Convergence rates of posterior distributions for dirichlet mixtures of normal densities

C. Scricciolo

2001.21

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

Dicembre 2001

CONVERGENCE RATES OF POSTERIOR DISTRIBUTIONS FOR DIRICHLET MIXTURES OF NORMAL DENSITIES

CATIA SCRICCIOLO

ABSTRACT. In this paper convergence rates of posterior distributions of Dirichlet mixtures of normal densities are determined when the true density is a mixture of normals over a compact set of locations. The scale parameter is allowed to take on values in an interval $(0, \sigma_2]$, with σ_2 known. For location mixtures the rate depends on the tail behavior of the base measure as well as on the tail behavior of the prior for the scale parameter. For location-scale mixtures, if the location and scale parameters are independently distributed according to Dirichlet processes, then the rate depends on the tail behavior of the base measures. If the scale and location parameters are components of the overall mixing parameter, then the rate is governed by the tail behavior of the base measure of the unique Dirichlet process prior.

1. INTRODUCTION

In a mixture model, a random density is generated by convoluting a known kernel with a distribution which is given a prior. Let $F(\cdot)$ stand for the cumulative distribution function (cdf) of a probability measure P on $\mathbb{R}$. In the sequel, F will be used also to indicate the corresponding probability measure P. Let $\phi(\cdot)$ denote the standard normal density and $\phi_\sigma(\cdot)$ the density of a normal distribution with zero mean and standard deviation σ. The density is a location mixture of normals if

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

The scale parameter σ can be either fixed or random. Given P_0, $X_1, \ldots, X_n$ are independent and identically distributed (i.i.d.) observations from $f_0 = \mathrm{d}P_0/\mathrm{d}\lambda$, where λ is the Lebesgue measure. It is assumed that the true density $f_0 = f_{F_0,\sigma_0}$, for some $F_0 \in \mathcal{M}(\mathbb{R})$ and $\sigma_0 \in (0, \infty)$.

Suppose that σ is distributed according to μ and that F is chosen independently of σ according to Π, then the measure $\pi = \Pi \times \mu$ on $\mathcal{M}(\mathbb{R}) \times (0, \infty)$ induces a prior on the set of all probability measures on $\mathbb{R}$ absolutely continuous w.r.t. λ 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 $\bigcup_{0<\sigma\leq\sigma_2} \mathcal{F}_\sigma$, where $\mathcal{F}_\sigma = \{\phi_\sigma * F,\ F \in \mathcal{M}(\mathbb{R})\}$. The prior Π is typically chosen to be a Dirichlet process with base measure α. This model was introduced by Lo (1984), who studied the problem of density estimation from a Bayesian point of view. He gave the expression of the Bayes estimator of the density function, see also Ghorai *et al.* (1982).

The problem of determining rates of convergence for posterior distributions of Dirichlet mixtures of normal densities has been recently addressed by Ghosal *et* *al.* (2000) under the restrictive assumption that the scale parameter takes on values in an interval with a strictly positive lower bound. This assumption, however, is not realistic: this bound is not known in general whilst it is needed to elicit the prior distribution for σ. The same problem is herein considered removing the aforesaid restriction. As it is intuitively expected, slower rates of convergence are obtained. We emphasize that the optimal rate of convergence relative to the Hellinger distance is still not known.

The strategy consists in verifying that the conditions of a general theorem on rates of convergence of posterior distributions are fullfilled. In order to quote the theorem, some definitions are recalled. The ε-packing number of a set of densities $\mathcal{F}$ with respect to a distance d, denoted by $D(\varepsilon, \mathcal{F}, d)$, is the maximum number of points in $\mathcal{F}$ such that the distance between each pair is at least ε. The ε-covering number $N(\varepsilon, \mathcal{F}, d)$ is the minimum number of ε-balls needed to cover $\mathcal{F}$. The ε-bracketing number $N_{[]}(\varepsilon, \mathcal{F}, d)$ is the minimum number of brackets of size ε necessary to cover $\mathcal{F}$. (An ε-bracket is a set $[f^L, f^U] = \{f : f^L \le f \le f^U, \text{ a.s. } [\lambda]\}$ and $d(f^L, f^U) < \varepsilon$.) Recall that the Hellinger distance between any pair of probability measures P and Q dominated by a σ-finite measure λ, f_P and f_Q being the respective densities, is defined as

$$d_H(P, Q) = \left\{ \int_{\mathcal{X}} \left(\sqrt{f_P(x)} - \sqrt{f_Q(x)} \right)^2 \, d\lambda(x) \right\}^{1/2}.$$

Theorem 1.1. [Ghosal and van der Vaart (2000)]. *Let $\{\pi_n\}_{n=1}^{\infty}$ be a sequence of prior distributions on a set $\mathcal{F} = \{P : P \ll \lambda\}$. Suppose there exist positive sequences $\bar{\varepsilon}_n$ and $\tilde{\varepsilon}_n$ such that $\varepsilon_n = \max\{\bar{\varepsilon}_n, \tilde{\varepsilon}_n\} \to 0$ and $n \min\{\bar{\varepsilon}_n^2, \tilde{\varepsilon}_n^2\} \to \infty$, as $n \to \infty$, positive constants c_1, c_2, c_3, c_4 and sets $\mathcal{F}_n \subseteq \mathcal{F}$ such that*

$$\ln D(\bar{\varepsilon}_n, \mathcal{F}_n, d_H) \le c_1 n \bar{\varepsilon}_n^2, \tag{1.1}$$

$$\pi_n(\mathcal{F}_n^c) \le c_3 e^{-(c_2+4) n \tilde{\varepsilon}_n^2}, \tag{1.2}$$

$$\pi_n(B(\tilde{\varepsilon}_n^2)) \ge c_4 e^{-c_2 n \tilde{\varepsilon}_n^2}, \tag{1.3}$$

where $B(\tilde{\varepsilon}_n^2) = \{P : \int f_0 \ln(f_0/f_P) < \tilde{\varepsilon}_n^2,\ \int f_0 (\ln(f_0/f_P))^2 < \tilde{\varepsilon}_n^2\}$. Then, for a sufficiently large constant $K > 0$, the posterior probability $\pi_n(\{P : d_H(f_0, f_P) > K\varepsilon_n\} | X^n)$ tends to zero in P_0^n-probability.

The paper is organized as follows. In Section 2, location mixtures are considered. In Section 3, location-scale mixtures are studied.

2. Convergence rates of posteriors for location mixtures

The generic density is here modeled as a location mixture of normal densities. A Dirichlet process prior is assumed for F. 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 on the k-dimensional simplex with parameters $(\alpha(A_1), \alpha(A_2), \ldots, \alpha(A_k))$.

If the true mixing distribution is assumed to have compact support and the mixing distribution is distributed according to a Dirichlet process prior with a base measure having sufficiently light tails, then the posterior converges at the rate $n^{-(1/2-\beta)}(\ln n)^{\kappa}$, where κ depends on the tail behavior of the base measure and β on the tail behavior of the prior distribution for σ. This rate differs for an n^{β}-factor from the rate of convergence obtained by Ghosal *et al.* (2000) for the case when σ is positively lower bounded.

