

Collegio Carlo Alberto



Bayesian asymptotics with misspecified models

Pierpaolo De Blasi
Stephen G. Walker

No. 270
October 2012

Carlo Alberto Notebooks

www.carloalberto.org/research/working-papers

Bayesian asymptotics with misspecified models

Pierpaolo De Blasi*

University of Torino and Collegio Carlo Alberto, Italy

Stephen G. Walker

School of Mathematics, Statistics, & Actuarial Science,
University of Kent, Canterbury, UK

Abstract

In this paper, we study the asymptotic properties of a sequence of posterior distributions based on an independent and identically distributed sample and when the Bayesian model is misspecified. We find a sufficient condition on the prior for the posterior to accumulate around the densities in the model closest in the Kullback–Leibler sense to the true density function. Examples are presented.

Key words and phrases: Asymptotics, Consistency, Misspecified model.

1 Introduction

This paper is concerned with asymptotics for Bayesian nonparametric models. In particular, we consider generalizations of the recent literature on consistency; see for example, Barron, Schervish and Wasserman (1999), Ghosal, Ghosh and Ramamoorthi (1999), and Walker (2004). The standard assumption for consistency is that the true density function, which we denote by f_0 , is in the Kullback–Leibler support of the prior, denoted by Π . Further sufficient conditions on the prior are then established in order to ensure that the sequence of posterior distributions accumulate in suitable neighborhoods of f_0 . The three papers just cited deviate in the precise form of the further sufficient conditions.

We make the support of the prior assumption more general now by assuming that the closest density in the support of the prior is a possibly non-zero Kullback–Leibler divergence away from f_0 ; specifically, if f_1 is the closest density, in the Kullback–Leibler sense, in the support \mathcal{F} of the prior (to be made more precise later), then δ_1 is defined to be the Kullback–Leibler divergence between f_0 and f_1 . We then look for further sufficient conditions under which the posterior distributions accumulate in suitable neighborhoods of f_1 . In particular, we work around the ideas presented in Walker (2004) and the sufficient conditions for accumulation at f_1 can be seen as a generalization of the condition appearing in Walker (2004).

The convenience of working in this setting is quite evident. When considering asymptotics, there are two possible scenarios: $\delta_1 = 0$ or $\delta_1 > 0$. The former involves

*address for correspondence: pierpaolo.deblasi@unito.it

a well specified model and the latter a misspecified model. Typically, the latter is more likely, though in reality it will be unknown. However, one can assume $\delta_1 = 0$ and derive conditions on the prior for posterior accumulation at f_0 , and then assume $\delta_1 > 0$ and derive another set of conditions on the prior for posterior accumulation at f_1 . It is to be sure that the latter conditions will be stronger than the former. In this case, and the value of δ_1 unknown, it makes perfect sense to construct the prior under the condition that $\delta_1 \geq 0$, and hence under the misspecified case.

Early work for the misspecified problem has been done by Berk (1966) and more recently by Bunke and Milhaud (1998), Shalizi (2009), and Kleijn and van der Vaart (2006). In particular, the strategy followed by Kleijn and van der Vaart (2006) consists of defining a neighborhood around f_1 according to a suitable semi-metric on the space of densities that satisfies an entropy condition related to the Hellinger integral h_α (to be defined later); see their equation (2.2). Kleijn and van der Vaart concentrate on the notion of a single f_1 for which accumulation of the posterior takes place and extend this to a finite number of such f_1 .

On the other hand, we focus our efforts directly on a set \mathcal{F}_1 rather than on a single f_1 , acknowledging the fact that in general one does not know how big the set of densities associated with the minimum Kullback-Leibler distance δ_1 is. Hence, we find it appropriate to define

$$\mathcal{F}_1 = \{f \in \bar{\mathcal{F}} : D(f_0, f) \leq \delta_1\}, \quad (1.1)$$

where $\bar{\mathcal{F}}$ is the Hellinger closure of \mathcal{F} and $D(f_0, f)$ is the Kullback-Leibler divergence of f relative to f_0 . When \mathcal{F}_1 is non-empty, our working assumption throughout the paper, we show accumulation at \mathcal{F}_1 with respect to the Hellinger distance. When \mathcal{F}_1 is empty, we show that the posterior accumulates in a different set that we define and explain at the end of Section 3.

In reality, and in general, it is not known whether \mathcal{F}_1 is empty or not, since f_0 is not known. A notable exception is when \mathcal{F} is convex, in which case \mathcal{F}_1 reduces to a single density f_1 . However, our key prior condition, given in Section 3, covers both \mathcal{F}_1 empty or non-empty. Kleijn and van der Vaart (2006) only establish what the posterior does when \mathcal{F}_1 is a finite set. For a chosen Π it may be possible to find C_Π such that if $f_0 \in C_\Pi$ then \mathcal{F}_1 is empty, whereas if $f_0 \in C_\Pi^c$ then \mathcal{F}_1 is non-empty. But in spite of this being a difficult task, it would not even be known if f_0 was in C_Π or not, and hence the objective in this area would focus on finding Π for which C_Π can be shown to be empty. To date this is only known to be true when \mathcal{F} is convex. Therefore knowing what happens when $f_0 \in C_\Pi$ is important, and one of the contributions of the paper is to fill this gap in the literature.

The rest of the paper is organized as follows. In Section 2 we start with some notation, definitions, and preliminary results. The main results are presented in Section 3 and illustrations involving various priors are given in Section 4. We conclude with a discussion in Section 5.

2 Notation and preliminary results

We introduce the notation of the paper, together with essential preliminary results. Let \mathbb{X} be a separable metric space endowed with its Borel σ -field \mathcal{B} , and denote by Ω the space of density functions on $(\mathbb{X}, \mathcal{B})$ relative to some reference measure (that is omitted henceforth for ease of notation). For d a distance on Ω and $A \subset \Omega$, we denote by $N(\delta, A, d)$ the minimum number of balls of radius at most δ , with respect to the metric d , needed to cover A . In particular, on Ω we consider the Hellinger distance $H(f, g) = \{\int (\sqrt{f} - \sqrt{g})^2\}^{1/2}$ (which makes Ω a separable space) and the Kullback–Leibler divergence $D(f, g) = \int \log(f/g)f$. Moreover, we define, for $\alpha \in (0, 1)$, the Hellinger integral $h_\alpha(f, g) = \int f^{1-\alpha}g^\alpha$ and the α -divergence $d_\alpha(f, g) = \alpha^{-1}[1 - h_\alpha(f, g)]$, see Liese and Vajda (2006) and the references therein. We recall here that $h_\alpha(f, g) \leq 1$ for any $f, g \in \Omega$ and that $d_\alpha(f, g)$ is decreasing in α with $\lim_{\alpha \rightarrow 0} d_\alpha(f, g) = D(f, g)$ whenever $D(f, g) < \infty$. The case $\alpha = 1/2$ yields the Hellinger distance, since $d_{1/2}(f, g) = H^2(f, g)$. A lemma, whose proof is deferred to the Appendix, provides a useful inequality that relates the Hellinger distance and the Hellinger integral h_α . It is used in the proofs of Theorem 1 and Lemma 2.

Lemma 1. *For any $f, g, f_0 \in \Omega$ and $0 \leq \alpha \leq \frac{1}{2}$,*

$$|h_\alpha(f_0, f) - h_\alpha(f_0, g)| \leq [H(f, g)]^{2\alpha}.$$

Note that for $\alpha = 1/2$, we have $H(f, g) \geq |H^2(f_0, f) - H^2(f_0, g)|/2$, which is weaker than the standard triangle inequality, as it can be proved by using (A.2) in the Appendix and the fact that $H(f, g) \leq \sqrt{2}$.

