) = y\}(q(y) - q(t)), \tag{1.44}$$

потому что $\sum_{y\in\mathbb{Y}}\mathbb{I}\{f_{PA}(x,\xi_N(\overline{S_k(N)}))=y\}=1$, и Q является симметричной матрицей. Для $y,t\in\mathbb{Y}$ согласно (1.17) и (1.18) имеем

$$w^T(x)(q(y)-q(t))=\begin{cases}-L_{y+1}(x)-\ldots-L_t(x), & y<t,\\ 0, & y=t,\\ L_{t+1}(x)+\ldots+L_y(x), & y>t.\end{cases}\tag{1.45}$$

Если $t=m$, тогда $\sum_{t+1\leqslant r\leqslant y}L_r=0$; если $t=-m$, тогда $\sum_{y+1\leqslant r\leqslant -m}L_r=0$, так как сумма по пустому множеству равна нулю. Изменяя порядок суммирования, получаем

$$\begin{aligned}\sum_{y<t}\sum_{r=y+1}^{t}&\mathbb{I}\{f_{PA}(x,\xi_N(\overline{S_k(N)}))=y\}L_r(x)\\ &=\sum_{r=-m+1}^{t}\sum_{y=-m}^{r-1}\mathbb{I}\{f_{PA}(x,\xi_N(\overline{S_k(N)}))=y\}L_r(x)\\ &=\sum_{r=-m+1}^{t}\mathbb{I}\{f_{PA}(x,\xi_N(\overline{S_k(N)}))\leqslant r-1\}L_r(x).\end{aligned}\tag{1.46}$$

Аналогично, имеем

$$\begin{aligned}\sum_{y>t}\sum_{r=t+1}^{y}&\mathbb{I}\{f_{PA}(x,\xi_N(\overline{S_k(N)}))=y\}L_r(x)\\ &=\sum_{r=t+1}^{m}\mathbb{I}\{f_{PA}(x,\xi_N(\overline{S_k(N)}))\geqslant r\}L_r(x).\end{aligned}\tag{1.47}$$

Таким образом, из (1.46) и (1.47) следует, что

$$w^{\top}(x)\,Q\,\delta(N,x,k)=L^{\top}(x)I(N,x,k,t).\tag{1.48}$$

Комбинируя выражения (1.43) и (1.48), мы приходим к (1.42). Доказательство завершено. □

Следствие 2. *Пусть* ψ *– штрафная функция,* $f:\mathbb{X}\to\mathbb{Y}$*, и предсказательный алгоритм определяется функцией* f_{PA}*. Предположим, что для некоторого множества* $U\subset\mathbb{X}$ *выполняется условие* (1.27)*. Будем считать, что для каждого* $t\in\mathbb{Y}$ *и произвольного* $x\in\mathbb{X}(t,U)$ *существуют* $i=i(x),j=j(x)$*, принадлежащие* $\mathbb{Y}$*,* $i<j$*, такие, что*

$$i\leqslant f_{PA}(x,\xi_N(\overline{S_k(N)}))\leqslant j\ \ \textit{п.н. для}\ \ k=1,\ldots,K\tag{1.49}$$

при достаточно больших N. Тогда условие

$$L_{\min\{t,i\}+1}(x) = \ldots = L_{\max\{t,j\}}(x) = 0. \tag{1.50}$$

влечет выполнение (1.42).

Доказательство. Очевидно, что для каждого $x \in \mathbb{X}(t,U)$ мы имеем

$$L^{\top}(x)I(N,x,k,t) = \sum_{y=-m+1}^{\min\{t,i\}} L_y^{\top}(x)I_y(N,x,k,t) + \sum_{y=\max\{t,j\}+1}^{m} L_y^{\top}(x)I_y(N,x,k,t). \tag{1.51}$$

Ввиду (1.41) каждое слагаемое в правой части (1.51) будет равно нулю п.н. для достаточно больших N, так как $i \leqslant f_{PA}(x,\xi_N(\overline{S_k(N)})) \leqslant j$ п.н. для $k = 1,\ldots,K$ (если N достаточно велико). Принимая во внимание тот факт, что $\#\,\mathbb{X} < \infty$, мы получаем искомое утверждение. $\square$

Пример 1. Пусть ψ – штрафная функция, а f_{opt} задана в (1.20). Для $x \in \mathbb{X}$ и множества $W_N \subset \{1,\ldots,N\}$ введем случайный вектор $\widetilde{w}^{W_N}(x,\omega)$ с компонентами

$$\widetilde{w}_y^{W_N}(x,\omega) = \frac{\psi(y)}{\sharp W_N} \sum_{j \in W_N} I\{Y^j = y, X^j = x\}, \;\; y \in \mathbb{Y}, \tag{1.52}$$

здесь $0/0 := 0$ (если W_N пусто). Положим

$$f_{PA}(x,\xi_N(W_N),w) := \sum_{y \in \mathbb{Y}} y\, \mathbb{I}\{x \in \widetilde{A}_y^{W_N}(\omega)\}, \tag{1.53}$$

где $\xi_N(W_N) = \{\xi_i(\omega), \omega \in \Omega, i \in W_N\}$,

$$x \in \widetilde{A}_y^{W_N}(\omega) \Longleftrightarrow \begin{cases} \widetilde{L}_{-m+1}^{W_N}(x,\omega) \geqslant 0, & y = -m, \\ \widetilde{L}_{y+1}^{W_N}(x,\omega) \geqslant 0, \widetilde{L}_y^{W_N}(x,\omega) < 0, & y \neq \pm m, \\ \widetilde{L}_m^{W_N}(x,\omega) < 0, & y = m, \end{cases} \tag{1.54}$$

и

$$\widetilde{L}_y^{W_N}(x,\omega) := (\widetilde{w}^{W_N}(x,\omega))^{\top}\Delta(y). \tag{1.55}$$

Мы используем ω в (1.53)–(1.55) для того, чтобы подчеркнуть случайность рассматриваемых величин. Ясно, что можно определить $f_{PA}(x,v_m)$ для $x \in \mathbb{X}$, $v_m \in \mathbb{V}_m$, а затем ввести $f_{PA}(x,\xi_N(\overline{S_k(N)}))$, встречающуюся в (1.53).

Покажем, что f и f_{PA} удовлетворяют условиям следствия 2, если мы возьмем

$$U = \{x \in \mathbb{X} : L_y(x) \neq 0 \text{ для всех } y = -m+1,\ldots,m\}. \tag{1.56}$$

Можно проверить, что выполняется не только (1.27), но и

$$f_{PA}(x,\xi_N(W_N)) \to f(x) \quad \text{п.н.}, \ N \to \infty,$$

для всех $W_N \subset \{1,\dots,N\}$ таких, что $\#W_N \to \infty$ ($N \to \infty$). В самом деле, для всех $x \in \mathbb{X}$, $y \in \mathbb{Y}$, $y > -m$, и описанных можеств W_N имеем

$$\widetilde{L}_y^{W_N}(x,\omega) \to L_y(x) \quad \text{п.н.}, \ N \to \infty. \tag{1.57}$$

Пусть $x \in U$, тогда для каждого $y > -m$, $y \in \mathbb{Y}$, мы можем утверждать, что $L_y(x) < 0$ или $L_y(x) > 0$. Поэтому для почти всех $\omega \in \Omega$ существует такое $N_2 = N_2(\omega)$, что $\widetilde{L}_y(x,\xi_N(W_N),\omega) < 0$ или $\widetilde{L}_y(x,\xi_N(W_N),\omega) > 0$, соответственно, при $N > N_2(\omega)$. Таким образом, (1.54) влечет выполнение условия (1.27) для U, заданного в (1.56). Рассмотрим теперь $x \in \mathbb{X} \setminus U$. Тогда $L_v(x) = 0$ для некоторого $v \in \mathbb{Y}$, $v > -m$. В этом случае, согласно замечанию 2, в $\mathbb{Y}$ найдется такое подмножество $J = J(x)$ с $\#J(x) > 1$, что $x \in B_J$ (см. (1.10)). Тогда $v \in J(x)$. Ввиду (1.20) мы имеем, что $x \in \mathbb{X}(t,U)$, где $t = \min\{y : y \in J(x)\} - 1$. Для произвольного $k \in \{1,\dots,K\}$ и всех достаточно больших N

$$f_{PA}(x,\xi_N(\overline{S_k(N)})) \in J(x) \cup \{t\} \ \text{ п.н.}$$

В самом деле, согласно лемме 2 можно утверждать, что $L_y(x) \neq 0$ для $y \in \mathbb{Y} \setminus J(x)$, и, таким образом, для произвольного $k \in \{1,\dots,K\}$ и всех достаточно больших N в силу (1.54) и (1.57) значение $f_{PA}(x,\xi_N(\overline{S_k(N)}))$ не может принадлежать $\mathbb{Y} \setminus (J(x) \cup \{t\})$. Значит, мы можем применить следствие 2 с $i = t$ и $j = \max\{y : y \in J(x)\}$, потому что согласно лемме 2 такой выбор гарантирует выполнение (1.50). $\square$

Пример 2. Предположим теперь, что нам не известна штрафная функция ψ, но для каждого $k \in \{1,\dots,K\}$ существует последовательность ее сильно состоятельных оценок, то есть таких, что (1.25) выполняется. Тогда для $x \in \mathbb{X}$ зададим случайный вектор $\widetilde{w}^{W_N}(x,\omega)$ с компонентами

$$\widehat{w}_y^{W_N}(x,\omega) = \frac{\widehat{\psi}(y,\xi_N(W_N))}{\sharp W_N} \sum_{j \in W_N} \mathbb{I}\{Y^j = y, X^j = x\}, \ y \in \mathbb{Y}. \tag{1.58}$$

Для $x \in \mathbb{X}$ и $N \in \mathbb{N}$ рассмотрим случайный вектор $\widehat{L}^{W_N}(x,\omega)$ с компонентами

$$\widehat{L}_y^{W_N}(x,\omega) := (\widehat{w}^{W_N}(x,\omega))^\top \Delta(y), \ y \in \mathbb{Y}. \tag{1.59}$$

Введем $\widehat{A}_y^{W_N}(\omega)$ так же, как и в (1.54), где вместо $\widetilde{L}_y^{W_N}(x,\omega)$ используется $\widehat{L}_y^{W_N}(x,\omega)$, $y \in \mathbb{Y}$, $x \in \mathbb{X}$, $\omega \in \Omega$ и $N \in \mathbb{N}$. Положим

$$\widehat{f}_{PA}(x,\xi_N(W_N)) = \sum_{y\in\mathbb{Y}} y\, \mathbb{I}\{x \in \widehat{A}_y^{W_N}(\omega)\}. \tag{1.60}$$

Аналогично примеру 1 можно показать, что $f = f_{opt}$, где f_{opt} определена в (1.20), и $\widetilde{f}_{PA}$, заданная в (1.60), удовлетворяют условиям следствия 2.

В [68] был предложен следующий выбор штрафной функции ψ, когда переменная отклика Y принимает значения -1 и 1.

$$\psi(y) = c(\mathbb{P}(Y = y))^{-1}, \;\; y \in \{-1,1\}, \;\; c = const > 0. \tag{1.61}$$

В [25] было показано, что такой выбор имеет определенное обоснование. Здесь мы предположим, что $\mathbb{P}(Y = y) > 0$ при $y \in \{-1,1\}$, без потери общности можно считать, что $c = 1$ в (1.61). Для более общего случая $\mathbb{Y} = \{-m,\ldots,m\}$ ($m \in \mathbb{N}$) введем события

$$A_{W_N}(y) = \{Y^j \neq y,\, j \in W_N\}, \;\; N \in \mathbb{N}, \;\; k \in \{1,\ldots,K\}, \;\; y \in \mathbb{Y},$$

и случайные величины

$$\widehat{\psi}(y,\xi_N(W_N)) := \frac{\mathbb{I}\{\Omega \setminus A_{N,k}(y)\}}{\widehat{\mathbb{P}}_{W_N}(Y = y)}. \tag{1.62}$$

Для $W_N \subset \{1,\ldots,N\}$, $N \in \mathbb{N}$, $y \in \mathbb{Y}$, $C \subset \mathbb{X}$ положим

$$\widehat{\mathbb{P}}_{W_N}(Y = y|X \in C) := \frac{\sum_{j\in W_N} \mathbb{I}\{Y^j = y, X^j \in C\}}{\sum_{j\in W_N} \mathbb{I}\{X^j \in C\}}. \tag{1.63}$$

Когда $C = \mathbb{X}$, мы будем использовать обозначение $\widehat{\mathbb{P}}_{W_N}(Y = y)$. Согласно УЗБЧ для массивов (теорема 2) при произвольном $C \subset \mathbb{X}$ с $\mathbb{P}(X \in C) > 0$,

$$\widehat{\mathbb{P}}_{W_N}(Y = y|X \in C) \to \mathbb{P}(Y = y|X \in C) \;\text{ п.н.}, \;\; N \to \infty. \tag{1.64}$$

Тривиальные случаи $\mathbb{P}(Y = y) = 0$, $y \in \mathbb{Y}$, исключаются из рассмотрения, и мы формально полагаем, что $0/0 := 0$. Тогда

$$\begin{aligned}\widehat{\psi}(y,\xi_N(W_N)) - \psi(y) \\ = \frac{\mathbb{P}(Y = y) - \widehat{\mathbb{P}}_{W_N}(Y = y)}{\widehat{\mathbb{P}}_{W_N}(Y = y)\mathbb{P}(Y = y)}\mathbb{I}\{\Omega \setminus A_{W_N}(y)\} - \frac{1}{\mathbb{P}(Y = y)}\mathbb{I}\{A_{W_N}(y)\}.\end{aligned} \tag{1.65}$$

Ясно, что

$$\mathbb{I}\{A_{W_N}(y)\} \to 1 \quad \text{п.н.,} \ N \to \infty, \tag{1.66}$$

и выполняется следующее соотношение:

$$\frac{\mathbb{I}\{\Omega \setminus A_{W_N}(y)\}}{\widehat{\mathbb{P}}_{W_N}(Y=y)} \to \frac{1}{\mathbb{P}(Y=y)} \quad \text{п.н.,} \ N \to \infty. \tag{1.67}$$

Таким образом, используя (1.65) – (1.67), мы получаем, что для $k = 1, \ldots, K$ имеет место (1.25). □

Замечание 6. В медицинских приложениях функция отклика Y описывает состояние здоровья пациента. А именно, для бинарной переменной значения 1 и -1 соответственно означают "болен" и "здоров" ("контроль"). Если Y принимает значения в множестве $\{-1,0,1\}$, тогда значения 1 и -1 имеют аналогичный смысл, а значение 0 интерпретируется как "промежуточное состояние", когда невозможно определить болен человек или здоров. В данном важном случае тернарного отклика следствие 1 дает критерий выполнения (1.28) в терминах асимптотического поведения f_{PA} и свойств функций $L_0(x)$, $L_1(x)$ для $x \in \mathbb{X} \setminus U$.

Замечание 7. В [85] рассматривается случай, когда Y принимает конечный набор неупорядоченных значений.

Теперь мы перейдем к обсуждению выбора штрафной функции ψ. Как упоминалось ранее, выбор (1.61) оправдан в случае бинарной функции отклика Y со значениями -1 и 1. А именно, в [25] было показано, что если мы предположим, что f_{PA} не улавливает зависимость между Y и X вне "хорошего" множества U (то есть такого множества, что (1.27) выполняется), тогда независимость событий $\{Y = 1\}$ и $\{X = x\}$ для $x \in \mathbb{X} \setminus U$ естественным образом ведет (см. следствие 1 в [25]) к формуле (1.61). Однако, для Y, принимающего значения в множестве $\mathbb{Y} = \{-m, \ldots, m\}$ с $m \in \mathbb{N}$, ситуация более сложная. Если мы хотим (аналогично случаю бинарного Y) иметь $L_y(x) = 0$ для любого $y \in \mathbb{Y}$, $y > -m$, и $x \in \mathbb{X} \setminus U$, тогда можно видеть, что подобное требование эквивалентно соотношениям

$$\psi(y)\mathbb{P}(Y = y, X = x) = 0, \quad -m < y < m, \tag{1.68}$$

$$\psi(-m)\mathbb{P}(Y = -m, X = x) = \psi(m)\mathbb{P}(Y = m, X = x). \tag{1.69}$$

Таким образом, если мы предположим независимость событий $\{Y = y\}$ и $\{X = x\}$ для $y \in \mathbb{Y}$ и $x \in \mathbb{X} \setminus U$, то (1.69) выполняется, когда (1.61) справедливо

для $y \in \{-m,m\}$. В то же время, (1.68) означает, что $\psi(y)\mathbb{P}(Y = y) = 0$, если $\mathbb{P}(X \in \mathbb{X}\backslash U) > 0$. Следовательно, (1.68) верно, если $\psi(y) = 0$ при $\mathbb{P}(Y = y) \neq 0$. В этом случае для общего $\mathbb{Y}$ не удается обосновать выбор $\psi(y)$, приведенный в (1.61). Если $\mathbb{Y} = \{-1,0,1\}$, тогда выбор $\psi(y)$ согласно (1.61) для $y \in \{-1,1\}$ и $\psi(0) = 0$ может рассматриваться как ситуация, в которой мы ничего не теряем в "промежуточном" случае, соответствующем $Y = 0$. Заметим, что выбор ψ в (1.61) подходит, если мы хотим в значительной степени учесть редкие значения Y. Мы также подчеркиваем, что в следствии 2 мы не предполагаем, что $L_y(x) = 0$ для произвольных $y \in \mathbb{Y}$, $y > -m$, и $x \in \mathbb{X} \setminus U$.

1.2.3 Выбор набора значимых факторов.

Для многих моделей естественным предположением является то, что функция отклика Y зависит лишь от некоторых факторов $X_{k_1},\ldots,X_{k_r}$, где $1 \leqslant k_1 < \ldots < k_r \leqslant n$. Другими словами, для произвольных $x=(x_1,\ldots,x_n)\in M$ и $y \in \mathbb{Y}$,

$$\mathbb{P}(Y = y|X_1 = x_1,\ldots,X_n = x_n) = \mathbb{P}(Y = y|X_{k_1} = x_{k_1},\ldots,X_{k_r} = x_{k_r}), \quad (1.70)$$

если $P(X_1 = x_1,\ldots,X_n = x_n) \neq 0$. В рамках медицинских приложений это означает, что факторы $X_{k_1},\ldots,X_{k_r}$ могут рассматриваться, как оказывающие существенный вклад в развитие комплексного заболевания, в то время как влиянием остальных факторов можно пренебречь. Любой такой набор индексов $\{k_1,\ldots,k_r\}$ будем называть *значимым*. Ясно, что если набор $\{k_1,\ldots,k_r\}$ значимый и если $\{k_1,\ldots,k_r\} \subset \{m_1,\ldots,m_p\} \subset \{1,\ldots,n\}$, то $\{m_1,\ldots,m_p\}$ тоже является значимым. Отметим, что имеются и другие подходы к определению значимых, избыточных и взаи- модополняющих факторов, см., например, [70].

Для $r = 1,\ldots,n$ положим $\mathbb{X}_r = \{0,1\ldots,s\}^r$. Тогда $\mathbb{X} = \mathbb{X}_n$. Далее, будем писать $\alpha = (k_1,\ldots,k_r)$, $X_\alpha = (X_{k_1},\ldots,X_{k_r})$ и $x_\alpha = (x_{k_1},\ldots,x_{k_r})$, где $x_i \in \{0,\ldots,s\}$, $i = 1,\ldots,n$. Для $x \in M$ и $y \in \mathbb{Y}$ формула (1.70) может быть записана в виде

$$\mathbb{P}(Y = y|X = x) = \mathbb{P}(Y = y|X_\alpha = x_\alpha). \quad (1.71)$$

Здесь $\mathbb{P}(X = x_\alpha) \geqslant \mathbb{P}(X = x) > 0$, так как $x \in M$. Для $x \in \mathbb{X}$ и $y \in \mathbb{Y}$ определим вектор $w^\alpha(x)$ с компонентами

$$w_y^\alpha(x) = \begin{cases} \psi(y)\mathbb{P}(Y = y, X_\alpha = x_\alpha), & x \in M, \\ 0, & x \notin M. \end{cases}$$

Заметим, что (1.16) и (1.18) влекут $L_y(x) = 0$ для $x \in M$ и всех $y \in \mathbb{Y}$, $y > -m$ (здесь, как и ранее, $\mathbb{Y} = \{-m, \ldots, m\}$). Зададим функции $L_y^\alpha(x)$ аналогично (1.18), где вместо $w(x)$ используется $w^\alpha(x)$. Другими словами, для $x \in \mathbb{X}$

$$L_y^\alpha(x) = (w^\alpha(x))^\top \Delta(y) = w_{-m}^\alpha(x) + \ldots + w_{y-1}^\alpha(x) - w_y^\alpha(x) - \ldots - w_m^\alpha(x),$$

где $y \in \mathbb{Y}$, $y > -m$. Тогда (1.71) означает, что для всех $x \in \mathbb{X}$ и $y \in \{-m+1, \ldots, m\}$ мы имеем $L_y(x) = (w_y^\alpha(x))^\top \Delta(y)\mathbb{P}(X = x)/\mathbb{P}(X_\alpha = x_\alpha)$. Следовательно, $L_y(x)$ и $L_y^\alpha(x)$ для таких x и y принимают положительные или отрицательные значения, или равны нулю одновременно.

Если (1.71) выполнено, то, согласно (1.20), оптимальная функция f_{opt} запишется в виде

$$f^\alpha(x) = \sum_{y \in \mathbb{Y}} y\, \mathbb{I}\{x \in A_y^\alpha\}, \tag{1.72}$$

где

$$x \in A_y^\alpha \Longleftrightarrow \begin{cases} L_{-m+1}^\alpha(x) \geqslant 0, & y = -m, \\ L_{y+1}^\alpha(x) \geqslant 0, H_y^\alpha(x) < 0, & y \neq \pm m, \\ L_m^\alpha(x) < 0, & y = m. \end{cases} \tag{1.73}$$

Следовательно, для всех $x \in \mathbb{X}$ имеем $f_{opt}(x) = f^\alpha(x)$. В действительности, $f^\alpha(x)$ зависит только от x_α.

Теперь рассмотрим произвольное $\beta = (m_1, \ldots, m_r)$, где $1 \leqslant m_1 < \ldots < m_r \leqslant n$ и применим (1.72), (1.73) с β вместо α (мы не предполагаем, что набор $\{m_1, \ldots, m_r\}$ значимый). Таким образом, мы получаем функцию $f^\beta(x)$. Несложно проверить, что для каждого $y \in \mathbb{Y}$, $y > -m$, функция $L_y^\beta(x)$, $x \in \mathbb{X}$, имеет такие же свойства, что и $L_y(x)$, потому что $w_y(x) = w_y^\beta(x)\frac{\mathbb{P}(X=x)}{\mathbb{P}(X_\beta = x_\beta)}$, где

$$w_y^\beta(x) = \begin{cases} \psi(y)\mathbb{P}(Y = y, X_\beta = x_\beta), & x \in M, \\ 0, & x \notin M, \end{cases}$$

и $\mathbb{P}(X_\beta = x_\beta) > 0$ при $x \in M$. Следовательно, A_y^β, $y \in \mathbb{Y}$, образуют разбиение $\mathbb{X}$, и мы имеем корректно определенную функцию $f^\beta(x)$ для $x \in \mathbb{X}$. Более того, если набор индексов α является значимым, то из оптимальности f^α следует, что для любого $\beta = (m_1, \ldots, m_r)$ с $1 \leqslant m_1 < \ldots < m_r \leqslant n$ выполняется неравенство

$$Err(f^\alpha) \leqslant Err(f^\beta). \tag{1.74}$$

Пусть ψ – штрафная функция. Для произвольного $\beta = (m_1, \ldots, m_r)$, где $1 \leqslant m_1 < \ldots, < m_r \leqslant n$, $x \in \mathbb{X}$ и множества $W_N \subset \{1, \ldots, N\}$, введем случайный вектор $\widetilde{w}^{\beta, W_N}(x, \omega)$ с компонентами

$$\widetilde{w}_y^{\beta, W_N}(x, \omega) = \frac{\psi(y)}{\sharp W_N} \sum_{j \in W_N} \mathbb{I}\{Y^j = y, X_\beta^j = x_\beta\}, \ y \in \mathbb{Y}. \tag{1.75}$$

Пусть предсказательный алгоритм определяется такой функцией f_{PA}^β, что

$$\widetilde{f}_{PA}^\beta(x, \xi_N(W_N)) = \sum_{y \in \mathbb{Y}} y \, \mathbb{I}\{x \in \widetilde{A}_y^{\beta, W_N}(\omega)\}, \tag{1.76}$$

где

$$x \in \widetilde{A}_y^{\beta, W_N}(\omega) \Longleftrightarrow \begin{cases} \widetilde{L}_{-m+1}^{\beta, W_N}(x, \omega) \geqslant 0, & y = -m, \\ \widetilde{L}_{y+1}^{\beta, W_N}(x, \omega) \geqslant 0, \widetilde{L}_y^{\beta, W_N}(x, \omega) < 0, & y \neq \pm m, \\ \widetilde{L}_m^{\beta, W_N}(x, \omega) < 0, & y = m, \end{cases} \tag{1.77}$$

и

$$\widetilde{L}_y^{\beta, W_N}(x, \omega) := (\widetilde{w}^{\beta, W_N}(x, \omega))^\top \Delta(y). \tag{1.78}$$

Мы используем ω в (1.76) – (1.78) для того, чтобы подчеркнуть случайность рассматриваемых величин.

Лемма 3. *Пусть $f = f^\beta$ задается (1.72) (с β вместо α). Тогда для произвольных $\beta = (m_1, \ldots, m_r)$, $1 \leqslant m_1 < \ldots < m_r \leqslant n$, и $f_{PA} = \widetilde{f}_{PA}^\beta$ соотношение (1.27) выполняется, когда множества $W_N \subset \{1, \ldots, N\}$ таковы, что $\# W_N \to \infty$ при $N \to \infty$. Более того, если условие (1.40) выполнено, тогда $\widehat{Err}_K(f_{PA}, \xi_N)$ – асимптотически несмещенная оценка $Err(f)$ при $N \to \infty$.*

Доказательство. Для $u \in \mathbb{X}_r$ и $v_m \in (\mathbb{X} \times \mathbb{Y})^m$ введем функции

$$f^*(u) := f(x), \ \ f_{PA}^*(u, v_m) := f_{PA}(x, v_m), \tag{1.79}$$

где $x_\beta = u$. Заметим, что для каждого $x \in \mathbb{X}$ величина $w^\beta(x)$ зависит только от x_β. Следовательно, f^* и f^*_{PA} определены корректно, так как $f(x) = f(z)$ и $f_{PA}(x,v_m) = f_{PA}(z,v_m)$ для произвольных $v_m \in (\mathbb{X}\times\mathbb{Y})^m$ $(1 \leqslant m \leqslant N)$, $x,z \in \mathbb{X}$ таких, что $x_\beta = z_\beta$. Положим

$$U = \{x \in \mathbb{X} : L^\beta_y(x) \neq 0 \ \text{ для всех } \ y \in \mathbb{Y},\ y > -m\}. \tag{1.80}$$

Для $x,z \in \mathbb{X}$ и $y \in \mathbb{Y}$ имеем $L^\beta_y(x) = L^\beta_y(z)$, если $x_\beta = z_\beta$. Введем $U^* = \{x_\beta : x \in U\}$. Тогда мы можем применить следствие 2 для f^* и f^*_{PA}, определенных на $\mathbb{X}_r$ с X_β, U^*, $\mathbb{X}_r(t,U^*)$ вместо X, U и $\mathbb{X}(t,U)$, соответственно. Для того, чтобы получить второе утверждение данной леммы, мы используем замечание 5. Доказательство завершено. □

Теперь аналогично (1.58) определим

$$\widehat{w}^{\beta,W_N}_y(x,\omega) = \frac{\widehat{\psi}(y,\xi_N(W_N))}{\sharp W_N} \sum_{j\in W_N} \mathbb{I}\{Y^j = y, X^j_\beta = x_\beta\},\ y \in \mathbb{Y}. \tag{1.81}$$

Для $x \in \mathbb{X}$, $y \in \mathbb{Y}$, $y > -m$, и $N \in \mathbb{N}$ положим

$$\widehat{L}^{\beta,W_N}_y(x,\omega) := (\widehat{w}^{\beta,W_N}(x,\omega))^\top \Delta(y). \tag{1.82}$$

Введем

$$\widehat{f}^\beta_{PA}(x,\xi_N(W_N)) = \sum_{y\in\mathbb{Y}} y\, \mathbb{I}\{x \in \widehat{A}^{\beta,W_N}_y(\omega)\}, \tag{1.83}$$

где $\widehat{A}^{\beta,N}_y(\omega)$ определяется, как и (1.77), с заменой $\widetilde{L}^{\beta,W_N}_y(x,\omega)$ на $\widehat{L}^{\beta,W_N}_y(x,\omega)$.

Замечание 8. Легко видеть, что утверждение леммы 3 выполняется, если мы будем использовать $f_{PA} = \widehat{f}^\beta_{PA}$ вместо $f_{PA} = f^\beta_{PA}$.

Теорема 3. *Пусть $\alpha = (k_1,\dots,k_r)$ является значимым набором $\{k_1,\dots,k_r\} \subset \{1,\dots,n\}$. Тогда для произвольного $\varepsilon > 0$ и всех $\beta = (m_1,\dots,m_r)$ с $\{m_1,\dots,m_r\} \subset \{1,\dots,n\}$ выполняется следующее неравенство:*

$$\widehat{Err}_K(\widehat{f}^\alpha_{PA}) \leqslant \widehat{Err}_K(\widehat{f}^\beta_{PA}) + \varepsilon \ \ \textit{п.н.} \tag{1.84}$$

для всех достаточно больших N.

Доказательство. Ввиду замечания 8 утверждение теоремы немедленно следует из следствия леммы 3 и соотношения (1.74). □

Из теоремы 3 вытекает, что для дальнейшего анализа естественно в качестве значимого набора выбирать такие наборы $\{k_1,\ldots,k_r\} \subset \{1,\ldots,n\}$, что $\widehat{Err}_K(\widehat{f}^{\alpha}_{PA})$ с $\alpha = (k_1,\ldots,k_r)$ имеет минимальное (или почти минимальное) значение среди всех $\widehat{Err}_K(\widehat{f}^{\beta}_{PA})$, где $\beta = (m_1,\ldots,m_r)$ и $\{m_1,\ldots,m_r\} \subset \{1,\ldots,n\}$. Важно отметить, что мы установили сходимость почти наверное в (1.84), так как нам необходимо сравнивать $\widehat{Err}_K(\widehat{f}^{\beta}_{PA})$ для различных $\beta = (m_1,\ldots,m_r)$ одновременно. Если бы была доказана только сходимость по вероятности, то потребовалось бы использовать неравенство Бонферрони. Это не позволило бы осуществить с большой вероятностью одновременное сравнение многих наборов факторов. Также заметим, что в целом ряде задач имеется набор из небольшого количества объясняющих факторов $X_{k_1},\ldots,X_{k_r}$, в то время как общее количество значимых факторов $X_1,\ldots,X_n$ может быть достаточно велико.

1.3 Состоятельность MDR-EFE метода в случае непрерывных объясняющих переменных

Пусть теперь X принимает значения в $\mathbb{X} = \mathbb{R}^n$. Будем считать, что у случайного вектора X есть плотность ρ по мере Лебега в $\mathbb{R}^n$. Положим $M = \{x \in \mathbb{X} : \rho_X(x) > 0\}$. В данном разделе мы будем рассматривать бинарную функцию отклика Y с $\mathbb{Y} = \{-1,1\}$. Введем обозначение

$$F(x) = \psi(-1)\mathbb{P}(Y = -1|X = x) - \psi(1)\mathbb{P}(Y = 1|X = x), \quad x \in M, \qquad (1.85)$$

где $\mathbb{P}(Y = y|X) = \mathbb{E}(\mathbb{I}\{Y = y\}|X) = g(X)$, g – борелевская функция, а запись $\mathbb{P}(Y = y|X = x)$ означает, что берется $g(x)$ ($y \in \mathbb{Y}$, $x \in \mathbb{X}$). Об условных математических ожиданиях см., например, [6], гл. 2, раздел 8).

