Noname manuscript No.
(will be inserted by the editor)

A semiparametric Bernstein - von Mises theorem for Gaussian process priors

Ismaël Castillo

Received: date / Accepted: date

Abstract This paper is a contribution to the Bayesian theory of semiparametric estimation. We are interested in the so-called Bernstein-von Mises theorem, in a semiparametric framework where the unknown quantity is (θ, f), with θ the parameter of interest and f an infinite-dimensional nuisance parameter. Two theorems are established, one in the case with no loss of information and one in the information loss case with Gaussian process priors. The general theory is applied to three specific models: the estimation of the center of symmetry of a symmetric function in Gaussian white noise, a time-discrete functional data analysis model and Cox's proportional hazards model. In all cases, the range of application of the theorems is investigated by using a family of Gaussian priors parametrized by a continuous parameter.

Keywords Bayesian non and semiparametrics, Bernstein-von Mises Theorems, Gaussian process priors, Estimation of the center of symmetry, Cox's proportional hazards model

Mathematics Subject Classification (2000) 62G05 · 62G20

In a Bayesian estimation framework, the Bernstein-von Mises phenomenon (hereafter abbreviated BVM) is concerned with the fact that posterior distributions often asymptotically look like normal distributions. In the parametric i.i.d. case, according to Le Cam [20], the phenomenon was discovered by Laplace and further studied by Bernstein, von Mises, Le Cam and many others since then. For a statement in the parametric case under fairly minimal assumptions, we refer to [27].

A natural question is then to know if this phenomenon occurs in more general settings. In non- and semiparametric models, the question of consistency of the posterior distribution is already very interesting, see for instance [23, 1]. As pointed out

I. Castillo
CNRS, LPMA Paris & Vrije Universiteit Amsterdam
175, rue du Chevaleret
75013 Paris, France
E-mail: ismael.castillo@upmc.fr

in [8], some care must be taken when dealing with infinite-dimensional prior distributions since even consistency might fail for innocent looking priors, and this is still the case when looking at asymptotic normality of the posterior, see [9]. However, positive BVM results are known in particular nonparametric settings, see [10, 17].

From the practical point of view, Bayesian methods are broadly used, and the development of computer capacities has enabled the use of computational methods like Markov chain Monte Carlo algorithms to simulate from the posterior. Also, the use of infinite-dimensional prior distributions is becoming more and more common, for instance Gaussian process priors are widely used in machine learning, see [22]. However, often, in infinite-dimensional contexts, little is known about how to prove that posterior distributions have the desired behavior. In semiparametric frameworks, the BVM theorem constitutes a very suitable theoretical guarantee. The importance of this result lies in the fact that it gives the explicit asymptotic form of the relevant part of the posterior distribution, from which one easily derives in particular the consistency of the Bayes procedure, together with the fact that Bayesian confidence intervals have the desired asymptotic coverage probability.

A few semiparametric BVM theorems are known in specific models with well-chosen families of priors, often exploiting the fact that the prior has certain conjugacy properties, see for instance [16], where the author establishes the BVM theorem in the proportional hazards model for Lévy-type priors. To the best of our knowledge, the only work considering a general semiparametric framework is [24]. Here we would like to acknowledge the importance of this work, which contains many interesting ideas, for the present paper. However, as we explain in more details in the sequel, some of the conditions in [24] are rather implicit and hard to check in practice. Our goal in this paper is to obtain simple interpretable conditions which give more insight into the problem and are usable in practice in a variety of situations. Although we do not consider the most general semiparametric framework possible, our results already enable to treat very different practical examples, especially cases where the prior is not conjugate.

Two BVM theorems are proved, depending on whether or not there is a loss of information in the model, that is whether the efficient information coincides with the information in the associated parametric model, see [27], Section 25.4. In the case of loss of information, we restrict our investigations to Gaussian process priors for the nonparametric component of the model. Our assumptions naturally extend in their spirit the ones for proving the parametric BVM theorem, that is the concentration of the posterior in neighborhoods of the true and a proper "shape" of the model locally around the true parameter. As an application, we obtain new BVM theorems in three different models, the problem of estimating the translation parameter of a symmetric signal in Gaussian noise, see [15, 7], a time-discrete functional data analysis model and the proportional hazards model introduced by Cox in [6].

Two main tools are used which both correspond to recent advances in the Bayesian nonparametric theory. The first one is the theory of nonparametric posterior concentration, as presented in [13] and [25] in i.i.d. settings, see also [12] for several non-i.i.d. extensions. The second tool is the use of Gaussian process priors, which provide a flexible framework in which to analyze our results and for which concentration properties of the posterior have been recently investigated by [29] and [5].

The article is organized as follows. In Section 1, the semiparametric framework we study is described and the general theorems are stated. Applications and results in special models are given in Section 2. Proofs of the main theorems are in Section 3, while Section 4 is devoted to checking the assumptions in the particular models. Section 5 gathers useful properties of Gaussian processes used throughout the paper.

Let us denote by $N(\mu,\sigma^2)$ the 1-dimensional Gaussian distribution with mean μ and variance σ^2 and by Φ the cumulative distribution function associated to a $N(0,1)$ variable. For a distribution P and a measurable set B, we denote by $P(B)$ the mass that P puts on B. Let $K(f,g)=\int f\log(f/g)d\mu$ stand for the Kullback-Leibler divergence between the two non-negative densities f and g relative to a measure μ. Let us also define $V(f,g)=\int f|\log(f/g)-K(f,g)|^2d\mu$. The ε-covering number of a set Θ for a semi-metric d, denoted by $N(\varepsilon,\Theta,d)$, is the minimal number of d-balls of radius ε needed to cover Θ. Let $\mathscr{C}^\beta[0,1]$ denote the Hölder space of order β of continuous functions on $[0,1]$ that have $\underline{\beta}$ continuous derivatives for $\underline{\beta}$ the largest integer strictly smaller than β with the $\underline{\beta}$th derivative being Lipschitz-continuous of order $\beta-\underline{\beta}$. The notation $\lesssim$ is used for 'smaller than or equal to a universal constant times' and $\triangleq$ means 'equal by definition to'.

1 Statement

1.1 Bayesian semiparametric estimators

Let us consider a sequence of statistical models $(\mathscr{X}^{(n)},\mathscr{G}^{(n)},\mathbf{P}_\eta^{(n)},\ \eta\in\mathscr{E})$, with observations $X^{(n)}$, where $\mathscr{E}$ is a parameter set of the form $\Theta\times\mathscr{F}$ with Θ an interval of $\mathbb{R}$ (the results of this paper extend without much effort to $\mathbb{R}^k$) and $\mathscr{F}$ a subset of a separable Banach space. The "true" value of the parameter is denoted by $\eta_0=(\theta_0,f_0)$ and is assumed to be an interior point of $\mathscr{E}$. We assume that the measures $\mathbf{P}_\eta^{(n)}$ admit densities $p_\eta^{(n)}$ with respect to a σ-finite measure $\mu^{(n)}$ on $(\mathscr{X}^{(n)},\mathscr{G}^{(n)})$. The log-likelihood $\log p_\eta^{(n)}$ is denoted by $\ell_n(\eta)$ and $\Lambda_n(\eta)=\ell_n(\eta)-\ell_n(\eta_0)$. The space $\mathscr{E}=\Theta\times\mathscr{F}$ is equipped with a product σ-field $\mathscr{T}\otimes\mathscr{B}$ and we assume that $(x,\eta)\to p_\eta^{(n)}(x)$ is jointly measurable.

Prior, condition (P). Let us put a probability measure Π, called *prior*, on the pair (θ,f), of the form $\Pi=\pi_\theta\otimes\pi_f$. For π_θ we choose any probability measure on Θ having a density λ with respect to Lebesgue's measure on Θ, with λ *positive* and *continuous* at the point θ_0.

Bayes formula. In the preceding framework, Bayes' theorem asserts that the posterior distribution is given explicitly for any C in $\mathscr{T}\otimes\mathscr{B}$ by

$$\Pi(C|X^{(n)})=\frac{\int_C p_\eta^{(n)}(X^{(n)})d\Pi(\eta)}{\int p_\eta^{(n)}(X^{(n)})d\Pi(\eta)}. \tag{1}$$

For any $\varepsilon > 0$, let us define a Kullback-Leibler-type neighborhood of η_0 in $\mathscr{E}$ as

$$B_{KL,n}(\eta_0,\varepsilon) = \{\eta \in \mathscr{E} : K(P^{(n)}_{\eta_0}, P^{(n)}_{\eta}) \le n\varepsilon^2,\ \ V(P^{(n)}_{\eta_0}, P^{(n)}_{\eta}) \le n\varepsilon^2\}.$$

Let $\Pi^{\theta=\theta_0}(\cdot|X^{(n)})$ be the posterior distribution in the model where θ is known to be equal to θ_0 and one takes π_f as prior on f. By Bayes' theorem, for any $B \in \mathscr{B}$,

$$\Pi^{\theta=\theta_0}(B|X^{(n)}) = \frac{\int_B p^{(n)}_{\theta_0,f}(X^{(n)})d\pi_f(f)}{\int p^{(n)}_{\theta_0,f}(X^{(n)})d\pi_f(f)}.$$

As above let us also define a neighborhood of f_0 in $\mathscr{F}$ by

$$B^{\theta=\theta_0}_{KL,n}(f_0,\varepsilon) = \{f \in \mathscr{F} : K(P^{(n)}_{\eta_0}, P^{(n)}_{\theta_0,f}) \le n\varepsilon^2,\ \ V(P^{(n)}_{\eta_0}, P^{(n)}_{\theta_0,f}) \le n\varepsilon^2\}.$$

1.2 A specific semiparametric framework

A natural way to study efficiency in a semiparametric model is to study estimation along a maximal collection of 1-dimensional paths locally around the true parameter, as explained for instance in [27], Chap. 25, see also [21] where some tools are developed in non-i.i.d. situations. The likelihood ratios along the paths might then for instance be well approximated by the likelihood ratios in the case of a Gaussian shift experiment, which leads to the notion of local asymptotic normality (LAN).

In this paper we take a slightly different approach. Given a true $\eta_0 = (\theta_0, f_0)$ in $\mathscr{E}$, for any $\eta = (\theta, f)$ in $\mathscr{E}$ (possibly restricted to a subset of $\mathscr{E}$, possibly close enough to η_0), let us assume that the pair $(\theta - \theta_0, f - f_0)$ can be embedded in a product Hilbert space of the form $\mathscr{V}_{\eta_0} = \mathbb{R} \times \mathscr{G}_{\eta_0}$ equipped with an inner-product $\langle\, ,\, \rangle_L$ with associated norm $\|\cdot\|_L$. Locally around the true parameter, we shall compare the log-likelihood differences to a quadratic term plus a stochastic term. We set

$$R_n(\theta, f) = \Lambda_n(\theta, f) + n\|\theta - \theta_0, f - f_0\|^2_L/2 - \sqrt{n}W_n(\theta - \theta_0, f - f_0), \qquad (2)$$

where we use the following notation

- $\Lambda_n(\theta, f) = \ell_n(\theta, f) - \ell_n(\theta_0, f_0)$ is the difference of log-likelihoods between the points (θ, f) and (θ_0, f_0).
- $\{W_n(v),\ v \in \mathscr{V}_{\eta_0}\}$ is a collection of random variables measurable with respect to the observations $X^{(n)}$ and satisfying the following properties.
  - ⋄ For any $v_1, \ldots, v_d$ in $\mathscr{V}_{\eta_0}$, the d-tuple $(W_n(v_1), \ldots, W_n(v_d))$ converges in distribution to the d-dimensional centered Gaussian distribution with covariance structure given by the matrix $(\langle v_i, v_j \rangle_L)_{1 \le i,j \le d}$.
  - ⋄ The map $v \to W_n(v)$ is linear.

We will further assume that one has a form of uniform control of the R_n's over (sieved) shrinking neighborhoods of the true η_0, see the assumptions **(N)** and **(N')** of the theorems below.

The inner-product and the stochastic term introduced above are often identified from LAN-type expansions. For instance, one might be in a situation where the model is LAN with linear paths (see e.g. [21]) in that for each $v = (s,g) \in \mathscr{V}_{\eta_0}$, as $n \to +\infty$,

$$\Lambda_n(\theta_0 + s/\sqrt{n}, f_0 + g/\sqrt{n}) = -\|s,g\|_L^2/2 + W_n(s,g) + o_{P_{\eta_0}^{(n)}}(1),$$

where $\|\cdot\|_L$, W_n and $\mathscr{V}_{\eta_0}$ are as above. To define the notions of information and efficiency in our model (let us recall that it is not necessarily i.i.d.), we assume for simplicity that the considered model is LAN with linear paths, which falls in the framework considered in [21], so we can borrow from that paper the definitions and their implications for efficiency. In fact, such an assumption is essentially weaker than the *uniform* type of control on $R_n(\theta, f)$ required below, see Section 1.7. All models considered in Section 2 admit such a LAN expansion, at least for a well-chosen parametrization of the model.

Semi-parametric structure. Here we define the notions of least favorable direction and efficient Fisher information following [21]. Let $\overline{\mathscr{F}}$ be the closure in $\mathscr{V}_{\eta_0}$ of the linear span of all elements of the type $(0, f - f_0)$, where f belongs to $\mathscr{F}$. Let us define the element $(0, \gamma(\cdot)) \in \overline{\mathscr{F}}$ as the orthogonal projection of the vector $(1,0)$ onto the closed subspace $\overline{\mathscr{F}}$. The element γ is called *least favorable direction*. For any $(s,g) \in \mathscr{V}_{\eta_0}$, one has the following decomposition

$$\|s,g\|_L^2 = (\|1,0\|_L^2 - \|0,\gamma\|_L^2)s^2 + \|0, g + s\gamma\|_L^2.$$

The coefficient of s^2 is called *efficient Fisher information* and is denoted by $\tilde{I}_{\eta_0} = \|1,0\|_L^2 - \|0,\gamma\|_L^2$. If γ is zero, we say there is *no loss of information* and denote the information simply by I_{η_0}. Note also that since $\|\cdot\|_L$ is a norm, $I_{\eta_0} = \|1,0\|_L^2$ is always nonzero. If $\tilde{I}_{\eta_0}$ itself is nonzero, let us also denote

$$\Delta_{n,\eta_0} = \tilde{I}_{\eta_0}^{-1} W_n(1, -\gamma).$$

In particular, an estimator $\hat{\theta}_n$ of θ_0 is asymptotically linear and efficient if it satisfies $\sqrt{n}(\hat{\theta}_n - \theta_0) = \Delta_{n,\eta_0} + o_{P_{\eta_0}^{(n)}}(1)$.

Local parameters. Throughout the paper, we use the following shorthand notation $h = \sqrt{n}(\theta - \theta_0)$ and $a = \sqrt{n}(f - f_0)$.

1.3 The case without loss of information

If there is no information loss, it holds $\|h,a\|_L^2 = \|h,0\|_L^2 + \|0,a\|_L^2$ and $I_{\eta_0} = \tilde{I}_{\eta_0} = \|1,0\|_L^2$ and $\Delta_{n,\eta_0} = W_n(1,0)/\|1,0\|_L^2$.

Concentration (C). Let $\varepsilon_n > 0$ be a sequence such that $\varepsilon_n \to 0$ and $n\varepsilon_n^2 \to +\infty$. The statistical model and the prior Π satisfy condition **(C)** with rate ε_n if there exists a sequence of measurable sets $\mathscr{F}_n$ in $\mathscr{F}$ such that

$$\Pi\left(\{\eta \in \Theta \times \mathscr{F}_n, \quad \|\eta - \eta_0\|_L \leq \varepsilon_n\} \mid X^{(n)}\right) \to 1,$$

$$\Pi^{\theta=\theta_0}\left(\{f \in \mathscr{F}_n, \quad \|0, f - f_0\|_L \leq \varepsilon_n/\sqrt{2}\} \mid X^{(n)}\right) \to 1,$$

as $n \to +\infty$, in $\mathbf{P}^{(n)}_{\eta_0}$-probability.

Local shape (N). Let R_n be defined by (2) and let ε_n and $\mathscr{F}_n$ be as in **(C)**. Let us denote $V_n = \{(\theta, f) \in \Theta \times \mathscr{F}_n, \quad \|\theta - \theta_0, f - f_0\|_L \leq \varepsilon_n\}$. The model satisfies **(N)** with rate ε_n over the sieve $\mathscr{F}_n$ if

$$\sup_{(\theta,f)\in V_n} \frac{|R_n(\theta, f) - R_n(\theta_0, f)|}{1 + n(\theta - \theta_0)^2} = o_{P^{(n)}_{\eta_0}}(1).$$

Theorem 1 *Let us assume that the prior Π on (θ, f) satisfies* **(P)** *and that the model and prior verify conditions* **(C)**, **(N)**. *Suppose there is no information loss, then it holds*

$$\sup_B \left| \Pi\left(B \times \mathscr{F} \mid X^{(n)}\right) - N\left(\theta_0 + \frac{1}{\sqrt{n}}\Delta_{n,\eta_0}, \frac{1}{n} I_{\eta_0}^{-1}\right)(B) \right| \to 0,$$

as $n \to +\infty$, in $\mathbf{P}^{(n)}_{\eta_0}$-probability, where the supremum is taken over all measurable sets B in Θ. In words, the total variation distance between the marginal in θ of the posterior distribution and a Gaussian distribution centered at $\theta_0 + \frac{1}{\sqrt{n}}\Delta_{n,\eta_0}$, of variance $\frac{1}{n} I_{\eta_0}^{-1}$, converges to zero, in $\mathbf{P}^{(n)}_{\eta_0}$-probability.

Condition **(C)** means that the posterior concentrates at ε_n-rate around the true η_0 in terms of $\|\cdot\|_L$. Sufficient conditions for **(C)** are discussed in Section 1.6. Condition **(N)** controls how much the likelihood ratio differs locally from the one of a Gaussian experiment and is studied in Section 1.7. Note that assumptions **(P)** -the parametric part of the prior must charge θ_0-, **(N)** -which is about the shape of the model- and **(C)** -which enables us to localize in a neighborhood of the true η_0- compare in their spirit to the assumptions for the *parametric* Bernstein-von Mises theorem as stated in [20][§7.3, Prop.1]. One can also note that Theorem 1 actually yields a result in the particular case where f is known. In this case, the Theorem implies that if posterior concentration occurs at rate $\varepsilon_n = M_n n^{-1/2}$ for some $M_n \to \infty$ (for instance $M_n = \log n$ say) and if the uniform LAN property **(N)** in θ holds in that neighborhood of size $M_n n^{-1/2}$ then the (parametric) BVM theorem holds.

1.4 The case with information loss

Here we shall restrict our investigations to Gaussian priors for π_f. Roughly, this enables us to change variables in the expression $\|0, f - f_0 + (\theta - \theta_0)\gamma\|_L^2$ by setting $g = f + (\theta - \theta_0)\gamma$ (or $g = f + (\theta - \theta_0)\gamma_n$, where γ_n is close enough to γ).

Suppose π_f is the distribution associated to a centered Gaussian process taking its values almost surely in a separable Banach space $\mathbb{B}$. For an overview of the basic properties of these objects and applications to Bayesian nonparametrics, we refer to [30, 29]. Two examples of families of Gaussian priors are considered in Section 2, see (10)-(14).

Let $\mathbb{H}$ be the Reproducing Kernel Hilbert Space (RKHS) of the Gaussian process. For a zero-mean real-valued Gaussian stochastic process $(W_t,\ t\in T)$ for some index set T, this space is defined through the covariance function $K: T\times T\to\mathbb{R}$, $K(s,t) = E(W_sW_t)$ as the completion $\mathbb{H}$ of the set of all linear functions

$$t\to\sum_{i=1}^{k}\lambda_i K(s_i,t),\quad \lambda_1,\dots,\lambda_k\in\mathbb{R},\ s_1,\dots,s_k\in\mathbb{R},\ k\in\mathbb{N},$$

with respect to the norm $\|\cdot\|_{\mathbb{H}}$ induced by the inner product

$$\langle\sum_{i=1}^{k}\lambda_i K(s_i,t),\sum_{j=1}^{l}\mu_j K(t_j,t)\rangle_{\mathbb{H}} = \sum_{i=1}^{k}\sum_{j=1}^{l}\lambda_i\mu_j K(s_i,t_j).$$

If the Gaussian process is given as a Borel measurable map in a Banach space, $\mathbb{H}$ is generally defined in a more abstract way through the Pettis integral, but both definition coincide in general, see [30], Section 2, for a detailed discussion.

