На правах рукописи

Панов Максим Евгеньевич

Неасимптотические свойства апостериорных распределений в семипараметрических задачах байесовского оценивания

01.01.05 – Теория вероятностей и математическая статистика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

кандидата физико-математических наук

Москва – 2015

Работа выполнена в *Московском физико-техническом институте (государственном университете)*

Научный руководитель: *кандидат физико-математических наук, профессор Гумбольдтского университета и Московского физико-технического института,*
Спокойный Владимир Григорьевич

Официальные оппоненты: *доктор физико-математических наук, профессор кафедры Статистического моделирования Математико-механического факультета Санкт-Петербургского государственного университета,*
Ермаков Михаил Сергеевич

PhD (Университет Бристоля), старший преподаватель Математической школы Университета Эдинбурга,
Бочкина Наталья Александровна

Ведущая организация: *Национальный исследовательский университет "Высшая школа экономики"*

Защита состоится «____» ____________ 2015 г. в ____ часов на заседании диссертационного совета *Д 002.077.03* при *федеральном государственном бюджетном учреждении науки Институте проблем передачи информации им. А.А. Харкевича РАН*, расположенном по адресу: *127051, г. Москва, Большой Каретный переулок, д.19 стр. 1.*

С диссертацией можно ознакомиться в библиотеке *Института проблем передачи информации им. А.А. Харкевича РАН.*

Автореферат разослан «____» ____________ 2015 г.

Отзывы и замечания по автореферату в двух экземплярах, заверенные печатью, просьба высылать по вышеуказанному адресу на имя ученого секретаря диссертационного совета.

Ученый секретарь
диссертационного совета,
д. ф.-м. н. *Соболевский А.Н.*

Общая характеристика работы

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. Знаменитая 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 данной 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, являет-

[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.

ся 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 Шеня[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

[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].

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 теорема БфМ для достаточно широкого класса параметрических и семипараметрических моделей. Важной особенностью нашего исследования является предположение о фиксированном размере выборки. 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[23]). В данной работе рассматривается семипараметрическая задача, в которой размерность полного параметра велика или бесконечна, а целевой параметр имеет небольшую 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[24]). 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 че-

[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] 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.

[24] 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.

рез эффективную информацию Фишера (см. работу Бикеля и Кляйна[25]). Обычно методы изучения в данном случае основываются на понятии наихудшей параметрической подмодели (см. обзор в книге Косорока[26]). Более того, предполагается, что существует метод оценивания мешающего параметра, достигающий определенной скорости сходимости оценки к истинному значению (см. работу Ченга и Косорока[27]). 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.

[25] 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.

[26] Kosorok M. R. Introduction to empirical processes and semiparametric inference. Springer Series in Statistics. New York, NY., 2008.

[27] 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.

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.

Общая методика исследования. Для решения поставленных задач в работе используются методы математической статистики, теории эмпирических процессов, теории вероятности, аппарат анализа Фурье.

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^3/n мала, где p – размерность задачи и n – размер выборки. Таким образом, размерность $p = O(n^{1/3})$ 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 статисти-

ческой 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 и размера выборки n как $\sqrt{p^3/n}$, 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 усечения базиса 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).

Также результаты работы обсуждались на семинарах Лаборатории структурных методов анализа данных в предсказательном моделировании МФТИ (2013-2015) и семинаре Международной лаборатории стохастического анализа и его приложений НИУ ВШЭ (2015).

Публикации. Основные результаты по теме диссертации изложены в 6 печатных работах, из которых 2 изданы в журналах, рекомендованных ВАК [1, 2].

Личный вклад автора. Содержание диссертации и основные положения, выносимые на защиту, отражают персональный вклад автора в опубликованные работы. Подготовка к публикации полученных результатов проводилась совместно с соавторами, причем вклад диссертанта был определяющим.

Структура и объем диссертации. Диссертация состоит из введения, обзора литературы, 4 глав, заключения и библиографии. Общий объем диссертации 93 страницы, включая 1 рисунок. Библиография включает 48 наименований.

Благодарности. Автор благодарен своему научному руководителю Владимиру Григорьевичу Спокойному за постановки задач, 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.

Содержание работы

Во Введении обоснована актуальность диссертационной работы, сформулирована цель и аргументирована научная новизна исследований, показана теоретическая и практическая значимость полученных результатов, представлены выносимые на защиту научные положения.

В первой главе рассматривается задача изучения свойств апостериорного распределения в случае, когда параметр модели имеет конечную размерность. Обозначим через $\boldsymbol{Y}$ наблюдаемые случайные данные и через $I\!\!P$ – их распределение. Параметрическая статистическая модель предполагает, что неизвестное распределение данных $I\!\!P$ принадлежит к заданному параметрическому семейству $(I\!\!P_{\boldsymbol{\upsilon}})$:

$$\boldsymbol{Y} \sim I\!\!P = I\!\!P_{\boldsymbol{\upsilon}^*} \in (I\!\!P_{\boldsymbol{\upsilon}}, \, \boldsymbol{\upsilon} \in \Upsilon),$$

где $\varUpsilon$ – это пространство параметров и $\boldsymbol{\upsilon}^* \in \varUpsilon$ – истинное значение параметра. В семипараметрическом случае целью оценивания является только низкоразмерная компонента $\boldsymbol{\theta}$ полного параметра $\boldsymbol{\upsilon}$. Таким образом целью оценивания является $\boldsymbol{\theta}^* \stackrel{\text{def}}{=} \varPi_0 \boldsymbol{\upsilon}^*$ для некоторого отображения $\varPi_0 : \varUpsilon \to I\!\!R^q$, где $q \in \mathbb{N}$ – размерность целевого параметра. Обычно в классическом семипараметрическом подходе вектор $\boldsymbol{\upsilon}$ представляется в виде $\boldsymbol{\upsilon} = (\boldsymbol{\theta}, \boldsymbol{\eta})$, где $\boldsymbol{\theta}$ является целью оценивания, а $\boldsymbol{\eta}$ является *мешающим параметром.*

Кроме того, в данной работе мы обращаемся к проблеме неверной спецификации модели. Это означает, что распределение $I\!\!P$ не принадлежит к рассматриваемому семейству $(I\!\!P_{\boldsymbol{\upsilon}}, \boldsymbol{\upsilon} \in \varUpsilon)$. "Истинное" значение $\boldsymbol{\upsilon}^*$ параметра $\boldsymbol{\upsilon}$ может быть определено как

$$\boldsymbol{\upsilon}^* = \operatorname*{argmax}_{\boldsymbol{\upsilon} \in \varUpsilon} I\!\!E \mathcal{L}(\boldsymbol{\upsilon}),$$

где $\mathcal{L}(\boldsymbol{\upsilon}) = \mathcal{L}(\boldsymbol{\upsilon} \,\big|\, \boldsymbol{Y}) = \log \frac{dI\!\!P_{\boldsymbol{\upsilon}}}{d\boldsymbol{\mu}_0}(\boldsymbol{Y})$ – логарифм правдоподобия семейства $(I\!\!P_{\boldsymbol{\upsilon}})$ для доминирующей меры $\boldsymbol{\mu}_0$. В случае неверной спецификации модели $\boldsymbol{\upsilon}^*$ определяет наилучшее параметрическое приближение к мере $I\!\!P$ в рассматриваемом семействе. Цель оценивания $\boldsymbol{\theta}^*$ по-прежнему определяется отображением $\varPi_0$:

$$\boldsymbol{\theta}^* \stackrel{\text{def}}{=} \varPi_0 \boldsymbol{\upsilon}^*.$$

В первом разделе вводятся необходимые для формулировки основных результатов определения, а также приводится ряд базовых неасимптотических статистических результатов, которые будут основой нашего анализа. Мы предполагаем, что большая положительная константа $\mathtt{x}$ фиксирована таким образом, чтобы задать множество случайных событий $\varOmega(\mathtt{x})$ *доминирующей вероятности*: $I\!\!P\bigl(\varOmega(\mathtt{x})\bigr) \geq 1 - \mathtt{C}\mathrm{e}^{-\mathtt{x}}$.