1.3.1 Доказательство асимптотической состоятельности оценок функционала ошибки

Как и прежде, нас интересуют асимптотические свойства оценки функционала ошибки, введенного формулой (1.7). Ясно, что во многом свойства этой оценки определяются свойствами оценок $\widehat{\mathbb{P}}(Y = 1|X = x)$ и $\widehat{\gamma}(\psi)$.

Замечание 9. В следующей теореме будет фигурировать функция от ω

$$\sup_{x\in\mathbb{R}^n} |\widehat{\mathbb{P}}_N(Y = 1|X = x) - \mathbb{P}(Y = 1|X = x)|,$$

при $N \in \mathbb{N}$. Мы будем предполагать, что эта верхняя грань является случайной величиной. В следующем разделе приведем примеры оценок $\widehat{\mathbb{P}}_N(Y = 1|X = x)$, для которых это предположение выполняется.

Для случайных величин нам также потребуется понятие *сходимости вполне*, которое было определено в [42].

Определение 1. *Последовательность случайных величин* $\{\xi_n\}_{n\in\mathbb{N}}$ *сходится вполне к случайной величине* ξ *при* $n \to \infty$, *если для любого* $\varepsilon > 0$ *сходится ряд*

$$\sum_n \mathbb{P}(|\xi_n - \xi| \geqslant \varepsilon).$$

Несложно показать, что сходимость вполне влечет сходимость почти наверное (см. [5], глава 3, с. 156). Сформулируем основной результат данного раздела.

Теорема 4. *Пусть* $\xi^1,\xi^2,\ldots$ – *последовательность н.о.р. случайных векторов, имеющих тот же закон распределения, что и вектор* (X, Y), ψ – *некоторая штрафная функция,* $f : \mathbb{X} \to \{-1,1\}$ *и* f_{PA} *задает алгоритм предсказания согласно* (1.8). *Предположим, что при* $N \to \infty$ *последовательность случайных величин* $\widehat{\gamma}_N(\psi)$ *сходится вполне к* $\gamma(\psi)$. *Пусть у вектора* X *существует плотность* $\rho(x)$, $x \in \mathbb{R}^n$. *Кроме того, пусть функция*

$$\kappa(\delta) = \int_{\mathbb{R}^n} \mathbb{I}\{x : \mathbb{P}(Y = 1|X = x) \in [\gamma(\psi) - \delta,\gamma(\psi) + \delta]\}\, \rho(x)dx \qquad (1.86)$$

непрерывна в окрестности точки $\delta = 0$. *Потребуем*[1], *чтобы* $\kappa(0) = 0$, *а также чтобы последовательность* $\sup_{x\in\mathbb{R}^n} |\widehat{\mathbb{P}}_N(Y = 1|X = x) - \mathbb{P}(Y = 1|X = x)|$

[1] с учетом замечания 9 выражение в (1.87) определено корректно.

сходилась к 0 *вполне при* $N \to \infty$, *то есть, выполняется неравенство*

$$\sum_{N=1}^{\infty} \mathbb{P}\left(\sup_{x \in \mathbb{R}^n} |\widehat{\mathbb{P}}_N(Y=1|X=x) - \mathbb{P}(Y=1|X=x)| > \delta \right) < \infty \tag{1.87}$$

для произвольного $\delta > 0$. *Тогда справедливо соотношение*

$$\widehat{Err}_K(f_{PA},\xi_N) \to Err(f) \quad \text{п.н., } N \to \infty.$$

Доказательство. Достаточно убедиться в том, что

$$\frac{2}{K} \sum_{k=1}^{K} \sum_{y \in \{-1,1\}} \psi(y) \sum_{j \in S_k(N)} \frac{\mathbb{I}\left\{ Y^j = y, f_{PA}\big(X^j, \xi_N(\overline{S_k(N)})\big) \neq y \right\}}{\# S_k(N)} \to Err(f) \tag{1.88}$$

почти наверное при $N \to \infty$. Действительно, $\widehat{\psi}(y,S_k(N)) \to \psi(y)$ п.н. для $y \in \{-1,1\}$, когда $N \to \infty$, и при этом для всех $\omega \in \Omega$ и $k = 1, \ldots, K$ выполнено неравенство

$$\frac{1}{\# S_k(N)} \sum_{j \in S_k(N)} \mathbb{I}\left\{ Y^j = y, f_{PA}\big(X^j, \xi_N(\overline{S_k(N)})\big) \neq y \right\} \leqslant 1.$$

Согласно УЗБЧ для массивов (теорема 2) при каждом $y \in \{-1,1\}$ имеем

$$\frac{1}{\# S_k(N)} \sum_{j \in S_k(N)} \mathbb{I}\left\{ Y^j = y, f(X^j) \neq y \right\} \to \mathbb{P}(Y = y, f(x) \neq y) \quad \text{п.н., } N \to \infty.$$

Условия УЗБЧ для серий выполнены, так как случайные величины $\mathbb{I}\left\{ Y^j = y, f(X^j) \neq y \right\} / \# S_k(N)$, $j \in S_k(N)$, независимы, одинаково распределены и ограничены 1.

Следовательно, учитывая формулу (1.5) для $Err(f)$, при каждом $k = 1, \ldots, K$ получаем

$$2 \sum_{y \in \{-1,1\}} \sum_{j \in S_k(N)} \frac{\mathbb{I}\left\{ Y^j = y, f(X^j) \neq y \right\}}{\# S_k(N)} \to Err(f) \quad \text{п.н., } N \to \infty. \tag{1.89}$$

Для $y \in \{-1,1\}$, $N \in \mathbb{N}$, $j \in S_k(N)$ и $k = 1, \ldots, K$ введем случайные величины

$$F_{N,k}(x,y) := \mathbb{I}\left\{ f_{PA}\big(x, \xi_N(\overline{S_k(N)})\big) \neq y \right\} - \mathbb{I}\left\{ f(x) \neq y \right\}. \tag{1.90}$$

$$Q_{N,k}(y) := \frac{1}{\# S_k(N)} \sum_{j \in S_k(N)} \mathbb{I}\left\{ Y^j = y \right\} F_{N,k}(X^j, y). \tag{1.91}$$

Принимая во внимание (1.89), (1.88) и (1.90), видим, что соотношение (1.88) эквивалентно следующему:

$$\sum_{k=1}^{K} \sum_{y\in\{-1,1\}} \psi(y) Q_{N,k}(y) \to 0 \quad \text{п.н., } N \to \infty.$$

В случае, когда вектор X принимал конечное множество значений, доказательство существенно опиралось на конечность множества $\mathbb{X}$. А именно, утверждалось, что существует такое достаточно большое $N_0 < \infty$, что при $N > N_0$ имеем $Q_{N,k}(y) = 0$ для каждого $y \in \{-1,1\}$, $k = 1,\ldots,K$ и почти всех $\omega \in \Omega$.

В непрерывном случае мы воспользуемся следствием из леммы Бореля-Кантели (см., например, [6], гл. 10, с. 357). Согласно этому следствию, сходимость $Q_{N,k}(y) \to 0$ п.н., когда $N \to \infty$, имеет место, если при каждом $\varepsilon > 0$

$$\sum_{N=1} \mathbb{P}(|Q_{N,k}(y)| > \varepsilon) < \infty. \tag{1.92}$$

Для произвольных $\varepsilon > 0, k = 1,\ldots,K, y \in \{-1,1\}$ и $N \in \mathbb{N}$ найдем оценку сверху слагаемого $\mathbb{P}(|Q_{N,k}(y)| > \varepsilon)$, фигурирующего в (1.92). По определению (1.91) величина $Q_{N,k}(y)$ при каждом k представляет сумму зависимых одинаково распределенных случайных величин. Зависимость обусловлена тем, что оценки $f_{PA}(X^j,\xi_N(\overline{S_k(N)})$ при разных j зависит от набора случайных величин $\xi_N(\overline{S_k(N)})$. Кроме того, заметим что математическое ожидание слагаемых в (1.91) конечно, но, вообще говоря, необязательно равно нулю. Идея первой части доказательства состоит в том, чтобы от исследования суммы нецентрированных зависимых случайных величин перейти к изучению определенной суммы центрированных случайных величин, являющихся мартингалом относительно некоторой фильтрации.

Покажем, что для любых $i,j \in S_k(N)$ при произвольном $k = 1,\ldots,K$ справедливо равенство условных математических ожиданий

$$\mathbb{E}\Big(\mathbb{I}\left\{Y^i = y\right\} F_{N,k}(X^i,y)\Big|\xi_N(\overline{S_k(N)})\Big) = \mathbb{E}\Big(\mathbb{I}\left\{Y^j = y\right\} F_{N,k}(X^j,y)\Big|\xi_N(\overline{S_k(N)})\Big) \quad \text{п.н.} \tag{1.93}$$

Это соотношение вытекает из утверждения, приведенного в [7], с. 306. Пусть случайные векторы ξ и η независимы, и $\varphi = \varphi(x,y)$ такая борелевская функция, что $\mathbb{E}|\varphi(\xi,\eta)| < \infty$, тогда

$$\mathbb{E}(\varphi(\xi,\eta)|\eta = y) = \mathbb{E}(\varphi(\xi,y)) \quad \mathbb{P}_\eta\text{-п.н.} \tag{1.94}$$

Для интегрируемой случайной величины ζ существует такая борелевская функция $g : \mathbb{R} \to \mathbb{R}$, что $\mathbb{E}(\zeta|\eta) = g(\eta)$. Тогда, если $\zeta = \varphi(\xi,\eta)$, где ξ и η – независимые векторы, φ – ограниченная борелевская функция, то $g(x) = \mathbb{E}(\varphi(\xi,x))$. Последнее означает, что $\mathbb{E}(\varphi(\xi,\eta)|\eta = x) = g(x)$. Если ξ и γ имеют одинаковое распределение, то при каждом x величины $\varphi(\xi,x)$ и $\varphi(\gamma,x)$ будут одинаково распределены. Поэтому при каждом x получаем $\mathbb{E}\varphi(\xi,x) = \mathbb{E}\varphi(\gamma,x)$. Остается учесть, что если γ и η независимы, то $\mathbb{E}(\varphi(\gamma,\eta)|\eta = x) = \mathbb{E}(\varphi(\gamma,x)) = g(x)$.

В нашем случае мы имеем дело со случайными векторами ξ^i, ξ^j, $\{\xi^k\}_{k\in\overline{S_k(N)}}$, где $i,j \in S_k(N)$, которые независимы в совокупности (про условные математические ожидания вектором см.

Тогда в силу сказанного выше для ограниченной борелевской функции φ справедливо:

$$\mathbb{E}(\varphi(\xi^i,\{\xi^k\}_{k\in\overline{S_k(N)}})|\{\xi^k\}_{k\in\overline{S_k(N)}}) = \mathbb{E}(\varphi(\xi^j,\{\xi^k\}_{k\in\overline{S_k(N)}})|\{\xi^k\}_{k\in\overline{S_k(N)}}) \quad \text{п.н.},$$

где $i,j \in S_k(N)$. Вспоминая, что $\xi^i = (X^i,Y^i)$, $i = 1,\ldots,N$, и полагая

$$\varphi(\xi^i,\{\xi^k\}_{k\in\overline{S_k(N)}}) = \mathbb{I}\left\{Y^i = y\right\} F_{N,k}(X^i,y),$$

мы получим соотношение (1.93). Здесь следует отметить, что по построению $F_{N,k}$ зависит от $\{\xi^k\}_{k\in\overline{S_k(N)}}$.

Обозначим $\mu_{N,k}(y)$ условное математическое ожидание, фигурирующее в (1.93). Оно одинаково для всех $i,j \in S_k(N)$. Пусть $S_k(N) = \{j_1,\ldots,j_{\#S_k(N)}\}$. Чтобы не усложнять запись, мы не указываем зависимость этих индексов от k и N. Тогда в силу (1.91) имеем

$$Q_{N,k}(y) = \frac{1}{\#S_k(N)} \sum_{i=1}^{\#S_k(N)} \Big(\mathbb{I}\left\{Y^{j_i} = y\right\} F_{N,k}(X^{j_i},y) - \mu_{N,k}(y)\Big) + \mu_{N,k}(y). \tag{1.95}$$

Рассмотрим набор случайных величин

$$\eta_{N,k,m} := \frac{1}{\#S_k(N)} \sum_{i=1}^{m} \Big(\mathbb{I}\left\{Y^{j_i} = y\right\} F_{N,k}(X^{j_i},y) - \mu_{N,k}(y)\Big)$$

для $m = 1,\ldots,\#S_k(N)$. Кроме того, положим $\eta_0 = 0$ почти наверное. Несложно видеть, что $\{\eta_{N,m}\}$, $m = 0,1,\ldots,\#S_k(N)$, – мартингал относительно фильтрации

$$\mathfrak{F}_{N,k,m} := \sigma\left\{\xi_N(\overline{S_k(N)}) \cup \{\xi_{j_1},\ldots,\xi_{j_m}\}\right\}.$$

Действительно, конечность первого абсолютного момента следует из ограниченности функций $\mathbb{I}\{\cdot\}$ и $F_{N,k}(\cdot,\cdot)$. Кроме того, опуская для удобства индексы N и k у $\eta_{N,k,m}$ и $\mathfrak{F}_{N,k,m}$ имеем

$$\begin{aligned}\mathbb{E}\Big(\eta_{m+1}\Big|\mathfrak{F}_m\Big) &= \mathbb{E}\Big(\frac{1}{\#S_k(N)}\sum_{i=1}^{m+1}\big(\mathbb{I}\left\{Y^{j_i}=y\right\}F_{N,k}(X^{j_i},y)-\mu_{N,k}(y)\big)\Big|\mathfrak{F}_m\Big)\\ &= \mathbb{E}\Big(\frac{1}{\#S_k(N)}\sum_{i=1}^{m}\big(\mathbb{I}\left\{Y^{j_i}=y\right\}F_{N,k}(X^{j_i},y)-\mu_{N,k}(y)\big)\Big|\mathfrak{F}_m\Big)\\ &\quad+\mathbb{E}\Big(\frac{1}{\#S_k(N)}\big(\mathbb{I}\left\{Y^{j_{m+1}}=y\right\}F_{N,k}(X^{j_{m+1}},y)-\mu_{N,k}(y)\big)\Big|\mathfrak{F}_m\Big). \quad (1.96)\end{aligned}$$

Заметим, что $\mu_{N,k}(y)$ по определению условного математического ожидания является измеримой случайной величиной относительно сигма-алгебры $\sigma\{\xi_N(\overline{S_k(N)})\} \subset \mathfrak{F}_m$ для $m = j_1,\ldots,j_{\#S_k(N)}$. Далее, случайная величина $F_{N,k}(X^{j_i},y)$ есть борелевская функция случайных величин $\xi_N(\overline{S_k(N)})$ и X^{j_i}. Действительно, согласно (1.90) функция $F_{N,k}(X^{j_i},y)$ является композицией борелевских функций и случайной величины $f_{PA}(x,\xi_N(\overline{S_k(N)}))$. Поэтому, очевидно, $F_{N,k}(X^{j_i},y)$ измерима относительно $\mathfrak{F}_m$, $i \geqslant m$. Следовательно, мы можем продолжить (1.96):

$$\begin{aligned}&\mathbb{E}\Big(\eta_{m+1}\Big|\mathfrak{F}_m\Big) = \eta_m + \mathbb{E}\Big(\frac{1}{\#S_k(N)}\big(\mathbb{I}\left\{Y^{j_{m+1}}=y\right\}F_{N,k}(X^{j_{m+1}},y)-\mu_{N,k}(y)\big)\Big|\mathfrak{F}_m\Big)\\ &= \eta_m + \mathbb{E}\Big(\frac{1}{\#S_k(N)}\mathbb{I}\left\{Y^{j_{m+1}}=y\right\}F_{N,k}(X^{j_{m+1}},y)\Big|\mathfrak{F}_m\Big) - \frac{1}{\#S_k(N)}\mathbb{E}\Big(\mu_{N,k}(y)\Big|\mathfrak{F}_m\Big)\\ &= \eta_m+\mathbb{E}\Big(\frac{1}{\#S_k(N)}\mathbb{I}\left\{Y^{j_{m+1}}=y\right\}F_{N,k}(X^{j_{m+1}},y)\Big|\sigma\left\{\xi_N(\overline{S_k(N)})\cup\{\xi_{j_1},\ldots,\xi_{j_m}\}\right\}\Big)\\ &\qquad-\frac{\mu_{N,k}(y)}{\#S_k(N)}. \quad (1.97)\end{aligned}$$

Теперь воспользуемся следующей леммой.

Лемма 4. *Пусть U,W,Z – независимые случайные векторы, а f – некоторая ограниченная борелевская функция. Тогда справедливо*

$$\mathbb{E}(f(Z,U)|U,W) = \mathbb{E}(f(Z,U)|U) \quad \textit{п.н.} \quad (1.98)$$

Доказательство. Заметим, что $\mathbb{E}(f(Z,U)|U)$ является измеримой функцией относительно сигма-алгебры $\sigma\{U\}$. Следовательно, эта же функция измерима и относительно $\sigma\{U,W\}$. Таким образом, требуется проверить, что для каждого $A \in \sigma\{U,W\}$ верно равенство

$$\mathbb{E}(\mathbb{E}(f(Z,U)|U)\mathbb{I}_A) = \mathbb{E}(f(Z,U)\mathbb{I}_A). \tag{1.99}$$

Возьмем вместо $\mathbb{I}_A$ функцию $h(U)g(W)$, где h и g – ограниченные борелевские функции в соответствующих пространствах. Тогда в силу независимости U,W и Z правая часть (1.99) запишется следующим образом:

$$\mathbb{E}(f(Z,U)h(U)g(W)) = \mathbb{E}(f(Z,U)h(U))\mathbb{E}g(W).$$

При этом левая часть (1.99) равна

$$\mathbb{E}(\mathbb{E}(f(Z,U)|U)h(U)g(W)) = \mathbb{E}(\mathbb{E}(h(U)f(Z,U)|U)g(W))$$

$$= \mathbb{E}(\mathbb{E}(h(U)f(Z,U)|U))\mathbb{E}(g(W))$$

$$= \mathbb{E}(\mathbb{E}(h(U)f(Z,U))\mathbb{E}g(W)).$$

Согласно лемме 2 на с. 182 [2] рассмотрение $\mathbb{I}_A$ в (1.99) сводится к произведенному выше с помощью должных аппроксимаций конечными линейными комбинациями функций, измеримых относительно $\sigma\{U,W\}$. Доказательство леммы завершено. □

Воспользуемся тем, что $\xi^{j_{m+1}}$, $\xi_N(\overline{S_k(N)})$ и $\{\xi_{j_1},\ldots,\xi_{j_m}\}$ независимы. Положим $Z = \xi^{j_{m+1}}$, $W = \{\xi_{j_1},\ldots,\xi_{j_m}\}$ и $U = \xi_N(\overline{S_k(N)})$. Введем

$$f(Z,U) = \frac{1}{\#S_k(N)}\mathbb{I}\left\{Y^{j_{m+1}} = y\right\} F_{N,k}(X^{j_{m+1}},y).$$

Тогда f является ограниченной борелевской функцией. Ограниченность следует из ограниченности функций $\mathbb{I}$ и $F_{N,k}$, $k = 1,\ldots,K$, $N \in \mathbb{N}$. $F_{N,k}$ по модулю не превосходит 1, так как является разностью двух индикаторных функций. Из доказанной выше леммы вытекает, что

$$\mathbb{E}\Big(\frac{1}{\#S_k(N)}\mathbb{I}\left\{Y^{j_{m+1}} = y\right\} F_{N,k}(X^{j_{m+1}},y)\Big|\sigma\left\{\xi_N(\overline{S_k(N)}) \cup \{\xi_{j_1},\ldots,\xi_{j_m}\}\right\}\Big)$$

$$= \mathbb{E}\Big(\frac{1}{\#S_k(N)}\mathbb{I}\left\{Y^{j_{m+1}} = y\right\} F_{N,k}(X^{j_{m+1}},y)\Big|\sigma\left\{\xi_N(\overline{S_k(N)})\right\}\Big).$$

Тогда

$$\mathbb{E}\Big(\eta_{m+1}\Big|\mathfrak{F}_m\Big)=\eta_m+\mathbb{E}\Big(\frac{1}{\#S_k(N)}\mathbb{I}\left\{Y^{j_{m+1}}=y\right\}F_{N,k}(X^{j_{m+1}},y)\Big|\sigma\left\{\xi_N(\overline{S_k(N)})\right\}\Big)$$
$$-\frac{\mu_{N,k}(y)}{\#S_k(N)}=\eta_m+\frac{\mu_{N,k}(y)}{\#S_k(N)}-\frac{\mu_{N,k}(y)}{\#S_k(N)}=\eta_m. \quad (1.100)$$

Для произвольного $\varepsilon>0$ имеем

$$\mathbb{P}\big(\big|Q_{N,k}(y)\big|>\varepsilon\big)=\mathbb{P}\left(\big|\eta_{\#S_k(N)}+\mu_{N,k}(y)\big|>\varepsilon\right)$$
$$\leqslant\mathbb{P}\left(\big|\eta_{\#S_k(N)}\big|>\varepsilon/2\right)+\mathbb{P}\left(\big|\mu_{N,k}(y)\big|>\varepsilon/2\right). \quad (1.101)$$

Рассмотрим отдельно каждое слагаемое. Оценим сверху $\mathbb{P}\left(|\eta_{\#S_k(N)}|>\varepsilon/2\right)$ с помощью неравенства Хефдинга-Азума для мартингалов.

Неравенство Хефдинга-Азума для мартингалов [12]. Пусть η_m, $m=0,1,2,\ldots$ – мартингал относительно некоторой фильтрации и $|\eta_m-\eta_{m-1}|<c_m$ почти наверное для некоторых положительных констант c_m и всех рассматриваемых m. Тогда для любых натуральных N и всех положительных t справедливо неравенство

$$\mathbb{P}\left(\big|\eta_N-\eta_0\big|\geqslant t\right)\leqslant 2\exp\Big(\frac{-t^2}{2\sum_{j=1}^N c_j^2}\Big).$$

Применительно к нашему случаю имеем

$$\big|\eta_m-\eta_{m-1}\big|=\frac{1}{\#S_k(N)}\big|\mathbb{I}\left\{Y^{j_m}=y\right\}F_{N,k}(X^{j_m},y)-\mu_{N,k}(y)\big|\leqslant\frac{2}{\#S_k(N)}, \quad (1.102)$$

Здесь мы воспользовались тем, что $F_{N,k}(\cdot,\cdot)\in[-1,1]$ и $\mu_{N,k}(\cdot)\in[-1,1]$. Таким образом,

$$\mathbb{P}\left(\big|\eta_{\#S_k(N)}\big|>\varepsilon/2\right)$$
$$\leqslant 2\exp\Big(-\frac{\varepsilon^2}{2\sum_{j=1}^{\#S_k(N)}4/(\#S_k(N))^2}\Big)=2\exp\Big(-\frac{\varepsilon^2*\#S_k(N)}{8}\Big). \quad (1.103)$$

Поскольку $\#S_k(N)\geqslant[N/K]$, экспоненциальный ряд из слагаемых, фигурирующих в правой части (1.103), сходится, и мы имеем

$$\sum_{N=1}^{\infty}\mathbb{P}\left(\big|\eta_{\#S_k(N)}\big|>\varepsilon/2\right)<\infty. \quad (1.104)$$

Теперь разберемся со вторым слагаемым из (1.101):

$$\mathbb{P}\left(\left|\mu_{N,k}(y)\right| > \varepsilon/2\right) = \mathbb{P}\left(\left|\mathbb{E}\left(\mathbb{I}\left\{Y^{j_1} = y\right\} F_{N,k}(X^{j_1},y)\middle|\xi_N(\overline{S_k(N)})\right)\right| > \varepsilon/2\right). \quad (1.105)$$

Заметим, что $F_{N,k}(\cdot,\cdot)$ всегда принимает не более трех значений: -1, 0 или 1. Поэтому можем продолжить (1.105):

$$\begin{aligned}
&\mathbb{P}\left(\left|\mathbb{E}\left(\mathbb{I}\left\{Y^{j_1} = y\right\} F_{N,k}(X^{j_1},y)|\xi_N(\overline{S_k(N)})\right)\right| < \varepsilon/2\right) \\
&\leqslant \mathbb{P}\left(\mathbb{E}\left(\left|\mathbb{I}\left\{Y^{j_1} = y\right\} F_{N,k}(X^{j_1},y)\right|\middle|\xi_N(\overline{S_k(N)})\right) > \varepsilon/2\right) \\
&\leqslant \mathbb{P}\left(\mathbb{E}\left(\left|F_{N,k}(X^{j_1},y)\right|\middle|\xi_N(\overline{S_k(N)})\right) > \varepsilon/2\right) \\
&= \mathbb{P}\left(\mathbb{E}\left(\mathbb{I}\left\{\left|F_{N,k}(X^{j_1},y)\right| = 1\right\}\middle|\xi_N(\overline{S_k(N)})\right) > \varepsilon/2\right). \quad (1.106)
\end{aligned}$$

По условию теоремы случайный вектор X^{j_1} имеет плотность $\rho(x)$ и по лемме о группировке не зависит от случайных векторов $\xi_N(\overline{S_k(N)})$. Тогда, используя (1.94), мы можем продолжить (1.106):

$$\begin{aligned}
&\mathbb{P}\left(\mathbb{E}\left(\mathbb{I}\left\{\left|F_{N,k}(X^{j_1},y)\right| = 1\right\}\middle|\xi_N(\overline{S_k(N)})\right) > \varepsilon/2\right) \\
&= \mathbb{P}\left(\int_{\mathbb{R}^n} \mathbb{I}\left\{|F_{N,k}(x,y)| = 1\right\}\rho(x)dx > \varepsilon/2\right). \quad (1.107)
\end{aligned}$$

Рассмотрим более подробно индикатор $\mathbb{I}\left\{|F_{N,k}(x,y)| = 1\right\}$ из (1.107). Напомним, что

$$F_{N,k}(x,y) := \mathbb{I}\left\{f_{PA}\left(x,\xi_N(\overline{S_k(N)})\right) \neq y\right\} - \mathbb{I}\left\{f(x) \neq y\right\}.$$

Следовательно,

$$\mathbb{I}\left\{|F_{N,k}(x,y)| = 1\right\} = \mathbb{I}\left\{f_{PA}\left(x,\xi_N(\overline{S_k(N)})\right) \neq f(x)\right\}.$$

Для произвольного $\delta > 0$ запишем

$$\begin{aligned}
&\mathbb{I}\left\{f_{PA}\left(x,\xi_N(\overline{S_k(N)})\right) \neq f(x)\right\} \\
&= \mathbb{I}\left\{f_{PA}\left(x,\xi_N(\overline{S_k(N)})\right) \neq f(x), \left|\widehat{\gamma}_N(\psi) - \gamma(\psi)\right| \leqslant \delta\right\} \\
&+ \mathbb{I}\left\{f_{PA}\left(x,\xi_N(\overline{S_k(N)})\right) \neq f(x), \left|\widehat{\gamma}_N(\psi) - \gamma(\psi)\right| > \delta\right\}. \quad (1.108)
\end{aligned}$$

Продолжим оценку (1.107), пользуясь (1.108). Имеем

$$\mathbb{P}\left(\int_{\mathbb{R}^n}\mathbb{I}\left\{f_{PA}\left(x,\xi_N(\overline{S_k(N)})\right)\neq f(x),\left|\widehat{\gamma}_N(\psi)-\gamma(\psi)\right|>\delta\right\}\rho(x)dx>\varepsilon/2\right)$$
$$\leqslant\mathbb{P}\left(\int_{\mathbb{R}^n}\mathbb{I}\left\{\left|\widehat{\gamma}_N(\psi)-\gamma(\psi)\right|>\delta\right\}\rho(x)dx>\varepsilon/2\right)$$
$$=\mathbb{P}\left(\mathbb{I}\left\{\left|\widehat{\gamma}_N(\psi)-\gamma(\psi)\right|>\delta\right\}>\varepsilon/2\right)$$
$$\leqslant\mathbb{P}\left(\left|\widehat{\gamma}_N(\psi)-\gamma(\psi)\right|>\delta\right). \quad (1.109)$$

По условию теоремы $\widehat{\gamma}_N(\psi)\to\gamma(\psi)$ вполне при $N\to\infty$. Сходимость вполне означает, что

$$\sum_{N=1}^{\infty}\mathbb{P}\left(\left|\widehat{\gamma}_N(\psi)-\gamma(\psi)\right|>\delta\right)<\infty \quad (1.110)$$

для любого $\delta>0$. Итак, чтобы обеспечить выполнение (1.92), учитывая (1.101), нам осталось разобраться с первым слагаемым в правой части (1.108) и его вкладом в оценку (1.107). Используя непрерывность $\kappa(\delta)$ в окрестности 0, и то, что по условию теоремы $\kappa(0)=0$, выберем для положительного $\varepsilon>0$ такое $\delta>0$, что

$$\int_{\mathbb{R}^n}\mathbb{I}\left\{|P(Y=1|X=x)-\gamma(\psi)|<2\delta\right\}\rho(x)dx<\varepsilon/8. \quad (1.111)$$

Заметим, что $f_{PA}\left(x,\xi_N(\overline{S_k(N)})\right)$ и $f(x)$, $x\in\mathbb{X}$, могут принимать лишь значения -1 или 1. Поэтому

$$\mathbb{I}\left\{f_{PA}\left(x,\xi_N(\overline{S_k(N)})\right)\neq f(x)\right\}=\mathbb{I}\left\{f_{PA}\left(x,\xi_N(\overline{S_k(N)})\right)=1,f(x)=-1\right\}$$
$$+\mathbb{I}\left\{f_{PA}\left(x,\xi_N(\overline{S_k(N)})\right)=-1,f(x)=1\right\}.$$

Согласно определению функции f (см. (1.2)) и построению предсказательного алгоритма f_{PA} в (1.8) имеем

$$\mathbb{I}\left\{f_{PA}\left(x,\xi_N(\overline{S_k(N)})\right)=1,f(x)=-1\right\}$$
$$=\mathbb{I}\left\{\widehat{\mathbb{P}}(Y=1|X=x)>\widehat{\gamma}_N(\psi),\mathbb{P}(Y=1|X=x)\leqslant\gamma(\psi)\right\}$$

и

$$\mathbb{I}\left\{f_{PA}\left(x,\xi_N(\overline{S_k(N)})\right)=-1,f(x)=1\right\}$$
$$=\mathbb{I}\left\{\widehat{\mathbb{P}}(Y=1|X=x)\leqslant\widehat{\gamma}_N(\psi),\mathbb{P}(Y=1|X=x)>\gamma(\psi)\right\}$$

.

