Journal of Machine Learning Research 26 (2025) 1-81 Submitted 4/24; Revised 2/25; Published 5/25

How good is your Laplace approximation of the Bayesian posterior? Finite-sample computable error bounds for a variety of useful divergences

Mikołaj J. Kasprzak KASPRZAK@ESSEC.EDU
Department of Information Systems, Data Analytics and Operations
ESSEC Business School
3 Avenue Bernard Hirsch
95021 Cergy-Pontoise Cedex, France

Ryan Giordano RGIORDANO@BERKELEY.EDU
Department of Statistics
University of California, Berkeley
367 Evans Hall
Berkeley, CA 94720-3860, USA

Tamara Broderick TAMARAB@MIT.EDU
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
32 Vassar Street, 32-D608
Cambridge, MA 02139, USA

Editor: Matthew Hoffman

Abstract

The Laplace approximation is a popular method for constructing a Gaussian approximation to the Bayesian posterior and thereby approximating the posterior mean and variance. But approximation quality is a concern. One might consider using rate-of-convergence bounds from certain versions of the Bayesian Central Limit Theorem (BCLT) to provide quality guarantees. But existing bounds require assumptions that are unrealistic even for relatively simple real-life Bayesian analyses; more specifically, existing bounds either (1) require knowing the true data-generating parameter, (2) are asymptotic in the number of samples, (3) do not control the Bayesian posterior mean, or (4) require strongly log concave models to compute. In this work, we provide the first computable bounds on quality that simultaneously (1) do not require knowing the true parameter, (2) apply to finite samples, (3) control posterior means and variances, and (4) apply generally to models that satisfy the conditions of the asymptotic BCLT. Moreover, we substantially improve the dimension dependence of existing bounds; in fact, we achieve the lowest-order dimension dependence possible in the general case. We compute exact constants in our bounds for a variety of standard models, including logistic regression, and numerically demonstrate their utility. We provide a framework for analysis of more complex models.

Keywords: Bayesian inference, Laplace approximation, Bernstein–von Mises theorem, approximate inference, log-Sobolev inequality

1. Introduction

1.1 Motivation

Bayesian inference is widely used in modern statistical practice to provide both point estimates and uncertainties of unknown quantities; Abbott et al. (2023); Flaxman et al. (2020); Freedman (2021); Jones et al. (2021) give just a few especially influential recent examples. In applications, practitioners typically report posterior means and (co)variances. These expectations under the posterior distribution are, however, often intractable to compute, so practitioners must use approximations. The *Laplace approximation* is popular in many communities (e.g., Bishop 2006, Section 4.4; Murphy 2022, Section 4.6.8.2; Barber et al. 2016; Thygesen et al. 2017; Gomez-Rubio et al. 2021; Ritter et al. 2018; Riihimäki and Vehtari 2014; Long et al. 2013; Wang et al. 2018) in large part due to its ease of use and computational speed. It approximates the Bayesian posterior by a suitably chosen Gaussian distribution and is grounded in the celebrated *Bernstein–von Mises theorem*. The Bernstein–von Mises theorem is often colloquially referred to as the Bayesian Central Limit Theorem, or BCLT; see e.g. the recent work of Goehle (2024). Studies have shown the appealing empirical performance of the Laplace approximation, for instance in the context of Bayesian neural networks (Daxberger et al., 2021).

However, approximation quality remains a concern. If users could check the quality of their approximation directly, with high confidence, they could better decide whether to trust any conclusions based on the approximation. And if the approximation could not be certified, the user might invest further computational power into a more expensive approximation. Notably, the BCLT on its own cannot be expected to provide such a check. For instance, convergence in the BCLT is normally expressed in terms of the total variation distance, and indeed the Laplace approximation is typically justified by the fact that the total variation distance between the rescaled posterior and the Gaussian vanishes in the limit. The total variation distance, however, does not control the difference of means or the difference of covariances in general. But it is exactly posterior means and (co)variances that are most often reported by users of approximate Bayesian inference. We therefore need finite-sample guarantees on the quality of Laplace approximation, expressed in terms of metrics that control the error in mean and covariance approximation.

1.2 Our contribution

Our work provides an important step forward toward achieving fully practical and ready-to-use quality guarantees. Namely, our work provides the first bounds for the Laplace approximation that simultaneously

- do not require knowing the true parameter in advance,
- apply to finite data sets (i.e., not just in the limit of infinite data),
- control the difference in means and the difference in covariances,
- do not require the posterior to be log concave.

We achieve control over difference in means by controlling the 1-Wasserstein distance, and we achieve control over the difference in covariances via another integral probability metric.

We are also able to control the total variance distance. Moreover, our bounds are closed form.

More specifically, let n be the sample size and d be the dimension. Let $\bar{\theta}_n$ denote the maximum a posteriori (MAP) estimate. Let $\tilde{\boldsymbol{\theta}}_n$ denote a random variable distributed according to the posterior. Let $\bar{J}_n(\bar{\theta}_n)$ denote the *posterior empirical Fisher information at the MAP*. Let $\nu^\pi := \mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n\right)\right)$ (where $\mathcal{L}$ denotes the law) and $\nu^{\mathcal{N}} := \mathcal{N}(0, \bar{J}_n(\bar{\theta}_n)^{-1})$. Our main bounds, presented in full rigor in Section 3, may be (very informally and in a very simplified way) described as taking the following form:

$$TV\left(\nu^\pi, \nu^{\mathcal{N}}\right) \leq H_d d n^{-1/2} + \exp\left(-\overline{H}_d\left(\sqrt{n} - \sqrt{d}\right)^2\right) + \hat{H}_d n^{d/2} e^{-n\bar{\kappa}},$$

$$W_1\left(\nu^\pi, \nu^{\mathcal{N}}\right) \leq H_d d n^{-1/2} + \sqrt{n}\exp\left(-\overline{H}_d\left(\sqrt{n} - \sqrt{d}\right)^2\right) + \hat{H}_d n^{d/2+1/2} e^{-n\bar{\kappa}},$$

$$D\left(\nu^\pi, \nu^{\mathcal{N}}\right) \leq H_d d^{3/2} n^{-1/2} + n\exp\left(-\overline{H}_d\left(\sqrt{n} - \sqrt{d}\right)^2\right) + \hat{H}_d n^{d/2+1} e^{-n\bar{\kappa}},$$

for appropriate model-dependent factors $H_d, \overline{H}_d, \hat{H}_d, \bar{\kappa}$. Here TV is the total variation distance, W_1 is the 1-Wasserstein distance, and D is an appropriate distance that controls the difference of covariances. All the constants and factors in our bounds are explicit and presented in full detail in Section 3.

The sample-size and dimension dependence of our bounds on the total-variation and the 1-Wasserstein distance are tight and such that they *cannot be improved in general*, as we discuss in Section 4.2. Our bounds' dimension and sample-size dependence are also *strictly better* than those appearing in all the previous works on the Laplace approximation, despite our results holding under weaker assumptions than those of the previous works.

We highlight that our bounds do not require access to the true parameter or exact integrals with respect to the posterior, which would be unrealistic for practical checks. Rather, our bounds are expressed in terms of the data and work under any distribution of the data; importantly, our bounds apply when the model is misspecified, as we expect to be the case in practice. Our results are fully applicable to models involving generalized likelihoods and the resulting generalized posteriors; see Bissiri et al. (2016); Miller (2021); Chernozhukov and Hong (2003). Our assumptions on the generalized likelihood and the prior are standard and no stronger than the assumptions of the classical proofs of the Bernstein–von Mises Theorem; see e.g. Ghosh and Ramamoorthi (2003, Section 1.4) or Miller (2021) for details. We compute our bounds explicitly for a variety of Bayesian models, including logistic regression with a Student's t prior.

Our contribution lies also in our proof techniques. In order to control the discrepancy inside a ball around the maximum likelihood estimator (MLE) or the maximum a posteriori (MAP) estimator, we use the log-Sobolev inequality or Stein's method. In order to control the discrepancy over the rest of the parameter space, we carefully bound the tail growth using standard assumptions of the Bernstein–von Mises Theorem.

A number of important directions remain for future work. Though beyond the scope of the present paper, we believe that our approach to proving computable non-asymptotic bounds could be extended so as to cover more general statistical models satisfying the conditions of the local asymptotic normality (LAN) theory (Ibragimov and Has'minskii,

Table 1: Very brief summary of related results. See section 1.3 for full details.

	Present work	PS 15	HK 22	H+ 18	D 19	S 22	YK 20
True parameter not needed	✓	–	✓	✓	✓	✓	–
Global log-concavity not needed	✓	✓	–	–	–	–	✓
Generic priors / likelihoods	✓	–	–	✓	✓	–	–
Explicit bounds (no order notation)	✓	✓	✓	✓	–	✓	✓
Controls TV distance	✓	✓	✓	–	✓	✓	✓
Controls means and variances	✓	✓	–	✓	–	–	–

1981, Chapters 1–3). We also note that, while our experiments demonstrate that we can use our bounds to make non-vacuous conclusions about real-data analyses, exact computation of our bounds requires analytical derivations and a series of numerical optimization procedures. Their current instantiations can become onerous as models increase in complexity and size. We believe our current work, though, points to fruitful directions to continue building these tools.

1.3 Related work

A number of recent papers have studied the non-asymptotic properties of the Laplace approximation and the Bernstein–von Mises theorem for Bayesian posteriors. We next contrast the present work with these analyses in terms of the computability of the bounds, strength of the results, and restrictiveness of the assumptions. A brief summary of the differences can be found in Table 1.

Recently, Huggins et al. (2018, Section 6.1), Dehaene (2019), and Spokoiny (2022) have offered ways of obtaining guarantees on the quality of Laplace approximation under log-concavity of the posterior. In Proposition 6.1, Huggins et al. (2018) assume that the posterior is strongly log-concave and obtain a computable bound on the 1- and 2-Wasserstein distances between the posterior and the approximating Gaussian. In Proposition 6.2, Huggins et al. (2018) relax the assumption on the posterior to weak log-concavity, but the bound they obtain is not computable in practice, for finite data. Dehaene (2019) assumes weak, yet strict, log-concavity of the posterior. In the paper's main result (Theorem 5), Dehaene (2019) obtains a bound on the Kullback-Leibler (KL) divergence between the posterior and the Gaussian approximation. In this bound, only the leading terms are computable, while the higher order terms are presented using the big-O notation. Spokoiny (2022) assumes a Gaussian prior and a log-concave likelihood, which yields a strongly log-concave posterior. The author relaxes this assumption only in a very specific case of a nonlinear inverse problem with a certain "warm start condition." The bounds of Spokoiny (2022) are on the total variation distance. The author also provides a bound on the difference of means – yet this bound involves generic abstract constants and it is not clear to us whether it is computable in applications. Schillings et al. (2020) have proved a qualitative, asymptotic result about convergence of the Laplace approximation for inverse problems. Motivated by this work, Helin and Kretschmann (2022) have provided non-asymptotic bounds for the

Laplace approximation on the total variation distance – in the context of Bayesian inverse problems and only under the assumption that the posterior distribution is sub-Gaussian. Moreover, in concurrent work, Fischer et al. (2022) have provided bounds on the Gaussian approximation of the posterior, yet only for i.i.d. data coming from a regular k-parameter exponential family. The dimension dependence of the bounds of Helin and Kretschmann (2022); Spokoiny (2022); Dehaene (2019); Fischer et al. (2022) is worse than ours, as we describe in Section 4.2. Additionally, the recent paper by Hasenpflug et al. (2022) proved interesting *asymptotic* convergence results for Bayesian posteriors in setups in which the MAP estimator may not be not unique. We also mention two very recent pieces of work (Katsevich, 2023a,b), which appeared after the first version of the present article was put on ArXiv. Katsevich (2023a) studies the leading order contribution to the total variation distance between the posterior and the Laplace approximation and thus gives *necessary* conditions under which the Laplace approximation is valid. For the sufficient conditions, Katsevich (2023a) refers the reader to the ArXiv version of the present article, in particular to our Theorems 17 and 18. Katsevich (2023b) proposes a skew adjustment to the Laplace approximation so as to improve its quality in high dimensions.

Non-asymptotic analyses of the Bernstein–von Mises (BvM) Theorem (discussed in Appendix B) have previously been performed by Panov and Spokoiny (2015); Yano and Kato (2020). The aim and focus of those analyses are however significantly different from those of the present paper, as we concentrate on the quality of the Laplace approximation (of the form of Equation (B.4), as discussed in Appendix B). Panov and Spokoiny (2015) consider semiparametric inference and prove results whose purpose it is to provide insight into the *critical dimension* of the parameter for which the BvM Theorem holds. They prove their bounds only for the non-informative and the Gaussian prior. Computing their bounds requires knowledge of the true parameter, which is the object of inference. Additionally, Yano and Kato (2020) derive a Berry–Esseen-type bound (which depends on the true parameter value) on the total variation distance between the posterior and the approximating normal in the approximately linear regression model.

In contrast to the references mentioned above, our bounds hold and may be computed without access to the true parameter, for general posteriors satisfying assumptions analogous to the classical assumptions of the Bernstein–von Mises theorem (see e.g. Miller 2021, Section 4 for a recent reference or Ghosh and Ramamoorthi 2003, Section 1.4 for a more classical one). In particular, we do not require (weak or strong) log-concavity or sub-Gaussianity of the posterior. Indeed, we compute our bounds explicitly for examples of commonly used non-log-concave and heavy-tailed posteriors in Section 7. The reason we can avoid imposing the assumption of log-concavity is that all we need to control the tail behaviour of the posterior is the assumption of the strict optimality of the MLE or MAP (Assumption 9 or Assumption 6). This requirement is much weaker than assuming log-concavity or strong unimodality of the posterior. The array of priors our bounds are available for is also much wider than just the Gaussian family. All we require is differentiability, boundedness, and boundedness away from zero of the prior density in a small neighbourhood around the MAP or the MLE. We do not make any assumption about the true distribution of the data, and we cover generalized likelihoods not coming from any particular family. Moreover, we bound a variety of divergences including those that control means and variances.

Finally, it is worth noting that bounds similar to ours are not widely available for approximate Bayesian inference techniques in general. Indeed, they are not available for variational inference (VI) methods (Blei et al., 2017; Wainwright and Jordan, 2008), except for the recently studied case of Gaussian VI (Katsevich and Rigollet, 2024). In the area of Markov Chain Monte Carlo methods, progress has recently been made on deriving convergence guarantees for the Unadjusted Langevin Algorithm under different sets of assumptions on the tail growth of the target distribution (Chewi et al., 2022; Balasubramanian et al., 2022; Erdogdu et al., 2022). However, the popular Metropolis-adjusted Langevin Algorithm and Hamiltonian Monte Carlo are not equipped with such guarantees beyond the case of log-concave targets. Flexible and computable post hoc checks measuring a discrepancy between the empirical distribution of a sample and the target distribution are given by graph and kernel Stein discrepancies. Graph Stein discrepancies (Gorham and Mackey, 2015; Gorham et al., 2019) metrize weak convergence and control the difference of means for distantly dissipative targets. They are however not known to control the difference of variances. Graph diffusion Stein discrepancies (Gorham et al., 2019) possess the same properties under a slightly weaker (yet technical) assumption of the underlying diffusion having a rapid Wasserstein decay rate. The fast and popular kernel Stein discrepancies (Chwialkowski et al., 2016; Liu et al., 2016; Oates et al., 2017; Gorham and Mackey, 2017) metrize weak convergence for certain choices of kernels and for distantly-dissipative targets. In comparison, we derive control over the rate of weak convergence and the differences of means and variances under interpretable assumptions that do not require any particular tail behavior of the target posterior.

1.4 Structure of the paper

In Section 2 we describe our setup, introduce the necessary notation, and present our assumptions. We also give examples of popular models satisfying the assumptions. In Section 3 we present and discuss our main bounds. In Section 4 we provide a comprehensive discussion of the main results and their dependence on the sample size, dimension, and data, as well as computability. Section 5 discusses our proof techniques. In Section 6 we present results analogous to those of Section 3, yet focused on different types of approximations. In Section 7 we show how to compute our bounds for the (non-log-concave) posterior in the logistic regression model with Student's t prior. We also present plots of our bounds computed numerically in this case. Moreover, we numerically compare our control over the difference of means and the difference of variances to the ground truth for certain conjugate prior models. In Section 8 we present conclusions of our work. The proofs of the results of Sections 3 and 6 are postponed to the appendices.

2. Setup, assumptions and notation

2.1 Setup

Our setup and assumptions are similar to those found for instance in Miller (2021). We fix $n \in \mathbb{N}$ and study probability measures on $\mathbb{R}^d$ having Lebesgue densities of the form:

$$\Pi_n(\theta) = e^{L_n(\theta)}\pi(\theta)/z_n, \tag{2.1}$$

where $\pi : \mathbb{R}^d \to \mathbb{R}$ is a Lebesgue probability density function, $L_n : \mathbb{R}^d \to \mathbb{R}$ and $z_n \in \mathbb{R}_+$ is the normalizing constant. Throughout the paper, we call π *the prior density* (or simply *the prior*), L_n *the generalized log-likelihood*, and Π_n *the generalized posterior*. By

$$\overline{L}_n(\theta) := \log\left(\Pi_n(\theta)\right)$$

we denote the *generalized log-posterior* and let:

$$\hat{\theta}_n := \arg\max_{\theta\in\mathbb{R}^d} L_n(\theta), \qquad \bar{\theta}_n := \arg\max_{\theta\in\mathbb{R}^d} \overline{L}_n(\theta),$$

whenever those quantities exist. If those quantities are unique, we call $\hat{\theta}_n$ the *maximum likelihood estimator* (*MLE*) and $\bar{\theta}_n$ the *maximum a posteriori* (*MAP*) estimator. Here and elsewhere, overbars will refer to quantities derived from the MAP ($\bar{J}_n(\bar{\theta}_n)$, $\bar{\delta}$, $\overline{M}_2$, etc., introduced below) and hats will refer to quantities derived from the MLE (for instance $\hat{J}_n(\hat{\theta}_n)$, introduced below). For any twice-differentiable function $f : \mathbb{R}^d \to \mathbb{R}$, we let f' stand for its gradient and f'' for its Hessian. We shall write

$$\hat{J}_n(\theta) = -\frac{L_n''(\theta)}{n}, \qquad \bar{J}_n(\theta) = -\frac{\overline{L}_n''(\theta)}{n},$$

whenever those expressions make sense (i.e. when the Hessians exist). For any three times differentiable function $f : \mathbb{R}^d \to \mathbb{R}$, we will also write f''' for its third (Fréchet) derivative, defined as the following multilinear 3-form on $\mathbb{R}^d$:

$$f'''(\theta)[u, v, w] = \sum_{i,j,k=1}^{d} \frac{\partial f}{\partial\theta_i\partial\theta_j\partial\theta_k}(\theta) u_i v_j w_k.$$

The norm $\|\cdot\|^*$ of this third derivative will be defined in the following way:

$$\|f'''(\theta)\|^* := \sup_{\|u\|\le 1, \|v\|\le 1, \|w\|\le 1} \left| f'''(\theta)[u, v, w] \right|.$$

Throughout the paper, $\|\cdot\|$ denotes the Euclidean norm. We will also let $\lambda_{\min}(\hat{\theta}_n)$ be the minimal eigenvalue of $\hat{J}_n(\hat{\theta}_n)$ and $\overline{\lambda}_{\min}(\bar{\theta}_n)$ be the minimal eigenvalue of $\bar{J}_n(\bar{\theta}_n)$. Moreover, throughout the paper $\|\cdot\|_{op}$ will denote the operator (i.e. spectral) norm, $\langle\cdot,\cdot\rangle$ will be the Euclidean inner product and $\tilde{\boldsymbol{\theta}}_n$ will always denote a random variable distributed according to the generalized posterior measure with density given by Eq. (2.1). $\mathcal{N}(\mu, \Sigma)$ will denote the normal distribution with mean μ and covariance Σ, and the function $\mathcal{L}(\cdot)$ will return the law of its argument. $I_{d\times d}$ will always denote the d-dimensional identity matrix.

Our bounds will be derived for two types of approximations. The first type is what we call the *MAP-centric approach*. Within this approach, $\mathcal{L}(\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n))$ is approximated by $\mathcal{N}(0, \bar{J}_n(\bar{\theta}_n)^{-1})$. On the other hand, what we call the *MLE-centric approach* is the approximation of $\mathcal{L}(\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n))$ by $\mathcal{N}(0, \hat{J}_n(\hat{\theta}_n)^{-1})$. The *MLE-centric* approach is similar to the classical Bernstein–von Mises Theorem (see Appendix B for details). The approximation it yields is a Gaussian distribution whose parameters depend only on the likelihood while ignoring the prior completely. The *MAP-centric* approach is arguably more popular and standard among practitioners using the Laplace approximation. The approximating

Gaussian's parameters depend on the posterior, thus depending on the prior as well. Thus, this approximation may be expected to be more accurate for many commonly used models. We derive our results for both approaches in order to make them applicable in a variety of circumstances. Indeed, our bounds can be used by practitioners who are interested in assessing the applicability of one of those approximation approaches. At the same time, we believe they are also interesting for researchers who study theoretical aspects of Bayesian Central Limit Theorems.

The bounds we obtain are on the following distances:

1. The Total Variation (TV) distance, which, for two probability measures ν_1 and ν_2 on a measurable space $(\Omega, \mathcal{F})$ is defined by
$$TV(\nu_1, \nu_2) := \sup_{A \in \mathcal{F}} |\nu_1(A) - \nu_2(A)| .$$
2. The 1-Wasserstein distance, which, for probability measures ν_1 and ν_2 and the set $\Gamma(\nu_1, \nu_2)$ of all couplings between them, is defined by
$$W_1(\nu_1, \nu_2) := \inf_{\gamma \in \Gamma(\nu_1, \nu_2)} \int \|x - y\| d\gamma(x, y).$$
Kantorovich duality (see, e.g. Villani 2009, Theorem 5.10) provides an equivalent definition. Let $\|\cdot\|_L$ return the Lipschitz constant of the input. Then
$$W_1(\nu_1, \nu_2) = \sup_{\substack{f \text{ Lipschitz:} \\ \|f\|_L = 1}} |\mathbb{E}_{\nu_1} f - \mathbb{E}_{\nu_2} f| .$$
3. The following integral probability metric, which for $Y_1 \sim \nu_1$ and $Y_2 \sim \nu_2$ is defined by
$$\sup_{v : \|v\| \leq 1} \left| \mathbb{E} \langle v, Y_1 \rangle^2 - \mathbb{E} \langle v, Y_2 \rangle^2 \right| .$$
For zero-mean ν_1 and ν_2, this integral probability metric controls the operator norm of the difference of the covariance matrices of ν_1 and ν_2. For more general ν_1 and ν_2, it may be combined with the 1-Wasserstein distance in order to provide control over the operator norm of the difference between the covariance matrices of ν_1 and ν_2. Our proof techniques allow us to upper-bound integral probability metrics. Obtaining a bound on the above integral probability metric lets us reach our goal of controlling the difference between the covariance of the posterior and that of the Laplace approximation.

2.2 Assumptions made throughout the paper

Now we list the assumptions that we will need to prove our finite-sample bounds and define constants used therein. We will first present those assumptions that will stand for both approaches described above and then others, which are divided between those relevant for the MAP-centric approach (Section 2.3) and the MLE-centric approach (Section 2.4). We reiterate that the conditions we require are similar to the classical assumptions of the

Bernstein–von Mises theorem, as given in Ghosh and Ramamoorthi (2003, Section 1.4) and Miller (2021, Theorem 5). The first assumption that will be used throughout the paper is the following:

Assumption 1 *There exists a unique MLE $\hat{\theta}_n$. There also exists a real number $\delta > 0$ such that the generalized log-likelihood L_n is three times differentiable inside $\{\theta : \|\theta - \hat{\theta}_n\| \le \delta\}$. For the same $\delta > 0$ there exists a real number $M_2 > 0$, such that:*

$$\sup_{\|\theta - \hat{\theta}_n\| \le \delta} \frac{\|L_n'''(\theta)\|^*}{n} \le M_2. \tag{2.2}$$

Remark 1 *Assumption 1 is needed to ensure that the posterior looks Gaussian inside the δ-neighborhood around the MLE. Indeed, Assumption 1, combined with Taylor's theorem, implies that $L_n(n^{-1/2}\theta + \hat{\theta}_n) - L_n(\hat{\theta}_n)$ can be approximated by $-\frac{1}{2}\theta^T \hat{J}_n(\hat{\theta}_n)\theta$ for θ in a neighborhood of $\hat{\theta}_n$. Note that $-\frac{1}{2}\theta^T \hat{J}_n(\hat{\theta}_n)\theta$ is the logarithm of the $\mathcal{N}(0, \hat{J}_n(\hat{\theta}_n))$ density (up to an additive constant). When proving our bounds in the MLE-centric approach, we also use Assumption 1 to prove that the posterior satisfies the log-Sobolev inequality inside this neighborhood (see Appendix F for details).*

Moreover, we make the following assumption on the prior:

Assumption 2 *For the same $\delta > 0$ as in Assumption 1, there exists a real number $\hat{M}_1 > 0$, such that*

$$\sup_{\theta : \|\theta - \hat{\theta}_n\| \le \delta} \left| \frac{1}{\pi(\theta)} \right| \le \hat{M}_1.$$

Remark 2 *Note that for Assumption 2 to be satisfied, it suffices to assume that π is continuous and positive in the δ-ball around $\hat{\theta}_n$. Assumption 2 essentially ensures that the prior puts a non-negligible amount of mass in the δ-neighbourhood of the MLE.*

2.3 Additional assumptions in the MAP-centric approach

In the MAP-centric approach we keep Assumptions 1 and 2 and additionally assume the following:

Assumption 3 *There exists a unique MAP $\bar{\theta}_n$. There also exists a real number $\bar{\delta} > 0$, such that the log-prior, $\log \pi$, is three times differentiable inside $\{\theta : \|\theta - \bar{\theta}_n\| \le \bar{\delta}\}$. Moreover, for the same $\bar{\delta}$, there exists a real number $\overline{M}_2 > 0$, such that*

$$\sup_{\theta : \|\theta - \bar{\theta}_n\| \le \bar{\delta}} \frac{\|\overline{L}_n'''(\theta)\|^*}{n} \le \overline{M}_2. \tag{2.3}$$

Remark 3 *Assumption 3 is very similar to Assumption 1. The difference is that we now consider a ball around the MAP rather than the MLE, and we require additional differentiability of the prior density inside this ball. We will use Assumption 3 in the MAP-centric approach (together with Assumption 5) to show that, inside the $\bar{\delta}$-ball around the MAP, the posterior satisfies the log-Sobolev inequality.*

In the MAP-centric approach, we use both Assumptions 1 and 3. Assumption 1 will play an important role in the process of controlling the posterior in the region $\{\theta : \|\theta - \bar\theta_n\| > \bar\delta\}$. *More specifically, we will use Assumption 1 to lower-bound the normalizing constant of the posterior after controlling it with the integral* $\int_{\|t-\hat\theta_n\|\le\delta} \pi(t) e^{L_n(t)-L_n(\hat\theta_n)} dt$. *See Appendix D.5 for more detail.*

Assumption 4 *For the same* $\delta > 0$ *as in Assumption 1 and for the same* $\bar\delta$ *as in Assumption 3,*

$$\max\left\{\|\hat\theta_n - \bar\theta_n\|, \sqrt{\frac{\operatorname{Tr}\left[\bar J_n(\bar\theta_n)^{-1}\right]}{n}}\right\} < \bar\delta \quad \text{and} \quad \sqrt{\frac{\operatorname{Tr}\left[\left(\hat J_n(\hat\theta_n) + \frac{\delta M_2}{3} I_{d\times d}\right)^{-1}\right]}{n}} < \delta.$$

Remark 4 *Assumption 4 is an assumption on the size of* n *and the choice of* $\bar\delta$ *and* δ. *Indeed, as long as the MLE and MAP converge to the same limit (which is true in the majority of commonly used modelling setups), we expect* $\max\left\{\|\hat\theta_n - \bar\theta_n\|, \sqrt{\frac{\operatorname{Tr}\left(\bar J_n(\bar\theta_n)^{-1}\right)}{n}}\right\}$ *to go to zero as* $n \to \infty$. *We similarly expect* $\sqrt{\frac{\operatorname{Tr}\left[\left(\hat J_n(\hat\theta_n)+\frac{\delta M_2}{3} I_{d\times d}\right)^{-1}\right]}{n}}$ *to go to zero as* $n \to \infty$. *Moreover* $\sqrt{\frac{\operatorname{Tr}\left[\left(\hat J_n(\hat\theta_n)+\frac{\delta M_2}{3} I_{d\times d}\right)^{-1}\right]}{n}} < \delta$ *will be satisfied if* $\sqrt{\frac{\operatorname{Tr}\left(\hat J_n(\hat\theta_n)^{-1}\right)}{n}} < \delta$, *which might be an easier condition to check. Assumption 4 allows us to use appropriate Gaussian concentration inequalities in our proofs. Moreover, the assumption* $\|\hat\theta_n - \bar\theta_n\| < \bar\delta$ *is necessary for Assumption 6 below to be satisfied.*

Assumption 5 *For the same* $\bar\delta > 0$ *and* $\overline{M}_2 > 0$ *as in Assumption 3,*

$$\overline{\lambda}_{\min}(\bar\theta_n) > \bar\delta\, \overline{M}_2. \tag{2.4}$$

Remark 5 *If* $\bar J_n(\bar\theta_n)$ *is positive definite and Assumption 1 is satisfied then one can adjust the choice of* $\bar\delta$ *so that both Eqs. (2.3) and (2.4) hold. Indeed, one can adjust the value of* $\bar\delta$ *accordingly because decreasing the value of* $\bar\delta$ *in Assumption 3 does not lead to an increase in the value of* $\overline{M}_2$. *At the same time, decreasing the value of* $\bar\delta$ *in Assumption 5, while keeping* $\overline{M}_2$ *fixed, decreases the right-hand side of Eq. (2.4).*

Assumption 5, combined with Assumption 3 and with Taylor's theorem, will allow us to prove that the posterior is strongly log-concave inside the $\bar\delta$*-neighborhood around the MAP. As a result, we will be able to show that the posterior satisfies the log-Sobolev inequality inside this neighborhood.*

Assumption 6 *For the same* $\bar\delta > 0$ *as in Assumption 3, there exists* $\bar\kappa > 0$, *such that*

$$\sup_{\theta : \|\theta - \hat\theta_n\| > \bar\delta - \|\hat\theta_n - \bar\theta_n\|} \frac{L_n(\theta) - L_n(\hat\theta_n)}{n} \le -\bar\kappa.$$

Remark 6 *Assumption 6 ensures that any local maxima of L_n achieved outside of the $\left(\bar{\delta} - \|\hat{\theta}_n - \bar{\theta}_n\|\right)$-ball around the MLE do not get arbitrarily close to the global maximum achieved at the MLE. It also ensures that the posterior puts asymptotically negligible mass outside the $\left(\bar{\delta} - \|\hat{\theta}_n - \bar{\theta}_n\|\right)$-neighborhood around the MLE. As a result, only the locally Gaussian part around the MLE remains as the sample size n grows. Note that for the vast majority of commonly used parametric models and data generating distributions, $\|\hat{\theta}_n - \bar{\theta}_n\|$, which appears in the expression for the radius of the ball, will tend to 0 as $n \to \infty$, a.s. Moreover, note that we do not strictly require that $\bar{\kappa}$ not depend on the sample size n. Under certain conditions, our bounds will converge to zero as $n \to \infty$ even if $\bar{\kappa} \xrightarrow{n\to\infty} 0$, as long as $\bar{\kappa}$ vanishes strictly slower than $\frac{d \log n}{n}$ (see Sections 4.1 and 4.2 for more discussion).*

This kind of assumption is standard in the discussion of the Bernstein–von Mises Theorem, both in classical references (Ghosh and Ramamoorthi, 2003) and in more recent ones (Miller, 2021). Nevertheless, we note that there are references in which this assumption is replaced with certain weaker probabilistic separation conditions (such as uniformly consistent tests), see e.g. van der Vaart (1998). However, it is not clear how to adapt similar conditions to our setup, in which we seek to obtain computable non-asymptotic bounds. Similarly, Miller (2021) remarked that it is already not clear how to use probabilistic separation conditions for proving the asymptotic convergence of the posterior to Gaussianity when the studied convergence is almost sure, rather than in probability.

Remark 7 *In Assumptions 2 and 4 – 6 we require the stated conditions to hold for the same δ as in Assumption 1 and for the same $\bar{\delta}$ as in Assumption 3. In practice, we may, however, verify those assumptions separately and, for each of them, find the ranges for $\delta > 0$ and $\bar{\delta} > 0$ for which it holds. We may then set the values of $\delta > 0$ and $\bar{\delta} > 0$ equal to (one of) the values of $\delta > 0$ and $\bar{\delta} > 0$ for which all the Assumptions 1 – 6 are satisfied.*

2.4 Additional assumptions in the MLE-centric approach

Besides Assumption 1 and Assumption 2, in the MLE-centric approach, we have the following assumptions:

Assumption 7 *For the same $\delta > 0$ as in Assumption 1,*

$$\sqrt{\frac{\operatorname{Tr}\left[\hat{J}_n(\hat{\theta}_n)^{-1}\right]}{n}} < \delta.$$

Remark 8 *Assumption 7 is an assumption on the size of n and the choice of δ. Indeed, for typical applications we expect $\sqrt{\frac{\operatorname{Tr}\left[\hat{J}_n(\hat{\theta}_n)^{-1}\right]}{n}}$ to go to zero as $n \to \infty$. This is a technical assumption, necessary to ensure that the Gaussian concentration inequalities we use in our proofs are valid.*

Assumption 8 *For the same $\delta > 0$ and $M_2 > 0$ as in Assumption 1,*

$$\lambda_{\min}(\hat{\theta}_n) > \delta M_2. \tag{2.5}$$

Remark 9 *Assumption 8 is the analogue of Assumption 5 for the MLE-centric approach. Combined with Assumption 1 and with Taylor's theorem, this assumption will allow us to prove that the likelihood is strongly log-concave inside the δ-neighborhood around the MLE. As a result, we will be able to show that the posterior satisfies the log-Sobolev inequality inside this neighborhood.*

Assumption 9 *For the same $\delta > 0$, as in Assumption 1, there exists $\kappa > 0$, such that*

$$\sup_{\theta:\|\theta-\hat{\theta}_n\|>\delta} \frac{L_n(\theta) - L_n(\hat{\theta}_n)}{n} \leq -\kappa.$$

Remark 10 *Assumption 9 is similar to Assumption 6 and ensures that any local maxima of L_n achieved outside of the δ-ball around the MLE do not get arbitrarily close to the global maximum achieved at the MLE. In other words, this assumption ensures that the posterior puts asymptotically negligible mass outside the δ-neighborhood around the MLE.*

Assumption 10 *For the same δ as in Assumption 1, there exist real numbers $M_1 > 0$ and $\widetilde{M}_1 > 0$, such that*

$$\sup_{\theta:\|\theta-\hat{\theta}_n\|\leq\delta} \left\| \frac{\pi'(\theta)}{\pi(\theta)} \right\| \leq M_1 \qquad \text{and} \qquad \sup_{\theta:\|\theta-\hat{\theta}_n\|\leq\delta} |\pi(\theta)| \leq \widetilde{M}_1.$$

Remark 11 *Note that for Assumption 10 to be satisfied, it suffices that π is continuously differentiable and positive inside the δ-ball around $\hat{\theta}_n$. Assumption 10 is a technical assumption that we use in the MLE-centric approach when showing that the posterior satisfies the log-Sobolev inequality inside the δ-ball around the MLE.*

Remark 12 *As in Remark 7, we note that one may first verify Assumptions 1, 2 and 7 – 10 separately and then set the value of $\delta > 0$ equal to (one of) the values of $\delta > 0$ for which all of those assumptions are satisfied.*

2.5 Additional notation

Certain quantities occur repeatedly in our bounds. By giving these quantities special symbols, we can express the bounds more compactly and readably.

First, we define a set of matrices closely related to $\hat{J}_n(\hat{\theta}_n)$ and $\bar{J}_n(\bar{\theta}_n)$.

$$\begin{aligned} \hat{J}_n^p(\hat{\theta}_n, \delta) &:= \hat{J}_n(\hat{\theta}_n) + (\delta M_2/3) I_{d\times d} & \qquad \hat{J}_n^m(\hat{\theta}_n, \delta) &:= \hat{J}_n(\hat{\theta}_n) - (\delta M_2/3) I_{d\times d} \\ \bar{J}_n^p(\bar{\theta}_n, \bar{\delta}) &:= \bar{J}_n(\bar{\theta}_n) + (\bar{\delta}\,\overline{M}_2/3) I_{d\times d} & \qquad \bar{J}_n^m(\bar{\theta}_n, \bar{\delta}) &:= \bar{J}_n(\bar{\theta}_n) - (\bar{\delta}\overline{M}_2/3) I_{d\times d}. \end{aligned}$$

The superscript p in those symbols refers to the *plus* sign appearing in the definition of the symbol and the superscript m refers to the *minus* sign. Each of these matrices is positive definite by Assumptions 5 and 8.

