Asymptotic properties of a sequence of posterior distributions with application to the dirichlet process mixture

C. Scricciolo

2001.19

Dipartimento di Scienze Statistiche
Università degli Studi
Via C. Battisti 241-243
35121 Padova

Dicembre 2001

ASYMPTOTIC PROPERTIES OF A SEQUENCE OF POSTERIOR DISTRIBUTIONS WITH APPLICATION TO THE DIRICHLET PROCESS MIXTURE

CATIA SCRICCIOLO

ABSTRACT. In this paper a sequence of distributions on the set of all probability measures absolutely continuous with respect to a σ-finite measure on a sample space is considered. The generic distribution can be viewed either as a pseudo-posterior distribution obtained by integrating the likelihood raised to a positive power less than one with respect to a legitimate prior or as the posterior distribution corresponding to a certain data-dependent prior. The distribution is Hellinger consistent at each probability measure in the Kullback-Leibler support of the legitimate prior. Theoretical justification is provided for using the *ad hoc* data-dependent prior and the derived posterior, whose rate of convergence is assessed. It is shown how recourse to this distribution can be made to establish sufficient conditions for consistency of the posterior of location mixtures of normal densities, when the scale parameter has a sample-size dependent prior and the mixing measure has any distribution. If the mixing distribution is the trajectory of a Dirichlet process, then the posterior converges at the best known rate $n^{-1/2}(\ln n)$ with respect to the Hellinger distance.

1. INTRODUCTION

Let $X_1, X_2, \ldots, X_n$ be n independent and identically distributed (i.i.d.) observations from an unknown distribution P_0 belonging to the space $\mathcal{F}$ of all probability measures dominated by a σ-finite measure λ on a sample space $(\mathcal{X}, \mathcal{B})$, where $\mathcal{X}$ is a Polish metric space. Assume a prior distribution π is elicited on the measurable space $(\mathcal{F}, \mathcal{C})$, where $\mathcal{C}$ is the Borel σ-field of subsets of $\mathcal{F}$ generated by open sets under the Hellinger metric. The Hellinger distance between any pair $P, Q \in \mathcal{F}$ having densities f_P and f_Q, respectively, is defined as $d_H(P, Q) = \{\int_{\mathcal{X}}(\sqrt{f_P(x)} - \sqrt{f_Q(x)})^2 \, \mathrm{d}\lambda(x)\}^{1/2}$. If $d_{TV}(P, Q) = \sup_{B\in\mathcal{B}} |P(B) - Q(B)|$ is the total variation distance, which can be expressed in terms of the distance induced by the L_1-norm as $d_{TV}(P, Q) = \frac{1}{2}\|f_P - f_Q\|_1$, then

$$\frac{1}{2}d_H(P, Q)^2 \le d_{TV}(P, Q) \le d_H(P, Q), \qquad (P, Q) \in \mathcal{F}\times\mathcal{F}. \tag{1.1}$$

The Hellinger distance and the total variation distance induce the same topology on $\mathcal{F}$ which is referred to as strong topology. We shall consider also the Kullback-Leibler divergence of P and Q

$$K(P\|Q) = \begin{cases} \int_{\mathcal{X}} f_P \ln \frac{f_P}{f_Q} \, \mathrm{d}\lambda, & \text{if } P \ll Q, \\ \infty, & \text{if } P \not\ll Q. \end{cases} \tag{1.2}$$

For any fixed $\eta > 0$, let $K_\eta \equiv K_\eta(P_0) = \{P \in \mathcal{F} : K(P_0\|P) < \eta\}$ denote the generic Kullback-Leibler neighborhood of P_0. The Kullback-Leibler support $\mathcal{S}_K(\pi)$ of π is the set $\mathcal{S}_K(\pi) = \{P_0 \in \mathcal{F} : \forall \eta > 0,\ \pi(K_\eta) > 0\}$. Throughout, we shall write $\mathcal{S}_K$ when the prior is clear from context.

The questions to be answered are the following. As more and more data come in, does the posterior concentrate around the true distribution? What is the rate of convergence?

Let $H_\varepsilon \equiv H_\varepsilon(P_0) = \{P \in \mathcal{F} : d_H(P_0, P) < \varepsilon\}$ be the ε-Hellinger open ball centered at P_0. Denote by P_0^∞ the infinite product measure when sampling from P_0. A sequence $\{\pi(\cdot|x^n)\}_{n=1}^\infty$ of posterior distributions on $(\mathcal{F}, \mathcal{C})$ is said to be *consistent at* $P_0 \in \mathcal{F}$ if, for each $\varepsilon > 0$,

$$\lim_{n\to\infty} \pi(H_\varepsilon^c|x^n) = 0, \tag{1.3}$$

along almost all sample paths. The approach followed by Barron *et al.* (1999) and Ghosal *et al.* (1999) in establishing conditions that guarantee consistency a posteriori, as well as the one pursued by Ghosal, Ghosh *et al.* (2000) and Shen *et al.* (2000) in determining rates of convergence is based on the idea of approximating the parameter space with finite-dimensional subsets. Following this way, a sieve with sets satisfying both an entropy condition and a tail condition is needed to be found. In some cases, the ascertainment of these conditions might be hard to perform or the requirements on the prior be too restrictive, therefore, an alternative approach is desirable. A different attack to both the above problems can be based on the asymptotic behavior of a sequence of distributions introduced by Walker *et al.* (2001). Let $f_P^n(x^n)$ be the joint density $\prod_{i=1}^n f_P(x_i)$. Let $\alpha \in (0, 1)$ be fixed. For any $A \in \mathcal{C}$, let

$$Q^\alpha(A|x^n) = \frac{\int_A f_P^n(x^n)^\alpha \, d\pi(P)}{\int_{\mathcal{F}} f_P^n(x^n)^\alpha \, d\pi(P)}, \qquad x^n \in \mathcal{X}^n,\ n \geq 1$$

if $m_n^\alpha(x^n) = \int_{\mathcal{F}} f_P^n(x^n)^\alpha \, d\pi(P) \in (0, \infty)$, P_0^n-almost surely, otherwise arbitrarily set $Q^\alpha(A|x^n) = \pi(A)$. If P_0 belongs to the Kullback-Leibler support of the prior, then, for each n, $P_0^n(\{x^n : m_n^\alpha(x^n) \in \{0, \infty\}\}) = 0$, a consequence of Lemma 1 by Barron *et al.* (1999, page 541). In the sequel, for ease of notation, also Q_n^α will be used to indicate the pseudo-posterior distribution.

Given x^n, let $\mathcal{L}(f_P; x^n)$ stand for the likelihood function. Q_n^α can be regarded either as the resulting distribution when the pseudo-likelihood function

$$\mathcal{L}^\alpha(f_P; x^n) \propto [\mathcal{L}(f_P; x^n)]^\alpha, \qquad P \in \mathcal{F},$$

is adopted, or as the true posterior based on the data-dependent prior

$$\pi^\alpha(A; x^n) = \frac{\int_A f_P^n(x^n)^{-(1-\alpha)} \, d\pi(P)}{\int_{\mathcal{F}} f_P^n(x^n)^{-(1-\alpha)} \, d\pi(P)}, \qquad A \in \mathcal{C}, \tag{1.4}$$

provided that, for each n, $P_0^n(\{x^n : \int_{\mathcal{F}} f_P^n(x^n)^{-(1-\alpha)} \, d\pi(P) \in (0, \infty)\}) = 1$. We show that the consistency result of Walker *et al.* (2001, page 5, Theorem 1) for the case $\alpha = 1/2$ extends to every $\alpha \in (0, 1)$. By using Q_n^α, assumptions on the prior assuring consistency can be remarkably weakened.

It seems intuitively plausible that the closer is α to one, the closer are the data-dependent prior and the derived posterior to the real ones. We show that for α sufficiently close to one, the true posterior is approximated by the pseudo-posterior with arbitrarily small error in terms of total variation distance, and, that, contextually, the data-dependence of $\pi^{\alpha}(\cdot;\, x^n)$ disappears. This provides a theoretical justification for a direct use of the pseudo-posterior distribution in those cases when no conclusive assertion on consistency of the true posterior can be easily reached. Then, inference can be based on the pseudo-posterior itself.

In non-parametric problems, posteriors may converge at suboptimal rates. We study the rate at which the pseudo-posterior probability of a shrinking Hellinger neighborhood of the sampling distribution tends to one. We show that $Q_n^{\alpha_n}$ converges at least as fast as $\pi(\cdot;\, x^n)$. The open question remains how to choose priors that lead to posteriors converging at good rates. This question is crucial for the pseudo-posterior, whose rate of convergence is shown to be exclusively driven by the concentration rate of the prior distribution on suitable neighborhoods of P_0.

Theoretical results on the asymptotic behavior of Q_n^{α} are used to study posterior consistency of location mixtures of normal densities when the scale parameter is assigned an *ad hoc* chosen sample-size dependent prior. If the mixing distribution is the trajectory of a Dirichlet process, then the posterior is shown to converge at the best known rate $n^{-1/2}(\ln n)$.

The paper is organized as follows. In Section 2, consistency results are obtained. In Section 3, we study the limiting behavior of Q_n^{α} for α tending to one, while in Section 4, the asymptotic behavior of $Q_n^{\alpha_n}$ is considered. Section 5 reports results on the rate of convergence of $Q_n^{\alpha_n}$. Section 6 is devoted to illustrate the case of mixture models. Final remarks and comments are contained in Section 7. Technical proofs and auxiliary results are deferred to the appendices.

2. Consistency results

In this section the consistency result of Walker *et al.* (2001) is extended to the case when α is any number in the open unit interval. The choice $\alpha = 1/2$ well agrees with the Hellinger distance. This metric is a member of a family of divergences, whose definition is introduced. A preliminary result is reported, references are Liese *et al.* (1987) and Vajda (1989).

Definition 2.1. For any $\alpha \in (0,\, 1)$, the *I_α-divergence* of P and Q is defined as

$$I_\alpha(P\|Q) = \frac{1}{\alpha(1-\alpha)}\left[1 - \int_{\mathcal{X}} f_P^{\alpha} f_Q^{1-\alpha}\, \mathrm{d}\lambda\right],$$

where $\int_{\mathcal{X}} f_P^{\alpha} f_Q^{1-\alpha}\, \mathrm{d}\lambda$ is called the *Hellinger integral* of order α.

Note that $I_{1/2}(P\|Q)$ is the only member of the family which is symmetric, moreover, $I_{1/2}(P\|Q) = 2d_H(P,\, Q)^2$.

Proposition 2.2. *For each pair $P,\, Q \in \mathcal{F}$,*

$$\gamma I_\gamma(P\|Q) \le \delta I_\delta(P\|Q), \qquad \text{for } 0 < \gamma \le \delta < \infty, \tag{2.1}$$

$$(1-\delta) I_\delta(P\|Q) \le (1-\gamma) I_\gamma(P\|Q), \qquad \text{for } -\infty < \gamma \le \delta < 1. \tag{2.2}$$

Theorem 2.3. *Let $\alpha \in (0, 1)$ be fixed. If $P_0 \in \mathcal{S}_K$, then, for each $\varepsilon > 0$, $Q^\alpha(H_\varepsilon^c|x^n) \to 0$ exponentially fast, as $n \to \infty$, with P_0^∞-probability one.*

Proof. Choose $\varepsilon > 0$ and write

$$Q^\alpha(H_\varepsilon^c|x^n) = \frac{\int_{H_\varepsilon^c} R(f_P;\, x^n)^\alpha \, \mathrm{d}\pi(P)}{\int_{\mathcal{F}} R(f_P;\, x^n)^\alpha \, \mathrm{d}\pi(P)} = \frac{N_n(x^n)}{L_n(x^n)}, \qquad x^n \in \mathcal{X}^n, \quad n \geq 1,$$

where the likelihood ratio $R(f_P;\, x^n) = (f_P^n(x^n)/f_0^n(x^n))$ is defined because, for each n, $P_0^n(\{x^n : f_0^n(x^n) \in \{0,\, \infty\}\}) = 0$, see Lemma 2 by Barron *et al.* (1999, page 541). Consider the denominator L_n. If, for each n, $x^n \in \mathcal{X}^n$ and $P \in \mathcal{F}$, the position $K(f_P;\, x^n) = n^{-1}\sum_{i=1}^n \ln(f_0(x_i)/f_P(x_i))$ is set, $L_n(x^n)$ can be rewritten as $L_n(x^n) = \int_{\mathcal{F}} \exp\{-n\alpha K(f_P;\, x^n)\}\, \mathrm{d}\pi(P)$. For any $\eta > 0$, arguing as in Lemma 4 of Barron *et al.* (1999, pages 542-543), we draw that $L_n(x^n) > \exp\{-n\eta\}$ for all large n, a.s. $[P_0^\infty]$.