Одним из основных элементов нашей конструкции является $(p \times p)$-матрица $\mathcal{D}_0^2$, которая определяется аналогично информационной матрице Фишера:

$$\mathcal{D}_0^2 \stackrel{\text{def}}{=} -\nabla^2 I\!\!E \mathcal{L}(\boldsymbol{\upsilon}^*). \tag{1}$$

Также положим $\boldsymbol{\xi} \stackrel{\text{def}}{=} \mathcal{D}_0^{-1} \nabla \mathcal{L}(\boldsymbol{\upsilon}^*)$.

Для $(\boldsymbol{\theta}, \boldsymbol{\eta})$-модели рассмотрим блочное представление вектора $\nabla \mathcal{L}(\boldsymbol{\upsilon}^*)$ и матрицы $\mathcal{D}_0^2$ из (1):

$$\nabla \mathcal{L}(\boldsymbol{\upsilon}^*) = \begin{pmatrix} \nabla_{\boldsymbol{\theta}} \\ \nabla_{\boldsymbol{\eta}} \end{pmatrix}, \quad \mathcal{D}_0^2 = \begin{pmatrix} D_0^2 & A_0 \\ A_0^\top & H_0^2 \end{pmatrix}.$$

Определим также $(q \times q)$-матрицу $\breve{D}_0^2$ и случайный вектор $\breve{\boldsymbol{\xi}} \in I\!R^q$:

$$\breve{D}_0^2 \stackrel{\text{def}}{=} D_0^2 - A_0 H_0^{-2} A_0^\top, \tag{2}$$

$$\breve{\boldsymbol{\xi}} \stackrel{\text{def}}{=} \breve{D}_0^{-1} (\nabla_{\boldsymbol{\theta}} - A_0 H_0^{-2} \nabla_{\boldsymbol{\eta}}). \tag{3}$$

Матрица $\breve{D}_0^2$ размера $q \times q$ обычно называется *эффективной информационной матрицей Фишера*. Везде по ходу изложения $\|\boldsymbol{a}\|$ обозначает евклидову норму вектора $\boldsymbol{a}$, 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.

Сформулируем утверждение классической теоремы Бернштейна – фон Мизеса в семипараметрическом случае. Пусть на множестве параметров $\varUpsilon$ задано априорное распределение с положительной плотностью $\pi(\boldsymbol{\upsilon})$ по отношению к мере Лебега. В данной работе мы исследуем свойства апостериорного распределения для целевого параметра $\boldsymbol{\vartheta} = \varPi_0 \boldsymbol{\upsilon}$, которое может быть записано как

$$\boldsymbol{\vartheta} \mid \boldsymbol{Y} \propto \int \exp\bigl\{ \mathcal{L}(\boldsymbol{\upsilon}) \bigr\} \, \pi(\boldsymbol{\upsilon}) d\boldsymbol{\eta}.$$

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

$$\breve{D}_0 (\boldsymbol{\vartheta} - \widetilde{\boldsymbol{\theta}}) \mid \boldsymbol{Y} \xrightarrow{w} \mathcal{N}(0, I_q),$$

где I_q – единичная матрица размерности q, а $\breve{D}_0^2$ задается формулой (2).

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 и матрицу ковариаций.

Второй раздел посвящен формулировке основного результат гл. 1, а именно семипараметрической теоремы Бернштейна - фон Мизеса в случае конечной размерности мешающего параметра для фиксированного размера выборки.

Наш подход предполагает выполнение некоторого количества условий, которые можно разделить на локальные и глобальные. Локальные условия описывают поведение процесса $\mathcal{L}(\boldsymbol{\upsilon})$ в локальной области $\boldsymbol{\upsilon} \in \Upsilon_0(\mathtt{r}_0)$ при некотором фиксированном значении $\mathtt{r}_0$:

$$\Upsilon_0(\mathtt{r}_0) \stackrel{\text{def}}{=} \bigl\{\boldsymbol{\upsilon} \in \Upsilon\colon \|\mathcal{D}_0(\boldsymbol{\upsilon} - \boldsymbol{\upsilon}^*)\| \le \mathtt{r}_0\bigr\}.$$

Глобальные условия должны выполняться на всем Υ. Определим стохастическую компоненту логарифма правдоподобия $\zeta(\boldsymbol{\upsilon}) \stackrel{\text{def}}{=} \mathcal{L}(\boldsymbol{\upsilon}) - I\!\!E\mathcal{L}(\boldsymbol{\upsilon})$. Начнем с условий на конечность экспоненциальных моментов.

$(\boldsymbol{ED_0})$ Существует константа $\nu_0 > 0$, положительно определенная $(p{\times}p)$-матрица $\mathcal{V}_0^2$, удовлетворяющая $\operatorname{Var}\{\nabla\zeta(\boldsymbol{\upsilon}^*)\} \le \mathcal{V}_0^2$, и константа $\mathtt{g} > 0$ такие, что

$$\sup_{\boldsymbol{\gamma} \in I\!\!R^p} \log I\!\!E \exp\left\{\mu \frac{\langle \nabla\zeta(\boldsymbol{\upsilon}^*), \boldsymbol{\gamma}\rangle}{\|\mathcal{V}_0\boldsymbol{\gamma}\|}\right\} \le \frac{\nu_0^2\mu^2}{2},\ \forall \mu\colon |\mu| \le \mathtt{g}.$$

$(\boldsymbol{ED_2})$ Существуют константы $\nu_0, \omega > 0$ и для каждого $\mathtt{r} > 0$ константа $\mathtt{g}(\mathtt{r}) > 0$ такие, что для всех $\boldsymbol{\upsilon} \in \Upsilon_0(\mathtt{r})$:

$$\sup_{\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in I\!\!R^p} \log I\!\!E \exp\left\{\frac{\mu}{\omega} \frac{\boldsymbol{\gamma}_1^\top \nabla^2\zeta(\boldsymbol{\upsilon})\boldsymbol{\gamma}_2}{\|\mathcal{D}_0\boldsymbol{\gamma}_1\| \cdot \|\mathcal{D}_0\boldsymbol{\gamma}_2\|}\right\} \le \frac{\nu_0^2\mu^2}{2},\ \forall \mu\colon |\mu| \le \mathtt{g}(\mathtt{r}).$$

Определим $\mathcal{D}_0^2(\boldsymbol{\upsilon}) \stackrel{\text{def}}{=} -\nabla^2 I\!\!E\mathcal{L}(\boldsymbol{\upsilon})$. Тогда $\mathcal{D}_0^2 = \mathcal{D}_0^2(\boldsymbol{\upsilon}^*)$. Следующее условие необходимо, чтобы обеспечить гладкость математического ожидания логарифма правдоподобия $I\!\!E\mathcal{L}(\boldsymbol{\upsilon})$ в локальной области $\boldsymbol{\upsilon} \in \Upsilon_0(\mathtt{r}_0)$:

$(\boldsymbol{\mathcal{L}_0})$ Существует константа $\delta(\mathtt{r})$ такая, что на множестве $\Upsilon_0(\mathtt{r})$ для всех $\mathtt{r} \le \mathtt{r}_0$

$$\bigl\|\mathcal{D}_0^{-1}\mathcal{D}_0^2(\boldsymbol{\upsilon})\mathcal{D}_0^{-1} - I_p\bigr\| \le \delta(\mathtt{r}).$$

Обозначим $L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) = \mathcal{L}(\boldsymbol{\upsilon}) - \mathcal{L}(\boldsymbol{\upsilon}^*)$. Условие глобальной идентификации:

$(\boldsymbol{\mathcal{L}}\mathtt{r})$ Для любого $\mathtt{r} > 0$ существует константа $\mathtt{b}(\mathtt{r}) > 0$ такая, что $\mathtt{r}\mathtt{b}(\mathtt{r}) \to \infty$ при $\mathtt{r} \to \infty$ и

$$-I\!\!E L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) \ge \mathtt{r}^2\mathtt{b}(\mathtt{r}),\ \mathtt{r} = \|\mathcal{D}_0(\boldsymbol{\upsilon} - \boldsymbol{\upsilon}^*)\|.$$

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

$$\mathcal{D}_0^2 = \begin{pmatrix} D_0^2 & A_0 \\ A_0^\top & H_0^2 \end{pmatrix}, \qquad \mathcal{V}_0^2 = \begin{pmatrix} V_0^2 & B_0 \\ B_0^\top & Q_0^2 \end{pmatrix}.$$

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

$$\mathfrak{a}^2 D_0^2 \ge V_0^2, \qquad \mathfrak{a}^2 H_0^2 \ge Q_0^2, \qquad \mathfrak{a}^2 \mathcal{D}_0^2 \ge \mathcal{V}_0^2$$

и $\|D_0^{-1} A_0 H_0^{-2} A_0^\top D_0^{-1}\| \le \nu$.

Матрица $\breve{D}_0^2$ положительно определена при условии выполнения $(\mathcal{I})$.

Далее рассмотрим предложенную в работе Спокойного[28] оценку брэкетинга, которая описывает качество квадратичной аппроксимации логарифма правдоподобия $\mathcal{L}(\boldsymbol{\upsilon})$ в локальной окрестности точки $\boldsymbol{\upsilon}^*$. Определим квадратичный процесс $\mathbb{L}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*)$:

$$\mathbb{L}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) \stackrel{\text{def}}{=} (\boldsymbol{\upsilon} - \boldsymbol{\upsilon}^*)^\top \nabla \mathcal{L}(\boldsymbol{\upsilon}^*) - \|\mathcal{D}_0(\boldsymbol{\upsilon} - \boldsymbol{\upsilon}^*)\|^2/2.$$

Теорема 1 (Спокойный 2012)**.** *Пусть условия* (ED_0), (ED_2), $(\mathcal{L}_0)$, *и* $(\mathcal{I})$ *выполняются для некоторого* $\mathtt{r}_0 > 0$. *Тогда на множестве случайных событий* $\Omega(\mathtt{x})$ *доминирующей вероятности не менее* $1 - 4\mathrm{e}^{-\mathtt{x}}$

$$|L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) - \mathbb{L}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*)| \le \Delta(\mathtt{r}_0, \mathtt{x}), \quad \boldsymbol{\upsilon} \in \Upsilon_0(\mathtt{r}_0), \tag{4}$$

где

$$\Delta(\mathtt{r}_0, \mathtt{x}) \stackrel{\text{def}}{=} \bigl\{\delta(\mathtt{r}_0) + 6\nu_0\, z_{\mathbb{H}}(\mathtt{x})\, \omega\bigr\}\, \mathtt{r}_0^2 \tag{5}$$

и $z_{\mathbb{H}}(\mathtt{x}) \stackrel{\text{def}}{=} 2p^{1/2} + \sqrt{2\mathtt{x}} + \mathtt{g}^{-1}(\mathtt{g}^{-2}\mathtt{x} + 1)4p$.

Далее мы обращаемся к семипараметрической теореме Бернштейна – фон Мизеса в случае конечного размера выборки и конечной размерности мешающего параметра. Первым шагом нашего анализа является проверка того, что $\boldsymbol{\vartheta} \,\big|\, \boldsymbol{Y}$ концентрируется в малой окрестности $\Theta_0(\mathtt{r}_0) = \bigl\{\boldsymbol{\theta} \colon \|\breve{D}_0(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\| \le \mathtt{r}_0\bigr\}$ центральной точки $\boldsymbol{\theta}^* = \Pi_0 \boldsymbol{\upsilon}^*$ при правильном выборе $\mathtt{r}_0$. Предположим, что априорное распределение является равномерным, т.е. $\pi(\boldsymbol{\upsilon}) \equiv 1$, $\boldsymbol{\upsilon} \in \Upsilon$. Концентрационные свойства апостериорного распределения будут описываться с помощью случайной величины

$$\rho(\mathtt{r}_0) \stackrel{\text{def}}{=} \frac{\int_{\Upsilon} \exp\bigl\{L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*)\bigr\} \mathbb{1}\bigl\{\boldsymbol{\theta} \notin \Theta_0(\mathtt{r}_0)\bigr\} d\boldsymbol{\upsilon}}{\int_{\Upsilon} \exp\bigl\{L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*)\bigr\} \mathbb{1}\bigl\{\boldsymbol{\theta} \in \Theta_0(\mathtt{r}_0)\bigr\} d\boldsymbol{\upsilon}}.$$

Очевидно, что $I\!\!P\bigl(\boldsymbol{\vartheta} \notin \Theta_0(\mathtt{r}_0) \,\big|\, \boldsymbol{Y}\bigr) \le \rho(\mathtt{r}_0)$.

[28] Spokoiny V. Parametric estimation. Finite sample theory // The Annals of Statistics. 2012. Vol. 40, no. 6. P. 2877–2909.

Теорема 2. *Пусть выполнено неравенство* (4)*. Тогда при* $\mathtt{b}\mathtt{r}_0^2 \ge z^2(p, \mathtt{x} + \frac{p}{2}\log\frac{e}{b})$ *на множестве* $\Omega(\mathtt{x})$ *вероятности не менее* $1 - 4\mathrm{e}^{-\mathtt{x}}$

$$\rho(\mathtt{r}_0) \le \exp\{2\Delta(\mathtt{r}_0, \mathtt{x}) + 2\mathrm{e}^{-\mathtt{x}} - \mathtt{x}\}.$$

Заметим, что похожие результаты для больших уклонений в частном случае независимых наблюдений можно найти в книге Боровкова и Могульского[29].

