On the Bernstein-von Mises phenomenon in the Gaussian white noise model

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

On the Bernstein-von Mises phenomenon in the Gaussian white noise model[*]

Haralambie Leahu

Department of Mathematics and Computer Science
Technical University Eindhoven (TU/e)
Den Dolech 2, 5600 MB, Eindhoven, The Netherlands
e-mail: h.leahu@tue.nl

Abstract: We study the Bernstein-von Mises (BvM) phenomenon, i.e., Bayesian credible sets and frequentist confidence regions for the estimation error coincide asymptotically, for the infinite-dimensional Gaussian white noise model governed by Gaussian prior with diagonal-covariance structure. While in parametric statistics this fact is a consequence of (a particular form of) the BvM Theorem, in the nonparametric setup, however, the BvM Theorem is known to fail even in some, apparently, elementary cases. In the present paper we show that BvM-like statements hold for this model, provided that the parameter space is suitably embedded into the support of the prior. The overall conclusion is that, unlike in the parametric setup, positive results regarding frequentist probability coverage of credible sets can only be obtained if the prior assigns null mass to the parameter space.

AMS 2000 subject classifications: Primary 62G08, 62G20; secondary 60B12, 60F05, 62J05, 28C20.
Keywords and phrases: Nonparametric Bernstein-von Mises Theorem.

Contents

Introduction

In parametric statistics, the celebrated Bernstein-von Mises (BvM) Theorem states that in a statistical model with finite-dimensional parameter $\theta \in \Theta$, if the observed variable X follows some known distribution $\mathbb{P}_\theta := \mathcal{L}(X|\theta)$ and π is a prior probability on the parameter space Θ then, under fairly general conditions over the true parameter, the model and the prior, the centered Bayesian posterior and the sampling distribution of any asymptotically efficient estimator centered at truth will be close for a large number of observations, where "close" means with respect to total variation norm; see, e.g., [13]. In fact, both distributions will approach the normal distribution with null mean and covariation matrix given by the inverse Fisher information. In particular, since the posterior mean is known to achieve, under some regularity conditions, asymptotic efficiency, cf. [9], one can conclude that the centered posterior distribution and the sampling distribution of the posterior mean centered at truth are asymptotically the same. If $\theta \in \Theta$ denotes the true parameter, ϑ denotes a generic random variable on Θ distributed according to the prior π and $\hat{\vartheta} := \mathbb{E}[\vartheta|X]$ denotes the posterior mean then the BvM statement can be re-phrased as

$$\|\mathcal{L}(\Delta|X) - \mathcal{L}(\Delta|\theta)\|_V \xrightarrow{\mathbb{P}_\theta} 0. \tag{1}$$

In the above display, $\Delta := \hat{\vartheta} - \vartheta$, $\mathcal{L}(\Delta|\theta)$ denotes the sampling distribution of the estimation error Δ (under $\mathbb{P}_\theta$) and $\|\cdot\|_V$ denotes the total variation norm.

The main importance of the BvM Theorem is that it allows one to use Bayesian credible sets, i.e., sets which receive a fixed fraction of the total mass under $\mathcal{L}(\Delta|X)$, which are available via Markov-chain Monte Carlo techniques, to derive confidence regions for θ based on asymptotically efficient frequentist estimators, e.g., MLE. A natural question is whether such a result can be also established in a nonparametric framework, i.e., for infinite-dimensional parameter space Θ. The first difficulty in answering this question is to formulate a generalization of the parametric statement. For instance, one of the conditions in the classical statement of the theorem is that the prior should be (Lebesgue) absolutely continuous, but it is known that no infinite-dimensional counterpart of the Lebesgue measure exists. Moreover, one of the key assumptions of the BvM Theorem is that the model is smooth in the sense of Hellinger differentiability but differentiability w.r.t. infinite-dimensional parameters is a rather restrictive condition, hence a weaker concept of differentiability might be suitable. Finally, asymptotically efficient estimators are rarely available in infinite-dimensional models; also a tractable concept of Fisher information (operator) is lacking. Nevertheless, it makes sense to consider the simpler version, in which the posterior mean plays the role of the estimator for θ; see (1).

In this paper we study the BvM phenomenon for the infinite-dimensional Gaussian white noise model, described by the linear equation

$$X = \theta + \sigma_n \varepsilon, \tag{2}$$

where θ is a square-summable sequence, X is a noisy observation of θ, ε denotes a sequence of standard, i.i.d. Gaussian variables and $\sigma_n \downarrow 0$; typically, we have $\sigma_n^2 = 1/n$. For a Bayesian approach, one chooses a Gaussian prior π on $\mathbb{R}^\infty$ with diagonal covariance structure, i.e., π makes the coordinates independent non-degenerate Gaussian variables. The posterior distribution $\mathfrak{L}_n(\vartheta|X)$ is Gaussian and its centered version $\mathfrak{L}_n(\vartheta - \hat{\vartheta}|X) = \mathfrak{L}_n(-\Delta|X) = \mathfrak{L}_n(\Delta|X)$ is non-random. Moreover, $\mathfrak{L}_n(\Delta|\theta)$ is also a Gaussian measure, not depending on the observation X but depending on θ only through its mean, cf. [2, 3]. Hence, in this case, one may disregard convergence in $\mathbb{P}_\theta$-probability in (1), which reduces to

$$\|\mathfrak{L}_n(\Delta|X) - \mathfrak{L}_n(\Delta|\theta)\| \to 0. \tag{3}$$

The validity of (3), which obviously depends on θ, has been extensively studied in [3] in the π-a.s. sense; specifically, the prior is chosen such that ϑ is a square-summable sequence a.s. and it is shown that (3) fails in this case, in the sense that for almost all θ's drawn from π the expression in (3) does not converge to 0. The main argument is that the r.v. $\sum_{k \geq 1} \Delta_k^2$, which is properly defined, has different asymptotic behavior when regarded from Bayesian and frequentist perspectives; this is explained in Section 3. A similar result has been obtained in [2] in a slightly more general setup, where the parameter θ belongs to some abstract Hilbert space and the observations lie in a Banach space. Although some positive results regarding the validity of BvM-like statements in semi- and nonparametric models, see, e.g., [1, 11, 12] for semiparamatric and [5, 6] for nonparametric, the results in [2] and [3] led to the widely accepted belief that the BvM phenomenon does not occur in the nonparametric framework.

Nonetheless, some questions, of both theoretical and practical interest, are left unanswered in [2] and [3] and we aim to elucidate these issues in this paper. A first question is regarding the choice of the prior. Denoting by ℓ^2 the space of square-summable sequences in $\mathbb{R}^\infty$, we have $\theta \in \ell^2$. Moreover, for the prior considered in [3], it holds $\vartheta \in \ell^2$ a.s. Since ℓ^2 coincides with the reproducing Hilbert space (RHS) of the noise ε, it follows that ε, hence the observation X, lies almost surely in some larger (Banach) space $\mathbb{H}$ in which ℓ^2 appears as a dense subspace. Furthermore, since in the Bayesian paradigm we have

$$X = \vartheta + \sigma_n \varepsilon,$$

by the Cameron-Martin Theorem, the distribution of the data X is equivalent to that of the Gaussian noise $\sigma_n \varepsilon$, hence orthogonal w.r.t. the prior π. In other words, the randomness induced by the prior distribution in the model is rather insignificant w.r.t. the distribution of the data, unlike in finite-dimensional framework where the prior is required to be Lebesgue continuous; a quasi-similar choice is made in [2]. This abnormality occurs only in the infinite-dimensional framework; in finite-dimensional spaces the RHS of some Gaussian measure coincides with its support. Therefore, it may come as no surprise that the BvM statement, as formulated in (3), for the models considered in [2] and [3] fails.

On the other hand, in Bayesian statistics, the statistician must choose the prior distribution π, based on some (apriori) subjective beliefs, so that one may always question these beliefs. Therefore, a statement which is true π-a.s.

is not always satisfactory since sets of parameters of null prior probability are actually ignored. For example, it is known that the classical Wiener measure does not charge the space of differentiable paths, hence statements which are true "Wiener almost surely" are, in fact, disregarding smooth paths; this issue appears for any infinite-dimensional Gaussian distribution. In order to cope with this problem, we shall consider analytic BvM statements; specifically, if Θ is some given parameter set, we investigate pointwise convergence in (3), w.r.t. $\theta \in \Theta$, rather than π-a.s. convergence. Probabilistic BvM statements can be easily derived, provided that $\pi(\Theta) = 1$. The results in [3] show that, if the prior π is supported by ℓ^2, no $\Theta \subset \ell^2$ with $\pi(\Theta) > 0$ exists such that the BvM statement in (3) holds for all $\theta \in \Theta$ but no relevant conclusion can be drawn if $\pi(\Theta) = 0$. The interest in parameter sets Θ of null prior probability is raised by the fact that, in this model, the Bayes estimator $\hat{\vartheta}$ achieves optimal minimax rate of convergence when the parameter θ belongs to some linear subspace $\Theta_0 \subset \ell^2$ (to be defined in Section 3) of null prior mass; see [14]. Therefore, it would be of some interest to know whether (3) holds true for $\theta \in \Theta_0$.

The present paper is aimed to perform a thorough investigation and to provide answers to the questions stated above. There will be three main conclusions:

- If the prior π makes the coordinates ϑ_k centered Gaussian variables with variance τ_k^2, for $k \geq 1$, then the BvM statement in (3) holds true, provided that $\tau_k \to \infty$ sufficiently fast. In fact, it turns out that both $\mathfrak{L}_n(\Delta|X)$ and $\mathfrak{L}_n(\Delta|\theta)$, when re-scaled by $1/\sigma_n$, approach the Gaussian white noise (centered) distribution whose covariance operator can be formally regarded as the inverse Fisher information of the linear model defined by (2).
- Unfortunately, when the prior π is supported by ℓ^2, having diagonal power-covariance structure as in [3], there is no sensible subspace Θ (not even of null prior mass) such that (3) holds for all parameters $\theta \in \Theta$.
- The good news, however, is that if $\theta \in \Theta_0$, i.e., the Bayes estimator $\hat{\vartheta}$ achieves optimal minimax rates, then the Bayesian credible ℓ^2-balls have good frequentist probability coverage, for large n, so they may be used to derive confidence regions for θ based on the posterior mean $\hat{\vartheta}$.

The paper is organized as follows: Section 1 gives a brief overview of results on Gaussian measures in Hilbert spaces which will be relevant for our analysis. In Section 2 we formulate and prove some BvM-like statements and explain why the ℓ^2-space is too small for dealing with BvM-related issues. Also, we provide conditions over π and θ such that (3) holds true. In Section 3 we zoom in to the framework of [3], where the prior distribution is supported by ℓ^2, and show that there is no reasonable parameter set $\Theta \subset \ell^2$ such that (3) holds true for all $\theta \in \Theta$; to this end, we consider a Hilbert scale (increasing family of Hilbert spaces) $\{\Theta_\delta\}_\delta$ in ℓ^2 and prove that there is no Θ_δ satisfying the requirement, proving analytic BvM statements rather than probabilistic statements as in [3]. We also investigate in Section 3 the asymptotic frequentist probability coverage of Bayesian credible ℓ^2-balls for various classes of parameters θ. Some technical facts and results are detailed in Section 4 (Appendix) while the proofs of the main results are deferred to Section 5.

Notations and conventions

Throughout this paper, $\mathbb{R}^\infty$ will denote the linear space of all real-valued sequences. For convenience, we identify the sequence with components $\{x_k\}_{k \geq 1}$ with an element $(x_k) \in \mathbb{R}^\infty$, when no confusion occurs. For sequences with double (or multi-) index $\{x_{nk}\}_{n,k \geq 1}$, we use the notation $(x_{nk})_k$ to emphasize that we are referring to the sequence labeled w.r.t. k. We denote by $(\ell^2, \|\cdot\|)$ the space of square-summable sequences endowed with the usual Hilbert space structure. For $k \geq 1$ we denote by $\mathbf{e}_k$ the kth unit vector (direction) in $\mathbb{R}^\infty$, having the kth entry equal to 1, the other elements being null. The family $\{\mathbf{e}_k : k \geq 1\}$ defines an orthonormal basis (complete orthonormal system) in ℓ^2. Also, we shall denote by ℓ^1 the space of (absolutely) summable sequences and by ℓ^∞ the space of bounded sequences endowed with the usual norms; recall that $\ell^1 \subset \ell^2 \subset \ell^\infty$.

If $\{u_n\}_{n \geq 1}$ and $\{v_n\}_{n \geq 1}$ are sequences of positive numbers we write $u_n \approx v_n$ if $\lim_n (u_n/v_n) = 1$ and we write $u_n \sim v_n$ if there exist some positive constant c such that $\lim_n (u_n/v_n) = c$. If $\lim_n (u_n/v_n) = 0$ then we write $u_n \ll v_n$.

If X is a r.v. on some measure space $\mathbb{X}$ we denote by $\mathfrak{L}(X)$ its distribution on $\mathbb{X}$ and denote by $\mathfrak{L}(X|\cdot)$ a conditional distribution of X. Also, we shall denote by $\mathbb{E}[X|\cdot]$ and $\mathrm{Var}[X|\cdot]$ the (conditional) expectation, resp. variance, of X. If $\mathbb{X}$ is a Banach/Hilbert space we shall denote by $\mathcal{N}(b; S)$ the Gaussian measure on $\mathbb{X}$ with mean $b \in \mathbb{X}$ and covariation operator $S : \mathbb{X}^* \to \mathbb{X}$; in particular, $\mathcal{N}(b; \sigma^2)$ denotes the one-dimensional Gaussian measure with mean b and variance σ^2.

Let $(\mathbb{X}, d)$ be a metric space. On the space of signed measures on $\mathbb{X}$ we define by $\|\cdot\|_V$ and $\|\cdot\|_H$ the total variation and Hellinger[1] norms, respectively. The metrics induced by these norms on the class of probability measures are known to generate the same topology. If $\{P_n\}_n$ and $\{Q_n\}_n$ are two sequences of probability measures, by $P_n \simeq Q_n$ we mean that $P_n - Q_n$ converges to 0 in this (common) topology. Both $\|P - Q\|_V$ and $\|P - Q\|_H$ attain their maximum whenever P and Q are orthogonal measures. In that sense, each distance can be used to measure the degree of overlapping between P and Q. The expression

$$\mathcal{A}(P, Q) := 1 - \|P - Q\|_H^2 = \int_{\mathbb{X}} \sqrt{dP}\sqrt{dQ}$$

is called the Hellinger affinity of P and Q. If P and Q are equivalent measures then $\mathcal{A}(P, Q) > 0$; null affinity means orthogonality between P and Q. Finally, we shall denote by $\twoheadrightarrow$ the weak convergence on the class of (probability) measures on $\mathbb{X}$. Weak convergence is weaker than convergence in total variation/Hellinger distance, one of the main differences being that a sequence P_n may converge weakly to P even if P_n and P are orthogonal measures, for arbitrarily large n; this is not possible if convergence holds in total variation/Hellinger distance. As a consequence, if $P_n \twoheadrightarrow P$ then the property $P_n(A) \to P(A)$ is restricted to the class of those Borel measurable sets $A \subset \mathbb{X}$ having P-negligible boundary, whereas convergence in total variation/Hellinger distance implies $P_n(A) \to P(A)$, for any Borel measurable $A \subset \mathbb{X}$.

---

[1] The Hellinger norm is defined such that $\sqrt{2}\|P\|_2 = 1$ for any probability measure P.

1. Preliminary results on Gaussian measures in Hilbert spaces

There is a rich literature treating Gaussian measures on separable Hilbert or, more generally, Banach spaces. For a nice, comprehensive overview of Gaussian measures on Banach spaces and related concepts we refer to [7, 8]. Here we briefly present some facts which will be relevant for our analysis. To avoid a rather technical exposition we restrict ourselves to the Hilbert space setting.

In the following $\mathbb{H}$ is a separable Hilbert space, $b, d \in \mathbb{H}$ and $S, T : \mathbb{H} \to \mathbb{H}$ are covariance operators[2] on $\mathbb{H}$. The following theorems, which establish conditions under which Gaussian measures are equivalent, will be useful in our analysis.

Cameron-Martin Theorem: $\mathcal{N}(b; S)$ and $\mathcal{N}(\mathbf{0}; S)$ are equivalent if and only if $b \in \mathcal{H}$, where $\mathcal{H} := \sqrt{S}\mathbb{H}$ denotes the RHS of $\mathcal{N}(\mathbf{0}; S)$ endowed with the usual Hilbert space structure. In addition, the Radon-Nikodym derivative is given by

$$\frac{d\mathcal{N}(b; S)}{d\mathcal{N}(\mathbf{0}; S)}(h) = \exp\left[\langle b|h\rangle_{\widetilde{\mathcal{H}}} - \frac{1}{2}\|b\|_{\mathcal{H}}^2\right], \ \mathcal{N}(\mathbf{0}; S) - a.s.,$$

where $b \in \mathcal{H} \mapsto \langle b|\cdot\rangle_{\widetilde{\mathcal{H}}} \in \mathcal{L}^2(\mathcal{N}(0; S))$ denotes the extension of the linear isometry $b \in \mathcal{H} \mapsto \langle b|\cdot\rangle_{\mathcal{H}}$; that is, $\langle b|h\rangle_{\widetilde{\mathcal{H}}} = \mathcal{L}^2 - \lim_n \langle b|h_n\rangle_{\mathcal{H}}$, for $h_n \to h$, $\{h_n\}_n \subset \mathcal{H}$.

