Байесовский выбор модели и концентрация апостериорного распределения для гиперпараметров*

Н. П. БАЛДИН
Лаборатория структурных методов анализа данных в предсказательном моделировании при Московском физико-техническом институте (государственном университете)
email: baldin.np@gmail.com

В. Г. СПОКОЙНЫЙ
Институт прикладного анализа и стохастики им. Вейерштрасса, Берлин, Германия, Берлинский университет им. Гумбольдта, Германия, Лаборатория структурных методов анализа данных в предсказательном моделировании при Московском физико-техническом институте (государственном университете)
e-mail: spokoiny@wias-berlin.de

УДК 519.22

Ключевые слова: максимум правдоподобия, квадратичная регуляризация, локальный брэкетинг.

Аннотация

Данная работа предлагает конструкцию априорного распределения гиперпараметра, которая может использоваться в задаче байесовского выбора модели. Конструкция основывается на идее несмещённой оценки риска в методе максимума правдоподобия с регуляризацией. Главный результат работы показывает одностороннюю концентрацию апостериорного распределения гиперпараметра: апостериорная масса концентрируется в области моделей сложности ниже, чем сложность модели, соответствующей сложности оракульной модели.

Abstract

N. P. Baldin, V. G. Spokoiny, Bayesian model selection and the concentration of the posterior of hyperparameters, Fundamentalnaya i prikladnaya matematika, vol. 18 (2013), no. 2, pp. 13—34.

The present paper offers a construction of a hyperprior that can be used for Bayesian model selection. This construction is inspired by the idea of the unbiased model selection in a penalized maximum likelihood approach. The main result shows a one-sided contraction of the posterior: the posterior mass is allocated on models of lower complexity than the oracle one.

*Работа выполнена при поддержке лаборатории структурных методов анализа данных в предсказательном моделировании, грант Правительства Российской Федерации 11.G34.31.0073.

Фундаментальная и прикладная математика, 2013, том 18, № 2, с. 13—34.

1. Введение

1.1. Основные определения и обозначения

Пусть наблюдения $\boldsymbol{Y}$ определены на некотором вероятностном пространстве и описываются некоторой мерой $\mathbb{P}_{\boldsymbol{f}}$. Параметрическое предположение состоит в том, что данная мера принадлежит семейству $(\mathbb{P}_{\boldsymbol{\theta}},\ \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p)$ с доминирующей мерой $\boldsymbol{\mu}_0$. Принцип максимума правдоподобия предлагает оценивать вектор $\boldsymbol{\theta}$, параметризующий истинную меру, максимизируя функцию правдоподобия $L(\boldsymbol{\theta}) = L(\boldsymbol{Y}, \boldsymbol{\theta}) \stackrel{\text{def}}{=} \log \frac{d\mathbb{P}_{\boldsymbol{\theta}}}{d\boldsymbol{\mu}_0}(\boldsymbol{Y})$:

$$\tilde{\boldsymbol{\theta}} \stackrel{\text{def}}{=} \operatorname*{argmax}_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta}).$$

Параметрические предположение может оказаться неверным: $\mathbb{P}_{\boldsymbol{f}} \notin \mathbb{P}_{\boldsymbol{\theta}}$. В этом случае $\tilde{\boldsymbol{\theta}}$ оценивает лучшую аппроксимацию истинной меры мерой параметрического семейства. Вектор, определяющий лучшую аппроксимацию, определяется как

$$\boldsymbol{\theta}^* \stackrel{\text{def}}{=} \operatorname*{argmax}_{\boldsymbol{\theta} \in \Theta} \mathbb{E}_{\boldsymbol{f}} L(\boldsymbol{\theta}).$$

Математическое ожидание во всей работе берётся по истинной мере $\mathbb{P}_{\boldsymbol{f}}$. Для простоты будем писать $\mathbb{E}$ вместо $\mathbb{E}_{\boldsymbol{f}}$. Некоторые полезные свойства оценки максимума правдоподобия можно найти, например, в [11].

Для повышения устойчивости оценки $\tilde{\boldsymbol{\theta}}$ в модель искусственно вводится регуляризация, которую можно рассмотреть как штраф за сложность модели. Зафиксируем некоторое семейство функций $\operatorname{pen}(\boldsymbol{\theta}, \varkappa)$ для $\varkappa \in \mathcal{K}$ и $\mathcal{K} \subseteq \mathbb{R}$, которые выполняют роль штрафа. Тогда логарифмическая функция правдоподобия со штрафом описывает $\varkappa$-модель

$$L_{\varkappa}(\boldsymbol{\theta}) \stackrel{\text{def}}{=} L(\boldsymbol{\theta}) - \operatorname{pen}(\boldsymbol{\theta}, \varkappa).$$

Для $\varkappa$-модели оценка $\tilde{\boldsymbol{\theta}}_{\varkappa}$ и величина $\boldsymbol{\theta}^*_{\varkappa}$, которую оценивает оценка, могут быть определены следующим образом:

$$\tilde{\boldsymbol{\theta}}_{\varkappa} \stackrel{\text{def}}{=} \operatorname*{argmax}_{\boldsymbol{\theta} \in \Theta} L_{\varkappa}(\boldsymbol{\theta}), \quad \boldsymbol{\theta}^*_{\varkappa} \stackrel{\text{def}}{=} \operatorname*{argmax}_{\boldsymbol{\theta} \in \Theta} \mathbb{E} L_{\varkappa}(\boldsymbol{\theta}).$$

В данной работе рассматривается специальный случай квадратичной пенализации. Соответствующая функция штрафа есть $\operatorname{pen}(\boldsymbol{\theta}, \varkappa) = \varkappa \|G\boldsymbol{\theta}\|^2/2$ с евклидовой нормой, где G^2 — некоторая диагональная матрица с элементами на диагонали $0 < g_1^2 \leqslant g_2^2 \leqslant \ldots \leqslant g_p^2$. Частный случай такой регуляризации при $G^2 = I_p$ есть регуляризация Тихонова, изученная, например, в [7]. Там же исследованы различные методы регуляризации в обратных задачах. Главный вопрос после выбора фиксированного семейства функций штрафа $\operatorname{pen}(\boldsymbol{\theta}, \varkappa)$, какую модель выбрать или, что эквивалентно, как выбрать параметр $\varkappa$.

Регуляризация вектора $\boldsymbol{\theta}$, как правило, эквивалентна предположению, что параметр $\boldsymbol{\theta}$ есть случайный элемент на некотором вероятностном пространстве с мерой $\Pi_\varkappa$, которая имитирует структуру штрафа. Распределение данных описывается как

$$\boldsymbol{Y} \mid \boldsymbol{\vartheta} \sim p(\boldsymbol{y} \mid \boldsymbol{\theta}) = \exp L(\boldsymbol{\theta}), \quad \boldsymbol{\vartheta} \sim \Pi_\varkappa(\cdot).$$

Объект изучения — это апостериорная мера, которая описывает условное распределение $\boldsymbol{\vartheta}$ при данных $\boldsymbol{Y}$. Используя формулу Байеса, получаем

$$\boldsymbol{\vartheta} \mid \boldsymbol{Y} \propto \exp\{L(\boldsymbol{\theta})\}\, \Pi_\varkappa(d\boldsymbol{\theta}).$$

Объект зависит от распределения $\Pi_\varkappa$, но при асимптотическом анализе зависимость несущественна. Априорное распределение $\Pi_\varkappa(\cdot)$ параметризуется другим вектором $\varkappa \in \mathcal{K} \subseteq \mathbb{R}$. Наиболее изучены два способа выбора значения $\varkappa$:

— оценить $\varkappa$;
— предполагать некоторое распределение $\varkappa$ с плотностью $\pi(\cdot)$, которое в конечном итоге будет зависеть от параметров, влияние которых на анализ мало.

В данной работе изучаются оба метода и исследуется их эквивалентность. Основной объект изучения в работе есть

$$P(\boldsymbol{\theta} \in A,\ \varkappa \in B \mid \boldsymbol{Y}) \propto \int\limits_A \exp\bigl(L(\boldsymbol{\theta})\bigr) \int\limits_B \pi(\varkappa)\Pi_\varkappa(d\boldsymbol{\theta})\, d\varkappa,$$

где $A \subseteq \Theta$ и $B \subseteq \mathcal{K}$.

1.2. Обзор известных результатов

В литературе встречаются несколько способов выбора функции $\operatorname{pen}(\boldsymbol{\theta}, \varkappa)$. Информационный критерий Акаике, изученный в [1], использует удвоенное количество параметров модели как функцию штрафа. Другой популярный метод, предложенный в [12], основан на идее несмещённой оценки риска Штайна. Этот метод был изучен подробно в [5, 6] среди многих других. В этом случае функция $\operatorname{pen}(\boldsymbol{\theta}, \varkappa)$ зависит от гладкости целевого параметра и дисперсии шума. Ещё один популярный метод основан на сглаживающих сплайнах (см. [8]).

К сожалению, методы, основанные на идее несмещённой оценки риска, на практике обнаруживают существенные недостатки. В частности, такие методы теряют устойчивость в задачах с плохой обусловленностью или с большим количеством различных моделей для выбора. Как следствие, подобные методы теряют концентрационные и оракульные свойства. Таким образом, выбор семейства функций регуляризации и выбор конкретной функции из рассматриваемого семейства являются главными вопросами в задачах регуляризации модели. Выбор подходящей функции штрафа в гауссовских моделях был изучен в [2, 3]. Эффективный метод для линейных гауссовских обратных задач был изучен в [4].

2. Выбор лучшей модели

Далее рассматривается конкретная модель $\boldsymbol{Y} = \boldsymbol{f} + \boldsymbol{\varepsilon} \in \mathbb{R}^p$, где $\boldsymbol{\varepsilon} \sim \sim \mathcal{N}(0, \sigma^2 I_p)$, и функция штрафа $\operatorname{pen}(\boldsymbol{\theta}, \varkappa) = \varkappa \|G\boldsymbol{\theta}\|^2/2$.

2.1. Выбор модели, основанный на несмещённой оценке риска

Оценка $\tilde{\boldsymbol{\theta}}_\varkappa$ внутри $\varkappa$-модели получается максимизацией $L_\varkappa(\boldsymbol{\theta})$:

$$\begin{aligned}\tilde{\boldsymbol{\theta}}_\varkappa &= \operatorname*{argmax}_{\boldsymbol{\theta}} L_\varkappa(\boldsymbol{\theta}) = \operatorname*{argmin}_{\boldsymbol{\theta}} \{\sigma^{-2}\|\boldsymbol{Y} - \boldsymbol{\theta}\|^2 + \varkappa\|G\boldsymbol{\theta}\|^2\} = \\ &= \sigma^{-2}(\sigma^{-2}I_p + \varkappa G^2)^{-1}\boldsymbol{Y} = \sigma^{-2}D_\varkappa^{-2}\boldsymbol{Y},\end{aligned}$$

где $D_\varkappa^2 = \sigma^{-2}I_p + \varkappa G^2$. Величина, которую оценивает оценка $\tilde{\boldsymbol{\theta}}_\varkappa$ внутри $\varkappa$-модели, есть

$$\boldsymbol{\theta}^*_\varkappa = \operatorname*{argmax}_{\boldsymbol{\theta}} \mathbb{E}L_\varkappa(\boldsymbol{\theta}) = \sigma^{-2}(\sigma^{-2}I_p + \varkappa G^2)^{-1}\mathbb{E}\boldsymbol{Y} = \sigma^{-2}D_\varkappa^{-2}\boldsymbol{f},$$

где $\boldsymbol{f} \stackrel{\text{def}}{=} \mathbb{E}\boldsymbol{Y}$. Разложение Тейлора $L_\varkappa(\boldsymbol{\theta})$ при фиксированном $\varkappa$ ведёт к понятию *эксцесс*: $\mathcal{X}_\varkappa \stackrel{\text{def}}{=} L_\varkappa(\tilde{\boldsymbol{\theta}}_\varkappa) - L_\varkappa(\boldsymbol{\theta}^*_\varkappa)$,

$$\mathcal{X}_\varkappa = L_\varkappa(\tilde{\boldsymbol{\theta}}_\varkappa) - L_\varkappa(\boldsymbol{\theta}^*_\varkappa) = \frac{1}{2}\|D_\varkappa(\tilde{\boldsymbol{\theta}}_\varkappa - \boldsymbol{\theta}^*_\varkappa)\|^2 = \frac{1}{2}\|\sigma^{-2}D_\varkappa^{-1}(\boldsymbol{Y} - \boldsymbol{f})\|^2.$$

Качество оценивания внутри $\varkappa$-модели может быть измерено как математическое ожидание эксцесса

$$\mathbb{E}\mathcal{X}_\varkappa = \frac{1}{2\sigma^2}\operatorname{tr}(D_\varkappa^{-2}) = \frac{1}{2}\sum_{j=1}^{p}\frac{1}{1 + \sigma^2\varkappa g_j^2}.$$

Для линейных моделей математическое ожидание эксцесса внутри $\varkappa$-модели не зависит от истинного вектора $\boldsymbol{f}$ и может быть вычислено, если известна дисперсия шума. Заметим, что качество оценки определяется именно в терминах значения функции правдоподобия $L_\varkappa$ в точках $\tilde{\boldsymbol{\theta}}_\varkappa$ и $\boldsymbol{\theta}^*_\varkappa$, а не расстоянием $\mathbb{E}\|\tilde{\boldsymbol{\theta}}_\varkappa - \boldsymbol{\theta}^*_\varkappa\|^2$.

