Critical dimension in semiparametric Bernstein - von Mises Theorem

Maxim Panov*

Moscow Institute of Physics and Technology,
Institute for Information Transmission Problems of RAS,
Datadvance Company,
Pokrovsky blvd. 3 building 1B,
109028 Moscow, Russia
maxim.panov@datadvance.net

Vladimir Spokoiny†

Weierstrass Institute and
Humboldt University Berlin,
Moscow Institute of Physics and Technology
Mohrenstr. 39,
10117 Berlin, Germany
spokoiny@wias-berlin.de

August 23, 2013

*The author is partially supported by Laboratory for Structural Methods of Data Analysis in Predictive Modeling, MIPT, RF government grant, ag. 11.G34.31.0073.

†The author is partially supported by Laboratory for Structural Methods of Data Analysis in Predictive Modeling, MIPT, RF government grant, ag. 11.G34.31.0073. Financial support by the German Research Foundation (DFG) through the Collaborative Research Center 649 "Economic Risk" is grate-

Abstract

The classical Bernstein- von Mises (BvM) result is considered in finite sample non-asymptotic setup. The approach follows the idea of Spokoiny (2012) and it is mainly based on the local majorization device which is introduced in that paper. The main attention is paid to notion of *critical dimension* which determines maximum allowed problem dimension (p) given sample size (n). It is shown that condition "(p^3/n) is small" is sufficient for BvM result to be valid. Also we provide an example which shows the failure of BvM result in case, when this condition is not satisfied .

AMS 2000 Subject Classification: Primary 62F15. Secondary 62F25,62H12
Keywords: bayesian inference, local quadratic majorization, concentration, semiparametric

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. Also it justifies 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 Le Cam and Yang (1990); van der Vaart (2000) for a detailed overview. However, in modern statistic applications one often faces very complicated models involving a lot of parameters and with a limited sample size. This requires an extension of the classical results to such non-classical situation. We mention Cox (1993); Freedman (1999); Ghosal (1999); Johnstone (2010) and references therein for some special phenomena arising in the Bayesian analysis when the parameter dimension increases. Already consistency of the posterior distribution in nonparametric and semiparametric models is a nontrivial problem; cf. Schwartz (1965) and Barron et al. (1996). Asymptotic normality of the posterior measure for these classes of models is even more challenging; see e.g. Shen (2002). Some results for particular semi and nonparametric problems are available from Kim and Lee (2004); Kim (2006). Cheng and Kosorok (2008) obtained a version of the BvM statement based on a high order expansion of the profile sampler. The recent paper

fully acknowledged.

Bickel and Kleijn (2012) extends the BvM statement from the classical parametric case to a rather general i.i.d. framework. Castillo (2012) studies the semiparametric BvM result for Gaussian process functional priors. In Rivoirard and Rousseau (2012) semiparametric BvM theorem is derived for linear functionals of density and in forthcoming work Castillo and Rousseau (2013) the 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.

In this paper we reconsider the BvM result for a general parametric 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 the most of notions, methods and tools in the classical theory are formulated in the asymptotic setup with the growing sample size. Only few finite sample general results are available; see e.g. the recent paper Boucheron and Massart (2011). This paper focuses on the semiparametric problem when the full parameter is large or infinite dimensional but the target is low dimension. In the Bayesian framework, the aim is the marginal of the posterior corresponding to the target parameter; cf. Castillo (2012). Typical examples are provided by functional estimation, estimation of a function at a point, or simply by estimating 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. Bickel and Kleijn (2012). The methods of study heavily rely 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 Cheng and Kosorok (2008) or Bickel and Kleijn (2012). This essentially simplifies the study but does not allow 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 Ghosal (1999, 2000), Boucheron and Gassiat (2009), Johnstone (2010) for specific examples. See the discussion after Theorem 3.1 below for more details.

In this paper we show that the *bracketing* approach of Spokoiny (2012) can be used for obtaining a finite sample semiparametric version of Bernstein – von Mises 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 parameter nor it involves the notion of the hardest parametric submodel. 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 to 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 $\boldsymbol{Y}$ denote the observed random data, and $I\!P$ denote the data distribution. The parametric statistical model assumes that the unknown data distribution $I\!P$ belongs to a given parametric family $(I\!P_{\boldsymbol{\upsilon}})$:

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

where Υ is some parameter space and $\boldsymbol{\upsilon}^* \in \Upsilon$ is the true value of 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

$$\boldsymbol{\theta}^* \stackrel{\text{def}}{=} P\boldsymbol{\upsilon}^*,$$

for some mapping $P : \Upsilon \to I\!R^q$, and $q \in \mathbb{N}$ stands for the dimension of the target. Usually in the classical semiparametric setup, the vector $\boldsymbol{\upsilon}$ is 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 $(\boldsymbol{\theta}, \boldsymbol{\eta})$-setup and our presentation follows this setting. An extension to the $\boldsymbol{\upsilon}$-setup with $\boldsymbol{\theta} = P\boldsymbol{\upsilon}$ 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 dimensionality p.

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

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

where $\mathcal{L}(\boldsymbol{\upsilon}) = \log \frac{dI\!P_{\boldsymbol{\upsilon}}}{d\boldsymbol{\mu}_0}(\boldsymbol{Y})$ is the log-likelihood function of the family $(I\!P_{\boldsymbol{\upsilon}})$ for some dominating measure $\boldsymbol{\mu}_0$. Under model misspecification, $\boldsymbol{\upsilon}^*$ defines the best parametric fit to $I\!P$ by the considered family; cf. Chernozhukov and Hong (2003), Kleijn and van der Vaart (2006, 2012) and references therein. The target $\boldsymbol{\theta}^*$ is defined by the mapping P:

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

2 Main results

In what follows, by $\mathtt{C}$ we denote a generic fixed constant which does not depend on the dimensions q, p of the target and of the full parameter. In most of statements below this constant can be made explicit. We also 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

$$I\!\!P\bigl(\Omega(\mathtt{x})\bigr) \geq 1 - \mathtt{C}e^{-\mathtt{x}}.$$

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

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

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 $\nabla^2 I\!\!E \mathcal{L}(\boldsymbol{\upsilon})$ for the Hessian of the expectation $I\!\!E \mathcal{L}(\boldsymbol{\upsilon})$. Also denote $\nabla \stackrel{\text{def}}{=} \nabla \mathcal{L}(\boldsymbol{\upsilon}^*)$ and define the score vector

$$\boldsymbol{\xi} \stackrel{\text{def}}{=} \mathcal{D}_0^{-1} \nabla. \tag{2.2}$$

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

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

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

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

$$\begin{aligned} \breve{D}_0^2 &= D_0^2 - A_0 H_0^{-2} A_0^\top, \\ \breve{\nabla}_{\boldsymbol{\theta}} &= \nabla_{\boldsymbol{\theta}} - A_0 H_0^{-2} \nabla_{\boldsymbol{\eta}}, \\ \breve{\boldsymbol{\xi}} &= \breve{D}_0^{-1} \breve{\nabla}_{\boldsymbol{\theta}}. \end{aligned} \tag{2.3}$$

The $q \times q$ matrix $\breve{D}_0^2$ is usually called the efficient Fisher information matrix, while the random vector $\breve{\boldsymbol{\xi}} \in I\!\!R^q$ is the efficient score.

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 $\boldsymbol{Y}$ and obtained by normalization of the product $\exp\bigl\{\mathcal{L}(\boldsymbol{\upsilon})\bigr\} \Pi(d\boldsymbol{\upsilon})$. This relation is usually written as

$$\boldsymbol{\upsilon} \mid \boldsymbol{Y} \propto \exp\bigl\{\mathcal{L}(\boldsymbol{\upsilon})\bigr\} \, \Pi(d\boldsymbol{\upsilon}). \tag{2.4}$$

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 (2.4) describes a *quasi posterior*; Chernozhukov and Hong (2003). 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 the other important characteristics of the model.

We suppose that the prior measure Π has a positive density $\pi(\boldsymbol{\upsilon})$ w.r.t. to the Lebesgue measure on Υ: $\Pi(d\boldsymbol{\upsilon}) = \pi(\boldsymbol{\upsilon})d\boldsymbol{\upsilon}$. Then (2.4) can be written as

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

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 MLE) and scaled by the total Fisher information matrix is nearly standard normal:

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

where I_p is an 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 and credible sets.

In this work we investigate the properties of the posterior distribution for the target parameter $\boldsymbol{\vartheta} = P\boldsymbol{\upsilon}$. In this case (2.4) can be written as

$$\boldsymbol{\vartheta} \mid \boldsymbol{Y} \propto \int \exp\bigl\{\mathcal{L}(\boldsymbol{\upsilon})\bigr\}\,\pi(\boldsymbol{\upsilon})d\boldsymbol{\eta}. \tag{2.5}$$

The BvM result in this case transforms into

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

where I_q is an identity matrix of dimension q, $\widetilde{\boldsymbol{\theta}} = P\widetilde{\boldsymbol{\upsilon}}$, and $\breve{D}_0^2$ is given in (2.3).

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 target parameter (2.5) 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.

First we state the BvM result about the properties of the $\boldsymbol{\vartheta}$-posterior given by (2.5) in case of uniform prior, that is, $\pi(\boldsymbol{\upsilon}) \equiv 1$ on $\varUpsilon$. Define

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

Also define

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

Below we present a version of the BvM result in the considered nonasymptotic setup which claims that $\overline{\boldsymbol{\vartheta}}$ is close to $\breve{\boldsymbol{\theta}}$, $\mathfrak{S}^2$ is nearly equal to $\breve{D}_0^{-2}$, and $\breve{D}_0(\boldsymbol{\vartheta} - \breve{\boldsymbol{\theta}})$ is nearly standard normal conditionally on $\boldsymbol{Y}$. 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 $\varOmega(\mathtt{x})$ we denote a random event of dominating probability with $I\!\!P\bigl(\varOmega(\mathtt{x})\bigr) \ge 1 - \mathtt{C}\mathrm{e}^{-\mathtt{x}}$.

Our results require some conditions to be satisfied, see Section 4.1 for a precise formulation. The conditions include among others the quantity $\boldsymbol{\epsilon}$ which is defined in terms of the imposed conditions on the model; see (4.3). This value is rather small in typical situations. For example, in i.i.d case it holds $\boldsymbol{\epsilon} = \mathtt{C}\sqrt{(p + \mathtt{x})/n}$.

Theorem 2.1. *Suppose the conditions of Section 4.1. Define the random quantity*

$$\mathfrak{p} \stackrel{\text{def}}{=} p + \|\boldsymbol{\xi}\|^2 + \mathtt{x},$$

where $\boldsymbol{\xi}$ *is defined by* (2.2). *There exists a random event* $\varOmega(\mathtt{x})$ *of a dominating probability such that it holds on* $\varOmega(\mathtt{x})$

$$\|\breve{D}_0(\overline{\boldsymbol{\vartheta}} - \breve{\boldsymbol{\theta}})\|^2 \le \mathtt{C}\,\boldsymbol{\epsilon}\,\mathfrak{p},$$

$$\bigl\| I_q - \breve{D}_0 \mathfrak{S}^2 \breve{D}_0 \bigr\|_{\infty} \le \mathtt{C}\,\boldsymbol{\epsilon}\,\mathfrak{p},$$

and also $\|\boldsymbol{\xi}\|^2 \le \mathtt{C}(p + \mathtt{x})$. *Moreover, for any measurable set* $A \subset I\!\!R^q$

$$\begin{aligned} &\exp\bigl(-\mathtt{C}\,\boldsymbol{\epsilon}\,\mathfrak{p}\bigr)\Bigl\{ I\!\!P\bigl(\boldsymbol{\gamma} \in A\bigr) - \sqrt{\mathtt{C}(\boldsymbol{\epsilon}\,\mathfrak{p} + \boldsymbol{\epsilon}^2\,\mathfrak{p}^2\,q)} \Bigr\} - \mathtt{C}\mathrm{e}^{-\mathtt{x}} \\ &\quad \le I\!\!P\bigl(\mathfrak{S}^{-1}(\boldsymbol{\vartheta} - \overline{\boldsymbol{\vartheta}}) \in A \,\big|\, \boldsymbol{Y}\bigr) \\ &\qquad \le \exp\bigl(\mathtt{C}\,\boldsymbol{\epsilon}\,\mathfrak{p}\bigr)\Bigl\{ I\!\!P\bigl(\boldsymbol{\gamma} \in A\bigr) + \sqrt{\mathtt{C}(\boldsymbol{\epsilon}\,\mathfrak{p} + \boldsymbol{\epsilon}^2\,\mathfrak{p}^2\,q)} \Bigr\} + \mathtt{C}\mathrm{e}^{-\mathtt{x}}. \end{aligned} \tag{2.6}$$

If q is fixed then the condition "$\boldsymbol{\epsilon}\,\mathfrak{p}$ 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. All the error terms are given explicitly up to absolute constants. Moreover, the statement can be extended to situations when q grows but $\boldsymbol{\epsilon}^2\,\mathfrak{p}^2\,q$ 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 Section 4.3 for details.

3 The i.i.d. case and critical dimension

This section comments how the previously obtained general results 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: once checked one can directly apply any of the mentioned results. Typical examples include i.i.d., GLM, 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.

Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$ be an i.i.d. sample from a measure P. Here we suppose the conditions of Section 5.1 in Spokoiny (2012) on P and $(P_{\boldsymbol{\upsilon}})$ to be fulfilled. We admit that the parametric assumption $P \in (P_{\boldsymbol{\upsilon}}, \boldsymbol{\upsilon} \in \varUpsilon)$ 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 3.1. *Suppose the conditions of Theorem 5.1 in Spokoiny (2012). Let also* $p_n \to \infty$ *and* $p_n^3/n \to 0$. *Then the result of Theorem 2.1 holds with* $\boldsymbol{\epsilon} = \mathtt{C}\sqrt{p_n/n}$, $\mathcal{D}_0^2 = n\mathbb{F}_{\boldsymbol{\upsilon}^*}$, *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 Ghosal (1999). However, the convergence is proved under the condition $p_n^4 \log(p_n)/n \to 0$ which appears to be too strong. Ghosal (2000) showed that the dimensionality constraint can be relaxed to $p_n^3/n \to 0$ for exponential models with a product structure. Boucheron and Gassiat (2009) proved the BvM result in a specific class of i.i.d. model 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 Johnstone (2010).

3.1 Critical dimension

This section 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 3.1 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 $p_n^3/n \geq \beta^2 > 0$ 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 $\mathcal{I}_j \stackrel{\text{def}}{=} \{i : \lceil i/M_n \rceil = j\}$ and $j = 1, \ldots, p_n$. Let also $u_j = \log \upsilon_j$ be the canonical parameter. The log-likelihood $\mathcal{L}(\boldsymbol{u})$ with $\boldsymbol{u} = (u_1, \ldots, u_{p_n})$ reads as

$$\mathcal{L}(\boldsymbol{u}) = \sum_{j=1}^{p_n} \bigl(Z_j u_j - M_n e^{u_j}\bigr),$$

where

$$Z_j \stackrel{\text{def}}{=} \sum_{i \in \mathcal{I}_j} Y_i .$$

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

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

Below we study this problem in the asymptotic setup with $p_n \to \infty$ as $n \to \infty$ when the underlying measure $I\!\!P$ corresponds to $u_1^* = \ldots = u_{p_n}^* = u^*$ for some u^* yielding $\theta^* = u^*$. The value u^* will be specified later. Define $\beta_n = p_n / M_n^{1/2} = p_n^{3/2}/n^{1/2}$. If $n = p_n^3$, then $\beta_n = 1$. We consider an i.i.d. exponential prior on the parameters υ_j of Poisson distribution:

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

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

$$\upsilon_j \bigm| \boldsymbol{Y} \sim \text{Gamma}(\alpha_j, \mu_j),$$

where $\alpha_j = 1 + \sum_{i \in \mathcal{I}_j} Y_i$, $\mu_j = \frac{\mu}{M_n \mu + 1}$.

First we describe the profile maximum likelihood estimator $\widetilde{\theta}_n$ 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

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

Furthermore, the efficient Fisher information $\breve{D}_0^2$ is equal to $\beta_n^{-2} p_n^2$; see Lemma 5.4 below.

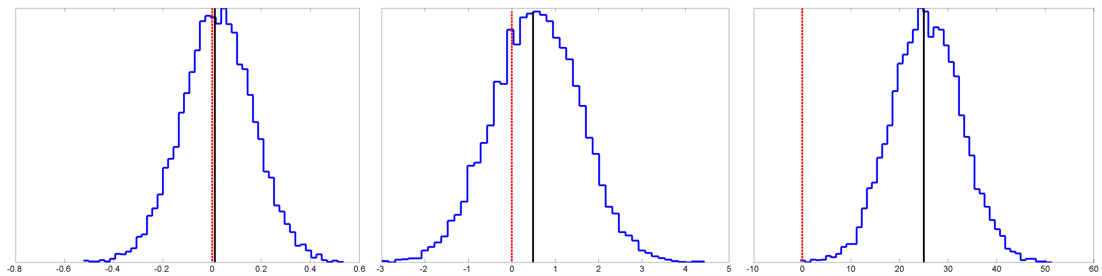

Figure 3.1: Posterior distribution of $\beta_n^{-1} p_n \bigl(\theta - \widetilde{\theta}_n\bigr)$ for $\beta_n = 1/\log(p_n)$, $\beta_n = 1$, and $\beta_n = \log(p_n)$. Solid line is for posterior mean and dashed line is for true mean.

Theorem 3.2. *Let* $Y_i \sim \operatorname{Poisson}(v^*)$ *for all* $i = 1, \ldots, n$, $v^* = 1/p_n$. *Then*

1. *If* $\beta_n \to 0$ *as* $p_n \to \infty$, *then*

$$\beta_n^{-1} p_n \bigl(\theta - \widetilde{\theta}_n\bigr) \, \big| \, \boldsymbol{Y} \xrightarrow{w} \mathcal{N}(0, 1).$$

2. *Let* $\beta_n \equiv \beta > 0$. *Then*

$$\beta^{-1} p_n \bigl(\theta - \widetilde{\theta}_n\bigr) \, \big| \, \boldsymbol{Y} \xrightarrow{w} \mathcal{N}(\beta/2, 1).$$

3. *If* $\beta_n \to \infty$ *but* $\beta_n^3/\sqrt{p_n} \to 0$, *then*

$$\beta_n^{-1} p_n \bigl(\theta - \widetilde{\theta}_n\bigr) \, \big| \, \boldsymbol{Y} \xrightarrow{w} \infty.$$

We carried out a series of experiments to numerically demonstrate the results of Theorem 3.2. The dimension of parameter space was fixed $p_n = 10000$. Three cases were considered:

1. $\beta_n = \frac{1}{\log p_n}$, which corresponds to $\beta_n \to 0, n \to \infty$.

2. $\beta_n = 1$.

3. $\beta_n = \log p_n$, which corresponds to $\beta_n \to \infty, n \to \infty$.

For each sample 10000 realizations of $\boldsymbol{Y}$ were generated from the exponential distribution $\operatorname{Exp}(v_*)$ and so were corresponding posterior values $\theta \, \big| \, \boldsymbol{Y}$. The resulting posterior distribution for three cases is demonstrated on Figure 3.1. It can be easily seen that results of Theorem 3.2 are numerically confirmed.

4 Supplementary

This section contains the imposed conditions and some supplementary statements which are of some interest by itself.

4.1 Conditions

Our approach assumes a number of conditions to be satisfied. The list is essentially as in Spokoiny (2012), one can find there some discussion and examples showing that the conditions are not restrictive and are fulfilled in most of classical models used in statistical studies like 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 $\mathtt{r}_0$:

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

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

$$\zeta(\boldsymbol{\upsilon}) \stackrel{\text{def}}{=} \mathcal{L}(\boldsymbol{\upsilon}) - I\!E \mathcal{L}(\boldsymbol{\upsilon}).$$

We start with the local conditions.

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

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

$(\mathcal{E}\mathcal{D}_1)$ For all $0 < \mathtt{r} < \mathtt{r}_0$, there exists a constant $\omega(\mathtt{r}) \le 1/2$ such that for all $\boldsymbol{\upsilon} \in \Upsilon_0(\mathtt{r})$ and $|\mu| \le \mathtt{g}$

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

$(\mathcal{L}_0)$ There exist a symmetric $p \times p$ matrix $\mathcal{D}_0^2$ and a constant $\delta(\mathtt{r}) \le 1/2$ such that it holds on the set $\Upsilon_0(\mathtt{r})$ for all $\mathtt{r} \le \mathtt{r}_0$

$$\left| \frac{-2 I\!E \mathcal{L}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*)}{\|\mathcal{D}_0(\boldsymbol{\upsilon} - \boldsymbol{\upsilon}^*)\|^2} - 1 \right| \le \delta(\mathtt{r}).$$

The global conditions are:

$(\mathcal{L}\mathtt{r})$ For any $\mathtt{r} > \mathtt{r}_0$ there exists a value $\mathtt{b}(\mathtt{r}) > 0$, such that

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

$(\mathcal{E}\mathtt{r})$ For any $\mathtt{r} \ge \mathtt{r}_0$ there exists a constant $\nu_0 > 0$ and a constant $\mathtt{g}(\mathtt{r}) > 0$ such that

$$\sup_{\boldsymbol{\upsilon} \in \Upsilon_0(\mathtt{r})} \sup_{\mu \le \mathtt{g}(\mathtt{r})} \sup_{\boldsymbol{\gamma} \in I\!R^p} \log I\!E \exp\left\{ \mu \frac{\langle \nabla\zeta(\boldsymbol{\upsilon}), \boldsymbol{\gamma} \rangle}{\|\mathcal{V}_0 \boldsymbol{\gamma}\|} \right\} \le \frac{\nu_0^2 \mu^2}{2}.$$

Condition $(\mathcal{E}\mathtt{r})$ will be made more precise by specifying the rate of decay of the function $\mathtt{g}(\mathtt{r})$; (see section 4.2).

Finally we specify the regularity conditions. We begin by representing the information and the covariance matrices in 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 Spokoiny (2012) ensure that the matrix $\mathcal{D}_0$ is positive and satisfied $\mathfrak{a}^2\mathcal{D}_0^2 \geq \mathcal{V}_0^2$ 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.

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

$$\mathfrak{a}^2 D_0^2 \geq V_0^2, \qquad \mathfrak{a}^2 H_0^2 \geq Q_0^2, \qquad \mathfrak{a}^2\mathcal{D}_0^2 \geq \mathcal{V}_0^2. \tag{4.1}$$

and

$$\|D_0^{-1} A_0 H_0^{-2} A_0^\top D_0^{-1}\|_\infty \leq \nu. \tag{4.2}$$

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 (4.1) are given with the same constant $\mathfrak{a}$ only for simplifying the notation. One can show that the last bound on $\mathcal{D}_0^2$ follows from the first two and (4.2) with another constant $\mathfrak{a}'$ depending on $\mathfrak{a}$ and ν only.

Also we introduce a constant $\boldsymbol{\epsilon}$, which depends on the choice of the local zone radius $\mathtt{r}_0$:

$$\boldsymbol{\epsilon} = 3\nu_0\, \mathfrak{a}^2\, \omega(\mathtt{r}_0) + \delta(\mathtt{r}_0). \tag{4.3}$$

4.2 Bracketing and upper function devices

This section briefly overviews the main constructions of Spokoiny (2012) 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

$$\Upsilon_0(\mathtt{r}) \stackrel{\text{def}}{=} \bigl\{\boldsymbol{\upsilon} : (\boldsymbol{\upsilon} - \boldsymbol{\upsilon}^*)^\top \mathcal{D}_0^2 (\boldsymbol{\upsilon} - \boldsymbol{\upsilon}^*) \leq \mathtt{r}^2\bigr\}.$$

For ϵ, define the bracketing quadratic processes $\mathbb{L}_{\epsilon}(\upsilon, \upsilon^*)$ and $\mathbb{L}_{\underline{\epsilon}}(\upsilon, \upsilon^*)$:

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

$$\mathcal{D}_{\epsilon}^2 \stackrel{\text{def}}{=} \mathcal{D}_0^2(1-\epsilon), \qquad \xi_{\epsilon} \stackrel{\text{def}}{=} \mathcal{D}_{\epsilon}^{-1} \nabla \mathcal{L}(\upsilon^*)$$

and accordingly for $\underline{\epsilon} = -\epsilon$.

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 select $\mathtt{x} = \mathtt{C} \log(p)$. Also a value $\mathtt{r}_0$ has to be fixed which separated the local and global zones.

Theorem 4.1. *Suppose the conditions* (ED_0)*,* (ED_1)*,* $(\mathcal{L}_0)$*, and* $(\mathcal{I})$ *from Section 4.1 with* $\mathtt{r}_0^2 \ge \mathtt{C}(\mathfrak{a}^2 + 1)(p + \mathtt{x})$ *for a fixed constant* $\mathtt{C}$*. Then*

$$\mathbb{L}_{\underline{\epsilon}}(\upsilon, \upsilon^*) - \diamondsuit_{\underline{\epsilon}}(\mathtt{r}_0) \le L(\upsilon, \upsilon^*) \le \mathbb{L}_{\epsilon}(\upsilon, \upsilon^*) + \diamondsuit_{\epsilon}(\mathtt{r}_0), \quad \upsilon \in \Upsilon_0(\mathtt{r}_0), \tag{4.4}$$

where the random variables $\diamondsuit_{\epsilon}(\mathtt{r}_0), \diamondsuit_{\underline{\epsilon}}(\mathtt{r}_0)$ *fulfill on a random set* $\Omega(\mathtt{x})$ *of dominating probability*

$$\diamondsuit_{\epsilon}(\mathtt{r}_0) \le \mathtt{C}\, \epsilon\, (p + \mathtt{x}), \qquad \diamondsuit_{\underline{\epsilon}}(\mathtt{r}_0) \le \mathtt{C}\, \epsilon\, (p + \mathtt{x}). \tag{4.5}$$

Moreover, the random vector $\xi = \mathcal{D}_0^{-1} \nabla \mathcal{L}(\upsilon^*)$ *fulfills on* $\Omega(\mathtt{x})$

$$\|\xi\|^2 \le \mathtt{C} I\!\!E \|\xi\|^2 \le \mathtt{C}\mathfrak{a}^2 (p + \mathtt{x}).$$

Furthermore, assume $(E\mathtt{r})$ *and* $(\mathcal{L}\mathtt{r})$ *with* $\mathtt{b}(\mathtt{r}) \equiv \mathtt{b}$ *yielding*

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

for each $\upsilon \in \Upsilon \setminus \Upsilon_0(\mathtt{r}_0)$*. Let also* $\mathtt{r}_0^2 \ge \mathtt{C}(\mathfrak{a}^2 \vee \mathtt{b}^{-1})\,(p + \mathtt{x})$ *and* $\mathtt{g}(\mathtt{r}) \ge \mathtt{C}\mathtt{b}$ *for all* $\mathtt{r} \ge \mathtt{r}_0$*; see* $(E\mathtt{r})$*. Then*

$$L(\upsilon, \upsilon^*) \le -\mathfrak{u}(\upsilon), \qquad \upsilon \in \Upsilon \setminus \Upsilon_0(\mathtt{r}_0). \tag{4.6}$$

holds on a random set $\Omega(\mathtt{x})$ *with* $\mathfrak{u}(\upsilon) = \mathtt{b}\, \|\mathcal{D}_0(\upsilon - \upsilon^*)\|^2/2$*.*

The results (4.4) and (4.5) are stated in Spokoiny (2012) but for a slightly different definition of $\mathcal{D}_{\epsilon}^2$ and $\mathcal{D}_{\underline{\epsilon}}^2$. Namely, for $\epsilon = (\delta, \varrho)$, one defines $\mathcal{D}_{\epsilon}^2 = \mathcal{D}_0^2(1-\delta) - \varrho \mathcal{V}_0^2$ and similarly for $\underline{\epsilon} = -\epsilon$. However, these two constructions are essentially equivalent due to the identifiability condition $(\mathcal{I})$. Indeed, $\mathfrak{a}^2 \mathcal{D}_0^2 \ge \mathcal{V}_0^2$ implies

$$\mathcal{D}_0^2(1-\delta) - \varrho \mathcal{V}_0^2 \ge (1-\epsilon)\mathcal{D}_0^2, \qquad \mathcal{D}_0^2(1+\delta) + \varrho \mathcal{V}_0^2 \le (1+\epsilon)\mathcal{D}_0^2$$

with $\epsilon = \delta + \mathfrak{a}^2 \varrho$.

