МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ М.В. ЛОМОНОСОВА
МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ

На правах рукописи

Линке Юлиана Юрьевна

Универсальные ядерные оценки в непараметрической регрессии с приложениями к нелинейным регрессионным моделям

1.1.4 — теория вероятностей и математическая статистика

ДИССЕРТАЦИЯ

на соискание ученой степени

доктора физико-математических наук

Научный консультант:

д. ф.-м. н., профессор

Е. Б. Яровая

Москва — 2024

Содержание

ВВЕДЕНИЕ

Обозначения и соглашения

• Всюду в работе пределы, если не оговорено противное, берутся при $n \to \infty$.

• Запись $\zeta_n = O_p(\eta_n)$ означает ограниченность по вероятности последовательности случайных величин ζ_n/η_n, т.е. для некоторой величины $\eta_n > 0$ (возможно, случайной) и всех чисел $M > 0$ выполнено $\limsup \mathbb{P}(|\zeta_n|/\eta_n > M) \le \beta(M)$, где функция $\beta(M)$ не зависит от параметров рассматриваемой модели и $\lim_{M\to\infty} \beta(M) = 0$. Символ $\widetilde{O}_p(\eta_n)$ будет использоваться в случае, если функция $\beta(M)$ может зависеть от параметров модели (за исключением объема наблюдений n).

• По умолчанию арифметические операции рассматриваются на расширенной числовой прямой при естественных соглашениях типа $c/0 = \infty$ при $c > 0$, $c/\infty = 0$ и специальном соглашении $0/0 = 0$.

• Символ $\Lambda_k(\cdot)$ обозначает меру Лебега в $\mathbb{R}^k$.

• Условимся, что векторы обозначаются полужирными буквами, а матрицы — прямыми заглавными буквами. По умолчанию в качестве векторов, если не оговорено иное, будут рассматриваться вектор-столбцы.

• Запись вида $\boldsymbol{\xi}_n \xrightarrow{d} \mathcal{N}_m(\mathbf{0}, \Sigma)$ означает слабую сходимость распределений $\boldsymbol{\xi}_n$ к m-мерному нормальному распределению с нулевым средним и ковариационной матрицей Σ.

• Будем говорить, что оценка $\boldsymbol{\theta}_n^*$ является α_n-состоятельной для параметра $\boldsymbol{\theta}$, если для некоторой числовой последовательности $\alpha_n \to \infty$ имеет место сходимость по вероятности $\alpha_n(\boldsymbol{\theta}_n^* - \boldsymbol{\theta}) \xrightarrow{p} \mathbf{0}$.

• Для любого вектора $\mathbf{x}$ символ $\|\mathbf{x}\|$ обозначает норму этого вектора, а через $\|\mathrm{A}\|$ обозначается норма матрицы A. Считаем, что матричная норма согласована с векторной нормой и субмультипликативна, т.е. $\|\mathrm{A}\mathbf{x}\| \le \|\mathrm{A}\|\|\mathbf{x}\|$ и $\|\mathrm{AB}\| \le \|\mathrm{A}\|\|\mathrm{B}\|$ для любых матриц A, B и любого вектор-столбца $\mathbf{x}$ соответствующих размерностей. Символ $^\top$ обозначает транспонирование вектора или матрицы. Определитель произвольной квадратной матрицы A обозначается $\det(\mathrm{A})$. Символ I обозначает единичную матрицу, а через $\mathrm{diag}\{d_1, \ldots, d_m\}$ обозначается диагональная матрица размерности $m \times m$ с соответствующими элементами на главной диагонали.

• Условимся, что во всех утверждениях те или иные условия, содержащие неизвестные параметры, нужно проверять при всевозможных значениях этих параметров.

• Завершение доказательств, а также окончание некоторых примеров и замечаний, обозначается символом $\square$.

Общая характеристика работы

Актуальность темы исследования. Регрессионный анализ является одной из широко востребованных и активно развивающихся областей статистической обработки данных (см., например, монографии [16], [21], [38], [49], [51], [60], [70], [74], [91], [92], [94], [98], [104], [105], [110], [136], [158], [161], [167], [186], [204], [222], [223], [229], [234], [236], [248], [253], [255], [259], [261], [266], [275], [281], [282], [295], [299], [310], библиографические подборки [61]–[66], а также приводимые далее ссылки). Диссертационная работа посвящена методологии оценивания в задачах непараметрической и нелинейной регрессии в случае так называемых плотных данных. Подтверждением актуальности рассматриваемых задач может служить большое число публикаций в ведущих научных журналах непосредственно по тематике исследования (см. библиографические ссылки далее).

В разделе диссертации, посвященном непараметрическому оцениванию, рассматриваются классические задачи регрессии: оценивание регрессионной функции по наблюдениям ее зашумленных значений в некотором известном наборе точек из области ее определения, называемых регрессорами, а также оценивание функций среднего и ковариации случайного процесса в схеме, когда каждая из независимых копий этого процесса наблюдается в зашумленном варианте в том или ином наборе регрессоров. Эти задачи общепризнаны фундаментальными и многие работы по непараметрическому оцениванию были посвящены их решению. В качестве ориентира укажем недавние работы [50], [71], [113], [121], [125], [126], [129], [159], [163], [172], [173], [180], [182], [276], [300], [305], [308], [312]–[314], в которых применяются методы ядерного сглаживания (более подробные библиографические сведения по каждой из задач мы приведем в основной части введения).

Для того, чтобы оценить интересующие нас функции, нужно накладывать те или иные ограничения на регрессоры. В многочисленных работах предшественников относительно регрессоров предполагается, что они либо фиксированы и в известном смысле регулярно заполняют область определения регрессионной функции или случайного процесса, либо случайны и состоят из независимых или слабо зависимых величин. В диссертации предложены концепция плотных данных и классы состоятельных оценок в моделях непараметрической и нелинейной регрессии, универсальных относительно стохастической природы набора регрессоров. Эта концепция позволила в задачах непараметрического оценивания существенно ослабить известные условия на регрессоры без какой-либо спецификации их типа. В частности, в работе построены новые непараметрические ядерные оценки для регрессионной функции, равномерно состоятельные лишь при условии асимптотически (при растущем объеме наблюдений) плотного заполнения регрессорами области определения регрессионной функции. В отличие от известных ранее предположений новое условие нечувствительно к характеру зависимости регрессоров, по существу является необходимым для восстановления функции с той или иной точностью и включает в себя как ситуацию детерминированных регрессоров без дополнительного требования регулярности, так и случайных регрессоров, которые могут не удовлетворять условиям слабой зависимости. Концепция плотных данных и универсальности оценок реализуется и в различных регрессионных постановках задачи оценивания функций среднего и ковариации случайного процесса. Доказать равномерную состоятельность новых ядерных оценок лишь при указанных ограничениях на регрессоры (в терминах плотных данных) во многом удается благодаря специальной структуре оценок, содержащей конструкции сумм определенным образом взвешенных наблюдений со структурой интегральных сумм Римана. Конструкции интегральных сумм открывают возможность исследовать асимптотические свойства оценок за счет близости интегральных сумм и соответствующих интегралов, а не предельных теорем теории вероятностей.

В рамках нашей концепции мы исследуем также два наиболее востребованных (см., например, [4], [306], [123]) варианта ядерных оценок: оценки Надарая–Ватсона и классические локально–линейные ядерные оценки. В диссертации для асимптотически плотных данных доказаны как поточечная, так и равномерная состоятельность этих наиболее популярных ядерных оценок. При этом мы, в отличие от предшественников, не используем те или иные эргодические свойства наблюдаемой выборки регрессоров.

Таким образом, в задачах непараметрической регрессии выделяется некоторое свойство регрессоров (в терминах плотных данных), которое в оценивании играет роль «по существу». Принципиальная новизна этих результатов заключается, на наш взгляд, в возможности оценить интересующие нас функции без использования какой-либо информации о структуре регрессоров и характере их зависимости. Полезно отметить, что характер зависимости реальных выборочных данных в статистике, если зависимость наблюдений имеет место по природе стохастического эксперимента, определить бывает достаточно трудно. В этой связи создание и развитие новых методов и подходов статистического анализа зависимых наблюдений, не удовлетворяющих классическим условиям перемешивания и другим известным формам корреляции, а также исследование новых форм зависимостей, которые были бы статистически более наглядными и обоснованными, представляет интерес не только с теоретической точки зрения, но и является актуальным и особенно важным для приложений.

В задачах нелинейной регрессии асимптотически оптимальные оценки как правило задаются неявно и нередко определяются как решения тех или иных уравнений (см., например, монографии [38], [60], [104], [136], [122], [161], [223], [253], [261], [281], [295], [310]). Существуют различные численные методы приближенного поиска таких оценок (см., например, [60], [68], [105], [236], [248], [253], [277]). Стоит отметить, что для задач нелинейной регрессии весьма типична ситуация, когда имеется несколько корней того или иного уравнения, определяющего оценку (см., например, [266], [267], [268], а также рис. 25–26 в главе 3). Данное обстоятельство является главной проблемой, затрудняющей использование численных методов: при неудачном выборе начального приближения итерационные процедуры обнаруживают лишь корень, ближайший к стартовой точке, а не к параметру. Более того, если корней несколько, то как среди всех выбрать корень, приближающий параметр? Эта проблематика, известная в англоязычной литературе как *multiple root problem*, обсуждается, например, в монографиях [122], [266], [278] (см. также ссылки далее, связанные с одношаговыми оценками).

Помимо нелинейного регрессионного анализа, имеется немало разделов статистического оценивания, также связанных с поиском корней тех или иных уравнений (см., например, монографии [80], [84], [114], [122], [149], [150], [151], [196], [209], [240], [247], [261], [266], [278], [281], а также приводимые далее библиографические ссылки). При этом уравнения лишь в исключительных случаях разрешаются в явном виде и зачастую найти нужные корни или их приближения бывает технически сложно, особенно в случае существования нескольких корней исходного уравнения. Один из подходов при решении указанной проблемы состоит в использовании одношаговых оценок. Идея одношагового оценивания, восходящая к работам Р. Фишера [89] и получившая дальнейшее развитие в работах Л. Ле Кама, П. Бикела, П. Хьюбера и др., заключается в следующем: в качестве стартовой точки итерационной процедуры ньютоновского типа используется предварительная состоятельная оценка, сходящаяся к параметру с нужной скоростью. Оказывается, в этом случае зачастую достаточно лишь одного шага итерационной процедуры, чтобы получить явную оценку (так называемую «одношаговую», в англоязычной литературе — *one–step*), имеющую ту же асимптотическую точность, как и искомая статистика (см., например, [278]). Отметим также, что Р. Фишер использовал указанную конструкцию одношаговых оценок для аппроксимации оценок максимального правдоподобия в случае однородных выборок.

Одношаговые процедуры являются весьма популярным современным инструментом статистического оценивания в различных задачах, связанных с поиском корней тех или иных уравнений, и в последние годы интерес к одношаговому оцениванию в статистической литературе только нарастает. В этой связи отметим, например, работы последней четверти века [1], [5], [9], [15], [17], [18], [24], [25], [28]–[30], [40], [41], [52], [67], [73], [76], [78], [79], [81]–[83], [85], [103], [112], [131], [134], [141], [144], [145], [147], [155], [170], [174], [175], [183], [185], [187], [197], [221], [233], [274], [290], [294], [316], [317], [323], а также более ранние работы [19], [48], [59], [88], [148], [152], [184], [198], [210], [262], [263], [264], [271], [273], [293], [296], [297] и монографии [80], [84], [114], [149], [150], [151], [196], [209], [240], [247], [261], [266], [278], [281]. Подчеркнем следующий ключевой момент: конструкция одношаговой оценки предполагает наличие некоторой предварительной оценки. В некоторых задачах такую оценку построить несложно (например, для приближенного вычисления оценок максимального правдоподобия в случае однородных выборок в качестве предварительной оценки, как правило, может выступать оценка метода моментов). Построение такой оценки в задачах нелинейной регрессии может быть отдельной непростой задачей (см., например, [211]). В специальных постановках задач нелинейной регрессии одношаговые оценки исследуются в работах [47], [212], [228], [251], [266], но почти во всех известных автору работах, связанных с одношаговым оцениванием в нелинейной регрессии, существование предварительной оценки лишь постулируется. Исключением является вышеупомянутая недавняя работа [211], посвященная построению предварительной состоятельной оценки для одной частной модели нелинейной регрессии. Важность разработки методологии одношагового оценивания именно для задач нелинейной регрессии подчеркивается, например, в монографии [266].

В диссертации предложены методы построения явных состоятельных с некоторой скоростью (так называемых α_n-состоятельных) или асимптотически нормальных оценок конечномерных параметров в моделях нелинейной регрессии в случае плотного заполнения регрессорами некоторой области и без требования полного контроля над ними. В статистической литературе подобной регулярной методологии построения оценок в задачах нелинейной регрессии до настоящего времени, по-видимому, не было, и лишь для ограниченного круга моделей нелинейной регрессии были известны явные состоятельные оценки. Применительно к упомянутым выше одношаговым процедурам оценивания эти новые оценки могут быть использованы в качестве предварительных, так что наличие подобных методик открывает возможность использования одношаговых оценок в задачах нелинейной регрессии.

Исследования в области нелинейной и непараметрической регрессии, представленные в диссертационной работе, тесно связаны между собой единым подходом к решению поставленных задач. В частности, построение оценок в обоих случаях основано на конструкциях сумм специальным образом взвешенных наблюдений со структурой интегральных сумм Римана, при этом используются схожие универсальные относительно стохастической природы условия на регрессоры. Кроме того, в одном из подходов к построению явных оценок в нелинейной регрессии используются непараметрические ядерные оценки.

В диссертационной работе исследуются также и некоторые варианты одношаговых оценок, которые могут найти применение в задачах нелинейной регрессии. Одним из общих методов получения статистических оценок как в случае однородных, так и неоднородных выборок, является M-оценивание (см., например, монографию [325]). С точки зрения задач нелинейной регрессии этот подход включает в себя метод квазиправдоподобия (см., например, [122]), метод наименьших квадратов, метод максимального правдоподобия (в регулярных случаях) и др. В диссертации проведен подробный асимптотический анализ одношаговых M-оценок, построенных по выборке разнораспределенных наблюдений и имеющих ту же точность, что и асимптотически нормальные M-оценки. Необходимость такого исследования во многом обусловлена приложениями к задачам нелинейной регрессии. Ранее одношаговые M-оценки и, в частности, оценки Фишера были исследованы достаточно полно лишь в случае однородных выборок и, нередко, одномерного параметра (подробные библиографические ссылки мы приведем далее во введении).

Цели диссертационной работы. Основная цель состоит в построении универсальных относительно стохастической природы регрессоров оценок ядерного типа в задачах непараметрической регрессии, являющихся равномерно состоятельными при близких к минимальным и наглядных условиях на регрессоры, а также в получении более общих, чем ранее известные, и универсальных условий состоятельности некоторых классических ядерных оценок. Другая цель состоит в разработке методики построения явных асимптотически близких к оптимальным оценок для широкого класса моделей нелинейной регрессии в случае плотных данных. Эти оценки основаны на применении одношаговых процедур.

Научная новизна. Все результаты диссертации являются новыми. В регрессионном анализе, как правило, модели с фиксированными или случайными регрессорами принято рассматривать отдельно. В диссертации предложена концепция плотных данных и универсальности статистических оценок, открывающая возможность в едином подходе рассматривать ситуацию детерминированных и случайных регрессоров. В непараметрической регрессии эта концепция реализуется как для новых классов ядерных оценок, так и известных ранее, и позволяет существенно ослабить известные ограничения на регрессоры. В задачах нелинейной регрессии предлагаемая концепция позволяет решить открытую проблему поиска явных оценок конечномерных параметров. Остановимся подробнее на полученных в диссертации результатах.

• Построены универсальные относительно стохастической природы регрессоров равномерно состоятельные ядерные оценки для регрессионной функции. Предложено наглядное, близкое к минимальному и универсальное достаточное условие равномерной состоятельности оценок — условие асимптотически плотного заполнения регрессорами области задания регрессионной функции. В отличие от известных ранее результатов новое условие нечувствительно к стохастической природе регрессоров и включает в себя как ситуацию детерминированных регрессоров без дополнительного требования регулярности, так и случайных регрессоров, которые могут не удовлетворять условиям слабой зависимости.

• Построены универсальные ядерные оценки для функций среднего и ковариации случайного процесса, когда зашумленные значения независимых копий этого процесса наблюдаются в некоторых наборах точек (регрессорах), имеющих как плотную, так и разреженную структуру. Универсальные условия, накладываемые на регрессоры и гарантирующие равномерную состоятельность новых ядерных оценок, сформулированы в терминах плотных данных.

• Получены более общие, чем ранее известные, и универсальные относительно стохастической природы регрессоров условия состоятельности оценок Надарая–Ватсона и классических локально–линейных ядерных оценок в терминах плотных данных. Новые условия равномерной состоятельности этих классических ядерных оценок предполагают «более равномерное» (т.е. без резких перепадов) плотное заполнение регрессорами области задания функции, чем требуется для новых универсальных ядерных оценок, предложенных в диссертационной работе. В отличие от работ предшественников, новые условия универсальны и позволяют выйти за рамки слабо зависимых наблюдений, что обосновывает использование этих популярных ядерных оценок при качественном ослаблении условий на регрессоры.

• Разработаны методы построения явных α_n-состоятельных или асимптотически нормальных оценок конечномерных параметров в задачах нелинейной регрессии в случае плотного заполнения регрессорами некоторой области, что решает проблему поиска предварительных оценок для рассматриваемых моделей. Ранее явные оценки были известны лишь для очень ограниченного круга моделей, и проблема построения предварительных оценок для широких классов моделей нелинейной регрессии оставалась открытой. Так что достаточно общие методы построения явных оценок в нелинейной регрессии предложены, по-видимому, впервые. Предлагаемые методы являются еще и универсальными относительно стохастической природы регрессоров.

• Проведен асимптотический анализ одношаговых M-оценок, построенных по разнораспределенным выборочным данным. В такой общности в случае разнораспределенных наблюдений рассматриваемые типы одношаговых оценок исследуются впервые. Эти результаты, вместе с методами построения предварительных оценок, позволяют для широкого класса моделей нелинейной регрессии находить явные одношаговые оценки, асимптотически эквивалентные оценкам метода наименьших квадратов, квазиправдоподобия, максимального правдоподобия и др.

Методы исследования. Основными методами исследования являются общие методы теории вероятностей и математической статистики. В основе доказательств равномерной состоятельности всех изучаемых в работе непараметрических оценок лежит метод диадических цепочек (см., например, [43]), предложенный А. Н. Колмогоровым для оценки хвоста распределения супремальной нормы стохастического процесса с почти наверное непрерывными траекториями. Кроме того, используются методы математического анализа и линейной алгебры.

Теоретическая и практическая значимость. Диссертационная работа имеет теоретический характер. Ее результаты могут быть использованы специалистами в области математической статистики и, в частности, регрессионного анализа. Предлагаемые идеи, методы и подходы могут получить дальнейшее развитие в тех или иных статистических задачах и уже применяются авторами, работающими в Московском государственном университете имени М. В. Ломоносова, Математическом институте имени В. А. Стеклова РАН, Международном математическом центре в Академгородке, Институте математики имени С. Л. Соболева СО РАН, Новосибирском государственном университете и других научных центрах.

Положения, выносимые на защиту. Основные результаты, выносимые на защиту, состоят в следующем.

1. Теоремы о равномерной состоятельности новых универсальных локально–постоянных и локально–линейных ядерных оценок для регрессионной функции скалярного и векторного аргумента в условиях плотного заполнения регрессорами области задания регрессионной функции. Теоремы об асимптотической нормальности новых универсальных локально–постоянных оценок.

2. Теоремы о поточечной и равномерной состоятельности оценок Надарая–Ватсона и классических локально–линейных ядерных оценок. Универсальные достаточные условия состоятельности в терминах плотного заполнения регрессорами области задания регрессионной функции.

3. Теоремы о равномерной состоятельности новых универсальных оценок для функций среднего и ковариации непрерывного случайного процесса как в случае разреженных данных, так и плотных.

4. Новый подход получения явных состоятельных оценок конечномерных параметров в задачах нелинейной регрессии, основанный на использовании ядерных оценок регрессионной функции. Теорема об α_n-состоятельности оценок.

5. Метод построения явных состоятельных оценок конечномерных параметров, основанный на аддитивных преобразованиях откликов. Теоремы об α_n-состоятельности и асимптотической нормальности оценок.

6. Асимптотический анализ одношаговых M-оценок, построенных по разнораспределенным выборочным данным: теорема об асимптотической нормальности одношаговых оценок, теорема о минимальном достаточном условии на точность предварительной оценки, теорема о сходимости к нормальному закону погрешностей оценивания при замене асимптотических дисперсий их оценками.

7. Алгоритм построения асимптотически эффективных одношаговых оценок неизвестного параметра в случае однородной выборки из однопараметрического семейства распределений

и достаточно медленно сходящейся к параметру предварительной оценки. Теорема об асимптотической эффективности новых оценок.

Апробация результатов. Результаты докладывались на Большом семинаре кафедры теории вероятностей МГУ (2021, 2022, 2023), на семинаре проекта МГУ «Диалог о настоящем и будущем» (2023), на научно–исследовательском семинаре кафедры математической статистики ВМК МГУ (2024), на Городском семинаре по теории вероятностей и математической статистике в Санкт-Петербургском отделении Математического института им. В.А. Стеклова РАН (2016), на семинаре по непараметрической статистике в Высшей школе экономике (2023), на семинаре «Прикладная статистика» Института математики им. С.Л. Соболева СО РАН (2022-2024, неоднократно), на объединенном семинаре лаборатории теории вероятностей и математической статистики ИМ СО РАН и кафедры теории вероятностей и математической статистики НГУ (2014-2021, неоднократно). Результаты диссертации докладывались также на конференциях «Третья конференция Математических центров России» (Майкоп, 2023), «Вторая конференция Математических центров России» (Москва, 2022), «Научная конференция сотрудников ИМ СО РАН, посвящённая подведению итогов 2022 года» (Новосибирск, 2022) и следующих международных конференциях: «Limit theorems of probability theory and mathematical statistics» (Ташкент, 2022), «Modern challenges of inverse problems» (Новосибирск, 2022), «Марчуковские научные чтения» (Новосибирск, 2022), «Математика в современном мире» (Новосибирск, 2017), «Modern problems in theoretical and applied probability» (Новосибирск, 2016), «Предельные теоремы в теории вероятностей и их приложения» (Новосибирск, 2011), «Стохастические модели в биологии и предельные алгебры» (Омск, 2010).

Публикации. Основные результаты диссертации опубликованы в 19 статьях, из которых 10 — без соавторов. Все работы опубликованы в журналах, индексируемых в базах данных Web of Science и/или Scopus, список работ приведен в конце диссертации. В работах, написанных в соавторстве, вклад автора диссертации является решающим. Все представленные в диссертации основные результаты и положения, выносимые на защиту, получены лично автором.

Благодарность. Автор выражает особую благодарность научному консультанту профессору Елене Борисовне Яровой за ценные советы и постоянное внимание к работе. Автор признателен профессору Игорю Семеновичу Борисову за научные дискуссии и полезные замечания к работе. Кроме того, автор благодарит Владимира Александровича Куценко и Павла Сергеевича Рузанкина за помощь в подготовке графических иллюстраций первой главы диссертации.

Обзор работ по теме исследований и краткое содержание по главам

Диссертационная работа состоит из введения, трех глав, заключения и библиографического списка. Далее во введении кратко остановимся на содержании диссертации по главам, приведем необходимые библиографические комментарии по теме исследований и некоторые полученные результаты. Отметим, что во введении некоторые утверждения приведены при упрощающих предположениях. Нумерация формул, предположений и утверждений во введении не совпадает с нумерацией в основном тексте диссертации.

Первая глава диссертации посвящена задачам непараметрической регрессии. В разделах 1.1–1.4 рассматривается следующая регрессионная модель: даны наблюдения $X_1, \ldots, X_n$ (так называемые отклики, зависимые или объясняемые переменные), которые представимы в виде

$$X_i = f(\mathbf{z}_i) + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{1}$$

где неизвестная скалярная функция $f(\mathbf{t})$, $\mathbf{t} \in [0,1]^k$, непрерывна; ненаблюдаемые погрешности $\{\varepsilon_i,\ i = 1, \ldots, n\}$ являются центрированными случайными величинами, вообще говоря, не обязательно независимыми или одинаково распределенными; k-мерные векторы $\{\mathbf{z}_i,\ i = 1, \ldots, n\}$ нам известны и могут быть как случайными, так и детерминированными. Таким образом, мы наблюдаем лишь зашумленные значения $\{X_i\}$ регрессионной функции f в некотором известном наборе ее аргументов $\{\mathbf{z}_i\}$, для обозначения элементов которого в научной литературе принято использовать целый ряд терминов: регрессоры, независимые переменные, предикторы, объясняющие переменные, факторы и план эксперимента. Совокупность $\{\mathbf{z}_i, i = 1 \ldots n\}$ мы будем называть набором регрессоров или просто регрессорами. Задача состоит в том, чтобы по наблюдениям $\{(\mathbf{z}_i, X_i);\ i = 1, \ldots, n\}$ восстановить (оценить) функцию f. В случае, если известна функциональная форма зависимости регрессионной функции от конечномерного неизвестного параметра, задача сводится к оцениванию этого параметра и относится к параметрической регрессии. В главе 2 мы рассмотрим такие задачи в случае нелинейной зависимости функции от конечномерного параметра. В главе 1 речь идет исключительно о непараметрической постановке.

Начиная с пионерских работ [243] и [224], посвященных оцениванию плотности распределения, методы ядерного сглаживания широко используются в различных статистических задах непараметрического оценивания (см., например, [22], [36], [127], [123], [193], [194], [195], [282]). Методы ядерного сглаживания, к которым относятся такие известные оценки, как Надарая–Ватсона, Пристли–Чжао, Гассера–Мюллера, локально–полиномиальные оценки и др., весьма популярны и в задачах непараметрической регрессии (см., например, монографии [74], [77], [84], [110], [116], [118], [186], [214], [234], [255], [281], [310]).

Приведем определения некоторых ядерных оценок, которые нам потребуются. В случае одномерных регрессоров оценки Надарая–Ватсона, Присли–Чжао и Гассера–Мюллера задаются следующими соотношениями:

$$\begin{aligned}
\widehat{f}_{NW}(t) &= \frac{\sum_{i=1}^{n} X_i K_h(t - z_i)}{\sum_{i=1}^{n} K_h(t - z_i)}, \\
\widehat{f}_{PC}(t) &= \sum_{i=1}^{n} (z_i - z_{i-1}) K_h(t - z_i) X_i, \\
\widehat{f}_{GM}(t) &= \sum_{i=1}^{n} \left[\int_{s_{i-1}}^{s_i} K_h(t - z_i) \right] X_i,
\end{aligned} \tag{2}$$

где $K_h(t) = h^{-1}K(h^{-1}t)$ и $K(\cdot)$ — некоторая ядерная функция (как правило, это плотность симметричного распределения с носителем $[-1, 1]$ или $\mathbb{R}$), $h > 0$ — размер окна (параметр «сглаживания» в ядерных оценках); регрессоры в двух последних оценках предполагаются упорядоченными по возрастанию, $s_0 = 0$ и $s_i = (z_i + z_{i-1})/2$. Классическая локально–полиномиальная оценка порядка p для функции f определяется как первая координата $(p + 1)$-мерной точки, на которой достигается следующий минимум (см., например, [77]):

$$\min_{(a,b_1,\ldots,b_p)} \sum_{i=1}^{n} (X_i - a - \ldots - b_p(z_i - t)^p)^2 K_h(t - z_i)$$

(остальные координаты указанного вектора оценивают производные f', f'', …, $f^{(p)}$ функции f). Отметим, что оценки Надарая–Ватсона (т.е. классические локально–постоянные оценки) и классические локально–линейные оценки являются представителями класса локально–полиномиальных оценок соответственно порядка $p = 0$ и $p = 1$. При использовании локально–полиномиальных оценок в случае оценивания лишь самой регрессионной функции, а не ее производных, специалисты зачастую рекомендуют ограничиваться именно этими значениями p (см., например, [44], [118]). Дело в том, что при бóльших значениях p у локально–полиномиальных оценок могут проявиться некоторые недостатки. Например, такие оценки могут вообще не существовать при малых значениях размера окна h из-за недостаточного количества выборочных точек. Среди всех вариантов ядерных оценок наиболее популярными оценками принято считать (см., например, [4], [123], [306]) оценки Надарая–Ватсона и классические локально–линейные оценки. На основе этих двух классов оценок часто происходит оценивание и в специальных задачах непараметрической регрессии (например, в задаче оценивания функций среднего и ковариации случайного процесса, см. подробные библиографические ссылки далее). Именно поэтому далее этим двум классам оценок мы уделяем особое внимание. Отметим еще, что ядерные оценки относят к «локальным» методам непараметрического оценивания, поскольку благодаря свойствам ядерной функции при построении оценки в точке t используется «локальная» информация: регрессоры из некоторой окрестности этой точки и соответствующие им отклики.

Для того, чтобы иметь возможность оценить регрессионную функцию, необходимо накладывать те или иные ограничения на параметры модели. Нас в методах ядерного сглаживания в непараметрической регрессии интересуют условия на регрессоры, которые обеспечивают возможность восстановления функции. Поэтому далее в обзоре публикаций, связанных с ядерными оценками в рассматриваемой задаче, остановимся лишь на основных ограничениях на регрессоры.

В статистической литературе регрессионные модели с детерминированными и стохастическими регрессорами принято рассматривать отдельно [1]. Отчасти подобное деление связано, по-видимому, с различными подходами к исследованию оценок в одном и в другом случае. В работах, имеющих дело со случайными регрессорами, часто рассматриваются независимые одинаково распределенные величины (см., например, приведенные выше монографии, а также работы [45], [58], [72], [96], [108], [117], [171], [177], [181], [192], [217], [220], [249]). Стоит отметить, что за несколько последних десятилетий в стохастике было предложено множество форм зависимости случайных величин и доказаны разнообразные вероятностные неравенства и предельные теоремы для последовательностей с такими свойствами. Данные обстоятельства в полной мере отразились и на задачах непараметрической регрессии, в которых все чаще рассматриваются величины (как правило, стационарно связанные), удовлетворяющие тем или иным известным формам зависимости. В частности, для задания регрессоров используются различные варианты условий перемешивания, схемы скользящих средних, ассоциированных случайных величин, марковские или мартингальные свойства и др. (см., например, [77], [115], [125], [126], [142], [163], [164], [166], [172], [199], [200], [201], [202], [208], [245], [260]). В недавних публикациях [37], [39], [95], [154], [182], [286], [288], [289] рассматриваются нестационарные последовательности с теми или иными специальными формами зависимости (марковские цепи, авторегрессия, частичные суммы скользящих средних и др.).

В случае моделей с детерминированными регрессорами в подавляющем большинстве работ предполагаются те или иные условия регулярности (см., например, [11], [12], [13], [69], [109], [117], [135], [276], [298], [322], а также многомерные варианты в [2], [99], [100], [216]). Так, в одномерном случае точки z_i чаще всего задаются формулой $z_i = g(i/n) + o(1/n)$ с некоторой функцией g ограниченной вариации, где погрешность $o(1/n)$ равномерна по всем $i = 1, \ldots, n$. Если g линейна, то получаем так называемый *эквидистантный план.* Другой вариант условия регулярности — соотношение $\max\limits_{i\le n}(z_i - z_{i-1}) = O(1/n)$ (здесь предполагается, что регрессоры упорядочены по возрастанию). Изредка (см., например, недавние работы

[1]В англоязычной литературе нередко используются термины *random design model* и *fixed design model* (см., например, [214]).

[121], [300], [305]) при исследовании оценок типа Пристли–Чжао используется более общее предположение $\max_{i\le n}(z_i - z_{i-1}) \to 0$. Кроме того, иногда (например, при исследовании так называемых взвешенных оценок) накладываются некоторые условия на поведение функций от регрессоров, но содержательные примеры выполнения этих условий зачастую ограничиваются случаями того или иного варианта регулярных регрессоров (см., например, [75], [101], [176], [246], [312], [313]).

Отметим еще, что изначально между теми или иными ядерными оценками существовала некоторая спецификация в зависимости от стохастической природы регрессоров. Так, оценки Надарая–Ватсона использовались лишь в моделях со случайными регрессорами, а оценки Присли–Чжао и Гассера–Мюллера — в моделях с одномерными неслучайными регрессорами. В дальнейшем в указанном направлении был получен ряд обобщений и отмеченные границы между ядерными оценками несколько размылись. Например, оценки Надарая– Ватсона в ситуации неслучайных регулярных многомерных регрессоров исследовались в [99]. Оценки типа Присли–Чжао и Гассера–Мюллера для многомерных неслучайных регулярных регрессоров предложены, соответственно, в [2] и [100]. Оценки типа Пристли–Чжао со случайными одномерными регрессорами рассматривались в [190] (см. также [46], [146]), но только в случае независимых и одинаково распределенных элементов. Имеется точка зрения (см., например, [45]), что оценки типа Присли–Чжао и Гассера–Мюллера в ситуации многомерных случайных регрессоров использовать затруднительно, если вообще возможно.

Задача равномерной аппроксимации оценок ядерного типа изучалась многими авторами (см., например, [37], [58], [72], [95], [109], [115], [117], [135], [172], [176], [177], [192], [200], [220], [260], [286] и ссылки там же).

Помимо оценивания собственно самой регрессионной функции f в модели (1), в случае управляемого эксперимента естественным образом возникает задача планирования (оптимальной организации) эксперимента и, в частности, оптимального в том или ином смысле выбора регрессоров в этой модели. В непараметрической регрессии подобные задачи достаточно сложны (см., например, [20]), поскольку те или иные критерии оптимальности (например, основанные на минимизации интегральной среднеквадратичной погрешности) зависят, как правило, от неизвестных величин. Тем не менее, имеется ряд исследований в этом направлении. Не претендуя на исчерпывающий обзор (и не ограничиваясь здесь лишь направлением ядерного оценивания), для модели (1) с фиксированными регрессорами укажем в качестве примеров работы [11], [20], [71], [87], [215], [256], а со случайными (речь идет об оптимальном в том или ином смысле выборе плотности распределения независимых одинаково распределенных регрессоров), — работы [42], [71] и [318].

В диссертационной работе в модели (1) предполагается, что неизвестная функция f представляет собой случайный процесс с почти наверное непрерывными траекториями. Такую более общую постановку, нежели классическая, мы отчасти рассматриваем для того, чтобы в качестве приложения использовать полученные результаты в задаче оценивания функций среднего и ковариации случайного процесса. В связи со случайной регрессионной функцией отметим, например, недавние работы [113], [159], [173], [180], [306], [308], [311], [315], [319]–[321]. Во введении для простоты будем считать, что в модели (1) функция f неслучайна.

Перейдем к некоторым полученным результатам. Всюду далее, если не оговорено иное, будем предполагать, что регрессоры $\{\mathbf{z}_i;\ i = 1, \ldots, n\}$ в модели (1) — это наблюдаемые случайные величины со значениям в $[0,1]^k$ (вообще говоря, с неизвестными распределениями), не обязательно независимые и одинаково распределенные. Регрессоры можно рассматривать в схеме серий (т.е. они могут зависеть от n), так что данное предположение включает в себя и модели с фиксированными регрессорами (например, модели с эквидистантным планом эксперимента).

В разделах 1.1 и 1.2 рассматривается случай $k = 1$. Обозначим через $K(t)$ плотность симметричного распределения с носителем $[-1, 1]$. Считаем, что ядерная функция $K(t)$ удовлетворяет условию Липшица с константой $L \geq 1$ и $K(\pm 1) = 0$. Определим также плотность $K_h(t) = h^{-1}K(h^{-1}t)$ с носителем на отрезке $[-h, h]$. Обозначим через $z_{n:1} \leq \ldots \leq z_{n:n}$ элементы вариационного ряда, построенного по элементам $\{z_i;\ i = 1, \ldots, n\}$. Положим $z_{n:0} = 0$, $z_{n:n+1} = 1$ и $\Delta z_{ni} = z_{n:i} - z_{n:i-1}$, $i = 1, \ldots, n+1$. Отклик из (1), ассоциированный с порядковой статистикой $z_{n:i}$, обозначим через X_{ni}. Такие ассоциированные случайные величины нередко называют *конкомитантами*. Наконец, определим оценку $f^*_{n,h}(t)$ равенством

$$f^*_{n,h}(t) = \frac{\sum_{i=1}^{n} X_{ni} K_h(t - z_{n:i}) \Delta z_{ni}}{\sum_{i=1}^{n} K_h(t - z_{n:i}) \Delta z_{ni}}. \tag{3}$$

Заметим, что имеет место соотношение

$$f^*_{n,h}(t) = \arg\min_a \sum_{i=1}^{n} (X_{ni} - a)^2 K_h(t - z_{n:i}) \Delta z_{ni},$$

т.е. ядерная оценка $f^*_{n,h}(t)$, подобно классическим оценкам Надарая-Ватсона, является оценкой взвешенного метода наименьших квадратов и принадлежит классу локально–постоянных оценок. Но в методе наименьших квадратов предлагается использовать другие веса, определяемые порядковыми статистиками $\{z_{n:i}\}$, а исходные наблюдения X_i заменить на конкомитанты X_{ni}, ассоциированные с указанными порядковыми статистиками. Новые оценки близки к оценкам, предложенным Пристли и Чжао [231], поскольку числитель дроби в (3) представляет собой оценку типа Пристли–Чжао (см. определение в (2)). Важно подчеркнуть, что оценки Пристли–Чжао не являются равномерно состоятельными (см. [214]) и их исследование за редким исключением ограничивается лишь примерами неслучайных регрессоров.

Условие на регрессоры, обеспечивающее существование равномерно состоятельной оценки в классе оценок (3), состоит в следующем.

(П1) *Имеет место предельное соотношение* $\delta_n = \max\limits_{1\le i\le n+1} \Delta z_{ni} \xrightarrow{p} 0$.

Таким образом, относительно регрессоров требуется лишь, чтобы они образовывали измельчающееся разбиение области определения регрессионной функции, диаметр которого стремится к нулю по вероятности с увеличением объема выборки. На наш взгляд, условие вида (П1) весьма наглядно и по сути является необходимым для восстановления регрессионной функции. Очевидно, что неслучайные регрессоры, регулярно заполняющие $[0,1]$, удовлетворяют условию (П1). Если $\{z_i\}$ независимы и одинаково распределены, а отрезок $[0,1]$ является носителем их общего распределения, то условие (П1) также выполнено. Если $\{z_i\}$ — стационарная последовательность с условием α-перемешивания и маргинальным распределением с носителем $[0,1]$, то условие (П1) также выполнено. Все другие известные в литературе формы слабой зависимости регрессоров также влекут за собой условие (П1). Но выполнение этого условия вполне возможно и для других типов зависимости, которая может быть более сильной, нежели классические условия слабой зависимости (например, когда не выполнены предельные теоремы типа законов больших чисел, см. подробности в примерах 1.1 и 1.2 в главе 1).

Чтобы сформулировать основной результат раздела 1.1, нам дополнительно потребуется следующее ограничение на погрешности в модели (1): при всех $n \ge 1$ ненаблюдаемые случайные величины $\{\varepsilon_i;\, i = 1, \ldots, n\}$ при всех $i, j \le n$ и $i \ne j$ с вероятностью 1 удовлетворяют условиям $\mathbb{E}_{\mathcal{F}_n}\varepsilon_i = 0$, $\sup_{i\le n} \mathbb{E}_{\mathcal{F}_n}\varepsilon_i^2 \le \sigma^2$, $\mathbb{E}_{\mathcal{F}_n}\varepsilon_i\varepsilon_j = 0$, где константа $\sigma^2 > 0$ не зависит от n, а символ $\mathbb{E}_{\mathcal{F}_n}$ обозначает условное математическое ожидание относительно σ-алгебры $\mathcal{F}_n$, порожденной набором $\{z_i;\, i = 1, \ldots, n\}$. В приведенных условиях (за исключением (П1)) справедлива

Теорема 1 (Теорема 1.1). *Для любого фиксированного* $h \in (0, 1/2)$ *с вероятностью* 1 *выполнено*

$$\sup_{t\in[0,1]} |f^*_{n,h}(t) - f(t)| \le \omega_f(h) + \zeta_n(h),$$

где $\omega_f(h) = \sup\limits_{t,s\in[0,1]:|t-s|\le h} |f(t) - f(s)|$, *а случайная величина* $\zeta_n(h)$ *такова, что*

$$\mathbb{P}\left(\zeta_n(h) > y\right) \le C\sigma^2 L^2 y^{-2} h^{-2}\mathbb{E}\delta_n + \mathbb{P}(\delta_n > h/(8L)),$$

и C *есть абсолютная положительная константа.*

Формально в теореме 1 не требуется выполнение условия (П1), но оценки для супремальной нормы, полученные в теореме, становятся, на наш взгляд, содержательными лишь в этом случае. Аналогичное замечание справедливо во всех нижеприводимых утверждениях, в которых получены оценки для соответствующей супремальной нормы. Если выполнено (П1), то $\zeta_n(h) = O_p\big((\sigma+1)Lh^{-1}(\mathbb{E}\delta_n)^{1/2}\big)$ и в качестве размера окна $h = h_n$ можно взять, например, решение уравнения $h^{-1}(\mathbb{E}\delta_n)^{1/2} = \omega_f(h)$. Фактически таким образом выбранный размер окна уравнивает по h порядок малости обоих слагаемых в правой части оценки для супремальной нормы, приведенной в теореме 1.

Теорема 2 (Следствие 1.1). *Пусть выполнены условия теоремы* 1, *предположение* (П1) *и* $\mathcal{C}$ *есть произвольное подмножество равностепенно непрерывных функций в* $C[0,1]$. *Тогда*

$$\gamma_n(\mathcal{C}) = \sup_{f\in\mathcal{C}} \sup_{t\in[0,1]} |f^*_{n,h_n}(t) - f(t)| \xrightarrow{p} 0,$$

где h_n *есть решение уравнения* $h_n^{-1}(\mathbb{E}\delta_n)^{1/2} = \omega_f^{\mathcal{C}}(h_n)$ *при* $\omega_f^{\mathcal{C}}(h) = \sup_{f\in\mathcal{C}}\omega_f(h)$. *Кроме того, выполнено* $\gamma_n(\mathcal{C}) = O_p\big((\sigma+1)L\omega_f^{\mathcal{C}}(h_n)\big)$.

Например, если $\mathbb{E}\delta_n = O(1/n)$ и $\mathcal{C}$ состоит из функций $f(t)$, удовлетворяющих условию Гельдера с показателем $\alpha \in (0,1]$ и универсальной константой, то $h_n = O\big(n^{-\frac{1}{2(1+\alpha)}}\big)$ и $\omega_f^{\mathcal{C}}(h_n) = O\big(n^{-\frac{\alpha}{2(1+\alpha)}}\big)$. В частности, если функции из $\mathcal{C}$ удовлетворяют условию Липшица ($\alpha = 1$) с универсальной константой, то $\gamma_n(\mathcal{C}) = \widetilde{O}_p(n^{-1/4})$.

Следующее утверждение (теорема 3 об асимптотической нормальности) может быть полезно при построении доверительных интервалов для значений регрессионной функции f. Положим

$$B^2_{n,h}(t) = J^{-2}_{n,h}(t)\sum_{i=1}^{n} K_h^2\left(t - z_{n:i}\right)\left(\Delta z_{ni}\right)^2,$$

$$J_{n,h}(t) = \sum_{i=1}^{n} K_h(t - z_{n:i})\Delta z_{ni},$$

$$r_{n,h}(t) = J^{-1}_{n,h}(t)\sum_{i=1}^{n}(f(z_{n:i}) - f(t))K_h\left(t - z_{n:i}\right)\Delta z_{ni}.$$

Теорема 3 (см. теорему 1.2). *Пусть погрешности* $\{\varepsilon_i\}$ *независимы и одинаково распределены, не зависят от регрессоров* $\{z_i\}$, $\mathbb{E}\varepsilon_1 = 0$, $\mathbb{E}\varepsilon_1^2 = \sigma^2$, $\mathbb{E}|\varepsilon_1|^3 < \infty$. *Кроме того, выполнены условия на ядро из теоремы* 1, *условие* (П1), *и при некотором* $t \in [0,1]$ *и* $h \equiv h_n$ *имеет место соотношение*

$$\max_{k\le n} K_h^2\left(t - z_{n:k}\right)\left(\Delta z_{nk}\right)^2\left(\sum\nolimits_{j=1}^{n} K_h^2\left(t - z_{n:j}\right)\left(\Delta z_{nj}\right)^2\right)^{-1} \xrightarrow{p} 0.$$

Тогда $\quad B^{-1}_{n,h}(t)\left(f^*_{n,h}(t) - f(t) - r_{n,h}(t)\right) \xrightarrow{d} \mathcal{N}_1(0,\sigma^2)$.

Таким образом, в работе предлагается класс оценок ядерного типа, асимптотические свойства которого малочувствительны к структуре корреляции регрессоров. Важно подчеркнуть, что новые локально–постоянные оценки универсальны относительно стохастической природы регрессоров и являются равномерно состоятельными не только в вышеуказанных в обзоре случаях зависимости, но и для существенно иной корреляции наблюдений, когда не выполняются условия эргодичности (например, классические условия перемешивания и другие известные условия слабой зависимости). Отметим, что конструкция новых локально–постоянных оценок (3) включает в себя суммы определенных образом взвешенных наблюдений со структурой интегральных сумм Римана, что открывает возможность асимптотические свойства оценок исследовать благодаря сближению интегральных сумм и соответствующих интегралов. В работах предшественников выполнение по сути необходимых условий плотного заполнения регрессорами области задания регрессионной функции в ситуации случайных регрессоров достигалось, как правило, за счет тех или иных форм слабой зависимости, а асимптотические свойства оценок исследовались, как правило, с использованием тех или иных предельных теорем.

Интересно сравнить новые универсальные локально–постоянные оценки с классическими локально–постоянными оценками, т.е. оценками Надарая–Ватсона. Оценки сравним в среднеквадратичном смысле в следующей простой ситуации: регрессоры $\{z_i\}$ независимы, одинаково распределены и не зависят от $\{\varepsilon_i\}$, погрешности $\{\varepsilon_i\}$ независимы, одинаково распределены, центрированы, $0 < \mathbb{E}\varepsilon_1^2 = \sigma^2 < \infty$, а функция распределения $F(t)$ случайной величины z_1 имеет непрерывно дифференцируемую положительную на $(0,1)$ плотность $p(t)$. Положим $\kappa_2 = \int_{-1}^{1} u^2 K(u)du$, $\|K\|^2 = \int_{-1}^{1} K^2(u)du$. Хорошо известен (см., например, [244], [190]) следующий результат, связанный с асимптотическим поведением смещения и дисперсии оценок Надарая–Ватсона $\widehat{f}_{NW}(t)$, определенных в (2).

Предложение 1 (Предложение 1.1; Rosenblatt, 1969). *Если $n \to \infty$ и $h \to 0$ так, что $h^3 n \to \infty$, то для любого $t \in (0,1)$ имеют место соотношения*

$$\mathbb{E}\widehat{f}_{NW}(t) - f(t) = \frac{h^2\kappa_2}{2p(t)}\left(f''(t)p(t) + 2f'(t)p'(t)\right) + o(h^2), \quad \mathbb{D}\widehat{f}_{NW}(t) \sim \frac{\sigma^2}{hnp(t)}\|K\|^2.$$

Асимптотическое представление соответствующих величин для оценки $f^*_{n,h}(t)$ содержится в следующем утверждении.

Предложение 2 (Предложение 1.2). *Пусть $\inf\limits_{t\in[0,1]} p(t) > 0$, $n \to \infty$ и $h \to 0$ так, что $h\sqrt{n}/\log n \to \infty$, $h^{-2}\mathbb{E}\delta_n \to 0$ и $h^{-3}\mathbb{E}\delta_n^2 \to 0$. Тогда при любом $t \in (0,1)$ справедливы соотношения*

$$\mathbb{E}f^*_{n,h}(t) - f(t) = \frac{h^2\kappa_2}{2}f''(t) + o(h^2), \quad \mathbb{D}f^*_{n,h}(t) \sim \frac{2\sigma^2}{hnp(t)}\|K\|^2.$$

Приведенный в предложениях 1 и 2 асимптотический анализ показывает, что в рассматриваемых условиях дисперсия оценки Надарая–Ватсона $\widehat{f}_{NW}(t)$ асимптотически в два раза меньше дисперсии оценки $f^*_{n,h}(t)$. Но среднеквадратичная погрешность любой оценки равна сумме дисперсии и квадрата смещения, которое для двух сравниваемых оценок асимптотически определяется величинами $f''(t)p(t) + 2f'(t)p'(t)$ и $f''(t)p(t)$ соответственно, так что за счет разницы в смещениях можно нивелировать влияние дисперсии на погрешность. Другими словами, если стандартное отклонение погрешностей σ не сильно велико и справедливо неравенство $|f''(t)p(t) + 2f'(t)p'(t)| > |f''(t)p(t)|$, то оценка $f^*_{n,h}(t)$ может быть лучше (в смысле поточечной среднеквадратичной погрешности), чем $\widehat{f}_{NW}(t)$. В противоположной ситуации предпочтительнее оценка Надарая–Ватсона $\widehat{f}_{NW}(t)$. Указанный эффект подтверждают результаты компьютерного моделирования, приведенные в разделе 1.1.5). Ранее близкий эффект в случае независимых одинаково распределенных регрессоров и сравнении оценок Гассера–Мюллера и Надарая–Ватсона отмечался в [96]. Отметим, что коэффициент 2 в асимптотическом представлении для дисперсии новой оценки может быть уменьшен до 1.5, если в определении оценки $f^*_{n,h}(t)$ использовать иной вариант разбиения с отмеченными точками (разбиение Вороного; см. подробности в замечании 1.15).

В разделе 1.2 предложен и исследуется второй вариант новых универсальных оценок, относящихся к классу локально–линейных ядерных оценок. Эти оценки проще всего определить как первую координату вектора, на котором достигается минимум

$$\min_{(a,b)} \sum_{i=1}^{n} (X_{ni} - a - b(t - z_{n:i}))^2 K_h(t - z_{n:i}) \Delta z_{ni},$$

где величины $z_{n:i}$, Δ_{ni} и X_{ni} введены выше перед формулой (3). Универсальные локально–линейные оценки имеют близкие свойства с универсальными локально-постоянными оценками, определенными в (3). В частности, равномерную состоятельность оценок в обоих случаях обеспечивает (относительно регрессоров) условие (П1). Качественное отличие этих двух новых классов оценок наблюдается, например, в окрестностях граничных точек 0 и 1: для оценки $f^*_{n,h}(t)$ в h-окрестностях указанных точек порядок малости смещения h, а для универсальной локально–линейной оценки — порядок h^2. Такая связь между новыми универсальными оценками представляется весьма естественной ввиду известной взаимосвязи в граничных точках между классическими оценками Надарая–Ватсона и классическими локально–линейными оценками.

В разделе 1.3 рассматривается модель (1) при произвольном фиксированном $k \geq 1$ и получены обобщения некоторых результатов разделов 1.1 и 1.2 на случай оценивания случайной вещественнозначной регрессионной функции нескольких переменных (случайного поля). Приведем один из основных результатов. Для простоты считаем, что регрессоры $\{\mathbf{z}_i, i = 1, \ldots, n\}$ в модели (1) попарно различны. В этом случае относительно этого набора предполагается выполненным следующее условие (см. условие (D_2) и замечание 1.16).

(П2) *Для каждого n существует такое разбиение k-мерного куба $[0,1]^k$ на n измеримых по Жордану подмножеств $\{\mathcal{P}_i,\ i = 1, \ldots, n\}$, что каждый элемент этого разбиения содержит ровно по одной точке из набора $\{\mathbf{z}_i;\ i = 1, \ldots, n\}$ (нумерация элементов разбиения такова, что $\mathbf{z}_i \in \mathcal{P}_i$), при этом $\delta_n = \max_{i \le n} d(\mathcal{P}_i) \xrightarrow{p} 0$, где $d(A) = \sup_{\mathbf{x},\mathbf{y} \in A} \|\mathbf{x} - \mathbf{y}\|$ – диаметр множества, $\|\cdot\|$ — супремальная норма в $\mathbb{R}^k$.*

Другими словами, условие (П2) означает, что случайные величины $\{\mathbf{z}_i, i \le n\}$ образуют ε-сеть множества $[0,1]^k$ при $\varepsilon = \delta_n$ с условием $\delta_n \xrightarrow{p} 0$. Один из вариантов новых универсальных оценок для функции $f : [0,1]^k \to \mathbb{R}$ определяется равенством

$$f^*_{n,h}(\mathbf{t}) = \frac{\sum_{i=1}^{n} X_i K_h(\mathbf{t} - \mathbf{z}_i) \Lambda_k(\mathcal{P}_i)}{\sum_{i=1}^{n} K_h(\mathbf{t} - \mathbf{z}_i) \Lambda_k(\mathcal{P}_i)}, \tag{4}$$

где $\{\mathcal{P}_i\}$ есть разбиение множества $[0,1]^k$, введенное в условии (П2), $K_h(\mathbf{s}) = h^{-k} K(h^{-1}\mathbf{s})$ и $K(\mathbf{s})$, $\mathbf{s} \in \mathbb{R}^k$ — ядерная функция. Нетрудно видеть, что

$$f^*_{n,h}(\mathbf{t}) = \arg\min_a \sum_{i=1}^{n} (X_i - a)^2 K_h(\mathbf{t} - \mathbf{z}_i) \Lambda_k(\mathcal{P}_i),$$

так что предлагаемые оценки относятся к классу локально–постоянных, но с некоторыми иными весами, нежели в классическом варианте. Эти веса задаются мерой Лебега элементов некоторого конечного случайного разбиения выборочного пространства регрессоров, введенного в (П2). Подобные веса позволяют нам образовать конструкцию кратных интегральных сумм Римана в структуре предлагаемых ядерных оценок. Указанное разбиение с отмеченными точками можно построить, например, методом последовательных покоординатно-медианных сечений или с помощью мозаики Вороного (см. подробности в разделе 1.3.3).

В диссертации доказана равномерная состоятельность и асимптотическая нормальность оценок $f^*_{n,h}(\mathbf{t})$. Остановимся подробнее на первом свойстве оценок. Относительно ядра будем считать, что функция $K(\mathbf{s})$ является плотностью распределения с носителем $[-1,1]^k$, удовлетворяет условию Липшица с константой $L \ge 1$ и $K(-\mathbf{s}) = K(\mathbf{s})$. Кроме того, существуют константы $\rho > 0$ и $h_0 \in (0,1]$ такие, что $\int_{[0,1]^k} K_h(\mathbf{t} - \mathbf{x})\, \Lambda_k(\mathbf{dx}) \ge \rho$ при всех $\mathbf{t} \in [0,1]^k$ и $0 < h \le h_0$. Предполагается, что погрешности $\{\varepsilon_i\}$ не зависят от $\{\mathbf{z}_i\}$ и при всех $n \ge 1$ случайные величины $\{\varepsilon_i;\ i = 1, \ldots, n\}$ образуют последовательность мартингал-разностей с условием $M_p = \sup_{i \le n} \mathbb{E}|\varepsilon_i|^p < \infty$ при некотором $p > k$ и $p \ge 2$, где M_p не зависит от n. В приведенных условиях справедлива

Теорема 4 (см. теорему 1.4). *Для любого фиксированного $h \in (0, h_0)$ с вероятностью 1 выполнено*

$$\sup_{\mathbf{t} \in \mathcal{P}} |f^*_{n,h}(\mathbf{t}) - f(\mathbf{t})| \le \omega_f(h) + \zeta_n(h), \tag{5}$$

где $\omega_f(h)$ — модуль непрерывности f, а случайная величина $\zeta_n(h) > 0$ такова, что

$$\mathbb{P}\left(\zeta_n(h) > y\right) \le C\rho^{-p} M_p L^p y^{-p} h^{-k(p/2+1)} \mathbb{E}(\delta_n^{kp/2}) + \mathbb{P}(\delta_n > h \min\{1,\ \rho(k2^{k+1}L)^{-1}\});$$

константа C зависит от k и p.

В следующем утверждении величина h_n выбирается так, чтобы минимизировать по h порядок малости правой части соотношения (5).

Теорема 5 (Следствие 1.5). *Пусть $\mathcal{C}$ — множество неслучайных равностепенно непрерывных функций из пространства $C\big([0,1]^k\big)$. Тогда в условиях теоремы 4*

$$\gamma_n(\mathcal{C}) = \sup_{f \in \mathcal{C}} \sup_{\mathbf{t} \in [0,1]^k} |f^*_{n,h_n}(\mathbf{t}) - f(\mathbf{t})| \xrightarrow{p} 0,$$

где h_n определяется как решение уравнения $\mathbb{E}(\delta_n^{kp/2}) = h^{k(p/2+1)}\big(\omega_f^{\mathcal{C}}(h)\big)^p$ при

$$\omega_f^{\mathcal{C}}(h) = \sup\nolimits_{f \in \mathcal{C}} \omega_f(h).$$

Кроме того, $\gamma_n(\mathcal{C}) = \widetilde{O}_p(\omega_f^{\mathcal{C}}(h_n))$.

Второй класс новых оценок, исследуемых в работе, — универсальные локально–линейные оценки. Эти оценки можно определить как первую координату $(k+1)$-мерного вектора, на котором достигается минимум

$$\min_{a,\mathbf{b}} \sum\nolimits_{i=1}^{n} \left(X_i - \left(a + \mathbf{b}^\top(\mathbf{t} - \mathbf{z}_i)\right)\right)^2 K_h(\mathbf{t} - \mathbf{z}_i)\Lambda_k(\mathcal{P}_i).$$

Явное представление для этих оценок, требующее введения матричных обозначений, а также аналоги теорем 4 и 5, мы опускаем. Отметим лишь, что единственное условие на k-мерные регрессоры $\{\mathbf{z}_i\}$, гарантирующее существование равномерно состоятельных оценок в данном классе — это условие (П2) плотного заполнения регрессорами области определения функции f.

В разделе 1.4 исследуются оценки Надарая–Ватсона и классические локально–линейные оценки. Цель данного раздела — ослабить известные ограничения на регрессоры, гарантирующие состоятельность (в смысле как поточечной, так и равномерной сходимости по вероятности) указанных популярных ядерных оценок, а также реализовать для этих оценок идеи об общих и универсальных условиях на регрессоры в терминах плотных данных. Доказано, что рассматриваемые классические ядерные оценки могут быть равномерно состоятельными не только в вышеуказанных в обзоре случаях слабой зависимости, но и для существенно иной корреляции наблюдений, когда не выполняются свойства эргодичности. Каким образом удалось достичь этих целей? При изучении, например, асимптотических свойств оценки Надарая–Ватсона, нередко принято отдельно исследовать асимптотическое поведение числителя и знаменателя дроби, определяющей эту оценку (см. определение (2)). При этом в ситуации случайных регрессоров используются, как правило, те или иные формы предельных теорем, что и объясняет известные в литературе условия на корреляцию наблюдений. Оказывается, если указанную дробь исследовать «в целом» как аддитивную статистику с нормированными весами и при этом для вывода равномерной состоятельности вместо оценок в теоремах типа Гливенко–Кантелли (см., например, [220]) использовать метод диадических цепочек, то условия на регрессоры удается получить лишь в терминах асимптотического поведения числа регрессоров, попавших в ту или иную окрестность точек из области задания f.

Приведем результаты раздела 1.4, связанные с оценками Надарая–Ватсона, при этом для простоты рассмотрим модель (1) при $k = 1$. Сохраним обозначение $K(t)$ для плотности симметричного распределения с носителем $[-1, 1]$, при этом $\sup_{t\in[0,1]} K(t) \le L < \infty$ и существует положительное $\delta \le 1$ такое, что $\inf_{|t|\le\delta} K(t) \ge l > 0$. Кроме того, относительно погрешностей $\{\varepsilon_i\}$ выполнено условие, введенное перед теоремой 1. Для любых $z_1, \ldots, z_n$, $h \in (0, 1)$ и $t \in [0, 1]$ положим $N_{n,h}(t) = \{i : |t - z_i| \le h, \ 1 \le i \le n\}$. Обозначим через $\#(\cdot)$ стандартную считающую меру и введем следующее условие.

(П3) При всех фиксированных $t \in [0, 1]$ и $h \in (0, 1)$ имеет место предельное соотношение $\#(N_{n,h}(t)) \xrightarrow{p} \infty$.

В приведенных условиях (за исключением (П3)) справедлива

Теорема 6 (Следствие 1.7). *Для любого* $h \in (0, 1)$ *и любого* $t \in [0, 1]$ *с вероятностью* 1 *выполнено*

$$\widehat{f}_{NW}(t) - f(t) = O_p\left(\omega_f(h) + \frac{\sigma\sqrt{L}}{\sqrt{l\#(N_{n,\delta h}(t))}}\right).$$

Теорема 7 (Следствие 1.8) *Если в условиях теоремы* 6 *выполнено предположение* (П3), *то существует последовательность* $h = h_n \to 0$, *для которой при всех* $t \in [0, 1]$ *имеет место предельное соотношение* $\widehat{f}_{NW}(t) \xrightarrow{p} f(t)$.

Отметим, что условие (П3), обеспечивающее *поточечную* состоятельность оценок Надарая–Ватсона, эквивалентно условию (П1), гарантирующему *равномерную* состоятельность новых универсальных оценок.

Рассмотрим теперь вопрос о равномерной состоятельности оценки Надарая–Ватсона. Нам потребуется следующее ограничение.

(П4) *При всех* $h \in (0,1)$ *имеет место сходимость*

$$\Delta_{n,h} = \sup_{t\in[0,1]} \frac{\sup_{|s|\le h} \#^3(N_{n,h}(t+s))}{\#^4(N_{n,\delta h}(t))} \xrightarrow{p} 0.$$

Теорема 8 (Теорема 1.10). *Пусть в условиях теоремы* 6 *функция* $K(t)$ *непрерывно дифференцируема на* $(0,1)$ *и* $\sup_{t\in[0,1]}|K'(t)| \le L < \infty$. *Тогда для любого фиксированного* $h \in (0,1)$ *с вероятностью* 1 *выполнено*

$$\sup_{t\in[0,1]} |\widehat{f}_{NW}(t) - f(t)| \le \omega_f(h) + \zeta_n(h),$$

где $\omega_f(\cdot)$ — *модуль непрерывности функции* f, *а* $\zeta_n(h)$ *есть последовательность положительных случайных величин, удовлетворяющих соотношению*

$$\zeta_n(h) = O_p\left(\sigma(L/l)^2\sqrt{\Delta_{n,h}h^{-1}}\right).$$

Теорема 9 (Следствие 1.10). *Если в условиях теоремы* 8 *выполнено предположение* (П4), *то существует последовательность* $h = h_n \to 0$, *для которой* $\sup_{t\in[0,1]}|\widehat{f}_{NW}(t) - f(t)| \xrightarrow{p} 0$.

Таким образом, условие (П4) — это единственное ограничение на регрессоры $\{z_i\}$, обеспечивающее равномерную состоятельность оценок Надарая–Ватсона. В частности, (П4) выполнено, если

$$\Delta'_{n,h} = \frac{\sup_{t\in[0,1]} \#^3(N_{n,h}(t))}{\inf_{t\in[0,1]} \#^4(N_{n,\delta h}(t))} \xrightarrow{p} 0.$$

В широких условиях величина $\#(N_{n,h}(t))$ имеет порядок (в известном смысле) nh. Например, в случае независимых одинаково распределенных регрессоров с распределением z_1, имеющим ограниченную и отделенную от нуля на отрезке $[0,1]$ плотность, а также в случае, когда регрессоры удовлетворяют условию φ-перемешивания. Аналогичная оценка в широких условиях справедлива, например, для неслучайных регулярных регрессоров. Можно построить примеры (см. пример 1.16), когда указанная оценка имеет место в случае сильно зависимых $\{z_i\}$ для нестационарной последовательности, не удовлетворяющей закону больших чисел. Для указанных $\{z_i\}$ величина $\Delta'_{n,h}$ имеет порядок $O((hn)^{-1})$, что гарантирует выполнение (П4). Сравнивая условия (П1) и (П4) отметим, что условие (П4) по сути предполагает более равномерное плотное заполнение регрессорами области определения регрессионной функции, нежели требуется в условии (П1) (см. пример 1.18). Полезно отметить, что результаты компьютерного моделирования, приведенные в разделах 1.1–1.3 показывают, что новые оценки в тех или иных ситуациях могут быть в известном смысле точнее классических аналогов (оценок Надарая–Ватсона и локально–линейных оценок).

Раздел 1.5 завершает первую главу диссертации и посвящен оцениванию некоторых характеристик случайного процесса. Постановка задачи следующая. Пусть $f_1(t), \ldots, f_n(t)$ — это независимые копии некоторого почти наверное непрерывного случайного процесса $f(t)$, определенного на $[0, 1]$. Задача состоит в оценивании функций среднего $\mu(t) = \mathbb{E}f(t)$ и ковариации $\psi(t, s) = \mathbb{C}ov\{f(t), f(s)\}$, когда сами случайные функции $\{f_i(t), i = 1, \ldots, n\}$ нам неизвестны и мы наблюдаем лишь зашумленные их значения в некотором известном наборе точек (вообще говоря, своем для каждой из этих функций). Обозначим через X_{ij} зашумленное значение i-ой функции $f_i(t)$ при $t = Z_{ij}$, $j = 1, \ldots, m_i$. Таким образом, нам даны пары наблюдений $\{(Z_{ij}, X_{ij}),\ i = 1, \ldots n,\ j = 1, \ldots, m_i\}$ со следующей структурой:

$$X_{ij} = f_i(Z_{ij}) + \varepsilon_{ij}, \qquad i = 1, \ldots, n, \quad j = 1, \ldots, m_i, \tag{6}$$

где $\{\varepsilon_{ij}\}$ — ненаблюдаемые случайные погрешности, известные величины $\{Z_{ij}\}$ могут быть как случайными, так и детерминированными.

Оценивание функций среднего и ковариации случайного процесса $f(t)$ по выборочным данным со структурой (6) является фундаментальной задачей в так называемом функциональном анализе данных (в широком смысле включающем в себя статистический анализ данных, представленных в виде функций; см., например, [50], [159], [213], [129]) и многие недавние работы по непараметрическому оцениванию были посвящены ее решению. Неполный список таких работ включает, например, [27], [31]–[33], [113], [129], [157], [173], [180], [188], [269], [285], [299], [306], [308], [311], [314], [315], [319], [320]. Оценки для функций среднего и ковариации могут представлять как самостоятельный интерес, так и играть важную вспомогательную роль в том или ином последующем анализе (см., например, [129], [137], [159], [173], [213], [287], [315]).

Приведем краткий обзор публикаций, связанных с рассматриваемой задачей. Мы не стремимся представить всесторонний обзор данной активно развивающейся (особенно последние два десятилетия) области непараметрического оценивания, и укажем лишь некоторые публикации, представляющие те или иные направления. Подходы к решению указанной задачи можно условно разделить на две основные группы: методы ядерного сглаживания ([113], [157], [173], [269], [306], [308], [311], [314], [315], [319]–[321]) и сглаживание сплайнами ([31]–[33], [102], [188], [238], [239], [285], [307]). Для рассматриваемой задачи наиболее часто исследуются три вида асимптотических свойств оценок: равномерная состоятельность, ([129], [173], [308], [315], [320]), L_2-состоятельность ([31], [113], [315]), асимптотическая нормальность ([102], [157], [306], [311], [315]). Вопросы интервального оценивания рассматриваются, например, в

[32], [33], [53], [157], [188], [269], [285], [319]. В контексте рассматриваемой задачи исключительная важность свойства равномерной состоятельности оценок отмечается, например, в работах [173] и [308].

Нас в первую очередь интересуют ограничения на регрессоры $\{Z_{ij}\}$, которые используются в модели (6), поэтому подробнее остановимся на основных ограничениях на элементы $\{Z_{ij}\}$. Набор регрессоров в модели (6) может как изменяться от серии к серии, так и быть одинаковыми для всех серий (в этом случае говорят об *общем* наборе регрессоров для каждой из независимых реализаций случайного процесса). Те или иные общие наборы регрессоров рассматривались, например, в [31]–[33], [53], [102], [238], [269], [285]. По стохастической природе регрессоры принято рассматривать либо случайными ([27], [31], [113], [157], [173], [188], [306], [308], [311], [314], [315], [319]–[321]), либо детерминированными ([31]–[33], [53], [102], [238], [269], [284], [285]). Как и в классической постановке задачи непараметрической регрессии (1), модели (6) со случайными или детерминированными регрессорами рассматриваются отдельно. Отметим, что в первом случае, как правило, предполагается, что случайные величины $\{Z_{ij}\}$ являются независимыми и одинаково распределенными (это условие используется во всех указанных выше работах). Некоторые авторы подчеркивают (см., например, [113]), что их результаты можно перенести и на слабо зависимые величины. Для детерминированных регрессоров зачастую дополнительно требуется то или иное условие регулярности. Например, один из популярных вариантов таких условий (см., например, [32], [33], [113], [269], [284], [285]) — это требование так называемого *общего эквидистантного плана*, т.е. когда для любого i выполнено $m_i = m$ и $Z_{ij} = j/m$ при $j = 1, \ldots, m$. Другой вариант условий регулярности можно найти, например, в [53].

Данные в модели (6) подразделяются на те или иные типы в зависимости от количества наблюдений для той или иной реализации случайного процесса. Так, данные могут быть в некотором смысле *плотными*, или *разреженными* (в английской терминологии — *dense functional data* и *sparse functional data*, соответственно), или смешанными. Хотя в рассматриваемой задаче не существует строгого разделения типов данных (см., например, [287], [315]), тем не менее, данные принято относить к разреженным (неплотным) в одном из двух случаях: либо когда количество наблюдений в каждой серии неслучайно и равномерно ограничено, т.е. $\max_{1\le i\le n} m_i \le c$ и константа c не зависит от n ([31], [113], [173], [320]), либо когда m_i случайны и являются независимыми копиями некоторой положительной целочисленной случайной величины ([306], [308], [319]). К плотным данным относят, например, случаи, когда $\min_{1\le i\le n} m_i \ge m(n) \to \infty$ ([31], [32], [53], [173], [269], [285], [311], [320]). Стоит отметить, что в литературе основное внимание уделяется этим двум типам данных (разреженным или плотным). Иные данные, в том числе смешанного типа, когда для некоторых реализаций случайного процесса данные могут быть плотными, а для других разреженными, рассматривались в [31], [173], [315].

Методологии оценивания функции среднего, используемые для плотных или разреженных данных, как правило, различны (см., например, [213], [287]). В ситуации растущего объема наблюдений в каждой серии с увеличением числа серий (т.е. количества случайных функций), естественно предварительно оценить случайную регрессионную функцию в каждой серии, а затем провести усреднение по всем сериям (см., например, [31], [113], [311]). Для разреженных данных такой способ построения оценки не будет работать в силу недостаточности информации, относящейся к каждой реализации, и зачастую наблюдения каким-либо образом объединяют для заимствования информации друг у друга (см., например, [113], [308], [315]). Имеется точка зрения (см., например, [287]), что оценивание для неплотных данных нередко требует больше усилий, чем для плотных. Некоторые унифицированные подходы, которые годятся как для плотных, так и для разреженных функциональных данных, можно найти в [31], [173], [315].

В разделе 1.5 диссертации предложены новые универсальные классы равномерно состоятельных оценок ядерного типа для функций среднего и ковариации как в случае разреженных, так и плотных данных. В отличие от работ предшественников, мы не накладываем на регрессоры общепринятых условий регулярности или независимости, предлагаемые условия нечувствительны к характеру корреляции регрессоров как внутри серии, так и между сериями. Так, при оценивании функции среднего в случае разреженных данных относительно регрессоров лишь требуется, чтобы вся их совокупность из всех серий с ростом объема наблюдений с высокой вероятностью образовывала измельчающееся разбиение отрезка $[0,1]$ — области определения случайного процесса, а для плотных данных подобное условие должно быть выполнено для набора регрессоров каждой из серий.

Перейдем к некоторым полученным результатам. Считаем, что в модели (6) набор регрессоров $\{Z_{ij}\ i=1,\dots n,\ j=1,\dots,m_i\}$ состоит из наблюдаемых случайных величин со значениями в $[0,1]$ и, вообще говоря, с неизвестными распределениями, при этом не обязательно независимых или одинаково распределенных. Для каждого i случайные величины i-й серии $\{Z_{ij},\ j=1,\dots,m_i\}$ могут зависеть от m_i (если m_i неслучайно) и n. В частности, рассматриваемая нами схема включает в себя и ситуацию детерминированных регрессоров.

Рассмотрим сначала более сложную ситуацию — случай разреженных данных. Считаем, что случайные функции $\{f_i(t)\}$ не зависят от $\{Z_{ij}\}$ и $\sup_{t\in[0,1]}\mathbb{D}f_1(t)\le\sigma_f^2<\infty$. Ограничения на ядро K здесь те же, что и в теореме 1.

Приведем вкратце идею построения оценки для функции среднего $\mu(t)$. Пусть случайные величины $\{m_i\}$ (не обязательно независимые или одинаково распределенные) не зависят от

$\{f_i(t)\}$ и $\{Z_{ij}\}$, а также не зависят от n. Положим $N = m_1+\ldots+m_n$ и по N элементам выборки $\{Z_{ij};\ i = 1,\ldots,n,\ j = 1,\ldots,m_i\}$ образуем вариационный ряд, элементы которого обозначим через $Z_{N:1} \le \ldots \le Z_{N:N}$. Пусть $N = lr + s$, где l, r и s — целые, r неслучайно, а случайные величины l и s таковы, что $1 \le s < r$ почти наверное (т.е. s — это остаток от деления N на l, который во введении для простоты полагаем равным нулю). Считаем, что $r = r(n) \to \infty$ и $r = o(n)$, так что $l = l(n) \ge n/r$ также неограниченно возрастает с ростом n. Положим $Z_{N:0} = 0$, $Z_{N:N+1} = 1$, $\Delta Z_{Nl} = Z_{N:N+1} - Z_{N:r(l-1)}$, $\Delta Z_{Nk} = Z_{N:rk} - Z_{N:r(k-1)}$, $k = 1,\ldots,l-1$. Таким образом, отрезок $[0,1]$ мы разбили на l попарно несовместных отрезков с длинами $\Delta Z_{N1},\ldots,\Delta Z_{Nl}$, каждый из которых содержит по r точек. Определим теперь множества $H_1 = \{(i,j) :\ \ Z_{ij} \in [Z_{N:0},\ Z_{N:r}]\}$, $H_k = \{(i,j) :\ \ Z_{ij} \in (Z_{N:r(k-1)},\ Z_{N:rk}]\}$, $k = 2,\ldots,l$, содержащие ровно по r пар индексов. Введем обозначения

$$\overline{X}_k = r^{-1}\sum_{(i,j)\in H_k} X_{ij}, \quad \overline{f}_k = r^{-1}\sum_{(i,j)\in H_k} f_i(Z_{ij}), \quad \overline{\varepsilon}_k = r^{-1}\sum_{(i,j)\in H_k}\varepsilon_{ij}, \quad k = 1,\ldots,l.$$

Из уравнения (6) имеем $\overline{X}_k = \overline{f}_k + \overline{\varepsilon}_k$, $k = 1,\ldots,l$. Но в силу введенных условий и определений можно ожидать, что $\overline{f}_k \approx \mu(t)$, где $t \in (Z_{N:r(k-1)},\ Z_{N:rk}]$ (например, можно положить $t = Z_{N:rk}$). Иными словами, исходную задачу мы свели к следующей классической модели непараметрической регрессии: $\overline{X}_k \approx \mu\big(Z_{N:rk}\big) + \overline{\varepsilon}_k$, $k = 1,\ldots,l$. Мы предлагаем оценить функцию среднего $\mu(t)$ в этой модели регрессии с помощью метода ядерного сглаживания, предложенного в разделе 1.1, что приводит нас к следующей оценке:

$$\widehat{\mu}_1(t) = \frac{\sum_{k=1}^{l}\overline{X}_k K_h(t - Z_{N:rk})\Delta Z_{Nk}}{\sum_{k=1}^{l} K_h(t - Z_{N:rk})\Delta Z_{Nk}}.$$

Относительно погрешностей считаем, что случайные величины $\{\varepsilon_{ij};\ i = 1,\ldots,n,\ j = 1,\ldots,m_i\}$ при всех i,j, а также $(i_1,j_1) \ne (i_2,j_2)$, с вероятностью 1 удовлетворяют следующим условиям: $\mathbb{E}_{\mathcal{F}}\varepsilon_{ij} = 0$, $\max_{i,j}\mathbb{E}_{\mathcal{F}}\varepsilon_{ij}^2 \le \sigma_\varepsilon^2$, $\mathbb{E}_{\mathcal{F}}\varepsilon_{i_1j_1}\varepsilon_{i_2j_2} = 0$, где константа $\sigma_\varepsilon^2 > 0$ не зависит от n, символ $\mathbb{E}_{\mathcal{F}}$ обозначает условное математическое ожидание при фиксации σ-алгебры $\mathcal{F}$, порожденной случайными величинами $\{Z_{ij};\ i = 1,\ldots,n,\ j = 1,\ldots,m_i\}$ и $\{m_i, i = 1,\ldots,n\}$. В приведенных условиях справедлива

Теорема 10 (см. теорему 1.13). *Для любого фиксированного $h \in (0,1/2)$ с вероятностью 1 выполнено*

$$\sup_{t\in[0,1]}|\widehat{\mu}_1(t) - \mu(t)| \le \omega_\mu(h) + \omega_\mu(\delta_l) + \zeta_{l,r,h} + \eta_{l,r},$$

где $\delta_l = \max_{1\le k\le l}\Delta Z_{Nk}$, $\omega_\mu(\cdot)$ — модуль непрерывности функции $\mu(t)$, а случайные величины

$\zeta_{l,r,h}$ и $\eta_{l,r}$ таковы, что

$$\mathbb{P}\big(\zeta_{l,r,h} > y,\, \delta_l \le h/(8L)\big) \le C\sigma_\varepsilon^2 L^2 y^{-2} r^{-1} h^{-2} \mathbb{E}\delta_l,$$
$$\mathbb{P}(\eta_{l,r} > y) \le 2\sigma_f^2 M^3 n r^{-2} y^{-2} + \mathbb{P}\big(\max_{i\le n} m_i > M\big);$$

здесь $C > 0$ — абсолютная постоянная, а M — произвольное положительное число.

Из теоремы 10 вытекает, что условие на регрессоры, гарантирующее в случае разреженных данных существование равномерно состоятельной оценки для функции среднего, состоит в следующем: $\delta_l \xrightarrow{p} 0$ с ростом n. Теорема 10 универсальна и относительно стохастической природы величин $\{m_i\}$ — количества наблюдений по сериям. Если $\delta_l \xrightarrow{p} 0$ и выполнены все ограничения, введенные перед теоремой 10, то имеют место следующие два утверждения, включающие в себя оба популярных варианта разреженных данных, рассматриваемых в работах предшественников.

Теорема 11 (Следствие 1.15). *Пусть $\{m_i\}$ случайны и являются независимыми копиями положительной целочисленной случайной величины, $\mathbb{E}m_1^\alpha < \infty$ при некотором $\alpha > 3$ и $h \to 0$, $r^{-1}h^{-2}\mathbb{E}\delta_l \to 0$, $\mathbb{P}(\delta_l > h/(8L)) \to 0$, $n^{(3+\alpha)/(2\alpha)}r^{-1} \to 0$. Тогда $\sup\limits_{t\in[0,1]} |\widehat{\mu}_1(t) - \mu(t)| \xrightarrow{p} 0$.*

Теорема 12 (Следствие 1.17). *Пусть $\{m_i\}$ неслучайны, $\max_{i\le n} m_i \le c$ для некоторой константы c, не зависящей от n, и $h \to 0$, $r^{-1}h^{-2}\mathbb{E}\delta_l \to 0$, $\mathbb{P}(\delta_l > h/(8L)) \to 0$, $l/r \to 0$. Тогда $\sup_{t\in[0,1]} |\widehat{\mu}_1(t) - \mu(t)| \xrightarrow{p} 0$.*

Определим теперь в случае разреженных данных оценку для второго смешанного момента $\varphi(t,s) = \mathbb{E}f_1(t)f_1(s)$. Пусть $\{m_i\}$ неслучайны, $m_i \ge 2$ при всех i и $\max_{i\le n} m_i \le c$, где константа c не зависит от n. Положим $\widetilde{N} = \widetilde{N}(n) = m_1(m_1-1) + \ldots + m_n(m_n-1)$. Без ограничения общности считаем, что $\widetilde{N} = \tilde{l} \times \tilde{r}$, где $\tilde{l}$ и $\tilde{r}$ — целые, при этом $\tilde{l} = \tilde{l}(n) \to \infty$ $\tilde{r} = \tilde{r}(n) \to \infty$. Пусть случайные погрешности $\{\varepsilon_{ij}\}$ независимы, одинаково распределены и не зависят от $\{Z_{ij}\}$ и $\{f_i(\cdot)\}$, при этом $\mathbb{E}\varepsilon_{11} = 0$, $\mathbb{E}|\varepsilon_{11}|^p < \infty$ при некотором $p > 2$. Кроме того, $\mathbb{E}\sup_{t\in[0,1]} |f_1(t)|^p < \infty$, $\sup_{t\in[0,1]} \mathbb{D}f_1^2(t) < \infty$. Относительно регрессоров предполагается выполненным следующее несколько более сильное предположение, нежели используемое выше условие плотного заполнения всей совокупностью точек $\{Z_{ij}\}$ отрезка $[0,1]$.

(П5) *Все двумерные точки из набора $\{(Z_{ij_1}, Z_{ij_2}), 1 \le j_1 \ne j_2 \le m_i, i = 1,\ldots,n\}$ попарно различны и для каждого $\widetilde{N}$ существует случайное разбиение множества $[0,1]^2$ на $\tilde{l}$ измеримых по Жордану подмножеств $\{\mathcal{P}_k;\ k = 1,\ldots,\tilde{l}\}$ таких, что каждое подмножество $\mathcal{P}_k$ содержит ровно по $\tilde{r}$ двумерных точек из указанного набора, при этом $\widetilde{\delta}_{\tilde{l}} = \max_{k\le\tilde{l}} d(\mathcal{P}_k) \xrightarrow{p} 0$, где $d(A) = \sup_{\boldsymbol{x},\boldsymbol{y}\in A} \|\boldsymbol{x} - \boldsymbol{y}\|$ и $\|\cdot\|$ — супремальная норма в $\mathbb{R}^2$.*

При $k = 1,\ldots,\tilde{l}$ введем обозначения

$$\widetilde{H}_k = \big\{\big((i,j_1),(i,j_2)\big):\ \big(Z_{ij_1}, Z_{ij_2}\big) \in \mathcal{P}_k\big\}, \quad \widetilde{X}_k = \tilde{r}^{-1} \sum\nolimits_{((i,j_1),(i,j_2))\in\widetilde{H}_k} X_{ij_1}X_{ij_2}.$$

Оценку для $\varphi(t,s)$ зададим равенством

$$\widehat{\varphi}_1(t,s)=\frac{\sum_{k=1}^{\tilde{l}}\widetilde{X}_k K_h(t-Z_{i_kj_{1k}})K_h(s-Z_{i_kj_{2k}})\Lambda_2(\mathcal{P}_k)}{\sum_{k=1}^{\tilde{l}}K_h(t-Z_{i_kj_{1k}})K_h(s-Z_{i_kj_{2k}})\Lambda_2(\mathcal{P}_k)},$$

где $\left(Z_{i_kj_{1k}},Z_{i_kj_{2k}}\right)$ — произвольная фиксированная точка, принадлежащую множеству $\mathcal{P}_k$. Отметим, что при построении оценки $\widehat{\varphi}_1(t,s)$ существенно используются методы ядерного сглаживания из раздела 1.3. При выполнении вышеприведенных условий справедлива

Теорема 13 (Следствие 1.18). *Пусть справедливы соотношения*

$$h\to 0,\quad \tilde{r}^{-p/2}h^{-(p+2)}\mathbb{E}\widetilde{\delta}_{\tilde{l}}^{p}\to 0,\quad \tilde{l}/\tilde{r}\to 0,\quad \mathbb{P}\big(\widetilde{\delta}_{\tilde{l}}>h/(8L)^2\big)\to 0.$$

Тогда $\sup_{(t,s)\in[0,1]^2}|\widehat{\varphi}_1(t,s)-\varphi(t,s)|\xrightarrow{p}0.$

Оценку для функции ковариации $\psi(t,s)$ можно определить теперь равенством $\widehat{\psi}_1(t,s)=\widehat{\varphi}_1(t,s)-\widehat{\mu}_1(t)\widehat{\mu}_1(s)$.

В более простой ситуации плотных данных в модели (6), т.е. когда $m_i=m_i(n)\to\infty$, мы следуем общепринятому подходу: сначала оцениваем случайные функции f_i по наблюдениям i-ой серии, а затем проводим соответствующие усреднения по всем сериям. Но при оценивании функций f_i, $i=1,\dots,n$, мы используем методы ядерного сглаживания, предложенные в диссертационной работе. Последнее обстоятельство и позволяет нам накладывать весьма слабые ограничения на регрессоры. Точный вид оценок и формулировки утверждений об их равномерной состоятельности мы во введении опускаем. Отметим лишь, что оценки для функций среднего и ковариации имеют следующую структуру

$$\widehat{\mu}_2(t)=\frac{1}{n}\sum_{i=1}^{n}\widehat{f}_i(t),\qquad \widehat{\psi}_2(t,s)=\frac{1}{n}\sum_{i=1}^{n}\widehat{f}_i(t)\widehat{f}_i(s)-\widehat{\mu}_2(t)\widehat{\mu}_2(s),$$

где $\widehat{f}_i(t)$ — универсальная локально–постоянная оценка для $f_i(t)$, построенная по наблюдениям i-ой серии. Условие на регрессоры, гарантирующее равномерную состоятельность этих оценок, состоит в следующем: набор регрессоров каждой из серий с высокой вероятностью образует измельчающееся разбиение области определения случайного процесса.

Разделы 1.1.7, 1.2.3 и 1.3.3 главы 1 содержат примеры компьютерного моделирования и обработки реальных данных.

Вторая глава диссертационной работы посвящена задаче оценивания конечномерных параметров моделей нелинейной регрессии. Суть проблемы, решению которой посвящен данный раздел, поясним следующим простым примером. Пусть требуется оценить одномерный параметр θ по выборке $X_1, \ldots, X_n$ в случае, когда наблюдения имеют следующую структуру:

$$X_i = f(\theta, \mathbf{z}_i) + \varepsilon_i, \qquad \mathbb{E}\varepsilon_i = 0, \quad \mathbb{E}\varepsilon_i^2 = \sigma^2/w_i(\theta), \quad i = 1, \ldots, n,$$

где f и $\{w_i\}$ — известные функции, точки $\{\mathbf{z}_i\}$ неслучайны и известны, погрешности $\{\varepsilon_i\}$ независимы. Оценка квазиправдоподобия, являющая в такой постановке неоднородной модели нелинейной регрессии наилучшей в некотором классе оценок (см., например, [122]), определяется уравнением

$$\psi_n(t) = 0 \qquad \text{при} \quad \psi_n(t) = \sum_{i=1}^{n} w_i(t) f'(t, \mathbf{z}_i)\big(X_i - f(t, \mathbf{z}_i)\big).$$

На рисунке 1 представлены графики функции $\psi_n(t)$ для трех независимых реализаций вы-

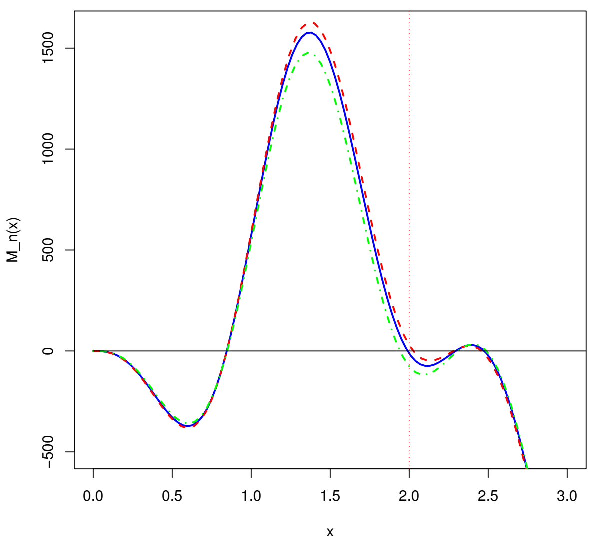


Рис. 1: Графики функции $\psi_n(t)$ для трех независимых реализаций выборки $\{X_i, i = 1, \ldots, 100\}$.

борки $\{X_i\}$ в случае, когда $f(t, \mathbf{z}_i) = w_i(t) = z_{1i}t + z_{2i}\cos t$, $\mathbf{z}_i = (z_{1i}, z_{2i})$, $\sigma^2 = 0.25$, $\theta = 1$, погрешности имеют нормальное распределение и $n = 100$. Проблема оценивания в этом примере связана с наличием нескольких корней уравнения $\psi_n(t) = 0$, определяющего оценку. Предположим, что мы смогли определить все корни этого уравнения (точнее, их приближения). Как из них выбрать корень, приближающий параметр? Надо сказать, что подобная ситуация наличия нескольких корней у тех или иных уравнений, определяющих в известном смысле оптимальные оценки, весьма типична для задач нелинейной регрессии (см., например, [266], [267], [268], а также рис. 25–26 в главе 3). Отмеченная проблема выбора состоятельного корня — это принципиальная статистическая сложность (см., например, монографии [122], [266],

[278]), поэтому естественно ожидать, что и ее решение лежит в области статистических (а не вычислительных) методов.

Во второй главе диссертационной работы предложены достаточно общие подходы к построению явных α_n-состоятельных или асимптотически нормальных оценок для широких классов моделей нелинейной регрессии. Эти подходы тесно связаны с результатами главы 1 и предполагают плотное заполнение регрессорами некоторой области. Предлагаемые оценки могут быть использованы в качестве предварительных для одношаговых процедур. По сути, если мы располагаем некоторой предварительной состоятельной оценкой, то мы получаем возможность выделить тот из корней интересующего нас уравнения, который, собственно, и приближает параметр. Так, в вышеприведенном примере одношаговая оценка $\theta_n^{**} = \theta_n^* - \psi_n(\theta_n^*)/\psi_n'(\theta_n^*)$, представляющая собой одну итерацию метода Ньютона для приближенного поиска корней уравнения $\psi_n(t) = 0$ с состоятельной оценкой θ_n^* в качестве начальной точки, в широких условиях оказывается асимптотически эквивалентной оценке квазиправдоподобия.

Одношаговые оценки в тех или иных задачах нелинейной регрессии исследовались, например, в [47], [211], [212], [228], [251], [266] и др. Но во всех известных нам работах, за исключением статьи [211], посвященной дробно–линейным моделям, существование предварительной оценки лишь постулируется. Что касается явных оценок в задачах нелинейной регрессии, то ранее такие оценки были известны, по-видимому, лишь для нескольких простых регрессионных моделей с аддитивными погрешностями (см., например, [60], [160], [236], [250], [331], [333], [339]–[341], [346]), которые, как оказалось, относятся к очень узкому классу внутренне-линейных моделей (см. подробности далее и в разделе 2.1). Так что предлагаемые в диссертации достаточно общие подходы к построению предварительных оценок параметров нелинейной регрессии, по-видимому, не имеют аналогов в литературе.

Говоря о задачах нелинейной регрессии, нельзя не упомянуть такое важное направление стохастического анализа, как планирование эксперимента. Это направление относится, в основном, к случаю, когда регрессоры неслучайны и управляемы. Имеется ряд публикаций, посвященных задачам планирования эксперимента в моделях нелинейной регрессии. В качестве ориентира укажем монографии [6], [14], [86], [204], [330], [343] и статьи [55], [90], [106], [107], [203], [205], [207], [206], [303], [304], в которых можно найти более подробные библиографические сведения в указанной области. Надо сказать, что задачи планирования эксперимента для нелинейных регрессионных моделей достаточно сложны и основная трудность таких задач заключается в зависимости асимптотической ковариационной матрицы оценок метода наименьших квадратов от истинных значений параметров. Таким образом, оптимальные решающие правила, которые минимизируют, например, те или иные функционалы от ковариационной матрицы, также зависят от истинных значениях параметров. Чтобы преодолеть эту трудность и уменьшить зависимость плана эксперимента от неизвестного параметра, существует целый ряд подходов (локально–оптимальный, последовательный, минимаксный, максиминный или байесовский). Например, в локально–оптимальных планах неизвестное значение параметра заменяется некоторым приближением, при этом эффективность такой конструкции во многом зависит от качества этого начального приближения. При последовательном подходе эксперимент разбивается на несколько этапов: на каждом этапе параметр оценивается (например, численно с помощью метода наименьших квадратов), и полученная оценка используется на последующем этапе планирования. Подчеркнем, что все подобные подходы требуют некоторой дополнительной информации о параметре (например, некоторой локализации истинного значения параметра в достаточно узкой области, знание достаточно хорошей оценки параметра или априорного распределения неизвестного параметра). В монографии [343, стр. 132] можно найти следующую рекомендацию: если некоторый функционал от информационной матрицы, участвующий в построении оптимального плана, зависит от неизвестного параметра, то можно истратить малый относительно общего числа экспериментов объем выборки на предварительную состоятельную оценку параметра, и далее в вышеупомянутом функционале параметр заменить оценкой. В этом смысле идеология планирования эксперимента в моделях нелинейной регрессии близка к идее одношагового оценивания, основанной на использовании предварительной состоятельной оценки.

Отметим, что в задачах планирования эксперимента в нелинейной регрессии число так называемых опорных точек планов нередко совпадает с размерностью параметра. Наша постановка задачи несколько иная, чем в задачах планирования эксперимента, поскольку мы допускаем и ситуацию неуправляемого эксперимента. Как и в задачах непараметрической регрессии, предлагаемые в диссертации подходы построения предварительных оценок универсальны относительно стохастической природы регрессоров: регрессоры могут быть как детерминированными, так и случайными. Одно из основных предположений, которое мы нередко будем использовать, состоит (как и в главе 1) в плотном заполнении регрессорами некоторой области. Это условие представляется естественным, например, для моделей со случайными регрессорами. Стоит отметить, что подобные модели весьма популярны, при этом вычислительные сложности поиска оценок, о которых говорилось выше, в равной степени присущи как моделям с детерминированными, так и со случайными регрессорами. Например, в ситуации случайных регрессоров широко используется метод наименьших квадратов или его модификации (см., например, [165], [130], [230], [232], [265], [283]). Так что построение тех или иных предварительных оценок (с целью обойти вычислительные сложности поиска в известном смысле оптимальных оценок) в полной мере актуально и для моделей нелинейной регрессии со случайными регрессорами.

Перейдем к краткому описанию результатов главы 2. В разделе 2.1 обсуждается понятие внутренней линейности моделей. Напомним (см., например, [60], [236]), что к *внутренне линейным* (intrinsically linear) или *внешне нелинейным* (nonintrinsically nonlinear) моделям принято относить модели регрессии, для которых регрессионное уравнение можно трансформировать к линейному. Для таких моделей явные оценки параметров зачатую могут быть построены методами линейного регрессионного анализа. Ранее считалось, что внутренне линейными могут быть только модели с мультипликативными погрешностями (см., например, [236]). В диссертации уточнено и расширено определение внутренне линейных моделей и установлено, что несколько известных моделей нелинейной регрессии с аддитивными погрешностями удовлетворяют этому определению. Показано, что известные ранее явные оценки для этих моделей, построенные исходя из тех или иных эвристических соображений, можно построить методами линейного регрессионного анализа, если под внутренней линейностью понимать новое расширенное определение. Кроме того, показано, что одна из известных интерпретаций внутренней линейности, основанная на трансформации регрессионного уравнения без учета погрешностей (см., например, монографию [10, стр. 34]), может приводить к несостоятельным оценкам. Отметим, что возможность преобразования модели регрессии с аддитивными погрешностями к линейной является скорее редким исключением, нежели правилом. Поэтому и возникает необходимость разработки подходов к построению явных оценок для собственно нелинейных моделей.

В разделах 2.2 и 2.3 предложены два подхода к построению явных оценок. Постановка задачи следующая: наблюдения $\{X_i,\ i=1,\dots,n\}$ имеют структуру

$$X_i = f(\boldsymbol{\theta}, \mathbf{z}_i) + \varepsilon_i, \qquad i = 1, \dots, n, \tag{7}$$

где f — некоторая известная функция, набор k-мерных регрессоров $\{\mathbf{z}_i\}$ состоит из наблюдаемых случайных величин в схеме серий со значениями в некотором множестве $\mathcal{P} \subset \mathbb{R}^k$, $\{\varepsilon_i\}$ — ненаблюдаемые случайные погрешности, которые также могут зависеть от n. Подчеркнем, что условия на регрессоры включают в себя и ситуацию с детерминированными регрессорами. Задача состоит в оценивании m-мерного параметра $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

В разделе 2.2 предложен способ построения оценок, непосредственно использующий непараметрические ядерные методы главы 1. Эта методика близка к методологии метода моментов. Оценку для параметра $\boldsymbol{\theta}$ в модели (7) предлагается искать как решение системы уравнений

$$f(\boldsymbol{\theta}, \mathbf{t}_j) = \widehat{f}_n(\mathbf{t}_j), \quad j = 1, \dots, m,$$

где $\widehat{f}_n(\mathbf{t})$ — непараметрическая ядерная оценка для регрессионной функции $f(\boldsymbol{\theta}_0, \mathbf{t})$ и $\boldsymbol{\theta}_0$

— истинное значение параметра $\boldsymbol{\theta}$. Набор точек $\{\mathbf{t}_j\}$ подбирается так, чтобы эта система была разрешима единственным образом и обратное отображение было бы непрерывным. По сути, предлагается приравнять значения регрессионной функции в тех или иных точках к соответствующим значениям ее состоятельной непараметрической оценки и решить эту систему уравнений относительно неизвестного m-мерного параметра. Число таких уравнений (или что то же — указанных точек) должно совпадать с размерностью m параметра $\boldsymbol{\theta}$. В качестве оценки $\widehat{f}_n(\mathbf{t})$ рекомендуется использовать универсальные ядерные оценки главы 1.

Приведем условия α_n-состоятельности оценок в случае, когда в качестве $\widehat{f}_n(\mathbf{t})$ используется универсальная локально-постоянная ядерная оценка $f^*_{n,h}(\mathbf{t})$, введенная в (4) (нормировка α_n, как правило, будет иметь порядок не больше, чем $n^{1/(k+2)}$). В приводимой далее теореме 14 предполагается, что погрешности $\{\varepsilon_i\}$ и ядро K удовлетворяют условиям, используемым в теореме 4, справедливо предположение (П2), $\mathcal{P} = [0,1]^k$ и $\|f\| = \sup_{\mathbf{t}\in[0,1]^k} |f(\mathbf{t})|$.

Теорема 14 (Следствие 2.1). *Пусть имеется непрерывное отображение $\boldsymbol{G}$ банахова пространства $(C[0,1]^k, \|\cdot\|)$ в $\mathbb{R}^m$, для которого векторная функция $\boldsymbol{g}(\boldsymbol{\theta}) \equiv \boldsymbol{G}\big(f(\boldsymbol{\theta},\cdot)\big)$ является гомеоморфизмом открытого множества $\boldsymbol{\Theta}$ на некоторую область пространства $\mathbb{R}^m$, отображение $\boldsymbol{G}$ и обратное отображение $\boldsymbol{g}^{-1}$ удовлетворяют условию Липшица в своих пространствах, для каждого $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ регрессионная функция $f(\boldsymbol{\theta},\boldsymbol{t})$ удовлетворяет условию Липшица по второму аргументу и не является постоянной функцией. Тогда оценка $\boldsymbol{\theta}^*_n = \boldsymbol{g}^{-1}\big(\boldsymbol{G}(f^*_{n,h})\big)$ определена с вероятностью, стремящейся к 1, и является α_n-состоятельной, когда $\alpha_n = o(h_n^{-1})$ и $h_n = \big(\mathbb{E}(\delta_n^{kp/2})\big)^{\frac{1}{p(k/2+1)+k}}$.*

В разделе 2.3 предложен другой подход к построению оценок для конечномерного параметра $\boldsymbol{\theta}$ в модели (7), основанный на использовании сумм определенным образом взвешенных откликов и по сути повторяющий методологию первой главы. Как правило, этот подход позволяет получать α_n-состоятельные оценки с более высокой скоростью сходимости, чем оценки первого метода. Одну из основных идей этого подхода опишем сначала в простейшей ситуации, когда параметр $\theta \in \Theta = (a,b)$ одномерный (границы a и b могут быть бесконечными с соответствующими знаками), регрессоры $\{z_i\}$ одномерны ($k=1$) и с вероятностью 1 принадлежат некоторому конечному отрезку $[c,d]$. Прежде всего при каждом фиксированном $n \geq 1$ упорядочим $z_1,\ldots,z_n$ по возрастанию: $z_{n:1} \leq \ldots \leq z_{n:n}$. Отклики и погрешности из (7), ассоциированные с порядковой статистикой $z_{n:i}$, обозначим соответственно X_{ni} и ε_{ni}. В этом случае модель (7) примет вид

$$X_{ni} = f(\theta, z_{n:i}) + \varepsilon_{ni}, \qquad i = 1,\ldots,n. \tag{8}$$

Положим $\Delta z_{ni} = z_{n:i} - z_{n:i-1}$, $z_{n:0} = c$, $z_{n:n+1} = d$.

Основным предположением является следующее (первая часть этого условия при $c = 0$ и

$d = 1$ уже фигурировала в предположении (П1) первой главы).

(П6) *Совокупность* $\{z_{ni}\}$ *с ростом* n *образует измельчающееся разбиение отрезка* $[c, d]$ *в том смысле, что* $\max_{1 \le i \le n+1} \Delta z_{ni} \xrightarrow{p} 0$. *Функция* $f(\theta, z)$ *интегрируема по Риману по* z *на отрезке* $[c, d]$, *при этом интеграл Римана* $T(t) = \int_c^d f(t, z)dz$ *является строго монотонной и непрерывной функцией.*

Предположим дополнительно, что $\sum_{i=1}^n \Delta z_{ni}\varepsilon_{ni} \xrightarrow{p} 0$. Например, в силу (П5) это условие выполнено, если при всех $n \ge 1$ с вероятностью 1 $\sup_{i \le n} \mathbb{E}_{\mathcal{F}_n}\varepsilon_i^2 \le \sigma^2$ и $\mathbb{E}_{\mathcal{F}_n}\varepsilon_i\varepsilon_j = 0$ при всех $i, j \le n$ и $i \ne j$, где константа $\sigma^2 > 0$ не зависит от n, а $\mathbb{E}_{\mathcal{F}_n}$ есть условное математическое ожидание относительно σ-алгебры $\mathcal{F}_n$, порожденной случайными величинами $\{z_i;\ i = 1, \dots, n\}$ (т.е. выполнено в классической постановке, когда $\{z_i\}$ неслучайны, а погрешности есть центрированная последовательность независимых одинаково распределенных случайных величин с конечным вторым моментом). Заметим, что с учетом (8) и (П6)

$$\sum_{i=1}^n \Delta z_{ni}X_{ni} = \sum_{i=1}^n \Delta z_{ni} f(\theta, z_{n:i}) + \sum_{i=1}^n \Delta z_{ni}\varepsilon_{ni}, \qquad \sum_{i=1}^n \Delta z_{ni} f(\theta, z_{n:i}) \xrightarrow{p} T(\theta),$$

а потому при введенных условиях оценку θ_n^* параметра θ естественно определить равенством

$$\theta_n^* = T^{-1}\Big(\sum_{i=1}^n \Delta z_{ni}X_{ni}\Big) \tag{9}$$

всюду, когда корректно задана правая часть (т.е. аргумент обратной функции T^{-1} лежит в области ее определения). Отметим, что в указанных условиях оценка θ_n^* из (9) определена с вероятностью, стремящейся к 1, и является состоятельной.

Чтобы использовать предлагаемые оценки в одношаговых процедурах, требуется бóльшая точность, нежели просто состоятельность (см. результаты главы 3). Отметим, что в широких условиях в приводимых далее теоремах 15 и 16 величина α_n имеет порядок $o(\sqrt{n})$, а величина D_n^{-1} — порядок $\sqrt{n}$.

Теорема 15 (см. следствие 2.10). *Пусть выполнено предположение* (П6), *функция* $T^{-1}(t)$ *удовлетворяет условию Гельдера с показателем* $p \in (0, 1]$ *и для некоторой последовательности* $\alpha_n \to \infty$ *справедливы соотношения*

$$\alpha_n^{1/p}\sum_{i=1}^n \Delta z_{ni}\varepsilon_{ni} \xrightarrow{p} 0, \quad \alpha_n^{1/p}\sum_{i=1}^n \omega_f(\Delta z_{ni})\Delta z_{ni} \xrightarrow{p} 0, \quad \alpha_n^{1/p}\Delta z_{nn+1} \xrightarrow{p} 0, \tag{10}$$

где $\omega_f(\delta) = \sup_{t_1, t_2: |t_1 - t_2| \le \delta} |f(\theta, t_1) - f(\theta, t_2)|$. *Тогда оценка* θ_n^* *является* α_n*-состоятельной.*

Найдены также условия асимптотической нормальности оценки θ_n^* той или иной степени общности. Если регрессоры неслучайны, погрешности $\{\varepsilon_i\}$ независимы, центрированы и имеют конечные и отличные от нуля вторые моменты, то справедливо следующее утверждение.

Теорема 16 (см. следствие 2.3). *Пусть выполнено предположение* (П6), *функция* $T(\theta)$ *непрерывно дифференцируема и* $T'(\theta) \neq 0$. *Кроме того, справедливы соотношения*

$$D_n^2 = \sum_{i=1}^n \bigl(\Delta z_{ni}\bigr)^2 \mathbb{E}\varepsilon_{ni}^2 \to 0 \qquad \textit{и} \qquad D_n^{-1} \sum_{i=1}^n \Delta z_{ni}\varepsilon_{ni} \xrightarrow{d} \mathcal{N}_1(0,1),$$

а второе и третье условия из (10) *выполнены при замене* $\alpha_n^{1/p}$ *на* D_n^{-1}. *Тогда имеет место предельное соотношение* $D_n^{-1}(\theta_n^* - \theta) \xrightarrow{d} \mathcal{N}_1\bigl(0, [T'(\theta)]^{-2}\bigr)$.

Важно подчеркнуть, что приведенная выше методология позволяет отказаться от условия плотного заполнения регрессорами *всей* области своих значений. Иногда при построении оценки следует убрать из рассмотрения некоторое количество «лишних» наблюдений, чтобы обеспечить выполнение нужных нам условий (в частности, (П6)). Конечно, мы теряем при этом часть выборочной информации, но поскольку задача состоит в построении лишь некоторой *предварительной* состоятельной оценки параметра, то данное обстоятельство не столь существенно. Например, если регрессоры образуют измельчающееся разбиение только части области своего задания (обозначим эту область «плотного» заполнения регрессорами через $A \subset \mathbb{R}$), то для построения оценки следует использовать только регрессоры, принадлежащие A, и соответствующие им отклики. Или, скажем, если функция $T(t) = \int_A f(t,z)dz$ не является строго монотонной и непрерывной, но существует подмножество $A_o \subset A$ такое, что функция $T_o(t) = \int_{A_o} f(t,z)dz$ строго монотонна и непрерывна, то разумно при построении оценки использовать лишь регрессоры, принадлежащие множеству A_o, и соответствующие им наблюдения. Отметим, что указанное множество A может быть не обязательно односвязным, набор регрессоров может разбиваться на кластеры с условием измельчения в каждом. Отметим еще, что вышеприведенные результаты распространяются на случай, когда $d = \infty$ и/или $c = -\infty$; в частности, если $d = \infty$, то требуется, чтобы $z_{n,n} \to \infty$ и $\max_{1 \le i \le n} \Delta z_{ni} \xrightarrow{p} 0$, а интеграл Римана $T(\theta) = \int_c^\infty f(t,z)dz$ понимается как несобственный.

Более общая идея построения оценки θ_n^* для одномерного параметра (иногда без условия плотного заполнения регрессорами некоторой области) состоит в том, чтобы подобрать такие $h_{ni} \ge 0$, A_n и функции $T_n(\theta)$, для которых

$$\sum_{i=1}^n h_{ni} f(\theta, z_{n:i}) - A_n - T_n(\theta) \xrightarrow{p} 0 \qquad \text{и} \qquad \sum_{i=1}^n h_{ni}\varepsilon_{ni} \xrightarrow{p} 0, \tag{11}$$

где начиная с некоторого n функции $T_n(\theta)$ с вероятностью 1 строго монотонны и непрерывны, а обратные функции $T_n^{-1}(t)$ равностепенно непрерывны. В этом случае оценка определяется равенством

$$\theta_n^* = T_n^{-1}\Bigl(\sum_{i=1}^n h_{ni} X_{ni} - A_n\Bigr). \tag{12}$$

Например, выше при определении оценки (9) мы положили $h_{ni} = \Delta z_{ni}$, $A_n = 0$ и $T_n(\theta) = T(\theta)$. Можно определить веса h_{ni} иначе: $h_{ni} = g(z_{n:i})\Delta z_{ni}$ для некоторой подходящей функции g такой, чтобы обеспечить выполнение нужных условий (в том числе и строгую монотонность и непрерывность функции $T(\theta)$, определяемой в этом случае равенством $T(\theta) = \int_c^d g(z)f(t,z)dz$). Используя конструкцию (12) с условиями (11), в ряде моделей удается оценить параметр θ при нарушении условия $\max_{1\le i\le n}\Delta z_{ni} \xrightarrow{p} 0$ в случае, когда $z_{n:n} \xrightarrow{p} \infty$ и $\max_{1\le i\le n}\Delta z_{ni}/z_{n:n} \xrightarrow{p} 0$ (например, для $z_i = i$). При этом функция $f(\theta, z)$ может быть неинтегрируемой по z на луче $[0,\infty)$.

Приведем теперь схематично вид оценок в случае многомерных параметра и регрессоров, опуская достаточные условия α_n-состоятельности или асимптотической нормальности предлагаемых оценок. Рассмотрим вначале случай, когда $\boldsymbol{\theta} \in \boldsymbol{\Theta} = \prod_{j=1}^m (a_j, b_j) \subset \mathbb{R}^m$, а случайные величины $\{z_i\}$ по-прежнему с вероятностью 1 принадлежат конечному отрезку $[c,d]$ и образуют измельчающееся разбиение этого отрезка (см. (П6)). Пусть для простоты $m = 2$. В предположении интегрируемости по Риману функции $f(\boldsymbol{\theta}, z)$ по переменной z на $[c,d]$, выберем $c' \in [c,d]$ так, чтобы гарантировать гомеоморфность двумерного отображения $\mathbf{T}(\boldsymbol{\theta}) = (T_1(\boldsymbol{\theta}), T_2(\boldsymbol{\theta}))$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, где

$$T_1(\boldsymbol{\theta}) = \int_c^{c'} f(\boldsymbol{\theta}, z)dz, \qquad T_2(\boldsymbol{\theta}) = \int_{c'}^{d} f(\boldsymbol{\theta}, z)dz.$$

И пусть дополнительно $\sum_{i=1}^{n_1}\Delta z_{ni}\varepsilon_{ni} \xrightarrow{p} 0$ и $\sum_{i=n_1+1}^{n}\Delta z_{ni}\varepsilon_{ni} \xrightarrow{p} 0$, где n_1 таково, что $z_{n:i} \in [c, c']$ при $i \le n_1$. В этом случае оценку $\boldsymbol{\theta}_n^*$ двумерного параметра $\boldsymbol{\theta}$ определим равенством

$$\boldsymbol{\theta}_n^* = \mathbf{T}^{-1}\Big(\sum_{i=1}^{n_1}\Delta z_{ni}X_{ni}, \quad \sum_{i=n_1+1}^{n}\Delta z_{ni}X_{ni}\Big).$$

Здесь мы учли, что в сделанных предположениях имеют место предельные соотношения $\sum_{i=1}^{n_1}\Delta z_{ni} f(\boldsymbol{\theta}, z_{n:i}) \xrightarrow{p} T_1(\boldsymbol{\theta})$, $\sum_{i=n_1+1}^{n}\Delta z_{ni} f(\boldsymbol{\theta}, z_{n:i}) \xrightarrow{p} T_2(\boldsymbol{\theta})$ и воспользовались тождествами

$$\sum_{i=1}^{n_1}\Delta z_{ni}X_{ni} = \sum_{i=1}^{n_1}\Delta z_{ni} f(\boldsymbol{\theta}, z_{n:i}) + \sum_{i=1}^{n_1}\Delta z_{ni}\varepsilon_{ni},$$
$$\sum_{i=n_1+1}^{n}\Delta z_{ni}X_{ni} = \sum_{i=n_1+1}^{n}\Delta z_{ni} f(\boldsymbol{\theta}, z_{n:i}) + \sum_{i=n_1+1}^{n}\Delta z_{ni}\varepsilon_{ni}.$$

Таким образом, по построению $\boldsymbol{\theta}_n^* \xrightarrow{p} \mathbf{T}^{-1}\big(T_1(\boldsymbol{\theta}), T_2(\boldsymbol{\theta})\big) \equiv \boldsymbol{\theta}$. Оценивание многомерного параметра $\boldsymbol{\theta}$ произвольной размерности происходит аналогично.

В случае многомерных регрессоров $\{\mathbf{z}_i\}$ в предлагаемом подходе оценивания используются конструкции кратных интегральных сумм Римана. Считаем, что при всех i с вероятностью 1 выполнено $\mathbf{z}_i \in \mathcal{P} \subset \mathbb{R}^k$, где множество $\mathcal{P}$ измеримо по Жордану, векторы $\{\mathbf{z}_i\}$ попарно различны и «плотно» заполняют $\mathcal{P}$ (подобное требование при $\mathcal{P} = [0,1]^k$ фигурировало в (П2)). В случае скалярного θ считаем, что $f(\theta, \mathbf{z})$ как функция k-мерного аргумента $\mathbf{z}$ интегрируема по Риману на множестве $\mathcal{P}$ и интеграл Римана $T(t) = \int_{\mathcal{P}} f(t, \mathbf{z}) d\mathbf{z}$ есть строго монотонная непрерывная функция. В предположении, что $\sum_{i=1}^n \varepsilon_i \Lambda_k(\mathcal{P}_i) \xrightarrow{p} 0$, состоятельная оценка θ_n^* задается равенством

$$\theta_n^* = T^{-1}\Big(\sum_{i=1}^n X_i \Lambda_k(\mathcal{P}_i)\Big).$$

Обоснованием выбора этой статистики является следующее соотношение

$$\sum_{i=1}^n X_i \Lambda_k(\mathcal{P}_i) = \sum_{i=1}^n f(\theta, \mathbf{z}_i) \Lambda_k(\mathcal{P}_i) + \sum_{i=1}^n \varepsilon_i \Lambda_k(\mathcal{P}_i) \xrightarrow{p} T(\theta),$$

а потому $\theta_n^* \xrightarrow{p} T^{-1}\big(T(\theta)\big) \equiv \theta$. Обобщение на случай m-мерного параметра $\boldsymbol{\theta}$ происходит согласно схеме, изложенной выше.

Таким образом, второй из предлагаемых нами подходов основан на использовании сумм определенным образом взвешенных откликов в зависимости от конфигурации k-мерных регрессоров, главная часть которых представляет собой структуру интегральных сумм кратного интеграла Римана. В итоге оценки имеют вид некоторых функций от сумм специальным образом взвешенных откликов, при этом веса определяются лишь конфигурацией набора регрессоров и не зависят от их распределений. В случае одномерных, вообще говоря, случайных регрессоров оценка представляет собой некоторую функцию от суммы взвешенных откликов, ассоциированных с порядковыми статистиками, построенными по набору регрессоров, при этом в качестве весов можно брать спейсинги вариационного ряда, построенного по выборке этих самых регрессоров. Стоит отметить, что предлагаемый нами подход отчасти близок к методологии метода моментов. Но при этом лишь при одном условии «плотного» заполнения регрессорами некоторой области устанавливается факт сходимости к кратному риманову интегралу, который *не зависит от распределений регрессоров*. В этом состоит принципиальное отличие от классического метода моментов, в котором, скажем, в случае одномерных независимых одинаково распределенных наблюдений (регрессоров) с *известным* распределением интегральный функционал представляет собой математическое ожидание регрессионной функции относительно распределения регрессоров (см. замечание 2.24 в разделе 2.3.7). Так что интегральные функционалы метода моментов (среднее регрессионной функции по распределению регрессоров) и предлагаемого нами подхода (среднее, с точностью до константы, регрессионной функции по мере Лебега) принципиально различаются.

Третья глава диссертационной работы посвящена асимптотическому анализу одношаговых оценок, связанных с M-оцениванием. Пусть $X_1, \ldots, X_n$ — не обязательно независимые и одинаково распределенные наблюдения произвольной природы, распределения которых зависят от неизвестного параметра $\theta \in \Theta$ (чтобы упростить изложение, ограничимся во введении лишь случаем, когда $\Theta \subset \mathbb{R}$). Задача состоит в оценивании этого параметра по наблюдениям $X_1, \ldots, X_n$. Одним из общих методов получения оценок является M-оценивание (см., например, монографии [325], [350], [38], [94], [149] [235], [254], [278], [279], [281] и др.), которое нередко сводится к поиску корней уравнения

$$\sum_{i=1}^{n} \psi_i(t, X_i) = 0 \qquad (13)$$

для того или иного набора функций $\{\psi_i(t,x)\}$ с условием $\mathbb{E}\psi_i(\theta, X_i) = 0$ при всех i и $\theta \in \Theta$ (в дальнейшем символом θ мы будем обозначать истинное значение параметра, при этом зависимость от него в операторах усреднения $\mathbb{E}$ и вероятностях $\mathbb{P}$ указываться не будет).

Лишь в редких случаях мы имеем дело с корнями уравнения (13) как с обычными статистиками, т.е. измеримыми функциями от выборки. Корни уравнения (13) вычисляются при *фиксированной* выборке, при этом уравнение (13) может иметь случайное число корней (различное число корней для различных выборок). Например, рисунок 26 (см. модель 5) в подразделе 3.4.2 главы 3 иллюстрирует ситуацию, когда для одной выборки уравнение (13) имеет один корень, для другой – пять корней, а для третьей – семь (см. также модель 8). Что представляют собой корни с вероятностной точки зрения? Они могут быть определены как измеримые отображения лишь *части* (измеримой) выборочного пространства (возможно, очень незначительной по вероятностной мере). Именно такая ситуация реализуется, например, в случае оценивания параметра сдвига распределения Коши, когда мы исследуем производную логарифмической функции правдоподобия. В этом случае уравнение (13) — это алгебраическое уравнение степени $2n-1$. Количество интересующих нас *действительных* корней этого уравнения есть случайная величина $1+2\zeta_n$, при этом неотрицательная случайная величина ζ_n в пределе имеет распределение Пуассона с параметром $1/\pi$ (см. [237]). Иначе говоря, в этом примере при всех достаточно больших n лишь один корень (совпадающий с оценкой максимального правдоподобия) будет вещественным с вероятностью 1.

С точки зрения оценивания параметра естественно рассматривать лишь те из корней уравнения (13), которые удовлетворяют этому уравнению на множестве выборок асимптотически полной меры (на оставшейся части выборочного пространства значения этих функций можно доопределять произвольным образом). Таким образом, M-оценки разумно определить как статистики $\widetilde{\theta}_n = \widetilde{\theta}_n(X_1, \ldots, X_n)$, которые на множестве асимптотически полной меры являются решениями уравнения (13), т.е. $\mathbb{P}\Big(\sum\limits_{i=1}^{n}\psi_i(\widetilde{\theta}_n, X_i)=0\Big)=1-\delta_n$ и $\delta_n\to 0$ (аналогичное определение можно найти, например, в [8]; см. также пример 3.2 из раздела 3.1.6). Подчеркнем, что обычно функции $\{\psi_i(t,x)\}$ в той или иной статистической постановке выбираются таким образом, чтобы обеспечить желаемые свойства соответствующих M-оценок. Пусть, например, $X_i=f_i(\theta)+\varepsilon_i$, $\mathbb{E}\varepsilon_i=0$ и $\mathbb{E}\varepsilon_i=\sigma^2 w_i^{-1}(\theta)$ при всех i, где функции $\{f_i(\cdot)\}$ и $\{w_i(\cdot)\}$ известны, а параметр σ^2 может быть неизвестным. Оценки квазиправдоподобия (оптимальные в некотором классе; см., например, [122], а также раздел 3.4.1) определяются здесь уравнением (13) при $\psi_i(t,X_i)=w_i(t)f_i'(t)(X_i-f_i(t))$.

Впервые M-оценки были введены П. Хьюбером [132] в качестве обобщения оценок максимального правдоподобия в случае однородной выборки (в [133] — для случая разнораспределенных наблюдений). С одной стороны, вопросы существования и асимптотические свойства M-оценок при той или иной общности хорошо изучены (см., например, [325], [7], [120], [122], [124], [178], [179], [226], [227], [278], [302] и библиографические ссылки там же, а также в [228]). В частности, при некоторых условиях регулярности (см., например, [7]) M-оценки асимптотически нормальны:

$$Q_{n,\theta}\big(\widetilde{\theta}_n-\theta\big)\xrightarrow{d}\mathcal{N}_1(0,1), \tag{14}$$

где величины $Q_{n,\theta}$ мы определим далее. С другой стороны, как уже отмечалось выше, вычисление M-оценок может быть связано не только с техническими трудностями, но и с принципиальными: если уравнение (13) имеет несколько корней, то возникает проблема выбора состоятельного корня. Например, какой из корней уравнения (13) приближает параметр на рисунках 25 и 26, приведенных в разделе 3.4.2 главы 3? Некоторые обсуждения, связанные с этой проблематикой, можно найти, например, в [122], [266]–[268]).

В диссертации в первую очередь изучаются одношаговые M-оценки вида

$$\theta_n^{**}=\theta_n^*-\sum_{i=1}^{n}\psi_i(\theta_n^*,X_i)\left(\sum_{i=1}^{n}\psi_i'(\theta_n^*,X_i)\right)^{-1}, \tag{15}$$

где θ_n^* — некоторая предварительная оценка для θ, приближающая параметр с нужной нам скоростью сходимости, а через $\psi_i'(t,X_i)$ обозначены производные этих функций по первому аргументу. Оценка θ_n^{**} представляет собой *одну итерацию* метода Ньютона с начальной точкой $t=\theta_n^*$ для приближенного вычисления одного из корней уравнения (13). В силу состоятельности θ_n^* этот корень в широких условиях будет ближайшим корнем к θ (и этот корень, тем не менее, может не быть M-оценкой; см. комментарии далее). Оценка θ_n^{**} является одним из вариантов так называемых *одношаговых оценок.*

В третьей главе диссертационной работы получены весьма общие условия для выполнения предельного соотношения

$$Q_{n,\theta}\big(\theta_n^{**}-\theta\big) \xrightarrow{d} \mathcal{N}_1(0,1). \tag{16}$$

Таким образом, одношаговая M-оценка θ_n^{**} асимптотически нормальна с той же асимптотической дисперсией, что и M-оценка $\widetilde{\theta}_n$ со свойством (14). Более того, достаточные условия для сходимости (16) не влекут выполнение (14) (см. замечание 3.6 и комментарии далее).

Приведем обзор известных ранее результатов, связанных с одношаговыми оценками вида (15) и некоторыми аналогами этой оценки (см. приводимую далее оценку (18)). В случае *независимых и одинаково распределенных наблюдений* исследования таких оценок (в частности, оценок Фишера, определенных далее в (22)) содержатся в ряде монографий и статей (см., например, [138], [153], [168], [169], [254], [278], [280], [325], [332], [337], [338], [350] и ссылки там же). Условия асимптотической нормальности той или иной степени общности можно найти в [168], [169], [254], [278], [332], [337], [338]. Помимо задачи оценивания параметра, рассматривается и сближение одношаговых M-оценок (а также их k-шаговых версий) и состоятельных M-оценок. Так, асимптотическая эквивалентность $\sqrt{n}(\widetilde{\theta}_n-\theta_n^{**}) \xrightarrow{p} 0$ одношаговых оценок Фишера и состоятельных оценок максимального правдоподобия доказывается, например, в [325] (см. также [138] и ссылки в [151]). В [138] и [280] установлено, что эта скорость сходимости при некоторых дополнительных предположениях может быть улучшена. В частности, там доказано, что $n\big(\theta_n^{**}-\widetilde{\theta}_n\big)=O_p(1)$. В [153] найдено предельное распределение для $n\big(\theta_n^{**}-\widetilde{\theta}_n\big)$, не являющееся нормальным.

Известно, что предварительная оценка, используемая в одношаговых процедурах, может иметь различную точность. Например, в [152] установлено следующее: если выполнено $n^{\tau}|\theta_n^*-\theta|=O_p(1)$ при $\tau\in(1/4,1/2]$, то $\widetilde{\theta}_n^{**}-\widetilde{\theta}_n=O_p(n^{-2\tau})$, где θ_n^{**} — одношаговая оценка из [19], а $\widetilde{\theta}_n$ — M-оценка из [133]. Но практически во всех указанных выше работах, связанных с одношаговыми оценками вида (15), в качестве предварительной оценки θ_n^* рассматриваются либо асимптотически нормальные оценки, либо чуть более широкий класс так называемых $\sqrt{n}$*-ограниченных* оценок θ_n^*, т.е. оценок, удовлетворяющих соотношению $\sqrt{n}(\theta_n^*-\theta)=O_p(1)$. Исключение составляют монография [332] и работы [337], [338]. В частности, в [332] (теорема 5.5.4) при доказательстве асимптотической нормальности одношаговой оценки Фишера предполагается выполненным условие $n^{1/4}$-состоятельности оценки θ_n^*, т.е. сходимость $n^{1/4}(\theta_n^*-\theta) \xrightarrow{p} 0$. В [337] и [338] асимптотическая нормальность одношаговых оценок доказана в широких условиях на точность предварительных оценок (от $\sqrt{n}$-ограниченных до $n^{1/4}$-состоятельных). Отметим, что условие $n^{1/4}$-состоятельности предварительной оценки является минимальным достаточным ограничением на точность θ_n^* для асимптотической нормальности одношаговых M-оценок (см., например, [337], [338]) и необходимым для асимптотической эквивалентности одношаговой M-оценки θ_n^{**} и состоятельной M-оценки $\widetilde{\theta}_n$. Отметим еще, что методологически параграф 3.1 диссертации близок к работе [338], поскольку некоторые утверждения указанного раздела диссертационной работы являются обобщениями результатов из [338] на случай разнораспределенных и не обязательно независимых наблюдений и многомерного параметра.

В случае разнораспределенных и независимых выборочных данных некоторый вариант условий асимптотической нормальности одношаговых оценок (15) приведен в монографии [267] (стр. 135, предложение 5.7) и статье [47], при этом рассматривается лишь частный вариант нормировок и точности предварительной оценки (условие $\sqrt{n}$-ограниченности). Другие исследования асимптотического поведения одношаговых оценок (15) (как и рассматриваемых далее одношаговых оценок) в случае неоднородных выборок автору не известны.

Перейдем теперь к результатам третьей главы. В разделе 3.1 исследуются асимптотические свойства оценки θ_n^{**}, определенной в (15). В частности, доказано предельное соотношение (16). Чтобы сформулировать указанное утверждение об асимптотической нормальности θ_n^{**} (для простоты в случае одномерного параметра), нам потребуется следующее условие.

(П7). *Даны наблюдения $X_1,\ X_2,\ldots,X_n$ со значениями в произвольном измеримом пространстве $\mathcal{X}$ и распределениями $\mathcal{L}_{1,\theta},\ \mathcal{L}_{2,\theta},\ldots,\mathcal{L}_{n,\theta}$, зависящими от параметра $\theta\in\Theta\subset\mathbb{R}$, где Θ — открытое множество (распределения могут зависеть от n и дополнительного параметра $\sigma\in\Xi$ произвольной природы). При любом i на множестве $\Theta\times\mathcal{X}$ заданы зависящие, вообще говоря, от n измеримые функции $\psi_i(t,x)$ и $\psi_i'(t,x)$, при этом для каждого интервала (t_1,t_2), целиком лежащего в Θ, при любом $x\in\mathcal{X}$ выполнено $\psi_i(t_2,x)-\psi_i(t_1,x)=\int\limits_{t_1}^{t_2}\psi_i'(t,x)dt$ и справедливы соотношения $\mathbb{E}\psi_i(\theta,X_i)=\ 0,\ \mathbb{E}\psi_i^2(\theta,X_i)<\ \infty,\ \mathbb{E}|\psi_i'(\theta,X_i)|<\ \infty$. Кроме того, начиная с некоторого n выполнено $I_{n,\theta}=\mathbb{E}\Big(\sum\limits_{i=1}^n\psi_i(\theta,X_i)\Big)^2>0$ и $J_{n,\theta}=\sum\limits_{i=1}^n\mathbb{E}\psi_i'(\theta,X_i)\neq 0$, и имеют место предельные соотношения*

$$J_{n,\theta}^{-1}\sum_{i=1}^n\psi_i'(\theta,X_i)\xrightarrow{p}1,\qquad I_{n,\theta}^{-1/2}\sum_{i=1}^n\psi_i(\theta,X_i)\xrightarrow{d}\mathcal{N}_1(0,1),$$
$$I_{n,\theta}^{-1/2}|J_{n,\theta}|\to\infty,\qquad \limsup\mathcal{E}_{n,\theta}(\delta)\to 0\quad \textit{при}\quad \delta\to 0,$$

где $\mathcal{E}_{n,\theta}(\delta)=|J_{n,\theta}|^{-1}\sum_{i=1}^n\mathbb{E}\sup\limits_{t\in\Theta:\ |t-\theta|\le\delta}\big|\psi_i'(t,X_i)-\psi_i'(\theta,X_i)\big|$.

Теорема 17 (см. теорему 3.1). *Пусть выполнено условие* (П7) и оценка θ_n^* такова, что

$$I_{n,\theta}^{-1/2}|J_{n,\theta}||\theta_n^*-\theta|\mathcal{E}_{n,\theta}(|\theta_n^*-\theta|)\xrightarrow{p}0.\tag{17}$$

Тогда одношаговая M-оценка θ_n^{**} определена с вероятностью, стремящейся к 1, и имеет место сходимость (16) при $Q_{n,\theta}=I_{n,\theta}^{-1/2}J_{n,\theta}$.

Положим $Q_n^* = \left(\sum_{i=1}^n \psi_i^2(\theta_n^{**}, X_i)\right)^{-1/2} \sum_{i=1}^n \psi_i'(\theta_n^*, X_i)$.

Теорема 18 (Теорема 3.3). *Пусть выполнены условия* (П7), (17) *и*

$$J_{n,\theta}^{-2} \sum\nolimits_{i=1}^n \left(\psi_i'(\theta, X_i)\right)^2 \xrightarrow{p} 0, \qquad I_{n,\theta}^{-1} \sum\nolimits_{i=1}^n \psi_i^2(\theta, X_i) \xrightarrow{p} 1.$$

Тогда $Q_n^* \bigl(\theta_n^{**} - \theta\bigr) \xrightarrow{d} \mathcal{N}_1(0,1)$.

Утверждение теоремы 18 может быть полезным при построении доверительных интервалов и проверке гипотез, поскольку нормирующий множитель Q_n^* является статистикой. Отметим еще, что условие (17) включает в себя широкий спектр ограничений на точность предварительной оценки θ_n^*. Это условие связывает гладкость функций $\{\psi_i(\cdot,\cdot)\}$, определяющих одношаговую M-оценку, и скорость сближения предварительной оценки θ_n^* и параметра θ, которые нужны для асимптотической нормальности одношаговых M-оценок, при этом точность θ_n^* и гладкость функций $\{\psi_i(\cdot,\cdot)\}$ в известном смысле обратно пропорциональны друг другу. В частности, условие (17) выполнено, когда величина $\mathcal{E}_{n,\theta}(\delta)$ и оценка θ_n^* одновременно (для некоторого α из указанной области) удовлетворяют одному из следующих двух предположений:

1) при некотором $0 \le \alpha < 1$ выполнено
$\limsup \mathcal{E}_{n,\theta}(\delta) = o(\delta^\alpha)$ при $\delta \to 0$ и $\left(I_{n,\theta}^{-1/2} |J_{n,\theta}|\right)^{1/(1+\alpha)} |\theta_n^* - \theta| = O_p(1)$;

2) при некотором $0 < \alpha \le 1$ выполнено
$\limsup \mathcal{E}_{n,\theta}(\delta) = O(\delta^\alpha)$ при $\delta \to 0$ и $\left(I_{n,\theta}^{-1/2} |J_{n,\theta}|\right)^{1/(1+\alpha)} |\theta_n^* - \theta| \xrightarrow{p} 0$.

В частности, в двух крайних случаях получаем следующие ограничения на точность θ_n^*: $|J_{n,\theta}| I_{n,\theta}^{-1/2} |\theta_n^* - \theta| = O_p(1)$ при $\alpha = 0$ и $|J_{n,\theta}|^{1/2} I_{n,\theta}^{-1/4} |\theta_n^* - \theta| \xrightarrow{p} 0$ при $\alpha = 1$. Таким образом, если $|J_{n,\theta}| I_{n,\theta}^{-1/2}$ имеет порядок $\sqrt{n}$ (в частности, в случае независимых одинаково распределенных наблюдений), то указанные ограничения на точность θ_n^* есть либо предположение о $\sqrt{n}$-ограниченности, либо об $n^{1/4}$-состоятельности предварительной оценки.

В разделе 3.1 исследуется также вопрос о минимальном достаточном ограничении на точность предварительной оценки θ_n^*. Доказано (см. теорему 3.5), что при широких ограничениях указанное выше условие $\left(J_{n,\theta}^2 I_{n,\theta}^{-1}\right)^{1/4} |\theta_n^* - \theta| \xrightarrow{p} 0$ является необходимым для справедливости соотношения $Q_{n,\theta}\bigl(\theta_n^{**} - \theta\bigr) + \zeta_n \xrightarrow{p} 0$, где $\zeta_n \xrightarrow{d} \mathcal{N}_1(0,1)$. Таким образом, данное ограничение на точность θ_n^* по сути необходимо для доказательства асимптотической нормальности (16) одношаговой M-оценки θ_n^{**}.

Одношаговым оценкам присуще еще одно интересное и несколько неожиданное свойство: они могут обладать некоторыми нужными свойствами *вне зависимости от поведения оценок, определяемых соответствующим уравнением*. Подобный факт отмечается, например, в [19], [82], [168], [169], [228], [337], [338]. В частности, сходимость (16) может иметь место, в то время как для соответствующей M-оценки $\widetilde{\theta}_n$ соотношение (14) не выполнено. По-видимому, впервые этот эффект был отмечен Л. Ле Камом [169] в случае оценок максимального правдоподобия для однородных выборок. Чем объясняется этот эффект? Оказывается, при весьма слабых ограничениях (не гарантирующих состоятельность M-оценок из некоторого набора) в окрестности θ с вероятностью, стремящейся к 1, существует единственный корень $\widetilde{\theta}_n(\theta)$ уравнения (13), имеющий требуемую точность. Этот корень уравнения $\widetilde{\theta}_n(\theta)$, не являющийся статистикой, собственно и приближают одношаговые M-оценки. Если же существует состоятельная M-оценка $\widetilde{\theta}_n$, то $\mathbb{P}\big(\widetilde{\theta}_n(\theta) \neq \widetilde{\theta}_n\big) \to 0$ и одношаговые M-оценки аппроксимируют состоятельную M-оценку. В работе исследуется асимптотическое поведение ближайшего к θ корня $\widetilde{\theta}_n(\theta)$ уравнения (13) и некоторые смежные вопросы. В частности, доказано (см. теорему 3.6), что при некоторых весьма слабых ограничениях имеет место предельное соотношение $Q_{n,\theta}\big(\widetilde{\theta}_n(\theta) - \theta\big) \xrightarrow{d} \mathcal{N}_1(0,1)$. В качестве иллюстрации вышесказанного в работе также построен пример, когда имеются две M-оценки, они не состоятельны, в то время как ближайшее к θ решение $\widetilde{\theta}_n(\theta)$ соответствующего уравнения почти наверное сходится к θ (см. пример 3.2 из раздела 3.1.6).

В диссертационной работе исследуются еще два типа одношаговых оценок. Во-первых, рассматривается следующая модификация оценки θ_n^{**}:

$$\widehat{\theta}_n^{**} = \theta_n^* - J_n^{*-1} \sum_{i=1}^{n} \psi_i(\theta_n^*, X_i), \tag{18}$$

где статистика J_n^* такова, что $J_n^* J_{n,\theta}^{-1} \xrightarrow{p} 1$. Выбор J_n^* может быть очевиден, например, если $J_{n,\theta} = J_{n,\theta,\sigma} \equiv J_n(\theta)$. Здесь проще всего положить $J_n^* = J_n(\theta_n^*)$. Подобная ситуация реализуется, например, в случае построения одношаговых приближений для оценок квазиправдоподобия и метода наименьших квадратов в задачах нелинейной регрессии.

Во-вторых, исследуются одношаговые взвешенные M-оценки. Эти оценки рассчитаны на ситуацию, когда функции $\{\psi_i(t, X_i)\}$ из (13) имеют структуру

$$\psi_i(t, X_i) = h_i(t)\widetilde{\psi}_i(t, X_i) \tag{19}$$

для некоторых функций $h_i(t)$. Подобная факторизация для функций $\{\psi_i(t, X_i)\}$ естественным образом возникает в некоторых статистических задачах. Например, оценка метода наименьших квадратов в задаче нелинейной регрессии при широких ограничениях определяется уравнением $\sum_{i=1}^n f_i'(t)\big(X_i - f_i(t)\big) = 0$, где $\{f_i(\cdot)\}$ — известные функции. Тем самым, можно положить $h_i(t) = f_i'(t)$ и $\widetilde{\psi}_i(t, X_i) = X_i - f_i(t)$.

