ISSN 0081-5438, Proceedings of the Steklov Institute of Mathematics, 2014, Vol. 287, pp. 232–255. © Pleiades Publishing, Ltd., 2014.
Published in Russian in Trudy Matematicheskogo Instituta imeni V.A. Steklova, 2014, Vol. 287, pp. 242–266.

Critical Dimension in the Semiparametric Bernstein–von Mises Theorem

Maxim E. Panov [a,b,c] **and Vladimir G. Spokoiny** [d,e,a]

Received June 2014

Abstract—The classical parametric and semiparametric Bernstein–von Mises (BvM) results are reconsidered in a nonclassical setup allowing finite samples and model misspecification. In the parametric case and in the case of a finite-dimensional nuisance parameter, we establish an upper bound on the error of Gaussian approximation of the posterior distribution of the target parameter; the bound depends explicitly on the dimension of the full and target parameters and on the sample size. This helps to identify the so-called *critical dimension* p_n of the full parameter for which the BvM result is applicable. In the important special i.i.d. case, we show that the condition "p_n^3/n is small" is sufficient for the BvM result to be valid under general assumptions on the model. We also provide an example of a model with the phase transition effect: the statement of the BvM theorem fails when the dimension p_n approaches $n^{1/3}$.

DOI: 10.1134/S0081543814080148

1. INTRODUCTION

The prominent Bernstein–von Mises (BvM) theorem claims that the posterior measure is asymptotically normal with the mean close to the maximum likelihood estimator (MLE) and the posterior variance is nearly the inverse of the total Fisher information matrix. The BvM result provides a theoretical background for Bayesian computations of the MLE and its variance. It also justifies the usage of elliptic credible sets based on the first two moments of the posterior. The classical version of the BvM theorem is stated for the standard parametric setup with a fixed parametric model and large samples (see [21, 29] for a detailed overview). However, in modern statistical applications one often faces very complicated models involving a lot of parameters with a limited sample size. This requires an extension of the classical results to such a nonclassical situation. See [12–14, 16] and references therein for some special phenomena arising in the Bayesian analysis when the parameter dimension increases. Even consistency of the posterior distribution in the nonparametric and semiparametric models is a nontrivial problem (cf. [25, 1]). Asymptotic normality of the posterior measure for these classes of models is still more challenging (see, e.g., [26]). Some results for particular semi- and nonparametric problems are available in [18, 17, 22, 8]. In [10] a version of the BvM statement based on a high order expansion of the profile sampler is obtained. The recent paper [2] extends the BvM statement from the classical parametric case to a rather general i.i.d. framework. In [7] the semiparametric BvM result for Gaussian process functional priors is studied. In [24] the semiparametric BvM theorem is derived for linear functionals of density, and in [9] the

[a] Moscow Institute of Physics and Technology (State University), Institutskii per. 9, Dolgoprudnyi, Moscow oblast, 141700 Russia.

[b] Institute for Information Transmission Problems (Kharkevich Institute), Russian Academy of Sciences, Bol'shoi Karetnyi per. 19, Moscow, 127994 Russia.

[c] Datadvance Company, Pokrovskii bul'var 3, stroenie 1B, Moscow, 109028 Russia.

[d] Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstr. 39, 10117 Berlin, Germany.

[e] Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany.

E-mail addresses: panov.maxim@gmail.com (M.E. Panov), spokoiny@wias-berlin.de (V.G. Spokoiny).

result is generalized to a broad class of models and functionals. However, all these results are limited to the asymptotic setup and to some special classes of models like i.i.d. or Gaussian ones.

In this paper we reconsider the BvM result for the parametric component of a general semiparametric model. An important feature of the study is that the sample size is fixed; we proceed with just one sample. A finite sample theory is especially challenging because most notions, methods, and tools in the classical theory are formulated in the asymptotic setup with a growing sample size. Only a few finite sample general results are available (see, e.g., the recent paper [6]). This paper focuses on the semiparametric problem when the full parameter is high- or infinite-dimensional but the target is low-dimensional. In the Bayesian framework, the aim is the marginal of the posterior corresponding to the target parameter (cf. [7]). Typical examples are provided by functional estimation, estimation of a function at a point, or simply estimation of a given subvector of the parameter vector. An interesting feature of the semiparametric BvM result is that the nuisance parameter appears only via the effective score and the related efficient Fisher information (cf. [2]). The methods of study rely heavily on the notion of the hardest parametric submodel. In addition, one assumes that an estimate of the nuisance parameter is available which ensures a certain accuracy of estimation (see [10] or [2]). This considerably simplifies the study but does not allow one to derive a qualitative relation between the full dimension of the parameter space and the total available information in the data.

Some recent results study the impact of a growing parameter dimension p_n on the quality of Gaussian approximation of the posterior. We mention [4, 5, 14–16] as specific examples (see the discussion after Theorem 5.1 below for more details).

In this paper we show that the *bracketing* approach of [27] can be used for obtaining a finite sample semiparametric version of the BvM theorem even if the full parameter dimension grows with the sample size. The ultimate goal of this paper is to quantify the so-called critical parameter dimension for which the BvM result can be applied. Our approach neither relies on a pilot estimate of the nuisance and target parameters nor involves the notion of the hardest parametric submodel. In the case of finite-dimensional nuisance the obtained results only require some smoothness of the log-likelihood function, its finite exponential moments, and some identifiability conditions. Further we specify this result in the i.i.d. setup and show that the imposed conditions are satisfied if p_n^3/n is small. We present an example showing that the dimension $p_n = O(n^{1/3})$ is indeed critical and the BvM result starts to fail if p_n grows over $n^{1/3}$.

Now we describe our setup. Let $\mathbf{Y}$ denote the observed random data and P denote the data distribution. The parametric statistical model assumes that the unknown data distribution P belongs to a given parametric family $(\mathsf{P}_{\boldsymbol{\upsilon}})$:

$$\mathbf{Y} \sim \mathsf{P} = \mathsf{P}_{\boldsymbol{\upsilon}^*} \in (\mathsf{P}_{\boldsymbol{\upsilon}},\, \boldsymbol{\upsilon} \in \Upsilon),$$

where Υ is some parameter space and $\boldsymbol{\upsilon}^* \in \Upsilon$ is the true value of the parameter. In the semiparametric framework, one attempts to recover only a low-dimensional component $\boldsymbol{\theta}$ of the whole parameter $\boldsymbol{\upsilon}$. This means that the target of estimation is

$$\boldsymbol{\theta}^* \stackrel{\mathrm{def}}{=} \Pi_0 \boldsymbol{\upsilon}^*$$

for some mapping $\Pi_0 \colon \Upsilon \to \mathbb{R}^q$, and $q \in \mathbb{N}$ stands for the dimension of the target. In the classical semiparametric setup, the vector $\boldsymbol{\upsilon}$ is usually represented as $\boldsymbol{\upsilon} = (\boldsymbol{\theta}, \boldsymbol{\eta})$, where $\boldsymbol{\theta}$ is the target of analysis while $\boldsymbol{\eta}$ is the *nuisance parameter*. We refer to this situation as a $(\boldsymbol{\theta}, \boldsymbol{\eta})$-*setup*, and our presentation follows this setting. An extension to the $\boldsymbol{\upsilon}$-setup with $\boldsymbol{\theta} = \Pi_0 \boldsymbol{\upsilon}$ is straightforward. Also for simplicity we first develop our results for the case when the total parameter space Υ is a subset of the Euclidean space of dimension p.

Another issue addressed in this paper is the model misspecification. In the majority of practical problems, it is unrealistic to expect that the model assumptions are exactly fulfilled, even if some rich nonparametric models are used. This means that the true data distribution P does not belong to the considered family $(\mathsf{P}_{\boldsymbol{\upsilon}},\, \boldsymbol{\upsilon} \in \Upsilon)$. The "true" value $\boldsymbol{\upsilon}^*$ of the parameter $\boldsymbol{\upsilon}$ can defined by

$$\boldsymbol{\upsilon}^* = \operatorname*{arg\,max}_{\boldsymbol{\upsilon}\in\Upsilon} \mathsf{E}\,\mathcal{L}(\boldsymbol{\upsilon}),$$

where $\mathcal{L}(\boldsymbol{\upsilon}) = \log \frac{d\mathsf{P}_{\boldsymbol{\upsilon}}}{d\boldsymbol{\mu}_0}(\mathbf{Y})$ is the log-likelihood function of the family $(\mathsf{P}_{\boldsymbol{\upsilon}})$ for some dominating measure $\boldsymbol{\mu}_0$. Under model misspecification, $\boldsymbol{\upsilon}^*$ defines the best parametric fit to P by the considered family (see [11, 19, 20] and references therein). The target $\boldsymbol{\theta}^*$ is defined by the mapping Π_0:

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

2. PARAMETRIC ESTIMATION: MAIN DEFINITIONS

We suppose that a large constant $\mathtt{x}$ is fixed which specifies random events $\Omega(\mathtt{x})$ of dominating probability. We say that a generic random set $\Omega(\mathtt{x})$ is *of dominating probability* if

$$\mathsf{P}(\Omega(\mathtt{x})) \ge 1 - \mathtt{C}e^{-\mathtt{x}}.$$

One of the main elements of our construction is a $p \times p$ matrix $\mathcal{D}_0^2$ which is defined similarly to the Fisher information matrix:

$$\mathcal{D}_0^2 \stackrel{\text{def}}{=} -\nabla^2\, \mathsf{E}\,\mathcal{L}(\boldsymbol{\upsilon}^*). \tag{2.1}$$

Here and in what follows we implicitly assume that the log-likelihood function $\mathcal{L}(\boldsymbol{\upsilon})$ is sufficiently smooth in $\boldsymbol{\upsilon}$, $\nabla\mathcal{L}(\boldsymbol{\upsilon})$ stands for its gradient, and $\nabla^2\, \mathsf{E}\,\mathcal{L}(\boldsymbol{\upsilon})$, for the Hessian of the expectation $\mathsf{E}\,\mathcal{L}(\boldsymbol{\upsilon})$. Define also the score vector

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

The definition of $\boldsymbol{\upsilon}^*$ implies that $\nabla\, \mathsf{E}\,\mathcal{L}(\boldsymbol{\upsilon}^*) = 0$ and hence $\mathsf{E}\,\boldsymbol{\xi} = 0$.

For the $(\boldsymbol{\theta}, \boldsymbol{\eta})$-setup, we consider the block representation of the vector $\nabla\mathcal{L}(\boldsymbol{\upsilon}^*)$ and matrix $\mathcal{D}_0^2$ from (2.1):

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

Define also a $q \times q$ matrix $\breve{D}_0^2$ and random vectors $\breve{\nabla}_{\boldsymbol{\theta}}, \breve{\boldsymbol{\xi}} \in \mathbb{R}^q$ as

$$\breve{D}_0^2 \stackrel{\text{def}}{=} D_0^2 - A_0 H_0^{-2} A_0^\top, \tag{2.2}$$

$$\breve{\nabla}_{\boldsymbol{\theta}} \stackrel{\text{def}}{=} \nabla_{\boldsymbol{\theta}} - A_0 H_0^{-2} \nabla_{\boldsymbol{\eta}}, \qquad \breve{\boldsymbol{\xi}} \stackrel{\text{def}}{=} \breve{D}_0^{-1} \breve{\nabla}_{\boldsymbol{\theta}}. \tag{2.3}$$

The $q \times q$ matrix $\breve{D}_0^2$ is usually called the *efficient Fisher information matrix*, while the random vector $\breve{\boldsymbol{\xi}} \in \mathbb{R}^q$ is the *efficient score*. Everywhere in the text for a vector $\mathbf{a}$ we denote by $\|\mathbf{a}\|$ its Euclidean norm and for a matrix A we denote by $\|A\|$ its operator norm.

3. CLASSICAL BERNSTEIN–VON MISES THEOREM

Let π be a prior measure on the parameter set Υ. Below we study the properties of the posterior measure which is the random measure on Υ describing the conditional distribution of $\boldsymbol{\upsilon}$ given $\mathbf{Y}$ and obtained by normalization of the product $\exp\{\mathcal{L}(\boldsymbol{\upsilon})\}\,\pi(d\boldsymbol{\upsilon})$. This relation is usually written as

$$\boldsymbol{\upsilon} \mid \mathbf{Y} \propto \exp\{\mathcal{L}(\boldsymbol{\upsilon})\}\,\pi(d\boldsymbol{\upsilon}). \tag{3.1}$$

An important feature of our analysis is that $\mathcal{L}(\boldsymbol{\upsilon})$ is not assumed to be the true log-likelihood. This means that a model misspecification is possible and the underlying data distribution can be beyond the considered parametric family. In this sense, the Bayes formula (3.1) describes a *quasi posterior* (see [11]). Below we show that smoothness of the log-likelihood function $\mathcal{L}(\boldsymbol{\upsilon})$ ensures a kind of a Gaussian approximation of the posterior measure. Our focus is to describe the accuracy of such approximation as a function of the parameter dimension p and other important characteristics of the model.

We suppose that the prior measure π has a positive density $\pi(\boldsymbol{\upsilon})$ with respect to the Lebesgue measure on Υ: $\pi(d\boldsymbol{\upsilon}) = \pi(\boldsymbol{\upsilon})\,d\boldsymbol{\upsilon}$. Then (3.1) can be written as

$$\boldsymbol{\upsilon} \mid \mathbf{Y} \propto \exp\{\mathcal{L}(\boldsymbol{\upsilon})\}\,\pi(\boldsymbol{\upsilon}). \tag{3.2}$$

The famous Bernstein–von Mises (BvM) theorem claims that the posterior centered by any efficient estimator $\widetilde{\boldsymbol{\upsilon}}$ of the parameter $\boldsymbol{\upsilon}^*$ (for example, the MLE) and scaled by the total Fisher information matrix is nearly standard normal:

$$\mathcal{D}_0(\boldsymbol{\upsilon} - \widetilde{\boldsymbol{\upsilon}}) \mid \mathbf{Y} \xrightarrow{\mathrm{w}} \mathcal{N}(0, I_p),$$

where I_p is the identity matrix of dimension p.

An important feature of the posterior distribution is that it is entirely known and can be numerically assessed. If we know in addition that the posterior is nearly normal, it suffices to compute its mean and variance for building the concentration set and credible set. It is also worth noting that the BvM theorem itself does not require the prior distribution to be proper, and the phenomenon can be observed in the case of improper priors as well (for examples, see [3]).

In this work we investigate the properties of the posterior distribution for the target parameter $\boldsymbol{\vartheta} = \Pi_0\boldsymbol{\upsilon}$. In this case (3.2) can be written as

$$\boldsymbol{\vartheta} \mid \mathbf{Y} \propto \int \exp\{\mathcal{L}(\boldsymbol{\upsilon})\}\,\pi(\boldsymbol{\upsilon})\,d\boldsymbol{\eta}. \tag{3.3}$$

The BvM result in this case transforms into

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

where I_q is the identity matrix of dimension q, $\widetilde{\boldsymbol{\theta}} = \Pi_0\widetilde{\boldsymbol{\upsilon}}$, and $\breve{D}_0^2$ is given in (2.2).

We consider two important classes of priors, namely, non-informative and continuous priors. Our goal is to prove that under reasonable conditions, the posterior measure for the target parameter (3.3) is close to a Gaussian distribution with properly chosen mean and variance even for finite samples. The other important issue is to specify the conditions on the sample size and the dimension of the parameter space for which the BvM result is still applicable.