Далее перейдем к верхней оценке гауссовской аппроксимации апостериорного распределения. Удобно ввести понятие *локального условного математического ожидания* $I\!\!E^{\circ}$: для случайной величины η определим

$$I\!\!E^{\circ}\eta \stackrel{\mathrm{def}}{=} I\!\!E\Big[\eta 1\!\!1\big\{\boldsymbol{\vartheta} \in \Theta_0(\mathtt{r}_0)\big\} \,\Big|\, \boldsymbol{Y}\Big].$$

Введем обозначение $\boldsymbol{\theta}^{\circ} \stackrel{\mathrm{def}}{=} \boldsymbol{\theta}^* + \breve{D}_0^{-1}\breve{\boldsymbol{\xi}}$, где матрица $\breve{D}_0$ определена в (2) и вектор $\breve{\boldsymbol{\xi}}$ определен в (3). Вектор $\boldsymbol{\theta}^{\circ}$ может рассматриваться как аппроксимация оценки максимума правдоподобия до первого члена разложения в ряд Тейлора. Для случайного события, состоящего в том, что $\eta \in A \subseteq \mathbb{R}^q$, определим

$$I\!\!P^{\circ}(\eta \in A) \stackrel{\mathrm{def}}{=} I\!\!E^{\circ} 1\!\!1\{\eta \in A\}.$$

Теорема 3. *Для любого* $\boldsymbol{\lambda} \in I\!\!R^q$ *на* $\Omega(\mathtt{x})$ *справедлива оценка*

$$I\!\!E^{\circ}\big|\boldsymbol{\lambda}^{\top}\breve{D}_0(\boldsymbol{\vartheta} - \boldsymbol{\theta}^{\circ})\big|^2 \le \exp\big\{\Delta^{+}(\mathtt{r}_0, \mathtt{x})\big\}\|\boldsymbol{\lambda}\|^2.$$

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

$$I\!\!P^{\circ}\big(\breve{D}_0(\boldsymbol{\vartheta} - \boldsymbol{\theta}^{\circ}) \in A\big) \le \exp\big\{\Delta^{+}(\mathtt{r}_0, \mathtt{x})\big\} I\!\!P\big(\boldsymbol{\gamma} \in A\big).$$

Также на $\Omega(\mathtt{x})$ *верна оценка* $\Delta^{+}(\mathtt{r}_0, \mathtt{x}) \le 2\Delta(\mathtt{r}_0, \mathtt{x}) + 2\mathrm{e}^{-\mathtt{x}} + 2\exp\big\{\Delta(\mathtt{r}_0, \mathtt{x}) + 4\mathrm{e}^{-\mathtt{x}} - \mathtt{x}\big\}$.

Теперь укажем нижнюю границу для апостериорного математического ожидания.

Теорема 4. *Для любого* $\boldsymbol{\lambda} \in I\!\!R^q$ *на* $\Omega(\mathtt{x})$ *справедлива оценка*

$$I\!\!E^{\circ}\big|\boldsymbol{\lambda}^{\top}\breve{D}_0(\boldsymbol{\vartheta} - \boldsymbol{\theta}^{\circ})\big|^2 \ge \exp\big\{-\Delta^{-}(\mathtt{r}_0, \mathtt{x})\big\}\|\boldsymbol{\lambda}\|^2.$$

[29] Боровков А.А., Могульский А.А. Большие уклонения и проверка статистических гипотез. Наука. Сибирское отделение, 1992.

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

$$I\!P^{\circ}\bigl(\breve{D}_0(\boldsymbol{\vartheta} - \boldsymbol{\theta}^{\circ}) \in A\bigr) \ge \exp\bigl\{\Delta^{-}(\mathtt{r}_0, \mathtt{x})\bigr\} I\!P\bigl(\boldsymbol{\gamma} \in A\bigr) - \mathrm{e}^{-\mathtt{x}},$$

причем на $\Omega(\mathtt{x})$ *имеем* $\Delta^{-}(\mathtt{r}_0, \mathtt{x}) \le 2\Delta(\mathtt{r}_0, \mathtt{x}) + 4\mathrm{e}^{-\mathtt{x}} + 4\exp\bigl\{\Delta(\mathtt{r}_0, \mathtt{x}) + 4\mathrm{e}^{-\mathtt{x}} - \mathtt{x}\bigr\}$.

Результаты теорем 3 и 4 позволяют получить основной результат первой главы. Положим

$$\overline{\boldsymbol{\vartheta}} \stackrel{\mathrm{def}}{=} I\!E\bigl(\boldsymbol{\vartheta} \,\big|\, \boldsymbol{Y}\bigr), \qquad \mathfrak{S}^2 \stackrel{\mathrm{def}}{=} \operatorname{Cov}(\boldsymbol{\vartheta} \,|\, \boldsymbol{Y}) \stackrel{\mathrm{def}}{=} I\!E\bigl\{(\boldsymbol{\vartheta} - \overline{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta} - \overline{\boldsymbol{\vartheta}})^{\top} \,\big|\, \boldsymbol{Y}\bigr\}. \tag{6}$$

Ниже мы представляем версию теоремы БфМ в рассматриваемом неасимптотическом подходе, которая утверждает, что вектор $\overline{\boldsymbol{\vartheta}}$ близок к $\boldsymbol{\theta}^{\circ}$, $\mathfrak{S}^2$ примерно равна $\breve{D}_0^{-2}$ и распределение вектора $\breve{D}_0\bigl(\boldsymbol{\vartheta} - \boldsymbol{\theta}^{\circ}\bigr) \,\big|\, \boldsymbol{Y}$ близко к стандартному нормальному распределению.

Теорема 5. *Пусть выполнены условия данного раздела (см. выше). Пусть априорное распределение равномерно на* Υ . *Тогда существует случайное событие* $\Omega(\mathtt{x})$ *доминирующей вероятности не меньше* $1 - 4\mathrm{e}^{-\mathtt{x}}$ *такое, что на* $\Omega(\mathtt{x})$ *выполняются неравенства*

$$\begin{aligned} \|\breve{D}_0(\overline{\boldsymbol{\vartheta}} - \boldsymbol{\theta}^{\circ})\|^2 &\le 4\Delta(\mathtt{r}_0, \mathtt{x}) + 16\mathrm{e}^{-\mathtt{x}}, \\ \bigl\|I_q - \breve{D}_0 \mathfrak{S}^2 \breve{D}_0\bigr\| &\le 4\Delta(\mathtt{r}_0, \mathtt{x}) + 16\mathrm{e}^{-\mathtt{x}}, \end{aligned}$$

где $\overline{\boldsymbol{\vartheta}}$ *и* $\mathfrak{S}^2$ *определены в* (6) *и* $\Delta(\mathtt{r}_0, \mathtt{x})$ *определено в* (5).

Кроме того, на $\Omega(\mathtt{x})$ *для любого измеримого множества* $A \subset I\!R^q$

$$\begin{aligned} &\exp\bigl(-2\Delta(\mathtt{r}_0, \mathtt{x}) - 8\mathrm{e}^{-\mathtt{x}}\bigr) I\!P\bigl(\boldsymbol{\gamma} \in A\bigr) - \mathrm{e}^{-\mathtt{x}} \\ &\le I\!P\bigl(\breve{D}_0(\boldsymbol{\vartheta} - \boldsymbol{\theta}^{\circ}) \in A \,\big|\, \boldsymbol{Y}\bigr) \\ &\qquad \le \exp\bigl(2\Delta(\mathtt{r}_0, \mathtt{x}) + 5\mathrm{e}^{-\mathtt{x}}\bigr) I\!P\bigl(\boldsymbol{\gamma} \in A\bigr), \end{aligned}$$

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

Условие “ $\Delta(\mathtt{r}_0, \mathtt{x})$ 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 выведены как следствия для многих классических моделей (см. обсуждение в гл. 2). Полученный результат можно расширить следующим образом.