Theorem 2.1. *Assume that the prior distribution μ for σ possesses a continuous and positive density on an interval containing σ_0 and that, for a constant $\beta \in (0, 1/2)$, $\mu(\{\sigma : 0 < \sigma < s\}) \lesssim e^{-ds^{-2/\beta}}$ for all $s \in (0, \sigma_2]$, d being a positive constant. Suppose that F_0 has a compact support, the base measure α has a continuous and positive density on an interval containing the support of F_0 and, for some $\delta \geq 2$, satisfies the tail condition $\alpha(\{\theta : |\theta| > t\}) \lesssim e^{-bt^{\delta}}$ for every $t > 0$ and a constant $b > 0$. Denoted by $\kappa = \max\{\frac{1}{2}, \frac{2}{\delta}\} + \frac{1}{2}$, the posterior distribution converges in probability to δ_{P_0} at the rate $\varepsilon_n = n^{-(\frac{1}{2}-\beta)}(\ln n)^{\kappa}$.*

Proof. It is shown that conditions of Theorem 1.1 are fulfilled. Let $\eta \in (0, 1/2)$ and $s \in (0, \sigma_2]$ be given. Denote by $\eta' = s^{-1}\eta$. For a positive constant L to be chosen later, let $a \leq L(\ln \frac{1}{\eta})^{2/\delta}$. We shall consider the sets $\mathcal{F}_{a,s} = \{f_{F,\sigma} : F([-a, a]) = 1,\ s \leq \sigma \leq \sigma_2\}$ and $\mathcal{F}_{a,\eta,s} = \{f_{F,\sigma} : F([-a, a]) \geq 1 - \eta,\ s \leq \sigma \leq \sigma_2\}$.

The first step consists in providing an estimate for the η'-metric entropy number of $\mathcal{F}_{a,\eta,s}$. In view of the general relationship $D(\eta', \mathcal{F}, d) \leq N(\eta'/2, \mathcal{F}, d) \leq N_{[\,]}(\eta', \mathcal{F}, d)$, valid for any metric d on $\mathcal{F}$, it suffices to find an estimate for the bracketing metric entropy of $\mathcal{F}_{a,\eta,s}$. Set $\vartheta = \max\{\frac{1}{2}, \frac{2}{\delta}\}$, it is

$$\begin{aligned}
\ln N(s^{-1}\eta, \mathcal{F}_{a,\eta,s}, d_H) &\leq \ln N((s^{-1}\eta)^2, \mathcal{F}_{a,\eta,s}, \|\cdot\|_1)\\
&\leq \ln N((s^{-1}\eta)^2/3, \mathcal{F}_{a,s}, \|\cdot\|_1)\\
&\leq \ln N_{[\,]}(2(s^{-1}\eta)^2/3, \mathcal{F}_{a,s}, \|\cdot\|_1)\\
&\leq Cs^{-2}\left(\ln \frac{1}{\eta}\right)^{2\vartheta+1}\\
&\lesssim s^{-2}\left(\ln \frac{1}{\eta}\right)^{2\kappa},
\end{aligned}$$

where the first inequality descends from $d_H(f, g)^2 \leq \|f - g\|_1$, the second line from Lemma A.3 by Ghosal *et al.* (2000, pages 26-27) and the fourth line from Lemma A.4. Choose $\tilde{\varepsilon}_n = (\ln n)/\sqrt{n}$, $a_n = L(\ln \frac{1}{\tilde{\varepsilon}_n})^{2/\delta}$, $\eta_n = (\ln n)/\sqrt{n}$ and $s_n = n^{-\beta}$ and $\bar{\varepsilon}_n = (\ln n)^{\kappa}/n^{(1/2-\beta)}$. Let $\mathcal{F}_n = \mathcal{F}_{a_n,\eta_n,s_n}$. Then, the condition $\eta_n < s_n$ is fulfilled and, as $(\ln \frac{1}{\eta_n})^2 \lesssim n\eta_n^2$, condition (1.1) is fulfilled:

$$\ln D(\bar{\varepsilon}_n, \mathcal{F}_n, d_H) \lesssim s_n^{-2}\left(\ln \frac{1}{\eta_n}\right)^{2\kappa} \lesssim s_n^{-2}(n\eta_n^2)^{\kappa} = n\bar{\varepsilon}_n^2.$$

The second step consists in estimating $\pi(\mathcal{F}^c_{a,\eta,s})$. Using the independence of F and σ, we get that

$$\begin{aligned}
\pi(\mathcal{F}^c_{a,\eta,s}) &= \pi(\{(F,\sigma):\, F([-a,\,a]) \geq 1-\eta,\ s \leq \sigma \leq \sigma_2\}^c) \\
&\leq \pi(\{(F,\sigma):\, F([-a,\,a]) \geq 1-\eta,\ 0 < \sigma < s\}) \\
&\qquad\qquad + \pi(\{(F,\sigma):\, F([-a,\,a]) < 1-\eta,\ 0 < \sigma \leq \sigma_2\}) \\
&= \Pi(\{F:\, F([-a,\,a]) \geq 1-\eta\})\,\mu(\{\sigma:\, 0<\sigma<s\}) \\
&\qquad\qquad + \Pi(\{F:\, F([-a,\,a]) < 1-\eta\})\,\mu(\{\sigma:\, 0<\sigma\leq\sigma_2\}) \\
&\leq \mu(\{\sigma:\, 0<\sigma<s\}) + \Pi(\{F:\, F([-a,\,a]^c) > \eta\}) \\
&\leq \mu(\{\sigma:\, 0<\sigma<s\}) + \frac{\alpha([-a,\,a]^c)}{\eta\alpha(\mathbb{R})} \qquad (2.1) \\
&\lesssim e^{-ds^{-2/\beta}} + \frac{e^{-ba^\delta}}{\eta},
\end{aligned}$$

where in (2.1) we used Markov's inequality together with the fact that $F([-a,\,a]^c)$ has a beta distribution with parameters $\alpha([-a,\,a]^c)$ and $\alpha([-a,\,a])$. Take a_n, η_n, s_n and $\mathcal{F}_n$ as before. Let $L > [(4(c_2+4)+1)/b]^{1/\delta}$, then

$$\begin{aligned}
\pi(\mathcal{F}^c_n) &\lesssim e^{-ds_n^{-2/\beta}} + \frac{e^{-ba_n^\delta}}{\eta_n} \\
&\lesssim \exp\left\{-(c_2+4)\frac{d}{c_2+4}n^2\right\} + \exp\left\{\left(\ln\frac{1}{\tilde{\varepsilon}_n}\right) - [4(c_2+4)+1]\left(\ln\frac{1}{\tilde{\varepsilon}_n}\right)^2\right\} \\
&\lesssim \exp\{-(c_2+4)(\ln n)^2\} + \exp\left\{-4(c_2+4)\left(\ln\frac{1}{\tilde{\varepsilon}_n}\right)^2\right\} \\
&\lesssim \exp\{-(c_2+4)n\tilde{\varepsilon}_n^2\} + \exp\left\{-4(c_2+4)\left(\ln\frac{\sqrt{n}}{\ln n}\right)^2\right\} \\
&\lesssim \exp\{-(c_2+4)n\tilde{\varepsilon}_n^2\} + \exp\left\{-4(c_2+4)\left(\ln n^{-1/2}\right)^2\right\} \\
&\lesssim \exp\{-(c_2+4)n\tilde{\varepsilon}_n^2\}.
\end{aligned}$$