4. FINITE SAMPLE SEMIPARAMETRIC BERNSTEIN–VON MISES THEOREM WITH FINITE-DIMENSIONAL NUISANCE

First we state the BvM result about the properties of the $\boldsymbol{\vartheta}$-posterior given by (3.3) in the case of a uniform prior, that is, $\pi(\boldsymbol{\upsilon}) \equiv 1$ on Υ. Define

$$\overline{\boldsymbol{\vartheta}} \stackrel{\mathrm{def}}{=} \mathsf{E}(\boldsymbol{\vartheta} \mid \mathbf{Y}), \qquad \mathfrak{S}^2 \stackrel{\mathrm{def}}{=} \operatorname{Cov}(\boldsymbol{\vartheta} \mid \mathbf{Y}) \stackrel{\mathrm{def}}{=} \mathsf{E}\bigl\{(\boldsymbol{\vartheta} - \overline{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta} - \overline{\boldsymbol{\vartheta}})^\top \mid \mathbf{Y}\bigr\}. \tag{4.1}$$

Define also

$$\boldsymbol{\theta}^\circ \stackrel{\mathrm{def}}{=} \boldsymbol{\theta}^* + \breve{D}_0^{-1}\breve{\boldsymbol{\xi}}.$$

Below we present a version of the BvM result in the considered nonasymptotic setup which claims that $\overline{\boldsymbol{\vartheta}}$ is close to $\boldsymbol{\theta}^{\circ}$, $\mathfrak{S}^2$ is nearly equal to $\breve{D}_0^{-2}$, and $\breve{D}_0(\boldsymbol{\vartheta} - \boldsymbol{\theta}^{\circ})$ is nearly standard normal conditionally on $\mathbf{Y}$. Recall the notation $\mathtt{C}$ for a generic absolute constant and $\mathtt{x}$ for a positive value ensuring that $\mathrm{e}^{-\mathtt{x}}$ is negligible. By $\Omega(\mathtt{x})$ we denote a random event of dominating probability with $\mathsf{P}(\Omega(\mathtt{x})) \ge 1 - \mathtt{C}\mathrm{e}^{-\mathtt{x}}$. The exact values of $\mathtt{C}$ will be specified below.

Theorem 4.1. *Suppose the conditions of Subsection* 4.1 (*see below*) *hold. Let the prior be uniform on* Υ. *Then there exists a random event* $\Omega(\mathtt{x})$ *of dominating probability at least* $1 - 4\mathrm{e}^{-\mathtt{x}}$ *such that on* $\Omega(\mathtt{x})$

$$\|\breve{D}_0(\overline{\boldsymbol{\vartheta}} - \boldsymbol{\theta}^{\circ})\|^2 \le 4\Delta(\mathtt{r}_0, \mathtt{x}) + 16\mathrm{e}^{-\mathtt{x}},$$

$$\|I_q - \breve{D}_0\mathfrak{S}^2\breve{D}_0\| \le 4\Delta(\mathtt{r}_0, \mathtt{x}) + 16\mathrm{e}^{-\mathtt{x}},$$

where $\overline{\boldsymbol{\vartheta}}$ *and* $\mathfrak{S}^2$ *are from* (4.1) *and* $\Delta(\mathtt{r}_0, \mathtt{x})$ *is from* (6.3) (*see below*).

Moreover, on $\Omega(\mathtt{x})$ *for any measurable set* $A \subset \mathbb{R}^q$

$$\exp\bigl\{-2\Delta(\mathtt{r}_0, \mathtt{x}) - 8\mathrm{e}^{-\mathtt{x}}\bigr\}\, \mathsf{P}(\boldsymbol{\gamma} \in A) - \mathrm{e}^{-\mathtt{x}} \le \mathsf{P}\bigl(\breve{D}_0(\boldsymbol{\vartheta} - \boldsymbol{\theta}^{\circ}) \in A \mid \mathbf{Y}\bigr)$$
$$\le \exp\bigl\{2\Delta(\mathtt{r}_0, \mathtt{x}) + 5\mathrm{e}^{-\mathtt{x}}\bigr\}\, \mathsf{P}(\boldsymbol{\gamma} \in A),$$

where $\boldsymbol{\gamma}$ *is a standard Gaussian vector in* $\mathbb{R}^q$.

The condition "$\Delta(\mathtt{r}_0, \mathtt{x})$ is small" yields the desirable BvM result, that is, the posterior measure after centering and standardization is close in total variation to the standard normal law. The classical asymptotic results immediately follow for many classical models (see the discussion in Section 5 below). The next corollary extends the previous result by using empirically computable objects.

Corollary 4.1. *Under the conditions of Theorem* 4.1, *for any measurable set* $A \subset \mathbb{R}^q$ *on a random set* $\Omega(\mathtt{x})$ *of dominating probability at least* $1 - 4\mathrm{e}^{-\mathtt{x}}$

$$\exp\bigl\{-2\Delta(\mathtt{r}_0, \mathtt{x}) - 8\mathrm{e}^{-\mathtt{x}}\bigr\}\{\mathsf{P}(\boldsymbol{\gamma} \in A) - \tau\} - \mathrm{e}^{-\mathtt{x}} \le \mathsf{P}(\mathfrak{S}^{-1}(\boldsymbol{\vartheta} - \overline{\boldsymbol{\vartheta}}) \in A \mid \mathbf{Y})$$
$$\le \exp\bigl\{2\Delta(\mathtt{r}_0, \mathtt{x}) + 5\mathrm{e}^{-\mathtt{x}}\bigr\}\{\mathsf{P}(\boldsymbol{\gamma} \in A) + \tau\},$$

where $\tau = (\Delta(\mathtt{r}_0, \mathtt{x})^2 q + (1 + \Delta(\mathtt{r}_0, \mathtt{x}))^2 \Delta(\mathtt{r}_0, \mathtt{x})^2)/2$ *and* $\boldsymbol{\gamma}$ *is a standard Gaussian vector in* $\mathbb{R}^q$.

This corollary is important because in practical applications we do not know the matrix $\breve{D}_0$ and vector $\boldsymbol{\theta}^{\circ}$, but the matrix $\mathfrak{S}^{-1}$ and vector $\overline{\boldsymbol{\vartheta}}$ can be found by numerical computations. If the dimension q is fixed, the result is valid under the same condition, namely, when "$\Delta(\mathtt{r}_0, \mathtt{x})$ is small." Moreover, the statement can be extended to situations when q grows but $\Delta(\mathtt{r}_0, \mathtt{x})q^{1/2}$ is still small. The results for a non-informative prior can be extended to the case of a general prior $\pi(d\boldsymbol{\upsilon})$ with a density $\pi(\boldsymbol{\upsilon})$ which is uniformly continuous (see Subsection 4.2 below for details).

4.1. Conditions. Our approach assumes a number of conditions to be satisfied. The list is essentially as in [27]; one can find there some discussion and examples showing that the conditions are not restrictive and are fulfilled in most classical models used in statistical studies like the i.i.d., regression, or generalized linear models. The conditions are split into local and global. The local conditions only describe the properties of the process $\mathcal{L}(\boldsymbol{\upsilon})$ for $\boldsymbol{\upsilon} \in \Upsilon_0(\mathtt{r}_0)$ with some fixed value $\mathtt{r}_0$:

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

The global conditions have to be fulfilled on the whole Υ. Define the stochastic component $\zeta(\boldsymbol{\upsilon})$ of $\mathcal{L}(\boldsymbol{\upsilon})$:

$$\zeta(\boldsymbol{\upsilon}) \stackrel{\mathrm{def}}{=} \mathcal{L}(\boldsymbol{\upsilon}) - \mathsf{E}\,\mathcal{L}(\boldsymbol{\upsilon}).$$

We start with some exponential moment conditions.

(ED_0) There exists a constant $\nu_0 > 0$, a positive symmetric $p \times p$ matrix $\mathcal{V}_0^2$ satisfying the condition $\operatorname{Var}\{\nabla\zeta(\boldsymbol{\upsilon}^*)\} \le \mathcal{V}_0^2$, and a constant $\mathtt{g} > 0$ such that

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

(ED_2) There exist constants $\nu_0, \omega > 0$ and, for each $\mathtt{r} > 0$, a constant $\mathtt{g}(\mathtt{r}) > 0$ such that for all $\boldsymbol{\upsilon} \in \Upsilon_0(\mathtt{r})$

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

The next condition is needed to ensure some smoothness properties of the expected log-likelihood $\mathsf{E}\,\mathcal{L}(\boldsymbol{\upsilon})$ in the local zone $\boldsymbol{\upsilon} \in \Upsilon_0(\mathtt{r}_0)$. Define

$$\mathcal{D}_0^2(\boldsymbol{\upsilon}) \stackrel{\mathrm{def}}{=} -\nabla^2\,\mathsf{E}\,\mathcal{L}(\boldsymbol{\upsilon}).$$

Then $\mathcal{D}_0^2 = \mathcal{D}_0^2(\boldsymbol{\upsilon}^*)$.

($\mathcal{L}_0$) There exists a constant $\delta(\mathtt{r})$ such that on the set $\Upsilon_0(\mathtt{r})$ for all $\mathtt{r} \le \mathtt{r}_0$ it holds that

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

The global identification condition is as follows:

($\mathcal{L}\mathtt{r}$) For any $\mathtt{r}$ there exists a value $\mathtt{b}(\mathtt{r}) > 0$ such that $\mathtt{r}\mathtt{b}(\mathtt{r}) \to \infty$ as $\mathtt{r} \to \infty$ and

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

Finally we specify the regularity conditions. We begin by representing the information and covariance matrices in a block form:

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

The *identifiability conditions* in [27] ensure that the matrix $\mathcal{D}_0$ is positive and satisfies $\mathfrak{a}^2\mathcal{D}_0^2 \ge \mathcal{V}_0^2$ for some $\mathfrak{a} > 0$. Here we restate these conditions in the special block form which is specific for the $(\boldsymbol{\theta}, \boldsymbol{\eta})$-setup.

($\mathcal{I}$) There are constants $\mathfrak{a} > 0$ and $\nu < 1$ such that

$$\mathfrak{a}^2 D_0^2 \ge V_0^2, \qquad \mathfrak{a}^2 H_0^2 \ge Q_0^2, \qquad \mathfrak{a}^2\mathcal{D}_0^2 \ge \mathcal{V}_0^2 \tag{4.2}$$

and

$$\|D_0^{-1}A_0H_0^{-2}A_0^\top D_0^{-1}\| \le \nu. \tag{4.3}$$

The quantity ν bounds the angle between the target and nuisance subspaces in the tangent space. The regularity condition ($\mathcal{I}$) ensures that this angle is not too small and, hence, the target and nuisance parameters are identifiable. In particular, the matrix $\breve{D}_0^2$ is well posed under ($\mathcal{I}$). The bounds in (4.2) are given with the same constant $\mathfrak{a}$ only to simplify the notation. One can show that the last bound on $\mathcal{D}_0^2$ follows from the first two and (4.3) with another constant $\mathfrak{a}'$ depending on $\mathfrak{a}$ and ν only.

4.2. Extension of Theorem 4.1 to a flat prior. The results of Theorem 4.1 for a non-informative prior can be extended to the case of a general prior $\pi(d\boldsymbol{\upsilon})$ with a density $\pi(\boldsymbol{\upsilon})$ which is uniformly continuous and sufficiently flat on the local set $\Upsilon_0(\mathtt{r}_0)$. More precisely, let $\pi(\boldsymbol{\upsilon})$ satisfy

$$\sup_{\boldsymbol{\upsilon}\in\Upsilon_0(\mathtt{r}_0)} |\log\pi(\boldsymbol{\upsilon}) - \log\pi(\boldsymbol{\upsilon}^*)| \le \alpha(\mathtt{r}_0), \qquad \sup_{\boldsymbol{\upsilon}\in\Upsilon} \frac{\pi(\boldsymbol{\upsilon})}{\pi(\boldsymbol{\upsilon}^*)} \le \mathtt{C}(\mathtt{r}_0),$$

where $\alpha(\mathtt{r}_0)$ is a small constant while $\mathtt{C}(\mathtt{r}_0)$ is any fixed constant. Then the results of Theorem 4.1 continue to apply with an obvious correction of the approximation error.

Indeed, for any local set $A \subseteq \Upsilon_0(\mathtt{r}_0)$ one can apply the bounds

$$\int_A \exp\{L(\boldsymbol{\upsilon})\}\pi(\boldsymbol{\upsilon})\,d\boldsymbol{\upsilon} \le \mathrm{e}^{\alpha(\mathtt{r}_0)}\pi(\boldsymbol{\upsilon}^*)\int_A \exp\{L(\boldsymbol{\upsilon})\}\,d\boldsymbol{\upsilon},$$

$$\int_A \exp\{L(\boldsymbol{\upsilon})\}\pi(\boldsymbol{\upsilon})\,d\boldsymbol{\upsilon} \ge \mathrm{e}^{-\alpha(\mathtt{r}_0)}\pi(\boldsymbol{\upsilon}^*)\int_A \exp\{L(\boldsymbol{\upsilon})\}\,d\boldsymbol{\upsilon}.$$

This, in particular, implies that for each $A \subset \Upsilon_0(\mathtt{r}_0)$

$$\mathsf{P}_\pi(A \mid \mathbf{Y}) \le \exp\{2\alpha(\mathtt{r}_0)\}\,\mathsf{P}(A \mid \mathbf{Y}). \tag{4.4}$$

The tail probability of the complement $\Upsilon_0^{\mathrm{c}}(\mathtt{r}_0) \stackrel{\mathrm{def}}{=} \Upsilon \setminus \Upsilon_0(\mathtt{r}_0)$ of $\Upsilon_0(\mathtt{r}_0)$ can be enlarged by a factor of $\mathtt{C}(\mathtt{r}_0)$ relative to the uniform prior:

$$\int_{\Upsilon_0^{\mathrm{c}}(\mathtt{r}_0)} \exp\{L(\boldsymbol{\upsilon})\}\pi(\boldsymbol{\upsilon})\,d\boldsymbol{\upsilon} \le \mathtt{C}(\mathtt{r}_0)\,\pi(\boldsymbol{\upsilon}^*)\int_{\Upsilon_0^{\mathrm{c}}(\mathtt{r}_0)} \exp\{L(\boldsymbol{\upsilon})\}\,d\boldsymbol{\upsilon};$$

hence

$$\mathsf{P}_\pi(\Upsilon_0^{\mathrm{c}}(\mathtt{r}_0) \mid \mathbf{Y}) \le \mathtt{C}(\mathtt{r}_0)\,\mathsf{P}(\Upsilon_0^{\mathrm{c}}(\mathtt{r}_0) \mid \mathbf{Y}). \tag{4.5}$$

In particular, if the tail of the non-informative posterior satisfies $\mathsf{P}(\Upsilon_0^{\mathrm{c}}(\mathtt{r}_0) \mid \mathbf{Y}) \le \mathrm{e}^{-\mathtt{x}}$, then $\mathsf{P}_\pi(\Upsilon_0^{\mathrm{c}}(\mathtt{r}_0) \mid \mathbf{Y}) \le \mathtt{C}(\mathtt{r}_0)\mathrm{e}^{-\mathtt{x}}$.

Theorem 4.2. *Suppose the conditions of Theorem* 4.1 *are satisfied. Let also* $\Pi \sim \mathcal{N}(0, G^{-2})$ *be a Gaussian prior measure on* $\mathbb{R}^q$ *such that*

$$\|\mathcal{D}_0^{-1}G^2\mathcal{D}_0^{-1}\| \le \epsilon^2, \tag{4.6}$$

where ϵ *is a given constant. Then* (4.4) *and* (4.5) *hold with* $\mathtt{C}(\mathtt{r}_0) \le \exp\{\|G\boldsymbol{\upsilon}^*\|^2/2\}$ *and* $\alpha(\mathtt{r}_0) = \max\{\epsilon\mathtt{r}_0\|G\boldsymbol{\upsilon}^*\|, \epsilon^2\mathtt{r}_0^2/2\}$.

The result of Theorem 4.2 tells us that the BvM result holds if the prior distribution is flat enough. The case of regularizing prior is not considered in this work but can be easily handled within the same framework. We are going to consider this case in forthcoming papers.