4.3 Extension of Theorem 2.1 to a continuous prior

The results of Theorem 2.1 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 on the local set $\Upsilon_0(\mathtt{r}_0)$. More precisely, let $\pi(\boldsymbol{\upsilon})$ satisfy

$$\sup_{\boldsymbol{\upsilon}\in\Upsilon_0(\mathtt{r}_0)}\Bigl|\frac{\pi(\boldsymbol{\upsilon})}{\pi(\boldsymbol{\upsilon}^*)}-1\Bigr|\le\alpha,\qquad\sup_{\boldsymbol{\upsilon}\in\Upsilon}\frac{\pi(\boldsymbol{\upsilon})}{\pi(\boldsymbol{\upsilon}^*)}\le\mathtt{C},\tag{4.7}$$

where α is a small constant while $\mathtt{C}$ is any fixed constant. Then the results of Theorem 2.1 continue to apply with an obvious correction of the approximation error.

As an example, consider the case of a Gaussian prior $\Pi=\mathcal{N}(0,G^{-2})$ with the density $\pi(\boldsymbol{\upsilon})\propto\exp\bigl\{-\|G\boldsymbol{\upsilon}\|^2/2\bigr\}$. In addition, suppose that the value $\|G\boldsymbol{\upsilon}^*\|$ is bounded by a fixed constant. Then

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

and the condition (4.7) is fulfilled if $\|G(\boldsymbol{\upsilon}-\boldsymbol{\upsilon}^*)\|$ is a small number for all $\boldsymbol{\upsilon}\in\Upsilon_0(\mathtt{r}_0)$. The non-informative prior can be viewed as a limiting case of a Gaussian prior as $G\to0$. We are interested in quantifying this relation and addressing the question, how small should G be to ensure the BvM result. It is obvious from the definition of $\Upsilon_0(\mathtt{r}_0)$ that

$$\|G(\boldsymbol{\upsilon}-\boldsymbol{\upsilon}^*)\|=\|G\mathcal{D}_0^{-1}\mathcal{D}_0(\boldsymbol{\upsilon}-\boldsymbol{\upsilon}^*)\|\le\|G\mathcal{D}_0^{-1}\|_\infty\,\mathtt{r}_0\,.$$

Similarly

$$\bigl|(\boldsymbol{\upsilon}-\boldsymbol{\upsilon}^*)^\top G^2\boldsymbol{\upsilon}^*\bigr|\le\|G\boldsymbol{\upsilon}^*\|\cdot\|G(\boldsymbol{\upsilon}-\boldsymbol{\upsilon}^*)\|\le\|G\boldsymbol{\upsilon}^*\|\cdot\|G\mathcal{D}_0^{-1}\|_\infty\,\mathtt{r}_0.$$

Therefore, (4.7) effectively requires that $\|G\mathcal{D}_0^{-1}\|_\infty\,\mathtt{r}_0$ is small. A proper choice of $\mathtt{r}_0$ is given by $\mathtt{r}_0^2=\mathtt{C}(p+\mathtt{x})$ yielding the rule " $\|G\mathcal{D}_0^{-1}\|_\infty\,(p+\mathtt{x})^{1/2}$ is small".

Theorem 4.2. *Suppose the conditions of Theorem 4.1. Let also* $\Pi=\mathcal{N}(0,G^{-2})$ *be a Gaussian prior measure on* $I\!R^p$ *such that*

$$\|G\boldsymbol{\upsilon}^*\|\le\mathtt{C},\qquad G^2\le\mathtt{C}\,\boldsymbol{\epsilon}\,\mathcal{D}_0^2,$$

and $\boldsymbol{\epsilon}\,(p+\mathtt{x})$ *is small. Then the BvM result* (2.6) *of Theorem 2.1 holds.*

4.4 Tail posterior probability for 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

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

Theorem 4.3. *Suppose the conditions of Theorem 4.1. Then it holds on* $\Omega(\mathtt{x})$

$$\rho^*(\mathtt{r}_0) \le \exp\{\diamondsuit_{\underline{\boldsymbol{\epsilon}}}(\mathtt{r}_0) + \nu(\mathtt{r}_0)\} \left(\frac{1+\boldsymbol{\epsilon}}{\mathtt{b}}\right)^{p/2} I\!\!P\bigl(\|\boldsymbol{\gamma}\|^2 \ge \mathtt{b}\mathtt{r}_0^2\bigr),$$

with

$$\nu(\mathtt{r}_0) \stackrel{\text{def}}{=} -\log I\!\!P\bigl(\|\boldsymbol{\gamma} + \boldsymbol{\xi}\| \le \mathtt{r}_0 \bigm| \boldsymbol{Y}\bigr). \tag{4.8}$$

Similarly, for each $m \ge 0$

$$\begin{aligned}\rho_m^*(\mathtt{r}_0) &\stackrel{\text{def}}{=} I\!\!E\bigl[\|\mathcal{D}_{\underline{\boldsymbol{\epsilon}}}(\boldsymbol{\upsilon} - \boldsymbol{\upsilon}_{\underline{\boldsymbol{\epsilon}}})\|^m \, 1\!\!I\bigl\{\boldsymbol{\upsilon} \notin \Upsilon_0(\mathtt{r}_0)\bigr\} \bigm| \boldsymbol{Y}\bigr] \\ &\le \exp\{\diamondsuit_{\underline{\boldsymbol{\epsilon}}}(\mathtt{r}_0) + \nu(\mathtt{r}_0)\} \left(\frac{1+\boldsymbol{\epsilon}}{\mathtt{b}}\right)^{p/2} I\!\!E\bigl[\|\boldsymbol{\gamma}\|^m \, 1\!\!I\bigl(\|\boldsymbol{\gamma}\|^2 \ge \mathtt{b}\mathtt{r}_0^2\bigr)\bigr],\end{aligned}$$

where $\boldsymbol{\upsilon}_{\underline{\boldsymbol{\epsilon}}} = \boldsymbol{\upsilon}^* + \mathcal{D}_{\underline{\boldsymbol{\epsilon}}}^{-1} \boldsymbol{\xi}_{\underline{\boldsymbol{\epsilon}}}$.

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

Corollary 4.4. *Assume the conditions of Theorem 4.3. Then inequality* $\mathtt{b}\mathtt{r}_0^2 \ge \mathtt{C}(p + \mathtt{x})$ *ensures*

$$\rho_m^*(\mathtt{r}_0) \le \mathrm{e}^{-\mathtt{x}}, \quad m = 0, 1, 2.$$

4.5 Tail posterior probability for target parameter

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

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

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

$$\rho(\mathtt{r}_0) = \frac{\int_{\Upsilon} \exp\bigl\{L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*)\bigr\} \, 1\!\!I\bigl\{\boldsymbol{\theta} \notin \Theta_0(\mathtt{r}_0)\bigr\} d\boldsymbol{\upsilon}}{\int_{\Upsilon} \exp\bigl\{L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*)\bigr\} \, 1\!\!I\bigl\{\boldsymbol{\theta} \in \Theta_0(\mathtt{r}_0)\bigr\} d\boldsymbol{\upsilon}} . \tag{4.9}$$

Obviously $I\!P\bigl(\boldsymbol{\theta} \notin \Theta_0(\mathtt{r}_0) \,\big|\, \boldsymbol{Y}\bigr) \le \rho(\mathtt{r}_0)$. Therefore, small values of $\rho(\mathtt{r}_0)$ indicate a small posterior probability of the large deviation set $\{\boldsymbol{\theta} \notin \Theta_0(\mathtt{r}_0)\}$. Define

$$\breve{D}_{\boldsymbol{\epsilon}}^2 = (1-\boldsymbol{\epsilon})\breve{D}_0^2, \qquad \breve{D}_{\underline{\boldsymbol{\epsilon}}}^2 = (1+\boldsymbol{\epsilon})\breve{D}_0^2.$$

Theorem 4.5. *Suppose* (4.4). *Let a random set* $\Omega(\mathtt{x})$ *be such that* $I\!P\bigl(\Omega(\mathtt{x})\bigr) \ge 1 - \mathtt{C}\mathrm{e}^{-\mathtt{x}}$ *and* (4.6) *holds on* $\Omega(\mathtt{x})$. *Then on* $\Omega(\mathtt{x})$

$$\rho(\mathtt{r}_0) \le \rho^*(\mathtt{r}_0) \le \exp\{\diamondsuit_{\underline{\boldsymbol{\epsilon}}}(\mathtt{r}_0) + \nu(\mathtt{r}_0)\} \left(\frac{1+\boldsymbol{\epsilon}}{\mathtt{b}}\right)^{p/2} I\!P\bigl(\|\boldsymbol{\gamma}\|^2 \ge \mathtt{b}\mathtt{r}_0^2\bigr),$$

where $\nu(\mathtt{r}_0)$ *is from* (4.8). *Similarly, for each* $m \ge 0$

$$\begin{aligned}\rho_m^*(\mathtt{r}_0) &\stackrel{\mathrm{def}}{=} I\!E\bigl[\|\breve{D}_{\underline{\boldsymbol{\epsilon}}}(\boldsymbol{\vartheta} - \boldsymbol{\theta}_{\underline{\boldsymbol{\epsilon}}})\|^m \, \mathrm{I\!I}\bigl\{\boldsymbol{\vartheta} \notin \Theta_0(\mathtt{r}_0)\bigr\} \,\big|\, \boldsymbol{Y}\bigr] \\ &\le \exp\{\diamondsuit_{\underline{\boldsymbol{\epsilon}}}(\mathtt{r}_0) + \nu(\mathtt{r}_0)\} \left(\frac{1+\boldsymbol{\epsilon}}{\mathtt{b}}\right)^{p/2} I\!E\bigl[\|\boldsymbol{\gamma}\|^m \, \mathrm{I\!I}\bigl(\|\boldsymbol{\gamma}\|^2 \ge \mathtt{b}\mathtt{r}_0^2\bigr)\bigr].\end{aligned}$$

4.6 Local Gaussian approximation of the posterior. Upper bound

It is convenient to introduce local conditional expectation: for a random variable η, define

$$I\!E^{\circ}\eta \stackrel{\mathrm{def}}{=} I\!E\Bigl[\eta \, \mathrm{I\!I}\bigl\{\boldsymbol{\theta} \in \Theta_0(\mathtt{r}_0)\bigr\} \,\big|\, \boldsymbol{Y}\Bigr].$$

The following theorem gives exact statement about upper bound of this posterior probability. Let

$$\begin{aligned}\breve{\boldsymbol{\theta}} &\stackrel{\mathrm{def}}{=} \boldsymbol{\theta}^* + \breve{D}_0^{-1}\breve{\boldsymbol{\xi}}, \\ \boldsymbol{\theta}_{\boldsymbol{\epsilon}} &\stackrel{\mathrm{def}}{=} \boldsymbol{\theta}^* + \breve{D}_{\boldsymbol{\epsilon}}^{-1}\breve{\boldsymbol{\xi}}_{\boldsymbol{\epsilon}} = \boldsymbol{\theta}^* + (1-\boldsymbol{\epsilon})^{-1}\breve{D}_0^{-1}\breve{\boldsymbol{\xi}},\end{aligned}$$

where $\breve{\boldsymbol{\xi}}_{\boldsymbol{\epsilon}} = \breve{D}_{\boldsymbol{\epsilon}}^{-1}\breve{\nabla}_{\boldsymbol{\theta}} = (1-\boldsymbol{\epsilon})^{-1/2}\breve{\boldsymbol{\xi}}$.