The last step consists in finding an estimate of the probability $\pi(B(\varepsilon^2))$ for small ε. From Theorem 3.1 by Ghosal *et al.* (*ibidem*, pages 9-10), it is known that, for positive constants c_1 and c_2, $\pi(B(\varepsilon^2)) \geq c_1 \exp\{-c_2 (\ln \frac{1}{\varepsilon})^2\}$. The sequence $\widetilde{\varepsilon}_n = (\ln n)/\sqrt{n}$ satisfies condition (1.3): $\pi(B(\widetilde{\varepsilon}_n^2)) \gtrsim \exp\{-c_2 n \widetilde{\varepsilon}_n^2\}$. □

If $\delta = 2$ is taken, then a normal base measure α is elicited and the rate of convergence is $n^{-(1/2-\beta)}(\ln n)^{3/2}$. If the hypothesis on the compact support of the true mixing distribution is relaxed to the assumption that F_0 has sub-Gaussian tails, then the same rate is obtained. Note that the best rate $\varepsilon_n = n^{-(1/2-\beta)}(\ln n)$ is obtained for $\delta \geq 4$.

3. Convergence rates of posteriors for location-scale mixtures

The generic density is herein assumed to be modeled as a location-scale mixture of normal densities, where the location and scale parameters are either distributed according to independent Dirichlet process priors,

$$f_{F,G}(x) = \int_0^{\sigma_2} \int_{-\infty}^{\infty} \phi_\sigma(x-\theta)\, \mathrm{d}F(\theta)\, \mathrm{d}G(\sigma), \qquad x \in \mathbb{R},$$

or to a unique Dirichlet process,

$$f_F(x) = \int_0^{\sigma_2} \int_{-\infty}^{\infty} \phi_\sigma(x-\theta)\, \mathrm{d}F(\theta,\, \sigma) \qquad x \in \mathbb{R}.$$

Theorem 3.1. *Suppose the true mixing distribution F_0 to have a compact support. Assume that the base measure α is as in Theorem 2.1. Assume G_0 to be supported on an interval $[\sigma_1, \sigma_2]$, with $\sigma_1 > 0$. Let the base measure γ of the Dirichlet process prior for G to be supported on $(0, \sigma_2]$, wherein it has a positive and continuous density. Assume that it satisfies the tail condition $\gamma(\{\sigma :\, 0 < \sigma < s\}) \lesssim e^{-ds^{-2/\beta}}$, for all $s \in (0, \sigma_2]$ and constants $d > 1$ and $\beta \in (0, 1/2)$. Then, the posterior probability $\pi(H^c_{\varepsilon_n}|X^n)$ tends to zero in P_0^n-probability at the rate $\varepsilon_n = n^{-(\frac{1}{2}-\beta)}(\ln n)^\kappa$, with $\kappa = \max\{\frac{1}{2}, \frac{2}{\delta}\} + \frac{1}{2}$.*

Proof. The proof develops along the same lines as those in Theorem 2.1. The proof is split into three steps each one corresponding to the verification of a condition of Theorem 1.1.

For $\eta \in (0, 1/2)$ and L a sufficiently large constant, let $a \leq L(\ln \frac{1}{\eta})^{2/\delta}$. Let $s \in (0, 1)$. Define the sets $\mathcal{F}_{a,s} = \{f_{F,G} :\, F([-a,\, a]) = 1,\, G([s,\, \sigma_2]) = 1\}$ and $\mathcal{F}_{a,\eta,s} = \{f_{F,G} :\, F([-a,\, a]) \geq 1-\eta,\, G([s,\, \sigma_2]) \geq 1-s\}$. In view of the arguments laid out in Theorem 2.1, we need an estimate of $\ln N_{[\,]}(2(s^{-1}\eta)^2/3,\, \mathcal{F}_{a,s},\, \|\cdot\|_1)$. The following developments show how a net over $\mathcal{F}_{a,s}$ can be constructed and how an estimate of its cardinality can be obtained. Let F', G' be discrete distributions on $[-a,\, a]$, $[s,\, \sigma_2]$, respectively, that satisfy the following constraints

$$\int_{-a}^{a} (x-\theta)^{2j}\, \mathrm{d}F'(\theta) = \int_{-a}^{a} (x-\theta)^{2j}\, \mathrm{d}F(\theta), \qquad j = 1, \ldots, N-1,$$

$$\int_{s}^{\sigma_2} \sigma^{-(2j+1)}\, \mathrm{d}G'(\sigma) = \int_{s}^{\sigma_2} \sigma^{-(2j+1)}\, \mathrm{d}G(\sigma), \qquad j = 0, \ldots, N-1.$$

Set $M=\max\{2a,\,\sqrt{8}\sigma_2(\ln\frac{1}{\eta})^{1/2}\}$ and taken $N\le Ds^{-2}(\ln\frac{1}{\eta})^{2\vartheta}$, with D a positive constant and $\vartheta=\max\{\frac{1}{2},\,\frac{2}{\delta}\}$, by reasoning as in Lemma A.1, we get that

$$
\begin{aligned}
\|f_{F,G}-f_{F',G'}\|_\infty &\le \sup_{|x|\le M}|f_{F,G}(x)-f_{F',G'}(x)|+\sup_{|x|\ge M}|f_{F,G}(x)-f_{F',G'}(x)|\\
&\le \sup_{|x|\le M}\Bigg|f_{F,G}(x)\mp\int_s^{\sigma_2}\int_{-a}^{a}\sum_{j=0}^{N-1}a_j\sigma^{-(2j+1)}(x-\theta)^{2j}\,\mathrm{d}F(\theta)\,\mathrm{d}G(\sigma)\\
&\qquad\mp\int_s^{\sigma_2}\int_{-a}^{a}\sum_{j=0}^{N-1}a_j\sigma^{-(2j+1)}(x-\theta)^{2j}\,\mathrm{d}F'(\theta)\,\mathrm{d}G'(\sigma)-f_{F',G'}(x)\Bigg|\\
&\qquad+\int_s^{\sigma_2}\phi_\sigma(M-a)\,\mathrm{d}G(\sigma)+\int_s^{\sigma_2}\phi_\sigma(M-a)\,\mathrm{d}G'(\sigma)\\
&\le \int_s^{\sigma_2}\sup_{\substack{|x|\le M\\|\theta|\le a}}\Bigg|\phi_\sigma(x-\theta)-\sum_{j=0}^{N-1}a_j\sigma^{-(2j+1)}(x-\theta)^{2j}\Bigg|\,\mathrm{d}G(\sigma)\\
&\qquad+\int_s^{\sigma_2}\sup_{\substack{|x|\le M\\|\theta|\le a}}\Bigg|\phi_\sigma(x-\theta)-\sum_{j=0}^{N-1}a_j\sigma^{-(2j+1)}(x-\theta)^{2j}\Bigg|\,\mathrm{d}G'(\sigma)\\
&\qquad+\sum_{j=0}^{N-1}a_j\sup_{|x|\le M}\Bigg|\int_s^{\sigma_2}\int_{-a}^{a}\sigma^{-(2j+1)}(x-\theta)^{2j}\,\mathrm{d}F(\theta)\,\mathrm{d}G(\sigma)\\
&\qquad\qquad-\int_s^{\sigma_2}\int_{-a}^{a}\sigma^{-(2j+1)}(x-\theta)^{2j}\,\mathrm{d}F'(\theta)\,\mathrm{d}G'(\sigma)\Bigg|\\
&\qquad+\sqrt{2/\pi}s^{-1}\eta\\
&\le 2\frac{(2\pi)^{-1/2}s^{-1}\left(\sqrt{e/2}\,s^{-1}\,|x-\theta|\right)^{2N}}{N^N}+\sqrt{2/\pi}s^{-1}\eta\\
&\lesssim s^{-1}\eta.
\end{aligned}
$$