Feldman-Hajek Theorem: Assume that S and T commute; in this case there exists some orthonormal basis $\{\phi_n\}_{n \geq 1}$ consisting of common eigenvectors of S and T. If S has eigenvalues $\{\lambda_n^2\}_{n \geq 1}$ and T has eigenvalues $\{\mu_n^2\}_{n \geq 1}$ w.r.t. $\{\phi_n\}_{n \geq 1}$ then $\mathcal{N}(\mathbf{0}; S)$ and $\mathcal{N}(\mathbf{0}; T)$ are equivalent if and only if

$$\sum_{n=1}^{\infty} \left(\frac{\lambda_n^2 - \mu_n^2}{\lambda_n^2 + \mu_n^2}\right)^2 < \infty.$$

Otherwise, $\mathcal{N}(\mathbf{0}; S)$ and $\mathcal{N}(\mathbf{0}; T)$ are orthogonal.

Equivalence Theorem: $\mathcal{N}(b; S)$ and $\mathcal{N}(d; T)$ are equivalent if and only if

$$\mathcal{N}(b; S) \equiv \mathcal{N}(b; T) \equiv \mathcal{N}(d; T) \equiv \mathcal{N}(d; S).$$

Otherwise $\mathcal{N}(b; S)$ and $\mathcal{N}(d; T)$ are orthogonal.

Finally, the following theorem gives necessary and sufficient conditions for a sequence of Gaussian measures to converge weakly on $\mathbb{H}$; see, e.g., [10].

Convergence Theorem: Let $\{b_n\}_{n \geq 1} \subset \mathbb{H}$ and S_n, for $n \geq 1$, be a family of covariation operators on $\mathbb{H}$. Then the sequence $\mathcal{N}(b_n; S_n)$ converges weakly on $\mathbb{H}$ if and only if there exist some $b \in \mathbb{H}$ and some covariation operator S on $\mathbb{H}$ such that $b_n \to b$ in $\mathbb{H}$ and $S_n \to S$ in trace-class norm, i.e., $\text{Tr}(S_n - S) \to 0$. In this case, we have $\mathcal{N}(b_n; S_n) \twoheadrightarrow \mathcal{N}(b; S)$.

---

[2]By covariance operator we mean a positive, self-adjoint, trace-class operator $S : \mathbb{H} \to \mathbb{H}$.

2. The Gaussian white noise model

In this section we consider the linear model, defined by the equation

$$X = \theta + \sigma_n \varepsilon, \tag{4}$$

where the equality holds in $\mathbb{R}^\infty$, $\theta := (\theta_k) \in \ell^2$ is an unknown parameter, $\varepsilon := (\varepsilon_k)$ is a sequence of i.i.d. standard Gaussian variables (noise) and $\sigma_n > 0$ are chosen such that $\sigma_n \downarrow 0$ as $n \to \infty$; typically one chooses $\sigma_n^2 = 1/n$. As usual, the problem is to estimate θ from the noisy observation X.

For a Bayesian approach, one considers a prior π on $\mathbb{R}^\infty$ which makes the coordinates centered independent (nondegenerate) Gaussian variables; that is, one assumes that the parameter θ is a realization of some centered Gaussian r.v. $\vartheta = (\vartheta_k)$ such that $\mathrm{Cov}[\vartheta_k, \vartheta_l] = \mathbb{E}[\vartheta_k \vartheta_l] = \tau_k^2 \delta_{kl}$, for any $k, l \geq 1$. Hence, $\mathfrak{L}_n(X_k) = \mathcal{N}(0; \sigma_n^2 + \tau_k^2)$ and $\mathfrak{L}_n(X_k|\vartheta_k) = \mathcal{N}(\vartheta_k; \sigma_n^2)$ so that the posterior distribution $\mathfrak{L}_n(\vartheta_k|X_k)$ can be described, cf. [3], as follows:

- the posterior mean $\hat{\vartheta} = (\hat{\vartheta}_k) \in \mathbb{R}^\infty$ satisfies

$$\forall k \geq 1: \ \hat{\vartheta}_k = \mathbb{E}[\vartheta_k|X] = \frac{\tau_k^2}{\sigma_n^2 + \tau_k^2} X_k.$$

- the Bayesian estimation error $\Delta := \hat{\vartheta} - \vartheta$ satisfies

$$\forall k \geq 1: \ \Delta_k = \frac{\sigma_n \tau_k^2}{\sigma_n^2 + \tau_k^2} \varepsilon_k - \frac{\sigma_n^2}{\sigma_n^2 + \tau_k^2} \vartheta_k.$$

- the centered posterior $\mathfrak{L}_n(\Delta|X)$ is independent of $\mathfrak{L}_n(X)$ and satisfies

$$\forall k \geq 1: \ \mathrm{Var}[\vartheta_k|X] = \mathbb{E}\left[(\vartheta_k - \hat{\vartheta}_k)^2 \,\middle|\, X \right] = \frac{\sigma_n^2 \tau_k^2}{\sigma_n^2 + \tau_k^2}. \tag{5}$$

In fact, $\mathfrak{L}_n(\vartheta_k|X) = \mathfrak{L}_n(\vartheta_k|X_k)$ and the posterior $\mathfrak{L}_n(\vartheta|X)$ can be expressed as

$$\mathfrak{L}_n(\vartheta|X) = \bigotimes_{k \geq 1} \mathfrak{L}_n(\vartheta_k|X_k) = \bigotimes_{k \geq 1} \mathcal{N}\left(\frac{\tau_k^2}{\sigma_n^2 + \tau_k^2} X_k; \frac{\sigma_n^2 \tau_k^2}{\sigma_n^2 + \tau_k^2} \right).$$

Definition. Let $\Theta \subset \mathbb{R}^\infty$ and ψ be some functional. Then we say that:

- the BvM statement holds for the parameter set Θ if for all $\theta \in \Theta$ we have

$$\|\mathfrak{L}_n(\Delta|X) - \mathfrak{L}_n(\Delta|\theta)\|_H \to 0; \ (\sigma_n \downarrow 0); \tag{6}$$

- the BvM statement holds for the functional ψ and parameter set Θ if for all $\theta \in \Theta$ it holds that

$$\|\mathfrak{L}_n(\psi(\Delta)|X) - \mathfrak{L}_n(\psi(\Delta)|\theta)\|_H \to 0; \ (\sigma_n \downarrow 0); \tag{7}$$

- the BvM statement holds (for the functional ψ) π-a.s. if (6), resp. (7), holds for almost all θ's drawn from the prior π.

The support of a Gaussian measure in $\mathbb{R}^\infty$ is, in general, a Banach space $\mathbb{B}$; see [4]. Note that, if the BvM statement holds true for the parameter set Θ then the BvM statement holds true for any (norm-measurable) functional ψ defined on the support of Δ, for the same parameter set. Also, if the BvM statement holds (for the functional ψ) for some parameter set Θ such that $\pi(\Theta) = 1$ then the BvM statement holds (for the functional ψ) π-a.s.

In the following we aim to investigate the validity, in general, of the above statements for the model under discussion. We first consider linear functionals. Recall that, from a frequentist perspective, we have

$$\forall k \geq 1 : \Delta_k = \frac{\sigma_n \tau_k^2}{\sigma_n^2 + \tau_k^2}\varepsilon_k - \frac{\sigma_n^2 \theta_k}{\sigma_n^2 + \tau_k^2}. \tag{8}$$

Therefore, from (5) and (8) we deduce that (for all $k \geq 1$)

$$\mathfrak{L}_n\left(\Delta_k|X\right) = \mathcal{N}\left(0; \frac{\sigma_n^2 \tau_k^2}{\sigma_n^2 + \tau_k^2}\right), \quad \mathfrak{L}_n\left(\Delta_k|\theta\right) = \mathcal{N}\left(\frac{-\sigma_n^2 \theta_k}{\sigma_n^2 + \tau_k^2}; \frac{\sigma_n^2 \tau_k^4}{(\sigma_n^2 + \tau_k^2)^2}\right).$$

Now we aim to prove that the BvM statement holds for any suitable linear functional ψ. First, however, we need to make clear the kind of linear functionals ψ we are considering and this requires a slightly technical discussion. Assume that $\mathbb{B} \subset \mathbb{R}^\infty$ is a Banach space which supports a Gaussian measure μ. Then it makes sense to consider bounded linear functionals in the topological dual $\mathbb{B}^*$; that is, if $\mathfrak{L}(Y) = \mu$ then, for any $\psi \in \mathbb{B}^*$, $\psi(Y)$ is a r.v. finite a.s. Note, however, that the definition of the support of a measure is closely related to the topology under consideration, so that the support is not unique. To avoid this inconvenience, we need to consider a class of "universal" bounded linear functionals related to the measure itself, rather than to its topological support.

Let P be a regular probability measure on $\mathbb{R}^\infty$ and recall that any linear functional ψ on $\mathbb{R}^\infty$ is identified with a sequence $(\psi_k) \in \mathbb{R}^\infty$ such that

$$\psi(x) = \sum_{k \geq 1} \psi_k x_k, \tag{9}$$

the set of x's for which the above series is convergent being a linear (sub)space. We say that the linear functional $\psi \in \mathbb{R}^\infty$ is defined P-a.s. if the series in (9) converges for P-almost all $x \in \mathbb{R}^\infty$. In addition, we say that the P-a.s. defined linear functional ψ is bounded, or that ψ is a bounded linear functional defined P-a.s., if the series in (9) converges absolutely for P-almost all x's. In the following we will denote by γ the probability distribution on $\mathbb{R}^\infty$ which makes the coordinates standard i.i.d. Gaussian variables. By Kolmogorov's three-series Theorem it follows that a linear functional $\psi \in \mathbb{R}^\infty$ is defined γ-a.s. if and only if $\psi \in \ell^2$ and ψ is bounded if and only if $\psi \in \ell^1$. Finally, if $P = \otimes_{k \geq 1} \mathcal{N}(\nu_k; \varsigma_k^2)$ then $\psi \in \mathbb{R}^\infty$ is a bounded linear functional defined P-a.s. if and only if $(\psi_k \nu_k) \in \ell^1$ and $(\psi_k \varsigma_k) \in \ell^1$. Indeed, if $\mathfrak{L}(Y) = P$ then $Y_k = \nu_k + \varsigma_k \xi_k$, with $\mathfrak{L}(\xi) = \gamma$, i.e.,

$$\sum_{k \geq 1} \psi_k Y_k = \sum_{k \geq 1} \psi_k \nu_k + \sum_{k \geq 1} \psi_k \varsigma_k \xi_k,$$

so the two series in the r.h.s. converge absolutely if $(\psi_k \nu_k) \in \ell^1$ and $(\psi_k \varsigma_k) \in \ell^1$.

2.1. The BvM statement for linear functionals

The following result shows that the bounded linear functionals defined γ-a.s. are also bounded linear functionals defined $\mathfrak{L}_n(\Delta|X)$ and $\mathfrak{L}_n(\Delta|\theta)$-a.s., for all n, for almost all θ's drawn from the prior π. In addition, a π-a.s. BvM statement holds for such linear functionals ψ, i.e., for $\psi = (\psi_k) \in \ell^1$.

Lemma 1. *Let $\psi = (\psi_k) \in \ell^1$ be a bounded linear functional defined γ-a.s. Then it holds that*

(i) *ψ is a bounded linear functional defined $\mathfrak{L}_n(\Delta|X)$-a.s., for any $n \geq 1$.*
(ii) *ψ is a bounded linear functional defined $\mathfrak{L}_n(\Delta|\theta)$-a.s., for any $n \geq 1$, π-a.s.*
(iii) *If $\gamma \circ \psi^{-1}$ denotes the pushforward measure[3] of γ through ψ then (π-a.s.)*

$$\left\| \mathfrak{L}_n \left(\psi(\sigma_n^{-1}\Delta)|X \right) - \gamma \circ \psi^{-1} \right\|_H \to 0, \left\| \mathfrak{L}_n \left(\psi(\sigma_n^{-1}\Delta)|\theta \right) - \gamma \circ \psi^{-1} \right\|_H \to 0.$$

(iv) *The BvM statement holds true π-a.s. for ψ, i.e.,*

$$\left\| \mathfrak{L}_n \left(\psi(\Delta)|X \right) - \mathfrak{L}_n \left(\psi(\Delta)|\theta \right) \right\|_H \to 0, \ \pi - a.s.$$

Lemma 1 shows that the finite-dimensional projections of $\mathfrak{L}_n \left(\sigma_n^{-1}\Delta|X \right)$ and $\mathfrak{L}_n \left(\sigma_n^{-1}\Delta|\theta \right)$ converge to those of γ. Indeed, it is straightforward that

$$\forall k \geq 1 : \ \mathfrak{L}_n \left(\psi(\sigma_n^{-1}\Delta_k)|X \right) \simeq \mathcal{N}(0;1) \simeq \mathfrak{L}_n \left(\psi(\sigma_n^{-1}\Delta_k)|\theta \right).$$

Consequently, if any of the sequences $\mathfrak{L}_n \left(\sigma_n^{-1}\Delta|X \right)$ or $\mathfrak{L}_n \left(\sigma_n^{-1}\Delta|\theta \right)$ converges in some sense, e.g., either weakly or in total variation/Hellinger distance, then the limit is necessarily γ. In particular, we see that if the distributions under discussion are supported by ℓ^2, which is the case when $(\tau_k) \in \ell^2$, then the convergence can not hold in total variation/Hellinger norm since γ is not supported by ℓ^2; in fact, we have $\gamma(\ell^2) = 0$. In order to assess convergence to γ, it will be useful to construct a Hilbert space $\mathbb{H}$ which supports all the measures under consideration, i.e., γ, $\mathfrak{L}_n(\Delta|X)$ and $\mathfrak{L}_n(\Delta|\theta)$, for all n. Such a space is given by

$$\mathbb{H} := \left\{ x \in \mathbb{R}^\infty : \ \|x\|_{\mathbb{H}}^2 := \sum_{k=1}^{\infty} \lambda_k^2 x_k^2 < \infty \right\}, \tag{10}$$

for some sequence of positive numbers $(\lambda_k) \in \ell^2$. Indeed, if $\{\mathbf{e}_k\}_{k \geq 1}$ are the canonical unit vectors in $\mathbb{R}^\infty$ then an orthonormal system in $\mathbb{H}$ is given by $\{\mathbf{h}_k := \lambda_k^{-1}\mathbf{e}_k\}_{k \geq 1}$. If S denotes the covariance operator of γ in $\mathbb{H}$ then we have

$$\langle S\mathbf{h}_k|\mathbf{h}_l \rangle_{\mathbb{H}} = \int_{\mathbb{H}} \langle t|\mathbf{h}_k \rangle_{\mathbb{H}} \langle t|\mathbf{h}_l \rangle_{\mathbb{H}} \gamma(dt).$$

Now $t = (t_k) \in \mathbb{R}^\infty$ and t_k are i.i.d. $\mathcal{N}(0;1)$ variables under γ. Therefore,

$$\langle t|\mathbf{h}_k \rangle_{\mathbb{H}} = \lambda_k t_k \Rightarrow \langle S\mathbf{h}_k|\mathbf{h}_l \rangle_{\mathbb{H}} = \lambda_k^2 \delta_{kl},$$

---

[3]Provided that $\psi : \mathbb{R}^\infty \to \mathbb{R}$ is measurable, $\gamma \circ \psi^{-1}(B) := \gamma\{x : \psi(x) \in B\}$ always defines a measure on the Borel sets of $\mathbb{R}$, having total mass at most 1. If, in addition, ψ is defined γ-a.s. then $\gamma\{x : \psi(x) \in \mathbb{R}\} = 1$, hence $\gamma \circ \psi^{-1}$ defines a probability measure on $\mathbb{R}$.

i.e., $S\mathbf{h}_k = \lambda_k^2 \mathbf{h}_k$, for any k, hence S is a linear operator defined by the eigenvalues $\{\lambda_k^2\}_{k\geq 1}$ w.r.t. $\{\mathbf{h}_k\}_{k\geq 1}$. Therefore, the condition $(\lambda_k) \in \ell^2$ guarantees that the covariance operator of γ in $\mathbb{H}$ is of trace class, hence $\gamma = \mathcal{N}(\mathbf{0}; S)$ in $\mathbb{H}$. In the same vein, one can easily check that the covariance operators of $\mathfrak{L}_n\left(\sigma_n^{-1}\Delta|X\right)$ and $\mathfrak{L}_n\left(\sigma_n^{-1}\Delta|\theta\right)$ are defined by the eigenvalues

$$\left\{\frac{\lambda_k^2 \tau_k^2}{\sigma_n^2 + \tau_k^2}\right\}_{k\geq 1}, \ \left\{\frac{\lambda_k^2 \tau_k^4}{(\sigma_n^2 + \tau_k^2)^2}\right\}_{k\geq 1}, \tag{11}$$

respectively, and are of trace-class if $(\lambda_k) \in \ell^2$. Clearly, $\ell^2 \subset \mathbb{H}$, the inclusion being proper, and both $\mathfrak{L}_n(\Delta|X)$ and $\mathfrak{L}_n(\Delta|\theta)$ are supported by $\mathbb{H}$, for any prior π and any $n \geq 1$. The space $\mathbb{H}$ has rather theoretical significance and will play little role in what follows; one can take, for instance, $\lambda_k = 1/k$, for $k \geq 1$.