Theorem 4.6. *Suppose* (4.4). *Let a random set* $\Omega(\mathtt{x})$ *be such that* $I\!P\bigl\{\Omega(\mathtt{x})\bigr\} \ge 1 - \mathtt{C}\mathrm{e}^{-\mathtt{x}}$ *and* (4.6) *holds on* $\Omega(\mathtt{x})$. *Then for any* $f \colon \mathbb{R}^q \to \mathbb{R}_+$

$$I\!E^{\circ} f\bigl(\breve{D}_{\boldsymbol{\epsilon}}(\boldsymbol{\vartheta} - \boldsymbol{\theta}_{\boldsymbol{\epsilon}})\bigr) \le \exp\bigl\{\Delta_{\boldsymbol{\epsilon}}^{+}(\mathtt{r}_0)\bigr\} \, I\!E f(\boldsymbol{\gamma}), \tag{4.10}$$

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

$$\begin{aligned}\Delta_{\boldsymbol{\epsilon}}^{+}(\mathtt{r}_0) &\stackrel{\mathrm{def}}{=} \Delta_{\boldsymbol{\epsilon}}(\mathtt{r}_0) + \frac{p}{2}\log\left(\frac{1+\boldsymbol{\epsilon}}{1-\boldsymbol{\epsilon}}\right) + \nu(\mathtt{r}_0) + \rho_f(\mathtt{r}_0), \\ \Delta_{\boldsymbol{\epsilon}}(\mathtt{r}_0) &\stackrel{\mathrm{def}}{=} \diamondsuit_{\boldsymbol{\epsilon}}(\mathtt{r}_0) + \diamondsuit_{\underline{\boldsymbol{\epsilon}}}(\mathtt{r}_0) + \bigl(\|\boldsymbol{\xi}_{\boldsymbol{\epsilon}}\|^2 - \|\boldsymbol{\xi}_{\underline{\boldsymbol{\epsilon}}}\|^2\bigr)/2\,, \\ \rho_f(\mathtt{r}_0) &\stackrel{\mathrm{def}}{=} \frac{\int_{\Upsilon \setminus \Upsilon_0(\mathtt{r}_0)} \exp\bigl\{L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*)\bigr\} \, f\bigl(\breve{D}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta} - \boldsymbol{\theta}_{\boldsymbol{\epsilon}})\bigr)\, d\boldsymbol{\upsilon}}{\int_{\Upsilon_0(\mathtt{r}_0)} \exp\bigl\{L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*)\bigr\} \, f\bigl(\breve{D}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta} - \boldsymbol{\theta}_{\boldsymbol{\epsilon}})\bigr)\, d\boldsymbol{\upsilon}}.\end{aligned}$$

The next result considers a special case with $f(\boldsymbol{u}) = \left|\boldsymbol{\lambda}^{\top}\boldsymbol{u}\right|^2$.

Corollary 4.7. *For any* $\boldsymbol{\lambda} \in I\!R^q$

$$I\!E^{\circ}\left|\boldsymbol{\lambda}^{\top}\breve{D}_0(\boldsymbol{\vartheta} - \breve{\boldsymbol{\theta}})\right|^2 \le \exp(\Delta_{\boldsymbol{\epsilon}}^{\oplus}(\mathtt{r}_0))\|\boldsymbol{\lambda}\|^2, \tag{4.11}$$

where

$$\Delta_{\boldsymbol{\epsilon}}^{\oplus}(\mathtt{r}_0) = \Delta_{\boldsymbol{\epsilon}}^{+}(\mathtt{r}_0) - \log(1-\boldsymbol{\epsilon}) + \mathtt{C}\boldsymbol{\epsilon}\exp\bigl(\Delta_{\boldsymbol{\epsilon}}^{+}(\mathtt{r}_0)/2\bigr)\,\|\breve{\boldsymbol{\xi}}\| + \mathtt{C}\boldsymbol{\epsilon}^2\,\|\breve{\boldsymbol{\xi}}\|^2.$$

On $\Omega(\mathtt{x})$ *one obtains* $\Delta_{\boldsymbol{\epsilon}}^{\oplus}(\mathtt{r}_0) \le \mathtt{C}\,\boldsymbol{\epsilon}\,\mathfrak{p}$.

Define for random event $\eta \in A \subseteq \mathbb{R}^q$:

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

The next corollary describes an upper bound for the posterior probability.

Corollary 4.8. *Let* D_1 *be symmetric* $q \times q$ *matrix such that* $\|I - D_1^{-1}\breve{D}_0^2 D_1^{-1}\|_{\infty} \le \mathtt{C}\,\boldsymbol{\epsilon}\,\mathfrak{p}$ *and let* $\widehat{\boldsymbol{\theta}} \in \mathbb{R}^q$ *be such that* $\|\breve{D}_0(\breve{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}})\|^2 \le \mathtt{C}\,\boldsymbol{\epsilon}\,\mathfrak{p}$. *Then for any measurable set* $A \subset I\!R^q$, *it holds on* $\Omega(\mathtt{x})$ *with* $\boldsymbol{\delta}_{\boldsymbol{\epsilon}} \stackrel{\mathrm{def}}{=} D_1(\boldsymbol{\theta}_{\boldsymbol{\epsilon}} - \widehat{\boldsymbol{\theta}})$

$$I\!P^{\circ}\bigl(D_1(\boldsymbol{\vartheta} - \widehat{\boldsymbol{\theta}}) \in A\bigr) \le \exp\bigl\{\Delta_{\boldsymbol{\epsilon}}^{+}(\mathtt{r}_0)\bigr\} I\!P\bigl(D_1\breve{D}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{\gamma} + \boldsymbol{\delta}_{\boldsymbol{\epsilon}} \in A\bigr) \tag{4.12}$$

$$\le \exp\bigl\{\Delta_{\boldsymbol{\epsilon}}^{+}(\mathtt{r}_0)\bigr\}\bigl\{I\!P\bigl(\boldsymbol{\gamma} \in A\bigr) + \sqrt{\mathtt{C}(\boldsymbol{\epsilon}\,\mathfrak{p} + \boldsymbol{\epsilon}^2\mathfrak{p}^2 q)}\bigr\}. \tag{4.13}$$

4.7 Local Gaussian approximation of the posterior. Lower bound

Let $\boldsymbol{\theta}_{\underline{\boldsymbol{\epsilon}}} = \boldsymbol{\theta}^* + \breve{D}_{\underline{\boldsymbol{\epsilon}}}^{-1}\breve{\boldsymbol{\xi}}_{\underline{\boldsymbol{\epsilon}}}$, where $\breve{\boldsymbol{\xi}}_{\underline{\boldsymbol{\epsilon}}} = \breve{D}_{\underline{\boldsymbol{\epsilon}}}^{-1}\breve{\nabla}_{\boldsymbol{\theta}}$. Now we present a local lower bound for the posterior probability:

Theorem 4.9. *Suppose* (4.4). *Let a random set* $\Omega(\mathtt{x})$ *be such that* $I\!P\bigl\{\Omega(\mathtt{x})\bigr\} \ge 1 - \mathtt{C}\mathrm{e}^{-\mathtt{x}}$ *and* (4.6) *holds on* $\Omega(\mathtt{x})$. *Then it holds on* $\Omega(\mathtt{x})$

$$I\!E^{\circ} f\bigl(\breve{D}_{\underline{\boldsymbol{\epsilon}}}(\boldsymbol{\vartheta} - \boldsymbol{\theta}_{\underline{\boldsymbol{\epsilon}}})\bigr) \ge \exp\bigl\{-\Delta_{\boldsymbol{\epsilon}}^{-}(\mathtt{r}_0)\bigr\}\, I\!E f(\boldsymbol{\gamma})\,\mathrm{I\!I}\bigl\{\|\boldsymbol{\gamma}\| \le \mathtt{C}\mathtt{r}_0\bigr\}, \tag{4.14}$$

where

$$\Delta_{\boldsymbol{\epsilon}}^{-}(\mathtt{r}_0) \stackrel{\mathrm{def}}{=} \Delta_{\boldsymbol{\epsilon}}(\mathtt{r}_0) + \frac{p}{2}\log\left(\frac{1+\boldsymbol{\epsilon}}{1-\boldsymbol{\epsilon}}\right) + \nu(\mathtt{r}_0) + \rho^*(\mathtt{r}_0) + 2\widetilde{\rho}_f(\mathtt{r}_0),$$

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

This result means that posterior has lower bound which is nearly standard normal up to (small) multiplicative and additive constants. As a corollary, we state the result for a quadratic function f. Here we need an additional condition that the $\mathtt{r}_0^2 \ge \mathtt{C}(p+\mathtt{x})$ for $\mathtt{C}$ sufficiently large.

Corollary 4.10. *Let* $\mathtt{r}_0^2 \ge \mathtt{C}(p+\mathtt{x})$. *For any* $\boldsymbol{\lambda} \in I\!R^q$

$$I\!E^{\circ}\bigl|\boldsymbol{\lambda}^{\top}\breve{D}_0(\boldsymbol{\vartheta}-\breve{\boldsymbol{\theta}})\bigr|^2 \ge \exp(-\Delta_{\boldsymbol{\epsilon}}^{\ominus}(\mathtt{r}_0))\|\boldsymbol{\lambda}\|^2,$$

where

$$\Delta_{\boldsymbol{\epsilon}}^{\ominus}(\mathtt{r}_0) = \Delta_{\boldsymbol{\epsilon}}^{-}(\mathtt{r}_0) + \log(1+\boldsymbol{\epsilon}) + \mathtt{C}\boldsymbol{\epsilon}\exp\bigl(\Delta_{\boldsymbol{\epsilon}}^{-}(\mathtt{r}_0)/2\bigr)\,\|\breve{\boldsymbol{\xi}}\| + \mathtt{C}\boldsymbol{\epsilon}^2\,\|\breve{\boldsymbol{\xi}}\|^2 + \mathrm{e}^{-\mathtt{x}}.$$

Let D_1^2 *be a symmetric* $q \times q$ *matrix such that* $\|I - D_1^{-1}\breve{D}_0^2 D_1^{-1}\|_{\infty} \le \mathtt{C}\,\boldsymbol{\epsilon}\,\mathfrak{p}$ *and let* $\widehat{\boldsymbol{\theta}} \in \mathbb{R}^q$ *be such that* $\|\breve{D}_0(\breve{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}})\|^2 \le \mathtt{C}\,\boldsymbol{\epsilon}\,\mathfrak{p}$. *Then for any measurable subset* A *in* $I\!R^q$, *it holds on* $\Omega(\mathtt{x})$ *with* $\boldsymbol{\delta}_{\underline{\boldsymbol{\epsilon}}} = D_1(\boldsymbol{\theta}_{\underline{\boldsymbol{\epsilon}}} - \widehat{\boldsymbol{\theta}})$

$$\begin{aligned} I\!P^{\circ}\bigl(D_1(\boldsymbol{\vartheta}-\widehat{\boldsymbol{\theta}}) \in A\bigr) &\ge \exp\bigl\{\Delta_{\boldsymbol{\epsilon}}^{-}(\mathtt{r}_0)\bigr\} I\!P\bigl(D_1\breve{D}_{\underline{\boldsymbol{\epsilon}}}^{-1}\boldsymbol{\gamma} + \boldsymbol{\delta}_{\underline{\boldsymbol{\epsilon}}} \in A\bigr) - \mathrm{e}^{-\mathtt{x}} \\ &\ge \exp\bigl\{\Delta_{\boldsymbol{\epsilon}}^{-}(\mathtt{r}_0)\bigr\}\Bigl\{I\!P\bigl(\boldsymbol{\gamma} \in A\bigr) - \sqrt{\mathtt{C}(\boldsymbol{\epsilon}\,\mathfrak{p} + \boldsymbol{\epsilon}^2\mathfrak{p}^2 q)}\Bigr\} - \mathrm{e}^{-\mathtt{x}}. \end{aligned}$$

On $\Omega(\mathtt{x})$, *it holds* $\Delta_{\boldsymbol{\epsilon}}^{-}(\mathtt{r}_0) \le \mathtt{C}\,\boldsymbol{\epsilon}\,\mathfrak{p}$, $\Delta_{\boldsymbol{\epsilon}}^{\ominus}(\mathtt{r}_0) \le \mathtt{C}\,\boldsymbol{\epsilon}\,\mathfrak{p}$.

The proof of this Corollary is similar to Corollary 4.8 and Corollary 4.7.

5 Proofs

This appendix collects the proofs of the results.

5.1 Proof of Theorem 4.5

Obviously

$$\bigl\{\boldsymbol{\theta} \notin \Theta_0(\mathtt{r}_0),\, \boldsymbol{\upsilon} \in \varUpsilon\bigr\} \subset \bigl\{\varUpsilon \setminus \varUpsilon_0(\mathtt{r}_0)\bigr\}.$$

Therefore, it holds for the integral in the nominator of (4.9) in a view of (4.6)

$$\int_{\varUpsilon} \exp\bigl\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\bigr\}\, \mathrm{1\!I}\bigl\{\boldsymbol{\theta} \notin \Theta_0(\mathtt{r}_0)\bigr\}\, d\boldsymbol{\upsilon} \le \int_{\varUpsilon\setminus\varUpsilon_0(\mathtt{r}_0)} \exp\bigl\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\bigr\}\, d\boldsymbol{\upsilon}.$$

For the local integral in the denominator, the inclusion $\varUpsilon_0(\mathtt{r}_0) \subset \bigl\{\boldsymbol{\theta} \in \Theta_0(\mathtt{r}_0), \boldsymbol{\upsilon} \in \varUpsilon\bigr\}$ and (4.6) imply

$$\int_{\varUpsilon} \exp\bigl\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\bigr\}\, \mathrm{1\!I}\bigl\{\boldsymbol{\theta} \in \Theta_0(\mathtt{r}_0)\bigr\}\, d\boldsymbol{\upsilon} \ge \int_{\varUpsilon_0(\mathtt{r}_0)} \exp\bigl\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\bigr\}\, d\boldsymbol{\upsilon}.$$

Finally

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

and the assertion follows from Theorem 4.3.