We analogously define the minimum eigenvalues of each matrix in the preceding display:

$$\hat{\lambda}_{\min}^p(\hat{\theta}_n, \delta) := \left[\left\| \hat{J}_n^p(\hat{\theta}_n, \delta)^{-1} \right\|_{op} \right]^{-1}; \qquad \hat{\lambda}_{\min}^m(\hat{\theta}_n, \delta) := \left[\left\| \hat{J}_n^m(\hat{\theta}_n, \delta)^{-1} \right\|_{op} \right]^{-1};$$

$$\overline{\lambda}^p_{\min}(\bar{\theta}_n, \bar{\delta}) := \left[\left\|\bar{J}^p_n(\bar{\theta}_n, \bar{\delta})^{-1}\right\|_{op}\right]^{-1}; \qquad \overline{\lambda}^m_{\min}(\bar{\theta}_n, \bar{\delta}) := \left[\left\|\bar{J}^m_n(\bar{\theta}_n, \bar{\delta})^{-1}\right\|_{op}\right]^{-1}.$$

Finally, we define a set of quantities of the following form:

$$\bar{\mathscr{D}}(n, \bar{\delta}) := \exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n} - \sqrt{\operatorname{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]}\right)^2 \overline{\lambda}_{\min}(\bar{\theta}_n)\right];$$

$$\bar{\mathscr{D}}^p(n, \bar{\delta}) := \exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n} - \sqrt{\operatorname{Tr}\left[\bar{J}^p_n(\bar{\theta}_n, \bar{\delta})^{-1}\right]}\right)^2 \overline{\lambda}^p_{\min}(\bar{\theta}_n, \bar{\delta})\right];$$

$$\hat{\mathscr{D}}(n, \delta) := \exp\left[-\frac{1}{2}\left(\delta\sqrt{n} - \sqrt{\operatorname{Tr}\left[\hat{J}_n(\hat{\theta}_n)^{-1}\right]}\right)^2 \lambda_{\min}(\hat{\theta}_n)\right];$$

$$\hat{\mathscr{D}}^p(n, \delta) := \exp\left[-\frac{1}{2}\left(\delta\sqrt{n} - \sqrt{\operatorname{Tr}\left[\hat{J}^p_n(\hat{\theta}_n, \delta)^{-1}\right]}\right)^2 \hat{\lambda}^p_{\min}(\hat{\theta}_n, \delta)\right].$$

The preceding terms are labeled with a script "D" for "decay," because they decrease to zero exponentially in n when all other quantities are held constant.

2.6 Examples of popular models satisfying the assumptions

We now present results stating that, under standard regularity conditions, and for large enough n, exponential families and generalized linear models satisfy the assumptions listed above with constants that do not depend on n.

2.6.1 Exponential families

We have the following result, whose proof can be found in Appendix C.1.

Proposition 13 (cf. Miller 2021, Theorem 12) *Consider an exponential family that is full, regular, nonempty, identifiable and in natural form. In particular, let this exponential family have density $q(y|\eta) = \exp\left(\eta^T s(y) - \alpha(\eta)\right)$ with respect to a sigma-finite Borel measure λ on $\mathcal{Y} \subseteq \mathbb{R}^k$, where $s : \mathcal{Y} \to \mathbb{R}^d$, $\eta \in \mathbb{R}^d$ and $\alpha(\eta) = \log \int_{\mathcal{Y}} \exp\big(\eta^T s(y)\big)\lambda(dy)$. Let $Q_\eta(E) = \int_E q(y|\eta)\lambda(dy)$ and denote $\mathbb{E}_\eta s(Y) = \int_{\mathcal{Y}} s(y) Q_\eta(dy)$. Let $\mathcal{E} := \{\eta \in \mathbb{R}^d : |\alpha(\eta)| < \infty\}$. Assume that $\mathcal{E}$ is open and nonempty, let the parameter space be given by $\Theta := \mathcal{E}$ and assume that $\eta \mapsto Q_\eta$ is one-to-one.*

Suppose $Y_1, Y_2, \dots \in \mathcal{Y}$ are i.i.d. random vectors, such that $\mathbb{E}s(Y_i) = \mathbb{E}_{\theta_0} s(Y)$ for some $\theta_0 \in \Theta := \mathcal{E}$. Let $L_n(\theta) = \sum_{i=1}^n \log q(Y_i|\theta)$. Then, almost surely, for large enough n, Assumptions 1, 7, 8 and 9 are satisfied, with constants δ, M_2, κ independent of n.

If, in addition, the prior density π does not depend on n and is continuous and positive in a neighborhood around θ_0 then, almost surely, for large enough n, Assumption 2 is satisfied with for constant $\hat{M}_1$ independent of n.

If, in addition, the prior density π is continuously differentiable in a neighborhood of θ_0 then, almost surely, Assumption 10 is satisfied for large enough n, with $M_1, \widetilde{M}_1$ independent of n.

If, in addition, the prior density π *is thrice continuously differentiable on* Θ*, then, almost surely, Assumptions 3 – 6 are satisfied for all large enough* n*, with constants* $\bar{\delta}, \overline{M}_2, \bar{\kappa}$ *independent of* n*.*

Remark 14 *Our Proposition 13 is very similar to Miller (2021, Theorem 12). The assumptions of Proposition 13 are, firstly, that the exponential family is full, regular, nonempty, identifiable and in natural form. Such conditions are standard and hold for typical commonly used exponential families (Miller and Harrison, 2014). Secondly, we assume that* $\mathbb{E}s(Y_i) = \mathbb{E}_{\theta_0} s(Y)$ *for some* θ_0*, which is a standard assumption and ensures that matching the expected sufficient statistics to the observed sufficient statistics is possible asymptotically. Note that we do not assume that the model is correctly specified.*

2.6.2 Generalized Linear Models

We have the following result, whose proof can be found in Appendix C.2.

Proposition 15 (cf. Miller 2021, Theorem 13) *Consider a regression model of the form* $p(y_i|\theta, x_i) \propto_\theta q(y_i|\theta^T x_i)$ *for covariates* $x_i \in \mathcal{X} \subseteq \mathbb{R}^d$ *and coefficients* $\theta \in \Theta \subseteq \mathbb{R}^d$*, where* $q(y|\eta) = \exp(\eta s(y) - \alpha(\eta))$ *is a one-parameter exponential family, with respect to a sigma-finite Borel measure* λ *on* $\mathcal{Y} \subseteq \mathbb{R}^d$*. Note that proportionality* $\propto_\theta$ *is with respect to* θ*, not* y_i*.*

Let $s : \mathcal{Y} \to \mathbb{R}^d$*,* $\eta \in \mathbb{R}^d$ *and* $\alpha(\eta) = \log \int_{\mathcal{Y}} \exp(\eta^T s(y))\lambda(dy)$*. Moreover, let* $Q_\eta(E) = \int_E q(y|\eta)\lambda(dy)$ *and* $\mathcal{E} := \{\eta \in \mathbb{R}^d : |\alpha(\eta)| < \infty\}$*. Assume* Θ *is open,* Θ *is convex, and* $\theta^T x \in \mathcal{E}$ *for all* $\theta \in \Theta$*,* $x \in \mathcal{X}$*. Moreover, assume* $\mathcal{E}$ *is non-empty and open and* $\eta \mapsto Q_\eta$ *is one-to-one. Suppose* $(X_1, Y_1), (X_2, Y_2), \dots \in \mathcal{X} \times \mathcal{Y}$ *are i.i.d. such that:*

1. $f'(\theta_0) = 0$ *for some* $\theta_0 \in \Theta$*, where* $f(\theta) = \mathbb{E} \log q(Y_i \,|\, \theta^T X_i)$
2. $\mathbb{E}\,|X_i s(Y_i)| < \infty$ *and* $\mathbb{E}\left|\alpha(\theta^T X_i)\right| < \infty$ *for all* $\theta \in \Theta$*,*
3. *for all* $a \in \mathbb{R}^d$*, if* $a^T X_i \overset{a.s.}{=} 0$ *then* $a = 0$
4. *There is* $\epsilon > 0$ *such that for all* $j, k, l \in \{1, \dots, d\}$*,* $\mathbb{E}\left[\sup_{\theta: \|\theta - \theta_0\| \le \epsilon} \left|\alpha'''(\theta^T X_i) X_{ij} X_{ik} X_{il}\right|\right] < \infty.$

Let $L_n(\theta) = \sum_{i=1}^n \log p(Y_i|\theta, X_i)$*. Then, almost surely, for all large enough* n*, Assumptions 1, 7, 8 and 9 are satisfied with constants* δ, M_2, κ *independent of* n*.*

If, in addition, the prior density π *does not depend on* n *and is continuous and positive in a neighborhood around* θ_0 *then, almost surely, Assumption 2 is satisfied for large enough* n *and constant* $\hat{M}_1$ *independent of* n*.*

If, in addition, the prior density π *is continuously differentiable in a neighborhood of* θ_0 *then, almost surely, Assumption 10 is satisfied for large enough* n *and* $M_1, \widetilde{M}_1$ *independent of* n*.*

If, in addition, the prior density π *is thrice continuously differentiable on* Θ*, then, almost surely, Assumptions 3 – 6 are satisfied for all large enough* n *and for constants* $\bar{\delta}, \overline{M}_2, \bar{\kappa}$ *independent of* n*.*

Remark 16 *Our Proposition 15 is very similar to Miller (2021, Theorem 13). Condition 1 in Proposition 15 says essentially that the MLE exists asymptotically. Conditions 2 and 4 are moment conditions. They will be satisfied in many situations — for instance if the covariates are bounded and* $\mathbb{E}s(Y_i)$ *exists (since* $\alpha \in C^\infty$*). Condition 3 is necessary to ensure identifiability. When* $\mathbb{E}X_iX_i^T$ *exists and is finite, condition 3 is equivalent to* $\mathbb{E}X_iX_i^T$ *being non-singular, which is often assumed to ensure identifiability for GLMs (van der Vaart, 1998, Example 16.8). Note that Proposition 15 does not assume that the model is correctly specified.*

3. Main results

In this section we present our bounds on the quality of Laplace approximation in the *MAP-centric* approach, as described in Section 2. This approach is arguably the most popular one among users of Laplace approximation (e.g. Bishop 2006, Section 4.4, Murphy 2022, Sections 4.6.8.2 and 10.5.1). The proofs of the results from this section are presented in Appendices D and E. A discussion of our bounds can be found in Section 4 below. In particular, Section 4 discusses the dependence on the sample size n and dimension d of our bounds, their dependence on the data, their computability and the way in which they control credible sets, means and variances. Our bounds in the *MLE-centric* approach will be presented in Section 6.

3.1 Control over the total variation distance

We start with a bound over the total variation distance.

Theorem 17 *Fix* $n \in \mathbb{N}$*, suppose that Assumptions 1 – 6 hold and retain the notation thereof. Let TV denote the total variation distance. Then:*

$$TV\left(\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n\right)\right), \mathcal{N}(0, \bar{J}_n(\bar{\theta}_n)^{-1})\right) \leq A_1 n^{-1/2} + 2\bar{\mathscr{D}}(n, \bar{\delta}) + A_2 n^{d/2} e^{-n\bar{\kappa}},$$

where

$$A_1 = \frac{\sqrt{3}\ \mathrm{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right] \overline{M}_2}{4\sqrt{\left(\overline{\lambda}_{\min}(\bar{\theta}_n) - \bar{\delta}\,\overline{M}_2\right)\left(1 - \bar{\mathscr{D}}(n, \bar{\delta})\right)}}; \qquad A_2 = \frac{2\left|\det\left(\hat{J}_n^p(\hat{\theta}_n, \delta)\right)\right|^{1/2} \hat{M}_1}{(2\pi)^{d/2}\left(1 - \hat{\mathscr{D}}^p(n, \delta)\right)}.$$

3.2 Control over the 1-Wasserstein distance

Now, we bound the 1-Wasserstein distance, which is known to control the difference of means.

Theorem 18 *Fix* $n \in \mathbb{N}$*, suppose that Assumptions 1 – 6 hold and retain the notation thereof. Then:*

$$\begin{aligned} &W_1\left(\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n\right)\right), \mathcal{N}(0, \bar{J}_n(\bar{\theta}_n)^{-1})\right) \\ &\qquad \leq B_1 n^{-1/2} + B_3\left(B_2 + \sqrt{n}\,B_4\right) n^{d/2} e^{-n\bar{\kappa}} + \left(\bar{\delta}\sqrt{n} + \sqrt{\frac{2\pi}{\overline{\lambda}_{\min}(\bar{\theta}_n)}} + B_2\right)\bar{\mathscr{D}}(n, \bar{\delta}), \end{aligned}$$

where

$$B_1 := \frac{\sqrt{3}\,\mathrm{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]\overline{M}_2}{2\left(\bar{\lambda}_{\min}(\bar{\theta}_n) - \bar{\delta}\overline{M}_2\right)\sqrt{1-\bar{\mathscr{D}}(n,\bar{\delta})}};$$

$$B_2 := \frac{\left|\det\left(\bar{J}_n^p(\bar{\theta}_n,\bar{\delta})\right)\right|^{1/2}\left|\det\left(\bar{J}_n^m(\bar{\theta}_n,\bar{\delta})\right)\right|^{-1/2}\sqrt{\mathrm{Tr}\left[\bar{J}_n^m(\bar{\theta}_n,\bar{\delta})^{-1}\right]}}{1-\bar{\mathscr{D}}^p(n,\bar{\delta})};$$

$$B_3 := \frac{\left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2}\hat{M}_1}{(2\pi)^{d/2}\left(1-\hat{\mathscr{D}}^p(n,\delta)\right)}; \qquad B_4 := \int_{\|u-\hat{\theta}_n\|>\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|}\|u-\bar{\theta}_n\|\pi(u)du.$$

Remark 19 *Note that, in order for the bound in Theorem 18 to be finite, we need the following integral with respect to the prior to be finite:* $\int_{\|u-\hat{\theta}_n\|>\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|}\|u-\bar{\theta}_n\|\pi(u)du.$

3.3 Control over the difference of covariances

Finally, we upper bound an integral probability metric that lets us control the difference of covariances.

Theorem 20 *Fix $n \in \mathbb{N}$, suppose that Assumptions 1 – 6 hold and retain the notation thereof. Let $\boldsymbol{Z}_n \sim \mathcal{N}\left(0, \bar{J}_n(\bar{\theta}_n)^{-1}\right)$ Then:*

$$\begin{aligned}
&\sup_{v:\|v\|\le 1}\left|\mathbb{E}\left[\left\langle v, \sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n\right)\right\rangle^2\right] - \mathbb{E}\left[\langle v, \boldsymbol{Z}_n\rangle^2\right]\right| \\
&\le C_1 n^{-1/2} + \frac{\sqrt{3\,\mathrm{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]}C_1\overline{M}_2}{4\left(\bar{\lambda}_{\min}(\bar{\theta}_n) - \bar{\delta}\,\overline{M}_2\right)}n^{-1} \\
&\quad + C_3\left(C_2 + n\,C_4\right)n^{d/2}e^{-n\bar{\kappa}} + \left(\bar{\delta}^2 n + \frac{\sqrt{2\pi}}{\bar{\lambda}_{\min}(\bar{\theta}_n)} + C_2\right)\bar{\mathscr{D}}(n,\bar{\delta}),
\end{aligned}$$

for

$$C_1 := \frac{\sqrt{3}\left(\mathrm{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]\right)^{3/2}\overline{M}_2}{\left(\bar{\lambda}_{\min}(\bar{\theta}_n) - \bar{\delta}\overline{M}_2\right)\left(1-\bar{\mathscr{D}}(n,\bar{\delta})\right)}$$

$$C_2 := \frac{\left|\det\left(\bar{J}_n^p(\bar{\theta}_n,\bar{\delta})\right)\right|^{1/2}\left|\det\left(\bar{J}_n^m(\bar{\theta}_n,\bar{\delta})\right)\right|^{-1/2}\mathrm{Tr}\left[\bar{J}_n^m(\bar{\theta}_n,\bar{\delta})^{-1}\right]}{1-\bar{\mathscr{D}}^p(n,\bar{\delta})}$$

$$C_3 := \frac{\left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2}\hat{M}_1}{(2\pi)^{d/2}\left(1-\hat{\mathscr{D}}^p(n,\delta)\right)}; \qquad C_4 := \int_{\|v-\hat{\theta}_n\|>\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|}\|u-\bar{\theta}_n\|^2\pi(u)du.$$

Remark 21 *Note that, in order for the bound in Theorem 20 to be finite, we need the following integral with respect to the prior to be finite:* $\int_{\|u-\hat{\theta}_n\|>\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|}\|u-\bar{\theta}_n\|^2\pi(u)du.$

4. Discussion of the bounds

We make some remarks about the bounds presented in Section 3 and their applicability in approximate inference.

4.1 Dependence on the sample size n

The quantities A_1, A_2, A_3, B_1, B_2, B_3, B_4, C_1, C_2, C_3, and C_4 appearing in the above bounds depend on n but, for models and data generating distributions that satisfy the assumptions of the Bernstein–von Mises theorem (as in Ghosh and Ramamoorthi 2003, Theorem 1.4.2 or Miller 2021, Theorem 5), they are bounded as n grows. In particular, they are bounded in n, as long as the constants $M_2, \overline{M}_2, \hat{M}_1$ are bounded from above and $\bar{\delta}, \delta$ are bounded from below by a positive number. In this scenario, keeping the dimension d fixed, all our bounds will vanish as $n \to \infty$ at the rate of $\frac{1}{\sqrt{n}}$, as long as asymptotically the ratio of $\frac{d+3}{2} \cdot \frac{\log n}{n}$ to $\bar{\kappa}$ goes to zero. See Appendices H.2.2, H.3.6 and H.4.4 for a discussion of the choice of $\bar{\kappa}$ in our experiments. We reiterate that our bounds are fully non-asymptotic and computable.

In the numerical examples in Section 7 below, we observed that our bounds are often coarse for small and moderate sample sizes. The main reason for this looseness at small and moderate sample sizes can be understood by looking at our assumptions. Indeed, in the cases where $\overline{\lambda}_{\min}(\bar{\theta}_n)$ is very small, Assumption 5 pushes $\bar{\delta}$ to also become very small. A small $\bar{\delta}$ means, in turn, that we need to choose a very small $\bar{\kappa}$ to make Assumption 6 satisfied. Subsequently, a very small $\bar{\kappa}$, combined with a small or moderate sample size n, results in the bound in Theorem 17 being dominated by the third summand $A_2 n^{d/2} e^{-n\bar{\kappa}}$, the bound in Theorem 18 being dominated by $B_3(B_2 + \sqrt{n} B_4) n^{d/2} e^{-n\bar{\kappa}}$ and the bound in Theorem 20 being dominated by $C_3(C_2 + nC_4) n^{d/2} e^{-n\bar{\kappa}}$. As a consequence, we need a large sample size n in order to start seeing the asymptotic convergence rate of our bounds.

Katsevich (2024, Section 2.2) notes that our bounds are not affine invariant. Indeed, passing the parameter of interest through the following transformation $\theta \mapsto \bar{J}_n(\bar{\theta}_n)^{1/2}(\theta - \bar{\theta}_n)$ and applying our analysis to the transformed parameter would change the value of our bounds. In particular, such a transformation makes the negative Hessian of the generalized log posterior, divided by n and evaluated at zero (the new mode of the posterior), equal to the identity matrix. Importantly, this means that, after the transformation, the left-hand side of Equation (2.4) becomes equal to 1, irrespective of the dimension d and one can thus have a much better control over the range of $\bar{\delta}$'s that satisfy Assumption 5. We conjecture that such a transformation thus leads in many cases to an improvement in the tightness of our bounds for smaller sample sizes n. A detailed analysis of the consequences of applying the transformation proposed in Katsevich (2024, Section 2.2) is, however, beyond the scope of the present paper. We refer the interested reader to Katsevich (2024) and, more specifically, to Section 2.2 and Appendix A.3 therein.

We hypothesized that the looseness of our bounds might be a function of the condition number of $\bar{J}_n(\bar{\theta}_n)$ (defined as the ratio of the largest eigenvalue of $\bar{J}_n(\bar{\theta}_n)$ and the smallest eigenvalue $\overline{\lambda}_{\min}(\bar{\theta}_n)$). In our experiments, we investigated this hypothesis for logistic regression with a Student's t prior. We found that while it may be the case that a smaller condition number is associated with a smaller bound, the relationship is not definitive. For instance, in Figures 6a – 6c, we can see both very large bound values and very small bound values for the same condition number; see Section 7.5 for more discussion. We leave a more detailed investigation for future work.

4.2 Dependence on the dimension d

Several terms in our bounds depend on d and the exact dependence differs depending on the analyzed model. This is not surprising; the fact that the dimension dependence of the Laplace approximation is model-dependent has already been observed by other authors, including Katsevich (2023a). Nevertheless, as observed in Katsevich (2023a), which was written after the first version of the current paper appeared on ArXiv, the dimension dependence in our total-variation and 1-Wasserstein bounds is such that it cannot be improved in general. It is also better than the dimension dependence appearing in the previous works on this topic by Helin and Kretschmann (2022); Spokoiny (2022); Dehaene (2019); Fischer et al. (2022). Below, we elaborate more on our dimension dependence and on the comparison to the dependence other authors have achieved in the past.

Let us take a closer look at our bound on the total variation distance presented in Theorem 17. Our bound consists of three summands. The first summand is of the same order as $\frac{\text{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]\overline{M}_2}{\sqrt{n}\cdot\sqrt{\bar{\lambda}_{\min}(\bar{\theta}_n)-\overline{M}_2\bar{\delta}}}$ as d and n grow. The second summand is of a lower order, provided, for instance, that $\bar{\delta} \gg \frac{\sqrt{\log n}}{\sqrt{n\,\bar{\lambda}_{\min}(\bar{\theta}_n)}}$. This condition is reasonable to expect as for models satisfying the assumptions of the Bernstein–von Mises theorem (Ghosh and Ramamoorthi, 2003, Theorem 1.4.2), $\bar{\delta}$ is required to be positive and constant in n. The third summand is of the same order as $\hat{M}_1\left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2} n^{d/2}e^{-n\bar{\kappa}}$. Therefore, if $\bar{\kappa} \gg \frac{\log n}{n}\cdot\frac{d+1}{2}+\frac{1}{n}\log\left(\hat{M}_1\left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2}\right)$ and $\bar{\delta} \gg \frac{\sqrt{\log n}}{\sqrt{n\,\bar{\lambda}_{\min}(\bar{\theta}_n)}}$, then the first summand is of the leading order as n and d grow.

Let us then look at the dimension dependence of the first summand in Theorem 17 through the lens of the recent analysis presented by Katsevich (2023a) and thus make a comparison to the previous work. Our first summand in Theorem 17 is of the same order as $C_d\sqrt{d^2/n}$, where

$$C_d = \frac{\text{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]}{d}\cdot\frac{\overline{M}_2}{\sqrt{\bar{\lambda}_{\min}(\bar{\theta}_n)-\overline{M}_2\bar{\delta}}} \leq \frac{\overline{M}_2}{\bar{\lambda}_{\min}(\bar{\theta}_n)\sqrt{\bar{\lambda}_{\min}(\bar{\theta}_n)-\overline{M}_2\bar{\delta}}},$$

where we recall from Assumption 3 that $\overline{M}_2$ controls the third derivative of the log posterior. As noted by Katsevich (2023a), all the recent works by Helin and Kretschmann (2022); Spokoiny (2022); Dehaene (2019) have obtained finite-sample bounds on the total variation distance of order $c_d\sqrt{d^3/n}$, where c_d is a ratio of the third derivative of the log posterior and the (3/2)-power of the second derivative of the log-posterior, with the definition varying slightly from paper to paper[1] (while the bounds provided by Fischer et al. 2022 are of

1. To be precise, we note that the bound appearing in Spokoiny (2022) is actually expressed in terms of the *effective dimension* d_{eff}, which may sometimes be smaller than the actual dimension d and depends on the strength of regularization by a Gaussian prior. The effective dimension d_{eff} however approaches the true dimension d as the sample size n goes to infinity, as long as the prior does not depend on n. For such a prior, the asymptotic order of the total-variation-distance bound in Spokoiny (2022) is still $c_d\sqrt{d^3/n}$.

a higher order). As described in Section 1.3, all the works of Helin and Kretschmann (2022); Spokoiny (2022); Dehaene (2019); Fischer et al. (2022) obtain their bounds under assumptions significantly stronger than ours. Still, our bound offers a tighter dimension dependence as it is of order $C_d\sqrt{d^2/n}$, where C_d again represents the ratio of a bound on the third derivative of the log posterior and the (3/2)-power of the second derivative of the log-posterior.

As noted above, Katsevich (2023a) refers specifically to our paper and states that the order of our bound cannot be improved in general. Indeed, Katsevich (2023a, Theorem V1) states that the condition $\bar{C}_d d \ll \sqrt{n}$ is necessary for accurate Laplace approximation, where $\bar{C}_d$ is a model-specific term involving ratios of derivatives of the log-posterior. We refer the interested reader to Katsevich (2023a) for a more precise statement of this result and a more thorough discussion.

In Example 1, we discuss a popular model which satisfies the assumptions of Spokoiny (2022) and for which our bound on the total variation distance is of smaller order than the one of Spokoiny (2022).

Example 1 *Let us consider the following example, inspired by Katsevich (2023a, Section 3) and discussed in more detail in Appendix H.1. Suppose that* $X_i \overset{i.i.d.}{\sim} \mathcal{N}(0, I_{d\times d})$, $\tilde{Y}_i \,|\, X_i \sim$ *Bernoulli* $\big(s(\theta_0^T X_i)\big)$ *and*

$$Y_i = \begin{cases} 1, & \textit{if } \tilde{Y}_i = 1 \\ -1, & \textit{if } \tilde{Y}_i = 0, \end{cases}$$

where s is the sigmoid $s(t) = (1+e^{-t})^{-1}$ and θ_0 is the ground truth value of the parameter. For simplicity, take $\theta_0 = (1, 0, \dots, 0)$. Now, let $\rho(t) = -\log s(t)$ and take $L_n(\theta) = -\sum_{i=1}^n \rho\left(Y_i X_i^T \theta\right)$, which corresponds to logistic regression. Consider a standard Gaussian prior on θ given by $\pi(\theta) \propto \exp\big(-\|\theta\|^2/2\big)$ and assume $d^{3/2} \leq n \leq e^{\sqrt{d}}$. This model satisfies the assumptions of Theorem 17 with high probability for sufficiently large n and the first summand in the bound in Theorem 17 is of the leading order. See Appendix H.1.3 for a discussion of those facts.

In Spokoiny (2022), bounds are proved on the quality of the Laplace approximation for models in which the likelihood is log-concave and the prior is Gaussian. The model we consider here in this example (logistic regression with a Gaussian prior) satisfies these conditions as well as the additional local smoothness conditions imposed in Spokoiny (2022). Spokoiny (2022) notes in their paper that their bounds on the total-variation distance from the Laplace approximation converge to zero only if $\sqrt{\frac{(d_{eff})^3}{n}} \to 0$. For logistic regression with a Gaussian prior, as we consider it here in this example, we have that:

$$d_{eff} := \mathrm{Tr}\left\{ \left(\bar{J}_n(\bar{\theta}_n) + \frac{(\log \pi)''(\bar{\theta}_n)}{n} \right) \bar{J}_n(\bar{\theta}_n)^{-1} \right\} \geq d \left(1 - \frac{1}{n\,\overline{\lambda}_{\min}(\bar{\theta}_n)} \right),$$

and $\overline{\lambda}_{\min}(\bar{\theta}_n)$ is lower-bounded by a positive constant not depending on n or d with high probability (see Appendix H.1.2 for more detail). Therefore $\frac{d_{eff}}{d} \xrightarrow{n\to\infty} 1$ in probability. As we show in Appendix H.1.1, our bound on the total variation distance is therefore of a smaller order than the one of Spokoiny (2022); our bound is of order $\sqrt{\frac{d^2}{n}}$ with high *probability. Moreover, we note that Katsevich (2023a, Section 3) has shown numerically that this rate is optimal for logistic regression.*

Furthermore, we note that in Theorem 18, the first summand (which will be the leading-order term in the majority of standard modeling setups) in our 1-Wasserstein bound is of the order of $\bar{C}_d\sqrt{d^2/n}$, where

$$\bar{C}_d = \frac{\operatorname{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]\overline{M}_2}{d\cdot\left(\bar{\lambda}_{\min}(\bar{\theta}_n)-\overline{M}_2\bar{\delta}\right)} \leq \frac{\overline{M}_2}{\bar{\lambda}_{\min}(\bar{\theta}_n)\left(\bar{\lambda}_{\min}(\bar{\theta}_n)-\overline{M}_2\bar{\delta}\right)}.$$

Moreover, in the bound of Theorem 20, the first summand is of the order of $\tilde{C}_d\sqrt{d^3/n}$, where

$$\tilde{C}_d \leq \frac{\overline{M}_2}{\bar{\lambda}_{\min}(\bar{\theta}_n)^{3/2}\left(\bar{\lambda}_{\min}(\bar{\theta}_n)-\overline{M}_2\bar{\delta}\right)},$$

and the second summand is of the order $\hat{C}_d\, d^2/n$, where

$$\hat{C}_d \leq \frac{\overline{M}_2}{\bar{\lambda}_{\min}(\bar{\theta}_n)^{2}\left(\bar{\lambda}_{\min}(\bar{\theta}_n)-\overline{M}_2\bar{\delta}\right)^2}$$

and those two summands form together the term that will be of the leading order in the majority of standard modelling setups.

In Section 5 we explain how our proof techniques allowed us to obtain the bounds of order $C_d\sqrt{d^2/n}$ and $\bar{C}_d\sqrt{d^2/n}$ in Theorems 17 and 18, thus achieving a sharper dimension dependence as compared to all earlier works. In Section 5 we also explain why our covariance error bound in Theorem 20 is of the order of $\tilde{C}_d\sqrt{d^3/n}$ and why we could not achieve the order of $\tilde{C}_d\sqrt{d^2/n}$ in Theorem 20 using our techniques.

4.3 Dependence on the data

In the typical applications, the bounds in Theorems 17, 18 and 20 depend on the data. For a practitioner, such dependence on data is very natural because the bound is computable using the data and therefore usable in practice. However, for a theoretical statistician, interested in Bernstein–von Mises phenomena, it might be interesting to assume the data come from a certain distribution. For instance, researchers investigating frequentist properties of Bayesian estimators may make such an assumption in order to compare the coverage of Bayesian credible sets with that of frequentist confidence sets. It might be interesting to make such an assumption for instance in order to investigate how fast the coverage of Bayesian credible sets approaches the coverage of frequentist confidence sets as $n \to \infty$. Having made such an assumption, one can use our bounds in order to quantify the speed of almost sure convergence or convergence in probability of the distances between the prior and the Gaussian. To this end, one only needs to control the speed of the relevant mode of convergence of our bounds, which, in most cases, should be achievable using standard results, similar to those quantifying the rate of convergence in the law of large numbers.

4.4 Computability

Our bounds are computable from the constants and quantities introduced in Section 2, such as $\hat{M}_1$, M_2, $\overline{M}_2$, $\bar{\delta}$, $\bar{\kappa}$, $\hat{\theta}_n$, $\bar{\theta}_n$, $\bar{J}_n(\bar{\theta}_n)$, $\hat{J}_n(\hat{\theta}_n)$. Computing some of those constants and quantities, including the bounds on the third derivatives of the log-likelihood and log-posterior given by M_2 and $\overline{M}_2$, may require additional work. It is often possible to obtain such bounds analytically. Sharper bounds can typically be obtained numerically. A robust approach to doing so is to obtain an analytical bound on the fourth derivative and then run a grid-search optimizer and apply the mean value theorem. Running a grid-search optimizer can, however, be very expensive computationally, especially in high dimensions. Another option is to run a simpler global optimizer. This is much faster but can occasionally return results that are inaccurate or incorrect. Indeed, a global numerical optimizer might output a value smaller than the maximum, rather than producing the desired upper bound. There is, therefore, a statistical–computational trade-off that needs to be taken into consideration when calculating our bounds in practice, on real data sets. Moreover, there is typically a range of values for δ and $\bar{\delta}$ for which the assumptions of Theorems 17, 18 and 20 are satisfied. The user may find the optimal choice of δ and $\bar{\delta}$ within the appropriate ranges by running a numerical optimizer on the bounds. We describe how we compute our bounds in practice in Section 7.3.1 and Appendices H.2 and H.3. We present our bounds computed for several example Bayesian models in Section 7.

4.5 Our bounds control the quality of credible-set approximations

Our bound on the total variation distance provides quality guarantees on the approximate computation of posterior credible sets. Indeed, suppose, for instance, that one is interested in finding a value b_α, such that $\mathbb{P}\left(\|\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n\| \leq \frac{b_\alpha}{\sqrt{n}}\right) \geq 1-\alpha$, for a fixed value α, where the probability $\mathbb{P}$ is *conditional on the observed data*. Let $A(n)$ denote the value of our upper bound in Theorem 17. If n is sufficiently large and $A(n)$ is smaller that α, then one could choose $b_\alpha = \tilde{b}_\alpha$, such that, for $\boldsymbol{Z}_n \sim \mathcal{N}(0, \bar{J}_n(\bar{\theta}_n)^{-1})$, $\mathbb{P}\left(\|\boldsymbol{Z}_n\| \leq \tilde{b}_\alpha\right) = 1-\alpha+A(n)$. Our bound implies that:

$$\left|\mathbb{P}\left(\|\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n\| \leq \frac{\tilde{b}_\alpha}{\sqrt{n}}\right) - \mathbb{P}\left(\|\boldsymbol{Z}_n\| \leq \tilde{b}_\alpha\right)\right| \leq A(n)$$

and so $\mathbb{P}\left(\|\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n\| \leq \frac{\tilde{b}_\alpha}{\sqrt{n}}\right) \geq 1-\alpha$.

4.6 Our bounds control the difference of means

Our bound on the 1-Wasserstein distance controls the difference of means in the Laplace approximation, in the following way. The upper bound in Theorem 18 controls $\sqrt{n}\|\mathbb{E}[\tilde{\boldsymbol{\theta}}_n] - \bar{\theta}_n\|$. In order to obtain an upper bound on $\|\mathbb{E}[\tilde{\boldsymbol{\theta}}_n] - \bar{\theta}_n\|$, one needs to divide our bound from Theorem 18 by $\sqrt{n}$.

4.7 Our bounds control the difference of covariances

Theorem 20 together with Theorem 18 let us control the difference of covariances. Suppose, for instance, that we are interested in the operator norm of the difference of the posterior covariance matrix and the covariance matrix of the Laplace approximation. Let $B(n)$ denote the value of our bound from Theorem 18 and let $C(n)$ be the value of our bound from

Theorem 20. Then, a straightforward calculation reveals that:

$$\left\| \mathrm{Cov}(\tilde{\boldsymbol{\theta}}_n) - \frac{\bar{J}_n(\bar{\theta}_n)^{-1}}{n} \right\|_{op} \leq \frac{1}{n} \left(B(n)^2 + C(n) \right).$$

5. Our proof techniques

Now, we briefly describe our proof strategies. We discuss how we prove our theory for the MAP-centric approach (in Theorems 17, 18 and 20 above), the MLE-centric approach (in Theorems 22, 23 and 25 below), and additional theory for general test functions in one dimension (in Theorem 27 in the appendix).

All the proofs of our results are provided in full detail in Appendices D – G.

5.1 The MAP-Centric Approach

In our proofs for the MAP-centric approach, we compare the distribution of $\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n\right)$ to $\mathcal{N}\left(0, \bar{J}_n(\bar{\theta}_n)^{-1}\right)$ by concentrating separately on the region $\{\theta : \|\theta\| \leq \bar{\delta}\sqrt{n}\}$ and the region $\{\theta : \|\theta\| > \bar{\delta}\sqrt{n}\}$.

In order to compare the rescaled posterior with the Gaussian in the outer region $\{\theta : \|\theta\| > \sqrt{n}\bar{\delta}\}$, we simply upper bound tail integrals with respect to the posterior and with respect to the Gaussian. We use Assumption 6, together with Gaussian concentration inequalities, which let us upper-bound integrals with respect to the posterior with integrals with respect to the prior, multiplied by $n^{d/2}e^{-n\bar{\kappa}}$ and suitable constants. Integrals with respect to the Gaussian are upper-bounded using Gaussian concentration inequalities.

With the tail integrals controlled, we can focus on the two distributions truncated to the region $\{\theta : \|\theta\| \leq \sqrt{n}\bar{\delta}\}$. As we will describe in more detail shortly, we prove our main results, Theorems 17, 18 and 20 by (1) controlling the Fisher divergence using a Taylor series expansion, (2) controlling the Kullback-Leibler (KL) divergence using strong log-concavity in the inner region together with the log-Sobolev inequality, and (3) controlling the total variation and Wasserstein distances in terms of the KL divergence using Pinsker's inequality and the transportation-entropy (Talagrand) inequality.

We use the following notation to represent the truncated versions of the two probability measures. First, for any probability measure μ and measurable set A, let $[\mu]_A$ denote μ truncated to set A. $[\mu]_A$ is constructed by restricting μ to set A and then renormalizing it so that $[\mu]_A$ is a well-defined probability measure. Now, let $B_0\left(\bar{\delta}\sqrt{n}\right) := \{\theta : \|\theta\| \leq \bar{\delta}\sqrt{n}\}$ denote the inner region, and define the truncated posterior and normal measures respectively as $\mu_n^{\pi} := \left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n\right)\right)\right]_{B_0\left(\bar{\delta}\sqrt{n}\right)}$ and $\mu_n^{\mathcal{N}} = \left[\mathcal{N}\left(0, \bar{J}_n(\bar{\theta}_n)^{-1}\right)\right]_{B_0\left(\bar{\delta}\sqrt{n}\right)}$, with densities with respect to the Lebesgue measure given respectively by f_n^{π} and $f_n^{\mathcal{N}}$. By Taylor expanding $(\log f_n^{\pi})'$ around $\bar{\theta}_n$ and applying Assumption 3 we can successfully control the following *Fisher divergence* between the two truncated distributions:

$$\mathrm{Fisher}\left(\mu_n^{\mathcal{N}} \| \mu_n^{\pi}\right) := \int \left\| \left(\log f_n^{\pi}\right)'(t) - \left(\log f_n^{\mathcal{N}}\right)'(t) \right\|^2 \mu_n^{\mathcal{N}}(dt) \leq \frac{3\left(\mathrm{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right] \overline{M}_2\right)^2}{4n}. \tag{5.1}$$

Now, by Assumptions 3 and 5, combined with Taylor's theorem, we show that the density of μ_n^π is $\left(\overline{\lambda}_{\min}(\bar{\theta}_n) - \overline{\delta M}_2\right)$-strongly log-concave on $B_0\left(\bar{\delta}\sqrt{n}\right)$ (i.e. the logarithm of its density is a strongly concave function). The celebrated *Bakry-Émery* criterion (Bakry and Émery, 1985) (see also Schlichting 2019, Theorem A1) says that any strongly log-concave measure satisfies the *log-Sobolev inequality*, which is an inequality between the KL divergence (or relative entropy) and the Fisher divergence (see Bakry et al. 2016, Chapter 5 for a comprehensive treatment of the topic or Vempala and Wibisono 2019, Section 2 for a brief and intuitive explanation of the main ideas). Specifically, in our context, the log-Sobolev inequality for μ_n^π on the set $B_0\left(\bar{\delta}\sqrt{n}\right)$ implies that:

$$\mathrm{KL}[\mu_n^{\mathcal{N}}||\mu_n^\pi] := \int \log\left(\frac{f_n^{\mathcal{N}}(x)}{f_n^\pi(x)}\right)\mu_n^{\mathcal{N}}(dx) \leq \frac{\mathrm{Fisher}(\mu_n^{\mathcal{N}}\|\mu_n^\pi)}{2\left(\overline{\lambda}_{\min}(\bar{\theta}_n) - \overline{\delta M}_2\right)}. \tag{5.2}$$

Note that we could also use the log-Sobolev inequality for $\mu_n^{\mathcal{N}}$ and obtain an bound on $\mathrm{KL}(\mu_n^\pi||\mu_n^{\mathcal{N}})$ expressed in terms of $\mathrm{Fisher}(\mu_n^\pi\|\mu_n^{\mathcal{N}})$. However, such a bound would not be as useful as it would involve integrating with respect to μ_n^π, which is often intractable. This is why we intentionally use the log-Sobolev inequality for μ_n^π, which yields the upper bound in Eq. (5.2) expressed in terms of $\mathrm{Fisher}(\mu_n^{\mathcal{N}}\|\mu_n^\pi)$, thus involving only Gaussian integration. Because of that, we derive a bound expressed in terms of the right-hand side of Eq. (5.1).