As we are going to deal with convergence of sets of densities, we consider the metric space (Ω, H) , and define $H(A, f) = \inf_{g \in A} H(g, f)$ to be the Hellinger distance between A and $f \in \Omega$, and $H(A, B) = \max\{\sup_{f \in B} H(A, f), \sup_{f \in A} H(B, f)\}$ to be the Hausdorff distance (relative to the Hellinger) between A and B . In particular, it can be shown that

$$H(A, f) \leq H(B, f) + H(A, B). \quad (2.1)$$

Since the Hellinger is a bounded distance, convergence in the Hausdorff metric of a sequence (A_n) to A is equivalent to $H(A_n, f) \rightarrow H(A, f)$ for every $f \in \Omega$ (known as Kuratowski convergence). Moreover, in case of a decreasing sequence (A_n) , the limit is given by $\bigcap_n \overline{A_n}$, see Rockafellar and Wets (2009, Chapter 4).

Now let X_1, X_2, \dots be independent and identically distributed random variables taking values in $(\mathbb{X}, \mathcal{B})$ with common density function $f_0 \in \Omega$ and, given $\mathcal{F} \subseteq \Omega$, let Π be a prior probability measure on \mathcal{F} . The Bayesian posterior measure is given by

$$\Pi_n(A) = \int_A R_n(f) \Pi(df) / \int_{\mathcal{F}} R_n(f) \Pi(df), \quad (2.2)$$

where A is a measurable subset of \mathcal{F} and $R_n(f) = \prod_{i=1}^n f(X_i)/f_0(X_i)$. We denote by F_0^∞ the infinite product measure relative to f_0 . Finally, upon definition of $\delta_1 = \inf_{f \in \mathcal{F}} D(f_0, f)$, we refer to \mathcal{F}_1 in (1.1) as the set of pseudo-true densities f_1 . An associate editor has suggested an alternative definition of the minimum Kullback–Leibler distance as $\delta_1 = \inf\{t : \Pi(f : D(f_0, f) \leq t) > 0\}$; we explicitly work with the

former but note the latter is effectively equivalent for the purposes of our paper, see also (3.5) below. Existence of a pseudo-true density is a delicate issue. Given the lower semicontinuity of $D(f_0, \cdot)$ as a map from the metric space (Ω, H) to \mathbb{R} , see Lemma 8.2 in Kleijn and van der Vaart (2006), a (rather) stringent sufficient condition is the compactness of \mathcal{F} . To our knowledge, most of the theoretical results are based on the hypothesis of convexity of \mathcal{F} , $\sup_{f \in \mathcal{F}} \int \log(f) f_0$ being finite the essential additional requirement. See Liese and Vajda (1987, Chapter 8), Pfanzagl (1990) and Patilea (2001).

3 Main results

In the well-specified case with $\delta_1 = 0$, strong consistency corresponds to

$$\Pi_n(f : H(f_0, f) > \epsilon) \rightarrow 0 \quad F_0^\infty - \text{a.s.} \quad (3.1)$$

for any $\epsilon > 0$, entailing that the posterior concentrates all the mass in an arbitrarily small Hellinger neighborhood of the true f_0 . In the misspecified case, it is reasonable to ask that the posterior concentrates mass around the set of pseudo-true densities \mathcal{F}_1 in (1.1),

$$\Pi_n(\{f : H(\mathcal{F}_1, f) > \epsilon\}) \rightarrow 0 \quad F_0^\infty - \text{a.s.}$$

In order to establish this result, we follow a route different from the one in Kleijn and van der Vaart (2006) in that we keep on working, although in an instrumental way, on neighborhoods around the true f_0 . Specifically, we consider the sets

$$A_{\alpha, \epsilon} = \{f \in \mathcal{F} : d_\alpha(f_0, f) > \delta_1 + \epsilon/\alpha\}, \quad (3.2)$$

$$A_\alpha = \{f \in \mathcal{F} : d_\alpha(f_0, f) > \delta_1 + \alpha\}. \quad (3.3)$$

Note that the latter can be recovered from (3.2) with $\epsilon = \alpha^2$, although any $\epsilon = \epsilon(\alpha)$ decreasing in α with $\epsilon(\alpha)/\alpha \rightarrow 0$ as $\alpha \rightarrow 0$ would work. The idea is that A_α^c is monotonically decreasing in α to \mathcal{F}_1 .

Lemma 2. *Let A_α and \mathcal{F}_1 be as in (3.3) and (1.1), respectively. Then, $\bigcap_\alpha \overline{A_\alpha^c} \subseteq \mathcal{F}_1$.*

The proof of Lemma 2 is provided in the Appendix. We aim now at establishing sufficient conditions for

$$\Pi_n(A_{\alpha, \epsilon}) \rightarrow 0 \quad F_0^\infty - \text{a.s.} \quad (3.4)$$

for any α and ϵ sufficiently small. To this aim, we first adapt the Kullback–Leibler property to the misspecified case as

$$\Pi(f \in \mathcal{F} : D(f_0, f) \leq \delta_1 + \eta) > 0 \quad (3.5)$$

for any $\eta > 0$, see Theorem 2.1 in Kleijn and van der Vaart (2006). In fact, a simple corollary of Lemma 3 and 4 in Barron, Schervish and Wasserman (1999) implies that, for all large n and for any $c > 0$,

$$I_n \geq e^{-n(\delta_1 + c)}, \quad F_0^\infty - \text{a.s.}, \quad (3.6)$$

where $I_n = \int_{\mathcal{F}} R_n(f) \Pi(df)$ is the denominator of (2.2). As for the numerator, the key condition can be stated, similar to Walker (2004), in terms of summability of powers of prior probabilities. To this end, for a given $\alpha \in (0, 1)$, let $(B_{j,\varepsilon})_{j \geq 1}$ be Hellinger balls of size $\varepsilon > 0$ that cover \mathcal{F} such that

$$\sum_{j \geq 1} \Pi(B_{j,\varepsilon})^\alpha < \infty. \quad (3.7)$$

We are now ready to state and prove our main result.

Theorem 1. *Suppose Π satisfies (3.5) and that (3.7) holds for some $\alpha \in (0, 1/2)$, where the sets $B_{j,\varepsilon}$ are Hellinger balls of size $\varepsilon = 2(\varepsilon/2)^{1/(2\alpha)}$ whose union covers \mathcal{F} . Then (3.4) holds.*

Proof. Let $(A_j)_{j \geq 1}$ be a partition of $A_{\alpha,\varepsilon}$ (to be specified later) and define $f_{n,j}$ to be the predictive density with posterior distribution restricted, and normalized, to the set A_j . Note that

$$f_{n,j}(x) = \int_{A_j} f(x) \frac{\Pi_n(df)}{\Pi_n(A_j)} = \frac{\int_{A_j} f(x) R_n(f) \Pi(df)}{\int_{A_j} R_n(f) \Pi(df)},$$

so that, letting $L_{n,j} = \int_{A_j} R_n(f) \Pi(df)$,

$$L_{n+1,j}/L_{n,j} = f_{n,j}(X_{n+1})/f_0(X_{n+1}),$$

see also Walker (2004). Then, we have

$$\Pi_n(A_{\alpha,\varepsilon}) = \sum_{j \geq 1} \Pi_n(A_j) \leq \sum_{j \geq 1} \Pi_n(A_j)^\alpha = \sum_{j \geq 1} L_{n,j}^\alpha / I_n^\alpha, \quad (3.8)$$

$$\mathbb{E}(L_{n+1,j}^\alpha | X_1, \dots, X_n) = h_\alpha(f_0, f_{n,j}) L_{n,j}^\alpha. \quad (3.9)$$