Следствие 1. *В случае выполнения условий Теоремы 5 для любого измеримого множества* $A \subset I\!R^q$ *на множестве случайных событий* $\Omega(\mathtt{x})$ *доминирующей вероятности не менее* $1 - 4\mathrm{e}^{-\mathtt{x}}$ *выполняются неравенства*

$$\begin{aligned}
&\exp\bigl(-2\Delta(\mathtt{r}_0, \mathtt{x}) - 8\mathrm{e}^{-\mathtt{x}}\bigr)\bigl\{I\!P\bigl(\boldsymbol{\gamma} \in A\bigr) - \tau\bigr\} - \mathrm{e}^{-\mathtt{x}} \\
&\quad \le I\!P\bigl(\mathfrak{S}^{-1}(\boldsymbol{\vartheta} - \overline{\boldsymbol{\vartheta}}) \in A \bigm| \boldsymbol{Y}\bigr) \\
&\qquad\quad \le \exp\bigl(2\Delta(\mathtt{r}_0, \mathtt{x}) + 5\mathrm{e}^{-\mathtt{x}}\bigr)\bigl\{I\!P\bigl(\boldsymbol{\gamma} \in A\bigr) + \tau\bigr\},
\end{aligned}$$

где $\tau = \frac{1}{2}\Delta(\mathtt{r}_0, \mathtt{x})^2\bigl(q + (1 + \Delta(\mathtt{r}_0, \mathtt{x}))^2\bigr)$ *и* $\boldsymbol{\gamma} \in \mathbb{R}^q$ *имеет стандартное нормальное распределение.*

Это следствие является очень важным, так как в приложениях матрица $\breve{D}_0$ и вектор $\boldsymbol{\theta}^{\circ}$ неизвестны, в то время как матрица $\mathfrak{S}^{-1}$ и вектор $\overline{\boldsymbol{\vartheta}}$ могут быть оценены численно.

В третьем разделе результаты теоремы 5 для неинформативного априорного распределения распространяются на случай гауссовского распределения с плотностью $\pi(\boldsymbol{\upsilon}) \propto \exp\bigl\{-\|G\boldsymbol{\upsilon}\|^2/2\bigr\}$ для некоторой симметричной матрицы G^2. Неинформативное априорное распределение может рассматриваться как предельный случай гауссовского априорного распределения при $G \to 0$. Придадим данному факту точный количественный смысл.

Теорема 6. *Предположим, что условия теоремы 5 выполнены. Пусть также* $\mathcal{P} = \mathcal{N}(0, G^{-2})$ *является гауссовской априорной мерой на* $I\!R^p$ *такой, что*

$$\|\mathcal{D}_0^{-1} G^2 \mathcal{D}_0^{-1}\| \le \boldsymbol{\epsilon} \le 1/2, \quad \operatorname{tr}\bigl(\mathcal{D}_0^{-1} G^2 \mathcal{D}_0^{-1}\bigr)^2 \le \delta^2, \quad \|\mathcal{D}_G^{-1} G^2 \boldsymbol{\upsilon}^*\| \le \beta,$$

где δ *и* β *– заданные константы. Тогда на множестве* $\Omega(\mathtt{x})$ *вероятности* $1 - 5\mathrm{e}^{-\mathtt{x}}$ *выполняются неравенства*

$$\begin{aligned}
I\!P\bigl(\breve{D}_0(\boldsymbol{\vartheta} - \boldsymbol{\theta}^{\circ}) \in A \bigm| \boldsymbol{Y}\bigr) &\ge \exp\bigl(-2\Delta(\mathtt{r}_0, \mathtt{x}) - 8\mathrm{e}^{-\mathtt{x}}\bigr)\bigl\{I\!P\bigl(\boldsymbol{\gamma} \in A\bigr) - \tau\bigr\} - \mathrm{e}^{-\mathtt{x}}, \\
I\!P\bigl(\breve{D}_0(\boldsymbol{\vartheta} - \boldsymbol{\theta}^{\circ}) \in A \bigm| \boldsymbol{Y}\bigr) &\le \quad \exp\bigl(2\Delta(\mathtt{r}_0, \mathtt{x}) + 5\mathrm{e}^{-\mathtt{x}}\bigr)\bigl\{I\!P\bigl(\boldsymbol{\gamma} \in A\bigr) + \tau\bigr\} + \mathrm{e}^{-\mathtt{x}},
\end{aligned}$$

где

$$\tau \stackrel{\text{def}}{=} \frac{1}{2}\sqrt{(1 + \boldsymbol{\epsilon})\bigl(3\beta + \boldsymbol{\epsilon}\sqrt{6(p + \mathtt{x})}\bigr)^2 + \delta^2}.$$

Похожие условия и результаты могут быть найдены в литературе для более конкретных статистических моделей; см. работу Джонстона[30] и разд. 1 гл. 3 ниже для 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.

Во второй главе основные результаты применяются к модели независимых одинаково распределенных случайных величин и используются для определения критической размерности.

В первом разделе рассматривается модель, в которой случайные величины $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$ независимы и одинаково распределены согласно мере P. Мы также предположим выполнение условий на P и $(P_{\boldsymbol{\upsilon}})$, которые являются естественным аналогом условий из раздела 1 гл. 1, но для правдоподобия одного измерения. Отметим, что параметрическое предположение $P \in (P_{\boldsymbol{\upsilon}}, \boldsymbol{\upsilon} \in \varUpsilon)$ может быть неверно специфицировано, и рассмотрим асимптотический случай с $n \to \infty$, и одновременно $p = p_n \to \infty$. В этом случае верна следующая

Теорема 7. *Пусть выполнены условия из раздела 5.1 в работе Спокойного*[31]*. Пусть также* $p_n \to \infty$ *и* $p_n^3/n \to 0$*. Тогда результат теоремы 5 справедлив с* $\varDelta(\mathtt{r}_0, \mathtt{x}) = \mathtt{C}\sqrt{p_n^3/n}$*,* $\mathbb{D}_0^2 = n\mathbb{F}_{\boldsymbol{\upsilon}^*}$*, где* $\mathbb{F}_{\boldsymbol{\upsilon}^*}$ *– информационная матрица Фишера для распределения* $(P_{\boldsymbol{\upsilon}})$ *в точке* $\boldsymbol{\upsilon}^*$*.*

Похожие результаты об асимптотической нормальности апостериорного распределения могут быть найдены в литературе; см., например, работу Госала[32]. Однако в них сходимость доказывается либо при слишком строгих условиях (например, $p_n^4 \log(p_n)/n \to 0$), либо для очень специфичных моделей. Заметим, что теорема 7 требует, чтобы $p_n = o(n^{1/3})$.

Во втором разделе мы приводим пример простой пуассоновской модели, в которой при $p_n^3/n \geq \beta^2 > 0$ апостериорное распределение не концентрирует-

[30] 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.

[31] Spokoiny V. Parametric estimation. Finite sample theory // The Annals of Statistics. 2012. Vol. 40, no. 6. P. 2877–2909.

[32] 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.

ся вокруг оценки максимума правдоподобия. В дополнение к теоретическому результату была проведена серия экспериментов, которые полностью подтвердили возникновение фазового перехода при превышении уровня критической размерности.