Тогда с использованием (1.111) для первого слагаемого в правой части (1.108) можно записать

$$\mathbb{P}\Big(\int_{\mathbb{R}^n}\mathbb{I}\Big\{f_{PA}\left(x,\xi_N(\overline{S_k(N)})\right)\neq f(x),|\widehat{\gamma}_N(\psi)-\gamma(\psi)|\leqslant\delta\Big\}\rho(x)dx>\varepsilon/4\Big)$$

$$\leqslant\mathbb{P}\Big(\int_{\mathbb{R}^n}\Big(\mathbb{I}\Big\{\mathbb{P}(Y=1|X=x)\leqslant\gamma(\psi),\widehat{\gamma}_N(\psi)<\widehat{\mathbb{P}}(Y=1|X=x),$$

$$|\widehat{\gamma}_N(\psi)-\gamma(\psi)|\leqslant\delta,|\mathbb{P}(Y=1|X=x)-\gamma(\psi)|>2\delta\Big\}$$

$$+\mathbb{I}\Big\{\widehat{\mathbb{P}}_N(Y=1|X=x)\leqslant\widehat{\gamma}_N(\psi),\gamma(\psi)<\mathbb{P}(Y=1|X=x),$$

$$|\widehat{\gamma}_N(\psi)-\gamma(\psi)|\leqslant\delta,|\mathbb{P}(Y=1|X=x)-\gamma(\psi)|>2\delta\Big\}\Big)\rho(x)dx>\varepsilon/4\Big)$$

$$\leqslant\mathbb{P}\Big(\int_{\mathbb{R}^n}\mathbb{I}\Big\{|\widehat{\mathbb{P}}_N(Y=1|X=x)-\mathbb{P}(Y=1|X=x)|>\delta\Big\}\rho(x)dx>\varepsilon/4\Big). \tag{1.112}$$

Здесь мы использовали то, что

$$\mathbb{P}\Big(\mathbb{I}\Big\{f_{PA}\left(x,\xi_N(\overline{S_k(N)})\right)\neq f(x),|\widehat{\gamma}_N(\psi)-\gamma(\psi)|\leqslant\delta,$$

$$|\mathbb{P}(Y=1|X=x)-\gamma(\psi)|>2\delta\Big\}\rho(x)dx>\varepsilon/4\Big)$$

$$\leqslant\mathbb{P}\Big(\int_{\mathbb{R}^n}\mathbb{I}\Big\{|\mathbb{P}(Y=1|X=x)-\gamma(\psi)|>2\delta\Big\}\rho(x)dx>\varepsilon/4\Big)=0,$$

так как согласно (1.111) имеем

$$\int_{\mathbb{R}^n}\mathbb{I}\left\{|\mathbb{P}(Y=1|X=x)-\gamma(\psi)|>2\delta\right\}\rho(x)dx<\varepsilon/8<\varepsilon/4.$$

Чтобы показать справедливость последнего неравенства в (1.112), рассмотрим первое слагаемое под знаком интегрирования. Индикаторная функция будет равна единице в том и только том случае, когда выполнены все условия, задающие множество, индикатор которого рассматривается. Из того, что $\mathbb{P}(Y=1|X=x)\leqslant\gamma(\psi)$, $|\widehat{\gamma}_N(\psi)-\gamma(\psi)|\leqslant\delta$ и $|\mathbb{P}(Y=1|X=x)-\gamma(\psi)|>2\delta$ следует, что $\widehat{\gamma}_N(\psi)-\mathbb{P}(Y=1|X=x)>\delta$. Неравенство $\widehat{\mathbb{P}}(Y=1|X=x)-\mathbb{P}(Y=1|X=x)>\delta$ выполняется, так как справедливо неравенство $\widehat{\gamma}_N(\psi)<\widehat{\mathbb{P}}(Y=1|X=x)$. Аналогично рассматривается второй индикатор и показывается, что $\mathbb{P}(Y=1|X=x)-\widehat{\mathbb{P}}(Y=1|X=x)>\delta$.

Применяя неравенство Маркова и теорему Фубини ([6], с. 276), получаем оценку сверху для выражения из (1.112):

$$\mathbb{P}\Big(\int_{\mathbb{R}^n}\big(\mathbb{I}\left\{|\widehat{\mathbb{P}}_N(Y=1|X=x)-\mathbb{P}(Y=1|X=x)|>\delta\right\}\rho(x)dx>\varepsilon/4\Big)$$

$$\leqslant\frac{4}{\varepsilon}\mathbb{E}\left[\int_{\mathbb{R}^n}\big(\mathbb{I}\left\{|\widehat{\mathbb{P}}_N(Y=1|X=x)-\mathbb{P}(Y=1|X=x)|>\delta\right\}\rho(x)dx\right]$$

$$\leqslant\frac{4}{\varepsilon}\int_{\mathbb{R}^n}\mathbb{P}(|\widehat{\mathbb{P}}_N(Y=1|X=x)-\mathbb{P}(Y=1|X=x)|>\delta)\rho(x)dx$$

$$\leqslant\frac{4}{\varepsilon}\int_{\mathbb{R}^n}\mathbb{P}(\sup_{x\in\mathbb{R}^n}|\widehat{\mathbb{P}}_N(Y=1|X=x)-\mathbb{P}(Y=1|X=x)|>\delta)\rho(x)dx$$

$$\leqslant\frac{4}{\varepsilon}\mathbb{P}\left(\sup_{x\in\mathbb{R}^n}|\widehat{\mathbb{P}}_N(Y=1|X=x)-\mathbb{P}(Y=1|X=x)|>\delta\right). \tag{1.113}$$

Из условия (1.87) следует, что величины, входящие в последнюю оценку (1.113), суммируются по N. Доказательство теоремы завершено. □

Теперь мы покажем, что условия теоремы выполняются для различных оценок, применяющихся на практике. Начнем с требования о сходимости вполне последовательности случайных величин $\widehat{\gamma}_N(\psi)$ к $\gamma(\psi)$ при $N\to\infty$.

Замечание 10. В [25] в замечании 4 объясняется, почему предложенная в [68] штрафная функция

$$\psi(y)=\frac{c}{\mathbb{P}(Y=y)},\quad y\in\{-1,1\},\quad c>0, \tag{1.114}$$

является хорошим выбором. Если штрафная функция задана согласно (1.114), то справедливо

$$\gamma(\psi)=\frac{\psi(-1)}{\psi-1+\psi(1)}=\frac{c/\mathbb{P}(Y=-1)}{c/\mathbb{P}(Y=-1)+c/\mathbb{P}(Y=1)}=\mathbb{P}(Y=1).$$

В качестве оценки вероятности $\mathbb{P}(Y=1)$ естественно использовать

$$\widehat{\gamma}_N(\psi,\xi_N(\overline{S_k(N)})):=\frac{1}{\#\overline{S_k(N)}}\sum_{j\in\overline{S_k(N)}}\mathbb{I}\left\{Y^j=1\right\},\quad k=1,\ldots,K. \tag{1.115}$$

Предложение 1. *Предположим, что штрафная функция* $\psi(y)$ *задана согласно* (1.114), *а оценка* $\widehat{\gamma}_N(\psi,\xi_N(\overline{S_k(N)}))$, $k=1,\ldots,K$, *определена в* (1.115). *Тогда для любого положительного* δ *справедливо неравенство*

$$\mathbb{P}(|\widehat{\gamma}_N(\psi,\xi_N(\overline{S_k(N)}))-\gamma(\psi)|\geqslant\delta)\leqslant 2\exp(-2\delta^2\#\overline{S_k(N)}),\quad k=1,\ldots,K. \tag{1.116}$$

В частности, из этого следует, что для любого $k = 1, \ldots, K$ *оценка* $\widehat{\gamma}_N(\psi, \xi_N(\overline{S_k(N)}))$ *сходится к* $\gamma(\psi)$ *вполне.*

Доказательство. Неравенство (1.116) вытекает непосредственно из неравенства Хефдинга [41] для независимых одинаково распределенных ограниченных случайных величин, которыми являются $\mathbb{I}\left\{Y^j = 1\right\}$, $j = 1, \ldots, N$. Сходимость вполне следует и того, что правые части (1.116) суммируются по N, так как $\#S_k(N) \geqslant [N/K]$. $\square$

Замечание 11. В теореме 4 фигурируют условия на функцию $\varkappa$. А именно, требуется, чтобы $\varkappa$ была непрерывна в 0 и $\varkappa(0) = 0$. Как было показано выше, если штрафная функция ψ задана согласно (1.114), то $\gamma(\psi) = \mathbb{P}(Y = 1)$. Тогда условие $\varkappa(0) = 0$ можно переписать в виде

$$\int_{\mathbb{R}^n} \mathbb{I}\left\{x : \mathbb{P}(Y = 1|X = x) = \mathbb{P}(Y = 1)\right\} \rho(x)dx = 0.$$

Другими словами, индуцированная случайным вектором X мера точек $x \in \mathbb{X}$, для которых условная вероятность $\mathbb{P}(Y = 1|X = x)$ совпадает с безусловной $\mathbb{P}(Y = 1)$, равна нулю. Это условие естественно, так как фактически означает, что мера значений X, при которых функция отклика Y не зависит от X, равна 0. Второе условие на непрерывность в точке 0 выполняется для широкого класса распределений. Например, это справедливо для нормально распределенного вектора X и модели логистической регрессии:

$$\mathbb{P}(Y = 1|X = x) = \frac{\exp\{\beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n\}}{1 + \exp\{\beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n\}},$$

где $\beta_i > 0$, $i = 1, \ldots, n$. Действительно, в таком случае

$$\varkappa(\delta) = \int_{\mathbb{R}^n} \mathbb{I}\left\{x : \mathbb{P}(Y = 1|X = x) \in [\gamma(\psi) - \delta, \gamma(\psi) + \delta]\right\} \rho(x)dx$$

$$= \int_{\mathbb{R}^n} \mathbb{I}\left\{x : \frac{\exp\{\beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n\}}{1 + \exp\{\beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n\}} \in [\gamma(\psi) - \delta, \gamma(\psi) + \delta]\right\}$$

$$= \int_{\beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n \in [\log \frac{\gamma(\psi) - \delta}{1 - \gamma(\psi) + \delta}, \log \frac{\gamma(\psi) + \delta}{1 - \gamma(\psi) - \delta}]} \rho(x)dx \to 0$$

при $\delta \to 0$. Здесь $\rho(x)$ будет соответствовать плотности многомерного нормального распределения. Такие предположения на распределения встречаются в различных моделях биологических и генетических данных, в частности, при анализе генной экспрессии [44].

1.3.2 Варианты оценок условных вероятностей $\widehat{\mathbb{P}}(Y = 1|X = x)$

В данном разделе мы рассмотрим варианты оценок условных вероятностей $\mathbb{P}(Y = 1|X = x)$, которые удовлетворяют условию равномерной сходимости вполне (1.87) из теоремы 4. Так как изначально MDR-EFE метод разрабатывался как непараметрический метод выявления значимых факторов, естественно использовать в качестве $\widehat{\mathbb{P}}_N(Y = 1|X = x)$ непараметрические оценки. Мы остановимся на ядерных оценках плотности распределения, использующих метод k-ближайших соседей (см., например, [15]).

Ядерные оценки плотности распределения

Случайная величина $\widehat{\mathbb{P}}_N(Y = 1|X = x)$ есть оценка условной вероятности $\mathbb{P}(Y = 1|X = x)$. Пусть распределение (X,Y) абсолютно непрерывно относительно меры $\mu \otimes \lambda$, где μ – мера Лебега в $\mathbb{R}^n$, λ – считающая мера на $\mathbb{Y}$. Иначе говоря, мы рассматриваем смешанную модель. Такие модели исследуются в различных работах, например, [21]. Обозначим

$$f(x,y) := \frac{dP_{(X,Y)}}{\mu \otimes \lambda}, v \in \mathbb{R}^n, y \in \{-1,1\}.$$

Пусть $\rho_X(x)$ обозначает плотность случайного вектора X, тогда $\rho_X(x) = \sum_{y\in\{-1,1\}} f(x,y)$, $x \in \mathbb{R}^n$. Кроме того, условная плотность вектора X при заданном значении Y имеет вид

$$f_{X|Y}(x|y) = \frac{f(x,y)}{\mathbb{P}(Y = y)} = \frac{f_{Y|X}(y|x)\rho_X(x)}{\mathbb{P}(Y = y)},$$

поскольку $f_{Y|X}(y|x) = \frac{f(x,y)}{\rho_X(x)}$. Мы всюду полагаем, что (версия) плотности $f(x,y) > 0$ при всех $x \in \mathbb{R}^n$ и $y \in \{-1,1\}$. Тогда существует версия плотности $\rho_X(x) > 0$ для всех $x \in \mathbb{R}^n$ и $\mathbb{P}(Y = y) > 0$ для $y \in \{-1,1\}$.

Как и ранее, случайная величина

$$\widehat{\mathbb{P}}_N(Y = y) = \frac{1}{N}\sum_{j=1}^{N} \mathbb{I}\left\{Y^j = y\right\} \tag{1.117}$$

будет использоваться для оценки $\mathbb{P}(Y = y)$.

Наши дальнейшие рассуждения будут опираться на результаты статьи [33]. В указанной работе исследуются оценки плотности $\rho(\cdot)$ (по мере Лебега) вектора со значениями в $\mathbb{R}^n$, построенные по выборке независимых одинаково распределенных векторов $X_1, \ldots, X_N$ с плотностью ρ на $\mathbb{R}^n$. Авторы рассматривают ядерную оценку специального вида:

$$\widehat{\rho}(x,k,N) := \frac{1}{N}\sum_{i=1}^{N} \frac{1}{(H_{N,k,i})^n} K\left(\frac{X_i - x}{H_{N,k,i}}\right),\ x \in \mathbb{R}^n,\ N \in \mathbb{N}, \tag{1.118}$$

где $N > 1$, $H_{N,k,i}$ – расстояние от X_i до $k-$го ближайшего соседа среди X_j, $j \neq i$, $k = k_N$ – положительные целые числа. Далее мы будем опускать индекс k в обозначениях $\rho(x,k,N)$ и $H_{N,k,i}$. В [33] доказывается следующий важный результат.

Теорема 5 ([33]). *Если*

$$K(x) = \mathbb{I}\{||x|| \leqslant 1/c\}, \tag{1.119}$$

где c – положительная константа,

$$\lim_{N\to\infty} \frac{k}{N} = 0, \tag{1.120}$$

$$\lim_{N\to\infty} \frac{k}{\log N} = \infty, \tag{1.121}$$

и функция ρ равномерно непрерывна, то оценка плотности, заданная в (1.118) с помощью ядра (1.119), обладает следующим свойством. Для всех $\varepsilon > 0$ существуют $\delta > 0$ и N_0 такие, что

$$\mathbb{P}(\sup_x |\widehat{\rho}(x,N) - \rho(x)| > \varepsilon) \leqslant \exp(-\delta k), \qquad N > N_0. \tag{1.122}$$

Заметим, что в качестве ядра K используется индикаторная функция. Из этого следует, что заданная в (1.118) случайная величина принимает конечное количество значений при каждом $N \in \mathbb{N}$. Поэтому $\sup_x |\widehat{\rho}(x,N) - \rho(x)|$ в утверждении теоремы 5 является измеримой случайной величиной и определен корректно.

Опираясь на теорему 5, построим оценку для $\mathbb{P}(Y = 1|X = x)$:

$$\widehat{\mathbb{P}}(Y = 1|X = x) = \frac{\widehat{f}_{Y|X}(x|1,N)\widehat{\mathbb{P}}_N(Y = 1)}{\widehat{\rho}_X(x,N)}. \tag{1.123}$$

Докажем следующее утверждение.

Теорема 6. *Пусть вектор X имеет абсолютно-непрерывную плотность $\rho_X(x)$ с носителем M. Пусть функция κ из (1.86) непрерывна в окрестности 0 и $\kappa(0) = 0$. Предположим, что для некоторых положительных констант C_1,C_2*

$$C_1 < \rho_X(x) < C_2, \quad \forall x \in M. \tag{1.124}$$

$$C_1 < f_{X|Y}(x|1) < C_2, \quad \forall x \in M. \tag{1.125}$$

Кроме того, будем считать, что выполнены условия (1.119) - (1.121). Тогда справедливо соотношение

$$\widehat{Err}_K(f_{PA},\xi_N) \to Err(f) \quad п.н.,\ N \to \infty.$$

Доказательство. Нам достаточно проверить, что для каждого $\delta > 0$

$$\sum_{N=1}^{\infty} \mathbb{P}\left(\sup_{x\in\mathbb{R}^n} |\widehat{\mathbb{P}}_N(Y=1|X=x) - \mathbb{P}(Y=1|X=x)| > \delta \right) < \infty. \tag{1.126}$$

Мы хотим оценить сверху вероятность

$$\mathbb{P}\Big(\sup_{x\in\mathbb{R}^n} \Big| \frac{\widehat{\rho}_{X|Y=1}(x,N)\widehat{\mathbb{P}}_N(Y=1)}{\widehat{\rho}_X(x,N)} - \frac{\rho_{X|Y=1}(x)\mathbb{P}(Y=1)}{\rho_X(x)} \Big| > \delta \Big).$$

Выберем некоторое положительное δ^* такое, что $\delta^* < C_1$. Для произвольных положительных δ_1, δ_2, δ_3 таких, что $|\delta_i| < \delta^*$, $i = 1,2,3$, имеем

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{x\in\mathbb{R}^n} \Big| \frac{\widehat{f}_{X|Y}(x|1,N)\widehat{\mathbb{P}}_N(Y=1)}{\widehat{\rho}_X(x,N)} - \frac{f_{X|Y}(x|1)\mathbb{P}(Y=1)}{\rho_X(x)} \Big| > \delta \Big) \\
&\leqslant \mathbb{P}\Big(\sup_{x\in\mathbb{R}^n} \Big| \widehat{f}_{X|Y}(x|1,N) - f_{X|Y}(x|1) \Big| > \delta_1 \Big) \\
&+\mathbb{P}\Big(\sup_{x\in\mathbb{R}^n} \Big| \widehat{\rho}_X(x,N) - \rho_X(x) \Big| > \delta_2 \Big) + \mathbb{P}\Big(\Big| \widehat{\mathbb{P}}_N(Y=1) - \mathbb{P}(Y=1) \Big| > \delta_3 \Big) \\
&+ \mathbb{P}\Big(\sup_{x\in\mathbb{R}^n} \Big| \frac{\widehat{f}_{X|Y}(x|1,N)\widehat{\mathbb{P}}_N(Y=1)}{\widehat{\rho}_X(x,N)} - \frac{f_{X|Y=1}(x|1)\mathbb{P}(Y=1)}{\rho_X(x)} \Big| > \delta, \\
&\sup_{x\in\mathbb{R}^n} \Big| \widehat{f}_{X|Y}(x|1,N) - f_{X|Y}(x|1) \Big| \leqslant \delta_1,\ \sup_{x\in\mathbb{R}^n} \Big| \widehat{\rho}_X(x,N) - \rho_X(x) \Big| \leqslant \delta_2, \\
&\Big| \widehat{\mathbb{P}}_N(Y=1) - \mathbb{P}(Y=1) \Big| \leqslant \delta_3 \Big).
\end{aligned} \tag{1.127}$$

Первые три слагаемых из последнего выражения суммируются по N. Оценки плотностей $\widehat{f}_{X|Y}(x|1,N)$ и $\widehat{\rho}_X(x,N)$ удовлетворяют условиям теоремы 5.

Поэтому для первых двух слагаемых $\mathbb{P}\Big(\sup_{x\in\mathbb{R}^n}\Big|\widehat{f}_{X|Y}(x|1,N)-f_{X|Y}(x|1)\Big|>\delta_1\Big)$ и $\mathbb{P}\Big(\sup_{x\in\mathbb{R}^n}\Big|\widehat{\rho}_X(x,N)-\rho_X(x)\Big|>\delta_2\Big)$ справедливы оценки сверху из (1.122). Согласно замечанию 1 из [33] такие оценки сверху суммируются по N. Оценка $\widehat{\mathbb{P}}_N(Y=1)$ из (1.117) является суммой независимых ограниченных одинаково распределенных случайных величин. Поэтому мы можем применить неравенство Хефдинга:

$$\mathbb{P}\Big(\Big|\widehat{\mathbb{P}}_N(Y=1)-\mathbb{P}(Y=1)\Big|>\delta_3\Big)\leqslant 2\exp\Big(-\frac{N\delta_3^2}{2\mathsf{var}(\mathbb{I}\{Y^1=1\})}\Big).$$

Отсюда следует суммируемость третьего слагаемого по N.

Рассмотрим третье слагаемое. Для некоторых $|\tilde{\delta}_i|<\delta_i$, $i=1,2,3$, справедливо

$$\begin{gathered}\Big|\frac{\widehat{f}_{X|Y}(x|1,N)\widehat{\mathbb{P}}_N(Y=1)}{\widehat{\rho}_X(x,N)}-\frac{f_{X|Y}(x|1)\mathbb{P}(Y=1)}{\rho_X(x)}\Big|\\ =\Big|\frac{\big(f_{X|Y}(x|1)+\tilde{\delta}_1\big)\big(\mathbb{P}(Y=1)+\tilde{\delta}_3\big)}{\rho_X(x)+\tilde{\delta}_2}-\frac{f_{X|Y}(x|1)\mathbb{P}(Y=1)}{\rho_X(x)}\Big|\\ =\Big|\frac{f_{X|Y}(x|1)\rho_X(x)\tilde{\delta}_3+\rho_X(x)\mathbb{P}(Y=1)\tilde{\delta}_1+\rho_X(x)\tilde{\delta}_1\tilde{\delta}_3-f_{X|Y}(x|1)\mathbb{P}(Y=1)\tilde{\delta}_2}{(\rho_X(x)+\tilde{\delta}_2)\rho_X(x)}\Big|\\ \leqslant\frac{C_2^2\delta^*+C_2\delta^*(\delta^*+2)}{(C_1-\delta^*)C_1}.\end{gathered}$$

Здесь мы воспользовались тем, что $\max(f_{X|Y}(x|1),\rho_X(x))\leqslant C_2$, $x\in M$, а также $|\tilde{\delta}_i|<\delta^*$, $i=1,2,3$, и $(\rho_X(x)+\tilde{\delta}_2)\rho_X(x)>(C_1-\delta^*)C_1>0$.

Несложно выбрать такое δ^*, что

$$\frac{C_2^2\delta^*+C_2\delta^*(\delta^*+2)}{(C_1-\delta^*)C_1}<\delta.$$

Тогда последнее слагаемое в сумме (1.127) будет равно 0. Тем самым показано, что соотношение (1.126) выполняется. Доказательство теоремы завершено. □

Замечание 12. Для оценки из (1.123) измеримость $\sup_{x\in\mathbb{R}^n}|\widehat{\mathbb{P}}_N(Y=1|X=x)-\mathbb{P}(Y=1|X=x)|$ следует из того, что при каждом натуральном N эта оценка принимает конечное число значений.

Глава 2. Скорость сходимости оценок функционала ошибки в MDR-EFE методе

При подготовке данной главы диссертации использован материал публикаций [81—83]. Работа [81] выполнена автором в соавторстве с профессором А.В.Булинским. В публикации А.В.Булинскому принадлежит постановка задач и общий подход к их решению, им также доказана лемма 2 (лемма 2 в диссертации), следствие 1 (следствие 1 в диссертации) и следствие 4 (теорема 8 в диссертации), все остальные результаты доказаны автором диссертации. Работа [82] выполнена автором в соавторстве с профессором А.В.Булинским. В публикации А.В.Булинскому принадлежит постановка задач и общий подход к их решению, им также доказана лемма 1 (лемма 5 в диссертации), теорема 3 (теорема 13 в диссертации) и следствие 1 (следствие 4 в диссертации), все остальные результаты доказаны автором диссертации. Работа [83] выполнена автором в соавторстве с П. Алонсо-Руиз. В публикации П. Алонсо-Руиз принадлежит предложение 1 (предложение 2 в диссертации) и следствие 1 (следствие 3 в диссертации), все остальные результаты доказаны автором диссертации.

2.1 Центральная предельная теорема для регуляризованных оценок $\widehat{Err}_K(\widehat{f}^{\beta}_{PA})$

Для практических приложений важно понимать, как ведет себя разность $\widehat{Err}_K(\widehat{f}^{\beta}_{PA})$ и $Err(f^{\beta})$ при $N \to \infty$. Данная задача рассматривается в этой главе.

Пусть $\beta = (m_1, \ldots, m_r)$ с $\{m_1, \ldots, m_r\} \subset \{1, \ldots, n\}$. Мы зададим функции, которые будут рассматриваться как *регуляризованные версии* оценок $\widehat{f}^{\beta}_{PA}$ функции f^{β} (см. (1.83) и (1.72)). А именно, для $\varepsilon = (\varepsilon_N)_{N \in \mathbb{N}}$, где неслучайные положительные $\varepsilon_N \to 0$ при $N \to \infty$, положим

$$\widehat{f}^{\beta}_{PA,\varepsilon}(x, \xi_N(W_N)) = \sum_{y \in \mathbb{Y}} y \, \mathbb{I}\{x \in \widehat{A}^{\beta, W_N}_{y,\varepsilon}\}, \tag{2.1}$$

Здесь $\widehat{A}^{\beta,W_N}_{y,\varepsilon} = \widehat{A}^{\beta,W_N}_{y,\varepsilon}(\omega)$ задается следующим образом:

$$x \in \widehat{A}^{\beta,W_N}_{y,\varepsilon}(\omega) \Longleftrightarrow \begin{cases} \widehat{L}^{\beta,W_N}_{-m+1}(x,\omega) + \varepsilon_N \geqslant 0, & y = -m, \\ \widehat{L}^{\beta,W_N}_{y+1}(x,\omega) + \varepsilon_N \geqslant 0, \widehat{L}^{\beta,W_N}_{y}(x,\omega) + \varepsilon_N < 0, & y \neq \pm m, \\ \widehat{L}^{\beta,W_N}_{m}(x,\omega) + \varepsilon_N < 0, & y = m, \end{cases} \tag{2.2}$$

$\widehat{L}^{\beta,W_N}_{y}(x,\omega)$ для $y \in \mathbb{Y}$, $y > -m$, и $\omega \in \Omega$ определены в (1.82). Для упрощения обозначений мы будем писать $\widehat{A}^{\beta,W_N}_{y,\varepsilon}(\omega)$ вместо $\widehat{A}^{\beta,W_N}_{y,\varepsilon_N}(\omega)$.

Теперь рассмотрим U, возникающее в (1.80). Утверждения, аналогичные лемме 3 и теореме 3, справедливы и для регуляризованных версий оценок, введенных выше. Далее мы будем рассматривать штрафную функцию из (1.61) в качестве ψ. Под $\widehat{\psi}_{N,k}$ мы понимаем оценку ψ, заданную в (1.62). Теперь обратимся к основному результату этой главы.

Теорема 7. *Пусть $\varepsilon_N \to 0$ и $N^{1/2}\varepsilon_N \to \infty$ при $N \to \infty$. Тогда для каждого $K \in \mathbb{N}$, произвольного вектора $\beta = (m_1, \dots m_r)$ с $1 \leqslant m_1 < \dots < m_r \leqslant n$, соответствующей функции $f = f^\beta$ и предсказательного алгоритма $f_{PA} = \widehat{f}^{\beta}_{PA,\varepsilon}$, заданного формулой (2.1), выполняется следующее соотношение:*

$$\sqrt{N}(\widehat{Err}_K(f_{PA},\xi_N) - Err(f)) \xrightarrow{law} Z \sim N(0,\sigma^2), \ N \to \infty. \tag{2.3}$$

Здесь σ^2 – дисперсия случайной величины

$$V = \sum_{i=0}^{2m-1} \sum_{i-m<|y|\leqslant m} \frac{\mathbb{I}\{Y=y\}}{\mathbb{P}(Y=y)} \left(\mathbb{I}\{|f(X)-y| > i\} - \mathbb{P}(|f(X)-y| > i \big| Y = y) \right). \tag{2.4}$$

Доказательство. Для фиксированного $K \in \mathbb{N}$ и произвольного $N \in \mathbb{N}$ положим

$$T_N(f) := \frac{1}{K} \sum_{k=1}^{K} \frac{1}{\sharp S_k(N)} \sum_{i=0}^{2m-1} \sum_{i-m<|y|\leqslant m} \psi(y) \sum_{j \in S_k(N)} \mathbb{I}\{Y^j = y, |f(X^j)-y| > i\},$$

$$\widehat{T}_N(f) := \frac{1}{K} \sum_{k=1}^{K} \frac{1}{\sharp S_k(N)} \sum_{i=0}^{2m-1} \sum_{i-m<|y|\leqslant m} \widehat{\psi}_{N,k}(y) \sum_{j \in S_k(N)} \mathbb{I}\{Y^j = y, |f(X^j)-y| > i\},$$

где $\widehat{\psi}_{N,k}(y) = \widehat{\psi}(y, \xi_N(S_k(N)))$. Имеем

$$\widehat{Err}_K(f_{PA},\xi_N) - Err(f) = (\widehat{Err}_K(f_{PA},\xi_N) - \widehat{T}_N(f))$$

$$+(\widehat{T}_N(f)-T_N(f))+(T_N(f)-Err(f)). \tag{2.5}$$

Прежде всего, мы покажем, что

$$\sqrt{N}(\widehat{Err}_K(f_{PA},\xi_N)-\widehat{T}_N(f)) \xrightarrow{\mathbb{P}} 0,\ \ N\to\infty. \tag{2.6}$$

Используя (1.31), можно записать

$$\widehat{Err}_K(f_{PA},\xi_N)-\widehat{T}_N(f) = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{\sharp S_k(N)}\sum_{i=0}^{2m-1}\sum_{i-m<|y|\leqslant m}\widehat{\psi}_{N,k}(y)\sum_{j\in S_k(N)}\mathbb{I}\{Y^j=y\}F^{(i)}_{N,k}(X^j,y). \tag{2.7}$$

Для $i=0,\ldots,2m-1$, $k=1,\ldots,K$ и $N\in\mathbb{N}$ определим случайные величины

$$G^{(i)}_{N,k}(y):=\frac{1}{\sqrt{\sharp S_k(N)}}\sum_{j\in S_k(N)}\mathbb{I}\{Y^j=y\}F^{(i)}_{N,k}(X^j,y)$$

и убедимся, что для каждого $k=1,\ldots,K$

$$\sum_{i=0}^{2m-1}\sum_{i-m<|y|\leqslant m}\widehat{\psi}_{N,k}(y)G^{(i)}_{N,k}(y) \xrightarrow{\mathbb{P}} 0,\ \ N\to\infty. \tag{2.8}$$

Ясно, что (2.8) влечет (2.6) в силу (2.7), так как $\sharp S_k(N)=[N/K]$ для $k=1,\ldots,K-1$ и $[N/K]\leqslant\sharp S_K(N)<[N/K]+K$. Для всех рассматриваемых i,N,k и U, заданном в (1.80), имеем $G^{(i)}_{N,k}(y)=G^{(i),U}_{N,k}(y)+G^{(i),\mathbb{X}\setminus U}_{N,k}(y)$, где

$$G^{(i),U}_{N,k}(y)=\frac{1}{\sqrt{\sharp S_k(N)}}\sum_{j\in S_k(N)}\mathbb{I}\{X^j\in U)\}\mathbb{I}\{Y^j=y\}F_{N,k}(X^j,y),$$

$$G^{(i),\mathbb{X}\setminus U}_{N,k}(y)=\frac{1}{\sqrt{\sharp S_k(N)}}\sum_{j\in S_k(N)}\mathbb{I}\{X^j\notin U\}\mathbb{I}\{Y^j=y\}F_{N,k}(X^j,y).$$

Очевидно, что

$$|G^{(i),U}_{N,k}(y)|\leqslant\sum_{x\in U}\sqrt{\sharp S_k(N)}\,\big|\mathbb{I}\{|f_{PA}(x,\xi_N(\overline{S_k(N)}))-y|>i\}-\mathbb{I}\{|f(x)-y|>i\}\big|.$$

Функции f_{PA} и f принимают значения в множестве $\mathbb{Y}=\{-m,\ldots,m\}$. Таким образом, для любых $x\in U$, $k=1,\ldots,K$ и почти всех $\omega\in\Omega$ соотношение (1.27) обеспечивает существование такого целого $N_0(x,k,\omega)$, что $f_{PA}(x,\xi_N(\overline{S_k(N)}))=f(x)$ при $N\geqslant N_0(x,k,\omega)$. Следовательно, $G^{(i),U}_{N,k}(y)=0$ для

любого y, принадлежащего $\mathbb{Y}$, каждых $i = 0,\dots,2m-1$, $k = 1,\dots,K$ и почти всех $\omega \in \Omega$, когда $N \geqslant N_{0,k}(\omega) = \max_{x\in U} N_0(x,k,\omega)$. Далее, $N_{0,k} < \infty$ п.н., поскольку $\sharp\mathbb{X} < \infty$. Мы получили, что

$$\sum_{i=0}^{2m-1} \sum_{i-m<|y|\leqslant m} \widehat{\psi}_{N,k}(y) G_{N,k}^{(i),U}(y) \to 0 \quad \text{п.н.,} \quad N \to \infty. \tag{2.9}$$

Если $U = \mathbb{X}$, тогда $G_{N,k}^{(i),\mathbb{X}\setminus U}(y) = 0$ для всех рассматриваемых N, k и y. Следовательно, (2.8) проверено и, таким образом, для $U = \mathbb{X}$ выполнено соотношение (2.6). Пусть теперь $U \neq \mathbb{X}$. Как было отмечено ранее, $\cup_{J\subset\mathbb{Y},J\neq\varnothing} B_J = \mathbb{X}$. Принимая во внимание определение множества U, мы утверждаем, что $\cup_{J\subset\mathbb{Y},\sharp J>1} B_J = \mathbb{X}\setminus\mathbb{U}$. Согласно лемме 2 для $J = \{y\in\mathbb{Y} : t \leqslant y \leqslant z\}$ имеем $B_J = D_{t,z}$, где $D_{t,z} := \{x \in \mathbb{X}\setminus U : L_y^{\beta}(x) = 0, t < y \leqslant z;\ L_y^{\beta}(x) < 0, -m < y \leqslant t;\ L_y^{\beta}(x) > 0, y > z\}$, $t < z$, $t,z \in \mathbb{Y}$. Тогда для $k = 1,\dots,K$ и $N \in \mathbb{N}$ получаем

$$\sum_{i=0}^{2m-1} \sum_{i-m<|y|\leqslant m} \widehat{\psi}_{N,k}(y) G_{N,k}^{(i),\mathbb{X}\setminus U}(y) = \sum_{z=-m+1}^{m} \sum_{t=-m}^{z-1} \sum_{x\in D_{t,z}} \sum_{i=0}^{2m-1} \sum_{i-m<|y|\leqslant m} \Phi_{N,k}^{(i)}(x,y).$$

Здесь

$$\Phi_{N,k}^{(i)}(x,y) := \frac{\widehat{\psi}_{N,k}(y)}{\sqrt{\sharp S_k(N)}} \sum_{j\in S_k(N)} \mathbb{I}\{A^j(x,y)\} \times (\mathbb{I}\{|f_{PA}(x,\xi_N(\overline{S_k(N)})) - y| > i\} - \mathbb{I}\{|f(x) - y| > i\}),$$

а $A^j(x,y) = \{X^j = x, Y^j = y\}$.

