

Федеральное государственное образовательное учреждение
высшего профессионального образования
«Московский физико-технический институт
(государственный университет)»

На правах рукописи
УДК 519.221, 519.226.3

Панов Максим Евгеньевич

**Точность гауссовской аппроксимации
апостериорного распределения в теореме
Бернштейна - фон Мизеса**

01.01.05 – теория вероятностей и математическая статистика

ДИССЕРТАЦИЯ

на соискание ученой степени

кандидата физико-математических наук

Научный руководитель

кандидат физико-математических наук

Спокойный Владимир Григорьевич

Москва – 2015

Оглавление

Введение	4
Глава 1. Теорема Бернштейна – фон Мизеса для конечномерного случая	12
1.1. Постановка задачи	12
1.2. Семипараметрическая теорема Бернштейна – фон Мизеса в случае конечного размера выборки и конечной размерности мешающего параметра	17
1.3. Расширение результата теоремы БфМ на случай гладкого априорного распределения	32
1.4. Доказательства	34
Глава 2. Модель независимых одинаково распределенных случайных величин и критическая размерность	53
2.1. Независимые одинаково распределенные случайные величины	53
2.2. Доказательства	60
Глава 3. Теорема Бернштейна – фон Мизеса в случае бесконечной размерности мешающего параметра	62
3.1. Метод проекционных оценок	63
3.2. Доказательство теоремы 12	68
Глава 4. Примеры	74
4.1. Линейная гауссовская регрессия и плоское гауссовское априорное распределение	74
4.2. Линейная негауссовская регрессия	75
4.3. Семипараметрическая негауссовская линейная регрессия	78
4.4. Обобщенные линейные модели	80

4.5. Доказательства	85
Заключение	89
Список литературы	90

Введение

Актуальность темы исследования. Байесовский подход является одним из центральных направлений развития современной математической статистики. В данном подходе изучается апостериорное распределение параметров модели, т.е. распределение, получаемое в результате уточнения априорного распределения по результатам наблюдения данных. Теорема Бернштейна — фон Мизеса (БфМ) утверждает асимптотическую близость апостериорного распределения к нормальному со средним, близким к оценке максимума правдоподобия, и с апостериорной ковариационной матрицей, близкой к обратной информационной матрице Фишера. Теорема БфМ дает теоретическое обоснование байесовских вычислений оценки максимума правдоподобия и ее ковариации. Также она обосновывает использование эллиптических доверительных множеств, основанных на первых двух моментах апостериорного распределения. Классическая версия теоремы БфМ формулируется для стандартной параметрической постановки с фиксированной параметрической моделью и большими размерами выборки (см. подробный обзор в книгах Ле Кама [1] и Ван дер Ваарта [2]). Однако в современных статистических приложениях часто встречаются очень сложные модели, включающие большое количество параметров, причем доступный размер выборки, как правило, очень ограничен (см. подробный обзор современной статистики для данных большой размерности в книге Бюльманна и Ван де Гир [3]). Таким образом, возникает необходимость расширения классических результатов на такие неклассические ситуации. Отметим работы Кокса [4], Фридмана [5], Бушерона и Гассья [6] и Госала [7], в которых рассмотрены некоторые особенности байесовского анализа в моделях с растущей размерностью параметра. Уже решение вопроса о том, является ли апостериорное распределение в непараметрических и семипараметрических моделях состоятельным, представляется непростой задачей (см. работы Шварца [8], Баррона [9] и Бочкиной [10]). Еще более трудным является вопрос асимптотической нормальности апостериорной меры (см., например, работу Ше-

ня [11]). Некоторые результаты для конкретных семи- и непараметрических моделей можно найти в работах Кима [12] [13], Леу [14], Кастилло и Никля [15]. В работе Ченга и Косорока [16] получен вариант теоремы БфМ, основанный на разложении профайл-правдоподобия (profile likelihood). В недавней работе Бикеля и Кляйна [17] теорема БфМ доказана для достаточно широкого класса моделей с независимыми одинаково распределенными случайными величинами. В работе Кастилло [18] изучается асимптотическая нормальность апостериорного распределения целевого параметра в семипараметрических моделях, в которых функциональный параметр порожден гауссовским процессом. В работе Ривуарара и Руссо [19] семипараметрическая теорема БфМ доказана для линейных функционалов плотности распределения, а в работе Кастилло и Руссо [20] результат обобщен для более широкого класса моделей и функционалов. Также в другой работе Ривуарара и Руссо [21] изучена скорость концентрации апостериорного распределения в случае распределения данных из экспоненциального семейства. Беллони и Черножуков [22] изучили асимптотическую нормальность апостериорного распределения для экспоненциальных семейств в случае растущей размерности. Однако все эти результаты ограничены их применимостью только к асимптотическому случаю или к отдельным классам моделей, таких как гауссовские модели, модели из экспоненциального семейства или модели с независимыми одинаково распределенными наблюдениями.

В данной работе доказывается вариант теоремы БфМ для достаточно широкого класса параметрических и семипараметрических моделей. Важной особенностью нашего исследования является предположение о фиксированном размере выборки. В классической теории обычно предполагается выполнение условий *локальной асимптотической нормальности*, причем рассматриваются модели с фиксированной конечной размерностью полного параметра, а размер выборки предполагается стремящимся к бесконечности, см. книги Ле Кама и Янга [1] и Ибрагимова и Хасьминского [23]. Отметим также работы Гусева [24, 25], в которых в модели независимых одинаково распределенных случайных величин

были подробно рассмотрены асимптотические разложения апостериорных плотностей распределения, моментов случайных величин и рисков байесовских оценок. В дальнейшем асимптотические разложения второго порядка для байесовских оценок в схеме независимых наблюдений были подробно исследованы Бурнашевым [26]. Построение теории для работы с конечными выборками является сложной задачей, так как большинство подходов и методов в классической теории разработаны для асимптотического случая, подразумевающего стремящийся к бесконечности размер выборки. Известно лишь небольшое число результатов для конечных размеров выборки (см., например, недавнюю статью Бушерона и Массара [27]). Другой особенностью нашего исследования являются учет возможной неверной спецификации модели, т.е. ситуации, в которой истинное распределение данных не принадлежит рассматриваемому параметрическому семейству. Учет неверной спецификации модели также слабо представлен в литературе, см. работу Кляйна и ван дер Варта [28].

В данной работе рассматривается семипараметрическая задача, в которой размерность полного параметра велика или бесконечна, а целевой параметр имеет меньшую размерность. Компоненту полного вектора параметров, ортогональную пространству целевого параметра, называют мешающим параметром. В байесовском подходе целью семипараметрического оценивания является маргинальное распределение целевого параметра (см. работу Кастилло [18]). Типичными примерами являются оценивание функционалов, оценивание значения функции в точке или просто оценивание заданного подвектора вектора параметров. Интересной особенностью семипараметрической теоремы БфМ является тот факт, что мешающий параметр входит в результат только через проекцию нормированного градиента логарифма правдоподобия на целевое подпространство и через эффективную информацию Фишера (см. работу Бикеля и Кляйна [17]). Обычно методы изучения в данном случае основываются на понятии наилучшей параметрической подмодели (см. обзор в книге Косорока [29]). Более того, предполагается, что существует метод оценивания мешающего параметра, достигающий определенной

скорости сходимости оценки к истинному значению (см. работу Ченга и Косорока [16]). Такое предположение сильно упрощает работу с задачей, но не позволяет вывести качественные соотношения между полной размерностью целевого пространства и содержащейся в данных информацией.

Сформулируем цели данной работы:

1. Разработать подход к построению неасимптотических оценок близости апостериорного распределения к нормальному для широкого класса статистических моделей.
2. Исследовать особенности семипараметрического байесовского оценивания и их влияние на апостериорное распределение целевого параметра.
3. Математически исследовать границы применимости теоремы БфМ в моделях с большой, в том числе растущей размерностью полного параметра.

Для достижения поставленных целей были определены следующие **задачи** исследования:

1. Вычислить ошибку аппроксимации апостериорного распределения гауссовским распределением для общего случая гладкой семипараметрической статистической модели с конечной размерностью мешающего параметра и равномерного априорного распределения параметров.
2. Исследовать зависимость полученной ошибки аппроксимации от размерности задачи и размера выборки для ряда статистических моделей в случае конечной размерности полного параметра.
3. Рассмотреть случай гауссовского априорного распределения, которое приводит к смещению апостериорного распределения, и количественно изучить эффект смещения.
4. Обобщить полученные результаты на случай семипараметрических моделей с бесконечной размерностью мешающего параметра.

5. Показать применимость общих теоретических результатов к конкретным статистическим моделям.

Научная новизна результатов, полученных в диссертации, состоит в том, что разработан новый метод оценки близости апостериорного распределения к гауссовскому распределению в параметрических и семипараметрических задачах. Основной особенностью подхода является оценка ошибки аппроксимации в случае конечного размера выборки даже для тех ситуаций, когда размерность параметра увеличивается с ростом размера выборки, а параметрическая модель может быть неверно специфицирована. Впервые для настолько широкого класса статистических моделей показано, что ошибка аппроксимации мала, если величина p^2q/n мала, где p – полная размерность задачи, q – размерность целевого параметра и n – размер выборки. Таким образом, размерность $p^2q = O(n)$ является критической для результата теоремы БфМ. Также получены новые условия для выполнения теоремы БфМ в случае гауссовского априорного распределения, а также в семипараметрических моделях с бесконечномерным мешающим параметром при дополнительном предположении о гладкости непараметрической части.

Теоретическая и практическая значимость. Результаты диссертации дают основу для анализа байесовских методов статистики с учетом конечного размера наблюдаемой выборки и возможной неверной спецификации модели. С практической точки зрения результаты позволяют дать обоснование применению методов построения доверительных множеств на основе первых двух моментов апостериорного распределения.

На защиту выносятся следующие результаты:

1. Вычислена ошибка аппроксимации апостериорного распределения гауссовским распределением для случая гладкой семипараметрической статистической модели со стохастической частью, удовлетворяющей условиям типа конечности экспоненциальных моментов, в случае конечной размерности мешающего параметра и равномерного априорного распределения параметров.

2. Показано, что для модели независимых одинаково распределенных случайных величин, линейных и обобщенных линейных моделей полученная ошибка аппроксимации зависит от размерности задачи p , размерности целевого параметра q и размера выборки n как $\sqrt{p^2q/n}$, что позволяет определить критическую для выполнения теоремы БфМ размерность параметрического множества.
3. Показано, что если гауссовское распределение является достаточно плоским, то результат теоремы БфМ остается в силе, как и в случае равномерного распределения.
4. С помощью метода проекционных оценок результаты обобщены на случай семипараметрических моделей с бесконечномерным мешающим параметром.
5. Показана применимость общих теоретических результатов к линейным и обобщенным линейным моделям с мешающим параметром, принадлежащим соболевскому классу гладкости.

Апробация результатов. Результаты диссертации докладывались и обсуждались на следующих конференциях:

- 2nd Conference of International Society of Nonparametric Statistics (2014, Кадис, Испания);
- SAMSI-CRM Workshop on Geometric Aspects of High-dimensional Inference (2014, Дурхэм, Северная Каролина, США);
- Meeting in Mathematical Statistics: New Procedures for New Data (2014, Люмини, Франция);
- Conference on Structural Inference in Statistics (2013, Потсдам, Германия);
- 36-я Международная конференция молодых ученых “Информационные технологии и системы” (2013, Калининград, Россия);

- 55-я Всероссийская Научная конференция Московского физико-технического института (2012, Долгопрудный, Россия).
- 57-я Всероссийская Научная конференция Московского физико-технического института (2014, Долгопрудный, Россия).

Также результаты работы обсуждались на семинарах Лаборатории структурных методов анализа данных в предсказательном моделировании МФТИ (2013-2015), семинаре Международной лаборатории стохастического анализа и его приложений НИУ ВШЭ (2015), городском семинаре по теории вероятностей и математической статистике города Санкт-Петербурга (2015).

Публикации. Основные результаты по теме диссертации изложены в 7 печатных работах, из которых 3 изданы в журналах, рекомендованных ВАК [30–32].

Личный вклад автора. Содержание диссертации и основные положения, выносимые на защиту, отражают персональный вклад диссертанта в опубликованные работы. Постановка задач и предложение общих подходов к их решению осуществлялась научным руководителем. Подготовка к публикации полученных результатов проводилась совместно с соавторами, причем вклад диссертанта был определяющим.

Так, в работе [30] идея доказательства оценки ошибки аппроксимации и принципа больших уклонений для полного параметра восходит к работе Спокойного [33], но доказательства теорем, строго обосновывающих перечисленные идеи и конструкции для семипараметрической постановки, получены лично диссертантом.

В работе [31] все основные результаты, включая обобщение результатов, полученных для случая равномерного априорного распределение, на случай плоского гауссовского распределения, а также обобщение результатов на бесконечномерного полного параметра, принадлежат диссертанту.

Идея подхода к улучшению оценок аппроксимации в работе [32] и его прак-

тическая реализация также принадлежат лично диссертанту.

Структура и объем диссертации. Диссертация состоит из введения, обзора литературы, 4 глав, заключения и библиографии. Общий объем диссертации 94 страницы. Библиография включает 59 наименований.

Благодарности. Автор благодарен своему научному руководителю Владимиру Григорьевичу Спокойному за всегда интересные и плодотворные обсуждения, постоянную поддержку и участие.

Работа выполнена при поддержке Лаборатории структурного анализа данных в предсказательном моделировании, МФТИ, грант правительства РФ дог. 11.G34.31.0073.

Теорема Бернштейна – фон Мизеса для конечномерного случая

1.1. Постановка задачи

Обозначим через $\mathbf{Y} \in \mathbb{R}^n$ наблюдаемые случайные данные и через \mathbb{P} – их распределение. Параметрическая статистическая модель предполагает, что неизвестное распределение данных \mathbb{P} принадлежит к заданному параметрическому семейству $(\mathbb{P}_{\mathbf{v}})$:

$$\mathbf{Y} \sim \mathbb{P} = \mathbb{P}_{\mathbf{v}^*} \in (\mathbb{P}_{\mathbf{v}}, \mathbf{v} \in \mathcal{Y}),$$

где \mathcal{Y} – это пространство параметров и $\mathbf{v}^* \in \mathcal{Y}$ – истинное значение параметра. В семипараметрическом случае целью оценивания является только низкоразмерная компонента $\boldsymbol{\theta}$ полного параметра \mathbf{v} . Таким образом целью оценивания является $\boldsymbol{\theta}^* \stackrel{\text{def}}{=} \Pi_0 \mathbf{v}^*$ для некоторого отображения $\Pi_0: \mathcal{Y} \rightarrow \mathbb{R}^q$, где $q \in \mathbb{N}$ – размерность целевого параметра. Обычно в классическом семипараметрическом подходе вектор \mathbf{v} представляется в виде $\mathbf{v} = (\boldsymbol{\theta}, \boldsymbol{\eta})$, где $\boldsymbol{\theta}$ является целью оценивания, а $\boldsymbol{\eta}$ является *мешающим параметром*. Мы будем называть такую ситуацию $(\boldsymbol{\theta}, \boldsymbol{\eta})$ -*моделью* и в дальнейшем будем работать в таком предположении. Переход к более общей \mathbf{v} -модели, в которой $\boldsymbol{\theta} = \Pi_0 \mathbf{v}$, также может быть выполнен в рамках нашего подхода, но требует выполнения некоторых технических условий на оператор Π_0 (см., например, работу Ермакова [34]). Также для простоты мы сначала выведем наши результаты для случая, когда пространство полного параметра \mathcal{Y} является подмножеством евклидова пространства размерности p .

Кроме того, в данной работе мы обращаемся к проблеме неверной спецификации модели. В большинстве практических задач представляется нереалистичным, что модельные предположения выполнены в точности даже в случае использова-

ния богатых параметрических семейств. Это означает, что распределение \mathbb{P} не принадлежит к рассматриваемому семейству $(\mathbb{P}_{\mathbf{v}}, \mathbf{v} \in \mathcal{Y})$. “Истинное” значение \mathbf{v}^* параметра \mathbf{v} может быть определено как

$$\mathbf{v}^* = \operatorname{argmax}_{\mathbf{v} \in \mathcal{Y}} \mathbb{E} \mathcal{L}(\mathbf{v}),$$

где $\mathcal{L}(\mathbf{v}) = \mathcal{L}(\mathbf{v} \mid \mathbf{Y}) \stackrel{\text{def}}{=} \log \frac{d\mathbb{P}_{\mathbf{v}}}{d\mu_0}(\mathbf{Y})$ – логарифм правдоподобия семейства $(\mathbb{P}_{\mathbf{v}})$ для некоторой доминирующей меры μ_0 . В случае неверной спецификации модели \mathbf{v}^* определяет наилучшее параметрическое приближение к мере \mathbb{P} в рассматриваемом семействе; см. [28, 35, 36] и ссылки в этих работах. Цель оценивания $\boldsymbol{\theta}^*$ по-прежнему определяется отображением Π_0 :

$$\boldsymbol{\theta}^* \stackrel{\text{def}}{=} \Pi_0 \mathbf{v}^*.$$

1.1.1. Параметрическое оценивание: основные определения

Мы предполагаем, что большая положительная константа \mathbf{x} фиксирована таким образом, чтобы задать на пространстве наблюдений множество случайных событий $\Omega(\mathbf{x})$ доминирующей вероятности. Говорят, что множество случайных событий $\Omega(\mathbf{x})$ является *множеством доминирующей вероятности*, если

$$\mathbb{P}(\Omega(\mathbf{x})) \geq 1 - \mathbf{c}e^{-\mathbf{x}}.$$

Одним из основных элементов нашей конструкции является $(p \times p)$ матрица \mathcal{D}^2 , которая определяется аналогично информационной матрице Фишера:

$$\mathcal{D}^2 \stackrel{\text{def}}{=} -\nabla^2 \mathbb{E} \mathcal{L}(\mathbf{v}^*). \quad (1.1)$$

Здесь и далее мы работаем при условиях, которые близки к классическим условиям *регулярного параметрического семейства* (см. книгу Ибрагимова и Хасьминского [23]). Мы неявно предполагаем, что логарифмическая функция правдоподобия $\mathcal{L}(\mathbf{v})$ достаточно гладкая по \mathbf{v} , и обозначаем через $\nabla \mathcal{L}(\mathbf{v})$ ее градиент, а

через $\nabla^2 \mathbb{E}\mathcal{L}(\mathbf{v})$ гессиан математического ожидания $\mathbb{E}\mathcal{L}(\mathbf{v})$. Также положим

$$\boldsymbol{\xi} \stackrel{\text{def}}{=} \mathcal{D}^{-1} \nabla \mathcal{L}(\mathbf{v}^*).$$

Определение \mathbf{v}^* подразумевает, что $\nabla \mathbb{E}\mathcal{L}(\mathbf{v}^*) = 0$ и, следовательно, $\mathbb{E}\boldsymbol{\xi} = 0$.

Для $(\boldsymbol{\theta}, \boldsymbol{\eta})$ -модели рассмотрим блочное представление вектора $\nabla \mathcal{L}(\mathbf{v}^*)$ и матрицы \mathcal{D}^2 из (1.1):

$$\nabla \mathcal{L}(\mathbf{v}^*) = \begin{pmatrix} \nabla_{\boldsymbol{\theta}} \mathcal{L}(\mathbf{v}^*) \\ \nabla_{\boldsymbol{\eta}} \mathcal{L}(\mathbf{v}^*) \end{pmatrix}, \quad \mathcal{D}^2 = \begin{pmatrix} D^2 & A \\ A^\top & H^2 \end{pmatrix}.$$

Определим также $(q \times q)$ -матрицу \check{D}^2 и случайный вектор $\check{\boldsymbol{\xi}} \in \mathbb{R}^q$:

$$\check{D}^2 \stackrel{\text{def}}{=} D^2 - AH^{-2}A^\top, \tag{1.2}$$

$$\check{\nabla}_{\boldsymbol{\theta}} \stackrel{\text{def}}{=} \nabla_{\boldsymbol{\theta}} - AH^{-2}\nabla_{\boldsymbol{\eta}},$$

$$\check{\boldsymbol{\xi}} \stackrel{\text{def}}{=} \check{D}^{-1} \check{\nabla}_{\boldsymbol{\theta}} \mathcal{L}(\mathbf{v}^*). \tag{1.3}$$

Матрица \check{D}^2 размера $q \times q$ обычно называется *эффективной информационной матрицей Фишера*.

Далее везде по ходу изложения $\|\mathbf{a}\|$ обозначает евклидову норму вектора \mathbf{a} , а для матрицы A ее операторная норма будет обозначаться через $\|A\|$. Порядок на квадратных матрицах определяется стандартным образом, т.е. $A > B$ означает, что матрица $A - B$ положительно определена.

1.1.2. Классическая теорема Бернштейна – фон Мизеса

Пусть задано априорное распределение Π на множестве параметров $\mathcal{Y} \subseteq \mathbb{R}^p$. Ниже мы будем изучать свойства апостериорной меры, которая является случайной мерой на \mathcal{Y} , описывающей условное распределение \mathbf{v} при заданном \mathbf{Y} и полученной нормированием произведения $\exp\{\mathcal{L}(\mathbf{v})\} \Pi(d\mathbf{v})$. Это соотношение может быть записано как

$$\mathbf{v} \mid \mathbf{Y} \propto \exp\{\mathcal{L}(\mathbf{v})\} \Pi(d\mathbf{v}). \tag{1.4}$$

Важной особенностью нашего анализа является отсутствие предположения о том, что $\mathcal{L}(\mathbf{v})$ является истинным логарифмом правдоподобия. Это означает, что неверная спецификация модели возможна и распределение данных может не принадлежать параметрическому семейству. В таком случае формула Байеса (1.4) описывает *квази-апостериорное распределение* (см. работу Черножукова [35]). Ниже мы покажем, что гладкость логарифма правдоподобия $\mathcal{L}(\mathbf{v})$ обеспечивает гауссовскую аппроксимацию апостериорной меры. Мы сконцентрируемся на описании точности такой аппроксимации как функции размерности параметра p и других важных характеристик модели.

Предположим, что для априорной меры Π существует положительная плотность $\pi(\mathbf{v})$ по отношению к мере Лебега на \mathcal{Y} : $\Pi(d\mathbf{v}) = \pi(\mathbf{v})d\mathbf{v}$. Тогда (1.4) может быть записано как

$$\mathbf{v} \mid \mathbf{Y} \propto \exp\{\mathcal{L}(\mathbf{v})\} \pi(\mathbf{v}). \quad (1.5)$$

Знаменитая теорема Бернштейна – фон Мизеса (БфМ) утверждает, что апостериорное распределение, центрированное с помощью любой эффективной оценки $\tilde{\mathbf{v}}$ параметра \mathbf{v}^* (например, с помощью оценки максимума правдоподобия) и нормированное с помощью информационной матрицы Фишера, близко к стандартному нормальному распределению:

$$\mathcal{D}(\mathbf{v} - \tilde{\mathbf{v}} \mid \mathbf{Y}) \rightarrow \mathcal{N}(0, I_p),$$

где I_p – единичная матрица размерности p , а сходимость понимается в смысле сходимости по полной вариации.

Важной особенностью апостериорного распределения является тот факт, что оно полностью известно и значения из него можно численно генерировать. Если мы знаем, что апостериорное распределение близко к нормальному, то для построения множеств концентрации и доверительных множеств достаточно подсчитать его среднее значение и матрицу ковариаций. Важно также отметить, что теорема Бернштейна – фон Мизеса не требует, чтобы априорное распределение было

собственным, и это явление может наблюдаться также в случае несобственных распределений (см., например, работу Бочкиной и Грина [37]).

В данной работе мы исследуем свойства апостериорного распределения для целевого параметра $\boldsymbol{\vartheta} = \Pi_0 \boldsymbol{\nu}$ при наблюдении данных \mathbf{Y} . Заметим, что здесь и далее мы будем обозначать целевой параметр как $\boldsymbol{\vartheta}$, когда он имеет смысл апостериорной случайной величины, и как $\boldsymbol{\theta}$ во всех остальных ситуациях. В этом случае (1.5) может быть записано как

$$\boldsymbol{\vartheta} | \mathbf{Y} \propto \int \exp\{\mathcal{L}(\boldsymbol{\nu})\} \pi(\boldsymbol{\nu}) d\boldsymbol{\eta}. \quad (1.6)$$

Утверждение теоремы БфМ в данном случае преобразуется в

$$\check{D}(\boldsymbol{\vartheta} - \tilde{\boldsymbol{\theta}}) | \mathbf{Y} \rightarrow \mathcal{N}(0, I_q),$$

где I_q – единичная матрица размерности q , $\tilde{\boldsymbol{\theta}} = \Pi_0 \tilde{\boldsymbol{\nu}}$, и \check{D}^2 задается формулой (1.2).

Мы рассмотрим два важных класса априорных распределений: неинформативные априорные распределения и гладкие априорные распределения. Нашей целью является доказательство того, что при разумных условиях апостериорное распределение целевого параметра (1.6) близко к нормальному распределению с правильно выбранным средним и дисперсией даже для конечного размера выборки. Другой важным аспектом работы является описание ограничений на размер выборки и размерность параметрического множества, при которых результат теоремы БфМ остается верным.

1.2. Семипараметрическая теорема Бернштейна – фон Мизеса в случае конечного размера выборки и конечной размерности мешающего параметра

1.2.1. Условия

Наш подход предполагает выполнение некоторого количества условий. Список условий близок к работе [38], в которой можно найти обсуждение условий и примеры, показывающие, что условия не ограничительны и выполняются для большинства классических моделей, используемых в статистике, таких как модель независимых одинаково распределенных случайных величин, модель регрессии и обобщенные линейные модели. Условия можно разделить на локальные и глобальные. Локальные условия описывают поведение процесса $\mathcal{L}(\mathbf{v})$ на локальном множестве $\mathbf{v} \in \mathcal{Y}_0(\mathbf{r}_0)$ при некотором фиксированном значении \mathbf{r}_0 , где

$$\mathcal{Y}_0(\mathbf{r}) \stackrel{\text{def}}{=} \{\mathbf{v} \in \mathcal{Y} : \|\mathcal{D}(\mathbf{v} - \mathbf{v}^*)\| \leq \mathbf{r}\}. \quad (1.7)$$

Заметим, что везде ниже будет неявно предполагаться, что точка \mathbf{v}^* является внутренней точкой множества \mathcal{Y} . Глобальные условия должны выполняться на всем \mathcal{Y} . Определим стохастическую компоненту $\zeta(\mathbf{v})$ логарифма правдоподобия $\mathcal{L}(\mathbf{v})$:

$$\zeta(\mathbf{v}) \stackrel{\text{def}}{=} \mathcal{L}(\mathbf{v}) - \mathbb{E}\mathcal{L}(\mathbf{v}).$$

Начнем с условий на конечность экспоненциальных моментов.

(ED₀) Существует константа $\nu_0 > 0$, положительно определенная $(p \times p)$ -матрица \mathcal{V}^2 , удовлетворяющая $\text{Var}\{\nabla\zeta(\mathbf{v}^*)\} \leq \mathcal{V}^2$, и константа $\mathbf{g} > 0$ такие, что

$$\sup_{\gamma \in \mathbb{R}^p} \log \mathbb{E} \exp \left\{ \mu \frac{\gamma^\top \nabla \zeta(\mathbf{v}^*)}{\|\mathcal{V}\gamma\|} \right\} \leq \frac{\nu_0^2 \mu^2}{2}, \quad \forall \mu : |\mu| \leq \mathbf{g}.$$

(ED₂) Существуют константы $\nu_0, \omega > 0$ и для каждого $\mathbf{r} > 0$ константа $\mathbf{g}(\mathbf{r}) >$

0 такие, что для всех $\mathbf{v} \in \mathcal{Y}_0(\mathbf{r})$:

$$\sup_{\gamma_1, \gamma_2 \in \mathbb{R}^p} \log \mathbb{E} \exp \left\{ \frac{\mu}{\omega} \frac{\gamma_1^\top \nabla^2 \zeta(\mathbf{v}) \gamma_2}{\|\mathcal{D}\gamma_1\| \cdot \|\mathcal{D}\gamma_2\|} \right\} \leq \frac{\nu_0^2 \mu^2}{2}, \quad \forall \mu: |\mu| \leq \mathbf{g}(\mathbf{r}).$$

Определим

$$\mathcal{D}^2(\mathbf{v}) \stackrel{\text{def}}{=} -\nabla^2 \mathbb{E} \mathcal{L}(\mathbf{v}).$$

Тогда $\mathcal{D}^2 = \mathcal{D}^2(\mathbf{v}^*)$. Следующее условие необходимо, чтобы обеспечить гладкость математического ожидания логарифма правдоподобия $\mathbb{E} \mathcal{L}(\mathbf{v})$ в локальной области $\mathbf{v} \in \mathcal{Y}_0(\mathbf{r}_0)$:

(\mathcal{L}_0) Для любого $\mathbf{r} \leq \mathbf{r}_0$ существует константа $\delta(\mathbf{r}) > 0$ такая, что на множестве $\mathcal{Y}_0(\mathbf{r})$ выполняется неравенство:

$$\|\mathcal{D}^{-1} \mathcal{D}^2(\mathbf{v}) \mathcal{D}^{-1} - I_p\| \leq \delta(\mathbf{r}).$$

Введем обозначение $L(\mathbf{v}, \mathbf{v}^*) \stackrel{\text{def}}{=} \mathcal{L}(\mathbf{v}) - \mathcal{L}(\mathbf{v}^*)$ для логарифма отношения правдоподобия. Условие глобальной идентификации выглядит следующим образом:

(\mathcal{L}_r) Существует константа $\mathbf{b} > 0$ такая, что для любого $\mathbf{r} > 0$ выполняется неравенство:

$$-\mathbb{E} L(\mathbf{v}, \mathbf{v}^*) \geq \mathbf{b} \mathbf{r}^2, \quad \mathbf{r} = \|\mathcal{D}(\mathbf{v} - \mathbf{v}^*)\|.$$

Также необходимо ввести некоторые условия идентифицируемости. Сначала запишем информационную и ковариационную матрицы в блочной форме:

$$\mathcal{D}^2 = \begin{pmatrix} D^2 & A \\ A^\top & H^2 \end{pmatrix}, \quad \mathcal{V}^2 = \begin{pmatrix} V^2 & B \\ B^\top & Q^2 \end{pmatrix}.$$

Условия идентифицируемости в [38] гарантируют, что матрица \mathcal{D}^2 положительная определена и удовлетворяет условию $\mathbf{a}^2 \mathcal{D}^2 \geq \mathcal{V}^2$ для некоторого $\mathbf{a} > 0$. Здесь мы заново выпишем эти условия в блочной форме, которая является особенностью $(\boldsymbol{\theta}, \boldsymbol{\eta})$ -модели.

(\mathcal{I}) Существуют константы $\mathbf{a} > 0$ и $0 \leq \nu < 1$ такие, что

$$\mathbf{a}^2 D^2 \geq V^2, \quad \mathbf{a}^2 H^2 \geq Q^2, \quad \mathbf{a}^2 \mathcal{D}^2 \geq \nu^2 \quad (1.8)$$

и

$$\|D^{-1} A H^{-2} A^\top D^{-1}\| \leq \nu. \quad (1.9)$$

Величина ν ограничивает угол между подпространствами целевого и мешающего параметров в касательном пространстве. Условие регулярности (\mathcal{I}) гарантирует, что угол не является слишком маленьким, целевой и мешающий параметры идентифицируемы. В частности, матрица \check{D}^2 положительно определена при условии выполнения (\mathcal{I}). Ограничения (1.8) даны для одной и той же константы \mathbf{a} только для упрощения обозначений. Можно показать, что последняя граница для \mathcal{D}^2 следует из первых двух и (1.9) с другой константой \mathbf{a}' , зависящей только от \mathbf{a} и ν .

Наряду с локальным множеством $\Upsilon_0(\mathbf{r})$ определим отдельно локальные множества для целевого и мешающего параметров:

$$\begin{aligned} \Theta_0(\mathbf{r}) &\stackrel{\text{def}}{=} \{\boldsymbol{\theta}: \|\check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\| \leq \mathbf{r}\}, \\ \mathcal{H}_0(\mathbf{r}) &\stackrel{\text{def}}{=} \{\check{\boldsymbol{\eta}}: \|H(\check{\boldsymbol{\eta}} - \check{\boldsymbol{\eta}}^*)\| \leq \mathbf{r}\}, \end{aligned}$$

где $\check{\boldsymbol{\eta}} \stackrel{\text{def}}{=} \boldsymbol{\eta} + H^{-2} A^\top \boldsymbol{\theta}$ и $\check{\boldsymbol{\eta}}^* \stackrel{\text{def}}{=} \boldsymbol{\eta}^* + H^{-2} A^\top \boldsymbol{\theta}^*$. Переход от параметров $(\boldsymbol{\theta}, \boldsymbol{\eta})$ к модифицированным параметрам $(\boldsymbol{\theta}, \check{\boldsymbol{\eta}})$ позволяет учесть взаимодействие между целевым и мешающим параметром. Заметим, что выполняется равенство $\|\mathcal{D}(\mathbf{v} - \mathbf{v}^*)\|^2 = \|\check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\|^2 + \|H(\check{\boldsymbol{\eta}} - \check{\boldsymbol{\eta}}^*)\|^2$. Таким образом можно получить следующее представление

$$\Upsilon_0(\mathbf{r}) = \Theta_0(\mathbf{r}) \times \mathcal{H}_0(\mathbf{r}). \quad (1.10)$$

Также определим анизотропное локальное множество, основанное на представлении (1.10):

$$\Upsilon_0(\mathbf{h}, \mathbf{r}) = \Theta_0(\mathbf{h}) \times \mathcal{H}_0(\mathbf{r}).$$

Заметим, что при $\mathbf{h} < \mathbf{r}$ выполняется

$$\mathcal{Y}_0(\mathbf{h}, \mathbf{r}) \subset \mathcal{Y}_0(\mathbf{r}).$$

Использование анизотропного локального множества $\mathcal{Y}_0(\mathbf{h}, \mathbf{r})$ позволяет учесть особенности семипараметрической задачи.