В третьей главе описанные выше результаты применяются к случаю, когда мешающий параметр имеет бесконечную размерность. Более конкретно, мы рассмотрим $(\boldsymbol{\theta}, \boldsymbol{f})$-модель, в которой $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^q$ и $\boldsymbol{f} \in \mathcal{H}$ для некоторого гильбертова пространства $\mathcal{H}$. Предположим, что в $\mathcal{H}$ существует счетный базис $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots$. Тогда

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

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

Обозначим функцию правдоподобия для полной модели через $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{f})$. Также обозначим $\boldsymbol{\upsilon} = (\boldsymbol{\theta}, \boldsymbol{\phi})$ и $\mathcal{L}(\boldsymbol{\upsilon}) = \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}) = \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{f}(\boldsymbol{\phi}))$. “Истинные” значения полного и целевого параметра могут быть определены при помощи максимизации математического ожидания правдоподобия:

$$\boldsymbol{\upsilon}^* \stackrel{\text{def}}{=} \operatorname*{argmax}_{\boldsymbol{\upsilon}=(\boldsymbol{\theta},\boldsymbol{\phi})} I\!\!E \mathcal{L}(\boldsymbol{\upsilon}), \qquad \boldsymbol{\theta}^* \stackrel{\text{def}}{=} \Pi_0 \boldsymbol{\upsilon}^*.$$

Мы применим метод усечения базиса и рассмотрим случай применения конечномерного неинформативного априорного распределения для параметров $\boldsymbol{\theta}$ и $\boldsymbol{\phi}$. Главным вопросом исследования является то, как усечение базиса влияет на свойства апостериорного распределения.

Пусть $\boldsymbol{\eta} = \{\eta_j\}_{j=1}^{m}$ является проекцией мешающего параметра $\boldsymbol{\phi}$ на конечномерное подпространство первых m компонент мешающего параметра. Для удобства обозначений представим $\boldsymbol{\phi} = (\boldsymbol{\eta}, \boldsymbol{\varkappa})$. В таком случае аппроксимация усечением базиса соответствует ситуации с $\boldsymbol{\varkappa} \equiv 0$. Запишем “истинную” точку $\boldsymbol{\upsilon}^*$ в виде $\boldsymbol{\upsilon}^* = (\boldsymbol{\theta}^*, \boldsymbol{\eta}^*, \boldsymbol{\varkappa}^*)$. Аппроксимация мешающего параметра $\boldsymbol{\phi}$ с помощью m-мерного параметра $\boldsymbol{\eta}$ приводит к двум источникам смещения. Первый из них связан с тем, что “усеченный” целевой параметр $\boldsymbol{\theta}_m^*$, определенный как

$$\boldsymbol{\theta}_m^* \stackrel{\text{def}}{=} \operatorname*{argmax}_{\boldsymbol{\theta}} \max_{\boldsymbol{\eta}} I\!\!E L(\boldsymbol{\theta}, \boldsymbol{\eta}, 0)$$

может быть отличен от истинного значения $\boldsymbol{\theta}^*$. Другой источник смещения связан с заменой эффективной информационной матрицы Фишера $\breve{D}^2$ ее аналогом для случая усечения базиса $\breve{D}_m^2$. Величины смещения могут быть оценены при предположениях гладкости на модель и на функциональный мешающий параметр $\boldsymbol{f}$ с использованием стандартных методов теории аппроксимации.

Обозначения упрощаются, если мы также предположим, что базис $\boldsymbol{e}_m$ в пространстве $\mathcal{H}$ выбран таким образом, чтобы обеспечить ортогональность блока H^2 информационной матрицы Фишера, т.е. $H^2 = \boldsymbol{I_\phi}$. Очевидным образом такая же структура сохраняется и для ее блоков $H_{\boldsymbol{\eta}}^2$ и $H_{\boldsymbol{\varkappa}}^2$. Таким образом, полная информационная матрица Фишера может быть представлена в виде

$$\mathcal{D}_0^2 = \begin{pmatrix} D^2 & A_m & C_m \\ A_m^\top & \boldsymbol{I_\eta} & 0 \\ C_m^\top & 0 & I_\varkappa \end{pmatrix}.$$

В модели с усеченным базисом $\boldsymbol{\upsilon}_m = (\boldsymbol{\theta}, \boldsymbol{\eta}, 0)$, рассмотрим матрицу Фишера

$$\mathcal{D}_m^2 \stackrel{\text{def}}{=} \begin{pmatrix} D^2 & A_m \\ A_m & \boldsymbol{I_\eta} \end{pmatrix}.$$

Также обозначим $\mathcal{D}_m^2(\boldsymbol{\upsilon}_m) = -\nabla_m^2 \mathbb{E}\mathcal{L}(\boldsymbol{\upsilon}_m)$, где ∇_m обозначает проекцию градиента на подпространство переменных $(\boldsymbol{\theta}, \boldsymbol{\eta})$.

Мы начнем с формулировки необходимых условий. Первое условие гарантирует семипараметрическую идентифицируемость и позволяет отделить целевой и мешающий параметры:

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

$$\|D^{-1} A_m A_m^\top D^{-1}\| \le \nu.$$

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

$(\boldsymbol{\mathcal{L}_m})$ Для любого $\mathtt{r} \le \mathtt{r}_0$ существует константа $\delta(\mathtt{r}) > 0$ такая, что на множестве $\Upsilon_0(\mathtt{r})$ выполняется неравенство:

$$\left\| \mathcal{D}_m^{-1} \mathcal{D}_m^2(\boldsymbol{\upsilon}_m) \mathcal{D}_m^{-1} - I \right\| \le \delta(\mathtt{r}).$$

Условия гладкости для параметров $\boldsymbol{\theta}$ и $\boldsymbol{\phi}$ выражаются через компоненту $\boldsymbol{\varkappa}^*$ полного параметра $\boldsymbol{\upsilon}^*$ и блок C_m матрицы $\mathcal{D}_0^2$.

$(\boldsymbol{B})$ Существуют ρ_m, $b_m \le 1/2$ такие, что выполняются неравенства

$$\|D^{-1} C_m \varkappa^*\| \le \rho_m, \qquad \|D^{-1} C_m C_m^\top D^{-1}\| \le b_m \le 1/2.$$

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 $\boldsymbol{f}$. Пример вычисления величин ρ_m и b_m приводится в разд. 3 гл. 3 ниже.

Рассмотрим неинформативное априорное распределение, определенное на пространстве параметров $(\boldsymbol{\theta}, \boldsymbol{\eta})$, задающее равномерную плотность для параметров усеченного базиса и задающее сингулярную массу в точке 0 для остальных компонент мешающего параметра $\varkappa$. Предположим, что условия теоремы 5 и следствия 1 выполнены для данного априорного распределения. Определим эффективную информационную матрицу Фишера $\breve{D}_m^2$ и вектор $\boldsymbol{\theta}_m^\circ$ как

$$\breve{D}_m^2 \stackrel{\text{def}}{=} D^2 - A_m A_m^\top, \ \boldsymbol{\theta}_m^\circ \stackrel{\text{def}}{=} \boldsymbol{\theta}_m^* + \breve{D}_m^{-1} \breve{\boldsymbol{\xi}}_m,$$

где $\breve{\boldsymbol{\xi}}_m \stackrel{\text{def}}{=} \nabla_{\boldsymbol{\theta}} - A_m \nabla_{\boldsymbol{\eta}}$.