Из определения $D_{t,z}$ следует, что $f(x) = t$ для $x \in M \cap D_{t,z}$. Для всех $x \in D_{t,z}$, $t < z$ и $t,z \in \mathbb{Y}$ мы проверим соотношение

$$\sum_{i=0}^{2m-1} \sum_{i-m<|y|\leqslant m} \Phi_{N,k}^{(i)}(x,y) = \sqrt{\sharp S_k(N)} \cdot (\widehat{L}^{,S_k(N)}(x,\omega))^{\top} I(N,x,k,t), \tag{2.10}$$

где компоненты случайных векторов $I(N,x,k,t)$, $N \in \mathbb{N}$, $x \in \mathbb{X}$, $k \in \{1,\dots,K\}$, $t \in \mathbb{Y}$, введены в (1.41). Прежде всего, заметим, что

$$\Phi_{N,k}^{(i)}(x,y) = \sqrt{\sharp S_k(N)} \cdot \widehat{w}_y^{S_k(N)}(x,\omega) F_{N,k}^{(i)}(x,y),$$

где $\widehat{w}_y^{S_k(N)}(x,\omega)$ определено в (1.58).

Принимая во внимание, что $\mathbb{I}\{\cup_{j\in J} D_j\} = \sum_{j=1}^{J} \mathbb{I}\{D_j\}$ для попарно не пересекающихся множеств $D_1,\ldots,D_J$, мы можем записать

$$\sum_{i=0}^{2m-1} \sum_{i-m<|y|\leqslant m} \sum_{|g-y|>i} \widehat{w}_y^{S_k(N)}(x,\omega)\delta_g(N,x,k) = \sum_{y\in\mathbb{Y}} \widehat{w}_y^{S_k(N)}(x,\omega) \sum_{i=0}^{I(y)} \sum_{|g-y|>i} \delta_g(N,x,k),$$

где $I(y) = 2m-1$ для $y = \pm m$ и $I(y) = m-1+|y|$ для $|y| < m$ ($y \in \mathbb{Y}$). Заметим, что

$$\sum_{i=0}^{I(y)} \sum_{|g-y|>i} \delta_g(N,x,k) = \sum_{g=-m}^{m} \sum_{i<|g-y|} \delta_g(N,x,k)$$
$$= \sum_{g=-m}^{m} |y-g|\delta_g(N,x,k) = (Q\delta(N,x,k))_y, \quad (2.11)$$

так как $|g-y|-1 \leqslant I(y)$ для всех $g,y \in \mathbb{Y}$. Здесь $(Q\delta(N,x,k))_y$, $y \in \mathbb{Y}$ – координаты вектора $Q\delta(N,x,k)$. Значит,

$$\sum_{i=0}^{2m-1} \sum_{i-m<|y|\leqslant m} \Phi_{N,k}^{(i)}(x,y) = \sqrt{\sharp S_k(N)} \cdot (\widehat{w}^{S_k(N)}(x,\omega))^{\top} Q\delta(N,x,k).$$

Для $x \in D_{t,z}$, $t < z$, $t,z \in \mathbb{Y}$

$$Q\,\delta(N,x,k) = \sum_{y\in\mathbb{Y}} \mathbb{I}\{f_{PA}(x,\xi_N(\overline{S_k(N)})) = y\}(q(y)-q(t)), \quad (2.12)$$

поскольку $\sum_{y\in\mathbb{Y}} \mathbb{I}\{f_{PA}(x,\xi_N(\overline{S_k(N)})) = y\} = 1$, и Q – симметричная матрица Для $y,t \in \mathbb{Y}$, согласно (1.59), имеем

$$(\widehat{w}^{S_k(N)}(x,\omega))^{\top}(q(y)-q(t)) = \begin{cases} -\widehat{L}_{y+1}^{S_k(N)}(x) - \ldots - \widehat{L}_{t}^{S_k(N)}(x), & y < t, \\ 0, & y = t, \\ \widehat{L}_{t+1}^{S_k(N)}(x) + \ldots + \widehat{L}_{y}^{S_k(N)}(x), & y > t. \end{cases} \quad (2.13)$$

Если $t = m$, то $\sum_{t+1\leqslant g\leqslant y} \widehat{L}_g^{S_k(N)} = 0$; если $t = -m$, тогда $\sum_{y+1\leqslant g\leqslant -m} L_g^{S_k(N)} = 0$, так как сумма по пустому множеству равна нулю. Изменяя порядок суммирования, приходим к формуле

$$\sum_{y<t} \sum_{g=y+1}^{t} \mathbb{I}\{f_{PA}(x,\xi_N(\overline{S_k(N)})) = y\}\widehat{L}_g^{S_k(N)}(x)$$

$$= \sum_{g=-m+1}^{t} \sum_{y=-m}^{g-1} \mathbb{I}\{f_{PA}(x,\xi_N(\overline{S_k(N)})) = y\} \widehat{L}_g^{S_k(N)}(x)$$

$$= \sum_{g=-m+1}^{t} \mathbb{I}\{f_{PA}(x,\xi_N(\overline{S_k(N)})) \leqslant g-1\} \widehat{L}_g^{S_k(N)}(x). \tag{2.14}$$

Аналогичным образом получается

$$\sum_{y>t} \sum_{g=t+1}^{y} \mathbb{I}\{f_{PA}(x,\xi_N(\overline{S_k(N)})) = y\} \widehat{L}_g^{S_k(N)}(x)$$

$$= \sum_{g=t+1}^{m} \mathbb{I}\{f_{PA}(x,\xi_N(\overline{S_k(N)})) \geqslant g\} \widehat{L}_g^{S_k(N)}(x). \tag{2.15}$$

Таким образом, из (2.14) и (2.15) следует, что

$$(\widehat{w}^{S_k(N)})^\top(x)\, Q\, \delta(N,x,k) = (\widehat{L}^{S_k(N)})^\top(x) I(N,x,k,t), \tag{2.16}$$

и (2.10) установлено.

Для произвольных фиксированных $t,z \in \mathbb{Y}$ таких, что $t < z$ и $x \in D_{t,z}$, проверим соотношение

$$\sqrt{\sharp S_k(N)} \cdot (\widehat{L}^{\beta,S_k(N)}(x,\omega))^\top I(N,x,k,t) \xrightarrow{\mathbb{P}} 0, \ \ N \to \infty. \tag{2.17}$$

Заметим, что $\widehat{L}_y^{\beta,S_k(N)}(x,\omega) = 0$ п.н. для всех $x \in \overline{M}$, $y \in \mathbb{Y}$, $y > -m$, $k = 1,\ldots,K$, и $N \in \mathbb{N}$. Получаем $(\widehat{L}^{\beta,S_k(N)}(x,\omega))^\top I(N,x,k,t) = \widehat{R}_{N,k}^{\beta,(1)}(x,t,z) + \widehat{R}_{N,k}^{\beta,(2)}(x,t,z)$, где

$$\widehat{R}_{N,k}^{\beta,(1)}(x,t,z) = \sum_{y \notin \{-m\} \cup (t,z]} \widehat{L}_y^{\beta,S_k(N)}(x,\omega) I_y(N,x,k,t),$$

$$\widehat{R}_{N,k}^{\beta,(2)}(x,t,z) = \sum_{y \in (t,r]} \widehat{L}_y^{\beta,S_k(N)}(x,\omega) I_y(N,x,k,t).$$

Ясно, что

$$|\widehat{R}_{N,k}^{\beta,(1)}(x,t,z)| \leqslant \sum_{y \notin \{-m\} \cup (t,z]} \mathbb{I}\{f_{PA}(x,\xi_N(\overline{S_k(N)})) \notin [t,z]\}$$

$$\times |\widehat{L}_y^{\beta,S_k(N)}(x,\omega) I_y(N,x,k,t)|.$$

Для произвольных $x \in D_{t,z}$, $k = 1,\ldots,K$ и почти всех $\omega \in \Omega$ соотношение (1.82) обеспечивает существование такого целого $N_3(x,k,\omega)$, что

$f_{PA}(x,\xi_N(\overline{S_k(N)})) \in [t,z]$ для $N \geqslant N_3(x,k,\omega)$. Следовательно, $\widehat{R}^{\beta,(1)}_{N,k}(x,t,z) = 0$ для любого $x \in D_{t,z}$, каждого $k = 1,\ldots,K$ и почти всех $\omega \in \Omega$, где $N \geqslant N_{3,k}(\omega) = \max_{x\in D_{t,z}} N_3(x,k,\omega)$. Заметим, что

$$\widehat{R}^{\beta,(2)}_{N,k}(x,t,z) = \sum_{y\in(t,z]} \mathbb{I}\{f_{PA}(x,\xi_N(\overline{S_k(N)})) \neq t\} \cdot \widehat{L}^{\beta,S_k(N)}_y(x,\omega) I_y(N,x,k,t). \quad (2.18)$$

Докажем, что для произвольных $x \in M \cap D_{t,z}$ и $k = 1,\ldots,K$,

$$\mathbb{I}\{f_{PA}(x,\xi_N(\overline{S_k(N)})) \neq t\} \xrightarrow{\mathbb{P}} 0, \ \ N \to \infty. \quad (2.19)$$

При любых $\nu > 0$ и $x \in M \cap D_{t,z}$ верно равенство

$$\mathbb{P}(\mathbb{I}\{f_{PA}(x,\xi_N(\overline{S_k(N)})) \neq t\} > \nu)$$

$$= \mathbb{P}\left(\widehat{L}^{\beta,\overline{S_k(N)}}_t(x,\omega) + \varepsilon_N \geqslant 0, \widehat{L}^{\beta,\overline{S_k(N)}}_{t+1}(x,\omega) + \varepsilon_N < 0\right).$$

Очевидно, существует такое $N_4 \in \mathbb{N}$, что для $x \in D_{t,z}$ и $N > N_4$ имеем $\widehat{L}^{\beta,\overline{S_k(N)}}_t(x,\omega) + \varepsilon_N < 0$. Следовательно, если $N > N_3$, то

$$\mathbb{P}\left(\widehat{L}^{\beta,\overline{S_k(N)}}_t(x,\omega) + \varepsilon_N \geqslant 0, \widehat{L}^{\beta,\overline{S_k(N)}}_{t+1}(x,\omega) + \varepsilon_N < 0\right)$$

$$= \mathbb{P}\left(\widehat{L}^{\beta,\overline{S_k(N)}}_{t+1}(x,\omega) + \varepsilon_N < 0\right).$$

Теперь мы покажем, что для $k = 1,\ldots,K$ данная вероятность стремится к 0 при $N \to \infty$. Соотношение (2.19) эквивалентно следующему:

$$\mathbb{P}\Big(\sqrt{\sharp\overline{S_k(N)}} \cdot \widehat{L}^{\beta,\overline{S_k(N)}}_{t+1}(x,\omega) < -\sqrt{\sharp\overline{S_k(N)}}\varepsilon_N\Big) \to 0, \quad N \to \infty. \quad (2.20)$$

Рассмотрим такие множества $W_N \subset \{1,\ldots,N\}$, что $\sharp W_N \to \infty$ при $N \to \infty$. Для всех $x \in M \cap D_{t,z}$, $t<z$, $t,z \in \mathbb{Y}$ и произвольного $y \in [t,z]$ имеем

$$\widehat{L}^{\beta,W_N}_y(x,\omega) = \widetilde{L}^{\beta,W_N}_y(x,\omega) + \big(\widehat{L}^{\beta,W_N}_y(x,\omega) - \widetilde{L}^{\beta,W_N}_y(x,\omega)\big).$$

Заметим, что $\mathbb{E}\widetilde{L}^{\beta,j}_y(x,\omega) = 0$ при всех $x \in D_{t,z}$, $j \in \mathbb{N}$ и $y \in (t,z]$ (мы пишем $\widetilde{L}^{\beta,j}_y(x,\omega)$ вместо $\widetilde{L}^{\beta,\{j\}}_y(x,\omega)$). $\{\widetilde{L}^{\beta,j}_y(x,\omega), j \in W_N, N \in \mathbb{N}\}$ представляют треугольный массив равномерно ограниченных центрированных случайных величин. Причем в каждой строке этого массива (при некотором N) случайные величины независимы и одинаково распределены. Поэтому применима теорема Линдеберга (теорема 27.2, [17]). Выполнение условия Линдеберга обусловлено

равномерной ограниченностью рассматриваемых случайных величин. Принимая во внимание, что $\widetilde{L}_y^{\beta,W_N}(x,\omega) = \frac{1}{\sharp W_N}\sum_{j\in W_N}\widetilde{L}_y^{\beta,j}(x,\omega)$, можем утверждать, что

$$\sqrt{\sharp W_N}\,\widetilde{L}_y^{\beta,W_N}(x,\omega) \xrightarrow{law} Z_1 \sim N(0,\sigma_{1,y}^2(x)), \quad y\in(t,z],\ N\to\infty, \tag{2.21}$$

где $\sigma_{1,y}^2(x) = var\big[\widetilde{L}_y^{\beta,j}(x,\omega)\big]$, $x\in D_{t,z}$, $y\in(t,z]$, $j\in W_N$. Для каждого $y\in\mathbb{Y}$

$$(\widehat{\psi}(y,\xi_N(W_N)) - \psi(y))\frac{1}{\sqrt{\sharp W_N}}\sum_{j\in W_N}\mathbb{I}\{X^j=x\}\mathbb{I}\{Y^j=y\}$$

$$= (\widehat{\psi}(y,\xi_N(W_N)) - \psi(y))\frac{1}{\sqrt{\sharp W_N}}$$

$$\times\sum_{j\in W_N}(\mathbb{I}\{X^j=x\}\mathbb{I}\{Y^j=y\} - \mathbb{E}\mathbb{I}\{X^j=x\}\mathbb{I}\{Y^j=y\})$$

$$+(\widehat{\psi}(y,\xi_N(W_N)) - \psi(y))\sqrt{\sharp W_N}\mathbb{P}(X=x,Y=y).$$

В силу ЦПТ для треугольного массива независимых одинаково распределенных случайных величин (теорема 27.2, [17])

$$\sum_{j\in W_N}\frac{\mathbb{I}\{X^j=x\}\mathbb{I}\{Y^j=y\} - \mathbb{E}\mathbb{I}\{X^j=x\}\mathbb{I}\{Y^j=y\}}{\sqrt{\sharp W_N}} \xrightarrow{law} Z_2\sim N(0,\sigma_{2,y}^2(x))$$

при $N\to\infty$, где $\sigma_{2,y}^2(x) = var\big[\mathbb{I}\{X^j=x,Y^j=y\}\big]$, $j\in W_N$. Выполнение условия Линдеберга следует из равномерной ограниченности рассматриваемых случайных величин. Ввиду (1.25) имеем

$$\frac{\widehat{\psi}(y,\xi_N(W_N)) - \psi(y)}{\sqrt{\sharp W_N}}\sum_{j\in W_N}(\mathbb{I}\{X^j=x\}\mathbb{I}\{Y^j=y\} - \mathbb{E}\mathbb{I}\{X^j=x\}\mathbb{I}\{Y^j=y\}) \xrightarrow{\mathbb{P}} 0$$

при $N\to\infty$. Теперь мы воспользуемся (1.65) – (1.67) еще раз для того, чтобы получить

$$\sqrt{\sharp W_N}(\widehat{\psi}(y,\xi_N(W_N)) - \psi(y)) \xrightarrow{law} Z_3\sim N(0,\sigma_{3,y}^2), \quad y\in\mathbb{Y},\ N\to\infty,$$

с $\sigma_{3,y}^2 = \mathbb{P}(Y=-y)(\mathbb{P}(Y=y))^{-3}$. Таким образом, для каждого $y\in(t,z]$ имеем

$$\sqrt{\sharp W_N}\big(\widehat{L}_y^{\beta,W_N}(x) - \widetilde{L}_y^{\beta,W_N}(x)\big) \xrightarrow{\mathbb{P}} 0, \quad N\to\infty. \tag{2.22}$$

По леммы Слуцкого

$$\sqrt{\sharp W_N}\,\widehat{L}_y^{\beta,W_N}(x) \xrightarrow{law} Z_1\sim N(0,\sigma_{1,y}^2(x)), \quad y\in(t,r],\ N\to\infty. \tag{2.23}$$

Подставим $W_N = \overline{S_k(N)}$ с $k = 1,\dots,K$ в (2.23). Тогда $\sharp\overline{S_k(N)} \geqslant (K-1)[N/K]$ для $k = 1,\dots,K$, и мы заключаем, что (2.20) справедливо, когда $\varepsilon_N N^{1/2} \to \infty$ при $N \to \infty$. Таким образом, приходим к (2.19). В силу (2.18), (2.19) и (2.22) с $W_N = S_k(N)$ получаем

$$\sum_{i=0}^{2m-1} \sum_{i-m<|y|\leqslant m} \widehat{\psi}_{N,k}(y) Q_{N,k}^{(i),\mathbb{X}\setminus U}(y) \to 0 \quad \text{п.н.}, \ N \to \infty. \tag{2.24}$$

Принимая во внимание (2.9) и (2.24), мы приходим к (2.8). Следовательно, (2.6) установлено.

Теперь обратимся к исследованию величин $\widehat{T}_N(f) - T_N(f)$, фигурирующих в (2.5). Заметим, что

$$\begin{aligned}\sqrt{N}(\widehat{T}_N(f) - T_N(f)) &\\ = \frac{\sqrt{N}}{K} \sum_{k=1}^{K} \frac{1}{\sharp S_k(N)} \sum_{i=0}^{2m-1} \sum_{i-m<|y|\leqslant m} &(\widehat{\psi}_{N,k}(y) - \psi(y)) \\ \times \sum_{j\in S_k(N)} &\mathbb{I}\{Y^j = y, |f_(X^j) - y| > i\}.\end{aligned}$$

Положим $Z^{i,j} = \mathbb{I}\{Y^j = y, |f(X^j) - y| > i\}$, $i = 0,\dots,2m-1$, $j = 1,\dots,N$. Для каждого $k = 1,\dots,K$

$$\sum_{i=0}^{2m-1} \sum_{i-m<|y|\leqslant m} (\widehat{\psi}_{N,k}(y) - \psi(y)) \frac{1}{\sqrt{\sharp S_k(N)}} \sum_{j\in S_k(N)} \mathbb{I}\{Y^j = y, |f_N(X^j) - y| > i\}$$

$$= \sum_{i=0}^{2m-1} \sum_{i-m<|y|\leqslant m} (\widehat{\psi}_{N,k}(y) - \psi(y)) \frac{1}{\sqrt{\sharp S_k(N)}} \sum_{j\in S_k(N)} (Z^{i,j} - \mathbb{E}Z^{i,j})$$

$$+\sqrt{\sharp S_k(N)} \sum_{i=0}^{2m-1} \sum_{i-m<|y|\leqslant m} (\widehat{\psi}_{N,k}(y) - \psi(y)) \mathbb{P}(Y = y, |f_N(X) - y| > i).$$

В силу (1.25) и ЦПТ (теорема 27.2, [17]) для треугольного массива независимых одинаково распределенных ограниченных случайных величин $\{Z^{i,j}, j \in S_k(N), N \in \mathbb{N}\}$ при каждом $i \in \{0,\dots,2m-1\}$ имеем

$$\sum_{i-m<|y|\leqslant m} (\widehat{\psi}_{N,k}(y) - \psi(y)) \frac{1}{\sqrt{\sharp S_k(N)}} \sum_{j\in S_k(N)} (Z^{i,j} - \mathbb{E}Z^{i,j}) \xrightarrow{\mathbb{P}} 0, \quad N \to \infty.$$

Следовательно, предельное (при $N \to \infty$) распределение

$$\sqrt{N}[(\widehat{T}_N(f) - T_N(f)) + (T_N(f) - Err(f))]$$

будет совпадать с предельным распределением случайных величин

$$\sqrt{N}\big[(T_N(f) - Err(f)) + \frac{1}{K}\sum_{k=1}^{K}\sum_{i=0}^{2m-1}\sum_{i-m<|y|\leqslant m}(\widehat{\psi}_{N,k}(y) - \psi(y))\mathbb{P}(Y = y, |f(X) - y| > i)\big]. \quad (2.25)$$

Для каждого $y \in \mathbb{Y}$ и $k = 1, \ldots, K$

$$\widehat{\mathbb{P}}_{S_k(N)}(Y = y) - \mathbb{P}(Y = y) \xrightarrow{\mathbb{P}} 0,$$

$$\sqrt{\sharp S_k(N)}(\widehat{\mathbb{P}}_{S_k(N)}(Y = y) - \mathbb{P}(Y = y)) \xrightarrow{law} Z_4 \sim N(0, \sigma_4^2),$$

$N \to \infty$, где $\sigma_4^2 = \mathbb{P}(Y \neq y)\mathbb{P}(Y = y)$.

Теперь лемма Слуцкого обеспечивает, что предельное поведение случайных величин, заданных в (2.25), такое же, как у случайных величин

$$\sqrt{N}(T_N(f) - Err(f))$$

$$-\frac{\sqrt{N}}{K}\sum_{k=1}^{K}\sum_{i=0}^{2m-1}\sum_{i-m<|y|\leqslant m}\frac{(\widehat{\mathbb{P}}_{S_k(N)}(Y = y) - \mathbb{P}(Y = y))\mathbb{P}(Y = y, |f(X) - y| > i)}{\mathbb{P}(Y = y)^2}$$

$$= \frac{\sqrt{N}}{K}\sum_{k=1}^{K}\sum_{i=0}^{2m-1}\sum_{i-m<|y|\leqslant m}\frac{1}{\sharp S_k(N)}\sum_{j\in S_k(N)}\Big(\frac{\mathbb{I}\{Y^j = y, |f(X^j) - y| > i\}}{\mathbb{P}(Y = y)}$$

$$-\frac{\mathbb{P}(Y = y, |f(X) - y| > i)}{\mathbb{P}(Y = y)} - \frac{\mathbb{I}\{Y^j = y\} - \mathbb{P}(Y = y)\mathbb{P}(Y = y, |f(X) - y|)}{\mathbb{P}(Y = y)^2}\Big)$$

$$= \frac{\sqrt{N}}{K}\sum_{k=1}^{K}\frac{1}{\sharp S_k(N)}\sum_{j\in S_k(N)}(V^j - \mathbb{E}V^j),$$

где

$$V^j = \sum_{i=0}^{2m-1}\sum_{i-m<|y|\leqslant m}\frac{\mathbb{I}\{Y^j = y\}}{\mathbb{P}(Y = y)}\left(\mathbb{I}\{|f(X^j) - y| > i\} - \frac{\mathbb{P}(Y = y, |f(X) - y| > i)}{\mathbb{P}(Y = y)}\right).$$

Для каждого $k = 1, \ldots, K$ из ЦПТ (теорема 27.2, [17]) для массива независимых одинаково распределенных ограниченных случайных величин $\{V^j, j \in S_k(N), N \in \mathbb{N}\}$ следует соотношение

$$Z_{N,k} := \frac{1}{\sqrt{\sharp S_k(N)}}\sum_{j\in S_k(N)}(V^j - \mathbb{E}V^j) \xrightarrow{law} Z \sim N(0, \sigma^2), \ N \to \infty,$$

где $\sigma^2 = var\, V$, а V было определено в (2.4). Так как $Z_{N,1},\ldots,Z_{N,K}$ независимы, а $\sqrt{N}/\sqrt{\sharp S_k(N)} \to \sqrt{K}$ для $k = 1,\ldots,K$, когда $N \to \infty$, мы приходим к (2.3). Доказательство теоремы завершено. $\square$

Напомним, что для последовательности случайных величин $(\eta_N)_{N\in\mathbb{N}}$ и последовательности положительных чисел $(a_N)_{N\in\mathbb{N}}$ мы пишем $\eta_N = o_P(a_N)$, если $\eta_N/a_N \xrightarrow{\mathbb{P}} 0,\ N \to \infty$.

Замечание 13. Как известно, ЦПТ можно рассматривать как результат, описывающий точность аппроксимации исследуемых случайных величин. Из теоремы 7 следует, что

$$\widehat{Err}_K(f_{PA},\xi_N) - Err(f) = o_P(a_N),\ N \to \infty, \tag{2.26}$$

где $a_N = o(N^{-1/2})$.

Замечание 14. Ввиду (1.63) несложно построить состоятельные оценки $\widehat{\sigma}_N$ неизвестного σ, возникающего в (2.3). Следовательно, (если $\sigma^2 \neq 0$) мы можем утверждать, что в условиях теоремы 1 справедливо соотношение

$$\frac{\sqrt{N}}{\widehat{\sigma}_N}(\widehat{Err}_K(f_{PA},\xi_N) - Err(f)) \xrightarrow{law} \frac{Z}{\sigma} \sim N(0,1),\ N \to \infty.$$

Теперь рассмотрим многомерную версию теоремы 7. Используя прием Крамера-Уолда (см., например, [2], дополнение к главе V, пункт 8) и доказательство теоремы 7, мы приходим к следующему утверждению.

Теорема 8. *Пусть $\varepsilon_N \to 0$ и $N^{1/2}\varepsilon_N \to \infty$ при $N \to \infty$. Тогда для каждого $K \in \mathbb{N}$ и произвольного $\alpha(l) = \{m_1^{(l)},\ldots,m_r^{(l)}\} \subset \{1,\ldots,n\}$, где $l = 1,\ldots,s$, имеем*

$$\sqrt{N}(Z_N^{(1)},\ldots,Z_N^{(s)})^\top \xrightarrow{law} \mathsf{Z} \sim N(0,C),\ N \to \infty.$$

Здесь $Z_N^{(l)} = \widehat{Err}_K(\widehat{f}_{PA,\varepsilon}^{\alpha(l)},\xi_N) - Err(f^{\alpha(l)})$, $l = 1,\ldots,s$, и элементы ковариационной матрицы $C = (c_{l,p})$ имеют вид

$$c_{l,p} = cov(V(\alpha(l)),V(\alpha(p))),\ l,p = 1,\ldots,s,$$

случайные величины $V(\alpha(l))$ определены аналогично V в (2.4) с f^β, замененной на $f^{\alpha(l)}$.

В заключение отметим (см. также замечание 14), что возможно построить состоятельные оценки $\widehat{C}_N$ неизвестной (невырожденной) матрицы C для

того, чтобы получить статистическую версию последней теоремы. А именно, в условиях теоремы 8 выполняется следующее соотношение

$$(\widehat{C}_N)^{-1/2}(Z_N^{(1)},\ldots,Z_N^{(s)})^\top \xrightarrow{law} C^{-1/2}\mathsf{Z} \sim N(0,I),\ \ N\to\infty,$$

где I обозначает единичную матрицу порядка s.

2.2 Теорема типа Эрдеша-Каца для перестановочных случайных величин

В [34] Эрдеш и Кац установили несколько фундаментальных результатов для распределения максимума частичных сумм $S_k := \sum_{i=1}^k X_i$, где $\{X_n\}_{n\in\mathbb{N}}$ – последовательность независимы одинаково распределенных случайных величин с нулевым математическим ожиданием и единичной дисперсией. В частности, ими было показано, что предельное (при $n\to\infty$) распределение $n^{-\frac{1}{2}}\max_{1\leqslant k\leqslant n} S_k$ равно $(2\Phi(x)-1)\mathbb{I}_{[0,\infty)}(x)$, где $\Phi(\cdot)$ обозначает функцию распределения стандартного нормального распределения.

Мы обобщим классический результат Эрдеша и Каца на случай последовательностей перестановочных случайных величин. Это позволит использовать данную статистику в более общих стохастических моделях. Перестановочные случайные величины были введены Де Финетти в [32] как случайные величины с определенным свойством условной независимости. Другими словами, их можно представлять как смеси независимых одинаково распределенных случайных величин с помощью некоторой случайной меры. Одним из первых результатов для перестановочных случайных величин была ЦПТ, доказанная Блюмом, Черновым и Тейчером в [18]. Другие результаты можно найти, например, в [26; 35; 76]. Перестановочные случайные величины интересны потому, что встречаются в различных стохастических моделях [11; 24], находят применение в генетике [48], байесовском анализе [29] и многих других областях статистики [75], [82].

Предельная теорема, которая доказывается далее, развивает теорию максимальных значений в отношении перестановочных случайных величин. В последние годы вышел ряд работ, посвященных теории максимальных значений для зависимых переменных (см., например, [45]). Ранее Берман в [13]

установил предельное распределение для максимума последовательности перестановочных случайных величин.

Наши результаты показывают, что если классические условия ЦПТ Блюма и соавторов выполнены, то мы приходим к пределу из упомянутой теоремы Эрдеша и Каца для независимых одинаково распределенных случайных величин (см. предложение 2). Если же отказаться от условия на дисперсию управляющей случайной меры, то в теореме 10 возникает распределение, которое использует функцию распределения смеси гауссовских случайных величин. Далее в следствии 3, когда отсутствуют условия на управляющую случайную меру, предельное распределение $n^{-1/2}\max_{1\leqslant k\leqslant n} S_k$ зависит от условного смещения и условной дисперсии случайных величин $\{X_n\}_{n\in\mathbb{N}}$. В частности, мы увидим, что вероятность отрицательного смещения у случайных величин $\{X_n\}_{n\in\mathbb{N}}$ вносит существенный вклад в предельное распределение.

2.2.1 Определения и вспомогательные результаты

Пусть $\Pi(n)$ обозначает множество перестановок элементов $\{1,\dots,n\}$. Последовательность случайных величин $\{X_n\}_{n\in\mathbb{N}}$ называется перестановочной, если для любого $n\in\mathbb{N}$, $X_1,\dots,X_n$ перестановочны, то есть для каждой перестановки $\pi\in\Pi(n)$,

$$Law(X_1,\dots,X_n) = Law(X_{\pi(1)},\dots,X_{\pi(n)}).$$

Иначе говоря, $(X_1,\dots,X_n)\stackrel{d}{=}(X_{\pi(1)},\dots,X_{\pi(n)})$. Концепция перестановочности была предложена Де Финетти, который, в частности, доказал, что такие последовательности условно независимы и одинаково распределены при условии σ-алгебры перестановочных событий [32].

Один из основных инструментов доказательства теорем в данной области – теорема Де Финетти, которую мы сейчас приведем. Пусть $\mathfrak{F}$ обозначает набор всех функций распределения на $\mathbb{R}$ с топологией, задаваемой слабой сходимостью функций распределения. Теорема Де Финетти утверждает, что для бесконечных последовательностей перестановочных случайных величин $\{X_n\}_{n\in\mathbb{N}}$ существует единственная вероятностная мера μ на борелевской

σ-алгебре $\mathfrak{A}$ подмножеств $\mathfrak{F}$ такая, что для любого $n \geqslant 1$, соотношение

$$\mathbb{P}\big(g(X_1,\ldots,X_n) \in B\big) = \int_{\mathfrak{F}} \mathbb{P}_F\big(g(X_1,\ldots,X_n) \in B\big)\,\mu(dF) \tag{2.27}$$

выполняется для каждого борелевского множества $B \in \mathcal{B}(\mathbb{R})$ и произвольной борелевской функции $g : \mathbb{R}^n \to \mathbb{R}$. Здесь $\mathbb{P}_F\big(g(X_1,\ldots,X_n) \in B\big)$ – вероятность события в предположении, что случайные величины $X_1,\ldots,X_n$ независимы и имеют общею функцию распределения F. Среднее $\mathbb{E}_F g(X_1,\ldots,X_n)$ получается интегрированием g по вероятностной мере F, отвечающей функции распределения F. Обозначим $\mathsf{F}\colon \Omega \to \mathfrak{F}$ случайную величину, имеющую распределение μ, фигурирующее в теореме Де Финетти. Условное математическое ожидание $\mathbb{E}_{\mathsf{F}} g(X_1,\ldots,X_n)$ определяется аналогично $\mathbb{E}_F g(X_1,\ldots,X_n)$ и само по себе является случайной величиной, так как функция распределения F случайна. Следует отметить, что теорема Де Финетти не работает для конечных наборов перестановочных случайных величин. Более детально об этом написано в [9].

