HAL
open science

Concentration rate and consistency of the posterior distribution for selected priors under monotonicity constraints

Jean-Bernard Salomond

To cite this version:

Jean-Bernard Salomond. Concentration rate and consistency of the posterior distribution for selected priors under monotonicity constraints. Electronic Journal of Statistics , 2014, 8 (1), http://projecteuclid.org/euclid.ejs/1408540291. 10.1214/14-EJS929 . hal-01065564

HAL Id: hal-01065564

https://hal.science/hal-01065564v1

Submitted on 18 Sep 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Electronic Journal of Statistics
Vol. 8 (2014) 1380–1404
ISSN: 1935-7524
DOI: 10.1214/14-EJS929

Concentration rate and consistency of the posterior distribution for selected priors under monotonicity constraints

Jean-Bernard Salomond*

CREST – Université Paris Dauphine
3 avenue Pierre Larousse
92245 Malakoff, France
e-mail: jean.bernard.salomond@ensae.fr

Abstract: In this paper, we consider the well known problem of estimating a density function under qualitative assumptions. More precisely, we estimate monotone non-increasing densities in a Bayesian setting and derive concentration rate for the posterior distribution for a Dirichlet process and finite mixture prior. We prove that the posterior distribution based on both priors concentrates at the rate $(n/\log(n))^{-1/3}$, which is the minimax rate of estimation up to a $\log(n)$ factor. We also study the behaviour of the posterior for the point-wise loss at any fixed point of the support of the density and for the sup-norm. We prove that the posterior distribution is consistent for both loss functions.

AMS 2000 subject classifications: 62G07, 62G20.
Keywords and phrases: Density estimation, Bayesian inference, concentration rate.

Received January 2013.

1. Introduction

The non-parametric problem of estimating monotone curves, and monotone densities in particular, has been well studied in the literature both from theoretical and applied perspectives. Shape constrained estimation is fairly popular in the non-parametric literature and widely used in practice (see Robertson et al., 1988, for instance). Monotone densities appear in a wide variety of applications such as survival analysis, where it is natural to assume that the uncensored survival time has a monotone non-increasing density. In these problems, estimating the survival function is equivalent to estimating the survival time density say f and the point-wise estimate $f(0)$. It is thus interesting to have a better understanding of the behaviour of the estimation procedures in this case. An interesting property of monotone non-increasing densities on $\mathbb{R}^+$ is that they have a mixture representation pointed out by Williamson (1956)

$$f(x) = \int_0^\infty \frac{\mathbb{I}_{[0,\theta]}(x)}{\theta} dP(\theta), \tag{1}$$

*This work is part of my PhD thesis under the supervision of Judith Rousseau.

where P is a probability distribution on $\mathbb{R}^+$ called the mixing distribution. In order to emphasize the dependence in P, we will denote f_P the functions admitting representation (1). This representation allows for inference based on the likelihood. Grenander (1956) derived the non-parametric maximum likelihood estimator of a monotone density and Prakasa Rao (1970) studied the behaviour of the Grenander estimator at a fixed point. Groeneboom (1985) and more recently, Balabdaoui and Wellner (2007) studied very precisely the asymptotic properties of the non parametric maximum likelihood estimator. It is proved to be consistent and to converge at the minimax rate $n^{-1/3}$ when the support of the distribution is compact. In their paper, Durot et al. (2012) get some refined asymptotic results for the supremum norm.

The mixture representation of monotone densities lead naturally to a mixture type prior on the set of monotone non increasing densities with support on $[0, L]$ or $\mathbb{R}^+$. For example, Ferguson (1983) and Lo (1984) introduced the Dirichlet Process prior (DP) and Brunner and Lo (1989) considered the special case of unimodal densities with a prior based on a Dirichlet Process mixture. The problem of deriving concentration rates for mixtures models has received a huge interest in the past decade. Wu and Ghosal (2008) studied properties of general mixture models, Ghosal and van der Vaart (2001) studied the well known problem of Gaussian mixtures, Rousseau (2010) derived concentration rates for mixtures of beta distributions, Kruijer et al. (2010) proved good adaptive properties of mixtures of Gaussian. Extensions to the multivariate case have recently been introduced (e.g. Shen et al. (2013)).