Now, consider the numerator N_n. With regard to the possible values of α, two cases can be distinguished: $0 < \alpha \leq 1/2$ and $1/2 \leq \alpha < 1$. To fix ideas, the case $1/2 \leq \alpha < 1$ is treated in detail; the other can be dealt with analogously. For every $c > 0$, by applying Markov's inequality, the inequality $\ln(1-x) \leq -x$, for $0 \leq x < 1$, and then taking into account (2.1), with $\gamma = 1/2$, and $I_{1/2}(P\|Q)/2 = d_H(P,\, Q)^2$, the following chain of inequalities is seen to hold

$$\begin{aligned}
P_0^n(\{x^n : N_n(x^n) \geq \exp\{-nc\}\}) &\leq \exp\{nc\} \int_{H_\varepsilon^c} \left\{ \mathrm{E}_0\left[\left(\frac{f_P(X_1)}{f_0(X_1)} \right)^\alpha \right] \right\}^n \mathrm{d}\pi(P) \\
&= \exp\{nc\} \int_{H_\varepsilon^c} \left\{ \int_{\mathcal{X}} f_P^\alpha f_0^{1-\alpha} \, \mathrm{d}\lambda \right\}^n \mathrm{d}\pi(P) \\
&= \exp\{nc\} \int_{H_\varepsilon^c} \{1 - \alpha(1-\alpha) I_\alpha(P\|P_0)\}^n \, \mathrm{d}\pi(P) \\
&\leq \exp\{nc\} \int_{H_\varepsilon^c} \exp\{-n\alpha(1-\alpha) I_\alpha(P\|P_0)\} \, \mathrm{d}\pi(P) \\
&\leq \exp\{nc\} \int_{H_\varepsilon^c} \exp\left\{-n\frac{(1-\alpha)}{2} I_{1/2}(P\|P_0)\right\} \mathrm{d}\pi(P) \\
&= \exp\{nc\} \int_{H_\varepsilon^c} \exp\{-n(1-\alpha) d_H(P_0,\, P)^2\} \, \mathrm{d}\pi(P) \\
&\leq \exp\{n[c - (1-\alpha)\varepsilon^2]\},
\end{aligned}$$

where, for each $\alpha \in (0, 1)$, $I_\alpha(\cdot\|P_0)$ is $\mathcal{C}$-measurable. More precisely, $I_\alpha(0) + I_\alpha(\infty)/\!/\infty = [\alpha(1-\alpha)]^{-1} < \infty$, therefore $I_\alpha(\cdot\|P_0)$ is continuous in the total variation topology, hence measurable, see Vajda (1989, pages 272-273, Theorem 9.28). For $0 < \alpha \leq 1/2$, by applying (2.2), with $\delta = 1/2$,

$$P_0^n(\{x^n : N_n(x^n) \geq \exp\{-nc\}\}) \leq \exp\{n[c - \alpha\varepsilon^2]\}, \qquad n \geq 1.$$

If $1-\alpha$ and α are renamed to β, with $\beta \in (0,\, 1/2]$, then the two cases can be given a unified treatment for the rest of the proof. Chosen c to be $0 < c < \beta\varepsilon^2$, the series $\sum_{n=1}^{\infty} \exp\{-n[\beta\varepsilon^2 - c]\} < \infty$, so that, by the first Borel-Cantelli lemma,

$$P_0^{\infty}\left(\left\{x^{\infty} :\, \int_{H_\varepsilon^c} R(f_P;\, x^n)^{\alpha}\, \mathrm{d}\pi(P) \geq \exp\{-nc\} \text{ i.o.}\right\}\right) = 0, \tag{2.3}$$

namely, $N_n(x^n) < \exp\{-nc\}$ for all large n, a.s. $[P_0^{\infty}]$. Now, combining bounds on N_n and L_n, it is $Q^{\alpha}(H_\varepsilon^c|x^n) < \exp\{-n[c-\eta]\}$ for all large n, a.s. $[P_0^{\infty}]$. Set $\delta = c - \eta$. Then, for any $0 < \eta < c$, it is $\delta > 0$, and the result follows. □

For $\alpha = 1/2$, the theorem specializes to Theorem 1 of Walker *et al.* (2001, page 814). What is surprising with this result is the fact that, apart from the support condition required to guarantee that the denominator remains exponentially bounded below, no condition is needed to upper bound the numerator as long as α is strictly less than one. For $\alpha = 1$, that is to say when the true posterior distribution is considered, unless the parameter space is totally bounded for the Hellinger distance, hence relatively compact, in general consistency does not hold without additional conditions. Roughly speaking, there is a sort of "discontinuity" in the asymptotic behavior of Q_n^{α} with respect to the possible values of α.

The expectation of f_P with respect to Q_n^{α} defines an estimator of f_0

$$\widehat{f}_n^{\alpha}(x) = \int_{\mathcal{F}} f_P(x)\, \mathrm{d}Q^{\alpha}(P|x^n), \qquad x \in \mathcal{X}.$$

It is straightforward to verify that, for each n, with P_0^n-probability one, $\widehat{f}_n^{\alpha}$ is a density function.

Proposition 2.4. *Let $\alpha \in (0,\, 1)$ be fixed. If $P_0 \in \mathcal{S}_K$, then, $\widehat{f}_n^{\alpha} \xrightarrow{L_1} f_0$, as $n \to \infty$, with P_0^{∞}-probability one.*

Proof. In view of (1.1), the assertion is proved if it is shown that $d_H(f_0,\, \widehat{f}_n^{\alpha}) \to 0$, as $n \to \infty$, a.s. $[P_0^{\infty}]$. Let $\mathcal{X}_0^{\infty} = \{x^{\infty} :\, Q_n^{\alpha} \Rightarrow_n \delta_{P_0}\}$. It descends from Theorem 2.3 that $P_0^{\infty}(\mathcal{X}_0^{\infty}) = 1$. Fix $x^{\infty} \in \mathcal{X}_0^{\infty}$. For each $\varepsilon > 0$, in force of the same theorem, the following chain of inequalities holds true for a suitable $\delta > 0$:

$$d_H(f_0,\, \widehat{f}_n^{\alpha})^2 \leq \int_{H_\varepsilon \cup H_\varepsilon^c} d_H(f_0,\, f_P)^2\, \mathrm{d}Q^{\alpha}(P|x^n) \leq \varepsilon^2 + 2\exp\{-n\delta\}$$

for all large n. Thus, along the fixed sample path x^{∞}, $\lim_{n\to\infty} d_H(f_0,\, \widehat{f}_n^{\alpha}) = 0$, and this holds for all $x^{\infty} \in \mathcal{X}_0^{\infty}$. The assertion follows. □

Note that under the condition $P_0 \in \mathcal{S}_K$, the pseudo-Bayes density estimator $\widehat{f}_n^{\alpha}$ is L_1-consistent at f_0 and, *a fortiori*, L_1-consistent in the Cesàro sense, whereas, the estimator $\widehat{f}_n = \int_{\mathcal{X}} f_P\, \mathrm{d}\pi(P|x^n)$ is only L_1-consistent in the Cesàro sense.

3. Limiting behavior of Q_n^α with respect to α

Under the only assumption that the prior assigns positive mass to all Kullback-Leibler neighborhoods of P_0, Q_n^α is consistent at P_0, no matter how "close" is α to one. The same requirement alone is not sufficient for π_n to be consistent. It is thus natural to inquire about the kind of relationship existing between corresponding terms of the two sequences of distributions. For each n, one expects that, when α tends to one, Q_n^α "tends" to π_n, in a sense to be made precise. Here, it is shown that Q_n^α converges to π_n in the total variation distance. Analogue results hold for the data-dependent prior and the pseudo-Bayes density estimator.

Theorem 3.1. *If $P_0 \in \mathcal{S}_K$, then*

$$P_0^\infty(\{x^\infty : \forall n \in \mathbb{N},\ \lim_{\alpha\to 1} d_{TV}(\pi_n, Q_n^\alpha) = 0\}) = 1.$$

Proof. The assertion holds true if and only if, for each n,

$$P_0^n(\{x^n : \lim_{\alpha\to 1} d_{TV}(\pi_n, Q_n^\alpha) = 0\}) = 1.$$

Let n be fixed and let $\mathcal{X}_1^n = \{x^n : m_n(x^n) \in (0, \infty)\}$. From Lemma 1 of Barron *et al.* (1999, page 541), it is known that $P_0^n(\mathcal{X}_1^n) = 1$. Let $x^n \in \mathcal{X}_1^n$ be fixed. For an easier exposition, the conditioning on x^n will be omitted. It is $\lim_{\alpha\to 1} d_{TV}(\pi_n, Q_n^\alpha) = 0$ if and only if, for each real-valued sequence $\{\alpha_k\}_{k=1}^\infty$, with all $\alpha_k \in (0, 1)$ and $\alpha_k \to 1$, as $k \to \infty$, it holds true that $\lim_{k\to\infty} d_{TV}(\pi_n, Q_n^{\alpha_k}) = 0$, see, e.g., Checcucci *et al.* (1968, page 175, Proposition 1.6). Let $\{\alpha_k\}_{k=1}^\infty$ be any sequence as above. For each k and any Borel set A, let

$$Q_n^{\alpha_k}(A) = \int_A \frac{(f_P^n)^{\alpha_k}}{m_n^{\alpha_k}}\, \mathrm{d}\pi(P) = \int_A \delta_k(P)\, \mathrm{d}\pi(P),$$

$$\pi_n(A) = \int_A \frac{f_P^n}{m_n}\, \mathrm{d}\pi(P) = \int_A \delta(P)\, \mathrm{d}\pi(P),$$

having set the positions $\delta_k(P) = (f_P^n)^{\alpha_k}/m_n^{\alpha_k}$ and $\delta(P) = f_P^n/m_n$, where $m_n^{\alpha_k} \in (0, \infty)$ as a consequence of Lemma 1 by Barron *et al.* (1999, page 541). By definition, $d_{TV}(\pi_n, Q_n^{\alpha_k}) = \sup_{A\in\mathcal{C}} |\pi_n(A) - Q_n^{\alpha_k}(A)|$. Since $Q_n^{\alpha_k}(\mathcal{F}) = \pi_n(\mathcal{F}) = 1 < \infty$ for $k = 1, 2, \ldots$, in order to appeal to Scheffé's theorem, see, e.g., Billingsley (1995, page 215, Theorem 16.12), it suffices to show that $\delta_k \to \delta$, as $k \to \infty$, a.s. $[\pi]$. First, note that

$$\lim_{k\to\infty} (f_P^n)^{\alpha_k} = f_P^n, \qquad \text{a.s. } [\pi]. \tag{3.1}$$

In order to show that $m_n^{\alpha_k}$ converges to m_n, as $k \to \infty$, it is checked that the hypotheses of the Lebesgue's dominated convergence theorem are fulfilled. From the inequality $x^\alpha < 2(1+x)$, valid for all $\alpha \in (0, 1]$ and $x \geq 0$, it follows that, for each k, $(f_P^n)^{\alpha_k} < 2(1 + f_P^n)$ for all $P \in \mathcal{F}$, where $2(1 + f_P^n)$ is π-integrable (recall that $m_n < \infty$). Taking into account (3.1), it results

$$\lim_{k\to\infty} m_n^{\alpha_k} = \lim_{k\to\infty} \int_{\mathcal{F}} (f_P^n)^{\alpha_k}\, \mathrm{d}\pi(P) = \int_{\mathcal{F}} f_P^n\, \mathrm{d}\pi(P) = m_n. \tag{3.2}$$

By combining (3.1) and (3.2), it is seen that $\delta_k \to \delta$, as $k \to \infty$, a.s. $[\pi]$. Now, by applying Scheffé's theorem,

$$\sup_{A\in\mathcal{C}} |\pi_n(A) - Q_n^{\alpha_k}(A)| \le \int_{\mathcal{F}} |\delta(P) - \delta_k(P)|\, \mathrm{d}\pi(P) \to 0, \qquad (k \to \infty),$$

whence $\lim_{k\to\infty} d_{TV}(\pi_n, Q_n^{\alpha_k}) = 0$. This holds for every sequence $\{\alpha_k\}_{k=1}^{\infty}$, hence

$$\lim_{\alpha\to 1} d_{TV}(\pi_n, Q_n^{\alpha}) = 0.$$

The same reasoning applies to every x^n in $\mathcal{X}_1^n$. This completes the proof. □

By taking into account the previous finding, for each n, the posterior probability of any Borel set can be approximated with an arbitrarily small error, in terms of total variation distance, by choosing α sufficiently close to one. This provides a justification for replacing π_n with Q_n^{α}. In the following statement, a consequence for $\widehat{f}_n^{\alpha}$ is derived.

Corollary 3.1. *If $P_0 \in \mathcal{S}_K$, then*

$$P_0^{\infty}(\{x^{\infty} : \forall n \in \mathbb{N},\ \lim_{\alpha\to 1} \|\widehat{f}_n - \widehat{f}_n^{\alpha}\|_1 = 0\}) = 1.$$

Proof. The basic arguments parallel those of Theorem 3.1. For fixed n, let $\mathcal{X}_1^n$ be the same set as above. Fix $x^n \in \mathcal{X}_1^n$. It suffices to show that for each sequence $\{\alpha_k\}_{k=1}^{\infty}$ as before, $\lim_{k\to\infty} \|\widehat{f}_n - \widehat{f}_n^{\alpha_k}\|_1 = 0$. This descends from the following chain of inequalities, where the fact that $\delta_k \to \delta$, as $k \to \infty$, π-almost surely, a by-product of Theorem 3.1, is considered. Then,

$$\begin{aligned} 0 \le \limsup_{k\to\infty} \left\| \widehat{f}_n - \widehat{f}_n^{\alpha_k} \right\|_1 &= \limsup_{k\to\infty} \int_{\mathcal{X}} \left| \widehat{f}_n - \widehat{f}_n^{\alpha_k} \right| \mathrm{d}\lambda \\ &= \limsup_{k\to\infty} \int_{\mathcal{X}} \left| \int_{\mathcal{F}} f_P [\delta(P) - \delta_k(P)]\, \mathrm{d}\pi(P) \right| \mathrm{d}\lambda \\ &\le \limsup_{k\to\infty} \int_{\mathcal{X}} \int_{\mathcal{F}} f_P\, |\delta(P) - \delta_k(P)|\, \mathrm{d}\pi(P)\, \mathrm{d}\lambda \\ &\le \int_{\mathcal{X}} \int_{\mathcal{F}} f_P \limsup_{k\to\infty} |\delta(P) - \delta_k(P)|\, \mathrm{d}\pi(P)\, \mathrm{d}\lambda = 0. \end{aligned}$$

□

The next assertion provides an analogue result for the data-dependent prior.