Закон больших чисел для последовательностей перестновочных случайных величин был установлен Хью и Тейлером в [50]. Они показали, что для перестановочной последовательности $\{X_n\}_{n\in\mathbb{N}}$ такой, что $\mathbb{E}_{\mathsf{F}}|X_1| < \infty$ μ-п.н.,

$$\frac{1}{n}S_n \xrightarrow{\text{п.н.}} 0 \quad \text{при } n \to \infty \quad \text{тогда и только тогда, когда} \quad \mathbb{E}X_1X_2 = 0. \tag{2.28}$$

Несложно видеть, что $\mathbb{E}X_1X_2 = 0$ эквивалентно условию $\mathbb{E}_{\mathsf{F}}X_1 = 0$ μ-п.н. Как уже упоминалось, Блюм, Чернов, Розенблат и Тейчер в [18] доказали, что для перестановочных последовательностей с нулевым средним и единичной дисперсией справедлива ЦПТ тогда и только тогда, когда

$$\mathbb{E}X_iX_j = 0 \qquad \text{и} \qquad \mathbb{E}X_i^2X_j^2 = 1 \qquad \forall\, i \neq j. \tag{2.29}$$

Теорема Де Финетти может быть сформулирована (см. [9, Theorem 3.1]) в следующем виде: бесконечная последовательность перестановочных случайных величин $\{X_n\}_{n\in\mathbb{N}}$ является смесью независимых одинаково распределенных случайных величин управляемых случайной мерой F, чье вероятностное распределение определяется μ из (2.27). В этих обозначениях

$$\mathbb{P}_{\mathsf{F}}((X_1,\ldots,X_n) \in A) = \prod_{i=1}^{n} \mathsf{F}(A_i), \qquad A = A_1 \times \cdots \times A_n \in \mathcal{B}(\mathbb{R}^n),$$

и $\mathbb{E}_{\mathsf{F}}X_1 = \int_{\mathbb{R}} x\,\mathsf{F}(dx)$. Более того, если $\mathbb{E}|X_1| < \infty$, то выполняется усиленный закон больших чисел

$$\frac{1}{n}S_n \xrightarrow{\text{п.н.}} \mathbb{E}_{\mathsf{F}}X_1 \qquad \text{при} \quad n \to \infty \tag{2.30}$$

(см., например, [9], с. 17). Дополнительное условие $\mathbb{E}X_1X_2 = 0$ означает, что случайная мера F имеет нулевое среднее, то есть $\mathbb{E}_{\mathsf{F}}X_1 = 0$ μ-п.н., и мы приходим к (2.28). Далее, если $0 < \mathbb{E}X_1^2 < \infty$, то ЦПТ

$$\frac{S_n - n\mathbb{E}_{\mathsf{F}}X_1}{\sqrt{n}\sigma_{\mathsf{F}}} \xrightarrow{law} \mathcal{N}(0,1), \quad n \to \infty, \tag{2.31}$$

выполняется, где $\sigma_{\mathsf{F}}^2 := \mathbb{E}_{\mathsf{F}}(X_1 - \mathbb{E}_{\mathsf{F}}X_1)^2$. Если выполняются условия (2.29), то F имеет нулевое среднее и единичную дисперсию μ-п.н., то есть $\sigma_{\mathsf{F}}^2 = 1$ п.н., и мы приходим к ЦПТ для перестановочных случайных величин в формулировке Блюма, Чернова, Розенблата и Тейчера.

2.2.2 Предельная теорема для максимума сумм перестановочных случайных величин

В этом разделе мы устанавливаем предельное распределение для наибольшей частичной суммы последовательности перестановочных случайных величин. Вначале мы получим предельное распределение в случае, когда управляющая случайная мера F имеет нулевое среднее и единичную дисперсию. Затем мы получим предельное распределение, отказавшись от условия единичной дисперсии. Наконец, мы обобщим результат для предела $\lim_{n\to\infty}\mathbb{P}(\max(S_1,\dots,S_n) < x\sqrt{n})$ на случай перестановочных случайных величин с произвольной управляющей случайной мерой.

Для начала приведем классическую теорему Эрдеша и Каца.

Теорема 9. *[34] Пусть $\{X_n\}_{n\in\mathbb{N}}$ – последовательность независимых одинаково распределенных случайных величин с нулевым средним и единичной дисперсией, а также пусть $S_k := \sum_{i=1}^k X_i$. Тогда*

$$\lim_{n\to\infty}\mathbb{P}(\max(S_1,\dots,S_n) < x\sqrt{n}) = G(x),\ x \in \mathbb{R}, \tag{2.32}$$

где

$$G(x) := (2\Phi(x) - 1)\mathbb{I}_{[0,\infty)}(x),\ x \in \mathbb{R}, \tag{2.33}$$

и Φ обозначает функцию распределения стандартной нормальной случайной величины.

Непосредственное обобщение данной теоремы справедливо для перестановочных случайных величин. Оно получается в условиях ЦПТ, то есть когда выполняется (2.29).

Предложение 2. *Пусть* $\{X_n\}_{n\in\mathbb{N}}$ – *последовательность перестановочных случайных величин с нулевым средним и единичной дисперсией, удовлетворяющая* (2.29). *Тогда имеет место соотношение* (2.32).

Доказательство. Так как $\mathbb{P}_F(\max(S_1,\ldots,S_n) < x\sqrt{n})$ равномерно ограничена единицей, а μ – вероятностная мера, то применяя теорему Де Финетти, теорему Лебега о мажорируемой сходимости и теорему 9 для $x \in \mathbb{R}$ получаем

$$\lim_{n\to\infty} \mathbb{P}(\max(S_1,\ldots,S_n) < x\sqrt{n}) = \int_{\mathfrak{F}} \lim_{n\to\infty} \mathbb{P}_F(\max(S_1,\ldots,S_n) < x\sqrt{n})\, \mu(dF) = G(x).$$

□

Замечание 15. Заметим, что и другие предельные теоремы, установленные Эрдешем и Кацем для независимых случайных величин в [34], могут быть обобщены на случай перестановочных случайных величин аналогичным образом.

Следующим естественным шагом является обобщение предложения 2 на случай перестановочных случайных величин $\{X_n\}_{n\in\mathbb{N}}$, для которых случайная F удовлетворяет лишь условию центрированности. Согласно (2.31)

$$\frac{1}{\sqrt{n}} S_n \xrightarrow{d} Z \cdot \sigma_{\mathsf{F}}, \tag{2.34}$$

где $Z \sim \mathcal{N}(0,1)$ не зависит от σ_{F}, и F имеет распределение μ из теоремы Де Финетти. Определим $G_\mu\colon \mathbb{R} \to \mathbb{R}$ формулой

$$G_\mu(x) := \int_{\mathfrak{F}} \mathbb{I}_{(0,\infty)}(\sigma_F^2) G(x/\sigma_F)\, \mu(dF), \tag{2.35}$$

где G была задана в (2.33). Тогда мы приходим к следующему результату.

Теорема 10. *Пусть* $\{X_n\}_{n\in\mathbb{N}}$ – *такая последовательность перестановочных случайных величин с нулевым средним и дисперсией* $0 < \mathbb{E}X_1^2 < \infty$, *что* $\mathbb{E}X_1X_2 = 0$. *Тогда*

$$\lim_{n\to\infty} \mathbb{P}(\max(S_1,\ldots,S_n) < x\sqrt{n}) = \mathbb{P}(\sigma_{\mathsf{F}}^2 = 0)\mathbb{I}_{[0,\infty)}(x) + G_\mu(x), \tag{2.36}$$

где $G_{\mu}\colon \mathbb{R} \to \mathbb{R}$ *введено в* (2.35), μ – *распределение управляющей случайной меры последовательности* $\{X_n\}_{n\in\mathbb{N}}$.

Доказательство. Ввиду (2.34), теоремы Де Финетти, теоремы Лебега о мажорируемой сходимостии и теоремы 9 имеем

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}(\max(S_1,\ldots,S_n) < x\sqrt{n}) \\
&= \mathbb{P}(\sigma_{\mathsf{F}}^2 = 0)\mathbb{I}_{[0,\infty)}(x) + \int_{\mathfrak{F}} \lim_{n\to\infty} \mathbb{I}_{(0,\infty)}(\sigma_F^2)\mathbb{P}_F(\max(S_1,\ldots,S_n) < x\sqrt{n})\,\mu(dF) \\
&= \mathbb{P}(\sigma_{\mathsf{F}}^2 = 0)\mathbb{I}_{[0,\infty)}(x) + \int_{\mathfrak{F}} \mathbb{I}_{(0,\infty)}(\sigma_F^2)G(x/\sigma_F)\,\mu(dF).
\end{aligned}$$

□

Замечание 16. Заметим, что в случае перестановочных величин можно привести пример последовательности не постоянных случайных величин с $\mathbb{P}(\sigma_{\mathsf{F}}^2 = 0) > 0$ (Пример 4). В частности, теорема 10 показывает, что если последовательность случайных величин не вырождена в смысле того, что условная дисперсия положительна почти наверное, то есть $\mathbb{P}(\sigma_{\mathsf{F}}^2 > 0) = 1$, то предельное распределение в (2.36) становится смесью

$$\int_{\mathfrak{F}} G(x/\sigma_F)\,\mu(dF). \tag{2.37}$$

Замечание 17. Предельные распределения из [34] в случае перестановочных случайных величин и при условиях теоремы 10 могут быть получены аналогичным образом.

В заключение данного раздела мы применим теорему 10 для получения предельного распределения $n^{-1/2}\max_{1\leqslant k\leqslant n} S_k$ без наложения ограничений на управляющую меру последовательности перестановочных случайных величин $\{X_n\}_{n\in\mathbb{N}}$.

Следствие 3. *Пусть* $\{X_n\}_{n\in\mathbb{N}}$ – *последовательность перестановочных случайных величин с нулевым средним и дисперсией* $0 < \mathbb{E}X_1^2 < \infty$. *Тогда*

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}(\max(S_1,\ldots,S_n) < x\sqrt{n}) \\
&\qquad = \mathbb{P}(\mathbb{E}_F X_1 < 0) + \mathbb{P}(\mathbb{E}_F X_1 = 0, \sigma_F^2 = 0)\mathbb{I}_{[0,\infty)}(x) + \widetilde{G}_{\mu}(x),
\end{aligned}$$

где $\widetilde{G}_{\mu}\colon \mathbb{R} \to \mathbb{R}$ *задается формулой*

$$\widetilde{G}_{\mu}(x) = \int_{\mathfrak{F}} \mathbb{I}_{\{0\}}(\mathbb{E}_F X_1)\mathbb{I}_{(0,\infty)}(\sigma_F^2)\, G(x/\sigma_F)\, \mu(dF), \quad x \in \mathbb{R},$$

и G *определена в* (2.33).

Доказательство. Согласно теореме Де Финетти

$$\begin{aligned}\mathbb{P}(\max(S_1,\ldots,S_n) < x\sqrt{n}) &= \int_{\mathfrak{F}_+} \mathbb{P}_F(\max_{1\leqslant k\leqslant n} S_k < x\sqrt{n})\, \mu(dF)\\ &+ \int_{\mathfrak{F}_0} \mathbb{P}_F(\max_{1\leqslant k\leqslant n} S_k < x\sqrt{n})\, \mu(dF) + \int_{\mathfrak{F}_-} \mathbb{P}_F(\max_{1\leqslant k\leqslant n} S_k < x\sqrt{n})\, \mu(dF)\\ &=: I_{n,+}(x) + I_{n,0}(x) + I_{n,-}(x),\end{aligned}$$

где $\mathfrak{F}_+ := \{F \in \mathfrak{F} \,|\, \mathbb{E}_F X_1 > 0\}$, $\mathfrak{F}_0 := \{F \in \mathfrak{F} \,|\, \mathbb{E}_F X_1 = 0\}$, и $\mathfrak{F}_- := \{F \in \mathfrak{F} \,|\, \mathbb{E}_F X_1 < 0\}$.

С другой стороны, для любой $F \in \mathfrak{F}_+$ и каждого $x \neq 0$,

$$\mathbb{P}_F(\max_{1\leqslant k\leqslant n} S_k < x\sqrt{n}) \leqslant \mathbb{P}_F(S_n - n\mathbb{E}_F X_1 < x\sqrt{n}) \leqslant \frac{\sigma_F^2}{nx^2},$$

что стремится к 0 при $n \to \infty$. По теореме Лебега о мажорируемой сходимости $\lim\limits_{n\to\infty} I_{n,+}(x) = 0$. С другой стороны, следуя доказательству теоремы 10, мы имеем

$$\begin{aligned}\lim_{n\to\infty} I_{n,0}(x) &= \int_{\mathfrak{F}_0} \mathbb{I}_{(0,\infty)}(\sigma_F)\mathbb{I}_{[0,\infty)}(x)\, \mu(dF) + \int_{\mathfrak{F}_0} \mathbb{I}_{(0,\infty)}(\sigma_F)\, G(x/\sigma_F)\, \mu(dF)\\ &= \mathbb{P}(\mathbb{E}_{\mathsf{F}} X_1 = 0, \sigma_{\mathsf{F}}^2 = 0)\mathbb{I}_{[0,\infty)}(x) + \widetilde{G}_{\mu}(x).\end{aligned}$$

Наконец, если $F \in \mathfrak{F}_-$, по неравенству Гаека-Реньи для $x \neq 0$

$$\mathbb{P}_F(\max_{1\leqslant k\leqslant n} S_k < x\sqrt{n}) \geqslant 1 - \sum_{k=1}^{n} \frac{\sigma_F^2}{(x\sqrt{n} - k\mathbb{E}_F X_1)^2} \geqslant 1 - \frac{\sigma_F^2}{\mathbb{E}_F X_1 x\sqrt{n}},$$

что стремится к единице при $n \to \infty$. Согласно теореме Лебега о мажорируемой сходимости $\lim\limits_{n\to\infty} I_{n,-}(x) = \mathbb{P}(\mathbb{E}_{\mathsf{F}} X_1 < 0)$, и теорема доказана. □

Замечание 18. Заметим, что предел, возникающий в следствии 3, не является функцией распределения, когда не выполнены условия теоремы 10. Таким образом, сходимость по распределению имеет место в предложении 2 и теореме 10, но не в общем случае следствия 3.

Приведем несколько примеров, которые иллюстрируют результаты, полученные в предыдущем разделе.

Пример 3. Пусть $\{Y_n\}_{n\in\mathbb{N}}$ – последовательность случайных величин с $\mathbb{E}Y_1Y_2 = 0$ и $\mathbb{E}Y_1^2Y_2^2 = 1$, которые принимают значения $\{-1,1\}$, и пусть $\{Z_n\}_{n\in\mathbb{N}}$ – последовательность независимых случайных величин, имеющих стандартное нормальное распределение и не зависящих от $\{Y_n\}_{n\in\mathbb{N}}$. Последовательность

$$\{X_n\}_{n\in\mathbb{N}} := \{Y_n + Z_n\}_{n\in\mathbb{N}}$$

перестановочна, а случайный процесс $S_n =: \sum_{k=1}^n X_k$ может быть назван перестановочным случайным блужданием с шумом. Эта модель возникает, например, в динамическом байесовском моделировании [30], см. главу 8. Для этой последовательности $\mathbb{E}X_n = \mathbb{E}Y_n + \mathbb{E}Z_n = 0$ и

$$\mathbb{E}X_1X_2 = \mathbb{E}(Y_1 + Z_1)(Y_2 + Z_2) = \mathbb{E}Y_1Y_2 + \mathbb{E}Y_1\mathbb{E}Z_2 + \mathbb{E}Z_1\mathbb{E}Y_2 + \mathbb{E}Z_1\mathbb{E}Z_2 = 0.$$

Более того, данная последовательность не вырождена в том смысле, что

$$\mathbb{P}(\sigma_{\mathsf{F}}^2 > 0) = \mathbb{P}(\mathbb{E}_{\mathsf{F}}X_1^2 > 0) \geqslant \mathbb{P}(\mathbb{E}_{\mathsf{F}}Y_1^2 > 0) = 1,$$

и, следоватлельно, $\mathbb{P}(\sigma_{\mathsf{F}}^2 = 0) = 0$. Таким образом, мы можем получить асимптотическое распределение моментов остановки

$$T_n(x) := \inf\{k \geqslant 1 \,:\, S_k > x\sqrt{n}\},$$

изучая асимптотическое распределение максимума частичных сумм S_k. Применяя теорему 10, мы получаем

$$\begin{aligned}\lim_{n\to\infty}\mathbb{P}(T_n(x) \leqslant n) &= \lim_{n\to\infty}\mathbb{P}(\max(S_1,\ldots,S_n) > x\sqrt{n}) \\ &= 1 - \lim_{n\to\infty}\mathbb{P}(\max(S_1,\ldots,S_n) \leqslant x\sqrt{n}) \\ &= 1 - G_{\mu}(x),\end{aligned}$$

где G_{μ} задана в (2.37).

Как отмечалось в замечании 16, возможно получить перестаночную последовательность из непостоянных случайных величин с $\mathbb{P}(\sigma_{\mathsf{F}}^2 = 0) > 0$. Следующий пример демонстрирует, как подобная ситуация может возникать в приложениях.

Пример 4. В финансовом моделировании риски финансовых активов могут быть представлены как последовательность независимых одинаково распределенных случайных величин $\{\xi_n\}_{n\in\mathbb{N}}$, например, с нулевым средним и положительной дисперсией $\sigma^2 > 0$. Тогда величина $\max_{1\leqslant k\leqslant n}\sum_{i=1}^{k}\xi_i$ отображает максимальные потери втечение некоторого периода. Перестановочные модели возникают в данной задаче, если мы, например, вводим независимый индикатор Y такой, что $\mathbb{P}(Y=1)=1-\mathbb{P}(Y=0)=p$, $p\in(0{,}1)$. В этом случае последовательность

$$\{X_n\}_{n\in\mathbb{N}} := \{Y\xi_n\}_{n\in\mathbb{N}}$$

перестановочна, с $\mathbb{E}X_1 = p\,\mathbb{E}\xi_1 = 0$ и $\mathbb{E}X_1^2 = p\,\mathbb{E}\xi_1^2 = p\,\sigma^2 > 0$. Этот пример особенно нагляден, поскольку теорема 10 позволяет получить предельные вероятности максимальных потерь явно. Так как ξ_n – независимые одинаково распределенные случайные величины, мы имеем $\mathbb{E}X_1X_2 = \mathbb{E}Y^2\xi_1\xi_2 = p\,\mathbb{E}\xi_1\xi_2 = p(\mathbb{E}\xi_1)^2 = 0$ и $\mathbb{E}X_1^2X_2^2 = \mathbb{E}Y^4\xi_1^2\xi_2^2 = p\,\mathbb{E}\xi_1^2\xi_2^2 > 0$. При этом, $\mathbb{E}X_1^2X_2^2$ необязательно равняется единице.

Для частичных сумм $S_n := \sum_{i=1}^{n} X_i$ предельная вероятность

$$\lim_{n\to\infty}\mathbb{P}(\max(S_1,\ldots,S_n)\leqslant x\sqrt{n})$$

может быть получена следующим образом. Пусть $\tilde{S}_n := \sum_{k=1}^{n}\xi_k$. Применяя классический результат Эрдеша-Каца к $\tilde{S}_n/\sigma$, мы имеем

$$\lim_{n\to\infty}\mathbb{P}(\max_{1\leqslant k\leqslant n} S_k\leqslant x\sqrt{n}) = p\,G(x/\sigma) + (1-p)\mathbb{I}_{[0,\infty)}(x).$$

Покажем, что это совпадает с предельным выражением в нашей теореме 10. В нашем случае мера F дискретна и принимает значения F_1 и F_2 с вероятностями p и $1-p$ соответственно,

$$X_n = \begin{cases} \xi_n & \text{при } \mathsf{F}=F_1,\\ \\ 0 & \text{при } \mathsf{F}=F_2, \end{cases}$$

для каждого $n\in\mathbb{N}$. Таким образом, $\mathbb{E}_{F_1}X_1 = \mathbb{E}\xi_1 = 0 = \mathbb{E}_{F_2}X_2$ и $\mathsf{var}_{F_1}X_1 = \sigma^2$, $\mathsf{var}_{F_2}X_1 = 0$. В частности, $X_1\equiv 0$ при F_2, следовательно,

$$\mathbb{P}(\mathbb{E}_{\mathsf{F}}X_1 = 0, \sigma_{\mathsf{F}}^2 = 0) = \mathbb{P}(\mathsf{F}=F_2) = 1-p$$

и

$$G_{\mu}(x) = \int_{\mathfrak{F}}\mathbb{I}_{\{0\}}(\mathbb{E}_F X_1)\mathbb{I}_{(0,\infty)}(\sigma_F^2)\,G(x/\sigma_F)\,\mu(dF) = p\,G(x/\sigma),$$

как и утверждалось.

2.3 Новая версия центральной предельной теоремы для перестановочных случайных величин

В последние годы активно изучаются асимптотические свойства последовательностей перестановочных случайных величин [54], симметричных функций [38], перестановочных графов [31]. В этом разделе мы применим теорию перестановочных случайных величин для доказательства нового варианта ЦПТ для оценки функционала ошибки.

Теорема 7 утверждает, что асимптотическое распределение случайных величин $\sqrt{N}(\widehat{Err}_K(f_{PA},\xi_N) - Err(f))$ совпадает с предельным законом распределения

$$\sqrt{N}(\widehat{T}_N(f)-Err(f)) = \frac{\sqrt{N}}{K}\sum_{k=1}^{K}\sum_{i=0}^{2m-1}\sum_{i-m<|y|\leqslant m}\frac{1}{\sharp S_k(N)}\sum_{j\in S_k(N)} h_N(y,i,k,j), \quad (2.38)$$

при $N \to \infty$, где

$$h_N(y,i,k,j) = \widehat{\psi}(y,S_k(N))\mathbb{I}\{Y^j = y,|f(X^j)-y| > i\}-\psi(y)\mathbb{P}(Y = y,|f(X)-y| > i)$$

и $\widehat{\psi}(y,S_k(N)) := \widehat{\psi}(y,\xi_N(S_k(N)))$.

Очевидно, что слагаемые здесь зависимые из-за наличия множителя $\widehat{\psi}(\cdot,S_k(N))$. Для доказательства ЦПТ для случайных величин, возникающих в (2.38), мы использовали асимптотическую нормальность вектора, состоящего из двух подвекторов, одним из которых является $\sqrt{N}(\widehat{\psi}(\cdot,S_k(N)) - \psi(\cdot))$. Теперь же мы будем применять иной подход, основанный на предположении симметричности оценок $\widehat{\psi}(\cdot,S_k(N))$ штрафной функции.

Пусть $K \in \mathbb{N}$, и предположим, что $N/K = q$, где $q \in \mathbb{N}$. Тогда $\sharp S_k(N) = q$ для каждого $k = 1,\ldots,K$. Рассмотрим последовательность $K \times q$-матриц $(C^{(N)})_{N\in\mathbb{N}}$ с элементами

$$\xi_{k,j}^{(N)} := \sum_{i=0}^{2m-1}\sum_{i-m<|y|\leqslant m}\widehat{\psi}(y,S_k(N))\cdot\mathbb{I}\{Y^{j+(k-1)q} = y, |f(X^{j+(k-1)q}) - y| > i\}, \quad (2.39)$$

где $k = 1,\ldots,K$ и $j = 1,\ldots,q$. Введем

$$X_{N,j} := \frac{1}{\sqrt{K}}\sum_{k=1}^{K}\xi_{k,j}^{(N)}, \quad j = 1,\ldots,q. \quad (2.40)$$

Тогда

$$\sqrt{N}(\widehat{T}_N(f) - Err(f)) = \frac{1}{\sqrt{q}} \sum_{j=1}^{q} (X_{N,j} - \sqrt{K} Err(f)). \tag{2.41}$$

Возьмем функции $\widehat{\psi}(y,\cdot)$, которые симметричны для каждого $y \in \mathbb{Y}$. Тогда любая строка и любой столбец $C^{(N)}$ состоит из перестановочных случайных величин. Ясно, что треугольный массив $\{X_{N,j},\ 1 \leqslant j \leqslant q, N \in \mathbb{N}\}$ является построчно перестановочным.

Мы установим ЦПТ для сумм из (2.41). В [14] можно найти несколько результатов, которые обеспечивают выполнение ЦПТ, когда слагаемые $\{X_i\}_{i=1}^n$ условно одинаково распределены. А именно,

$$\frac{1}{\sqrt{n}}\big(f(X_1) + \ldots + f(X_n) - L_n\big) \xrightarrow{law} Z_{0,\sigma^2} \sim \mathcal{N}(0,\sigma^2), \tag{2.42}$$

где f – такая измеримая функция, что $\mathbb{E}|f^2(X_1)| < \infty$, и $L_n = L_n(X_1, \ldots, X_n)$. В упомянутой статье авторы использовали мартингалы. Подобный подход был развит для перестановочных величин в [71]. Чтобы доказать ЦПТ в форме (2.42) с $f(x) = x$ для построчно перестановочных массивов мы применим иной подход, основанный на результатах Röllin [62]. Пусть $\mathsf{Y} = (Y_1, \ldots, Y_m)$ – набор перестановочных случайных величин такой, что

$$\mathbb{E}Y_1 = 0, \quad \mathbb{E}|Y_1|^3 < \infty. \tag{2.43}$$

Рассмотрим $\Sigma = (\sigma_{i,j})_{1 \leqslant i,j \leqslant m}$ с $\sigma_{i,j} = \mathbb{E}(Y_i Y_j)$ (ковариационная матрица Y). Пусть $\sigma_{i,i} = \sigma^2$. Предположим, что

$$\sum_{i=1}^{m} Y_i = 0 \quad \text{п.н.} \tag{2.44}$$

Для функции $h : \mathbb{R}^d \to \mathbb{R}$ и $k \in \mathbb{N}$ положим

$$C_h^{(k)} := \max_{i_1,\ldots,i_d \geqslant 0, \sum_{j=1}^d i_j = k} \left\| \frac{\partial^k h}{\partial x_1^{i_1} \ldots \partial x_d^{i_d}} \right\|_\infty .$$

Теорема 11 (Röllin [62]). *Пусть Y – вектор, состоящий из перестановочных случайных величин и имеющий ковариационную матрицу Σ. Предположим, что условия (2.43) и (2.44) выполнены. Тогда*

$$|\mathbb{E}h(\mathsf{Y}) - \mathbb{E}h(Z)| \leqslant C_h^{(2)} \Big[\mathsf{var}\Big(\sum_{i=1}^{m} Y_i^2 \Big) \Big]^{\frac{1}{2}} + 16 m C_h^{(3)} \mathbb{E}|Y_1|^3, \tag{2.45}$$

где $Z \sim \mathcal{N}(0,\Sigma)$.

Для массива $\{X_{n,i}, 1 \leqslant i \leqslant k_n, n \in \mathbb{N}\}$ будем использовать следующие обозначения

$$\widehat{\mu}_{k_n} := \frac{1}{k_n}\sum_{i=1}^{k_n} X_{n,i}, \quad \widehat{\sigma}^2_{k_n} := \frac{1}{k_n}\sum_{i=1}^{k_n}(X_{n,i} - \widehat{\mu}_{k_n})^2. \tag{2.46}$$

Применим (2.45) для доказательства следующего результата.

Лемма 5. *Пусть $\{X_{n,i}, 1 \leqslant i \leqslant k_n, n \in \mathbb{N}\}$ – построчно перестановочный массив случайных величин, где положительные числа $k_n \to \infty$ при $n \to \infty$. Предположим, что*

1°. $\sup_{n\in\mathbb{N}} \mathbb{E}X^4_{n,1} < \infty$,

2°. $\mathbb{E}X^2_{n,1} - \mathbb{E}X_{n,1}X_{n,2} \to \sigma^2 > 0$, $n \to \infty$,

3°. $\mathsf{cov}(X^2_{n,1},X^2_{n,2}) + \mathsf{cov}(X_{n,1}X_{n,2},X_{n,3}X_{n,4}) - 2\,\mathsf{cov}(X^2_{n,1},X_{n,2}X_{n,3}) \to 0$,

при $n \to \infty$.