We shall assume that the space $\mathbb{H}$ is "large enough" so that the least favorable direction γ introduced above can be approximated by elements of $\mathbb{H}$. Suppose that there exists $\rho_n\to 0$ and a sequence γ_n of elements in $\mathbb{H}$ such that for all $n>0$, $\gamma_n-\gamma$ belongs to $\mathscr{G}_{\eta_0}$ and

$$\|\gamma_n\|_{\mathbb{H}}^2\le 2n\rho_n^2\quad\text{and}\quad \|0,\gamma_n-\gamma\|_L\le\rho_n. \tag{3}$$

Concentration (C'). The model verifies condition **(C')** with rate ε_n if there exists a sequence of measurable sets $\mathscr{F}_n$ in $\mathscr{F}$ such that, if $\mathscr{F}_n(\theta) = (\mathscr{F}_n + (\theta-\theta_0)\gamma_n)$,

$$\Pi\left(\{\eta\in\Theta\times\mathscr{F}_n,\quad \|\eta-\eta_0\|_L\le\varepsilon_n\}\mid X^{(n)}\right)\to 1,$$

$$\inf_{|\theta-\theta_0|\tilde{I}_{\eta_0}^{1/2}\le\varepsilon_n}\Pi^{\theta=\theta_0}\left(\{f\in\mathscr{F}_n(\theta),\quad \|0,f-f_0\|_L\le\varepsilon_n/2\}\mid X^{(n)}\right)\to 1,$$

as $n\to+\infty$, in $P_{\eta_0}^{(n)}$-probability. We also assume that for some $c,d>0$, it holds

$$\Pi(B_{KL,n}(\eta_0,d\varepsilon_n))\ge\exp(-cn\varepsilon_n^2)\quad\text{and}\quad \pi_f(B_{KL,n}^{\theta=\theta_0}(f_0,d\varepsilon_n))\ge\exp(-cn\varepsilon_n^2).$$

Local Shape (N'). Let $V_n = \{(\theta,f)\in\Theta\times\mathscr{F}_n,\quad \|\theta-\theta_0,f-f_0\|_L\le 2\varepsilon_n\}$. Assume that for any (θ,f) in V_n, the function $f-(\theta-\theta_0)\gamma_n$ belongs to $\mathscr{F}$ and that

$$\sup_{(\theta,f)\in V_n}\frac{|R_n(\theta,f)-R_n(\theta_0,f-(\theta-\theta_0)\gamma_n)|}{1+n(\theta-\theta_0)^2} = o_{P_{\eta_0}^{(n)}}(1).$$

Our last assumption is related to how well the least favorable direction is approximated by elements of $\mathbb{H}$. As $n\to+\infty$ suppose

$$\textbf{(E)}\qquad \sqrt{n}\varepsilon_n\rho_n = o(1)\quad\text{and}\quad W_n(0,\gamma-\gamma_n) = o_{P_{\eta_0}^{(n)}}(1).$$

Theorem 2 *Let us assume that the prior $\Pi = \pi_\theta \otimes \pi_f$ on (θ, f) satisfies* **(P)***, that π_f is a Gaussian prior, that $\tilde{I}_{\eta_0} > 0$ and that the least favorable direction γ can be approximated according to (3). Suppose that conditions* **(C')**, **(N')** *and* **(E)** *are satisfied. Then it holds*

$$\sup_B \left| \Pi\left(B \times \mathscr{F} \mid X^{(n)}\right) - N\left(\theta_0 + \frac{1}{\sqrt{n}} \Delta_{n,\eta_0}, \frac{1}{n} \tilde{I}_{\eta_0}^{-1}\right)(B)\right| \to 0,$$

as $n \to +\infty$, in $P_{\eta_0}^{(n)}$-probability, where the supremum is taken over all measurable sets B in Θ.

The assumptions are similar in nature to the ones of Theorem 1, with additional requirements about the least favorable direction γ and Gaussianity of π_f. These assumptions are further discussed in Sections 1.8 and 2.4. For the moment let us only note that γ might in fact lie in $\mathbb{H}$, in which case (3) and **(E)** are trivially satisfied.

1.5 Comments on centering and applications to confidence intervals

Centering. The Bernstein-von Mises theorem is sometimes stated with the target Gaussian distribution centered at the maximum likelihood estimator. Here the approach we consider is slightly more general. As we show in the next paragraph, the conclusions of Theorems 1 or 2 imply a statement with a centering at any arbitrary asymptotically linear and efficient estimator of θ. The centering is thus in a way the best possible one can expect. We note that it does not require to prove any property of the maximum likelihood estimator. This approach was first introduced by Le Cam, see [26], pp. 678-679 and [27], Chapter 10 for a discussion.

Let $\tilde{\theta}_n$ be any estimator such that $\sqrt{n}(\tilde{\theta}_n - \theta_0) = \Delta_{n,\eta_0} + o_{P_{\eta_0}}(1)$, as $n \to +\infty$, where as above we denote $\Delta_{n,\eta_0} = \tilde{I}_{\eta_0}^{-1} W_n(1, -\gamma)$. In other words, $\tilde{\theta}_n$ is asymptotically linear and efficient (the asymptotic variance of $\sqrt{n}(\tilde{\theta}_n - \theta_0)$ equals the information bound $\tilde{I}_{\eta_0}^{-1}$). Due to the invariance by location and scale changes of the total variation distance on $\mathbb{R}$ and the assumed property of $\tilde{\theta}_n$, one has

$$\sup_B \left| N\left(\theta_0 + \frac{1}{\sqrt{n}} \Delta_{n,\eta_0}, \frac{1}{n} \tilde{I}_{\eta_0}^{-1}\right)(B) - N\left(\tilde{\theta}_n, \frac{1}{n} \tilde{I}_{\eta_0}^{-1}\right)(B)\right| \to 0,$$

in probability. If the conclusion of Theorems 1 or 2 holds, the posterior is in the limit like $N(\theta_0 + \Delta_{n,\eta_0}/\sqrt{n}, \tilde{I}_{\eta_0}^{-1}/n)$ in the total variation sense, and thus one gets automatically the same result for the Gaussian target distribution replaced by $N(\tilde{\theta}_n, \tilde{I}_{\eta_0}^{-1}/n)$.

In particular, if the maximum likelihood estimator $\hat{\theta}^{MLE}$ is asymptotically linear and efficient, that is $\sqrt{n}(\hat{\theta}^{MLE} - \theta_0) = \Delta_{n,\eta_0} + o_{P_{\eta_0}}(1)$ then one deduces the form of the BVM theorem with the centering at $\hat{\theta}^{MLE}$. But as seen above, the result can in fact be stated with centering at any other (asymptotically linear and) efficient estimator.

Application to confidence intervals. Let us consider the random interval having as endpoints the 2.5% and 97.5% percentiles of the posterior marginal $\Pi(\cdot \times \mathscr{F} \mid X^{(n)})$. It is the interval $[A_n, B_n]$ such that

$$\Pi((-\infty, A_n) \times \mathscr{F} \mid X^{(n)}) = 0.025, \qquad \Pi((B_n, +\infty) \times \mathscr{F} \mid X^{(n)}) = 0.025.$$

Note that $[A_n, B_n]$ is accessible in practice as soon as simulation from the posterior marginal is feasible. Now, the conclusion of the BVM theorem implies that $[A_n, B_n]$ coincides asymptotically with the interval having as endpoints the same percentiles but for the distribution $N(\theta_0 + \Delta_{n,\eta_0}/\sqrt{n}, \tilde{I}_{\eta_0}^{-1}/n)$. Simple verifications reveal that the latter interval contains θ_0 with probability 95%. Hence $[A_n, B_n]$ is asymptotically a 95%-confidence interval in the frequentist sense. Moreover, it holds

$$[A_n, B_n] = \left[\tilde{\theta}_n + \frac{\Phi^{-1}(0.025)}{\sqrt{n}\tilde{I}_{\eta_0}^{1/2}} + o_P(n^{-1/2}), \tilde{\theta}_n + \frac{\Phi^{-1}(0.975)}{\sqrt{n}\tilde{I}_{\eta_0}^{1/2}} + o_P(n^{-1/2})\right],$$

for any asymptotically linear and efficient $\tilde{\theta}_n$. In particular, Bayes and frequentist credible regions asymptotically coincide. An advantage of the Bayes approach is that, to build $[A_n, B_n]$, estimation of $\tilde{I}_{\eta_0}$ is not required.

1.6 Tests and concentration: about conditions **(C)**, **(C')**

In this subsection, we give sufficient conditions for **(C)** and **(C')** following [12, 13]. Suppose that there exist two semi-metrics d_n and e_n (possibly depending on n) on $\mathscr{E}$ satisfying the following property. There exist universal constants $\xi > 0$ and $K > 0$ such that for every $\varepsilon > 0$, $n > 0$, for each $\eta_1 \in \mathscr{E}$ with $d_n(\eta_1, \eta_0) > \varepsilon$, there exists a test ϕ_n such that

$$\mathbf{P}_{\eta_0}^{(n)} \phi_n \leq e^{-Kn\varepsilon^2}, \qquad \sup_{\eta \in \mathscr{E},\ e_n(\eta, \eta_1) < \xi\varepsilon} \mathbf{P}_{\eta}^{(n)}(1 - \phi_n) \leq e^{-Kn\varepsilon^2}. \tag{4}$$

Let $\varepsilon_n \to 0$, $n\varepsilon_n^2 \to +\infty$, let $\mathscr{A}_n \subset \mathscr{E}$ be a sequence of measurable sets and $C_1, \ldots, C_4$ positive constants such that, as $n \to +\infty$,

$$\log N(C_1\varepsilon_n, \mathscr{A}_n, e_n) \leq C_2 n\varepsilon_n^2 \tag{5}$$

$$\Pi(\mathscr{E} \backslash \mathscr{A}_n) \leq \exp(-n\varepsilon_n^2(C_3 + 4C_4^2)), \tag{6}$$

$$\Pi\left(B_{KL,n}(\eta_0, C_4\varepsilon_n)\right) \geq \exp(-n\varepsilon_n^2 C_3) \tag{7}$$

Lemma 1 *Let us assume that (5), (6), (7) hold. Then for M large enough, the posterior* $\Pi(\{\eta \in \mathscr{E},\ d_n(\eta, \eta_0) \leq M\varepsilon_n\} \mid X^{(n)})$ *tends to 1 as* $n \to +\infty$*, in* $\mathbf{P}_{\eta_0}^{(n)}$*-probability.*

This lemma is a version of Theorem 1 in [12], with the constants C_i adding some extra flexibility. More refined versions of this result could also be used, see [12]. More generally, any approach providing posterior rates could be used. For instance, in the i.i.d. setting for Hellinger's distance, one can also check the set of conditions considered in [25] to obtain posterior rates.

Let us now briefly discuss the meaning of the previous conditions. Condition (7) is in terms of the prior only and means that there should be enough prior mass in balls of size equal to the desired rate of convergence ε_n around the true parameter η_0. Conditions of the same flavor, without the rate, appear already when proving consistency results, see for instance [23], and they enable to control the denominator of the Bayes ratio. Condition (5) requires the entropy over a *sieve* set $\mathscr{A}_n$ with respect to e_n to be controlled. This is a way of controlling the complexity of the model on a set which should capture most prior mass, as required by Condition (6). Note also that the latter condition can be checked independently using the next Lemma.

Lemma 2 *If $\varepsilon_n \to 0$ and $n\varepsilon_n^2 \to +\infty$, if $\mathscr{A}_n \subset \mathscr{E}$ is a measurable set such that*

$$\Pi(\mathscr{E}\backslash\mathscr{A}_n)/\Pi(B_{KL,n}(\eta_0,\varepsilon_n)) \leq \exp(-2n\varepsilon_n^2),$$

then $\Pi(\mathscr{A}_n|X^{(n)}) \to 1$ as $n \to +\infty$, in $\mathbf{P}_{\eta_0}^{(n)}$-probability.

Finally, with (4) we require that it is possible to test a point versus a ball using semi-distances d_n, e_n with exponential decrease of the error probabilities. This testing condition is particularly useful when dealing with non i.i.d. data and it has been checked in the literature for a variety of frameworks. For a detailed discussion and examples we refer the reader to [12] (see also Lemma 3 and Section 4.3). In fact, in some frameworks condition (4) will be automatically satisfied. In the case of i.i.d. data for instance, general results in [2, 19] on existence of tests between convex sets of probability measures imply that such tests always exist when $d_n = e_n$ is the total variation or Hellinger metric, see for instance [13][Section 7].

Lemma 1 returns a result in terms of the semi-distance d_n appearing in the testing condition (think, for instance, of Hellinger's distance in an i.i.d. framework). To obtain **(C)**, some link has to be made between the target metric $\|\cdot\|_L$ appearing in the LAN expansion and d_n. In some cases, those might be equivalent or at least there might exist a $D > 0$ such that for any sequence $r_n \to 0$ as $n \to +\infty$, for n large enough it holds

$$\{\eta \in \mathscr{E},\ d_n(\eta,\eta_0)^2 \leq r_n^2\} \subset \{\eta = (\theta, f),\ \|\theta-\theta_0, f-f_0\|_L^2 \leq Dr_n^2\}. \tag{8}$$

If (5), (6), (7) and (8) hold, then **(C)** holds. For instance, in the Gaussian white noise model considered in Section 2.1, one can take d_n equal to the L^2-distance, and it can be checked that $\|\cdot\|_L$ satisfies (8), see Lemma 4. However, in some situations the relation between those distances might be more involved (see for instance, in the case of Cox's model, the proof of Theorem 5). Then it might not be straightforward at all to compare the above distances and this might result in a lower rate than r_n for the concentration of the posterior in terms of $\|\cdot\|_L$.

The condition involving $\Pi^{\theta=\theta_0}$ can be checked similarly. In fact, often, one starts by choosing a prior π_f which is adapted to the nonparametric situation where θ is fixed, then check the condition in the model where $\theta = \theta_0$ and in a second step check the condition on (θ, f) about semiparametric concentration. Regarding **(C')**, the assumption on $\Pi^{\theta=\theta_0}$ is not much more difficult to check than **(C)**, especially if the sieves have ball-shapes (this is the case for the sieve used for Gaussian priors in this paper), since a translation by a small quantity can be absorbed by considering a ball of slightly larger radius. The assumption about KL-neighborhoods is classical in the literature, see also above.

We also underline the fact that for Gaussian priors on π_f, recent results of [29] provide tools for checking concentration conditions in many estimation frameworks. In particular, explicit evaluations of the rate of convergence ε_n are often available at least for some distance d and a natural choice of sieves $\mathscr{F}_n$ is provided by Borell's inequality. Detailed statements can be found in Section 5.

1.7 About the shape conditions (N), (N')

First note that if $\Delta R_n(\theta,f)$ denotes $R_n(\theta,f)-R_n(\theta_0,f)$, it holds

$$\ell_n(\theta,f)-\ell_n(\theta_0,f) = -\frac{n}{2}I_{\eta_0}(\theta-\theta_0)^2+\sqrt{n}(\theta-\theta_0)W_n(1,0)+\Delta R_n(\theta,f),$$

and a similar identity holds for $\ell_n(\theta,f)-\ell_n(\theta_0,f-(\theta-\theta_0)\gamma)$ in the information loss case, involving $\tilde{I}_{\eta_0}$ and $W_n(1,-\gamma)$. Thus **(N)**, **(N')** quantify how much these likelihood differences differ from a "parametric"-type likelihood. Note also that **(N)** is weaker than a condition where the difference of R_n's would be replaced by the single $R_n(\theta,f)$. The two conditions can nevertheless be equivalent, for instance if $R_n(\theta_0,f)=0$ as is the case for the model (9) of translation parameter estimation considered in Section 2.

Note that there is no term $\|0,f-f_0\|_L$ appearing in the control of $R_n(\theta,f)$. One could ask whether it is possible to get results under a weaker condition involving $\|0,f-f_0\|_L$. Let us consider for instance the assumption

$$\textbf{(M)} \qquad \sup_{(\theta,f)\in V_n}\frac{|R_n(\theta,f)-R_n(\theta_0,f)|}{1+h^2+\delta_n\|0,a\|_L^2}=o_{P_{\eta_0}^{(n)}}(1),$$

where δ_n is a sequence tending to zero. We believe that it is possible to replace **(N)** by **(M)** only at the cost of assuming that $n\varepsilon_n^2\delta_n\to 0$, as $n\to+\infty$. But then **(M)** does not improve on **(N)** because on V_n, we would then have $\delta_n\|0,a\|_L^2\le n\varepsilon_n^2\delta_n=o(1)$. Here is a brief heuristical justification of the preceding claim. The effect of a perturbation of the likelihood of the order $\delta_n\|0,a\|_L^2$ makes appear additional terms of the type $\Pi\left(\exp\left\{\delta_n\|0,a\|_L^2\right\}|X^{(n)}\right)$ in the proofs. If the posterior contracts at rate exactly ε_n, those are of the order of $\exp(\delta_n n\varepsilon_n^2)$ hence the condition.

Let us still give more insight into condition **(N)** by considering a special case. In the i.i.d. case with observations $(X_1,\ldots,X_n)$, let us assume that the likelihood can be written in the form $(\theta,f)\to\sum_{i=1}^n\ell(\theta,f(X_i))$. Then, provided enough regularity is present in the model, we might want to see $R_n(\theta,f)$ as the remainder of a Taylor expansion of the preceding function around (θ_0,f_0). If we neglect the terms of Taylor's expansion of order higher than 3, we are left with

$$R_n(\theta,f) \approx \sum_{i=1}^{n} (\theta-\theta_0)^3 g_{3,0}(\theta_0,f_0,X_i) + (\theta-\theta_0)^2 (f-f_0)(X_i) g_{2,1}(\theta_0,f_0,X_i)$$
$$+(\theta-\theta_0)(f-f_0)^2(X_i) g_{1,2}(\theta_0,f_0,X_i) + (f-f_0)^3(X_i) g_{0,3}(\theta_0,f_0,X_i),$$

where the g's denote (up to numerical constants) the partial derivatives at the order 3. By subtracting $R_n(\theta_0,f)$ the last term vanishes and

$$R_n(\theta,f) - R_n(\theta_0,f) \approx (\theta-\theta_0)h^2\overline{g_{3,0}} + h^2\overline{(f-f_0)g_{2,1}} + \sqrt{n}h\overline{(f-f_0)^2 g_{1,2}},$$

where the bar denotes the empirical mean. Since we are in a neighborhood of size ε_n of (θ_0,f_0), the first term in the preceding display is $o(h^2)$, the second term is expected to be $h^2 O(\|f-f_0\|) = o(h^2)$, while the last one should have order $O(h\sqrt{n}\|f-f_0\|^2)$. This might lead to a condition of the type $\sqrt{n}\varepsilon_n^2 \to 0$, which limits the range of ε_n. This appears to be an analog of the "no-bias" condition arising in frequentist contexts (see e.g. [27], Section 25.8). This is exactly what happens in Cox's model (one just needs to replace $R_n(\theta_0,f)$ by $R_n(\theta_0,f-(\theta-\theta_0)\gamma)$ but the argument remains unchanged), see Section 4.3. By contrast in model (9) we will show that no special assumption on ε_n is needed, at least if f_0 is smooth enough.

Verification of **(N')** is similar to checking **(N)**. A possibility to obtain **(N')** is to establish that **(N)** holds together with the fact that a similar condition involving $R_n(\theta_0,f) - R_n(\theta_0,f-(\theta-\theta_0)\gamma_n)$ holds.

Finally, let us comment on the assumption of LAN with linear paths and its link to **(N)**-**(N')**. The first one requires a pointwise control of the remainder $R_n(\theta,f)$ in a neighborhood of size $1/\sqrt{n}$ of the true while the second requires a uniform control over sieves of the differences $\Delta R_n(\theta,f)$ in a neighborhood of larger size ε_n. Though the second does not imply the first in general, the first is generally much easier to check. We also note that having LAN with linear paths is not needed for Theorems 1 and 2 to hold but assuming it enables to interpret the results in terms of efficiency (convolution theorems and resulting efficiency notions being established for models with LAN with linear paths in [21]).

1.8 About Gaussian priors, (3) and **(E)**

In the information loss case in Theorem 2, we assume that π_f belongs to the class of Gaussian priors. This choice of prior entails some restrictions on the models we can deal with, the main one being that the random f's generated by the prior must be almost surely in $\mathscr{F}$. This excludes the modelling of nuisances f with too many restrictions. Simple types of restrictions such as symmetry or periodicity can nevertheless be handled, see for instance the example of prior (10).

A natural way to obtain (3) and **(E)** is as follows. The support of a Gaussian prior in $\mathbb{B}$ is the closure of its RKHS $\mathbb{H}$ in $\mathbb{B}$, see e.g. [30], Lemma 5.1. Thus if γ belongs to this support, then (3) is verified with $\|\cdot\|_L$ replaced by $\|\cdot\|_{\mathbb{B}}$. If these two norms combine correctly, one deduces a rate ρ_n satisfying (3).

Even if the Gaussianity assumption is not satisfied, our proof can be adapted in some cases. For instance, as a referee pointed out, Theorem 2 carries over to the case of a finite mixture of Gaussian processes for π_f by conditioning, as long as the assumptions are satisfied for each individual component of the mixture. An even further extension would be the case of continuous mixtures, but extra work would be needed to control uniformly the individual results on each component. Finally, in the special case of truncated series priors - with non-Gaussian components-, in the case explicit expressions would be available for the posterior, an analog of the change of variables used here could be considered component by component, although one would lose the interpretation in terms of RKHS.

2 Applications

2.1 Translation parameter estimation in Gaussian white noise

One observes sample paths of the process $\{X^{(n)}(t)\}$ such that

$$dX^{(n)}(t) = f(t-\theta)dt + \frac{1}{\sqrt{n}}dW(t), \quad t \in [-1/2, 1/2], \tag{9}$$

where the unknown function f is *symmetric* (that is $f(-x) = f(x)$ for all x), 1-periodic and when restricted to $[0,1]$, belongs to $L^2[0,1]$. The unknown parameter θ is the center of symmetry of the *signal* $f(\cdot - \theta)$ and is supposed to belong to $\Theta = [-\tau_0, \tau_0] \subset]-1/4, 1/4[$. Here W is standard Brownian motion on $[-1/2, 1/2]$. We shall work in the asymptotic framework $n \to +\infty$.

Model (9) can be seen as an idealized version in continuous time of a signal processing problem where one would observe discretely sampled and noisy observations of a symmetric signal, see for instance [15] and references therein for the parametric case and [7] for frequentist estimators in the semiparametric case.