1.2.2. Брэкетинг и метод верхней функции

В данном пункте представлен краткий обзор основных элементов конструкции из работы [38], включая оценку брэкетинга и результаты, основанные на методе верхней функции. Также данные результаты обобщены и уточнены для семипараметрического случая. Оценка брэкетинга описывает качество квадратичной аппроксимации логарифма правдоподобия $\mathcal{L}(\mathbf{v})$ в локальной окрестности точки \mathbf{v}^* , а метод верхней функции используется, чтобы показать, что оценка максимума правдоподобия $\tilde{\mathbf{v}}$ принадлежит этой окрестности с доминирующей вероятностью. Вспомним определение локального множества (1.7):

$$\mathcal{Y}_0(\mathbf{r}) = \{\mathbf{v} \in \mathcal{Y} : \|\mathcal{D}(\mathbf{v} - \mathbf{v}^*)\|^2 \leq \mathbf{r}^2\}.$$

Формулировки результатов подразумевают, что зафиксировано такое значение \mathbf{x} , что величина $e^{-\mathbf{x}}$ является достаточно малой. Если размерность p велика, то можно положить $\mathbf{x} = \mathbf{C} \log p$. Предположим, что значение $\mathbf{r} = \mathbf{r}_0$ фиксировано таким образом, чтобы разделить локальную и глобальную зоны. Как базовый результат мы будем использовать следующую теорему из [39]:

Теорема 1 (Спокойный, 2013). *Предположим, что выполнены условия (ED_2) и (\mathcal{L}_0) . Тогда для всех $\mathbf{r} \leq \mathbf{r}_0$*

$$\mathbb{P}\left\{ \sup_{\mathbf{v} \in \mathcal{Y}_0(\mathbf{r})} \|\mathcal{D}^{-1}\{\nabla \mathcal{L}(\mathbf{v}) - \nabla \mathcal{L}(\mathbf{v}^*)\} + \mathcal{D}(\mathbf{v} - \mathbf{v}^*)\| \geq \diamond(\mathbf{r}, \mathbf{x}) \right\} \leq e^{-\mathbf{x}}, \quad (1.11)$$

где

$$\begin{aligned} \diamond(\mathbf{r}, \mathbf{x}) &\stackrel{\text{def}}{=} \{\delta(\mathbf{r}) + 6\nu_0 z_{\mathbb{H}}(\mathbf{x}) \omega\} \mathbf{r}, \\ z_{\mathbb{H}}(\mathbf{x}) &\stackrel{\text{def}}{=} 2p^{1/2} + \sqrt{2\mathbf{x}} + \mathbf{g}^{-1}(\mathbf{g}^{-2}\mathbf{x} + 1)4p. \end{aligned} \quad (1.12)$$

Данный результат позволяет ограничить ошибку локальной линейной аппроксимации градиента логарифма правдоподобия равномерно для $\mathbf{v} \in \mathcal{Y}_0(\mathbf{r}_0)$. Величина $\diamond(\mathbf{r}_0, \mathbf{x})$ измеряет качество аппроксимации градиента логарифма отношения правдоподобия $\nabla L(\mathbf{v}, \mathbf{v}^*)$ линейным процессом $\mathcal{D}(\mathbf{v} - \mathbf{v}^*)$. Первый член $\delta(\mathbf{r}_0)\mathbf{r}_0$ измеряет ошибку линейной аппроксимации математического ожидания градиента логарифма отношения правдоподобия $\nabla \mathbb{E}L(\mathbf{v}, \mathbf{v}^*)$ согласно условию (\mathcal{L}_0) , а второй член $6\nu_0 z_{\mathbb{H}}(\mathbf{x}) \omega \mathbf{r}_0$ контролирует градиент стохастической части и пропорционален энтропии локального множества, которая вовлечена в определение величины $z_{\mathbb{H}}(\mathbf{x})$.

Определим квадратичный процесс $\mathbb{L}(\mathbf{v}, \mathbf{v}^*)$:

$$\mathbb{L}(\mathbf{v}, \mathbf{v}^*) \stackrel{\text{def}}{=} (\mathbf{v} - \mathbf{v}^*)^\top \nabla \mathcal{L}(\mathbf{v}^*) - \|\mathcal{D}(\mathbf{v} - \mathbf{v}^*)\|^2/2.$$

Оценка брэккетинга является следствием аппроксимации градиента (1.11) и описывает качество квадратичной аппроксимации логарифма правдоподобия $\mathcal{L}(\mathbf{v})$ в локальной окрестности точки \mathbf{v}^* .

Следствие 1 (Спокойный, 2012-2013). Пусть условия (ED_2) и (\mathcal{L}_0) из п. 1.2.1 выполняются для некоторого $\mathbf{r}_0 > 0$. Тогда на множестве случайных событий $\Omega_{\mathbf{r}_0}(\mathbf{x})$ доминирующей вероятности не менее $1 - e^{-\mathbf{x}}$ выполняется неравенство

$$|L(\mathbf{v}, \mathbf{v}^*) - \mathbb{L}(\mathbf{v}, \mathbf{v}^*)| \leq \mathbf{r}_0 \diamond(\mathbf{r}_0, \mathbf{x}), \quad \mathbf{v} \in \mathcal{Y}_0(\mathbf{r}_0), \quad (1.13)$$

где $\diamond(\mathbf{r}_0, \mathbf{x})$ определено в 1.12 и $\mathcal{Y}_0(\mathbf{r}_0)$ определено в (1.7).

Результат (1.13) является улучшенной версией аппроксимационной границы из [38] (теорема 3.1). Результат теоремы 1 может быть расширен на проекцию градиента в целевом направлении:

Следствие 2. Предположим выполнение условий (\mathcal{L}_0) , (ED_2) и (\mathcal{I}) . Тогда для всех $\mathbf{r} \leq \mathbf{r}_0$ и всех $\mathbf{v} \in \Upsilon_0(\mathbf{r})$ выполняется неравенство

$$\mathbb{P} \left\{ \sup_{\boldsymbol{\theta} \in \Theta_0(\mathbf{r})} \left\| \check{D}^{-1} \{ \check{\nabla}_{\boldsymbol{\theta}} \mathcal{L}(\mathbf{v}) - \check{\nabla}_{\boldsymbol{\theta}} \mathcal{L}(\mathbf{v}^*) \} + \check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^*) \right\| \geq \diamond(\mathbf{r}, \mathbf{x}) \right\} \leq e^{-\mathbf{x}}, \quad (1.14)$$

где величина $\diamond(\mathbf{r}, \mathbf{x})$ определяется уравнением (1.12).

Основываясь на результате следствия 2, можно построить аппроксимацию отношения правдоподобия в целевом направлении соответствующей квадратичной формой. Во-первых, определим

$$L_{\circ}(\boldsymbol{\theta}, \check{\boldsymbol{\eta}}) \stackrel{\text{def}}{=} \mathcal{L}(\boldsymbol{\theta}, \check{\boldsymbol{\eta}} - H^{-2} A^{\top} \boldsymbol{\theta}),$$

где введено обозначение $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\eta}) \stackrel{\text{def}}{=} \mathcal{L}(\mathbf{v})$ для $\mathbf{v} = (\boldsymbol{\theta}, \boldsymbol{\eta})$. Далее рассмотрим следующее представление отношения правдоподобия

$$\begin{aligned} L(\mathbf{v}, \mathbf{v}^*) &= L_{\circ}(\boldsymbol{\theta}, \check{\boldsymbol{\eta}}) - L_{\circ}(\boldsymbol{\theta}^*, \check{\boldsymbol{\eta}}^*) \\ &= [L_{\circ}(\boldsymbol{\theta}, \check{\boldsymbol{\eta}}) - L_{\circ}(\boldsymbol{\theta}^*, \check{\boldsymbol{\eta}})] + [L_{\circ}(\boldsymbol{\theta}^*, \check{\boldsymbol{\eta}}) - L_{\circ}(\boldsymbol{\theta}^*, \check{\boldsymbol{\eta}}^*)]. \end{aligned}$$

Также определим

$$\check{\mathbb{L}}(\boldsymbol{\theta}, \boldsymbol{\theta}^*) \stackrel{\text{def}}{=} \check{\boldsymbol{\xi}}^{\top} \check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^*) - \|\check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\|^2/2,$$

где $\check{\boldsymbol{\xi}} = \check{D}^{-1} \check{\nabla}_{\boldsymbol{\theta}} L(\mathbf{v}^*)$.

Основываясь на следствии 2, можно доказать следующий результат.

Следствие 3. Пусть выполнены условия следствия 2. Тогда на множестве $\Omega_{\mathbf{r}_0}(\mathbf{x})$ вероятности не менее $1 - e^{-\mathbf{x}}$ для всех $\mathbf{v} \in \Upsilon_0(\mathbf{r}_0)$ выполняется неравенство

$$\left| L(\mathbf{v}, \mathbf{v}^*) - \check{\mathbb{L}}(\boldsymbol{\theta}, \boldsymbol{\theta}^*) - [L_{\circ}(\boldsymbol{\theta}^*, \check{\boldsymbol{\eta}}) - L_{\circ}(\boldsymbol{\theta}^*, \check{\boldsymbol{\eta}}^*)] \right| \leq \|\check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\| \diamond(\mathbf{r}_0, \mathbf{x}).$$

Данное неравенство позволяет оценить логарифм отношения правдоподобия в направлении целевого параметра соответствующей квадратичной формой. Этот результат является основой для получения точных оценок аппроксимации апостериорного распределения целевого параметра.

Теорема 2 (Спокойный и Жилова, 2013). Пусть выполнено условие (ED_0) . Тогда случайный вектор $\boldsymbol{\xi} = \mathcal{D}^{-1}\nabla\mathcal{L}(\mathbf{v}^*)$ удовлетворяет на множестве $\Omega_B(\mathbf{x})$ доминирующей вероятности не менее $1 - 2e^{-x}$ неравенству

$$\|\boldsymbol{\xi}\|^2 \leq z_B^2(\mathbf{x}), \quad (1.15)$$

где $z_B^2(\mathbf{x}) \stackrel{\text{def}}{=} \mathbf{p}_B + 6\lambda_B\mathbf{x}$,

$$B \stackrel{\text{def}}{=} \mathcal{D}^{-1}\mathcal{V}^2\mathcal{D}^{-1}, \quad \mathbf{p}_B \stackrel{\text{def}}{=} \text{tr}(B), \quad \lambda_B \stackrel{\text{def}}{=} \lambda_{\max}(B).$$

Заметим, что результат (1.15) может быть найден в работе [40]. Данная теорема может быть легко расширена на случай проекции на целевое подпространство.

Следствие 4. Пусть выполнены условия (ED_0) и (\mathcal{I}) . Тогда случайный вектор $\check{\boldsymbol{\xi}} = \check{D}^{-1}\check{\nabla}_{\boldsymbol{\theta}}L(\mathbf{v}^*)$ на множестве $\Omega_B(\mathbf{x})$ доминирующей вероятности не менее $1 - 2e^{-x}$ удовлетворяет неравенству

$$\|\check{\boldsymbol{\xi}}\|^2 \leq z_{\check{B}}^2(\mathbf{x}),$$

где

$$z_{\check{B}}^2(\mathbf{x}) \stackrel{\text{def}}{=} \mathbf{p}_{\check{B}} + 6\lambda_{\check{B}}\mathbf{x}, \quad \check{V}^2 \stackrel{\text{def}}{=} V^2 - B^\top Q^{-2}B,$$

$$\check{B} \stackrel{\text{def}}{=} \check{D}^{-1}\check{V}^2\check{D}^{-1}, \quad \mathbf{p}_{\check{B}} \stackrel{\text{def}}{=} \text{tr}(\check{B}), \quad \lambda_{\check{B}} \stackrel{\text{def}}{=} \lambda_{\max}(\check{B}).$$

Следующая теорема позволяет построить оценки больших уклонений при помощи метода верхней функции.

Теорема 3. (Спокойный, 2012) Предположим выполнение условия (ED_2) . Также предположим, что выполнено условие $(\mathcal{L}\mathbf{r})$, т.е. для всех $\mathbf{v} \in \Upsilon \setminus \Upsilon_0(\mathbf{r}_0)$ выполняется неравенство

$$-\mathbb{E}L(\mathbf{v}, \mathbf{v}^*) \geq \mathbf{b} \|\mathcal{D}(\mathbf{v} - \mathbf{v}^*)\|^2$$

Пусть также $\mathbf{b}\mathbf{r}_0 \geq 2(z_B(\mathbf{x}) + \varrho(\mathbf{r}_0, \mathbf{x}))$, где $\varrho(\mathbf{r}, \mathbf{x}) = 6\nu_0 z_{\mathbb{H}}(\mathbf{x} + \log(2\mathbf{r}/\mathbf{r}_0))\omega$.

Тогда неравенство

$$L(\mathbf{v}, \mathbf{v}^*) \leq -\mathbf{b} \|\mathcal{D}(\mathbf{v} - \mathbf{v}^*)\|^2/2, \quad \mathbf{v} \in \Upsilon \setminus \Upsilon_0(\mathbf{r}_0). \quad (1.16)$$

выполняется на множестве $\Omega(\mathbf{x})$ доминирующей вероятности не менее $1 - 4e^{-x}$.

Результат (1.16) близок к утверждению теоремы 4.2 из [38].

1.2.3. Апостериорная вероятность хвостов распределения полного параметра

Следующим шагом в нашем анализе является проверка того, что \mathbf{v} концентрируется в малой окрестности $\mathcal{Y}_0(\mathbf{r}_0)$ центральной точки \mathbf{v}^* при правильном выборе \mathbf{r}_0 . Рассмотрим равномерное априорное распределение $\pi(\mathbf{v}) \equiv 1$, $\mathbf{v} \in \mathcal{Y}$. Концентрационные свойства апостериорного распределения будут описываться с помощью случайной величины

$$P(\mathbf{v} \notin \mathcal{Y}_0(\mathbf{r}_0) \mid \mathbf{Y}) = \frac{\int_{\mathcal{Y} \setminus \mathcal{Y}_0(\mathbf{r}_0)} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v}}{\int_{\mathcal{Y}} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v}}.$$

Теорема 4. Пусть выполнены условия раздела 1.2.1. Тогда на $\Omega_{\mathbf{r}_0}(\mathbf{x})$ справедливо неравенство

$$P(\mathbf{v} \notin \mathcal{Y}_0(\mathbf{r}_0) \mid \mathbf{Y}) \leq \exp\{\mathbf{r}_0 \diamond(\mathbf{r}_0, \mathbf{x}) + \nu_0(\mathbf{r}_0)\} \mathbf{b}^{-p/2} P(\|\boldsymbol{\gamma}\|^2 \geq \mathbf{b}\mathbf{r}_0^2) \quad (1.17)$$

с

$$\nu_0(\mathbf{r}_0) \stackrel{\text{def}}{=} -\log P(\|\boldsymbol{\gamma} + \boldsymbol{\xi}\| \leq \mathbf{r}_0 \mid \mathbf{Y}),$$

где $\boldsymbol{\gamma} \in \mathbb{R}^p$ – стандартный нормальный вектор.

Если $\mathbf{r}_0 \geq z_B(\mathbf{x}) + z(p, \mathbf{x})$, то на $\Omega(\mathbf{x})$ выполняется неравенство

$$\nu_0(\mathbf{r}_0) \leq 2e^{-x}. \quad (1.18)$$

Этот результат дает простое достаточное условие на значение \mathbf{r}_0 , которое обеспечивает концентрацию апостериорного распределения на множестве $\mathcal{Y}_0(\mathbf{r}_0)$.

Следствие 5. Пусть выполнены условия теоремы 4. Тогда дополнительное неравенство $\mathbf{b}\mathbf{r}_0^2 \geq z^2(p, \mathbf{x} + \frac{p}{2} \log \frac{e}{\mathbf{b}})$ обеспечивает справедливость неравенства

$$P(\mathbf{v} \notin \mathcal{Y}_0(\mathbf{r}_0) \mid \mathbf{Y}) \leq \exp\{\mathbf{r}_0 \diamond(\mathbf{r}_0, \mathbf{x}) + 2e^{-x} - \mathbf{x}\} \quad (1.19)$$

на множестве $\Omega(\mathbf{x})$ вероятности не менее $1 - 4e^{-x}$

Этот результат следует из теоремы 4 в силу леммы 2 (см. разд. 1.4.1 ниже). Заметим, что похожие результаты для больших уклонений в частном случае независимых наблюдений можно найти в книге Боровкова и Могульского [41].

1.2.4. Апостериорная вероятность хвостов распределения целевого параметра

Следующим важным шагом нашего анализа является проверка того, что апостериорное распределение целевого параметра $\boldsymbol{\vartheta} \mid \mathbf{Y}$ концентрируется в малой окрестности $\Theta_0(\mathbf{h}_0) = \{\boldsymbol{\theta}: \|\check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\| \leq \mathbf{h}_0\}$ центральной точки $\boldsymbol{\theta}^* = \Pi_0 \mathbf{v}^*$ при правильном выборе \mathbf{h}_0 . Концентрационные свойства апостериорного распределения будут описываться с помощью случайной величины

$$P(\boldsymbol{\vartheta} \notin \Theta_0(\mathbf{h}_0) \mid \mathbf{Y}) = \frac{\int_{\mathcal{Y}} \mathbb{1}\{\boldsymbol{\theta} \notin \Theta_0(\mathbf{h}_0)\} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} \pi(\mathbf{v}) d\mathbf{v}}{\int_{\mathcal{Y}} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} \pi(\mathbf{v}) d\mathbf{v}}.$$

Предположим, что априорное распределение является равномерным, т.е. $\pi(\mathbf{v}) \equiv 1$, $\mathbf{v} \in \mathcal{Y}$. Тогда получается следующее выражение для вероятности больших уклонений:

$$P(\boldsymbol{\vartheta} \notin \Theta_0(\mathbf{h}_0) \mid \mathbf{Y}) = \frac{\int_{\mathcal{Y}} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} \mathbb{1}\{\boldsymbol{\theta} \notin \Theta_0(\mathbf{h}_0)\} d\mathbf{v}}{\int_{\mathcal{Y}} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v}}. \quad (1.20)$$

Заметим, что для $\mathbf{h}_0 \leq \mathbf{r}_0$ выполняется неравенство

$$P(\mathbf{v} \notin \mathcal{Y}_0(\mathbf{r}_0) \mid \mathbf{Y}) \leq P(\mathbf{v} \notin \mathcal{Y}_0(\mathbf{h}_0, \mathbf{r}_0) \mid \mathbf{Y}) \stackrel{\text{def}}{=} \frac{\int_{\mathcal{Y} \setminus \mathcal{Y}_0(\mathbf{h}_0, \mathbf{r}_0)} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v}}{\int_{\mathcal{Y}_0(\mathbf{h}_0, \mathbf{r}_0)} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v}}.$$

Покажем, что можно выбрать \mathbf{h}_0 намного меньшее чем \mathbf{r}_0 таким образом, что полное апостериорное распределение концентрируется на множестве $\mathcal{Y}_0(\mathbf{h}_0, \mathbf{r}_0)$ и соответственно апостериорное распределение целевого параметра концентрируется на множестве $\Theta_0(\mathbf{h}_0)$.

Теорема 5. *Предположим выполнение условий теоремы 4. В дополнение предположим, что $\mathbf{h}_0 \geq z_{\check{B}}(\mathbf{x}) + \sqrt{q} + \sqrt{2\mathbf{x}}$, тогда на множестве $\Omega(\mathbf{x})$ выполняется*

неравенство

$$\mathbb{P}(\mathbf{v} \notin \mathcal{Y}_0(\mathbf{h}_0, \mathbf{r}_0) \mid \mathbf{Y}) \leq \rho_0(\mathbf{h}_0, \mathbf{r}_0, \mathbf{x})e^{-\mathbf{x}},$$

где

$$\rho_0(\mathbf{h}_0, \mathbf{r}_0, \mathbf{x}) \stackrel{\text{def}}{=} e^{\mathbf{r}_0 \diamond (\mathbf{r}_0, \mathbf{x}) + 2e^{-\mathbf{x}}} + 4e^{2\mathbf{h}_0 \diamond (\mathbf{r}_0, \mathbf{x}) + 2e^{-\mathbf{x}}}. \quad (1.21)$$

Более того, на множестве $\Omega(\mathbf{x})$ выполняется неравенство

$$\mathbb{P}(\boldsymbol{\vartheta} \notin \Theta_0(\mathbf{h}_0) \mid \mathbf{Y}) \leq \mathbb{P}(\mathbf{v} \notin \mathcal{Y}_0(\mathbf{h}_0, \mathbf{r}_0) \mid \mathbf{Y}) \leq \rho_0(\mathbf{h}_0, \mathbf{r}_0, \mathbf{x})e^{-\mathbf{x}}. \quad (1.22)$$

Результат (1.19) показывает концентрацию полного параметра на множестве $\mathcal{Y}_0(\mathbf{r}_0)$ с $\mathbf{r}_0^2 \geq \mathbf{C}(p + \mathbf{x})$. Результат (1.22) показывает, что концентрационные свойства могут быть показаны даже для меньшего множества $\mathcal{Y}_0(\mathbf{h}_0, \mathbf{r}_0) = \Theta_0(\mathbf{h}_0) \times \mathcal{H}_0(\mathbf{r}_0)$, что позволяет улучшить оценки для множеств концентрации как для полного параметра $\mathbf{v} = (\boldsymbol{\theta}, \boldsymbol{\eta})$, так и для целевого параметра $\boldsymbol{\theta}$. Также полезным является расширение данного результата на случай локального математического ожидания квадратичной формы от целевого параметра. Для произвольного $\boldsymbol{\lambda} \in \mathbb{R}^q$ с $\|\boldsymbol{\lambda}\| = 1$ определим величину

$$\rho_2(\mathbf{h}_0, \mathbf{r}_0) \stackrel{\text{def}}{=} \frac{\int_{\mathcal{Y} \setminus \mathcal{Y}_0(\mathbf{h}_0, \mathbf{r}_0)} |\boldsymbol{\lambda}^\top \check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)|^2 \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v}}{\int_{\mathcal{Y}} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v}},$$

где $\boldsymbol{\theta}^\circ \stackrel{\text{def}}{=} \boldsymbol{\theta}^* + \check{D}^{-1}\check{\boldsymbol{\xi}}$.

Теорема 6. *Предположим выполнение условий теоремы 4. В дополнение предположим, что $\mathbf{h}_0 \geq z_{\check{B}}(\mathbf{x}) + \sqrt{q} + \sqrt{2\mathbf{x}}$, тогда на множестве $\Omega(\mathbf{x})$ выполняется неравенство*

$$\rho_2(\mathbf{h}_0, \mathbf{r}_0) \leq 3\rho_0(\mathbf{h}_0, \mathbf{r}_0, \mathbf{x})e^{-\mathbf{x}}. \quad (1.23)$$

1.2.5. Локальная гауссовская аппроксимация апостериорного распределения: верхняя граница

В данном разделе мы установим верхние границы для апостериорного математического ожидания. Удобно ввести понятие *условного математического ожидания* \mathbb{E}° : для случайной величины η определим

$$\mathbb{E}^\circ \eta \stackrel{\text{def}}{=} \mathbb{E}[\eta \mathbb{1}\{\mathbf{v} \in \mathcal{Y}_0(\mathbf{h}_0, \mathbf{r}_0)\} \mid \mathbf{Y}].$$

Следующая теорема дает точное утверждение о верхней границе для апостериорного математического ожидания. Введем обозначение

$$\boldsymbol{\theta}^\circ \stackrel{\text{def}}{=} \boldsymbol{\theta}^* + \check{D}^{-1} \check{\boldsymbol{\xi}}, \quad (1.24)$$

где матрица \check{D} определена в (1.2), а вектор $\check{\boldsymbol{\xi}}$ определен в (1.3). Величина $\boldsymbol{\theta}^\circ$ может рассматриваться как аппроксимация оценки максимума правдоподобия до первого члена разложения в ряд Тейлора и используется в работе для удобства и краткости записи результатов. Все результаты могут быть легко обобщены на случай использования вместо $\boldsymbol{\theta}^\circ$ оценки максимума правдоподобия или другой асимптотически эффективной оценки.

Теорема 7. Пусть выполнено неравенство (1.13). Тогда для любой функции $f: \mathbb{R}^q \rightarrow \mathbb{R}_+$ на $\Omega_{\mathbf{r}_0}(\mathbf{x})$ справедлива оценка

$$\mathbb{E}^\circ f(\check{D}(\boldsymbol{\vartheta} - \boldsymbol{\theta}^\circ)) \leq \exp\{2\mathbf{h}_0 \diamond(\mathbf{r}_0, \mathbf{x}) + \nu(\mathbf{h}_0)\} \mathbb{E} f(\boldsymbol{\gamma}), \quad (1.25)$$

где $\boldsymbol{\gamma} \sim \mathcal{N}(0, I_q)$.

В следующем результате рассмотрены частные случаи $f(\mathbf{u}) = |\boldsymbol{\lambda}^\top \mathbf{u}|^2$ и $f(\mathbf{u}) = \mathbb{1}(\mathbf{u} \in A)$ для любого измеримого множества A .

Следствие 6. Для любого $\boldsymbol{\lambda} \in \mathbb{R}^q$ с $\|\boldsymbol{\lambda}\| = 1$ на $\Omega(\mathbf{x})$ справедлива оценка

$$\mathbb{E} \left[\left| \boldsymbol{\lambda}^\top \check{D}(\boldsymbol{\vartheta} - \boldsymbol{\theta}^\circ) \right|^2 \mid \mathbf{Y} \right] \leq \exp\{2\mathbf{h}_0 \diamond(\mathbf{r}_0, \mathbf{x}) + 2e^{-x}\} + 3\rho_0(\mathbf{h}_0, \mathbf{r}_0, \mathbf{x})e^{-x}.$$

Для любого измеримого множества $A \subseteq \mathbb{R}^q$ на $\Omega(\mathbf{x})$ справедлива оценка

$$\begin{aligned} \mathbb{P}(\check{D}(\boldsymbol{\vartheta} - \boldsymbol{\theta}^\circ) \in A \mid \mathbf{Y}) &\leq \exp\{2\mathbf{h}_0 \diamond(\mathbf{r}_0, \mathbf{x}) + 2e^{-x}\} \mathbb{P}(\boldsymbol{\gamma} \in A) \\ &+ \rho_0(\mathbf{h}_0, \mathbf{r}_0, \mathbf{x})e^{-x}, \end{aligned} \quad (1.26)$$

где $\boldsymbol{\gamma}$ – стандартный нормальный вектор в \mathbb{R}^q .

Результат данного следствия легко получается из результата теоремы 7 и оценок больших уклонений (1.22) и (1.23). Приведенное ниже следствие описывает верхнюю оценку апостериорной вероятности в случае, когда происходит изменение центрирования и нормирования.

Следствие 7. Пусть D_1 – симметричная $(q \times q)$ -матрица, удовлетворяющая условию $\|I - D_1^{-1} \check{D}^2 D_1^{-1}\| \leq \alpha$, пусть вектор $\widehat{\boldsymbol{\theta}} \in \mathbb{R}^q$ таков, что $\|\check{D}(\boldsymbol{\theta}^\circ - \widehat{\boldsymbol{\theta}})\| \leq \beta$, и пусть $\boldsymbol{\delta}_0 \stackrel{\text{def}}{=} D_1(\boldsymbol{\theta}^\circ - \widehat{\boldsymbol{\theta}})$. Тогда для любого измеримого множества $A \subset \mathbb{R}^q$ на $\Omega(\mathbf{x})$ выполняется неравенство

$$\begin{aligned} \mathbb{P}(D_1(\boldsymbol{\vartheta} - \widehat{\boldsymbol{\theta}}) \in A \mid \mathbf{Y}) & \\ &\leq \exp\{2\mathbf{h}_0 \diamond(\mathbf{r}_0, \mathbf{x}) + 2e^{-x}\} \mathbb{P}(D_1 \check{D}^{-1} \boldsymbol{\gamma} + \boldsymbol{\delta}_0 \in A) + \rho_0(\mathbf{h}_0, \mathbf{r}_0, \mathbf{x})e^{-x} \\ &\leq \exp\{2\mathbf{h}_0 \diamond(\mathbf{r}_0, \mathbf{x}) + 2e^{-x}\} \left\{ \mathbb{P}(\boldsymbol{\gamma} \in A) + \frac{1}{2} \sqrt{\alpha^2 q + (1 + \alpha)^2 \beta^2} \right\} + \rho_0(\mathbf{h}_0, \mathbf{r}_0, \mathbf{x})e^{-x}. \end{aligned} \quad (1.27)$$

1.2.6. Локальная гауссовская аппроксимация апостериорного распределения: нижняя граница

Теперь укажем нижнюю границу для апостериорного математического ожидания.

Теорема 8. Пусть выполнено неравенство (1.13). Тогда для любой функции $f: \mathbb{R}^q \rightarrow \mathbb{R}_+$ на $\Omega_{\mathbf{r}_0}(\mathbf{x})$ справедлива оценка

$$\mathbb{E}^\circ f(\check{D}(\boldsymbol{\vartheta} - \boldsymbol{\theta}^\circ)) \geq \exp\{-2\mathbf{h}_0 \diamond(\mathbf{r}_0, \mathbf{x})\} \mathbb{E} \left\{ f(\boldsymbol{\gamma}) \mathbb{1}(\|\boldsymbol{\gamma} + \check{\boldsymbol{\xi}}\| \leq \mathbf{h}_0) \right\}, \quad (1.28)$$

где $\boldsymbol{\gamma} \sim \mathcal{N}(0, I_q)$.

Данный результат показывает, что апостериорное математическое ожидание имеет нижнюю границу, которая близка к математическому ожиданию функции стандартной нормальной случайной величины с точностью до (малых) мультипликативных и аддитивных величин. В качестве следствия мы установим результат для квадратичной и индикаторной функций $f(\mathbf{u})$.

Следствие 8. Пусть $\mathbf{x} > 3$. Для любого $\boldsymbol{\lambda} \in \mathbb{R}^q$ с $\|\boldsymbol{\lambda}\| = 1$ на $\Omega(\mathbf{x})$ справедлива оценка

$$\mathbb{E} \left[\left| \boldsymbol{\lambda}^\top \check{D}(\boldsymbol{\vartheta} - \boldsymbol{\theta}^\circ) \right|^2 \mid \mathbf{Y} \right] \geq \exp\{-2\mathbf{h}_0 \diamond(\mathbf{r}_0, \mathbf{x})\} - 3e^{-\mathbf{x}}.$$

Для любого измеримого множества $A \subseteq \mathbb{R}^q$ на $\Omega(\mathbf{x})$ справедлива оценка

$$\mathbb{P}(\check{D}(\boldsymbol{\vartheta} - \boldsymbol{\theta}^\circ) \in A \mid \mathbf{Y}) \geq \exp\{-2\mathbf{h}_0 \diamond(\mathbf{r}_0, \mathbf{x})\} \mathbb{P}(\boldsymbol{\gamma} \in A) - e^{-\mathbf{x}}. \quad (1.29)$$

Пусть D_1^2 – симметричная $(q \times q)$ -матрица такой, что $\|I - D_1^{-1} \check{D}^2 D_1^{-1}\| \leq \alpha$, и пусть вектор $\widehat{\boldsymbol{\theta}} \in \mathbb{R}^q$ таков, что $\|\check{D}(\boldsymbol{\theta}^\circ - \widehat{\boldsymbol{\theta}})\| \leq \beta$. Определим $\boldsymbol{\delta}_0 \stackrel{\text{def}}{=} D_1(\boldsymbol{\theta}^\circ - \widehat{\boldsymbol{\theta}})$. Тогда для любого измеримого множества A в \mathbb{R}^q на $\Omega(\mathbf{x})$ выполняется неравенство

$$\begin{aligned} \mathbb{P}(D_1(\boldsymbol{\vartheta} - \widehat{\boldsymbol{\theta}}) \in A \mid \mathbf{Y}) &\geq \exp\{-2\mathbf{h}_0 \diamond(\mathbf{r}_0, \mathbf{x})\} \mathbb{P}(D_1 \check{D}^{-1} \boldsymbol{\gamma} + \boldsymbol{\delta}_0 \in A) - e^{-\mathbf{x}} \quad (1.30) \\ &\geq \exp\{-2\mathbf{h}_0 \diamond(\mathbf{r}_0, \mathbf{x})\} \left\{ \mathbb{P}(\boldsymbol{\gamma} \in A) - \frac{1}{2} \sqrt{\alpha^2 q + (1 + \alpha)^2 \beta^2} \right\} - e^{-\mathbf{x}}. \end{aligned}$$

Доказательство этого следствия близко к доказательствам следствий 6 и 7.