Finally, we can then control the total variation and Wasserstein distances in terms of the KL divergence. By Pinsker's inequality (Massart, 2007, Theorem 2.16) we have that

$$\mathrm{TV}\left(\mu_n^\pi, \mu_n^{\mathcal{N}}\right) \leq \sqrt{\frac{1}{2}\mathrm{KL}[\mu_n^{\mathcal{N}}||\mu_n^\pi]},$$

which gives us the desired control over the total variation distance. In order to control the 1-Wasserstein distance and the integral probability metric introduced in Theorem 20, we bound the 2-Wasserstein distance given by:

$$\mathrm{W}_2\left(\mu_n^\pi, \mu_n^{\mathcal{N}}\right) := \inf_\Gamma \sqrt{\mathbb{E}_\Gamma\left[\|X-Y\|^2\right]},$$

where the infimum is taken over all distributions Γ of (X, Y) with the correct marginals $X \sim \mu_n^\pi$ and $Y \sim \mu_n^{\mathcal{N}}$. Indeed, the 2-Wasserstein distance is known to upper bound the 1-Wasserstein distance and a transformation of it upper-bounds the integral probability metric appearing in Theorem 20. We upper-bound the 2-Wasserstein distance by the KL divergence, which we have previously controlled in Eq. (5.1). We use the *transportation-entropy (Talagrand) inequality*, which is implied by the log-Sobolev inequality (see Gozlan 2009, Theorem 4.1 for the specific result we use or Vempala and Wibisono 2019, Section 2.2.1 for an intuitive explanation). In our context, this inequality says that

$$\mathrm{W}_2\left(\mu_n^\pi, \mu_n^{\mathcal{N}}\right) \leq \sqrt{\frac{2\mathrm{KL}\left[\mu_n^{\mathcal{N}}||\mu_n^\pi\right]}{\overline{\lambda}_{\min}(\bar{\theta}_n) - \overline{\delta M}_2}}.$$

Combined with Eq. (5.2) it yields the desired control over the 1-Wasserstein distance and the metric considered in Theorem 20, inside the inner region $B_0\left(\bar{\delta}\sqrt{n}\right)$.

It is precisely the bounds on the total-variational distance and the 1-Wasserstein distance between the truncated measures μ_n^π and $\mu_n^{\mathcal{N}}$ that yield the leading summands of orders

$C_d\sqrt{d^2/n}$ and $\bar{C}_d\sqrt{d^2/n}$ in the bounds in Theorems 17 and 18 respectively, described in detail in Section 4.2. The leading summand in Theorem 20 is of order $\tilde{C}_d\sqrt{d^3/n}$ (using the notation of Section 4.2) because of the following part of our proof, which can be found in full detail in Appendix E.3.2. Suppose that $X \sim \mu_n^\pi$ and $Y \sim \mu_n^{\mathcal{N}}$. Consider any vector $v \in \mathbb{R}^d$ with $\|v\| \leq 1$ and any coupling $(\tilde{X}, \tilde{Y})$ between μ_n^π and $\mu_n^{\mathcal{N}}$. Then,

$$\begin{aligned}\left|\mathbb{E}\left[\langle v, X\rangle^2\right] - \mathbb{E}\left[\langle v, Y\rangle^2\right]\right| =& \left|\mathbb{E}\left[\left\langle v, \tilde{X}-\tilde{Y}\right\rangle \left\langle v, \tilde{X}+\tilde{Y}\right\rangle\right]\right| \\ =& \left|\mathbb{E}\left[\left\langle v, \tilde{X}-\tilde{Y}\right\rangle^2\right] + 2\mathbb{E}\left[\left\langle v, \tilde{X}-\tilde{Y}\right\rangle \left\langle v, \tilde{Y}\right\rangle\right]\right| \\ \leq& \mathbb{E}\left[\left\|\tilde{X}-\tilde{Y}\right\|^2\right] + 2\sqrt{\mathbb{E}\left[\left\|\tilde{X}-\tilde{Y}\right\|^2\right]}\sqrt{\mathbb{E}\left[\|Y\|^2\right]},\end{aligned}$$

from which it follows that

$$\sup_{\|v\|\leq 1}\left|\mathbb{E}\left[\langle v, X\rangle^2\right] - \mathbb{E}\left[\langle v, Y\rangle^2\right]\right| \leq \left(\mathrm{W}_2\left(\mu_n^\pi, \mu_n^{\mathcal{N}}\right)\right)^2 + 2\mathrm{W}_2\left(\mu_n^\pi, \mu_n^{\mathcal{N}}\right)\sqrt{\mathbb{E}\left[\|Y\|^2\right]}.$$

We obtain a bound on $\mathrm{W}_2\left(\mu_n^\pi, \mu_n^{\mathcal{N}}\right)$ that is of order $\bar{C}_d\sqrt{d^2/n}$ (and appears as the leading summand in the final bound in Theorem 18). Yet, at the same time, $\sqrt{\mathbb{E}\left[\|Y\|^2\right]}$ is of order $\frac{\sqrt{d}}{\overline{\lambda}_{\min}(\bar{\theta}_n)}$. As a result, the leading term in the bound of Theorem 20 is of order $\tilde{C}_d\sqrt{d^3/n}$.

5.2 The MLE-Centric Approach

Theorems 22, 23 and 25 establish our bounds in the MLE-centric approach. Our proof technique is very similar to the strategy we used in the MAP-centric approach. We compare the distribution of $\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n\right)$ to $\mathcal{N}\left(0, \hat{J}_n(\hat{\theta}_n)^{-1}\right)$ by concentrating separately on the region $\{\theta : \|\theta\| \leq \delta\sqrt{n}\}$ and the region $\{\theta : \|\theta\| > \delta\sqrt{n}\}$. In the outer region, we use Assumption 9 and Gaussian concentration inequalities. In the inner region, we use the log-Sobolev inequality.

5.3 Our theory for general test functions in one dimension

Theorem 27 in the appendix establishes our bounds for general test functions in a single dimension. Its proof also divides the domain into an inner and outer region, but proceeds using Stein's method instead of via the Fisher divergence, using known properties of solutions to the Stein equation in one dimension. As with our control over the Fisher divergence, the key to our proof is using Taylor's expansions and expressing the bound as an integral over the known measure $\mu_n^{\mathcal{N}}$.

6. Further results: MLE-centered approach

In this section we present some further results on the quality of Laplace approximation in the *MLE-centric* approach. The proofs of Theorems 22, 23 and 25 presented below can be found in Appendices D and F.

6.1 Control over the TV distance in the MLE-centric approach

We start by controlling the total variation distance.

Theorem 22 *Fix $n \in \mathbb{N}$ and suppose that Assumptions 1, 2 and 7 – 10 hold. Let TV denote the total variation distance. We have the following upper bound:*

$$TV\Big(\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n\right)\right), \mathcal{N}(0, \hat{J}_n(\hat{\theta}_n)^{-1})\Big) \leq D_1 n^{-1/2} + D_2 n^{d/2} e^{-n\kappa} + 2\hat{\mathscr{D}}(n,\delta),$$

where

$$D_1 := \frac{\sqrt{\widetilde{M}_1 \hat{M}_1}}{2\sqrt{\lambda_{\min}(\hat{\theta}_n) - \delta M_2}} \left(\frac{\sqrt{3}\,\mathrm{Tr}\left[\hat{J}_n(\hat{\theta}_n)^{-1}\right]}{2\sqrt{1 - \hat{\mathscr{D}}(n,\delta)}} M_2 + M_1 \right);$$

$$D_2 := \frac{2\hat{M}_1 \left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2}}{(2\pi)^{d/2}\left(1 - \hat{\mathscr{D}}^p(n,\delta)\right)}.$$

6.2 Control over the 1-Wasserstein distance in the MLE-centric approach

Theorem 23 *Fix $n \in \mathbb{N}$ and suppose that Assumptions 1, 2 and 7 – 10 hold. We have the following upper bound:*

$$\begin{aligned} &W_1\left(\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n\right)\right), \mathcal{N}(0, \hat{J}_n(\hat{\theta}_n)^{-1})\right) \\ &\quad \leq E_1 n^{-1/2} + E_3\left(E_2 + \sqrt{n}E_4\right) n^{d/2} e^{-n\kappa} + \left(\delta\sqrt{n} + \sqrt{\frac{2\pi}{\lambda_{\min}(\hat{\theta}_n)}} + E_2\right)\hat{\mathscr{D}}(n,\delta), \end{aligned}$$

where

$$E_1 := \frac{\widetilde{M}_1 \hat{M}_1}{\lambda_{\min}(\hat{\theta}_n) - \delta M_2} \left(\frac{\sqrt{3}\,\mathrm{Tr}\left[\hat{J}_n(\hat{\theta}_n)^{-1}\right]}{2\sqrt{1 - \hat{\mathscr{D}}(n,\delta)}} M_2 + M_1 \right)$$

$$E_2 := \frac{\hat{M}_1 \widetilde{M}_1 \left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2} \left|\det\left(\hat{J}_n^m(\hat{\theta}_n,\delta)\right)\right|^{-1/2} \sqrt{\mathrm{Tr}\left[\hat{J}_n^m(\hat{\theta}_n,\delta)^{-1}\right]}}{1 - \hat{\mathscr{D}}^p(n,\delta)}$$

$$E_3 = \frac{\hat{M}_1 \left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2}}{(2\pi)^{d/2}\left(1 - \hat{\mathscr{D}}^p(n,\delta)\right)}; \qquad E_4 := \int_{\|u - \hat{\theta}_n\| > \delta} \|u - \hat{\theta}_n\| \pi(u) du.$$

Remark 24 *Note that, in order for the bound in Theorem 23 to be finite, we need the following integral with respect to the prior: $\int_{\|u-\hat{\theta}_n\|>\delta} \|u - \hat{\theta}_n\|\pi(u)du$ to be finite.*

6.3 Control over the difference of covariances in the MLE-centric approach

Theorem 25 *Fix $n \in \mathbb{N}$ and suppose that Assumptions 1, 2 and 7 – 10 hold. Let $\boldsymbol{Z}_n \sim \mathcal{N}(0, \hat{J}_n(\hat{\theta}_n)^{-1})$. We have that*

$$\sup_{v:\|v\|\leq 1} \left| \mathbb{E}\left[\left\langle v, \sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n\right)\right\rangle^2 \right] - \mathbb{E}\left[\langle v, \boldsymbol{Z}_n \rangle^2 \right] \right|$$
$$\leq (F_1)^2 n^{-1} + F_1 F_2 n^{-1/2} + \frac{F_5 (F_3)^2 + F_3 F_4 \sqrt{n}}{(2\pi)^{d/2}} n^{d/2} e^{-n\kappa}$$
$$+ \left(\delta^2 n + \sqrt{\frac{2\pi}{\lambda_{\min}(\hat{\theta}_n)}} + F_3 F_5 \right) \hat{\mathscr{D}}(n, \delta)$$

where

$$F_1 := \frac{\widetilde{M_1} \hat{M}_1}{\lambda_{\min}(\hat{\theta}_n) - \delta M_2} \left(\frac{\sqrt{3}\,\mathrm{Tr}\left[\hat{J}_n(\hat{\theta}_n)^{-1} \right]}{\sqrt{1 - \hat{\mathscr{D}}(n,\delta)}} M_2 + M_1 \right); \qquad F_2 := \frac{2\sqrt{\mathrm{Tr}\left[\hat{J}_n(\hat{\theta}_n)^{-1} \right]}}{\sqrt{1 - \hat{\mathscr{D}}(n,\delta)}};$$

$$F_3 := \frac{\hat{M}_1 \left| \det\left(\hat{J}_n^p(\hat{\theta}_n, \delta) \right) \right|^{1/2}}{1 - \hat{\mathscr{D}}^p(n,\delta)}; \qquad F_4 := \int_{\|u\|>\delta} \|u\|^2 \pi(u + \hat{\theta}_n) du;$$

$$F_5 := \widetilde{M_1} \left| \det\left(\hat{J}_n^m(\hat{\theta}_n, \delta) \right) \right|^{-1/2} \mathrm{Tr}\left[\hat{J}_n^m(\hat{\theta}_n, \delta)^{-1} \right].$$

Remark 26 *Note that, in order for the bound in Theorem 25 to be finite, we need the following integral with respect to the prior to be finite: $\int_{\|u\|>\delta} \|u\|^2 \pi(u + \hat{\theta}_n) du$.*

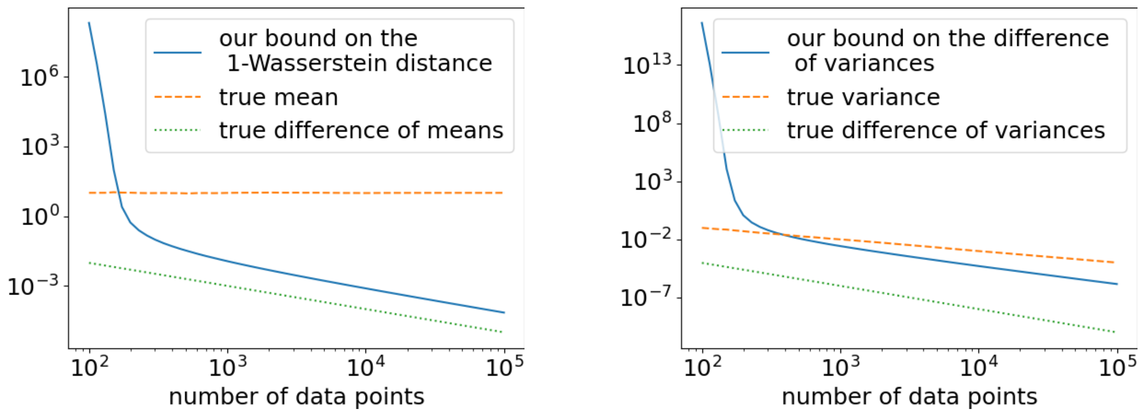
7. Example applications

Now, we present examples of our bounds computed for different models.[2] None of these examples correspond to strongly log-concave posteriors, making them challenging for the Laplace approximation. Indeed, the first one yields a weakly log-concave posterior, and the second and third one produce non-log-concave posteriors. They all show that our bounds are explicitly computable for a variety of commonly used models, including heavy-tailed ones, to which the methods currently available in the literature do not apply. Our examples also show that the computations may be executed in a practical amount of time. Our bounds also go well below the true values of the mean and the norm of the covariance, for reasonable sample sizes. This comparison indicates that they are applicable for practitioners who wish to assess how confident they should be in their mean and variance estimates.

7.1 Our bounds work under misspecification: Poisson likelihood with gamma prior and exponential data

First, we look at a one-dimensional conjugate model, for which we can compare our bounds on the difference of means and the difference of variances to the ground truth. We consider

2. The code for our experiments is available at `https://github.com/mikkasprzak/laplace_approximation.git`



(a) Posterior mean

(b) Posterior variance

Figure 1: Poisson likelihood with gamma prior and exponential data (MAP-centric approach)

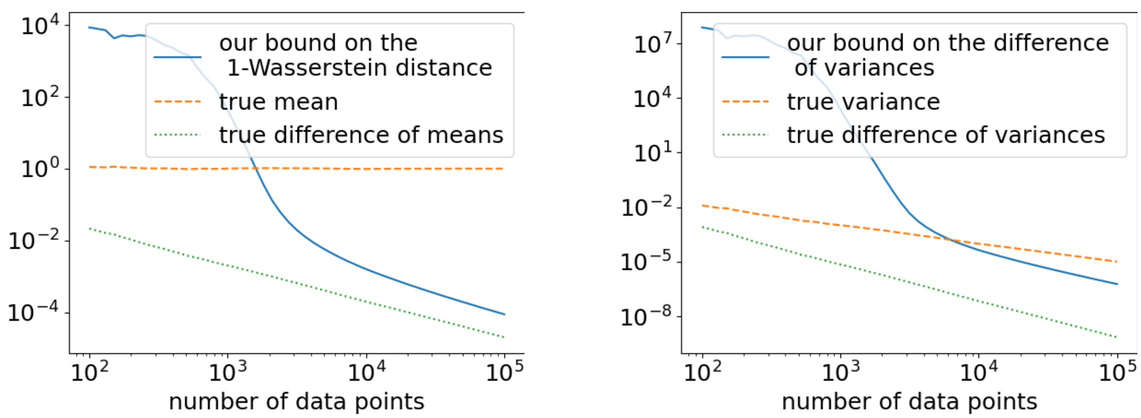
a Poisson likelihood and a gamma prior with shape equal to 0.1 and rate equal to 3, which yields a weakly log-concave posterior. Our data are generated from the exponential distribution with mean 10. Figures 1a and 1b show that our bounds (in the MAP-centric approach) on the difference of means and the difference of variances get close to the true difference of means and the true difference of variances for sample sizes above around 250. More detail on how one can compute the constants appearing in the bounds can be found in Appendix H.2.

7.2 Our bounds work for heavy-tailed posteriors: Weibull likelihood with inverse-gamma prior

Now, we consider another conjugate model and compare our bounds to the ground truth. In this case, the posterior is heavy-tailed, which represents one common way that non-log-concavity arises in practice. In our experiment, we set the shape of the Weibull to $\frac{1}{2}$, and we make inference about the scale. The prior is inverse-gamma with shape equal to 3 and scale equal to 10. The data are Weibull with shape 1/2 and scale 1. Figures 2a and 2b demonstrate that our bounds on the difference of means and the difference of variances (in the MAP-centric approach) get close to the true difference of means and the true difference of variances for sample sizes in low thousands. Our bounds take large values for small sample sizes — a phenomenon explained in detail in Section 4.1 above. More detail on how one can compute the constants appearing in the bounds can be found in Appendix H.3.

7.3 Our bounds work for multivariate non-log-concave posteriors: logistic regression with Student's t prior

While the examples above have a one-dimensional parameter, we next show that our bounds are computable and non-vacuous for a multivariate posterior. Like the Weibull example



(a) Posterior mean (b) Posterior variance

Figure 2: Weibull likelihood with inverse-gamma prior (MAP-centric approach)

above, it is not log-concave. But in this case, the non-log-concavity arises closer to the central mass of the distribution rather than in the tails.

7.3.1 Setup

Suppose $X_1, \ldots, X_n \in \mathbb{R}^d$ and $Y_1, \ldots, Y_n \in \{-1, 1\}$. We will study the following log-likelihood:

$$L_n(\theta) := L_n\left(\theta \middle| \left(Y_i\right)_{i=1}^n, \left(X_i\right)_{i=1}^n\right) = -\sum_{i=1}^n \log\Big(1 + e^{-X_i^T \theta Y_i}\Big). \tag{7.1}$$

For a covariance matrix Σ, a vector $\mu \in \mathbb{R}^d$ and the number of degrees of freedom $\nu > 0$, we consider a d-dimensional Student's t prior on θ, given by

$$\pi(\theta) = \frac{\Gamma((\nu + d)/2)}{\Gamma(\nu/2)\nu^{d/2}\pi^{d/2}|\Sigma|^{1/2}} \left[1 + \frac{1}{\nu}(\theta - \mu)^T \Sigma^{-1}(\theta - \mu)\right]^{-(\nu+d)/2}. \tag{7.2}$$

Combining the log-likelihood given by Eq. (7.1) with the prior given by Eq. (7.2) yields a posterior that is known not to be log-concave. In Appendix H.4 we compute analytically the constants arising in our bounds in the case of the posterior coming from the likelihood given in Eq. (7.1) and prior given in Eq. (7.2). In our present experiments, we choose the Student's t prior to have mean zero, identity covariance, and four degrees of freedom.

We performed our experiments for the MAP-centric approach. We derived the values for the MAP and MLE numerically; recall that we use both the MAP and MLE in our MAP-centric bounds. Moreover, in order to sharpen our bounds, we ran a built-in global optimizer scipy.optimize.shgo to derive the maximum of the third derivative inside a ball around the MLE and the MAP, i.e. to derive M_2 and $\overline{M}_2$. As there was a certain degree of choice for δ and $\bar{\delta}$, we also optimized the choice thereof numerically. The data $(Y_i)_{i=1}^n$ we used came from logistic regression with parameter $(1, 1, 1, 1, 1)$, where we simulated the $(X_i)_{i=1}^n$ i.i.d. from the 5-dimensional normal distribution with mean $(0, 0, 0, 0, 0)$ and covariance $\frac{0.15}{\sqrt{5}} I_{5\times 5}$, where $I_{5\times 5}$ is the 5×5 identity matrix. We compared our bound from Theorem 18 with the Euclidean norm of the difference between the MAP and the simulated true posterior mean. We obtained our simulated true posterior mean through MCMC, using the standard *Stan* implementation (with 4 parallel chains). Similarly, we compared our bound from Theorem 20 with the operator norm of the difference between $\frac{\bar{J}_n(\bar{\theta}_n)^{-1}}{n}$ and the simulated true posterior covariance, obtained again through the standard *Stan* MCMC implementation.

7.3.2 Results

Figures 3a and 3b demonstrate that, in this case, our bounds on the 1-Wasserstein distance and the 2-norm of the difference of covariances tightly control the true difference of means and the true difference of covariances, respectively, for sample sizes of about 10,000 or higher. Our bounds are still computable and accurate at lower sample sizes, but they can become loose enough as to lose practicality.

7.4 Our bounds become looser and more difficult to compute as parameter dimension increases: logistic regression with Student's t prior

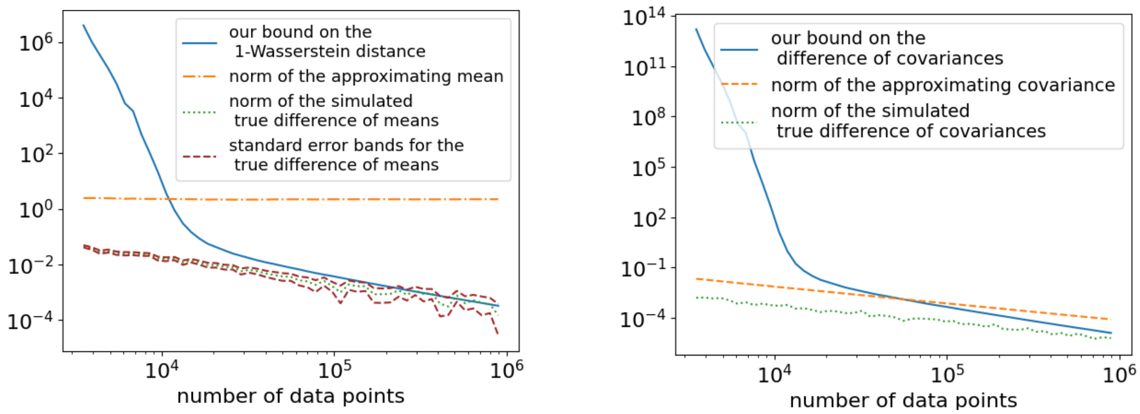
We use a similar setup as in Section 7.3.1, but now we vary the dimension d between 1 and 9. We find that increasing d loosens our bounds and makes them more difficult to compute. But we still find that they are computable and tight for reasonable data set sizes.

Specifically, we follow the setup in Section 7.3.1 except that now we simulate $(X_i)_{i=1}^n$ i.i.d. from the d-dimensional normal distribution with mean $(0, \ldots, 0)$ and covariance $\frac{0.15}{\sqrt{d}} I_{d\times d}$. We simulate the corresponding $(Y_i)_{i=1}^n$ from d-dimensional logistic regression with parameter $(0, \ldots, 0)$. We compare our bounds across different dimensions in Figure 4. While we see that the bounds generally increase with dimension, we also see that they remain practically small for some portion of the plotted data range for every dimension we consider.

There are two ways in which our bounds become more difficult to compute with increasing dimension. First, we cover the ability to compute the bounds. Second, we cover computational speed in cases where the bounds are possible to compute.

On the first point, we are able to compute our bounds when Assumption 5 is satisfied. The sample sizes where Assumption 5 is satisfied vary by dimension. In Figure 4, we plot our bounds for every number of data points where Assumption 5 is satisfied. For the same data sets and priors, we calculated the minimum number of data points required to satisfy Assumption 5, for the same range of dimensions: d between 1 and 9. See Figure 5. We find that the minimum number of data points remains practical in this range.

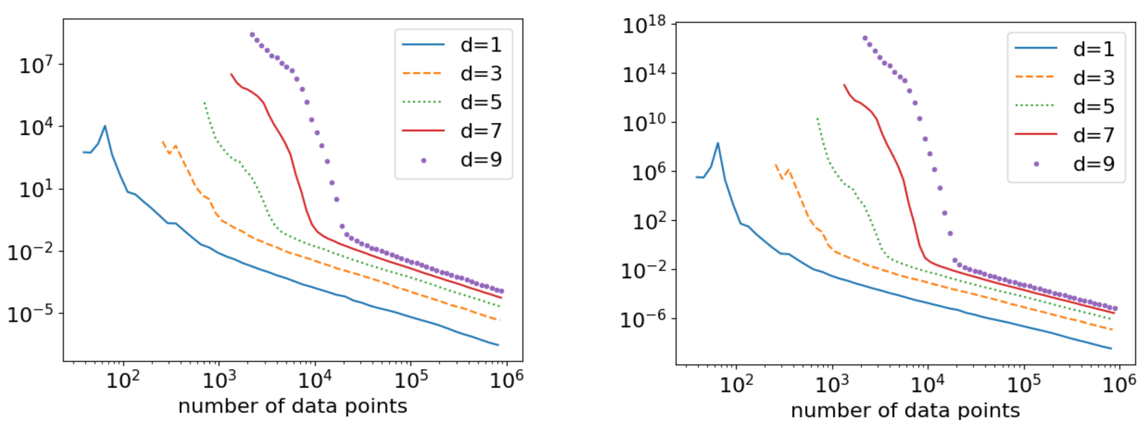
When we are able to compute our bounds, computation still slows down as dimension increases due to the numerical optimization involved in calculating $\overline{M}_2$. The slowdown is mainly caused by the global optimizer scipy.optimize.shgo becoming slow in higher dimensions. To create Figure 4, we computed our bounds on the 1-Wasserstein distance and the 2-norm of the difference of covariances, simultaneously, on 51 points spread uniformly on the logarithmic scale between the minimum sample size given by Figure 5 and the sample size of 10^6. Performing this experiment on a MacBook Pro with the Apple M3 Max Chip and 16 cores in total for $d = 1$ took us about 1.5 total minutes, for $d = 3$ about 3.35 minutes, for $d = 5$ about 10 minutes, for $d = 7$ about 36 minutes and for $d = 9$ about 147 minutes.



(a) Posterior mean

(b) Posterior covariance

Figure 3: Our bounds for 5-dimensional logistic regression with t prior (MAP-centric approach) compared to the simulated true difference of means and covariances (with Monte Carlo standard error bands included for the posterior mean). The data $(Y_i)_{i=1}^n$ come from logistic regression with parameter $(1, 1, 1, 1, 1)$, where the $(X_i)_{i=1}^n$ are simulated i.i.d. from the 5-dimensional normal distribution with mean $(0, 0, 0, 0, 0)$ and covariance $\frac{0.15}{\sqrt{5}} I_{5\times 5}$.



(a) Our bound on 1-Wasserstein distance

(b) Our bound on difference of covariances

Figure 4: Comparison of our bounds for the logistic regression with t prior (MAP-centric approach) for different dimensions d. The data $(Y_i)_{i=1}^n$ come from logistic regression with parameter $(0, \ldots, 0)$, where the $(X_i)_{i=1}^n$ are simulated i.i.d. from the normal distribution with mean $(0, \ldots, 0)$ and covariance $\frac{0.15}{\sqrt{d}} I_{d\times d}$.

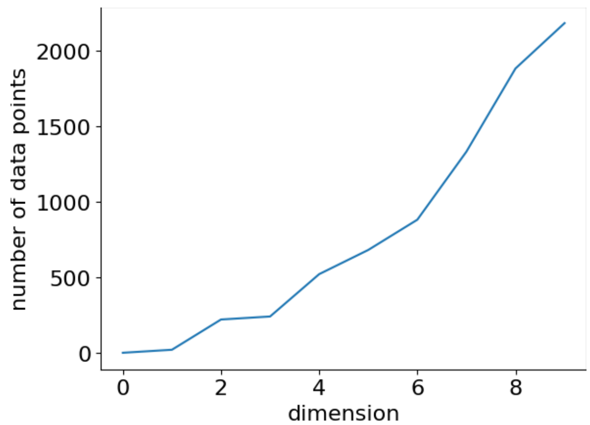
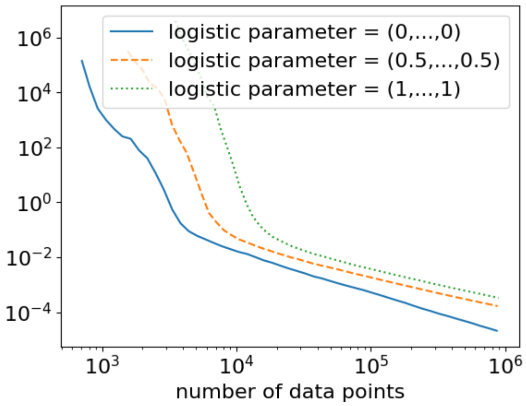


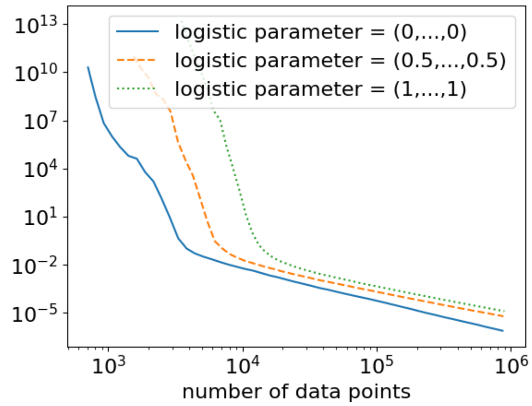
Figure 5: Minimum number of data points required to compute our bounds for logistic regression with t prior (MAP-centric approach). The data $(Y_i)_{i=1}^n$ come from logistic regression with parameter $(0, \ldots, 0)$, where the $(X_i)_{i=1}^n$ are simulated i.i.d. from the normal distribution with mean $(0, \ldots, 0)$ and covariance $\frac{0.15}{\sqrt{d}} I_{d \times d}$.

7.5 Are our bounds smaller for small condition numbers? Logistic regression with Student's t prior

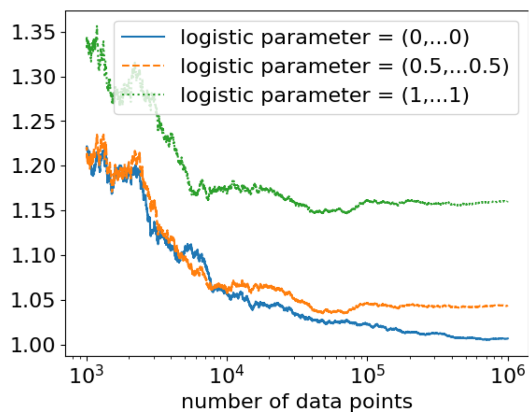
In all of our experiments above, we see that our bounds are tighter for larger sample sizes and become loose at small numbers of data points; see Section 4.1. We hypothesized that the looseness of our bounds might be a function of condition number (defined as the ratio of the largest and the smallest eigenvalue of $\bar{J}_n(\bar{\theta}_n)$). To test this hypothesis, we calculated and compared our bounds from Theorems 18 and 20 for 5-dimensional logistic regression on three different datasets. As before, the prior was Student's t with 4 degrees of freedom, mean zero, and identity covariance. We simulated the data $(X_i)_{i=1}^n$ i.i.d. from the 5-dimensional normal distribution with mean $(0, 0, 0, 0, 0)$ and covariance $\frac{0.15}{\sqrt{5}} I_{5 \times 5}$. We simulated the corresponding $(Y_i)_{i=1}^n$ from logistic regression but now with three different parameter values: $(0, 0, 0, 0, 0)$ in the first case, $(0.5, 0.5, 0.5, 0.5, 0.5)$ in the second case, and $(1, 1, 1, 1, 1)$ in the third case. In Figures 6a and 6b, we plot how our bound varies both by number of data points and by the logistic parameter. In Figure 6c we plot how the condition number varies by number of data points and logistic parameter. As expected, the condition number is higher for smaller numbers of data points in Figure 6c, and that range of data points roughly corresponds to the range where our bound is looser in Figures 6a and 6b. But we do not see that condition number is decisive. For instance, we see that our bound is loose for condition numbers above 1.15 on the orange curve (parameter $(0.5, 0.5, 0.5, 0.5, 0.5)$); compare Figures 6a and 6b and Figure 6c. But we also see that the condition numbers for the green curve (parameter $(1, 1, 1, 1, 1)$) are generally at or above 1.15 (Figure 6c), but our bounds are still tight for high sample sizes (Figures 6a and 6b). See Section 4.1 for an additional discussion.



(a) Our bound on 1-Wasserstein distance



(b) Our bound on difference of covariances



(c) Condition number

Figure 6: Comparison of our bounds for the 5-dimensional logistic regression with t prior (MAP-centric approach) for different data sets and the comparison of the condition number calculated for those data sets. The data $(X_i)_{i=1}^n$ were in each case simulated i.i.d. from the 5-dimensional normal distribution with mean $(0,0,0,0,0)$ and covariance $\frac{0.15}{\sqrt{5}}I_{5\times 5}$. The corresponding $(Y_i)_{i=1}^n$ were simulated from logistic regression with parameter $(0,0,0,0,0)$ in the first case, $(0.5,0.5,0.5,0.5,0.5)$ in the second case and $(1,1,1,1,1)$ in the third case.

8. Conclusions and future work

We provide bounds on the quality of the Laplace approximation that are computable and hold under the standard assumptions of the Bernstein–von Mises Theorem. We control the total-variation distance, the 1-Wasserstein distance, and another useful integral probability metric. Our bounds on the total-variation and 1-Wasserstein distance are such that their sample-size and dimension dependence cannot be improved in general. However, the constants in our bounds are evidently sub-optimal, which makes our bounds unnecessarily large for small and moderate sample sizes. We hope that the values of those constants may be reduced in the future.

An interesting question is whether, for multivariate posteriors, we could derive bounds on more general integral probability metrics, in a way similar to our univariate Theorem 27. We proved Theorem 27 using Stein's method and we present it as part of the appendix. Whether or not this result could be extended to the multivariate context is to a certain extent a question about the applicability of techniques like Stein's method in dimension greater than one for measures truncated to a bounded convex set. So far we have struggled to find enough theory that would let us compute useful bounds using this approach.

Acknowledgments

The authors would like to thank Krishna Balasubramanian, Xiao Fang, Max Fathi, Jonathan Huggins, Anya Katsevich, Tin D. Nguyen, Jeff Miller, Daniel Paulin, Gesine Reinert, Håvard Rue, Adil Salim, Vladimir Spokoiny and Yvik Swan for interesting discussions. We are particularly grateful to Anya Katsevich for a discussion concerning the dimension dependence. This work is part of project **Stein-ML** that has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement **No 101024264**. Additionally, this work was supported by France 2030. This work was also supported in part by an NSF Career Award and an ONR Early Career Grant.

Appendix A. An additional general univariate bound

Now we present the following bound, which works for pretty arbitrary test functions, but only in dimension one. The proof of this bound is different from the proofs of the results presented in the main body of the paper as it relies on Stein's method instead of the log-Sobolev inequality. Stein's method allows us to achieve control over the difference of expectations with respect to the rescaled posterior and with respect to the approximating Gaussian of very general test functions. The particular properties of Stein's method we use, however, only yield bounds in dimension one. We believe it would be interesting and useful to derive such general bounds in higher dimensions in the future. The proof of Theorem 27 presented below can be found in Appendices D and G.