Let us define $\mathscr{F}$ as the linear space of all *symmetric* square-integrable functions $f : [-1/2, 1/2] \to \mathbb{R}$ and, for simplicity in the definitions of the classes of functions below, such that $\int_0^1 f(u)du = 0$. We extend any $f \in \mathscr{F}$ by 1-periodicity and denote its real Fourier coefficients by $f_k = \sqrt{2}\int_0^1 f(u)\cos(2\pi ku)du$, $k \geq 1$. Note that we can still denote by f_0 the "true" function f. Let us denote $\varepsilon_k(\cdot) \triangleq \cos(2\pi k \cdot)$ for $k \geq 0$. Finally $\|\cdot\|$ is the L^2-norm over $[-1/2, 1/2]$.

Conditions (R). A function $f = (f_k)_{k\geq 1}$ is said to fulfill conditions **(R)** if there exist reals $\rho > 0$, $L > 0$ and $\beta > 1$ such that $|f_1| \geq \rho$ and $\sum_{k\geq 1} k^{2\beta} f_k^2 \leq L^2$. In the sequel it is assumed that the true f_0 satisfies conditions **(R)**.

The likelihood associated to a path $X^{(n)}$ in (9) is defined with the help of a common dominating measure, here the probability $\mathbf{P}_0^{(n)}$ generated by $1/\sqrt{n}$ times Brownian motion on $[-1/2, 1/2]$. Thanks to Girsanov's formula,

$$\frac{d\mathbf{P}_{\theta,f}^{(n)}}{d\mathbf{P}_0^{(n)}}(X^{(n)}) = \exp\left\{ n\int_{-1/2}^{1/2} f(u-\theta)dX^{(n)}(u) - \frac{n}{2}\int_{-1/2}^{1/2} f(u-\theta)^2 du \right\}.$$

and $2K(\mathbf{P}_{\theta,f}^{(n)}, \mathbf{P}_{\tau,g}^{(n)}) = V(\mathbf{P}_{\theta,f}^{(n)}, \mathbf{P}_{\tau,g}^{(n)}) = n\|f(\cdot-\theta) - g(\cdot-\tau)\|^2$ as simple calculations reveal. Moreover, in model (9) there is no information loss (see Section 4.1), thus we are in the framework of Theorem 1.

Prior. For the parametric part π_θ of the prior, we choose any probability measure which satisfies condition (**P**) (for instance the uniform measure on $[-1/4, 1/4]$). The nonparametric part π_f is chosen among a family of Gaussian priors parametrized by a real parameter α. Let $\{\nu_k\}_{k\geq 1}$ be a sequence of independent standard normal random variables and for any $k > 0$ and $\alpha > 1$ let $\sigma_k = k^{-1/2-\alpha}$. The prior π_f^α is the distribution generated by

$$f(\cdot) = \sum_{k=1}^{+\infty} \sigma_k \nu_k \varepsilon_k(\cdot). \tag{10}$$

Let us also define $\pi_{f,k(n)}^\alpha$ as the distribution of $g(\cdot) = \sum_{k=1}^{k(n)} \sigma_k \nu_k \varepsilon_k(\cdot)$, where $k(n)$ is a strictly increasing sequence of integers. The latter prior depends on n, which is often viewed as non-desirable, but the entropy bounds involved to get posterior convergence are easier to obtain since just a finite number of Fourier coefficients are involved for all n. This also explains why for this prior the domain where the BVM-theorem holds is slightly larger in the following theorem. In both cases α can be seen as the “regularity” of the prior and the rate convergence of the associated posterior for estimating f of Sobolev-regularity β is quantified in Lemma 14.

Theorem 3 *Suppose that f_0 satisfies* (**R**) *with regularity $\beta > 1$. Let the prior π_θ satisfy* (**P**) *and let π_f be defined by* (10) *for some $\alpha > 1$. Then conditions* (**C**) *and* (**N**) *of Theorem 1 are satisfied for pairs (β, α) such that the corresponding point in Figure 1 lies in the shaded area. In particular, the BVM theorem holds in this region. For the prior $\pi_{f,k(n)}^\alpha$ with $k(n) = \lfloor n^{1/(2\alpha+1)} \rfloor$, the same holds in the region delimited by the ‘triangle’-curve.*

The region for which $\beta > 1$ and $\alpha > 1$ delimited by the ‘square’-curve in Figure 1 can be regarded as the ‘best possible’ region, since it describes true functions and priors which have at least one derivatives in a weak (L^2-) sense. This condition on β is necessary to have a finite Fisher information, which here equals $\|f_0'\|_2$. Thus with this respect the results of Theorem 3 are quite sharp, in that only a small strip in the region where α or β are very close to 1 is not covered by Theorem 1. More precisely, the region where BVM holds is defined by $\alpha > 1 + \sqrt{3}/2$ (resp $\alpha > 3/2$ for the truncated prior), $\beta > 3/2$ and, finally, $\alpha < (3\beta - 2)/(4 - 2\beta)$, which corresponds to the non-linear curve in Figure 1. These mild conditions arise when checking (**N**).

For instance for $\beta = 2$, any prior of the type (10) with $\alpha > 1 + \sqrt{3}/2$ will do. This means also that in model (9) for $\beta \geq 2$, *no condition on the nonparametric concentration rate ε_n of the posterior is needed* to get the semiparametric BVM theorem. For example, it can be easily seen that if $\beta = 2$ and α increases, ε_n becomes slower and slower (in fact if $\beta = 2$ and $\alpha \geq 2$, then ε_n can be as slow as $n^{-2/(2\alpha+1)}$, see [5]).

The results of Theorem 3 can be extended to other families of priors. For instance, for Gaussian series priors for which just an equivalent of σ_k is available in the form of a power of k^{-1} as $k \to \infty$, the method presented here could be adapted, since equivalents of the small ball probabilities are still available for these priors. One could also consider non-Gaussian priors in form of infinite series, in this case however, understanding the concentration properties of the posterior for π_f in the nonparametric setting would be very desirable and will be the object of future research.

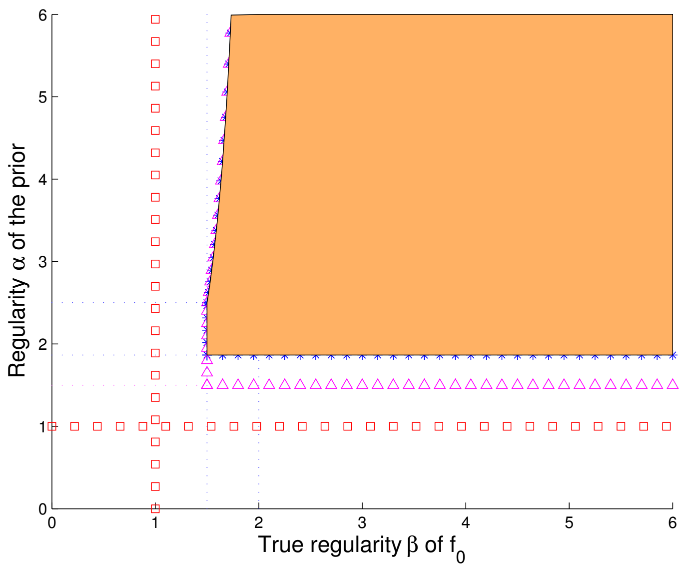

Fig. 1 Translation model. Possible choices for π_f

2.2 A functional data analysis model

A referee suggested, as a follow-up to the white noise model (9), to investigate what the theory would give in the following functional data analysis case of non-parametric regression. For $i = 1, \ldots, n$, one observes independent realizations of

$$Y_i = f(t_i) + \sigma\varepsilon_i \tag{11}$$

$$Z_i = f(t_i - \theta) + \tau\zeta_i,, \tag{12}$$

with ε_i, ζ_i independent standard normal random variables. The pair (θ, f) is unknown and t_i are fixed design points (say $t_i = i/n$, for $i = 1, \ldots, n$). Here, the parameter of interest θ represents the relative translation between the two groups observed through the Y's and Z's respectively.

This example indeed nicely illustrates how the theory can be applied in a framework with discrete observations, more directly related to real data applications. Another important difference with model (9) is that a loss of information occurs, as can be seen from the LAN expansion with linear paths arising for this model. Though, due to the similarity in structure it is still possible to use or adapt several of the arguments used for model (9). We shall now briefly describe and discuss the results.

Regularity conditions. Let us assume that the parameter set $\mathscr{E} = \Theta \times \mathscr{F}$ is such that Θ is a compact interval $[-\tau_0, \tau_0] \subset]-1/2, 1/2[$. The elements f of $\mathscr{F}$ are square integrable functions on $[0,1]$ and extended by 1-periodicity to $\mathbb{R}$. The complex Fourier basis is denoted $\psi_k(t) = e^{2ik\pi t}$ and $g_k = \int_0^1 g(t)\psi_{-k}(t)dt$ for any $g \in \mathscr{F}$.

We further assume that the true function f_0 fulfills some regularity conditions. For some $\rho > 0, \beta > 1, L_1 > 0, L_2 > 0$, the function f_0 belongs to the following class of continuously differentiable functions ($\mathscr{C}^1$ functions)

$$\mathscr{C}(\beta,\rho,L_1,L_2) = \{g \in \mathscr{C}^1,\ g_0 = 0,\ |g_1| \geq \rho,\ \sum_{k\in\mathbb{Z}} k^{2\beta}|g_k|^2 \leq L_1^2,\ \sum_{k\in\mathbb{Z}} |kg_k| \leq L_2\}.$$

We also assume that the noise variances σ, τ are known and we set $\sigma = \tau = 1$.

Prior. The prior Π on (θ, f) is chosen of the form $\pi_\theta \otimes \pi_f$ with π_f a Gaussian series prior. To simplify we shall take finite series, which suffice to illustrate our point, but the case of infinite series could also be considered, so we set

$$g(\cdot) = \sigma_0 \nu_0 + \sum_{k=1}^{k(n)} \sigma_{2k}\nu_{2k}\cos(2\pi k\cdot) + \sigma_{2k-1}\nu_{2k-1}\sin(2\pi k\cdot), \tag{13}$$

where $k(n) \to +\infty$ is an increasing sequence of integers, ν_k is a sequence of independent standard normal variables, and σ_k the sequence $\sigma_{2k} = \sigma_{2k+1} = k^{-\alpha-1/2}$ for $k \geq 1$ and $\sigma_0 = 1$. For $k(n)$ we choose $k(n) = \lfloor n^{1/(2\alpha+1)} \rfloor$.

Theorem 4 *Suppose that f_0 satisfies the above regularity conditions with $\beta > 3/2$. Let the prior π_θ satisfy* **(P)** *and let π_f be defined by* (13) *for some $\alpha > 3/2$. Then the semiparametric BVM theorem holds in model* (11)-(12) *for all pairs (β, α) such that $\alpha > 3/2$ and $\alpha < 2\beta - 3/2$.*

The region in (β, α) defined by these two conditions is an (affine) convex cone (see the next section for a picture of a similar region for Cox model). One can also notice that this region is included in the region obtained for the same prior in the previous model and depicted in Figure 1 by the 'triangle'-curve. The proof of Theorem 4 is in the same spirit as the one of Theorem 3. We shall not give the full proof here but only check **(E)**. This will be done in Section 4.2, where we show that the limiting condition $\alpha < 2\beta - 3/2$ originates from **(E)**. More precisely, $\sqrt{n}\varepsilon_n\rho_n \to 0$ is only satisfied if the least favorable direction can be well enough approximated by the RKHS of the prior.

As long as we choose α not too large, the BVM theorem holds for any β larger than a reasonably small value. For instance, for $\alpha = 2$ we need $\beta > 1.75$ and for $\alpha = 3$ we need $\beta > 2.25$. Since we have the choice of the parameter α of the prior, the best choice is the smallest possible α. Since we need $\alpha > 3/2$ one can take $\alpha = 3/2 + \delta$ for some small δ in which case Theorem 4 implies that the BVM theorem holds from $\beta > 3/2 + \delta/2$.

2.3 Cox's proportional hazards model

The observations are a random sample from the distribution of the variable (T,δ,Z), where $T = X \wedge Y$, $\delta = \mathbf{1}_{X \leq Y}$, for some real-valued random variables X, Y and Z. We assume that the variable Z, called covariate, is bounded by M and admits a continuous density φ with respect to Lebesgue's measure on $[-M,M]$. Suppose that given Z, the variables X and Y are independent and that there exists a real $\tau > 0$ such that $\mathbf{P}_{\eta_0}(X > \tau) > 0$ and $\mathbf{P}_{\eta_0}(Y \geq \tau) = \mathbf{P}_{\eta_0}(Y = \tau) > 0$. The conditional hazard function α of X knowing Z is defined by $\alpha(x)dx = \mathbf{P}(X \in [x, x+dx] \mid X \geq x,\ Z)$. Cox's model assumes that $\alpha(x) = e^{\theta Z}\lambda(x)$, where λ is an unknown hazard function and θ a real parameter. For simplicity, we have assumed that Z and θ are one-dimensional.

Let us assume that λ_0 is continuous and that there exists a $\rho > 0$ such that, for all x in $[0,\tau]$, one has $\lambda_0(x) \geq \rho > 0$. We will denote $\Lambda(x) = \int_0^x \lambda(u)du$. We also assume that Y given $Z = z$ admits a continuous density g_z with respect to Lebesgue's measure on $[0,\tau)$ with distribution function G_z and that there exists $\rho' > 0$ such that $g_z(t) \geq \rho'$ for almost all z,t. Finally we assume that the possible values of θ lie in some compact interval $[-\theta_M, \theta_M]$.

In this semiparametric framework the unknown parameter η will be taken equal to $\eta = (\theta, \log\lambda)$. In the sequel we use the notation $r \triangleq \log\lambda$. Under these assumptions, the triplet (T,δ,Z) admits a density $f_\eta(t,d,z)$ with respect to the dominating measure $\mu = \{\mathscr{L}_{[0,\tau]} + \delta_\tau\} \otimes \{\delta_0 + \delta_1\} \otimes \mathscr{L}_{[-M,M]}$, where $\mathscr{L}_I$ denotes Lebesgue's measure on the interval I and

$$f_\eta(t,d,z) = \left\{g_z(t)e^{-\Lambda(t)e^{\theta z}}\right\}^{1-d}\left\{(1-G_z(t-))\lambda(t)e^{\theta z - \Lambda(t)e^{\theta z}}\right\}^{d}\varphi(z)\mathbf{1}_{t<\tau}(t) + \left\{(1-G_z(\tau-))e^{-\Lambda(\tau)e^{\theta z}}\right\}\varphi(z)\mathbf{1}_{d=0,\ t=\tau}(d,t).$$

The log likelihood-ratio associated to the data $(T_i,\delta_i,Z_i)_{i=1,\ldots,n}$ is

$$\Lambda_n(\theta,r) = \sum_{i=1}^{n} \delta_i\{(r-r_0)(T_i) + (\theta-\theta_0)Z_i\} - \Lambda(T_i)e^{\theta Z_i} + \Lambda_0(T_i)e^{\theta_0 Z_i}.$$

Prior. We construct Π as $\pi_\theta \otimes \pi_f$ with π_θ having a positive continuous density with respect to Lebesgue's measure on a compact interval containing $[-\theta_M,\theta_M]$. As prior π_f on $r = \log\lambda$, we take a Gaussian prior belonging to a 1-parameter family. For $\alpha > 0$ and W standard Brownian motion, we define the *Riemann-Liouville type process* with parameter $\alpha > 0$ as

$$X_t = \int_0^t (t-s)^{\alpha-1/2}dW_s + \sum_{k=0}^{\underline{\alpha}+1} Z_k t^k, \quad 0 \leq t \leq \tau, \tag{14}$$

where $Z_0,\ldots,Z_{\underline{\alpha}+1},R_t$ are independent and Z_i is standard normal. If $\alpha = 1/2+k$, with $k \in \mathbb{N}$, then, up to the polynomial part, the process X is simply k-times integrated Brownian motion. This prior generates a distribution on the space of continuous functions equipped with the sup-norm $\|\cdot\|_\infty$. Its properties for Bayesian estimation are studied in [29, 5] and summarized in Section 5, see in particular Lemma 16.

Theorem 5 *Suppose that* $\log \lambda_0$ *belongs to* $\mathscr{C}^\beta[0,\tau]$ *with* $\beta > 3/2$. *Suppose the least favorable direction* γ *has Hölder regularity at least* $2\beta/3$. *Let the prior* π_θ *be defined as described above and* π_f *be a Riemann-Liouville type process with parameter* $\alpha > 3/2$. *Then the conditions* **(P)**, **(C')**, **(N')** *and* **(E)** *of Theorem 2 are satisfied for pairs* (β, α) *such that the corresponding point in Figure 2 lies in the shaded area. In particular, the BVM theorem holds when* $\alpha > 3/2$ *and* $\alpha < 4\beta/3 - 1/2$.

The main difference with Theorem 3 is the triangular shape of the region we obtain. The origin of this shape is that, as an examination of the proof reveals, the rate ζ_n of estimation of the nonparametric part r in terms of the LAN-norm must verify $\zeta_n = o(n^{-1/4})$ to satisfy condition **(N')**. Since we are able to choose $\zeta_n = \sqrt{n}\varepsilon_n^2$, where ε_n is the rate in terms of the Hellinger distance, the condition is $n^{3/4}\varepsilon_n^2 \to 0$. So, if α is much larger than β, the concentration rate of the posterior becomes too small and the condition cannot be fulfilled. As discussed in Section 1.7, this phenomenon corresponds to a "no-bias"-type condition. Note also that **(E)** imposes a somewhat similar condition by asking $\sqrt{n}\zeta_n\rho_n \to 0$, where ρ_n is the approximation rate of γ. The assumption that γ is at least $2\beta/3$-Hölder actually makes this condition always weaker than the previous one, as can be checked from the expression of the rates ε_n and ζ_n, see Section 4.3. However, if one would allow for much less regular γ's, then **(E)** would become also here the limiting condition, as was the case for model (11)-(12). Finally note that the regularity condition on γ is not difficult to check, due to the existence of an explicit form (24) for γ. For instance if Y and Z are independent, in which case $G_z(u)$ does not depend on z, then γ has the same regularity as Λ_0 so its Hölder regularity is at least $\beta + 1 > 2\beta/3$ and the condition is verified.

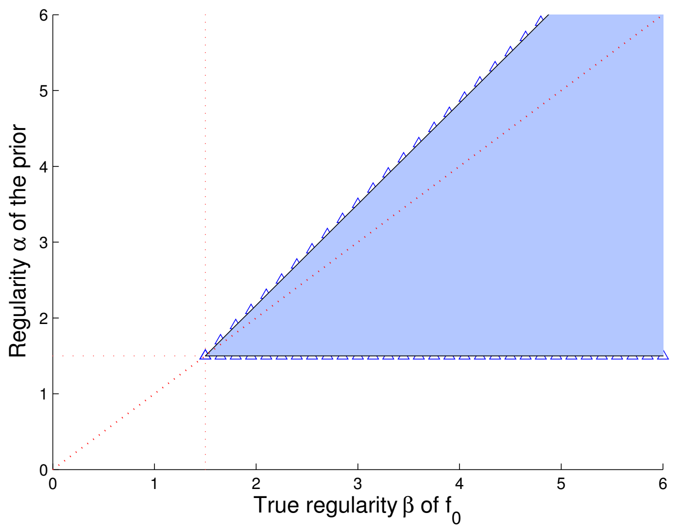

Fig. 2 Cox's model. Possible choices for π_f

2.4 Discussion and perspectives

Comparison with the approach in [24]. First, the approach relies on properties of the non-parametric maximum likelihood $\hat{\theta}^{MLE}$ or rather, as the author notes, on an appropriately penalized version of it, where the penalty has to be tuned. As discussed above, it provides more generality to see the BVM theorem as centered at Δ_{n,θ_0} (or at any efficient estimator of θ). This way one does not have to carry out the work of showing that the "MLE" is efficient which would be in a way "doing the work twice". Moreover, sufficient conditions ensuring consistency and efficiency of the MLE might turn out slightly stronger than the ones for required for the (general form of the) BVM theorem to hold.

More importantly, several conditions in [24] are implicit, for instance, Conditions (14)-(15) p. 229. Roughly, Condition (14) says that the dependence in the parameter θ of the likelihood integrated with respect to f should have a negligible contribution. This is an important step in the proof and putting it as an assumption appears problematic as soon as no explicit expressions for the quantities at stake are available. Also, computing the rate of $g_n(\hat{\theta})$ in [24], Conditions (9) and (14), as the author himself notes on p. 230 "may require considerable effort" and seems a rather difficult task if no explicit formulas are available. By contrast, though we restrict our investigations to Gaussian process priors for π_f, we are able to give a simple interpretable condition (condition **(E)**) to treat the integrated likelihood in f, in terms of approximation properties of the least favorable direction by the RKHS of the prior.

To illustrate the relative ease of our conditions, let us take the example of Cox Model of Section 2.3. Note that the form of the prior (14) prevents from using explicit expressions. To verify **(C')**, the general techniques of [13] can be used to obtain rates of posterior concentration. Since the RKHS of the prior (14) is well-known, checking **(E)** is not difficult, see Section 4. We do not have to deal with a nonparametric $\hat{\theta}^{MLE}$, which would in this example turn the study of the $g_n(\hat{\theta})$ in [24] even more complex, even putting aside the fact that no direct computations of the integrated likelihood seem possible for the considered model and prior, at least in a simple way.