Если функции $h_i(t)$ достаточно регулярные, то в окрестности истинного значения параметра θ функции $\psi_i(t, X_i)$ вида (19) и $h_i(\theta_n^*)\widetilde{\psi}_i(t, X_i)$ будут в известном смысле близкими. Это позволяет надеяться, что в этой окрестности будут близки корни уравнения (13) в случае

(19) и корни уравнения

$$\sum_{i=1}^{n} h_i(\theta_n^*)\widetilde{\psi}_i(t, X_i) = 0 \qquad (20)$$

(см. также рисунок 30 в разделе 3.4.2 главы 2). Используя один шаг метода Ньютона с начальной точкой $t = \theta_n^*$ для приближенного поиска решения уравнения (20), получаем так называемую одношаговую *взвешенную* M-оценку

$$\theta_{n,h}^{**} = \theta_n^* - \sum_{i=1}^{n} h_i(\theta_n^*)\widetilde{\psi}_i(\theta_n^*, X_i)\Big(\sum_{i=1}^{n} h_i(\theta_n^*)\widetilde{\psi}_i'(\theta_n^*, X_i)\Big)^{-1}.$$

В диссертационной работе найдены условия для выполнения предельных соотношений $Q_{n,\theta}\big(\widehat{\theta}_n^{**} - \theta\big) \xrightarrow{d} \mathcal{N}_1(0,1)$ и $Q_{n,\theta}\big(\theta_{n,h}^{**} - \theta\big) \xrightarrow{d} \mathcal{N}_1(0,1)$. Таким образом, все три варианта введенных выше одношаговых оценок эквивалентны в смысле асимптотической точности (см. также (16)), но та или иная оценка может быть предпочтительнее. Например, в случае (19) оценка θ_n^{**} из (15) имеет вид

$$\theta_n^{**} = \theta_n^* - \frac{\sum_{i=1}^{n} h_i(\theta_n^*)\widetilde{\psi}_i(\theta_n^*, X_i)}{\sum_{i=1}^{n} h_i(\theta_n^*)\widetilde{\psi}_i'(\theta_n^*, X_i) + \sum_{i=1}^{n} h_i'(\theta_n^*)\widetilde{\psi}_i(\theta_n^*, X_i)}.$$

Тем самым, оценки $\theta_{n,h}^{**}$ могут быть проще, нежели θ_n^{**}, при этом конструкция $\theta_{n,h}^{**}$ не предполагает дифференцируемость функций $h_i(t)$. Для отмеченной выше задачи приближения оценки метода наименьших квадратов в нелинейной регрессии последнее означает меньшие условия на гладкость функций $\{f_i(\theta)\}$. Отметим, что для двух из трех вариантов одношаговых оценок, указанных выше, условия асимптотической нормальности получены и в случае многомерного основного параметра.

Раздел 3.3 посвящен одношаговому оцениванию по однородной выборке из однопараметрического семейства распределений в случае, когда предварительная оценка сближается с параметром достаточно медленно. Этот раздел применительно к задачам регрессии может включать лишь очень частный случай, когда регрессоры и погрешности независимы и одинаково распределены, а их распределения известны. Однако, на наш взгляд, предлагаемый подход улучшения точности медленно сходящихся предварительных оценок может быть перенесен и на более общие ситуации (например, те или иные варианты разнораспределенных наблюдений). Итак, рассматривается задача оценивания числового параметра $\theta \in \Theta \subset \mathbb{R}$ по выборке объема n из последовательности $X_1, X_2, \ldots$ независимых одинаково распределенных случайных величин произвольной природы, распределение которых имеет плотность $f_\theta(\cdot)$ относительно некоторой σ-конечной меры μ. В широких условиях некоторым эталоном точности при оценивании неизвестного параметра θ принято считать оценку максимального правдоподобия $\widetilde{\theta}_n$. Хорошо известно, что при выполнении некоторых условий регулярности имеет место следующая сходимость:

$$\sqrt{n}\big(\widetilde{\theta}_n - \theta\big) \xrightarrow{d} \mathcal{N}_1(0, I^{-1}(\theta)), \tag{21}$$

где $I(t)$ — информация Фишера, соответствующая плотности $f_t(x)$. Оценки, удовлетворяющие соотношению (21), будем называть *асимптотически эффективными*. Стоит отметить, что оценка $\widetilde{\theta}_n$ вычисляется в явном виде лишь для ограниченного круга распределений, при этом поиск приближений для $\widetilde{\theta}_n$ итерационными процедурами может быть дополнительно затруднен наличием большого числа локальных максимумов у логарифмической функции правдоподобия. С вычислительной точки зрения более удобными, по сравнению $\widetilde{\theta}_n$, могут быть одношаговые оценки Фишера, определяемые равенствами

$$\theta^{**}_{n1} = \theta^*_n - L'_n(\theta^*_n)/L''_n(\theta^*_n) \qquad \text{или} \qquad \theta^{**}_{n2} = \theta^*_n + L'_n(\theta^*_n)(nI(\theta^*_n))^{-1}, \tag{22}$$

где θ^*_n — некоторая предварительная оценка параметра θ (как правило, оценка метода моментов), $L_n(t) = \sum_{i=1}^n \ln f_t(X_i)$ — логарифмическая функция правдоподобия. Одношаговые оценки Фишера асимптотически эффективны, как известно, в случае, когда предварительная оценка θ^*_n является n^β-состоятельной (т.е. $n^\beta(\theta^*_n - \theta) \xrightarrow{p} 0$ с ростом n) при $\beta \geq 1/4$. Но задача поиска явной асимптотически эффективной оценки может быть актуальной и в ситуации, когда скорость сближения предварительной оценки и параметра оказывается медленнее, чем $n^{-1/4}$. Например, если плотность $f_\theta(x)$ элементов выборки есть $f_\theta(x) = \frac{(h+1)}{\theta(1+x/\theta)^{2+h}}$ при $h \in \big(0, 1/3\big]$ и $x \geq 0$, то оценка по методу моментов $\theta^*_n = h\sum\limits_{i=1}^{n} X_i/n$ является лишь n^β-состоятельной при $\beta < h/(h+1) \in (0, 1/4]$.

Известно, что в случае медленно сходящей предварительной оценки ее точность можно улучшить многократным применением метода Ньютона (см. подробности в разделе 3.1.8 и замечании 3.18). В диссертационной работе предлагается другой подход: приведен алгоритм построения новых одношаговых оценок, специально ориентированных на медленно сближающиеся с параметром предварительные оценки. Эти оценки «одношаговые» в том смысле, что они позволяют *за один шаг* n^β-состоятельную оценку при $\beta < 1/4$ улучшить до асимптотически эффективной, но шаг определяется уже не методом Ньютона, а несколько иным преобразованием. Иными словами, нам хотелось бы понять, как можно «подправить», скажем, информацию Фишера в (22), чтобы существенно улучшить точность соответствующей оценки. Опишем кратко структуру этих новых оценок. Предположим, что при некотором натуральном k нам известна $n^{1/2(k+2)}$-состоятельная оценка $\theta^*_{n,k}$ параметра θ, т.е.

$$n^{1/2(k+2)}(\theta^*_{n,k} - \theta) \xrightarrow{p} 0.$$

Одношаговая асимптотически эффективная оценка задается соотношением

$$\theta^{**}_{n,k} = \theta^*_{n,k} + L'_n(\theta^*_{n,k}) \left(\sum_{i=1}^{n} \lambda_k(\theta^*_{n,k}, X_i) \right)^{-1}.$$

Здесь функция $\lambda_k(\theta, x)$ имеет структуру

$$\lambda_k(\theta, x) = I(\theta) + c_{k1,\theta} \frac{f'_\theta(x)}{f_\theta(x)} + c_{k2,\theta} \frac{f''_\theta(x)}{f_\theta(x)} + \ldots + c_{kk,\theta} \frac{f^{(k)}_\theta(x)}{f_\theta(x)},$$

где $c_{k1,\theta}, ..., c_{kk,\theta}$ — решения некоторой системы линейных уравнений, дифференцирование производится по переменной θ и через $f^{(k)}_t(x)$ обозначена k-я производная плотности $f_t(x)$ по t. В частности, при $k = 1$ (т.е. в случае $n^{1/6}$- состоятельности предварительной оценки) одношаговая оценка определяется соотношением

$$\theta^{**}_{n,1} = \theta^*_{n,1} + \frac{L'_n(\theta^*_{n,1})}{nI(\theta^*_{n,1}) + L'_n(\theta^*_{n,1})c(\theta^*_{n,1})} \quad \text{при} \quad c(\theta) = \frac{1}{2I(\theta)} \int \frac{f'_\theta(x) f''_\theta(x)}{f_\theta(x)} \mu(dx).$$

Доказано (см. теоремы 3.10, 3.11 и следствие 3.2), что в широких условиях одношаговые оценки $\theta^{**}_{n,k}$ асимптотически эффективны. Как видно из приведенных выше определений, оценки $\theta^{**}_{n,k}$ отличаются от одношаговой оценки Фишера θ^{**}_{n2} наличием некоторых дополнительных слагаемых в знаменателе соответствующей дроби, при этом количество этих слагаемых определяется скоростью сходимости предварительной оценки $\theta^*_{n,k}$ (т.е. задается значением k). Кроме того, от значения k зависят и коэффициенты $c_{k1,\theta}, ..., c_{kk,\theta}$, задающие эти аддитивные добавки. Тем самым, с ростом k не только добавляется соответствующее количество слагаемых, но при каждом новом k заново пересчитываются и коэффициенты $c_{k1,\theta}, ..., c_{kk,\theta}$.

Заключительный раздел 3.4 главы 3 диссертации посвящен компьютерному моделированию, связанному с одношаговыми процедурами приближения оценок квазиправдоподобия и метода наименьших квадратов в задачах нелинейной регрессии.

1 Непараметрическая регрессия

В этой главе в первую очередь (в разделах 1.1-1.4) рассматривается следующая модель непараметрической регрессии.

($\mathbf{M}_1$) *Даны наблюдения $X_1, \ldots, X_n$, которые представимы в виде*

$$X_i = f(\mathbf{z}_i) + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{23}$$

где неизвестная (возможно, случайная) вещественнозначная функция $f(\mathbf{t})$, $\mathbf{t} \in \mathcal{P}$, с вероятностью 1 непрерывна на некотором компакте $\mathcal{P} \subset \mathbb{R}^k$ (значение $k \geq 1$ фиксировано); $\{\mathbf{z}_i;\ i = 1, \ldots, n\}$ есть набор наблюдаемых k-мерных случайных величин с, вообще говоря, неизвестными распределениями со значениями в $\mathcal{P}$, не обязательно независимых или одинаково распределенных, при этом случайные величины $\{\mathbf{z}_i;\ i = 1, \ldots, n\}$ могут зависеть от n.

Одна из наших задач состоит в том, чтобы при минимальных ограничениях на регрессоры $\{\mathbf{z}_i\}$ по наблюдениям $\{(\mathbf{z}_i, X_i);\ i = 1, \ldots, n\}$ построить равномерно состоятельные оценки ядерного типа для регрессионной функции $f(\mathbf{t})$.

Относительно погрешностей мы будем предполагать выполненным одно из следующих двух ограничений.

($\mathbf{E}_1$) *При всех $n \geq 1$ ненаблюдаемые случайные погрешности $\{\varepsilon_i;\ i = 1, \ldots, n\}$ с вероятностью 1 при всех $i, j \leq n$, $i \neq j$, удовлетворяют следующим условиям:*

$$\mathbb{E}_{\mathcal{F}_n} \varepsilon_i = 0, \qquad \sup_{i \leq n} \mathbb{E}_{\mathcal{F}_n} \varepsilon_i^2 \leq \sigma^2 < \infty, \qquad \mathbb{E}_{\mathcal{F}_n} \varepsilon_i \varepsilon_j = 0,$$

где константа $\sigma^2 > 0$ неизвестна и не зависит от n, а символ $\mathbb{E}_{\mathcal{F}_n}$ обозначает условное математическое ожидание при фиксации σ-алгебры $\mathcal{F}_n$, порожденной набором регрессоров $\{\mathbf{z}_i;\ i = 1, \ldots, n\}$.

($\mathbf{E}_2$) *При всех $n \geq 1$ ненаблюдаемые случайные погрешности $\{\varepsilon_i;\ i = 1, \ldots, n\}$ образуют последовательность мартингал-разностей с условием*

$$M_p = \sup_{i \leq n} \mathbb{E}|\varepsilon_i|^p < \infty \quad \textit{при} \quad \textit{некотором } p > k \textit{ и } p \geq 2,$$

где M_p не зависит от n. Предполагается также, что набор $\{\varepsilon_i\}$ не зависит от $\{\mathbf{z}_i\}$ и случайной регрессионной функции $f(\cdot)$, при этом элементы $\{\varepsilon_i, i = 1, \ldots, n\}$ могут зависеть от n.

Относительно ядра сглаживания всюду считаем выполненным следующее условие.

($\mathbf{K}$) *Ядерная функция $K(\mathbf{t})$, $\mathbf{t} \in \mathbb{R}^k$ является плотностью распределения с носителем*

$[-1,1]^k$, *т.е.* $K(\mathbf{t}) \geq 0$ *при всех* $\mathbf{t} \in [-1,1]^k$ *и* $\int_{[-1,1]^k} K(\mathbf{t})d\mathbf{t} = 1$. *Кроме того,* $K(\mathbf{t}) = K(-\mathbf{t})$ *при всех* $\mathbf{t} \in [-1,1]^k$.

Нам также потребуется обозначение $K_h(\mathbf{t}) = h^{-k}K(h^{-1}\mathbf{t})$. Понятно, что $K_h(\mathbf{t})$ является плотностью распределения на $[-h,h]^k$.

З а м е ч а н и е 1.1. Подчеркнем, что модель (M_1) включает в себя и ситуацию фиксированных регресоров $\{\mathbf{z}_i\}$. Исключительно для простоты в качестве $\mathcal{P}$ мы будем рассматривать, как правило, куб $[0,1]^k$. В общем случае в качестве $\mathcal{P}$ можно рассматривать произвольное измеримое по Жордану множество $\mathcal{P} \subset \mathbb{R}^k$. Отметим еще, что в качестве ядра $K(\mathbf{t})$ можно выбирать функции вида $K(\mathbf{t}) = \prod\limits_{j=1}^{k} K_o(t_j)$, где $K_o(\cdot)$ есть плотность симметричного распределения с носителем $[-1,1]$. $\square$

Условимся, что символ $\mathbb{P}_{\mathcal{F}_n}$ обозначает условную вероятность при фиксации σ-алгебры $\mathcal{F}_n$, введенной в условии (E_1). Для любого $x \in \mathbb{R}^+$ символ $\lceil x \rceil$ есть минимальное натуральное число, большее или равное x. Нам потребуется обозначение для модуля непрерывности той или иной функции g одного или k аргументов, заданной соответственно на отрезке $[0,1]$ или единичном k-мерном кубе $[0,1]^k$:

$$\omega_g(h) = \sup_{\mathbf{x},\mathbf{y}:\, \|\mathbf{x}-\mathbf{y}\| \leq h} |g(\mathbf{x}) - g(\mathbf{y})|,$$

где для k-мерного аргумента обозначение $\|\cdot\|$ — это *супремальная норма* в $\mathbb{R}^k$, а для одномерного — обычный модуль.

Глава устроена следующим образом. В разделах 1.1 и 1.2 изучаются новые универсальные локально–постоянные и локально–линейные оценки ядерного типа в случае $k = 1$. В разделе 1.3 приведены обобщения некоторых результатов разделов 1.1 и 1.2 на случай оценивания случайной вещественнозначной регрессионной функции нескольких переменных (случайного поля). В разделе 1.4 рассматриваются вопросы состоятельности оценок Надарая–Ватсона и классических локально–линейных оценок. Задаче оценивания функций среднего и ковариации непрерывного случайного процесса посвящен раздел 1.5.

Результаты главы опубликованы в работах [351]–[355] и [357]–[360].

1.1 Универсальные локально–постоянные оценки

1.1.1 Основные результаты. Равномерная состоятельность

Всюду в этом разделе предполагаем выполненным условие (M_1) при $k = 1$ и $\mathcal{P} = [0,1]$. Обозначим через $z_{n:1} \leq \ldots \leq z_{n:n}$ элементы вариационного ряда, построенного по выборке

$\{z_i;\, i = 1, \ldots, n\}$. Положим

$$z_{n:0} = 0, \qquad z_{n:n+1} = 1 \quad \text{и} \quad \Delta z_{ni} = z_{n:i} - z_{n:i-1}, \quad i = 1, \ldots, n+1.$$

Наблюдения (отклики) из регрессионного уравнения (23), ассоциированные с порядковой статистикой $z_{n:i}$, обозначим через X_{ni}.

Нам потребуется еще два предположения.

($\mathbf{K}_1$) *Выполнено условие* (K) *при* $k = 1$, *ядерная функция* $K(t)$ *удовлетворяет условию Липшица с константой* $L \geq 1$ *и* $K(\pm 1) = 0$.

($\mathbf{D}_1$) *Имеет место предельное соотношение* $\delta_n = \max\limits_{1 \leq i \leq n+1} \Delta z_{ni} \xrightarrow{p} 0$.

Подчеркнем, что условие (D_1) — это единственное условие на регрессоры, которое гарантирует существование состоятельных в равномерной норме оценок для регрессионной функции $f \in C[0,1]$. Это центральное условие мы подробнее обсудим далее.

Оценку для функции f определим равенством

$$f^*_{n,h}(t) = \frac{\sum_{i=1}^n X_{ni} K_h(t - z_{n:i}) \Delta z_{ni}}{\sum_{i=1}^n K_h(t - z_{n:i}) \Delta z_{ni}}. \tag{24}$$

Заметим, что

$$f^*_{n,h}(t) = \arg\min_a \sum_{i=1}^n (X_{ni} - a)^2 K_h(t - z_{n:i}) \Delta z_{ni},$$

т.е. оценки (24), как и классические оценки Надарая–Ватсона, являются оценками взвешенного метода наименьших квадратов и относятся к классу локально–постоянных оценок. Но в методе наименьших квадратов используются другие веса, определяемые порядковыми статистиками, построенным по набору регрессоров, а вместо исходных наблюдений X_i участвуют наблюдения X_{ni}, ассоциированные с соответствующими порядковыми статистиками.

Основной результат раздела содержит следующее утверждение.

Теорема 1.1. *Пусть выполнено условие* (M_1) *при* $k = 1$ *и* $\mathcal{P} = [0,1]$, *а также условия* (E_1) *и* (K_1). *Тогда для любого фиксированного* $h \in (0, 1/2)$ *с вероятностью* 1

$$\sup_{t \in [0,1]} |f^*_{n,h}(t) - f(t)| \leq \omega_f(h) + \zeta_n(h),$$

где случайная величина $\zeta_n(h)$ *такова, что*

$$\mathbb{P}\big(\zeta_n(h) > y\big) \leq C\sigma^2 L^2 y^{-2} h^{-2} \mathbb{E}\delta_n + \mathbb{P}(\delta_n > h/(8L))$$

и C *– абсолютная положительная константа.*

З а м е ч а н и е 1.2. Отметим, что формально в теореме 1.1 не требуется выполнения условия (D_1), но оценки, полученные в теореме, становятся, на наш взгляд, содержательными лишь в этом случае. Поскольку $\delta_n \le 1$, то при выполнении условия (D_1) имеет место предельное соотношение $\mathbb{E}\delta_n \to 0$. Следовательно, с учетом теоремы 1.1 можно утверждать, что $\zeta_n(h) = O_p\big((\sigma+1)Lh^{-1}(\mathbb{E}\delta_n)^{1/2}\big)$. Таким образом, h можно определить, к примеру, соотношением

$$h_n = \sup\left\{h > 0 :\, \mathbb{P}\left(\omega_f(h) \ge h^{-1}(\mathbb{E}\delta_n)^{1/2}\right) \le h^{-1}(\mathbb{E}\delta_n)^{1/2}\right\}. \tag{25}$$

Нетрудно видеть, что при выполнении (D_1) имеют место предельные соотношения $h_n \to 0$ и $h_n^{-1}(\mathbb{E}\delta_n)^{1/2} \to 0$. Фактически величина h_n уравнивает по h порядок малости по вероятности обоих слагаемых в оценке для супремальной нормы в теореме 1.1.

Заметим, что в ситуации неслучайной f можно выбрать $h \equiv h_n$ как решение уравнения

$$h^{-1}(\mathbb{E}\delta_n)^{1/2} = \omega_f(h). \tag{26}$$

Понятно, что это решение будет стремиться к нулю с ростом n.

Соотношения (25) и (26) позволяют нам получить порядок малости оптимального h, но не само оптимальное h. На практике h может быть выбрано, например, так называемой перекрестной проверкой (кросс-валидацией). □

Из теоремы 1.1 и замечания 1.2 нетрудно извлечь

Следствие 1.1. *Пусть функция f в модели (M_1) при $k=1$, $\mathcal{P}=[0,1]$ неслучайна, выполнены условия (E_1), (K_1), (D_1) и $\mathcal{C}$ есть произвольное подмножество равностепенно непрерывных функций в $C[0,1]$. Тогда*

$$\gamma_n(\mathcal{C}) = \sup_{f\in\mathcal{C}}\sup_{t\in[0,1]} |f^*_{n,\tilde{h}_n}(t) - f(t)| \xrightarrow{p} 0,$$

где размер окна $\tilde{h}_n$ определяется уравнением (26) при замене модуля непрерывности $\omega_f(h)$ универсальным модулем $\omega_f^{\mathcal{C}}(h) = \sup\limits_{f\in\mathcal{C}}\omega_f(h)$. Кроме того, $\gamma_n(\mathcal{C}) = O_p\big((\sigma+1)L\omega_f^{\mathcal{C}}(\tilde{h}_n)\big)$.

З а м е ч а н и е 1.3. Нетрудно видеть, что для неслучайной f модуль непрерывности в (26) можно заменить той или иной верхней границей для $\omega_f^{\mathcal{C}}(h)$, получая соответствующую верхнюю границу для $\gamma_n(\mathcal{C})$. Рассмотрим случай, когда $\mathbb{E}\delta_n = O(1/n)$. Если $\mathcal{C}$ состоит из функций f, удовлетворяющих условию Гельдера с показателем $\alpha \in (0,1]$ и универсальной константой, то $\tilde{h}_n = O\left(n^{-\frac{1}{2(1+\alpha)}}\right)$ и $\omega_f^{\mathcal{C}}(\tilde{h}_n) = O\left(n^{-\frac{\alpha}{2(1+\alpha)}}\right)$. В частности, если функции из $\mathcal{C}$ удовлетворяют условию Липшица ($\alpha = 1$) с универсальной константой, то $\gamma_n(\mathcal{C}) = \widetilde{O}_p\big(n^{-1/4}\big)$.

Из теоремы 1.1 и замечания 1.2 получаем

Следствие 1.2. *Пусть выполнено условие* (M_1) *при* $k = 1$ *и* $\mathcal{P} = [0,1]$, *условия* (E_1), (K_1), (D_1), *а модуль непрерывности случайной регрессионной функции* f *допускает оценку* $\omega_f(h) \le \zeta d(h)$ *п.н., где* $\zeta > 0$ *– собственная случайная величина и* $d(h)$ *положительная непрерывная неслучайная функция с условием* $d(h) \to 0$ *при* $h \to 0$. *Тогда*

$$\sup_{t\in[0,1]} |f^*_{n,\hat h(n)}(t) - f(t)| \xrightarrow{p} 0,$$

где величина $\hat h_n$ *определена в* (26) *при замене* $\omega_f(h)$ *на функцию* $d(h)$.

З а м е ч а н и е 1.4. Обсудим подробнее условие (D_1). Понятно, что неслучайные (это есть случай неодинаково распределенных $\{z_i\}$, зависящих от n) и в том или ином смысле регулярные регрессоры удовлетворяет условию (D_1) . Хорошо известно, что в случае независимых и одинаково распределенных $\{z_i\}$ и существования отделенной от нуля на $[0,1]$ плотности распределения z_1, с вероятностью 1 выполнено $\delta_n = O\left(\frac{\log n}{n}\right)$ (см. детали в подразделе 1.1.5, лемма 1.5).

Вообще в случае одинаково распределенных $\{z_i\}$ для выполнения (D_1) можно требовать, чтобы при всех $\delta \in (0,1)$

$$p_n(\delta) \equiv \sup_{|\Delta|=\delta} \mathbb{P}\left(\bigcap_{i\le n}\{z_i \notin \Delta\}\right) \to 0, \tag{27}$$

где супремум берется по всем интервалам $\Delta \subset [0,1]$ длины δ. В самом деле, для любого натурального $N > 1$ разобьем отрезок $[0,1]$ на N интервалов Δ_k, $k = 1,\dots,N$, длины $1/N$. Имеем

$$\mathbb{P}\left(\max_{1\le i\le n+1} \Delta z_{ni} > \frac{2}{N}\right) \le \sum_{k=1}^{N} \mathbb{P}\left(\bigcap_{i\le n}\{z_i \notin \Delta_k\}\right) \le N \max_k \mathbb{P}\left(\bigcap_{i\le n}\{z_i \notin \Delta_k\}\right) \le N p_n(1/N),$$

поскольку событие $\left\{\max_{1\le i\le n+1} \Delta z_{ni} > 2/N\right\}$ влечет существование интервала Δ_k длины $1/N$, не содержащего точек $\{z_i\}$. Тем самым из условия (27) следует предельное соотношение $\max_{i\le n+1} \Delta z_{ni} \xrightarrow{p} 0$. В частности, если $\{z_i\}$ независимы, то $p_n(\delta) = e^{-c(\delta)n}$ и $c(\delta) > 0$, т.е. $\{z_i\}$ с ростом n образуют измельчающееся разбиение отрезка $[0,1]$.

Если $\{z_i;\, i \ge 1\}$ — стационарная последовательность с условием α-перемешивания и маргинальным распределением с носителем $[0,1]$, то условие (D_1) также выполнено. Действительно, соотношение вида (27) следует из аналогичного соотношения для любой подпоследовательности случайных величин $\{z_{j_r}\}$, в частности, для «разреженного» набора индексов $\{j_r\}$ с условиями $(j_{r+1} - j_r) = m \to \infty$ при $j_1 = 1$ и $m = o(n)$. Так что при условии α-перемешивания исходной последовательности $\{z_i\}$ соотношение (27) также будет выполнено, поскольку

$$\mathbb{P}\left(\bigcap_{r=1}^{\tilde n}\{z_{j_r}\notin\Delta_k\}\right)\le \mathbb{P}\big(z_{j_1}\notin\Delta_k\big)\mathbb{P}\left(\bigcap_{r=2}^{\tilde n}\{z_{j_r}\notin\Delta_k\}\right)+\alpha(m)\le\ldots$$
$$\ldots\le\mathbb{P}^{\tilde n}\big(z_1\notin\Delta_k\big)+\alpha(m)\left(1+\mathbb{P}\big(z_1\notin\Delta_k\big)+\mathbb{P}^2\big(z_1\notin\Delta_k\big)+\ldots+\mathbb{P}^{\tilde n-1}\big(z_1\notin\Delta_k\big)\right),$$

в силу чего

$$N\sup_{\Delta:\,|\Delta|=1/N}\mathbb{P}\left(\bigcap_{r=1}^{\tilde n}\{z_{j_r}\notin\Delta\}\right)\le Np^{\tilde n}+\alpha(m)\frac{N}{1-p}.$$

Здесь $p=\sup_{\Delta:\,|\Delta|=1/N}\mathbb{P}(z_1\notin\Delta)<1$, $\alpha(m)\to 0$ при $m\to\infty$ и число $\tilde n$ есть максимальное натуральное такое, что $1+(\tilde n-1)m\le n$.

Если регрессоры $\{z_i\}$ не зависят от n, то сходимость по вероятности в условии (D_1) эквивалентна сходимости с вероятностью 1. $\square$

Отметим, что все известные автору типы зависимости регрессоров влекут за собой условие (D_1) и выполнение подобного условия в ситуации случайных регрессоров в работах предшественников достигается, как правило, за счет тех или иных форм слабой зависимости случайных величин. Заметим, что зависимость случайных величин $\{z_i\}$ в условии (D_1) может быть более сильной. Этот факт мы продемонстрируем в нижеследующих примерах 1.1 и 1.2.

П р и м е р 1.1. Рассмотрим последовательность равномерно распределенных на $[0,1]$ случайных величин, каждую из которых можно представить в виде равновесной смеси случайных величин U_i^l и U_i^r, равномерно распределенных на отрезках $[0,1/2]$ и $[1/2,1]$ соответственно, т. е.

$$z_i=\nu_iU_i^l+(1-\nu_i)U_i^r,$$

где ν_i – бернуллиевская случайная величина с вероятностью успеха $1/2$, при этом ν_i, U_i^l и U_i^r независимы. Предположим, что компоненты смеси $\{U_i^l,U_i^r;i=1,\ldots,n\}$ независимы в совокупности, а не зависящие от указанной последовательности бернуллиевские переключатели этих смесей для любого $i\ge 1$ связаны следующими соотношениями: $\nu_{2i-1}=\nu_1$ $\nu_{2i}=1-\nu_1$. То есть мы разыгрываем, на какой из отрезков бросаем наудачу первую точку, и далее чередуем по одному бросанию на каждый из двух указанных отрезков. Нетрудно видеть, что $\{z_i\}$ образуют стационарную последовательность равномерно распределенных на отрезке $[0,1]$ случайных величин. Кроме того, при всех натуральных m и n

$$\mathbb{P}(z_{2m}\le 1/2,\,z_{2n-1}\le 1/2)=0,$$
$$\mathbb{C}ov(z_{2m},\,z_{2n-1})=-1/16,\qquad \mathbb{C}ov(z_{2m},\,z_{2n})=\mathbb{C}ov(z_{2m-1},\,z_{2n-1})=1/16,\quad n\ne m.$$

Таким образом, наиболее популярные условия слабой зависимости случайных величин (перемешивания, асимптотическая некоррелируемость) в данном примере не выполнены. В частности, в случае существования ковариационной функции подобные условия предполагают выполнение предельного соотношения $\lim\limits_{|n-m|\to\infty} \mathbb{C}ov(z_n, z_m) = 0$. Более того, нетрудно видеть, что цепь Маркова $\{z_i;\ i = 1, \ldots, n\}$ не эргодична. Тем не менее, стационарная последовательность $\{z_i;\ i = 1, \ldots, n\}$ удовлетворяет условию (D_1). $\square$

П р и м е р 1.2. По схеме примера 1.1 можно конструировать различные последовательности зависимых равномерно распределенных на $[0, 1]$ случайных величин $\{z_i\}$ исходя из выбора различных последовательностей бернуллиевских переключателей с условиями $\nu_{j_k} = 1$ и $\nu_{l_k} = 0$ для неограниченного числа индексов $\{j_k\}$ и $\{l_k\}$ соответственно. В этом случае условие (D_1) также будет выполнено. Но соответствующая последовательность $\{z_i\}$ (не обязательно стационарная) может не удовлетворять, например, усиленному закону больших чисел. Например, подобная ситуация имеет место в случае, когда $\nu_j = 1 - \nu_1$ при $j = 2^{2k-1}, \ldots, 2^{2k} - 1$ и $\nu_j = \nu_1$ при $j = 2^{2k}, \ldots, 2^{2k+1} - 1$, где $k = 1, 2, \ldots$. Иными словами, на первом шаге стохастического эксперимента мы разыгрываем с равными шансами ту или иную половинку единичного отрезка, на которую наудачу бросаем первую точку. Затем мы бросаем наудачу две точки на другую половинку отрезка $[0, 1]$ и вновь возвращаемся на исходную половинку, куда теперь бросаем наудачу четыре точки. Процесс чередования половинок единичного отрезка продолжается, и на k-м шаге процедуры мы бросаем наудачу на очередную половинку единичного отрезка 2^{k-1} точек. Положим теперь $n_k = 2^{2k} - 1$, $\tilde{n}_k = 2^{2k+1} - 1$, $S_m = \sum_{i=1}^{m} z_i$ и заметим, что для почти всех элементарных исходов, составляющих событие $\{\nu_1 = 1\}$, выполнено

$$\frac{S_{n_k}}{n_k} = \frac{1}{n_k} \sum_{i \in A_{1,k}} U_i^l + \frac{1}{n_k} \sum_{i \in A_{2,k}} U_i^r, \tag{28}$$

где $A_{1,k}$ и $A_{2,k}$ — наборы индексов, для которых наблюдения $\{z_i, i \le n_k\}$ расположены в отрезке $[0, 1/2]$ или $[1/2, 1]$ соответственно. Нетрудно видеть, что $\#(A_{2,k}) = 2\#(A_{1,k})$. Следовательно, $S_{n_k}/n_k \to 7/12$ п.н. при $k \to \infty$ ввиду усиленного закона больших чисел для последовательностей $\{U_i^l\}$ и $\{U_i^r\}$. С другой стороны, для почти всех элементарных исходов, принадлежащих событию $\{\nu_1 = 1\}$, с вероятностью 1 выполнено $S_{\tilde{n}_k}/\tilde{n}_k \to 5/12$ при $k \to \infty$ (здесь для аналога (28) мы использовали соотношение $\#(\tilde{A}_{1,k}) = 2\#(\tilde{A}_{2,k}) + 1$, где $\tilde{A}_{1,k}$ и $\tilde{A}_{2,k}$ — наборы индексов, для которых наблюдения $\{z_i, i \le \tilde{n}_k\}$ расположены в отрезке $[0, 1/2]$ или $[1/2, 1]$ соответственно). Схожие выводы справедливы и для почти всех элементарных исходов, составляющих событие $\{\nu_1 = 0\}$. Таким образом, построенная нестационарная последовательность равномерно распределенных на $[0, 1]$ случайных величин $\{z_i\}$ не удовлетворяет усиленному закону больших чисел. Нетрудно видеть, что S_n/n сходится почти наверное к двухточечной случайной величине, так что закон больших чисел не выполнен и в смысле сходимости по вероятности. $\square$

1.1.2 Доказательство теоремы 1.1

Всюду в этом подразделе считаем выполненными условия теоремы 1.1. Погрешности из регрессионного уравнения (23), ассоциированные с порядковой статистикой $z_{n:i}$, обозначим через ε_{ni}. Нетрудно видеть, что случайные величины $\{\varepsilon_{ni};\ i = 1, \ldots, n\}$ также удовлетворяют условию (E_1), а регрессионное уравнение (23) можно переписать следующим образом:

$$X_{ni} = f(z_{n:i}) + \varepsilon_{ni}, \qquad i = 1, \ldots, n.$$

Принимая во внимание это соотношение, получаем тождество

$$f^*_{n,h}(t) = I_{n,h}(f, t) + \nu_{n,h}(t), \tag{29}$$

где

$$\begin{aligned} I_{n,h}(f,t) &= J^{-1}_{n,h}(t) \sum_{i=1}^{n} f(z_{n:i}) K_h\left(t - z_{n:i}\right) \Delta z_{ni}, \\ J_{n,h}(t) &= \sum_{i=1}^{n} K_h(t - z_{n:i}) \Delta z_{ni}, \\ \nu_{n,h}(t) &= J^{-1}_{n,h}(t) \sum_{i=1}^{n} K_h\left(t - z_{n:i}\right) \Delta z_{ni} \varepsilon_{ni}. \end{aligned}$$

Заметим теперь, что ввиду свойств ядра K, область суммирования в трех вышеуказанных суммах есть множество $\{i : |t - z_{n:i}| \le h\}$. Этот факт является принципиальным для дальнейших рассуждений.

Для вывода теоремы 1.1 нам потребуется три вспомогательных леммы.

Лемма 1.1. *Имеет место следующая оценка:*

$$\sup_{t \in [0,1]} |I_{n,h}(f,t) - f(t)| \le \omega_f(h).$$

Д о к а з а т е л ь с т в о немедленно следует из тождества

$$I_{n,h}(f,t) = f(t) + J^{-1}_{n,h}(t) \sum_{i:\, |t - z_{n:i}| \le h} (f(z_{n:i}) - f(t)) K_h\left(t - z_{n:i}\right) \Delta z_{ni}.$$

Лемма доказана. $\square$

Введем обозначение

$$J_h(t) = \int\limits_0^1 K_h(t-z)dz \equiv \mathbb{E}I_{[0,1]}(t-\lambda_h), \tag{30}$$

где через λ_h обозначена случайная величина с плотностью $K_h(t)$.

Лемма 1.2. *Для любого $h < 1/2$ имеют место соотношения*

$$J_h(t) = 1 \text{ при всех } t \in [h, 1-h],$$
$$\inf_{t\in[0,1]} J_h(t) = 1/2, \qquad \sup_{t\in[0,1]} |J_{n,h}(t) - J_h(t)| \le 2L\delta_n/h.$$

Д о к а з а т е л ь с т в о. С учетом (30) нетрудно видеть, что для любого $h < 1/2$ справедливо равенство

$$J_h(t) = \begin{cases} 1 & \text{если } t \in [h, 1-h]; \\ \mathbb{P}(\lambda_h < t) & \text{если } t \in [0, h]; \\ \mathbb{P}(\lambda_h > t-1) & \text{если } t \in [1-h, 1], \end{cases}$$

при этом распределение λ_h симметрично. Очевидно, что функция $J_h(t)$ достигает минимального значения в точках 0 и 1. Для вывода третьего утверждения леммы положим $\Delta_i = [z_{n:i-1}, z_{n:i}]$. Имеем

$$|J_{n,h}(t) - J_h(t)| \le \sum_i \Delta z_{ni} L h^{-2} \Lambda_1\big(\Delta_i \cap [t-h, t+h]\big).$$

Это соотношение завершает вывод леммы. □

Лемма 1.3. *Для любых $y > 0$ и $h \in (0, 1/2)$ на подмножестве элементарных исходов, определяемых соотношением $\delta_n < h/(8L)$, имеет место следующая оценка:*

$$\mathbb{P}_{\mathcal{F}_n}\left((\sigma\, L)^{-1}\delta_n^{-1/2} h \sup_{t\in[0,1]} |\nu_{n,h}(t)| > y\right) \le \frac{1028}{\big(1 - 4L\delta_n/h\big)^2} y^{-2}.$$

Д о к а з а т е л ь с т в о. Из леммы 1.2 нетрудно получить оценку

$$|\nu_{n,h}(t)| \le \frac{2}{1 - 4L\delta_n/h} |\mu_{n,h}(t)| \quad \text{при} \quad \mu_{n,h}(t) = \sum_{i=1}^n K_h\left(t - z_{n:i}\right) \Delta z_{ni} \varepsilon_{ni}. \tag{31}$$

Хвост распределения величины $\sup\limits_{t\in[0,1]} |\mu_{n,h}(t)|$ оценим с помощью метода диадических цепочек для оценки хвоста распределения супремальной нормы стохастического процесса с почти наверное непрерывными траекториями (см. [43]). Заметим, что интервал $[0,1]$ под знаком вышеупомянутого супремума может быть заменен на следующее множество двоично-рациональных точек:

$$\mathcal{R} = \{j/2^k;\ j = 1, \ldots, 2^k - 1; k \geq 1\}.$$

Таким образом, при некотором натуральном m

$$\sup_{t \in [0,1]} |\mu_{n,h}(t)| = \sup_{t \in \mathcal{R}} |\mu_{n,h}(t)| \leq$$
$$\leq \max_{j=1,\ldots,2^m-1} |\mu_{n,h}(j2^{-m})| + \sum_{k=m+1}^{\infty} \max_{j=1,\ldots,2^k-2} \left|\mu_{n,h}((j+1)2^{-k}) - \mu_{n,h}(j2^{-k})\right|.$$

Следовательно,

$$\mathbb{P}_{\mathcal{F}_n}\Big(\sup_{t \in [0,1]} |\mu_{n,h}(t)| > y\Big) \leq \mathbb{P}_{\mathcal{F}_n}\Big(\max_{j=1,\ldots,2^m-1} |\mu_{n,h}(j2^{-m})| > a_m y\Big) +$$
$$+ \sum_{k=m+1}^{\infty} \mathbb{P}_{\mathcal{F}_n}\Big(\max_{j=1,\ldots,2^k-2} \left|\mu_{n,h}((j+1)2^{-k}) - \mu_{n,h}(j2^{-k})\right| > a_k y\Big) \leq$$
$$\leq \sum_{j=1}^{2^m-1} \mathbb{P}_{\mathcal{F}_n}(|\mu_{n,h}(j2^{-m})| > a_m y) +$$
$$+ \sum_{k=m+1}^{\infty} \sum_{j=1}^{2^k-2} \mathbb{P}_{\mathcal{F}_n}\Big(\left|\mu_{n,h}((j+1)2^{-k}) - \mu_{n,h}(j2^{-k})\right| > a_k y\Big), \quad (32)$$

где $a_m, a_{m+1}, \ldots$ — последовательность положительных чисел, удовлетворяющих условию $a_m + a_{m+1} + \ldots = 1$.

Далее, с помощью неравенства Маркова со вторым моментом получаем

$$\mathbb{P}_{\mathcal{F}_n}(|\mu_{n,h}(j2^{-m})| > a_m y) \leq (a_m y)^{-2} \sum_{i:\,|j2^{-m} - z_{n:i}| \leq h} K_h^2(j2^{-m} - z_{n:i})(\Delta z_{ni})^2 \sigma^2 \leq$$
$$\leq \sigma^2 L^2 (a_m y)^{-2} \delta_n (2h + \delta_n) h^{-2}. \quad (33)$$

Кроме того,

$$\mathbb{P}_{\mathcal{F}_n}\Big(\left|\mu_{n,h}((j+1)2^{-k}) - \mu_{n,h}(j2^{-k})\right| > a_k y\Big) \leq$$
$$\leq (a_k y)^{-2} \sum_{i=1}^{n} \mathbb{E}_{\mathcal{F}}\Big(\big(K_h((j+1)2^{-k} - z_{n:i}) - K_h(j2^{-k} - z_{n:i})\big)\Delta z_{ni} \varepsilon_{ni}\Big)^2 \leq$$
$$\leq (a_k y)^{-2} \cdot 2^{-2k} \sigma^2 L^2 \delta_n 2(2h + \delta_n) h^{-4} \quad (34)$$

ввиду оценки

$$|K_h(u - z_{n:i}) - K_h(v - z_{n:i})| \le Lh^{-2}|u - v|$$

и того факта, что область суммирования в первом неравенстве (34) совпадает с множеством $\{i: |(j+1)2^{-k} - z_{n:i}| \le h\} \cup \{i: |j2^{-k} - z_{n:i}| \le h\}$.

Теперь из (32)–(34) получаем

$$\mathbb{P}_{\mathcal{F}_n}\left(\sup_{t\in[0,1]} |\mu_{n,h}(t)| > y\right) \le y^{-2}\sigma^2 L^2 \delta_n (2h + \delta_n) h^{-2} \left(2^m a_m^{-2} + h^{-2} \sum_{k=m+1}^{\infty} 2^{-k+1} a_k^{-2}\right).$$

Оптимальная последовательность a_k, минимизирующая правую часть этого неравенства, есть $a_m = c2^{m/3}$ и $a_k = ch^{-2/3}2^{(-k+1)/3}$ при $k = m+1, m+2, ...$, где c определяется соотношением $a_m + a_{m+1} + ... = 1$. Для указанной последовательности заключаем, что

$$\mathbb{P}_{\mathcal{F}_n}\left(\sup_{t\in[0,1]} |\mu_{n,h}(t)| > y\right) \le y^{-2}\sigma^2 L^2 \delta_n (2h + \delta_n) h^{-2} \left(2^{m/3} + h^{-2/3}2^{-m/3}\bigl(2 + 2^{1/3} + 2^{2/3}\bigr)\right)^3.$$

Положим $m = \lceil |\log_2 h| + b \rceil$, где $b > 0$. Имеем

$$\mathbb{P}_{\mathcal{F}_n}\left(\sup_{t\in[0,1]} |\mu_{n,h}(t)| > y\right) \le$$
$$\le y^{-2}\sigma^2 L^2 \delta_n (2h + \delta_n) h^{-3} \left(2^{(b+1)/3} + 2^{-b/3}\bigl(2 + 2^{1/3} + 2^{2/3}\bigr)\right)^3 \le 257 y^{-2}\sigma^2 L^2 \delta_n h^{-2},$$

где b выбирается так, чтобы $2^{b/3} = 2^{-1/6}\sqrt{2 + 2^{1/3} + 2^{2/3}}$. Заметим еще, что в силу условий леммы и ограничения $L \ge 1$ выполнено $2h + \delta_n < 17h/8$. Утверждение леммы следует теперь из (31). $\square$

Завершим доказательство теоремы 1.1. Положим $\zeta_n(h) = \sup_{t\in[0,1]} |\nu_{n,h}(t)|$. Нам остается воспользоваться тождеством (29), утверждениями лемм 1.1, 1.3 и учесть соотношение

$$\mathbb{P}\bigl(\zeta_n(h) > y,\, \delta_n \le h/(8L)\bigr) = \mathbb{E}I\bigl(\delta_n \le h/(8L)\bigr)\mathbb{P}_{\mathcal{F}_n}\bigl(\zeta_n(h) > y\bigr).$$

Теорема 1.1 доказана. $\square$

1.1.3 Асимптотическая нормальность

В этом разделе приведем достаточные условия для асимптотической нормальности оценок $f^*_{n,h}(t)$. Результаты такого типа могут быть использованы при построении доверительных интервалов для значений регрессионной функции f при любом t.

Теорема 1.2. *Пусть в модели* (M_1) *при* $k = 1$ *и* $\mathcal{P} = [0,1]$ *погрешности* $\{\varepsilon_i\}$ *независи-*

мы и одинаково распределены, имеют нулевое среднее, дисперсию σ^2 и конечный третий момент, при этом $\{\varepsilon_i\}$ не зависят от регрессоров $\{z_i\}$ и случайного процесса $f(\cdot)$. Кроме того, выполнены условия (K_1), (D_1) и при некотором $t \in [0,1]$ и $h \equiv h_n$

$$\frac{\max\limits_{k\le n} K_h^2\left(t - z_{n:k}\right)\left(\Delta z_{nk}\right)^2}{\sum_{j=1}^{n} K_h^2\left(t - z_{n:j}\right)\left(\Delta z_{nj}\right)^2} \xrightarrow{p} 0. \tag{35}$$

Тогда

$$\eta_{n,h}(t) = B_{n,h}^{-1}(t)\left(f_{n,h}^*(t) - f(t) - r_{n,h}(t)\right) \xrightarrow{d} \mathcal{N}_1(0,\sigma^2), \tag{36}$$

где

$$\begin{aligned}
B_{n,h}^2(t) &= J_{n,h}^{-2}(t)\sum_{i=1}^{n} K_h^2\left(t - z_{n:i}\right)\left(\Delta z_{ni}\right)^2,\\
J_{n,h}(t) &= \sum_{i=1}^{n} K_h(t - z_{n:i})\Delta z_{ni},\\
r_{n,h}(t) &= J_{n,h}^{-1}(t)\sum_{i=1}^{n}(f(z_{n:i}) - f(t))K_h\left(t - z_{n:i}\right)\Delta z_{ni}.
\end{aligned}$$

Кроме того, для любых $t_1 < \ldots < t_m$ таких, что при всех достаточно больших n выполнено $\min_k |t_k - t_{k-1}| > 2h_n$, величины $\eta_{n,h}(t_1), \ldots, \eta_{n,h}(t_m)$ асимптотически независимы. В частности, если условие (35) выполнено при всех $t \in [0,1]$, то конечномерные распределения случайного процесса $\eta_{n,h}(t)$ слабо сходятся к конечномерным распределениям стационарного центрированного гауссовского процесса с независимыми значениями.

З а м е ч а н и е 1.5. Из определения величины $r_{n,h}(t)$ и лемм 1.1, 1.2 можно извлечь оценки

$$\begin{aligned}
&\sup_{0\le t\le 1} |r_{n,h}(t)| \le \omega_f(h),\\
&\sup_{0\le t\le 1} B_{n,h}^2(t) \le 2\delta_n h^{-1}\int K^2(x)dx + O_p(L^2\delta_n^2 h^{-2}).
\end{aligned} \tag{37}$$

Говорят, что случайная регрессионная функция $f(t)$ принадлежит пространству Гельдера с показателем $\alpha \in (0,1]$, если $\omega_f(h) \le \zeta h^\alpha$ п.н. для всех достаточно малых $h > 0$ и некоторой собственной случайной величины ζ. В условиях теоремы 1.2 и с учетом (36) и (37) для таких случайных регрессионных функций $f(t)$ можно строить асимптотические доверительные интервалы уровня $1-\beta$ с границами

$$f_{n,h}^*(t) \pm (x_\beta \hat{\sigma}_n B_{n,h}(t) + h^{\alpha'}) \tag{38}$$

для любого положительного $\alpha' < \alpha$, где

$$\hat{\sigma}_n^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (X_{ni+1} - X_{ni})^2,$$

а величина x_β определяется уравнением

$$\int_{-\infty}^{-x_\beta} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = \beta/2.$$

В частности, если $f(t)$ — винеровский процесс, то в (38) можно рассматривать произвольное $\alpha' < 1/2$. В (38) предполагается, что $h \to 0$ и $\delta_n h^{-1} \to 0$ с вероятностью 1.

Отметим, что оценка $\hat{\sigma}_n^2$ аналогична хорошо известным оценкам в случае фиксированных регрессоров (см., например, [26]). Состоятельность оценки $\hat{\sigma}_n^2$ в условии (D_1) может быть доказана аналогично случаю детерминированных регрессоров. □

З а м е ч а н и е 1.6. Если дополнительно к ограничениям на ядерную функцию $K(s)$, введенным в (K_1), предполагать, что $\inf\limits_{s:\,|s| \le 1/N} K(s) > 0$ при некотором $N > 1$, то для проверки (35) достаточно установить, что

$$\delta_n^2 \Big(\sum_{|t - z_{n:j}| \le h/N} (\Delta z_{nj})^2 \Big)^{-1} \xrightarrow{p} 0. \tag{39}$$

Условие (35) не требует выполнение предельного соотношения $h_n \to 0$. В (35) и (39) можно рассматривать h фиксированным. Тем не менее, можно выбирать последовательность $h_n \to 0$, удовлетворяющую соотношениям (35) или (39). Например, если в (39) при некотором неслучайном $c > 0$ имеется оценка снизу $\inf\limits_{j:\,|t - z_{n:j}| \le h/N} \Delta z_{nj} \ge c\delta_n$, то соотношение (39) справедливо для любой последовательности $h_n \to 0$ с условием $\delta_n / h_n \xrightarrow{p} 0$. Отметим также, что в случае независимых одинаково распределенных регрессоров некоторые достаточные условия для выполнения (35) будут предложены в разделе 1.1.5 (см. замечание 1.7). □

1.1.4 Доказательство теоремы 1.2

Как и в начале доказательства теоремы 1.1, воспользуемся равенством

$$f_{n,h}^*(t) - f(t) - r_{n,h}(t) = \nu_{n,h}(t),$$

где

$$\nu_{n,h}(t) = J_{n,h}^{-1}(t) \sum_{k=1}^{n} K_h\left(t - z_{n:k}\right) \varepsilon_k \Delta z_{nk} \equiv \sum_{k=1}^{n} b_{nk}(t) \varepsilon_k,$$

$$r_{n,h}(t) = J_{n,h}^{-1}(t) \sum_{i=1}^{n} (f(z_{n:i}) - f(t)) K_h\left(t - z_{n:i}\right) \Delta z_{ni}.$$

Заметим, что $B_{n,h}^2(t) = \sum\limits_{k=1}^{n} b_{nk}^2(t)$, а соотношение (35) можно переписать в виде

$$B_{n,h}^{-1}(t) \max_{k \le n} b_{nk}(t) \xrightarrow{p} 0.$$

Рассмотрим теперь следующее представление характеристической функции случайной величины $\eta_{n,h}(t)$:

$$\mathbb{E}e^{is\eta_{n,h}(t)} = \mathbb{E}\mathbb{E}_{\mathcal{F}_n} \exp\left\{ isB_{n,h}^{-1}(t) \sum_{k=1}^{n} b_{nk}(t)\varepsilon_k \right\} =$$

$$= \mathbb{E} \prod_{k=1}^{n} \left\{ 1 - 2^{-1}s^2\sigma^2 B_{n,h}^{-2}(t) b_{nk}^2(t) \left[1 + \sigma^{-2}\gamma_s\left(B_{n,h}^{-1}(t) b_{nk}(t) \right) \right] \right\}, \quad (40)$$

где

$$\gamma_s(x) = 2\mathbb{E}\left(\varepsilon_1^2 \int_0^1 (1-u)(e^{isxu\varepsilon_1} - 1) du \right).$$

Далее, по теореме Лебега о мажорируемой сходимости $\gamma_s(x) \to 0$ при $x \to 0$ и всех фиксированных действительных s. С учетом элементарных оценок

$$B_{n,h}^{-1}(t) b_{nk}(t) \le 1, \qquad |\gamma_s(x)| \le |sx| \mathbb{E}|\varepsilon_1|^3/3$$

и асимптотического представления

$$\prod_{i=1}^{n} (1 + \beta_i) = \exp\Big\{ \sum_{i=1}^{n} \beta_i + O\Big(\sum_{i=1}^{n} \beta_i^2 \Big) \Big\},$$

в (40) можно воспользоваться вышеупомянутой теоремой Лебега для внесения предела по n под знак математического ожидания. Отсюда получаем окончательное предельное соотношение $\mathbb{E}e^{is\eta_{n,h}} \to e^{-\sigma^2 s^2/2}$, которое и требовалось доказать.

Для вывода последнего утверждения теоремы нужно исследовать предельное поведение произвольной линейной комбинации вида $\sum\limits_{k \le m} c_k \eta_{n,h}(t_k)$. Для этого прежде всего заметим, что

$$K_h\left(t_i - z_{n:k}\right) K_h\left(t_j - z_{n:k}\right) = 0,$$

если только $|t_i - t_j| > 2h$. Так что в условиях теоремы $\mathbb{E}\eta_{n,h}(t_i)\eta_{n,h}(t_j) = 0$ при всех $t_i \ne t_j$ и достаточно большом n. Повторяя эти аргументы в (40) нетрудно получить, что предельная характеристическая функция вышеуказанной линейной комбинации представима в виде характеристической функции аналогичной линейной комбинации m независимых одинаково распределенных гауссовских случайных величин с параметрами 0 и σ^2. $\square$

1.1.5 Сравнение с оценками Надарая–Ватсона

Интересно сравнить новые универсальные локально–постоянные оценки с классическим представителем локально–постоянных оценок — оценками Надарая–Ватсона. В этом разделе считаем, что в модели (M_1) при $k = 1$ регрессионная функция f неслучайна, случайные величины $\{z_i\}$ независимы и одинаково распределены, погрешности $\{\varepsilon_i\}$ независимы, одинаково распределены, центрированы, не зависят от регрессоров $\{z_i\}$ и $0 < \mathbb{E}\varepsilon_1^2 = \sigma^2 < \infty$. Оценки мы сравним поточечно в среднеквадратичном смысле. Напомним, что оценки Надарая–Ватсона определяются равенством

$$\hat{f}_{NW}(t) = \frac{\sum_{k=1}^{n} X_k K_h(t - z_k)}{\sum_{k=1}^{n} K_h(t - z_k)}.$$

Как и при выводе теоремы 1.2, имеем

$$\hat{f}_{NW}(t) = f(t) + \hat{r}_{n,h}(t) + \hat{\nu}_{n,h}(t),$$

где

$$\hat{r}_{n,h}(t) = \frac{\sum_{k=1}^{n} (f(z_k) - f(t)) K_h(t - z_k)}{\sum_{k=1}^{n} K_h(t - z_k)}, \qquad \hat{\nu}_{n,h}(t) = \frac{\sum_{k=1}^{n} K_h(t - z_k)\, \varepsilon_k}{\sum_{k=1}^{n} K_h(t - z_k)}.$$

Нетрудно видеть, что справедливы равенства

$$\begin{gathered} \mathbb{E}\hat{f}_{NW}(t) - f(t) = \mathbb{E}\hat{r}_{n,h}(t), \qquad \mathbb{E}\hat{f}^*_{n,h}(t) - f(t) = \mathbb{E}\hat{r}_{n,h}(t), \\ \mathbb{D}\hat{f}_{NW}(t) = \mathbb{D}\hat{\nu}_{n,h}(t) + \mathbb{D}\hat{r}_{n,h}(t), \qquad \mathbb{D}f^*_{n,h}(t) = \mathbb{D}\nu_{n,h}(t) + \mathbb{D}r_{n,h}(t) \end{gathered} \tag{41}$$

(напомним, что величины $r_{n,h}(t)$ и $\nu_{n,h}(t)$ введены в подразделе 1.1.4). Сравним смещения $\mathbb{E}\hat{r}_{n,h}(t)$ и $\mathbb{E}r_{n,h}(t)$ оценок $\hat{f}_{NW}(t)$ и $f^*_{n,h}(t)$, а также дисперсии этих оценок. Здесь и далее в подразделе считаем, что функция распределения $F(t)$ случайной величины z_1 имеет непрерывно дифференцируемую положительную на $(0, 1)$ плотность $p(t)$. Далее для любой ограниченной на $[0, 1]$ функции положим $\bar{g} = \sup\limits_{0 \le t \le 1} |g(t)|$ и $\underline{g} = \inf\limits_{0 \le t \le 1} |g(t)|$. Нам также потребуются обозначения

$$\kappa_2 = \int\limits_{-1}^{1} u^2 K(u) du, \qquad \|K\|^2 = \int\limits_{-1}^{1} K^2(u) du.$$

Напомним, прежде всего, хорошо известные результаты (см., например, [244], [189], [190] и [191]), связанные с асимптотикой величин $\mathbb{E}\hat{r}_{n,h}(t)$ и $\mathbb{D}\hat{f}_{NW}(t)$.

Предложение 1.1. *Пусть функция $f(t)$ дважды непрерывно дифференцируема на $[0,1]$. Если $n \to \infty$ и $h \to 0$ так, что $h^3 n \to \infty$, то для любого $t \in (0,1)$ выполнено*

$$\mathbb{E}\hat{f}_{NW}(t) - f(t) = \frac{h^2 \kappa_2}{2p(t)} \left(f''(t)p(t) + 2f'(t)p'(t)\right) + o(h^2), \quad \mathbb{D}\hat{f}_{NW}(t) \sim \frac{\sigma^2}{hnp(t)} \|K\|^2.$$

Исследуем теперь асимптотику соответствующих величин для оценки $f^*_{n,h}(t)$. Обозначим через $\zeta_j \equiv \zeta_j(n,h,\{z_i\})$, $j = 0,1,\ldots$, некоторые случайные величины, удовлетворяющие с вероятностью 1 условию $|\zeta_j| \le 1$. В следующем утверждении не требуется никакой детализации характера распределения регрессоров.

Лемма 1.4. *Пусть функция $f(t)$ дважды непрерывно дифференцируема на $[0,1]$. Если $h \to 0$, то при любом $t \in (0,1)$ имеет место следующее асимптотическое представление:*

$$r_{n,h}(t) = \frac{h^2 \kappa_2}{2} f''(t) + 2\zeta_0 |\bar{f}'| L (3\delta_n + h^{-1}\delta_n^2) + o(h^2),$$

где $o(h^2)$ неслучайно и не зависит от n. Кроме того, если $h^{-2}\mathbb{E}\delta_n \to 0$ и $h^{-3}\mathbb{E}\delta_n^2 \to 0$, то

$$\mathbb{E}r_{n,h}(t) = \frac{h^2 \kappa_2}{2} f''(t) + o(h^2), \qquad \mathbb{D}r_{n,h}(t) = O(\mathbb{E}\delta_n^2 + h^{-2}\mathbb{E}\delta_n^4).$$

Лемма 1.5. *Пусть $\underline{p} > 0$. Если $(\log n)^{-1} h\sqrt{n} \to \infty$, то*

$$\mathbb{D}\nu_{n,h}(t) \sim \frac{2\sigma^2}{hnp(t)} \|K\|^2. \tag{42}$$

Более того, с вероятностью 1 выполнено

$$\limsup_n \frac{n\underline{p}}{3\log n} \delta_n < 1$$

и при всех достаточно больших n

$$\frac{\log n}{2n\bar{p}} \le \mathbb{E}\delta_n \le \frac{2\log n}{n\underline{p}}. \tag{43}$$

Лемма 1.6. *Пусть выполнены условия леммы 1.4 (за исключением дополнительных ограничений) и условия леммы 1.5. Тогда*

$$\mathbb{D}r_{n,h}(t) = o(\mathbb{D}\nu_{n,h}(t)) \qquad \text{и} \qquad \mathbb{D}f^*_{n,h}(t) \sim \frac{2\sigma^2}{hnp(t)} \|K\|^2.$$

З а м е ч а н и е 1.7. В условиях леммы 1.5 справедливо предположение (35), поскольку, как нетрудно получить, частное в (35) с вероятностью 1 имеет порядок $O(1/(nh))$ (см. приводимые далее формулы (54) и (55)).

Из лемм 1.4, 1.6 и соотношений (41) получаем следующее утверждение.

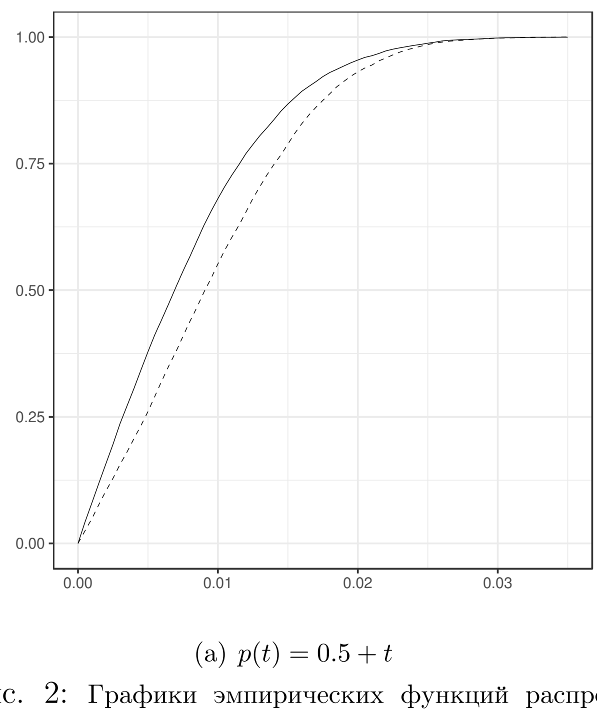
Предложение 1.2. *Пусть* $\underline{p} > 0$, $n \to \infty$ *и* $h \to 0$ *так, что* $h\sqrt{n}/\log n \to \infty$, $h^{-2}\mathbb{E}\delta_n \to 0$ *и* $h^{-3}\mathbb{E}\delta_n^2 \to 0$. *Тогда при любом* $t \in (0,1)$

$$\mathbb{E}f^*_{n,h}(t) - f(t) = \frac{h^2\kappa_2}{2}f''(t) + o(h^2), \qquad \mathbb{D}f^*_{n,h}(t) \sim \frac{2\sigma^2}{hnp(t)}\|K\|^2.$$

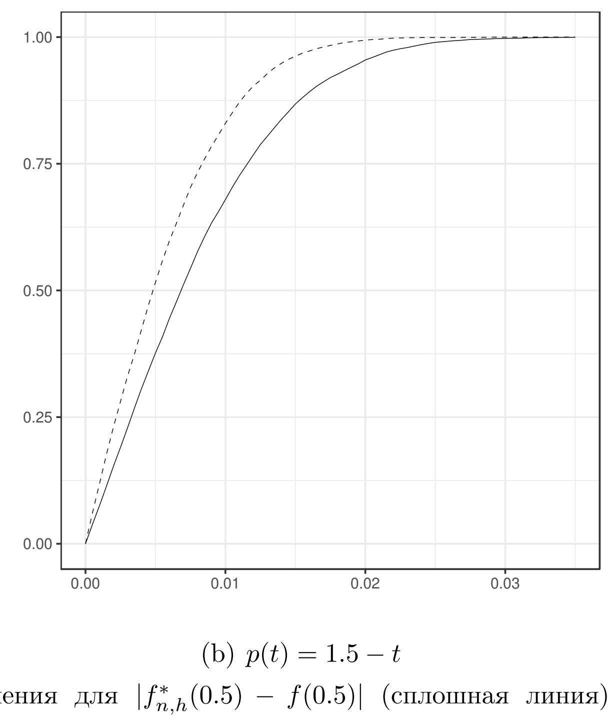
З а м е ч а н и е 1.8. Приведенный выше асимптотический анализ (в частности, предложения 1.1 и 1.2) показывает, что в случае гладкой регрессионной функции и независимых одинаково распределенных регрессоров дисперсия оценки Надарая–Ватсона $\hat{f}_{NW}(t)$ асимптотически меньше дисперсии универсальной локально–постоянной оценки $f^*_{n,h}(t)$. Но выбор одной из этих оценок определяется не только дисперсией, но и смещением, задаваемым величинами $f''(t)p(t) + 2f'(t)p'(t)$ и $f''(t)p(t)$ соответственно. Другими словами, если стандартное отклонение погрешностей σ не сильно велико и справедливо неравенство

$$|f''(t)p(t) + 2f'(t)p'(t)| > |f''(t)p(t)| \tag{44}$$

то оценка $f^*_{n,h}(t)$ может быть лучше, чем $\hat{f}_{NW}(t)$ в среднеквадратичном смысле. В противном случае предпочтительнее оценка Надарая–Ватсона $\hat{f}_{NW}(t)$. Отметим еще, что для выбора оптимального размера окна h можно приравнять порядки малости смещения и стандартного отклонения оценки. Таким образом, нужно решить уравнение $h^2 \approx (nh)^{-1/2}$, т.е. взять стандартный оптимальный размер окна $h \approx n^{-1/5}$ (с точностью до константы).



(a) $p(t) = 0.5 + t$



(b) $p(t) = 1.5 - t$

Рис. 2: Графики эмпирических функций распределения для $|f^*_{n,h}(0.5) - f(0.5)|$ (сплошная линия) и $|\hat{f}_{n,h}(0.5) - f(0.5)|$ (пунктирная линия), построенные по 10000 реализаций при $n = 1000$.

Рисунок 2 иллюстрирует влияние (44) на точность аппроксимации. Графики изображают эмпирические функции распределения для $|f^*_{n,h}(0.5) - f(0.5)|$ (сплошная линия) и $|\hat{f}_{NW}(0.5) - f(0.5)|$ (пунктирная линия), построенные по 10000 реализаций, где $f(t) = t^2$, $h = 0.15$, погрешности нормально распределены со средним 0 и $\sigma = 0.1$, размер выборки $n = 1000$, $K(t)$ — ядро Епанечникова, плотность $p(t)$ определяется одним из двух соотношений: $p(t) = 0.5+t$ или $p(t) = 1.5-t$. Для плотности $p(t) = 0.5+t$ неравенство (44) выполнено и оценка $f^*_{n,h}(t)$ оказывается лучше, нежели $\hat{f}_{NW}(t)$. Если $p(t) = 1.5 - t$, то ситуация обратная.

1.1.6 Доказательства утверждений раздела 1.1.5

Доказательство леммы 1.4. Прежде всего, условие (D_1) позволяет нам воспользоваться хорошо известной аппроксимацией сумм Римана соответствующим интегралом со стандартной оценкой погрешности:

$$\sum_{i:\,|t-z_{n:i}|\le h} (f(z_{n:i}) - f(t))K_h\left(t - z_{n:i}\right)\Delta z_{ni} =$$

$$= \int\limits_0^1 (f(x) - f(t))K_h(t-x)dx + 2\zeta_1 h^{-1}|\bar{f}'|L \sum_{i:\,|t-z_{n:i}|\le h} (\Delta z_{ni})^2. \quad (45)$$

Остается заметить, что

$$\sum_{i:\,|t-z_{n:i}|\le h} (\Delta z_{ni})^2 \le \delta_n(2h + \delta_n),$$

$$\int\limits_0^1 (f(x) - f(t))K_h(t-x)dx = \int\limits_{-1}^1 (f(t+hu) - f(t))K(u)du = h^2\kappa_2\frac{1}{2}f''(t) + o(h^2).$$

Окончательно, с учетом леммы 1.2, при всех фиксированных $t \in (0,1)$ и достаточно малых $h > 0$ получаем

$$J_{n,h}(t) = 1 + \zeta_2 L\delta_n h^{-1}. \quad (46)$$

Таким образом, с учетом первой оценки в (37) (см. замечание 1.5), выполнено

$$|J_{n,h}(t)r_{n,h}(t) - r_{n,h}(t)| \le 2|\bar{f}'|L\delta_n.$$

Вышеприведенные оценки и представления доказывают первое утверждение леммы. Остальные утверждения очевидным образом извлекаются из первого. □

Доказательство леммы 1.5. Без ограничения общности можно считать, что $z_{n:i} = F^{-1}(U_{n:i})$, где $F^{-1}(x)$ при $x \in [0,1]$ есть обратная функция для $F(t)$ и через

$U_{n:1} \leq \ldots \leq U_{n:n}$ обозначен вариационный ряд, построенный по выборке $\{U_i;\, i = 1, \ldots, n\}$ равномерно распределенных на $[0, 1]$ случайных величин. Нам понадобится следующее известное представление вариационного ряда по выборке объема n из равномерного распределения (см., например, [325]):

$$U_{n:k} \stackrel{d}{=} S_k / S_{n+1},$$

где $S_k = \sum_{i=1}^{k} \tau_i$ и независимые случайные величины $\{\tau_i\}$ имеют показательное распределение с параметром 1. В дальнейшем считаем, что $U_{n:k} = S_k/S_{n+1}$. Заметим, что по неравенству Бернштейна для хвостов распределения максимума n частичных сумм центрированных случайных величин с конечными экспоненциальными моментами (см., например, [225]) выполнено

$$\mathbb{P}\left(\max_{k \leq n} |S_k - \mathbb{E}S_k| \geq \alpha_n n\right) \leq e^{-c_o \alpha_n \sqrt{n}} \tag{47}$$

при всех $n \geq n_0$ и произвольной последовательности положительных чисел $\alpha_n \to 0$ с условием $\alpha_n \sqrt{n} \to \infty$, где $c_o > 0$ — некоторая абсолютная константа. Положим $\alpha_n = 2 \log n / (c_o \sqrt{n})$. По лемме Бореля–Кантелли отсюда заключаем, что с вероятностью 1 при всех достаточно больших n супремальная норма между векторами $\{S_k/n;\, k = 1, \ldots, n\}$ и $\{k/n;\, k = 1, \ldots, n\}$ меньше, чем α_n. Используя приведенное выше представление для $U_{n:k}$ и усиленный закон больших чисел для $\{\tau_i\}$, получаем аналогичное утверждение для случайных векторов $\{U_{n:k};\, k = 1, \ldots, n\}$ и $\{k/n;\, k = 1, \ldots, n\}$.

Далее, по формуле конечных приращений получаем следующее соотношение:

$$\Delta z_{nk} = F^{-1}(U_{n:k}) - F^{-1}(U_{n:k-1}) = \frac{\tau_k}{(n+1) p(F^{-1}(\theta_{nk}))} (1 + \gamma_n), \tag{48}$$

где $\gamma_n = ((n+1)/S_{n+1} - 1)$ и $U_{n:k-1} \leq \theta_{nk} \leq U_{n:k}$. А поскольку S_{n+1} имеет гамма-распределение с параметрами 1 и $n+1$, то после несложных вычислений получаем

$$\mathbb{E}\gamma_n^2 = O(n^{-1}), \;\; \mathbb{E}\gamma_n^4 = O(n^{-2}). \tag{49}$$

Более того, из вышеприведенных аргументов, (48) и (47) следует, что с вероятностью 1 выполнено $\gamma_n = O(\alpha_n)$ и равномерно по $k \leq n$ с вероятностью 1 имеет место асимптотическое представление

$$\Delta z_{nk} = \frac{1}{n} \frac{\tau_k}{p(F^{-1}(k/n))} + o(1/n). \tag{50}$$

Положим $\delta_n(t, h) = \max\limits_{k:\, |t - z_{n:k}| \leq h} \Delta z_{nk}$ в случае, когда существует хотя бы одна точка $z_{n:k}$ в h-окрестности точки t. В противном случае полагаем $\delta_n(t, h) = 0$. По формуле конечных приращений при всех достаточно больших n с вероятностью 1 выполнено

$$\max_{k\le n}|F^{-1}(U_{n:k})-F^{-1}(k/n)|\le \alpha_n\underline{p}.$$

Следовательно, виду (50) получаем

$$\begin{aligned}\delta_n(t,h)&\le \frac{1}{n}\max_{k:\,|t-F^{-1}(k/n)|\le h+\alpha_n/\underline{p}}\frac{\tau_k}{p(F^{-1}(k/n))}+o(1/n)\le \frac{3\log n}{n\inf_{|s|\le h+\alpha_n/\underline{p}}p(t+s)}+o(1/n),\\ \delta_n&\le \frac{3\log n}{n\underline{p}}+o(1/n),\end{aligned}\tag{51}$$

Здесь мы учли, что $\sup\limits_t \delta_n(t,h)=\delta_n$ при $\delta_n\le h$ и тот факт, что по лемме Бореля–Кантелли с вероятностью 1 имеет место неравенство $\max\limits_{k\le n}\tau_k\le 3\log n$ при всех достаточно больших n. Символ $o(1/n)$ есть стандартный символ «o-малое» почти наверное.

При изучении асимптотики моментов мы будем использовать следующее двойное моментное неравенство, справедливое для произвольной ограниченной случайной величины η:

$$\mathbb{E}\eta^*-2b_o\mathbb{P}(\bar A_n)\le \mathbb{E}\eta\le \mathbb{E}\eta^{**}+2b_o\mathbb{P}(\bar A_n),\tag{52}$$

где A_n есть некоторое событие, η^* и η^{**} являются соответственно нижней и верхней оценками для η на множестве A_n, при этом $\max\{|\eta|,|\eta^*|,|\eta^{**}|\}\le b_o$ с вероятностью 1. Отметим еще, что в случае $\eta^*\ge 0$ п.н. множитель 2 в обеих частях (52) можно опустить.

Сначала изучим математическое ожидание δ_n. В соотношении (52) положим

$$A_n=\Big\{\max_{k\le n+1}|S_k-k|\le \alpha_n n\Big\}.$$

Хорошо известно, что n-ая порядковая статистика $\max\limits_{k\le n}\tau_k$ совпадает по распределению с суммой $\sum_{k=1}^n k^{-1}\tau_k$. Следовательно, $\mathbb{E}\max\limits_{k\le n}\tau_k\le\log n$ при всех $n\ge 2$ и, более того, $\mathbb{E}\max\limits_{k\le n}\tau_k\sim \log n$ при $n\to\infty$. Тем самым, из (47), (48) и (52) получаем, что при всех $n>n_0$

$$\frac{\log n}{2n\sup\limits_{|s|\le h+\alpha_n/\underline{p}}p(t+s)}\le \mathbb{E}\delta_n(t,h)\le \frac{2\log n}{n\inf\limits_{|s|\le h+\alpha_n/\underline{p}}p(t+s)},$$

$$\frac{\log n}{2n\bar p}\le \mathbb{E}\delta_n\le\frac{2\log n}{n\underline{p}},$$

что доказывает (43). Для вывода (42) заметим, прежде всего, что

$$\mathbb{D}\nu_{n,h}(t)=\mathbb{E}\frac{\sum_{k\le n}K_h^2\,(t-z_{n:k})\,(\Delta z_{nk})^2}{\Big(\sum_{j\le n}K_h\,(t-z_{n:j})\,\Delta z_{nj}\Big)^2}\equiv \mathbb{E}\frac{D^2_{n,h}(t)}{J^2_{n,h}(t)}\tag{53}$$

и дробь под знаком математического ожидания ограничена единицей. Тем самым, можно воспользоваться вышеприведенными аргументами, которые были использованы для оценивания подобных математических ожиданий с применением (52), событий A_n и неравенства (47) для получения соответствующих верхних и нижних оценок для числителя и знаменателя дроби в (53). С этой целью прежде всего точно определим o-символ в (51). Итак, ввиду (48) для всех элементарных событий из множества A_n получаем

$$\Delta z_{nk} = \frac{\tau_k}{(n+1)p(F^{-1}(k/n))}\left(1+\zeta_3 2\bar{p}'\underline{p}^{-3}\alpha_n\right)(1+\gamma_n), \tag{54}$$

$$\begin{aligned} D^2_{n,h}(t) &= \sum_{k\colon |t-F^{-1}(k/n)|\le h} K_h^2\left(t-F^{-1}(k/n)\right)(\Delta z_{nk})^2 + \sum_{k\colon |t-F^{-1}(k/n)|\le h} 2\zeta_4\alpha_n L^2 h^{-3}(\Delta z_{nk})^2 = \\ &= \sum_{k\colon |t-F^{-1}(k/n)|\le h} K_h^2\left(t-F^{-1}(k/n)\right)\frac{\tau_k^2}{n^2p^2(F^{-1}(k/n))}(1+4\zeta_5\bar{p}'\underline{p}^{-3}(\tau_k^2\alpha_n^2+\tau_k^2\gamma_n^2))+ \\ &\quad + \sum_{k\colon |t-F^{-1}(k/n)|\le h} 2\zeta_4\alpha_n L^2h^{-3}\frac{\tau_k^2}{n^2p^2(F^{-1}(k/n))}(1+4\zeta_5\bar{p}'\underline{p}^{-3}(\tau_k^2\alpha_n^2+\tau_k^2\gamma_n^2)). \end{aligned} \tag{55}$$

Для завершения асимптотического анализа математического ожидания дроби в правой части в (53) заметим, что ввиду (46), (47) и (54) соответствующую дробь можно изучать только на множестве элементарных исходов

$$A_n^* = A_n \cap \left\{\max_{i\le n}\tau_i \le 3\log n\right\}$$

(в соотношении (52) нужно заменить A_n на A_n^*). Но для указанных элементарных исходов в условиях леммы $J_{n,h}(t) = 1+o(n^{-1/2})$ п.н. и $|o(n^{-1/2})|$ допускает универсальную неслучайную оценку такого же порядка малости. Таким образом, оценка правой части (53) сводится к оценке математического ожидания $\mathbb{E}D^2_{n,h}(t)$. Далее, с учетом оценки $\sup\limits_x(F^{-1}(x))' \le 1/\underline{p}$, число слагаемых в двух последних суммах в (55) имеет порядок $O(nh)$. Следовательно, из (49) и (55) получаем следующее соотношение, завершающее доказательство леммы:

$$\begin{aligned} \mathbb{E}D^2_{n,h}(t) &= \sum_{k\colon |t-F^{-1}(k/n)|\le h} K_h^2\left(t-F^{-1}(k/n)\right)\frac{2}{n^2p^2(F^{-1}(k/n))} + O(\alpha_n^2/(nh^2)) = \\ &= \frac{2}{n}\int_0^1 \frac{K_h^2\left(t-F^{-1}(x)\right)}{p^2(F^{-1}(x))}dx + O(\alpha_n^2/(nh^2)) = \frac{2}{n}\int_0^1\frac{K_h^2\left(t-y\right)}{p(y)}dy + O(\alpha_n^2/(nh^2)) = \\ &= \frac{2}{nh}\int_{-1}^1\frac{K_h^2\left(u\right)}{p(t+uh)}du + O(\alpha_n^2/(nh^2)) = \frac{2}{nhp(t)}\|K\|^2 + O(\log^2 n/(n^2h^2)). \end{aligned}$$

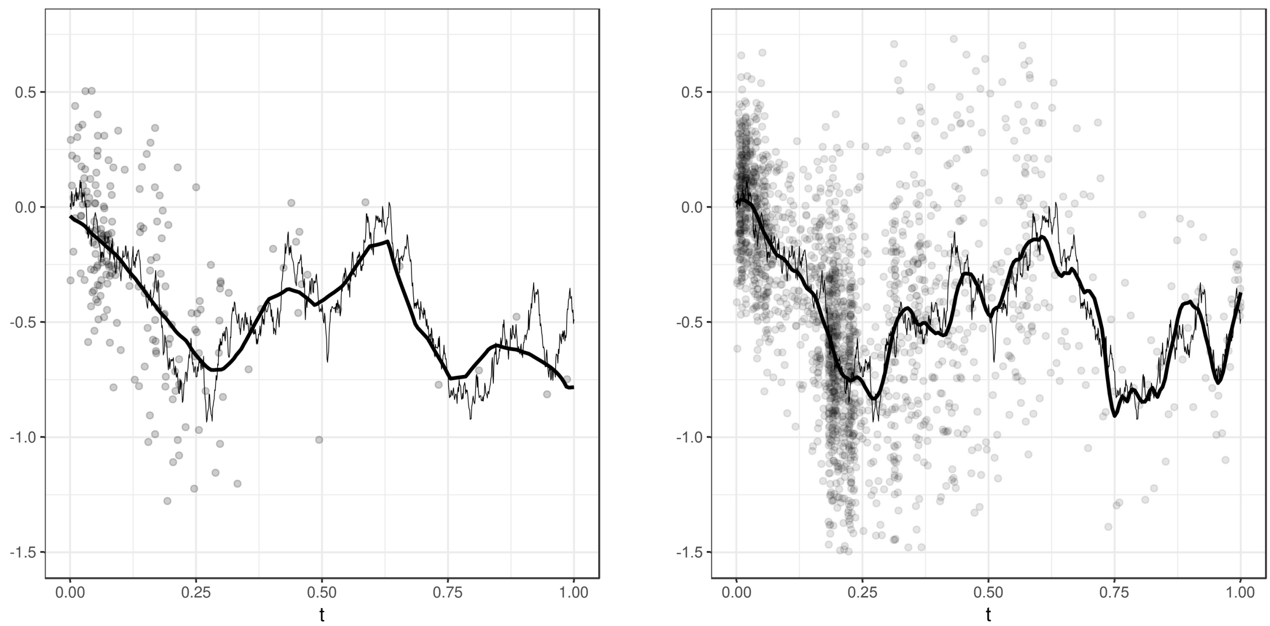
Лемма доказана. □

Д о к а з а т е л ь с т в о леммы 1.6. Повторяя с очевидными изменениями доказательство леммы 1.5 (см. формулу (52) при $\eta = \delta_n^r(t,h)$), нетрудно получить, что $\mathbb{E}\delta_n = O\big(n^{-1}\log n\big)$ при $r > 0$. Это соотношение позволяет проверить дополнительные условия, участвующую в лемме 1.4. Утверждение леммы следует из лемм 1.4 и 1.5. $\square$

1.1.7 Примеры компьютерного моделирования

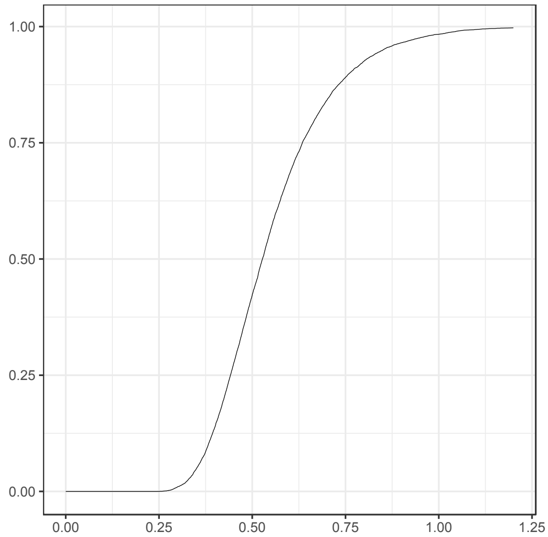
В этом разделе мы приведем два небольших примера. Более обстоятельное компьютерное моделирование, включающее сравнение новых оценок и некоторых известных ранее подходов, будет приведено далее, в разделе 1.2.3, после изучения второго класса новых оценок (универсальных локально–линейных ядерных оценок).

Рассмотрим модель из условия (M_1) при $k = 1$ со стандартным винеровским процессом $\mathrm{w}(t)$ в качестве регрессионной функции $f(t)$. Хорошо известно, что для некоторой собственной случайной величины ζ с вероятностью 1 выполнено $\omega_{\mathrm{w}}(h) \le \zeta(h|\log h|)^{1/2}$. Таким образом, можно воспользоваться утверждением следствия 1.2. Считаем, что погрешности $\{\varepsilon_i\}$ не зависят от винеровского процесса $\mathrm{w}(\cdot)$ и имеют соответствующие нормальные распределения со средним 0 и стандартными уклонениями $\sigma_i = 0.2(3 - 4|z_i - 0.5|)$. Далее рассмотрим два примера случайных регрессоров $\{z_i;\ i = 1, \ldots, n\}$ и проиллюстрируем точность аппроксимации траектории винеровского процесса. В определении универсальной локально–постоянной оценки $f^*_{n,h}(t)$ используется ядро Епанечникова.

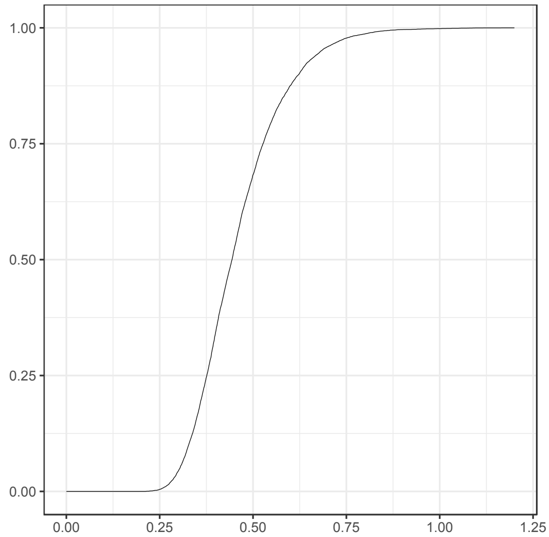


(a) $n = 200,\ \ A = 0.005,\ h = 0.1$ (b) $n = 2000,\ \ A = 0.0005,\ h = 0.03$

Рис. 3: Иллюстрация к примеру 1.3. Изображены графики траектории винеровского процесса (тонкая линия) и функции $f^*_{n,h}(t)$ (жирная линия). Выборка двумерных наблюдений (z_i, X_i) отмечена серыми точками.



(a) $n = 200, \quad A = 0.005, \quad h = 0.1$



(b) $n = 2000, \quad A = 0.0005, \quad h = 0.03$

Рис. 4: Иллюстрация к примеру 1.3. Изображены графики эмпирических функции распределения для $\sup_{t\in[0,1]} |f^*_{n,h}(t) - f(t)|$, построенные по 10000 реализаций.

П р и м е р 1.3. В качестве регрессоров рассматривается последовательность зависимых величин вида $z_i = s(Ai)$, $i = 1, ..., n$, где число $A > 0$ таково, что A/π иррационально,

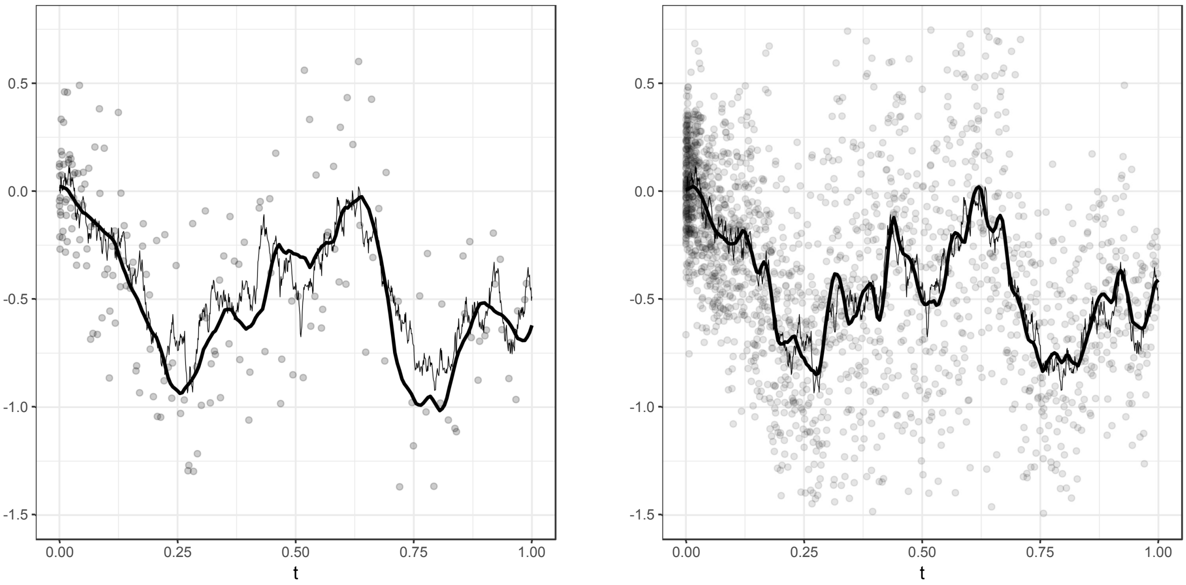
$$s(t) = \Big|\sum_{k=1}^{100} \eta_k \cos(tk)\Big| \quad \text{при} \quad \eta_k = k^{-1}\psi_k \left(\sum_{j=1}^{100} j^{-1}\psi_j\right)^{-1},$$

$\{\psi_j\}$ независимые и равномерно распределенные на отрезке $[0, 1]$ случайные величины, не зависящие от $\mathrm{w}(\cdot)$ и погрешностей $\{\varepsilon_i\}$.

В этом примере случайная последовательность $s(Ai)$ с вероятностью 1 плотно заполняет отрезок $[0, 1]$. В самом деле, при изменении t от 0 до 2π все значения непрерывной периодической функции $s(t)$ с периодом 2π с вероятностью 1 заполняют отрезок $[0, 1]$. Этот факт следует из того, что $s(0) = 1$ и ввиду соотношения $\int_0^{2\pi} \sum\limits_{k=1}^{100} \eta_k \cos(tk)dt = 0$ с вероятностью 1 существует такое $t_0 \in (0, 2\pi)$, что $s(t_0) = 0$. Далее, $s(Ai) = s(\{Ai/(2\pi)\}2\pi)$, где $\{a\}$ обозначает дробную часть a. Тот факт, что последовательность $\{Ai/(2\pi)\}$, $i = 1, 2, \ldots$, образует разбиение отрезка $[0, 1]$, хорошо известен.

На рисунке 3 изображена типичная конфигурация выборки двумерных наблюдений (z_i, X_i) (серые точки) и график оценивающей функции $f^*_{n,h}(t)$ (жирная линия) для траектории винеровского процесса (тонкая линия). На рисунке 4 изображены эмпирические функции распределения для случайной величины $\sup\limits_{t\in[0,1]} |f^*_{n,h}(t) - f(t)|$, построенные по 10000 реализаций (прогонок). $\square$

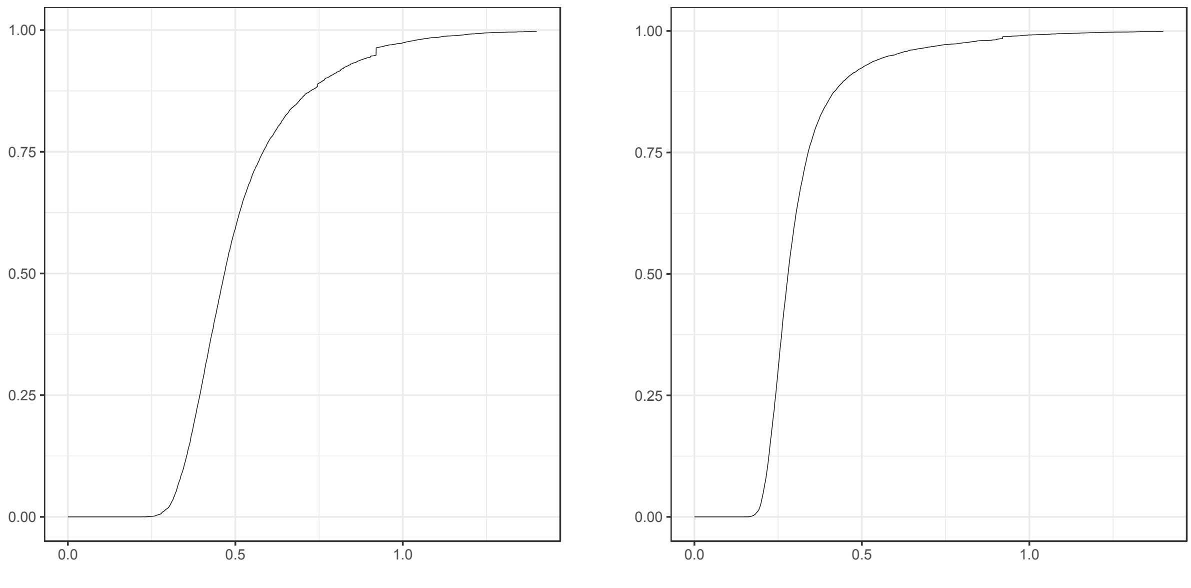
П р и м е р 1.4. Регрессоры задаются формулой $z_i = \{\eta i\}^2$, $i = 1, ..., n$, где случайная величина η равномерна распределена на $[0, 1]$ и не зависит от винеровского процесса $\mathrm{w}(\cdot)$ и



(a) $n = 200, \quad h = 0.05$

(b) $n = 2000, \quad h = 0.02$

Рис. 5: Иллюстрация к примеру 1.4. Изображены графики траектории винеровского процесса (тонкая линия) и функции $f^*_{n,h}(t)$ (жирная линия). Выборка двумерных наблюдений (z_i, X_i) отмечена серыми точками.



(a) $n = 200, \quad h = 0.05$

(b) $n = 2{,}000, \quad h = 0.02$

Рис. 6: Иллюстрация к примеру 1.4. Изображены эмпирические функции распределения для $\sup_{t\in[0,1]} |f^*_{n,h}(t) - f(t)|$, построенные по 10000 реализаций.

погрешностей $\{\varepsilon_i\}$.

На рисунке 5 показаны выборка двумерных наблюдений (серые точки) и график оценивающей функции $f^*_{n,h}(t)$ (жирная линия) для той же траектории винеровского процесса (тонкая линия). На рисунке 6 изображены эмпирические функции распределения для $\sup\limits_{t\in[0,1]} |f^*_{n,h}(t) - f(t)|$, построенные по 10000 реализаций. $\square$

1.2 Универсальные локально–линейные оценки

1.2.1 Основные результаты

Всюду в этом разделе, как и в разделе 1.1.1, предполагаем выполненным условие (M_1) при $k = 1$ и $\mathcal{P} = [0, 1]$. Нам потребуются предположения (E_1) и (K_1) на погрешности и ядро сглаживания, которые мы использовали ранее в разделе 1.1, посвященном универсальным локально–постоянным оценкам.

Как и ранее, через $z_{n:1} \leq \ldots \leq z_{n:n}$ обозначим элементы вариационного ряда, построенного по выборке $\{z_i;\, i = 1, \ldots, n\}$. Положим $z_{n:0} = 0$, $z_{n:n+1} = 1$ и $\Delta z_{ni} = z_{n:i} - z_{n:i-1}$, $i = 1, \ldots, n + 1$. Отклики из уравнения (23), ассоциированные с порядковой статистикой $z_{n:i}$, обозначим через X_{ni}. Наконец, единственное условие на регрессоры $\{z_i\}$, как и прежде, состоит в следующем.

$(\mathbf{D}_1)$ *Имеет место предельное соотношение* $\delta_n = \max\limits_{1\leq i\leq n+1} \Delta z_{ni} \xrightarrow{p} 0$.

В дальнейшем в этом подразделе символом κ_j, $j = 0, 1, 2, 3$, будем обозначать j-й абсолютный момент распределения с плотностью $K(t)$:

$$\kappa_j = \int_{-1}^{1} |u|^j K(u)du. \tag{56}$$

Кроме того, нам потребуются обозначения

$$w_{nj}(t) = \sum_{i=1}^{n} (t - z_{n:i})^j K_h(t - z_{n:i})\Delta z_{ni}, \quad j = 0, 1, 2, 3.$$

Для любого $h \in (0, 1)$ введем в рассмотрение следующий класс оценок для регрессионной функции f:

$$\widetilde{f}^*_{n,h}(t) = I(\delta_n \leq c_* h) \sum_{i=1}^{n} \frac{w_{n2}(t) - (t - z_{n:i})w_{n1}(t)}{w_{n0}(t)w_{n2}(t) - w_{n1}^2(t)} X_{ni} K_h(t - z_{n:i})\Delta z_{ni}, \tag{57}$$

где $I(\cdot)$ — индикаторная функция,

$$c_* \equiv c_*(K) = \frac{\kappa_2 - \kappa_1^2}{96L(6L + \kappa_2 + \kappa_1/2)} < \frac{1}{864L}. \tag{58}$$

З а м е ч а н и е 1.9. Легко видеть, что разность $\kappa_2 - \kappa_1^2$ представляет собой дисперсию невырожденного распределения, тем самым она строго положительна.

З а м е ч а н и е 1.10. Нетрудно проверить, что без индикаторного множителя ядерная оценка (57) является первой координатой двумерной оценки взвешенного метода наименьших квадратов, т.е. двумерной точки, на которой достигается минимум

$$\min_{(a,b)} \sum_{i=1}^{n} \left(X_{ni} - a - b(t - z_{n:i})\right)^2 K_h(t - z_{n:i})\Delta z_{ni}. \tag{59}$$

Таким образом, предлагаемый класс оценок в известном смысле близок к классическим локально–линейным оценкам, но во взвешенном методе наименьших квадратов используются несколько иные веса, определяемые порядковыми статистиками, построенными по набору регрессоров, а вместо исходных наблюдений X_i участвуют наблюдения X_{ni}, ассоциированные с соответствующими порядковыми статистиками.

З а м е ч а н и е 1.11. В случае, когда в наборе $\{z_i;\, i = 1, \dots, n\}$ имеются кратные точки, некоторые спейсинги Δz_{ni} обращаются в ноль, и мы теряем часть выборочной информации в оценке $\widetilde{f}^*_{n,h}(t)$, определенной в (57). В этом случае предлагается прежде, чем использовать оценку (57), несколько сократить выборку, заменив наблюдения X_i с одинаковыми точками z_i их средним арифметическим и оставляя в новой выборке лишь одну точку из кратных. При этом усредненные наблюдения будут иметь меньшее зашумление, так что, несмотря на меньший объем новой выборки, мы не теряем информацию, содержащуюся в исходной выборке. Данное замечание в полной мере относится и к универсальным локально–постоянным оценкам, введенным в разделе 1.1.

Условимся далее в разделе обозначать через C_j, $j \ge 1$, абсолютные положительные постоянные, а через C^*_j — положительные константы, зависящие только от ядра K. Основной результат этого раздела состоит в следующем.

Теорема 1.3. *Пусть выполнено условие (M_1) при $k = 1$ и $\mathcal{P} = [0, 1]$, а также выполнены условия (E_1) и (K_1). Тогда для любого фиксированного $h \in (0, 1/2)$ с вероятностью* 1 *справедлива оценка*

$$\sup_{t \in [0,1]} \left|\widetilde{f}^*_{n,h}(t) - f(t)\right| \le C^*_1 \omega_f(h) + \widetilde{\zeta}_n(h),$$

где случайная величина $\widetilde{\zeta}_n(h)$ такова, что

$$\mathbb{P}\left(\widetilde{\zeta}_n(h) > y,\, \delta_n \le c_* h\right) \le C^*_2 \sigma^2 h^{-2} y^{-2} \mathbb{E}\delta_n,$$

постоянная c_* определена в (58).

З а м е ч а н и е 1.12. Как следует из доказательства теоремы 1.3, константы C^*_1 и C^*_2 имеют следующую структуру:

$$C^*_1 = C_1 \frac{L^2}{\kappa_2 - \kappa_1^2}, \qquad C^*_2 = C_2 \frac{L^4}{(\kappa_2 - \kappa_1^2)^2}.$$

Сравнивая утверждения теорем 1.3 и 1.1, нетрудно видеть, что для оценок $\widetilde{f}^*_{n,t}(t)$ имеют имеют место аналоги следствий 1.1 и 1.2. Кроме того, справедливы замечания 1.4, 1.2 и 1.3 из раздела 1.1. Качественное же отличие универсальных локально–постоянных оценок $f^*_{n,h}(t)$ и универсальных локально–линейных оценок $\widetilde{f}^*_{n,h}(t)$, как мы уже отмечали во введении, наблюдается, например, в окрестностях граничных точек 0 и 1: для оценки $f^*_{n,h}(t)$ в h-окрестностях указанных точек порядок малости смещения h, а для $\widetilde{f}^*_{n,h}(t)$ этот порядок h^2 (см. подробности в следующем разделе).

1.2.2 Сравнение с некоторыми другими ядерными оценками

В разделе 1.1 исследовались универсальные локально–постоянные оценки

$$f^*_{n,h}(t) = \frac{\sum_{i=1}^n X_{ni} K_h(t - z_{n:i}) \Delta z_{ni}}{\sum_{i=1}^n K_h(t - z_{n:i}) \Delta z_{ni}} \equiv \frac{\sum_{i=1}^n X_{ni} K_h(t - z_{n:i}) \Delta z_{ni}}{w_{n0}(t)}, \tag{60}$$

определяемые также соотношением

$$f^*_{n,h}(t) \equiv \arg\min_a \sum_{i=1}^n (X_{ni} - a)^2 K_h(t - z_{n:i}) \Delta z_{ni}. \tag{61}$$

Интересно сравнить универсальные локально–линейные оценки $\widetilde{f}^*_{n,h}(t)$ с универсальными локально–постоянными оценками $f^*_{n,h}(t)$, а также другими оценками (например, оценками Надарая–Ватсона $\widehat{f}_{NW}(t)$ и классическими локально–линейными оценками $\widehat{f}_{LL}(t)$). Всюду в этом подразделе считаем, что в условии (M_1) при $k = 1$ регрессионная функция f неслучайна и справедливо предположение (K_1) относительно ядра сглаживания. Кроме того, нам потребуется следующее ограничение.

(**IID**) *Регрессионная функция f дважды непрерывно дифференцируема, погрешности $\{\varepsilon_i\}$ независимы, одинаково распределены, центрированы и не зависят от регрессоров $\{z_i\}$, которые, в свою очередь, независимы и одинаково распределены. Кроме того, функция распределения случайной величины z_1 имеет непрерывно дифференцируемую на $(0,1)$ строго положительную плотность $p(t)$.*

Столь жесткие ограничения на параметры регрессионной модели объясняются как проблемами в вычислении асимптотического представления для дисперсий оценок $\widetilde{f}^*_{n,h}(t)$ и $f^*_{n,h}(t)$, так и известными свойствами оценок Надарая–Ватсона.

Для любой статистической оценки $\tilde{f}_n(t)$ регрессионной функции $f(t)$ будем использовать обозначение $\mathrm{Bias}\tilde{f}_n(t)$ для ее смещения:

$$\mathrm{Bias}\tilde{f}_n(t) = \mathbb{E}\tilde{f}_n(t) - f(t).$$

Положим $\overline{f} = \sup_{t\in[0,1]} |f(t)|$ и при $j = 0,1,2,3$ введем обозначение

$$w_j(t) = \int\limits_0^1 (t-z)^j K_h(t-z)dz = \int\limits_{z\in[0,1]:|t-z|\le h} (t-z)^j K_h(t-z)dz, \quad t\in[0,1]. \tag{62}$$

Нам также потребуются обозначения

$$\|K\|^2 = \int_{-1}^1 K^2(u)du, \qquad \kappa_j(\alpha) = \int_{-1}^\alpha t^j K(t)dt, \quad \alpha\in[0,1], \quad j=0,1,2,3.$$

Асимптотическое представление для смещения и дисперсии оценки $f^*_{n,h}(t)$ установлено в разделе 1.1:

Предложение 1.3. *Пусть выполнено условие* (IID) *и* $\inf_{t\in[0,1]} p(t) > 0$. *Если* $n\to\infty$ *и* $h\to 0$ *так, что выполнено* $(\log n)^{-1}h\sqrt{n}\to\infty$, $h^{-2}\mathbb{E}\delta_n\to 0$ *и* $h^{-3}\mathbb{E}\delta_n^2\to 0$, *то при любом* $t\in(0,1)$ *справедливы соотношения*

$$\mathrm{Bias} f^*_{n,h}(t) = \frac{h^2\kappa_2}{2}f''(t) + o(h^2), \qquad \mathbb{D}f^*_{n,h}(t) \sim \frac{2\sigma^2}{hnp(t)}\|K\|^2.$$

Отметим, что первое утверждение касательно асимптотики смещения в предложении 1.3 на самом деле было доказано для произвольно зависимых регрессоров при выполнении условия (D_1). Два нижеследующих предложения и следствия также получены без каких-либо предположений касательно корреляции регрессоров, используется лишь условное центрирование и условная ортогональность погрешностей из условия (E_1).

Предложение 1.4. *Пусть* $h < 1/2$. *Тогда для любого фиксированного* $t\in[h,1-h]$

$$\mathrm{Bias}\widetilde{f}^*_{n,h}(t) = \mathrm{Bias} f^*_{n,h}(t) + \gamma_{n,h}(t), \quad \mathbb{D}\widetilde{f}^*_{n,h}(t) = \mathbb{D}f^*_{n,h}(t) + \rho_{n,h}(t),$$

где

$$|\gamma_{n,h}(t)| \le C_3^*\overline{f}h^{-1}\mathbb{E}\delta_n, \quad |\rho_{n,h}(t)| \le C_4^*\big(\sigma^2 + \overline{f}^2\big)h^{-1}\mathbb{E}\delta_n.$$

Предложение 1.5. *Пусть регрессионная функция* f *дважды непрерывно дифференцируема. Тогда для любого фиксированного* $t\in(0,1)$

$$\mathrm{Bias}\widetilde{f}^*_{n,h}(t) = \frac{f''(t)}{2}B_0(t) + O(\mathbb{E}\delta_n/h) + o(h^2), \tag{63}$$

где

$$B_0(t) = \frac{w_2^2(t) - w_3(t)w_1(t)}{w_0(t)w_2(t) - w_1^2(t)}. \tag{64}$$

Кроме того,

$$\mathrm{Bias} f^*_{n,h}(t) = -f'(t)\frac{w_1(t)}{w_0(t)} + \frac{f''(t)}{2}\frac{w_2(t)}{w_0(t)} + O(\mathbb{E}\delta_n) + o(h^2), \tag{65}$$

при этом погрешности $o(\cdot)$ *и* $O(\cdot)$ *в* (63) *и* (65) *равномерны по* t.

Следствие 1.3. *Пусть регрессионная функция* f *дважды непрерывно дифференцируема,* $h \to 0$ *и* $h^{-3}\mathbb{E}\delta_n \to 0$. *Тогда при любом фиксированном* $t \in (0,1)$, *для которого* $f''(t) \neq 0$, *имеет место цепочка асимптотических соотношений*

$$\mathrm{Bias}\widetilde{f}^*_{n,h}(t) \sim \mathrm{Bias} f^*_{n,h}(t) \sim \frac{f''(t)}{2}\kappa_2 h^2.$$

Следствие 1.4. *Пусть в условиях следствия* 1.3 *функция* f *имеет в окрестности нуля ненулевые первую и вторую производные. Тогда для любого фиксированного положительного* $\alpha < 1$ *такого, что* $\kappa_1(\alpha) < 0$, *имеют место асимптотические соотношения*

$$\mathrm{Bias}\widetilde{f}^*_{n,h}(\alpha h) \sim \frac{1}{2}h^2 D(\alpha) f''(0+), \qquad \mathrm{Bias} f^*_{n,h}(\alpha h) \sim -h\frac{\kappa_1(\alpha)}{\kappa_0(\alpha)} f'(0+),$$

где

$$D(\alpha) = \frac{\kappa_2^2(\alpha) - \kappa_3(\alpha)\kappa_1(\alpha)}{\kappa_0(\alpha)\kappa_2(\alpha) - \kappa_1^2(\alpha)}.$$

Отметим, что в силу неравенства Коши–Буняковского и свойств плотности $K(\cdot)$ имеет место строгое неравенство $\kappa_0(\alpha)\kappa_2(\alpha) - \kappa_1^2(\alpha) > 0$ для любого $\alpha \in [0,1]$.

З а м е ч а н и е 1.13. Аналогичные соотношения имеют место и в окрестности правой границы отрезка $[0,1]$, когда $t = 1 - \alpha h$ при любом $\alpha \leq 1$. В этом случае в вышеприведенных асимптотиках нужно просто заменить правосторонние производные в нуле на аналогичные (ненулевые) левосторонние производные в точке 1, а вместо величин $\kappa_j(\alpha)$ надо подставить $\tilde{\kappa}_j(\alpha) = \int_{-\alpha}^{1} v^i K(v)dv = (-1)^j \kappa_j(\alpha)$. При этом коэффициент $D(\alpha)$ не изменится, а соответствующий коэффициент в правой части второй асимптотики изменит лишь знак.

Таким образом, в сделанных предположениях качественное отличие новых универсальных оценок $f^*_{n,h}(t)$ и $\widetilde{f}^*_{n,h}(t)$ наблюдается только в окрестностях граничных точек 0 и 1: для оценки $f^*_{n,h}(t)$ в h-окрестностях указанных точек порядок малости смещения h, а для $\widetilde{f}^*_{n,h}(t)$ этот порядок h^2. Такая связь между оценками $f^*_{n,h}(t)$ и $\widetilde{f}^*_{n,h}(t)$, определенными соответственно в (57) и (60), представляется весьма естественной ввиду соотношений (59) и (61) и известной взаимосвязи в граничных точках между классическими оценками Надарая–Ватсона $\hat{f}_{NW}(t)$ и локально–линейными оценками $\widehat{f}_{LL}(t)$.

З а м е ч а н и е 1.14. Если выполнено условие (IID), то для смещения и дисперсии классических локально–постоянных и локально–линейных оценок $\widehat{f}_{NW}(t)$ и $\widehat{f}_{LL}(t)$ хорошо известны (см., например, [77] и предложение 1.1) следующие асимптотические представления, справедливые в широких условиях для любого $t \in (0,1)$:

$$
\begin{aligned}
\mathrm{Bias}\widehat{f}_{NW}(t) &= \frac{h^2\kappa_2}{2p(t)}\left(f''(t)p(t) + 2f'(t)p'(t)\right) + o(h^2), & \mathbb{D}\widehat{f}_{NW}(t) &\sim \frac{\sigma^2}{hnp(t)}\|K\|^2,\\
\mathrm{Bias}\widehat{f}_{LL}(t) &= \frac{h^2\kappa_2}{2}f''(t) + o(h^2), & \mathbb{D}\widehat{f}_{LL}(t) &\sim \frac{\sigma^2}{hnp(t)}\|K\|^2.
\end{aligned}
$$

Приведенные асимптотические представления показывают, что при выполнении предположения (IID) дисперсия оценки Надарая–Ватсона $\hat{f}_{NW}(t)$ и локально-линейной оценки $\hat{f}_{LL}(t)$ в широких условиях асимптотически в два раза меньше дисперсии оценок $f^*_{n,h}(t)$ и $\widetilde{f}^*_{n,h}(t)$. Но среднеквадратичная погрешность любой оценки равна сумме дисперсии и квадрата смещения, так что если стандартное отклонение погрешностей σ не сильно велико и справедливо неравенство

$$
|f''(t)p(t) + 2f'(t)p'(t)| > |f''(t)p(t)|\,,
$$

то оценка $f^*_{n,h}(t)$ или $\widetilde{f}^*_{n,h}(t)$ может быть точнее, чем $\hat{f}_{NW}(t)$ с среднеквадратичном смысле. Указанный эффект для оценки $f^*_{n,h}(t)$ подтверждают результаты компьютерного моделирования, приведенные в разделе 1.1. $\square$

З а м е ч а н и е 1.15. Универсальные локально–линейные и локально–постоянные оценки $\widetilde{f}^*_{n,h}(t)$ и $f^*_{n,h}(t)$, заданные в (57) и (60), можно определить несколько иначе, в зависимости от выбора того или иного разбиения с выделенными точками $\{z_i;\ i = 1,\dots,n\}$ области определения регрессионной функции, лежащего в основе этих оценок. Например, используя разбиение Вороного отрезка $[0,1]$, универсальную локально–постоянную оценку можно задать равенством

$$
f^*_{n,h}(t) = \frac{\sum_{i=1}^{n} X_{ni}K_h(t - z_{n:i})\widetilde{\Delta} z_{ni}}{\sum_{i=1}^{n} K_h(t - z_{n:i})\widetilde{\Delta} z_{ni}}, \tag{66}
$$

где $\widetilde{\Delta} z_{n1} = \Delta z_{n2}$, $\widetilde{\Delta} z_{nn} = \Delta z_{nn}$, $\widetilde{\Delta} z_{ni} = (\Delta z_{ni} + \Delta z_{ni+1})/2$ при $i = 2,\dots,n-1$. Просматривая доказательства утверждений раздела 1.1, нетрудно видеть, что в этом случае все свойства оценки $f^*_{n,h}(t)$ сохраняются, за исключением асимптотического представления для дисперсии. Повторяя с очевидными изменениями вывод предложений 1.5 и 1.6, имеем

$$
\mathbb{D}f^*_{n,h}(t) \sim \frac{1.5\sigma^2}{hnp(t)}\|K\|^2.
$$

Таким образом, асимптотическая дисперсия оценки может быть несколько уменьшена за счет выбора того или иного разбиения (по-крайней мере, в случае независимых и одинаково распределенных регрессоров).

Аналогично, в определении (57) оценки $\widetilde{f}^*_{n,h}(t)$ величины $\{\Delta z_{ni}\}$ можно заменить разбиением Вороного $\{\widetilde{\Delta} z_{ni}\}$. Стоит также отметить, что индикаторный множитель, участвующий в определении (57) оценки $\widetilde{f}^*_{n,h}(t)$, не влияет на асимптотические свойства оценки, приведенные в теореме 1.3, и потребовался нам лишь для вычисления точной асимптотики смещения оценки. Так что с практической точки зрения универсальные локально–линейные оценки можно задавать равенством

$$\widetilde{f}^*_{n,h}(t) = (1,0)\ \arg\min_{a,b} \sum_{i=1}^{n} \left(X_{ni} - a - b(t - z_{n:i})\right)^2 K_h(t - z_{n:i})\widetilde{\Delta} z_{ni}, \tag{67}$$

где величины $\widetilde{\Delta} z_{ni}$ определены после формулы (66). $\square$

1.2.3 Примеры компьютерного моделирования и обработки реальных данных

В данном разделе рассматривается несколько примеров регрессионных моделей. С помощью компьютерного моделирования сравниваются новые универсальные локально–постоянные (ULC), новые универсальные локально–линейные оценки (ULL) и два популярных представителя известных ранее ядерных оценок: оценки Надарая–Ватсона (NW) и классические локально–линейные оценки (LOESS). Моделирование также включает тестирование следующих основных алгоритмов статистического обучения: линейная регрессия (LR), обобщенная аддитивная модель (GAM) и случайный лес (RF) (см., например, [119]). В скобках указана аббревиатура, которую мы будем использовать на соответствующих графических иллюстрациях.

При моделировании универсальные локально–постоянные оценки $f^*_{n,h}(t)$ и универсальные локально–линейные оценки $\widetilde{f}^*_{n,h}(t)$ вычислялись по формулам (66) и (67). Для вычисления классических локально–линейных оценок использовалась стандартная функция R-библиотеки loess(). Во всех оценках ядерного типа, к которым относятся два класса новых оценок, оценки Надарая-Ватсона и классические локально–линейные оценки, использовалось трикубическое ядро

$$K(t) = \frac{70}{81}\max\{0, (1 - |t|^3)^3\}.$$

Выбор трикубического ядра объясняется тем, что это ядро используется в функции R loess(), которая применялась при расчетах. Обучение обобщенной аддитивной модели и случайного леса проводилось с использованием R-библиотек mgcv и randomForest.

Точность рассматриваемых оценок регрессионной функции сравнивалась в смысле максимальной и среднеквадратичной ошибок. Вычисление максимальной ошибки проводилось на равномерной сетке из 1001 точки на отрезке $[0, 10]$ по формуле

$$\max_{j=1,\ldots,1001} |\check{f}(t_j) - f(t_j)|,$$

где t_j — точки равномерной сетки из отрезка $[0, 10]$, $t_1 = 0$, $t_{1001} = 10$, $\check{f}(t_j)$ — значения той или иной оценки в точках сетки разбиения, $f(t_j)$ — истинные значения оцениваемой функции. Точность приближения в среднеквадратичном смысле рассчитывалась для одного случайного разбиения исходной выборки на обучающую и проверочную в пропорции 60% к 40%, по формуле

$$\frac{1}{m} \sum_{j \in N_m} \left(\check{f}(z_j) - X_j) \right)^2,$$

где m — размер проверочной выборки, N_m — множество индексов наблюдений, составляющих проверочную выборку объема m, z_j — регрессоры проверочной выборки, X_j — зашумленные наблюдения значений регрессионной функции в точках проверочной выборки, $\check{f}$ — оценка, построенная по обучающей выборке. Разбиение на обучающую и проверочную выборки было одно и то же для всех рассматриваемых оценок регрессионной функции.

Для каждой из ядерных оценок размер окна h определялся при помощи перекрестной проверки (кросс-валидации). При вычислении среднеквадратичной ошибки кросс-валидация проводилась по обучающей выборке, а при вычислении максимальной ошибки — по всей выборке. Для новых универсальных оценок и оценок Надарая-Ватсона параметр h выбирался из 20 значений, расположенных на логарифмической сетке от $\max\{0.0001, 1.1 \max_i \Delta z_{ni}\}$ до 0.9. Для классических локально–линейных оценок параметр `span` выбирался таким же образом из 20 значений, расположенных на логарифмической сетке от 0.0001 до 0.9.

В каждом примере было выполнено по 1000 реализаций (прогонок), в каждой из реализаций моделировалось по $n = 5000$ наблюдений. По каждой реализации вычислялись максимальная и среднеквадратичная ошибки для каждой из оценок. Результаты вычислений представлены ниже на графиках «ящик с усами», где «ящик» отображает медиану и 25% и 75% процентили ($0,25$ и $0,75$ квантили) эмпирического распределения. На графиках не приводятся результаты для линейной регрессии, потому что в этих примерах они оказывались значительно хуже, чем результаты для всех остальных вариантов оценок функции. Среднеквадратичные и максимальные ошибки тех или иных оценок сравнивались также при помощи парного теста Вилкоксона.

Во всех приводимых далее примерах погрешности $\{\varepsilon_i\}$ гауссовские со стандартным отклонением $\sigma = 2$. Примеры этого раздела были выбраны таким образом, чтобы распределение регрессоров оказалось «сильно неравномерным». Потенциально это могло продемонстрировать преимущество новых оценок перед известными ранее.

П р и м е р 1.5. Зададим регрессионную функцию

$$f(z) = (z-5)^2 + 10, \quad 0 \le z \le 10. \tag{68}$$

На отрезке $z \in [0, 5]$ расположено 4500 точек, независимых равномерно распределенных на этом отрезке; а на отрезке $z \in [5, 10]$ — 500 точек, также независимых и равномерно распределенных на этом отрезке. График регрессионной функции и одна из реализаций наблюдений представлены на рисунке 7.

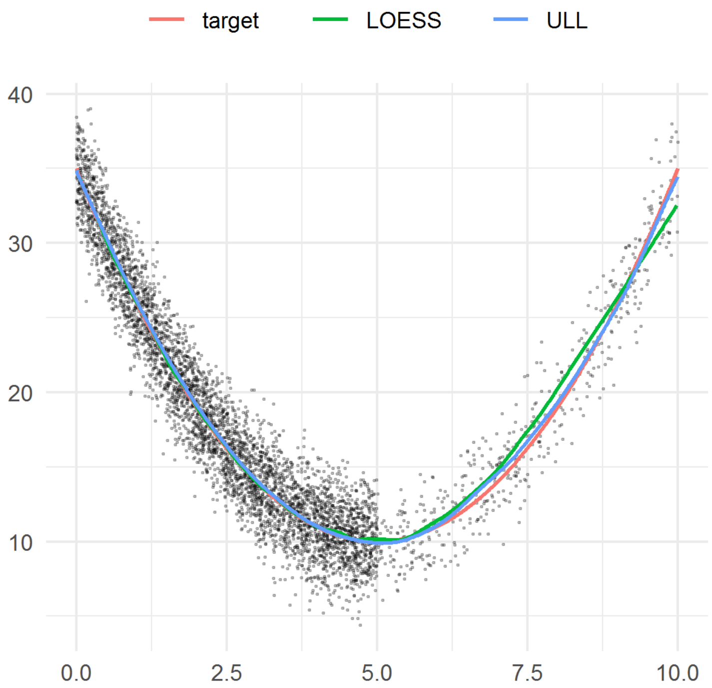


Рис. 7: Пример 1.5. Регрессионная функция, реализация наблюдений и локально–линейные оценки (универсальная ULL и классическая LOESS), построенные по этой реализации.

Результаты расчетов приведены на рисунке 8. В этом примере при сравнении максимальных ошибок заметно преимущество локально–линейных оценок (универсальных и классических) перед локально–постоянными ядерными оценками (оценками Надарая-Ватсона и универсальными), при этом новая универсальная локально–линейная оценка оказалась точнее всех рассматриваемых оценок. В частности, универсальная локально–линейная оценка точнее классической локально–линейной оценки (0.6357 (0.4993, 0.8224) против 0.6582 (0.5205, 0.8508), p=0.019). При сравнении оценок в среднеквадратичном смысле все оценки, кроме случайного леса и линейной регрессии, имеют близкие результаты. Универсальная локально–линейная оценка оказалась несколько точнее остальных, хотя ее отличие от классической локально–линейной статистически незначимо (4.017 (3.896, 4.139) против 4.030 (3.906, 4.154), p=0.11).

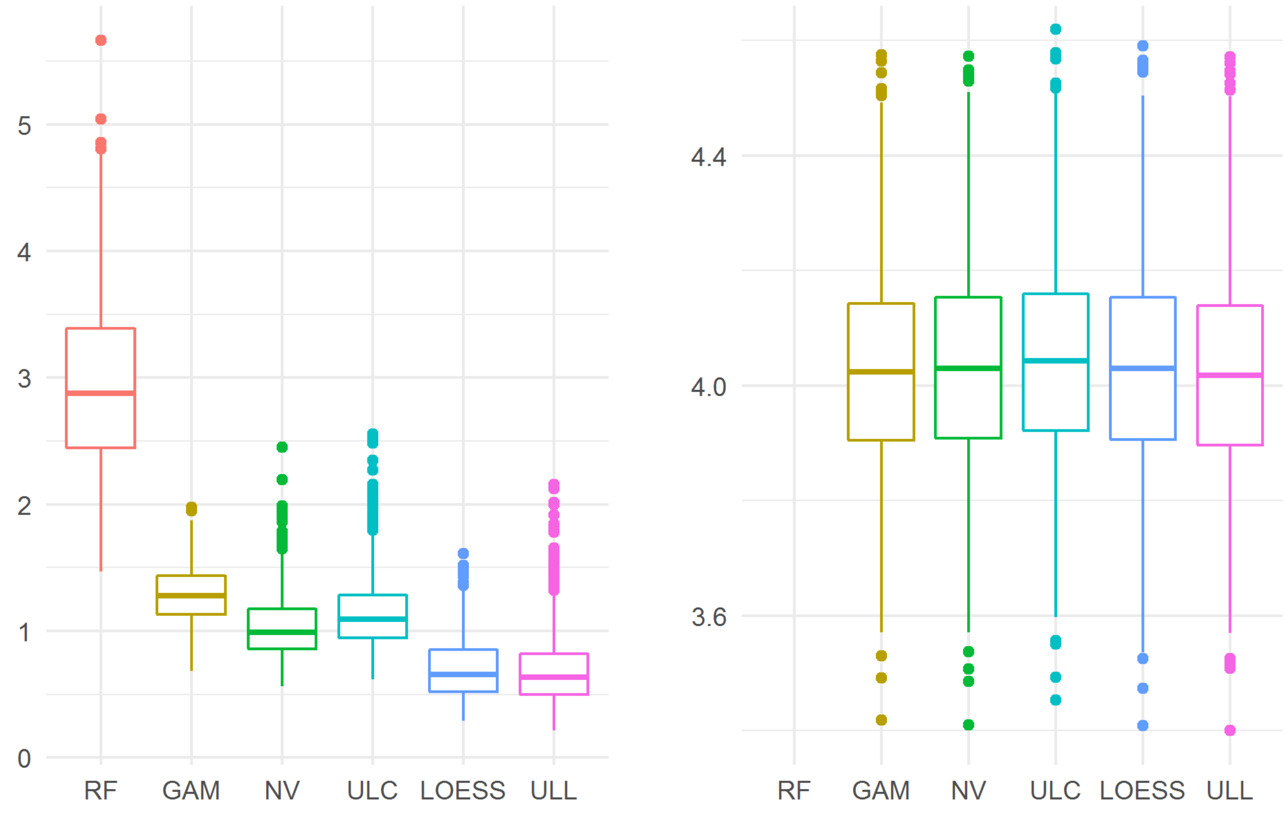


Рис. 8: Максимальная (слева) и среднеквадратичная (справа) ошибки примера 1.5. Для среднеквадратичной ошибки модель случайного леса показала худшие результаты (10.97 (10.55, 11.39)), поэтому характеристики модели случайного леса не удалось представить на графике.

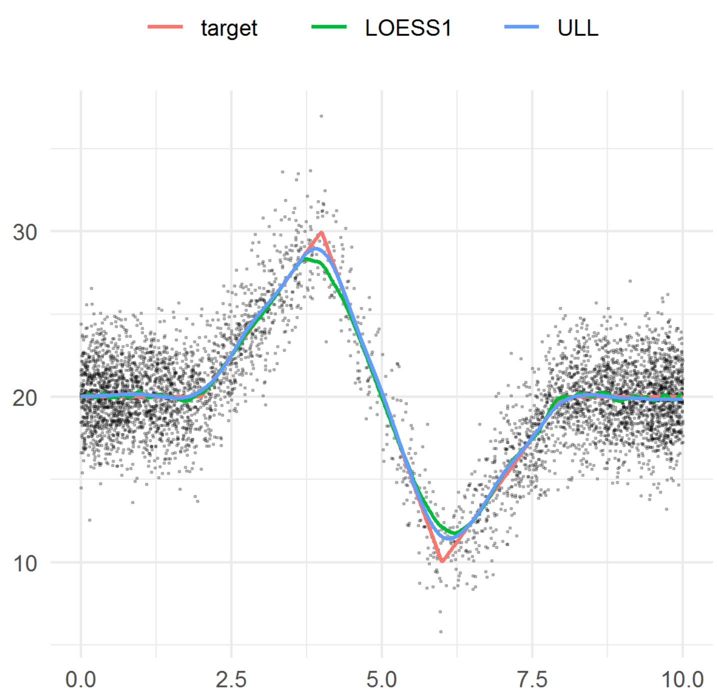


Рис. 9: Пример 1.6. Регрессионная функция, реализация наблюдений и локально–линейные оценки (ULL и LOESS), построенные по этой реализации.

П р и м е р 1.6. Кусочно-линейная регрессионная функция изображена на рисунке 9. Регрессоры в этом примере независимы и одинаково распределены с плотностью, пропорциональной $(z-5)^2+2$, $0 \le z \le 10$.

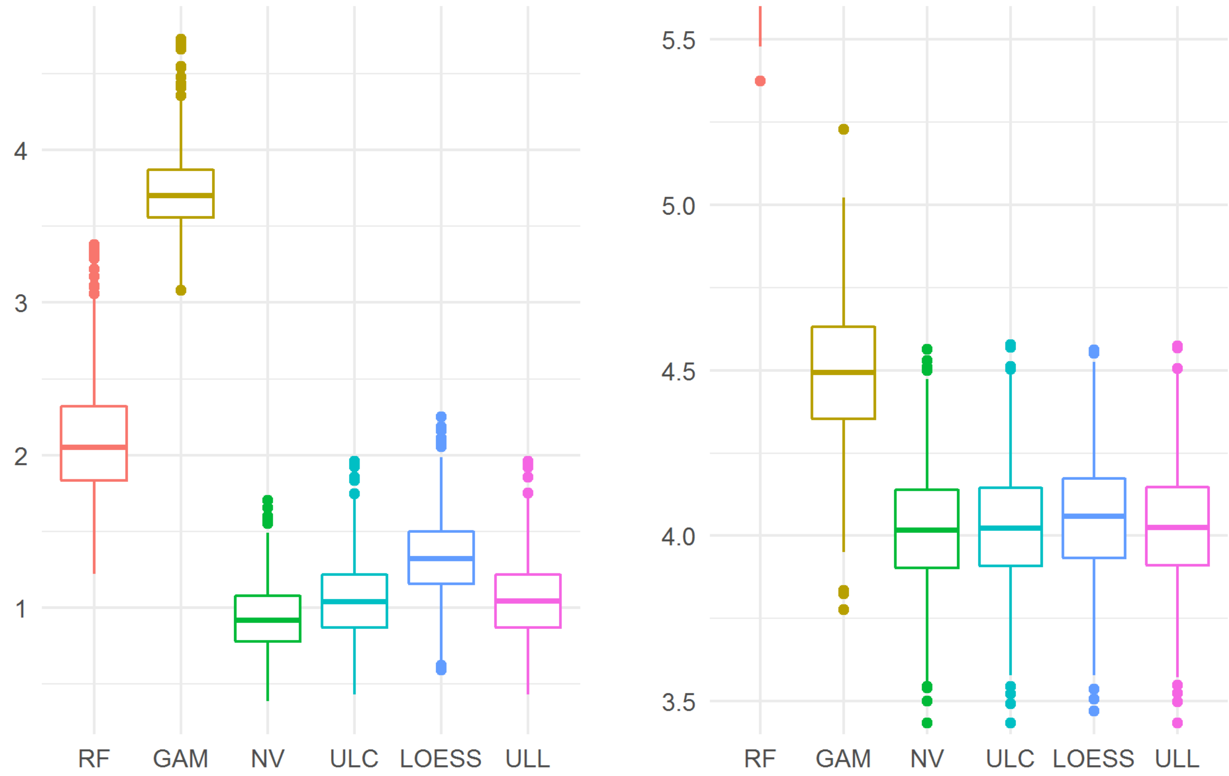


Рис. 10: Максимальная (слева) и среднеквадратичная (справа) ошибки примера 1.6. Для среднеквадратичной ошибки модель случайного леса показала худшие результаты (6.699 (6.412, 7.046)) по сравнению с моделью GAM и ядерными оценками, поэтому результаты модели случайного леса не удалось представить на графике.

Результаты расчетов приведены на рисунке 10. В этом примере наилучшей оказалась оценка Надарая-Ватсона как при сравнении максимальных ошибок, так и среднеквадратичном смысле. В обоих случаях универсальная локально–линейная оценка точнее классической локально–линейной оценки (p<0.0001 для максимальной ошибки и p=0.0030 для среднеквадратичной ошибки).

П р и м е р 1.7. Рассматривается следующая регрессионная функция

$$f(z) = 0.2\ \big(((z-5)^2+25)\ \cos((z-5)^2/2)\ +\ 60\big), \qquad z \in [0, 10],$$

график которой представлен на рисунке 11. В этом примере регрессоры определены следующим образом: $z_i = 10 \cdot s(Ai)$, $i = 1, \ldots, n$, где положительное число A таково, что A/π иррационально (мы выбрали $A = 0.0002$),

$$s(t) = \Big|\sum_{k=1}^{100} \eta_k \cos(tk)\Big| \quad \text{при} \quad \eta_k = k^{-1}\psi_k \left(\sum_{j=1}^{100} j^{-1}\psi_j\right)^{-1},$$

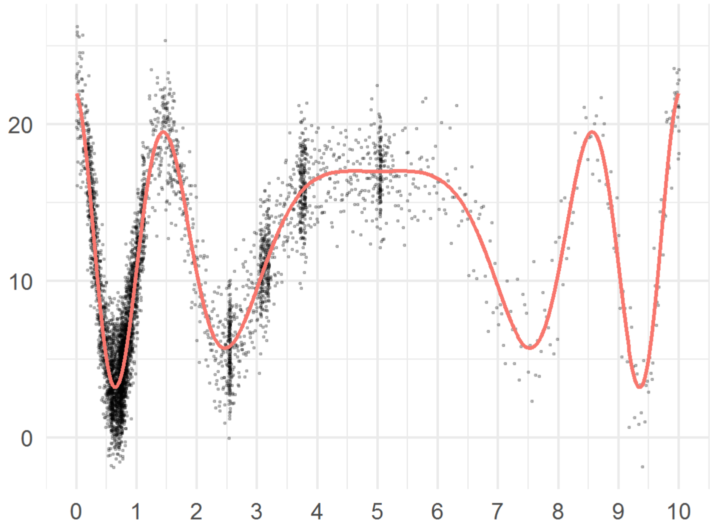


Рис. 11: Пример 1.7. Регрессионная функция и реализация наблюдений

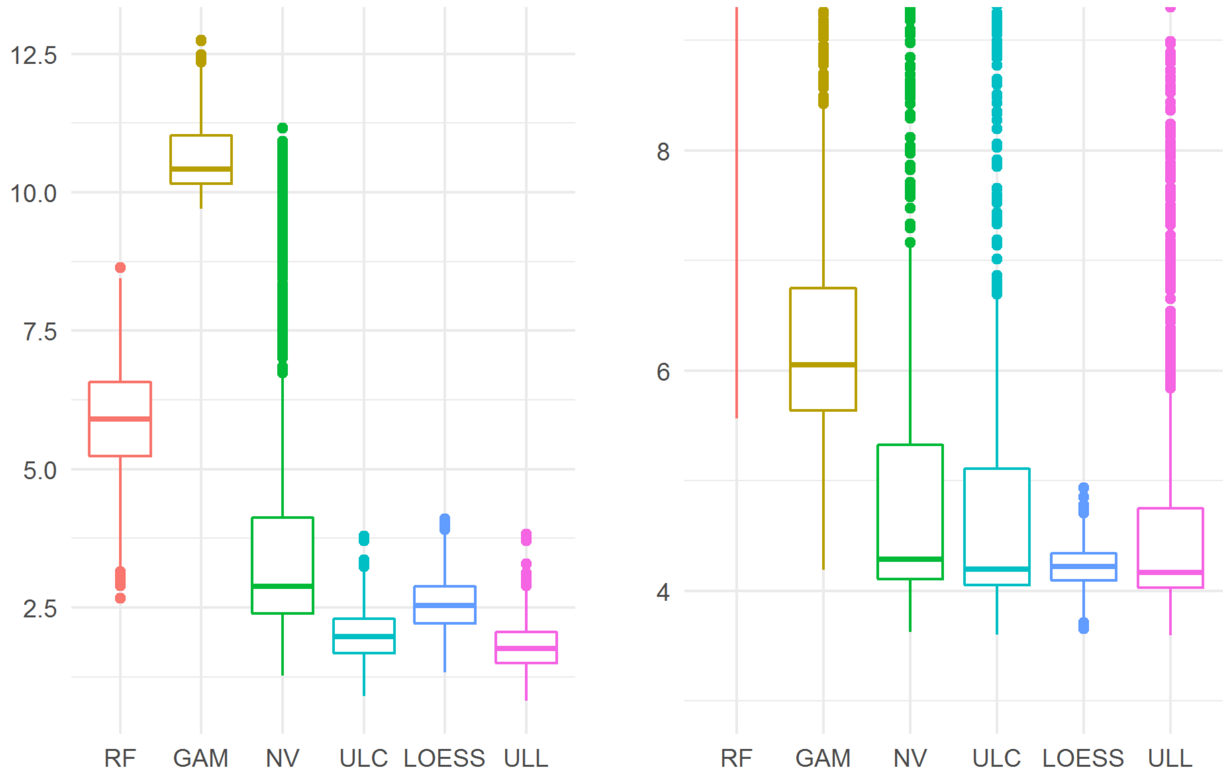


Рис. 12: Максимальная (слева) и среднеквадратичная (справа) ошибка примера 1.7. Как и ранее, для среднеквадратичной ошибки результаты модели случайного леса (13.95 (11.69, 16.18)) не показаны полностью на графике. Кроме того, на этом графике «обрезаны» выбросы для оценок GAM, NV, ULC, ULL.

$\{\psi_j\}$ независимые и равномерно распределенные на отрезке $[0,1]$ случайные величины, не зависящие от погрешностей $\{\varepsilon_i\}$. Отметим, что случайная последовательность $s(Ai)$ с вероятностью 1 плотно заполняет отрезок $[0,1]$ (см. подробности в примере 1.3).

Результаты расчетов приведены на рисунке 12. Сравнивая в этом примере оценки в смысле максимальной ошибки, отметим, что универсальная локально–линейная оценка точнее других рассматриваемых оценок. В частности, эта оценка лучше классической локально–

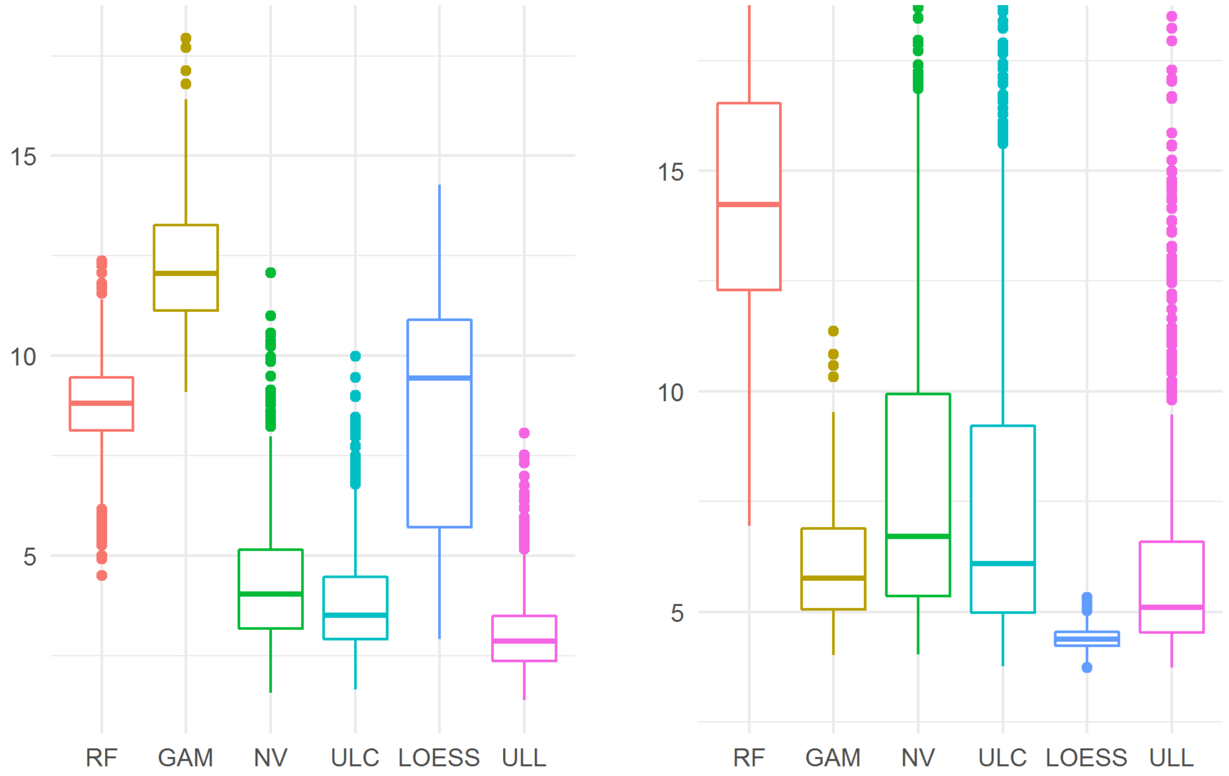


Рис. 13: Максимальная (слева) и среднеквадратичная (справа) ошибка примера 1.8. Как и ранее, для среднеквадратичной ошибки результаты модели случайного леса не показаны полностью на графике. Кроме того, на этом графике «обрезаны» выбросы для оценок NV, ULC, ULL.

линейной оценки (1.757 (1.491, 2.053) против 2.538 (2.216, 2.886), $p<0.0001$). Медиана среднеквадратичной ошибки новой универсальной локально–линейной оценки также оказывается наименьшей из рассмотренных. В частности, в этом смысле универсальная локально–линейная оценка точнее классической локально–линейной оценки, но это различие незначимо (4.166 (4.025, 4.751) против 4.219 (4.096, 4.338), $p=0.92$).

П р и м е р 1.8. Регрессионная функция в этом примере та же, что и в примере 1.7. Далее, по алгоритму примера 1.7 генерировалось 50000 точек $\{z_i\}$, из которых выбиралось 5000 точек. Подобный алгоритм позволил сделать заполнение регрессорами области задания f более равномерным, чем в предыдущем примере, при этом сохраняя «кластеры».