Take $A_j \subseteq A_j^* = \{f : H(f_j, f) < \varepsilon/2\}$, where $\varepsilon = 2(\varepsilon/2)^{1/(2\alpha)}$ and $(f_j)_{j \geq 1}$ are densities in $A_{\alpha,\varepsilon}$. By using Lemma 1, we have that

$$h_\alpha(f_0, f_{n,j}) - h_\alpha(f_0, f_j) \leq [H(f_j, f_{n,j})]^{2\alpha} < \varepsilon/2$$

which, together with $h_\alpha(f_0, f_j) < 1 - \alpha(\delta_1 + \varepsilon/\alpha)$ (since $f_j \in A_{\alpha,\varepsilon}$), yields

$$h_\alpha(f_0, f_{n,j}) < 1 - \alpha\delta_1 - \varepsilon/2.$$

Hence, from (3.9), we get

$$\mathbb{E}(L_{n,j}^\alpha) < (1 - \alpha\delta_1 - \varepsilon/2)^n \Pi(A_j)^\alpha < e^{-n(\alpha\delta_1 + \varepsilon/2)} \Pi(A_j)^\alpha.$$

As for the numerator of (3.8), by the Markov Inequality,

$$\mathbb{P}\left(\sum_{j \geq 1} L_{n,j}^\alpha > e^{nd} e^{-n(\alpha\delta_1 + \varepsilon/2)}\right) < e^{-nd} \sum_{j \geq 1} \Pi(A_j)^\alpha.$$

Since (3.7) implies that $\sum_{j \geq 1} \Pi(A_j)^\alpha < \infty$, we get that $\sum_{j \geq 1} L_{nj}^\alpha < e^{-n(\alpha\delta_1 + \epsilon/2 - d)}$ F_0^∞ -a.s. for all large n , for any $d > 0$. As for the denominator of (3.8), note that (3.6) implies that, for all large n and for any $c > 0$, $I_n^\alpha \geq e^{-n\alpha(\delta_1 + c)}$ F_0^∞ -a.s. Therefore,

$$\Pi_n(A_{\alpha, \epsilon}) \leq \sum_{j \geq 1} L_{nj}^\alpha / I_n^\alpha \leq e^{-n(\epsilon/2 - d - \alpha c)} \rightarrow 0 \quad F_0^\infty - \text{a.s.}$$

by taking c and d sufficiently small. \square

A corollary to Theorem 1 and Lemma 2 provides the sufficient condition for accumulation of the posterior at \mathcal{F}_1 in the Hellinger sense.

Corollary 1. *Suppose Π satisfies (3.5) and that*

$$\sum_{j \geq 1} \Pi(B_{j, \varepsilon_\alpha})^\alpha < \infty \quad (3.10)$$

for all $\alpha \in (0, 1/2)$, where $\varepsilon_\alpha = 2(\alpha^2/2)^{1/(2\alpha)}$. Then, as $n \rightarrow \infty$, if \mathcal{F}_1 is non-empty, $\Pi_n(\{f : H(\mathcal{F}_1, f) > \epsilon\}) \rightarrow 0$ F_0^∞ -a.s. for any $\epsilon > 0$.

Proof. Condition (3.10) implies that, for all $\alpha \in (0, 1/2)$, $\Pi_n(A_\alpha) \rightarrow 0$ F_0^∞ -a.s. So, clearly,

$$\Pi_n(f \in \mathcal{F} : H(\overline{A_\alpha^c}, f) \leq \xi) \rightarrow 1 \quad F_0^\infty - \text{a.s.}$$

for any $\xi > 0$ and all $\alpha \in (0, 1/2)$. Now define $A = \bigcap_\alpha \overline{A_\alpha^c}$ and use (2.1) to get $H(A, f) \leq H(\overline{A_\alpha^c}, f) + H(\overline{A_\alpha^c}, A)$. Since $A_\alpha^c \rightarrow A$ in Kuratowski sense, and because of the equivalence of Kuratowski and Hausdorff convergence, $H(\overline{A_\alpha^c}, A) \rightarrow 0$ as $\alpha \rightarrow 0$. It follows that, for any $\epsilon > 0$, there are ξ and α sufficiently small such that $\{f \in \mathcal{F} : H(A, f) \leq \epsilon\} \subseteq \{f \in \mathcal{F} : H(\overline{A_\alpha^c}, f) \leq \xi\}$, and we can conclude that

$$\Pi_n(f \in \mathcal{F} : H(A, f) \leq \epsilon) \rightarrow 1 \quad F_0^\infty - \text{a.s.}$$

for any $\epsilon > 0$. The thesis follows from Lemma 2, since $A \subseteq \mathcal{F}_1$. \square

Remark 1. *We discuss the alteration of Corollary 1 if \mathcal{F}_1 is empty. Suppose Π satisfies (3.5) and (3.10) for all $\alpha \in (0, 1/2)$, where $\varepsilon_\alpha = 2(\alpha^2/2)^{1/(2\alpha)}$. Then, as $n \rightarrow \infty$, the posterior still accumulates at A_α^c for any $\alpha > 0$. The lack of elements in \mathcal{F}_1 now means there is no further development possible. However, it is to be noted that (3.10) is the key condition we need from the prior in order to establish what happens to the posterior no matter the state of \mathcal{F}_1 . Hence, it is this condition we examine in the examples of Section 4.*

At this point it is useful to see how consistency is recovered in the well-specified case. First note that (3.4) for $\alpha = 1/2$ and any $\epsilon > 0$ corresponds to strong consistency in the well-specified case. In fact it is easy to check that the conditions of Theorem 1 for $\delta_1 = 0$ and $\alpha = 1/2$ correspond to Theorem 4 of Waker (2004). However, as noted by Walker, Lijoi and Prünster (2005), the prior summability condition can be replaced with an arbitrary power α .

Theorem 2. (Walker, 2004) *Let $\delta_1 = 0$. If Π satisfies (3.5) and (3.7) for some $\alpha \in (0, 1)$, then $\Pi_n(\{f : H(f_0, f) > \varepsilon\}) \rightarrow 0$ F_0^∞ -a.s.*

For completeness, the proof of Theorem 2 is provided in the Appendix. It is now clear how the sufficient condition in the misspecified case differs from that in the well-specified case: for the latter, according to Theorem 2, condition (3.7) needs to be satisfied for a single $\alpha \in (0, 1)$ (and any $\varepsilon > 0$) in order to have Hellinger consistency.

4 Examples

In this section we consider a number of examples. In each case we consider an infinite-dimensional model and find the prior summability conditions in Section 3 established on each prior Π . If Π has full support and inference is possible for the infinite-dimensional model, then we revert to the well-specified case and the required condition is weaker ((3.7) only needs to hold for a single α). However, typically infinite-dimensional models are truncated, or do not have full support, and if f_0 is out of the range of the truncation or the support of the prior, then our results are required for all α , (3.10). Specifically, truncation has to be intended in terms of: the number of components in the mixture of priors of Section 4.1; the number of elements in the orthonormal basis of the infinite-dimensional exponential family example of Section 4.2; the support of the prior for the scale parameters σ and λ in Section 4.3 and 4.4, respectively. We find the condition on the infinite-dimensional model as this obviously covers all levels of truncation.

4.1. Mixture of priors. Consider the prior on Ω given by

$$\Pi = \sum_{N \geq 1} p_N \Pi_N,$$

where $\sum_{N \geq 1} p_N = 1$ and Π_N is supported on a set of densities $C_N \subset \Omega$. This example has been considered in Walker (2004). Let $C_N \subseteq C_{N+1}$, C_N increasing to some $\mathcal{F} \subseteq \Omega$ as $N \rightarrow \infty$. We assume that, for each N , C_N is totally bounded with respect to the Hellinger metric, $N(\epsilon, C_N, H) < \infty$ for any $\epsilon > 0$. An example is given by the Bernstein polynomial prior of Petrone and Wasserman (2002).