Also, our approach enables a more thorough investigation of regularity issues. Indeed, while results in [24] are typically given for one fixed prior, here we pushed the analysis one step further by considering a priori all pairs (α,β) and providing positive results in full regions of the space with (β,α) coordinates, as represented in Figures 1, 2. In the case of the Gaussian white noise model for example, the region seems fairly optimal.

Shape of the satisfaction regions in (β,α). Several conclusions can be drawn from Figures 1, 2. The triangle-type shapes of the regions for Cox model (and for model (11)-(12), that we did not depict) suggest one should avoid choosing too smooth priors, for which α is very large compared to β. Two different cases can occur. First, the uniform control **(N')** might in some cases (e.g. Cox model) require that the size of the shrinking neighborhood, controlled by ε_n, be not too large. Due to the typical form of rates for Gaussian priors as established in [29],[5], this fails for α large and β small. Second, again for α large and β small, the approximation of γ by the RKHS of the prior can become too slow, preventing **(E)** to be true. This happens for model

(11)-(12), and can also happen for Cox model if γ is much less smooth than $\log \lambda_0$, as noted above.

The preceding suggests choosing a prior with α as small as allowed by the theory. For this α the semiparametric BVM theorem then holds for most β's. For instance, in model (11)-(12), the BVM theorem holds for all regularities β's of the true f_0 larger than $3/2+\delta/2$ if one chooses $\alpha = 3/2+\delta$ and a small $\delta > 0$. Note also that we do not use here the knowledge of the "true" regularity β of f_0. Our result says that the BVM theorem holds for some universal choice of α. The effective concentration rate of the nonparametric part of the correponding posterior will in general be slower than the "adaptive" rate for nonparametric estimation (the minimax rate on the corresponding class of functions of regularity β) which here would typically be in $n^{-\beta/(2\beta+1)}$.

At least some care in the choice of the nonparametric part of the prior is needed in general, especially if a loss of information occurs. For some special models however, it can happen that most priors work. Model (9), where all priors with $\alpha > 2$ work if β is at least 2, is an example. Nevertheless, we point out that in a forthcoming work we show that condition **(E)** is necessary in general in that, in most cases, failure of **(E)** will result in the BVM theorem not being satisfied, due to an extra bias term appearing in this case. The triangular shape of the satisfaction region thus appears to be unavoidable in general in the loss of information case.

Future work. Apart from the models studied here, the methods presented can already be applied to a variety of models, among which other Gaussian white noise models (e.g. the period estimation considered in [4]), the density estimation version of model (9), the class of partially linear models with Gaussian errors, see [14], to name a few. Beyond the scope of this paper, it would be interesting to extend Theorem 2 to non-Gaussian priors, which would probably require an equivalent of the change of variables used here, in terms of measuring the influence of translations of small amplitude on the nonparametric part of the prior. Also, concentration results including construction of new types of sieves would have to be obtained for those priors. Future results in the Bayesian nonparametrics field will hopefully enable to answer these questions. Beyond "separated" semiparametric models with unknown (θ, f), obtaining results about finite-dimensional functionals of a nonparametric distribution is also of great interest. For sufficiently smooth functionals, the methods presented here can presumably be adapted, essentially by conditioning on the value of the functional of interest in the nonparametric prior and thus in a way recovering a type of separated structure in two parts for the integrals at stake. However, this framework deserves special attention, in particular due to the variety of situations which can arise (e.g. rates slower that $n^{-1/2}$ even for simple functionals like the square norm of a density) and will be treated in future contributions.

3 Proof of the main results

3.1 Proof of Theorem 1

First one localizes the problem around the true parameter using the concentration result **(C)**. Then one can use the local shape property **(N)** and further split up the likelihood in a "θ-part" and a "f-part". For any B Borel set in Θ, let us denote $D = B \times \mathscr{F}$, then using the local sieve V_n introduced for condition **(N)**,

$$\Pi(D|X^{(n)}) = \Pi(D \cap V_n|X^{(n)}) + \Pi(D \cap V_n^c|X^{(n)}).$$

But $\Pi(D \cap V_n^c|X^{(n)}) \le \Pi(V_n^c|X^{(n)})$ and $\Pi(V_n^c|X^{(n)})$ tends to 0 in probability as $n \to +\infty$ due to **(C)**. Note that if Π^A denotes the restriction of the prior Π to A, with the corresponding posterior we have

$$\Pi(D \cap V_n|X^{(n)}) = \Pi^{V_n}(D|X^{(n)})\Pi(V_n|X^{(n)}),$$

and $\Pi(V_n|X^{(n)})$ tends to 1 as $n \to +\infty$ in probability again due to **(C)**. Thus it is enough to focus on $\Pi^{V_n}(D|X^{(n)})$, which explicitly is

$$q_n(B) \triangleq \frac{\int \mathbf{1}_B(\theta) \int \mathbf{1}_{V_n}(\theta, f) \exp \Lambda_n(\theta, f)\, d\pi_f(f) d\pi_\theta(\theta)}{\int\int \mathbf{1}_{V_n}(\theta, f) \exp \Lambda_n(\theta, f)\, d\pi_f(f) d\pi_\theta(\theta)} \triangleq \frac{q_1(B)}{q_0},$$

where $\Lambda_n(\theta, f)$ denotes $\ell_n(\theta, f) - \ell_n(\theta_0, f_0)$. On the local neighborhood V_n we expand the log-likelihood according to (2). Using $h = \sqrt{n}(\theta - \theta_0)$ and $a = \sqrt{n}(f - f_0)$ as shorthand notation,

$$n\|\theta - \theta_0, f - f_0\|_L^2 = \|h, a\|_L^2 = \|h, 0\|_L^2 + \|0, a\|_L^2,$$

since by assumption there is no loss of information. The linearity of W_n now implies

$$\sqrt{n} W_n(\theta - \theta_0, f - f_0) = W_n(h, a) = W_n(h, 0) + W_n(0, a).$$

The numerator $q_1(B)$ of $q_n(B)$ can thus be written,

$$q_1(B) = \int_B \exp(-\|h, 0\|_L^2/2 + W_n(h, 0)) \times$$
$$\int \mathbf{1}_{V_n}(\theta, f) \exp\left\{-\|0, a\|_L^2/2 + W_n(0, a) + R_n(\theta, f)\right\} d\pi_f(f) d\pi_\theta(\theta).$$

By linearity of the inner product we have $\|h, 0\|_L^2 = h^2\|1, 0\|_L^2$ and since there is no information loss, it holds $\tilde{I}_{\eta_0} = I_{\eta_0} = \|1, 0\|_L^2$. Let us now split the indicator of V_n using inequalities. For any $s > 0$, it holds

$$\{0 \le x \le s/2,\ 0 \le y \le s/2\} \subset \{x \ge 0,\ y \ge 0,\ x + y \le s\} \subset \{0 \le x \le s,\ 0 \le y \le s\}.$$

Applying this with $x = h^2 I_{\eta_0}$, $y = \|0, a\|_L^2$ and $s = n\varepsilon_n^2$, one obtains

$$\mathbf{1}_{h^2 I_{\eta_0} \le n\varepsilon_n^2/2}(\theta) \int \mathbf{1}_{\|0,a\|_L^2 \le n\varepsilon_n^2/2} \exp\left\{-\|0, a\|_L^2/2 + W_n(0, a) + R_n(\theta, f)\right\} d\pi_f(f)$$
$$\le \int \mathbf{1}_{V_n}(\theta, f) \exp\left\{-\|0, a\|_L^2/2 + W_n(0, a) + R_n(\theta, f)\right\} d\pi_f(f)$$
$$\le \mathbf{1}_{h^2 I_{\eta_0} \le n\varepsilon_n^2}(\theta) \int \mathbf{1}_{\|0,a\|_L^2 \le n\varepsilon_n^2} \exp\left\{-\|0, a\|_L^2/2 + W_n(0, a) + R_n(\theta, f)\right\} d\pi_f(f).$$

Now denoting $S_n = \sup_{(\theta,f)\in V_n} |R_n(\theta,f) - R_n(\theta_0,f)|/(1+h^2)$, it holds

$$-(1+h^2)S_n \leq R_n(\theta,f) - R_n(\theta_0,f) \leq (1+h^2)S_n,$$

and condition **(N)** tells us that $S_n = o_{P^{(n)}_{\eta_0}}(1)$. Further denoting

$$\zeta_n(t) = \int \mathbf{1}_{\|0,a\|_L^2 \leq nt}(a) \exp\left\{-\|0,a\|_L^2/2 + W_n(0,a) + R_n(\theta_0,f)\right\} d\pi_f(f),$$

for small enough positive reals t, one obtains

$$\zeta_n(\varepsilon_n^2/2) \int_B \mathbf{1}_{h^2 I_{\eta_0} \leq n\varepsilon_n^2/2} \exp\left\{-\|h,0\|_L^2/2 + W_n(h,0) - (1+h^2)S_n\right\} d\pi_\theta(\theta)$$

$$\leq q_1(B) \leq \zeta_n(\varepsilon_n^2) \int_B \mathbf{1}_{h^2 I_{\eta_0} \leq n\varepsilon_n^2} \exp\left\{-\|h,0\|_L^2/2 + W_n(h,0) + (1+h^2)S_n\right\} d\pi_\theta(\theta).$$

The same calculation can be done for q_0 and thus we obtain

$$\frac{\zeta_n(\varepsilon_n^2/2)}{\zeta_n(\varepsilon_n^2)} q_n^{P,-}(B) \leq q_n(B) \leq \frac{\zeta_n(\varepsilon_n^2)}{\zeta_n(\varepsilon_n^2/2)} q_n^P(B),$$

where we have denoted

$$q_n^P(B) = \frac{\int_B \mathbf{1}_{h^2 I_{\eta_0} \leq n\varepsilon_n^2} \exp\{-\|h,0\|_L^2/2 + W_n(h,0) + (1+h^2)S_n\} d\pi_\theta(\theta)}{\int \mathbf{1}_{h^2 I_{\eta_0} \leq n\varepsilon_n^2/2} \exp\{-\|h,0\|_L^2/2 + W_n(h,0) - (1+h^2)S_n\} d\pi_\theta(\theta)},$$

and $q_n^{P,-}(B)$ has a similar expression. Now $q_n(B)$ is bounded from above and below by similar quantities. Both are a product of a nonparametric part, involving π_f through the function ζ_n, and a parametric part, involving π_θ through the terms $q_n^P(B)$ and $q_n^{P,-}(B)$. The nonparametric part is handled noticing that

$$\zeta_n(\varepsilon_n^2/2)/\zeta_n(1) = \Pi^{\theta=\theta_0}\left(\{f \in \mathscr{F}_n,\ \|0, f - f_0\|_L^2 \leq \varepsilon_n^2/2\} \mid X^{(n)}\right),$$

and similarly for $\zeta_n(\varepsilon_n^2)/\zeta_n(1)$. Thus the second part of assumption **(C)** implies that $\zeta_n(\varepsilon_n^2)/\zeta_n(\varepsilon_n^2/2)$ tends to 1 in probability.

Now we handle the parametric part of the bounds on $q_n(B)$. Let us introduce the notation $B'_n = \{u,\ \theta_0 + (u+\Gamma_n)/\sqrt{n} \in B\}$ and

$$\Gamma_n = \frac{W_n(0,1)}{\|1,0\|_L^2 - 2S_n}, \qquad \overline{\Gamma}_n = \frac{W_n(0,1)}{\|1,0\|_L^2 + 2S_n},$$

$$D_n = \left\{u,\ (u+\Gamma_n)^2 \leq \frac{n\varepsilon_n^2}{\|1,0\|_L^2}\right\}, \quad \overline{D}_n = \left\{u,\ (u+\overline{\Gamma}_n)^2 \leq \frac{n\varepsilon_n^2}{2\|1,0\|_L^2}\right\}.$$

Then $q_n^P(B)$ can be rewritten as follows

$$q_n^P(B) = \frac{\int_{B'_n} \mathbf{1}_{D_n}(u) \exp\left(\{-\frac{\|1,0\|_L^2}{2} + S_n\}u^2\right) \lambda(\theta_0 + \frac{u+\Gamma_n}{\sqrt{n}}) du}{\int \mathbf{1}_{\overline{D}_n}(u) \exp\left(\{-\frac{\|1,0\|_L^2}{2} - S_n\}u^2\right) \lambda(\theta_0 + \frac{u+\overline{\Gamma}_n}{\sqrt{n}}) du} \times \frac{e^{\Gamma_n^2\left\{\frac{\|1,0\|_L^2}{2} - S_n\right\}}}{e^{\overline{\Gamma}_n^2\left\{\frac{\|1,0\|_L^2}{2} + S_n\right\}}},$$

and the lower bound $q_n^{P,-}(B)$ admits a similar expression. To conclude the proof it is enough to see that

$$\sup_B \left| q_n^P(B) - \frac{\int_B \exp\left(-\frac{\|1,0\|_L^2}{2}\{\sqrt{n}(\theta-\theta_0) - \frac{W_n(1,0)}{\|1,0\|_L^2}\}^2\right) d\theta}{\int \exp\left(-\frac{\|1,0\|_L^2}{2}\{\sqrt{n}(\theta-\theta_0) - \frac{W_n(1,0)}{\|1,0\|_L^2}\}^2\right) d\theta} \right| \to 0,$$

as $n \to +\infty$, in $P_{\eta_0}^{(n)}$-probability and that the same holds for $q_n^{P,-}(B)$. This can be checked using that $S_n = o_P(1)$, that $n\varepsilon_n^2 \to +\infty$ and that $\Gamma_n, \overline{\Gamma}_n$ are bounded in probability, together with the use of the continuous mapping theorem. The verifications are slightly technical but not difficult and are left to the reader. □

3.2 Proof of Theorem 2

The main difference with Theorem 1 is the split of the inner product

$$\|\theta-\theta_0, f-f_0\|_L^2 = \tilde{I}_{\eta_0}(\theta-\theta_0)^2 + \|0, f-f_0+(\theta-\theta_0)\gamma\|_L^2.$$

Since here γ is nonzero, the last term still contains a dependence in θ. We shall use the fact that the prior is Gaussian to change variables by setting $g = f + (\theta-\theta_0)\gamma_n$, where γ_n is a sequence approximating γ in $\mathbb{H}$ and satisfying (3).

We start by setting some notation for the change of variables. The prior π_f is the law of a centered Gaussian process Z in a Banach space $(\mathbb{B}, \|\cdot\|)$ of real functions with RKHS $(\mathbb{H}, \|\cdot\|_{\mathbb{H}})$ and covariance kernel $K(s,t) = \mathbf{E}(Z(s)Z(t))$. Denoting by Ω the probability space on which Z is defined, one can define a map $U : \mathbb{H} \to L^2(\Omega)$ as

$$U : K(t,\cdot) \to Z(t)$$

and extending linearly and continuously, see Section 5.4 for details. The change of variable formula for Gaussian measures, see for instance [30], Lemma 3.1, states that if P^Z denotes the law of the Gaussian process Z and if ϕ is an element of its RKHS $\mathbb{H}$, then the measures P^Z and $P^{Z-\phi}$ are absolutely continuous with $dP^{Z-\phi}/dP^Z = \exp(U(-\phi) - \|\phi\|_{\mathbb{H}}^2/2)$. Below we will slightly restrict the range of U and use the adapted formula given in Lemma 17. In the sequel it is also technically useful to see the variable f as a random variable $\omega \to f(\omega)$, for ω in Ω. In particular for any measurable ψ we write $\int \psi(f(\omega))dP_f(\omega)$ instead of $\int \psi(f)d\pi_f(f)$. The reason is that the random variable $U(g)$ for $g \in \mathbb{H}$ is naturally defined on Ω.

Let us now turn to the proof of Theorem 2. First we introduce the subset $\mathscr{V}_n = \{(\theta,\omega) \in \Theta \times \Omega,\ (\theta, f(\omega)) \in V_n\}$. By the same argument as in the proof of Theorem 1, due to **(C')** one can restrict the study of the posterior to $\mathscr{V}_n$. One can also restrict slightly the set of $\omega's$ by considering, for a large enough constant M, the set

$$\mathscr{C}_n = \left\{\omega \in \Omega, \quad |U\gamma_n|(\omega) \le M\sqrt{n}\varepsilon_n\|\gamma_n\|_{\mathbb{H}}\right\}.$$

Since $U\gamma_n$ is centered, normally distributed with variance $\|\gamma_n\|_{\mathbb{H}}^2$, we have the bound $P_f(\mathscr{C}_n^c) \lesssim \exp(-n\varepsilon_n^2 M^2/2)$. The assumption on $B_{KL,n}$ together with Lemma 2 imply that for M large enough, it holds $(\pi_\theta \otimes P_f)(\Theta \times \mathscr{C}_n | X^{(n)}) \to 1$.

The preceding shows that it is enough to focus on the ratio

$$\frac{\int_B \int_{\mathscr{C}_n} \mathbf{1}_{\mathscr{V}_n}(\theta,\omega) \exp \Lambda_n(\theta, f(\omega))\, dP_f(\omega) d\pi_\theta(\theta)}{\int \int_{\mathscr{C}_n} \mathbf{1}_{\mathscr{V}_n}(\theta,\omega) \exp \Lambda_n(\theta, f(\omega))\, dP_f(\omega) d\pi_\theta(\theta)} = \frac{s_1(B)}{s_0}$$

and bound it from above and below. In total there are four inequalities to establish (upper and lower bound for $s_1(B)$ and s_0). Let us prove that

$$\begin{aligned} s_1(B) \le & \int_B \exp\left\{-h^2/2 + hW_n(1,-\gamma) + o_P(1+h^2)\right\} d\pi_\theta(\theta) \qquad (15) \\ & \times E_{P_f}\left(\mathbf{1}_F \exp\left\{-\|0,a\|_L^2/2 + W_n(0,a) + R_n(\theta_0, f)\right\}\right), \end{aligned}$$

where the set F is defined as

$$F = \{\omega \in \Omega,\ \ |U\gamma_n|(\omega) \le 2M\sqrt{n}\varepsilon_n \|\gamma_n\|_{\mathbb{H}},\ \ g(\omega) \in \mathscr{F}_n(\theta),\ \ \|0, g(\omega) - f_0\|_L \le 2\varepsilon_n\},$$

the other three inequalities being established in a similar way.

Since $\|\theta-\theta_0, f-f_0\|_L^2 = \tilde{I}_{\eta_0}(\theta-\theta_0)^2 + \|0, f+(\theta-\theta_0)\gamma - f_0\|_L^2$ and, for n large enough, $\|0,(\theta-\theta_0)(\gamma-\gamma_n)\|_L \le \varepsilon_n$, it holds

$$\{\|\theta-\theta_0, f-f_0\|_L \le \varepsilon_n\} \subset \{\tilde{I}_{\eta_0}(\theta-\theta_0)^2 \le \varepsilon_n^2,\ \|0, f+(\theta-\theta_0)\gamma_n - f_0\|_L \le 2\varepsilon_n\}.$$

Thus by factorizing the terms in θ one obtains as upper bound for $s_1(B)$ the first part of the right hand-side of (15) times an integral with respect to P_f. This latter integral, which depends on θ, equals

$$\begin{aligned} &\int \mathbf{1}\{\omega \in \mathscr{C}_n,\ \ f(\omega) \in \mathscr{F}_n,\ \|0, f(\omega) + (\theta-\theta_0)\gamma_n - f_0\|_L \le 2\varepsilon_n\} \times \\ &\exp\left\{-\frac{n}{2}\|0, f+(\theta-\theta_0)\gamma - f_0\|_L^2 + \sqrt{n}W_n(0, f+(\theta-\theta_0)\gamma - f_0) + R_n(\theta, f)\right\} dP_f(\omega) \end{aligned}$$

Let us rewrite the term in the exponential introducing γ_n as

$$\begin{aligned} & -n\|0, f+(\theta-\theta_0)\gamma - f_0\|_L^2/2 + \sqrt{n}W_n(0, f+(\theta-\theta_0)\gamma - f_0) + R_n(\theta, f) \\ & = -n\|0, f+(\theta-\theta_0)\gamma_n - f_0\|_L^2/2 + \sqrt{n}W_n(0, f+(\theta-\theta_0)\gamma_n - f_0) \\ & \quad + R_n(\theta_0, f+(\theta-\theta_0)\gamma_n) + \{R_n(\theta, f) - R_n(\theta_0, f+(\theta-\theta_0)\gamma_n)\} + Q_n(\theta, f), \end{aligned}$$

where $Q_n(\theta, f)$ is the remainder term, that is

$$\begin{aligned} Q_n(\theta, f) = & -n(\theta-\theta_0)^2\|0, \gamma-\gamma_n\|_L^2/2 + \sqrt{n}(\theta-\theta_0)W_n(0, \gamma-\gamma_n) \\ & -n(\theta-\theta_0)\langle\ (0, f-f_0+(\theta-\theta_0)\gamma_n)\ ,\ (0, \gamma-\gamma_n)\ \rangle_L. \end{aligned}$$

Due to **(E)**, $\|0, \gamma-\gamma_n\|_L$ tends to zero and $W_n(0, \gamma-\gamma_n)$ is a $o_P(1)$. Using Cauchy-Schwarz inequality and the fact that V_n is a ε_n-neighborhood of η_0, the crossed term in $Q_n(\theta, f)$ is bounded from above by $(\sqrt{n}\varepsilon_n\rho_n)h$. Thus due to **(E)**, the term $Q_n(\theta, f)$ is a $o_P(1+h^2)$ uniformly over the considered neighborhood. The same holds for the remainder terms difference due to **(N')**.