Remark 1. Let $\mathbb{H}$ be defined by (10) for some positive sequence $(\lambda_k) \in \ell^2$, i.e., $\mathbb{H}$ supports γ, and let $\psi = (\psi_k)$ denote some bounded linear functional on $\mathbb{H}$. Define $x = (x_k)$, with $x_k = \text{sign}(\psi_k)$, for $k \geq 1$. Since $|x_k| = 1$ for any k, it follows that $x \in \mathbb{H}$, hence $\psi(x) = \|\psi\|_{\ell^1} < \infty$; that is, $\psi \in \ell^1$. Conversely, let $\psi \in \ell^1$ and set $\lambda_k := \sqrt{|\psi_k| + (1/k^2)}$, for all $k \geq 1$. It is easy to check that these λ_k's are positive, such that $(\lambda_k) \in \ell^2$ and ψ defines a linear functional on $\mathbb{H}$ (defined by these λ_k's) with norm less than $\sqrt{\|\psi\|_{\ell^1}}$. Indeed, for any $x \in \mathbb{H}$,

$$|\psi(x)| \leq \sum_{k\geq 1} |\psi_k x_k| = \sum_{k\geq 1} \frac{|\psi_k|}{\lambda_k} \cdot (\lambda_k|x_k|) \leq \sqrt{\sum_{k\geq 1} \frac{\psi_k^2}{\lambda_k^2}} \cdot \|x\|_{\mathbb{H}} \leq \sqrt{\|\psi\|_{\ell^1}} \cdot \|x\|_{\mathbb{H}}.$$

This justifies the terminology "bounded" used for linear functionals $\psi \in \ell^1$.

Let now $\mathbb{H} \subset \mathbb{R}^\infty$ be defined by (10), for some positive sequence $(\lambda_k) \in \ell^2$, i.e., $\mathbb{H}$ supports all the measures under discussion. For simplicity, we denote by T_n and S_n the linear operators on $\mathbb{H}$ having eigenvalues given by (11) w.r.t. the canonical unit vectors in $\mathbb{R}^\infty$ and set

$$b_n^\theta := \mathbb{E}\left[\sigma_n^{-1}\Delta|\theta\right] = \left(\frac{-\sigma_n \theta_k}{\sigma_n^2 + \tau_k^2}\right)_k.$$

With these notations, $\mathfrak{L}_n(\sigma_n^{-1}\Delta|X) = \mathcal{N}(\mathbf{0}; T_n)$ and $\mathfrak{L}_n(\sigma_n^{-1}\Delta|\theta) = \mathcal{N}(b_n^\theta; S_n)$. The following result shows, in particular, that a weak version of the π-a.s. BvM statement holds for Gaussian priors π having diagonal covariance structure.

Theorem 1. $\mathcal{N}(\mathbf{0}; T_n)$ *converges weakly to* γ *in* $\mathbb{H}$ *and, for almost all* θ's *drawn from* π, $\mathcal{N}(b_n^\theta; S_n)$ *converges weakly to* γ *in* $\mathbb{H}$, *as well. In particular, for any measurable set* $B \subset \mathbb{H}$ *satisfying* $\gamma(\partial B) = 0$ *it holds that*

$$\lim_{n\to\infty} \mathbb{P}\left\{\Delta \in \sigma_n B | X\right\} - \mathbb{P}\left\{\Delta \in \sigma_n B | \theta\right\} = 0, \ \pi - a.s. \tag{12}$$

It is important to note that, even when $\mathcal{N}(\mathbf{0}; T_n)$ and $\mathcal{N}(b_n^\theta; S_n)$ are supported by ℓ^2, the weak convergence in Theorem 1 does not hold in ℓ^2, but in the larger space $\mathbb{H}$ which supports γ. In fact, although supported by ℓ^2, the two sequences are not tight in ℓ^2. The above analysis shows, in particular, that any reasonable BvM statement for this model can only be obtained beyond the ℓ^2-framework.

2.2. The BvM statement

In this section, we seek necessary/sufficient conditions for π and θ such that

$$\|\mathfrak{L}_n\left(\Delta|X\right) - \mathfrak{L}_n\left(\Delta|\theta\right)\|_H \to 0; \tag{13}$$

in other words, given a prior π with diagonal covariance structure on $\mathbb{R}^\infty$, we aim to characterize/determine a (maximal) parameter set Θ for which the BvM statement holds. For a π-a.s. statement, we check whether $\pi(\Theta) = 1$.

Since the Hellinger distance is invariant to re-scaling, (13) is equivalent to

$$\|\mathcal{N}(\mathbf{0}; T_n) - \mathcal{N}(b_n^\theta; S_n)\|_H \to 0. \tag{14}$$

A necessary condition for the convergence in the last display is that $\mathcal{N}(\mathbf{0}; T_n)$ and $\mathcal{N}(b_n^\theta; S_n)$ are equivalent measures, for n large; otherwise, if they are orthogonal along some subsequence of n's then the limit along this subsequence will be strictly larger than 0. By the Equivalence Theorem, Gaussian measures $\mathcal{N}(\mathbf{0}; T_n)$ and $\mathcal{N}(b_n^\theta; S_n)$ are either equivalent or orthogonal and equivalence obtains if and only if both of them are equivalent to $\mathcal{N}(\mathbf{0}; S_n)$. By the Cameron-Martin Theorem, equivalence between $\mathcal{N}(b_n^\theta; S_n)$ and $\mathcal{N}(\mathbf{0}; S_n)$ requires that $b_n^\theta \in \sqrt{S_n}\mathbb{H}$. Now recall that $\sqrt{S_n}\mathbb{H}$ is also a Hilbert space and a complete orthonormal system in this space can be obtained as follows:

$$\forall k \geq 1 : \mathbf{f}_k := \sqrt{S_n}\mathbf{h}_k = \frac{\lambda_k \tau_k^2}{\sigma_n^2 + \tau_k^2}\mathbf{h}_k = \frac{\tau_k^2}{\sigma_n^2 + \tau_k^2}\mathbf{e}_k.$$

Hence, $b_n^\theta \in \sqrt{S_n}\mathbb{H}$ iff $b_n^\theta = \sum_{k \geq 1} u_k \mathbf{f}_k$, for some sequence $(u_k) \in \ell^2$. Since

$$b_n^\theta = \sum_{k \geq 1} \frac{-\sigma_n \theta_k}{\sigma_n^2 + \tau_k^2}\mathbf{e}_k = \sum_{k \geq 1} u_k \mathbf{f}_k = \sum_{k \geq 1} \frac{u_k \tau_k^2}{\sigma_n^2 + \tau_k^2}\mathbf{e}_k,$$

one concludes, after identifying the coefficients, that $u_k = -\sigma_n \theta_k/\tau_k^2$. Hence, by Cameron-Martin Theorem, the equivalence $\mathcal{N}(b_n^\theta; S_n) \equiv \mathcal{N}(\mathbf{0}; S_n)$ obtains iff

$$\left(\frac{\theta_k}{\tau_k^2}\right) \in \ell^2. \tag{15}$$

On the other hand, by Feldman-Hajek Theorem, $\mathcal{N}(\mathbf{0}; T_n) \equiv \mathcal{N}(\mathbf{0}; S_n)$ requires

$$\mathcal{S}_n(\pi) := \sum_{k \geq 1} \left(\frac{\frac{\lambda_k^2 \tau_k^2}{\sigma_n^2 + \tau_k^2} - \frac{\lambda_k^2 \tau_k^4}{(\sigma_n^2 + \tau_k^2)^2}}{\frac{\lambda_k^2 \tau_k^2}{\sigma_n^2 + \tau_k^2} + \frac{\lambda_k^2 \tau_k^4}{(\sigma_n^2 + \tau_k^2)^2}}\right)^2 = \sum_{k \geq 1} \frac{\sigma_n^4}{(\sigma_n^2 + 2\tau_k^2)^2} < \infty; \tag{16}$$

for the last inequality it suffices that $(1/\tau_k^2) \in \ell^2$. Therefore, (15) and (16) provide necessary and sufficient conditions for the equivalence between $\mathcal{N}(\mathbf{0}; T_n)$ and $\mathcal{N}(b_n^\theta; S_n)$, for large n, which is a necessary condition for the validity of (14). In particular, the conditions $(\theta_k/\tau_k^2) \in \ell^2$ and $(1/\tau_k^2) \in \ell^2$ guarantee the equivalence $\mathcal{N}(\mathbf{0}; T_n) \equiv \mathcal{N}(b_n^\theta; S_n)$. It turns out that these conditions are both necessary and sufficient for the validity of (14). More specifically, we have:

Lemma 2. *Let $\theta = (\theta_k) \in \mathbb{R}^\infty$ be arbitrary and let $\pi = \bigotimes_{k \geq 1} \mathcal{N}(0; \tau_k^2)$, for some arbitrary $\tau_k > 0$; $k \geq 1$. Then the following statements are equivalent:*

(i) *$(1/\tau_k^2) \in \ell^2$ and $(\theta_k/\tau_k^2) \in \ell^2$.*
(ii) *$\|\mathcal{N}(\mathbf{0}; T_n) - \gamma\|_H \to 0$ and $\|\mathcal{N}(b_n^\theta; S_n) - \gamma\|_H \to 0$.*
(iii) *$\|\mathcal{N}(\mathbf{0}; T_n) - \mathcal{N}(b_n^\theta; S_n)\|_H \to 0$.*

Lemma 2 shows that the condition $(1/\tau_k^2) \in \ell^2$ is crucial for the validity of the BvM statements, regardless of θ. Namely, if the condition holds then the BvM statement holds true for the parameter set

$$\Theta := \left\{ \theta \in \mathbb{R}^\infty : \left(\theta_k/\tau_k^2 \right) \in \ell^2 \right\} \tag{17}$$

In particular, we note that in this case $\ell^2 \subset \Theta$, hence if the true parameter θ belongs to ℓ^2 then the BvM statement holds for any prior satisfying $(1/\tau_k^2) \in \ell^2$. Comparing the result in Lemma 2 to that in Theorem 1 we see that the above condition is needed to obtain convergence in total variation/Hellinger norm instead of weak convergence. Both statements, however, show that whenever a BvM statement holds, the two measures must necessarily converge to γ.

On the other hand, if the condition is not fulfilled, e.g., if the sequence $\{\tau_k\}_{k \geq 1}$ is upper-bounded, as it is in [3], then the BvM statement does not hold, for any (nonempty) parameter set Θ. This, in particular, shows that there is no parameter $\theta \in \ell^2$ such that (13) holds; the results in [3] only show that for most of the θ's in ℓ^2 (in both probabilistic and topological sense) the statement fails.

Finally, the parameter set Θ defined in (17) satisfies $\pi(\Theta) = 1$ if and only if $(1/\tau_k) \in \ell^2$; otherwise, we have $\pi(\Theta) = 0$. Indeed, recall that, under the prior π, we have $\vartheta_k = \tau_k \xi_k$, with $\{\xi_k\}_{k \geq 1}$ i.i.d. standard Gaussian variables. Hence, by Kolmogorov's three-series Theorem, one concludes that the random series $\sum_k (\vartheta_k^2/\tau_k^4) = \sum_k (\xi_k/\tau_k)^2$ either converges or diverges with probability 1 and convergence obtains if and only if $(1/\tau_k) \in \ell^2$. One can synthesize this analysis into the following statement which gives a complete overview of the validity of the BvM statements; the proof follows by Lemma 2 and by previous remarks.

Theorem 2. *If the prior π in Lemma 2 satisfies $(1/\tau_k) \in \ell^2$ then the BvM statement holds true π-a.s. If $(1/\tau_k^2) \in \ell^2$ but $(1/\tau_k) \notin \ell^2$ the BvM statement holds for the parameter set Θ in (17), having null prior probability. Finally, if $(1/\tau_k^2) \notin \ell^2$ the BvM statement fails for any nonempty parameter set Θ.*

Let us consider $\mathcal{T} := \left\{ (x_k) \in \mathbb{R}^\infty : (x_k/\tau_k) \in \ell^2 \right\}$ and Θ defined by (17). Note that the linear spaces $\mathcal{T}$ and Θ become Hilbert spaces when endowed with the norms $\|(x_k)\|_{\mathcal{T}} := \|(x_k/\tau_k)\|$ and $\|(x_k)\|_\Theta := \|(x_k/\tau_k^2)\|$, respectively, and $\mathcal{T}$ is the RHS of π. If $\tau_k \to \infty$ then $\ell^2 \subset \mathcal{T} \subset \Theta$, the embeddings being continuous. The condition $(1/\tau_k^2) \in \ell^2$ is equivalent to the fact that the covariation operator of γ in Θ is of trace-class, hence γ is supported by Θ. Under the stronger condition $(1/\tau_k) \in \ell^2$, the covariation operator of γ in $\mathcal{T}$ (the RHS of π) becomes of trace-class; that is, the noise ε is supported by the RHS of π or, yet, the prior π is equivalent to the distribution of the data X (Cameron-Martin Theorem), γ-a.s. By virtue of Theorem 2, this condition seems both necessary and sufficient for the validity of the π-a.s. BvM statement for the model under discussion.

2.3. Conclusions and remarks

Typically, the prior π is supported by ℓ^2; see, e.g., [2, 3]. Although the BvM statement holds for bounded linear functionals and in a weak sense, it does not hold in the sense of (13), for any θ. In fact, in this case many irregularities occur due to infinite-dimensional nature of the problem. For instance, the probability measures $\mathfrak{L}_n(\sigma_n^{-1}\Delta|X)$ and $\mathfrak{L}_n(\sigma_n^{-1}\Delta|\theta)$ are always, orthogonal and this is the most evident reason why the BvM Theorem does not hold in this case. Moreover, although the two corresponding sequences of measures are supported also by ℓ^2, they are not tight in ℓ^2 since their finite-dimensional projections converge to those of γ, which is not supported by ℓ^2. Intuitively, this means that compact credible sets/confidence regions, for arbitrarily large n, do not exist in ℓ^2. One may then think of embedding ℓ^2 into some larger Hilbert space $\mathbb{H}$, which supports the limiting measure γ, and use the weak version in Theorem 1. Although such a result holds for π-almost all θ, this is of no avail in terms of ℓ^2-confidence regions since one can only apply it for (credible) sets whose boundary is not charged by γ and this is not the case for "most of" the subsets of ℓ^2; recall that $\gamma(\ell^2) = 0$, so that γ is concentrated on the boundary of ℓ^2 in $\mathbb{H}$. Nevertheless, by virtue of Lemma 1, one can still use credible sets of $\mathbb{H}$ with γ-negligible boundary, e.g., open balls in $\mathbb{H}$, as confidence regions, cf. (12). The open balls in $\mathbb{H}$, however, are much wider than their ℓ^2-counterparts. In fact, the centered $\mathbb{H}$-ball of radius δ appears intuitively as a huge ellipsoid in ℓ^2, with semi-axes $\{\delta/\lambda_k\}_{k\geq 1}$ tending to infinity and this might be inconvenient in applications as it yields very slow convergence rates for many functionals of interest.