Смещение $b_\varkappa$ фиксированной $\varkappa$-модели может быть определено через расстояние Кульбака—Лейблера

$$\operatorname{KL}(\mathbb{P}_{\boldsymbol{\theta}^*_\varkappa}, \mathbb{P}_{\boldsymbol{f}}) = -\mathbb{E}L(\boldsymbol{\theta}^*_\varkappa) = \frac{1}{2\sigma^2}\|\boldsymbol{f} - \boldsymbol{\theta}^*_\varkappa\|^2.$$

Геометрический смысл заключается в том, что расстояние между мерой, параметризованной $\boldsymbol{\theta}^*_\varkappa$, и истинной мерой не должно быть слишком большим. Альтернативно можно определить модельное смещение величиной $-\mathbb{E}L_\varkappa(\boldsymbol{\theta}^*_\varkappa)$, которая отличается от расстояния Кульбака—Лейблера на величину $\operatorname{pen}(\boldsymbol{\theta}^*_\varkappa, \varkappa)$:

$$b_\varkappa \stackrel{\text{def}}{=} -\mathbb{E}L_\varkappa(\boldsymbol{\theta}^*_\varkappa) = \operatorname{KL}(\mathbb{P}_{\boldsymbol{\theta}^*_\varkappa}, \mathbb{P}_{\boldsymbol{f}}) - \operatorname{pen}(\boldsymbol{\theta}^*_\varkappa, \varkappa).$$

Общий риск оценивания $\mathcal{R}_\varkappa$ можно определить как сумму математического ожидания эксцесса и смещения между моделями:

$$\mathcal{R}_\varkappa = \mathbb{E}\mathcal{X}_\varkappa + b_\varkappa.$$

Естественно определить *оракульный* выбор $\varkappa^\circ$ как минимум риска $\mathcal{R}_\varkappa$:

$$\varkappa^\circ = \operatorname*{argmin}_{\varkappa \in \mathcal{K}} \mathcal{R}_\varkappa.$$

Заметим, что предлагаемый метод существенно отличается от широко распространённого метода оценивания $\mathbb{E}\|\tilde{\boldsymbol{\theta}}_\varkappa - \boldsymbol{f}\|^2$ как раз из-за того, что рассматриваются значения функций правдоподобия. То есть идея состоит в исследовании искусственно созданных $\varkappa$-моделей, когда $\mathbb{E}L(\boldsymbol{\theta}^*_\varkappa)$ не сильно отличается от $\mathbb{E}L(f)$.

Иначе можно определять лучшую точку $\varkappa$ через оптимизацию математического ожидания правдоподобия $\mathbb{E}L_\varkappa(\boldsymbol{\theta})$ по обоим параметрам как

$$\varkappa^* = \operatorname*{argmax}_{\varkappa \in \mathcal{K}} \max_{\boldsymbol{\theta} \in \Theta} \mathbb{E}L_\varkappa(\boldsymbol{\theta}) = \operatorname*{argmax}_{\varkappa \in \mathcal{K}} \mathbb{E}L_\varkappa(\boldsymbol{\theta}^*_\varkappa)$$

К сожалению, данная величина никак не учитывает «положительный эффект» от регуляризации внутри $\varkappa$-модели. Но если получится как-то внести в модель эту информацию и рассматривать $\mathbb{E}L_\varkappa(\boldsymbol{\theta}^*_\varkappa) - \mathbb{E}\mathcal{X}_\varkappa$ как правдоподобие модели, то очевидно, что $\varkappa^* = \varkappa^\circ$, и тогда можно использовать все хорошо известные свойства оценки максимума правдоподобия.

2.2. Байесовский выбор модели

В байесовском выборе модели предполагается, что параметр $\boldsymbol{\theta}$ случайный. Рассматриваемая регуляризация эквивалентна гауссовскому распределению $\boldsymbol{\theta}$. Модель задаётся следующим распределением:

$$\boldsymbol{Y} \mid \boldsymbol{\vartheta} \sim \exp\bigl(L(\boldsymbol{\theta})\bigr), \quad \boldsymbol{\vartheta} \mid \varkappa \sim \Pi_\varkappa(\cdot).$$

Для апостериорного распределения $\boldsymbol{\vartheta}$ при условии данных $\boldsymbol{Y}$ имеем

$$\mathbb{P}(\boldsymbol{\vartheta} \in A \mid \boldsymbol{Y}) \propto \int\limits_A \exp\{L(\boldsymbol{\theta})\}\Pi_\varkappa(d\boldsymbol{\theta}) = \int\limits_A \exp\{L(\boldsymbol{\theta})\}\pi_\varkappa(\boldsymbol{\theta})\, d\boldsymbol{\theta}. \tag{2.1}$$

Неизвестный параметр $\varkappa$ можно выбирать, максимизируя данное выражение относительно $\varkappa$:

$$\hat{\varkappa} = \operatorname*{argmax}_{\varkappa \in \mathcal{K}} \int\limits_A \exp\{L(\boldsymbol{\theta})\}\pi_\varkappa(\boldsymbol{\theta})\, d\boldsymbol{\theta}.$$

Данный метод называется методом максимума апостериорной вероятности и изучался, например, в [9], где были рассмотрены асимптотические свойства апостериорного распределения, в том числе его концентрация вокруг истинного значения $\boldsymbol{\theta}^*$.

Другой метод основан на максимизации по $\varkappa$ максимального значения по параметру $\boldsymbol{\theta}$ подынтегрального выражения в (2.1):

$$\hat{\varkappa} = \operatorname*{argmax}_{\varkappa \in \mathcal{K}} \max_{\boldsymbol{\theta}} \{L(\boldsymbol{\theta}) + \log \pi_{\varkappa}(\boldsymbol{\theta})\}.$$

Видно, что данная оценка $\hat{\varkappa}$ совпадает с

$$\operatorname*{argmax}_{\varkappa \in \mathcal{K}} \max_{\boldsymbol{\theta}} L_{\varkappa}(\boldsymbol{\theta})$$

при

$$\operatorname{pen}(\boldsymbol{\theta}, \varkappa) = -\log \pi_{\varkappa}(\boldsymbol{\theta}).$$

При полном байесовском подходе предполагается, что $\varkappa$ тоже случайный элемент, определённый на некотором вероятностном пространстве, из некоторого распределения с плотностью $\pi(\cdot)$, которое в конечном итоге будет зависеть от параметров, влияние которых мало на анализ. Для такой модели апостериорное распределение есть

$$\begin{aligned}\mathbb{P}(\boldsymbol{\vartheta} \in A,\ \varkappa \in B \mid \boldsymbol{Y}) \propto \int\limits_A \exp\{L(\boldsymbol{\theta})\} \int\limits_B \pi_{\varkappa}(\boldsymbol{\theta})\pi(\varkappa)\, d\boldsymbol{\theta}\, d\varkappa = \\ = \int\limits_A \int\limits_B \exp\{L(\boldsymbol{\theta}) + \log \pi_{\varkappa}(\boldsymbol{\theta}) + \log \pi(\varkappa)\}\, d\boldsymbol{\theta}\, d\varkappa. \qquad (2.2)\end{aligned}$$

Для гауссовской модели

$$\begin{aligned}\boldsymbol{Y} \mid \boldsymbol{\vartheta} &\sim \mathcal{N}(\boldsymbol{\theta}, \sigma^2 I_p), \\ \boldsymbol{\vartheta} \mid \varkappa &\sim \mathcal{N}(0, \varkappa^{-1} G^{-2}), \\ \varkappa &\sim \pi(d\varkappa)\end{aligned}$$

подынтегральное выражение в (2.2) — это совместная плотность $(\boldsymbol{Y}, \boldsymbol{\vartheta}, \varkappa)$:

$$(\boldsymbol{Y}, \boldsymbol{\vartheta}, \varkappa) \sim \exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{Y} - \boldsymbol{\theta}\|^2 - \frac{\varkappa}{2}\|G\boldsymbol{\theta}\|^2 - \frac{1}{2}\log\det(\varkappa^{-1}G^{-2}) + \log \pi(\varkappa)\right), \qquad (2.3)$$

где член $(1/2)\log\det(\varkappa^{-1}G^{-2})$ — нормализация плотности $\pi_{\varkappa}(\boldsymbol{\theta})$.

Далее в работе показывается, что апостериорное распределение (2.2) при $A = \Theta$ концентрируется вокруг эмпирической версии оракульной оценки при подходящем априорном распределении $\pi(\varkappa)$. Как было отмечено, оценка максимума правдоподобия совпадает с оракульной оценкой, если сделать небольшую коррекцию функции правдоподобия. Естественным представляется использование байесовского подхода и осуществление данной коррекции путём предположения соответствующего априорного распределения.

3. Основные результаты

Исследуется априорное распределение $\pi(\varkappa)$, которое имитирует идею оракульного оценивания.

3.1. Построение априорного распределения

Рассмотрим распределение

$$\pi_0(\varkappa) = \varkappa^{-p/2}\exp(-\mathbb{E}\mathcal{X}_\varkappa).$$

Очевидно, что соответствующая логарифмическая плотность (2.3) равна $L_\varkappa(\boldsymbol{\theta})$ без учёта постоянного члена с матрицей G^2. К сожалению, данное распределение не влечёт концентрации, что будет показано далее. Ниже рассматривается несколько модифицированное распределение

$$\pi(d\varkappa) \propto \pi_0(\varkappa)\exp(-\varepsilon\mathbb{E}\mathcal{X}_\varkappa + c\log\varkappa)\,d\varkappa,$$

где ε и c — некоторые положительные константы. Априорное распределение с параметрами $c = \varepsilon = 0$ не является достаточным для концентрации. Существует «худший» случай, когда риск становится «плоским». Для любой модели из рассматриваемого класса существует сигнал $f_i = \sigma$, который влечёт $\mathcal{R}_\varkappa = p/2$ независимо от $\varkappa$.

3.2. Односторонняя концентрация апостериорного распределения $\varkappa$

В этом разделе описывается главный результат для апостериорного распределения $\varkappa$. Обозначим $\Omega(\mathtt{x})$ случайное событие с вероятностью $\mathbb{P}(\Omega(\mathtt{x})) \geqslant 1-\mathtt{C}\mathrm{e}^{-x}$. Удобно ввести функцию

$$\mathcal{L}(\boldsymbol{\theta},\varkappa) \stackrel{\text{def}}{=} L(\boldsymbol{\theta}) + \log\pi_\varkappa(\boldsymbol{\theta}) + \log\pi(\varkappa)$$

из (2.2). Также будем использовать обозначение $\boldsymbol{\nu} = (\boldsymbol{\theta},\varkappa)$ для простоты изложения. Таким образом, будем писать $\mathcal{L}(\boldsymbol{\nu})$ вместо $\mathcal{L}(\boldsymbol{\theta},\varkappa)$, если нет необходимости выделять компоненты вектора $\boldsymbol{\nu}$. Также введём обозначение

$$\mathcal{D}_0^2 = -\nabla^2\mathbb{E}\mathcal{L}(\boldsymbol{\nu}^\star),$$

где

$$\boldsymbol{\nu}^\star = (\boldsymbol{\theta}^*,\varkappa^*) = \operatorname*{argmax}_{\boldsymbol{\theta},\varkappa}\mathbb{E}\mathcal{L}(\boldsymbol{\theta},\varkappa),$$

и зададим локальную окрестность точки $\varkappa^*$:

$$\mathcal{K}_\circ(\mathtt{r}) \stackrel{\text{def}}{=} \{\varkappa\colon \|(\boldsymbol{\nu}-\boldsymbol{\nu}^\star)^\top\mathcal{D}_0^2(\boldsymbol{\nu}-\boldsymbol{\nu}^\star)\| \leqslant \mathtt{r}\}.$$

Для удобства обозначим границы множества $\mathcal{K}_\circ(\mathtt{r}_0)$ через $\varkappa_1$ и $\varkappa_2$, $\varkappa_1 < \varkappa_2$, и введём множества $\mathcal{K}_1 = \{\varkappa \in \mathbb{R}_+ \colon \varkappa \leqslant \varkappa_1\}$ и $\mathcal{K}_2 = \{\varkappa \in \mathbb{R}_+ \colon \varkappa \geqslant \varkappa_2\}$. Далее предполагается, что локальный радиус $\mathtt{r}_0$ зафиксирован определённым образом. Условия на выбор $\mathtt{r}_0$ будут указаны ниже в разделе D. Удобно рассмотреть матрицу $\mathcal{D}_0^2$ в блочном виде:

$$\mathcal{D}_0^2 = \begin{pmatrix} D_0^2 & A \\ A^\top & H_0^2 \end{pmatrix},$$

где $D_0^2 = D_{\varkappa^*}^2 = \sigma^{-2} I_p + \varkappa^* G^2$. Обозначим $d_0^2 = H_0^2 - A^\top D_0^{-2} A$, и пусть $\xi = d_0^{-1} \breve{\nabla}_\varkappa$, где $\breve{\nabla}_\varkappa = -D_0^{-2} A \nabla_{\boldsymbol{\theta}}$ и $\nabla_{\boldsymbol{\theta}} = \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\nu}^\star)$.