1.2.7. Основная теорема

Мы начнем с формулировки результата БфМ для апостериорного распределения параметра $\boldsymbol{\vartheta}$, заданного формулой (1.6), в случае равномерного априорного распределения, т.е. $\pi(\mathbf{v}) \equiv 1$ на \mathcal{Y} . Определим апостериорное математическое ожидание целевого параметра $\bar{\boldsymbol{\vartheta}}$ и апостериорную ковариационную матрицу \mathfrak{S}^2 :

$$\bar{\boldsymbol{\vartheta}} \stackrel{\text{def}}{=} \mathbb{E}(\boldsymbol{\vartheta} \mid \mathbf{Y}), \quad \mathfrak{S}^2 \stackrel{\text{def}}{=} \text{Cov}(\boldsymbol{\vartheta} \mid \mathbf{Y}) \stackrel{\text{def}}{=} \mathbb{E}\{(\boldsymbol{\vartheta} - \bar{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta} - \bar{\boldsymbol{\vartheta}})^\top \mid \mathbf{Y}\}. \quad (1.31)$$

Также вспомним определение вектора $\boldsymbol{\theta}^\circ$ из (1.24):

$$\boldsymbol{\theta}^\circ = \boldsymbol{\theta}^* + \check{D}^{-1}\check{\boldsymbol{\xi}}.$$

Ниже мы представляем версию теоремы БфМ в рассматриваемом неасимптотическом подходе, которая утверждает, что вектор $\bar{\boldsymbol{\vartheta}}$ близок к $\boldsymbol{\theta}^\circ$, \mathfrak{S}^2 примерно равна \check{D}^{-2} и распределение вектора $\check{D}(\boldsymbol{\vartheta} - \boldsymbol{\theta}^\circ)$ при условии наблюдения \mathbf{Y} близко к стандартному нормальному распределению. Напомним, что \mathbf{C} является общим обозначением для абсолютных констант и \mathbf{x} является положительным числом, обеспечивающим пренебрежимо малое значение $e^{-\mathbf{x}}$. Обозначим $\Omega(\mathbf{x})$ множество случайных событий доминирующей вероятности с $\mathbb{P}(\Omega(\mathbf{x})) \geq 1 - \mathbf{C}e^{-\mathbf{x}}$. Точные значения \mathbf{C} будут определены ниже.

Теорема 9. *Пусть выполнены условия разд. 1.2.1. В дополнение предположим, что $\mathbf{h}_0 \geq z_{\check{B}}(\mathbf{x}) + \sqrt{q} + \sqrt{2\mathbf{x}}$. Пусть априорное распределение равномерно на \mathcal{Y} . Тогда существует случайное событие $\Omega(\mathbf{x})$ доминирующей вероятности не меньше $1 - 4e^{-\mathbf{x}}$ такое, что на $\Omega(\mathbf{x})$ выполняются неравенства*

$$\|\check{D}(\bar{\boldsymbol{\vartheta}} - \boldsymbol{\theta}^\circ)\|^2 \leq 4\mathbf{h}_0 \diamond(\mathbf{r}_0, \mathbf{x}) + 6(1 + \rho_0(\mathbf{h}_0, \mathbf{r}_0, \mathbf{x}))e^{-\mathbf{x}},$$

$$\|I_q - \check{D}\mathfrak{S}^2\check{D}\| \leq 4\mathbf{h}_0 \diamond(\mathbf{r}_0, \mathbf{x}) + 6(1 + \rho_0(\mathbf{h}_0, \mathbf{r}_0, \mathbf{x}))e^{-\mathbf{x}},$$

где $\bar{\boldsymbol{\vartheta}}$ и \mathfrak{S}^2 определены в (1.31), величина $\diamond(\mathbf{r}_0, \mathbf{x})$ определена в (1.12), а величина $\rho_0(\mathbf{h}_0, \mathbf{r}_0, \mathbf{x})$ определена в (1.21).

Кроме того, на $\Omega(\mathbf{x})$ для любого измеримого множества $A \subset \mathbb{R}^q$ справедливы неравенства

$$\exp(-2\mathbf{h}_0 \diamond(\mathbf{r}_0, \mathbf{x})) \mathbb{P}(\boldsymbol{\gamma} \in A) - e^{-\mathbf{x}}$$

$$\leq \mathbb{P}(\check{D}(\boldsymbol{\vartheta} - \boldsymbol{\theta}^\circ) \in A \mid \mathbf{Y})$$

$$\leq \exp(2\mathbf{h}_0 \diamond(\mathbf{r}_0, \mathbf{x}) + 2e^{-\mathbf{x}}) \mathbb{P}(\boldsymbol{\gamma} \in A) + \rho_0(\mathbf{h}_0, \mathbf{r}_0, \mathbf{x})e^{-\mathbf{x}},$$

где вектор $\boldsymbol{\gamma} \in \mathbb{R}^q$ имеет стандартное нормальное распределение.

Условие “ $\mathbf{h}_0 \diamond (\mathbf{r}_0, \mathbf{x})$ мало” влечет за собой результат теоремы БфМ, т.е. близость центрированной и нормированной апостериорной меры к стандартной нормальной в смысле полной вариации. Классические асимптотические результаты могут быть легко выведены как следствия для многих классических моделей (см. обсуждение в гл. 2). Полученный результат можно расширить следующим образом.

Следствие 9. *В случае выполнения условий теоремы 9 для любого измеримого множества $A \subset \mathbb{R}^q$ на множестве случайных событий $\Omega(\mathbf{x})$ доминирующей вероятности не менее $1 - 4e^{-x}$ справедливы неравенства*

$$\begin{aligned} & \exp(-2\mathbf{h}_0 \diamond (\mathbf{r}_0, \mathbf{x})) \{IP(\boldsymbol{\gamma} \in A) - \tau\} - e^{-x} \\ & \leq IP(\mathfrak{S}^{-1}(\boldsymbol{\vartheta} - \bar{\boldsymbol{\vartheta}}) \in A \mid \mathbf{Y}) \\ & \leq \exp(2\mathbf{h}_0 \diamond (\mathbf{r}_0, \mathbf{x}) + 2e^{-x}) \{IP(\boldsymbol{\gamma} \in A) + \tau\} + \rho_0(\mathbf{h}_0, \mathbf{r}_0, \mathbf{x})e^{-x}, \end{aligned}$$

где $\tau = \frac{1}{2}\Delta(\mathbf{h}_0, \mathbf{r}_0, \mathbf{x})\sqrt{q + (1 + \Delta(\mathbf{h}_0, \mathbf{r}_0, \mathbf{x}))^2}$, $\Delta(\mathbf{h}_0, \mathbf{r}_0, \mathbf{x}) \stackrel{\text{def}}{=} 4\mathbf{h}_0 \diamond (\mathbf{r}_0, \mathbf{x}) + 6(1 + \rho_0(\mathbf{h}_0, \mathbf{r}_0, \mathbf{x}))e^{-x}$ и вектор $\boldsymbol{\gamma} \in \mathbb{R}^q$ имеет стандартное нормальное распределение.

Это следствие является очень важным, так как в приложениях матрица \check{D} и вектор $\boldsymbol{\theta}^\circ$ неизвестны, в то время как матрица \mathfrak{S}^{-1} и вектор $\bar{\boldsymbol{\vartheta}}$ могут быть оценены численно. В случае, когда размерность q фиксирована, условие применимости результата остается неизменным: “ $\mathbf{h}_0 \diamond (\mathbf{r}_0, \mathbf{x})$ мало”. Более того, утверждение может быть расширено на случаи, когда размерность q растет, но $\mathbf{h}_0 \diamond (\mathbf{r}_0, \mathbf{x}) q^{1/2}$ остается малым. Результаты для неинформативного априорного распределения могут быть расширены на случай априорного распределения общего вида $\Pi(d\boldsymbol{\nu})$, плотность которого $\pi(\boldsymbol{\nu})$ является гладкой функцией (см. разд. 1.3).

1.3. Расширение результата теоремы БфМ на случай гладкого априорного распределения

Результаты теоремы 9 для неинформативного априорного распределения могут быть расширены на случай априорного распределения общего вида $\Pi(d\mathbf{v})$ с плотностью $\pi(\mathbf{v})$, которая является гладкой функцией на $\mathcal{Y}_0(\mathbf{r}_0)$. Более точно, предположим, что $\pi(\mathbf{v})$ удовлетворяет условиям

$$\sup_{\mathbf{v} \in \mathcal{Y}_0(\mathbf{r}_0)} |\log \pi(\mathbf{v}) - \log \pi(\mathbf{v}^*)| \leq \alpha(\mathbf{r}_0), \quad \sup_{\mathbf{v} \in \mathcal{Y}} \frac{\pi(\mathbf{v})}{\pi(\mathbf{v}^*)} \leq \mathbf{C}(\mathbf{r}_0),$$

где $\alpha(\mathbf{r}_0)$ – это маленькая константа, а $\mathbf{C}(\mathbf{r}_0)$ – любая фиксированная константа. Тогда результаты теоремы 9 остаются верными с очевидными поправками к ошибке аппроксимации.

Действительно, для любого локального множества $A \subseteq \mathcal{Y}_0(\mathbf{r}_0)$ можно воспользоваться неравенствами:

$$\begin{aligned} \int_A \exp\{\mathcal{L}(\mathbf{v})\} \pi(\mathbf{v}) d\mathbf{v} &\leq e^{\alpha(\mathbf{r}_0)} \pi(\mathbf{v}^*) \int_A \exp\{\mathcal{L}(\mathbf{v})\} d\mathbf{v}, \\ \int_A \exp\{\mathcal{L}(\mathbf{v})\} \pi(\mathbf{v}) d\mathbf{v} &\geq e^{-\alpha(\mathbf{r}_0)} \pi(\mathbf{v}^*) \int_A \exp\{\mathcal{L}(\mathbf{v})\} d\mathbf{v}, \end{aligned}$$

Из этого, в частности, следует для любого $A \subset \mathcal{Y}_0(\mathbf{r}_0)$, что

$$P_\pi(A \mid \mathbf{Y}) \leq \exp\{2\alpha(\mathbf{r}_0)\} P(A \mid \mathbf{Y}).$$

Вероятность дополнения $\mathcal{Y}_0^c(\mathbf{r}_0) \stackrel{\text{def}}{=} \mathcal{Y} \setminus \mathcal{Y}_0(\mathbf{r}_0)$ множества $\mathcal{Y}_0(\mathbf{r}_0)$ увеличивается в $\mathbf{C}(\mathbf{r}_0)$ раз по сравнению со случаем равномерного априорного распределения:

$$\int_{\mathcal{Y}_0^c(\mathbf{r}_0)} \exp\{\mathcal{L}(\mathbf{v})\} \pi(\mathbf{v}) d\mathbf{v} \leq \mathbf{C}(\mathbf{r}_0) \pi(\mathbf{v}^*) \int_{\mathcal{Y}_0^c(\mathbf{r}_0)} \exp\{\mathcal{L}(\mathbf{v})\} d\mathbf{v};$$

таким образом

$$P_\pi(\mathcal{Y}_0^c(\mathbf{r}_0) \mid \mathbf{Y}) \leq \mathbf{C}(\mathbf{r}_0) P(\mathcal{Y}_0^c(\mathbf{r}_0) \mid \mathbf{Y}).$$

В частности, если хвост апостериорного распределения в неинформативном случае удовлетворяет условию $P(\Upsilon_0^c(\mathbf{r}_0) | \mathbf{Y}) \leq e^{-x}$, то $P_\pi(\Upsilon_0^c(\mathbf{r}_0) | \mathbf{Y}) \leq \mathbf{C}(\mathbf{r}_0)e^{-x}$.

Заметим, что приведенные выше оценки являются достаточно грубыми, т.к. в них априорное распределение аппроксимируется константой на локальном множестве. Для случая априорного гауссовского распределения $\Pi(d\mathbf{v})$ можно предложить более точный метод учета его влияния на результат теоремы БфМ. Случай гауссовского распределения является прототипической ситуацией, так как любое гладкое распределение может быть локально аппроксимировано гауссовским. Без ограничения общности будем считать, что среднее значение априорного распределения равно нулю:

$$\Pi = \mathcal{N}(0, G^{-2})$$

с плотностью

$$\pi(\mathbf{v}) \propto \exp\{-\|G\mathbf{v}\|^2/2\}$$

для некоторой симметричной матрицы G^2 .

Неинформативное априорное распределение может рассматриваться как предельный случай гауссовского априорного распределения при $G \rightarrow 0$. Придадим данному факту точный количественный смысл. Насколько малой должна быть матрица G , чтобы результат теоремы БфМ оставался верен? Чтобы объяснить результат первым делом рассмотрим гауссовское распределение данных $\mathcal{N}(\mathbf{v}^*, \mathcal{D}^{-2})$ с центром в истинной точке \mathbf{v}^* и ковариационной матрицей \mathcal{D}^{-2} . Хорошо известно, что в данной ситуации неинформативное априорное распределение приводит к тому, что апостериорное распределение также является гауссовским $\mathcal{N}(\mathbf{v}^*, \mathcal{D}^{-2})$, в то время как гауссовское априорное распределение $\Pi = \mathcal{N}(0, G^{-2})$ приводит также к гауссовскому апостериорному распределению, но с ковариационной матрицей \mathcal{D}_G^{-2} , где $\mathcal{D}_G^2 = \mathcal{D}^2 + G^2$, и со средним $\mathbf{v}_G^* = \mathcal{D}_G^{-2}\mathcal{D}^2\mathbf{v}^*$. Таким образом, априорное распределение не влияет на апостериорное распределение, если две гауссовские меры $\mathcal{N}(\mathbf{v}^*, \mathcal{D}^{-2})$ и $\mathcal{N}(\mathbf{v}_G^*, \mathcal{D}_G^{-2})$ примерно эквивалент-

ны. Соответствующее условие представлено в лемме 3. Оно требует, чтобы величины $\|\mathcal{D}^{-1}\mathcal{D}_G^2\mathcal{D}^{-1} - I_q\| = \|\mathcal{D}^{-1}G^2\mathcal{D}^{-1}\|$, $\text{tr}(\mathcal{D}^{-1}\mathcal{D}_G^2\mathcal{D}^{-1} - I_q)^2 = \text{tr}(\mathcal{D}^{-1}G^2\mathcal{D}^{-1})^2$ и $\|\mathcal{D}_G(\mathbf{v}^* - \mathbf{v}_G^*)\| \asymp \|\mathcal{D}_G^{-1}G^2\mathbf{v}^*\|$ были малы.

Теорема 10. *Предположим, что условия теоремы 9 выполнены. Пусть также $\Pi = \mathcal{N}(0, G^{-2})$ является гауссовской априорной мерой на \mathbb{R}^p такой, что*

$$\|\mathcal{D}^{-1}G^2\mathcal{D}^{-1}\| \leq \epsilon \leq 1/2, \quad \text{tr}(\mathcal{D}^{-1}G^2\mathcal{D}^{-1})^2 \leq \delta^2, \quad \|\mathcal{D}_G^{-1}G^2\mathbf{v}^*\| \leq \beta,$$

где δ и β – заданные константы. Тогда на множестве $\Omega(\mathbf{x})$ вероятности $1 - 4e^{-x}$ выполняются неравенства

$$\mathbb{P}(\check{\mathcal{D}}(\boldsymbol{\vartheta} - \boldsymbol{\theta}^\circ) \in A \mid \mathbf{Y}) \geq \exp(-2\mathbf{h}_0 \diamond(\mathbf{r}_0, \mathbf{x})) \{ \mathbb{P}(\boldsymbol{\gamma} \in A) - \tau \} - e^{-x},$$

$$\mathbb{P}(\check{\mathcal{D}}(\boldsymbol{\vartheta} - \boldsymbol{\theta}^\circ) \in A \mid \mathbf{Y}) \leq \exp(2\mathbf{h}_0 \diamond(\mathbf{r}_0, \mathbf{x}) + 2e^{-x}) \{ \mathbb{P}(\boldsymbol{\gamma} \in A) + \tau \} + \rho_0(\mathbf{h}_0, \mathbf{r}_0, \mathbf{x})e^{-x},$$

где

$$\tau \stackrel{\text{def}}{=} \frac{1}{2} \sqrt{(1 + \epsilon)(3\beta + \epsilon z_B(\mathbf{x}))^2 + \delta^2}$$

и квантильная функция $z_B(\mathbf{x})$ определяется уравнением (1.15).

Похожие условия и результаты могут быть найдены в литературе для конкретных статистических моделей. В частности, в работах [42, 43] исследуется гауссовский случай; см. разд. 4.1 ниже для более тщательного сравнения.

Заметим, что ситуация, в которой матрица G^2 не является малой, также представляет большой теоретический интерес и является объектом дальнейших исследований. Методы, разработанные в данной работе, могут быть применены к этому случаю при помощи рассмотрения пенализированного правдоподобия.

1.4. Доказательства

В этом разделе собраны доказательства результатов главы 1.

1.4.1. Некоторые неравенства для нормального распределения

Этот пункт содержит некоторые простые, но полезные факты о свойствах многомерного стандартного нормального распределения. Многие похожие результаты могут быть найдены в литературе, но мы приводим их с доказательствами, чтобы сохранить самостоятельность нашего изложения. Везде в данном разделе $\boldsymbol{\gamma}$ обозначает стандартный нормальный вектор в пространстве \mathbb{R}^p .

Лемма 1. Для любого $\mathbf{u} \in \mathbb{R}^p$, любого единичного вектора $\mathbf{a} \in \mathbb{R}^p$ и любого $z > 0$ справедливы неравенства

$$\mathbb{P}(\|\boldsymbol{\gamma} - \mathbf{u}\| \geq z) \leq \exp\{-z^2/4 + p/2 + \|\mathbf{u}\|^2/2\}, \quad (1.32)$$

$$\mathbb{E}\{|\boldsymbol{\gamma}^\top \mathbf{a}|^2 \mathbb{1}(\|\boldsymbol{\gamma} - \mathbf{u}\| \geq z)\} \leq (2 + |\mathbf{u}^\top \mathbf{a}|^2) \exp\{-z^2/4 + p/2 + \|\mathbf{u}\|^2/2\}. \quad (1.33)$$

Доказательство. Согласно экспоненциальному неравенству Чебышева, для любого $\lambda < 1$

$$\begin{aligned} \mathbb{P}(\|\boldsymbol{\gamma} - \mathbf{u}\| \geq z) &\leq \exp(-\lambda z^2/2) \mathbb{E} \exp(\lambda \|\boldsymbol{\gamma} - \mathbf{u}\|^2/2) \\ &= \exp\left\{-\frac{\lambda z^2}{2} - \frac{p}{2} \log(1 - \lambda) + \frac{\lambda}{2(1 - \lambda)} \|\mathbf{u}\|^2\right\}. \end{aligned}$$

В частности, при $\lambda = 1/2$ получаем (1.32). Далее, для $\|\mathbf{a}\| = 1$

$$\begin{aligned} \mathbb{E}\{|\boldsymbol{\gamma}^\top \mathbf{a}|^2 \mathbb{1}(\|\boldsymbol{\gamma} - \mathbf{u}\| \geq z)\} &\leq \exp(-z^2/4) \mathbb{E}\{|\boldsymbol{\gamma}^\top \mathbf{a}|^2 \exp(\|\boldsymbol{\gamma} - \mathbf{u}\|^2/4)\} \\ &\leq (2 + |\mathbf{u}^\top \mathbf{a}|^2) \exp(-z^2/4 + p/2 + \|\mathbf{u}\|^2/2) \end{aligned}$$

и получаем (1.33). □

Следующая лемма описывает концентрационные свойства нормы $\|\boldsymbol{\gamma}\|$ нормального случайного вектора. Мы используем версию из [44].

Лемма 2. Для любого $\mathbf{x} > 0$ выполняется неравенство

$$\mathbb{P}(\|\boldsymbol{\gamma}\| \geq z(p, \mathbf{x})) \leq \exp(-\mathbf{x}), \quad (1.34)$$

где

$$z(p, \mathbf{x}) \stackrel{\text{def}}{=} \sqrt{p} + \sqrt{2\mathbf{x}}. \quad (1.35)$$

Доказательство. Доказательство использует следующий хорошо известный факт, который может быть получен непосредственными вычислениями: для $\mu < 1$

$$\log \mathbb{E} \exp(\mu(\|\gamma\|^2 - p)/2) = -\mu p/2 - (p/2) \log(1 - \mu).$$

С учетом $\mu < 1$ получим

$$-\mu - \log(1 - \mu) = \sum_{k=2}^{\infty} \frac{\mu^k}{k} \leq \frac{(\mu)^2}{2} \sum_{k=0}^{\infty} \mu^k \leq \frac{\mu^2}{2(1 - \mu)},$$

Далее для любого $\mathbf{x} > 0$ выполняется

$$\inf_{\mu > 0} \left\{ -\mu(p^{1/2}\mathbf{x}^{1/2} + \mathbf{x}) + \frac{\mu^2 p}{4(1 - \mu)} \right\} \leq -\mathbf{x}.$$

Действительно возьмем

$$\mu = 1 - \frac{1}{2\mathbf{x}^{1/2}/p^{1/2} + 1} = \frac{\mathbf{x}^{1/2}}{\mathbf{x}^{1/2} + p^{1/2}/2},$$

таким образом, что $\mu/(1 - \mu) = 2\mathbf{x}^{1/2}/p^{1/2}$. Тогда

$$\begin{aligned} & -\mu(p^{1/2}\mathbf{x}^{1/2} + \mathbf{x}) + \frac{\mu^2 p}{4(1 - \mu)} \\ &= -\mu(p^{1/2}\mathbf{x}^{1/2} + \mathbf{x} + p/4) + \frac{\mu p}{4(1 - \mu)} \\ &= -\frac{\mathbf{x}^{1/2}}{\mathbf{x}^{1/2} + p^{1/2}/2} (\mathbf{x}^{1/2} + p^{1/2}/2)^2 + \frac{2\mathbf{x}^{1/2}p^{1/2}}{4} = -\mathbf{x}, \end{aligned}$$

что и требовалось. Теперь применим неравенство Маркова

$$\begin{aligned} \log \mathbb{P}(\|\gamma\|^2 > p + 2p^{1/2}\mathbf{x}^{1/2} + 2\mathbf{x}) &= \log \mathbb{P}((\|\gamma\|^2 - p)/2 > p^{1/2}\mathbf{x}^{1/2} + \mathbf{x}) \\ &\leq \inf_{\mu > 0} \left\{ -\mu(p^{1/2}\mathbf{x}^{1/2} + \mathbf{x}) + \log \mathbb{E} \exp\{\mu(\|\gamma\|^2 - p)/2\} \right\} \\ &\leq \inf_{\mu > 0} \left\{ -\mu(p^{1/2}\mathbf{x}^{1/2} + \mathbf{x}) + \frac{\mu^2 p}{4(1 - \mu)} \right\} \leq -\mathbf{x} \end{aligned}$$

и утверждение (1.34) следует. □

Следующая лемма дает границы на расстояние Кульбака-Лейблера $\mathcal{K}(P, P^\circ)$ между двумя нормальными распределениями P и P° .

Лемма 3. Пусть $P = \mathcal{N}(\mathbf{b}, \Sigma)$ и $P^\circ = \mathcal{N}(\mathbf{b}^\circ, \Sigma^\circ)$ для некоторых невырожденных матриц Σ и Σ° . Если

$$\|\Sigma^{-1/2}\Sigma^\circ\Sigma^{-1/2} - I_p\| \leq \epsilon \leq 1/2, \quad \text{tr}(\Sigma^{-1/2}\Sigma^\circ\Sigma^{-1/2} - I_p)^2 \leq \delta^2,$$

то

$$\begin{aligned} \mathcal{K}(P, P^\circ) &= -\mathbb{E}_0 \log \frac{dP^\circ}{dP} \leq \frac{\delta^2}{2} + \frac{1}{2}(\mathbf{b} - \mathbf{b}^\circ)^\top \Sigma^\circ (\mathbf{b} - \mathbf{b}^\circ) \\ &\leq \frac{\delta^2}{2} + \frac{1 + \epsilon}{2}(\mathbf{b} - \mathbf{b}^\circ)^\top \Sigma (\mathbf{b} - \mathbf{b}^\circ). \end{aligned}$$

Для любого измеримого множества $A \subset \mathbb{R}^p$ верна цепочка неравенств

$$|P(A) - P^\circ(A)| \leq \sqrt{\mathcal{K}(P, P^\circ)/2} \leq \frac{1}{2} \sqrt{\delta^2 + (1 + \epsilon)(\mathbf{b} - \mathbf{b}^\circ)^\top \Sigma (\mathbf{b} - \mathbf{b}^\circ)}.$$

Доказательство. Замена переменных $\mathbf{u} = \Sigma^{-1/2}(\mathbf{x} - \mathbf{b})$ позволяет свести общий случай к ситуации, когда P – стандартное нормальное распределение в \mathbb{R}^p , в то время как $P_1 = \mathcal{N}(\boldsymbol{\beta}, B)$ с $\boldsymbol{\beta} = \Sigma^{1/2}(\mathbf{b}^\circ - \mathbf{b})$ и $B \stackrel{\text{def}}{=} \Sigma^{-1/2}\Sigma^\circ\Sigma^{-1/2}$. По определению

$$2 \log \frac{dP^\circ}{dP}(\boldsymbol{\gamma}) = \log \det(B) - (\boldsymbol{\gamma} - \boldsymbol{\beta})^\top B(\boldsymbol{\gamma} - \boldsymbol{\beta}) + \|\boldsymbol{\gamma}\|^2,$$

где $\boldsymbol{\gamma}$ – стандартный нормальный вектор и

$$2\mathcal{K}(P, P^\circ) = -2\mathbb{E}_0 \log \frac{dP^\circ}{dP} = -\log \det(B) + \text{tr}(B - I_p) + \boldsymbol{\beta}^\top B\boldsymbol{\beta}.$$

Пусть a_j есть j -ое собственное значение матрицы $B - I_p$. Тогда из условия $\|B - I_p\| \leq 1/2$ следуют неравенства $|a_j| \leq 1/2$ и

$$\begin{aligned} 2\mathcal{K}(P, P^\circ) &= \boldsymbol{\beta}^\top B\boldsymbol{\beta} + \sum_{j=1}^p \{a_j - \log(1 + a_j)\} \leq \boldsymbol{\beta}^\top B\boldsymbol{\beta} + \sum_{j=1}^p a_j^2 \\ &\leq \boldsymbol{\beta}^\top B\boldsymbol{\beta} + \text{tr}(B - I_p)^2 \leq \boldsymbol{\beta}^\top B\boldsymbol{\beta} + \delta^2. \end{aligned}$$

С помощью неравенства Пинскера [45, 46] получим

$$\sup_A |\mathbb{P}(A) - \mathbb{P}^\circ(A)| \leq \sqrt{\frac{1}{2} \mathcal{K}(\mathbb{P}, \mathbb{P}^\circ)} \leq \frac{1}{2} \sqrt{\delta^2 + \boldsymbol{\beta}^\top B \boldsymbol{\beta}},$$

что и требовалось доказать. □

1.4.2. Доказательство следствия 2

Заметим, что верна следующая цепочка равенств

$$\begin{aligned} & \|\mathcal{D}^{-1}[\nabla L(\mathbf{v}) - \nabla L(\mathbf{v}^*) - \mathcal{D}^2(\mathbf{v} - \mathbf{v}^*)]\| \\ &= \|K^{-1}[\nabla L(\mathbf{v}) - \nabla L(\mathbf{v}^*) - \mathcal{D}^2(\mathbf{v} - \mathbf{v}^*)]\| \\ &= \|K^{-1}[\nabla L(\mathbf{v}) - \nabla L(\mathbf{v}^*)] - K^\top(\mathbf{v} - \mathbf{v}^*)\|, \end{aligned}$$

где матрица K определяется как

$$K \stackrel{\text{def}}{=} \begin{pmatrix} \check{D} & AH^{-1} \\ O & H \end{pmatrix}$$

и соответственно

$$K^{-1} = \begin{pmatrix} \check{D}^{-1} & -\check{D}^{-1}AH^{-2} \\ O & H^{-1} \end{pmatrix}.$$

Проектируя на направление параметра $\boldsymbol{\theta}$, получим

$$\begin{aligned} & \|\check{D}^{-1}[\check{\nabla}_{\boldsymbol{\theta}} L(\mathbf{v}) - \check{\nabla}_{\boldsymbol{\theta}} L(\mathbf{v}^*) - \check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)]\| \\ & \leq \|K^{-1}[\nabla L(\mathbf{v}) - \nabla L(\mathbf{v}^*)] - K^\top(\mathbf{v} - \mathbf{v}^*)\| \\ & = \|\mathcal{D}^{-1}[\nabla L(\mathbf{v}) - \nabla L(\mathbf{v}^*) - \mathcal{D}^2(\mathbf{v} - \mathbf{v}^*)]\|. \end{aligned}$$

В итоге получаем неравенство (1.14) как прямое следствие неравенства (1.11).

1.4.3. Доказательство следствия 3

Рассмотрим величину

$$\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^*, \check{\boldsymbol{\eta}}) \stackrel{\text{def}}{=} L_\circ(\boldsymbol{\theta}, \check{\boldsymbol{\eta}}) - L_\circ(\boldsymbol{\theta}^*, \check{\boldsymbol{\eta}}) - L_0(\boldsymbol{\theta}, \boldsymbol{\theta}^*).$$

Подсчитаем градиент в направлении $\boldsymbol{\theta}$:

$$\nabla_{\boldsymbol{\theta}}\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^*, \check{\boldsymbol{\eta}}) = \check{\nabla}_{\boldsymbol{\theta}}L(\boldsymbol{v}) - \check{\nabla}_{\boldsymbol{\theta}}L(\boldsymbol{v}^*) + \check{D}^2(\boldsymbol{\theta} - \boldsymbol{\theta}^*).$$

Получим

$$\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^*, \check{\boldsymbol{\eta}}) = (\boldsymbol{\theta} - \boldsymbol{\theta}^*)^\top \nabla_{\boldsymbol{\theta}}\alpha(\boldsymbol{\theta}', \boldsymbol{\theta}^*, \check{\boldsymbol{\eta}}),$$

где $\boldsymbol{\theta}'$ – некоторая точка из отрезка между точками $\boldsymbol{\theta}^*$ и $\boldsymbol{\theta}$. Мы можем ограничить

$$|\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^*, \check{\boldsymbol{\eta}})| \leq \|\check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\| \cdot \|\check{D}^{-1}[\nabla_{\boldsymbol{\theta}}\alpha(\boldsymbol{\theta}', \boldsymbol{\theta}^*, \check{\boldsymbol{\eta}})]\|$$

и с помощью (1.14) получаем требуемый результат.