Theorem 27 *Assume that we study a* ***univariate posterior****, i.e. that* $d = 1$*. Let* $\sigma_n^2 := \hat{J}_n(\hat{\theta}_n)^{-1}$*. Suppose that Assumptions 1, 2 and 8 – 10 hold. Let* $\boldsymbol{Z}_n \sim \mathcal{N}(0, \sigma_n^2)$*. Then, for any function* $g : \mathbb{R} \to \mathbb{R}$ *which is integrable with respect to the posterior and with respect to*

$\mathcal{N}(0,\sigma_n^2)$,

$$\begin{aligned}&\left|\mathbb{E}\left[g\left(\sqrt{n}(\tilde{\boldsymbol{\theta}}_n-\hat{\theta}_n)\right)\right]-\mathbb{E}\left[g(\boldsymbol{Z}_n)\right]\right|\\ \leq&(G_1+G_2)n^{-1/2}+\left|\int_{|u|>\delta\sqrt{n}}g(u)\frac{e^{-u^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}}du\right|\\ &+\left(G_3\int_{|u|>\delta}|g(u\sqrt{n})|\pi(u+\hat{\theta}_n)du+G_4\right)n^{1/2}e^{-n\kappa}+G_5e^{-\delta^2n/(2\sigma_n^2)},\end{aligned}$$

for

$$\begin{aligned}C_n^{(1)}&:=\frac{1}{2\sigma_n^2}-\frac{\delta M_2}{3}>0; & C_n^{(2)}&:=\left(\frac{1}{2\sigma_n^2}-\frac{\delta M_2}{6}\right)>0;\\ C_n^{(3)}&:=\left(\frac{1}{2\sigma_n^2}+\frac{\delta M_2}{3}\right)>0; & C_n^{(4)}&:=\left(\frac{1}{2\sigma_n^2}+\frac{\delta M_2}{6}\right)>0\end{aligned}$$

and

$$\begin{aligned}G_1&:=\frac{2\widetilde{M}_1\hat{M}_1}{\sqrt{2\pi\sigma_n^2}}\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}}|ug(u)|\left[\left(M_1+\frac{3}{\delta}\right)e^{-C_n^{(1)}u^2}-\frac{3}{\delta}e^{-C_n^{(2)}u^2}\right]du;\\ G_2&:=\frac{2\sqrt{C_n^{(4)}}\left(\widetilde{M}_1\hat{M}_1\right)^2\left(M_1+\frac{3}{\delta}\right)\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}}|g(u)|e^{-C_n^{(2)}u^2}du}{C_n^{(1)}\pi\sqrt{\sigma_n^2}\left(1-2e^{-\delta^2nC_n^{(4)}}\right)}\left(\frac{M_1+\frac{3}{\delta}}{C_n^{(1)}}-\frac{3}{\delta C_n^{(2)}}\right);\\ G_3&:=\frac{\hat{M}_1\sqrt{C_n^{(4)}}}{\sqrt{\pi}\left\{1-2\exp\left[-C_n^{(4)}\delta^2n\right]\right\}};\\ G_4&:=\frac{\hat{M}_1^2\widetilde{M}_1C_n^{(4)}\int_{|t|\leq\delta\sqrt{n}}|g(t)|e^{-C_n^{(2)}t^2}dt}{\pi\left\{1-2\exp\left[-C_n^{(4)}\delta^2n\right]\right\}^2};\\ G_5&:=\frac{2\hat{M}_1\widetilde{M}_1\sqrt{C_n^{(4)}}\int_{|t|\leq\delta\sqrt{n}}|g(t)|e^{-C_n^{(2)}t^2}dt}{\sqrt{\pi}\left\{1-2\exp\left[-C_n^{(4)}\delta^2n\right]\right\}}.\end{aligned}$$

Remark 28 *The bound in Theorem 27 is for the MLE-centric approach. However, its proof can easily be modified in order to yield analogous bounds on the quality of approximation in the MAP-centric approach, as described in Section 2 and presented in Section 3.*

Constants G_1, G_2, G_4, G_5 appearing above may be straightforwardly controlled by controlling the involved Gaussian integrals. Moreover, the fact that $C_n^{(1)}, C_n^{(2)}, C_n^{(3)}, C_n^{(4)}$ are positive follows from Assumption 8.

Appendix B. Introduction to the Laplace approximation and the Bernstein–von Mises theorem

The foundations of Laplace approximation date back to the work of Laplace (1774) (see Laplace 1986 for an English translation and Bach 2021 for an intuitive discussion). It was originally introduced as a method of approximating integrals of the form

$$\mathrm{Int}(n) := \int_K e^{-nf(x)} dx, \quad n \in \mathbb{N},$$

where K is a subset of $\mathbb{R}^d$ and f is a real-valued function on $\mathbb{R}^d$. Suppose that $x^* \in K$ is a strict global maximizer of f on K. Heuristically, under appropriate smoothness assumptions on f, we can use Taylor's expansion to obtain:

$$\begin{aligned} f(x) \approx & f(x^*) + f'(x^*)(x - x^*) + \frac{1}{2}(x - x^*)^T f''(x^*)(x - x^*) \\ = & f(x^*) + \frac{1}{2}(x - x^*)^T f''(x^*)(x - x^*). \end{aligned}$$

We therefore have:

$$\mathrm{Int}(n) \approx \int_K \exp\left[-nf(x^*) - \frac{n}{2}(x - x^*)^T f''(x^*)(x - x^*)\right] dx$$

As a result, heuristically, up to a constant not depending on K, $\mathrm{Int}(n)$ can be approximated by the integral of the density of the Gaussian measure with mean x^* and covariance matrix given by $\frac{1}{n} f''(x^*)^{-1}$. Now, suppose that $e^{-nf(x)}$ is an unnormalized posterior density, i.e. that $\Pi_n(\cdot) \propto e^{-nf(\cdot)}$ for some posterior Π_n. Writing $\boldsymbol{\Pi}_n$ for the posterior probability measure and $\bar{\theta}_n$ for the posterior mode (i.e. the maximum a posteriori or MAP), the above statement suggests that

$$\boldsymbol{\Pi}_n \approx \mathcal{N}(\bar{\theta}_n, \left(\log(\Pi_n)''(\bar{\theta}_n)\right)^{-1}), \tag{B.1}$$

where $\mathcal{N}(\mu, \Sigma)$ denotes the normal law with mean μ and covariance Σ. The computation of the mean and covariance of the above Gaussian is in the majority of cases easy numerically. It can be achieved using standard optimization schemes and does not require access to integrals with respect to the posterior, the normalizing constant of the posterior density or the true parameter. This is why Laplace approximation is a popular tool in approximate Bayesian inference.

While the above heuristic considerations may be turned into rigorous statements under certain conditions, a proper probabilistic grounding for the Laplace approximation is provided by the Bernstein–von Mises (BvM) theorem. As described in numerous classical references, including Ghosh and Ramamoorthi (2003, Section 1.4) or van der Vaart (1998, Section 10.2), the BvM theorem says that under mild assumptions on the likelihood and the prior, the posterior distribution converges to a Gaussian law in the following sense. Suppose that $\tilde{\boldsymbol{\theta}}_n$ is distributed according to the posterior, obtained after observing n data points. Let θ_0 be the true parameter, $\hat{\theta}_n$ be the MLE and $I(\cdot)$ be the Fisher information matrix. Let TV denote the total variation distance and let the function $\mathcal{L}(\cdot)$ return the law of its argument. Then, if the model is well-specified and certain regularity conditions are satisfied,

$$\mathrm{TV}\left(\mathcal{L}\left(\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n)\right), \mathcal{N}(0, I(\theta_0)^{-1})\right) \xrightarrow{\mathcal{P}} 0, \qquad \text{as } n \to \infty, \tag{B.2}$$

where n is the number of data and the convergence occurs in probability with respect to the law of the data.

While the model being well-specified is a crucial assumption in the above statement, Kleijn and van der Vaart (2012) proved its modified version, under model misspecification. The main difference is in the limiting covariance matrix. Specifically, in this context, the authors assume the model is of the form $\theta \mapsto p_\theta$ and the observations are sampled from a density p_0 that is not necessarily of the form p_{θ_0} for some θ_0. They show that under certain regularity conditions,

$$\mathrm{TV}\left(\mathcal{L}\left(\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n)\right), \mathcal{N}(0, V(\theta^*)^{-1})\right) \xrightarrow{\mathcal{P}} 0, \qquad \text{as } n \to \infty, \tag{B.3}$$

where θ^* minimizes the Kullback-Leibler divergence $\theta \mapsto \int \log(p_0(x)/p_\theta(x))p_0(x)dx$ and $V(\theta^*)$ is minus the second derivative of this map, evaluated at θ^*.

Let us now denote by L_n the generalized log-likelihood. A closer look at the classical proofs, including Le Cam's one (see e.g. Ghosh and Ramamoorthi 2003, Section 1.4), or more recent ones, including that of Miller (2021, Appendix B), reveals that, under standard regularity conditions:

$$\mathrm{TV}\left(\mathcal{L}\left(\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n)\right), \mathcal{N}\left(0, \left[-\frac{L_n''(\hat{\theta}_n)}{n}\right]^{-1}\right)\right) \xrightarrow{\mathcal{P}} 0, \qquad \text{as } n \to \infty, \tag{B.4}$$

no matter if the model is well-specified or not. It is known that under mild assumptions the MLE $\hat{\theta}_n$ and the maximum a posteriori (MAP) $\bar{\theta}_n$ get arbitrarily close to each other as the number of data n goes to infinity. Similarly, denoting by $\overline{L}_n$ the logarithm of the posterior density, $\overline{L}_n$ and L_n get arbitrarily close as n goes to infinity. It can be shown, in a similar fashion to Eq. (B.4), that under standard regularity assumptions,

$$\mathrm{TV}\left(\mathcal{L}\left(\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n)\right), \mathcal{N}\left(0, \left[-\frac{\overline{L}_n''(\bar{\theta}_n)}{n}\right]^{-1}\right)\right) \xrightarrow{\mathcal{P}} 0, \qquad \text{as } n \to \infty. \tag{B.5}$$

Equation (B.5) gives a rigorous justification and meaning to the Laplace approximation given by Eq. (B.1) and Eq. (B.4) provides its alternative version. The approximating covariance in both Eqs. (B.4) and (B.5) is computable without access to the posterior normalizing constant or the true parameter.

The recent paper Miller (2021) proves almost sure versions of the statements given in Eqs. (B.2) and (B.3) for a large collection of commonly used models. Naturally, similar almost sure convergence statements can be obtained for the approximations appearing in Eqs. (B.4) and (B.5).

Appendix C. Proofs of Propositions 13 and 15

C.1 Proof of Proposition 13

The proof is inspired by the proof of Miller (2021, Theorem 12). Note that $L_n(\theta) = -n\alpha(\theta) + \theta^T S_n$, where $S_n = \sum_{i=1}^n s(Y_i)$. By standard exponential family theory (Miller and Harrison, 2014, Proposition 19), α is C^∞, strictly convex on Θ, $\alpha'(\theta) = \mathbb{E}_\theta s(Y)$ and $\alpha''(\theta)$ is symmetric positive definite for all $\theta \in \Theta$. Let $s_0 := \mathbb{E}_{\theta_0} s(Y)$. By the strong law of large numbers, for all $\theta \in \Theta$,

$$\frac{L_n(\theta)}{n} \xrightarrow{n\to\infty} -\alpha(\theta) + \theta^T s_0 =: f(\theta), \quad \text{almost surely.}$$

Note that due to the almost sure convergence of the sufficient statistics, we actually have a stronger statement. Indeed, it holds that, almost surely, for all $\theta \in \Theta$, $\frac{L_n(\theta)}{n} \xrightarrow{n\to\infty} f(\theta)$.

Now, note that $f'(\theta_0) = 0$. Note also that the MLE $\hat{\theta}_n$ satisfying $L_n'(\hat{\theta}_n) = 0$ is unique (if it exists) because L_n is strictly concave. By Miller (2021, Theorems 12 and 5), we have that the MLE $\hat{\theta}_n$ almost surely exists and

$$\hat{\theta}_n \to \theta_0 \text{ almost surely} \quad \text{and} \quad \frac{L_n(\hat{\theta}_n)}{n} \xrightarrow{n\to\infty} f(\theta_0) \text{ almost surely.} \tag{C.1}$$

Let $\delta > 0$ and $\bar{E} := \{\theta : \|\theta - \theta_0\| \le 2\delta\}$ be such that $\bar{E} \subseteq \Theta$. Then α''' is bounded on $\bar{E}$, since α''' is continuous and $\bar{E}$ is compact. Hence, $\frac{L_n'''}{n}$ is uniformly bounded on $\bar{E}$ because $L_n'''(\theta) = -n\alpha'''(\theta)$. Therefore, almost surely, Assumption 1 is satisfied for large enough n and small enough δ with M_2 not depending on n. This is because $\hat{\theta}_n \to \theta_0$ almost surely and so, almost surely, for large enough n, $\{\theta : \|\theta - \hat{\theta}_n\| \le \delta\} \subseteq \bar{E}$. Now, let ξ_θ be the point on the line connecting θ and θ_0 that lies on $\{t : \|t - \theta_0\| = \delta/2\}$. The strict concavity of L_n implies that, almost surely,

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{\theta:\|\theta-\hat{\theta}_n\|>\delta} \frac{L_n(\theta) - L_n(\hat{\theta}_n)}{n} \\
\le &\limsup_{n\to\infty} \sup_{\theta:\|\theta-\theta_0\|>\delta/2} \frac{L_n(\theta) - L_n(\theta_0)}{n} \\
\le &\limsup_{n\to\infty} \sup_{\theta:\|\theta-\theta_0\|>\delta/2} \frac{\|\theta_0 - \theta\|}{\delta/2} \cdot \frac{L_n(\xi_\theta) - L_n(\theta_0)}{n} \\
\le &\limsup_{n\to\infty} \sup_{t:\|t-\theta_0\|=\delta/2} \frac{L_n(t) - L_n(\theta_0)}{n} \\
= &\limsup_{n\to\infty} \sup_{t:\|t-\theta_0\|=\delta/2} \left(-\alpha(t) + \frac{t^T S_n}{n}\right) - f(\theta_0) \\
\overset{(*)}{=} &\sup_{t:\|t-\theta_0\|=\delta/2} f(t) - f(\theta_0) \overset{(**)}{<} 0.
\end{aligned} \tag{C.2}$$

Equality $(*)$ follows from the fact that the sequence of functions $\left(t \mapsto \frac{t^T S_n}{n}\right)_n$ almost surely converges to $\left(t \mapsto t^T s_0\right)$ uniformly on the set $\{t : \|t - \theta_0\| = \delta/2\}$. Now, we note that the function $t \mapsto f(t) - f(\theta_0)$ is continuous and negative on the set $\{t : \|t-\theta_0\| = \delta/2\}$, by Miller (2021, Lemma 27 (1)). Therefore, its supremum on the compact set $\{t : \|t - \theta_0\| = \delta/2\}$ is negative, which justifies the last inequality $(**)$. Therefore almost surely Assumption 9 is satisfied for large enough n, for a κ not depending on n.

Note that the above considerations are valid for any fixed $\delta > 0$, such that $\bar{E} \subseteq \Theta$. Since α'' is continuous on Θ and $\alpha''(\theta)$ is symmetric, positive definite for all θ and $\hat{\theta}_n \to \theta_0$, almost surely, Assumption 8 is also satisfied for n large enough and for $\delta > 0$ small enough, almost surely. Moreover, note that $\mathrm{Tr}\left(\hat{J}_n(\hat{\theta}_n)^{-1}\right) = \mathrm{Tr}\left(\alpha''(\hat{\theta}_n)^{-1}\right) \leq \frac{d}{\lambda_{\min}(\hat{\theta}_n)}$. Since α'' is continuous, $\alpha''(\theta)$ is positive definite for all $\theta \in \Theta$ and $\hat{\theta}_n \to \theta_0$ almost surely, we have that almost surely for large n, $\frac{d}{\lambda_{\min}(\hat{\theta}_n)}$ is uniformly bounded from above. Therefore, Assumption 7 is satisfied for large enough n, almost surely.

Furthermore, Assumption 2 is satisfied immediately for large n (with $\hat{M}_1$ independent of n) if the prior density π is continuous and positive in a neighborhood around θ_0 and if $\hat{\theta}_n \to \theta_0$, which holds almost surely (Eq. (C.1)). Similarly Assumption 10 is immediately satisfied for large n (with M_1 and $\widetilde{M}_1$ independent of n) if, additionally, π is continuously differentiable in a neighborhood of θ_0.

Now, assume that π is thrice continuously differentiable on Θ. Then $\overline{L}_n$ is upper-bounded and has at least one global maximum $\theta_n^* \in \Theta$. Using Eq. (C.2), we note that $\|\theta_n^* - \hat{\theta}_n\| \xrightarrow{n\to\infty} 0$ almost surely (as π is uniformly upper bounded on Θ). We also note that, almost surely, for any fixed neighborhood of θ_0 and for sufficiently large n, $-\frac{\overline{L}_n''(\theta)}{n} = \alpha''(\theta) - \frac{(\log\pi)''(\theta)}{n}$ is strictly positive definite for θ in that neighborhood. This means that, almost surely, for sufficiently large n, the MAP $\bar{\theta}_n$ is unique. Moreover, $\|\bar{\theta}_n - \hat{\theta}_n\| \xrightarrow{n\to\infty} 0$ almost surely,which implies that $\bar{\theta}_n \to \theta_0$ almost surely, by Eq. (C.1). Moreover, note that

$$\frac{\overline{L}_n'''(\theta)}{n} = -\alpha'''(\theta) + \frac{(\log\pi)'''(\theta)}{n}.$$

Note that $(\log\pi)'''$ is continuous in a neighbourhood of θ_0. Then, almost surely, for large enough n and small enough $\bar{\delta}$, $\frac{\overline{L}_n'''(\theta)}{n}$ is uniformly bounded inside $\{\theta : \|\theta - \bar{\theta}_n\| \leq \bar{\delta}\}$ (which follows from the fact that Assumption 1 is satisfied). Thus, Assumption 3 is satisfied for large enough n and small enough $\bar{\delta}$, almost surely. Using the same assumption that $(\log\pi)'''$ is continuous in a neighbourhood of θ_0, we have that $\alpha'' - \frac{(\log\pi)''}{n}$ is continuous in a neighbourhood of θ_0 and for large enough n, $\alpha''(\theta) - \frac{(\log\pi)''(\theta)}{n}$ is symmetric, positive definite for all θ in a (potentially smaller) neighborhood of θ_0. Moreover, as we showed above, $\bar{\theta}_n \to \theta_0$, almost surely. Therefore, Assumption 5 is satisfied for n large enough and for $\bar{\delta} > 0$ small enough, almost surely.

Now, keeping the same assumptions, note that $\max\left\{\|\hat{\theta}_n - \bar{\theta}_n\|, \sqrt{\frac{\mathrm{Tr}\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}{n}}\right\} < \bar{\delta}$ for large enough n and small enough $\bar{\delta}$, almost surely. This is because if $\overline{\lambda}_{\min}(\bar{\theta}_n) > 0$ then $\mathrm{Tr}\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right) \leq \frac{d}{\overline{\lambda}_{\min}(\bar{\theta}_n)}$, which is uniformly bounded from above for large enough n. Note also that $\sqrt{\frac{\mathrm{Tr}\left[\left(\hat{J}_n(\hat{\theta}_n) + \frac{\delta M_2}{3} I_{d\times d}\right)^{-1}\right]}{n}} < \delta$ follows from Assumption 7, which holds for large n and small $\delta > 0$, almost surely, as we showed above. Therefore, Assumption 4 is satisfied for large enough n and small enough δ and $\bar{\delta}$, almost surely. Moreover, as $\bar{\theta}_n \to \theta_0$ almost surely, Assumption 6 is satisfied for large enough n and small enough $\bar{\delta}$, almost surely, by an argument analogous to Eq. (C.2).■

C.2 Proof of Proposition 15

The proof is inspired by the proof of Miller (2021, Theorem 13). Note that, for all $\theta \in \Theta$, $L_n(\theta) = -\sum_{i=1}^n \alpha(\theta^T X_i) + \theta^T S_n$, where $S_n = \sum_{i=1}^n X_i s(Y_i)$. Thus, L_n is C^∞ on Θ by the chain rule, since α is C^∞ on $\mathcal{E}$ by Miller and Harrison (2014, Proposition 19). Also, L_n is strictly concave since α is strictly convex (Miller and Harrison, 2014, Proposition 19). Now, note that, for all $\theta \in \Theta$, $\frac{L_n(\theta)}{n} \xrightarrow{n\to\infty} f(\theta)$ almost surely (by our assumed condition 2). This implies that, almost surely, $\frac{L_n(\theta)}{n} \xrightarrow{n\to\infty} f(\theta)$, for all $\theta \in \Theta$. In order to show this implication, let us fix a countable dense subset C of Θ. Then, almost surely, $\frac{L_n(\theta)}{n} \xrightarrow{n\to\infty} f(\theta)$ for all $\theta \in C$. Since $\frac{L_n}{n}$ is concave, it follows from Rockafellar (1970, Theorem 10.8) that, almost surely, the limit $\tilde{f}(\theta) := \lim_n \frac{L_n(\theta)}{n}$ exists and is finite for all $\theta \in \Theta$ and $\tilde{f}$ is concave. As f is also concave, then $\tilde{f}$ and f are continuous functions (Rockafellar, 1970, Theorem 10.1) that agree on a dense subset of Θ so they are equal on Θ.

Now, note that, by Miller (2021, Theorem 5 and Theorem 13), almost surely, there exists MLE $\hat{\theta}_n$, such that $L_n'(\hat{\theta}_n) = 0$ and

$$\hat{\theta}_n \xrightarrow{n\to\infty} \theta_0, \quad \text{almost surely.} \tag{C.3}$$

As L_n is strictly concave, this MLE is almost surely unique.

We will show that, with probability 1, $\frac{L_n'''}{n}$ is uniformly bounded on $\bar{E} = \{\theta : \|\theta - \theta_0\| \leq \epsilon\}$, where $\epsilon > 0$ satisfies condition 4 of Proposition 15. Fix $j, k, l \in \{1, \ldots, d\}$ and define $T(\theta, x) = \alpha'''(\theta^T x) x_j x_k x_l$ for $\theta \in \Theta$, $x \in \mathcal{X}$. For all $x \in \mathcal{X}$, $\theta \mapsto T(\theta, x)$ is continuous and for all $\theta \in \Theta$, $x \mapsto T(\theta, x)$ is measurable. Since $L_n'''(\theta)_{j,k,l} = -\sum_{i=1}^n T(\theta, X_i)$, condition 4 above implies that with probability 1, $\frac{L_n'''(\theta)_{j,k,l}}{n}$ is uniformly bounded on $\bar{E}$, by the uniform law of large numbers (Ghosh and Ramamoorthi, 2003, Theorem 1.3.3). Letting $C_{j,k,l}(X_1, X_2, \ldots)$ be such a uniform bound for each j, k, l, we have that with probability 1, for all $n \in \mathbb{N}$, $\theta \in \bar{E}$, $\left\|\frac{L_n'''(\theta)}{n}\right\|^2 = \sum_{j,k,l} \frac{L_n'''(\theta)_{j,k,l}^2}{n^2} \leq \sum_{j,k,l} C_{j,k,l}(X_1, X_2, \ldots)^2 < \infty$. Hence $\frac{L_n'''}{n}$ is uniformly bounded on $\bar{E}$. Therefore, almost surely, for large enough n and small enough δ, Assumption 1 is satisfied with a constant M_2 not depending on n. This is because $\hat{\theta}_n \to \theta_0$ almost surely and so, almost surely, for large enough n and $\delta = \epsilon/2$, $\{\theta : \|\theta - \hat{\theta}_n\| \leq \delta\} \subseteq \bar{E}$.

Now, using Miller (2021, Theorem 7),

$$f''(\theta_0) \overset{a.s.}{=} \lim_{n\to\infty} \frac{L_n''(\theta_0)}{n} = \lim_{n\to\infty} \left(-\frac{1}{n}\sum_{i=1}^n \alpha''(\theta_0^T X_i) X_i X_i^T\right). \tag{C.4}$$

This limit exists and is finite almost surely. Therefore, by the strong law of large numbers,

$$f''(\theta_0) \overset{a.s.}{=} -\mathbb{E}\left(\alpha''(\theta_0^T X_i) X_i X_i^T\right).$$

Therefore, $-f''(\theta_0)$ is positive definite almost surely because $\alpha''(\eta)$ is positive definite for any η and so for any nonzero $a \in \mathbb{R}^d$, we have

$$\mathbb{E}\left[\alpha''(\theta_0^T X_i) a^T X_i X_i^T a\right] > 0$$

by our assumption that $a^T X_i$ is distinct from zero with a positive probability. Now, by Miller (2021, Theorem 7), $\frac{L_n''}{n}$ is L-equi-Lipschitz for some $L < \infty$. Therefore, we have that:

$$\left\|\frac{L_n''(\hat{\theta}_n)}{n} - f''(\theta_0)\right\| \leq L\left\|\hat{\theta}_n - \theta_0\right\| + \left\|\frac{L_n''(\theta_0)}{n} - f''(\theta_0)\right\|. \tag{C.5}$$

It follows that $\frac{L_n''(\hat{\theta}_n)}{n} \xrightarrow{n\to\infty} f''(\theta_0)$ almost surely by Eqs. (C.3) and (C.4). This means that, almost surely, for large enough n, the minimum eigenvalue of $-\frac{L_n''(\hat{\theta}_n)}{n}$ stays positive and lower bounded by a positive number, not depending on n, as $f''(\theta_0)$ is positive definite. Therefore, almost surely, Assumption 8 is satisfied for large enough n and small enough δ. Moreover, for the same reason, Assumption 7 is satisfied, for large enough n and small enough $\delta > 0$, almost surely. Indeed, note that $\mathrm{Tr}\left(\hat{J}_n(\hat{\theta}_n)^{-1}\right) = \mathrm{Tr}\left(\alpha''(\hat{\theta}_n)^{-1}\right) \leq \frac{d}{\lambda_{\min}(\hat{\theta}_n)}$. Since $\frac{L_n''(\hat{\theta}_n)}{n} \xrightarrow{n\to\infty} f''(\theta_0)$ almost surely, we have that, almost surely, for large n, $\frac{d}{\lambda_{\min}(\hat{\theta}_n)}$ is uniformly bounded from above.

Now, let ξ_θ be the point on the line connecting θ and θ_0 that lies on $\{t : \|t-\theta_0\| = \delta/2\}$. The strict concavity of L_n implies that, almost surely,

$$
\begin{aligned}
&\limsup_{n\to\infty} \sup_{\theta:\|\theta-\hat{\theta}_n\|>\delta} \frac{L_n(\theta) - L_n(\hat{\theta}_n)}{n} \\
\leq &\limsup_{n\to\infty} \sup_{\theta:\|\theta-\theta_0\|>\delta/2} \frac{L_n(\theta) - L_n(\theta_0)}{n} \\
\leq &\limsup_{n\to\infty} \sup_{\theta:\|\theta-\theta_0\|>\delta/2} \frac{\|\theta_0 - \theta\|}{\delta/2} \cdot \frac{L_n(\xi_\theta) - L_n(\theta_0)}{n} \\
\leq &\limsup_{n\to\infty} \sup_{t:\|t-\theta_0\|=\delta/2} \frac{L_n(t) - L_n(\theta_0)}{n} \\
= &\limsup_{n\to\infty} \sup_{t:\|t-\theta_0\|=\delta/2} \frac{L_n(t)}{n} - f(\theta_0) \\
\overset{(*)}{=} &\sup_{t:\|t-\theta_0\|=\delta/2} f(t) - f(\theta_0) < 0.
\end{aligned} \tag{C.6}
$$

Equality $(*)$ follows from the fact that, by Miller (2021, Theorem 7), the sequence of functions $\left(\frac{L_n}{n}\right)_n$ almost surely converges to f uniformly on the set $\{t : \|t - \theta_0\| = \delta/2\}$. Now, we note that the function $t \mapsto f(t) - f(\theta_0)$ is continuous and negative on the set $\{t : \|t - \theta_0\| = \delta/2\}$, by Miller (2021, Lemma 27 (1)). Therefore, its supremum on the compact set $\{t : \|t - \theta_0\| = \delta/2\}$ is negative, which justifies the last inequality. Therefore, almost surely, Assumption 9 is satisfied, for large enough n, small enough δ and for some $\kappa > 0$ not depending on n.

Furthermore, Assumption 2 is satisfied immediately, almost surely for large n and small enough δ, if the prior density π is continuous and positive in a neighborhood around θ_0 since $\hat{\theta}_n \to \theta_0$ almost surely (Eq. (C.3)). Similarly Assumption 10 is immediately satisfied almost surely if, additionally, π is continuously differentiable in a neighborhood of θ_0.

Now, assume that π is thrice continuously differentiable on Θ. Then $\overline{L}_n$ is upper-bounded and has at least one global maximum $\theta_n^* \in \Theta$. Using Eq. (C.6), we note that $\|\theta_n^* - \hat{\theta}_n\| \xrightarrow{n\to\infty} 0$ almost surely (as π is uniformly upper bounded on Θ). We also note that, almost surely, for any fixed neighborhood of θ_0 and for sufficiently large n, $-\frac{\overline{L}_n''(\theta)}{n}$ is strictly positive definite for θ in that neighborhood. This means that, almost surely, for sufficiently large n, the MAP $\bar{\theta}_n$ is unique. Moreover, $\|\bar{\theta}_n - \hat{\theta}_n\| \xrightarrow{n\to\infty} 0$ almost surely,which implies that $\bar{\theta}_n \to \theta_0$ almost surely. Moreover, note that

$$\frac{\overline{L}_n'''(\theta)}{n} = \frac{L_n'''(\theta)}{n} + \frac{(\log \pi)'''(\theta)}{n}.$$

Hence, if $(\log \pi)'''$ is continuous in a neighbourhood of θ_0 then, almost surely, for large enough n and small enough $\bar{\delta}$, $\frac{\overline{L}_n'''(\theta)}{n}$ is uniformly bounded inside $\{\theta : \|\theta - \bar{\theta}_n\| \le \bar{\delta}\}$ (which follows from the fact that Assumption 1 holds almost surely and that $\|\hat{\theta}_n - \bar{\theta}_n\| \to 0$ almost surely). Thus, almost surely, Assumption 3 is satisfied for large enough n and small anough $\bar{\delta}$ with $\overline{M}_2$ independent of n.

Now, as in Eq. (C.5), we have that

$$\left\| \frac{L_n''(\bar{\theta}_n)}{n} - f''(\theta_0) \right\| \le L \left\| \bar{\theta}_n - \theta_0 \right\| + \left\| \frac{L_n''(\theta_0)}{n} - f''(\theta_0) \right\|$$

and so $\frac{L_n''(\bar{\theta}_n)}{n} \xrightarrow{n\to\infty} f''(\theta_0)$ almost surely. As a result, almost surely, for large enough n, the minimum eigenvalue of $-\frac{L_n''(\bar{\theta}_n)}{n}$ stays lower bounded by a positive number not depending on n. It follows that, if $(\log \pi)'''$ is continuous in a neighbourhood of θ_0, then, almost surely, for large n, the minimum eigenvalue of $-\frac{L_n''(\bar{\theta}_n)}{n} - \frac{(\log \pi)''(\bar{\theta}_n)}{n}$ stays lower bounded by a positive number not depending on n. Therefore, almost surely, Assumption 5 is also satisfied, for n large enough and for $\bar{\delta} > 0$ small enough.

Still assuming that $(\log \pi)'''$ is continuous in a neighborhood of θ_0, note that, almost surely, $\max\left\{ \|\hat{\theta}_n - \bar{\theta}_n\|, \sqrt{\frac{\mathrm{Tr}\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}{n}} \right\} < \bar{\delta}$ for large enough n and small enough $\bar{\delta}$. This is because $\mathrm{Tr}\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right) \le \frac{d}{\bar{\lambda}_{\min}(\bar{\theta}_n)}$ and $\|\hat{\theta}_n - \bar{\theta}_n\| \to 0$ almost surely. Moreover, note that $\sqrt{\frac{\mathrm{Tr}\left[\left(\hat{J}_n(\hat{\theta}_n) + \frac{\delta M_2}{3} I_{d\times d}\right)^{-1}\right]}{n}} < \delta$ if Assumption 7 is satisfied. Therefore, almost surely, Assumption 4 is satisfied for large enough n and small enough δ and $\bar{\delta}$. Furthermore, as $\bar{\theta}_n \to \theta_0$ almost surely then also almost surely Assumption 6 is satisfied for large enough n and small enough $\bar{\delta}$, by an argument analogous to Eq. (C.6).■

Appendix D. Introductory arguments for the proofs of Theorems 17 – 20, 22 – 25 and 27

D.1 Introduction

In order to prove the main results of this paper, we let $g : \mathbb{R}^d \to \mathbb{R}$ and consider:

$$D_g^{MLE} := \left| \mathbb{E}\left[g\left(\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n) \right) \right] - \mathbb{E}_{Z_n \sim \mathcal{N}(0, \hat{J}_n(\hat{\theta}_n)^{-1})}\left[g(Z_n) \right] \right|; \tag{D.1}$$

$$D_g^{MAP} := \left| \mathbb{E}\left[g\left(\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n) \right) \right] - \mathbb{E}_{\bar{Z}_n \sim \mathcal{N}(0, \bar{J}_n(\bar{\theta}_n)^{-1})}\left[g(\bar{Z}_n) \right] \right|. \tag{D.2}$$

The following lemma will be useful in the sequel:

Lemma 29 *Let $Z \sim \mathcal{N}(0, \Sigma)$. Then, for any $t > 0$,*

$$\mathbb{P}\left[\|Z\| - \sqrt{\mathrm{Tr}(\Sigma)} \ge \sqrt{2 \|\Sigma\|_{op}\, t} \right] \le e^{-t}; \tag{D.3}$$

$$\mathbb{P}\left[\|Z\|^2 - \mathrm{Tr}(\Sigma) \geq 2\sqrt{\mathrm{Tr}\left(\Sigma^2\right) t} + 2\left\|\Sigma\right\|_{op} t\right] \leq e^{-t}. \tag{D.4}$$

Proof The proof follows an argument similar to the one of Vershynin (2018, page 135). Specifically, Eq. (D.4) comes from Hsu et al. (2012, Proposition 1). In order to prove Eq. (D.3), we note that

$$\begin{aligned}\mathbb{P}\left[\|Z\| - \sqrt{\mathrm{Tr}(\Sigma)} \geq \sqrt{2\left\|\Sigma\right\|_{op} t}\right] =& \mathbb{P}\left[\|Z\|^2 - \mathrm{Tr}(\Sigma) \geq 2\left\|\Sigma\right\|_{op} t + 2\sqrt{2\,\mathrm{Tr}\left(\Sigma\right)\left\|\Sigma\right\|_{op} t}\right] \\ \leq& \mathbb{P}\left[\|Z\|^2 - \mathrm{Tr}(\Sigma) \geq 2\sqrt{\mathrm{Tr}\left(\Sigma^2\right) t} + 2\left\|\Sigma\right\|_{op} t\right] \leq e^{-t}.\end{aligned}$$

■

D.2 Initial decomposition of the distances D_g^{MLE} and D_g^{MAP}

Now, let

$$\begin{aligned} h_g^{MLE}(u) &:= g(u) - \frac{n^{-d/2}}{C_n^{MLE}} \int_{\|t\|\leq\delta\sqrt{n}} g(t)\Pi_n(n^{-1/2}t + \hat{\theta}_n)dt, \\ h_g^{MAP}(u) &:= g(u) - \frac{n^{-d/2}}{C_n^{MAP}} \int_{\|t\|\leq\bar{\delta}\sqrt{n}} g(t)\Pi_n(n^{-1/2}t + \bar{\theta}_n)dt, \end{aligned}$$

for

$$\begin{aligned} C_n^{MLE} &:= n^{-d/2} \int_{\|u\|\leq\delta\sqrt{n}} \Pi_n(n^{-1/2}u + \hat{\theta}_n)du \\ C_n^{MAP} &:= n^{-d/2} \int_{\|u\|\leq\bar{\delta}\sqrt{n}} \Pi_n(n^{-1/2}u + \bar{\theta}_n)du. \end{aligned} \tag{D.5}$$

Note that, for

$$F_n^{MLE} := \int_{\|u\|\leq\delta\sqrt{n}} \frac{\sqrt{\left|\det\left(\hat{J}_n(\hat{\theta}_n)\right)\right|} e^{-u^T \hat{J}_n(\hat{\theta}_n)u/2}}{(2\pi)^{d/2}} du, \tag{D.6}$$

we have

$$\begin{aligned} &D_g^{MLE} \\ &= \left| \int_{\mathbb{R}^d} h_g^{MLE}(u) \frac{\sqrt{\left|\det\left(\hat{J}_n(\hat{\theta}_n)\right)\right|} e^{-u^T \hat{J}_n(\hat{\theta}_n)u/2}}{(2\pi)^{d/2}} du \right. \\ &\qquad\qquad \left. - n^{-d/2} \int_{\|u\|>\delta\sqrt{n}} h_g^{MLE}(u)\Pi_n(n^{-1/2}u + \hat{\theta}_n)du \right| \\ &\leq \frac{1}{F_n^{MLE}} \left| \int_{\|u\|\leq\delta\sqrt{n}} h_g^{MLE}(u) \frac{\sqrt{\left|\det\left(\hat{J}_n(\hat{\theta}_n)\right)\right|} e^{-u^T \hat{J}_n(\hat{\theta}_n)u/2}}{(2\pi)^{d/2}} du \right| \end{aligned}$$

$$
+ \left| \int_{\|u\| > \delta\sqrt{n}} h_g^{MLE}(u) \left[\frac{\sqrt{\left|\det\left(\hat{J}_n(\hat{\theta}_n)\right)\right|} e^{-u^T \hat{J}_n(\hat{\theta}_n) u/2}}{(2\pi)^{d/2}} - n^{-d/2}\, \Pi_n(n^{-1/2}u + \hat{\theta}_n) \right] du \right|
$$

$$
= \left| \int_{\|u\| \leq \delta\sqrt{n}} g(u) \frac{\sqrt{\left|\det\left(\hat{J}_n(\hat{\theta}_n)\right)\right|} e^{-u^T \hat{J}_n(\hat{\theta}_n) u/2}}{F_n^{MLE}(2\pi)^{d/2}} du \right.
$$

$$
\left. - \frac{n^{-d/2}}{C_n^{MLE}} \int_{\|u\| \leq \delta\sqrt{n}} g(u) \Pi_n(n^{-1/2}u + \hat{\theta}_n) du \right|
$$

$$
+ \left| \int_{\|u\| > \delta\sqrt{n}} h_g^{MLE}(u) \left[\frac{\sqrt{\left|\det\left(\hat{J}_n(\hat{\theta}_n)\right)\right|} e^{-u^T \hat{J}_n(\hat{\theta}_n) u/2}}{(2\pi)^{d/2}} - n^{-d/2}\, \Pi_n(n^{-1/2}u + \hat{\theta}_n) \right] du \right|
$$

$$
=: I_1^{MLE} + I_2^{MLE}. \tag{D.7}
$$

In a similar manner, for

$$
F_n^{MAP} := \int_{\|u\| \leq \bar{\delta}\sqrt{n}} \frac{\sqrt{\left|\det\left(\bar{J}_n(\bar{\theta}_n)\right)\right|} e^{-u^T \bar{J}_n(\bar{\theta}_n) u/2}}{(2\pi)^{d/2}} du, \tag{D.8}
$$

we have

$$
D_g^{MAP}
$$

$$
\leq \left| \int_{\|u\| \leq \bar{\delta}\sqrt{n}} g(u) \frac{\sqrt{\left|\det\left(\bar{J}_n(\bar{\theta}_n)\right)\right|} e^{-u^T \bar{J}_n(\bar{\theta}_n) u/2}}{F_n^{MAP}(2\pi)^{d/2}} du \right.
$$

$$
\left. - \frac{n^{-d/2}}{C_n^{MAP}} \int_{\|u\| \leq \bar{\delta}\sqrt{n}} g(u) \Pi_n(n^{-1/2}u + \bar{\theta}_n) du \right|
$$

$$
+ \left| \int_{\|u\| > \bar{\delta}\sqrt{n}} h_g^{MAP}(u) \left[\frac{\sqrt{\left|\det\left(\bar{J}_n(\bar{\theta}_n)\right)\right|} e^{-u^T \bar{J}_n(\bar{\theta}_n) u/2}}{(2\pi)^{d/2}} - n^{-d/2}\, \Pi_n(n^{-1/2}u + \bar{\theta}_n) \right] du \right|
$$

$$
=: I_1^{MAP} + I_2^{MAP}. \tag{D.9}
$$

D.3 Strategies for controlling terms I_1^{MLE} and I_1^{MAP}

In order to control I_1^{MAP} in Theorems 17, 18 and 20 and I_1^{MLE} in Theorems 22, 23 and 25, we apply the log-Sobolev inequality and the associated transportation inequalities. Before we introduce those concepts, we define the following Kullback-Leibler divergence (or relative entropy, see e.g. Bishop 2006, Section 1.6.1) between two measures ν and μ such that $\nu \ll \mu$:

$$
\mathrm{KL}(\nu\|\mu) = \int \log\left(\frac{d\nu}{d\mu}(x)\right) \nu(dx). \tag{D.10}
$$

We also define the following Fisher divergence between two measures ν and μ such that $\nu \ll \mu$

$$\text{Fisher}(\nu\|\mu) = \int \left\| \left(\log \left(\frac{d\nu}{d\mu} \right) \right)' (x) \right\|^2 \nu(dx). \tag{D.11}$$

The following definition, which can be found in Bakry et al. (2016, Chapter 5) and Vempala and Wibisono (2019, Section 2.2), will play a central role in our proofs:

Definition 30 (Log-Sobolev inequality (LSI)) *Let μ be a probability measure on $\Omega \subset \mathbb{R}^d$ and let $\alpha > 0$. We say that μ satisfies the log-Sobolev inequality with constant α, or LSI(α) for short, if*

$$\int f^2 \log f^2 d\mu - \int f^2 d\mu \log \left(\int f^2 d\mu \right) \leq \frac{2}{\alpha} \int \|f'\|^2 d\mu, \quad \text{for all } f : \Omega \to \mathbb{R}^+. \tag{D.12}$$

Equivalently, we say that μ satisfies LSI(α) if for any probability measure ν on Ω, such that $\mu \ll \nu$ and $\nu \ll \mu$,

$$KL(\nu\|\mu) \leq \frac{1}{2\alpha} Fisher(\nu\|\mu). \tag{D.13}$$

Remark 31 *The equivalence between the two definitions of the log-Sobolev inequality LSI(α), given by Eqs.* (D.12) *and* (D.13)*, follows by the following argument. If ν is a probability measure on Ω such that μ and ν are equivalent then we obtain Eq.* (D.13) *by setting $f^2 = \frac{d\nu}{d\mu}$ in Eq.* (D.12)*. On the other hand, for a function $f : \Omega \to \mathbb{R}^+$, let ν be a probability measure on Ω with density $\nu(dx) \propto f^2(x)\mu(dx)$. Plugging this ν into Eq.* (D.13) *yields Eq.* (D.12).