Определим

$$\varkappa_{\boldsymbol{\epsilon}} \stackrel{\text{def}}{=} \bar{\mathbb{E}} \varkappa, \quad \mathfrak{S}^2 \stackrel{\text{def}}{=} \operatorname{Cov}(\varkappa) \stackrel{\text{def}}{=} \bar{\mathbb{E}} (\varkappa - \varkappa_{\boldsymbol{\epsilon}})^2,$$

где $\bar{\mathbb{E}}$ — условное математическое ожидание:

$$\bar{\mathbb{E}} \mu \stackrel{\text{def}}{=} \mathbb{E}[\mu \mid \boldsymbol{Y}, \ \varkappa \in \mathcal{K}_1 \cup \mathcal{K}_\circ(\mathtt{r}_0)]$$

для случайной величины η. Обе величины зависят от данных и могут быть вычислены.

Далее формулируется результат в неасимптотической постановке, который утверждает, что $\varkappa_{\boldsymbol{\epsilon}}$ близко к оценке $\tilde{\varkappa}$, $\mathfrak{S}^2$ примерно равно обратной дисперсии d_0^{-2} оценки $\tilde{\varkappa}$ и $\mathfrak{S}^{-1}(\varkappa_{\boldsymbol{\epsilon}} - \tilde{\varkappa})$ при условии данных имеет примерно стандартно нормальное распределение. Под константой $\mathtt{C}$ будем подразумевать абсолютную константу, не обязательно одну и ту же во всех результатах.

Теорема 3.1. *При условии* $(\mathcal{L}_0)$ *на множестве* $\mathcal{K}_\circ(\mathtt{r}_0)$ (*см. раздел* A) *на множестве доминирующей вероятности* $\Omega(\mathtt{x})$ *с* $\delta(\mathtt{r}_0) > 0$ *из условия* $(\mathcal{L}_0)$ *выполняется*

$$d_0^2 (\varkappa_{\boldsymbol{\epsilon}} - \tilde{\varkappa})^2 \leqslant \mathtt{C}\, \delta(p + 1 + \mathtt{x}), |1 - d_0^2 \mathfrak{S}^2| \leqslant \mathtt{C}\, \delta(p + 1 + \mathtt{x}).$$

Более того, для любого $\lambda \in \mathbb{R}$ *с* $\lambda^2 \leqslant \mathtt{x}$

$$|\log \bar{\mathbb{E}}[\exp\{\lambda d_0 (\varkappa - \tilde{\varkappa})\}] - \lambda^2/2| \leqslant \mathtt{C}\, \delta(p + 1 + \mathtt{x}).$$

Для любого измеримого множества $A \subset \mathbb{R}$ *и* $\mathfrak{q} = 1 + |\xi|^2$

$$\begin{aligned}
&\mathbb{P}\bigl(d_0(\varkappa - \tilde{\varkappa}) \in A \mid \boldsymbol{Y}, \varkappa \in \mathcal{K}_1 \cup \mathcal{K}_\circ(\mathtt{r}_0)\bigr) \geqslant \\
&\quad \geqslant \exp\bigl(-\mathtt{C}\, \delta(p + 1 + \mathtt{x})\bigr) \{\mathbb{P}(\gamma \in A) - \mathtt{C}\delta \mathfrak{q}^{1/2}\} - \mathtt{C}\mathrm{e}^{-\mathtt{x}}, \\
&\mathbb{P}\bigl(d_0(\varkappa - \tilde{\varkappa}) \in A \mid \boldsymbol{Y}, \varkappa \in \mathcal{K}_1 \cup \mathcal{K}_\circ(\mathtt{r}_0)\bigr) \leqslant \\
&\quad \leqslant \exp\bigl(\mathtt{C}\, \delta(p + 1 + \mathtt{x})\bigr) \{\mathbb{P}(\gamma \in A) + \mathtt{C}\delta\, \mathfrak{q}^{1/2}\} + \mathtt{C}\mathrm{e}^{-\mathtt{x}}.
\end{aligned}$$

Полученный результат описывает апостериорное распределение $\varkappa$ в зоне $\varkappa < \varkappa^*$. Согласно последнему утверждению теоремы это распределение мажорируется сверху и снизу двумя нормальными распределениями, что можно рассматривать как одностороннюю теорему Бернштейна—фон Мизеса (см. [13]).

A. Общий подход

Приложение содержит условия и доказательства основной теоремы и вспомогательных утверждений. Напомним, что

$$\mathcal{L}(\boldsymbol{\theta}, \varkappa) \stackrel{\text{def}}{=} L(\boldsymbol{\theta}) + \log \pi_\varkappa(\boldsymbol{\theta}) + \log \pi(\varkappa) \tag{A.1}$$

из (2.2).

Рассмотрим разложение процесса $\mathcal{L}(\boldsymbol{\nu})$ ($\boldsymbol{\nu} = (\boldsymbol{\theta}, \varkappa)$) как сумму его математического ожидания и стохастической компоненты $\mathcal{L}(\boldsymbol{\nu}) = \mathbb{E}\mathcal{L}(\boldsymbol{\nu}) + \boldsymbol{\zeta}(\boldsymbol{\nu})$. Обозначим $\mathcal{L}(\boldsymbol{\nu}, \boldsymbol{\nu}^{\star}) = \mathcal{L}(\boldsymbol{\nu}) - \mathcal{L}(\boldsymbol{\nu}^{\star})$, где

$$\boldsymbol{\nu}^{\star} = (\boldsymbol{\theta}^{*}, \varkappa^{*}) = \operatorname*{argmax}_{\boldsymbol{\theta}, \varkappa} \mathbb{E}\mathcal{L}(\boldsymbol{\theta}, \varkappa).$$

Разложение вблизи истинной точки $\boldsymbol{\nu}^{\star}$ Тейлора даёт

$$\mathbb{E}\mathcal{L}(\boldsymbol{\nu}, \boldsymbol{\nu}^{\star}) \approx -\frac{1}{2}(\boldsymbol{\nu} - \boldsymbol{\nu}^{\star})^{\top} \mathcal{D}_0^2 (\boldsymbol{\nu} - \boldsymbol{\nu}^{\star}), \tag{A.2}$$

где используется, что $\nabla \mathbb{E}\mathcal{L}(\boldsymbol{\nu}^{\star}) = 0$ и $\mathcal{D}_0^2 = -\nabla^2 \mathbb{E}\mathcal{L}(\boldsymbol{\nu}^{\star})$. Введём

$$\tilde{\boldsymbol{\nu}} = (\tilde{\boldsymbol{\theta}}, \tilde{\varkappa}) = \operatorname*{argmax}_{\boldsymbol{\theta}, \varkappa} \mathcal{L}(\boldsymbol{\theta}, \varkappa).$$

Член $\log \pi_{\varkappa}(\boldsymbol{\theta}) + \log \pi(\varkappa)$ в (A.1) детерминированный, поэтому стохастическая компонента $\boldsymbol{\zeta}(\boldsymbol{\nu})$ функции $\mathcal{L}(\boldsymbol{\nu})$ совпадает со стохастической компонентой процесса $L(\boldsymbol{\theta})$ и не зависит от $\varkappa$. Более того, для гауссовской модели (2.3) стохастическая компонента линейная:

$$\boldsymbol{\zeta}(\boldsymbol{\nu}) - \boldsymbol{\zeta}(\boldsymbol{\nu}^{\star}) = (\boldsymbol{\nu} - \boldsymbol{\nu}^{\star})^{\top} \nabla \boldsymbol{\zeta}(\boldsymbol{\nu}^{\star}), \quad \nabla \boldsymbol{\zeta}(\boldsymbol{\nu}^{\star}) = \begin{pmatrix} \sigma^{-2}(\boldsymbol{Y} - \boldsymbol{f}) \\ 0 \end{pmatrix}. \tag{A.3}$$

Матрица $\mathcal{D}_0^2$ — гессиан со знаком минус математического ожидания $\mathbb{E}\mathcal{L}(\boldsymbol{\nu})$ в точке $(\boldsymbol{\nu}^{\star})$. Для гауссовской модели она имеет вид

$$\mathcal{D}_0^2 = \begin{pmatrix} \sigma^{-2} I_p + \varkappa^{*} G^2 & \sigma^{-2} G^2 (\sigma^2 I_p + \varkappa^{*} G^2)^{-1} \boldsymbol{f} \\ (\sigma^{-2} G^2 (\sigma^{-2} I_p + \varkappa^{*} G^2)^{-1} \boldsymbol{f})^{\top} & (2\varkappa^{*2})^{-2} p - \frac{d^2}{d\varkappa^2} \log \pi(\varkappa) \end{pmatrix}.$$

Для дальнейшего анализа удобно использовать более компактную запись

$$\mathcal{D}_0^2 = \begin{pmatrix} D_0^2 & A \\ A^{\top} & H_0^2 \end{pmatrix}$$

где $D_0^2 = D_{\varkappa^{*}}^2 = \sigma^{-2} I_p + \varkappa^{*} G^2$.

Используя (A.2) и (A.3), можно разложить процесс $\mathcal{L}(\boldsymbol{\nu})$, введя $\boldsymbol{\nu}_{\varkappa}^{\star} = (\boldsymbol{\theta}_{\varkappa}^{*}, \varkappa)$:

$$\mathbb{E}\mathcal{L}(\boldsymbol{\nu}) - \mathbb{E}\mathcal{L}(\boldsymbol{\nu}^{\star}) = \mathbb{E}\mathcal{L}(\boldsymbol{\nu}) - \mathbb{E}\mathcal{L}(\boldsymbol{\nu}_{\varkappa}^{\star}) + \mathbb{E}\mathcal{L}(\boldsymbol{\nu}_{\varkappa}^{\star}) - \mathbb{E}\mathcal{L}(\boldsymbol{\nu}^{\star}),$$

$$\mathbb{E}\mathcal{L}(\boldsymbol{\nu}) - \mathbb{E}\mathcal{L}(\boldsymbol{\nu}_{\varkappa}^{\star}) = -\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_{\varkappa}^{*})^{\top} D_{\varkappa}^2 (\boldsymbol{\theta} - \boldsymbol{\theta}_{\varkappa}^{*}), \tag{A.4}$$

$$\mathbb{E}\mathcal{L}(\boldsymbol{\nu}_{\varkappa}^{\star}) - \mathbb{E}\mathcal{L}(\boldsymbol{\nu}^{\star}) \approx -\frac{1}{2}(\boldsymbol{\nu}_{\varkappa}^{\star} - \boldsymbol{\nu}^{\star})^{\top} \mathcal{D}_0^2 (\boldsymbol{\nu}_{\varkappa}^{\star} - \boldsymbol{\nu}^{\star}), \tag{A.5}$$

где для $\mathbb{E}\mathcal{L}(\boldsymbol{\nu}) - \mathbb{E}\mathcal{L}(\boldsymbol{\nu}_{\varkappa}^{\star})$ использовано, что процесс гауссовский при фиксированном $\varkappa$.

Подход [10] предполагает выполнение нескольких достаточных условий. Для гауссовской модели условия существенно упрощаются. Локальная окрестность $\mathcal{K}_{\circ}(\mathtt{r})$ лучшей точки $\varkappa^{*}$ зависит только от параметра $\varkappa$. Введём $\Upsilon_{\circ}(\mathtt{r}) = \Theta \times \mathcal{K}_{\circ}(\mathtt{r})$. Локальное условие описывает свойства процесса $\mathcal{L}(\boldsymbol{\nu})$ для $\varkappa \in \mathcal{K}_{\circ}(\mathtt{r}_0)$ при некотором фиксированном значении $\mathtt{r}_0$.

B. Локальное условие

$(\mathcal{L}_0)$ для любого $\mathtt{r} \leqslant \mathtt{r}_0$ найдутся $\delta(\mathtt{r}) \leqslant 1/2$ и положительно определённая матрица $\mathcal{D}_0^2$, такие что для любого $\varkappa \in \mathcal{K}_\circ(\mathtt{r})$ справедливо

$$\left| \frac{-2\mathbb{E}\mathcal{L}(\boldsymbol{\nu}, \boldsymbol{\nu}^\star)}{(\boldsymbol{\nu} - \boldsymbol{\nu}^\star)^\top \mathcal{D}_0^2 (\boldsymbol{\nu} - \boldsymbol{\nu}^\star)} - 1 \right| \leqslant \delta(\mathtt{r}). \tag{B.1}$$

Локальное условие нужно, чтобы контролировать ошибку в разложении Тейлора (A.2) вблизи окрестности точки $\boldsymbol{\nu}^\star$.

C. Локальный брэкетинг и большие уклонения

Основной шаг в подходе [10] — это локальный брэкетинг, или зажатие процесса $\mathcal{L}(\boldsymbol{\nu})$ в локальной окрестности точки $\boldsymbol{\nu}^\star$ двумя квадратичными функциями:

$$\mathbb{L}_{\underline{\boldsymbol{\epsilon}}}(\boldsymbol{\nu}, \boldsymbol{\nu}^\star) \leqslant \mathcal{L}(\boldsymbol{\nu}, \boldsymbol{\nu}^\star) \leqslant \mathbb{L}_{\boldsymbol{\epsilon}}(\boldsymbol{\nu}, \boldsymbol{\nu}^\star), \quad \boldsymbol{\nu} \in \Upsilon_\circ(\mathtt{r}), \tag{C.1}$$

где

$$\mathbb{L}_{\boldsymbol{\epsilon}}(\boldsymbol{\nu}, \boldsymbol{\nu}^\star) \stackrel{\text{def}}{=} (\boldsymbol{\nu} - \boldsymbol{\nu}^\star)^\top \nabla \boldsymbol{\zeta}(\boldsymbol{\nu}) - \frac{1}{2} (\boldsymbol{\nu} - \boldsymbol{\nu}^\star)^\top \mathcal{D}_{\boldsymbol{\epsilon}}^2 (\boldsymbol{\nu} - \boldsymbol{\nu}^\star), \tag{C.2}$$

$\mathcal{D}_{\boldsymbol{\epsilon}}^2 = (1-\delta)\mathcal{D}_0^2$. Разложение напрямую следует из условия $(\mathcal{L}_0)$. Аналогично для $\underline{\boldsymbol{\epsilon}} = -\boldsymbol{\epsilon}$.