Finally, we note that if $\mathbb{P}_\theta$ denotes the true distribution of the data X then $\mathbb{P}_\theta = \mathcal{N}(\theta; \sigma_n^2 I)$, where I denotes the identity operator on $\mathbb{R}^\infty$, and the statistical model $\{\mathbb{P}_\theta : \theta \in \ell^2\}$ is dominated by $\mathbb{P}_\mathbf{0}$, by the Cameron-Martin Theorem. The log-likelihood w.r.t. $\mathbb{P}_\mathbf{0}$ is given by (below $\langle\cdot|\cdot\rangle^\sim$ is relative to $\langle\cdot|\cdot\rangle$ in ℓ^2)

$$\ell_\theta(X) = \frac{\langle\theta|X\rangle^\sim}{\sigma_n^2} - \frac{\|\theta\|^2}{2\sigma_n^2}.$$

The above expression is differentiable w.r.t. θ, in the Malliavin sense (hence in quadratic mean) and we have $\dot{\ell}_\theta(X) = (X - \theta)/\sigma_n^2$. Since under $\mathbb{P}_\theta$ each $(X - \theta)_k$ is a $\mathcal{N}(0; \sigma_n^2)$-variable, it readily follows that the covariance operator of the score $\dot{\ell}_\theta(X)$ is formally given by $\sigma_n^{-2}I$, so that $\sigma_n^2 I$ appears, in some sense, as the inverse Fisher information. Provided that $(1/\tau_k^2) \in \ell^2$, Lemma 2 implies $\mathfrak{L}_n(\vartheta - \hat{\vartheta}|X) \simeq \mathcal{N}(0; \sigma_n^2 I)$, or $\mathfrak{L}_n(\vartheta|X) \simeq \mathcal{N}(\hat{\vartheta}; \sigma_n^2 I)$ which, for $\sigma_n^2 = 1/n$, looks quasi-similar to the standard parametric statement in which the posterior mean plays the role of the asymptotically efficient estimator. Although formal, the above reasoning may suggest the lines along which the BvM Theorem can be generalized to this nonparametric model. Also, statement (iii) in Lemma 1 shows that for any ℓ^1-functional ψ the frequentist distribution $\mathfrak{L}_n(\psi(\Delta)|\theta)$ is asymptotically normal with variance $\sigma_n^2\|\psi\|^2$. This suggests that, for well-behaved functionals, the projected posterior mean $\psi(\hat{\vartheta})$ is an asymptotically efficient estimator for $\psi(\theta)$. Going further on this track, one may establish a semi-parametric BvM Theorem for such functionals, obtaining results similar to those in [1, 11].

3. The BvM statement for the squared-norm functional

Throughout this section we shall assume that the prior π satisfies $\tau_k^2 \sim k^{-(1+2\alpha)}$, for some $\alpha > 0$, and we shall investigate the validity of the BvM statement for the functional $\psi(\Delta) = \|\Delta\|^2$, for a certain class of parameter subsets in ℓ^2. The corresponding π-a.s. statement was treated in detail in [3] and it has been proved to be invalid, as shall be explained below.

To proceed to our analysis, we set $A_\pi := \lim_k k^{1+2\alpha} \tau_k^2$; by assumption, we have $0 < A_\pi < \infty$. Moreover, for notational convenience, we define the family of constants $K_{\varpi,\eta}^\lambda$, for $\lambda \geq 0$, $\varpi > 0$ and $\eta > 1$ satisfying $1 + \lambda < \varpi\eta$, as follows:

$$K_{\varpi,\eta}^\lambda := \int_0^\infty \frac{t^\lambda}{(1 + t^\varpi)^\eta} dt.$$

For later reference we note that for suitable λ, ϖ, η we have $K_{\varpi,\eta}^\lambda > 0$ and for any integer p satisfying $1 \leq p < \eta$ and $\varpi(\eta - p) > 1$ it holds that

$$\sum_{i=0}^p \binom{p}{i} K_{\varpi,\eta}^{i\cdot\varpi} = K_{\varpi,\eta-p}^0, \tag{18}$$

The results obtained in [3] can be aggregated into the following statement.

Theorem 3. *(Freedman 99) Let $\tau_k^2 \approx k^{-(1+2\alpha)} A_\pi$, for some $\alpha > 0$ and $A_\pi > 0$ and consider the following representation:*

$$\|\Delta\|^2 = M_n + Q_n(\theta) + Z_n(\theta, \varepsilon); \tag{19}$$

analytic expressions for M_n, Q_n, and Z_n are provided in the Appendix. Then:

(i) *M_n are real numbers satisfying $M_n \approx A_\pi^{\frac{1}{1+2\alpha}} K_{1+2\alpha,1}^0 \, n^{-\frac{2\alpha}{1+2\alpha}}$.*
(ii) *Under π, $Q_n(\theta)$ are random variables with null mean satisfying*

$$\mathrm{Var}[Q_n(\vartheta)] \approx 2A_\pi^{\frac{1}{1+2\alpha}} K_{1+2\alpha,4}^{2+4\alpha} \, n^{-\frac{1+4\alpha}{1+2\alpha}}.$$

Furthermore, $Q_n(\theta) \simeq \mathcal{N}(0; \mathrm{Var}[Q_n(\vartheta)])$, π-a.s., and it holds that

$$\liminf \frac{Q_n(\theta)}{\sqrt{\mathrm{Var}[Q_n(\vartheta)]}} = -\infty, \ \limsup \frac{Q_n(\theta)}{\sqrt{\mathrm{Var}[Q_n(\vartheta)]}} = \infty, \ \pi - a.s.$$

(iii) *For each θ, $Z_n(\theta, \varepsilon)$ are random variables with null mean satisfying*

$$\mathrm{Var}[Z_n(\theta, \varepsilon)] \approx 2A_\pi^{\frac{1}{1+2\alpha}} \left(K_{1+2\alpha,4}^0 + 2K_{1+2\alpha,4}^{1+2\alpha} \right) n^{-\frac{1+4\alpha}{1+2\alpha}}, \ \pi - a.s.$$

In addition, it holds that $Z_n(\theta, \varepsilon) \simeq \mathcal{N}(0; \mathrm{Var}[Z_n(\theta, \varepsilon)])$, π-a.s.
(iv) *Under π, $Z_n(\theta, \varepsilon)$ are random variables uncorrelated with $Q_n(\theta)$, such that*

$$\frac{Q_n(\vartheta) + Z_n(\vartheta, \varepsilon)}{\sqrt{\mathrm{Var}[Q_n(\vartheta)] + \mathrm{Var}[Z_n(\vartheta, \varepsilon)]}} \simeq \mathcal{N}(0; 1).$$

By Theorem 3, the Bayesian expectation and variance of $\|\Delta\|^2$ are given by

$$\mathbb{E}_n\left[\|\Delta\|^2|X\right] = M_n, \ \text{Var}_n\left[\|\Delta\|^2|X\right] = \text{Var}[Q_n(\vartheta)] + \text{Var}[Z_n(\vartheta,\varepsilon)],$$

whereas their frequentist counterparts are readily given by

$$\mathbb{E}_n\left[\|\Delta\|^2|\theta\right] = M_n + Q_n(\theta), \ \text{Var}_n\left[\|\Delta\|^2|\theta\right] = \text{Var}[Z_n(\theta,\varepsilon)].$$

Moreover, (iv) shows that the asymptotic behavior of $\mathfrak{L}_n(\|\Delta\|^2|X)$ satisfies

$$\mathfrak{L}_n(\|\Delta\|^2|X) \simeq \mathcal{N}\left(M_n; \text{Var}[Q_n(\vartheta)] + \text{Var}[Z_n(\vartheta,\varepsilon)]\right), \quad (20)$$

while, according to (iii), the asymptotic behavior of $\mathfrak{L}_n(\|\Delta\|^2|\theta)$ is described by

$$\mathfrak{L}_n(\|\Delta\|^2|\theta) \simeq \mathcal{N}\left(M_n + Q_n(\theta); \text{Var}[Z_n(\theta,\varepsilon)]\right), \ \pi - a.s. \quad (21)$$

Defining now for any $n \geq 1$

$$D_n^2(\theta) := \frac{\text{Var}[Z_n(\theta,\varepsilon)]}{\text{Var}\left[\|\Delta\|^2|X\right]}, \ C_n^2(\theta) := \frac{Q_n^2(\theta)}{(1 + D_n^2(\theta))\,\text{Var}\left[\|\Delta\|^2|X\right]},$$

one can approximate the Hellinger affinity of $\mathfrak{L}_n(\|\Delta\|^2|X)$ and $\mathfrak{L}_n(\|\Delta\|^2|\theta)$ by the corresponding affinity of their Gaussian approximations given by (20) and (21), respectively; more specifically, we have

$$\mathcal{A}_n(\pi,\theta) := \sqrt{\frac{2D_n(\theta)}{1 + D_n^2(\theta)}} \cdot e^{-\frac{1}{4}C_n^2(\theta)}. \quad (22)$$

To see now that $\mathcal{A}_n(\pi,\theta) \nrightarrow 1$, π-a.s., note first that

$$\text{Var}[Z_n(\vartheta,\varepsilon)] = \int \text{Var}[Z_n(\theta,\varepsilon)]\pi(d\theta) \approx 2A_\pi^{\frac{1}{1+2\alpha}}\left(K_{1+2\alpha,4}^0 + 2K_{1+2\alpha,4}^{1+2\alpha}\right)\,n^{-\frac{1+4\alpha}{1+2\alpha}}.$$

Therefore, taking $\varpi = 1 + 2\alpha$, $\eta = 4$ and $p = 2$ in (18) one obtains the estimate

$$\text{Var}\left[\|\Delta\|^2|X\right] = \text{Var}[Q_n(\vartheta)] + \text{Var}[Z_n(\vartheta,\varepsilon)] \approx 2A_\pi^{\frac{1}{1+2\alpha}}K_{1+2\alpha,2}^0\,n^{-\frac{1+4\alpha}{1+2\alpha}},$$

hence, according to (iii), we obtain $\lim_n D_n^2(\theta) = D^2(\theta)$, π-a.s., with

$$D^2(\theta) := \frac{K_{1+2\alpha,4}^0 + 2K_{1+2\alpha,4}^{1+2\alpha}}{K_{1+2\alpha,2}^0} = 1 - \frac{K_{1+2\alpha,4}^{2+4\alpha}}{K_{1+2\alpha,2}^0} \in (0,1).$$

In particular, we have $D_n^2(\theta) \sim 1$ and $\text{Var}\left[\|\Delta\|^2|X\right] \sim \text{Var}[Q_n(\vartheta)]$, hence one concludes by (ii) that $\limsup C_n^2(\theta) = \infty$, π-a.s. This leads to

$$0 = \liminf \mathcal{A}_n(\pi,\theta) \leq \limsup \mathcal{A}_n(\pi,\theta) \leq \sqrt{\frac{2D(\theta)}{1 + D^2(\theta)}} < 1, \ \pi - a.s. \quad (23)$$

We also note that, provided that the Gaussian approximation (21) holds true for θ, $\lim_n \mathcal{A}_n(\pi, \theta) = 1$ if and only if $C_n^2(\theta) \to 0$ and $D_n^2(\theta) \to 1$, or, equivalently, $Q_n^2(\theta) \ll \mathrm{Var}[\|\Delta\|^2|\theta] \approx \mathrm{Var}[\|\Delta\|^2|\theta]$. Unfortunately, this is not the case π-a.s. We conclude that, for almost all θ's drawn from the prior π, the Hellinger affinity $\mathcal{A}_n(\pi, \theta)$ converges to 0 along some subsequence of n's; that is, $\mathfrak{L}_n(\|\Delta\|^2|X)$ and $\mathfrak{L}_n(\|\Delta\|^2|\theta)$ will be almost orthogonal for arbitrarily large n, π-a.s. Moreover, even for "nice" subsequences, the limiting affinity between the two measures is strictly below 1. Intuitively, the degree of overlapping between the two measures may not exceed a certain threshold $D^2(\theta) < 1$. Although formal, this argument can be made precise. The conclusion is that for π-almost all θ's the asymptotic behavior of the frequentist distribution $\mathfrak{L}_n(\|\Delta\|^2|\theta)$ is essentially different from that of the Bayesian distribution $\mathfrak{L}_n(\|\Delta\|^2|X)$, in the sense that the two distributions concentrate their mass on disjoint intervals.

3.1. Parameters with given level of smoothness

A typical assumption made by statisticians is that the true parameter θ has some pre-specified level of smoothness; see, e.g., [14], so that would be interesting to investigate whether the BvM statement for the squared ℓ^2-norm holds for sets of parameters having certain smoothness properties. Throughout this section we consider the Hilbert scale $\{(\Theta_\delta, \| \cdot \|_\delta) : \delta \leq \alpha\} \subset \ell^2$ defined by

$$\Theta_\delta := \left\{ \theta = (\theta_k) : \|\theta\|_\delta^2 := \sum_{k=1}^\infty k^{2(\alpha-\delta)} \theta_k^2 < \infty \right\},$$

and check if the BvM statement holds for the squared ℓ^2-norm for some Θ_δ. Note that, for $\delta = \alpha$ we have $\Theta_\alpha = \ell^2$ while the choice $\delta = -1/2$ corresponds to the RHS of the prior π. In addition, $\pi(\Theta_\delta) = 0$, for $\delta \leq 0$ and $\pi(\Theta_\delta) = 1$, for $0 < \delta \leq \alpha$; hence Θ_0 appears as the largest Θ_δ of null prior probability. In what follows, we investigate the validity of the BvM statement for the squared ℓ^2-norm, for the parameter set Θ_δ, for $\delta \leq \alpha$. Since the family $\{\Theta_\delta\}_\delta$ is increasing, if the BvM statement holds for some Θ_δ, then it holds for any $\Theta_{\delta'}$, with $\delta' \leq \delta$.

Remark 2. By Theorem 3, there exists some (unknown) set $\Omega \subset \ell^2$, such that $\pi(\Omega) = 1$ and (23) holds true for $\theta \in \Omega$. Since $\pi(\Theta_\delta) = 1$, for $\delta > 0$, it follows that $\Omega \cap \Theta_\delta$ has π-probability 1, hence it is certainly a non-empty set. This shows that the BvM statement for the squared ℓ^2-norm fails, for any parameter set Θ_δ, with $\delta > 0$; that is, there exists parameters $\theta \in \Theta_\delta$ for which (23) holds true.

In the light of the above remark, one could only hope that the BvM statement holds true for a parameter set Θ_δ, with $\delta \leq 0$. In the reminder of this section we shall prove that the BvM statement for the squared ℓ^2-norm does not hold for any Θ_δ, with $\delta \leq 0$, either. To this end, we consider the sets

$$\mathcal{B}_\omega := \left\{ \theta = (\theta_k) : \theta_k^2 \sim k^{-(1+2\omega)} \right\},$$

for $\omega > 0$. Note that $\mathcal{B}_\omega \subset \ell^2$ are mutually disjoint sets with $\pi(\mathcal{B}_\omega) = 0$. The connection between Θ_δ and $\mathcal{B}_\omega$ is established by the following statement.

Lemma 3. *Let $\delta \leq \alpha$. If $\omega > \alpha - \delta$ then $\mathcal{B}_\omega$ is a dense subset of $(\Theta_\delta, \|\cdot\|_\delta)$. Otherwise, if $\omega \leq \alpha - \delta$, for some $\delta < \alpha$, it holds that $\mathcal{B}_\omega \cap \Theta_\delta = \emptyset$.*

The main reason for considering these sets is that for $\theta \in \mathcal{B}_\omega$ one can obtain exact asymptotics for $\mathbb{E}_n[\|\Delta\|^2|\theta]$ and $\mathrm{Var}_n[\|\Delta\|^2|\theta]$, via Lemma 8 (Appendix). Namely, assume that $\theta \in \mathcal{B}_\omega$ and let $L_\theta := \lim_k k^{1+2\omega}\theta_k^2$. Using the expressions in (31) and (32) (Appendix), for $\sigma_n^2 = 1/n$ and $\tau_k^2 = A_\pi k^{-(1+2\alpha)}$ we obtain[4]

$$M_n + Q_n(\theta) = \sum_{k \geq 1} \frac{A_\pi^2 n}{(A_\pi n + k^{1+2\alpha})^2} + \sum_{k \geq 1} \frac{k^{2+4\alpha}\theta_k^2}{(A_\pi n + k^{1+2\alpha})^2} = T_n(\pi) + U_n(\pi, \theta),$$

respectively,

$$\mathrm{Var}[Z_n(\theta, \varepsilon)] = \sum_{k \geq 1}^{\infty} \frac{2A_\pi^4 n^2}{(A_\pi n + k^{1+2\alpha})^4} + \sum_{k \geq 1}^{\infty} \frac{4A_\pi^2 n k^{2+4\alpha}\theta_k^2}{(A_\pi n + k^{1+2\alpha})^4} = V_n(\pi) + W_n(\pi, \theta).$$

For the choices $\varpi = 1 + 2\alpha$, $\eta = 2$ and $\lambda = 0$, respectively $\lambda = 1 + 4\alpha - 2\omega$, we obtain by Lemma 8 the following estimates:

$$T_n(\pi) \approx A_\pi^{\frac{1}{1+2\alpha}} K_{1+2\alpha,2}^0\, n^{-\frac{2\alpha}{1+2\alpha}}, \ U_n(\pi,\theta) \approx L_\theta A_\pi^{-\frac{2\omega}{1+2\alpha}} K_{1+2\alpha,2}^{1+4\alpha-2\omega}\, n^{-\frac{2\omega}{1+2\alpha}},$$

while for $\varpi = 1 + 2\alpha$, $\eta = 4$ and $\lambda = 0$, respectively $\lambda = 1 + 4\alpha - 2\omega$, we obtain

$$V_n(\pi) \approx 2A_\pi^{\frac{1}{1+2\alpha}} K_{1+2\alpha,4}^0\, n^{-\frac{1+4\alpha}{1+2\alpha}}, \ W_n(\pi,\theta) \approx 4L_\theta A_\pi^{-\frac{2\omega}{1+2\alpha}} K_{1+2\alpha,4}^{1+4\alpha-2\omega}\, n^{-\frac{1+2\alpha+2\omega}{1+2\alpha}}.$$