For fixed ϵ , let $(B_{j,\epsilon})$ be the Hellinger balls of size ϵ that cover \mathcal{F} . We may assume without loss of generality that, for $I_N := N(\epsilon, C_N, H)$, $C_N \subseteq \bigcup_{j \leq I_N} B_{j,\epsilon}$, so that $\Pi_N(B_{j,\epsilon}) = 0$ for any $j > I_N$. For $\alpha \in (0, 1)$, we consider

$$\sum_{j \geq 1} \Pi(B_{j,\epsilon})^\alpha = \sum_{j \geq 1} \left(\sum_{N: I_N \geq j} p_N \Pi_N(B_{j,\epsilon}) \right)^\alpha \leq \sum_{j \geq 1} \left(\sum_{N \geq M_j} p_N \right)^\alpha$$

where $M_j = \min\{N : I_N \geq j\}$. Since I_N depends on ϵ , so does M_j , hence we write $M_j(\epsilon)$. Consequently, by defining $\bar{P}(m) = \sum_{N \geq m} p_N$, if $\sum_{j \geq 1} \bar{P}(M_j(\epsilon))^\alpha < \infty$, then (3.7) holds. Hence, it is sufficient that

$$\bar{P}(M_j(\epsilon)) < a j^{-\frac{1}{\alpha}-r} \tag{4.1}$$

for some $r > 0$ and $a > 0$ for all large j .

For example if $I_N = (c/\epsilon)^N$, for some c not depending on ϵ , as in the case of Bernstein polynomial prior, then $M_j(\epsilon) = \lfloor \log j / \log(c/\epsilon) \rfloor$, so that (4.1) yields

$$\bar{P}(N) < a(c/\epsilon)^{-\left(\frac{1}{\alpha}+r\right)N} = a \exp\{-\psi(\epsilon, \alpha)N\}, \quad (4.2)$$

where $\psi(\epsilon, \alpha) = \log(c/\epsilon)(\alpha^{-1} + r)$. Note that (4.2) puts a constraint on the prior mass on large C_N . Since $\psi(\epsilon, \alpha)$ increases to ∞ as ϵ decreases to zero, (3.7) holds for any $\epsilon > 0$ (consistency in the well-specified case) if $\bar{P}(N) < ae^{-\psi N}$ for any $\psi > 0$ and all large N , which holds if $N^{-1} \log \bar{P}(N) \rightarrow -\infty$. Hence, we recover the condition of Section 6 in Walker (2004). Establishing (3.10) does not change the result, in fact $\psi(\epsilon, \alpha)$ also increases to ∞ as α decreases to zero. Therefore, in this example, the required condition for the misspecified case is that for the well-specified case.

4.2. Infinite-dimensional exponential family. Let $\Theta = (\theta_j)_{j \geq 1}$ be a sequence of independent random variables with $\theta_j \sim N(0, \sigma_j^2)$, and let $(\phi_j)_{j \geq 1}$ be a sequence of orthogonal polynomials on $[0, 1]$. Define the family of densities as

$$f_{\Theta}(x) = \exp \left\{ \sum_{j \geq 1} \theta_j \phi_j(x) - c(\Theta) \right\}$$

where $c(\Theta)$ makes f_{Θ} a density. This example has been considered in Barron, Schervish and Wasserman (1999) and Walker (2004). For illustration, we work with the orthonormal basis

$$\phi_1(x) = 1 \quad \text{and} \quad \phi_j(x) = \sqrt{2} \cos(j\pi x) \quad \text{for } j \geq 2,$$

so that $\|\phi_j\|_{\infty} = \sqrt{2}$ and $\|\phi_j'\|_{\infty} = j$, for any $j \geq 2$. To ensure that f_{Θ} is a density with probability 1, it is sufficient that $\sum_j \sigma_j < \infty$.

We next consider how to construct a Hellinger covering $(B_{j,\epsilon})_{j \geq 1}$ in (3.7) for the density set \mathcal{F} on which Π is supported, the prior being induced by the distribution on the infinite sequence Θ . Suppose, for $i = 1, 2$, we put

$$f_i(x) = \frac{e^{w_i(x)}}{\int e^{w_i(y)} dy}.$$

Then, $\|w_1 - w_2\|_{\infty} \leq \epsilon$ implies $H(f_1, f_2) \leq \epsilon e^{\epsilon/2}$ (see Lemma 3.1 in van der Vaart and van Zanten (2008)). Now take $f_1 = f_{\Theta_1}$ and $f_2 = f_{\Theta_2}$ for sequences $\Theta_1 = (\theta_{1j})$ and $\Theta_2 = (\theta_{2j})$ such that $w_i(x) = \sum_j \theta_{ij} \phi_j(x)$. Also, take Θ_1 and Θ_2 close in the sense that

$$|\theta_{1j} - \theta_{2j}| < \delta_j = \delta \omega_j / \sum_{j \geq 1} \omega_j$$

for some sequence (ω_j) satisfying $\sum_j \omega_j < \infty$. Then

$$\left\| \sum_{j \geq 1} \theta_{1j} \phi_j - \sum_{j \geq 1} \theta_{2j} \phi_j \right\|_{\infty} = \sqrt{2} \delta$$

and $H(f_{\Theta_1}, f_{\Theta_2}) \leq \sqrt{2} \delta e^{\delta/\sqrt{2}}$. It follows that $B_{j,\epsilon}$ can be taken as set of the type

$$\{f_{\Theta} : n_j \delta_j < \theta_j < (n_j + 1) \delta_j\}$$

for $\delta_j = \delta(\epsilon)\omega_j/\sum_{j \geq 1} \omega_j$ with $\delta(\epsilon) = g^{-1}(\epsilon)/\sqrt{2}$, g^{-1} being the inverse of $g(x) = xe^{x/2}$, and integers (n_j) that vary between $-\infty$ and $+\infty$. Note that $\delta(\epsilon)$ is monotonic increasing in ϵ , with $\delta(0) = 0$. Since $\theta_j \sim N(0, \sigma_j^2)$, with independence across j , in view of (3.7), we are interested in the finiteness of

$$\sum_{n_1=-\infty}^{\infty} \dots \sum_{n_M=-\infty}^{\infty} \prod_{j=1}^{\infty} (\Pr[n_j \delta_j < \theta_j < (n_j + 1)\delta_j])^\alpha.$$

Due to symmetry, this holds if

$$\prod_{j=1}^{\infty} \sum_{n=0}^{\infty} (\Pr[n\delta_j < \theta_j < (n+1)\delta_j])^\alpha.$$

Next, we have

$$\begin{aligned} \sum_{n=0}^{\infty} (\Pr[n\delta_j < \theta_j < (n+1)\delta_j])^\alpha &\leq 1 + (2\pi)^{-\frac{\alpha}{2}} (\delta_j/\sigma_j)^\alpha \sum_{n=1}^{\infty} \exp\{-\alpha\delta_j^2 n/(2\sigma_j^2)\} \\ &\leq 1 + (2\pi)^{-\frac{\alpha}{2}} (\delta_j/\sigma_j)^\alpha [\exp\{\alpha\delta_j^2/(2\sigma_j^2)\} - 1]^{-1}. \end{aligned}$$