Changing variables according to Lemma 17, the nonparametric integral equals, up to a $\exp\{o_P(1+h^2)\}$ term,

$$\int \mathbf{1}\{\omega,\ |U\gamma_n(\omega)+(\theta-\theta_0)\|\gamma_n\|^2_{\mathbb{H}}| \le M\sqrt{n}\varepsilon_n\|\gamma_n\|_{\mathbb{H}}\}$$

$$\times\ \mathbf{1}\{g(\omega)\in\mathscr{F}_n(\theta),\ \|0,g(\omega)-f_0\|_L \le 2\varepsilon_n\}$$

$$\times\ \exp\left\{-n\|0,g-f_0\|_L^2/2+\sqrt{n}W_n(0,g-f_0)+R_n(\theta_0,g)\right\}$$

$$\times\ \exp\left\{(\theta-\theta_0)U\gamma_n(\omega)-(\theta-\theta_0)^2\|\gamma_n\|^2_{\mathbb{H}}/2\right\}\ dP_f(\omega).$$

Now using the fact that $\|\gamma_n\|_{\mathbb{H}} \lesssim \sqrt{n}\rho_n$ and that, on the considered set, due to **(E)**, $|U\gamma_n|(\omega) \lesssim (\sqrt{n}\varepsilon_n\rho_n)\sqrt{n} = o(\sqrt{n})$, we obtain that the last term in the preceding display is $\exp\left\{o_P(1+h^2)\right\}$. The upper bound we obtain for the nonparametric integral is, up to $\exp\{o_P(1+h^2)\}$,

$$\int \mathbf{1}\left\{\omega,\ |U\gamma_n(\omega)| \le 2M\sqrt{n}\varepsilon_n\|\gamma_n\|_{\mathbb{H}},\ g(\omega)\in\mathscr{F}_n(\theta),\ \|0,g(\omega)-f_0\|_L \le 2\varepsilon_n\right\}$$

$$\times\ \exp\left\{-n\|0,g-f_0\|_L^2/2+\sqrt{n}W_n(0,g-f_0)+R_n(\theta_0,g)\right\}dP_f(\omega),$$

that is the second term of the right-hand side of (15). Note that the expression inside the exponential is nothing but $\ell_n(\theta_0,g)-\ell_n(\theta_0,f_0)$. Thus we are almost in position to use the second part of $(\mathbf{C}')$ about $\Pi^{\theta=\theta_0}$, except for the extra indicator of $\{\omega,\ |U\gamma_n(\omega)| \le 2M\sqrt{n}\varepsilon_n\|\gamma_n\|_{\mathbb{H}}\}$. But the posterior probability of this event tends to one due to the assumption on $B^{\theta=\theta_0}_{KL,n}$ together with Lemma 2, thus one can delete this event without alterating the bound at stake for the ratio $s_1(B)/s_0$, as in the beginning of the proof. One concludes like in the proof of Theorem 1. □

4 Applications: proofs

4.1 Proof of Theorem 3

We first obtain the LAN-expansion for model (9) and introduce the norm $\|\cdot\|_L$. Then a testing distance d_T which verifies (4) is defined and related to $\|\cdot\|_L$. A sieve $\mathscr{A}_n$ can then be constructed such that (5)-(6) are verified, which enables one to obtain **(C)** for some rate ε_n. Finally we conclude by showing that **(N)** is fulfilled for our choices of ε_n and $\mathscr{F}_n$. For any real θ and any symmetric 1-periodic function f in $L^2[0,1]$, simple calculations reveal that the log-likelihood in model (9) can be expanded as follows,

$$\Lambda_n(\theta,f) = -\|h,a\|_L^2/2+W(h,a)+R_n(\theta,f),$$

with $\|h,a\|_L^2 = \int(h^2 f_0'(u)^2+a(u)^2)du$ and $W(h,a)=\int(-hf_0'(t-\theta_0)+a(t-\theta_0))dW(t)$, the integration domain being $[-1/2,1/2]$. The Hilbert space on which the inner product is defined is $\mathbb{R}\times\mathscr{G}_{\eta_0}$, with $\mathscr{G}_{\eta_0}$ the Hilbert space of even, square integrable functions on $(-1/2,1/2)$, extended by 1-periodicity. In particular, note that there is no information loss. Denoting

$$\Delta_n(t,h) \triangleq \sqrt{n}(f_0(t-\theta_0) - f_0(t-\theta_0+h/\sqrt{n})) + hf_0'(t-\theta_0),$$

the remainder term $R_n(\theta, f)$ of the expansion has the expression

$$R_n(\theta,f) = -\frac{1}{2}\int \Delta_n(t,h)^2 dt - \int \Delta_n(t,-h)dW(t) \tag{16}$$

$$-\int \Delta_n(t,h)\left[-hf_0'(t-\theta_0) + a(t-\theta_0)\right]dt \tag{17}$$

$$+\int \left[a(t-\theta_0-h/\sqrt{n}) - a(t-\theta_0)\right]dW(t). \tag{18}$$

For any $\eta = (\theta, f)$ and $\lambda = (\tau, g)$ in $\mathscr{E}$, let us introduce the distance d_T as

$$d_T^2(\eta,\lambda) = \int_{-1/2}^{1/2} (g(t-\tau) - f(t-\theta))^2 dt.$$

Lemma 3 *Equation (4) is satisfied with $d_n = e_n = d_T$ and $\xi = 1/4$.*

Proof Let us consider the test

$$\phi_n = \mathbf{1}\{2\int_{-1/2}^{1/2} \{f_1(t-\theta_1) - f_0(t-\theta_0)\} dX^{(n)}(t) > \|f_1\|^2 - \|f_0\|^2\}.$$

The verifications are then as in the proof of Lemma 5 in [12] and are omitted. □

Lemma 4 *Let f be a symmetric 1-periodic function in $L^2[0,1]$ and let g satisfy conditions* (**R**)*. There exist positive D_1, D_2, D_3 and $\mu > 0$ depending only on β, ρ, L, τ_0 such that, for all $\theta, \tau \in \Theta$,*

$$D_3(\theta-\tau)^2 \le d_T^2((\theta,f),(\tau,g)) \le D_2\left\{(\theta-\tau)^2 + \|f-g\|^2\right\} \tag{19}$$

$$\text{if } |\theta-\tau| \le \delta, \quad d_T^2((\theta,f),(\tau,g)) \ge D_1\left\{(\theta-\tau)^2 + \|f-g\|^2\right\}. \tag{20}$$

Proof First we prove (19). Denoting $\delta = \theta - \tau$, it holds

$$d_T^2((\theta,f),(\tau,g)) = \sum_{k\ge 0} (g_k\cos(2\pi k\delta) - f_k)^2 + g_k^2 \sin^2(2\pi k\delta).$$

Since $|\delta| \le 2\tau_0 < 1/2$, there exists a constant $d > 0$ depending on τ_0 only such that for any $|\delta| < 2\tau_0$, $|\sin(2\pi\delta)| \ge d|\delta|$. Since g satisfies (**R**), we have $g_1^2 \ge \rho^2$, thus the first inequality is obtained by setting $D_3 = d^2\rho^2$. Then note that

$$d_T^2((\theta,f),(\tau,g)) \le 2d_T^2((\theta,f),(\theta,g)) + 2d_T^2((\theta,g),(\tau,g))$$

$$\le 2\|f-g\|_2^2 + 8\sum_{k\ge 0} g_k^2 \sin^2(\pi k\delta) \le D_2(\|f-g\|_2^2 + \delta^2),$$

with $D_2 = 2 \vee 8\pi^2 L^2$, using **(R)**. In order to check (20), let us write

$$\begin{aligned} d_T^2\left((\theta,f),(\tau,g)\right) &= \sum_{k\geq 0} (g_k - f_k)^2 + 4 f_k g_k \sin^2(\pi k \delta) \\ &= \|f-g\|_2^2 + 4\sum_{k\geq 0}(f_k - g_k) g_k \sin^2(\pi k\delta) + 4\sum_{k\geq 0} g_k^2 \sin^2(\pi k \delta). \end{aligned}$$

Since g satisfies **(R)** and using the bound $|\sin(x)| \geq 2|x|/\pi$ valid for $|x| \leq \pi/2$, the last term is bounded from below by $4g_1^2\sin^2(\pi\delta) \geq 4\rho^2(2\delta)^2$. Using Cauchy-Schwarz inequality, conditions **(R)** and denoting $\nu = (\beta - 1)\wedge 1$,

$$\begin{aligned} &\left|\sum_{k\geq 0}(f_k - g_k) g_k \sin^2(\pi k\delta)\right| \leq \|f-g\|_2 \left(\sum_{k\geq 0} g_k^2 \sin^4(\pi k\delta)\right)^{1/2} \\ &\leq \|f-g\|_2 \left(\pi^{2\nu+2}\delta^{2\nu+2}\sum_{k\geq 0} k^{2\beta} g_k^2\right)^{1/2} \leq \pi^{\nu+1}\delta^\nu L(\|f-g\|_2^2 + \delta^2)/2, \end{aligned}$$

where we have used the inequality $2\|f-g\|_2\delta \leq \|f-g\|_2^2 + \delta^2$. Thus

$$d_T\left((\theta,f),(\tau,g)\right)^2 \geq \left\{1 - 2\pi^{\nu+1}\delta^\nu L\right\}\|f-g\|_2^2 + \left\{16\rho^2 - 2\pi^{\nu+1}\delta^\nu L\right\}\delta^2.$$

Let us choose δ such that $\delta^\nu \leq (1/2)(1\wedge 16\rho^2)/(2\pi^{\nu+1}L)$. Then it suffices to choose $D_1 = (1\wedge 16\rho^2)/2$ to obtain (20). □

A consequence of Lemma 4 is that (8) is fulfilled. Indeed, f_0 satisfies **(R)**, thus if $d_T(\eta,\eta_0) \leq \gamma_n$, due to (19) one has $(\theta-\theta_0)^2 \lesssim \gamma_n^2$ and if $\gamma_n \to 0$, one can apply (20).

4.1.1 Concentration result, translation model

In order to check conditions (5)-(7) for the prior (10), we first define the rate ε_n. According to Lemma 14, an appropriate choice is $\varepsilon_n = Dn^{-\alpha\wedge\beta/(2\alpha+1)}$ for D large enough. Then we define sieves $\mathscr{F}_n$ using the RKHS $\mathbb{H}^\alpha$ of the prior and the Hilbert spaces $\mathbb{B}^p$, both defined in Section 5.2. Given a sequence $\alpha_n \to 0$ and a real $1 \leq p < \alpha$, both to be specified later, let us define $\mathscr{A}_n = \Theta \times \mathscr{F}_n$ and

$$\mathscr{F}_n = \left\{\varepsilon_n \mathbb{B}_1^0 + \sqrt{10Cn}\varepsilon_n \mathbb{H}_1^\alpha\right\} \cap \left\{\alpha_n \mathbb{B}_1^p + \sqrt{n}\alpha_n \mathbb{H}_1^\alpha\right\}, \tag{21}$$

with C constant large enough. We make the following restriction on α_n,

$$(\mathbf{C1}) \quad \alpha_n \gtrsim \varepsilon_n \qquad \text{and} \qquad (\mathbf{C2}) \quad \alpha_n \gtrsim n^{-1/(2+\frac{1}{\alpha-p})},$$

so that the conclusion of Lemma 13 holds and $\exp(-n\alpha_n^2) \leq \exp(-Cn\varepsilon_n^2)$. Further restrictions on p, α_n arise when dealing with remainder terms in the LAN expansion.

Verification of (5). For any $\eta_1 = (\theta_1, f_1)$ and $\eta_2 = (\theta_2, f_2)$ in $\mathscr{A}_n$, by the same calculation as in the proof of Lemma 4,

$$d_T^2(\eta_1,\eta_2)/2 \leq \|f_1 - f_2\|^2 + 4\pi^2(\theta_1-\theta_2)^2\sum_{k\geq 1} k^2 f_{2,k}^2.$$

Since f_2 belongs to the set $\alpha_n \mathbb{B}_1^p + \sqrt{n}\alpha_n \mathbb{H}_1^\alpha \subset 2\sqrt{n}\alpha_n \mathbb{B}_1^p$ and $p > 1$, we have that $\sum_{k\geq 1} k^2 f_{2,k}^2 \lesssim n\alpha_n^2$. Thus $d_T(\eta_1,\eta_2)^2 \leq K(n\alpha_n^2(\theta_1-\theta_2)^2 + \|f_1 - f_2\|_2^2)$ for some positive constant K. Hence

$$N(2K\varepsilon_n, \mathscr{A}_n, d_T) \leq N\left(\frac{\varepsilon_n}{\sqrt{n}\alpha_n}, \Theta, |\cdot|\right) \times N\left(\varepsilon_n, \varepsilon_n \mathbb{B}_1^0 + \sqrt{10Cn}\varepsilon_n \mathbb{H}_1^\alpha, \|\cdot\|_2\right)$$

$$\lesssim (\sqrt{n}\alpha_n \varepsilon_n^{-1}) \exp(6Cn\varepsilon_n^2) \lesssim \exp(7Cn\varepsilon_n^2),$$

using Lemma 14, as long as $\log n = o(n\varepsilon_n^2)$, which is the case here.

Verification of (6). Due to the definition of $\mathscr{A}_n$,

$$\pi_f(\mathscr{E} \backslash \mathscr{A}_n) \leq \pi_f\left(f \notin \varepsilon_n \mathbb{B}_1^0 + \sqrt{10Cn}\varepsilon_n \mathbb{H}_1^\alpha\right) + \pi_f\left(f \notin \alpha_n \mathbb{B}_1^p + \sqrt{n}\alpha_n \mathbb{H}_1^\alpha\right).$$

Lemmas 13 and 14 and the conditions on α_n give us the desired inequality.

Verification of (7). Note that $B_{KL,n}$ neighborhoods are simply d_T-balls here. Thus (19) implies that

$$\Pi\left(B_{KL}(\eta_0, \sqrt{5D_2}\varepsilon_n)\right) \geq \pi_\theta\left((\theta-\theta_0)^2 < \varepsilon_n^2\right) \times \pi_f\left(\|f - f_0\|^2 \leq 4\varepsilon_n^2\right)$$

$$\gtrsim \varepsilon_n^{-1} \exp(-n\varepsilon_n^2) \gtrsim \exp(-2n\varepsilon_n^2),$$

using Lemma 14, which is enough to check (7).

Thus Lemma 1 gives that $\Pi(\eta \in \mathscr{E},\ d_T(\eta,\eta_0) < M\varepsilon_n | X^{(n)})$ tends to 1. Since (8) holds, this is also true in terms of $\|\cdot\|_L$. An application of Lemma 2 leads to the fact that $\Pi(\eta \in \mathscr{A}_n,\ \|\eta - \eta_0\|_L < M\varepsilon_n | X^{(n)}) \to 1$ and the first part of assumption **(C)** is established. The concentration result about $\Pi^{\theta=\theta_0}$ follows by the same techniques, in an even simpler way since no θ is involved, which concludes the verification of **(C)**.

4.1.2 Translation model, LAN-type conditions

Let us check **(N)** by proving that each of the terms composing $R_n(\theta,f)/(1+h^2)$ is uniformly a $o_P(1)$ over V_n. The next Lemma deals with the term (16).

Lemma 5 *As $n \to +\infty$ it holds*

$$\sup_{(\theta,f)\in V_n} \int_{-1/2}^{1/2} \frac{\Delta_n(t,h)^2}{1+h^2} dt = o(1) \quad \textit{and} \quad \sup_{(\theta,f)\in V_n} \int_{-1/2}^{1/2} \frac{\Delta_n(t,-h)}{1+h^2} dW(t) = o_P(1).$$

Proof Let us denote $\delta = (\theta - \theta_0) = h/\sqrt{n}$, then using the Fourier series expansion of f_0 and f_0', one obtains

$$\int_0^1 \Delta_n(t,h)^2(u)du = n \sum_{k\geq 1} f_{0,k}^2 \{\cos(2\pi k\delta) - 1\}^2 + f_{0,k}^2\{2\pi k\delta - \sin(2\pi k\delta)\}^2.$$

Using that $0 \le x - \sin(x) \le 2(x \wedge \frac{x^3}{6})$ for any $x > 0$, denoting $\nu = (\beta - 1) \wedge 1$,

$$\int_0^1 \Delta_n(t,h)^2(u)du \lesssim n \sum_{k\ge 1} f_{0,k}^2 \left[\sin^4(\pi k\delta) + k^6\delta^6 \mathbf{1}_{k|\delta|\le 1} + k^2\delta^2 \mathbf{1}_{k|\delta|>1}\right]$$
$$\lesssim n \sum_{k\ge 1} f_{0,k}^2 k^{2+2\nu} \left[\delta^{2+2\nu} \sin^{2-2\nu}(\pi k\delta) + \delta^{2+2\nu}(k\delta)^{4-2\nu} \mathbf{1}_{k|\delta|\le 1} + \delta^2 k^{-2\nu} \mathbf{1}_{k|\delta|>1}\right]$$
$$\lesssim n\delta^{2+2\nu} \sum_{k\ge 1} f_{0,k}^2 k^{2+2\nu}.$$

Since on V_n, we have $|h| \lesssim \varepsilon_n$, using **(R)** the last display can be further bounded from by $n^{-\nu} h^{2+2\nu} \lesssim \varepsilon_n^{2\nu} h^2 = o(h^2)$, hence the first statement. The proof of the second result is not difficult using Lemma 12 and the first statement and is omitted. □

Lemma 6 *For any $\beta \ge 2$, $|(17)|/(1+h^2)$ tends to 0 uniformly over V_n. This also holds for $1 < \beta < 2$ as soon as $\varepsilon_n = o(n^{-1+\frac{\beta}{2}})$ as $n \to +\infty$.*

Proof Cauchy-Schwarz inequality, the fact that $\sqrt{x+y} \le \sqrt{x} + \sqrt{y}$ for all positive x and y, and Lemma 5 together imply that on V_n,

$$(17) \le \left(\int \Delta_n(t,h)^2 dt\right)^{1/2} \left(\int [-hf_0'(t-\theta_0) + a(t-\theta_0)]^2 dt\right)^{1/2}$$
$$\lesssim n^{-\nu/2} h^{1+\nu} \left(h^2 + \|a\|_2^2\right)^{1/2} \lesssim \varepsilon_n^{\nu} h^2 + n^{-\nu/2} h^{1+\nu} \|a\|_2.$$

If $\beta \ge 2$, using that (θ, f) lies in V_n, the last display is bounded by $\varepsilon_n h^2 = o(h^2)$, whereas if $1 < \beta < 2$, we get the bound $n^{1-\beta/2} h^{1+\nu} \varepsilon_n$, which is $o(1+h^2)$ as soon as $\varepsilon_n = o(n^{-1+\beta/2})$. □

Finally let us focus on (18). By definition of $\mathscr{F}_n$, we can set $f = \alpha_n g_n + h_n$, with $g_n \in \mathbb{B}_1^p$ and $h_n \in \sqrt{n}\alpha_n \mathbb{H}_1^{\alpha}$. Thus

$$\frac{(18)}{1+h^2} = \frac{\sqrt{n}\alpha_n}{1+\sqrt{n}(\theta-\theta_0)} \int_{-1/2}^{1/2} (g_n(t-\theta) - g_n(t-\theta_0)) dW(t) \tag{22}$$
$$+ \frac{\sqrt{n}}{1+\sqrt{n}(\theta-\theta_0)} \int_{-1/2}^{1/2} ((h_n - f_0)(t-\theta) - (h_n - f_0)(t-\theta_0)) dW(t). \tag{23}$$

To bound these terms we use Lemma 12 applied to processes of the form $\int \phi_{n,\theta,f}(u)dW(u)$. The entropies involved are bounded using Lemma 15 as follows.

Bounding (22). Let us write (22)$= \int \varphi_n(u)dW(u)$ and see this process as indexed by L^2-functions φ_n. Thus the distance d in Lemma 12 is simply the L^2-norm and δ^2 can be chosen proportional to an upper bound on the quantities $\int \varphi_n^2$. Note that

$$g_n(t-\theta) - g_n(t-\theta_0) = 2\sum_{k\ge 1} \sin(\pi k(\theta-\theta_0)) \sin(2\pi kt - \pi k(\theta+\theta_0)) g_{n,k}.$$

Since $g_n \in \mathbb{B}_1^p$ and $p>1$, it holds $\int \varphi_n^2 \lesssim \alpha_n^2 \sum_{k\geq 1} k^2 g_{n,k}^2 \lesssim \alpha_n^2$ and, denoting by $\varphi_{n,k}$ the Fourier coefficients of φ_n, we have $\sum_{k\geq 1} k^{2p}\varphi_{n,k}^2 \lesssim n\alpha_n^2 \sum_{k\geq 1} k^{2p} g_{n,k}^2 \lesssim n\alpha_n^2$. This means that all φ_n's belong to $\sqrt{n}\alpha_n \mathbb{B}_1^p$. Thus

$$\int_0^{\alpha_n} \sqrt{\log N(\eta, \sqrt{n}\alpha_n \mathbb{B}_1^p, \|\cdot\|_2)}d\eta \lesssim \int_0^{\alpha_n} \left(\sqrt{n}\alpha_n/\eta\right)^{1/2p} d\eta \lesssim \alpha_n n^{1/4p}.$$

This leads to the condition $\quad$ **(C3)** $\quad \alpha_n = o(n^{-1/4p})$.

Bounding (23). By similar arguments as for (22), one obtains that (23) is uniformly a $o_P(1)$ over V_n as soon as

$$\textbf{(C4)} \quad \alpha_n = o(n^{-3/2(2\alpha+1)}).$$

We include the detailed calculations for completeness.