Consider the following set of densities $f_{F,G}$, with F and G discrete distributions: F is supported on at most N points $\{0,\,\pm s^2\eta^2,\,\pm 2s^2\eta^2,\ldots\}$ of $[-a,\,a]$, each point being associated a weight coming from an η-net for the l_1-distance over the N-simplex; analogously, G is supported on at most N points $\{s^2\eta^2,\,2s^2\eta^2,\,\ldots\}$ of $[s,\,\sigma_2]$, with weights coming from an η-net for the N-simplex. In force of the previous arguments, this set of densities is a $k\eta$-net for $\mathcal{F}_{a,s}$, where k is some positive constant, and its cardinality can be bounded above by

$$\left(\frac{\sigma_2 - s}{s^2\eta^2}\right)^N \times \left(\frac{2a}{s^2\eta^2}\right)^N \times \left(\frac{5}{\eta}\right)^N \times \left(\frac{5}{\eta}\right)^N = (50a(\sigma_2 - s))^N \eta^{-6N} s^{-4N}.$$

Therefore, if $\eta < s$, noting that $0 \le (2/\delta) \le 1$,

$$\begin{aligned}\ln N(c_1 s^{-1}\eta, \mathcal{F}_{a,s}, \|\cdot\|_\infty) &\le N\ln(50a) + 6N\left(\ln\frac{1}{\eta}\right) + 4N\left(\ln\frac{1}{s}\right)\\ &\lesssim s^{-2}\left(\ln\frac{1}{\eta}\right)^{2\vartheta+2/\delta} + s^{-2}\left(\ln\frac{1}{\eta}\right)^{2\vartheta+1} + s^{-2}\left(\ln\frac{1}{\eta}\right)^{2\vartheta}\left(\ln\frac{1}{s}\right)\\ &\lesssim s^{-2}\left(\ln\frac{1}{\eta}\right)^{2\vartheta+1}.\end{aligned}$$

Consequently, since Lemma A.4 goes through immediately,

$$\ln N(s^{-1}\eta, \mathcal{F}_{a,\eta,s}, d_H) \le s^{-2}\left(\ln\frac{1}{\eta}\right)^{2\kappa}.$$

Choose $\tilde{\varepsilon}_n = (\ln n)/\sqrt{n}$, $\eta_n = (\ln n)/\sqrt{n}$, $a_n = L(\ln\frac{1}{\tilde{\varepsilon}_n})^{2/\delta}$, $s_n = n^{-\beta}$ and $\bar{\varepsilon}_n = n^{-(1/2-\beta)}(\ln n)^\kappa$. Then, the condition $\eta_n < s_n$ is satisfied. Let $\mathcal{F}_n = \mathcal{F}_{a_n,\eta_n,s_n}$. As $(\ln\frac{1}{\eta_n})^2 \lesssim n\eta_n^2$,

$$\ln D(\bar{\varepsilon}_n, \mathcal{F}_n, d_H) \lesssim s_n^{-2}\left(\ln\frac{1}{\eta_n}\right)^{2\kappa} \lesssim n\bar{\varepsilon}_n^2.$$

Now, an estimate of $\pi(\mathcal{F}^c_{a,\eta,s})$ is provided. Using the independence of the Dirichlet process priors for F and σ, we get that

$$\begin{aligned}
\pi(\mathcal{F}^c_{a,\eta,s}) &= \pi(\{(F,\sigma):\, F([-a,\,a]) \geq 1-\eta,\ G([s,\,\sigma_2]) \geq 1-s\}^c)\\
&\leq \Pi_1(\{F:\, F([-a,\,a]) \geq 1-\eta\})\,\Pi_2(\{G:\, G([s,\,\sigma_2]) < 1-s\})\\
&\qquad\qquad + \Pi_1(\{F:\, F([-a,\,a]) < 1-\eta\})\,\Pi_2(\{G:\, G([s,\,\sigma_2]) \geq 1-s\})\\
&\leq \Pi_2(\{G:\, G([s,\,\sigma_2]) < 1-s\}) + \Pi_1(\{F:\, F([-a,\,a]) < 1-\eta\})\\
&= \Pi_2(\{G:\, G([s,\,\sigma_2]^c) > s\}) + \Pi_1(\{F:\, F([-a,\,a]^c) > \eta\})\\
&\leq \frac{\gamma((0,\,s))}{s\gamma((0,\,\sigma_2])} + \frac{\alpha([-a,\,a]^c)}{\eta\alpha(\mathbb{R})}\\
&\lesssim \frac{e^{-ds^{-2/\beta}}}{s} + \frac{e^{-ba^\delta}}{\eta}.
\end{aligned}$$

For $\mathcal{F}_n = \mathcal{F}_{a_n,\eta_n,s_n}$, taken $L > [(4(c_2+4)+1)/b]^{1/\delta}$, condition (1.2) is satisfied:

$$\begin{aligned}
\pi(\mathcal{F}^c_n) &\lesssim \exp\left\{\left(\ln\frac{1}{s_n}\right) - d\left(\frac{1}{s_n}\right)^{2/\beta}\right\} + \exp\{-(c_2+4)n\tilde{\varepsilon}^2_n\}\\
&\lesssim \exp\left\{-(d-1)\left(\frac{1}{s_n}\right)^{2/\beta}\right\} + \exp\{-(c_2+4)n\tilde{\varepsilon}^2_n\}\\
&\lesssim \exp\{-(c_2+4)n\tilde{\varepsilon}^2_n\}.
\end{aligned}$$