Результаты расчетов приведены на рисунке 13. При сравнении максимальных ошибок новая универсальная локально–линейная оценка оказалась точнее других оценок. В частности, универсальная локально–линейная оценка точнее классической локально–линейной оценки (2.872 (2.369, 3.488) против 9.435 (5.719, 10.9), $p<0.0001$). При сравнении оценок в среднеквадратичном смысле самой точной оказалась классическая локально–линейная оценка. Универсальная локально–линейная оценка проигрывает здесь классической (5.108 (4.535, 6.597) против 4.378 (4.229, 4.541), $p<0.0001$), но оказывается точнее остальных рассматриваемых оценок.

Полезно отметить, что во всех приведенных примерах оценки ядерного типа (как новые, так и классические) оказались точнее остальных рассмотренных оценок, при этом нередко новые оценки (универсальные локально–постоянные или универсальные локально–линейные) оказались наиболее точными из ядерных.

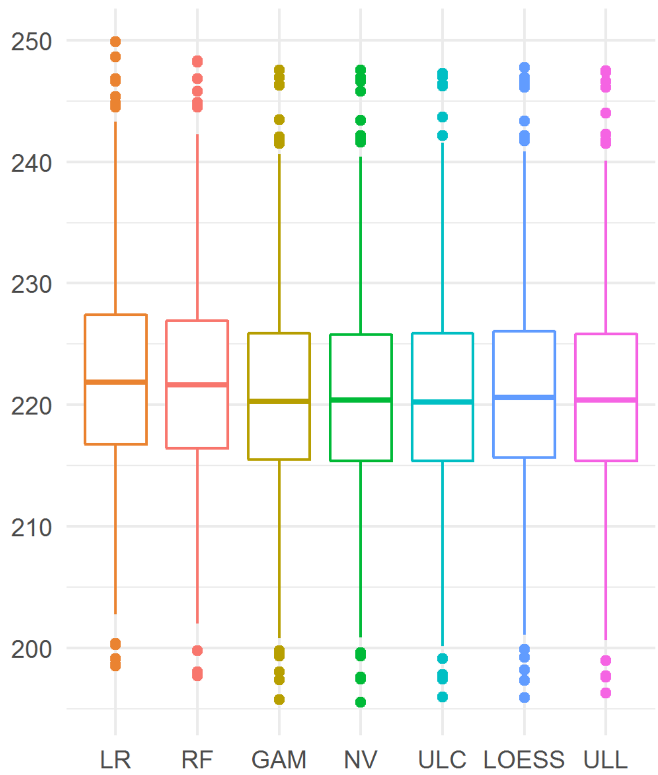


Рис. 14: Пример 1.9. Среднеквадратичная ошибка предсказания зависимости АД от ЧСС

В следующем примере сравнение оценок проводилось на реальных данных.

П р и м е р 1.9. В данном примере использовались данные исследования «Эпидемиология сердечно-сосудистых заболеваний в регионах Российской Федерации (ЭССЕ-РФ)». Это исследование, в котором были собраны те или иных характеристики неорганизованного мужского и женского населения в возрасте от 25 до 64 лет из 13 регионов РФ, подробно представлено в [257].

Мы выбрали в качестве отклика (предсказываемой переменной) уровень систолического артериального давления (АД), а в качестве регрессора — частоту сердечных сокращений (ЧСС). Взаимосвязь между этими переменными ранее оценивалась как нелинейная (см. [258]). В анализе используются данные 6597 участников исследования из четырех регионов РФ. Уровни АД и ЧСС в выбранных регионах попарно статистически значимо различались. Тем самым, нарушалась гипотеза о независимости набора регрессоров.

Поскольку точный вид регрессионной функции, участвующий в определении максимальной ошибки, нам неизвестен, то оценки в этом примере сравнивались только в среднеквадратичном смысле. Среднеквадратичная ошибка вычислялась для 1000 случайных разбиений всей совокупности наблюдений на обучающую (60%) и проверочную (40%) выборки.

В данном примере оценки GAM и ядерные оценки показали близкие результаты, которые оказались лучше результатов моделей линейной регрессии и случайного леса.

Лучшей по точности оценкой оказалась новая универсальная локально–постоянная оценка (66), хотя ее отличие от оценки Надарая-Ватсона было статистически незначимо (220.2 (215.4, 225.9) против 220.4 (215.4, 225.8), p=0.91). Различие между универсальной локально–линейной оценкой (67) и классической локально–линейной оценкой также было статистически незначимым (220.4 (215.4, 225.9) против 220.6 (215.6, 226.1), p=0.52).

Таким образом, в данном примере статистической обработки реальных данных точность новых ядерных оценок сравнима с точностью двух наиболее популярных представителей классических ядерных оценок.

1.2.4 Доказательства утверждений разделов 1.2.1 и 1.2.2

Для доказательства теоремы 1.3 нам потребуется ряд вспомогательных утверждений. В приводимых далее в этом подразделе леммах считаем, что выполнены условия теоремы 1.3. Обозначим

$$\beta_{n,i}(t) = \frac{w_{n2}(t) - (t - z_{n:i})w_{n1}(t)}{w_{n0}(t)w_{n2}(t) - w_{n1}^2(t)}. \tag{69}$$

Сохраним обозначение ε_{ni} для погрешностей, ассоциированных с порядковой статистикой $z_{n:i}$. Принимая во внимание соотношение $X_{ni} = f(z_{n:i}) + \varepsilon_{ni},\ i = 1, \ldots, n$, и тождество

$$\sum_{i=1}^{n} \beta_{n,i}(t) K_h(t - z_{n:i}) \Delta z_{ni} \equiv 1, \tag{70}$$

получаем представление

$$\widetilde{f}^*_{n,h}(t) = f(t) + f(t) I(\delta_n > c_* h) + \widetilde{r}_{n,h}(f,t) + \widetilde{\nu}_{n,h}(t), \tag{71}$$

где

$$\widetilde{r}_{n,h}(f,t) = I(\delta_n \le c_* h) \sum_{i=1}^{n} \beta_{n,i}(t)(f(z_{n:i}) - f(t)) K_h(t - z_{n:i}) \Delta z_{ni},$$
$$\widetilde{\nu}_{n,h}(t) = I(\delta_n \le c_* h) \sum_{i=1}^{n} \beta_{n,i}(t) K_h(t - z_{n:i}) \Delta z_{ni} \varepsilon_{ni}.$$

Подчеркнем, что ввиду свойств плотности $K_h(\cdot)$ область суммирования в двух последних суммах, а также во всех суммах, определяющих величины $w_{nj}(t)$ из (58), совпадает с множеством $A_{n,h}(t) = \{i : |t - z_{n:i}| \le h, 1 \le i \le n\}$, что является принципиальным моментом для дальнейшего анализа.

Лемма 1.7. *При $h < 1/2$ имеют место равенства*

$$\inf_{t\in[0,1]}(w_0(t)w_2(t) - w_1^2(t)) = \frac{1}{4}(\kappa_2 - \kappa_1^2)h^2, \quad \inf_{t\in[0,1]} w_0(t) = 1/2, \tag{72}$$

$$\sup_{t\in[0,1]} |w_j(t)| = \left(\frac{1}{2}\right)^{j-2[j/2]} \kappa_j h^j, \quad j = 0, 1, 2, 3, \tag{73}$$

где κ_j и $w_j(t)$ введены в (56) и (62). Кроме того, на подмножестве элементарных исходов, определяемых соотношением $\delta_n \le c_* h$, справедливы неравенства

$$\sup_{t\in[0,1]} |w_{nj}(t)| \le 3Lh^j, \quad \sup_{t\in[0,1]} |w_{nj}(t) - w_j(t)| \le 12L\delta_n h^{j-1}, \quad j = 0, 1, 2, 3, \tag{74}$$

$$\inf_{t\in[0,1]}(w_{n0}(t)w_{n2}(t) - w_{n1}^2(t)) \ge \frac{1}{8}(\kappa_2 - \kappa_1^2)h^2, \quad \inf_{t\in[0,1]} w_{n0}(t) \ge 1/4, \tag{75}$$

$$\forall t_1, t_2 \in [0,1] \quad |w_{nj}(t_2) - w_{nj}(t_1)| \le 18Lh^{j-1}|t_2 - t_1|, \quad j = 0, 1, 2. \tag{76}$$

Доказательство. Докажем (72) и (73). Прежде всего отметим, что в силу неравенства Коши–Буняковского–Шварца $w_0(t)w_2(t) - w_1^2(t) \ge 0$ при всех $t \in [0,1]$, и указанная разность непрерывна по t. Сначала рассмотрим простейший случай, когда $h \le t \le 1-h$. Для таких t после замены переменной интегрирования в определении (62) величин $w_j(t)$ имеем

$$w_j(t) = \int\limits_{t-h}^{t+h} (t-z)^j K_h(t-z)dz = h^j \int\limits_{-1}^{1} v^j K(v)dv, \tag{77}$$

т.е. $w_0(t) \equiv 1$, $w_1(t) \equiv 0$ и $w_2(t) \equiv h^2\kappa_2$. Иначе говоря, на отрезке $[h, 1-h]$ имеет место тождество

$$w_0(t)w_2(t) - w_1^2(t) \equiv h^2\kappa_2. \tag{78}$$

Теперь рассмотрим случай $t = \alpha h$ для любых $\alpha \in [0,1]$. Имеем

$$w_j(\alpha h) = \int\limits_0^{(1+\alpha)h} (\alpha h - z)^j K_h(\alpha h - z)dz = h^j \kappa_j(\alpha). \tag{79}$$

Далее, с помощью (79) получаем

$$\frac{d}{d\alpha}h^{-2}(w_0(\alpha h)w_2(\alpha h) - w_1^2(\alpha h)) = \frac{d}{d\alpha}(\kappa_0(\alpha)\kappa_2(\alpha) - \kappa_1^2(\alpha)) =$$
$$= K(\alpha)\left(\alpha^2 \int\limits_{-1}^{\alpha} K(v)dv + \int\limits_{-1}^{\alpha} v^2 K(v)dv - 2\alpha \int\limits_{-1}^{\alpha} vK(v)dv\right) \ge 0,$$

поскольку $\int_{-1}^{\alpha} vK(v)dv \le 0$ в силу четности функции $K(v)$. Аналогично рассматривается симметричный случай $t = 1 - \alpha h$ для всех $\alpha \in [0,1]$. Отсюда и из (78) следует первое соотношение из (72):

$$\inf_{t\in[0,1]}\{w_0(t)w_2(t) - w_1^2(t)\} = w_0(0)w_2(0) - w_1^2(0) = \frac{1}{4}h^2(\kappa_2 - \kappa_1^2).$$

Второе соотношение из (72) очевидным образом следует из (79). Кроме того, вышеприведенные рассуждения и представления (77), (79) влекут (73).

Первая оценка в (74) очевидна в силу вышеприведенного замечания об области суммирования в определении функций $w_{nj}(t)$ и соотношений

$$\sup_{s\in[0,1]} K(s) \le L, \qquad \sum_{i\in A_{n,h}(t)} \Delta z_{ni} \le 2h + \delta_n \le 3h. \tag{80}$$

Вторая оценка в (74) сразу следует из известной оценки погрешности аппроксимации интегральными суммами Римана соответствующих интегралов от гладких функций на отрезке:

$$\left|\sum_{i\in A_{n,h}(t)} g_{t,j}(z_{n:i})\Delta z_{ni} - \int\limits_{z\in[0,1]:|t-z|\le h} g_{t,j}(z)dz\right| \le (2h+\delta_n)\delta_n L_{g_{t,j}}, \tag{81}$$

где функции $g_{t,j}(z) = (t-z)^j K_h(t-z)$, $j = 0,1,2,3$, заданы при значениях аргумента $z \in [0 \vee t-h, 1 \wedge t+h]$, а $L_{g_{t,j}}$ — константа Липшица функции $g_{t,j}(z)$. Нетрудно показать, что $\sup_{t\in[0,1]} L_{g_{t,j}} \le 4Lh^{j-2}$ для всех $h \in (0,1/2)$ и $j = 0,1,2,3$, так что на множестве элементарных исходов $\{\delta_n \le c_* h\}$ (напомним, что $c_* < 1$) правую часть в (81) можно заменить на $12L\delta_n h^{j-1}$.

Далее, с учетом (73) и (74) получаем, что

$$\begin{aligned}&|w_{n0}(t)w_{n2}(t) - w_0(t)w_2(t)| \le \\ &\qquad\le w_{n0}(t)|w_{n2}(t) - w_2(t)| + w_2(t)|w_{n0}(t) - w_0(t)| \le 9L\delta_n(3L+\kappa_2)h,\end{aligned}$$

$$|w_{n1}^2(t) - w_1^2(t)| \le |w_{n1}(t) - w_1(t)|(|w_{n1}(t)| + |w_1(t)|) \le 9L\delta_n(3L + \kappa_1/2)h.$$

Тем самым,

$$|w_{n0}(t)w_{n2}(t) - w_{n1}^2(t) - w_0(t)w_2(t) + w_1^2(t)| \le 9L\delta_n(6L + \kappa_2 + \kappa_1/2)h. \tag{82}$$

Неравенства в (75) следуют теперь из (72), (82) и определения константы c_*.

Для доказательства (76) заметим, что

$$|(t_2 - z_{n:i})^j - (t_1 - z_{n:i})^j| \le 2h^{j-1}|t_2 - t_1| \quad \text{при} \quad j = 0,1,2,$$

$|t_k - z_{n:i}| \le h$ при $k = 1, 2$. Нам остается учесть, что

$$w_{nj}(t_2) - w_{nj}(t_1) = \sum_{i=1}^{n} \left\{(t_2 - z_{n:i})^j K_h(t_2 - z_{n:i}) - (t_1 - z_{n:i})^j K_h(t_1 - z_{n:i})\right\} \Delta z_{ni} =$$
$$= \sum_{i \in A_{n,h}(t_1) \cup A_{n,h}(t_2)} \left\{(t_2 - z_{n:i})^j K_h(t_2 - z_{n:i}) - (t_1 - z_{n:i})^j K_h(t_1 - z_{n:i})\right\} \Delta z_{ni}$$

и воспользуемся неравенствами

$$|K_h(t_2 - z_{n:i}) - K_h(t_1 - z_{n:i})| \le L h^{-2} |t_2 - t_1|$$

и

$$\sum_{i \in A_{n,h}(t_1) \cup A_{n,h}(t_2)} \Delta z_{ni} \le 4h + 2\delta_n \le 6h. \tag{83}$$

Таким образом, лемма 1.7 доказана. □

Лемма 1.8. *Для любого положительного $h < 1/2$ имеет место оценка*

$$\sup_{t \in [0,1]} |\widetilde{r}_{n,h}(f, t)| \le C_1^* \omega_f(h) \quad \text{при} \quad C_1^* = C_1 \frac{L^2}{\kappa_2 - \kappa_1^2}.$$

Д о к а з а т е л ь с т в о. Без ограничения общности требуемую оценку можно выводить на множестве элементарных исходов, определяемых условием $\delta_n \le c_* h$. Тогда утверждение леммы следует из соотношения

$$|\widetilde{r}_{n,h}(f, t)| \le \frac{\omega_f(h) w_{n2}(t)}{w_{n0}(t) w_{n2}(t) - w_{n1}^2(t)} \sum_{i \in A_{n,h}(t)} K_h(t - z_{n:i}) \Delta z_{ni} +$$
$$+ \frac{\omega_f(h) |w_{n1}(t)|}{w_{n0}(t) w_{n2}(t) - w_{n1}^2(t)} \sum_{i \in A_{n,h}(t)} |t - z_{n:i}| K_h(t - z_{n:i}) \Delta z_{ni}, \tag{84}$$

оценок из (80) и утверждений леммы 1.7. □

Лемма 1.9. *Для любых $y > 0$ и $h < 1/2$ на подмножестве элементарных исходов, определяемых соотношением $\delta_n \le c_* h$, имеет место следующая оценка*:

$$\mathbb{P}_{\mathcal{F}_n}\left(\sup_{t \in [0,1]} |\widetilde{\nu}_{n,h}(t)| > y\right) \le C_2^* \sigma^2 \frac{\delta_n}{h^2 y^2} \quad \text{при} \quad C_2^* = C_2 \frac{L^4}{(\kappa_2 - \kappa_1^2)^2}.$$

Д о к а з а т е л ь с т в о. Положим

$$\mu_{n,h}(t) = \sum_{i \in N_{n,h}(t)} h^{-2} \alpha_{n,i}(t) K_h\left(t - z_{n:i}\right) \Delta z_{ni} \varepsilon_{ni}, \tag{85}$$

где $\alpha_{n,i}(t) = w_{n2}(t) - (t - z_{n:i})w_{n1}(t)$. Заметим, что из леммы 1.7 и условий леммы 1.9 следуют два факта. Во-первых, $h^{-2}|\alpha_{n,i}(t)| \le 6L$, если только $i \in A_{n,h}(t)$. Во-вторых,

$$|\widetilde{\nu}_{n,h}(t)| \le 8(\kappa_2 - \kappa_1^2)^{-1}|\mu_{n,h}(t)|. \tag{86}$$

Хвост распределения величины $\sup\limits_{t\in[0,1]} |\mu_{n,h}(t)|$ оценим с помощью метода диадических цепочек [43]. Введем множество двоично-рациональных точек

$$\mathcal{R} = \{j/2^k;\ j = 1, \dots, 2^k - 1; k \ge 1\}.$$

Имеем

$$\sup_{t\in[0,1]} |\mu_{n,h}(t)| = \sup_{t\in\mathcal{R}} |\mu_{n,h}(t)| \le$$
$$\le \max_{j=1,\dots,2^m-1} |\mu_{n,h}(j2^{-m})| + \sum_{k=m+1}^{\infty} \max_{j=1,\dots,2^k-2} \big|\mu_{n,h}((j+1)2^{-k}) - \mu_{n,h}(j2^{-k})\big|,$$

где натуральное m зададим равенством $m = \lceil |\log_2 h| \rceil$. Имеем

$$\mathbb{P}_{\mathcal{F}_n}\Big(\sup_{t\in[0,1]} |\mu_{n,h}(t)| > y\Big) \le \mathbb{P}_{\mathcal{F}_n}\Big(\max_{j=1,\dots,2^m-1} |\mu_{n,h}(j2^{-m})| > a_m y\Big) +$$
$$+ \sum_{k=m+1}^{\infty} \mathbb{P}_{\mathcal{F}_n}\Big(\max_{j=1,\dots,2^k-2} \big|\mu_{n,h}((j+1)2^{-k}) - \mu_{n,h}(j2^{-k})\big| > a_k y\Big) \le$$
$$\le \sum_{j=1}^{2^m-1} \mathbb{P}_{\mathcal{F}_n}(|\mu_{n,h}(j2^{-m})| > a_m y) +$$
$$+ \sum_{k=m+1}^{\infty} \sum_{j=1}^{2^k-2} \mathbb{P}_{\mathcal{F}_n}\Big(\big|\mu_{n,h}((j+1)2^{-k}) - \mu_{n,h}(j2^{-k})\big| > a_k y\Big), \tag{87}$$

где $a_m, a_{m+1}, \dots$ — последовательность положительных чисел, удовлетворяющих условию $a_m + a_{m+1} + \dots = 1$.

Оценим теперь каждое из слагаемых в правой части (87). С помощью неравенства Маркова со вторым моментом и оценок (80), получаем

$$\mathbb{P}_{\mathcal{F}_n}(|\mu_{n,h}(j2^{-m})| > a_m y) \le \frac{(6L)^2}{(a_m y)^2} \sum_{i\in A_{n,h}(j2^{-m})} K_h^2(j2^{-m} - z_{n:i})(\Delta z_{ni})^2\sigma^2 \le$$
$$\le (6L)^2\sigma^2(a_m y)^{-2}\delta_n(2h + \delta_n)h^{-2} \le C_3 L^2\sigma^2(a_m y)^{-2}\delta_n h^{-1}. \tag{88}$$

Далее,

$$\mathbb{P}_{\mathcal{F}_n}\Big(\big|\mu_{n,h}((j+1)2^{-k})-\mu_{n,h}(j2^{-k})\big|>a_k y\Big)\le (a_k y)^{-2}h^{-4}\times$$
$$\times\sum_{i=1}^{n}\mathbb{E}_{\mathcal{F}}\Big(\big(\alpha_{n,i}((j+1)2^{-k})K_h((j+1)2^{-k}-z_{n:i})-\alpha_{n,i}(j2^{-k})K_h(j2^{-k}-z_{n:i})\big)\Delta z_{ni}\varepsilon_{ni}\Big)^2\le$$
$$\le\sigma^2(a_k y)^{-2}h^{-4}\times$$
$$\times\sum_{i=1}^{n}\Big(\alpha_{n,i}((j+1)2^{-k})K_h((j+1)2^{-k}-z_{n:i})-\alpha_{n,i}(j2^{-k})K_h(j2^{-k}-z_{n:i})\Big)^2(\Delta z_{ni})^2\le$$
$$\le C_4\sigma^2L^4(a_k y)^{-2}2^{-2k}\delta_n(4h+2\delta_n)h^{-4}\le C_5\sigma^2L^4(a_k y)^{-2}2^{-2k}\delta_n h^{-3}. \quad (89)$$

Здесь мы учли, что область суммирования в (89) совпадает с множеством

$$\big\{i:\ i\in A_{n,h}((j+1)2^{-k})\cup A_{n,h}(j2^{-k})\big\},$$

а потому в силу соотношения $|(j+1)/2^k-j/2^k|=2^{-k}\le h$ при $k>m$ имеет место оценка (83) при $t_1=j2^{-k}$ и $t_2=(j+1)2^{-k}$. Кроме того, мы использовали оценки

$$\sup_t K_h(t)\le Lh^{-1},\quad |K_h(u)-K_h(v)|\le Lh^{-2}|u-v|$$

и учли следующие неравенства в вышеуказанной области изменения параметров, вытекающие из леммы 1.7:

$$|\alpha_{n,i}((j+1)2^{-k})-\alpha_{n,i}(j2^{-k})|\le C_6Lh2^{-k},\qquad |\alpha_{n,i}(j2^{-k})|\le C_7Lh^2,$$
$$|\alpha_{n,i}((j+1)2^{-k})K_h((j+1)2^{-k}-z_{n:i})-\alpha_{n,i}(j2^{-k})K_h(j2^{-k}-z_{n:i})|\le C_8L2^{-k}.$$

Теперь из (87)–(89) получаем

$$\mathbb{P}_{\mathcal{F}_n}\left(\sup_{t\in[0,1]}|\mu_{n,h}(t)|>y\right)\le C_9y^{-2}\sigma^2L^4\delta_nh^{-1}\left(2^m a_m^{-2}+h^{-2}\sum_{k=m+1}^{\infty}2^{-k+1}a_k^{-2}\right).$$

Оптимальная последовательность a_k, минимизирующая правую часть этого неравенства, есть $a_m=c2^{m/3}$ и $a_k=ch^{-2/3}2^{(-k+1)/3}$ при $k=m+1,m+2,...$, где c определяется соотношением $a_m+a_{m+1}+...=1$. Для указанной последовательности заключаем, что

$$\mathbb{P}_{\mathcal{F}_n}\left(\sup_{t\in[0,1]}|\mu_{n,h}(t)|>y\right)\le$$
$$\le C_{10}y^{-2}\sigma^2L^4\delta_nh^{-1}\left(2^{m/3}+h^{-2/3}2^{-m/3}\big(2+2^{1/3}+2^{2/3}\big)\right)^3\le C_{11}y^{-2}\sigma^2L^4\delta_nh^{-2}.$$

Утверждение леммы следует теперь из (86). □

Д о к а з а т е л ь с т в о теоремы 1.3. Это утверждение следует из лемм 1.8 и 1.9, если только мы положим

$$\widetilde{\zeta}_n(h) = \sup_{t\in[0,1]} |\widetilde{\nu}_{n,h}(t)| + \sup_{t\in[0,1]} |f(t)| I(\delta_n > c_* h)$$

и учтем соотношение

$$\mathbb{P}\big(\widetilde{\zeta}_n(h) > y,\, \delta_n \le c_* h\big) = \mathbb{E} I\big(\delta_n \le c_* h\big) \mathbb{P}_{\mathcal{F}_n}\big(\widetilde{\zeta}_n(h) > y\big).$$

Теорема доказана. □

Д о к а з а т е л ь с т в о предложения 1.4. Для оценки $f^*_{n,h}(t)$, определенной в (60), нам потребуется следующее представление:

$$f^*_{n,h}(t) = f(t) + r^*_{n,h}(f,t) + \nu^*_{n,h}(t), \tag{90}$$

где

$$r^*_{n,h}(f,t) = w_{n0}^{-1}(t) \sum_{i=1}^{n} (f(z_{n:i}) - f(t)) K_h(t - z_{n:i}) \Delta z_{ni},$$
$$\nu^*_{n,h}(t) = w_{n0}^{-1}(t) \sum_{i=1}^{n} K_h(t - z_{n:i}) \Delta z_{ni} \varepsilon_{ni}.$$

В силу представлений (71) и (90) получаем, что

$$\begin{aligned} \mathrm{Bias}\widetilde{f}^*_{n,h}(t) &= \mathbb{E}\widetilde{r}_{n,h}(f,t) + f(t)\mathbb{P}(\delta_n > c_* h) = \\ &= \sum_{i=1}^{n} \mathbb{E}\{I(\delta_n \le c_* h)\beta_{n,i}(t)(f(z_{n:i}) - f(t))K_h(t - z_{n:i})\Delta z_{ni}\} + f(t)\mathbb{P}(\delta_n > c_* h), \end{aligned} \tag{91}$$

$$\mathrm{Bias} f^*_{n,h}(t) = \mathbb{E} r^*_{n,h}(f,t) = \sum_{i=1}^{n} \mathbb{E}\{I(\delta_n \le c_* h) w_{n0}^{-1}(t)(f(z_{n:i}) - f(t))K_h(t - z_{n:i})\Delta z_{ni}\} + \tau_n, \tag{92}$$

где $|\tau_n| \le \omega_f(h)\mathbb{P}(\delta_n > c_* h)$. Далее, из леммы 1.7 следует, что для любой точки $t \in [h, 1-h]$ при условии $\delta_n \le c_* h$ справедлива оценка

$$\sup_{i\in A_{n,h}(t)} |\beta_{n,i}(t) - w_{n0}^{-1}(t)| \le C_5^* \delta_n h^{-1}. \tag{93}$$

При выводе соотношения (93) мы также учли, что $w_0(t) = 1$ и $w_1(t) = 0$ при всех $t \in [h, 1-h]$ (см. доказательство леммы 1.7). Теперь с учетом (80), (91), (92), (93) и леммы 1.7 нетрудно вывести первое утверждение леммы, поскольку

$$|\mathrm{Bias}\widetilde{f}^*_{n,h}(t)-\mathrm{Bias}f^*_{n,h}(t)|\le C_5^*h^{-1}\omega_f(h)\mathbb{E}\left\{\delta_n I(\delta_n\le c_*h)\sum_{i=1}^n K_h(t-z_{n:i})\Delta z_{ni}\right\}+$$
$$+(|f(t)|+\omega_f(h))\mathbb{P}(\delta_n>c_*h)\le C_6^*\omega_f(h)h^{-1}\mathbb{E}\delta_n+(|f(t)|+\omega_f(h))\mathbb{P}(\delta_n>c_*h).\quad(94)$$

Для доказательства второго утверждения прежде всего отметим, что

$$\mathbb{D}\widetilde{f}^*_{n,h}(t)=\mathbb{D}\widetilde{\nu}_{n,h}(t)+\mathbb{D}\left(\widetilde{r}_{n,h}(f,t)+f(t)I(\delta_n>c_*h)\right)=$$
$$=\mathbb{D}\widetilde{\nu}_{n,h}(t)+\mathbb{D}\widetilde{r}_{n,h}(f,t)+f^2(t)\mathbb{P}(\delta_n>c_*h)\mathbb{P}(\delta_n\le c_*h),$$

$$\mathbb{D}f^*_{n,h}(t)=\mathbb{D}\nu^*_{n,h}(t)+\mathbb{D}r^*_{n,h}(f,t).$$

Таким образом, нам нужно сравнить две дисперсии в правой части первого равенства с соответствующими дисперсиями второго. С помощью (80) и (93) получаем

$$|\mathbb{D}\widetilde{\nu}_{n,h}(t)-\mathbb{D}\nu^*_{n,h}(t)|\le\sigma^2\left|\mathbb{E}\sum_{i=1}^n I(\delta_n\le c_*h)(\beta_{n,i}^2(t)-w_{n0}^{-2}(t))K_h^2(t-z_{n:i})(\Delta z_{ni})^2\right|+$$
$$+\sigma^2\mathbb{P}(\delta_n>c_*h)\le C_7^*\sigma^2h^{-1}\mathbb{E}\left\{\delta_n I(\delta_n\le c_*h)\sum_{i=1}^n hK_h^2(t-z_{n:i})\Delta z_{ni}\right\}+$$
$$+\sigma^2\mathbb{P}(\delta_n>c_*h)\le C_8^*\sigma^2h^{-1}\mathbb{E}\delta_n.$$

При выводе последней оценки мы учли, что $\sum_{i=1}^n w_{n0}^{-2}(t))K_h^2(t-z_{n:i})(\Delta z_{ni})^2\le 1$.

Чтобы получить оценку для разности $|\mathbb{D}\widetilde{r}_{n,h}(t)-\mathbb{D}r^*_{n,h}(t)|$ отметим, что верхняя оценка $C_9^*\overline{f}^2h^{-1}\mathbb{E}\delta_n$ для модуля разности квадратов смещений случайных величин $\widetilde{r}_{n,h}(f,t)$ и $r^*_{n,h}(f,t)$ по существу содержится в (84) и (94). Оценивание разности вторых моментов указанных случайных величин проводится аналогично с помощью (80), (93) и (94):

$$|\mathbb{E}\widetilde{r}^2_{n,h}(f,t)-\mathbb{E}r^{*2}_{n,h}(f,t)|\le\mathbb{E}|\widetilde{r}_{n,h}(f,t)-r^*_{n,h}(f,t)||\widetilde{r}_{n,h}(f,t)+r^*_{n,h}(f,t)|\le C_{10}^*\overline{f}^2h^{-1}\mathbb{E}\delta_n,$$

что завершает доказательство предложения. □

Д о к а з а т е л ь с т в о предложения 1.5. Из определения величины $\beta_{n,i}(t)$ в (69) следует, что при любом $t\in[0,1]$

$$\sum_{i=1}^n\beta_{n,i}(t)(z_{n:i}-t)K_h(t-z_{n:i})\Delta z_{ni}=0,$$
$$\sum_{i=1}^n\beta_{n,i}(t)(z_{n:i}-t)^2K_h(t-z_{n:i})\Delta z_{ni}=D_n^{-1}(t)(w_{n2}^2(t)-w_{n3}(t)w_{n1}(t))\equiv B_n(t),$$

где $D_n(t) = w_{n0}(t)w_{n2}(t) - w_{n1}^2(t)$. Раскладывая функцию $f(\cdot)$ по формуле Тейлора в окрестности точки t (до второй производной), из вышеприведенных тождеств получаем с помощью (69), (91) и леммы 1.7, что для любой точки t выполнено

$$\mathrm{Bias}\widetilde{f}^*_{n,h}(t) = \mathbb{E}I(\delta_n \le c_* h)\sum_{i=1}^n \{\beta_{n,i}(t)(f(z_{n:i}) - f(t))K_h(t - z_{n:i})\Delta z_{ni}\} +$$
$$+ f(t)\mathbb{P}(\delta_n > c_* h) = \frac{f''(t)}{2}\mathbb{E}I(\delta_n \le c_* h)B_n(t) + f(t)\mathbb{P}(\delta_n > c_* h) + o(h^2) =$$
$$= \frac{f''(t)}{2}B_0(t) + O(\mathbb{E}\delta_n/h) + o(h^2), \quad (95)$$

при этом O- и o-символы в правой части (95) равномерны по t. Отметим, что при любом t имеет место соотношение $B_0(t) = O(h^2)$.

Далее, так как для $j = 1, 2$ и при всех натуральных n выполнено

$$|w_j(t)|w_0^{-1}(t) \le h^j, \qquad |w_{nj}(t)|w_{n0}^{-1}(t) \le h^j,$$

то справедливо асимптотическое представление

$$\mathrm{Bias}f^*_{n,h}(t) = \sum_{i=1}^n \mathbb{E}w_{n0}^{-1}(t)(f(z_{n:i}) - f(t))K_h(t - z_{n:i})\Delta z_{ni} =$$
$$= -f'(t)\mathbb{E}\frac{w_{n1}(t)}{w_{n0}(t)}I(\delta_n \le c_* h) + \frac{f''(t)}{2}\mathbb{E}\frac{w_{n2}(t)}{w_{n0}(t)}I(\delta_n \le c_* h) + O(h\mathbb{P}(\delta_n > c_* h)) + o(h^2)$$
$$= -f'(t)\frac{w_1(t)}{w_0(t)} + \frac{f''(t)}{2}\frac{w_2(t)}{w_0(t)} + O(\mathbb{E}\delta_n) + o(h^2). \qquad \square$$

Д о к а з а т е л ь с т в о следствия 1.3. Без ограничения общности можно считать, что $t \in [h, 1-h]$. Тогда, как уже было отмечено при выводе леммы 1.7, для указанных t будет выполнено $w_0(t) = 1$, $w_1(t) = 0$ и $w_2(t) = \kappa_2 h^2$, т.е. $B_0(t) = \kappa_2 h^2$. $\square$

Д о к а з а т е л ь с т в о следствия 1.4. Это утверждение непосредственно вытекает из предложения 1.5 и формулы (79). $\square$

1.3 Универсальные оценки для случайных полей

1.3.1 Универсальные локально-постоянные оценки

В этом разделе считаем выполненным предположение (M_1) о регрессионной модели и условие (E_2), касающееся погрешностей. Нам дополнительно потребуются следующие предположения на ядро и регрессоры.

($\mathbf{K}_2$) *Выполнено условие* (K), *ядерная функция* $K(\mathbf{s})$ *удовлетворяет условию Липшица с константой* $L \geq 1$, *т.е.* $|K(\mathbf{x}) - K(\mathbf{y})| \leq L\,(|x_1 - y_1| + \cdots + |x_k - y_k|)$ *при всех* $\mathbf{x} = (x_1, \ldots, x_k)$ *и* $\mathbf{y} = (y_1, \ldots, y_k)$. *Кроме того, существуют константы* $\rho > 0$ *и* $h_0 \in (0,1]$ *такие, что* $J_h(\mathbf{t}) \geq \rho$ *при всех* $\mathbf{t} \in \mathcal{P}$ *и* $0 < h \leq h_0$, *где* $J_h(\mathbf{t}) = \int\limits_{\mathcal{P}} K_h(\mathbf{t} - \mathbf{x})\,\Lambda_k(\mathbf{dx})$ *и* $\mathbf{t} \in \mathcal{P}$.

В ядерных оценках мы будем использовать конструкцию кратных интегральных сумм Римана, поэтому нам потребуется следующее условие плотного заполнения регрессорами $\{\mathbf{z}_i;\, i = 1, \ldots, n\}$ области $\mathcal{P}$ определения регрессионной функции f.

($\mathbf{D}_2$) *Для каждого* n *существует такое разбиение множества* $\mathcal{P}$ *на* n *измеримых по Жордану подмножеств* $\{\mathcal{P}_i,\ i = 1, \ldots, n\}$, *что каждый элемент этого разбиения содержит ровно по одной точке из набора* $\{\mathbf{z}_i;\, i = 1, \ldots, n\}$ (*нумерация элементов разбиения такова, что* $\mathbf{z}_i \in \mathcal{P}_i$), *при этом* $\delta_n = \max_{i \leq n} d(\mathcal{P}_i) \xrightarrow{p} 0$, *где* $d(A) = \sup_{\mathbf{x},\mathbf{y} \in A} \|\mathbf{x} - \mathbf{y}\|$ – *диаметр множества,* $\|\cdot\|$ — *супремальная норма в* $\mathbb{R}^k$.

З а м е ч а н и е 1.16. Традиционно семейство непустых множеств $\mathcal{P}_1, \ldots, \mathcal{P}_n$ образует разбиение множества $\mathcal{P}$, если элементы семейства $\{\mathcal{P}_i\}$ попарно не пересекаются и $\bigcup\limits_{i=1}^{n} \mathcal{P}_i = \mathcal{P}$. Условимся, что в ($D_2$) допускается, чтобы элементы набора $\{\mathcal{P}_i\}$ пересекались по множествам нулевой меры Лебега (например, по границам). Подобная оговорка позволяет нам не исключать ситуацию кратных точек в наборе регрессоров. В случае попарно различных регрессоров подобная оговорка не требуется. Отметим еще, что без вышеуказанного соглашения условие (D_2) можно сформулировать следующим образом: *для каждого* n *существует разбиение множества* $\mathcal{P}$ *на* n *измеримых по Жордану подмножеств* $\{\mathcal{P}_i;\ i = 1, \ldots, n\}$ *таких, что* $\delta_n = \max\limits_{i \leq n} d(\mathcal{P}_i \cup \mathbf{z}_i) \xrightarrow{p} 0$.

З а м е ч а н и е 1.17. Условие (D_2) означает, что для любого n регрессоры $\{\mathbf{z}_i;\, i = 1, \ldots, n\}$ образуют ε-сеть компактного множества $\mathcal{P}$ при $\varepsilon = \delta_n$ с условием $\delta_n \xrightarrow{p} 0$. Отметим, что неслучайные регрессоры с условиями регулярного плотного заполнения области задания регрессионной функции в широких условиях удовлетворяют условию (D_2). Напомним, что в регулярном случае неслучайные точки $\mathbf{z}_i$ задаются, например, формулой $\mathbf{z}_i = g(\mathbf{v}_{in})$ для некоторой k-мерной функции g ограниченной вариации и k-мерных точек решетки вида $\mathbf{v}_{in} = (j_1/n^{1/k}, \ldots, j_k/n^{1/k}) \in [0,1]^k$ (если $\mathcal{P} = [0,1]^k$ и g есть тождественная функция, то получаем эквидистантный план). Другой вариант задания регулярных неслучайных регрессоров — условие вида $\max\limits_{i \leq n} d(\mathcal{P}_i) = O(1/n)$. В случае независимых одинаково распределенных регрессоров с плотностью распределения случайной величины $\mathbf{z}_1$, отделенной от нуля на $[0,1]^k$, с вероятностью 1 выполнено $\delta_n = O\left(\frac{\log n}{n^{1/k}}\right)$. Если $\{\mathbf{z}_i;\, i \geq 1\}$ есть стационарная последовательность с условием α-перемешивания и множество $\mathcal{P} = [0,1]^k$ является носителем распределения $\mathbf{z}_1$, то условие (D_2) также выполнено (см. замечание 1.4). Отметим, что зависимость случайных величин $\{\mathbf{z}_i\}$ в условии (D_2) может быть более сильной (см. примеры 1.1, 1.2 и приводимые далее примеры 1.16, 1.18). □

З а м е ч а н и е 1.18. Без ограничения общности считаем, что $\mathcal{P} \subseteq [0,1]^k$. Заметим еще, что при сделанных предположениях $\sup_{\mathbf{s}} K(\mathbf{s}) \le L$. Если $\mathcal{P} = [0,1]^k$ и $K(\mathbf{t}) = \prod_{j=1}^k K_o(t_j)$ при $\mathbf{t} = (t_1, \ldots, t_k)^\top$, где $K_o(\cdot)$ есть симметричная плотность распределения с носителем $[-1,1]$, то $\rho \ge 2^{-k}$.

Введем в рассмотрение следующий класс ядерных оценок для функции f:

$$f^*_{n,h}(\mathbf{t}) = \frac{\sum_{i=1}^n X_i K_h(\mathbf{t} - \mathbf{z}_i)\Lambda_k(\mathcal{P}_i)}{\sum_{i=1}^n K_h(\mathbf{t} - \mathbf{z}_i)\Lambda_k(\mathcal{P}_i)}. \tag{96}$$

Отметим, что

$$f^*_{n,h}(\mathbf{t}) = \arg\min_{a \in \mathbb{R}} \sum_{i=1}^n (X_i - a)^2 K_h(\mathbf{t} - \mathbf{z}_i)\Lambda_k(\mathcal{P}_i),$$

т.е. оценки (96) принадлежат классу оценок взвешенного метода наименьших квадратов и являются локально–постоянными оценками. Сравнивая $f^*_{n,h}(\mathbf{t})$ с классическими локально–линейными оценками $\widehat{f}_{NW}(\mathbf{t})$ (т.е. оценками Надарая–Ватсона, см. формулу (163) далее), отметим, что для новых оценок в методе наименьших квадратов используются иные веса, задаваемые мерой Лебега элементов конечного случайного разбиения выборочного пространства регрессоров, введенного в условии (D_2). Подобные веса позволяют нам получить конструкцию интегральных сумм Римана в структуре предлагаемых ядерных оценок. Данное обстоятельство существенно для наших целей, поскольку при асимптотическом анализе оценок открывается возможность вместо тех или иных предельных теорем использовать сближение интегральных сумм и соответствующих кратных интегралов Римана.

Основной результат подраздела содержится в следующем утверждении.

Теорема 1.4. *Пусть выполнены условия* (M_1), (E_2), (K_2) *и* (D_2). [2] *Тогда для любого фиксированного* $h \in (0, h_0)$ *с вероятностью* 1 *справедливо соотношение*

$$\sup_{\mathbf{t} \in \mathcal{P}} |f^*_{n,h}(\mathbf{t}) - f(\mathbf{t})| \le \omega_f(h) + \zeta_n(h), \tag{97}$$

где случайная величина $\zeta_n(h) > 0$ *такова, что*

$$\mathbb{P}\left(\zeta_n(h) > y\right) \le G(k,p)\,\rho^{-p}\, M_p\, L^p\, y^{-p}\, h^{-k(p/2+1)}\,\mathbb{E}(\delta_n^{kp/2}) + \mathbb{P}(\delta_n > h\min\{1,\, \rho(k2^{k+1}L)^{-1}\}) \tag{98}$$

при $0 < G(k,p) < (p-1)^{p/2} 2^{p(k+(3/2))} \left(1 + \frac{k}{2^{(p-k)/(p+1)} - 1}\right)^{p+1}$.

[2]Утверждение теоремы имеет место из без выполнения второй части условия (D_2). Но поскольку оценки, полученные в теореме, становятся содержательными лишь при выполнении предельного соотношения из (D_2), то нам удобнее включить это условие в формулировку теоремы целиком. Аналогичная ситуация имеет место и в приводимой далее теореме 1.6.

З а м е ч а н и е 1.19. Пусть, например, f – неслучайная функция. Положим в (98) $y = \left(h^{-k(p/2+1)}\,\mathbb{E}(\delta_n^{kp/2})\right)^{1/p}$. Применяя степенное неравенство Маркова с показателем $kp/2$ для второго слагаемого в (98), нетрудно видеть, что в условиях теоремы 1.4

$$\zeta_n(h) = \widetilde{O}_p\left(\left(h^{-k(p/2+1)}\,\mathbb{E}(\delta_n^{kp/2})\right)^{1/p}\right)$$

и существует решение $h \equiv h_n$ уравнения

$$\mathbb{E}(\delta_n^{kp/2}) = h^{k(p/2+1)}\omega_f^p(h). \tag{99}$$

Нетрудно видеть, что это решение будет стремиться к нулю с ростом n. Фактически величина h_n минимизирует по h порядок малости правой части соотношения (97). Отметим, что ввиду (99) имеют место соотношения $\left(h_n\right)^{-k(p/2+1)}\,\mathbb{E}(\delta_n^{kp/2}) \to 0$ и $\delta_n/h_n \xrightarrow{p} 0$.

Принимая во внимание замечание 1.19, из теоремы 1.4 получаем следующие два утверждения.

Следствие 1.5. *Пусть выполнены условия* (M_1), (E_2), (K_2) *и* (D_2), *а* $\mathcal{C}$ — *множество неслучайных равностепенно непрерывных функций из пространства* $C\left([0,1]^k\right)$ *Тогда*

$$\gamma_n(\mathcal{C}) = \sup_{f\in\mathcal{C}}\ \sup_{\mathbf{t}\in[0,1]^k} |f^*_{n,\tilde{h}_n}(\mathbf{t}) - f(\mathbf{t})| \xrightarrow{p} 0,$$

где $\tilde{h}_n$ *есть решение уравнения* (99), *в котором модуль непрерывности* $\omega_f(h)$ *заменен универсальным модулем* $\omega_f^{\mathcal{C}}(h) = \sup\limits_{f\in\mathcal{C}}\omega_f(h)$. *Кроме того, имеет место соотношение* $\gamma_n(\mathcal{C}) = \widetilde{O}_p(\omega_f^{\mathcal{C}}(\tilde{h}_n))$.

Следствие 1.6. *Если выполнены условия* (M_1), (E_2), (K_2) *и* (D_2), *а модуль непрерывности случайной регрессионной функции* $f(\mathbf{t})$, $\mathbf{t} \in [0,1]^k$ *с вероятностью* 1 *удовлетворяет условию* $\omega_f(h) \le \zeta\varphi(h)$ *для некоторой собственной случайной величины* $\zeta > 0$ *и положительной неслучайной функции* $\varphi(h)$ *с условием* $\varphi(h) \to 0$ *при* $h \to 0$, *то справедливо соотношение*

$$\sup_{\mathbf{t}\in[0,1]^k} |f^*_{n,\hat{h}_n}(\mathbf{t}) - f(\mathbf{t})| \xrightarrow{p} 0, \tag{100}$$

где $\hat{h}_n$ — *решение уравнения* (99), *в котором модуль непрерывности* $\omega_f(h)$ *заменен на* $\varphi(h)$.

П р и м е р 1.10. Пусть $\mathcal{P} = [0,1]^k$, $\delta_n \le \nu n^{-1/k}$ при $\mathbb{E}\nu^{kp/2} < \infty$ и $\omega_f(h) \le \zeta h^\alpha$, где $\alpha \in (0,1]$ и ζ — некоторая собственная случайная величина. Тогда

$$h_n = O\left(n^{-\frac{1}{2k(1/p+1/2)+2\alpha}}\right) \qquad \text{и} \qquad \sup_{\mathbf{t}\in[0,1]^k} |f^*_{n,h}(\mathbf{t}) - f(\mathbf{t})| = \widetilde{O}_p\left(n^{-\frac{\alpha}{2k(1/p+1/2)+2\alpha}}\right).$$

В частности, если в одномерном случае $f(\cdot) = W(\cdot)$ есть Винеровский процесс на $[0,1]$, а независимые одинаково распределенные случайные величины $\{\varepsilon_i\}$ имеют нормальное распределение с нулевым средним, то для любого достаточно малого $\nu_0 > 0$ выполнено

$$\sup_{t\in[0,1]} |f^*_{n,h}(t) - f(t)| = \widetilde{O}_p(n^{-1/3+\nu_0}).$$

Здесь $k = 1$, $\alpha = 1/2 - \nu_1$ и $1/p < \nu_2$ для произвольных достаточно малых ν_1 и ν_2. $\square$

Пусть $\mathcal{F}_0$ — тривиальная σ-алгебра. Для любого $i = 1, \ldots, n$ символом $\mathcal{F}_{n,i}$ обозначим σ-алгебру, порожденную наборами $\{\varepsilon_1, \ldots, \varepsilon_i\}$, $\{\mathbf{z}_1, \ldots, \mathbf{z}_n\}$ и случайным полем f. В следующем утверждении приведены достаточные условия асимптотической нормальности оценок $f^*_{n,h}(\mathbf{t})$.

Теорема 1.5. *Пусть регрессоры $\{\mathbf{z}_i\}$ не зависят от n, выполнены условия (M_1), (E_2), (K_2), (D_2) и для некоторого $\mathbf{t} \in \mathcal{P}$ и последовательности $h \equiv h_n$ справедливы следующие соотношения:*

$$\tau_n = \frac{\max\limits_{i\le n} \big(K_h(\mathbf{t} - \mathbf{z}_i)\,\Lambda_k(\mathcal{P}_i)\big)^2}{\sum_{i=1}^n \big(K_h(\mathbf{t} - \mathbf{z}_i)\,\Lambda_k(\mathcal{P}_i)\big)^2} \xrightarrow{p} 0,$$

$$\mathbb{E}(\varepsilon_i^2 \mid \mathcal{F}_{n,i-1}) = \sigma^2 \quad \textit{п.н.} \quad \textit{при всех } i \le n,$$

$$\max_{i\le n} \mathbb{E}\left(\varepsilon_i^2\, I(\varepsilon_i^2 > a/\tau_n) \mid \mathcal{F}_{n,i-1}\right) \xrightarrow{p} 0 \quad \textit{при всех } a > 0.$$

Тогда

$$B^{-1}_{n,h}(\mathbf{t})\left(f^*_{n,h}(\mathbf{t}) - f(\mathbf{t}) - r_{n,h}(\mathbf{t})\right) \xrightarrow{d} \mathcal{N}_1(0, \sigma^2),$$

где

$$B^2_{n,h}(\mathbf{t}) = J^{-2}_{n,h}(\mathbf{t}) \sum_{i=1}^n \big(K_h(\mathbf{t} - \mathbf{z}_i)\Lambda_k(\mathcal{P}_i)\big)^2, \quad J_{n,h}(t) = \sum_{i=1}^n K_h(\mathbf{t} - \mathbf{z}_i)\Lambda_k(\mathcal{P}_i),$$
$$r_{n,h}(\mathbf{t}) = J^{-1}_{n,h}(\mathbf{t}) \sum_{i=1}^n (f(\mathbf{z}_i) - f(\mathbf{t})) K_h(\mathbf{t} - \mathbf{z}_i)\,\Lambda_k(\mathcal{P}_i). \tag{101}$$

1.3.2 Доказательство результатов раздела 1.3.1

Д о к а з а т е л ь с т в о теоремы 1.4. Принимая во внимание соотношение (23), получаем тождество

$$f^*_{n,h}(\mathbf{t}) = R_{n,h}(f, \mathbf{t}) + \nu_{n,h}(\mathbf{t}),$$

где

$$R_{n,h}(f, \mathbf{t}) = J^{-1}_{n,h}(\mathbf{t}) \sum_{i=1}^n f(\mathbf{z}_i) K_h(\mathbf{t} - \mathbf{z}_i)\,\Lambda_k(\mathcal{P}_i),$$

$$\nu_{n,h}(\mathbf{t}) = J^{-1}_{n,h}(\mathbf{t}) \sum_{i=1}^n K_h(\mathbf{t} - \mathbf{z}_i)\,\varepsilon_i \Lambda_k(\mathcal{P}_i),$$

а величина $J_{n,h}(\mathbf{t})$ определена в (101). Важно подчеркнуть, в силу условия (K) областью суммирования в трех указанных суммах является множество $\{i : \|\mathbf{t} - \mathbf{z}_i\| \le h\}$.

Утверждение теоремы 1.4 нетрудно извлечь из указанного тождества и приводимых далее лемм 1.10 и 1.12. Отметим также, что утверждение леммы 1.11 нам потребуется при доказательстве леммы 1.12.

Лемма 1.10. *Имеет место оценка*

$$\sup_{\mathbf{t}\in\mathcal{P}} |R_{n,h}(f,\mathbf{t}) - f(\mathbf{t})| \le \omega_f(h). \tag{102}$$

Д о к а з а т е л ь с т в о утверждения следует из равенства

$$R_{n,h}(f,\mathbf{t}) = f(\mathbf{t}) + J_{n,h}^{-1}(\mathbf{t}) \sum_{i:\, \|\mathbf{t}-\mathbf{z}_i\|\le h} (f(\mathbf{z}_i) - f(\mathbf{t}))K_h\left(\mathbf{t}-\mathbf{z}_i\right)\Lambda_k(\mathcal{P}_i).$$

Лемма доказана. □

Лемма 1.11. *Если $\delta_n \le h \le h_0$, то для любого $\mathbf{t} \in \mathcal{P}$ справедливо соотношение*

$$J_{n,h}(\mathbf{t}) \ge \rho - k2^k L\delta_n h^{-1}.$$

Д о к а з а т е л ь с т в о. Имеем

$$J_{n,h}(\mathbf{t}) = \int g(\mathbf{y})\mathbf{dy},$$

где $g(\mathbf{y}) = \sum\limits_{i=1}^{n} K_h(\mathbf{t}-\mathbf{z}_i)\, I(\mathbf{y}\in\mathcal{P}_i)$ и $I(\cdot)$ есть индикатор события. Далее, поскольку

$$|K_h\left(\mathbf{t}-\mathbf{x}\right) - K_h\left(\mathbf{t}-\mathbf{y}\right)| \le kLh^{-k-1}\delta_n \quad \text{при} \quad \|\mathbf{x}-\mathbf{y}\| \le \delta_n, \tag{103}$$

то $g(\mathbf{y}) \ge K_h(\mathbf{t}-\mathbf{y}) - kLh^{-k-1}\delta_n$ для всех $\mathbf{y}\in\mathcal{P}$. Следовательно, с учетом условия (K_2),

$$J_{n,h}(\mathbf{t}) \ge \int\limits_{\mathbf{y}\in\mathcal{P}:\ \|\mathbf{t}-\mathbf{y}\|\le h} \big(K_h(\mathbf{t}-\mathbf{y}) - kLh^{-k-1}\delta_n\big)\mathbf{dy} \ge \rho - k2^k L\delta_n h^{-1}$$

и лемма доказана. □

Лемма 1.12. *При всех $y > 0$ и $h \in (0, h_0)$ на множестве элементарных исходов, определяемых соотношением $\delta_n/h \le \min\{1,\, \rho(k2^{k+1}L)^{-1}\}$, имеет место оценка*

$$\mathbb{P}_{\mathcal{F}_n}\left(\left(\delta_n^{k/2} h^{-k((1/2)+(1/p))}\right)^{-1} \sup_{\mathbf{t}\in\mathcal{P}} |\nu_{n,h}(\mathbf{t})| > y\right) \le G(k,p)\,\rho^{-p}\, M_p\, L^{p/2} y^{-p}, \tag{104}$$

где

$$G(k,p) < (p-1)^{p/2} 2^{p(k+(3/2))} \left(1 + \frac{k}{2^{(p-k)/(p+1)} - 1}\right)^{p+1}.$$

Д о к а з а т е л ь с т в о. Из условия $\delta_n/h \le \rho(k2^{k+1}L)^{-1}$ и леммы 1.11 получаем неравенство

$$|\nu_{n,h}(\mathbf{t})| \le 2\rho^{-1}|\mu_{n,h}(\mathbf{t})|, \quad \text{где} \quad \mu_{n,h}(\mathbf{t}) = \sum_{i=1}^{n} K_h(\mathbf{t}-\mathbf{z}_i)\Lambda_k(\mathcal{P}_i)\varepsilon_i.$$

Далее с помощью метода диадических цепочек оценим хвост распределения $\sup\limits_{\mathbf{t}\in\mathcal{P}}|\mu_{n,h}(\mathbf{t})|$.

Заметим, прежде всего, что множество $\mathcal{P}$ под знаком супремума можно заменить множеством двоично-рациональных точек $\mathcal{R} = \cup_{l\ge 1}\mathcal{R}_l$, где

$$\mathcal{R}_l = \{(j_1/2^l, \dots, j_k/2^l) :\ j_1 = 1, \dots, 2^l - 1; \dots; j_k = 1, \dots, 2^l - 1\}.$$

Таким образом,

$$\sup_{\mathbf{t}\in\mathcal{P}}|\mu_{n,h}(\mathbf{t})| \le \sup_{\mathbf{t}\in\mathcal{R}}|\mu_{n,h}(\mathbf{t})| \le \max_{\mathbf{t}\in\mathcal{R}_m}|\mu_{n,h}(\mathbf{t})| + \sum_{l=m+1}^{\infty}\sum_{r=1}^{k}\max_{\mathbf{t}\in\mathcal{R}_l}\left|\mu_{n,h}(\mathbf{t}+2^{-l}\mathbf{e}_r) - \mu_{n,h}(\mathbf{t})\right|,$$

где натуральное m будет выбрано далее, $\mathbf{e}_r$ есть k-мерный вектор с единичной r-ой компонентой и нулями в качестве остальных компонент. Следовательно,

$$\begin{aligned}
&\mathbb{P}_{\mathcal{F}_n}\left(\sup_{\mathbf{t}\in\mathcal{P}}|\mu_{n,h}(\mathbf{t})| > y\right) \le \mathbb{P}_{\mathcal{F}_n}\left(\max_{\mathbf{t}\in\mathcal{R}_m}|\mu_{n,h}(\mathbf{t})| > a_m y\right) + \\
&\qquad + \sum_{l=m+1}^{\infty}\sum_{r=1}^{k}\mathbb{P}_{\mathcal{F}_n}\Big(\max_{\mathbf{t}\in\mathcal{R}_l}\left|\mu_{n,h}(\mathbf{t}+2^{-l}\mathbf{e}_r) - \mu_{n,h}(\mathbf{t})\right| > a_l y/k\Big) \le \\
&\le \sum_{\mathbf{t}\in\mathcal{R}_m}\mathbb{P}_{\mathcal{F}_n}(|\mu_{n,h}(\mathbf{t})| > a_m y) + \sum_{l=m+1}^{\infty}\sum_{r=1}^{k}\sum_{\mathbf{t}\in\mathcal{R}_l}\mathbb{P}_{\mathcal{F}_n}\Big(\left|\mu_{n,h}(\mathbf{t}+2^{-l}\mathbf{e}_r) - \mu_{n,h}(\mathbf{t})\right| > a_l y/k\Big),
\end{aligned} \tag{105}$$

где $a_m, a_{m+1}, \dots$ — последовательность положительных чисел с условием $a_m + a_{m+1} + \dots = 1$.

Чтобы оценить вероятность $\mathbb{P}_{\mathcal{F}_n}(|\mu_{n,h}(\mathbf{t})| > a_m y)$, воспользуемся следующим мартингальным неравенством (см. [241], теорема 2.1):

$$\mathbb{E}\Big|\sum_{i=1}^{n}\eta_i\Big|^p \le \left((p-1)\sum_{i=1}^{n}\left(\mathbb{E}|\eta_i|^p\right)^{2/p}\right)^{p/2}, \tag{106}$$

где $\{\eta_i\}$ — последовательность мартингал-разностей с конечными моментами порядка $p \ge 2$.

Положим

$$\eta_i = K_h(\mathbf{t}-\mathbf{z}_i)\Lambda_k(\mathcal{P}_i)\varepsilon_i.$$

Из (106) и элементарных оценок

$$K_h(\mathbf{t}-\mathbf{z}_i)\Lambda_k(\mathcal{P}_i) \le Lh^{-k}\delta_n^k, \qquad \sum_{i=1}^n K_h(\mathbf{t}-\mathbf{z}_i)\Lambda_k(\mathcal{P}_i) \le Lh^{-k}(2h+2\delta_n)^k$$

получаем, что с вероятностью 1 выполнено

$$\sum_{i=1}^n \left(\mathbb{E}_{\mathcal{F}_n}|\eta_i|^p\right)^{2/p} \le M_p^{2/p}\, 2^k L^2\, (1+\delta_n/h)^k\, (\delta_n/h)^k.$$

Далее, последнее неравенство и (106), с учетом оценки $\delta_n \le h \le 1$, влекут соотношение

$$\mathbb{P}_{\mathcal{F}_n}(|\mu_{n,h}(\mathbf{t})| > a_m y) \le \frac{\mathbb{E}_{\mathcal{F}_n}(|\mu_{n,h}(\mathbf{t})|^p)}{(a_m y)^p} \le G_1 \frac{(\delta_n/h)^{kp/2}}{(a_m y)^p} \quad \text{п.н.}, \tag{107}$$

где $G_1 = (p-1)^{p/2}\, 2^{kp} L^p\, M_p$.

Чтобы оценить $\mathbb{P}_{\mathcal{F}_n}\Big(\big|\mu_{n,h}(\mathbf{t}+2^{-l}\mathbf{e}_r)-\mu_{n,h}(\mathbf{t})\big| > a_l y/k\Big)$, воспользуемся неравенством (106) при

$$\eta_i = \Big(K_h(\mathbf{t}-\mathbf{z}_i+2^{-l}\mathbf{e}_r) - K_h(\mathbf{t}-\mathbf{z}_i)\Big)\Lambda_k(\mathcal{P}_i)\varepsilon_i.$$

Имеем

$$\Big|K_h(\mathbf{t}-\mathbf{z}_i+2^{-l}\mathbf{e}_r) - K_h(\mathbf{t}-\mathbf{z}_i)\Big|\Lambda_k(\mathcal{P}_i) \le Lh^{-k-1}\; 2^{-l}\; \delta_n^k,$$

$$\sum_{i=1}^n \Big|K_h(\mathbf{t}-\mathbf{z}_i+2^{-l}\mathbf{e}_r) - K_h(\mathbf{t}-\mathbf{z}_i)\Big|\Lambda_k(\mathcal{P}_i) \le Lh^{-k-1}\; 2^{-l+1}\; (2h+2\delta_n)^k.$$

Таким образом,

$$\sum_{i=1}^n \left(\mathbb{E}_{\mathcal{F}_n}|\eta_i|^p\right)^{2/p} \le M_p^{2/p}\, 2^k L^2\, 2^{-2l+1}\, (1+\delta_n/h)^k\, (\delta_n/h)^k.$$

Кроме того, с учетом соотношения $\delta_n \le h \le 1$, из последнего неравенства и (106) следует, что

$$\mathbb{P}_{\mathcal{F}_n}\Big(\big|\mu_{n,h}(\mathbf{t}+2^{-l}\mathbf{e}_r)-\mu_{n,h}(\mathbf{t})\big| > a_l y/k\Big) \le \frac{G_2}{k}\; \frac{(\delta_n/h)^{kp/2}\; h^{-p}\; 2^{-lp}}{(a_l y)^p}, \tag{108}$$

где $G_2 = 2^{p/2}\, k^{p+1}\, (p-1)^{p/2}\, 2^{kp} L^p\, M_p = 2^{p/2}\, k^{p+1}\, G_1$.

Используя теперь (105), (107) и (108) заключаем, что

$$\mathbb{P}_{\mathcal{F}_n}\left(\sup_{\mathbf{t}\in\mathcal{P}} |\mu_{n,h}(\mathbf{t})| > y\right) < y^{-p}(\delta_n/h)^{kp/2}\left(G_1 2^{km} a_m^{-p} + G_2\, h^{-p} \sum_{l=m+1}^{\infty} 2^{-(p-k)l}\, a_l^{-p}\right).$$

Оптимальная последовательность a_l, минимизирующая правую часть этого неравенства, есть

$a_m = c(G_1 2^{km})^{1/(p+1)}$, $a_l = c\big(G_2\, h^{-p}\, 2^{-(p-k)l}\big)^{1/(p+1)}$ при $l = m+1, m+2, \ldots$, где коэффициент c определяется соотношением $a_m + a_{m+1} + \cdots = 1$. Для указанной последовательности

$$\mathbb{P}_{\mathcal{F}_n}\left(\sup_{\mathbf{t}\in\mathcal{P}} |\mu_{n,h}(\mathbf{t})| > y\right) \le y^{-p}(\delta_n/h)^{kp/2}\left((G_1 2^{km})^{1/(p+1)} + \sum_{l=m+1}^{\infty}\Big(G_2\, h^{-p} 2^{-(p-k)l}\Big)^{1/(p+1)}\right)^{p+1}.$$

Положим $m = \lceil -\log_2 h\rceil$. В этом случае

$$\begin{aligned}
\mathbb{P}_{\mathcal{F}_n}\left(\sup_{\mathbf{t}\in\mathcal{P}} |\mu_{n,h}(\mathbf{t})| > y\right) \le \\
&\le y^{-p}\,\delta_n^{kp/2}\, h^{-k(1+(p/2))}\left((2G_1)^{1/(p+1)} + \big(G_2 2^{-(p-k)}\big)^{1/(p+1)}\sum_{l=0}^{\infty} 2^{-(p-k)l/(p+1)}\right)^{p+1} \le \\
&\le y^{-p}\,\delta_n^{kp/2}\, h^{-k(1+(p/2))} 2^{p/2}\, G_1\left(1 + \frac{k}{2^{(p-k)/(p+1)} - 1}\right)^{p+1}.
\end{aligned}$$

Это соотношение завершает вывод леммы 1.12 при

$$G(k,p) = 2^p\, 2^{p/2}\, \frac{G_1}{L^p\, M_p}\left(1 + \frac{k}{2^{(p-k)/(p+1)} - 1}\right)^{p+1},$$

а вместе с тем и доказательство теоремы 1.4. □

Теорема 1.5 по существу вытекает из следующего утверждения, доказанного в [111, следствие 3.1].

Лемма 1.13. *Пусть $\{S_{ni}, \widetilde{\mathcal{F}}_{n,i},\ 1 \le i \le k_n, n \ge 1\}$ — квадратично интегрируемый мартингал с нулевым средним и мартингал-разностью Y_{ni} и η — некоторая почти наверное конечная случайная величина. Предположим, что σ-алгебры $\{\widetilde{\mathcal{F}}_{n,i}\}$ образуют возрастающий поток по аргументу n, т.е. $\widetilde{\mathcal{F}}_{n,i} \subseteq \widetilde{\mathcal{F}}_{n+1,i}$ для всех $1 \le i \le k_n$, $n \ge 1$, и, кроме того,*

$$\sum_{i=1}^{k_n} \mathbb{E}[Y_{ni}^2 I(Y_{ni} > \varepsilon)|\widetilde{\mathcal{F}}_{n,i-1}] \xrightarrow{p} 0 \quad \forall\varepsilon > 0, \quad \sum_{i=1}^{k_n} \mathbb{E}(Y_{ni}^2|\widetilde{\mathcal{F}}_{n,i-1}) \xrightarrow{p} \eta^2.$$

Тогда $S_{nk_n} = \sum_{i=1}^{k_n} Y_{ni} \xrightarrow{d} Z$, где характеристическая функция случайной величины Z определяется соотношением $\mathbb{E}\exp\{-\eta^2 t^2/2\}$.

Доказательство теоремы 1.5. Имеем

$$B_{n,h}^{-1}(\mathbf{t})\left(f_{n,h}^*(\mathbf{t}) - f(\mathbf{t}) - r_{n,h}(\mathbf{t})\right) = B_{n,h}^{-1}(\mathbf{t})\nu_{n,h}(\mathbf{t}) = \frac{\sum_{i=1}^{n} K_h\left(\mathbf{t} - \mathbf{z}_i\right)\varepsilon_i \Lambda_k(\mathcal{P}_i)}{\left(\sum_{i=1}^{n}\big(K_h(\mathbf{t} - \mathbf{z}_i)\Lambda_k(\mathcal{P}_i)\big)^2\right)^{1/2}}. \tag{109}$$

Теперь в формулировке леммы 1.13 положим $k_n = n$, $\widetilde{\mathcal{F}}_{n,i} = \mathcal{F}_{n,i}$ и

$$Y_{ni} = \frac{K_h(\mathbf{t}-\mathbf{z}_i)\,\varepsilon_i\Lambda_k(\mathcal{P}_i)}{\left(\sum_{i=1}^n \left(K_h(\mathbf{t}-\mathbf{z}_i)\Lambda_k(\mathcal{P}_i)\right)^2\right)^{1/2}}.$$

Заметим, что в силу условий теоремы

$$\sum_{i=1}^n \mathbb{E}(Y_{ni}^2|\widetilde{\mathcal{F}}_{n,i-1}) = \sum_{i=1}^n \frac{K_h^2(\mathbf{t}-\mathbf{z}_i)\,\Lambda_k^2(\mathcal{P}_i)\mathbb{E}(\varepsilon_i^2 \mid \mathcal{F}_{n,i-1})}{\sum_{i=1}^n \left(K_h(\mathbf{t}-\mathbf{z}_i)\Lambda_k(\mathcal{P}_i)\right)^2} \equiv \sigma^2$$

почти наверное. Кроме того, с вероятностью 1 имеет место оценка

$$\sum_{i=1}^n \mathbb{E}[Y_{ni}^2 I(Y_{ni} > \varepsilon)|\widetilde{\mathcal{F}}_{n,i-1}] \le \sum_{i=1}^n \frac{K_h^2(\mathbf{t}-\mathbf{z}_i)\,\Lambda_k^2(\mathcal{P}_i)\mathbb{E}(\varepsilon_i^2 I(\varepsilon_i^2 > a/\tau_n)|\mathcal{F}_{n,i-1})}{\sum_{i=1}^n \left(K_h(\mathbf{t}-\mathbf{z}_i)\Lambda_k(\mathcal{P}_i)\right)^2} \le$$
$$\le \max_{i\le n} \mathbb{E}\left(\varepsilon_i^2\, I(\varepsilon_i^2 > a/\tau_n)|\mathcal{F}_{n,i-1}\right).$$

Иными словами, в условиях леммы 1.13 случайная величина $\eta^2 \equiv \sigma^2$ вырождена, т.е. имеет место центральная предельная теорема. □

1.3.3 Примеры компьютерного моделирования и обработки реальных данных

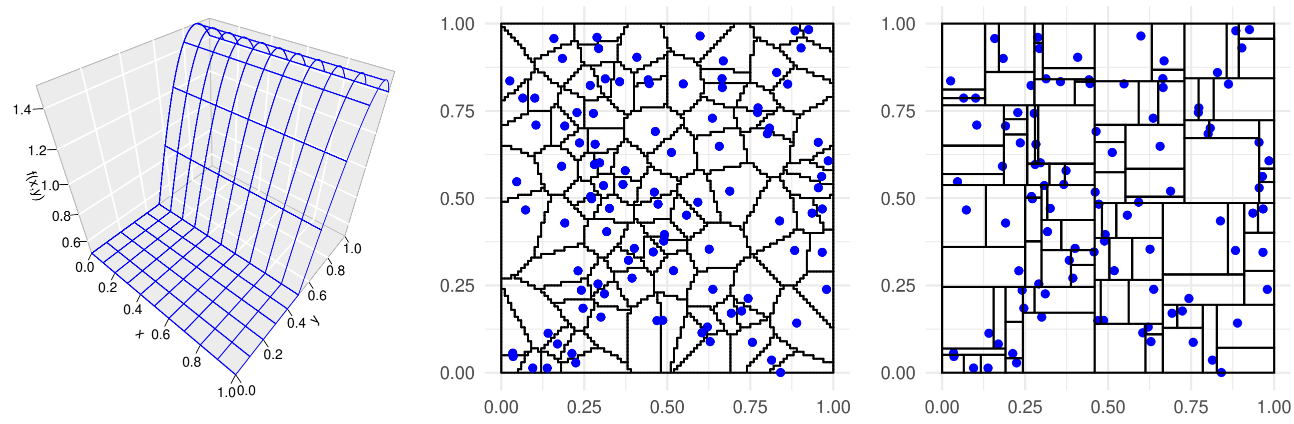
В этом разделе с помощью компьютерного моделирования проведено некоторое сравнение новых универсальных локально–постоянных оценок $f^*_{n,h}(\mathbf{t})$, введенных в (96), и оценок Надарая–Ватсона. Кроме того, оценки сравниваются и на примере обработки реальных данных. Всюду в этом разделе считаем, что $k = 2$.

Рассматриваются следующие два алгоритма вычисления разбиения $\{\mathcal{P}_i\}$ множества $\mathcal{P}$, которое участвует в определении оценки $f^*_{n,h}(\mathbf{t})$.

Первый алгоритм — это мозаика Вороного. Напомним, что мозаика (диаграмма) Вороного некоторого конечного множества точек на плоскости представляет собой такое разбиение плоскости, при котором каждый элемент этого разбиения содержит ровно по одной точке из указанного набора и образует множество всех точек, для которых указанная точка из набора является ближайшей. Таким образом, для каждого i множество $\mathcal{P}_i$ — это ячейка Вороного, соответствующая $\mathbf{z}_i$, т. е. множество всех точек $\mathcal{P}$, расположенных ближе к $\mathbf{z}_i$, чем к любой другой точке. Для расчета площадей ячеек использовался пакет deldir R-библиотеки. Пример разбиения множества $\mathcal{P}$ с помощью мозаики Вороного приведен на рисунках 15b) и 17a).

Второй — алгоритм последовательных покоординатно-медианных сечений. Прежде всего, разделим $\mathcal{P}$ на два прямоугольника прямой $t_1 = \mathrm{median}\{z_{11}, \ldots, z_{1n}\}$, равной середине интервала $\left(z_{n:1\lfloor n/2\rfloor},\ z_{n:1\lfloor n/2\rfloor+1}\right)$, где точки $z_{n:11}, \ldots, z_{n:1n}$ есть вариационный ряд, построенный по проекциям регрессоров на ось абсцисс. Далее каждый из двух прямоугольников делится рекуррентным образом следующим образом: если на некотором шаге прямоугольник содержит две или более точек из набора регрессоров и ширина прямоугольника больше его высоты, то прямоугольник разбивается на два прямой $t_1 = \mathrm{median}\{z_{1j} : j \in B\}$, где B есть множество индексов регресоров, попадающих в прямоугольник. В противном случае прямоугольник разбивается прямой $t_2 = \mathrm{median}\{z_{2j} : j \in B\}$. Как только в прямоугольнике оказывается одна точка $\mathbf{z}_i$ из набора регрессоров, он полагается равным $\mathcal{P}_i$. Пример такого разбиения приведен на рисунках 15c) и 17b).

П р и м е р 1.11. В этом примере аппроксимируется неслучайная регрессионная функция $f(x, y) = \max\{0, \sin(4y - 2)\} + 0.5$, изображенная на рисунке 15a). Регрессоры $\{\mathbf{z}_i\}$ независимы и имеют равномерное распределение в единичном квадрате $\mathcal{P} = [0, 1] \times [0, 1]$, погрешности $\{\varepsilon_i\}$ независимы и имеют нормальное распределение со стандартным отклонением $\sigma = 0.1$.



(a) $f(x, y)$ (b) Метод Вороного (c) Метод медианных сечений

Рис. 15: Иллюстрация к примеру 1.11 (график функции f и разбиения множества $\mathcal{P}$).

В супремальной норме сравниваются три варианта оценок: новая универсальная локально–постоянная оценка $f^*_{n,h}(\mathbf{t})$ с использование разбиения Вороного, новая универсальная локально–постоянная оценка $f^*_{n,h}(\mathbf{t})$ с использование разбиения методом последовательных покоординатно-медианных сечений, классическая локально–постоянная оценка $\hat{f}_{NW}(\mathbf{t})$ (оценка Надарая–Ватсона). Рассматриваются два случая в зависимости от объема выборки: $n = 100$ и $n = 1000$. Результаты компьютерного моделирования представлены на рисунке 16, содержащем графики эмпирических функций распределения для следующих трех величин:

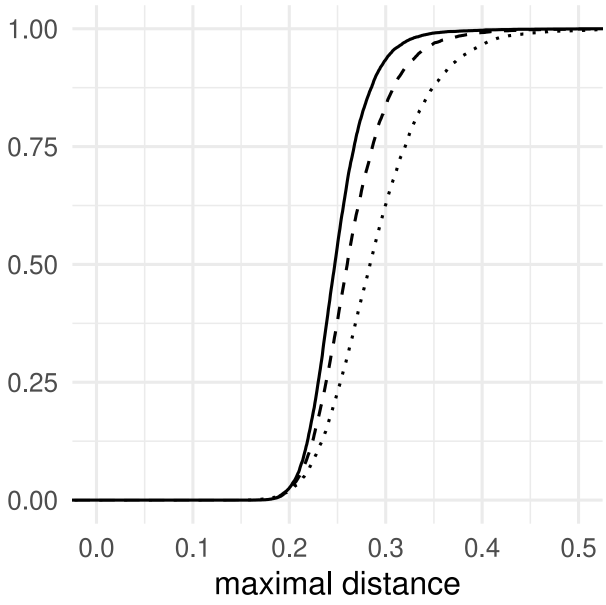
1) $\max_{\mathbf{t}\in\mathcal{P}} |f^*_{n,h}(\mathbf{t}) - f(\mathbf{t})|$ для оценки с разбиением Вороного (сплошная линия),

2) $\max_{\mathbf{t}\in\mathcal{P}} |f^*_{n,h}(\mathbf{t}) - f(\mathbf{t})|$ для оценки с разбиением медианными сечениями (пунктирная линия),

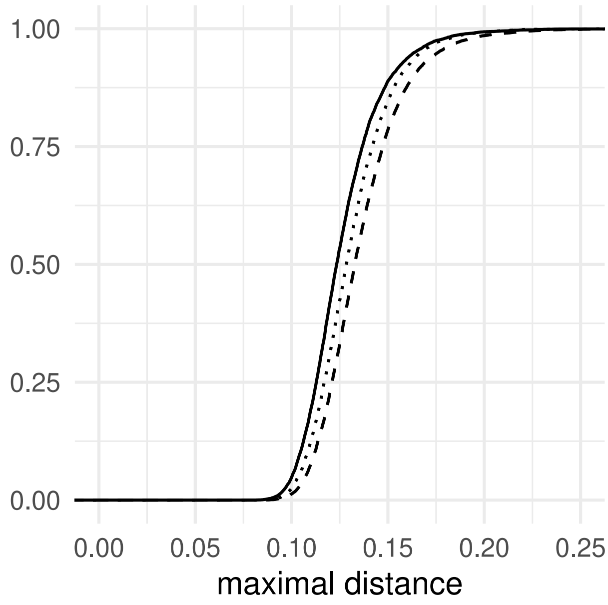
3) $\max\limits_{\mathbf{t}\in\mathcal{P}} |\hat{f}_{NW}(\mathbf{t}) - f(\mathbf{t})|$ (точечная линия).

Всюду максимум вычислен по сетке 1000×1000. Каждый из графиков эмпирических функций распределения построен по 10000 реализациям (прогонкам).

В этом примере новая локально–постоянная оценка $f^*_{n,h}(\mathbf{t})$ оказалась точнее в случае



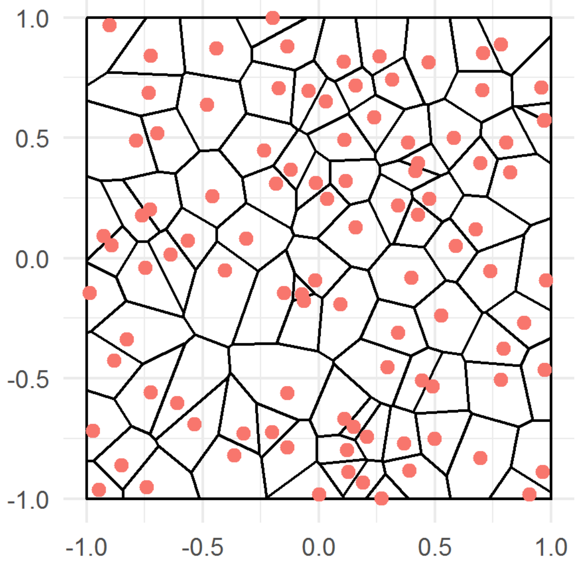
(a) $n = 100$, $h = 0.3$



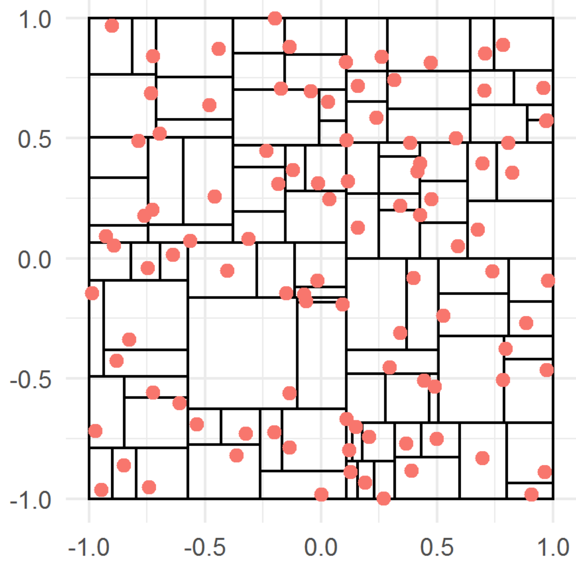
(b) $n = 1000$, $h = 0.1$

Рис. 16: Иллюстрация к примеру 1.11. На рисунках изображены графики эмпирических функций распределения, построенных по 10000 прогонкам, для следующих величин: $\max_{\mathbf{t}\in\mathcal{P}} |f^*_{n,h}(\mathbf{t}) - f(\mathbf{t})|$ с разбиением Вороного (сплошная линия), $\max_{\mathbf{t}\in\mathcal{P}} |f^*_{n,h}(\mathbf{t}) - f(\mathbf{t})|$ с разбиением методом последовательных покоординатно-медианных сечений (пунктирная линий), $\max_{\mathbf{t}\in\mathcal{P}} |\hat{f}_{NW}(\mathbf{t}) - f(\mathbf{t})|$ (точечная линия).

использования алгоритма Вороного для разбиения $\mathcal{P}$, нежели алгоритма последовательных покоординатно-медианных сечений. В обоих случаях оценка $f^*_{n,h}(\mathbf{t})$ с разбиением Вороного оказалась точнее двух других оценок (оценки $f^*_{n,h}(t)$ со вторым алгоритмом разбиения и оценки Надарая–Ватсона).



(a)



(b)

Рис. 17: Разбиение методом Вороного (слева) и методом покоординатно–медианных сечений (справа) множества $\mathcal{P} = [-1, 1] \times [-1, 1]$ для одного и того же набора точек.

Как и ранее в примере 1.11, далее сравниваются три варианта ядерных оценок: оценка Надарая–Ватсона (NW), универсальная локально–постоянная оценка с разбиением Вороного (ULCV), универсальная локально–постоянная оценка с разбиением покоординатно–медианными сечениями (ULCV) (в скобках указана аббревиатура, которая используется далее в графических иллюстрациях). Но во всех приводимых далее примерах алгоритм моделирования усложняется следующим образом. В каждом примере было выполнено по 1000 реализаций (прогонок), при этом в каждой из реализаций моделировалось $n = 5000$ расчетных точек (регрессоров) $\{\mathbf{z}_i;\ i = 1, \dots 5000\}$. Для расчетных точек $\mathbf{z}_i$ наблюдения X_i моделировались с помощью независимых гауссовских погрешностей со стандартным отклонением $\sigma = 0.5$. Выборка $\{(\mathbf{z}_i, X_i),\ i = 1, \dots, 5000\}$ случайным образом делились на обучающую (80%) и проверочную (20%). Для каждой из рассматриваемых ядерных оценок по обучающей выборке методом перекрестной проверки (кросс–валидации) вычислялся оптимальный размер окна h. Случайное разбиение для перекрестной проверки было одинаковым для всех оценок. Далее оптимальный параметр сглаживания h использовался при вычислении двух вариантов погрешностей оценивания, но уже по наблюдениям проверочной выборки: среднеквадратичной ошибки (MSE) и максимальной абсолютной ошибки (MaxE). Указанные характеристики для каждой из ядерных оценок определялись следующим образом:

$$\mathrm{MSE} = \frac{1}{m}\sum_j (\widetilde{f}_h(\mathbf{z}_j) - X_j)^2, \qquad \mathrm{MaxE} = \max_j |\widetilde{f}_h(\boldsymbol{\gamma}_j) - f(\boldsymbol{\gamma}_j)|, \tag{110}$$

где суммирование ведется по индексам j, соответствующим элементам проверочной выборки, максимум вычисляется по элементам $\boldsymbol{\gamma}_j$ равномерной сетки 100×100 на множестве $\mathcal{P}$, m — размер проверочной выборки, $\widetilde{f}_h$ — оценка для f. Результаты моделирования и вычислений представлены далее графически в виде диаграмм «ящик с усами», где «ящик» отображает медиану, $0,25$ и $0,75$ квантили (первый и третий квартили). Кроме того, производится сравнение некоторых оценок с помощью парного теста Вилкоксона.

В приведенных ниже примерах 1.12–1.14 мы намеренно выбрали конфигурацию расчетных точек (регрессоров) с высокой неравномерностью, чтобы продемонстрировать возможные преимущества новой оценки. В указанных примерах моделирования $\mathcal{P} = [-1, 1] \times [-1, 1]$. Во всех приведенных далее примерах использовалось трикубическое ядро

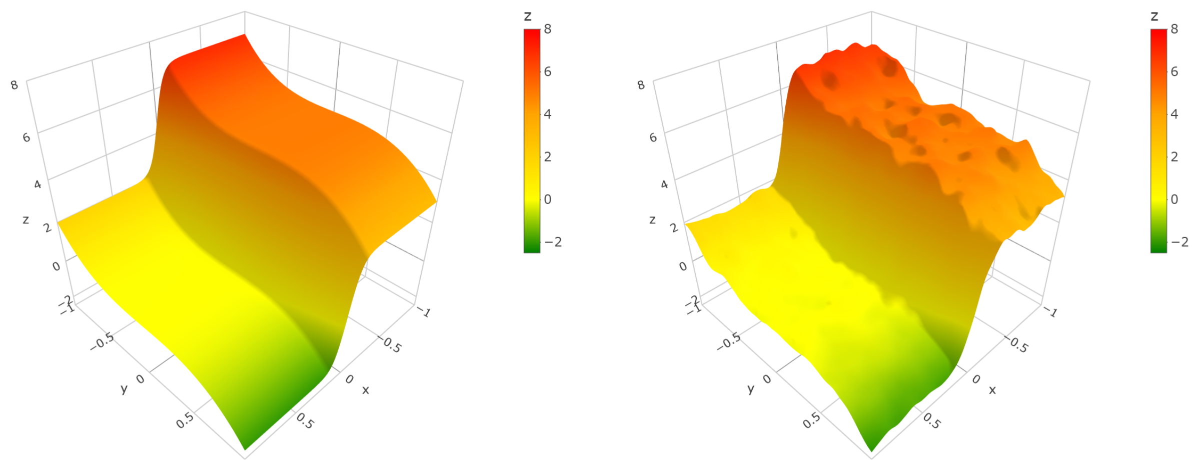
$$K(x, y) = \frac{440}{162\pi} \max\left\{0, \left(1 - \sqrt{x^2 + y^2}^{\,3}\right)^3\right\}.$$

П р и м е р 1.12. Регрессионная функция задается соотношением

$$f(x, y) = \frac{5}{1 + e^{-20x}} - 2y^3.$$

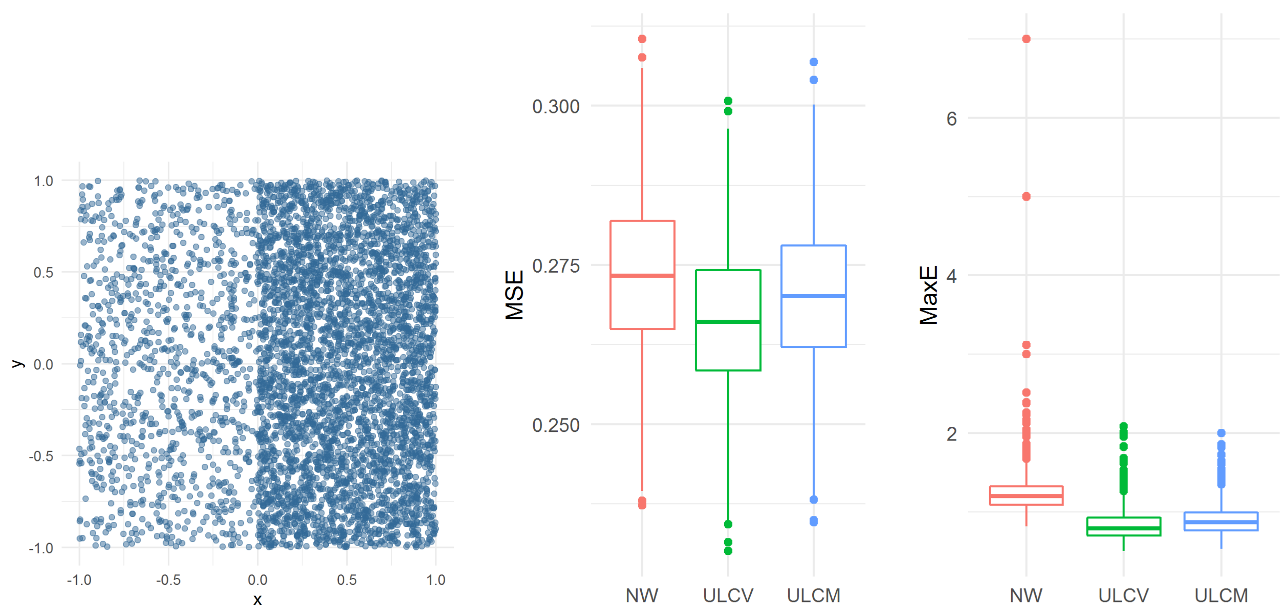
Равномерно распределенные в квадрате $\mathcal{P} = [-1, 1] \times [-1, 1]$ случайные величины $\{\mathbf{z}_i\}$ задаются по алгоритму примера 1.2: вначале мы разыгрываем в какой из двух прямоугольников

($[-1,0)\times[-1,1]$ или $[0,1]\times[-1,1]$) бросаем наудачу первую точку, а далее чередуем количество бросаний в каждый из двух указанных прямоугольников следующим образом: 10, 10^2, 10^3 и т.д. На рисунке 19a) показана одна реализация набора $\{\mathbf{z}_i;\, i=1,\ldots,5000\}$. Регрессионная функция $f(x,y)$, а также вычисленная по одной из реализаций универсальная локально-постоянная оценка с разбиением Вороного (ULCV) для этой функции, приведены на рисунке 18.



(a) функция $f(x,y)$ (b) ULCV-оценка для f

Рис. 18: Иллюстрация к примеру 1.12. На рисунке представлены регрессионная функция f и универсальная локально–постоянная оценка с разбиением Вороного (ULCV) для этой функции (справа).



(a) регрессоры $\{\mathbf{z}_i\}$ (b) MSE (c) MaxE

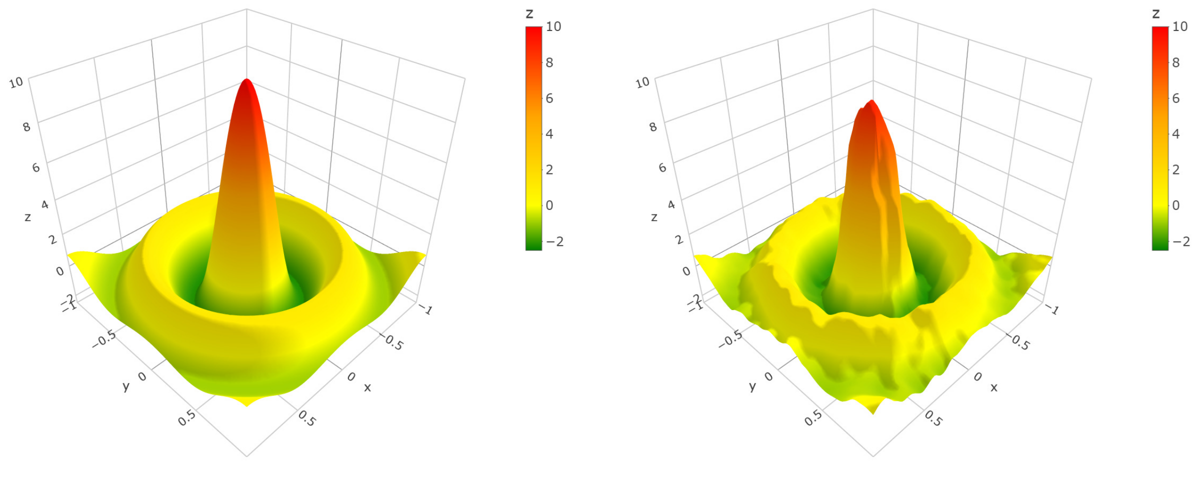
Рис. 19: Иллюстрация к примеру 1.12. На рисунке представлены набор регрессоров $\{\mathbf{z}_i\}$ (слева), среднеквадратичная ошибка оценок (в центре) и максимальная абсолютная ошибка оценок (справа).

Результаты расчетов представлены на рисунке 19. Универсальная локально–постоянная оценка с разбиением Вороного (ULCV) оказалась точнее двух других оценок как с точки зрения среднеквадратичной ошибки (MSE), так и максимальной (MaxE). В частности, универсальная локально–постоянная оценка с разбиением Вороного (ULCV) была лучше оценки Надарая–Ватсона (NW): MSE 0,2661 (0,2584, 0,2742) против 0,2734 (0,2650, 0,2819), $p < 0,0001$; MaxE 0,7878 (0,7013, 0,9230) против 1,1998 (1,0911, 1,3250), $p < 0,0001$. В этом примере и универсальная локально–постоянная оценка с разбиением медианными сечениями (ULCM) оказалась точнее оценки Надарая–Ватсона.

П р и м е р 1.13. Рассматривается следующая регрессионная функция

$$f(x,y) = \sin\left(10\sqrt{x^2+y^2}\right)/\sqrt{x^2+y^2}. \tag{111}$$

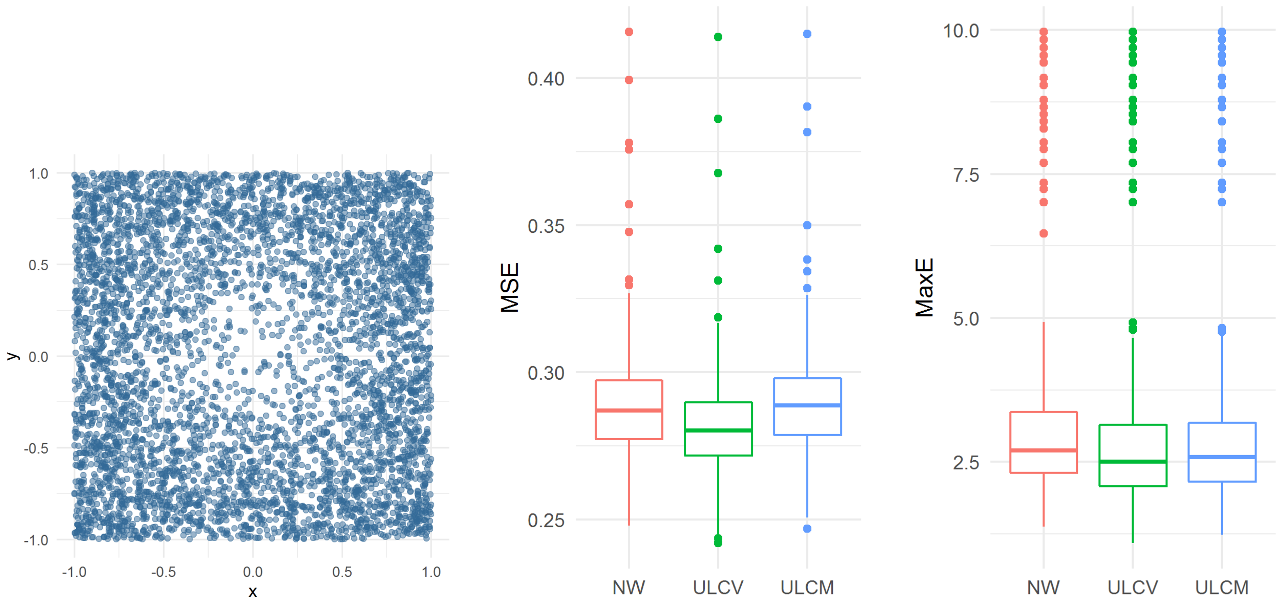
Регрессоры $\{\mathbf{z}_i\}$ моделировались полярными координатами (ρ,φ), где случайные величины ρ и φ независимы, при этом плотность ρ пропорциональна $r^2(2-r)^{1/10}$, $0 \le r \le 2$, а φ равномерно распределена на $[0, 2\pi)$. Регрессоры, не попавшие в множество $\mathcal{P}$, исключались из рассмотрения, при этом генерировались дополнительные величины, чтобы общее количество «подходящих» разыгранных точек оставалось равным $n = 5000$. Одна из реализаций набора $\{\mathbf{z}_i\}$ расчетных точек показана на рисунке 21. График регрессионной функции $f(x,y)$ этого примера, а также универсальная локально–постоянная оценка с разбиением Вороного (ULCV), построенная по одной из реализаций, показаны на рисунке 20.



(a) функция $f(x,y)$ (b) ULCV-оценка для f

Рис. 20: Иллюстрация к примеру 1.13. На рисунке представлены регрессионная функция f и универсальная локально–постоянная оценка с разбиением Вороного (ULCV) для этой функции (справа).

Результаты вычислений представлены на рисунке 21. Универсальная локально–постоянная оценка с разбиением Вороного (ULCV) оказалась точнее двух других оценок как с точки зрения среднеквадратичной ошибки (MSE), так и максимальной (MaxE). В частности, универсальная локально–постоянная оценка с разбиением Вороного (ULCV) лучше оценки Надарая–Ватсона (NW): MSE 0.2803 (0.2718, 0.2898) против 0.2870 (0.2774, 0.2974),



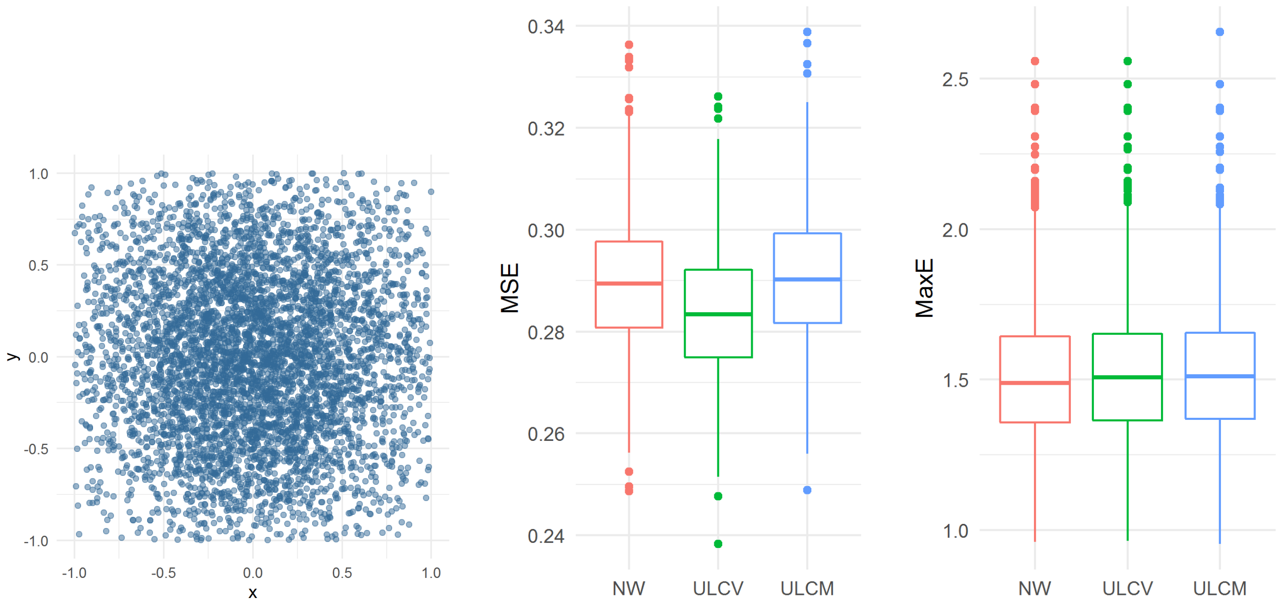
(a) регрессоры $\{\mathbf{z}_i\}$ (b) MSE (c) MaxE

Рис. 21: Иллюстрация к примеру 1.13. На рисунке представлены набор регрессоров $\{\mathbf{z}_i\}$ (слева), среднеквадратичная ошибка оценок (в центре) и максимальная абсолютная ошибка оценок (справа).

$p < 0.0001$; MaxE 2.505 (2.072, 3.140) против 2.695 (2.303, 3.361), $p < 0.0001$. В этом примере универсальная локально–постоянная оценка с разбиением медианными сечениями (ULCM) оказалась точнее оценки Надарая–Ватсона (NW) в среднеквадратичном смысле, и хуже оценки Надарая–Ватсона при сравнении максимальных ошибок (MaxE).

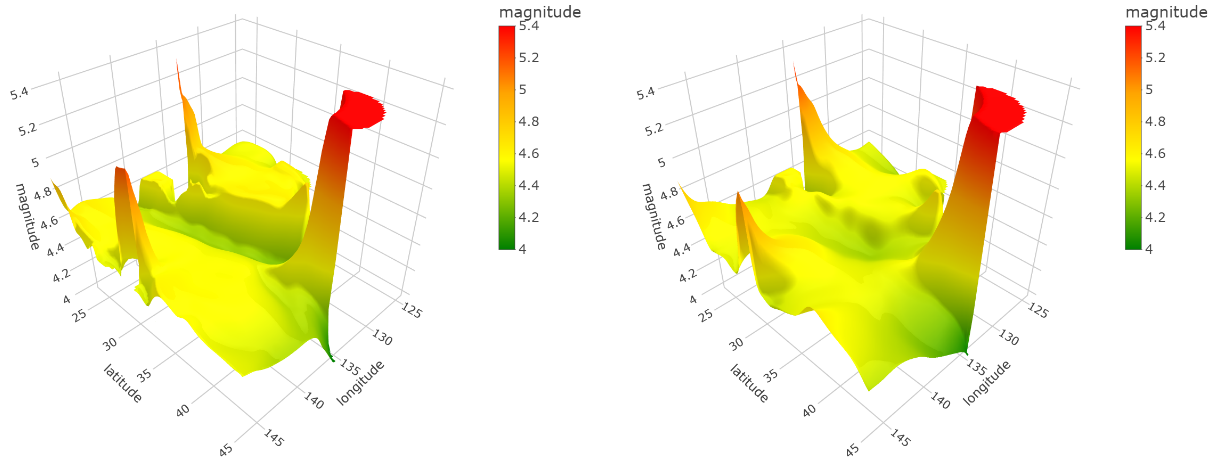
П р и м е р 1.14. В этом примере рассматривается та же неслучайная функция регрессии (см. формулу (111)), как и в примере 1.13. Единственное отличие этого примера от предыдущего состоит в том, что координаты точек $\{\mathbf{z}_i\}$ моделировались как независимые гауссовские случайные величины со средним значением 0 и стандартным отклонением 0.5. Как и в предыдущем примере, использовались только регрессоры, попавшие в множество $\mathcal{P}$. Одна из реализаций набора $\{\mathbf{z}_i;\ i = 1, \ldots, 5000\}$ показана на рисунке 22a).

Результаты вычислений представлены на рисунке 22. Универсальная локально–постоянная оценка с разбиением Вороного (ULCV) оказалась точнее двух других оценок при сравнении среднеквадратичных ошибок (MSE). В частности, в указанном смысле эта новая оценка оказалась точнее оценки Надарая–Ватсона (NW): MSE 0.2834 (0.2750, 0.2922) против 0.2895 (0.2808, 0.2977), $p < 0.0001$. При сравнении максимальных абсолютных ошибок (MaxE) универсальная локально–постоянная оценка с разбиением Вороного (ULCV) оказалась хуже оценки Надарая–Ватсона (NW): 1.507 (1.364, 1.653) против 1.488 (1.357, 1.643), $p < 0.0001$. В этом примере универсальная локально–постоянная оценка с разбиением медианными сечениями (ULCM) оказалась наименее точной в обоих смыслах. Нельзя не отметить, что при сравнении максимальных абсолютных ошибок все три оценки показали близкую точность.



(a) регрессоры $\{\mathbf{z}_i\}$ (b) MSE (c) MaxE

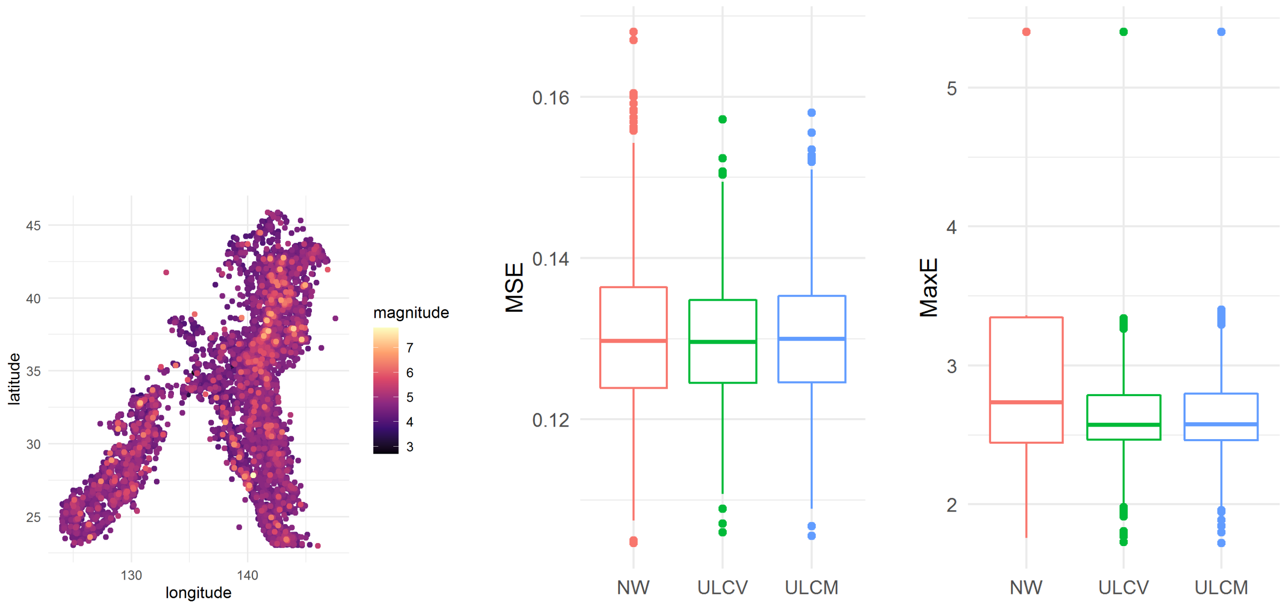
Рис. 22: Иллюстрация к примеру 1.14. На рисунке представлены набор регрессоров $\{\mathbf{z}_i\}$ (слева), среднеквадратичная ошибка оценок (в центре) и максимальная абсолютная ошибка оценок (справа).



(a) NW (b) ULCV

Рис. 23: Иллюстрация к примеру 1.15. Оценка Надарая–Ватсона (слева) и универсальная локально–постоянная оценка с разбиением Вороного (справа) средней магнитуды землетрясений.

Пример 1.15. В этом примере мы сравнили три рассматриваемые ядерные оценки (ULCV, ULCM, NW), используя данные о землетрясениях в Японии, произошедших в 2012-2021 гг. (данные взяты из открытого каталога землетрясений ANSS Comprehensive Earthquake Catalog, 2022 г.). В этом каталоге каждое из 10184 зафиксированных землетрясений представлено своими координатами (долгота и широта) и магнитудой (в диапазоне от 2,7 до 7,8). Данные наблюдений приведены на рисунке 24. С помощью трех вышеуказанных ядерных оценок мы аппроксимировали среднюю магнитуду землетрясений в зависимости от координат. Как и в приведенных выше примерах моделирования, было выбрано (случайным об-



(a) данные о землетрясениях (b) MSE (c) MaxE

Рис. 24: Иллюстрация к примеру 1.15. Данные о землетрясениях в Японии в 2012-2021 гг. (слева), среднеквадратичные ошибки оценок (в центре), максимальные абсолютные ошибки оценок (справа).

разом) 1000 реализаций, в каждой из которых данные были случайным образом разделены на обучающую (80%) и проверочную (20%) выборки. Для каждой из ядерных оценок на обучающей выборке был рассчитан оптимальный размер окна h методом кросс-валидации. Случайное разбиение для кросс-валидации было одинаковым для всех ядерных оценок. Отличие этого примера от примеров 1.12–1.14 состоит в том, что функция f при обработке реальных данных неизвестна. Поэтому максимальная ошибка (MaxE) оценивалась не на истинных значениях оцениваемой функции в точках $\boldsymbol{\gamma}_j$ некоторой решетки (см. формулу (110)), а на проверочной выборке каждой из реализаций по формуле $\mathrm{MaxE} = \max_j |\widetilde{f}_h(\mathbf{z}_j) - X_j|$, где максимум берется по индексам j, соответствующим элементам проверочной выборки. На рисунке 23 изображены универсальная локально–постоянная оценка с разбиением Вороного (ULCV) и оценка Надарая–Ватсона.

Результаты вычислений представлены на рисунке 24. Универсальная локально–постоянная оценка с разбиением Вороного (ULCV) оказалась точнее двух других оценок как с точки зрения среднеквадратичных ошибок, так и максимальных. В частности, универсальная локально–постоянная оценка с разбиением Вороного (ULCV) точнее оценки Надарая–Ватсона (NW): MSE 0.1296 (0.1245, 0.1348) против 0.1297 (0.1239, 0.1364), $p < 0.0001$; MaxE 2.573 (2.464, 2.785) против 2.736 (2.442, 3.346), $p = 0.0005$. В этом примере универсальная локально–постоянная оценка с медианными разбиениями (ULCM) оказалась точнее оценки Надарая–Ватсона (NW) при сравнении максимальных ошибок (MaxE), и проигрывает оценке Надарая–Ватсона с точки зрения среднеквадратичных ошибок (MSE). Отметим, что все три оценки показали практически одинаковую медиану в смысле среднеквадратичной ошибки.

1.3.4 Универсальные локально–линейные оценки

Всюду в разделе считаем выполненными условия (M_1) и (E_2). Без ограничения общности будем считать, что $\mathcal{P}=[0,1]^k$. Нам также потребуется следующее условие на ядро.

$(\mathbf{K}_3)$ *Выполнено условие* (K) *и ядро* $K(\mathbf{t})$ *представимо в виде* $K(\mathbf{t})=\prod_{j=1}^k K_o(t_j)$ *при* $\mathbf{t}=(t_1,\ldots,t_k)^\top$, *где* $K_o(\cdot)$ *есть симметричная плотность распределения с носителем* $[-1,1]$, *т.е.* $K_o(t)\ge 0$, $K_o(t)=K_o(-t)$ *при всех* $t\in[-1,1]$ *и* $\int_{-1}^1 K_o(t)dt=1$. *Считаем, что функция* $K_o(t)$ *удовлетворяет условию Липшица с константой* $1\le L<\infty$ *и* $K_o(\pm1)=0$.

Напомним условие (D_2) относительно набора регрессоров, которое мы будем использовать. Условимся также, что верно соглашение, приведенное в замечании 1.16.

$(\mathbf{D}_2)$ *Для каждого* n *существует такое разбиение множества* $\mathcal{P}$ *на* n *измеримых по Жордану подмножеств* $\{\mathcal{P}_i,\ i=1,\ldots,n\}$, *что каждый элемент этого разбиения содержит ровно по одной точке из набора* $\{\mathbf{z}_i;\, i=1,\ldots,n\}$ *(нумерация элементов разбиения такова, что* $\mathbf{z}_i\in\mathcal{P}_i$*), при этом* $\delta_n=\max_{i\le n} d(\mathcal{P}_i)\xrightarrow{p}0$, *где* $d(A)=\sup_{\mathbf{x},\mathbf{y}\in A}\|\mathbf{x}-\mathbf{y}\|$ *– диаметр множества,* $\|\cdot\|$ *— супремальная норма в* $\mathbb{R}^k$.

Введем ряд соглашений. Для любого вектора $\mathbf{x}=(x_1,\ldots,x_k)^\top$ символ $\|\mathbf{x}\|$ означает супремальную норму в $\mathbb{R}^k$, т.е. $\|\mathbf{x}\|=\max_{j=1,\ldots k}|x_j|$. Для произвольной матрицы X в качестве матричной нормы рассматривается норма, подчиненную супремальной векторной норме, т.е. $\|\mathrm{X}\|=\max_{l\le k}\sum_{m\le k}|\mathrm{X}_{lm}|$, где символ X_{lm} здесь и далее обозначает элемент матрицы X на пересечении l-ой строки и m-го столбца. Поскольку любой k-мерный вектор-столбец $\mathbf{x}$ можно рассматривать как матрицу размерности $k\times 1$, при этом супремальная норма $\|\mathbf{x}\|$ совпадает с введенной выше матричной нормой, то мы будем использовать один символ $\|\cdot\|$ для векторной и матричной норм.

Сохраним обозначение $K_h(\mathbf{t})=h^{-k}K(h^{-1}\mathbf{t})$ и положим

$$\mathbf{x}=(X_1,\ldots,X_n)^\top,\quad \mathrm{W}_{\mathbf{t}}=\mathrm{diag}\{K_h(\mathbf{t}-\mathbf{z}_1)\Lambda_k(\mathcal{P}_1),\ldots,K_h(\mathbf{t}-\mathbf{z}_n)\Lambda_k(\mathcal{P}_n)\},$$
$$\mathrm{Z}_{\mathbf{t}}=\begin{bmatrix}1 & (\mathbf{t}-\mathbf{z}_1)^\top\\ \vdots & \vdots\\ 1 & (\mathbf{t}-\mathbf{z}_n)^\top\end{bmatrix}. \tag{112}$$

Введем в рассмотрение следующий класс оценок для регрессионной функции f:

$$\widetilde{f}^*_{n,h}(\mathbf{t})=\mathbf{e}_1^\top(\mathrm{Z}_{\mathbf{t}}^\top\mathrm{W}_{\mathbf{t}}\mathrm{Z}_{\mathbf{t}})^{-1}\mathrm{Z}_{\mathbf{t}}^\top\mathrm{W}_{\mathbf{t}}\mathbf{x}, \tag{113}$$

где $\mathbf{e}_1=(1,0,\ldots,0)^\top$ — вектор размерности $k+1$, первая координата которого единица, а остальные координаты — нули.

З а м е ч а н и е 1.20. Нетрудно проверить, что ядерная оценка (113) является первой координатой $(k+1)$-мерной оценки следующего варианта взвешенного метода наименьших квадратов:

$$\min_{a,\mathbf{b}} \sum_{i=1}^{n} \left(X_i - \left(a + \mathbf{b}^\top(\mathbf{t} - \mathbf{z}_i)\right)\right)^2 K_h(\mathbf{t} - \mathbf{z}_i)\Lambda_k(\mathcal{P}_i).$$

Таким образом, предлагаемый класс оценок в известном смысле близок к классическим многомерным локально–линейным ядерным оценкам, но во взвешенном методе наименьших квадратов, как и ранее, мы используем несколько иные веса.

Сформулируем теперь основное утверждение раздела.

Теорема 1.6. *Пусть выполнены условия* (M_1), (E_2), (K_3) *и* (D_2). *Тогда для любого фиксированного* $h \in (0, 1/2)$ *с вероятностью* 1 *справедливо соотношение*

$$\sup_{\mathbf{t}\in[0,1]^k} |\widetilde{f}^*_{n,h}(\mathbf{t}) - f(\mathbf{t})| \le C_1^*\omega_f(h) + \widetilde{\zeta}_n(h),$$

где случайная величина $\widetilde{\zeta}_n(h) > 0$ *такова, что*

$$\mathbb{P}\left(\widetilde{\zeta}_n(h) > y\right) \le C_2^* y^{-p}\, h^{-k(p/2+1)}\, \mathbb{E}(\delta_n^{kp/2}) + \ \mathbb{P}(\delta_n > c^*h),$$

где константы $c^* < 1$, C_1^* *и* C_2^*, *определенные далее в* (128) *и леммах* 1.19 *и* 1.21, *зависят от размерности* k *и ядра* K, *а константа* C_2^* *дополнительно зависит еще и от* p *и* M_p.

Сравнивая теоремы 1.4 и 1.6, нетрудно видеть, что с очевидными изменениями сохраняются все утверждения, приведенные после теоремы 1.4 (см. замечание 1.19, следствия 1.5, 1.6 и пример 1.10).

1.3.5 Доказательство теоремы 1.6

Всюду в этом разделе считаем, что выполнены условия теоремы 1.6. Для доказательства теоремы нам потребуется ряд вспомогательных утверждений. Положим

$$\begin{gathered}
\kappa_j = \int_{-1}^{1} |u|^j K_o(u)du,\ j = 1, 2, \qquad \lambda_i(\mathbf{t}) = K_h(\mathbf{t} - \mathbf{z}_i)\Lambda_k(\mathcal{P}_i), \qquad w_{n0}(\mathbf{t}) = \sum_{i=1}^{n} \lambda_i(\mathbf{t}),\\
w_{nuv}(\mathbf{t}) = \sum_{i=1}^{n} (t_u - z_{ui})(t_v - z_{vi})\lambda_i(\mathbf{t}), \quad w_{nu}(\mathbf{t}) = \sum_{i=1}^{n} (t_u - z_{ui})\lambda_i(\mathbf{t}), \quad u, v = 1, \ldots, k,\\
w_0(\mathbf{t}) = \int K_h(\mathbf{t} - \mathbf{z})d\mathbf{z}, \quad w_u(\mathbf{t}) = \int (t_u - z_u)K_h(\mathbf{t} - \mathbf{z})d\mathbf{z}, \quad u = 1, \ldots, k,\\
w_{uv}(\mathbf{t}) = \int (t_u - z_u)(t_v - z_v)K_h(\mathbf{t} - \mathbf{z})d\mathbf{z}, \quad u, v = 1, \ldots, k,
\end{gathered} \tag{114}$$

где $\mathbf{t} = (t_1, \ldots, t_k)^\top$, $\mathbf{z}_i = (z_{i1}, \ldots, z_{ik})^\top$, $\mathbf{z} = (z_1, \ldots, z_k)^\top$.

З а м е ч а н и е 1.21. Подчеркнем, что ввиду свойств плотности $K_h(\cdot)$ область суммирования во всех суммах в (114), а также во всех суммах в приводимых далее формулах (133), (135) и (143) совпадает с множеством

$$N_{n,h}(\mathbf{t}) = \{i : \|\mathbf{t} - \mathbf{z}_i\| \le h, 1 \le i \le n\}, \tag{115}$$

а область интегрирования в (114) соответственно совпадает с множеством

$$A_h(\mathbf{t}) = \{\mathbf{z} \in [0,1]^k : \|\mathbf{t} - \mathbf{z}\| \le h\}.$$

Указанные факты являются принципиальными для дальнейшего анализа.

Лемма 1.14. *При $h < 1/2$ имеют место соотношения*

$$\inf_{\mathbf{t} \in [0,1]^k} \left(w_0(\mathbf{t}) w_{uu}(\mathbf{t}) - w_u^2(\mathbf{t})\right) \ge 2^{-k-1} \left(\kappa_2 - \kappa_1^2\right) h^2,\ u = 1, \ldots, k,$$

$$\forall \mathbf{t} \in [0,1]^k \quad w_0(\mathbf{t}) w_{uv}(\mathbf{t}) - w_u(\mathbf{t}) w_v(\mathbf{t}) = 0,\ u \ne v,\ u, v = 1, \ldots, k,$$

$$\inf_{\mathbf{t} \in [0,1]^k} w_0(\mathbf{t}) = 2^{-k}, \qquad \sup_{\mathbf{t} \in [0,1]^k} w_0(\mathbf{t}) = 1,$$

$$\sup_{\mathbf{t} \in [0,1]^k} w_u(\mathbf{t}) = 2^{-1} \kappa_1 h, \qquad \sup_{\mathbf{t} \in [0,1]^k} w_{uv}(\mathbf{t}) = 2^{-2} \kappa_1^2 h^2,\ u, v = 1, \ldots, k.$$

Д о к а з а т е л ь с т в о. Второе утверждение леммы очевидным образом следует из определений в (114) и представления $K(\mathbf{t}) = \prod_{j=1}^k K_o(t_j)$, введенного в условии (K_3). Остальные утверждения следуют из леммы 1.7 и указанного представления для ядра. □

Лемма 1.15. *При $h < 1/2$ на подмножестве элементарных исходов, определяемых соотношением $\delta_n \le h$, при любых $u, v = 1, \ldots, k$ справедливы неравенства*

$$\sup_{\mathbf{t} \in [0,1]^k} w_{n0}(\mathbf{t}) \le 2^{2k} L^k, \qquad \sup_{\mathbf{t} \in [0,1]^k} |w_{n0}(\mathbf{t}) - w_0(\mathbf{t})| \le k 2^{2k} L^k h^{-1} \delta_n, \tag{116}$$

$$\sup_{\mathbf{t} \in [0,1]^k} |w_{nu}(\mathbf{t})| \le 2^{2k} L^k h, \qquad \sup_{\mathbf{t} \in [0,1]^k} |w_{nu}(\mathbf{t}) - w_u(\mathbf{t})| \le (k+1) 2^{2k} L^k \delta_n, \tag{117}$$

$$\sup_{\mathbf{t} \in [0,1]^k} |w_{nuv}(\mathbf{t})| \le 2^{2k} L^k h^2, \qquad \sup_{\mathbf{t} \in [0,1]^k} |w_{nuv}(\mathbf{t}) - w_{uv}(\mathbf{t})| \le (k+2) 2^{2k} L^k h \delta_n. \tag{118}$$

А на подмножестве элементарных исходов $\delta_n \le (k 2^{3k+1} L^k)^{-1} h$ выполнено

$$\inf_{\mathbf{t} \in [0,1]^k} w_{n0}(\mathbf{t}) \ge 2^{-k-1}, \tag{119}$$

$$\sup_{\mathbf{t} \in [0,1]^k} \left| w_{nuv}(\mathbf{t}) - \tfrac{w_{nu}(\mathbf{t}) w_{nv}(\mathbf{t})}{w_{n0}(\mathbf{t})} - w_{uv}(\mathbf{t}) + \tfrac{w_u(\mathbf{t}) w_v(\mathbf{t})}{w_0(\mathbf{t})} \right| \le (k+2) 2^{6k+3} L^{k+k \wedge 2} h \delta_n. \tag{120}$$

Д о к а з а т е л ь с т в о. Для вывода первых оценок в (116)–(118) нужно учесть замечание 1.21 об области суммирования соответствующих функций и следующие соотношения

$$\sup_{\mathbf{s}\in[-1,1]^k} K(\mathbf{s}) \le L^k, \quad \sum_{i=1}^{n} K_h(\mathbf{t}-\mathbf{z}_i)\Lambda_k(\mathcal{P}_i) \le h^{-k}L^k(2h+2\delta_n)^k \le 2^{2k}L^k. \tag{121}$$

Вторые оценки в (116)–(118) следуют из известной оценки погрешности аппроксимации интегральными суммами Римана соответствующих интегралов от функций, удовлетворяющих условию Липшица:

$$\Big| \sum_{i\in N_{n,h}(\mathbf{t})} g^{u,v}_{\mathbf{t},\alpha,\beta}(\mathbf{z}_i)\Lambda_k(\mathcal{P}_i) - \int\limits_{\mathbf{z}\in A_h(\mathbf{t})} g^{u,v}_{\mathbf{t},\alpha,\beta}(\mathbf{z})d\mathbf{z}\Big| \le \delta_n L_{g^{u,v}_{\mathbf{t},\alpha,\beta}} \sum_{i=1}^{n}\Lambda_k(\mathcal{P}_i), \tag{122}$$

где функции $g^{u,v}_{\mathbf{t},\alpha,\beta}(\mathbf{z}) = (t_u - z_u)^\alpha (t_v - z_v)^\beta K_h(\mathbf{t}-\mathbf{z})$, $\alpha,\beta = 0,1$, заданы при значениях аргумента $z_j \in [0 \vee t_j - h, 1 \wedge t_j + h]$, $j = 1,\ldots,k$, а $L_{g^{u,v}_{\mathbf{t},\alpha,\beta}}$ — константа Липшица функции $g^{u,v}_{\mathbf{t},\alpha,\beta}(\mathbf{z})$. Нетрудно проверить, что в условиях леммы

$$\sup_{t\in[0,1]^k} L_{g^{u,v}_{\mathbf{t},0,0}} \le kL^k h^{-k-1}, \quad \sup_{t\in[0,1]^k} L_{g^{u,v}_{\mathbf{t},1,0}} \le (k+1)L^k h^{-k}, \quad \sup_{t\in[0,1]^k} L_{g^{u,v}_{\mathbf{t},1,1}} \le (k+2)L^k h^{-k+1}.$$

Для завершения вывода (116)–(118) нам остается вместе с определениями (114) учесть оценки (121) и (122).

Наконец, (119) следует из леммы 1.14 и второго соотношения в (116). При выводе (120) мы использовали утверждения лемм 1.14 и 1.15. □

В дальнейшем нам также потребуется свойство липшицивости функций $w_{n0}(\mathbf{t})$, $w_{nu}(\mathbf{t})$ и $w_{nuv}(\mathbf{t})$. Справедлива

Лемма 1.16. *При $h < 1/2$ на множестве элементарных исходов, определяемых соотношением $\delta_n \le h$, для любых $\mathbf{t},\mathbf{t}' \in [0,1]^k$ справедливы неравенства*

$$|w_{n0}(\mathbf{t}') - w_{n0}(\mathbf{t})| \le k2^{2k+1}L^k h^{-1}\|\mathbf{t}'-\mathbf{t}\|, \tag{123}$$

$$|w_{nu}(\mathbf{t}') - w_{nu}(\mathbf{t})| \le (k+1)2^{2k+1}L^k\|\mathbf{t}'-\mathbf{t}\|, \quad u = 1,\ldots,k, \tag{124}$$

$$|w_{nuv}(\mathbf{t}') - w_{nuv}(\mathbf{t})| \le (k+2)2^{2k+1}L^k h\|\mathbf{t}'-\mathbf{t}\|, \quad u,v = 1,\ldots,k. \tag{125}$$

На множестве элементарных исходов с условием $\delta_n \le (k2^{3k+1}L^k)^{-1}h$ выполнено

$$\left| w_{nuv}(\mathbf{t}') - \frac{w_{nu}(\mathbf{t}')w_{nv}(\mathbf{t}')}{w_{n0}(\mathbf{t}')} - w_{nuv}(\mathbf{t}) + \frac{w_{nu}(\mathbf{t})w_{nv}(\mathbf{t})}{w_{n0}(\mathbf{t})} \right| \le 5(k+1)2^{8k+2}L^{3k}h\|\mathbf{t}'-\mathbf{t}\|. \tag{126}$$

Д о к а з а т е л ь с т в о. Прежде всего отметим, что нормированное ядро K_h удовлетворяет условию Липшица с константой $k2^{2k+1}L^k h^{-k-1}$, и для множеств $N_{n,h}(\mathbf{t})$, определенных в (115), в условиях леммы справедлива оценка

$$\sum\nolimits_{i\in N_{n,h}(\mathbf{t})} \Lambda_k(\mathcal{P}_i) \le 2^k(h+\delta_n)^k \le 2^{2k}h^k.$$

Отсюда легко выводим (123):

$$|w_{n0}(\mathbf{t}') - w_{n0}(\mathbf{t})| \le \sum_{i\in N_{n,h}(\mathbf{t}')\cup N_{n,h}(\mathbf{t})} |\lambda_i(\mathbf{t}') - \lambda_i(\mathbf{t})| \le$$
$$\le kL^k h^{-k-1}\|\mathbf{t}'-\mathbf{t}\|\Big(\sum_{i\in N_{n,h}(\mathbf{t}')}\Lambda_k(\mathcal{P}_i) + \sum_{i\in N_{n,h}(\mathbf{t})}\Lambda_k(\mathcal{P}_i)\Big) \le k2^{2k+1}L^k h^{-1}\|\mathbf{t}'-\mathbf{t}\|.$$

Немногим от только что продемонстрированного отличается вывод оценки (124):

$$|w_{nu}(\mathbf{t}') - w_{nu}(\mathbf{t})| \le \sum_{i\in N_{n,h}(\mathbf{t}')\cup N_{n,h}(\mathbf{t})} \lambda_i(\mathbf{t})|t'_u - t_u| + \sum_{i\in N_{n,h}(\mathbf{t}')\cup N_{n,h}(\mathbf{t})} |\lambda_i(\mathbf{t}') - \lambda_i(\mathbf{t})||t_u - z_{ui}| \le$$
$$\le 2^{2k+1}L^k\|\mathbf{t}'-\mathbf{t}\| + k2^{2k+1}L^k\|\mathbf{t}'-\mathbf{t}\|.$$

Точно так же выводится оценка (125) после разбивки исходной суммы на три. Неравенство (125) непосредственно следует из верхних оценок (116) и (117) для рассматриваемых величин, а также из (123)–(124). Приведем лишь основную оценку:

$$|w_{nu}(\mathbf{t}')w_{nv}(\mathbf{t}')w_{n0}(\mathbf{t}) - w_{nu}(\mathbf{t})w_{nv}(\mathbf{t})w_{n0}(\mathbf{t}')| \le |w_{nu}(\mathbf{t}') - w_{nu}(\mathbf{t})||w_{nv}(\mathbf{t}')|w_{n0}(\mathbf{t}) +$$
$$+ |w_{nv}(\mathbf{t}') - w_{nv}(\mathbf{t})||w_{nu}(\mathbf{t})|w_{n0}(\mathbf{t}) + |w_{n0}(\mathbf{t}') - w_{n0}(\mathbf{t})||w_{nu}(\mathbf{t})||w_{nv}(\mathbf{t})| \le$$
$$\le (5k+4)2^{6k}L^{3k}h\|\mathbf{t}'-\mathbf{t}\|.$$

Для доказательства (126) нам осталось использовать оценку (125), а также равномерную нижнюю оценку $\inf\limits_{\mathbf{t}\in[0,1]^k} w_{n0}(\mathbf{t}) \ge 2^{-k-1}$ на множестве элементарных исходов, задаваемых неравенством $\delta_n \le (k2^{3k+1}L^k)^{-1}h$ (см. (119)). После этого надо просто сложить полученные оценки и провести элементарные выкладки, которые мы опускаем. □

Определим элементы квадратной матрицы H размерности $k\times k$ следующим образом:

$$\mathrm{H}_{uv} = w_{nuv}(\mathbf{t}) - w_{n0}^{-1}(\mathbf{t})w_{nu}(\mathbf{t})w_{nv}(\mathbf{t}), \qquad u,v = 1,\ldots,k. \tag{127}$$

Отметим, что в силу неравенства Коши-Буняковского $(\mathrm{H})_{uv} \ge 0$ для всех u,v. Нам также потребуется обозначение

$$c^* = \frac{\left(\kappa_2 - \kappa_1^2\right)^k}{(k+2)!2^{6k^2+4k+2}L^{2k^2-k+k\wedge 2}}. \tag{128}$$

Легко видеть, что разность $\kappa_2 - \kappa_1^2$ представляет собой дисперсию невырожденного распределения, тем самым она строго положительна. Кроме того, $c^* < 1$. Справедлива

Лемма 1.17. *При $h < 1/2$ на подмножестве элементарных исходов, определяемых соотношением $\delta_n \le c^* h$, выполнено*

$$\sup_{t\in[0,1]^k} \|\mathrm{H}^{-1}\| \le k! 2^{6k^2-2k-1} L^{2k(k-1)} (\kappa_2 - \kappa_1^2)^{-k} h^{-2}.$$

Д о к а з а т е л ь с т в о. Определим матрицу H^0 размерности $k \times k$:

$$\mathrm{H}^0_{uv} = w_{uv}(\mathbf{t}) - w_0^{-1}(\mathbf{t}) w_u(\mathbf{t}) w_v(\mathbf{t}), \qquad u, v = 1, \ldots, k.$$

В силу леммы 1.14

$$\mathrm{H}^0 = \mathrm{diag}\left\{ w_{11}(\mathbf{t}) - w_0(\mathbf{t}) w_1^2(\mathbf{t}), \ldots, w_{kk}(\mathbf{t}) - w_0(\mathbf{t}) w_k^2(\mathbf{t}) \right\}$$

и имеет место следующая оценка для детерминанта этой диагональной матрицы с положительными элементами:

$$\inf_{\mathbf{t}\in[0,1]^k} \det\left(\mathrm{H}^0\right) \ge 2^{-k(k+1)} \left(\kappa_2 - \kappa_1^2\right)^k h^{2k}. \tag{129}$$

Далее,

$$\sup_{\mathbf{t}\in[0,1]^k} \left|\det(\mathrm{H}) - \det(\mathrm{H}^0)\right| \le k! k(k+2) 2^{6k+3} L^{k+k\wedge 2} \left(L^{2k} 2^{5k+2}\right)^{k-1} \delta_n h h^{2(k-1)} \le$$
$$\le (k+2)! 2^{5k^2+3k+1} L^{2k^2-k+k\wedge 2} \delta_n h^{2k-1}. \tag{130}$$

При выводе (130) мы использовали определение детерминанта матрицы, оценку (120) леммы 1.15 и учли, что в силу лемм 1.14 и 1.15 при всех u, v

$$\sup_{\mathbf{t}\in[0,1]^k} \mathrm{H}_{uv} \le 2^{5k+2} L^{2k} h^2, \qquad \sup_{\mathbf{t}\in[0,1]^k} \mathrm{H}^0_{uv} \le 2^{k-1} L^2 h^2.$$

Следовательно, на множестве элементарных исходов $\delta_n \le c^* h$ выполнено

$$\inf_{\mathbf{t}\in[0,1]^k} \det(\mathrm{H}) \ge 2^{-k(k+1)-1} \left(\kappa_2 - \kappa_1^2\right)^k h^{2k}. \tag{131}$$

Утверждение леммы нетрудно извлечь теперь непосредственно из определения обратной матрицы и соотношения (131). □

Нам потребуются следующие n-мерные векторы:

$$\mathbf{1} = (1, \ldots, 1)^\top, \quad \boldsymbol{\epsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^\top, \quad \boldsymbol{\delta}_f = \big(f(\mathbf{z}_1) - f(\mathbf{t}), \ldots, f(\mathbf{z}_n) - f(\mathbf{t})\big)^\top.$$

В этом случае $\mathbf{x} = f(\mathbf{t})\mathbf{1} + \boldsymbol{\delta}_f + \boldsymbol{\epsilon}$ и для оценки $\widetilde{f}^*_{n,h}(\mathbf{t})$, определенной в (113), справедливо представление

$$\widetilde{f}^*_{n,h}(\mathbf{t}) = f(\mathbf{t})\mathbf{e}_1^\top(\mathrm{Z}_\mathbf{t}^\top \mathrm{W}_\mathbf{t} \mathrm{Z}_\mathbf{t})^{-1}\mathrm{Z}_\mathbf{t}^\top \mathrm{W}_\mathbf{t}\mathbf{1} + \mathbf{e}_1^\top(\mathrm{Z}_\mathbf{t}^\top \mathrm{W}_\mathbf{t} \mathrm{Z}_\mathbf{t})^{-1}\mathrm{Z}_\mathbf{t}^\top \mathrm{W}_\mathbf{t}\boldsymbol{\delta}_f + \mathbf{e}_1^\top(\mathrm{Z}_\mathbf{t}^\top \mathrm{W}_\mathbf{t} \mathrm{Z}_\mathbf{t})^{-1}\mathrm{Z}_\mathbf{t}^\top \mathrm{W}_\mathbf{t}\boldsymbol{\epsilon}. \quad (132)$$

Используя определения (112) и обозначение из (114), квадратную матрицу $\mathrm{Z}_\mathbf{t}^\top \mathrm{W}_\mathbf{t} \mathrm{Z}_\mathbf{t}$ размера $(k+1) \times (k+1)$ и $k+1$-мерные векторы $\mathrm{Z}_\mathbf{t}^\top \mathrm{W}_\mathbf{t}\mathbf{1}$, $\mathrm{Z}_\mathbf{t}^\top \mathrm{W}_\mathbf{t}\boldsymbol{\delta}_f$ и $\mathrm{Z}_\mathbf{t}^\top \mathrm{W}_\mathbf{t}\boldsymbol{\epsilon}$ можно переписать в виде

$$\begin{aligned}
\mathrm{Z}_\mathbf{t}^\top \mathrm{W}_\mathbf{t} \mathrm{Z}_\mathbf{t} &= \begin{bmatrix} \sum_{i=1}^n \lambda_i(\mathbf{t}) & \sum_{i=1}^n (\mathbf{t} - \mathbf{z}_i)^\top \lambda_i(\mathbf{t}) \\ \sum_{i=1}^n (\mathbf{t} - \mathbf{z}_i)\lambda_i(\mathbf{t}) & \sum_{i=1}^n (\mathbf{t} - \mathbf{z}_i)(\mathbf{t} - \mathbf{z}_i)^\top \lambda_i(\mathbf{t}) \end{bmatrix}, \quad & \mathrm{Z}_\mathbf{t}^\top \mathrm{W}_\mathbf{t}\mathbf{1} &= \begin{bmatrix} \sum_{i=1}^n \lambda_i(\mathbf{t}) \\ \sum_{i=1}^n (\mathbf{t} - \mathbf{z}_i)\lambda_i(\mathbf{t}) \end{bmatrix}, \\
\mathrm{Z}_\mathbf{t}^\top \mathrm{W}_\mathbf{t}\boldsymbol{\delta}_f &= \begin{bmatrix} \sum_{i=1}^n \lambda_i(\mathbf{t})\big(f(\mathbf{z}_i) - f(\mathbf{t})\big) \\ \sum_{i=1}^n (\mathbf{t} - \mathbf{z}_i)\lambda_i(\mathbf{t})\big(f(\mathbf{z}_i) - f(\mathbf{t})\big) \end{bmatrix}, & \mathrm{Z}_\mathbf{t}^\top \mathrm{W}_\mathbf{t}\boldsymbol{\epsilon} &= \begin{bmatrix} \sum_{i=1}^n \lambda_i(\mathbf{t})\varepsilon_i \\ \sum_{i=1}^n (\mathbf{t} - \mathbf{z}_i)\lambda_i(\mathbf{t})\varepsilon_i \end{bmatrix}.
\end{aligned} \quad (133)$$

Лемма 1.18. *Имеет место соотношение*

$$\mathbf{e}_1^\top(\mathrm{Z}_\mathbf{t}^\top \mathrm{W}_\mathbf{t} \mathrm{Z}_\mathbf{t})^{-1}\mathrm{Z}_\mathbf{t}^\top \mathrm{W}_\mathbf{t}\mathbf{1} = 1.$$

Доказательство. Воспользуемся формулой Фробениуса для обращения невырожденной блочной матрацы:

$$\begin{bmatrix} \mathrm{A} & \mathrm{B} \\ \mathrm{C} & \mathrm{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathrm{A}^{-1} + \mathrm{A}^{-1}\mathrm{B}\mathrm{H}^{-1}\mathrm{C}\mathrm{A}^{-1} & -\mathrm{A}^{-1}\mathrm{B}\mathrm{H}^{-1} \\ -\mathrm{H}^{-1}\mathrm{C}\mathrm{A}^{-1} & \mathrm{H}^{-1} \end{bmatrix} \qquad \text{при} \quad \mathrm{H} = \mathrm{D} - \mathrm{C}\mathrm{A}^{-1}\mathrm{B},$$

где A — невырожденная квадратная матрица размера $m_1 \times m_1$ и D — квадратная матрица размера $m_2 \times m_2$. В этой формуле положим $m_1 = 1$, $m_2 = k$,

$$\begin{aligned}
\mathrm{A} &= \sum_{i=1}^n \lambda_i(\mathbf{t}) \equiv w_{n0}(\mathbf{t}), \quad & \mathrm{B} &= \sum_{i=1}^n (\mathbf{t} - \mathbf{z}_i)^\top \lambda_i(\mathbf{t}), \\
\mathrm{C} &= \sum_{i=1}^n (\mathbf{t} - \mathbf{z}_i)\lambda_i(\mathbf{t}), & \mathrm{D} &= \sum_{i=1}^n (\mathbf{t} - \mathbf{z}_i)(\mathbf{t} - \mathbf{z}_i)^\top \lambda_i(\mathbf{t}).
\end{aligned} \quad (134)$$

Тогда ввиду (133) и формулы Фробениуса первая координата интересующего нас вектора выражается соотношением

$$\left(\mathrm{A}^{-1}+\mathrm{A}^{-1}\mathrm{B}\mathrm{H}^{-1}\mathrm{C}\mathrm{A}^{-1}\right)\mathrm{A}-\mathrm{A}^{-1}\mathrm{B}\mathrm{H}^{-1}\mathrm{C}\equiv 1,$$

что доказывает лемму. □

Лемма 1.19. *При $h<1/2$ на подмножестве элементарных исходов, определяемых соотношением $\delta_n\le c^*h$, выполнено*

$$\sup_{t\in[0,1]^k}\left|\mathbf{e}_1^\top(\mathrm{Z}_\mathbf{t}^\top\mathrm{W}_\mathbf{t}\mathrm{Z}_\mathbf{t})^{-1}\mathrm{Z}_\mathbf{t}^\top\mathrm{W}_\mathbf{t}\boldsymbol{\delta}_f\right|\le C_1^*\omega_f(h)\quad \textit{при}\quad C_1^*=1+k\cdot k!2^{6k^2}L^{2k^2-k}(\kappa_2-\kappa_1^2)^{-k}.$$

Д о к а з а т е л ь с т в о. Сохраним определения (134), введенные в лемме 1.18. Используя формулу Фробениуса (см. доказательство леммы 1.18), получаем тождество

$$\mathbf{e}_1^\top(\mathrm{Z}_\mathbf{t}^\top\mathrm{W}_\mathbf{t}\mathrm{Z}_\mathbf{t})^{-1}\mathrm{Z}_\mathbf{t}^\top\mathrm{W}_\mathbf{t}\boldsymbol{\delta}_f=\mathrm{A}^{-1}\mathrm{A}_f+\mathrm{A}^{-1}\mathrm{B}\mathrm{H}^{-1}\mathrm{C}\mathrm{A}^{-1}\mathrm{A}_f-\mathrm{A}^{-1}\mathrm{B}\mathrm{H}^{-1}\mathrm{C}_f,\tag{135}$$

где

$$\mathrm{A}_f=\sum_{i=1}^n\lambda_i(\mathbf{t})\big(f(\mathbf{z}_i)-f(\mathbf{t})\big),\qquad \mathrm{C}_f=\sum_{i=1}^n(\mathbf{t}-\mathbf{z}_i)\lambda_i(\mathbf{t})\big(f(\mathbf{z}_i)-f(\mathbf{t})\big).$$

Рассмотрим каждое из слагаемых в правой части соотношения (135). Отметим, что для скалярного произведения $(\mathbf{x},\mathbf{y})$ в $\mathbb{R}^k$ и рассматриваемой супремальной нормы $\|\cdot\|$ справедлива оценка $|(\mathbf{x},\mathbf{y})|\le k\|\mathbf{x}\|\|\mathbf{y}\|$. Следовательно,

$$|\mathrm{A}^{-1}\mathrm{B}\mathrm{H}^{-1}\mathrm{C}\mathrm{A}^{-1}\mathrm{A}_f|\le k\mathrm{A}^{-1}|\mathrm{A}^{-1}\mathrm{A}_f|\|\mathrm{B}^\top\|\|\mathrm{H}^{-1}\|\|\mathrm{C}\|,$$
$$|\mathrm{A}^{-1}\mathrm{B}\mathrm{H}^{-1}\mathrm{C}_f|\le k\mathrm{A}^{-1}\|\mathrm{B}^\top\|\|\mathrm{H}^{-1}\|\|\mathrm{C}_f\|.$$

Далее, с учетом замечания 1.21 и леммы 1.15 выполнено

$$|\mathrm{A}^{-1}\mathrm{A}_f|\le\omega_f(h)|,\quad \|\mathrm{C}_f\|\le\omega_f(h)\mathrm{A}h,\quad \|\mathrm{C}\|\le\mathrm{A}h,\quad \sup_{t\in[0,1]^k}\|\mathrm{B}^\top\|\le 2^{2k}L^kh.\tag{136}$$

Приведенные соотношения вместе с леммой 1.17 завершают доказательство. □

Лемма 1.20. *Для любых $h<1/2$, $\mathbf{t}\in[0,1]^k$ и $m\in N_{n,h}(\mathbf{t})$ на множестве элементарных исходов, определяемых соотношением $\delta_n\le c^*h$, имеют место следующие оценки:*

$$|\alpha_{n,m}(\mathbf{t})|\le 1+(k+1)!2^{6k^2-1}L^{2k^2-k}(\kappa_2-\kappa_1^2)^{-k}\equiv\alpha,$$
$$\left|\alpha_{n,m}(\mathbf{t}+2^{-l}\mathbf{e}_r)-\alpha_{n,m}(\mathbf{t})\right|\le\overline{\alpha}2^{-l}h^{-1}\quad \textit{при}\quad 2^{-l}\le h\ \textit{и}$$
$$\overline{\alpha}=30(k!)^2k^{3-2k}L^{10k-2}2^{8k^2+21k+5}\left(\kappa_2-\kappa_1^2\right)^{-2k},$$

где $\alpha_{n,m}(\mathbf{t})=(1+\mathrm{B}\mathrm{H}^{-1}\mathrm{C}\mathrm{A}^{-1}-\mathrm{B}\mathrm{H}^{-1}(\mathbf{t}-\mathbf{z}_m))$, а $\mathbf{e}_r$ есть k-мерный вектор с единичной r-ой компонентой и нулями в качестве остальных компонент.