Therefore, one has to distinguish between the following three situations which arise naturally when comparing $T_n(\pi)$ vs. $U_n(\pi, \theta)$ and $V_n(\pi)$ vs. $W_n(\pi, \theta)$:

(i) the *over-smoothing* case corresponds to the situation $0 < \omega < \alpha$. In this case, cf. Lemma 8, it holds that $U_n(\pi, \theta) \gg T_n(\pi)$ and $W_n(\pi, \theta) \gg V_n(\pi)$, hence the sampling mean satisfies

$$\mathbb{E}_n\left[\|\Delta\|^2|\theta\right] \approx L_\theta A_\pi^{-\frac{2\omega}{1+2\alpha}} K_{1+2\alpha,2}^{1+4\alpha-2\omega}\, n^{-\frac{2\omega}{1+2\alpha}}, \tag{24}$$

whereas the sampling variance satisfies

$$\mathrm{Var}_n\left[\|\Delta\|^2|\theta\right] \approx 4L_\theta A_\pi^{-\frac{2\omega}{1+2\alpha}} K_{1+2\alpha,4}^{1+4\alpha-2\omega}\, n^{-\frac{1+2\alpha+2\omega}{1+2\alpha}}. \tag{25}$$

(ii) the *under-smoothing* case corresponds to the situation $\omega > \alpha$. In this case, cf. Lemma 8, it holds that $U_n(\pi, \theta) \ll T_n(\pi)$ and $W_n(\pi, \theta) \ll V_n(\pi)$, hence the sampling mean satisfies

$$\mathbb{E}_n\left[\|\Delta\|^2|\theta\right] \approx A_\pi^{\frac{1}{1+2\alpha}} K_{1+2\alpha,2}^0\, n^{-\frac{2\alpha}{1+2\alpha}}, \tag{26}$$

---

[4]By Lemma 8, if $\tau_k^2 \approx A_\pi k^{-(1+2\alpha)}$ then we have $M_n + Q_n(\theta) \approx T_n(\pi) + U_n(\pi, \theta)$ and $\mathrm{Var}[Z_n(\theta, \varepsilon)] \approx V_n(\pi) + W_n(\pi, \theta)$, so the estimates in (24)–(29) extend easily to this case.

whereas the sampling variance satisfies

$$\mathrm{Var}_n\left[\|\Delta\|^2|\theta\right] \approx 2A_\pi^{\frac{1}{1+2\alpha}} K_{1+2\alpha,4}^0 \, n^{-\frac{1+4\alpha}{1+2\alpha}}. \tag{27}$$

(iii) the *correct smoothing* case corresponds to the situation $\omega = \alpha$. In this case, cf. Lemma 8, it holds that $U_n(\pi,\theta) \sim T_n(\pi)$ and $W_n(\pi,\theta) \sim V_n(\pi)$, hence the sampling mean satisfies

$$\mathbb{E}_n\left[\|\Delta\|^2|\theta\right] \approx A_\pi^{\frac{1}{1+2\alpha}} \left[K_{1+2\alpha,2}^0 + \frac{L_\theta}{A_\pi} K_{1+2\alpha,2}^{1+2\alpha}\right] n^{-\frac{2\alpha}{1+2\alpha}}, \tag{28}$$

and the sampling variance satisfies

$$\mathrm{Var}_n\left[\|\Delta\|^2|\theta\right] \approx 2A_\pi^{\frac{1}{1+2\alpha}} \left[K_{1+2\alpha,4}^0 + 2\frac{L_\theta}{A_\pi} K_{1+2\alpha,4}^{1+2\alpha}\right] n^{-\frac{1+4\alpha}{1+2\alpha}}. \tag{29}$$

The above results suggest that the asymptotic behavior of the frequentist distribution $\mathfrak{L}_n(\|\Delta\|^2|\theta)$, for $\theta \in \mathcal{B}_\omega$, is invariant w.r.t. both θ and ω as long as $\omega > \alpha$ (under-smoothing). Now recall the definition of the space Θ_δ and note that $\delta \leq 0$ entails $\alpha - \delta \geq \alpha$. Since $\mathcal{B}_\omega \subset \Theta_\delta$ if and only if $\omega > \alpha - \delta$ it follows that for $\delta \leq 0$ the inclusion $\mathcal{B}_\omega \subset \Theta_\delta$ is true only if $\omega > \alpha$ and there is no $\omega \leq \alpha$ such that $\mathcal{B}_\omega \cap \Theta_\delta \neq \emptyset$; see Lemma 3. In other words, if $\delta \leq 0$ then Θ_δ may only contain $\mathcal{B}_\omega$'s with $\omega > \alpha$. Moreover, since on these $\mathcal{B}_\omega$'s, which are dense subsets of Θ_δ, the asymptotic behavior of $\mathfrak{L}_n(\|\Delta\|^2|\theta)$ does not depend neither on θ nor on ω, but on the smoothness of the prior only, one would expect the same behavior on the whole space Θ_δ. Our next statement uses a continuity argument to establish this fact.

Lemma 4. *Let $\delta \leq 0$. Then for any $\theta \in \Theta_\delta$ it holds that*

$$\mathbb{E}_n\left[\|\Delta\|^2|\theta\right] \approx A_\pi^{\frac{1}{1+2\alpha}} K_{1+2\alpha,2}^0 \, n^{-\frac{2\alpha}{1+2\alpha}}, \ \mathrm{Var}_n\left[\|\Delta\|^2|\theta\right] \approx 2A_\pi^{\frac{1}{1+2\alpha}} K_{1+2\alpha,4}^0 \, n^{-\frac{1+4\alpha}{1+2\alpha}}.$$

In particular, it holds that $W_n(\pi,\theta) \ll V_n(\pi) \approx \mathrm{Var}_n\left[\|\Delta\|^2|\theta\right]$.

To conclude our analysis, we need to prove that the Gaussian approximation in (21) holds true for $\theta \in \Theta_\delta$, for $\delta \leq 0$. The following result provides sufficient conditions over θ for such an approximation, based on Lindeberg-Lévy CLT.

Lemma 5. *If $\theta \in \ell^2$ is such that*

$$\lim_{n\to\infty} \frac{\max_{k\geq 1} \frac{nk^{2+4\alpha}\theta_k^2}{(A_\pi n + k^{1+2\alpha})^4}}{\mathrm{Var}_n\left[\|\Delta\|^2|\theta\right]} = 0,$$

then the Gaussian approximation in (21) holds true.

An immediate consequence of Lemma 5 is the following corollary.

Corollary 1. *Let $\omega > 0$. Then for any $\theta \in \mathcal{B}_\omega$ the Gaussian approximation in (21) holds true. Moreover, if $\delta \leq 0$, then the Gaussian approximation in (21) holds true for any $\theta \in \Theta_\delta$.*

Now recall the definitions of $\mathcal{A}_n(\pi, \theta)$ and $D_n^2(\theta)$. The following statement shows that, for $\theta \in \Theta_\delta$, with $\delta \leq 0$, the frequentist variance of $\|\Delta\|^2$ is asymptotically smaller than the Bayesian variance. Moreover, the asymptotic variance ratio $D^2(\theta)$, for such θ, is even worse (smaller) than the π-a.s. one.

Theorem 4. *Let $\delta \leq 0$. Then for any $\theta \in \Theta_\delta$ it holds that*

$$\lim_{n \to \infty} D_n^2(\theta) = \lim_{n \to \infty} \frac{\mathrm{Var}_n\left[\|\Delta\|^2|\theta\right]}{\mathrm{Var}_n\left[\|\Delta\|^2|X\right]} = \frac{K_{1+2\alpha, 4}^0}{K_{1+2\alpha, 2}^0} < 1.$$

In particular, it holds that $\limsup \mathcal{A}_n(\pi, \theta) \leq \sqrt{2D(\theta)/[1 + D^2(\theta)]} < 1$.

Theorem 4, complemented by Remark 2, shows that a BvM statement for the squared ℓ^2-norm, with parameter set Θ_δ, may not hold, for any $\delta \leq \alpha$.

3.2. Asymptotic frequentist probability coverage of credible balls

As pointed out at the beginning of this paper, one of the main features of the BvM Theorem in the parametric framework is that it allows one to use any Bayesian credible set as a frequentist confidence region. Specifically, let $p \in (0, 1)$ be some number close to 1. A measurable set $B_n \subset \ell^2$ is called a credible set if $\mathbb{P}_n\{\Delta \in B_n|X\} \geq p$ and is called a confidence region if $\mathbb{P}_n\{\Delta \in B_n|\theta\} \geq p$. The classical BvM Theorem asserts that, for n large enough, credible sets are also confidence regions and viceversa, provided that the true parameter θ and the prior π satisfies some regularity conditions. In applications, however, one is happy if (certain) credible sets can be employed as confidence regions, for large n. In the following we investigate whether centered ℓ^2-balls, which are credible sets in the sense of the above definition, can be employed as confidence regions. In other words, we investigate whether Bayesian credible (centered) ℓ^2-balls have good frequentist probability coverage, for large n, i.e., if

$$\forall p \in (0, 1): \ \lim_n \mathbb{P}_n\{\Delta \in B_n|X\} > p \Rightarrow \liminf \mathbb{P}_n\{\Delta \in B_n|\theta\} > p.$$

Throughout this section $\Phi : [-\infty, \infty] \to [0, 1]$ will denote the c.d.f. of the standard normal distribution $\mathcal{N}(0; 1)$. Based on Theorem 3 one can construct a credible ℓ^2-ball B_n^p as follows: take some $p \in (0, 1)$, close to 1 and let $\kappa_p > 0$ be such that $\Phi(\kappa_p) > p$; that is, κ_p must be larger than the p-quantile $\Phi^{-1}(p)$ of the standard Gaussian distribution. One can see now that the sets

$$B_n^p := \left\{x \in \ell^2 : \ \|x\|^2 < \mathbb{E}_n\left[\|\Delta\|^2|X\right] + \kappa_p \sqrt{\mathrm{Var}_n\left[\|\Delta\|^2|X\right]}\right\}$$

satisfy $\lim_n \mathbb{P}_n\{\Delta \in B_n^p|X\} = \Phi(\kappa_p) > p$; in particular, $\mathbb{P}_n\{\Delta \in B_n^p|X\} \geq p$, for large n. Consequently, for large n, B_n^p is a credible set which will be called a Bayesian credible ℓ^2-ball. It is interesting to note that the asymptotic behavior of the radius ρ_n of the Bayesian credible ℓ^2-balls B_n^p is given by

$$\rho_n = \left(\mathbb{E}_n\left[\|\Delta\|^2|X\right] + \kappa_p \sqrt{\mathrm{Var}_n\left[\|\Delta\|^2|X\right]}\right)^{1/2} \sim \left(\mathbb{E}_n\left[\|\Delta\|^2|X\right]\right)^{1/2} \sim n^{-\frac{\alpha}{1+2\alpha}},$$

where the above estimates follow from Theorem 3.

To investigate the asymptotic frequentist probability of the sets B_n^p, note that

$$\mathbb{P}_n\left\{\Delta \in B_n^p|\theta\right\} = \mathbb{P}_n\left\{\left.\frac{\|\Delta\|^2 - \mathbb{E}_n[\|\Delta\|^2|\theta]}{\sqrt{\mathrm{Var}_n[\|\Delta\|^2|\theta]}} < \frac{\kappa_p\sqrt{\mathrm{Var}_n[\|\Delta\|^2|X]} - Q_n(\theta)}{\sqrt{\mathrm{Var}_n[\|\Delta\|^2|\theta]}}\right|\theta\right\};$$

that is, we normalize $\|\Delta\|^2$ under its conditional law w.r.t. θ and recall that we have $\mathbb{E}_n[\|\Delta\|^2|\theta] - \mathbb{E}_n[\|\Delta\|^2|X] = Q_n(\theta)$. Provided that $\mathfrak{L}_n(\|\Delta\|^2|\theta)$ satisfies the Gaussian approximation in (21), the normalized variables appearing in the last display converge in distribution to $\mathcal{N}(0;1)$, conditionally on θ. Let us define

$$t_n(\theta) := \frac{\kappa_p\sqrt{\mathrm{Var}_n[\|\Delta\|^2|X]} - Q_n(\theta)}{\sqrt{\mathrm{Var}_n[\|\Delta\|^2|\theta]}}. \tag{30}$$

Our next result links the asymptotic behavior of $\mathbb{P}_n\{\Delta \in B_n^p|\theta\}$ to that of $t_n(\theta)$.

Lemma 6. *Let $\{\Gamma_n\}_{n\geq 1}$ be a sequence of r.v. such that $\mathfrak{L}(\Gamma_n) \twoheadrightarrow \mathcal{N}(0;1)$ and $\{t_n\}_{n\geq 1} \subset \mathbb{R}$. If $\underline{t} := \liminf t_n$ and $\bar{t} := \limsup t_n$ then it holds that*

$$\liminf \mathbb{P}\{\Gamma_n < t_n\} = \Phi(\underline{t}), \; \limsup \mathbb{P}\{\Gamma_n < t_n\} = \Phi(\bar{t});$$

In particular, if $\lim_n t_n = t \in [-\infty, \infty]$ then $\lim_n \mathbb{P}\{\Gamma_n < t_n\} = \Phi(t)$.

By Lemma 6, the asymptotic behavior of $t_n(\theta)$ leads to relevant conclusions on the asymptotic frequentist probability coverage of the credible ℓ^2-balls B_n^p. In the reminder of this section we analyze the asymptotic behavior of $t_n(\theta)$ in (30) for various levels of smoothness of θ, as well as the π-a.s. behavior.

The case $\delta \leq 0$
If $\delta \leq 0$, then for any $\theta \in \Theta_\delta$ the Gaussian approximation in (21) holds true, cf. Corollary 1. In addition, by Lemma 4 and Theorem 3, for such θ we have $\mathrm{Var}_n[\|\Delta\|^2|\theta] \sim \mathrm{Var}_n[\|\Delta\|^2|X] \sim n^{-\frac{1+4\alpha}{1+2\alpha}}$. Also, since $Q_n(\theta) = \mathbb{E}\left[\|\Delta\|^2|\theta\right] - M_n$,

$$\lim_{n\to\infty} n^{\frac{2\alpha}{1+2\alpha}} Q_n(\theta) = A_\pi^{\frac{1}{1+2\alpha}}\left(K_{1+2\alpha,2}^0 - K_{1+2\alpha,1}^0\right).$$

Cf. (18), for $p = 1$, $K_{1+2\alpha,2}^0 - K_{1+2\alpha,1}^0 = -K_{1+2\alpha,2}^{1+2\alpha} < 0$, hence $t_n(\theta) \to \infty$, for any $\theta \in \Theta_\delta$. Consequently, $\lim_n \mathbb{P}_n\{\Delta \in B_n^p|\theta\} = 1 > p$, by Lemma 6, i.e., the Bayesian credible ℓ^2-balls B_n^p have good frequentist probability coverage.

The π-a.s. behavior
In this case $\mathrm{Var}_n[\|\Delta\|^2|\theta] \sim \mathrm{Var}_n[\|\Delta\|^2|X] \sim \mathrm{Var}[Q_n(\vartheta)] \sim n^{-\frac{1+4\alpha}{1+2\alpha}}$, for almost all θ's drawn from π, cf. Theorem 3 (ii) and (iii). It follows that

$$\liminf t_n(\theta) = -\infty, \; \limsup t_n(\theta) = \infty,$$

for almost all θ's drawn from π, hence, cf. Lemma 6 we conclude that

$$\liminf \mathbb{P}_n\{\Delta \in B_n^p|\theta\} = 0, \; \limsup \mathbb{P}_n\{\Delta \in B_n^p|\theta\} = 1, \; \pi - a.s.$$

A similar result is obtained in [2], in a slightly more general framework.