Now, a lower bound on $\pi(B(\varepsilon^2))$ is provided. We need to show that analogue assertions to those of Lemma 3.3, Lemma 3.6 and Lemma 3.5 of Ghosal *et al.* (2000, page 8) hold true. Lemma 3.3 goes over immediately. Lemma 3.6 requires some care. Let F_0' and G_0' be discrete distributions on $[-k_0,\,k_0]$ and $[\sigma_1,\,\sigma_2]$, respectively, such that $\left\|f_{F_0,G_0} - f_{F_0',G_0'}\right\|_\infty \lesssim \varepsilon$. They can be taken to have at most $N \lesssim \ln\frac{1}{\varepsilon}$ support points. Represent F_0' and G_0' as $\sum_{j=1}^N p_j\delta_{\theta_j}$ and $\sum_{k=1}^N q_k\delta_{\sigma_k}$, respectively. Without loss of generality, the points θ_j and σ_k can be taken to be 2ε-separated, i.e., $|\theta_j - \theta_{j'}| > 2\varepsilon$ for all $j \neq j'$ and $|\sigma_k - \sigma_{k'}| > 2\varepsilon$ for all $k \neq k'$. It is

$$
\begin{aligned}
\|f_{F,G} - f_{F_0',G_0'}\|_1 &= \int_{-\infty}^{\infty} \bigg| \int_0^{\sigma_2} \int_{-\infty}^{\infty} \phi_\sigma(x-\theta)\,\mathrm{d}F(\theta)\,\mathrm{d}G(\sigma) \\
&\quad \mp \sum_{k=1}^N \sum_{j=1}^N \int_{|\theta-\theta_j|\le\varepsilon,\ |\sigma-\sigma_k|\le\varepsilon} \phi_{\sigma_k}(x-\theta)\,\mathrm{d}F(\theta)\,\mathrm{d}G(\sigma) \\
&\quad \mp \sum_{k=1}^N \sum_{j=1}^N \phi_{\sigma_k}(x-\theta_j) F([\theta_j-\varepsilon,\, \theta_j+\varepsilon])\, G([\sigma_k-\varepsilon,\, \sigma_k+\varepsilon]) \\
&\quad - \sum_{k=1}^N \sum_{j=1}^N \phi_{\sigma_k}(x-\theta_j) p_j q_k \bigg|\,\mathrm{d}x \\
&\lesssim (F\times G)(\{(\theta,\sigma) : |\theta-\theta_j| > \varepsilon\ \forall j \text{ or } |\sigma-\sigma_k| > \varepsilon\ \forall k\}) \\
&\quad + \sum_{k=1}^N \sum_{j=1}^N \int_{|\theta-\theta_j|\le\varepsilon,\ |\sigma-\sigma_k|\le\varepsilon} \|\phi_\sigma(\cdot-\theta) - \phi_{\sigma_k}(\cdot-\theta)\|_1\,\mathrm{d}F(\theta)\,\mathrm{d}G(\sigma) \\
&\quad + \sum_{k=1}^N \sum_{j=1}^N \int_{|\theta-\theta_j|\le\varepsilon,\ |\sigma-\sigma_k|\le\varepsilon} \|\phi_{\sigma_k}(\cdot-\theta) - \phi_{\sigma_k}(\cdot-\theta_j)\|_1\,\mathrm{d}F(\theta)\,\mathrm{d}G(\sigma) \\
&\quad + \sum_{k=1}^N \sum_{j=1}^N |F([\theta_j-\varepsilon,\, \theta_j+\varepsilon])\, G([\sigma_k-\varepsilon,\, \sigma_k+\varepsilon]) - p_j q_k| \\
&\lesssim 1 - \sum_{k=1}^N \sum_{j=1}^N F([\theta_j-\varepsilon,\, \theta_j+\varepsilon])\, G([\sigma_k-\varepsilon,\, \sigma_k+\varepsilon]) \\
&\quad + \sum_{k=1}^N \sum_{j=1}^N \int_{|\theta-\theta_j|\le\varepsilon,\ |\sigma-\sigma_k|\le\varepsilon} |\sigma-\sigma_k|\,\mathrm{d}F(\theta)\,\mathrm{d}G(\sigma) \\
&\quad + \sum_{k=1}^N \sum_{j=1}^N \int_{|\theta-\theta_j|\le\varepsilon,\ |\sigma-\sigma_k|\le\varepsilon} \frac{1}{\sigma_k}|\theta-\theta_j|\,\mathrm{d}F(\theta)\,\mathrm{d}G(\sigma) \\
&\quad + \sum_{k=1}^N \sum_{j=1}^N |F([\theta_j-\varepsilon,\, \theta_j+\varepsilon])\, G([\sigma_k-\varepsilon,\, \sigma_k+\varepsilon]) - p_j q_k| \\
&\lesssim \varepsilon + \sum_{j=1}^N |F([\theta_j-\varepsilon,\, \theta_j+\varepsilon]) - p_j| + \sum_{k=1}^N |G([\sigma_k-\varepsilon,\, \sigma_k+\varepsilon]) - q_k|\,.
\end{aligned}
$$

Also Lemma 3.5 goes through, consequently,

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

$$\subseteq B\left(c\varepsilon\left(\ln\frac{1}{\varepsilon}\right)^{5/2}\right)$$

and

$$\pi\left(B\left(\varepsilon\left(\ln\frac{1}{\varepsilon}\right)^{5/2}\right)\right)\geq\Pi_1\left(\left\{F:\,\sum_{j=1}^{N}|F([\theta_j-\varepsilon,\,\theta_j+\varepsilon])-p_j|\leq\varepsilon\right\}\right)$$

$$\times\Pi_2\left(\left\{G:\,\sum_{k=1}^{N}|G([\sigma_k-\varepsilon,\,\sigma_k+\varepsilon])-q_k|\leq\varepsilon\right\}\right)$$

$$\geq c_1\exp\left\{-c_2\left(\ln\frac{1}{\varepsilon}\right)^2\right\},$$

for $c_1,\,c_2>0$. Condition (1.3) is verified for the sequence $\tilde{\varepsilon}_n=(\ln n)/\sqrt{n}$. □

Theorem 3.2. *Suppose that F_0 has a compact support, say $[-k_0,\,k_0]\times[\sigma_1,\,\sigma_2]$. Assume that the base measure α of the Dirichlet process prior for F has a positive and continuous density on a rectangle containing $[-k_0,\,k_0]\times(0,\,\sigma_2]$ and satisfies the tail condition: $\alpha(\{(\theta,\,\sigma):\,|\theta|>t,\,0<\sigma<s\})\lesssim e^{-bt^{\delta}}e^{-ds^{-3/\beta}}$ for all $s,\,t>0$ and constants $b>0$, $\delta\geq3$, $d>1$ and $\beta\in(0,\,1/2)$. Then, the posterior probability tends to zero in P_0^n-probability at the rate $\varepsilon_n=n^{-(\frac{1}{2}-\beta)}(\ln n)^{\kappa}$, where $\kappa=2\max\{\frac{1}{2},\,\frac{3}{\delta}\}+\frac{1}{2}$.*

Proof. The claim can be proved along the same lines as those of the proof of Theorem 3.1. Attention is called on those steps requiring some changes. Given $\eta>0$, let $a\leq L(\ln\frac{1}{\eta})^{3/\delta}$, with L chosen later. We need an estimate of the entropy of the class $\mathcal{F}_{a,s}=\{f_F:\,F([-a,\,a]\times[s,\,\sigma_2])=1\}$. In view of Lemma A.1 by Ghosal *et al.* (*ibidem*, pages 25-26), for any $F\in\mathcal{F}_{a,s}$, a discrete distribution F' supported on points of $[-a,\,a]\times[s,\sigma_2]$ can be found such that

$$\int_s^{\sigma_2}\int_{-a}^{a}(x-\theta)^{2j}\sigma^{-(2k+1)}\,\mathrm{d}F(\theta,\,\sigma)$$

$$=\int_s^{\sigma_2}\int_{-a}^{a}(x-\theta)^{2j}\sigma^{-(2k+1)}\,\mathrm{d}F'(\theta,\,\sigma),\qquad\forall\,j,\,k\in\{0,\,\ldots,\,N-1\}.$$

Set $\vartheta = \max\{\frac{1}{2}, \frac{3}{\delta}\}$, F' can be chosen to have at most $N \lesssim s^{-2}(\ln \frac{1}{\eta})^{4\vartheta}$ support points. Defined the set $\mathcal{F}_{a,\eta,s} = \{f_F : F([-a, a] \times [s, \sigma_2]) \geq 1 - s\eta\}$, by standard arguments $\ln N(s^{-1}\eta, \mathcal{F}_{a,\eta,s}, d_H) \lesssim s^{-2}(\ln \frac{1}{\eta})^{4\vartheta+1}$.