Д о к а з а т е л ь с т в о. Поскольку $\|\mathbf{t}-\mathbf{z}_m\| \le h$ при $m \in N_{n,h}(\mathbf{t})$, то первое утверждение леммы нетрудно извлечь из леммы 1.17, соотношений (136) и следующей оценки:

$$|\alpha_{n,m}(\mathbf{t})| \le 1 + k\mathrm{A}^{-1}\|\mathrm{B}^\top\|\|\mathrm{H}^{-1}\|\|\mathrm{C}\| + \|\mathrm{B}^\top\|\|\mathrm{H}^{-1}\|\|\mathbf{t}-\mathbf{z}_m\|.$$

Докажем второе утверждение. Прежде всего отметим, что

$$\mathrm{H}^{-1} = \frac{\mathrm{adj\ H}}{\det(\mathrm{H})},$$

где элементами присоединенной матрицы adj H являются с точностью до ± 1 дополнительные миноры исходной матрицы H. Как известно, детерминант любой матрицы M размера $k \times k$ с элементами M_{ij}, $i, j = 1, \ldots, k$, представим в виде полилинейной формы

$$\det(\mathrm{M}) = \sum_{\alpha_1,\ldots,\alpha_k} (-1)^{I(\bar{\alpha})}\mathrm{M}_{1\alpha_1}\ldots\mathrm{M}_{k\alpha_k}, \tag{137}$$

где суммирование проводится по всем перестановкам $\bar{\alpha} \equiv (\alpha_1 \ldots \alpha_k)$ натуральных чисел $1, \ldots, k$, а $I(\bar{\alpha})$ — число инверсий в конкретной перестановке $\bar{\alpha}$. Формула (137) нам необходима для представления детерминанта det(H) и дополнительных миноров матрицы H в конструкции обратной матрицы H^{-1} соответственно в виде k- и $k-1$-линейных форм вида (137), построенных по матричным элементам (127).

Доказательство липшицивости слагаемого $\mathrm{BH}^{-1}\mathrm{CA}^{-1}$ и локальной липшицивости слагаемого $\mathrm{BH}^{-1}(\mathbf{t}-\mathbf{z}_m)$ в определении функций $\alpha_{n,m}(\mathbf{t})$ проводится по существу так же, как и в лемме 1.16 (см. вывод (126)). Сначала отметим, что имеет место представление

$$\mathrm{BH}^{-1}\mathrm{CA}^{-1} = \big(A\det(\mathrm{H})\big)^{-1}\sum_{i,j=1}^{k} w_{ni}(\mathbf{t})w_{nj}(\mathbf{t})(-1)^{i+j}\det\big(\mathrm{H}^{(ij)}\big), \tag{138}$$

где $\det(\mathrm{H}^{(ij)})$ – дополнительный минор к элементу

$$\mathrm{H}_{ij} = w_{nij}(\mathbf{t}) + \frac{w_{ni}(\mathbf{t})w_{nj}(\mathbf{t})}{w_{n0}(\mathbf{t})}$$

матрицы H. Заметим, что в силу обозначения (134) соотношение (119) леммы 1.15 можно переписать в виде

$$\inf_{t\in[0,1]^k} \mathrm{A} \ge 2^{-k-1}. \tag{139}$$

Далее, для вычисления константы Липшица нам понадобятся нижние оценки (139) и (131) для знаменателя дроби в правой части (138), которые справедливы на множестве элементарных исходов $\delta_n \le c_* h$ и вторая из которых имеет порядок малости h^{2k}. Кроме того, леммы 1.15 и 1.16 вместе с представлением (137) позволяют утверждать, что и верхняя оценка для det(H) имеет такой же порядок малости по h, а константа Липшица для $\det(\mathrm{H})$ оценивается сверху как $C_3^* h^{2k-1}$. В то же самое время дополнительные миноры $\det\left(\mathrm{H}^{(ij)}\right)$ оцениваются сверху как $C_4^* h^{2k-2}$ с константой Липшица порядка h^{2k-3} равномерно по всем $i, j = 1, \ldots, k$. Нам осталось по сути только повторить выкладки при выводе оценки (126) для вычисления константы Липшица дроби в правой части (138), откуда и следует, что упомянутая константа имеет вид $C_5^* h^{-1}$, где в качестве константы C_5^* можно взять величину

$$C_5^* = 15(k!)^2 k^{3-2k} L^{10k-2} 2^{8k^2+21k+5} \left(\kappa_2 - \kappa_1^2\right)^{-2k}. \tag{140}$$

Для слагаемого $\mathrm{BH}^{-1}(\mathbf{t} - \mathbf{z}_m))$ мы установим, что в h-окрестности точки $\mathbf{t}$ оно обладает свойством локальной липшицивости с константой вида $C_6^* h^{-1}$. По аналогии с (138) представим указанное слагаемое как

$$\mathrm{BH}^{-1}(\mathbf{t} - \mathbf{z}_m) = \left(\det(\mathrm{H})\right)^{-1} \sum_{i,j=1}^{k} w_{ni}(\mathbf{t})(t_j - z_{mj})(-1)^{i+j} \det\left(\mathrm{H}^{(ij)}\right). \tag{141}$$

Повторяя почти дословно предыдущие рассуждения этой леммы, мы приходим к заключению, что в h-окрестности любой точки $\mathbf{t} \in [0,1]^k$ (при условии $\|\mathbf{t} - \mathbf{z}_m\| \leq h$) функция $\mathrm{BH}^{-1}(\mathbf{t} - \mathbf{z}_m)$ удовлетворяет условию Липшица с константой C_6^*, существенно меньшей, чем в (140). Так что итоговую постоянную в лемме можно положить равной $2C_5^*$. $\square$

Лемма 1.21. *Для любых $y > 0$ и $h < 1/2$ на подмножестве элементарных исходов, определяемых соотношением $\delta_n \leq c^* h$, имеет место следующая оценка*:

$$\mathbb{P}_{\mathcal{F}_n}\left(\sup_{\mathbf{t} \in [0,1]^k} \left|\mathbf{e}_1^\top (\mathrm{Z}_\mathbf{t}^\top \mathrm{W}_\mathbf{t} \mathrm{Z}_\mathbf{t})^{-1} \mathrm{Z}_\mathbf{t}^\top \mathrm{W}_\mathbf{t} \boldsymbol{\epsilon}\right| > y\right) \leq C_2^* y^{-p}\, \delta_n^{kp/2}\, h^{-k(1+(p/2))},$$

где константа C_2^* определена в (153).

Д о к а з а т е л ь с т в о. Сохраним обозначения (134). Тогда с учетом (133) и формулы Фробениуса получаем следующее тождество:

$$\widetilde{\nu}_{n,h}(\mathbf{t}) \equiv \mathbf{e}_1^\top (\mathrm{Z}_\mathbf{t}^\top \mathrm{W}_\mathbf{t} \mathrm{Z}_\mathbf{t})^{-1} \mathrm{Z}_\mathbf{t}^\top \mathrm{W}_\mathbf{t} \boldsymbol{\epsilon} = \left(\mathrm{A}^{-1} + \mathrm{A}^{-1}\mathrm{BH}^{-1}\mathrm{CA}^{-1}\right) \mathrm{A}_\varepsilon - \mathrm{A}^{-1}\mathrm{BH}^{-1}\mathrm{C}_\varepsilon, \tag{142}$$

где

$$\mathrm{A}_\varepsilon = \sum_{i=1}^{n} \lambda_i(\mathbf{t})\varepsilon_i, \qquad \mathrm{C}_\varepsilon = \sum_{i=1}^{n} (\mathbf{t} - \mathbf{z}_i)\lambda_i(\mathbf{t})\varepsilon_i. \tag{143}$$

Положим

$$\mu_{n,h}(\mathbf{t}) = \sum_{i \in N_{n,h}(\mathbf{t})} \alpha_{n,i}(\mathbf{t})\lambda_i(\mathbf{t})\varepsilon_i.$$

В силу оценки (119) леммы 1.15 и замечания 1.21, имеем

$$\left|\widetilde{\nu}_{n,h}(\mathbf{t})\right| \leq 2^{k+1}|\mu_{n,h}(\mathbf{t})|. \tag{144}$$

Хвост распределения $\sup\limits_{\mathbf{t}\in[0,1]^k} |\mu_{n,h}(\mathbf{t})|$ оценим с помощью метода диадических цепочек. Заметим, прежде всего, что множество $[0,1]^k$ под знаком супремума можно заменить множеством двоично-рациональных точек $\mathcal{R} = \cup_{l\geq 1}\mathcal{R}_l$, где

$$\mathcal{R}_l = \{(j_1/2^l, \ldots, j_k/2^l) : \ j_1 = 1, \ldots, 2^l - 1; \ldots ; j_k = 1, \ldots, 2^l - 1\}.$$

Таким образом,

$$\sup_{\mathbf{t}\in[0,1]^k} |\mu_{n,h}(\mathbf{t})| = \sup_{\mathbf{t}\in\mathcal{R}} |\mu_{n,h}(\mathbf{t})| \leq \max_{\mathbf{t}\in\mathcal{R}_m} |\mu_{n,h}(\mathbf{t})| + \sum_{l=m+1}^{\infty}\sum_{r=1}^{k} \max_{\mathbf{t}\in\mathcal{R}_l} \left|\mu_{n,h}(\mathbf{t}+2^{-l}\mathbf{e}_r) - \mu_{n,h}(\mathbf{t})\right|,$$

где $m = \lceil -\log_2 h\rceil$. Следовательно,

$$\begin{aligned}\mathbb{P}_{\mathcal{F}_n}\Big(\sup_{\mathbf{t}\in[0,1]^k} |\mu_{n,h}(\mathbf{t})| > y\Big) \leq \mathbb{P}_{\mathcal{F}_n}\Big(\max_{\mathbf{t}\in\mathcal{R}_m} |\mu_{n,h}(\mathbf{t})| > a_m y\Big) + \\ + \sum_{l=m+1}^{\infty}\sum_{r=1}^{k} \mathbb{P}_{\mathcal{F}_n}\Big(\max_{\mathbf{t}\in\mathcal{R}_l} \left|\mu_{n,h}(\mathbf{t}+2^{-l}\mathbf{e}_r) - \mu_{n,h}(\mathbf{t})\right| > a_l y/k\Big) \leq \\ \leq \sum_{\mathbf{t}\in\mathcal{R}_m} \mathbb{P}_{\mathcal{F}_n}(|\mu_{n,h}(\mathbf{t})| > a_m y) + \sum_{l=m+1}^{\infty}\sum_{r=1}^{k}\sum_{\mathbf{t}\in\mathcal{R}_l} \mathbb{P}_{\mathcal{F}_n}\Big(\left|\mu_{n,h}(\mathbf{t}+2^{-l}\mathbf{e}_r) - \mu_{n,h}(\mathbf{t})\right| > a_l y/k\Big),\end{aligned} \tag{145}$$

где $a_m, a_{m+1}, \ldots$ — последовательность положительных чисел с условием $a_m + a_{m+1} + \cdots = 1$.

Чтобы оценить вероятности в правой части (145), нам потребуется мартингальное неравенство (см. [241], теорема 2.1)

$$\mathbb{E}\Big|\sum_{i=1}^{n} \eta_i\Big|^p \leq \left((p-1)\sum_{i=1}^{n}\left(\mathbb{E}|\eta_i|^p\right)^{2/p}\right)^{p/2}, \tag{146}$$

где $\{\eta_i\}$ — последовательность мартингал-разностей с конечными моментами порядка $p \geq 2$.

Для получения оценки для $\mathbb{P}_{\mathcal{F}_n}(|\mu_{n,h}(\mathbf{t})| > a_m y)$, воспользуемся (146) при

$$\eta_i = \alpha_{n,i}(\mathbf{t})K_h(\mathbf{t}-\mathbf{z}_i)\Lambda_k(\mathcal{P}_i)\varepsilon_i.$$

Из (121), леммы 1.20 и элементарной оценки $K_h(\mathbf{t}-\mathbf{z}_i)\Lambda_k(\mathcal{P}_i) \leq L^k(\delta_n/h)^k$ получаем, что с вероятностью 1 выполнено

$$\sum_{i=1}^{n}\left(\mathbb{E}_{\mathcal{F}_n}|\eta_i|^p\right)^{2/p}\le M_p^{2/p}\,\alpha^2 2^{2k}L^{2k}\;(\delta_n/h)^k.$$

Далее, последнее неравенство и (146), с учетом оценки $\delta_n\le h\le 1$, влекут соотношение

$$\mathbb{P}_{\mathcal{F}_n}(|\mu_{n,h}(\mathbf{t})|>a_m y)\le\frac{\mathbb{E}_{\mathcal{F}_n}(|\mu_{n,h}(\mathbf{t})|^p)}{(a_m y)^p}\le G_1\frac{(\delta_n/h)^{kp/2}}{(a_m y)^p}\quad\text{п.н.},\tag{147}$$

где

$$G_1=(p-1)^{p/2}\,\alpha^p 2^{kp}L^{kp}\,M_p.\tag{148}$$

Чтобы оценить $\mathbb{P}_{\mathcal{F}_n}\Big(\big|\mu_{n,h}(\mathbf{t}+2^{-l}\mathbf{e}_r)-\mu_{n,h}(\mathbf{t})\big|>a_l y/k\Big)$, воспользуемся неравенством (106) при

$$\eta_i=\Big(\alpha_{n,i}(\mathbf{t}+2^{-l}\mathbf{e}_r)K_h(\mathbf{t}-\mathbf{z}_i+2^{-l}\mathbf{e}_r)-\alpha_{n,i}(\mathbf{t})K_h(\mathbf{t}-\mathbf{z}_i)\Big)\Lambda_k(\mathcal{P}_i)\varepsilon_i.$$

Нам потребуются следующие соотношения:

$$\begin{aligned}\Big|\alpha_{n,i}(\mathbf{t}+2^{-l}\mathbf{e}_r)K_h(\mathbf{t}-\mathbf{z}_i+2^{-l}\mathbf{e}_r)-\alpha_{n,i}(\mathbf{t})K_h(\mathbf{t}-\mathbf{z}_i)\Big|\Lambda_k(\mathcal{P}_i)\le&\\ \le\big|\alpha_{n,i}(\mathbf{t}+2^{-l}\mathbf{e}_r)\big|\Big|K_h(\mathbf{t}-\mathbf{z}_i+2^{-l}\mathbf{e}_r)-K_h(\mathbf{t}-\mathbf{z}_i)\Big|\Lambda_k(\mathcal{P}_i)+&\\ +K_h(\mathbf{t}-\mathbf{z}_i)\Big|\alpha_{n,i}(\mathbf{t}+2^{-l}\mathbf{e}_r)-\alpha_{n,i}(\mathbf{t})\Big|\Lambda_k(\mathcal{P}_i)\le&\\ \le\alpha L^k h^{-k-1}\;2^{-l}\;\delta_n^k+\overline{\alpha}h^{-1}2^{-l}L^k h^{-k}\delta_n^k=(\alpha+\overline{\alpha})L^k h^{-k-1}\;2^{-l}\;\delta_n^k,&\end{aligned}\tag{149}$$

$$\begin{aligned}\sum_{i=1}^{n}\Big|\alpha_{n,i}(\mathbf{t}+2^{-l}\mathbf{e}_r)K_h(\mathbf{t}-\mathbf{z}_i+2^{-l}\mathbf{e}_r)-\alpha_{n,i}(\mathbf{t})K_h(\mathbf{t}-\mathbf{z}_i)\Big|\Lambda_k(\mathcal{P}_i)\le&\\ \le(\alpha+\overline{\alpha})L^k h^{-k-1}\;2^{-l+1}\;(2h+2\delta_n)^k\le(\alpha+\overline{\alpha})L^k h^{-1}2^{-l+1}2^{2k}.&\end{aligned}\tag{150}$$

Отметим, что $|\mathbf{t}-\mathbf{z}_i+2^{-l}\mathbf{e}_r-\mathbf{t}+\mathbf{z}_i|=2^{-l}\le h$ при $l>m$. Кроме того, область суммирования в (150) совпадает с множеством

$$N_{n,h}(\mathbf{t}+2^{-l}\mathbf{e}_r)\cup N_{n,h}(\mathbf{t})\qquad\text{при}\quad\mathbf{t}\in\mathcal{R}_l.$$

Эти факты вместе с леммой 1.20 и соотношением $\delta_n\le h\le 1$ мы учли при выводе (150). Для доказательства (149) мы использовали второе утверждение леммы 1.20. Таким образом,

$$\sum_{i=1}^{n}\left(\mathbb{E}_{\mathcal{F}_n}|\eta_i|^p\right)^{2/p}\le M_p^{2/p}(\alpha+\overline{\alpha})^2 2^{2k}L^{2k}2^{-2l+1}\,h^{-2}\,(\delta_n/h)^k,$$

и в силу неравенства (146) получаем

$$\mathbb{P}_{\mathcal{F}_n}\Big(\big|\mu_{n,h}(\mathbf{t}+2^{-l}\mathbf{e}_r)-\mu_{n,h}(\mathbf{t})\big|>a_l y/k\Big)\le\frac{G_2}{k}\;\frac{(\delta_n/h)^{kp/2}\;h^{-p}\;2^{-lp}}{(a_l y)^p},\tag{151}$$

где

$$G_2=2^{p/2}(\alpha+\overline{\alpha})^p\,k^{p+1}\,(p-1)^{p/2}\,2^{kp}L^{kp}\,M_p=2^{p/2}\,k^{p+1}\,\alpha^{-p}(\alpha+\overline{\alpha})^pG_1.\tag{152}$$

Используя теперь (145), (147) и (151) заключаем, что

$$\mathbb{P}_{\mathcal{F}_n}\left(\sup_{\mathbf{t}\in\mathcal{P}}|\mu_{n,h}(\mathbf{t})|>y\right)<y^{-p}(\delta_n/h)^{kp/2}\left(G_1 2^{km}a_m^{-p}+G_2\,h^{-p}\sum_{l=m+1}^{\infty}2^{-(p-k)l}\,a_l^{-p}\right).$$

Оптимальная последовательность a_l, минимизирующая правую часть этого неравенства, есть $a_m=c(G_1 2^{km})^{1/(p+1)}$, $a_l=c\big(G_2\,h^{-p}\,2^{-(p-k)l}\big)^{1/(p+1)}$ при $l=m+1,m+2,\ldots$, где коэффициент c определяется соотношением $a_m+a_{m+1}+\cdots=1$. Для указанной последовательности получаем

$$\begin{aligned}
&\mathbb{P}_{\mathcal{F}_n}\left(\sup_{\mathbf{t}\in\mathcal{P}}|\mu_{n,h}(\mathbf{t})|>y\right)\le\\
&\qquad\le y^{-p}(\delta_n/h)^{kp/2}\left((G_1 2^{km})^{1/(p+1)}+\sum_{l=m+1}^{\infty}\Big(G_2\,h^{-p}2^{-(p-k)l}\Big)^{1/(p+1)}\right)^{p+1}\le\\
&\quad\le y^{-p}\,\delta_n^{kp/2}\,h^{-k(1+(p/2))}\left((2G_1)^{1/(p+1)}+\big(G_2 2^{-(p-k)}\big)^{1/(p+1)}\sum_{l=0}^{\infty}2^{-(p-k)l/(p+1)}\right)^{p+1}\le\\
&\qquad\le y^{-p}\,\delta_n^{kp/2}\,h^{-k(1+(p/2))}\Big((2G_1)^{1/(p+1)}+\big(G_2 2^{-(p-k)}\big)^{1/(p+1)}\big(1-2^{-(p-k)/(p+1)}\big)^{-1}\Big)^{p+1}.
\end{aligned}$$

Это соотношение вместе с оценкой (144) завершает вывод леммы 1.21 при

$$C_2^*=2^{p(k+1)}\Big((2G_1)^{1/(p+1)}+\big(G_2 2^{-(p-k)}\big)^{1/(p+1)}\big(1-2^{-(p-k)/(p+1)}\big)^{-1}\Big)^{p+1},\tag{153}$$

где константы G_1 и G_2 определены соответственно в (148) и (152). □

Д о к а з а т е л ь с т в о теоремы 1.6. Положим $\widetilde{\zeta}_n(h)=\sup_{\mathbf{t}\in[0,1]^k}\big|\widetilde{\nu}_{n,h}(\mathbf{t})\big|$, где величина $\widetilde{\nu}_{n,h}(\mathbf{t})$ определена в (142), и вместе с леммой 1.21 учтем соотношение

$$\mathbb{P}\big(\widetilde{\zeta}_n(h)>y,\,\delta_n\le c^*h\big)=\mathbb{E}I\big(\delta_n\le c^*h\big)\mathbb{P}_{\mathcal{F}_n}\big(\widetilde{\zeta}_n(h)>y\big).$$

Тождество (132) вместе с утверждениями лемм 1.18 и 1.19 завершают доказательство. □

1.3.6 Оценивание функции среднего случайного регрессионного поля

В этом подразделе в качестве приложения теоремы 1.6 мы рассмотрим один из вариантов задачи оценивания функции среднего почти наверное непрерывного случайного процесса. Другие постановки этой задачи мы рассмотрим далее в разделе 1.5.

Перейдем к постановке задачи. Пусть имеется N независимых копий модели (23) из условия (M_1):

$$X_{i,j} = f_j(\mathbf{z}_{i,j}) + \varepsilon_{i,j}, \qquad i = 1, \ldots, n, \quad j = 1, \ldots, N, \tag{154}$$

где $f_1(\mathbf{t}), \ldots, f_N(\mathbf{t})$, $\mathbf{t} \in [0,1]^k$, есть независимые неизвестные с вероятностью 1 непрерывные случайные процессы, наборы $\{\varepsilon_{i,j};\, i = 1, \ldots, n\}$ и $\{\mathbf{z}_{i,j};\, i = 1, \ldots, n\}$ при любом j удовлетворяет условиям (E_2) и (D_2). Здесь и далее в подразделе индекс j означает номер копии модели (23). В предположении, что количество наблюдений в каждой серии растет с ростом числа серий, определим оценку для функции среднего $\mathbb{E}f_1(\mathbf{t})$ равенством

$$\overline{f^*}_{N,n,h}(\mathbf{t}) = \frac{1}{N}\sum_{j=1}^{N} \widetilde{f}^*_{n,h,j}(\mathbf{t}),$$

где оценки $\widetilde{f}^*_{n,h,j}(\mathbf{t})$ задаются по формуле (113) при замене величин из (23) соответствующими характеристиками из (154).

В ситуации растущего количества наблюдений в каждой серии, естественно предварительно оценить случайную регрессионную функцию в каждой серии, а затем провести усреднение по всем сериям. Именно так мы и поступаем, выбирая в качестве оценки для функции среднего статистику $\overline{f^*}_{N,n,h}(\mathbf{t})$. В приводимой далее теореме 1.7 доказана равномерная состоятельность этой оценки для функции среднего $\mathbb{E}f_1(\mathbf{t})$, при этом от регрессоров требуется лишь выполнение условия (D_2) в каждой из независимых серий. Отметим, что предлагаемое условие на регрессоры является более слабым, чем известные ранее ограничения (см. подробности и библиографические ссылки во введении), и является универсальным относительно стохастической природы регрессоров.

Следствием теоремы 1.6 является следующее утверждение.

Теорема 1.7. *Пусть для модели* (154) *выполнены условия* (M_1), (K_2), (D_2) *и*

$$\mathbb{E} \sup_{\mathbf{t}\in[0,1]^k} |f_1(\mathbf{t})| < \infty. \tag{155}$$

Кроме того, последовательность $h \equiv h_n \to 0$ *и последовательность натуральных чисел* $N \equiv N_n \to \infty$ *удовлетворяют условиям*

$$h^{-k(p/2+1)}\,\mathbb{E}(\delta_n^{kp/2}) \to 0 \ \text{ и } \ N\mathbb{P}(\delta_n > c^*h\}) \to 0. \tag{156}$$

Тогда

$$\sup_{\mathbf{t}\in[0,1]^k} |\overline{f^*}_{N,n,h}(\mathbf{t}) - \mathbb{E}f_1(\mathbf{t})| \xrightarrow{p} 0.$$

З а м е ч а н и е 1.22. Если условие (155) заменить чуть более сильным ограничением

$\mathbb{E}\sup_{\mathbf{t}\in[0,1]^k} f_1^2(\mathbf{t}) < \infty$, то при выполнении (156) можно показать равномерную состоятельность оценки

$$M^*_{N,n,h}(\mathbf{t}_1,\mathbf{t}_2) = \frac{1}{N}\sum_{j=1}^{N} \widetilde{f}^*_{n,h,j}(\mathbf{t}_1)\widetilde{f}^*_{n,h,j}(\mathbf{t}_2), \quad \mathbf{t}_1,\mathbf{t}_2 \in [0,1]^k,$$

для смешанного второго момента $\mathbb{E}f_1(\mathbf{t}_1)f_1(\mathbf{t}_2)$, где $h \equiv h_n$ и $N \equiv N_n$ определены в (156). Доказательство этого факта аналогично выводу теоремы 1.7, поэтому подробные рассуждения мы опускаем. Другими словами, при выполнении вышеуказанных условий оценка

$$\mathbb{C}ov^*_{N,n,h}(\mathbf{t}_1,\mathbf{t}_2) = M^*_{N,n,h}(\mathbf{t}_1,\mathbf{t}_2) - \overline{f^*}_{N,n,h}(\mathbf{t}_1)\overline{f^*}_{N,n,h}(\mathbf{t}_2)$$

является равномерно состоятельной для ковариационной функции случайного регрессионного поля $f_1(\mathbf{t})$. □

1.3.7 Доказательство теоремы 1.7

Заметим, что условие (155) и теорема Лебега о мажорируемой сходимости влекут соотношение

$$\lim_{\nu\to 0} \mathbb{E}\omega_f(\nu) = 0. \tag{157}$$

Покажем, что сходимость (157) гарантирует выполнение следующего равномерного закона больших чисел:

$$\sup_{\mathbf{t}\in[0,1]^k} |\overline{f}_N(\mathbf{t}) - \mathbb{E}f_1(\mathbf{t})| \xrightarrow{p} 0 \quad \text{при} \quad N\to\infty, \qquad \text{где} \quad \overline{f}_N(\mathbf{t}) = N^{-1}\sum\nolimits_{j=1}^{N} f_j(\mathbf{t}). \tag{158}$$

Положим

$$\omega_{\overline{f}_N}(\varepsilon) = \sup_{\mathbf{t},\mathbf{s}:\|\mathbf{t}-\mathbf{s}\|\le\varepsilon} \big|\overline{f}_N(\mathbf{t}) - \overline{f}_N(\mathbf{s})\big|, \qquad \omega_{\mathbb{E}f_1}(\varepsilon) = \sup_{\mathbf{t},\mathbf{s}:\|\mathbf{t}-\mathbf{s}\|\le\varepsilon} \big|\mathbb{E}f_1(\mathbf{t}) - \mathbb{E}f_1(\mathbf{s})\big|.$$

Пусть для простоты $k = 2$. При произвольном фиксированном $r > 0$ и $u, v = 0,\ldots,r$, имеют место соотношения

$$\begin{aligned}
\sup_{\mathbf{t}\in[0,1]^2} \big|\overline{f}_N(\mathbf{t}) - \mathbb{E}f_1(\mathbf{t})\big| &\le \max_{0\le u,v\le r} \big|\overline{f}_N\big(u/r, v/r\big) - \mathbb{E}f_1\big(u/r,v/r\big)\big| + \\
&+ \max_{1\le u,v\le r} \sup_{\frac{u-1}{r}\le t_1\le\frac{u}{r},\ \frac{v-1}{r}\le t_2\le\frac{v}{r}} \big|\overline{f}_N(\mathbf{t}) - \overline{f}_N\big(u/r,v/r\big)\big| + \\
&\max_{1\le u,v\le r} \sup_{\frac{u-1}{r}\le t_1\le\frac{u}{r},\ \frac{v-1}{r}\le t_2\le\frac{v}{r}} \big|\mathbb{E}f_1(\mathbf{t}) - \mathbb{E}f_1\big(u/r,v/r\big)\big| \le \\
&\le \max_{0\le u,v\le r} \big|\overline{f}_N\big(u/r,v/r\big) - \mathbb{E}f_1\big(u/r,v/r\big)\big| + \omega_{\overline{f}_N}(1/r) + \omega_{\mathbb{E}f_1}(1/r)\,.
\end{aligned} \tag{159}$$

Заметим теперь, что $\omega_{\mathbb{E}f_1}(\varepsilon) \le \mathbb{E}\omega_{f_1}(\varepsilon)$ и

$$\overline{f}_N\big(u/r, v/r\big) \xrightarrow{p} \mathbb{E}f_1\big(u/r, v/r\big), \quad \omega_{\overline{f}_N}(\varepsilon) \le \frac{1}{N}\sum_{i=1}^{N}\omega_{f_i}(\varepsilon) \xrightarrow{p} \mathbb{E}\omega_{f_1}(\varepsilon).$$

Следовательно, правая часть в (159) не превосходит $2\mathbb{E}\omega_{f_1}(1/r)+o_p(1)$ и в силу произвольности r и (157) соотношение (158) при $k=2$ доказано. При произвольном k вывод аналогичен. Введем теперь обозначение

$$\Delta_{n,h,j} = \sup_{\mathbf{t}\in[0,1]^k}\big|\widetilde{f}^*_{n,h,j}(\mathbf{t}) - f_j(\mathbf{t})\big| \tag{160}$$

и установим следующую версию закона больших чисел для величин $\{\Delta_{n,h,j}\}$:

$$\frac{1}{N}\sum_{j=1}^{N}\Delta_{n,h,j} \xrightarrow{p} 0 \qquad \text{при} \quad N\to\infty, \tag{161}$$

где последовательности $h=h_n$ и $N=N_n$ заданы в (156). Введем в рассмотрение события $A_{n,h,j}=\{\delta_{n,j}\le c^*h\}$, $j=1,\ldots,N$, и заметим, что для любого положительного ν выполнено

$$\mathbb{P}\Big\{\frac{1}{N}\sum_{j=1}^{N}\Delta_{n,h,j} > \nu\Big\} \le \mathbb{P}\Big\{\frac{1}{N}\sum_{j=1}^{N}\Delta_{n,h,j}I(A_{n,h,j}) > \nu\Big\} + N\mathbb{P}(\overline{A_{n,h,1}}). \tag{162}$$

Далее, из теоремы 1.6 следует, что

$$\mathbb{E}\Delta_{n,h,j}I(A_{n,h,j}) \le C_1^*\mathbb{E}\omega_f(h) + \int\limits_0^\infty \mathbb{P}\left(\zeta_n(h)>y,\, \delta_n\le c^*h\right)dy \le$$

$$\le C_1^*\mathbb{E}\omega_f(h) + \gamma_n + \int\limits_{\gamma_n}^\infty \mathbb{P}\left(\zeta_n(h)>y,\, \delta_n\le c^*h\right)dy \le C_1^*\mathbb{E}\omega_f(h) + (1+C_2^*)\gamma_n,$$

где $\gamma_n = \Big(h^{-k(p/2+1)}\,\mathbb{E}\delta_n^{kp/2}\Big)^{1/p}$. Для завершения доказательства (161) остается с помощью неравенство Маркова оценить первое слагаемое в правой части (162), учесть предельные соотношения из (156) и последнюю оценку.

Утверждение теоремы 1.7 следует из очевидной оценки

$$\sup_{\mathbf{t}\in[0,1]^k}\left|N^{-1}\sum_{j=1}^{N}f^*_{n,h,j}(\mathbf{t}) - \mathbb{E}f_1(\mathbf{t})\right| \le \sup_{\mathbf{t}\in[0,1]^k}\big|\overline{f}_N(\mathbf{t}) - \mathbb{E}f_1(\mathbf{t})\big| + N^{-1}\sum_{j=1}^{N}\Delta_{n,h,j}.$$

и предельных соотношений (158) и (161). Теорема 1.7 доказана. $\square$

1.4 Оценки Надарая–Ватсона и классические локально–линейные оценки

В этом разделе доказана состоятельность двух классических вариантов ядерных оценок: Надарая–Ватсона и локально–линейных. Получены новые ограничения на регрессоры (более слабые, чем были известны ранее), гарантирующие как поточечную, так и равномерную состоятельность этих оценок. В отличие от известных ранее результатов, новые условия универсальны относительно стохастической природы регрессоров и сформулированы в терминах плотных данных, а именно — асимптотического поведения числа регрессоров, попавших в ту или иную окрестность точек из области задания регрессионной функции. В новых условиях, гарантирующих равномерную состоятельность оценок Надарая–Ватсона или локально–линейных оценок, требуется, по сути, более равномерное плотное заполнение регрессорами области определения регрессионной функции, чем в условиях (D_1) или (D_2), гарантирующих равномерную состоятельность новых классов оценок, предложенных в разделах 1.1–1.3.

1.4.1 Оценки Надарая–Ватсона

Всюду в этом разделе, не оговаривая это особо, считаем, что в регрессионной модели, определяемой условием (M_1), функция f неслучайна. Условимся, что в качестве векторной нормы $\|\cdot\|$ рассматривается супремальная норма в $\mathbb{R}^k$. Нам потребуется следующее ограничение на ядро сглаживания.

$(\mathbf{K}_4)$ *Выполнено условие (K), ядерная функция $K(\mathbf{t})$ удовлетворяет соотношению*

$$\sup_{\mathbf{t}\in[0,1]^k} K(\mathbf{t}) \le L < \infty$$

и существует положительное $\delta \le 1$ такое, что

$$\inf_{\|\mathbf{t}\|\le\delta} K(\mathbf{t}) \ge l > 0.$$

Введем в рассмотрение оценку Надарая–Ватсона

$$\widehat{f}_{NW}(\mathbf{t}) = \frac{\sum_{i=1}^{n} X_i K_h(\mathbf{t}-\mathbf{z}_i)}{\sum_{i=1}^{n} K_h(\mathbf{t}-\mathbf{z}_i)} \tag{163}$$

и для любых $\mathbf{z}_1,\ldots,\mathbf{z}_n$, $h\in(0,1)$ и $\mathbf{t}\in[0,1]^k$ положим

$$N_{n,h}(\mathbf{t}) = \{i:\ \|\mathbf{t}-\mathbf{z}_i\|\le h,\ 1\le i\le n\}. \tag{164}$$

Символом $\#(\cdot)$ обозначим стандартную считающую меру.

Рассмотрим прежде всего вопрос о поточечной состоятельности оценок $\widehat{f}_{NW}(\mathbf{t})$.

Теорема 1.8. *Пусть выполнено условие* (M_1) *при* $\mathcal{P} = [0,1]^k$ *и условия* (E_1) *и* (K_4). *Тогда для любого* $h \in (0,1)$ *и любого* $\mathbf{t} \in [0,1]^k$ *с вероятностью* 1 *выполнено*

$$\mathbb{E}_{\mathcal{F}_n}(\widehat{f}_{NW}(\mathbf{t}) - f(\mathbf{t}))^2 \le \omega_f^2(h) + \frac{\sigma^2 L}{l\#(N_{n,\delta h}(\mathbf{t}))}.$$

Следствие 1.7. *В условиях теоремы* 1.8 *имеет место соотношение*

$$\widehat{f}_{NW}(\mathbf{t}) - f(\mathbf{t}) = O_p\left(\omega_f(h) + \frac{\sigma\sqrt{L}}{\sqrt{l\#(N_{n,\delta h}(\mathbf{t}))}}\right). \tag{165}$$

Введем еще одно предположение.

$(\mathbf{D}_3)$ *При всех фиксированных* $\mathbf{t} \in [0,1]^k$ *и* $h \in (0,1)$ *имеет место предельное соотношение* $\#(N_{n,h}(\mathbf{t})) \xrightarrow{p} \infty$.

Следствие 1.8. *Пусть выполнены условия* (M_1) *при* $\mathcal{P} = [0,1]^k$, (E_1), (K_4) *и* (D_3). *Тогда существует последовательность* $h = h_n \to 0$, *для которой при всех* $\mathbf{t} \in [0,1]^k$ *имеет место предельное соотношение*

$$\widehat{f}_{NW}(\mathbf{t}) \xrightarrow{p} f(\mathbf{t}).$$

Таким образом, условие (D_3) — это единственное условие на регрессоры $\{\mathbf{z}_i\}$, обеспечивающее поточечную состоятельность оценок Надарая–Ватсона. Понятно, что условие (D_3) выполнено для неслучайных регулярных регрессоров. Если $\{\mathbf{z}_i\}$ независимы и одинаково распределены и множество $[0,1]^k$ является носителем распределения $\mathbf{z}_1$, то условие (D_3) также выполнено. Например, если в случае независимых и одинаково распределенных $\{\mathbf{z}_i\}$ распределение $\mathbf{z}_1$ имеет отделенную от нуля и ограниченную на $[0,1]^k$ плотность, то по теореме Гливенко–Кантелли $\#(N_{n,h}(\mathbf{t})) \sim nh^k$ почти наверное равномерно по $\mathbf{t}$.[3] Следовательно, если функция f удовлетворяет условию Липшица, то можно положить $h = n^{-1/(2+k)}$. В этом случае оценка поточечной сходимости из следствия 1.7 будет иметь порядок $\widetilde{O}_p(n^{-1/(2+k)})$. Отметим, что таким образом выбранный размер окна h фактически уравнивает порядок малости по h обоих слагаемых в правой части соотношения (165). Если $\{\mathbf{z}_i;\ i \ge 1\}$ — стационарная последовательность с условием α-перемешивания и маргинальным распределением с носителем $[0,1]^k$, то условие (D_3) также выполнено. Отметим, что зависимость случайных величин $\{\mathbf{z}_i\}$, удовлетворяющих условию (D_3), может быть более сильной, чем классические условия слабой зависимости (см. пример 1.16, а также замечание 1.23).

[3] Данная запись означает, что с вероятностью 1 выполнено $c_1 h^k \le \limsup_{\mathbf{t}\in[0,1]^k} \#(N_{n,h}(\mathbf{t}))/n \le c_2 h^k$, где константы c_1 и c_2 не зависят от t и h.

З а м е ч а н и е 1.23. Нетрудно видеть, что условие (D_3) эквивалентно предположению (D_2), а при $k = 1$ эквивалентно условию (D_1). Тем самым, справедливы все замечания, сделанные ранее относительно указанных условий плотного заполнения регрессорами области определения регрессионной функции. Подчеркнем, что условия (D_1) или (D_2) гарантируют существование *равномерно* состоятельных оценок (в новых классах универсальных локально–постоянных или локально линейных оценок), а условие (D_3) для оценок Надарая–Ватсона гарантирует лишь *поточечную* состоятельность. Кроме того, новые ядерные оценки могут быть в известном смысле точнее оценок Надарая–Ватсона даже в классической ситуации независимых и одинаково распределенных регрессоров (см. раздел 1.1.5, а также результаты компьютерного моделирования в разделах 1.2.3 и 1.3.3).

П р и м е р 1.16. Рассмотрим следующее обобщение примеров 1.1 и 1.2 на случай равномерно распределенных в единичном квадрате $[0, 1]^2$ случайных величин. Последовательность двумерных случайных величин $\{\mathbf{z}_i;\ i \geq 1\}$ определяется соотношением

$$\mathbf{z}_i = \nu_i \mathbf{u}_i^l + (1 - \nu_i)\mathbf{u}_i^r, \tag{166}$$

где $\{\mathbf{u}_i^l\}$ и $\{\mathbf{u}_i^r\}$ независимы и равномерно распределены на $[0, 1/2] \times [0, 1]$ и $[1/2, 1] \times [0, 1]$ соответственно, последовательность $\{\nu_i\}$ не зависит от $\{\mathbf{u}_i^l\}$, $\{\mathbf{u}_i^r\}$ и состоит из бернуллиевских случайных величин с вероятностью успеха $1/2$, т.е. распределение случайных величин $\mathbf{z}_i$ представляет собой равновесную смесь двух равномерных распределений в соответствующих прямоугольниках. Зависимость между случайными величинами ν_i для любого натурального i определим равенствами $\nu_{2i-1} = \nu_1$, $\nu_{2i} = 1 - \nu_1$. В этом случае случайные величины $\{\mathbf{z}_i;\ i \geq 1\}$ из (166) образуют стационарную последовательность случайных величин, равномерно распределенных на $[0, 1]^2$, но, скажем, все известные условия перемешивания здесь не выполнены, поскольку при всех натуральных m и n

$$\mathbb{P}\big(\mathbf{z}_{2m} \in [0, 1/2] \times [0, 1],\ \mathbf{z}_{2n-1} \in [0, 1/2] \times [0, 1]\big) = 0$$

(см. также подробности в примере 1.1 в случае одномерных случайных величин). С другой стороны, нетрудно проверить, что стационарная последовательность $\{\mathbf{z}_i\}$ удовлетворяет теореме Гливенко–Кантелли. Это означает, что для любого фиксированного $h > 0$ выполнено соотношение $\#(N_{n,h}(\mathbf{t})) \sim 4h^2 n$ почти наверное равномерно по t. Другими словами, последовательность $\{\mathbf{z}_i\}$ удовлетворяет условию (D_3).

В общем случае, как мы уже отмечали в примере 1.2, по схеме этого примера можно конструировать различные последовательности зависимых равномерно распределенных на $[0, 1]^2$ случайных величин исходя из выбора различных последовательностей бернуллиевских переключателей с условиями $\nu_{j_k} = 1$ и $\nu_{l_k} = 0$ для неограниченного числа индексов $\{j_k\}$ и $\{l_k\}$ соответственно. В этом случае условие (D_3) также будет выполнено. Но соответствующая последовательность $\{\mathbf{z}_i\}$ (не обязательно стационарная) может не удовлетворять даже усиленному закону больших чисел. Например, подобная ситуация имеет место в случае, когда $\nu_j = 1 - \nu_1$ при $j = 2^{2k-1}, \dots, 2^{2k} - 1$ и $\nu_j = \nu_1$ при $j = 2^{2k}, \dots, 2^{2k+1} - 1$, где $k = 1, 2, \dots$ (т.е. мы разыгрываем, в какой из прямоугольников бросаем наудачу первую точку, и далее чередуем количество бросаний на каждый из двух указанных прямоугольников следующим образом: 2, 4, 8, 16 и т.д.). В самом деле, введем обозначения $n_k = 2^{2k} - 1$, $\tilde{n}_k = 2^{2k+1} - 1$, $S_m = \sum_{i=1}^m z_{1i}$ при $\mathbf{z}_i = (z_{1i}, z_{2i})$ и заметим, что для всех элементарных исходов, составляющих событие $\{\nu_1 = 1\}$, выполнено

$$\frac{S_{n_k}}{n_k} = \frac{1}{n_k} \sum_{i \in A_{1,k}} u_{1i}^l + \frac{1}{n_k} \sum_{i \in A_{2,k}} u_{1i}^r, \tag{167}$$

где $\mathbf{u}_i^l = (u_{1i}^l, u_{2i}^l)$, $\mathbf{u}_i^r = (u_{1i}^r, u_{2i}^r)$, а $A_{1,k}$ и $A_{2,k}$ — наборы индексов, для которых наблюдения $\{\mathbf{z}_i;\ i \le n_k\}$ расположены в прямоугольнике $[0, 1/2] \times [0, 1]$ или $[1/2, 1] \times [0, 1]$ соответственно. Нетрудно видеть, что $\#(A_{1,k}) = n_k/3$, $\#(A_{2,k}) = 2\#(A_{1,k})$. Следовательно, $S_{n_k}/n_k \to 7/12$ почти наверное при $k \to \infty$ ввиду усиленного закона больших чисел для последовательностей $\{u_{1i}^l\}$ и $\{u_{1i}^r\}$. С другой стороны, для всех элементарных исходов, принадлежащих событию $\{\nu_1 = 1\}$, при $k \to \infty$ с вероятностью 1 выполнено

$$\frac{S_{\tilde{n}_k}}{\tilde{n}_k} = \frac{1}{\tilde{n}_k} \sum_{i \in \tilde{A}_{1,k}} u_{1i}^l + \frac{1}{\tilde{n}_k} \sum_{i \in \tilde{A}_{2,k}} u_{1i}^r \to \frac{5}{12}, \tag{168}$$

где $\tilde{A}_{1,k}$ и $\tilde{A}_{2,k}$ — наборы индексов, для которых наблюдения $\{\mathbf{z}_i;\ i \le \tilde{n}_k\}$ расположены в прямоугольнике $[0, 1/2] \times [0, 1]$ или $[1/2, 1] \times [0, 1]$ соответственно. При получении сходимости в (168) мы учли, что $\#(\tilde{A}_{1,k}) = (2^{2k+2} - 1)/3$, $\#(\tilde{A}_{2,k}) = 2n_k/3$, т.е. $\#(\tilde{A}_{1,k}) = 2\#(\tilde{A}_{2,k}) + 1$.

Схожие выводы справедливы и для всех элементарных исходов, составляющих событие $\{\nu_1 = 0\}$. $\square$

Рассмотрим теперь вопрос о равномерной состоятельности оценок Надарая–Ватсона. Нам потребуются дополнительные предположения на ядро и регрессоры.

$(\mathbf{K}_5)$ *Выполнено условие (K_4), ядерная функция $K(\mathbf{t})$ представима в виде $K(\mathbf{t}) = \prod_{j=1}^k K_o(t_j)$, где $K_o(\cdot)$ — плотность некоторого симметричного распределения с носителем $[-1, 1]$, при этом функция $K_o(t)$ непрерывно дифференцируема на открытом интервале $(0, 1)$ и*

$$\sup_{t \in [0,1]} \left(K_o(t) + |K_o'(t)|\right) \le L^{1/k} < \infty.$$

($\mathbf{D}_4$) *Для любых* $h \in (0,1)$ *имеет место предельное соотношение*

$$\Delta_{n,h} = \sup_{\mathbf{t}\in[0,1]^k} \frac{\sup_{\|\mathbf{s}\|\le h} \#^3(N_{n,h}(\mathbf{t}+\mathbf{s}))}{\#^4(N_{n,\delta h}(\mathbf{t}))} \xrightarrow{p} 0.$$

З а м е ч а н и е 1.24. В условии (D_4) по умолчанию считаем, что под внутренним знаком супремума область изменения переменной $\mathbf{s}$ при каждом фиксированном $\mathbf{t} \in [0,1]^k$ определяется соотношениями $\mathbf{s} \in [0,1]^k$ и $\mathbf{s}+\mathbf{t} \in [0,1]^k$. Отметим, что условие ($K_5$) не требует гладкости в нуле, так что плотности типа треугольных удовлетворяют этому условию.

Приведем теперь основной результат подраздела.

Теорема 1.9. *Пусть выполнены условия* (M_1) *при* $\mathcal{P} = [0,1]^k$, (E_2) *и* (K_5). *Тогда для любого фиксированного* $h \in (0,1)$ *с вероятностью* 1 *справедлива оценка*

$$\sup_{\mathbf{t}\in[0,1]^k} |\widehat{f}_{NW}(\mathbf{t}) - f(\mathbf{t})| \le \omega_f(h) + \zeta_n(h),$$

где последовательность случайных величин $\zeta_n(h) > 0$ *удовлетворяет соотношению*

$$\zeta_n(h) = O_p\left(G\sqrt{\Delta_{n,h}/h^{2k/p}}\right)$$

при $0 < G = G(M_p, K, k, p) \le C^{1/p}(k,p)M_p^{1/p}(p-1)^{1/2}(L/l)^2$ *и*

$$C(k,p) = \left(2^{k/(p+1)} + \frac{k2^{(p+2k)/2(p+1)}}{1-2^{-(p-k)/(p+1)}}\right)^{p+1}.$$

Следствие 1.9. *Пусть выполнены условия* (M_1) *при* $\mathcal{P} = [0,1]^k$, (E_2), (K_5) *и* (D_4). *Тогда существует последовательность* $h = h_n \to 0$, *для которой имеет место предельное соотношение*

$$\sup_{\mathbf{t}\in[0,1]^k} |\widehat{f}_{NW}(\mathbf{t}) - f(\mathbf{t})| \xrightarrow{p} 0.$$

З а м е ч а н и е 1.25. Нетрудно видеть, что выполнение условия (D_4) гарантирует следующее чуть более сильное ограничение:

$$\Delta'_{n,h} = \frac{\sup_{\mathbf{t}\in[0,1]^k} \#^3(N_{n,h}(\mathbf{t}))}{\inf_{\mathbf{t}\in[0,1]^k} \#^4(N_{n,\delta h}(\mathbf{t}))} \xrightarrow{p} 0.$$

Как отмечалось выше, в случае независимых и одинаково распределенных регрессоров с распределением $\mathbf{z}_1$, имеющим отделенную от нуля и ограниченную на $[0,1]^k$ плотность, по теореме Гливенко–Кантелли $\#(N_{n,h}(\mathbf{t})) \sim nh^k$ почти наверное равномерно по $\mathbf{t}$. Аналогичная оценка справедлива и в случае, когда $\{\mathbf{z}_i\}$ удовлетворяют условию φ-перемешивания (при широких условиях на коэффициенты перемешивания), а также в случае сильно зависимых $\{\mathbf{z}_i\}$ из примера 1.16. Тем самым, для таких наблюдений $\{\mathbf{z}_i\}$ с вероятностью 1 выполнено $\Delta'_{n,h} \sim (nh^k)^{-1}$. Понятно, что указанная оценка для $\Delta'_{n,h}$ в широких условиях справедлива также для различных вариантов неслучайных регрессоров регулярного типа (например, вида $\mathbf{z}_i = \mathbf{g}(j_{i1}/n^{1/k}, \ldots, j_{ik}/n^{1/k})$ с гладкой функцией $\mathbf{g} : \mathbb{R}^k \to \mathbb{R}^k$).

П р и м е р 1.17. Для функции f, удовлетворяющей условию Липшица, в случае, когда с вероятностью 1 справедливо соотношение $\Delta'_{n,h} = O\big((nh^k)^{-1}\big)$, можно положить $h = n^{-p/(2p+kp+2k)}$. В этом случае

$$\sup_{\mathbf{t}\in[0,1]^k} |\widehat{f}_{NW}(\mathbf{t}) - f(\mathbf{t})| = \widetilde{O}_p(n^{-p/(2p+kp+2k)}).$$

В частности, при $k = 1$ и $p = 2$ получаем, что $h = n^{-1/4}$, а вышеуказанный супремум есть $\widetilde{O}_p(n^{-1/4})$.

В следующем примере построена последовательность сильно зависимых регрессоров, для которой предположение (D_3) (как и эквивалентное условие (D_2) из раздела 1.3, гарантирующее *равномерную* состоятельность новых универсальных локально–постоянных и локально–линейных оценок) выполнено, а предположение (D_4) — нет.

П р и м е р 1.18. Пусть для последовательности $\{\mathbf{z}_i\}$ из примера 1.16, определенной в (166), зависимость между случайными величинами ν_i для любого натурального i задается равенствами $\nu_{[e^k]} = \nu_1$ при $k \geq 1$ и $\nu_i = 1 - \nu_1$ при всех $i \neq [e^k]$, $k \geq 1$ (т.е. мы разыгрываем в какой из прямоугольников бросаем наудачу первую точку, а далее чередуем количество бросаний таким образом, чтобы с вероятностями 1/2 в один из прямоугольников попадало $\log n$ точек, а в другой — остальные). В этом случае нестационарная последовательность $\{\mathbf{z}_i\}$ не удовлетворяет усиленному закону больших чисел, так как нормированная сумма первой координаты элементов $\{\mathbf{z}_i\}$ почти наверное сходится к некоторой случайной величине, имеющей двухточечное распределение (1/4 и 3/4 с равными вероятностями). Но условие (D_3) для этой последовательности выполнено. Покажем, что для рассматриваемой последовательности сильно зависимых случайных величин не выполнено условие (D_4). Действительно, при положительном фиксированном достаточно малом h на почти всех элементарных исходах из множества $\{\nu_1 = 1\}$ мы имеем $\#(N_{n,h}(1/2, 0)) \sim 2h^2 n$, $\#(N_{n,\delta h}(1/2 - h, 0)) \sim 4h^2\delta^2 \log n$. Таким образом, на множестве указанных элементарных исходов, имеющем вероятность 1/2, выполнено

$$\Delta_{n,h} \geq \frac{\#^3(N_{n,h}(1/2, 0))}{\#^4(N_{n,\delta h}(1/2 - h, 0))} \sim \frac{n^3}{2^5\delta^8 h^2 \log^4 n} \to \infty.$$

$\square$

Сравнивая условия (D_3) и (D_4) отметим, что несколько более громоздкое условие (D_4) по сути предполагает более равномерное плотное заполнение регрессорами области определения регрессионной функции, нежели требуется в условии (D_3) (а также (D_1) и (D_2)).

Отметим, что в случае одномерных регрессоров утверждения теоремы 1.9 и следствия 1.9 можно получить при замене условия (E_2) на (E_1). Справедлива

Теорема 1.10. *Пусть $k = 1$ и выполнены условия (M_1) при $\mathcal{P} = [0,1]$, (E_1) и (K_5). Тогда для любого фиксированного $h \in (0,1)$ с вероятностью 1 выполнено*

$$\sup_{t\in[0,1]} |\widehat{f}_{NW}(t) - f(t)| \le \omega_f(h) + \zeta_n(h),$$

где $\zeta_n(h)$ есть последовательность положительных случайных величин, удовлетворяющих соотношению $\zeta_n(h) = O_p\left(\sigma(L/l)^2\sqrt{\Delta_{n,h}h^{-1}}\right)$.

Следствие 1.10. *Если в условиях теоремы 1.10 выполнено условие (D_4), то существует последовательность $h = h_n \to 0$, для которой имеет место предельное соотношение*

$$\sup_{\mathbf{t}\in[0,1]^k} |\widehat{f}_{NW}(\mathbf{t}) - f(\mathbf{t})| \xrightarrow{p} 0.$$

1.4.2 Доказательства утверждений раздела 1.4.1

Всюду в этом подразделе считаем выполненными условия (M_1) и (K_4). Учитывая соотношения (23) и (163), получаем тождество

$$\widehat{f}_{NW}(\mathbf{t}) = I_{n,h}(f,\mathbf{t}) + \nu_{n,h}(\mathbf{t}), \tag{169}$$

где

$$I_{n,h}(f,\mathbf{t}) = J_{n,h}^{-1}(\mathbf{t})\sum_{i=1}^{n} f(\mathbf{z}_i)K_h\left(\mathbf{t}-\mathbf{z}_i\right),$$
$$J_{n,h}(\mathbf{t}) = \sum_{i=1}^{n} K_h(\mathbf{t}-\mathbf{z}_i),$$
$$\nu_{n,h}(\mathbf{t}) = J_{n,h}^{-1}(\mathbf{t})\sum_{i=1}^{n} K_h\left(\mathbf{t}-\mathbf{z}_i\right)\varepsilon_i.$$

Подчеркнем, что ввиду свойств плотности $K_h(\cdot)$ область суммирования в этих трех суммах совпадает с множеством $N_{n,h}(\mathbf{t}) = \{i:\ \|\mathbf{t}-\mathbf{z}_i\| \le h, 1 \le i \le n\}$. Этот факт мы неоднократно будем использовать далее.

Лемма 1.22. *Имеет место оценка*

$$\sup_{\mathbf{t}\in[0,1]^k} |I_{n,h}(f,\mathbf{t}) - f(\mathbf{t})| \le \omega_f(h).$$

Д о к а з а т е л ь с т в о этой леммы непосредственно вытекает из равенства

$$I_{n,h}(f,\mathbf{t}) = f(\mathbf{t}) + J_{n,h}^{-1}(\mathbf{t}) \sum_{i\in N_{n,h}(\mathbf{t})} (f(\mathbf{z}_i) - f(\mathbf{t}))K_h\left(\mathbf{t} - \mathbf{z}_i\right).$$

Лемма доказана. □

Лемма 1.23. *Если выполнено условие* (E_1), *то для любого* $\mathbf{t} \in [0,1]^k$ *с вероятностью* 1

$$\mathbb{E}_{\mathcal{F}_n}\nu_{n,h}^2(\mathbf{t}) \le \frac{\sigma^2 L}{l\#(N_{n,\delta h}(\mathbf{t}))}.$$

Д о к а з а т е л ь с т в о. Это утверждение влечет следующая цепочка оценок, справедливых с вероятностью 1:

$$\mathbb{E}_{\mathcal{F}_n}\nu_{n,h}^2(\mathbf{t}) \le \sigma^2 J_{n,h}^{-2}(\mathbf{t}) \sum_{i=1}^{n} K_h^2\left(\mathbf{t} - \mathbf{z}_i\right) \le \sigma^2 h^{-k} L J_{n,h}^{-1}(\mathbf{t}) \le \frac{\sigma^2 L}{l\#(N_{n,\delta h}(\mathbf{t}))}.$$

Лемма доказана. □

Д о к а з а т е л ь с т в о теоремы 1.8. Нетрудно видеть, что с учетом (E_1) с вероятностью 1 имеет место равенство

$$\mathbb{E}_{\mathcal{F}_n}(\widehat{f}_{NW}(\mathbf{t}) - f(\mathbf{t}))^2 = (I_{n,h}(f,\mathbf{t}) - f(\mathbf{t}))^2 + \mathbb{E}_{\mathcal{F}_n}\nu_{n,h}^2(\mathbf{t}). \tag{170}$$

Для завершения доказательства теоремы остается оценить каждое из слагаемых в правой части этого соотношения, используя леммы 1.22 и 1.23. □

Д о к а з а т е л ь с т в о следствия 1.7. Положим

$$\zeta_n = \widehat{f}_{NW}(\mathbf{t}) - f(\mathbf{t}), \qquad \eta_n = \omega_f(h) + \sigma\sqrt{L}/\sqrt{l\#(N_{n,\delta h}(\mathbf{t}))}.$$

Поскольку по теореме 1.8 выполнено $\mathbb{E}_{\mathcal{F}_n}\zeta_n^2 \le \eta_n^2$, то

$$\mathbb{P}\big(|\zeta_n|/\eta_n > M\big) = \mathbb{E}\mathbb{P}_{\mathcal{F}_n}\big(|\zeta_n|/\eta_n > M\big) \le \mathbb{E}\frac{\mathbb{E}_{\mathcal{F}_n}\zeta_n^2}{M^2\eta_n^2} \le \frac{1}{M^2},$$

что доказывает утверждение. □

Д о к а з а т е л ь с т в о следствия 1.8. Нетрудно видеть, что при выполнении условия (D_3) для любого $\mathbf{t}$ существует последовательность $h(\mathbf{t}) \equiv h_n(\mathbf{t}) \to 0$, удовлетворяющая предельному соотношению условия (D_3). Диагональным методом можно выбрать универсальную последовательность $h_n \to 0$, удовлетворяющую предельному соотношению условия (D_3) для всех рациональных точек $\mathbf{t} \in [0,1]^k$. Поскольку и оценка $\widehat{f}_{NW}(\mathbf{t})$ и регрессионная функция $f(\mathbf{t})$ непрерывны, то ввиду следствия 1.7 эта последовательность h_n обеспечивает поточечную состоятельность оценок (163) при всех $\mathbf{t} \in [0,1]^k$. □

Для вывода теоремы 1.9 нам потребуется еще одно вспомогательное утверждение.

Лемма 1.24. *В условиях теоремы* 1.9 *для любых* $y > 0$ *и* $h \in (0,1)$ *справедлива оценка*

$$\mathbb{P}_{\mathcal{F}_n}\left(\sup_{\mathbf{t}\in[0,1]^k} |\nu_{n,h}(\mathbf{t})| > y\right) \leq \frac{C(k,p)M_p h^{-k}}{y^p}\left(\frac{(p-1)L^4\Delta_{n,h}}{l^4}\right)^{p/2}.$$

Д о к а з а т е л ь с т в о. Хвост распределения величины $\sup\limits_{\mathbf{t}\in[0,1]^k} |\nu_{n,h}(\mathbf{t})|$ оценим с помощью метода диадических цепочек (см. [43]), который мы неоднократно использовали выше. Как и ранее, прежде всего заменим множество $[0,1]^k$ под знаком супремума множеством двоично-рациональных точек $\mathcal{R} = \cup_{l\geq 1}\mathcal{R}_s$, где

$$\mathcal{R}_s = \{(j_1/2^s, \ldots, j_k/2^s):\ j_1 = 1, \ldots, 2^s - 1; \ldots; j_k = 1, \ldots, 2^s - 1\}.$$

Таким образом,

$$\sup_{\mathbf{t}\in[0,1]^k} |\nu_{n,h}(\mathbf{t})| = \sup_{\mathbf{t}\in\mathcal{R}} |\nu_{n,h}(\mathbf{t})| \leq \max_{\mathbf{t}\in\mathcal{R}_m} |\nu_{n,h}(\mathbf{t})| + \sum_{s=m+1}^{\infty}\sum_{r=1}^{k}\max_{\mathbf{t}\in\mathcal{R}_s}\big|\nu_{n,h}(\mathbf{t}+2^{-s}\mathbf{e}_r) - \nu_{n,h}(\mathbf{t})\big|,$$

где натуральное m будет выбрано далее, $\mathbf{e}_r$ есть k-мерный вектор с единичной r-ой компонентой и нулями в качестве остальных компонент. Следовательно,

$$\begin{aligned}\mathbb{P}_{\mathcal{F}_n}(\sup_{\mathbf{t}\in[0,1]^k} |\nu_{n,h}(\mathbf{t})| > y) \leq \mathbb{P}_{\mathcal{F}_n}(\max_{\mathbf{t}\in\mathcal{R}_m} |\nu_{n,h}(\mathbf{t})| > a_m y) + \\ + \sum_{s=m+1}^{\infty}\sum_{r=1}^{k}\mathbb{P}_{\mathcal{F}_n}\Big(\max_{\mathbf{t}\in\mathcal{R}_s}\big|\nu_{n,h}(\mathbf{t}+2^{-s}\mathbf{e}_r) - \nu_{n,h}(\mathbf{t})\big| > a_s y/k\Big) \leq \\ \leq \sum_{\mathbf{t}\in\mathcal{R}_m}\mathbb{P}_{\mathcal{F}_n}(|\nu_{n,h}(\mathbf{t})| > a_m y) + \sum_{s=m+1}^{\infty}\sum_{r=1}^{k}\sum_{\mathbf{t}\in\mathcal{R}_s}\mathbb{P}_{\mathcal{F}_n}\Big(\big|\nu_{n,h}(\mathbf{t}+2^{-s}\mathbf{e}_r) - \nu_{n,h}(\mathbf{t})\big| > a_s y/k\Big),\end{aligned} \quad (171)$$

где $a_m, a_{m+1}, \ldots$ есть последовательность положительных чисел, удовлетворяющих условию $a_m + a_{m+1} + \cdots = 1$.

Чтобы оценить вероятности в правой части соотношения (171), воспользуемся следующим мартингальным неравенством (см. [241]):

$$\mathbb{E}\Big|\sum_{i=1}^{n}\eta_i\Big|^p \leq \left((p-1)\sum_{i=1}^{n}\big(\mathbb{E}|\eta_i|^p\big)^{2/p}\right)^{p/2}, \quad (172)$$

где $\{\eta_i\}$ — последовательность мартингал-разностей с конечными моментами порядка $p \geq 2$.

Используя неравенство (172) при $\eta_i = J_{n,h}^{-1} K_h(\mathbf{t} - \mathbf{z}_i)\varepsilon_i$, с вероятностью 1 имеем

$$\mathbb{P}_{\mathcal{F}_n}\left(|\nu_{n,h}(\mathbf{t})| > a_m y\right) \le \frac{\mathbb{E}_{\mathcal{F}_n}|\nu_{n,h}(\mathbf{t})|^p}{(a_m y)^p} \le \frac{M_p}{(a_m y)^p}\left(\frac{(p-1)\sum_{i=1}^n K_h^2(\mathbf{t}-\mathbf{z}_i)}{J_{n,h}^{-2}}\right)^{p/2} \le$$
$$\le \frac{M_p}{(a_m y)^p}\left(\frac{(p-1)Lh^{-1}}{J_{n,h}^{-1}}\right)^{p/2} \le \frac{M_p}{(a_m y)^p}\left(\frac{(p-1)L}{l\#(N_{n,\delta h}(\mathbf{t}))}\right)^{p/2} \le$$
$$\le \frac{M_p}{(a_m y)^p}\left(\frac{(p-1)L\Delta_{n,h}}{l}\right)^{p/2}. \quad (173)$$

Положим $m = \lceil |\log_2 h| \rceil$ и оценим вторую вероятность в правой части соотношения (171). Отметим, что под знаком вероятности $\mathbf{t} \in \mathcal{R}_s$ и $s \ge m+1$, так что

$$\mathbf{t} = (t_1, \ldots, t_k) = (j_1/2^s, \ldots, j_k/2^s), \qquad \text{где } 1 \le j_r \le 2^s - 1 \quad \text{при } r = 1, \ldots, k.$$

Введем обозначение $\xi_{i,r} = J_{n,h}^{-1}(\mathbf{t} + 2^{-s}\mathbf{e}_r)K_h(\mathbf{t} + 2^{-s}\mathbf{e}_r - \mathbf{z}_i) - J_{n,h}^{-1}(\mathbf{t})K_h(\mathbf{t} - \mathbf{z}_i)$. Используя неравенство (172) при $\eta_i = \xi_{i,r}\varepsilon_i$, с вероятностью 1 имеем

$$\mathbb{P}_{\mathcal{F}_n}\left(\left|\nu_{n,h}(\mathbf{t} + 2^{-s}\mathbf{e}_r) - \nu_{n,h}(\mathbf{t})\right| > a_s y/k\right) \le \frac{\mathbb{E}_{\mathcal{F}_n}\left|\sum_{i=1}^n \eta_i\right|^p}{(a_s y/k)^p} \le \frac{M_p\left((p-1)\sum_{i=1}^n \xi_{i,r}^2\right)^{p/2}}{(a_s y/k)^p}. \quad (174)$$

Заметим теперь, что

$$|\xi_{i,r}| \le \Psi_{n,r} 2^{-s} h^{-1},$$

где

$$\Psi_{n,r} \le \max_{i \le n} \sup_{\frac{j_r}{2^l} \le y_r \le \frac{j_r+1}{2^l}} \left\{ \frac{\left|\widetilde{K_i'}(y_r)\right|}{\sum \widetilde{K_j}(y_r)} + \frac{\widetilde{K_i}(y_r)\sum\left|\widetilde{K_j'}(y_r)\right|}{\left(\sum \widetilde{K_j}(y_r)\right)^2} \right\} \le$$
$$\le 2\left(\frac{L}{l}\right)^2 \sup_{\frac{j_r}{2^s} \le y_r \le \frac{j_r+1}{2^s}} \frac{\#\left(N_{n,h}(\mathbf{t} + \mathbf{e}_r(y_r - t_r))\right)}{\#^2\left(N_{n,\delta h}(\mathbf{t} + \mathbf{e}_r(y_r - t_r))\right)},$$

$$\widetilde{K_i}(y) = K\left(\frac{\mathbf{t} + \mathbf{e}_r(y - t_r) - \mathbf{z}_i}{h}\right),$$
$$\widetilde{K_i'}(y) = K'\left(\frac{\mathbf{t} + \mathbf{e}_r(y - t_r) - \mathbf{z}_i}{h}\right) \quad \text{при} \quad K'(\mathbf{s}) = \frac{\partial K(\mathbf{s})}{\partial s_r},$$

а суммирование в вышеприведенной оценке для $\Psi_{n,r}$ производится по множеству индексов $j \in N_{n,h}(\mathbf{t} + \mathbf{e}_r(y_r - t_r))$. Отметим еще, что $|(j_r+1)/2^s - j_r/2^s| = 2^{-s} \le h$ при $s > m$. Кроме того, порядок суммирования в (174) совпадает при $\mathbf{t} \in \mathcal{R}_s$ с множеством

$$N_{n,h}(\mathbf{t} + 2^{-s}\mathbf{e}_r) \cup N_{n,h}(\mathbf{t}).$$

Таким образом,

$$\mathbb{P}_{\mathcal{F}_n}\Big(\big|\nu_{n,h}(\mathbf{t}+2^{-s}\mathbf{e}_r)-\nu_{n,h}(\mathbf{t})\big|>a_s y/k\Big)\le\frac{M_p\left((p-1)2^{3-2s}\big(L/l\big)^4h^{-2}\Delta_{n,h}\right)^{p/2}}{(a_s y/k)^p}. \qquad (175)$$

При выводе (175) мы учли, что

$$\left(\sup_{\frac{j_r}{2^s}\le y_r\le\frac{j_r+1}{2^s}}\frac{\#\big(N_{n,h}(\mathbf{t}+\mathbf{e}_r(y_r-t_r))\big)}{\#^2\big(N_{n,\delta h}(\mathbf{t}+\mathbf{e}_r(y_r-t_r))\big)}\right)^2\#\left(N_{n,h}(\mathbf{t}+2^{-s}\mathbf{e}_r)\cup N_{n,h}(\mathbf{t})\right)\le$$

$$\le 2\sup_{\frac{j_r}{2^s}\le y_r\le\frac{j_r+1}{2^s}}\frac{\sup\limits_{|s_r|\le h}\#^3\big(N_{n,h}(\mathbf{t}+\mathbf{e}_r(y_r+s_r-t_r))\big)}{\#^4\big(N_{n,\delta h}(\mathbf{t}+\mathbf{e}_r(y_r-t_r))\big)}\le$$

$$\le 2\sup_{\frac{j_1}{2^s}\le y_1\le\frac{j_1+1}{2^s},\dots,\frac{j_k}{2^s}\le y_k\le\frac{j_k+1}{2^s}}\frac{\sup\limits_{\|\mathbf{s}\|\le h}\#^3\big(N_{n,h}(\mathbf{y}+\mathbf{s})\big)}{\#^4\big(N_{n,\delta h}(\mathbf{y})\big)}\le 2\Delta_{n,h}.$$

Теперь из (171), (173) и (175) получаем, что

$$\mathbb{P}_{\mathcal{F}_n}\left(\sup_{\mathbf{t}\in[0,1]^k}|\nu_{n,h}(\mathbf{t})|>y\right)\le\frac{M_p}{y^p}\left(\frac{(p-1)L^4\Delta_{n,h}}{l^4}\right)^{p/2}\left(\frac{2^{mk}}{a_m^p}+C_o h^{-p}\sum_{s=m+1}^{\infty}\frac{2^{-(p-k)s}}{a_s^p}\right)$$

при $C_o=k^{1+p}2^{3p/2}$. Оптимальная последовательность a_s, минимизирующая правую часть неравенства, есть $a_m=c(2^{km})^{1/(p+1)}$ и $a_s=c\big(C_o\,h^{-p}2^{-(p-k)s}\big)^{1/(p+1)}$ при $s=m+1,m+2,\dots,$ где коэффициент c определяется соотношением $a_m+a_{m+1}+\dots=1$. Следовательно,

$$c^{-1}=(2^{km})^{1/(p+1)}+\sum_{s=m+1}^{\infty}\left(C_o h^{-p}2^{-(p-k)s}\right)^{1/(p+1)}$$

и для указанной последовательности получаем, что

$$\frac{2^{mk}}{a_m^p}+C_o h^{-p}\sum_{s=m+1}^{\infty}\frac{2^{-(p-k)s}}{a_s^p}=\left((2^{km})^{1/(p+1)}+\sum_{s=m+1}^{\infty}\left(C_o h^{-p}2^{-(p-k)s}\right)^{1/(p+1)}\right)^{p+1}\le$$

$$\le h^{-k}\left(2^{k/(p+1)}+\frac{C_o^{1/(p+1)}2^{-(p-k)/(p+1)}}{1-2^{-(p-k)/(p+1)}}\right)^{p+1}\equiv h^{-k}C(k,p).$$

При выводе этой оценки мы учли, что $m=\lceil|\log_2 h|\rceil$, а также использовали двойное неравенство $|\log_2 h|\le\lceil|\log_2 h|\rceil\le|\log_2 h|+1$. Лемма доказана. □

Д о к а з а т е л ь с т в о теоремы 1.9. Положим

$$\zeta_n=\sup_{\mathbf{t}\in[0,1]^k}|\nu_{n,h}(\mathbf{t})|,\qquad \eta_n=G\Delta_{n,h}^{1/2}h^{-k/p}$$

и заметим, что в силу леммы 1.24 выполнено $\zeta_n = O_p(\eta_n)$, поскольку

$$\mathbb{P}\big(\zeta_n/\eta_n > M\big) = \mathbb{E}\mathbb{P}_{\mathcal{F}_n}\big(\zeta_n/\eta_n > M\big) \le M^{-p}.$$

Утверждение теоремы следует теперь из представления (169) и леммы 1.22. $\square$

Д о к а з а т е л ь с т в о теоремы 1.10 повторяет вывод теоремы 1.9, в котором вместо леммы 1.24 нужно использовать следующее утверждение.

Лемма 1.25. *В условиях теоремы* 1.10 *для любых* $y > 0$ *и* $h \in (0,1)$ *выполнено*

$$\mathbb{P}_{\mathcal{F}_n}\left(\sup_{t\in[0,1]} |\nu_{n,h}(t)| > y\right) \le 2^4 \left(1 - 2^{-1/3}\right)^{-3} y^{-2}\sigma^2 (L/l)^4 h^{-1}\Delta_{n,h}.$$

Д о к а з а т е л ь с т в о. Как и ранее, воспользуемся методом диадических цепочек. Введем множество двоично-рациональных точек $\mathcal{R} = \{j/2^k;\ j = 1, \dots, 2^k - 1; k \ge 1\}$. Имеем

$$\sup_{t\in[0,1]} |\nu_{n,h}(t)| = \sup_{t\in\mathcal{R}} |\nu_{n,h}(t)| \le$$
$$\le \max_{j=1,\dots,2^m-1} |\nu_{n,h}(j2^{-m})| + \sum_{k=m+1}^{\infty} \max_{j=1,\dots,2^k-2} \big|\nu_{n,h}((j+1)2^{-k}) - \nu_{n,h}(j2^{-k})\big|.$$

Следовательно,

$$\mathbb{P}_{\mathcal{F}_n}\left(\sup_{t\in[0,1]} |\nu_{n,h}(t)| > y\right) \le \sum_{j=1}^{2^m-1} \mathbb{P}_{\mathcal{F}_n}\left(|\nu_{n,h}(j2^{-m})| > a_m y\right) +$$
$$+ \sum_{k=m+1}^{\infty} \sum_{j=1}^{2^k-2} \mathbb{P}_{\mathcal{F}_n}\Big(\big|\nu_{n,h}((j+1)2^{-k}) - \nu_{n,h}(j2^{-k})\big| > a_k y\Big), \quad (176)$$

где $a_m + a_{m+1} + ... = 1$.

С помощью неравенства Чебышева и леммы 1.23 получаем оценку

$$\mathbb{P}_{\mathcal{F}_n}(|\nu_{n,h}(j2^{-m})| > a_m y) \le (a_m y)^{-2} \frac{\sigma^2 L}{l\#(N_{n,\delta h}(j2^{-m}))} \quad \text{п.н.} \quad (177)$$

Положим $m = \lceil |\log_2 h| \rceil$. Тогда

$$\mathbb{P}_{\mathcal{F}_n}\Big(\big|\nu_{n,h}((j+1)2^{-k}) - \nu_{n,h}(j2^{-k})\big| > a_k y\Big) \le \sigma^2 (a_k y)^{-2} G^{(n)}_{j,k,h} \quad (178)$$

при

$$G^{(n)}_{j,k,h} = \sum_{i=1}^{n} \Big(J^{-1}_{n,h}((j+1)2^{-k})\big(K_h((j+1)2^{-k} - z_i) - J^{-1}_{n,h}(j2^{-k})K_h(j2^{-k} - z_i)\big)\Big)^2.$$

Далее, для любых $u, v \in [0,1]$ выполнено

$$|J_{n,h}^{-1}(u)K_h(u-z_i) - J_{n,h}^{-1}(v)K_h(v-z_i)| \le \Psi_n(K,u,v)h^{-1}|u-v|,$$

где

$$\Psi_n(K,u,v) = \max_{i\le n}\sup_{u\le t\le v}\left\{\frac{\left|K'\left(\frac{t-z_i}{h}\right)\right|}{\sum K\left(\frac{t-z_j}{h}\right)} + \frac{K\left(\frac{t-z_i}{h}\right)\sum\left|K'\left(\frac{t-z_j}{h}\right)\right|}{\left(\sum K\left(\frac{t-z_j}{h}\right)\right)^2}\right\} \le$$
$$\le 2\left(\frac{L}{l}\right)^2 \sup_{u\le t\le v}\frac{\#(N_{n,h}(t))}{\#^2(N_{n,\delta h}(t))}$$

и суммирование производится по $j \in N_{n,h}(t)$. Отметим теперь, что $|u-v| \le h$ при $u = (j+1)2^{-k}$, $v = j2^{-k}$ и $k > m$. Кроме того, порядок суммирования в первом неравенстве соотношения (178) совпадает с множеством $N_{n,h}((j+1)2^{-k}) \bigcup N_{n,h}(j2^{-k})$. С учетом этих фактов заключаем, что

$$G_{j,k,h}^{(n)} \le 2^{3-2k}(L/l)^4\Delta_{n,h}h^{-2}. \tag{179}$$

Теперь из (176)–(179) получаем

$$\mathbb{P}_{\mathcal{F}_n}\Big(\sup_{t\in[0,1]}|\nu_{n,h}(t)| > y\Big) \le 2^3 y^{-2}\sigma^2(L/l)^4\Delta_{n,h}\Big(2^m a_m^{-2} + h^{-2}\sum_{k=m+1}^{\infty}2^{-k+1}a_k^{-2}\Big). \tag{180}$$

Оптимальная последовательность a_k, минимизирующая правую часть этого неравенства, есть $a_m = c2^{m/3}$ и $a_k = ch^{-2/3}2^{(-k+1)/3}$ при $k = m+1, m+2, ...$, где

$$c^{-1} = 2^{m/3} + h^{-2/3}\sum_{k>m}2^{(-k+1)/3} = 2^{m/3} + h^{-2/3}\frac{2^{-m/3}}{1-2^{-1/3}}.$$

Для указанной последовательности получаем, что

$$2^m a_m^{-2} + h^{-2}\sum_{k=m+1}^{\infty}2^{-k+1}a_k^{-2} = \left(2^{m/3} + h^{-2/3}\frac{2^{-m/3}}{1-2^{-1/3}}\right)^3. \tag{181}$$

Соотношения $m = \lceil|\log_2 h|\rceil$, $|\log_2 h| \le \lceil|\log_2 h|\rceil \le |\log_2 h| + 1$, а также (180) и (181), влекут за собой оценку

$$\mathbb{P}_{\mathcal{F}_n}\Big(\sup_{t\in[0,1]}|\nu_{n,h}(t)| > y\Big) \le 2^4\left(1-2^{-1/3}\right)^{-3}y^{-2}\sigma^2(L/l)^4\Delta_{n,h}h^{-1}.$$

Таким образом, лемма 1.25, а вместе с ней и теорема 1.10, доказаны. □

1.4.3 Классические локально–линейные оценки

Классические локально–линейные оценки мы рассмотрим в случае скалярного аргумента регрессионной функции. Напомним (см., например, [77]), что классическая локально–линейная ядерная оценка $\widehat{f}_{LL}(t)$ определяется в этом случае как первая координата двумерной точки, на которой достигается следующий минимум:

$$\min_{a,b}\sum_{i=1}^{n}\bigl(X_i-\bigl(a+b(t-z_i)\bigr)\bigr)^2 K_h(t-z_i).$$

Нетрудно видеть, что имеет место представление

$$\widehat{f}_{LL}(t)=\sum_{i=1}^{n}\frac{S_{n,2}(t)-(t-z_i)S_{n,1}(t)}{S_{n,0}(t)S_{n,2}(t)-S_{n,1}^2(t)}X_iK_h(t-z_i), \tag{182}$$

где

$$S_{n,j}(t)=\sum_{i=1}^{n}(t-z_i)^jK_h(t-z_i),\quad j=0,1,2. \tag{183}$$

Далее считаем, что при $k=1$ и $\mathcal{P}=[0,1]$ выполнены условия (M_1) и (K_4), при этом регрессионная функция f неслучайна. Для любых $z_1,\dots,z_n$, $h\in(0,1)$ и $t\in[0,1]$ сохраним следующее обозначение, введенное ранее в (164):

$$N_{n,h}(t)=\{i:\ |t-z_i|\le h,\ 1\le i\le n\}.$$

Рассмотрим прежде всего вопрос о поточечной состоятельности оценок (182). Нам потребуется еще одно предположение, являющееся единственным условием на регрессоры, гарантирующим поточечную состоятельность классических локально–линейных оценок $\widehat{f}_{LL}(t)$.

$(\mathbf{D}_5)$ *При всех фиксированных* $t\in[0,1]$ *и* $h\in(0,1)$ *имеет место предельное соотношение*

$$\Lambda_{n,h}(t)=\frac{\#^3(N_{n,h}(t))}{\#^2(N_{n,\delta h}(t))\widetilde{N}^2_{n,h,\delta}(t)}\xrightarrow{p}0,$$

где величина $\widetilde{N}_{n,h,\delta}(t)$ *определяется равенством*

$$\widetilde{N}_{n,h,\delta}(t)=\min\Bigl\{\#(N_{n,\frac{\delta h}{2}}(t)),\ \#(N_{n,\frac{\delta h}{4}}(t-3\delta h/4),\ \#(N_{n,\frac{\delta h}{4}}(t+3\delta h/4)\Bigr\}.$$

Теорема 1.11. *Пусть выполнены условия* (M_1) *и* (K_4) *при* $k=1$ *и условие* (E_1)*. Тогда для любого* $h\in(0,1)$ *и любого* $t\in[0,1]$ *с вероятностью* 1 *выполнено*

$$\mathbb{E}_{\mathcal{F}_n}\bigl(\widehat{f}_{LL}(t)-f(t)\bigr)^2\le\omega_f^2(h)+64\sigma^2L^4(l\delta)^{-4}\,\Lambda_{n,h}(t).$$

Следствие 1.11. *В условиях теоремы* 1.11 *имеет место соотношение*

$$\widehat{f}_{LL}(t) - f(t) = O_p\left(\omega_f(h) + 8\sigma L^2(l\delta)^{-2}\Lambda_{n,h}^{1/2}(t)\right).$$

Следствие 1.12. *Пусть выполнены условия* (M_1) *и* (K_4) *при* $k=1$, *а также условия* (E_1) *и* (D_5). *Тогда существует последовательность* $h = h_n \to 0$, *для которой при всех* $t \in [0,1]$ *имеет место предельное соотношение*

$$\widehat{f}_{LL}(t) \xrightarrow{p} f(t).$$

З а м е ч а н и е 1.26. Легко видеть, что в условиях теоремы 1.11 для любого $h \in (0,1)$ и любого $t \in [0,1]$ справедливо неравенство

$$\mathbb{E}\big(\widehat{f}_{LL}(t) - f(t)\big)^2 \le \omega_f^2(h) + 64\sigma^2 L^4(l\delta)^{-4}\,\mathbb{E}\Lambda_{n,h}(t).$$

Однако условие (D_5) без уточнения характера корреляции регрессоров еще не гарантирует выполнение предельного соотношения $\mathbb{E}\Lambda_{n,h}(t) \to 0$ при всех фиксированных t и h. Но если это соотношение справедливо, то существует последовательность h_n, для которой поточечный риск оценки $\widehat{f}_{LL}(t)$ сходится к нулю:

$$\mathbb{E}\big(\widehat{f}_{LL}(t) - f(t)\big)^2 \to 0 \qquad \text{при всех} \quad t \in [0,1].$$

Аналогичное замечание справедливо и для оценок Надарая–Ватсона (см. теорему 1.8).

З а м е ч а н и е 1.27. Обсудим подробнее (D_5). В широких условиях для всех известных нам неслучайных регулярных регрессоров выполнено $\#(N_{n,h}(t)) \sim nh$ равномерно по t, что гарантирует выполнение (D_5). Если $\{z_i\}$ независимы и одинаково распределены и отрезок $[0,1]$ является носителем распределения z_1 (функция распределения z_1 является строго монотонной на $[0,1]$), то условие (D_5) также выполнено, поскольку по теореме Гливенко–Кантелли $\#(N_{n,h}(t)) \sim nh$ равномерно по t почти наверное[4]. Если $\{z_i\}$ — стационарная последовательность с условием α-перемешивания и маргинальным распределением с носителем $[0,1]$, то в широких условиях $\#(N_{n,h}(t)) \sim nh$ по вероятности[5] для любой точки t и условие (D_5) также выполнено. Подчеркнем, что выполнение условия (D_5) вполне возможно и для других типов зависимости случайных величин $\{z_i\}$ (см. пример 1.19).

Отметим еще, что во всех указанных в данном замечании случаях величина $\Lambda_{n,h}(t)$ имеет порядок $(nh)^{-1}$. Если дополнительно функция f удовлетворяет условию Липшица, то мож-

[4]Данная запись означает, что с вероятностью 1 выполнено $c_1 h \le \limsup_{t\in[0,1]} \#(N_{n,h}(t))/n \le c_2 h$, где константы c_1 и c_2 не зависят от t и h.

[5]Данная запись означает, что $\mathbb{P}\big(\#(N_{n,h}(t))/(nh) \in [c_1, c_2]\big) \to 1$, где $c_1 > 0$ и $c_2 > 0$.

но положить $h = n^{-1/3}$. Выбранный таким образом размер окна h фактически уравнивает порядок малости по h обоих слагаемых в правой части утверждения следствия 1.11, а соответствующее поточечное расстояние из следствия 1.11 имеет порядок $O_p(n^{-1/3})$.

П р и м е р 1.19. Вернемся к примерам 1.1 и 1.2 раздела 1.1.1: последовательность равномерно распределенных на отрезке $[0,1]$ случайных величин $\{z_i;\ i \geq 1\}$ определяется соотношением

$$z_i = \nu_i U_i^l + (1-\nu_i) U_i^r, \tag{184}$$

где $\{U_i^l\}$ и $\{U_i^r\}$ независимы и равномерно распределены на $[0,1/2]$ и $[1/2,1]$ соответственно, последовательность $\{\nu_i\}$ не зависит от $\{U_i^l\}$, $\{U_i^r\}$ и состоит из бернуллиевских случайных величин с вероятностью успеха $1/2$.

Пусть зависимость между случайными величинами ν_i для любого натурального i определяется равенствами $\nu_{2i-1} = \nu_1$ и $\nu_{2i} = 1 - \nu_1$. В этом случае, как мы уже отмечали выше, для стационарной последовательности случайных величин $\{z_i\}$ наиболее популярные условия слабой зависимости не выполнены. В частности, не выполнены те или иные условия перемешивания, так как при всех натуральных m и n $\mathbb{P}(z_{2m} \leq 1/2,\ z_{2n-1} \leq 1/2) = 0$. С другой стороны, нетрудно показать, что для эмпирической функции распределения, построенной по выборке растущего объема из стационарной последовательности $\{z_i;\ i \geq 1\}$, справедлива теорема Гливенко–Кантелли. Это означает, что для любого фиксированного $h > 0$ выполнено $N_{n,h}(t) \sim nh$ почти наверное равномерно по t, что влечет выполнение условия (D_5).

Предположим теперь, что зависимость между случайными величинами ν_j определяется следующими соотношениями: $\nu_j = 1 - \nu_1$ при $j = 2^{2k-1}, \ldots, 2^{2k} - 1$ и $\nu_j = \nu_1$ при $j = 2^{2k}, \ldots, 2^{2k+1} - 1$, где $k = 1, 2, \ldots$. Иными словами, на первом шаге стохастического эксперимента мы разыгрываем с равными шансами ту или иную половинку единичного отрезка, на которую наудачу бросаем первую точку. Затем мы бросаем наудачу две точки на другую половинку отрезка $[0,1]$ и вновь возвращаемся на исходную половинку, куда теперь бросаем наудачу четыре точки. Процесс чередования половинок единичного отрезка продолжается, и на k-м шаге процедуры мы бросаем наудачу на очередную половинку единичного отрезка 2^{k-1} точек. Как мы показали выше (см. пример 1.2), данная нестационарная последовательность равномерно распределенных на $[0,1]$ случайных величин $\{z_i\}$ не удовлетворяет усиленному закону больших чисел. Кроме того, эта последовательность не удовлетворяет и теореме Гливенко–Кантелли, но используя эту теорему отдельно для каждой из последовательностей $\{U_i^l\}$ и $\{U_i^r\}$, нетрудно получить следующие две оценки, справедливые с вероятностью 1:

$$\limsup \frac{1}{n} \sup_{t\in[0,1]} \#(N_{n,h}(t)) \leq 6h, \qquad \liminf \frac{1}{n} \inf_{t\in[0,1]} \#(N_{n,h}(t)) \geq 2h/3,$$

где h произвольное фиксированное число из интервала $(0,1)$. Таким образом, для рассматриваемой последовательности сильно зависимых случайных величин условие (D_5) выполнено. $\square$

Рассмотрим теперь вопрос о равномерной состоятельности классических локально–линейных оценок. Нам потребуется еще одно предположение.

($\mathbf{D}_6$) *Для любых* $t \in [0,1]$ *и* $h \in (0,1)$ *имеет место предельное соотношение*

$$\widetilde{\Delta}_{n,h} = \sup_{t\in[0,1]} \frac{\sup_{|s|\le h} \#^7(N_{n,h}(t+s))}{\#^4(N_{n,\delta h}(t))\widetilde{N}^4_{n,h,\delta}(t)} \xrightarrow{p} 0.$$

По умолчанию в (D_6) под внутренним знаком супремума область изменения переменной s при каждом фиксированном $t \in [0,1]$ определяется соотношениями $s \in [0,1]$ и $s+t \in [0,1]$.

Теорема 1.12. *Пусть при* $k = 1$ *выполнены условия* (M_1), (E_1) *и* (K_5). *Тогда для любого фиксированного* $h \in (0,1)$ *с вероятностью* 1

$$\sup_{t\in[0,1]} \left|\widehat{f}_{LL}(t) - f(t)\right| \le \omega_f(h) + \widetilde{\zeta}_n(h),$$

где $\widetilde{\zeta}_n(h)$ *есть последовательность положительных случайных величин, удовлетворяющих соотношению*

$$\widetilde{\zeta}_n(h) = O_p\left(\sigma L^4(l\delta)^{-4}\sqrt{\widetilde{\Delta}_{n,h}h^{-1}}\right).$$

Следствие 1.13. *Пусть выполнены условия теоремы* 1.12 *и условие* (D_6). *Тогда существует последовательность* $h = h_n \to 0$, *для которой*

$$\sup_{t\in[0,1]} \left|\widehat{f}_{LL}(t) - f(t)\right| \xrightarrow{p} 0.$$

З а м е ч а н и е 1.28. Нетрудно видеть, что условие (D_6) выполнено, если для всех достаточно малых $h > 0$

$$\widetilde{\Delta}'_{n,h} = \frac{\sup_{t\in[0,1]} \#^7(N_{n,h}(t))}{\inf_{t\in[0,1]} \#^4(N_{n,\delta h}(t))\widetilde{N}^4_{n,\delta,h}(t)} \xrightarrow{p} 0.$$

Скажем, пусть $\#(N_{n,h}(t)) \sim nh$ с вероятностью 1 равномерно по t (см. замечание 1.27 и пример 1.19 сильно зависимых $\{z_i\}$). Тогда с вероятностью 1

$$\widetilde{\Delta}'_{n,h} = O((hn)^{-1}).$$

Если $\tilde{\Delta}'_{n,h} \sim (nh)^{-1}$ и функция f удовлетворяет условию Липшица, то можно положить $h = n^{-1/4}$. В этом случае $\sup_{t\in[0,1]} \left|\widehat{f}_{LL}(t) - f(t)\right| = \widetilde{O}_p(n^{-1/4})$. $\square$

З а м е ч а н и е 1.29. В силу теоремы 1.12 имеем

$$\mathbb{E}\sup_{t\in[0,1]}\left|\widehat{f}_{LL}(t)-f(t)\right|^2\le 2\omega_f(h)+4\int\limits_0^\infty y\mathbb{P}\big(\widetilde{\zeta}_n(h)>y\big)dy. \tag{185}$$

В условиях теоремы 1.12 (см. доказательство теоремы и, в частности, утверждение леммы 1.28) величина $\mathbb{P}\big(\widetilde{\zeta}_n(h)>y\big)$ имеет порядок y^{-2}. Если в условии (E_1) считать $\{\varepsilon_i\}$ условно относительно $\mathcal{F}_n$ независимыми и центрированными случайными величинами с более сильным моментным ограничением $\sup_{i\le n}\mathbb{E}_{\mathcal{F}_n}|\varepsilon_i|^{2+\gamma}\le\mu<\infty$ при некотором $\gamma>0$ и неслучайном $\mu>0$, то в доказательстве теоремы 1.12 можно использовать моментное (порядка $2+\gamma$) неравенство Розенталя. Как следствие, для величины $\mathbb{P}\big(\widetilde{\zeta}_n(h)>y\big)$ можно получить оценку порядка $y^{-(2+\gamma)}$, которая ввиду соотношения (185) позволит получить и оценку для равномерного риска $\mathbb{E}\sup\limits_{t\in[0,1]}\left|\widehat{f}_{LL}(t)-f(t)\right|^2$. Аналогичные рассуждения справедливы и для оценок Надарая–Ватсона (см. теорему 1.9).

З а м е ч а н и е 1.30. Условие (D_6), гарантирующее равномерную состоятельность локально–линейных оценок (182), влечет за собой выполнение условия (D_5) для поточечной состоятельности этих оценок, что является весьма ожидаемым и с учетом приводимого далее замечания 1.31 следует из очевидного соотношения

$$\widetilde{\Delta}_{n,h}=\sup_{t\in[0,1]}\frac{\sup_{|s|\le h}\#^7(N_{n,h}(t+s))}{\#^4(N_{n,\delta h}(t))\widetilde{N}^4_{n,h,\delta}(t)}\ge\#(N_{n,h}(t))\Lambda^2_{n,h}(t),$$

справедливого для любого t.

З а м е ч а н и е 1.31. Отметим, что требование $\#(N_{n,h}(t))\ne 0$ при всех фиксированных $t\in[0,1]$ и $h\in(0,1)$ (начиная с некоторого n), которое влекут условия (D_5) и (D_6) (см. соглашения в начале работы), в силу аддитивности меры $\#(\cdot)$ может быть переписано в эквивалентной форме следующим образом: при всех фиксированных $t\in[0,1]$ и $h\in(0,1)$ имеет место предельное соотношение $\#(N_{n,h}(t))\xrightarrow{p}\infty$. Другими словами, выполнено условие (D_3) (или эквивалентное ему условие (D_1)).

Напомним, что в начале главы предложены новые классы оценок ядерного типа для регрессионной функции f, являющиеся равномерно состоятельными лишь при выполнении условия (D_1) относительно регрессоров, означающего, что регрессоры $\{z_i\}$ с высокой вероятностью образуют измельчающееся разбиение отрезка $[0,1]$ — области определения регрессионной функции f. Сравнивая условия (D_1) или (D_3) и предположения (D_5), (D_6), отметим, что условия (D_5) и (D_6) по сути требуют более равномерное плотное заполнение регрессорами области определения регрессионной функции, нежели в условиях (D_3) или (D_1). Приведем соответствующий пример.

П р и м е р 1.20. Как и в примере 1.19 считаем, что последовательность случайных величин $\{z_i;\ i \geq 1\}$ определяется соотношением (184), где $\{U_i^l\}$ и $\{U_i^r\}$ независимы и равномерно распределены на $[0, 1/2]$ и $[1/2, 1]$ соответственно, последовательность $\{\nu_i\}$ не зависит от $\{U_i^l\}$, $\{U_i^r\}$ и состоит из бернуллиевских случайных величин с вероятностью успеха $1/2$, а зависимость между случайными величинами ν_i для любого натурального i зададим теперь равенствами $\nu_{[e^k]} = \nu_1$ при $k \geq 1$ и $\nu_i = 1 - \nu_1$ при всех $i \neq [e^k]$, $k \geq 1$. Иначе говоря, сначала мы разыгрываем, на какой из двух отрезков бросаем наудачу первую точку, и далее чередуем количество бросаний таким образом, что с вероятностями $1/2$ на один из отрезков попадает $\log n$ точек, а на другой — остальные (т.е. порядка n). В этом случае нестационарная последовательность $\{z_i\}$ не удовлетворяет усиленному закону больших чисел, так как нормированная сумма элементов $\{z_i\}$ почти наверное сходится к некоторой случайной величине, имеющей двухточечное распределение ($1/4$ и $3/4$ с равными вероятностями). Понятно, что условие (D_3) (или эквивалентное ему условие (D_1)) в данном примере выполнено.

Покажем, что для рассматриваемой последовательности сильно зависимых случайных величин не выполнено условие (D_5). Действительно, на почти всех элементарных исходах из множества $\{\nu_1 = 1\}$ мы имеем $\#(N_{n,h}(1/2 - \delta h)) \sim hn$, $\#(N_{n,\delta h}(1/2 - \delta h)) \sim h \log n$, $\widetilde{N}^4_{n,h,\delta}(1/2 - \delta h) \sim h \log n$. Таким образом, на множестве указанных элементарных исходов, имеющем вероятность $1/2$,

$$\Lambda_{n,h}(1/2 - \delta h) = \frac{\#^3(N_{n,h}(1/2 - \delta h))}{\#^2(N_{n,\delta h}(1/2 - \delta h))\widetilde{N}^2_{n,h,\delta}(1/2 - \delta h)} \sim \frac{n^3}{h \log^4 n} \to \infty,$$

так что при $t = 1/2 - \delta h$ предельное соотношение (D_5) не выполнено. $\square$

1.4.4 Доказательства утверждений раздела 1.4.3

Для оценки $\widehat{f}_{LL}(t)$, определенной в (182), мы будем использовать следующее представление:

$$\widehat{f}_{LL}(t) = \Big(\sum\nolimits_{i=1}^n w_i(t)\Big)^{-1} \sum\nolimits_{i=1}^n w_i(t) X_i, \tag{186}$$

где

$$w_i(t) = K_h(t - z_i)\big(S_{n,2}(t) - (t - z_i) S_{n,1}(t)\big). \tag{187}$$

Учитывая соотношения (186) и (23), получаем тождество

$$\widehat{f}_{LL}(t) = f(t) + \widetilde{\eta}_{n,h}(t) + \widetilde{\nu}_{n,h}(t), \tag{188}$$

$$\widetilde{\eta}_{n,h}(t) = \frac{\sum_{i=1}^{n} \big(f(z_i) - f(t)\big) w_i(t)}{\sum_{i=1}^{n} w_i(t)}, \qquad \widetilde{\nu}_{n,h}(t) = \frac{\sum_{i=1}^{n} w_i(t)\varepsilon_i}{\sum_{i=1}^{n} w_i(t)}. \tag{189}$$

З а м е ч а н и е 1.32. Подчеркнем, что ввиду свойств плотности $K_h(\cdot)$ область суммирования во всех суммах в (183) и (189), а также во всех приводимых далее формулах, когда при суммировании функция $K_h(\cdot)$ участвует в качестве множителя, совпадает с множеством $N_{n,h}(t)$. Указанный факт является принципиальным для дальнейшего анализа.

Лемма 1.26. *Имеет место следующая оценка*:

$$\sup_{t\in[0,1]} |\widetilde{\eta}_{n,h}(t)| \le \omega_f(h).$$

Д о к а з а т е л ь с т в о. Утверждение леммы следует из (189) и замечания 1.32. □

Лемма 1.27. *Для любого* $t \in [0,1]$ *с вероятностью* 1 *имеет место неравенство*

$$\mathbb{E}_{\mathcal{F}_n} \widetilde{\nu}^2_{n,h}(t) \le 64\sigma^2 L^4 (l\delta)^{-4}\, \Lambda_{n,h}(t),$$

где величина $\Lambda_{n,h}(t)$ *определена в условии* (D_5).

Д о к а з а т е л ь с т в о. С вероятностью 1 имеет место соотношение

$$\mathbb{E}_{\mathcal{F}_n} \widetilde{\nu}^2_{n,h}(t) \le \frac{\sigma^2 \sum_{i\in N_{n,h}(t)} w_i^2(t)}{w^2(t)}, \quad \text{где} \quad w(t) = \sum_{i\in N_{n,h}(t)} w_i(t). \tag{190}$$

С учетом определений (187), (183) и замечания 1.32, для любого $t \in [0,1]$ выполнено

$$\max_{i\in N_{n,h}(t)} |w_i(t)| \le 2L^2\, \#(N_{n,h}(t)), \qquad \sum_{i\in N_{n,h}(t)} w_i^2(t) \le 4L^4\, \#^3(N_{n,h}(t)). \tag{191}$$

Нам остается оценить снизу величину $w(t)$. Из определений (187) и (183) имеем

$$\begin{aligned} w(t) = S^2_{n,0}(t) \left(\frac{S_{n,2}(t)}{S_{n,0}(t)} - \frac{S^2_{n,1}(t)}{S^2_{n,0}(t)} \right) = \\ = S_{n,0}(t) h^2 \sum_{i\in N_{n,h}(t)} \left(\frac{z_i - t}{h} - \frac{S_{n,1}(t)}{h S_{n,0}(t)} \right)^2 K_h(z_i - t) \ge \\ \ge S_{n,0}(t) h^2 \sum_{i\in N_{n,\delta h}(t)} \left(\frac{z_i - t}{h} - \frac{S_{n,1}(t)}{h S_{n,0}(t)} \right)^2 K_h(z_i - t) \ge \\ \ge 4^{-1} l^2 \delta^2 \#(N_{n,\delta h}(t)) \min\left\{ \#(N_{n,\frac{\delta h}{2}}(t)),\ \#(N_{n,\frac{\delta h}{4}}(t - 3\delta h/4),\ \#(N_{n,\frac{\delta h}{4}}(t + 3\delta h/4) \right\}. \end{aligned} \tag{192}$$

При выводе этого соотношения мы учли, что для любого $t \in [0,1]$ и всех $i \in N_{n,\delta h}(t)$ выполнено $K_h(z_i - t) \geq h^{-1} l$, а потому

$$S_{n,0}(t) \geq \sum_{i \in N_{n,\delta h}(t)} K_h(z_i - t) \geq h^{-1} l \, \#(N_{n,\delta h}(t)).$$

Кроме того, мы учли, что при всех $t \in [0,1]$ справедливы соотношения

$$S(t) = \frac{S_{n,1}(t)}{h S_{n,0}(t)} \in [-1,1], \qquad \frac{z_i - t}{h} \in [-\delta, \delta] \quad \text{при} \quad i \in N_{n,\delta h}(t).$$

Следовательно, если $|S(t)| \geq \delta$, то

$$\widetilde{S} = \sum_{i \in N_{n,\delta h}(t)} \left(\frac{z_i - t}{h} - \frac{S_{n,1}(t)}{h S_{n,0}(t)} \right)^2 \geq \sum_{i \in N_{n,\frac{\delta h}{2}}(t)} \left(\frac{z_i - t}{h} - \frac{S_{n,1}(t)}{h S_{n,0}(t)} \right)^2 \geq \frac{\delta^2}{4} \, \#(N_{n,\frac{\delta h}{2}}(t)).$$

Если же $0 \leq S(t) < \delta$, то

$$\widetilde{S} \geq \sum_{i \in N_{n,\frac{\delta h}{4}}(t - 3\delta h/4)} \left(\frac{z_i - t}{h} - \frac{S_{n,1}(t)}{h S_{n,0}(t)} \right)^2 \geq \frac{\delta^2}{4} \, \#(N_{n,\frac{\delta h}{4}}(t - 3\delta h/4)).$$

Аналогично, если $-\delta \leq S(t) < 0$, то

$$\widetilde{S} \geq \sum_{i \in N_{n,\frac{\delta h}{4}}(t + 3\delta h/4)} \left(\frac{z_i - t}{h} - \frac{S_{n,1}(t)}{h S_{n,0}(t)} \right)^2 \geq \frac{\delta^2}{4} \, \#(N_{n,\frac{\delta h}{4}}(t + 3\delta h/4)).$$

Приведенные соотношения доказывают (192). Утверждение леммы следует теперь из оценок (190), (191), (192) и обозначений в условии (D_5). ☐

Д о к а з а т е л ь с т в о теоремы 1.11. Это утверждение следует из равенства

$$\mathbb{E}_{\mathcal{F}_n} \big(\widehat{f}_{LL}(t) - f(t) \big)^2 = \widetilde{\eta}^2_{n,h}(t) + \mathbb{E}_{\mathcal{F}_n} \widetilde{\nu}^2_{n,h}(t) \;\; \text{п.н.},$$

при выводе которого мы учли тождество (188) и утверждения лемм 1.26 и 1.27. ☐

Д о к а з а т е л ь с т в о следствия 1.11. Положим

$$\zeta_n = \widehat{f}_{LL}(t) - f(t), \qquad \eta_n = \omega_f(h) + 8\sigma L^2 (l\delta)^{-2} \, \Lambda^{1/2}_{n,h}(t)$$

и заметим, что в силу теоремы 1.11 имеет место оценка $\mathbb{E}_{\mathcal{F}_n} \zeta_n^2 \leq \eta_n^2$. Следовательно, справедлива цепочка соотношений

$$\mathbb{P}\big(|\zeta_n| / \eta_n > M \big) = \mathbb{E} \mathbb{P}_{\mathcal{F}_n} \big(|\zeta_n| / \eta_n > M \big) \leq \mathbb{E} \frac{\mathbb{E}_{\mathcal{F}_n} \zeta_n^2}{M^2 \eta_n^2} \leq \frac{1}{M^2},$$

которая и доказывает утверждение следствия. ☐

Д о к а з а т е л ь с т в о следствия 1.12 с очевидными изменениями повторяет вывод следствия 1.8, поэтому подробности мы опускаем. □

Для вывода теоремы 1.12 нам потребуется следующее ключевое утверждение.

Лемма 1.28. *В условиях теоремы* 1.12 *для любых* $y > 0$ *и* $h \in (0,1)$ *выполнено*

$$\mathbb{P}_{\mathcal{F}_n}\Big(\sup_{t\in[0,1]}\big|\widetilde{\nu}_{n,h}(t)\big| > y\Big) \le 2^9\cdot 23\left(1-2^{-1/3}\right)^{-3} y^{-2}\sigma^2 L^8 (l\delta)^{-8}\widetilde{\Delta}_{n,h}h^{-1},$$

где величина $\widetilde{\Delta}_{n,h}$ *определена в условии* (D_6).

Д о к а з а т е л ь с т в о. Используя метод диадических цепочек для оценивания хвоста распределения величины $\sup_{t\in[0,1]}|\nu_{n,h}(t)|$, заменим интервал $[0,1]$ множеством двоично-рациональных точек $\mathcal{R} = \{j/2^k;\ j = 1,\dots,2^k-1; k \ge 1\}$. Таким образом, при некотором натуральном m справедливо соотношение

$$\sup_{t\in[0,1]}\big|\widetilde{\nu}_{n,h}(t)\big| = \sup_{t\in\mathcal{R}}\big|\widetilde{\nu}_{n,h}(t)\big| \le$$
$$\le \max_{j=1,\dots,2^m-1}\big|\widetilde{\nu}_{n,h}(j2^{-m})\big| + \sum_{k=m+1}^{\infty}\max_{j=1,\dots,2^k-2}\big|\widetilde{\nu}_{n,h}((j+1)2^{-k}) - \widetilde{\nu}_{n,h}(j2^{-k})\big|.$$

Следовательно,

$$\mathbb{P}_{\mathcal{F}_n}\Big(\sup_{t\in[0,1]}\big|\widetilde{\nu}_{n,h}(t)\big| > y\Big) \le \mathbb{P}_{\mathcal{F}_n}\Big(\max_{j=1,\dots,2^m-1}\big|\widetilde{\nu}_{n,h}(j2^{-m})\big| > a_m y\Big) +$$
$$+ \sum_{k=m+1}^{\infty}\mathbb{P}_{\mathcal{F}_n}\Big(\max_{j=1,\dots,2^k-2}\big|\widetilde{\nu}_{n,h}((j+1)2^{-k}) - \widetilde{\nu}_{n,h}(j2^{-k})\big| > a_k y\Big) \le$$
$$\le \sum_{j=1}^{2^m-1}\mathbb{P}_{\mathcal{F}_n}\left(\big|\widetilde{\nu}_{n,h}(j2^{-m})\big| > a_m y\right) +$$
$$+ \sum_{k=m+1}^{\infty}\sum_{j=1}^{2^k-2}\mathbb{P}_{\mathcal{F}_n}\Big(\big|\widetilde{\nu}_{n,h}((j+1)2^{-k}) - \widetilde{\nu}_{n,h}(j2^{-k})\big| > a_k y\Big), \quad (193)$$

где последовательность $a_m, a_{m+1}, \ldots$ положительных чисел такова, что $\sum_{i\ge m} a_i = 1$.

С помощью неравенства Маркова со вторым моментом и леммы 1.27, получаем

$$\mathbb{P}_{\mathcal{F}_n}\left(\big|\widetilde{\nu}_{n,h}(j2^{-m})\big| > a_m y\right) \le 64\sigma^2 L^4\big(l^2\delta^2 a_m y\big)^{-2}\,\Lambda_{n,h}(j2^{-m}) \quad \text{п.н.} \quad (194)$$

Положим $m = \lceil|\log_2 h|\rceil$. В этом случае

$$\mathbb{P}_{\mathcal{F}_n}\Big(\big|\widetilde{\nu}_{n,h}((j+1)2^{-k}) - \widetilde{\nu}_{n,h}(j2^{-k})\big| > a_k y\Big) \le \sigma^2(a_k y)^{-2}G^{(n)}_{j,k,h}, \quad (195)$$

где

$$G^{(n)}_{j,k,h} = \sum_{i=1}^{n} \left(\frac{w_i((j+1)2^{-k})}{w((j+1)2^{-k})} - \frac{w_i(j2^{-k})}{w(j2^{-k})} \right)^2, \tag{196}$$

а величина $w(t)$ определена в (190). Оценим величину $G^{(n)}_{j,k,h}$. Для любых $u, v \in [0,1]$ с условием $|u-v| \le h$ и всех $i \in N_{n,h}(u) \cup N_{n,h}(v)$ выполнено

$$|w^{-1}(u)w_i(u) - w^{-1}(v)w_i(v)| \le \Psi_n(K,u,v)h^{-1}|u-v|,$$

где

$$\Psi_n(K,u,v) = \max_{i \in N_{n,h}(u) \cup N_{n,h}(v)} \sup_{u \le t \le v} \left\{ \frac{|w_i'(t)|\, w(t) + |w_i(t)| \sum_{j \in N_{n,h}(t)} \left|w_j'(t)\right|}{w^2(t)} \right\} \le$$

$$\le \frac{184 L^4}{l^4 \delta^4 h} \sup_{u \le t \le v} \frac{\#^3(N_{n,h}(t))}{\#^2(N_{n,\delta h}(t)) \widetilde{N}^2_{n,h,\delta}(t)}.$$

При выводе последней оценки мы использовали (191), (192) и соотношения

$$|w_j'(t)| \le 8L^2 h^{-1} \#(N_{n,h}(t)) \quad \text{при} \quad j \in N_{n,h}(t),$$

$$K_h(t - z_i) \le h^{-1}L,\ |t - z_i| \le 2h,\ |w_i(t)| \le 3L^2 \#(N_{n,h}(t)),\ |w_i'(t)| \le 11 L^2 h^{-1} \#(N_{n,h}(t))$$

при $i \in N_{n,h}(u) \cup N_{n,h}(v)$, $|u-v| \le h$ и $u \le t \le v$. Принимая во внимание, во-первых, что $|u-v| \le h$ при $u = (j+1)2^{-k}$, $v = j2^{-k}$ и $k > m$, и во-вторых, что множество индексов суммирования в первом неравенстве соотношения (196) совпадает с множеством

$$N_{n,h}((j+1)2^{-k}) \bigcup N_{n,h}(j2^{-k}),$$

заключаем, что

$$G^{(n)}_{j,k,h} \le 2 \cdot 184^2 2^{-2k} L^8 (l\delta)^{-8} \widetilde{\Delta}_{n,h} h^{-2}. \tag{197}$$

Теперь из (193)–(197) нетрудно получить, что

$$\mathbb{P}_{\mathcal{F}_n}\Big(\sup_{t \in [0,1]} \left| \widetilde{\nu}_{n,h}(t) \right| > y \Big) \le 2^8 \cdot 23\, y^{-2} \sigma^2 L^8 (l\delta)^{-8} \widetilde{\Delta}_{n,h} \Big(2^m a_m^{-2} + h^{-2} \sum_{k=m+1}^{\infty} 2^{-k+1} a_k^{-2} \Big).$$

Оптимальная последовательность a_k, минимизирующая правую часть этого неравенства, есть $a_m = c2^{m/3}$ и $a_k = ch^{-2/3} 2^{(-k+1)/3}$ при $k = m+1, m+2, \ldots$, где

$$c^{-1} = 2^{m/3} + h^{-2/3} \sum_{k > m} 2^{(-k+1)/3} = 2^{m/3} + h^{-2/3} \frac{2^{-m/3}}{1 - 2^{-1/3}}.$$

Для указанной последовательности получаем, что

$$2^m a_m^{-2} + h^{-2} \sum_{k=m+1}^{\infty} 2^{-k+1} a_k^{-2} = \left(2^{m/3} + h^{-2/3} \frac{2^{-m/3}}{1 - 2^{-1/3}} \right)^3. \tag{198}$$

Учтем, что $m = \lceil |\log_2 h| \rceil$. Соотношение (198) и неравенство $|\log_2 h| \le \lceil |\log_2 h| \rceil \le |\log_2 h| + 1$ влекут за собой оценку

$$\mathbb{P}_{\mathcal{F}_n}\Big(\sup_{t\in[0,1]} \big|\widetilde{\nu}_{n,h}(t)\big| > y \Big) \le 2^9 \cdot 23 \left(1 - 2^{-1/3}\right)^{-3} y^{-2} \sigma^2 L^8 (l\delta)^{-8} \widetilde{\Delta}_{n,h} h^{-1}.$$

Таким образом, лемма доказана. □

Д о к а з а т е л ь с т в о теоремы 1.12. Положим

$$\zeta_n(h) = \sup_{t\in[0,1]} \big|\widetilde{\nu}_{n,h}(t)\big|, \qquad \eta_n = \sigma L^4 (l\delta)^{-4} \sqrt{\widetilde{\Delta}_{n,h} h^{-1}}$$

и заметим, что в силу леммы 1.28 выполнено $\zeta_n(h) = O_p(\eta_n)$, поскольку

$$\mathbb{P}\big(\zeta_n(h)/\eta_n > M\big) = \mathbb{E}\mathbb{P}_{\mathcal{F}_n}\big(\zeta_n(h)/\eta_n > M\big) \le 2^9 \cdot 23 \left(1 - 2^{-1/3}\right)^{-3} M^{-2}.$$

Утверждение теоремы следует теперь из представления (188) и леммы 1.26. □

1.5 Универсальные оценки для функций среднего и ковариации случайного процесса

В этом разделе мы рассматриваем следующую регрессионную модель.

($\mathbf{M}_2$) *Даны пары наблюдений* $\{(Z_{ij}, X_{ij}),\ i = 1, \dots n,\ j = 1, \dots, m_i\}$, *представимые в виде*

$$X_{ij} = f_i(Z_{ij}) + \varepsilon_{ij}, \qquad i = 1, \dots, n, \quad j = 1, \dots, m_i, \tag{199}$$

где $f_1(\cdot)$, ..., $f_n(\cdot)$ *— неизвестные независимые одинаково распределенные с вероятностью* 1 *непрерывные случайные процессы, заданные на* $[0,1]$, $\{\varepsilon_{ij},\ i = 1, \dots n,\ j = 1, \dots, m_i\}$ *— ненаблюдаемые случайные погрешности. Для каждого фиксированного* $i = 1, \dots, n$ *регрессоры* $\{Z_{ij};\ j = 1, \dots, m_i\}$ *представляют собой набор наблюдаемых случайных величин со значениями в* $[0,1]$, *имеющих, вообще говоря, неизвестные распределения, не обязательно независимых или одинаково распределенных. Случайные величины* $\{Z_{ij};\ j = 1, \dots, m_i\}$ *могут зависеть от* m_i *(если* m_i *неслучайно) и* n.

Введем обозначения

$$\mu(t) = \mathbb{E} f_1(t), \quad \varphi(t,s) = \mathbb{E}\{f_1(t) f_1(s)\}, \quad \psi(t,s) = \varphi(t,s) - \mu(t)\mu(s). \tag{200}$$

Задача этого раздела состоит в оценивании функций среднего $\mu(t)$ и ковариации $\psi(t,s)$ случайного процесса (в предположении их существования) в рамках условия (M_2), т.е. когда мы наблюдаем зашумленные значения каждой из n независимых копий $f_1 \ldots, f_n$ некоторого почти наверное непрерывного случайного процесса в некотором наборе точек (вообще говоря, своем для каждой реализации), состоящем из m_i элементов для i-ой случайной функции f_i. Отметим, что предположение (M_2) включает в себя ситуацию как случайных, так и детерминированных регрессоров $\{Z_{ij}\}$. Отрезок $[0,1]$ в качестве области определения мы рассматриваем исключительно с целью простоты изложения подхода.

В разделе 1.5.1 мы рассмотрим случай разреженных данных (количество наблюдений для каждой серии в том или ином смысле «мало»), а в разделе 1.5.3 — плотных (объем наблюдений в каждой из n серий растет с ростом числа серий).

1.5.1 Случай разреженных данных

В этом разделе относительно ядра сглаживания считаем выполненным условие (K_1), введенное в начале главы (см. раздел 1.1). Нам потребуется ряд дополнительных предположений, так что приведем прежде всего условия на параметры модели (M_2), которые мы будем использовать в тех или иных сочетаниях.

$(\mathbf{E}_3)$ *Ненаблюдаемые случайные погрешности* $\{\varepsilon_{ij};\ i = 1, \ldots, n,\ j = 1, \ldots, m_i\}$ *при всех* i, j, *а также* $(i_1, j_1) \neq (i_2, j_2)$, *с вероятностью* 1 *удовлетворяют следующим условиям*:

$$\mathbb{E}_{\mathcal{F}} \varepsilon_{ij} = 0, \qquad \max_{i,j} \mathbb{E}_{\mathcal{F}} \varepsilon_{ij}^2 \leq \sigma_\varepsilon^2, \qquad \mathbb{E}_{\mathcal{F}} \varepsilon_{i_1 j_1} \varepsilon_{i_2 j_2} = 0,$$

где константа (возможно, неизвестная) $\sigma_\varepsilon^2 > 0$ *не зависит от* n, *символ* $\mathbb{E}_{\mathcal{F}}$ *обозначает условное математическое ожидание при фиксации* σ*-алгебры* $\mathcal{F}$, *порожденной случайными величинами* $\{Z_{ij};\ i = 1, \ldots, n,\ j = 1, \ldots, m_i\}$ *и* $\{m_i, i = 1, \ldots, n\}$.

$(\mathbf{B}_1)$ *Случайные функции* $\{f_i(t)\}$ *не зависят от регрессоров* $\{Z_{ij}\}$, *при этом*

$$\sup_{t \in [0,1]} \mathbb{D} f_1(t) \leq \sigma_f^2 < \infty.$$

$(\mathbf{C}_1)$ *Положительные целочисленные случайные величины* $\{m_i\}$ *(не обязательно независимые или одинаково распределенные) не зависят от* $\{f_i(t)\}$ *и* $\{Z_{ij}\}$, *а также не зависят от* n.

($\mathbf{C}_2$) *Выполнено условие* (C_1) *и при некотором* $\alpha > 3$ *справедливо соотношение*

$$\max_{i \le n} \mathbb{E} m_i^\alpha \le \lambda_\alpha < \infty,$$

где константа λ_α *может быть неизвестна и не зависит от* n.

($\mathbf{C}_2'$) *Выполнено условие* (C_1), *случайные величины* $\{m_i\}$ *являются независимыми копиями некоторой целочисленной случайной величины и при некотором* $\alpha > 3$ *справедливо соотношение* $\mathbb{E} m_1^\alpha < \infty$.

($\mathbf{C}_3$) *Выполнено условие* (C_1), *справедливо равенство* $m_1 = \ldots = m_n$ *и при некотором* $\alpha > 0$ *имеет место ограничение* $\mathbb{E} m_1^\alpha < \infty$.

($\mathbf{C}_4$) Величины $\{m_i\}$ неслучайны и равномерно ограничены: $\max_{i \le n} m_i \le c < \infty$, где константа c не зависит от n.

З а м е ч а н и е 1.33. В случае, когда в исходной выборке в той или иной серии наблюдений i имеются кратные регрессоры, предлагается несколько сократить выборку, заменив наблюдения X_{ij} с одинаковыми точками Z_{ij} их средним арифметическим и оставляя в новой выборке лишь один регрессор из кратных.

Перейдем к построению оценки для функции $\mu(t)$. Пусть выполнено условие (C_1). Положим $N = m_1 + \ldots + m_n$ и по выборке $\{Z_{ij};\ i = 1, \ldots, n,\ j = 1, \ldots, m_i\}$ образуем вариационный ряд, элементы которого обозначим через $Z_{N:1} \le \ldots \le Z_{N:N}$. Положим $Z_{N:0} = 0$, $Z_{N:N+1} = 1$. Пусть $N = lr + s$, где l, r и s — целые, r неслучайно, а случайные величины l и s таковы, что $1 \le s < r$ почти наверное (т.е. s — это остаток от деления N на l). Считаем, что $r = r(n) \to \infty$ и $r = o(n)$. Определим следующие величины:

$$\Delta Z_{Nl} = Z_{N:N+1} - Z_{N:r(l-1)}, \qquad \Delta Z_{Nk} = Z_{N:rk} - Z_{N:r(k-1)}, \quad k = 1, \ldots, l-1. \tag{201}$$

Таким образом, отрезок $[0, 1]$ мы разбили на l попарно несовместных отрезков с длинами $\Delta Z_{N1}, \ldots, \Delta Z_{Nl}$, каждый из которых (за исключением последнего отрезка) содержит по r точек, а последний отрезок длины ΔZ_{Nl} содержит $r + s < 2r$ точек. Отметим, что в введенных условиях количество (возможно, случайное) отрезков $l = l(n) \ge n/r - 1$ должно неограниченно возрастать с ростом n.

Нам также потребуются следующие обозначения:

$$H_1 = \{(i,j):\ Z_{ij} \in [Z_{N:0},\ Z_{N:r}]\},$$
$$H_k = \{(i,j):\ Z_{ij} \in (Z_{N:r(k-1)},\ Z_{N:rk}]\}, \quad k = 2, \ldots, l,$$
$$\overline{X}_l = (r+s)^{-1} \sum_{(i,j) \in H_l} X_{ij}, \quad \overline{X}_k = r^{-1} \sum_{(i,j) \in H_k} X_{ij}, \qquad k = 1, \ldots, l.$$

Отметим, что множества H_k, $k = 1, \ldots, l-1$, содержат ровно по r пар индексов, а множество H_l содержит $r + s < 2r$ пар индексов.

Оценку для $\mu(t)$ определим равенством

$$\widehat{\mu}_1(t) = \frac{\sum_{k=1}^{l} \overline{X}_k K_{h_\mu}(t - Z_{N:rk})\Delta Z_{Nk}}{\sum_{k=1}^{l} K_{h_\mu}(t - Z_{N:rk})\Delta Z_{Nk}}. \tag{202}$$

З а м е ч а н и е 1.34. Поясним идею выбора статистики $\widehat{\mu}_1(t)$ из (202) в качестве оценки для функции среднего $\mu(t)$. Введем обозначения

$$\begin{gathered} \overline{f}_k = r^{-1} \sum_{(i,j)\in H_k} f_i(Z_{ij}), \quad \overline{\varepsilon}_k = r^{-1} \sum_{(i,j)\in H_k} \varepsilon_{ij}, \quad k = 1, \ldots, l-1, \\ \overline{f}_l = (r+s)^{-1} \sum_{(i,j)\in H_l} f_i(Z_{ij}), \qquad \overline{\varepsilon}_l = (r+s)^{-1} \sum_{(i,j)\in H_l} \varepsilon_{ij}. \end{gathered} \tag{203}$$

Из уравнения (199) имеем $\overline{X}_k = \overline{f}_k + \overline{\varepsilon}_k$, $k = 1, \ldots, l$. Но в силу условий (M_2), (B_1), определения множеств H_k, закона больших чисел, а также условия (C_1) или (C_4), можно ожидать, что $\overline{f}_k \approx \mu(t)$, где $t \in (Z_{N:r(k-1)},\ Z_{N:rk}]$ (например, можно положить $t = Z_{N:rk}$). Иными словами,

$$\overline{X}_k \approx \mu\big(Z_{N:rk}\big) + \overline{\varepsilon}_k, \qquad k = 1, \ldots, l.$$

Мы предлагаем оценить функцию $\mu(t)$ в такой модели непараметрической регрессии с помощью метода ядерного сглаживания из раздела 1.1. Это и приводит нас к оценке (202). □

Основное условие на регрессоры, гарантирующее в случае разреженных данных существование равномерно состоятельной оценки для функции среднего, состоит в следующем.

$(\mathbf{D}_7)$ *С ростом n имеет место предельное соотношение*

$$\delta_l = \max_{1\le k\le l} \Delta Z_{Nk} \xrightarrow{p} 0.$$

З а м е ч а н и е 1.35. Другими словами, условие (D_7) означает, что набор точек $\{Z_{ij}\}$ с высокой вероятностью образует измельчающееся разбиение отрезка $[0,1]$. Если $\{Z_{ij}\}$ независимы и одинаково распределены, а отрезок $[0,1]$ является носителем распределения, то условие (D_7) выполнено. В частности, в случае существования отделенной от нуля на $[0,1]$ плотности распределения Z_{11} и неслучайном l, с вероятностью 1 справедливо соотношение $\delta_l = O\left(\log l/l\right)$. Если последовательность бесконечномерных векторов $\{m_i, Z_{i1}, Z_{i2}, \ldots\}$, $i = 1, 2, \ldots$, удовлетворяет условию α-перемешивания, причем для любого фиксированного i все конечномерные распределения последовательности $\{Z_{ij};\ j \ge 1\}$ имеют строго положительные плотности, а сама эта последовательность не зависит от m_i, то условие (D_7) также будет выполнено. Но выполнение условия (D_7) вполне возможно и для других типов зависимости, которая может быть более сильной, нежели те или иные условия слабой зависимости. Соответствующие примеры и обсуждения такого рода условий содержатся в вышеприведенных разделах диссертации (например, см. раздел 1.1, примеры 1.1, 1.2). $\square$

Теорема 1.13. *Пусть выполнены условия* (M_2), (E_3), (B_1), (C_1) *и* (K_1). *Тогда для любого фиксированного* $h_\mu \in (0, 1/2)$ *с вероятностью* 1

$$\sup_{t\in[0,1]} |\widehat{\mu}_1(t) - \mu(t)| \leq \omega_\mu(h_\mu) + \omega_\mu(\delta_l) + \zeta_{l,r,h_\mu} + \eta_{l,r} \tag{204}$$

и случайные величины ζ_{l,r,h_μ} *и* $\eta_{l,r}$ *таковы, что*

$$\mathbb{P}\big(\zeta_{l,r,h_\mu} > y,\, \delta_l \leq h_\mu/(8L)\big) \leq C\sigma_\varepsilon^2 L^2 y^{-2} {h_\mu}^{-2} r^{-1}\mathbb{E}\delta_l,$$
$$\mathbb{P}(\eta_{l,r} > y) \leq 2\sigma_f^2 M^3 n r^{-2} y^{-2} + \mathbb{P}\Big(\max_{i\leq n} m_i > M\Big),$$

где C *— абсолютная положительная постоянная, а* M *— произвольное положительное число. Если дополнительно при некотором* $\alpha > 0$ *и всех* $i \leq n$ *выполнено* $\mathbb{E}m_i^\alpha < \infty$, *то*

$$\mathbb{P}(\eta_{l,r} > y) \leq 2(2\sigma_f^2)^{\frac{\alpha}{3+\alpha}} \lambda_{\alpha,n}^{\frac{3}{3+\alpha}} n^{\frac{\alpha}{3+\alpha}} (ry)^{-\frac{2\alpha}{3+\alpha}}, \qquad \text{где } \lambda_{\alpha,n} = \mathbb{E}\big(\max_{i\leq n} m_i^\alpha\big).$$

З а м е ч а н и е 1.36. Условие (C_1) теоремы 1.13 универсально относительно стохастической природы количества наблюдений в сериях, а также их корреляции, и включает в себя оба предположения, используемых в литературе в случае разреженных данных: либо $\{m_i\}$ случайны и являются независимыми копиями целочисленной случайной величины, либо неслучайны и равномерно ограничены (т.е. выполнено (C_4)). Отметим, что во всех известных нам работах, связанных с разреженными данными, рассматривается только какой-либо один из указанных вариантов. Приводимые далее элементарные следствия теоремы 1.13, содержащие простые достаточные условия равномерной состоятельности оценки для среднего, включают в себя указанные предположения.

З а м е ч а н и е 1.37. Отметим, что $\delta_l \leq 1$, а потому при выполнении условия (D_7) имеет место предельное соотношение $\mathbb{E}\delta_l \to 0$. Кроме того, справедливо соотношение

$$\zeta_{l,r,h_\mu} = \widetilde{O}_p\left(h_\mu^{-1}\big(r^{-1}\mathbb{E}\delta_l\big)\right)^{1/2}\Big) + O\big(h_\mu^{-1}\mathbb{E}\delta_l\big). \tag{205}$$

Далее, если выполнено (C_2), то в силу очевидной оценки $\max_{i\leq n} m_i^\alpha \leq \sum_{i=1}^n m_i^\alpha$ получаем, что $\lambda_{\alpha,n} \leq n\lambda_\alpha$, а потому

$$\mathbb{P}(\eta_{l,r} > y) \leq 2(2\sigma_f^2)^{\frac{\alpha}{3+\alpha}} \lambda_\alpha^{\frac{3}{3+\alpha}}\ n(ry)^{-\frac{2\alpha}{3+\alpha}}.$$

Если выполнено (C_3) или (C_4), то соответственно

$$\mathbb{P}(\eta_{l,r} > y) \le 2(2\sigma_f^2)^{\frac{\alpha}{3+\alpha}} \left(\mathbb{E}m_1^\alpha\right)^{\frac{3}{3+\alpha}} n^{\frac{\alpha}{3+\alpha}} (ry)^{-\frac{2\alpha}{3+\alpha}}, \qquad \mathbb{P}(\eta_{l,r} > y) \le 2\sigma_f^2 c^3 n r^{-2} y^{-2}.$$

Таким образом, если выполнено (C_2), то $\eta_{l,r} = \widetilde{O}_p\left(n^{(3+\alpha)/(2\alpha)} r^{-1}\right)$, а если справедливо условие (C_3) или (C_4), то $\eta_{l,r} = \widetilde{O}_p\left(\sqrt{n}/r\right)$.

Из теоремы 1.13 и замечания 1.37 получаем следующие утверждения.

Следствие 1.14. *Пусть выполнены условия* (M_2), (E_3), (B_1), (C_2), (D_7), (K_1) *и*

$$h_\mu \to 0, \quad h_\mu^{-2} r^{-1} \mathbb{E}\delta_l \to 0, \quad \mathbb{P}(\delta_l > h_\mu/(8L)) \to 0, \quad n^{(3+\alpha)/(2\alpha)} r^{-1} \to 0. \tag{206}$$

Тогда

$$\sup_{t \in [0,1]} |\widehat{\mu}_1(t) - \mu(t)| \xrightarrow{p} 0. \tag{207}$$

Следствие 1.15. *Пусть выполнены условия* (M_2), (E_3), (B_1), (C_2'), (D_7), (K_1) *и соотношения* (206). *Тогда имеет место сходимость* (207).

Следствие 1.16. *Пусть выполнены условия* (M_2), (E_3), (B_1), (C_3), (D_7), (K_1), *а также справедливы первые три соотношения в* (206) *и* $\sqrt{n}/r \to 0$. *Тогда имеет место сходимость* (207).

Следствие 1.17. *Пусть выполнены условия* (M_2), (E_3), (B_1), (C_4), (D_7), (K_1), *справедливы первые три соотношения в* (206) *и* $\sqrt{n}/r \to 0$. *Тогда имеет место сходимость* (207).

З а м е ч а н и е 1.38. В условиях следствия 1.14 количество точек $r = r(n) \to \infty$ нужно выбирать таким образом, чтобы (среди прочих) выполнялось соотношение $n^{(3+\alpha)/(2\alpha)} r^{-1} \to 0$. Отметим, что в рамках следствия 1.14 мы имеем $(3+\alpha)/(2\alpha) < 1$ при $\alpha > 3$. В условиях одного из следствий 1.16 или 1.17 количество точек $r = r(n) \to \infty$ нужно выбирать так, чтобы $\sqrt{n}/r \to 0$. Понятно, что с учетом условия $r = o(n)$ количество отрезков $l = l(n) \ge N/r - 1 \ge n/r - 1$ также должно неограниченно возрастать с ростом n.

П р и м е р 1.21. Пусть $r = n^{1/2+\varepsilon}$ при некотором $\varepsilon > 0$ и $\mathbb{E}\delta_l = O(r/n)$. Последнее условие выполнено, например, если вся совокупность временных точек образует равномерную решетку на $[0,1]$, или если набор $\{Z_{ij}\}$ состоит из независимых и одинаково распределенных случайных величин с отделенной от нуля плотностью на $[0,1]$. Предположим, что функция $\mu(t)$ удовлетворяет условию Гельдера, т.е. $\omega_\mu(h) \le C h^\gamma$ при всех $h > 0$ и некоторых фиксированных $\gamma \in (0,1]$ и $C > 0$. В этом случае величина $h_\mu = n^{-\frac{1}{2(\gamma+1)}}$ уравнивает по h_μ порядок малости слагаемого $\omega_\mu(h_\mu)$ и первой компоненты слагаемого ζ_{l,r,h_μ} (см. формулу (205)) в правой части соотношения (204), зависящих от размера окна h_μ. В сделанных предположениях

$\mathbb{E}\delta_l/h_\mu = O\Big(n^{\frac{\gamma-2\varepsilon(1+\gamma)}{2(\gamma+1)}}\Big)$, так что третье условие в (206) выполнено при $0<\varepsilon<\gamma/(2+2\gamma)$. При этом для выполнения четвертого условия в (206) нужно, чтобы $\alpha>3/(2\varepsilon)$.