Матрицу $\mathcal{D}_{\boldsymbol{\epsilon}}^2$ можно диагонализировать. Используя замену переменных $\gamma = \boldsymbol{\theta} - \boldsymbol{\theta}^* + D_0^{-2} A(\varkappa - \varkappa^*)$, запишем (C.2) как

$$\mathbb{L}_{\boldsymbol{\epsilon}}(\boldsymbol{\nu}, \boldsymbol{\nu}^\star) = (\gamma - \gamma^*)^\top \nabla_{\boldsymbol{\theta}} \boldsymbol{\zeta} - \frac{1}{2} (\gamma - \gamma^*)^\top D_{\boldsymbol{\epsilon}}^2 (\gamma - \gamma^*) + (\varkappa - \varkappa^*) \breve{\nabla}_\varkappa \boldsymbol{\zeta} - \frac{1}{2} (\varkappa - \varkappa^*)^2 d_{\boldsymbol{\epsilon}}^2,$$

где $\breve{\nabla}_\varkappa = -D_0^{-2} A \nabla_{\boldsymbol{\theta}}$, $d_{\boldsymbol{\epsilon}}^2 = (1-\delta) d_0^2$, $d_0^2 = H_0^2 - A^\top D_0^{-2} A$ и $D_{\boldsymbol{\epsilon}}^2 = (1-\delta) D_0^2$. Обозначим $\breve{\mathcal{D}}_{\boldsymbol{\epsilon}}^2$ диагонализированную матрицу $\mathcal{D}_{\boldsymbol{\epsilon}}^2$. Переобозначив $\boldsymbol{\nu} = (\gamma, \varkappa)$, можно переписать $\mathbb{L}_{\boldsymbol{\epsilon}}(\boldsymbol{\nu}, \boldsymbol{\nu}^\star)$ в более компактном виде:

$$\mathbb{L}_{\boldsymbol{\epsilon}}(\boldsymbol{\nu}, \boldsymbol{\nu}^\star) = (\boldsymbol{\nu} - \boldsymbol{\nu}^\star)^\top \breve{\mathcal{D}}_{\boldsymbol{\epsilon}} \boldsymbol{\xi}_{\boldsymbol{\epsilon}} - \frac{1}{2} (\boldsymbol{\nu} - \boldsymbol{\nu}^\star)^\top \breve{\mathcal{D}}_{\boldsymbol{\epsilon}}^2 (\boldsymbol{\nu} - \boldsymbol{\nu}^\star), \tag{C.3}$$

где $\boldsymbol{\xi}_{\boldsymbol{\epsilon}} = \breve{\mathcal{D}}_{\boldsymbol{\epsilon}}^{-1} \nabla \boldsymbol{\zeta}$ и $\xi^* = d_{\boldsymbol{\epsilon}}^{-1} \breve{\nabla}_\varkappa$. Аналогично для $\underline{\boldsymbol{\epsilon}} = -\boldsymbol{\epsilon}$.

Далее покажем, что распределение $d_{\boldsymbol{\epsilon}}(\varkappa - \varkappa_{\boldsymbol{\epsilon}})$ при данных $\boldsymbol{Y}$ и при условии, что $\varkappa \in \mathcal{K}_1 \cup \mathcal{K}_\circ(\mathtt{r}_0)$, примерно стандартное нормальное, где

$$\varkappa_{\boldsymbol{\epsilon}} = \varkappa^* + d_{\boldsymbol{\epsilon}}^{-1} \xi^*.$$

Вместе с разложением $d_{\boldsymbol{\epsilon}}(\tilde{\varkappa} - \varkappa^*) - \xi^* \approx 0$ это показывает, что $d_{\boldsymbol{\epsilon}}(\tilde{\varkappa} - \varkappa_{\boldsymbol{\epsilon}}) \approx 0$, и теорема утверждает, что распределение $d_{\boldsymbol{\epsilon}}(\varkappa - \tilde{\varkappa})$ примерно стандартное нормальное при условии данных $\boldsymbol{Y}$ и при условии $\varkappa \in \mathcal{K}_1 \cup \mathcal{K}_\circ(\mathtt{r}_0)$.

Для неотрицательной функции f локальный брэкетинг (C.1) даёт

$$\int\limits_{\Upsilon_\circ(\mathtt{r}_0)} \exp\{\mathcal{L}(\boldsymbol{\nu}, \boldsymbol{\nu}^\star)\} f\bigl(d_{\boldsymbol{\epsilon}}(\varkappa - \varkappa_{\boldsymbol{\epsilon}})\bigr)\, d\boldsymbol{\nu} \leqslant \int\limits_{\Upsilon_\circ(\mathtt{r}_0)} \exp\{\mathbb{L}_{\boldsymbol{\epsilon}}(\boldsymbol{\nu}, \boldsymbol{\nu}^\star)\} f\bigl(d_{\boldsymbol{\epsilon}}(\varkappa - \varkappa_{\boldsymbol{\epsilon}})\bigr)\, d\boldsymbol{\nu}.$$

Аналогично для $\varkappa_{\underline{\epsilon}} = \varkappa^* + d_{\underline{\epsilon}}^{-1}\xi_{\underline{\epsilon}}$

$$\int\limits_{\Upsilon_\circ(\mathtt{r}_0)} \exp\{\mathcal{L}(\boldsymbol{\nu},\boldsymbol{\nu}^\star)\} f\bigl(d_{\underline{\epsilon}}(\varkappa-\varkappa_{\underline{\epsilon}})\bigr)\, d\boldsymbol{\nu} \geqslant \int\limits_{\Upsilon_\circ(\mathtt{r}_0)} \exp\{\mathbb{L}_{\underline{\epsilon}}(\boldsymbol{\nu},\boldsymbol{\nu}^\star)\} f\bigl(d_{\underline{\epsilon}}(\varkappa-\varkappa_{\underline{\epsilon}})\bigr)\, d\boldsymbol{\nu}.$$

Основной смысл данного брэкетинга в том, что $\mathbb{L}_{\boldsymbol{\epsilon}}(\boldsymbol{\nu},\boldsymbol{\nu}^\star)$ и $\mathbb{L}_{\underline{\epsilon}}(\boldsymbol{\nu},\boldsymbol{\nu}^\star)$ — квадратичные функции по $\boldsymbol{\nu}$ и, значит, соответствуют некоторым гауссовским распределениям.

Напомним, что $\Omega(\mathtt{x})$ обозначает случайное событие с вероятностью $\mathbb{P}\bigl(\Omega(\mathtt{x})\bigr) \geqslant 1 - \mathtt{C}\mathrm{e}^{-x}$.

Теорема C.1. *Для любого $\delta \geqslant \delta(\mathtt{r}_0)$ при условии $(\boldsymbol{\mathcal{L}_0})$ на $\mathcal{K}_\circ$ для неотрицательной функции $f(\cdot)$ на $\mathbb{R}$ на множестве доминирующей вероятности $\Omega(\mathtt{x})$ выполняется*

$$\mathbb{E}\bigl[f\bigl(d_{\boldsymbol{\epsilon}}(\varkappa-\varkappa_{\boldsymbol{\epsilon}})\bigr)\mathbb{I}\{\varkappa\in\mathcal{K}_\circ(\mathtt{r}_0)\}\mid \boldsymbol{Y}\bigr] \leqslant \exp\{\Delta_{\boldsymbol{\epsilon}}^+(\mathtt{r}_0)\}\,\mathbb{E}f(\boldsymbol{\gamma}), \tag{C.4}$$

где

$$\Delta_{\boldsymbol{\epsilon}}^+(\mathtt{r}_0) = \Delta_{\boldsymbol{\epsilon}}(\mathtt{r}_0) + \frac{p+1}{2}\log\frac{1+\delta}{1-\delta} + \tau_{\underline{\epsilon}}(\mathtt{r}_0)$$

с $\Delta_{\boldsymbol{\epsilon}}(\mathtt{r}_0) \stackrel{\mathrm{def}}{=} \|\boldsymbol{\xi}_{\boldsymbol{\epsilon}}\|^2/2 - \|\boldsymbol{\xi}_{\underline{\epsilon}}\|^2/2$, и для $\gamma \sim \mathcal{N}(0,1)$

$$\tau_{\underline{\epsilon}}(\mathtt{r}_0) \stackrel{\mathrm{def}}{=} -\log\mathbb{P}(|\gamma+\xi_{\underline{\epsilon}}| \leqslant \mathtt{r}_0 \mid \boldsymbol{Y}). \tag{C.5}$$

Более того, на $\Omega(\mathtt{x})$ выполняется $\Delta_{\boldsymbol{\epsilon}}^+(\mathtt{r}_0) \leqslant \mathtt{C}\,\delta\,(p+1+\mathtt{x})$.

Доказательство. Определим

$$m_{\boldsymbol{\epsilon}}(\boldsymbol{\xi}_{\boldsymbol{\epsilon}}) \stackrel{\mathrm{def}}{=} -\frac{\|\boldsymbol{\xi}_{\boldsymbol{\epsilon}}\|^2}{2} + \log(\det\breve{\mathcal{D}}_{\boldsymbol{\epsilon}}) - (p+1)\log(\sqrt{2\pi}). \tag{C.6}$$

Тогда с учётом (C.3) имеем

$$m_{\boldsymbol{\epsilon}}(\boldsymbol{\xi}_{\boldsymbol{\epsilon}})+\mathbb{L}_{\boldsymbol{\epsilon}}(\boldsymbol{\nu},\boldsymbol{\nu}^\star) = -\frac{\|\breve{\mathcal{D}}_{\boldsymbol{\epsilon}}\{\boldsymbol{\nu}-\boldsymbol{\nu}^\star\}-\boldsymbol{\xi}_{\boldsymbol{\epsilon}}\|^2}{2} + \log(\det\breve{\mathcal{D}}_{\boldsymbol{\epsilon}}) - (p+1)\log(\sqrt{2\pi}), \tag{C.7}$$

что есть логарифмическая плотность нормального распределения со средним $\boldsymbol{\nu}_{\boldsymbol{\epsilon}} = \breve{\mathcal{D}}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{\xi}_{\boldsymbol{\epsilon}} + \boldsymbol{\nu}^\star$ и матрицей ковариации $\breve{\mathcal{D}}_{\boldsymbol{\epsilon}}^{-2}$. Используя замену переменных $u = d_{\boldsymbol{\epsilon}}(\varkappa-\varkappa^*) - \xi^*$ и (C.7), получаем для любой неотрицательной функции f

$$\begin{gathered}\int\limits_{\Upsilon_\circ} \exp\{\mathcal{L}(\boldsymbol{\nu},\boldsymbol{\nu}^\star)+m_{\boldsymbol{\epsilon}}(\boldsymbol{\xi}_{\boldsymbol{\epsilon}})\} f\bigl(d_{\boldsymbol{\epsilon}}(\varkappa-\varkappa_{\boldsymbol{\epsilon}})\bigr)\, d\boldsymbol{\nu} \leqslant \\ \leqslant \int \exp\{\mathbb{L}_{\boldsymbol{\epsilon}}(\boldsymbol{\nu},\boldsymbol{\nu}^\star)+m_{\boldsymbol{\epsilon}}(\boldsymbol{\xi}_{\boldsymbol{\epsilon}})\} f\bigl(d_{\boldsymbol{\epsilon}}(\varkappa-\varkappa_{\boldsymbol{\epsilon}})\bigr)\, d\boldsymbol{\nu} = \int \phi(u) f(u)\, du = \mathbb{E}f(\boldsymbol{\gamma}).\end{gathered} \tag{C.8}$$

Аналогично если $m_{\underline{\epsilon}}(\boldsymbol{\xi}_{\underline{\epsilon}})$ определена в (C.6) с $\underline{\boldsymbol{\epsilon}}$ вместо $\boldsymbol{\epsilon}$, то значение $m_{\underline{\epsilon}}(\boldsymbol{\xi}_{\underline{\epsilon}}) + \mathbb{L}_{\underline{\epsilon}}(\boldsymbol{\nu},\boldsymbol{\nu}^\star)$ при условии данных есть плотность нормального закона со средним $\boldsymbol{\nu}_{\underline{\epsilon}} = \breve{\mathcal{D}}_{\underline{\epsilon}}^{-1}\boldsymbol{\xi}_{\underline{\epsilon}} + \boldsymbol{\nu}^{\star}$ и ковариационной матрицей $\breve{\mathcal{D}}_{\underline{\epsilon}}^{-2}$. Для любой неотрицательной функции f

$$\int\limits_{\Upsilon_{\circ}(\mathtt{r}_0)} \exp\{\mathcal{L}(\boldsymbol{\nu},\boldsymbol{\nu}^{\star})\} f\bigl(d_{\underline{\epsilon}}(\varkappa - \varkappa_{\underline{\epsilon}})\bigr)\, d\boldsymbol{\nu} \geqslant \\ \geqslant \exp\{-m_{\underline{\epsilon}}(\boldsymbol{\xi}_{\underline{\epsilon}})\} \int \phi(u) f(u) \mathbb{I}\{d_{\underline{\epsilon}}^{-1}(u + \xi_{\underline{\epsilon}}) \in \mathcal{K}_{\circ}(\mathtt{r}_0)\}\, du. \quad (\mathrm{C}.9)$$