Theorem 3.2. *Suppose that, for each n, $P_0^n(\{x^n : f_P^n(x^n) > 0$ a.s. $[\pi]\}) = 1$. Then,*

$$P_0^\infty(\{x^\infty : \forall n \in \mathbb{N},\ \lim_{\alpha\to 1} d_{TV}(\pi,\ \pi_n^\alpha) = 0\}) = 1.$$

Proof. In account of the hypothesis, for each n, $P_0^n(\{x^n : f_P^n(x^n) \in (0,\ \infty)$ a.s. $[\pi]\}) = 1$ and also, for each $\alpha \in (0,\ 1)$, $P_0^n(\{x^n : \int_{\mathcal{F}} f_P^n(x^n)^{-(1-\alpha)}\,\mathrm{d}\pi(P) \in (0,\ \infty)\}) = 1$ so that π_n^α is well-defined. Fix any x^n in the former set. As in the proof of Theorem 3.1, the key step consists in showing that, for each sequence $\{\alpha_k\}_{k=1}^\infty$ such that $\alpha_k \in (0,\ 1)$ for all k, and $\alpha_k \to 1$, as $k \to \infty$, $\lim_{k\to\infty} d_{TV}(\pi,\ \pi_n^{\alpha_k}) = 0$. In virtue of Scheffé's theorem, for each k,

$$\begin{aligned} d_{TV}(\pi,\ \pi_n^{\alpha_k}) &= \sup_{A\in\mathcal{C}} |\pi(A) - \pi_n^{\alpha_k}(A)| \\ &\leq \int_{\mathcal{F}} \left| 1 - \frac{f_P^n(x^n)^{-(1-\alpha_k)}}{\int_{\mathcal{F}} f_P^n(x^n)^{-(1-\alpha_k)}\,\mathrm{d}\pi(P)} \right| \mathrm{d}\pi(P). \end{aligned}$$

It is therefore enough to show that

$$\lim_{k\to\infty} \frac{f_P^n(x^n)^{-(1-\alpha_k)}}{\int_{\mathcal{F}} f_P^n(x^n)^{-(1-\alpha_k)}\,\mathrm{d}\pi(P)} = 1, \qquad \text{a.s. } [\pi].$$

This follows from the fact that $\lim_{k\to\infty} f_P^n(x^n)^{-(1-\alpha_k)} = 1$ a.s. $[\pi]$, and

$$\begin{aligned} 1 = \int_{\mathcal{F}} \liminf_{k\to\infty} f_P^n(x^n)^{-(1-\alpha_k)}\,\mathrm{d}\pi(P) &\leq \liminf_{k\to\infty} \int_{\mathcal{F}} f_P^n(x^n)^{-(1-\alpha_k)}\,\mathrm{d}\pi(P) \\ &\leq \limsup_{k\to\infty} \int_{\mathcal{F}} f_P^n(x^n)^{-(1-\alpha_k)}\,\mathrm{d}\pi(P) \\ &\leq \int_{\mathcal{F}} \limsup_{k\to\infty} f_P^n(x^n)^{-(1-\alpha_k)}\,\mathrm{d}\pi(P) = 1. \end{aligned}$$

□

4. Consistency of $Q_n^{\alpha_n}$

A sequence of pseudo-posterior distributions with a constant α has been considered so far. In what follows, the asymptotic behavior of a sequence $Q_n^{\alpha_n}$, obtained by taking a sample-size dependent sequence $\{\alpha_n\}_{n=1}^\infty$ which converges to one as n goes to infinity, is studied. Conditions under which α_n converges to one, meanwhile $Q_n^{\alpha_n}$ is consistent, are established. Interest in this study stems from the observation that, when α_n is close to one, the influence of data over the data-dependent prior is disappearing.

As a consequence of Lemma 1 of Barron *et al.* (1999, page 541), it is seen that, if $P_0 \in \mathcal{S}_K$, then $P_0^\infty(\{x^\infty : \exists n \in \mathbb{N},\ m_n^{\alpha_n}(x^n) \in \{0,\ \infty\}\}) = 0$ and $Q_n^{\alpha_n}$ is well-defined.

By inspecting the proof of Theorem 2.3, it is seen that, when α_n tends to one, different arguments are needed to upper bound the numerator of $Q^{\alpha_n}(H_\varepsilon^c|x^n)$. To this aim, a couple of inequalities concerning the I_α-divergences are established.

Let $\mathcal{P}$ be the set of all probability measures on $(\mathcal{X}, \mathcal{B})$, while $\mathcal{F}$ is the subset comprising all probability measures dominated by a common σ-finite measure λ. For any pair P and Q in $\mathcal{P}$, let $V(P, Q) = 2d_{TV}(P, Q)$. Recall that if $P, Q \in \mathcal{F}$, then $V(P, Q) = \|f_P - f_Q\|_1$. Let $\alpha \in (0, 1)$ be fixed and let the function

$$(\underline{I_\alpha})_V : [0, 2] \longrightarrow \left[0, \frac{1}{\alpha(1-\alpha)}\right]$$

be defined as

$$(\underline{I_\alpha})_V(v) = \inf_{\{(P, Q) \in \mathcal{P}\times\mathcal{P}:\, V(P, Q) = v\}} I_\alpha(P\|Q).$$

It is known that the inequality

$$(4.1) \qquad (\underline{I_\alpha})_V(v) \geq \frac{1}{\alpha(1-\alpha)}\left[1 - \left(1 + \frac{v}{2}\right)^{\alpha\wedge(1-\alpha)}\left(1 - \frac{v}{2}\right)^{\alpha\vee(1-\alpha)}\right]$$

holds true for all $v \in [0, 2]$, see, e.g., Vajda (1989, page 294). From the strict convexity of the function $I_\alpha(\cdot)$ on $(0, \infty)$, it follows that $(\underline{I_\alpha})_V(\cdot)$ is continuous and increasing on the domain $[0, 2]$, Vajda (*ibidem*, pages 289-292, Proposition 9.49(i)). A lemma is stated.

Lemma 4.1. *Let $P_0 \in \mathcal{F}$. For any $v \in [0, 2]$, let $U_v^c = \{P \in \mathcal{F} : \|f_0 - f_P\|_1 \geq v\}$. For a fixed $\alpha \in (0, 1)$, it results*

$$(4.2) \qquad I_\alpha(P_0\|P) \geq \frac{1}{\alpha(1-\alpha)}\left[1 - \left(1 + \frac{v}{2}\right)^{\alpha\wedge(1-\alpha)}\left(1 - \frac{v}{2}\right)^{\alpha\vee(1-\alpha)}\right]$$

for all $P \in U_v^c$. Consequently, if $\alpha > \frac{1}{2}$,

$$(4.3) \qquad I_\alpha(P_0\|P) \geq \frac{1}{\alpha(1-\alpha)}\left[(2\alpha - 1)\frac{v}{2}\right], \qquad P \in U_v^c.$$

Proof. Inequality (4.2) follows from (4.1) by straightforward calculations. □

Theorem 4.2. *Let $\{\alpha_n\}_{n=1}^\infty$ be any real-valued sequence such that $\alpha_n \in (0, 1)$ for all n and $\alpha_n \to 1$, as $n \to \infty$. If $P_0 \in \mathcal{S}_K$, then, for each $\varepsilon > 0$, $\lim_{n\to\infty} Q^{\alpha_n}(H_\varepsilon^c|x^n) = 0$ a.s. $[P_0^\infty]$. The convergence has exponential rate.*

Proof. The proof follows the same pattern as the proof of Theorem 2.3: the only step where some changes are required is when an exponential upper bound on $\int_{H_\varepsilon^c} R(f_P; x^n)^{\alpha_n}\, d\pi(P)$ has to be established. Recall that $d_H(f_P, f_Q)^2 \leq \|f_P - f_Q\|_1$ for all P, Q in $\mathcal{F}$, cf. (1.1). Let $\varepsilon > 0$ be fixed and let $U_{\varepsilon^2}^c = \{P \in \mathcal{F} : \|f_0 - f_P\|_1 \geq \varepsilon^2\}$. The inclusion $H_\varepsilon^c \subseteq U_{\varepsilon^2}^c$ holds true. By Markov's inequality, for n sufficiently large so that $\alpha_n > \frac{3}{4}$ and, consequently, (4.3) can be invoked, it is

$$
\begin{aligned}
&P_0^n\left(\left\{x^n : \int_{H_\varepsilon^c} R(f_P; x^n)^{\alpha_n}\, \mathrm{d}\pi(P) > \exp\left\{-\frac{n\varepsilon^2}{8}\right\}\right\}\right)\\
&\quad \le \exp\left\{\frac{n\varepsilon^2}{8}\right\}\int_{H_\varepsilon^c}\left\{\mathrm{E}_0\left[\left(\frac{f_P(X_1)}{f_0(X_1)}\right)^{\alpha_n}\right]\right\}^n \mathrm{d}\pi(P)\\
&\quad \le \exp\left\{\frac{n\varepsilon^2}{8}\right\}\int_{U_{\varepsilon^2}^c}\left\{\mathrm{E}_0\left[\left(\frac{f_P(X_1)}{f_0(X_1)}\right)^{\alpha_n}\right]\right\}^n \mathrm{d}\pi(P)\\
&\quad \le \exp\left\{\frac{n\varepsilon^2}{8}\right\}\int_{U_{\varepsilon^2}^c\cap\{P\perp P_0\}^c} \exp\{-n\alpha_n(1-\alpha_n)I_{\alpha_n}(P\|P_0)\}\, \mathrm{d}\pi(P)\\
&\quad \le \exp\left\{\frac{n\varepsilon^2}{8}\right\}\exp\left\{-n\left[1-\left(1+\frac{\varepsilon^2}{2}\right)^{(1-\alpha_n)\wedge\alpha_n}\left(1-\frac{\varepsilon^2}{2}\right)^{(1-\alpha_n)\vee\alpha_n}\right]\right\}\pi(U_{\varepsilon^2}^c\cap\{P\perp P_0\}^c)\\
&\quad \le \exp\left\{\frac{n\varepsilon^2}{8}\right\}\exp\left\{-\frac{n(2\alpha_n-1)\varepsilon^2}{2}\right\}\\
&\quad < \exp\left\{-\frac{n\varepsilon^2}{8}\right\}.
\end{aligned}
$$

By the first Borel-Cantelli lemma,

$$\int_{H_\varepsilon^c} R(f_P; x^n)^{\alpha_n}\, \mathrm{d}\pi(P) \le \exp\left\{-\frac{n\varepsilon^2}{8}\right\}$$

for all large n, along almost all sample paths when sampling from P_0. On other hand, for any $\eta > 0$,

$$\int_{\mathcal{F}} \exp\{-n\alpha_n K(f_P; x^n)\}\, \mathrm{d}\pi(P) > \exp\{-n\eta\}$$

for all large n, a.s. $[P_0^\infty]$. Thus, $Q^{\alpha_n}(H_\varepsilon^c|x^n) < \exp\{-n[\varepsilon^2/8 - \eta]\}$ for sufficiently large n, a.s. $[P_0^\infty]$. Chosen $\eta < \varepsilon^2/8$, the thesis follows. □

It is surprising that no condition on the rate of convergence of α_n need to be required. The Kullback-Leibler support condition is strong enough to yield consistency even when the exponent is allowed to vary with sample size.

5. Convergence rate of $Q_n^{\alpha_n}$

The motivating reason for considering $Q_n^{\alpha_n}$ is that consistency can be established under mild conditions on the prior. However, consistency is by itself a weak property; it is desirable that $Q_n^{\alpha_n}$ enjoys other large sample properties as well. If it is studied how $Q_n^{\alpha_n}$ concentrates in shrinking neighborhoods of P_0, then the rate of convergence can be assessed and compared with that of the true posterior. Moreover, a pseudo-posterior distribution converging at a good rate yields a density estimator converging at least as quickly.

We say that the (point-wise) rate of convergence of $Q_n^{\alpha_n}$ is ε_n if, for every sequence K_n such that $K_n \to \infty$ and $K_n\varepsilon_n \to 0$, as $n \to \infty$, $Q^{\alpha_n}(H_{\varepsilon_n}^c|x^n)$ tends to zero in probability or almost surely when sampling from P_0.

The problem addressed is the following: determine the rate of convergence of $Q_n^{\alpha_n}$, in other terms, establish the speed at which H_{ε_n} shrinks to P_0, meanwhile still capturing almost all the posterior probability mass. We show that the rate of convergence of $Q_n^{\alpha_n}$ depends only on the prior assignment in suitable neighborhoods of P_0.

When the parameter space is totally bounded for the Hellinger distance or it is a countable union of totally bounded subspaces, general results on rates of convergence of posterior distributions are available from the works of Ghosal, Ghosh *et al.* (2000) and Shen *et al.* (2001). Convergence rates of posterior distributions are driven by two factors: the "size" r_n of the parameter space, as measured by some entropy number, and the concentration rate t_n of the prior mass in appropriate neighborhoods of P_0. More precisely, posterior distributions converge at rate $\varepsilon_n = \max\{t_n^{1/2}, r_n\}$. If t_n is large, that is a small amount of probability is given a priori to neighborhoods of P_0, and $t_n^{1/2} > r_n$, then the posterior converges at a suboptimal rate $\varepsilon_n = t_n^{1/2}$. In fact, if r_n is the minimal possible value satisfying the bracketing integral equation, then it is the best achievable rate by sieve maximum likelihood estimators, see apropos Wong *et al.* (1995, pages 348-350, Theorem 1). In this case there is no advantage in replacing the posterior with the pseudo-posterior: they converge at the same rate. In principle, the pseudo-posterior may be of help in cases which cannot be easily handled by appealing to the above mentioned results.