Тогда, для любой последовательности положительных чисел $(m_n)_{n\in\mathbb{N}}$ такой, что $m_n \to \infty$ и $m_n/k_n \to \alpha < 1$ при $n \to \infty$, выполняется следующее соотношение:

$$\frac{1}{\sqrt{m_n}}\sum_{i=1}^{m_n}(X_{n,i} - \widehat{\mu}_{k_n}) \xrightarrow{law} Z_{0,(1-\alpha)\sigma^2} \sim \mathcal{N}(0,(1-\alpha)\sigma^2), \ n \to \infty.$$

Доказательство. Прежде всего, для каждого $n \in \mathbb{N}$ введем вспомогательные случайные величины

$$Y_{n,i} := X_{n,i} - \widehat{\mu}_{k_n}, \quad i = 1,\ldots,k_n.$$

Набор $\{Y_{n,1},\ldots,Y_{n,k_n}\}$ перестановочен, так как $\{X_{n,1},\ldots,X_{n,k_n}\}$ обладает тем же свойством. Очевидно, $\sum_{i=1}^{k_n} Y_{n,i} = 0$ п.н. для всех $n \in \mathbb{N}$. Более того, $\mathbb{E}Y_{n,1} = 0$ для произвольного $n \in \mathbb{N}$. Можно показать, что

$$\mathbb{E}Y^2_{n,1} = \left(1 - \frac{1}{k_n}\right)(\mathbb{E}X^2_{n,1} - \mathbb{E}X_{n,1}X_{n,2}), \quad \mathbb{E}Y_{n,1}Y_{n,2} = -\frac{1}{k_n}(\mathbb{E}X^2_{n,1} - \mathbb{E}X_{n,1}X_{n,2}).$$

Для каждого $n \in \mathbb{N}$ возмем вектор $\mathbf{Z} = (Z_{n,1},\ldots,Z_{n,m_n})$, независимый от $(X_{n,1},\ldots,X_{n,k_n})$ и такой, что $\mathbf{Z} \sim \mathcal{N}(0,\Sigma)$. Здесь Σ – ковариационная матрица $\mathbf{Y} = (Y_{n,1},\ldots,Y_{n,m_n})$, причем $\mathsf{cov}(Z_{n,i},Z_{n,j}) = \mathsf{cov}(Y_{n,i},Y_{n,j})$, $1 \leqslant i,j \leqslant m_n$. Нетрудно видеть, что

$$\frac{1}{\sqrt{m_n}}\sum_{i=1}^{m_n}(X_{n,i} - \widehat{\mu}_{k_n}) = \frac{1}{\sqrt{m_n}}\sum_{i=1}^{m_n} Y_{n,i} =: S_{\mathbf{Y},m_n}.$$

Положим $S_{\mathbf{Z},m_n} := \frac{1}{\sqrt{m_n}} \sum_{i=1}^{m_n} Z_{n,i}$. В силу 2° и условия $m_n/k_n \to \alpha$ $(n \to \infty)$ имеем

$$\mathsf{var} S_{\mathbf{Z},m_n} = \mathbb{E}Y_{n,1}^2 + (m_n - 1)\mathbb{E}Y_{n,1}Y_{n,2}$$
$$= \left(1 - \frac{m_n}{k_n}\right)(\mathbb{E}X_{n,1}^2 - \mathbb{E}X_{n,1}X_{n,2}) \to (1-\alpha)\sigma^2.$$

Следовательно, $S_{\mathbf{Z},m_n} \xrightarrow{law} \mathcal{N}(0,(1-\alpha)\sigma^2)$, $n \to \infty$. Теперь мы покажем, что $S_{\mathbf{Y},m_n}$ и $S_{\mathbf{Z},m_n}$ имеют одинаковое предельное распределение. Согласно теореме 7.1 [16] достаточно проверить, что

$$\mathbb{E}f(S_{\mathbf{Y},m_n}) - \mathbb{E}f(S_{\mathbf{Z},m_n}) \to 0, \quad n \to \infty, \tag{2.47}$$

для любой трижды непрерывно дифференцируемой функции $f : \mathbb{R} \to \mathbb{R}$ такой, что

$$c_f^{(j)} := \left\| \frac{d^j f}{dx^j} \right\|_\infty < \infty, \ j = 1,2,3.$$

Для фиксированного $n \in \mathbb{N}$ применим теорему 11 с $m = m_n$, $Y_i = \frac{1}{\sqrt{m_n}} Y_{n,i}$, $i = 1,\ldots,m_n$, и

$$h(x_1,\ldots,x_{m_n}) := f(x_1 + \ldots + x_{m_n}).$$

Тогда можно записать

$$\left|\mathbb{E}f(S_{\mathbf{Y},m_n}) - \mathbb{E}f(S_{\mathbf{Z},m_n})\right| = \left|\mathbb{E}h(\mathbf{Y}) - \mathbb{E}h(\mathbf{Y})\right|$$
$$\leqslant C_f^{(2)} m_n^{-1} \Big[\mathsf{var}\Big(\sum_{i=1}^{m_n} Y_{n,i}^2\Big)\Big]^{\frac{1}{2}} + 16 C_f^{(3)} m_n^{-1/2} \mathbb{E}|Y_{n,1}|^3.$$

Заметим, что

$$\mathsf{var}\Big(\sum_{i=1}^{m_n} Y_{n,i}^2\Big) = m_n \mathbb{E}Y_{n,1}^4 + m_n(m_n-1)\mathbb{E}Y_{n,1}^2 Y_{n,2}^2 - m_n^2\big(\mathbb{E}Y_{n,1}^2\big)^2$$
$$= m_n\big(\mathbb{E}Y_{n,1}^4 - (\mathbb{E}Y_{n,1}^2)^2\big) + m_n(m_n-1)\mathsf{cov}(Y_{n,1}^2, Y_{n,2}^2).$$

Справедливо соотношение

$$\mathsf{cov}(Y_{n,1}^2, Y_{n,2}^2) - \Big[\mathsf{cov}(X_{n,1}^2, X_{n,2}^2) + \mathsf{cov}(X_{n,1}X_{n,2}, X_{n,3}X_{n,4}) - 2\mathsf{cov}(X_{n,1}^2, X_{n,2}X_{n,3})\Big] \to 0$$

при $n \to \infty$. Далее, положим $S_n = \frac{1}{k_n}\sum_{i=1}^{k_n} X_{n,i}$. Используя свойство перестановочности $(X_{n,1},\ldots,X_{n,k_n})$ и принимая во внимание билинейность ковариационной функции, имеем

$$\mathsf{cov}(Y_{n,1}^2,Y_{n,2}^2) = \mathbb{E}Y_{n,1}^2Y_{n,2}^2 - (\mathbb{E}Y_{n,1}^2)^2$$

$$= \mathsf{cov}(X_{n,1}^2,X_{n,2}^2) + 2\,\mathsf{cov}(X_{n,1}^2,S_n^2) - 4\,\mathsf{cov}(X_{n,1}^2,X_{n,2}S_n)$$
$$-4\,\mathsf{cov}(X_{n,1}S_n,S_n^2) + 4\,\mathsf{cov}(X_{n,1}S_n,X_{n,2}S_n) + \mathsf{cov}(S_n^2,S_n^2).$$

Для $n \to \infty$ в силу 1° верно

$$\mathsf{cov}(X_{n,1}^2,S_n^2) = \mathsf{cov}(X_{n,1}^2,X_{n,2}X_{n,3}) + O(k_n^{-1}),$$

$$\mathsf{cov}(X_{n,1}^2,X_{n,2}S_n) = \mathsf{cov}(X_{n,1}^2,X_{n,2}X_{n,3}) + O(k_n^{-1}),$$
$$\mathsf{cov}(X_{n,1}S_n,S_n^2) = \mathsf{cov}(X_{n,1}X_{n,2},X_{n,3}X_{n,4}) + O(k_n^{-1}),$$
$$\mathsf{cov}(X_{n,1}S_n,X_{n,2}S_n) = \mathsf{cov}(X_{n,1}X_{n,2},X_{n,3}X_{n,4}) + O(k_n^{-1}).$$

Следовательно, условие 3° влечет $\mathsf{cov}(Y_{n,1}^2,Y_{n,2}^2) \to 0$ при $n \to \infty$. Таким образом, соотношение (2.47) выполняется и доказательство завершено. $\square$

Замечание 19. Предположим, что

$$\sup_{n\in\mathbb{N}} \mathbb{E}\big((X_{n,1} - \widehat{\mu}_{k_n})/\widehat{\sigma}_{k_n}\big)^4 < \infty.$$

Тогда из леммы 5 следует, что для последовательности $\{m_n\}_{n\in\mathbb{N}}$ справедлива следующая ЦПТ:

$$\frac{1}{\sqrt{m_n}}\sum_{i=1}^{m_n}\left(\frac{X_{n,i} - \widehat{\mu}_{k_n}}{\widehat{\sigma}_{k_n}}\right) \xrightarrow{law} Z_{0,1-\alpha} \sim \mathcal{N}(0,1-\alpha), \quad n \to \infty.$$

Замечание 20. Похожий на лемму 1 результат был установлен в [27], но важный случай $\alpha = 0$ (который мы рассмотрим позднее) не исследовался. Результаты леммы 1 можно также получить, применяя мартингальный подход из [71]. Однако, теорема 11 из [62] позволяет оценить скорость сходимости к гауссовскому распределению. Более того, можно доказать, что при некоторых условиях асимптотическое поведение частичных сумм описывается смесью нормальных распределений.

Рассмотрим треугольный массив $\{X_{N,i}, 1 \leqslant i \leqslant q, N \in \mathbb{N}\}$ с элементами, определенными в (2.40). Положим $k_n = q$ в лемме 1 и будем писать N вместо n.

Лемма 6. *Предположим, что для каждого $N \in \mathbb{N}$, произвольного $y \in \mathbb{Y}$ и всех $k = 1, \ldots, K$*

$$\sup_{y\in\mathbb{Y},\, N\in\mathbb{N},\, k\in\{1,\ldots,K\}} \mathbb{E}\left(\widehat{\psi}(y, S_k(N))\right)^4 < \infty. \tag{2.48}$$

Пусть $(m_N)_{N\in\mathbb{N}}$ – последовательность положительных целых чисел таких, что $m_N \leqslant q$, $m_N \to \infty$ и $m_N/N \to \alpha < 1$ при $N \to \infty$. Тогда

$$\frac{1}{\sqrt{m_N}} \sum_{i=1}^{m_N} (X_{N,i} - \widehat{\mu}_N) \xrightarrow{law} Z_{0,(1-\alpha)\sigma^2} \sim \mathcal{N}(0, (1-\alpha)\sigma^2),$$

где μ_N введено в (2.46) (с $k_n = q$ и N вместо n) и

$$\sigma^2 = \mathbb{E}\left[\sum_{i=0}^{2m-1} \sum_{i-m<|y|\leqslant m} \psi(y)(\mathbb{I}\{Y=y, |f(X)-y|>i\} - \mathbb{P}(Y=y, |f(X)-y|>i))\right]^2. \tag{2.49}$$

Доказательство. Мы покажем, что условия леммы 5 выполняются. 1° следует из (2.38), (2.40) и (2.48), так как индикаторная функция принимает значения в $\{0,1\}$. Теперь обратимся к 2°. Перестановочность столбцов массива $\{\xi_{k,j}^{(N)}\}$ влечет

$$\mathbb{E}X_{N,1}X_{N,2} = \frac{1}{K}\mathbb{E}\left(\sum_{k=1}^{K} \xi_{k,1}^{(N)}\right)\left(\sum_{k=1}^{K} \xi_{k,2}^{(N)}\right) = \mathbb{E}\xi_{1,1}^{(N)}\xi_{1,2}^{(N)} + (K-1)\mathbb{E}\xi_{1,1}^{(N)}\xi_{2,2}^{(N)}.$$

По теореме Лебега о мажорируемой сходимости получаем, что предельное поведение $\mathbb{E}\xi_{1,1}^{(N)}\xi_{1,2}^{(N)}$ при $N \to \infty$ будет таким же, как и у $\mathbb{E}\zeta_{1,1}^{(N)}\zeta_{1,2}^{(N)}$, где

$$\zeta_{k,j}^{(N)} := \sum_{i=0}^{2m-1} \sum_{i-m<|y|\leqslant m} \psi(y)\, \mathbb{I}\{Y^{j+(k-1)q} = y, |f(X^{j+(k-1)q}) - y| > i\}.$$

Случайные векторы $(X^1, Y^1), (X^2, Y^2), \ldots$ независимы. Следовательно, $\mathbb{E}\zeta_{1,1}^{(N)}\zeta_{1,2}^{(N)} = \mathbb{E}\zeta_{1,1}^{(N)}\mathbb{E}\zeta_{1,2}^{(n)}$. Мы можем переписать $Err(f)$ как

$$Err(f) = \sum_{i=0}^{2m-1} \sum_{i-m<|y|\leqslant m} \psi(y)\mathbb{P}(Y = y, |f(X) - y| > i), \tag{2.50}$$

и в силу (2.50)

$$\lim_{N\to\infty} \mathbb{E}\xi_{1,1}^{(N)}\xi_{1,2}^{(N)} = \lim_{N\to\infty} \left(\mathbb{E}\zeta_{1,1}^{(N)}\right)^2 = \left(Err(f)\right)^2.$$

Аналогично приходим к соотношению

$$\lim_{N\to\infty}\mathbb{E}\xi_{1,1}^{(N)}\xi_{2,2}^{(N)}=\lim_{N\to\infty}\big(\mathbb{E}\zeta_{1,1}^{(N)}\big)^2=\big(Err(f)\big)^2.$$

Таким образом, $\mathbb{E}X_{N,1}X_{N,2}\to K\big(Err(f)\big)^2$ при $N\to\infty$. Применяя теорему Лебега еще раз, имеем

$$\lim_{N\to\infty}\mathbb{E}(X_{N,j})^2=\lim_{N\to\infty}\mathbb{E}(Z_{N,j})^2=\lim_{N\to\infty}\Big[\mathbb{E}(\zeta_{1,1}^{(N)})^2+(K-1)\mathbb{E}\zeta_{1,1}^{(N)}\zeta_{1,2}^{(N)}\Big],$$

где

$$Z_{N,j}:=\frac{1}{\sqrt{K}}\sum_{k=1}^{K}\zeta_{k,j}^{(N)},\quad j=1,\ldots,q.$$

Принимая во внимание то, что $\mathbb{E}\zeta_{1,1}^{(N)}\zeta_{1,2}^{(N)}=\big(Err(f)\big)^2$ (для каждого $N\geqslant 2K$), получаем

$$\sigma^2=\mathbb{E}\Big[\sum_{i=0}^{2m-1}\sum_{i-m<|y|\leqslant m}\psi(y)\,\mathbb{I}\{Y=y,|f(X)-y|>i\}\Big]^2-\big(Err(f)\big)^2$$

$$=\mathbb{E}\left[\sum_{i=0}^{2m-1}\sum_{i-m<|y|\leqslant m}\psi(y)\big(\mathbb{I}\{Y=y,|f(X)-y|>i\}-\mathbb{P}(Y=y,|f(X)-y|>i)\big)\right]^2.$$

Для завершения доказательства проверим условие 3°. Согласно теореме Лебега

$$\lim_{N\to\infty}\mathsf{cov}(X_{N,1}^2,X_{N,2}^2)=\lim_{k\to\infty}\mathsf{cov}(Z_{k,1}^2,Z_{k,2}^2)=0,$$

так как $Z_{k,1}$ и $Z_{k,2}$ независимые. Из аналогичных соображений следует, что $\mathsf{cov}(X_{N,1}X_{N,2},X_{N,3}X_{N,4})\to 0$ и $\mathsf{cov}(X_{N,1}^2,X_{N,2}X_{N,3})\to 0$ при $N\to\infty$. $\square$

Обсудим установленный результат. Вместо того, чтобы исследовать асимптотическое поведение $\sqrt{N}(\widehat{T}_N(f)-Err(f))$, мы можем рассмотреть предельное распределение разности двух оценок $Err(f)$. А именно, положим

$$\widehat{L}_{m_N}=\frac{1}{m_N}\sum_{j=1}^{m_N}\frac{1}{\sqrt{K}}\sum_{k=1}^{K}\sum_{i=0}^{2m-1}\sum_{i-m<|y|\leqslant m}\Big[\widehat{\psi}(y,S_k(N))\times\mathbb{I}\{Y^{j+(k-1)q}=y,|f(X^{j+(k-1)q})-y|>i\}\Big]\quad(2.51)$$

и введем $\widehat{L}_q$, как ранее, с q вместо m_n. Тогда лемма 6 утверждает, что $\sqrt{m_N}(\widehat{L}_{m_N}-\widehat{L}_q)\xrightarrow{law}Z_{0,(1-\alpha)\sigma^2}\sim\mathcal{N}(0,(1-\alpha)\sigma^2)$ при $N\to\infty$. Следовательно,

если мы обеспечим, что $\sqrt{m_N}\big(\widehat{L}_q - Err(f)\big) \xrightarrow{\mathsf{P}} 0$ при $N \to \infty$, то тогда сможем построить асимптотические доверительные интервалы для $Err(f)$. Мы ранее показали, что это возможно для регуляризованных статистик, введенных в [81].

Для последовательности случайных величин $(\eta_N)_{N\in\mathbb{N}}$ мы будем писать $\eta_N = O_{\mathbb{P}}(1)$, если для всех $\gamma > 0$ существует $M(\gamma) > 0$ такая, что $\mathbb{P}(|\eta_N| \geqslant M(\gamma)) \leqslant \gamma$ для всех достаточно больших N. Пусть $(m_N)_{N\in\mathbb{N}}$ – последовательность положительных целых чисел таких, что $m_N \leqslant q$ для $q = [N/K]$, и

$$m_N \to \infty,\ m_N/N \to 0,\ \text{когда}\ N \to \infty.$$

Теорема 12. *Пусть $(m_N)_{N\in\mathbb{N}}$ – последовательность, определенная выше. Предположим, что $\varepsilon = (\varepsilon_N)_{N\in\mathbb{N}}$ – такая последовательность положительных чисел, что $\varepsilon_N \to 0$ и $m_N^{1/2}\varepsilon_N \to \infty$ при $N \to \infty$. Рассмотрим произвольный вектор $\beta = (k_1, \dots k_r)$ с $1 \leqslant k_1 < \dots < k_r \leqslant n$, соответствующую функцию $f = f^\beta$ и предсказательный алгоритм $f_{PA} = \widehat{f}^{\beta}_{PA,\varepsilon}$. Пусть для всех $y \in \mathbb{Y}$ и $k \in \{1, \dots, K\}$ оценка $\widehat{\psi}(y, S_k(N))$ сильно состоятельна и*

$$\sqrt{\sharp S_k(N)}(\widehat{\psi}(y, S_k(N)) - \psi(y)) = O_{\mathbb{P}}(1),\ N \to \infty. \tag{2.52}$$

Пусть также выполнено (2.48). Тогда при $N \to \infty$ справедливо соотношение

$$\sqrt{m_N}\Big(\frac{1}{m_N}\sum_{j=1}^{m_N}\frac{1}{\sqrt{K}}\sum_{k=1}^{K}\sum_{i=0}^{2m-1}\sum_{i-m<|y|\leqslant m}\Big[\widehat{\psi}(y, S_k(N))\mathbb{I}\{A_N(i,j,k,y)\}\Big] - Err(f)\Big) \xrightarrow{law} Z_{0,\sigma^2}.$$

Здесь $A_N(i,j,k,y) = \{Y^{j+(k-1)q} = y, |f_{PA}(X^{j+(k-1)q}) - y| > i\}$, $Z_{0,\sigma^2} \sim \mathcal{N}(0,\sigma^2)$ и σ^2 были определены в (2.49).

Доказательство. Можно показать, что

$$\sqrt{m_N}\big(\widehat{L}_q - Err(f)\big)$$

$$-\frac{\sqrt{m_N}}{\sqrt{K}}\sum_{k=1}^{K}\sum_{i=0}^{2m-1}\sum_{i-m<|y|\leqslant m}\Big[(\widehat{\psi}(y, S_k(N)) - \psi(y))\mathbb{P}(Y = y, |f(X) - y| > i)$$

$$+\psi(y)\Big(\frac{1}{\sharp S_k(N)}\sum_{j\in S_k(N)} H_N(y,i,j)\Big)\Big] \xrightarrow{\mathsf{P}} 0$$

при $N \to \infty$. Здесь

$$H_N(y,i,j) = \mathbb{I}\{Y^j = y, |f(X^j) - y| > i\} - \mathbb{P}(Y = y, |f(X) - y| > i).$$

Для любых $i \in \{0,\dots,2m-1\}$ и $y \in \mathbb{Y}$ из теоремы Линдеберга (теорема 27.2, [17]) для массивов независимых одинаково распределенных случайных величин следует

$$\frac{1}{\sqrt{\sharp S_k(N)}} \sum_{j \in S_k(N)} H_N(y,i,j) = O_{\mathbb{P}}(1), \quad N \to \infty.$$

Выполнение условия Линдеберга обусловлено равномерной ограниченностью рассматриваемых случайных величин. Так как $m_N/\sharp S_k(N) \to 0$, мы имеем

$$\frac{\sqrt{m_N}}{\sqrt{K}} \sum_{k=1}^{K} \sum_{i=0}^{2m-1} \sum_{i-m<|y|\leqslant m} \psi(y) \frac{1}{\sharp S_k(N)} \sum_{j \in S_k(N)} H_N(y,i,j) \xrightarrow{\mathsf{P}} 0, \quad N \to \infty.$$

Аналогично, в силу (2.52),

$$\frac{\sqrt{m_N}}{\sqrt{K}} \sum_{k=1}^{K} \sum_{i=0}^{2m-1} \sum_{i-m<|y|\leqslant m} \big(\widehat{\psi}(y,S_k(N)) - \psi(y)\big)\mathbb{P}(Y = y, |f(X) - y| > i) \xrightarrow{\mathsf{P}} 0,$$

при $N \to \infty$. В условиях теоремы 12 асимптотическое поведение $\sqrt{m_N}\big(\widehat{L}_{m_N} - \widehat{L}_q\big)$ такое же, как у $\sqrt{m_N}\big(\widehat{L}_{m_N} - Err(f)\big)$. $\square$

Как уже отмечалось в главе 1, в [68] была введена следующая штрафная функция ψ

$$\psi(y) = \frac{c}{\mathbb{P}(Y = y)}, \ y \in \mathbb{Y}, \ c = \text{const} > 0.$$

Этот выбор был обоснован в [19] в случае бинарного отклика Y. Мы также будем использовать подобную штрафную функцию и в многомерном случае, а именно, когда$\mathbb{Y} = \{-m,\dots,0,\dots,m\}$. Далее мы предполагаем, что $\mathbb{P}(Y = y) > 0$ для всех $y \in \mathbb{Y}$, и без ограничения общности положим $c = 1$.

Введем $A_N(y,S_k(N)) = \{Y^j \neq y, \ j \in S_k(N)\}$, $N \in \mathbb{N}$, $k \in \{1,\dots,K\}$, $y \in \mathbb{Y}$ и положим (считаем по определению $0/0 := 0$)

$$\widehat{\mathbb{P}}_{S_k(N)}(Y = y) := \frac{\sum_{j \in S_k(N)} \mathbb{I}\{Y^j = y\}}{\sharp S_k(N)}, \ \widehat{\psi}(y,S_k(N)) := \frac{\mathbb{I}\{\Omega \setminus A_N(y,S_k(N))\}}{\widehat{\mathbb{P}}_{S_k(N)}(Y = y)}. \tag{2.53}$$

Следствие 4. *Оценка* $\widehat{\psi}$*, определенная в* (2.53)*, удовлетворяет условиям теоремы 12.*

Доказательство. Зафиксируем произвольные $y \in \mathbb{Y}$ и $k = 1,\dots,K$. Несложно проверить, что $\widehat{\psi}(y,S_k(N))$ – сильно состоятельная оценка $\psi(y)$. Кроме того, из теоремы Линдеберга (теорема 27.2, [17]) для массивов независимых одинаково распределенных случайных величин имеем

$$\sqrt{\sharp\, S_k(N)}\big(\widehat{\mathbb{P}}_{S_k(N)}(Y=y)-\mathbb{P}(Y=y)\big) \xrightarrow{law} Z_{0,\sigma_1^2(y)} \sim \mathcal{N}\big(0,\sigma_1^2(y)\big), \quad N\to\infty,$$

где $\sigma_1^2(y) = \mathbb{P}(Y=y)(1-\mathbb{P}(Y=y))$. Выполнение условия Линдеберга обусловлено равномерной ограниченностью рассматриваемых случайных величин. Принимая во внимание, что $\widehat{\mathbb{P}}_{S_k(N)}(Y=y)\to\mathbb{P}(Y=y)$ п.н. и $\sqrt{S_k(N)}\mathbb{I}\{A_N(y,S_k(N))\} \xrightarrow{\mathsf{P}} 0$ при $N\to\infty$, по лемме Слуцкого получаем, что

$$\sqrt{\sharp S_k(N)}\big(\widehat{\psi}_{S_k(N)}(y)-\psi(y)\big) \xrightarrow{law} Z_{0,\sigma_2^2(y)} \sim \mathcal{N}\big(0,\sigma_2^2(y)\big), \quad N\to\infty,$$

где $\sigma_2^2(y) = (1-\mathbb{P}(Y=y))\mathbb{P}(Y=y)^{-3}$. Таким образом, (2.52) выполняется. Теперь мы проверим (2.48). Ясно, что $\widehat{\psi}(y,S_k(N)) \leqslant \sharp S_K(N)$ для всех $N\in\mathbb{N}$. Положим $\varepsilon := \min_{y\in\mathbb{Y}}\mathbb{P}(Y=y)$. Тогда по неравенству Хефдинга

$$\begin{aligned}\mathbb{E}|\widehat{\psi}(y,S_k(N))|^4 &= \mathbb{E}\Big[|\widehat{\psi}_{N,k}(y)|^4\mathbb{I}\Big\{\big|\widehat{\mathbb{P}}_{S_k(N)}(Y=y)-\mathbb{P}(Y=y)\big|>\varepsilon/2\Big\}\Big]\\ &+\mathbb{E}\Big[(\widehat{\psi}(y,S_k(N)))^4\mathbb{I}\Big\{\big|\widehat{\mathbb{P}}_{S_k(N)}(Y=y)-\mathbb{P}(Y=y)\big|\leqslant\varepsilon/2\Big\}\Big]\\ &\leqslant 2(\sharp S_K(N))^4\exp\{-\sharp S_1(N)\varepsilon^2/2\}+2^4/\varepsilon^4.\end{aligned}$$

Таким образом мы приходим к (2.48). □

С целью упрощения обозначений мы будем писать в следующей теореме $\widehat{Err}_K(f_{PA},\xi_N)$ для случайной величины, введенной в (1.24), заменяя $\widehat{\psi}(y,S_k(N))$ на $\widehat{\psi}(y,\overline{S_k(N)})$, $y\in\mathbb{Y}$, $k=1,\dots,K$. После замены в (2.39) – (2.41) мы получим новый построчно перестановочный массив $\{X_{N,j}, 1\leqslant j\leqslant q, N\in\mathbb{N}\}$, а, значит, все установленные результаты остаются справедливыми и в этом случае.

Теорема 13. *Пусть* $\varepsilon_N\to 0$ *и* $N^{1/2}\varepsilon_N\to\infty$ *при* $N\to\infty$*. Если для любого* $K\in\mathbb{N}$ *и произвольного вектора* $\beta = (k_1,\dots k_r)$ *с* $1\leqslant k_1<\dots<k_r\leqslant n$ *соответствующие* $f = f^\beta$ *и предсказательный алгоритм определяются как* $f_{PA} = \widehat{f}^{\beta}_{PA,\varepsilon}$*, то выполняется следующее соотношение:*

$$\sqrt{N}(\widehat{Err}_K(f_{PA},\xi_N) - Err(f)) \xrightarrow{law} Z_{0,\sigma^2} \sim \mathcal{N}(0,\sigma^2), \ N\to\infty. \tag{2.54}$$

Здесь σ^2 *– дисперсия случайной величины* V*, заданной в* (2.4).

Доказательство. Введем для $f : \mathbb{X} \to \mathbb{Y}$ и $N \in \mathbb{N}$

$$T_N(f) := \frac{1}{K}\sum_{k=1}^{K}\frac{1}{\sharp S_k(N)}\sum_{i=1}^{2m-1}\sum_{i-m<|y|\leqslant m}\psi(y)\sum_{j\in S_k(N)}\mathbb{I}\{Y^j = y,|f(X^j)-y|>i\}.$$

Согласно лемме Слуцкого предельное поведение случайных величин, определенных в (2.38), будет совпадать с предельным распределением

$$\rho_N := \sqrt{N}(T_N(f) - Err(f))$$

$$-\frac{\sqrt{N}}{K}\sum_{k=1}^{K}\sum_{i=0}^{2m-1}\sum_{i-m<|y|\leqslant m}\frac{(\widehat{\mathsf{P}}_{\overline{S_k(N)}}(Y=y)-\mathsf{P}(Y=y))\mathsf{P}(Y=y,|f(X)-y|>i)}{\mathsf{P}(Y=y)^2}$$

$$=\frac{\sqrt{N}}{K}\sum_{k=1}^{K}\sum_{i=0}^{2m-1}\sum_{i-m<|y|\leqslant m}\frac{1}{\sharp\, S_k(N)}\sum_{j\in S_k(N)}\frac{\mathbb{I}\{Y^j=y,|f(X^j)-y|>i)}{\mathsf{P}(Y=y)}-$$

$$-\frac{\sqrt{N}}{K}\sum_{k=1}^{K}\sum_{i=0}^{2m-1}\sum_{i-m<|y|\leqslant m}\frac{1}{\sharp\,\overline{S_k(N)}}\sum_{j\in\overline{S_k(N)}}\mathbb{I}\{Y^j=y\}\frac{\mathsf{P}(Y=y,|f(X)-y|>i\}}{(\mathsf{P}(Y=y))^2}.$$

Пусть a_k, b_k, $k = 1,\ldots,K$, – произвольные вещественные числа. Используем следующее наблюдение:

$$\frac{1}{\sharp\, S_k(N)}\sum_{k=1}^{K}a_k+\frac{1}{\sharp\,\overline{S_k(N)}}\sum_{l=1,\ldots,K;l\neq k}b_l=\sum_{k=1}^{K}\left(\frac{a_k}{\sharp\, S_k(N)}+b_k\sum_{l=1,\ldots,K;l\neq k}\frac{1}{\sharp\,\overline{S_l(N)}}\right).$$

Комбинируя последние выражения, получаем

$$\rho_N=\frac{\sqrt{N}}{K}\sum_{k=1}^{K}\sum_{j\in S_k(N)}\left(\frac{V_1^j}{\sharp\, S_k(N)}+V_2^j\sum_{l=1,\ldots,K;l\neq k}\frac{1}{\sharp\,\overline{S_l(N)}}\right),$$

где

$$V_1^j=\sum_{i=0}^{2m-1}\sum_{i-m<|y|\leqslant m}\frac{\mathbb{I}\{Y^j=y,|f(X^j)-y|>i\}}{\mathsf{P}(Y=y)},$$

$$V_2^j=-\sum_{i=0}^{2m-1}\sum_{i-m<|y|\leqslant m}\frac{\mathbb{I}\{Y^j=y\}\mathsf{P}(Y=y,|f(X)-y|>i)}{(\mathsf{P}(Y=y))^2}.$$

Возьмем произвольное $k \in \{1,\ldots,K\}$ и воспользуемся ЦПТ для массивов ограниченных центрированных независимых одинаково распределенных случайных величин $\{V_1^j + V_2^j, j \in S_k(N), N \in \mathbb{N}\}$. Тогда

$$\frac{1}{\sqrt{\sharp\, S_k(N)}}\sum_{j\in S_k(N)}(V_1^j+V_2^j)\xrightarrow{law}Z_{0,\sigma^2}\sim\mathcal{N}(0,\sigma^2),\ N\to\infty,$$

где $\sigma^2 = \mathsf{var}(V_1^j + V_2^j)$. Заметим, что для каждого $k = 1, \ldots, K$

$$\frac{N}{\sharp\, S_k(N)} \to \frac{1}{K}, \quad \sum_{l=1,\ldots,K; l\neq k} \frac{N}{\sharp\, \overline{S_l(N)}} \to \frac{1}{K}, \quad N \to \infty.$$

Для всех $N \in \mathbb{N}$ семейства случайных величин $\{V_1^j + V_2^j, j \in S_k(N)\}$, $k = 1, \ldots, K$, независимы. Таким образом, мы приходим к следующему соотношению:

$$\rho_N \xrightarrow{law} Z_{0,\sigma^2} \sim \mathcal{N}(0, \sigma^2), \ N \to \infty.$$

Очевидно, что $\sigma^2 = \mathsf{var}\, V$, где V было введено в (2.4). Теорема доказана. □

Замечание 21. Несложно построить состоятельные оценки $\widehat{\sigma}_N$ неизвестного параметра σ, участвующего в (2.54). Следовательно, (при $\sigma^2 \neq 0$) мы можем утверждать, что в условиях теоремы 13

$$\frac{\sqrt{N}}{\widehat{\sigma}_N}(\widehat{Err}_K(f_{PA}, \xi_N) - Err(f)) \xrightarrow{law} \frac{Z}{\sigma} \sim \mathcal{N}(0,1), \ N \to \infty.$$

Глава 3. Последовательный отбор переменных в MDR-EFE методе.

При подготовке данной главы диссертации использован материал публикаций [89; 84]. Обе публикации выполнены автором без соавторов.

Преимуществом MDR-EFE является его непараметричность, в отличие, например, от LASSO [66] или классического метода наименьших квадратов. Однако, это влечет высокую вычислительную сложность связанную с тем, что необходимо находить оценку функционала ошибки для всевозможных поднаборов факторов. В работах [40; 57] рассматриваются реализации классического MDR метода для высокопроизводительных кластерах. Тем не менее, применение этих реализации времязатратные и ресурсоемкие. В связи с этим в данной главе предлагается версия MDR-EFE метода с последовательным отбором факторов. Методы с последовательным выбором переменных (sequential search, sequential forward selection) находят широкое применение в практике благодаря своей простоте и скорости работы [46]. Однако, для таких алгоритмов редко существуют строгие теоретические результаты, выявляющие качество их работы, когда данные подчинены той или иной модели. В работе [72] устанавливается ряд асимптотических результатов для последовательного отбора переменных в модели линейной регрессии. Мы рассмотрим модель наивного байесовского классификатора. В первом разделе этой главы выводится оценка снизу вероятности выбора правильного набора значимых факторов.

3.1 Логистическая регрессия и наивный байесовский классификатор.

Пусть, как и ранее Y, обозначает функцию отклика, а вектор $X = (X_1, \ldots, X_n)$ – набор факторов (признаков, переменных). В частности, Y может характеризовать клинический статус человека (болен или здоров), а набор факторов X – описательные характеристики индивида (возраст, давление, наличие определенных мутаций и тд.). Здесь и далее $n \in \mathbb{N}$ – количество наблюдаемых факторов.

Мы будем предполагать, что выполняются следующие *условия*:

1. $X_i : \Omega \to \{0,1\}$, $i = 1, \ldots, n$; $Y : \Omega \to \{-1,1\}$;

2. Y зависит только от $X_1, \ldots, X_r$ для некоторого натурального $r < n$ и не зависит от $X_{r+1}, \ldots, X_n$;

3. $\mathbb{P}(X|Y) = \prod_{i=1}^{n} \mathbb{P}(X_i|Y)$.