Note that, for any $m \geq 1$, $e^z - 1 \geq z^m/m!$, so that we can use the inequality

$$[\exp\{\alpha\delta_j^2/(2\sigma_j^2)\} - 1]^{-1} \leq m!(2\sigma_j^2)^m/\alpha^m \delta_j^{2m}$$

to get

$$\begin{aligned} \sum_{n=0}^{\infty} (\Pr[n\delta_j < \theta_j < (n+1)\delta_j])^\alpha &\leq 1 + (2\pi)^{-\frac{\alpha}{2}} (\delta_j/\sigma_j)^\alpha m!(2\sigma_j^2)^m/\alpha^m \delta_j^{2m} \\ &= 1 + (2/\alpha)^m m! (2\pi)^{-\frac{\alpha}{2}} (\sigma_j/\delta_j)^{2m-\alpha}. \end{aligned}$$

Substituting for $\delta_j = \delta(\epsilon)\omega_j/\sum_j \omega_j$, we get

$$\sum_{n=0}^{\infty} (\Pr[n\delta_j < \theta_j < (n+1)\delta_j])^\alpha \leq 1 + \psi(\epsilon, \alpha) (\sigma_j/\omega_j)^{2m-\alpha},$$

where

$$\psi(\epsilon, \alpha) = (2/\alpha)^m m! (2\pi)^{-\frac{\alpha}{2}} (\delta(\epsilon)/\sum_{j \geq 1} \omega_j)^{2m-\alpha}.$$

The required condition on the (σ_j) is then that

$$\prod_{j=1}^{\infty} \{1 + \psi(\epsilon, \alpha) (\sigma_j/\omega_j)^{2m-\alpha}\} < \infty,$$

or that $\sum_{j=1}^{\infty} \log \{1 + \psi(\epsilon, \alpha) (\sigma_j/\omega_j)^{2m-\alpha}\} < \infty$, which holds if

$$\psi(\epsilon, \alpha) \sum_{j=1}^{\infty} (\sigma_j/\omega_j)^{2m-\alpha} < \infty.$$

The convergence of this series requires a restriction on how the sequence (σ_j) grows as $j \rightarrow \infty$. Moreover, we see that the size ϵ of the Hellinger covering does not play any role. Now, if we put $\omega_j \propto j^{-1-r}$ for any $r > 0$, then the condition

$$\sum_{j=1}^{\infty} (\sigma_j j^{1+r})^{2m-\alpha} < \infty \quad (4.3)$$

is sufficient. Therefore, we can actually have $\sigma_j \propto j^{-1-q}$ for any $q > 0$, by choosing $r < q$ and m large enough such that $(r - q)2m$ is sufficiently smaller than -1 . We also see that α does not affect condition (4.3), therefore $\sigma_j \propto j^{-1-q}$ for any $q > 0$ is sufficient for (3.7) to hold for any $\epsilon > 0$ and for (3.10) to hold for any α sufficiently small. This means that the condition in Section 6 of Walker (2004) for consistency works also in the misspecified case.

4.3. Mixtures of normal densities. We consider priors obtained via a nonparametric mixture of normal densities; see Ghosal, Ghosh and Ramamoorthi (1999) and Lijoi, Pruenster and Walker (2005). Let \tilde{P} be a discrete random probability distribution on \mathbb{R} with law Λ and prior guess P_0 . For ϕ_σ , the density function of the normal with mean 0 and variance σ^2 , we model the density as

$$\tilde{f}_{\sigma, \tilde{P}}(x) = \phi_\sigma * \tilde{P} = \int \phi_\sigma(x - \theta) \tilde{P}(d\theta)$$

and σ , with prior distribution μ , is supported on the interval $[0, \bar{\sigma}]$.

We follow the proof of Theorem 1 in Lijoi, Pruenster and Walker (2005) by defining the sets

$$\mathcal{F}_{\underline{\sigma}, a, \delta}^{\bar{\sigma}} = \bigcup_{\underline{\sigma} < \sigma < \bar{\sigma}} \{\phi_\sigma * P : P([-a, a]) \geq 1 - \delta\},$$

where $\underline{\sigma}, a > 0$. Recall that $H^2(f, g) \leq \|f - g\|_1$, where $\|f - g\|_1 = \int |f - g|$ is the L_1 -distance between f and g . Hence we have $N(\sqrt{\delta}, \mathcal{G}, H) \leq N(\delta, \mathcal{G}, \|\cdot\|_1)$. From Ghosal, Ghosh and Ramamoorthi (1999), the upper bound for the L_1 -metric entropy of set $\mathcal{F}_{\underline{\sigma}, a, \delta}^{\bar{\sigma}}$ is given by

$$\log N(\delta, \mathcal{F}_{\underline{\sigma}, a, \delta}^{\bar{\sigma}}, \|\cdot\|_1) \leq C_\delta a / \underline{\sigma},$$

where $C_\delta = K\delta^{-1} \log(1/\delta)$ for some constant K . Hence, $\log N(\delta, \mathcal{F}_{\underline{\sigma}, a, \delta}^{\bar{\sigma}}, H) \leq C_{\delta^2} a / \underline{\sigma}$. Now let $(a_n)_{n \geq 1}$ be an increasing sequence of positive numbers such that $a_0 = 0$ and $\lim_n a_n = \infty$, and let $(\sigma_n)_{n \geq 1}$ be a decreasing sequence of positive numbers such that $\sigma_0 = \bar{\sigma}$ and $\lim_n \sigma_n = 0$. Set

$$\mathcal{G}_{\sigma_k, a_j, \delta^2}^{\bar{\sigma}} = \bigcup_{\sigma_k < \sigma < \bar{\sigma}} \{\phi_\sigma * P : P([-a_j, a_j]) \geq 1 - \delta^2, P([-a_{j-1}, a_{j-1}]) < 1 - \delta^2\},$$

so that $\bigcup_{j,k} \mathcal{G}_{\sigma_k, a_j, \delta^2}^{\bar{\sigma}} = \mathcal{F}$. Reasoning as in Lijoi, Pruenster and Walker (2005), for any δ there is an integer N such that $\mathcal{G}_{\sigma_k, a_j, \delta^2}^{\bar{\sigma}}$ is included in $\mathcal{F}_{\sigma_k, a_N, \delta^2}^{\bar{\sigma}}$, so

$$\log N(\delta, \mathcal{G}_{\sigma_k, a_j, \delta^2}^{\bar{\sigma}}, H) \leq C_{\delta^2} a_N / \sigma_k.$$

This means that for each j and k , $\mathcal{G}_{\sigma_k, a_j, \epsilon^2}^{\bar{\sigma}}$ has a finite Hellinger ϵ -covering $\{B_{jkl, \epsilon} : l = 1, \dots, N_{j,k}\}$, where $N_{j,k} \leq \exp\{C_{\epsilon^2} a_N / \sigma_k\}$, so that we consider $(B_{jkl, \epsilon})$ in establishing (3.7). Now, for each $j \geq 1$, define the sets

$$B'_{j, \epsilon} = \{P : P([-a_j, a_j]) \geq 1 - \epsilon^2, P([-a_{j-1}, a_{j-1}]) < 1 - \epsilon^2\}.$$