5. THE I.I.D. CASE AND CRITICAL DIMENSION

This section shows how the general results from the previous sections can be linked to the classical asymptotic results in the statistical literature. The nice feature of the whole approach based on the local bracketing is that all the results are stated under the same list of conditions: having checked the conditions once, one can directly apply any of the mentioned results. Typical examples include the i.i.d., generalized linear, and median regression models. Here we briefly discuss how the BvM result can be applied to one typical case, namely, to an i.i.d. experiment.

5.1. The i.i.d. case. Let $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top$ be an i.i.d. sample from a measure P. Here we suppose the conditions of [27, Sect. 5.1] on P and $(P_{\boldsymbol{\upsilon}})$ to be fulfilled. These conditions are essential analogues of the conditions of Subsection 4.1 but for likelihood of one observation. We admit that the parametric assumption $P \in (P_{\boldsymbol{\upsilon}},\, \boldsymbol{\upsilon} \in \Upsilon)$ can be misspecified and consider the asymptotic setup with n growing to infinity and simultaneously $p = p_n$ growing to infinity. In this setup the following theorem is valid.

Theorem 5.1. *Suppose the conditions of* [27, Sect. 5.1] *hold. Let also* $p_n \to \infty$ *and* $p_n^3/n \to 0$. *Then the result of Theorem* 4.1 *holds with* $\Delta(\mathtt{r}_0, \mathtt{x}) = \mathtt{C}\sqrt{p_n^3/n}$ *and* $\mathcal{D}_0^2 = n\mathbb{F}_{\boldsymbol{\upsilon}^*}$, *where* $\mathbb{F}_{\boldsymbol{\upsilon}^*}$ *is the Fisher information of* $(P_{\boldsymbol{\upsilon}})$ *at* $\boldsymbol{\upsilon}^*$.

A similar result about asymptotic normality of the posterior in a linear regression model can be found in [14]. However, the convergence is proved under the condition $(p_n^4 \log p_n)/n \to 0$, which appears to be too strong. In [15] it is shown that the dimensionality constraint can be relaxed to $p_n^3/n \to 0$ for exponential models with a product structure. In [5] the BvM result is proved in a specific class of i.i.d. models with discrete probability distribution under the condition $p_n^3/n \to 0$. Further examples and the related conditions for Gaussian models are presented in [16].

5.2. Critical dimension. This subsection discusses the issue of a *critical dimension*. Namely, we assume that the total dimension p grows with the sample size n and write $p = p_n$. Theorem 5.1 requires that $p_n = o(n^{1/3})$. Here we show that this condition is essential and cannot be dropped or relaxed. Namely, we present an example for which $p_n^3/n \ge \beta^2 > 0$ and the posterior distribution does not concentrate around MLE.

Let n and p_n be such that $M_n = n/p_n$ is an integer. We consider a simple Poissonian model with $Y_i \sim \text{Poisson}(\upsilon_j)$ for $i \in \mathcal{I}_j$, where $\mathcal{I}_j \stackrel{\text{def}}{=} \{i\colon \lceil i/M_n \rceil = j\}$ for $j = 1, \ldots, p_n$ and $\lceil x \rceil$ is the nearest integer greater than or equal to x. Let also $u_j = \log \upsilon_j$ be the canonical parameter. The log-likelihood $\mathcal{L}(\mathbf{u})$ with $\mathbf{u} = (u_1, \ldots, u_{p_n})$ reads as

$$\mathcal{L}(\mathbf{u}) = \sum_{j=1}^{p_n} \bigl(Z_j u_j - M_n e^{u_j}\bigr), \qquad \text{where} \qquad Z_j \stackrel{\text{def}}{=} \sum_{i \in \mathcal{I}_j} Y_i.$$

We consider the problem of estimating the mean of the u_j's:

$$\theta = \frac{1}{p_n}(u_1 + \ldots + u_{p_n}).$$

Below we study this problem in the asymptotic setup with $p_n \to \infty$ as $n \to \infty$ when the underlying measure P corresponds to $u_1^* = \ldots = u_{p_n}^* = u^*$ for some u^*, which yields $\theta^* = u^*$. The value u^* will be specified later. We consider an i.i.d. exponential prior on the parameters υ_j of the Poisson distribution:

$$\upsilon_j \sim \text{Exp}(\mu).$$

Below we admit that μ may depend on n. Our results are valid for $\mu \le \mathtt{C}\sqrt{n/\log n}$. The posterior is Gamma distributed:

$$\upsilon_j \mid \mathbf{Y} \sim \text{Gamma}(\alpha_j, \mu_j),$$

where $\alpha_j = 1 + \sum_{i \in \mathcal{I}_j} Y_i$ and $\mu_j = \mu/(M_n\mu + 1)$.

First we describe the profile MLE $\widetilde{\theta}_n$ of the target parameter θ. The MLE for the full parameter $\boldsymbol{\upsilon}$ reads as $\widetilde{\boldsymbol{\upsilon}} = (\widetilde{\upsilon}_1, \ldots, \widetilde{\upsilon}_{p_n})^\top$ with $\widetilde{\upsilon}_j = Z_j/M_n$. Thus, the profile MLE $\widetilde{\theta}_n$ reads as

$$\widetilde{\theta}_n = \frac{1}{p_n} \sum_{j=1}^{p_n} \log \widetilde{\upsilon}_j.$$

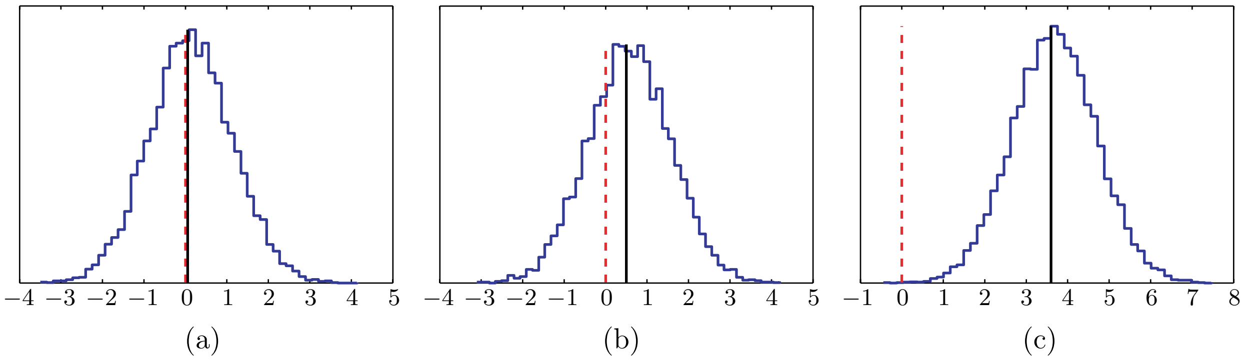

Fig. 1. Posterior distribution of $\beta_n^{-1} p_n(\theta - \widetilde{\theta}_n)$ for $\beta_n = 1/\log p_n$ (a), $\beta_n = 1$ (b), and $\beta_n = \log p_n$ (c). Solid lines are for the posterior mean, and dashed lines are for the true mean.

Furthermore, the efficient Fisher information $\breve{D}_0^2$ is equal to $p_n^{-1} n$ (see Lemma 7.6 below). Let us note that as $\widetilde{\theta}_n$ is a profile MLE, it is an efficient estimator with asymptotic variance equal to $1/\breve{D}_0^2$.

Theorem 5.2. *Let* $Y_i \sim \text{Poisson}(\upsilon^*)$ *for all* $i = 1, \ldots, n$ *with* $\upsilon^* = 1/p_n$. *Then the following holds*:

1. *If* $p_n^3/n \to 0$ *as* $p_n \to \infty$, *then*

$$p_n^{1/2} n^{1/2}(\theta - \widetilde{\theta}_n) \mid \mathbf{Y} \xrightarrow{\text{w}} \mathcal{N}(0, 1).$$

2. *Let* $p_n^3/n \equiv \beta > 0$. *Then*

$$p_n^{1/2} n^{1/2}(\theta - \widetilde{\theta}_n) \mid \mathbf{Y} \xrightarrow{\text{w}} \mathcal{N}(\beta/2, 1).$$

3. *If* $p_n^3/n \to \infty$, *but* $p_n^4/n^{3/2} \to 0$, *then*

$$p_n^{1/2} n^{1/2}(\theta - \widetilde{\theta}_n) \mid \mathbf{Y} \xrightarrow{\text{w}} \infty.$$

We carried out a series of experiments to numerically demonstrate the results of Theorem 5.2. The dimension of the parameter space was fixed at $p_n = 10\,000$. Three cases were considered:

(1) $p_n^{3/2}/n^{1/2} = 1/\log p_n$, which corresponds to $p_n^3/n \to 0$, $n \to \infty$;

(2) $p_n^{3/2}/n^{1/2} \equiv 1$;

(3) $p_n^{3/2}/n^{1/2} = \log p_n$, which corresponds to $p_n^3/n \to \infty$, $n \to \infty$.

For each sample, 10 000 realizations of $\mathbf{Y}$ were generated from the exponential distribution $\text{Exp}(\upsilon_*)$ and so were corresponding posterior values $\theta \mid \mathbf{Y}$. The resulting posterior distribution for the three cases is demonstrated in Fig. 1. It can be easily seen that the results of Theorem 5.2 are numerically confirmed.

6. SUPPLEMENTARY MATERIAL

This section contains some supplementary statements which are of interest in themselves.

6.1. Bracketing and upper function devices. This subsection briefly overviews the main constructions of [27] including the bracketing bound and the upper function results. The bracketing bound describes the quality of quadratic approximation of the log-likelihood process $\mathcal{L}(\boldsymbol{\upsilon})$ in a local vicinity of the point $\boldsymbol{\upsilon}^*$, while the upper function method is used to show that the full MLE $\widetilde{\boldsymbol{\upsilon}}$ belongs to this vicinity with a dominating probability. Introduce the notation $L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) = \mathcal{L}(\boldsymbol{\upsilon}) - \mathcal{L}(\boldsymbol{\upsilon}^*)$ for the (quasi) log-likelihood ratio. Given $\mathtt{r} > 0$, define the local set

$$\Upsilon_0(\mathtt{r}) \stackrel{\text{def}}{=} \{\boldsymbol{\upsilon}\colon\ (\boldsymbol{\upsilon} - \boldsymbol{\upsilon}^*)^\top \mathcal{D}_0^2 (\boldsymbol{\upsilon} - \boldsymbol{\upsilon}^*) \le \mathtt{r}^2\}. \tag{6.1}$$

Define the quadratic processes $\mathbb{L}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*)$:

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

The next result states the local bracketing bound. The formulation assumes that some value $\mathtt{x}$ is fixed such that $\mathrm{e}^{-\mathtt{x}}$ is sufficiently small. If the dimension p is large, one can set $\mathtt{x} = \mathtt{C} \log p$. We also assume that a value $\mathtt{r} = \mathtt{r}_0$ is fixed which separates the local and global zones.

Theorem 6.1. *Suppose that conditions* (ED_0), (ED_2), $(\mathcal{L}_0)$, *and* $(\mathcal{I})$ *from Subsection* 4.1 *hold for some* $\mathtt{r}_0 > 0$. *Then on a random set* $\Omega_{\mathtt{r}_0}(\mathtt{x})$ *of dominating probability at least* $1 - \mathrm{e}^{-\mathtt{x}}$

$$|L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) - \mathbb{L}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*)| \le \Delta(\mathtt{r}_0, \mathtt{x}), \qquad \boldsymbol{\upsilon} \in \Upsilon_0(\mathtt{r}_0), \tag{6.2}$$

where

$$\Delta(\mathtt{r}_0, \mathtt{x}) \stackrel{\mathrm{def}}{=} \bigl\{\delta(\mathtt{r}_0) + 6\nu_0\, z_{\mathbb{H}}(\mathtt{x})\omega\bigr\}\mathtt{r}_0^2, \tag{6.3}$$

$z_{\mathbb{H}}(\mathtt{x}) = 2p^{1/2} + \sqrt{2\mathtt{x}} + 4\mathtt{g}^{-1}(\mathtt{g}^{-2}\mathtt{x} + 1)p$, *and* $\Upsilon_0(\mathtt{r}_0)$ *is defined in* (6.1). *Moreover, on a random set* $\Omega_B(\mathtt{x})$ *of dominating probability at least* $1 - 2\mathrm{e}^{-\mathtt{x}}$, *the random vector* $\boldsymbol{\xi} = \mathcal{D}_0^{-1} \nabla \mathcal{L}(\boldsymbol{\upsilon}^*)$ *satisfies the inequality*

$$\|\boldsymbol{\xi}\|^2 \le z_B^2(\mathtt{x}),$$

where $z_B^2(\mathtt{x}) \stackrel{\mathrm{def}}{=} \mathtt{p}_B + 6\lambda_B \mathtt{x}$ *with*

$$B \stackrel{\mathrm{def}}{=} \mathcal{D}_0^{-1} \mathcal{V}_0^2 \mathcal{D}_0^{-1}, \qquad \mathtt{p}_B \stackrel{\mathrm{def}}{=} \operatorname{tr} B, \qquad \lambda_B \stackrel{\mathrm{def}}{=} \lambda_{\max}(B).$$

Furthermore, assume that condition $(\mathcal{L}\mathtt{r})$ *holds with* $\mathtt{b}(\mathtt{r}) \equiv \mathtt{b}$, *i.e.,*

$$-\mathsf{E}\, L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) \ge \mathtt{b}\|\mathcal{D}_0(\boldsymbol{\upsilon} - \boldsymbol{\upsilon}^*)\|^2$$

for all $\boldsymbol{\upsilon} \in \Upsilon \setminus \Upsilon_0(\mathtt{r}_0)$. *Let also* $\mathtt{r} \ge 2(z_B(\mathtt{x}) + \varrho(\mathtt{r}, \mathtt{x}))/\mathtt{b}$, *where* $\varrho(\mathtt{r}, \mathtt{x}) = 6\nu_0\, z_{\mathbb{H}}(\mathtt{x} + \log(2\mathtt{r}/\mathtt{r}_0))\omega$. *Then the inequality*

$$L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) \le -\frac{\mathtt{b}\|\mathcal{D}_0(\boldsymbol{\upsilon} - \boldsymbol{\upsilon}^*)\|^2}{2}, \qquad \boldsymbol{\upsilon} \in \Upsilon \setminus \Upsilon_0(\mathtt{r}_0), \tag{6.4}$$

holds on a random set $\Omega(\mathtt{x})$ *of dominating probability at least* $1 - 4\mathrm{e}^{-\mathtt{x}}$.

The result (6.2) is an improved version of the approximation bound obtained in [27, Theorem 3.1]. The result (6.3) can be found in [28]. The result (6.4) is very similar to [27, Theorem 4.2].

6.2. Tail posterior probability for the full parameter space. The next step in our analysis is to check that $\boldsymbol{\upsilon}$ concentrates in a small vicinity $\Upsilon_0(\mathtt{r}_0)$ of the central point $\boldsymbol{\upsilon}^*$ with a properly selected $\mathtt{r}_0$. The concentration properties of the posterior will be described by using the random quantity

$$\rho^*(\mathtt{r}_0) = \frac{\int_{\Upsilon \setminus \Upsilon_0(\mathtt{r}_0)} \exp\{L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*)\}\, d\boldsymbol{\upsilon}}{\int_{\Upsilon_0(\mathtt{r}_0)} \exp\{L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*)\}\, d\boldsymbol{\upsilon}}.$$

Theorem 6.2. *Suppose the conditions of Theorem* 6.1 *are satisfied. Then*

$$\rho^*(\mathtt{r}_0) \le \exp\{2\Delta(\mathtt{r}_0, \mathtt{x}) + \nu(\mathtt{r}_0)\}\, \mathtt{b}^{-p/2}\, \mathsf{P}\bigl(\|\boldsymbol{\gamma}\|^2 \ge \mathtt{b}\mathtt{r}_0^2\bigr) \tag{6.5}$$

on $\Omega_{\mathtt{r}_0}(\mathtt{x})$ *with*

$$\nu(\mathtt{r}_0) \stackrel{\mathrm{def}}{=} -\log \mathsf{P}\bigl(\|\boldsymbol{\gamma} + \boldsymbol{\xi}\| \le \mathtt{r}_0 \mid \mathbf{Y}\bigr).$$

If $\mathtt{r}_0 \ge z_B(\mathtt{x}) + z(p, \mathtt{x})$, *then on* $\Omega(\mathtt{x})$

$$\nu(\mathtt{r}_0) \le 2\mathrm{e}^{-\mathtt{x}}. \tag{6.6}$$

This result yields a simple sufficient condition on the value $\mathtt{r}_0$ which ensures the concentration of the posterior on $\Upsilon_0(\mathtt{r}_0)$.