One of the fundamental results we will use is the *Bakry-Émery* criterion. Its first general version was proved in Bakry and Émery (1985). Below we state its version for measures on $\mathbb{R}^d$ truncated to convex sets (see Kolesnikov and Milman 2016, Theorem 2.1 and Schlichting 2019, Theorem A1):

Proposition 32 (Bakry-Émery criterion) *Let $\Omega \subset \mathbb{R}^d$ be convex, let $H : \Omega \to \mathbb{R}$ and consider a probability measure $\mu(dx) \propto e^{-H(x)} \mathbb{1}_\Omega(x) dx$. Assume that $H''(x) \succeq \alpha I_{d\times d}$, for some $\alpha > 0$. Then μ satisfies LSI(α) (see Definition Theorem 30).*

The following Holley-Stroock perturbation principle (see Holley and Stroock 1987, page 1184, Schlichting 2019, Theorem A2) will be crucial in our proofs concerning the MLE-centric approach:

Proposition 33 (Holley-Stroock perturbation principle) *Let $\Omega \subset \mathbb{R}^d$ and $H : \Omega \to \mathbb{R}$. Let $\psi : \Omega \to \mathbb{R}^d$ be a bounded function. Let μ and $\tilde{\mu}$ be probability measures with densities of the form $\mu(dx) \propto e^{-H(x)} \mathbb{1}_\Omega(x) dx$ and $\tilde{\mu}(dx) \propto e^{-H(x)-\psi(x)} \mathbb{1}_\Omega(x) dx$. Suppose that μ satisfies the LSI(α) (see Definition Theorem 30). Then $\tilde{\mu}$ satisfies LSI($e^{-osc\psi}\alpha$), where $osc\psi := \sup_\Omega \psi - \inf_\Omega \psi$.*

In order to control I_1^{MAP} and I_1^{MLE} in our proofs, we control the Fisher divergence between the rescaled posterior and the Gaussian inside the ball around the MAP or the MLE, establish the log-Sobolev inequality inside that ball for the rescaled posterior and thus control the KL divergence. In order to obtain our bounds on the total variation distance in Theorems 17 and 22, we will use the following result:

Proposition 34 (Pinsker's inequality, e.g. Massart 2007, Theorem 2.16) *For any two probability measures μ and ν such that $\nu \ll \mu$, we have that*

$$TV(\mu, \nu) \leq \sqrt{\frac{1}{2} KL(\nu \| \mu)}.$$

In order to control the 1-Wasserstein distance in Theorems 18 and 23 and the integral probability metric appearing in Theorems 20 and 25, we first control I_1^{MAP} or I_1^{MLE} by controlling the 2-Wasserstein distance inside the appropriate neighborhood of the MAP or the MLE. The 2-Wasserstein distance for two measures μ and ν on the same measurable space is defined in the following way:

$$W_2(\nu, \mu) = \inf_{\Gamma} \sqrt{\mathbb{E}_{\Gamma}\left[\|X - Y\|^2\right]} \tag{D.14}$$

where the infimum is over all distributions Γ of (X, Y) with the correct marginals $X \sim \nu$, $Y \sim \mu$. It is an easy consequence of Jensen's inequality that for any probability measures μ, ν,

$$W_1(\mu, \nu) \leq W_2(\mu, \nu). \tag{D.15}$$

We also have the following Talagrand inequality:

Proposition 35 *Let μ be a probability measure on a closed ball $\{u : \|u\| \leq \eta\}$ for some $\eta > 0$ such that μ is absolutely continuous with respect to the Lebesgue measure. Suppose that μ satisfies LSI(α) on $\{u : \|u\| \leq \eta\}$. Then, for any probability measure $\nu \ll \mu$ on $\{u : \|u\| \leq \eta\}$, we have that*

$$W_2(\mu, \nu)^2 \leq \frac{2}{\alpha} KL(\nu \| \mu).$$

Proof This result follows from Gozlan (2009, Theorem 4.1) by an argument analogous to the one that let the authors prove the result in Gozlan (2009, Corollary 4.2). Indeed, for a natural number $n \geq 1$ and $x = (x_1, \ldots, x_n) \in \left(\mathbb{R}^d\right)^n$, consider $L_n^x = n^{-1} \sum_{i=1}^n \delta_{x_i}$, where δ_{x_i}'s are Dirac deltas. In the proof of Gozlan (2009, Corollary 4.2) it is shown that the function $x \mapsto W_2(L_n^x, \mu)$ is $\frac{1}{\sqrt{n}}$-Lipschitz for the Euclidean distance. When restricted to the set $\{u : \|u\| \leq \eta\}^n \subset \left(\mathbb{R}^d\right)^n$, this function is still $\frac{1}{\sqrt{n}}$-Lipschitz. As in the proof of Gozlan (2009, Corollary 4.2), we note that, by Rademacher's theorem the function $F_n(x) = W_2(L_n^x, \mu)$ is differentiable Lebesgue-almost everywhere on $\left(\mathbb{R}^d\right)^n$. When restricted to $\{u : \|u\| \leq \eta\}^n \subset \left(\mathbb{R}^d\right)^n$ this function stays differentiable Lebesgue-almost everywhere. Therefore, the condition

$$\sum_{i=1}^n \|\nabla_i F_n\|^2 (x) \leq \frac{1}{n} \quad \text{for } \mu^n\text{-almost every } x \in \{u : \|u\| \leq \eta\}^n \subset \left(\mathbb{R}^d\right)^n$$

is fulfilled, as required by Gozlan (2009, Theorem 4.1). ■

Proposition 35 together with Eq. (D.15) and the log-Sobolev inequality will let us control I_1^{MAP} and I_1^{MLE} in the proofs of Theorems 18, 20, 23 and 25.

D.4 Controlling term I_2^{MLE}

Note that

$$
\begin{aligned}
I_2^{MLE} \le & \left| \int_{\|u\|>\delta\sqrt{n}} g(u) \left(n^{-d/2}\Pi_n(n^{-1/2}u+\hat{\theta}_n) - \frac{\sqrt{\left|\det\left(\hat{J}_n(\hat{\theta}_n)\right)\right|} e^{-u^T \hat{J}_n(\hat{\theta}_n)u/2}}{(2\pi)^{d/2}} \right) du \right| \\
& + \frac{n^{-d/2}}{C_n^{MLE}} \int_{\|t\|\le\delta\sqrt{n}} |g(t)| \Pi_n(n^{-1/2}t+\hat{\theta}_n)dt \\
& \quad \cdot \left| \int_{\|u\|>\delta\sqrt{n}} \left(n^{-d/2}\Pi_n(n^{-1/2}u+\hat{\theta}_n) - \frac{\sqrt{\left|\det\left(\hat{J}_n(\hat{\theta}_n)\right)\right|} e^{-u^T \hat{J}_n(\hat{\theta}_n)u/2}}{(2\pi)^{d/2}} \right) du \right| \\
=: & I_{2,1}^{MLE} + I_{2,2}^{MLE}.
\end{aligned}
$$

Now, note that $L_n'(\hat{\theta}_n) = 0$. Therefore, for $\|t\| \le \delta\sqrt{n}$, Assumption 1 implies that

$$
\left| L_n(n^{-1/2}t+\hat{\theta}_n) - L_n(\hat{\theta}_n) + \frac{1}{2} t^T \hat{J}_n(\hat{\theta}_n) t \right| \le \frac{1}{6} n^{-1/2} M_2 \|t\|^3 \le \frac{\delta M_2}{6} \|t\|^2. \tag{D.16}
$$

Therefore, under Assumptions 1, 2 and 9, and using the notation of Section 2.5,

$$
\begin{aligned}
& \int_{\|u\|>\delta\sqrt{n}} |g(u)| n^{-d/2} \Pi_n(n^{-1/2}u+\hat{\theta}_n) du \\
= & \frac{\int_{\|u\|>\delta\sqrt{n}} |g(u)| \pi(n^{-1/2}u+\hat{\theta}_n) e^{L_n(n^{-1/2}u+\hat{\theta}_n)} du}{\int_{\mathbb{R}^d} \pi(n^{-1/2}t+\hat{\theta}_n) e^{L_n(n^{-1/2}t+\hat{\theta}_n)} dt} \\
\le & \frac{\int_{\|u\|>\delta\sqrt{n}} |g(u)| \pi(n^{-1/2}u+\hat{\theta}_n) e^{L_n(n^{-1/2}u+\hat{\theta}_n)} du}{\int_{\|t\|\le\delta\sqrt{n}} \pi(n^{-1/2}t+\hat{\theta}_n) e^{L_n(n^{-1/2}t+\hat{\theta}_n)} dt} \\
= & \frac{\int_{\|u\|>\delta\sqrt{n}} |g(u)| \pi(n^{-1/2}u+\hat{\theta}_n) e^{L_n(n^{-1/2}u+\hat{\theta}_n) - L_n(\hat{\theta}_n)} du}{\int_{\|t\|\le\delta\sqrt{n}} \pi(n^{-1/2}t+\hat{\theta}_n) e^{L_n(n^{-1/2}t+\hat{\theta}_n) - L_n(\hat{\theta}_n)} dt} \\
\overset{Eq.\ (D.16)}{\le} & \frac{e^{-n\kappa} \int_{\|u\|>\delta\sqrt{n}} |g(u)| \pi(n^{-1/2}u+\hat{\theta}_n) du}{\int_{\|t\|\le\delta\sqrt{n}} \pi(n^{-1/2}t+\hat{\theta}_n) e^{-t^T(\hat{J}_n(\hat{\theta}_n)+(\delta M_2/3) I_{d\times d})t/2} dt} \\
\le & \frac{n^{d/2} e^{-n\kappa} \hat{M}_1 \left| \det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right) \right|^{1/2} \int_{\|u\|>\delta} |g(u\sqrt{n})| \pi(u+\hat{\theta}_n) du}{(2\pi)^{d/2} \left(1 - \hat{\mathscr{D}}^p(n,\delta)\right)},
\end{aligned} \tag{D.17}
$$

if $n > \frac{\mathrm{Tr}\left[\hat{J}_n^p(\hat{\theta}_n,\delta)^{-1}\right]}{\delta^2}$, where the last inequality follows from Lemma 29. Therefore, if $n > \frac{\mathrm{Tr}\left[\hat{J}_n^p(\hat{\theta}_n,\delta)^{-1}\right]}{\delta^2}$ and if Assumptions 1, 2 and 9 are satisfied,

$$I_{2,1}^{MLE} \leq \left| \int_{\|u\|>\delta\sqrt{n}} g(u) \frac{\sqrt{\left|\det \hat{J}_n(\hat{\theta}_n)\right|} e^{-u^T \hat{J}_n(\hat{\theta}_n)u/2}}{(2\pi)^{d/2}} du \right| + \frac{n^{d/2} e^{-n\kappa} \hat{M}_1 \left|\det\left(\bar{J}_n^p(\bar{\theta}_n,\bar{\delta})\right)\right|^{1/2} \int_{\|u\|>\delta} |g(u\sqrt{n})| \pi(u+\hat{\theta}_n) du}{(2\pi)^{d/2}\left(1-\hat{\mathscr{D}}^p(n,\delta)\right)}. \tag{D.18}$$

Now, under Assumptions 1, 2 and 7 – 10,

$$\begin{aligned}
&\frac{n^{-d/2}}{C_n^{MLE}} \int_{\|t\|\leq\delta\sqrt{n}} |g(t)| \Pi_n(n^{-1/2}t+\hat{\theta}_n) dt \\
\leq & \frac{\hat{M}_1 \widetilde{M}_1 \int_{\|t\|\leq\delta\sqrt{n}} |g(t)| e^{L_n(n^{-1/2}t+L_n)-L_n(\hat{\theta}_n)} dt}{\int_{\|u\|\leq\delta\sqrt{n}} e^{L_n(n^{-1/2}u+\hat{\theta}_n)-L_n(\hat{\theta}_n)} du} \\
\leq & \frac{\hat{M}_1 \widetilde{M}_1 \int_{\|t\|\leq\delta\sqrt{n}} |g(t)| e^{-\frac{1}{2}t^T\left(\hat{J}_n(\hat{\theta}_n)-\frac{M_2\delta}{3} I_{d\times d}\right)t} dt}{\int_{\|u\|\leq\delta\sqrt{n}} e^{-u^T(\hat{J}_n(\hat{\theta}_n)+(\delta M_2/3)I_{d\times d})u/2} du} \\
\leq & \frac{\hat{M}_1 \widetilde{M}_1 \left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2} \int_{\|t\|\leq\delta\sqrt{n}} |g(t)| e^{-\frac{1}{2}t^T \hat{J}_n^m(\hat{\theta}_n,\delta)t} dt}{(2\pi)^{d/2}\left(1-\hat{\mathscr{D}}^p(n,\delta)\right)},
\end{aligned} \tag{D.19}$$

where the last inequality follows from Lemma 29. A bound on $I_{2,2}$ can be obtained by combining Eq. (D.19) with Eq. (D.18) applied to $g = 1$. Indeed, we thus obtain:

$$I_{2,2}^{MLE} \leq \frac{\hat{M}_1 \widetilde{M}_1 \left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2} \int_{\|u\|\leq\delta\sqrt{n}} |g(u)| e^{-\frac{1}{2}u^T \hat{J}_n^m(\hat{\theta}_n,\delta)u} du}{(2\pi)^{d/2}\left(1-\hat{\mathscr{D}}^p(n,\delta)\right)} \cdot \left\{ \hat{\mathscr{D}}(n,\delta) + \frac{n^{d/2} e^{-n\kappa} \hat{M}_1 \left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2}}{(2\pi)^{d/2}\left(1-\hat{\mathscr{D}}^p(n,\delta)\right)} \right\}, \tag{D.20}$$

where we used Lemma 29. A bound on I_2^{MLE} is obtained by adding together the bounds on $I_{2,1}^{MLE}$ (Eq. (D.18)) and $I_{2,2}^{MLE}$ (Eq. (D.20)).

Remark 36 *Note that, for g, such that $|g| \leq U$, for some $U > 0$, we have that*

$$\frac{n^{-d/2}}{C_n^{MLE}} \int_{\|t\|\leq\delta\sqrt{n}} |g(t)| \Pi_n(n^{-1/2}t+\hat{\theta}_n) dt \leq U.$$

Therefore, for $|g| \le U$, the same argument as above yields a simpler bound:

$$I_{2,2}^{MLE} \le U \left\{ \hat{\mathscr{D}}(n,\delta) + \frac{n^{d/2} e^{-n\kappa} \hat{M}_1 \left| det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2}}{(2\pi)^{d/2}\left(1-\hat{\mathscr{D}}^p(n,\delta)\right)} \right\}.$$

D.5 Controlling term I_2^{MAP}

Note that

$$\begin{aligned} I_2^{MAP} \le & \left| \int_{\|u\|>\bar{\delta}\sqrt{n}} g(u) \left(n^{-d/2}\Pi_n(n^{-1/2}u+\bar{\theta}_n) - \frac{\sqrt{\left|\det \bar{J}_n(\bar{\theta}_n)\right|} e^{-u^T \bar{J}_n(\bar{\theta}_n)u/2}}{(2\pi)^{d/2}} \right) du \right| \\ & + \frac{n^{-d/2}}{C_n^{MAP}} \int_{\|t\| \le \bar{\delta}\sqrt{n}} |g(t)| \Pi_n(n^{-1/2}t+\bar{\theta}_n)dt \\ & \quad \cdot \left| \int_{\|u\|>\bar{\delta}\sqrt{n}} \left(n^{-d/2}\Pi_n(n^{-1/2}u+\bar{\theta}_n) - \frac{\sqrt{\left|\det \bar{J}_n(\bar{\theta}_n)\right|} e^{-u^T \bar{J}_n(\bar{\theta}_n)u/2}}{(2\pi)^{d/2}} \right) du \right| \\ =: & I_{2,1}^{MAP} + I_{2,2}^{MAP}. \end{aligned}$$

Note that, using Assumptions 1, 2 and 6, a calculation similar to Eq. (D.17), yields

$$\begin{aligned} & \int_{\|u\|>\bar{\delta}\sqrt{n}} |g(u)| n^{-d/2}\Pi_n(n^{-1/2}u+\bar{\theta}_n)du \\ = & \int_{\|v-\bar{\theta}_n\|>\bar{\delta}} \left| g\left(\sqrt{n}(v-\bar{\theta}_n)\right)\right| \Pi_n(v)dv \\ \le & \frac{\int_{\|v-\hat{\theta}_n\|>\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|} \left| g\left(\sqrt{n}(v-\bar{\theta}_n)\right)\right| \pi(v)e^{L_n(v)}dv}{\int_{\mathbb{R}^d} \pi(t)e^{L_n(t)}dt} \\ \le & \frac{\int_{\|v-\hat{\theta}_n\|>\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|} \left| g\left(\sqrt{n}(v-\bar{\theta}_n)\right)\right| \pi(v)e^{L_n(v)-L_n(\hat{\theta}_n)}dv}{\int_{\|t-\hat{\theta}_n\|\le\delta} \pi(t)e^{L_n(t)-L_n(\hat{\theta}_n)}dt} \\ \le & \frac{n^{d/2}e^{-n\bar{\kappa}} \int_{\|v-\hat{\theta}_n\|>\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|} \left| g\left(\sqrt{n}(v-\bar{\theta}_n)\right)\right| \pi(v)dv}{\int_{\|t\|\le\delta\sqrt{n}} \pi(n^{-1/2}t+\hat{\theta}_n)e^{-t^T(\hat{J}_n(\hat{\theta}_n)+(\delta M_2/3)I_{d\times d})t/2}dt} \\ \le & \frac{n^{d/2}e^{-n\bar{\kappa}}\hat{M}_1 \left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2} \int_{\|v-\hat{\theta}_n\|>\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|} \left| g\left(\sqrt{n}(v-\bar{\theta}_n)\right)\right| \pi(v)du}{(2\pi)^{d/2}\left(1-\hat{\mathscr{D}}^p(n,\delta)\right)}, \end{aligned}$$

if $n > \frac{\mathrm{Tr}\left[\hat{J}_n^p(\hat{\theta}_n,\delta)^{-1}\right]}{\delta^2}$, where the last inequality follows from Lemma 29. Therefore, under Assumptions 1, 2 and 6 and if $n > \frac{\mathrm{Tr}\left[\hat{J}_n^p(\hat{\theta}_n,\delta)^{-1}\right]}{\delta^2}$,

$$I_{2,1}^{MAP} \le \left| \int_{\|u\|>\bar{\delta}\sqrt{n}} g(u) \frac{\sqrt{\left|\det \bar{J}_n(\bar{\theta}_n)\right|} e^{-u^T \bar{J}_n(\bar{\theta}_n)u/2}}{(2\pi)^{d/2}} du \right|$$

$$+\frac{n^{d/2}e^{-n\bar{\kappa}}\hat{M}_1\left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2}\int_{\|v-\hat{\theta}_n\|>\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|}\left|g\left(\sqrt{n}(v-\bar{\theta}_n)\right)\right|\pi(v)du}{(2\pi)^{d/2}\left(1-\hat{\mathscr{D}}^p(n,\delta)\right)}. \tag{D.21}$$

Now, under Assumptions 1 – 6,

$$\begin{aligned}
&\frac{n^{-d/2}}{C_n^{MAP}}\int_{\|t\|\le\bar{\delta}\sqrt{n}}|g(t)|\Pi_n(n^{-1/2}t+\bar{\theta}_n)dt\\
\le&\frac{\int_{\|t\|\le\bar{\delta}\sqrt{n}}|g(t)|e^{l\overline{L}_n(n^{-1/2}t+\bar{\theta}_n)-\overline{L}_n(\bar{\theta}_n)}dt}{\int_{\|u\|\le\bar{\delta}\sqrt{n}}e^{\overline{L}_n(n^{-1/2}u+\bar{\theta}_n)-\overline{L}_n(\bar{\theta}_n)}du}\\
\le&\frac{\int_{\|t\|\le\bar{\delta}\sqrt{n}}|g(t)|e^{-\frac{1}{2}t^T\left(\bar{J}_n(\bar{\theta}_n)-\frac{\overline{M}_2\bar{\delta}}{3}I_{d\times d}\right)t}dt}{\int_{\|u\|\le\bar{\delta}\sqrt{n}}e^{-u^T(\bar{J}_n(\bar{\theta}_n)+(\bar{\delta}\overline{M}_2/3)I_{d\times d})u/2}du}\\
\le&\frac{\left|\det\left(\bar{J}_n^p(\bar{\theta}_n,\bar{\delta})\right)\right|^{1/2}\int_{\|u\|\le\bar{\delta}\sqrt{n}}|g(u)|e^{-\frac{1}{2}u^T\bar{J}_n^m(\bar{\theta}_n,\bar{\delta})u}du}{(2\pi)^{d/2}\left(1-\bar{\mathscr{D}}^p(n,\bar{\delta})\right)},
\end{aligned}\tag{D.22}$$

where the last inequality follows from Lemma 29. A bound on $I_{2,2}^{MAP}$ can be obtained by combining Eq. (D.22) with Eq. (D.21) applied to $g=1$. Indeed, we obtain:

$$\begin{aligned}
I_{2,2}^{MAP}\le&\frac{\left|\det\left(\bar{J}_n^p(\bar{\theta}_n,\bar{\delta})\right)\right|^{1/2}\int_{\|u\|\le\bar{\delta}\sqrt{n}}|g(u)|e^{-\frac{1}{2}u^T\bar{J}_n^m(\bar{\theta}_n,\bar{\delta})u}du}{(2\pi)^{d/2}\left(1-\bar{\mathscr{D}}^p(n,\bar{\delta})\right)}\\
&\cdot\left\{\bar{\mathscr{D}}(n,\bar{\delta})+\frac{n^{d/2}e^{-n\bar{\kappa}}\hat{M}_1\left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2}}{(2\pi)^{d/2}\left(1-\hat{\mathscr{D}}^p(n,\delta)\right)}\right\},
\end{aligned}\tag{D.23}$$

where we applied Lemma 29.

Remark 37 *As in Remark 36, our bound gets simpler if $|g|\le U$, for some $U>0$. In that case, instead of Eq.* (D.23)*, we can write:*

$$I_{2,2}^{MAP}\le U\left\{\bar{\mathscr{D}}(n,\bar{\delta})+\frac{n^{d/2}e^{-n\bar{\kappa}}\hat{M}_1\left|det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2}}{(2\pi)^{d/2}\left(1-\hat{\mathscr{D}}^p(n,\delta)\right)}\right\}.$$

Appendix E. Proofs of Theorems 17, 18 and 20

Throughout this section we adopt the notation of Appendix D. Additionally, for any probability measure μ, we let $[\mu]_{B_0(\bar{\delta}\sqrt{n})}$ denote its restriction (truncation) to the ball of radius $\bar{\delta}\sqrt{n}$ around 0.

In all the proofs below, we wish to control the quantity D_g^{MAP} of Eq. (D.2) for all functions g that satisfy certain prescribed criteria. In the proof of Theorem 17, we look at functions g that are indicators of measurable sets, in the proof of Theorem 18 we look at 1-Lipschitz functions g and in the proof of Theorem 20 at those that are of the form $g(x) = \langle v, x\rangle^2$ for some $v \in \mathbb{R}^d$ with $\|v\| = 1$. In order to prove Theorems 17, 18 and 20, we will bound terms I_2^{MAP} and I_1^{MAP} of Eq. (D.9) separately.

E.1 Proof of Theorem 17

E.1.1 Controlling term I_2^{MAP}

We wish to obtain a uniform bound on I_2^{MAP} for all functions g that are indicators of measurable sets. Every indicator function is upper-bounded by one, so we can use Remark 37 to obtain:

$$I_{2,2}^{MAP} \leq \bar{\mathscr{D}}(n,\bar{\delta}) + \frac{n^{d/2}e^{-n\bar{\kappa}}\hat{M}_1\left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2}}{(2\pi)^{d/2}\Big(1-\hat{\mathscr{D}}^p(n,\delta)\Big)}. \tag{E.1}$$

Similarly, since $|g| \leq 1$, we can use Eq. (D.21) and Lemma 29 to obtain:

$$I_{2,1}^{MAP} \leq \bar{\mathscr{D}}(n,\bar{\delta}) + \frac{n^{d/2}e^{-n\bar{\kappa}}\hat{M}_1\left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2}}{(2\pi)^{d/2}\Big(1-\hat{\mathscr{D}}^p(n,\delta)\Big)}. \tag{E.2}$$

From Eqs. (E.1) and (E.2), it follows that

$$I_{2}^{MAP} \leq 2\,\bar{\mathscr{D}}(n,\bar{\delta}) + \frac{2n^{d/2}e^{-n\bar{\kappa}}\hat{M}_1\left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2}}{(2\pi)^{d/2}\Big(1-\hat{\mathscr{D}}^p(n,\delta)\Big)}. \tag{E.3}$$

E.1.2 Controlling term I_1^{MAP} using the log-Sobolev inequality

Let KL $(\cdot\|\cdot)$ denote the Kullback-Leibler divergence (Eq. (D.10)) and Fisher $(\cdot\|\cdot)$ denote the Fisher divergence (Eq. (D.11)). Let F_n^{MAP} be given by Eq. (D.8).

Note that, by Assumption 3, for t such that $\|t\| < \sqrt{n}\bar{\delta}$, we have

$$-n^{-1}\overline{L}_n''(\bar{\theta}_n + n^{-1/2}t) \succeq \bar{J}_n(\bar{\theta}_n) - \overline{\delta M}_2 I_{d\times d} \succeq \left(\overline{\lambda}_{\min}(\bar{\theta}_n) - \overline{\delta M}_2\right) I_{d\times d}.$$

This means that, inside the convex set $\{t \in \mathbb{R}^d : \|t\| < \sqrt{n}\bar{\delta}\}$, the density of $\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n)$ is $\left(\overline{\lambda}_{\min}(\bar{\theta}_n) - \overline{\delta M}_2\right)$-strongly log-concave (see e.g. Saumard and Wellner 2014). Using Proposition 32 (the **Bakry-Émery criterion**), we have that $\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n\right)\right)\right]_{B_0(\bar{\delta}\sqrt{n})}$ satisfies the **log-Sobolev inequality** LSI$\left(\overline{\lambda}_{\min}(\bar{\theta}_n) - \overline{\delta M}_2\right)$ (see Definition 30). By combining the log-Sobolev inequality with Pinsker's inequality (Proposition 34) we obtain that, for all functions g, which are indicators of measurable sets,

$$\begin{aligned} I_1^{MAP} &\leq \mathrm{TV}\left(\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n\right)\right)\right]_{B_0(\bar{\delta}\sqrt{n})}, \left[\mathcal{N}(0,\bar{J}_n(\bar{\theta}_n)^{-1})\right]_{B_0(\bar{\delta}\sqrt{n})}\right) \\ &\underset{\text{inequality}}{\overset{\text{Pinsker's}}{\leq}} \sqrt{\frac{1}{2}\mathrm{KL}\left(\left[\mathcal{N}(0,\bar{J}_n(\bar{\theta}_n)^{-1})\right]_{B_0(\bar{\delta}\sqrt{n})}\middle\|\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n\right)\right)\right]_{B_0(\bar{\delta}\sqrt{n})}\right)} \end{aligned}$$

$$\overset{\text{log-Sobolev}}{\underset{\text{inequality}}{\leq}} \frac{\sqrt{\text{Fisher}\left(\left[\mathcal{N}(0, \bar{J}_n(\bar{\theta}_n)^{-1})\right]_{B_0(\bar{\delta}\sqrt{n})} \middle\| \left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n\right)\right)\right]_{B_0(\bar{\delta}\sqrt{n})}\right)}}{2\sqrt{\overline{\lambda}_{\min}(\bar{\theta}_n) - \bar{\delta}\,\overline{M}_2}}$$

$$= \frac{1}{2\sqrt{\overline{\lambda}_{\min}(\bar{\theta}_n) - \bar{\delta}\,\overline{M}_2}} \cdot \sqrt{\int_{\|u\| \leq \bar{\delta}\sqrt{n}} \frac{\sqrt{\left|\det\left(\bar{J}_n(\bar{\theta}_n)\right)\right|} e^{-u^T \bar{J}_n(\bar{\theta}_n) u/2}}{F_n^{MAP}(2\pi)^{d/2}} \left\| \bar{J}_n(\bar{\theta}_n) u + \frac{\overline{L}_n'(n^{-1/2}u + \bar{\theta}_n)}{\sqrt{n}} \right\|^2 du}$$

$$\overset{\text{Taylor's}}{\underset{\text{theorem}}{\leq}} \frac{\overline{M}_2}{4\sqrt{n\left(\overline{\lambda}_{\min}(\bar{\theta}_n) - \bar{\delta}\,\overline{M}_2\right)}} \sqrt{\int_{\|u\| \leq \bar{\delta}\sqrt{n}} \frac{\left|\det\left(\bar{J}_n(\bar{\theta}_n)\right)\right|^{1/2} e^{-u^T \bar{J}_n(\bar{\theta}_n) u/2}}{F_n^{MAP}(2\pi)^{d/2}} \|u\|^4 du}$$

$$\leq \frac{\sqrt{3}\ \mathrm{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right] \overline{M}_2}{4\sqrt{n\left(\overline{\lambda}_{\min}(\bar{\theta}_n) - \bar{\delta}\overline{M}_2\right)\left(1 - \bar{\mathscr{D}}(n, \bar{\delta})\right)}}, \tag{E.4}$$

as long as $n > \frac{\mathrm{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]}{\bar{\delta}^2}$, where the last inequality follows from Lemma 29 and the fact that, for $U \sim \mathcal{N}(0, \bar{J}_n(\bar{\theta}_n)^{-1})$, $\mathbb{E}\|U\|^4 \leq 3\,\mathrm{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]^2$. To see this last fact, use the spectral theorem and write $\bar{J}_n(\bar{\theta}_n)^{-1} = P^T \Lambda P$, for an orthogonal matrix P and diagonal matrix Λ. Then, it follows that, for $N \sim \mathcal{N}(0, I_{d\times d})$, $U^T U = N^T P^T \Lambda P N = (PN)^T \Lambda (PN)$. Since $(PN) \sim \mathcal{N}(0, I_{d\times d})$, note that $\mathbb{E}(U^T U)^2 = \mathbb{E}\left[\left(\sum_{i=1}^d N_{\lambda_i}^2\right)^2\right] \leq 3\left(\sum_{i=1}^d \lambda_i\right)^2$, where λ_i's are the diagonal elements of Λ and N_{λ_i}'s are independent such that $N_{\lambda_i} \sim \mathcal{N}(0, \lambda_i)$ for $i = 1, \dots, d$.

E.1.3 Conclusion

The result now follows from adding together bounds in Eqs. (E.3) and (E.4).

E.2 Proof of Theorem 18

E.2.1 Controlling term I_2^{MAP}

Now we wish to control I_2^{MAP} uniformly over all functions g which are 1-Lipschitz. Let us fix a function g that is 1-Lipschitz and WLOG set $g(0) = 0$. In that case $|g(u)| \leq \|u\|$ and, using the notation of Appendix D and Eq. (D.21),

$$I_{2,1}^{MAP} \leq \int_{\|u\| > \bar{\delta}\sqrt{n}} \|u\| \frac{\sqrt{\left|\det \bar{J}_n(\bar{\theta}_n)\right|} e^{-u^T \bar{J}_n(\bar{\theta}_n) u/2}}{(2\pi)^{d/2}} du$$
$$+ \frac{n^{d/2+1/2} e^{-n\bar{\kappa}} \hat{M}_1 \left|\det\left(\hat{J}_n^p(\hat{\theta}_n, \delta)\right)\right|^{1/2} \int_{\|v - \hat{\theta}_n\| > \bar{\delta} - \|\hat{\theta}_n - \bar{\theta}_n\|} \|v - \bar{\theta}_n\| \pi(v) du}{(2\pi)^{d/2}\left(1 - \hat{\mathscr{D}}^p(n, \delta)\right)}.$$

Now, for $U \sim \mathcal{N}(0, \bar{J}_n(\bar{\theta}_n)^{-1})$, and assuming that $n > \frac{\mathrm{Tr}\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}{\bar{\delta}^2}$,

$$
\begin{aligned}
&\int_{\|u\|>\bar{\delta}\sqrt{n}} \|u\| \frac{\sqrt{\left|\det \bar{J}_n(\bar{\theta}_n)\right|} e^{-u^T \bar{J}_n(\bar{\theta}_n) u/2}}{(2\pi)^{d/2}} du \\
&= \int_0^\infty \mathbb{P}\left[\|U\| \mathbb{1}_{[\|U\|>\bar{\delta}\sqrt{n}]} > t\right] dt \\
&\leq \int_0^\infty \mathbb{P}\left[\|U\| > \max(t, \bar{\delta}\sqrt{n})\right] dt \\
&\overset{\text{Theorem 29}}{\leq} \int_{\bar{\delta}\sqrt{n}}^\infty \exp\left[-\frac{1}{2}\left(t - \sqrt{\mathrm{Tr}\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}\right)^2 \overline{\lambda}_{\min}(\bar{\theta}_n)\right] dt \\
&\quad + \bar{\delta}\sqrt{n} \exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n} - \sqrt{\mathrm{Tr}\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}\right)^2 \overline{\lambda}_{\min}(\bar{\theta}_n)\right] \\
&\leq \left(\bar{\delta}\sqrt{n} + \sqrt{\frac{2\pi}{\overline{\lambda}_{\min}(\bar{\theta}_n)}}\right) \exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n} - \sqrt{\mathrm{Tr}\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}\right)^2 \overline{\lambda}_{\min}(\bar{\theta}_n)\right]. \qquad (E.5)
\end{aligned}
$$

This means that

$$
\begin{aligned}
I_{2,1}^{MAP} \leq &\left(\bar{\delta}\sqrt{n} + \sqrt{\frac{2\pi}{\overline{\lambda}_{\min}(\bar{\theta}_n)}}\right) \bar{\mathscr{D}}(n, \bar{\delta}) \\
&+ \frac{n^{d/2+1/2} e^{-n\bar{\kappa}} \hat{M}_1 \left|\det\left(\hat{J}_n^p(\hat{\theta}_n, \delta)\right)\right|^{1/2} \int_{\|v-\hat{\theta}_n\|>\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|} \|v - \bar{\theta}_n\| \pi(v) dv}{(2\pi)^{d/2}\left(1 - \hat{\mathscr{D}}^p(n, \delta)\right)}. \qquad (E.6)
\end{aligned}
$$

Now, using Eq. (D.23), we have

$$
\begin{aligned}
I_{2,2}^{MAP} \leq &\frac{\left|\det\left(\bar{J}_n^p(\bar{\theta}_n, \bar{\delta})\right)\right|^{1/2} \left|\det\left(\bar{J}_n^m(\bar{\theta}_n, \bar{\delta})\right)\right|^{-1/2} \sqrt{\mathrm{Tr}\left[\bar{J}_n^m(\bar{\theta}_n, \bar{\delta})^{-1}\right]}}{1 - \bar{\mathscr{D}}^p(n, \bar{\delta})} \\
&\cdot \left\{\bar{\mathscr{D}}(n, \bar{\delta}) + \frac{n^{d/2} e^{-n\bar{\kappa}} \hat{M}_1 \left|\det\left(\hat{J}_n^p(\hat{\theta}_n, \delta)\right)\right|^{1/2}}{(2\pi)^{d/2}\left(1 - \hat{\mathscr{D}}^p(n, \delta)\right)}\right\}. \qquad (E.7)
\end{aligned}
$$

Adding together bounds in Equations (E.6) and (E.7) now yields a bound on I_2^{MAP}.