1.4.4. Доказательство теоремы 4

Положим $\mathbf{u}(\boldsymbol{v}) = \mathbf{b} \|\mathcal{D}(\boldsymbol{v} - \boldsymbol{v}^*)\|^2/2$. Теперь с помощью замены переменных получим

$$\begin{aligned} & \frac{\mathbf{b}^{p/2} \det(\mathcal{D})}{(2\pi)^{p/2}} \int_{r \setminus \mathcal{r}_0(\mathbf{r}_0)} \exp\{-\mathbf{u}(\boldsymbol{v})\} d\boldsymbol{v} \\ & \leq \frac{\mathbf{b}^{p/2} \det(\mathcal{D})}{(2\pi)^{p/2}} \int_{r \setminus \mathcal{r}_0(\mathbf{r}_0)} \exp\{-\mathbf{b} \|\mathcal{D}(\boldsymbol{v} - \boldsymbol{v}^*)\|^2/2\} d\boldsymbol{v} = IP(\|\boldsymbol{\gamma}\|^2 \geq \mathbf{b}r_0^2). \end{aligned}$$

Для интеграла в числителе (1.20) на $\Omega(\mathbf{x})$ с использованием (1.16) получим

$$\int_{r \setminus \mathcal{r}_0(\mathbf{r}_0)} \exp\{L(\boldsymbol{v}, \boldsymbol{v}^*)\} d\boldsymbol{v} \leq \int_{r \setminus \mathcal{r}_0(\mathbf{r}_0)} \exp\{-\mathbf{u}(\boldsymbol{v})\} d\boldsymbol{v}.$$

Далее используем тот факт, что функция $\mathbb{L}(\boldsymbol{v}, \boldsymbol{v}^*) = \boldsymbol{\xi}^\top \mathcal{D}(\boldsymbol{v} - \boldsymbol{v}^*) - \|\mathcal{D}(\boldsymbol{v} - \boldsymbol{v}^*)\|^2/2$ пропорциональна плотности нормального распределения. Положим

$$m(\boldsymbol{\xi}) \stackrel{\text{def}}{=} -\|\boldsymbol{\xi}\|^2/2 + \log(\det \mathcal{D}) - p \log(\sqrt{2\pi}).$$

Тогда

$$\begin{aligned} m(\boldsymbol{\xi}) + \mathbb{L}(\mathbf{v}, \mathbf{v}^*) \\ = -\|\mathcal{D}(\mathbf{v} - \mathbf{v}^*) - \boldsymbol{\xi}\|^2/2 + \log(\det \mathcal{D}) - p \log(\sqrt{2\pi}) \end{aligned}$$

является (условно при наблюдении \mathbf{Y}) логарифмом плотности нормального распределения со средним $\mathbf{v}_0 = \mathbf{v}^* + \mathcal{D}^{-1}\boldsymbol{\xi}$ и ковариационной матрицей \mathcal{D}^{-2} . Тогда для интеграла в знаменателе выполняется неравенство

$$\begin{aligned} \int_{\mathcal{Y}_0(\mathbf{r}_0)} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v} \\ \geq \exp\{-\mathbf{r}_0 \diamond(\mathbf{r}_0, \mathbf{x}) - m(\boldsymbol{\xi})\} \int_{\mathcal{Y}_0(\mathbf{r}_0)} \exp\{\mathbb{L}(\mathbf{v}, \mathbf{v}^*) + m(\boldsymbol{\xi})\} d\mathbf{v}. \end{aligned} \quad (1.36)$$

Из неравенства (1.36) по определению $\nu(\mathbf{r}_0)$ следует, что

$$\int_{\mathcal{Y}_0(\mathbf{r}_0)} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v} \geq \exp\{-\mathbf{r}_0 \diamond(\mathbf{r}_0, \mathbf{x}) - m(\boldsymbol{\xi}) - \nu(\mathbf{r}_0)\}. \quad (1.37)$$

Из оценки (1.39) для локального интеграла $\int_{\mathcal{Y}_0(\mathbf{r}_0)} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v}$ получаем

$$\mathbb{P}(\mathbf{v} \notin \mathcal{Y}_0(\mathbf{r}_0) \mid \mathbf{Y}) \leq \exp\{\mathbf{r}_0 \diamond(\mathbf{r}_0, \mathbf{x}) + \nu_0(\mathbf{r}_0) + m(\boldsymbol{\xi})\} \int_{\mathcal{Y} \setminus \mathcal{Y}_0(\mathbf{r}_0)} \exp\{-\mathbf{u}(\mathbf{v})\} d\mathbf{v}.$$

Наконец,

$$\exp\{m(\boldsymbol{\xi})\} = \exp\{-\|\boldsymbol{\xi}\|^2/2\} (2\pi)^{-p/2} \det(\mathcal{D}) \leq (2\pi)^{-p/2} \det(\mathcal{D})$$

и мы получаем утверждение (1.17). Оценка (1.18) также очевидна:

$$\begin{aligned} \nu_0(\mathbf{r}_0) &= -\log \mathbb{P}(\|\boldsymbol{\gamma} + \boldsymbol{\xi}\| \leq \mathbf{r}_0 \mid \mathbf{Y}) \leq -\log \mathbb{P}(\|\boldsymbol{\gamma}\| + \|\boldsymbol{\xi}\| \leq \mathbf{r}_0 \mid \mathbf{Y}) \\ &\leq -\log \mathbb{P}(\|\boldsymbol{\gamma}\| \leq z(p, \mathbf{x}) \mid \mathbf{Y}) \leq 2e^{-\mathbf{x}}. \end{aligned}$$

1.4.5. Доказательство теоремы 5

Покажем, что на множестве доминирующей вероятности апостериорная мера множества $\mathcal{Y}_0(\mathbf{r}_0) \setminus \mathcal{Y}_0(\mathbf{h}_0, \mathbf{r}_0)$ экспоненциально мала при условии, что \mathbf{h}_0^2 превысит

уровень $\mathbf{C}(q + \mathbf{x})$.

Лемма 4. *Предположим выполнение неравенства (1.11). Если $\mathbf{h}_0 \geq z_{\check{B}}(\mathbf{x}) + \sqrt{q} + \sqrt{2\mathbf{x}}$ и $\mathbf{x} > 2$, то на множестве $\Omega(\mathbf{x})$ выполняется неравенство*

$$\int_{\mathcal{Y}_0(\mathbf{r}_0)} e^{L(\mathbf{v}, \mathbf{v}^*)} d\mathbf{v} \leq (1 + 4\delta_0 e^{-\mathbf{x}}) \int_{\mathcal{Y}_0(\mathbf{h}_0, \mathbf{r}_0)} e^{L(\mathbf{v}, \mathbf{v}^*)} d\mathbf{v},$$

$$\delta_0 \stackrel{\text{def}}{=} e^{2\mathbf{h}_0 \diamond(\mathbf{r}_0, \mathbf{x}) + 2e^{-\mathbf{x}}}.$$

Доказательство. Пусть $\mathbf{h}_0 < \mathbf{h}_1 < \dots < \mathbf{h}_K = \mathbf{r}_0$ является возрастающей последовательностью. Определим $U_k \stackrel{\text{def}}{=} \Theta_0(\mathbf{h}_k) \times \mathcal{H}_0(\mathbf{r}_0)$ и применим оценку (1.25), чтобы оценить интеграл по множеству $U_k \setminus U_{k-1}$. Определим $T \stackrel{\text{def}}{=} \int_{U_0} e^{L(\mathbf{v}, \mathbf{v}^*)} d\mathbf{v}$ и получим

$$\begin{aligned} \frac{1}{T} \int_{U_K \setminus U_0} e^{L(\mathbf{v}, \mathbf{v}^*)} d\mathbf{v} &= \sum_{k=1}^K \frac{1}{T} \int_{U_k \setminus U_{k-1}} e^{L(\mathbf{v}, \mathbf{v}^*)} d\mathbf{v} \\ &= \sum_{k=1}^K \frac{\int_{U_k} e^{L(\mathbf{v}, \mathbf{v}^*)} \mathbb{1}(\|\check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\| > \mathbf{h}_{k-1}) d\mathbf{v}}{\int_{U_0} e^{L(\mathbf{v}, \mathbf{v}^*)} d\mathbf{v}} \\ &\leq \sum_{k=1}^K e^{2\mathbf{h}_k \diamond(\mathbf{r}_0, \mathbf{x}) + \nu(\mathbf{h}_0)} \mathbb{P}(\|\boldsymbol{\gamma} + \check{\boldsymbol{\xi}}\| > \mathbf{h}_{k-1}). \end{aligned}$$

Выбор $\mathbf{h}_0 = z_{\check{B}}(\mathbf{x}) + \sqrt{q} + \sqrt{2\mathbf{x}}$ обеспечивает согласно (1.35) выполнение неравенства

$$\mathbb{P}(\|\boldsymbol{\gamma} + \check{\boldsymbol{\xi}}\| \leq \mathbf{h}_0) \geq 1 - e^{-\mathbf{x}}.$$

Далее определим $\mathbf{h}_k = z_{\check{B}}(\mathbf{x}) + \sqrt{q} + \sqrt{2\mathbf{x} + k^2}$. Тогда выполняется неравенство

$h_k - h_0 \leq k$. При условии $2\Diamond(\mathbf{r}_0, \mathbf{x}) \leq 1/2$ выполняется цепочка неравенств:

$$\begin{aligned}
\frac{1}{T} \int_{U_K \setminus U_0} e^{L(\mathbf{v}, \mathbf{v}^*)} d\mathbf{v} &\leq \sum_{k=1}^K e^{2h_k \Diamond(\mathbf{r}_0, \mathbf{x}) + \nu(h_0)} \mathbb{P}(\|\gamma + \check{\xi}\| > h_{k-1}) \\
&\leq \sum_{k=1}^K e^{2h_k \Diamond(\mathbf{r}_0, \mathbf{x}) + \nu(h_0)} e^{-x - (k-1)^2/2} \\
&\leq e^{2h_0 \Diamond(\mathbf{r}_0, \mathbf{x}) + \nu(h_0) - x} \sum_{k=1}^{\infty} e^{k/2 - (k-1)^2/2} \\
&\leq 4e^{2h_0 \Diamond(\mathbf{r}_0, \mathbf{x}) + \nu(h_0) - x}.
\end{aligned}$$

□

Далее легко видеть, что выполняется включение

$$\{\boldsymbol{\theta} \notin \Theta_0(h_0), \mathbf{v} \in \mathcal{Y}\} \subset \{\mathcal{Y} \setminus \mathcal{Y}_0(h_0, \mathbf{r}_0)\}.$$

Таким образом, для интеграла в числителе (1.20) с учетом (1.16) имеем

$$\int_{\mathcal{Y}} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} \mathbb{1}\{\boldsymbol{\theta} \notin \Theta_0(h_0)\} d\mathbf{v} \leq \int_{\mathcal{Y} \setminus \mathcal{Y}_0(h_0, \mathbf{r}_0)} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v}.$$

Основываясь на лемме 4, получим:

$$\int_{\mathcal{Y}_0(\mathbf{r}_0) \setminus \mathcal{Y}_0(h_0, \mathbf{r}_0)} e^{L(\mathbf{v}, \mathbf{v}^*)} d\mathbf{v} \leq 4\delta_0 e^{-x} \int_{\mathcal{Y}_0(h_0, \mathbf{r}_0)} e^{L(\mathbf{v}, \mathbf{v}^*)} d\mathbf{v} \leq 4\delta_0 e^{-x} \int_{\mathcal{Y}_0(\mathbf{r}_0)} e^{L(\mathbf{v}, \mathbf{v}^*)} d\mathbf{v}.$$

Наконец,

$$\begin{aligned}
\mathbb{P}(\boldsymbol{\theta} \notin \Theta_0(h_0) \mid \mathbf{Y}) &= \frac{\int_{\mathcal{Y}} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} \mathbb{1}\{\boldsymbol{\theta} \notin \Theta_0(\mathbf{r}_0)\} d\mathbf{v}}{\int_{\mathcal{Y}} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v}} \\
&\leq \frac{\int_{\mathcal{Y} \setminus \mathcal{Y}_0(h_0, \mathbf{r}_0)} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v}}{\int_{\mathcal{Y}} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v}} \\
&\leq \frac{\int_{\mathcal{Y} \setminus \mathcal{Y}_0(\mathbf{r}_0)} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v} + \int_{\mathcal{Y}_0(\mathbf{r}_0) \setminus \mathcal{Y}_0(h_0, \mathbf{r}_0)} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v}}{\int_{\mathcal{Y}} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v}} \\
&\leq e^{2\mathbf{r}_0 \Diamond(\mathbf{r}_0, \mathbf{x}) + 2e^{-x} - x} + 4e^{2h_0 \Diamond(\mathbf{r}_0, \mathbf{x}) + 2e^{-x} - x}.
\end{aligned}$$

где последний переход сделан с использованием теоремы 4.

1.4.6. Доказательство теоремы 6

Мы оценим числитель и знаменатель отдельно. Введем обозначение

$$\boldsymbol{\lambda}_0 = \mathcal{D}^{-1} \begin{pmatrix} \check{D}\boldsymbol{\lambda} \\ \mathbf{0} \end{pmatrix},$$

где $\mathbf{0}$ – нулевой вектор размерности $p - q$. Заметим, что $\|\boldsymbol{\lambda}_0\|^2 = \|\boldsymbol{\lambda}\|^2 = 1$.

Разобьем числитель на два слагаемых:

$$\begin{aligned} \rho_2(\mathbf{h}_0, \mathbf{r}_0) &= \int_{\mathcal{r} \setminus \mathcal{r}_0(\mathbf{h}_0, \mathbf{r}_0)} |\boldsymbol{\lambda}^\top \check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)|^2 \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v} \\ &= \int_{\mathcal{r} \setminus \mathcal{r}_0(\mathbf{r}_0)} |\boldsymbol{\lambda}^\top \check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)|^2 \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v} \\ &\quad + \int_{\mathcal{r}_0(\mathbf{r}_0) \setminus \mathcal{r}_0(\mathbf{h}_0, \mathbf{r}_0)} |\boldsymbol{\lambda}^\top \check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)|^2 \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v} \end{aligned}$$

Для первого слагаемого с помощью (1.16) получаем неравенство

$$\begin{aligned} &\int_{\mathcal{r} \setminus \mathcal{r}_0(\mathbf{r}_0)} |\boldsymbol{\lambda}^\top \check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)|^2 \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v} \\ &\leq \int_{\mathcal{r} \setminus \mathcal{r}_0(\mathbf{r}_0)} |\boldsymbol{\lambda}^\top \check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)|^2 \exp\{-\mathbf{b}\|\mathcal{D}(\mathbf{v} - \mathbf{v}^*)\|^2/2\} d\mathbf{v} \\ &= \int_{\mathcal{r} \setminus \mathcal{r}_0(\mathbf{r}_0)} |\boldsymbol{\lambda}_0^\top \mathcal{D}(\mathbf{v} - \mathbf{v}^*)|^2 \exp\{-\mathbf{b}\|\mathcal{D}(\mathbf{v} - \mathbf{v}^*)\|^2/2\} d\mathbf{v} \\ &= \exp\{-(p/2) \log \mathbf{b} - \log(\det \mathcal{D}) + p \log(\sqrt{2\pi})\} \mathbb{E} |\boldsymbol{\lambda}_0^\top \boldsymbol{\gamma}|^2 \mathbb{1}(\|\boldsymbol{\gamma}\|^2 \geq \mathbf{b}r_0^2). \end{aligned}$$

Далее с помощью (1.33) получим

$$\begin{aligned} &\int_{\mathcal{r} \setminus \mathcal{r}_0(\mathbf{r}_0)} |\boldsymbol{\lambda}^\top \check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)|^2 \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v} \\ &\leq 2 \exp\left\{-\frac{p}{2} \log \mathbf{b} - \log(\det \mathcal{D}) + p \log(\sqrt{2\pi}) - \mathbf{b}r_0^2/4 + p/2\right\} \|\boldsymbol{\lambda}_0\|^2. \end{aligned}$$

Теперь возьмем $\mathbf{br}_0^2 \geq (2p+2)\log(e/b) + 4\mathbf{x}$ на $\Omega(\mathbf{x})$ и получим

$$\begin{aligned} & \int_{\mathcal{r} \setminus \mathcal{r}_0(\mathbf{r}_0)} |\boldsymbol{\lambda}^\top \check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)|^2 \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v} \\ & \leq \exp\{-\log(\det \mathcal{D}) + p \log(\sqrt{2\pi}) - \mathbf{x}\} \|\boldsymbol{\lambda}_0\|^2 \\ & = \exp\{-\log(\det \mathcal{D}) + p \log(\sqrt{2\pi}) - \mathbf{x}\}. \end{aligned}$$

Теперь оценим второе слагаемое числителя. Пусть аналогично лемме 4: $\mathbf{h}_0 < \mathbf{h}_1 < \dots < \mathbf{h}_K = \mathbf{r}_0$ является возрастающей последовательностью. Определим $U_k \stackrel{\text{def}}{=} \Theta_0(\mathbf{h}_k) \times \mathcal{H}_0(\mathbf{r}_0)$ и применим оценку (1.25), чтобы оценить интеграл по множеству $U_k \setminus U_{k-1}$. Определим $T = \int_{U_0} e^{L(\mathbf{v}, \mathbf{v}^*)}$. Получим цепочку неравенств:

$$\begin{aligned} & \frac{1}{T} \int_{U_K \setminus U_0} |\boldsymbol{\lambda}^\top \check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)|^2 e^{L(\mathbf{v}, \mathbf{v}^*)} d\mathbf{v} \\ & = \sum_{k=1}^K \frac{1}{T} \int_{U_k \setminus U_{k-1}} |\boldsymbol{\lambda}^\top \check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)|^2 e^{L(\mathbf{v}, \mathbf{v}^*)} d\mathbf{v} \\ & \leq \sum_{k=1}^K \frac{\int_{U_k} |\boldsymbol{\lambda}^\top \check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)|^2 e^{L(\mathbf{v}, \mathbf{v}^*)} \mathbb{1}(\|\check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\| > \mathbf{h}_{k-1}) d\mathbf{v}}{\int_{U_0} e^{L(\mathbf{v}, \mathbf{v}^*)} d\mathbf{v}} \\ & \leq \sum_{k=1}^K e^{2\mathbf{h}_k \diamond(\mathbf{r}_0, \mathbf{x}) + \nu(\mathbf{h}_0)} \mathbb{E} |\boldsymbol{\lambda}^\top \boldsymbol{\gamma}|^2 \mathbb{1}\{\|\boldsymbol{\gamma} + \check{\boldsymbol{\xi}}\| > \mathbf{h}_{k-1}\} \\ & \leq (2 + |\boldsymbol{\lambda}^\top \check{\boldsymbol{\xi}}|^2) 4e^{2\mathbf{h}_0 \diamond(\mathbf{r}_0, \mathbf{x}) + \nu(\mathbf{h}_0) - \mathbf{x}} \\ & \leq 8(1 + \|\boldsymbol{\lambda}\|^2 e^{-\mathbf{x}}) e^{2\mathbf{h}_0 \diamond(\mathbf{r}_0, \mathbf{x}) + \nu(\mathbf{h}_0) - \mathbf{x}} \\ & \leq 9e^{2\mathbf{h}_0 \diamond(\mathbf{r}_0, \mathbf{x}) + \nu(\mathbf{h}_0) - \mathbf{x}}. \end{aligned}$$

для $\mathbf{h}_0 \geq z_{\check{B}}(\mathbf{x}) + \sqrt{q} + \sqrt{2\mathbf{x}}$ на $\Omega(\mathbf{x})$, $\mathbf{x} > 3$.

Далее для знаменателя получим с помощью (1.37):

$$\int_{\mathcal{r}_0(\mathbf{r}_0)} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v} \geq \exp\{-\mathbf{r}_0 \diamond(\mathbf{r}_0, \mathbf{x}) - m(\boldsymbol{\xi}) - \nu_0(\mathbf{r}_0)\}.$$

Наконец, на $\Omega(\mathbf{x})$ получаем цепочку неравенств

$$\begin{aligned}
\rho_2(\mathbf{h}_0, \mathbf{r}_0) &= \frac{\int_{\mathcal{Y} \setminus \mathcal{Y}_0(\mathbf{h}_0, \mathbf{r}_0)} |\boldsymbol{\lambda}^\top \check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)|^2 \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v}}{\int_{\mathcal{Y}} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v}} \\
&\leq \frac{2 \exp\{-\log(\det \mathcal{D}) + p \log(\sqrt{2\pi})\} - \mathbf{x}}{\exp\{-\mathbf{r}_0 \diamond (\mathbf{r}_0, \mathbf{x}) - m(\check{\boldsymbol{\xi}}) - \nu_0(\mathbf{r}_0)\}} \\
&= 2e^{\mathbf{r}_0 \diamond (\mathbf{r}_0, \mathbf{x}) - \|\check{\boldsymbol{\xi}}\|^2/2 + \nu_0(\mathbf{r}_0) - \mathbf{x}} + 9e^{2\mathbf{h}_0 \diamond (\mathbf{r}_0, \mathbf{x}) + \nu(\mathbf{h}_0) - \mathbf{x}} \\
&\leq 2e^{\mathbf{r}_0 \diamond (\mathbf{r}_0, \mathbf{x}) + \nu_0(\mathbf{r}_0) - \mathbf{x}} + 9e^{2\mathbf{h}_0 \diamond (\mathbf{r}_0, \mathbf{x}) + \nu(\mathbf{h}_0) - \mathbf{x}} \\
&\leq 2e^{\mathbf{r}_0 \diamond (\mathbf{r}_0, \mathbf{x}) + 2e^{-\mathbf{x}} - \mathbf{x}} + 9e^{2\mathbf{h}_0 \diamond (\mathbf{r}_0, \mathbf{x}) + 2e^{-\mathbf{x}} - \mathbf{x}},
\end{aligned}$$

которая влечет за собой результат теоремы 6.

1.4.7. Доказательство теоремы 7

Заметим, что величина $\check{\mathbb{L}}(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = \check{\boldsymbol{\xi}}^\top \check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^*) - \|\check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\|^2/2$ пропорциональна плотности нормального распределения. Более конкретно, определим

$$m(\check{\boldsymbol{\xi}}) \stackrel{\text{def}}{=} -\|\check{\boldsymbol{\xi}}\|^2/2 + \log(\det \check{D}) - q \log(\sqrt{2\pi}).$$

Тогда

$$m(\check{\boldsymbol{\xi}}) + \check{\mathbb{L}}(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = -\|\check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^*) - \check{\boldsymbol{\xi}}\|^2/2 + \log(\det \check{D}) - q \log(\sqrt{2\pi})$$

является (условно при наблюдении \mathbf{Y}) нормальным распределением со средним $\boldsymbol{\theta}^\circ = \boldsymbol{\theta}^* + \check{D}^{-1}\check{\boldsymbol{\xi}}$ и ковариационной матрицей \check{D}^{-2} . Таким образом, для любой неотрицательной функции $f: \mathbb{R}^q \rightarrow \mathbb{R}_+$ получим

$$\begin{aligned}
&\int_{\mathcal{Y}_0(\mathbf{h}_0, \mathbf{r}_0)} \exp\{L(\mathbf{v}, \mathbf{v}^*) + m(\check{\boldsymbol{\xi}})\} f(\check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)) d\mathbf{v} \\
&\leq \int_{\boldsymbol{\theta}_0(\mathbf{h}_0)} \exp\{\|\check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\| \diamond (\mathbf{r}_0, \mathbf{x}) + \check{\mathbb{L}}(\boldsymbol{\theta}, \boldsymbol{\theta}^*) + m(\check{\boldsymbol{\xi}})\} f(\check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)) d\boldsymbol{\theta} \\
&\quad \cdot \int_{\mathcal{H}_0(\mathbf{r}_0)} \exp\{\mathcal{L}_\circ(\check{\boldsymbol{\eta}}, \check{\boldsymbol{\eta}}^*)\} d\check{\boldsymbol{\eta}},
\end{aligned}$$

где $\mathcal{L}_\circ(\check{\eta}, \check{\eta}^*) = L_\circ(\theta^*, \check{\eta}) - L_\circ(\theta^*, \check{\eta}^*)$. Далее получим

$$\begin{aligned} & \int_{\mathcal{r}} \exp\{L(\mathbf{v}, \mathbf{v}^*) + m(\check{\xi})\} f(\check{D}(\theta - \theta^\circ)) d\mathbf{v} \\ & \leq \mathbb{E} e^{\|\gamma + \check{\xi}\| \diamond(\mathbf{r}_0, \mathbf{x})} \mathbb{1}\{\|\gamma + \check{\xi}\| \leq h_0\} f(\gamma) \int_{\mathcal{H}_0(\mathbf{r}_0)} \exp\{\mathcal{L}_\circ(\check{\eta}, \check{\eta}^*)\} d\check{\eta} \\ & \leq e^{h_0 \diamond(\mathbf{r}_0, \mathbf{x})} \mathbb{E} f(\gamma) \int_{\mathcal{H}_0(\mathbf{r}_0)} \exp\{\mathcal{L}_\circ(\check{\eta}, \check{\eta}^*)\} d\check{\eta}, \end{aligned}$$

где γ является стандартным нормальным вектором в \mathbb{R}^q . Тогда

$$\begin{aligned} & \int_{\mathcal{r}} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} f(\check{D}(\theta - \theta^\circ)) d\mathbf{v} \\ & \leq \exp\{h_0 \diamond(\mathbf{r}_0, \mathbf{x}) - m(\check{\xi})\} \mathbb{E} f(\gamma) \int_{\mathcal{H}_0(\mathbf{r}_0)} \exp\{\mathcal{L}_\circ(\check{\eta}, \check{\eta}^*)\} d\check{\eta}. \end{aligned} \quad (1.38)$$

Для интеграла в знаменателе выполняется

$$\begin{aligned} & \int \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v} \\ & \geq e^{-h_0 \diamond(\mathbf{r}_0, \mathbf{x}) - m(\check{\xi})} \int_{\mathcal{r}_0(h_0, \mathbf{r}_0)} \exp\{\check{L}(\theta, \theta^*) + m(\check{\xi})\} \exp\{\mathcal{L}_\circ(\check{\eta}, \check{\eta}^*)\} d\theta d\check{\eta}. \end{aligned}$$

Далее мы можем ограничить снизу

$$\begin{aligned} & \int_{\mathcal{r}_0(\mathbf{r}_0)} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v} \\ & \geq e^{-h_0 \diamond(\mathbf{r}_0, \mathbf{x}) - m(\check{\xi})} \int_{\Theta_0(h_0)} \exp\{\check{L}(\theta, \theta^*) + m(\check{\xi})\} d\theta \int_{\mathcal{H}_0(\mathbf{r}_0)} \exp\{\mathcal{L}_\circ(\check{\eta}, \check{\eta}^*)\} d\check{\eta} \\ & = e^{-h_0 \diamond(\mathbf{r}_0, \mathbf{x}) - m(\check{\xi})} \mathbb{E} \mathbb{1}\{\|\gamma + \check{\xi}\| \leq h_0\} \int_{\mathcal{H}_0(\mathbf{r}_0)} \exp\{\mathcal{L}_\circ(\check{\eta}, \check{\eta}^*)\} d\check{\eta}. \end{aligned} \quad (1.39)$$

Из (1.38) и (1.39) получаем

$$\begin{aligned} & \frac{\int_{\mathcal{r}} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} f(\check{D}(\theta - \theta^\circ)) d\mathbf{v}}{\int_{\mathcal{r}} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v}} \\ & \leq \exp\{2h_0 \diamond(\mathbf{r}_0, \mathbf{x}) + \nu(h_0)\} \mathbb{E} f(\gamma), \end{aligned}$$

что и требовалось доказать.

1.4.8. Доказательство следствия 7

Первое утверждение в (1.30) следует из теоремы 7 с $f(\mathbf{u}) = \mathbb{1}(D_1\check{D}^{-1}\mathbf{u} + \boldsymbol{\delta}_0 \in A)$. Далее, на $\Omega(\mathbf{x})$ для $\boldsymbol{\delta}_0 \stackrel{\text{def}}{=} D_1(\boldsymbol{\theta}^\circ - \widehat{\boldsymbol{\theta}})$ выполняется соотношение

$$\begin{aligned} \|\boldsymbol{\delta}_0\|^2 &= \|D_1(\boldsymbol{\theta}^\circ - \widehat{\boldsymbol{\theta}})\|^2 \leq (1 + \alpha)\|\check{D}(\boldsymbol{\theta}^\circ - \widehat{\boldsymbol{\theta}})\|^2 \\ &\leq (1 + \alpha)\beta^2. \end{aligned}$$

Для доказательства (1.27) мы подсчитаем расстояние Кульбака-Лейблера и применим неравенство Пинскера. Пусть $\boldsymbol{\gamma}$ – стандартный нормальный вектор в \mathbb{R}^q . Случайная величина $D_1\check{D}^{-1}\boldsymbol{\gamma} + \boldsymbol{\delta}_0$ нормальна со средним $\boldsymbol{\delta}_0$ и ковариационной матрицей $B_1^{-1} \stackrel{\text{def}}{=} D_1\check{D}^{-2}D_1$. Очевидно,

$$\|I_q - B_1\| = \|I_q - D_1^{-1}\check{D}^2D_1^{-1}\| \leq \alpha.$$

Таким образом, по лемме 3 для любого измеримого множества A

$$IP(D_1\check{D}^{-1}\boldsymbol{\gamma} + \boldsymbol{\delta}_0 \in A \mid \mathbf{Y}) \leq IP(\boldsymbol{\gamma} \in A) + \frac{1}{2}\sqrt{\alpha^2q + (1 + \alpha)^2\beta^2}.$$

1.4.9. Доказательство теоремы 8

Воспользуемся неравенством (1.13). Как и в доказательстве теоремы 7, для любой неотрицательной функции $f : \mathbb{R}^q \rightarrow \mathbb{R}_+$ верна следующая цепочка неравенств:

$$\begin{aligned} &\int_{\mathcal{Y}} \exp\{L(\mathbf{v}, \mathbf{v}^*) + m(\check{\boldsymbol{\xi}})\} f(\check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)) d\mathbf{v} \\ &\geq \int_{\mathcal{Y}_0(\mathbf{h}_0, \mathbf{r}_0)} \exp\{L(\mathbf{v}, \mathbf{v}^*) + m(\check{\boldsymbol{\xi}})\} f(\check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)) d\mathbf{v} \\ &\geq e^{-\mathbf{h}_0 \diamond (\mathbf{r}_0, \mathbf{x})} \int_{\mathcal{Y}_0(\mathbf{h}_0, \mathbf{r}_0)} \exp\{\check{\mathbb{L}}(\boldsymbol{\theta}, \boldsymbol{\theta}^*) + m(\check{\boldsymbol{\xi}})\} \exp\{\mathcal{L}_\circ(\check{\boldsymbol{\eta}}, \check{\boldsymbol{\eta}}^*)\} f(\check{D}(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)) d\mathbf{v}, \end{aligned}$$

где $\mathcal{L}_\circ(\check{\eta}, \check{\eta}^*) = L_\circ(\theta^*, \check{\eta}) - L_\circ(\theta^*, \check{\eta}^*)$. Далее получим

$$\begin{aligned} & \int_{\mathcal{Y}} \exp\{L(\mathbf{v}, \mathbf{v}^*) + m(\check{\xi})\} f(\check{D}(\theta - \theta^\circ)) d\mathbf{v} \\ & \geq e^{-h_0 \diamond (\mathbf{r}_0, \mathbf{x})} \mathbb{E} \mathbb{1}\{\|\gamma + \check{\xi}\| \leq h_0\} f(\gamma) \int_{\mathcal{H}_0(\mathbf{r}_0)} \exp\{\mathcal{L}_\circ(\check{\eta}, \check{\eta}^*)\} d\check{\eta}, \end{aligned} \quad (1.40)$$

где γ – стандартный нормальный вектор в \mathbb{R}^q . Аналогично для знаменателя

$$\begin{aligned} & \int \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v} \\ & \leq e^{h_0 \diamond (\mathbf{r}_0, \mathbf{x}) - m(\check{\xi})} \int_{\mathcal{Y}_0(h_0, \mathbf{r}_0)} \exp\{\check{L}(\theta, \theta^*) + m(\check{\xi})\} \exp\{\mathcal{L}_\circ(\check{\eta}, \check{\eta}^*)\} d\theta d\check{\eta}. \end{aligned}$$

и, наконец,

$$\begin{aligned} & \int \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v} \\ & \leq e^{h_0 \diamond (\mathbf{r}_0, \mathbf{x}) - m(\check{\xi})} \int_{\Theta_0(h_0)} \exp\{\check{L}(\theta, \theta^*) + m(\check{\xi})\} d\theta \int_{\mathcal{H}_0(\mathbf{r}_0)} \exp\{\mathcal{L}_\circ(\check{\eta}, \check{\eta}^*)\} d\check{\eta} \\ & = e^{h_0 \diamond (\mathbf{r}_0, \mathbf{x}) - m(\check{\xi})} \mathbb{E} \mathbb{1}\{\|\gamma + \check{\xi}\| \leq h_0\} \int_{\mathcal{H}_0(\mathbf{r}_0)} \exp\{\mathcal{L}_\circ(\check{\eta}, \check{\eta}^*)\} d\check{\eta} \\ & \leq e^{h_0 \diamond (\mathbf{r}_0, \mathbf{x}) - m(\check{\xi})} \int_{\mathcal{H}_0(\mathbf{r}_0)} \exp\{\mathcal{L}_\circ(\check{\eta}, \check{\eta}^*)\} d\check{\eta}. \end{aligned} \quad (1.41)$$

Из неравенств (1.40) и (1.41) следует, что

$$\begin{aligned} & \frac{\int_{\mathcal{Y}} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} f(\check{D}(\theta - \theta^\circ)) d\mathbf{v}}{\int_{\mathcal{Y}} \exp\{L(\mathbf{v}, \mathbf{v}^*)\} d\mathbf{v}} \\ & \geq \frac{\exp\{-h_0 \diamond (\mathbf{r}_0, \mathbf{x}) - m(\check{\xi})\} \mathbb{E} f(\gamma) \mathbb{1}\{\|\gamma + \check{\xi}\| \leq h_0\}}{\exp\{h_0 \diamond (\mathbf{r}_0, \mathbf{x}) - m(\check{\xi})\}} \\ & \geq \exp\{-2h_0 \diamond (\mathbf{r}_0, \mathbf{x})\} \mathbb{E} f(\gamma) \mathbb{1}\{\|\gamma + \check{\xi}\| \leq h_0\}. \end{aligned}$$

Таким образом, получаем (1.28).