Используя $d_{\underline{\epsilon}}^2 \geqslant d_0^2$ и $|\xi_{\underline{\epsilon}}| \leqslant |\xi|$, получаем

$$\{d_{\underline{\epsilon}}^{-1}(u + \xi_{\underline{\epsilon}}) \in \mathcal{K}_{\circ}(\mathtt{r}_0)\} = \{|d_0 d_{\underline{\epsilon}}^{-1}(u + \xi_{\underline{\epsilon}})| \leqslant \mathtt{r}_0\} \supset \{|u + \xi| \leqslant \mathtt{r}_0\}.$$

Специальный случай (C.9) при $f(u) \equiv 1$ по определению $\tau_{\underline{\epsilon}}$ даёт

$$\int\limits_{\Upsilon_{\circ}} \exp\{\mathcal{L}(\boldsymbol{\nu},\boldsymbol{\nu}^{\star})\}\, d\boldsymbol{\nu} \geqslant \exp\{-m_{\underline{\epsilon}}(\boldsymbol{\xi}_{\underline{\epsilon}}) - \tau_{\underline{\epsilon}}\}. \quad (\mathrm{C}.10)$$

При (C.8) и (C.10) получаем

$$\frac{\int\limits_{\Upsilon_{\circ}(\mathtt{r}_0)} \exp\{\mathcal{L}(\boldsymbol{\nu},\boldsymbol{\nu}^{\star})\} f\bigl(d_{\epsilon}(\varkappa - \varkappa_{\epsilon}^{*})\bigr)\, d\boldsymbol{\nu}}{\int \exp\{\mathcal{L}(\boldsymbol{\nu},\boldsymbol{\nu}^{\star})\}\, d\boldsymbol{\nu}} \leqslant \exp\{m_{\underline{\epsilon}}(\boldsymbol{\xi}_{\underline{\epsilon}}) - m_{\epsilon}(\boldsymbol{\xi}_{\epsilon}) + \tau_{\underline{\epsilon}}(\mathtt{r}_0)\}\, \mathbb{E} f(\gamma) = \\ = \exp\left\{\Delta_{\epsilon}(\mathtt{r}_0) + \frac{p+1}{2}\log\frac{1+\delta}{1-\delta} + \tau_{\underline{\epsilon}}(\mathtt{r}_0)\right\} \mathbb{E} f(\gamma),$$

что доказывает (C.4). □

Далее покажем, что члены $\Delta_{\epsilon}(\mathtt{r}_0)$, $(p+1)\log\bigl((1+\delta)/(1-\delta)\bigr)$ и $\tau_{\epsilon}(\mathtt{r}_0)$ малы с большой вероятностью при выполнении локального условия. Важно, что если забыть о них, приравнять матрицы $\breve{\mathcal{D}}_{\epsilon}$ и $\breve{\mathcal{D}}_{\underline{\epsilon}}$ и векторы $\boldsymbol{\xi}_{\epsilon}$ и $\boldsymbol{\xi}_{\underline{\epsilon}}$, то теорема говорит о примерно стандартном нормальном распределении для $\breve{\mathcal{D}}_{\epsilon}(\boldsymbol{\nu} - \boldsymbol{\nu}^{\star}) - \boldsymbol{\xi}_{\epsilon}$ при условии данных $\boldsymbol{Y}$. Значение

$$\varkappa_{\epsilon} \stackrel{\mathrm{def}}{=} (p+1)\log\frac{1+\delta}{1-\delta}$$

может быть ограничено $1{,}1(p+1)\delta$, если $\delta \leqslant 0{,}5$. По определению $d_{\underline{\epsilon}}^2 d_0^{-2} \leqslant 1 + \delta$. Тогда для $\boldsymbol{\gamma} \sim \mathcal{N}(0,1)$ и $\mathtt{r}_0^2 \geqslant \mathtt{C}$

$$\tau_{\underline{\epsilon}}(\mathtt{r}_0) = -\log \mathbb{P}(|d_0 d_{\underline{\epsilon}}^{-1}(\gamma + \xi_{\underline{\epsilon}})| \leqslant \mathtt{r}_0 \mid \boldsymbol{Y}) \leqslant -\log \mathbb{P}(\gamma^2 \leqslant 1 + \mathtt{C}\mathtt{x}) \leqslant \mathrm{e}^{-\mathtt{x}}.$$

В [10] было доказано в чуть более общей формулировке, что $\boldsymbol{\xi}_{\epsilon}^2 \leqslant \mathtt{C}\delta(p + 1 + \mathtt{x})$ на $\Omega(\mathtt{x})$. Суммируя, можно получить

$$\Delta_{\epsilon}^{+}(\mathtt{r}_0) \leqslant \mathtt{C}\delta(p + 1 + \mathtt{x}).$$

Нижняя граница находится с помощью нескольких вспомогательных утверждений о концах апостериорного распределения. В конечном итоге будет использован специальный случай этой границы с $f(u) = \exp(\lambda u)$ при фиксированном векторе $\lambda \in \mathbb{R}$. Удобно ввести локальное условное математическое ожидание: для случайной величины η определим

$$\mathbb{E}^{\circ}\eta \stackrel{\text{def}}{=} \mathbb{E}[\eta\mathbb{I}\{\varkappa \in \mathcal{K}_{\circ}(\mathtt{r}_0)\} \mid \boldsymbol{Y}].$$

Тогда границу (C.4) можно записать как

$$\mathbb{E}^{\circ} f\bigl(\lambda d_{\boldsymbol{\epsilon}}(\varkappa - \varkappa_{\boldsymbol{\epsilon}})\bigr) \leqslant \exp\{\Delta_{\boldsymbol{\epsilon}}^{+}(\mathtt{r}_0)\}\mathbb{E} f(\gamma).$$

В следующем результате рассмотрен специальный случай с $f(u) = \exp(\lambda u)$ и $f(u) = \mathbb{I}(u \in A)$ для измеримого множества $A \subset \mathbb{R}$.

Следствие C.2. *Для любого* $\lambda \in \mathbb{R}$

$$\log \mathbb{E}^{\circ} \exp\{\lambda d_{\boldsymbol{\epsilon}}(\varkappa - \varkappa_{\boldsymbol{\epsilon}})\} \leqslant \frac{\lambda^2}{2} + \Delta_{\boldsymbol{\epsilon}}^{+}(\mathtt{r}_0). \tag{C.11}$$

Более того, для $\lambda^2 \leqslant \mathtt{x}$ *и для* $\breve{\varkappa} = \varkappa^* + d_0^{-2}\breve{\nabla}_{\varkappa}\mathcal{L}(\boldsymbol{\nu}^{\star})$

$$\log \mathbb{E}^{\circ} \exp\{\lambda d_0(\varkappa - \breve{\varkappa})\} \leqslant \frac{\lambda^2}{2} + \Delta_{\boldsymbol{\epsilon}}^{\oplus}(\mathtt{r}_0), \tag{C.12}$$

где

$$\Delta_{\boldsymbol{\epsilon}}^{\oplus}(\mathtt{r}_0) \stackrel{\text{def}}{=} \Delta_{\boldsymbol{\epsilon}}^{+}(\mathtt{r}_0) + \delta + 2\delta|\xi|.$$

На множестве $\Omega(\mathtt{x})$ *выполняется* $\Delta_{\boldsymbol{\epsilon}}^{\oplus}(\mathtt{r}_0) \leqslant \mathtt{C}\,\delta(p + 1 + \mathtt{x})$.

Доказательство. Первое утверждение есть прямое следствие теоремы с учётом того, что

$$\mathbb{E}\exp(\lambda u) = \frac{\lambda^2}{2}$$

для $u \sim \mathcal{N}(0, 1)$.

Для доказательства второго утверждения воспользуемся тем, что

$$\exp\{\lambda d_0(\varkappa - \breve{\varkappa})\} = \exp\{\lambda_1 d_{\boldsymbol{\epsilon}}(\varkappa - \varkappa_{\boldsymbol{\epsilon}})\}\exp\{\lambda d_0(\varkappa_{\boldsymbol{\epsilon}} - \breve{\varkappa})\}, \tag{C.13}$$

где $\lambda_1 = (1 - \delta)^{-1/2}\lambda$. Так как $\lambda^2 \leqslant 1$, имеем

$$\lambda_1^2 = (1 - \delta)^{-1}\lambda^2 \leqslant (1 + 2\delta)\lambda^2 \leqslant \lambda^2 + 2\delta.$$

Осталось ограничить $\lambda d_0(\varkappa_{\boldsymbol{\epsilon}} - \breve{\varkappa})$ для $\breve{\varkappa} = \varkappa^* + d_0^{-2}\breve{\nabla}_{\varkappa}\mathcal{L}(\boldsymbol{\nu}^{\star}) = \varkappa^* + d_0^{-1}\xi$ и $\varkappa_{\boldsymbol{\epsilon}} = \varkappa^* + d_{\boldsymbol{\epsilon}}^{-1}\xi^*$. По определению $d_{\boldsymbol{\epsilon}}$ при $\delta \leqslant 1/2$

$$|d_0(\varkappa_{\boldsymbol{\epsilon}} - \breve{\varkappa})| = |d_0 d_{\boldsymbol{\epsilon}}^{-1}\xi^* - \xi| = |(d_0^2 d_{\boldsymbol{\epsilon}}^{-2} - 1)\xi| \leqslant \left(\frac{1}{1 - \delta} - 1\right)|\xi| \leqslant 2\delta|\xi|. \tag{C.14}$$

Это вместе с (C.11) и (C.13) в случае $\lambda^2 \leqslant 1$ доказывает второе утверждение. $\square$

Аналогично предыдущему результату условие $\delta \to 0$ влечёт малость $\Delta_{\boldsymbol{\epsilon}}^{+}(\mathtt{r}_0)$ в (C.11) и $\Delta_{\boldsymbol{\epsilon}}^{\oplus}(\mathtt{r}_0)$ в (C.12).

Следствие C.3. *Для любого измеримого множества* $A \subset \mathbb{R}$ *на множестве* $\Omega_{\mathtt{x}}$ *с* $\boldsymbol{\delta}_{\boldsymbol{\epsilon}} \stackrel{\text{def}}{=} d_0(\varkappa_{\boldsymbol{\epsilon}} - \breve{\varkappa})$, *где* $\breve{\varkappa} = \varkappa^* + d_0^{-2}\breve{\nabla}_{\varkappa}\mathcal{L}(\boldsymbol{\nu}^{\star})$ *и* $\mathtt{q} = 1 + |\xi|^2$, *выполняется*

$$\mathbb{P}^{\circ}(d_0(\varkappa - \breve{\varkappa}) \in A) \leqslant \exp\{\Delta_{\boldsymbol{\epsilon}}^{+}(\mathtt{r}_0)\}\mathbb{P}(d_0 d_{\boldsymbol{\epsilon}}^{-1}\gamma + \boldsymbol{\delta}_{\boldsymbol{\epsilon}} \in A) \tag{C.15}$$

$$\leqslant \exp\{\Delta_{\boldsymbol{\epsilon}}^{+}(\mathtt{r}_0)\}\bigl\{\mathbb{P}(\gamma \in A) + \mathtt{C}\delta\mathtt{q}^{1/2}\bigr\}. \tag{C.16}$$

Доказательство. Первое утверждение (C.15) следует из теоремы при $f(u) = \mathbb{I}(d_0 d_{\boldsymbol{\epsilon}}^{-1} u + \boldsymbol{\delta}_{\boldsymbol{\epsilon}} \in A)$. По определению при $\delta \leqslant 1/2$

$$\boldsymbol{\delta}_{\boldsymbol{\epsilon}} \leqslant 2\delta|\xi|. \tag{C.17}$$

Для доказательства (C.16) используем неравенство Пинскера для расстояния Кульбака—Лейблера между двумя нормальными распределениями. Пусть γ — стандартная нормальная случайная величина на $\mathbb{R}$ из распределения $\mathbb{P}_0$.

Случайная величина $d_0 d_{\boldsymbol{\epsilon}}^{-1}\gamma + \boldsymbol{\delta}_{\boldsymbol{\epsilon}}$ нормальная со средним $\boldsymbol{\delta}_{\boldsymbol{\epsilon}}$ и дисперсией $b_{\boldsymbol{\epsilon}}^{-1} \stackrel{\mathrm{def}}{=} d_0^2 d_{\boldsymbol{\epsilon}}^{-2}$. Обозначим это распределение $\mathbb{P}_{\boldsymbol{\epsilon}}$. Очевидно, что $b_{\boldsymbol{\epsilon}} = d_0^{-2} d_{\boldsymbol{\epsilon}}^2 = (1-\delta)$. Используем следующую техническую лемму.