E.2.2 Controlling term I_1^{MAP} using the log-Sobolev inequality and the transportation-entropy inequality

As in Appendix E.1.2, we shall use the log-Sobolev inequality $\mathrm{LSI}\left(\overline{\lambda}_{\min}(\bar{\theta}_n) - \overline{\delta M_2}\right)$ for the measure $\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n\right)\right)\right]_{B_0(\bar{\delta}\sqrt{n})}$, implied by Proposition 32. A consequence of the log-Sobolev inequality is that we can apply the **transportation-entropy inequality** for $\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n\right)\right)\right]_{B_0(\bar{\delta}\sqrt{n})}$, given by Proposition 35. It lets us upper bound the 1- and 2-Wasserstein distances (see Eq. (D.14)) by a constant times the KL divergence. The

KL divergence is in turn bounded by a constant times the Fisher divergence by the log-Sobolev inequality. Let $W_2(\cdot,\cdot)$ denote the 2-Wasserstein distance and $W_1(\cdot,\cdot)$ denote the 1-Wasserstein distance. We have that for all 1-Lipschitz functions g,

$$
\begin{aligned}
I_1^{MAP} \leq & W_1\left(\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n-\bar{\theta}_n\right)\right)\right]_{B_0(\bar{\delta}\sqrt{n})}, \left[\mathcal{N}(0,\bar{J}_n(\bar{\theta}_n)^{-1})\right]_{B_0(\bar{\delta}\sqrt{n})}\right)\\
\overset{Eq.\ (D.15)}{\leq} & W_2\left(\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n-\bar{\theta}_n\right)\right)\right]_{B_0(\bar{\delta}\sqrt{n})}, \left[\mathcal{N}(0,\bar{J}_n(\bar{\theta}_n)^{-1})\right]_{B_0(\bar{\delta}\sqrt{n})}\right)\\
\underset{\text{inequality}}{\overset{\text{transportation-entropy}}{\leq}} & \frac{\sqrt{2\,\mathrm{KL}\left(\left[\mathcal{N}(0,\bar{J}_n(\bar{\theta}_n)^{-1})\right]_{B_0(\bar{\delta}\sqrt{n})}\,\middle\|\,\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n-\bar{\theta}_n\right)\right)\right]_{B_0(\bar{\delta}\sqrt{n})}\right)}}{\sqrt{\overline{\lambda}_{\min}(\bar{\theta}_n)-\bar{\delta}\,\overline{M}_2}}\\
\underset{\text{inequality}}{\overset{\text{log-Sobolev}}{\leq}} & \frac{\sqrt{\mathrm{Fisher}\left(\left[\mathcal{N}(0,\bar{J}_n(\bar{\theta}_n)^{-1})\right]_{B_0(\bar{\delta}\sqrt{n})}\,\middle\|\,\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n-\bar{\theta}_n\right)\right)\right]_{B_0(\bar{\delta}\sqrt{n})}\right)}}{\overline{\lambda}_{\min}(\bar{\theta}_n)-\bar{\delta}\,\overline{M}_2}\\
\leq & \frac{1}{\overline{\lambda}_{\min}(\bar{\theta}_n)-\bar{\delta}\,\overline{M}_2}\\
& \cdot\sqrt{\int_{\|u\|\leq\bar{\delta}\sqrt{n}}\frac{\sqrt{\left|\det\left(\bar{J}_n(\bar{\theta}_n)\right)\right|}e^{-u^T\bar{J}_n(\bar{\theta}_n)u/2}}{F_n^{MAP}(2\pi)^{d/2}}\left\|\bar{J}_n(\bar{\theta}_n)u+\frac{\overline{L}_n'(n^{-1/2}u+\bar{\theta}_n)}{\sqrt{n}}\right\|^2du}\\
\underset{\text{theorem}}{\overset{\text{Taylor's}}{\leq}} & \frac{\overline{M}_2}{2\sqrt{n}\left(\overline{\lambda}_{\min}(\bar{\theta}_n)-\bar{\delta}\,\overline{M}_2\right)}\sqrt{\int_{\|u\|\leq\bar{\delta}\sqrt{n}}\frac{\left|\det\left(\bar{J}_n(\bar{\theta}_n)\right)\right|^{1/2}e^{-u^T\bar{J}_n(\bar{\theta}_n)u/2}}{F_n^{MAP}(2\pi)^{d/2}}\|u\|^4\,du}\\
\leq & \frac{\sqrt{3}\ \mathrm{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]\overline{M}_2}{2\left(\overline{\lambda}_{\min}(\bar{\theta}_n)-\bar{\delta}\overline{M}_2\right)\sqrt{n\left(1-\bar{\mathscr{D}}(n,\bar{\delta})\right)}},
\end{aligned}
\tag{E.8}
$$

as long as $n > \frac{\mathrm{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]}{\bar{\delta}^2}$, where the last iequality follows from Lemma 29.

E.2.3 Conclusion

The result now follows from adding together bounds in Equations (E.6) – (E.8).

E.3 Proof of Theorem 20

E.3.1 Controlling term I_2^{MAP}

Now we wish to control I_2^{MAP} uniformly over all functions g which are of the form $g(u) = \langle v,u\rangle^2$ for some $v \in \mathbb{R}^d$ with $\|v\| = 1$. Such functions satisfy the following property: $|g(u)| \leq \|u\|^2$. Using the notation of Appendix D and Eq. (D.21), we therefore have that:

$$
I_{2,1}^{MAP} \leq \int_{\|u\|>\bar{\delta}\sqrt{n}}\|u\|^2\frac{\sqrt{\left|\det\bar{J}_n(\bar{\theta}_n)\right|}e^{-u^T\bar{J}_n(\bar{\theta}_n)u/2}}{(2\pi)^{d/2}}du
$$

$$+\frac{n^{d/2+1}e^{-n\bar{\kappa}}\hat{M}_1\left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2}\int_{\|v-\hat{\theta}_n\|>\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|}\|v-\bar{\theta}_n\|^2\pi(v)du}{(2\pi)^{d/2}\left(1-\hat{\mathscr{D}}^p(n,\delta)\right)}.$$

Now, for $U\sim\mathcal{N}(0,\bar{J}_n(\bar{\theta}_n)^{-1})$, and assuming that $n>\frac{\mathrm{Tr}\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}{\bar{\delta}^2}$,

$$\begin{aligned}
&\int_{\|u\|>\bar{\delta}\sqrt{n}}\|u\|^2\frac{\sqrt{\left|\det\bar{J}_n(\bar{\theta}_n)\right|}e^{-u^T\bar{J}_n(\bar{\theta}_n)u/2}}{(2\pi)^{d/2}}du\\
&=\int_0^\infty\mathbb{P}\left[\|U\|^2\mathbb{1}_{[\|U\|>\bar{\delta}\sqrt{n}]}>t\right]dt \qquad \text{(E.9)}\\
&\le\int_0^\infty\mathbb{P}\left[\|U\|>\max(\sqrt{t},\bar{\delta}\sqrt{n})\right]dt\\
&\le\int_{\bar{\delta}^2n}^\infty\exp\left[-\frac{1}{2}\left(\sqrt{t}-\sqrt{\mathrm{Tr}\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}\right)^2\overline{\lambda}_{\min}(\bar{\theta}_n)\right]dt\\
&\quad+\bar{\delta}^2n\,\exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n}-\sqrt{\mathrm{Tr}\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}\right)^2\overline{\lambda}_{\min}(\bar{\theta}_n)\right]\\
&\le\left(\bar{\delta}^2n+\sqrt{\frac{2\pi}{\overline{\lambda}_{\min}(\bar{\theta}_n)}}\right)\exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n}-\sqrt{\mathrm{Tr}\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}\right)^2\overline{\lambda}_{\min}(\bar{\theta}_n)\right], \qquad \text{(E.10)}
\end{aligned}$$

where we have used Lemma 29. This means that

$$\begin{aligned}
I_{2,1}^{MAP}\le&\left(\bar{\delta}^2n+\sqrt{\frac{2\pi}{\overline{\lambda}_{\min}(\bar{\theta}_n)}}\right)\bar{\mathscr{D}}(n,\bar{\delta})\\
&+\frac{n^{d/2+1}e^{-n\bar{\kappa}}\hat{M}_1\left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2}\int_{\|v-\hat{\theta}_n\|>\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|}\|v-\bar{\theta}_n\|^2\pi(v)du}{(2\pi)^{d/2}\left(1-\hat{\mathscr{D}}^p(n,\delta)\right)}. \qquad \text{(E.11)}
\end{aligned}$$

Now, using Eq. (D.23), we have

$$\begin{aligned}
I_{2,2}^{MAP}\le&\frac{\left|\det\left(\bar{J}_n^p(\bar{\theta}_n,\bar{\delta})\right)\right|^{1/2}\left|\det\left(\bar{J}_n^m(\bar{\theta}_n,\bar{\delta})\right)\right|^{-1/2}\mathrm{Tr}\left[\bar{J}_n^m(\bar{\theta}_n,\bar{\delta})^{-1}\right]}{1-\bar{\mathscr{D}}^p(n,\bar{\delta})}\\
&\cdot\left\{\bar{\mathscr{D}}(n,\bar{\delta})+\frac{n^{d/2}e^{-n\bar{\kappa}}\hat{M}_1\left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2}}{(2\pi)^{d/2}\left(1-\hat{\mathscr{D}}^p(n,\delta)\right)}\right\}. \qquad \text{(E.12)}
\end{aligned}$$

A bound on I_2^{MAP} now follows from adding up the bounds in Eqs. (E.11) and (E.12).

E.3.2 Controlling term I_1^{MAP} using the log-Sobolev inequality and the transportation-information inequality

Note that calculation from Eq. (E.8) yields that

$$W_2\left(\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n-\bar{\theta}_n\right)\right)\right]_{B_0(\bar{\delta}\sqrt{n})},\left[\mathcal{N}(0,\bar{J}_n(\bar{\theta}_n)^{-1})\right]_{B_0(\bar{\delta}\sqrt{n})}\right)$$

$$\leq \frac{\sqrt{3}\ \mathrm{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]\overline{M}_2}{2\left(\bar{\lambda}_{\min}(\bar{\theta}_n)-\bar{\delta}\overline{M}_2\right)\sqrt{n\left(1-\bar{\mathscr{D}}(n,\bar{\delta})\right)}}. \tag{E.13}$$

Now, let us fix two random vectors: $X \sim \left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n-\bar{\theta}_n\right)\right)\right]_{B_0(\bar{\delta}\sqrt{n})}$ and $Y \sim \left[\mathcal{N}(0,\bar{J}_n(\bar{\theta}_n)^{-1})\right]_{B_0(\bar{\delta}\sqrt{n})}$ and a vector v such that $\|v\|=1$. Let γ denote the set of all couplings between $\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n-\bar{\theta}_n\right)\right)\right]_{B_0(\bar{\delta}\sqrt{n})}$ and $\left[\mathcal{N}(0,\bar{J}_n(\bar{\theta}_n)^{-1})\right]_{B_0(\bar{\delta}\sqrt{n})}$. Let $(\tilde{X},\tilde{Y})\in\gamma$ be such that

$$\inf_{(Z_1,Z_2)\in\gamma}\mathbb{E}\left[\|Z_1-Z_2\|^2\right]=\mathbb{E}\left[\|\tilde{X}-\tilde{Y}\|^2\right].$$

It follows that:

$$\begin{aligned}\mathbb{E}\left[\langle v,X\rangle^2\right]-\mathbb{E}\left[\langle v,Y\rangle^2\right] =&\mathbb{E}\left[\left\langle v,\tilde{X}-\tilde{Y}\right\rangle\left\langle v,\tilde{X}+\tilde{Y}\right\rangle\right]\\ =&\mathbb{E}\left[\left\langle v,\tilde{X}-\tilde{Y}\right\rangle^2\right]+2\mathbb{E}\left[\left\langle v,\tilde{X}-\tilde{Y}\right\rangle\left\langle v,\tilde{Y}\right\rangle\right]\\ \leq&\mathbb{E}\left[\left\|\tilde{X}-\tilde{Y}\right\|^2\right]+2\sqrt{\mathbb{E}\left[\left\|\tilde{X}-\tilde{Y}\right\|^2\right]}\sqrt{\mathbb{E}\left[\|Y\|^2\right]}.\end{aligned}$$

Therefore, for all functions g which are of the form $g(u)=\langle v,u\rangle^2$ for some $v\in\mathbb{R}^d$ with $\|v\|=1$, we have that

$$\begin{aligned}I_1^{MAP} \leq& W_2\left(\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n-\bar{\theta}_n\right)\right)\right]_{B_0(\bar{\delta}\sqrt{n})},\left[\mathcal{N}(0,\bar{J}_n(\bar{\theta}_n)^{-1})\right]_{B_0(\bar{\delta}\sqrt{n})}\right)^2\\ &+2W_2\left(\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n-\bar{\theta}_n\right)\right)\right]_{B_0(\bar{\delta}\sqrt{n})},\left[\mathcal{N}(0,\bar{J}_n(\bar{\theta}_n)^{-1})\right]_{B_0(\bar{\delta}\sqrt{n})}\right)\\ &\cdot\frac{\sqrt{\mathrm{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]}}{\sqrt{1-\bar{\mathscr{D}}(n,\bar{\delta})}},\end{aligned} \tag{E.14}$$

if $n>\frac{\mathrm{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]}{\bar{\delta}^2}$. The final bound on I_1^{MAP} may now be obtained by using Eq. (E.13).

E.3.3 Conclusion

The result now follows by combining Eq. (E.14) with Eq. (E.13) and then summing together with Eqs. (E.11) and (E.12).

Appendix F. Proofs of Theorems 22, 23 and 25

Throughout this section we adopt the notation of Appendix D. For any probability measure μ, we let $[\mu]_{B_0(\delta\sqrt{n})}$ denote its restriction (truncation) to the ball of radius $\delta\sqrt{n}$ around 0. In all the proofs below, we wish to control the quantity D_g^{MLE} of Eq. (D.1) for all functions g that satisfy certain prescribed criteria. In the proof of Theorem 17, we look at functions g that are indicators of measurable sets, in the proof of Theorem 18 we look at 1-Lipschitz functions g and in the proof of Theorem 20 at those which are of the form $g(x) = \langle v, x \rangle^2$ for some $v \in \mathbb{R}^d$ with $\|v\| = 1$. In order to prove Theorems 22, 23 and 25, we will bound terms I_2^{MLE} and I_1^{MAP} of Eq. (D.7) separately.

F.1 Proof of Theorem 22

F.1.1 Controlling term I_2^{MLE}

We wish to obtain a uniform bound on I_2^{MLE} for all functions g which are indicators of measurable sets. Every indicator function is upper-bounded by one, so we can use Remark 36 to obtain:

$$I_{2,2}^{MLE} \leq \hat{\mathscr{D}}(n,\delta) + \frac{n^{d/2} e^{-n\kappa} \hat{M}_1 \left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2}}{(2\pi)^{d/2}\left(1-\hat{\mathscr{D}}^p(n,\delta)\right)}.$$

Similarly, since $|g| \leq 1$, we can use Eq. (D.18) and Lemma 29 to obtain

$$I_{2,1}^{MLE} \leq \hat{\mathscr{D}}(n,\delta) + \frac{n^{d/2} e^{-n\kappa} \hat{M}_1 \left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2}}{(2\pi)^{d/2}\left(1-\hat{\mathscr{D}}^p(n,\delta)\right)},$$

which implies that

$$I_2^{MLE} \leq 2\,\hat{\mathscr{D}}(n,\delta) + \frac{2n^{d/2} e^{-n\kappa} \hat{M}_1 \left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2}}{(2\pi)^{d/2}\left(1-\hat{\mathscr{D}}^p(n,\delta)\right)}. \tag{F.1}$$

F.1.2 Controlling term I_1^{MLE} using the log-Sobolev inequality

We shall proceed as we did in Appendix E.1.2. Let $\mathrm{KL}(\cdot\|\cdot)$ denote the Kullback-Leibler divergence (Eq. (D.10)) and $\mathrm{Fisher}(\cdot\|\cdot)$ denote the Fisher divergence (Eq. (D.11)). Let F_n^{MLE} be given by Eq. (D.6).

Note that, by Assumption 1, for t such that $\|t\| < \sqrt{n}\delta$, we have

$$-n^{-1} L_n''(\hat{\theta}_n + n^{-1/2}t) \succeq \hat{J}_n(\hat{\theta}_n) - \delta M_2 I_{d\times d} \succeq \left(\lambda_{\min}(\hat{\theta}_n) - \delta M_2\right) I_{d\times d}.$$

This means that, inside the convex set $\{t \in \mathbb{R}^d : \|t\| < \sqrt{n}\delta\}$, the measure whose density, up to a normalizing constant, is given by $t \mapsto e^{L_n(\hat{\theta}_n + n^{-1/2}t)} \mathbb{1}_{\{\|t\|<\sqrt{n}\delta\}}(t)$, is $\left(\lambda_{\min}(\hat{\theta}_n) - \delta M_2\right)$-strongly log-concave (see e.g. Saumard and Wellner 2014). Also, note that

$$\frac{\sup_{\|t\|\leq\delta\sqrt{n}} \pi(n^{-1/2}t + \hat{\theta}_n)}{\inf_{\|t\|\leq\delta\sqrt{n}} \pi(n^{-1/2}t + \hat{\theta}_n)} \leq \widetilde{M}_1 \hat{M}_1.$$

Using the **Bakry-Émery criterion**, given by Proposition Theorem 32, and the **Holley-Stroock perturbation principle**, given by Proposition 33, we therefore have that

$\left[\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n\right)\right]_{B_0(\delta\sqrt{n})}$ satisfies the **log-Sobolev inequality** $\mathrm{LSI}\left(\frac{\lambda_{\min}(\hat{\theta}_n)-\delta M_2}{\widetilde{M}_1 \hat{M}_1}\right)$ (see Definition 30). By combining the log-Sobolev inequality with **Pinsker's inequality** (Proposition 34) we obtain that, for all functions g, which are indicators of measurable sets,

$$
\begin{aligned}
I_1^{MLE} \leq& \mathrm{TV}\left(\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n\right)\right)\right]_{B_0(\delta\sqrt{n})}, \left[\mathcal{N}(0, \hat{J}_n(\hat{\theta}_n)^{-1})\right]_{B_0(\delta\sqrt{n})}\right) \\
\overset{\text{Pinsker's}}{\underset{\text{inequality}}{\leq}}& \sqrt{\frac{1}{2}\mathrm{KL}\left(\left[\mathcal{N}(0, \hat{J}_n(\hat{\theta}_n)^{-1})\right]_{B_0(\delta\sqrt{n})} \middle\| \left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n\right)\right)\right]_{B_0(\delta\sqrt{n})}\right)} \\
\overset{\text{log-Sobolev}}{\underset{\text{inequality}}{\leq}}& \frac{\sqrt{\widetilde{M}_1 \hat{M}_1}}{2\sqrt{\lambda_{\min}(\hat{\theta}_n) - \delta M_2}} \sqrt{\mathrm{Fisher}\left(\left[\mathcal{N}(0, \hat{J}_n(\hat{\theta}_n)^{-1})\right]_{B_0(\delta\sqrt{n})} \middle\| \left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n\right)\right)\right]_{B_0(\delta\sqrt{n})}\right)} \\
\leq& \frac{\sqrt{\widetilde{M}_1 \hat{M}_1}}{2\sqrt{\lambda_{\min}(\hat{\theta}_n) - \delta M_2}} \\
&\cdot \sqrt{\int_{\|u\| \leq \delta\sqrt{n}} \frac{\sqrt{\left|\det\left(\hat{J}_n(\hat{\theta}_n)\right)\right|} e^{-u^T \hat{J}_n(\hat{\theta}_n) u/2}}{F_n^{MLE}(2\pi)^{d/2}} \left\| \hat{J}_n(\hat{\theta}_n) u + \frac{L_n'(n^{-1/2}u + \hat{\theta}_n)}{\sqrt{n}} \right\|^2 du} \\
&+ \frac{\sqrt{\widetilde{M}_1 \hat{M}_1}}{2\sqrt{n\left(\lambda_{\min}(\hat{\theta}_n) - \delta M_2\right)}} \\
&\quad \cdot \sqrt{\int_{\|u\| \leq \delta\sqrt{n}} \frac{\sqrt{\left|\det\left(\hat{J}_n(\hat{\theta}_n)\right)\right|} e^{-u^T \hat{J}_n(\hat{\theta}_n) u/2}}{F_n^{MLE}(2\pi)^{d/2}} \left\| \frac{\pi'(n^{-1/2}u + \hat{\theta}_n)}{\pi(n^{-1/2}u + \hat{\theta}_n)} \right\|^2 du} \\
\overset{\text{Taylor's}}{\underset{\text{theorem}}{\leq}}& \frac{\sqrt{\widetilde{M}_1 \hat{M}_1} M_2}{4\sqrt{n\left(\lambda_{\min}(\hat{\theta}_n) - \delta M_2\right)}} \sqrt{\int_{\|u\| \leq \delta\sqrt{n}} \frac{\sqrt{\left|\det\left(\hat{J}_n(\hat{\theta}_n)\right)\right|} e^{-u^T \hat{J}_n(\hat{\theta}_n) u/2}}{F_n^{MLE}(2\pi)^{d/2}} \|u\|^4 du} \\
&+ \frac{M_1 \sqrt{\widetilde{M}_1 \hat{M}_1}}{2\sqrt{n\left(\lambda_{\min}(\hat{\theta}_n) - \delta M_2\right)}} \\
\leq& \frac{\sqrt{3}\ \mathrm{Tr}\left[\hat{J}_n(\hat{\theta}_n)^{-1}\right] \sqrt{\widetilde{M}_1 \hat{M}_1}\, M_2}{4\sqrt{n\left(\lambda_{\min}(\hat{\theta}_n) - \delta M_2\right)\left(1 - \hat{\mathscr{D}}(n, \delta)\right)}} + \frac{M_1 \sqrt{\widetilde{M}_1 \hat{M}_1}}{2\sqrt{n\left(\lambda_{\min}(\hat{\theta}_n) - \delta M_2\right)}},
\end{aligned}
\tag{F.2}
$$

as long as $n > \frac{\mathrm{Tr}\left[\hat{J}_n(\hat{\theta}_n)^{-1}\right]}{\delta^2}$, where we have used Lemma 29.

F.1.3 Conclusion

The result now follows by adding together the bounds in Eqs. (F.1) and (F.2).

F.2 Proof of Theorem 23

F.2.1 Controlling term I_2^{MLE}

Now we wish to control I_2^{MLE} uniformly over all functions g which are 1-Lipschitz and WLOG set $g(0) = 0$. It follows that $|g(u)| \leq \|u\|$ and, using the notation of Appendix D and Eq. (D.18),

$$I_{2,1}^{MLE} \leq \int_{\|u\|>\delta\sqrt{n}} \|u\| \frac{\sqrt{\left|\det \hat{J}_n(\hat{\theta}_n)\right|} e^{-u^T \hat{J}_n(\hat{\theta}_n)u/2}}{(2\pi)^{d/2}} du + \frac{n^{d/2+1/2} e^{-n\kappa} \hat{M}_1 \left|\det\left(\hat{J}_n^p(\hat{\theta}_n, \delta)\right)\right|^{1/2} \int_{\|u\|>\delta} \|u\| \pi(u + \hat{\theta}_n) du}{(2\pi)^{d/2}\left(1 - \hat{\mathscr{D}}^p(n, \delta)\right)}.$$

A calculation similar to Eq. (E.5) reveals that

$$\int_{\|u\|>\delta\sqrt{n}} \|u\| \frac{\sqrt{\left|\det \hat{J}_n(\hat{\theta}_n)\right|} e^{-u^T \hat{J}_n(\hat{\theta}_n)u/2}}{(2\pi)^{d/2}} du \leq \left(\delta\sqrt{n} + \sqrt{\frac{2\pi}{\lambda_{\min}(\hat{\theta}_n)}}\right) \hat{\mathscr{D}}(n, \delta)$$

and so

$$I_{2,1}^{MLE} \leq \left(\delta\sqrt{n} + \sqrt{\frac{2\pi}{\lambda_{\min}(\hat{\theta}_n)}}\right) \hat{\mathscr{D}}(n, \delta) + \frac{n^{d/2+1/2} e^{-n\kappa} \hat{M}_1 \left|\det\left(\hat{J}_n^p(\hat{\theta}_n, \delta)\right)\right|^{1/2} \int_{\|u\|>\delta} \|u\| \pi(u + \hat{\theta}_n) du}{(2\pi)^{d/2}\left(1 - \hat{\mathscr{D}}^p(n, \delta)\right)}. \tag{F.3}$$

Now, using Eq. (D.20), we obtain

$$I_{2,2}^{MLE} \leq \frac{\hat{M}_1 \widetilde{M}_1 \left|\det\left(\hat{J}_n^p(\hat{\theta}_n, \delta)\right)\right|^{1/2} \left|\det\left(\hat{J}_n^m(\hat{\theta}_n, \delta)\right)\right|^{-1/2} \sqrt{\mathrm{Tr}\left[\hat{J}_n^m(\hat{\theta}_n, \delta)^{-1}\right]}}{1 - \hat{\mathscr{D}}^p(n, \delta)} \cdot \left\{\hat{\mathscr{D}}(n, \delta) + \frac{n^{d/2} e^{-n\kappa} \hat{M}_1 \left|\det\left(\hat{J}_n^p(\hat{\theta}_n, \delta)\right)\right|^{1/2}}{(2\pi)^{d/2}\left(1 - \hat{\mathscr{D}}^p(n, \delta)\right)}\right\}. \tag{F.4}$$

Adding together bounds from Eqs. (F.3) and (F.4) yields a bound on I_2^{MLE}.

F.2.2 Controlling term I_1^{MLE} using the log-Sobolev inequality and the transportation-entropy inequality

As in Appendix F.1.2, we shall use the log-Sobolev inequality for the measure $\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n\right)\right)\right]_{B_0(\delta\sqrt{n})}$. A consequence of the log-Sobolev inequality is that we can apply the transportation-entropy inequality for $\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n\right)\right)\right]_{B_0(\delta\sqrt{n})}$ (Proposition 35), which lets us upper bound the 1- and 2-Wasserstein distances by a constant times the KL divergence. The log-Sobolev inequality then upper-bounds the KL divergence in terms of the Fisher divergence. Let $W_2(\cdot,\cdot)$ denote the 2-Wasserstein distance and $W_1(\cdot,\cdot)$ denote the 1-Wasserstein distance. We have that, for all 1-Lipschitz test functions g,

$$
\begin{aligned}
I_1^{MLE} \leq& W_1\left(\left[\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n\right)\right]_{B_0(\delta\sqrt{n})}, \left[\mathcal{N}(0, \hat{J}_n(\hat{\theta}_n)^{-1})\right]_{B_0(\delta\sqrt{n})}\right) \\
\overset{Eq.\ (D.15)}{\leq}& W_2\left(\left[\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n\right)\right]_{B_0(\delta\sqrt{n})}, \left[\mathcal{N}(0, \hat{J}_n(\hat{\theta}_n)^{-1})\right]_{B_0(\delta\sqrt{n})}\right) \\
\underset{\text{inequality}}{\overset{\text{transportation-entropy}}{\leq}}& \sqrt{\frac{2\widetilde{M}_1\hat{M}_1 \ \mathrm{KL}\left(\left[\mathcal{N}(0, \hat{J}_n(\hat{\theta}_n)^{-1})\right]_{B_0(\delta\sqrt{n})} \middle\| \left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n\right)\right)\right]_{B_0(\delta\sqrt{n})}\right)}{\lambda_{\min}(\hat{\theta}_n) - \delta M_2}} \\
\underset{\text{inequality}}{\overset{\text{log-Sobolev}}{\leq}}& \frac{\widetilde{M}_1\hat{M}_1\sqrt{\mathrm{Fisher}\left(\left[\mathcal{N}(0, \hat{J}_n(\hat{\theta}_n)^{-1})\right]_{B_0(\delta\sqrt{n})} \middle\| \left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n\right)\right)\right]_{B_0(\delta\sqrt{n})}\right)}}{\lambda_{\min}(\hat{\theta}_n) - \delta M_2} \\
\leq& \frac{\widetilde{M}_1\hat{M}_1}{\lambda_{\min}(\hat{\theta}_n) - \delta M_2} \\
&\cdot \sqrt{\int_{\|u\|\leq\delta\sqrt{n}} \frac{\sqrt{\left|\det\left(\hat{J}_n(\hat{\theta}_n)\right)\right|} e^{-u^T\hat{J}_n(\hat{\theta}_n)u/2}}{F_n^{MLE}(2\pi)^{d/2}} \left\| \hat{J}_n(\hat{\theta}_n)u + \frac{L_n'(n^{-1/2}u + \hat{\theta}_n)}{\sqrt{n}} \right\|^2 du} \\
&+ \frac{\widetilde{M}_1\hat{M}_1}{\sqrt{n}\left(\lambda_{\min}(\hat{\theta}_n) - \delta M_2\right)} \\
&\quad \cdot \sqrt{\int_{\|u\|\leq\delta\sqrt{n}} \frac{\sqrt{\left|\det\left(\hat{J}_n(\hat{\theta}_n)\right)\right|} e^{-u^T\hat{J}_n(\hat{\theta}_n)u/2}}{F_n^{MLE}(2\pi)^{d/2}} \left\| \frac{\pi'(n^{-1/2}u + \hat{\theta}_n)}{\pi(n^{-1/2}u + \hat{\theta}_n)} \right\|^2 du} \\
\underset{\text{theorem}}{\overset{\text{Taylor's}}{\leq}}& \frac{\widetilde{M}_1\hat{M}_1M_2}{2\sqrt{n}\left(\lambda_{\min}(\hat{\theta}_n) - \delta M_2\right)} \sqrt{\int_{\|u\|\leq\delta\sqrt{n}} \frac{\sqrt{\left|\det\left(\hat{J}_n(\hat{\theta}_n)\right)\right|} e^{-u^T\hat{J}_n(\hat{\theta}_n)u/2}}{F_n^{MLE}(2\pi)^{d/2}} \|u\|^4 \, du} \\
&+ \frac{M_1\widetilde{M}_1\hat{M}_1}{\sqrt{n}\left(\lambda_{\min}(\hat{\theta}_n) - \delta M_2\right)}
\end{aligned}
$$

$$\leq \frac{\sqrt{3}\ \mathrm{Tr}\left[\hat{J}_n(\hat{\theta}_n)^{-1}\right]\widetilde{M}_1\hat{M}_1\, M_2}{2\left(\lambda_{\min}(\hat{\theta}_n)-\delta M_2\right)\sqrt{n\left(1-\hat{\mathscr{D}}(n,\delta)\right)}}+\frac{M_1\widetilde{M}_1\hat{M}_1}{\sqrt{n}\left(\lambda_{\min}(\hat{\theta}_n)-\delta M_2\right)}. \tag{F.5}$$

F.2.3 Conclusion

The result now follows from adding together the bounds in Eqs. (F.3) – (F.5).

F.3 Proof of Theorem 25

F.3.1 Controlling term I_2^{MLE}

Now we want to control I_2^{MLE} uniformly over all functions g which are of the form $g(u) = \langle v, u\rangle$, for some $v \in \mathbb{R}^d$ with $\|v\| = 1$. For such functions we have that $|g(u)| \leq \|u\|^2$. Using the notation of Appendix D and Eq. (D.18), we have that

$$I_{2,1}^{MLE} \leq \int_{\|u\|>\delta\sqrt{n}} \|u\|^2 \frac{\sqrt{\left|\det \hat{J}_n(\hat{\theta}_n)\right|}e^{-u^T\hat{J}_n(\hat{\theta}_n)u/2}}{(2\pi)^{d/2}}du$$
$$+\frac{n^{d/2+1}e^{-n\kappa}\hat{M}_1\left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2}\int_{\|u\|>\delta}\|u\|^2\pi(u+\hat{\theta}_n)du}{(2\pi)^{d/2}\left(1-\hat{\mathscr{D}}^p(n,\delta)\right)}.$$

A calculation similar to Eq. (E.10) reveals that

$$\int_{\|u\|>\delta\sqrt{n}} \|u\|^2 \frac{\sqrt{\left|\det \hat{J}_n(\hat{\theta}_n)\right|}e^{-u^T\hat{J}_n(\hat{\theta}_n)u/2}}{(2\pi)^{d/2}}du \leq \left(\delta^2 n + \sqrt{\frac{2\pi}{\lambda_{\min}(\hat{\theta}_n)}}\right)\hat{\mathscr{D}}(n,\delta)$$

and so

$$I_{2,1}^{MLE} \leq \left(\delta^2 n + \sqrt{\frac{2\pi}{\lambda_{\min}(\hat{\theta}_n)}}\right)\hat{\mathscr{D}}(n,\delta)$$
$$+\frac{n^{d/2+1}e^{-n\kappa}\hat{M}_1\left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2}\int_{\|u\|>\delta}\|u\|^2\pi(u+\hat{\theta}_n)du}{(2\pi)^{d/2}\left(1-\hat{\mathscr{D}}^p(n,\delta)\right)}. \tag{F.6}$$

Now, using Eq. (D.20), we obtain

$$I_{2,2}^{MLE} \leq \frac{\hat{M}_1\widetilde{M}_1\left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2}\left|\det\left(\hat{J}_n^m(\hat{\theta}_n,\delta)\right)\right|^{-1/2}\mathrm{Tr}\left[\hat{J}_n^m(\hat{\theta}_n,\delta)^{-1}\right]}{1-\hat{\mathscr{D}}^p(n,\delta)}$$
$$\cdot\left\{\hat{\mathscr{D}}(n,\delta)+\frac{n^{d/2}e^{-n\kappa}\hat{M}_1\left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2}}{(2\pi)^{d/2}\left(1-\hat{\mathscr{D}}^p(n,\delta)\right)}\right\}. \tag{F.7}$$

Adding together bounds from Eqs. (F.6) and (F.7) yields a bound on I_2^{MLE}.

F.3.2 Controlling term I_1^{MLE} using the log-Sobolev inequality and the transportation-entropy inequality

Note that the calculation in Eq. (F.5) yields that

$$W_2\left(\left[\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n-\hat{\theta}_n\right)\right]_{B_0(\delta\sqrt{n})},\left[\mathcal{N}(0,\hat{J}_n(\hat{\theta}_n)^{-1})\right]_{B_0(\delta\sqrt{n})}\right)$$
$$\leq\frac{\sqrt{3}\ \mathrm{Tr}\left[\hat{J}_n(\hat{\theta}_n)^{-1}\right]\widetilde{M_1}\hat{M}_1\,M_2}{2\left(\lambda_{\min}(\hat{\theta}_n)-\delta M_2\right)\sqrt{n\left(1-\hat{\mathscr{D}}(n,\delta)\right)}}+\frac{M_1\widetilde{M_1}\hat{M}_1}{\sqrt{n}\left(\lambda_{\min}(\hat{\theta}_n)-\delta M_2\right)}. \tag{F.8}$$

Therefore, for all functions g which are of the form $g(u)=\langle v,u\rangle^2$ for some $v\in\mathbb{R}^d$ with $\|v\|=1$, an argument similar to the one that led to Eq. (E.14) yields:

$$I_1^{MLE}\leq W_2\left(\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n-\hat{\theta}_n\right)\right)\right]_{B_0(\delta\sqrt{n})},\left[\mathcal{N}(0,\hat{J}_n(\hat{\theta}_n)^{-1})\right]_{B_0(\delta\sqrt{n})}\right)^2$$
$$+\frac{2\sqrt{\mathrm{Tr}\left[\hat{J}_n(\hat{\theta}_n)^{-1}\right]}}{\sqrt{1-\hat{\mathscr{D}}(n,\delta)}}W_2\left(\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n-\hat{\theta}_n\right)\right)\right]_{B_0(\delta\sqrt{n})},\left[\mathcal{N}(0,\hat{J}_n(\hat{\theta}_n)^{-1})\right]_{B_0(\delta\sqrt{n})}\right), \tag{F.9}$$

and the final bound on I_1^{MLE} follows from Eq. (F.8).

F.3.3 Conclusion

The result now follows from combining Eqs. (F.8) and (F.9) and adding together with Eqs. (F.6) and (F.7).

Appendix G. Proof of Theorem 27

In this section, we concentrate on the univariate context (i.e. on $d=1$). We shall apply Stein's method, in the framework described in Ernst et al. (2020, Section 2.1). Before we do that, however, let us recall that we want to upper-bound the quantity D_g^{MLE} given by Eq. (D.1) for all functions g for which the two expectations in Eq. (D.1) exist. Recall the definition of C_n^{MLE} from Eq. (D.5) and let:

$$h(t)=h_g^{MLE}(t)=g(t)-\frac{n^{-1/2}}{C_n}\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}}g(u)\Pi_n(n^{-1/2}u+\hat{\theta}_n)du. \tag{G.1}$$

We can repeat the calculation leading to Eq. (D.7), without dividing the first term after the first inequality by F_n^{MLE}. We then obtain:

$$D_g^{MLE}\leq\left|\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}}h(u)\frac{e^{-t^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}}\right|+\left|\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}}h(u)\left[\frac{e^{-u^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}}-n^{-1/2}\Pi_n(n^{-1/2}u+\hat{\theta}_n)\right]\right|$$
$$=:\tilde{I}_1+\tilde{I}_2.$$

We will bound $\tilde{I}_1$ and $\tilde{I}_2$ separately.