Corollary 6.1. *Suppose the conditions of Theorem* 6.2 *hold. Then the additional inequality* $\mathtt{b}\mathtt{r}_0^2 \ge z^2(p, \mathtt{x} + (p/2)\log(\mathrm{e}/\mathtt{b}))$ *ensures the inequality*

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

on a random set $\Omega(\mathtt{x})$ *of probability at least* $1 - 4\mathrm{e}^{-\mathtt{x}}$.

The result follows from Theorem 6.2 by virtue of Lemma 7.2 (see below).

6.3. Tail posterior probability for the target parameter. The next major step in our analysis is to check that $\boldsymbol{\theta}$ concentrates in a small vicinity $\Theta_0(\mathtt{r}_0) = \{\boldsymbol{\theta} \colon \|\breve{D}_0(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\| \le \mathtt{r}_0\}$ of the central point $\boldsymbol{\theta}^* = \Pi_0 \boldsymbol{\upsilon}^*$ with a properly selected $\mathtt{r}_0$. The concentration properties of the posterior will be described by using the random quantity

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

In what follows we suppose that the prior is uniform, i.e., $\pi(\boldsymbol{\upsilon}) \equiv 1$, $\boldsymbol{\upsilon} \in \Upsilon$. This results in the following representation for $\rho(\mathtt{r}_0)$:

$$\rho(\mathtt{r}_0) = \frac{\int_\Upsilon \exp\{L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*)\}\,\mathbb{1}\{\boldsymbol{\theta} \notin \Theta_0(\mathtt{r}_0)\}\,d\boldsymbol{\upsilon}}{\int_\Upsilon \exp\{L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*)\}\,\mathbb{1}\{\boldsymbol{\theta} \in \Theta_0(\mathtt{r}_0)\}\,d\boldsymbol{\upsilon}}. \tag{6.7}$$

Obviously, $\mathsf{P}(\boldsymbol{\vartheta} \notin \Theta_0(\mathtt{r}_0) \mid \mathbf{Y}) \le \rho(\mathtt{r}_0)$. Therefore, small values of $\rho(\mathtt{r}_0)$ indicate a small posterior probability of the large deviation set $\{\boldsymbol{\vartheta} \notin \Theta_0(\mathtt{r}_0)\}$.

Theorem 6.3. *Suppose* (6.2) *holds. Then for* $\mathtt{b}\mathtt{r}_0^2 \ge z^2(p, \mathtt{x} + (p/2)\log(\mathrm{e}/\mathtt{b}))$ *on a set* $\Omega(\mathtt{x})$ *of probability at least* $1 - 4\mathrm{e}^{-\mathtt{x}}$

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

6.4. Local Gaussian approximation of the posterior: Upper bound. It is convenient to introduce *local conditional expectation* E°: for a random variable η, define

$$\mathsf{E}^\circ\,\eta \stackrel{\text{def}}{=} \mathsf{E}\bigl[\eta\,\mathbb{1}\{\boldsymbol{\vartheta} \in \Theta_0(\mathtt{r}_0)\} \mid \mathbf{Y}\bigr].$$

The following theorem gives an exact statement about the upper bound of this posterior expectation. Let $\boldsymbol{\theta}^\circ \stackrel{\text{def}}{=} \boldsymbol{\theta}^* + \breve{D}_0^{-1}\breve{\boldsymbol{\xi}}$, where $\breve{\boldsymbol{\xi}}$ is from (2.3).

Theorem 6.4. *Suppose* (6.2) *holds. Then on* $\Omega_{\mathtt{r}_0}(\mathtt{x})$ *for any* $f \colon \mathbb{R}^q \to \mathbb{R}_+$

$$\mathsf{E}^\circ f\bigl(\breve{D}_0(\boldsymbol{\vartheta} - \boldsymbol{\theta}^\circ)\bigr) \le \exp\{\Delta^+(\mathtt{r}_0, \mathtt{x})\}\,\mathsf{E} f(\boldsymbol{\gamma}), \tag{6.8}$$

where $\boldsymbol{\gamma} \sim \mathcal{N}(0, I_q)$ *and*

$$\Delta^+(\mathtt{r}_0, \mathtt{x}) \stackrel{\text{def}}{=} 2\Delta(\mathtt{r}_0, \mathtt{x}) + \nu(\mathtt{r}_0) + \rho_f(\mathtt{r}_0),$$

$$\rho_f(\mathtt{r}_0) \stackrel{\text{def}}{=} \frac{\int_{\Upsilon \setminus \Upsilon_0(\mathtt{r}_0)} \exp\{L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*)\} f(\breve{D}_0(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ))\,d\boldsymbol{\upsilon}}{\int_{\Upsilon_0(\mathtt{r}_0)} \exp\{L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*)\} f(\breve{D}_0(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ))\,d\boldsymbol{\upsilon}}. \tag{6.9}$$

For a random event $\eta \in A \subseteq \mathbb{R}^q$ define

$$\mathsf{P}^{\circ}(\eta \in A) = \mathsf{E}^{\circ}\, \mathbb{1}\{\eta \in A\}.$$

The next result considers special cases with $f(\mathbf{u}) = |\boldsymbol{\lambda}^{\top}\mathbf{u}|^2$ and $f(\mathbf{u}) = \mathbb{1}\{\mathbf{u} \in A\}$ for any measurable set A.

Corollary 6.2. *For any $\boldsymbol{\lambda} \in \mathbb{R}^q$, it holds on $\Omega_{\mathtt{r}_0}(\mathtt{x})$ that*

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

For any measurable set $A \subseteq \mathbb{R}^q$, it holds on $\Omega_{\mathtt{r}_0}(\mathtt{x})$ that

$$\mathsf{P}^{\circ}\bigl(\breve{D}_0(\boldsymbol{\vartheta} - \boldsymbol{\theta}^{\circ}) \in A\bigr) \le \exp\{\Delta^{+}(\mathtt{r}_0, \mathtt{x})\}\, \mathsf{P}(\boldsymbol{\gamma} \in A). \tag{6.10}$$

On $\Omega(\mathtt{x})$ one obtains $\Delta^{+}(\mathtt{r}_0, \mathtt{x}) \le 2\Delta(\mathtt{r}_0, \mathtt{x}) + 2\mathrm{e}^{-\mathtt{x}} + 2\exp\{\Delta(\mathtt{r}_0, \mathtt{x}) + 4\mathrm{e}^{-\mathtt{x}} - \mathtt{x}\}$.

The next corollary describes an upper bound for the posterior probability in the case of a change of scaling.

Corollary 6.3. *Let D_1 be a symmetric $q \times q$ matrix such that $\|I - D_1^{-1}\breve{D}_0^2 D_1^{-1}\| \le \alpha$, let $\widehat{\boldsymbol{\theta}} \in \mathbb{R}^q$ be such that $\|\breve{D}_0(\boldsymbol{\theta}^{\circ} - \widehat{\boldsymbol{\theta}})\| \le \beta$, and let $\boldsymbol{\delta}_0 \stackrel{\mathrm{def}}{=} D_1(\boldsymbol{\theta}^{\circ} - \widehat{\boldsymbol{\theta}})$. Then, for any measurable set $A \subseteq \mathbb{R}^q$, it holds on $\Omega(\mathtt{x})$ that*

$$\begin{aligned}\mathsf{P}^{\circ}\bigl(D_1(\boldsymbol{\vartheta} - \widehat{\boldsymbol{\theta}}) \in A\bigr) &\le \exp\{\Delta^{+}(\mathtt{r}_0, \mathtt{x})\}\, \mathsf{P}\bigl(D_1\breve{D}_0^{-1}\boldsymbol{\gamma} + \boldsymbol{\delta}_0 \in A\bigr)\\ &\le \exp\{\Delta^{+}(\mathtt{r}_0, \mathtt{x})\}\biggl\{\mathsf{P}(\boldsymbol{\gamma} \in A) + \frac{1}{2}\sqrt{\alpha^2 q + (1+\alpha)^2\beta^2}\biggr\}.\end{aligned} \tag{6.11}$$

6.5. Local Gaussian approximation of the posterior: Lower bound. Now we present a local lower bound for the posterior expectation:

Theorem 6.5. *Suppose* (6.2) *is satisfied. Then for any $f\colon \mathbb{R}^q \to \mathbb{R}_+$ it holds on $\Omega_{\mathtt{r}_0}(\mathtt{x})$ that*

$$\mathsf{E}^{\circ}\, f\bigl(\breve{D}_0(\boldsymbol{\vartheta} - \boldsymbol{\theta}^{\circ})\bigr) \ge \exp\{-\Delta^{-}(\mathtt{r}_0, \mathtt{x})\}\, \mathsf{E}\bigl\{f(\boldsymbol{\gamma})\,\mathbb{1}\{\|\boldsymbol{\gamma} + \breve{\boldsymbol{\xi}}\| \le \mathtt{r}_0\}\bigr\}, \tag{6.12}$$

where

$$\Delta^{-}(\mathtt{r}_0, \mathtt{x}) \stackrel{\mathrm{def}}{=} 2\Delta(\mathtt{r}_0, \mathtt{x}) + \nu(\mathtt{r}_0) + \rho^{*}(\mathtt{r}_0) + 2\widetilde{\rho}_f(\mathtt{r}_0),$$

$$\widetilde{\rho}_f(\mathtt{r}_0) \stackrel{\mathrm{def}}{=} \frac{\int_{\mathbb{R}^p \setminus \Upsilon_0(\mathtt{r}_0)} \exp\{\mathbb{L}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^{*})\} f(\breve{D}_0(\boldsymbol{\theta} - \boldsymbol{\theta}^{\circ}))\, d\boldsymbol{\upsilon}}{\int_{\Upsilon_0(\mathtt{r}_0)} \exp\{\mathbb{L}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^{*})\} f(\breve{D}_0(\boldsymbol{\theta} - \boldsymbol{\theta}^{\circ}))\, d\boldsymbol{\upsilon}}.$$

This result means that the posterior expectation has a lower bound which is nearly equal to the expectation of a function of a standard normal random variable up to (small) multiplicative and additive constants. As a corollary, we state the result for quadratic and indicator functions $f(\mathbf{u})$.

Corollary 6.4. *For any $\boldsymbol{\lambda} \in \mathbb{R}^q$ it holds on $\Omega_{\mathtt{r}_0}(\mathtt{x})$ that*

$$\mathsf{E}^{\circ}|\boldsymbol{\lambda}^{\top}\breve{D}_0(\boldsymbol{\vartheta} - \boldsymbol{\theta}^{\circ})|^2 \ge \exp\bigl\{-\Delta^{-}(\mathtt{r}_0, \mathtt{x}) + \mathrm{e}^{-\mathtt{x}}\bigr\}\|\boldsymbol{\lambda}\|^2.$$

For any measurable set $A \subseteq \mathbb{R}^q$, it holds on $\Omega_{\mathtt{r}_0}(\mathtt{x})$ that

$$\mathsf{P}^{\circ}\bigl(\breve{D}_0(\boldsymbol{\vartheta} - \boldsymbol{\theta}^{\circ}) \in A\bigr) \ge \exp\{\Delta^{-}(\mathtt{r}_0, \mathtt{x})\}\, \mathsf{P}(\boldsymbol{\gamma} \in A) - \mathrm{e}^{-\mathtt{x}}. \tag{6.13}$$

Let D_1^2 be a symmetric $q \times q$ matrix such that $\|I - D_1^{-1} \breve{D}_0^2 D_1^{-1}\| \le \alpha$, and let $\widehat{\boldsymbol{\theta}} \in \mathbb{R}^q$ be such that $\|\breve{D}_0(\boldsymbol{\theta}^\circ - \widehat{\boldsymbol{\theta}})\| \le \beta$. Define $\boldsymbol{\delta}_0 \stackrel{\text{def}}{=} D_1(\boldsymbol{\theta}^\circ - \widehat{\boldsymbol{\theta}})$. Then, for any measurable subset A in $\mathbb{R}^q$, it holds on $\Omega(\mathtt{x})$ that

$$
\begin{aligned}
\mathsf{P}^\circ\bigl(D_1(\boldsymbol{\vartheta} - \widehat{\boldsymbol{\theta}}) \in A\bigr) &\ge \exp\{\Delta^-(\mathtt{r}_0, \mathtt{x})\}\, \mathsf{P}\bigl(D_1 \breve{D}_0^{-1} \boldsymbol{\gamma} + \boldsymbol{\delta}_0 \in A\bigr) - \mathrm{e}^{-\mathtt{x}} \\
&\ge \exp\{\Delta^-(\mathtt{r}_0, \mathtt{x})\} \biggl\{ \mathsf{P}(\boldsymbol{\gamma} \in A) - \frac{1}{2}\sqrt{\alpha^2 q + (1+\alpha)^2 \beta^2} \biggr\} - \mathrm{e}^{-\mathtt{x}}. \qquad (6.14)
\end{aligned}
$$

On $\Omega(\mathtt{x})$ one obtains $\Delta^-(\mathtt{r}_0, \mathtt{x}) \le 2\Delta(\mathtt{r}_0, \mathtt{x}) + 3\mathrm{e}^{-\mathtt{x}} + 4\exp\{\Delta(\mathtt{r}_0, \mathtt{x}) + 4\mathrm{e}^{-\mathtt{x}} - \mathtt{x}\}$.

The proof of this corollary is similar to those of Corollaries 6.2 and 6.3.

7. PROOFS

This section collects the proofs of the results formulated above.

7.1. Some inequalities for the normal law. This subsection collects some simple but useful facts about the properties of the multivariate standard normal distribution. Many similar results can be found in the literature; we present the proofs to keep the presentation self-contained. Everywhere in this section $\boldsymbol{\gamma}$ means a standard normal vector in $\mathbb{R}^q$.

Lemma 7.1. *For any $\mathbf{u} \in \mathbb{R}^q$, any unit vector $\mathbf{a} \in \mathbb{R}^q$, and any $z > 0$, it holds that*

$$
\mathsf{P}\bigl(\|\boldsymbol{\gamma} - \mathbf{u}\| \ge z\bigr) \le \exp\biggl\{-\frac{z^2}{4} + \frac{q}{2} + \frac{\|\mathbf{u}\|^2}{2}\biggr\}, \qquad (7.1)
$$

$$
\mathsf{E}\bigl\{|\boldsymbol{\gamma}^\top \mathbf{a}|^2\, \mathbb{1}\{\|\boldsymbol{\gamma} - \mathbf{u}\| \ge z\}\bigr\} \le (2 + |\mathbf{u}^\top \mathbf{a}|^2) \exp\biggl\{-\frac{z^2}{4} + \frac{q}{2} + \frac{\|\mathbf{u}\|^2}{2}\biggr\}. \qquad (7.2)
$$

Proof. By the exponential Chebyshev inequality, for any $\lambda < 1$

$$
\begin{aligned}
\mathsf{P}\bigl(\|\boldsymbol{\gamma} - \mathbf{u}\| \ge z\bigr) &\le \exp\biggl\{-\frac{\lambda z^2}{2}\biggr\}\, \mathsf{E}\exp\biggl\{\frac{\lambda\|\boldsymbol{\gamma} - \mathbf{u}\|^2}{2}\biggr\} \\
&= \exp\biggl\{-\frac{\lambda z^2}{2} - \frac{q}{2}\log(1-\lambda) + \frac{\lambda}{2(1-\lambda)}\|\mathbf{u}\|^2\biggr\}.
\end{aligned}
$$

In particular, with $\lambda = 1/2$, this implies (7.1). Further, for $\|\mathbf{a}\| = 1$

$$
\begin{aligned}
\mathsf{E}\bigl\{|\boldsymbol{\gamma}^\top \mathbf{a}|^2\, \mathbb{1}\{\|\boldsymbol{\gamma} - \mathbf{u}\| \ge z\}\bigr\} &\le \exp\biggl\{-\frac{z^2}{4}\biggr\}\, \mathsf{E}\biggl\{|\boldsymbol{\gamma}^\top \mathbf{a}|^2 \exp\biggl\{\frac{\|\boldsymbol{\gamma} - \mathbf{u}\|^2}{4}\biggr\}\biggr\} \\
&\le (2 + |\mathbf{u}^\top \mathbf{a}|^2) \exp\biggl\{-\frac{z^2}{4} + \frac{q}{2} + \frac{\|\mathbf{u}\|^2}{2}\biggr\},
\end{aligned}
$$

and (7.2) follows. $\square$

The next result explains the concentration effect for the norm $\|\boldsymbol{\gamma}\|^2$ of a Gaussian vector. We use a version from [27].