5.2 Proof of Theorem 4.6

We use that $\mathbb{L}_{\boldsymbol{\epsilon}}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) = \boldsymbol{\xi}_{\boldsymbol{\epsilon}}^\top \mathcal{D}_{\boldsymbol{\epsilon}}(\boldsymbol{\upsilon} - \boldsymbol{\upsilon}^*) - \|\mathcal{D}_{\boldsymbol{\epsilon}}(\boldsymbol{\upsilon} - \boldsymbol{\upsilon}^*)\|^2/2$ is proportional to the density of a Gaussian distribution and similarly for $\mathbb{L}_{\underline{\boldsymbol{\epsilon}}}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*)$. More precisely, define

$$m_{\boldsymbol{\epsilon}}(\boldsymbol{\xi}_{\boldsymbol{\epsilon}}) \stackrel{\text{def}}{=} -\|\boldsymbol{\xi}_{\boldsymbol{\epsilon}}\|^2/2 + \log(\det \mathcal{D}_{\boldsymbol{\epsilon}}) - p \log(\sqrt{2\pi}). \tag{5.1}$$

Then

$$\begin{aligned} & m_{\boldsymbol{\epsilon}}(\boldsymbol{\xi}_{\boldsymbol{\epsilon}}) + \mathbb{L}_{\boldsymbol{\epsilon}}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) \\ & \quad = -\|\mathcal{D}_{\boldsymbol{\epsilon}}(\boldsymbol{\upsilon} - \boldsymbol{\upsilon}^*) - \boldsymbol{\xi}_{\boldsymbol{\epsilon}}\|^2/2 + \log(\det \mathcal{D}_{\boldsymbol{\epsilon}}) - p \log(\sqrt{2\pi}) \end{aligned}$$

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

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

Thus,

$$\int_{\Upsilon} \exp\bigl\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^{*})\bigr\}\, f\bigl(\breve{D}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}-\boldsymbol{\theta}_{\boldsymbol{\epsilon}})\bigr)\, d\boldsymbol{\upsilon} \\ \leq \exp\{\diamondsuit_{\boldsymbol{\epsilon}}(\mathtt{r}_0) - m_{\boldsymbol{\epsilon}}(\boldsymbol{\xi}_{\boldsymbol{\epsilon}}) + \rho_f(\mathtt{r}_0)\}\, I\!E f(\boldsymbol{\gamma}). \tag{5.2}$$

Similarly, $m_{\underline{\boldsymbol{\epsilon}}}(\boldsymbol{\xi}_{\underline{\boldsymbol{\epsilon}}})$ is defined by (5.1) with $\underline{\boldsymbol{\epsilon}}$ in place of $\boldsymbol{\epsilon}$, and the value $m_{\underline{\boldsymbol{\epsilon}}}(\boldsymbol{\xi}_{\underline{\boldsymbol{\epsilon}}}) + \mathbb{L}_{\underline{\boldsymbol{\epsilon}}}(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^{*})$ is (conditionally on $\boldsymbol{Y}$) the density of the normal law with the mean $\boldsymbol{\upsilon}_{\underline{\boldsymbol{\epsilon}}} = \mathcal{D}_{\underline{\boldsymbol{\epsilon}}}^{-1}\boldsymbol{\xi}_{\underline{\boldsymbol{\epsilon}}} + \boldsymbol{\upsilon}^{*}$ and the covariance matrix $\mathcal{D}_{\underline{\boldsymbol{\epsilon}}}^{-2}$. So, it holds

$$\int_{\Upsilon} \exp\bigl\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^{*})\bigr\}\, d\boldsymbol{\upsilon} \geq \int_{\Upsilon_0(\mathtt{r}_0)} \exp\bigl\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^{*})\bigr\}\, d\boldsymbol{\upsilon} \\ \geq \exp\{-\diamondsuit_{\underline{\boldsymbol{\epsilon}}}(\mathtt{r}_0) - m_{\underline{\boldsymbol{\epsilon}}}(\boldsymbol{\xi}_{\underline{\boldsymbol{\epsilon}}})\} \int_{\Upsilon_0(\mathtt{r}_0)} \exp\bigl\{\mathbb{L}_{\underline{\boldsymbol{\epsilon}}}(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^{*}) + m_{\underline{\boldsymbol{\epsilon}}}(\boldsymbol{\xi}_{\underline{\boldsymbol{\epsilon}}})\bigr\}\, d\boldsymbol{\upsilon}. \tag{5.3}$$

Further, by construction, $\mathcal{D}_{\underline{\boldsymbol{\epsilon}}}^2 \geq \mathcal{D}_0^2$ and $\|\boldsymbol{\xi}_{\underline{\boldsymbol{\epsilon}}}\| \leq \|\boldsymbol{\xi}\|$, yielding

$$\bigl\{\mathcal{D}_{\underline{\boldsymbol{\epsilon}}}^{-1}\bigl(\boldsymbol{u}+\boldsymbol{\xi}_{\underline{\boldsymbol{\epsilon}}}\bigr) \in \Upsilon_0(\mathtt{r}_0)\bigr\} = \bigl\{\bigl\|\mathcal{D}_0\mathcal{D}_{\underline{\boldsymbol{\epsilon}}}^{-1}\bigl(\boldsymbol{u}+\boldsymbol{\xi}_{\underline{\boldsymbol{\epsilon}}}\bigr)\bigr\| \leq \mathtt{r}_0\bigr\} \supset \bigl\{\|\boldsymbol{u}+\boldsymbol{\xi}\| \leq \mathtt{r}_0\bigr\}.$$

Hence, inequality (5.3) implies by definition of $\nu(\mathtt{r}_0)$:

$$\int_{\Upsilon} \exp\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^{*})\}\, d\boldsymbol{\upsilon} \geq \exp\bigl\{-\diamondsuit_{\underline{\boldsymbol{\epsilon}}}(\mathtt{r}_0) - m_{\underline{\boldsymbol{\epsilon}}}(\boldsymbol{\xi}_{\underline{\boldsymbol{\epsilon}}}) - \nu(\mathtt{r}_0)\bigr\}. \tag{5.4}$$

Now (5.2) and (5.4) imply

$$\frac{\int_{\Upsilon} \exp\bigl\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^{*})\bigr\} f\bigl(\breve{D}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}-\boldsymbol{\theta}_{\boldsymbol{\epsilon}})\bigr)\, d\boldsymbol{\upsilon}}{\int_{\Upsilon} \exp\bigl\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^{*})\bigr\}\, d\boldsymbol{\upsilon}} \\ \leq \exp\bigl\{\diamondsuit_{\boldsymbol{\epsilon}}(\mathtt{r}_0) + \diamondsuit_{\underline{\boldsymbol{\epsilon}}}(\mathtt{r}_0) + m_{\underline{\boldsymbol{\epsilon}}}(\boldsymbol{\xi}_{\underline{\boldsymbol{\epsilon}}}) - m_{\boldsymbol{\epsilon}}(\boldsymbol{\xi}_{\boldsymbol{\epsilon}}) + \nu(\mathtt{r}_0) + \rho_f(\mathtt{r}_0)\bigr\}\, I\!E f(\boldsymbol{\gamma})$$

and (4.10) follows by definition of $m_{\boldsymbol{\epsilon}}(\boldsymbol{\xi}_{\boldsymbol{\epsilon}})$, $m_{\underline{\boldsymbol{\epsilon}}}(\boldsymbol{\xi}_{\underline{\boldsymbol{\epsilon}}})$, $\Delta_{\boldsymbol{\epsilon}}(\mathtt{r}_0)$ and $\Delta_{\boldsymbol{\epsilon}}^{+}(\mathtt{r}_0)$.

5.3 Proof of Corollary 4.7

As a direct implication of (4.10) one easily gets

$$I\!E^{\circ}\bigl|\boldsymbol{\lambda}^{\top}\breve{D}_{\boldsymbol{\epsilon}}(\boldsymbol{\vartheta}-\boldsymbol{\theta}_{\boldsymbol{\epsilon}})\bigr|^{2} \leq \exp(\Delta_{\boldsymbol{\epsilon}}^{+}(\mathtt{r}_0))\|\boldsymbol{\lambda}\|^{2}. \tag{5.5}$$

Furthermore,

$$I\!E^{\circ}\bigl|\boldsymbol{\lambda}^{\top}\breve{D}_0(\boldsymbol{\vartheta}-\breve{\boldsymbol{\theta}})\bigr|^{2} \\ = (1-\boldsymbol{\epsilon})^{-1} I\!E^{\circ}\Bigl|\boldsymbol{\lambda}^{\top}\breve{D}_{\boldsymbol{\epsilon}}(\boldsymbol{\vartheta}-\boldsymbol{\theta}_{\boldsymbol{\epsilon}}) + \frac{\boldsymbol{\epsilon}}{1-\boldsymbol{\epsilon}}\,\boldsymbol{\lambda}^{\top}\breve{\boldsymbol{\xi}}\Bigr|^{2} \\ \leq \frac{1}{1-\boldsymbol{\epsilon}} I\!E^{\circ}\bigl|\boldsymbol{\lambda}^{\top}\breve{D}_{\boldsymbol{\epsilon}}(\boldsymbol{\vartheta}-\boldsymbol{\theta}_{\boldsymbol{\epsilon}})\bigr|^{2} + \frac{2\boldsymbol{\epsilon}|\boldsymbol{\lambda}^{\top}\breve{\boldsymbol{\xi}}|}{(1-\boldsymbol{\epsilon})^{2}}\, I\!E^{\circ}\bigl|\boldsymbol{\lambda}^{\top}\breve{D}_{\boldsymbol{\epsilon}}(\boldsymbol{\vartheta}-\boldsymbol{\theta}_{\boldsymbol{\epsilon}})\bigr| + \frac{\boldsymbol{\epsilon}^{2}|\boldsymbol{\lambda}^{\top}\breve{\boldsymbol{\xi}}|^{2}}{(1-\boldsymbol{\epsilon})^{3}}.$$

This, bound (5.5) and the elementary inequality $\mathrm{e}^{a} + b \le \mathrm{e}^{a+b}$ for $a, b \ge 0$, imply for $\epsilon \le 1/2$

$$\begin{aligned} & I\!E^{\circ}\bigl|\boldsymbol{\lambda}^{\top}\breve{D}_0(\boldsymbol{\vartheta} - \breve{\boldsymbol{\theta}})\bigr|^2 \\ & \quad \le \Bigl[\exp\bigl(\Delta_{\epsilon}^{+}(\mathtt{r}_0) - \log(1-\epsilon)\bigr) + \mathtt{C}\epsilon\exp\bigl(\Delta_{\epsilon}^{+}(\mathtt{r}_0)/2\bigr)\|\breve{\boldsymbol{\xi}}\| + \mathtt{C}\epsilon^2\|\breve{\boldsymbol{\xi}}\|^2\Bigr]\|\boldsymbol{\lambda}\|^2 \\ & \quad \le \exp\Bigl(\Delta_{\epsilon}^{+}(\mathtt{r}_0) - \log(1-\epsilon) + \mathtt{C}\epsilon\exp\bigl(\Delta_{\epsilon}^{+}(\mathtt{r}_0)/2\bigr)\|\breve{\boldsymbol{\xi}}\| + \mathtt{C}\epsilon^2\|\breve{\boldsymbol{\xi}}\|^2\Bigr)\|\boldsymbol{\lambda}\|^2 \end{aligned}$$

and (4.11) follows from definition of $\Delta_{\epsilon}^{+}(\mathtt{r}_0)$. It can be easily shown that on $\Omega(\mathtt{x})$ $\Delta_{\epsilon}^{+}(\mathtt{r}_0) \le \mathtt{C}\epsilon\mathfrak{p}$. The only important step is to show that $\rho_{x^2}(\mathtt{r}_0)$ is small. It follows