G.1 Controlling term $\tilde{I}_2$

Note that $\tilde{I}_2$ is the same as I_2^{MLE} defined by Eq. (D.7), for $d = 1$. We will use the calculations leading to Eqs. (D.18) and (D.19). Instead of using Lemma 29, we will, however, apply the standard one-dimensional Gaussian concentration inequality, which says that, for $Z_n \sim \mathcal{N}\left(0, \sigma_n^2\right)$,

$$\mathbb{P}\left[|Z_n| > \delta\sqrt{n}\right] \leq 2e^{-\delta^2 n/(2\sigma_n^2)}.$$

We obtain

$$\begin{aligned}
\tilde{I}_2 \leq & \left| \int_{|u|>\delta\sqrt{n}} g(u) \frac{e^{-u^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} du \right| \\
& + \frac{n^{1/2} e^{-n\kappa} \hat{M}_1 \left(\frac{1}{\sigma_n^2} + \frac{\delta M_2}{3}\right)^{1/2} \int_{|u|>\delta} |g(u\sqrt{n})| \pi(u + \hat{\theta}_n) du}{\sqrt{2\pi}\left\{1 - 2\exp\left[-\frac{1}{2}\left(\frac{1}{\sigma_n^2} + \frac{M_2\delta}{3}\right)\delta^2 n\right]\right\}} \\
& + \frac{\hat{M}_1 \widetilde{M}_1 \left(\frac{1}{\sigma_n^2} + \frac{\delta M_2}{3}\right)^{1/2} \int_{|t|\leq\delta\sqrt{n}} |g(t)| e^{-\frac{1}{2}\left(\frac{1}{\sigma_n^2} - \frac{M_2\delta}{3}\right)t^2} dt}{\sqrt{2\pi}\left\{1 - 2\exp\left[-\frac{1}{2}\left(\frac{1}{\sigma_n^2} + \frac{M_2\delta}{3}\right)\delta^2 n\right]\right\}} \\
& \qquad \cdot \left\{ 2e^{-\delta^2 n/(2\sigma_n^2)} + \frac{n^{1/2} e^{-n\kappa} \hat{M}_1 \left(\frac{1}{\sigma_n^2} + \frac{\delta M_2}{3}\right)^{1/2}}{\sqrt{2\pi}\left\{1 - 2\exp\left[-\frac{1}{2}\left(\frac{1}{\sigma_n^2} + \frac{M_2\delta}{3}\right)\delta^2 n\right]\right\}} \right\}.
\end{aligned} \tag{G.2}$$

G.2 Controlling term $\tilde{I}_1$ using Stein's method

In this section we will use Stein's method in the framework of Ernst et al. (2020, Section 2.1). Note that, by integration by parts, for all continuous functions $f : [-\delta\sqrt{n}, \delta\sqrt{n}] \to \mathbb{R}$ which are differentiable on $(-\delta\sqrt{n}, \delta\sqrt{n})$ and satisfy $f(-\delta\sqrt{n}) = f(\delta\sqrt{n})$, we have

$$\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} f'(t) \frac{e^{-t^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} dt = \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} t f(t) \frac{e^{-t^2/(2\sigma_n^2)}}{\sigma_n^2 \sqrt{2\pi\sigma_n^2}} dt \tag{G.3}$$

Now, for our function h, given by Eq. (G.1), let

$$f(t) := \begin{cases} \frac{1}{\Pi_n(n^{-1/2}t + \hat{\theta}_n)} \int_{-\delta\sqrt{n}}^{t} h(u) \Pi_n(n^{-1/2}u + \hat{\theta}_n) du, & \text{if } t \in (-\delta\sqrt{n}, \delta\sqrt{n}) \\ 0, & \text{otherwise.} \end{cases} \tag{G.4}$$

Note that f is continuous on $[-\delta\sqrt{n}, \delta\sqrt{n}]$, differentiable on $(-\delta\sqrt{n}, \delta\sqrt{n})$ and $f(-\delta\sqrt{n}) = f(\delta\sqrt{n}) = 0$. Moreover, on $(-\delta\sqrt{n}, \delta\sqrt{n})$, f solves the Stein equation associated to the distribution of $\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n\right)$ for test function h, as described in Ernst et al. (2020, Section 2.1). In other words,

$$h(t) = f'(t) + f(t)\left(\frac{\mathrm{d}}{\mathrm{d}t} \log \Pi_n(n^{-1/2}t + \hat{\theta}_n)\right), \qquad t \in \left(-\delta\sqrt{n}, \delta\sqrt{n}\right). \tag{G.5}$$

Now, by Taylor's theorem, we obtain that, for some $c \in (0,1)$,

$$
\begin{aligned}
&\tilde{I}_1 \\
&\overset{Eq.~(G.5)}{=} \left| \int_{-\delta/\sqrt{n}}^{\delta\sqrt{n}} \left[f'(t) + f(t) \left(\frac{\mathrm{d}}{\mathrm{d}t} L_n(n^{-1/2}t + \hat{\theta}_n) + \frac{\mathrm{d}}{\mathrm{d}t} \log \pi(n^{-1/2}t + \hat{\theta}_n) \right) \right] \frac{e^{-t^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} dt \right| \\
&\leq \left| \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} \left[f'(t) + f(t) \left(\frac{L_n'(\hat{\theta}_n)}{\sqrt{n}} + \frac{tL_n''(\hat{\theta}_n)}{n} + \frac{t^2}{2n^{3/2}} L_n'''(\hat{\theta}_n + cn^{-1/2}t) \right) \right] \frac{e^{-t^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} dt \right| \\
&\quad + \frac{1}{\sqrt{n}} \left| \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} f(t) \frac{\pi'(n^{-1/2}t + \hat{\theta}_n)}{\pi(n^{-1/2}t + \hat{\theta}_n)} \cdot \frac{e^{-t^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} dt \right| \\
&= \left| \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} \left[f'(t) + f(t) \left(-\frac{t}{\sigma_n^2} + \frac{t^2}{2n^{3/2}} L_n'''(\hat{\theta}_n + cn^{-1/2}t) \right) \right] \frac{e^{-t^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} dt \right| \\
&\quad + \frac{1}{\sqrt{n}} \left| \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} f(t) \frac{\pi'(n^{-1/2}t + \hat{\theta}_n)}{\pi(n^{-1/2}t + \hat{\theta}_n)} \cdot \frac{e^{-t^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} dt \right| \\
&\leq \left| \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} \left[f'(t) - \frac{tf(t)}{\sigma_n^2} \right] \frac{e^{-t^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} dt \right| + \frac{M_2}{2\sqrt{n}} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} |t^2 f(t)| \frac{e^{-t^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} dt \\
&\quad + \frac{M_1}{\sqrt{n}} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} |f(t)| \frac{e^{-t^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} dt \\
&\overset{Eq.~(G.3)}{=} \frac{M_2}{2\sqrt{n}} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} |t^2 f(t)| \frac{e^{-t^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} dt + \frac{M_1}{\sqrt{n}} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} |f(t)| \frac{e^{-t^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} dt \\
&\overset{Eq.~(G.4)}{=} \frac{M_2}{2\sqrt{2\pi\sigma_n^2\, n}} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} \frac{t^2 e^{-t^2/(2\sigma_n^2)}}{\Pi_n(n^{-1/2}t + \hat{\theta}_n)} \left| \int_{-\delta\sqrt{n}}^{t} h(u) \Pi_n(n^{-1/2}u + \hat{\theta}_n) du \right| dt \\
&\quad + \frac{M_1}{\sqrt{2\pi\sigma_n^2\, n}} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} \frac{e^{-t^2/(2\sigma_n^2)}}{\Pi_n(n^{-1/2}t + \hat{\theta}_n)} \left| \int_{-\delta\sqrt{n}}^{t} h(u) \Pi_n(n^{-1/2}u + \hat{\theta}_n) du \right| dt \\
&= \frac{M_2}{2\sqrt{2\pi\sigma_n^2\, n}} \int_{-\delta\sqrt{n}}^{0} \frac{t^2 e^{-t^2/(2\sigma_n^2)}}{\Pi_n(n^{-1/2}t + \hat{\theta}_n)} \left| \int_{-\delta\sqrt{n}}^{t} h(u) \Pi_n(n^{-1/2}u + \hat{\theta}_n) du \right| dt \\
&\quad + \frac{M_2}{2\sqrt{2\pi\sigma_n^2\, n}} \int_{0}^{\delta\sqrt{n}} \frac{t^2 e^{-t^2/(2\sigma_n^2)}}{\Pi_n(n^{-1/2}t + \hat{\theta}_n)} \left| \int_{t}^{\delta\sqrt{n}} h(u) \Pi_n(n^{-1/2}u + \hat{\theta}_n) du \right| dt \\
&\quad + \frac{M_1}{\sqrt{2\pi\sigma_n^2\, n}} \int_{-\delta\sqrt{n}}^{0} \frac{e^{-t^2/(2\sigma_n^2)}}{\Pi_n(n^{-1/2}t + \hat{\theta}_n)} \left| \int_{-\delta\sqrt{n}}^{t} h(u) \Pi_n(n^{-1/2}u + \hat{\theta}_n) du \right| dt \\
&\quad + \frac{M_1}{\sqrt{2\pi\sigma_n^2\, n}} \int_{0}^{\delta\sqrt{n}} \frac{e^{-t^2/(2\sigma_n^2)}}{\Pi_n(n^{-1/2}t + \hat{\theta}_n)} \left| \int_{t}^{\delta\sqrt{n}} h(u) \Pi_n(n^{-1/2}u + \hat{\theta}_n) du \right| dt \\
&=: \tilde{I}_{1,1} + \tilde{I}_{1,2} + \tilde{I}_{1,3} + \tilde{I}_{1,4}.
\end{aligned}
$$

Now, note that, for some $c_1, c_2 \in (0,1)$,

$$
\tilde{I}_{1,1} \leq \frac{M_2}{2\sqrt{2\pi\sigma_n^2\, n}} \int_{-\delta\sqrt{n}}^{0} \frac{t^2 e^{-t^2/(2\sigma_n^2)}}{\Pi_n(n^{-1/2}t + \hat{\theta}_n)} \int_{-\delta\sqrt{n}}^{t} |h(u)| \Pi_n(n^{-1/2}u + \hat{\theta}_n)\, du\, dt
$$

$$
\begin{aligned}
=&\frac{M_2}{2\sqrt{2\pi\sigma_n^2\, n}}\int_{-\delta\sqrt{n}}^{0}|h(u)|\int_u^0 t^2 e^{-t^2/(2\sigma_n^2)}\frac{\Pi_n(n^{-1/2}u+\hat{\theta}_n)}{\Pi_n(n^{-1/2}t+\hat{\theta}_n)}\,dt\,du\\
\le&\frac{\widetilde{M}_1\hat{M}_1 M_2}{2\sqrt{2\pi\sigma_n^2\, n}}\int_{-\delta\sqrt{n}}^{0}|h(u)|\int_u^0 t^2 e^{-t^2/(2\sigma_n^2)}\exp\Bigg[L_n(\hat{\theta}_n)+\frac{u}{\sqrt{n}}L_n'(\hat{\theta}_n)+\frac{u^2}{2n}L_n''(\hat{\theta}_n)\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad+\frac{u^3}{6n^{3/2}}L_n'''(\hat{\theta}_n+c_1 n^{-1/2}u)\Bigg]\\
&\cdot\exp\Bigg[-L_n(\hat{\theta}_n)-\frac{t}{\sqrt{n}}L_n'(\hat{\theta}_n)-\frac{t^2}{2n}L_n''(\hat{\theta}_n)-\frac{t^3}{6n^{3/2}}L_n'''(\hat{\theta}_n+c_2 n^{-1/2}t)\Bigg]dt\,du\\
\le&\frac{\widetilde{M}_1\hat{M}_1 M_2}{2\sqrt{2\pi\sigma_n^2 n}}\int_{-\delta\sqrt{n}}^{0}|h(u)|e^{-u^2/(2\sigma_n^2)}e^{\delta M_2 u^2/6}\int_u^0 t^2 e^{\delta M_2 t^2/6}\,dt\,du\\
\le&\frac{3\widetilde{M}_1\hat{M}_1}{\delta\sqrt{2\pi\sigma_n^2\, n}}\int_{-\delta\sqrt{n}}^{0}|uh(u)|\left(e^{-\left(\frac{1}{2\sigma_n^2}-\frac{\delta M_2}{3}\right)u^2}-e^{-\left(\frac{1}{2\sigma_n^2}-\frac{\delta M_2}{6}\right)u^2}\right)du.
\end{aligned}
$$

By a similar argument,

$$
\begin{aligned}
\tilde{I}_{1,2}&\le\frac{3\widetilde{M}_1\hat{M}_1}{\delta\sqrt{2\pi\sigma_n^2\, n}}\int_0^{\delta\sqrt{n}}|uh(u)|\left(e^{-\left(\frac{1}{2\sigma_n^2}-\frac{\delta M_2}{3}\right)u^2}-e^{-\left(\frac{1}{2\sigma_n^2}-\frac{\delta M_2}{6}\right)u^2}\right)du.\\
\tilde{I}_{1,3}&\le\frac{\widetilde{M}_1\hat{M}_1 M_1}{\sqrt{2\pi\sigma_n^2\, n}}\int_{-\delta\sqrt{n}}^{0}|uh(u)|e^{-\left(\frac{1}{2\sigma_n^2}-\frac{\delta M_2}{3}\right)u^2}du.\\
\tilde{I}_{1,4}&\le\frac{\widetilde{M}_1\hat{M}_1 M_1}{\sqrt{2\pi\sigma_n^2\, n}}\int_0^{\delta\sqrt{n}}|uh(u)|e^{-\left(\frac{1}{2\sigma_n^2}-\frac{\delta M_2}{3}\right)u^2}du.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\tilde{I}_1\le&\frac{2\widetilde{M}_1\hat{M}_1\left(M_1+\frac{3}{\delta}\right)}{\sqrt{2\pi\sigma_n^2\, n}}\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}}|uh(u)|e^{-\left(\frac{1}{2\sigma_n^2}-\frac{\delta M_2}{3}\right)u^2}du\\
&-\frac{6\widetilde{M}_1\hat{M}_1}{\delta\sqrt{2\pi\sigma_n^2\, n}}\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}}|uh(u)|e^{-\left(\frac{1}{2\sigma_n^2}-\frac{\delta M_2}{6}\right)u^2}du.
\end{aligned}
\tag{G.6}
$$

Now, by Taylor's expansion:

$$
\begin{aligned}
\frac{n^{-1/2}}{C_n^{MLE}}\left|\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}}g(t)\Pi_n(n^{-1/2}t+\hat{\theta}_n)dt\right|\le&\frac{\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}}|g(t)|\Pi_n(n^{-1/2}t+\hat{\theta}_n)dt}{\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}}\Pi_n(n^{-1/2}u+\hat{\theta}_n)du}\\
\le&\widetilde{M}_1\hat{M}_1\frac{\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}}|g(t)|e^{-(1/(2\sigma_n^2)-\delta M_2/6)t^2}dt}{\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}}e^{-(1/(2\sigma_n^2)+\delta M_2/6)u^2}du}\\
\le&\widetilde{M}_1\hat{M}_1\frac{\sqrt{\frac{1}{2\sigma_n^2}+\frac{\delta M_2}{6}}\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}}|g(t)|e^{-(1/(2\sigma_n^2)-\delta M_2/6)t^2}dt}{\sqrt{2\pi}\left(1-2e^{-\delta^2 n(1/(2\sigma_n^2)+\delta M_2/6)}\right)}.
\end{aligned}
\tag{G.7}
$$

Equations (G.6) and (G.7), together with a standard expression for the normal first absolute moment now yield that

$$\begin{aligned}\tilde{I}_1 \leq & \frac{2\widetilde{M}_1\hat{M}_1}{\sqrt{2\pi\sigma_n^2\, n}}\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} |ug(u)| \left[\left(M_1+\frac{3}{\delta}\right)e^{-\left(\frac{1}{2\sigma_n^2}-\frac{\delta M_2}{3}\right)u^2} - \frac{3}{\delta}e^{-\left(\frac{1}{2\sigma_n^2}-\frac{\delta M_2}{6}\right)u^2}\right]du \\ & + \frac{2\sqrt{\frac{1}{2\sigma_n^2}+\frac{\delta M_2}{6}}\left(\widetilde{M}_1\hat{M}_1\right)^2\left(M_1+\frac{3}{\delta}\right)\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}}|g(u)|e^{-(1/(2\sigma_n^2)-\delta M_2/6)u^2}du}{\left(\frac{1}{2\sigma_n^2}-\frac{\delta M_2}{3}\right)\pi\sqrt{\sigma_n^2}\left(1-2e^{-\delta^2 n(1/(2\sigma_n^2)+\delta M_2/6)}\right)\sqrt{n}} \\ & \cdot\left(\frac{M_1+\frac{3}{\delta}}{\frac{1}{2\sigma_n^2}-\frac{\delta M_2}{3}} - \frac{3}{\delta\left(\frac{1}{2\sigma_n^2}-\frac{\delta M_2}{6}\right)}\right). \qquad (G.8)\end{aligned}$$

G.3 Conclusion

The final bound now follows from Eqs. (G.2) and (G.8).

Appendix H. More detail on the examples

H.1 More detail on Example 1

Note that the function L_n is concave, so its global maximum (i.e. the MLE) $\hat{\theta}_n$ exists as long as the data are not linearly separable. The log-posterior is given by

$$\overline{L}_n(\theta) = -\sum_{i=1}^{n}\rho\left(Y_iX_i^T\theta\right) - \frac{\|\theta\|^2}{2} - \frac{d}{2}\log(2\pi).$$

Note that $\overline{L}_n$ is also concave so its global maximum (i.e. the MAP) $\bar{\theta}_n$ exists as long as the data are not linearly separable. Note also that $L_n''(\theta) = -\sum_{i=1}^n \rho''(Y_iX_i^T\theta)X_iX_i^T$ and $(\log\pi)''(\theta) = -I_{d\times d}$.

H.1.1 Deriving the order of our total variation distance bound

Consider the following two results:

Lemma 38 (cf. Sur et al. 2019, Lemma 4) *Suppose that $d/n < 1$. Let $H(\epsilon) = -\epsilon\log\epsilon - (1-\epsilon)\log(1-\epsilon)$. Then there exists a constant ϵ_0 such that for all $0 \leq \epsilon \leq \epsilon_0$, with probability at least $1 - 2\exp(-nH(\epsilon)) - 2\exp(-n/2)$, the following matrix inequality*

$$-\frac{L_n''(\theta)}{n} \succeq \frac{\exp(3\|\theta\|/\sqrt{\epsilon})}{\left(1+\exp(3\|\theta\|/\sqrt{\epsilon})\right)^2}\left(\sqrt{1-\epsilon} - \frac{d}{n} - 2\sqrt{\frac{H(\epsilon)}{1-\epsilon}}\right)I_{d\times d}$$

holds simultaneously for all $\theta \in \mathbb{R}^d$.

Proof This follows directly from Sur et al. (2019, Lemma 4), using the fact that

$$\inf_{z:|z|\leq\frac{3\|\theta\|}{\sqrt{\epsilon}}}\rho''(z) = \frac{\exp(3\|\theta\|/\sqrt{\epsilon})}{\left(1+\exp(3\|\theta\|/\sqrt{\epsilon})\right)^2}.$$

■

Lemma 39 (cf. Katsevich 2024, Corollary 6.4) *Fix any small constant $\epsilon > 0$. Then, there exist universal constants $c_1, c_2, C_2 > 0$, independent of n or d such that if $d^{3/2} \le n$ then, for large enough n, the MLE $\hat{\theta}_n$ associated to L_n satisfies*

$$\|\hat{\theta}_n\| < c_1,$$

with probability at least $1 - \exp\big(-c_2 (nd)^{1/8}\big) - e^{-C_2 n^{1/6}}$.

Proof This follows directly from Katsevich (2024, Corollary 6.4) upon choosing $s = c_3 n^{1/6}$ in their setup, for a sufficiently small absolute constant $c_3 > 0$ not depending on n or d. ■

Note that functions $\frac{L_n}{n}$ and $\frac{\overline{L}_n}{n}$ are continuously differentiable and strictly concave and converge almost surely to the same (concave) limit as $n \to \infty$. Therefore, if $\|\hat{\theta}_n\|$ is uniformly bounded, then so is $\|\bar{\theta}_n\|$. Therefore, by Lemma 39, if $d^{3/2} \le n$ then $\|\bar{\theta}_n\|$ is upper-bounded, uniformly in n and d, with high probability. As $\overline{L}_n''(\theta) = L_n''(\theta) - I_{d\times d}$, we can now apply Lemma 38 to conclude that, when $d^{3/2} \le n$, then $\overline{\lambda}_{\min}(\bar{\theta}_n)$ is lower bounded uniformly in d and n, with high probability.

Now, consider the following result, which is a direct consequence of Katsevich (2023a, Lemma 3.2):

Lemma 40 (cf. Katsevich 2023a, Lemma 3.2) *Suppose that $d \le n \le e^{\sqrt{d}}$. Then, there exist absolute constants $B_1, B_2, C > 0$, such that*

$$\sup_{\theta \in \mathbb{R}^d} \left\| \frac{\overline{L}_n'''(\theta)}{n} \right\| \le C \left(1 + \frac{d^{3/2}}{n}\right); \qquad \sup_{\theta \in \mathbb{R}^d} \left\| \frac{L_n'''(\theta)}{n} \right\| \le C \left(1 + \frac{d^{3/2}}{n}\right)$$

with probability at least $1 - B_1 \exp\Big(-B_2 \sqrt{nd} / \log(2n/d)\Big)$.

As discussed in Section 4.2 and in Appendix H.1.3 below, as long as $\frac{d \log n}{n} \xrightarrow{n\to\infty} 0$, the leading term in in our bound on the total variation distance from Theorem 17 is of order $C_d \sqrt{d^2/n}$, where $C_d \le \frac{\overline{M}_2}{\overline{\lambda}_{\min}(\bar{\theta}_n)\sqrt{\overline{\lambda}_{\min}(\bar{\theta}_n) - \bar{\delta}\overline{M}_2}}$. As described above, we have established via Lemmas 38 and 39 that $\overline{\lambda}_{\min}(\bar{\theta}_n)$ is lower-bounded with high probability if $n \ge d^{3/2}$. We have also established via Lemma 40 that $\overline{M}_2$ is upper-bounded with high probability. Moreover, we can choose $\bar{\delta} > 0$ to be independent of n and d and arbitrarily small. Therefore C_d is with high probability upper-bounded by a finite constant not depending on d or n. It follows that the leading order term in our bound on the total variation distance from Theorem 17 is with high probability upper bounded by a universal constant multiplied by $\sqrt{d^2/n}$.

H.1.2 Lower bound on the effective dimension

Now, we shall show that, in the setup of Example 1,

$$d_{eff} := \mathrm{Tr}\left\{ \left(\bar{J}_n(\bar{\theta}_n) + \frac{(\log \pi)''(\bar{\theta}_n)}{n} \right) \bar{J}_n(\bar{\theta}_n)^{-1} \right\} \ge d \left(1 - \frac{1}{n\, \overline{\lambda}_{\min}(\bar{\theta}_n)}\right),$$

where d_{eff} denotes the *effective dimension* introduced in Spokoiny (2022). Note that

$$
\begin{aligned}
d_{eff} =& \operatorname{Tr}\left\{\left(\bar{J}_n(\bar{\theta}_n) - \frac{1}{n} I_{d\times d}\right)\bar{J}_n(\bar{\theta}_n)^{-1}\right\} \\
=& \operatorname{Tr}\left\{I_{d\times d} - \frac{1}{n}\bar{J}_n(\bar{\theta}_n)^{-1}\right\} \\
=& d - \frac{1}{n}\operatorname{Tr}\left\{\bar{J}_n(\bar{\theta}_n)^{-1}\right\} \\
\geq& d\left(1 - \frac{1}{n\,\overline{\lambda}_{\min}(\bar{\theta}_n)}\right).
\end{aligned}
$$

The fact that $\overline{\lambda}_{\min}(\bar{\theta}_n)$ is lower-bounded by a positive number not depending on n of d with high probability follows from Lemmas 38 and 39 presented above and is discussed directly below Lemma 39 in Appendix H.1.1 above.

H.1.3 Checking that the model satisfies the assumptions of Theorem 17

We retain the assumption that $2d < n < e^{\sqrt{d}}$ and assume that $\frac{d\log n}{n} \xrightarrow{n\to\infty} 0$. As discussed above, the MLE $\hat{\theta}_n$ and MAP $\bar{\theta}_n$ exist and are bounded with high probability by Lemma 39. Assumption 1 is satisfied with high probability, (for instance with $\delta = 1$), with M_2 not depending on n or d, by Lemma 40. The Gaussian prior trivially satisfies Assumption 2 for any $\delta \leq 1$. Moreover, $\hat{M}_1$ is with high probability bounded by a constant multiplied by $(2\pi)^{d/2}$, as $\|\hat{\theta}_n\|$ is upper bounded by a constant independent of n and d with high probability. By Lemma 40, with high probability Assumption 3 is also satisfied for any fixed $\bar{\delta}$ not depending on n or d and with $\overline{M}_2$ not depending on n or d.

As discussed above, $\frac{\overline{L}_n}{n}$ and $\frac{L_n}{n}$ are continuously differentiable and concave and converge almost surely to the same (concave) limit as $n \to \infty$. By Lemma 39, $\hat{\theta}_n$ and $\bar{\theta}_n$ are bounded with high probability. Therefore, with high probability, $\|\bar{\theta}_n - \hat{\theta}_n\| < \bar{\delta}$, for any fixed $\bar{\delta} > 0$ and for large enough n. Moreover, with high probability, $\overline{\lambda}_{\min}(\bar{\theta}_n)$ and $\lambda_{\min}(\hat{\theta}_n)$ are lower bounded by a positive number not depending on n or d, by Lemmas 38 and 39. Moreover $\frac{\operatorname{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]}{n} \leq \frac{d}{n\,\overline{\lambda}_{\min}(\bar{\theta}_n)}$ and $\frac{\operatorname{Tr}\left[\hat{J}_n(\hat{\theta}_n)^{-1}\right]}{n} \leq \frac{d}{n\,\lambda_{\min}(\hat{\theta}_n)}$. Therefore, with high probability, Assumption 4 is satisfied, for large enough n and for any fixed δ and $\bar{\delta}$, not depending on n or d.

Recall that $\overline{\lambda}_{\min}(\bar{\theta}_n)$ is lower bounded by a positive number not depending on n or d with high probability, if $d < \frac{n}{2}$. Therefore, Assumption 5 is satisfied for small enough $\bar{\delta}$. Assume n is large enough so that $\bar{\delta} - \|\hat{\theta}_n - \bar{\theta}_n\| > 0$. Assumption 6 is satisfied with high probability by the following reasoning, which uses the strict concavity of L_n:

$$
\begin{aligned}
&\sup_{\theta:\|\theta-\hat{\theta}_n\|>\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|} \frac{L_n(\theta) - L_n(\hat{\theta}_n)}{n} \leq \sup_{\theta:\|\theta-\hat{\theta}_n\|=\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|} \frac{L_n(\theta) - L_n(\hat{\theta}_n)}{n} \\
&\leq \sup_{\theta:\|\theta-\hat{\theta}_n\|=\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|} \left\{-\frac{1}{2}\left(\theta - \hat{\theta}_n\right)^T \hat{J}_n(\hat{\theta}_n)\left(\theta - \hat{\theta}_n\right)\right\} + \frac{M_2\left(\bar{\delta} - \|\hat{\theta}_n - \bar{\theta}_n\|\right)^3}{6} \\
&\leq -\frac{1}{2}\lambda_{\min}(\hat{\theta}_n)\left(\bar{\delta} - \|\hat{\theta}_n - \bar{\theta}_n\|\right)^2 + \frac{M_2\left(\bar{\delta} - \|\hat{\theta}_n - \bar{\theta}_n\|\right)^3}{6} =: -\bar{\kappa}.
\end{aligned}
$$

With high probability, $\bar{\kappa}$ is lower bounded by a positive number not depending on n or d, for any fixed $\bar{\delta}$ independent of n and d, as long as n is large enough.

Now, we have shown that we can make choices of $\bar{\delta}$ and $\bar{\kappa}$ independent of n and d, such that with high probability the assumptions of Theorem 17 hold for large enough n. For such a choice of $\bar{\delta}$ we have that $\bar{\delta} \gg \frac{\sqrt{\log n}}{\sqrt{n\bar{\lambda}_{\min}(\bar{\theta}_n)}}$. By our assumption $\frac{d\log n}{n} \xrightarrow{n\to\infty} 0$, we also have that $\bar{\kappa} \gg \frac{\log n}{n} \cdot \frac{d+1}{2}$. Moreover, we have established that $\hat{M}_1$ is with high probability bounded by a universal constant multiplied by $(2\pi)^{d/2}$. Also, note that

$$\left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right| = \left|\det\left(\frac{\sum_{i=1}^n \rho''(Y_iX_i^T\hat{\theta}_n)X_iX_i^T}{n} + (\delta M_2/3)I_{d\times d}\right)\right|$$

Note that

$$\left\|\frac{\sum_{i=1}^n \rho''(Y_iX_i^T\hat{\theta}_n)X_iX_i^T}{n}\right\|_{op} \leq \frac{1}{4n}\sum_{i=1}^n \left\|X_iX_i^T\right\|_{op} \leq \frac{1}{4n}\sum_{i=1}^n \|X_i\|^2.$$

Therefore,

$$\log\left[\left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2}\right] \leq \frac{d}{2}\log\left[\frac{1}{4n}\sum_{i=1}^n \|X_i\|^2 + \delta M_2/3\right].$$

It follows that $\bar{\kappa} \gg \frac{1}{n}\log\left(\hat{M}_1\left|\det\left(\hat{J}_n^p(\hat{\theta}_n,\delta)\right)\right|^{1/2}\right)$ with high probability. Hence, by the discussion of Section 4.2, with high probability, the bound in Theorem 17 is of the same order as that of the first summand $A_1 n^{-1/2}$.

H.2 Calculations for Section 7.1

H.2.1 The MLE-centric approach

Let $X_1,\ldots,X_n \geq 0$ be our data and assume that their sum is positive. We have

$$L_n(\theta) = -n\theta + (\log\theta)\left(\sum_{i=1}^n X_i\right) - \sum_{i=1}^n \log(X_i!)$$

The MLE is given by $\hat{\theta}_n = \overline{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$. Also,

$$L_n'(\theta) = -n + \frac{n\overline{X}_n}{\theta}, \quad L_n''(\theta) = -\frac{n\overline{X}_n}{\theta^2}, \quad L_n'''(\theta) = \frac{2n\overline{X}_n}{\theta^3}.$$

We have $\sigma_n^2 := \hat{J}_n(\hat{\theta}_n)^{-1} = \hat{\theta}_n^2/\ \overline{X}_n = \left|\overline{X}_n\right|$.

Now, for $c \in (0,1)$ and $\delta = c\hat{\theta}_n$, we have that for $\theta \in (\hat{\theta}_n - \delta, \hat{\theta}_n + \delta) = \left(\overline{X}_n - c\overline{X}_n, \overline{X}_n + c\overline{X}_n\right)$,

$$\frac{|L_n'''(\theta)|}{n} = \frac{2\overline{X}_n}{|\theta|^3} \leq \frac{2}{(1-c)^3\left(\overline{X}_n\right)^2} =: M_2$$

Moreover, for θ, such that $\left|\theta - \hat{\theta}_n\right| > \delta = c\overline{X}_n$, i.e. for $\theta > \overline{X}_n + c\overline{X}_n$ or $\theta < \overline{X}_n - c\overline{X}_n$,

$$\begin{aligned}\frac{L_n(\theta) - L_n(\hat{\theta}_n)}{n} \leq& \max\left\{\frac{L_n((1+c)\overline{X}_n) - L_n(\overline{X}_n)}{n}, \frac{L_n((1-c)\overline{X}_n) - L_n(\overline{X}_n)}{n}\right\} \\ \leq& \overline{X}_n \cdot \max\left\{\log(1+c) - c, \log(1-c) + c\right\} = \left[\log(1+c) - c\right]\overline{X}_n =: -\kappa.\end{aligned}$$

We also need to make sure that $\hat{J}_n(\hat{\theta}_n) > \delta M_2$, i.e. that $\frac{1}{\overline{X}_n} > \frac{2c}{(1-c)^3\overline{X}_n}$, which is true for all $0 < c \leq 0.229$.

Now, the gamma prior with shape α and rate β satisfies

$$\pi'(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\left[(\alpha-1)\theta^{\alpha-2}e^{-\beta\theta} - \beta\theta^{\alpha-1}e^{-\beta\theta}\right].$$

Suppose that $\alpha < 1$, then

$$\begin{aligned}\sup_{\theta\in((1-c)\overline{X}_n,(1+c)\overline{X}_n)} |\pi'(\theta)| \leq& \frac{\beta^\alpha}{\Gamma(\alpha)}e^{-\beta(1-c)\overline{X}_n}\overline{X}_n^{\alpha-2}(1-c)^{\alpha-2}\left[(1-\alpha) + \beta(1-c)\overline{X}_n\right] \\ \sup_{\theta\in((1-c)\overline{X}_n,(1+c)\overline{X}_n)} |\pi(\theta)| \leq& \frac{\beta^\alpha}{\Gamma(\alpha)}(1-c)^{\alpha-1}\overline{X}_n^{\alpha-1}e^{-\beta(1-c)\overline{X}_n} \\ \sup_{\theta\in((1-c)\overline{X}_n,(1+c)\overline{X}_n)} \frac{1}{|\pi(\theta)|} \leq& \frac{\Gamma(\alpha)}{\beta^\alpha}(1+c)^{1-\alpha}\overline{X}_n^{1-\alpha}e^{\beta(1+c)\overline{X}_n}.\end{aligned}$$

Finally, the bounds in the MLE-centric approach are computed under the assumption $\sqrt{\frac{\operatorname{Tr}\left[\hat{J}_n(\hat{\theta}_n)^{-1}\right]}{n}} < \delta$. In our case, it says $\sqrt{\frac{\overline{X}_n}{n}} < c\overline{X}_n$, i.e. that $c > \frac{1}{\sqrt{n\overline{X}_n}}$. Therefore, assuming $\frac{1}{\sqrt{n\overline{X}_n}} < 0.299$, letting $c \in \left(\frac{1}{\sqrt{n\overline{X}_n}}, 0.229\right]$, and assuming the shape α of the gamma prior is smaller than one, we can set

a) $\delta = c\overline{X}_n$

b) $M_1 = (1-c)^{\alpha-2}(1+c)^{1-\alpha}\overline{X}_n^{-1}e^{2\beta c\overline{X}_n}\left[(1-\alpha) + \beta(1-c)\overline{X}_n\right]$

c) $\widetilde{M_1} = \frac{\beta^\alpha}{\Gamma(\alpha)}(1-c)^{\alpha-1}\overline{X}_n^{\alpha-1}e^{-\beta(1-c)\overline{X}_n}$

d) $\hat{M}_1 = \frac{\Gamma(\alpha)}{\beta^\alpha}(1+c)^{1-\alpha}\overline{X}_n^{1-\alpha}e^{\beta(1+c)\overline{X}_n}$

e) $M_2 = \frac{2}{(1-c)^3(\overline{X}_n)^2}$

f) $\hat{J}_n(\hat{\theta}_n) = \overline{X}_n^{-1}$

g) $\kappa = \left[c - \log(1+c)\right]\overline{X}_n$.

The concrete choice of c may be optimized numerically.

H.2.2 The MAP-centric approach

Let us still assume $\alpha < 1$. We note that

$$\begin{aligned}
\overline{L}_n(\theta) &= -n\theta + n(\log\theta)\overline{X}_n - \sum_{i=1}^n \log(X_i!) + \alpha\log(\beta) - \log(\Gamma(\alpha)) + (\alpha-1)\log\theta - \beta\theta; \\
\overline{L}_n'(\theta) &= -n + \frac{n\overline{X}_n}{\theta} + \frac{(\alpha-1)}{\theta} - \beta = 0 \quad \text{iff} \quad \theta = \bar{\theta}_n := \frac{n\overline{X}_n + (\alpha-1)}{(n+\beta)}; \\
\overline{L}_n''(\theta) &= -\frac{n\overline{X}_n + \alpha - 1}{\theta^2} \quad \text{and so} \quad \bar{J}_n(\bar{\theta}_n) = \frac{(n+\beta)^2}{n(n\overline{X}_n + \alpha - 1)}; \\
\overline{L}_n'''(\theta) &= 2\frac{n\overline{X}_n + \alpha - 1}{\theta^3}
\end{aligned}$$

and so for $\bar{c} \in (0,1), \bar{\delta} = \bar{c}\bar{\theta}_n$ and $\theta \in (\bar{\theta}_n - \bar{\delta}, \bar{\theta}_n + \bar{\delta}) = \big((1-\bar{c})\bar{\theta}_n, (1+\bar{c})\bar{\theta}_n\big)$, we have that

$$\frac{1}{n}|\overline{L}_n'''(\theta)| \leq \frac{2(n+\beta)^3}{n(n\overline{X}_n + \alpha - 1)^2(1-\bar{c})^3} =: \overline{M}_2.$$

Now, we require that $\bar{J}_n(\bar{\theta}_n) > \bar{\delta}\,\overline{M}_2$, i.e. that

$$\frac{(n+\beta)^2}{n(n\overline{X}_n + \alpha - 1)} > \frac{2\bar{c}(n+\beta)^2}{n(n\overline{X}_n + \alpha - 1)(1-\bar{c})^3}, \quad \text{which holds for } \bar{c} \in (0, 0.229], \text{ if } n\overline{X}_n > 1 - \alpha.$$

We will also want to make sure that $\bar{\delta} = \bar{c}\frac{n\overline{X}_n + (\alpha-1)}{n+\beta} \geq \|\bar{\theta}_n - \hat{\theta}_n\| = \frac{\beta\overline{X}_n + 1 - \alpha}{n+\beta}$. Assuming that $n\overline{X}_n > 1 - \alpha$, this translates to $\bar{c} \geq \frac{1}{n} \cdot \frac{\beta\overline{X}_n + 1 - \alpha}{\overline{X}_n + (\alpha-1)/n}$. Moreover, we require that $\bar{\delta} = \bar{c}\frac{n\overline{X}_n + (\alpha-1)}{n+\beta} > \sqrt{\frac{1}{n\bar{J}_n(\bar{\theta}_n)}} = \sqrt{\frac{n\overline{X}_n + \alpha - 1}{(n+\beta)^2}}$, which is equivalent to saying that $\bar{c} > \sqrt{\frac{1}{n\overline{X}_n + \alpha - 1}}$.