Последне условие означает, что данные удовлетворяют *модели наивного байесовского классификатора*. Иными словами, компоненты случайного вектора X условно независимы относительно случайного отклика Y. Для упрощения записи мы пишем $\mathbb{P}(X|Y)$ вместо $\mathbb{P}(X = x|Y = y)$, $x \in \mathbb{X}$, $y \in \mathbb{Y}$. Далее мы будем использовать обозначения

$$\theta_{v,y}^{(i)} := \mathbb{P}(X_i = v|Y = y), \tag{3.1}$$

где $v \in \{0,1\}$, $y \in \{-1,1\}$. Тогда можно записать

$$\mathbb{P}(X_i|Y = 1) = (\theta_{1,1}^{(i)})^{X_i} (\theta_{0,1}^{(i)})^{1-X_i}. \tag{3.2}$$

Покажем, что из введенных выше условий следует, что данные удовлетворяют модели логистической регрессии. Для доказательства теорем нам удобно будет иметь выражения для коэффициентов логистической регрессии через $\theta_{v,y}^{(i)}$.

Применяя формулу Байеса, запишем выражение для условной вероятности $\mathbb{P}(Y = 1|X)$:

$$\begin{aligned}\mathbb{P}(Y = 1|X) &= \frac{\mathbb{P}(X|Y = 1)\mathbb{P}(Y = 1)}{\mathbb{P}(X|Y = -1)\mathbb{P}(Y = -1) + \mathbb{P}(X|Y = 1)\mathbb{P}(Y = 1)} \\ &= \frac{1}{1 + \frac{\mathbb{P}(X|Y=-1)\mathbb{P}(Y=-1)}{\mathbb{P}(X|Y=1)\mathbb{P}(Y=1)}} = \frac{1}{1 + \exp\Big(\log\big(\frac{\mathbb{P}(X|Y=-1)\mathbb{P}(Y=-1)}{\mathbb{P}(X|Y=1)\mathbb{P}(Y=1)}\big)\Big)} \\ &= \frac{1}{1 + \exp\Big(\log\big(\frac{\mathbb{P}(Y=-1)}{\mathbb{P}(Y=1)}\big) + \sum_{i=1}^{n} \log\big(\frac{\mathbb{P}(X_i|Y=-1)}{\mathbb{P}(X_i|Y=1)}\big)\Big)}.\end{aligned} \tag{3.3}$$

Используя (3.2), имеем

$$\begin{aligned}\sum_{i=1}^{n} \log \frac{\mathbb{P}(X_i|Y = -1)}{\mathbb{P}(X_i|Y = 1)} &= \sum_{i=1}^{n} \log \frac{(\theta_{1,-1}^{(i)})^{X_i}(1 - \theta_{1,-1}^{(i)})^{1-X_i}}{(\theta_{1,1}^{(i)})^{X_i}(1 - \theta_{1,1}^{(i)})^{1-X_i}} \\ &= \sum_{i=1}^{n} \Big(X_i \log \frac{\theta_{1,-1}^{(i)}}{\theta_{1,1}^{(i)}} + (1 - X_i) \log \frac{(1 - \theta_{1,-1}^{(i)})}{(1 - \theta_{1,1}^{(i)})}\Big).\end{aligned}$$

Наконец, мы получаем

$$logit(\mathbb{P}(Y = 1|X)) = -\Big[\log \frac{\mathbb{P}(Y = -1)}{\mathbb{P}(Y = 1)} + \sum_{i=1}^{n} \frac{1 - \theta_{1,-1}^{(i)}}{1 - \theta_{1,1}^{(i)}}\Big]$$

$$+\sum_{i=1}^{n} X_i\Big(-\log\frac{\theta_{1,-1}^{(i)}(1-\theta_{1,1}^{(i)})}{\theta_{1,1}^{(i)}(1-\theta_{1,-1}^{(i)})}\Big)$$
$$=\beta_0+\sum_{i=1}^{n}\beta_i X_i,$$

что соответствует модели логистической регрессии. Здесь

$$logit(x):=\log(x/(1-x)),\quad x\in(0,1),$$

обозначает логистическое преобразование. Коэффициенты β_i, $i=0,1,\ldots,n$, определяются следующим образом:

$$\beta_0=-\Big[\log\frac{\mathbb{P}(Y=-1)}{\mathbb{P}(Y=1)}+\sum_{i=1}^{n}\frac{1-\theta_{1,-1}^{(i)}}{1-\theta_{1,1}^{(i)}}\Big], \tag{3.4}$$

$$\beta_i=\Big(-\log\frac{\theta_{1,-1}^{(i)}(1-\theta_{1,1}^{(i)})}{\theta_{1,1}^{(i)}(1-\theta_{1,-1}^{(i)})}\Big),\quad i=1,\ldots,n. \tag{3.5}$$

Переобозначая компоненты вектора X, всегда можно добиться того, что $\beta_i>0$, $i=1,\ldots,r$. Без ограничения общности, всюду далее будем считать, что $\beta_1>\beta_2>\ldots>\beta_r$. Так как Y зависит только от $X_1,\ldots,X_r$, легко видеть, что $\beta_{r+1}=0,\ldots,\beta_n=0$.

Как уже было отмечено ранее, мы считаем, что Y зависит не от всех $X_1,\ldots,X_n$, а лишь от некоторого поднабора факторов. Наша задача – по выборке из распределения $Law(X,Y)$ научиться определять, от каких факторов зависит Y. В предыдущих главах мы рассматривали MDR-EFE метод, основанный на оценке *функционала ошибки* $Err(f)$ (см. (1.5)). Для оценки $Err(f)$ используется статистика $\widehat{Err}_K(f_{PA},\xi_N)$ из (1.7). Будем использовать штрафную функцию ψ, заданную в (1.61). А для оценки ее значений применять (1.62).

3.2 MDR-EFE с последовательным отбором переменных.

Далее мы будем рассматривать модификацию метода MDR-EFE, в которой не требуется вычислять оценки функционала ошибки для всех возможных комбинаций факторов. Взамен этого отбор факторов производится последовательно (в англоязычной литературе пишут forward selection, если факторы

последовательно добавляются на каждом шаге, если же из полного набора $X_1,\ldots,X_n$ факторы последовательно удаляются, то говорят о backward selection). Будем считать, что нам известно число значимых факторов r. На первом шаге находится оценка функционала ошибки для всех наборов из одного фактора и выбирается фактор X_{i_1} с наименьшей ошибкой $\widehat{Err}_K(\widehat{f}_{PA}^{(i_1)})$. Здесь $\widehat{f}_{PA}^{(i_1)}$ обозначает предсказательный алгоритм, который для оценок Y использует только i_1 компоненту вектора факторов X (см. (1.76)). На следующем шаге рассматриваются наборы из двух факторов вида $(X_{i_1},\cdot)$, и выбирается такой фактор X_{i_2}, что среди всех указанных пар набор (X_{i_1},X_{i_2}) имеет наименьшую ошибку. Аналогично, на каждом из оставшихся $r-2$ шагов добавляется по одному фактору. В итоге мы получаем набор $(X_{i_1},\ldots,X_{i_r})$, который считаем значимым. Естественным образом возникает вопрос, какова вероятность того, что набор $(X_{i_1},\ldots,X_{i_r})$ будет состоять из значимых факторов и только из них. Далее устанавливаются оценки для упомянутой вероятности.

Будем считать, что выполняются условия 1-3 из раздела 3.1, и $\beta_1>\beta_2>\ldots>\beta_r>0$, а также $\beta_j=0$ при $j=r+1,\ldots,n$. Пусть s_i – номер фактора, выбранного в качестве значимого на i-ом шаге алгоритма ($i=1,\ldots,r$). Для описанной выше модели можно ожидать, что наибольшее значение будет иметь вероятность $\mathbb{P}(s_1=1,\ldots,s_r=r)$ при достаточно больших N. Для данной вероятности мы получим оценку снизу и увидим, что она стремится к 1 при росте числа наблюдений N.

Теорема 14. *Пусть выполнены условия, достаточные для справедливости утверждения теоремы 7. Тогда*

$$\mathbb{P}(s_1=1,\ldots,s_r=r)\geqslant\prod_{k=1}^{r}\Big(1-\frac{1}{N}\sum_{t=k+2}^{n}\frac{8V_{max}+o(1)}{\big(c_{k+1,t}^{(1,\ldots,k)}+o(1/\sqrt{N})\big)^2}\Big),N\to\infty,\quad(3.6)$$

где V_{max} и $c_{k+1,t}^{(1,\ldots,k)}$ – некоторые положительные константы.

Доказательство. Очевидно, что справедливо равенство

$$\mathbb{P}(s_1=1,\ldots,s_r=r)=\mathbb{P}(s_1=1)\mathbb{P}(s_2=2|s_1=1)\times\ldots$$

$$\times\mathbb{P}(s_r=r|s_1=1,\ldots,s_{r-1}=r-1).$$

Для начала оценим вероятность $\mathbb{P}(s_{k+1}=k+1|s_1=1,\ldots,s_k=k)$, $k=1,\ldots,r-1$, того, что на $k+1$ шаге алгоритма будет выбран фактор X_{k+1}

при условии, что на предыдущих шагах были выбраны факторы $X_1, \ldots, X_k$ именно в таком порядке. Ясно, что

$$\mathbb{P}(s_{k+1} = k+1 | s_1 = 1, \ldots, s_k = k) = \mathbb{P}\Big(\widehat{Err}_K(\widehat{f}_{PA}^{(1,\ldots,k,k+1)})$$

$$< \widehat{Err}_K(\widehat{f}_{PA}^{(1,\ldots,k,k+2)}), \ldots, \widehat{Err}_K(\widehat{f}_{PA}^{(1,\ldots,k,k+1)}) < \widehat{Err}_K(\widehat{f}_{PA}^{(1,\ldots,k,n)})\Big)$$

$$\geqslant 1 - \sum_{t=k+2}^{n} \mathbb{P}\Big(\widehat{Err}_K(\widehat{f}_{PA}^{(1,\ldots,k,k+1)}) \geqslant \widehat{Err}_K(\widehat{f}_{PA}^{(1,\ldots,k,t)})\Big) \tag{3.7}$$

Для получения оценки снизу преобразуем слагаемое из правой части (3.7) для произвольного $t = k+2, \ldots, n$:

$$\mathbb{P}\Big(\widehat{Err}_K(\widehat{f}_{PA}^{(1,\ldots,k,k+1)}) < \widehat{Err}_K(\widehat{f}_{PA}^{(1,\ldots,k,t)})\Big) = \mathbb{P}\Big(\widehat{Err}_K(\widehat{f}_{PA}^{(1,\ldots,k,k+1)})$$

$$-\mathbb{E}\widehat{Err}_K(\widehat{f}_{PA}^{(1,\ldots,k,k+1)}) + \mathbb{E}\widehat{Err}_K(\widehat{f}_{PA}^{(1,\ldots,k,k+1)}) - Err(f_{PA}^{(1,\ldots,k,k+1)})$$

$$+Err(f_{PA}^{(1,\ldots,k,t)}) - Err(f_{PA}^{(1,\ldots,k,t)}) + Err(f_{PA}^{(1,\ldots,k,t)})$$

$$-\mathbb{E}\widehat{Err}_K(\widehat{f}_{PA}^{(1,\ldots,k,t)}) + \mathbb{E}\widehat{Err}_K(\widehat{f}_{PA}^{(1,\ldots,k,t)}) - \widehat{Err}_K(\widehat{f}_{PA}^{(1,\ldots,k,t)}) < 0\Big)$$

$$= \mathbb{P}\Big(\xi_0^{(1,\ldots,k,k+1)} + \xi_{exp}^{(1,\ldots,k,k+1)} - \xi_{exp}^{(1,\ldots,k,t)} - \xi_0^{(1,\ldots,k,t)} < c_{k+1,t}^{(1,\ldots,k)}\Big),$$

где

$$c_{k+1,t}^{(1,\ldots,k)} := Err(f_{PA}^{1,\ldots,k,t}) - Err(f_{PA}^{1,\ldots,k,k+1}) > 0,$$

$$\xi_0^{(1,\ldots,k)} := \widehat{Err}_K(\widehat{f}_{PA}^{(1,\ldots,k)}) - \mathbb{E}\widehat{Err}_K(\widehat{f}_{PA}^{(1,\ldots,k)}),$$

$$\xi_{exp}^{(1,\ldots,k)} := \mathbb{E}\widehat{Err}_K(\widehat{f}_{PA}^{(1,\ldots,k)}) - Err(f_{PA}^{(1,\ldots,k)}),$$

где $k = 1, \ldots, n$.

Из асимптотической нормальности $\widehat{Err}_K(\widehat{f}_{PA}^{((1,\ldots,k))})$ следует, что $\xi_{exp}^{(1,\ldots,k)} = o(1/\sqrt{N})$ при $N \to \infty$. Тогда

$$\mathbb{P}\Big(\xi_0^{(1,\ldots,k,k+1)} + \xi_{exp}^{(1,\ldots,k,k+1)} - \xi_{exp}^{(1,\ldots,k,t)} - \xi_0^{(1,\ldots,k,t)} < c_{k+1,t}^{(1,\ldots,k)}\Big)$$

$$\geqslant 1 - \mathbb{P}\Big(|\xi_0^{(1,\ldots,k,k+1)}| > \frac{c_{k+1,t}^{(1,\ldots,k)} + o(1/\sqrt{N})}{2}\Big) - \mathbb{P}\Big(|\xi_0^{(1,\ldots,k,t)}| > \frac{c_{k+1,t}^{(1,\ldots,k)} + o(1/\sqrt{N})}{2}\Big).$$

Согласно неравенству Маркова

$$\mathbb{P}\Big(|\xi_0^{(1,\ldots,k,k+1)}| > \frac{c_{k+1,t}^{(1,\ldots,k)} + o(1/\sqrt{N})}{2}\Big) \leqslant \frac{4\mathbb{E}(\xi_0^{(1,\ldots,k,k+1)})^2}{(c_{k+1,t}^{(1,\ldots,k)} + o(1/\sqrt{N}))^2}$$

$$= \frac{4(V^{(1,\ldots,k,k+1)} + o(1))}{N(c_{k+1,t}^{(1,\ldots,k)} + o(1/\sqrt{N}))^2}.$$

Здесь $V^{(1,\ldots,k,k+1)}$ – предел $N\mathbb{E}(\xi_0^{(1,\ldots,k,k+1)})^2$ при $N \to \infty$. Этот предел существует в силу теоремы 7.

Введем обозначение $V_{max} := \max_{\alpha} V^{(\alpha)}$, где максимум находится по всем поднаборам α из не более чем r значений из множества $\{1,\ldots,n\}$. Таким образом, мы показали, что

$$\mathbb{P}(s_{k+1} = k+1 | s_1 = 1,\ldots,s_k = k) \geqslant 1 - \sum_{t=k+2}^{n} \frac{8V_{max} + o(1)}{N\big(c_{k+1,t}^{(1,\ldots,k)} + o(1/\sqrt{N})\big)^2}.$$

□

Из доказанной теоремы легко выводится следующий результат.

Следствие 5. *Пусть выполнены условия теоремы 14. Пусть C_1,C_2 – некоторые константы. Тогда для достаточно больших N справедливы неравенства*

$$\Big(1 - \frac{C_1C_2}{N}\Big)^r \leqslant \mathbb{P}(s_1 = 1,\ldots,s_r = r) \leqslant \Big(1 - \frac{C_1}{N}\Big)^r \quad (3.8)$$

Для $v \in \{0,1\}$ и $y \in \{-1,1\}$ введем величины

$$\theta_{v,y}^{(i)} := \mathbb{P}(X_i = v | Y = y), \quad (3.9)$$

где $i = 1,\ldots,n$.

Теорема 15. *В (3.6) константы $c_{k+1,t}^{(1,\ldots,k)}$ могут быть выражены через коэффициенты логистической регрессии:*

$$c_{k+1,t}^{(1,\ldots,k)} = c_t^{(1,\ldots,k)} - c_{k+1}^{(1,\ldots,k)},$$

и для $l = (k+1),\ldots,n$

$$c_l^{(1,\ldots,k)} = \sum_{v \in \{0,1\}^{(k+1)}} \theta_{v_1,-1}^{(1)} \cdot \ldots \cdot \theta_{v_k,-1}^{(k)} \cdot \theta_{v_{k+1},-1}^{(l)}$$

$$\times \mathbb{I}\{(-1)^{v_1}\beta_1 + \ldots + (-1)^{v_k}\beta_k + (-1)^{v_{k+1}}\beta_l < 0\}$$

$$+ \sum_{v \in \{0,1\}^{(k+1)}} \theta_{v_1,1}^{(1)} \cdot \ldots \cdot \theta_{v_k,1}^{(k)} \cdot \theta_{v_{k+1},1}^{(t)}$$

$$\times \mathbb{I}\{(-1)^{v_1}\beta_1 + \ldots + (-1)^{v_k}\beta_k + (-1)^{v_{k+1}}\beta_t \geqslant 0\}.$$

β_i для $i = 1,\ldots,r$ заданы в (3.5) и $\beta_i = 0$ для $i = r+1,\ldots,n$.

Доказательство. Чтобы понять механизм, сперва рассмотрим простой случай для $k = 0$:

$$c_{1,r+1} = Err(f_{PA}^{(r+1)}) - Err(f_{PA}^{(1)}) =$$

$$= 2 \sum_{y \in \{-1,1\}} \psi(y) \big[\mathbb{P}(Y = y, f_{PA}^{(r+1)}(X) \neq y) - \mathbb{P}(Y = y, f_{PA}^{(1)}(X) \neq y) \big].$$

Воспользуемся определением штрафной функции ψ:

$$Err(f_{PA}^{(r+1)}) = 2 \sum_{y \in \{-1,1\}} \psi(y) \big[\mathbb{P}(Y = y, f_{PA}^{(r+1)}(X) \neq y) \big]$$

$$= 2 \sum_{y \in \{-1,1\}} \frac{1}{\mathbb{P}(Y = y)} \big[\mathbb{P}(Y = y, f_{PA}^{(r+1)}(X) \neq y) \big]$$

$$= 2 \Big[\frac{\mathbb{P}(Y = -1, \mathbb{P}(\widetilde{Y} = 1 | \widetilde{X}_{r+1} = X_{r+1}) > \gamma(\psi))}{\mathbb{P}(Y = -1)}$$

$$+ \frac{\mathbb{P}(Y = 1, \mathbb{P}(\widetilde{Y} = 1 | \widetilde{X}_{r+1} = X_{r+1}) \leqslant \gamma(\psi))}{\mathbb{P}(Y = 1)} \Big]$$

$$= 2 \Big[\frac{\mathbb{P}(Y = -1, \mathbb{P}(\widetilde{Y} = 1) > \gamma(\psi))}{\mathbb{P}(Y = -1)} + \frac{\mathbb{P}(Y = 1, \mathbb{P}(\widetilde{Y} = 1) \leqslant \gamma(\psi))}{\mathbb{P}(Y = 1)} \Big]$$

$$= 2 \Big[\frac{0}{\mathbb{P}(Y = -1)} + \frac{\mathbb{P}(Y = 1)}{\mathbb{P}(Y = 1)} \Big] = 2.$$

Здесь случайный вектор $(\widetilde{X}, \widetilde{Y})$ независим от случайного вектора (X,Y) и совпадает с ним по распределению. Для X_1 из набора значимых факторов имеем

$$Err(f_{PA}^{(1)}) = 2 \sum_{y \in \{-1,1\}} \psi(y) \big[\mathbb{P}(Y = y, f_{PA}^{(1)}(X) \neq y) \big]$$

$$= 2 \Big[\frac{\mathbb{P}(Y = -1, \mathbb{P}(\widetilde{Y} = 1 | \widetilde{X}_1 = X_1) > \gamma(\psi))}{\mathbb{P}(Y = -1)}$$

$$+ \frac{\mathbb{P}(Y = 1, \mathbb{P}(\widetilde{Y} = 1 | \widetilde{X}_1 = X_1) \leqslant \gamma(\psi))}{\mathbb{P}(Y = 1)} \Big].$$

Применяем формулу полной вероятности, получаем

$$\mathbb{P}(Y = -1, \mathbb{P}(\widetilde{Y} = 1 | \widetilde{X}_1 = X_1) > \gamma(\psi))$$

$$= \mathbb{P}(Y = -1, \mathbb{P}(\widetilde{Y} = 1 | \widetilde{X}_1 = 1) > \gamma(\psi), X_1 = 1)$$

$$+ \mathbb{P}(Y = -1, \mathbb{P}(\widetilde{Y} = 1 | \widetilde{X}_1 = 0) > \gamma(\psi), X_1 = 0).$$

Теперь перепишем условие

$$\mathbb{P}(\widetilde{Y}=1|\widetilde{X}_1=X_1)>\gamma(\psi). \tag{3.10}$$

В силу того, что $\gamma(\psi)/\mathbb{P}(\widetilde{Y}=1)=1$, выражение (3.10) эквивалентно следующему:

$$\mathbb{P}(\widetilde{X}_1=1|\widetilde{Y}=1)>\mathbb{P}(\widetilde{X}_1=1)$$

$$=\mathbb{P}(\widetilde{X}_1=1|\widetilde{Y}=-1)\mathbb{P}(\widetilde{Y}=-1)+\mathbb{P}(\widetilde{X}_1=1|\widetilde{Y}=1)\mathbb{P}(\widetilde{Y}=1).$$

Продолжаем набор эквивалентных неравенств

$$\mathbb{P}(\widetilde{X}_1=1|\widetilde{Y}=1)(1-\mathbb{P}(\widetilde{Y}=1))>\mathbb{P}(\widetilde{X}_1=1|\widetilde{Y}=-1)\mathbb{P}(\widetilde{Y}=-1),$$

$$\mathbb{P}(\widetilde{X}_1=1|\widetilde{Y}=1)>\mathbb{P}(\widetilde{X}_1=1|\widetilde{Y}=-1),$$

$$\theta^{(1)}_{1,1}>\theta^{(1)}_{1,-1},$$

$$\frac{\theta^{(1)}_{1,1}}{1-\theta^{(1)}_{1,1}}>\frac{\theta^{(1)}_{1,-1}}{1-\theta^{(1)}_{1,-1}},$$

$$\frac{\theta^{(1)}_{1,1}(1-\theta^{(1)}_{1,-1})}{\theta^{(1)}_{1,-1}(1-\theta^{(1)}_{1,1})}>1,$$

$$\log\Big(\frac{\theta^{(1)}_{1,1}(1-\theta^{(1)}_{1,-1})}{\theta^{(1)}_{1,-1}(1-\theta^{(1)}_{1,1})}\Big)>0,$$

$$\beta_1>0.$$

Таким образом,

$$\mathbb{P}(Y=-1,\mathbb{P}(\widetilde{Y}=1|\widetilde{X}_1=1)>\gamma(\psi),X_1=1)=\mathbb{P}(Y=-1,\beta_1>0,X_1=1)$$

$$=\mathbb{P}(Y=-1,X_1=1).$$

Аналогично можно показать, что

$$\mathbb{P}(Y=1,\mathbb{P}(\widetilde{Y}=1|\widetilde{X}_1=0)\leqslant\gamma(\psi),X_1=0)=\mathbb{P}(Y=1,\beta_1\geqslant 0,X_1=0)$$

$$=\mathbb{P}(Y=1,X_1=-1).$$

В итоге мы получаем, что

$$Err(f^{(1)}_{PA})=2\Big[\mathbb{P}(X_1=1|Y=-1)+\mathbb{P}(X_1=0|Y=1)\Big].$$

Отсюда следует, что

$$c_{1,r+1} = 2\Big[1 - \mathbb{P}(X_1 = 1|Y = -1) - \mathbb{P}(X_1 = 0|Y = 1)\Big]$$

$$= 2\Big[\mathbb{P}(X_1 = 0|Y = -1) - \mathbb{P}(X_1 = 0|Y = 1)\Big]$$

$$= 2\Big[\theta^{(1)}_{0,-1} - \theta^{(1)}_{0,1}\Big] > 0.$$

Ясно, что чем больше β_1, тем больше будет разность $\theta^{(1)}_{0,-1} - \theta^{(1)}_{0,1}$. Повторяя рассуждения, получаем выражение

$$c_{1,k} = 2\Big[\big(\theta^{(k)}_{1,-1} + \theta^{(k)}_{0,1}\big) - \big(\theta^{(1)}_{1,-1} + \theta^{(1)}_{1,-1}\big)\Big]$$

для $k = 2,\ldots,r$.

Теперь рассмотрим общий случай, когда на первых k шагах алгоритма уже были выбраны факторы $X_1,\ldots,X_k$. По определению $c^{(1,\ldots,k)}_{k+1,t} := Err(f^{(1,\ldots,k,k+1)}_{PA}) - Err(f^{(1,\ldots,k,t)}_{PA})$. Пусть α обозначает вектор индексов $(1,\ldots,k,t)$, $k < t \leqslant n$. Как и ранее,

$$Err(f^{(\alpha)}_{PA}) = 2\sum_{y\in\{-1,1\}} \frac{1}{\mathbb{P}(Y=y)}\mathbb{P}\Big(Y = y, f^{(\alpha)}_{PA}(X) \neq y\Big)$$

$$= 2\Big[\mathbb{P}\Big(\mathbb{P}(\widetilde{Y} = 1|\widetilde{X}^{(\alpha)} = X^{(\alpha)}) > \gamma(\psi)\Big|Y = -1\Big)$$
$$+\mathbb{P}\Big(\mathbb{P}(\widetilde{Y} = 1|\widetilde{X}^{(\alpha)} = X^{(\alpha)}) \leqslant \gamma(\psi)\Big|Y = 1\Big)\Big].$$

Рассмотрим подробнее первое слагаемое:

$$\mathbb{P}\Big(\mathbb{P}(\widetilde{Y} = 1|\widetilde{X}^{(\alpha)} = X^{(\alpha)}) > \gamma(\psi)\Big|Y = -1\Big)$$

$$= \sum_{v\in\{0,1\}^{(k+1)}} \mathbb{P}(X^{(\alpha)} = v|Y = -1)\mathbb{I}\{\mathbb{P}(Y = 1|X^{(\alpha)} = v) > \gamma(\psi)\}$$

$$= \sum_{v\in\{0,1\}^{(k+1)}} \mathbb{P}(X_1 = v_1,\ldots,X_k = v_k, X_t = v_{k+1}|Y = -1)$$
$$\times\mathbb{I}\{(-1)^{v_1}\beta_1 + \ldots + (-1)^{v_k}\beta_k + (-1)^{v_{k+1}}\beta_t < 0\}$$
$$= \sum_{v\in\{0,1\}^{(k+1)}} \theta^{(1)}_{v_1,-1}\cdot\ldots\cdot\theta^{(k)}_{v_k,-1}\cdot\theta^{(t)}_{v_{k+1},-1}$$
$$\times\mathbb{I}\{(-1)^{v_1}\beta_1 + \ldots + (-1)^{v_k}\beta_k + (-1)^{v_{k+1}}\beta_t < 0\}.$$

Аналогично устанавливается, что

$$\mathbb{P}\Big(\mathbb{P}(\widetilde{Y}=1|\widetilde{X}^{(\alpha)}=X^{(\alpha)})\leqslant\gamma(\psi)\Big|Y=-1\Big)$$

$$=\sum_{v\in\{0,1\}^{(k+1)}}\theta^{(1)}_{v_1,1}\cdot\ldots\cdot\theta^{(k)}_{v_k,1}\cdot\theta^{(t)}_{v_{k+1},1}$$

$$\times\mathbb{I}\{(-1)^{v_1}\beta_1+\ldots+(-1)^{v_k}\beta_k+(-1)^{v_{k+1}}\beta_t\geqslant 0\}.$$

Таким образом, найдено требуемое выражение для констант $c^{(1,\ldots,k)}_{k+1,t}$ и теорема 15 доказана. □

3.3 Реализация MDR-EFE алгоритма в виде программного кода и применение к данным компьютерного моделирования

В данном разделе приводится реализация MDR-EFE метода в виде программного кода. Чтобы продемонстрировать возможность применения MDR-EFE метода в практических приложениях, мы сравним работу MDR-EFE метода с классической логистической регрессией и MDR методом, реализованном в [52]. Работа алгоритмов сравнивается на данных компьютерного моделирования в рамках различных моделей эпистаза.

3.3.1 Программный код MDR-EFE алгоритма

Ниже приводится реализация функции вычисления значения функционала ошибки на языке программирования R [59].

```
# Err_k вычисляет функционал ошибки для k-го фолда
# dt -- таблица с данными
# k -- номер фолда в кросс-валидации
# alpha -- индексы тестируемых факторов
Err_k <- function(dt = dt, k = k, alpha = alpha){
  # разделим датасет на обучающий и тестирующий
  dt[-flds[[k]], ] %>% select(c(alpha, "Class")) -> dt_train
  dt[flds[[k]], ] %>% select(c(alpha, "Class")) -> dt_test
```

```
  # оценим P(Y=1) и значения функции psi
  p1 <- sum(dt_train$Class == 1) / nrow(dt_train)
  psi1 <- 1/(sum(dt_test$Class) / nrow(dt_test))
  psi0 <- 1/(sum(dt_test$Class == 0) / nrow(dt_test))

  # для каждого уникального набора значений факторов X^(alpha)
  # вычислим P(Y=1|X^(alpha)=x^(alpha))
  dt_train %>%
    group_by_at(alpha) %>%
    summarise(p1g=sum(Class)/n(),.groups="drop") -> dt_train

  # применим предсказательный алгоритм для тестовой подвыборки
  suppressMessages(
  dt_test %>%
    group_by_all() %>%
    summarise(n_test = n(), .groups = "drop") %>%
    left_join(dt_train) %>%
    mutate(ClassPredicted = ifelse(p1g > p1,1,0)) %>%
    mutate(err = ifelse(Class == ClassPredicted, 0,
                 ifelse(Class == 1, psi1, psi0))) -> dt_test)

  # вычисление и вывод значеия функционала ошибки
  return(sum(dt_test$err * dt_test$n_test, na.rm = T) /
           sum(dt_test$n_test, na.rm = T))
}
```

Ниже приводится программный код на языке R, который использовался для того, чтобы найти значимый набор факторов с помощью EFE-MDR метода, примененного к данным компьютерного моделирования.

```
# загружаем необходимые пакеты и вспомогательные файлы
library(data.table)
library(dplyr)
library(caret)
library(parallel)
source("functions.R")

# фиксируем seed для воспроизводимости экспериментов
set.seed(1234)

# название модели, данные которой используем
modelname <- "Model1"
model <- paste0("./data/", modelname, "/")
files <- list.files(model)
```

```
file.create(paste0(modelname, "_MDR-EFE.txt"))

# цикл по эксперимента
for(file in files){
  # чтение данных
  dt <- fread(paste0(model, file),
              data.table = FALSE)
  factorNames <- names(dt)[-ncol(dt)]

  # создаем индексы разбиения для кросс-валидации
  flds <- createFolds(dt$Class, k = 10, list = TRUE,
                      returnTrain = FALSE)

  # генерируем все комбинации из m факторов
  factorCombinations <- combn(factorNames, m = 3, simplify =
   FALSE)

  # засекаем время на итерацию
  t <- Sys.time()

  # вычисляем значение функционала ошибки
  mclapply(factorCombinations, function(alpha){
    lapply(1:10, function(k) Err_k(dt, k, alpha)) %>% unlist()
   %>% mean()
  }, mc.cores = 30) -> err

  # выводим время итерации
  print(Sys.time() - t)
  factorCombinations %>% do.call(rbind, .) %>% as.data.frame()
   -> res
  res$err <- unlist(err)

  # сохраням лучшую модель в файл
  write(x = paste(as.character(unlist(res[which.min(res$err),
   1:3])),
                  collapse = " "),
        file = paste0(modelname, "_MDR-EFE.txt"),
        append = T)
}
```

3.3.2 Генерация данных в модели эпистаза

Как уже отмечалось ранее, одно из преимуществ MDR-EFE метода заключается в возможности находить наборы значимых факторов в случае, когда наблюдается взаимодействие факторов в отношении случайного отклика Y. В частности, по отдельности факторы могут не влиять на Y. Однако, их специфичные комбинации значений могут быть более характерны для случайного отклика $Y = 1$. В биологии тип взаимодействия генов, при котором проявление одного гена находится под влиянием другого гена, называется *эпистазом*. Можно ожидать, что MDR-EFE алгоритм будет точнее находить наборы значимых факторов при анализе данных, удовлетворяющих модели эпистаза, чем, например, логистическая регрессия.