Теорема 5 гарантирует результат БфМ для неинформативного априорного распределения на пространстве параметров усеченного базиса $\boldsymbol{\theta}$ и $\boldsymbol{\eta}$: апостериорное распределение $\boldsymbol{\theta}$ аппроксимируется гауссовским распределением $\mathcal{N}(\boldsymbol{\theta}_m^\circ, \breve{D}_m^{-2})$. Основной вопрос состоит в том, вносит ли усечение базиса значительный сдвиг в апостериорное распределение. Для полной семипараметрической модели определим

$$\breve{D}^2 \stackrel{\text{def}}{=} D^2 - A_m A_m^\top - C_m C_m^\top, \ \boldsymbol{\theta}^\circ \stackrel{\text{def}}{=} \boldsymbol{\theta}^* + \breve{D}^{-1} \breve{\boldsymbol{\xi}}.$$

Теорема 8. *Предположим выполнение условий $(\mathcal{I}^m)$, $(\mathcal{L}_m)$ и условия гладкости (B). Тогда эффективные информационные матрицы Фишера $\breve{D}^2$ и $\breve{D}_m^2$ в полной и усеченной моделях удовлетворяют неравенствам*

$$\begin{gathered}\bigl\|\breve{D}_m^{-1} \breve{D}^2 \breve{D}_m^{-1} - I_q\bigr\| \le (1-\nu)^{-1} \rho_m, \\ \operatorname{tr}\bigl\{(\breve{D}_m^{-1} \breve{D}^2 \breve{D}_m^{-1} - I_q)^2\bigr\} \le (1-\nu)^{-2} q\, \rho_m^2.\end{gathered}$$

Целевой параметр $\boldsymbol{\theta}^*$ и его аналог в усеченной модели $\boldsymbol{\theta}_m^*$ связаны соотношением

$$\bigl\|\breve{D}(\boldsymbol{\theta}^* - \boldsymbol{\theta}_m^*)\bigr\| \le (1-\nu)^{-1/2} \bigl\{\rho_m + 2\delta(\mathtt{r}_m)\mathtt{r}_m\bigr\} + 3\,\delta(3\mathtt{r}_m)\mathtt{r}_m\,,$$

где $\mathtt{r}_m = \|\varkappa^*\|$. *Более того, выполняется неравенство*

$$\begin{aligned}\bigl\|\breve{D}(\boldsymbol{\theta}^\circ - \boldsymbol{\theta}^\circ_m)\bigr\| &\le (1-\nu)^{-1}\rho_m\bigl(\|\breve{\boldsymbol{\xi}}_m\| + \|\boldsymbol{\xi}_\varkappa\|\bigr) \\ &\quad + (1-\nu)^{-1/2}\bigl\{\rho_m + 2\delta(\mathtt{r}_m)\mathtt{r}_m\bigr\} + 3\,\delta(3\mathtt{r}_m)\mathtt{r}_m\,,\end{aligned}$$

где $\boldsymbol{\xi}_\varkappa \stackrel{\mathrm{def}}{=} D^{-1}C_m\nabla_\varkappa$. *Наконец, при выполнении условий* (ED_0) *и* $(\mathcal{I})$ *на множестве* $\varOmega(\mathtt{x})$ *вероятности* $I\!\!P\bigl(\varOmega(\mathtt{x})\bigr) \ge 1 - 4\mathrm{e}^{-\mathtt{x}}$ *выполняется неравенство*

$$\begin{aligned}\bigl\|\breve{D}(\boldsymbol{\theta}^\circ - \boldsymbol{\theta}^\circ_m)\bigr\| &\le 2\mathfrak{a}(1-\nu)^{-1}\rho_m\bigl(q^{1/2} + 2\mathtt{x}\bigr) \\ &\quad + (1-\nu)^{-1/2}\bigl\{\rho_m + 2\delta(\mathtt{r}_m)\mathtt{r}_m\bigr\} + 3\,\delta(3\mathtt{r}_m)\mathtt{r}_m\,.\end{aligned}$$

Мы заключаем, что усеченное априорное распределение работает правильным образом, если величины $q^{1/2}\rho_m$ и b_m малы, а величина $\mathtt{r}_m = \|\varkappa^*\|$ не слишком большая.

В четвертой главе рассматривается ряд примеров, иллюстрирующих общие результаты предыдущих глав.

В первом разделе рассматривается модель линейной гауссовской регрессии и плоские гауссовские априорные распределения. Пусть $\boldsymbol{Y} = (Y_1,\ldots,Y_n)^\top$ – случайный вектор в $\mathbb{R}^n$, который удовлетворяет уравнению

$$\boldsymbol{Y} = \boldsymbol{f} + \boldsymbol{\varepsilon} = \Psi^\top \boldsymbol{\upsilon}^* + \boldsymbol{\varepsilon}, \qquad (7)$$

где ошибки $\boldsymbol{\varepsilon} = (\varepsilon_1,\ldots,\varepsilon_n)^\top$ – независимые с нулевым средним, а матрица плана Ψ размера $p\times n$ задана. Вектор средних значений $\boldsymbol{f}\in\mathbb{R}^n$ неизвестен и второе уравнение из (7) означает, что он принадлежит некоторому заданному p-мерному линейному подпространству в $\mathbb{R}^n$: $\boldsymbol{f} = \Psi^\top\boldsymbol{\upsilon}^*$ для неизвестного целевого вектора $\boldsymbol{\upsilon}^*\in\mathbb{R}^p$. Запишем матрицу Ψ в виде $\Psi = \{\Psi_1,\ldots,\Psi_n\}$, таким образом $f_i = \Psi_i^\top\boldsymbol{\upsilon}^*$. 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. $\varepsilon_i \in \mathcal{N}(0,\sigma_n^2 I_n)$, $i = 1,\ldots,n$, 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 $\mathcal{D}_0^2 = \sigma_n^{-2} I_p$.

Для гауссовского априорного распределения апостериорное распределение является в точности гауссовским. При использовании неинформативного априорного распределения единственным условием, необходимым для выполнения результата теоремы БфМ, является $p = p_n = o(n)$; см. работу Бонтама[33]. В работе Джонстона[34] 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 $(\sigma_n/\tau_n^2)\|\boldsymbol{\upsilon}^*\|$. Наши общие результаты из раздела 6 покрывают случай гауссовского априорного распределения при аналогичных условиях, т.е., как легко видеть, являются настолько же точными, как и результаты, полученные для очень специального гауссовского случая.

Во втором разделе рассматривается модель линейной негауссовской регрессии. В этой, более общей, ситуации ошибки ε_i в (7) распределены согласно распределению общего вида с плотностью $f(\cdot)\colon\ \varepsilon_i \sim P_f$. Обозначим $h(x) = \log f(x)$. Функция логарифма правдоподобия для данной задачи записывается следующим образом

$$\mathcal{L}(\boldsymbol{\upsilon}) = \sum_{i=1}^{n} \log f(Y_i - \Psi_i^\top \boldsymbol{\upsilon}) = \sum_{i=1}^{n} h(Y_i - \Psi_i^\top \boldsymbol{\upsilon}). \tag{8}$$

Предположим, что функция $h(z)$ дважды непрерывно дифференцируема и пусть $\mathrm{h}^2 \stackrel{\mathrm{def}}{=} \int h''(z) f(z) dz < \infty$. Если модель корректно специфицирована, то

$$\mathcal{D}_0^2 = \mathrm{h}^2 \sum_{i=1}^{n} \Psi_i \Psi_i^\top. \tag{9}$$

Для применения общих результатов из разд. 3 гл. 1 требуется проверка условий из разд. 1 гл. 1. Предположим выполнение некоторых условий типа конечности экспоненциальных моментов на распределение P_f:

$(\boldsymbol{e_0})$ *Существуют константы ν_0 и $\mathsf{g}_1 > 0$ такие, что для случайной величины $\varepsilon \sim P_f$ выполняется неравенство*

$$\log I\!\!E \exp\bigl(\mu h'(\varepsilon)/\mathrm{h}\bigr) \le \nu_0^2 \mu^2/2, \qquad \forall \mu\colon |\mu| \le \mathsf{g}_1.$$

Условие (e_0) означает, что распределение ошибок ε имеет экспоненциально убывающий хвост. При выполнении условия e_0 условие (ED_0) выполняется в соответствии со следующей леммой.

[33] 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.