$$\begin{aligned} \rho_{x^2}(\mathtt{r}_0) &= \frac{\int_{\Upsilon\setminus\Upsilon_0(\mathtt{r}_0)}\bigl|\boldsymbol{\lambda}^{\top}\breve{D}_{\epsilon}(\boldsymbol{\theta}-\boldsymbol{\theta}_{\epsilon})\bigr|^2\,\exp\bigl\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\bigr\}\,d\boldsymbol{\upsilon}}{\int_{\Upsilon_0(\mathtt{r}_0)}\bigl|\boldsymbol{\lambda}^{\top}\breve{D}_{\epsilon}(\boldsymbol{\theta}-\boldsymbol{\theta}_{\epsilon})\bigr|^2\,\exp\bigl\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\bigr\}d\boldsymbol{\upsilon}} \\ &\le \frac{\int_{\Upsilon\setminus\Upsilon_0(\mathtt{r}_0)}\bigl|\boldsymbol{\lambda}^{\top}\breve{D}_{\epsilon}(\boldsymbol{\theta}-\boldsymbol{\theta}_{\epsilon})\bigr|^2\,\exp\bigl\{-\mathtt{b}\|\mathcal{D}_0(\boldsymbol{\upsilon}-\boldsymbol{\upsilon}^*)\|^2/2\bigr\}\,d\boldsymbol{\upsilon}}{\int_{\Upsilon_0(\mathtt{r}_0)}\bigl|\boldsymbol{\lambda}^{\top}\breve{D}_{\epsilon}(\boldsymbol{\theta}-\boldsymbol{\theta}_{\epsilon})\bigr|^2\,\exp\bigl\{\mathbb{L}_{\underline{\epsilon}}(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\bigr\}d\boldsymbol{\upsilon}} \\ &= \frac{\int_{\Upsilon\setminus\Upsilon_0(\mathtt{r}_0)}\bigl|\boldsymbol{\lambda}_{\epsilon}^{\top}\mathcal{D}_{\epsilon}(\boldsymbol{\upsilon}-\boldsymbol{\upsilon}^*)\bigr|^2\,\exp\bigl\{-\mathtt{b}\|\mathcal{D}_0(\boldsymbol{\upsilon}-\boldsymbol{\upsilon}^*)\|^2/2\bigr\}d\boldsymbol{\upsilon}}{\int_{\Upsilon_0(\mathtt{r}_0)}\bigl|\boldsymbol{\lambda}_{\epsilon}^{\top}\mathcal{D}_{\epsilon}(\boldsymbol{\upsilon}-\boldsymbol{\upsilon}_{\underline{\epsilon}})\bigr|^2\,\exp\bigl\{\mathbb{L}_{\underline{\epsilon}}(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\bigr\}d\boldsymbol{\upsilon}} \\ &\le \mathtt{C}\frac{\int_{\mathbb{R}^p\setminus\Upsilon_0(\mathtt{r}_0)}\bigl|\boldsymbol{\lambda}_0^{\top}\mathcal{D}_0(\boldsymbol{\upsilon}-\boldsymbol{\upsilon}^*)\bigr|^2\,\exp\bigl\{-\mathtt{b}\|\mathcal{D}_0(\boldsymbol{\upsilon}-\boldsymbol{\upsilon}^*)\|^2/2\bigr\}d\boldsymbol{\upsilon}}{\int_{\Upsilon_0(\mathtt{r}_0)}\bigl|\boldsymbol{\lambda}_{\underline{\epsilon}}^{\top}\mathcal{D}_{\underline{\epsilon}}(\boldsymbol{\upsilon}-\boldsymbol{\upsilon}_{\underline{\epsilon}})\bigr|^2\,\exp\bigl\{\mathbb{L}_{\underline{\epsilon}}(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\bigr\}\,d\boldsymbol{\upsilon}} \\ &\le \mathtt{C}\;\frac{I\!E\bigl|\boldsymbol{\lambda}_0^{\top}\boldsymbol{\gamma}\bigr|^2\,\mathbb{1}\{\|\boldsymbol{\gamma}\| > \mathtt{C}\mathtt{r}_0\}}{I\!E\bigl|\boldsymbol{\lambda}_{\underline{\epsilon}}^{\top}\boldsymbol{\gamma}\bigr|^2\,\mathbb{1}\{\|\boldsymbol{\gamma}\| < \mathtt{C}\mathtt{r}_0\}} = \mathtt{C}(1+\epsilon)\;\frac{I\!E\bigl|\boldsymbol{\lambda}_0^{\top}\boldsymbol{\gamma}\bigr|^2\,\mathbb{1}\{\|\boldsymbol{\gamma}\| > \mathtt{C}\mathtt{r}_0\}}{I\!E\bigl|\boldsymbol{\lambda}_0^{\top}\boldsymbol{\gamma}\bigr|^2\,\mathbb{1}\{\|\boldsymbol{\gamma}\| < \mathtt{C}\mathtt{r}_0\}}, \end{aligned}$$

where

$$\boldsymbol{\lambda}_0 = \mathcal{D}_0^{-1}\begin{pmatrix}\breve{D}_{\epsilon}\boldsymbol{\lambda}\\ \mathbf{0}\end{pmatrix}, \qquad \boldsymbol{\lambda}_{\epsilon} = \mathcal{D}_{\epsilon}^{-1}\begin{pmatrix}\breve{D}_{\epsilon}\boldsymbol{\lambda}\\ \mathbf{0}\end{pmatrix} = (1-\epsilon)^{-1/2}\boldsymbol{\lambda}_0,$$

and similarly for $\boldsymbol{\lambda}_{\underline{\epsilon}}$, and $\mathbf{0}$ is a zero vector of dimension $(p-q)$. A choice of $\mathtt{r}_0^2 = \mathtt{C}(p+\mathtt{x})$ with a proper absolute constant $\mathtt{C}$ ensures that $\rho_{x^2}(\mathtt{r}_0) \le \mathrm{e}^{-\mathtt{x}}$.

5.4 Proof of Corollary 4.8

The first statement (4.12) follows from Theorem 4.6 with $f(\boldsymbol{u}) = \mathbb{1}\bigl(D_1\breve{D}_{\epsilon}^{-1}\boldsymbol{u} + \boldsymbol{\delta}_{\epsilon} \in A\bigr)$. Further, it holds on $\Omega(\mathtt{x})$ for $\boldsymbol{\delta}_{\epsilon} \stackrel{\mathrm{def}}{=} D_1(\boldsymbol{\theta}_{\epsilon} - \widehat{\boldsymbol{\theta}})$

$$\begin{aligned} \|\boldsymbol{\delta}_{\epsilon}\|^2 &= \|D_1(\boldsymbol{\theta}_{\epsilon} - \widehat{\boldsymbol{\theta}})\|^2 \le (1 + \mathtt{C}\,\epsilon\,\mathfrak{p})\|\breve{D}_0(\boldsymbol{\theta}_{\epsilon} - \widehat{\boldsymbol{\theta}})\|^2 \\ &\le (1 + \mathtt{C}\,\epsilon\,\mathfrak{p})\bigl(\|\breve{D}_0(\boldsymbol{\theta}_{\epsilon} - \breve{\boldsymbol{\theta}})\|^2 + \|\breve{D}_0(\breve{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}})\|^2\bigr) \\ &\le (1 + \mathtt{C}\,\epsilon\,\mathfrak{p})\bigl(4\epsilon^2\|\breve{\boldsymbol{\xi}}\|^2 + \mathtt{C}\,\epsilon\,\mathfrak{p}\bigr) \le \mathtt{C}(\epsilon^2\mathtt{q} + \epsilon\,\mathfrak{p}). \end{aligned} \tag{5.6}$$

For proving (4.13), we compute the Kullback–Leibler divergence between two multivariate normal distributions and apply Pinsker's inequality. Let $\boldsymbol{\gamma}$ be standard normal in $I\!R^q$, and $I\!P_0$ stand for its distribution. The random variable $D_1 \breve{D}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{\gamma} + \boldsymbol{\delta}_{\boldsymbol{\epsilon}}$ is normal with mean $\boldsymbol{\delta}_{\boldsymbol{\epsilon}}$ and variance $B_{\boldsymbol{\epsilon}}^{-1} \stackrel{\mathrm{def}}{=} D_1 \breve{D}_{\boldsymbol{\epsilon}}^{-2} D_1$. Denote this distribution by $I\!P_{\boldsymbol{\epsilon}}$. Obviously

$$\|\boldsymbol{I}_q - B_{\boldsymbol{\epsilon}}\|_\infty = \|\boldsymbol{I}_q - D_1^{-1}\breve{D}_{\boldsymbol{\epsilon}}^2 D_1^{-1}\|_\infty = \|\boldsymbol{I}_q - (1-\boldsymbol{\epsilon})D_1^{-1}\breve{D}_0^2 D_1^{-1}\|_\infty$$

which implies $\|B_{\boldsymbol{\epsilon}} - \boldsymbol{I}_q\|_\infty \le \mathtt{C}\,\boldsymbol{\epsilon}\,\mathfrak{p}$. We use the following technical lemma.

Lemma 5.1. *Let* $\|B_{\boldsymbol{\epsilon}} - \boldsymbol{I}_q\|_\infty \le \alpha_{\boldsymbol{\epsilon}} \le 1/2$. *Then*

$$\begin{aligned} 2\mathcal{K}(I\!P_0, I\!P_{\boldsymbol{\epsilon}}) &= -2I\!E_0 \log \frac{dI\!P_{\boldsymbol{\epsilon}}}{dI\!P_0}(\boldsymbol{\gamma}) \\ &\le \operatorname{tr}(B_{\boldsymbol{\epsilon}} - \boldsymbol{I}_q)^2 + (1+\alpha_{\boldsymbol{\epsilon}})\|\boldsymbol{\delta}_{\boldsymbol{\epsilon}}\|^2 = \alpha_{\boldsymbol{\epsilon}}^2\, q + (1+\alpha_{\boldsymbol{\epsilon}})\|\boldsymbol{\delta}_{\boldsymbol{\epsilon}}\|^2. \end{aligned}$$

Proof. It holds

$$2\log \frac{dI\!P_{\boldsymbol{\epsilon}}}{dI\!P_0}(\boldsymbol{\gamma}) = -\log\det(B_{\boldsymbol{\epsilon}}) - (\boldsymbol{\gamma} - \boldsymbol{\delta}_{\boldsymbol{\epsilon}})^\top B_{\boldsymbol{\epsilon}}(\boldsymbol{\gamma} - \boldsymbol{\delta}_{\boldsymbol{\epsilon}}) + \|\boldsymbol{\gamma}\|^2$$

and

$$\begin{aligned} 2\mathcal{K}(I\!P_0, I\!P_{\boldsymbol{\epsilon}}) &= -2I\!E_0 \log \frac{dI\!P_{\boldsymbol{\epsilon}}}{dI\!P_0}(\boldsymbol{\gamma}) \\ &= \log\det(B_{\boldsymbol{\epsilon}}) + \operatorname{tr}(B_{\boldsymbol{\epsilon}} - \boldsymbol{I}_q) + \boldsymbol{\delta}_{\boldsymbol{\epsilon}}^\top B_{\boldsymbol{\epsilon}} \boldsymbol{\delta}_{\boldsymbol{\epsilon}}. \end{aligned}$$

Denote by a_j the jth eigenvalue of $B_{\boldsymbol{\epsilon}} - \boldsymbol{I}_q$. Then $\|B_{\boldsymbol{\epsilon}} - \boldsymbol{I}_q\|_\infty \le \alpha_{\boldsymbol{\epsilon}} \le 1/2$ yields $|a_j| \le 1/2$ and

$$\begin{aligned} 2\mathcal{K}(I\!P_0, I\!P_{\boldsymbol{\epsilon}}) &= \boldsymbol{\delta}_{\boldsymbol{\epsilon}}^\top B_{\boldsymbol{\epsilon}} \boldsymbol{\delta}_{\boldsymbol{\epsilon}} + \sum_{j=1}^{q} \bigl\{ a_j - \log(1+a_j) \bigr\} \le (1+\alpha_{\boldsymbol{\epsilon}})\|\boldsymbol{\delta}_{\boldsymbol{\epsilon}}\|^2 + \sum_{j=1}^{q} a_j^2 \\ &\le (1+\alpha_{\boldsymbol{\epsilon}})\|\boldsymbol{\delta}_{\boldsymbol{\epsilon}}\|^2 + \operatorname{tr}(B_{\boldsymbol{\epsilon}} - \boldsymbol{I}_q)^2 \le (1+\alpha_{\boldsymbol{\epsilon}})\|\boldsymbol{\delta}_{\boldsymbol{\epsilon}}\|^2 + \alpha_{\boldsymbol{\epsilon}}^2\, q. \end{aligned}$$

as required. □

This lemma with $\alpha_{\boldsymbol{\epsilon}} = \mathtt{C}\,\boldsymbol{\epsilon}\,\mathfrak{p} \le \mathtt{C}$ and (5.6) imply by Pinsker's inequality

$$\|I\!P_0 - I\!P_{\boldsymbol{\epsilon}}\|_{TV}^2 \le \frac{1}{2}\mathcal{K}(I\!P_0, I\!P_{\boldsymbol{\epsilon}}) \le \mathtt{C}(\boldsymbol{\epsilon}^2 \mathfrak{q} + \boldsymbol{\epsilon}\,\mathfrak{p}) + \mathtt{C}\,\boldsymbol{\epsilon}^2\mathfrak{p}^2\, q \le \mathtt{C}(\boldsymbol{\epsilon}\,\mathfrak{p} + \boldsymbol{\epsilon}^2\mathfrak{p}^2 q).$$

Equivalently, for any measurable set A, it holds

$$I\!P\bigl(D_1 \breve{D}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{\gamma} + \boldsymbol{\delta}_{\boldsymbol{\epsilon}} \in A \bigm| \boldsymbol{Y}\bigr) \le I\!P\bigl(\boldsymbol{\gamma} \in A\bigr) + \sqrt{\mathtt{C}(\boldsymbol{\epsilon}\,\mathfrak{p} + \boldsymbol{\epsilon}^2\mathfrak{p}^2 q)}.$$