Let us write (23)$= \int \psi_n(u)dW(u)$ and see this process as indexed by the functions ψ_n and apply Lemma 12. The distance d is the L^2-norm and the parameter δ^2 can still be chosen proportional to an upper bound on the variances $\int \psi_{n,\theta,f}(u)^2 du$. First we wish to bound from above these variances. Note that due to **(C1)** we have the inequalities $\|h_n - f_0\|_2^2 \leq \|f-f_0\|_2^2 + \alpha_n^2\|g_n\|_2^2 \leq \varepsilon_n^2 + \alpha_n^2 \leq \alpha_n^2$, thus

$$\begin{aligned}
&\frac{n}{1+n(\theta-\theta_0)^2}\int_{-1/2}^{1/2}\{(h_n-f_0)(t-\theta)-(h_n-f_0)(t-\theta_0)\}^2 dt\\
&\lesssim \frac{n}{1+n(\theta-\theta_0)^2}\sum_{k\geq 1}\sin^2(\pi k(\theta-\theta_0))(h_{n,k}-f_{0,k})^2 \lesssim \sum_{k\geq 1} k^2 (h_{n,k}-f_{0,k})^2.
\end{aligned}$$

Let us denote $2\chi = 2\beta \wedge (2\alpha+1)$ and $r_n = \sqrt{n}\alpha_n$. Hölder inequality implies

$$\sum_{k\geq 1} k^2 (h_{n,k}-f_{0,k})^2 \leq \left\{\sum_{k\geq 1}(h_{n,k}-f_{0,k})^2\right\}^{1-1/\chi}\left\{\sum_{k\geq 1}k^{2\chi}(h_{n,k}-f_{0,k})^2\right\}^{1/\chi}.$$

If $2\beta > 2\alpha+1$, then $\sum_{k\geq 1} k^{2\chi}(h_{n,k}-f_{0,k})^2$ is bounded from above by Cr_n^2 using the fact that h_n belongs to $r_n \mathbb{H}_1^\alpha$. We obtain

$$\sum_{k\geq 1} k^2 (h_{n,k}-f_{0,k})^2 \lesssim \alpha_n^{2-2/\chi}\zeta_n^{2/\chi} \leq \alpha_n^2 n^{2/(1+2\alpha)}.$$

If $2\beta \leq 2\alpha+1$ then $\chi = \beta$ and for any integer $K_n > 0$ it holds

$$\begin{aligned}
&\sum_{k\geq 1} k^{2\beta}(h_{n,k}-f_{0,k})^2\\
&\lesssim K_n^{2\beta}\sum_{1\leq k\leq K_n}(h_{n,k}-f_{0,k})^2 + \sum_{k\geq 1}k^{2\beta}f_{0,k}^2 + K_n^{-1-2\alpha+2\beta}\sum_{k>K_n}k^{1+2\alpha}h_{n,k}^2\\
&\lesssim K_n^{2\beta}\alpha_n^2 + L^2 + r_n^2 K_n^{-1-2\alpha+2\beta} \;\lesssim\; 1 \vee \alpha_n^2 n^{2\beta/(1+2\alpha)},
\end{aligned}$$

where we obtained the last inequality by optimizing in K_n. Since $\varepsilon_n = Dn^{-\alpha\wedge\beta/(2\alpha+1)}$, using **(C1)**, we have that $\alpha_n^2 n^{2\beta/(1+2\alpha)}$ is always larger than some positive constant, thus $1 \vee \alpha_n^2 n^{2\beta/(1+2\alpha)}$ reduces to the second term. Thus in this case

$$\sum_{k\geq 1} k^2 (h_{n,k} - f_{0,k})^2 \lesssim \alpha_n^2 n^{2/(1+2\alpha)}.$$

Hence in all cases, the obtained bound on the variances is $\chi_n^2 = \alpha_n^2 n^{2/(1+2\alpha)}$. Now let us show that the ψ_n's are in a set whose entropy is well-controlled.

$$\begin{aligned}\psi_n(\cdot) &= \frac{\sqrt{n}}{1+\sqrt{n}(\theta-\theta_0)}\{h_n(\cdot-\theta) - h_n(\cdot-\theta_0)\}\\ &\quad - \frac{\sqrt{n}}{1+\sqrt{n}(\theta-\theta_0)}\{f_0(\cdot-\theta) - f_0(\cdot-\theta_0)\} = H_{n,\theta,h_n}(\cdot) - F_{0,\theta}(\cdot).\end{aligned}$$

We deal separately which each of these two terms. First,

$$\begin{aligned}&\int_{-1/2}^{1/2} (F_{0,\theta_1}(u) - F_{0,\theta_2}(u))^2 du\\ &\leq 2\int_{-1/2}^{1/2}\left[\frac{\sqrt{n}}{1+\sqrt{n}(\theta_1-\theta_0)} - \frac{\sqrt{n}}{1+\sqrt{n}(\theta_2-\theta_0)}\right]^2 f_0(u-\theta_0)^2 du\\ &\quad +2\int_{-1/2}^{1/2}\left[\frac{\sqrt{n}}{1+\sqrt{n}(\theta_1-\theta_0)} f_0(u-\theta_1) - \frac{\sqrt{n}}{1+\sqrt{n}(\theta_2-\theta_0)} f_0(u-\theta_2)\right]^2 du\\ &\leq 6n^2(\theta_1-\theta_2)^2\|f_0\|^2 + \left[\frac{2\sqrt{n}}{1+\sqrt{n}(\theta_1-\theta_0)}\right]^2 \int_0^1 (f_0(u-\theta_1) - f_0(u-\theta_2))^2 du\\ &\leq 6n^2(\theta_1-\theta_2)^2\|f_0\|^2 + 4n(4\pi^2)(\theta_1-\theta_2)^2 \sum_{k\geq 1} k^2 f_{0,k}^2 \lesssim n^2(\theta_1-\theta_2)^2.\end{aligned}$$

Thus for a universal constant C (independent of η),

$$N(\eta, \{F_{0,\theta},\ \theta\in\Theta\}, \|\cdot\|_2) \leq Cn/\eta.$$

Now let us cover the set of functions $\{H_{n,\theta,h_n}\}$ by L^2-balls. If q_k denote the Fourier coefficients of any function in this set, one easily sees that $|q_k| \leq |k||h_{n,k}|$, thus $\sum_{k\geq 1} k^{2\alpha-1} q_k^2 \leq r_n^2$, which means that all functions H_{n,θ,h_n} belong to $r_n \cdot \mathbb{H}_1^{\alpha-1}$. Due to Lemma 15, there exists a universal constant D such that, for n large enough, for any $\eta > 0$,

$$N(\eta, r_n\cdot\mathbb{H}_1^{\alpha-1}, \|\cdot\|_2) \leq \exp\{D(r_n/\eta)^{\frac{2}{2\alpha-1}}\}.$$

Combining the two preceding results, there exists a universal constant D_1 such that

$$\begin{aligned}&N(\eta, \{F_{0,\theta} - H_{n,\theta,h_n}\}, \|\cdot\|_2)\\ &\quad\leq N\left(\frac{\eta}{2}, \{F_{0,\theta}\}, \|\cdot\|_2\right) \times N\left(\frac{\eta}{2}, \{H_{n,\theta,h_n}\}, \|\cdot\|_2\right) \lesssim \frac{n}{\eta} e^{D_1 \eta^{-\frac{2}{2\alpha-1}}}.\end{aligned}$$

If χ_n^2 is the bound on the variances obtained above, due to Lemma 12,

$$\begin{aligned}
&\int_0^{\chi_n} \sqrt{\log N(\eta, \{F_{0,\theta} - H_{n,\theta,h_n}\}, \|\cdot\|_2)} d\eta \\
\lesssim &\qquad \int_0^{\chi_n} \log^{1/2}(n/\eta)\, d\eta + \int_0^{\chi_n} (r_n/\eta)^{\frac{1}{2\alpha-1}}\, d\eta. \\
\lesssim &\qquad \chi_n \log^{1/2}(n/\chi_n) + \chi_n + r_n^{\frac{1}{2\alpha-1}} \chi_n^{1-\frac{1}{2\alpha-1}}
\end{aligned}$$

As long as $\chi_n \log n \to 0$, the first two terms tend to zero as $n \to +\infty$, while the third one leads to the condition $r_n^{\frac{1}{2\alpha-1}} \chi_n^{1-\frac{1}{2\alpha-1}} = o(1)$. That is, $\alpha_n n^{\frac{3}{2(2\alpha+1)}} = o(1)$. □

4.1.3 Conclusion

Let us recall that we have chosen $\varepsilon_n = Dn^{-\alpha\wedge\beta/(2\alpha+1)}$. Simple verifications reveal, setting $p = 1 \vee (\alpha/2)$, that for any α, β such that $\alpha > 1 + \sqrt{3}/2$ and $\beta > 3/2$, it is possible to find a sequence $\alpha_n \to 0$ such that **(C1)-(C4)** are satisfied. Lemma 6 gives another condition in the case $\beta < 2$, where one also needs $\varepsilon_n = o(n^{-1+\beta/2})$. This imposes, if $\alpha \geq \beta$, that $\alpha < (3\beta - 2)/(4 - 2\beta)$, whereas if $\alpha < \beta$ the condition is trivial. Putting these conditions together gives us the area presented in Figure 1 and concludes the proof of Theorem 3 for the prior π_f^α.

4.1.4 Case of the prior $\pi_{f,k(n)}^\alpha$

The proof for this prior is similar, though easier. We explain it briefly. The concentration step **(C)** can be done following the methods of [29], leading to a concentration rate $\varepsilon_n \approx n^{-\frac{\alpha\wedge\beta}{2\alpha+1}}$ if one chooses $k(n) = \lfloor n^{1/(2\alpha+1)} \rfloor$. The step **(N)** is similar to the case of prior (10). The control of the third term of the LAN expansion is simpler, since the sum defining the prior is finite, and it leads to the condition $k(n)^3 \varepsilon_n^2 \to 0$ as $n \to +\infty$, that is $\alpha \wedge \beta > 3/2$, which concludes the proof.

4.2 Functional data analysis model

Here we shall only give the LAN expansion in model (11)-(12) and check **(E)**. Let $\mathscr{A}$ be the set of all square integrable functions on $[0,1]$ extended to $\mathbb{R}$ by 1-periodicity. Let us define the inner product $\langle \cdot, \cdot \rangle_L$ on $\mathbb{R} \times \mathscr{A}$ by

$$\|(h_1, a_1), (h_2, a_2)\|_L^2 = \langle a_1, a_2 \rangle_2 + \langle a_1 - h_1 f_0', a_2 - h_2 f_0' \rangle_2,$$

with $\langle \cdot, \cdot \rangle_2$ the usual inner product on $L^2(\mathbb{R})$. For any $(h, a) \in \mathbb{R} \times \mathscr{A}$, we denote

$$W_n((h,a)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[a(t_i)\varepsilon_i + (a(t_i - \theta_0) - h f_0'(t_i - \theta_0))\zeta_i \right].$$

Note that for any $d \geq 1$ and any fixed $v_1, \ldots, v_d$, the variable $W_n(v_1, \ldots, v_d)$ converges in distribution towards a Gaussian variable of covariance structure $(\langle v_i, v_j \rangle_L)_{1 \leq i,j \leq d}$. Direct calculations lead to express $\Lambda_n(\theta, f) = \ell_n(\theta, f) - \ell_n(\theta_0, f_0)$ as

$$\Lambda_n(\theta, f) = -\frac{n}{2} \|\theta - \theta_0, f - f_0\|_L^2 + \sqrt{n} W_n(\theta - \theta_0, f - f_0) + R_n(\theta, f),$$

where the remainder term $R_n(\theta, f)$ can be written in a similar way as for model (9), with additional terms accounting for the approximation of integrals by their discrete counterparts on the design points (t_i).

From the structure of the inner product $\langle \cdot, \cdot \rangle_L$ on the Hilbert space $\mathbb{R} \times \mathscr{A}$ one easily deduces that the least favorable direction and the efficient Fisher information are respectively given by $\gamma = -f_0'/2$ and $\tilde{I}_{\eta_0} = \|(1, -\gamma)\|_L^2 = \|f_0'\|_2^2/2$. In particular, there is a loss of information. The RKHS of the prior (13), again denoted $\mathbb{H}^\alpha$, is the space of functions spanned by the first $2k(n)+1$ elements of the Fourier basis equipped with the norm $\langle f, g \rangle_{\mathbb{H}^\alpha} = \sum_{k=0}^{2k(n)} \sigma_k^{-2} f_k g_k$, see [30], Theorem 4.1. This space enables to approximate the least favorable direction $\gamma = -f_0'/2$ by a truncated version of it. Denoting by γ_k the real Fourier coefficients of γ, we set $\gamma_n(\cdot) = \sum_{k=0}^{2k(n)} \gamma_k \varepsilon_k(\cdot)$. Since f_0 is $\mathscr{C}^1$, the complex Fourier coefficients of f_0' are given by $(2i\pi k)c_k$. Thus

$$\|\gamma_n\|_{\mathbb{H}}^2 \lesssim \sum_{k=0}^{k(n)} k^{3+2\alpha} |c_k|^2 \lesssim k(n)^{(2\alpha+3-2\beta)\vee 0}$$

$$\|0, \gamma_n - \gamma\|_L^2 \lesssim \sum_{k > k(n)} k^2 |c_k|^2 \lesssim k(n)^{2-2\beta}.$$

In view of the preceding and of (3) , we define $\rho_n \triangleq k(n)^{1-\beta} \vee n^{-1/2}$. Now we can check **(E)** and the condition $\sqrt{n}\varepsilon_n \rho_n \to 0$. Due to the expression of the rate $\varepsilon_n = n^{-\frac{\alpha \wedge \beta}{2\alpha+1}}$, obtained in the same way as for model (9), and the expressions of ρ_n and $k(n) = \lfloor n^{1/(2\alpha+1)} \rfloor$, we have three cases

- if $2\alpha < 2\beta - 3$, then $\rho_n = n^{-1/2}$ and the condition is satisfied since $\varepsilon_n \to 0$.
- if $2\beta - 3 < 2\alpha < 2\beta$, the condition becomes $\beta > 3/2$.
- if $\alpha \geq \beta$, the condition is $\alpha < 2\beta - 3/2$.

Thus we are left with the conditions $\alpha < 2\beta - 3/2$ and $\beta > 3/2$. One then checks that $W_n(0, \gamma - \gamma_n)$ is a $o_P(1)$ using its explicit expression, which leads to **(E)**. We omit the verifications for **(N')**, which lead to the condition $\alpha \wedge \beta > 3/2$.

4.3 Proof of Theorem 5

We start by introducing the LAN expansion and $\|\cdot\|_L$. Testing can be done using Hellinger's distance h and Lemmas 7-8 enable to relate h to other metrics. The concentration of the posterior associated to the prior (14) is first obtained in terms of h at a rate ε_n. Then we obtain **(C')** in terms of $\|\cdot\|_L$ at a slightly slower rate ζ_n and finally check **(N')** and **(E)**. Throughout this section, we work with the local parameters $h = \sqrt{n}(\theta - \theta_0)$ and $a = \sqrt{n}(r - r_0)$. For any $u \in [0, \tau]$ and any integer i, let

$$M_i(u) = \mathbf{E}_{\eta_0}(\mathbf{1}_{u \leq T} Z^i e^{\theta_0 Z}) \quad \text{and} \quad M_0(\theta)(u) = \mathbf{E}_{\eta_0}(\mathbf{1}_{u \leq T} e^{\theta Z}).$$

As simple calculations reveal, the log-likelihood in Cox's model can be expanded as follows $\Lambda_n(\theta,r) = -\|h,a\|_L^2/2 + W_n(h,a) + R_n(\theta,r)$, with

$$\|h,a\|_L^2 = \int_0^\tau \{h^2 M_2(u) + 2ha(u)M_1(u) + a(u)^2 M_0(u)\} d\Lambda_0(u)$$

$$W_n(h,a) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\delta_i \{hZ_i + a(T_i)\} - e^{\theta_0 Z_i} \{hZ_i \Lambda_0(T_i) + (\Lambda_0 a)(T_i)\} \right],$$

the remainder term $R_n(\theta,r)$ being studied in Section 4.3.2. From the expression of the norm, one deduces that the least favorable direction γ (the projection of the vector $(1,0)$ onto the nuisance space) is $u \to M_1(u)/M_0(u)$. Explicitly, for any $u \in [0,\tau]$,

$$\gamma(u) = \frac{M_1}{M_0}(u) = \frac{\int_0^\tau (1-G_z(u-)) z e^{\theta_0 z - \Lambda_0(u) e^{\theta_0 z}} \varphi(z) dz}{\int_0^\tau (1-G_z(u-)) e^{\theta_0 z - \Lambda_0(u) e^{\theta_0 z}} \varphi(z) dz}. \tag{24}$$

As noted in Section 1.6, for independent identically distributed observations, tests satisfying (4) exist if one chooses $d_n = e_n$ equal to Hellinger's distance h and $\xi = 1/2$, $K = 1/8$, see [13], Section 7. Let us denote $P = P_\eta = P_{\theta,r}$ and $P_0 = P_{\eta_0} = P_{\theta_0,r_0}$, then

$$\begin{aligned} h(P,P_0)^2 = & \iint \Big[\{ e^{(r(t)+\theta z - \Lambda(t) e^{\theta z})/2} - e^{(r_0(t)+\theta_0 z - \Lambda_0(t) e^{\theta_0 z})/2} \}^2 (1-G_z(t-)) \\ & + \left\{ e^{-\Lambda(t) e^{\theta z}/2} - e^{-\Lambda_0(t) e^{\theta_0 z}/2} \right\}^2 g_z(t) \Big] dt \varphi(z) dz \\ & + \int \left\{ \Lambda e^{\theta z} - \Lambda_0 e^{\theta_0 z} \right\} (\tau)(1-G_z(\tau-)) \varphi(z) dz \\ = & \, h_1(P,P_0)^2 + h_2(P,P_0)^2 + h_3(P,P_0)^2. \end{aligned}$$

The Kullback-Leibler divergence $KL(P_0,P)$, the quantity $V(P_0,P)$ and the second moment $W(P_0,P) = \int \log^2(p_0/p) p_0 d\mu$ can all be expressed in a similar way.

Lemma 7 *Fix $\lambda_1,\lambda_2,\theta_1,\theta_2$ and let P_1,P_2 be the distributions associated to (θ_1,r_1) and (θ_2,r_2) respectively. Let us assume that there exists a constant $0 < Q \le 1/4$ such that $\|r_1 - r_2\|_\infty + M|\theta_1 - \theta_2| \le Q$. Then there exists a constant c depending on τ, M only such that*

$$h^2(P_1,P_2) \le cQ^2 e^{2Q}.$$

Proof For all u in $[0,\tau]$, using the inequality $|e^x - 1| \le |x| e^{|x|}$ valid for all real x,

$$\left| 1 - e^{(r_2 - r_1)(u) + (\theta_2 - \theta_1)z} \right| \le |(r_2 - r_1)(u) + (\theta_2 - \theta_1)z| e^Q \le Q e^Q.$$

Thus, for all t in $[0,\tau]$, $|\Lambda_1(t) e^{\theta_1 z} - \Lambda_2(t) e^{\theta_2 z}|$ is bounded from above by

$$\int_0^t \left| 1 - e^{(r_2 - r_1)(u) + (\theta_2 - \theta_1)z} \right| e^{\theta_1 z} d\Lambda_1(u) \le Q e^Q \Lambda_1(t) e^{\theta_1 z}.$$

Hence using the inequality $|1 - e^y| \le e^{|y|} - 1$ valid for all real y,

$$\begin{aligned} & \left\{ \sqrt{e}^{r_1 + \theta_1 z - \Lambda_1 e^{\theta_1 z}} - \sqrt{e}^{r_2 + \theta_2 z - \Lambda_2 e^{\theta_2 z}} \right\}^2 \le \lambda_1 e^{\theta_1 z - \Lambda_1 e^{\theta_1 z}} \left\{ \sqrt{e}^{Q + Q e^Q \Lambda_1 e^{\theta_1 z}} - 1 \right\}^2 \\ & \lesssim \lambda_1 e^{\theta_1 z - \Lambda_1 e^{\theta_1 z}} (\sqrt{e}^Q - 1)^2 e^{Q e^Q \Lambda_1 e^{\theta_1 z}} + \lambda_1 e^{\theta_1 z} \left\{ \sqrt{e}^{-\Lambda_1 e^{\theta_1 z}} - \sqrt{e}^{-(1 - Q e^Q)\Lambda_1 e^{\theta_1 z}} \right\}^2. \end{aligned}$$

The mean value theorem applied to the function $s \to e^{-s\Lambda_1(t)e^{\theta_1 z}/2}$ ensures the existence of a real ζ_t such that $1 - Qe^Q \leq \zeta_t \leq 1$ and

$$\left| e^{-\Lambda_1 e^{\theta_1 z}/2} - e^{-(1-Qe^Q)\Lambda_1 e^{\theta_1 z}/2} \right| = Qe^Q \Lambda_1(t) e^{\theta_1 z} e^{-\zeta_t \Lambda_1 e^{\theta_1 z}/2}/2$$
$$\lesssim Qe^Q \Lambda_1(t) e^{\theta_1 z} e^{-(1-Qe^Q)\Lambda_1 e^{\theta_1 z}/2}.$$

Since $Qe^Q \leq 1/2$, using the inequality $0 \leq xe^{-x/4} \lesssim e^{-x/8}$ for positive x, we obtain

$$h_1^2(P_1,P_2) \lesssim Q^2 e^{2Q} \iint \lambda_1(t) e^{\theta_1 z} e^{-\Lambda_1(t)e^{\theta_1 z}/4} dt \varphi(z) dz$$
$$\lesssim \int Q^2 e^{2Q} \left[-4e^{-\Lambda_1(t)e^{\theta_1 z}/4} \right]_0^\tau \varphi(z) dz \lesssim Q^2 e^{2Q}.$$

The same bound is obtained for $h_2^2(P_1,P_2)$ and $h_3^2(P_1,P_2)$ in a similar way. □

Lemma 8 *Suppose that $|\theta_0 - \theta| + \|r - r_0\|_\infty$ is bounded in $\mathbb{R}$. Then*

$$K(P_0,P) = P_0 \log p_0/p \lesssim h^2(p_0,p) \quad \textit{and} \quad V(P_0,P) \leq P_0 \log^2 p_0/p \lesssim h^2(p_0,p).$$

Proof First note that

$$\log p_0/p = \{\log \lambda_0/\lambda + (\theta_0 - \theta)z - \Lambda_0 e^{\theta_0 z} + \Lambda e^{\theta z}\} \mathbf{1}_{d=1,t<\tau}$$
$$+ \{\Lambda e^{\theta z} - \Lambda_0 e^{\theta_0 z}\} \mathbf{1}_{d=0,t<\tau} + \{\Lambda(\tau) e^{\theta z} - \Lambda_0(\tau) e^{\theta_0 z}\} \mathbf{1}_{d=0,t=\tau}.$$

By assumption $M|\theta_0 - \theta| + \|\log \lambda_0/\lambda\|_\infty$ is bounded by some constant Q_1. Then using the same technique as in the proof of Lemma 7, one obtains

$$\log \|p_0/p\|_\infty \leq \|\log(p_0/p)\|_\infty \leq Q_1 + e^{\theta_0 M} \|\Lambda_0\|_\infty Q_1 e^{Q_1}.$$

An application of Lemma 8 in [11] gives the result. □

4.3.1 Semi-parametric concentration

For any $\alpha > 0$, let $\mathscr{H}^\alpha$ be the RKHS of the prior (14), seen as a random element in $\mathbb{B} = (\mathscr{C}^0([0,\tau]), \|\cdot\|_\infty)$. Since $\log \lambda_0 = r_0$ is β-Hölder, it follows from the results stated in Section 5 that the equation $\varphi_{r_0}(\varepsilon_n) \leq n\varepsilon_n^2$ is solved for ε_n equal to $n^{-\alpha\wedge\beta/(2\alpha+1)}$, possibly up to some log-factor that we omit to write in the sequel. Moreover, Equations (26) and (27) - (29) are then satisfied and (28) is fulfilled with $B_n = \varepsilon_n \mathbb{B}_1 + \sqrt{10Cn}\varepsilon_n \mathscr{H}_1^\alpha$, where C is a large enough constant.