The case $\delta > 0$

As δ grows larger than 0, Θ_δ accommodates more $\mathcal{B}_\omega$'s with $\alpha - \delta < \omega \le \alpha$; see Lemma 3, and for any $\theta \in \mathcal{B}_\omega$ the Gaussian approximation in (21) holds true, by virtue of Corollary 1. For $\omega \in (\alpha - \delta, \alpha)$, based on the estimates in (24),

$$\lim_{n \to \infty} n^{\frac{2\omega}{1+2\alpha}} Q_n(\theta) = L_\theta A_\pi^{-\frac{2\omega}{1+2\alpha}} K_{1+2\alpha,2}^{1+4\alpha-2\omega} > 0,$$

for all $\theta \in \mathcal{B}_\omega$, provided that $\theta_k^2 \approx L_\theta k^{-(1+2\omega)}$. Also, for $\theta \in \mathcal{B}_\omega$ it holds that

$$\sqrt{\mathrm{Var}_n[\|\Delta\|^2|X]} \ll \sqrt{\mathrm{Var}_n[\|\Delta\|^2|\theta]} \sim n^{-\frac{1/2+\alpha+\omega}{1+2\alpha}} \ll Q_n(\theta).$$

Therefore, if $\theta \in \mathcal{B}_\omega$, for some $\omega \in (\alpha - \delta, \alpha)$, then $t_n(\theta)$ defined in (30) converges to $-\infty$ and one concludes by Lemma 6 that $\lim_n \mathbb{P}_n\{\Delta \in B_n^p|\theta\} = 0$. In words, if $\delta > 0$ then there exist many θ's in Θ_δ (in fact, a dense subset) such that the Bayesian credible ℓ^2-balls B_n^p have asymptotically null frequentist coverage probability. On the other hand, if $\omega = \alpha$, taking in (18) $\varpi = 1 + 2\alpha$, $\eta = 4$ and $p = 1$, we obtain from (28) and (29)

$$\lim_{n \to \infty} n^{\frac{2\alpha}{1+2\alpha}} Q_n(\theta) = A_\pi^{-\frac{2\omega}{1+2\alpha}} \left[\frac{L_\theta}{A_\pi} - 1\right] K_{1+2\alpha,2}^{1+2\alpha}.$$

Since in this case we have

$$\mathrm{Var}_n[\|\Delta\|^2|\theta] \sim \mathrm{Var}_n[\|\Delta\|^2|X] \sim n^{-\frac{1+4\alpha}{1+2\alpha}},$$

it follows that for $L_\theta > A_\pi$ we have $t_n(\theta) \to -\infty$, hence the same phenomenon as for $\omega < \alpha$ occurs. If $L_\theta < A_\pi$, on the other hand, then $t_n(\theta) \to \infty$, hence by Lemma 6 $\lim_n \mathbb{P}_n\{\Delta \in B_n^p|\theta\} = 1$, so the Bayesian credible ℓ^2-balls B_n^p have good frequentist probability coverage. Finally, if $L_\theta = A_\pi$ then the limit

$$\lim_{n \to \infty} \frac{Q_n(\theta)}{\sqrt{\mathrm{Var}_n[\|\Delta\|^2|\theta]}},$$

can take any value in $[-\infty, \infty]$, depending on how fast the sequence $(\theta_k/\tau_k)^2$ converges to 1; in the special case $\theta_k^2 = \tau_k^2$ for all but finitely-many k's, the limit is null, hence, using (18) for $p = 2$ and the estimates in (29), we obtain

$$\lim_{n \to \infty} t_n(\theta) = \kappa_p \lim_{n \to \infty} \sqrt{\frac{\mathrm{Var}_n[\|\Delta\|^2|X]}{\mathrm{Var}_n[\|\Delta\|^2|\theta]}} = \kappa_p \sqrt{\frac{K_{1+2\alpha,2}^0}{K_{1+2\alpha,4}^0 + 2K_{1+2\alpha,4}^{1+2\alpha}}} > \kappa_p.$$

Therefore, by Lemma 6, $\lim_n \mathbb{P}_n\{\Delta \in B_n^p|\theta\} > p$ in this case, hence the Bayesian credible ℓ^2-balls B_n^p have again good frequentist probability coverage, provided that the prior π approximates well enough the true parameter θ. One concludes that, by considering parameter sets Θ_δ with $\delta > 0$, virtually anything is possible in terms of asymptotic frequentist probability coverage of the credible sets B_n^p.

3.3. Conclusions and remarks

In [3] a probabilistic analysis of the BvM statement for $\|\Delta\|^2$ w.r.t. the prior π was performed and the answer was negative, the main reason being that the variance of the frequentist distribution is asymptotically smaller than the variance of the Bayesian one, π-a.s. Here we have performed a rather analytic investigation, assuming that the true parameter belongs to some Sobolev subspaces $\Theta_\delta \subset \ell^2$, for $\delta \le \alpha$, hoping that such a BvM statement would hold for some of these parameter sets. While the choice $\delta > 0$, which in this context coincide with $\pi(\Theta_\delta) > 0$, is already ruled out by the results in [3], the choice $\delta \le 0$, which corresponds to $\pi(\Theta_\delta) = 0$, does not lead to a positive answer, either. Essentially, a quasi-similar behavior (to the π-a.s. one) for the ratio of the two variances was observed, for all $\theta \in \Theta_\delta$, for $\delta \le 0$, which led to the conclusion that neither analytic nor probabilistic BvM statements for $\|\Delta\|^2$ hold for this model.

Nevertheless, the good news is that if θ is assumed to belong to Θ_δ, with $\delta \le 0$, then the Bayesian credible ℓ^2-balls have good frequentist probability coverage, hence one can use them to derive frequentist confidence regions for the true parameter θ. Of particular interest is the space Θ_0 which appears to be the largest space on the Hilbert scale $\{\Theta_\delta\}_{\delta \le \alpha}$ having null prior probability and also the largest Θ_δ on which the positive result stated above remains valid. Another interesting property of the space Θ_0 is that the Bayes estimator $\hat{\vartheta}$ computed according to the prior π, achieves the optimal minimax rate if $\theta \in \Theta_0$; see [14]. We complement this result by showing that, in this setup, the Bayesian credible (centered) ℓ^2-balls can be employed as frequentist confidence regions for θ.

When $0 < \delta \le \alpha$, the asymptotic behavior of $\mathfrak{L}_n(\|\Delta\|^2|\theta)$ seems to be rather irregular for $\theta \in \Theta_\delta$. In fact, as δ grows larger than 0, more and more $\mathcal{B}_\omega$'s with $\omega < \alpha$ (over-smoothing) will lie inside Θ_δ, contributing with slower and slower rates. More specifically, there will be θ's in Θ_δ for which $\|\Delta\|$ converges to 0 at rate $n^{-\frac{\omega}{1+2\alpha}}$, for any $\omega \in (\alpha - \delta, \alpha)$, each set of such θ's (which includes $\mathcal{B}_\omega$) being dense in both Θ_δ and ℓ^2. In particular, if $\delta = \alpha$, i.e., $\Theta_\delta = \ell^2$, then the Bayes estimator $\hat{\vartheta}$ may converge to θ at arbitrarily slow rates since any $\mathcal{B}_\omega$, with $\omega > 0$, lies in ℓ^2. This shows that a result such as Lemma 4, establishing a constant convergence rate for $\hat{\vartheta}$, provided that $\theta \in \Theta_\delta$, may not hold for $\delta > 0$. However, one can establish without much effort that for any $\theta \in \ell^2$ it holds that

$$\liminf n^{\frac{2\alpha}{1+2\alpha}} \mathbb{E}\left[\|\Delta\|^2|\theta\right] \ge A_\pi^{\frac{1}{1+2\alpha}} K_{1+2\alpha,2}^0,$$

thus obtaining an upper-bound for the convergence rates. We conclude that, although consistent for any $\theta \in \ell^2$, the Bayes estimator $\hat{\vartheta}$ may converge to θ at arbitrarily slow rates when $\theta \in \ell^2 \setminus \Theta_0$ whereas for $\theta \in \Theta_0$ the convergence rates are the fastest possible. Regarding the frequentist probability coverage of Bayesian credible balls B_n^p, for most of the aforementioned θ's (in a topological sense) it holds that $\lim_n \mathbb{P}_n\{\Delta \in B_n^p|\theta\} = 0$. In addition, for any $p \in [0,1]$ one can find a $\theta \in \Theta_\delta$ such that $\lim_n \mathbb{P}_n\{\Delta \in B_n^p|\theta\} = p$. Finally, for most of the θ's in Θ_δ (in a probabilistic sense) it holds that $\liminf \mathbb{P}_n\{\Delta \in B_n^p|\theta\} = 0$ and $\limsup \mathbb{P}_n\{\Delta \in B_n^p|\theta\} = 1$. This completes the picture of the irregularity of the asymptotic behavior of $\mathfrak{L}_n(\|\Delta\|^2|\theta)$ on parameter sets of prior probability 1.

4. Appendix

The expressions of M_n, $Q_n(\theta)$ and $Z_n(\theta, \varepsilon)$ appearing in (19) are:

$$M_n := \sum_{k=1}^{\infty} \frac{\sigma_n^2 \tau_k^2}{\sigma_n^2 + \tau_k^2}, \; Q_n(\theta) := \sum_{k=1}^{\infty} \frac{\sigma_n^4 \tau_k^2}{(\sigma_n^2 + \tau_k^2)^2} \left(\xi_k^2(\theta) - 1 \right),$$

where $\xi_k(\theta) := \theta_k / \tau_k$ i.i.d. $\mathcal{N}(0; 1)$ variables relative to π. Furthermore, we have

$$Z_n(\theta, \varepsilon) := \sum_{k=1}^{\infty} \frac{\sigma_n^2 \tau_k^4}{(\sigma_n^2 + \tau_k^2)^2} (\varepsilon_k^2 - 1) + \sum_{k=1}^{\infty} \frac{2\sigma_n^3 \tau_k^3}{(\sigma_n^2 + \tau_k^2)^2} \xi_k(\theta) \varepsilon_k.$$

Since ε_k and $\varepsilon_k^2 - 1$ are uncorrelated variables and $\mathrm{Var}[\varepsilon_k^2 - 1] = 2$, we obtain

$$\mathrm{Var}[Z_n(\theta, \varepsilon)] = \sum_{k=1}^{\infty} \frac{2\sigma_n^4 \tau_k^8}{(\sigma_n^2 + \tau_k^2)^4} + \sum_{k=1}^{\infty} \frac{4\sigma_n^6 \tau_k^6}{(\sigma_n^2 + \tau_k^2)^4} \xi_k^2(\theta).$$

The sampling distribution of $\|\Delta\|^2$ has mean and variance given by

$$\mathbb{E}_n \left[\|\Delta\|^2 | \theta \right] = M_n + Q_n(\theta) = \sum_{k=1}^{\infty} \frac{\sigma_n^2 \tau_k^4}{(\sigma_n^2 + \tau_k^2)^2} + \sum_{k=1}^{\infty} \frac{\sigma_n^4 \theta_k^2}{(\sigma_n^2 + \tau_k^2)^2}, \tag{31}$$

$$\mathrm{Var} \left[\|\Delta\|^2 | \theta \right] = \mathrm{Var}[Z_n(\theta, \varepsilon)] = \sum_{k=1}^{\infty} \frac{2\sigma_n^4 \tau_k^8}{(\sigma_n^2 + \tau_k^2)^4} + \sum_{k=1}^{\infty} \frac{4\sigma_n^6 \tau_k^4 \theta_k^2}{(\sigma_n^2 + \tau_k^2)^4}, \tag{32}$$

respectively. The next results can be used to investigate the asymptotic behavior of the above expressions. Lemma 7 shows convergence to 0 when $(\tau_k), (\theta_k) \in \ell^2$ while Lemma 8 gives convergence rates for given asymptotics of (τ_k) and (θ_k).

Lemma 7. *Let $\{\sigma_n\}_{n \geq 1}, \{\tau_k\}_{k \geq 1}$ be positive numbers s.t. $\sigma_n \to 0$ and let $\eta > 0$. Then for any $w := (w_k) \in \ell^1$ it holds that*

$$\lim_{n \to \infty} \sum_{k \geq 1} \frac{\sigma_n^{2\eta} w_k}{(\sigma_n^2 + \tau_k^2)^\eta} \to 0.$$

Lemma 8. *Let $\varpi, \eta > 1$ and $\lambda \geq 0$ be s.t. $1 + \lambda < \varpi \eta$ and $f(t) := (1 + t^\varpi)^{-\eta} t^\lambda$, for $t \geq 0$. Then, it holds that*

$$\lim_{h \to 0} \sum_{k=1}^{\infty} h \, f(kh) = \int_0^\infty f(t) dt. \tag{33}$$

In addition, if $\zeta_n \to \infty$, $u_k \approx k^\lambda$ and $v_k \approx k^\varpi$ it follows that

(i) *For any positive integer $k_0 \geq 1$ we have*

$$\lim_{n \to \infty} \zeta_n^{\eta - (1+\lambda)/\varpi} \sum_{k=k_0}^{\infty} \frac{u_k}{(\zeta_n + v_k)^\eta} = \lim_{n \to \infty} \zeta_n^{\eta - (1+\lambda)/\varpi} \sum_{k=k_0}^{\infty} \frac{k^\lambda}{(\zeta_n + k^\varpi)^\eta} = K_{\varpi, \eta}^\lambda.$$

(ii) *In addition, there exist some positive constant $l > 0$ s.t.*

$$\lim_{n \to \infty} \zeta_n^{\eta - (\lambda/\varpi)} \max_{k \geq 1} \frac{u_k}{(\zeta_n + v_k)^\eta} = l.$$

5. Proofs of the results

Proof of Lemma 1. Let $\psi \in \mathbb{R}^\infty$. For the ease of writing, we define

$$\forall n, k \geq 1: \ \phi_{nk} := \frac{\tau_k \psi_k}{\sqrt{\sigma_n^2 + \tau_k^2}}, \ \varphi_{nk} := \frac{\tau_k^2 \psi_k}{\sigma_n^2 + \tau_k^2}, \ \beta_{nk}^\theta := \frac{\sigma_n \theta_k \psi_k}{\sigma_n^2 + \tau_k^2},$$

and $\phi_n := (\phi_{nk})_k$, $\varphi_n := (\varphi_{nk})_k$ and $\beta_n^\theta := (\beta_{nk}^\theta)_k$. Then it is easy to see that our hypothesis is equivalent to $\psi \in \ell^1$, (i) is equivalent to $\phi_n \in \ell^1$, for all n, while (ii) is equivalent to $\varphi_n \in \ell^1$ and $\beta_n^\theta \in \ell^1$, for all n, for almost all θ's drawn from π. Since $|\varphi_{nk}| < |\phi_{nk}| < |\psi_k|$, for all n, k, entails $\|\varphi_n\|_{\ell^1} \leq \|\phi_n\|_{\ell^1} \leq \|\psi\|_{\ell^1}$, for all n, hence $\phi_n \in \ell^1$ and $\varphi_n \in \ell^1$, for all n. This proves (i) and the variance condition in (ii). Moreover, under the prior π, $\vartheta_k = \tau_k \xi_k$, for any $k \geq 1$, with $\{\xi_k\}_{k \geq 1}$ being i.i.d. standard Gaussian variables. Therefore, we obtain

$$\|\beta_n^\vartheta\|_{\ell^1} = \sum_{k \geq 1} \frac{\sigma_n \tau_k |\psi_k \xi_k|}{\sigma_n^2 + \tau_k^2} \leq \sum_{k \geq 1} \frac{\sigma_n}{\sqrt{\sigma_n^2 + \tau_k^2}} |\psi_k \xi_k| \leq \|(\psi_k \xi_k)\|_{\ell^1}. \tag{34}$$

The last expression in the above display is a random variable with finite mean, hence finite almost surely. That is, $(\psi_k \xi_k) \in \ell^1$ almost surely; in particular, (34) shows that $\beta_n^\theta \in \ell^1$, for all n, for π-almost all θ's, which concludes the proof of (ii). To prove (iii), we assume that $\psi \neq \mathbf{0}$ (otherwise the statement is trivial) and note that by re-scaling the distributions under consideration, we have

$$\mathfrak{L}_n \left(\psi \left(\sigma_n^{-1} \Delta \right) | X \right) = \mathcal{N}(0; \|\phi_n\|^2), \ \mathfrak{L}_n \left(\psi \left(\sigma_n^{-1} \Delta \right) | \theta \right) = \mathcal{N} \left(-\sum_{k \geq 1} \beta_{nk}^\theta; \|\varphi_n\|^2 \right)$$

and $\gamma \circ \psi^{-1} = \mathcal{N}(0; \|\psi\|^2)$. Therefore, since $|\sum_{k \geq 1} \beta_{nk}^\theta| \leq \|\beta_n^\theta\|_{\ell^1}$, it suffices to show that $\|\beta_n^\theta\|_{\ell^1} \ll \|\phi_n\| \approx \|\varphi_n\| \approx \|\psi\|$, π-a.s. First we prove the $\approx$ relations. Indeed, applying Lemma 7 for $\eta = 1$ and $w_k = |\psi_k|$, we obtain

$$\|\varphi_n - \psi\|_{\ell^1} = \sum_{k \geq 1} \frac{\sigma_n^2 |\psi_k|}{\sigma_n^2 + \tau_k^2} \to 0.$$