5.5 Proof of Theorem 4.9

As in proof of Theorem 4.6, consider $m_{\underline{\epsilon}}(\boldsymbol{\xi}_{\underline{\epsilon}})$ which is defined by (5.1) with $\underline{\boldsymbol{\epsilon}}$ in place of $\boldsymbol{\epsilon}$. The sum $m_{\underline{\epsilon}}(\boldsymbol{\xi}_{\underline{\epsilon}}) + \mathbb{L}_{\underline{\epsilon}}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*)$ is conditionally on $\boldsymbol{Y}$ the log-density of the normal law of $\boldsymbol{\upsilon} = (\boldsymbol{\theta}, \boldsymbol{\eta})$. Its $\boldsymbol{\theta}$-marginal obtained by integration w.r.t. $\boldsymbol{\eta}$ is again the normal density with the mean $\boldsymbol{\theta}_{\underline{\epsilon}} = \breve{D}_{\underline{\epsilon}}^{-1} \breve{\boldsymbol{\xi}}_{\underline{\epsilon}} + \boldsymbol{\theta}^*$ and the covariance matrix $\breve{D}_{\underline{\epsilon}}^{-2}$. So, for any nonnegative function $f: \mathbb{R}^q \to \mathbb{R}_+$, it holds

$$
\begin{aligned}
& \int_{\Upsilon} \exp\bigl\{ L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) \bigr\} \, f\bigl(\breve{D}_{\underline{\epsilon}} (\boldsymbol{\theta} - \boldsymbol{\theta}_{\underline{\epsilon}}) \bigr) \, \mathbb{I}\bigl\{ \boldsymbol{\theta} \in \Theta_0(\mathtt{r}_0) \bigr\} \, d\boldsymbol{\upsilon} \\
& \quad \geq \int_{\Upsilon_0(\mathtt{r}_0)} \exp\bigl\{ L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) \bigr\} \, f\bigl(\breve{D}_{\underline{\epsilon}} (\boldsymbol{\theta} - \boldsymbol{\theta}_{\underline{\epsilon}}) \bigr) \, d\boldsymbol{\upsilon} \\
& \quad \geq \exp\{ -\diamondsuit_{\underline{\epsilon}} - m_{\underline{\epsilon}}(\boldsymbol{\xi}_{\underline{\epsilon}}) \} \int_{\Upsilon_0(\mathtt{r}_0)} \exp\bigl\{ \mathbb{L}_{\underline{\epsilon}}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) \bigr\} \, f\bigl(\breve{D}_{\underline{\epsilon}} (\boldsymbol{\theta} - \boldsymbol{\theta}_{\underline{\epsilon}}) \bigr) \, d\boldsymbol{\upsilon} \\
& \quad \geq \exp\{ -\diamondsuit_{\underline{\epsilon}} - m_{\underline{\epsilon}}(\boldsymbol{\xi}_{\underline{\epsilon}}) \} \int_{\mathbb{R}^p} \exp\bigl\{ \mathbb{L}_{\underline{\epsilon}}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) \bigr\} \, f\bigl(\breve{D}_{\underline{\epsilon}} (\boldsymbol{\theta} - \boldsymbol{\theta}_{\underline{\epsilon}}) \bigr) \, d\boldsymbol{\upsilon} \\
& \qquad - \exp\{ -\diamondsuit_{\underline{\epsilon}} - m_{\underline{\epsilon}}(\boldsymbol{\xi}_{\underline{\epsilon}}) \} \int_{\mathbb{R}^p \setminus \Upsilon_0(\mathtt{r}_0)} \exp\bigl\{ \mathbb{L}_{\underline{\epsilon}}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) \bigr\} \, f\bigl(\breve{D}_{\underline{\epsilon}} (\boldsymbol{\theta} - \boldsymbol{\theta}_{\underline{\epsilon}}) \bigr) \, d\boldsymbol{\upsilon} \\
& \quad = \exp\{ -\diamondsuit_{\underline{\epsilon}} - m_{\underline{\epsilon}}(\boldsymbol{\xi}_{\underline{\epsilon}}) \} (1 - \widetilde{\rho}_f) \int_{\mathbb{R}^p} \exp\bigl\{ \mathbb{L}_{\underline{\epsilon}}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) \bigr\} \, f\bigl(\breve{D}_{\underline{\epsilon}} (\boldsymbol{\theta} - \boldsymbol{\theta}_{\underline{\epsilon}}) \bigr) \, d\boldsymbol{\upsilon} \\
& \quad \geq \exp\{ -\diamondsuit_{\underline{\epsilon}} - m_{\underline{\epsilon}}(\boldsymbol{\xi}_{\underline{\epsilon}}) \} (1 - \widetilde{\rho}_f) \int_{\Theta_0(\mathtt{r}_0) \times \mathbb{R}^{(p-q)}} \exp\bigl\{ \mathbb{L}_{\underline{\epsilon}}(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) \bigr\} \, f\bigl(\breve{D}_{\underline{\epsilon}} (\boldsymbol{\theta} - \boldsymbol{\theta}_{\underline{\epsilon}}) \bigr) \, d\boldsymbol{\upsilon} \\
& \quad \geq \exp\{ -\diamondsuit_{\underline{\epsilon}} - m_{\underline{\epsilon}}(\boldsymbol{\xi}_{\underline{\epsilon}}) - 2\widetilde{\rho}_f \} \, \mathbb{E} f(\boldsymbol{\gamma}) \, \mathbb{I}\{ \|\boldsymbol{\gamma}\| \leq \mathtt{C}\mathtt{r}_0 \},
\end{aligned}
\tag{5.7}
$$

Here we used that $1 - \alpha \geq \mathrm{e}^{-2\alpha}$ for $0 \leq \alpha \leq \frac{1}{2}$. Similarly,

$$
\begin{aligned}
& \int_{\Upsilon} \exp\bigl\{ L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) \bigr\} \, d\boldsymbol{\upsilon} = \int_{\Upsilon_0(\mathtt{r}_0)} \exp\bigl\{ L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) \bigr\} \, d\boldsymbol{\upsilon} + \int_{\Upsilon \setminus \Upsilon_0(\mathtt{r}_0)} \exp\bigl\{ L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) \bigr\} \, d\boldsymbol{\upsilon} \\
& \quad = \{ 1 + \rho^*(\mathtt{r}_0) \} \int_{\Upsilon_0(\mathtt{r}_0)} \exp\bigl\{ L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) \bigr\} \, d\boldsymbol{\upsilon} \\
& \quad \leq \{ 1 + \rho^*(\mathtt{r}_0) \} \exp\{ \diamondsuit_{\boldsymbol{\epsilon}} - m_{\boldsymbol{\epsilon}}(\boldsymbol{\xi}_{\boldsymbol{\epsilon}}) \} \, \mathbb{P}\bigl(\bigl\| \mathcal{D}_0 \mathcal{D}_{\boldsymbol{\epsilon}}^{-1} (\boldsymbol{\gamma} + \boldsymbol{\xi}_{\boldsymbol{\epsilon}}) \bigr\| \leq \mathtt{r}_0 \bigm| \boldsymbol{Y} \bigr).
\end{aligned}
$$

Further, by construction, $\mathcal{D}_{\boldsymbol{\epsilon}}^2 \leq \mathcal{D}_0^2$ and $\|\boldsymbol{\xi}_{\underline{\epsilon}}\| \geq \|\boldsymbol{\xi}\|$, yielding

$$
\bigl\{ \mathcal{D}_{\boldsymbol{\epsilon}}^{-1} \bigl(\boldsymbol{u} + \boldsymbol{\xi}_{\boldsymbol{\epsilon}} \bigr) \in \Upsilon_0(\mathtt{r}_0) \bigr\} = \bigl\{ \bigl\| \mathcal{D}_0 \mathcal{D}_{\boldsymbol{\epsilon}}^{-1} \bigl(\boldsymbol{u} + \boldsymbol{\xi}_{\boldsymbol{\epsilon}} \bigr) \bigr\| \leq \mathtt{r}_0 \bigr\} \subset \bigl\{ \|\boldsymbol{u} + \boldsymbol{\xi}\| \leq \mathtt{r}_0 \bigr\},
$$

and finally

$$
\int_{\Upsilon} \exp\bigl\{ L(\boldsymbol{\upsilon}, \boldsymbol{\upsilon}^*) \bigr\} \, d\boldsymbol{\upsilon} \leq \exp\{ \diamondsuit_{\boldsymbol{\epsilon}} - m_{\boldsymbol{\epsilon}}(\boldsymbol{\xi}_{\boldsymbol{\epsilon}}) + \nu(\mathtt{r}_0) + \rho^*(\mathtt{r}_0) \}. \tag{5.8}
$$

The bounds (5.7) and (5.8) imply

$$
\begin{aligned}
&\frac{\int_{\Upsilon} \exp\bigl\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\bigr\} f\bigl(\breve{D}_{\underline{\boldsymbol{\epsilon}}}(\boldsymbol{\theta}-\boldsymbol{\theta}_{\underline{\boldsymbol{\epsilon}}})\bigr)\, d\boldsymbol{\upsilon}}{\int_{\Upsilon} \exp\bigl\{L(\boldsymbol{\upsilon},\boldsymbol{\upsilon}^*)\bigr\}\, d\boldsymbol{\upsilon}} \\
&\qquad \ge \frac{\exp\{-\diamondsuit_{\underline{\boldsymbol{\epsilon}}} - m_{\underline{\boldsymbol{\epsilon}}}(\boldsymbol{\xi}_{\underline{\boldsymbol{\epsilon}}}) - 2\widetilde{\rho}_f\}\, I\!E f(\boldsymbol{\gamma})\, \mathbb{I}\{\|\boldsymbol{\gamma}\| \le \mathtt{C}\mathtt{r}_0\}}{\exp\bigl\{\diamondsuit_{\boldsymbol{\epsilon}} - m_{\boldsymbol{\epsilon}}(\boldsymbol{\xi}_{\boldsymbol{\epsilon}}) + \nu(\mathtt{r}_0) + \rho^*(\mathtt{r}_0)\bigr\}} \\
&\qquad \ge \frac{\exp\{-\Delta_{\boldsymbol{\epsilon}} - \varkappa_{\boldsymbol{\epsilon}} - 2\widetilde{\rho}_f\} I\!E f(\boldsymbol{\gamma})\, \mathbb{I}\{\|\boldsymbol{\gamma}\| \le \mathtt{C}\mathtt{r}_0\}}{\exp\{\nu(\mathtt{r}_0) + \rho^*(\mathtt{r}_0)}\,.
\end{aligned}
$$

This yields (4.14).

5.6 Proof of Theorem 2.1

Due to our previous results, it is convenient to decompose the r.v. $\boldsymbol{\vartheta}$ in the form

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

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

$$
\overline{\boldsymbol{\vartheta}}_{\circ} \stackrel{\text{def}}{=} I\!E^{\circ}\boldsymbol{\vartheta}, \qquad \mathfrak{S}_{\circ}^2 \stackrel{\text{def}}{=} \operatorname{Cov}(\boldsymbol{\vartheta}^{\circ}) \stackrel{\text{def}}{=} I\!E^{\circ}\bigl\{(\boldsymbol{\vartheta}-\overline{\boldsymbol{\vartheta}}_{\circ})(\boldsymbol{\vartheta}-\overline{\boldsymbol{\vartheta}}_{\circ})^{\top}\bigr\}.
$$

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

$$
\begin{aligned}
\|\breve{D}_0(\overline{\boldsymbol{\vartheta}}_{\circ} - \breve{\boldsymbol{\theta}})\|^2 &\le \mathtt{C}\Delta_{\boldsymbol{\epsilon}}^* \\
\bigl\|I_q - \breve{D}_0\mathfrak{S}_{\circ}^2\breve{D}_0\bigr\|_{\infty} &\le \mathtt{C}\Delta_{\boldsymbol{\epsilon}}^*,
\end{aligned}
$$

where $\Delta_{\boldsymbol{\epsilon}}^* = \max\bigl\{\Delta_{\boldsymbol{\epsilon}}^{\oplus}, \Delta_{\boldsymbol{\epsilon}}^{\ominus}\bigr\} \le \mathtt{C}\,\boldsymbol{\epsilon}\,\mathfrak{p}$.

Consider $\boldsymbol{\eta} \stackrel{\text{def}}{=} \breve{D}_0(\boldsymbol{\vartheta}-\breve{\boldsymbol{\theta}})$. Corollaries 4.7 and 4.10 yield for any $\boldsymbol{\lambda} \in I\!R^q$ that

$$
\|\boldsymbol{\lambda}\|^2 \exp(-\Delta^-) \le I\!E^{\circ}\bigl|\boldsymbol{\lambda}^{\top}\boldsymbol{\eta}\bigr|^2 \le \|\boldsymbol{\lambda}\|^2 \exp(\Delta^+) \tag{5.9}
$$

with $\Delta^- = \Delta_{\boldsymbol{\epsilon}}^{\ominus}$ and $\Delta^+ = \Delta_{\boldsymbol{\epsilon}}^{\oplus}$. Define the first two moments of $\boldsymbol{\eta}$:

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

Use the following technical statement.

Lemma 5.2. *Assume* (5.9). *Then with* $\Delta^* = \max\bigl\{\Delta^+, \Delta^-\bigr\} \le 1/2$

$$
\|\overline{\boldsymbol{\eta}}\|^2 \le \mathtt{C}\Delta^*, \qquad \|S_{\circ}^2 - I_q\|_{\infty} \le \mathtt{C}\Delta^*. \tag{5.10}
$$

Proof. Let $\boldsymbol{u}$ be any unit vector in $I\!\!R^q$. We obtain from (5.9)

$$\exp(-\Delta^{-}) \le I\!\!E^{\circ}\bigl|\boldsymbol{u}^{\top}\boldsymbol{\eta}\bigr|^2 \le \exp(\Delta^{+}).$$

Note now that

$$I\!\!E^{\circ}\bigl|\boldsymbol{u}^{\top}\boldsymbol{\eta}\bigr|^2 = \boldsymbol{u}^{\top} S_{\circ}^2 \boldsymbol{u} + |\boldsymbol{u}^{\top}\overline{\boldsymbol{\eta}}|^2.$$

Hence

$$\exp(-\Delta^{-}) \le \boldsymbol{u}^{\top} S_{\circ}^2 \boldsymbol{u} + |\boldsymbol{u}^{\top}\overline{\boldsymbol{\eta}}|^2 \le \exp(\Delta^{+}). \tag{5.11}$$

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

$$I\!\!E^{\circ}\bigl|\boldsymbol{u}^{\top}(\boldsymbol{\eta} - \overline{\boldsymbol{\eta}})\bigr|^2 \ge \mathrm{e}^{-\Delta^{-}} I\!\!E\bigl|\boldsymbol{u}^{\top}(\boldsymbol{\gamma} - \overline{\boldsymbol{\eta}})\bigr|^2 = \mathrm{e}^{-\Delta^{-}}\bigl(1 + \|\overline{\boldsymbol{\eta}}\|^2\bigr)$$

yielding

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

This inequality contradicts (5.11) if $\|\overline{\boldsymbol{\eta}}\|^2 > 2\Delta^{*} > 1$, and (5.10) follows. □

The bound for the first moment implies with $\overline{\boldsymbol{\vartheta}}_{\circ} = I\!\!E^{\circ}\boldsymbol{\vartheta}$

$$\bigl\|\breve{D}_0(\overline{\boldsymbol{\vartheta}}_{\circ} - \breve{\boldsymbol{\theta}})\bigr\|^2 \le \mathtt{C}\Delta_{\boldsymbol{\epsilon}}^{*}$$

while the second bound yields with

$$\bigl\|\breve{D}_0 \mathfrak{S}_{\circ}^2 \breve{D}_0 - I_q\bigr\|_{\infty} \le \mathtt{C}\Delta_{\boldsymbol{\epsilon}}^{*}.$$

The last result follows from Corollary 4.8 and 4.10 with $D_1 = \mathfrak{S}^{-1}$ and $\widehat{\boldsymbol{\theta}} = \overline{\boldsymbol{\vartheta}}$.