Lemma 7.2. *For each $\mathtt{x}$,*

$$
\mathsf{P}\bigl(\|\boldsymbol{\gamma}\| \ge z(q, \mathtt{x})\bigr) \le \exp\{-\mathtt{x}\}, \qquad \mathsf{P}\bigl(\|\boldsymbol{\gamma}\| \le z_1(q, \mathtt{x})\bigr) \le \exp\{-\mathtt{x}\},
$$

where

$$
z^2(q, \mathtt{x}) \stackrel{\text{def}}{=} q + \sqrt{6.6 q \mathtt{x}} \vee (6.6\mathtt{x}), \qquad z_1^2(q, \mathtt{x}) \stackrel{\text{def}}{=} q - 2\sqrt{q\mathtt{x}}.
$$

The next lemma bounds from above the Kullback–Leibler divergence $\mathcal{K}(\mathsf{P}_0, \mathsf{P}_1)$ between two normal distributions P_0 and P_1.

Lemma 7.3. *Let* $\mathsf{P}_0 = \mathcal{N}(0, I_q)$ *and* $\mathsf{P}_1 = \mathcal{N}(\boldsymbol{\beta}, (U^\top U)^{-1})$ *with some nondegenerate matrix* U. *If*

$$\|U^\top U - I_q\| \le \boldsymbol{\epsilon} \le \frac{1}{2},$$

then

$$2\mathcal{K}(\mathsf{P}_0, \mathsf{P}_1) = -2\,\mathsf{E}_0 \log \frac{d\mathsf{P}_1}{d\mathsf{P}_0} \le \operatorname{tr}(U^\top U - I_q)^2 + (1+\boldsymbol{\epsilon})\|\boldsymbol{\beta}\|^2 \le \boldsymbol{\epsilon}^2 q + (1+\boldsymbol{\epsilon})\|\boldsymbol{\beta}\|^2.$$

For any measurable set $A \subset \mathbb{R}^q$, *it holds with* $\boldsymbol{\gamma} \sim \mathcal{N}(0, I_q)$ *that*

$$|\mathsf{P}_0(A) - \mathsf{P}_1(A)| = \bigl|\mathsf{P}(\boldsymbol{\gamma} \in A) - \mathsf{P}\bigl(U(\boldsymbol{\gamma} - \boldsymbol{\beta}) \in A\bigr)\bigr| \le \sqrt{\frac{\mathcal{K}(\mathsf{P}_0, \mathsf{P}_1)}{2}}.$$

Proof. We have

$$2 \log \frac{d\mathsf{P}_1}{d\mathsf{P}_0}(\boldsymbol{\gamma}) = \log \det(U^\top U) - (\boldsymbol{\gamma} - \boldsymbol{\beta})^\top U^\top U (\boldsymbol{\gamma} - \boldsymbol{\beta}) + \|\boldsymbol{\gamma}\|^2$$

with $\boldsymbol{\gamma}$ standard normal and

$$2\mathcal{K}(\mathsf{P}_0, \mathsf{P}_1) = -2\,\mathsf{E}_0 \log \frac{d\mathsf{P}_1}{d\mathsf{P}_0} = -\log \det(U^\top U) + \operatorname{tr}(U^\top U - I_q) + \boldsymbol{\beta}^\top U^\top U \boldsymbol{\beta}.$$

Let a_j be the jth eigenvalue of $U^\top U - I_q$. Then $\|U^\top U - I_q\| \le \boldsymbol{\epsilon} \le 1/2$ yields $|a_j| \le 1/2$ and

$$\begin{aligned} 2\mathcal{K}(\mathsf{P}_0, \mathsf{P}_1) &= \boldsymbol{\beta}^\top U^\top U \boldsymbol{\beta} + \sum_{j=1}^{q} \bigl\{a_j - \log(1 + a_j)\bigr\} \le (1+\boldsymbol{\epsilon})\|\boldsymbol{\beta}\|^2 + \sum_{j=1}^{q} a_j^2 \\ &\le (1+\boldsymbol{\epsilon})\|\boldsymbol{\beta}\|^2 + \operatorname{tr}(U^\top U - I_q)^2 \le (1+\boldsymbol{\epsilon})\|\boldsymbol{\beta}\|^2 + \boldsymbol{\epsilon}^2 q. \end{aligned}$$

By Pinsker's inequality, this implies

$$\sup_A |\mathsf{P}_0(A) - \mathsf{P}_1(A)| \le \sqrt{\frac{1}{2}\mathcal{K}(\mathsf{P}_0, \mathsf{P}_1)},$$

as required. $\square$

7.2. Proof of Theorem 6.2. Define $\mathfrak{u}(\boldsymbol{v}) = \mathtt{b}\|\mathcal{D}_0(\boldsymbol{v} - \boldsymbol{v}^*)\|^2/2$. Now, by a change of variables, one obtains

$$\begin{aligned} \frac{\mathtt{b}^{p/2} \det \mathcal{D}_0}{(2\pi)^{p/2}} \int\limits_{\Upsilon \setminus \Upsilon_0(\mathtt{r}_0)} \exp\{-\mathfrak{u}(\boldsymbol{v})\}\, d\boldsymbol{v} &\le \frac{\mathtt{b}^{p/2} \det \mathcal{D}_0}{(2\pi)^{p/2}} \int\limits_{\Upsilon \setminus \Upsilon_0(\mathtt{r}_0)} \exp\biggl\{-\frac{\mathtt{b}\|\mathcal{D}_0(\boldsymbol{v} - \boldsymbol{v}^*)\|^2}{2}\biggr\} d\boldsymbol{v} \\ &= \mathsf{P}\bigl(\|\boldsymbol{\gamma}\|^2 \ge \mathtt{b}\mathtt{r}_0^2\bigr). \end{aligned}$$

For the integral in the numerator of (6.7), it holds on $\Omega(\mathtt{x})$ by (6.4) that

$$\int\limits_{\Upsilon \setminus \Upsilon_0(\mathtt{r}_0)} \exp\{L(\boldsymbol{v}, \boldsymbol{v}^*)\}\, d\boldsymbol{v} \le \int\limits_{\Upsilon \setminus \Upsilon_0(\mathtt{r}_0)} \exp\{-\mathfrak{u}(\boldsymbol{v})\}\, d\boldsymbol{v}.$$

For the integral in the denominator one has

$$\int_{\Upsilon_0(\mathtt{r}_0)} \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\}\,d\boldsymbol{\upsilon} \ge \exp\{-\Delta(\mathtt{r}_0,\mathtt{x}) - m(\boldsymbol{\xi})\} \int_{\Upsilon_0(\mathtt{r}_0)} \exp\{\mathbb{L}(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*) + m(\boldsymbol{\xi})\}\,d\boldsymbol{\upsilon}. \tag{7.3}$$

By the definition of $\nu(\mathtt{r}_0)$, inequality (7.3) implies

$$\int_{\Upsilon_0(\mathtt{r}_0)} \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\}\,d\boldsymbol{\upsilon} \ge \exp\bigl\{-\Delta(\mathtt{r}_0,\mathtt{x}) - m(\boldsymbol{\xi}) - \nu(\mathtt{r}_0)\bigr\}. \tag{7.4}$$

The bound (7.4) for the local integral $\int_{\Upsilon_0(\mathtt{r}_0)} \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\}\,d\boldsymbol{\upsilon}$ implies that

$$\rho^*(\mathtt{r}_0) \le \exp\bigl\{\Delta(\mathtt{r}_0,\mathtt{x}) + \nu(\mathtt{r}_0) + m(\boldsymbol{\xi})\bigr\} \int_{\Upsilon\setminus\Upsilon_0(\mathtt{r}_0)} \exp\{-\mathfrak{u}(\boldsymbol{\upsilon})\}\,d\boldsymbol{\upsilon}.$$

Finally,

$$\exp\{m(\boldsymbol{\xi})\} = \exp\biggl\{-\frac{\|\boldsymbol{\xi}\|^2}{2}\biggr\}(2\pi)^{-p/2}\det\mathcal{D}_0 \le (2\pi)^{-p/2}\det\mathcal{D}_0,$$

and the assertion (6.5) follows. The bound (6.6) is also straightforward:

$$\begin{aligned}\nu(\mathtt{r}_0) &= -\log\mathsf{P}\bigl(\|\boldsymbol{\gamma}+\boldsymbol{\xi}\| \le \mathtt{r}_0 \mid \mathbf{Y}\bigr) \le -\log\mathsf{P}\bigl(\|\boldsymbol{\gamma}\| + \|\boldsymbol{\xi}\| \le \mathtt{r}_0 \mid \mathbf{Y}\bigr)\\ &\le -\log\mathsf{P}\bigl(\|\boldsymbol{\gamma}\| \le z(p,\mathtt{x}) \mid \mathbf{Y}\bigr) \le 2\mathrm{e}^{-\mathtt{x}}.\end{aligned}$$

7.3. Proof of Theorem 6.3. Obviously,

$$\{\boldsymbol{\theta} \notin \Theta_0(\mathtt{r}_0),\ \boldsymbol{\upsilon} \in \Upsilon\} \subset \{\Upsilon \setminus \Upsilon_0(\mathtt{r}_0)\}.$$

Therefore, it holds for the integral in the numerator of (6.7) in view of (6.4) that

$$\int_{\Upsilon} \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\}\,\mathbb{1}\{\boldsymbol{\theta} \notin \Theta_0(\mathtt{r}_0)\}\,d\boldsymbol{\upsilon} \le \int_{\Upsilon\setminus\Upsilon_0(\mathtt{r}_0)} \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\}\,d\boldsymbol{\upsilon}.$$

For the local integral in the denominator, the inclusion $\Upsilon_0(\mathtt{r}_0) \subset \bigl\{\boldsymbol{\theta} \in \Theta_0(\mathtt{r}_0), \boldsymbol{\upsilon} \in \Upsilon\bigr\}$ and (6.4) imply

$$\int_{\Upsilon} \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\}\,\mathbb{1}\{\boldsymbol{\theta} \in \Theta_0(\mathtt{r}_0)\}\,d\boldsymbol{\upsilon} \ge \int_{\Upsilon_0(\mathtt{r}_0)} \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\}\,d\boldsymbol{\upsilon}.$$

Finally,

$$\rho(\mathtt{r}_0) = \frac{\int_{\Upsilon} \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\}\,\mathbb{1}\{\boldsymbol{\theta} \notin \Theta_0(\mathtt{r}_0)\}\,d\boldsymbol{\upsilon}}{\int_{\Upsilon} \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\}\,\mathbb{1}\{\boldsymbol{\theta} \in \Theta_0(\mathtt{r}_0)\}\,d\boldsymbol{\upsilon}} \le \frac{\int_{\Upsilon\setminus\Upsilon_0(\mathtt{r}_0)} \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\}\,d\boldsymbol{\upsilon}}{\int_{\Upsilon_0(\mathtt{r}_0)} \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\}\,d\boldsymbol{\upsilon}} = \rho^*(\mathtt{r}_0),$$

and the assertion follows from Theorem 6.2.

7.4. Proof of Theorem 6.4. Note that $\mathbb{L}(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*) = \boldsymbol{\xi}^\top \mathcal{D}_0(\boldsymbol{\upsilon}-\boldsymbol{\upsilon}^*) - \|\mathcal{D}_0(\boldsymbol{\upsilon}-\boldsymbol{\upsilon}^*)\|^2/2$ is proportional to the density of a Gaussian distribution. More precisely, define

$$m(\boldsymbol{\xi}) \stackrel{\mathrm{def}}{=} -\frac{\|\boldsymbol{\xi}\|^2}{2} + \log(\det \mathcal{D}_0) - p\log\sqrt{2\pi}.$$

Then

$$m(\boldsymbol{\xi}) + \mathbb{L}(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*) = -\frac{\|\mathcal{D}_0(\boldsymbol{\upsilon}-\boldsymbol{\upsilon}^*) - \boldsymbol{\xi}\|^2}{2} + \log(\det \mathcal{D}_0) - p\log\sqrt{2\pi}$$

is (conditionally on $\mathbf{Y}$) the log-density of the normal law with the mean $\boldsymbol{\upsilon}_0 = \boldsymbol{\upsilon}^* + \mathcal{D}_0^{-1}\boldsymbol{\xi}$ and the covariance matrix $\mathcal{D}_0^{-2}$. If we integrate and leave only the $\boldsymbol{\theta}$ part of $\boldsymbol{\upsilon}$, then $m(\boldsymbol{\xi}) + \mathbb{L}(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)$ is (conditionally on $\mathbf{Y}$) the log-density of the normal law with the mean $\boldsymbol{\theta}^\circ = \breve{D}_0^{-1}\breve{\boldsymbol{\xi}} + \boldsymbol{\theta}^*$ and the covariance matrix $\breve{D}_0^{-2}$. So, for any nonnegative function $f\colon \mathbb{R}^q \to \mathbb{R}_+$ we get

$$\begin{aligned}
&\int_{\Upsilon} \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*) + m(\boldsymbol{\xi})\} f\bigl(\breve{D}_0(\boldsymbol{\theta}-\boldsymbol{\theta}^\circ)\bigr)\, d\boldsymbol{\upsilon} \\
&= \int_{\Upsilon_0(\mathtt{r}_0)} \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*) + m(\boldsymbol{\xi})\} f\bigl(\breve{D}_0(\boldsymbol{\theta}-\boldsymbol{\theta}^\circ)\bigr)\, d\boldsymbol{\upsilon} + \int_{\Upsilon\setminus\Upsilon_0(\mathtt{r}_0)} \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*) + m(\boldsymbol{\xi})\} f\bigl(\breve{D}_0(\boldsymbol{\theta}-\boldsymbol{\theta}^\circ)\bigr)\, d\boldsymbol{\upsilon} \\
&= \bigl(1+\rho_f(\mathtt{r}_0)\bigr) \int_{\Upsilon_0(\mathtt{r}_0)} \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*) + m(\boldsymbol{\xi})\} f\bigl(\breve{D}_0(\boldsymbol{\theta}-\boldsymbol{\theta}^\circ)\bigr)\, d\boldsymbol{\upsilon} \\
&\le \mathrm{e}^{\Delta(\mathtt{r}_0,\mathtt{x})+\rho_f(\mathtt{r}_0)} \int_{\Upsilon_0(\mathtt{r}_0)} \exp\{\mathbb{L}(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*) + m(\boldsymbol{\xi})\} f\bigl(\breve{D}_0(\boldsymbol{\theta}-\boldsymbol{\theta}^\circ)\bigr)\, d\boldsymbol{\upsilon} \\
&\le \mathrm{e}^{\Delta(\mathtt{r}_0,\mathtt{x})+\rho_f(\mathtt{r}_0)} \int_{\mathbb{R}^p} \exp\{\mathbb{L}(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*) + m(\boldsymbol{\xi})\} f\bigl(\breve{D}_0(\boldsymbol{\theta}-\boldsymbol{\theta}^\circ)\bigr)\, d\boldsymbol{\upsilon} \\
&= \mathrm{e}^{\Delta(\mathtt{r}_0,\mathtt{x})+\rho_f(\mathtt{r}_0)}\, \mathsf{E}\, f(\boldsymbol{\gamma}).
\end{aligned}$$

Thus,

$$\int_{\Upsilon} \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\} f\bigl(\breve{D}_0(\boldsymbol{\theta}-\boldsymbol{\theta}^\circ)\bigr)\, d\boldsymbol{\upsilon} \le \exp\bigl\{\Delta(\mathtt{r}_0,\mathtt{x}) - m(\boldsymbol{\xi}) + \rho_f(\mathtt{r}_0)\bigr\}\, \mathsf{E}\, f(\boldsymbol{\gamma}). \tag{7.5}$$

Now (7.5) and (7.4) imply

$$\frac{\int_{\Upsilon} \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\} f(\breve{D}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}-\boldsymbol{\theta}_{\boldsymbol{\epsilon}}))\, d\boldsymbol{\upsilon}}{\int_{\Upsilon} \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\}\, d\boldsymbol{\upsilon}} \le \exp\bigl\{2\Delta(\mathtt{r}_0,\mathtt{x}) + \nu(\mathtt{r}_0) + \rho_f(\mathtt{r}_0)\bigr\}\, \mathsf{E}\, f(\boldsymbol{\gamma}),$$

and (6.8) follows by the definition of $\Delta^+(\mathtt{r}_0,\mathtt{x})$.