Under monotonicity constrained, we derive an upper bound for the posterior concentration rate with respect to some metric or semi-metric $d(\cdot, \cdot)$, that is a positive sequence $(\epsilon_n)_n$ that goes to 0 when n goes to infinity such that

$$\mathrm{E}_0^n \left(\Pi(d(f, f_0) > \epsilon_n | \mathbf{X^n})\right) \to 0,$$

where the expectation is taken under the true distribution P_0 of the data $\mathbf{X^n}$ and where f_0 is the density of P_0 with respect to the Lebesgue measure. Following Khazaei et al. (2010), we study two families of non-parametric priors on the class of monotone non-increasing densities. Interestingly in our setting, the so called Kullback-Leibler property, that is the fact that the prior puts enough mass on Kulback-Leibler neighbourhoods of the true density, is not satisfied. Thus, the approach based on the seminal paper of Ghosal et al. (2000) cannot be directly applied. We therefore use a modified version of their results and obtain for the two families of prior a concentration rate of order $(n/\log(n))^{-1/3}$ which is the minimax estimation rate up to a $\log(n)$ factor under the L_1 or Hellinger distance. We extend these results to densities with support on $\mathbb{R}^+$ and prove that under some conditions on the tail of the distribution, the posterior still concentrates at an almost optimal rate. To the author's knowledge, no concentration rates have been derived for monotone densities on $\mathbb{R}^+$.

Interestingly, the non-parametric maximum likelihood estimator of $f_P(x)$ is not consistent for $x = 0$ (see Sun and Woodroofe (1996) and Balabdaoui and Wellner (2007) for instance). However, we prove that the posterior distribution of f is still consistent at this point under a specific family of non-parametric

mixture priors. In fact, we prove the pointwise consistency of the posterior for all x in $[0, L]$ with $L \leq \infty$. We then derive a consistent Bayesian estimator of the density at any fixed point of the support. This is particularly interesting as the point-wise loss is usually difficult to study in a Bayesian framework as Bayesian approaches are well suited to losses related to the Kullback-Leiber divergence. We also study the behaviour of the posterior distribution for the sup-norm when the density has a compact support. This problem has been addressed recently in the frequentist literature by Durot et al. (2012). They derive refined asymptotic results on the sup-norm of the difference between a Grenander-type estimator and the true density on sub intervals of the form $[\epsilon, L-\epsilon]$, where $\epsilon > 0$ avoiding the problems at the boundaries. Here, we prove that the posterior distribution is consistent in sup-norm on the whole support of f_0 when it has compact support. We also derive concentration rates for the posterior of the density taken at a fixed point and for the sup-norm on subsets of $[0, L]$ for $L < \infty$. We derive an upper bound for the concentration rate of $f(x)$ for $x \in (0, L)$ but only get suboptimal rates using a testing approach as in Giné and Nickl (2010). It is to be noted that for this problem the modulus of continuity for the point-wise and Hellinger losses defined for $f_0 \in \mathcal{F}$ and $x \in (0, L)$ by

$$m(\epsilon) := \sup\{|f(x) - f_0(x)| : f \in \mathcal{F},\ h(f, f_0) \leq \epsilon\}$$

is of the order $\epsilon^{2/3}$ (see Donoho and Liu, 1991). Given the discussion in Hoffmann et al. (2013), it is to be expected that the usual approach of Ghosal et al. (2000) based on tests will lead to suboptimal concentration rates. We now introduce some notations used throughout the paper.