The way we attack the problem addressed closely follows the approach taken by the above cited authors to determine rates of convergence of posterior distributions. In view of Theorem 4.2, in order to find an exponential upper bound on $Q^{\alpha_n}(H_{\varepsilon_n}^c|x^n)$, rewritten as the ratio N_n/L_n, we only need to bound the denominator. The numerator stays bounded above along almost all sample paths when sampling from P_0. We expect that conditions implying an exponential lower bound on the denominator of the true posterior would imply the same bound on L_n, the factor α_n being absorbed. Thus, efforts are concentrated in searching lower bounds on L_n.

5.1. **Main Results.** We show that $Q_n^{\alpha_n}$ converges at a rate that is equal to the square rooted radius of balls centered at P_0 which are given a priori probability mass not exponentially small.

We need to introduce a Kullback-Leibler type neighborhood of P_0. For any $t > 0$, let $M(t) = \{P \in \mathcal{F} : \max\{K(P_0\|P), D(P_0, P)\} < t\}$, where $D(P_0, P) = \int_{\mathcal{X}} f_0 \left(\ln(f_0/f_P)\right)^2 \, d\lambda - K(P_0\|P)^2$. In the sequel, for a positive sequence t_n, let M_n stand for $M(t_n)$. A further convention. When writing $a_n \gtrsim b_n$, we mean that the inequality $a_n \geq cb_n$ holds for all sufficiently large n, with $c > 0$ a constant that is fixed throughout. Analogue meaning has $a_n \lesssim b_n$. If both $a_n \gtrsim b_n$ and $a_n \lesssim b_n$ hold true, then we write $a_n \asymp b_n$.

Theorem 5.1. *Suppose there exists a positive sequence $\{t_n\}_{n=1}^{\infty}$ such that $t_n \to 0$ and $nt_n \to \infty$, as $n \to \infty$. Let $\{\alpha_n\}_{n=1}^{\infty}$ be a sequence such that $\alpha_n \in (0, 1)$ and $\alpha_n \to 1$, as $n \to \infty$. If $\pi(M_n) \succeq e^{-2nt_n}$, then, for $\varepsilon_n = t_n^{1/2}$ and a large enough constant $K > 0$, $Q^{\alpha_n}(H_{\varepsilon_n}^c|x^n) \lesssim \exp\{-K^2 n\varepsilon_n^2/16\}$ on a Borel set of P_0^n-probability tending to one.*

Proof. In force of Lemma 1 by Shen *et al.* (2001, pages 690-691) and in virtue of the hypothesis on the prior assignment to M_n, it turns out to be $L_n(x^n) \gtrsim e^{-4nt_n}$, except on set of P_0^n-probability tending to zero.

The numerator N_n can be bounded above in probability reasoning as in Theorem 4.2: just replace ε with $K\varepsilon_n$, where K is any positive constant and $\{\varepsilon_n\}_{n=1}^{\infty}$ any positive sequence such that $\varepsilon_n \to 0$ and $n\varepsilon_n^2 \to \infty$, as $n \to \infty$. Therefore, $N_n(x^n) \lesssim e^{-K^2 n\varepsilon_n^2/8}$ in P_0^n-probability. In account of the position $\varepsilon_n^2 = t_n$, for K sufficiently large so that $4t_n < K^2\varepsilon_n^2/16$, it turns out to be

$$\begin{aligned} Q^{\alpha_n}(H_{\varepsilon_n}^c|x^n) &\lesssim \frac{\exp\{-K^2 n\varepsilon_n^2/8\}}{\exp\{-4nt_n\}} \\ &\lesssim \exp\left\{-\frac{K^2 n\varepsilon_n^2}{16}\right\}, \end{aligned}$$

on a set of P_0^n-probability tending to one. □

The condition $\pi(M_n) \succeq e^{-2nt_n}$ can be replaced by similar conditions involving different neighborhoods of P_0, see Lemma A.1, Lemma A.3 and Lemma A.4. All results agree to say that if certain shrinking balls centered at P_0 are given enough prior mass, then $Q_n^{\alpha_n}$ converges at a rate equal to their square rooted radii. Hence, ε_n is governed by the prior assignment only. A pseudo-posterior with rate ε_n leads to a pseudo-Bayes estimator converging to f_0 in the Hellinger distance at least as fast as ε_n. Thus, ε_n is an upper bound on the actual rate of convergence of $\widehat{f}_n^{\alpha_n}$.

Proposition 5.2. *Under the conditions of Theorem 5.1,* $d_H(f_0, \widehat{f}_n^{\alpha_n})^2 \preceq \varepsilon_n^2 + 2\,e^{-K^2 n\varepsilon_n^2/16}$ *in probability.*

Proof. Bringing obvious changes to the proof of Proposition 2.4, it is seen that

$$d_H(f_0, \widehat{f}_n^{\alpha_n})^2 \leq \varepsilon_n^2\, Q^{\alpha_n}(H_{\varepsilon_n}|x^n) + 2\,Q^{\alpha_n}(H_{\varepsilon_n}^c|x^n) \leq \varepsilon_n^2 + 2\exp\{-K^2 n\varepsilon_n^2/16\}$$

for large n, on a set of P_0^n-probability tending to one . □

If, on the one hand, $\widehat{f}_n^{\alpha_n}$ converges at least as quickly as $Q_n^{\alpha_n}$, on the other hand, the pseudo-posterior distribution cannot converge at a rate faster than the minimax optimal rate for estimators. This argument can be helpful to conclude that $Q_n^{\alpha_n}$ converges at the best possible rate.

Remark 1. It is striking that as long as each term α_n is strictly less than one, the sequence $\{\alpha_n\}_{n=1}^{\infty}$ can be chosen in a completely free way. Thus, the case $\alpha_n = \alpha$, briefly considered by Walker *et al.* (2001, page 815), is contemplated by the present treatment.

Remark 2. Suppose that the posterior π_n converges at rate $\varepsilon_n^{\pi_n} = \max\{r_n, t_n^{1/2}\}$ and $Q_n^{\alpha_n}$ at rate $\varepsilon_n^{Q_n} = t_n^{1/2}$. Then, $\varepsilon_n^{Q_n} \leq \varepsilon_n^{\pi_n}$. Two cases can be distinguished.

If $t_n^{1/2} > r_n$, then $\varepsilon_n^{\pi_n} = t_n^{1/2} = \varepsilon_n^{Q_n}$, that is, π_n and $Q_n^{\alpha_n}$ converge at the same suboptimal rate.

If $t_n^{1/2} \le r_n$, then $\varepsilon_n^{\pi_n} = r_n$ and $\varepsilon_n^{Q_n} = t_n^{1/2}$. Suppose that r_n is the optimal rate for the model in the minimax sense. On the one hand, $\varepsilon_n^{Q_n} \le r_n$, on the other hand, $\varepsilon_n^{Q_n} \ge r_n$, because r_n is the best possible rate for estimators in the Hellinger distance, hence, it necessarily is $\varepsilon_n^{Q_n} = \varepsilon_n^{\pi_n} = r_n$. This means that $t_n^{1/2}$ cannot be faster than r_n. This means that $t_n^{1/2}$ cannot be faster than r_n. Again, the posterior and the pseudo-posterior have the same rate.

To sum up, when the parameter space is totally bounded for the Hellinger metric, there is no advantage to raise the likelihood to a power α_n: the set is relatively compact and there is no need for any trick to get consistency.

6. Location mixtures of normal densities

A way of defining a prior living on a space of densities is to induce it via a mixture model. In a mixture model, a random density is generated by convoluting a known kernel with a distribution which is given a prior. More precisely, if $F(\cdot)$ is the cumulative distribution function (cdf) of a probability measure P on a space Θ, endowed with its Borel σ-field $\mathcal{A}$, the random "density" has the form

$$f_F(x) = \int_\Theta K(x, \theta)\, \mathrm{d}F(\theta), \qquad x \in \mathcal{X}, \tag{6.1}$$

where $K(\cdot, \cdot)$ is a kernel density on $(\mathcal{X} \times \Theta, \mathcal{B} \otimes \mathcal{A})$, that is a non-negative, measurable function such that, for each $\theta \in \Theta$, $K(\cdot, \theta)$ is a density with respect to λ. From Fubini's theorem, it follows that f_F is indeed a density function. Given P, $X_1, \ldots, X_n$ are i.i.d. observations from f_F. In the sequel, F will be also used to indicate the corresponding probability measure P. If Π denotes a prior for F, a probability measure is induced on the space of densities via the map $F \mapsto f_F$.

This model was introduced by Lo (1984), who studied the problem of density estimation from a Bayesian point of view. He focused attention on the case when the mixing distribution is given a Dirichlet process prior. Note that a Dirichlet process prior cannot be directly used on $\mathcal{F}$, because it "picks" discrete distributions with probability one. He gave the expression of the Bayes density estimator, see also Ghorai *et al.* (1982).

The choice of the kernel is a delicate question: K determines the class of the f_F's when F varies in $\mathcal{P}$. A discussion on this issue can be found in Lo (1984, pages 354-357, §3). A general choice is to take $\sigma^{-1}K((x-\theta)/\sigma)$, with K a symmetric density around 0, typically the normal density. In this case, $\mathcal{X} = \Theta = \mathbb{R}$ and λ is the Lebesgue measure. Let $\phi(\cdot)$ stand for the standard normal density and $\phi_\sigma(\cdot)$ for the density of a normal distribution with zero mean and standard deviation σ. If

$$f_{F,\sigma}(x) = \int_{-\infty}^{\infty} \phi_\sigma(x-\theta)\, \mathrm{d}F(\theta), \qquad x \in \mathbb{R},$$

then the random density is a location mixture of normal densities, formally, is the convolution $\phi_\sigma * F$. The scale parameter σ can be either fixed or random.

Since Ferguson (1983), mixtures of normal densities have been conveniently used in many Bayesian density estimation problems, see West (1992), West *et al.* (1994), Roeder *et al.* (1997). The issue of consistency, however, had been left open until recently, when Ghosal *et al.* (1999) have established sufficient conditions for weak and strong consistency of Dirichlet mixtures of normal densities. Weak consistency is established by verifying Schwartz's support condition. Strong consistency is handled by using a sieve. A different approach is herein taken to prove strong consistency. Developing an idea of Walker *et al.* (2001, pages 818-819), it is illustrated that, for location mixtures of normal densities with *any* prior on $\mathcal{P}$, not necessarily a Dirichlet process prior, and a random scale parameter σ, strong consistency is obtained under the only assumption that the true distribution belongs to the Kullback-Leibler support of the prior, provided that a special sample-size dependent prior for σ is adopted.

Suppose that σ is distributed according to μ and that F is chosen independently of σ according to Π. The measure $\pi = \Pi \times \mu$ on $\mathcal{P} \times (0, \infty)$ induces a prior on $\mathcal{F}$ via the map $(F, \sigma) \mapsto f_{F,\sigma}$. If $(0, \sigma_2]$ is the support of μ, then the prior is concentrated on the set $\cup_{0<\sigma\leq\sigma_2} \mathcal{F}_\sigma$, where $\mathcal{F}_\sigma = \{f_{F,\sigma},\ F \in \mathcal{P}\}$.

Some preliminary facts are stated. Let $\alpha \in (0, 1)$ be fixed. For any pair (F, σ), set the position

$$g_{F,\sigma}(x) = \left\{ \int_{-\infty}^{\infty} \phi_\sigma(x-\theta)^\alpha \, \mathrm{d}F(\theta) \right\}^{1/\alpha}, \qquad x \in \mathbb{R},$$

it is

$$g_{F,\sigma}(x) \leq f_{F,\sigma}(x), \qquad x \in \mathbb{R}. \tag{6.2}$$

Since $\phi_\sigma(z) = [\alpha^\alpha (2\pi)^{-(1-\alpha)}]^{1/2} \sigma^{-(1-\alpha)} \phi_{\sqrt{\alpha}\sigma}(z)^\alpha$, $z \in \mathbb{R}$, $f_{F,\sigma}$ can be rewritten as

$$f_{F,\sigma}(x) = [\alpha^\alpha (2\pi)^{-(1-\alpha)}]^{1/2} \sigma^{-(1-\alpha)} g_{F,\sqrt{\alpha}\sigma}(x)^\alpha, \qquad x \in \mathbb{R}. \tag{6.3}$$

For $\eta > 0$, let

$$G_\eta = \left\{ (F, \sigma) : \int_{-\infty}^{\infty} f_0(x) \ln \frac{f_0(x)}{g_{F,\sqrt{\alpha}\sigma}(x)} \, \mathrm{d}x < \eta \right\}$$

be a Kullback-Leibler type neighborhood of P_0. In order to prove the main result on consistency, densities f_0 such that, for each $\eta > 0$, G_η is assigned a priori positive probability, need to be identified. In force of the inequality

$$\int_{-\infty}^{\infty} f_0(x) \ln \frac{f_0(x)}{f_{F,\sqrt{\alpha}\sigma}(x)} \, \mathrm{d}x \leq \int_{-\infty}^{\infty} f_0(x) \ln \frac{f_0(x)}{g_{F,\sqrt{\alpha}\sigma}(x)} \, \mathrm{d}x,$$

which follows from (6.2), neither Theorem 3 nor Theorem 4 of Ghosal *et al.* (1999, pages 146-147, 148) is helpful to the purpose. Nonetheless, as subsequently shown, at least in the case when the true density is a mixture of normals over a compact set of locations, an additional mild condition on the true value σ_0 of σ allows to conclude that, for each $\eta > 0$, $\pi(G_\eta) > 0$, under the same conditions as those of the aforesaid Theorem 3.