7.5. Proof of Corollary 6.2. As a direct implication of (6.8), one easily gets

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

The only important step is to show that $\rho_{x^2}(\mathtt{r}_0)$ is small. We estimate the numerator and denominator (see (6.9)) separately. Define

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

where $\mathbf{0}$ is the zero vector of dimension $p-q$. First, for the numerator, with the use of (7.1) and (7.2) we have

$$\int\limits_{\Upsilon\setminus\Upsilon_0(\mathtt{r}_0)} |\boldsymbol{\lambda}^\top \breve{D}_0(\boldsymbol{\theta}-\boldsymbol{\theta}^\circ)|^2 \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\}\,d\boldsymbol{\upsilon} \le \int\limits_{\Upsilon\setminus\Upsilon_0(\mathtt{r}_0)} |\boldsymbol{\lambda}^\top \breve{D}_0(\boldsymbol{\theta}-\boldsymbol{\theta}^\circ)|^2 \exp\biggl\{-\frac{\mathtt{b}\|\mathcal{D}_0(\boldsymbol{\upsilon}-\boldsymbol{\upsilon}^*)\|^2}{2}\biggr\}\,d\boldsymbol{\upsilon}$$

$$= \int\limits_{\Upsilon\setminus\Upsilon_0(\mathtt{r}_0)} |\boldsymbol{\lambda}_0^\top \mathcal{D}_0(\boldsymbol{\upsilon}-\boldsymbol{\upsilon}_0)|^2 \exp\biggl\{-\frac{\mathtt{b}\|\mathcal{D}_0(\boldsymbol{\upsilon}-\boldsymbol{\upsilon}^*)\|^2}{2}\biggr\}\,d\boldsymbol{\upsilon}$$

$$= \exp\Bigl\{-\Bigl(\frac{p}{2}+2\Bigr)\log\mathtt{b} - \log(\det\mathcal{D}_0) + p\log\sqrt{2\pi}\Bigr\}\,\mathsf{E}|\boldsymbol{\lambda}_0^\top(\boldsymbol{\gamma}+\boldsymbol{\xi})|^2\,\mathbb{1}\bigl\{\|\boldsymbol{\gamma}\|^2 \ge \mathtt{b}\mathtt{r}_0^2\bigr\}$$

$$\le (4+2\|\boldsymbol{\xi}\|^2)\exp\Bigl\{-\Bigl(\frac{p}{2}+2\Bigr)\log\mathtt{b} - \log(\det\mathcal{D}_0) + p\log\sqrt{2\pi} - \frac{\mathtt{b}\mathtt{r}_0^2}{4} + \frac{p}{2}\Bigr\}\|\boldsymbol{\lambda}_0\|^2$$

$$= \exp\biggl\{2\log 2 + \log\biggl(1+\frac{\|\boldsymbol{\xi}\|^2}{2}\biggr) - \Bigl(\frac{p}{2}+2\Bigr)\log\mathtt{b} - \log(\det\mathcal{D}_0) + p\log\sqrt{2\pi} - \frac{\mathtt{b}\mathtt{r}_0^2}{4} + \frac{p}{2}\biggr\}\|\boldsymbol{\lambda}_0\|^2$$

$$\le \exp\biggl\{\frac{\|\boldsymbol{\xi}\|^2}{2} + \Bigl(\frac{p}{2}+2\Bigr)\log\frac{\mathrm{e}}{\mathtt{b}} - \log(\det\mathcal{D}_0) + p\log\sqrt{2\pi} - \frac{\mathtt{b}\mathtt{r}_0^2}{4}\biggr\}\|\boldsymbol{\lambda}_0\|^2$$

$$\le \exp\bigl\{-\log(\det\mathcal{D}_0) + p\log\sqrt{2\pi} - \mathtt{x}\bigr\}\|\boldsymbol{\lambda}_0\|^2$$

for $\mathtt{b}\mathtt{r}_0^2 \ge (2p+4)\log(\mathrm{e}/\mathtt{b}) + 2z_B(\mathtt{x}) + 4\mathtt{x}$ on $\Omega(\mathtt{x})$.

Second, for the denominator we obtain

$$\int\limits_{\Upsilon_0(\mathtt{r}_0)} |\boldsymbol{\lambda}^\top \breve{D}_0(\boldsymbol{\theta}-\boldsymbol{\theta}^\circ)|^2 \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\}\,d\boldsymbol{\upsilon} \ge \mathrm{e}^{-\Delta(\mathtt{r}_0,\mathtt{x})} \int\limits_{\Upsilon_0(\mathtt{r}_0)} |\boldsymbol{\lambda}^\top \breve{D}_0(\boldsymbol{\theta}-\boldsymbol{\theta}^\circ)|^2 \exp\{\mathbb{L}(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\}\,d\boldsymbol{\upsilon}$$

$$= \mathrm{e}^{-\Delta(\mathtt{r}_0,\mathtt{x})} \int\limits_{\Upsilon_0(\mathtt{r}_0)} |\boldsymbol{\lambda}_0^\top \mathcal{D}_0(\boldsymbol{\upsilon}-\boldsymbol{\upsilon}_0)|^2 \exp\{\mathbb{L}(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\}\,d\boldsymbol{\upsilon}$$

$$= \mathrm{e}^{-\Delta(\mathtt{r}_0,\mathtt{x})-m(\boldsymbol{\xi})}\,\mathsf{E}|\boldsymbol{\lambda}_0^\top(\boldsymbol{\gamma}+\boldsymbol{\xi})|^2\,\mathbb{1}\bigl\{\|\boldsymbol{\gamma}+\boldsymbol{\xi}\|^2 \le \mathtt{r}_0^2\bigr\}$$

$$= \mathrm{e}^{-\Delta(\mathtt{r}_0,\mathtt{x})-m(\boldsymbol{\xi})}\Bigl\{\mathsf{E}|\boldsymbol{\lambda}_0^\top(\boldsymbol{\gamma}+\boldsymbol{\xi})|^2 - \mathsf{E}|\boldsymbol{\lambda}_0^\top(\boldsymbol{\gamma}+\boldsymbol{\xi})|^2\,\mathbb{1}\bigl\{\|\boldsymbol{\gamma}+\boldsymbol{\xi}\|^2 \ge \mathtt{r}_0^2\bigr\}\Bigr\}$$

$$\ge \mathrm{e}^{-\Delta(\mathtt{r}_0,\mathtt{x})-m(\boldsymbol{\xi})}\biggl\{\|\boldsymbol{\lambda}_0\|^2 + |\boldsymbol{\lambda}_0^\top\boldsymbol{\xi}|^2 - 2\|\boldsymbol{\lambda}_0\|^2\exp\biggl\{-\frac{\mathtt{r}_0^2}{4} + \frac{p}{2} + \frac{\|\boldsymbol{\xi}\|^2}{2}\biggr\}\biggr\}$$

$$\ge \mathrm{e}^{-\Delta(\mathtt{r}_0,\mathtt{x})-m(\boldsymbol{\xi})}\|\boldsymbol{\lambda}_0\|^2\{1-2\mathrm{e}^{-\mathtt{x}}\} \ge \exp\bigl\{-\Delta(\mathtt{r}_0,\mathtt{x}) - m(\boldsymbol{\xi}) - 4\mathrm{e}^{-\mathtt{x}}\bigr\}\|\boldsymbol{\lambda}_0\|^2$$

for $\mathtt{r}_0^2 \ge 2p + 2z_B(\mathtt{x}) + 4\mathtt{x}$ on $\Omega(\mathtt{x})$.

Finally, on $\Omega(\mathtt{x})$

$$\rho_{x^2}(\mathtt{r}_0) = \frac{\int_{\Upsilon\setminus\Upsilon_0(\mathtt{r}_0)}|\boldsymbol{\lambda}^\top \breve{D}_0(\boldsymbol{\theta}-\boldsymbol{\theta}^\circ)|^2\exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\}\,d\boldsymbol{\upsilon}}{\int_{\Upsilon_0(\mathtt{r}_0)}|\boldsymbol{\lambda}^\top \breve{D}_0(\boldsymbol{\theta}-\boldsymbol{\theta}^\circ)|^2\exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\}\,d\boldsymbol{\upsilon}}$$

$$\le \frac{2\exp\{-\log(\det\mathcal{D}_0) + p\log\sqrt{2\pi} - \mathtt{x}\}\|\boldsymbol{\lambda}_0\|^2}{\exp\{-\Delta(\mathtt{r}_0,\mathtt{x}) + m(\boldsymbol{\xi}) - 4\mathrm{e}^{-\mathtt{x}}\}\|\boldsymbol{\lambda}_0\|^2}$$

$$= 2\exp\biggl\{\Delta(\mathtt{r}_0,\mathtt{x}) - \frac{\|\boldsymbol{\xi}\|^2}{2} + 4\mathrm{e}^{-\mathtt{x}} - \mathtt{x}\biggr\} \le 2\exp\{\Delta(\mathtt{r}_0,\mathtt{x}) + 4\mathrm{e}^{-\mathtt{x}} - \mathtt{x}\}.$$

7.6. Proof of Corollary 6.3. The first statement in (6.11) follows from Theorem 6.4 with $f(\mathbf{u}) = \mathbb{1}\{D_1\breve{D}_0^{-1}\mathbf{u} + \boldsymbol{\delta}_0 \in A\}$. Further, on $\Omega(\mathrm{x})$ for $\boldsymbol{\delta}_0 \stackrel{\mathrm{def}}{=} D_1(\boldsymbol{\theta}^\circ - \widehat{\boldsymbol{\theta}})$ one has

$$\|\boldsymbol{\delta}_0\|^2 = \|D_1(\boldsymbol{\theta}^\circ - \widehat{\boldsymbol{\theta}})\|^2 \le (1+\alpha)\|\breve{D}_0(\boldsymbol{\theta}^\circ - \widehat{\boldsymbol{\theta}})\|^2 \le (1+\alpha)\beta^2.$$

To prove (6.11), we compute the Kullback–Leibler divergence between two multivariate normal distributions and apply Pinsker's inequality. Let $\boldsymbol{\gamma}$ be a standard normal vector in $\mathbb{R}^q$. The random variable $D_1\breve{D}_0^{-1}\boldsymbol{\gamma} + \boldsymbol{\delta}_0$ is normal with mean $\boldsymbol{\delta}_0$ and variance $B_1^{-1} \stackrel{\mathrm{def}}{=} D_1\breve{D}_0^{-2}D_1$. Obviously,

$$\|I_q - B_1\| = \|I_q - D_1^{-1}\breve{D}_0^2 D_1^{-1}\| \le \alpha.$$

Thus, by Lemma 7.3, for any measurable set A

$$\mathsf{P}\bigl(D_1\breve{D}_0^{-1}\boldsymbol{\gamma} + \boldsymbol{\delta}_0 \in A \mid \mathbf{Y}\bigr) \le \mathsf{P}(\boldsymbol{\gamma} \in A) + \frac{1}{2}\sqrt{\alpha^2 q + (1+\alpha)^2\beta^2}.$$

7.7. Proof of Theorem 6.5. As in the proof of Theorem 6.4, for any nonnegative function $f\colon \mathbb{R}^q \to \mathbb{R}_+$, it holds that

$$\begin{aligned}
&\int_{\Upsilon} \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\} f\bigl(\breve{D}_0(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)\bigr)\,\mathbb{1}\{\boldsymbol{\theta} \in \Theta_0(\mathtt{r}_0)\}\,d\boldsymbol{\upsilon} \ge \int_{\Upsilon_0(\mathtt{r}_0)} \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\} f\bigl(\breve{D}_0(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)\bigr)\,d\boldsymbol{\upsilon} \\
&\qquad \ge \exp\{-\Delta(\mathtt{r}_0,\mathrm{x}) - m(\boldsymbol{\xi})\} \int_{\Upsilon_0(\mathtt{r}_0)} \exp\{\mathbb{L}(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\} f\bigl(\breve{D}_0(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)\bigr)\,d\boldsymbol{\upsilon} \\
&\qquad \ge \exp\{-\Delta(\mathtt{r}_0,\mathrm{x}) - m(\boldsymbol{\xi})\} \int_{\mathbb{R}^p} \exp\{\mathbb{L}(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\} f\bigl(\breve{D}_0(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)\bigr)\,d\boldsymbol{\upsilon} \\
&\qquad\quad - \exp\{-\Delta(\mathtt{r}_0,\mathrm{x}) - m(\boldsymbol{\xi})\} \int_{\mathbb{R}^p\setminus\Upsilon_0(\mathtt{r}_0)} \exp\{\mathbb{L}(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\} f\bigl(\breve{D}_0(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)\bigr)\,d\boldsymbol{\upsilon} \\
&\qquad = \exp\{-\Delta(\mathtt{r}_0,\mathrm{x}) - m(\boldsymbol{\xi})\}(1 - \widetilde{\rho}_f(\mathtt{r}_0)) \int_{\mathbb{R}^p} \exp\{\mathbb{L}(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\} f\bigl(\breve{D}_0(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)\bigr)\,d\boldsymbol{\upsilon} \\
&\qquad \ge \exp\{-\Delta(\mathtt{r}_0,\mathrm{x}) - m(\boldsymbol{\xi})\}(1 - \widetilde{\rho}_f(\mathtt{r}_0)) \int_{\Theta_0(\mathtt{r}_0)\times\mathbb{R}^{(p-q)}} \exp\{\mathbb{L}(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\} f\bigl(\breve{D}_0(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)\bigr)\,d\boldsymbol{\upsilon} \\
&\qquad \ge \exp\{-\Delta(\mathtt{r}_0,\mathrm{x}) - m(\boldsymbol{\xi}) - 2\widetilde{\rho}_f(\mathtt{r}_0)\}\,\mathsf{E}\,f(\boldsymbol{\gamma})\,\mathbb{1}\bigl\{\|\boldsymbol{\gamma} + \breve{\boldsymbol{\xi}}\| \le \mathtt{r}_0\bigr\}.
\end{aligned} \tag{7.6}$$