Notations For $0 < L \leq \infty$ define the set $\mathcal{F}_L$ by

$$\mathcal{F}_L = \left\{ f \text{ s.t. } 0 \leq f < \infty,\ f \searrow \int_0^L f = 1 \right\},$$

We also define $\mathfrak{S}_k$ be the k-simplex that is the set $\{(s_1, \ldots, s_k) \in [0,1]^k, \sum_{i=1}^k s_i = 1\}$. Let $KL(p_1, p_2)$ be the Kullback-Leibler deviation between the densities p_1 and p_2 with respect to some measure λ

$$KL(p_1, p_2) = \int \log\left(\frac{p_1}{p_2}\right) p_1 d\lambda.$$

We also define the Hellinger distance $h(p_1, p_2)$ between p_1 and p_2 as

$$h^2(p_1, p_2) = \frac{1}{2} \int (\sqrt{p_1} - \sqrt{p_2})^2 d\lambda.$$

We will say that $\Xi^n = o_{p_0}(1)$ if $\Xi^n \to 0$ under P_0. Finally we will denote by f' the derivative of f.

Construction of a prior distribution on $\mathcal{F}_L$ Using the mixture representation of monotone non-increasing densities (1) we construct non-parametric priors on the set $\mathcal{F}_L$ by considering a prior on the mixing distribution P. Let $\mathcal{P}$

be the set of probability measures on $[0, L]$. Thus, we fall in the well known set up of non-parametric mixture priors models. We consider two types of prior on the set $\mathcal{P}$.

Type 1: Dirichlet Process prior $P \sim DP(A, \alpha)$ where A is a positive constant and α a probability density on $[0, L]$.

Type 2: Finite mixture $P = \sum_{j=1}^{K} p_j \delta_{x_j}$ with K a non zero integer and δ_x the dirac function on x. We choose a prior distribution Q on K and given K, define distributions $\pi_{x,K}$ on $(x_1, \ldots, x_K) \in [0, L]^K$ and $\pi_{p,K}$ on $(p_1, \ldots, p_K) \in \mathfrak{S}_K$.

For $\mathbf{X^n} = (X_1, \ldots, X_n)$, a sample of n independent and identically distributed random variables with common probability distribution function f in $\mathcal{F}_L$ with respect to the Lebesgue measure, we denote $\Pi(\cdot|\mathbf{X^n})$ the posterior probability measure associated with the prior Π.

The paper is organised as follow: the main results are given in Section 2, where conditions on the priors are discussed. The proofs are presented in Section 3.

2. Main results

Concentration rates of the posterior distributions have been well studied in the literature and some general results link the rate to the prior distribution (see Ghosal et al. (2000)). However, in our setting, the Kullback Leibler property is not satisfied in its usual form and thus the standard Theorems do not hold. In fact an interesting feature of mixture distributions whose kernels have varying support is that the prior mass of the sets $\{f, KL(f_0, f) = +\infty\}$ is 1 for most $f_0 \in \mathcal{F}_L$ given that f and f_0 will have different support. One could prevent this by imposing that the support of the mixing distribution is wider than the support of f_0; however, this could lead to a deterioration of the concentration rate. Here, we use a modified version of the results of Ghosal et al. (2000) considering truncated versions of the density f. This idea has been considered in Khazaei et al. (2010) in a similar setting. We impose some conditions on the prior under which the posterior distribution concentrates at the minimax rate up to a $\log(n)$ term.