The other relevant difference is in the estimate of the prior probability of a Kullback-Leibler type neighborhood. A discrete distribution F_0' on $\mathrm{supp}(F_0)$ having $N \lesssim (\ln \frac{1}{\varepsilon})^2$ support points can be found such that $\|f_{F_0} - f_{F_0'}\|_\infty \lesssim \varepsilon$. It can be represented as $F_0' = \sum_{j=1}^N p_j \delta_{(\theta_j, \sigma_j)}$. Assuming the sets $[\theta_j - \varepsilon, \theta_j + \varepsilon] \times [\sigma_j - \varepsilon, \sigma_j + \varepsilon]$, $j = 1, \ldots, N$, to be disjoint, it can be seen that, for any F,

$$\|f_F - f_{F_0'}\|_1 \lesssim \varepsilon + \sum_{j=1}^N |F([\theta_j - \varepsilon, \theta_j + \varepsilon] \times [\sigma_j - \varepsilon, \sigma_j + \varepsilon]) - p_j| .$$

The prior probability of $B(\varepsilon^2)$ can be lower bounded by the prior probability of the set $\{F : \sum_{j=1}^N |F([\theta_j - \varepsilon, \theta_j + \varepsilon] \times [\sigma_j - \varepsilon, \sigma_j + \varepsilon]) - p_j| \leq \varepsilon\}$, which is in turn bounded below by $\exp\{-c(\ln \frac{1}{\varepsilon})^3\}$, for some constant $c > 0$.

The result follows by choosing $\widetilde{\varepsilon}_n = (\ln n)^{3/2}/\sqrt{n}$, $a_n = L(\ln \frac{1}{\widetilde{\varepsilon}_n})^{3/\delta}$, with $L > [(c_2 + 4 + 1)/b]^{1/\delta}$, $s_n = n^{-\beta}$, $\eta_n = (\ln n)/\sqrt{n}$, $\bar{\varepsilon}_n = n^{-(\frac{1}{2} - \beta)}(\ln n)^\kappa$. □

The slower rate of convergence obtained when the location and scale parameters are not independent is due to the many more moments to be matched.

In all the cases taken into account, the Bayes' density estimator converges at least as fast as the posterior distribution. As already emphasized, we do not know the optimal rate of convergence for point estimators of mixtures of normals in the Hellinger distance. Until the optimal rate is established, we can only try to improve the already known best rates, for instance by replacing the Dirichlet process with a Pólya tree.

Previous results are heavily based on the hypothesis that the true mixing distribution has a compact support. This is a restrictive assumption and it would be better to remove it. As a first step, one could investigate whether previous theorems can be restated for the case when the true mixing distribution has sub-Gaussian tails. A further step would consist in studying the case when the sampling density is compactly supported.

Appendix A. Auxiliary results

Some lemmas providing basic tools used to derive the main results are herein stated and proved. They are similar to lemmas in Ghosal *et al.* (2000, page 8).

Lemma A.1. *Let $\varepsilon \in (0, 1/2)$ and $s \in (0, 1)$ be fixed. Let $\sigma, \sigma' \in [s, \sigma_2]$ be such that $|\sigma' - \sigma| < \varepsilon$. Let a be a positive number such that $a \leq L\left(\ln \frac{1}{\varepsilon}\right)^\gamma$, where $L > 0$ and $0 \leq \gamma \leq 1$ are constants. For any probability measure F on $[-a, a]$, there exists a discrete probability measure F' supported on at most $N \lesssim s^{-2}\left(\ln \frac{1}{\varepsilon}\right)^{2\vartheta}$ points of $[-a, a]$, with $\vartheta = \max\{\frac{1}{2}, \gamma\}$, such that*

$$\|f_{F,\sigma} - f_{F',\sigma'}\|_\infty \lesssim s^{-1}\varepsilon, \tag{A.1}$$

with $s^{-1}\varepsilon$ arbitrarily close to zero.

Proof. In what follows, let ε' stand for $s^{-1}\varepsilon$. In view of the triangle inequality

$$\|f_{F,\sigma} - f_{F',\sigma'}\|_\infty \le \|f_{F,\sigma} - f_{F',\sigma}\|_\infty + \|f_{F',\sigma} - f_{F',\sigma'}\|_\infty ,$$

and the fact that, for any probability measure F',

$$\|f_{F',\sigma} - f_{F',\sigma'}\|_\infty \le \|\phi_\sigma - \phi_{\sigma'}\|_\infty \lesssim |\sigma - \sigma'| < \varepsilon < s^{-1}\varepsilon = \varepsilon', \tag{A.2}$$

it is enough to prove that $\|f_{F,\sigma} - f_{F',\sigma}\|_\infty \lesssim \varepsilon'$. First, note that, for any $M > 0$,

$$\begin{aligned} \|f_{F,\sigma} - f_{F',\sigma}\|_\infty &\le \sup_{|x|\le M} |f_{F,\sigma}(x) - f_{F',\sigma}(x)| + \sup_{|x|\ge M} |f_{F,\sigma}(x) - f_{F',\sigma}(x)| \\ &= S_1 + S_2. \end{aligned} \tag{A.3}$$

Set the position $M = \max\{2a, \sqrt{8}\sigma_2 \left(\ln \frac{1}{\varepsilon}\right)^{1/2}\}$, if F' is a probability measure on $[-a, a]$, then the second term in (A.3) turns out to be bounded above by ε':

$$S_2 = \sup_{|x|\ge M} |f_{F,\sigma}(x) - f_{F',\sigma}(x)| \le \sqrt{\frac{2}{\pi}}\, s^{-1}\varepsilon \lesssim \varepsilon'. \tag{A.4}$$