Here we used the fact that $1 - \alpha \ge \mathrm{e}^{-2\alpha}$ for $0 \le \alpha \le 1/2$. Similarly,

$$\begin{aligned}
\int_{\Upsilon} \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\}\,d\boldsymbol{\upsilon} &= \int_{\Upsilon_0(\mathtt{r}_0)} \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\}\,d\boldsymbol{\upsilon} + \int_{\Upsilon\setminus\Upsilon_0(\mathtt{r}_0)} \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\}\,d\boldsymbol{\upsilon} \\
&= \{1 + \rho^*(\mathtt{r}_0)\} \int_{\Upsilon_0(\mathtt{r}_0)} \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\}\,d\boldsymbol{\upsilon} \\
&\le \{1 + \rho^*(\mathtt{r}_0)\} \exp\{\Delta(\mathtt{r}_0,\mathrm{x}) - m(\boldsymbol{\xi})\}\,\mathsf{P}\bigl(\|\boldsymbol{\gamma} + \boldsymbol{\xi}\| \le \mathtt{r}_0 \mid \mathbf{Y}\bigr),
\end{aligned}$$

and finally

$$\int\limits_{\Upsilon} \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\}\,d\boldsymbol{\upsilon} \le \exp\bigl\{\Delta(\mathtt{r}_0,\mathtt{x}) - m(\boldsymbol{\xi}) + \nu(\mathtt{r}_0) + \rho^*(\mathtt{r}_0)\bigr\}. \tag{7.7}$$

The bounds (7.6) and (7.7) imply

$$\frac{\int_\Upsilon \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\} f(\breve{D}_0(\boldsymbol{\theta}-\boldsymbol{\theta}^\circ))\,d\boldsymbol{\upsilon}}{\int_\Upsilon \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\}\,d\boldsymbol{\upsilon}} \ge \frac{\exp\{-\Delta(\mathtt{r}_0,\mathtt{x}) - m(\boldsymbol{\xi}) - 2\widetilde{\rho}_f(\mathtt{r}_0)\}\,\mathsf{E}\,f(\gamma)\,\mathbb{1}\{\|\gamma+\breve{\boldsymbol{\xi}}\|\le \mathtt{r}_0\}}{\exp\{\Delta(\mathtt{r}_0,\mathtt{x}) - m(\boldsymbol{\xi}) + \nu(\mathtt{r}_0) + \rho^*(\mathtt{r}_0)\}}$$

$$\ge \exp\bigl\{-2\Delta(\mathtt{r}_0,\mathtt{x}) - 2\widetilde{\rho}_f(\mathtt{r}_0) - \nu(\mathtt{r}_0) - \rho^*(\mathtt{r}_0)\bigr\}\,\mathsf{E}\,f(\gamma)\,\mathbb{1}\bigl\{\|\gamma+\breve{\boldsymbol{\xi}}\|\le \mathtt{r}_0\bigr\}.$$

This yields (6.12).

7.8. Proof of Theorem 4.1. Due to our previous results, it is convenient to decompose the random variable $\boldsymbol{\vartheta}$ in the form

$$\boldsymbol{\vartheta} = \boldsymbol{\vartheta}\,\mathbb{1}\bigl\{\boldsymbol{\vartheta}\in\Theta_0(\mathtt{r}_0)\bigr\} + \boldsymbol{\vartheta}\,\mathbb{1}\bigl\{\boldsymbol{\vartheta}\notin\Theta_0(\mathtt{r}_0)\bigr\} = \boldsymbol{\vartheta}^\circ + \boldsymbol{\vartheta}^{\mathrm{c}}.$$

The large deviation result yields that the posterior distribution of the part $\boldsymbol{\vartheta}^{\mathrm{c}}$ is negligible under a proper choice of $\mathtt{r}_0$. Below we show that $\boldsymbol{\vartheta}^\circ$ is nearly normal, which yields the BvM result. Define

$$\bar{\boldsymbol{\vartheta}}^\circ \overset{\mathrm{def}}{=} \mathsf{E}^\circ\,\boldsymbol{\vartheta}, \qquad \mathfrak{S}_\circ^2 \overset{\mathrm{def}}{=} \mathrm{Cov}(\boldsymbol{\vartheta}^\circ) \overset{\mathrm{def}}{=} \mathsf{E}^\circ\bigl\{(\boldsymbol{\vartheta}-\bar{\boldsymbol{\vartheta}}^\circ)(\boldsymbol{\vartheta}-\bar{\boldsymbol{\vartheta}}^\circ)^\top\bigr\}.$$

It suffices to show that on $\Omega(\mathtt{x})$

$$\|\breve{D}_0(\bar{\boldsymbol{\vartheta}}^\circ - \boldsymbol{\theta}^\circ)\|^2 \le 2\Delta^*, \qquad \|I_q - \breve{D}_0\mathfrak{S}_\circ^2\breve{D}_0\| \le 2\Delta^*,$$

where $\Delta^* = \max\{\Delta^+,\Delta^-\}$ with $\Delta^+ = \Delta^+(\mathtt{r}_0,\mathtt{x})$ and $\Delta^- = \Delta^-(\mathtt{r}_0,\mathtt{x}) + \mathrm{e}^{-\mathtt{x}}$.

Consider $\boldsymbol{\eta} \overset{\mathrm{def}}{=} \breve{D}_0(\boldsymbol{\vartheta}-\boldsymbol{\theta}^\circ)$. Corollaries 6.2 and 6.4 imply that for any $\boldsymbol{\lambda}\in\mathbb{R}^q$

$$\|\boldsymbol{\lambda}\|^2\exp\{-\Delta^-\} \le \mathsf{E}^\circ|\boldsymbol{\lambda}^\top\boldsymbol{\eta}|^2 \le \|\boldsymbol{\lambda}\|^2\exp\{\Delta^+\}. \tag{7.8}$$

Define the first two moments of $\boldsymbol{\eta}$:

$$\overline{\boldsymbol{\eta}} \overset{\mathrm{def}}{=} \mathsf{E}^\circ\,\boldsymbol{\eta}, \qquad S_\circ^2 \overset{\mathrm{def}}{=} \mathsf{E}^\circ\bigl\{(\boldsymbol{\eta}-\overline{\boldsymbol{\eta}})(\boldsymbol{\eta}-\overline{\boldsymbol{\eta}})^\top\bigr\} = \breve{D}_0\mathfrak{S}_\circ^2\breve{D}_0.$$

We use the following technical statement.

Lemma 7.4. *Assume that* (7.8) *holds. Then with* $\Delta^* = \max\{\Delta^+,\Delta^-\} \le 1/2$

$$\|\overline{\boldsymbol{\eta}}\|^2 \le 2\Delta^*, \qquad \|S_\circ^2 - I_q\| \le 2\Delta^*. \tag{7.9}$$

Proof. Let $\mathbf{u}$ be any unit vector in $\mathbb{R}^q$. From (7.8) we obtain

$$\exp\{-\Delta^-\} \le \mathsf{E}^\circ|\mathbf{u}^\top\boldsymbol{\eta}|^2 \le \exp\{\Delta^+\}.$$

Note now that

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

Hence

$$\exp\{-\Delta^-\} \le \mathbf{u}^\top S_\circ^2\mathbf{u} + |\mathbf{u}^\top\overline{\boldsymbol{\eta}}|^2 \le \exp\{\Delta^+\}. \tag{7.10}$$

In a similar way with $\mathbf{u} = \overline{\boldsymbol{\eta}}/\|\overline{\boldsymbol{\eta}}\|$ and $\boldsymbol{\gamma} \sim \mathcal{N}(0, I_q)$

$$\mathsf{E}^{\circ}|\mathbf{u}^{\top}(\boldsymbol{\eta} - \overline{\boldsymbol{\eta}})|^2 \geq \mathrm{e}^{-\Delta^-}\, \mathsf{E}|\mathbf{u}^{\top}(\boldsymbol{\gamma} - \overline{\boldsymbol{\eta}})|^2 = \mathrm{e}^{-\Delta^-}\bigl(1 + \|\overline{\boldsymbol{\eta}}\|^2\bigr),$$

which yields

$$\mathbf{u}^{\top} S_{\circ}^2 \mathbf{u} \geq \bigl(1 + \|\overline{\boldsymbol{\eta}}\|^2\bigr) \exp\{-\Delta^-\}.$$

This inequality contradicts (7.10) if $\|\overline{\boldsymbol{\eta}}\|^2 > 2\Delta^* > 1$, and (7.9) follows. □

The bound for the first moment with $\boldsymbol{\vartheta}^{\circ} = \mathsf{E}^{\circ}\,\boldsymbol{\vartheta}$ implies

$$\|\breve{D}_0(\boldsymbol{\vartheta}^{\circ} - \boldsymbol{\theta}^{\circ})\|^2 \leq 2\Delta^*,$$

while the second bound yields

$$\|\breve{D}_0 \mathfrak{S}_{\circ}^2 \breve{D}_0 - I_q\| \leq 2\Delta^*.$$

The last result follows from (6.10) and (6.13) under an additional assumption that $\mathtt{x}$ is large enough to ensure $\Delta^+(\mathtt{r}_0, \mathtt{x}) \leq 2\Delta(\mathtt{r}_0, \mathtt{x}) + 5\mathrm{e}^{-\mathtt{x}}$ and $\Delta^-(\mathtt{r}_0, \mathtt{x}) \geq 2\Delta(\mathtt{r}_0, \mathtt{x}) - 8\mathrm{e}^{-\mathtt{x}}$.

7.9. Proof of Theorem 4.2. It suffices to check (4.4). First we evaluate the ratio $\pi(\boldsymbol{\upsilon})/\pi(\boldsymbol{\upsilon}^*)$ for any $\boldsymbol{\upsilon} \in \Upsilon_0(\mathtt{r}_0)$. It holds that

$$\log \frac{\pi(\boldsymbol{\upsilon})}{\pi(\boldsymbol{\upsilon}^*)} = -\frac{\|G\boldsymbol{\upsilon}\|^2}{2} + \frac{\|G\boldsymbol{\upsilon}^*\|^2}{2} = -(\boldsymbol{\upsilon} - \boldsymbol{\upsilon}^*)^{\top} G^2 \boldsymbol{\upsilon}^* - \frac{\|G(\boldsymbol{\upsilon} - \boldsymbol{\upsilon}^*)\|^2}{2}.$$

It follows from the definition of $\Upsilon_0(\mathtt{r}_0)$ and (4.6) that for $\boldsymbol{\upsilon} \in \Upsilon_0(\mathtt{r}_0)$

$$\|G(\boldsymbol{\upsilon} - \boldsymbol{\upsilon}^*)\|^2 = \|G D_0^{-1} \mathcal{D}_0(\boldsymbol{\upsilon} - \boldsymbol{\upsilon}^*)\|^2 \leq \|\mathcal{D}_0^{-1} G^2 \mathcal{D}_0^{-1}\| \mathtt{r}_0^2 \leq \epsilon^2 \mathtt{r}_0^2.$$

Similarly

$$|(\boldsymbol{\upsilon} - \boldsymbol{\upsilon}^*)^{\top} G^2 \boldsymbol{\upsilon}^*| \leq \|G\boldsymbol{\upsilon}^*\| \cdot \|G(\boldsymbol{\upsilon} - \boldsymbol{\upsilon}^*)\| \leq \|G\boldsymbol{\upsilon}^*\| \cdot \|G\mathcal{D}_0^{-1}\| \mathtt{r}_0 \leq \epsilon \mathtt{r}_0 \|G\boldsymbol{\upsilon}^*\|.$$

This obviously implies

$$-\frac{1}{2}\epsilon^2 \mathtt{r}_0^2 \leq \log \frac{\pi(\boldsymbol{\upsilon})}{\pi(\boldsymbol{\upsilon}^*)} \leq \epsilon \mathtt{r}_0 \|G\boldsymbol{\upsilon}^*\|,$$

and (4.4) follows with $\alpha(\mathtt{r}_0) = \max\{\epsilon \mathtt{r}_0 \|G\boldsymbol{\upsilon}^*\|, \epsilon^2 \mathtt{r}_0^2/2\}$.

7.10. Proof of Theorem 5.1. The bracketing bound and the large deviation result of Theorem 6.1 apply if the sample size n satisfies $n \geq \mathtt{C}(p_n + \mathtt{x})$ for a fixed constant $\mathtt{C}$. It appears that the BvM result requires a stronger condition. Indeed, in the regular i.i.d. case

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

where $a \asymp b$ means that $a = O(b)$ and $b = O(a)$ as $n \to \infty$. The radius $\mathtt{r}_0$ should satisfy the condition $\mathtt{r}_0^2 \geq \mathtt{C}(p_n + \mathtt{x})$ to ensure the large deviation result. This yields

$$\Delta(\mathtt{r}_0, \mathtt{x}) = \bigl(\delta(\mathtt{r}_0) + 3\nu_0 z_{\mathbb{H}}^2(\mathtt{x}_n)\omega\bigr)\mathtt{r}_0^2 \geq \mathtt{C}\sqrt{\frac{(p_n + \mathtt{x})^3}{n}}.$$

If we fix $\mathtt{x} = \mathtt{C}p_n$, our BvM result effectively requires the condition $p_n^3/n \to 0$ as $n \to \infty$.

7.11. Proof of Theorem 5.2. First we check that the required conditions of Subsection 4.1 are fulfilled in the considered example. This can be easily done if we slightly change the definition of the local set $\Upsilon_0(\mathtt{r}_0)$. Namely, for $\mathbf{u}^* = (u_1^*, \dots, u_{p_n}^*)^\top$, define $\Upsilon_0(\sqrt{\mathfrak{z}})$ as a rectangle

$$\Upsilon_0(\sqrt{\mathfrak{z}}) \stackrel{\text{def}}{=} \bigl\{\mathbf{u}\colon\ M_n\mathcal{K}(u_j, u_j^*) \le \mathfrak{z},\ j = 1, \dots, p_n\bigr\}.$$

Here $\mathcal{K}(u, u^*)$ is the Kullback–Leibler divergence for the Poisson family:

$$\mathcal{K}(u, u^*) = \mathrm{e}^u(u - u^*) - \mathrm{e}^u + \mathrm{e}^{u^*}.$$

Lemma 7.5. *Let $\mathfrak{z}_n$ be such that $2p_n\mathrm{e}^{-\mathfrak{z}_n} \le 1/2$. Then*

$$\mathsf{P}\bigl(\widetilde{\mathbf{u}} \in \Upsilon_0(\sqrt{\mathfrak{z}_n})\bigr) \ge 1 - 4p_n\mathrm{e}^{-\mathfrak{z}_n}. \tag{7.11}$$

In particular, the choice $\mathfrak{z}_n = \mathtt{x}_n + \log p_n$ with $\mathtt{x}_n = \mathtt{C}\log n$ provides the estimate

$$\mathsf{P}\bigl(\widetilde{\mathbf{u}} \in \Upsilon_0(\sqrt{\mathfrak{z}_n})\bigr) \ge 1 - 4\mathrm{e}^{-\mathtt{x}_n}. \tag{7.12}$$

Proof. We use the following bound from [23]:

$$\mathsf{P}\bigl(M_n\mathcal{K}(\widetilde{u}_j, u_j^*) > \mathfrak{z}_n\bigr) \le 2\mathrm{e}^{-\mathfrak{z}_n}.$$

This yields

$$\mathsf{P}\bigl(\widetilde{\mathbf{u}} \in \Upsilon_0(\sqrt{\mathfrak{z}_n})\bigr) \ge (1 - 2\mathrm{e}^{-\mathfrak{z}_n})^{p_n}.$$

Now the elementary inequalities $\log(1-\alpha) \ge -2\alpha$ for $0 \le \alpha \le 1/2$ and $\mathrm{e}^{-\delta} \ge 1 - \delta$ for $\delta \ge 0$ applied with $\alpha_n = 2\mathrm{e}^{-\mathfrak{z}_n}$ and $\delta_n = 2\alpha_n p_n$ imply

$$(1 - \alpha_n)^{p_n} = \mathrm{e}^{\log(1-\alpha_n)p_n} \ge \mathrm{e}^{-2\alpha_n p_n} \ge 1 - 2\alpha_n p_n,$$

and (7.11) follows. □

In the special case $u_1^* = \ldots = u_{p_n}^* = u^*$, the set $\Upsilon_0(\sqrt{\mathfrak{z}})$ is a cube, which can also be viewed as a ball in the sup-norm. Moreover, if $\mathfrak{z}_n/(M_n\mathrm{e}^{u^*}) \le 1/2$, then this cube is contained in the cube $\{\mathbf{u}\colon \|\mathbf{u} - \mathbf{u}^*\| \le \sqrt{\mathfrak{z}_n/(M_n\mathrm{e}^{u^*})}\}$ since $\mathrm{e}^x - 1 - x \le a^2 \le 1/2$ for $|x| \le a \le 1$. The concentration bound (7.12) enables us to check the local conditions only on the cube $\Upsilon_0(\sqrt{\mathfrak{z}_n})$. In particular, condition (ED_2) is trivially fulfilled because $\boldsymbol{\zeta}(\mathbf{u}) = \mathcal{L}(\mathbf{u}) - \mathsf{E}\,\mathcal{L}(\mathbf{u})$ is linear in $\mathbf{u}$ and θ is a linear functional of $\mathbf{u}$. Condition $(\mathcal{L}_0)$ can be checked on $\Upsilon_0(\sqrt{\mathfrak{z}_n})$ with $\delta(\mathfrak{z}_n) = \sqrt{\mathfrak{z}_n/(M_n\mathrm{e}^{u^*})}$.