Conditions on the prior

C1 condition on α Let α be a probability density on $\mathbb{R}^+$ such that for all $\theta \in (0, L)$, $\alpha(\theta) > 0$. Consider the following conditions on α

- for $0 < t_2 \leq t_1$ and θ small enough
$$\theta^{t_1} \lesssim \alpha(\theta) \lesssim \theta^{t_2} \tag{2a}$$
- for $1 < a_1 \leq a_2$ and θ small enough
$$e^{-a_1/\theta} \lesssim \alpha(\theta) \lesssim e^{-a_2/\theta} \tag{2b}$$
- for $1 < b_1 \leq b_2$ and θ small enough
$$e^{-b_1/\theta} \lesssim \alpha(L - \theta) \lesssim e^{-b_2/\theta}. \tag{2c}$$

C2 condition for Type I prior For $P \sim DP(\alpha, M)$ with α satisfying C1.
C3 condition for the Type II prior The following conditions holds

- For some positive constants $C_1, C_2, a_1, \ldots, a_k, c$

$$e^{-C_1 K \log(K)} \geq Q(K) \geq e^{-C_2 K \log(K)} \tag{3}$$

$$\pi_{p,k}(p_1, \ldots, p_K) \geq K^{-K} c^K p_1^{a_1} \ldots p_K^{a_K} \tag{4}$$

- $\pi_{x,K}$ is the distribution of K independent and identically distributed random variables sampled from α.

C4 condition for densities on $\mathbb{R}^+$ If $f_0 \in \mathcal{F}_\infty$ then for some fixed positive constants β and τ we have for x large enough

$$f_0(x) \leq e^{-\beta x^\tau}. \tag{5}$$

2.1. Posterior concentration rate for the L_1 and Hellinger metric

The following Theorems gives the posterior concentration rate for the L_1 and Hellinger metrics for monotone non-increasing densities on $[0, L]$ with $L < \infty$ and $L = \infty$. For both Theorems the proofs are postponed to section 3.

Theorem 1. *Let $\mathbf{X^n} = (X_1, \ldots, X_n)$ be an independent and identically distributed sample with a common probability distribution function f_0 such that $f_0 \in \mathcal{F}_L$ with $0 < L < \infty$. Let Π be either a Type I or Type II prior satisfying **C2** or **C3** respectively with α satisfying (2a). If $d(\cdot, \cdot)$ is either the L^1 or Hellinger distance, then there exists a positive constant C such that*

$$\Pi\left(f, d(f, f_0) \geq C\left(\frac{n}{\log(n)}\right)^{-1/3} |\mathbf{X^n}\right) \to 0, \qquad P_0 \text{ a.e.} \tag{6}$$

when n goes to infinity, where C depends on f_0 only through L and an upper bound on $f_0(0)$. Furthermore, if for $\delta > 0$, $\sup_{[0,\delta]} |f_0'(x)| < \infty$ and α satisfies (2b), or $\sup_{[L,L-\delta]} |f_0'(x)| < \infty$ and α satisfies (2c), then (6) still holds.

Conditions C1 and C2 are roughly the same as in Khazaei et al. (2010). Theorem 1 is thus an extension of their results to concentration rates. We also extend their results to mixture priors satisfying (2b) or (2c) under some additional conditions on f_0. This will prove useful for the estimation of $f(0)$ and $f(L)$. Under condition C3 on the tail of the true density, i.e. we require exponential tails, we get the posterior concentration rate for a density with support on $\mathbb{R}^+$.

Theorem 2. *Let $\mathbf{X^n} = (X_1, \ldots, X_n)$ be an independent and identically distributed sample with a common probability distribution density f_0, such that $f_0 \in \mathcal{F}_\infty$ and f_0 satisfy **C3**. Let Π be either a Type I or Type II prior satisfying **C2** or **C3** respectively with α satisfying (2a). Then, for some positive constant C, we have for $d(\cdot, \cdot)$ either the L_1 or Hellinger metric*

$$\Pi\left(d(f_P, f_0) \geq C\left(n/\log(n)\right)^{-1/3} \log(n)^{1/\tau} |\mathbf{X^n}\right) \to 0, \; P_0 \text{ a.e.} \tag{7}$$

when n goes to infinity. Similarly, if for $\delta > 0$, $\sup_{[0,\delta]} |f_0'(x)| < \infty$ and α satisfies (2b), (7) *still holds.*