Therefore, $\varphi_n \to \psi$ in ℓ^1 (hence also in ℓ^2) and it follows that $\|\varphi_n\| \to \|\psi\|$. Moreover, $\|\varphi_n\| \leq \|\phi_n\| \leq \|\psi\|$ proves that $\|\phi_n\| \to \|\psi\|$, which proves the claim (recall that $\|\psi\| > 0$). Finally, to prove that $\|\beta_n^\vartheta\|_{\ell^1} \to 0$ a.s., we use again Lemma 7, with $\eta = 1/2$ and $w_k = |\psi_k \xi_k|$ (recall that $(\psi_k \xi_k) \in \ell^1$ a.s.) to prove that the first majorant in (34) converges to 0 almost surely; this proves (iii). Finally, ψ being linear, we have

$$\|\mathfrak{L}_n(\sigma_n^{-1}\psi(\Delta)|X) - \mathfrak{L}_n(\sigma_n^{-1}\psi(\Delta)|\theta)\|_H = \|\mathfrak{L}_n(\psi(\sigma_n^{-1}\Delta)|X) - \mathfrak{L}_n(\psi(\sigma_n^{-1}\Delta)|\theta)\|_H.$$

Since the Hellinger distance is invariant to re-scaling, and the r.h.s. in the last display converges to 0, π-a.s., this concludes the proof of (iv). $\qquad\square$

Proof of Theorem 1. Recall that a sequence of Gaussian measures $\mathcal{N}(b_n; S_n)$, for $n \geq 1$, converges weakly on a Hilbert space if b_n converges to some $b \in \mathbb{H}$ and S_n converges in trace-class norm to some (trace-class) operator S on $\mathbb{H}$, in which case the limit is $\mathcal{N}(b; S)$. For $w_k = \lambda_k^2$ and $\eta = 1$ in Lemma 7, we obtain

$$\|T_n - S\|_1 = \sum_{k \geq 1} \left| \frac{\lambda_k^2 \tau_k^2}{\sigma_n^2 + \tau_k^2} - \lambda_k^2 \right| = \sum_{k \geq 1} \frac{\sigma_n^2 \lambda_k^2}{\sigma_n^2 + \tau_k^2} \to 0;$$

In the same vein, we obtain $\|S_n - S\|_1 \to 0$ as follows:

$$\|S_n - S\|_1 = \sum_{k \geq 1} \left| \frac{\lambda_k^2 \tau_k^4}{(\sigma_n^2 + \tau_k^2)^2} - \lambda_k^2 \right| = \sum_{k \geq 1} \frac{\sigma_n^2 \lambda_k^2 (\sigma_n^2 + 2\tau_k^2)}{(\sigma_n^2 + \tau_k^2)^2} \leq 2 \sum_{k \geq 1} \frac{\sigma_n^2 \lambda_k^2}{\sigma_n^2 + \tau_k^2} \to 0.$$

This proves that $\mathcal{N}(\mathbf{0}; T_n)$ converges weakly to γ and S_n converges in trace-class norm to S. To conclude now that $\mathcal{N}(b_n^\theta; S_n)$ converges weakly to γ for π-almost all θ's, we need to show that $\|b_n^\vartheta\|_\mathbb{H} \to 0$, almost surely. Again, since under π we have $\vartheta_k = \tau_k \xi_k$, with $\{\xi_k\}_{k \geq 1}$ standard i.i.d. Gaussian variables, taking $\eta = 1$ and $w = (\lambda_k^2 \xi_k^2)$ (note that $w \in \ell^1$ with probability 1) in Lemma 7 yields

$$\|b_n^\vartheta\|_\mathbb{H}^2 = \sum_{k \geq 1} \frac{\sigma_n^2 \lambda_k^2 \tau_k^2 \xi_k^2}{(\sigma_n^2 + \tau_k^2)^2} \leq \sum_{k \geq 1} \frac{\sigma_n^2 (\lambda_k^2 \xi_k^2)}{\sigma_n^2 + \tau_k^2} \to 0;$$

For the last statement, note that both probabilities in (12) approach $\gamma(B)$. $\square$

Proof of Lemma 2. (i)→(ii) Let $\mathcal{A}_n$ denote the Hellinger affinity of $\mathfrak{L}(\Delta|\theta)$ and γ. The statement is now equivalent to $\mathcal{A}_n \to 1$ or $\log(1/\mathcal{A}_n) \to 0$. Using the fact that both measures are independent products of independent Gaussian distributions and the multiplicative property of the Hellinger affinity, we obtain

$$\mathcal{A}_n = \prod_{k \geq 1} \sqrt{\frac{2\tau_k^2(\sigma_n^2 + \tau_k^2)}{\sigma_n^4 + 2\sigma_n^2 \tau_k^2 + 2\tau_k^4}} \exp \left[-\frac{\sigma_n^2 \theta_k^2}{4(\sigma_n^4 + 2\sigma_n^2 \tau_k^2 + 2\tau_k^4)} \right].$$

Now note that $\mathcal{A}_n \leq 1$, hence $\log(1/\mathcal{A}_n) \geq 0$. Therefore, we have

$$\log(1/\mathcal{A}_n) \leq \frac{\sigma_n^4}{4} \left\| \left(\frac{1}{\tau_k^2} \right) \right\|^2 + \frac{\sigma_n^2}{8} \left\| \left(\frac{\theta_k}{\tau_k^2} \right) \right\|^2;$$

we used the fact that $\log(1 + x) \leq x$, for all $x > 0$. Letting $n \to \infty$ proves that $\|\mathfrak{L}(\sigma_n^{-1}\Delta|\theta) - \gamma\|_H \to 0$. A similar argument leads to $\|\mathfrak{L}(\sigma_n^{-1}\Delta|X) - \gamma\|_H \to 0$.

(ii)→(iii) Follows by the scaling invariance property of the Hellinger distance.

(iii)→(i) As already noted, the statement in (iii) implies conditions (15) and (16). To prove now that (16) implies $(1/\tau_k^2) \in \ell^2$, note first that $\mathcal{S}_n(\pi) < \infty$ entails $\tau_k^2 \to \infty$; in particular, the sequence $1/\tau_k^2$ is bounded by some constant $M > 0$. Next use the inequality $\|(1/\tau_k^2)\|^2 \leq (M + 2/\sigma_n^2)^2 \cdot \mathcal{S}_n(\pi)$ to deduce that $(1/\tau_k^2) \in \ell^2$. This concludes the proof. $\square$

Proof of Lemma 3. Let $\omega > \alpha - \delta$. The inclusion $\mathcal{B}_\omega \subset \Theta_\delta$ is immediate. Let now $\theta \in \Theta_\delta$ and choose some arbitrary $\epsilon > 0$. Define $\beta := (\beta_k)$ as follows

$$\beta_k := \left\{ \begin{array}{ll} \theta_k, & k \leq n_\epsilon \\ k^{-(1/2+\omega)}, & k > n_\epsilon, \end{array} \right.$$

with n_ϵ chosen such that $\sum_{k>n_\epsilon} k^{2(\alpha-\delta)}\theta_k^2 < \epsilon$ and $\sum_{k>n_\epsilon} k^{-[1+2(\omega+\delta-\alpha)]} < \epsilon$; this is possible since $\theta \in \Theta_\delta$ and $\omega + \delta - \alpha > 0$, hence both expressions are remainder terms from convergent series. Obviously, $\beta \in \mathcal{B}_\omega$ and we have

$$\|\beta - \theta\|_\delta^2 = \sum_{k>n_\epsilon} k^{2(\alpha-\delta)}\left[k^{-(1/2+\omega)} - \theta_k\right]^2 \leq 2\sum_{k>n_\epsilon} k^{2(\alpha-\delta)}\left[k^{-(1+2\omega)} + \theta_k^2\right] < 4\epsilon.$$

Finally, if $\omega \leq \alpha - \delta$ then $\theta \in \mathcal{B}_\omega$ yields $k^{2(\alpha-\delta)}\theta_k^2 \sim k^{-[1+2(\omega+\delta-\alpha)]}$, the last sequence defining a divergent series. Therefore, one concludes that $\mathcal{B}_\omega \cap \Theta_\delta = \emptyset$, in this case. This concludes the proof. $\square$

Proof of Lemma 4. Recall first that $\mathbb{E}_n[\|\Delta\|^2|\theta] = T_n(\pi) + U_n(\pi,\theta)$. For the choices $\lambda = 0$, $\varpi = 1+2\alpha$ and $\eta = 2$ in Lemma 8 (Appendix), one obtains

$$\lim_{n\to\infty} n^{\frac{2\alpha}{1+2\alpha}} T_n(\pi) = A_\pi^{\frac{1}{1+2\alpha}} K_{1+2\alpha,2}^0.$$

Letting $E_n(\theta) := n^{\frac{2\alpha}{1+2\alpha}} U_n(\pi,\theta)$, the first statement is equivalent to $E_n(\theta) \to 0$, for any $\theta \in \Theta_\delta$. To prove the last claim, note that for any $\omega > \alpha - \delta \geq \alpha$ we have $\mathcal{B}_\omega \subset \Theta_\delta$ and the statement holds true for any $\theta \in \mathcal{B}_\omega$. Fix now $\omega > \alpha - \delta$ and note that for arbitrary $\theta, \beta \in \Theta_\delta$ it holds that

$$|E_n(\beta) - E_n(\theta)| \leq \sum_{k=1}^\infty \frac{n^{\frac{2\alpha}{1+2\alpha}} k^{2+4\alpha}}{(A_\pi n + k^{1+2\alpha})^2}|\beta_k^2 - \theta_k^2| = \sum_{k=1}^\infty \nu_{nk} k^{2(\alpha-\delta)}|\beta_k - \theta_k| \cdot |\beta_k + \theta_k|,$$

where, for simplicity, for $n, k \geq 1$ we denote

$$\nu_{nk} := \frac{n^{\frac{2\alpha}{1+2\alpha}} k^{2(1+\alpha+\delta)}}{(A_\pi n + k^{1+2\alpha})^2}.$$

Set now $\nu_n^* := \|(\nu_{nk})_k\|_{\ell^\infty}$. Using Hölder Inequality twice, according to the scheme $\|x \cdot y \cdot z\|_{\ell^1} \leq \|x\|_{\ell^\infty}\|y \cdot z\|_{\ell^1} \leq \|x\|_{\ell^\infty}\|y\|_{\ell^2}\|z\|_{\ell^2}$, yields

$$|E_n(\beta) - E_n(\theta)| \leq \nu_n^* \|\beta + \theta\|_\delta \|\beta - \theta\|_\delta, \tag{35}$$

where (take $\lambda = 2(1+\alpha+\delta)$, $\varpi = 1+2\alpha$, $\eta = 2$ in Lemma 8 (ii))

$$\nu_n^* := n^{\frac{2\alpha}{1+2\alpha}}\left\|\left(\frac{k^{2(1+\alpha+\delta)}}{(A_\pi n + k^{1+2\alpha})^2}\right)_k\right\|_{\ell^\infty} \sim n^{\frac{2\alpha}{1+2\alpha}} \cdot n^{\frac{2(1+\alpha+\delta)}{1+2\alpha}-2} = n^{\frac{2\delta}{1+2\alpha}}. \tag{36}$$

By hypothesis, $\delta \leq 0$, hence

$$\forall n \geq 1: \ |E_n(\beta) - E_n(\theta)| \leq C\|\beta + \theta\|_\delta\|\beta - \theta\|_\delta, \tag{37}$$

for some constant $C > 0$, independent of n. Fix now some $\theta \in \Theta_\delta$ and let $\epsilon > 0$ be arbitrary. Choose $\eta > 0$ s.t. $(2\|\theta\|_\delta + \eta)C\eta < \epsilon/2$ and let $\beta \in \mathcal{B}_\omega$ be such that $\|\beta - \theta\|_\delta < \eta$. For such β, the inequality $\|\beta + \theta\|_\delta \leq 2\|\theta\|_\delta + \|\beta - \theta\|_\delta$ implies

$$|E_n(\beta) - E_n(\theta)| \leq C\|\beta + \theta\|_\delta \|\beta - \theta\|_\delta < \epsilon/2.$$

Finally, take $n_0 \geq 1$ large enough to guarantee that $E_n(\beta) < \epsilon/2$, $n \geq n_0$. Then,

$$\forall n \geq n_0 : \ 0 \leq E_n(\theta) \leq E_n(\beta) + |E_n(\beta) - E_n(\theta)| < \epsilon.$$

Since $\epsilon > 0$ was arbitrary, letting $n \to \infty$ proves the claim.

For the second statement, let $F_n(\theta) := n^{\frac{1+4\alpha}{1+2\alpha}} W_n(\pi, \theta)$, note that

$$\mathrm{Var}_n[\|\Delta\|^2|\theta] = V_n(\pi) + W_n(\pi, \theta), \ V_n(\pi) \approx 2A_\pi^{\frac{1}{1+2\alpha}} K_{1+2\alpha,4}^0 n^{-\frac{1+4\alpha}{1+2\alpha}},$$

and prove that $F_n(\theta) \to 0$, for all $\theta \in \Theta_\delta$, using a similar reasoning. Now

$$\nu_n^* = n^{\frac{1+4\alpha}{1+2\alpha}} \left\| \left(\frac{4A_\pi^2 n k^{2(1+\alpha+\delta)}}{(A_\pi n + k^{1+2\alpha})^2} \right)_k \right\|_{\ell^\infty} \sim n^{\frac{1+4\alpha}{1+2\alpha}} \cdot n^{\frac{2(1+\alpha+\delta)}{1+2\alpha} - 3} = n^{\frac{2\delta}{1+2\alpha}},$$

instead of (36), however. The last statement is straightforward. $\qquad\square$

Proof of Lemma 5. We check the Lindeberg condition. For simplicity, let

$$\forall n, k \geq 1 : \ \lambda_{nk} := \frac{A_\pi^2 n}{(A_\pi n + k^{1+2\alpha})^2}, \ \mu_{nk} := -\frac{2A_\pi \sqrt{n} k^{1+2\alpha}\theta_k}{(A_\pi n + k^{1+2\alpha})^2}.$$

Then we have $Z_n(\theta, \varepsilon) = \sum_{k \geq 1} \lambda_{nk}(\varepsilon_k^2 - 1) + \sum_{k \geq 1} \mu_{nk}\varepsilon_k$, for $n \geq 1$, hence

$$S_n^2 := \mathrm{Var}[Z_n(\theta, \varepsilon)] = 2\sum_{k \geq 1} \lambda_{nk}^2 + \sum_{k \geq 1} \mu_{nk}^2 = V_n(\pi) + W_n(\pi, \theta). \qquad (38)$$

Letting $\lambda_n^* := \|(\lambda_{nk})_k\|_{\ell^\infty}$, $\mu_n^* := \|(\mu_{nk})_k\|_{\ell^\infty}$, the Lindeberg condition becomes

$$\forall t > 0 : \ \lim_{n \to \infty} \frac{1}{S_n^2} \sum_{k \geq 1} \int_{\{|X_{nk}| > 2tS_n\}} X_{nk}^2 d\mathbb{P} = 0, \qquad (39)$$

where $X_{nk} = \lambda_{nk}(\varepsilon_k^2 - 1) + \mu_{nk}\varepsilon_k$. The inequality $(u + v)^2 \leq 2(u^2 + v^2)$ shows that the expression under the limit in (39) is bounded by

$$2\sum_{k \geq 1} \frac{\lambda_{nk}^2}{S_n^2} \int_{\{|X_{nk}| > 2tS_n\}} (\varepsilon_k^2 - 1)^2 d\mathbb{P} + 2\sum_{k \geq 1} \frac{\mu_{nk}^2}{S_n^2} \int_{\{|X_{nk}| > 2tS_n\}} \varepsilon_k^2 d\mathbb{P}. \qquad (40)$$

By (38) we conclude that for each $n \geq 1$ it holds that

$$0 < \max \left\{ \frac{\sum_{k \geq 1} \lambda_{nk}^2}{S_n^2}, \frac{\sum_{k \geq 1} \mu_{nk}^2}{S_n^2} \right\} < 1,$$

hence for the Lindeberg condition in (39) to be verified it suffices that the integrals in (40) to converge to 0, uniformly in k. Now note that $|u + v| > 2a$ entails (at least) one of the conditions $|u| > a$ or $|v| > a$. Consequently, we have

$$
\begin{aligned}
\{|X_{nk}| > 2tS_n\} &\subset \left\{|\varepsilon_k^2 - 1| > \frac{tS_n}{|\lambda_{nk}|}\right\} \cup \left\{|\varepsilon_k| > \frac{tS_n}{|\mu_{nk}|}\right\} \\
&\subset \left\{|\varepsilon_k^2 - 1| > \frac{tS_n}{\lambda_n^*}\right\} \cup \left\{|\varepsilon_k| > \frac{tS_n}{\mu_n^*}\right\}.
\end{aligned}
$$

Hence, a sufficient condition for the integrals in (40) to converge to 0, uniformly in k, is that $\lambda_n^*/S_n \to 0$ and $\mu_n^*/S_n \to 0$. Indeed, this follows from the fact that, for a $\mathcal{N}(0; 1)$ variable Z, if $t_n \to \infty$ then, denoting $C(Z, t) := \{|Z| > t\}$ and $D(Z, t) := \{|Z^2 - 1| > t\}$, for $t > 0$, it holds that

$$
\lim_{n \to \infty} \int_{C(Z, t_n)} Z^2 d\mathbb{P} = \lim_{n \to \infty} \int_{D(Z, t_n)} Z^2 d\mathbb{P} = 0
$$

and

$$
\lim_{n \to \infty} \int_{C(Z, t_n)} (Z^2 - 1)^2 d\mathbb{P} = \lim_{n \to \infty} \int_{D(Z, t_n)} (Z^2 - 1)^2 d\mathbb{P} = 0,
$$

all the integrals depending only on the distribution $\mathcal{N}(0; 1)$, but not on Z itself.