Lemma 6.1. *Let $\alpha \in [1/2, 1)$ be fixed. Let f_0 be of the form $f_0 = f_{F_0,\sigma_0}$. Suppose $F_0 \in \mathcal{S}_W(\Pi)$. Furthermore, assume that F_0 has a compact support. Let $(0, \sigma_2]$ be the support of μ. If $\sigma_0 \in (0, \sigma_2/\sqrt{2}]$, then, for each $\eta > 0$, $\pi(G_\eta) > 0$.*

Proof. See the Appendix. □

In the preceding lemma f_0 was a mixture of normal densities over a compact set of locations. In the following, f_0 itself has a compact support. Let

$$g_{0,\sigma}(x) = \left\{ \int_{-\infty}^{\infty} \phi_\sigma(x-\theta)^\alpha f_0(x)\, dx \right\}^{1/\alpha}, \qquad x \in \mathbb{R}.$$

Lemma 6.2. *Let $\alpha \in [1/2, 1)$ be fixed. Suppose that $P_0 \in \mathcal{S}_W(\Pi)$ and has a compact support. Let $(0, \sigma_2]$ be the support of μ. If $\lim_{\sigma \to 0} \int f_0 \ln(f_0/g_{0,\sigma}) = 0$, then, for each $\eta > 0$, $\pi(G_\eta) > 0$.*

Proof. It follows the same lines as the proof of Theorem 4 in Ghosal *et al.* (1999, page 148). □

Let $H_\varepsilon = \{(F, \sigma) : d_H(f_0, f_{F,\sigma}) < \varepsilon\}$ be the ε-Hellinger open ball and let $\mathcal{F} = \mathcal{P} \times (0, \infty)$.

Theorem 6.3. *Suppose that, for each $\eta > 0$, $\pi(G_\eta) > 0$. Assume a prior density for σ to be of the form $\mu_n(\sigma) \propto \sigma^{(1-\alpha)n}\mu(\sigma)$, with $\alpha \in [1/2, 1)$. Let $(0, \sigma_2]$ be the support of μ. Suppose that $\int_0^{\sigma_2} \sigma^\xi \mu(\sigma)\, d\sigma < \infty$ for all $\xi > 0$. Then, for each $\varepsilon > 0$, $\lim_{n\to\infty} \pi(H_\varepsilon^c|x^n) = 0$ a.s. $[P_0^\infty]$.*

Proof. Choose $\varepsilon > 0$. Taking into account (6.3) and (6.2), the posterior probability of H_ε^c can be bounded above as follows

$$\begin{aligned}
\pi(H_\varepsilon^c|x^n) &= \frac{\int_{H_\varepsilon^c} R(f_{F,\sigma}; x^n)\, \mu_n(\sigma)\, d\sigma\, d\Pi(F)}{\int_{\mathcal{F}} R(f_{F,\sigma}; x^n)\, \mu_n(\sigma)\, d\sigma\, d\Pi(F)} \\
&= \frac{\int_{H_\varepsilon^c} R(g_{F,\sqrt{\alpha}\sigma}; x^n)^\alpha\, \sigma^{-(1-\alpha)n}\, \mu_n(\sigma)\, d\sigma\, d\Pi(F)}{\int_{\mathcal{F}} R(g_{F,\sqrt{\alpha}\sigma}; x^n)^\alpha\, \sigma^{-(1-\alpha)n}\, \mu_n(\sigma)\, d\sigma\, d\Pi(F)} \\
&= \frac{\int_{H_\varepsilon^c} R(g_{F,\sqrt{\alpha}\sigma}; x^n)^\alpha\, \mu(\sigma)\, d\sigma\, d\Pi(F)}{\int_{\mathcal{F}} R(g_{F,\sqrt{\alpha}\sigma}; x^n)^\alpha\, \mu(\sigma)\, d\sigma\, d\Pi(F)} \\
&\le \frac{\int_{H_\varepsilon^c} R(f_{F,\sqrt{\alpha}\sigma}; x^n)^\alpha\, \mu(\sigma)\, d\sigma\, d\Pi(F)}{\int_{\mathcal{F}} R(g_{F,\sqrt{\alpha}\sigma}; x^n)^\alpha\, \mu(\sigma)\, d\sigma\, d\Pi(F)},
\end{aligned} \tag{6.4}$$

where the third line follows from the assumption on μ_n. Now, for any $0 < c < (1-\alpha)\varepsilon^2$, in force of (2.3),

$$\int_{H_\varepsilon^c} R(f_{F,\sqrt{\alpha}\sigma}; x^n)^\alpha\, \mu(\sigma)\, d\sigma\, d\Pi(F) < \exp\{-nc\} \tag{6.5}$$

for all large n, with P_0^∞-probability one. The denominator in (6.4) can be bounded below using standard arguments. Thus, for each $\eta > 0$,

$$(6.6) \qquad \int_{\mathcal{F}} R(g_{F,\sqrt{\alpha}\sigma};\, x^n)^\alpha\, \mu(\sigma)\, d\sigma\, d\Pi(F) > \exp\{-n\eta\}$$

for all large n, on a set of P_0^∞-probability one. Combine bounds in (6.5) and (6.6). Choose η such that $\delta = c - \eta > 0$. Then, $\pi(H_\varepsilon^c|x^n) < \exp\{-n\delta\}$ for all large n, a.s. $[P_0^\infty]$, and the assertion follows. □

Suppose that the mixing measure has a prior distribution depending on a parameter which is itself given a prior, say ρ. The introduction of a hyperparameter yields the additional condition

$$(6.7) \qquad \rho(\{\tau:\, \forall\eta > 0,\, \pi_\tau(G_\eta) > 0\}) > 0.$$

Theorem 6.3 goes through to this model.

Theorem 6.4. *Suppose that F_0 (P_0) has a compact support. Assume that*

$$(6.8) \qquad \rho(\{\tau:\, F_0(P_0) \in \mathcal{S}_W(\Pi_\tau)\}) > 0.$$

Let the prior density for σ be of the form $\mu_n(\sigma) \propto \sigma^{(1-\alpha)n}\mu(\sigma)$, with $\alpha \in [1/2, 1)$. Let $(0, \sigma_2]$ be the support of μ. Suppose that $\int_0^{\sigma_2} \sigma^\xi \mu(\sigma)\, d\sigma < \infty$ for all $\xi > 0$. If $\sigma_0 \in (0, \sigma_2/\sqrt{2}]$ $(\lim_{\sigma\to 0} \int f_0 \ln(f_0/g_{0,\sigma}) = 0)$, then, for each $\varepsilon > 0$, $\lim_{n\to\infty} \pi(H_\varepsilon^c|x^n) = 0$ a.s. $[P_0^\infty]$.

Results concerning the pseudo-posterior distribution have been here used as a mathematical device to establish consistency for the true posterior. This approach will also be taken to study the rate of convergence of the posterior distribution. The use of a sample-size dependent prior for σ produces the same effect as the use of a sieve and a tail condition combined together. Indeed, μ_n acts like a Bayesian sieve, giving sufficiently low weight to small values of σ as the sample size increases. This approach also results in a good rate of convergence for the posterior distribution when the true mixing distribution is compactly supported and is the trajectory of a Dirichlet process. If we choose $\alpha_n = 1 - (\ln n)^2/n$, then the posterior converges at the best known rate $n^{-1/2}(\ln n)$ without imposing any restriction on the tail behavior of the base measure.

Recall that a Dirichlet process on a measurable space $(\Theta, \mathcal{A})$ with base measure α is a random probability measure F on Θ such that, for every finite, measurable partition $(A_1, A_2, \ldots, A_k)$ of Θ, the probability vector $(F(A_1), F(A_2), \ldots, F(A_k))$ has a Dirichlet distribution with parameters $(\alpha(A_1), \alpha(A_2), \ldots, \alpha(A_k))$ on the k-dimensional simplex.

Theorem 6.5. *Let $\varepsilon_n = (\ln n)/\sqrt{n}$ and let $\alpha_n = 1 - \varepsilon_n^2$. Assume that the prior density for σ is of the form $\mu_n(\sigma) \propto \sigma^{(1-\alpha_n)n}\mu(\sigma)$, with μ a continuous and positive density on an interval containing σ_0. Furthermore, assume that $\int_0^{\sigma_2} \sigma^\xi \mu(\sigma)\, d\sigma < \infty$ for all $\xi > 0$. Suppose that F_0 has a compact support $[-k_0, k_0]$, for some $k_0 > 0$. Suppose that the base measure β of the Dirichlet process prior for F has a continuous and positive density on an interval containing $[-k_0, k_0]$. Then, the posterior probability $\pi(H_{\varepsilon_n}^c|x^n)$ tends to zero in P_0^n-probability at the rate $(\ln n)/\sqrt{n}$.*

Proof. Let n be fixed and let ε_n be as in the statement. Recall that $H^c_{\varepsilon_n} = \{(F,\sigma): d_H(f_0, f_{F,\sigma}) \ge K\varepsilon_n\}$, for a sufficiently large constant $K>0$ to be suitably chosen later on. Reasoning as in Theorem 6.3, we get that

$$\begin{aligned}\pi(H^c_{\varepsilon_n}|x^n) &= \frac{\int_{H^c_{\varepsilon_n}} R(f_{F,\sigma};\, x^n)\,\mu_n(\sigma)\,\mathrm{d}\sigma\,\mathrm{d}\Pi(F)}{\int_{\mathcal{F}} R(f_{F,\sigma};\, x^n)\,\mu_n(\sigma)\,\mathrm{d}\sigma\,\mathrm{d}\Pi(F)}\\
&\le \frac{\int_{H^c_{\varepsilon_n}} R(f_{F,\sqrt{\alpha_n}\sigma};\, x^n)^{\alpha_n}\,\mu(\sigma)\,\mathrm{d}\sigma\,\mathrm{d}\Pi(F)}{\int_{\mathcal{F}} R(g_{F,\sqrt{\alpha_n}\sigma};\, x^n)^{\alpha_n}\,\mu(\sigma)\,\mathrm{d}\sigma\,\mathrm{d}\Pi(F)}.\end{aligned} \tag{6.9}$$

Arguing as in Theorem 4.2, we get that, for n large enough so that $\alpha_n > \frac{3}{4}$,

$$\begin{aligned}P_0^n\left(\left\{x^n:\ \int_{H^c_{\varepsilon_n}} R(f_{F,\sqrt{\alpha_n}\sigma};\, x^n)^{\alpha_n}\,\mu(\sigma)\,\mathrm{d}\sigma\,\mathrm{d}\Pi(F) > \exp\left\{-\frac{K^2 n\varepsilon_n^2}{8}\right\}\right\}\right)\\
\le \exp\left\{-\frac{K^2 n\varepsilon_n^2}{8}\right\}.\end{aligned} \tag{6.10}$$

The purpose of the following arguments is to establish a lower bound on the denominator of (6.9). Set the positions

$$K_n(P_0\|P_{F,\sigma}) = \int_{-\infty}^{\infty} f_0(x)\ln\frac{f_0(x)}{g_{F,\sqrt{\alpha_n}\sigma}(x)}\,\mathrm{d}x,$$

$$D_n(P_0,\, P_{F,\sigma}) = \int_{-\infty}^{\infty} f_0(x)\left(\ln\frac{f_0(x)}{g_{F,\sqrt{\alpha_n}\sigma}(x)}\right)^2\mathrm{d}x - \left(\int_{-\infty}^{\infty} f_0(x)\,\ln\frac{f_0(x)}{g_{F,\sqrt{\alpha_n}\sigma}(x)}\,\mathrm{d}x\right)^2,$$

$$G(\varepsilon_n^2) = \{(F,\sigma):\ \max\{K_n(P_0\|P_{F,\sigma}),\, D_n(P_0,\, P_{F,\sigma})\} < \varepsilon_n^2\},$$

$$F(\varepsilon_n^2) = \left\{(F,\sigma):\ \max\left\{K_n(P_0\|P_{F,\sigma}),\, \int_{-\infty}^{\infty} f_0(x)\left(\ln\frac{f_0(x)}{g_{F,\sqrt{\alpha_n}\sigma}(x)}\right)^2\mathrm{d}x\right\} < \varepsilon_n^2\right\},$$

by laying out the same arguments as in Lemma 1 by Shen *et al.* (2001, pages 690-691), it can be seen that, if $\pi(G(\varepsilon_n^2)) > 0$ for all large n, then, for any $\xi \in (0,1)$,

$$P_0^n\left(\left\{x^n:\ \int_{\mathcal{F}} R(g_{F,\sqrt{\alpha_n}\sigma};\, x^n)^{\alpha_n}\,\mu(\sigma)\,\mathrm{d}\sigma\,\mathrm{d}\Pi(F) \le \xi\,\pi(G(\varepsilon_n^2))\,\exp\{-2n\varepsilon_n^2\}\right\}\right) \le \frac{1}{(1-\xi)n\varepsilon_n^2}$$

for sufficiently large n. Hence, since $n\varepsilon_n^2 \to \infty$, as $n\to\infty$,

$$\begin{aligned}\int_{\mathcal{F}} R(g_{F,\sqrt{\alpha_n}\sigma};\, x^n)^{\alpha_n}\,\mu(\sigma)\,\mathrm{d}\sigma\,\mathrm{d}\Pi(F) &> \xi\,\pi(G(\varepsilon_n^2))\,\exp\{-2n\varepsilon_n^2\}\\
&\gtrsim \pi(F(\varepsilon_n^2))\,\exp\{-2n\varepsilon_n^2\},\end{aligned}$$

on a Borel set of P_0^n-probability tending to one. The next developments show that $\pi(F(\varepsilon_n^2)) \gtrsim \exp\{-c_2 n\varepsilon_n^2\}$, for some constant $c_2>0$. We shall show that, for a fixed $\varepsilon>0$ and $\alpha = 1-\varepsilon^2$, $\pi(F(\varepsilon^2)) \gtrsim \exp\{-c_2\left(\ln\frac{1}{\varepsilon}\right)^2\}$. From Theorem 3.1 in Ghosal *et al.* (2000, pages 9-10), it is known that for a constant $d>0$,

$$\left\{(F,\,\sigma):\,\sum_{j=1}^{N}|F([\theta_j-\varepsilon,\,\theta_j+\varepsilon])-p_j|\leq\varepsilon,\ |\sigma-\sigma_0|\leq\varepsilon\right\}$$

$$\subseteq\left\{(F,\,\sigma):\,d_H(f_0,\,f_{F,\sigma})\leq d\varepsilon^{1/2}\left(\ln\frac{1}{\varepsilon}\right)^{1/4}\right\}. \tag{6.11}$$