Перейдем теперь к оцениванию функции $\varphi(t,s)$. Для простоты и определенности будем считать, что выполнено (C_4), при этом $m_i\ge 2$ при всех i. Нам потребуются ряд дополнительных ограничений. Положим $\widetilde{N}=\widetilde{N}(n)=m_1(m_1-1)+\ldots+m_n(m_n-1)$. Без ограничения общности считаем, что $\widetilde{N}=\tilde{l}\times\tilde{r}$, где $\tilde{l}$ и $\tilde{r}$ — целые, при этом $\tilde{l}=\tilde{l}(n)\to\infty$ $\tilde{r}=\tilde{r}(n)\to\infty$ при $n\to\infty$. Для произвольного ограниченного множества $A\subset\mathbb{R}^2$ определим его диаметр следующим равенством: $d(A)=\sup_{\mathbf{x},\mathbf{y}\in A}\|\mathbf{x}-\mathbf{y}\|$, где $\|\cdot\|$ — супремальная норма в $\mathbb{R}^2$.

Помимо (C_4), предполагаются выполненными следующие три условия.

$(\mathbf{D}_8)$ *Все двумерные точки из набора* $\{(Z_{ij_1},Z_{ij_2}),1\le j_1\ne j_2\le m_i, i=1,\ldots,n\}$ *попарно различны и для каждого* $\widetilde{N}$ *существует случайное разбиение множества* $[0,1]^2$ *на* $\tilde{l}$ *измеримых по Жордану подмножеств* $\{\mathcal{P}_k;\ k=1,\ldots,\tilde{l}\}$ *таких, что каждое подмножество* $\mathcal{P}_k$ *содержит ровно по* $\tilde{r}$ *двумерных точек из указанного набора, при этом* $\widetilde{\delta}_{\tilde{l}}=\max_{k\le\tilde{l}}d(\mathcal{P}_k)\xrightarrow{p}0$.

$(\mathbf{E}_4)$ *Погрешности* $\{\varepsilon_{ij};\ i=1,\ldots,n,\ j=1,\ldots,m_i\}$ *— независимые одинаково распределенные случайные величины, удовлетворяющие моментным ограничениям*

$$\mathbb{E}\varepsilon_{11}=0,\quad \lambda_p=\mathbb{E}|\varepsilon_{11}|^p<\infty\quad \textit{при некотором } p>2,$$

где константа λ_p *может быть неизвестна и не зависит от* n. *Погрешности* $\{\varepsilon_{ij}\}$ *не зависят от* $\{Z_{ij}\}$ *и* $\{f_i(\cdot)\}$.

$(\mathbf{B}_2)$ *Выполнено условие* (B_1) *и для* p *из условия* (E_4)

$$\gamma_p=\mathbb{E}\sup_{t\in[0,1]}|f_1(t)|^p<\infty,\quad \widetilde{\sigma}_f^2=\sup_{t\in[0,1]}\mathbb{D}f_1^2(t),$$

где константы (возможно, неизвестные) $\gamma_p>0$ *и* $\widetilde{\sigma}_f^2>0$ *не зависят от* n.

Введем обозначения

$$\widetilde{H}_k=\left\{\big((i,j_1),(i,j_2)\big):\ \big(Z_{ij_1},Z_{ij_2}\big)\in\mathcal{P}_k\right\},$$
$$\widetilde{X}_k=\tilde{r}^{-1}\sum_{((i,j_1),(i,j_2))\in\widetilde{H}_k}X_{ij_1}X_{ij_2},\qquad k=1,\ldots,\tilde{l}.$$

Отметим, что множества $\widetilde{H}_k$, $k=1,\ldots,\tilde{l}$, содержат ровно по $\tilde{r}$ элементов. Выберем некоторую точку, принадлежащую множеству $\mathcal{P}_k$, и обозначим ее через $\big(Z_{i_kj_{1k}},Z_{i_kj_{2k}}\big)$. Оценку для $\varphi(t,s)$ определим равенством

$$\widehat{\varphi}_1(t,s)=\frac{\sum_{k=1}^{\tilde{l}}\widetilde{X}_kK_{h_\varphi}(t-Z_{i_kj_{1k}})K_{h_\varphi}(s-Z_{i_kj_{2k}})\Lambda_2(\mathcal{P}_k)}{\sum_{k=1}^{\tilde{l}}K_{h_\varphi}(t-Z_{i_kj_{1k}})K_{h_\varphi}(s-Z_{i_kj_{2k}})\Lambda_2(\mathcal{P}_k)}.\tag{208}$$

З а м е ч а н и е 1.39. Поясним идею построения оценки $\widehat{\varphi}_1(t,s)$ из (208). Положим

$$\begin{aligned}
\widetilde{f}_k &= \tilde{r}^{-1} \sum_{((i,j_1),(i,j_2))\in\widetilde{H}_k} f_i(Z_{ij_1})f_i(Z_{ij_2}),\\
\widetilde{\varepsilon}_k &= \tilde{r}^{-1} \sum_{((i,j_1),(i,j_2))\in\widetilde{H}_k} \left(f_i(Z_{ij_1})\varepsilon_{ij_2} + f_i(Z_{ij_2})\varepsilon_{ij_1} + \varepsilon_{ij_1}\varepsilon_{ij_2}\right), \quad k=1,\ldots,\tilde{l}.
\end{aligned} \tag{209}$$

Из уравнения (199) имеем

$$\widetilde{X}_k = \widetilde{f}_k + \widetilde{\varepsilon}_k, \qquad k=1,\ldots,\tilde{l}.$$

Но в силу условий (M_2) и (B_2), а также определения множества $\widetilde{H}_k$ и закона больших чисел можно ожидать, что $\widetilde{f}_k \approx \varphi(t,s)$, где $(t,s)\in\mathcal{P}_k$ (например, можно положить $(t,s) = (Z_{i_kj_{1k}}, Z_{i_kj_{2k}})$). Таким образом, мы получаем модель непараметрической регрессии

$$\widetilde{X}_k \approx \varphi\big(Z_{i_kj_{1k}}, Z_{i_kj_{2k}}\big) + \widetilde{\varepsilon}_k, \qquad k=1,\ldots,\tilde{l}.$$

Оценка (208) — это универсальная локально–постоянная оценка (см. раздел 1.3) для полученной двухфакторной регрессионной модели. $\square$

З а м е ч а н и е 1.40. Нетрудно видеть, что условие (D_8) влечет предположение о том, что совокупность временных точек по всем сериям с высокой вероятностью образует измельчающееся разбиение области определения исходного случайного процесса. Обратное неверно, и соответствующий пример несложно построить. Например, пусть $m_i = 2$ при всех i и парные точки, участвующие в условии (D_8), сосредоточены на некоторой кривой таким образом, что проекции этих точек на ось абсцисс, собственно и представляющие собор совокупность временных точек из всех серий, образуют измельчающееся разбиение.

Теорема 1.14. *Пусть выполнены условия (M_2), (D_8), (E_4), (C_4), (B_2) и (K_1), при этом $m_i \ge 2$ при всех i. Тогда для любого фиксированного $h_\varphi \in (0,1/2)$ с вероятностью 1*

$$\sup_{(t,s)\in[0,1]^2} |\widehat{\varphi}_1(t,s) - \varphi(t,s)| \le \omega_\varphi(h_\varphi) + \omega_\varphi(\widetilde{\delta}_{\tilde{l}}) + \widetilde{\zeta}_{\tilde{l},\tilde{r},h_\varphi} + \widetilde{\eta}_{\tilde{l},\tilde{r}}, \tag{210}$$

где случайные величины $\widetilde{\eta}_{\tilde{l},\tilde{r}}$ и $\widetilde{\zeta}_{\tilde{l},\tilde{r},h_\varphi}$ таковы, что

$$\begin{aligned}
\mathbb{P}(\widetilde{\eta}_{\tilde{l},\tilde{r}} > y) &\le \widetilde{\sigma}_f^2(2c^4+1)\tilde{l}\tilde{r}^{-1}y^{-2},\\
\mathbb{P}\big(\widetilde{\zeta}_{\tilde{l},\tilde{r},h_\varphi} > y\big) &\le \frac{\widetilde{C}L^{2p}}{\rho^p y^p \tilde{r}^{p/2}} \cdot \frac{\mathbb{E}\widetilde{\delta}_{\tilde{l}}^p}{h_\varphi^{p+2}} + \mathbb{P}\big(\widetilde{\delta}_{\tilde{l}} > h_\varphi/(8L)^2\big)
\end{aligned} \tag{211}$$

и $\widetilde{C}$ – положительная константа, зависящая от p, λ_p, γ_p и $\widetilde{\sigma}_\varepsilon^2 = \mathbb{E}\varepsilon_{11}^2$.

З а м е ч а н и е 1.41. Положим в (211) $y = \Big(h_\varphi^{-(p+2)}\,\mathbb{E}(\widetilde{\delta}_{\tilde{l}}^p)\Big)^{1/p}$. Применяя степенное неравенство Маркова с показателем p для второго слагаемого в (211), нетрудно видеть, что при выполнении условий теоремы 1.14

$$\widetilde{\zeta}_{\tilde{l},\tilde{r},h_\varphi} = \widetilde{O}_p\left(\left(\tilde{r}^{-p/2}h_\varphi^{-(p+2)}\,\mathbb{E}\widetilde{\delta}_{\tilde{l}}^p\right)^{1/p}\right) + O\big(h_\varphi^{-1}\,\mathbb{E}\widetilde{\delta}_{\tilde{l}}\big), \qquad \widetilde{\eta}_{\tilde{l},\tilde{r}} = \widetilde{O}_p((\tilde{l}/\tilde{r})^{1/2}).$$

Следствие 1.18. *Пусть выполнены условия* (M_2), (D_8), (E_4), (B_2), (K_1) *и*

$$h_\varphi \to 0, \quad \tilde{r}^{-p/2}h_\varphi^{-(p+2)}\,\mathbb{E}\widetilde{\delta}_{\tilde{l}}^p \to 0, \quad \tilde{l}/\tilde{r} \to 0, \quad \mathbb{P}\big(\widetilde{\delta}_{\tilde{l}} > h_\varphi/(8L)^2\big) \to 0.$$

Тогда имеет место предельное соотношение

$$\sup_{(t,s)\in[0,1]^2} |\widehat{\varphi}_1(t,s) - \varphi(t,s)| \xrightarrow{p} 0.$$

Оценку для функции ковариации можно определить теперь равенством

$$\widehat{\psi}_1(t,s) = \widehat{\varphi}_1(t,s) - \widehat{\mu}_1(t)\widehat{\mu}_1(s).$$

Условия равномерной состоятельности этой оценки можно получить из вышеприведенных утверждений. Подробности мы опускаем.

1.5.2 Доказательство результатов раздела 1.5.1

Для вывода теоремы 1.13 нам потребуются следующие обозначения:

$$\begin{aligned}
J_{l,h_\mu}(t) &= \sum_{k=1}^{l} K_{h_\mu}(t - Z_{N:rk})\Delta Z_{Nk},\\
\nu_{l,r,h_\mu}(t) &= J_{l,h_\mu}^{-1}(t)\sum_{k=1}^{l} K_{h_\mu}(t - Z_{N:rk})\Delta Z_{Nk}\overline{\varepsilon}_k,\\
\chi_{l,h_\mu}(t) &= J_{l,h_\mu}^{-1}(t)\sum_{k=1}^{l} \big(\mu(Z_{N:rk}) - \mu(t)\big)K_{h_\mu}(t - Z_{N:rk})\Delta Z_{Nk},\\
\tau_{l,h_\mu}(t) &= J_{l,h_\mu}^{-1}(t)\sum_{k=1}^{l} \big(\overline{f}_k - \mu(Z_{N:rk})\big)K_{h_\mu}(t - Z_{N:rk})\Delta Z_{Nk}.
\end{aligned} \tag{212}$$

Подчеркнем, что ввиду свойств плотности $K_h(\cdot)$ область суммирования во введенных величинах совпадает с множеством $\{k:\ |t - Z_{N:rk}| \le h_\mu, 1 \le k \le l\}$. Данное обстоятельство мы неоднократно будем использовать в дальнейших рассуждениях. Имеем

$$\widehat{\mu}_1(t) = \mu(t) + \chi_{l,h_\mu}(t) + \tau_{l,h_\mu}(t) + \nu_{l,r,h_\mu}(t). \tag{213}$$

Лемма 1.29. *В условиях теоремы* 1.13 *имеют место оценки*

$$\sup_{t\in[0,1]}\left|\chi_{l,h_\mu}(t)\right|\le\omega_\mu(h_\mu),\quad \sup_{t\in[0,1]}\left|\tau_{l,h}(t)\right|\le\omega_\mu(\delta_l)+\eta_{l,r},$$

где $\mathbb{P}(\eta_{l,r}>y)\le 2\sigma_f^2 nM^3r^{-2}y^{-2}+\mathbb{P}\Big(\max\limits_{i=1,\dots,n} m_i>M\Big)$.

Д о к а з а т е л ь с т в о. Первое утверждение леммы очевидно. Для вывода второго соотношения заметим прежде всего, что ввиду (212) выполнено

$$\sup_{t\in[0,1]}|\tau_{l,h}(t)|\le\max_{1\le k\le l}|\overline{f}_k-\mu(Z_{N:rk})|.$$

Пусть символы $\mathbb{D}_\mathcal{F}$, $\mathbb{C}ov_\mathcal{F}$ и $\mathbb{P}_\mathcal{F}$ обозначают соответственно условную дисперсию, условную ковариацию и условную вероятность при фиксации σ-алгебры $\mathcal{F}$, порожденной случайными величинами из наборов $\{Z_{ij};\ i=1,\dots,n,\ j=1,\dots,m_i\}$ и $\{m_i;\ i=1,\dots,n\}$. В силу определений (203) и равенства $\mathbb{E}_\mathcal{F}f_1(Z_{ij})=\mu(Z_{ij})$ имеем

$$\overline{f}_k-\mu(Z_{N:rk})=\rho_{1k}+\rho_{2k},\quad k=1,\dots,l,$$

где

$$\rho_{1k}=r^{-1}\sum_{(i,j)\in H_k}\big(f_i(Z_{ij})-\mathbb{E}_\mathcal{F}f_1(Z_{ij})\big),\qquad \rho_{2k}=r^{-1}\sum_{(i,j)\in H_k}\mu(Z_{ij})-\mu(Z_{N:rk}),\ k\le l-1,$$
$$\rho_{1l}=(r+s)^{-1}\sum_{(i,j)\in H_l}\big(f_i(Z_{ij})-\mathbb{E}_\mathcal{F}f_1(Z_{ij})\big),\quad \rho_{2l}=(r+s)^{-1}\sum_{(i,j)\in H_l}\mu(Z_{ij})-\mu(Z_{N:rl}).$$

С учетом определения множеств H_k, $k=1,\dots,l$, и условия (D_7) получаем, что

$$\max_{1\le k\le l}|\rho_{2k}|\le\max_{1\le k\le l}\omega_\mu\big(\Delta Z_{Nk}\big)\le\omega_\mu(\delta_l),\qquad k=1,\dots,l.$$

Введем событие $A_n=\bigcap_{i=1}^n\{m_i<M\}$ при некотором $M>0$. Имеем

$$\mathbb{P}\left(\overline{A}_n\right)=\mathbb{P}\left(\bigcup_{i=1}^n\{m_i\ge M\}\right)=\mathbb{P}\left(\max_{i=1,\dots,n}m_i\ge M\right).$$

Далее, поскольку $\sum_{i=1}^n m_i=rl+s$, то на множестве A_n выполнено $l<nM/r$

$$\mathbb{P}\left(\max_{1\le k\le l}|\rho_{1k}|>y,\ A_n\right)\le\mathbb{P}\left(\max_{1\le k<nM/r}|\rho_{1k}|>y,\ A_n\right)\le$$
$$\le nMr^{-1}\max_{1\le k<nM/r}\mathbb{E}\mathbb{P}_\mathcal{F}\left(\left|\overline{\rho}_{1k}\right|>y\right)\le nMy^{-2}r^{-1}\max_{1\le k<nM/r}\mathbb{E}\mathbb{D}_\mathcal{F}\overline{\rho}_{1k},$$

где $\overline{\rho}_{1k} = \rho_{1k} I(A_n)$ и $I(\cdot)$ — индикатор события.

Далее, справедливо представление

$$r^2 \mathbb{D}_{\mathcal{F}} \overline{\rho}_{1k} = I(A_n) \sum_{(i,j)\in H_k} \mathbb{D}_{\mathcal{F}} f_1(Z_{ij}) + I(A_n) \sum_{(i,j)\neq(i_1,j_1)\in H_k} \mathbb{C}ov_{\mathcal{F}} \left\{ f_i(Z_{ij}), f_{i_1}(Z_{i_1 j_1}) \right\}. \qquad (214)$$

С учетом условия (B_1) для любых $j, j_1 \le m_i$ имеем

$$\mathbb{D}_{\mathcal{F}} f_1(Z_{ij}) \le \sigma_f^2, \quad \mathbb{C}ov_{\mathcal{F}} \left\{ f_i(Z_{ij}), f_{i_1}(Z_{i_1 j_1}) \right\} \le \sigma_f^2.$$

Кроме того, если $i \neq i_1$ во второй сумме в (214), то соответствующие ковариации равны нулю. Если же $i = i_1$, то количество пар индексов (i, j) и (i, j_1) в двойной сумме в (214) при выполнении события A_n будет меньше, чем $m_i^2 \le M^2$. Если же произойдет дополнительное событие $\overline{A}_n$, то все члены в (214) обратятся в ноль. Следовательно, с вероятностью 1

$$r^2 \mathbb{D}_{\mathcal{F}} \overline{\rho}_{1k} \le r\sigma_f^2 + rM^2\sigma_f^2, \qquad \mathbb{D}_{\mathcal{F}} \overline{\rho}_{1k} \le 2r^{-1}M^2\sigma_f^2.$$

Аналогично, для $k = l \le [nM/r]$ при выполнении события A_n имеем

$$(r+s)^2 \mathbb{D}_{\mathcal{F}} \overline{\rho}_{1l} = \sum_{(i,j)\in H_l} \mathbb{D}_{\mathcal{F}} f_1(Z_{ij}) + \sum_{(i,j)\neq(i_1,j_1)\in H_l} \mathbb{C}ov_{\mathcal{F}} \left\{ f_i(Z_{ij}), f_{i_1}(Z_{i_1 j_1}) \right\},$$

$$(r+s)^2 \mathbb{D}_{\mathcal{F}} \overline{\rho}_{1k} \le 2(r+s)\sigma_f^2 + (r+s)M^2\sigma_f^2, \qquad \mathbb{D}_{\mathcal{F}} \overline{\rho}_{1k} \le 2r^{-1}M^2\sigma_f^2.$$

Таким образом,

$$\mathbb{P}\left(\max_{1\le k\le l} |\rho_{1k}| > y,\ A_n \right) \le 2\sigma_f^2 n M^3 r^{-2} y^{-2},$$

а потому

$$\mathbb{P}\left(\max_{1\le k\le l} |\rho_{1k}| > y \right) \le 2\sigma_f^2 n M^3 r^{-2} y^{-2} + \mathbb{P}\Big(\max_{i=1,\dots,n} m_i > M \Big).$$

Для завершения доказательства леммы остается положить $\eta_{l,r} = \max\limits_{1\le k\le l} |\rho_{1k}|$. □

Лемма 1.30. *Для любых $y > 0$ и $h_\mu \in (0, 1/2)$ на подмножестве элементарных исходов, определяемых соотношением $\delta_l \le h_\mu/(8L)$, имеет место следующая оценка:*

$$\mathbb{P}_{\mathcal{F}}\left(\sup_{t\in[0,1]} |\nu_{l,r,h_\mu}(t)| > y \right) \le C\sigma_\varepsilon^2 L^2 r^{-1} \delta_l h_\mu^{-2} y^{-2},$$

где C — абсолютная положительная константа, а символ $\mathbb{P}_{\mathcal{F}}$ обозначает условную вероятность при фиксации σ-алгебры $\mathcal{F}$, определенной в условии (E_3).

Д о к а з а т е л ь с т в о. Для любых $k \le l$ и $u, v \le l$, $u \ne v$, имеем

$$\mathbb{E}_{\mathcal{F}} \bar{\varepsilon}_k = 0, \quad \sup_{k \le l} \mathbb{E}_{\mathcal{F}} \bar{\varepsilon}_k^2 \le r^{-1} \sigma_\varepsilon^2, \quad \mathbb{E}_{\mathcal{F}} \bar{\varepsilon}_u \bar{\varepsilon}_v = 0.$$

Здесь мы учли, что $1/(r+s) \le 1/r$. Таким образом, доказательство этого утверждения с очевидными изменениями повторяет вывод леммы 1.3 из раздела 1.1. $\square$

Д о к а з а т е л ь с т в о теоремы 1.13. Положим $\zeta_{l,r,h_\mu} = \sup_{t \in [0,1]} |\nu_{l,r,h_\mu}(t)|$ и заметим, что

$$\mathbb{P}\big(\zeta_{l,r,h_\mu} > y,\, \delta_l \le h_\mu/(8L)\big) = \mathbb{E} I\big(\delta_l \le h_\mu/(8L)\big) \mathbb{P}_{\mathcal{F}}\big(\zeta_{l,r,h_\mu} > y\big).$$

Первое утверждение теоремы следует теперь из тождества (213) и лемм 1.29, 1.30.

В силу оценки $m_1^\alpha \le \max_{i \le n} m_i^\alpha \le \sum_i^n m_i^\alpha$ предположение о том, что при некотором α и всех $i \le n$ выполнено $\mathbb{E} m_i^\alpha < \infty$, эквивалентно соотношению $\lambda_{\alpha,n} \equiv \mathbb{E}\big(\max_{i \le n}\big) m_i^\alpha < \infty$. Поэтому с учетом неравенства Маркова

$$\mathbb{P}\left(\eta_{l,r} > y\right) \le 2\sigma_f^2 n M^3 r^{-2} y^{-2} + \lambda_{\alpha,n}\, M^{-\alpha}.$$

Приравнивая теперь оба слагаемых в правой части этого соотношения, находим оптимальный уровень «срезки», равный $M = \big(r^2 y^2 \lambda_{\alpha,n}/(2\sigma_f^2 n)\big)^{1/(3+\alpha)}$. В итоге

$$\mathbb{P}\left(\eta_{l,r} > y\right) \le 2(2\sigma_f^2)^{\frac{\alpha}{3+\alpha}} \lambda_{\alpha,n}^{\frac{3}{3+\alpha}} n^{\frac{\alpha}{3+\alpha}} (ry)^{-\frac{2\alpha}{3+\alpha}}.$$

Теорема доказана. $\square$

Для доказательства теоремы 1.14 нам понадобятся дополнительные определения и ряд вспомогательных утверждений. Положим

$$\mathbf{u} = (t, s), \quad \mathbf{z}_k = (Z_{i_k j_{1k}}, Z_{i_k j_{2k}}), \quad \widetilde{K}_{h_\varphi}(\mathbf{u} - \mathbf{z}_k) = K_{h_\varphi}(t - Z_{i_k j_{1k}}) K_{h_\varphi}(s - Z_{i_k j_{2k}})$$

$$\begin{aligned}
\widetilde{J}_{\tilde{l},h_\varphi}(\mathbf{u}) &= \sum_{k=1}^{\tilde{l}} \widetilde{K}_{h_\varphi}(\mathbf{u} - \mathbf{z}_k) \Lambda_2(\mathcal{P}_k), \\
\widetilde{\nu}_{\tilde{l},\tilde{r},h_\varphi}(\mathbf{u}) &= \widetilde{J}_{\tilde{l},h_\varphi}^{-1}(\mathbf{u}) \sum_{k=1}^{\tilde{l}} \widetilde{K}_{h_\varphi}(\mathbf{u} - \mathbf{z}_k) \Lambda_2(\mathcal{P}_k) \widetilde{\varepsilon}_k, \\
\widetilde{\chi}_{\tilde{l},h_\varphi}(\mathbf{u}) &= \widetilde{J}_{\tilde{l},h_\varphi}^{-1}(\mathbf{u}) \sum_{k=1}^{\tilde{l}} \big(\varphi(\mathbf{z}_k) - \varphi(\mathbf{u})\big) \widetilde{K}_{h_\varphi}(\mathbf{u} - \mathbf{z}_k) \Lambda_2(\mathcal{P}_k), \\
\widetilde{\tau}_{\tilde{l},h_\varphi}(\mathbf{u}) &= \widetilde{J}_{\tilde{l},h_\varphi}^{-1}(\mathbf{u}) \sum_{k=1}^{\tilde{l}} \big(\widetilde{f}_k - \varphi(\mathbf{z}_k)\big) \widetilde{K}_{h_\varphi}(\mathbf{u} - \mathbf{z}_k) \Lambda_2(\mathcal{P}_k).
\end{aligned} \tag{215}$$

Заметим, что в силу свойств ядра K область суммирования в четырех вышеуказанных суммах есть $\{k : \|\mathbf{u}-\mathbf{z}_k\| \le h_\varphi\}$. С учетом введенных обозначений имеет место следующее представление, являющееся ключевым при выводе теоремы 1.14:

$$\widehat{\varphi}_1(\mathbf{u}) = \varphi(\mathbf{u}) + \widetilde{\chi}_{\tilde{l},h_\varphi}(\mathbf{u}) + \widetilde{\tau}_{\tilde{l},h_\varphi}(\mathbf{u}) + \widetilde{\nu}_{\tilde{l},\tilde{r},h_\varphi}(\mathbf{u}). \tag{216}$$

Лемма 1.31. *В условиях теоремы* 1.14 *имеют место оценки*

$$\sup_{\mathbf{u}\in[0,1]^2}\left|\widetilde{\chi}_{\tilde{l},h_\varphi}(\mathbf{u})\right| \le \omega_\varphi(h_\varphi), \qquad \sup_{\mathbf{u}\in[0,1]^2}\left|\widetilde{\tau}_{\tilde{l},h_\varphi}(\mathbf{u})\right| \le \omega_\varphi(\widetilde{\delta}_{\tilde{l}}) + \widetilde{\eta}_{\tilde{l},\tilde{r}},$$

где $\mathbb{P}(\widetilde{\eta}_{\tilde{l},\tilde{r}} > y) \le \widetilde{\sigma}_f^2(2c^4+1)\tilde{l}\tilde{r}^{-1}y^{-2}$.

Д о к а з а т е л ь с т в о. Первое утверждение леммы вытекает из определения (215) и вышеприведенного замечания. Аналогично выводу леммы 1.29, имеем

$$\sup_{\mathbf{u}\in[0,1]^2}\left|\widetilde{\tau}_{\tilde{l},h_\varphi}(\mathbf{u})\right| \le \max_{1\le k\le \tilde{l}}|\widetilde{f}_k - \varphi(\mathbf{z}_k)|.$$

В силу (209) и равенства $\mathbb{E}_{\mathcal{F}} f_1(Z_{ij_1})f_1(Z_{ij_2}) = \varphi(Z_{ij_1,Z_{ij_2}})$ справедливо представление

$$\widetilde{f}_k - \varphi(\mathbf{z}_k) = \widetilde{\rho}_{1k} + \widetilde{\rho}_{2k},$$

$$\widetilde{\rho}_{1k} = \tilde{r}^{-1}\sum_{((i,j_1),(i,j_2))\in\widetilde{H}_k}\big(f_i(Z_{ij_1})f_i(Z_{ij_2}) - \mathbb{E}_{\mathcal{F}} f_1(Z_{ij_1})f_1(Z_{ij_2})\big),$$
$$\widetilde{\rho}_{2k} = \tilde{r}^{-1}\sum_{((i,j_1),(i,j_2))\in\widetilde{H}_k}\varphi(Z_{ij_1},Z_{ij_2}) - \varphi(\mathbf{z}_k).$$

Далее, с учетом определения множества $\widetilde{H}_k$ и условия (D_8) получаем, что

$$\max_{1\le k\le \tilde{l}}|\widetilde{\rho}_{2k}| \le \max_{1\le k\le \tilde{l}}\omega_\varphi\big(d(\mathcal{P}_k)\big) \le \omega_\varphi\big(\widetilde{\delta}_{\tilde{l}}\big).$$

Кроме того,

$$\mathbb{P}\left(\max_{1\le k\le \tilde{l}}|\widetilde{\rho}_{1k}| > y\right) \le \tilde{l}\max_{1\le k\le \tilde{l}}\mathbb{E}\mathbb{P}_{\mathcal{F}}(|\widetilde{\rho}_{1k}|>y) \le \tilde{l}y^{-2}\max_{1\le k\le \tilde{l}}\mathbb{E}\mathbb{D}_{\mathcal{F}}\widetilde{\rho}_{1k},$$

при этом

$$\tilde{r}^2\mathbb{D}_{\mathcal{F}}\widetilde{\rho}_{1k} = \sum_{((i,j_1),(i,j_2))\in\widetilde{H}_k}\mathbb{D}_{\mathcal{F}} f_1(Z_{ij_1})f_1(Z_{ij_2}) +$$
$$+ \sum_{((i,j_1),(i,j_2))\ne((\tilde{i},\tilde{j}_1),(\tilde{i},\tilde{j}_2))\in\widetilde{H}_k}\mathbb{C}ov_{\mathcal{F}}\left\{f_i(Z_{ij_1})f_i(Z_{ij_2}), f_{\tilde{i}}(Z_{\tilde{i}\tilde{j}_1})f_{\tilde{i}}(Z_{\tilde{i}\tilde{j}_2})\right\}, \tag{217}$$

$$\mathbb{D}_{\mathcal{F}} f_1(Z_{ij_1}) f_1(Z_{ij_2}) \le \widetilde{\sigma}_f^2, \qquad \mathbb{C}ov_{\mathcal{F}} \left\{ f_i(Z_{ij_1}) f_i(Z_{ij_2}), f_{\tilde{i}}(Z_{\tilde{i}\tilde{j}_1}) f_{\tilde{i}}(Z_{\tilde{i}\tilde{j}_2}) \right\} \le \widetilde{\sigma}_f^2.$$

Учтем теперь, что если $i \neq \tilde{i}$ во второй сумме в (217), то соответствующие ковариации равны нулю. Если же $i = \tilde{i}$, то количество пар двойных индексов $((i,j_1),(i,j_2))$ и $((i,\tilde{j}_1),(i,\tilde{j}_2))$ в двойной сумме в (217) будет меньше, чем $2m_i^4$. Таким образом, для завершения доказательства остается учесть, что

$$\tilde{r}^2 \mathbb{D}_{\mathcal{F}} \widetilde{\tau}_{1k} \le \tilde{r} \widetilde{\sigma}_f^2 + 2\tilde{r} c^4 \widetilde{\sigma}_f^2, \quad \widetilde{\eta}_{\tilde{l},\tilde{r}} = \max_{1 \le k \le \tilde{l}} |\widetilde{\tau}_{1k}|.$$

Лемма доказана. □

Лемма 1.32. *Если $\widetilde{\delta}_{\tilde{l}} \le h_\varphi \le 1/2$, то для любого $\mathbf{u} \in [0,1]^2$ справедливо соотношение*

$$\widetilde{J}_{\tilde{l},h_\varphi}(\mathbf{u}) \ge 1/4 - 8L^2 \widetilde{\delta}_{\tilde{l}} h_\varphi^{-1}.$$

Доказательство. Заметим прежде всего, что при всех $\mathbf{u} = (t,s) \in [0,1]^2$ и $h_\varphi \le 1/2$ справедлива оценка снизу

$$J_{h_\varphi}(\mathbf{u}) = \int\limits_{[0,1]^2} K_{h_\varphi}(t-x) K_{h_\varphi}(s-y)\, dxdy \ge 1/4. \tag{218}$$

Действительно, если $h_\varphi \le t \le 1 - h_\varphi$, то в силу свойств ядра K

$$\int\limits_0^1 K_{h_\varphi}(t-x)dx = \int\limits_{t-h_\varphi}^{t+h_\varphi} K_{h_\varphi}(t-x)dx = \int\limits_{-1}^1 K(v)dv \equiv 1.$$

Если же $t = \alpha h$ для любых $\alpha \in [0,1]$, то

$$\int\limits_0^1 K_{h_\varphi}(t-x)dx = \int\limits_0^{(1+\alpha)h_\varphi} K_h(\alpha h_\varphi - x)dx = \int\limits_{-1}^{\alpha} K(t)dt \ge 1/2.$$

Симметричный случай $t = 1 - \alpha h$ при $\alpha \in [0,1]$ рассматривается аналогично и приведенные соотношения доказывают (218). Далее, имеет место соотношение $\widetilde{J}_{\tilde{l},h_\varphi}(\mathbf{u}) = \int g(\mathbf{v})\mathbf{dv}$, где $g(\mathbf{v}) = \sum_{k=1}^{\tilde{l}} \widetilde{K}_{h_\varphi}(\mathbf{u} - \mathbf{z}_k)\, I(\mathbf{v} \in \mathcal{P}_k)$, а через $I(\cdot)$ обозначен индикатор события. Поскольку

$$|\widetilde{K}_{h_\varphi}(\mathbf{u}-\mathbf{x}) - \widetilde{K}_{h_\varphi}(\mathbf{u}-\mathbf{y})| \le 2L^2 h^{-3} \widetilde{\delta}_{\tilde{l}} \;\text{ при }\; \|\mathbf{x}-\mathbf{y}\| \le \widetilde{\delta}_{\tilde{l}}, \tag{219}$$

то $g(\mathbf{v}) \ge \widetilde{K}_{h_\varphi}(\mathbf{u}-\mathbf{v}) - 2L^2 h_\varphi^{-3} \widetilde{\delta}_{\tilde{l}}$ для всех $\mathbf{v} \in [0,1]^2$. Следовательно, с учетом (218) получаем

$$\widetilde{J}_{\tilde{l},h_\varphi}(\mathbf{u}) \ge \int\limits_{\mathbf{v}\in[0,1]^2:\ \|\mathbf{u}-\mathbf{v}\|\le h_\varphi} \left(\widetilde{K}_{h_\varphi}(\mathbf{u}-\mathbf{v}) - 2L^2 h_\varphi^{-3}\widetilde{\delta}_{\tilde{l}}\right)\mathbf{dv} \ge 1/4 - 8L^2\widetilde{\delta}_{\tilde{l}} h_\varphi^{-1}.$$

Тем самым, лемма доказана. □

Лемма 1.33. *При всех $y > 0$ и $h \in (0, 1/2)$ на множестве элементарных исходов, определяемых соотношением $\widetilde{\delta}_{\tilde{l}}/h_\varphi \le (64L^2)^{-1}$, имеет место оценка*

$$\mathbb{P}_{\widetilde{\mathcal{F}}}\left(\sup_{\mathbf{u}\in[0,1]^2} |\widetilde{\nu}_{\tilde{l},\tilde{r},h_\varphi}(\mathbf{u})| > y\right) \le \widetilde{C}_p L^{2p}\left(M_2^{p/2} + M_p\right)\widetilde{\delta}_{\tilde{l}}^p h_\varphi^{-p-2} y^{-p}, \tag{220}$$

где символ $\mathbb{P}_{\widetilde{\mathcal{F}}}$ означает условную вероятность при фиксации σ-алгебры $\widetilde{\mathcal{F}}$, порожденной набором $\{Z_{ij}\}$ и случайными функциями $\{f_i(\cdot)\}$, положительная константа $\widetilde{C}_p$ зависит только от p,

$$\begin{aligned}
M_p &= \tilde{r}^{1-p}\left(\lambda_p(A_p + B_p) + \lambda_p^2\right) + \tilde{r}^{-p/2}\left(\tilde{\sigma}_\varepsilon^2\left(A_2^{p/2} + B_2^{p/2}\right) + \tilde{\sigma}_\varepsilon^4\right),\\
M_2 &= \tilde{\sigma}_\varepsilon^2 \tilde{r}^{-1}\left(\tilde{\sigma}_\varepsilon^2 + A_2 + B_2\right),\\
A_p &= \tilde{r}^{-1}\sum_{((i,j_1),(i,j_2))\in\widetilde{H}_k} |f_i(Z_{ij_1})|^p,\\
B_p &= \tilde{r}^{-1}\sum_{((i,j_1),(i,j_2))\in\widetilde{H}_k} |f_i(Z_{ij_2})|^p.
\end{aligned}$$

Д о к а з а т е л ь с т в о этого утверждения идейно близко к выводу леммы 1.12. Из условия $\tilde{\delta}_{\tilde{l}}/h_\varphi \le (64L^2)^{-1}$ и леммы 1.32 получаем неравенство

$$\left|\widetilde{\nu}_{\tilde{l},\tilde{r},h_\varphi}(\mathbf{u})\right| \le 8|\eta(\mathbf{u})|, \quad \text{где} \quad \eta(\mathbf{u}) = \sum_{k=1}^{\tilde{l}} \widetilde{K}_{h_\varphi}(\mathbf{u}-\mathbf{z}_k)\Lambda_2(\mathcal{P}_k)\widetilde{\varepsilon}_k. \tag{221}$$

Хвост распределения $\sup_{\mathbf{u}\in[0,1]^2}|\eta(\mathbf{u})|$ оценим с помощью метода диадических цепочек. Заметим, прежде всего, что множество $[0,1]^2$ под знаком супремума можно заменить множеством двоично-рациональных точек $\mathcal{R} = \cup_{l\ge1}\mathcal{R}_l$, где

$$\mathcal{R}_l = \{(j_1/2^l, j_2/2^l):\ j_1 = 1, \ldots, 2^l - 1; j_2 = 1, \ldots, 2^l - 1\}.$$

Таким образом,

$$\sup_{\mathbf{u}\in[0,1]^2}|\eta(\mathbf{u})| = \sup_{\mathbf{u}\in\mathcal{R}}|\eta(\mathbf{u})| \le \max_{\mathbf{u}\in\mathcal{R}_m}|\eta(\mathbf{u})| + \sum_{l=m+1}^{\infty}\sum_{r=1}^{2}\max_{\mathbf{u}\in\mathcal{R}_l}\left|\eta(\mathbf{u}+2^{-l}\mathbf{e}_r) - \eta(\mathbf{u})\right|,$$

где натуральное m будет выбрано далее, $\mathbf{e}_r$ есть двумерный вектор с единичной r-ой компонентой и другой компоненты — нулем.

Следовательно,

$$\mathbb{P}_{\widetilde{\mathcal{F}}}\left(\sup_{\mathbf{u}\in[0,1]^2}|\eta(\mathbf{u})|>y\right)\le\mathbb{P}_{\widetilde{\mathcal{F}}}\left(\max_{\mathbf{u}\in[0,1]^2}|\eta(\mathbf{u})|>a_m y\right)+$$
$$+\sum_{l=m+1}^{\infty}\sum_{r=1}^{2}\mathbb{P}_{\widetilde{\mathcal{F}}}\Big(\max_{\mathbf{u}\in\mathcal{R}_l}\big|\eta(\mathbf{u}+2^{-l}\mathbf{e}_r)-\eta(\mathbf{u})\big|>a_l y/2\Big)+$$
$$\le\sum_{\mathbf{u}\in\mathcal{R}_m}\mathbb{P}_{\widetilde{\mathcal{F}}}(|\eta(\mathbf{u})|>a_m y)+\sum_{l=m+1}^{\infty}\sum_{r=1}^{2}\sum_{\mathbf{u}\in\mathcal{R}_l}\mathbb{P}_{\widetilde{\mathcal{F}}}\Big(\big|\eta(\mathbf{u}+2^{-l}\mathbf{e}_r)-\eta(\mathbf{u})\big|>a_l y/2\Big),\quad(222)$$

где $\{a_j,\ j\ge m\}$ — последовательность положительных чисел с условием $\sum_{j\ge m}a_j=1$.

В дальнейшем нам потребуется классическое неравенство Розенталя

$$\mathbb{E}\Big|\sum_{k=1}^{l}\xi_k\Big|^p\le C_p\left(\sum_{k=1}^{l}\mathbb{E}|\xi_k|^p+\left(\sum_{k=1}^{l}\mathbb{E}\xi_k^2\right)^{p/2}\right),\quad(223)$$

где $\{\xi_k\}$ — последовательность независимых центрированных случайных величин с конечными моментами порядка $p\ge 2$, а C_p — некоторая константа, зависящая от p. Чтобы оценить вероятность $\mathbb{P}_{\widetilde{\mathcal{F}}}(|\eta(\mathbf{u})|>a_m y)$, воспользуемся неравенством (223) при

$$\xi_k=\widetilde{K}_{h_\varphi}(\mathbf{u}-\mathbf{z}_k)\Lambda_2(\mathcal{P}_k)\widetilde{\varepsilon}_k.$$

В силу элементарного соотношения $|x+y+z|^p\le 3^{p-1}(|x|^p+|y|^p+|z|^p)$ и неравенства (223), имеем

$$\mathbb{E}_{\widetilde{\mathcal{F}}}|\widetilde{\varepsilon}_k|^p\le 3^{p-1}C_pM_p\equiv\widetilde{M}_p,\quad\mathbb{E}_{\widetilde{\mathcal{F}}}|\widetilde{\varepsilon}_k|^2\le 3M_2\equiv\widetilde{M}_2,\quad(224)$$

где символ $\mathbb{E}_{\widetilde{\mathcal{F}}}$ означает условное математическое ожидание при фиксации σ-алгебры $\widetilde{\mathcal{F}}$. Из элементарных оценок

$$\widetilde{K}_{h_\varphi}(\mathbf{u}-\mathbf{z}_k)\Lambda_2(\mathcal{P}_k)\le L^2h_\varphi^{-2}\widetilde{\delta}_{\tilde{l}}^2,\quad\sum_{k=1}^{\tilde{l}}\widetilde{K}_{h_\varphi}(\mathbf{u}-\mathbf{z}_k)\Lambda_2(\mathcal{P}_k)\le L^2h_\varphi^{-2}(2h_\varphi+2\widetilde{\delta}_{\tilde{l}})^2$$

и соотношения $\widetilde{\delta}_{\tilde{l}}\le h_\varphi\le 1$ получаем, что с вероятностью 1

$$\mathbb{P}_{\widetilde{\mathcal{F}}}(|\eta(\mathbf{u})|>a_m y)\le\frac{\mathbb{E}_{\widetilde{\mathcal{F}}}(|\eta(\mathbf{u})|^p)}{(a_m y)^p}\le G_1\frac{\widetilde{M}_2^{p/2}(\widetilde{\delta}_{\tilde{l}}/h_\varphi)^p+\widetilde{M}_p(\widetilde{\delta}_{\tilde{l}}/h_\varphi)^{2(p-1)}}{(a_m y)^p}\le$$
$$\le G_1\frac{(\widetilde{M}_2^{p/2}+\widetilde{M}_p)(\widetilde{\delta}_{\tilde{l}}/h_\varphi)^p}{(a_m y)^p},\quad(225)$$

где $G_1 = C_p 2^{2p} L^{2p}$. В последнем неравенстве в (225) мы также учли, что $2(p-1) > p$ при $p > 2$.

Чтобы оценить $\mathbb{P}_{\widetilde{\mathcal{F}}}\Big(\big|\eta(\mathbf{u}+2^{-l}\mathbf{e}_r)-\eta(\mathbf{u})\big| > a_l y/2\Big)$, воспользуемся неравенством (223) при

$$\xi_k = \Big(\widetilde{K}_{h_\varphi}(\mathbf{u}-\mathbf{z}_k+2^{-l}\mathbf{e}_r)-\widetilde{K}_{h_\varphi}(\mathbf{u}-\mathbf{z}_k)\Big)\Lambda_2(\mathcal{P}_k)\widetilde{\varepsilon}_k.$$

Имеем

$$\Big|\widetilde{K}_{h_\varphi}(\mathbf{u}-\mathbf{z}_k+2^{-l}\mathbf{e}_r)-\widetilde{K}_{h_\varphi}(\mathbf{u}-\mathbf{z}_k)\Big|\Lambda_2(\mathcal{P}_k) \le L^2 h_\varphi^{-3} 2^{-l}\widetilde{\delta}_{\tilde{l}}^2,$$

$$\sum_{k=1}^{\tilde{l}}\Big|\widetilde{K}_{h_\varphi}(\mathbf{u}-\mathbf{z}_k+2^{-l}\mathbf{e}_r)-\widetilde{K}_{h_\varphi}(\mathbf{u}-\mathbf{z}_k)\Big|\Lambda_2(\mathcal{P}_k) \le L^2 h_\varphi^{-3} 2^{-l+1}(2h_\varphi+2\widetilde{\delta}_{\tilde{l}})^2.$$

Таким образом, с учетом аргументов, используемых при выводе (225), получаем, что с вероятностью 1

$$\mathbb{P}_{\widetilde{\mathcal{F}}}\left(\big|\eta(\mathbf{u}+2^{-l}\mathbf{e}_r)-\eta(\mathbf{u})\big| > a_l y/2\right) \le G_2\Big(\widetilde{M}_2^{p/2}+\widetilde{M}_p\Big)\,(\widetilde{\delta}_{\tilde{l}}/h_\varphi)^p h_\varphi^{-p} 2^{-lp-1}(a_l y)^{-p}, \tag{226}$$

где $G_2 = 2^{3p/2+1} G_1$.

Используя теперь (222), (225) и (226), заключаем следующее:

$$\mathbb{P}_{\widetilde{\mathcal{F}}}\left(\sup_{\mathbf{u}\in[0,1]^2}|\eta(\mathbf{u})| > y\right) \le y^{-p}\left(\frac{\widetilde{\delta}_{\tilde{l}}}{h_\varphi}\right)^p\left(\widetilde{M}_2^{p/2}+\widetilde{M}_p\right)\left(G_1 2^{2m} a_m^{-p} + G_2 h_\varphi^{-p}\sum_{l=m+1}^{\infty} 2^{-(p-2)l} a_l^{-p}\right).$$

Оптимальная последовательность a_l, минимизирующая правую часть этого неравенства, есть $a_m = c(G_1 2^{2m})^{1/(p+1)}$, $a_l = c\big(G_2 h_\varphi^{-p} 2^{-(p-2)l}\big)^{1/(p+1)}$ при $l = m+1, m+2, \ldots$, где коэффициент c определяется соотношением $a_m + a_{m+1} + \cdots = 1$. Для указанной последовательности получаем

$$G_1 2^{2m} a_m^{-p} + G_2 h_\varphi^{-p}\sum_{l=m+1}^{\infty} 2^{-(p-2)l} a_l^{-p} \le$$

$$\le \left((G_1 2^{2m})^{1/(p+1)} + \sum_{l=m+1}^{\infty}\Big(G_2 h_\varphi^{-p} 2^{-(p-2)l}\Big)^{1/(p+1)}\right)^{p+1} =: G.$$

Положим $m = \lceil -\log_2 h\rceil$. В этом случае

$$G \le h_\varphi^{-2}\left((2G_1)^{1/(p+1)} + \big(G_2 2^{-(p-2)}\big)^{1/(p+1)}\sum_{l=0}^{\infty} 2^{-(p-2)l/(p+1)}\right)^{p+1} \le$$

$$\le h_\varphi^{-2} 2^{p/2} G_1\left(1+\frac{2}{2^{(p-2)/(p+1)}-1}\right)^{p+1}.$$

Таким образом,

$$\mathbb{P}_{\widetilde{\mathcal{F}}}\left(\sup_{\mathbf{u}\in[0,1]^2}|\eta(\mathbf{u})|>y\right)\le 2^{3p}\left(1+\frac{2}{2^{(p-2)/(p+1)}-1}\right)^{p+1}C_pL^{2p}\left(\widetilde{M}_2^{p/2}+\widetilde{M}_p\right)\widetilde{\delta}_{\tilde{l}}^{p}h_\varphi^{-p-2}y^{-p}.$$

Это соотношение вместе с (221) и (224) завершает доказательство леммы. □

Д о к а з а т е л ь с т в о теоремы 1.14. Положим $\widetilde{\zeta}_{\tilde{l},\tilde{r},h_\varphi}=\sup_{\mathbf{t}\in[0,1]^2}|\widetilde{\nu}_{\tilde{l},\tilde{r},h_\varphi}(\mathbf{t})|$ и через $\mathbb{E}_{\mathcal{F}}$ обозначим условное математическое ожидание при фиксации σ-алгебры $\mathcal{F}$, порожденной случайными величинами $\{Z_{ij}:i=1,\dots,n,\ j=1,\dots,m_i\}$. Утверждение теоремы следует теперь из тождества (216), лемм 1.31, 1.33 и соотношения

$$\mathbb{P}\big(\widetilde{\zeta}_{\tilde{l},\tilde{r},h_\varphi}>y,\ \widetilde{\delta}_{\tilde{l}}\le h_\varphi(64L^2)^{-1}\big)=$$
$$=\mathbb{E}\mathbb{E}_{\mathcal{F}}I\big(\widetilde{\delta}_{\tilde{l}}\le h_\varphi(64L^2)^{-1}\big)\mathbb{P}_{\widetilde{\mathcal{F}}}\big(\widetilde{\zeta}_{\tilde{l},\tilde{r},h_\varphi}>y\big)\le\mathbb{E}\mathbb{E}_{\mathcal{F}}\mathbb{P}_{\widetilde{\mathcal{F}}}\big(\widetilde{\zeta}_{\tilde{l},\tilde{r},h_\varphi}>y\big).$$

При этом в последнем соотношении мы также учли, что

$$\mathbb{E}_{\mathcal{F}}A_2^{p/2}\le\Bigg(\tilde{r}^{-1}\sum_{((i,j_1),(i,j_2))\in\widetilde{H}_k}\sup_{t\in[0,1]}|f_i(t)|^2\Bigg)^{p/2}\le$$
$$\le\mathbb{E}_{\mathcal{F}}\tilde{r}^{-1}\sum_{((i,j_1),(i,j_2))\in\widetilde{H}_k}\sup_{t\in[0,1]}|f_i(t)|^p\equiv\gamma_p,\qquad \mathbb{E}_{\mathcal{F}}A_p\le\gamma_p.$$

Аналогичные оценки справедливы для величин $\mathbb{E}_{\mathcal{F}}B_2^{p/2}$ и $\mathbb{E}_{\mathcal{F}}B_p$. Теорема 1.14 доказана. □

1.5.3 Случай плотных данных

Введем прежде всего ряд дополнительных условий и обозначений, которые потребуются в этом разделе.

($\mathbf{E}_5$) *Для любого $i=1,\dots,n$ и всех $m_i\ge 2$ ненаблюдаемые случайные погрешности $\{\varepsilon_{ij};\ j=1,\dots,m_i\}$ с вероятностью 1 при всех $j,j_1,j_2\le m_i$, $j_1\ne j_2$, удовлетворяют следующим условиям:*

$$\mathbb{E}_{\mathcal{F}_i}\varepsilon_{ij}=0,\quad \max_{j\le m_i}\mathbb{E}_{\mathcal{F}_i}\varepsilon_{ij}^2\le\sigma^2,\quad \mathbb{E}_{\mathcal{F}_i}\varepsilon_{ij_1}\varepsilon_{ij_2}=0,$$

где константа $\sigma^2>0$ не зависит от m_i и n и может быть неизвестной, символ $\mathbb{E}_{\mathcal{F}_i}$ обозначает условное математическое ожидание при фиксации σ-алгебры $\mathcal{F}_i$, порожденной случайными величинами $\{Z_{ij};\ j=1,\dots,m_i\}$.

Для любого i обозначим через $Z_{i,m_i:1}\le\ldots\le Z_{i,m_i:m_i}$ элементы вариационного ряда, построенного по выборке i-ой серии $\{Z_{ij};\ j=1,\dots,m_i\}$. Положим

$$Z_{i,m_i:0}=0,\quad Z_{i,m_i:m_i+1}=1,\quad \Delta Z_{i,m_i,j}=Z_{i,m_i:j}-Z_{i,m_i:j-1},\quad j=1,\dots,m_i+1.$$

Для любого i отклики и погрешности из (199), ассоциированные с порядковой статистикой $Z_{i,m_i:j}$, обозначим соответственно через $X_{i,m_i,j}$ и $\varepsilon_{i,m_i,j}$. Нетрудно видеть, что погрешности $\{\varepsilon_{i,m_i,j};\ j = 1,\ldots,m_i\}$ также удовлетворяют условию (E_5). Мы предполагаем, что величины m_i зависят от n и $m_i = m_i(n) \to \infty$ при $n \to \infty$.

Центральное условие на регрессоры состоит в следующем.

$(\mathbf{D}_9)$ *При любом* $i = 1,\ldots,n$ *имеет место предельное соотношение*

$$\delta_{m_i} = \max_{1\le j\le m_i+1} \Delta Z_{i,m_i,j} \xrightarrow{p} 0 \quad \textit{при } m_i = m_i(n) \to \infty.$$

З а м е ч а н и е 1.42. Условие (D_9) означает, что для любой серии i набор регрессоров $\{Z_{ij}, j = 1,\ldots,m_i\}$ с высокой вероятностью образуют измельчающееся разбиение отрезка $[0,1]$. Понятно, что детерминированные регрессоры регулярного типа удовлетворяет условию (D_9). В случае независимых и одинаково распределенных регрессоров и существования отделенной от нуля на $[0,1]$ плотности распределения Z_{i1}, с вероятностью 1 выполнено $\delta_{mi} = O\left(\log m_i/m_i\right)$ (см. детали в разделе 1.1). Если $\{Z_{ij};\ j \ge 1\}$ — стационарная последовательность с условием α-перемешивания и маргинальным распределением с носителем $[0,1]$, то условие (D_9) также выполнено. Как мы уже отмечали, зависимость случайных величин в условии (D_9) может быть более сильной (см. примеры 1.1 и 1.2 в разделе 1.1). $\square$

Наконец, для любых $h_1,\ldots,h_n \in (0,1)$ введем в рассмотрение следующие классы оценок:

$$\widehat{f}_i(t) = \frac{\sum_{j=1}^{m_i} X_{i,m_i,j} K_{h_i}(t - Z_{i,m_i:j})\Delta Z_{i,m_i,j}}{\sum_{j=1}^{m_i} K_{h_i}(t - Z_{i,m_i:j})\Delta Z_{i,m_i,j}}, \quad i = 1,\ldots,n, \tag{227}$$

$$\widehat{\mu}_2(t) = \frac{1}{n}\sum_{i=1}^{n}\widehat{f}_i(t), \quad \widehat{\varphi}_2(t,s) = \frac{1}{n}\sum_{i=1}^{n}\widehat{f}_i(t)\widehat{f}_i(s), \quad \widehat{\psi}_2(t,s) = \widehat{\varphi}_2(t,s) - \widehat{\mu}_2(t)\widehat{\mu}_2(s). \tag{228}$$

З а м е ч а н и е 1.43. Как мы уже отмечали во введении, в ситуации плотных данных представляется естественным предварительно оценить случайные функции f_i по наблюдениям i-ой серии, а затем для оценивания функции среднего провести усреднение по всем сериям (аналогичным образом можно можно оценить и функцию ковариации). Именно так мы и поступаем в (228), следуя этому общепринятому подходу. Отметим, что при $n = 1$ статистики вида (227) были введены и исследованы в разделе 1.1. В частности, там доказано, что условие $\delta_{m_1} \xrightarrow{p} 0$ при $m_1 \to \infty$, содержащееся в предположении (D_9), гарантирует существование в классе оценок (227) равномерно состоятельной оценки для случайного процесса $f_1(t)$.

З а м е ч а н и е 1.44. В случае, когда имеются кратные точки в той или иной серии наблюдений i, некоторые спейсинги $\Delta Z_{i,m_i,j}$ обращаются в ноль, и мы теряем часть выборочной информации в оценках (227) (и, как следствие, оценках (228)). В этом случае предлагается прежде, чем использовать оценку (227), несколько сократить выборку в каждой серии i, содержащей кратные регрессоры, заменив наблюдения X_{ij} с одинаковыми точками Z_{ij} их средним арифметическим и оставляя в новой выборке лишь одну точку из кратных. При этом усредненные наблюдения будут иметь меньшее зашумление. Так что, несмотря на меньший объем новой выборки, мы не теряем информацию, содержащуюся в исходной выборке. $\square$

З а м е ч а н и е 1.45. В случае независимых одинаково распределенных регрессоров $\{Z_{ij}\}$ естественно рассматривать один и тот же размер окна при оценивании каждой из функций $\{f_i(t)\}$ (см., например, [311]). В указанной ситуации подобная стратегия оправдана, поскольку уменьшает вычислительные сложности по определению оптимального размера окна и упрощает асимптотический анализ оценки среднего. При выполнении условия (D_9), когда допускается, что в каждой серии регрессоры могут вести себя различным образом, может быть полезным рассматривать различный размер окна при сглаживании в каждой серии наблюдений. $\square$

Теорема 1.15. *Пусть выполнены условия* (M_2), (E_5), (K_1), (D_9) *и*

$$\mathbb{E}\sup_{t\in[0,1]}|f_1(t)|<\infty. \tag{229}$$

Пусть последовательности $h_i=h_i(m_i,n)$ *удовлетворяют условиям*

$$\max_{1\le i\le n}h_i\to 0,\quad \frac{1}{n}\sum_{i=1}^{n}\frac{\left(\mathbb{E}\delta_{m_i}\right)^{1/2}}{h_i}\to 0,\quad \sum_{i=1}^{n}\mathbb{P}(\delta_{m_i}>h_i/(8L))\to 0. \tag{230}$$

Тогда

$$\sup_{t\in[0,1]}|\widehat{\mu}_2(t)-\mu(t)|\xrightarrow{p}0.$$

Теорема 1.16. *Пусть выполнены условия* (M_2), (E_5), (K_1), (D_9) *и*

$$\mathbb{E}\sup_{t\in[0,1]}|f_1(t)|^2<\infty. \tag{231}$$

Пусть последовательности $h_i=h_i(m_i,n)$ *удовлетворяют* (230) *и при некотором* $\tau\in(0,1)$

$$n^{-\tau}\sum_{i=1}^{n}\mathbb{E}\omega_{f_1}^{2\tau}(h_i)\to 0,\quad n^{-\tau}\sum_{i=1}^{n}\frac{\left(\mathbb{E}\delta_{m_i}\right)^{\tau}}{h_i^{2\tau}}\to 0. \tag{232}$$

Тогда

$$\sup_{t,s\in[0,1]}|\widehat{\varphi}_2(t,s)-\varphi(t,s)|\xrightarrow{p}0,\quad \sup_{t,s\in[0,1]}\left|\widehat{\psi}_2(t,s)-\psi(t,s)\right|\xrightarrow{p}0.$$

Приведем два следствия, в которых предполагается, что $m_j = m$ и $h_j = h$ при всех $j \le n$. Нам потребуются следующие дополнительные условия.

($\mathbf{M}_2'$) *Выполнено условие* (M_2) *при* $m_1 = \ldots = m_n = m$ *и для каждого* $i = 2, \ldots, n$ *набор* $\{Z_{ij};\ j = 1, \ldots, m\}$ *является независимой копией* $\{Z_{1j};\ j = 1, \ldots, m\}$.

($\mathbf{D}_9'$) *Имеет место предельное соотношение*

$$\delta_m = \max_{1 \le j \le m+1} \Delta Z_{1,m,j} \xrightarrow{p} 0 \quad \textit{при} \quad m = m(n) \to \infty.$$

Следствие 1.19. *Пусть выполнены условия* (M_2'), (E_5), (K_1), (D_9'), (229), *а последовательность* $h = h(m,n)$ *удовлетворяет условиям*

$$h^{-2}\mathbb{E}\delta_m \to 0, \quad n\mathbb{P}(\delta_m > h/(8L)) \to 0. \tag{233}$$

Тогда имеет место утверждение теоремы 1.15.

Если вместо (229) *и первого условия в* (233) *выполнено условие* (231) *и дополнительно при некотором* $\tau \in (0,1)$

$$n^{1-\tau}\mathbb{E}\omega_{f_1}^{2\tau}(h) \to 0, \quad n^{(1-\tau)/\tau}h^{-2}\mathbb{E}\delta_m \to 0, \tag{234}$$

то справедливо утверждение теоремы 1.16.

П р и м е р 1.22. Пусть в условиях следствия 1.19 выполнено $\mathbb{E}\delta_{m_1} = O(1/m)$. Тогда условия $mh^2 \to \infty$ и $n/(mh) \to 0$ влекут (233). Если $\omega_{f_1}(h) \le \zeta h^\gamma$ с вероятностью 1 при некотором неслучайном $\gamma \in (0,1]$ и $\mathbb{E}\zeta < \infty$, то соотношения $n^{1-\tau}h^{2\tau\gamma} \to 0$ и $n^{(1-\tau)/\tau}/(mh^2) \to 0$ гарантируют выполнение условий из (234).

Рассмотрим другой частный случай — общий набор регрессоров для всех серий.

($\mathbf{M}_2''$)*Выполнено условие* (M_2) *при* $m_1 = \ldots = m_n = m$ *и для любого* $j = 1, \ldots, m$ *выполнено* $Z_{1j} = Z_{2j} = ...Z_{nj} = Z_j$.

Обозначим через $Z_{m:1} \le \ldots \le Z_{m:m}$ элементы вариационного ряда, построенного по выборке $\{Z_j;\ j = 1, \ldots, m\}$. Положим

$$Z_{m:0} = 0, \quad Z_{m:m+1} = 1, \quad \Delta Z_{mj} = Z_{m:j} - Z_{m:j-1}, \quad j = 1, \ldots, m_i + 1.$$

($\mathbf{D}_9''$) *Имеет место предельное соотношение*

$$\delta_m = \max_{1 \le j \le m+1} \Delta Z_{mj} \xrightarrow{p} 0 \quad \textit{при} \quad m = m(n) \to \infty.$$

Для любого i отклики из (199), ассоциированные с порядковой статистикой $Z_{m:j}$, обозначим через $X_{i,m,j}$. В этом случае оценка $\widehat{f}_i(t)$ из (227) примет вид

$$\widehat{f}_i(t) = \frac{\sum_{j=1}^{m} X_{i,m,j} K_h(t - Z_{m:j}) \Delta Z_{mj}}{\sum_{j=1}^{m} K_h(t - Z_{m:j}) \Delta Z_{mj}}, \quad i = 1, \ldots, n.$$

Следствие 1.20. *Пусть выполнены условия* (M_2''), (E_5), (K_1), (D_9''), (229), *а последовательность* $h = h(m, n)$ *удовлетворяет условиям*

$$h^{-2}\mathbb{E}\delta_m \to 0, \quad n\mathbb{P}(\delta_m > h/(8L)) \to 0. \tag{235}$$

Тогда имеет место утверждение теоремы 1.15.

Если вместо (229) *и первого условия в* (235) *справедливо условие* (231) *и при некотором* $\tau \in (0, 1)$

$$n^{1-\tau}\mathbb{E}\omega_{f_1}^{2\tau}(h) \to 0, \quad n^{(1-\tau)/\tau} h^{-2}\mathbb{E}\delta_m \to 0, \tag{236}$$

то справедливо утверждение теоремы 1.16.

1.5.4 Доказательство результатов раздела 1.5.3

Для доказательства теорем 1.15 и 1.16 нам также потребуется несколько вспомогательных утверждений.

Лемма 1.34. *Пусть выполнены условия* (M_2), (E_5) *и* (K_1). *Тогда для любого* $i = 1, \ldots, n$ *и любого фиксированного* $h_i \in (0, 1)$ *с вероятностью* 1

$$\Delta_{1i} = \sup_{t \in [0,1]} |\widehat{f}_i(t) - f_i(t)| \le \omega_{f_i}(h_i) + \zeta_{m_i}(h_i),$$

где случайная величина $\zeta_{m_i}(h_i)$ *такова, что для любых* $y > 0$

$$\mathbb{P}\big(\zeta_{m_i}(h_i) > y,\, B_i\big) \le C\sigma^2 L^2 y^{-2} h_i^{-2} \mathbb{E}\delta_{m_i}, \quad B_i = \{\delta_{m_i} < h_i/(8L)\} \tag{237}$$

и C *– абсолютная положительная константа. В частности, на подмножестве элементарных исходов* B_i

$$\mathbb{P}_{\mathcal{F}_i}\left(\zeta_{m_i}(h_i) > y\right) \le C\sigma^2 L^2 y^{-2} h_i^{-2} \delta_{m_i},$$

а символ $\mathbb{P}_{\mathcal{F}_i}$ *обозначает условную вероятность при фиксации* σ*-алгебры* $\mathcal{F}_i$*, введенной в условии* (E_5).

Это утверждение следует из теоремы 1.1, приведенной в разделе 1.1.

Лемма 1.35. *Если выполнено условие* (229), *то* $\lim_{\varepsilon\to 0}\mathbb{E}\omega_{f_1}(\varepsilon) = 0$ *и для независимых копий* $f_1(t),\ldots,f_n(t)$ *почти наверное непрерывного случайного процесса имеет место следующий равномерный закон больших чисел:*

$$\sup_{t\in[0,1]}\left|\overline{f}_n(t)-\mu(t)\right| \xrightarrow{p} 0, \quad \textit{где}\ \ \overline{f}_n(t) = n^{-1}\sum_{i=1}^{n} f_i(t). \tag{238}$$

Д о к а з а т е л ь с т в о. Первое утверждение леммы следует из (229) и теоремы Лебега о мажорируемой сходимости. Положим

$$\omega_{\overline{f}_n}(\varepsilon) = \sup_{t,s:|t-s|\le\varepsilon}\left|\overline{f}_n(t)-\overline{f}_n(s)\right|.$$

При произвольном фиксированном $k>0$ и $u=0,\ldots,k$, имеют место соотношения

$$\begin{aligned}\sup_{t\in[0,1]}\left|\overline{f}_n(t)-\mu(t)\right| &\le \max_{0\le u\le k}\left|\overline{f}_n\big(u/k\big)-\mu\big(u/k\big)\right| + \\ &+ \max_{1\le u\le k}\sup_{(u-1)/k\le t\le u/k}\left|\overline{f}_n(t)-\overline{f}_n\big(u/k\big)\right| + \max_{1\le u\le k}\sup_{(u-1)/k\le t\le u/k}\left|\mu(t)-\mu\big(u/k\big)\right| \le \\ &\le \max_{0\le u\le k}\left|\overline{f}_n\big(u/k\big)-\mu\big(u/k\big)\right| + \omega_{\overline{f}_n}(1/k) + \omega_{\mu}(1/k)\,.\end{aligned} \tag{239}$$

Заметим теперь, что $\omega_\mu(\varepsilon)\le\mathbb{E}\omega_{f_1}(\varepsilon)$ и

$$\overline{f}_n\big(u/k\big) \xrightarrow{p} \mu\big(u/k\big), \quad \omega_{\overline{f}_n}(\varepsilon) \le \frac{1}{n}\sum_{i=1}^{n}\omega_{f_i}(\varepsilon) \xrightarrow{p} \mathbb{E}\omega_{f_1}(\varepsilon).$$

Следовательно, правая часть в (239) не превосходит $2\mathbb{E}\omega_{f_1}(1/k)+o_p(1)$ и в силу произвольности k и первого утверждения леммы соотношение (238) доказано. □

Лемма 1.36. *В условиях теоремы* 1.15 *имеет место предельное соотношение*

$$\frac{1}{n}\sum_{i=1}^{n}\Delta_{1i} \xrightarrow{p} 0, \quad \textit{где}\ \Delta_{1i} = \sup_{t\in[0,1]}|\widehat{f}_i(t)-f_i(t)|. \tag{240}$$

Д о к а з а т е л ь с т в о. Для любого положительного ε и событий B_i, $i=1,\ldots,n$, определенных в лемме 1.34, имеем

$$\begin{aligned}\mathbb{P}\left\{\frac{1}{n}\sum_{i=1}^{n}\Delta_{1i}>\varepsilon\right\} &\le \mathbb{P}\left\{\frac{1}{n}\sum_{j=1}^{n}\Delta_{1i}>\varepsilon,\ \bigcap_{i=1}^{n}B_i\right\} + \mathbb{P}\left(\bigcup_{i=1}^{n}\overline{B}_i\right) \le \\ &\le \mathbb{P}\left\{\frac{1}{n}\sum_{i=1}^{n}\Delta_{1i}I(B_i)>\varepsilon\right\} + \sum_{i=1}^{n}\mathbb{P}(\overline{B_i}),\end{aligned} \tag{241}$$

при этом второе слагаемое в правой части (241) сходится к нулю ввиду третьего условия в (230). Покажем, что к нулю сходится и первое слагаемое. Действительно, из леммы 1.34 получаем

$$\mathbb{E}\big\{\Delta_{1i}I(B_i)\big\} \le \mathbb{E}\omega_{f_1}(h_i) + \int\limits_0^\infty \mathbb{P}\left(\zeta_{m_i}(h_i) > y,\, B_i\right) dy \le$$

$$\le \mathbb{E}\omega_{f_1}(h_i) + h_i^{-1}(\mathbb{E}\delta_{m_i})^{1/2} + \int\limits_{{h_i}^{-1}(\mathbb{E}\delta_{m_i})^{1/2}}^\infty \mathbb{P}\left(\zeta_{m_i}(h_i) > y,\, B_i\right) dy \le$$

$$\le \mathbb{E}\omega_{f_1}(h_i) + (1 + C\sigma^2L^2)h_i^{-1}(\mathbb{E}\delta_{m_i})^{1/2}. \quad (242)$$

Для завершения доказательства леммы остается для первой вероятности в правой части (241) применить неравенство Маркова, использовать оценку (242), предельные соотношения (230) и первое утверждение леммы 1.35. □

Д о к а з а т е л ь с т в о теоремы 1.15 следует из лемм 1.35 и 1.36 и очевидной оценки

$$\sup_{t\in[0,1]} \big|\widehat{\mu}_2(t) - \mu(t)\big| \le \sup_{t\in[0,1]} \big|\overline{f}_n(t) - \mathbb{E}f_1(t)\big| + \frac{1}{n}\sum_{i=1}^n \Delta_{1i},$$

при выводе которой мы учли определения (200), (228), (238) и (240). □

Лемма 1.37. *Если выполнено условие* (231), *то* $\lim_{\varepsilon\to 0}\mathbb{E}\omega^2_{f_1}(\varepsilon) = 0$ *и для независимых копий* $f_1(t), \ldots, f_n(t)$ *почти наверное непрерывного случайного процесса имеет место следующий равномерный закон больших чисел:*

$$\sup_{t,s\in[0,1]} \big|\overline{f}_n(t,s) - \varphi(t,s)\}\big| \xrightarrow{p} 0, \quad \text{где } \overline{f}_n(t,s) = n^{-1}\sum_{i=1}^n f_i(t)f_i(s). \quad (243)$$

Д о к а з а т е л ь с т в о. Первое утверждение леммы следует из (231) и теоремы Лебега о мажорируемой сходимости. Далее, аналогично выводу леммы 1.35, при произвольном фиксированном $k > 0$ и $u, v = 0, \ldots, k$, имеют место соотношения

$$\sup_{t,s\in[0,1]} \big|\overline{f}_n(t,s) - \varphi(t,s)\big| \le \max_{0\le u,v\le k} \big|\overline{f}_n\big(u/k, v/k\big) - \varphi\big(u/k, v/k\big)\big| +$$

$$+ \max_{1\le u,v\le k} \sup_{\frac{u-1}{k}\le t\le\frac{u}{k},\, \frac{v-1}{k}\le s\le\frac{v}{k}} \big|\overline{f}_n(t,s) - \overline{f}_n\big(u/k, v/k\big)\big| +$$

$$+ \max_{1\le u,v\le k} \sup_{\frac{u-1}{k}\le t\le\frac{u}{k},\, \frac{v-1}{k}\le s\le\frac{v}{k}} \big|\varphi\big(t,s\big) - \varphi\big(u/k, v/k\big)\big| \le$$

$$\le \max_{0\le u,v\le k} \big|\overline{f}_n\big(u/k, v/k\big) - \varphi\big(u/k, v/k\big)\big| +$$

$$+ \frac{1}{n}\sum_{i=1}^n \sup_{t\in[0,1]} |f_i(t)|\omega_{f_i}(1/k) + \mathbb{E}\left\{\sup_{t\in[0,1]} |f_1(t)|\omega_{f_1}\left(1/k\right)\right\}. \quad (244)$$

Учтем теперь, что $\overline{f}_n\big(u/k, v/k\big) \xrightarrow{p} \varphi\big(u/k, v/k\big)$,

$$\frac{1}{n}\sum_{i=1}^{n}\sup_{t\in[0,1]} f_i(t)\omega_{f_i}(\varepsilon) \xrightarrow{p} \mathbb{E}\left\{\sup_{t\in[0,1]}|f_1(t)|\omega_{f_1}(\varepsilon)\right\} \leq \left(\mathbb{E}\sup_{t\in[0,1]}|f_1(t)|^2\right)^{1/2}\big(\mathbb{E}\omega^2_{f_1}(\varepsilon)\big)^{1/2}.$$

Таким образом, ввиду (231), правая часть в (244) не превосходит $C\big(\mathbb{E}\omega^2_{f_1}(1/k)\big)^{1/2} + o_p(1)$ и в силу произвольности k и первого утверждения леммы предельное соотношение (243) доказано. $\square$

Лемма 1.38. *В условиях теоремы* 1.16 *имеет место предельное соотношение*

$$\frac{1}{n}\sum_{i=1}^{n}\Delta_{2i} \xrightarrow{p} 0 \qquad \textit{при} \quad \Delta_{2i} = \sup_{t,s\in[0,1]}|\widehat{f}_i(t)\widehat{f}_i(s) - f_i(t)f_i(s)|. \tag{245}$$

Д о к а з а т е л ь с т в о. Заметим прежде всего, что с учетом (240) и (245), выполнено

$$\Delta_{2i} \leq \big(\Delta_{1i}\big)^2 + 2\sup_{t\in[0,1]}|f_i(t)|\Delta_{1i}. \tag{246}$$

Отметим, что аналогично (241), для любого фиксированного $\varepsilon > 0$ выполнено

$$\mathbb{P}\left(\sum_{i=1}^{n}\Delta^2_{1i} > n\varepsilon\right) \leq \mathbb{P}\left(\sum_{i=1}^{n}\Delta^2_{1i}I(B_i) > n\varepsilon\right) + \sum_{i=1}^{n}\mathbb{P}(\overline{B_i}) \xrightarrow{p} 0 \tag{247}$$

в силу условий (232). При доказательстве сходимости в (247) для первой вероятности в правой части мы использовали также неравенство

$$\mathbb{P}\left(\sum_{i=1}^{n}\Delta^2_{1i}I(B_i) > n\varepsilon\right) \leq (n\varepsilon)^{-\tau}\mathbb{E}\left(\sum_{i=1}^{n}\Delta^2_{1i}I(B_i)\right)^{\tau} \leq (n\varepsilon)^{-\tau}\sum_{i=1}^{n}\mathbb{E}\Delta^{2\tau}_{1i}I(B_i)$$

при $\tau \in (0,1)$ и следующую цепочку соотношений, справедливую в силу леммы 1.34:

$$\mathbb{E}\Delta^{2\tau}_{1i}I(B_i) \leq 2\mathbb{E}\omega^{2\tau}_{f_1}(h_i) + 4\tau\int_0^\infty y^{2\tau-1}\mathbb{P}\left(\zeta_{m_i}(h_i) > y,\, B_i\right)dy \leq$$

$$\leq 2\mathbb{E}\omega^{2\tau}_{f_1}(h_i) + 4\tau\int_0^{h_i^{-1}(\mathbb{E}\delta_{m_i})^{1/2}} y^{2\tau-1}dy + 4\tau\int_{{h_i}^{-1}(\mathbb{E}\delta_{m_i})^{1/2}}^{\infty} y^{2\tau-1}\mathbb{P}\left(\zeta_{m_i}(h_i) > y,\, B_i\right)dy \leq$$

$$\leq 2\mathbb{E}\omega^{2\tau}_{f_1}(h_i) + 2(1 + \tau(\tau-1)^{-1}\sigma^2L^2)h_i^{-2\tau}(\mathbb{E}\delta_{m_i})^{\tau}.$$

Далее, для любого $\varepsilon > 0$ с учетом леммы 1.34 имеем

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\sup_{t\in[0,1]}|f_i(t)|\Delta_{1i} > \varepsilon\right) \le \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\sup_{t\in[0,1]}|f_i(t)|\Delta_{1i} > \varepsilon,\ \bigcap_{i=1}^{n} B_i\right) + \sum_{i=1}^{n}\mathbb{P}(\overline{B_i}) \le$$
$$\le \frac{1}{n\varepsilon}\sum_{i=1}^{n}\mathbb{E}\sup_{t\in[0,1]}|f_i(t)|\Delta_{1i}I(B_i) + \sum_{i=1}^{n}\mathbb{P}(\overline{B_i}) \le \frac{1}{n\varepsilon}\sum_{i=1}^{n}\mathbb{E}\sup_{t\in[0,1]}|f_i(t)|\omega_{f_i}(h_i) +$$
$$+ \frac{1}{n\varepsilon}\sum_{i=1}^{n}\mathbb{E}\sup_{t\in[0,1]}|f_i(t)|\mathbb{E}_{\mathcal{F}_i}\zeta_{m_i}(h_i)I(B_i) + \sum_{i=1}^{n}\mathbb{P}(\overline{B_i}) \xrightarrow{p} 0. \quad (248)$$

Действительно, первое слагаемое в правой части соотношения (248) сходится к нулю ввиду первого утверждения леммы 1.37, условия (231), первого условия в (230) и неравенства Коши–Буняковского

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\sup_{t\in[0,1]}|f_i(t)|\omega_{f_i}(h_i) \le \left(\mathbb{E}\sup_{t\in[0,1]} f_1^2(t)\right)^{1/2}\left(\mathbb{E}\omega_{f_1}^2\big(\max_i h_i\big)\right)^{1/2}.$$

Для второго слагаемого в правой части (248) учтем, что в силу леммы 1.34

$$\mathbb{E}_{\mathcal{F}_i}\zeta_{m_i}(h_i)I(B_i) = \int_0^\infty \mathbb{P}_{\mathcal{F}_i}\left(\zeta_{m_i}(h_i) > y,\, B_i\right)dy \le$$
$$\le \int_0^{h_i^{-1}\delta_{m_i}^{1/2}} dy + \int_{h_i^{-1}\delta_{m_i}^{1/2}}^{\infty} \mathbb{P}\left(\zeta_{m_i}(h_i) > y,\, B_i\right)dy \le (1 + C\sigma^2L^2)h_i^{-1}\delta_{m_i}^{1/2},$$

а потому из (230) и (231) получаем, что

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\sup_{t\in[0,1]}|f_i(t)|\mathbb{E}_{\mathcal{F}_i}\zeta_{m_i}(h_i)I(B_i) \le (1 + C\sigma^2L^2)\left(\mathbb{E}\sup_{t\in[0,1]} f_1^2(t)\right)^{1/2}\frac{1}{n}\sum_{i=1}^{n}\frac{\left(\mathbb{E}\delta_{m_i}\right)^{1/2}}{h_i} \to 0$$

и соотношение (248) доказано. Утверждение леммы следует из (246)–(248). □

Д о к а з а т е л ь с т в о теоремы 1.16. Нетрудно видеть (см. (200), (228), (243) и (245)), что имеет место оценка

$$\sup_{t,s\in[0,1]}\left|\widehat{\varphi}_2(t,s) - \varphi(t,s)\right| \le \sup_{t,s\in[0,1]}\left|\overline{f}_n(t,s) - \varphi\big(t,s\big)\right| + \frac{1}{n}\sum_{i=1}^{n}\Delta_{2i},$$

которая вместе с леммами 1.37 и 1.38 влечет выполнение первого утверждения теоремы. Второе утверждение следует из первого и теоремы 1.15. □

2 Построение явных оценок в нелинейной регрессии

Постановка задачи следующая: наблюдения $X_1, \ldots, X_n$ представимы в виде

$$X_i = f(\boldsymbol{\theta}, \mathbf{z}_i) + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{249}$$

где f — некоторая известная функция, значения набора k-мерных регрессоров $\{\mathbf{z}_i\}$ (детерминированных или случайных) известны, погрешности $\{\varepsilon_i\}$ представляют собой последовательность центрированных ненаблюдаемых случайных величин с неизвестными распределениями, при этом наблюдения, регрессоры и погрешности могут зависеть от n. Задача состоит в построении в известном смысле явных оценок m-мерного параметра $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ ($\boldsymbol{\Theta}$ — открытое подмножество $\mathbb{R}^m$) по выборке $(k+1)$-мерных наблюдений $\{(\mathbf{z}_i, X_i);\ i = 1, \ldots, n\}$. Наличие таких оценок решает проблему поиска предварительных (начальных) оценок для ньютоновских одношаговых процедур оценивания параметров рассматриваемых моделей.

Структура главы следующая. В разделе 2.1 обсуждается один из подклассов задач нелинейной регрессии — внутренне линейные модели. В разделе 2.2 предложен способ оценивания параметра, основанный на использовании непараметрических оценок регрессионной функции. Раздел 2.3 посвящен описанию другого метода построения оценок, основанного на использовании сумм определенным образом взвешенных откликов и конструкций интеграла Римана. Оба новых подхода к оцениванию предполагают условие плотного заполнения регрессорами некоторой области.

Отметим, что в разделе 2.3.1 модели с детерминированными и случайными регрессорами рассматриваются отдельно. Дело в том, что условия состоятельности предлагаемых новых оценок легко сформулировать, как и в предыдущей главе, единообразно как для детерминированных, так и для случайных регрессоров. Но для доказательства асимптотической нормальности оценок в ситуации случайных регрессоров с неизбежностью приходится накладывать более жесткие ограничения на параметры модели. Поэтому нам иногда удобнее модели со случайными и детерминированными регрессорами рассматривать отдельно.

Результаты главы опубликованы в работах [363], [367], [362], [356], [361] и [364].

2.1 О внутренне линейных моделях

1. Традиционно в качестве *внутренне линейных* (intrinsically linear) или *внешне нелинейных* (nonintrinsically nonlinear) моделей рассматривают модели регрессии с мультипликативными погрешностями. Более того, существует точка зрения, что внутренне линейными могут быть только модели с мультипликативными погрешностями (см., например, [236, стр. 487]). Поэтому нам потребуется некоторое обобщение модели (7). А именно, считаем, что отклики

$\{X_i\}$ имеют структуру

$$X_i = f(\boldsymbol{\theta}, \mathbf{z}_i, \varepsilon_i), \qquad i = 1, \ldots, n, \tag{250}$$

где функция f известна, значения набора k-мерных регрессоров $\{\mathbf{z}_i\}$ также известны, погрешности $\{\varepsilon_i\}$ — ненаблюдаемые случайные величины. Общепринято, что регрессионная модель (250) является внутренне линейной, если отклики и параметр могут быть преобразованы в некоторую новую модель вида (250) с некоторой другой функцией f, являющейся линейной относительно нового параметра (см., например, [60], [236], [93], [223]). Отметим еще, что в известных нам работах размерность нового и исходного параметров предполагается одинаковой, а соответствующее преобразование между исходным и новым параметром — взаимно-однозначным (см., например, [223]). Таким образом, для внутренне линейных моделей явные оценки исходных параметров могут быть найдены, как правило, методами линейного регрессионного анализа.

Рассмотрим, например, следующую модель с мультипликативным шумом, приведенную в [236]:

$$X_i = \theta_1 e^{\theta_2 z_i} \varepsilon_i, \qquad i = 1, \ldots, n,$$

где $\{\varepsilon_i\}$ одинаково распределены и $\mathbb{E} \log \varepsilon_1 = 0$. Заменой переменных $\widetilde{X}_i = \log X_i$, $\widetilde{\theta}_1 = \log \theta_1$, $\widetilde{\varepsilon}_i = \log \varepsilon_i$ регрессионное уравнение преобразуется к виду

$$\widetilde{X}_i = \widetilde{\theta}_1 + \theta_2 z_i + \widetilde{\varepsilon}_i, \qquad i = 1, \ldots, n$$

с условием $\mathbb{E}\widetilde{\varepsilon}_i = 0$ при всех i. Таким образом, вышеприведенная модель является внутренне линейной. Далее, методом наименьших квадратов получаем оценки $\widetilde{\theta}^*_{n1}$ и θ^*_{n2} для параметров $\widetilde{\theta}_1$ и θ_2, а с помощью обратного преобразования — оценки для исходных параметров θ_1 и θ_2. Таким образом, $\theta^*_{n1} = \exp\{\widetilde{\theta}^*_{n1}\}$.

Аналогичным образом, при тех или иных ограничениях внутренне линейными являются, например, регрессионные модели

$$\begin{aligned} &X_i = \theta_1 z_i^{\theta_2} \varepsilon_i, && X_i = e^{\theta_1 + \theta_2 z_i + \varepsilon_i}, \\ &X_i = 1/\big(1 + \theta_1 e^{-\theta_2 z_i} \varepsilon_i\big), && X_i = \theta_1 z_{1i}^{\theta_2\theta_3} z_{2i}^{\theta_2(1-\theta_3)} z_{3i}^{\theta_2\theta_3(1-\theta_3)} g(\varepsilon_i), \end{aligned}$$

приведенные в [60], [236] и [223].

2. Определение внутренней линейности моделей мы предлагаем уточнить следующим образом. Будем говорить, что модель регрессии (250) *внутренне линейна*, если n-мерный вектор $(X_1, \ldots, X_n)$ может быть преобразован в N-мерный вектор $(\widetilde{X}_1, \ldots, \widetilde{X}_N)$ вида (250) с, возможно, бóльшим числом параметров, включающих исходные, при этом новая преобразованная регрессионная модель *с точностью до взаимно-однозначного соответствия своих параметров* является линейной. Последнее уточнение связано, в частности, с методикой получения оценок наименьших квадратов в моделях нелинейной регрессии: глобальный экстремум той или иной функции инвариантен относительно взаимно-однозначного преобразования ее переменных.

Подчеркнем, что в отличие от известных ранее работ предлагается преобразовывать весь вектор наблюдений, а не координаты этого вектора отдельно (см. приводимый далее пример 2.1). Кроме того, после преобразования вектора наблюдений допускается появление дополнительных неизвестных параметров, не являющихся функциями от основных исходных параметров (см. приводимые далее примеры 2.2 и 2.3). Отметим еще, что в уточненной трактовке внутренне-линейными оказываются и модели с аддитивными погрешностями (см. примеры 2.1–2.3).

П р и м е р 2.1. [6] Пусть

$$X_i = z_i\theta + r_i f(\theta) + \varepsilon_i, \qquad i = 1, \dots, n,$$

где z_i и r_i – известные числа, функция f может быть неизвестной. Положим

$$\widetilde{X}_i = \frac{X_{2i}}{r_{2i}} - \frac{X_{2i-1}}{r_{2i-1}}, \quad \widetilde{z}_i = \frac{z_{2i}}{r_{2i}} - \frac{z_{2i-1}}{r_{2i-1}}, \quad \widetilde{\varepsilon}_i = \frac{\varepsilon_{2i}}{r_{2i}} - \frac{\varepsilon_{2i-1}}{r_{2i-1}}, \quad i = 1, \dots, [n/2];$$

(без ограничения общности можно считать, что $r_j \neq 0$ и $\widetilde{z}_i \neq 0$ при всех j). Получаем простую неоднородную модель линейной регрессии

$$\widetilde{X}_i = \widetilde{z}_i\theta + \widetilde{\varepsilon}_i, \qquad i = 1, \dots, [n/2].$$

Стандартная оценка взвешенного метода наименьших квадратов в этом случае имеет вид

$$\theta_n^* = \sum_{i=1}^{[n/2]} w_i \widetilde{X}_i \widetilde{z}_i \left(\sum_{i=1}^{[n/2]} w_i \widetilde{z}_i^2 \right)^{-1},$$

где w_i – произвольные положительные веса. □

В примерах 2.2 и 2.3 считаем, что $\theta > 0$, $z_i \geq 0$, центрированные погрешности $\{\varepsilon_i\}$ независимы и одинаково распределены.

[6] Задача построения явных оценок для пяти моделей из примеров 2.1–2.4 рассматривалась в [250], см. также подробности в [331], [333], [341], [346]. Нетрудно показать, что построенные здесь оценки совпадают с оценками из указанных работ с точностью до переобозначений, но в [250] использовались несколько иные рассуждения, не связанные с внутренней линейностью этих моделей и со стандартными оценками метода наименьших квадратов в линейной регрессии.

П р и м е р 2.2. Пусть

$$X_i = (1+\theta z_i)^r + \varepsilon_i, \qquad i = 1, \ldots, n; \quad r = 1/2.$$

Возведем это регрессионное уравнение в квадрат: $X_i^2 = 1 + \theta z_i + \varepsilon_i^2 + 2\sqrt{1+\theta z_i}\varepsilon_i$. Легко видеть, что это есть неоднородная двумерная (по параметру) модель линейной регрессии

$$\widetilde{X}_i = \theta_0 + \theta z_i + \widetilde{\varepsilon}_i, \qquad i = 1, \ldots, n \tag{251}$$

с условием $\mathbb{E}\widetilde{\varepsilon}_i = 0$ при всех i, если определить новый параметр $\theta_0 = \mathbb{E}\varepsilon_1^2 + 1$ и положить $\widetilde{\varepsilon}_i = 2\varepsilon_i\sqrt{1+\theta z_i} + \varepsilon_i^2 - \mathbb{E}\varepsilon_1^2$, $\widetilde{X}_i = X_i^2$. Используя хорошо известные (см., например, [254]) оценки взвешенного метода наименьших квадратов с некоторыми известными весами $\{w_i\}$, $\sum_{i=1}^n w_i = 1$, получаем следующую оценку для интересующего нас параметра θ:

$$\theta_n^* = S_w^{-2}(z)\sum_{i=1}^n w_i\widetilde{X}_i(z_i - \overline{z}_w), \tag{252}$$

где

$$\overline{z}_w = \sum_{i=1}^n z_i w_i, \quad S_w^2(z) = \sum_{i=1}^n w_i z_i(z_i - \overline{z}_w) = \sum_{i=1}^n w_i z_i^2 - (\overline{z}_w)^2 = \sum_{i=1}^n w_i(z_i - \overline{z}_w)^2.$$

Подчеркнем, что условие $r = 1/2$ здесь существенно: если $r \neq 1/2$ и r не является положительным целым, то модель не является внутренне линейной. При $r = 1/2$ модель является внутренне линейной и в случае, когда $\{\varepsilon_i\}$ неодинаково распределены с условием $\mathbb{E}\varepsilon_i^2 = \sigma^2/w_{oi}$ при всех i и некоторых известных $\{w_{oi}\}$. □

П р и м е р 2.3. 1) Пусть

$$X_i = \log(1+\theta z_i) + \varepsilon_i, \qquad i = 1, \ldots, n.$$

Преобразуем уравнение следующим образом: $e^{X_i} = (1+\theta z_i)\big(e^{\varepsilon_i} - \mathbb{E}e^{\varepsilon_1}\big) + (1+\theta z_i)\mathbb{E}e^{\varepsilon_1}$. Положим $\widetilde{\varepsilon}_i = (1+\theta z_i)\big(e^{\varepsilon_i} - \mathbb{E}e^{\varepsilon_1}\big)$, $\widetilde{X}_i = e^{X_i}$ и введем новый параметр $\theta_0 = \mathbb{E}e^{\varepsilon_1}$. Получаем регрессионную модель

$$\widetilde{X}_i = \theta_0 + \theta\theta_0 z_i + \widetilde{\varepsilon}_i, \qquad i = 1, \ldots, n$$

с условием $\mathbb{E}\widetilde{\varepsilon}_i = 0$ при всех i. Обозначим $\theta_1 \equiv \theta_1(\theta, \theta_0) = \theta\theta_0$ и заметим, что двумерная функция $G(\theta_0, \theta) \equiv (\theta_0, \theta_1(\theta_0, \theta))$ является взаимно-однозначным отображением положительного квадранта $\{(\theta_0, \theta) : \theta_0 > 0, \theta > 0\}$ на себя. Далее, используя взвешенный метод наименьших квадратов с некоторыми известными весами $\{w_i\}$, $\sum_{i=1}^n w_i = 1$, найдем оценки $(\theta_{n0}^*, \theta_{n1}^*)$ двумерного параметра (θ_0, θ_1) линейной неоднородной модели регрессии $\widetilde{X}_i = \theta_0 + \theta_1 z_i + \widetilde{\varepsilon}_i$,

$i = 1, \ldots, n$. Далее, обратным преобразованием $G^{-1}(\theta_0, \theta_1)$ находим оценку для θ:

$$\theta_n^* = \theta_{n1}^*/\theta_{n0}^* == \frac{\sum_{i=1}^n w_i e^{X_i}(z_i - \overline{z}_w)}{S_w^2(z)\sum_{i=1}^n w_i e^{X_i} - \overline{z}_w \sum_{i=1}^n w_i e^{X_i}(z_i - \overline{z}_w)}.$$

Для внутренней линейности модели 1) существенно, чтобы величины $\{\mathbb{E}e^{\varepsilon_i}\}$ не зависели от i.

2) Пусть $X_i = (1 + \theta z_i)^r g(\varepsilon_i)$, $i = 1, \ldots, n$, при $r \neq 0$. Эта модель с мультипликативным шумом также является внутренне линейной, поскольку ввиду следующего тождества $r^{-1}\log X_i = \log(1+\theta z_i) + r^{-1}\log g(\varepsilon_i)$ она эквивалентна модели из 1). Для внутренней линейности здесь принципиально предположение о мультипликативной структуре погрешностей. $\square$

В модели (250), как и предлагаемом выше определении внутренней линейности, мы не конкретизировали стохастическую природу регрессоров $\{\mathbf{z}_i\}$ (хотя во всех вышеприведенных примерах предполагалось, что регрессоры детерминированы). Вообще говоря, регрессоры в модели (250) могут быть как детерминированными, так и случайными. Но если в определении внутренней линейности допустить, что стохастическая природа регрессоров после преобразования может измениться (например, регрессоры исходной модели детерминированы, а преобразованной — случайны), то внутренне-линейными оказываются и модели дробно–линейной регрессии (см. приводимые далее пример 2.4 и замечание 2.1).

П р и м е р 2.4. Пусть

$$X_i = \frac{r_i}{1 + \theta z_i} + \varepsilon_i, \qquad i = 1, \ldots, n,$$

где $\theta > 0$, регрессоры $r_i \geq 0$ и $z_i \geq 0$ неслучайны, а погрешности ε_i центрированы. Домножая правую и левую часть этого уравнения на знаменатель $(1 + \theta z_i)$, получаем регрессионную модель $r_i - X_i = X_i z_i \theta + (1 + \theta z_i)\varepsilon_i$. Положим $\widetilde{X}_i = r_i - X_i$, $\widetilde{z}_i = X_i z_i$ и $\widetilde{\varepsilon}_i = (1 + \theta z_i)\varepsilon_i$. В итоге исходное уравнение преобразуется к линейной неоднородной регрессионной модели

$$\widetilde{X}_i = \theta \widetilde{z}_i + \widetilde{\varepsilon}_i, \qquad i = 1, \ldots, n$$

со случайными регрессорами $\{\widetilde{z}_i\}$ и условием $\mathbb{E}\widetilde{\varepsilon}_i = 0$ при всех i. Используя взвешенный метод наименьших квадратов с некоторыми известными весами $\{w_i\}$, получаем следующую оценку для интересующего нас параметра:

$$\theta_n^* = \sum_{i=1}^n w_i \widetilde{X}_i \widetilde{z}_i \left(\sum_{i=1}^n w_i \widetilde{z}_i^2\right)^{-1}, \tag{253}$$

где $\{w_i\}$ – произвольные величины (возможно, случайные и не обязательно положительные). Отметим, что данная модель регрессии исследовалась в [341]. Если в представлении (253) положить $w_i = c_i/\widetilde{z}_i$ (здесь c_i — некоторые константы), то получаем оценку из [341].

З а м е ч а н и е 2.1. Нетрудно видеть, что в широких условиях внутренне линейными (в смысле предлагаемой в диссертации новой трактовки этого понятия) являются и следующие обобщения примеров 2.1–2.4 на случай многомерного основного параметра:

$$X_i = \sqrt{1 + \theta_1 z_{1i} + \ldots + \theta_m z_{mi}} + \varepsilon_i,$$
$$X_i = \log(1 + \theta_1 z_{1i} + \ldots + \theta_m z_{mi}) + \varepsilon_i,$$
$$X_i = \theta_1 z_{1i} + \ldots + \theta_m z_{mi} + r_{1i} f_1(\theta_1) + \ldots + r_{mi} f_m(\theta_m) + \varepsilon_i,$$
$$X_i = \frac{\theta z_i}{\theta + z_i} + \varepsilon_i,$$
$$X_i = \frac{r_{0i} + \theta_1 r_{1i} + \ldots + \theta_m r_{mi}}{\theta_1 z_{1i} + \ldots + \theta_m z_{mi}} + \varepsilon_i,$$

где значения векторов $(z_{1i}, \ldots, z_{mi})$ и $(r_{0i}, r_{1i}, \ldots, r_{mi})$, $i = 1, \ldots, n$, известны, а функции $f_1 \ldots, f_m$ могут быть неизвестны.

Задача построения явных оценок для логарифмической модели из этого замечания рассматривалась в [160], а для дробно–линейных моделей — в [339] и [340]. Кроме того, все примеры этого замечания упомянуты и в [250]. Но в указанных работах использовались несколько иные рассуждения, не связанные со стандартными оценками взвешенного метода наименьших квадратов в линейной регрессии. Хотя сам феномен преобразования дробно–линейной модели к линейной был отмечен в [339] и, по-сути, эта взаимосвязь двух моделей и использовалась при построении оценки. Нетрудно показать, что оценки, построенные благодаря внутренней линейности данных моделей, совпадают с оценками из указанных работ с точностью до переобозначений. □

З а м е ч а н и е 2.2. В примерах 2.1–2.3 полученные в результате преобразования новые шумы $\{\widetilde{\varepsilon}_i\}$ линейных уравнений являются разнораспределенными, но независимы и центрированы, что при широких условиях обеспечивает состоятельность или асимптотическую нормальность оценок θ_n^*. Например, рассмотрим оценки θ_n^* из примера 2.2.[7] Учитывая (251), (252) и очевидное соотношение $\sum\limits_{i=1}^{n} w_i(z_i - \overline{z}_w) = 0$, имеем

$$\theta_n^* - \theta = \frac{\sum_{i=1}^{n} w_i \widetilde{\varepsilon}_i (z_i - \overline{z}_w)}{S_w^2(z)} =$$
$$= \frac{\sum_{i=1}^{n} w_i 2\varepsilon_i \sqrt{1+\theta z_i}(z_i - \overline{z}_w)}{S_w^2(z)} + \frac{\sum_{i=1}^{n} w_i (\varepsilon_i^2 - \mathbb{E}\varepsilon_1^2)(z_i - \overline{z}_w)}{S_w^2(z)}. \quad (254)$$

Далее, из (254) получаем оценку

$$\mathbb{E}\Big(\sum_{i=1}^{n} \widetilde{\varepsilon}_i w_i (z_i - \overline{z}_w)\Big)^2 \le 8 w_n^* S_w^2(z)\left(\sigma^2 + \theta\sigma^2 \overline{z}_w + \mathbb{D}(\varepsilon_1^2)\right) + 8\theta\sigma^2 w_n^* \mu_w^3(z),$$

[7] Отметим, что эти оценки подробно исследовались в [346].

где $w_n^* = \max\limits_{i\le n} w_i$, $\mu_w^3(z) = \sum\limits_{i=1}^n w_i|z_i - \overline{z}_w|^3$, $\sigma^2 = \mathbb{E}\varepsilon_1^2$ и $\mathbb{D}(\varepsilon_1^2) = \mathbb{E}\varepsilon_1^4 - \sigma^4$. Следовательно, $D_n^2 \equiv \mathbb{E}(\theta_n^* - \theta)^2 \le 2w_n^* S_w^{-2}(z)\,(4\sigma^2 + \theta\overline{z}_w + \mathbb{D}(\varepsilon_1^2)) + 2w_n^*\theta\mu_w^3(z)S_w^{-4}(z)$. Другими словами, если $\beta_n^2 \equiv w_n^* S_w^{-2}(z)\bigl(1 + \overline{z}_w + \mu_w^3(z)S_w^{-2}(z)\bigr) \to 0$, то оценка θ_n^* состоятельна и $\beta_n^{-1}(\theta_n^* - \theta) = O_p(1)$. В частности, если $w_i = 1/n$ при всех i, то $\overline{z}_w$ можно считать выборочным первым моментом, а $S_w^2(z)$ и $\mu_w^3(z)$ — выборочными абсолютными центральными моментами второго и третьего порядка соответственно, построенными по выборке $z_1, \ldots, z_n$. Таким образом, если величины $\overline{z}_w$ и $\mu_w^3(z)$ ограничены, а $S_w^2(z)$ отделена от нуля равномерно по n, то $\beta_n^2 = O(1/n)$, т.е. оценка θ_n^* является $\sqrt{n}$-ограниченной.

Рассмотрим также вопрос об асимптотической нормальности θ_n^*. Покажем, что при широких ограничениях $D_n^{-1}(\theta_n^* - \theta) \xrightarrow{d} \mathcal{N}_1(0,1)$. Для этого достаточно установить, что имеет место следующее условие Ляпунова третьего порядка:

$$L_{3,n} \equiv \frac{\sum_{i=1}^n w_i^3 \mathbb{E}|\widetilde{\varepsilon}_i|^3 |z_i - \overline{z}_w|^3}{\left(\sum_{i=1}^n w_i^2 \mathbb{E}\widetilde{\varepsilon}_i^2 |z_i - \overline{z}_w|^2\right)^{3/2}} \to 0.$$

Для простоты предположим, что $\mathbb{E}\varepsilon_1^3 \ge 0$ и $\mathbb{E}\varepsilon_1^6 < \infty$. Тогда $\mathbb{E}\widetilde{\varepsilon}_i^2 \ge 4\sigma^2$ при всех i и

$$L_{3,n} \le C_0(1 + \mathbb{E}\varepsilon_1^6)\frac{(w_n^*)^2}{(w_n^o)^{3/2}} \cdot \frac{\mu_w^3(z) + \sum_{i=1}^n w_i(\theta z_i)^{3/2}|z_i - \overline{z}_w|^3}{\sigma^3 S_w^3(z)}, \tag{255}$$

где $w_n^o = \min\limits_{i\le n} w_i$ и C_0 некоторая абсолютная константа. Таким образом, если правая часть в (255) стремится к нулю, то оценка θ_n^* асимптотически нормальна. Например, если $w_i = 1/n$ при всех i, величина $S_w^2(x)$ ограничена от нуля равномерно по n и $\sup_n n^{-1}\sum\limits_{i=1}^n |z_i|^{9/2} < \infty$, то $L_{3,n} = O(1/\sqrt{n})$. $\square$

3. Иногда *внутренне-линейными* или *трансформирующимися к линейным* называют модели нелинейной регрессии вида (249) в случае, когда *регрессионную функцию* $f(\boldsymbol{\theta}, \mathbf{z})$ можно преобразовать к линейной относительно некоторых параметров (см., например, монографии [10, стр. 34], [38, стр. 153], а также [310], [93] и [281]). Отмечается, что указанное свойство трансформации регрессионной функции позволяет методами линейного регрессионного анализа оценивать параметры исходной регрессионной модели (или хотя бы находить стартовые точки-приближения параметров исходной модели [10, стр. 72]). Так, к внутренне-линейным в указанном смысле (см., например, [10]) часто относят модель Михаэлиса–Ментен

$$X_i = \theta_1 z_i/(\theta_2 + z_i) + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{256}$$

поскольку

$$\frac{1}{f(\boldsymbol{\theta}, z)} = \frac{1}{\theta_1} + \frac{\theta_2}{\theta_1}\frac{1}{z} \equiv \beta_1 + \beta_2 u$$

при $\beta_1 = 1/\theta_1$, $\beta_2 = \theta_2/\theta_1$, $u = 1/z$. Далее предлагается найти оценки β^*_{n1} и β^*_{n2}, минимизирующие следующую сумму квадратов

$$\min_{\beta_1,\beta_2} \sum_{i=1}^{n} (1/X_i - \beta_1 + \beta_2 u_i)^2, \tag{257}$$

и, окончательно, положить $\theta^*_{n1} = 1/\beta^*_{n1}$, $\theta^*_{n2} = \beta^*_{n2}/\beta^*_{n1}$.

Отметим, что указанный подход построения оценок (стартовых точек) не гарантирует даже свойство состоятельности получающихся таким способом оценок. Проиллюстрируем этот факт простейшим примером.

Рассмотрим модель (256) при $\theta_1 = 1$ и предположим, что $\theta_2 \equiv \theta \in (0, 1/2)$, а $z_i \geq 1$ при всех i. Пусть погрешности $\{\varepsilon_i\}$ независимы и, для простоты, имеют равномерное распределение на отрезке $[-1/2, 1/2]$. Реализуя указанный выше алгоритм из [10], имеем

$$\theta^*_n = \frac{1}{B_n} \sum_{i=1}^{n} (1/(X_i z_i) - 1/z_i) \qquad \text{при} \quad B_n = \sum_{i=1}^{n} z_i^{-2}. \tag{258}$$

Предложение 2.1. *В условиях, введенных выше, оценка θ^*_n из* (258) *несостоятельна. Более того, существуют регрессоры $\{z_i\}$ и последовательность $\delta_n \to \infty$ такие, что*

$$\liminf \mathbb{P}(\theta^*_n \geq \delta_n) > 0. \tag{259}$$

Таким образом, обсуждаемый подход построения оценок не корректен. Отметим еще, что состоятельные и асимптотически нормальные оценки для модели Михаэлиса–Ментен были предложены выше, а также в [340].

З а м е ч а н и е 2.3. Настороженность по-поводу предлагаемой в п.3 трансформации модели к линейной и последующего применения метода наименьших квадратов высказывается многими авторами (см, например, монографии [35] и [253], а также библиографические ссылки в [253]). Так, например, в [23] приведен численный пример, показывающий, что полученные обсуждаемым методом оценки оказываются противоречащими физическому смыслу параметров. Причина искажений кроется в существенной неоднородности погрешностей трансформированной модели и применение к этой модели обычного метода наименьших квадратов. Тот факт, что обсуждаемая трансформация регрессионной функции к линейной может не гарантировать построение подходящих оценок, упоминается также и в монографии [10]. В [345] доказана несостоятельность оценки Иохансена–Ламри для одного из параметров уравнения Михаэлиса–Ментен. Если обратиться к первоисточникам (см., например, [335]), то нетрудно видеть, что при построении указанных оценок используется конструкция (257) с дополнительными весовыми множителями под знаком суммы.

Д о к а з а т е л ь с т в о предложения 2.1. Заметим, прежде всего, что имеет место следующее тождество:

$$\theta_n^* - \theta = -\frac{1}{B_n}\sum_{i=1}^{n}\frac{1}{z_i}\frac{c_i^2\varepsilon_i}{1+c_i\varepsilon_i}, \tag{260}$$

где $c_i = (\theta + z_i)/z_i$. Далее, вычислим среднее каждого слагаемого, составляющего вышеприведенную сумму. Имеем

$$\mathbb{E}\left(\frac{c_i^2\varepsilon_i}{1+c_i\varepsilon_i}\right) = \int\limits_{-1/2}^{1/2}\frac{c_i^2 z}{1+c_i z}dz = c_i + \log\frac{1-c_i/2}{1+c_i/2}. \tag{261}$$

В условиях предложения

$$\frac{1}{2} < \frac{c_i}{2} = \frac{\theta+z_i}{2z_i} < \frac{3}{4}. \tag{262}$$

Используя теперь элементарную оценку $\log(1-t)/(1+t) \le -2t - 2t^3/3$ при всех $t \in (0,1)$, где $t = c_i/2 = (\theta+z_i)/(2z_i)$, из (261) и (262) получаем оценку сверху

$$\mathbb{E}\left(\frac{c_i^2\varepsilon_i}{1+c_i\varepsilon_i}\right) < -\frac{c_i^3}{12} < -\frac{1}{12}.$$

Следовательно, при всех n имеем

$$\mathbb{E}(\theta_n^* - \theta) = -\frac{1}{B_n}\sum_{i=1}^{n}\frac{1}{z_i}\mathbb{E}\left(\frac{c_i^2\varepsilon_i}{1+c_i\varepsilon_i}\right) > \frac{A_n}{12B_n}, \quad \text{где} \quad A_n = \sum_{i=1}^{n} 1/z_i. \tag{263}$$

С другой стороны, в силу (260) и (262), с вероятностью 1

$$|\theta_n^* - \theta| \le 9A_n/(2B_n). \tag{264}$$

Далее, из (263) легко получить, что

$$\mathbb{E}(\theta_n^* - \theta)I\left(\theta_n^* - \theta > A_n/(24B_n)\right) > A_n/(24B_n), \tag{265}$$

где $I(A)$ есть индикатор события A. Соотношения (264) и (265) влекут неравенство

$$\mathbb{P}\left(|\theta_n^* - \theta| > A_n/(24B_n)\right) > 1/108. \tag{266}$$

Нетрудно видеть, что $B_n \le A_n$ при $z_i \ge 1$. Следовательно, из (266) заключаем, что

$$\mathbb{P}\left(|\theta_n^* - \theta| > 1/24\right) > 1/108,$$

т.е. в условиях предложения оценка θ_n^* всегда несостоятельна. С другой стороны, если $A_n \to \infty$, но $\lim B_n < \infty$ (например, при $z_i = i$), то из (266) получаем соотношение (259). $\square$

Стоит отметить, что обсуждаемое в этом разделе свойство внутренней линейности задействует очень небольшое число моделей. Внутренняя линейность той или иной модели нелинейной регрессии (с аддитивными погрешностями) является скорее редким исключением, нежели правилом. Поэтому и возникает необходимость разработки подходов к построению явных оценок для собственно нелинейных моделей, к рассмотрению которых мы и переходим.

2.2 Построение оценок с использованием непараметрических методов

2.2.1 Методика построения оценок

В данном разделе рассматривается задача построения состоятельных оценок конечномерных параметров нелинейной регрессии с использованием различных ядерных оценок в моделях непараметрической регрессии. Рассматривается регрессионная модель (249), которую в этом разделе нам удобнее записать в виде

$$X_i = f_\theta(\mathbf{z}_i) + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{267}$$

где $\{f_\theta(\mathbf{t});\, \boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)^\top \in \boldsymbol{\Theta}\}$ — параметрическое семейство вещественнозначных непрерывных функций k переменных, $\boldsymbol{\Theta}$ – открытое подмножество $\mathbb{R}^m$, $\mathbf{t} = (t_1, \ldots, t_k)^\top \in \mathcal{P}$, где $\mathcal{P}$ – измеримое по Жордану компактное множество. Набор регрессоров $\{\mathbf{z}_i;\, i = 1, \ldots, n\}$ состоит из наблюдаемых случайных k-мерных векторов с возможно неизвестными распределениями и со значениями в $\mathcal{P}$, при этом не обязательно независимых или одинаково распределенных. Регрессоры мы рассматриваем в схеме серий, т.е. случайные векторы $\{\mathbf{z}_i;\ i = 1, \ldots, n\}$ могут зависеть от n. В частности, данная схема включает в себя модели с фиксированными регрессорами (например, схему эквидистантного плана эксперимента). Относительно случайных погрешностей $\{\varepsilon_i;\, i = 1, \ldots, n\}$ предполагаем выполненным условие (E_2), введенное в главе 1, т.е. $\{\varepsilon_i;\, i = 1, \ldots, n\}$ образуют последовательность мартингал-разностей с условием $M_p = \sup_{i \le n} \mathbb{E}|\varepsilon_i|^p < \infty$ при некотором $p > k$ и $p \ge 2$, где M_p не зависит от n. Предполагается также, что случайные величины $\{\varepsilon_i\}$ не зависят от $\{\mathbf{z}_i\}$, но могут зависеть от n. Задача состоит в оценивании параметра $\boldsymbol{\theta}$ по выборке $(m+1)$-мерных наблюдений $\{(X_i, \mathbf{z}_i);\, i = 1, \ldots, n\}$.

Далее в этом разделе мы предполагаем, что с вероятностью 1 каждая точка $\mathbf{z}_i$ в выборке имеет кратность 1. Отметим, что если в наборе $\{\mathbf{z}_i;\, i = 1, \ldots, n\}$ имелись кратные точки, то можно рассматривать среднее арифметическое откликов X_i с одинаковыми регрессорами и сводить задачу к исходной. Для каждого n обозначим через δ_n минимально возможный размер ε-сети, образованной набором регрессоров $\{\mathbf{z}_1, \ldots, \mathbf{z}_n\}$ в компактном множестве $\mathcal{P}$. Итак, единственным ограничением на регрессоры является следующее условие.

(D) *Имеет место предельное соотношение* $\delta_n \xrightarrow{p} 0$.

З а м е ч а н и е 2.4. Если все $\mathbf{z}_i$ не зависят от n, то сходимость по вероятности в приведенном условии будет эквивалентна сходимости почти наверное в силу монотонности последовательности $\{\delta_n\}$. Если $\{\mathbf{z}_i;\ i = 1, 2, \ldots\}$ – последовательность одинаково распределенных случайных векторов (не обязательно стационарная) с условием сильного перемешивания и куб $[0,1]^k \equiv \mathcal{P}$ представляет собой носитель маргинального распределения, то условие (D) будет выполнено. Примеры более сильной корреляции регрессоров, когда не выполняются все известные условия слабой зависимости, но имеет место условие (D), приведены в главе 1, в которой мы уже использовали подобные ограничения (см. условия (D_2) и (D_3), эквивалентные (D) и введенные в разделах 1.3 и 1.4). □

Перейдем к методологии получения оценок. Для простоты изложения далее считаем, что $\mathcal{P} = [0,1]^k$. Обозначим через $C[0,1]^k$ пространство непрерывных функций на $[0,1]^k$. Предположим, что в пространстве $C[0,1]^k$ задана некоторая норма $\|\cdot\|$ и далее мы будем рассматривать линейное нормированное пространство $(C[0,1]^k, \|\cdot\|)$. Нас будут интересовать, главным образом, два случая, когда

$$\|f\| \equiv \|f\|_{sup} = \sup_{\mathbf{t}\in[0,1]^k} |f(\mathbf{t})|, \qquad \|f\| \equiv \|f\|_{pw} = \sum_{i=1}^{\infty} \frac{|f(\mathbf{t}_i)|}{2^i},$$

где суммирование берется по всем точкам $\mathbf{t}_i \in [0,1]^k$ с рациональными координатами, занумерованным произвольным образом. Отметим, что сходимость в норме $\|\cdot\|_{pw}$ эквивалентна поточечной сходимости функций из $C[0,1]^k$.

В основе предлагаемого подхода лежат следующие два предположения.

(I) *существует непрерывное отображение* $\mathbf{G} : (C[0,1]^k, \|\cdot\|) \to \mathbb{R}^m$, *для которого векторная функция* $\mathbf{g}(\boldsymbol{\theta}) \equiv \mathbf{G}(f_\theta)$ *является гомеоморфизмом открытого множества* $\boldsymbol{\Theta}$ *на некоторую область пространства* $\mathbb{R}^m$;

(II) *существует* $\|\cdot\|$*-состоятельная (или сильно* $\|\cdot\|$*-состоятельная) непараметрическая оценка* $\widehat{f}_n(\mathbf{t}) \in C[0,1]^k$ *для регрессионной функции* $f_{\theta_0}(\mathbf{t})$, *где* $\boldsymbol{\theta}_0$ *— истинное значение параметра в модели* (267).

Определим оценку для параметра $\boldsymbol{\theta}$ соотношением

$$\boldsymbol{\theta}_n^* = \mathbf{g}^{-1}\big(\mathbf{G}\big(\widehat{f}_n\big)\big),$$

где $\mathbf{g}^{-1}$ — обратное преобразование для $\mathbf{g}$. Условимся, что мы по определению полагаем $\mathbf{g}^{-1}\big(\mathbf{G}\big(\widehat{f}_n\big)\big) = \mathbf{0}$, если $\mathbf{G}\big(\widehat{f}_n\big) \notin \{\mathbf{g}(\boldsymbol{\theta});\ \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$.

Справедливо следующее утверждение.

Теорема 2.1. *При выполнении предположений* (I) *и* (II) *оценка* $\boldsymbol{\theta}_n^*$ *является состоятельной* (*или сильно состоятельной в соответствии с условием в* (II)).

Д о к а з а т е л ь с т в о этого утверждения достаточно прозрачное. В самом деле, скажем, $\|\cdot\|$-состоятельность непараметрической оценки $\widehat{f}_n(\mathbf{t})$ означает, что

$$\left\|\widehat{f}_n - f_{\theta_0}\right\| \xrightarrow{p} 0.$$

Иными словами, с вероятностью, стремящейся к 1, в силу непрерывности преобразования $\mathbf{G}$ в норме $\|\cdot\|$ пространства $C[0,1]^k$, выполнено $\mathbf{G}\big(\widehat{f}_n\big) \in S_{g(\theta_0)}(\varepsilon)$ для любого наперед заданного $\varepsilon > 0$, где $S_{g(\theta_0)}(\varepsilon)$ — открытый шар радиуса ε с центром в точке $\mathbf{g}(\boldsymbol{\theta}_0)$ пространства $\mathbb{R}^m$. Поскольку образ открытого множества при гомеоморфном преобразовании будет открытым, то для достаточно малого $\varepsilon > 0$ имеет место вложение $S_{g(\theta_0)}(\varepsilon) \subseteq \{\mathbf{g}(\boldsymbol{\theta});\, \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$. Остается воспользоваться непрерывностью обратного преобразования $\mathbf{g}^{-1}$, откуда и следует, что $\boldsymbol{\theta}_n^* \xrightarrow{p} \boldsymbol{\theta}_0$. Аналогичные рассуждения используются и для доказательства $\|\cdot\|$-сильной состоятельности. $\square$

З а м е ч а н и е 2.5. Предложенная в теореме 2.1 методика построения оценок конечномерных параметров в задачах нелинейной регрессии близка к методологии метода моментов. Теорема 2.1 по сути предлагает приравнять значения регрессионной функции в тех или иных точках из области ее определения к соответствующим значениям ее состоятельной непараметрической оценки. При этом число таких уравнений (или что то же — указанных точек) должно совпадать с размерностью параметра регрессионной функции. Именно это происходит в методе моментов, когда для построения оценок параметров приравнивают истинные моменты (как функции от рассматриваемого параметра) к соответствующим выборочным моментам, которые, в свою очередь, будут состоятельными оценками для истинных моментов. Так что подход, предложенный в теореме 2.1, по аналогии с методом моментов может быть изложен следующим образом: в качестве оценки параметра $\boldsymbol{\theta}$ предлагается выбрать решение системы уравнений

$$f_\theta(\mathbf{t}_j) = \widehat{f}_n(\mathbf{t}_j), \quad j = 1, \ldots, \dim\theta, \tag{268}$$

где набор точек $\{\mathbf{t}_j\}$ подбирается так, чтобы эта система была разрешима единственным образом и обратное отображение было бы непрерывным. Здесь мы не планируем обсуждение вопроса оптимального выбора точек $\{\mathbf{t}_j\}$. $\square$

Отметим, что в качестве $\|\cdot\|$-состоятельной оценки $\widehat{f}_n(\mathbf{t})$ в предположении (II) (а также в теореме 2.1 и приводимых далее примерах 2.5, 2.6 и 2.7) можно рассмотреть либо одну из новых универсальных ядерных оценок (локально– постоянную $f^*_{n,h}(\mathbf{t})$ или локально–линейную

$\widetilde{f}^*_{n,h}(\mathbf{t})$, определенные соответственно в (96) и (113)), либо оценку Надарая–Ватсона $f^*_{NW}(\mathbf{t})$, определенную в (163), если рассматривать сходимость в норме $\|\cdot\|_{pw}$. Как уже было доказано, при выполнении лишь условия (D) эти три оценки являются состоятельными по вероятности в норме $\|\cdot\|_{pw}$, а новые оценки — и в норме $\|\cdot\|_{sup}$.

В качестве иллюстрации подхода рассмотрим несколько примеров широко известных в приложениях регрессионных моделей.

П р и м е р 2.5. Рассмотрим модель Михаэлиса–Ментен. В этом случае регрессионная функция в (267) задается равенством $f_\theta(t) = \theta_1 t/(\theta_2 + t)$, где $t \in \mathbb{R}_+$ и $\boldsymbol{\theta} = (\theta_1, \theta_2) \in \mathbb{R}^2_+$. Введем в рассмотрение следующее непрерывное отображение из $(C[0,1], \|\cdot\|_{pw})$ в $\mathbb{R}^2_+$:

$$\mathbf{G}(f) = (f(1), f(1/2)).$$

Теперь покажем, что суперпозиция $\mathbf{g}(\boldsymbol{\theta}) \equiv \mathbf{G}(f_\theta)$ — это гомеоморфизм открытого положительного квадранта $\mathbb{R}^2_+$ в открытый конус $C^2_+ = \{(r_1, r_2);\ r_2 > 0,\ r_2 < r_1 < 2r_2\}$. Нам достаточно доказать, что $\mathbf{g}(\boldsymbol{\theta})$ — биекция. В самом деле, рассмотрим систему уравнений

$$\frac{\theta_1}{1+\theta_2} = r_1, \qquad \frac{2^{-1}\theta_1}{2^{-1}+\theta_2} = r_2$$

при любом векторе $(r_1, r_2) \in C^2_+$. Эта система очевидным образом сводится к следующей системе двух линейных уравнений с двумя неизвестными θ_1 и θ_2:

$$\theta_1 - r_1\theta_2 = r_1, \qquad \theta_1 - 2r_2\theta_2 = r_2,$$

матрица которой невырождена всюду в вышеуказанном открытом конусе. В результате получаем

$$\theta^*_{n1} = r_1\left(1 + \frac{r_1 - r_2}{2r_2 - r_1}\right), \quad \theta^*_{n2} = \frac{r_1 - r_2}{2r_2 - r_1}. \tag{269}$$

Очевидно, что построенное взаимно-однозначное отображение двусторонне непрерывно, т.е. является гомеоморфизмом указанных выше областей.

Теперь можно рассмотреть в качестве $\widehat{f}_n(t)$ одну из ядерных оценок, состоятельных в норме $\|\cdot\|_{pw}$ при выполнении условия (D). Так что, например, можно положить в (269)

$$r_1 = \widehat{f}_n(1), \quad r_2 = \widehat{f}_n(1/2).$$

В этом случае двумерная оценка $\boldsymbol{\theta}^*_n = (\theta^*_{n1}, \theta^*_{n2})$ в (269) будет корректно определена на множестве элементарных исходов асимптотически полной меры с ростом n и, более того, будет состоятельной в силу теоремы 2.1. $\square$

П р и м е р 2.6. Рассмотрим модель (267) с регрессионной функцией $f_\theta(\mathbf{t}) = \theta_1 t_1^{\theta_2} t_2^{\theta_3}$, где $\mathbf{t} = (t_1, t_2) \in \mathbb{R}_+^2$ и $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) \in \mathbb{R}_+^3$ — векторы с положительными координатами. Это так называемая модель Кобба–Дугласа, достаточно популярная в эконометрике. Рассмотрим отображение из $(C[0,1]^2, \|\cdot\|_{pw})$ в $\mathbb{R}_+^3$

$$\mathbf{G}(f) = (f(2^{-1}, 2^{-1}), f(2^{-1}, 3^{-1}), f(3^{-1}, 3^{-1})),$$

которое, очевидно, будет непрерывным. Теперь покажем, что суперпозиция $\mathbf{g}(\boldsymbol{\theta}) \equiv \mathbf{G}(f_\theta)$ осуществляет гомеоморфизм $\mathbb{R}_+^3$ в коническую область

$$C_+^3 = \{(r_1, r_2, r_3) \in R_+^3 : 0 < r_3 < r_2 < r_1\}.$$

Нам достаточно доказать, что $\mathbf{g}(\boldsymbol{\theta})$ — биекция, так как непрерывность отображения очевидна. В самом деле, рассмотрим систему уравнений

$$\begin{cases} f_\theta(2^{-1}, 2^{-1}) \equiv \theta_1 2^{-\theta_2} 2^{-\theta_3} = r_1, \\ f_\theta(2^{-1}, 3^{-1}) \equiv \theta_1 2^{-\theta_2} 3^{-\theta_3} = r_2, \\ f_\theta(3^{-1}, 3^{-1}) \equiv \theta_1 3^{-\theta_2} 3^{-\theta_3} = r_3, \end{cases}$$

где (r_1, r_2, r_3) – произвольная точка в C_+^3. Логарифмируя эту систему уравнений, приводим ее к эквивалентной форме

$$\begin{cases} \widetilde{\theta}_1 - \theta_2 \log 2 - \theta_3 \log 2 = s_1, \\ \widetilde{\theta}_1 - \theta_2 \log 2 - \theta_3 \log 3 = s_2, \\ \widetilde{\theta}_1 - \theta_2 \log 3 - \theta_3 \log 3 = s_3, \end{cases}$$

где $\widetilde{\theta}_1 = \log \theta_1$, $s_j = \log r_j$, $j = 1, 2, 3$. Легко проверить, что матрица A этой системы линейных уравнений невырождена. Единственное ее решение легко находится методом последовательного исключения переменных:

$$\widetilde{\theta}_{n1}^* = \frac{s_1 \log 3 - s_3 \log 2}{\log(3/2)}, \quad \theta_{n2}^* = \frac{s_1 - s_2}{\log(3/2)}, \quad \theta_{n3}^* = \frac{s_2 - s_3}{\log(3/2)}.$$

Положим

$$s_1 = \log \widehat{f}_n(2^{-1}, 2^{-1}), \quad s_2 = \log \widehat{f}_n(2^{-1}, 3^{-1}), \quad s_3 = \log \widehat{f}_n(3^{-1}, 3^{-1}).$$

В силу сказанного на множестве элементарных исходов асимптотически полной меры для всех достаточно больших n выполнено двойное неравенство $s_1 > s_2 > s_3$, т.е. трехмерная оценка $(\widetilde{\theta}^*_{n1}, \theta^*_{n2}, \theta^*_{n3})$ корректно задана, и на основании теоремы 2.1 она будет состоятельной для трехмерного параметра $(\log\theta_1, \theta_2, \theta_3)$. Тем самым, оценка $\boldsymbol{\theta}^*_n = \big(\exp\{\widetilde{\theta}^*_{n1}\}, \theta^*_{n2}, \theta^*_{n3}\big)$ будет состоятельной для интересующего нас исходного параметра $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$. □

П р и м е р 2.7. Рассмотрим регрессионную функцию

$$f_\theta(\mathbf{t}) = \Big(1 + e^{-(\mathbf{t}, \boldsymbol{\theta})}\Big)^{-1},$$

где $\dim\mathbf{t} = \dim\boldsymbol{\theta} = m$, а $(\cdot,\cdot)$ – стандартное евклидово скалярное произведение в $\mathbb{R}^m$.

Для произвольного набора точек $\{\mathbf{t}_j;\ j = 1, \dots, m\}$ система уравнений (268) очевидным образом приводится к системе линейных уравнений вида

$$\begin{cases} (\mathbf{t}_1, \boldsymbol{\theta}) = r_1, \\ \dots\dots\dots\dots \\ (\mathbf{t}_m, \boldsymbol{\theta}) = r_m. \end{cases} \tag{270}$$

Единственное ограничение на векторы $\{\mathbf{t}_j\}$ — их линейная независимость, т.е. невырожденность матрицы T приведенной системы. Теперь положим

$$r_j = \log\frac{\widehat{f}_n(\mathbf{t}_j)}{1 - \widehat{f}_n(\mathbf{t}_j)},$$

где $\widehat{f}_n(\mathbf{t})$ — любая из ядерных оценок, определенных в (96), (113) или (163). Отметим, что в силу $\|\cdot\|_{pw}$-состоятельности указанных оценок при выполнении условия (D) имеет место неравенство $\widehat{f}_n(\mathbf{t}_j) < 1$ с вероятностью, стремящейся к 1 с ростом n для любого фиксированного $\mathbf{t}_j$. Если указанное неравенство нарушено, полагаем $r_j = 0$. Окончательно получаем следующую конструкцию состоятельной оценки для этой модели:

$$\boldsymbol{\theta}^*_n = \mathrm{T}^{-1}\cdot(r_1, \dots, r_m)^\top, \tag{271}$$

где матрица T определяется системой уравнений линейных уравнений (270). □

З а м е ч а н и е 2.6. Возможно, модели Михаэлиса–Ментен и Кобба–Дугласа не самые удачные для наших целей, поскольку нередко они относятся, по-видимому, к моделям, контролируемым экспериментатором. В частности, вопросы оптимального планирования для этих важных для приложений моделей исследовались в целом ряде работ (см., например, [14], [54], [56], [57], [106], [207], [270], [344]). Стоит отметить, что в оптимальных планах количество опорных точек нередко совпадает с размерностью параметра, но при этом в нелинейных моделях велика зависимость оптимального планирования от неизвестного параметра (см. ссылки во введении). Мы же используем иное предположение: набор регрессоров плотно заполняет некоторую область. Возможно, те или иные подходы к построению предварительных оценок могут быть полезны и с точки зрения оптимального планирования эксперимента в ситуации, когда нет никакой информации о неизвестном параметре и имеется возможность провести некоторое количество экспериментов для построения предварительной оценки (подобную рекомендацию можно найти, например, в [343]). Отметим еще, что при наличии растущей (с ростом n) кратности тех или иных точек из набора регрессоров задача построения предварительной оценки иногда может быть сведена к классической статистической постановке метода моментов (при определенных ограничениях на регресионную функцию).

Стоит отметить, что явные оценки для параметров модели Михаэлиса–Ментен были известны и ранее (см. [340]). В частности, оценки в работе [340] построены, по сути, за счет внутренней линейности этой модели и выполнение условия типа (D) в [340] не требуется. Однако принципиальным отличием этой работы от результатов предлагаемого исследования состоит в том, что в [340] регрессоры $\{\mathbf{z}_i\}$ неслучайны, а случайные погрешности $\{\varepsilon_i\}$ независимы. Другие явные оценки для моделей типа Кобба–Дугласа и модели Михаэлиса–Ментен будут предложены далее, в разделе 2.3.6 (см. примеры 2.16 и 2.17). □

2.2.2 α_n-состоятельность оценок

Состоятельные оценки $\boldsymbol{\theta}_n^*$ могут рассматриваться в качестве стартовых точек в итерационных процедурах и позволяют решать проблему множества корней. Но для того, чтобы использовать ту или иную оценку $\boldsymbol{\theta}_n^*$ в качестве предварительной в одношаговых процедурах оценивания, не достаточно лишь свойства состоятельности (требуется свойство состоятельности с некоторой скоростью). Обсудим вопрос об уточнении результатов теоремы 2.1. Напомним, что оценка $\boldsymbol{\theta}_n^*$ параметра $\boldsymbol{\theta} \in \mathbb{R}^m$ называется α_n-состоятельной, если $\alpha_n(\boldsymbol{\theta}_n^* - \boldsymbol{\theta}) \xrightarrow{p} \mathbf{0}$ и $\alpha_n \to \infty$. Это определение без труда переносится на случай бесконечномерных параметров.

Определение. Пусть $(\mathcal{C}, d)$ – сепарабельное метрическое пространство. Последовательность случайных элементов $g_n^* \in \mathcal{C}$ называется α_n-состоятельной оценкой случайного элемента $g \in \mathcal{C}$, если $\alpha_n d(g_n^*, g) \xrightarrow{p} 0$ и $\alpha_n \to \infty$.

В качестве следствия теоремы 2.1. приведем следующее утверждение.

Теорема 2.2. *Пусть в условиях теоремы* 2.1 *отображения* $\mathbf{G}$ *и* $\mathbf{g}^{-1}$ *удовлетворяют условию Липшица в своих пространствах и существует непараметрическая* α_n*-состоятельная (в норме пространства* $(C[0,1]^k, \|\cdot\|)$*) оценка* $\widehat{f}_n$ *для неизвестной регрессионной функции* f_{θ_o}*. Тогда оценка* $\boldsymbol{\theta}_n^*$ *будет* α_n*-состоятельной.*

Доказательство этого утверждения следует из представления $\boldsymbol{\theta}_n^* = \mathbf{g}^{-1}\big(G\big(\widehat{f}_n\big)\big)$ и вышеприведенного определения α_n-состоятельной непараметрической оценки $\widehat{f}_n$.

Как нетрудно видеть, основное условие в теореме 2.2 — существование α_n-состоятельной непараметрической оценки $\widehat{f}_n$. Следующее утверждение дает пример такой непараметрической оценки.

В следующем утверждении нам вместо условия (D) удобно использовать эквивалентное ограничение (D_2), введенное в главе 1.

Теорема 2.3. *Пусть в модели* (267) *для каждого* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ *регрессионная функция* $f_\theta(\mathbf{t})$ *удовлетворяет условию Липшица и не является постоянной функцией. Тогда при выполнении ограничений* (D_2), (E_2) *и* (K_2) *универсальная локально-постоянная ядерная оценка* $f^*_{n,h}(\mathbf{t})$, *определенная в* (96), *будет* α_n*-состоятельной при условии*

$$\alpha_n = o(h_n^{-1}), \qquad \text{где} \quad h_n = \left(\mathbb{E}(\delta_n^{kp/2})\right)^{\frac{1}{p(k/2+1)+k}}. \tag{272}$$

В качестве следствия теорем 2.2 и 2.3 получаем следующее утверждение.

Следствие 2.1. *Пусть в модели* (267) *для погрешностей и регрессоров выполнены условия* (E_2) *и* (D_2), *для каждого* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ *регрессионная функция* $f_\theta(\mathbf{t})$ *удовлетворяет условию Липшица и не является постоянной функцией. Пусть имеется непрерывное отображение* $\boldsymbol{G}$ *банахова пространства* $(C[0,1]^k, \|\cdot\|_{\sup})$ *в* $\mathbb{R}^m$, *для которого векторная функция* $\boldsymbol{g}(\boldsymbol{\theta}) \equiv \boldsymbol{G}(f_\theta)$ *является гомеоморфизмом открытого множества* $\boldsymbol{\Theta}$ *на некоторую область пространства* $\mathbb{R}^m$, *отображение* $\boldsymbol{G}$ *и обратное отображение* $\boldsymbol{g}^{-1}$ *удовлетворяют условию Липшица в своих пространствах. Тогда при выполнении условия* (K_2) *оценка* $\boldsymbol{\theta}^*_n = \boldsymbol{g}^{-1}\big(\boldsymbol{G}(f^*_{n,h})\big)$ *определена с вероятностью, стремящейся к* 1, *и является* α_n*-состоятельной, когда* $\alpha_n = o(h_n^{-1})$ *и* $h_n = \left(\mathbb{E}(\delta_n^{kp/2})\right)^{\frac{1}{p(k/2+1)+k}}$.

Таким образом, во всех вышеприведенных примерах при подстановке в полученные формулы непараметрической универсальной локально–постоянной ядерной оценки f^*_{n,h_n} мы с помощью следствия 2.1 получим α_n-состоятельные оценки для многомерных параметров рассматриваемых моделей нелинейной регрессии при $\alpha_n = o(h_n^{-1})$. Например, для оценки (271) это утверждение следует из того факта, что функция $\phi(z) = \log\frac{z}{1-z}$ в окрестности каждой точки открытого интервала $(0,1)$ удовлетворяет условию Липшица, и, кроме того, в силу (271) на множестве элементарных исходов асимптотически полной меры при всех достаточно больших n имеет место неравенство

$$\|\boldsymbol{\theta}^*_n - \boldsymbol{\theta}_0\| \le C \max_{j \le m} |f^*_n(\mathbf{t}_j) - f_{\theta_0}(\mathbf{t}_j)|, \tag{273}$$

где неслучайная постоянная C зависит от набора $\{\mathbf{t}_j\}$ и m.

З а м е ч а н и е 2.7. Если радиус δ_n-сети в условии (D) допускает детерминированную оценку сверху $\delta_n \le \widehat{\delta}_n$, то формула (272) для размера окна может быть преобразована как

$$h_n \le \widehat{h}_n = \widehat{\delta}_n^{\frac{k}{k+2+2/p}},$$

при этом при достаточно большом p (например, когда шум гауссов) порядок малости размера окна можно сделать сколь угодно близким к величине $\widehat{\delta}_n^{\frac{k}{k+2}}$. В качестве примера можно рассмотреть одномерный ($k = 1$) эквидистантный план эксперимента, когда в универсальной локально–постоянной оценке (24) нужно положить $z_{n:i} = i/n$ и $\Delta z_{ni} = 1/n$, при этом $\widehat{h}_n = \widehat{\delta}_n^{1/3} = n^{-1/3}$. Стоит также отметить, что в случае независимых одинаково распределенных регрессоров $\{\mathbf{z}_i;\ i = 1, \dots, n\}$ нетрудно установить факт асимптотической нормальности рассмотренных выше оценок, используя известные результаты об асимптотической нормальности соответствующих ядерных оценок. Например, при $k = 1$ нормировка невязки $|f_n^*(\mathbf{t}_j) - f_{\theta_0}(\mathbf{t}_j)|$ в широких условиях будет иметь порядок $\sqrt{nh_n}$ для некоторой стремящейся к нулю со степенной скоростью по n последовательности h_n, что является верхней границей (недостижимой) для нормирующего множителя α_n в определении α_n-состоятельности. □

Д о к а з а т е л ь с т в о теоремы 2.3. В разделе 1.3 главы 1 показано, что если в наших условиях в качестве такой оценки взять универсальную локально–постоянную ядерную оценку (96), то с вероятностью 1

$$\left\|\widehat{f}_{n,h} - f_{\theta_0}\right\|_{sup} \le \omega_{f_{\theta_0}}(h) + \zeta_n(h), \tag{274}$$

где $\omega_{f_{\theta_0}}(h)$ — модуль непрерывности регрессионной функции, а случайная величина $\zeta_n(h)$ имеет порядок

$$\zeta_n(h) = \widetilde{O}_p\left(\left(h^{-k(p/2+1)}\,\mathbb{E}(\delta_n^{kp/2})\right)^{1/p}\right).$$

Величина $h \equiv h_n$, выравнивающая порядки обоих слагаемых в правой части (274), находится как решение $h \equiv h_n$ уравнения $\mathbb{E}(\delta_n^{kp/2}) = h^{k(p/2+1)}\omega^p_{f_{\theta_0}}(h)$. Для липшицевой регрессионной функции в этом случае получаем представление для оптимального (по порядку малости) размера окна (272), что и требовалось показать. □

2.3 Построение оценок с помощью сумм взвешенных откликов

В данном разделе мы рассмотрим другой подход к оцениванию неизвестных параметров в нелинейной регрессии, который, как правило, позволяет получать оценки с более высокой скоростью сходимости (по сравнению с методом предыдущего раздела). Приведем ряд вспомогательных утверждений, которые нам потребуются.

2.3.1 Вспомогательные утверждения. Одномерный параметр

В дальнейшем нам будет удобно рассматривать регрессионную модель (249) в схеме серий с объемом наблюдений n в качестве параметра этой схемы. Так что в случае одномерного параметра мы будем предполагать следующее.

(**R**) *При каждом $n \geq 1$ наблюдения $X_{n1}, \ldots, X_{nn}$ представимы в виде*

$$X_{ni} = f_{ni}(\theta) + \varepsilon_{ni}, \qquad i = 1, \ldots, n,$$

где $\{\varepsilon_{ni}\}$ — ненаблюдаемые случайные погрешности, известные неслучайные функции $\{f_{ni}(\theta)\}$ заданы на $\Theta = (a, b)$, при этом границы a и b (одна или обе) могут быть бесконечными с соответствующими знаками.