Let B denote the Hölder regularity of the least favorable direction γ given by (24). Due to Lemma 16, if, up to a log-factor, we choose $\rho_n = n^{-\alpha\wedge B/(2\alpha+1)}$, then $\varphi_\gamma(\rho_n) \leq n\rho_n^2$. Hence there exists a sequence γ_n in $\mathscr{H}^\alpha$ such that

$$\|\gamma_n - \gamma\|_\infty \leq \rho_n \quad \text{and} \quad \|\gamma_n\|_{\mathscr{H}^\alpha}^2 \leq 2n\rho_n^2,$$

Since $\|0, h - \gamma\|_L \lesssim \|h - \gamma\|_\infty$, the preceding display also holds with the $\|\cdot\|_\infty$-norm replaced by the $\|\cdot\|_L$-norm, which implies (3). Note also that $\|\gamma_n\|_\infty$ is bounded.

Now let us check (5)-(7). Lemmas 7 and 8 imply

$$\begin{aligned}&\{\theta:\ |\theta-\theta_0|\le\varepsilon_n\}\times\{r:\ \|r-r_0\|_\infty\le\varepsilon_n\}\\ &\quad\subset\{(\theta,r):\ h(p,p_0)\lesssim\varepsilon_n,\ \|\log p_0/p\|_\infty\lesssim\varepsilon_n\}\\ &\quad\subset\left\{(\theta,r):\ K(p_0,p)\lesssim\varepsilon_n^2,\ V(p_0,p)\lesssim\varepsilon_n^2\right\}.\end{aligned}$$

Due to (27), $\Pi(B_{KL,n}(\eta_0,c\varepsilon_n))\gtrsim\varepsilon_n^{-1}\exp(-n\varepsilon_n^2)\gtrsim\exp(-2n\varepsilon_n^2)$ for some $c>0$, which gives (7). Equation (28) implies (6) for the sieve $\mathscr{A}_n'=\Theta\times B_n$. Lemma 7 implies that ε_n-balls for the metric induced by $|\theta_1-\theta_2|+\|r_1-r_2\|_\infty$ are included in $C\varepsilon_n$-Hellinger balls for a universal constant C. Thus to obtain (5) it suffices to cover $\mathscr{A}_n'$ with balls for the first-mentioned metric and to use (27). Then Lemma 1 yields, for M large enough, in $P_{\eta_0}^{(n)}$-probability,

$$\Pi\left((\theta,r)\in\Theta\times B_n,\ h(P,P_0)\le M\varepsilon_n|X^{(n)}\right)\to 1.$$

Now let us translate this result in terms of the norm $\|\cdot\|_L$ while also slightly modifying the sieve. Let us denote $\zeta_n=C\sqrt{n}\varepsilon_n^2$ with C large enough and $\mathscr{A}_n=\Theta\times\mathscr{F}_n$, with

$$\mathscr{F}_n=B_n\cap\{r:\ \int_0^\tau e^{r(u)}du\le C_5,\ \int_0^\tau e^{r(u)}(r-r_0)^2(u)du\le C_6\zeta_n^2\},$$

for C_5,C_6 large enough contants. It follows from Lemmas 9, 10 and 11 below that

$$\Pi\left((\theta,r)\in\Theta\times\mathscr{F}_n,\ \|\theta-\theta_0,r-r_0\|_L\le\zeta_n|X^{(n)}\right)\to 1.$$

So, our definitive choice of sieve is $\mathscr{A}_n=\Theta\times\mathscr{F}_n$ and our definitive rate is ζ_n.

To check the second part of **(C')**, one has to deal with $\mathscr{F}_n(\theta)=\mathscr{F}_n+(\theta-\theta_0)\gamma_n$. Note that those sets are, up to constants, of the same type as $\mathscr{F}_n$, since on the set $|\theta-\theta_0|\lesssim\varepsilon_n$, the norm $\|(\theta-\theta_0)\gamma_n\|_{\mathscr{H}^\alpha}\lesssim\varepsilon_n\sqrt{n}\zeta_n=o(\sqrt{n}\varepsilon_n)$. The concentration condition on $\Pi^{\theta=\theta_0}$ easily follows. Finally, the conditions on Kullback-Leibler neighborhoods are checked as above.

Before giving the proof of Lemmas 9-11, let us summarize what we have obtained up to now. The posterior has been seen to converge at rate ε_n for Hellinger's distance and at rate ζ_n for the LAN-distance. In particular, condition **(C')** is satisfied with the rate ζ_n. Below, in checking **(N')**, a mild condition on ζ_n will appear, namely $\zeta_n=o(n^{-1/4})$, necessary to control the remainder term $Q_{n,2}$ below.

Lemma 9 *There exist universal constants $C_3,C_4>0$ such that if $h(P_0,P)\le C_3$, then $\|\Lambda\|_\infty\le C_4$, where the sup-norm is taken over the interval $[0,\tau]$.*

Proof Let us first check that

$$\sup_{u\in[0,\tau]}\int(1-G_z(u-))\left|e^{-e^{\theta z}\Lambda(u)}-e^{-e^{\theta_0 z}\Lambda_0(u)}\right|\varphi(z)dz\le 2h(P_0,P). \tag{25}$$

Let us denote $\Psi_{\eta,z}(u)=\mathbf{P}_\eta(T\ge u|Z=z)=(1-G_z(u-))e^{-e^{\theta z}\Lambda(u)}$ for $0\le u<\tau$. For any $0\le s<\tau$, let $\psi_{\eta,z}(s)$ be the derivative of $\Psi_{\eta,z}(u)$ with respect to u at the point s. Since $\Psi_{\eta,z}(0)=\Psi_{\eta_0,z}(0)$, it holds $\Psi_{\eta,z}(u)-\Psi_{\eta_0,z}(u)=\int_0^u\left[\psi_{\eta_0,z}(s)-\psi_{\eta,z}(s)\right]ds$

for $0 \le u < \tau$. Using Cauchy-Schwarz inequality, the inequality $(a+b)^2 \le 2a^2 + 2b^2$ and the fact that $\Psi_{\eta,z}(u) \le 1$, we get that for all $0 \le u < \tau$ it holds

$$\left(\iint_0^u \left| \psi_{\eta_0,z}(s) - \psi_{\eta,z}(s) \right| \varphi(z) ds dz \right)^2 \le 2 \iint_0^u \left(\sqrt{\psi_{\eta_0,z}(s)} - \sqrt{\psi_{\eta,z}(s)} \right)^2 \varphi(z) ds dz.$$

Now note that for any positive reals a_1, a_2, b_1, b_2, it holds

$$\left\{ \sqrt{a_1 + b_1} - \sqrt{a_2 + b_2} \right\}^2 \le 2 \left\{ (\sqrt{a_1} - \sqrt{a_2})^2 + (\sqrt{b_1} - \sqrt{b_2})^2 \right\}.$$

Thus for $0 \le u < \tau$, the penultimate display is bounded above by

$$4 \iint \left[(1 - G_z(t-)) \left\{ \sqrt{e^{r + \theta z - \Lambda e^{\theta z}}} - \sqrt{e^{r_0 + \theta_0 z - \Lambda_0 e^{\theta_0 z}}} \right\}^2 \right.$$
$$\left. + g_z(t) \left\{ \sqrt{e^{-\Lambda e^{\theta z}}} - \sqrt{e^{-\Lambda_0 e^{\theta_0 z}}} \right\}^2 \right] \varphi(z) dt dz \le 4h^2(P_0, P).$$

For $u = \tau$, similarly, one gets $\int |\Psi_{\eta,z}(\tau) - \Psi_{\eta_0,z}(\tau)| \varphi(z) dz \le h(P_0, P)$, which proves (25). Now let $d\mu_u(z)$ denote the measure $(1 - G_z(u-))\varphi(z)dz$. For each fixed u in $[0, \tau]$, the mean value theorem applied to the function $z \to e^{-e^{\theta z}\Lambda(u)} - e^{-e^{\theta_0 z}\Lambda_0(u)}$ with respect to the measure $d\mu_u$ implies that there exists a real $z^*(u) \in [-M, M]$ such that

$$\int \left| e^{-e^{\theta z}\Lambda(u)} - e^{-e^{\theta_0 z}\Lambda_0(u)} \right| d\mu_u(z) = \left| e^{-e^{\theta z^*(u)}\Lambda(u)} - e^{-e^{\theta_0 z^*(u)}\Lambda_0(u)} \right| \int d\mu_u(z).$$

Since $\int (1 - G_z(u-)) \varphi(z) dz = \mathbf{P}_{\eta_0}(Y \ge u) \ge \mathbf{P}_{\eta_0}(Y \ge \tau)$ is bounded away from zero, this means that, due to (25),

$$\sup_{u \in [0,\tau]} \left| e^{-e^{\theta z^*(u)}\Lambda(u)} - e^{-e^{\theta_0 z^*(u)}\Lambda_0(u)} \right| \le C h(P_0, P).$$

From this it is easily seen that, since θ, Z and $\|\Lambda_0\|_\infty$ are bounded, Λ must be bounded from above by a finite constant if $h(P_0, P)$ is small enough. □

Lemma 10 *For any $(\theta, r) \in \Theta \times \mathscr{F}_n$ such that $h(P_{\theta,r}, P_0) \le \varepsilon_n$, it holds*

$$\|\theta - \theta_0, \log\lambda - \log\lambda_0\|_L^2 \lesssim (n\varepsilon_n^2)\varepsilon_n^2 = \zeta_n^2.$$

Moreover, $(\theta - \theta_0)^2 \lesssim \zeta_n^2$ and $\int \lambda_0 \log^2 \lambda/\lambda_0 \lesssim \zeta_n^2$.

Proof First, due to the form of $\mathscr{F}_n$, on this set it holds $\|r\|_\infty = \|\log\lambda\|_\infty \lesssim \sqrt{n}\varepsilon_n$. Thus using the expression of $\log(p/p_0)$ obtained in the proof of Lemma 8, one gets $\|\log p/p_0\|_\infty \lesssim \sqrt{n}\varepsilon_n$ on the sieve. An application of Lemma 8 in [11] then gives that the second moment $W(P_0, P)$ verifies $W(P_0, P) \lesssim h^2(P_0, P) n\varepsilon_n^2 \lesssim n\varepsilon_n^4$. Now simple calculations reveal that $\|\theta - \theta_0, r - r_0\|_L^2$ is bounded from above by

$$W(P_0, P) + \iint \left(\Lambda_0 e^{\theta_0 z} - \Lambda e^{\theta z} \right)^2 e^{\theta_0 z - \Lambda_0 e^{\theta_0 z}} (1 - G_z(t-)) d\Lambda_0(t) \varphi(z) dz.$$

Since $g_z(t)$ is uniformly bounded away from zero, the last term in the preceding display is bounded from above by

$$\iint (r-r_0+(\theta-\theta_0)z)^2 e^{\theta_0 z-\Lambda_0 e^{\theta_0 z}} g_z(t)\varphi(z)d\Lambda_0(t)dz \lesssim W(P_0,P).$$

The last two statements of the lemma are simple consequences of the first result and of the explicit expression of the norm $\|\cdot\|_L$. □

Lemma 11 *For any* $(\theta,r)\in\mathscr{A}_n$ *such that* $h(P_{\theta,r},P_0)\le\varepsilon_n$*, it holds*

$$\int \lambda\log^2\lambda/\lambda_0 \lesssim (n\varepsilon_n^2)\varepsilon_n^2=\zeta_n^2.$$

Proof Let us denote $S(\lambda)^2=\lambda e^{\theta z-\Lambda e^{\theta z}}(1-G_z(t-))\varphi(z)$ and define $S(\lambda_0)^2$ similarly. Further set $X(\lambda)=\iint(r-r_0+(\theta-\theta_0)z)^2S(\lambda)^2$. Due to Lemma 10 and Cauchy-Schwarz inequality,

$$\begin{aligned}
X(\lambda) &= \iint [r-r_0+(\theta-\theta_0)z]^2\,(S(\lambda_0)^2+\{S(\lambda)^2-S(\lambda_0)^2\})\\
&\lesssim \zeta_n^2+\left(\iint\{S(\lambda)-S(\lambda_0)\}^2\right)^{1/2}\left(\iint [r-r_0+(\theta-\theta_0)z]^4\,\{S(\lambda)^2+S(\lambda_0)^2\}\right)^{1/2}\\
&\lesssim \zeta_n^2+h(P_0,P)\,\|r-r_0+(\theta-\theta_0)z\|_\infty\left(X(\lambda)+\iint [r-r_0+(\theta-\theta_0)z]^2\,S(\lambda_0)^2\right)^{1/2}.
\end{aligned}$$

We deduce, using that $\|r\|_\infty\lesssim\sqrt{n}\varepsilon_n$ on the sieve,

$$X(\lambda)\ \lesssim\ \zeta_n^2+h(P_0,P)\sqrt{n}\varepsilon_n\sqrt{X(\lambda)+\zeta_n^2}\ \lesssim\ \zeta_n^2+\zeta_n\sqrt{X(\lambda)+\zeta_n^2}.$$

Hence $X(\lambda)\lesssim\zeta_n^2$, from which, using the facts that Λ is bounded due to Lemma 9 and that $(\theta-\theta_0)^2\lesssim\zeta_n^2$ (Lemma 10), we obtain the result. □

4.3.2 Verification of **(N')**

Let V_n be defined as in **(N')** but with f replaced by r. Let the sieve $\mathscr{A}_n$ and the rate ζ_n be as defined above. Let us write $R_n(\theta,r)=R_{n,1}(\theta,r)+R_{n,2}(\theta,r)$, where $R_{n,1}(\theta,r)=-\mathbb{G}_n\psi_n(\theta,r)$, with

$$\begin{aligned}
\psi_n(\theta,r)(T_i,Z_i) &= \sqrt{n}\Big\{e^{\theta Z_i}\Lambda_0\{e^{r-r_0}\}(T_i)-e^{\theta_0 Z_i}\Lambda_0(T_i)\\
&\qquad -(\theta-\theta_0)Z_ie^{\theta_0 Z_i}\Lambda_0(T_i)-e^{\theta_0 Z_i}\Lambda_0\{r-r_0\}(T_i)\Big\}\\
R_{n,2}(\theta,r) &= -n\Lambda_0\Big\{M_0(\theta)e^{r-r_0}-M_0-(\theta-\theta_0)M_1-(r-r_0)M_0\\
&\qquad -\frac{1}{2}\left[(\theta-\theta_0)^2M_2+2(\theta-\theta_0)(r-r_0)M_1+(r-r_0)^2M_0\right]\Big\}.
\end{aligned}$$

Let us use the decomposition

$$\begin{aligned}
&R_n(\theta,r)-R_n(\theta_0,r-(\theta-\theta_0)\gamma_n)\\
&\quad=R_n(\theta,r)-R_n(\theta_0,r)+R_n(\theta_0,r)-R_n(\theta_0,r-(\theta-\theta_0)\gamma_n).
\end{aligned}$$

We first deal with the term $Q_{n,2}(\theta,r)=R_{n,2}(\theta,r)-R_{n,2}(\theta_0,r)$. Due to Taylor's theorem, $\exp((\theta-\theta_0)Z)-1=(\theta-\theta_0)Z+(\theta-\theta_0)^2Z^2/2+(\theta-\theta_0)^3Z^3Y/6$, with Y bounded and $\exp(r-r_0)=1+(r-r_0)+(r-r_0)^2/2+T_1/2$, where

$$T_1=\int_0^{r-r_0}(r-r_0-t)^2e^t\,dt\leq(r-r_0)^2\left|1-e^{r-r_0}\right|\lesssim(r-r_0)^2(1+e^{r-r_0}).$$

Simple calculations yield that, on V_n,

$$\begin{aligned}Q_{n,2}(\theta,r)=-n(\theta-\theta_0)\Big[(\theta-\theta_0)\int(r-r_0)M_2d\Lambda_0/2+\int(r-r_0)^2M_1d\Lambda_0/2\\ +(\theta-\theta_0)\int(r-r_0)^2M_2d\Lambda_0/4+\Lambda_0\left\{(M_1+o(1))T_1+(\theta-\theta_0)^2o(1)\right\}\Big].\end{aligned}$$

The dominating term is the one involving T_1. The upper bound on T_1 obtained above together with Lemmas 10, 11 implies that $|Q_{n,2}(\theta,r)|\leq\zeta_nh^2+(\sqrt{n}\zeta_n^2)h$. As announced above, we now impose $\zeta_n=o(n^{-1/4})$, so one obtains $Q_{n,2}(\theta,r)=o(1+h^2)$.

Now let us deal with $Q_{n,1}(\theta,r)=(R_{n,1}(\theta,r)-R_{n,1}(\theta_0,r))/(1+h^2)=-\mathbb{G}_nf_n$. Note that $f_n(t,z)(\theta,r)=g_{1,n,\theta}(z)\Lambda(t)-g_{2,n,\theta}(z)\Lambda_0(t)$, where

$$g_{1,n,\theta}(z)=\sqrt{n}e^{\theta_0z}\frac{e^{(\theta-\theta_0)z}-1}{1+n(\theta-\theta_0)^2}\quad\text{and}\quad g_{2,n,\theta}(z)=\sqrt{n}e^{\theta_0z}\frac{(\theta-\theta_0)z}{1+n(\theta-\theta_0)^2}.$$

Let us denote by $\mathscr{P},\mathscr{L},\mathscr{G}_1$ and $\mathscr{G}_2$ the respective collections of all functions $f_n(t,z)(\theta,r)$, $\Lambda(t)$, $g_{1,n,\theta}(z)$, $g_{2,n,\theta}(z)$, for $n\geq1$ and (θ,r) varying in V_n. Denoting by $N_{[]}$ and $J_{[]}$ the bracketing number and bracketing integral (see e.g. [27], Chapter 19), it follows from the decomposition of f_n that

$$N_{[]}(\varepsilon,\mathscr{P},L^2(P_{\eta_0}))\lesssim N_{[]}(\varepsilon,\mathscr{L},L^2(P_{\eta_0}))\cdot N_{[]}(\varepsilon,\mathscr{G}_1,L^2(P_{\eta_0}))\cdot N_{[]}(\varepsilon,\mathscr{G}_2,L^2(P_{\eta_0})).$$

Since θ and z are bounded, $\|g_i\|_\infty+\|g_i'\|_\infty$ is bounded from above for $i=1,2$ by a universal constant. Note also that Λ is nondecreasing and uniformly bounded over the sieve due to Lemma 9. Thus (see e.g. Corollary 2.7.2 and Theorem 2.7.5 of [28]), the preceding display is bounded above by $e^{C/\varepsilon}$. Hence $J_{[]}(\delta,\mathscr{P},L^2(P_{\eta_0}))\lesssim\sqrt{\delta}$ for any $\delta>0$. Using Taylor's theorem, there exists a (bounded) real r_z such that

$$\begin{aligned}f_n(t,z)(\theta,r)\lesssim\frac{|h|e^{\theta_0z}}{1+h^2}\Big(|z|\Lambda_0\left(|r-r_0|+(r-r_0)^2+(r-r_0)^2e^{r-r_0}\right)(t)\\ +\frac{|h|z^2e^{r_z}}{2\sqrt{n}}\Lambda_0\left(1+|r-r_0|+(r-r_0)^2+(r-r_0)^2e^{r-r_0}\right)(t)\Big),\end{aligned}$$

from which, using Lemmas 10 and 11, we easily deduce that $\|f_n\|_\infty\lesssim\zeta_n$ and thus also $P_{\eta_0}f^2\lesssim\zeta_n^2$. We apply Lemma 3.4.2 in [28] to conclude that $\|\mathbb{G}_nf_n\|_{\mathscr{P}}=o_P(1)$.