The condition (3.7) is implied by the finiteness of the sum

$$\begin{aligned} \sum_{j, k \geq 1} \sum_{l=1}^{N_{j,k}} \Pi(B_{jkl, \epsilon})^\alpha &\leq \sum_{j, k \geq 1} N_{j,k}^{1-\alpha} \Pi(\mathcal{G}_{\sigma_k, a_j, \epsilon^2}^{\bar{\sigma}})^\alpha \\ &\leq \sum_{k \geq 1} e^{(1-\alpha)C_{\epsilon^2} a_N / \sigma_k} \mu(\sigma_k < \sigma \leq \sigma_{k-1})^\alpha \sum_{j \geq 1} \Lambda(B'_{j, \epsilon})^\alpha, \end{aligned} \quad (4.4)$$

where the inequality (4.4) follows by the monotonicity of power means. We deal with the inner sum first, showing that $\sum_{j \geq 1} \Lambda(B'_{j, \epsilon})^\alpha < \infty$ for any $\epsilon > 0$ and any $\alpha \in (0, 1)$ is implied by

$$P_0([-a, a]^c) \leq e^{-\eta a} \quad (4.5)$$

for some $\eta > 0$ and a sufficiently large. In fact, since $B'_{j, \epsilon} \subset \{P : P([-a_{j-1}, a_{j-1}]^c) > \epsilon^2\}$, an application of Markov Inequality, together with (4.5), yields

$$\sum_{j \geq 1} \Lambda(B'_{j, \epsilon})^\alpha \leq \epsilon^{-2\alpha} \sum_{j \geq 1} P_0([-a_{j-1}, a_{j-1}]^c)^\alpha \leq \epsilon^{-2\alpha} \sum_{j \geq 1} e^{-\alpha \eta a_{j-1}}.$$

Then, by taking $a_j \sim j$ as $j \rightarrow \infty$, $a_j > (\alpha \eta)^{-1} (1+s) \log j$ for j sufficiently large and some $s > 0$, so that $\sum_{j \geq 1} e^{-\alpha \eta a_{j-1}} < \infty$ for any α arbitrarily small (the size of ϵ does not play any role). Note that (4.5) is stronger than the tail condition on P_0 needed for consistency, $\int |\theta| P_0(d\theta) < \infty$, see Theorem 1 in Lijoi, Pruenster and Walker (2005). In fact, the latter only implies $P_0([-a, a]^c) = O(a^{-(1+r)})$ for $r > 0$, and so we get the convergence of the series $\sum_{j \geq 1} \Lambda(B_{j, \epsilon})^\alpha$ only for $\alpha > (1+r)^{-1}$. At this stage we are left to establish that

$$\sum_{k \geq 1} e^{(1-\alpha)C_{\epsilon^2} a_N / \sigma_k} \mu(\sigma_k < \sigma \leq \sigma_{k-1})^\alpha < \infty.$$

If we assume that

$$\mu\{\sigma < k^{-1}\} \leq e^{-\gamma k} \quad (4.6)$$

then

$$\sum_{k \geq 1} e^{(1-\alpha)C_{\epsilon^2} a_N / \sigma_k} \mu(\sigma_k < \sigma \leq \sigma_{k-1})^\alpha \leq \sum_{k \geq 1} e^{-(\gamma \alpha - (1-\alpha)C_{\epsilon^2} a_N) / \sigma_k}$$

Let $\psi(\epsilon, \alpha) = \alpha^{-1}(1-\alpha)C_{\epsilon^2} a_N$, which goes to infinity as either α or ϵ go to zero. Now set $\sigma_k < \psi(\epsilon, \alpha)(1-s)^{-1} / \log k$ for k sufficiently large and for some $s > 0$; this is possible for any α and ϵ by taking, for example, $\sigma_k \sim 1/k$ as $k \rightarrow \infty$. Then the finiteness of the series in the r.h.s. of the last display is implied by $\gamma > \psi(\epsilon, \alpha)$ in (4.6), so that in the well-specified case and the misspecified case we need (4.6) to hold for any $\gamma > 0$.

4.4. *Gaussian process priors.* We consider random densities obtained as logistic transformations of Gaussian processes; see Leonard (1978), Lenk (1988, 1991), Tokdar and Ghosh (2007), and van der Vaart and van Zanten (2008). Let I be a fixed bounded interval I in \mathbb{R} , and let

$$f(x) = \frac{e^{W(x)}}{\int_I e^{W(s)} ds},$$

where $\{W(x), x \in I\}$ is a Gaussian process with mean function $\mu(t)$ and covariance kernel $\sigma(s, t) = \text{Cov}(W(s), W(t))$. Without loss of generality we take $\mu = 0$ and $I = (0, 1)$. Moreover, we let σ depend on a parameter $\lambda > 0$ via

$$\sigma(s, t) = \sigma_0(\lambda s, \lambda t),$$

where σ_0 is a fixed covariance kernel and λ has prior distribution Π_λ , supported on \mathbb{R}_+ . Let W_0 be the Gaussian process with covariance σ_0 so that $W_0(\lambda t)$ has covariance $\sigma_0(\lambda s, \lambda t)$. Then, the random density f is modeled by

$$\begin{aligned} f|W, \lambda &= \frac{e^{W(x)}}{\int_I e^{W(s)} ds} \\ W|\lambda &= W_0(\lambda \cdot) \end{aligned}$$

with λ and W_0 independent. This defines a prior distribution Π on Ω , the space of densities on I . As before, \mathcal{F} is the support of Π .

The Kullback-Leibler support of Π has been studied in Tokdar and Ghosh (2007), see also Ghosal and Roy (2006). With a regularity condition on σ_0 , which we assume to be satisfied (see, e.g., Theorem 5 in Ghosal and Roy (2006)), $W_0(\cdot)$ has differentiable sample paths and the derivative process $DW_0(\cdot)$ is Gaussian with continuous sample paths. If we take

$$\sigma_1(s, t) = \frac{\partial^2}{\partial s \partial t} \sigma_0(s, t),$$

to be the covariance kernel of $DW_0(t)$, then the derivative process $DW_0(\cdot)$ is sub-Gaussian with respect to the Euclidean distance

$$\mathbb{E}(DW_0(s) - DW_0(t))^2 \leq c_1(s - t)^2$$

for some constant c_1 depending on σ_1 . Define $\sigma_1^2(W_0) = \sup_{t \in I} \text{Var}(DW_0(t)) < \infty$. Then, an application of Proposition A.2.7 of van der Vaart and Wellner (1996), yields

$$\Pi\left(\sup_{t \in I} |DW(t)| > a \mid \lambda\right) \leq c_2 \exp\{-c_3 a^2 \lambda^{-2}\} \quad (4.7)$$

for some positive constants c_2 and c_3 . Finally, Theorem 2.7.1 of van der Vaart and Wellner (1996) gives the entropy bound

$$\log N\left(\epsilon, \left\{w : \sup_{t \in I} |Dw(t)| \leq a\right\}, \|\cdot\|_\infty\right) \leq c_4 \frac{a}{\epsilon} \quad (4.8)$$

for some positive constant c_4 . Let $(a_n)_{n \geq 1}$ be an increasing sequence of positive real numbers such that $a_0 = 0$ and $\lim_n a_n = \infty$. For

$$B_j = \left\{ w : a_{j-1} < \sup_{t \in I} |Dw(t)| < a_j \right\},$$

define $F_j = \{f(x) = e^{w(x)} / \int_I e^{w(s)} ds : w \in B_j\}$. Then the sets $(F_j)_{j \geq 1}$ are pairwise disjoint and form a partition of \mathcal{F} . Since $N(\epsilon e^\epsilon, F_j, H) \leq N(\epsilon, B_j, \|\cdot\|_\infty)$, see Section 4.2, there is no loss of generality in working with the sets B_j and the sup norm.