Чтобы продемонстрировать это, мы провели компьютерное моделирование генетических данных, удовлетворяющих модели эпистаза. В работах [82; 85] есть результаты численных экспериментов для MDR-EFE метода. Там зависимость между X и Y задавалась набором формул. Здесь же мы воспользуемся программным обеспечением GAMETES 2.0 ([67]) для генерации данных. Эта программа многократно применялась в недавних исследования для моделирования генетических данных в предположении модели эпистаза [63—65]. Мы использовали четыре разные модели для генерации данных, чтобы сравнить работу алгоритмов в различных конфигурациях. Каждую модель можно задать рядом параметров через интерфейс программы GAMETES 2.0 (см. рисунок 3.1). В таблице 1 приводятся параметры, задающие каждую из четырех моделей. Параметр наследуемости h^2 отвечает за величину вклада генетических факторов в риск заболевания. Подробно смысл этого параметра описывается в [67]. Точность работы алгоритма определялась как доля случаев, в которых алгоритмом был выявлен истинный набор значимых факторов. Мы моделировали данные по 100 раз для каждой из моделей, чтобы добиться более стабильных оценок точности алгоритмов при выборе значимых наборов факторов. Для каждого сгенерированного массива данных запускались алгоритмы, которые на входе выдали наиболее вероятную комбинацию индексов значимых факторов. В первых трех моделях от алгоритмов требовалось найти значимый набор из двух факторов, а в четвертой модели – набор из трех значимых факторов.

Таблица 1 — Параметры моделей, используемые для симуляции данных в программе GAMETES 2.0

Параметр	Мод. 1	Мод. 2	Мод. 3	Мод. 4
Число факторов	20	20	20	10
Число значимых факторов	2	2	2	3
Наследуемость (h^2)	0.01	0.01	0.1	0.1
Число больных	400	800	400	400
Число здоровых	400	800	400	400

3.3.3 Применение к данным компьютерного моделирования

На рисунке 3.2 приведены результаты работы алгоритмов на данных компьютерного моделирования. Мы видим, что во всех четырех моделях логистическая регрессия уступает по точности MDR и MDR-EFE методам. Во всех экспериментах мы использовали сбалансированную выборку, то есть количество больных совпадало с количеством здоровых. Модель 2 отличается от модели 1 только удвоенным размером выборки. Можно видеть, что увеличение размера выборки существенно повышает точность MDR и MDR-EFE методов: с 33-37% до 87%. Увеличение параметра наследуемости h^2, отвечающего за вклад значимых генетических факторов в заболевание, так же приводит к более высокой точности как MDR, так и MDR-EFE метода. В целом, в проведенных экспериментах MDR-EFE и MDR алгоритмы показали сопоставимую точность, что говорит о возможности их применения в прикладных задачах.

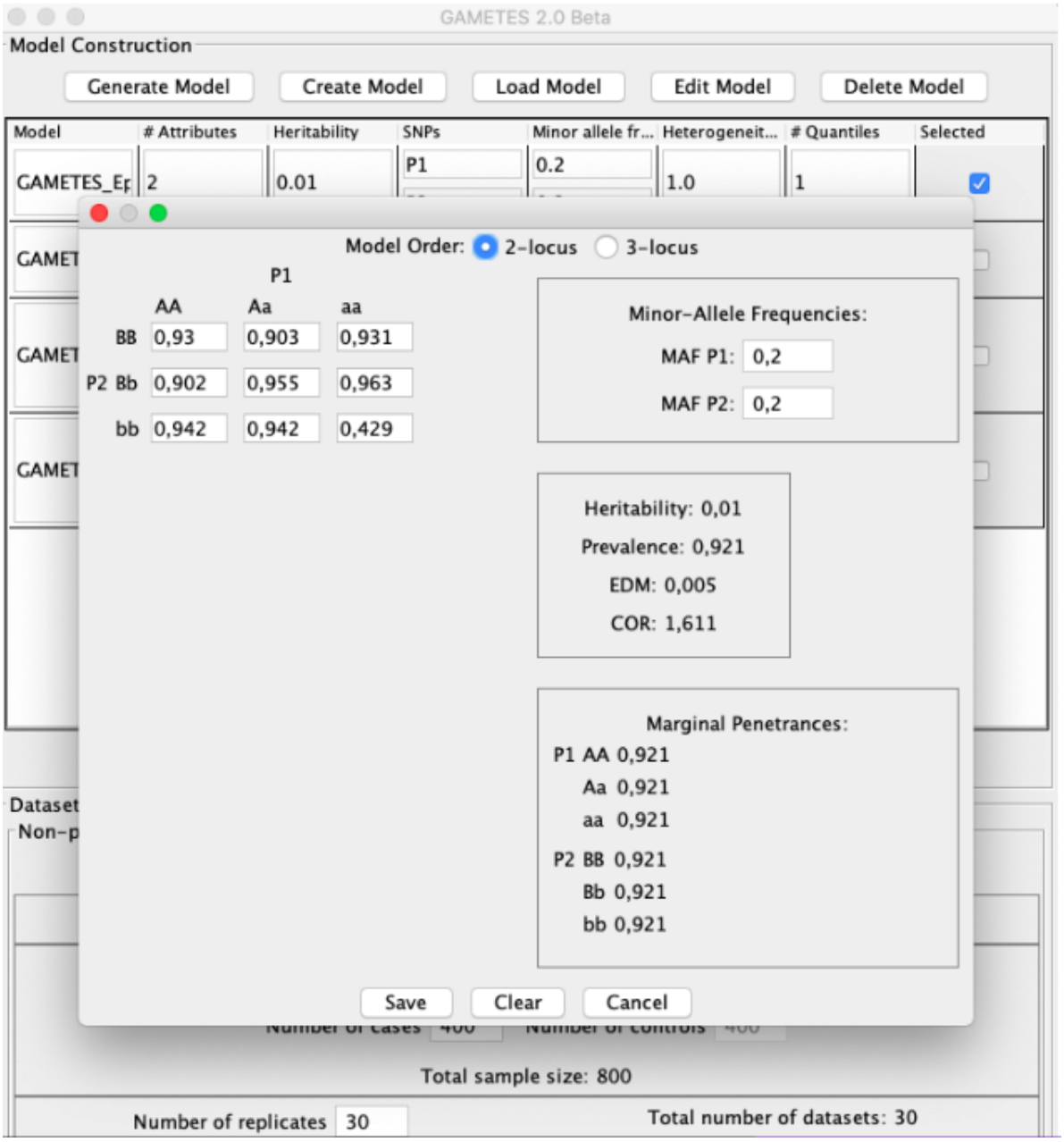

Рисунок 3.1 — Интерфейс программы GAMETES 2.0 для коммпьютерного симулирования генетических данных, удовлетворяющих модели эпистаза.

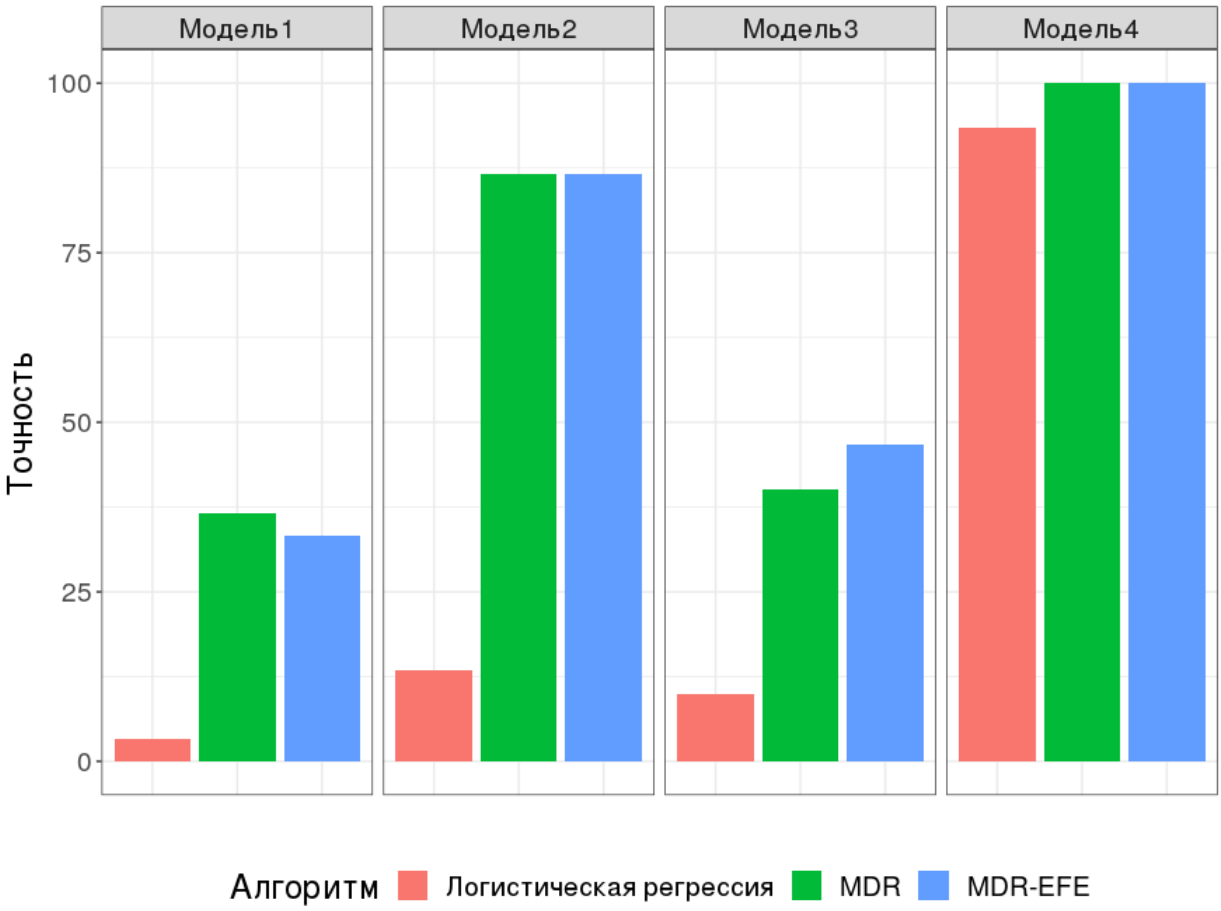

Рисунок 3.2 — Результат работы алгоритмов на данных компьютерного моделирования

Заключение

Обзор проведенного исследования. Тематика диссертации относится к области разработки вероятностно-статистических методов выявления значимых факторов, влияющих на изучаемую случайную функцию отклика. Проведено исследование асимптотических свойств MDR-EFE (multifactorial dimensionality reduction – error function estimation) метода и его модификаций. Основные результаты диссертации состоят в следующем.

1. В случае небинарной функции отклика построена оценка функционала ошибки в MDR-EFE методе по имеющейся выборке с помощью процедуры кросс-валидации. Установлен критерий сильной состоятельности построенной оценки. Доказана теорема, обосновывающая возможность использования оценок функционала ошибки для выбора наборов значимых факторов.

2. Найдены достаточные условия сильной состоятельности оценок функционала ошибки в случае объясняющих факторов, имеющих абсолютно непрерывное распределение относительно меры Лебега в пространстве $\mathbb{R}^n$. Показано, что построенные ядерные оценки условных плотностей с помощью статистик k ближайших соседей удовлетворяют условиям полученной теоремы.

3. Доказана центральная предельная теорема для регуляризованных версий оценок функционала ошибки в случае небинарной функции отклика.

4. С помощью техники перестановочных случайных величин доказан новый вариант центральной теоремы для оценок функционала ошибки в MDR-EFE методе.

5. Установлен аналог теоремы Эрдеша и Каца для перестановочных случайных величин.

6. Разработан вариант MDR-EFE метода с последовательным отбором переменных. В случае модели наивного байесовского классификатора получены оценки снизу для вероятности выбора значимого набора факторов MDR-EFE методом с последовательным отбором переменных.

7. Применение MDR-EFE метода проиллюстрировано на данных компьютерного моделирования.

Рекомендации и перспективы по дальнейшей разработке темы. Дальнейшие исследования по тематике диссертации могут проводиться в следующих направлениях:

По главе 1. Доказательство критерия сильной состоятельности оценки функционала ошибки в случае функции отклика, имеющей абсолютно непрерывное распределение относительно меры Лебега в пространстве $\mathbb{R}$.

По главе 2. Получение оценок скорости сходимости в центральной предельной теореме для оценок функционала ошибок в MDR-EFE методе.

По главе 3. Вывод оценок вероятности выбора значимого набора факторов для MDR-EFE метода с последовательным отбором переменных в различных моделях (отличных от модели наивного байесовского классификатора). Исследование случая, когда количество значимых факторов неизвестно. Применение данного метода к реальным данным и данным компьютерного моделирования.

Список литературы

1. Бобрикова Е. В., Платонова А. А., Гайдамака Ю. В., Шоргин С. Я. Пример применения аппарата нейронных сетей при назначении модуляционно-кодовой схемы планировщиком базовой станции сети 5G // Системы и средства информатики. — 2021. — Т. 31, № 3. — С. 135—143.
2. Булинский А. В., Ширяев А. Н. Теория случайных процессов. — ФИЗМАТЛИТ, 2005. — 400 с.
3. Грушо А. А., Грушо Н. А., Забежайло М. И., Смирнов Д. В. [и др.]. Поиск аномалий в больших данных // Системы и средства информатики. — 2022. — Т. 32, № 1. — С. 160—167.
4. Ребриков Д. В., Коростин Д. О., Шубина Е. С., Ильинский В. В. NGS: высокопроизводительное секвенирование. — БИНОМ. Лаборатория знаний, 2015. — 232 с.
5. Стоянов Й. Контрпримеры в теории вероятностей. — МЦНМО, 2014. — 300 с.
6. Ширяев А. Н. Вероятность-1. — МЦНМО, 2007. — 552 с.
7. Ширяев А. Н. Вероятность-2. — МЦНМО, 2007. — 416 с.
8. Abegaz F., Van Lishout F., Mahachie John J. M., Chiachoompu K., [et al.]. Performance of model-based multifactor dimensionality reduction methods for epistasis detection by controlling population structure // BioData Mining. — 2021. — Vol. 14, no. 1. — P. 1—20.
9. Aldous D. J. Exchangeability and related topics // École d'Été de Probabilités de Saint-Flour XIII—1983. — Springer, 1985. — P. 1—198.
10. Arlot S., Celisse A. A survey of cross-validation procedures for model selection // Statistics Surveys. — 2010. — Vol. 4. — P. 40—79.
11. Austin T. On exchangeable random variables and the statistics of large graphs and hypergraphs // Probability Surveys. — 2008. — Vol. 5. — P. 80—145.
12. Azuma K. Weighted sums of certain dependent random variables // Tohoku Mathematical Journal, Second Series. — 1967. — Vol. 19, no. 3. — P. 357—367.

13. Berman S. M. Limiting distribution of the maximum term in sequences of dependent random variables // The Annals of Mathematical Statistics. — 1962. — Vol. 33, no. 3. — P. 894—908.

14. Berti P., Pratelli L., Rigo P. Limit theorems for a class of identically distributed random variables // The Annals of Probability. — 2004. — Vol. 32, no. 3. — P. 2029—2052.

15. Biau G., Devroye L. Lectures on the Nearest Neighbor Method. Vol. 246. — Springer, 2015. — 290 p.

16. Billingsley P. Convergence of probability measures. — John Wiley & Sons, 2013. — 304 p.

17. Billingsley P. Probability and measure. — Wiley, 1995. — 593 p.

18. Blum J., Chernoff H., Rosenblatt M., Teicher H. Central limit theorems for interchangeable processes // Canadian Journal of Mathematics. — 1958. — Vol. 10. — P. 222—229.

19. Bulinski A., Butkovsky O., Sadovnichy V., Shashkin A., [et al.]. Statistical Methods of SNP Data Analysis and Applications // Open Journal of Statistics. — 2012. — Vol. 2, no. 1. — P. 73—87.

20. Bulinski A. Central limit theorem related to MDR-method // Asymptotic Laws and Methods in Stochastics. — Springer, 2015. — P. 113—128.

21. Bulinski A., Kozhevin A. New version of the MDR method for stratified samples // Statistics, Optimization & Information Computing. — 2017. — Vol. 5, no. 1. — P. 1—18.

22. Bulinski A., Kozhevin A. Statistical estimation of conditional Shannon entropy // ESAIM: Probability and Statistics. — 2019. — Vol. 23. — P. 350—386.

23. Bulinski A., Kozhevin A. Statistical estimation of mutual information for mixed model // Methodology and Computing in Applied Probability. — 2021. — Vol. 23. — P. 123—142.

24. Bulinski A., Slepov N. Sharp Estimates for Proximity of Geometric and Related Sums Distributions to Limit Laws // Mathematics. — 2022. — Vol. 10, no. 24. — P. 4747.

25. Bulinski A. On foundation of the dimensionality reduction method for explanatory variables // Journal of Mathematical Sciences. — 2014. — Vol. 199, no. 2. — P. 113—122.

26. Chatterjee S. A generalization of the Lindeberg principle // The Annals of Probability. — 2006. — Vol. 34, no. 6. — P. 2061—2076.

27. Chernoff H., Teicher H. A central limit theorem for sums of interchangeable random variables // The Annals of Mathematical Statistics. — 1958. — P. 118—130.

28. Climente-González H., Azencott C.-A., Kaski S., Yamada M. Block HSIC Lasso: model-free biomarker detection for ultra-high dimensional data // Bioinformatics. — 2019. — Vol. 35, no. 14. — P. i427—i435.

29. Coen A., Mena R. H. Ruin probabilities for Bayesian exchangeable claims processes // Journal of Statistical Planning and Inference. — 2015. — Vol. 166. — P. 102—115.

30. Damien P., Dellaportas P., Polson N. G., Stephens D. A. Bayesian Theory and Applications. — OUP Oxford, 2013. — 702 p.

31. Das B., Wang T., Dai G. Asymptotic Behavior of Common Connections in Sparse Random Networks // Methodology and Computing in Applied Probability. — 2022. — P. 1—22.

32. De Finetti B. Funzione caratteristica di un fenomeno aleatorio // Atti del Congresso Internazionale dei Matematici: Bologna del 3 al 10 de settembre di 1928. — 1929. — P. 179—190.

33. Devroye L., Penrod C. S. The strong uniform convergence of multivariate variable kernel estimates // Canadian Journal of Statistics. — 1986. — Vol. 14, no. 3. — P. 211—220.

34. Erdős P., Kac M. On certain limit theorems of the theory of probability // Bulletin of the American Mathematical Society. — 1946. — Vol. 52. — P. 292—302.

35. Fortini S., Ladelli L., Regazzini E. Central limit theorem with exchangeable summands and mixtures of stable laws as limits // arXiv preprint arXiv:1204.4357. — 2012.

36. Fujikoshi Y., Ulyanov V. V. Non-asymptotic Analysis of Approximations for Multivariate Statistics. — Springer Singapore, 2020.

37. Gola D., Mahachie John J. M., Steen K. van, König I. R. A roadmap to multifactor dimensionality reduction methods // Briefings in Bioinformatics. — 2016. — Vol. 17, no. 2. — P. 293—308.

38. Götze F., Naumov A., Ulyanov V. Asymptotic analysis of symmetric functions // Journal of Theoretical Probability. — 2017. — Vol. 30. — P. 876—897.

39. Guo H., Yu Z., An J., Han G., [et al.]. A two-stage mutual information based Bayesian Lasso algorithm for multi-locus genome-wide association studies // Entropy. — 2020. — Vol. 22, no. 3. — P. 329.

40. Hamacher K., Kussel T., Schneider T., Tkachenko O. PEA: Practical Private Epistasis Analysis Using MPC // Computer Security–ESORICS 2022: 27th European Symposium on Research in Computer Security, Copenhagen, Denmark, September 26–30, 2022, Proceedings, Part III. — Springer. 2022. — P. 320—339.

41. Hoeffding W. Probability inequalities for sums of bounded random variables // Journal of the American Statistical Association. — 1963. — Vol. 58, no. 301. — P. 13—30.

42. Hsu P.-L., Robbins H. Complete convergence and the law of large numbers // Proceedings of the National Academy of Sciences. — 1947. — Vol. 33, no. 2. — P. 25—31.

43. Hu T.-C., Moricz F., Taylor R. Strong laws of large numbers for arrays of rowwise independent random variables // Acta Mathematica Hungarica. — 1989. — Vol. 54, no. 1/2. — P. 153—162.

44. Huang H., Gao Y., Zhang H., Li B. Weighted Lasso estimates for sparse logistic regression: Non-asymptotic properties with measurement errors // Acta Mathematica Scientia. — 2021. — Vol. 41, no. 1. — P. 207—230.

45. Isaev M., Rodionov I. V., Zhang R.-R., Zhukovskii M. E. Extreme value theory for triangular arrays of dependent random variables // Russian Mathematical Surveys. — 2020. — Oct. — Vol. 75, no. 5. — P. 968—970.

46. Jia W., Sun M., Lian J., Hou S. Feature dimensionality reduction: a review // Complex & Intelligent Systems. — 2022. — Vol. 8, no. 3. — P. 2663—2693.

47. Khandezamin Z., Naderan M., Rashti M. J. Detection and classification of breast cancer using logistic regression feature selection and GMDH classifier // Journal of Biomedical Informatics. — 2020. — Vol. 111. — P. 103591.

48. Kingman J. F. C. Mathematics of Genetic Diversity. — SIAM, 1980.

49. Kozhevin A. A. Feature selection based on statistical estimation of mutual information // Siberian Electronic Mathematical Reports. — 2021. — Vol. 18, no. 1. — P. 720—728.

50. Lee Taylor R., Hu T.-C. On laws of large numbers for exchangeable random variables // Stochastic Analysis and Applications. — 1987. — Vol. 5, no. 3. — P. 323—334.

51. Mills M. C., Barban N., Tropf F. C. An introduction to statistical genetic data analysis. — MIT Press, 2020.

52. Moor J. Scikit-MDR. — 2013. — https://github.com/EpistasisLab/scikit-mdr.

53. Murphy K. P. Probabilistic machine learning: an introduction. — MIT press, 2022.

54. Naderi H., Jafari M., Matuła P., Mohammadi M. On the Jajte weak law of large numbers for exchangeable random variables // Communications in Statistics-Theory and Methods. — 2022. — P. 1—9.

55. Nassif A. B., Shahin I., Attili I., Azzeh M., Shaalan K. Speech recognition using deep neural networks: A systematic review // IEEE access. — 2019. — Vol. 7. — P. 19143—19165.

56. Park M., Jeong H.-B., Lee J.-H., Park T. Spatial rank-based multifactor dimensionality reduction to detect gene–gene interactions for multivariate phenotypes // BMC bioinformatics. — 2021. — Vol. 22, no. 1. — P. 1—21.

57. Ponte-Fernandez C., Gonzalez-Dominguez J., Martin M. J. Fiuncho: a program for any-order epistasis detection in CPU clusters // The Journal of Supercomputing. — 2022. — Vol. 78, no. 13. — P. 15338—15357.

58. Pudjihartono N., Fadason T., Kempa-Liehr A. W., O'Sullivan J. M. A review of feature selection methods for machine learning-based disease risk prediction // Frontiers in Bioinformatics. — 2022. — Vol. 2. — P. 927312.

59. R Core Team. R: A Language and Environment for Statistical Computing / R Foundation for Statistical Computing. — Vienna, Austria, 2019. — URL: https://www.R-project.org/.

60. Ritchie M. D., Hahn L. W., Roodi N., Bailey L. R., [et al.]. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer // The American Journal of Human Genetics. — 2001. — Vol. 69, no. 1. — P. 138—147.

61. Roberts D. A., Yaida S., Hanin B. The principles of deep learning theory. — Cambridge University Press Cambridge, MA, USA, 2022.

62. Röllin A. Stein's method in high dimensions with applications // Annales de l'IHP Probabilités et Statistiques. Vol. 49. — 2013. — P. 529—549.

63. Russ D., Williams J. A., Cardoso V. R., Bravo-Merodio L., [et al.]. Evaluating the detection ability of a range of epistasis detection methods on simulated data for pure and impure epistatic models // Plos One. — 2022. — Vol. 17, no. 2. — e0263390.

64. Shang J., Cai X., Zhang T., Sun Y., [et al.]. EpiReSIM: A Resampling Method of Epistatic Model without Marginal Effects Using Under-Determined System of Equations // Genes. — 2022. — Vol. 13, no. 12. — P. 2286.

65. Sun Y., Gu Y., Ren Q., Li Y., [et al.]. MDSN: A Module Detection Method for Identifying High-Order Epistatic Interactions // Genes. — 2022. — Vol. 13, no. 12. — P. 2403.

66. Tibshirani R. Regression shrinkage and selection via the lasso // Journal of the Royal Statistical Society. Series B (Methodological). — 1996. — P. 267—288.

67. Urbanowicz R. J., Kiralis J., Sinnott-Armstrong N. A., Heberling T., [et al.]. GAMETES: a fast, direct algorithm for generating pure, strict, epistatic models with random architectures // BioData Mining. — 2012. — Vol. 5, no. 1. — P. 1—14.

68. Velez D. R., White B. C., Motsinger A. A., Bush W. S., [et al.]. A balanced accuracy function for epistasis modeling in imbalanced datasets using multifactor dimensionality reduction // Genetic Epidemiology. — 2007. — Vol. 31, no. 4. — P. 306—315.

69. Venter J. C., Adams M. D., Myers E. W., Li P. W., [et al.]. The sequence of the human genome // Science. — 2001. — Vol. 291, no. 5507. — P. 1304—1351.

70. Vergara J. R., Estévez P. A. A review of feature selection methods based on mutual information // Neural Computing and Applications. — 2014. — Vol. 24, no. 1. — P. 175—186.

71. Weber N. A martingale approach to central limit theorems for exchangeable random variables // Journal of Applied Probability. — 1980. — Vol. 17, no. 3. — P. 662—673.

72. Wieczorek J., Lei J. Model selection properties of forward selection and sequential cross-validation for high-dimensional regression // Canadian Journal of Statistics. — 2022. — Vol. 50, no. 2. — P. 454—470.

73. Yang C.-H., Hou M.-F., Chuang L.-Y., Yang C.-S., Lin Y.-D. Dimensionality reduction approach for many-objective epistasis analysis // Briefings in Bioinformatics. — 2023. — Vol. 24, no. 1. — bbac512.

74. Yang C.-H., Huang H.-C., Hou M.-F., Chuang L.-Y., Lin Y.-D. Fuzzy-based multiobjective multifactor dimensionality reduction for epistasis analysis // IEEE/ACM Transactions on Computational Biology and Bioinformatics. — 2022.

75. Yu C., Zelterman D. Sums of exchangeable Bernoulli random variables for family and litter frequency data // Computational Statistics & Data Analysis. — 2008. — Vol. 52, no. 3. — P. 1636—1649.

76. Yuan D.-M., Li S.-J. Extensions of several classical results for independent and identically distributed random variables to conditional cases // Journal of the Korean Mathematical Society. — 2015. — Vol. 52, no. 2. — P. 431—445.

77. Zhang C., Qin Q., Li Y., Zheng X., [et al.]. Multifactor dimensionality reduction reveals the effect of interaction between ERAP1 and IFIH1 polymorphisms in psoriasis susceptibility genes // Frontiers in Genetics. — 2022. — Vol. 13.

78. Zhang H. A Review of Convolutional Neural Network Development in Computer Vision // EAI Endorsed Transactions on Internet of Things. — 2022. — Vol. 7, no. 28.

79. Zhao Y., Zhu H., Lu Z., Knickmeyer R. C., Zou F. Structured genome-wide association studies with Bayesian hierarchical variable selection // Genetics. — 2019. — Vol. 212, no. 2. — P. 397—415.

Публикации автора по теме диссертации

Статьи в научных журналах Web of Science, SCOPUS, RSCI

80. Bulinski A. V., Rakitko A. S. Estimation of nonbinary random response // Doklady Mathematics. — 2014. — Vol. 89, no. 2. — P. 225—229.

81. Bulinski A. V., Rakitko A. S. MDR method for nonbinary response variable // Journal of Multivariate Analysis. — 2015. — Vol. 135. — P. 25—42.

82. Bulinski A. V., Rakitko A. S. Simulation and analytical approach to the identification of significant factors // Communications in Statistics-Simulation and Computation. — 2016. — Vol. 45, no. 5. — P. 1430—1450.

83. Ruiz P. A., Rakitko A. S. The limit theorem for maximum of partial sums of exchangeable random variables // Statistics & Probability Letters. — 2016. — Vol. 119. — P. 357—362.

Статьи в трудах научных конференций

84. Rakitko A. S. MDR-EFE method with forward selection // The 5th International Conference on Stochastic Methods (ICSM-5). — 2020. — P. 163—167.

85. Rakitko A. S. Multifactorial Dimensionality Reduction for Disordered Trait // Proceedings of the International Joint Conference on Biomedical Engineering Systems and Technologies. Vol. 3. — 2015. — P. 232—236.

Тезисы докладов в материалах научных конференций

86. Ракитько А. С. Последовательный отбор переменных в MDR-EFE методе // Сборник тезисов XXIV Международной научной конференции студентов, аспирантов и молодых учёных «Ломоносов–2017». — Макс Пресс Москва, 2017. — С. 1—2.

87. Ракитько А. С. Центральные предельные теоремы для массивов перестановочных случайных величин // Сборник тезисов XXI Международной конференции студентов, аспирантов и молодых учёных «Ломоносов-2014». — 2014. — С. 1—2.

88. Rakitko A. S. Multifactor dimensionality reduction method and simulation techniques // Abstracts of XXXII International Seminar on Stability Problems for Stochastic Models and VIII International Workshop «Applied Problems in Theory of Probabilities and Mathematical Statistics related to modeling of information systems». — Institute of Informatics Problems, Russian Academy of Sciences, 2014. — P. 94—95.

89. Rakitko A. S. On the application of MDR-EFE method for relevant feature selection // Abstract of communications for international conference «Limit Theorems of Probability Theory and Mathematical Statistics». — 2022. — P. 98.

Другие публикации автора, относящиеся к приложениям статистических методов к анализу генетических данных

90. Berseneva A., Kovalenko E., Vergasova E., Prohorov A., [et al.]. Association of common genetic variants with body mass index in Russian population // European Journal of Clinical Nutrition. — 2023.

91. Boev A., Rakitko A., Usmanov S., Kobzeva A., [et al.]. Genome assembly using quantum and quantum-inspired annealing // Scientific Reports. — 2021. — Vol. 11, no. 1. — P. 13183.

92. Borisevich D., Schnurr T. M., Engelbrechtsen L., Rakitko A., [et al.]. Nonlinear interaction between physical activity and polygenic risk score of body mass index in Danish and Russian populations // Plos One. — 2021. — Vol. 16, no. 10. — e0258748.

93. COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19 // Nature. — 2021. — Vol. 600, no. 7889. — P. 472—477.

94. Kasyanov E., Rakitko A., Rukavishnikov G., Golimbet V., [et al.]. Contemporary Genome-Wide Association Studies in Depression: The Critical Role of Phenotyping // Neuroscience and Behavioral Physiology. — 2022. — Vol. 52, no. 6. — P. 826—835.

95. Kibitov A., Rakitko A., Kasyanov E., Yermakovich D., [et al.]. Genome-wide association study of depression symptoms using online self-questionnaires in the Russian population cohort: preliminary results // European Psychiatry. — 2022. — Vol. 65, S1. — S327—S327.

96. Pinakhina D., Yermakovich D., Vergasova E., Kasyanov E., [et al.]. GWAS of depression in 4,520 individuals from the Russian population highlights the role of MAGI2 (S-SCAM) in the gut-brain axis // Frontiers in Genetics. — 2023. — Vol. 13. — P. 3571.

97. Tsyrulnikov M., Rakitko A. A hierarchical Bayes ensemble Kalman filter // Physica D: Nonlinear Phenomena. — 2017. — Vol. 338. — P. 1—16.

98. Tsyrulnikov M., Rakitko A. Impact of non-stationarity on hybrid ensemble filters: A study with a doubly stochastic advection-diffusion-decay model // Quarterly Journal of the Royal Meteorological Society. — 2019. — Vol. 145, no. 722. — P. 2255—2271.

99. Verbenko D. A., Karamova A. E., Artamonova O. G., Deryabin D. G., [et al.]. Apremilast pharmacogenomics in Russian patients with moderate-to-severe and severe psoriasis // Journal of Personalized Medicine. — 2020. — Vol. 11, no. 1. — P. 20.

100. Weiner 3rd J., Suwalski P., Holtgrewe M., Rakitko A., [et al.]. Increased risk of severe clinical course of COVID-19 in carriers of HLA-C*04:01 // EClinicalMedicine. — 2021. — Vol. 40. — P. 101099.