Лемма C.4. *Пусть* $|b_{\boldsymbol{\epsilon}} - 1| \leqslant \delta \leqslant 1/2$. *Тогда*

$$2\mathcal{K}(\mathbb{P}_0, \mathbb{P}_{\boldsymbol{\epsilon}}) = -2\mathbb{E}_0 \log \frac{d\mathbb{P}_{\boldsymbol{\epsilon}}}{d\mathbb{P}_0}(\boldsymbol{\gamma}) \leqslant (b_{\boldsymbol{\epsilon}} - 1)^2 + (1+\delta)|\boldsymbol{\delta}_{\boldsymbol{\epsilon}}|^2 = \delta^2 + (1+\delta)|\boldsymbol{\delta}_{\boldsymbol{\epsilon}}|^2.$$

По определению

$$2 \log \frac{d\mathbb{P}_{\boldsymbol{\epsilon}}}{d\mathbb{P}_0}(\gamma) = -\log b_{\boldsymbol{\epsilon}} - (\gamma - \boldsymbol{\delta}_{\boldsymbol{\epsilon}})^2 b_{\boldsymbol{\epsilon}} + |\gamma|^2.$$

Далее,

$$2\mathcal{K}(\mathbb{P}_0, \mathbb{P}_{\boldsymbol{\epsilon}}) = -2\mathbb{E}_0 \log \frac{d\mathbb{P}_{\boldsymbol{\epsilon}}}{d\mathbb{P}_0}(\gamma) = \log(b_{\boldsymbol{\epsilon}}) + (b_{\boldsymbol{\epsilon}} - 1)^2 + \boldsymbol{\delta}_{\boldsymbol{\epsilon}}^2 b_{\boldsymbol{\epsilon}} \leqslant \delta^2 + (1+\delta)|\boldsymbol{\delta}_{\boldsymbol{\epsilon}}|^2. \ \square$$

Используя данную лемму, (C.17) и неравенство Пинскера, получаем

$$\|\mathbb{P}_0 - \mathbb{P}_{\boldsymbol{\epsilon}}\|_{TV} \leqslant \sqrt{\frac{1}{2}\mathcal{K}(\mathbb{P}_0, \mathbb{P}_{\boldsymbol{\epsilon}})} \leqslant \mathtt{C}\delta \sqrt{1 + |\xi|^2} \leqslant \mathtt{C}\delta\, \mathtt{q}^{1/2},$$

где $\mathtt{q} = 1 + |\xi|^2$. Эквивалентно для любого измеримого множества A

$$\mathbb{P}(d_0 d_{\boldsymbol{\epsilon}}^{-1}\gamma + \boldsymbol{\delta}_{\boldsymbol{\epsilon}} \in A \mid \boldsymbol{Y}) \leqslant \mathbb{P}(\gamma \in A) + \mathtt{C}\boldsymbol{\epsilon}\, \mathtt{q}^{1/2}.$$

Следующее утверждение описывает локальные концентрационные свойства апостериорного распределения. А именно, центрированный и скалированный вектор $\boldsymbol{\eta} \stackrel{\mathrm{def}}{=} d_{\boldsymbol{\epsilon}}(\varkappa - \varkappa_{\boldsymbol{\epsilon}})$ концентрируется на множестве $\bigl\{u\colon |u^2 - 1| \leqslant \sqrt{2\mathtt{x}}\bigr\}$ с вероятностью $\mathbb{P}^{\circ}$ порядка $1 - 2\mathrm{e}^{-\mathtt{x}/4}$.

Следствие C.5. *Пусть* $\mathtt{x} \leqslant 1/2$. *Тогда*

$$\mathbb{P}^{\circ}\bigl(|d_{\boldsymbol{\epsilon}}(\varkappa - \varkappa_{\boldsymbol{\epsilon}})|^2 - 1 > \sqrt{2\mathtt{x}}\bigr) \leqslant \exp\Bigl(-\frac{\mathtt{x}}{4} + \Delta_{\boldsymbol{\epsilon}}^{+}(\mathtt{r}_0)\Bigr), \tag{C.18}$$

$$\mathbb{P}^{\circ}\bigl(|d_{\boldsymbol{\epsilon}}(\varkappa - \varkappa_{\boldsymbol{\epsilon}})|^2 - 1 > -\sqrt{2\mathtt{x}}\bigr) \leqslant \exp\Bigl(-\frac{\mathtt{x}}{2} + \Delta_{\boldsymbol{\epsilon}}^{+}(\mathtt{r}_0)\Bigr). \tag{C.19}$$

Доказательство. Обозначим $\boldsymbol{\eta}_{\boldsymbol{\epsilon}} \stackrel{\mathrm{def}}{=} d_{\boldsymbol{\epsilon}}(\varkappa - \varkappa_{\boldsymbol{\epsilon}})$. Используя экспоненциальное неравенство Чебышёва для положительного $\lambda \leqslant 1/2$, с учётом того, что

$-\log(1-\lambda) \leqslant \lambda + \lambda^2$, получаем

$$\begin{aligned}\mathbb{P}^{\circ}\bigl(\boldsymbol{\eta}_{\boldsymbol{\epsilon}}^2 > 1 + \sqrt{2\mathtt{x}}\bigr) &\leqslant \exp\left\{-\frac{\lambda}{2}\bigl(1+\sqrt{2\mathtt{x}}\bigr)\right\}\mathbb{E}^{\circ}\exp\left(\frac{\lambda}{2}\boldsymbol{\eta}_{\boldsymbol{\epsilon}}^2\right) \leqslant \\ &\leqslant \exp\left\{-\frac{\lambda}{2}\bigl(1+\sqrt{2\mathtt{x}}\bigr) + \Delta_{\boldsymbol{\epsilon}}^{+}(\mathtt{r}_0)\right\}\mathbb{E}\exp\left(\frac{\lambda}{2}\gamma^2\right) = \\ &= \exp\left\{-\frac{\lambda}{2}\bigl(1+\sqrt{2\mathtt{x}}\bigr) - \frac{1}{2}\log(1-\lambda) + \Delta_{\boldsymbol{\epsilon}}^{+}(\mathtt{r}_0)\right\} \leqslant \\ &\leqslant \exp\left\{-\lambda\sqrt{\frac{\mathtt{x}}{2}} + \frac{\lambda^2}{2} + \Delta_{\boldsymbol{\epsilon}}^{+}(\mathtt{r}_0)\right\}.\end{aligned}$$

Выбор $\lambda = \sqrt{\mathtt{x}/2}$ даёт (C.18). Аналогично

$$\begin{aligned}\mathbb{P}\bigl(\boldsymbol{\eta}_{\boldsymbol{\epsilon}}^2 - 1 < -\sqrt{2\mathtt{x}}\bigr) &\leqslant \exp\left\{-\frac{\lambda}{2}\bigl(\sqrt{2\mathtt{x}}-1\bigr)\right\}\mathbb{E}\exp\left(-\frac{\lambda}{2}\boldsymbol{\eta}_{\boldsymbol{\epsilon}}^2\right) \leqslant \\ &\leqslant \exp\left\{-\frac{\lambda}{2}\bigl(\sqrt{2\mathtt{x}}-1\bigr) + \Delta_{\boldsymbol{\epsilon}}^{+}(\mathtt{r}_0)\right\}\mathbb{E}\exp\left(-\frac{\lambda}{2}\gamma^2\right) = \\ &= \exp\left\{-\frac{\lambda}{2}\bigl(\sqrt{2\mathtt{x}}-1\bigr) - \frac{1}{2}\log(1+\lambda) + \Delta_{\boldsymbol{\epsilon}}^{+}(\mathtt{r}_0)\right\} \leqslant \\ &\leqslant \exp\left\{-\lambda\sqrt{\frac{\mathtt{x}}{2}} + \frac{\lambda^2}{4} + \Delta_{\boldsymbol{\epsilon}}^{+}(\mathtt{r}_0)\right\}.\end{aligned}$$

В данном случае выбор $\lambda = \sqrt{2\mathtt{x}}$ влечёт результат. □

D. «Хвосты» апостериорного распределения

Следующий основной шаг в анализе — показать, что апостериорное распределение концентрируется в локальной окрестности истинной точки.

Напомним, что мы ввели множества $\mathcal{K}_1 = \{\varkappa \in \mathbb{R}_+ \colon \varkappa \leqslant \varkappa_1\}$ и $\mathcal{K}_2 = \{\varkappa \in \mathbb{R}_+ \colon \varkappa \geqslant \varkappa_2\}$. Рассмотрим случайную величину

$$\rho(\mathtt{r}_0) \stackrel{\mathrm{def}}{=} \frac{\int_{\Theta}\int_{\mathcal{K}_1\cup\mathcal{K}_2} \exp\{\mathcal{L}(\boldsymbol{\nu},\boldsymbol{\nu}^{\star})\}\,d\boldsymbol{\nu}}{\int_{\Upsilon_\circ}\exp\{\mathcal{L}(\boldsymbol{\nu},\boldsymbol{\nu}^{\star})\}\,d\boldsymbol{\nu}} = \sum_{i=1}^{2}\frac{\int_{\Upsilon_i}\exp\{\mathcal{L}(\boldsymbol{\nu},\boldsymbol{\nu}^{\star})\}\,d\boldsymbol{\nu}}{\int_{\Upsilon_\circ}\exp\{\mathcal{L}(\boldsymbol{\nu},\boldsymbol{\nu}^{\star})\}\,d\boldsymbol{\nu}} = \rho_1(\mathtt{r}_0) + \rho_2(\mathtt{r}_0), \tag{D.1}$$

где $\Upsilon_i = \Theta \times \mathcal{K}_i$. Ясно, что $\mathbb{P}\{\varkappa \notin \mathcal{K}_\circ \mid \boldsymbol{Y}\} \leqslant \rho(\mathtt{r}_0)$. Представляется естественным оценить отдельно $\rho_1(\mathtt{r}_0)$ и $\rho_2(\mathtt{r}_0)$ из-за несимметричности процесса по $\varkappa$. Используем (A.4) и (A.5) для доказательства следующего утверждения.

Теорема D.1. *Пусть*

$$6\sqrt{\mathtt{x} + c_1(p+1)} \leqslant \mathtt{r}_0(1+\varepsilon)$$

с некоторой константой c_1, *такой что* $\mathtt{x} + c_1 p \geqslant 2{,}5$. *Тогда на множестве* $\Omega(\mathtt{x})$ *выполняется*

$$\rho_1(\mathtt{r}_0) \leqslant \exp\left(\tau_{\underline{\epsilon}}(\mathtt{r}_0) + \frac{\sigma^2}{2}\|\nabla_{\boldsymbol{\theta}}\boldsymbol{\zeta}\|^2\right)(1+\delta)^{p+1}\frac{d_0}{\sqrt{2\pi}}\beta(\mathtt{r}_0), \tag{D.2}$$

где $\tau_{\underline{\epsilon}}(\mathtt{r}_0)$ *из* (C.5) *и*

$$\beta(\mathtt{r}_0) = \frac{2}{(p+2)}\left(\frac{\varkappa_1}{\varkappa^*}\right)^{p/2}\varkappa_1.$$

Доказательство. Для интеграла в числителе в (D.1) для любой неотрицательной функции f на множестве $\Omega(\mathtt{x})$ выполняется

$$\begin{aligned}
&\int\limits_{\Upsilon_1} \exp\bigl\{\mathcal{L}\bigl((\boldsymbol{\theta},\varkappa),(\boldsymbol{\theta}^*,\varkappa^*)\bigr)\bigr\} f(\varkappa)\, d\boldsymbol{\theta}\, d\varkappa = \\
&\quad = \int\limits_{\mathcal{K}_1} \exp\{\mathcal{L}(\boldsymbol{\nu}^\star_\varkappa,\boldsymbol{\nu}^\star)\} f(\varkappa) \int\limits_{\Theta} \exp\{\mathcal{L}(\boldsymbol{\nu},\boldsymbol{\nu}^\star_\varkappa)\}\, d\boldsymbol{\theta}\, d\varkappa = \\
&\quad = \int\limits_{\mathcal{K}_1} \exp\{\mathcal{L}(\boldsymbol{\nu}^\star_\varkappa,\boldsymbol{\nu}^\star)\} f(\varkappa) \exp\left(-\log(\det D_\varkappa) + p\log(\sqrt{2\pi}) + \frac{1}{2}\|D_\varkappa^{-1}\nabla_{\boldsymbol{\theta}}\boldsymbol{\zeta}\|^2\right) d\varkappa,
\end{aligned} \tag{D.3}$$

Мы использовали, что

$$\nabla_{\boldsymbol{\theta}}\mathcal{L} = \nabla_{\boldsymbol{\theta}}\boldsymbol{\zeta} + \nabla_{\boldsymbol{\theta}}\mathbb{E}\mathcal{L} = \nabla_{\boldsymbol{\theta}}\boldsymbol{\zeta}$$

в точке $(\boldsymbol{\theta}^*_\varkappa, \varkappa)$. Между двумя $\varkappa$-моделями используется стандартное разложение

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\nu}^\star_\varkappa,\boldsymbol{\nu}^\star) &= \boldsymbol{\zeta}(\boldsymbol{\nu}^\star_\varkappa,\boldsymbol{\nu}^\star) + \mathbb{E}\mathcal{L}(\boldsymbol{\nu}^\star_\varkappa,\boldsymbol{\nu}^\star) = \\
&= (1+\varepsilon)(\mathbb{E}\mathcal{X}_{\varkappa^*} - \mathbb{E}\mathcal{X}_\varkappa) - b_\varkappa + b_{\varkappa^*} + c\log\frac{\varkappa}{\varkappa^*} + \frac{\boldsymbol{\varepsilon}^\top}{\sigma^2}(\boldsymbol{\theta}^*_\varkappa - \boldsymbol{\theta}^*).
\end{aligned}$$

Используем следствие 4.2 из [10] для того, чтобы подавить стохастическую компоненту квадратичной функцией $-b\|\boldsymbol{\theta}^*_\varkappa - \boldsymbol{\theta}^*\|^2$, где константа b должна удовлетворять условию $6\sqrt{\mathtt{x} + c_1(p+1)} \leqslant \mathtt{r}_0 b$.