From (4.8), each B_j has a finite ϵ -covering $\{C_{jl, \epsilon} : l = 1, 2, \dots, N_j\}$, where $N_j \leq \exp(c_4 a_j / \epsilon)$. Hence we consider $(C_{jl, \epsilon})$ in establishing (3.7), that is, we are interested in the finiteness of the sum

$$\sum_{j \geq 1} \sum_{l=1}^{N_j} \Pi(C_{jl})^\alpha \leq \sum_{j \geq 1} N_j^{1-\alpha} \Pi(B_j)^\alpha, \quad (4.9)$$

where we have used the monotonicity of power means as in (4.4). Since $N_j^{1-\alpha}$ grows exponentially in a_j , we need to control the behavior of $\Pi(B_j)$ as j goes to infinity. Now, from (4.7),

$$\Pi(B_j) \leq \int_0^\infty \Pi\left(\sup_{t \in I} |DW(t)| > a_{j-1} \mid \lambda\right) \Pi_\lambda(d\lambda) \leq \int_0^\infty c_2 \exp\{-c_3 a_{j-1}^2 \lambda^{-2}\} \Pi_\lambda(d\lambda),$$

and therefore we need to study the behavior, as $a_j \rightarrow \infty$, of a Laplace-type transform of the prior Π_λ . To this end, we resort to a suitable version of the Tauberian Theorem to show that if

$$\Pi_\lambda(\lambda > t) \sim e^{-\gamma t^2} \quad (4.10)$$

for some $\gamma > 0$ and t sufficiently large, then there exists a positive constant c_5 such that

$$\Pi(B_j) \leq e^{-c_5 \gamma a_{j-1}} \quad (4.11)$$

as $j \rightarrow \infty$. The result follows by an application of Theorem 4.12.9 of Bingham, Godie and Teugels (1987). Let μ be a measure on $(0, \infty)$ whose Laplace transform $M(\tau) = \int_0^\infty e^{-\tau x} d\mu(x)$ converges for all $\tau > 0$. With their notation, we choose $\alpha = -1$ and $\phi(\tau) \sim \tau^{-1}$ as $\tau \rightarrow 0^+$. Then, for $B > 0$,

$$-\log \mu(0, x] \sim Bx^{-1} \quad (x \rightarrow 0^+) \quad \text{iff} \quad -\log M(\tau) \sim 2B^{1/2} \tau^{1/2} \quad (\lambda \rightarrow \infty).$$

Now, with $y = 1/\sqrt{x}$, write $M(\tau) = \int_0^\infty e^{-\tau y^{-2}} d\eta(y)$ for the measure η on $(0, \infty)$ defined as $\eta(0, y] = \mu[y^{-2}, \infty)$. Then

$$-\log \eta(y, \infty) \sim By^2 \quad (y \rightarrow \infty) \quad \text{iff} \quad -\log M(\tau) \sim 2B^{1/2} \tau^{1/2} \quad (\lambda \rightarrow \infty). \quad (4.12)$$

Put $\eta(y, \infty) = \Pi_\lambda(y, \infty)$ in (4.12) and assume that (4.10) is in force. Also put $B = \gamma$ and $\tau = c_3 a_{j-1}^2$ in (4.12). Then

$$M(c_3 a_{j-1}^2) = \int_0^\infty e^{-c_3 a_{j-1}^2 \lambda^{-2}} \Pi_\lambda(d\lambda) \sim \exp\{-2\gamma^{1/2} c_3^{1/2} a_{j-1}\}$$

as $j \rightarrow \infty$. It is now easy to see that (4.10) implies (4.11). Back to (4.9), by using (4.8) and (4.11) we get

$$\sum_{j \geq 1} N_j^{1-\alpha} \Pi(B_j)^\alpha \leq \sum_{j \geq 1} \exp \{ - (\alpha c_5 \gamma - c_4(1-\alpha)/\epsilon) a_j \}.$$

Thus, under (4.10), finiteness of the series in (4.9) and, in turn (3.7), is implied by $\gamma > \psi(\alpha, \epsilon)$, where

$$\psi(\alpha, \epsilon) = \frac{c_4(1-\alpha)}{\epsilon \alpha c_5}.$$

Since $\psi(\epsilon, \alpha)$ goes to infinity as either α or ϵ go to zero, we conclude that the same sufficient condition applies for (3.7) for any $\epsilon > 0$ (consistency in the well-specified case) and for (3.10) for any $\alpha > 0$, namely that (4.10) is satisfied for any $\gamma > 0$.

5 Discussion

In this paper, we have generalized the condition for consistency in the well-specified case to asymptotic results in the misspecified case. Illustrations considered suggest that it is not too problematic to implement the sufficient conditions on the prior.

The sufficient conditions we find for the prior allow us to say what happens to the posterior under all scenarios. Basically, whether \mathcal{F}_1 is empty or not, we establish asymptotics for either case. The search when \mathcal{F}_1 is finite and non-empty has been important due to the lack of general theory for the case when \mathcal{F}_1 is empty. To our knowledge we are the first to describe some general theory for the asymptotics when \mathcal{F}_1 is empty; though we are aware of the special case of the Bernstein polynomial prior studied in Petrone and Wasserman (2002).

In each case of misspecified prior Π we would find it difficult to find the C_Π such that if $f_0 \in C_\Pi$ then \mathcal{F}_1 is empty, whereas if $f_0 \in C_\Pi^c$ then \mathcal{F}_1 is non-empty. But of course we would also find it difficult to use this information since the location of f_0 is not known. It would therefore be useful to identify Π for which C_Π is empty, something which has been established only when \mathcal{F} is convex.

Appendix

Proof of Lemma 1. Start with the inequality

$$|h_\alpha(f_0, f) - h_\alpha(f_0, g)| \leq \int |(f/f_0)^\alpha - (g/f_0)^\alpha| f_0. \quad (\text{A.1})$$

Next, note that, for any $a, b > 0$,

$$|a^\alpha - b^\alpha|^{1/\alpha} \leq |a^\beta - b^\beta|^{1/\beta}, \quad \text{for } 0 \leq \alpha \leq \beta. \quad (\text{A.2})$$

To see this, it is sufficient to take $a > b$ and show that $\tau(\alpha) = (a^\alpha - b^\alpha)^{1/\alpha}$ is non decreasing for $\alpha > 0$. The latter is equivalent to $(e^\alpha - 1)^{1/\alpha}$ non decreasing for $\alpha \geq 0$,

and, in turn, to $\log(x-1)/\log(x)$ non decreasing for $x > 1$. The last statement can be easily checked by looking at the first derivative and using the inequality $(x-1)\log x \leq x \log x$ for $x > 1$. Since, by hypothesis, $\alpha \leq 1/2$, (A.2) yields an upper bound on the right hand side of (A.1):

$$\int |(f/f_0)^\alpha - (g/f_0)^\alpha| f_0 \leq \int |(f/f_0)^{1/2} - (g/f_0)^{1/2}|^{2\alpha} f_0.$$

Now use $E|X|^p \leq (E|X|)^p$ for any $0 < p < 1$ to get

$$|h_\alpha(f_0, f) - h_\alpha(f_0, g)| \leq \left(\int |(f/f_0)^{1/2} - (g/f_0)^{1/2}|^2 f_0 \right)^\alpha.$$

Finally, note that $\int |(f/f_0)^{1/2} - (g/f_0)^{1/2}|^2 f_0 = \int (f^{1/2} - g^{1/2})^2 = H^2(f, g)$. \square

Proof of Lemma 2. Let α_m and ϵ_m be two positive sequences decreasing to 0 such that $\epsilon_m^{2\alpha_m}/\alpha_m \rightarrow 0$. Next, let $f \in \bigcap_m \overline{A_{\alpha_m}^c}$ so that, for each m , there exists a sequence $(f_{m,r_m})_{r_m=1}^\infty \in \mathcal{F}$ such that

$$\begin{aligned} H(f, f_{m,r_m}) &\rightarrow 0 \quad \text{as } r_m \rightarrow \infty, \\ d_{\alpha_m}(f_0, f_{m,r_m}) &\leq \delta_1 + \alpha_m \quad \text{for all } r_m. \end{aligned}$$