It remains to compute the value $\breve{D}_0^2$. Define $\beta_n = p_n/M_n^{1/2} = p_n^{3/2}/n^{1/2}$. If $n = p_n^3$, then $\beta_n = 1$.

Lemma 7.6. *Let $v^* = 1/p_n$. Then $\breve{D}_0^2 = p_n^2\beta_n^{-2}$.*

The proof of the lemma is given in the next subsection.

Now we are ready to finalize the proof Theorem 5.2. Let β_n be bounded. The definition implies

$$p_n(\theta - \widetilde{\theta}_n) = \sum_{j=1}^{p_n} \log\frac{\upsilon_j}{Z_j/M_n}.$$

The posterior distribution $\upsilon_j \mid \mathbf{Y}$ is $\mathrm{Gamma}(\alpha_j, \mu_j)$ with $\alpha_j = 1 + Z_j$ and $\mu_j = \mu/(M_n\mu + 1)$. We use the following decomposition:

$$\frac{\upsilon_j}{Z_j/M_n} = \frac{M_n\mu_j\alpha_j}{\alpha_j - 1}\bigl(1 + \alpha_j^{-1/2}\gamma_j\bigr),$$

where $\gamma_j \stackrel{\text{def}}{=} (\alpha_j \mu_j^2)^{-1/2}(\upsilon_j - \alpha_j \mu_j)$ has zero mean and unit variance. We can use the Taylor expansion

$$p_n(\theta - \widetilde{\theta}_n) = \sum_{j=1}^{p_n} \log\Bigl(1 - \frac{1}{M_n\mu + 1}\Bigr) + \sum_{j=1}^{p_n} \log\Bigl(1 + \frac{1}{\alpha_j - 1}\Bigr) + \sum_{j=1}^{p_n} \log\Bigl(1 + \alpha_j^{-1/2}\gamma_j\Bigr).$$

Now let us take into account the properties of the real data distribution:

$$\alpha_j = \frac{M_n}{p_n}\Bigl(1 + \sqrt{\frac{p_n}{M_n}}\delta_j\Bigr),$$

where δ_j is asymptotically standard normal.

Suppose now that $\beta_n^3/\sqrt{p_n} \to 0$ as $p_n \to \infty$. Then $M_n/p_n = \bigl(\sqrt{p_n}/\beta_n^3\bigr)^{2/3} p_n^{2/3} \to \infty$ as $p_n \to \infty$. Thus, for sufficiently large p_n, $\alpha_j \approx M_n/p_n$. Moreover, for sufficiently large p_n one has $\max_{j=1,\dots,p_n} \alpha_j^{-1/2}|\gamma_j| \le 1/2$ with high probability. Below we can restrict ourselves to the case when $\alpha_j^{-1/2}|\gamma_j| \le 1/2$. This allows us to use the Taylor expansion

$$\begin{aligned} p_n(\theta - \widetilde{\theta}_n) &= \sum_{j=1}^{p_n} \log\Bigl(1 - \frac{1}{M_n\mu + 1}\Bigr) + \sum_{j=1}^{p_n} \log\Bigl(1 + \frac{1}{\alpha_j - 1}\Bigr) + \sum_{j=1}^{p_n} \log\Bigl(1 + \frac{\gamma_j}{\sqrt{\alpha_j}}\Bigr) \\ &= \sum_{j=1}^{p_n} \frac{1}{\alpha_j - 1} + \sum_{j=1}^{p_n} \frac{1}{\sqrt{\alpha_j}}\gamma_j - \sum_{j=1}^{p_n} \frac{1}{2\alpha_j}\gamma_j^2 + R. \end{aligned}$$

One can easily check that the remainder R is of order $\beta_n^3/\sqrt{p_n} \to 0$. Moreover, $p_n^{-1/2}\sum_{j=1}^{p_n}\gamma_j$ is asymptotically standard normal, while $p_n^{-1}\sum_{j=1}^{p_n}\gamma_j^2 \xrightarrow{\mathsf{P}} 1$. The conditions of the central limit theorem can be easily checked here because the Lyapunov condition is valid. Also $\sum_{j=1}^{p_n}(\alpha_j - 1)^{-1} = p_n^2/M_n + o_n(\beta_n^2)$.

Now consider the case $\beta_n \to 0$:

$$\beta_n^{-1} p_n(\theta - \widetilde{\theta}_n) = \beta_n + \frac{1}{\sqrt{p_n}} \sum_{j=1}^{p_n} \gamma_j - \frac{\beta_n}{2p_n} \sum_{j=1}^{p_n} \gamma_j^2 + o_n(1) \xrightarrow{\text{w}} \mathcal{N}(0,1).$$

Similarly, with $\beta_n \equiv \beta$,

$$\beta^{-1} p_n(\theta - \widetilde{\theta}_n) = \beta + \frac{1}{\sqrt{p_n}} \sum_{j=1}^{p_n} \gamma_j - \frac{\beta_n}{2p_n} \sum_{j=1}^{p_n} \gamma_j^2 + o_n(1) \xrightarrow{\text{w}} \mathcal{N}(\beta/2,1).$$

This proves the result for $\beta_n \equiv \beta$.

Finally, in the case when β_n grows to infinity but $\beta_n^3/\sqrt{p_n} \to 0$, one has $\beta_n^{-1}(\theta - \widetilde{\theta}_n) \xrightarrow{\mathsf{P}} \infty$.

7.12. Proof of Lemma 7.6. Let $\bar{u}_j = u_j - u_j^*$. Then

$$L(\mathbf{u}, \mathbf{u}^*) = \mathcal{L}(\mathbf{u}) - \mathcal{L}(\mathbf{u}^*) = \sum_{j=1}^{p_n} \bigl\{ Z_j \bar{u}_j - M_n p_n^{-1}(\mathrm{e}^{\bar{u}_j} - 1) \bigr\}.$$

The expected value of Z_j is M_n/p_n, which leads to the following expectation of likelihood:

$$\mathsf{E}\, L(\mathbf{u}, \mathbf{u}^*) = \frac{M_n}{p_n} \sum_{j=1}^{p_n} \bigl(\bar{u}_j - (e^{\bar{u}_j} - 1)\bigr) = -\frac{M_n}{p_n} \sum_{j=1}^{p_n} \frac{\bar{u}_j^2}{2} + O\bigl(\|\bar{\mathbf{u}}\|^3\bigr).$$

Then we make the substitution $\bar{u}_1 = p_n\bar{\theta} - \sum_{j=2}^{p_n} \bar{u}_j$, where $\bar{\theta} = \theta - \theta^*$. Thus we get

$$\mathsf{E}\, L(\mathbf{u}, \mathbf{u}^*) = -\frac{M_n}{p_n}\frac{1}{2}\left(p_n\bar{\theta} - \sum_{j=2}^{p_n} \bar{u}_j\right)^2 - \frac{M_n}{p_n}\sum_{j=2}^{p_n}\frac{\bar{u}_j^2}{2} + O\big(\|\bar{\mathbf{u}}\|^3\big).$$

This Taylor expansion allows us to compute the components of the Fisher information matrix:

$$\mathcal{D}_0^2 = -\nabla^2\,\mathsf{E}\,\mathcal{L}(\mathbf{u}^*) = \frac{M_n}{p_n}\begin{pmatrix} p_n^2 & -p_n & \dots & \dots & -p_n \\ -p_n & 2 & 1 & \dots & 1 \\ \vdots & 1 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 1 \\ -p_n & 1 & \dots & 1 & 2 \end{pmatrix}.$$

The Fisher information for the target parameter θ can be computed as follows:

$$\breve{D}_0^2 = M_n p_n(1 - \mathbf{e}^\top \mathbb{H}^{-1}\mathbf{e}),$$

where $\mathbf{e} = (1, \dots, 1)^\top$ and $\mathbb{H} = I + E$ with $E = \mathbf{e}\mathbf{e}^\top$ being the $(p_n - 1) \times (p_n - 1)$ matrix of ones.

It follows that

$$\mathbf{e}^\top \mathbb{H}^{-1}\mathbf{e} = \operatorname{tr}(\mathbf{e}^\top \mathbb{H}^{-1}\mathbf{e}) = \operatorname{tr}(\mathbb{H}^{-1}\mathbf{e}\mathbf{e}^\top) = \operatorname{tr}\big((E + I)^{-1}E\big).$$

Further, since $(E + I)^{-1}E = I - (E + I)^{-1}$, one has

$$\mathbf{e}^\top \mathbb{H}^{-1}\mathbf{e} = \operatorname{tr}\{I - (E + I)^{-1}\} = (p_n - 1) - \operatorname{tr}\{(E + I)^{-1}\} = (p_n - 1) - \sum_{j=1}^{p_n}\lambda_j,$$

where λ_j are the eigenvalues of the matrix $(E + I)^{-1}$. It is easy to see that $\lambda_1 = p_n^{-1}$ while $\lambda_2 = \ldots = \lambda_{p_n - 1} = 1$. Thus

$$\mathbf{e}^\top \mathbb{H}^{-1}\mathbf{e} = (p_n - 1) - \{p_n^{-1} + (p_n - 2)\} = 1 - p_n^{-1},$$

$$\breve{D}_0^2 = M_n p_n(1 - \mathbf{e}^\top \mathbb{H}^{-1}\mathbf{e}) = M_n p_n\{1 - (1 - p_n^{-1})\} = M_n = p_n^2\beta_n^{-2},$$

which completes the proof.

ACKNOWLEDGMENTS

The work was partially supported by the Laboratory for Structural Methods of Data Analysis in Predictive Modeling, Moscow Institute of Physics and Technology (contract no. 11.G34.31.0073 with the Ministry of Education and Science of the Russian Federation). The second author was also supported by the German Research Foundation (DFG) through the Collaborative Research Center 649 “Economic Risk.”

REFERENCES

1. A. Barron, M. J. Schervish, and L. Wasserman, “The consistency of posterior distributions in nonparametric problems,” Ann. Stat. **27** (2), 536–561 (1996).
2. P. J. Bickel and B. J. K. Kleijn, “The semiparametric Bernstein–von Mises theorem,” Ann. Stat. **40** (1), 206–237 (2012).
3. N. Bochkina and P. J. Green, “The Bernstein–von Mises theorem and nonregular models,” Ann. Stat. **42** (5), 1850–1878 (2014).

4. D. Bontemps, "Bernstein–von Mises theorems for Gaussian regression with increasing number of regressors," Ann. Stat. **39** (5), 2557–2584 (2011).
5. S. Boucheron and E. Gassiat, "A Bernstein–von Mises theorem for discrete probability distributions," Electron. J. Stat. **3**, 114–148 (2009).
6. S. Boucheron and P. Massart, "A high-dimensional Wilks phenomenon," Probab. Theory Relat. Fields **150** (3–4), 405–433 (2011).
7. I. Castillo, "A semiparametric Bernstein–von Mises theorem for Gaussian process priors," Probab. Theory Relat. Fields **152** (1–2), 53–99 (2012).
8. I. Castillo and R. Nickl, "Nonparametric Bernstein–von Mises theorems in Gaussian white noise," Ann. Stat. **41** (4), 1999–2028 (2013).
9. I. Castillo and J. Rousseau, "A general Bernstein–von Mises theorem in semiparametric models," arXiv: 1305.4482v1 [math.ST].
10. G. Cheng and M. R. Kosorok, "General frequentist properties of the posterior profile distribution," Ann. Stat. **36** (4), 1819–1853 (2008).
11. V. Chernozhukov and H. Hong, "An MCMC approach to classical estimation," J. Econom. **115** (2), 293–346 (2003).
12. D. D. Cox, "An analysis of Bayesian inference for nonparametric regression," Ann. Stat. **21** (2), 903–923 (1993).
13. D. Freedman, "On the Bernstein–von Mises theorem with infinite-dimensional parameters," Ann. Stat. **27** (4), 1119–1140 (1999).
14. S. Ghosal, "Asymptotic normality of posterior distributions in high-dimensional linear models," Bernoulli **5** (2), 315–331 (1999).
15. S. Ghosal, "Asymptotic normality of posterior distributions for exponential families when the number of parameters tends to infinity," J. Multivariate Anal. **74** (1), 49–68 (2000).
16. I. M. Johnstone, "High dimensional Bernstein–von Mises: simple examples," in *Borrowing Strength: Theory Powering Applications—A Festschrift for Lawrence D. Brown* (Inst. Math. Stat., Beachwood, OH, 2010), Inst. Math. Stat. Collect. **6**, pp. 87–98.
17. Y. Kim, "The Bernstein–von Mises theorem for the proportional hazard model," Ann. Stat. **34** (4), 1678–1700 (2006).
18. Y. Kim and J. Lee, "A Bernstein–von Mises theorem in the nonparametric right-censoring model," Ann. Stat. **32** (4), 1492–1512 (2004).
19. B. J. K. Kleijn and A. W. van der Vaart, "Misspecification in infinite-dimensional Bayesian statistics," Ann. Stat. **34** (2), 837–877 (2006).
20. B. J. K. Kleijn and A. W. van der Vaart, "The Bernstein–von Mises theorem under misspecification," Electron. J. Stat. **6**, 354–381 (2012).
21. L. Le Cam and G. L. Yang, *Asymptotics in Statistics: Some Basic Concepts* (Springer, New York, 1990), Springer Ser. Stat.
22. H. Leahu, "On the Bernstein–von Mises phenomenon in the Gaussian white noise model," Electron. J. Stat. **5**, 373–404 (2011).
23. J. Polzehl and V. Spokoiny, "Propagation–separation approach for local likelihood estimation," Probab. Theory Relat. Fields **135** (3), 335–362 (2006).
24. V. Rivoirard and J. Rousseau, "Bernstein–von Mises theorem for linear functionals of the density," Ann. Stat. **40** (3), 1489–1523 (2012).
25. L. Schwartz, "On Bayes procedures," Z. Wahrscheinlichkeitstheor. Verw. Geb. **4** (1), 10–26 (1965).
26. X. Shen, "Asymptotic normality of semiparametric and nonparametric posterior distributions," J. Am. Stat. Assoc. **97** (457), 222–235 (2002).
27. V. Spokoiny, "Parametric estimation. Finite sample theory," Ann. Stat. **40** (6), 2877–2909 (2012); arXiv: 1111.3029 [math.ST].
28. V. Spokoiny and M. Zhilova, "Sharp deviation bounds for quadratic forms," Math. Methods Stat. **22** (2), 100–113 (2013); arXiv: 1302.1699 [math.PR].
29. A. W. van der Vaart, *Asymptotic Statistics* (Cambridge Univ. Press, Cambridge, 2000), Cambridge Ser. Stat. Probab. Math. **3**.

Translated by the authors