We show that also the set inclusion

$$\left\{(F,\,\sigma):\,\sum_{j=1}^{N}|F([\theta_j-\varepsilon,\,\theta_j+\varepsilon])-p_j|\leq\varepsilon,\ |\sigma-\sigma_0|\leq\varepsilon\right\}$$

$$\subseteq\left\{(F,\,\sigma):\,d_H(f_0,\,f_{F,\sqrt{\alpha}\sigma})\leq d_3\varepsilon^{1/2}\left(\ln\frac{1}{\varepsilon}\right)^{1/4}\right\}$$

holds true. To this aim, note that $1-\sqrt{\alpha}=1-\sqrt{1-\varepsilon^2}<1-(1-\varepsilon^2)=\varepsilon^2<\varepsilon$. Therefore, if σ satisfies $|\sigma-\sigma_0|\leq\varepsilon$, then, for a suitable constant $c>0$,

$$\left|\sqrt{\alpha}\sigma-\sigma_0\right|\leq\sqrt{\alpha}\,|\sigma-\sigma_0|+(1-\sqrt{\alpha})\,\sigma_0<\varepsilon+\sigma_0\varepsilon\leq c\varepsilon.$$

It can be seen that, for constants $d_1,\,d_2,\,d_3>0$,

$$\begin{aligned}
&\left\{(F,\,\sigma):\,\sum_{j=1}^{N}|F([\theta_j-\varepsilon,\,\theta_j+\varepsilon])-p_j|\leq\varepsilon,\ |\sigma-\sigma_0|\leq\varepsilon\right\}\\
&\qquad\subseteq\left\{(F,\,\sigma):\,\sum_{j=1}^{N}|F([\theta_j-\varepsilon,\,\theta_j+\varepsilon])-p_j|\leq\varepsilon,\ \left|\sqrt{\alpha}\sigma-\sigma_0\right|\leq c\varepsilon\right\}\\
&\qquad\subseteq\left\{(F,\,\sigma):\,\left\|f_{F_0',\sigma_0}-f_{F,\sqrt{\alpha}\sigma}\right\|_1\leq d_1\varepsilon\right\}\\
&\qquad\subseteq\left\{(F,\,\sigma):\,\left\|f_0-f_{F,\sqrt{\alpha}\sigma}\right\|_1\leq d_2\varepsilon\left(\ln\frac{1}{\varepsilon}\right)^{1/2}\right\}\\
&\qquad\subseteq\left\{(F,\,\sigma):\,d_H(f_0,\,f_{F,\sqrt{\alpha}\sigma})^2\leq d_3\varepsilon\left(\ln\frac{1}{\varepsilon}\right)^{1/2}\right\}.
\end{aligned}\tag{6.12}$$

In force of (6.11) and (6.12), by Lemma B.1, it is

$$\left\{(F,\,\sigma):\,\sum_{j=1}^{N}|F([\theta_j-\varepsilon,\,\theta_j+\varepsilon])-p_j|\leq\varepsilon,\ |\sigma-\sigma_0|\leq\varepsilon\right\}\subseteq F\left(\bar{c}\varepsilon\left(\ln\frac{1}{\varepsilon}\right)^{5/2}\right),$$

for a suitable constant $\bar{c}>0$. It is known that, for some constant $c_2>0$,

$$\pi(F(\varepsilon^2))\gtrsim\exp\left\{-c_2\left(\ln\frac{1}{\varepsilon}\right)^2\right\}.$$

Using the sequence ε_n, it is seen that $\pi(F(\varepsilon_n^2)) \gtrsim \exp\{-c_2 n\varepsilon_n^2\}$. The thesis follows. □

The hypothesis on the compactness of the support of F_0 can be relaxed. The above conclusion still holds when F_0 has sub-Gaussian tails and α is a normal density.

7. Discussion

The study of consistency based on the pseudo-posterior distribution can lead to considerable advantages when the parameter space is not totally bounded. If the parameter space is totally bounded there is no need for any trick to get consistency. At this stage of knowledge, this approach requires a case-by-case treatment, a unifying theory is still lacking. Consistency difficulties can be avoided if one is available to adopt the data-dependent prior (1.4). Even though Walker *et al.* (2001, pages 813-814) give a suggestive intuitive explanation about the reason this prior works well, its asymptotic behavior remains on open question.

It would be of interest to see whether an instrumental use of the general properties of Q_n^α can be of help in establishing consistency for other distributions such as the random histograms and an improvement in the rate of convergence can be obtained by suitably choosing α_n in $Q_n^{\alpha_n}$.

Appendix A. Lower Bounds on $m_n^{\alpha_n}/f_0^n$

Conditions on the prior probability of neighborhoods of P_0 implying the existence of in-probability or almost-sure exponential lower bounds on $m_n^{\alpha_n}/f_0^n$ are herein provided.

First, a condition on the prior probability of shrinking Hellinger balls of P_0 is given. The idea is to consider a weaker metric than the Kullback-Leibler divergence. One such condition is used also in Ghosal *et al.* (1997, pages 127-129, Theorem 4.1), another involving the total variation distance on $\mathcal{F}^n$ can be found in Barron (1988, pages 24-25, Lemma 11).

Lemma A.1. *Let $\{\alpha_n\}_{n=1}^\infty$ be a sequence such that $\alpha_n \in (0, 1)$ and $\alpha_n \to 1$, as $n \to \infty$. Suppose there exists a positive sequence $\{t_n\}_{n=1}^\infty$ such that $n^{1/2}t_n \to 0$, as $n \to \infty$. Let $H_{t_n} = \{P \in \mathcal{F} : d_H(P_0, P) < t_n\}$. If $\pi(H_{t_n}) > 0$ for all large n, then, for any $\xi \in (0, 1)$,*

$$\int_{\mathcal{F}} \left(\frac{f_P^n(x^n)}{f_0^n(x^n)}\right)^{\alpha_n} \mathrm{d}\pi(P) > \xi\, \pi(H_{t_n}),$$

on a Borel set of P_0^n-probability tending to one.

Proof. First, observe that, for any $P \in \mathcal{F}$,

$$\|f_0^n - f_P^n\|_1 \le 2\sqrt{2\left[1 - \left(1 - \frac{1}{2}d_H(P_0, P)^2\right)^n\right]} \le 2\sqrt{n}\, d_H(P_0, P), \qquad n \ge 1.$$

For details, make reference to Ghosal *et al.* (1997, page 128). For each n, if $P \in H_{t_n}$, then

$$\text{(A.1)} \qquad d_{TV}(P_0^n, P^n) = \frac{1}{2}\|f_0^n - f_P^n\|_1 \le \sqrt{n}\, d_H(P_0, P) < \sqrt{n}\, t_n.$$

Let n be large enough so that $\pi(H_{t_n}) > 0$. By applying Markov's inequality, Hölder's inequality and Tonelli's theorem, the following chain of inequalities can be derived:

$$P_0^n\left(\left\{x^n : \int_{\mathscr{F}}\left(\frac{f_P^n(x^n)}{f_0^n(x^n)}\right)^{\alpha_n} \mathrm{d}\pi(P) \le \xi\,\pi(H_{t_n})\right\}\right) \le$$

$$\begin{aligned}
&\le P_0^n\left(\left\{x^n : \int_{H_{t_n}}\left(\frac{f_P^n(x^n)}{f_0^n(x^n)}\right)^{\alpha_n} \mathrm{d}\pi(P) \le \xi\,\pi(H_{t_n})\right\}\right)\\
&\le P_0^n\left(\left\{x^n : \left|\int_{H_{t_n}}\left[\left(\frac{f_P^n(x^n)}{f_0^n(x^n)}\right)^{\alpha_n} - 1\right] \mathrm{d}\pi(P)\right| \ge (1-\xi)\,\pi(H_{t_n})\right\}\right)\\
&\le P_0^n\left(\left\{x^n : \int_{H_{t_n}}\left|\left(\frac{f_P^n(x^n)}{f_0^n(x^n)}\right)^{\alpha_n} - 1\right| \mathrm{d}\pi(P) \ge (1-\xi)\,\pi(H_{t_n})\right\}\right)\\
&\le \frac{1}{(1-\xi)\,\pi(H_{t_n})}\,\mathrm{E}_0^n\left[\int_{H_{t_n}}\left|\left(\frac{f_P^n(x^n)}{f_0^n(x^n)}\right)^{\alpha_n} - 1\right| \mathrm{d}\pi(P)\right]\\
&\le \frac{1}{(1-\xi)}\left\{\mathrm{E}_0^n\left[\left(\int_{H_{t_n}}\left|\left(\frac{f_P^n(x^n)}{f_0^n(x^n)}\right)^{\alpha_n} - 1\right| \frac{\mathrm{d}\pi(P)}{\pi(H_{t_n})}\right)^{1/\alpha_n}\right]\right\}^{\alpha_n}\\
&\le \frac{1}{(1-\xi)}\left\{\mathrm{E}_0^n\left[\int_{H_{t_n}}\left|\left(\frac{f_P^n(x^n)}{f_0^n(x^n)}\right)^{\alpha_n} - 1\right|^{1/\alpha_n} \frac{\mathrm{d}\pi(P)}{\pi(H_{t_n})}\right]\right\}^{\alpha_n}\\
&= \frac{1}{(1-\xi)}\left[\int_{H_{t_n}}\int_{\mathcal{X}^n}|f_0^n(x^n)^{\alpha_n} - f_P^n(x^n)^{\alpha_n}|^{1/\alpha_n}\,\mathrm{d}\lambda^n(x^n)\,\frac{\mathrm{d}\pi(P)}{\pi(H_{t_n})}\right]^{\alpha_n}\\
&= \frac{1}{(1-\xi)}\left[\int_{H_{t_n}} M_{\alpha_n}(P_0^n\|P^n)\,\frac{\mathrm{d}\pi(P)}{\pi(H_{t_n})}\right]^{\alpha_n}\\
&\le \frac{1}{(1-\xi)}\left[2\int_{H_{t_n}} d_{TV}(P_0^n, P^n)\,\frac{\mathrm{d}\pi(P)}{\pi(H_{t_n})}\right]^{\alpha_n}\\
&\le \frac{1}{(1-\xi)}\,(2\sqrt{n}t_n)^{\alpha_n}.
\end{aligned}$$

The second line from the bottom descends from the fact that the Matusita distances which, for each $\alpha \in (0, 1]$, are defined as $M_\alpha(P\|Q) = \int_{\mathcal{X}} |p^\alpha - q^\alpha|^{1/\alpha}\,\mathrm{d}\lambda$, satisfy the relationship $M_\alpha(P\|Q) \le 2d_{TV}(P, Q)$, see Liese *et al.* (1987, page 48, Proposition 2.37). The last line follows from inequality (A.1). Finally, by taking into account that $n^{1/2}t_n \to 0$ and $\alpha_n \to 1$, as $n \to \infty$, the assertion follows. □

Almost-sure lower bounds on $m_n^{\alpha_n}/f_0^n$ are herein derived. Some preliminary definitions and results are needed. First of all, a "family of indexes of discrepancy" of probability measures called ρ_α-divergences, whose definition is substantially due to Wong *et al.* (1995, pages 351-352), is presented. For $y > 0$, let

$$g_\alpha(y) = \begin{cases} \dfrac{y^\alpha - 1}{\alpha}, & \text{if } (-1 < \alpha < 0) \text{ or } (0 < \alpha \leq 1), \\ \ln y, & \text{if } \alpha = 0^+. \end{cases}$$

For probability measures P and Q, set $y = f_P/f_Q$, the ρ_α-divergence is defined as $\rho_\alpha(P\|Q) = \mathrm{E}_P[g_\alpha(Y)]$. For specific values of α, notable divergences of statistical interest are obtained. For $\alpha = 0^+$, ρ_{0^+} is the Kullback-Leibler divergence; for $\alpha = 1$, ρ_1 is the Pearson χ^2-divergence. Wong *et al.* (1995, page 351) note that $\rho_{0^+} \leq \rho_\alpha$ for all $0 < \alpha \leq 1$. A more refined order relationship holds true.