Finally, the value of $\bar{\kappa}$ may be obtained in the following way:

$$\begin{aligned}
\bar{\kappa} := -\max\Bigg\{ &-\bar{\delta} + \|\bar{\theta}_n - \hat{\theta}_n\| + \hat{\theta}_n \log\left(\frac{\hat{\theta}_n + \bar{\delta} - \|\bar{\theta}_n - \hat{\theta}_n\|}{\hat{\theta}_n}\right), \\
&\bar{\delta} - \|\bar{\theta}_n - \hat{\theta}_n\| + \hat{\theta}_n \log\left(\frac{\hat{\theta}_n - \bar{\delta} + \|\bar{\theta}_n - \hat{\theta}_n\|}{\hat{\theta}_n}\right)\Bigg\} \\
= -\overline{X}_n &\left\{\frac{\beta - \bar{c}n + (1-\alpha)(1+\bar{c})/\overline{X}_n}{n+\beta} + \log\left(1 + \frac{\bar{c}n - \beta - (1-\alpha)(1+\bar{c})/\overline{X}_n}{n+\beta}\right)\right\}.
\end{aligned}$$

Therefore, in addition to the values we listed at the end of Appendix H.2.2, we have the following. We assume that $\alpha < 1$, $n\overline{X}_n > 1 - \alpha$ and $\max\left\{\sqrt{\frac{1}{n\overline{X}_n + \alpha - 1}}, \frac{1}{n} \cdot \frac{\beta\overline{X}_n + 1 - \alpha}{\overline{X}_n + (\alpha-1)/n}\right\} < 0.229$. We let $\bar{c} \in \left(\frac{1}{n} \cdot \frac{\beta\overline{X}_n + 1 - \alpha}{\overline{X}_n + (\alpha-1)/n}, 0.229\right] \cap \left(\sqrt{\frac{1}{n\overline{X}_n + \alpha - 1}}, 0.229\right]$. Then

i) $\bar{\delta} = \bar{c}\frac{n\overline{X}_n + (\alpha-1)}{(n+\beta)}$

ii) $\bar{J}_n(\bar{\theta}_n) = \frac{(n+\beta)^2}{n(n\overline{X}_n + \alpha - 1)}$

iii) $\overline{M}_2 = \frac{2\cdot(n+\beta)^3}{(1-\bar{c})^3 n(n\overline{X}_n+\alpha-1)^2}$

iv) $\bar{\kappa} = -\overline{X}_n\left\{\frac{\beta-\bar{c}n+(1-\alpha)(1+\bar{c})/\overline{X}_n}{n+\beta} + \log\left(1+\frac{\bar{c}n-\beta-(1-\alpha)(1+\bar{c})/\overline{X}_n}{n+\beta}\right)\right\}.$

H.3 Calculations for Section 7.2: the MAP-centric approach

Let k be the shape of the Weibull and let $X_1,\ldots,X_n \geq 0$ be our data. Our log-likelihood is given by:

$$L_n(\theta) = n\left[\log(k) - \log(\theta)\right] + (k-1)\sum_{i=1}^n \log(X_i) - \frac{\sum_{i=1}^n X_i^k}{\theta}$$

H.3.1 Calculating $\hat{\theta}_n$ and $\hat{J}_n(\hat{\theta}_n)$

Now

$$L_n'(\theta) = -\frac{n}{\theta} + \frac{\sum_{i=1}^n X_i^k}{\theta^2}, \quad L_n''(\theta) = \frac{n}{\theta^2} - \frac{2\sum_{i=1}^n X_i^k}{\theta^3}$$

The MLE is $\hat{\theta}_n = \frac{\sum_{i=1}^n X_i^k}{n} =: \overline{X^k}(n)$. We have that

$$\hat{J}_n(\hat{\theta}_n) = -\frac{1}{\hat{\theta}_n^2} + \frac{2}{\hat{\theta}_n^2} = \frac{1}{\hat{\theta}_n^2}.$$

H.3.2 Calculating M_2

Now, note that

$$L_n'''(\theta) = \frac{6\sum_{i=1}^n X_i^k - 2n\theta}{\theta^4}, \quad L_n^{(4)}(\theta) = \frac{6n\theta - 24\sum_{i=1}^n X_i^k}{\theta^5}, \quad L_n^{(5)}(\theta) = \frac{120\sum_{i=1}^n X_i^k - 24n\theta}{\theta^6}.$$

Therefore, for $0 < \theta < 2\hat{\theta}_n$, L_n''' is decreasing and positive. This means that, if we let $0 < \delta < \hat{\theta}_n$, then

$$\sup_{\theta\in(\hat{\theta}_n-\delta,\hat{\theta}_n+\delta)} \frac{|L_n'''(\theta)|}{n} \leq \frac{L_n'''(\hat{\theta}_n-\delta)}{n} =: M_2.$$

H.3.3 Calculating $\hat{M}_1$

Now, for a given shape $\alpha > 0$ and scale $\beta > 0$,

$$\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\left(\frac{1}{\theta}\right)^{\alpha+1}\exp\left(-\beta/\theta\right), \quad \pi'(\theta) = \frac{\beta^\alpha \exp\left(-\beta/\theta\right)}{\Gamma(\alpha)\theta^{\alpha+2}}\left(-\alpha-1+\frac{\beta}{\theta}\right)$$

and it follows that, for $\theta > \frac{\beta}{\alpha+1}$, $\frac{1}{\pi(\theta)}$ is increasing and it is decreasing otherwise. Therefore:

$$\sup_{\theta\in(\hat{\theta}_n-\hat{\delta},\hat{\theta}_n+\delta)}\left|\frac{1}{\pi(\theta)}\right| \leq \max\left\{\frac{1}{\pi(\hat{\theta}_n-\delta)}, \frac{1}{\pi(\hat{\theta}_n+\delta)}\right\} =: \hat{M}_1.$$

H.3.4 Calculating $\bar{\theta}_n$ and $\bar{J}_n(\bar{\theta}_n)$

Now

$$\overline{L}_n(\theta) = n\left[\log(k) - \log(\theta)\right] + (k-1)\sum_{i=1}^n \log(X_i) - \frac{\sum_{i=1}^n X_i^k}{\theta} - (\alpha+1)\log(\theta) - \frac{\beta}{\theta}.$$

Therefore,

$$\begin{aligned}\overline{L}_n(\theta)' &= -\frac{n}{\theta} + \frac{\sum_{i=1}^n X_i^k}{\theta^2} - \frac{\alpha+1}{\theta} + \frac{\beta}{\theta^2}\\ \overline{L}_n(\theta)'' &= \frac{n}{\theta^2} - \frac{2\sum_{i=1}^n X_i^k}{\theta^3} + \frac{\alpha+1}{\theta^2} - \frac{2\beta}{\theta^3}\end{aligned}$$

and the MAP $\bar{\theta}_n$ is given by

$$(n+\alpha+1)\bar{\theta}_n = \beta + \sum_{i=1}^n X_i^k \quad \Leftrightarrow \quad \bar{\theta}_n = \frac{\beta + \sum_{i=1}^n X_i^k}{n+\alpha+1}.$$

Moreover,

$$\bar{J}_n(\bar{\theta}_n) = -\frac{1}{\bar{\theta}_n^2} + \frac{2\sum_{i=1}^n X_i^k}{n\bar{\theta}_n^3} - \frac{\alpha+1}{n\bar{\theta}_n^2} + \frac{2\beta}{n\bar{\theta}_n^3}.$$

H.3.5 Calculating $\bar{M}_2$

Now, note that

$$\overline{L}_n'''(\theta) = -\frac{2n}{\theta^3} + \frac{6\sum_{i=1}^n X_i^k}{\theta^4} - \frac{2(\alpha+1)}{\theta^3} + \frac{6\beta}{\theta^4}, \quad \overline{L}_n^{(4)}(\theta) = \frac{6n}{\theta^4} - \frac{24\sum_{i=1}^n X_i^k}{\theta^5} + \frac{6(\alpha+1)}{\theta^4} - \frac{24\beta}{\theta^5}.$$

Therefore $\overline{L}_n'''$ is increasing if and only if

$$6(n+\alpha+1)\theta > 24\left(\beta + \sum_{i=1}^n X_i^k\right) \quad \Leftrightarrow \theta > 4\bar{\theta}_n$$

This means that, for $\bar{\delta} \in \left(0, \bar{\theta}_n\right)$ and $\theta \in (\bar{\theta}_n - \bar{\delta}, \bar{\theta}_n + \bar{\delta})$, $\overline{L}_n'''$ is decreasing and

$$\sup_{|\theta-\bar{\theta}_n|<\bar{\delta}} \frac{|\overline{L}_n'''(\theta)|}{n} \leq \frac{|\overline{L}_n'''(\bar{\theta}_n - \bar{\delta})|}{n} =: \overline{M}_2.$$

H.3.6 Calculating $\bar{\kappa}$

Now, note that, for $\theta < \hat{\theta}_n$, L_n, is increasing and otherwise it's decreasing. This means that,

$$\begin{aligned}&\sup_{|\theta-\hat{\theta}_n|>\bar{\delta}-|\hat{\theta}_n-\bar{\theta}_n|} \frac{L_n(\theta) - L_n(\hat{\theta}_n)}{n}\\ &\leq \max\left\{\frac{L_n(\hat{\theta}_n - \bar{\delta} + |\hat{\theta}_n - \bar{\theta}_n|) - L_n(\hat{\theta}_n)}{n}, \frac{L_n(\hat{\theta}_n + \bar{\delta} - |\hat{\theta}_n - \bar{\theta}_n|) - L_n(\hat{\theta}_n)}{n}\right\} =: -\bar{\kappa}.\end{aligned}$$

H.3.7 Constraints on $\bar{\delta}$ and δ

Finally, we need the derive the constraints on $\bar{\delta}$. We first assume that $0 < \bar{\delta} < \bar{\theta}_n$. We also suppose that $\bar{\delta} > \max\left(\|\bar{\theta}_n - \hat{\theta}_n\|, \frac{1}{\sqrt{n\bar{J}_n(\bar{\theta}_n)}}\right)$ and conditions on how large n needs to be for this to hold are easy to obtain numerically. We also require $\bar{\delta} < \frac{\lambda(\bar{\theta}_n)}{\overline{M}_2}$. Looking closer at this last condition, we require that:

$$\bar{\delta} < \overline{\lambda}_{\min}(\bar{\theta}_n)\left(\bar{\theta}_n - \bar{\delta}\right)^4 \left[\left(-2 - \frac{2\alpha+2}{n}\right)\left(\bar{\theta}_n - \bar{\delta}\right) + 6\left(\frac{\beta}{n} + \hat{\theta}_n\right)\right]^{-1}$$

Letting $0 < \tilde{\delta} := \bar{\theta}_n - \bar{\delta} < \bar{\theta}_n$, we therefore require

$$\tilde{\delta} + \overline{\lambda}_{\min}(\bar{\theta}_n)\tilde{\delta}^4 \left[\left(-2 - \frac{2\alpha+2}{n}\right)\tilde{\delta} + 6\left(\frac{\beta}{n} + \hat{\theta}_n\right)\right]^{-1} > \bar{\theta}_n$$

which is equivalent to:

$$\tilde{\delta} + \tilde{\delta}^4\overline{\lambda}_{\min}(\bar{\theta}_n)\left(2 + \frac{2\alpha+2}{n}\right)^{-1}\left[3\bar{\theta}_n - \tilde{\delta}\right]^{-1} > \bar{\theta}_n.$$

This condition will be satisfied if

$$\tilde{\delta} + \tilde{\delta}^4\overline{\lambda}_{\min}(\bar{\theta}_n)\left(2 + \frac{2\alpha+2}{n}\right)^{-1}\left[3\bar{\theta}_n\right]^{-1} > \bar{\theta}_n. \tag{H.1}$$

The left-hand side of Eq. (H.1) is increasing in $\tilde{\delta}$ and is clearly strictly greater than $\bar{\theta}_n$ for $\tilde{\delta} = \bar{\theta}_n$. This means that there exists a choice of $\bar{\delta}$ that yields $\bar{\delta} < \frac{\overline{\lambda}_{\min}(\bar{\theta}_n)}{\overline{M}_1}$ and the set of such choices can be obtained by solving Eq. (H.1) numerically. Finally, in order to make sure that the condition on δ is satisfied, we just need to check numerically how large n needs to be so that $\delta > \frac{1}{\sqrt{n\hat{J}_n(\hat{\theta}_n)}}$.

H.4 Calculations for Section 7.3

H.4.1 Calculating $\hat{J}_n(\hat{\theta}_n)$ and $\bar{J}_n(\bar{\theta}_n)$

Note that:

$$\begin{aligned}
\hat{J}_n(\hat{\theta}_n) =& \frac{1}{n}\sum_{k=1}^{n} \frac{e^{X_k^T\hat{\theta}_n Y_k}}{(1+e^{X_k^T\hat{\theta}_n Y_k})^2} X_k\left(X_k\right)^T; \\
\bar{J}_n(\bar{\theta}_n) =& \frac{1}{n}\sum_{k=1}^{n} \frac{e^{X_k^T\bar{\theta}_n Y_k}}{(1+e^{X_k^T\bar{\theta}_n Y_k})^2} X_k\left(X_k\right)^T \\
&+ \frac{\nu+d}{2\nu n}\left[1 + \frac{1}{\nu}(\bar{\theta}_n - \mu)^T\Sigma^{-1}(\bar{\theta}_n - \mu)\right]^{-1}\left(\Sigma^{-1} + \mathrm{diag}\left(\Sigma^{-1}\right)\right) \\
&- \frac{\nu+d}{2\nu^2 n}\left[1 + \frac{1}{\nu}(\bar{\theta}_n - \mu)^T\Sigma^{-1}(\bar{\theta}_n - \mu)\right]^{-2} \\
&\qquad \cdot \left[\left(\Sigma^{-1} + \mathrm{diag}\left(\Sigma^{-1}\right)\right)\left(\bar{\theta}_n - \mu\right)\right]\left[\left(\Sigma^{-1} + \mathrm{diag}\left(\Sigma^{-1}\right)\right)\left(\bar{\theta}_n - \mu\right)\right]^T.
\end{aligned}$$

H.4.2 Calculating M_1, $\widetilde{M}_1$ and $\hat{M}_1$

Note that

$$\begin{aligned}\pi'(\theta) =&\frac{\Gamma((\nu+d)/2)}{\Gamma(\nu/2)\nu^{d/2}\pi^{d/2}|\Sigma|^{1/2}}\left(-\frac{\nu+d}{2}\right)\left[1+\frac{1}{\nu}(\theta-\mu)^T\Sigma^{-1}(\theta-\mu)\right]^{-(\nu+d)/2-1}\\ &\cdot\left[\frac{1}{\nu}\left(\Sigma^{-1}+\operatorname{diag}\left(\Sigma^{-1}_{1,1},\ldots,\Sigma^{-1}_{d,d}\right)\right)\right](\theta-\mu). \qquad \text{(H.2)}\end{aligned}$$

It follows from the expressions in Eqs. (7.2) and (H.2) that

$$\begin{aligned}\sup_{\theta:\|\theta-\hat{\theta}\|<\delta}|\pi(\theta)| \leq&\frac{\Gamma((\nu+d)/2)}{\Gamma(\nu/2)\nu^{d/2}\pi^{d/2}|\Sigma|^{1/2}} =: \widetilde{M}_1;\\ \sup_{\theta:\|\theta-\hat{\theta}\|<\delta}\frac{1}{|\pi(\theta)|} \leq& \sup_{\theta:\|\theta-\hat{\theta}\|<\delta}\frac{\Gamma(\nu/2)\nu^{d/2}\pi^{d/2}|\Sigma|^{1/2}}{\Gamma((\nu+d)/2)}\left[1+\frac{1}{\nu\lambda_{\min}(\Sigma)}\|\theta-\mu\|^2\right]^{(\nu+d)/2}\\ \leq&\frac{\Gamma(\nu/2)\nu^{d/2}\pi^{d/2}|\Sigma|^{1/2}}{\Gamma((\nu+d)/2)}\left[1+\frac{2\delta^2+2\|\hat{\theta}-\mu\|^2}{\nu\lambda_{\min}(\Sigma)}\right]^{(\nu+d)/2} =: \hat{M}_1;\\ \sup_{\theta:\|\theta-\hat{\theta}\|<\delta}\frac{\|\pi'(\theta)\|}{|\pi(\theta)|} =& \sup_{\theta:\|\theta-\hat{\theta}\|<\delta}\left(\frac{\nu+d}{2}\right)\frac{\left\|\frac{1}{\nu}\left(\Sigma^{-1}+\operatorname{diag}\left(\Sigma^{-1}_{1,1},\ldots,\Sigma^{-1}_{d,d}\right)\right)(\theta-\mu)\right\|}{\left[1+\frac{1}{\nu}(\theta-\mu)^T\Sigma^{-1}(\theta-\mu)\right]}\\ \leq&\left(\frac{\nu+d}{2}\right)\frac{\left(\delta+\|\hat{\theta}-\mu\|\right)\left\|\Sigma^{-1}+\operatorname{diag}\left(\Sigma^{-1}_{1,1},\ldots,\Sigma^{-1}_{d,d}\right)\right\|}{\nu}\\ \leq&\frac{(\nu+d)\left(\delta+\|\hat{\theta}-\mu\|\right)}{\nu\lambda_{\min}(\Sigma)} =: M_1.\end{aligned}$$

H.4.3 Calculating M_2 and $\overline{M}_2$

Note that, for all $\theta\in\mathbb{R}^d$,

$$L_n'''(\theta)[u_1,u_2,u_3] = \sum_{i=1}^n Y_i^3\sum_{j,k,l=1}^d\frac{e^{X_i^T\theta Y_i}\left(e^{X_i^T\theta Y_i}-1\right)}{\left(1+e^{X_i^T\theta Y_i}\right)^3}X_i^{(j)}X_i^{(k)}X_i^{(l)}u_1^{(j)}u_1^{(k)}u_1^{(l)}$$

and therefore, for all $\theta\in\mathbb{R}^d$,

$$\frac{1}{n}\|L_n'''(\theta)\| \leq\frac{1}{n}\sum_{k=1}^n\|X_k\|^3\frac{e^{X_k^T\theta Y_k}\left|e^{X_k^T\theta Y_k}-1\right|}{\left(1+e^{X_k^T\theta Y_k}\right)^3} \leq \frac{1}{6\sqrt{3}\,n}\sum_{k=1}^n\|X_k\|^3 =: M_2. \qquad \text{(H.3)}$$

Now, a straightforward calculation reveals that, for $\|u\|\leq 1, \|v\|\leq 1, \|w\|\leq 1$ and any $\theta\in\mathbb{R}^d$, and $\bar{\delta}\leq 1$,

$$\sup_{\|\theta-\bar{\theta}_n\|\leq\bar{\delta}}\left|\sum_{i,j,k=1}^d\left(\frac{\partial^3}{\partial\theta_j\partial\theta_i\partial\theta_k}\log\pi(\theta)\right)u_iv_jw_k\right|$$

$$
\begin{aligned}
\leq & \sup_{\|\theta-\bar{\theta}_n\|\leq\bar{\delta}} \left\{ \frac{3(\nu+d)}{[\nu+(\theta-\mu)^T\Sigma^{-1}(\theta-\mu)]^2} \left\|\Sigma^{-1}+\operatorname{diag}(\Sigma^{-1})\right\|_{op}^2 \|\theta-\mu\| \right. \\
& \left. + \frac{2(\nu+d)}{[\nu+(\theta-\mu)^T\Sigma^{-1}(\theta-\mu)]^3} \left\|\Sigma^{-1}+\operatorname{diag}(\Sigma^{-1})\right\|_{op}^3 \|\theta-\mu\|^3 \right\} \\
\leq & \frac{3(\nu+d)}{\nu^2} \left\|\Sigma^{-1}+\operatorname{diag}(\Sigma^{-1})\right\|_{op}^2 \left(1+\|\bar{\theta}_n-\mu\|\right) \\
& + \frac{2(\nu+d)}{\nu^2} \left\|\Sigma^{-1}+\operatorname{diag}(\Sigma^{-1})\right\|_{op}^3 \left(1+\|\bar{\theta}_n-\mu\|\right)^3.
\end{aligned}
$$

Combining this with Eq. (H.3), we obtain

$$
\begin{aligned}
\frac{1}{n}\left\|\overline{L}_n'''(\bar{\theta}_n)\right\| \leq & \frac{1}{6\sqrt{3}\,n}\sum_{k=1}^{n}\|X_k\|^3 + \frac{3(\nu+d)}{\nu^2\,n}\left\|\Sigma^{-1}+\operatorname{diag}(\Sigma^{-1})\right\|_{op}^2\left(1+\|\bar{\theta}_n-\mu\|\right) \\
& + \frac{2(\nu+d)}{\nu^3\,n}\left\|\Sigma^{-1}+\operatorname{diag}(\Sigma^{-1})\right\|_{op}^3\left(1+\|\bar{\theta}_n-\mu\|\right)^3 =: \overline{M}_2. \qquad \text{(H.4)}
\end{aligned}
$$

H.4.4 Calculating κ and $\bar{\kappa}$

Note that L_n is strictly concave. Therefore,

$$
\begin{aligned}
\sup_{\theta:\|\theta-\hat{\theta}_n\|>\delta} \frac{L_n(\theta)-L_n(\hat{\theta})}{n} \leq & \sup_{\theta:\|\theta-\hat{\theta}_n\|=\delta} \frac{L_n(\theta)-L_n(\hat{\theta})}{n} \\
\leq & \sup_{\theta:\|\theta-\hat{\theta}_n\|=\delta} \left\{-\frac{1}{2}\left(\theta-\hat{\theta}_n\right)^T \hat{J}_n(\hat{\theta}_n)\left(\theta-\hat{\theta}_n\right)\right\} + \frac{M_2\delta^3}{2} \\
\leq & -\frac{1}{2}\lambda_{\min}(\hat{\theta}_n)\delta^2 + \frac{M_2\delta^3}{2} =: -\kappa.
\end{aligned}
$$

Since M_2 in Eq. (H.3) provides a uniform bound on $\frac{1}{n}\|L_n'''(\theta)\|$ over $\theta\in\mathbb{R}^d$, a similar calculation shows:

$$
\begin{aligned}
& \sup_{\theta:\|\theta-\hat{\theta}_n\|>\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|} \frac{L_n(\theta)-L_n(\hat{\theta}_n)}{n} \\
\leq & -\frac{1}{2}\lambda_{\min}(\hat{\theta}_n)\left(\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|\right)^2 + \frac{M_2\left(\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|\right)^3}{2} =: -\bar{\kappa}.
\end{aligned}
$$

H.4.5 Finding appropriate values of δ and $\bar{\delta}$

In order to apply our results in the MLE-centric approach, we need

$$
\sqrt{\frac{\operatorname{Tr}\left[\hat{J}_n(\hat{\theta}_n)^{-1}\right]}{n}} < \delta < \frac{\lambda_{\min}(\hat{\theta}_n)}{M_2}. \qquad \text{(H.5)}
$$

This assumption also ensures that κ from Appendix H.4.4 is positive. In the MAP-centric approach we also require

$$\max\left\{\|\hat{\theta}_n - \bar{\theta}_n\|, \sqrt{\frac{\operatorname{Tr}\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}{n}}\right\} < \bar{\delta} < \frac{\overline{\lambda}_{\min}(\bar{\theta}_n)}{\overline{M}_2},$$

which also ensures that $\bar{\kappa}$ from Appendix H.4.4 is positive. Such choices of $\bar{\delta}$ and δ will be available for sufficiently large n. In order to choose the appropriate concrete values of $\bar{\delta}$ and δ one can run a numerical optimization scheme.

References

Richard Abbott et al. Population of merging compact binaries inferred using gravitational waves through GWTC-3. *Physical Review X*, 13(1):011048, 2023.

Francis Bach. Approximating integrals with Laplace's method. `https://francisbach.com/laplace-method/`, 2021. Accessed: 2022-08-29.

Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and Geometry of Markov Diffusion Operators*. Springer Cham, 2016.

Dominique Bakry and Michel Émery. Diffusions hypercontractives. *Séminaire de probabilités de Strasbourg*, 19:177–206, 1985. URL `http://eudml.org/doc/113511`.

Krishna Balasubramanian, Sinho Chewi, Murat A Erdogdu, Adil Salim, and Shunshi Zhang. Towards a Theory of Non-Log-Concave Sampling:First-Order Stationarity Guarantees for Langevin Monte Carlo. In Po-Ling Loh and Maxim Raginsky, editors, *Proceedings of Thirty Fifth Conference on Learning Theory*, volume 178 of *Proceedings of Machine Learning Research*, pages 2896–2923. PMLR, 02–05 Jul 2022. URL `https://proceedings.mlr.press/v178/balasubramanian22a.html`.

Rina Foygel Barber, Mathias Drton, and Kean Ming Tan. Laplace Approximation in High-Dimensional Bayesian Regression. In Arnoldo Frigessi, Peter Bühlmann, Ingrid K. Glad, Mette Langaas, Sylvia Richardson, and Marina Vannucci, editors, *Statistical Analysis for High-Dimensional Data*, pages 15–36, Cham, 2016. Springer International Publishing. ISBN 978-3-319-27099-9.

Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer New York, NY, 2006.

P. G. Bissiri, C. C. Holmes, and S. G. Walker. A general framework for updating belief distributions. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 78 (5):1103–1130, 2016. ISSN 13697412, 14679868. URL `http://www.jstor.org/stable/44682909`.

David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. Variational Inference: A Review for Statisticians. *Journal of the American Statistical Association*, 112(518):859–877, 2017. doi: 10.1080/01621459.2017.1285773.

Victor Chernozhukov and Han Hong. An MCMC approach to classical estimation. *Journal of Econometrics*, 115(2):293–346, 2003. ISSN 0304-4076. doi: https://doi.org/10.1016/S0304-4076(03)00100-3. URL https://www.sciencedirect.com/science/article/pii/S0304407603001003.

Sinho Chewi, Murat A Erdogdu, Mufan Li, Ruoqi Shen, and Shunshi Zhang. Analysis of Langevin Monte Carlo from Poincare to Log-Sobolev. In Po-Ling Loh and Maxim Raginsky, editors, *Proceedings of Thirty Fifth Conference on Learning Theory*, volume 178 of *Proceedings of Machine Learning Research*, pages 1–2. PMLR, 02–05 Jul 2022. URL https://proceedings.mlr.press/v178/chewi22a.html.

Kacper Chwialkowski, Heiko Strathmann, and Arthur Gretton. A Kernel Test of Goodness of Fit. In Maria Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 2606–2615, New York, New York, USA, 20–22 Jun 2016. PMLR. URL https://proceedings.mlr.press/v48/chwialkowski16.html.

Erik Daxberger, Agustinus Kristiadi, Alexander Immer, Runa Eschenhagen, Matthias Bauer, and Philipp Hennig. Laplace Redux - Effortless Bayesian Deep Learning. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 20089–20103. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper/2021/file/a7c9585703d275249f30a088cebba0ad-Paper.pdf.

Guillaume Dehaene. A deterministic and computable Bernstein-von Mises theorem. arXiv:1904.02505v2, 2019.

Murat A. Erdogdu, Rasa Hosseinzadeh, and Shunshi Zhang. Convergence of Langevin Monte Carlo in Chi-Squared and Rényi Divergence. In Gustau Camps-Valls, Francisco J. R. Ruiz, and Isabel Valera, editors, *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 8151–8175. PMLR, 28–30 Mar 2022. URL https://proceedings.mlr.press/v151/erdogdu22a.html.

Marie Ernst, Gesine Reinert, and Yvik Swan. First-order covariance inequalities via Stein's method. *Bernoulli*, 26(3):2051 – 2081, 2020. doi: 10.3150/19-BEJ1182. URL https://doi.org/10.3150/19-BEJ1182.

Adrian Fischer, Robert E. Gaunt, Gesine Reinert, and Yvik Swan. Normal approximation for the posterior in exponential families. arXiv:2209.08806v1, 2022.

Seth Flaxman, Swapnil Mishra, Axel Gandy, H Juliette T Unwin, Thomas A Mellan, Helen Coupland, Charles Whittaker, Harrison Zhu, Tresnia Berah, Jeffrey W Eaton, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, 584(7820):257–261, 2020.

Wendy L Freedman. Measurements of the Hubble constant: tensions in perspective. *The Astrophysical Journal*, 919(1):16, 2021.

J. K. Ghosh and R. V. Ramamoorthi. *Bayesian Nonparametrics*. Springer Science+Business Media, 2003.

Geoff Goehle. The Bayesian central limit theorem for exponential family distributions: a geometric approach. *Information Geometry*, 2024. doi: 10.1007/s41884-024-00152-1.

Virgilio Gomez-Rubio, Roger S Bivand, and Håvard Rue. Spatial models using Laplace approximation methods. In *Handbook of regional science*, pages 1943–1959. Springer, 2021.

Jackson Gorham and Lester Mackey. Measuring Sample Quality with Stein's Method. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'15, page 226–234, Cambridge, MA, USA, 2015. MIT Press.

Jackson Gorham and Lester Mackey. Measuring Sample Quality with Kernels. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1292–1301. PMLR, 06–11 Aug 2017. URL `https://proceedings.mlr.press/v70/gorham17a.html`.

Jackson Gorham, Andrew B. Duncan, Sebastian J. Vollmer, and Lester Mackey. Measuring sample quality with diffusions. *The Annals of Applied Probability*, 29(5):2884 – 2928, 2019. doi: 10.1214/19-AAP1467. URL `https://doi.org/10.1214/19-AAP1467`.

Nathael Gozlan. A characterization of dimension free concentration in terms of transportation inequalities. *The Annals of Probability*, 37(6):2480 – 2498, 2009. doi: 10.1214/09-AOP470. URL `https://doi.org/10.1214/09-AOP470`.

Mareike Hasenpflug, Daniel Rudolf, and Björn Sprungk. Wasserstein convergence rates of increasingly concentrating probability measures. arXiv:2207.08551v1, 2022.

Tapio Helin and Remo Kretschmann. Non-asymptotic error estimates for the Laplace approximation in Bayesian inverse problems. *Numerische Mathematik*, 150:521—-549, 2022.

R. Holley and D. Stroock. Logarithmic Sobolev inequalities and stochastic Ising models. *Journal of Statistical Physics*, 46:1159–1194, 1987.

Daniel Hsu, Sham Kakade, and Tong Zhang. A tail inequality for quadratic forms of subgaussian random vectors. *Electronic Communications in Probability*, 17(none):1 – 6, 2012. doi: 10.1214/ECP.v17-2079. URL `https://doi.org/10.1214/ECP.v17-2079`.

Jonathan H. Huggins, Mikołaj Kasprzak, Trevor Campbell, and Tamara Broderick. Practical Bounds on the Error of Bayesian Posterior Approximations: a Nonasymptotic Approach. arXiv:1809.09505v2, 2018.

I. A. Ibragimov and R. Z. Has'minskii. *Statistical Estimation: Asymptotic Theory*. Stochastic Modelling and Applied Probability. Springer New York, NY, 1981.

Terry C Jones, Guido Biele, Barbara Mühlemann, Talitha Veith, Julia Schneider, Jörn Beheim-Schwarzbach, Tobias Bleicker, Julia Tesch, Marie Luisa Schmidt, Leif Erik Sander, et al. Estimating infectiousness throughout SARS-CoV-2 infection course. *Science*, 373(6551):eabi5273, 2021.

Anya Katsevich. Tight Dimension Dependence of the Laplace Approximation. arXiv:2305.17604v2, 2023a.

Anya Katsevich. Tight skew adjustment to the Laplace approximation in high dimensions. arXiv:2306.07262v1, 2023b.

Anya Katsevich. Improved dimension dependence in the Bernstein von Mises Theorem via a new Laplace approximation bound. arXiv:2308.06899v2, 2024.

Anya Katsevich and Philippe Rigollet. On the Approximation Accuracy of Gaussian Variational Inference. *Annals of Statistics*, 52(4):1384–1409, 2024. doi: 10.1214/24-AOS2393.

B.J.K. Kleijn and A.W. van der Vaart. The Bernstein-Von-Mises theorem under misspecification. *Electronic Journal of Statistics*, 6(none):354 – 381, 2012. doi: 10.1214/12-EJS675. URL `https://doi.org/10.1214/12-EJS675`.

A.V. Kolesnikov and E. Milman. Riemannian metrics on convex sets with applications to Poincaré and log-Sobolev inequalities. *Calculus of Variations and Partial Differential Equations*, 55:1–36, 2016.

Pierre Simon Laplace. Mémoire sur la probabilité des causes par les événements. In *Mémoires de l'Académie royale des sciences de Paris (Savants étrangers)*, pages 621–656. 1774.

Pierre Simon Laplace. Memoir on the Probability of the Causes of Events. *Statistical Science*, 1(3):364–378, 1986. ISSN 08834237. URL `http://www.jstor.org/stable/2245476`.

Qiang Liu, Jason Lee, and Michael Jordan. A Kernelized Stein Discrepancy for Goodness-of-fit Tests. In Maria Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 276–284, New York, New York, USA, 20–22 Jun 2016. PMLR. URL `https://proceedings.mlr.press/v48/liub16.html`.

Quan Long, Marco Scavino, Raúl Tempone, and Suojin Wang. Fast estimation of expected information gains for Bayesian experimental designs based on Laplace approximations. *Computer Methods in Applied Mechanics and Engineering*, 259:24–39, 2013.

Pascal Massart. Concentration inequalities and model selection. In Jean Picard, editor, *Ecole d'Eté de Probabilités de Saint-Flour XXXIII - 2003*, Lecture Notes in Mathematics. Springer Berlin, Heidelberg, 2007.

Jeffrey W. Miller. Asymptotic Normality, Concentration, and Coverage of Generalized Posteriors. *Journal of Machine Learning Research*, 22(168):1–53, 2021. URL `http://jmlr.org/papers/v22/20-469.html`.

Jeffrey W. Miller and Matthew T. Harrison. Inconsistency of Pitman-Yor Process Mixtures for the Number of Components. *Journal of Machine Learning Research*, 15(96):3333–3370, 2014. URL `http://jmlr.org/papers/v15/miller14a.html`.

Kevin P. Murphy. *Probabilistic Machine Learning: An introduction*. MIT Press, 2022. URL probml.ai.

Chris J. Oates, Mark Girolami, and Nicolas Chopin. Control functionals for Monte Carlo integration. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3):695–718, 2017. doi: https://doi.org/10.1111/rssb.12185. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssb.12185.

Maxim Panov and Vladimir Spokoiny. Finite Sample Bernstein – von Mises Theorem for Semiparametric Problems. *Bayesian Analysis*, 10(3):665 – 710, 2015. doi: 10.1214/14-BA926. URL https://doi.org/10.1214/14-BA926.

Jaakko Riihimäki and Aki Vehtari. Laplace approximation for logistic Gaussian process density estimation and regression. 2014.

Hippolyt Ritter, Aleksandar Botev, and David Barber. Online structured Laplace approximations for overcoming catastrophic forgetting. *Advances in Neural Information Processing Systems*, 31, 2018.

Ralph Tyrell Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970. ISBN 9781400873173. doi: doi:10.1515/9781400873173. URL https://doi.org/10.1515/9781400873173.

Adrien Saumard and Jon A. Wellner. Log-concavity and strong log-concavity: A review. *Statistics Surveys*, 8(none):45 – 114, 2014. doi: 10.1214/14-SS107. URL https://doi.org/10.1214/14-SS107.

Claudia Schillings, Björn Sprungk, and Philipp Wacker. On the convergence of the Laplace approximation and noise-level-robustness of Laplace-based Monte Carlo methods for Bayesian inverse problems. *Numerische Mathematik*, 145:917–971, 2020.

André Schlichting. Poincaré and Log-Sobolev Inequalities for Mixtures. *Entropy*, 21(89), 2019.

Vladimir Spokoiny. Dimension free non-asymptotic bounds on the accuracy of high dimensional Laplace approximation. arXiv:2204.11038v3, 2022.

P. Sur, Y. Chen, and E.J. Candès. The likelihood ratio test in high-dimensional logistic regression is asymptotically a rescaled Chi-square. *Probab. Theory Relat. Fields*, 175: 487—558, 2019.

Uffe Høgsbro Thygesen, Christoffer Moesgaard Albertsen, Casper Willestofte Berg, Kasper Kristensen, and Anders Nielsen. Validation of ecological state space models using the Laplace approximation. *Environmental and Ecological Statistics*, 24:317–339, 2017.

A.W. van der Vaart. *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.

Santosh Vempala and Andre Wibisono. Rapid Convergence of the Unadjusted Langevin Algorithm: Isoperimetry Suffices. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL `https://proceedings.neurips.cc/paper_files/paper/2019/file/65a99bb7a3115fdede20da98b08a370f-Paper.pdf`.

Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018. doi: 10.1017/9781108231596.

Cédric Villani. *Optimal Transport Old and New*. Grundlehren der mathematischen Wissenschaften. Springer Berlin, Heidelberg, 2009.

Martin J. Wainwright and Michael I. Jordan. Graphical Models, Exponential Families, and Variational Inference. *Foundations and Trends® in Machine Learning*, 1(1–2):1–305, 2008. ISSN 1935-8237. doi: 10.1561/2200000001. URL `http://dx.doi.org/10.1561/2200000001`.

Jian Wang, William M Wells, Polina Golland, and Miaomiao Zhang. Efficient Laplace Approximation for Bayesian Registration Uncertainty Quantification. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I*, pages 880–888. Springer, 2018.

Keisuke Yano and Kengo Kato. On frequentist coverage errors of Bayesian credible sets in moderately high dimensions. *Bernoulli*, 26(1):616 – 641, 2020. doi: 10.3150/19-BEJ1142. URL `https://doi.org/10.3150/19-BEJ1142`.