Заметим, что из $\mathcal{R}_\varkappa \geqslant \mathcal{R}_{\varkappa^*}$ следует, что

$$\mathbb{E}\mathcal{X}_{\varkappa^*} - \mathbb{E}\mathcal{X}_\varkappa \leqslant b_\varkappa - b_{\varkappa^*},$$

и по определению смещения $b_\varkappa$

$$\begin{aligned}
b_\varkappa - b_{\varkappa^*} &= -\mathbb{E}L_\varkappa(\boldsymbol{\theta}^*_\varkappa) + \mathbb{E}L_{\varkappa^*}(\boldsymbol{\theta}^*_{\varkappa^*}) \leqslant \\
&\leqslant -\mathbb{E}L_\varkappa(\boldsymbol{\theta}^*_\varkappa) + \mathbb{E}L_\varkappa(\boldsymbol{\theta}^*_{\varkappa^*}) = -\|D_\varkappa(\boldsymbol{\theta}^*_\varkappa - \boldsymbol{\theta}^*_{\varkappa^*})\|^2.
\end{aligned}$$

Используя это и следствие 4.2 из [10] при условии $6\sqrt{x + c_1(p+1)} \leqslant \mathtt{r}_0(1+\varepsilon)$, получаем, что можно выбрать

$$\mathfrak{u}(\boldsymbol{\nu}) = \frac{(1+\varepsilon)\|(\boldsymbol{\theta}^*_\varkappa - \boldsymbol{\theta}^*_{\varkappa^*})\|^2}{2}.$$

Тогда при $c = 0$

$$\mathbb{E}\mathcal{L}(\boldsymbol{\nu}^\star_\varkappa, \boldsymbol{\nu}^\star) + \mathfrak{u}(\boldsymbol{\nu}^\star_\varkappa) \leqslant -\frac{(1+\varepsilon)\|(\boldsymbol{\theta}^*_\varkappa - \boldsymbol{\theta}^*_{\varkappa^*})\|^2}{2}.$$

Далее используем следствие, в котором вместо $\mathtt{b}(\mathtt{r})$ стоит $(1+\varepsilon)/2$. Этот результат показывает, что $\mathcal{L}(\boldsymbol{\nu}, \boldsymbol{\nu}^\star)$ можно ограничить $\mathfrak{u}(\boldsymbol{\nu})$ при $c = 0$ с большой вероятностью.

Соотношение (C.10) для локального интеграла

$$\int\limits_{\Upsilon_\circ} \exp\{\mathcal{L}(\boldsymbol{\nu}, \boldsymbol{\nu}^\star)\}\, d\boldsymbol{\nu}$$

даёт

$$\rho_1(\mathtt{r}_0) \leqslant \exp\{\tau_{\underline{\boldsymbol{\epsilon}}}(\mathtt{r}_0) + m_{\underline{\boldsymbol{\epsilon}}}(\boldsymbol{\xi}_{\underline{\boldsymbol{\epsilon}}})\} \times \\ \times \int\limits_{\mathcal{K}_1} \exp\left(c \log \frac{\varkappa}{\varkappa^*} - \log\det D_\varkappa + p \log\sqrt{2\pi} + \frac{1}{2}\|D_\varkappa^{-1}\nabla_{\boldsymbol{\theta}}\boldsymbol{\zeta}\|^2\right) f(\varkappa)\, d\varkappa.$$

Далее, используя

$$\exp\{m_{\underline{\boldsymbol{\epsilon}}}(\boldsymbol{\xi}_{\underline{\boldsymbol{\epsilon}}})\} = \exp\left(-\frac{\|\boldsymbol{\xi}_{\underline{\boldsymbol{\epsilon}}}\|^2}{2}\right)(2\pi)^{-(p+1)/2}\det(\breve{\mathcal{D}}_{\underline{\boldsymbol{\epsilon}}}) \leqslant (2\pi)^{-(p+1)/2}\det(\breve{\mathcal{D}}_{\underline{\boldsymbol{\epsilon}}}),$$

получаем

$$\rho_1(\mathtt{r}_0) \leqslant \exp\left(\tau_{\underline{\boldsymbol{\epsilon}}}(\mathtt{r}_0) + \frac{\sigma^2}{2}\|\nabla_{\boldsymbol{\theta}}\boldsymbol{\zeta}\|^2\right) \times \\ \times (1+\delta)^{p+1} d_0 \int\limits_{\mathcal{K}_1} \exp\left(c\log\frac{\varkappa}{\varkappa^*} + \log\left(\frac{\det D_{\varkappa^*}}{\det D_\varkappa}\right)\right) f(\varkappa)\, d\varkappa.$$

Для последнего члена в интеграле имеем

$$\log\left(\frac{\det D_{\varkappa^*}}{\det D_\varkappa}\right) = \frac{1}{2}\sum_{i=1}^{p}\log\left(\frac{1+\varkappa^*\sigma^2 g_i^2}{1+\varkappa\sigma^2 g_i^2}\right) \leqslant \frac{p}{2}\log\left(\frac{1+\varkappa^*\sigma^2 g_p^2}{1+\varkappa\sigma^2 g_p^2}\right) \leqslant \frac{c}{2}\log\frac{\varkappa^*}{\varkappa}$$

с достаточным $c \geqslant p$. Таким образом, справедливо (D.2) с

$$\beta(\mathtt{r}_0) = \frac{2}{(p+2)}\left(\frac{\varkappa_1}{\varkappa^*}\right)^{p/2}\varkappa_1. \qquad \square$$

Предыдущее утверждение предоставляет достаточное условие на значение $\mathtt{r}_0$, влекущее концентрацию апостериорного распределения на $\mathcal{K}_\circ(\mathtt{r}_0)$.

Для следующего утверждения удобно ввести обозначения: для случайной величины μ определим

$$\hat{\mathbb{E}}\mu \stackrel{\text{def}}{=} \mathbb{E}[\mu\mathbb{I}\{\varkappa \in \mathcal{K}_\circ(\mathtt{r}_0)\} \mid \boldsymbol{Y},\ \varkappa \in \mathcal{K}_\circ(\mathtt{r}_0) \cup \mathcal{K}_1].$$

Теорема D.2. *Для любой неотрицательной функции* $f(\cdot)$ *на* $\mathbb{R}$ *на множестве* $\Omega(\mathtt{x})$ *выполняется*

$$\hat{\mathbb{E}} f\big(d_{\underline{\epsilon}}(\varkappa - \varkappa_{\underline{\epsilon}})\big) \geqslant \exp\{-\Delta_{\epsilon}^{-}(\mathtt{r}_0)\}\, \mathbb{E}\{f(\gamma)\mathbb{I}(|\gamma| \leqslant \mathtt{C}\mathtt{r}_0)\}, \tag{D.4}$$

где

$$\Delta_{\epsilon}^{-}(\mathtt{r}_0) = \Delta_{\epsilon}(\mathtt{r}_0) + \frac{p+1}{2} \log \frac{1+\delta}{1-\delta} + \rho_1(\mathtt{r}_0).$$

На $\Omega(\mathtt{x})$ *также выполняется* $\Delta_{\epsilon}^{-}(\mathtt{r}_0) \leqslant \mathtt{C}\,\delta\,(p+1+\mathtt{x})$.

Доказательство. Из соотношения (C.8) с функцией $f(\cdot) \equiv 1$ и (D.3) следует, что на множестве $\Omega(\mathtt{x})$

$$\int\limits_{\Upsilon \setminus \Upsilon_2} \exp\{\mathcal{L}(\boldsymbol{\nu}, \boldsymbol{\nu}^{\star})\}\, d\boldsymbol{\nu} = \int\limits_{\Upsilon_{\circ}} \exp\{\mathcal{L}(\boldsymbol{\nu}, \boldsymbol{\nu}^{\star})\}\, d\boldsymbol{\nu} + \int\limits_{\Upsilon_1} \exp\{\mathcal{L}(\boldsymbol{\nu}, \boldsymbol{\nu}^{\star})\}\, d\boldsymbol{\nu} \leqslant$$

$$\leqslant \{1 + \rho_1(\mathtt{r}_0)\} \int\limits_{\Upsilon_{\circ}} \exp\{\mathcal{L}(\boldsymbol{\nu}, \boldsymbol{\nu}^{\star})\}\, d\boldsymbol{\nu} \leqslant$$

$$\leqslant \{1 + \rho_1(\mathtt{r}_0)\} \exp\{-m_{\epsilon}(\boldsymbol{\xi}_{\epsilon}) + \tau_{\underline{\epsilon}}(\mathtt{r}_0)\} \leqslant \exp\{-m_{\epsilon}(\boldsymbol{\xi}_{\epsilon}) + \tau_{\underline{\epsilon}}(\mathtt{r}_0) + \rho_1(\mathtt{r}_0)\}.$$

Используя соотношение (C.9), получаем

$$\frac{\int_{\Upsilon_{\circ}(\mathtt{r}_0)} \exp\{\mathcal{L}(\boldsymbol{\nu}, \boldsymbol{\nu}^{\star})\} f\big(d_{\underline{\epsilon}}(\varkappa - \varkappa_{\underline{\epsilon}})\big)\, d\boldsymbol{\nu}}{\int_{\Upsilon \setminus \Upsilon_2} \exp\{\mathcal{L}(\boldsymbol{\nu}, \boldsymbol{\nu}^{\star})\}\, d\boldsymbol{\nu}} \geqslant$$

$$\geqslant \frac{\exp\{-m_{\underline{\epsilon}}(\boldsymbol{\xi}_{\underline{\epsilon}})\} \int \phi(u) f(u) \mathbb{I}\{u \in \mathcal{B}_{\epsilon}(\mathtt{r}_0)\}\, du}{\exp\{-m_{\epsilon}(\boldsymbol{\xi}_{\epsilon}) + \tau_{\underline{\epsilon}}(\mathtt{r}_0) + \rho_1(\mathtt{r}_0)\}} \geqslant$$

$$\geqslant \exp\{-\Delta_{\epsilon}^{-}(\mathtt{r}_0)\}\, \mathbb{E}[f(\gamma)\mathbb{I}\{\gamma \in \mathcal{B}_{\epsilon}(\mathtt{r}_0)\}],$$

где

$$\mathcal{B}_{\epsilon}(\mathtt{r}_0) \stackrel{\text{def}}{=} \{u \in \mathbb{R}\colon |d_0 d_{\underline{\epsilon}}^{-1}(u + \xi_{\underline{\epsilon}})| \leqslant \mathtt{r}_0\}.$$

Используя $|\xi_{\underline{\epsilon}}|^2 = (1+\delta)^{-1}\|\xi\|^2$, $d_{\underline{\epsilon}}^2 = (1+\delta)d_0^2$, получаем

$$\mathcal{B}_{\epsilon}(\mathtt{r}_0) \supseteq \{u \in \mathbb{R}\colon |u| \leqslant \mathtt{C}\mathtt{r}_0\}. \tag{D.5}$$

Это доказывает (D.4). □

В качестве следствия сформулируем утверждение о производящей функции моментов $d_{\underline{\epsilon}}(\varkappa - \varkappa_{\underline{\epsilon}})$, соответствующей экспоненциальной функции f. Потребуется дополнительное условие, что $\mathtt{r}_0^2 \geqslant \mathtt{C}$ для достаточно большой $\mathtt{C}$.

Следствие D.3. *Пусть* $\mathtt{r}_0^2 \geqslant \mathtt{C}$. *Для* $\lambda \in \mathbb{R}$, $\lambda^2 \leqslant \mathtt{x}$,

$$\log \hat{\mathbb{E}} \exp\big(\lambda d_{\underline{\epsilon}}(\varkappa - \varkappa_{\underline{\epsilon}})\big) \geqslant \frac{\lambda^2}{2} - \Delta_{\epsilon}^{\ominus} \Delta_{\epsilon}^{\ominus} \stackrel{\text{def}}{=} \Delta_{\epsilon}^{-}(\mathtt{r}_0) + \mathrm{e}^{-\mathtt{x}}.$$

Более того, при $\lambda^2 \leqslant 1$

$$\log \hat{\mathbb{E}} \exp\{\lambda d_0(\varkappa - \breve{\varkappa})\} \geqslant \frac{\lambda^2}{2} - \Delta_{\epsilon}^{\ominus}(\mathtt{r}_0).$$

На множестве $\Omega(\mathtt{x})$ *выполняется* $\Delta_{\boldsymbol{\epsilon}}^{\ominus}(\mathtt{r}_0) \leqslant \mathtt{C}\,\delta\,(p{+}1{+}\mathtt{x})$. *Для любого измеримого* $A \subset \mathbb{R}$ *на множестве* $\Omega_{\mathtt{x}}$ *при* $\boldsymbol{\delta}_{\underline{\boldsymbol{\epsilon}}} = d_0(\varkappa_{\underline{\boldsymbol{\epsilon}}} - \breve{\varkappa})$ *выполняется*

$$\mathbb{P}^{\circ}(d_0(\varkappa - \breve{\varkappa}) \in A) \geqslant \exp\{\Delta_{\boldsymbol{\epsilon}}^{-}(\mathtt{r}_0)\}\mathbb{P}(d_0 d_{\underline{\boldsymbol{\epsilon}}}^{-1}\gamma + \boldsymbol{\delta}_{\underline{\boldsymbol{\epsilon}}} \in A) - \mathrm{e}^{-\mathtt{x}} \geqslant \\ \geqslant \exp\{\Delta_{\boldsymbol{\epsilon}}^{-}(\mathtt{r}_0)\}\{\mathbb{P}(\gamma \in A) - \mathtt{C}\delta\,\mathfrak{q}^{1/2}\} - \mathrm{e}^{-\mathtt{x}}$$

для $\mathfrak{q} = 1 + |\xi|^2$.