5.7 Proof of Theorem 3.1

The bracketing bound and the large deviation result of Theorem 4.1 apply if the sample size n fulfills $n \ge \mathtt{C}(p_n + \mathtt{x})$ for a fixed constant $\mathtt{C}$. It appears that the BvM result requires a stronger condition. Indeed, in the regular i.i.d. case it holds

$$\delta(\mathtt{r}_0) \asymp \mathtt{r}_0/\sqrt{n}, \qquad \omega(\mathtt{r}_0) \asymp \mathtt{r}_0/\sqrt{n}.$$

The radius $\mathtt{r}_0$ should fulfill $\mathtt{r}_0^2 \ge \mathtt{C}(p_n + \mathtt{x})$ to ensure the large deviation result. This yields

$$\boldsymbol{\epsilon} = \delta(\mathtt{r}_0) + 3\nu_0\mathfrak{a}^2\omega(\mathtt{r}_0) \ge \mathtt{C}\sqrt{(p_n + \mathtt{x})/n}.$$

If we fix $\mathtt{x} = \mathtt{C}p_n$, our BvM result requires the condition "$\boldsymbol{\epsilon}\, p_n$ is small", which effectively means that $p_n^3/n \to 0$ as $n \to \infty$.

5.8 Proof of Theorem 3.2

First we check that the required conditions of Section 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 $\boldsymbol{u}^* = (u_1^*, \ldots, u_{p_n}^*)^\top$, define $\Upsilon_0(\sqrt{\mathfrak{z}})$ as a rectangle

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

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

$$I\!P\bigl(\widetilde{\boldsymbol{u}} \in \Upsilon_0(\sqrt{\mathfrak{z}_n})\bigr) \ge 1 - 4p_n \mathrm{e}^{-\mathfrak{z}_n}. \tag{5.12}$$

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

$$I\!P\bigl(\widetilde{\boldsymbol{u}} \in \Upsilon_0(\sqrt{\mathfrak{z}_n})\bigr) \ge 1 - 4\mathrm{e}^{-\mathtt{x}_n}. \tag{5.13}$$

Proof. We use the bound from Polzehl and Spokoiny (2006)

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

This yields

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

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 (5.12) follows. □

In the special case $u_1^* = \ldots = u_{p_n}^* = u^*$, the set $\Upsilon_0(\sqrt{\mathfrak{z}})$ is a cube which can be also viewed as a ball in the sup-norm. Moreover, if $\mathfrak{z}_n/(M_n \mathrm{e}^{u^*}) \le 1/2$, this cube is contained in the cube $\bigl\{\boldsymbol{u}:\ \|\boldsymbol{u} - \boldsymbol{u}^*\|_\infty \le \sqrt{\mathfrak{z}_n/(M_n \mathrm{e}^{u^*})}\bigr\}$ in view of $\mathrm{e}^{x} - 1 - x \le a^2 \le 1/2$ for $|x| \le a \le 1$. The concentration bound (5.13) enables us to check the local conditions only on the cube $\Upsilon_0(\sqrt{\mathfrak{z}_n})$. Especially the condition $(\mathcal{E}\mathcal{D}_1)$ is trivially fulfilled because $\boldsymbol{\zeta}(\boldsymbol{u}) = \mathcal{L}(\boldsymbol{u}) - I\!E\mathcal{L}(\boldsymbol{u})$ is linear in $\boldsymbol{u}$ and θ is a linear functional of $\boldsymbol{u}$. 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 $\breve{D}_0^2$.

Lemma 5.4. *Let* $\upsilon^* = 1/p_n$. *Then it holds*

$$\breve{D}_0^2 = p_n^2 \beta_n^{-2}.$$

Now we are ready to finalize the proof Theorem 3.2.

Proof. Let β_n be bounded. The definition implies

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

The posterior distribution $\upsilon_j \bigm| \boldsymbol{Y}$ is $\mathrm{Gamma}(\alpha_j, \mu_j)$ with $\alpha_j = 1 + Z_j$ and $\mu_j = \frac{\mu}{M_n\mu+1}$. We use 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

$$\gamma_j \stackrel{\mathrm{def}}{=} (\alpha_j\mu_j^2)^{-1/2}\bigl(\upsilon_j - \alpha_j\mu_j\bigr)$$

has zero mean and unit variance. We can use the Taylor expansion

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

Now let's take into account properties of real data distribution.

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

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 p_n sufficient large, $\alpha_j \approx M_n/p_n$. Moreover, it holds for p_n sufficiently large that $\max_{j=1,\dots,p_n} \alpha_j^{-1/2}|\gamma_j| \le 1/2$ with a high probability. Below we can restrict ourselves to the case when $\alpha_j^{-1/2}|\gamma_j| \le 1/2$. This allows to use the Taylor expansion

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

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 $p_n^{-1} \sum_{j=1}^{p_n} \gamma_j^2 \stackrel{I\!P}{\longrightarrow} 1$. CLT here can be easily checked because of the Lyapunov condition being valid. Also $\sum_{j=1}^{p_n}(\alpha_j - 1)^{-1} = \frac{p_n^2}{M_n} + o_n(\beta_n^2)$. Now check what happens if $\beta_n \to 0$

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

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

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

This proves the result for $\beta_n \equiv \beta$. Finally in the case when β_n grows to infinity, but $\beta_n^3/\sqrt{p_n} \to 0$, then $\beta_n^{-1}(\theta - \widetilde{\theta}_n) \xrightarrow{I\!P} \infty$. □

5.9 Proof of Lemma 5.4

Proof. Let $\overline{u}_j = u_j - u_j^*$. Then

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

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

$$I\!E\mathcal{L}(\boldsymbol{u},\boldsymbol{u}^*) = \frac{M_n}{p_n}\sum_{j=1}^{p_n}\bigl(\overline{u}_j - (\mathrm{e}^{\overline{u}_j} - 1)\bigr) = -\frac{M_n}{p_n}\sum_{j=1}^{p_n}\frac{\overline{u}_j^2}{2} + O\bigl(\|\overline{\boldsymbol{u}}\|^3\bigr).$$

Then we substitute $\overline{u}_1 = p_n\overline{\theta} - \sum_{j=2}^{p_n}\overline{u}_j$, where $\overline{\theta} = \theta - \theta^*$. Thus we get

$$I\!E\mathcal{L}(\boldsymbol{u},\boldsymbol{u}^*) = -\frac{M_n}{p_n}\frac{1}{2}\bigl(p_n\overline{\theta} - \sum_{j=2}^{p_n}\overline{u}_j\bigr)^2 - \frac{M_n}{p_n}\sum_{j=2}^{p_n}\frac{\overline{u}_j^2}{2} + O\bigl(\|\overline{\boldsymbol{u}}\|^3\bigr).$$

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

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

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

$$\breve{D}_0^2 = M_n p_n\bigl(1 - \boldsymbol{e}^\top Q^{-1}\boldsymbol{e}\bigr),$$

where $\boldsymbol{e} = (1, \ldots, 1)^\top$ and $Q = \boldsymbol{I} + E$ with $E = \boldsymbol{e}\boldsymbol{e}^\top$ being the matrix of ones of size $(p_n - 1) \times (p_n - 1)$. It follows

$$\boldsymbol{e}^\top Q^{-1} \boldsymbol{e} = \operatorname{tr}\bigl(\boldsymbol{e}^\top Q^{-1} \boldsymbol{e}\bigr) = \operatorname{tr}\bigl(Q^{-1} \boldsymbol{e}\boldsymbol{e}^\top\bigr) = \operatorname{tr}\bigl((E + \boldsymbol{I})^{-1} E\bigr).$$

Further, $(E + \boldsymbol{I})^{-1} E = \boldsymbol{I} - (E + \boldsymbol{I})^{-1}$ yielding

$$\boldsymbol{e}^\top Q^{-1} \boldsymbol{e} = \operatorname{tr}\bigl\{\boldsymbol{I} - (E + \boldsymbol{I})^{-1}\bigr\} = (p_n - 1) - \operatorname{tr}\bigl\{(E + \boldsymbol{I})^{-1}\bigr\} = (p_n - 1) - \sum_{j=1}^{p_n} \lambda_j,$$

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

$$\boldsymbol{e}^\top Q^{-1} \boldsymbol{e} = (p_n - 1) - \bigl\{p_n^{-1} + (p_n - 2)\bigr\} = 1 - p_n^{-1},$$
$$\breve{D}_0^2 = M_n p_n \bigl(1 - \boldsymbol{e}^\top Q^{-1} \boldsymbol{e}\bigr) = M_n p_n \bigl\{1 - (1 - p_n^{-1})\bigr\} = M_n = p_n^2 \beta_n^{-2},$$

which completes the proof. □

References

Barron, A., Schervish, M. J., and Wasserman, L. (1996). The consistency of posterior distributions in nonparametric problems. *Ann. Statist*, 27:536–561.

Bickel, P. J. and Kleijn, B. J. K. (2012). The semiparametric Bernstein-von Mises theorem. *Ann. Statist.*, 40(1):206–237.

Boucheron, S. and Gassiat, E. (2009). A Bernstein-von Mises theorem for discrete probability distributions. *Electron. J. Stat.*, 3:114–148.

Boucheron, S. and Massart, P. (2011). A high-dimensional Wilks phenomenon. *Probability Theory and Related Fields*, 150:405–433. 10.1007/s00440-010-0278-7.

Castillo, I. (2012). A semiparametric Bernstein - von Mises theorem for Gaussian process priors. *Probability Theory and Related Fields*, 152:53–99. 10.1007/s00440-010-0316-5.

Castillo, I. and Rousseau, J. (2013). A general bernstein–von mises theorem in semiparametric models. Available at arXiv:1305.4482 [math.ST].

Cheng, G. and Kosorok, M. (2008). General frequentist properties of the posterior profile distribution. *Ann. Stat.*, 36(4):1819–1853.

Chernozhukov, V. and Hong, H. (2003). An mcmc approach to classical estimation. *Journal of Econometrics*, 115(2):293–346.

Cox, D. D. (1993). An analysis of Bayesian inference for nonparametric regression. *Ann. Stat.*, 21(2):903–923.

Freedman, D. (1999). On the Bernstein-von Mises theorem with infinite-dimensional parameters. *Ann. Stat.*, 27(4):1119–1140.

Ghosal, S. (1999). Asymptotic normality of posterior distributions in high-dimensional linear models. *Bernoulli*, 5(2):315–331.

Ghosal, S. (2000). Asymptotic normality of posterior distributions for exponential families when the number of parameters tends to infinity. *J. Multivariate Anal.*, 74(1):49–68.

Johnstone, I. M. (2010). High dimensional Bernstein–von Mises: simple examples. In *Borrowing strength: theory powering applications—a Festschrift for Lawrence D. Brown*, volume 6 of *Inst. Math. Stat. Collect.*, pages 87–98. Inst. Math. Statist., Beachwood, OH.

Kim, Y. (2006). The bernstein - von mises theorem for the proportional hazard model. *Ann. Statist*, 34(4):1678–1700.

Kim, Y. and Lee, J. (2004). A bernstein - von mises theorem in the nonparametric right-censoring model. *Ann. Statist*, 32(4):1492–1512.

Kleijn, B. J. K. and van der Vaart, A. W. (2006). Misspecification in infinite-dimensional Bayesian statistics. *Ann. Statist.*, 34(2):837–877.

Kleijn, B. J. K. and van der Vaart, A. W. (2012). The Bernstein-von-Mises theorem under misspecification. *Electronic J. Statist.*, 6:354–381.

Le Cam, L. and Yang, G. L. (1990). *Asymptotics in Statistics: Some Basic Concepts.* Springer in Statistics.

Polzehl, J. and Spokoiny, V. (2006). Propagation-separation approach for local likelihood estimation. *Probab. Theory Relat. Fields*, 135(3):335–362.

Rivoirard, V. and Rousseau, J. (2012). Bernsteinvon mises theorem for linear functionals of the density. *Ann. Statist.*, 40(3):1489–1523.

Schwartz, L. (1965). On Bayes procedures. *Probability Theory and Related Fields*, 4(1):10–26.

Shen, X. (2002). Asymptotic normality of semiparametric and nonparametric posterior distributions. *J. Am Stat. Assoc.*, 97(457):222–235.

Spokoiny, V. (2012). Parametric estimation. Finite sample theory. *Ann. Statist.*, 40(6):2877–2909. arXiv:1111.3029.

van der Vaart, A. W. (2000). *Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press.