Сформулируем ключевое утверждение для построения оценок в модели (R).

Теорема 2.4. *Пусть для модели (R) существуют числовая последовательность A_n, массив чисел $\{h_{ni};\ i = 1, \ldots, n\}$ и функции $T_n(\theta)$ такие, что*

$$\sum_{i=1}^{n} h_{ni} f_{ni}(\theta) - A_n - T_n(\theta) \to 0, \qquad \sum_{i=1}^{n} h_{ni}\varepsilon_{ni} \xrightarrow{p} 0, \tag{275}$$

при этом начиная с некоторого n функции $T_n(\theta)$ строго монотонны и непрерывны, а обратные функции $T_n^{-1}(t)$ равностепенно непрерывны. Тогда оценка

$$\theta_n^* = T_n^{-1}\left(\sum_{i=1}^{n} h_{ni} X_{ni} - A_n\right) \tag{276}$$

определена с вероятностью, стремящейся к 1, и состоятельна.

Если же функции $T_n^{-1}(t)$ удовлетворяют условию Гельдера с показателем $p \in (0, 1]$ и равномерно ограниченными по n константами, и для некоторой последовательности чисел $\alpha_n \to \infty$ выполнено

$$\alpha_n^{1/p}\left(\sum_{i=1}^{n} h_{ni} f_{ni}(\theta) - A_n - T_n(\theta)\right) \to 0 \qquad \text{и} \qquad \alpha_n^{1/p}\sum_{i=1}^{n} h_{ni}\varepsilon_{ni} \xrightarrow{p} 0, \tag{277}$$

то оценка θ_n^* является α_n-состоятельной.

Оценки вида (276) мы будем называть *явными*, если строго монотонные непрерывные функции $T_n(\theta)$ заданы явно. При этом нам не столь важно явное представления обратной функции $T_n^{-1}(t)$, поскольку значения этой функции могут быть легко вычислены с любой наперед заданной точностью.

З а м е ч а н и е 2.8. Утверждения теоремы 2.4 сохранятся, если функции $\{f_{ni}(\theta)\}$, $T_n(\theta)$, последовательность A_n и массив $\{h_{ni}\}$ случайны, при этом предельные соотношения в (275) или в (277) имеют место по вероятности, а дополнительные условия на функции $T_n(\theta)$ справедливы с вероятностью 1.

Теорема 2.5. *Пусть для модели* (R) *погрешности* $\{\varepsilon_{ni}\}$ *независимы и центрированы, выполнено условие* $0 < \mathbb{E}\varepsilon_{ni}^2 < \infty$ *при всех* $i \le n$ *и существуют числовая последовательность* A_n*, массив чисел* $\{h_{ni}\}$ *и последовательность функций* $T_n(\theta)$ *такие, что начиная с некоторого* n *функции* $T_n(\theta)$ *строго монотонны, непрерывно дифференцируемы и*

$$D_n^2 = \sum_{i=1}^{n} h_{ni}^2 \mathbb{E}\varepsilon_{ni}^2 \to 0, \quad D_n^{-1} \sum_{i=1}^{n} h_{ni}\varepsilon_{ni} \xrightarrow{d} \mathcal{N}_1(0,1), \tag{278}$$

$$D_n^{-1}\left(\sum_{i=1}^{n} h_{ni} f_{ni}(\theta) - A_n - T_n(\theta)\right) \to 0, \qquad T_n(\theta) \to T(\theta), \quad T'(\theta) \neq 0. \tag{279}$$

Тогда оценка θ_n^**, определенная в* (276)*, асимптотически нормальна*:

$$D_n^{-1}(\theta_n^* - \theta) \xrightarrow{d} \mathcal{N}_1\big(0, (T'(\theta))^{-2}\big). \tag{280}$$

З а м е ч а н и е 2.9. Утверждение теоремы 2.5 сохранится, если функции $\{f_{ni}(\theta)\}$, $T_n(\theta)$, последовательность A_n и массив $\{h_{ni}\}$ случайны и не зависят от $\{\varepsilon_{ni}\}$, при этом функция $T(\theta)$ неслучайна, вместо первого условия в (278) выполнено $\mathbb{E}D_n^2 \to 0$, первая сходимость в (279) имеет место по вероятности, а дополнительные условия на функции $T_n(\theta)$ справедливы с вероятностью 1. Условие независимости погрешностей $\{\varepsilon_{ni}\}$ в теореме 2.5 может быть ослаблено с соответствующим очевидным изменением условий в (278).

2.3.2 Вспомогательные утверждения. Многомерный параметр

Рассмотрим обобщение вышеприведенных результатов на случай m-мерного параметра $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$ в рамках следующей модели.[8]

($\bar{\mathbf{R}}$) *Пусть при каждом* $n \ge 1$ *наблюдения* $X_{n1}, \ldots, X_{nn}$ *представимы в виде*

$$X_{ni} = f_{ni}(\boldsymbol{\theta}) + \varepsilon_{ni}, \quad i = 1, \ldots, n, \tag{281}$$

где $\{\varepsilon_{ni}\}$ *— ненаблюдаемые случайные погрешности,* $\{f_{ni}(\boldsymbol{\theta})\}$ *— известные неслучайные функции, заданные на открытом множестве* $\boldsymbol{\Theta} = \prod_{j=1}^{m}(a_j, b_j) \subseteq \mathbb{R}^m$*, при этом для любого* j *границы* a_j *и* b_j *(одна или обе) могут быть бесконечными с соответствующими знаками.*

[8]Условимся, что всюду далее в этой главе векторы — это вектор-строки, а символ $\|\cdot\|_k$ обозначает евклидову векторную норму в $\mathbb{R}^k$.

Аналогами теорем 2.4 и 2.5 являются следующие два утверждения.

Теорема 2.6. *Пусть в условиях модели* $(\bar{R})$ *существуют массивы чисел* $\{A_{nj}; j = 1, \dots, m\}$ *и* $\{h_{ni}; i = 1, \dots, n\}$*, а также функции* $\{T_{nj}(\boldsymbol{\theta}); j = 1, \dots, m\}$ *и такое разбиение на группы* $[n_{j-1}+1, n_j] = [n_{j-1}(n)+1, n_j(n)]$*,* $j = 1, \dots, m$*, номеров наблюдений* $i = 1, \dots, n$ *(считаем, что* $n_0 = 0$ *и* $n_m = n$*), что при любом* $j = 1, \dots, m$

$$\sum_{i=n_{j-1}+1}^{n_j} h_{ni} f_{ni}(\boldsymbol{\theta}) - A_{nj} - T_{nj}(\boldsymbol{\theta}) \to 0 \quad \text{и} \quad \sum_{i=n_{j-1}+1}^{n_j} h_{ni}\varepsilon_{ni} \xrightarrow{p} 0, \tag{282}$$

при этом отображение $\mathbf{T}_n(\boldsymbol{\theta}) = (T_{n1}(\boldsymbol{\theta}), \dots, T_{nm}(\boldsymbol{\theta}))$ *взаимно однозначно при всех достаточно больших* n*, а обратные отображения* $\mathbf{T}_n^{-1}(\boldsymbol{\theta})$ *равностепенно непрерывны. Тогда оценка*

$$\boldsymbol{\theta}_n^* = \mathbf{T}_n^{-1}\left(\sum_{i=1}^{n_1} h_{ni}X_{ni} - A_{n1}, \dots, \sum_{i=n_{m-1}+1}^{n} h_{ni}X_{ni} - A_{nm}\right) \tag{283}$$

определена с вероятностью, стремящейся к 1*, и* $\boldsymbol{\theta}_n^* \xrightarrow{p} \boldsymbol{\theta}$*.*

Если дополнительно функции $\mathbf{T}_n^{-1}(\mathbf{t})$ *удовлетворяют условию Гельдера с показателем* $p \in (0, 1]$ *и равномерно ограниченными по* n *константами, и для некоторой последовательности чисел* $\alpha_n \to \infty$ *при любом* $j = 1, \dots, m$

$$\alpha_n^{1/p}\left(\sum_{i=n_{j-1}+1}^{n_j} h_{ni} f_{ni}(\boldsymbol{\theta}) - A_{nj} - T_{nj}(\boldsymbol{\theta})\right) \to 0 \quad \text{и} \quad \alpha_n^{1/p}\sum_{i=n_{j-1}+1}^{n_j} h_{ni}\varepsilon_{ni} \xrightarrow{p} 0, \tag{284}$$

то $\boldsymbol{\theta}_n^*$ *является* α_n*-состоятельной.*

Теорема 2.7. *Пусть в условиях теоремы* 2.6 *погрешности* $\{\varepsilon_{ni}\}$ *независимы и центрированы,* $0 < \mathbb{E}\varepsilon_{ni}^2 < \infty$ *при всех* $i \le n$*, функции* $T_{nj}(\boldsymbol{\theta})$ *непрерывно дифференцируемы и вместо предельных соотношений* (284) *для любого* $j = 1, \dots, m$

$$\begin{gathered} D_{nj}^2 = \sum_{i=n_{j-1}+1}^{n_j} h_{ni}^2 \mathbb{E}\varepsilon_{ni}^2 \to 0, \qquad D_{nj}^{-1}\sum_{i=n_{j-1}+1}^{n_j} h_{ni}\varepsilon_{ni} \xrightarrow{d} \mathcal{N}_1(0,1), \\ D_{nj}^{-1}\left(\sum_{i=n_{j-1}+1}^{n_j} h_{ni} f_{ni}(\boldsymbol{\theta}) - A_{nj} - T_{nj}(\boldsymbol{\theta})\right) \to 0, \qquad T_{nj}(\boldsymbol{\theta}) \to T_j(\boldsymbol{\theta}), \end{gathered} \tag{285}$$

при этом якобиан отображения $\mathbf{T}(\boldsymbol{\theta}) = (T_1(\boldsymbol{\theta}), \dots, T_m(\boldsymbol{\theta}))$ *всюду отличен от нуля. Тогда оценка* $\boldsymbol{\theta}_n^*$*, определенная в* (283)*, асимптотически нормальна:*

$$(\boldsymbol{\theta}_n^* - \boldsymbol{\theta})[\mathrm{T}'(\boldsymbol{\theta})]^\top \mathrm{D}_n^{-1} \xrightarrow{d} \mathcal{N}_m\big(\mathbf{0}, \mathrm{I}\big), \tag{286}$$

где $\mathrm{D}_n = \mathrm{diag}\{D_{n1}, \dots, D_{nm}\}$ *и* $\mathrm{T}'(\boldsymbol{\theta})$ *— матрица Якоби векторного отображения* $\mathbf{T}(\boldsymbol{\theta})$*.*

З а м е ч а н и е 2.10. Если якобиан отображения отличен от нуля в окрестности некоторой точки, то отображение взаимно-однозначно в окрестности этой точки. Но данное условие, выполненное для каждое точки области, не влечет существование глобального обратного отображения. Автору не известны общие условия, гарантирующие при $m > 1$ существование глобального обращения произвольного многомерного отображения. Обратимость того или иного отображения $\mathbf{T}(\boldsymbol{\theta})$ устанавливается непосредственно (см. пример 2.16).

З а м е ч а н и е 2.11. Утверждения, приведенные в замечаниях 2.8 и 2.9, с очевидными изменениями обобщаются на случай многомерного параметра.

З а м е ч а н и е 2.12. В теореме 2.6 вместо разбиения множества индексов $\{1, \ldots, n\}$ на непересекающиеся подмножества можно рассматривать и более общую ситуацию. А именно, пусть имеются m подмножеств $N_1, \ldots, N_m$ множества $\{1, \ldots, n\}$ (возможно, пересекающиеся), при этом в условиях (282) и (284) суммирование производится по индексам $i \in N_j$, $j = 1, \ldots, m$, а оценка $\boldsymbol{\theta}_n^*$ определяется соотношением

$$\boldsymbol{\theta}_n^* = \mathbf{T}_n^{-1} \left(\sum_{i \in N_1} h_{ni} X_{ni} - A_{n1}, \ldots, \sum_{i \in N_m} h_{ni} X_{ni} - A_{nm} \right).$$

Все остальные условия и утверждения в теореме 2.6 остаются в этом случае без изменений.

Нетрудно видеть, что можно получать и обобщения теоремы 2.7 на случай пересекающихся подмножеств $N_1, \ldots, N_m$ (а также зависимых погрешностей $\{\varepsilon_{ni}\}$), но нормирующая матрица D_n будет иметь уже более сложную структуру. □

Далее в разделах 2.3.4, 2.3.6, 2.3.7 и 2.3.9 мы покажем, каким образом для тех или иных классов регрессионных функций можно задавать величины, участвующие в построении оценки θ_n^* из (276) и ее многомерного аналога $\boldsymbol{\theta}_n^*$, определенного в (283), и приведем ряд следствий, вытекающих из теорем 2.4-2.7. Но прежде всего докажем вышеприведенные утверждения.

2.3.3 Доказательство утверждений разделов 2.3.1 и 2.3.2

Д о к а з а т е л ь с т в о теоремы 2.4. Состоятельность θ_n^* следует из условий (275) и равностепенной непрерывности последовательности функций $T_n^{-1}(t)$. В силу открытости множества $\Theta = (a, b)$ и равностепенной непрерывности рассматриваемых отображений оценка θ_n^* определена с вероятностью, стремящейся к 1. Следующая цепочка неравенств

$$|\theta_n^* - \theta| = \left| T_n^{-1} \left(\sum_{i=1}^n h_{ni} X_{ni} - A_n \right) - T_n^{-1}\big(T_n(\theta)\big) \right| \leq$$
$$\leq K \left| \sum_{i=1}^n h_{ni} X_{ni} - A_n - T_n(\theta) \right|^p \leq K \left| \sum_{i=1}^n h_{ni} f_{ni}(\theta) - A_n - T_n(\theta) \right|^p + K \left| \sum_{i=1}^n h_{ni} \varepsilon_{ni} \right|^p,$$

и оба условия в (277) влекут за собой α_n-состоятельность оценки θ_n^*. □

Д о к а з а т е л ь с т в о теоремы 2.5. В силу (278) и неравенства Чебышева имеет место соотношение $\sum\limits_{i=1}^{n} h_{ni}\varepsilon_{ni} \xrightarrow{p} 0$. Далее, с помощью формулы конечных приращений

$$\theta_n^* - \theta = T_n^{-1}\left(\sum_{i=1}^{n} h_{ni}X_{ni} - A_n\right) - T_n^{-1}\big(T_n(\theta)\big) =$$
$$= \big(T_n^{-1}(\widetilde{T}_n)\big)'\left(\sum_{i=1}^{n} h_{ni}f_{ni}(\theta) - A_n - T_n(\theta) + \sum_{i=1}^{n} h_{ni}\varepsilon_{ni}\right), \quad (287)$$

где $\widetilde{T}_n$ лежит между $\sum\limits_{i=1}^{n} h_{ni}X_{ni} - A_n$ и $T_n(\theta)$, а потому $\widetilde{T}_n \xrightarrow{p} T(\theta)$. Таким образом, $\big(T_n^{-1}(\widetilde{T}_n)\big)' = \big(T_n'\big(T_n^{-1}(\widetilde{T}_n)\big)\big)^{-1} \xrightarrow{p} \big(T'(\theta)\big)^{-1}$. Асимптотическая нормальность следует теперь из представления (287) и условий (278) и (279). □

Д о к а з а т е л ь с т в о теоремы 2.6. Аналогично одномерному случаю, состоятельность $\boldsymbol{\theta}_n^*$ непосредственно следует из условий (282) и равностепенной непрерывности обратного отображения $\mathbf{T}_n^{-1}(\mathbf{t})$. Следующая цепочка неравенств

$$\|\boldsymbol{\theta}_n^* - \boldsymbol{\theta}\|_m = \left\|\mathbf{T}_n^{-1}\left(\sum_{i=1}^{n_1} h_{ni}X_{ni} - A_{n1}, \ldots, \sum_{i=n_{m-1}+1}^{n} h_{ni}X_{ni} - A_{nm}\right) - \mathbf{T}_n^{-1}\big(\mathbf{T}_n(\boldsymbol{\theta})\big)\right\|_m \le$$
$$\le K\left\|\left(\sum_{i=1}^{n_1} h_{ni}X_{ni} - A_{n1} - T_{n1}(\boldsymbol{\theta}), \ldots, \sum_{i=n_{m-1}+1}^{n} h_{ni}X_{ni} - A_{nm} - T_{nm}(\boldsymbol{\theta})\right)\right\|_m^p \le$$
$$\le K\left\|\left(\sum_{i=1}^{n_1} h_{ni}f_{ni}(\boldsymbol{\theta}) - A_{n1} - T_{n1}(\boldsymbol{\theta}), \ldots, \sum_{i=n_{m-1}+1}^{n} h_{ni}f_{ni}(\boldsymbol{\theta}) - A_{nm} - T_{nm}(\boldsymbol{\theta})\right)\right\|_m^p +$$
$$+ K\left\|\left(\sum_{i=1}^{n_1} h_{ni}\varepsilon_{ni}, \ldots, \sum_{i=n_{m-1}+1}^{n} h_{ni}\varepsilon_{ni}\right)\right\|_m^p,$$

и условия из (284) влекут за собой α_n-состоятельность оценки $\boldsymbol{\theta}_n^*$. □

Д о к а з а т е л ь с т в о теоремы 2.7. В силу соотношений (285) и неравенства Чебышева имеем $\sum\limits_{i=1}^{n} h_{ni}\varepsilon_{ni} \xrightarrow{p} 0$. С помощью формулы Тейлора для векторных отображений получаем

$$\boldsymbol{\theta}_n^* - \boldsymbol{\theta} = \mathbf{T}_n^{-1}\Big(\sum_{i=1}^{n_1} h_{ni}X_{ni} - A_{n1}, \ldots, \sum_{i=n_{m-1}+1}^{n} h_{ni}X_{ni} - A_{nm}\Big) - \mathbf{T}_n^{-1}\big(\mathbf{T}_n(\boldsymbol{\theta})\big) =$$
$$= \Big(\sum_{i=1}^{n_1} h_{ni}f_{ni}(\boldsymbol{\theta}) - A_{n1} - T_{n1}(\boldsymbol{\theta}) + \sum_{i=1}^{n_1} h_{ni}\varepsilon_{ni}, \ldots, \sum_{i=n_{m-1}+1}^{n} h_{ni}f_{ni}(\boldsymbol{\theta}) - A_{nm} -$$
$$- T_{nm}(\boldsymbol{\theta}) + \sum_{i=n_{m-1}+1}^{n} h_{ni}\varepsilon_{ni}\Big)\int_0^1 \big[\mathrm{T}_n^{(-1)\,\prime}\big(\mathbf{T}_n(\boldsymbol{\theta}) + v(\mathbf{t}_n - \mathbf{T}_n(\boldsymbol{\theta}))\big]^\top dv,$$

где $\mathbf{t}_n = \Big(\sum\limits_{i=1}^{n_1} h_{ni}X_{ni} - A_{n1}, \ldots, \sum\limits_{i=n_{m-1}+1}^{n} h_{ni}X_{ni} - A_{nm} \Big)$, $\mathrm{T}_n^{(-1)\,\prime}(\mathbf{t})$ — матрица Якоби обратного векторного отображения $\mathbf{T}_n^{-1}(\mathbf{t})$. Для завершения доказательства остается учесть, что

$$\int_0^1 \mathrm{T}_n^{(-1)\,\prime}\big(\mathbf{T}_n(\boldsymbol{\theta}) + v(\mathbf{t}_n - \mathbf{T}_n(\boldsymbol{\theta}))\big)dv \xrightarrow{p} \big[\mathrm{T}'(\boldsymbol{\theta})\big]^{-1}.$$

Теорема доказана. □

2.3.4 Оценивание одномерного параметра

1. Рассмотрим следующий одномерный вариант регрессионной модели (249).

($\mathbf{R}_1$) *Наблюдения* $\{X_i\}$ *имеют структуру*

$$X_i = f(\theta, z_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{288}$$

где функция $f(\cdot,\cdot)$ *и числовая последовательность* $\{z_i\}$ *известны,* $\{\varepsilon_i\}$ *— ненаблюдаемые случайные погрешности и* $\theta \in \Theta = (a, b)$.

В первую очередь нам нужно привести модель к виду (R) при выполнении условий (275). Для этого при каждом фиксированном $n \geq 1$ упорядочим $z_1, \ldots, z_n$ по возрастанию: $z_{n:1} \leq \ldots \leq z_{n:n}$. В этом случае соответствующим образом перенумеруем отклики и погрешности (с сохранением условия независимости в случае независимости погрешностей). Обозначим отклики и погрешности, ассоциированные с i-ой порядковой статистикой $z_{n:i}$, через X_{ni} и ε_{ni} соответственно. Таким образом, модель (R_1) примет вид (R), где $f_{ni}(\theta) = f(\theta, z_{n:i})$.

Нам потребуется следующее центральное условие.

($\mathbf{C}_1$) *Имеет место соотношение* $\max\limits_{1 \leq i \leq n+1} \Delta z_{ni} \to 0$, *где*

$$\Delta z_{ni} = z_{n:i} - z_{n:i-1}, \quad z_{n:0} = c, \quad z_{n:n+1} = \begin{cases} d, & \text{\textit{если} } d < \infty \\ z_{n:n}, & \text{\textit{если} } d = \infty. \end{cases} \tag{289}$$

Отметим, что в случае $d < \infty$ предположение (C_1) означает, что набор $\{z_{n:i}\}$ образует измельчающееся разбиение отрезка $[c, d]$. Если $d = \infty$, то по умолчанию предполагается, что $z_{n:n} \to \infty$. Для простоты изложения мы не рассматриваем симметричный случай $c = -\infty$.

Построение оценки основано на том, что при выполнении (C_1) и минимальных дополнительных ограничениях справедливы соотношения

$$\sum_{i=1}^{n} f(\theta, z_{n:i})\Delta z_{ni} \to \int\limits_c^d f(\theta, z)dz \equiv T(\theta) \qquad \text{и} \qquad \sum_{i=1}^{n} \Delta z_{ni}\varepsilon_{ni} \xrightarrow{p} 0; \tag{290}$$

при этом интеграл Римана $T(\theta)$, понимаемый как несобственный при $d = \infty$, должен быть строго монотонной и непрерывной функцией на Θ. Тогда по теореме 2.4 при

$$h_{ni} = \Delta z_{ni}, \quad f_{ni}(\theta) = f(\theta, z_{n:i}), \quad A_n = 0 \quad \text{и} \quad T_n(\theta) = T(\theta)$$

оценка

$$\theta_n^* = T^{-1}\left(\sum_{i=1}^{n} \Delta z_{ni} X_{ni}\right) \tag{291}$$

будет состоятельной. Приведем утверждение касательно α_n-состоятельности и асимптотической нормальности оценок θ_n^*.

Следствие 2.2. *Пусть в условиях модели (R_1) функция $f(\theta, z)$ интегрируема по Риману по переменной z на $[c, d)$, а функция $T(\theta)$ из (290) строго монотонна и непрерывна. Кроме того, функция $T^{-1}(t)$ удовлетворяет условию Гельдера с показателем $p \in (0, 1]$ и для некоторой последовательности $\alpha_n \to \infty$*

$$\alpha_n^{1/p} \sum_{i=1}^{n} \Delta z_{ni} \varepsilon_{ni} \xrightarrow{p} 0, \quad \alpha_n^{1/p} \sum_{i=1}^{n} \omega_f(\Delta z_{ni}) \Delta z_{ni} \to 0, \quad \alpha_n^{1/p} \int\limits_{z_{n:n}}^{d} f(\theta, z) dz \to 0, \tag{292}$$

где $\omega_f(\Delta) = \sup_{t_1, t_2 : |t_1 - t_2| \le \Delta} |f(\theta, t_1) - f(\theta, t_2)|$. Тогда оценка θ_n^* из (291) определена с вероятностью, стремящейся к 1, и является α_n-состоятельной.

Следствие 2.3. *Пусть в условиях следствия 2.2 функция $T(\theta)$ непрерывно дифференцируема с ненулевой производной, центрированные погрешности $\{\varepsilon_i\}$ независимы, $0 < \mathbb{E}\varepsilon_i^2 < \infty$ при всех i, вместо первого условия в (292) выполнено*

$$D_n^2 = \sum_{i=1}^{n} \left(\Delta z_{ni}\right)^2 \mathbb{E}\varepsilon_{ni}^2 \to 0 \quad \textit{и} \quad D_n^{-1} \sum_{i=1}^{n} \Delta z_{ni} \varepsilon_{ni} \xrightarrow{d} \mathcal{N}_1(0, 1), \tag{293}$$

а остальные условия в (292) справедливы при замене $\alpha_n^{1/p}$ на D_n^{-1}, то θ_n^* асимптотически нормальна с параметрами, определенными в (280).

З а м е ч а н и е 2.13. Предположим, что элементы $\{z_{n:i}\}$ разбиваются на кластеры $\{z_{n:i} \in [c_1, d_1];\ i \in N_1\}, \ldots, \{z_{n:i} \in [c_k, d_k];\ i \in N_k\}$, где $\{N_j = N_j^{(n)}\}$ — последовательные отрезки (разбиение) конечного отрезка натурального ряда $1, \ldots, n$, а $[c_1, d_1], \ldots, [c_k, d_k]$ — попарно несовместные отрезки прямой, при этом $d_j < c_{j+1}$ для всех $j < k$ и допускается, что $d_k = \infty$. Тогда можно положить

$$T(\theta) = \sum_{j=1}^{k} \int\limits_{c_j}^{d_j} f(\theta, z) dz + \sum_{j=1}^{k-1} (c_{j+1} - d_j) f(\theta, d_j).$$

Приведем примеры известных регрессионных функций и соответствующих им $T(\theta)$, участвующих в построении θ_n^* из (291) и обладающих отмеченными выше свойствами монотонности и гладкости.

В примерах 2.8–2.14 считаем, что $\theta > 0$, $z \in [c, d]$, $c \geq 0$. По умолчанию считаем также, что $d < \infty$.

Пример 2.8. Пусть $f(\theta, z) = (1 + \theta z)^r$, $r \neq 0, -1$. В этом случае

$$T(\theta) = \int\limits_c^d (1 + \theta z)^r dz = \frac{(1 + d\theta)^{r+1} - (1 + c\theta)^{r+1}}{\theta(r+1)}.$$

В частности, при $c = 0$ и $d = 1$ имеем $T(\theta) = \theta^{-1}(r+1)^{-1}\big((1+\theta)^{r+1} - 1\big)$. Отметим также, что если $r < -1$, то можно рассматривать также случай $c = 0$ и $d = \infty$, для которого $T(\theta) = \theta^{-1}|r+1|^{-1}$, $T^{-1}(t) = t^{-1}|r+1|^{-1}$.

Пример 2.9. Пусть $f(\theta, z) = 1/(1 + \theta z)$. В этом случае

$$T(\theta) = \int\limits_c^d \frac{dz}{1 + \theta z} = \frac{1}{\theta} \log \frac{1 + d\theta}{1 + c\theta}.$$

При $c = 0$ и $d = 1$ имеем $T(\theta) = \theta^{-1} \log(1 + \theta)$. Очевидно, к рассмотренному случаю сводится модель, когда $f(\theta, z) = \theta z/(1 + \theta z)$.

Пример 2.10. Пусть $f(\theta, z) = 1/(\theta + z)$. В этом случае очевидным переобозначением параметра из предыдущего пункта получаем

$$T(\theta) = \int\limits_c^d \frac{dz}{\theta + z} = \log \frac{\theta + d}{\theta + c}, \qquad T^{-1}(t) = \frac{d - ce^t}{e^t - 1}.$$

Пример 2.11. Пусть $f(\theta, z) = 1/(1 + e^{-\theta z})$. В этом случае

$$T(\theta) = \int\limits_c^d \frac{dz}{1 + e^{-\theta z}} = (d - c) - \frac{1}{\theta} \log \frac{1 + e^{-\theta c}}{1 + e^{-\theta d}}.$$

При $c = 0$ и $d = 1$ имеем $T(\theta) = 1 + \theta^{-1} \log(1 + e^{-\theta})$. К этой модели преобразуется пример, когда $f(\theta, z) = z^\theta/(1 + z^\theta)$. В этом случае нужно перейти к новой переменной $\tilde{z} = \log z$ с соответствующим разбиением отрезка $[\log c, \log d]$ и рассмотреть интегральные суммы по приращениям $\Delta \tilde{z}_{ni}$.

Пример 2.12. Пусть $f(\theta, z) = e^{-\theta z}$. В этом случае

$$T(\theta) = \int\limits_c^d e^{-\theta z} dz = \theta^{-1}\big(e^{-\theta c} - e^{-\theta d}\big).$$

При $c = 0$ и $d = \infty$ имеем $T(\theta) = 1/\theta$ и $T^{-1}(t) = 1/t$. Аналогичные рассуждения справедливы для модели $f(\theta, z) = \theta^z \equiv e^{z \log \theta}$, $0 < \theta < 1$, $z \geq 0$.

П р и м е р 2.13. Пусть $f(\theta, z) = z^\theta$ и $c \leq 1$ и $d \leq 1$. В этом случае

$$T(\theta) = \int_c^d z^\theta dz = (\theta + 1)^{-1}\bigl(d^{\theta+1} - c^{\theta+1}\bigr).$$

При $c = 0$ и $d = 1$ имеем $T(\theta) = 1/(\theta + 1)$, $T^{-1}(t) = 1/t - 1$.

П р и м е р 2.14. Пусть $f(\theta, z) = \log(1 + \theta z)$. В этом случае

$$T(\theta) = \int_c^d \log(1 + \theta z)dz = \theta^{-1}\bigl((1 + d\theta)\log(1 + \theta d) - (1 + c\theta)\log(1 + \theta c)\bigr) - (d - c).$$

При $c = 0$ и $d = 1$ имеем $T(\theta) = \theta^{-1}(1 + \theta)\log(1 + \theta) - 1$.

Интеграл Римана $T(\theta)$ во всех приведенных выше примерах вычисляется в элементарных функциях, но в общем случае этого не требуется и значения $T(\theta)$ из (290), как и $T^{-1}(\cdot)$, легко могут быть вычислены приближенно с любой наперед заданной точностью.

З а м е ч а н и е 2.14. В следствиях 2.2 и 2.3 функция $f(\theta, z)$ предполагается непрерывной по второму аргументу на $[c, d)$. Подобное предположение позволяет использовать достаточно простые оценки близости интегральной суммы и соответствующего интеграла (см. первое соотношение в (290) и второе условие в (292)). Вместо указанного условия непрерывности можно предполагать, что функция $f(\theta, z)$ имеет конечное число разрывов. Но в этом в этом случае потребуется чуть более кропотливое доказательство и более громоздкие ограничения, хотя принципиальных проблем здесь нет.

З а м е ч а н и е 2.15. В некоторых случаях при построении оценки разумно несколько сократить выборку и убрать из рассмотрения часть регрессоров и соответствующих им откликов, чтобы обеспечить выполнение тех или иных условий. Например, если регрессоры образуют измельчающееся разбиение лишь части области своего задания (обозначим эту область плотного заполнения регрессорами через $A \subset \mathbb{R}$), то при построении оценки рекомендуется не использовать «лишние» регрессоры (все регрессоры вне области A) и соответствующие им наблюдения. Или, например, если функция $T(t) = \int_A f(t, z)dz$ не является строго монотонной и непрерывной, но существует подмножество $A_o \subset A$ такое, что функция $T_o(t) = \int_{A_o} f(t, z)dz$ строго монотонна и непрерывна, то разумно при построении оценки использовать лишь регрессоры, принадлежащие множеству A_o, и соответствующие им наблюдения. Множество A может быть не обязательно односвязным, набор регрессоров может разбиваться на кластеры с условием измельчения в каждом (см. замечание 2.13).

З а м е ч а н и е 2.16. Вернемся к требованиям (290) и рассмотрим некоторую более общую конструкцию. Предположим, что можно указать функцию $g : [c, d) \to \mathbb{R}$ такую, что

$$\sum_{i=1}^{n} f(\theta, z_{n:i}) g(z_{n:i}) \Delta z_{ni} \to \int_{c}^{d} f(\theta, z) g(z) dz \equiv T(\theta), \quad \sum_{i=1}^{n} g(z_{n:i}) \Delta z_{ni} \varepsilon_{ni} \xrightarrow{p} 0, \tag{294}$$

при этом интеграл Римана $T(\theta)$ обладает нужными нам свойствами (является строго монотонной и непрерывной функцией на Θ). Тогда в силу теоремы 2.4 при $h_{ni} = g(z_{n:i})\Delta z_{ni}$, $f_{ni}(\theta) = f(\theta, z_{n:i})$, $A_n = 0$ и $T_n(\theta) = T(\theta)$, оценку для параметра θ можно определить равенством

$$\theta_n^* = T^{-1}\left(\sum_{i=1}^{n} g(z_{n:i}) \Delta z_{ni} X_{ni}\right) \tag{295}$$

Использование дополнительной функции g может обеспечить выполнение тех или иных условий. Например, если интеграл $\int_c^d f(\theta, z)dz$ из (290) как функция от θ не обладает нужными свойствами, в то время как функция $T(\theta)$, определенная в (294), является строго монотонной и непрерывной. Кроме того, данное обобщение (вместе с замечанием 2.14) в известной степени формализует идею «отбрасывания» тех или иных наблюдений, упомянутую в замечании 2.15, если в качестве g рассматривать индикаторную функцию $g(z) = I(A_o)$. $\square$

2. В условиях модели (R_1) предлагаемый выше прием построения оценок параметра θ для некоторых регрессионных функций можно распространить на случай, когда $z_{n:n} \to \infty$, при этом функция $f(\theta, z)$ может быть неинтегрируемой по z на луче $[0, \infty)$ и может быть не выполнено условие $\max\limits_{1 \le i \le n} \Delta z_{ni} \to 0$. Вместо (C_1) в этом разделе, по сути, используется следующее условие.

$(\mathbf{C}_1^\infty)$ *Имеют место соотношения* $z_{n:n} \to \infty$ *и* $\max\limits_{1 \le i \le n} \Delta z_{ni}/z_{n:n} \to 0$, *где величины* Δz_{ni} *определены в* (289).

Это условие означает, что последовательность $\{z_{n:i}\}$ возрастает медленнее показательной функции.

Во всех утверждениях раздела считаем, что $\theta > 0$, $z_i \ge 0$, при этом $z_{n:n} \to \infty$. В следствиях 2.5 и 2.7 параметр θ отделен от нуля, т.е. $\theta > \varepsilon$ и $\varepsilon > 0$ известно.

Следствие 2.4. *Пусть* $f(\theta, z) = (1 + \theta z)^r$ *при* $r > 0$, *выполнено* (C_1^∞) *при* $c = 0$ *и имеет место предельное соотношение* $z_{n:n}^{-(r+1)} \sum\limits_{i=1}^{n} \Delta z_{ni} \varepsilon_{ni} \xrightarrow{p} 0$. *Тогда*

$$\theta_n^* = \left((r+1) z_{n:n}^{-(r+1)} \sum_{i=1}^{n} \Delta z_{ni} X_{ni}\right)^{1/r} \xrightarrow{p} \theta.$$

Следствие 2.5. *Пусть $f(\theta, z) = (1 + \theta z)^r$ при $r > 0$ и при $\alpha_n \to \infty$ выполнено*

$$\alpha_n z_{n:n}^{-2} \sum_{i=i_0}^{n} \bigl(\Delta z_{ni}\bigr)^2 \to 0, \qquad \alpha_n z_{n:n}^{-(r+1)} \sum_{i=i_0}^{n} \Delta z_{ni} \varepsilon_{ni} \xrightarrow{p} 0, \tag{296}$$

где $i_0 = \min\{i : 1 + \varepsilon z_{n:i} \geq \delta z_{n:n}\}$ при некотором $0 < \delta < \varepsilon$. Тогда оценка

$$\theta_n^* = T_n^{-1}\left(z_{n:n}^{-(r+1)} \sum_{i=i_0}^{n} \Delta z_{ni} X_{ni}\right) \quad \text{при} \;\; T_n(\theta) = \frac{(\theta + z_{n:n}^{-1})^{r+1} - \delta^{r+1}}{\theta(r+1)} \tag{297}$$

является α_n-состоятельной.

Если центрированные погрешности $\{\varepsilon_i\}$ независимы и вместо (296) *имеют место соотношения*

$$\forall i \quad 0 < \mathbb{E}\varepsilon_i^2 < \infty, \qquad D_n^2 = z_{n:n}^{-2(r+1)} \sum_{i=i_0}^{n} \bigl(\Delta z_{ni}\bigr)^2 \mathbb{E}\varepsilon_{ni}^2 \to 0,$$

$$D_n^{-1} z_{n:n}^{-2} \sum_{i=i_0}^{n} \bigl(\Delta z_{ni}\bigr)^2 \to 0, \qquad D_n^{-1} z_{n:n}^{-(r+1)} \sum_{i=i_0}^{n} \Delta z_{ni} \varepsilon_{ni} \xrightarrow{d} \mathcal{N}_1(0, 1), \tag{298}$$

то $D_n^{-1}(\theta_n^* - \theta) \xrightarrow{d} \mathcal{N}_1\bigl(0, (r+1)^2 (r\theta^{r-1} + \theta^{-2}\delta^{r+1})^{-2}\bigr)$.

П р и м е р 2.15. [9] Пусть в условиях следствия 2.4 или 2.5 погрешности $\{\varepsilon_i\}$ (а стало быть, и $\{\varepsilon_{ni}\}$) независимы и одинаково распределены и $z_{n:i} = i^\gamma$ при $\gamma > 0$. Тогда

$$\Delta z_{ni} \sim i^{\gamma-1}, \quad i_0 \sim \alpha n \quad \text{при} \quad \alpha = (\delta/\varepsilon)^{1/r} \in (0,1), \quad \sum_{i=i_0}^{n} (\Delta z_{ni})^2 \sim n^{2\gamma-1},$$

$$D_n^2 \sim 1/n^{1+2r\gamma}, \quad D_n^{-1} z_{n:n}^{-2} \sum_{i=i_0}^{n} \bigl(\Delta z_{ni}\bigr)^2 \sim 1/n^{1/2 - r\gamma}.$$

Кроме того,

$$\max_{i_0 \leq i \leq n} (\Delta z_{ni})^2 \left(\sum_{i=i_0}^{n} (\Delta z_{ni})^2\right)^{-1} \sim n^{-1},$$

что гарантирует (см., например, [326]) сходимость к нормальному закону в (298). Таким образом, оценка θ_n^* из (297) является асимптотически нормальной при $\gamma < (2r)^{-1}$, а также $D_n^{-1/2}$-состоятельной при $\gamma < 3(2r)^{-1}$. Более простая оценка θ_n^* из следствия 2.4 состоятельна при любом $\gamma > 0$. □

Следствие 2.6. *Пусть $f(\theta, z) = \theta z^r + g_1(\theta) z^{r_1} + \ldots + g_m(\theta) z^{r_m}$ при любых $r > r_1 > \ldots > r_m > 0$ и функциях $g_1, \ldots, g_m$, выполнено (C_1^∞) при $c = 0$ и имеет место предельное соотношение*

[9] В примере 2.15 запись вида $a_i \sim b_i$ понимается как $\lim_{i\to\infty} a_i/b_i = c > 0$.

$z_{n:n}^{-(r+1)} \sum\limits_{i=1}^{n} \Delta z_{ni} \varepsilon_{ni} \xrightarrow{p} 0$. *Тогда*

$$\theta_n^* = \left((r+1) z_{n:n}^{-(r+1)} \sum_{i=1}^{n} \Delta z_{ni} X_{ni} \right)^{1/r} \xrightarrow{p} \theta.$$

Для модели из следствия 2.6 можно строить и асимптотически нормальные оценки, по аналогии со следствием 2.5.

Рассмотри теперь логарифмическую регрессию.

Следствие 2.7. *Пусть* $f(\theta, z) = \log(1+\theta z)$, *центрированные погрешности* $\{\varepsilon_i\}$ *независимы,* $0 < \mathbb{E}\varepsilon_i^2 < \infty$ *при всех* i, *и*

$$D_n^2 = z_{n:n}^{-2} \sum_{i=i_0}^{n} \big(\Delta z_{ni}\big)^2 \mathbb{E}\varepsilon_{ni}^2 \to 0, \qquad D_n^{-1} z_{n:n}^{-2} \sum_{i=i_0}^{n} \big(\Delta z_{ni}\big)^2 \to 0,$$

$$D_n^{-1} z_{n:n}^{-1} \sum_{i=i_0}^{n} \Delta z_{ni} \varepsilon_{ni} \xrightarrow{d} \mathcal{N}_1(0,1),$$

где $i_0 = \min\{i : 1 + \varepsilon z_{n:i} \ge \delta z_{n:n}\}$ *и* $\delta \in (0,1)$ *таково, что* $\delta(1 + |\log \delta|) = \varepsilon$. *Тогда оценка*

$$\theta_n^* = T_n^{-1} \left(z_{n:n}^{-1} \sum_{i=i_0}^{n} \Delta z_{ni} (X_{ni} - \log z_{n:n}) \right)$$

при $T_n(\theta) = (1 + \theta^{-1} z_{n:n}^{-1})\big(\log(\theta + z_{n:n}) - 1\big) - \theta^{-1}(\delta \log \delta - \delta)$ *является асимптотически нормальной:*

$$D_n^{-1}(\theta_n^* - \theta) \xrightarrow{d} \mathcal{N}_1\big(0, \theta^4 (\theta - \varepsilon)^{-2}\big).$$

Состоятельная оценка (не обязательно асимптотически нормальная) для параметра логарифмической регрессии может быть найдена в следующей более простой форме.

Следствие 2.8. *Пусть* $f(\theta, z) = \log(1 + \theta z)$ *и вместе с условием* (C_1^∞) *при* $c = 0$ *выполнено*

$$z_{n:n}^{-1} \sum_{i=i_*}^{n} \Delta z_{ni} \varepsilon_{ni} \xrightarrow{p} 0,$$

где $i_* = \min\{i : z_{n:i} \ge \sqrt{z_{n:n} \max_{j \le n} \Delta z_{nj}}\}$. *Тогда имеет место соотношение*

$$\theta_n^* = \exp\left(1 + z_{n:n}^{-1} \sum_{i=i_*}^{n} \Delta z_{ni} (X_{ni} - \log z_{n:n}) \right) \xrightarrow{p} \theta.$$

З а м е ч а н и е 2.17. В следствии 2.5 при построении оценки θ_n^* не использовались наблюдения X_{ni} при $i < i_0$ по двум причинам: чтобы увеличить гладкость подынтегральной функции и чтобы расширить область выполнения условия Линдеберга. В следствиях 2.7 и 2.8 удаление наблюдений X_{ni} при $i < i_0$ или $i < i_*$ обусловлено особенностью функции $\log z$ в точке $z = 0$. Более того, возможна ситуация, когда условие (C_1^∞) будет выполнено, если предварительно убрать из рассмотрения элементы $z_{n:i}$ при близких к n значениях i (т.е. удалить элементы $z_{n:n-j}$, $j = 0, \ldots, k_n$ при $k_n = o(n)$) и изъять из рассмотрения соответствующие отклики X_{nn-j}.

Аналогично и в других утверждениях можно убирать из рассмотрения некоторое количество регрессоров и соответствующих им наблюдений, чтобы обеспечить выполнение тех или иных условий (см. замечание 2.15, а также замечание 2.23). □

З а м е ч а н и е 2.18. Подчеркнем, что модель логарифмической регрессии является внутренне–линейной только в случае, когда погрешности $\{\varepsilon_i\}$ независимы и одинаково распределены. Модель степенной регрессии является внутренне линейной лишь в случае, когда либо $r = 1/2$, либо r — положительное целое, при этом требуются дополнительные ограничения на погрешности $\{\varepsilon_i\}$ (см. подраздел 2.1).

З а м е ч а н и е 2.19. Отметим, что оценки, построенные предлагаемыми здесь приемами, отличаются от оценок метода наименьших квадратов в классической постановке задачи линейной регрессии. В самом деле, пусть $X_i = \theta z_i + \varepsilon_i$, погрешности $\{\varepsilon_i\}$ независимы и одинаково распределены, $\mathbb{E}\varepsilon_i = 0$ и $\mathbb{E}\varepsilon_i^2 = \sigma^2$.

Если выполнено условие (C_1) и $d < \infty$, то из теоремы 2.4 при $h_{ni} = \Delta z_{ni}$ и $A_n = 0$ получаем, что

$$\theta_n^* = 2(d^2 - c^2)^{-1} \sum_{i=1}^n \Delta z_{ni} X_{ni} \xrightarrow{p} \theta.$$

Если же выполнено условие (C_1^∞), то полагая $h_{ni} = \Delta z_{ni}/z_{n:n}^2$ и $A_n = 0$, из теоремы 2.4 следует, что

$$\theta_n^* = 2 z_{n:n}^{-2} \sum_{i=1}^n \Delta z_{ni} X_{ni} \xrightarrow{p} \theta.$$

Оценка метода наименьших квадратов для рассматриваемой линейной модели имеет другую структуру:

$$\theta_{n,MHK}^* = \sum_{i=1}^n X_i z_i \left(\sum\nolimits_{i=1}^n z_i^2\right)^{-1} \equiv \sum_{i=1}^n X_{ni} z_{n:i} \left(\sum_{i=1}^n z_{n:i}^2\right)^{-1}. \qquad \square$$

2.3.5 Доказательство утверждений раздела 2.3.4

Д о к а з а т е л ь с т в о следствий 2.2 и 2.3. Отметим, что если $\omega_f(\varepsilon) > 0$ при всех $\varepsilon > 0$, то условие (C_1) следует из (292). Воспользуемся утверждениями теорем 2.4 и 2.5 при $h_{ni} = \Delta z_{ni}$, $f_{ni}(\theta) = f(\theta, z_{n:i})$, $A_n = 0$ и $T_n(\theta) = T(\theta)$. В силу известной оценки погрешности приближения интеграла Римана интегральными суммами, имеем

$$\sum_{i=1}^{n} \Delta z_{ni} f(\theta, z_{n:i}) - \int\limits_{c}^{d} f(\theta, z) dz = \delta_n - \int\limits_{z_{n:n}}^{d} f(\theta, z) dz, \tag{299}$$

где $|\delta_n| \le \sum\limits_{i=1}^{n} \omega_f(\Delta z_{ni}) \Delta z_{ni}$. Следовательно, с учетом (292), выполнено

$$\alpha_n^{1/p} \left(\sum_{i=1}^{n} h_{ni} f_{ni}(\theta) - A_n - T(\theta) \right) = \alpha_n^{1/p} \left(\sum_{i=1}^{n} \Delta z_{ni} f(\theta, z_{n:i}) - \int\limits_{c}^{d} f(\theta, z) dz \right) \to 0,$$

т. е. справедливо первое условие в (277). Остальные условия теорем 2.4 и 2.5 проверяются очевидным образом. $\square$

Д о к а з а т е л ь с т в о следствий 2.4 и 2.6. Первое утверждение является частным случаем теоремы 2.4 при $h_{ni} = \Delta z_{ni} / z_{n:n}^{r+1}$, $f_{ni}(\theta) = f(\theta, z_{n:i})$, $T_n(\theta) = \theta^{-1} \int\limits_{0}^{\theta} z^r dz = \theta^r (r+1)^{-1}$ и $A_n \equiv 0$. Действительно, $T_n(\theta) \equiv T(\theta)$ — строго монотонная и непрерывная функция, при этом $T^{-1}(t) \equiv (r+1)^{1/r} t^{1/r}$ при $t > 0$. Далее, условия следствия гарантируют выполнение соотношений (275), если мы учтем равенство

$$\sum_{i=1}^{n} h_{ni}(\theta) f(\theta, z_{n:i}) - T(\theta) = \frac{1}{\theta} \sum_{i=1}^{n} \frac{\theta \Delta z_{ni}}{z_{n:n}} \left(\frac{1 + \theta z_{n:i}}{z_{n:n}} \right)^r - \frac{1}{\theta} \int\limits_{0}^{\theta} z^r dz =$$

$$= \frac{1}{\theta} \left(\sum_{i=1}^{n} \frac{\theta \Delta z_{ni}}{z_{n:n}} \left(\frac{1 + \theta z_{n:i}}{z_{n:n}} \right)^r - \int\limits_{z_{n:n}^{-1}}^{\theta + z_{n:n}^{-1}} z^r dz - \int\limits_{0}^{z_{n:n}^{-1}} z^r dz + \int\limits_{\theta}^{\theta + z_{n:n}^{-1}} z^r dz \right)$$

и используем следующую оценку $z_{n:n}^{-(\min\{r,1\}+1)} \sum\limits_{i=1}^{n} \left(\Delta z_{ni} \right)^{\min\{r,1\}+1} \to 0$ погрешности приближения интеграла Римана интегральными суммами для степенной функции, удовлетворяющей при $r \in (0,1)$ условию Гёльдера и условию Липшица при $r \ge 1$.

Для доказательства утверждения следствия 2.6 используем те же $h_{ni} = \Delta z_{ni} / z_{n:n}^{r+1}$ и $T(\theta)$, что и при доказательстве следствия 2.4. Теперь заметим, что

$$\sum_{i=1}^{n} h_{ni}(\theta) f(\theta, z_{n:i}) = \frac{1}{\theta} \sum_{i=1}^{n} \frac{\theta \Delta z_{ni}}{z_{n:n}} \left(\left(\theta \frac{z_{n:i}}{z_{n:n}} \right)^r + \sum_{k=1}^{m} \frac{1}{z_{n:n}^{r - r_k}} g_k(\theta) \left(\frac{z_{n:i}}{z_{n:n}} \right)^{r_k} \right).$$

Из вышеприведенных рассуждений следует, что

$$\frac{1}{\theta} \sum_{i=1}^{n} \frac{\theta \Delta z_{ni}}{z_{n:n}} \left(\theta \frac{z_{n:i}}{z_{n:n}} \right)^r \to T(\theta).$$

Из этих же соображений ясно, что оставшаяся сумма имеет порядок малости $O\left(\frac{1}{z_{n:n}^{r-r_1}}\right)$, что и требовалось показать. $\square$

Д о к а з а т е л ь с т в о следствия 2.5. Воспользуемся утверждением теоремы 2.5, полагая $A_n = 0$, $h_{ni} = 0$ при $i < i_0$ и $h_{ni} = \Delta z_{ni}/z_{n:n}^{r+1}$ при $i \geq i_0$. С учетом известной оценки погрешности приближения интеграла Римана интегральными суммами, имеем

$$T_n(\theta) = \frac{1}{\theta} \int\limits_{\delta}^{\theta + z_{n:n}^{-1}} z^r dz, \quad \frac{1}{D_n} \left(\frac{1}{\theta} \sum_{i=i_0}^{n} \frac{\theta \Delta z_{ni}}{z_{n:n}} \left(\frac{1 + \theta z_{n:i}}{z_{n:n}} \right)^r - T_n(\theta) \right) \to 0.$$

Нетрудно видеть, что $T_n(\theta) \to T(\theta)$, $T'(\theta) = (r+1)^{-1}(r\theta^{r-1} + \theta^{-2}\delta^{r+1}) \neq 0$, при этом $T_n'(\theta) > 0$ начиная с некоторых n, что гарантирует строгую монотонность $T_n(\theta)$. Остальные условия теоремы 2.5 выполнены очевидным образом, что доказывает асимптотическую нормальность θ_n^*. Первое утверждение следствия аналогичным образом извлекается из теоремы 2.4. $\square$

Д о к а з а т е л ь с т в о следствий 2.7 и 2.8 аналогично выводу следствий 2.4 и 2.5. Так для доказательства следствия 2.7 в теореме 2.5 нужно положить

$$h_{ni} = 0 \quad \text{при} \quad i < i_0, \qquad h_{ni} = \Delta z_{ni}/z_{n:n} \quad \text{при} \quad i \geq i_0,$$

$$A_n = \log z_{n:n} \sum_{i=i_0}^{n} \Delta z_{ni}/z_{n:n}, \qquad T_n(\theta) = \theta^{-1} \int\limits_{\delta}^{\theta + z_{n:n}^{-1}} \log z dz,$$

$$T(\theta) = \theta^{-1} \int\limits_{\delta}^{\theta} \log z dz \equiv \log\theta - \theta^{-1}(\delta \log\delta - \delta) - 1.$$

Поскольку

$$T'(\theta) = \theta^{-1}\big(1 - \theta^{-1}\delta(1 + |\log\delta|)\big) \equiv \theta^{-1}\big(1 - \varepsilon/\theta\big) > 0,$$

то функция $T(\theta)$ является строго монотонной, а функция $T_n(\theta)$ является строго монотонной, начиная с некоторых n. Далее, с нормировкой D_n^{-1} оценивается погрешность

$$\theta^{-1} \sum_{i=i_0}^{n} \frac{\theta \Delta z_{ni}}{z_{n:n}} \log\left(\frac{1 + \theta z_{n:i}}{z_{n:n}} \right) - T_n(\theta).$$

Для доказательства следствия 2.8 воспользуемся утверждением теоремы 2.4 при

$$h_{ni} = 0 \quad \text{при} \quad i < i_*, \qquad h_{ni} = \Delta z_{ni}/z_{n:n} \quad \text{при} \quad i \geq i_*,$$

$$A_n = \log z_{n:n} \sum_{i=i_*}^{n} \Delta z_{ni}/z_{n:n}, \qquad T(\theta) = \theta^{-1} \int\limits_{0}^{\theta} \log x dx = \log\theta - 1.$$

Полагая $\delta \equiv \delta_n = \sqrt{z_{n:n}^{-1} \max_{j\le n} \Delta z_{nj}}$, нам нужно только заметить, что погрешность замены интеграла $\int\limits_{\delta}^{c} \log x dx$ при $\delta < c$ римановой суммой вида $\sum\limits_{i: x_i \ge \delta} (x_i - x_{i-1}) \log x_i$ оценивается величиной $\delta^{-1} \max_i (x_i - x_{i-1})$, где $\{x_i\}$ – произвольное разбиение отрезка $[0, c]$. В нашем случае $x_i = z_{n:n}^{-1}(1 + \theta z_{n:i})$, $x_i - x_{i-1} \le \theta\delta^2$ и $c = \theta + z_{n:n}^{-1}$. Поэтому при выполнении условия (C_1^∞) имеем

$$\sum_{i=1}^{n} h_{ni}(\theta) f(\theta, z_{n:i}) - A_n - T(\theta) = z_{n:n}^{-1} \sum_{i=i_*}^{n} \Delta z_{ni} \log\left(\frac{1+\theta z_{n:i}}{z_{n:n}}\right) - T(\theta) =$$

$$= \theta^{-1} \sum_{i=i_*}^{n} \frac{\theta \Delta z_{ni}}{z_{n:n}} \log\left(\frac{1+\theta z_{n:i}}{z_{n:n}}\right) - \theta^{-1} \int\limits_{\delta}^{\theta + z_{n:n}^{-1}} \log x dx -$$

$$- \theta^{-1} \int\limits_{0}^{\delta} \log x dx + \theta^{-1} \int\limits_{\theta}^{\theta + z_{n:n}^{-1}} \log x dx \to 0,$$

что завершает доказательство следствия 2.8. □

2.3.6 Оценивание многомерного параметра

1. Основные идеи раздела 2.3.4 можно перенести на случай m-мерного параметра $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$. Нам потребуется следующее обобщение модели (R_1).

$(\bar{\mathbf{R}}_m)$ *Наблюдения* $\{X_i\}$ *имеют структуру*

$$X_i = f(\boldsymbol{\theta}, z_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{300}$$

где функция $f(\boldsymbol{\theta}, z)$ *и числовая последовательность* $\{z_i\}$ *известны,* $\{\varepsilon_i\}$ *— ненаблюдаемые случайные погрешности,* $\boldsymbol{\Theta} = \prod\limits_{j=1}^{m} (a_j, b_j) \subseteq \mathbb{R}^m$ *и при любом* j *границы* a_j *и* b_j *(одна или обе) могут быть бесконечными.*

Аналогично случаю $m = 1$, при каждом фиксированном $n \ge 1$ упорядочим $z_1, \ldots, z_n$ по возрастанию: $z_{n:1} \le \ldots \le z_{n:n}$. Далее соответствующим образом перенумеруем X_i и ε_i. В этом случае модель $(\bar{R}_m)$ примет вид $(\bar{R})$ при $f_{ni}(\boldsymbol{\theta}) = f(\boldsymbol{\theta}, z_{n:i})$.

В предположении, что выполнено условие (C_1), разобьем $\{z_{n:i}\}$ на m групп следующим образом: $z_{n:i} \in [c_{j-1}, c_j]$ при $i = n_{j-1} + 1, \ldots, n_j$, $j = 1, \ldots, m$ (считаем, что $c_0 = c$, $c_m = d$, $n_0 = 0$, $n_m = n$; здесь $n_j = n_j(n)$). Положим

$$T_j(\boldsymbol{\theta}) = \int\limits_{c_{j-1}}^{c_j} f(\boldsymbol{\theta}, z) dz, \quad j = 1, \ldots, m, \tag{301}$$

и будем использовать обозначение $\omega_f(\Delta) = \sup\limits_{|t_1 - t_2| \le \Delta} |f(\boldsymbol{\theta}, t_1) - f(\boldsymbol{\theta}, t_2)|$.

Из теорем 2.6 и 2.7 нетрудно извлечь

Следствие 2.9. *Пусть в модели* $(\bar{R}_m)$ *функция* $f(\boldsymbol{\theta}, z)$ *интегрируема по Риману по переменной* z *на* $[c, d]$, *функция* $\mathbf{T}(\boldsymbol{\theta}) = (T_1(\boldsymbol{\theta}), \ldots, T_m(\boldsymbol{\theta}))$ *гомеоморфно отображает* $\boldsymbol{\Theta}$ *на открытую область в* $\mathbb{R}^m$, *при этом обратная функция* $\mathbf{T}^{-1}(\mathbf{t})$ *удовлетворяет условию Гельдера с показателем* $p \in (0, 1]$. *Кроме того, пусть для некоторой последовательности* $\alpha_n \to \infty$ *и любом* $j = 1, \ldots, m$ *выполнено*

$$\alpha_n^{1/p} \sum_{i=n_{j-1}+1}^{n_j} \Delta z_{ni}\varepsilon_{ni} \xrightarrow{p} 0, \tag{302}$$

$$\alpha_n^{1/p} \sum_{i=n_{j-1}+1}^{n_j} \omega_f(\Delta z_{ni})\Delta z_{ni} \to 0, \qquad \alpha_n^{1/p} \int_{z_{n:n_j}}^{c_j} f(\boldsymbol{\theta}, z)dz \to 0. \tag{303}$$

Тогда оценка

$$\boldsymbol{\theta}_n^* = \mathbf{T}^{-1}\left(\sum_{i=1}^{n_1} \Delta z_{ni}X_{ni}, \ldots, \sum_{i=n_{m-1}+1}^{n} \Delta z_{ni}X_{ni}\right)$$

определена с вероятностью, стремящейся к 1, *и является* α_n-*состоятельной.*

Если дополнительно погрешности $\{\varepsilon_i\}$ *независимы и центрированы, при всех* i *выполнено* $0 < \mathbb{E}\varepsilon_i^2 < \infty$, *функции* $T_j(\boldsymbol{\theta})$ *непрерывно дифференцируемы по каждому аргументу, вместо* (302) *имеют место соотношения*

$$D_{nj}^2 = \sum_{i=n_{j-1}+1}^{n_j} \big(\Delta z_{ni}\big)^2 \mathbb{E}\varepsilon_{ni}^2 \to 0 \qquad \text{и} \qquad D_{nj}^{-1} \sum_{i=n_{j-1}+1}^{n_j} \Delta z_{ni}\varepsilon_{ni} \xrightarrow{d} \mathcal{N}_1(0, 1),$$

а условия (303) *выполнены при замене* $\alpha_n^{1/p}$ *на* D_{nj}^{-1}, *то* $\boldsymbol{\theta}_n^*$ *асимптотически нормальна с параметрами, определенными в* (286).

З а м е ч а н и е 2.20. Отметим, что отображение $\mathbf{T}(\boldsymbol{\theta})$ (тем самым, и оценка $\boldsymbol{\theta}_n^*$) задается не однозначно — все зависит от разбиения отрезка $[c, d]$ на m интервалов. В частности, это разбиение можно выбирать таким образом, чтобы обеспечить желаемые свойства для $\mathbf{T}(\boldsymbol{\theta})$.

2. В качестве иллюстрации подхода рассмотрим два примера — модель Михаэлиса–Ментен и частный случай модели Кобба–Дугласа.

Далее в двух примерах полагаем, что $\boldsymbol{\theta} = (\theta_1, \theta_2)$, $\theta_1, \theta_2 > 0$. Для определенности считаем, что $c = 0$, $c_1 = 1$ и $d = 2$.

П р и м е р 2.16. (Модель Михаэлиса–Ментен). Пусть $f(\boldsymbol{\theta}, z) = \theta_1 z/(\theta_2 + z)$. Тогда вышеупомянутые в примере 2.17 интегралы имеют вид

$$T_1(\boldsymbol{\theta}) = \theta_1\left(1 - \theta_2 \log\frac{\theta_2 + 1}{\theta_2}\right), \qquad T_2(\boldsymbol{\theta}) = \theta_1\left(1 - \theta_2 \log\frac{\theta_2 + 2}{\theta_2 + 1}\right).$$

Покажем, что двумерная функция $\mathbf{T}(\boldsymbol{\theta}) = (T_1(\boldsymbol{\theta}), T_2(\boldsymbol{\theta}))$ взаимно-однозначно отображает открытый положительный квадрант $Q_+ = \{(\theta_1, \theta_2) : \theta_1 > 0, \theta_2 > 0\}$ на открытый конус $C_+ = \{(t_1, t_2) \in Q_+ : t_2/3 < t_1 < t_2\}$. Иными словами, мы должны установить, что для любой точки $(t_1, t_2) \in C_+$ система уравнений

$$T_1(\boldsymbol{\theta}) = t_1, \qquad T_2(\boldsymbol{\theta}) = t_2 \tag{304}$$

имеет единственное решение. Отметим, что якобиан $J(\boldsymbol{\theta})$ отображения $\mathbf{T}(\boldsymbol{\theta})$ есть

$$J(\boldsymbol{\theta}) = \frac{\theta_1}{\theta_2 + 1}\left(\frac{3\theta_2 + 2}{\theta_2 + 2}\log\frac{\theta_2 + 1}{\theta_2} - \frac{2}{\theta_2 + 2} - \log\frac{\theta_2 + 2}{\theta_2 + 1}\right).$$

Положим $g(\theta_2) = \theta_1^{-1}(\theta_2 + 1)J(\boldsymbol{\theta})$. Нетрудно проверить, что $\lim_{x\to\infty} g(x) = 0$ и $g'(x) < 0$ при всех $x > 0$, откуда заключаем, что $g(x) > 0$, а значит и $J(\boldsymbol{\theta}) > 0$ при всех $\boldsymbol{\theta} \in Q_+$. Введем в рассмотрение функцию

$$\psi(x) = \frac{1 - x\log(\frac{x+1}{x})}{1 - x\log(\frac{x+2}{x+1})}, \;\; x > 0.$$

Легко видеть, что имеет место представление $\psi'(\theta_2) = -C(\boldsymbol{\theta})J(\boldsymbol{\theta})$, где $C(\boldsymbol{\theta}) > 0$ при всех $\boldsymbol{\theta} \in Q_+$. Иначе говоря, $\psi(x)$ – гладкая монотонно убывающая функция, при этом $\lim\limits_{x\to 0+}\psi(x) = 1$, $\lim\limits_{x\to\infty}\psi(x) = 1/3$. Из (304) получаем, что

$$\frac{T_1(\boldsymbol{\theta})}{T_2(\boldsymbol{\theta})} = \psi(\theta_2) = \frac{t_1}{t_2}.$$

Отсюда в силу вышеупомянутых свойств $\psi(x)$ следует существование единственного корня θ_2^* этого уравнения. Далее, из любого уравнения в (304) немедленно получаем и соответствующий единственный корень θ_1^*, что и требовалось показать.

Мы показали, что двумерное отображение $\mathbf{T}(\boldsymbol{\theta})$ является диффеоморфизмом между двумерными областями Q_+ и C_+, так что мы можем воспользоваться следствием 2.9. $\square$

П р и м е р 2.17. (Частный случай модели Кобба–Дугласа). Пусть $f(\boldsymbol{\theta}, z) = \theta_1 z^{\theta_2}$. В этом случае

$$T_1(\boldsymbol{\theta}) = \int\limits_0^1 f(\boldsymbol{\theta}, z)dz = \frac{\theta_1}{\theta_2 + 1}, \quad T_2(\boldsymbol{\theta}) = \int\limits_1^2 f(\boldsymbol{\theta}, z)dz = \frac{\theta_1}{\theta_2 + 1}(2^{\theta_2 + 1} - 1)$$

и $\mathbf{T}(\boldsymbol{\theta}) = (T_1(\boldsymbol{\theta}), T_2(\boldsymbol{\theta}))$ гомеоморфно отображает открытый положительный квадрант на открытый конус $\{(t_1, t_2) : 0 < t_1 < t_2\}$. Обратное преобразование имеет вид

$$\mathbf{T}^{-1}(t_1, t_2) = \Big(\frac{t_1}{\log 2}\log(1 + t_2/t_1), \frac{1}{\log 2}\log(1 + t_2/t_1) - 1\Big).$$

2.3.7 Случайные регрессоры

1. Рассмотрим следующую рандомизированную постановку.

($\mathbf{R}_1'$) *В условиях модели* (R_1) *набор регрессоров* $\{z_i,\ i=1,\dots,n\}$ *представляет собой совокупность, вообще говоря, зависимых и не обязательно одинаково распределенных случайных величин, при этом регрессоры могут рассматриваться в схеме серий.*

Результаты раздела 2.3.4 в широких условиях переносятся на модель (R_1'). Аналогом условия (C_1) является следующее предположение.

($\mathbf{C}_1'$) $\max\limits_{1\le i\le n+1} \Delta z_{ni} \xrightarrow{p} 0$, *где величины* Δz_{ni} *определены в* (289).

Имеет место следующее обобщение следствия 2.2 с почти дословным переносом соответствующего доказательства, если учесть еще замечание 2.8. Отметим, что предложение (R_1'), допускающее зависимость регрессоров от n, включает в себя и случай детерминированных регрессоров.

Следствие 2.10. *Пусть в условиях модели* (R_1') *выполнены предположения следствия* 2.2, *при этом предельные соотношения в* (292) *имеют место в смысле сходимости по вероятности. Тогда оценка* θ_n^* *из* (291) *определена с вероятностью, стремящейся к* 1, *и является* α_n*-состоятельной для параметра* θ.

З а м е ч а н и е 2.21. Нетрудно проверить, что если набор регрессоров $\{z_i\}$ не зависит от семейства $\{\varepsilon_i\}$, при этом исходные погрешности $\{\varepsilon_i\}$ независимы и одинаково распределены, то после случайной перестановки, ассоциированной с соответствующей перестановкой по возрастанию набора регрессоров, величины $\{\varepsilon_{ni}\}$ будут также независимыми. В этом случае (в предположении $\mathbb{E}\varepsilon_1^2<\infty$) достаточным условием для выполнения первой сходимости в (292) является, например, соотношение $\alpha_n^{2/p}\mathbb{E}D_n^2\to 0$, где случайная величина D_n^2 здесь и далее в этом разделе определена в (293).

З а м е ч а н и е 2.22. Предлагаемый метод позволяет строить оценки, не зависящие от распределений регрессоров. При этом зависимость элементов последовательности $\{z_i\}$ может быть довольно сильной (см. обсуждение и примеры в разделе 1.1.1). Нетрудно привести примеры последовательности одинаково распределенных ограниченных случайных величин $\{z_i\}$, которые не удовлетворяют усиленному закону больших чисел, но соответствующий вариационный ряд $\{z_{n:i}\}$ при $n\to\infty$ с вероятностью 1 будет образовывать измельчающееся разбиение общего носителя распределения (см. примеры 1.1 и 1.2).

З а м е ч а н и е 2.23. Напомним, что если $\{z_i\}$ одинаково распределены и носитель распределения z_1 представляет собой конечный отрезок $[c,d]$, то для выполнения (C_1') можно требовать, чтобы при всех $\delta \in (0, |d-c|)$

$$p_n(\delta) \equiv \sup_{|\Delta|=\delta} \mathbb{P}\left(\bigcap_{i\le n}\{z_i \notin \Delta\}\right) \to 0, \tag{305}$$

где супремум берется по всем интервалам $\Delta \subset [c,d]$ длины δ (см. замечание 1.4). Если элементы последовательности $\{z_i\}$ одинаково распределены и носитель распределения z_1 представляет из себя луч $[c,\infty)$, то условие (305) нужно требовать для отрезков $\Delta \subset [c,N]$ при всех положительных N. В этом случае диагональным методом можно выбрать последовательность $N = N(n) \to \infty$, для которой по-прежнему будет справедливо соотношение (305). В частности, для независимых или слабо зависимых $\{z_i\}$ из вышесказанного следует, что эти случайные точки будут измельчать с вероятностью 1 компакты $[c, N(n)]$ для некоторой достаточно медленно стремящейся к бесконечности последовательности $N(n)$. Иными словами, в этом случае можно удалить $k_n = o(n)$ крайних элементов $z_{n:n-j}$, $j = 0,\dots,k_n-1$ (т. е. в регрессионной модели изъять из рассмотрения соответствующие отклики X_{nn-j}) так, чтобы оставшиеся члены вариационного ряда $\{z_{n:i}\}$ образовывали бы измельчающееся разбиение расширяющихся компактов $[c, N(n)]$ с вероятностью 1.

Асимптотическую нормальность оценки θ_n^* из (291) в условиях модели (R_1') мы установим при некоторых упрощающих предположениях.

Следствие 2.11. *Пусть в условиях модели* (R_1') *погрешности* $\{\varepsilon_i\}$ *независимы, одинаково распределены, центрированы и не зависят от набора регрессоров* $\{z_i\}$, *функция* $f(\theta,z)$ *удовлетворяет условию Липшица по* z, *функция* $T(\theta)$ *строго монотонна и непрерывно дифференцируема, и справедливы соотношения*

$$\mathbb{E}|\varepsilon_1|^3 < \infty, \quad \mathbb{E}D_n^2 \to 0, \quad D_n^{-1}\max_{i\le n}\Delta z_{ni} \xrightarrow{p} 0, \quad D_n^{-1}\int_{z_{n:n}}^{d} f(\theta,z)dz \xrightarrow{p} 0. \tag{306}$$

Тогда имеет место сходимость (280).

Следствие 2.12. *Если в условиях следствия* 2.11 *набор регрессоров* $\{z_i\}$ *состоит из независимых одинаково распределенных на конечном отрезке* $[c,d]$ *случайных величин с функцией распределения* $F(t)$, *которая непрерывно дифференцируема с отделенной от нуля на* $[c,d]$ *производной* (*плотностью распределения*), *то*

$$nD_n^2 \xrightarrow{p} 2\sigma^2 \int_c^d (F'(x))^{-1}dx. \tag{307}$$

В частности,

$$\sqrt{n}(\theta_n^* - \theta) \xrightarrow{d} \mathcal{N}_1\big(0, B^2\big) \quad \textit{при} \quad B^2 = 2\sigma^2\big(T'(\theta)\big)^{-2} \int\limits_c^d (F'(x))^{-1} dx.$$

З а м е ч а н и е 2.24. Рассмотрим следующий частный случай. Предположим, что функция $f(\theta, t)$ непрерывна и ограничена, а также монотонно возрастает (или убывает) по θ при всех t. Обозначим $F_n^*(t) = n^{-1} \sum\limits_{i=1}^n I(z_i < t)$ и предположим, что с ростом n имеет место поточечная сходимость $F_n^*(t)$ к некоторой функции $F(t)$ во всех точках непрерывности последней. Например, это будет выполнено, если набор регрессоров $\{z_i\}$ представляет собой совокупность независимых (или слабо зависимых) одинаково распределенных случайных величин. Тогда свойством состоятельности обладает оценка

$$\theta_n^* = T_{n0}^{-1}\left(n^{-1} \sum_{i=1}^n X_i\right),$$

где $T_{n0}(\theta) = \int\limits_{-\infty}^{\infty} f(\theta, t) dF_n^*(t)$ строго монотонна и непрерывна. Кроме того, если функция F известна, то можно дополнительно рассматривать еще и оценку

$$\theta_n^* = T_0^{-1}\left(n^{-1} \sum_{i=1}^n X_i\right) \quad \text{при} \;\; T_0(\theta) = \int\limits_{-\infty}^{\infty} f(\theta, t) dF(t).$$

2. Нам потребуется следующий многомерный по параметру аналог рандомизированной модели (R_1').

$(\bar{\mathbf{R}}_m')$ *В условиях модели* $(\bar{R}_m)$ *набор регрессоров* $\{z_i\}$ *представляет собой совокупность, вообще говоря, зависимых и не обязательно одинаково распределенных случайных величин.*

В широких условиях результаты раздела 2.3.6 переносятся на модель $(\bar{R}_m')$ и, в частности, имеют место аналоги следствий 2.10-2.12. Приведем самое простое из таких утверждений.

Следствие 2.13. *Пусть в условиях модели* $(\bar{R}_m')$ *выполнены предположения первой части следствия* 2.9*, а предельные соотношения в* (303) *имеют место в смысле сходимости по вероятности. Тогда оценка* $\boldsymbol{\theta}_n^*$ *из следствия* 2.9 *определена с вероятностью, стремящейся к* 1*, и является* α_n*-состоятельной для параметра* $\boldsymbol{\theta}$*.*

2.3.8 Доказательство утверждений раздела 2.3.7

Д о к а з а т е л ь с т в о следствия 2.11. В силу замечания 2.9, воспользуемся утверждением теоремы 2.5 при $h_{ni} = \Delta z_{ni}$ и $A_n = 0$. Проводя рассуждения, аналогичные выводу следствия

2.9, нам достаточно установить справедливость второго соотношения в (293). Положим

$$\eta_n = D_n^{-1} \sum_{i=1}^{n} \Delta z_{ni} \varepsilon_i, \quad \gamma_t(x) = 2\mathbb{E}\left(\varepsilon_1^2 \int_0^1 (1-u)(e^{itxu\varepsilon_1} - 1)du\right)$$

и рассмотрим характеристическую функцию случайной величины η_n. Используя формулу Тейлора с остаточным членом в интегральной форме, имеем

$$\mathbb{E}e^{it\eta_n} = \mathbb{E}\mathbb{E}_{\{\varepsilon_i\}} \exp\left\{itD_n^{-1} \sum_{i=1}^{n} \Delta z_{ni}\varepsilon_i\right\} =$$
$$= \mathbb{E} \prod_{i=1}^{n} \left\{1 - \sigma^2 2^{-1} t^2 D_n^{-2} (\Delta z_{ni})^2 \left(1 + \sigma^{-2}\gamma_t\left(D_n^{-1}\Delta z_{ni}\right)\right)\right\}, \quad (308)$$

где $\sigma^2 = \mathbb{E}\varepsilon_1^2$, $\mathbb{E}_{\{\varepsilon_i\}}$ — оператор усреднения по набору случайных величин $\{\varepsilon_i\}$ при фиксации совокупности $\{z_i\}$. Далее, в силу оценки $|\gamma_t(x)| \le |tx|\mathbb{E}|\varepsilon_1|^3/3$ и теоремы Лебега о мажорируемой сходимости $\gamma_t(x) \to 0$ при $x \to 0$ для всех фиксированных t. С учетом оценки $\max\limits_{i\le n} |D_n^{-1}\Delta z_{ni}| \le \sigma^{-1}$ и представления для произведения

$$\prod_{i=1}^{n}(1+\beta_i) = \exp\left\{\sum_{i=1}^{n}\beta_i + O\left(\sum_{i=1}^{n}\beta_i^2\right)\right\} \quad \text{при} \quad \max_{i\le n}|\beta_i| \to 0$$

мы вправе вновь применить в (308) упомянутую теорему Лебега для внесения предела по n под знак $\mathbb{E}$. Таким образом, $\mathbb{E}e^{it\eta_n} \to e^{-t^2/2}$. □

Д о к а з а т е л ь с т в о следствия 2.12. Воспользуемся рассуждениями, приведенными в доказательстве леммы 1.5 главы 1. Без ограничения общности полагаем, что $z_{n:i} = F^{-1}(U_{n:i})$, где $F^{-1}(x)$ при $x \in [0,1]$ есть обратная функция к функции распределения $F(t)$, $U_{n:1} \le \ldots \le U_{n:n}$ – вариационный ряд, построенный по выборке $\{U_i\}$ равномерно распределенных величин. Нам понадобится следующее известное представление вариационного ряда по выборке объема n из равномерного распределения (см., например, [325]):

$$U_{n:k} \stackrel{d}{=} S_k S_{n+1}^{-1},$$

где $S_k = \sum\limits_{i=1}^{k} \tau_i$, $\{\tau_i\}$ — независимые случайные величины, экспоненциально распределенные с параметром 1. В силу неравенства Бернштейна для хвоста распределения максимума частичных сумм центрированных случайных величин с конечными экспоненциальными моментами (см., например, [225]), имеем

$$\mathbb{P}\left(\max_{k\le n}|S_k - \mathbb{E}S_k| \ge \varepsilon_n n\right) \le e^{-c\varepsilon_n\sqrt{n}}$$

для всех $n \geq n_0$ и любой последовательности $\varepsilon_n \to 0$ с условием $\varepsilon_n\sqrt{n} \to \infty$, где $c > 0$ — абсолютная постоянная. Отсюда с помощью леммы Бореля–Кантелли заключаем, что векторы

$$\{S_k/n;\, k = 1, \ldots, n\} \quad \text{и} \quad \{k/n;\, k = 1, \ldots, n\}$$

с вероятностью 1 будут сближаться в равномерной норме. В силу представления для $U_{n:k}$ и закона больших чисел для $\{\tau_i\}$ случайные векторы $\{U_{n:k};\, k = 1, \ldots, n\}$ и $\{k/n;\, k = 1, \ldots, n\}$ также будут сближаться по вероятности в равномерной норме. С помощью формулы конечных приращений и закона больших чисел получаем, что имеет место следующая эквивалентность по вероятности:

$$\sigma^{-2}D_n^2 = \sum_{i=1}^{n}(F^{-1}(U_{n:i}) - F^{-1}(U_{n:i-1}))^2 \sim \frac{1}{n^2}\sum_{i=1}^{n}\frac{\tau_i^2}{(F'(F^{-1}(U_{n:i})))^2} \sim$$
$$\sim \frac{1}{n^2}\sum_{i=1}^{n}\frac{\tau_i^2}{(F'(F^{-1}(i/n)))^2} = \frac{1}{n^2}\sum_{i=1}^{n}\frac{2}{(F'(F^{-1}(i/n)))^2} + \frac{1}{n}\delta_n, \quad (309)$$

где $\delta_n = n^{-1}\sum\limits_{i=1}^{n}(\tau_i^2 - 2)(F'(F^{-1}(i/n)))^{-2}$. При выводе (309) мы также учли, что в наших условиях функция $(F'(F^{-1}(x)))^2$ будет равномерно непрерывной на $[0, 1]$, совокупность точек $\{U_{n:i};\, i = 1, \ldots, n\}$ образует измельчающееся разбиение отрезка $[0, 1]$ с вероятностью 1. Нетрудно видеть, что в силу неравенства Чебышева $\delta_n \xrightarrow{p} 0$. Так что величина nD_n^2 эквивалентна по вероятности интегральной сумме Римана, сходящейся к соответствующему интегралу в (307). $\square$

2.3.9 Многомерные регрессоры

Рассмотрим более общую по сравнению с предложенной выше схему построения явных оценок для неизвестных параметров регрессионных моделей вида (R_1) или $(\bar{R}_m)$ в случае многомерных регрессоров $\{\mathbf{z}_i\}$. В условиях модели $(\bar{R}_m)$ считаем, что наблюдения $\{X_i\}$ имеют структуру

$$X_i = f(\boldsymbol{\theta}, \mathbf{z}_i) + \varepsilon_i, \qquad \mathbf{z}_i \in \mathcal{P} \subset \mathbb{R}^k, \qquad i = 1, \ldots, n, \quad (310)$$

где множество $\mathcal{P}$ измеримо по Жордану.

Как и в главе 1, в основе построения оценок лежит конструкция кратного интеграла Римана и соответствующих интегральных сумм Римана. Далее полагаем, что векторы $\{\mathbf{z}_i\}$ попарно различны. Аналогом свойства (C_1) «плотного» заполнения регрессорами некоторой области предлагается следующее условие, уже фигурировавшее в главе 1.

$(\mathbf{C}_k)$ *Для каждого n существует такое разбиение множества $\mathcal{P}$ на n измеримых по Жордану подмножеств $\{\mathcal{P}_i,\ i = 1, \ldots, n\}$, что каждый элемент этого разбиения содержит* *ровно по одной точке из набора* $\{\mathbf{z}_i\}$ *(нумерация элементов разбиения такова, что* $\mathbf{z}_i \in \mathcal{P}_i$*), при этом* $\max\limits_{i\le n} d(\mathcal{P}_i) \to 0$*, где* $d(A) = \sup\limits_{\mathbf{x},\mathbf{y}\in A} \|\mathbf{x}-\mathbf{y}\|_k$ *– диаметр множества.*

Опишем вкратце схему построения оценок. В случае скалярного параметра θ нас интересует предельное поведение сумм в правой части тождества

$$\sum_{i=1}^{n} X_i\Lambda_k(\mathcal{P}_i) = \sum_{i=1}^{n} f(\theta,\mathbf{z}_i)\Lambda_k(\mathcal{P}_i) + \sum_{i=1}^{n} \varepsilon_i\Lambda_k(\mathcal{P}_i),$$

где, напомним, $\Lambda_k(\cdot)$ есть мера Лебега в $\mathbb{R}^k$. Если при всех $\theta\in\Theta$ функция $f(\theta,\mathbf{z})$ как функция k-мерного аргумента $\mathbf{z}$ интегрируема по Риману на множестве $\mathcal{P}$ (скажем, она непрерывна и ограничена по $\mathbf{z}$), то с учетом (C_k) справедливо предельное соотношение

$$\sum_{i=1}^{n} f(\theta,\mathbf{z}_i)\Lambda_k(\mathcal{P}_i) \to \int\limits_{\mathcal{P}} f(\theta,\mathbf{z})d\mathbf{z} = T(\theta),$$

где величина $T(\theta)$ представляет собой кратный интеграл Римана. Далее, поскольку в силу (C_k) выполнено

$$\sum_{i=1}^{n} \Lambda_k^2(\mathcal{P}_i) \le \Lambda_k(\mathcal{P})\left(\max_{i\le n} d(\mathcal{P}_i)\right)^k \to 0,$$

то в случае центрированных погрешностей $\{\varepsilon_i\}$ условие $\sup_n \max_{i\le n}\mathbb{E}\varepsilon_i^2 < \infty$ гарантирует сходимость $\sum_{i=1}^n \varepsilon_i\Lambda_k(\mathcal{P}_i) \xrightarrow{p} 0$. Если функция $T(\theta)$ строго монотонна и непрерывна, то состоятельную оценку можно задать формулой

$$\theta_n^* = T^{-1}\left(\sum_{i=1}^{n} X_i\Lambda_k(\mathcal{P}_i)\right). \tag{311}$$

В широких условиях оценка θ_n^* будет α_n-состоятельной или асимптотически нормальной. Точные условия извлекаются из теорем 2.4 или 2.5 при $h_{ni} = \Lambda_k(\mathcal{P}_i)$ и $A_n \equiv 0$ и аналогичны вышеприведенным в подразделе 2.3.4, поэтому мы их опускаем.

З а м е ч а н и е 2.25. Алгоритмически найти указанное в (C_k) *разбиение с отмеченными точками* можно, например, методом последовательных покоординатно-медианных сечений (в случае, если $\mathcal{P}$ — параллелепипед) или с использованием диаграммы Вороного (см. подробности в разделе 1.3.3). Так, диаграмма Вороного (мозаика Вороного, разбиение Вороного) некоторого конечного множества точек в k-мерном пространстве представляет собой такое разбиение пространства $\mathbb{R}^k$ или его подмножества, содержащего указанный набор точек-узлов, при котором каждый элемент этого разбиения содержит одну узловую точку и все точки рассматриваемого подмножества, для которых данная узловая точка является ближайшей. $\square$

З а м е ч а н и е 2.26. Понятно, что оценка θ_n^* зависит от выбора того или иного разбиения множества $\mathcal{P}$. Например, если одномерные регрессоры лежат в отрезке $[c,d] \equiv \mathcal{P}$, то с использованием разбиения Вороного указанного отрезка, оценку параметра можно задать равенством $\theta_n^* = T^{-1}\left(\sum\limits_{i=1}^{n} \widetilde{\Delta} z_{ni} X_{ni}\right)$, где $\widetilde{\Delta} z_{n1} = \Delta z_{n1} + \Delta z_{n2}/2$, $\widetilde{\Delta} z_{nn} = \Delta z_{nn}/2 + \Delta z_{nn+1}$, $\widetilde{\Delta} z_{ni} = (\Delta z_{ni} + \Delta z_{ni+1})/2$ при $i = 2, \dots, n-1$, а величины X_{ni} и Δz_{ni} введены в подразделе 2.3.4 (см. формулу (289)). В некоторых случаях определенная здесь оценка θ^* может быть несколько точнее оценки, задаваемой формулой (291) (аналогичный эффект мы отмечали ранее в главе 1, см. замечание 1.15). В частности, соответствующие величины вида D_n^2 (см. (293)), отвечающие за асимптотический разброс оценок, могут быть несколько меньше. Но эти преимущества оказываются незначительными, особенно с точки зрения одношагового оценивания, поэтому задача выбора в том или ином смысле оптимального разбиения для наших целей не принципиальна. □

П р и м е р 2.18. Пусть $f(\theta, \mathbf{z}) = \theta z_1 + \theta^2 z_2$ при $\mathbf{z} = (z_1, z_2)$, $\theta > 0$, $z_1 \geq 0$, $z_2 \geq 0$. Пусть $\mathcal{P} = [c_1, d_1] \times [c_2, d_2]$. В этом случае

$$T(\theta) = \int\limits_{\mathcal{P}} f(\theta, \mathbf{z}) d\mathbf{z} = 2^{-1}\theta(d_1^2 - c_1^2)(d_2 - c_2) + 2^{-1}\theta^2(d_1 - c_1)(d_2^2 - c_2^2).$$

В частности, при $c_1 = c_2 = 0$ и $d_1 = d_2 = 1$ имеем $T(\theta) = 2^{-1}(\theta + \theta^2)$. □

Наконец, для оценивания многомерного параметра $\boldsymbol{\theta}$ в случае многомерных регрессоров нужно воспользоваться указанными в этом разделе построениями и схемой, изложенной в разделе 2.3.6 для скалярных регрессоров. Именно, рассмотрим конечное жорданово разбиение $\mathcal{P}^{(1)}, \dots, \mathcal{P}^{(m)}$ исходного множества $\mathcal{P}$. Для каждого подмножества $\mathcal{P}^{(l)}$ повторим вышеуказанные в этом разделе построения. В результате получим многомерное отображение $\mathbf{T}(\boldsymbol{\theta})$, определяемое соотношением

$$\mathbf{T}(\boldsymbol{\theta}) = (T_1(\boldsymbol{\theta}), \dots, T_m(\boldsymbol{\theta})), \qquad \text{где} \quad T_l(\boldsymbol{\theta}) = \int_{\mathcal{P}^{(l)}} f(\boldsymbol{\theta}, \mathbf{z}) d\mathbf{z}.$$

Если оно оказалось гомеоморфизмом, то оценку параметра $\boldsymbol{\theta}$ определим равенством

$$\boldsymbol{\theta}_n^* = \mathbf{T}^{-1}\Big(\sum_{i:\mathcal{P}_i^{(1)} \in \mathcal{P}^{(1)}} X_i \Lambda_k(\mathcal{P}_i^{(1)}), \dots, \sum_{i:\mathcal{P}_i^{(m)} \in \mathcal{P}^{(m)}} X_i \Lambda_k(\mathcal{P}_i^{(m)})\Big). \tag{312}$$

Здесь мы считаем, что нумерация элементов разбиения такова, что $\mathbf{z}_i \in \mathcal{P}_i^{(l)}$ при некотором l. Нетрудно видеть, что мы можем перенумеровать наблюдения $\{X_i\}$ таким образом, чтобы оценка (312) имела структуру (283). Следовательно, условия α_n-состоятельности или асимптотической нормальности оценки $\boldsymbol{\theta}_n^*$ легко извлечь из теорем 2.6 и 2.7.

П р и м е р 2.19. Рассмотрим в качестве примера известную в приложениях регрессионную функцию и покажем, что соответствующее отображение $\mathbf{T}(\boldsymbol{\theta})$ обладает требуемыми свойствами. Пусть $f(\boldsymbol{\theta}, \mathbf{z}) = z_1^{\theta_1} z_2^{\theta_2}$, где $\mathbf{z} = (z_1, z_2) \in \mathcal{P} = [0,1] \times [0,2]$, $\boldsymbol{\theta} = (\theta_1, \theta_2)$, $\theta_1, \theta_2 > 0$. Положим $\mathcal{P}^{(1)} = [0,1] \times [0,1]$, $\mathcal{P}^{(2)} = [0,1] \times [1,2]$. В этом случае

$$T_1(\boldsymbol{\theta}) = \int_{\mathcal{P}^{(1)}} f(\boldsymbol{\theta}, \mathbf{z}) d\mathbf{z} = (\theta_1 + 1)^{-1} (\theta_2 + 1)^{-1},$$
$$T_2(\boldsymbol{\theta}) = \int_{\mathcal{P}^{(2)}} f(\boldsymbol{\theta}, \mathbf{z}) d\mathbf{z} = (\theta_1 + 1)^{-1} (\theta_2 + 1)^{-1} (2^{\theta_2 + 1} - 1),$$

при этом якобиан двумерного преобразования $\mathbf{T}(\boldsymbol{\theta})$ всюду положителен, и система $T_1(\boldsymbol{\theta}) = t_1$, $T_2(\boldsymbol{\theta}) = t_2$ при любых $t_1, t_2 > 0$ однозначно разрешима в явном виде:

$$\boldsymbol{\theta} = \mathbf{T}^{-1}(t_1, t_2) = \left(t_1^{-1} \log_2^{-1}(1 + t_2/t_1),\ \log_2(1 + t_2/t_1) \right),$$

$\mathbf{T}(\boldsymbol{\theta})$ является диффеоморфизмом открытого положительного квадранта на себя. □

Рандомизация многомерных регрессоров проводится в точности так же, как и в разделе 2.3.7.

3 Асимптотический анализ одношаговых оценок

Пусть $X_1, \ldots, X_n$ — не обязательно независимые и одинаково распределенные наблюдения произвольной природы, распределения которых зависят от неизвестного параметра $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^m$, подлежащего оцениванию. Определим M-оценки как статистики $\widetilde{\boldsymbol{\theta}}_n$, которые на множестве асимптотически полной меры являются решениями уравнения

$$\sum_{i=1}^{n} \boldsymbol{\psi}_i(\mathbf{t}, X_i) = \mathbf{0}, \tag{313}$$

где $\{\boldsymbol{\psi}_i(\mathbf{t}, x)\}$ — известные вектор-функции с условием $\mathbb{E}\boldsymbol{\psi}_i(\boldsymbol{\theta}, X_i) = \mathbf{0}$ при всех i.

Глава посвящена асимптотическому анализу нескольких типов одношаговых оценок, имеющих асимптотически ту же точность, что и асимптотически нормальные M-оценки $\widetilde{\boldsymbol{\theta}}_n$.

Устройство главы следующее. В разделе 3.1 содержится асимптотический анализ двух типов одношаговых M-оценок. В частности, в подразделах 3.1.1 и 3.1.3 получены условия асимптотической нормальности оценок. В подразделе 3.1.5 исследуется вопрос о точности предварительной оценки, в подразделе 3.1.6 рассматривается поведение ближайшего к параметру корня уравнения (313). Комментарии о k-шаговом оценивании приведены в подразделе 3.1.8. Одношаговые приближения для взвешенных M-оценок исследуются в разделе 3.2. Раздел 3.3 посвящен уточнению одношаговых оценок Фишера в случае медленно сходящихся предварительных оценок. Результаты компьютерного моделирования, связанного с одношаговым приближением оценок квазиправдоподобия и метода наименьших квадратов в моделях нелинейной регрессии, приведены в разделе 3.4.

Результаты главы опубликованы в работах [361], [364], [365], [368] и [369].

3.1 Асимптотические свойства одношаговых M-оценок

3.1.1 Об условиях асимптотической нормальности

Нам потребуются следующие предположения.

$(\mathbf{A}_1)$ *Пусть нам даны наблюдения $X_1, \ldots, X_n$ со значениями в произвольном измеримом пространстве $\mathcal{X}$ и распределениями $\mathcal{L}_{1,\theta}, \ldots, \mathcal{L}_{n,\theta}$, зависящими от основного неизвестного параметра $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^m$, где $\boldsymbol{\Theta}$ — открытое множество. Кроме того, эти распределения могут зависеть от n и от некоторого второстепенного (мешающего) параметра $\sigma \in \Xi$ произвольной природы (зависимость от σ мы указывать не будем).*

$(\mathbf{A}_2)$ *При любом i на множестве $\boldsymbol{\Theta} \times \mathcal{X}$ заданы (зависящие, вообще говоря, от n) измеримые m-мерные векторы*[10] $\boldsymbol{\psi}_i(\mathbf{t}, x)$ *и $m \times m$-мерные матрицы $\mathrm{D}_i(\mathbf{t}, x)$ такие, что для*

[10]Условимся, что всюду в этой главе векторы — это вектор-столбцы.

каждого интервала $(\mathbf{t}_1, \mathbf{t}_2) \subset \Theta$ *при любом* $x \in \mathcal{X}$ *имеет место равенство*

$$\boldsymbol{\psi}_i(\mathbf{t}_2, x) - \boldsymbol{\psi}_i(\mathbf{t}_1, x) = \int\limits_0^1 \mathrm{D}_i(\mathbf{t}_1 + v(\mathbf{t}_2 - \mathbf{t}_1), x) dv \cdot (\mathbf{t}_2 - \mathbf{t}_1),$$

конечны математические ожидания каждой компоненты матрицы $\mathrm{D}_i(\theta, X_i)$ *и справедливы соотношения* $\mathbb{E}\boldsymbol{\psi}_i(\boldsymbol{\theta}, X_i) = \mathbf{0}$, $\mathbb{E}\|\boldsymbol{\psi}_i(\boldsymbol{\theta}, X_i)\|^2 < \infty$.

З а м е ч а н и е 3.1. Второстепенный параметр σ естественным образом возникает в моделях регрессии. Например, этим параметром может быть дисперсия наблюдений, не связанная с основным параметром. Математическое ожидание в (A_2) и далее вычисляется по отношению к распределению $\mathcal{L}_{i,\theta}$ (таким образом, оно может зависеть от σ). Условимся, что мы будем писать $\mathbb{E}$ и $\mathbb{P}$ вместо $\mathbb{E}_{\theta,\sigma}$ и $\mathbb{P}_{\theta,\sigma}$, соответственно. Заметим также, что для почти всех $\mathbf{t} \in \Theta$ матрица $\mathrm{D}_i(\mathbf{t}, X_i)$ с вероятностью 1 совпадает с матрицей Якоби вектора $\boldsymbol{\psi}_i(\mathbf{t}, X_i)$. □

Через $\mathrm{cond(A)}$ обозначим число обусловленности произвольной квадратной матрицы A: $\mathrm{cond(A)} \equiv \|\mathrm{A}^{-1}\|\|\mathrm{A}\|$. Положим

$$\mathrm{J}_{n,\theta} = \sum_{i=1}^n \mathbb{E}\mathrm{D}_i(\boldsymbol{\theta}, X_i), \qquad \mathrm{I}_{n,\theta} = \mathbb{E}\left(\sum_{i=1}^n \boldsymbol{\psi}_i(\boldsymbol{\theta}, X_i)\right)\left(\sum_{i=1}^n \boldsymbol{\psi}_i(\boldsymbol{\theta}, X_i)\right)^\top.$$

$(\mathbf{A}_3)$ *Для всех достаточно больших* n *матрица* $\mathrm{J}_{n,\theta}$ *невырождена, матрица* $\mathrm{I}_{n,\theta}$ *положительно определена. Кроме того,* $\|\mathrm{I}_{n,\theta}^{-1/2}\|\|\mathrm{J}_{n,\theta}\| \to \infty$, $\sup_n \mathrm{cond}(\mathrm{I}_{n,\theta}^{1/2}) < \infty$ *и*

$$\sum_{i=1}^n \mathrm{D}_i(\boldsymbol{\theta}, X_i)\mathrm{J}_{n,\theta}^{-1} \xrightarrow{p} \mathrm{I}, \qquad \mathrm{I}_{n,\theta}^{-1/2} \sum_{i=1}^n \boldsymbol{\psi}_i(\boldsymbol{\theta}, X_i) \xrightarrow{d} \mathcal{N}_m(\mathbf{0}, \mathrm{I}). \tag{314}$$

Положим $\mathcal{E}_{n,\theta}(\|\boldsymbol{\delta}\|) = \sum_{i=1}^n \mathbb{E}\omega_{i,\theta}(\|\boldsymbol{\delta}\|, X_i)$, где

$$\omega_{i,\theta}(\|\boldsymbol{\delta}\|, X_i) = \begin{cases} \sup\limits_{t:\, \|t-\theta\| \le \|\delta\|} \left\|\big(\mathrm{D}_i(\mathbf{t}, X_i) - \mathrm{D}_i(\boldsymbol{\theta}, X_i)\big)\mathrm{J}_{n,\theta}^{-1}\right\|, & \text{если } [\boldsymbol{\theta} - \boldsymbol{\delta}, \boldsymbol{\theta} + \boldsymbol{\delta}] \subset \boldsymbol{\Theta}, \\ \infty, \quad \text{иначе.} & \end{cases}$$

$(\mathbf{A}_4)$ *Имеет место соотношение* $\limsup \mathcal{E}_{n,\theta}(\|\boldsymbol{\delta}\|) \to 0$ *при* $\boldsymbol{\delta} \to \mathbf{0}$.

$(\mathbf{A}_5)$ *Имеется оценка* $\boldsymbol{\theta}_n^*$ *такая, что*

$$\|\mathrm{I}_{n,\theta}^{-1/2}\|\|\mathrm{J}_{n,\theta}\|\|\boldsymbol{\theta}_n^* - \boldsymbol{\theta}\|\mathcal{E}_{n,\theta}\big(\|\boldsymbol{\theta}_n^* - \boldsymbol{\theta}\|\big) \xrightarrow{p} 0.$$

Группу условий, состоящую из предположений $(A_1), \ldots, (A_5)$, обозначим символом (A). При выполнении условий (A) в качестве оценки параметра $\boldsymbol{\theta}$ рассмотрим статистику $\boldsymbol{\theta}_n^{**}$, определяемую соотношением

$$\boldsymbol{\theta}_n^{**} = \boldsymbol{\theta}_n^* - \left(\sum_{i=1}^n \mathrm{D}_i(\boldsymbol{\theta}_n^*, X_i)\right)^{-1} \sum_{i=1}^n \boldsymbol{\psi}_i(\boldsymbol{\theta}_n^*, X_i) \tag{315}$$

во всех случаях, когда определены все величины в правой части равенства в (315). Как будет установлено далее, при выполнении условий (A) справедливы соотношения

$$\mathbb{P}(\boldsymbol{\theta}_n^* \in \boldsymbol{\Theta}) \to 1 \qquad \text{и} \qquad \mathbb{P}\left(\det\Big(\sum_{i=1}^n \mathrm{D}_i(\boldsymbol{\theta}_n^*, X_i)\Big) \neq 0\right) \to 1.$$

Таким образом, при выполнении условий (A) статистика $\boldsymbol{\theta}_n^{**}$ определена с вероятностью, стремящейся к 1. Справедлива

Теорема 3.1. *Пусть выполнены условия (A). Тогда оценка $\boldsymbol{\theta}_n^{**}$ определена с вероятностью, стремящейся к* 1 *и*

$$\boldsymbol{\delta}_{n,\theta} = \mathrm{I}_{n,\theta}^{-1/2} \mathrm{J}_{n,\theta}\big(\boldsymbol{\theta}_n^{**} - \boldsymbol{\theta}\big) + \mathrm{I}_{n,\theta}^{-1/2} \sum_{i=1}^n \boldsymbol{\psi}_i(\boldsymbol{\theta}, X_i) \xrightarrow{p} \mathbf{0}. \tag{316}$$

В частности,

$$\mathrm{I}_{n,\theta}^{-1/2} \mathrm{J}_{n,\theta}\big(\boldsymbol{\theta}_n^{**} - \boldsymbol{\theta}\big) \xrightarrow{d} \mathcal{N}_m(\mathbf{0}, \mathrm{I}). \tag{317}$$

З а м е ч а н и е 3.2. Соотношения (314) в условии (A_3) — это соответственно варианты закона больших чисел и центральной предельной теоремы в схеме серий. Достаточные условия для выполнения таких предельных соотношений для тех или иных типов зависимостей хорошо известны. В частности, если наблюдения независимы и $m = 1$, то для выполнения указанных соотношений достаточно потребовать (см., например, [325]), чтобы

$$\sum_{i=1}^n \mathbb{E}\min\left\{\big|\psi_i'(\theta, X_i) - \mathbb{E}\psi_i'(\theta, X_i)\big|/|J_{n,\theta}|,\ \big|\psi_i'(\theta, X_i) - \mathbb{E}\psi_i'(\theta, X_i)\big|^2/J_{n,\theta}^2\right\} \to 0,$$

$$\sum_{i=1}^n \mathbb{E}\min\left\{\big|\psi_i^2(\theta, X_i)\big|/I_{n,\theta},\ \big|\psi_i(\theta, X_i)\big|^s/I_{n,\theta}^{s/2}\right\} \to 0$$

при некотором $s > 2$. □

З а м е ч а н и е 3.3. Центральное условие группы предположений (A) — это условие (A_5). Это условие является некоторым универсальным ограничением, связывающем гладкость функций $\boldsymbol{\psi}_i(\cdot, X_i)$, определяющих одношаговую M-оценку $\boldsymbol{\theta}_n^{**}$, и скорость сближения предварительной оценки $\boldsymbol{\theta}_n^*$ и параметра $\boldsymbol{\theta}$, которые нужны для асимптотической нормальности одношаговых M-оценок. При этом точность предварительной оценки $\boldsymbol{\theta}_n^*$ и гладкость функций $\boldsymbol{\psi}_i(\cdot, X_i)$ в известном смысле обратно пропорциональны друг другу: чем меньше точность, тем больше должна быть гладкость. Действительно, для выполнения (A_5) достаточно, чтобы величина $\mathcal{E}_{n,\theta}(|\boldsymbol{\delta}|)$ и оценка $\boldsymbol{\theta}_n^*$ одновременно (для некоторого α из указанной области) удовлетворяли одному из следующих двух условий:

$$\begin{gathered}\limsup \mathcal{E}_{n,\theta}(\|\boldsymbol{\delta}\|) = o(\|\boldsymbol{\delta}\|^{\alpha}) \quad \text{при} \qquad \boldsymbol{\delta} \to \mathbf{0} \quad \text{и} \\ \left(\|\mathrm{I}_{n,\theta}^{-1/2}\|\|\mathrm{J}_{n,\theta}\|\right)^{1/(1+\alpha)} \|\boldsymbol{\theta}_n^* - \boldsymbol{\theta}\| = O_p(1), \qquad 0 \le \alpha < 1;\end{gathered} \tag{318}$$

или

$$\begin{gathered}\limsup \mathcal{E}_{n,\theta}(\|\boldsymbol{\delta}\|) = O(\|\boldsymbol{\delta}\|^{\alpha}) \quad \text{при} \qquad \boldsymbol{\delta} \to \mathbf{0} \quad \text{и} \\ \left(\|\mathrm{I}_{n,\theta}^{-1/2}\|\|\mathrm{J}_{n,\theta}\|\right)^{1/(1+\alpha)} \|\boldsymbol{\theta}_n^* - \boldsymbol{\theta}\| \xrightarrow{p} 0, \qquad 0 < \alpha \le 1.\end{gathered} \tag{319}$$

В частности, в двух крайних случаях $\alpha = 0$ и $\alpha = 1$ получаем, что для справедливости (A_5) достаточно, чтобы

$$\limsup \mathcal{E}_{n,\theta}(\|\boldsymbol{\delta}\|) = o(1) \quad \text{при} \quad \boldsymbol{\delta} \to \mathbf{0} \quad \text{и} \quad \|\mathrm{I}_{n,\theta}^{-1/2}\|\|\mathrm{J}_{n,\theta}\|\|\boldsymbol{\theta}_n^* - \boldsymbol{\theta}\| = O_p(1)$$

или

$$\limsup \mathcal{E}_{n,\theta}(\|\boldsymbol{\delta}\|) = O(\|\boldsymbol{\delta}\|) \quad \text{при} \quad \boldsymbol{\delta} \to \mathbf{0} \quad \text{и} \quad \|\mathrm{I}_{n,\theta}^{-1/2}\|\|\mathrm{J}_{n,\theta}\|\|\boldsymbol{\theta}_n^* - \boldsymbol{\theta}\|^2 \xrightarrow{p} 0.$$

Таким образом, условие (A_5) включает в себя широкий спектр ограничений на точность предварительной оценки θ_n^*. Отметим еще, что если $\|\mathrm{I}_{n,\theta}^{-1/2}\|\|\mathrm{J}_{n,\theta}\|$ имеет порядок $\sqrt{n}$ (в частности, в случае независимых одинаково распределенных наблюдений), то условия на точность оценки $\boldsymbol{\theta}_n^*$ в указанных двух крайних случаях — это, соответственно, либо предположение о $\sqrt{n}$-ограниченности, либо об $n^{1/4}$-состоятельности этой оценки. □

З а м е ч а н и е 3.4. Помимо различных «регулярных» вариантов поведения двух рассматриваемых в замечании 3.3 характеристик, допускается и, в известной степени, «вырожденный» случай, когда с вероятностью 1 элементы матрицы $D_i(\mathbf{t}, X_i)$ постоянны в некоторой окрестности $\{\mathbf{t} : \|\mathbf{t} - \boldsymbol{\theta}\| \le \delta_o(\boldsymbol{\theta})\}$ параметра $\boldsymbol{\theta}$, т.е. $\omega_{i,\theta}(\|\boldsymbol{\delta}\|, X_i) = \mathcal{E}_{n,\theta}(\|\boldsymbol{\delta}\|) \equiv 0$ при $\|\boldsymbol{\delta}\| \le \delta_o(\boldsymbol{\theta})$. В этом случае статистика $\boldsymbol{\theta}_n^*$ может быть произвольной и даже не сходиться к $\boldsymbol{\theta}$, лишь бы $\boldsymbol{\theta}_n^*$ попадала в указанную окрестность $\boldsymbol{\theta}$ с вероятностью, стремящейся к 1. Отметим, что подобная ситуация реализуется для оценки максимального правдоподобия одномерного параметра сноса $\boldsymbol{\theta}$ для однородной выборки из нормального распределения. □

Рассмотрим также некоторую модификацию оценки $\boldsymbol{\theta}_n^{**}$ из (315). Положим

$$\widehat{\boldsymbol{\theta}}_n^{**} = \boldsymbol{\theta}_n^* - \mathrm{J}_n^{*-1} \sum_{i=1}^{n} \boldsymbol{\psi}_i(\boldsymbol{\theta}_n^*, X_i), \tag{320}$$

где матрица J_n^* является статистикой, в некотором смысле приближающей $\mathrm{J}_{n,\theta}$. Подобная оценка представляется весьма естественной, если сравнить определения (315) и (320) и учесть первое условие в (314). Выбор оценки J_n^* может быть достаточно очевиден, если $\mathrm{J}_{n,\theta} = \mathrm{J}_{n,\theta,\sigma} \equiv \mathrm{J}_n(\boldsymbol{\theta})$. Здесь проще всего положить $\mathrm{J}_n^* = \mathrm{J}_n(\boldsymbol{\theta}_n^*)$. Подобная ситуация реализуется, например, в случае построения одношаговых приближений для оценки максимального правдоподобия (в этом случае $-\mathrm{J}_{n,\theta} = -\mathrm{J}_n(\boldsymbol{\theta})$ есть сумма информационных матриц Фишера, построенных по элементам разнораспределенной выборки). Другие примеры приведены в разделе 3.4.1. Кроме того, возможны ситуации, когда можно построить множество различных приближений J_n^* для $\mathrm{J}_{n,\theta}$ (см. пример 3.4 в разделе 3.4.1). Поскольку выбор той или иной оценки J_n^* может быть осуществлен только в рамках конкретной статистической модели, то условия асимптотической нормальности одношаговой оценки $\widehat{\boldsymbol{\theta}}_n^{**}$ приведем в предположении, что существует статистика J_n^* такая, что $\mathrm{J}_n^*\mathrm{J}_{n,\theta}^{-1} \xrightarrow{p} \mathrm{I}$. Положим $\overline{\mathrm{D}_n}(\boldsymbol{\theta}) = \sum\limits_{i=1}^{n} \mathrm{D}_i(\boldsymbol{\theta}, X_i)\mathrm{J}_{n,\theta}^{-1} - \mathrm{I}$.

Теорема 3.2. *Пусть выполнены условия* (A) *и*

$$\|\mathrm{I}_{n,\theta}^{-1/2}\|\|\mathrm{J}_{n,\theta}\|\|\boldsymbol{\theta}_n^* - \boldsymbol{\theta}\|\Big(\|\mathrm{J}_n^*\mathrm{J}_{n,\theta}^{-1} - \mathrm{I}\| + \big\|\overline{\mathrm{D}_n}(\boldsymbol{\theta})\big\|\Big) \xrightarrow{p} 0, \qquad \mathrm{J}_n^*\mathrm{J}_{n,\theta}^{-1} \xrightarrow{p} \mathrm{I}, \tag{321}$$

Тогда оценка $\widehat{\boldsymbol{\theta}}_n^{**}$ *определена с вероятностью, стремящейся к* 1 *и*

$$\mathrm{I}_{n,\theta}^{-1/2}\mathrm{J}_{n,\theta}\big(\widehat{\boldsymbol{\theta}}_n^{**} - \boldsymbol{\theta}\big) \xrightarrow{d} \mathcal{N}_m(\mathbf{0}, \mathrm{I}). \tag{322}$$

Таким образом, при наличии некоторых приближений J_n^* для $\mathrm{J}_{n,\theta}$, мы располагаем одношаговыми оценками $\boldsymbol{\theta}_n^{**}$ и $\widehat{\boldsymbol{\theta}}_n^{**}$, эквивалентными в смысле асимптотической точности. Понятно, что среди этих оценок можно выбирать в том или ином смысле более предпочтительную оценку для параметра $\boldsymbol{\theta}$. Например, та или иная оценка может быть проще (см. раздел 3.4.1).

З а м е ч а н и е 3.5. Пусть $J_{n,\theta} = J_n(\boldsymbol{\theta})$. Положим $J_n^* = J_n(\boldsymbol{\theta}_n^*)$ и предположим, что выполнены условия (A_1)–(A_4). В этом случае для выполнения (A_5) и условия (321) достаточно, чтобы величины $\mathcal{E}_{n,\theta}(\|\boldsymbol{\delta}\|)$, $J_n(\mathbf{t})$ и оценка $\boldsymbol{\theta}_n^*$ одновременно удовлетворяли одному из следующих двух блоков условий:

(i) выполнено условие (318) и $\limsup \big\|\mathrm{J}_n(\mathbf{t})\mathrm{J}_n^{-1}(\boldsymbol{\theta}) - \mathrm{I}\big\| = o(\|\mathbf{t} - \boldsymbol{\theta}\|^\alpha)$ при $\mathbf{t} \to \boldsymbol{\theta}$,

$$\Big(\|\mathrm{I}_{n,\theta}^{-1/2}\|\|\mathrm{J}_n(\boldsymbol{\theta})\|\Big)^{\alpha/(1+\alpha)}\big\|\overline{\mathrm{D}_n}(\boldsymbol{\theta})\big\| \xrightarrow{p} 0;$$

(ii) выполнено (319) и $\limsup \big\|\mathrm{J}_n(t)\mathrm{J}_n^{-1}(\boldsymbol{\theta}) - \mathrm{I}\big\| = O(\|\mathbf{t} - \boldsymbol{\theta}\|^\alpha)$ при $\mathbf{t} \to \boldsymbol{\theta}$,

$$\Big(\|\mathrm{I}_{n,\theta}^{-1/2}\|\|\mathrm{J}_n(\boldsymbol{\theta})\|\Big)^{\alpha/(1+\alpha)}\big\|\overline{\mathrm{D}_n}(\boldsymbol{\theta})\big\| = O_p(1).$$

$\square$

З а м е ч а н и е 3.6. Интересно сравнить одношаговое оценивание и метод максимального правдоподобия. Оказывается, известны примеры, когда оценка максимального правдоподобия, удовлетворяющая M-уравнению, может не быть состоятельной, в то время как имеется другой состоятельный корень этого уравнения. Более того, возможна ситуация, когда оценка максимального правдоподобия не существует, а соответствующая одношаговая оценка является состоятельной с нужной скоростью сходимости (см., например, [162], [169]). Проиллюстрируем указанное преимущество одношагового оценивания одним простым примером (см., например, [168]). Пусть $\{X_i\}$ — независимые и одинаково распределенные случайные величины, имеющие плотность распределения

$$f_\theta(x) = (2\sigma)^{-1}\varphi\big((x-\alpha)/\sigma\big) + 2^{-1}\varphi(x),$$

где $\varphi(x) = e^{-x^2/2}/\sqrt{2\pi}$ и $\boldsymbol{\theta} = (\alpha, \sigma^2)^\top$, $\boldsymbol{\Theta} = \{\alpha \in \mathbb{R}, \sigma > 0\}$. С одной стороны, оценка максимального правдоподобия параметра $\boldsymbol{\theta}$ здесь не существует, поскольку функция правдоподобия неограничена в окрестностях n точек $(X_i, 0)$, лежащих на границе множества $\boldsymbol{\Theta}$. С другой стороны, соответствующие одношаговые оценки $\boldsymbol{\theta}_n^{**}$ или $\widehat{\boldsymbol{\theta}}_n^{**}$, где $\boldsymbol{\theta}_n^*$ оценка метода моментов, являются асимптотически эффективными в смысле сходимости (317) или (322). $\square$

3.1.2 Доказательство утверждений раздела 3.1.1

В доказательствах для краткости всюду положим

$$\mathrm{I}_n = \mathrm{I}_{n,\theta}, \quad \mathrm{J}_n = \mathrm{J}_{n,\theta}, \quad \boldsymbol{\delta}_n = \boldsymbol{\delta}_{n,\theta}, \quad \mathcal{E}_n(\cdot) = \mathcal{E}_{n,\theta}(\cdot), \quad \omega_i(\cdot, X_i) = \omega_{i,\theta}(\cdot, X_i). \tag{323}$$

Введем обозначения

$$\mathbf{u}_n^* = \boldsymbol{\theta}_n^* - \boldsymbol{\theta}, \qquad \mathrm{R}_n(\mathbf{u}) = \sum_{i=1}^n \big(\mathrm{D}_i(\boldsymbol{\theta}+\mathbf{u}, X_i) - \mathrm{D}_i(\boldsymbol{\theta}, X_i)\big)\mathrm{J}_n^{-1},$$
$$\boldsymbol{\rho}_n(\mathbf{u}) = \mathrm{I}_n^{-1/2}\sum_{i=1}^n \big(\mathrm{D}_i(\boldsymbol{\theta}+\mathbf{u}, X_i) - \mathrm{D}_i(\boldsymbol{\theta}, X_i)\big)\mathbf{u}, \tag{324}$$
$$\overline{\boldsymbol{\rho}}_n(\mathbf{u}) = \mathrm{I}_n^{-1/2}\sum_{i=1}^n \big(\boldsymbol{\psi}_i(\boldsymbol{\theta}+\mathbf{u}, X_i) - \boldsymbol{\psi}_i(\boldsymbol{\theta}, X_i) - \mathrm{D}_i(\boldsymbol{\theta}+\mathbf{u}, X_i)\mathbf{u}\big).$$

Нам потребуется два вспомогательных утверждения.

Лемма 3.1. *Пусть $\alpha_n(\cdot) : [0,\infty) \to [0,\infty)$, $n = 1, 2, \ldots,$ — последовательность случайных процессов с монотонно неубывающими траекториями и конечными средними $b_n(t) = \mathbb{E}\alpha_n(t)$. Пусть $b_n(\eta_n) \xrightarrow{p} 0$ для некоторой последовательности случайных величин $\eta_n \geq 0$. Тогда $\alpha_n(\eta_n) \xrightarrow{p} 0$.*

Д о к а з а т е л ь с т в о этого утверждения приведено, например, в [338].

Лемма 3.2. *Если выполнены условия* (A), *то*

$$\mathrm{R}_n(\mathbf{u}_n^*) \xrightarrow{p} 0, \qquad \boldsymbol{\rho}_n(\mathbf{u}_n^*) \xrightarrow{p} \mathbf{0}, \qquad \overline{\boldsymbol{\rho}}_n(\mathbf{u}_n^*) \xrightarrow{p} \mathbf{0} \tag{325}$$

и оценка $\boldsymbol{\theta}_n^{**}$ *определена с вероятностью, стремящейся к* 1. *Если дополнительно справедливы соотношения* (321), *то оценка* $\widehat{\boldsymbol{\theta}}_n^{**}$ *определена с вероятностью, стремящейся к* 1.

Д о к а з а т е л ь с т в о. В условиях леммы имеют место следующие предельные соотношения:

$$\mathcal{E}_n(\|\mathbf{u}_n^*\|) \xrightarrow{p} 0 \qquad \text{и} \qquad \|\mathrm{I}_n^{-1/2}\|\|\mathrm{J}_n\|\|\mathbf{u}_n^*\|\mathcal{E}_n(\|\mathbf{u}_n^*\|) \xrightarrow{p} 0. \tag{326}$$

Действительно, вторая сходимость в (326) совпадает с условием (A_5). Кроме того, с учетом этого соотношения для любого $\varepsilon > 0$

$$\begin{aligned}\mathbb{P}\Big(\mathcal{E}_n(\|\mathbf{u}_n^*\|) > \varepsilon\Big) = \mathbb{P}\Big(\mathcal{E}_n(\|\mathbf{u}_n^*\|) > \varepsilon,\ \|\mathrm{I}_n^{-1/2}\|\|\mathrm{J}_n\|\|\mathbf{u}_n^*\| \geq 1\Big) + \\ + \mathbb{P}\Big(\mathcal{E}_n(\|\mathbf{u}_n^*\|) > \varepsilon,\ \|\mathrm{I}_n^{-1/2}\|\|\mathrm{J}_n\|\|\mathbf{u}_n^*\| < 1\Big) \leq \mathbb{P}\Big(\|\mathrm{I}_n^{-1/2}\|\|\mathrm{J}_n\|\|\mathbf{u}_n^*\|\mathcal{E}_n(\|\mathbf{u}_n^*\|) > \varepsilon\Big) + \\ + \mathbb{P}\Big(\mathcal{E}_n(\|\mathbf{u}_n^*\|) > \varepsilon,\ \|\mathbf{u}_n^*\| < (\|\mathrm{I}_n^{-1/2}\|\|\mathrm{J}_n\|)^{-1}\Big) \to 0.\end{aligned} \tag{327}$$

При выводе сходимости в (327) мы использовали то, что начиная с некоторых n в правой части (327) под знаком вероятности стоит невозможное событие. Этот факт следует из монотонности функций $\mathcal{E}_n(\cdot)$ и условий (A_3) и (A_4).

Далее, ввиду (A_2),

$$\boldsymbol{\psi}_i(\boldsymbol{\theta} + \mathbf{u}, X_i) - \boldsymbol{\psi}_i(\boldsymbol{\theta}, X_i) - \mathrm{D}_i(\boldsymbol{\theta}, X_i)\mathbf{u} = \int\limits_0^1 \big(\mathrm{D}_i(\boldsymbol{\theta} + v\mathbf{u}) - \mathrm{D}_i(\boldsymbol{\theta}, X_i)\big) dv\mathbf{u}. \tag{328}$$

Следовательно, с учетом (324) и (328) получаем, что

$$\begin{aligned}\|\mathrm{R}_n(\mathbf{u})\| &\leq \sum_{i=1}^n \omega_i(\|\mathbf{u}\|, X_i), \\ \|\boldsymbol{\rho}_n(\mathbf{u})\| &\leq \|\mathrm{I}_n^{-1/2}\|\mathrm{J}_n\|\|\mathbf{u}\| \sum_{i=1}^n \omega_i(\|\mathbf{u}\|, X_i), \\ \|\overline{\boldsymbol{\rho}}_n(\mathbf{u})\| &\leq \|\mathrm{I}_n^{-1/2}\|\mathrm{J}_n\|\|\mathbf{u}\| \sum_{i=1}^n \omega_i(\|\mathbf{u}\|, X_i).\end{aligned} \tag{329}$$

Функции в правой части соотношений (329), рассматриваемые как функции $\alpha_n(\|\mathbf{u}\|)$, удовлетворяют условиям леммы 3.1. С учетом равенства $\mathcal{E}_n(\|\mathbf{u}\|) = \sum\limits_{i=1}^n \mathbb{E}\omega_i(\|\mathbf{u}\|, X_i)$ все предельные соотношения в (325) следуют теперь из леммы 3.1 при $\eta_n = \|\mathbf{u}_n^*\|$.

Докажем два последних утверждения леммы. Поскольку с вероятностью, стремящейся к 1, величина $\mathcal{E}_n(\|\boldsymbol{\theta}_n^* - \boldsymbol{\theta}\|)$ конечна в силу (326), то $\mathbb{P}(\boldsymbol{\theta}_n^* \in \boldsymbol{\Theta}) \to 1$. Следовательно, согласно условию (A_2), с вероятностью, стремящейся к 1, определены величины $\sum_{i=1}^n \boldsymbol{\psi}_i(\boldsymbol{\theta}_n^*, X_i)$ и $\sum_{i=1}^n \mathrm{D}_i(\boldsymbol{\theta}_n^*, X_i)$. А поскольку еще

$$\sum_{i=1}^n \mathrm{D}_i(\boldsymbol{\theta}_n^*, X_i)\mathrm{J}_n^{-1} = \sum_{i=1}^n \mathrm{D}_i(\boldsymbol{\theta}, X_i)\mathrm{J}_n^{-1} + \mathrm{R}_n(\mathbf{u}_n^*),$$

то из условия (A_3) и первого соотношения в (325) заключаем, что

$$\mathbb{P}\Big(\det\Big(\sum_{i=1}^n \mathrm{D}_i(\boldsymbol{\theta}_n^*, X_i)\Big) \neq 0\Big) \to 1.$$

Таким образом, с учетом определения (315), статистика $\boldsymbol{\theta}_n^{**}$ определена с вероятностью, стремящейся к 1. Вместе с определением (320) и условием (321) получаем, что статистика $\widehat{\boldsymbol{\theta}}_n^{**}$ также определена с вероятностью, стремящейся к 1. □

Д о к а з а т е л ь с т в о теоремы 3.1. Из условия (A_3) и определения (315) следует, что

$$\begin{gathered}\overline{\mathrm{R}}_n = \sum_{i=1}^n \mathrm{D}_i(\boldsymbol{\theta}, X_i)\mathrm{J}_n^{-1} - \mathrm{I} \xrightarrow{p} 0, \quad \boldsymbol{\rho}_{n0} = \mathrm{I}_n^{-1/2}\sum_{i=1}^n \boldsymbol{\psi}_i(\boldsymbol{\theta}, X_i) \xrightarrow{d} \mathcal{N}_m(\mathbf{0}, \mathrm{I}), \\ \sum_{i=1}^n \mathrm{D}_i(\boldsymbol{\theta}_n^*, X_i)(\boldsymbol{\theta}_n^{**} - \boldsymbol{\theta}) = \sum_{i=1}^n \mathrm{D}_i(\boldsymbol{\theta}_n^*, X_i)(\boldsymbol{\theta}_n^* - \boldsymbol{\theta}) - \sum_{i=1}^n \boldsymbol{\psi}_i(\boldsymbol{\theta}_n^*, X_i).\end{gathered} \tag{330}$$

Далее, учитывая определения (316) и (324), получаем

$$\boldsymbol{\delta}_n \equiv \mathrm{I}_n^{-1/2}\mathrm{J}_n(\boldsymbol{\theta}_n^{**} - \boldsymbol{\theta}) + \boldsymbol{\rho}_{n0} = \big(\mathrm{H}_n + \mathrm{I}\big)^{-1}\big(\mathrm{H}_n\boldsymbol{\rho}_{n0} + \boldsymbol{\rho}_n(\mathbf{u}_n^*) - \overline{\boldsymbol{\rho}}_n(\mathbf{u}_n^*)\big), \tag{331}$$

где $\mathrm{H}_n = \mathrm{I}_n^{-1/2}\big(\mathrm{R}_n(\mathbf{u}_n^*) + \overline{\mathrm{R}}_n\big)\mathrm{I}_n^{1/2}$. Из (330), леммы 3.2 и условия (A_3) заключаем, что

$$\|\mathrm{H}_n\| \le \big(\|\mathrm{R}_n(\mathbf{u}_n^*)\| + \|\overline{\mathrm{R}}_n\|\big)\sup_n \|\mathrm{I}_n^{-1/2}\|\|\mathrm{I}_n^{1/2}\| \xrightarrow{p} 0.$$

Таким образом, $\mathrm{H}_n \xrightarrow{p} 0$. Утверждения (316) и (317) теоремы следуют теперь из соотношения (331), леммы 3.2 и второго предельного соотношения в (330). Первое утверждение теоремы доказано в лемме 3.2. □

Д о к а з а т е л ь с т в о теоремы 3.2. Тот факт, что оценка $\widehat{\boldsymbol{\theta}}_n^{**}$ определена с вероятностью, стремящейся к 1, установлен в лемме 3.2. Далее, с учетом (321), имеем

$$\begin{aligned}\boldsymbol{\rho}_{n1} &= \mathrm{I}_n^{-1/2}(\mathrm{J}_n^* - \mathrm{J}_n)\mathbf{u}_n^* \xrightarrow{p} \mathbf{0}, \\ \boldsymbol{\rho}_{n2} &= \mathrm{I}_n^{-1/2}\Big(\sum_{i=1}^n \mathrm{D}_i(\boldsymbol{\theta}, X_i) - \mathrm{J}_n\Big)\mathbf{u}_n^* \xrightarrow{p} \mathbf{0}, \\ \mathrm{R}_{n1} &= \mathrm{I}_n^{-1/2}\big(\mathrm{J}_n^*\mathrm{J}_n^{-1} - \mathrm{I}\big)\mathrm{I}_n^{1/2} \xrightarrow{p} 0.\end{aligned} \tag{332}$$

Первые два предельных соотношения очевидны. Третья сходимость в (332) является следствием следующих соотношений:

$$\|\mathrm{R}_{n1}\| \le \|\mathrm{I}_n^{-1/2}\| \|\big(\mathrm{J}_n^* \mathrm{J}_n^{-1} - \mathrm{I}\big)\| \|\mathrm{I}_n^{1/2}\| \le \|\big(\mathrm{J}_n^* \mathrm{J}_n^{-1} - \mathrm{I}\big)\| \sup_n \|\mathrm{I}_n^{-1/2}\| \|\mathrm{I}_n^{1/2}\| \xrightarrow{p} 0.$$

Из определения (320) и обозначений из (324) и (332) получаем тождество

$$\mathrm{I}_n^{-1/2} \mathrm{J}_n (\widehat{\boldsymbol{\theta}}_n^{**} - \boldsymbol{\theta}) = \Big(\mathrm{R}_{n1} + \mathrm{I}\Big)^{-1} \Big(\boldsymbol{\rho}_{n1} - \boldsymbol{\rho}_{n2} - \overline{\boldsymbol{\rho}}_n(\mathbf{u}_n^*) - \boldsymbol{\rho}_{n0}\Big).$$

Сходимость (322) следует теперь из леммы 3.2 и соотношений (330) и (332). □

Д о к а з а т е л ь с т в о утверждений из замечания 3.3. Пусть выполнены ограничения из (318). В этом случае для всех достаточно больших n

$$\begin{aligned}\|\mathrm{I}_n^{-1/2}\| \|\mathrm{J}_n\| \|\mathbf{u}_n^*\| \mathcal{E}_n(\|\mathbf{u}_n^*\|) \le \|\mathrm{I}_n^{-1/2}\| \|\mathrm{J}_n\| \|\mathbf{u}_n^*\|^{1+\alpha} o_p(1) = \\ = \Big(\big(\|\mathrm{I}_n^{-1/2}\| \|\mathrm{J}_n\|\big)^{1/(1+\alpha)} \|\mathbf{u}_n^*\|\Big)^{1+\alpha} o_p(1) = \big(O_p(1)\big)^{1+\alpha} o_p(1) \xrightarrow{p} 0. \quad (333)\end{aligned}$$

Если же выполнены условия из (319), то нужно в этих рассуждениях поменять местами символы $o_p(1)$ и $O_p(1)$. □

Д о к а з а т е л ь с т в о утверждений из замечания 3.5. Пусть выполнены ограничения из (i). В этом случае имеет место (333), т.е. справедливо условие (A_5). Кроме того, поскольку $\|\mathrm{I}_n^{-1/2}\| \|\mathrm{J}_n\| \to \infty$, то в силу второго условия из (318) выполнено $\mathbf{u}_n^* \xrightarrow{p} \mathbf{0}$. Следовательно, для всех достаточно больших n

$$\|\mathrm{J}_n(\boldsymbol{\theta}^*) \mathrm{J}_n - \mathrm{I}\| = \|\mathbf{u}_n^*\|^{\alpha} o_p(1) \xrightarrow{p} 0,$$

$$\begin{aligned}\|\mathrm{I}_n^{-1/2}\| \|\mathrm{J}_n\| \|\mathbf{u}_n^*\| \big\|\mathrm{J}_n(\boldsymbol{\theta}^*) \mathrm{J}_n^{-1} - \mathrm{I}\big\| = \|\mathrm{I}_n^{-1/2}\| \|\mathrm{J}_n\| \|\mathbf{u}_n^*\|^{1+\alpha} o_p(1) = \\ = \Big(\big(\|\mathrm{I}_n^{-1/2}\| \|\mathrm{J}_n\|\big)^{1/(1+\alpha)} \|\mathbf{u}_n^*\|\Big)^{1+\alpha} o_p(1) = \big(O_p(1)\big)^{1+\alpha} o_p(1),\end{aligned}$$

$$\begin{aligned}\big(\|\mathrm{I}_n^{-1/2}\| \|\mathrm{J}_n\|\big) \|\mathbf{u}_n^*\| \big\|\overline{\mathrm{D}_n}(\boldsymbol{\theta})\big\| = \big(\|\mathrm{I}_n^{-1/2}\| \|\mathrm{J}_n\|\big)^{1/(1+\alpha)} \|\mathbf{u}_n^*\| \times \\ \times \big\|\overline{\mathrm{D}_n}(\boldsymbol{\theta})\big\| \big(\|\mathrm{I}_n^{-1/2}\| \|\mathrm{J}_n\|\big)^{\alpha/(1+\alpha)} = O_p(1) \cdot o_p(1) \xrightarrow{p} 0,\end{aligned}$$

т.е. выполнены все условия в (321).

Чтобы извлечь (321) в случае, когда выполнены условия (ii), нужно повторить вышеприведенные рассуждения, поменяв местами символы $o_p(1)$ и $O_p(1)$. □

3.1.3 Некоторые дополнения

Далее для упрощения изложения считаем, что основной параметр $\boldsymbol{\theta}$ одномерный ($m = 1$). В этом разделе мы приведем два утверждения. Первое касается оценивания коэффициента $Q_{n,\theta} = I_{n,\theta}^{-1/2} J_{n,\theta}$, нормирующего разность $\theta_n^{**} - \theta$ в теореме 3.1 об асимптотической нормальности оценки θ_n^{**}. Второе содержит равномерный по параметрам аналог теоремы 3.1.

Положим
$$Q_n^* = \left(\sum_{i=1}^n \psi_i^2(\theta_n^{**}, X_i)\right)^{-1/2} \sum_{i=1}^n D_i(\theta_n^*, X_i) \tag{334}$$
во всех случаях, когда определена правая часть этого равенства. Следующее утверждение может быть полезным при построении доверительных интервалов и проверке гипотез, поскольку множитель Q_n^*, нормирующий разность $\theta_n^{**} - \theta$, является статистикой (заданной на множестве выборок асимптотически полной меры).

Теорема 3.3. *Если выполнены условия* (A) *и*
$$I_{n,\theta}^{-1} \sum_{i=1}^n \psi_i^2(\theta, X_i) \xrightarrow{p} 1, \qquad J_{n,\theta}^{-2} \sum_{i=1}^n D_i^2(\theta, X_i) \xrightarrow{p} 0, \tag{335}$$
то величина Q_n^* *определена с вероятностью, стремящейся к* 1, *и*
$$Q_n^* \big(\theta_n^{**} - \theta\big) \xrightarrow{d} \mathcal{N}_1(0, 1).$$

З а м е ч а н и е 3.7. Первое условие в (335) является вариантом закона больших чисел и достаточные условия для его выполнения хорошо известны (см., например, [325], а также замечание 3.2). Заметим еще, что в случае разнораспределенных наблюдений соответствующего условия из (A_3) не достаточно, чтобы было выполнено второе соотношение в (335) (требуемый пример нетрудно построить).

Имеет место следующий *равномерный по параметрам* аналог последнего утверждения теоремы 3.1.

Теорема 3.4. *Пусть в условиях* (A) *предельные соотношения из* (314), (A_4) *и* (A_5) *заменены следующими:*
$$\sup_{\theta\in\Theta,\sigma\in\Xi,x\in\mathbb{R}} \left| \mathbb{P}\left(I_{n,\theta}^{-1/2} \sum_{i=1}^n \psi_i(\theta, X_i) < x \right) - \Phi(x) \right| \to 0$$
и для любого $\varepsilon > 0$ $$\sup_{\theta\in\Theta,\sigma\in\Xi} \mathbb{P}\left(\left| \sum_{i=1}^n D_i(\theta, X_i) J_{n,\theta}^{-1} - 1 \right| > \varepsilon \right) \to 0,$$
$$\sup_{\theta\in\Theta,\sigma\in\Xi} \Big(\mathbb{P}\left(I_{n,\theta}^{-1/2} |J_{n,\theta}| |\theta_n^* - \theta| \mathcal{E}_{n,\theta}(|\theta_n^* - \theta|) > \varepsilon \right) + \mathbb{P}\left(\mathcal{E}_{n,\theta}(|\theta_n^* - \theta|) > \varepsilon \right) \Big) \to 0,$$

где $\Phi(x)$ — функция распределения стандартного нормального закона. Тогда

$$\sup_{\theta\in\Theta,\sigma\in\Xi,x\in\mathbb{R}}\left|\mathbb{P}\left(I_{n,\theta}^{-1/2}J_{n,\theta}(\theta_n^{**}-\theta)<x\right)-\Phi(x)\right|\to 0.$$

3.1.4 Доказательство утверждений раздела 3.1.3

Для вывода теоремы 3.3 нам потребуется

Лемма 3.3. *Пусть выполнены условия (A) и (335). Тогда*

$$\rho_{n3}=I_n^{-1/2}\big(\psi_{n,2}(\theta_n^{**})-\psi_{n,2}(\theta)\big)\xrightarrow{p}0\quad \textit{при}\quad \psi_{n,2}^2(t)=\sum_{i=1}^n\psi_i^2(t,X_i) \tag{336}$$

и статистика Q_n^* определена с вероятностью, стремящейся к 1.

Д о к а з а т е л ь с т в о. Положим $u_n^{**}=\theta_n^{**}-\theta$. Имеют место следующие соотношения:

$$I_n^{-1/2}J_nu_n^{**}=O_p(1),\qquad \sum_{i=1}^n w_i(|u_n^{**}|,X_i)\xrightarrow{p}0. \tag{337}$$

Первое из этих соотношений следует из сходимости $I_n^{-1/2}J_nu_n^{**}\xrightarrow{d}\mathcal{N}_1(0,1)$, доказанной в теореме 3.1. Второе – из леммы 3.1 при $\eta_n=|u_n^{**}|$ и $\alpha_n(\cdot)=\sum_{i=1}^n w_i(\cdot,X_i)$, если только еще учесть равенство $\mathcal{E}_n(\cdot)=\sum_{i=1}^n\mathbb{E}w_i(\cdot,X_i)$ и тот факт, что $\mathcal{E}_n(|u_n^{**}|)\xrightarrow{p}0$ в силу первого соотношения в (337), сходимости $I_n^{-1/2}|J_n|\to\infty$ из (A_3) и условия (A_4).

Заметим теперь, что в силу определения (336) и условия (A_2), справедливы соотношения

$$\begin{aligned}
{I_n}^{1/2}|\rho_{n3}| &\le \left(\sum_{i=1}^n\left(\psi_i(\theta_n^{**},X_i)-\psi_i(\theta,X_i)\right)^2\right)^{1/2}=\left(\sum_{i=1}^n\left(\int_\theta^{\theta_n^{**}}D_i(t,X_i)dt\right)^2\right)^{1/2}\le\\
&\le\sqrt{2}\left(\sum_{i=1}^n\left(\int_\theta^{\theta_n^{**}}\left(D_i(t,X_i)-D_i(\theta,X_i)\right)dt\right)^2\right)^{1/2}+\sqrt{2}\left(\sum_{i=1}^n\left(\int_\theta^{\theta_n^{**}}D_i(\theta,X_i)dt\right)^2\right)^{1/2}\le\\
&\le\sqrt{2}|u_n^{**}|\left(\sum_{i=1}^n\left(\sup_{t:|t-\theta|\le|u_n^{**}|}|D_i(t,X_i)-D_i(\theta,X_i)|\right)^2\right)^{1/2}+\sqrt{2}|u_n^{**}|\left(\sum_{i=1}^nD_i^2(\theta,X_i)\right)^{1/2}\le\\
&\le\sqrt{2}|J_n||u_n^{**}|\sum_{i=1}^n\omega_i(|u_n^{**}|,X_i)+\sqrt{2}|u_n^{**}|\left(\sum_{i=1}^nD_i^2(\theta,X_i)\right)^{1/2}.
\end{aligned}$$

Следовательно, используя соотношения (335) и (337), получаем

$$|\rho_{n3}|\le\sqrt{2}I_n^{-1/2}|J_n||u_n^{**}|\sum_{i=1}^n\omega_i(|u_n^{**}|,X_i)+\sqrt{2}I_n^{-1/2}|J_n||u_n^{**}|\left(J_n^{-2}\sum_{i=1}^nD_i^2(\theta,X_i)\right)^{1/2}\xrightarrow{p}0$$

и первое утверждение леммы доказано.