Finally, denoting $\varphi_n=(\theta-\theta_0)\gamma_n$, we have to deal with the following terms

$$\begin{aligned}S_{n,1}&\triangleq\frac{R_{n,1}(\theta_0,r-\varphi_n)-R_{n,1}(\theta_0,r)}{1+h^2}=-\mathbb{G}_ng_n,\quad\text{with}\\ g_n&=\sqrt{n}\left\{\Lambda_0\left(e^{r-r_0}\{e^{\varphi_n}-1\}-\varphi_n\right)(T)\right\}e^{\theta_0Z}/(1+h^2)\quad\text{and}\\ S_{n,2}&\triangleq-n\Lambda_0\left(e^{r-r_0}\{e^{\varphi_n}-1\}M_0-\varphi_nM_0-(r-r_0)\varphi_nM_0-\varphi_n^2M_0/2\right),\end{aligned}$$

where $S_{n,2}$ is the analog of $Q_{n,2}$ between the points $(\theta_0, r-\varphi_n)$ and (θ_0, r). The term $S_{n,2}$ can be treated by the same method as $Q_{n,2}$ since, as noted above, $\|\gamma_n\|_\infty$ is bounded, so a Taylor expansion can be carried out.

To deal with the stochastic term $S_{n,1}$, we notice that g_n is a difference of two bounded functions of bounded total variation. Indeed, denoting for any s,t in $[0,\tau]$, $G_n(\Lambda)(t) = \sqrt{n}\int_0^t (e^{\varphi_n(u)}-1)d\Lambda(u)/(1+h^2)$, it holds

$$|G_n(\Lambda)(t) - G_n(\Lambda)(s)| \leq \frac{\sqrt{n}|\theta-\theta_0|}{1+h^2}\|\gamma_n\|_\infty(\Lambda(t)-\Lambda(s)) \lesssim \Lambda(t)-\Lambda(s).$$

Since Λ is nondecreasing and bounded on the sieve (see Lemma 9), we deduce that $G_n(\Lambda)$ belongs to the set of bounded functions (since $G_n(\Lambda)(0)=0$) of bounded total variation, whose entropy is well-controlled. On the other hand, using the notation $\Gamma_n(t) = \sqrt{n}\Lambda_0\varphi_n(t)/(1+h^2)$, it holds $|\Gamma_n(t)-\Gamma_n(s)| \leq \|\gamma_n\|_\infty\|\lambda_0\|_\infty|t-s|$. Thus Γ_n is again bounded of bounded total variation. To conclude that $S_{n,1} = o_P(1)$ using the entropy bounds of Chapter 2.7 of [28], it suffices to obtain an upper bound tending to zero for $\mathbf{E}(g_n(T,Z)^2)$. This can be done using Taylor expansions as for $S_{n,2}$ together with the fact that $\|\gamma_n\|_\infty$ is bounded and Lemma 11, which eventually leads to $\mathbf{E}(g_n(T,Z)^2) \lesssim \zeta_n^2$, which concludes the verification of **(N')**.

So far the only condition on the prior is that $\zeta_n = \sqrt{n}\varepsilon_n^2 = o(n^{-1/4})$ should hold. This leads, due to the fact that $\varepsilon_n \lesssim n^{-\alpha\wedge\beta/(2\alpha+1)}$ up to a log-factor, to the conditions $\alpha > 3/2$ and $8\beta > 6\alpha+3$, which correspond to the triangle shape in Figure 2.

4.3.3 Checking condition (E)

As noticed in 4.3.1, the condition on the rate ρ_n is $\rho_n \lesssim n^{-\alpha\wedge B/(2\alpha+1)}$, possibly up to a log-factor. Thus to satisfy $\sqrt{n}\zeta_n\rho_n = o(1)$, one must have $B > 2\beta/3$ if $\alpha \geq \beta$ and $B > 1$ if $\alpha < \beta$, which reduce to $B > 2\beta/3$. Finally note that $\sqrt{n}W_n(0,\gamma_n-\gamma)$ is a sum of n independent, centered random variables of variance bounded by a constant times $\int_0^\tau\{\gamma_n-\gamma\}^2 d\Lambda_0$, which tends to zero. This concludes the proof of Theorem 5. □

5 Appendix: tools for Gaussian processes

5.1 Concentration function, rate and suprema

Let Z be a separable Gaussian process with sample paths almost surely in the Banach space $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$ and let $\mathbb{H}$ be its Reproducing Kernel Hilbert space (RKHS). Let $\mathbb{B}_1$ and $\mathbb{H}_1$ denote the unit balls of the corresponding spaces. Suppose that f_0 belongs to the support of Z in $\mathbb{B}$ and define the *concentration function* of Z by

$$\varphi_{f_0}(\varepsilon) = \inf_{h\in\mathbb{H},\ \|h-f_0\|_{\mathbb{B}}\leq\varepsilon} \|h\|_{\mathbb{H}}^2 - \log\mathbf{P}(\|Z\|_{\mathbb{B}} \leq \varepsilon).$$

Then Theorem 2.1 in [29] states that, if ε_n is such that

$$\varphi_{f_0}(\varepsilon_n) \leq n\varepsilon_n^2, \tag{26}$$

then for any $C > 1$ there exists a measurable set B_n such that for large enough n,

$$\log N(3\varepsilon_n, B_n, \|\cdot\|_{\mathbb{B}}) \le 6Cn\varepsilon_n^2 \quad (27)$$
$$\mathbf{P}(Z \notin B_n) \le \exp(-Cn\varepsilon_n^2) \quad (28)$$
$$\mathbf{P}(\|Z - f_0\|_{\mathbb{B}} < 2\varepsilon_n) \ge \exp(-n\varepsilon_n^2). \quad (29)$$

and B_n can be chosen of the form $\varepsilon_n \mathbb{B}_1 + \sqrt{10Cn}\varepsilon_n \mathbb{H}_1$ for large enough C.

We often use the next Lemma to control suprema of Gaussian processes.

Lemma 12 (Corollary 2.2.8 in [28]) *Let Y be a separable sub-Gaussian process with intrinsic semi-metric $d(s,t) = \mathbf{E}((Y_s - Y_t)^2)^{1/2}$ on its index set. Then for every $\delta > 0$, for a universal constant K,*

$$\mathbf{E}\Big(\sup_{d(s,t)\le\delta} |Y_s - Y_t|\Big) \le K \int_0^{\delta} \sqrt{\log D(\eta, \mathscr{T}, d)}\, d\eta.$$

5.2 Gaussian priors given as series expansions

Let us define a scale of Hilbert spaces $(\mathbb{B}^\alpha, \|f\|_{2,\alpha})$ with parameter $\alpha > 0$ by

$$\mathbb{B}^\alpha = \Big\{ f = \sum_{k\ge1} f_k \varepsilon_k(\cdot), \quad \sum_{k\ge1} k^{2\alpha} f_k^2 < +\infty \Big\}, \quad \|f\|_{2,\alpha}^2 = \sum_{k\ge1} k^{2\alpha} f_k^2.$$

Note that for any $p < \alpha$, the process (10) has sample paths almost surely in $\mathbb{B}^p$, since $\sum_k k^{2p-1-2\alpha} v_k^2$ is positive with finite mean and thus is finite almost surely. Let $\mathbb{H}^\alpha$ be the RKHS of the process defined by (10). Due to Theorem 4.2 of [30], the space $\mathbb{H}^\alpha$ coincides with $(\mathbb{B}^{\alpha+1/2}, \|\cdot\|_{2,\alpha+1/2})$. Borell's inequality (see [3], here in the notation of [30], Theorem 5.1) then enables us to obtain concentration properties of the process (10) viewed as a random element of $\mathbb{B}^p$, as stated in the next Lemma.

Lemma 13 *Assume that f is distributed according to (10). For any $p < \alpha$, for $C > 0$ large enough, if $\alpha_n \to 0$ and $\alpha_n \ge Cn^{-1/(2+\frac{1}{\alpha-p})}$, then*

$$\mathbf{P}(f \notin \alpha_n \mathbb{B}_1^p + \sqrt{10n}\alpha_n \mathbb{H}_1^\alpha) \le \exp(-n\alpha_n^2).$$

Proof Let us denote $\varphi_{0,p}(\cdot) = -\log \pi_f(\|f\|_{\mathbb{B}^p} < \cdot)$ the small ball probability in $\mathbb{B}^p$ associated to the process (10). By Borell's inequality,

$$\mathbf{P}(f \notin \alpha_n \mathbb{B}_1^p + \sqrt{10n}\alpha_n \mathbb{H}_1^\alpha) \le 1 - \Phi(\Phi^{-1}(e^{-\varphi_{0,p}(\alpha_n)}) + \sqrt{10n}\alpha_n).$$

Theorem 4 in [18] implies that $\varphi_{0,p}(\alpha_n) \lesssim \alpha_n^{-1/(\alpha-p)}$. The condition on α_n then implies that for C large enough, $\varphi_{0,p}(\alpha_n) \le n\alpha_n^2$. Since Φ^{-1} is nondecreasing we have $\Phi^{-1}\left(e^{-\varphi_{0,p}(\alpha_n)}\right) \ge \Phi^{-1}(e^{-n\alpha_n^2})$. Now the inequality $-\sqrt{\frac{5}{2}\log(1/y)} \le \Phi^{-1}(y) \le 0$, valid for all $y \in (0,1/2)$, implies that $\sqrt{10n}\alpha_n \ge -2\Phi^{-1}(e^{-n\alpha_n^2})$, which, using the identity $\Phi(-\Phi^{-1}(x)) = 1 - x$, gives the result. □

Lemma 14 *Let π_f be the prior induced by (10), seen as a random element in $\mathbb{B}^0$. Let f_0 satisfy conditions* **(R)** *with $\beta > 1$. Then $\varepsilon_n = Dn^{-\frac{\alpha\wedge\beta}{2\alpha+1}}$ with D large enough, satisfies (26). With this choice of ε_n, Equations (27), (28), (29) are satisfied with $\mathbb{B} = \mathbb{B}^0$ and $B_n = \varepsilon_n \mathbb{B}_1^0 + \sqrt{10Cn}\varepsilon_n \mathbb{H}_1^\alpha$.*

Proof It suffices to establish that ε_n has the above expression, which can be done by bounding from above φ_{f_0}, see [5], Theorem 2 for the detailed calculations. □

To control the entropy of the RKHS unit ball $\mathbb{H}_1^\alpha$, we use a general link existing between the small ball probability of a Gaussian process and the entropy of its RKHS unit ball, see [18] or [30], Lemma 6.2.

Lemma 15 *Let Y be a centered Gaussian process in the space $(\mathbb{B}, \|\cdot\|)$ with associated RKHS $\mathbb{H}$. For any $\gamma > 0$, as $\varepsilon \to 0$ it holds $-\log \mathbf{P}(\|Y\| < \varepsilon) \asymp \varepsilon^{-\gamma}$ if and only $\log N(\varepsilon, \mathbb{H}_1, \|\cdot\|) \asymp \varepsilon^{-2\gamma/(2+\gamma)}$.*

Thus, if Y^α is a process distributed according to (10), due to [18], Theorem 4, the small ball probability of the process behaves as $-\log \mathbf{P}(\|Y^\alpha\|_2 < \varepsilon) \asymp \varepsilon^{-1/\alpha}$, as $\varepsilon \to 0$. Hence Lemma 15 implies that $\log N(\varepsilon, \mathbb{H}_1^\alpha, \|\cdot\|_2) \asymp \varepsilon^{-2/(2\alpha+1)}$ as $\varepsilon \to 0$.

5.3 Gaussian priors of the Riemann-Liouville type

Upper-bounds for the concentration function of the process prior (14) are given in the next Lemma. For a proof, see [5], Theorem 4.

Lemma 16 *Suppose f_0 belongs to $\mathscr{C}^\beta[0,1]$, with $\beta > 0$. The concentration function φ_{f_0} associated to the process X_t^α satisfies, if $0 < \alpha \le \beta$, that $\varphi_{f_0}(\varepsilon) = O(\varepsilon^{-1/\alpha})$ as $\varepsilon \to 0$. In the case that $\alpha > \beta$, if $\{\alpha\}$ denotes the integer part of α, as $\varepsilon \to 0$,*

$$\varphi_{f_0}(\varepsilon) = \begin{cases} O(\varepsilon^{-\frac{2\alpha-2\beta+1}{\beta}}) & \text{if} \quad \{\alpha\} = 1/2 \quad \text{or} \quad \alpha \notin \beta + \frac{1}{2} + \mathbb{N}, \\ O(\varepsilon^{-\frac{2\alpha-2\beta+1}{\beta}} \log(1/\varepsilon)) & \text{otherwise.} \end{cases}$$

This result implies that, for the Riemann-Liouville type process, the rate ε_n such that $\varphi_{f_0}(\varepsilon_n) \le n\varepsilon_n^2$ can be chosen equal to constant times $n^{-\alpha\wedge\beta/(2\alpha+1)}$, possibly with an additional logarithmic factor.

5.4 Changing variables in Gaussian measures

Let Z be a *centered* Gaussian process in a Banach space $(\mathbb{B}, \|\cdot\|)$ of real functions with RKHS $(\mathbb{H}, \|\cdot\|_{\mathbb{H}})$ and covariance function $K(s,t) = \mathbf{E}(Z(s)Z(t))$. Let us denote by P^Z the distribution of Z and by $\mathbf{E}_Z$ the expectation under this law. Given the collection of paths $t \to Z(t,\omega)$ for $\omega \in \Omega$ of the process Z, let us define

$$\begin{aligned} U : \langle\ \{t \to K(\cdot,t),\ t \in \mathbb{R}\}\ \rangle &\to L^2(\Omega) \\ \sum_{i=1}^p a_i K(\cdot, t_i) &\to \sum_{i=1}^p a_i Z(t_i, \omega). \end{aligned}$$

Note that U is an isometry and since by definition any $h \in \mathbb{H}$ is the limit of a sequence $\sum_{i=1}^{p(n)} a_{i,n} K(\cdot, t_{i,n})$, it can be extended into an isometry $U : \mathbb{H} \to L^2(\Omega)$. Then Uh is the limit in $L^2(\Omega)$ of the sequence $\sum_{i=1}^{p(n)} a_{i,n} Z(t_{i,n}, \omega)$.

Lemma 17 *Let $\Phi : \mathbb{B} \to \mathbb{R}$ be a measurable function. Then for any $g, h \in \mathbb{H}$ and $\rho > 0$ it holds*

$$\mathbf{E}_Z(\mathbf{1}_{|Ug| \le \rho} \Phi(Z - h)) = \mathbf{E}_Z(\mathbf{1}_{|Ug + \langle g,h \rangle_{\mathbb{H}}| \le \rho} \Phi(Z) \exp\{U(-h) - \|h\|_{\mathbb{H}}^2/2\}).$$

Proof The change of variable formula for Gaussian measures (see [30], Lemma 3.1) states that if h belongs to $\mathbb{H}$, then the measures P^{Z-h} and P^Z are absolutely continuous and $dP^{Z-h}/dP^Z = \exp(U(-h) - \|h\|_{\mathbb{H}}^2/2)$. On the other hand, since $g \in \mathbb{H}$, for any integer $m \ge 1$ there exist sequences $(a_{1,m}, \ldots, a_{p(m),m})$ and $(t_{1,m}, \ldots, t_{p(m),m})$ such that $p(m) \to +\infty$ and $g(\cdot)$ is the limit in $\mathbb{H}$ of $\sum_{i=1}^{p(m)} a_{i,m} K(\cdot, t_{i,m})$ as $m \to +\infty$. Moreover,

$$\sum_{i=1}^{p(m)} a_{i,m} h(t_{i,m}) = \langle \sum_{i=1}^{p(m)} a_{i,m} K(\cdot, t_{i,m}) \ , \ h(\cdot) \ \rangle_{\mathbb{H}} \to \langle g, h \rangle_{\mathbb{H}},$$

as $m \to +\infty$. By definition Ug is the limit in L^2 of $\sum_{i=1}^{p(m)} a_{i,m} Z(t_{i,m}, \omega)$, where Z is distributed according to P^Z. Thus this quantity is equal to the limit in L^2 of the sum $\sum_{i=1}^{p(m)} a_{i,m} \{Z'(t_{i,m}, \omega) + h(t_{i,m})\}$, where $Z' = Z - h$, that is $Ug + \langle g, h \rangle_{\mathbb{H}}$ if Z' is distributed according to P^Z. □

Acknowledgements Most of this work was carried out while the author was a Postdoc at the VU University of Amsterdam. The author would like to thank Professor Aad van der Vaart for his many helpful comments during the preparation of the paper and a referee for his very detailed and constructive comments.

References

1. Barron A, Schervish MJ, Wasserman L (1999) The consistency of posterior distributions in nonparametric problems. Ann Statist 27(2):536–561
2. Birgé L (1984) Sur un théorème de minimax et son application aux tests. Probab Math Statist 3(2):259–282
3. Borell C (1975) The Brunn-Minkowski inequality in Gauss space. Invent Math 30(2):207–216
4. Castillo I (2007) Semi-parametric second-order efficient estimation of the period of a signal. Bernoulli 13(4):910–932
5. Castillo I (2008) Lower bounds for posterior rates with Gaussian process priors. Electronic Journal of Statistics 2:1281–1299
6. Cox DR (1972) Regression models and life-tables. J Roy Statist Soc Ser B 34:187–220
7. Dalalyan AS, Golubev GK, Tsybakov AB (2006) Penalized maximum likelihood and semiparametric second-order efficiency. Ann Statist 34(1):169–201
8. Diaconis P, Freedman D (1986) On the consistency of Bayes estimates. Ann Statist 14(1):1–67, with a discussion and a rejoinder by the authors

9. Freedman D (1999) On the Bernstein-von Mises theorem with infinite-dimensional parameters. Ann Statist 27(4):1119–1140
10. Ghosal S (2000) Asymptotic normality of posterior distributions for exponential families when the number of parameters tends to infinity. J Multivariate Anal 74(1):49–68
11. Ghosal S, van der Vaart A (2007) Posterior convergence rates of Dirichlet mixtures at smooth densities. Ann Statist 35(2):697–723
12. Ghosal S, van der Vaart AW (2007) Convergence rates of posterior distributions for noniid observations. Ann Statist 35(1)
13. Ghosal S, Ghosh JK, van der Vaart AW (2000) Convergence rates of posterior distributions. Ann Statist 28(2):500–531
14. Härdle W, Liang H, Gao J (2000) Partially linear models. Contributions to Statistics, Physica-Verlag, Heidelberg
15. Ibragimov IA, Has′minskiĭ RZ (1981) Statistical estimation, Applications of Mathematics, vol 16. Springer-Verlag, New York
16. Kim Y (2006) The Bernstein-von Mises theorem for the proportional hazard model. Ann Statist 34(4):1678–1700
17. Kim Y, Lee J (2004) A Bernstein-von Mises theorem in the nonparametric right-censoring model. Ann Statist 32(4):1492–1512
18. Kuelbs J, Li WV (1993) Metric entropy and the small ball problem for Gaussian measures. J Funct Anal 116(1):133–157
19. Le Cam L (1986) Asymptotic methods in statistical decision theory. Springer Series in Statistics, Springer-Verlag, New York
20. Le Cam L, Yang GL (1990) Asymptotics in statistics. Springer Series in Statistics, Springer-Verlag, New York, some basic concepts
21. McNeney B, Wellner JA (2000) Application of convolution theorems in semiparametric models with non-i.i.d. data. J Statist Plann Inference 91(2):441–480
22. Rasmussen CE, Williams CK (2006) Gaussian processes for machine learning. MIT Press, Cambridge, MA
23. Schwartz L (1965) On Bayes procedures. Z Wahrscheinlichkeitstheorie und Verw Gebiete 4:10–26
24. Shen X (2002) Asymptotic normality of semiparametric and nonparametric posterior distributions. J Amer Statist Assoc 97(457):222–235
25. Shen X, Wasserman L (2001) Rates of convergence of posterior distributions. Ann Statist 29(3):687–714
26. van der Vaart A (2002) The statistical work of Lucien Le Cam. Ann Statist 30(3):631–682, dedicated to the memory of Lucien Le Cam
27. van der Vaart AW (1998) Asymptotic statistics, Cambridge Series in Statistical and Probabilistic Mathematics, vol 3. Cambridge University Press, Cambridge
28. van der Vaart AW, Wellner JA (1996) Weak convergence and empirical processes. Springer Series in Statistics, Springer-Verlag, New York
29. van der Vaart AW, van Zanten H (2008) Rates of contraction of posterior distributions based on Gaussian process priors. Ann Statist 36(3):1435–1463
30. van der Vaart AW, van Zanten H (2008) Reproducing kernel Hilbert spaces of Gaussian priors. IMS Collections 3:200–222