1.4.10. Доказательство теоремы 9

Достаточно показать, что на $\Omega(\mathbf{x})$ выполняются неравенства

$$\|\check{D}(\bar{\boldsymbol{\vartheta}} - \boldsymbol{\theta}^\circ)\|^2 \leq 2\Delta^*,$$

$$\|I_q - \check{D}\mathfrak{G}^2\check{D}\| \leq 2\Delta^*,$$

где $\Delta^* \stackrel{\text{def}}{=} \max\{\Delta^+, \Delta^-\}$ и введены обозначения $\Delta^+ \stackrel{\text{def}}{=} 2\mathbf{h}_0 \diamond(\mathbf{r}_0, \mathbf{x}) + 2e^{-\mathbf{x}} + 3\rho_0(\mathbf{h}_0, \mathbf{r}_0, \mathbf{x})e^{-\mathbf{x}}$ и $\Delta^- \stackrel{\text{def}}{=} 2\mathbf{h}_0 \diamond(\mathbf{r}_0, \mathbf{x}) + 3e^{-\mathbf{x}}$. Таким образом, $\Delta^* \leq 2\mathbf{h}_0 \diamond(\mathbf{r}_0, \mathbf{x}) + 3(1 + \rho_0(\mathbf{h}_0, \mathbf{r}_0, \mathbf{x}))e^{-\mathbf{x}}$.

Рассмотрим $\boldsymbol{\eta} \stackrel{\text{def}}{=} \check{D}(\boldsymbol{\vartheta} - \boldsymbol{\theta}^\circ)$. Из следствий 6 и 8 вытекает, что для любого $\boldsymbol{\lambda} \in \mathbb{R}^q$ с $\|\boldsymbol{\lambda}\| = 1$ выполняются неравенства

$$\exp(-\Delta^-) \leq \mathbb{E}^\circ |\boldsymbol{\lambda}^\top \boldsymbol{\eta}|^2 \leq \exp(\Delta^+). \quad (1.42)$$

Определим два первых момента величины $\boldsymbol{\eta}$:

$$\bar{\boldsymbol{\eta}} \stackrel{\text{def}}{=} \mathbb{E}^\circ \boldsymbol{\eta}, \quad S_\circ^2 \stackrel{\text{def}}{=} \mathbb{E}^\circ \{(\boldsymbol{\eta} - \bar{\boldsymbol{\eta}})(\boldsymbol{\eta} - \bar{\boldsymbol{\eta}})^\top\} = \check{D}\mathfrak{G}^2\check{D}.$$

Воспользуемся следующим техническим утверждением.

Лемма 5. Пусть выполнена оценка (1.42). Тогда при $\Delta^* = \max\{\Delta^+, \Delta^-\} \leq 1/2$ справедливы неравенства

$$\|\bar{\boldsymbol{\eta}}\|^2 \leq 2\Delta^*, \quad \|S_\circ^2 - I_q\| \leq 2\Delta^*. \quad (1.43)$$

Доказательство. Пусть \mathbf{u} - произвольный единичный вектор в \mathbb{R}^q . Из (1.42) получим

$$\exp(-\Delta^-) \leq \mathbb{E}^\circ |\mathbf{u}^\top \boldsymbol{\eta}|^2 \leq \exp(\Delta^+).$$

Заметим теперь, что

$$\mathbb{E}^\circ |\mathbf{u}^\top \boldsymbol{\eta}|^2 = \mathbf{u}^\top S_\circ^2 \mathbf{u} + |\mathbf{u}^\top \bar{\boldsymbol{\eta}}|^2.$$

Таким образом,

$$\exp(-\Delta^-) \leq \mathbf{u}^\top S_\circ^2 \mathbf{u} + |\mathbf{u}^\top \bar{\boldsymbol{\eta}}|^2 \leq \exp(\Delta^+). \quad (1.44)$$

Аналогично для $\mathbf{u} = \bar{\boldsymbol{\eta}}/\|\bar{\boldsymbol{\eta}}\|$ и $\boldsymbol{\gamma} \sim \mathcal{N}(0, I_q)$

$$\begin{aligned} \mathbb{E}^\circ |\mathbf{u}^\top (\boldsymbol{\eta} - \bar{\boldsymbol{\eta}})|^2 &\geq e^{-\Delta^-} \mathbb{E} |\mathbf{u}^\top (\boldsymbol{\gamma} - \bar{\boldsymbol{\eta}})|^2 \mathbb{1}\{\|\boldsymbol{\gamma} + \check{\boldsymbol{\xi}}\| \leq \mathbf{h}_0\} \\ &= e^{-\Delta^-} (1 + \|\bar{\boldsymbol{\eta}}\|^2) - e^{-\Delta^-} \mathbb{E} |\mathbf{u}^\top (\boldsymbol{\gamma} - \bar{\boldsymbol{\eta}})|^2 \mathbb{1}\{\|\boldsymbol{\gamma} + \check{\boldsymbol{\xi}}\| \geq \mathbf{h}_0\}. \end{aligned}$$

Рассмотрим остаточный член

$$\begin{aligned} e^{-\Delta^-} \mathbb{E} |\mathbf{u}^\top (\boldsymbol{\gamma} - \bar{\boldsymbol{\eta}})|^2 \mathbb{1}\{\|\boldsymbol{\gamma} + \check{\boldsymbol{\xi}}\| \geq \mathbf{h}_0\} \\ = e^{-\Delta^-} \mathbb{E} |\mathbf{u}^\top \boldsymbol{\gamma}|^2 \mathbb{1}\{\|\boldsymbol{\gamma} + \check{\boldsymbol{\xi}}\| \geq \mathbf{h}_0\} + e^{-\Delta^-} \|\bar{\boldsymbol{\eta}}\|^2 \mathbb{E} \mathbb{1}\{\|\boldsymbol{\gamma} + \check{\boldsymbol{\xi}}\| \geq \mathbf{h}_0\}. \end{aligned}$$

Воспользуемся неравенствами 1.32 и 1.33:

$$\begin{aligned} e^{-\Delta^-} \|\bar{\boldsymbol{\eta}}\|^2 \mathbb{E} \mathbb{1}\{\|\boldsymbol{\gamma} + \check{\boldsymbol{\xi}}\| \geq \mathbf{r}_0\} &\leq e^{-\Delta^-} \|\bar{\boldsymbol{\eta}}\|^2 \exp\{-\mathbf{r}_0^2/4 + q/2 + \|\check{\boldsymbol{\xi}}\|^2/2\}, \\ e^{-\Delta^-} \mathbb{E} |\mathbf{u}^\top \boldsymbol{\gamma}|^2 \mathbb{1}\{\|\boldsymbol{\gamma} + \check{\boldsymbol{\xi}}\| \geq \mathbf{r}_0\} &\leq e^{-\Delta^-} (2 + |\mathbf{u}^\top \check{\boldsymbol{\xi}}|^2) \exp(-\mathbf{r}_0^2/4 + q/2 + \|\check{\boldsymbol{\xi}}\|^2/2) \\ &\leq e^{-\Delta^-} (2 + \|\check{\boldsymbol{\xi}}\|^2) \exp(-\mathbf{r}_0^2/4 + q/2 + \|\check{\boldsymbol{\xi}}\|^2/2). \end{aligned}$$

Согласно (1.15) на $\Omega(\mathbf{x})$ выполняется, что $\|\check{\boldsymbol{\xi}}\|^2 \leq z_B^2(\mathbf{x})$. Рассмотрим $\mathbf{h}_0^2 \geq 4\mathbf{x} + 2q + 2z_B^2(\mathbf{x})$, что влечет за собой

$$\mathbf{u}^\top S_\circ^2 \mathbf{u} \geq \exp(-\Delta^-) (1 + \|\bar{\boldsymbol{\eta}}\|^2 - \exp\{-\mathbf{x} + \log \varkappa\}),$$

где $\varkappa \stackrel{\text{def}}{=} 2 + \|\bar{\boldsymbol{\eta}}\|^2 + z_B^2(\mathbf{x})$.

Заметим, что мы можем положить $\mathbf{x} > \log \Delta^* + \log \varkappa$. Пусть $\|\bar{\boldsymbol{\eta}}\|^2 > 2\Delta^*$, тогда

$$\mathbf{u}^\top S_\circ^2 \mathbf{u} \geq (1 + \Delta^*) \exp(-\Delta^*).$$

Это неравенство противоречит (1.44) при $\Delta^* \leq 1/2$. Действительно, при $\mathbf{u} = \bar{\boldsymbol{\eta}}/\|\bar{\boldsymbol{\eta}}\|$ из верхней границы в (1.44) следует, что $\mathbf{u}^\top S_\circ^2 \mathbf{u} \leq \exp\{\Delta^*\} - 2\Delta^*$. Получаем противоречие в силу того, что $e^x - 2x \leq (1+x)e^{-x}$ при $0 \leq x \leq 1/2$. Также заметим, что неравенство $\|S_\circ^2 - I_q\| \leq 2\Delta^*$ напрямую следует из неравенства (1.42) в силу элементарного неравенства $e^x - 1 \leq 2x$ при $0 \leq x \leq 1/2$. Поэтому справедливы неравенства (1.43). \square

Оценка на первый момент с $\bar{\boldsymbol{\vartheta}} = \mathbb{E}^\circ \boldsymbol{\vartheta}$ влечет за собой неравенство

$$\|\check{D}(\bar{\boldsymbol{\vartheta}} - \boldsymbol{\theta}^\circ)\|^2 \leq 2\Delta^*,$$

в то время как вторая оценка дает соотношение

$$\|\check{D}\check{\mathfrak{C}}^2\check{D} - I_q\| \leq 2\Delta^*.$$

Последний результат теоремы следует напрямую из неравенств (1.26) и (1.29).

1.4.11. Доказательство теоремы 10

Определим $\mathcal{L}_G(\boldsymbol{v}) = \mathcal{L}(\boldsymbol{v}) - \|G\boldsymbol{v}\|^2/2$. Стохастическая компонента $\mathcal{L}_G(\boldsymbol{v})$ совпадает с стохастической компонентой $\mathcal{L}(\boldsymbol{v})$. Также квадратичный член $\|G\boldsymbol{v}\|^2/2$ не ухудшает гладкость математического ожидания процесса $\mathbb{E}\mathcal{L}_G(\boldsymbol{v})$. В частности, можно локально аппроксимировать $\mathbb{E}\mathcal{L}_G(\boldsymbol{v}_G^\circ, \boldsymbol{v})$ квадратичной функцией $\|\mathcal{D}_G(\boldsymbol{v} - \boldsymbol{v}_G^\circ)\|^2/2$ с

$$\begin{aligned} \boldsymbol{v}_G^\circ &\stackrel{\text{def}}{=} \boldsymbol{v}_G^* + \mathcal{D}_G^{-1}\boldsymbol{\xi}_G, & \boldsymbol{\xi}_G &\stackrel{\text{def}}{=} \mathcal{D}_G^{-1}\nabla\mathbb{E}\mathcal{L}_G(\boldsymbol{v}^*) = \mathcal{D}_G^{-1}\nabla\mathbb{E}\mathcal{L}(\boldsymbol{v}^*) + \mathcal{D}_G^{-1}G^2\boldsymbol{v}^*, \\ \boldsymbol{v}_G^* &\stackrel{\text{def}}{=} \operatorname{argmax}_{\boldsymbol{v} \in \mathcal{Y}} \mathbb{E}\mathcal{L}_G(\boldsymbol{v}), & \mathcal{D}_G^2 &\stackrel{\text{def}}{=} -\nabla^2\mathbb{E}\mathcal{L}(\boldsymbol{v}^*) + G^2 = \mathcal{D}^2 + G^2. \end{aligned}$$

Теперь можно легко увидеть, что все условия теоремы 9 выполняются для процесса $\mathcal{L}_G(\boldsymbol{v})$, если заменить \boldsymbol{v}° на \boldsymbol{v}_G° и \mathcal{D} на \mathcal{D}_G . Результат позволяет аппроксимировать апостериорное распределение $\boldsymbol{v} \mid \boldsymbol{Y}$ для нормального априорного распределения Π с помощью нормального распределения $\mathcal{N}(\boldsymbol{v}_G^\circ, \mathcal{D}_G^{-2})$. Теперь окончательный результат следует из леммы 3, если мы сможем ограничить $\|\mathcal{D}^{-1}\mathcal{D}_G^2\mathcal{D}^{-1} - I_p\|$ и $\|\mathcal{D}_G(\boldsymbol{v}^\circ - \boldsymbol{v}_G^\circ)\|$. По определению

$$\mathcal{D}^{-1}\mathcal{D}_G^2\mathcal{D}^{-1} - I_p = \mathcal{D}^{-1}G^2\mathcal{D}^{-1}.$$

Далее по определению $\nabla\mathbb{E}\mathcal{L}_G(\boldsymbol{v}) = \nabla\mathbb{E}\mathcal{L}(\boldsymbol{v}) - G^2\boldsymbol{v}$ для любого \boldsymbol{v} , а также $\nabla\mathbb{E}\mathcal{L}(\boldsymbol{v}^*) = \nabla\mathbb{E}\mathcal{L}_G(\boldsymbol{v}_G^*) = 0$. Получим

$$\nabla\mathbb{E}\mathcal{L}_G(\boldsymbol{v}_G^*) - \nabla\mathbb{E}\mathcal{L}_G(\boldsymbol{v}^*) = \mathcal{D}_G^2(\check{\boldsymbol{v}})(\boldsymbol{v}_G^* - \boldsymbol{v}^*),$$

где $\check{\mathbf{v}}$ является некоторой точкой из интервала между \mathbf{v}^* и \mathbf{v}_G^* и $\mathcal{D}_G^2(\check{\mathbf{v}}) = -\nabla^2 \mathbb{E} \mathcal{L}(\check{\mathbf{v}}) + G^2$. Таким образом

$$\mathcal{D}_G^2(\check{\mathbf{v}})(\mathbf{v}_G^* - \mathbf{v}^*) = G^2 \mathbf{v}^*. \quad (1.45)$$

Пусть $\mathbf{r}_G = \|\mathcal{D}_G(\mathbf{v}^\circ - \mathbf{v}_G^\circ)\|$. Согласно условию (\mathcal{L}_0) и (1.45) с вероятностью $\geq 1 - 2e^{-x}$

$$\begin{aligned} \mathbf{r}_G &= \|\mathcal{D}_G(\mathbf{v}^\circ - \mathbf{v}_G^\circ)\| \leq \|\mathcal{D}_G(\mathbf{v}^* - \mathbf{v}_G^*)\| + \|\boldsymbol{\xi}_G - \mathcal{D}_G \mathcal{D}^{-1} \boldsymbol{\xi}\| \\ &\leq \|\mathcal{D}_G \mathcal{D}_G^{-2}(\check{\mathbf{v}}) \mathcal{D}_G \mathcal{D}_G^{-1} G^2 \mathbf{v}^*\| + \|\mathcal{D}_G^{-1} \mathcal{D} \boldsymbol{\xi} + \mathcal{D}_G^{-1} G^2 \mathbf{v}^* - \mathcal{D}_G \mathcal{D}^{-1} \boldsymbol{\xi}\| \\ &\leq \{1 - \delta(\mathbf{r}_G)\}^{-1} \beta + \beta + \|\mathcal{D}_G^{-1} (\mathcal{D}^2 - \mathcal{D}_G^2) \mathcal{D}^{-1}\| \cdot \|\boldsymbol{\xi}\| \\ &\leq \{1 - \delta(\mathbf{r}_G)\}^{-1} \beta + \beta + \epsilon z_B(\mathbf{x}), \end{aligned}$$

где $z_B(\mathbf{x})$ определена в (1.15). В частности, $\delta(\mathbf{r}_G) \leq 1/2$ и $\epsilon \leq 1/2$ подразумевают $\mathbf{r}_G \leq 3\beta + \epsilon z_B(\mathbf{x})$. Тогда согласно лемме 3 для любого $A^* \subset \mathbb{R}^p$:

$$|\mathbb{P}_0(A^*) - \mathbb{P}_G(A^*)| \leq \frac{1}{2} \sqrt{\delta^2 + (1 + \epsilon)(3\beta + \epsilon z_B(\mathbf{x}))}, \quad (1.46)$$

где мера \mathbb{P}_0 является стандартной гауссовской мерой $\mathcal{N}(\mathbf{v}^\circ, \mathcal{D}^{-2})$, а мера \mathbb{P}_G является гауссовской мерой $\mathcal{N}(\mathbf{v}_G^\circ, \mathcal{D}_G^{-2})$. Теперь мы можем взять $A^* = A \times \mathbb{R}^{(p-q)}$. Тогда согласно (1.46) получим:

$$|\mathbb{P}_0(A^*) - \mathbb{P}_G(A^*)| = |\check{\mathbb{P}}_0(A) - \check{\mathbb{P}}_G(A)| \leq \frac{1}{2} \sqrt{\delta^2 + (1 + \epsilon)(3\beta + \epsilon z_B(\mathbf{x}))},$$

где мера $\check{\mathbb{P}}_0$ является гауссовской мерой $\mathcal{N}(\boldsymbol{\theta}^\circ, \check{\mathcal{D}}^{-2})$ и мера $\check{\mathbb{P}}_G$ является гауссовской мерой $\mathcal{N}(\boldsymbol{\theta}_G^\circ, \check{\mathcal{D}}_G^{-2})$. Таким образом, результат теоремы получен.

Модель независимых одинаково распределенных случайных величин и критическая размерность

В данной главе показывается, как общие результаты из предыдущих разделов могут быть соотнесены с классическими асимптотическими результатами в статистической литературе. Приятной особенностью всего подхода, основанного на локальном брэкетинге, является тот факт, что все результаты доказываются при одних и тех же условиях: проверив условия один раз, можно использовать любой из перечисленных результатов. Типичные примеры включают модель независимых одинаково распределенных случайных величин, обобщенные линейные модели и медианную регрессию. В данной главе мы кратко обсудим как теорема Бернштейна – фон Мизеса может быть применена к одному типичному случаю, а именно модели с независимыми одинаково распределенными случайными величинами.

2.1. Независимые одинаково распределенные случайные величины

Рассмотрим параметрическую модель независимых одинаково распределенных случайных величин, в которой случайные величины $\mathbf{Y} = (Y_1, \dots, Y_n)$ являются реализациями распределения P , которое принадлежит заданному параметрическому семейству $(P_{\mathbf{v}}, \mathbf{v} \in \mathcal{T})$ на пространстве наблюдений \mathcal{Y}_1 . Каждое значение параметра $\mathbf{v} \in \mathcal{T}$ явным образом задает вероятностную меру в виде произведения одномерных мер $P_{\mathbf{v}} = P_{\mathbf{v}}^{\otimes n}$ на пространстве $\mathcal{Y} = \mathcal{Y}_1^n$. В данном разделе мы иллюстрируем как общие результаты могут быть применены к данному типу моделей в случае возможной неверной спецификации модели. Можно рассмотреть различные типы неверной спецификации. Каждое из предположений, а

именно независимость данных, одинаковые распределения, параметрическая форма маргинального распределения, могут быть нарушены. В данном разделе для определенности мы рассмотрим один тип неверной спецификации, а именно мы предполагаем, что распределение одного наблюдения Y_i не обязательно принадлежит к параметрическому семейству $(P_{\mathbf{v}})$. При этом мы предположим, что наблюдения Y_i независимы и одинаково распределены. Случай неидентично распределенных наблюдений может быть рассмотрен аналогичным образом, но потребует более сложных обозначений.

Мы предположим, что для семейства $(P_{\mathbf{v}})$ существует доминирующая мера μ_0 , и каждая плотность $p(y, \mathbf{v}) = dP_{\mathbf{v}}/d\mu_0(y)$ является дважды непрерывно дифференцируемой по \mathbf{v} для всех y . Обозначим $\ell(y, \mathbf{v}) \stackrel{\text{def}}{=} \log p(y, \mathbf{v})$. Параметрическое предположение $Y_i \sim P_{\mathbf{v}^*} \in (P_{\mathbf{v}})$ позволяет определить правдоподобие

$$\mathcal{L}(\mathbf{v}) = \sum_{i=1}^n \ell(Y_i, \mathbf{v}),$$

где суммирование ведется по номеру измерения $i = 1, \dots, n$. Оценка максимума правдоподобия $\tilde{\mathbf{v}}$ максимизирует правдоподобие по $\mathbf{v} \in \mathcal{Y}$:

$$\tilde{\mathbf{v}} \stackrel{\text{def}}{=} \operatorname{argmax}_{\mathbf{v} \in \mathcal{Y}} \mathcal{L}(\mathbf{v}) = \operatorname{argmax}_{\mathbf{v} \in \mathcal{Y}} \sum_{i=1}^n \ell(Y_i, \mathbf{v}).$$

Цель оценивания \mathbf{v}^* максимизирует математическое ожидание $\mathcal{L}(\mathbf{v})$:

$$\mathbf{v}^* \stackrel{\text{def}}{=} \operatorname{argmax}_{\mathbf{v} \in \mathcal{Y}} \mathbb{E} \mathcal{L}(\mathbf{v}) = \operatorname{argmax}_{\mathbf{v} \in \mathcal{Y}} \mathbb{E} \ell(Y_1, \mathbf{v}).$$

Пусть $\zeta_i(\mathbf{v}) \stackrel{\text{def}}{=} \ell(Y_i, \mathbf{v}) - \mathbb{E} \ell(Y_i, \mathbf{v})$. Тогда $\zeta(\mathbf{v}) = \sum_{i=1}^n \zeta_i(\mathbf{v})$. Уравнение $\nabla \mathbb{E} \mathcal{L}(\mathbf{v}^*) = 0$ подразумевает, что

$$\nabla \zeta(\mathbf{v}^*) = \sum_{i=1}^n \nabla \zeta_i(\mathbf{v}^*) = \sum_{i=1}^n \nabla \ell_i(\mathbf{v}^*).$$

Структура независимых одинаково распределенных случайных величин Y_i позволяет переписать условия $(\mathcal{L}\mathbf{r})$, (ED_0) , (ED_2) , (\mathcal{L}_0) и (\mathcal{I}) в терминах маргинальных распределений. Начнем с условия на конечность экспоненциальных моментов для стохастической компоненты правдоподобия в точке \mathbf{v}^* :

(ed₀) Существует положительно определенная симметричная матрица ν_0 и константа $\nu_0 > 0$ такие, что выполняется неравенство

$$\sup_{\gamma \in \mathbb{R}^p} \log \mathbb{E} \exp \left\{ \mu \frac{\gamma^\top \nabla \zeta_1(\mathbf{v}^*)}{\|\nu_0 \gamma\|} \right\} \leq \nu_0^2 \mu^2 / 2, \quad \forall \mu: |\mu| \leq \mathfrak{g}_1.$$

Естественным кандидатом на роль ν_0^2 является ковариационная матрица градиента $\nabla \ell(Y_1, \mathbf{v}^*)$, т.е. $\nu_0^2 = \text{Var} \{ \nabla \ell(Y_1, \mathbf{v}^*) \} = \text{Var} \{ \nabla \zeta_1(\mathbf{v}^*) \}$.

Определим матрицу $\mathbb{F}_0 \stackrel{\text{def}}{=} -\nabla^2 \mathbb{E} \ell(Y_i, \mathbf{v}^*)$. Следующим шагом рассмотрим локальные множества

$$\mathcal{Y}_{loc}(\mathbf{u}) = \{ \mathbf{v} \in \mathcal{Y}: \|\mathbb{F}_0^{1/2}(\mathbf{v} - \mathbf{v}^*)\| \leq \mathbf{u} \}.$$

Так как $\mathcal{D}^2 = n\mathbb{F}_0$, то выполняется $\mathcal{Y}_0(\mathbf{r}) = \mathcal{Y}_{loc}(\mathbf{u})$ с $\mathbf{r}^2 = n\mathbf{u}^2$.

Ниже мы будем различать локальные условия, которые выполняются при $\mathbf{u} \leq \mathbf{u}_0$ для некоторого фиксированного значения \mathbf{u}_0 , и глобальные условия, которые выполняются для всех $\mathbf{u} > 0$.

Введем глобальное условие на стохастическую часть правдоподобия $\zeta_1(\mathbf{v})$, а точнее на ее гессиан $\nabla^2 \zeta_1(\mathbf{v})$:

(ed₂) Существуют константы $\nu_0, \omega^* > 0$ и для любого $\mathbf{u} > 0$ существует константа $\mathfrak{g}_1(\mathbf{u}) > 0$ такая, что выполняется неравенство

$$\sup_{\gamma_1, \gamma_2 \in \mathbb{R}^p} \log \mathbb{E} \exp \left\{ \mu \frac{\gamma_1^\top \nabla^2 \zeta_1(\mathbf{v}) \gamma_2}{\omega^* \|\mathbb{F}_0^{1/2} \gamma_1\| \cdot \|\mathbb{F}_0^{1/2} \gamma_2\|} \right\} \leq \nu_0^2 \mu^2 / 2, \quad \forall \mu: |\mu| \leq \mathfrak{g}_1(\mathbf{u}).$$

Перейдем к рассмотрению локального условия гладкости (\mathcal{L}_0) , которое требует определения функции $\delta(\mathbf{r})$ для $\mathbf{r} \leq \mathbf{r}_0$, где $\mathbf{r}_0^2 = n\mathbf{u}_0^2$. Если функция логарифма правдоподобия $\ell(y, \mathbf{v})$ является достаточно гладкой по параметру \mathbf{v} , то функция $\delta(\mathbf{r})$ может быть выбрана пропорциональной величине $\mathbf{u} = \mathbf{r}/n^{1/2}$. Более конкретно, определим матрицу $\mathbb{F}(\mathbf{v}) = -\nabla^2 \mathbb{E} \ell_1(\mathbf{v})$, тогда $\mathbb{F}_0 = \mathbb{F}(\mathbf{v}^*)$. В качестве естественного аналога условия (\mathcal{L}_0) сформулируем следующее условие:

(l₀) Существует константа δ^* такая, что для всех $\mathbf{u} \leq \mathbf{u}_0$ выполняется

неравенство

$$\sup_{\mathbf{v} \in \mathcal{N}_{loc}(\mathbf{u})} \left\| \mathbb{F}_0^{-1} \mathbb{F}^2(\mathbf{v}) \mathbb{F}_0^{-1} - I \right\| \leq \delta^* \mathbf{u}.$$

Далее мы выразим глобальной идентификации $(\mathcal{L}\mathbf{r})$ в терминах информации Кульбака-Лейблера $\mathcal{K}(\mathbf{v}, \mathbf{v}^*) \stackrel{\text{def}}{=} -\mathbb{E}\{\ell(Y_i, \mathbf{v}) - \ell(Y_i, \mathbf{v}^*)\}$ для каждого i . В случае правильной спецификации модели $P = P_{\mathbf{v}^*}$, $\mathcal{K}(\mathbf{v}, \mathbf{v}^*)$ является информацией Кульбака-Лейблера между распределениями $P_{\mathbf{v}^*}$ и $P_{\mathbf{v}}$, а матрицы $\mathbf{v}_0^2 = \mathbb{F}_0$ совпадают и равны информационной матрице Фишера для параметрического семейства $(P_{\mathbf{v}})$ в точке \mathbf{v}^* .

(\ell\mathbf{u}) Для каждого $\mathbf{u} > 0$, существует $\mathbf{b}(\mathbf{u}) > 0$ такая, что выполняется неравенство

$$\mathcal{K}(\mathbf{v}, \mathbf{v}^*) \geq \mathbf{b}(\mathbf{u}) \mathbf{u}^2, \quad \mathbf{u} = \|\mathbf{v}_0(\mathbf{v} - \mathbf{v}^*)\|.$$

Остается сформулировать условия идентифицируемости. Во избежание введения дополнительных обозначений мы просто предположим выполнение условий (\mathcal{I}) , которые очевидным образом можно получить из соответствующих условий на маргинальное правдоподобие.

Лемма 6. Пусть случайные величины Y_1, \dots, Y_n независимы и одинаково распределены. Тогда из выполнения условий (ed_0) , (ed_2) , (ℓ_0) и $(\ell\mathbf{u})$ следует выполнение условий (ED_0) , (ED_2) , (\mathcal{L}_0) и $(\mathcal{L}\mathbf{r})$ с $\mathcal{V}^2 = n\mathbf{v}_0^2$, $\mathcal{D}^2 = n\mathbb{F}_0$, $\omega = \omega^*/n^{1/2}$, $\delta(\mathbf{r}) = \delta^*\mathbf{r}/n^{1/2}$, $\mathbf{g} = \mathbf{g}_1\sqrt{n}$ и $\mathbf{g}(\mathbf{r}) = \mathbf{g}_1(\mathbf{u})\sqrt{n}$.

Замечание 1. В данном замечании обсуждается вопрос того, как представленные условия соотносятся с условиями, которые обычно предполагаются в статистической литературе. Важное замечание касается выбора параметрического семейства $(P_{\mathbf{v}})$. В классической теории подразумевается, что истинная мера лежит в этом семействе, что позволяет рассмотреть наиболее слабые условия, необходимые для выполнения результата. В данной работе мы исходим из несколько другой точки

зрения: каким бы образом не было выбрано параметрическое семейство $(P_{\mathbf{v}})$, истинная мера в него никогда не попадет, и любая модель является только аппроксимацией реальности. С другой стороны, выбор параметрической модели $(P_{\mathbf{v}})$ всегда осуществляется статистиком. Иногда некоторые специальные свойства модели заставляют вносить нерегулярность в семейство. В остальных ситуациях условия гладкости на плотность $\ell(y, \mathbf{v})$ могут быть обеспечены подходящим выбором параметрического семейства $(P_{\mathbf{v}})$.

Предложенный список условий также включает условия на конечные экспоненциальные моменты (ed_0) и (ed_2) на градиент $\nabla\ell(Y_1, \mathbf{v})$ и гессиан $\nabla^2\ell(Y_1, \mathbf{v})$. Конечные экспоненциальные моменты необходимы, чтобы получить неасимптотические оценки риска оценивания. Заметим, что в классических концентрационных условиях требуются даже более сильные условия, в частности ограниченность рассматриваемых случайных величин.

Условие глобальной идентификации (ℓu) очень легко проверить в классическом асимптотическом подходе. Действительно, если параметрическое множество \mathcal{Y} компактно, то расстояние Кульбака-Лейблера $\mathcal{K}(\mathbf{v}, \mathbf{v}^*)$ непрерывно и положительно для всех $\mathbf{v} \neq \mathbf{v}^*$, т.е. (ℓu) выполняется автоматически с константой \mathbf{b} . Если \mathcal{Y} не является компактным, то условие также выполняется, но величина $\mathbf{b}(\mathbf{u})$ может зависеть от \mathbf{u} .

Ниже мы специфицируем общие результаты гл. 1 для случая независимых одинаково распределенных случайных величин.

2.1.1. Оценка больших уклонений

В данном разделе мы представляем некоторые достаточные условия, которые гарантируют малую вероятность события $\{\mathbf{v} \notin \mathcal{Y}_{loc}(\mathbf{u}_0) \mid \mathbf{Y}\}$ для фиксированного \mathbf{u}_0 .

В теореме 4 и следствии 5 накладывается ряд ограничений снизу на радиус локальной области \mathbf{r}_0 , которые совместно гарантируют экспоненциальную кон-

центрацию апостериорного распределения. Все эти условия выполняются, если взять $\mathbf{r}_0 > \mathbf{C}(z_B(\mathbf{x}) + z(p, \mathbf{x}))$, где \mathbf{C} является фиксированной константой, а величины $z_B(\mathbf{x})$ и $z(p, \mathbf{x})$ определены в (1.15) и в (1.35) соответственно. Вспоминая, что $\mathbf{r}_0 = n^{1/2}\mathbf{u}_0$, можно сформулировать следующую теорему, в которой мы неявно предположим, что при достаточно большом радиусе локальной окрестности \mathbf{r}_0 величина $\diamond(\mathbf{r}_0, \mathbf{x})$ может быть сделана малой. Порядок величины $\diamond(\mathbf{r}_0, \mathbf{x})$ при условии $\mathbf{r}_0 \geq \mathbf{C}(p + \mathbf{x})$ будет оценен ниже.

Утверждение 1. *Предположим выполнение условий данного раздела. Тогда, если для некоторого $\mathbf{u}_0 > 0$ выполняется неравенство*

$$n^{1/2}\mathbf{u}_0 > \mathbf{C}(z_B(\mathbf{x}) + z(p, \mathbf{x})), \quad (2.1)$$

то выполняется неравенство

$$P(\mathbf{v} \notin \Upsilon_{loc}(\mathbf{u}_0) \mid \mathbf{Y}) \leq e^{-\mathbf{x}}.$$

Замечание 2. Представленный результат помогает определить два значения \mathbf{u}_0 и n , которые предоставляют оценку на вероятность больших отклонений. С учетом условия (\mathcal{I}) условие (2.1) может быть записано в виде $n\mathbf{u}_0^2 \gg \mathbf{x} + p$. Другими словами, результат теоремы заявляет оценку больших отклонений для окрестности $\Upsilon_{loc}(\mathbf{u}_0)$ с \mathbf{u}_0^2 порядка p/n . В классической асимптотической статистике этот результат называется *корень из n состоятельность*. Наш подход позволяет получить данный результат в очень сильной форме и для конечного размера выборки.