Lemma A.2. *For any pair P and Q in $\mathcal{P}$, it holds*

$$\text{(A.2)} \qquad \rho_{0^+}(P\|Q) \leq \rho_\alpha(P\|Q) \leq \rho_\beta(P\|Q) \leq \rho_1(P\|Q), \qquad 0 < \alpha \leq \beta \leq 1.$$

Proof. If $P \not\ll Q$, then (A.2) is trivial. For P and Q such that $P \ll Q$, the claim is proved by relating the ρ_α-divergences to the I_α-divergences. For any $\alpha \in (0, 1]$,

$$1 + \alpha\rho_\alpha(P\|Q) = \int_{\{f_P f_Q > 0\}} f_P^{1+\alpha} f_Q^{-\alpha}\, \mathrm{d}\lambda = 1 + \alpha(1+\alpha)I_{1+\alpha}(P\|Q),$$

whence $\rho_\alpha(P\|Q) = (1+\alpha)I_{1+\alpha}(P\|Q)$. The conclusion follows from (2.1) by setting the positions $\gamma = 1 + \alpha$ and $\delta = 1 + \beta$. □

In the next lemma, $m_n^{\alpha_n}/f_0^n$ is bounded below by assigning not exponentially small prior probability to shrinking χ^2-balls.

Lemma A.3. *Let $\{\alpha_n\}_{n=1}^\infty$ and $\{t_n\}_{n=1}^\infty$ be positive sequences such that $\alpha_n \in (0, 1)$, $t_n \to 0$ and $nt_n \to \infty$, as $n \to \infty$. Let $S_n = \{P \in \mathcal{F} : \chi^2(P_0\|P) < t_n\}$. Suppose that $\pi(S_n) > 0$ for all large n. Then, for any $\xi \in (0, 1)$,*

$$\int_{\mathcal{F}} \left(\frac{f_P^n(x^n)}{f_0^n(x^n)} \right)^{\alpha_n} \mathrm{d}\pi(P) > \xi\, \pi(S_n)\, e^{-2nt_n},$$

for all sufficiently large n, along P_0^∞-almost all sample paths.

Proof. For each n, define the sets $W_n = \{(P, x^n) : \alpha_n K(f_P; x^n) \geq 2t_n\} \subset \mathcal{F} \times \mathcal{X}^n$, $V_{x^n} = \{P : (P, x^n) \in W_n\} \subset \mathcal{F}$ and $V_P = \{x^n : (P, x^n) \in W_n\} \subset \mathcal{X}^n$. For any $x^n \in \mathcal{X}^n$,

$$\text{(A.3)} \qquad L_n(x^n) \geq \int_{S_n \cap V_{x^n}^c} \exp\{-n\alpha_n K(f_P; x^n)\}\, \mathrm{d}\pi(P) > \pi(S_n \cap V_{x^n}^c)\, e^{-2nt_n}.$$

Let $P \in S_n$. Note that $P_0 \ll P$. From the chain of inequalities

$$\begin{aligned}
P_0^n(\{x^n : \alpha_n K(f_P; x^n) \geq 2t_n\}) &= P_0^n\left(\left\{x^n : \left(\frac{f_0^n(x^n)}{f_P^n(x^n)}\right)^{\alpha_n} \geq \exp\{2nt_n\}\right\}\right) \\
&\leq \exp\{-2nt_n\}\left\{\mathrm{E}_0\left[\left(\frac{f_0(X_1)}{f_P(X_1)}\right)^{\alpha_n}\right]\right\}^n \\
&= \exp\{-2nt_n\}\left[1 + \alpha_n \rho_{\alpha_n}(P_0\|P)\right]^n \\
&\leq \exp\{-2nt_n\}\exp\{n\alpha_n \rho_{\alpha_n}(P_0\|P)\} \\
&< \exp\{-2nt_n\}\exp\{n\chi^2(P_0\|P)\} \\
&< \exp\{-nt_n\},
\end{aligned}$$

where the second line from the bottom follows from (A.2) and the fact that $\alpha_n < 1$, it descends that $P_0^n(V_P) < e^{-nt_n}$. By means of Tonelli's theorem, this entails that

$$\mathrm{E}_0^n[\pi(S_n \cap V_{x^n})] \leq \pi(S_n)\, e^{-nt_n}. \tag{A.4}$$

Let n be large enough so that $\pi(S_n) > 0$. In account of (A.3), by Markov's inequality, using (A.4), it is seen that, for any $\xi \in (0, 1)$,

$$P_0^n\left(\{x^n : L_n(x^n) \leq \xi\, \pi(S_n)\, e^{-2nt_n}\}\right) \leq \frac{e^{-nt_n}}{(1-\xi)}.$$

Since $\sum_{n=1}^{\infty} e^{-nt_n} < \infty$, the result follows as a consequence of the first Borel-Cantelli lemma. □

For probability measures $P, Q \in \mathcal{F}$, let $\|f_P/f_Q\|_\infty = \sup_{x\in\mathcal{X}} |f_P(x)/f_Q(x)|$ denote the sup-norm. Ghosal, Ghosh *et al.* (2000, page 505, Theorem 2.2) use the Hellinger type neighborhood $B_n = \{P \in \mathcal{F} : \|f_0/f_P\|_\infty\, d_H(P_0, P)^2 < t_n\}$ to get the same result as above. However, B_n is even stronger than a χ^2-neighborhood.

Lemma A.4. *Let $\{\alpha_n\}_{n=1}^{\infty}$ and $\{t_n\}_{n=1}^{\infty}$ be sequences as in Lemma A.3. Suppose that $\pi(B_n) > 0$ for all large n. Then, for any constant $C > 4$ and $\xi \in (0, 1)$,*

$$\int_{\mathcal{F}} \left(\frac{f_P^n(x^n)}{f_0^n(x^n)}\right)^{\alpha_n} \mathrm{d}\pi(P) > \xi\, \pi(B_n)\, e^{-Cnt_n}$$

for all large n, with P_0^∞-probability one.

Proof. From Lemma 7 in Genovese *et al.* (2000, page 1118), it is known that, if $(f_P/f_Q) \leq V$, for some constant $V > 0$, then $\chi^2(P\|Q) \leq (1 + V^{1/2})^2\, d_H(P, Q)^2$. Consequently, if $V^{1/2} \geq 1$,

$$\chi^2(P\|Q) \leq (1 + V^{1/2})^2\, d_H(P, Q)^2 \leq (2V^{1/2})^2\, d_H(P, Q)^2 = 4V\, d_H(P, Q)^2.$$

Now, set the position $V = \|f_P/f_Q\|_\infty$. Since $\|f_P/f_Q\|_\infty \geq 1$, so is its square root and it can be stated that

$$\chi^2(P\|Q) \leq 4\left\|\frac{f_P}{f_Q}\right\|_\infty d_H(P, Q)^2. \tag{A.5}$$

In force of (A.5), the set inclusion

$$\{P : \|f_0/f_P\|_\infty\, d_H(P_0,\, P)^2 < t_n\} \subseteq \{P : \chi^2(P_0\|P) < 4t_n\}$$

holds true for every n. Denoted by S_n the set on the right-hand side of the above display, $B_n \subseteq S_n$, hence, $\pi(S_n) > 0$ for all large n. The result follows by suitably modifying the arguments of Lemma A.3. □

An alternative, but longer derivation of the previous bound can be developed along the same lines as the original proof presented by Ghosal, Ghosh *et al.* (*ibidem*, pages 526-527, Lemma 8.4).

Appendix B. Results on mixture models

Proof of Lemma 6.1 Fix $\eta > 0$. Let $[-k_0,\, k_0]$, with $k_0 > 0$, be the support of F_0. For any $(F,\, \sigma) \in \mathcal{P} \times (0,\, \infty)$, write

$$\int_{-\infty}^{\infty} f_0 \ln \frac{f_0}{g_{F,\sqrt{\alpha}\sigma}} = \int_{-\infty}^{\infty} f_0 \ln \frac{f_0}{g_{F_0,\sqrt{\alpha}\sigma}} + \int_{-\infty}^{\infty} f_0 \ln \frac{g_{F_0,\sqrt{\alpha}\sigma}}{g_{F,\sqrt{\alpha}\sigma}} = I_1 + I_2.$$

Consider I_1. Since

$$\frac{f_0(x)}{g_{F_0,\sqrt{\alpha}\sigma}(x)} = \frac{\int_{-k_0}^{k_0} \phi_{\sigma_0}(x-\theta)\, \mathrm{d}F_0(\theta)}{\left(\int_{-k_0}^{k_0} \phi_{\sqrt{\alpha}\sigma}(x-\theta)^\alpha\, \mathrm{d}F_0(\theta)\right)^{1/\alpha}} \leq \sup_{|\theta|\leq k_0} \frac{\phi_{\sigma_0}(x-\theta)}{\phi_{\sqrt{\alpha}\sigma}(x-\theta)}, \qquad x \in \mathbb{R},$$

it is

$$\lim_{\sqrt{\alpha}\sigma\to\sigma_0} \int_{-\infty}^{\infty} f_0(x) \ln \frac{f_0(x)}{g_{F_0,\sqrt{\alpha}\sigma}(x)}\, \mathrm{d}x = 0.$$

A neighborhood N_η of $\sigma_0/\sqrt{\alpha}$ can be chosen such that, if $\sigma \in N_\eta \equiv N_\eta(\sigma_0/\sqrt{\alpha})$, then $I_1 < \eta/2$. Note that $\sigma_0/\sqrt{\alpha} \in (\sigma_0,\, \sqrt{2}\sigma_0]$ and, by hypothesis, $\sigma_0 \leq \sigma_2/\sqrt{2}$, so that $\sigma_0/\sqrt{\alpha}$ is in the support of μ and $\mu(N_\eta) > 0$.

Now, consider I_2. Keeping in mind (6.3), by straightforward manipulations, it turns out to be

$$I_2 = \int_{-\infty}^{\infty} f_0(x) \ln \frac{g_{F_0,\sqrt{\alpha}\sigma}(x)}{g_{F,\sqrt{\alpha}\sigma}(x)}\, \mathrm{d}x = \int_{-\infty}^{\infty} f_0(x) \ln \frac{\left(\int_{-k_0}^{k_0} \phi_{\sqrt{\alpha}\sigma}(x-\theta)^\alpha\, \mathrm{d}F_0(\theta)\right)^{1/\alpha}}{\left(\int_{-\infty}^{\infty} \phi_{\sqrt{\alpha}\sigma}(x-\theta)^\alpha\, \mathrm{d}F(\theta)\right)^{1/\alpha}}\, \mathrm{d}x$$

$$= \frac{1}{\alpha}\int_{-\infty}^{\infty} f_0(x) \ln \frac{\int_{-k_0}^{k_0} \phi_\sigma(x-\theta)\, \mathrm{d}F_0(\theta)}{\int_{-\infty}^{\infty} \phi_\sigma(x-\theta)\, \mathrm{d}F(\theta)}\, \mathrm{d}x$$

$$= \frac{1}{\alpha}\int_{-\infty}^{\infty} f_0(x) \ln \frac{f_{F_0,\sigma}(x)}{f_{F,\sigma}(x)}\, \mathrm{d}x. \tag{B.1}$$

An upper bound on the integral in (B.1) is known as a by-product of Theorem 3 (*ibidem*). In fact, it can be seen that, for each $\sigma > 0$, $\zeta > 0$, $0 < \vartheta < 1/3$ (ζ and ϑ play the role of η and δ of the original proof) and F in a suitable weak neighborhood E of F_0, it is

$$\int_{-\infty}^{\infty} f_0(x)\ln\frac{f_{F_0,\sigma}(x)}{f_{F,\sigma}(x)}\,\mathrm{d}x < 2\left(\frac{2k_0}{\sigma^2}+\ln 2\right)\zeta+\frac{3\vartheta}{1-3\vartheta}.$$

Now, ζ and ϑ can be chosen so that, for each $\sigma\in N_\eta$,

$$\frac{1}{\alpha}\left[2\left(\frac{2k_0}{\sigma^2}+\ln 2\right)\zeta+\frac{3\vartheta}{1-3\vartheta}\right]<\eta/2.$$

It follows that $I_1+I_2<\eta$ for all $(F,\,\sigma)$ in $E_\eta\times N_\eta$ (E_η depends on η through ζ and ϑ). It results $\pi(E_\eta\times N_\eta)=\Pi(E_\eta)\mu(N_\eta)>0$, because $\Pi(E_\eta)>0$, E_η being a weak neighborhood of F_0, and $\mu(N_\eta)>0$. From $\pi(G_\eta)\geq\pi(E_\eta\times N_\eta)>0$, the thesis follows.