Clearly, $f \in \bar{\mathcal{F}}$. Moreover, by using Lemma 1 and the identity $d_\alpha(f, g) = \alpha^{-1}\{1 - h_\alpha(f, g)\}$, we have

$$d_{\alpha_m}(f_0, f) < d_{\alpha_m}(f_0, f_{m,r}) + \frac{1}{\alpha_m} H(f, f_{m,r_m})^{2\alpha_m}.$$

By the hypothesis made, we can take, for each m , r_m large enough such that $H(f, f_{m,r_m}) \leq \epsilon_m$ to get

$$d_{\alpha_m}(f_0, f) < \delta_1 + \alpha_m + \frac{1}{\alpha_m} \epsilon_m^{2\alpha_m}.$$

Since the last inequality holds for any m , hence for α_m going to 0, we exploit the convergence of $d_{\alpha_m}(f_0, f)$ to $D(f_0, f)$ to conclude that

$$D(f_0, f) < \delta_1 + \beta$$

for any β small enough. This implies that $f \in \mathcal{F}_1$, completing the proof. \square

Proof of Theorem 2. When $\alpha \leq 1/2$, $\Pi(B_{j,\epsilon})^{1/2} \leq \Pi(B_{j,\epsilon})^\alpha$, so that consistency follows from Theorem 4 in Walker (2004). Let then (3.7) be satisfied for $1/2 < \alpha < 1$. We aim at establishing that $\Pi_n(A_{1/2,\epsilon^2/2}) \rightarrow 0$ F_0^∞ -a.s. as $n \rightarrow \infty$, (3.2) when $\delta_1 = 0$ and (3.1). Reasoning as in the proof of Theorem 1, we consider a partition $(A_j)_{j \geq 1}$ of $A_{1/2,\epsilon^2/2}$ such that $A_j \subseteq A_j^* = \{f : H(f_j, f) < \epsilon/2\}$, $f_j \in A_{1/2,\epsilon^2/2}$ so that $d_{1/2}(f_0, f_j) > \epsilon^2$. Then, any $f \in A_j$ has $d_{1/2}(f_0, f) > \epsilon^2/2$ and $h_{1/2}(f_0, f) < 1 - \epsilon^2/4$. Now use the Hölder Inequality

$$\int (f_0/f_{n,j})^{1-\alpha} f_{n,j} \leq \left(\int (f_0/f_{n,j})^{1/2} f_{n,j} \right)^{2(1-\alpha)}$$

(since $2(1 - \alpha) < 1$) to conclude that $h_\alpha(f_0, f_{n,j}) \leq [h_{1/2}(f_0, f_{n,j})]^{2(1-\alpha)}$. Moreover, $f_{n,j} \in A_j$ implies that $h_\alpha(f_0, f_{n,j}) < (1 - \epsilon^2/4)^{2(1-\alpha)} < e^{-\epsilon^2(1-\alpha)/2}$. Now, similar to the proof of Theorem 1, we obtain that

$$P\left(\sum_{j \geq 1} L_{n,j}^\alpha > e^{-nd}\right) < e^{nd} e^{-n\epsilon^2(1-\alpha)/2} \sum_{j \geq 1} \Pi(A_j)^\alpha.$$

Thus condition (3.7) implies that $\sum_{j \geq 1} L_{n,j}^\alpha < e^{-nd}$ F_0^∞ -a.s. for all large n and for any $d < \epsilon^2(1 - \alpha)/2$. On the other hand, the Kullback-Leibler property (3.5) ensures that, for all large n and for any $c > 0$, $I_n^\alpha \geq e^{-nac}$ F_0^∞ -a.s. Therefore

$$\Pi_n(A_{1/2, \epsilon^2/2}) \leq \sum_{j \geq 1} L_{n,j}^\alpha / I_n^\alpha \leq e^{-n(d-\alpha c)} \rightarrow 0, \quad F_0^\infty - \text{a.s.}$$

by taking c sufficiently smaller than $\epsilon^2(1 - \alpha)/2\alpha$. □

Acknowledgment

S. Tokdar is gratefully acknowledged for some useful discussions. This work was supported by the European Research Council (ERC) through StG ‘‘N-BNP’’ 306406.

References

- Barron, A., Schervish, M.J., and Wasserman, L. (1999). The consistency of posterior distributions in nonparametric problems. *Ann. Statist.* **27**, 536–561.
- Berk, R.H. (1966). Limiting behaviour of posterior distributions when the model is incorrect. *Ann. Math. Statist.* **37**, 51–58. [Corrigendum **37**, 745–746]
- Bingham, N.H., Goldie, C.M., and Teugels, J.L. (1987). *Regular Variation*, Cambridge University Press, Cambridge, UK.
- Bunke, O. and Milhaud, X. (1998). Asymptotic behavior of Bayes estimates under possibly incorrect models. *Ann. Statist.* **26**, 617–644.
- Ghosal, S., Ghosh, J.K., and Ramamoorthi, R.V. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *Ann. Statist.* **27**, 143–158.
- Ghosal, S. and Roy, A. (2006). Posterior consistency of Gaussian process prior for nonparametric binary regression. *Ann. Statist.* **34**, 2413–2429.
- Kleijn, B.J.K. and van der Vaart, A.W. (2006). Misspecification in infinite-dimensional Bayesian statistics. *Ann. Statist.* **34**, 837–887.
- Lenk, P.J. (1988). The logistic normal distribution for Bayesian, nonparametric, predictive densities. *J. Amer. Statist. Assoc.* **83**, 509–516.
- Lenk, P.J. (1991). Towards a practicable Bayesian nonparametric density estimator. *Biometrika* **78**, 531–543.
- Leonard, T. (1978). Density estimation, stochastic processes and prior information (with discussion). *J. R. Stat. Soc. Ser. B* **40**, 113–146.

- Liese, F. and Vajda, I. (1987). *Convex Statistical Distances*, Teubner, Leipzig.
- Liese, F. and Vajda, I. (2006). On divergence and informations in statistics and information theory. *IEEE Transactions on Information Theory* **52**, 4394–4412.
- Lijoi, A., Prünster, I., and Walker, S.G. (2005). On consistency of nonparametric normal mixtures for Bayesian density estimation. *J. Amer. Statist. Assoc.* **100**, 1292–1296.
- Patilea, V. (2001). Convex models, MLE and misspecification. *Ann. Statist.* **29**, 94–123.
- Pfanzagl, J. (1990). Large deviation probabilities for certain nonparametric maximum likelihood estimators. *Ann. Statist.* **18**, 1868–1877.
- Petrone, S. and Wasserman, L. (2002). Consistency of Bernstein polynomial posteriors. *J. R. Stat. Soc. Ser. B* **64**, 79–100.
- Rockafellar, R.T., Wets, R.J-B (2009). *Variational Analysis*, 3rd printing. Springer-Verlag, Berlin.
- Shalizi, C.R. (2009). Dynamics of Bayesian updating with dependent data and misspecified models. *Electron. J. Stat.* **3**, 1039–1074,
- Tokdar, S. and Ghosh, J. (2007). Posterior consistency of Gaussian process priors in density estimation. *J. Statist. Plann. Inference* **137**, 34–42.
- van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.
- van der Vaart, A.W. and van Zanten, J.H. (2008). Rates of contraction of posterior distributions based on Gaussian process priors. *Ann. Statist.* **36**, 1435–1463.
- Walker, S.G. (2004). New approaches to Bayesian consistency. *Ann. Statist.* **32**, 2028–2043.
- Walker, S.G., Lijoi, A., and Prünster, I. (2005). Data tracking and the understanding of Bayesian consistency. *Biometrika* **92**, 765–778.