2.1.2. Локальное оценивание

Теперь нам необходимо оценить ошибку локальной аппроксимации в случае независимых одинаково распределенных случайных величин. Напомним, что данная ошибка определяется случайной величиной $\diamond(\mathbf{r}_0, \mathbf{x})$, заданной выражением (1.12).

Сначала опишем аппроксимирующие линейные модели. Матрицы \mathbf{v}_0^2 и \mathbb{F}_0 из условий (ed_0) , (ed_2) и (ℓ_0) определяют их компоненты сноса и ковариации.

Определим

$$\boldsymbol{\xi} \stackrel{\text{def}}{=} \mathcal{D}^{-1} \nabla \zeta(\mathbf{v}^*) = (n\mathbb{F}_0)^{-1/2} \sum_{i=1}^n \nabla \ell(Y_i, \mathbf{v}^*).$$

Аппроксимирующий процесс для метода брэкетинга записывается следующим образом:

$$\mathbb{L}(\mathbf{v}, \mathbf{v}^*) = \boldsymbol{\xi}^\top \mathcal{D}(\mathbf{v} - \mathbf{v}^*) - \|\mathcal{D}(\mathbf{v} - \mathbf{v}^*)\|^2/2.$$

Это выражение может рассматриваться как правдоподобие линейной модели $\boldsymbol{\xi} = \mathcal{D}\mathbf{v} + \boldsymbol{\varepsilon}$ со стандартной нормальной ошибкой $\boldsymbol{\varepsilon}$. Оценка максимума правдоподобия $\tilde{\mathbf{v}}$ для этой модели имеет вид $\tilde{\mathbf{v}} = \mathcal{D}^{-1}\boldsymbol{\xi}$.

Утверждение 2. *Предположим (ed₀). При заданном \mathbf{u}_0 предположим (ed₂), (ℓ_0) и (ι) на $\Upsilon_{loc}(\mathbf{u}_0)$ и пусть $\omega = \omega^*/n^{1/2}$, $\delta(\mathbf{r}) = \delta^*\mathbf{r}/n^{1/2}$. Тогда результаты теоремы (1.13) и все ее следствия выполняются в случае модели независимых одинаково распределенных случайных величин с $\mathbf{r}_0^2 = n\mathbf{u}_0^2$. В частности, на случайном множестве доминирующей вероятности не менее $1 - 4e^{-x}$ выполняется неравенство*

$$|L(\mathbf{v}, \mathbf{v}^*) - \mathbb{L}(\mathbf{v}, \mathbf{v}^*)| \leq \mathbf{r}_0 \diamond(\mathbf{r}_0, \mathbf{x}), \quad \mathbf{v} \in \Upsilon_0(\mathbf{r}_0),$$

где $\diamond(\mathbf{r}_0, \mathbf{x}) = \{\delta^*\mathbf{r}_0 + 6\nu_0 z_{\mathbb{H}}(\mathbf{x}) \omega^*\} \mathbf{r}_0/n^{1/2}$.

Теперь кратко обсудим применение полученных выше результатов к классическому асимптотическому подходу с $n \rightarrow \infty$. Допустим, что размерность также растет с размером выборки, т.е. $p = p_n \rightarrow \infty$. Зафиксируем $\mathbf{r}_0^2 = \mathbb{C}p_n$ для константы \mathbb{C} , обеспечивающей оценку больших уклонений из утверждения 1. Заметим, что $z_{\mathbb{H}}(\mathbf{x})$ имеет порядок $\sqrt{p_n}$. Таким образом, для достаточно большого размера выборки n получаем, что величина $\diamond(\mathbf{r}_0, \mathbf{x}) \asymp p_n/n^{1/2}$.

Таким образом получаем, что верна следующая теорема, более строгое доказательство которой можно найти ниже.

Теорема 11. Пусть выполнены условия из разд. 2.1. Пусть также $p_n \rightarrow \infty$ и $p_n^2/n \rightarrow 0$. Тогда результат теоремы 9 справедлив с $\mathbf{h}_0 \diamond (\mathbf{r}_0, \mathbf{x}) = \mathbf{C} \sqrt{p_n^2 q/n}$, $\mathcal{D}^2 = n\mathbb{F}_0$, где \mathbb{F}_0 – информационная матрица Фишера для распределения $(P_{\mathbf{v}})$ в точке \mathbf{v}^* .

Заметим, что в случае, когда целью оценивания является полный параметр, то согласно результату теоремы 11 достаточным условием для выполнения теоремы БфМ является условие $p_n^3/n \rightarrow 0$. Похожий результат об асимптотической нормальности апостериорного распределения в модели линейной регрессии может быть найден в [7]. Однако сходимость доказана при условии $p_n^4 \log(p_n)/n \rightarrow 0$, которое является слишком строгим. В работе [47] показано, что ограничение на рост размерности может быть смягчено до $p_n^3/n \rightarrow 0$ для экспоненциальных моделей. В работе [6] доказана теорема БфМ в особом классе моделей с независимыми одинаково распределенными случайными величинами в случае дискретных распределений вероятности при условии $p_n^3/n \rightarrow 0$. Некоторые дополнительные примеры и соответствующие условия для гауссовской модели представлены в [43], а для обобщенных линейных моделей в работе [48].

2.2. Доказательства

В этом разделе собраны доказательства результатов главы 2.

2.2.1. Доказательство леммы 6

Тождества $\mathcal{V}^2 = n\nu_0^2$, $\mathcal{D}^2 = n\mathbb{F}_0$ следуют из того, что выборка Y_i обладает структурой независимых одинаково распределенных наблюдений. Кратко прокомментируем вывод условия (ED_0) . С учетом того, что $\mathcal{V}^2 = n\nu_0^2$, можно показать, что

$$\log \mathbb{E} \exp \left\{ \mu \frac{\boldsymbol{\gamma}^\top \nabla \zeta(\mathbf{v})}{\|\mathcal{V}\boldsymbol{\gamma}\|} \right\} = n \log \mathbb{E} \exp \left\{ \frac{\mu}{n^{1/2}} \frac{\boldsymbol{\gamma}^\top \nabla \zeta_1(\mathbf{v})}{\|\mathbf{v}_0 \boldsymbol{\gamma}\|} \right\} \leq \nu_0^2 \mu^2 / 2,$$

так как $\mu \leq n^{1/2} \mathbf{g}_1 \leq \mathbf{g}$.

Обозначим $\omega = \omega^* n^{-1/2}$ и осуществим теперь проверку условия (ED_2) :

$$\begin{aligned} & \log \mathbb{E} \exp \left\{ \mu \frac{\boldsymbol{\gamma}_1^\top \nabla^2 \zeta(\mathbf{v}) \boldsymbol{\gamma}_2}{\omega \|\mathcal{D}\boldsymbol{\gamma}_1\| \cdot \|\mathcal{D}\boldsymbol{\gamma}_2\|} \right\} \\ &= n \log \mathbb{E} \exp \left\{ \frac{\mu}{n^{1/2}} \frac{\boldsymbol{\gamma}_1^\top \nabla^2 \zeta_1(\mathbf{v}) \boldsymbol{\gamma}_2}{\omega^* \|\mathbb{F}_0^{1/2} \boldsymbol{\gamma}_1\| \cdot \|\mathbb{F}_0^{1/2} \boldsymbol{\gamma}_2\|} \right\} \leq \nu_0^2 \mu^2 / 2, \end{aligned}$$

так как $\mu \leq n^{1/2} \mathbf{g}_1(\mathbf{u}) \leq \mathbf{g}(\mathbf{r})$.

2.2.2. Доказательство теоремы 11

Оценки брэккетинга из следствия 1 и больших уклонений из теоремы 3 применимы, если размер выборки n удовлетворяет $n \geq \mathbf{C}(p_n + \mathbf{x})$ для фиксированной константы \mathbf{C} . Оказывается, что для теоремы БфМ требуется более строгое условие. Действительно в регулярном случае независимых одинаково распределенных случайных величин выполняется

$$\delta(\mathbf{r}_0) \asymp \mathbf{r}_0 / \sqrt{n}, \quad z_{\mathbb{H}}^2(\mathbf{x}_n) \asymp p_n + \mathbf{x}_n, \quad \omega \asymp 1 / \sqrt{n},$$

где $a \asymp b$ означает, что $a = O(b)$ и $b = O(a)$ при $n \rightarrow \infty$. Радиус \mathbf{r}_0 должен удовлетворять условию $\mathbf{r}_0^2 \geq \mathbf{C}(p_n + \mathbf{x})$, чтобы обеспечить оценку больших уклонений для полного параметра, а для целевого параметра аналогично должно выполняться $\mathbf{h}_0^2 \geq \mathbf{C}(q + \mathbf{x})$. Из этого следует, что

$$\mathbf{h}_0 \diamond (\mathbf{r}_0, \mathbf{x}) = \mathbf{h}_0 (\delta(\mathbf{r}_0) + 3\nu_0 z_{\mathbb{H}}^2(\mathbf{x}_n) \omega) \mathbf{r}_0 \geq \mathbf{C} \sqrt{(p_n + \mathbf{x})^2 (q + \mathbf{x}) / n}.$$

Если зафиксировать $\mathbf{x} = \mathbf{C}q$, то теорема БфМ требует выполнения условия $p_n^2 q / n \rightarrow 0$ при $n \rightarrow \infty$.

Глава 3

Теорема Бернштейна – фон Мизеса в случае бесконечной размерности мешающего параметра

В данной главе исследуется вопрос применения предыдущих результатов к случаю, когда мешающий параметр имеет бесконечную размерность. Более конкретно, мы рассмотрим $(\boldsymbol{\theta}, \mathbf{f})$ -модель, в которой $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^q$ и $\mathbf{f} \in \mathcal{H}$ для некоторого гильбертова пространства \mathcal{H} . Предположим, что в \mathcal{H} существует счетный базис $\mathbf{e}_1, \mathbf{e}_2, \dots$. Тогда

$$\mathbf{f} = \mathbf{f}(\boldsymbol{\phi}) = \sum_{j=1}^{\infty} \phi_j \mathbf{e}_j \in \mathcal{H},$$

где вектор $\boldsymbol{\phi} = \{\phi_j\}_{j=1}^{\infty} \in \ell_2$ и $\phi_j = \langle \mathbf{f}, \mathbf{e}_j \rangle$.

Обозначим функцию правдоподобия для полной модели как $\mathcal{L}(\boldsymbol{\theta}, \mathbf{f})$. Также обозначим $\mathbf{v} = (\boldsymbol{\theta}, \boldsymbol{\phi})$ и

$$\mathcal{L}(\mathbf{v}) = \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}) = \mathcal{L}(\boldsymbol{\theta}, \mathbf{f}(\boldsymbol{\phi})).$$

“Истинные” значения полного и целевого параметра могут быть определены через максимизацию математического ожидания правдоподобия:

$$\mathbf{v}^* \stackrel{\text{def}}{=} \underset{\mathbf{v}=(\boldsymbol{\theta}, \boldsymbol{\phi})}{\operatorname{argmax}} \mathbb{E} \mathcal{L}(\mathbf{v}), \quad \boldsymbol{\theta}^* \stackrel{\text{def}}{=} \Pi_0 \mathbf{v}^*. \quad (3.1)$$

Также определим информационную матрицу \mathcal{D}^2 для полного параметра \mathbf{v} и эффективную информационную матрицу $\check{\mathcal{D}}^2$ для целевого параметра $\boldsymbol{\theta}$:

$$\begin{aligned} \mathcal{D}^2 &\stackrel{\text{def}}{=} -\nabla^2 \mathbb{E} \mathcal{L}(\mathbf{v}^*) \in \operatorname{Lin}(\ell_2, \ell_2), \\ \check{\mathcal{D}}^2 &\stackrel{\text{def}}{=} (\Pi_0 \mathcal{D}^{-2} \Pi_0^\top)^{-1} \in \mathbb{R}^{q \times q}, \end{aligned}$$

где $\operatorname{Lin}(\ell_2, \ell_2)$ – пространство линейных операторов из ℓ_2 в ℓ_2 .

3.1. Метод проекционных оценок

Мы применим метод проекционных оценок [49] и рассмотрим случай применения конечномерного неинформативного априорного распределения для параметров θ и ϕ . Заметим, что метод проекционных оценок является частным случаем метода решета (sieve approach), см. классическую работу Гренандера [50]. Для желающих подробнее ознакомиться с данным подходом в применении к семи- и непараметрическим задачам статистического оценивания рекомендуем обратиться к обзору [51]. Главным вопросом исследования является то, как аппроксимация влияет на свойства апостериорного распределения.

Пусть $\eta = \{\eta_j\}_{j=1}^m$ является проекцией мешающего параметра ϕ на конечномерное подпространство первых m компонент мешающего параметра. Для удобства обозначений представим $\phi = (\eta, \varkappa)$. В таком случае аппроксимация проекционной оценкой соответствует ситуации с $\varkappa \equiv 0$. Запишем “истинную” точку \mathbf{v}^* в виде $\mathbf{v}^* = (\theta^*, \eta^*, \varkappa^*)$. Аппроксимация функционального мешающего параметра ϕ с помощью m -мерного параметра η приводит к двум источникам смещения, связанным с проектированием функционального параметра на конечномерное пространство, порождаемое первыми m базисными функциями. Первый из них вызван игнорированием компоненты \varkappa . Аппроксимация целевого параметра θ_m^* , определенная как

$$\theta_m^* \stackrel{\text{def}}{=} \operatorname{argmax}_{\theta} \max_{\eta} \mathbb{E}L(\theta, \eta, 0), \quad (3.2)$$

может быть отлична от истинного значения θ^* . Другой источник смещения связан с заменой эффективной информационной матрицы Фишера \check{D}^2 ее аналогом для случая усечения базиса \check{D}_m^2 (см. определения ниже). Величины смещения могут быть оценены при предположениях гладкости на модель и на функциональный мешающий параметр ϕ с использованием стандартных методов теории аппроксимации. Чтобы избежать громоздких вычислений мы в некотором виде предположим, что аппроксимация с помощью проекционной оценки является со-

стоятельной. Обозначения упрощаются, если мы также предположим, что базис \mathbf{e}_m в пространстве \mathcal{H} выбран таким образом, чтобы обеспечить ортогональность блока H^2 информационной матрицы Фишера, т.е. $H^2 = I_\phi$. Очевидным образом такая же структура сохраняется и для ее блоков H_η^2 и H_ε^2 . Заметим, что общая ситуация может быть сведена к ортогональному случаю с помощью простого линейного преобразования мешающего параметра ϕ . Таким образом полная информационная матрица Фишера может быть представлена в виде

$$\mathcal{D}^2 = \begin{pmatrix} D^2 & \mathcal{A} \\ \mathcal{A}^\top & I_\phi \end{pmatrix} = \begin{pmatrix} D^2 & A_m & C_m \\ A_m^\top & I_\eta & 0 \\ C_m^\top & 0 & I_\varepsilon \end{pmatrix}. \quad (3.3)$$

В модели с усеченным базисом $\mathbf{v}_m = (\boldsymbol{\theta}, \boldsymbol{\eta}, 0)$ рассмотрим матрицу Фишера \mathcal{D}_m^2 с

$$\mathcal{D}_m^2 = \begin{pmatrix} D^2 & A_m \\ A_m & I_\eta \end{pmatrix}. \quad (3.4)$$

Также обозначим $\mathcal{D}_m^2(\mathbf{v}_m) = -\nabla_m^2 \mathbb{E} \mathcal{L}(\mathbf{v}_m)$, где ∇_m обозначает проекцию градиента на подпространство переменных $(\boldsymbol{\theta}, \boldsymbol{\eta})$. Предположим, что условия раздела 1.2.1 выполнены для аппроксимации правдоподобия $\mathcal{L}(\mathbf{v}_m)$. Необходимо также сформулировать некоторые дополнительные условия, которые связывают полный параметр \mathbf{v} и его аппроксимацию проекционной оценкой \mathbf{v}_m . Первое условие гарантирует семипараметрическую идентифицируемость и позволяет отделить целевой и мешающий параметры. Формально оно требует, чтобы угол между касательными подпространствами этих параметров был отделен от нуля:

(\mathcal{I}^f) Существует $\nu < 1$ такое, что выполняется неравенство

$$\|D^{-1} \mathcal{A} \mathcal{A}^\top D^{-1}\| \leq \nu.$$

Также для простоты формулировки результатов введем аналог условия (\mathcal{L}_0) для усеченной модели:

(\mathcal{L}_f) Для любого $\mathbf{r} \leq \mathbf{r}_0$ существует константа $\delta(\mathbf{r}) > 0$ такая, что на множестве $\mathcal{Y}_0(\mathbf{r})$ выполняется неравенство:

$$\|\mathcal{D}^{-1}\mathcal{D}^2(\mathbf{v})\mathcal{D}^{-1} - I\| \leq \delta(\mathbf{r}).$$

Заметим, что функция $\delta(\cdot)$ взята такой же, как и в условии (\mathcal{L}_0), для простоты обозначений. Также заметим, что в обозначениях данного раздела условие (\mathcal{L}_0) записывается для матриц \mathcal{D}_m^2 и $\mathcal{D}_m^2(\mathbf{v}_m)$.

Условия гладкости для параметров $\boldsymbol{\theta}$ и $\boldsymbol{\phi}$ выражаются через компоненты $\boldsymbol{\varkappa}^*$ полного параметра \mathbf{v}^* и блок C_m матрицы \mathcal{D}^2 .

(\mathcal{B}) Существуют $\rho_m, b_m \leq 1/2$ такие, что выполняются неравенства

$$\begin{aligned} \|D^{-1}C_m\boldsymbol{\varkappa}^*\| &\leq \rho_m, \\ \|D^{-1}C_mC_m^\top D^{-1}\| &\leq b_m \leq 1/2. \end{aligned} \quad (3.5)$$

Для состоятельности наших результатов необходимо, чтобы значение m было зафиксировано таким образом, чтобы величины ρ_m и b_m были достаточно малыми. Эти величины могут быть ограничены сверху при обычных условиях на гладкость функционального параметра \mathbf{f} , например в случае, когда \mathbf{f} принадлежит соболевскому шару определенной регулярности; см. например работы [17, 18, 42].

Также смотрите пример вычисления величин ρ_m и b_m в разделе 4.3 ниже.

Рассмотрим неинформативное априорное распределение, задающее равномерную плотность для параметров усеченного базиса $(\boldsymbol{\theta}, \boldsymbol{\eta})$ и задающее сингулярную массу в точке 0 для компонент мешающего параметра $\boldsymbol{\varkappa}$. Мы сфокусируемся на апостериорном распределении целевого параметра. Предположим, что условия теоремы 9 и следствия 9 выполнены для данного априорного распределения. Определим эффективную информационную матрицу Фишера \check{D}_m^2 и вектор

$\boldsymbol{\theta}_m^\circ$ как

$$\begin{aligned}\check{D}_m^2 &\stackrel{\text{def}}{=} D^2 - A_m A_m^\top, \\ \boldsymbol{\theta}_m^\circ &\stackrel{\text{def}}{=} \boldsymbol{\theta}_m^* + \check{D}_m^{-1} \check{\boldsymbol{\xi}}_m = \boldsymbol{\theta}_m^* + \check{D}_m^{-1} (\nabla_{\boldsymbol{\theta}} - A_m \nabla_{\boldsymbol{\eta}}) \mathcal{L}(\mathbf{v}_m^*).\end{aligned}$$

Теорема 9 гарантирует результат БФМ для неинформативного априорного распределения на пространстве параметров усеченного базиса $\boldsymbol{\theta}$ и $\boldsymbol{\eta}$: апостериорное распределение $\boldsymbol{\theta}$ аппроксимируется гауссовским распределением $\mathcal{N}(\boldsymbol{\theta}_m^\circ, \check{D}_m^{-2})$. Основной вопрос состоит в том, вносит ли усечение базиса значительный сдвиг в апостериорное распределение. Для полной семипараметрической модели определим

$$\begin{aligned}\check{D}^2 &\stackrel{\text{def}}{=} D^2 - \mathcal{A} \mathcal{A}^\top = D^2 - A_m A_m^\top - C_m C_m^\top, \\ \boldsymbol{\theta}^\circ &\stackrel{\text{def}}{=} \boldsymbol{\theta}^* + \check{D}^{-1} \check{\boldsymbol{\xi}} = \boldsymbol{\theta}^* + \check{D}^{-1} (\nabla_{\boldsymbol{\theta}} - \mathcal{A} \nabla_{\boldsymbol{\phi}}) \mathcal{L}(\mathbf{v}^*).\end{aligned}$$

Вектор $\boldsymbol{\theta}^\circ$ и эффективная матрица Фишера \check{D}^2 естественным образом возникают в случае бесконечномерного гауссовского распределения как апостериорное среднее и матрица влияния целевого параметра для несобственного неинформативного апостериорного распределения. Следующий результат дополняет теорему 9. При выполнении условия (\mathcal{I}^f) и условия гладкости (B) , он позволяет измерить расстояние между гауссовской мерой $\mathcal{N}(\boldsymbol{\theta}_m^\circ, \check{D}_m^{-2})$, которая аппроксимирует апостериорное распределение для усеченного базиса, и гауссовской мерой $\mathcal{N}(\boldsymbol{\theta}^\circ, \check{D}^{-2})$, соответствующей полноразмерному априорному распределению. Согласно лемме 3 эти две меры близки друг к другу, если отношение матриц $\check{D}_m^{-1} \check{D}^2 \check{D}_m^{-1}$ близко к единичной матрице, а нормализованная разница средних значений $\check{D}(\boldsymbol{\theta}^\circ - \boldsymbol{\theta}_m^\circ)$ мала.

Теорема 12. *Рассмотрим семипараметрическую модель с квази правдоподобием $L(\boldsymbol{\theta}, \boldsymbol{\phi})$. Истинное значение $(\boldsymbol{\theta}^*, \boldsymbol{\phi}^*)$ задано (3.1). Пусть $\mathcal{D}^2 \stackrel{\text{def}}{=} -\nabla^2 \mathbb{E} \mathcal{L}(\boldsymbol{\theta}^*, \boldsymbol{\phi}^*)$ является соответствующим фишеровским оператором. Предположим, что мешающий параметр $\boldsymbol{\phi}$ перенормирован таким образом, чтобы обеспечить, что*

соответствующий ϕ -блок матрицы \mathcal{D}^2 равен единичной матрица. Пусть $(\boldsymbol{\eta}, 0)$ является конечномерной аппроксимацией функционального мешающего параметра $\boldsymbol{\phi} = (\boldsymbol{\eta}, \boldsymbol{\varkappa})$, и уравнение (3.3) задает соответствующее блочное представление матрицы \mathcal{D}^2 . Предположим выполнение условий (\mathcal{I}^f) , (\mathcal{L}_f) , (\mathcal{I}) , (\mathcal{L}_0) и условия гладкости (B) . Тогда эффективные информационные матрицы Фишера \check{D}^2 и \check{D}_m^2 в полной и усеченной моделях удовлетворяют неравенствам

$$\|\check{D}_m^{-1}\check{D}^2\check{D}_m^{-1} - I_q\| \leq (1 - \nu)^{-1}\rho_m, \quad (3.6)$$

$$\text{tr}\{(\check{D}_m^{-1}\check{D}^2\check{D}_m^{-1} - I_q)^2\} \leq (1 - \nu)^{-2}q\rho_m^2. \quad (3.7)$$

Пусть выполняется, что величины $\delta(\mathbf{r}_m)$, $\delta(3\mathbf{r}_m) \leq 1/2$, где $\mathbf{r}_m = \|\boldsymbol{\varkappa}^*\|$. Целевой параметр $\boldsymbol{\theta}^*$ и его аналог в усеченной модели $\boldsymbol{\theta}_m^*$ связаны соотношением

$$\|\check{D}(\boldsymbol{\theta}^* - \boldsymbol{\theta}_m^*)\| \leq (1 - \nu)^{-1/2}\{\rho_m + 2\delta(\mathbf{r}_m)\mathbf{r}_m\} + 3\delta(3\mathbf{r}_m)\mathbf{r}_m. \quad (3.8)$$

Более того, выполняется неравенство

$$\begin{aligned} \|\check{D}(\boldsymbol{\theta}^\circ - \boldsymbol{\theta}_m^\circ)\| &\leq (1 - \nu)^{-1}\rho_m(\|\check{\boldsymbol{\xi}}_m\| + \|\boldsymbol{\xi}_\varkappa\|) \\ &+ (1 - \nu)^{-1/2}\{\rho_m + 2\delta(\mathbf{r}_m)\mathbf{r}_m\} + 3\delta(3\mathbf{r}_m)\mathbf{r}_m, \end{aligned} \quad (3.9)$$

где $\boldsymbol{\xi}_\varkappa \stackrel{\text{def}}{=} \rho_m^{-1}D^{-1}C_m\nabla_{\boldsymbol{\varkappa}}$. Наконец, при выполнении условий (ED_0) и (\mathcal{I}) на множестве $\Omega(\mathbf{x})$ вероятности $P(\Omega(\mathbf{x})) \geq 1 - 4e^{-\mathbf{x}}$ выполняется неравенство

$$\begin{aligned} \|\check{D}(\boldsymbol{\theta}^\circ - \boldsymbol{\theta}_m^\circ)\| &\leq 2\mathbf{a}(1 - \nu)^{-1}\rho_m(q^{1/2} + 2\mathbf{x}) \\ &+ (1 - \nu)^{-1/2}\{\rho_m + 2\delta(\mathbf{r}_m)\mathbf{r}_m\} + 3\delta(3\mathbf{r}_m)\mathbf{r}_m. \end{aligned} \quad (3.10)$$

Мы заключаем, что метод проекционных оценок работает правильным образом, если величины $q^{1/2}\rho_m$ и b_m малы, а величина $\mathbf{r}_m = \|\boldsymbol{\varkappa}^*\|$ не слишком большая.

Теперь прокомментируем, что изменится в модели, если соответствующий мешающему параметру блок H^2 фишеровского оператора \mathcal{D}^2 не равен единичному.

Мы запишем \mathcal{D}^2 в виде

$$\mathcal{D}^2 = \begin{pmatrix} D^2 & A \\ A^\top & H^2 \end{pmatrix} = \begin{pmatrix} D^2 & A_m & C_m \\ A_m^\top & H_m & B_m \\ C_m^\top & B_m^\top & K_m \end{pmatrix}.$$

Потребуем дополнительное условие, состоящее в том, что матрица

$$\check{K}_m \stackrel{\text{def}}{=} K_m - B_m H_m^{-1} B_m^\top$$

является положительно определенной. В трансформации, описанной в теореме 12, величины $A_m A_m^\top$ и $C_m C_m^\top$ заменяются величинами

$$\check{A}_m \check{A}_m^\top = A_m H_m^{-1} A_m^\top,$$

$$\check{C}_m \check{C}_m^\top = C_m \check{K}_m^{-1} C_m^\top = C_m (K_m - B_m H_m^{-1} B_m^\top)^{-1} C_m^\top.$$

Эффективная информационная матрица Фишера записывается как

$$\check{D}^2 = D^2 - \check{A}_m \check{A}_m^\top - \check{C}_m \check{C}_m^\top.$$

$$\check{D}_m^2 = D^2 - \check{A}_m \check{A}_m^\top.$$

Условие идентифицируемости (I) означает, что матрица \check{D}^2 является положительно определенной. Условия (B) записываются как

$$\|D^{-1} C_m \check{K}_m^{-1/2} \boldsymbol{\varkappa}^*\| \leq b_m,$$

$$\|D^{-1} \check{C}_m \check{C}_m^\top D^{-1}\| \leq \rho_m.$$

Итак, все утверждения теоремы 12 продолжают выполняться с описанными выше изменениями.

3.2. Доказательство теоремы 12

Доказательство состоит из двух частей. Сначала сравним неслучайные величины \check{D}^2 и $\boldsymbol{\theta}^*$ с их аналогами в случае усечения базиса. Затем проанализируем разницу между случайными векторами $\check{\boldsymbol{\xi}}$ и $\check{\boldsymbol{\xi}}_m$.

Условия идентифицируемости (\mathcal{I}^f) гарантирует для матрицы $\check{D}_m^2 = D^2 - A_m A_m^\top$, что

$$\|D^{-1}\check{D}_m^2 D^{-1}\| \geq 1 - \nu. \quad (3.11)$$

Далее с учетом того, что $\check{D}^2 = D^2 - \mathcal{A}\mathcal{A}^\top$ и $\check{D}_m^2 = D^2 - A_m A_m^\top$, получим

$$\check{D}_m^2 - \check{D}^2 = D^2 - A_m A_m^\top - (D^2 - \mathcal{A}\mathcal{A}^\top) = C_m C_m^\top.$$

Использование условия гладкости (B) позволяет оценить

$$\begin{aligned} \|D^{-1}(\check{D}_m^2 - \check{D}^2)D^{-1}\| &= \|D^{-1}C_m C_m^\top D^{-1}\| \leq \rho_m, \\ \|\check{D}^{-1}(\check{D}_m^2 - \check{D}^2)\check{D}^{-1}\| &\leq (1 - \nu)^{-1} \|D^{-1}C_m C_m^\top D^{-1}\| \leq (1 - \nu)^{-1} \rho_m. \end{aligned} \quad (3.12)$$

Использование условия гладкости (3.5) подразумевает, что выполняются неравенства (3.6) и (3.7). Аналогичным образом мы можем сравнить матрицу \mathcal{D}_m^{-2} из (3.4) и соответствующие блок в случае усечения базиса $\check{\mathcal{D}}_m^{-2}$ матрицы \mathcal{D}^{-2} . Очевидно,

$$\check{\mathcal{D}}_m^2 = \mathcal{D}_m^2 - \text{block}\{C_m C_m^\top, 0\}. \quad (3.13)$$

В свою очередь $\boldsymbol{\theta}$ -блок из \mathcal{D}_m^{-2} равен $\check{\mathcal{D}}_m^{-2}$ и выполняется неравенство

$$\|\mathcal{D}_m^{-1}(\check{\mathcal{D}}_m^2 - \mathcal{D}_m^2)\mathcal{D}_m^{-1}\| = \|\check{\mathcal{D}}_m^{-1}C_m C_m^\top \check{\mathcal{D}}_m^{-1}\| \leq (1 - \nu)^{-1} \rho_m.$$

Это означает, что при $\rho_m \leq \nu$ выполняется неравенство

$$\|\mathcal{D}_m^{-1}\check{\mathcal{D}}_m^2 \mathcal{D}_m^{-1}\| \leq 1 + (1 - \nu)^{-1} \rho_m \leq (1 - \nu)^{-1}. \quad (3.14)$$

Далее мы ограничим сдвиг $\boldsymbol{\theta}^* - \boldsymbol{\theta}_m^*$ в величине $\boldsymbol{\theta}^*$, к которому приводит усечение базиса. Для начала рассмотрим гауссовское правдоподобие с

$$\mathbb{E}L(\mathbf{v}^*) - \mathbb{E}L(\mathbf{v}) = \|\mathcal{D}(\mathbf{v}^* - \mathbf{v})\|^2/2.$$

Также определим точку $\mathbf{v}_s^* = (\boldsymbol{\theta}_s^*, \boldsymbol{\eta}_s^*, 0)$ как точку, в которой достигается минимума

по параметрам $\mathbf{v}^s = (\boldsymbol{\theta}, \boldsymbol{\eta}, 0)$ функция $\|\mathcal{D}(\mathbf{v}^s - \mathbf{v}^*)\|^2$:

$$\begin{aligned} \mathbf{v}_s^* &= \operatorname{argmin}_{\mathbf{v}^s=(\boldsymbol{\theta}, \boldsymbol{\eta}, 0)} \|\mathcal{D}(\mathbf{v}^s - \mathbf{v}^*)\|^2 \\ &= \operatorname{argmin}_{(\boldsymbol{\theta}, \boldsymbol{\eta})} \{ (\boldsymbol{\theta} - \boldsymbol{\theta}^*)^\top D^2 (\boldsymbol{\theta} - \boldsymbol{\theta}^*) + \|\boldsymbol{\eta} - \boldsymbol{\eta}^*\|^2 \\ &\quad + 2(\boldsymbol{\theta} - \boldsymbol{\theta}^*)^\top A_m (\boldsymbol{\eta} - \boldsymbol{\eta}^*) - 2(\boldsymbol{\theta} - \boldsymbol{\theta}^*)^\top C_m \boldsymbol{\varkappa}^* \}. \end{aligned}$$

Простые вычисления позволяют получить, $\mathbf{v}_s^* = (\boldsymbol{\theta}_m^*, \boldsymbol{\eta}_m^*, 0)$ удовлетворяет

$$\begin{aligned} \boldsymbol{\eta}_s^* - \boldsymbol{\eta}^* &= -A_m^\top (\boldsymbol{\theta}_m^* - \boldsymbol{\theta}^*), \\ \boldsymbol{\theta}_s^* - \boldsymbol{\theta}^* &= (D^2 - A_m A_m^\top)^{-1} C_m \boldsymbol{\varkappa}^* = \check{D}_m^{-2} C_m \boldsymbol{\varkappa}^*. \end{aligned} \tag{3.15}$$