Lemma B.1. *Suppose that F_0 is a probability measure supported on a compact set $[-k_0,\,k_0]$, with $k_0>0$. Let F be any probability measure such that $F([-B_0,\,B_0])>\frac{1}{2}$, for some constant $B_0>k_0$. Let $\sigma_0\in(0,\,\sigma_2]$ be given and assume ε to be sufficiently small so that $\varepsilon\leq\min\{\frac{\sigma_0}{2},\,(1-\frac{1}{e})/\sqrt{2}\}$. Let $\alpha=1-\varepsilon^2$. Let f_{F_0,σ_0}, $f_{F,\sigma}$, $f_{F,\sqrt{\alpha}\sigma}$ be densities with $\sigma\in(0,\,\sigma_2]$ and $|\sigma-\sigma_0|\leq\varepsilon$, such that*

$$\max\{d_H(f_{F_0,\sigma_0},\,f_{F,\sigma}),\,d_H(f_{F_0,\sigma_0},\,f_{F,\sqrt{\alpha}\sigma})\}\leq\varepsilon.$$

Then,

$$\int_{-\infty}^{\infty} f_{F_0,\sigma_0}\ln\frac{f_{F_0,\sigma_0}}{g_{F,\sqrt{\alpha}\sigma}}\lesssim\varepsilon^2\left(\ln\frac{1}{\varepsilon}\right), \tag{B.2}$$

$$\int_{-\infty}^{\infty} f_{F_0,\sigma_0}\left(\ln\frac{f_{F_0,\sigma_0}}{g_{F,\sqrt{\alpha}\sigma}}\right)^2\lesssim\left(\varepsilon\ln\frac{1}{\varepsilon}\right)^2. \tag{B.3}$$

Proof. Write

$$\begin{aligned}\int_{-\infty}^{\infty} f_{F_0,\sigma_0}\ln\frac{f_{F_0,\sigma_0}}{g_{F,\sqrt{\alpha}\sigma}} &= \int_{-\infty}^{\infty} f_{F_0,\sigma_0}\ln\frac{f_{F_0,\sigma_0}}{f_{F,\sigma}}+\int_{-\infty}^{\infty} f_{F_0,\sigma_0}\ln\frac{f_{F,\sigma}}{g_{F,\sqrt{\alpha}\sigma}}\\ &= K_1+K_2\\ &\lesssim\varepsilon^2\left(\ln\frac{1}{\varepsilon}\right)+K_2, \qquad \text{(B.4)}\end{aligned}$$

where the last line follows from Lemma 3.5 in Ghosal *et al.* (2000, page 8). We show that K_2 can be bounded above by the same quantity up to a constant. Note that

$$\begin{aligned}g_{F,\sqrt{\alpha}\sigma}(x) &= \left\{\int_{-\infty}^{\infty}\left(\frac{1}{\sqrt{2\alpha\pi}\,\sigma}\exp\left\{-\frac{(x-\theta)^2}{2\alpha\sigma^2}\right\}\right)^{\alpha}\mathrm{d}F(\theta)\right\}^{1/\alpha}\\ &> \left\{\int_{-B_0}^{B_0}\left(\frac{\sqrt{2}}{3\sqrt{\pi}\,\sigma_0}\exp\left\{-\frac{4(|x|+B_0)^2}{\sigma_0^2}\right\}\right)^{\alpha}\mathrm{d}F(\theta)\right\}^{1/\alpha}\\ &> \frac{1}{3\sqrt{2\pi}\,\sigma_0}\exp\left\{-\frac{4(|x|+B_0)^2}{\sigma_0^2}\right\},\qquad x\in\mathbb{R}. \qquad \text{(B.5)}\end{aligned}$$

Recalling (6.3) and using (B.5),

$$
\begin{aligned}
K_2 &= \int_{-\infty}^{\infty} f_{F_0,\sigma_0}(x)\, \ln \frac{f_{F,\sigma}(x)}{g_{F,\sqrt{\alpha}\sigma}(x)}\, dx \\
&= \int_{-\infty}^{\infty} f_{F_0,\sigma_0}(x) \left(\ln \frac{[\alpha^{\alpha}(2\pi)^{-(1-\alpha)}]^{1/2}\, \sigma^{-(1-\alpha)}\, g_{F,\sqrt{\alpha}\sigma}(x)^{\alpha}}{g_{F,\sqrt{\alpha}\sigma}(x)} \right) dx \\
&< \int_{-\infty}^{\infty} f_{F_0,\sigma_0}(x)\, \ln \left(\sqrt{2\pi}\, \sigma\, g_{F,\sqrt{\alpha}\sigma}(x) \right)^{-(1-\alpha)} dx \\
&< (1-\alpha) \int_{-\infty}^{\infty} f_{F_0,\sigma_0}(x)\, \ln \frac{\sqrt{2}}{\sqrt{\pi}\, \sigma_0\, g_{F,\sqrt{\alpha}\sigma}(x)}\, dx \\
&< (1-\alpha) \int_{-\infty}^{\infty} f_{F_0,\sigma_0}(x)\, \ln \left(6 \exp \left\{ \frac{4(|x|+B_0)^2}{\sigma_0^2} \right\} \right) dx \\
&\leq \frac{(1-\alpha)}{\delta_0} \int_{-\infty}^{\infty} f_{F_0,\sigma_0}(x) \left(6 \exp \left\{ \frac{4(|x|+B_0)^2}{\sigma_0^2} \right\} \right)^{\delta_0} dx \\
&= \frac{6^{\delta_0}(1-\alpha)}{\delta_0} \int_{-\infty}^{\infty} f_{F_0,\sigma_0}(x)\, \exp \left\{ \frac{4\delta_0(|x|+B_0)^2}{\sigma_0^2} \right\} dx \\
&\lesssim 1-\alpha \\
&\lesssim \varepsilon^2 \left(\ln \frac{1}{\varepsilon} \right), \qquad \text{(B.6)}
\end{aligned}
$$

where $\delta_0 \in (0, 1)$ is a number arbitrarily chosen so as to assure that the integral is finite, therefore, it depends on B_0 and σ_0 only. Combining (B.4) with (B.6), (B.2) is obtained.

Now, write

$$
\begin{aligned}
\int_{-\infty}^{\infty} f_{F_0,\sigma_0} \left(\ln \frac{f_{F_0,\sigma_0}}{g_{F,\sqrt{\alpha}\sigma}} \right)^2 &= \int_{-\infty}^{\infty} f_{F_0,\sigma_0} \left(\ln \frac{f_{F_0,\sigma_0}}{f_{F,\sqrt{\alpha}\sigma}} \right)^2 + \int_{-\infty}^{\infty} f_{F_0,\sigma_0} \left(\ln \frac{f_{F,\sqrt{\alpha}\sigma}}{g_{F,\sqrt{\alpha}\sigma}} \right)^2 \\
&\quad + 2 \int_{-\infty}^{\infty} f_{F_0,\sigma_0} \left(\ln \frac{f_{F_0,\sigma_0}}{f_{F,\sqrt{\alpha}\sigma}} \right) \left(\ln \frac{f_{F,\sqrt{\alpha}\sigma}}{g_{F,\sqrt{\alpha}\sigma}} \right) \\
&\leq Q_1 + Q_2 + 2\,(Q_1 Q_2)^{1/2}, \qquad \text{(B.7)}
\end{aligned}
$$

where the last line follows from Cauchy-Schwarz's inequality. Consider Q_1. By hypothesis, it is $d_H(f_{F_0,\sigma_0}, f_{F,\sqrt{\alpha}\sigma})^2 \leq \varepsilon^2$ and $|\sigma - \sigma_0| \leq \varepsilon$, hence, reasoning as in Lemma 3.5 in Ghosal *et al.* (2000, page 8),

$$
Q_1 \lesssim \left(\varepsilon \ln \frac{1}{\varepsilon} \right)^2. \qquad \text{(B.8)}
$$

Consider Q_2. First, note that in view of (6.2), $(f_{F,\sqrt{\alpha}\sigma}/g_{F,\sqrt{\alpha}\sigma}) \geq 1$, hence $\ln(f_{F,\sqrt{\alpha}\sigma}/g_{F,\sqrt{\alpha}\sigma}) \geq 0$. Taking into account that $f_{F,\sqrt{\alpha}\sigma} \leq f_{F,\sigma}/\sqrt{\alpha}$,

$$
\begin{aligned}
Q_2 &= \int_{-\infty}^{\infty} f_{F_0,\sigma_0}(x)\left(\ln\frac{f_{F,\sqrt{\alpha}\sigma}(x)}{g_{F,\sqrt{\alpha}\sigma}(x)}\right)^2 \mathrm{d}x \\
&< \int_{-\infty}^{\infty} f_{F_0,\sigma_0}(x)\left(\ln\frac{f_{F,\sigma}(x)}{\sqrt{\alpha}\, g_{F,\sqrt{\alpha}\sigma}(x)}\right)^2 \mathrm{d}x \\
&= \int_{-\infty}^{\infty} f_{F_0,\sigma_0}(x)\left((1-\alpha)\ln\frac{1}{\sqrt{2\alpha\pi}\,\sigma\, g_{F,\sqrt{\alpha}\sigma}(x)}\right)^2 \mathrm{d}x \\
&< \int_{-\infty}^{\infty} f_{F_0,\sigma_0}(x)\left((1-\alpha)\ln\frac{2}{\sqrt{\pi}\,\sigma_0\, g_{F,\sqrt{\alpha}\sigma}(x)}\right)^2 \mathrm{d}x \\
&< \int_{-\infty}^{\infty} f_{F_0,\sigma_0}(x)\left[(1-\alpha)\ln\left(6\sqrt{2}\exp\left\{4\frac{(|x|+B_0)^2}{\sigma_0^2}\right\}\right)\right]^2 \mathrm{d}x \\
&= 4\frac{(1-\alpha)^2}{\delta_0^2}\int_{-\infty}^{\infty} f_{F_0,\sigma_0}(x)\left[\ln\left(6\sqrt{2}\exp\left\{4\frac{(|x|+B_0)^2}{\sigma_0^2}\right\}\right)^{\delta_0/2}\right]^2 \mathrm{d}x \\
&< 4\frac{(1-\alpha)^2(6\sqrt{2})^{\delta_0}}{\delta_0^2}\int_{-\infty}^{\infty} f_{F_0,\sigma_0}(x)\,\exp\left\{4\delta_0\frac{(|x|+B_0)^2}{\sigma_0^2}\right\}\mathrm{d}x \\
&\lesssim (1-\alpha)^2 \\
&\lesssim \left(\varepsilon\ln\frac{1}{\varepsilon}\right)^2. \qquad \text{(B.9)}
\end{aligned}
$$

Inequality (B.3) follows by combining (B.7) with (B.8) and (B.9):

$$
\begin{aligned}
\int_{-\infty}^{\infty} f_{F_0,\sigma_0}\left(\ln\frac{f_{F_0,\sigma_0}}{g_{F,\sqrt{\alpha}\sigma}}\right)^2 &\le Q_1+Q_2+2\,(Q_1Q_2)^{1/2} \\
&\lesssim 2\left(\varepsilon\ln\frac{1}{\varepsilon}\right)^2+2\left(\varepsilon\ln\frac{1}{\varepsilon}\right)\left(\varepsilon\ln\frac{1}{\varepsilon}\right) \\
&\lesssim \left(\varepsilon\ln\frac{1}{\varepsilon}\right)^2.
\end{aligned}
$$

□

References

[1] Barron, A. R. (1988). The exponential convergence of posterior probabilities with implications for Bayes estimators of density functions. *Technical Report n. 7, Dept. Statist., Univ. of Illinois, Champaign, IL.*

[2] Barron, A., Schervish, M. J. and Wasserman, L. (1999). The consistency of posterior distributions in nonparametric problems. *Ann. Statist.*, **27**, 536–561.

[3] Billingsley, P. (1995). *Probability and Measure*, 3rd ed., Wiley, New York.

[4] Checcucci, V., Tognoli, A. and Vesentini, E. (1968). *Lezioni di topologia generale.* Feltrinelli Editore Milano.

[5] Ferguson, T. S. (1983). Bayesian density estimation by mixtures of Normal distributions. In *Recent Advances in Statistics* (M. Rizvi, J. Rustagi and D. Siegmund, Eds.), 287–302. Academic Press, New York.

[6] Ghorai, J. K. and Rubin, H. (1982). Bayes risk consistency of nonparametric Bayes density estimates. *Austral. J. Statist.*, **24**, 51–66.

[7] Ghosal, S., Ghosh, J. K. and Ramamoorthi, R. V. (1997). Non-informative priors via sieves and packing numbers. In *Advances in Statistical Decision Theory and Applications*, (S. Panchapakesan and N. Balakrishnan, Eds.), 119–132. Birkhauser, Boston.

[8] Ghosal, S., Ghosh, J. K. and Ramamoorthi, R. V. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *Ann. Statist.*, **27**, 143–158.

[9] Ghosal, S., Ghosh, J. K. and van der Vaart, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.*, **28**, 500–531.

[10] Ghosal, S. and van der Vaart, A. W. (2000). Rates of convergence for Bayes and maximum likelihood estimation for mixtures of normal densities. Preprint.

[11] Liese, F. and Vajda, I. (1987). *Convex Statistical Distances.* Teubner, Leipzig.

[12] Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *Ann. Statist.*, **12**, 351–357.

[13] Roeder, K. and Wasserman, L. (1997). Practical Bayesian density estimation using mixtures of normals. *J. Amer. Statist. Assoc.*, **92**, 894–902.

[14] Shen, X. and Wasserman, L. (2001). Rates of convergence of posterior distributions. *Ann. Statist.*, **29**, 687–714.

[15] Vajda, I. (1989). *Theory of Statistical Inference and Information.* Kluwer Academic Publishers, Dordrecht.

[16] Walker, S. and Hjort, N. L. (2001). On Bayesian consistency. *J. R. Statist. Soc.* B, **63**, 811–821.

[17] West, M. (1992). Modeling with mixtures. In *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, Eds.), 503–524. Oxford University Press.

[18] West, M., Muller, P. and Escobar, M. D. (1994). Hierarchical priors and mixture models, with applications in regression and density estimation. In *Aspects of Uncertainty: A Tribute to D. V. Lindley*, (P. R. Freeman and A. F. M. Smith, Eds.), 363–386. Wiley, Chichester.

C. Scricciolo,
Department of Statistical Sciences,
University of Padova,
Via C. Battisti, 241/243,
35121I - Padova,
Italy