Now, consider the term S_1:

$$
\begin{aligned}
S_1 &= \sup_{|x|\le M}\left| f_{F,\sigma}(x) \mp \int_{-a}^{a}\sum_{j=0}^{N-1} a_j\sigma^{-(2j+1)}(x-\theta)^{2j}\,\mathrm{d}F(\theta)\right.\\
&\qquad\left. \mp \int_{-a}^{a}\sum_{j=0}^{N-1} a_j\sigma^{-(2j+1)}(x-\theta)^{2j}\,\mathrm{d}F'(\theta) - f_{F',\sigma}(x)\right|\\
&\le \sup_{|x|\le M}\left|\int_{-a}^{a}\sum_{j=0}^{N-1} a_j\sigma^{-(2j+1)}(x-\theta)^{2j}\,\mathrm{d}F(\theta) - \int_{-a}^{a}\sum_{j=0}^{N-1} a_j\sigma^{-(2j+1)}(x-\theta)^{2j}\,\mathrm{d}F'(\theta)\right|\\
&\qquad + 2\sup_{\substack{|x|\le M\\ |\theta|\le a}}\left|\phi_\sigma(x-\theta) - \sum_{j=0}^{N-1} a_j\sigma^{-(2j+1)}(x-\theta)^{2j}\right|\\
&\le \sum_{j=0}^{N-1} a_j\sigma^{-(2j+1)} \sup_{|x|\le M}\left|\int_{-a}^{a}(x-\theta)^{2j}\,\mathrm{d}F(\theta) - \int_{-a}^{a}(x-\theta)^{2j}\,\mathrm{d}F'(\theta)\right|\\
&\qquad + 2\sup_{\substack{|x|\le M\\ |\theta|\le a}}\left|\phi_\sigma(x-\theta) - \sum_{j=0}^{N-1} a_j\sigma^{-(2j+1)}(x-\theta)^{2j}\right|, \qquad \text{(A.5)}
\end{aligned}
$$

where, for constants $a_j = (-1)^j 2^{-(j+1/2)}(\sqrt{\pi}j!)^{-1}$, $j = 0, \ldots, N-1$, the sum $\sum_{j=0}^{N-1} a_j\sigma^{-(2j+1)}(x-\theta)^{2j}$ is the Taylor expansion of order $N-1$ of ϕ_σ. In force of Lemma A.1 by Ghosal *et al.* (2000, pages 25-26), F' can be chosen to be a discrete distribution on $[-a, a]$ with at most N support points, such that the $N-1$ constraints

$$
\int_{-a}^{a}(x-\theta)^{2j}\,\mathrm{d}F'(\theta) = \int_{-a}^{a}(x-\theta)^{2j}\,\mathrm{d}F(\theta), \qquad j = 1, \ldots, N-1,
$$

are fulfilled. With this choice, the first addendum of (A.5) is zero. In order to upper bound (A.5), note that for $|x| \le M$ and $|\theta| \le a$, it is $|x-\theta| \le M + a \le 3M/2 \le \max\{3L, \sqrt{18}\sigma_2\}(\ln\frac{1}{\varepsilon})^{\vartheta}$. Taking into account that $\left|e^x - \sum_{j=0}^{N-1} x^j/j!\right| \le (e\,|x|\,/N)^N$, set the position $c = \sqrt{e/2}\max\{3L, \sqrt{18}\sigma_2\}$ and taken

$$
N = \left\lceil (1+c^2)s^{-2}\left(\ln\frac{1}{\varepsilon}\right)^{2\vartheta}\right\rceil,
$$

the following inequalities can be derived to see that (A.5) is bounded above by a constant multiple ε':

$$
\begin{aligned}
\left|\phi_\sigma(x-\theta)-\sum_{j=0}^{N-1} a_j\sigma^{-(2j+1)}(x-\theta)^{2j}\right| &\le \frac{(2\pi)^{-1/2}s^{-1}\left(\sqrt{e/2}\,s^{-1}\,|x-\theta|\right)^{2N}}{N^N}\\
&\le \frac{(2\pi)^{-1/2}s^{-1}\left(cs^{-1}\left(\ln\frac{1}{\varepsilon}\right)^{\vartheta}\right)^{2N}}{N^N}\\
&= (2\pi)^{-1/2}s^{-1}\exp\left\{-N\left[\ln\frac{N}{c^2s^{-2}\left(\ln\frac{1}{\varepsilon}\right)^{2\vartheta}}\right]\right\}\\
&\le (2\pi)^{-1/2}s^{-1}\exp\left\{-s^{-2}\left(\ln\frac{1}{\varepsilon}\right)^{2\vartheta}\ln\left(1+\frac{1}{c^2}\right)^{(1+c^2)}\right\}\\
\left(\text{set } c'=\ln\left(1+\frac{1}{c^2}\right)^{(1+c^2)}\right) \qquad &\le (2\pi)^{-1/2}s^{-1}\exp\left\{-c's^{-2}\ln\frac{1}{\varepsilon}\right\}\\
&= (2\pi)^{-1/2}s^{-1}\varepsilon^{c's^{-2}}\\
&\lesssim s^{-1}\varepsilon=\varepsilon'
\end{aligned}
$$

for s sufficiently small so that the inequality $c's^{-2}>1$ implies that $\varepsilon^{c's^{-2}}<\varepsilon$. It follows that

$$S_1=\sup_{|x|\le M}|f_{F,\sigma}(x)-f_{F',\sigma}(x)|\lesssim\varepsilon',$$

which combined with (A.4) and (A.2) yields (A.1). □

Reasoning as in the preceding lemma the following result can be derived.

Lemma A.2. *Let $\varepsilon\in(0,1/2)$ be fixed. Let $k_0>0$ be a constant and let $\sigma_0\in(0,\sigma_2]$ be given. For any probability measure F_0 on $[-k_0,k_0]$, there exists a discrete probability measure F_0' on $[-k_0,k_0]$ supported on at most $N\lesssim\ln\frac{1}{\varepsilon}$ points such that $\|f_{F_0,\sigma_0}-f_{F_0',\sigma_0}\|_\infty\lesssim\varepsilon$.*

Lemma A.3. *Let $\varepsilon\in(0,1/2)$ and $s\in(0,1)$ be fixed. Let $\mathfrak{F}_{a,s}=\{f_{F,\sigma}: F([-a,a])=1,\ s\le\sigma\le\sigma_2\}$, where $a\le L(\ln\frac{1}{\varepsilon})^{\gamma}$ and $0\le\gamma\le1$. If $\varepsilon<s$, then*

$$\ln N(s^{-1}\varepsilon,\,\mathfrak{F}_{a,s},\,\|\cdot\|_\infty)\lesssim s^{-2}\left(\ln\frac{1}{\varepsilon}\right)^{2\vartheta+1},$$

where $\vartheta=\max\{\frac{1}{2},\gamma\}$ and $\varepsilon'=s^{-1}\varepsilon$ is arbitrarily close to zero.

Proof. Let Σ_ε be an ε-net for $[s, \sigma_2]$. The set $\mathcal{F}_{\varepsilon'}$ of all densities $f_{F,\sigma}$ with σ in Σ_ε and F a discrete distribution on $[-a, a]$ supported on at most $N \leq Ds^{-2}(\ln \frac{1}{\varepsilon})^{2\vartheta}$ points, whose existence is guaranteed by Lemma A.1, is an ε'-net for $\mathcal{F}_{a,s}$ with respect to the distance induced by the sup-norm. An ε'-net for $\mathcal{F}_{\varepsilon'}$ is a $2\varepsilon'$-net for $\mathcal{F}_{a,s}$. Let $\mathcal{S}_\varepsilon$ be an ε-net for the N-dimensional simplex. Let $\mathcal{F}'_{\varepsilon'}$ be the set of all $f_{F',\sigma} \in \mathcal{F}_{\varepsilon'}$ such that $F' = \sum_{j=1}^N p'_j \delta_{\theta'_j}$ with $(p'_1, \ldots, p'_N) \in \mathcal{S}_\varepsilon$ and every $\theta'_j \in \{0, \pm s^2\varepsilon, \pm 2s^2\varepsilon, \ldots\}$. Note that if the interval $[-a, a]$ is divided for $s^2\varepsilon$, then there are at least N points in it. For any $f_{F,\sigma} \in \mathcal{F}_{\varepsilon'}$, an element of $\mathcal{F}'_{\varepsilon'}$ can be found such that their distance is a multiple of ε' and the cardinality of $\mathcal{F}'_{\varepsilon'}$ can be thus upper bounded

$$|\mathcal{F}'_{\varepsilon'}| \lesssim \frac{1}{\varepsilon} \times \left(\frac{5}{\varepsilon}\right)^N \times \left(\frac{2a}{s^2\varepsilon}\right)^N = (10a)^N \varepsilon^{-(2N+1)} s^{-2N}.$$