Далее, (3.15) с помощью простой алгебры можно получить

$$\mathcal{D}^2(\mathbf{v}^* - \mathbf{v}_s^*) = \left(0, 0, (I_{\boldsymbol{\varkappa}} - C_m^\top \check{D}_m^{-2} C_m) \boldsymbol{\varkappa}^* \right)$$

и, наконец,

$$\|\mathcal{D}(\mathbf{v}^* - \mathbf{v}_s^*)\|^2 = (\mathbf{v}^* - \mathbf{v}_s^*)^\top \mathcal{D}^2(\mathbf{v}^* - \mathbf{v}_s^*) \leq \|\boldsymbol{\varkappa}^*\|^2. \tag{3.16}$$

Вспомним определение истинной точки в усеченной модели $\mathbf{v}_m^* = (\boldsymbol{\theta}_m^*, \boldsymbol{\eta}_m^*, 0)$, задаваемой формулой (3.2). Заметим, что по определению

$$\mathbb{E}\mathcal{L}(\mathbf{v}_s^*) \leq \mathbb{E}\mathcal{L}(\mathbf{v}_m^*) \leq \mathbb{E}\mathcal{L}(\mathbf{v}^*).$$

Следовательно,

$$\mathbb{E}\mathcal{L}(\mathbf{v}^*) - \mathbb{E}\mathcal{L}(\mathbf{v}_m^*) \leq \mathbb{E}\mathcal{L}(\mathbf{v}^*) - \mathbb{E}\mathcal{L}(\mathbf{v}_s^*). \tag{3.17}$$

Воспользуемся гладкостью правдоподобия и получим для некоторых точек $\mathbf{v}_1 \in [\mathbf{v}_s^*, \mathbf{v}^*]$ и $\mathbf{v}_2 \in [\mathbf{v}_m^*, \mathbf{v}^*]$:

$$\mathbb{E}\mathcal{L}(\mathbf{v}^*) - \mathbb{E}\mathcal{L}(\mathbf{v}_s^*) = \|\mathcal{D}(\mathbf{v}_1)(\mathbf{v}^* - \mathbf{v}_s^*)\|^2/2,$$

$$\mathbb{E}\mathcal{L}(\mathbf{v}^*) - \mathbb{E}\mathcal{L}(\mathbf{v}_m^*) = \|\mathcal{D}(\mathbf{v}_2)(\mathbf{v}^* - \mathbf{v}_m^*)\|^2/2.$$

Из этих равенств и неравенства (3.17) получим неравенство:

$$\|\mathcal{D}(\mathbf{v}_2)(\mathbf{v}^* - \mathbf{v}_m^*)\|^2 \leq \|\mathcal{D}(\mathbf{v}_1)(\mathbf{v}^* - \mathbf{v}_s^*)\|^2.$$

Воспользуемся условием гладкости (\mathcal{L}_0) :

$$\begin{aligned} (1 - \delta(\|\mathcal{D}(\mathbf{v}_m^* - \mathbf{v}^*)\|))\|\mathcal{D}(\mathbf{v}_m^* - \mathbf{v}^*)\|^2 &\leq (1 + \delta(\|\mathcal{D}(\mathbf{v}_s^* - \mathbf{v}^*)\|))\|\mathcal{D}(\mathbf{v}_s^* - \mathbf{v}^*)\|^2 \\ &\leq (1 + \delta(\mathbf{r}_m))\mathbf{r}_m^2. \end{aligned}$$

Воспользуемся тем, что $\delta(\|\mathcal{D}(\mathbf{v}_m^* - \mathbf{v}^*)\|) \leq \frac{1}{2}$ и $\delta(\mathbf{r}_m) \leq \frac{1}{2}$ и получим

$$\|\mathcal{D}(\mathbf{v}^* - \mathbf{v}_m^*)\|^2 \leq 3\mathbf{r}_m^2. \quad (3.18)$$

Тогда с учетом (3.16) и (3.18) по неравенству треугольника получим

$$\|\mathcal{D}(\mathbf{v}_s^* - \mathbf{v}_m^*)\| \leq 3\mathbf{r}_m. \quad (3.19)$$

Мы воспользуемся линейной аппроксимацией градиента математического ожидания правдоподобия. Согласно (3.16) точка \mathbf{v}_s^* находится в окрестности $\mathcal{Y}_0(\mathbf{r}_m)$ точки \mathbf{v}^* с $\mathbf{r}_m = \|\boldsymbol{\varkappa}^*\|$. Согласно условию гладкости (\mathcal{L}_0)

$$\|\mathcal{D}^{-1}\{\nabla \mathbb{E}\mathcal{L}(\mathbf{v}^*) - \nabla \mathbb{E}\mathcal{L}(\mathbf{v}_s^*) - \mathcal{D}^2(\mathbf{v}^* - \mathbf{v}_s^*)\}\| \leq \delta(\mathbf{r}_m)\mathbf{r}_m. \quad (3.20)$$

Далее $\nabla \mathbb{E}\mathcal{L}(\mathbf{v}^*) = 0$ по определению точки $\mathbf{v}^* = (\boldsymbol{\theta}^*, \boldsymbol{\eta}^*, \boldsymbol{\varkappa}^*)$. Определим вектор

$$\mathbf{b} \stackrel{\text{def}}{=} \nabla \mathbb{E}\mathcal{L}(\mathbf{v}_s^*) + \mathcal{D}^2(\mathbf{v}^* - \mathbf{v}_s^*),$$

и \mathbf{b}_m как его проекцию на подпространство $(\boldsymbol{\theta}, \boldsymbol{\eta})$. Неравенство (3.20) означает, что $\|\mathcal{D}^{-1}\mathbf{b}\| \leq \delta(\mathbf{r}_m)\mathbf{r}_m$, и позволяет получить для проекции \mathbf{b}_m неравенство

$$\|\check{\mathcal{D}}_m^{-1}\nabla_m \mathbb{E}\mathcal{L}(\mathbf{v}_s^*)\| \leq \delta(\mathbf{r}_m)\mathbf{r}_m,$$

где $\check{\mathcal{D}}_m^2$ определяется уравнением (3.13). Совмещая с (3.14), получим

$$\|\mathcal{D}_m^{-1}\nabla_m \mathbb{E}\mathcal{L}(\mathbf{v}_s^*)\| \leq (1 - \nu)^{-1/2}\delta(\mathbf{r}_m)\mathbf{r}_m. \quad (3.21)$$

Аналогично в пространстве усеченного параметра с учетом (3.19) получим

$$\left\| \mathcal{D}_m^{-1} \left\{ \nabla_m \mathbb{E}\mathcal{L}(\mathbf{v}_m^*) - \nabla_m \mathbb{E}\mathcal{L}(\mathbf{v}_s^*) - \mathcal{D}_m^2(\mathbf{v}_m^* - \mathbf{v}_s^*) \right\} \right\| \leq 3\delta(3\mathbf{r}_m)\mathbf{r}_m$$

и таким образом с учетом $\nabla_m \mathbb{E}\mathcal{L}(\mathbf{v}_m^*) = 0$ и (3.21) получим неравенство

$$\|\mathcal{D}_m(\mathbf{v}_m^* - \mathbf{v}_s^*)\| \leq (1 - \nu)^{-1/2}\delta(\mathbf{r}_m)\mathbf{r}_m + 3\delta(3\mathbf{r}_m)\mathbf{r}_m.$$

Следующим шагом спроектируем на подпространство целевого параметра $\boldsymbol{\theta}$ и получим аналогично (3.15):

$$\|\check{D}_m(\boldsymbol{\theta}_m^* - \boldsymbol{\theta}_s^*) - \check{D}_m^{-1}C_m\boldsymbol{\varkappa}^*\| \leq \delta(\mathbf{r}_m)\mathbf{r}_m\{1 + (1 - \nu)^{-1/2}\} + 3\delta(3\mathbf{r}_m)\mathbf{r}_m.$$

Тогда получим

$$\|\check{D}_m(\boldsymbol{\theta}^* - \boldsymbol{\theta}_m^*)\| \leq (1 - \nu)^{-1/2}\{\rho_m + 2\delta(\mathbf{r}_m)\mathbf{r}_m\} + 3\delta(3\mathbf{r}_m)\mathbf{r}_m,$$

что завершает доказательство (3.8).

Теперь рассмотрим стохастическую часть $\check{D}_m^{-1}\check{\boldsymbol{\xi}}_m - \check{D}^{-1}\check{\boldsymbol{\xi}}$ вектора $\boldsymbol{\theta}^\circ - \boldsymbol{\theta}_m^\circ$. Оценки (3.12) и (3.11) подразумевают для $\check{\boldsymbol{\xi}} = \nabla_{\boldsymbol{\theta}} - A_m\nabla_{\boldsymbol{\eta}} - C_m\nabla_{\boldsymbol{\varkappa}}$ и $\check{\boldsymbol{\xi}}_m = \nabla_{\boldsymbol{\theta}} - A_m\nabla_{\boldsymbol{\eta}}$

$$\begin{aligned} & \left\| \check{D}\{\check{D}_m^{-1}\check{\boldsymbol{\xi}}_m - \check{D}^{-1}\check{\boldsymbol{\xi}}\} \right\| = \left\| \check{D}\check{D}_m^{-2}(\nabla_{\boldsymbol{\theta}} - A_m\nabla_{\boldsymbol{\eta}}) - \check{D}^{-1}(\nabla_{\boldsymbol{\theta}} - A_m\nabla_{\boldsymbol{\eta}} - C_m\nabla_{\boldsymbol{\varkappa}}) \right\| \\ & = \left\| \{\check{D}\check{D}_m^{-2} - \check{D}^{-1}\}(\nabla_{\boldsymbol{\theta}} - A_m\nabla_{\boldsymbol{\eta}}) + \check{D}^{-1}C_m\nabla_{\boldsymbol{\varkappa}} \right\| \\ & \leq \left\| \{I_q - \check{D}^{-1}\check{D}_m^2\check{D}^{-1}\}\check{D}\check{D}_m^{-2}(\nabla_{\boldsymbol{\theta}} - A_m\nabla_{\boldsymbol{\eta}}) \right\| + \|\check{D}^{-1}C_m\nabla_{\boldsymbol{\varkappa}}\| \\ & \leq (1 - \nu)^{-1}\rho_m\|\check{\boldsymbol{\xi}}_m\| + (1 - \nu)^{-1}\|D^{-1}C_m\nabla_{\boldsymbol{\varkappa}}\| \\ & = (1 - \nu)^{-1}\rho_m(\|\check{\boldsymbol{\xi}}_m\| + \|\boldsymbol{\xi}_{\boldsymbol{\varkappa}}\|). \end{aligned}$$

Здесь $\boldsymbol{\xi}_{\boldsymbol{\varkappa}} = \rho_m^{-1}D^{-1}C_m\nabla_{\boldsymbol{\varkappa}}$ и мы также используем, что $\check{D}_m^2 \geq \check{D}^2$. Это доказывает (3.9). Остается доказать (3.10). В полной модели выполняется, что $\text{Var}(\check{\boldsymbol{\xi}}_m) = I_q$ и при выполнении условия (ED_0) на множестве вероятности не менее $1 - 2e^{-x}$ верно неравенство

$$\|\check{\boldsymbol{\xi}}_m\| \leq \sqrt{q} + 2\mathbf{x}; \quad (3.22)$$

см. [38]. Аналогично, в случае правильной спецификации модели $\text{Var}(\nabla_{\boldsymbol{\varkappa}}) = I_{\boldsymbol{\varkappa}}$. Для $\boldsymbol{\xi}_{\boldsymbol{\varkappa}} = \rho_m^{-1}D^{-1}C_m\nabla_{\boldsymbol{\varkappa}}$ при выполнении условия (B) выполняется неравенство

$$\text{Var}(\boldsymbol{\xi}_{\boldsymbol{\varkappa}}) \leq \rho_m^{-2}\|D^{-1}C_mC_m^\top D^{-1}\| \leq 1,$$

и, таким образом, использование условия (ED_0) позволяет на множестве домини-

рующей вероятности $1 - 2e^{-x}$ сделать оценку

$$\|\xi_{\varkappa}\| \leq \sqrt{q} + 2x$$

аналогично (3.22). В общем случае, когда параметрическое предположение не выполнено в точности, мы тем не менее можем использовать условие (ED_0) . Оно гарантирует, что $\text{Var}(\check{\xi}_m) \leq \mathfrak{a}^2 I_q$, $\text{Var}(\xi_{\varkappa}) \leq \mathfrak{a}^2 I_q$ и

$$\|\check{\xi}_m\| \leq \mathfrak{a}(\sqrt{q} + 2x), \quad \|\xi_{\varkappa}\| \leq \mathfrak{a}(\sqrt{q} + 2x).$$

Это позволяет доказать последнее утверждение теоремы.

Глава 4

Примеры

В данной главе рассматриваются несколько примеров, которые иллюстрируют общие результаты раздела 1.2.

4.1. Линейная гауссовская регрессия и плоское гауссовское априорное распределение

Пусть $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$ – случайный вектор в \mathbb{R}^n , который удовлетворяет уравнению

$$\mathbf{Y} = \mathbf{f} + \boldsymbol{\varepsilon} = \Psi^\top \mathbf{v}^* + \boldsymbol{\varepsilon}, \quad (4.1)$$

где ошибки $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^\top$ – независимые с нулевым средним, а матрица плана Ψ размера $p \times n$ задана. Вектор средних значений $\mathbf{f} \in \mathbb{R}^n$ неизвестен и второе уравнение из (4.1) означает, что он принадлежит некоторому заданному p -мерному линейному подпространству в \mathbb{R}^n : $\mathbf{f} = \Psi^\top \mathbf{v}^*$ для неизвестного целевого вектора $\mathbf{v}^* \in \mathbb{R}^p$. Запишем матрицу Ψ в виде $\Psi = \{\Psi_1, \dots, \Psi_n\}$, таким образом $f_i = \Psi_i^\top \mathbf{v}^*$. Ниже мы предположим, что $n > p$ и ранг матрицы Ψ равен p , или, эквивалентно, строки Ψ являются линейно независимыми векторами в \mathbb{R}^n .

Сначала мы рассмотрим гауссовский случай, т.е. $\varepsilon_i \in \mathcal{N}(0, \sigma_n^2 I_n)$, $i = 1, \dots, n$, где I_n – единичная матрица размера $n \times n$. Дисперсия измерений σ_n^2 известна, но может зависеть от объема выборки n . Для простоты сравнения мы предположим, что матрица плана Ψ удовлетворяет условию $\Psi^\top \Psi = I_p$. В наших обозначениях это означает $\mathcal{D}^2 = \sigma_n^{-2} I_p$.

Для гауссовского априорного распределения апостериорное распределение является в точности гауссовским и аппроксимация функции логарифма правдоподобия квадратичной формой выполняется в точности. При использовании неин-

формативного априорного распределения единственным условием, необходимым для выполнения результата теоремы БфМ, является $p = p_n = o(n)$; см. [42]. В работе [43] показано, что результат БфМ может быть расширен на случай растущей размерности параметра $p = p_n$ и плоского гауссовского априорного распределения с ковариационной матрицей $\tau_n^2 I_n$ при условии малости величин $(\sigma_n/\tau_n)^4 p_n$ и $(\sigma_n/\tau_n^2) \|\mathbf{v}^*\|$. Наши результаты из раздела 10 покрывают случай гауссовского априорного распределения и единственными условиями, необходимыми для проверки теоремы БфМ, являются условия “ $\text{tr}(\mathcal{D}^{-2}G^2) = (\sigma_n/\tau_n)^4 p_n$ мало” и “ $\|\mathcal{D}_G^{-1}G^2\mathbf{v}^*\| = (\sigma_n/\tau_n^2) \|\mathbf{v}^*\|$ мало”. Легко видеть, что наши общие результаты легко применимы к случаю растущей размерности и являются настолько же точными, как и результаты, полученные для очень специального гауссовского случая.

4.2. Линейная негауссовская регрессия

Теперь рассмотрим более общую ситуацию, когда ошибки ε_i в (4.1) распределены согласно распределению общего вида с плотностью $f(\cdot)$: $\varepsilon_i \sim P_f$. Обозначим $g(x) = \log f(x)$. Функция логарифма правдоподобия для данной задачи записывается следующим образом

$$\mathcal{L}(\mathbf{v}) = \sum_{i=1}^n \log f(Y_i - \Psi_i^\top \mathbf{v}) = \sum_{i=1}^n g(Y_i - \Psi_i^\top \mathbf{v}). \quad (4.2)$$

Предположим, что функция $g(z)$ дважды непрерывно дифференцируема и пусть

$$k^2 \stackrel{\text{def}}{=} \int g''(z) f(z) dz < \infty.$$

Если модель корректно специфицирована, то

$$\mathcal{D}^2 = -\nabla^2 \mathbb{E} \sum_{i=1}^n g(Y_i - \Psi_i^\top \mathbf{v}^*) = \int g''(z) f(z) dz \cdot \sum_{i=1}^n \Psi_i \Psi_i^\top = k^2 \sum_{i=1}^n \Psi_i \Psi_i^\top. \quad (4.3)$$

Аналогично,

$$\mathcal{D}^2(\mathbf{v}) = -\nabla^2 \mathbb{E} \sum_{i=1}^n g(Y_i - \Psi_i^\top \mathbf{v}) = \sum_{i=1}^n \Psi_i \Psi_i^\top \int g''(z - \Psi_i^\top (\mathbf{v} - \mathbf{v}^*)) f(z) dz.$$

Для применения общих результатов из раздела 1.2.7 требуется проверка условий из раздела 1.2.1. Предположим выполнение некоторых условий типа конечности экспоненциальных моментов на распределение P_f :

(e₀) *Существуют константы ν_0 и $\mathbf{g}_1 > 0$ такие, что для случайной величины $\varepsilon \sim P_f$ выполняется неравенство*

$$\log \mathbb{E} \exp(\mu g'(\varepsilon)/k) \leq \nu_0^2 \mu^2 / 2, \quad \forall \mu: |\mu| \leq \mathbf{g}_1.$$

Условие (e₀) означает, что распределение ошибок ε имеет экспоненциально убывающий хвост. При выполнении условия (e₀) условие (ED₀) выполняется в соответствии со следующей леммой

Лемма 7. *Предположим выполнение условия (e₀) и пусть $\mathcal{V}^2 = \mathcal{D}^2 = k^2 \sum_{i=1}^n \Psi_i \Psi_i^\top$. Тогда условие (ED₀) следует из (e₀) с данной матрицей \mathcal{V}^2 и $\mathbf{g} = \mathbf{g}_1 N_1^{1/2}$, где*

$$N_1^{-1/2} \stackrel{\text{def}}{=} \max_i \sup_{\gamma \in \mathbb{R}^p} \frac{k |\Psi_i^\top \gamma|}{\|\mathcal{D}\gamma\|}. \quad (4.4)$$

Условия (\mathcal{L}_0) и (\mathcal{L}_r) также выполняются, если предположить, что функция $g''(\cdot)$ удовлетворяет условию Липшица.

Лемма 8. *Пусть*

$$|g''(z) - g''(z_0)| \leq L|z - z_0|, \quad z, z_0 \in \mathbb{R}. \quad (4.5)$$

Тогда

$$\delta(\mathbf{r}) \leq \frac{L \mathbf{r}}{k N_1^{1/2}},$$

где величина N_1 определена уравнением (4.4).

Следующий интересный вопрос заключается в проверке условия (ED₂). Стохастическая часть правдоподобия записывается следующим образом:

$$\zeta(\mathbf{v}) = \sum_{i=1}^n g(Y_i - \Psi_i^\top \mathbf{v}) - \mathbb{E} g(Y_i - \Psi_i^\top \mathbf{v}).$$

Чтобы проверить условие (ED_2) , необходимо подсчитать гессиан стохастической части правдоподобия $\zeta(\mathbf{v})$:

$$\nabla^2 \zeta(\mathbf{v}) = \sum_{i=1}^n \{g''(Y_i - \Psi_i^\top \mathbf{v}) - \mathbb{E}g''(Y_i - \Psi_i^\top \mathbf{v})\} \Psi_i \Psi_i^\top.$$

Наконец, нам необходимо наложить неограничительное условие на маргинальное правдоподобие:

(e₂) Существует константа ν_0 и для любого $\mathbf{r} > 0$ существует $\mathbf{g}(\mathbf{r}) > 0$ такие, что для любого δ с $|\delta| \leq N_2^{-1/2} \mathbf{r}$ выполняется

$$\log \mathbb{E} \exp \left(\frac{\mu}{\mathbf{s}_i} \{g''(Y_i + \delta) - \mathbb{E}g''(Y_i + \delta)\} \right) \leq \frac{\nu_0^2 \mu^2}{2}, \quad \forall \mu: |\mu| \leq \mathbf{g}(\mathbf{r}),$$

где \mathbf{s}_i – некоторые известные значения и

$$N_2^{-1/2} \stackrel{\text{def}}{=} \max_i \sup_{\gamma \in \mathbb{R}^p} \frac{\mathbf{s}_i |\Psi_i^\top \gamma|}{\|\mathcal{D}\gamma\|}.$$

Тогда

$$\begin{aligned} & \sup_{\gamma_1, \gamma_2 \in \mathbb{R}^p} \log \mathbb{E} \exp \left\{ \frac{\mu}{\omega} \frac{\gamma_1^\top \nabla^2 \zeta(\mathbf{v}) \gamma_2}{\|\mathcal{D}\gamma_1\| \cdot \|\mathcal{D}\gamma_2\|} \right\} \\ &= \sup_{\gamma_1, \gamma_2 \in \mathbb{R}^p} \log \mathbb{E} \exp \left\{ \frac{\mu}{\omega} \frac{\sum_{i=1}^n (g''(Y_i - \Psi_i^\top \mathbf{v}) - \mathbb{E}g''(Y_i - \Psi_i^\top \mathbf{v})) \gamma_1^\top \Psi_i \Psi_i^\top \gamma_2}{\|\mathcal{D}\gamma_1\| \cdot \|\mathcal{D}\gamma_2\|} \right\} \\ &\leq \sup_{\gamma_1, \gamma_2 \in \mathbb{R}^p} \frac{\nu_0^2 \mu^2 \sum_{i=1}^n \mathbf{s}_i^2 |\Psi_i^\top \gamma_1|^2 \cdot |\Psi_i^\top \gamma_2|^2}{2\omega^2 \|\mathcal{D}\gamma_1\|^2 \cdot \|\mathcal{D}\gamma_2\|^2} \\ &\leq \frac{\nu_0^2 \mu^2}{2\omega^2} \frac{n}{N_2^2}. \end{aligned}$$

Легко видеть, что достаточно положить $\omega = \frac{\sqrt{n}}{N_2}$, чтобы можно было показать выполнение условия (ED_2) . В регулярных случаях величина N_2 имеет порядок размера выборки n и таким образом $\omega \sim n^{-1/2}$.

Наконец, перейдем к семипараметрической модели, в которой $\mathbf{v} = (\boldsymbol{\theta}, \boldsymbol{\eta}) \in \mathbb{R}^p$ и $\boldsymbol{\theta} \in \mathbb{R}^q$. Предположим выполнение условия идентифицируемости (\mathcal{I}) из раздела 1.2.1. Тогда наши общие результаты позволяют получить семипараметрическую теорему БфМ для линейной модели (4.1).

Теорема 13. Пусть выполняются условия (e_0) , (e_2) , условия леммы 8 и условие (\mathcal{I}) для матрицы \mathcal{D}^2 из (4.3). Тогда результаты теоремы 9 выполняются для линейной модели (4.1) с $\mathbf{h}_0^2 \geq \mathbf{C}(q + \mathbf{x})$, $\mathbf{r}_0^2 \geq \mathbf{C}(p + \mathbf{x})$ и

$$\diamond(\mathbf{r}_0, \mathbf{x}) = \left\{ \delta(\mathbf{r}_0) + 6\nu_0 z_{\mathbb{H}}(\mathbf{x}) \omega \right\} \mathbf{r}_0 \leq \left\{ \frac{L}{k^2} \frac{\mathbf{r}_0}{N_1^{1/2}} + 6\nu_0 z_{\mathbb{H}}(\mathbf{x}) \frac{\sqrt{n}}{N_2} \right\} \mathbf{r}_0.$$

4.3. Семипараметрическая негауссовская линейная регрессия

В данном разделе полученные ранее результаты применяются к линейной негауссовской модели (4.1) с семипараметрической функцией регрессии:

$$\mathbf{f}^* = \Psi^\top \boldsymbol{\theta}^* + \mathbf{g}^*, \quad (4.6)$$

где $\boldsymbol{\theta}^* \in \mathbb{R}^q$ является неизвестным целевым вектором, а $\Psi = (\Psi_1, \dots, \Psi_n)$ – матрица размера $q \times n$ с $\Psi_i = (\psi_1(X_i), \dots, \psi_q(X_i))^\top \in \mathbb{R}^q$ для заданного набора базовых функций $\{\psi_j(\cdot), j = 1, \dots, q\}$ и точек плана эксперимента X_i , $i = 1, \dots, n$. Без ограничения общности можно предположить, что базисные функции ортонормированы относительно данного плана эксперимента:

$$\sum_{i=1}^n \psi_j(X_i) \psi_{j'}(X_i) = \delta_{j,j'}. \quad (4.7)$$

Общий случай можно свести к данному с помощью вращения и перенормирования. Аналогично мы предположим, что элементы вектора мешающего параметра $\mathbf{g}^* = \{g^*(X_1), \dots, g^*(X_n)\}^\top$ являются значениями в точках X_i функции g^* , которая является элементом функционального пространства. Это означает, что $g^* = g(x) = \sum_{k=1}^{\infty} \eta_k \varphi_k(x)$ для заданного функционального базиса $\{\varphi_k\}_{k=1}^{\infty}$ (например, Фурье, вейвлеты и так далее) и бесконечномерного вектора мешающего параметра $\boldsymbol{\eta} = \{\eta_1, \eta_2, \dots, \eta_m, \dots\}^\top$. Данная модель называется моделью частичной линейной регрессии, см. книгу Хердле и Лианга [52] для подробного обзора.

Более того, мы предположим, что g^* является гладкой, т.е. она может быть хорошо аппроксимирована конечной суммой $g_m^*(\cdot) = \sum_{k=1}^m \eta_k \varphi_k(\cdot)$ в следующем

СМЫСЛЕ

$$\|g^* - g_m^*\| \leq \gamma_m. \quad (4.8)$$

Например, если \mathcal{F}_s – соболевский шар, то есть

$$g^* \in \mathcal{F}_s(\mathbb{C}) \stackrel{\text{def}}{=} \left\{ g(x) = \sum_{k=1}^{\infty} \eta_k \varphi_k(x) : \sum_{k=1}^{\infty} \eta_k^2 k^{2s} \leq \mathbb{C} \right\},$$

то $\gamma_m \leq (m+1)^{-s}$.

В дополнение к предыдущему мы наложим такие же условия гладкости на каждую базисную функцию ψ_j для $j = 1, \dots, q$, то есть

$$\|\psi_j - \psi_{j,m}\| \leq \gamma_m, \quad (4.9)$$

где $\varphi_{j,m} = \Pi_m \varphi_j$ и Π_m является проектором на выпуклую оболочку первых m базисных функций $\varphi_1, \dots, \varphi_m$.

Чтобы избежать проблемы идентифицируемости мы ограничим разложение функции g до первых M коэффициентов для большого числа M , которое может зависеть от размера выборки n . Например, можно взять $M = n/\log(n)$. Или в качестве альтернативы можно взять $M = n^a$ для некоторого $a < 1$. Также для упрощения представления мы предположим, что базисные функции $\varphi_k(\cdot)$ ортонормальны в следующем смысле:

$$\sum_{i=1}^n \varphi_{k'}(X_i) \varphi_k(X_i) = \delta_{k',k}. \quad (4.10)$$

Также определим $\Phi_i = \{\varphi_1(X_i), \dots, \varphi_q(X_i)\}^\top \in \mathbb{R}^q$ для $i = 1, 2, \dots, n$. Тогда полный параметр модели записывается как $\mathbf{v} = (\boldsymbol{\theta}, \boldsymbol{\eta}) \in \mathbb{R}^p$, где $p = q + M$, и разложение (4.6) может быть переписано как

$$\mathbf{f} = \Xi \mathbf{v},$$

где $\Xi_i = (\Psi_i^\top, \Phi_i^\top)^\top$, $i = 1, \dots, n$ и $\Xi = (\Xi_1, \dots, \Xi_n)^\top$. Полная информационная

матрица Фишера записывается как

$$\mathcal{D}^2 = \begin{pmatrix} D^2 & A \\ A^\top & H^2 \end{pmatrix} = k^2 \begin{pmatrix} \sum_{i=1}^n \Psi_i \Psi_i^\top & \sum_{i=1}^n \Psi_i \Phi_i^\top \\ \sum_{i=1}^n \Phi_i \Psi_i^\top & \sum_{i=1}^n \Phi_i \Phi_i^\top \end{pmatrix}.$$

В силу условий ортогональности (4.7) и (4.10), блоки D^2 и H^2 пропорциональны единичной матрице: $D^2 = k^2 I_q$, $H^2 = k^2 I_M$. В дальнейшем мы предположим, что $k = 1$, расширение на общий случай представляется тривиальным. Условие идентифицируемости (1.9) может быть переписано как

$$\|AA^\top\| \leq \nu. \quad (4.11)$$

Условие (4.8) подразумевает ортогональность базисных функций φ_m

$$\|C_m \boldsymbol{\kappa}^*\|^2 = \|g^* - g_m^*\|^2 \leq \gamma_m^2.$$

Аналогично, условие гладкости (4.9) подразумевает, что норма каждой строки матрицы C_m может быть ограничена величиной γ_m . Таким образом

$$\|C_m C_m^\top\| \leq \text{tr}(C_m C_m^\top) \leq q \gamma_m^2.$$

Теорема 14. *Рассмотрим модель (4.1) с функцией логарифма правдоподобия (4.2) и семипараметрической регрессионной функцией из (4.6). Предположим выполнение условия идентифицируемости (4.11) и условий гладкости (4.8), (4.9). Тогда результат теоремы 12 выполняется с $\rho_m = b_m = (1 - \nu)^{-1} q^{1/2} \gamma_m^2$.*

Замечание 3. В случае, когда все функции $g^*, \psi_j, j = 1, \dots, q$ принадлежат к соболевскому шару гладкости s , мы можем заключить, что $\gamma_m \leq (m + 1)^{-s}$ и $\rho_m = b_m \leq (1 - \nu)^{-1} q^{1/2} (m + 1)^{-2s}$.

4.4. Обобщенные линейные модели

Теперь мы рассмотрим обобщенные линейные модели, которые часто используются для описания категориальных данных. Пусть $P = (P_w, w \in \mathcal{Y})$ являет-

ся экспоненциальным семейством с канонической параметризацией; см. например [53, 54]. Соответствующий логарифм плотности может быть представлен как $\ell(y, w) = yw - d(w)$, где $d(w)$ – выпуклая функция. Популярными примерами являются биномиальная (логистическая) модель с $d(w) = \log(e^w + 1)$, пуассоновская модель с $d(w) = e^w$ и экспоненциальная модель с $d(w) = -\log(w)$. Заметим, что линейная гауссовская регрессия является частным случаем с $d(w) = w^2/2$.

Определение обобщенной линейной модели подразумевает, что каждое наблюдение Y_i имеет распределение из семейства P с параметром w_i , который линейно зависит от регрессора $\Psi_i \in \mathbb{R}^p$:

$$Y_i \sim P_{\Psi_i^\top \mathbf{v}^*}. \quad (4.12)$$

Соответствующий логарифм плотности распределения для обобщенной линейной модели записывается следующим образом:

$$\mathcal{L}(\mathbf{v}) = \sum_{i=1}^n \{Y_i \Psi_i^\top \mathbf{v} - d(\Psi_i^\top \mathbf{v})\}.$$

Будучи распределенным согласно распределению $P_{\mathbf{v}^*}$, каждое наблюдение Y_i следует модели (4.12), в частности $\mathbb{E}Y_i = d'(\Psi_i^\top \mathbf{v}^*)$. Однако, аналогично предыдущим разделам неверная спецификация модели (4.12) допустима. Неверная спецификация функции отклика означает, что $\mathbf{f} \stackrel{\text{def}}{=} \mathbb{E}\mathbf{Y}$ не может быть представлена в виде $d'(\Psi^\top \mathbf{v})$ ни для какого \mathbf{v} . Другой тип неверной спецификации имеет отношение к распределению данных. Модель (4.12) предполагает, что наблюдаемые значения Y_i независимы и их маргинальные распределения принадлежат параметрическому семейству P . В дальнейшем мы будем предполагать только выполнение некоторых условий на экспоненциальные моменты. Цель оценивания \mathbf{v}^* определяется следующим образом:

$$\mathbf{v}^* \stackrel{\text{def}}{=} \underset{\mathbf{v}}{\operatorname{argmax}} \mathbb{E}\mathcal{L}(\mathbf{v}).$$

Оценка максимума правдоподобия $\tilde{\mathbf{v}}$ определяется максимизацией $\mathcal{L}(\mathbf{v})$:

$$\tilde{\mathbf{v}} = \operatorname{argmax}_{\mathbf{v}} \mathcal{L}(\mathbf{v}) = \operatorname{argmax}_{\mathbf{v}} \sum_{i=1}^n \{Y_i \Psi_i^\top \mathbf{v} - d(\Psi_i^\top \mathbf{v})\}.$$

Выпуклость $d(\cdot)$ подразумевает, что $\mathcal{L}(\mathbf{v})$ является вогнутой функцией параметра \mathbf{v} , что гарантирует единственность и эффективный способ решения оптимизационной задачи. Однако, решение в явном виде доступно только для константной или линейной гауссовской регрессии. Соответствующая цель оценивания \mathbf{v}^* является максимизатором математического ожидания логарифма правдоподобия:

$$\mathbf{v}^* = \operatorname{argmax}_{\mathbf{v}} \mathbb{E} \mathcal{L}(\mathbf{v}) = \operatorname{argmax}_{\mathbf{v}} \sum_{i=1}^n \{f_i \Psi_i^\top \mathbf{v} - d(\Psi_i^\top \mathbf{v})\}$$

с $f_i = \mathbb{E} Y_i$. Функция $\mathbb{E} \mathcal{L}(\mathbf{v})$ также является вогнутой и вектор \mathbf{v}^* хорошо определен.