Finally, we note that $\lambda_n^*/S_n \to 0$. Indeed, by taking $\lambda = 0$, $\varpi = 1 + 2\alpha$ and $\eta = 4$ in Lemma 8 we have the following estimates (as $n \to \infty$) and implication

$$
(\lambda_n^*)^2 \sim n^{-2}, \sum_{k \geq 1} \lambda_{nk}^2 \sim n^{-\frac{1+4\alpha}{1+2\alpha}} \implies \left(\frac{\lambda_n^*}{S_n}\right)^2 \leq \frac{(\lambda_n^*)^2}{2 \sum_{k \geq 1} \lambda_{nk}^2} \sim n^{-\frac{1}{1+2\alpha}} \to 0,
$$

hence the Lindeberg condition is fulfilled for any θ satisfying $\mu_n^*/S_n \to 0$; recall that (unlike λ_n^*) μ_n^* depends on θ. Now the fact that

$$
\frac{\max\limits_{k \geq 1} \frac{nk^{2+4\alpha}\theta_k^2}{(A_\pi n + k^{1+2\alpha})^4}}{\operatorname{Var}_n\left[\|\Delta\|^2|\theta\right]} = \frac{1}{4A_\pi^2} \left(\frac{\mu_n^*}{S_n}\right)^2,
$$

proves the Lindeberg condition, hence the claim, for the desired θ's. $\qquad\square$

Proof of Corollary 1. For $\theta \in \mathcal{B}_\omega$, with $\omega \leq \alpha$, taking $\lambda = 1 + 4\alpha - 2\omega$, $\varpi = 1 + 2\alpha$ and $\eta = 4$ in Lemma 8 yields

$$
\max_{k \geq 1} \frac{nk^{2+4\alpha}\theta_k^2}{(A_\pi n + k^{1+2\alpha})^4} \sim n^{-\frac{1}{1+2\alpha}} \sum_{k \geq 1} \frac{nk^{2+4\alpha}\theta_k^2}{(A_\pi n + k^{1+2\alpha})^4} = n^{-\frac{1}{1+2\alpha}} \frac{W_n(\pi, \theta)}{4A_\pi^2},
$$

while by the estimates in (25) and (29) we have $\operatorname{Var}_n\left[\|\Delta\|^2|X\right] \sim W_n(\pi, \theta)$. This proves that the conditions in Lemma 5 are satisfied for such θ's.

If $\theta \in \Theta_\delta$, for $\delta \leq 0$, the statement follows from $W_n(\pi, \theta) \ll \operatorname{Var}_n[\|\Delta\|^2|X]$; see Lemma 4. Finally, for $\theta \in \mathcal{B}_\omega$, for $\omega > \alpha$, note the inclusion $\mathcal{B}_\omega \subset \Theta_0$. $\qquad\square$

Proof of Lemma 6. Let $\Phi_n(t) := \mathbb{P}\{\Gamma_n < t\}$. Since $\mathfrak{L}(\Gamma_n) \twoheadrightarrow \mathcal{N}(0;1)$ and the limit has continuous distribution function Φ, it follows that $|\Phi_n(t) - \Phi(t)| \to 0$, uniformly in $t \in \mathbb{R}$; see [13]. The proof follows using the monotony of Φ_n and Φ and continuity of Φ, i.e., $\liminf \Phi(t_n) = \Phi(\underline{t})$ and $\limsup \Phi(t_n) = \Phi(\overline{t})$. $\quad\square$

Proof of Lemma 7. Let $\epsilon > 0$ and choose $k_\epsilon \geq 1$ such that $\sum_{k>k_\epsilon} |w_k| < \epsilon$; that is, k_ϵ only depends on the given sequence w. Therefore, we obtain

$$\sum_{k\geq 1} \frac{\sigma_n^{2\eta}|w_k|}{(\sigma_n^2 + \tau_k^2)^\eta} < \epsilon + \sum_{k=1}^{k_\epsilon} \frac{\sigma_n^{2\eta}|w_k|}{(\sigma_n^2 + \tau_k^2)^\eta}.$$

Now each term in the finite sum in the r.h.s. converges to 0, for $n \to \infty$, hence

$$\limsup \sum_{k\geq 1} \frac{\sigma_n^{2\eta}|w_k|}{(\sigma_n^2 + \tau_k^2)^\eta} \leq \epsilon.$$

Finally, since ϵ was arbitrary, this concludes the proof. Alternatively, note that the statement is equivalent with $\psi_n \to \mathbf{0}$ in the weak-* topology of ℓ^∞, where the sequence $\psi_n = (\psi_{nk})_k$, defined by $\psi_{nk} := \sigma_n^{2\eta}(\sigma_n^2 + \tau_k^2)^{-\eta}$, is norm-bounded by 1 in $\ell^\infty = (\ell^1)^*$. The Banach-Alaoglu Theorem can be used to conclude. $\quad\square$

Proof of Lemma 8. Let $T > 0$ be some arbitrarily large number. Then,

$$\lim_{h\to 0} \sum_{k=1}^{[T/h]} h\,f(kh) = \int_0^T f(t)dt. \tag{41}$$

On the other hand, integrability of f entails $\int_T^\infty f(t)dt \to 0$ for $T \to \infty$. Let $\gamma := \varpi\eta - \lambda$; by hypothesis, we have $\gamma > 1$. Since $f(t) < t^{-\gamma}$, it follows that

$$\sum_{k>[T/h]} h\,f(kh) < h \sum_{k>[T/h]} \left(\frac{1}{kh}\right)^\gamma = T^{1-\gamma}\left(\frac{T}{h}\right)^{\gamma-1} \sum_{k>[T/h]} \left(\frac{1}{k}\right)^\gamma \leq KT^{1-\gamma},$$

where $K > 0$ is some constant depending only on γ such that, for large $x > 0$, $x^{\gamma-1}\sum_{k>x} k^{-\gamma} < K$. Therefore, for large enough (T/h) one concludes that

$$\sum_{k=1}^\infty h\,f(kh) = \sum_{k=1}^{[T/h]} h\,f(kh) + \sum_{k>[T/h]} h\,f(kh) \leq \sum_{k=1}^{[T/h]} h\,f(kh) + KT^{1-\gamma}.$$

Letting $h \to 0$ in the above inequality and taking (41) into account leads to

$$\int_0^T f(t)dt \leq \liminf \sum_{k=1}^\infty h\,f(kh) \leq \limsup \sum_{k=1}^\infty h\,f(kh) \leq KT^{1-\gamma} + \int_0^T f(t)dt.$$

Finally, letting $T \to \infty$ in the above display concludes (33).

To prove now statement (i), let $x_n := \zeta_n^{\eta-(1+\lambda)/\varpi}$ and note that $\zeta_n \to \infty$ implies that for any fixed $k \geq 1$ we have

$$\lim_{n\to\infty} \frac{x_n u_k}{(\zeta_n + v_k)^\eta} = \lim_{n\to\infty} \frac{x_n k^\lambda}{(\zeta_n + k^\varpi)^\eta} = 0.$$

Therefore, for any $k_0 \geq 1$, it holds that

$$\lim_{n\to\infty} \sum_{k=k_0}^{\infty} \frac{x_n u_k}{(\zeta_n + v_k)^\eta} = \lim_{n\to\infty} \sum_{k=1}^{\infty} \frac{x_n u_k}{(\zeta_n + v_k)^\eta},$$

i.e., if any of the above limits exists the other one exists as well and they are necessary equal. In particular, this shows that the asymptotic behavior of the infinite sum is dictated by the asymptotic behavior of the sequences u_k and v_k.

Now let $h := \zeta_n^{-1/\varpi} \to 0$, i.e., $x_n k^\lambda (\zeta_n + k^\varpi)^{-\eta} = h\, f(kh)$. By the first part,

$$\lim_{n\to\infty} x_n \sum_{k=1}^{\infty} \frac{k^\lambda}{(\zeta_n + k^\varpi)^\eta} = K_{\varpi,\eta}^\lambda$$

and the same limit holds if summation starts from any fixed k_0, which proves the second equality in (i). To conclude, choose some arbitrary (small) $\epsilon > 0$ and take some large enough $k_0 \geq 1$ to ensure that

$$\forall k \geq k_0 : \ k^{-\lambda} u_k, k^{-\varpi} v_k \in (1-\epsilon, 1+\epsilon). \tag{42}$$

Then, for all $n \geq 1$, it holds the following double inequality

$$\sum_{k=k_0}^{\infty} \frac{(1-\epsilon)k^\lambda}{(\zeta_n + (1+\epsilon)k^\varpi)^\eta} \leq \sum_{k=k_0}^{\infty} \frac{u_k}{(\zeta_n + v_k)^\eta} \leq \sum_{k=k_0}^{\infty} \frac{(1+\epsilon)k^\lambda}{(\zeta_n + (1-\epsilon)k^\varpi)^\eta}. \tag{43}$$

We claim now that

$$\lim_{n\to\infty} x_n \sum_{k=k_0}^{\infty} \frac{(1+\epsilon)k^\lambda}{(\zeta_n + (1-\epsilon)k^\varpi)^\eta} = \frac{(1+\epsilon)}{(1-\epsilon)^{(1+\lambda)/\varpi}} K_{\varpi,\eta}^\lambda. \tag{44}$$

Indeed, if we take $h := \left(\frac{1-\varepsilon}{\zeta_n}\right)^{1/\varpi}$ then some elementary algebra shows that

$$(1-\epsilon)^{-(1+\lambda)/\varpi} h\, f(kh) = \frac{x_n k^\lambda}{(\zeta_n + (1-\epsilon)k^\varpi)^\eta}$$

and the claim in (44) follows by the first step. In the same vein, we obtain

$$\lim_{n\to\infty} x_n \sum_{k=k_0}^{\infty} \frac{(1-\epsilon)k^\lambda}{(\zeta_n + (1+\epsilon)k^\varpi)^\eta} = \frac{(1-\epsilon)}{(1+\epsilon)^{(1+\lambda)/\varpi}} K_{\varpi,\eta}^\lambda.$$

Going back to (43) we see that letting $n \to \infty$ yields

$$
\begin{aligned}
\frac{(1-\epsilon)}{(1+\epsilon)^{(1+\lambda)/\varpi}} K_{\varpi,\eta}^{\lambda} &\leq \liminf x_n \sum_{k=k_0}^{\infty} \frac{u_k}{(\zeta_n + v_k)^{\eta}} \\
&\leq \limsup x_n \sum_{k=k_0}^{\infty} \frac{u_k}{(\zeta_n + v_k)^{\eta}} \leq \frac{(1+\epsilon)}{(1-\epsilon)^{(1+\lambda)/\varpi}} K_{\varpi,\eta}^{\lambda}.
\end{aligned}
$$

Finally, recall that $\lim_n \sum_{k=1}^{k_0} \frac{x_n u_k}{(\zeta_n + v_k)^{\eta}} = 0$, hence one can replace in the last display k_0 by 1, to obtain a similar inequality for the sum started at $k = 1$ (which, unlike k_0, is independent of ε); since $\epsilon > 0$ was arbitrary it follows that

$$
\lim_{n \to \infty} x_n \sum_{k=1}^{\infty} \frac{u_k}{(\zeta_n + v_k)^{\eta}} = K_{\varpi,\eta}^{\lambda};
$$

again, the summation may start with any $k_0 \geq 1$ with no changes in the limit.

For (ii) note that $f : (0, \infty) \to \mathbb{R}$ has a unique maximum t^* which satisfies

$$
f'(t^*) = \frac{\lambda t^{\lambda-1}(1+t^{\varpi})^{\eta} - \varpi\eta(1+t^{\varpi})^{\eta-1}t^{\lambda+\varpi-1}}{(1+t^{\varpi})^{2\eta}} = 0 \implies t^* := \left(\frac{\lambda}{\varpi\eta - \lambda}\right)^{1/\varpi}.
$$

Therefore, the maximal value of f on $(0, \infty)$ is given by

$$
l := f(t^*) = \frac{\lambda^{(\lambda/\varpi)}(\varpi\eta - \lambda)^{\eta-(\lambda/\varpi)}}{(\varpi\eta)^{\eta}}.
$$

First, let $u_k = k^{\lambda}$ and $v_k = k^{\varpi}$ and note that in this case we have

$$
\frac{u_k}{(\zeta_n + v_k)^{\eta}} = \frac{k^{\lambda}}{(\zeta_n + k^{\varpi})^{\eta}} = \zeta_n^{(\lambda/\varpi)-\eta} f(\zeta_n^{-1/\varpi}k).
$$

Since f is increasing for $t < t^*$ and is decreasing for $t > t^*$, it follows that

$$
\max_{k \geq 1} \frac{k^{\lambda}}{(\zeta_n + k^{\varpi})^{\eta}} = \zeta_n^{(\lambda/\varpi)-\eta} \max_{k \geq 1} f(\zeta_n^{-1/\varpi}k)
$$

and the maximum in the r.h.s. above is attained for either $k_n := [\zeta_n^{1/\varpi}t^*]$ or $k_n + 1$. Now the fact that $\zeta_n \to \infty$ and continuity of f leads to

$$
\lim_{n \to \infty} \zeta_n^{-1/\varpi} k_n = t^* \implies \max_{k \geq 1} f(\zeta_n^{-1/\varpi}k) \to l = f(t^*),
$$

which concludes the proof in this case. Finally, for generic sequences (u_k) and (v_k) satisfying the asymptotic conditions in the hypothesis one can use a similar reasoning as in (i). This is possible because $k_n \to \infty$ for $n \to \infty$. Hence, if one chooses any small $\epsilon > 0$ and k_0 to satisfy (42) then for large n, such that $k_n > k_0$, the maximum is not affected if we ignore the first $k_0 - 1$ terms of the sequence. The proof is now complete. $\qquad\square$

Acknowledgment

The author is very grateful to Professor Harry van Zanten for raising some of the questions considered in this paper, as well as for taking the time to read the material and providing valuable feedback. I am also grateful to the referee for several comments/suggestions which led to the improvement of the presentation.

References

[1] CASTILLO, I. *A semiparametric Bernstein-von Mises Theorem for Gaussian process priors*, Probability Theory and Related Fields (to appear)

[2] COX, D. *An analysis of Bayesian inference for nonparametric regression*, Annals of Statistics 21(2), pp. 903–923, 1993. MR1232525

[3] FREEDMAN, D. *On the Bernstein-von Mises Theorem with infinite dimensional parameters*, Annals of Statistics 27(4), pp. 1119–1140, 1999. MR1740119

[4] ÎTO, K. *The topological support of Gauss measure on Hilbert space*, Nagoya Mathematical Journal, 38, pp. 181–183, 1970. MR0261332

[5] KIM, Y. and LEE, J. *A Bernstein-von Mises Theorem in the nonparametric right-censoring model*, Annals of Statistics 32(4), pp. 1492–1512, 2004. MR2089131

[6] KIM, Y. *The Bernstein-von Mises Theorem for the proportional hazard model*, Annals of Statistics 34(4), pp. 1678–1700, 2006. MR2283713

[7] KUELBS, J. *Gaussian measures on a Banach space*, Journal of Functional Analysis 5, pp. 354–367, 1970. MR0260010

[8] KUO, H-H. *Gaussian Measures in Banach Spaces*, Lecture Notes in Math., no. 463, Springer-Verlag Heidelberg, 1975. MR0461643

[9] LEHMANN, E.L. and CASELLA, G. *Theory of Point Estimation – 2nd Edition*, Springer Texts in Statistics, Springer Verlag NY, 1998. MR1639875

[10] PARTHASARATHY, K.R. *Probability Measures on Metric Spaces*, Academic Press New York – London, 1967. MR0226684

[11] RIVOIRARD, V. and ROUSSEAU, J. *Bernstein-von Mises Theorem for linear functionals of the density*, Annals of Statistics (to appear)

[12] SHEN, X. *Asymptotic normality of semiparametric and nonparametric posterior distributions*, J. of American Stat. Association 97, pp. 222–235, 2002. MR1947282

[13] VAN DER VAART, A. W. *Asymptotic Statistics*, Cambridge University Press, 1998. MR1652247

[14] ZHAO, L. H. *Bayesian aspects of some nonparametric problems*, Annals of Statistics 28(2), pp. 532–552, 2000. MR1790008