По теореме 3.1 оценка θ_n^{**} является асимптотически нормальной, поэтому справедливо соотношение $\mathbb{P}(\theta_n^{**} \in \Theta) \to 1$. А поскольку еще $\mathbb{P}(\theta_n^{*} \in \Theta) \to 1$ (см. доказательство леммы 3.2), то с вероятностью, стремящейся к 1 определены величины $\sum_{i=1}^{n} D_i(\theta_n^*, X_i)$ и $\psi_{n,2}(\theta_n^{**})$. Кроме того, $I_n^{-1/2}\psi_{n,2}(\theta_n^{**}) = \rho_{n3} + I_n^{-1/2}\psi_{n,2}(\theta)$, то из первого условия в (335) и сходимости (336) заключаем, что $\mathbb{P}\big(\psi_{n,2}(\theta_n^{**}) = 0\big) \to 0$. Таким образом, с учетом определения (334), статистика Q_n^* определена с вероятностью, стремящейся к 1. $\square$

Д о к а з а т е л ь с т в о теоремы 3.3. Используя тождество в (330), а также обозначения, введенные в (324) и (336), получаем следующее представление:

$$Q_n^*(\theta_n^{**} - \theta) = -\frac{I_n^{-1/2}\sum_{i=1}^{n}\psi_i(\theta, X_i) + \overline{\rho}_n(u_n^*) - \rho_n(u_n^*)}{I_n^{-1/2}\psi_{n,2}(\theta) + \rho_{n3}}. \tag{338}$$

Но $I_n^{-1/2}\psi_{n,2}(\theta) \xrightarrow{p} 1$ в силу первого условия в (335). Утверждения теоремы вытекают теперь из тождества (338), лемм 3.2, 3.3 и предельного соотношения (314) из условия (A_3). $\square$

Нам также потребуются следующие равномерные по параметру утверждения.

Лемма 3.4. *Пусть случайные последовательности $\zeta_{n1,\vartheta}$, $\zeta_{n2,\vartheta}$, $\zeta_{n3,\vartheta}$ и $\zeta_{n4,\vartheta}$, зависящие от параметра $\vartheta \in Q$ таковы, что*

$$\sup_{\vartheta\in Q, x\in\mathbb{R}} |\mathbb{P}(\zeta_{n1,\vartheta} < x) - \Phi(x)| \to 0$$

и для любого $\varepsilon > 0$

$$\sup_{\vartheta\in Q}\mathbb{P}(|\zeta_{n2,\vartheta} - 1| > \varepsilon) \to 0, \quad \sup_{\vartheta\in Q}\mathbb{P}(|\zeta_{n3,\vartheta}| > \varepsilon) \to 0, \quad \sup_{\vartheta\in Q}\mathbb{P}(|\zeta_{n4,\vartheta}| > \varepsilon) \to 0.$$

Тогда

$$\sup_{\vartheta\in Q, x\in\mathbb{R}} |\mathbb{P}(\zeta_{n,\vartheta} < x) - \Phi(x)| \to 0 \qquad \text{при} \quad \zeta_{n,\vartheta} = (\zeta_{n1,\vartheta} + \zeta_{n3,\vartheta})/(\zeta_{n2,\vartheta} + \zeta_{n4,\vartheta}).$$

Лемма 3.5. *Пусть $\alpha_{n,\vartheta}(\cdot) : [0,\infty) \to [0,\infty)$ – последовательность случайных процессов с монотонно неубывающими траекториями, зависящих дополнительно от некоторого параметра $\vartheta \in Q$ и имеющих конечные средние $b_{n,\vartheta}(\cdot) = \mathbb{E}\alpha_{n,\vartheta}(\cdot)$. Кроме того, при всех $\varepsilon > 0$ $\sup_{\vartheta\in Q}\mathbb{P}(b_{n,\vartheta}(\eta_n) > \varepsilon) \to 0$ для некоторой последовательности случайных величин $\eta_n \geq 0$. Тогда для любого $\varepsilon > 0$ имеет место равномерная сходимость $\sup_{\vartheta\in Q}\mathbb{P}(\alpha_{n,\vartheta}(\eta_n) > \varepsilon) \to 0$.*

Для доказательства леммы 3.5 нужно в новых условиях повторить вывод утверждения леммы 3.1. Доказательство леммы 3.4 элементарно, поэтому мы его опускаем.

Д о к а з а т е л ь с т в о теоремы 3.4. Используя представление из (330) и обозначения, введенные в (324), нетрудно извлечь тождество

$$I_n^{-1/2} J_n(\theta_n^{**} - \theta) = -\frac{I_n^{-1/2} \sum_{i=1}^n \psi_i(\theta, X_i) + \overline{\rho}_n(u_n^*) - \rho_n(u_n^*)}{\sum_{i=1}^n D_i(\theta, X_i) J_n^{-1} + R_n(u_n^*)}. \tag{339}$$

Остается воспользоваться леммой 3.4 при $\vartheta = (\theta, \sigma)$ и

$$\zeta_n = I_n^{-1/2} J_n(\theta_n^{**} - \theta), \quad \zeta_{n1} = I_n^{-1/2} \sum_{i=1}^n \psi_i(\theta, X_i), \quad \zeta_{n2} = \sum_{i=1}^n D_i(\theta, X_i) J_n^{-1},$$
$$\zeta_{n3} = \overline{\rho}_n(u_n^*) - \rho_n(u_n^*), \quad \zeta_{n4} = R_n(u_n^*),$$

при этом оценивание погрешностей ζ_{n3} и ζ_{n4} аналогично выводу леммы 3.2, если в этом доказательстве вместо леммы 3.1 использовать утверждение леммы 3.5. □

3.1.5 О точности предварительной оценки

Обсудим подробнее условия на точность предварительной оценки θ_n^*, используемые в теореме 3.1 об асимптотической нормальности одношаговой M-оценки θ_n^{**}. Оказывается, при некоторых достаточно широких ограничениях условие

$$\big(I_{n,\theta}^{-1} J_{n,\theta}^2\big)^{1/4} |\theta_n^* - \theta| \xrightarrow{p} 0 \tag{340}$$

является необходимым для справедливости следующего соотношения, введенного в (316):

$$\delta_{n,\theta} \equiv I_{n,\theta}^{-1/2} J_{n,\theta}\big(\theta_n^{**} - \theta\big) + I_{n,\theta}^{-1/2} \sum_{i=1}^n \psi_i(\theta, X_i) \xrightarrow{p} 0.$$

Таким образом, с учетом (A_3), условие (340) можно считать в некотором смысле минимальным достаточным ограничением для доказательства асимптотической нормальности (317) одношаговой оценки θ_n^{**}. Напомним, что в случае независимых и одинаково распределенных наблюдений ограничение (340) есть условие $n^{1/4}$-состоятельности оценки θ_n^*.

Чтобы сформулировать точное утверждение, нам потребуется дополнительное условие.

$(\mathbf{A}_6)$*При всех i функции $\psi_i(t, X_i)$ с вероятностью* 1 *дважды дифференцируемы по t и для любого $\varepsilon > 0$*

$$P(\delta) = \limsup \mathbb{P}\Big(\sum_{i=1}^n \sup_{t:|t-\theta|\le\delta} |\psi_i''(t, X_i) - \psi_i''(\theta, X_i)|/|J_{n,\theta}| > \varepsilon\Big) \to 0 \textit{ при } \delta \to 0.$$

Кроме того, $\overline{\psi_n''}(\theta) = \sum_{i=1}^n \big(\psi_i''(\theta, X_i) - \mathbb{E}\psi_i''(\theta, X_i)\big)/J_{n,\theta} \xrightarrow{p} 0.$

Теорема 3.5. *Пусть выполнены условия* (A), (A_6) *и предварительная оценка* θ_n^* *состоятельна. Тогда справедливо соотношение*

$$\delta_{n,\theta} = I_{n,\theta}^{-1/2} J_{n,\theta}(\theta_n^* - \theta)^2 \left(\sum_{i=1}^{n} \mathbb{E}\psi_i''(\theta, X_i)/(2J_{n,\theta}) + o_p(1) \right) + o_p(1). \tag{341}$$

Если дополнительно $\liminf \big| \sum\limits_{i=1}^{n} \mathbb{E}\psi_i''(\theta, X_i) \big| / |J_{n,\theta}| \geq \delta$ *при некотором* $\delta > 0$, *то соотношение* (316) *имеет место тогда и только тогда, когда выполнено* (340).

Таким образом, с учетом замечания 3.3, асимптотическая нормальность одношаговых M-оценок установлена в широком спектре ограничений на точность предварительной оценки θ_n^*. В завершении раздела приведем простой пример, касающийся различных скоростей сближения оценки θ_n^* и параметра θ.

П р и м е р 3.1. Пусть $\{X_i\}$ — последовательность независимых одинаково распределенных случайных величин, имеющих плотность распределения

$$f_\theta(x) = \frac{h+1}{\theta(1 + x/\theta)^{2+h}} \qquad \text{при} \quad x \geq 0,$$

где $h \in (0, 1]$ фиксировано. Поскольку $\mathbb{E}X_1 = \theta/h$, то оценка метода моментов для параметра θ есть $\theta_n^* = h\sum_{i=1}^n X_i/n$. Эта оценка может быть использована в качестве предварительной для аппроксимации оценки максимального правдоподобия. Заметим, что распределения $n^\alpha(\theta_n^* - \theta)$ при $\alpha = h/(h+1) \in (0, 1/2)$ слабо сходятся к устойчивому (см. раздел 3.3.2, замечание 3.16). Таким образом, значение h определяет скорость сближения θ_n^* и θ. Этот пример нетрудно модифицировать на случай разнораспределенных наблюдений. Например, положим $X_i = Y_i + \delta_i$, где $\{Y_i\}$ независимы и одинаково распределены с вышеуказанной плотностью и $\sum_{i=1}^n \delta_i/n^{1-\alpha} \to 0$ при $\alpha = h/(h+1)$. В этом случае распределения $n^\alpha(\theta_n^* - \theta)$ при $\theta_n^* = h\sum_{i=1}^n X_i/n$ слабо сходятся к устойчивому. Примеры предварительных оценок в задачах нелинейной регрессии, имеющих различные скорости сходимости, приведены в главе 2.

3.1.6 О ближайшем к параметру корне уравнения

Обозначим через $\widetilde{\theta}_n(t)$ ближайшее к t решение уравнения

$$\psi_n(t) = 0, \qquad \text{где} \quad \psi_n(t) = \sum_{i=1}^{n} \psi_i(t, X_i). \tag{342}$$

Нам также потребуется вариант закона больших чисел

$$\sum_{i=1}^{n} \sup_{t:|t-\theta|\leq\delta} D_i(t, X_i) \left(\sum_{i=1}^{n} \mathbb{E} \sup_{t:|t-\theta|\leq\delta} D_i(t, X_i) \right)^{-1} \xrightarrow{p} 1 \tag{343}$$

при некотором $\delta > 0$. Следующее утверждение описывает поведение $\widetilde{\theta}_n(\theta)$.

Теорема 3.6. *Пусть выполнены условия* (A_1)–(A_3) *и* $\limsup J_{n,\theta} < 0$. *Кроме того, при некотором* $0 < \delta_\theta \le \infty$ *справедливо неравенство*

$$\limsup \mathcal{E}_{n,\theta}(\delta_\theta) < 1 \tag{344}$$

и соотношение (343) *имеет место при* $\delta = \delta_\theta$. *Тогда с вероятностью, стремящейся к* 1, *на интервале* $(\theta-\delta_\theta, \theta+\delta_\theta)$ *функция* $\psi_n(t)$, *определенная в* (342), *строго убывает и меняет знак, т.е. с вероятностью, стремящейся к* 1, *на этом интервале существует единственное решение* $\widetilde{\theta}_n(\theta)$ *уравнения* (342).

Если дополнительно выполнено условие (A_4) *и сходимость* (343) *имеет место при всех* $\delta = \delta_\theta$ *таких, что выполнено* (344), *то*

$$I_{n,\theta}^{-1/2} J_{n,\theta}\big(\widetilde{\theta}_n(\theta) - \theta\big) \xrightarrow{d} \mathcal{N}_1\big(0, 1\big). \tag{345}$$

З а м е ч а н и е 3.8. Если $\liminf J_{n,\theta} > 0$ и в условиях теоремы 3.6 вместо предельного соотношения (343) выполнено

$$\sum_{i=1}^n \inf_{t:|t-\theta|\le\delta} D_i(t, X_i) \left(\sum_{i=1}^n \mathbb{E} \inf_{t:|t-\theta|\le\delta} D_i(t, X_i) \right)^{-1} \xrightarrow{p} 1, \tag{346}$$

то все утверждения теоремы 3.6 сохранятся (но функция $\psi_n(t)$ будет строго возрастающей). Заметим также (см., например, [324]), что в случае независимых наблюдений условие (343) выполнено, если имеет место сходимость $\sum\limits_{i=1}^n \mathbb{E} \min\big\{|\xi_{i,n}|, |\xi_{i,n}|^2\big\} \to 0$ при

$$\xi_{i,n} = \left(\sup_{t:|t-\theta|\le\delta} D_i(t, X_i) - \mathbb{E} \sup_{t:|t-\theta|\le\delta} D_i(t, X_i) \right) \left(\sum_{i=1}^n \mathbb{E} \sup_{t:|t-\theta|\le\delta} D_i(t, X_i) \right)^{-1}.$$

Аналогично, заменой супремума на инфимум в вышеприведенной формуле, получим достаточное условие для выполнения (346). □

Таким образом, согласно теореме 3.6, при весьма слабых ограничениях в окрестности параметра θ существует единственный корень $\widetilde{\theta}_n(\theta)$ уравнения (342), удовлетворяющий (345). Этот корень уравнения и приближают одношаговые M-оценки. Стоит отметить, что $\widetilde{\theta}_n(\theta)$ не является статистикой, поскольку эта величина зависит не только от выборки, но и от неизвестного параметра θ. Однако в случае, когда известна некоторая предварительная состоятельная оценка θ_n^*, суперпозиция $\widetilde{\theta}_n(\theta_n^*)$, являющаяся M-оценкой (статистикой), обладает свойством: $\mathbb{P}(\widetilde{\theta}_n(\theta) = \widetilde{\theta}_n(\theta_n^*)) \to 1$. Иными словами, алгоритм построения оценки состоит в следующем: мы выбираем корень уравнения (342), ближайший к состоятельной оценке θ_n^*.

Отметим еще, что если $\widetilde{\theta}_n$ — произвольная состоятельная M-оценка, то в условиях теоремы 3.6 выполнено $\mathbb{P}\big(\widetilde{\theta}_n = \widetilde{\theta}_n(\theta)\big) \to 1$. Тем самым, из теоремы 3.6 и замечания 3.8 можно извлечь следующее утверждение об асимптотическом поведении собственно M-оценок.

Следствие 3.1. *Пусть выполнены условия* (A_1)–(A_4) *и* M*-оценка* $\widetilde{\theta}_n$ *состоятельна. Кроме того, либо* $\limsup J_{n,\theta} < 0$ *и условие* (343) *выполнено при всех* $\delta = \delta_\theta$ *таких, что верно* (344), *либо* $\liminf J_{n,\theta} > 0$ *и условие* (346) *выполнено при всех* $\delta = \delta_\theta$ *таких, что верно* (344). *Тогда*

$$I_{n,\theta}^{-1/2} J_{n,\theta}\big(\widetilde{\theta}_n - \theta\big) \xrightarrow{d} \mathcal{N}_1\big(0, 1\big). \qquad (347)$$

З а м е ч а н и е 3.9. Несколько иные достаточные условия для выполнения соотношения (347) можно извлечь из теоремы 3.1. Действительно, в качестве предварительной оценки использовать M-оценку, то одношаговая M-оценка совпадает с M-оценкой. Другими словами, если $\theta_n^* \equiv \widetilde{\theta}_n$, то $\theta_n^{**} \equiv \widetilde{\theta}_n$ (подобный эффект использовался, например, в [309]). Тем самым, из теоремы 3.1 получаем следующее утверждение:

Если выполнены условия (A_1)–(A_4) *и* $I_{n,\theta}^{-1/2} J_{n,\theta}(\widetilde{\theta}_n - \theta) = O_p(1)$, *то справедливо* (347). □

Далее приведем пример, когда имеются две M-оценки, они не состоятельны, в то время как ближайшее к θ решение $\widetilde{\theta}_n(\theta)$ соответствующего уравнения почти наверное сходится к θ.

П р и м е р 3.2. Пусть $X_1, \ldots, X_n$ независимые одинаково распределенные наблюдения из параметрического семейства $\mathcal{L}_\theta(\cdot) = \theta L^{(1)}(\cdot) + \theta^2 L^{(2)}(\cdot) + (1 - \theta - \theta^2) L^{(3)}(\cdot)$, где $L^{(k)}$ есть некоторые распределения (в общем случае, не известные) с известными средними $a_k = \displaystyle\int t L^{(k)}(dt)$, $k = 1, 2, 3$. Предположим, для простоты, что $a_3 - a_1 = a_2 - a_3 = \delta > 0$. Определим множество

$$\Theta = \{\theta > 0 : \theta^2 + \theta < \ 1,\ \theta \neq 1/2\} \equiv \{\theta : 0 < \theta < (\sqrt{5} - 1)/2,\ \theta \neq 1/2\}$$

и положим $\psi(\theta, X_i) = X_i - \theta a_1 - \theta^2 a_2 - (1 - \theta - \theta^2) a_3$. Заметим, что при сделанных выше предположениях $\mathbb{E}\psi'(\theta, X_i) \neq 0$ при всех $\theta \in \Theta$. Для оценивания параметра θ используем следующее уравнение:

$$0 = \sum_{i=1}^n \psi(\theta, X_i) = n\overline{X} - n\big(\theta a_1 + \theta^2 a_2 + (1 - \theta - \theta^2) a_3\big), \qquad \text{где} \quad \overline{X} = \sum_{i=1}^n X_i / n.$$

Таким образом, нас интересует квадратное уравнение $\theta^2 \delta - \theta \delta + a_3 - \overline{X} = 0$. Поскольку $a_3 - \overline{X} \to (\theta - \theta^2)\delta$ почти наверное, то асимптотически дискриминант этого квадратного уравнения есть $(1 - 2\theta)^2$ и положителен при всех $t \in \Theta$, тем самым корни вещественны с вероятностью, стремящейся к 1. Имеем

$$\widetilde{\theta}_{n1,2} = \mathrm{Re}\Big(1 \pm \sqrt{1 - 4(a_3 - \overline{X})/\delta}\Big)/2 \xrightarrow{p} (1 \pm |1 - 2\theta|)/2.$$

То есть с вероятностью, стремящейся к 1, статистики $\widetilde{\theta}_{n1} < \widetilde{\theta}_{n2}$ являются корнями вышеуказанного квадратного уравнения. Согласно вышеприведенному определению, они являются M-оценками. Однако они не состоятельны. В самом деле, $\widetilde{\theta}_{n1} \to \theta$ почти наверное при $0 < \theta < 1/2$ и $\widetilde{\theta}_{n2} \to \theta$ почти наверное при $1/2 < \theta \le (\sqrt{5}-1)/2$. Ближайший к θ корень $\widetilde{\theta}_n(\theta)$ рассматриваемого уравнения при достаточно больших n определяется следующим соотношением:

$$\widetilde{\theta}_n(\theta) = \begin{cases} \widetilde{\theta}_{n1}, & \text{если} \;\; 0 < \theta < 1/2, \\ \widetilde{\theta}_{n2}, & \text{если} \;\; 1/2 < \theta \le (\sqrt{5}-1)/2, \end{cases}$$

при этом $\widetilde{\theta}_n(\theta) \to \theta$ почти наверное. $\square$

3.1.7 Доказательство утверждений разделов 3.1.5 и 3.1.6

Для доказательства теоремы 3.5 нам потребуется вспомогательное утверждение.

Лемма 3.6. *Пусть выполнены условия теоремы* 3.5. *Тогда*

$$\rho_{n4} = \left| \rho_n(u_n^*) - \overline{\rho}_n(u_n^*) - (u_n^*)^2 I_n^{-1/2} \sum_{i=1}^n \mathbb{E}\psi_i''(\theta, X_i)/2 \right| = I_n^{-1/2} J_n(u_n^*)^2 o_p(1).$$

Д о к а з а т е л ь с т в о. Заметим, что $D_i(\cdot, X_i) \equiv \psi_i'(\cdot, X_i)$ при всех i. Положим

$$\mu_i(u, X_i) = \begin{cases} \big(\psi_i'(\theta+u, X_i) - \psi_i'(\theta, X_i)\big)/u - \psi_i''(\theta, X_i), \; \text{если} \; u \ne 0, \\ 0, \; \text{если} \; u = 0, \end{cases} \tag{348}$$

$$\overline{\mu}_n(\delta) = |J_n|^{-1} \sum_{i=1}^n \sup_{u: |u| \le \delta, \theta+u \in \Theta} |\mu_i(u, X_i)|. \tag{349}$$

Покажем прежде всего, что

$$\overline{\mu}_n(u_n^*) \xrightarrow{p} 0. \tag{350}$$

В самом деле, поскольку $\mu_i(u, X_i) = \int\limits_0^1 \psi_i''(\theta + uv, X_i) dv - \psi_i''(\theta, X_i)$, то

$$\overline{\mu}_n(\delta) \le |J_n|^{-1} \sum_{i=1}^n \sup_{t \in \Theta: |t-\theta| \le \delta} |\psi_i''(t, X_i) - \psi_i''(\theta, X_i)|.$$

Таким образом, при всех неслучайных $\varepsilon > 0$ и $\alpha > 0$ и подходящего $\delta = \delta(\alpha) > 0$ выполнено

$$\mathbb{P}\big(\overline{\mu}_n(u_n^*) > \varepsilon\big) \le \mathbb{P}\big(|u_n^*| > \delta\big) + \mathbb{P}\big(\overline{\mu}_n(\delta) > \varepsilon\big),$$
$$\limsup \mathbb{P}\big(\overline{\mu}_n(u_n^*) > \varepsilon\big) \le \limsup \mathbb{P}\big(\overline{\mu}_n(\delta) > \varepsilon\big) \le P(\delta) < \alpha. \tag{351}$$

Здесь мы еще учли, что $P(\delta) \to 0$ при $\delta \to 0$ в силу (A_6). Предельное соотношение (350) установлено.

Завершим доказательство леммы. Сравнивая определения (324), (348) и обозначение $\overline{\psi_n''}(\theta)$ из (A_6), получаем

$$I_n^{1/2}\rho_n(vu_n^*)/v - v(u_n^*)^2\sum_{i=1}^n \mathbb{E}\psi_i''(\theta, X_i) = v(u_n^*)^2 J_n\overline{\psi_n''}(\theta) + v(u_n^*)^2\sum_{i=1}^n \mu_i(vu_n^*, X_i).$$

Вместе с (349) отсюда следует, что

$$\sup_{v\in[0,1]}\big|\rho_n(vu_n^*)/v - v(u_n^*)^2 I_n^{-1/2}\sum_{i=1}^n \mathbb{E}\psi_i''(\theta, X_i)\big| \le I_n^{-1/2}|J_n||u_n^*|^2\big(|\overline{\psi_n''}(\theta)| + \overline{\mu}_n(u_n^*)\big). \tag{352}$$

Далее, с учетом (324), (328) и (348), имеем

$$\rho_{n4} = \Big|\rho_n(u_n^*) - (u_n^*)^2 I_n^{-1/2}\sum_{i=1}^n \mathbb{E}\psi_i''(\theta, X_1)\Big) - \int_0^1 \Big(\rho_n(vu_n^*)/v - v(u_n^*)^2 I_n^{-1/2}\sum_{i=1}^n \mathbb{E}\psi_i''(\theta, X_1)\Big)dv\Big|.$$

Принимая во внимание (352), (A_6), и (350), получаем

$$\rho_{n4} \le 2\sup_{v\in[0,1]}\Big|\rho_n(vu_n^*)/v - v(u_n^*)^2 I_n^{-1/2}\sum_{i=1}^n \mathbb{E}\psi_i''(\theta, X_i)\Big| = I_n^{-1/2}|J_n|(u_n^*)^2 o_p(1),$$

и это соотношение завершает доказательство леммы. □

Д о к а з а т е л ь с т в о теоремы 3.5. Используя (330), (331), а также сходимость $H_n \xrightarrow{p} 0$, доказанную при выводе теоремы 3.1, получаем следующее соотношение:

$$\delta_n = \big(o_p(1) + \rho_n(u_n^*) - \overline{\rho}_n(u_n^*)\big)\big(1 + o_p(1)\big)^{-1}.$$

Утверждение (341) теоремы следует теперь из этого соотношения и леммы 3.6. Второе утверждение теоремы очевидно. □

Д о к а з а т е л ь с т в о теоремы 3.6. Положим $D_n(t) = \sum\limits_{i=1}^n D_i(t, X_i)$. Пусть $\delta = \delta_\theta$ таково, что выполнены соотношения (343) и (344). Имеем

$$\sup_{t:|t-\theta|\le\delta_\theta} D_n(t)/|J_n| \le \sum_{i=1}^n \sup_{t:|t-\theta|\le\delta_\theta} D_i(t, X_i)/|J_n| = \zeta_n\sum_{i=1}^n \mathbb{E}\sup_{t:|t-\theta|\le\delta_\theta} D_i(t, X_i)/|J_n|,$$

где через ζ_n обозначено выражение в левой части условия (343), ввиду которого $\zeta_n \xrightarrow{p} 1$. При этом, поскольку $\limsup \mathcal{E}_n(\delta_\theta) < 1$, то существует $\varepsilon > 0$ такое, что при всех достаточно больших n выполнено $\mathcal{E}_n(\delta_\theta) < 1 - \varepsilon$ и

$$\sum_{i=1}^n \mathbb{E} \sup_{t:|t-\theta|\le\delta_\theta} D_i(t, X_i)/|J_n| \le \sum_{i=1}^n \mathbb{E} D_i(\theta, X_i)/|J_n| + \mathcal{E}_n(\delta_\theta) \le -\varepsilon < 0.$$

Наконец, с учетом второй сходимости в (330) и условия $I_n^{1/2}|J_n|^{-1} \to 0$

$$|J_n|^{-1}\psi_n(t) = |J_n|^{-1}\psi_n(\theta) + |J_n|^{-1}\int_\theta^t D_n(z)dz = o_p(1) + |J_n|^{-1}\int_\theta^t D_n(z)dz.$$

Поскольку функция $D_n(t)$ отделена от нуля, то $\psi_n(\theta - \delta_\theta) > 0$ и $\psi_n(\theta + \delta_\theta) < 0$. Следовательно, с вероятностью, стремящейся к 1, строго убывающая на интервале $(\theta - \delta_\theta, \theta + \delta_\theta)$ функция $\psi_n(t)$ меняет знак на концах интервала. Тем самым, на этом интервале существует единственный нуль $\widetilde{\theta}_n(\theta)$ этой функции.

Из равенства

$$0 = \psi_n\big(\widetilde{\theta}_n(\theta)\big) = \psi_n(\theta) + \big(\widetilde{\theta}_n(\theta) - \theta\big)\int_0^1 D_n\big(\theta + v(\widetilde{\theta}_n(\theta) - \theta)\big)dv,$$

имеем

$$I_n^{-1/2} J_n\big(\widetilde{\theta}_n(\theta) - \theta\big) = -I_n^{-1/2}\psi_n(\theta)\left(J_n^{-1}D_n(\theta) + \int_0^1 R_n\big(v(\widetilde{\theta}_n(\theta) - \theta)\big)dv\right)^{-1}. \qquad (353)$$

Далее, $\limsup \mathcal{E}_n(\delta) \to 0$ при $\delta \to 0$ и условие (343) выполнено при всех δ таких, что справедливо (344). Следовательно, в силу доказанного при всех сколь угодно малых δ на интервале $(\theta - \delta, \theta + \delta)$ с вероятностью, стремящейся к 1, существует единственный локальный нуль $\widetilde{\theta}_n(\theta)$. Таким образом, $\widetilde{\theta}_n(\theta) \xrightarrow{p} \theta$. Нетрудно видеть, что $\mathcal{E}_n\big(|\widetilde{\theta}_n(\theta) - \theta|\big) \xrightarrow{p} 0$. А поскольку еще

$$\left|\int_0^1 R_n\big(v(\widetilde{\theta}_n(\theta) - \theta)\big)dv\right| \le \sum_{i=1}^n \omega_{i,\theta}\big(|\widetilde{\theta}_n(\theta) - \theta|, X_i\big) \quad \text{и} \quad \sum_{i=1}^n \mathbb{E}\omega_{i,\theta}(|u|, X_i) = \mathcal{E}_n(|u|),$$

то в силу леммы 3.1

$$\int_0^1 R_n\big(v(\widetilde{\theta}_n(\theta) - \theta)\big)dv \xrightarrow{p} 0.$$

Эта сходимость вместе с представлением (353) и предельными соотношениями из (330) влекут (345). $\square$

3.1.8 К вопросу о k-шаговом оценивании

Рассмотрим некоторые аспекты k-шагового оценивания, связанные с тематикой нашего исследования. Обозначим через $\theta^{**}_{n,(k)}$ результат k-ой итерации метода Ньютона со стартовой точкой θ^*_n. Таким образом, для оценки θ^{**}_n, определенной в (315), имеем $\theta^{**}_n \equiv \theta^{**}_{n,(1)}$. Как уже отмечалось выше, k-шаговые оценки $\theta^{**}_{n,(k)}$ можно использовать, во-первых, для приближения параметра θ. Во-вторых, k-шаговыми оценками $\theta^{**}_{n,(k)}$ можно аппроксимировать M-оценки $\widetilde{\theta}_n$ (точнее, приближать ближайший к θ корень уравнения (342), который с вероятностью, стремящейся к 1, совпадает с состоятельной M-оценкой $\widetilde{\theta}_n$ в случае существования последней).

Рассмотрим обе указанные ситуации более детально. Первая есть классическая статистическая задача оценивания параметра. Нетрудно видеть, что при некоторых ограничениях для одношаговой M-оценки $\theta^{**}_n = \theta^{**}_{n,(1)}$ из (315) справедливы следующие асимптотические представления:

$$\theta^{**}_{n,(1)} - \theta = \frac{(\theta^*_n - \theta)\psi'_n(\theta^*_n) - \psi_n(\theta^*_n)}{\psi'_n(\theta^*_n)} \sim \frac{\zeta_n}{Q_{n,\theta}} + (\theta^*_n - \theta)^2 \mu_n, \tag{354}$$

где функция $\psi_n(t)$ определена в (342), символ «$\sim$» означает эквивалентность по вероятности и

$$\begin{aligned} Q_{n,\theta} &= I^{-1/2}_{n,\theta} J_{n,\theta}, \quad |Q_{n,\theta}| \to \infty, \\ \zeta_n &= -I^{-1/2}_{n,\theta}\psi_n(\theta) \xrightarrow{d} \mathcal{N}_1(0,1), \\ \mu_n &= 2^{-1}\psi''_n(\theta)/\psi'_n(\theta), \quad \mathbb{P}(a \le |\mu_n| \le b) \to 1 \end{aligned} \tag{355}$$

для некоторых неслучайных $0 < a < b < \infty$. Отметим, что в случае независимых одинаково распределенных наблюдений $Q_{n,\theta} = O(\sqrt{n})$.

Из (354) и (355) следует, что близость по вероятности одношаговой оценки $\theta^{**}_{n,1}$ и параметра θ не может быть лучше чем $O_p(|Q_{n,\theta}|^{-1})$ и этот порядок достигается только для $|Q_{n,\theta}|^{1/2}$-ограниченных предварительных оценок θ^*_n, т.е. в случае, когда величина $|Q_{n,\theta}|^{1/2}\big(\theta^*_n - \theta\big)$ ограничена по вероятности равномерно по n. Более того, если оценка θ^*_n является $|Q_{n,\theta}|^{1/2}$-состоятельной, т.е. $|Q_{n,\theta}|^{1/2}\big(\theta^*_n - \theta\big) \xrightarrow{p} 0$, то

$$Q_{n,\theta}(\theta^{**}_{n,(1)} - \theta) \sim \zeta_n \xrightarrow{d} \mathcal{N}_1(0,1).$$

Другими словами, если предварительная оценка θ^*_n является $|Q_{n,\theta}|^{1/2}$-состоятельной, то нет необходимости в использовании многошаговой итерационной процедуры. При широких ограничениях достаточно использовать лишь один шаг метода Ньютона, чтобы построить оценку с искомыми свойствами.

Таким образом, имеет смысл использовать k-шаговые процедуры только в случае, когда предварительная оценка сближается с параметром медленнее, чем в условии $|Q_{n,\theta}|^{1/2}$-состоятельности. Например, пусть для некоторого $\varepsilon < 1/2$ оценка θ_n^* является $|Q_{n,\theta}|^{\varepsilon}$-состоятельной. С учетом (354) и (355), оценка $\theta_{n,(1)}^{**}$ является $|Q_{n,\theta}|^{2\varepsilon}$-состоятельной. Далее, подставим оценку $\theta_{n,(1)}^{**}$ вместо θ_n^* в (354). Тогда новая оценка $\theta_{n,(2)}^{**}$ будет $|Q_{n,\theta}|^{4\varepsilon}$-состоятельной, если $4\varepsilon < 1$. Нетрудно видеть, что минимальное число k итераций (354), необходимое для построения k-шаговой оценки $\theta_{n,(k)}^{**}$, удовлетворяющей соотношению $Q_{n,\theta}(\theta_{n,(k)}^{**} - \theta) \xrightarrow{d} \mathcal{N}_1(0,1)$, есть $k = [-\log\varepsilon/\log 2] + 1$.

Рассмотрим теперь вторую постановку, связанную с приближением ближайшего к параметру θ корня $\widetilde{\theta}_n(\theta)$ уравнения (342). Нетрудно получить, что при некоторых ограничениях

$$\theta_{n,(1)}^{**} - \widetilde{\theta}_n(\theta) \sim \big(\theta_n^* - \widetilde{\theta}_n(\theta)\big)^2 \mu_n, \qquad Q_{n,\theta}\big(\widetilde{\theta}_n(\theta) - \theta\big) \xrightarrow{d} \mathcal{N}_1(0,1), \tag{356}$$

где величина μ_n определена в (355). Первое рекуррентное соотношение в (356) фактически иллюстрирует хорошо известную теорему Канторовича, касающуюся квадратичной скорости сходимости итерационного метода Ньютона. Принципиальная разница между этим асимптотическим соотношением (356) и соответствующим соотношением в (354) состоит в наличии первого слагаемого в правой части (354), которое отсутствует в (356). Именно это слагаемое определяет порядок близости k-шаговой оценки и параметра. Таким образом, в отличие от оценивания истинного значения параметра в (354), аппроксимировать корень $\widetilde{\theta}_n(\theta)$ можно с любой точностью (при наличии состоятельной оценки θ_n^* с известной скоростью сходимости). Например, если θ_n^* является $|Q_{n,\theta}|^{\varepsilon}$-состоятельной при некотором положительном $\varepsilon < 1$, тогда $\theta_{n,(k)}^{**} - \widetilde{\theta}_n(\theta) = o_p\big(|Q_{n,\theta}|^{-2^k\varepsilon}\big)$. Но для классической статистической постановки оценивания истинного значения параметра подобная точность не играет роли, поскольку порядок близости этого корня к истинному значению, определяемый вторым соотношением в (356), задает необходимый порядок близости соответствующих итераций в первом соотношении в (356).

Некоторые версии k-шаговых оценок рассматривались, например, в [138], [148], [280] и др.

3.2 Взвешенные M-оценки и их одношаговые приближения

В этом разделе будут получены условия асимптотической нормальности одношаговых взвешенных M-оценок. Эти оценки являются приближениями для состоятельных взвешенных M-оценок. Напомним, что взвешенные M-оценки определяются уравнением

$$\sum_{i=1}^{n} h_i(\theta_n^*)\widetilde{\psi}_i(t, X_i) = 0,$$

где θ_n^* — некоторая предварительная оценка параметра θ. В дальнейшем символ «тильда» у функций $\widetilde{\psi}_i(t, X_i)$ мы для удобства опускаем.

3.2.1 Асимптотические свойства оценок

Нам потребуются следующие модификации $(A_1^h), \dots, (A_5^h)$ условий $(A_1), \dots, (A_5)$, введенных в разделе 3.1.

($\mathbf{A}_1^h$). *Наблюдается выборка объема n, состоящая из независимых элементов X_1, X_2, $\dots$, X_n со значениями в произвольном измеримом пространстве $\mathcal{X}$ и распределениями $\mathcal{L}_{1,\theta}$, $\mathcal{L}_{2,\theta}, \dots, \mathcal{L}_{n,\theta}$, зависящими от интересующего нас основного неизвестного параметра $\theta \in \Theta \subset \mathbb{R}$, где Θ — открытое множество. Кроме того, эти распределения могут зависеть от n и некоторого дополнительного параметра σ произвольной природы.*

($\mathbf{A}_2^h$). *На множестве $\Theta \times \mathcal{X}$ при любом i заданы зависящие, вообще говоря, от n измеримые функции $\psi_i(t,x)$ и $\psi_i'(t,x)$; при этом для каждого интервала (t_1,t_2), целиком лежащего в Θ, при любом $x \in \mathcal{X}$ имеет место равенство*

$$\psi_i(t_2,x) - \psi_i(t_1,x) = \int\limits_{t_1}^{t_2} \psi_i'(t,x)dt \tag{357}$$

и справедливы соотношения

$$\mathbb{E}\psi_i(\theta,X_i) = 0, \qquad \mathbb{E}\psi_i^2(\theta,X_i) < \infty, \qquad \mathbb{E}|\psi_i'(\theta,X_i)| < \infty. \tag{358}$$

Кроме того, на множестве Θ при любом i заданы зависящие от n функции $h_i(t)$, удовлетворяющие условию Гельдера с показателем $p \in (0,1]$:

$$|h_i(t_1) - h_i(t_2)| \le H_i|t_1 - t_2|^p \qquad \forall t_1, t_2 \in \Theta.$$

З а м е ч а н и е 3.10. Здесь и в дальнейшем символ θ обозначает истинное значение основного параметра распределений и математическое ожидание в (358) и далее берется по распределению $\mathcal{L}_{i,\theta}$. Отметим еще, что для почти всех $t \in \Theta$ функции $\psi_i'(t,X_i)$ с вероятностью 1 совпадают с частной производной функции $\psi_i(t,X_i)$ по первому аргументу. Зависимость от σ, как и зависимость от n, далее указываться не будет. В операторах усреднения $\mathbb{E}$ и вероятностях $\mathbb{P}$ указываться не будет и зависимость от θ.

($\mathbf{A}_3^h$). *Для всех достаточно больших n*

$$I_{n,\theta,h} = \sum_{i=1}^n h_i^2(\theta)\mathbb{E}\psi_i^2(\theta,X_i) > 0, \qquad J_{n,\theta,h} = \sum_{i=1}^n h_i(\theta)\mathbb{E}\psi_i'(\theta,X_i) \ne 0; \tag{359}$$

кроме того, имеют место предельные соотношения: $I_{n,\theta,h}^{-1/2}|J_{n,\theta,h}| \to \infty$,

$$J_{n,\theta,h}^{-1}\sum_{i=1}^n h_i(\theta)\psi_i'(\theta,X_i) \xrightarrow{p} 1, \qquad I_{n,\theta,h}^{-1/2}\sum_{i=1}^n h_i(\theta)\psi_i(\theta,X_i) \xrightarrow{d} \mathcal{N}_1(0,1). \tag{360}$$

$(\mathbf{A}_4^h)$. *Справедливо предельное соотношение* $\limsup \mathcal{E}_{n,\theta,h}(\delta) \to 0$ *при* $\delta \to 0$, *где*

$$\mathcal{E}_{n,\theta,h}(\delta) = \sum_{i=1}^{n} \big(|h_i(\theta)| + \delta^p H_i\big)\mathbb{E}\omega_{i,\theta}(\delta, X_i)$$

и величины $\omega_{i,\theta}(\delta, x)$ *определены следующим образом:*

$$\omega_{i,\theta}(\delta, x) = \begin{cases} \sup\limits_{t\colon |t-\theta|\le\delta} \big|\psi_i'(t,x) - \psi_i'(\theta,x)\big|/|J_{n,\theta,h}|, & \text{если } [\theta-\delta, \theta+\delta] \subset \Theta, \\ \infty, & \text{иначе.} \end{cases} \tag{361}$$

$(\mathbf{A}_5^h)$. *Предварительная оценка* θ_n^* *удовлетворяет условиям*

$$I_{n,\theta,h}^{-1/2} J_{n,\theta,h} |\theta_n^* - \theta| \mathcal{E}_{n,\theta,h}(|\theta_n^* - \theta|) \xrightarrow{p} 0,$$
$$J_{n,\theta,h}^{-1} |\theta_n^* - \theta|^p \sum_{i=1}^{n} H_i \mathbb{E}|\psi_i'(\theta, X_i)| \xrightarrow{p} 0.$$

Обозначим через (A^h) группу условий, которая включает в себя предположения (A_1^h), ..., (A_5^h). Одношаговую взвешенную M-оценку $\theta_{n,h}^{**}$ зададим равенством

$$\theta_{n,h}^{**} = \theta_n^* - \left(\sum_{i=1}^{n} h_i(\theta_n^*)\psi_i'(\theta_n^*, X_i)\right)^{-1} \sum_{i=1}^{n} h_i(\theta_n^*)\psi_i(\theta_n^*, X_i) \tag{362}$$

во всех случаях, когда определена правая часть в (362). Имеет место

Теорема 3.7. *Пусть выполнены предположения* (A^h) *и условие*

$$\rho_n = I_{n,\theta,h}^{-1/2} \sum_{i=1}^{n} \big(h_i(\theta_n^*) - h_i(\theta)\big)\psi_i(\theta, X_i) \xrightarrow{p} 0. \tag{363}$$

Тогда $\theta_{n,h}^{**}$ *определена с вероятностью, стремящейся к* 1, *и*

$$I_{n,\theta,h}^{-1/2} J_{n,\theta,h} \big(\theta_{n,h}^{**} - \theta\big) \xrightarrow{d} \mathcal{N}_1(0,1). \tag{364}$$

Положим

$$Q_{n,h}^* = \left(\sum_{i=1}^{n} h_i^2(\theta_{n,h}^{**})\psi_i^2(\theta_{n,h}^{**}, X_i)\right)^{-1/2} \sum_{i=1}^{n} h_i(\theta_n^*)\psi_i'(\theta_n^*, X_i) \tag{365}$$

во всех случаях, когда определена правая часть этого равенства.

Теорема 3.8. *Если выполнены условия теоремы* 3.7 *и*

$$I_{n,\theta,h}^{-1}\sum_{i=1}^{n}h_i^2(\theta)\psi_i^2(\theta,X_i)\xrightarrow{p}1,\quad J_{n,\theta,h}^{-2}\sum_{i=1}^{n}\big(H_i^2+h_i^2(\theta)\big)\big(\psi_i'(\theta,X_i)\big)^2\xrightarrow{p}0,$$
$$J_{n,\theta,h}^{-2p}I_{n,\theta,h}^{p-1}\sum_{i=1}^{n}H_i^2\psi_i^2(\theta,X_i)\xrightarrow{p}0, \tag{366}$$

то величина $Q^*_{n,h}$ *определена с вероятностью, стремящейся к* 1 *и*

$$Q^*_{n,h}\big(\theta^{**}_{n,h}-\theta\big)\xrightarrow{d}\mathcal{N}_1(0,1). \tag{367}$$

З а м е ч а н и е 3.11. Обсудим условия, участвующие в двух вышеприведенных теоремах. Все утверждения из замечании 3.3 с очевидными изменениями сохраняются в условиях данного раздела, и соотношения вида (318) и (319) выполнены при замене $I_{n,\theta}$, $J_{n,\theta}$ и $\mathcal{E}_{n,\theta}(\delta)$ на $I_{n,\theta,h}$, $J_{n,\theta,h}$ и $\mathcal{E}_{n,\theta,h}(\delta)$, соответственно. В частности, для справедливости (A_4^h) и (A_5^h) достаточно, чтобы имели место следующие простые условия:

$$|\psi_i'(t,X_i)-\psi_i'(\theta,X_i)|\le\mu_i|t-\theta|^q\quad\text{при } q\in(0,1],\qquad n^{\frac{1}{2(1+q)}}(\theta_n^*-\theta)\xrightarrow{p}0,$$

$$1/I_{n,\theta,h}+1/|J_{n,\theta,h}|=O\big(1/n\big),\quad\sup_{n,i}\{|h_i(\theta)|,H_i,\ \mathbb{E}\mu_i,\ \mathbb{E}|\psi_i'(\theta,X_i)|\}<\infty,\ p>0. \tag{368}$$

Дополнительные условия в теореме 3.8 не являются ограничительными, поскольку первая сходимость в (366) — это вариант закона больших чисел, а для справедливости двух других сходимостей из (366) достаточно, чтобы вместе с первым соотношением в (368) выполнялось условие

$$\sup_{n,i}\{|h_i(\theta)|,H_i,\ \mathbb{E}|\psi_i'(\theta,X_i)|,\ \mathbb{E}\psi_i^2(\theta,X_i)\}<\ \infty.$$

Что касается технического условия (363), то в случае достаточно гладких функций $h_i(t)$ с помощью формулы Тейлора нетрудно получить простые достаточные условия для (363). Например, если функции $h_i(t)$ непрерывно дифференцируемы, при этом $|h_i'(t)-h_i'(\theta)|\le\overline{h}_i|t-\theta|^r$ при всех i и некотором $r\in(0,1]$, то достаточно требовать, чтобы

$$(\theta_n^*-\theta)I_{n,h}^{-1/2}\sum_{i=1}^{n}h_i'(\theta)\psi_i(\theta,X_i)\xrightarrow{p}0,\quad(\theta_n^*-\theta)^{1+r}I_{n,h}^{-1/2}\sum_{i=1}^{n}\overline{h}_i\mathbb{E}|\psi_i(\theta,X_i)|\xrightarrow{p}0.$$

Если же функции $h_i(t)$ удовлетворяют лишь условию Гельдера, то для проверки (363) можно использовать предложения 3.1 и 3.2 из раздела 3.2.3. □

З а м е ч а н и е 3.12. Если при выполнении условий (A^h) у нас имеется оценка $J^*_{n,h}$ для величины $J_{n,\theta,h}$ такая, что $J^*_{n,h}/J_{n,\theta,h}\xrightarrow{p}1$, то одну из разновидностей одношаговых взвешенных

M-оценок можно определить соотношением

$$\widehat{\theta}^{**}_{n,h} = \theta^*_n - \sum_{i=1}^n h_i(\theta^*_n)\psi_i(\theta^*_n, X_i)\big/J^*_{n,h}.$$

При некоторых дополнительных ограничениях имеют место следующие аналоги утверждений двух вышеприведенных теорем:

$$I^{-1/2}_{n,\theta,h} J_{n,\theta,h}\big(\widehat{\theta}^{**}_{n,h} - \theta\big) \xrightarrow{d} \mathcal{N}_1(0,1),$$

$$\widehat{Q}^*_{n,h}\big(\widehat{\theta}^{**}_{n,h} - \theta\big) \xrightarrow{d} \mathcal{N}_1(0,1) \quad \text{при} \quad \widehat{Q}^*_{n,h} = J^*_{n,h}\left(\sum_{i=1}^n h_i^2(\widehat{\theta}^{**}_{n,h})\psi_i^2(\widehat{\theta}^{**}_{n,h}, X_i)\right)^{-1/2}.$$

Точные условия мы опускаем. □

3.2.2 Доказательства результатов раздела 3.2.1

Сохраним обозначение $u^*_n = \theta^*_n - \theta$ и положим $I_{n,h} = I_{n,\theta,h}$, $J_{n,h} = J_{n,\theta,h}$, $\mathcal{E}_{n,h} = \mathcal{E}_{n,\theta,h}$,

$$\begin{aligned}
\rho_{n5}(u) &= \sum_{i=1}^n h_i(\theta+u)\big(\psi_i'(\theta+u, X_i) - \psi_i'(\theta, X_i)\big),\\
\rho_{n6}(u) &= \sum_{i=1}^n h_i(\theta+u)\big(\psi_i(\theta+u, X_i) - \psi_i(\theta, X_i) - u\psi_i'(\theta, X_i)\big),\\
\rho_{n7}(u) &= \sum_{i=1}^n \big(h_i(\theta+u) - h_i(\theta)\big)\psi_i'(\theta, X_i).
\end{aligned} \tag{369}$$

Лемма 3.7. *Если выполнены условия* (A^h), *то*

$$J^{-1}_{n,h}\rho_{n5}(u^*_n) \xrightarrow{p} 0, \quad I^{-1/2}_{n,h} u^*_n \rho_{n5}(u^*_n) \xrightarrow{p} 0, \quad I^{-1/2}_{n,h}\rho_{n6}(u^*_n) \xrightarrow{p} 0, \quad J^{-1}_{n,h}\rho_{n7}(u^*_n) \xrightarrow{p} 0.$$

Доказательство. Заметим прежде всего, что при выполнении (A^h)

$$I^{-1/2}_{n,h} J_{n,h}|u^*_n|\mathcal{E}_{n,h}(|u^*_n|) \xrightarrow{p} 0, \qquad \mathcal{E}_{n,h}(|u^*_n|) \xrightarrow{p} 0. \tag{370}$$

Действительно, первая сходимость в (370) совпадает с первым условием в (A^h_5). Следовательно, для любого $\varepsilon > 0$

$$\begin{aligned}
\mathbb{P}\left(\mathcal{E}_{n,h}(|u^*_n|) > \varepsilon\right) &= \mathbb{P}\Big(\mathcal{E}_{n,h}(|u^*_n|) > \varepsilon,\ I^{-1/2}_{n,h}|J_{n,h}||u^*_n| \ge 1\Big) +\\
&\mathbb{P}\Big(\mathcal{E}_{n,h}(|u^*_n|) > \varepsilon,\ I^{-1/2}_{n,h}|J_{n,h}||u^*_n| < 1\Big) \le \mathbb{P}\Big(I^{-1/2}_{n,h}J_{n,h}|u^*_n|\mathcal{E}_{n,h}(|u^*_n|) > \varepsilon\Big) +\\
&\mathbb{P}\Big(\mathcal{E}_{n,h}(|u^*_n|) > \varepsilon,\ |u^*_n| < I^{1/2}_{n,h}/|J_{n,h}|\Big) \to 0.
\end{aligned} \tag{371}$$

При выводе сходимости в (371) мы учли также, что в силу условий (A_3^h), (A_4^h) и монотонности функции $\mathcal{E}_{n,h}(\cdot)$, во втором слагаемом в правой части (371) под знаком вероятности стоит, начиная с некоторых n, невозможное событие. Тем самым, вторая сходимость в (370) доказана.

Далее, в силу (A_2^h), $\quad |h_i(\theta+u)| \le |h_i(\theta)| + |u|^p H_i \quad$ и

$$\psi_i(\theta+u, X_i) - \psi_i(\theta, X_i) - u\psi_i'(\theta, X_i) = u\int\limits_0^1 \big(\psi_i'(\theta+vu, X_i) - \psi_i'(\theta, X_i)\big)dv,$$

а потому

$$|\rho_{n5}(u)| \le |J_{n,h}|\gamma_n(|u|), \qquad |\rho_{n6}(u)| \le |J_{n,h}||u|\gamma_n(|u|), \tag{372}$$

где

$$\gamma_n(|u|) = \sum_{i=1}^n \big(|h_i(\theta)| + |u|^p H_i\big)\omega_{i,\theta}(|u|, X_i) \quad \text{и} \quad \mathbb{E}\gamma_n(|u|) = \mathcal{E}_{n,h}(|u|), \tag{373}$$

а обозначение $\omega_{i,\theta}(\cdot, X_i)$ введено в (361).

Первые три утверждения леммы следуют теперь из леммы 3.1 при $\eta_n = |u_n^*|$, если положить $\alpha_n(|u|) = \gamma_n(|u|)$ или $\alpha_n(|u|) = I_{n,h}^{-1/2}|J_{n,h}||u|\gamma_n(|u|)$ и учесть оценки (372), соотношение из (373) и (370). Еще раз используя условие (A_2^h), имеем

$$|\rho_{n7}(u)| \le |u|^p \sum_{i=1}^n H_i|\psi_i'(\theta, X_i)|. \tag{374}$$

Последнее утверждение леммы следует из леммы 3.1 при $\eta_n = |u_n^*|$, если в качестве $\alpha_n(|u|)$ рассматривать правую часть в (374) и учесть второе условие в (A_5^h). $\square$.

Лемма 3.8. *Пусть выполнены условия теоремы* 3.8. *Тогда*

$$\rho_{n8} = I_{n,h}^{-1/2}\left(\left(\sum_{i=1}^n h_i^2(\theta_{n,h}^{**})\psi_i^2(\theta_{n,h}^{**}, X_i)\right)^{1/2} - \left(\sum_{i=1}^n h_i^2(\theta)\psi_i^2(\theta, X_i)\right)^{1/2}\right) \xrightarrow{p} 0.$$

Доказательство. Положим $u_{n,h}^{**} = \theta_{n,h}^{**} - \theta$. Нам потребуются соотношения

$$I_{n,h}^{-1/2}J_{n,h}u_{n,h}^{**} = O_p(1), \qquad u_{n,h}^{**} \xrightarrow{p} 0, \qquad \gamma_n(|u_{n,h}^{**}|) \xrightarrow{p} 0. \tag{375}$$

Первое из этих соотношений следует из теоремы 3.7, второе — из условия (A_3^h) и первого соотношения, а третье — из леммы 3.1 при $\eta_n = |u_{n,h}^{**}|$ и $\alpha_n(\cdot) = \gamma_n(\cdot)$, если только еще учесть (373) и условие (A_4^h). Далее, справедлива оценка

$$|\rho_{n8}| \le I_{n,h}^{-1/2}\left(\sum_{i=1}^n \big(h_i(\theta_{n,h}^{**})\psi_i(\theta_{n,h}^{**}, X_i) - h_i(\theta)\psi_i(\theta, X_i)\big)^2\right)^{1/2} \le \sqrt{2}\left(\rho_{n8}^{(1)} + \rho_{n8}^{(2)}\right), \tag{376}$$

где

$$\rho_{n8}^{(1)} = I_{n,h}^{-1/2} \left(\sum h_i^2(\theta_{n,h}^{**}) \left(\psi_i(\theta_{n,h}^{**}, X_i) - \psi_i(\theta, X_i) \right)^2 \right)^{1/2},$$
$$\rho_{n8}^{(2)} = I_{n,h}^{-1/2} \left(\sum \left(h_i(\theta_{n,h}^{**}) - h_i(\theta) \right)^2 \psi_i^2(\theta, X_i) \right)^{1/2}.$$

Оценим $\rho_{n8}^{(1)}$. С учетом предположения (A_2^h) и обозначений из (373) и (361), получаем

$$I_{n,h}^{1/2} \cdot \rho_{n8}^{(1)} = \left(\sum h_i^2(\theta_{n,h}^{**}) \left(\int\limits_{\theta}^{\theta_{n,h}^{**}} \psi_i'(t) dt \right)^2 \right)^{1/2} \leq$$
$$\sqrt{2} \left(\sum_{i=1}^n h_i^2(\theta_{n,h}^{**}) \left(\int\limits_{\theta}^{\theta_{n,h}^{**}} \left(\psi_i'(t, X_i) - \psi_i'(\theta, X_i) \right) dt \right)^2 \right)^{1/2} +$$
$$+ \sqrt{2} \left(\sum_{i=1}^n h_i^2(\theta_{n,h}^{**}) \left(\int\limits_{\theta}^{\theta_{n,h}^{**}} \psi_i'(\theta, X_i) dt \right)^2 \right)^{1/2} \leq$$
$$\leq \sqrt{2} |u_{n,h}^{**}| \left(\sum_{i=1}^n h_i^2(\theta_{n,h}^{**}) \left(\sup_{t: |t-\theta| \leq |u_{n,h}^{**}|} \left| \psi_i'(t, X_i) - \psi_i'(\theta, X_i) \right| \right)^2 \right)^{1/2} +$$
$$+ \sqrt{2} |u_{n,h}^{**}| \left(\sum_{i=1}^n h_i^2(\theta_{n,h}^{**}) \left(\psi_i'(\theta, X_i) \right)^2 \right)^{1/2} \leq \sqrt{2} |J_{n,h}| |u_{n,h}^{**}| \gamma_n(|u_{n,h}^{**}|) +$$
$$+ 2 |u_{n,h}^{**}| \left(\sum_{i=1}^n h_i^2(\theta) \left(\psi_i'(\theta, X_i) \right)^2 \right)^{1/2} + 2 |u_{n,h}^{**}|^{1+p} \left(\sum_{i=1}^n H_i^2 \left(\psi_i'(\theta, X_i) \right)^2 \right)^{1/2}.$$

Используя теперь соотношения (375) и условия из (366) заключаем, что

$$\rho_{n8}^{(1)} \leq \sqrt{2} I_{n,h}^{-1/2} |J_{n,h}| |u_{n,h}^{**}| \cdot \gamma_n(|u_{n,h}^{**}|) +$$
$$+ 2 I_{n,h}^{-1/2} |J_{n,h}| |u_{n,h}^{**}| \left(J_{n,h}^{-2} \sum_{i=1}^n h_i^2(\theta) \left(\psi_i'(\theta, X_i) \right)^2 \right)^{1/2} +$$
$$+ 2 I_{n,h}^{-1/2} |J_{n,h}| |u_{n,h}^{**}|^{1+p} \left(J_{n,h}^{-2} \sum_{i=1}^n H_i^2 \left(\psi_i'(\theta, X_i) \right)^2 \right)^{1/2} \xrightarrow{p} 0.$$

Из предположения (A_2^h), соотношения в (375) и условия из (366) следует, что

$$\rho_{n8}^{(2)} \leq \left(I_{n,h}^{-1/2} |J_{n,h}| |u_{n,h}^{**}| \right)^p \left(J_{n,h}^{-2p} I_{n,h}^{p-1} \sum_{i=1}^n H_i^2 \psi_i^2(\theta, X_i) \right)^{1/2} \xrightarrow{p} 0.$$

Полученные соотношения, вместе с оценкой (376), завершают доказательство. □

Д о к а з а т е л ь с т в о теоремы 3.7. Покажем, что статистика $\theta^{**}_{n,h}$ из (362) определена с вероятностью, стремящейся к 1. В самом деле, $\mathbb{P}(\theta^*_n \in \Theta) \to 1$, поскольку в условиях (A^h) с вероятностью, стремящейся к 1 величина $\mathcal{E}_{n,h}(|\theta^*_n - \theta|)$ конечна (см. (370)). Следовательно, согласно условию (A^h_2), с вероятностью, стремящейся к 1, определены функции $\{h_i(\theta^*_n)\}$, $\{\psi_i(\theta^*_n, X_i)\}$ и $\{\psi'_i(\theta^*_n, X_i)\}$, участвующие в представлении (362) оценки $\theta^{**}_{n,h}$. В силу обозначений (369), справедливо равенство

$$\sum_{i=1}^n h_i(\theta^*_n)\psi'_i(\theta^*_n, X_i) = \sum_{i=1}^n h_i(\theta)\psi'_i(\theta, X_i) + \rho_{n5}(u^*_n) + \rho_{n7}(u^*_n),$$

а потому, с учетом леммы 3.7 и условия (A^h_3), выполнено

$$\mathbb{P}\left(\sum_{i=1}^n h_i(\theta^*_n)\psi'_i(\theta^*_n, X_i) \neq 0\right) \to 1,$$

что завершает доказательство первого утверждения теоремы.

Далее, из определения (362) следует равенство

$$\theta^{**}_{n,h} - \theta = -\frac{h_i(\theta^*_n)\psi_i(\theta^*_n, X_i) - (\theta^*_n - \theta)\sum_{i=1}^n h_i(\theta^*_n)\psi'_i(\theta^*_n, X_i)}{\sum_{i=1}^n h_i(\theta^*_n)\psi'_i(\theta^*_n, X_i)}. \tag{377}$$

Используя представление (377) и обозначения, введенные в (363) и (369), получаем

$$I^{-1/2}_{n,h} J_{n,h}(\theta^{**}_{n,h} - \theta) = -\frac{I^{-1/2}_{n,h}\Big(\sum_{i=1}^n h_i(\theta)\psi_i(\theta, X_i) - u^*_n\rho_{n5}(u^*_n) + \rho_{n6}(u^*_n)\Big) + \rho_n}{J^{-1}_{n,h}\sum_{i=1}^n h_i(\theta)\psi'_i(\theta, X_i) + J^{-1}_{n,h}\big(\rho_{n5}(u^*_n) + \rho_{n7}(u^*_n)\big)}. \tag{378}$$

Чтобы теперь из (378) извлечь асимптотическую нормальность (364) оценки $\theta^{**}_{n,h}$, остается воспользоваться утверждениями леммы 3.7 и условиями (360), (363). $\square$

Д о к а з а т е л ь с т в о теоремы 3.8. Поскольку $\mathbb{P}(\theta^*_n \in \Theta) \to 1$ (см. доказательство теоремы 3.7) и $\mathbb{P}(\theta^{**}_{n,h} \in \Theta) \to 1$ в силу (375), то с вероятностью, стремящейся к 1, определены все величины в правой части равенства в (365). Кроме того, ввиду леммы 3.8 и условия (366)

$$I^{-1/2}_{n,h}\left(\sum_{i=1}^n h^2_i(\theta^{**}_{n,h})\psi^2_i(\theta^{**}_{n,h}, X_i)\right)^{1/2} \equiv I^{-1/2}_{n,h}\left(\sum_{i=1}^n h^2_i(\theta)\psi^2_i(\theta, X_i)\right)^{1/2} + \rho_{n8} \xrightarrow{p} 1, \tag{379}$$

а потому знаменатель в правой части равенства в (365) обращается в нуль с вероятностью, стремящейся к нулю. Таким образом, с учетом определения (365), статистика $Q^*_{n,h}$ определена с вероятностью, стремящейся к 1.

Учитывая представления (377), (378) и обозначение (365), имеем

$$Q^*_{n,h}(\theta^{**}_{n,h}-\theta) = -\frac{I^{-1/2}_{n,h}\Big(\sum_{i=1}^n h_i(\theta)\psi_i(\theta,X_i) - u^*_n\rho_{n5}(u^*_n) + \rho_{n6}(u^*_n)\Big) + \rho_n}{I^{-1/2}_{n,h}\left(\sum_{i=1}^n h_i^2(\theta)\psi_i^2(\theta,X_i)\right)^{1/2} + \rho_{n8}}.$$

Тот факт, что числитель в правой части этого представления по распределению сходится к стандартному нормальному закону, установлен при выводе теоремы 3.7. Вместе с (379) это доказывает (367). □

3.2.3 Некоторые дополнения

Приведем вначале два утверждения, содержащие достаточные условия для выполнения условия (363). В предложениях 3.1 и 3.2 считаем, что выполнено условие (A_1), первые два соотношения в (358) и первое условие в (359). Без ограничения общности будем также считать, что функции $h_i(t)$ продолжены на $\mathbb{R}$ с выполнением условия Гельдера (A_2^h) на всей числовой прямой.

Предложение 3.1. *Пусть оценка θ^*_n имеет структуру*

$$\theta^*_n - \theta = \left(1 + \sum_{i=1}^n v_{ni}(\theta,X_i)\right)^{-1} \sum_{i=1}^n u_{ni}(\theta,X_i), \tag{380}$$

при этом $\mathbb{E}u_{ni}(\theta,X_i) = \mathbb{E}v_{ni}(\theta,X_i) = 0$, $\mathbb{E}u^2_{ni}(\theta,X_i) < \infty$, $\mathbb{E}v^2_{ni}(\theta,X_i) < \infty$ *и*

$$\begin{gathered} d_{nu} + \sum_{i=1}^n \mathbb{E}v^2_{ni}(\theta,X_i) \to 0 \quad \textit{при} \quad d^2_{nu} = \sum_{i=1}^n \mathbb{E}u^2_{ni}(\theta,X_i), \\ I^{-1}_{n,\theta,h} d^{2p}_{nu}\left(\sum_{i=1}^n \left(H_i^2 \mathbb{E}\psi_i^2(\theta,X_i)\right)^{1/(2-p)}\right)^{2-p} \to 0. \end{gathered} \tag{381}$$

Тогда имеет место (363).

Отметим, что условию (380) удовлетворяют, например, оценки θ^*_n из раздела 2.1 и приводимого далее замечания 3.13. Для выполнения центрального условия (381) достаточно, чтобы

$$n^{(1/p-1)/2} d_{nu} \to 0, \quad I^{-1}_{n,\theta,h} = O\big(1/n\big), \quad \sup_{n,i}\{H_i,\ \mathbb{E}\psi_i^2(\theta,X_i)\} < \infty. \tag{382}$$

В частности, первое условие в (382) выполнено в следующих трех случаях:

$$\sqrt{n}d_{nu} = O(1) \text{ и } p > 1/2, \qquad n^{1/4}d_{nu} \to 0 \text{ и } p \ge 2/3, \qquad d_{nu} \to 0 \text{ и } p = 1.$$

В случае, когда θ_n^* имеет дробно-линейную структуру, достаточные условия для (363) можно получать и в терминах сближения θ_n^* и θ по вероятности (т.е. в терминах предельного соотношения $n^{(1/p-1)/2}(\theta_n^*-\theta)\xrightarrow{p}0$; см. подробности в [336]).

Предложение 3.2. *Пусть* $\mathbb{E}|\theta_n^*|^2<\infty$ *и*

$$I_{n,\theta,h}^{-1}\left(\mathbb{E}|\theta_n^*-\theta|^2\right)^p\sum_{i=1}^n H_i^2\mathbb{E}\psi_i^2(\theta,X_i)\to 0, \tag{383}$$

$$I_{n,\theta,h}^{-1/2}\sum_{i=1}^n H_i\left(\mathbb{E}|\theta_n^*-\mathbb{E}_{\neq i}\theta_n^*|^2\right)^{p/2}\left(\mathbb{E}\psi_i^2(\theta,X_i)\right)^{1/2}\to 0, \tag{384}$$

где $\mathbb{E}_{\neq i}$ *есть условное математическое ожидание, взятое при условии, что при всех* $j\neq i$ *фиксированы значения* X_j, $j=1,\ldots,n$. *Тогда имеет место* (363).

К вопросу о проверке технического условия (384) стоит отметить, что в простейших регулярных случаях $\mathbb{E}|\theta_n^*-\mathbb{E}_{\neq i}\theta_n^*|^2$ есть величина порядка $1/n^2$.

В следующем утверждении приведено некоторое свойство устойчивости оценок с асимптотической дисперсией $I_{n,\theta,h}J_{n,\theta,h}^{-2}$ как функционалов, зависящих от набора функций $\{h_i(\cdot)\}$: чем лучше $h_i(\cdot)$ приближают некоторые оптимальные функции $h_i^o(\cdot)$, тем меньше асимптотическая дисперсия $I_{n,\theta,h}J_{n,\theta,h}^{-2}$ отличается от оптимальной (аналог этого утверждения для асимптотических дисперсий специального вида приведен, например, в [342]).

Справедливо

Предложение 3.3. *Пусть* $J_{n,\theta,h}\equiv\sum_{i=1}^n h_i(\theta)\mathbb{E}\psi_i'(\theta,X_i)\neq 0$ *для всех достаточно больших* n. *В этом случае определена величина* $I_{n,\theta,h}J_{n,\theta,h}^{-2}$ *и*

$$\frac{I_{n,\theta,h}}{J_{n,\theta,h}^2}\geq\frac{I_{n,\theta,h^o}}{J_{n,\theta,h^o}^2}\equiv\left(\sum_{i=1}^n\frac{\left(\mathbb{E}\psi_i'(\theta,X_i)\right)^2}{\mathbb{E}\psi_i^2(\theta,X_i)}\right)^{-1}\quad \text{\emph{при}}\quad h_i^o(\theta)=\frac{\mathbb{E}\psi_i'(\theta,X_i)}{\mathbb{E}\psi_i^2(\theta,X_i)}. \tag{385}$$

Если дополнительно $\mathrm{sign}h_i(\theta)=\mathrm{sign}h_i^o(\theta)$, *то справедливо неравенство:*

$$1\leq\frac{I_{n,\theta,h}J_{n,\theta,h}^{-2}}{I_{n,\theta,h^o}J_{n,\theta,h^o}^{-2}}\leq 1+\frac{(\sqrt{H/h}-1)^2}{2\sqrt{H/h}}\leq\sqrt{\frac{H}{h}},\quad \text{\emph{где}} \tag{386}$$

$$h=\min_{\{i:\,\mathbb{E}\psi_i'(\theta,X_i)\neq 0\}}h_i(\theta)/h_i^o(\theta)>0,\quad H=\max_{\{i:\,\mathbb{E}\psi_i'(\theta,X_i)\neq 0\}}h_i(\theta)/h_i^o(\theta)<\infty. \tag{387}$$

Таким образом, если оптимальные функции $h_i^o(\cdot)$ известны (см., например, раздел 3.4), то в качестве $h_i(\cdot)$ разумно выбрать оптимальные $h_i^o(\cdot)$. Но если, скажем, оптимальные функции $h_i^o(\cdot)$ не удовлетворяют условиям теоремы 3.7, то можно использовать функции $\widetilde{h}_i^o(\cdot)$, полученные в результате некоторого сглаживания функций $h_i^o(\cdot)$. Кроме того, оптимальные функции $h_i^o(\cdot)$ могут быть неизвестны. Например, если в модели (463) $\mathbb{E}\varepsilon_i^2=\sigma_i^2$ при всех i

и значения $\{\sigma_i\}$ неизвестны, то при $\psi_i(\theta, X_i) = X_i - f_i(\theta)$ оптимальные функции $h_i^o(\cdot)$ с точностью до константы определяются равенством $h_i^o(\cdot) = f_i'(\cdot)/\sigma_i^2$. В подобных случаях можно использовать некоторые функции $\widetilde{h}_i^o(\cdot)$, которые должны, по-возможности, «не очень сильно отличаться» от $h_i^o(\cdot)$. При этом соотношение (386) гарантирует, что асимптотическая дисперсия $I_{n,\theta,\widetilde{h}^o} J^{-2}_{n,\theta,\widetilde{h}^o}$ будет «не очень сильно отличаться» от оптимальной асимптотической дисперсии $I_{n,\theta,h^o} J^{-2}_{n,\theta,h^o}$. Приведенные здесь соображения по выбору функций $h_i(\cdot)$ в полной мере относятся к выбору этих функций при построении M-оценок в случае (19) и взвешенных M-оценок.

З а м е ч а н и е 3.13. При широких ограничениях взвешенные M-оценки, определяемые двумя уравнениями

$$\sum_{i=1}^{n} h_i(\theta_n^*)\psi_i(t, X_i) = 0 \quad \text{и} \quad \sum_{i=1}^{n} h_i(\theta_n^*) g_i^{-1}(\theta_n^*) g_i(t) \psi_i(t, X_i) = 0, \tag{388}$$

асимптотически имеют одну и ту же точность, но при удачном выборе функций $g_i(t)$ второе уравнение в (388) может быть предпочтительнее, нежели первое. Более того, одношаговые взвешенные M-оценки можно строить исходя и из второго уравнения в (388), а потому за счет выбора функций $g_i(t)$ возникает множество различных явных оценок, эквивалентных в смысле асимптотической точности.

Например, в [341] для модели $X_i = 1/(1 + z_i\theta) + \varepsilon_i$ с условием $\mathbb{E}\varepsilon_i^2 = \sigma^2/w_i(\theta)$ построена в некотором смысле оптимальная оценка $\widetilde{\theta}_n^{**} = \widetilde{\theta}_n^{**}(\theta_n^*)$. В [341] при построении этой оценки использовались эвристические соображения, существенно использующие дробно-линейную структуру регрессионной функции и не связанные с одношаговым оцениванием. Как согласуется статистика $\widetilde{\theta}_n^{**}$ из [341] с предлагаемой здесь методологией? Оказывается, эта оценка есть частный случай предлагаемых здесь конструкций: $\widetilde{\theta}_n^{**}$ является явным решением второго уравнения в (388) при $g_i(t) = 1 + z_i t$ и $h_i(t) = w_i(t) f_i'(t)$ и $\psi_i(t, X_i) = X_i - f_i(t)$ при $f_i(t) = 1/(1 + z_i\theta)$, т.е. является взвешенной M-оценкой, имеющей асимптотически точность оценки квазиправдоподобия. Кроме того, $\widetilde{\theta}_n^{**}$ из [341] совпадает с одношаговой взвешенной M-оценкой, построенной по указанному уравнению. □

3.2.4 Вспомогательная теорема

В этом разделе мы приведем и докажем некоторое вспомогательное утверждение, которое может иметь и самостоятельный интерес. Нам же это утверждение понадобится для вывода предложения 3.1. Исключительно в этом разделе для произвольной случайной величины ξ будем полагать $\|\xi\| \equiv \left(\mathbb{E}\xi^2\right)^{1/2}$, а для произвольной последовательности $\{a_{ni}\}$ использовать обозначение

$$\mathbb{S}_m(a_{n\bullet}) = \left(\sum_{i=1}^{n} a_{ni}^m\right)^{1/m} \quad \text{при} \quad m > 0. \tag{389}$$

Перечислим теперь условия, которые нам потребуются в этом разделе.

($\mathbf{H}_1$). *Пусть $I_n = [\theta - \kappa_n, \theta + \kappa_n]$ — некоторый интервал, где $\kappa_n > 0$ и $\theta \in (-\infty, \infty)$ — действительные числа. При некотором фиксированном n и всех $i = 1, \ldots, n$ заданы случайные функции $\phi_{ni}(\cdot)$ и случайные величины $\overline{\phi}_{ni}$ такие, что $\|\overline{\phi}_{ni}\| < \infty$ и при некотором $q \in (0, 1]$ выполнено*

$$|\phi_{ni}(t_2) - \phi_{ni}(t_1)| \leq \overline{\phi}_{ni}|t_2 - t_1|^q \quad \text{при всех} \quad t_1, t_2 \in I_n. \tag{390}$$

($\mathbf{H}_2$). *При всех $i = 1, \ldots, n$ заданы случайные величины u_{ni} и v_{ni} с нулевыми средними такие, что случайные векторы $W_{ni} = \big(u_{ni}, v_{ni}, \overline{\phi}_{ni}, \phi_{ni}(\cdot)\big)$, $i = 1, \ldots, n$, независимы в совокупности и справедливо представление*

$$\theta_n^* - \theta = \Big(1 + \sum_{i=1}^{n} v_{ni}\Big)^{-1} \sum_{i=1}^{n} u_{ni}. \tag{391}$$

Положим

$$d_{nu}^2 = \mathbb{D}\sum_{i=1}^{n} u_{ni}^2, \qquad d_{nv}^2 = \mathbb{D}\sum_{i=1}^{n} v_{ni}^2, \tag{392}$$

$$\mathbb{C}_n = \sum_{i=1}^{n} \|\overline{\phi}_{ni}\|\big(\|u_{ni}\|^q + d_{nu}^q\|v_{ni}\|^q\big) + d_{nu}^q \mathbb{S}_2\big(\|\overline{\phi}_{n\bullet}\|\big), \tag{393}$$

$$\breve{\phi}_{ni}(t) = (\phi_{ni}(t) - \phi_{ni}(\theta)) - \mathbb{E}(\phi_{ni}(t) - \phi_{ni}(\theta)), \quad t \in I_n. \tag{394}$$

Последнее определение корректно, поскольку $\|\overline{\phi}_{ni}\| < \infty$, а потому ввиду (390) у случайной величины $\phi_{ni}(t) - \phi_{ni}(\theta)$ при $t \in I_n$ существует математическое ожидание.

Наряду с условиями (H_1)-(H_2), нам также потребуется более простое ограничение.

($\mathbf{H}_0$). *При некотором n случайные векторы (u_{ni}, v_{ni}), $i = 1, \ldots, n$, независимы в совокупности, имеют нулевые средние и при некотором действительном $\theta \in (-\infty, \infty)$ справедливо представление* (391).

Сформулируем теперь основное утверждение настоящего подраздела.

Теорема 3.9. *Пусть выполнено условие (H_0). В этом случае существует такая случайная величина $\widetilde{\theta}_n$, что одновременно верны следующие неравенства:*

$$\mathbb{P}(\widetilde{\theta}_n \neq \theta_n^*) \leq 4d_{nu}^2/\kappa_n^2 + 4d_{nv}^2, \qquad |\widetilde{\theta}_n - \theta| \leq \kappa_n, \qquad \mathbb{E}|\widetilde{\theta}_n - \theta|^{2q} \leq 4d_{nu}^{2q} \tag{395}$$

при некотором $\kappa_n > 0$. А если, дополнительно, справедливы условия $(H_1) - (H_2)$, то

$$\mathbb{E}\Big|\sum_{i=1}^{n} \breve{\phi}_{ni}(\widetilde{\theta}_n)\Big| \leq 3 \cdot 2^{2q}\mathbb{C}_n \leq 3 \cdot 2^{2q+1} d_{nu}^q (1 + d_{nv}^q) \mathbb{S}_{2/(2-q)}\big(\|\overline{\phi}_{n\bullet}\|\big). \tag{396}$$

З а м е ч а н и е 3.14. В конце раздела будет приведен пример функций $\{\phi_{ni}(\cdot)\}$, удовлетворяющих всем условиям теоремы 3.9 и таких, что

$$0 < \mathbb{E}\left|\sum_{i=1}^{n} \breve{\phi}_{ni}(\widetilde{\theta}_n)\right| \le 3 \cdot 2^{2q+1}(1 + d_{nv}^q) d_{nu}^q \mathbb{S}_{2/(2-q)}(\|\overline{\phi}_{n\bullet}\|) \le 3 \cdot 2^4 \mathbb{E} \sum_{i=1}^{n} \breve{\phi}_{ni}(\widetilde{\theta}_n) < \infty. \tag{397}$$

Таким образом, неравенство (396) неулучшаемо с точностью до константы. □

Остальная часть параграфа посвящена доказательству теоремы 3.9 и утверждения из замечания 3.14. Условимся далее в этом разделе опускать у всех используемых величин дополнительный индекс n. Отметим, что если дисперсии каких-то величин u_i и v_i не конечны, то нужно лишь доказать центральное неравенство в (396), поэтому всюду далее предполагаем конечность вторых моментов величин u_i и v_i, $i = 1, \ldots, n$.

Нам потребуется несколько вспомогательных лемм. Обозначим

$$f_u(x) = \begin{cases} x, & \text{если } |x| \le \kappa/2, \\ (\kappa/2)\mathrm{sign} x, & \text{если } |x| \ge \kappa/2, \end{cases} \qquad f_v(x) = \begin{cases} 1/2, & \text{если } x \le -1/2, \\ 1 + x, & \text{если } x \ge -1/2. \end{cases} \tag{398}$$

Поскольку, ввиду (391),

$$\theta^* = \theta + u/(1+v) \qquad \text{при} \qquad u = \sum_{i=1}^{n} u_i, \quad v = \sum_{i=1}^{n} v_i, \tag{399}$$

то «срезки» $\widetilde{\theta}$ величины θ^* введем, полагая

$$\begin{gathered} \widetilde{\theta} = \theta + f_u(u)/f_v(v), \qquad \widetilde{\theta}^{(i)} = \theta + f_u(u^{(i)})/f_v(v - v_i), \\ \widetilde{\theta}^{(ij)} = \widetilde{\theta}^{(ji)} = \theta + f_u(u^{(ij)})/f_v(v - v_i - v_j) \end{gathered} \tag{400}$$

при $u^{(i)} = u - u_i$ и $u^{(ij)} = u^{(ji)} = u - u_i - u_j$.

Лемма 3.9. *Справедливы первые два неравенства в* (395) *и при всех i и $j \ne i$*

$$|\widetilde{\theta} - \widetilde{\theta}^{(i)}| \le \tau_i = 2|u_i| + 4|u^{(i)}||v_i|, \qquad |\widetilde{\theta}^{(j)} - \widetilde{\theta}^{(ji)}| \le \tau_{ji} = 2|u_i| + 4|u^{(ji)}||v_i|. \tag{401}$$

Д о к а з а т е л ь с т в о. Ввиду условия (H_0) и обозначений из (392) и (399)

$$\mathbb{E}u = \mathbb{E}v = 0, \qquad \mathbb{E}u^2 = d_u^2, \qquad \mathbb{E}v^2 = d_v^2. \tag{402}$$

А потому, с учетом определений (399), (400) и неравенства Чебышева

$$\mathbb{P}(\widetilde{\theta} \ne \theta^*) \le \mathbb{P}(|u| \ge \kappa/2) + \mathbb{P}(v \le -1/2) \le 4\mathbb{E}u^2/\kappa^2 + 4\mathbb{E}v^2.$$

Тем самым, мы доказали первое неравенство из (395). Из определения (398) получаем, что $|f_u(u)| \le \kappa/2$ и $|f_v(v)| \ge 1/2$. Следовательно, с учетом (400), имеем $|\widetilde{\theta}-\theta| = |f_u(u)|/|f_v(v)| \le \kappa$, то есть выполнено второе неравенство из (395). Используя определение (400), получаем

$$\widetilde{\theta} - \widetilde{\theta}^{(i)} = \frac{f_u(u)}{f_v(v)} - \frac{f_u(u^{(i)})}{f_v(v-v_i)} = \frac{f_u(u) - f_u(u^{(i)})}{f_v(v)} - \frac{f_u(u^{(i)})\big(f_v(v-v_i) - f_v(v)\big)}{f_v(v)f_v(v-v_i)}. \tag{403}$$

Ясно, что $|f_u(u) - f_u(u^{(i)})| \le |u_i|, \quad |f_v(v) - f_v(v-v_i)| \le |v_i|,$

$$|f_v(x)| \ge 1/2, \qquad |f_u(x)| \le |x|, \tag{404}$$

Подставляя эти соотношения в (403), получаем первую оценку из (401). Для доказательства второго неравенства в (401) нужно еще раз повторить эти рассуждения при $u_j = v_j = 0$. $\square$

Лемма 3.10. *При всех i и $j \ne i$ справедливы следующие соотношения*

$$\mathbb{E}|\widetilde{\theta}^{(i)} - \theta|^{2q} \le 4^q d_u^{2q}, \quad \mathbb{E}\tau_i^{2q} \le \nu_i^2 = \big(2^q\|u_i\|^q + 4^q d_u^q \|v_i\|^q\big)^2, \quad \mathbb{E}\tau_{ji}^{2q} \le \nu_i^2. \tag{405}$$

Кроме того, верно последнее неравенство в (395).

Доказательство. Из (400) и (404), имеем

$$|\widetilde{\theta} - \theta| = \left|\frac{f_u(u)}{f_v(v)}\right| \le 2|u|, \qquad |\widetilde{\theta}^{(i)} - \theta| = \frac{|f_u(u^{(i)})|}{|f_v(v-v_i)|} \le 2|u^{(i)}|.$$

Следовательно, с учетом (402), $\mathbb{E}|\widetilde{\theta}-\theta|^2 \le 4\mathbb{E}|u|^2 = 4d_u^2$ и $\mathbb{E}|\widetilde{\theta}^{(i)}-\theta|^2 \le 4\mathbb{E}\big|u^{(i)}\big|^2 \le 4\mathbb{E}|u|^2 = 4d_u^2$, т. е. выполнены последнее неравенство в (395) и первое соотношение в (405) при $q = 1$. Кроме того, в силу независимости $u^{(i)}$ и v_i, а также очевидной оценки $\|u^{(i)}\| \le \|u\| = d_u$, получаем

$$\|\tau_i\| = \big\|2|u_i| + 4|u^{(i)}||v_i|\big\| \le 2\|u_i\| + 4d_u\|v_i\| \tag{406}$$

Это соотношение доказывает второе неравенство в (405) при $q = 1$. Чтобы получить третье неравенство в (405) при $q = 1$, достаточно в (406) положить $u_j = v_j = 0$. Тем самым, доказаны все утверждения леммы при $q = 1$. Из неравенства $\mathbb{E}\xi^{2q} \le (\mathbb{E}\xi^2)^q$ следует теперь справедливость всех утверждений леммы и при $0 < q \le 1$. $\square$

Далее, не ограничивая общности, считаем, что $\phi_i(\theta) = 0$. Действительно, пусть $\phi_{oi}(t) = \phi_i(t) - \phi_i(\theta)$. Тогда $\breve{\phi}_{oi}(t) = \breve{\phi}_i(t)$, при этом $\phi_{oi}(\theta) = 0$. Положим

$$\delta_i = \phi_i(\widetilde{\theta}) - \phi_i(\widetilde{\theta}^{(i)}), \qquad \delta_{ij} = \phi_i(\widetilde{\theta}^{(i)}) - \phi_i(\widetilde{\theta}^{(ij)}). \tag{407}$$

Лемма 3.11. *При всех i и $j \neq i$ верны следующие неравенства*

$$\mathbb{E}\big(\phi_i(\widetilde{\theta}^{(i)})\big)^2 \le 4^q d_u^{2q}\|\overline{\phi}_i\|^2, \qquad \mathbb{E}\delta_{ji}^2 \le \nu_i^2\|\overline{\phi}_j\|, \qquad \mathbb{E}|\delta_i| \le \|\overline{\phi}_i\|\nu_i.$$

Д о к а з а т е л ь с т в о. Используя определения из (407) и условие (H_1), имеем

$$\big|\phi_i(\widetilde{\theta}^{(i)})\big| = \big|\phi_i(\widetilde{\theta}^{(i)}) - \phi_i(\theta)\big| \le \overline{\phi}_i|\widetilde{\theta}^{(i)} - \theta|^q, \quad |\delta_{ji}| \le \overline{\phi}_i\tau_{ji}^q, \quad |\delta_i| \le \overline{\phi}_i|\widetilde{\theta} - \widetilde{\theta}^{(i)}|^q. \tag{408}$$

При выводе первого соотношения нужно учесть, что $\phi_i(\theta) = 0$, а при выводе второго неравенства в (408) воспользоваться еще второй оценкой в (401). Учитывая независимость величин $\widetilde{\theta}^{(i)}$ и $\overline{\phi}_i$, из первой оценки в (408) находим

$$\mathbb{E}\big(\phi_i(\widetilde{\theta}^{(i)})\big)^2 \le \mathbb{E}\big(\overline{\phi}_i^2|\widetilde{\theta}^{(i)} - \theta|^{2q}\big) = \mathbb{E}|\widetilde{\theta}^{(i)} - \theta|^{2q}\mathbb{E}\overline{\phi}_i^2 \le 4^q d_u^{2q}\|\overline{\phi}_i\|^2. \tag{409}$$

Выше при выводе заключительного неравенства в (409) использовалось первое утверждение леммы 3.10. Из (409) вытекает первое утверждение леммы 3.11. Второе неравенство в (408), независимость величин τ_{ji} и $\overline{\phi}_j$ и третья оценка в (405) влекут цепочку соотношений $\mathbb{E}\delta_{ji}^2 \le \mathbb{E}(\overline{\phi}_j^2\tau_{ji}^{2q}) = \mathbb{E}\overline{\phi}_j^2 \cdot \mathbb{E}\tau_{ji}^{2q} \le \|\overline{\phi}_i\|^2 \cdot \nu_i^2$. Из третьей оценки в (408), первого утверждения в (401) и второго утверждения леммы 3.10 получаем, что

$$\mathbb{E}|\delta_i| \le \mathbb{E}\left(\overline{\phi}_i|\widetilde{\theta} - \widetilde{\theta}^{(i)}|^q\right) \le \|\overline{\phi}_i\|\mathbb{E}\left(|\widetilde{\theta} - \widetilde{\theta}^{(i)}|^{2q}\right)^{1/2} \le \|\overline{\phi}_i\|\left(\mathbb{E}\tau_i^{2q}\right)^{1/2} \le \|\overline{\phi}_i\|\nu_i. \qquad \square$$

Лемма 3.12. *При всех i имеет место оценка $\mathbb{E}|\widetilde{\delta}_i| \le 2\nu_i\|\overline{\phi}_i\|$ при $\widetilde{\delta}_i = \breve{\phi}_i(\widetilde{\theta}) - \breve{\phi}_i(\widetilde{\theta}^{(i)})$.*

Д о к а з а т е л ь с т в о. Определения (394) (407) влекут равенство $\widetilde{\delta}_i = \delta_i - \mathbb{E}\delta_i$. Значит, $\mathbb{E}|\widetilde{\delta}_i| \le \mathbb{E}|\delta_i| + |\mathbb{E}\delta_i| \le 2\mathbb{E}|\delta_i|$. Используя теперь последнюю оценку из леммы 3.11, получаем утверждение леммы. $\square$

Введем следующие обозначения

$$\begin{gathered}\Delta_1 = \sum_{i=1}^n \big(\breve{\phi}_i(\widetilde{\theta}) - \breve{\phi}_i(\widetilde{\theta}^{(i)})\big) \equiv \sum_{i=1}^n \widetilde{\delta}_i,\\ \Delta_2 = \sum_{i=1}^n \breve{\phi}_i^2(\widetilde{\theta}^{(i)}), \qquad \Delta_3 = \sum_{i=1}^n \sum_{j\neq i} \breve{\phi}_i(\widetilde{\theta}^{(i)})\breve{\phi}_j(\widetilde{\theta}^{(j)}).\end{gathered} \tag{410}$$

Из определений (394) и (410) имеем

$$\sum_{i=1}^n \breve{\phi}_i(\widetilde{\theta}) = \Delta_1 + \sum_{i=1}^n \breve{\phi}_i(\widetilde{\theta}^{(i)}) \quad \text{и} \quad \left(\sum_{i=1}^n \breve{\phi}_i(\widetilde{\theta}^{(i)})\right)^2 = \Delta_2 + \Delta_3.$$

А потому $\mathbb{E}\left|\sum_{i=1}^n \breve{\phi}_i(\widetilde{\theta})\right| \le \mathbb{E}|\Delta_1| + \mathbb{E}(\Delta_2 + \Delta_3)^{1/2} \le \mathbb{E}|\Delta_1| + (\mathbb{E}\Delta_2 + \mathbb{E}\Delta_3)^{1/2}$. Значит,

$$\mathbb{E}\left|\sum_{i=1}^{n}\breve{\phi}_i(\widetilde{\theta})\right| \le \mathbb{E}\bigl|\Delta_1\bigr| + \bigl(\mathbb{E}\Delta_2\bigr)^{1/2} + \bigl|\mathbb{E}\Delta_3\bigr|^{1/2}. \tag{411}$$

Таким образом, доказательство соотношения (396) свелось к оцениванию трех слагаемых в правой части неравенства (411). Для оценивания $|\mathbb{E}\Delta_3|$ нам потребуется

Лемма 3.13. *Для любых i и $j \ne i$*

$$\mathbb{E}\breve{\phi}_i(\widetilde{\theta}^{(i)})\breve{\phi}_j(\widetilde{\theta}^{(j)}) = \mathbb{E}\widetilde{\delta}_{ij}\widetilde{\delta}_{ji} \qquad \textit{при} \qquad \widetilde{\delta}_{ij} = \breve{\phi}_i(\widetilde{\theta}^{(i)}) - \breve{\phi}_i(\widetilde{\theta}^{(ij)}). \tag{412}$$

Кроме того,

$$\mathbb{E}\breve{\phi}_i^2(\widetilde{\theta}^{(i)}) \le \mathbb{E}\phi_i^2(\widetilde{\theta}^{(i)}), \qquad \mathbb{E}\widetilde{\delta}_{ij}^2 \le \mathbb{E}\delta_{ij}^2, \qquad \mathbb{E}|\widetilde{\delta}_{ij}\widetilde{\delta}_{ji}| \le \bigl(\mathbb{E}\delta_{ij}^2\mathbb{E}\delta_{ji}^2\bigr)^{1/2}. \tag{413}$$

Д о к а з а т е л ь с т в о. Условимся через $\mathbb{E}_i Z$ обозначать условное математическое ожидание, взятое при условии, что при всех $j \ne i$ фиксированы значения независимых случайных векторов W_j, $j = 1, \dots, n$, определенных в условии (B_2). Нетрудно заметить, что в этом случае из определения (394) вытекает равенство

$$\breve{\phi}_i(Z) = \phi_i(Z) - \mathbb{E}_i\phi_i(Z) \quad \text{при} \quad Z = \widetilde{\theta}^{(i)} \quad \text{и} \quad Z = \widetilde{\theta}^{(ij)}, \tag{414}$$

поскольку во всех перечисленных в (414) вариантах случайная величина Z не зависит от случайного вектора W_i, что существенно для справедливости (414). Таким образом, из определений (407) и равенств (414) получаем, что $0 = \mathbb{E}_i\breve{\phi}_i(\widetilde{\theta}^{(i)}) = \mathbb{E}_i\breve{\phi}_i(\widetilde{\theta}^{(ij)}) = \mathbb{E}_i\widetilde{\delta}_{ij}$ и $0 = \mathbb{E}_j\breve{\phi}_j(\widetilde{\theta}^{(j)})$ при всех i и $j \ne i$. Следовательно,

$$\begin{aligned}\mathbb{E}\breve{\phi}_i(\widetilde{\theta}^{(ij)})\breve{\phi}_j(\widetilde{\theta}^{(j)}) = \mathbb{E}\mathbb{E}_j\breve{\phi}_i(\widetilde{\theta}^{(ij)})\breve{\phi}_j(\widetilde{\theta}^{(j)}) = \mathbb{E}\breve{\phi}_i(\widetilde{\theta}^{(ij)})\mathbb{E}_j\breve{\phi}_j(\widetilde{\theta}^{(j)}) = 0,\\ \mathbb{E}\widetilde{\delta}_{ij}\breve{\phi}_j(\widetilde{\theta}^{(ji)}) = \mathbb{E}\mathbb{E}_i\breve{\phi}_j(\widetilde{\theta}^{(ji)})\widetilde{\delta}_{ij} = \mathbb{E}\breve{\phi}_j(\widetilde{\theta}^{(ji)})\mathbb{E}_i\widetilde{\delta}_{ij} = 0.\end{aligned} \tag{415}$$

Далее, из определения (412) величины $\widetilde{\delta}_{ij}$ находим

$$\widetilde{\delta}_{ij}\widetilde{\delta}_{ji} = \widetilde{\delta}_{ij}\breve{\phi}_j(\widetilde{\theta}^{(j)}) - \widetilde{\delta}_{ij}\breve{\phi}_j(\widetilde{\theta}^{(ji)}) = \breve{\phi}_i(\widetilde{\theta}^{(i)})\breve{\phi}_j(\widetilde{\theta}^{(j)}) - \breve{\phi}_i(\widetilde{\theta}^{(ij)})\breve{\phi}_j(\widetilde{\theta}^{(j)}) - \widetilde{\delta}_{ij}\breve{\phi}_j(\widetilde{\theta}^{(ji)}).$$

Если теперь мы возьмем математические ожидания от обеих частей этого тождества и воспользуемся равенствами (415), то мы получим (412).

Докажем неравенства (413). Используя определения (394), (407) и (412) получаем, что

$$\begin{aligned}\mathbb{E}\breve{\phi}_i^2(\widetilde{\theta}^{(i)}) = \mathbb{E}\mathbb{E}_i\bigl(\phi_i(\widetilde{\theta}^{(i)}) - \mathbb{E}_i\phi_i(\widetilde{\theta}^{(i)})\bigr)^2 \le \mathbb{E}\mathbb{E}_i\phi_i^2(\widetilde{\theta}^{(i)}) = \mathbb{E}\phi_i^2(\widetilde{\theta}^{(i)}),\\ \mathbb{E}\widetilde{\delta}_{ij}^2 = \mathbb{E}\mathbb{E}_i(\delta_{ij} - \mathbb{E}_i\delta_{ij})^2 \le \mathbb{E}\mathbb{E}_i\delta_{ij}^2 = \mathbb{E}\delta_{ij}^2.\end{aligned} \tag{416}$$

При выводе (416) учтено, что дисперсия любой случайной величины не больше ее второго момента. С учетом (416), имеем $\mathbb{E}\big|\widetilde{\delta}_{ij}\widetilde{\delta}_{ji}\big| \le \big(\mathbb{E}\widetilde{\delta}_{ij}^2\big)^{1/2}\big(\mathbb{E}\widetilde{\delta}_{ji}^2\big)^{1/2} \le \big(\mathbb{E}\delta_{ij}^2\mathbb{E}\delta_{ji}^2\big)^{1/2}$. Тем самым мы вывели и третье утверждение в (413). □

Оценим величины, определенные в (410). Из лемм 3.11, 3.12 и 3.13 получаем, что

$$\mathbb{E}\big|\Delta_1\big| \le \sum_{i=1}^n \mathbb{E}\widetilde{\delta}_i \le 2\sum \nu_i\|\overline{\phi}_i\|, \qquad \mathbb{E}\Delta_2 \le \sum_{i=1}^n \mathbb{E}\phi_i^2(\widetilde{\theta}^{(i)}) \le 2^{2q}d_u^{2q}\sum_{i=1}^n\|\overline{\phi}_i\|^2, \tag{417}$$
$$\big|\mathbb{E}\Delta_3\big| \le \sum_{i=1}^n\sum_{j\ne i}\big|\mathbb{E}\breve{\phi}_i(\widetilde{\theta}^{(i)})\breve{\phi}_j(\widetilde{\theta}^{(j)})\big| \le \sum_{i=1}^n\sum_{j\ne i}\big(\|\overline{\phi}_i\|\nu_i\|\overline{\phi}_j\|\nu_j\big) \le \Big(\sum_{i=1}^n\|\overline{\phi}_i\|\nu_i\Big)^2.$$

При выводе последнего неравенства учтено, что из (412), (413) и леммы 3.11 следуют соотношения $\big|\mathbb{E}\breve{\phi}_i(\widetilde{\theta}^{(i)})\breve{\phi}_j(\widetilde{\theta}^{(j)})\big| = \big|\mathbb{E}\widetilde{\delta}_{ij}\widetilde{\delta}_{ji}\big| \le \mathbb{E}\big|\widetilde{\delta}_{ij}\widetilde{\delta}_{ji}\big| \le \big(\mathbb{E}\delta_{ij}^2\mathbb{E}\delta_{ji}^2\big)^{1/2} \le \|\overline{\phi}_i\|\nu_i\|\overline{\phi}_j\|\nu_j$. Теперь из (411) и (417), с учетом определений (405) и (393), находим, что

$$\mathbb{E}\left|\sum_{i=1}^n\breve{\phi}_i(\widetilde{\theta})\right| \le 3\sum_{i=1}^n\|\overline{\phi}_i\|\nu_i + 2^q d_u^q\mathbb{S}_2\big(\|\overline{\phi}_\bullet\|\big) \le 3\cdot 2^{2q}\mathbb{C}. \tag{418}$$

Тем самым мы вывели первое неравенство в (396). А чтобы получить теперь вторую оценку в (396), воспользуемся неравенством Гельдера:

$$\sum_{i=1}^n\|\overline{\phi}_i\|\cdot\|u_i\|^q \le \left(\sum_{i=1}^n\|\overline{\phi}_i\|^{2/(2-q)}\right)^{(2-q)/2}\left(\sum_{i=1}^n\|u_i\|^2\right)^{q/2} = d_u^q\mathbb{S}_{2/(2-q)}\big(\|\overline{\phi}_\bullet\|\big),$$
$$\sum_{i=1}^n\|\overline{\phi}_i\|\cdot\|v_i\|^q \le \left(\sum_{i=1}^n\|\overline{\phi}_i\|^{2/(2-q)}\right)^{(2-q)/2}\left(\sum_{i=1}^n\|v_i\|^2\right)^{q/2} = d_v^q\mathbb{S}_{2/(2-q)}\big(\|\overline{\phi}_\bullet\|\big).$$

Из этих оценок и определений (389), (393) следует, что $\mathbb{C} \le d_u^q(1+d_v^q)\mathbb{S}_{2/(2-q)}\big(\|\overline{\phi}_\bullet\|\big)+d_u^q\mathbb{S}_2\big(\|\overline{\phi}_\bullet\|\big)$. Для завершения вывода (396) учтем еще, что $\mathbb{S}_2\big(\|\overline{\phi}_\bullet\|\big) \le \mathbb{S}_{2/(2-q)}\big(\|\overline{\phi}_\bullet\|\big)$ ввиду монотонности по m норм $\mathbb{S}_m(\cdot)$. Неравенства из (395) установлены в леммах 3.9 и 3.10, тем самым доказательство теоремы 3.9 завершено.

П р и м е р 3.3. Пусть $\mathbb{P}(\xi = 1) = \mathbb{P}(\xi = -1) = 1/2$, $\sigma = \kappa/2$, $0 < \kappa < 1$, $K > 0$, $u_1 = \sigma\xi$, $v_1 = \sigma\xi/2$, $\gamma_1(t) = K|t-1|^q\mathrm{sign}(t-1)$, $\phi_1(t) = u_1\gamma_1(t)$. Кроме того, $u_i = v_i = 0 = \phi_i(\cdot)$ при $i \ge 2$. Покажем, что в этом случае при $\theta = 1$ имеет место (397), т.е. утверждение (396) теоремы 3.9 неулучшаемо с точностью до константы. Действительно, из определений (399) и (392) получаем, что

$$2v = \sigma\xi = u, \quad \theta^* - \theta \equiv \theta^* - 1 = \sigma\xi/(1+\sigma\xi/2), \quad |\xi| = 1, \quad 2d_v = \sigma = d_u. \tag{419}$$

В частности, $|v| \le |u| \le \sigma = \kappa/2 < 1/2$ при $0 < \kappa < 1$. Но в силу (400) это означает, что $\widetilde{\theta} = \theta^*$. Значит, при $\theta = 1$ из определения (394) и равенств (419) находим, что $\breve{\phi}_1(\widetilde{\theta}) = \breve{\phi}_1(\theta^*) = \phi_1(\theta^*) = \sigma\xi \cdot K|\sigma\xi|^q \mathrm{sign}(\xi)/|1+\sigma\xi/2|^q = K\sigma^{1+q}/|1+\sigma\xi/2|^q$. Из этого соотношения и неравенства Иенсена заключаем, что

$$\mathbb{E}\sum_{i=1}^{n}\breve{\phi}_i(\widetilde{\theta}) = \mathbb{E}\breve{\phi}_1(\widetilde{\theta}) = \mathbb{E}\frac{K\sigma^{1+q}}{|1+\sigma\xi/2|^q} \geq \frac{K\sigma^{1+q}}{|1+\sigma\mathbb{E}\xi/2|^q} = K\sigma^{1+q}. \tag{420}$$

Пусть теперь $t_2 - t_1 = 2h > 0$. Нетрудно понять, что в этом случае

$$\gamma_1(t_2) - \gamma_1(t_1) = \int\limits_{t_1}^{t_2}\gamma_1'(t)dt = K\int\limits_{t_1}^{t_1+2h} q|t|^{q-1}dt \leq K\int\limits_{-h}^{h} q|t|^{q-1}dt = 2Kh^q = 2^{1-q}K|t_2-t_1|^q.$$

Но отсюда находим, что $|\phi_1(t_2) - \phi_1(t_1)| \leq |u_1| \cdot 2^{1-q}K|t_2-t_1|^q = \sigma|\xi| \cdot 2^{1-q}K|t_2-t_1|^q = 2^{1-q}K\sigma|t_2-t_1|^q$. Таким образом, при $i = 1$ условие (390) выполнено при $\overline{\phi}_1 = 2^{1-q}K\sigma$. А поскольку $\overline{\phi}_i = \phi_i(\cdot) = 0$ при $i \geq 2$ по предположению, то из определений (389) и (393) получаем, что $\mathbb{S}_m\big(\|\overline{\phi}_\bullet\|\big) = \|\overline{\phi}_1\| = 2^{1-q}K\sigma$ при всех $m > 0$. Следовательно, с учетом (419) и условия $\sigma < 1/2$, имеем

$$3 \cdot 2^{2q+1}(1+d_v^q)d_u^q\mathbb{S}_{2/(2-q)}\big(\|\overline{\phi}_\bullet\|\big) = 3 \cdot 2^{2q+1}\big(1+(\sigma/2)^q\big)\sigma^q 2^{1-q}K\sigma < 3 \cdot 2^4 K\sigma^{1+q}, \tag{421}$$

Из (421), (420) и (396) вытекают все неравенства, требуемые в (397). □

3.2.5 Доказательства результатов раздела 3.2.3

Д о к а з а т е л ь с т в о предложения 3.1 следует из теоремы 3.9 (см. раздел 3.2.4).

Д о к а з а т е л ь с т в о предложения 3.2. Положим $\theta^*_{ni} = \mathbb{E}_{\neq i}\theta^*_n$, $\theta^*_{nij} = \mathbb{E}_{\neq i}\mathbb{E}_{\neq j}\theta^*_n$, $i, j = 1, \ldots, n$, $j \neq i$. Подчеркнем, что величина θ^*_{ni} не зависят от наблюдения X_i, а величина θ^*_{nij} не зависят еще и от X_j. Заметим прежде всего, что при всех i и $j \neq i$ справедливы оценки

$$\mathbb{E}\big|\theta^*_{ni} - \theta\big|^2 \leq 4\mathbb{E}\big|\theta^*_n - \theta\big|^2 \qquad \mathbb{E}\big|\theta^*_{ni} - \theta^*_{nij}\big|^2 \leq \mathbb{E}\big|\theta^*_n - \theta^*_{nj}\big|^2. \tag{422}$$

Действительно, $\mathbb{E}\big|\theta^*_{ni} - \theta\big|^2 \leq 2\mathbb{E}\big|\theta^*_n - \theta\big|^2 + 2\mathbb{E}\big|\theta^*_n - \theta^*_{ni}\big|^2$. Кроме того, с учетом формулы полной вероятности, $\mathbb{E}\big|\theta^*_n - \theta^*_{ni}\big|^2 = \mathbb{E}\mathbb{E}_{\neq i}\big|\theta^*_n - \mathbb{E}_{\neq i}\theta^*_n\big|^2 \leq \mathbb{E}\mathbb{E}_{\neq i}\big|\theta^*_n - h_{\neq i}\big|^2$, где величина $h_{\neq i}$ не зависит от X_i. В частности, полагая $h_{\neq i} = \theta$, получаем первое утверждение в (422). Еще раз используя формулу полной вероятности, получаем цепочку соотношений $\mathbb{E}\big|\theta^*_{ni} - \theta^*_{nij}\big|^2 = \mathbb{E}\big|\mathbb{E}_{\neq i}\theta^*_n - \mathbb{E}_{\neq i}\mathbb{E}_{\neq j}\theta^*_n\big|^2 \leq \mathbb{E}\mathbb{E}_{\neq i}\big|\theta^*_n - \mathbb{E}_{\neq j}\theta^*_n\big|^2 = \mathbb{E}\big|\theta^*_n - \theta^*_{nj}\big|^2$.

Обозначим через $\rho_n^{(1)}$ и $\rho_n^{(2)}$ слагаемые в правой части следующего тождества:

$$\rho_n \equiv I_{n,h}^{-1/2}\sum_{i=1}^{n}\big(h_i(\theta^*_n) - h_i(\theta^*_{ni})\big)\psi_i(\theta, X_i) + I_{n,h}^{-1/2}\sum_{i=1}^{n}\big(h_i(\theta^*_{ni}) - h_i(\theta)\big)\psi_i(\theta, X_i), \tag{423}$$

где, напомним, величина ρ_n определена в (363). В условиях предложения

$$I_{n,h}^{1/2}\mathbb{E}|\rho_n^{(1)}| \le \sum_{i=1}^n H_i\mathbb{E}\big(|\theta_n^* - \theta_{ni}^*|^p|\psi_i(\theta, X_i)|\big) \le \sum_{i=1}^n H_i\big(\mathbb{E}|\theta_n^* - \theta_{ni}^*|^2\big)^{p/2}\big(\mathbb{E}|\psi_i(\theta, X_i)|^2\big)^{1/2}.$$

Следовательно, с учетом условия (384), $\mathbb{E}\big|\rho_n^{(1)}\big| \to 0$, а потому $\rho_n^{(1)} \xrightarrow{p} 0$.

Покажем, что $\rho_n^{(2)} \xrightarrow{p} 0$. Нам потребуются следующие оценки:

$$\mathbb{E}\big((h_i(\theta_{ni}^*) - h_i(\theta))^2\psi_i^2(\theta, X_i)\big) = \mathbb{E}\big(h_i(\theta_{ni}^*) - h_i(\theta)\big)^2\mathbb{E}\psi_i^2(\theta, X_i) \le$$
$$\le H_i^2\mathbb{E}\big(|\theta_{ni}^* - \theta|^2\big)^p\mathbb{E}\psi_i^2(\theta, X_i) \le 4^p\mathbb{E}\big(|\theta_n^* - \theta|^2\big)^p H_i^2\mathbb{E}\psi_i^2(\theta, X_i),$$

$$\mathbb{E}\Big[\big(h_i(\theta_{ni}^*) - h_i(\theta)\big)\big(h_j(\theta_{nj}^*) - h_j(\theta)\big)\psi_i(\theta, X_i)\psi_j(\theta, X_j)\Big] =$$
$$= \mathbb{E}\Big[\big(h_i(\theta_{ni}^*) - h_i(\theta_{nij}^*) + h_i(\theta_{nij}^*) - h_i(\theta)\big)\big(h_j(\theta_{nj}^*) - h_j(\theta_{nji}^*) + h_j(\theta_{nji}^*) - h_j(\theta)\big)\times$$
$$\times\, \psi_i(\theta, X_i)\psi_j(\theta, X_j)\Big] = \mathbb{E}\Big[\big(h_i(\theta_{ni}^*) - h_i(\theta_{nij}^*)\big)\big(h_j(\theta_{nj}^*) - h_j(\theta_{nji}^*)\big)\psi_i(\theta, X_i)\psi_j(\theta, X_j)\Big] +$$
$$+\, \mathbb{E}\Big[\big(h_i(\theta_{ni}^*) - h_i(\theta_{nij}^*)\big)\big(h_j(\theta_{nji}^*) - h_j(\theta)\big)\psi_i(\theta, X_i)\psi_j(\theta, X_j)\Big] +$$
$$+\, \mathbb{E}\Big[\big(h_i(\theta_{nij}^*) - h_i(\theta)\big)\big(h_j(\theta_{nj}^*) - h_j(\theta_{nji}^*)\big)\psi_i(\theta, X_i)\psi_j(\theta, X_j)\Big] +$$
$$+\, \mathbb{E}\Big[\big(h_i(\theta_{nij}^*) - h_i(\theta)\big)\big(h_j(\theta_{nji}^*) - h_j(\theta)\big)\psi_i(\theta, X_i)\psi_j(\theta, X_j)\Big] =$$
$$= \mathbb{E}\Big[\big(h_i(\theta_{ni}^*) - h_i(\theta_{nij}^*)\big)\big(h_j(\theta_{nj}^*) - h_j(\theta_{nji}^*)\big)\psi_i(\theta, X_i)\psi_j(\theta, X_j)\Big] \le$$
$$\le \Big(\mathbb{E}\big(h_i(\theta_{ni}^*) - h_i(\theta_{nij}^*)\big)^2\mathbb{E}\psi_i^2(\theta, X_i)\Big)^{1/2}\Big(\mathbb{E}\big(h_j(\theta_{nj}^*) - h_j(\theta_{nji}^*)\big)^2\mathbb{E}\psi_j^2(\theta, X_j)\Big)^{1/2} \le$$
$$\le H_iH_j\big(\mathbb{E}|\theta_n^* - \theta_{ni}^*|^2\big)^{p/2}\big(\mathbb{E}|\theta_n^* - \theta_{nj}^*|^2\big)^{p/2}\big(\mathbb{E}\psi_i^2(\theta, X_i)\big)^{1/2}\big(\mathbb{E}\psi_j^2(\theta, X_j)\big)^{1/2}.$$

При выводе этих соотношений были учтены еще неравенства из (422). Следовательно, с учетом условий (383) и (384)

$$\mathbb{E}\big|\rho_n^{(2)}\big|^2 = I_{n,h}^{-1}\sum_{i=1}^n \mathbb{E}\big((h_i(\theta_{ni}^*) - h_i(\theta))^2\psi_i^2(\theta, X_i)\big) +$$
$$+\, I_{n,h}^{-1}\sum_{i,j=1,i\ne j}^n \mathbb{E}\Big[\big(h_i(\theta_{ni}^*) - h_i(\theta)\big)\big(h_j(\theta_{nj}^*) - h_j(\theta)\big)\psi_i(\theta, X_i)\psi_j(\theta, X_j)\Big] \le$$
$$\le 4^pI_{n,h}^{-1}\mathbb{E}\big(|\theta_n^* - \theta|^2\big)^p\sum_{i=1}^n H_i^2\mathbb{E}\psi_i^2(\theta, X_i) +$$
$$+\, \left(I_{n,h}^{-1/2}\sum_{i=1}^n H_i\big(\mathbb{E}|\theta_n^* - \theta_{ni}^*|^2\big)^{p/2}\big(\mathbb{E}\psi_i^2(\theta, X_i)\big)^{1/2}\right)^2 \to 0,$$

что завершает доказательство предложения. □

Д о к а з а т е л ь с т в о предложения 3.3. Подставляя значения $h_i(\theta) = h_i^o(\theta)$ из (385) в определение величины $I_{n,\theta,h}J_{n,\theta,h}^{-2}$ утверждение (385) можно переписать в следующем виде:

$$\left(\sum_{i=1}^n h_i(\theta)\mathbb{E}\psi_i'(\theta, X_i)\right)^2 \le \sum_{i=1}^n h_i^2(\theta)\mathbb{E}\psi_i^2(\theta, X_i) \cdot \sum_{i=1}^n \frac{\left(\mathbb{E}\psi_i'(\theta, X_i)\right)^2}{\mathbb{E}\psi_i^2(\theta, X_i)}.$$

Справедливость (385) теперь немедленно следует из неравенства Коши-Буняковского.

Положим $K_n = \sum\limits_{i=1}^n \left(h_i^o(\theta)\right)^2$. Поскольку по условиям теоремы при каждом i величины $h_i(\theta)$ и $h_i^o(\theta)$ имеют одинаковый знак, то справедливы неравенства (387). В частности, при каждом i значение $h_i(\theta)$ лежит между $hh_i^o(\theta)$ и $Hh_i^o(\theta)$. Значит, для величин $d_i = (h_i(\theta) - hh_i^o(\theta))(h_i(\theta) - Hh_i^o(\theta))$, $i = 1, \ldots, n$, выполнено $d_i \le 0$. Из этого факта с учетом обозначений из (R_3^h) получаем, что

$$I_{n,h} - (H + h)J_{n,h} + hHK_n = \sum_{i=1}^n d_i\mathbb{E}\psi_i^2(\theta, X_i) \le 0.$$

А поскольку функция $f(x) = x^{-1} - Hh(H + h)^{-1}K_nx^{-2}$ достигает своего максимального значения при $x = 2Hh(H + h)^{-1}K_n$, то

$$I_{n,h}J_{n,h}^{-2} \le (H + h)J_{n,h}^{-1} - HhK_nJ_{n,h}^{-2} = (H + h)f(J_{n,h}) \le$$
$$\le (H + h)f\left(\frac{2HhK_n}{H + h}\right) = \frac{(H + h)^2}{4HhK_n} = \frac{I_{n,h^o}}{J_{n,h^o}^2}\frac{(H + h)^2}{4Hh}.$$

Значит,

$$\frac{I_{n,h}J_{n,h}^{-2}}{I_{n,h^o}J_{n,h^o}^{-2}} \le \frac{H + h}{2\sqrt{Hh}} = \frac{H/h + 1}{2\sqrt{H/h}} = 1 + \frac{(\sqrt{H/h} - 1)^2}{2\sqrt{H/h}} \le \sqrt{\frac{H}{h}}$$

и доказательство предложения завершено. □

3.3 Уточнение одношаговых оценок Фишера

Рассматривается задача оценивания числового параметра $\theta \in \Theta \subset \mathbb{R}$ по выборке объема n из последовательности $X_1, X_2, \ldots$ независимых одинаково распределенных случайных величин произвольной природы, распределение которых имеет плотность $f_\theta(\cdot)$ относительно некоторой меры. Приводится алгоритм построения асимптотически эффективных одношаговых оценок неизвестного параметра в случае, когда имеется предварительная n^β-состоятельная оценка при $\beta < 1/4$. Известные ранее одношаговые оценки, предложенные Фишером как приближения для состоятельных оценок максимального правдоподобия, позволяют за один шаг находить асимптотически эффективные оценки лишь в случае, когда $\beta \ge 1/4$.

3.3.1 Алгоритм построения оценок и основные результаты

1. В этом разделе считаем выполненным следующее предположение.

(**B**). *Наблюдаются первые n элементов из последовательности независимых одинаково распределенных случайных величин $X_1,\ X_2, \ldots$ со значениями в произвольном измеримом пространстве $(\mathcal{X}, \mathcal{A})$ с плотностью $f_\theta(x)$ относительно некоторой меры μ на $\mathcal{A}$, где неизвестный параметр θ может принимать значения из некоторого открытого множества $\Theta \subset \mathbb{R}$. Кроме того, логарифмическая плотность $l(t,x) = \ln f_t(x)$ для почти всех (по мере μ) значений $x \in \mathcal{X}$ дважды непрерывно дифференцируема по t при всех $t \in \Theta$, и имеют место соотношения*

$$\mathbb{E}l'(\theta, X_1) = 0, \quad 0 < I(\theta) = \mathbb{E}\big(l'(\theta, X_1)\big)^2 = -\mathbb{E}l''(\theta, X_1) < \infty. \tag{424}$$

Математическое ожидание в (424) и далее вычисляется по отношению к распределению с плотностью $f(\theta, x)$. Условимся, что мы будем писать $\mathbb{E}$ и $\mathbb{P}$ вместо $\mathbb{E}_\theta$ и $\mathbb{P}_\theta$, соответственно.

При некотором фиксированном $k \geq 1$ нам также потребуется условие

($\mathbf{C}_k$). *Имеется некоторая $n^{1/2(k+2)}$-состоятельная оценка $\theta^*_{n,k} = \theta^*_{n,k}(X_1, \ldots, X_n)$ параметра θ, т.е. оценка такая, что $n^{1/2(k+2)}(\theta^*_{n,k} - \theta) \xrightarrow{p} 0$.*

Пусть выполнены предположения (B) и (C_k). Задача состоит в том, чтобы построить оценку $\theta^{**}_{n,k} = \theta^{**}_{n,k}(\theta^*_{n,k})$ такую, для которой имеет место сходимость

$$\sqrt{n}\big(\theta^{**}_{n,k} - \theta\big) \xrightarrow{d} \mathcal{N}_1(0, I^{-1}(\theta)), \tag{425}$$

т.е. за один шаг улучшить предварительную оценку $\theta^*_{n,k}$ параметра θ до асимптотически эффективной оценки $\theta^{**}_{n,k}$.

Одношаговую оценку $\theta^{**}_{n,k}$ предлагается искать в виде

$$\theta^{**}_{n,k} = \theta^*_{n,k} + L'_n(\theta^*_{n,k}) \left(\sum_{i=1}^n \lambda_k(\theta^*_{n,k}, X_i) \right)^{-1}, \tag{426}$$

где функция $\lambda_k(t,x)$ определяется скоростью сближения предварительной оценки $\theta^*_{n,k}$ и параметра θ, заданной в условии (C_k), $L_n(t) = \sum_{i=1}^n l(t, X_i)$. Забегая вперед, отметим, что функция $\lambda_k(t,x)$ из (426) будет иметь следующую структуру:

$$\lambda_k(\theta, x) = I(\theta) + c_{k1,\theta} \frac{f'_\theta(x)}{f_\theta(x)} + c_{k2,\theta} \frac{f''_\theta(x)}{f_\theta(x)} + \ldots + c_{kk,\theta} \frac{f^{(k)}_\theta(x)}{f_\theta(x)},$$

где $c_{k1,\theta}, \ldots, c_{kk,\theta}$ — решения некоторой системы линейных уравнений (см. приводимую далее формулу (438)), дифференцирование производится по переменной θ и через $f^{(k)}_t(x)$ обозначена k-я производная плотности $f_t(x)$ по t. Таким образом, предлагаемые одношаговые оценки $\theta^{**}_{n,k}$ отличаются от оценки Фишера θ^{**}_{n2}, определенной в (22), наличием некоторых дополнительных слагаемых в знаменателе соответствующей дроби, при этом количество этих слагаемых определяется скоростью сходимости $\theta^*_{n,k}$ (т.е. задается значением k). Кроме того, от значения k зависят и коэффициенты $c_{k1,\theta}, ..., c_{kk,\theta}$, задающие эти аддитивные добавки. Тем самым, с ростом k не только добавляется соответствующее количество слагаемых, но при каждом новом k заново пересчитываются и коэффициенты $c_{k1,\theta}, ..., c_{kk,\theta}$.

Вернемся к построению одношаговой оценки $\theta^{**}_{n,k}$. Как будет показано далее, для наших целей нужно, чтобы функция $\lambda_k(t,x)$ удовлетворяла соотношениям

$$\mathbb{E}\Big(m\lambda_k^{(m-1)}(\theta, X_1) + l^{(m+1)}(\theta, X_1)\Big) = 0 \qquad \text{при всех} \quad 1 \le m \le k+1. \tag{427}$$

Такие функции $\lambda_k(t,x)$ будут найдены в приводимой ниже теореме 3.11. Здесь же поясним идею построения одношаговых оценок $\theta^{**}_{n,k}$, приводящую к выбору в качестве функции $\lambda_k(t,x)$ решение системы интегро-дифференциальных уравнений (427). Воспользуемся следующим тождеством, которое нетрудно извлечь из (426):

$$\sqrt{n}\big(\theta^{**}_{n,k} - \theta\big) = \frac{n^{-1/2}\sum_{i=1}^n g_k(\theta^*_{n,k}, X_i)}{n^{-1}\sum_{i=1}^n \lambda_k(\theta^*_{n,k}, X_i)}, \quad \text{где} \quad g_k(t,x) = (t-\theta)\lambda_k(t,x) + l'(t,x). \tag{428}$$

Поскольку при широких ограничениях

$$n^{-1}\sum_{i=1}^n \lambda_k(\theta^*_{n,k}, X_i) \xrightarrow{p} \mathbb{E}\lambda_k(\theta, X_1), \tag{429}$$

то асимптотическое поведение разности $\sqrt{n}\big(\theta^{**}_{n,k} - \theta\big)$ определяется поведением числителя в правой части первого равенства в (428). При соответствующей гладкости функции $g_k(t,x)$ по первому аргументу имеем

$$\begin{aligned} n^{-1/2}\sum_{i=1}^n g_k(\theta^*_{n,k}, X_i) = n^{-1/2}\sum_{i=1}^n g_k(\theta, X_i) + \\ + \sum_{m=1}^{k+1} \frac{(\theta^*_{n,k} - \theta)^m}{m!} n^{-1/2}\sum_{i=1}^n g_k^{(m)}(\theta, X_i) + \sqrt{n}(\theta^*_{n,k} - \theta)^{k+2} G_{n,k}, \end{aligned} \tag{430}$$

где $g_k^{(m)}(t,x)$ — m-я производная функции $g_k(t,x)$ по t, $G_{n,k}$ — остаточный член формулы Тейлора, который будет определен в лемме 3.15. Кроме того, при некоторых естественных ограничениях

$$n^{-1/2}\sum_{i=1}^n g_k(\theta, X_i) \equiv n^{-1/2}\sum_{i=1}^n l'(\theta, X_i) \xrightarrow{d} \mathcal{N}_1(0, I(\theta)) \quad \text{и} \quad G_{n,k} = O_p(1). \tag{431}$$

Второе соотношение в (431) и условие (C_k) позволяют заключить, что последнее слагаемое в правой части (430) сходится по вероятности к нулю. Допустим еще, что функция $\lambda_k(\theta, X_1)$ такова, что выполнено следующее ключевое соотношение:

$$\mathbb{E}g_k^{(m)}(\theta, X_1) = 0 \qquad \text{при} \quad 1 \leq m \leq k+1. \tag{432}$$

В этом случае при соответствующих моментных ограничениях распределения сумм

$$n^{-1/2}\sum_{i=1}^{n} g_k^{(m)}(\theta, X_i)$$

сходятся к нормальному распределению и из первого соотношения в (431) и разложения (430) следует, что $n^{-1/2}\sum_{i=1}^{n} g_k(\theta^*_{n,k}, X_i) \xrightarrow{d} \mathcal{N}_1(0, I(\theta))$. Тем самым, выполнено (425), если еще учесть (428), (429) и считать, что справедливо равенство $\mathbb{E}\lambda_k(\theta, X_1) = I(\theta)$.

Нам остается заметить, что в силу представления (428) для любого натурального m имеет место равенство

$$g_k^{(m)}(t, x) = m\lambda_k^{(m-1)}(t, x) + l^{(m+1)}(t, x) + (t-\theta)\lambda_k^{(m)}(t, x), \tag{433}$$

а потому соотношение (432) эквивалентно (427).

Приведем теперь точное утверждения об асимптотических свойствах оценок $\theta^{**}_{n,k}$. Нам потребуется следующее предположение.

$(\mathbf{D}_k)$ 1) *функция* $\lambda_k(t, x)$ *для почти всех значений* x *(по мере* μ*)* $k+1$ *раз дифференцируема по* t *при всех* $t \in \Theta$ *и*

$$\begin{gathered}\mathbb{E}\lambda_k(\theta, X_1) = I(\theta), \quad \mathbb{E}|\lambda_k^{(m)}(\theta, X_1)|^2 < \infty \quad \text{при} \quad m = 0, \dots, k, \\ \mathbb{E}\sup_{t\in\Theta:|t-\theta|\leq\delta_1} |\lambda_k^{(k+1)}(t, X_1)| < \infty \quad \text{при некотором} \quad \delta_1 > 0;\end{gathered} \tag{434}$$

2) *функция* $l(t, x)$ *для почти всех значений* x *(по мере* μ*)* $k+2$ *раза дифференцируема по* t *при всех* $t \in \Theta$, $\mathbb{E}|l^{(m+2)}(\theta, X_1)|^2 < \infty$ *при* $0 \leq m \leq k$ *и*

$$\mathbb{E}\sup_{t\in\Theta:|t-\theta|<\delta_2, t\neq\theta} \frac{|l^{(k+2)}(t, X_1) - l^{(k+2)}(\theta, X_1)|}{|t-\theta|} < \infty \text{ при некотором } \delta_2 > 0. \tag{435}$$

Имеет место следующее утверждение.

Теорема 3.10. *Пусть выполнено предположение* (B) *и при некотором* $k \geq 1$ *справедливы предположения* (C_k) *и* (D_k)*. Кроме того, функция* $\lambda_k(\theta, x)$ *удовлетворяет системе уравнений* (427)*. Тогда одношаговая оценка* $\theta^{**}_{n,k}$*, определенная в* (426)*, асимптотически эффективна.*

2. В этом разделе найдем функции $\lambda_k(t,x)$, удовлетворяющие (427) для любого целого $k \geq 1$. В пространстве $L_2 = \{g(\cdot) : \mathbb{E}g^2(X_1) < \infty\}$ функций с конечным вторым моментом определим скалярное произведение $\langle \alpha(\cdot), \beta(\cdot) \rangle$ соотношением

$$\langle \alpha(\cdot), \beta(\cdot) \rangle = \mathbb{E}\alpha(X_1)\beta(X_1), \qquad \alpha(\cdot), \beta(\cdot) \in L_2. \tag{436}$$

Положим

$$e_{m,\theta}(\cdot) = \frac{f_\theta^{(m)}(\cdot)}{f_\theta(\cdot)}, \quad a_{m,\theta} = \frac{1}{m+1}\langle e_{1,\theta}(\cdot), e_{m+1,\theta}(\cdot) \rangle, \tag{437}$$

где $m \geq 0$. При любом фиксированном $k \geq 1$ обозначим через $c_{k1,\theta}$, ..., $c_{kk,\theta}$ решения следующей системы k линейных уравнений с k неизвестными:

$$c_{k1,\theta}\langle e_{1,\theta}(\cdot), e_{j,\theta}(\cdot)\rangle + \ldots + c_{kk,\theta}\langle e_{k,\theta}(\cdot), e_{j,\theta}(\cdot)\rangle = a_{j,\theta}, \quad j = 1, \ldots, k. \tag{438}$$

Отметим, что матрица, определяющая систему (438) (матрица Грама, построенная по элементам $\{e_{m,\theta}(\cdot);\ m \leq k\}$), является положительно определенной в случае линейной независимости функций $\{e_{m,\theta}(\cdot);\ m \leq k\}$ и, тем самым, система (438) однозначно задает элементы $c_{k1,\theta}$, ..., $c_{kk,\theta}$. Определим функцию $\lambda_k(\theta,x)$ равенством

$$\lambda_k(\theta,x) = I(\theta) + c_{k1,\theta}e_{1,\theta}(x) + c_{k2,\theta}e_{2,\theta}(x) + \ldots + c_{kk,\theta}e_{k,\theta}(x). \tag{439}$$

Почему функция $\lambda_k(\theta,x)$ из (439) удовлетворяет системе (427)? Оказывается, при достаточно общих условиях имеет место следующее ключевое утверждение (см. лемму 3.17): если функция $\lambda_k(\theta,x)$ является решением системы уравнений

$$\langle \lambda_k(\theta,\cdot), e_{m-1,\theta}(\cdot)\rangle = a_{m-1,\theta}, \qquad m = 1, \ldots, k+1, \tag{440}$$

то функция $\lambda_k(\theta,x)$ удовлетворяет системе (427). Тем самым, поиск подходящей функции $\lambda_k(\theta,x)$ сводится к отысканию решения системы (440). В предположении линейной независимости $e_{0,\theta}(x), \ldots,$ $e_{1,\theta}(x)$ система (440) задает проекции $a_{0,\theta}, \ldots,$ $a_{k,\theta}$ искомой функции $\lambda_k(\theta,x)$ на элементы базиса $e_{0,\theta}(x), \ldots,$ $e_{1,\theta}(x)$. Известно, что в этом случае линейная комбинация базисных элементов $\sum_{j=0}^k c_{kj,\theta}e_{j,\theta}(x)$ является решением (440), где коэффициенты $\{c_{kj,\theta}\}$ определяются системой линейных уравнений с матрицей Грама, построенной по элементам базиса $\{e_{j,\theta}(x)\}$ (см. доказательство леммы 3.18). Если же функции $f_\theta(\cdot)$, $f'_\theta(\cdot), \ldots,$ $f_\theta^{(k)}(\cdot)$ линейно зависимы (т.е. линейно зависимы функции $\{e_{m,\theta}(\cdot);\ m \leq k\}$), то, как хорошо известно, либо не существует решений системы линейных уравнений (438), либо их будет бесконечно много. Иными словами, в этих условиях либо невозможно представить искомую функцию $\lambda_k(\theta,x)$ в виде (439), либо таких функций бесконечно много, и алгоритм их поиска сводится к известной процедуре линейной алгебры.

Чтобы сформулировать основное утверждение раздела, нам потребуется следующее предположение.

($\mathbf{E}_k$) *Выполнено условие* (D_k) *без ограничений* (434) *и* (435), *функции* $f_\theta(\cdot)$, $f'_\theta(\cdot),\ldots,$ $f^{(k)}_\theta(\cdot)$ *линейно независимы и* $e_{2,\theta}(\cdot),\ldots,e_{k+1,\theta}(\cdot)\in L_2$. *Кроме того,* $\displaystyle\int f^{(j)}(\theta,x)\mu(dx)=0$, $j=3,\ldots,k+1$, *и можно менять порядок дифференцирования по переменной* θ *и интегрирования по переменой* x *в следующих случаях:*

$$\left(\int\lambda_k(\theta,x)f_\theta^{(m-1)}(x)\mu(dx)\right)^{(j)}=\int\left(\lambda_k(\theta,x)f_\theta^{(m-1)}(x)\right)^{(j)}\mu(dx),$$
$$\left(\int l'(\theta,x)f_\theta^{(k+1)}(x)\mu(dx)\right)^{(j)}=\int\left(l'(\theta,x)f_\theta^{(k+1)}(x)\right)^{(j)}\mu(dx)$$

при $m=1,\ldots,k+1$ *и* $j=1,\ldots,k+1-m$.

Имеет место

Теорема 3.11. *Пусть выполнено предположение* (B) *и условие* (E_k) *при некотором фиксированном* $k\ge 1$. *Тогда функция* $\lambda_k(t,x)$, *определенная в* (439), *является решением системы уравнений* (427).

3. Из теорем 3.10 и 3.11 вытекает следующее утверждение.

Следствие 3.2. *Пусть выполнено предположение* (B), *а при некотором фиксированном* $k\ge 1$ — *предположения* (C_k), (D_k) *и* (E_k). *Тогда одношаговая оценка* $\theta^{**}_{n,k}$, *определенная в* (426) *и* (439), *является асимптотически эффективной.*

Выделим в отдельное утверждение следствие из теорем 3.10 и 3.11 для случая $k=1$, то есть для случая, когда выполнено условие $n^{1/6}$-состоятельности предварительной оценки. Из определений (437), (438) и (439) нетрудно извлечь следующее представление для одношаговой статистики $\theta^{**}_{n,1}$:

$$\theta^{**}_{n,1}=\theta^*_{n,1}+\frac{L'_n(\theta^*_{n,1})}{nI(\theta^*_{n,1})+L'_n(\theta^*_{n,1})c(\theta^*_{n,1})},\quad\text{где}\quad c(\theta)=\frac{1}{2I(\theta)}\int\frac{f'_\theta(x)f''_\theta(x)}{f_\theta(x)}\mu(dx).\tag{441}$$

Следствие 3.3. *Пусть выполнены условия* (B), (C_1) *и следующие предположения:*

1) *функция* $l(t,x)$ *для почти всех значений* x *(по мере* μ*) трижды дифференцируема по* t, *функции* $I(t)$ *и* $a_{1,t}$ — *дважды дифференцируемы при всех* $t\in\Theta$, *а при дифференцировании интегралов* $\int l'(t,x)f'_t(x)\mu(dx)$ *и* $\int l'(t,x)f''_t(x)\mu(dx)$ *по* t *при всех* $t\in\Theta$ *можно менять порядок дифференцирования и интегрирования;*

2) *выполнено* (435) *при* $k=1$ *и* $\mathbb{E}|l''(\theta,X_1)|^2<\infty$, $\mathbb{E}|l'''(\theta,X_1)|^2<\infty$;

3) *функции* $f_\theta(\cdot)$ *и* $f'_\theta(\cdot)$ *линейно независимы.*

Тогда оценка $\theta^{**}_{n,1}$, *определенная в* (441), *асимптотически эффективна.*

3.3.2 Дополнительные комментарии

З а м е ч а н и е 3.15. Из теоремы 3.5 раздела 3.1.5 можно извлечь, что при достаточно слабых ограничениях условие $n^{1/4}$-состоятельности оценки θ_n^* является необходимым для сходимости $\sqrt{n}\big(\theta_{n1}^{**}-\theta\big)-n^{-1/2}L_n'(\theta)I^{-1}(\theta)\xrightarrow{p}0$. Таким образом, условие $n^{1/4}$-состоятельности предварительной оценки в широких условиях оказывается оказывается минимальным достаточным ограничением для асимптотической эффективности одношаговых оценок Фишера. Непосредственно для оценок Фишера этот факт был ранее установлен в [337].

Схематично покажем, что указанное условие на точность предварительной оценки θ_n^* является необходимым ограничением для асимптотической эквивалентности одношаговых оценок Фишера θ_{n1}^{**} и θ_{n2}^{**} и оценки максимального правдоподобия $\widetilde{\theta}_n$. Действительно, для производной логарифмической функции правдоподобия имеет место равенство $0=L_n'(\widetilde{\theta}_n)=L_n'(\theta_n^*)+(\widetilde{\theta}_n-\theta_n^*)L_n''(\theta_n^*)+(\widetilde{\theta}_n-\theta_n^*)^2L_n'''(\widetilde{\theta}_n^*)$, где $\widetilde{\theta}_n^*$ — промежуточная точка между θ_n^* и оценкой максимального правдоподобия $\widetilde{\theta}_n$. Используя это равенство и определение (22), получаем следующее тождество:

$$\sqrt{n}\big(\theta_{n1}^{**}-\widetilde{\theta}_n\big)=\sqrt{n}(\theta_n^*-\widetilde{\theta}_n)-\sqrt{n}\frac{L_n'(\theta_n^*)-L_n'(\widetilde{\theta}_n)}{L_n''(\theta_n^*)}=\frac{1}{2}\left(n^{1/4}(\theta_n^*-\widetilde{\theta}_n)\right)^2\frac{L_n'''(\widetilde{\theta}_n^*)}{L_n''(\theta_n^*)},$$

где в широких условиях $L_n'''(\widetilde{\theta}_n^*)/L_n''(\theta_n^*)\xrightarrow{p}-\mathbb{E}l'''(\theta,X_1)/I(\theta)$. А потому, если $\sqrt{n}\big(\theta_{n1}^{**}-\widetilde{\theta}_n\big)\xrightarrow{p}0$ и $\mathbb{E}l'''(\theta,X_1)\neq 0$, то с необходимостью $n^{1/4}(\theta_n^*-\theta)\xrightarrow{p}0$. Аналогичные рассуждения справедливы и для второй оценки θ_{n2}^{**} из (22). □

З а м е ч а н и е 3.16. Пусть распределение сосредоточено на положительной полуоси и плотность $f_\theta(x)$ при $x\geq 0$ задается соотношением

$$f_\theta(x)=\frac{(h+1)}{\theta(1+x/\theta)^{2+h}},\qquad h\in\big(0,1\big).\tag{442}$$

Нетрудно проверить, что $\mathbb{E}X_1=\theta/h$. А потому $\theta_n^*=h\sum\limits_{i=1}^{n}X_i/n$ есть оценка по методу моментов для параметра θ. Имеет место следующее утверждение: *если выполнено условие* (442), *то распределение* $n^\gamma(\theta_n^*-\theta)$ *при* $\gamma=h/(h+1)\in(0,1/2)$ *сходится к некоторому устойчивому распределению.*

Действительно, $\mathbb{E}X_1<\infty$ при $h>0$ и $\mathbb{E}X_1^2=\infty$ при $h<1$. Согласно [349, стр. 643] для сходимости соответствующих распределений к устойчивому закону нужно, чтобы выполнялось соотношение

$$\frac{n\mu(a_n)}{a_n^2}\to const,\quad \text{где}\quad \mu(x)=\int\limits_{-x}^{x}y^2f_\theta(y)dy\quad \text{и}\quad a_n=n^{1-\gamma}.$$

Учитывая (442), имеем

$$\mu(x) = (h+1)\int\limits_0^x \frac{y^2/\theta}{(1+y/\theta)^{h+2}}dy = C_0\int\limits_1^{x+1}\frac{(y-1)^2}{y^{h+2}}dy \sim C_1x^{-h+1}.$$

Следовательно, $\dfrac{n\mu(a_n)}{a_n^2} \sim n^{(1-\gamma)(1-h)+1-2(1-\gamma)} \sim const$, если $-(1-\gamma)(1+h)+1 = 0$. Из последнего соотношения находим, что $\gamma = h/(h+1)$.

Пусть $h \in \left(0, \frac{1}{3}\right]$. В этом случае $\gamma = h/(h+1) \in \left(0, \frac{1}{4}\right]$ и $n^\gamma(\theta_n^* - \theta) = O_p(1)$ при $\gamma \leq 1/4$. Следовательно, оценка θ_n^* является n^β-состоятельной при $\beta < \gamma \leq 1/4$. Кроме того, $\mathbb{E}l'''(\theta, X_1) \neq 0$. В такой ситуации для одношагового поиска *явной* асимптотически эффективной оценки не годятся оценки Фишера (22) (см. замечание 3.15), но можно использовать построенные выше одношаговые оценки $\theta_{n,k}^{**}$. Понятно, что индекс k при этом нужно выбирать в зависимости от скорости сближения θ_n^* к θ, определяемой значением h. Так, например, если $h \in (\frac{1}{5}, \frac{1}{3}]$, то выполнено условие $n^{1/6}$-состоятельности для $\theta_{n,1}^* = \theta_n^*$ (т.е. $k = 1$), и в качестве явной одношаговой асимптотически эффективной оценки можно использовать статистику $\theta_{n,1}^{**}$, определенную в (441).