Определим индивидуальные ошибки (остатки) $\varepsilon_i = Y_i - \mathbb{E} Y_i$. Ниже мы предположим, что эти ошибки удовлетворяют некоторым условиям на экспоненциальные моменты.

(e₀) *Существуют константы ν_0 и $\mathbf{g}_1 > 0$, и для любого i константа \mathbf{s}_i такие, что $\mathbb{E}(\varepsilon_i/\mathbf{s}_i)^2 \leq 1$ и*

$$\log \mathbb{E} \exp(\mu \varepsilon_i/\mathbf{s}_i) \leq \nu_0^2 \mu^2/2, \quad \forall \mu: |\mu| \leq \mathbf{g}_1. \quad (4.13)$$

Естественным кандидатом для \mathbf{s}_i являются σ_i , где $\sigma_i^2 = \mathbb{E} \varepsilon_i^2$ является дисперсией ε_i . При условии выполнения условия (4.13) введем $(p \times p)$ -матрицу \mathcal{V} определенную следующим образом:

$$\mathcal{V}^2 \stackrel{\text{def}}{=} \sum_{i=1}^n \mathbf{s}_i^2 \Psi_i \Psi_i^\top. \quad (4.14)$$

Условие (e_0) эффективно означает, что каждый остаток $\varepsilon_i = Y_i - \mathbb{E} Y_i$ имеет ограниченные экспоненциальные моменты: для $|\lambda| \leq \mathbf{g}_1$ выполняется, что $f(\lambda) \stackrel{\text{def}}{=} \log \mathbb{E} \exp(\lambda \varepsilon_i/\mathbf{s}_i) < \infty$. Другими словами, условие (e_0) требует, чтобы

маргинальное распределение каждой из величин ε_i имело легкий (экспоненциально убывающий) хвост.

Определим также

$$N_1^{-1/2} \stackrel{\text{def}}{=} \max_i \sup_{\gamma \in \mathbb{R}^p} \frac{\mathfrak{s}_i |\Psi_i^\top \gamma|}{\|\mathcal{V}\gamma\|}. \quad (4.15)$$

Теперь покажем выполнимость необходимых для теоремы БФМ условий.

Лемма 9. *Предположим выполнение условия (e_0) и пусть матрица \mathcal{V}^2 определена выражением (4.14), а величина N_1 – выражением (4.15). Тогда условие (ED_0) следует из условия (e_0) с данной матрицей \mathcal{V}^2 и $\mathfrak{g} = \mathfrak{g}_1 N_1^{1/2}$. Более того, стохастическая компонента $\zeta(\mathbf{v})$ линейна по параметру \mathbf{v} и условие (ED_2) выполняется с $\omega \equiv 0$.*

Остается только ограничить ошибку квадратичной аппроксимации для математического ожидания процесса $L(\mathbf{v}, \mathbf{v}^*)$ в окрестности \mathbf{v}^* . Интересной особенностью обобщенных линейных моделей является тот факт, что эффект неверной спецификации модели исчезает при рассмотрении математического ожидания $L(\mathbf{v}, \mathbf{v}^*)$.

Лемма 10. *Выполняется*

$$-EL(\mathbf{v}, \mathbf{v}^*) = \sum_{i=1}^n \{d(\Psi_i^\top \mathbf{v}) - d(\Psi_i^\top \mathbf{v}^*) - d'(\Psi_i^\top \mathbf{v}^*) \Psi_i^\top (\mathbf{v} - \mathbf{v}^*)\} = \mathcal{K}(\mathbb{P}_{\mathbf{v}^*}, \mathbb{P}_{\mathbf{v}}),$$

где $\mathcal{K}(\mathbb{P}_{\mathbf{v}^*}, \mathbb{P}_{\mathbf{v}})$ – расстояние Кульбака-Лейблера между мерами $\mathbb{P}_{\mathbf{v}^*}$ и $\mathbb{P}_{\mathbf{v}}$.

Более того,

$$-EL(\mathbf{v}, \mathbf{v}^*) = \|\mathcal{D}(\mathbf{v}^\circ)(\mathbf{v} - \mathbf{v}^*)\|^2/2, \quad (4.16)$$

где $\mathbf{v}^\circ \in [\mathbf{v}^*, \mathbf{v}]$ и

$$\mathcal{D}^2(\mathbf{v}^\circ) = \sum_{i=1}^n d''(\Psi_i^\top \mathbf{v}^\circ) \Psi_i \Psi_i^\top.$$

Определим теперь матрицу \mathcal{D}^2 следующим образом:

$$\mathcal{D}^2 \stackrel{\text{def}}{=} \mathcal{D}^2(\mathbf{v}^*) = \sum_{i=1}^n d''(\Psi_i^\top \mathbf{v}^*) \Psi_i \Psi_i^\top.$$

Пусть также \mathcal{V}^2 определяется выражением (4.14). Заметим, что матрицы \mathcal{D}^2 и \mathcal{V}^2 совпадают, если модель $Y_i \sim P_{\Psi_i^\top \mathbf{v}^*}$ правильно специфицирована и $\mathbf{s}_i^2 = d''(\Psi_i^\top \mathbf{v}^*)$. Матрица \mathcal{D}^2 описывает локальные эллиптические окрестности центральной точки \mathbf{v}^* вида $\mathcal{Y}_0(\mathbf{r}) = \{\mathbf{v} : \|\mathcal{D}(\mathbf{v} - \mathbf{v}^*)\| \leq \mathbf{r}\}$. Если матричная функция $\mathcal{D}^2(\mathbf{v})$ непрерывна в окрестности $\mathcal{Y}_0(\mathbf{r})$, то значение $\delta(\mathbf{r})$, которое является мерой качества аппроксимации $-\mathbb{E}L(\mathbf{v}, \mathbf{v}^*)$ квадратичной функцией $\|\mathcal{D}(\mathbf{v} - \mathbf{v}^*)\|^2/2$, мало и условие идентифицируемости (\mathcal{L}_0) выполняется на $\mathcal{Y}_0(\mathbf{r})$. Следующая лемма позволяет ограничить величину $\delta(\mathbf{r})$.

Лемма 11. Пусть $d''(z)$ удовлетворяет условию Липшица:

$$|d''(z) - d''(z_0)| \leq L|z - z_0|, \quad z, z_0 \in \mathbb{R}. \quad (4.17)$$

Тогда

$$\delta(\mathbf{r}) \leq L \frac{\mathbf{r}}{N_2^{1/2}},$$

где

$$N_2^{-1/2} \stackrel{\text{def}}{=} \max_i \sup_{\gamma \in \mathbb{R}^p} \frac{|\Psi_i^\top \gamma|}{d''(\Psi_i^\top \mathbf{v}^*) \cdot \|\mathcal{D}\gamma\|}. \quad (4.18)$$

Теперь мы готовы сформулировать результат БфМ для обобщенных линейных моделей.

Теорема 15. Пусть выполняется условие (e_0) и условия леммы 11. Тогда результаты теоремы 9 выполняются для обобщенных линейных моделей с

$$\diamond(\mathbf{r}_0, \mathbf{x}) \leq L \frac{\mathbf{r}_0^2}{N_2^{1/2}}.$$

Если функция $d(w)$ является квадратичной, то ошибка аппроксимации δ исчезает и квадратичная аппроксимация становится верной на всем простран-

стве, локализационный шаг не требуется. Однако, если $d(w)$ не является квадратичной, то результат применим только локально и должен быть дополнен оценкой больших уклонений. Структура обобщенных линейных помогает также и в зоне больших уклонений. Действительно, условие идентифицируемости $(\mathcal{L}\mathbf{r})$ легко следует из леммы 10: достаточно ограничить снизу матрицу $\mathcal{D}^2(\mathbf{v})$ для $\mathbf{v} \in \mathcal{Y}_0(\mathbf{r})$:

$$\mathcal{D}^2(\mathbf{v}) \geq \mathbf{b}\mathcal{D}^2, \quad \mathbf{v} \in \mathcal{Y}_0(\mathbf{r}).$$

Интересным вопросом, аналогично случаю независимых одинаково распределенных случайных величин, является минимальный радиус \mathbf{r}_0 локальной окрестности $\mathcal{Y}_0(\mathbf{r}_0)$, который гарантирует концентрацию на нем апостериорного распределения. Требуемые условия выполняются, если $\mathbf{r}^2 \geq \mathbf{r}_0^2 = \mathbf{C}(\mathbf{x} + p)$, где \mathbf{C} зависит только от ν_0 , \mathbf{b} и \mathbf{g} . Таким образом, результат теоремы БфМ верен, если величина

$$\delta(\mathbf{r}_0)\mathbf{r}_0 = \mathbf{C} \frac{\mathbf{r}_0^2}{N_1^{1/2}} = \mathbf{C} \frac{\mathbf{x} + p}{N_1^{1/2}}$$

является малой.

Обобщенные линейные модели также допускают семипараметрическое расширение; см. книгу Грина и Янделла [55]. Аналогично случаю частично линейной регрессии мы представляем параметр обобщенной линейной модели как сумму линейной и непараметрической части:

$$w_i = \Psi_i^\top \boldsymbol{\theta}^* + g^*(X_i),$$

где $g^*(\cdot)$ принадлежит, например, соболевскому классу. Данная ситуация отличается от раздела 4.3 только некоторыми техническими деталями и ведет к похожим теоретическим результатам.

4.5. Доказательства

В этом разделе собраны доказательства результатов главы 4.

4.5.1. Доказательство леммы 7

Выполняется неравенство

$$\begin{aligned}
& \sup_{\gamma \in \mathbb{R}^p} \log \mathbb{E} \exp \left\{ \mu \frac{\gamma^\top \nabla \zeta(\mathbf{v}^*)}{\|\mathcal{D}\gamma\|} \right\} \\
&= \sup_{\gamma \in \mathbb{R}^p} \log \mathbb{E} \exp \left\{ \mu \frac{\sum_{i=1}^n (g'(Y_i - \Psi_i^\top \mathbf{v}^*) - \mathbb{E} g'(Y_i - \Psi_i^\top \mathbf{v}^*)) \gamma^\top \Psi_i}{\|\mathcal{D}\gamma\|} \right\} \\
&= \sup_{\gamma \in \mathbb{R}^p} \log \mathbb{E} \exp \left\{ \frac{\mu k \sum_{i=1}^n \gamma^\top \Psi_i}{\|\mathcal{D}\gamma\|} g'(\varepsilon)/k \right\}.
\end{aligned}$$

По определению $k|\gamma^\top \Psi_i|/\|\mathcal{D}\gamma\| \leq N_1^{-1/2}$ и таким образом $\mu k|\gamma^\top \Psi_i|/\|\mathcal{D}\gamma\| \leq \mathbf{g}_1$.

Поэтому

$$\sup_{\gamma \in \mathbb{R}^p} \log \mathbb{E} \exp \left\{ \mu \frac{\gamma^\top \nabla \zeta(\mathbf{v}^*)}{\|\mathcal{D}\gamma\|} \right\} \leq \sup_{\gamma \in \mathbb{R}^p} \frac{\nu_0^2 \mu^2 k^2 \sum_{i=1}^n |\Psi_i^\top \gamma|^2}{2 \|\mathcal{D}\gamma\|^2} = \frac{\nu_0^2 \mu^2}{2}.$$

4.5.2. Доказательство леммы 8

По определению

$$\|\mathcal{D}^{-1}(\mathcal{D}^2(\mathbf{v}) - \mathcal{D}^2)\mathcal{D}^{-1}\| = \sup_{\gamma \in \mathbb{R}^p: \|\gamma\|=1} |\gamma^\top \mathcal{D}^{-1}(\mathcal{D}^2(\mathbf{v}) - \mathcal{D}^2)\mathcal{D}^{-1}\gamma|.$$

С учетом условия Липшица (4.5)

$$\begin{aligned}
\mathcal{D}^2(\mathbf{v}) - \mathcal{D}^2 &= \mathcal{D}^2(\mathbf{v}) - \mathcal{D}^2(\mathbf{v}^*) \\
&= \sum_{i=1}^n \Psi_i \Psi_i^\top \int \left[g''(z - \Psi_i^\top (\mathbf{v} - \mathbf{v}^*)) - g''(z) \right] f(z) dz \\
&= \sum_{i=1}^n \Psi_i \Psi_i^\top L |\Psi_i^\top (\mathbf{v} - \mathbf{v}^*)| \int f(z) dz = L \sum_{i=1}^n \Psi_i \Psi_i^\top |\Psi_i^\top (\mathbf{v} - \mathbf{v}^*)|.
\end{aligned}$$

Тогда с учетом определения (4.4) получим цепочку неравенств, которая доказывает требуемый результат:

$$\begin{aligned}
& \|\mathcal{D}^{-1}(\mathcal{D}^2(\mathbf{v}) - \mathcal{D}^2)\mathcal{D}^{-1}\| \\
& \leq \sup_{\gamma \in \mathbb{R}^p: \|\gamma\|=1} \sum_{i=1}^n L |\Psi_i^\top (\mathbf{v} - \mathbf{v}^*)| \gamma^\top \mathcal{D}^{-1} \Psi_i \Psi_i^\top \mathcal{D}^{-1} \gamma \\
& \leq L N_1^{-1/2} \|\mathcal{D}(\mathbf{v} - \mathbf{v}^*)\| \sup_{\gamma \in \mathbb{R}^p: \|\gamma\|=1} \gamma^\top \mathcal{D}^{-1} \left(\sum_{i=1}^n \Psi_i \Psi_i^\top \right) \mathcal{D}^{-1} \gamma \\
& = L \frac{\mathbf{r}}{\mathbf{k} N_1^{1/2}}.
\end{aligned}$$

4.5.3. Доказательство леммы 9

Градиент стохастической части $\zeta(\mathbf{v})$ правдоподобия $\mathcal{L}(\mathbf{v})$ не зависит от \mathbf{v} , а именно, $\nabla \zeta(\mathbf{v}) = \sum \Psi_i \varepsilon_i$ с $\varepsilon_i = Y_i - \mathbb{E} Y_i$. Для любого единичного вектора $\gamma \in \mathbb{R}^p$ и любого $\lambda \leq \mathbf{g}$, независимость остатков ε_i позволяет оценить

$$\log \mathbb{E} \exp \left\{ \frac{\lambda}{\|\mathcal{V}\gamma\|} \gamma^\top \sum_{i=1}^n \Psi_i \varepsilon_i \right\} = \sum_{i=1}^n \log \mathbb{E} \exp \left\{ \frac{\lambda \mathbf{s}_i \Psi_i^\top \gamma}{\|\mathcal{V}\gamma\|} \varepsilon_i / \mathbf{s}_i \right\}.$$

По определению $\mathbf{s}_i |\Psi_i^\top \gamma| / \|\mathcal{V}\gamma\| \leq N_1^{-1/2}$ и тогда, $\lambda \mathbf{s}_i |\Psi_i^\top \gamma| / \|\mathcal{V}\gamma\| \leq \mathbf{g}_1$. Таким образом (4.13) подразумевает

$$\log \mathbb{E} \exp \left\{ \frac{\lambda}{\|\mathcal{V}\gamma\|} \gamma^\top \sum_{i=1}^n \Psi_i \varepsilon_i \right\} \leq \frac{\nu_0^2 \lambda^2}{2 \|\mathcal{V}\gamma\|^2} \sum_{i=1}^n \mathbf{s}_i^2 |\Psi_i^\top \gamma|^2 = \frac{\nu_0^2 \lambda^2}{2},$$

и (ED_0) следует.

4.5.4. Доказательство леммы 10

По определению

$$\mathbb{E} L(\mathbf{v}, \mathbf{v}^*) = \sum_{i=1}^n \{ f_i \Psi_i^\top (\mathbf{v} - \mathbf{v}^*) - d(\Psi_i^\top \mathbf{v}) + d(\Psi_i^\top \mathbf{v}^*) \}.$$

По определению \mathbf{v}^* является точкой максимума для функции $\mathbb{E}L(\mathbf{v})$. Равенство $\nabla \mathbb{E}L(\mathbf{v}^*) = 0$ позволяет записать

$$\sum_{i=1}^n [f_i - d'(\Psi_i^\top \boldsymbol{\theta}^*)] \Psi_i = 0.$$

Теперь уравнение (4.16) получается с помощью разложения Тейлора до второго порядка в точке \mathbf{v}^* .

4.5.5. Доказательство леммы 11

По определению

$$\begin{aligned} \|\mathcal{D}^{-1}(\mathcal{D}^2(\mathbf{v}) - \mathcal{D}^2)\mathcal{D}^{-1}\| &= \sup_{\boldsymbol{\gamma} \in \mathbb{R}^p: \|\boldsymbol{\gamma}\|=1} |\boldsymbol{\gamma}^\top \mathcal{D}^{-1}(\mathcal{D}^2(\mathbf{v}) - \mathcal{D}^2)\mathcal{D}^{-1}\boldsymbol{\gamma}| \\ &= \sup_{\boldsymbol{\gamma} \in \mathbb{R}^p: \|\boldsymbol{\gamma}\|=1} \left| \sum_{i=1}^n (d''(\Psi_i^\top \mathbf{v}) - d''(\Psi_i^\top \mathbf{v}^*)) \boldsymbol{\gamma}^\top \mathcal{D}^{-1} \Psi_i \Psi_i^\top \mathcal{D}^{-1} \boldsymbol{\gamma} \right|. \end{aligned}$$

Тогда с использованием условия Липшица (4.17) и с учетом определения (4.18) получим цепочку неравенств, которая доказывает требуемый результат:

$$\begin{aligned} &\|\mathcal{D}^{-1}(\mathcal{D}^2(\mathbf{v}) - \mathcal{D}^2)\mathcal{D}^{-1}\| \\ &\leq \sup_{\boldsymbol{\gamma} \in \mathbb{R}^p: \|\boldsymbol{\gamma}\|=1} \sum_{i=1}^n |d''(\Psi_i^\top \mathbf{v}) - d''(\Psi_i^\top \mathbf{v}^*)| |\boldsymbol{\gamma}^\top \mathcal{D}^{-1} \Psi_i \Psi_i^\top \mathcal{D}^{-1} \boldsymbol{\gamma}| \\ &\leq \sup_{\boldsymbol{\gamma} \in \mathbb{R}^p: \|\boldsymbol{\gamma}\|=1} \sum_{i=1}^n L |\Psi_i^\top (\mathbf{v} - \mathbf{v}^*)| |\boldsymbol{\gamma}^\top \mathcal{D}^{-1} \Psi_i \Psi_i^\top \mathcal{D}^{-1} \boldsymbol{\gamma}| \\ &\leq LN_2^{-1/2} \|\mathcal{D}(\mathbf{v} - \mathbf{v}^*)\| \sup_{\boldsymbol{\gamma} \in \mathbb{R}^p: \|\boldsymbol{\gamma}\|=1} \boldsymbol{\gamma}^\top \mathcal{D}^{-1} \left(\sum_{i=1}^n d''(\Psi_i^\top \mathbf{v}^*) \Psi_i \Psi_i^\top \right) \mathcal{D}^{-1} \boldsymbol{\gamma} \\ &= L \frac{\mathbf{r}}{N_2^{1/2}}. \end{aligned}$$

Заключение

1. Вычислена ошибка аппроксимации апостериорного распределения гауссовским распределением для случая гладкой семипараметрической статистической модели со стохастической частью, удовлетворяющей условиям типа конечности экспоненциальных моментов, мешающего параметра конечной размерности и равномерного априорного распределения параметров.
2. Показано, что для модели независимых одинаково распределенных случайных величин, линейных и обобщенных линейных моделей полученная ошибка аппроксимации зависит от размерности задачи p , размерности целевого параметра q и размера выборки n как $\sqrt{p^2q/n}$, что позволяет определить критическую для выполнения теоремы БфМ размерность параметрического множества.
3. Показано, что если гауссовское распределение является достаточно плоским, то результат теоремы БфМ остается в силе как и в случае равномерного распределения.
4. Результаты обобщены на случай семипараметрических моделей с бесконечномерным мешающим параметром с помощью применения метода проекционных оценок.
5. Показана применимость общих теоретических результатов к линейным и обобщенным линейным моделям с мешающим параметром, принадлежащим соболевскому классу гладкости.

Список литературы

1. Le Cam L., Yang G. L. Asymptotics in Statistics: Some Basic Concepts. Springer in Statistics, 1990.
2. van der Vaart A. W. Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics). Cambridge University Press, 2000. ISBN: 0521784506.
3. Buhlmann P., van de Geer S. Statistics for High-Dimensional Data: Methods, Theory and Applications. 1st edition. Springer Publishing Company, Incorporated, 2011. ISBN: 3642201911, 9783642201912.
4. Cox D. D. An analysis of Bayesian inference for nonparametric regression. // The Annals of Statistics. 1993. Vol. 21, no 2. P. 903–923.
5. Freedman D. On the Bernstein-von Mises theorem with infinite-dimensional parameters. // The Annals of Statistics. 1999. Vol. 27, no 4. P. 1119–1140.
6. Boucheron S., Gassiat E. A Bernstein-von Mises theorem for discrete probability distributions // Electronic Journal of Statistics. 2009. Vol. 3. P. 114–148. URL <http://dx.doi.org/10.1214/08-EJS262>.
7. Ghosal S. Asymptotic normality of posterior distributions in high-dimensional linear models // Bernoulli. 1999. Vol. 5, no. 2. P. 315–331. URL <http://dx.doi.org/10.2307/3318438>.
8. Schwartz L. On Bayes Procedures // Probability Theory and Related Fields. 1965. Vol. 4, no. 1. P. 10–26.
9. Barron A., Schervish M. J., Wasserman L. The Consistency of Posterior Distributions in Nonparametric Problems // The Annals of Statistics. 1996. Vol. 27. P. 536–561.
10. Bochkina N. Consistency of the posterior distribution in generalized linear inverse problems // Inverse Problems. 2013. Vol. 29, no. 9. P. 095010. URL <http://stacks.iop.org/0266-5611/29/i=9/a=095010>.
11. Shen X. Asymptotic normality of semiparametric and nonparametric posterior distributions // Journal of American Statistical Association. 2002. Vol. 97(457). P. 222–235.
12. Kim Y., Lee J. A Bernstein - von Mises theorem in the nonparametric right-censoring model // The Annals of Statistics. 2004. Vol. 32(4). P. 1492–1512.
13. Kim Y. The Bernstein - von Mises theorem for the proportional hazard model // The Annals of Statistics. 2006. Vol. 34(4). P. 1678–1700.

14. Leahu H. On the Bernstein-von Mises phenomenon in the Gaussian white noise model // *Electronic Journal of Statistics*. 2011. Vol. 5. P. 373–404. URL <http://dx.doi.org/10.1214/11-EJS611>.
15. Castillo I., Nickl R. Nonparametric Bernstein–von Mises theorems in Gaussian white noise // *The Annals of Statistics*. 2013. Vol. 41, no. 4. P. 1999–2028. URL <http://dx.doi.org/10.1214/13-AOS1133>.
16. Cheng G., Kosorok M. R. General frequentist properties of the posterior profile distribution // *The Annals of Statistics*. 2008. — 08. Vol. 36, no. 4. P. 1819–1853. URL <http://dx.doi.org/10.1214/07-AOS536>.
17. Bickel P. J., Kleijn B. J. K. The semiparametric Bernstein-von Mises theorem // *The Annals of Statistics*. 2012. Vol. 40, no. 1. P. 206–237. URL <http://dx.doi.org/10.1214/11-AOS921>.
18. Castillo I. A semiparametric Bernstein - von Mises theorem for Gaussian process priors // *Probability Theory and Related Fields*. 2012. Vol. 152. P. 53–99. 10.1007/s00440-010-0316-5. URL <http://dx.doi.org/10.1007/s00440-010-0316-5>.
19. Rivoirard V., Rousseau J. Bernstein - von Mises theorem for linear functionals of the density // *The Annals of Statistics*. 2012. Vol. 40, no. 3. P. 1489–1523.
20. Castillo I., Rousseau J. A General Bernstein–von Mises Theorem in semiparametric models. Available at arXiv:1305.4482 [math.ST].
21. Rivoirard V., Rousseau J. Posterior Concentration Rates for Infinite Dimensional Exponential Families // *Bayesian Analysis*. 2012. Vol. 7, no. 2. P. 311–334. URL <http://dx.doi.org/10.1214/12-BA710>.
22. Belloni A., Chernozhukov V. Posterior inference in curved exponential families under increasing dimensions // *The Econometrics Journal*. 2014. Vol. 17, no. 2. P. S75–S100. URL <http://dx.doi.org/10.1111/ectj.12027>.
23. Ibragimov I., Khas'minskij R. Statistical estimation. Asymptotic theory. Translated from the Russian by Samuel Kotz. New York - Heidelberg -Berlin: Springer-Verlag, 1981.
24. Гусев С.И. Асимптотические разложения, связанные с некоторыми статистическими оценками в гладком случае. I. Разложения случайных величин // *Теория вероятностей и ее применения*. 1975. Vol. 20, no. 3. P. 488–514.
25. Гусев С.И. Асимптотические разложения, связанные с некоторыми

- статистическими оценками в гладком случае. II. Разложения моментов и распределений // Теория вероятностей и ее применения. 1976. Vol. 21, no. 1. P. 16–33.
26. Бурнашев М.В. Исследование свойств второго порядка статистических оценок в схеме независимых наблюдений // Изв. АН СССР. Сер. матем. 1981. Vol. 45, no. 3. P. 509–539.
 27. Boucheron S., Massart P. A high-dimensional Wilks phenomenon // Probability Theory and Related Fields. 2011. Vol. 150. P. 405–433. 10.1007/s00440-010-0278-7. URL <http://dx.doi.org/10.1007/s00440-010-0278-7>.
 28. Kleijn B. J. K., van der Vaart A. W. Misspecification in infinite-dimensional Bayesian statistics // The Annals of Statistics. 2006. Vol. 34, no. 2. P. 837–877. URL <http://dx.doi.org/10.1214/009053606000000029>.
 29. Kosorok M. R. Introduction to empirical processes and semiparametric inference. Springer Series in Statistics. New York, NY., 2008.
 30. Панов М.Е., Спокойный В.Г. Критическая размерность в семипараметрической теореме Бернштейна - фон Мизеса // Труды Математического Института им. В.А. Стеклова. 2014. Т. 287. С. 242–266.
 31. Panov M., Spokoiny V. Finite Sample Bernstein – von Mises Theorem for Semiparametric Problems // Bayesian Analysis. 2015. Vol. 10, no. 3. P. 665–710. URL <http://projecteuclid.org/euclid.ba/1422884986>.
 32. Панов М.Е. Неасимптотический подход к оцениванию в байесовских семипараметрических задачах // Доклады академии наук. 2015.
 33. Spokoiny V. Bernstein - von Mises Theorem for growing parameter dimension. Manuscript. arXiv:1302.3430.
 34. Ermakov M. On semiparametric statistical inferences in the moderate deviation zone // Journal of Mathematical Sciences. 2008. Vol. 152, no 6. P. 869–874. URL <http://dx.doi.org/10.1007/s10958-008-9104-5>.
 35. Chernozhukov V., Hong H. An MCMC approach to classical estimation // Journal of Econometrics. 2003. — Aug. Vol. 115, no. 2. P. 293–346. URL <http://www.sciencedirect.com/science/article/B6VC0-48F5NC7-1/2/1f7db9a39a87bad9b87709ae57ae6535>.
 36. Kleijn B. J. K., van der Vaart A. W. The Bernstein-von-Mises theorem under misspecification // Electronic Journal of Statistics. 2012. Vol. 6. P. 354–381.
 37. Bochkina N., Green P. J. The Bernstein-von Mises theorem and non-regular mod-

- els // The Annals of Statistics. 2014. Vol. 42, no. 5. P. 1850–1878. arXiv:math.ST/1211.3434.
38. Spokoiny V. Parametric estimation. Finite sample theory // The Annals of Statistics. 2012. Vol. 40, no. 6. P. 2877–2909.
 39. Spokoiny V. Wilks Theorem for penalized maximum likelihood estimators. Manuscript. arXiv:1205.0498.
 40. Spokoiny V., Zhilova M. Sharp deviation bounds for quadratic forms // Mathematical Methods of Statistics. 2013. Vol. 22, no. 2. P. 100–113. arXiv:1302.1699; doi:10.3103/S1066530713020026.
 41. Боровков А.А., Могульский А.А. Большие отклонения и проверка статистических гипотез. Наука. Сибирское отделение, 1992.
 42. Bontemps D. Bernstein - von Mises theorem for Gaussian regression with increasing number of regressors. // The Annals of Statistics. 2011. Vol. 39, No. 5. P. 2557–2584.
 43. Johnstone I. M. High dimensional Bernstein–von Mises: simple examples // Borrowing strength: theory powering applications—a Festschrift for Lawrence D. Brown. Beachwood, OH: Institute of Mathematical Statistics, 2010. Vol. 6 of Institute of Mathematical Statistics Collections. P. 87–98.
 44. Laurent B., Massart P. Adaptive estimation of a quadratic functional by model selection // Ann. Statist. 2000. — 10. Vol. 28, no. 5. P. 1302–1338. URL <http://dx.doi.org/10.1214/aos/1015957395>.
 45. Pinsker M. Information and Information Stability of Random Variables and Processes // The American Mathematical Monthly. 1966. no. 73.
 46. Csiszar I. Information-type measures of difference of probability distributions and indirect observations // Studia Sci. Math. Hungar. 1967. Vol. 2. P. 299–318.
 47. Ghosal S. Asymptotic normality of posterior distributions for exponential families when the number of parameters tends to infinity // Journal of Multivariate Analysis. 2000. Vol. 74, no. 1. P. 49–68. URL <http://dx.doi.org/10.1006/jmva.1999.1874>.
 48. Ghosal S. Normal approximation to the posterior distribution for generalized linear models with many covariates // Mathematical Methods of Statistics. 1997. Vol. 6, no. 3. P. 332–348.
 49. Ченцов Н.Н. Оценка неизвестной плотности распределения по наблюдениям // ДАН СССР. 1962. Vol. 147, no. 1. P. 45–48.

50. Grenander U. Abstract inference. Wiley New York, 1981.
51. Chen X. Large sample sieve estimation of semi-nonparametric models // Handbook of econometrics. 2007. Vol. 6. P. 5549–5632.
52. Härdle W., Liang H. Partially linear models. Springer, 2007.
53. Nelder J., Baker R. Generalized linear models // Encyclopedia of Statistical Sciences. 1972.
54. McCullagh P., Nelder J. Generalized linear models. 2nd ed. Monographs on Statistics and Applied Probability. 37. London etc.: Chapman and Hall., 1989.
55. Green P. J., Yandell B. S. Semi-parametric generalized linear models. Springer, 1985.
56. Панов М.Е., Спокойный В.Г. О семипараметрическом оценивании в байесовской постановке // Труды 55-й научной конференции МФТИ. Управление и прикладная математика. 2012. Т. 1. С. 104–105.
57. Panov M., Spokoyny V. Critical dimension in semiparametric Bernstein - von Mises Theorem // Proceedings of "Information technologies and systems - 2013" 36th conference of young scientists and specialists of ИТП RAS. Kaliningrad, Russia: 2013. P. 386–391.
58. Гончаров Ф.О., Панов М.Е., Спокойный В.Г. Теорема Бернштейна–фон Мизеса в непараметрическом случае // Труды 57-й научной конференции МФТИ. Управление и прикладная математика. 2014. Т. 1. С. 102–103.
59. Панов М.Е. О концентрации целевого параметра в статистических моделях с растущей размерностью // Труды 57-й научной конференции МФТИ. Управление и прикладная математика. 2014. Т. 1. С. 112–113.