Keeping in mind that $\varepsilon < s$, for suitable constants $c_1, c_2 > 0$,

$$\begin{aligned}
\ln N(c_1\varepsilon', \mathcal{F}_{a,s}, \|\cdot\|_\infty) &\leq N \ln(10a) + (2N+1)\left(\ln \frac{1}{\varepsilon}\right) + 2N\left(\ln \frac{1}{s}\right) + c_2 \\
&\leq 10aN + 2N\left(\ln \frac{1}{\varepsilon}\right) + \ln \frac{1}{\varepsilon} + 2N\left(\ln \frac{1}{s}\right) + c_2 \\
&\lesssim 10aN + 2N\left(\ln \frac{1}{\varepsilon}\right) \\
&\lesssim s^{-2}\left(\ln \frac{1}{\varepsilon}\right)^{2\vartheta+\gamma} + s^{-2}\left(\ln \frac{1}{\varepsilon}\right)^{2\vartheta+1} \\
&\lesssim s^{-2}\left(\ln \frac{1}{\varepsilon}\right)^{2\vartheta+1}.
\end{aligned}$$

□

Lemma A.4. *Let $\varepsilon \in (0, 1/2)$ and $s \in (0, 1)$ be fixed and such that $\varepsilon' = s^{-1}\varepsilon$ is arbitrarily close to zero. For $\mathcal{F}_{a,s} = \{f_{F,\sigma} : F([-a, a]) = 1,\ s \leq \sigma \leq \sigma_2\}$, with $a \leq L(\ln \frac{1}{\varepsilon})^\gamma$ and $0 \leq \gamma \leq 1$, it is*

$$H_{[\,]}(s^{-1}\varepsilon, \mathcal{F}_{a,s}) \lesssim s^{-2}\left(\ln \frac{1}{\varepsilon}\right)^{2\vartheta+1},$$

where, as usual, $\vartheta = \max\{\frac{1}{2}, \gamma\}$.

Proof. An ε'-Hellinger bracketing for $\mathcal{F}_{a,s}$ is constructed. For any $f_{F,\sigma} \in \mathcal{F}_{a,s}$, it is $f_{F,\sigma}(x) \leq s^{-1}\phi(0)$ for all $x \in \mathbb{R}$. Besides, for $|x| > 2a$, $f_{F,\sigma}(x) \leq s^{-1}\phi(x/2\sigma_2)$. Thus, an envelope function for $\mathcal{F}_{a,s}$ is

$$E(x)=\begin{cases} s^{-1}\phi(0), & |x|\le 2a,\\ s^{-1}\phi\left(\dfrac{x}{2\sigma_2}\right), & |x|>2a.\end{cases}$$

Let $f_1,\,\ldots,\,f_N$ be any ε'-net for $\mathcal{F}_{a,s}$ with respect to $\|\cdot\|_\infty$. For each $j\in\{1,\,\ldots,\,N\}$, define the functions

$$f_j^L(x)=\max\{f_j(x)-\varepsilon',\,0\},\qquad x\in\mathbb{R},$$

$$f_j^U(x)=\min\{f_j(x)+\varepsilon',\,E(x)\},\qquad x\in\mathbb{R},$$

and let

$$[f_j^L,\,f_j^U]=\{f_{F,\sigma}\in\mathcal{F}_{a,s}:\,f_j^L(x)\le f_{F,\sigma}(x)\le f_j^U(x),\ x\in\mathbb{R}\}$$

be the generic bracket. Then, $\mathcal{F}_{a,s}\subseteq\cup_{j=1}^N[f_j^L,\,f_j^U]$. For any $x\in\mathbb{R}$, $[f_j^U(x)-f_j^L(x)]\le\min\{2\varepsilon',\,E(x)\}$. If the position $B=\max\{2L,\,\sqrt{8}\sigma_2\}\left(\ln\frac{1}{\varepsilon}\right)^{\vartheta}$ is set, then $B\ge 2a$ and

$$\begin{aligned}\int_{-\infty}^{\infty}[f_j^U(x)-f_j^L(x)]\,\mathrm{d}x&\le\int_{-\infty}^{\infty}\min\{2\varepsilon',\,E(x)\}\,\mathrm{d}x\\ &\le 2\int_{|x|\le B}\varepsilon'\,\mathrm{d}x+\int_{|x|>B}E(x)\,\mathrm{d}x\\ &\lesssim\varepsilon'\left(\ln\frac{1}{\varepsilon}\right)^{\gamma}+E(B) \qquad\text{(A.6)}\\ &\lesssim\varepsilon'\left(\ln\frac{1}{\varepsilon}\right)^{\gamma}+\varepsilon'\\ &\lesssim\varepsilon'\left(\ln\frac{1}{\varepsilon}\right)^{\gamma},\end{aligned}$$

where (A.6) descends from $\int_{|x|>B}E(x)\,\mathrm{d}x\lesssim s^{-1}\phi(B/2\sigma_2)\le s^{-1}\varepsilon$. It follows that, for some positive constant c,

$$N_{[\,]}\left(c\varepsilon'\left(\ln\frac{1}{\varepsilon}\right)^{\gamma},\,\mathcal{F}_{a,s},\,\|\cdot\|_1\right)\le N.$$

In force of Lemma A.3, it is possible to choose $N\lesssim s^{-2}\left(\ln\frac{1}{\varepsilon}\right)^{2\vartheta+1}$. Putting $\eta^2=c\varepsilon\left(\ln\frac{1}{\varepsilon}\right)^{\gamma}$ and observing that $\ln\frac{1}{\eta}\sim\ln\frac{1}{\varepsilon}$, we have that

$$N_{[\,]}(s^{-1}\eta,\,\mathcal{F}_{a,s},\,d_H)\le N_{[\,]}\left(s^{-1/2}\eta,\,\mathcal{F}_{a,s},\,d_H\right)\le N_{[\,]}(s^{-1}\eta^2,\,\mathcal{F}_{a,s},\,\|\cdot\|_1)\lesssim s^{-2}\left(\ln\frac{1}{\eta}\right)^{2\vartheta+1},$$

where the second inequality from the left descends from the relationship between the Hellinger distance and the L_1-distance. The assertion follows. □

References

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

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

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

C. Scricciolo,
Department of Statistical Sciences,
University of Padova,
Via C. Battisti, 241/243,
35121I - Padova,
Italy
E-mail address: catia@stat.unipd.it