[34] 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.

Лемма 1. *Предположим выполнение условия* (e_0) *и пусть* $\mathcal{V}_0^2 = \mathcal{D}_0^2 = \mathrm{h}^2 \sum_{i=1}^n \varPsi_i \varPsi_i^\top$. *Тогда условие* (ED_0) *следует из* (e_0) *с данной матрицей* $\mathcal{V}_0^2$ *и* $\mathtt{g} = \mathtt{g}_1 N_1^{1/2}$, *где*

$$N_1^{-1/2} \stackrel{\mathrm{def}}{=} \max_i \sup_{\gamma \in I\!\!R^p} \frac{\mathrm{h}|\varPsi_i^\top \boldsymbol{\gamma}|}{\|\mathcal{D}_0 \boldsymbol{\gamma}\|} . \tag{10}$$

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

Лемма 2. *Пусть* $|h''(z) - h''(z_0)| \le L|z - z_0|,\ z, z_0 \in \mathbb{R}$. *Тогда*

$$\delta(\mathtt{r}) \le \frac{L\,\mathtt{r}}{\mathrm{h}\, N_1^{1/2}},$$

где величина N_1 *определена уравнением* (10).

Следующий вопрос заключается в проверке условия (ED_2) , для которой необходимо наложить условие на маргинальное правдоподобие:

$(\boldsymbol{e_2})$ *Существует константа* ν_0 *и для любого* $\mathtt{r} > 0$ *существует* $\mathtt{g}(\mathtt{r}) > 0$ *такие, что для любого* δ *с* $|\delta| \le N_2^{-1/2}\mathtt{r}$ *выполняется*

$$\log I\!\!E \exp\biggl(\frac{\mu}{\mathfrak{s}_i} \bigl\{ h''(Y_i + \delta) - I\!\!E h''(Y_i + \delta) \bigr\} \biggr) \le \frac{\nu_0^2 \mu^2}{2}, \qquad \forall \mu\colon |\mu| \le \mathtt{g}(\mathtt{r}),$$

где $\mathfrak{s}_i$ – некоторые известные значения и

$$N_2^{-1/2} \stackrel{\mathrm{def}}{=} \max_i \sup_{\gamma \in I\!\!R^p} \frac{\mathfrak{s}_i |\varPsi_i^\top \boldsymbol{\gamma}|}{\|\mathcal{D}_0 \boldsymbol{\gamma}\|} .$$

Легко видеть, что достаточно положить $\omega = \frac{\sqrt{n}}{N_2}$, чтобы при выполнении условия (e_2) 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 $\boldsymbol{\upsilon} = (\boldsymbol{\theta}, \boldsymbol{\eta}) \in \mathbb{R}^p$. Предположим выполнение условия идентифицируемости $(\mathcal{I})$ из разд. 1 гл. 1. Тогда наши общие результаты позволяют получить семипараметрическую теорему БфМ для линейной модели (7).

Теорема 9. *Пусть выполняются условия* (e_0) , (e_2) , *условия леммы 2 и условие* $(\mathcal{I})$ *для матрицы* $\mathcal{D}_0^2$ *из* (9). *Тогда результаты теоремы 5 выполняются для линейной модели* (7) *с* $\mathtt{r}_0^2 \ge \mathtt{C}(p + \mathtt{x})$ *и*

$$\varDelta(\mathtt{r}_0, \mathtt{x}) = \bigl\{ \delta(\mathtt{r}_0) + 6\nu_0\, z_{\mathbb{H}}(\mathtt{x})\, \omega \bigr\}\, \mathtt{r}_0^2 \le \biggl\{ \frac{L}{\mathrm{h}^2} \frac{\mathtt{r}_0}{N_1^{1/2}} + 6\nu_0\, z_{\mathbb{H}}(\mathtt{x})\, \frac{\sqrt{n}}{N_2} \biggr\}\, \mathtt{r}_0^2.$$

В третьем разделе полученные ранее результаты применяются к линейной негауссовской модели (7) с семипараметрической функцией регрессии:

$$\boldsymbol{f}^* = \Psi^\top \boldsymbol{\theta}^* + \boldsymbol{g}^*, \tag{11}$$

где $\boldsymbol{\theta}^* \in \mathbb{R}^q$ является неизвестным целевым вектором, а $\Psi = (\Psi_1, \ldots, \Psi_n)$ – матрица размера $q \times n$ с $\Psi_i = (\psi_1(X_i), \ldots, \psi_q(X_i))^\top \in \mathbb{R}^q$ для заданного набора базовых функций $\{\psi_j(\cdot), j = 1, \ldots, q\}$ и точек плана эксперимента X_i, $i = 1, \ldots, n$. Без ограничения общности можно предположить, что базисные функции ортонормированы относительно данного плана эксперимента, т.е. $\sum_i \psi_j(X_i)\psi_{j'}(X_i) = \delta_{j,j'}$. Общий случай можно свести к данному с помощью вращения и перенормирования. Аналогично мы предположим, что элементы вектора мешающего параметра $\boldsymbol{g}^* = \{g^*(X_1), \ldots, g^*(X_n)\}^\top$ 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 бесконечномерного вектора мешающего параметра $\boldsymbol{\eta} = \{\eta_1, \eta_2, \ldots, \eta_m, \ldots\}^\top$.

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

$$\|g^* - g_m^*\| \le \gamma_m. \tag{12}$$

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

$$g^* \in \mathcal{F}_s(\mathtt{C}) \stackrel{\text{def}}{=} \Big\{ g(x) = \sum_{k=1}^\infty \eta_k \varphi_k(x) \colon \sum_{k=1}^\infty \eta_k^2 k^{2s} \le \mathtt{C} \Big\},$$

то $\gamma_m \le (m+1)^{-s}$.

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

$$\|\psi_j - \psi_{j,m}\| \le \gamma_m, \tag{13}$$

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

Теорема 10. *Рассмотрим модель* (7) *с функцией логарифма правдоподобия* (8) *и семипараметрической регрессионной функцией из* (11). *Предположим выполнение условия идентифицируемости и условий гладкости* (12), (13). *Тогда результат теоремы 8 выполняется с* $\rho_m = b_m = (1-\nu)^{-1} q^{1/2} \gamma_m^2$.

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

В четвертом разделе рассматриваются обобщенные линейные модели, которые часто используются для описания категориальных данных. Необходимые для выполнения теоремы БфМ условия имеют свои особенности, но ведут к теоретическим результатам схожим с линейным случаем.

В Заключении перечислены следующие основные результаты:

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 и размера выборки n как $\sqrt{p^3/n}$, 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 усечения базиса.

5. Показана применимость общих теоретических результатов к линейным и обобщенным линейным моделям с мешающим параметром, принадлежащим соболевскому классу гладкости.

Список публикаций

1. 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.

2. Panov M., Spokoiny V. Finite Sample Bernstein – von Mises Theorem for Semiparametric Problems // Published electronically in Bayesian Analysis. 2015. URL `http://projecteuclid.org/euclid.ba/1422884986`.

3. Панов М.Е., Спокойный В.Г. О семипараметрическом оценивании в баейсовской постановке // Труды 55-й научной конференции МФТИ. Управление и прикладная математика. 2012. Т. 1. С. 104–105.

4. Panov M., Spokoiny V. Critical dimension in semiparametric Bernstein - von Mises Theorem // Proceedings of "Information technologies and systems - 2013" 36th conference of young scientists and specialists of IITP RAS. Kaliningrad, Russia: 2013. P. 386–391.

5. 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.

6. 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.