Первое утверждение следует из теоремы D.2. Единственный важный дополнительный шаг — вычисление интеграла $\mathbb{E}\{\exp(\lambda\gamma)\mathbb{I}(|\gamma| \leqslant \mathtt{r})\}$.

Лемма D.4. *Пусть* $\gamma \sim \mathcal{N}(0,1)$ *и* $\mu \in (0,1)$. *Тогда для любого* $\lambda \in \mathbb{R}$, $\lambda^2 \leqslant \mathtt{x}$,

$$\log \mathbb{E}\{\exp(\lambda\gamma)\mathbb{I}(|\gamma| > \mathtt{r})\} \leqslant -\frac{1-\mu}{2}\mathtt{r}^2 + \frac{1}{2\mu}\lambda^2 + \frac{1}{2}\log\left(\frac{1}{\mu}\right). \tag{D.6}$$

Доказательство. Так как для $\mu < 1$

$$\mathbb{E}\{\exp(\lambda\gamma)\mathbb{I}(|\gamma| > \mathtt{r})\} \leqslant \mathrm{e}^{-(1-\mu)\mathtt{r}^2/2}\mathbb{E}\exp\left\{\lambda\gamma + \frac{(1-\mu)\gamma^2}{2}\right\},$$

то

$$\mathbb{E}\exp\left\{\lambda\gamma + \frac{(1-\mu)\gamma^2}{2}\right\} = \\ = (2\pi)^{-1/2}\int \exp\left\{\lambda\gamma - \frac{\mu\gamma^2}{2}\right\}d\gamma = \mu^{-1/2}\exp\left(\frac{\mu^{-1}\lambda^2}{2}\right).$$

Следовательно, справедливо (D.6). □

Далее используем утверждение с $\mu = 1/2$.

Лемма D.5. *Пусть* $\gamma \sim \mathcal{N}(0,1)$ *и* $\mathtt{r}^2 \geqslant 4(1+\mathtt{x})$. *Тогда для* $\lambda \in \mathbb{R}$, $\lambda^2 \leqslant \mathtt{x}$,

$$\mathbb{E}\{\exp(\lambda\gamma)\mathbb{I}(|\gamma| \leqslant \mathtt{r})\} \geqslant \mathrm{e}^{\lambda^2/2}(1 - \mathrm{e}^{-\mathtt{x}}). \tag{D.7}$$

Доказательство. Из соотношения (D.6) при $\mu = 1/2$ с учётом того, что $\mathbb{E}\exp(\lambda\gamma) = \mathrm{e}^{\lambda^2/2}$, $\lambda^2 \leqslant \mathtt{x}$ и $1 + \log(2) < 2$, получаем, что

$$\mathrm{e}^{-\lambda^2/2}\mathbb{E}\{\exp(\lambda\gamma)\mathbb{I}(|\gamma| \leqslant \mathtt{r})\} \geqslant 1 - \exp\left(-\frac{\mathtt{r}^2}{4} + 1 + \frac{1}{2}\log(2)\right) \geqslant 1 - \exp(-\mathtt{x}).$$

Следовательно, верно (D.7). □

Последняя оценка доказывает следствие D.3. Второе утверждение доказывается аналогично следствию C.2, последнее утверждение — аналогично следствию C.3.

E. Доказательство теоремы 3.1

Удобно представить случайную величину $\varkappa$ в виде

$$\varkappa = \varkappa\mathbb{I}\{\varkappa \in \mathcal{K}_{\circ}(\mathtt{r}_0)\} + \varkappa\mathbb{I}\{\varkappa \in \mathcal{K}_1\} + \varkappa\mathbb{I}\{\varkappa \in \mathcal{K}_2\} = \varkappa^{\circ} + \varkappa_1^{c} + \varkappa_2^{c}.$$

Результат о больших уклонениях показывает, что апостериорная масса $\varkappa_1^c$ мала при соответствующем выборе $\mathtt{r}_0$. Далее покажем, что распределение $\varkappa^\circ$ примерно нормальное, и придём к главному результату. Определим

$$\bar{\varkappa} \stackrel{\text{def}}{=} \hat{\mathbb{E}}\varkappa, \quad \mathfrak{S}_\circ^2 \stackrel{\text{def}}{=} \operatorname{Cov}(\bar{\varkappa}) \stackrel{\text{def}}{=} \hat{\mathbb{E}}(\varkappa - \bar{\varkappa})^2.$$

Достаточно показать, что на $\Omega(\mathtt{x})$ выполняются соотношения

$$\bigl(d_{\underline{\epsilon}}(\bar{\varkappa} - \varkappa_{\underline{\epsilon}})\bigr)^2 \leqslant 2\Delta_{\boldsymbol{\epsilon}}^* |1 - d_{\underline{\epsilon}}^2 \mathfrak{S}_\circ^2| \leqslant 2\Delta_{\boldsymbol{\epsilon}}^*,$$

и для $\boldsymbol{\lambda} \in \mathbb{R}$, $\lambda^2 \leqslant \mathtt{x}$, справедливо

$$\left| \log \hat{\mathbb{E}} \exp\{\lambda \mathfrak{S}_\circ^{-1}(\varkappa - \varkappa_{\boldsymbol{\epsilon}})\} - \frac{\lambda^2}{2} \right| \leqslant \mathtt{C}\Delta_{\boldsymbol{\epsilon}}^*. \tag{E.1}$$

Рассмотрим $\eta \stackrel{\text{def}}{=} d_{\underline{\epsilon}}(\varkappa - \varkappa_{\underline{\epsilon}})$. Следствия C.2 и D.3 утверждают для $\lambda \in \mathbb{R}$, $\lambda^2 \leqslant \mathtt{x}$, что

$$\lambda^2 \exp\{-\Delta^-\} \leqslant \hat{\mathbb{E}}(\lambda^\top \eta)^2 \leqslant \lambda^2 \exp\{\Delta^+\}, \tag{E.2}$$

$$\frac{\lambda^2}{2} - \Delta^- \leqslant \log \hat{\mathbb{E}} \exp(\lambda \boldsymbol{\eta}) \leqslant \frac{\lambda^2}{2} + \Delta^+, \tag{E.3}$$

где $\Delta^- = \Delta_{\boldsymbol{\epsilon}}^{\ominus}$ и $\Delta^+ = \Delta_{\boldsymbol{\epsilon}}^{\oplus}$.

Определим первые два момента η:

$$\bar{\eta} \stackrel{\text{def}}{=} \hat{\mathbb{E}}\eta, \quad S_\circ^2 \stackrel{\text{def}}{=} \hat{\mathbb{E}}(\eta - \bar{\eta})^2 = d_{\underline{\epsilon}}^2 \mathfrak{S}_\circ^2.$$

Воспользуемся следующим техническим утверждением.

Лемма E.1. *Предположим, что верно* (E.2). *Тогда при* $\Delta^* = \max\{\Delta^+, \Delta^-\}$

$$|\bar{\eta}| \leqslant \sqrt{2\Delta^*}, \quad |S_\circ^2 - 1| \leqslant 2\Delta^*. \tag{E.4}$$

Доказательство. Пусть $u = 1$. При $\lambda = u$ для $\hat{\mathbb{E}}(u\eta)^2 = \hat{\mathbb{E}}(\lambda\eta)^2$ из (E.2) получаем, что

$$\hat{\mathbb{E}}(u\eta)^2 \leqslant \exp\{\Delta^+\}, \quad \hat{\mathbb{E}}(u\eta)^2 \geqslant \exp\{-\Delta^-\}.$$

Заметим, что

$$\hat{\mathbb{E}}(u\eta)^2 = u^2 S_\circ^2 + (u\bar{\eta})^2,$$

и поэтому

$$\exp\{-\Delta^-\} \leqslant u^2 S_\circ^2 + (u\bar{\eta})^2 \leqslant \exp\{\Delta^+\}. \tag{E.5}$$

Аналогично для $u = \bar{\eta}/|\bar{\eta}|$

$$\hat{\mathbb{E}}\bigl(u(\eta - \bar{\eta})\bigr)^2 \geqslant \exp\{-\Delta^-\} \mathbb{E}\bigl(u(\gamma - \bar{\eta})\bigr)^2 = (1 + \bar{\eta}^2) \exp\{-\Delta^-\},$$

откуда следует

$$u^2 S_\circ^2 \geqslant (1 + \bar{\eta}^2) \exp\{-\Delta^-\}.$$

Данное неравенство противоречит (E.5), если $\bar{\eta}^2 > 2\Delta^*$. Таким образом, справедливо (E.4). □

Граница для первого момента при $\bar{\varkappa} = \hat{\mathbb{E}}\varkappa$ даёт

$$|d_{\underline{\epsilon}}(\bar{\varkappa} - \varkappa_{\underline{\epsilon}})| \leqslant \sqrt{2\Delta^*}. \tag{E.6}$$

Используя границу для второго момента и $\mathfrak{S}_\circ^2 = \hat{\mathbb{E}}(\varkappa - \bar{\varkappa})^2$, получаем

$$|d_{\underline{\epsilon}}^2 \mathfrak{S}_\circ^2 - 1| \leqslant 2\Delta^*. \tag{E.7}$$

Далее рассмотрим производящую функцию моментов для $\boldsymbol{\eta}^\circ \stackrel{\text{def}}{=} \mathfrak{S}_\circ^{-1}(\varkappa - \varkappa_{\boldsymbol{\epsilon}})$. Воспользуемся соотношениями

$$\lambda\boldsymbol{\eta}^\circ = \lambda_1 d_{\underline{\epsilon}}(\varkappa - \varkappa_{\underline{\epsilon}}) + \lambda_1 d_{\underline{\epsilon}}(\varkappa_{\underline{\epsilon}} - \varkappa_{\boldsymbol{\epsilon}}) = \lambda_1\boldsymbol{\eta} + \lambda_1^\top d_{\underline{\epsilon}}(\varkappa_{\underline{\epsilon}} - \varkappa_{\boldsymbol{\epsilon}})$$

при $\lambda_1 = d_{\underline{\epsilon}}^{-1}\mathfrak{S}_\circ^{-1}\lambda$. Таким образом,

$$|\log\hat{\mathbb{E}}\exp\{\lambda\boldsymbol{\eta}^\circ\} - \log\hat{\mathbb{E}}\exp\{\lambda_1\boldsymbol{\eta}\}| \leqslant |\lambda\mathfrak{S}_\circ^{-1}(\varkappa_{\underline{\epsilon}} - \varkappa_{\boldsymbol{\epsilon}})|. \tag{E.8}$$

Используя (E.7), для $\lambda_1^2 = \lambda^2 d_{\underline{\epsilon}}^{-2}\mathfrak{S}_\circ^{-2}$ получаем

$$(1-\Delta^*)\lambda^2 \leqslant \lambda^2 d_{\underline{\epsilon}}^{-2}\mathfrak{S}_\circ^{-2} \leqslant (1-\Delta^*)^{-1}\lambda^2.$$

Это неравенство, (E.6), (E.8) и (E.3) влекут

$$\left|\log\hat{\mathbb{E}}\exp\{\lambda\boldsymbol{\eta}^\circ\} - \frac{\lambda^2}{2}\right| \leqslant \mathtt{C}\Delta^*.$$

Таким образом, справедливо (E.1).

Литература

[1] Akaike H. Information theory and an extension of the maximum likelihood principle // Second International Symposium on Information Theory (Tsahkadsor, 1971). — Budapest: Akadémiai Kiadó, 1973. — P. 267—281.

[2] Birgé L., Massart P. Gaussian model selection // J. Eur. Math. Soc. — 2001. — Vol. 3, no. 3. — P. 203—268.

[3] Birgé L., Massart P. Minimal penalties for Gaussian model selection // Probab. Theory Relat. Fields. — 2007. — Vol. 138, no. 1-2. — P. 33—73.

[4] Cavalier L., Golubev Y. Risk hull method and regularization by projections of ill-posed inverse problems // Ann. Statist. — 2006. — Vol. 34, no. 4. — P. 1653—1677.

[5] Donoho D. L., Johnstone I. M. Adapting to unknown smoothness via wavelet shrinkage // J. Am. Statist. Assoc. — 1995. — Vol. 90 (432). — P. 1200—1224.

[6] Efrojmovich S., Pinsker M. (1984). Learning algorithm for nonparametric filtering // Autom. Remote Control. — 1984. — Vol. 45. — P. 1434—1440.

[7] Engl H. W., Hanke M., Neubauer A. Regularization of Inverse Problems. — Dordrecht: Kluwer Academic, 1996.

[8] Green P. J., Silverman B. (1994). Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach. — London: Chapman & Hall, 1994.

[9] Petrone S., Rousseau J., Scricciolo C. Bayes and empirical bayes: do they merge? // Biometrika. — 2012. — Vol. 99, no. 1. — P. 1—21.

[10] Spokoiny V. Parametric estimation. Finite sample theory // Ann. Statist. — 2012. — Vol. 40, no. 6. — P. 2877—2909. — arXiv:1111.3029.

[11] Spokoiny V. Basics of Modern Parametric Statistics. — Berlin: Springer, 2013.

[12] Stein C. M. Estimation of the mean of a multivariate normal distribution // Ann. Statist. — 1981. — Vol. 9, no. 6. — P. 1135—1151.

[13] Van der Vaart A. W. Asymptotic Statistics. — Cambridge: Cambridge Univ. Press, 1998. — (Cambridge Ser. Statist. Probab. Math.).