Пусть $h \in (\frac{1}{3}, 1]$. В этом случае $n^\gamma(\theta_n^* - \theta) = O_p(1)$ при $\gamma \in (\frac{1}{4}, \frac{1}{2}]$. И этот диапазон h — область применения классических одношаговых оценок Фишера (22). □

З а м е ч а н и е 3.17. Чем медленнее предварительная оценка $\theta_{n,k}^*$ сближается с параметром θ (т.е. чем больше значение k), тем более гладкой должна быть плотность распределения, чтобы иметь возможность за один шаг, с помощью $\theta_{n,k}^{**}$, находить асимптотически эффективную оценку. Этот факт наглядно демонстрируют условия (C_k), (D_k) и определения (439), (437).

Аналогичная взаимосвязь между точностью предварительной оценки и гладкостью функций, определяющих одношаговые оценки, отмечена в замечании 3.3 раздела 3.1.1.

З а м е ч а н и е 3.18. Если имеется $n^{1/2(k+2)}$-состоятельная предварительная оценка, то для отыскания асимптотически эффективной оценки можно, с одной стороны, использовать одношаговую статистику $\theta_{n,k}^{**}$. С другой стороны, нельзя не отметить и альтернативную процедуру, связанную с многошаговым улучшением медленно сходящейся к параметру оценки (см., например, [30], [242]). Напомним, что многошаговые процедуры обсуждались выше, в разделе 3.1.8). Итак, применяя последовательно определенное количество раз преобразование Фишера, можно найти асимптотически эффективную оценку за число шагов $\ln(k+2)/\ln 2 = \log_2(k+2)$. Действительно, при некоторых дополнительных ограничениях для оценки Фишера θ_{n1}^{**} справедливо представление

$$\theta_{n1}^{**} - \theta = \frac{(\theta_n^* - \theta)L_n''(\theta_n^*) - L_n'(\theta_n^*)}{L_n''(\theta_n^*)} \sim \frac{\zeta_n}{\sqrt{n}} + (\theta_n^* - \theta)^2 \cdot C_n,$$

где $\zeta_n \xrightarrow{d} \mathcal{N}_1(0, I^{-1}(\theta))$ и $C_n \xrightarrow{p} -\dfrac{\mathbb{E}l'''(\theta, X_1)}{2I(\theta)}$. Таким образом, если мы располагаем n^{ε}-состоятельной оценкой θ_n^* параметра θ и $\mathbb{E}|l'''(\theta, X_1)| < \infty$, то $\theta_{n1}^{**} = \theta_{n1}^{**}(\theta_n^*)$ будет $n^{2\varepsilon}$-состоятельной оценкой (если $2\varepsilon < 1/2$). Далее, используя θ_{n1}^{**} в качестве предварительной оценки, мы получим $n^{4\varepsilon}$-состоятельную оценку (если $4\varepsilon < 1/2$) и т.д. Нетрудно видеть, что продолжая этот процесс некоторое число раз, мы достигнем нужной точности (т.е. $\sqrt{n}$-ограниченности), при этом количество итераций будет равно $[-\ln 2\varepsilon/\ln 2] + 1$. $\square$

3.3.3 Доказательства

В этом разделе предполагаем фиксированным некоторое значение $k \geq 1$ и считаем выполненным предположение (B), особо это не оговаривая. Для доказательства теоремы 3.10 нам потребуется три вспомогательных утверждения.

Лемма 3.14. *Пусть выполнено предположение (D_k). Тогда*

$$\tau_\lambda(\delta) = \mathbb{E} \sup_{t \in \Theta: |t-\theta| \leq \delta} |\lambda_k(t, X_1) - \lambda_k(\theta, X_1)| \to 0 \quad \textit{при} \quad \delta \to 0,$$

$$\tau_g(\delta) = \mathbb{E} \sup_{t \in \Theta: |t-\theta| \leq \delta, t \neq \theta} \frac{|g_k^{(k+1)}(t, X_1) - g_k^{(k+1)}(\theta, X_1)|}{|t-\theta|} < \infty, \; \delta = \min\{\delta_1, \delta_2\},$$

где функция $g_k(t,x)$ определена в (428).

Д о к а з а т е л ь с т в о. Воспользуемся разложением

$$\lambda_k(t, x) = \sum_{m=0}^{k} \frac{(t-\theta)^m}{m!} \lambda_k^{(m)}(\theta, x) + \frac{(t-\theta)^{k+1}}{(k+1)!} \lambda_k^{(k+1)}(\widetilde{t}, x),$$

где $\widetilde{t}$ — некоторое промежуточное значение между t и θ. При $\delta < \delta_1$, имеем

$$\tau_\lambda(\delta) \leq \sum_{m=1}^{k} \frac{\delta^m}{m!} \mathbb{E}|\lambda_k^{(m)}(\theta, X_1)| + \frac{\delta^{k+1}}{(k+1)!} \mathbb{E} \sup_{t: |t-\theta| \leq \delta} |\lambda_k^{(k+1)}(\widetilde{t}, x)|. \tag{443}$$

Эта оценка доказывает первое утверждение леммы, если еще учесть конечность математических ожиданий в правой части (443) ввиду условия (D_k). Второе утверждение леммы нетрудно извлечь из представления (433), если еще учесть оценки

$$\mathbb{E} \sup_{t \in \Theta: |t-\theta| \leq \delta} |\lambda_k^{(k+1)}(t, X_1) - \lambda_k^{(k+1)}(\theta, X_1)| \leq$$
$$\leq \mathbb{E} \sup_{t \in \Theta: |t-\theta| \leq \delta_1} |\lambda_k^{(k+1)}(t, X_1)| + \mathbb{E}|\lambda_k^{(k+1)}(\theta, X_1)| < \infty,$$

$$\mathbb{E} \sup_{t \in \Theta: |t-\theta| \leq \delta, t \neq \theta} \frac{|\lambda_k^{(k)}(t, X_1) - \lambda_k^{(k)}(\theta, X_1)|}{|t-\theta|} \leq \mathbb{E} \sup_{t \in \Theta: |t-\theta| \leq \delta_1} |\lambda_k^{(k+1)}(t, X_1)| < \infty. \quad \square$$

Лемма 3.15. *Пусть выполнены предположения* (C_k) *и* (D_k). *Тогда*

$$G_{n,k}=\sum_{i=1}^{n}\frac{\big(g_k^{(k+1)}(\widetilde{\theta}^*_{n,k},X_i)-g_k^{(k+1)}(\theta,X_i)\big)}{n(\theta^*_{n,k}-\theta)}=O_p(1),$$

где $\widetilde{\theta}^*_{n,k}$ — *некоторая промежуточная точка между* θ *и* $\theta^*_{n,k}$.

Д о к а з а т е л ь с т в о. Для любого $K>0$ и $\delta=\min\{\delta_1,\delta_2\}$, имеем

$$\begin{aligned}\mathbb{P}\big(|G_{n,k}|>K\big)&\le\mathbb{P}\big(|G_{n,k}|>K,\ |\theta^*_{n,k}-\theta|\le\delta\big)+\mathbb{P}(|\theta^*_{n,k}-\theta|>\delta)\le\\ &\le\mathbb{P}\left(\sup_{t:|t-\theta|\le\delta,t\ne\theta}\sum_{i=1}^{n}\frac{\big|g_k^{(k+1)}(t,X_i)-g_k^{(k+1)}(\theta,X_i)\big|}{n|t-\theta|}>K,\ |\theta^*_{n,k}-\theta|\le\delta\right)+\\ &+\mathbb{P}(|\theta^*_{n,k}-\theta|>\delta)\le\tau_g(\delta)/K+\mathbb{P}(|\theta^*_{n,k}-\theta|>\delta).\end{aligned}\tag{444}$$

При выводе (444) мы использовали формулу полной вероятности и неравенство Чебышева с первым моментом. Следовательно, в силу леммы 3.14 и состоятельности оценки $\theta^*_{n,k}$, выполнено соотношение $\limsup\mathbb{P}\big(|G_{n,k}|>K\big)\le\tau_g(\delta)/K\to0$ при $K\to\infty$, что доказывает утверждение леммы. □

Лемма 3.16. *Пусть выполнены предположения* (C_k) *и* (D_k). *Тогда*

$$\Lambda_{n,k}=n^{-1}\sum_{i=1}^{n}\lambda_k(\theta,X_i)\xrightarrow{p}I(\theta),\quad\overline{\Lambda}_{n,k}=n^{-1}\sum_{i=1}^{n}\big(\lambda_k(\theta^*_{n,k},X_i)-\lambda_k(\theta,X_i)\big)\xrightarrow{p}0.$$

Д о к а з а т е л ь с т в о. Первое утверждение леммы следует из закона больших чисел, поскольку функция $\lambda_k(\theta,x)$ такова, что $\mathbb{E}\lambda_k(\theta,X_1)=I(\theta)$.

Аналогично выводу (444), для любого $\varepsilon>0$ и $\delta>0$, имеем

$$\begin{aligned}\mathbb{P}\big(|\overline{\Lambda}_{n,k}|>\varepsilon\big)&\le\mathbb{P}\big(|\overline{\Lambda}_{n,k}|>\varepsilon,\ |\theta^*_{n,k}-\theta|\le\delta\big)+\mathbb{P}(|\theta^*_{n,k}-\theta|>\delta)\le\\ &\le\mathbb{P}\left(\sup_{t:|t-\theta|\le\delta}\sum_{i=1}^{n}\big|\lambda_k(t,X_i)-\lambda_k(\theta,X_i)\big|/n>\varepsilon,\ |\theta^*_{n,k}-\theta|\le\delta\right)+\\ &+\mathbb{P}(|\theta^*_{n,k}-\theta|>\delta)\le\tau_\lambda(\delta)/\varepsilon+\mathbb{P}(|\theta^*_{n,k}-\theta|>\delta).\end{aligned}$$

Таким образом, $\limsup\mathbb{P}\big(|\overline{\Lambda}_{n,k}|>\varepsilon\big)\le\tau_\lambda(\delta)/\varepsilon$. А поскольку по лемме 3.14 $\tau_\lambda(\delta)\to0$ при $\delta\to0$, то в силу произвольности ε и δ второе утверждение леммы доказано. □

Д о к а з а т е л ь с т в о теоремы 3.10. Заметим, что в условиях теоремы

$$\overline{G}_{nk}=\sum_{i=1}^{n}\big(g_k(\theta^*_{n,k},X_i)-g_k(\theta,X_i)\big)/\sqrt{n}\xrightarrow{p}0.\tag{445}$$

Действительно, с учетом обозначения из леммы 3.15, имеем

$$\overline{G}_{nk} \equiv \sum_{m=1}^{k+1} \frac{(\theta^*_{n,k} - \theta)^m}{m!} \sum_{i=1}^{n} g_k^{(m)}(\theta, X_i)/\sqrt{n} + \sqrt{n}(\theta^*_{n,k} - \theta)^{k+2} G_{n,k}. \tag{446}$$

Последнее слагаемое в правой части (446) сходится по вероятности к нулю в силу леммы 3.15 и условия (C_k). Далее, по построению (см. формулу (432)) имеем $\mathbb{E} g_k^{(m)}(\theta, X_i) = 0$ при $m = 1, \ldots, k+1$, а в в силу (433) и условия (D_k) выполнено $\mathbb{E}|g_k^{(m)}(\theta, X_i)|^2 < \infty$. Следовательно, в силу центральной предельной теоремы распределения $\sum_{i=1}^n g_k^{(m)}(\theta, X_i)/\sqrt{n}$ сходятся к нормальному, а потому все слагаемые по m в правой части (446) сходится по вероятности к нулю в силу состоятельности $\theta^*_{n,k}$.

Из (428) и (431) получаем, что

$$\sqrt{n}\big(\theta^{**}_{n,k} - \theta\big) = \frac{\sum_{i=1}^n g_k(\theta^*_{n,k}, X_i)/\sqrt{n}}{\sum_{i=1}^n \lambda_k(\theta^*_{n,k}, X_i)/n} = \frac{\sum_{i=1}^n l'(\theta, X_i)/\sqrt{n} + \overline{G}_{nk}}{\Lambda_{n,k} + \overline{\Lambda}_{n,k}}. \tag{447}$$

Сходимость $\sqrt{n}\big(\theta^{**}_{n,k} - \theta\big) \xrightarrow{d} \mathcal{N}_1(0, I^{-1}(\theta))$ следует теперь из представления (447), леммы 3.16, первой сходимости в (431), а также (445). $\square$

Вывод теоремы 3.11 проведем в два этапа. Нам потребуются

Лемма 3.17. *Если выполнено предположение* (E_k) *и функция* $\lambda_k(t, x)$ *является решением системы* (440), *то эта функция удовлетворяет и* (427).

Д о к а з а т е л ь с т в о. Уравнения (427) и (440) соответственно нам удобно переписать в следующем интегральном виде:

$$\int \lambda_k^{(m-1)} f = -\frac{1}{m} \int l^{(m+1)} f, \qquad m = 1, \ldots, k+1, \tag{448}$$

$$\int \lambda_k f^{(m-1)} = \frac{1}{m} \int l' f^{(m)}, \qquad m = 1, \ldots, k+1 \tag{449}$$

(условимся опускать аргументы у функций, а также переменную дифференцирования θ и переменную интегрирования x). Доказательство проведем методом индукции по m. Пусть при $m = 1$ функция λ_k удовлетворяет (449), т.е.

$$\int \lambda_k f = \int l' f'. \tag{450}$$

Дифференцируя тождество $\displaystyle\int l' f = 0$ по θ, имеем: $0 = \displaystyle\int l'' f + \int l' f'$. Следовательно, с учетом (450), $\displaystyle\int \lambda_k f = -\int l'' f$, т.е. выполнено (448) при $m = 1$.

Предположим теперь, что если λ_k удовлетворяет (449) при $m = 1, \dots, h-1$ для некоторого фиксированного $3 \le h \le k+1$, то λ_k удовлетворяет (448) при этих m. Покажем, что если, дополнительно, λ_k удовлетворяет (449) при $m = h$, то справедливо и (448) при $m = h$, т.е.

$$\int \lambda_k^{(h-1)} f = -\frac{1}{h} \int l^{(h+1)} f. \tag{451}$$

Дифференцируем $h-1$ раз равенство (449) при $m = 1$, $h-2$ раза — при $m = 2$ и т.д. (при всех $m = 1, \dots, h$ равенство (449) дифференцируем $h - m$ раз). В итоге, полагая $C_0^0 = 1$, получаем

$$\sum_{i=1}^{h-j} C_{h-j}^i \int \lambda_k^{(i)} f^{(h-1-i)} = \frac{1}{j} \sum_{i=0}^{h-j} C_{h-j}^i \int l^{(i+1)} f^{(h-i)}, \quad j = 1, \dots, h. \tag{452}$$

При выводе (452) мы учли, что в силу (E_k) можно менять порядок дифференцирования и интегрирования в соответствующих интегралах, а также использовали следующее представление для n-ой производной произведения двух функций:

$$(uv)^{(n)} = \sum_{i=0}^{n} C_n^i u^{(i)} v^{(n-i)} \quad \text{при} \qquad C_n^i = \frac{n!}{i!(n-i)!}. \tag{453}$$

Положим

$$x_i = \int \lambda_k^{(i)} f^{(h-1-i)}, \quad \nu_i = \int l^{(i+1)} f^{(h-i)}, \quad b_{h-j} = \frac{1}{j} \sum_{i=0}^{h-j} C_{h-j}^i \nu_i \tag{454}$$

при $i = 0, \dots, h-1$ и $j = 1, \dots, h$. В этом случае (452) примет следующий вид:

$$x_{h-j} + C_{h-j}^{h-j-1} x_{h-j-1} + \ldots + C_{h-j}^2 x_2 + C_{h-j}^1 x_1 + x_0 = b_{h-j}, \quad j = 1, \dots, h, \tag{455}$$

а равенство (451), которое нужно установить, можно переписать в виде

$$x_{h-1} = -\nu_h / h. \tag{456}$$

Нетрудно видеть, что соотношения (455) — это система линейных уравнений (с верхнетреугольной матрицей) относительно $x_0, \dots, x_{h-1}$. Используя обратный ход метода Гаусса и последовательно выражая $x_0 = b_0$, $x_1 = b_1 - b_0$, $x_2 = b_2 - 2b_1 + b_0$ и т.д., из (455) находим представление для x_{h-1} :

$$x_{h-1} = b_{h-1} - C_{h-1}^{h-2} b_{h-2} + C_{h-1}^{h-3} b_{h-3} - \ldots C_{h-1}^1 b_1 \pm b_0. \tag{457}$$

Чтобы теперь из (457) извлечь (456) и завершить доказательство леммы, нам потребуется несколько комбинаторных соотношений. Во-первых, при любом $i = 1, \ldots, h-1$

$$C_h^1 C_{h-1}^i - C_h^2 C_{h-2}^i + C_h^3 C_{h-3}^i - \ldots \pm C_h^{h-i} C_i^i = C_h^0 C_h^i \tag{458}$$

(см., например, в [348], стр.81). Во-вторых, имеют место равенства

$$\sum_{i=0}^{h-1} C_h^i \nu_i = -\nu_h, \qquad \frac{1}{j} C_{h-1}^{h-j} = \frac{1}{h} C_h^j. \tag{459}$$

Второе равенство в (459) нетрудно извлечь непосредственно из определения в (453), а чтобы получить первое равенство в (459), нужно h раз продифференцировать тождество $\displaystyle\int l' f = 0$ по θ, учесть, что в силу предположения (E_k) можно менять порядок дифференцирования и интегрирования, а также использовать соотношение (453) и обозначения (454). Наконец, подставляя значения b_{h-j} в соотношение (457) и используя тождества (458) и (459), имеем

$$\begin{aligned} x_{h-1} &= C_{h-1}^{h-1} \sum_{i=0}^{h-1} C_{h-1}^i \nu_i - \frac{1}{2} C_{h-1}^{h-2} \sum_{i=0}^{h-2} C_{h-2}^i \nu_i + \frac{1}{3} C_{h-1}^{h-3} \sum_{i=0}^{h-3} C_{h-3}^i \nu_i - \ldots \pm \frac{1}{h} \nu_0 = \\ &= \frac{1}{h} \Big(C_h^1 \sum_{i=0}^{h-1} C_{h-1}^i \nu_i - C_h^2 \sum_{i=0}^{h-2} C_{h-2}^i \nu_i + C_h^3 \sum_{i=0}^{h-3} C_{h-3}^i \nu_i - \ldots \pm \nu_0 \Big) = \\ &= \frac{1}{h} \sum_{i=0}^{h-1} \big(\sum_{j=1}^{h-i} (-1)^{j+1} C_h^j C_{h-j}^i \big) \nu_i = \frac{1}{h} \sum_{i=0}^{h-1} C_h^0 C_h^i \nu_i = -\frac{1}{h} \nu_h. \end{aligned}$$

Таким образом, выполнено (456) и доказательство леммы завершено. □

Лемма 3.18. *Если выполнено предположение* (E_k), *то функция* $\lambda_k(\theta, x)$, *определенная равенством* (439), *является решением системы* (440).

Д о к а з а т е л ь с т в о. В силу условий леммы и определения (437), нам заданы линейно независимые функции $e_{0,\theta}(\cdot), \ldots, e_{k,\theta}(\cdot)$, которые будем называть базисом, и известны проекции искомой функции $\lambda_k(\theta, \cdot)$ на элементы этого базиса, определяемые соотношениями (440):

$$\langle \lambda_k(\theta, \cdot), e_{m-1,\theta}(\cdot) \rangle = \frac{1}{m} \langle e_{1,\theta}(\cdot), e_{m,\theta}(\cdot) \rangle \equiv a_{m-1,\theta}, \qquad m = 1, \ldots, k+1. \tag{460}$$

Хорошо известно, что в этом случае функция

$$\lambda_k(\theta, x) = c_{k0,\theta} e_{0,\theta}(x) + c_{k1,\theta} e_{1,\theta}(x) + \ldots + c_{kk,\theta} e_{k,\theta}(x), \tag{461}$$

являющаяся линейной комбинацией элементов базиса со специально подобранными коэффициентами $c_{kj,\theta}$, $j = 0, \ldots, k$, удовлетворяет системе (460). Чтобы определить эти коэффициенты, домножим скалярно равенство (461) на элементы базиса $e_{0,\theta}(\cdot), \ldots, e_{k,\theta}(\cdot)$. Учитывая еще соотношения (460), получаем следующую линейную систему $(k+1)$-го уравнения относительно $(k+1)$-ой неизвестной $c_{kj,\theta},\ j=0,\dots,k$:

$$c_{k0,\theta}\langle e_{0,\theta}(\cdot), e_{j,\theta}(\cdot)\rangle + \ldots + c_{kk,\theta}\langle e_{k,\theta}(\cdot), e_{j,\theta}(\cdot)\rangle = a_{j,\theta}, \quad j=0,\dots,k. \tag{462}$$

При этом матрица с элементами $\langle e_{i,\theta}(\cdot), e_{j,\theta}(\cdot)\rangle$, $i,j=0,\dots,k$, определяющая эту систему линейных уравнений (матрица Грама, построенная по элементам базиса) положительно определена ввиду линейной независимости этих элементов.

В силу равенств $l'(\theta,x)=f'_\theta(x)/f_\theta(x)$, $l''(\theta,x)=f''_\theta(x)/f_\theta(x)-\big(f'_\theta(x)/f_\theta(x)\big)^2$ и условия (B) получаем, что $\int f'_\theta(x)\mu(dx)=0$ и $\int f''_\theta(x)\mu(dx)=0$. А потому, с учетом (436) и (437),

$$\langle e_{1,\theta}(\cdot), e_{1,\theta}(\cdot)\rangle \equiv a_{0,\theta} = I(\theta), \qquad \langle e_{0,\theta}(\cdot), e_{0,\theta}(\cdot)\rangle \equiv \int f_\theta(x)\mu(dx) = 1,$$
$$\langle e_{j,\theta}(\cdot), e_{0,\theta}(\cdot)\rangle = \langle e_{0,\theta}(\cdot), e_{j,\theta}(\cdot)\rangle = \int f_\theta^{(j)}(x)\mu(dx) = 0 \quad \text{при} \quad j\neq 0.$$

Следовательно, у матрицы Грама, определяющей систему уравнений (462), все элементы первой строки и первого столбца, за исключением диагонального — нулевые. А потому $c_{k0,\theta}=I(\theta)$ для любых k, а функция $\lambda_k(\theta,x)$, определенная в (438) и (439), является решением (440). □

Д о к а з а т е л ь с т в о теоремы 3.11. Это утверждение немедленно следует из лемм 3.17 и 3.18. □

3.4 Результаты компьютерного моделирования

3.4.1 Предварительные сведения

Рассмотрим некоторые приложения полученных в главах 2 и 3 результатов, связанные с одношаговыми оценками в задачах нелинейной регрессии. В этом разделе считаем, что наблюдения X_i представимы в виде

$$X_i = f_i(\theta)+\varepsilon_i \quad \text{при } f_i(\theta)=f(\theta,\mathbf{z}_i), \quad \mathbb{E}\varepsilon_i=0,\ 0<\mathbb{E}\varepsilon_i^2<\infty, \quad i=1,\dots,n, \tag{463}$$

где f — известная функция, набор $\{\mathbf{z}_i\}$ (скаляров или векторов) известен, погрешности $\{\varepsilon_i\}$ независимы, параметр θ одномерный и подлежит оцениванию. Для определенности и простоты будем рассматривать классический вариант модели (463), когда регрессоры детерминированы. Все приводимые далее рассуждения можно перенести и на модели со случайными регрессорами. Например, оценки метода наименьших квадратов в ситуации случайных регрессоров исследовались в [165], [130], [230], [232], [265], [283].

1. Пусть при всех i выполнено $\mathbb{E}\varepsilon_i^2 = \sigma^2 w_i^{-1}$, где числа $\{w_i\}$ известны, а параметр σ^2 может быть неизвестным. В этих условиях оценка взвешенного метода наименьших квадратов определяется соотношением

$$\widetilde{\theta}_n = \arg\min_t \sum_{i=1}^n w_i \big(X_i - f_i(t)\big)^2$$

и нас интересует тот регулярный случай, когда она удовлетворяет уравнению

$$\sum_{i=1}^n \psi_i(t, X_i) = 0 \qquad \text{при} \quad \psi_i(t, X_i) = w_i f_i'(t)(X_i - f_i(t)). \tag{464}$$

В случае $w_i = 1$ при всех i получаем классический метод наименьших квадратов.

Если регрессионная функция f из (463) линейна по первому аргументу, то оценка взвешенного метода наименьших квадратов имеет явное представление и в известном смысле оптимальна. В частности, является наилучшей линейной несмещенной оценкой, а в случае нормально распределенных ошибок — наилучшей и в классе всех несмещенных оценок (см., например, [60], [68], [74], [252], [295]). Эти свойства оценок не переносятся автоматически на случай нелинейной регрессионной функции f: в общем случае оценки могут не существовать, могут иметь смещение, оценка может быть не единственной и т.д. Тем не менее, при выполнении некоторых условий регулярности оценки взвешенного метода наименьших квадратов в случае нелинейной f все же обладают нужными свойствами по-крайней мере в асимптотике (см. подробности и библиографические ссылки, например, в [10], [60], [140], [236], [253], [295], [310], [328]). В частности, в случае нормально распределенных наблюдений оценки взвешенного метода наименьших квадратов и метода максимального правдоподобия совпадают и асимптотическая дисперсия $I_{n,\theta} J_{n,\theta}^{-2}$ оценок обратно пропорциональна суммарной информации Фишера, соответствующей $\{X_i\}$:

$$I_{n,\theta} J_{n,\theta}^{-2} = \sigma^2 \left(\sum_{i=1}^n w_i \big(f_i'(\theta)\big)^2 \right)^{-1} \equiv \left(\sum_{i=1}^n I_i(\theta) \right)^{-1}, \tag{465}$$

где $I_i(\theta)$ — информация Фишера, соответствующая наблюдению X_i. Таким образом, в силу обобщения неравенства Рао-Крамера на случай неоднородных наблюдений (см., например, [325]), можно сделать вывод о соответствующей точности оценки взвешенного метода наименьших квадратов в случае нормально распределенных наблюдений.

Подчеркнем еще, что в случае нелинейной регрессионной функции f оценка взвешенного метода наименьших квадратов, как правило, не может быть найдена в явном виде. В частности, уравнение (464) обычно не разрешается аналитически ([10], [60], [236], [253], [295], [310]). Существует ряд приближенных численных методов решения уравнения вида (464) и поиска минимума соответствующей функции. Выбор того или иного метода зависит от многих факторов, в том числе от вида регрессионной функции f. Напомним, что для задач нелинейной регрессии весьма типична ситуация наличия локальных экстремумов, а потому при неудачном выборе начального приближения параметра итерационный процесс обнаруживает лишь локальный минимум, ближайший к этой стартовой точке и при приближенном поиске оценки взвешенного метода наименьших квадратов возникает большинство проблем численного анализа: проблема поиска стартовой точки (особенно в случае большого числа локальных минимумов у изучаемой функции), проблема сходимости того или иного метода, проблема выбора момента прекращения итерационного процесса в связи с достижением требуемой точности и др. Подчеркнем, что задача поиска глобального экстремума функции в общем случае до настоящего времени не решена, сложность этой задачи общепризнана и решению подобных задач в тех или иных условиях посвящена обширная литература. В связи с вышесказанным отметим, например, работы [60], [68], [105], [248], [236], [253], [266], [267], [268], [272], [277], [327], [329]). Таким образом, в случае нелинейной регрессионной функции имеется некоторая проблема, касающаяся вычислительной стороны: как отыскать $\widetilde{\theta}_n$, удовлетворяющую (464).

Одношаговые процедуры позволяют избежать указанных выше вычислительных сложностей. В частности, при выполнении условий соответствующих утверждений (см. теоремы 3.1 и 3.2 из раздела 3.1.1) одношаговые M-оценки

$$\theta_n^{**} = \theta_n^* + \sum_{i=1}^n w_i(X_i - f_i(\theta_n^*))f_i'(\theta_n^*)\left(\sum_{i=1}^n w_i\big[(f_i'(\theta_n^*))^2 - (X_i - f_i(\theta_n^*))f_i''(\theta_n^*)\big]\right)^{-1}, \qquad (466)$$

$$\widehat{\theta}_n^{**} = \theta_n^* + \sum_{i=1}^n w_i(X_i - f_i(\theta_n^*))f_i'(\theta_n^*)\left(\sum_{i=1}^n w_i(f_i'(\theta_n^*))^2\right)^{-1} \qquad (467)$$

асимптотически нормальны с асимптотической дисперсией

$$I_{n,\theta}J_{n,\theta}^{-2} \equiv \sum_{i=1}^n \mathbb{E}\psi_i^2(\theta, X_i)\left(\sum_{i=1}^n \mathbb{E}\psi_i'(\theta, X_i)\right)^{-2} = \sigma^2\left(\sum_{i=1}^n w_i\big(f_i'(\theta)\big)^2\right)^{-2}, \qquad (468)$$

т.е. асимптотически имеют точность оценки взвешенного метода наименьших квадратов (см. (465) и (468)). Здесь в определениях (315) и (320) одношаговых оценок мы использовали ψ_i из (464) и выбрали $J_n^* = -\sum\limits_{i=1}^n w_i\big(f_i'(\theta_n^*)\big)^2$ в качестве оценки для $J_{n,\theta}$.

При выполнении условий теоремы 3.7 (см. раздел 3.2.1) асимптотически ту же точность имеет и следующая одношаговая взвешенная M-оценка

$$\theta_{n,h}^{**} = \theta_n^* + \sum_{i=1}^n w_i(X_i - f_i(\theta_n^*))f_i'(\theta_n^*)\left(\sum_{i=1}^n w_i(f_i'(\theta_n^*))^2\right)^{-1}, \qquad (469)$$

совпадающая с оценкой $\widehat{\theta}_n^{**}$ из (467). Здесь в определении (362) мы положили $h_i(t) = f_i'(t)$ и $\psi_i(t, X_i) = w_i(X_i - f_i(t))$.

В завершение обсуждения приведем пример неоднозначного определения одношаговой оценки $\widehat{\theta}_n^{**}$: все зависит от выбора оценки J_n^* для $J_{n,\theta} \equiv \sum\limits_{i=1}^n \mathbb{E}\psi_i'(\theta, X_i)$, участвующей в построении $\widehat{\theta}_n^{**}$.

Пример 3.4. Пусть для любого $i = 1, \ldots, n$ регрессионная модель (463) задается соотношениями $f_i(\theta) = \sqrt{1 + z_i\theta}$ и $\mathbb{E}\varepsilon_i^2 = \sigma^2$ при всех i, где $\theta > 0$ и числа $\{z_i > 0\}$ известны. Нас интересуют одношаговые M-оценки $\widehat{\theta}_n^{**}$, асимптотически имеющие точность оценки метода наименьших квадратов. Отметим, что эта модель с точки зрения трудности вычисления оценки метода наименьших квадратов упомянута в монографии [329]. Для функций $\psi_i(\theta, X_i) = f_i'(\theta)(X_i - f_i(\theta))$ имеем $\psi_i'(\theta, X_i) = f_i''(\theta)\varepsilon_i - (f_i'(\theta))^2 = f_i''(\theta)X_i - \big(f_i'(\theta)f_i(\theta)\big)_\theta'$, а потому

$$\mathbb{E}\psi_i'(\theta, X_i) = -(f_i'(\theta))^2 = f_i''(\theta)f_i(\theta) - \big(f_i'(\theta)f_i(\theta)\big)_\theta'.$$

Но для данной модели $f_i'(\theta)f_i(\theta)$ не зависит от θ. Следовательно в определении (320) одношаговой M-оценки $\widehat{\theta}_n^{**}$ можно один из следующих вариантов: $J_n^* = -\sum_{i=1}^n (f_i'(\theta_n^*))^2$, $J_n^* = \sum_{i=1}^n f_i''(\theta_n^*)\big(rf_i(\theta_n^*) - (r-1)X_i\big)$ или $J_n^* = \sum_{i=1}^n f_i''(\theta_n^*)\big(rX_i - (r-1)f_i(\theta_n^*)\big)$ при любом $r \in \mathbb{R}$. В первом случае мы получаем оценку $\widehat{\theta}_n^{**}$ из (467) при $w_i = 1$ для всех i. Отметим еще, что в качестве θ_n^* можно выбрать оценку из примера 2.2 в разделе 2.1. □

2. Важную роль для приложений играют неоднородные регрессионные модели, в которых дисперсии наблюдений зависят не только от номера наблюдения, но и от основного неизвестного параметра (см., например, [3], [34], [122], [128], [143], [270], [281]). Таким образом, в дополнение к (463) в этом подпункте считаем, что

$$\mathbb{E}\varepsilon_i^2 = \sigma^2/w_i(\theta), \qquad i = 1, \ldots, n, \tag{470}$$

где $\{w_i(\cdot)\}$ — известные функции, а параметр σ^2 может быть неизвестным.

Хорошо известно, что взвешенный метод наименьших квадратов не обеспечивает здесь нужной точности. Дело в том, что оптимальные веса этого метода должны в данной ситуации зависеть от основного параметра, но если веса, определяющие оценку взвешенного метода наименьших квадратов, зависят от θ, то в общем случае оценка оказывается не состоятельной (см., например, [122], [347]). Один из подходов оценивания в данной ситуации — метод квазиправдоподобия (см. [38], [122], [292]). Для модели (463) с условием (470) оценки квазиправдоподобия есть M-оценки, определяемые уравнением

$$\sum_{i=1}^n \psi_i(t, X_i) = 0 \qquad \text{при} \quad \psi_i(t, X_i) = w_i(t)f_i'(t)(X_i - f_i(t)). \tag{471}$$

Эти оценки являются наилучшими в классе M-оценок, являющихся решениями по t уравнений вида $\sum\limits_{i=1}^{n} h_i(t)(X_i - f_i(t)) = 0$ (см., например, [122], а также предложение 3.3 в разделе 3.2.3). Обсуждения, подробную библиографию и исследование асимптотических свойств оценок такого типа можно найти, например, в [8], [122], [301]. В частности, при широких ограничениях оценки $\widetilde{\theta}_n$ асимптотически нормальны с асимптотической дисперсией

$$\sigma^2 \left(\sum_{i=1}^{n} w_i(\theta) \big(f_i'(\theta)\big)^2 \right)^{-2}.$$

Отметим, что термин «квазиправдоподобие» отчасти связан с тем, что для функций

$$\psi_i(\theta, X_i) \equiv w_i(\theta) f_i'(\theta)(X_i - f_i(\theta))/\sigma^2$$

имеют место свойства, аналогичные свойствам производной логарифмической плотности:

$$\mathbb{E}\psi_i(\theta, X_i) = 0, \quad \mathbb{E}\psi_i^2(\theta, X_i) = w_i(\theta)\big(f_i'(\theta)\big)^2/\sigma^2, \quad -\mathbb{E}\partial\psi_i(\theta, X_i)/\partial\theta = w_i(\theta)\big(f_i'(\theta)\big)^2/\sigma^2.$$

Уместно отметить, что M-оценка $\widetilde{\theta}_n$, определяемая уравнением (471), не является оценкой метода наименьших квадратов.

Другой подход к оцениванию, все же основанный на взвешенном методе наименьших квадратов, состоит в том, чтобы вначале оценить неизвестные оптимальные веса $\{w_i(\theta)\}$, зависящие от θ, а затем использовать взвешенный метод наименьших квадратов с *известными* весами. Другими словами, оценку $\widetilde{\theta}_n$ для параметра θ предлагается искать в виде

$$\widetilde{\theta}_n = \arg\min_{\theta} \sum_{i=1}^{n} w_i(\theta_n^*)\big(X_i - f_i(\theta)\big)^2, \tag{472}$$

где θ_n^* — некоторая предварительная оценка для θ (например, оценка взвешенного метода наименьших квадратов с *некоторыми* известными весами). Подобная процедура оценивания обсуждается, например, в [34], [128]. Тем самым, в интересующем нас регулярном случае оценка $\widetilde{\theta}_n$ из (472) связана со взвешенной M-оценкой, определяемой уравнением

$$\sum_{i=1}^{n} h_i(\theta_n^*)\psi_i(t, X_i) = 0 \qquad \text{при } h_i(t) = w_i(t) \text{ и } \psi_i(t, X_i) = f_i'(t)\big(X_i - f_i(t)\big).$$

Понятно, что оба указанных подхода задают оценки, имеющие одну и ту же асимптотическую точность, при этом оба подхода связаны с вычислительными проблемами, которые затрагивались выше. Проблема поиска оценок квазиправдоподобия, особенно в случае существования нескольких корней уравнения (471), упоминается, например, в [122] и [266].

Одношаговые процедуры решают указанную проблему. В частности, мы можем использовать следующую одношаговую M-оценку

$$\widehat{\theta}_n^{**} = \theta_n^* + \sum_{i=1}^n w_i(\theta_n^*) f_i'(\theta_n^*)(X_i - f_i(\theta_n^*)) \left(\sum_{i=1}^n w_i(\theta_n^*) \big(f_i'(\theta_n^*)\big)^2 \right)^{-1}, \tag{473}$$

при широких ограничениях асимптотически эквивалентную оценке квазиправдоподобия. Наряду с $\widehat{\theta}_n^{**}$ мы можем использовать одношаговую M-оценку θ_n^{**} и одношаговые взвешенные M-оценки $\theta_{n,h}^{**}$, но уже при несколько иных ограничениях.

В качестве примера приведем простое следствие из теоремы 3.2 об асимптотической нормальности одношаговой M-оценки $\widehat{\theta}_n^{**}$ из (473).

Следствие 3.4. *Пусть выполнено* (463), (470) *и* $\varepsilon_i = \sigma\xi_i/\sqrt{w_i(\theta)}$ *при всех* i, *где* $\{\xi_i\}$ *независимы и одинаково распределены. Предположим, что при всех* i *функции* $w_i(t)$ *дважды непрерывно дифференцируемы, а функции* $f_i(t)$ *трижды непрерывно дифференцируемы. Кроме того, оценка* θ_n^* *является* $\sqrt{n}$*-ограниченной и*

$$\inf_i \min\{w_i(\theta), |f_i'(\theta)|\} > 0, \qquad \sup_i \max\{w_i(\theta), |w_i'(\theta)|, |f_i(\theta)|, |f_i'(\theta)|, |f_i''(\theta)|\} < \infty,$$
$$\sup_i \sup_{|t-\theta|\le\delta} \max\{|w_i''(t)|,\ |f_i'''(t)|\} < \infty$$

при всех достаточно малых δ. *Тогда для одношаговой* M*-оценки* $\widehat{\theta}_n^{**}$, *определенной в* (473), *справедливо соотношение* $\sigma^{-1}\left(\sum\limits_{i=1}^n w_i(\theta)(f_i'(\theta))^2\right)^{1/2}(\widehat{\theta}_n^{**} - \theta) \xrightarrow{d} \mathcal{N}_1(0,1)$.

З а м е ч а н и е 3.19. Нетрудно показать, что в условиях следствия выполнены предположения (A) и (321) при $\psi_i(t, X_i) = w_i(t) f_i'(t)(X_i - f_i(t))$. При этом для того, чтобы проверить условие Линдеберга, можно воспользоваться результатом Гаека-Шидака (см., например, [326, гл. V, § 1.2] или [151]). Таким образом, приведенное следствие является частным случаем теоремы 3.2. □

3.4.2 Приближение оценок квазиправдоподобия

Генерируются выборки $\{X_i;\, i \le n\}$ объема $n = 100$ нормально распределенных случайных величин со структурой (463) и условием $\mathbb{E}\varepsilon_i^2 = \sigma^2/w_i(\theta)$ при всех i. Параметры регрессионных моделей и результаты вычислений для трех различных выборок $\{X_i;\, i \le 100\}$ приведены в таблице 1. Для каждой из моделей, приведенной в таблице 1, оценки θ_n^* и $\widehat{\theta}_n^{**}$ из (473) вычисляются для 1000 различных выборок $\{X_i;\, i \le 100\}$. Предварительные оценки θ_n^* вычисляются с использованием результатов главы 1 (см. формулу (291) из раздела 2.3.4, а также раздел 2.1).

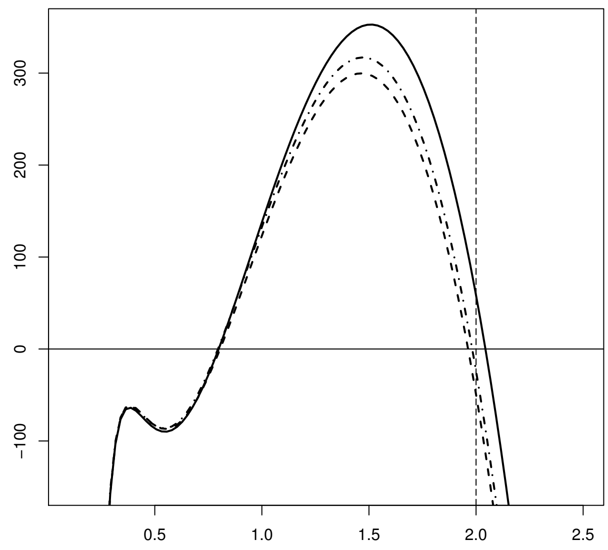
Регрессионные модели, приведенные в таблице 1, снабжены поясняющими иллюстрациями (см. рисунки 25–30). На этих рисунках графики соответствующих функций построены в некоторой окрестности истинного значения θ, обозначенного вертикальной линией. На рисунках 25 и 26 для моделей 1-8 из таблицы 1 приведены графики функции $\psi_n(t)$ из (342) при $\psi_i(t, X_i) = w_i(t) f_i'(t)(X_i - f_i(t))$ для трех независимых реализаций выборки $\{X_i;\, i \le 100\}$. На рисунках 27-29 приведены графики эмпирических функций распределения для θ_n^* и $\widehat{\theta}_n^{**}$, построенные по 1000 независимым выборкам. Рисунок 30 иллюстрирует поведение функций

$$\psi_n(x) = \sum\nolimits_{i=1}^n w_i(x) f_i'(x)(X_i - f_i(x)) \quad \text{и} \quad \psi_n^*(\cdot) = \sum\nolimits_{i=1}^n w_i(\theta_n^*) f_i'(\theta_n^*)(X_i - f_i(x))$$

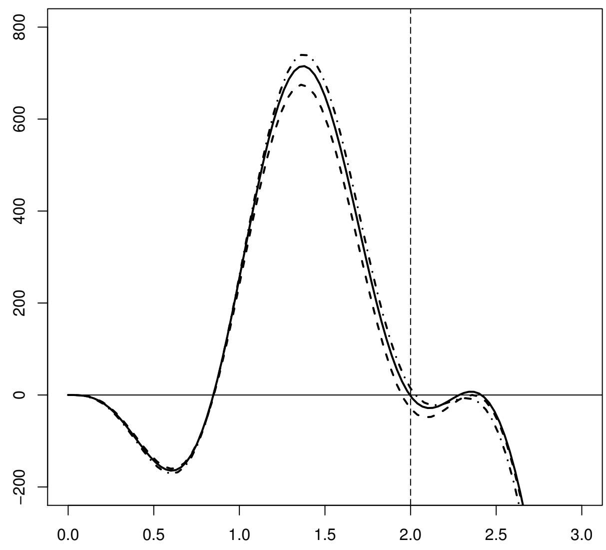
для одной реализации выборки в модели 6.

Таблица 1: Метод квазиправдоподобия. Результаты компьютерного моделирования при использовании одношаговых процедур для приближения оценок квазиправдоподобия.

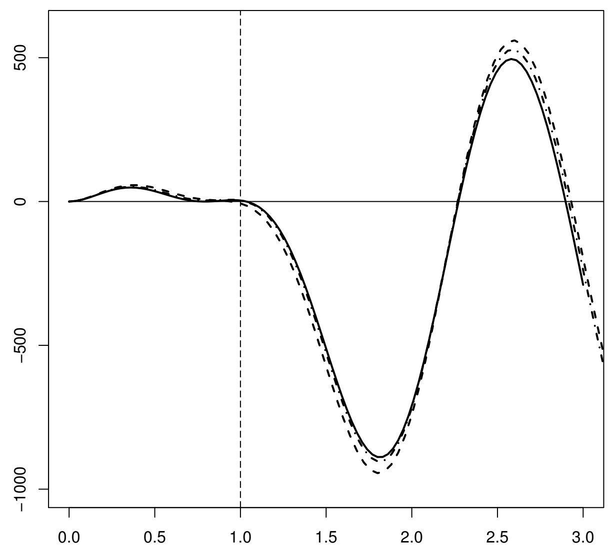
	$f_i(t)$	$\sigma^2/w_i(\theta) \equiv \mathbb{E}\varepsilon_i^2$	$z_i,\ r_i$	θ	θ_n^*	$\widehat{\theta}_n^{**}$
1.	$z_i t + r_i \log(t)$	$(1 + \theta z_i + r_i \log(\theta))^{-2}$	$z_i = 0.5 + \frac{i}{n}$ $r_i = 1.5 - \frac{i}{n}$	2	1.895 2.381 1.746	2.067 1.966 2.182
2.	$z_i t + r_i \sin(2t)$	$0.2\big(\theta z_i + r_i \sin(\theta)\big)^{-2}$	$z_i = \frac{i}{n}$ $r_i = 1.7 - \frac{i}{n}$	2	1.539 2.216 1.749	1.922 1.915 1.958
3.	$z_i t + r_i \sin(2t)$	$0.2\big(\theta z_i + r_i \sin(\theta)\big)^{-2}$	$z_i = \frac{i}{n}$ $r_i = 1.7 - \frac{i}{n}$	1	1.509 1.860 0.829	1.138 0.973 1.056
4.	$z_i t + r_i \exp(-t)$	$0.1(z_i\theta + r_i \exp(-\theta))^2$	$z_i = \frac{i}{n}$ $r_i = 2 - \frac{i}{n}$	5	8.164 3.743 6.297	5.009 4.657 4.946
5.	z_i^t	$0.1(1 + z_i^\theta)^{-1}(1.05 + z_i \cos(z_i\theta))^2$	$z_i = \frac{i}{n}$	0.3	0.259 0.367 0.404	0.305 0.302 0.315
6.	$z_i t + r_i t^{-1/5}$	$1/(z_i\theta + r_i\theta^{-1/5})$	$z_i = \frac{i}{n}$ $r_i = 2 - \frac{i}{n}$	3	1.804 3.941 2.074	3.204 2.928 3.017
7.	$\big(1 + \exp\{-z_i t\}\big)^{-1}$	$0.1\big(\exp\{-z_i\theta\}(1 + \exp\{-z_i\theta\})\big)^2$	$z_i = i/n$	3	2.834 3.299 2.667	2.955 3.036 2.959
8.	$z_i t + r_i \cos(t)$	$0.4(1 + \theta z_i + r_i \cos(\theta))^{-2}$	$z_i = i/n$ $r_i = 2 - i/n$	1	0.662 0.726 1.589	1.065 1.030 0.973
9.	$\big(1 + \exp\{-z_i t\}\big)^{-1}$	$0.2(\exp\{-z_i t\})^2$	$z_i = \frac{i}{n}$	1	0.608 1.233 1.507	1.010 0.935 1.197
10.	$(1 + z_i t)^{0.7}$	$0.4(1 + z_i\theta)^{-1.4}$	$z_i = \frac{i}{n}$	0.5	0.730 0.553 0.395	0.581 0.495 0.452
11.	$\log(1 + z_i t)$	$0.3\big(\log(1 + z_i\theta)\big)^{-2}$	$z_i = \frac{i}{n}$	5	7.541 3.586 7.167	5.304 4.992 4.919
12.	$z_i t + r_i t^{-1/2}$	$1/(z_i\theta + r_i\theta^{-1/2})$	$z_i = \frac{i}{n}$ $r_i = 2 - \frac{i}{n}$	3	3.726 2.583 2.709	2.953 3.054 3.011



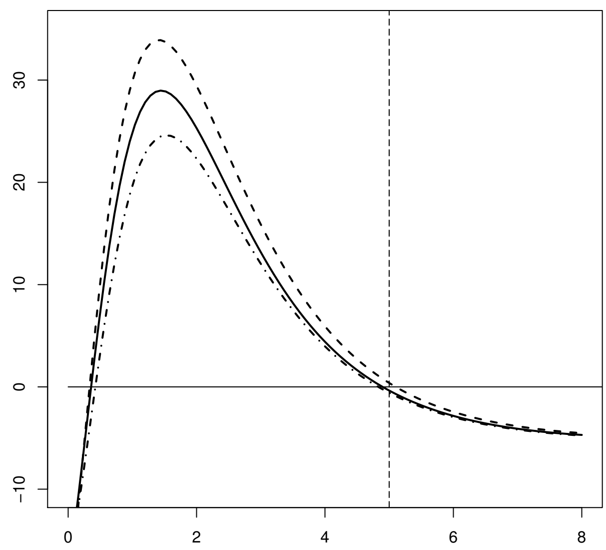
Модель 1



Модель 2

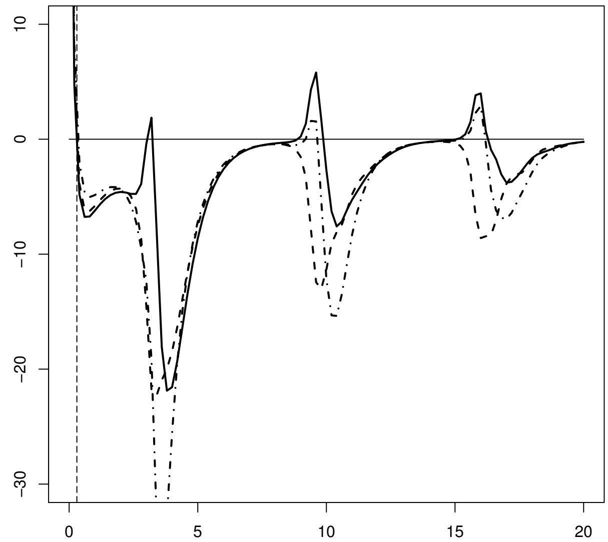


Модель 3

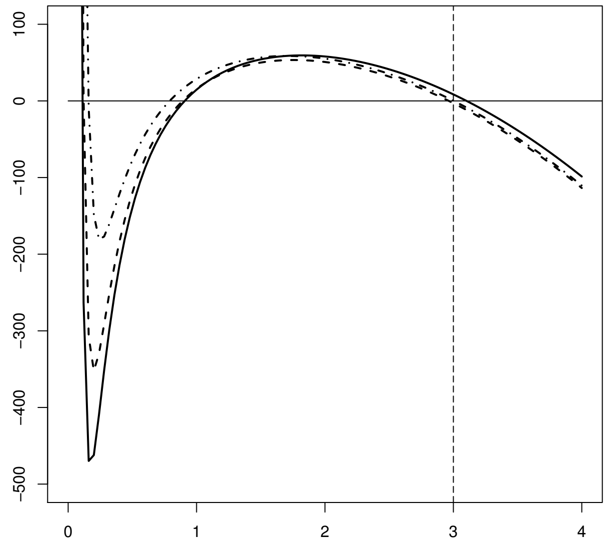


Модель 4

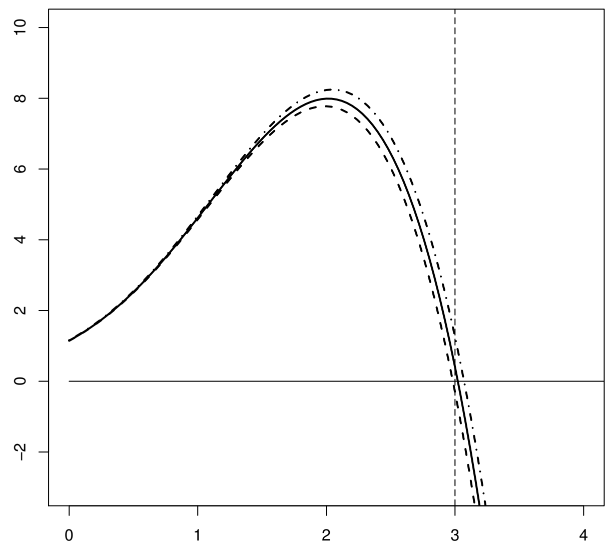
Рис. 25: Метод квазиправдоподобия. Для моделей 1-4 из таб. 1 иллюстрируется поведение функции $\psi_n(x) = \sum_{i=1}^{n} w_i(x) f_i'(x)(X_i - f_i(x))$ из (342) в окрестности истинного значения параметра при трех различных реализациях выборки.



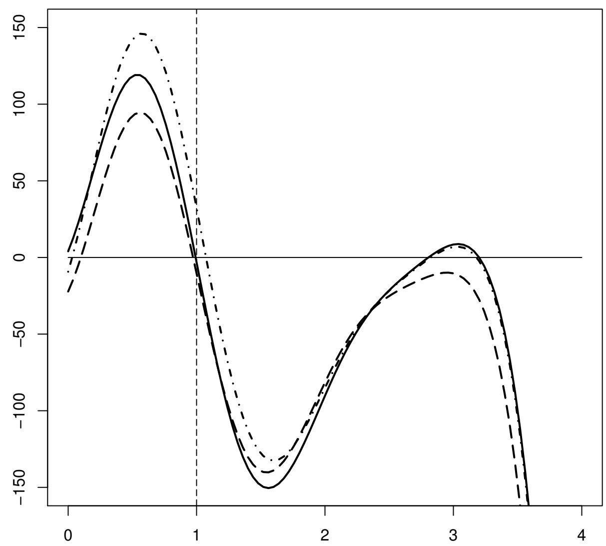
Модель 5



Модель 6

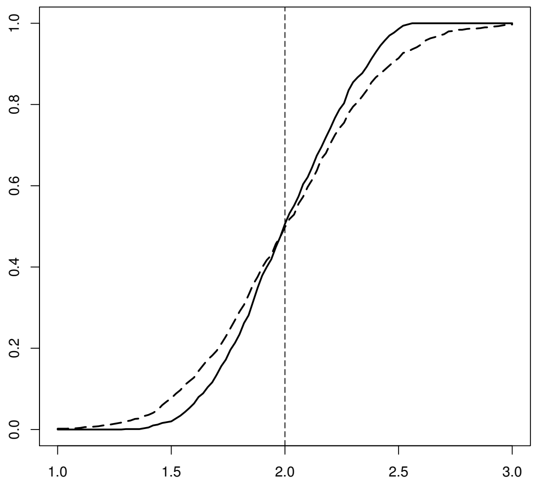


Модель 7

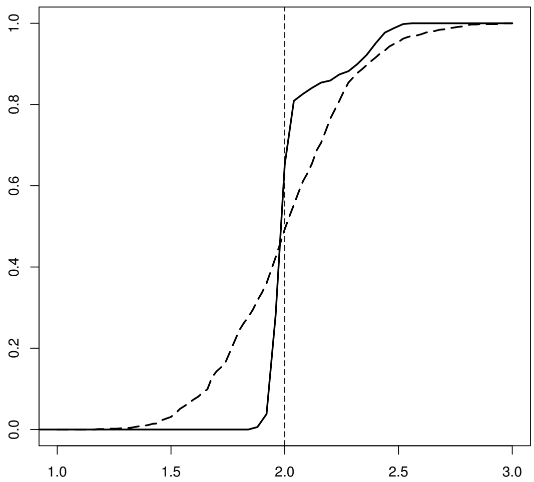


Модель 8

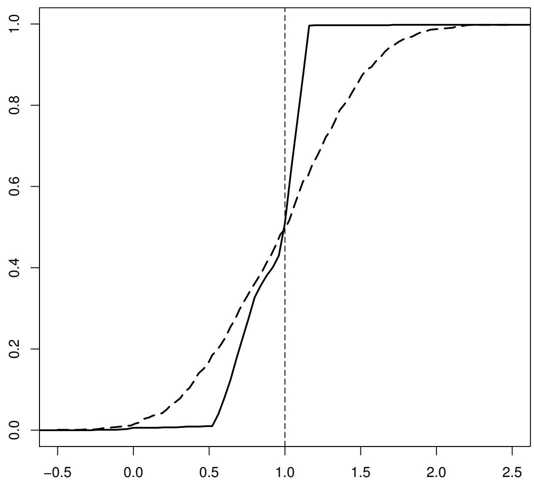
Рис. 26: Метод квазиправдоподобия. Для моделей 5-8 из таб. 1 иллюстрируется поведение функции $\psi_n(x) = \sum_{i=1}^{n} w_i(x) f_i'(x)(X_i - f_i(x))$ из (342) в окрестности истинного значения параметра при трех различных реализациях выборки.



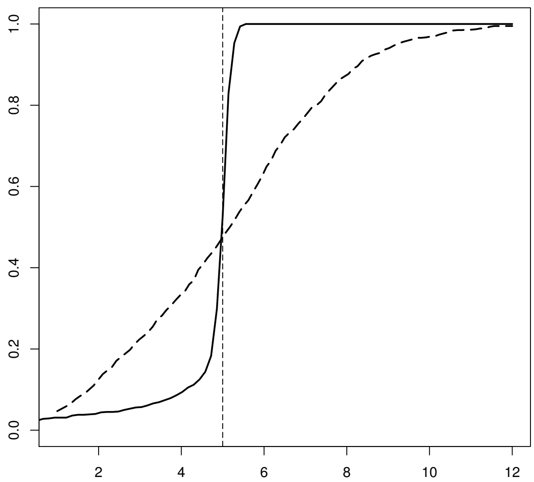
Модель 1



Модель 2

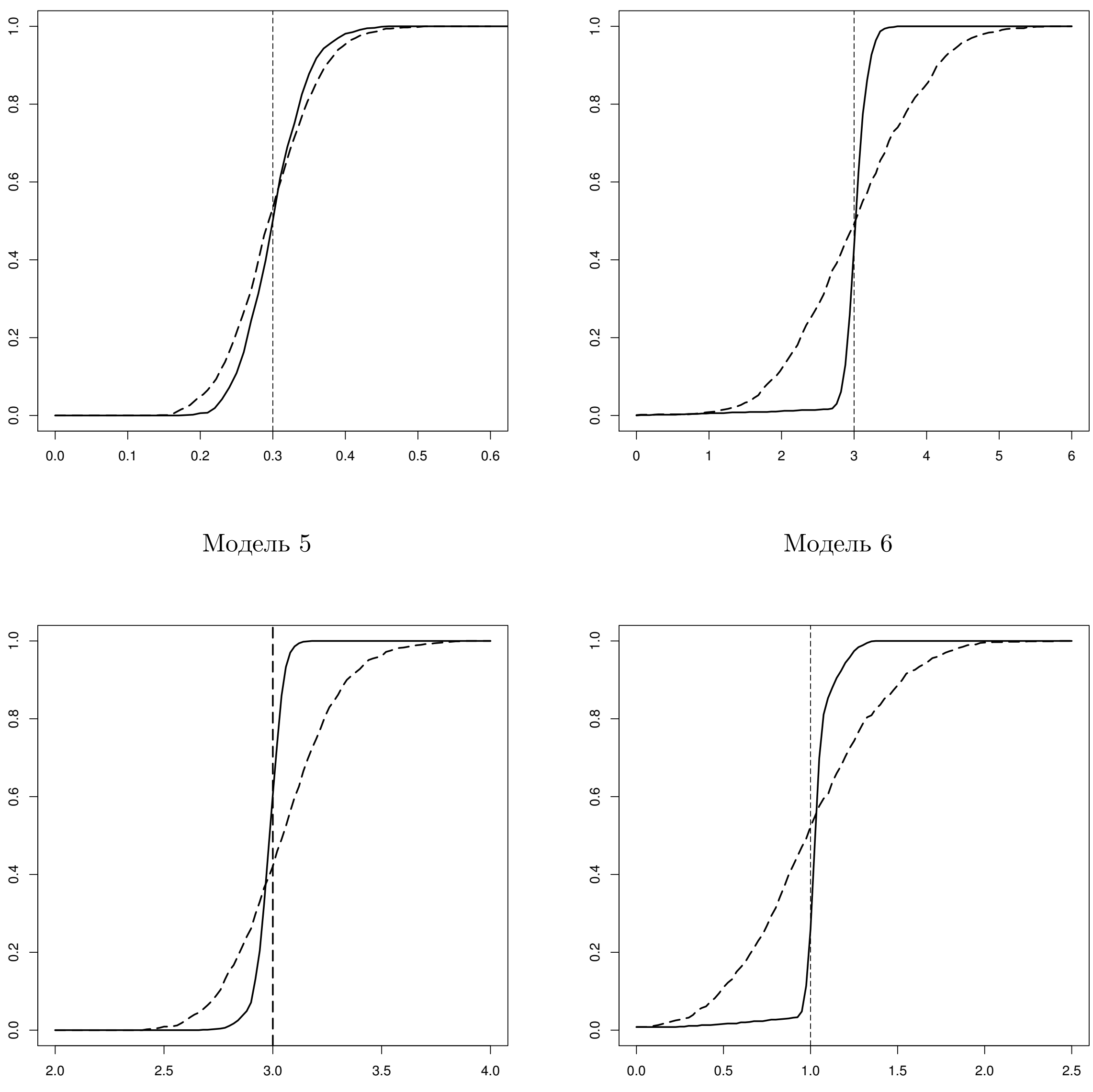


Модель 3



Модель 4

Рис. 27: Одношаговая аппроксимация оценок квазиправдоподобия. Для моделей 1-4 из таблицы таб. 1 представлены эмпирические функции распределения для θ_n^* (пунктирная линия) и $\widehat{\theta}_n^{**}$ (сплошная линия), построенные по 1000 независимых выборок.



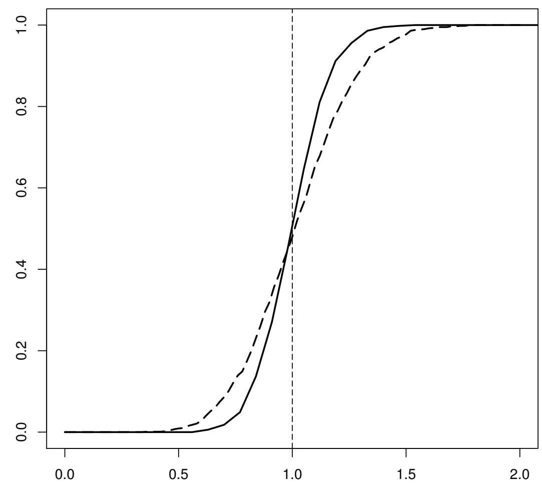
Модель 5

Модель 6

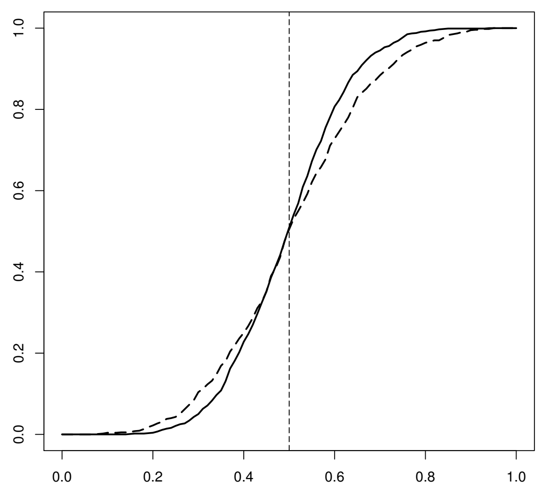
Модель 7

Модель 8

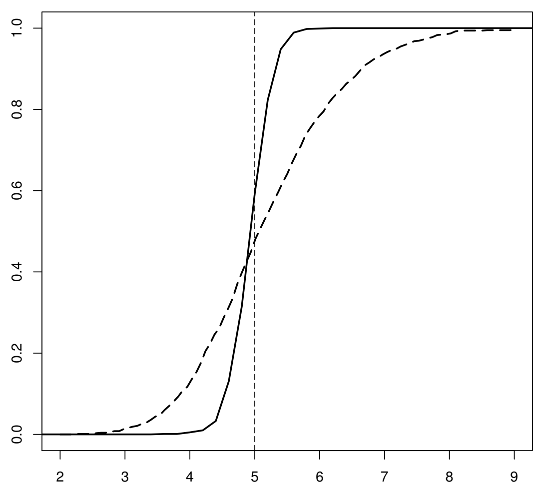
Рис. 28: Одношаговая аппроксимация оценок квазиправдоподобия. Для моделей 5-8 из таблицы таб. 1 представлены эмпирические функции распределения для θ_n^* (пунктирная линия) и $\widehat{\theta}_n^{**}$ (сплошная линия), построенные по 1000 независимых выборок.



Модель 9

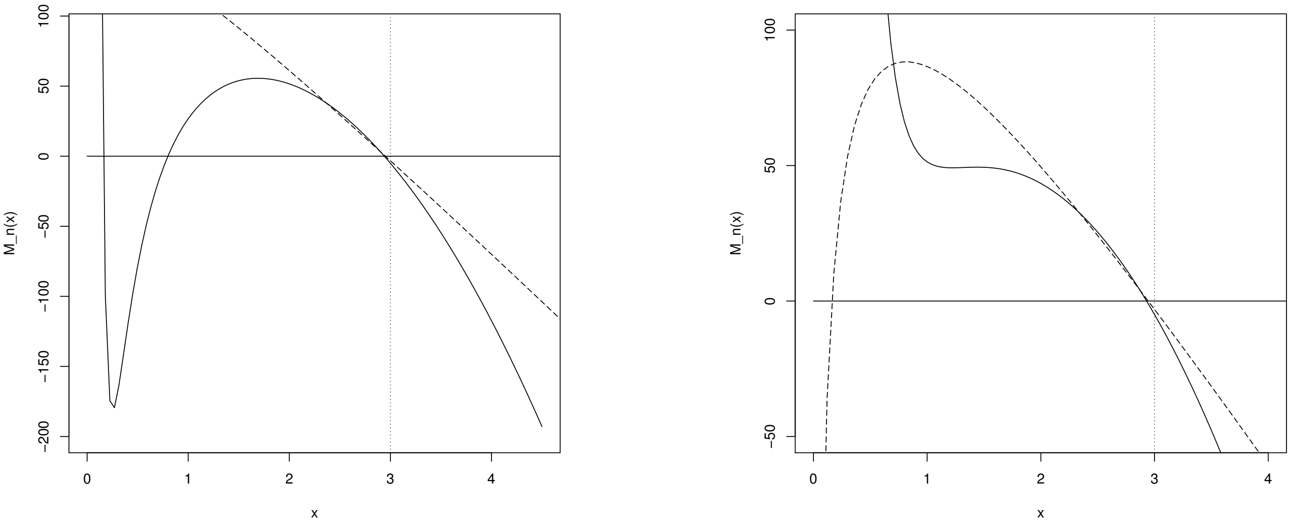


Модель 10



Модель 11

Рис. 29: Одношаговая аппроксимация оценок квазиправдоподобия. Для моделей 9-11 из таблицы таб. 1 представлены эмпирические функции распределения для θ_n^* (пунктирная линия) и $\widehat{\theta}_n^{**}$ (сплошная линия), построенные по 1000 независимых выборок.



а) модель 6 из таблицы 1

б) модель 12 из таблицы 1

Рис. 30: Метод квазиправдоподобия. Иллюстрируется поведение функций $\psi_n(x) = \sum_{i=1}^n w_i(x) f_i'(x)(X_i - f_i(x))$ (сплошная линия) и $\psi_n^*(\cdot) = \sum_{i=1}^n w_i(\theta_n^*) f_i'(\theta_n^*)(X_i - f_i(x))$ (пунктирная линия) в окрестности истинного значения параметра $\theta = 3$.

3.4.3 Приближение оценок метода наименьших квадратов

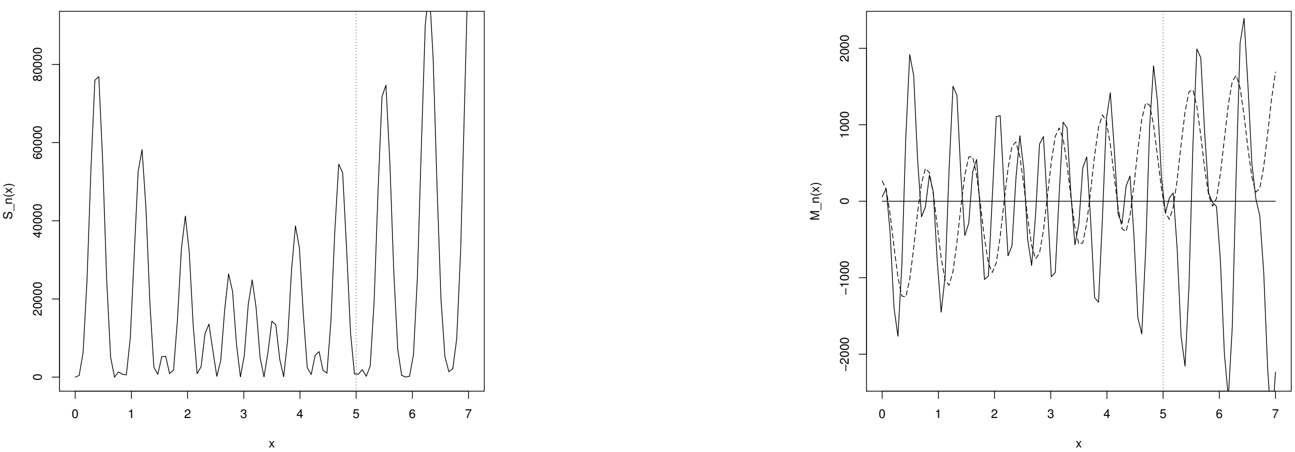
Генерируются выборки $\{X_i;\, i \le n\}$ объема $n = 100$ нормально распределенных случайных величин со структурой (463) и условием $\mathbb{E}\varepsilon_i = \sigma^2$ при всех i. Параметры регрессионных моделей и результаты вычислений для трех различных выборок $\{X_i;\, i \le 100\}$ приведены в таблице 2. Предварительные оценки θ_n^* вычисляются с использованием результатов главы 1 (см. формулу (291) из раздела 2.3.4, а также раздел 2.1). В таблице 2 приведены значения двух типов одношаговых оценок: одношаговой M-оценки θ_n^{**}, определенной в (466), и одношаговой взвешенной M-оценки $\theta_{n,h}^{**}$, определенной в (469) (всюду $w_i \equiv 1$). На рисунке 31 для модели 13 из таблицы 2 для одной реализации выборки приведены графики функций

$$S_n(\cdot) = \sum_{i=1}^n (X_i - f_i(x))^2 \quad \text{и} \quad \psi_n(\cdot) = \sum_{i=1}^n f_i'(x)(X_i - f_i(x)).$$

Графики указанных функций приведены в окрестности истинного значения параметра θ, обозначенного вертикальной линией.

Таблица 2: Метод наименьших квадратов. Результаты компьютерного моделирования при использовании одношаговых процедур для приближения оценок метода наименьших квадратов.

	$f_i(t)$	$\sigma^2 \equiv \mathbb{E}\varepsilon_i^2$	$z_i,\ r_i$	θ	θ_n^*	$\theta_{n,h}^{**}$	θ_n^{**}
					4.862	5.001	5.117
13.	$z_i t + r_i \cos(8t)$	0.4	$z_i \in (0,1]$	5	5.327	5.205	5.176
			$r_i \in (1,2]$		4.915	4.992	4.990
					3.240	2.990	3.006
14.	$z_i t + r_i t^2$	1	$z_i \in [0,1]$	3	3.459	3.014	3.061
			$r_i \in [0,1]$		2.405	3.047	3.242
					4.646	5.087	5.098
15.	$z_i t + r_i \log(t)$	0.5	$z_i \in [0.5, 1.5]$	5	4.777	5.089	5.095
			$r_i \in [0.5, 1.5]$		4.482	5.178	5.210
					1.005	0.916	0.896
16.	$(z_i + t)^{-1}$	0.8	$z_i \in (0,1]$	0.9	0.922	0.909	0.908
					0.784	0.874	0.858
					0.479	0.515	0.517
17.	t^{z_i}	1	$z_i \in [0, 0.5]$	0.5	0.401	0.469	0.464
					0.551	0.526	0.525
					0.799	1.094	1.007
18.	$(1 + z_i t)^{-1}$	0.6	$z_i \in [1,2]$	1	0.891	1.021	1.052
					0.911	1.061	1.041
					1.022	0.932	0.929
19.	$\left(1 + z_i t\right)^{0.7}$	1	$z_i \in [0.5, 2.5]$	0.9	0.975	0.933	0.932
					0.806	0.866	0.865
					0.481	0.511	0.510
20.	$\left(1 + e^{-z_i t}\right)^{-1}$	0.2	$z_i \in [0.5, 1.5]$	0.5	0.624	0.584	0.582
					0.540	0.487	0.488
					1.644	1.576	1.570
21.	z_i^t	0.5	$z_i \in (0,1]$	1.5	1.344	1.535	1.551
					1.260	1.504	1.474
					1.104	1.034	1.028
22.	$e^{-z_i t}$	1	$z_i \in (0,1]$	1	0.890	0.965	1.962
					1.095	1.017	1.010



а) $S_n(x) = \sum\limits_{i=1}^{n} (X_i - f_i(x))^2$

б) $\psi_n(x) = \sum\limits_{i=1}^{n} f_i'(x)(X_i - f_i(x))$

Рис. 31: Метод наименьших квадратов. Для модели 13 из таб. 2 иллюстрируется поведение функций $S_n(\cdot)$ и $\psi_n(\cdot)$ в окрестности истинного значения параметра $\theta = 5$. Пунктирной линией на графике б) отмечена функция $\psi_n^*(x) = \sum_{i=1}^n f_i'(\theta_n^*)(X_i - f_i(x))$.

Заключение

Регрессионный анализ относится к интенсивно развивающему разделу математической статистики. В диссертации установлен ряд взаимосвязанных результатов, относящихся к непараметрической и нелинейной регрессии. Все поставленные цели достигнуты, а результаты, включенные в диссертацию, являются новыми и актуальными. В качестве наиболее значимых результатов выделим следующее.

В диссертации предложены концепция плотных данных и классы универсальных относительно стохастической природы регрессоров состоятельных оценок в моделях непараметрической и нелинейной регрессии. Эта концепция позволила в задачах непараметрической регрессии значительно ослабить известные ограничения на регрессоры и выделить одно существенное их свойство (условие плотного заполнения), а в задачах нелинейной — решить открытую проблему поиска явных оценок конечномерных параметров. Кроме того, предлагаемая концепция позволила в едином подходе рассматривать ситуацию детерминированных и случайных регрессоров.

Важно подчеркнуть, что полученные в непараметрической регрессии результаты позволяют строить равномерно состоятельные оценки ядерного типа при отсутствии какой-либо информации о характере зависимости регрессоров, лишь бы они плотно заполняли нужную область. Условия на регрессоры в терминах плотных данных нечувствительны к типу регрессоров, характеру их корреляции и включают в себя как ситуацию детерминированных регрессоров без дополнительного требования регулярности, так и случайных, при этом не обязательно удовлетворяющих условиям слабой зависимости.

Методы построения явных оценок для широкого класса задач нелинейной регрессии, предложенные в диссертации, не имеют аналогов в научной литературе. Ранее явные оценки были известны лишь для небольшого числа нелинейных регрессионных моделей, и проблема их построения для достаточно широких классов моделей оставалась открытой. Отметим, что явные оценки играют важную роль в одношаговых процедурах статистического оценивания, позволяющих обойти вычислительные трудности поиска оценок, задаваемых уравнениями.

Рекомендации и перспективы дальнейшей разработки тематики исследований. Идеи и методы диссертационной работы могут получить дальнейшее развитие в различных статистических задачах.

1. Концепция универсальных относительно стохастической природы условий на регрессоры в терминах плотных данных может быть реализована, на наш взгляд, в различных постановках задач непараметрической регрессии. Например, когда вместе с регрессионной

функцией оцениваются и ее производные. Здесь может идти речь как о получении новых достаточных универсальных условий на регрессоры, обеспечивающих равномерную состоятельность классических локально–полиномиальных оценок степени $p \geq 2$, так и об исследовании новых классов универсальных локально–полиномиальных оценок, аналогичных рассмотренным в диссертации. Отметим, что при оценивании производных второго порядка и выше соответствующие локально–полиномиальные оценки будут неявными и в такой непараметрической постановке разумно использовать одношаговые процедуры оценивания.

2. Результаты, полученные в непараметрической регрессии, допускают обобщение на случай оценивания разрывных функций (без разрывов второго рода), но сближение оценки и функции в этом случае следует изучать в других метриках (например, интегральной). Так что представляется возможным получить результаты и в этом направлении, используя идеи исследования, представленного в диссертации.

3. В задаче оценивания функций среднего и ковариации случайного процесса существенно используются новые универсальные локально–постоянные оценки, предложенные в диссертационной работе. В этой задаче можно использовать и оценки локально–линейного типа. Кроме того, результаты могут быть перенесены на случай оценивания функций среднего и ковариции непрерывного случайного поля.

4. Идеи построения предварительных оценок в нелинейной регрессии могут получить дальнейшее развитие. Например, в случае разрывных функций. Остается неисследованным вопрос об условиях асимптотической нормальности явных оценок, построенных с использованием непараметрических методов ядерного сглаживания.

5. Методика оценивания в нелинейной регрессии, предложенная в диссертационной работе, может быть перенесена и адаптирована к различным специальным постановкам задач нелинейной регрессии (например, в моделях с разладкой).

6. Результаты, связанные с асимптотическим анализом одношаговых оценок, могут быть перенесены на другие классы одношаговых оценок, т.е. на классы оценок, порождаемых уравнениями, отличными от рассмотренных в диссертационной работе.

7. Те или иные результаты, полученные в случае скалярных регрессоров или одномерного параметра могут быть обобщены на случай векторных величин и, соответственно, функций многих переменных.

Литература

1. Acitas S., Kasap P., Senoglu B., Arslan O. One-step M-estimators: Jones and Faddy's skewed t-distribution // J. Appl. Stat. — 2013. — V.40. — No 7. — P.1545–1560.

2. Ahmad I. A., Lin P.-E. Fitting a multiple regression function // J. Statist. Plann. Infer. — 1984. — V.9. — No 2. — P.163–176.

3. Amemiya T. Regression analysis when the variances of the depended variable is proportional to the squares of its expectation // J. Amer. Stat. Assoc. — 1973. — V.68. — No 344. — P.928–934.

4. Anatolyev S. Nonparametric regression // Quantile. — 2009. — V.7. — P.37–52.

5. Antoniadis A., Fan J. Regularization of wavelet approximations // J. Amer. Stat. Assoc. — 2001. — V.96. — No 455. — P.939-955.

6. Atkinson A. C., Donev A. N., Tobias R. D. Optimum experimental design, with SAS. — Oxford University Press, Oxford, 2007.

7. Bai Z. D., Wu Y. General M-estimation // J. Multivar. Anal. — 1997. — V.63. — P.119– 135.

8. Balan R. M., Schiopu-Kratina I. Asymptotic results with generalized estimating equation for longitudinal data // Ann. Statist. — 2005. — V. 33. — No 2. — P.522–541.

9. Bassett R., Deride J. One-step estimation with scaled proximal methods // Math. Oper. Res. — 2022. — 47. — No 3. — P.2366–2386.

10. Bates D. M., Watts D. G. Nonlinear regression analysis and its applications. — Wiley, 1988.

11. Benelmadani D., Benhenni K., Louhichi S. Trapezoidal rule and sampling designs for the nonparametric estimation of the regression function in models with correlated errors // Statistics. — 2020. — V.54. — No 1. — P.59–96.

12. Benhenni K., Hedli-Griche S., Rachdi M. Estimation of the regression operator from functional fixed-design with correlated errors //J. Multivariate Anal. — 2010. — V.101. — P.476–490.

13. Beran J., Feng Y. Local polynomial estimation with a FARIMA-GARCH error process // Bernoulli. — 2001. — V.7. — No 5. — P.733–750.

14. Berger M., Wong W. K. An introduction to optimal designs for social and biomedical research. — John Wiley and Sons, 2009.

15. Bergesioa A., Yohaia V. Projection estimators for generalized linear models // J. Amer. Stat. Assoc. — 2011. V.106(494). — P.661–671.

16. Berk R.A. Statistical learning from a regression perspective. — Springer, 2020.

17. Bianco A., Boente G. On the asymptotic behavior of one-step estimates in heteroscedastic regression models // Stat. Probab. Lett. — 2002. — V.60. — No 1. — P.33–47.

18. Bianco A., Boente G., Martinez E. Robust tests in semiparametric partly linear models // Scand. J. Stat. — 2006. — V.33. — No 3. — P.435–450.

19. Bickel P. J. One-step Huber estimates in the linear model // J. Amer. Stat. Assoc. — 1975. — V.70. — P.428–434.

20. Biedermann S., Dette H. Minimax optimal designs for nonparametric regression - a further optimality property of the uniform distribution // In 6th International Workshop on Model-Oriented Design and Analysis (eds. A. C. Atkinson, P. Hackl and W. G. Muller). — 2001. — P.13–20.

21. Birkes D., Dodge Y. Alternative methods of regression. — Wiley, 1993.

22. Bowman A. W., Azzalini A. Applied smoothing techniques for data analysis: the kernel approach with S-plus illustrations. — Oxford University Press, 1997.

23. Box G. E. P., Hill W. J. Correcting inhomogeneity of variance with power transformation weighting // Technometrics. — 1974. — V.16. — No 3. — P.385–389.

24. Bradic J., Fan J., Wang W. Penalized composite quasi-likelihood for ultrahigh dimensional variable selection.// J. Roy. Statist. Soc. Ser. B. Stat. Methodol. — 2011. — V.73. — P.325–349.

25. Bradic J., Guo J. Generalized M-estimators for high-dimensional Tobit I models // Electron. J. Stat. — 2019. — V.13. — No 1. — P.582–645.

26. Brown L. D., Levine, M. Variance estimation in nonparametric regression via the difference sequence method // Ann. Statist. — 2007. — V.35. — P.2219–2232.

27. Bunea F., Ivanescu A. E., Wegkamp M. H. Adaptive inference for the mean of a Gaussian process in functional data // J. R. Stat. Soc. Ser. B Stat. Methodol. — 2011. — V.73. — P.531–558.

28. Cai Z., Chen L., Fang Y. A new forecasting model for USD/CNY exchange rate // Stud. Nonlinear Dyn. Econ. — 2012. — V.16. — No 3. — P.1–18.

29. Cai Z., Fan J., Li R. Z. Efficient estimation and inferences for varying-coefficient models // J. Amer. Statist. Assoc. — 2000. — V.95. — P.888–902.

30. Cai J., Fan J., Zhou H., Marginal hazard models with varying coefficients for multivariate failure time data // Ann. Statist. — 2007. — V.35. — P.324–354.

31. Cai T. T., Yuan M. Optimal estimation of the mean function based on discretely sampled functional data: phase transition // Ann. Statist. — 2011. — V. 39. — No 5. — P.2330–2355.

32. Cao G., Wang L., Li Y., Yang L. Oracle–efficient confidence envelopes for covariance functions in dense functional data // Stat. Sin. — 2016. — V.26.— P.359–383.

33. Cao G., Yang L., Todem D. Simultaneous inference for the mean function of dense functional data // J. Nonparametr. Statist. — 2012. — V.24. — P.359–377.

34. Carrol R.J., Ruppert D. A comparison between maximum likelihood and generalized least squares in heteroscedastic linear model // J. Amer. Stat. Assoc. — 1982. — V. 77. — P.878–882.

35. Carroll R.J., Ruppert D. Transformation and weighting in regression. — Springer, 1988.

36. Chacon J. E., Duong T. Multivariate kernel smoothing and its applications. — CRC Press, 2018.

37. Chan N., Wang Q. Uniform convergence for Nadaraya-Watson estimators with nonstationary data // Econ. Theory. — 2014. — V.30. — P.1110–1133.

38. Chatterjee S., Hadi A. S. Regression analysis by example. — Wiley, 2006.

39. Chen J., Gao J., Li D. Estimation in semi-parametric regression with non-stationary regressors // Bernoulli. — 2012. — V.18. — No 2. — P.678–702.

40. Chen J., Li D., Zhang L. Robust estimation in a nonlinear cointegration model // J. Multivariate Anal. — 2010. — V.101. — No 3. — P.706–717.

41. Chen K., Lin Y. Efficient estimation of censored linear regression model // Biometrika. — 2013. — V.100. — No 2. — P.525–530.

42. Cheng M.-Y., Hall P., Titterington D. M. Optimal design for curve estimation by local linear smoothing // Bernoulli. — 1998. — V.4(1). — P.3–14.

43. Chentsov N. N. Weak convergence of stochastic processes whose trajectories have no discontinuities of the second kind and the heuristic approach to the Kolmogorov-Smirnov tests // Theory Probab. Appl. — 1956. — V.1. — P.140–144.

44. Cheruiyot L. R. Local linear regression estimator on the boundary correction in nonparametric regression estimation // J. Stat. Theory Appl. — 2020. — V.19. — No 3. — P.460–471.

45. Chu C. K., Deng W.-S. An interpolation method for adapting to sparse design in multivariate nonparametric regression // J. Statist. Plann. Inference. — 2003. — V.116. — No 1. — P.91–111.

46. Chu C.-K., Marron J. S. Choosing a kernel regression estimator // Stat. Sci. — 1991. — V.6. — N4. — P.404–419.

47. Chung I. H., Kim K. H. Asymptotic properties of the one-step M-estimators in nonlinear regression model // Comm. Korean Math. Soc. — 1992. — V.7. — No 2. — P.293–306.

48. Coakley C. W., Hettmansperger T. P. A Bounded influence, high breakdown, efficient regression estimator // J. Amer. Statist. Assoc. — 1993. — V.88. — P.872–880.

49. Cook R. D., Weisberg S. Applied regression including computing and graphics. — Wiley, 1999.

50. Cuevas A. A partial overview of the theory of statistics with functional data // J. Stat. Plan. Inference. — 2014. — V.147. — P.1–23.

51. Das R. N. Robust response surfaces, regression, and positive data analyses. — Chapman and Hall/CRC, 2014.

52. Dattner I., Gugushvili S. Application of one-step method to parameter estimation in ODE models // Stat. Neerl. — 2018. — V.72. — No 2. — P.126–156.

53. Degras D. Asymptotics for the nonparametric estimation of the mean function of a random process // Stat. Prob. Lett. — 2008. — V.78. — No 17. — P.2976–2980.

54. Dette H., Melas V., Pepelyshev A. Standardized E-optimal designs for the Michaelis–Menten model // Stat. Sin. — 2003. — V.13. — P.1147–1163.

55. Dette H., Melas V. B. A note on the de la Garza phenomenon for locally optimal designs // Ann. Statist. — 2011. — V.39. — P.1266–1281.

56. Dette H., Melas V., Wong W. K. Optimal design for goodness-of-fit of the Michaelis–Menten enzyme kinetic function // J. Amer. Statist. Assoc. — 2005. — V.100. — N.472. — P.1370–1381.

57. Dette H., Wong W. K. E-optimal designs for the Michaelis-Menten model // Statist. Probab. Lett. — 1999. — V.44. — P.405–408.

58. Devroye L. P. The uniform convergence of the Nadaraya–Watson regression function estimate // Can. J. Stat. — 1979. — V.6. — P.179–191.

59. Dollinger M. B., Staudte R. G. Influence functions of iteratively reweighted least squares estimators // J. Amer. Statist. Assoc. — 1991. — V.86. — P.709–716.

60. Draper N. R., Smith H. Applied regression analysis. — Wiley, 1998.

61. Draper N. R. Applied regression analysis bibliography update 1988-1989 // Commun. Stat. Theory Methods — 1990. — V.19. — No 4. — P.1205–1229.

62. Draper N. R. Applied regression analysis bibliography update 1990-1991 // Commun. Stat. Theory Methods — 1992. — V.21. — No 9. — P.2415–2437.

63. Draper N. R. Applied regression analysis bibliography update 1992-1993 // Commun. Stat. Theory Methods — 1994. — V.23. — No 9. — P.2701–2731.

64. Draper N. R. Applied regression analysis bibliography update 1994-1997 // Commun. Stat. Theory Methods — 1998. — V.27. — No 10. — P.2581–2623.

65. Draper N. R. Applied regression analysis bibliography update 1998-1999 // Commun. Stat. Theory Methods — 2000. — V.29. — No 9,10. — P.2313–2341.

66. Draper N. R. Applied regression analysis bibliography update 2000-2001 // Commun. Stat. Theory Methods — 2002. — V.31. — No 11. — P.2051–2075.

67. Duchesne P., Micheaux P. L., Tatsinkou J.F.T. On strong consistency and asymptotic normality of one-step Gauss-Newton estimators in ARMA time series models // Statistics. — 2020. — V.54. — No 5. — P.1030–1057.

68. Dutter R., Huber P. Numerical methods for the nonlinear robust regression problem // J. Stat. Comput. Simulation. — 1981. — V.13. — P.79–114.

69. Eagleson G. K., Muller H.-G. Transformations for smooth regression models with multiplicative errors // J. R. Statist. Soc. B — 1997. — V.59. — No 1. — P.173–189.

70. Efromovich S. Nonparametric curve estimation: methods, theory, and applications. — New York: Springer, 1999.

71. Efromovich S. Optimal sequential design in a controlled non-parametric regression // Scand. Stat. Theory Appl. — 2008. — V.35. — P.266–285.

72. Einmahl U., Mason D. M. Uniform in bandwidth consistency of kernel-type function estimators // Ann. Statist. — 2005. — V.33. — P.1380–1403.

73. Eisenach C., Bunea F., Ning Y., Dinicu C. High-dimensional inference for cluster-based graphical models // J. Mach. Learn. Res. — 2020. — V.21. — P.1–55.

74. Fahrmeir L., Kneib T., Lang S., Marx B. Regression: models, methods and applications. — Springer, 2013.

75. Fan Y. Consistent nonparametric multiple regression for dependent heterogeneous processes: the fixed design case // J. Multivariate Anal. — 1990. — V.33. — P.72–88.

76. Fan J., Chen J. One-step local quasi-likelihood estimation // J. Roy. Statist. Soc. Ser. B. Stat. Methodol. — 1999. — V.61. — P.927–943.

77. Fan J., Gijbels I. Local polynomial modelling and its applications. — London: Chapman and Hall, 1996.

78. Fan J., Jiang J. Variable bandwidth and one-step local M-estimator // Science in China (Series A). — 2000. — V.43. — P.65–80.

79. Fan J., Li R. Variable selection for Cox's proportional hazards model and frailty model // Ann. Statist. — 2002. — V.30. — No 1. — P.74–99.

80. Fan J., Li R., Zhang C.-H., Zou H. Statistical foundations of data science. — Chapman and Hall/CRC, 2020.

81. Fan J., Lin H., Zhou Y. Local partial likelihood estimation for life time data // Ann. Statist. — 2006. — V.34. — P.290–325.

82. Fan J., Xue L., Zou H. Strong oracle optimality of folded concave penalized estimation // Ann. Statist. — 2014. — V.42. — P.819–849.

83. Fang F., Zhao J., Ahmed S. E., Qu A. A weak-signal-assisted procedure for variable selection and statistical inference with an informative subsample // Biometrics. — 2021. — V.77 — No 3. — P.996–1010.

84. Fan J., Yao Q. Nonlinear time series nonparametric and parametric methods. — Springer, 2003.

85. Fan J., Yao Q., Cai Z. Adaptive varying-coefficient linear models // J. R. Stat. Soc. Series B Stat. Methodol. — 2003. — V.65. — No 1. — P.57–80.

86. Fedorov V. V., Leonov S. L. Optimal design for nonlinear response models. — CRC Press, 2014.

87. Fedorov V. V., Montepiedra G., Nachtsheim C. J. Design of experiments for locally weighted regression // J. Stat. Plan. Inference. — 1999. — V.81. — P.363–382.

88. Field C. A, Wiens D. P. One-step M-estimators in the linear model, with dependent errors // Canad. J. Statist. — 1994. — V.22. — No 2. — P.219–231.

89. Fisher R. A. Theory of statistical estimation // Proc. Camb. Phil. Soc. — 1925. — Soc.22. — P.700–725.

90. Ford I., Torsney B., Wu, C. F. J. The use of a canonical form in the construction of locally optimal designs for non-linear problems // J. Roy. Statist. Soc. Ser. B. — 1992. — V.54. — P.569–583.

91. Fox J., Weisberg S. An R companion to applied regression. — SAGE, 2019.

92. Fox J. Applied regression analysis and generalized linear models. — SAGE, 2015.

93. Freund R. J., Wilson W. J., Sa P. Regression analysis: statistical modeling of a response variable. — Academic Press, 2006.

94. Furno M., Vistocco D. Quantile regression. — Wiley, 2018.

95. Gao J., Kanaya S., Li D., Tjostheim D. Uniform consistency for nonparametric estimators in null recurrent time series // Econ. Theory. — 2015. — V.31. — P.911–952.

96. Gasser T., Engel J. The choice of weights in kernel regression estimation // Biometrica. — 1990. — V.77. — P.277–381.

97. Gasser T., Müller H.-G. Kernel estimation of regression functions. // In Lecture Notes in Mathematics. — 1979. — 757. — P.23–68.

98. Gelman A. Data analysis using regression and multilevel/hierarchical models. — Cambridge University Press, 2008.

99. Georgiev A. A. Asymptotic properties of the multivariate Nadaraya-Watson regression function estimate: the fixed design case // Stat. Probab. Lett. — 1989. — V.7. — No 1. — P.35–40.

100. Georgiev A. A. Nonparametric multiple function fitting // Stat. Probab. Lett. — 1990. — V.10. — No 3. — P.203–211.

101. Georgiev A. A. Consistent nonparametric multiple regression: The fixed design case // J. Multivariate Anal. — 1988. — V.25. — No 1. — P.100–110.

102. Gervini D. Free-knot spline smoothing for functional data // J. R. Stat. Soc. Ser. B Stat. Methodol. — 2006. — V.68. — P. 671–687.

103. Giloni A., Simonoff J. S. The conditional breakdown properties of least absolute value local polynomial estimators // J. Nonparametr. Stat. — 2005. — V.17. — No 1. — P.15–30.

104. Ghilagaber G., Midi H., Riazoshams H. Robust nonlinear regression: with applications using R. — Wiley, 2019.

105. Glantz S. A., Slinker B., Neilands T. B. Primer of applied regression and analysis of variance. — McGraw-Hill Education, 2016.

106. Grigoriev Yu. D. Melas V. B., Shpilev P. V. Excess of locally D-optimal designs for Cobb–Douglas model // Stat. Pap. — 2018. — V.259.— P.1425–1439.

107. Grigoriev Yu. D. Melas V. B., Shpilev P. V. Excess and saturated D-optimal designs for the rational model // Stat. Pap. — 2021. — V.62. — P.1387–1405.

108. Gu J., Li Q., Yang J.-C. Multivariate local polynomial kernel estimators: leading bias and asymptotic distribution // Econom. Rev. — 2015. — V. 34. — P.979–1010.

109. Gu W., Roussas G. G., Tran L. T. On the convergence rate of fixed design regression estimators for negatively associated random variables // Stat. Probab. Lett. — 2007. — V.77. — No 12. — P.1214–1224.

110. Györfi L., Kohler M., Krzyzak A., Walk H. A distribution-free theory of nonparametric regression. — Springer, 2002.

111. Hall P., Heyde C. C. Martingale limit theory and its application. — Academic Press, 1980.

112. Hall P., Ma Y. Quick and easy one-step parameter estimation in differential equations // J. R. Stat. Soc. Series B Stat. Methodol. — 2014. — V.76. — No 4. — P.735–748.

113. Hall P., Müller H.-G., Wang, J.-L. Properties of principal component methods for functional and longitudinal data analysis // Ann. Statist. — 2006. — V.34. — P.1493–1517.

114. Hampel F. R., Ronchetti E. M., Rousseeuw P. J., Stahel W. A. Robust statistics: the approach based on influence functions. — Wiley, 2011.

115. Hansen B. E. Uniform convergence rates for kernel estimation with dependent data // Econ. Theory. — 2008. — V.24. — P.726–748.

116. Härdle W. Applied nonparametric regression. — Cambridge University Press, 1990.

117. Härdle W., Luckhaus S. Uniform consistency of a class of regression function estimators // Ann. Statist. — 1984. — V.12. — No 2. — P.612–623.

118. Härdle W., Muller M., Sperlich S., Werwatz A. Nonparametric and semiparametric models. — Berlin: Springer-Verlag, 2004.

119. Hastie T., Tibshirani R., Friedmanet J. The elements of statistical learning: data mining, inference, and prediction. — New York: Springer, 2009.

120. He X., Shao Q.-M. A general Bahadur representation of M-estimators and its application to linear regression with nonstochastic designs // Ann. Statist. — 1996. — V. 24. — No 6. — P.2608–2630.

121. He Q. Consistency of the Priestley–Chao estimator in nonparametric regression model with widely orthant dependent errors // J. Inequal. Appl. — 2019. — V.64. — P.2–13.

122. Heyde C. C. Quasi-likelihood and its application: a general approach to optimal parameter estimation. — Springer, 1997.

123. Hirukawa M. Asymmetric kernel smoothing. — Springer Singapore, 2018.

124. Hoadley B. Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case // Ann. Math. Statist. — 1971. — V. 42. — No 6. — P.1977–1991.

125. Honda T. Nonparametric regression for dependent data in the errors-in-variables problem // J. Stat. Plan. Inference. — 2010. — V. 140. — No 11. — P.3409–3424.

126. Hong S. Y., Linton O. B. Asymptotic properties of a Nadaraya-Watson type estimator for regression functions of infinite order // SSRN Electronic Journal. — 2016.

127. Horova I. Kolacek J., Zelinka J. Kernel smoothing in MATLAB: theory and practice of kernel smoothing. — World Scientific, 2012.

128. Houwelingen J. S. Use and abuse of variance models in regression // Biometrics. — 1988. — V.44. — P.1073–1081.

129. Hsing T., Eubank R. Theoretical foundations of functional data analysis, with an introduction to linear operators. — Wiley, 2015.

130. Hsu D., Kakade S. M., Zhang T. Random design analysis of ridge regression // Found. Comput. Math. — 2014. — V.14. — P.569–600.

131. Huang C., Huo X. A distributed one-step estimator // Math. Program. — 2019. — V.174. — P.41–76.

132. Huber P. J. Robust estimation of a location parameter // Ann. Math. Statist. — 1964. — V.35. — P.73–101.

133. Huber P. J. Robust regression: asymptotics, conjectures and Monte Carlo // Ann. Statist. — 1973. — V.1. — P.799–821.

134. Huo X., Huang C., Ni X. S. Scattered data and aggregated inference // In: Hardle W., Lu H., Shen X. (eds) Handbook of big data analytics. Springer, 2017. — P.75-102.

135. Ioannides D. A. Consistent nonparametric regression: some generalizations in the fixed design case // J. Nonparametr. Stat. — 1993. — V.2. — No 3. — P.203–213.

136. Ivanov A. V. Asymptotic theory of nonlinear regression. — Springer, 1997.

137. James G. M., Hastie T. J. Functional linear discriminant analysis for irregularly sampled curves // J. R. Stat. Soc. Ser. B Stat. Methodol. — 2001. — V.63. — P.533–550.

138. Janssen P., Jureckova J., Veraverbeke N. Rate of convergence of one- and two-step M-estimators with applications to maximum likelihood and Pitman estimators // Ann. Stat. — 1985. — V.13. — No 3. — P.1222–1229.

139. Jennen-Steinmetz C., Gasser T. A unifying approach for nonparametric regression estimation // J. American. Stat. Assoc. — 1989. — V.83. — P.1084–1089.

140. Jennrich R. I. Asymptotic properties of nonlinear least squares estimation // AMS. — 1969. — V.40. — No 2.

141. Jiao X., Nielsen B. Asymptotic analysis of iterated 1-Step Huber-skip M-estimators with varying cut-offs // Analytical methods in statistics. AMISTAT 2015. Springer Proceedings in Mathematics and Statistics, vol 193. — Springer, 2017.

142. Jiang J., Mack Y. P. Robust local polynomial regression for dependent data // Stat. Sin. — 2001 — V.11. — P.705-722.

143. Jobson J. D., Fuller W. A. Least squares estimation when the covariance matrix and the parameter vector are functionally related // J. Amer. Stat. Assoc. — 1980. — V.75. — 369. — P.176–181.

144. Johansen S., Nielsen B. An analysis of the indicator saturation estimator as a robust regression estimator // The methodology and practice of econometrics. Edited by Castle J., Shephard N. — Oxford University Press, 2009. — P.1-36.

145. Johansen S., Nielsen B. Asymptotic theory of outlier detection algorithms for linear time series regression models // Scand. J. Stats. — 2016. — V.43. — P.321–348.

146. Jones M. C., Davies S. J., Park B. U. Versions of kernel-type regression estimators // J. Americ. Stat. Assoc. — 1994. — V.89. — P.825–832.

147. Jureckova J. Tail-behavior of estimators and of their one-step versions // J. Soc. fr. stat. — 2012. — V.153. — No 1. — P.44–51.

148. Jureckova J., Maly M. The asymptotics for studentized k-step M-estimators of location // Sequen. Anal. — 1995. — V.14. — P.229–245.

149. Jureckova J., Picek J., Schindler M. Robust statistical methods with R. — Chapman and Hall, 2019.

150. Jureckova J., Sen P. K. Robust statistical procedures: asymptotics and interrelations. — Wiley, 1996.

151. Jureckova J., Sen P. K., Picek J. Methodology in robust and nonparametric statistics. — London: Chapman and Hall, 2012.

152. Jureckova J., Portnoy S. Asymptotics for one-step M-estimators in regression with application to combining efficiency and high breakdown point // Comm. Statist. Theory Methods. — 1987. — V.16. — P.2187–2200.

153. Jureckova J., Sen P. K. Effect of the initial estimator on the asymptotic behavior of the one-step M-estimator // Ann. Inst. Statist. Math. — 1990. — V.42. — No 2. — P. 345–357.

154. Karlsen H. A., Myklebust T., Tjostheim D. Nonparametric estimation in a nonlinear cointegration type model // Ann. Statist. — 2007. — V.35. — P.252–299.

155. Karunamuni R. J., Wu J. One-step minimum Hellinger distance estimation // Comput. Stat. Data An. — 2011. — V.55. — No 12. — P.3148–3164.

156. Khuri A. I., Mukherjee B., Sinha B. K., Ghosh M. Design issues for generalized linear models: a review // Statist. Sci. — 2006. — V.21. — P.376–399.

157. Kim S., Zhao Z. Unified inference for sparse and dense longitudinal models // Biometrika. — 2013. — V.100. — No 1. — P.203–212.

158. Klemela J. Multivariate nonparametric regression and visualization. — Wiley, 2014.

159. Kokoszka P., Reimherr M. Introduction to functional data analysis. — Chapman and Hall/CRC, 2017.

160. Koldaeva A., Sakhanenko A. Existence of explicit asymptotically normal estimators in a multiple logarithmic regression problem // Sib. Electron. Mat. Izv. — 2017. — V.14. — P. 972–979.

161. Knopov P. S., Korkhin A. S. Regression analysis under a priori parameter restrictions. — Springer, 2012.

162. Kraft C. and Le Cam L. A remark on the roots of the maximum likelihood equation // Ann. Math. Statist. — 1956. — V.27. — P.1174–1177.

163. Kulik R., Lorek P. Some results on random design regression with long memory errors and predictors // J. Statist. Plann. Infer. — 2011. — V.141. — P.508–523.

164. Kulik R., Wichelhaus C. Nonparametric conditional variance and error density estimation in regression models with dependent errors and predictors // Electr. J. Statist. — 2011. — V.5. — P.856–898.

165. Lai T. L. Asymptotic properties of nonlinear least squares estimates in stochastic regression models // Ann. Statist. — 1994. — V.22. — No 4. — P.1917–1930.

166. Laib N., Louani D. Nonparametric kernel regression estimation for stationary ergodic data: asymptotic properties // J. Multivar. Anal. — 2010. — V.101. — P.2266–2281.

167. LaRiccia V. N., Eggermont P. P. Maximum penalized likelihood estimation. Volume II: regression. — Springer, 2009.

168. Lehmann E., Casella G. Theory of point estimation. — Berlin: Springer, 1998.

169. Le Cam L. On the asymptotic theory of estimation and testing hypotheses // in Proc. Third Berkeley Sympos. Math. Statist. Probab. — 1956. — V.1. — 129.

170. Li H., Calder C. A., Cressie N. One-step estimation of spatial dependence parameters: properties and extensions of the APLE statistic // J. Multivar. Anal. — 2012. — V.105. — No 1. — P.68–84.

171. Li Q., Lu X., Ullah A. Multivariate local polynomial regression for estimating average derivatives // Nonparametric Statistics. — 2003. — V.15. — No 4-5. — P.607–624.

172. Li X, Yang W., Hu S. Uniform convergence of estimator for nonparametric regression with dependent data // J. Inequal. Appl. — V. 2016. —- Article number: 142.

173. Li Y., Hsing T. Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data // Ann. Statist. — 2010. — V.38. — P.3321–3351.

174. Li R., Marron J. S. Local likelihood SiZer Map // Sankhyia. — 2005. — V.67. — P.476–498.

175. Li J., Zheng M. Robust estimation of multivariate regression model // Stat. Pap. — 2007. — V.50. — No 1. — P.81–100.

176. Liang H.-Y., Jing B.-Y. Asymptotic properties for estimates of nonparametric regression models based on negatively associated sequences // J. Multivariate Anal. — 2005. — V.95. — No 2. — P.227–245.

177. Liero H. Strong uniform consistency of nonparametric regression function estimates // Probab. Th. Rel. Fields. — 1989. — V.82. — P.587–614.

178. Liese F., Vajda I. A general asymptotic theory of M-estimators. I // Math. Methods Stat. — 2004. — V.12. — No 4. — P.454–477.

179. Liese F., Vajda I. A general asymptotic theory of M-estimators. II // Math. Methods Stat. — 2004. — V.13. — No 1. — P.82–95.

180. Lin Z., Wang J.-L. Mean and covariance estimation for functional snippets // J. Amer. Statist. Assoc. — 2022. — V.117. — P.348–360.

181. Linton O. B., Jacho-Chavez D. T. On internally corrected and symmetrized kernel estimators for nonparametric regression // TEST. — 2010. — V.19. — No 1. — P.166–186.

182. Linton O., Wang Q. Nonparametric transformation regression with nonstationary data // Econ. Theory. — 2016. — V.32. — No 1. — P. 1–29.

183. Linton O., Xiao Z. A nonparametric regression estimator that adapts to error distribution of unknown form // Econ. Theory. — 2007. — V.23. — No 3. — P.371–413.

184. Lipsitz S. R., Dear K. B. G., Zhao L. Jackknife estimators of variance for parameter estimates from estimating equations with applications to clustered survival data // Biometrics. — 1994. — V.50. — No 3. — P.842–846.

185. Lipsitz S., Fitzmaurice G., Sinha D., Hevelone N., Hu J., Nguyen L. L. One-Step generalized estimating equations with large cluster sizes // J. Comput. Graph. Stat. — 2017. — V.26. — No 3. — P.734–737.

186. Loader C. Local regression and likelihood. — Springer, 1999.

187. Lopuhaa H. P. Asymptotics of reweighted estimators of multivariate location and scatter // Ann. Statist. — 1999. — V.27. — No 5. — P.1638–1665.

188. Ma S., Yang L., Carroll R. J. A simultaneous confidence band for sparse longitudinal regression // Statist. Sinica. — 2012. — V.22. — P.95–122.

189. Mack Y. P., Müller H.-G. Adaptive nonparametric estimation of a multivariate regression function // J. Multivariate Anal. — 1987. — V.23. — P.169–182.

190. Mack Y. P., Müller H.-G. Convolution type estimators for nonparametric regression // Stat. Probab. Lett. — 1988. — V.7. — P.229–239.

191. Mack Y. P., Müller H.-G. Erratum // Stat. Probab. Lett. — 1989. — V.8. — P.195.

192. Mack Y. P., Silvermann B.W. Weak and strong uniform consistency of kernel regression estimates // Z. Wahrscheinlichkeitstheor. Verw. Geb. — 1982. — V.61 — P.405–415.

193. Markovich N. M. Accuracy of transformed kernel density estimates for a heavy-tailed distribution // Autom. Remote Control. — 2005.— V.66. — No 2. — P. 217–232.

194. Markovich N. M. Transformed estimates of densities of heavy-tailed distributions and classification // Autom. Remote Control. — 2002. — V.63. — No 4. — P.627–640.

195. Markovich L. A. Nonparametric estimation of multivariate density and its derivative by dependent data using gamma kernels // Fundamentalnaya i prikladnaya matematika. — 2018. — V.22. — No 3. — P.145–177.

196. Maronna R. A., Martin R. D., Yohai V. Robust statistics: theory and methods. — Wiley, 2006.

197. Markatou M. Weighting games in robust linear regression // J. Multivariate Anal. — 1999. — V.70. — P.118–135.

198. Marx B. D., Smith E. P. Principal component estimation for generalized linear regression // Biometrika. — 1990. — V.77. — No 1. — P.23–31.

199. Masry E. Nonparametric regression estimation for dependent functional data // Stoch. Proc. Their Appl. — 2005. — V.115. — P.155–177.

200. Masry E. Multivariate local polynomial regression for time series: uniform strong consistency and rates // J. Time Series Analysis. — 1996. — V.17.— P.571–599.

201. Masry E. Long-range dependence: strong consistency and rates// IEEE Transactions on Information Theory. — 2001. — V.47(7), P.2863–2875.

202. Masry E., Fan J. Local polynomial estimation of regression functions for mixing processes // Scand. Stat. Theory Appl. — 1997. — V.24. — P.165–179.

203. Melas V. B. On the functional approach to optimal designs for nonlinear models // J. Stat. Plan. Inference. — 2005. — V.132. — P.93–116.

204. Melas V. B. Functional approach to optimal experimental design. — Heidelberg: Springer, 2006.

205. Melas V. B., Guchenko R., Strashko V. Standardized maximin criterion for discrimination and parameter estimation of nested models // Commun. Stat. Simul. Comput. — 2022. — V.51. — No 8.— P.4314–4325.

206. Melas V. B., Shpilev P. On the functional approach to bayesian efficient designs for nonlinear regression models // Commun. Stat. Theory Methods. — 2012. — V.41. — P.2908–2921.

207. Melas V. B., Staroselsky Yu. D-efficient Bayesian designs for a class of nonlinear models // J. Statist. Theor. Practice. — 2008. — V.2. — P.568–587.

208. Millionshchikov N. V. Asymptotic normality of regression estimates for weakly dependent random fields // Vestnik Moskov. Univ. Ser. 1. Mat. Mekh. — 2005. — N.2. — P.3–8.

209. Morgan B. J. T. Analysis of quantal response data. — Chapman and Hall/CRC, 2018.

210. Müller Ch. H. One-step-M-estimators in conditionally contaminated linear models // Stat. Decis. — 1994. — V.12. — P.331–342.

211. Mu B., Bai E.-W., Zheng W. X., Zhu Q. A globally consistent nonlinear least squares estimator for identification of nonlinear rational systems // Automatica. — 2017. — V.77. — P.322–335.

212. Müller Ch. H. Asymptotic behaviour of one-step-M-estimators in contaminated nonlinear models. — in Asymptotic Statistics, Physica-Verlag, Heidelberg, 1994. P.395–404.

213. Muller H.-G. Functional modelling and classification of longitudinal data // Scand. J. Statist. — 2005. — V.32. — P.223–246.

214. Müller H.-G. Nonparametric regression analysis of longitudinal data. — New York: Springer, 1988.

215. Muller H.-G. Optimal designs for nonparametric kernel regression // Stat. Probab. Lett. — 1984. — V.2. — No 5. — P.285–290.

216. Muller H.-G., Prewitt K. A. Multiparameter bandwidth processes and adaptive surface smoothing // J. Multivariate Anal. — 1993. — V.47. — No 1. — P.1–21.

217. Muller H.-G. Density adjusted kernel smoothers for random design nonparametric regression // Stat. Probab. Lett. — 1997. — V. 36. — No 2. — P.161-172.

218. Muller W. G. Optimal design for local fitting // J. Stat. Plan. Inference. — 1996. — V.55. — P.389–397.

219. Nadaraya E. A. On estimating regression // Theory Probab. Appl. — 1964. — V.9. — No 1. — P.141–142.

220. Nadaraya E. A. Remarks on non-parametric estimates for density functions and regression curves // Theory Probab. Appl. — 1970. — V.15. — P.134–137.

221. Ning Y. Liu H. A general theory of hypothesis tests and confidence regions for sparse high dimensional models // Ann. Statist. — 2017. — V.45. — No 1. — P.158–195.

222. Onyiah L. S. Design and analysis of experiments: classical and regression approaches with SAS. — Chapman and Hall/CRC, 2008.

223. Panik M. J. Regression modeling: methods, theory, and computation with SAS. — CRC Press, 2009.

224. Parzen E. On estimation of a probability density function and mode // Ann. Math. Statistics. — 1962. — V.33. — P.1065–1076.

225. Petrov V. V. Sums of independent random variables. — Springer, 1975.

226. Philippou A. N., Roussas G. G. Asymptotic distribution of the likelihood function in the independent not identically distributed case // Ann. Statist. — 1973. — V.1. — No 3. — P.454–471.

227. Philippou A. N., Roussas G. G. Asymptotic normality of the maximum likelihood estimate in the independent not identically distributed case // Ann. Inst. Statist. Math. — 1975. — V. 27. — No 1. — P.45–55.

228. Poetscher B. M., Prucha I. R. A class of partially adaptive one-step M-estimators for the non-linear regression model with dependent observations // J. Econometrics. — 1986. — V.32. — P.219–251.

229. Politis D. N. Model-free prediction and regression: a transformation-based approach to inference. — Springer International Publishing, 2015.

230. Pollard D., Radchenko P. Nonlinear least-squares estimation // J. Multivar. Anal. — 2006. — V.97. — No 2. — 548–562.

231. Priestley M. B., Chao M. T. Non-parametric function fitting // J. Roy. Statist. Soc. Ser. B. Stat. Methodol. — 1972. — V.34. — P.385–392.

232. Pronzato L. Asymptotic properties of nonlinear estimates in stochastic models with finite design space // Stat. Probab. Lett. — 2009. — V.79. — No 21. — P.2307–2313.

233. Qian L., Correa J. Estimation of weibull parameters for grouped data with competing risks // J. Stat. Comput. Simulation. — 2003. — V.73. — No 4. — P.261–275.

234. Qiu P. Image processing and jump regression analysis. — Wiley, 2005.

235. Rao C. R., Toutenburg H. Linear models: least squares and alternatives. — Springer, 1999.

236. Rawlings J. O., Pantula S. G., Dickey D. A. Applied regression analysis: a research tool. — Springer, 2001.

237. Reeds J. Asymptotic numbers of roots of Cauchy location likelihood equation // Ann. Statist. — 1985. — V.13. — P.775–784.

238. Rice J. A., Silverman B. W. Estimating the mean and covariance structure nonparametrically when the data are curves // J. Roy. Statist. Soc. Ser. B — 1991. — V.53. — P.233–243.

239. Rice J. A., Wu C. O. Nonparametric mixed effects models for unequally sampled noisy curves // Biometrics. — 2001. — V.57. — P.253–259.

240. Rieder H. Robust asymptotic statistics. — Springer, 1994.

241. Rio E. Moment inequalities for sums of dependent random variables under projective conditions // J. Theor. Probab. — 2009. — V.22. — P.146–163.

242. Robinson P. M. The stochastic difference between econometric statistics // Econometrica. — 1988. — V.56. — P.531–548.

243. Rosenblatt M. Remarks on some nonparametric estimates of a density function// Ann. Math. Statist. — 1956. — V.27. — No 3. — p.832–837.

244. Rosenblatt M. Conditional probability density and regression estimates // in P.R. Krishaiah, ed., Multivariate Analysis II, Wiley, New York. — 1969. — P.25–31.

245. Roussas G. G. Nonparametric regression estimation under mixing conditions // Stoch. Proc. Appl. — 1990. — V.36. — P.107–116.

246. Roussas G. G., Tran L. T., Ioannides D. A. Fixed design regression for time series: asymptotic normality // J. Multivariate Anal. — 1992. — V.40. — P.262–291.

247. Rousseeuw P. J., Leroy A. M. Robust regression and outlier detection. — Wiley, 1987.

248. Russell R. R. Nonlinear regression modeling for engineering applications: modeling, model validation, and enabling design of experiments. — Wiley, 2016.

249. Ruppert D., Wand M. P. Multivariate locally weighted least squares regression // Ann. Statist. — 1994. — V.22. — No 3. — P.1346–1370.

250. Sakhanenko A. I. On existence of explicit asymptotically normal estimators in nonlinear regression problems // Analytical Methods in Statistics (AMISTAT 2015) Springer Proceed. in Math. and Statist. 193, ed. J. Antoch, J. Jureckova, M. Maciak, M. Pesta — Springer, 2017. — P.159–187.

251. Savinkina E. N., Sakhanenko A. I. On improvement of statistical estimators in a power regression problem // Sib. Electron. Mat. Izv. — 2019. — V.16. — P.1901–1912. [Russian, English abstract]

252. Seber G., Lee A. Linear regression analysis. — Wiley, 2003.

253. Seber G. A. F., Wild C. J. Nonlinear regression. — Wiley, 2003.

254. Serfling R. J. Approximation theorems of mathematical statistics. — Wiley, 1980.

255. Schimek M. G. Smoothing and regression: approaches, computation, and application. — Wiley, 2000.

256. Schorning K., Konstantinou M., Dette H. Optimal designs for series estimation in nonparametric regression with correlated data // Stat. Sin. — 2021. — V.31. — P.1–25.

257. Shalnova S. A., Drapkina O. M. Significance of the ESSE-RF study for the development of prevention in Russia // Cardiovascular therapy and prevention. — 2020. — V.19. — No 3. — P.209–215.

258. Shalnova S. A., Kutsenko V. A., Kapustina A. V., Yarovaya E. B., et al. Associations of blood pressure and heart rate and their contribution to the development of cardiovascular complications and all-cause mortality in the Russian population of 25-64 years // Ration. Pharmacother. Cardiol. — 2020. — V.16. — No 5. — P.759–769.

259. Sheather S. A modern approach to regression with R. — Springer, 2009.

260. Shen J., Xie Y. Strong consistency of the internal estimator of nonparametric regression with dependent data // Stat. Probab. Lett. — 2013. — V.83. — No 8. — P.1915–1925.

261. Shults J., Hilbe J. M. Quasi-least squares regression. — CRC Press, 2014.

262. Simpson D. G., Chang Y.-C. I. Reweighting approximate GM estimators: asymptotics and residual-based graphics // J. Stat. Plan. Infer. — 1997. — V.57. — No 2. — P.273–293.

263. Simpson D. G., Ruppert D., Carroll R. J. On one-step GM estimates and stability of inferences in linear regression // J. Americ. Stat. Assoc. — 1992. — V.87. — P.439–450.

264. Simpson D. G., Yohai V. J. Functional stability of one-step GM-estimators in approximately linear regression // Ann. Statist. — 1998. — V.26. — No 3. — P.1147–1169.

265. Skouras K. Strong consistency in nonlinear regression model // Ann. Statist. — 2000. — V.28. — No 3. — P.871–879.

266. Small C. G., Wang J. Numerical methods for nonlinear estimating equations. — Oxford: Clarendon press, 2003.

267. Small C. G., Wang J., Yang Z. Eliminating multiple root problems in estimation // Statistical Science. — 2000. — V.15. — No 4. — P.313–341.

268. Small C. G., Yang Z. Multiple roots of estimating functions // Canad. J. Statist. — 1999. — V.27. — No 3. — P.585–598.

269. Song Q., Liu R., Shao Q., Yang L. A simultaneous confidence band for dense longitudinal regression // Comm. Statist. Theory Method. — 2014. — V.43. — No 24. — P.5195–5210.

270. Song D., Wong W. K. Optimal two-point designs for the Michaelis–Menten model with heteroscedastic errors // Comm. Statist. Theory Method. — 1998. — V.27. — P.1503–1516.

271. Stefanski L., Carroll L., Ruppert D. Optimally bounded score functions for generalized linear models with applications to logistic regression // Biometrika. — 1986. — V.73. — No 2. — P.413–424.

272. Strongin R. G., Sergeyev Ya. D. Global optimization with non-convex constraints. Sequential and parallel algorithms . — Kluwer Academic Publishers, 2000.

273. Tableman M. Bounded-influence rank regression: a one-step estimator based on Wilcoxon scores // J. Americ. Stat. Assoc. — 1990. — V.85. — P.508–513.

274. Taddy M. One-step estimator paths for concave regularization // J. Comput. Graph. Stat. — 2016. — V.26. — No 3. — P.525–536.

275. Takezawa K. Introduction to nonparametric regression. — Wiley, 2006.

276. Tang X., Xi M., Wu Y., Wang X. Asymptotic normality of a wavelet estimator for asymptotically negatively associated errors // Stat. Probab. Lett. — 2018. — V.140. — P.191–201.

277. Thisted R. A. Elements of statistical computing: numerical computation. — Chapman and Hall, 2000.

278. Van der Vaart A. W. Asymptotic statistics. — Cambridge University Press, 2000.

279. Van der Vaart A. W., Wellner J. A. Weak convergence and empirical processes. — Springer, 1996.

280. Verrill S. Rate of convergence of k-step Newton estimators to efficient likelihood estimators // Stat. Probab. Lett. — 2007. — V.77. — P.1371–1376.

281. Wakefield J. Bayesian and frequentist regression methods. — Springer, 2013.

282. Wand M. P., Jones M. C. Kernel smoothing. — London: Chapman and Hall, 1995.

283. Wang Q. Least squares estimation for nonlinear regression models with heteroscedasticity // Econ. Theory. — 2021. — V.37. — No 6. — P.1267–1289.

284. Wang S., Cao G., Shang Z. Estimation of the mean function of functional data via deep neural networks // Stat. — 2021. — V.10. — No 1. — e393.

285. Wang J., Cao G., Wang L., Yang L. Simultaneous confidence band for stationary covariance function of dense functional data // J. Multivariate Anal. — 2020. — V.176. — 104584.

286. Wang Q., Chan N. Uniform convergence rates for a class of martingales with application in non-linear cointegrating regression // Bernoulli. — 2014. — V.20. — No 1. — P.207–230.

287. Wang J.-L., Chiou J.-M., Muller H.-G. Review of functional data analysis // Annu. Rev. Statist. — 2016. — V.3. — P.257–295.

288. Wang Q. Y. Phillips P. C. B. Asymptotic theory for local time density estimation and nonparametric cointegrating regression // Econ. Theory. — 2009. — V.25. — P.710–738.

289. Wang Q., Phillips P. C. B. Structural nonparametric cointegrating regression // Econometrica. — 2009. — V.77. — P.1901–1948.

290. Wang K., Wang H. J. Optimally combined estimation for tail quantile regression // Stat. Sin. — 2016. — V.26. — No 1. — P.295–311.

291. Watson G. S. Smooth regression analysis // Sankhya: Indian J. Stat. Series A. — 1964. — V. 26. — No 4. — P. 359–372.

292. Wedderburn R. W. M. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method // Biometrika. — 1974. — V.61. — P.439–447.

293. Welsh A. H. On M-processes and M-estimation // Ann. Statist. — 1989. — V.17. — P.337–361.

294. Welsh A. H., Ronchetti E. A journey in single steps: robust one-step M-estimation in linear regression // J. Statist. Plann. Infer. — 2002. — V. 103. — No 1-2. — P.287–310.

295. Westfall P. H., Arias A. L. Understanding regression analysis: a conditional distribution approach. — Chapman and Hall/CRC, 2020.

296. White H. Some asymptotic results for learning in single hidden-layer feedforward network models // J. Americ. Stat. Assoc. — 1989. — V.84. — P.1003–1013.

297. Williams D. A. Generalized linear model diagnostics using the deviance and single case deletions // J. R. Stat. Soc. Ser. C. Appl. Stat. — 1987. — V.36. — No 2. — P.181–191.

298. Wu J. S., Chu C. K. Nonparametric estimation of a regression function with dependent observations // Stoch. Proc. Their Appl. — 1994. — V.50. — P.149–160.

299. Wu H., Zhang J.-T. Nonparametric regression methods for longitudinal data analysis: mixed-effects modeling approaches. — John Wiley and Sons, 2006.

300. Wu Y., Wang X., Balakrishnan N. On the consistency of the P–C estimator in a nonparametric regression model // Stat. Papers. — 2020. — V.61. — P.899–915.

301. Xie M., Yang Y. Asymptotics for generalized estimation equations with large clusters size // Ann. Statist. — 2003. — V.31. — No 1. — P.310–347.

302. Yang Y. Asymptotics of M-estimation in non-linear regression // Acta Math. Sinica. — 2004. — V. 20. — No 4. — P.749–760.

303. Yang M. On the de la Garza phenomenon // Ann. Statist. — 2010. — V.38. — P.2499–2524.

304. Yang M., Stufken J. Support points of locally optimal designs for nonlinear models with two parameters // Ann. Statist. — 2009. — V.37. — P.518–541.

305. Yang X., Yang S. Strong consistency of non parametric kernel regression estimator for strong mixing samples // Commun. Stat. Theory Methods — 2016. — V.46. — P.10537–10548.

306. Yao F. Asymptotic distributions of nonparametric regression estimators for longitudinal or functional data // J. Multivariate Anal. — 2007. — V.98. — No 1. — P.40–56.

307. Yao F., Lee T. Penalized spline models for functional principal component analysis // J. R. Stat. Soc. Ser. B. Stat. Methodol. — 2006. — V.68. — P.3–25.

308. Yao F., Muller H.-G., Wang J.-L. Functional data analysis for sparse longitudinal data // J. Amer. Statist. Assoc. — 2005. — V.100. — P.577–590.

309. Yohai V., Maronna R. Asymptotic behavior of M-estimators for linear model // Ann. Statist. — 1979. — V.7. — No 2. — P.258–268.

310. Young D. S. Handbook of regression methods. — Chapman and Hall, 2017.

311. Zhang J.-T., Chen J. Statistical inferences for functional data // Ann. Statist. — 2007. — V.35. — P.1052–1079.

312. Zhang S., Hou T., Qu C. Complete consistency for the estimator of nonparametric regression model based on martingale difference errors // Commun. Stat. Theory Methods. — 2019. — V.50. — No 2. — P.358–370.

313. Zhang S., Miao Y., Xu X., Gao Q. Limit behaviors of the estimator of nonparametric regression model based on martingale difference errors // J. Korean Stat. Soc. — 2018. — V.47. — No 4. — P.537–547.

314. Zhang X., Wang J.-L. Optimal weighting schemes for longitudinal and functional data // Stat. Prob. Lett. — 2018. — V.138. — P.165–170.

315. Zhang X., Wang J.-L. From sparse to dense functional data and beyond // Ann. Statist. — 2016. — V.44. — No 5. — P. 2281–2321.

316. Zhao Q. Restricted regression quantiles // J. Multivariate Anal. — 2000. — V.72. — P.78–99.

317. Zhao Q. Asymptotically efficient median regression in the presence of heteroskedasticity of unknown form // Econ. Theory. — 2001. — V.17. — No 4. — P.765–784.

318. Zhao Z., Yao W. Sequential design for nonparametric inference // Can. J. Stat. — V.40. — No 2. — P.362–377.

319. Zheng S., Yang L., Hardle W. A smooth simultaneous confidence corridor for the mean of sparse functional data // J. Amer. Statist. Assoc. — 2014. — V.109. — P.661–673.

320. Zhou L., Lin H., Liang H. Efficient estimation of the nonparametric mean and covariance functions for longitudinal and sparse functional data // J. Amer. Statist. Assoc. — 2018. — V.113. — P.1550–1564.

321. Zhou H., Yao F., Zhang H. Functional linear regression for discretely observed data: from ideal to reality // Biometrika. — 2023. — V.110. — No 2. — P. 381–393.

322. Zhou X., Zhu F. Asymptotics for L1-wavelet method for nonparametric regression // J. Inequal. Appl. — 2020. — V.216.

323. Zou H., Li R. One-step sparse estimates in nonconcave penalized likelihood models // Ann. Statist. — 2008. — V.36. — No 4.— P.1509–1533.

324. Боровков А.А. Теория вероятностей. — УРСС, 2009.

325. Боровков А.А. Математическая статистика. — М: Физматлит, 2007.

326. Гаек Я., Шидак З. Теория ранговых критериев. — М.: Наука, 1971.

327. Гилл Ф., Мюррей У., Райт М. Практическая оптимизация. — М.: Мир, 1985.

328. Демиденко Е.З. Линейная и нелинейная регрессия. — М.: Финансы и статистика, 1981.

329. Демиденко Е.З. Оптимизация и регрессия. — М.: Наука, 1989.

330. Ермаков С.М., Жиглявский. Математическая теория оптимального эксперимента. М.: Наука, 1987.

331. Ермоленко К.В., Саханенко А.И. Явные асимптотически нормальные оценки неизвестного параметра частично-линейной регрессии // Сиб. электрон. матем. изв. — 2013. — Т.10. — С.719–726.

332. Закс Ш. Теория статистических выводов. — М.:Мир, 1975.

333. Каленчук А.А., Саханенко А.И. О существовании явных асимптотически нормальных оценок неизвестного параметра логарифмической регрессии // Сиб. электрон. матем. изв. — 2015. — Т.12. —С.874–883.

334. Колмогоров А.Н., Фомин С.В. Элементы теории функций и функционального анализа. — М.: Наука, 1976.

335. Корниш–Боуден Э. Основы ферментативной кинетики. — М.: Мир, 1979.

336. Линке Ю.Ю., Саханенко А.И. Об асимптотике распределения одного класса двухшаговых статистических оценок многомерного параметра // Мат. труды. — 2013. — Т.16. — №1. — С.89–120.

337. Линке Ю.Ю., Саханенко А.И. Об условиях асимптотической нормальности одношаговых оценок Фишера для однопараметрических семейств распределений // Сиб. электрон. матем. изв. — 2014. — Т.11. — С.464–475.

338. Линке Ю.Ю., Саханенко А.И. Об условиях асимптотической нормальности одношаговых М-оценок // Сиб. журн. чист. и прикл. матем. — 2016. — Т.16. — №4. — С.46–64.

339. Линке Ю.Ю., Саханенко А.И. Асимптотически нормальное оценивание многомерного параметра в задаче дробно–линейной регрессии // Сиб. Матем. Жур. — 2001. — Т. 42. — № 2. — С.372–388.

340. Линке Ю.Ю., Саханенко А.И. Явное асимптотически нормальное оценивание параметров уравнения Михаэлиса–Ментен // Сиб. Матем. Жур. — 2001. — Т.42. — №3. — С.610–633.

341. Линке Ю.Ю., Саханенко А.И. Асимптотически нормальное оценивание параметра в задаче дробно–линейной регрессии // Сиб. Матем. Жур. — 2000. — Т.41. — №1. — С.150–163.

342. Линке Ю.Ю., Саханенко А.И. Асимптотически нормальное оценивание параметра в задаче дробно–линейной регрессии со случайными ошибками в коэффициентах // Сиб. Матем. Жур. — 2008. — Т.49. — №3. — С.592–619.

343. Математическая теория планирования эксперимента // Под ред. С.М. Ермакова. М.: Наука, 1983.

344. Мелас В.Б., Старосельский Ю.М. Исследование максиминно эффективных планов для модели Михаэлиса—Ментена // Вестн. С.–Петерб. ун-та, серия 1. — 2007. — вып. 2. — С.41-50.

345. Овчаренко С.С., Саханенко А.И. Несостоятельность оценок Йохансена–Лаври для параметров уравнения Михаэлиса–Ментен // Вестник ЮГУ. — 2008. — Вып.1(8). — С.95–99.

346. Савинкина Е.М., Саханенко А.И. Явные оценки неизвестного параметра в одной задаче степенной регрессии // Сиб. электрон. матем. изв. — 2014. — Т.11. — С.725–733.

347. Федоров В.В. Теория оптимального эксперимента. — М.: Наука, 1971.

348. Феллер В. Введение в теорию вероятностей и ее приложения. Т.1. — М.: Мир, 1984.

349. Феллер В. Введение в теорию вероятностей и ее приложения. Т.2. — М.: Мир, 1984.

350. Хьюбер П. Робастность в статистике. — М.:Мир, 1984.

Работы автора по теме диссертации

351. Линке Ю. Ю., Борисов И. С. Универсальные непараметрические ядерные оценки для функций среднего и ковариации случайного процесса // Теория вероятн. и ее примен. — 2024. — Т.69. — № 1. — С.46–75.

   Постановки задач и общие подходы к их решению принадлежат Ю. Ю. Линке, И. С. Борисовым предложена идея оценивания функции ковариации в случае разреженных данных. Доказательства выполнила Ю. Ю. Линке.

352. Линке Ю. Ю. К вопросу о нечувствительности оценок Надарая–Ватсона относительно корреляции элементов дизайна // Теория вероятн. и ее примен. — 2023. — Т.68. — № 2. — С.236–252.

   Linke Yu. Yu. Towards insensitivity of Nadaraya–Watson estimators to design correlation // Theory Probab. Appl. — 2023. — V.68 — № 2. — P.198–210.

353. Линке Ю. Ю. О достаточных условиях состоятельности локально–линейных ядерных оценок // Матем. заметки. — 2023. — Т.114. — № 3. — С.353–369.

   Linke Yu. Yu. On sufficient conditions for the consistency of local linear kernel estimators // Math. Notes. — 2023. — V.114. — № 3. — P.283–296.

354. Linke Yu. Yu., Borisov I. S., Ruzankin P. S. Universal kernel-type estimation of random fields. // Statistics. — 2023. — V.57. — № 4. — P.785–810.

   И. С. Борисовым предложена постановка задачи. Все доказательства выполнила Ю. Ю. Линке. П. С. Рузанкиным получены результаты численного анализа.

355. Линке Ю. Ю. Оценивание функции среднего для зашумленного случайного процесса при наличии разреженных данных // Чебышевский сб. – 2023. – Т.24 – № 5. – С.112–125.

356. Linke Yu. Yu., Borisov I. S. An approach to constructing explicit estimators in nonlinear regression // Siberian Adv. Math. — 2023. — V.33. — № 4. — P.338–346.

   И. С. Борисову принадлежит идея использования непараметрических ядерных методов в задаче оценивания конечномерных параметров нелинейной регрессии. Ю. Ю. Линке выполнила доказательства всех утверждений.

357. Linke Yu. Yu. Kernel estimators for the mean function of a stochastic process under sparse design conditions // Siberian Adv. Math. — 2022. — V.32. — № 4. — P.269–276.

358. Linke Y., Borisov I., Ruzankin P., Kutsenko V., Yarovaya E., Shalnova S. Universal local linear kernel estimators in nonparametric regression // Mathematics. — 2022. — V.10. — № 15. — P.2693.

Идея написания работы принадлежит Е. Б. Яровой. И. С. Борисовым доказаны результаты раздела 4 (*предложения* 2, 3 *и следствия* 3, 4). *Все остальные теоретические результаты* (*включая поставку задачи, конструкцию универсальных локально-линейных ядерных оценок и доказательства основных результатов*) *получены Ю. Ю. Линке. Компьютерное моделирование, а также обработку реальных данных провели В. А. Куценко и П. С. Рузанкин. Реальные данные предоставлены Е. Б. Яровой и С. А. Шальновой.*

359. Linke Yu. Yu., Borisov I. S. Insensitivity of Nadaraya–Watson estimators to design correlation // Commun. Stat. Theory Methods. — 2022. — V.51(19). — P.6909–6918.

Постановки задач предложены И. С. Борисовым, доказательства выполнила Ю. Ю. Линке.

360. Borisov I. S., Linke Yu. Yu., Ruzankin P. S. Universal weighted kernel-type estimators for some class of regression models // Metrika. — 2021. — V. 84. — № 2. — P.141–166.

И. С. Борисовым доказаны леммы 2 *и* 3 *из раздела* 5. *Постановка задачи, конструкция новых оценок и доказательства остальных результатов принадлежат Ю. Ю. Линке. П. С. Рузанкиным проведено компьютерное моделирование.*

361. Linke Yu. Yu. Asymptotic properties of one-step M-estimators // Commun. Stat. Theory Methods. — 2019. — V.48. — № 16. — P.4096–4118.

362. Linke Yu. Yu., Borisov I. S. Toward the notion of intrinsically linear models in nonlinear regression // Siberian Adv. Math. — 2019. — T.29. — № 3. — P.210–216.

И. С. Борисовым доказана несостоятельность оценок параметров трансформирующихся моделей. Все остальные результаты получены Ю. Ю. Линке.

363. Линке Ю. Ю., Борисов И. С. Построение явных оценок в задачах нелинейной регрессии // Теория вероятн. и ее примен. — 2018. — Т.63. — № 1. — С.29–56.

Linke Yu. Yu., Borisov I. S. Constructing explicit estimators in nonlinear regression models // Theory Probab. Appl. — 2018 — V.63 — № 1. — P.22–44.

И. С. Борисовым доказаны результаты разделов 4.2 (*следствия* 2–5) *и следствие* 9 *из раздела* 6.1. *Все остальные результаты получены автором диссертации.*

364. Линке Ю. Ю. Асимптотические свойства одношаговых взвешенных М-оценок с приложениями к задачам регрессии // Теория вероятн. и ее примен. — 2017. — Т.62. — № 3. — С.468–498.

Linke Yu. Yu. Asymptotic properties of one-step weighted M-estimators with application to some regression problems // Theory Probab. Appl. — 2018. — V.62. — № 3. — P.373–398.

365. Linke Yu.Yu. Asymptotic normality of one-step M-estimators based on non-identically distributed observations // Statist. Probab. Lett. — 2017. — V.129. — P.216–221.

366. Линке Ю. Ю. Двухшаговое оценивание параметра в одной неоднородной линейной регрессионной модели. // Сиб. журн. чист. и прикл. матем. — 2017. — Т.17. — № 2. — С.39–51.

Linke Yu. Yu. Two-step estimation in a heteroscedastic linear regression model // J. Math. Sci. — 2018. — V.231. — № 2. — P.206–217.

367. Linke Yu. Yu., Borisov I. S. Constructing initial estimators in one-step estimation procedures of nonlinear regression // Statist. Probab. Lett. — 2017. — V.120. — P.87–94.

И. С. Борисовым доказана теорема 5. *Ю. Ю. Линке предложена методика построения оценок и получены все остальные результаты.*

368. Линке Ю. Ю. Об уточнении одношаговых оценок Фишера в случае медленно сходящихся предварительных оценок // Теория вероятн. и ее примен. — 2015. — Т.60. — № 1. — С.80–98.

Linke Yu. Yu. Refinement of Fisher's one-step estimators in the case of slowly converging preliminary estimators // Theory Probab. Appl. — 2016. — V.60. — № 1. — P.88–102.

369. Линке Ю. Ю. Об асимптотике распределения двухшаговых статистических оценок // Сиб. матем. журн. — 2011. — Т. 52. – № 4 — С.841–860.

Linke Yu. Yu. On the asymptotics of distributions of two-step statistical estimates // Siberian Math. J. — 2011. — V.52. — № 4. — P.665–681.