Note that considering monotone non-increasing densities on $\mathbb{R}^+$ deteriorates the upper bound on the posterior concentration rate by a factor $\log(n)^{1/\tau}$. It is not clear whether it could be sharpened or not. For instance, in the frequentist literature, Reynaud-Bouret et al. (2011) observe a slower convergence rate when considering an infinite support for densities without any other conditions. In a Bayesian setting, a similar log term appears in Kruijer et al. (2010) when considering densities with non compact support. However this deterioration of the concentration rate does not have a great influence on the asymptotic behaviour of the posterior. Note also that the tail conditions are mild since τ can be taken as small as needed, and thus the considered densities can have almost polynomial tails.

The above results on the posterior concentration rate in terms of the L_1 or Hellinger metric are new to the best of our knowledge, but not surprising. The specificity of these results lies in the fact that the usual approach based on the approach of Ghosal et al. (2000) needs to bound the prior mass of Kullback Leibler neighbourhoods of the true density which cannot be done here as explained in section 1.

2.2. Consistency and posterior concentration rate for the point-wise and supremum loss

The following results consider the point-wise loss function for which only a few exist in the Bayesian non-parametric literature; see for instance the paper of Giné and Nickl (2010). The following Theorem proves consistency of the posterior distribution for all points in the interior of the support.

Theorem 3. *Let x be in $(0, L)$ with with $0 < L \leq \infty$ but $x < \infty$. Let $f_0 \in \mathcal{F}_L$ such that f_0' exists near x and $f_0'(x) < 0$. Let X_i, $i = 1, \ldots, n$ and Π be either a Type I or Type II prior satisfying **C2** or **C3** respectively with α satisfying **C1** with either* (2a), (2b) *or* (2c)*. Then, for all x in $(0, L)$ with $x < \infty$, and $\epsilon > 0$*

$$\Pi\big(|f_P(x) - f_0(x)| > \epsilon|\mathbf{X^n}\big) \to 0. \tag{8}$$

Consider the posterior median $\hat{f}_n^\pi(x) = \inf\{t, \Pi[f_P(x) \leq t|\mathbf{X^n}] > 1/2\}$, it follows that

$$P_0\big(|\hat{f}_n^\pi(x) - f_0(x)| > \epsilon|\mathbf{X^n}\big) \to 0. \tag{9}$$

We thus have a pointwise consistency of the posterior distribution of $f_0(x)$ for every x in the interior of the support of f_0. The maximum likelihood is not consistent at the boundaries of the support as pointed out in Sun and Woodroofe (1996). In particular it is not consistent at 0 and when $L < \infty$, it is not consistent at L. It is known that integrating the parameter as done in Bayesian approaches induces a penalisation. This is particularly useful in testing or model choice problems but can also be effective in estimation problems; see

for instance Rousseau and Mengersen (2011). Here we require that the base measure puts exponentially small mass at the boundaries. This induces enough penalization to achieve consistency of the posterior distribution of $f(0)$ and $f(L)$. The following Theorem gives consistency of the posterior distribution of f at every point on the support of f_0 including the boundaries.

Theorem 4. *Let x be in $[0, L]$ with with $0 < L \leq \infty$ but $x < \infty$. Let $f_0 \in \mathcal{F}_L$ such that f_0' exists at x and $f_0'(x) < 0$. Let X_i, $i = 1, \ldots, n$ and Π be either a Type I or Type II prior satisfying **C2** or **C3** with α satisfying condition (2b) if $x = 0$ or (2c) if $x = L$. Then, for all x in $[0, L]$ with $x < \infty$, and $\epsilon > 0$*

$$\Pi\big(|f_P(x) - f_0(x)| > \epsilon|\mathbf{X^n}\big) \to 0. \tag{10}$$

Consider the posterior median $\hat{f}_n^\pi(x) = \inf\{t, \Pi[f_P(x) \leq t|\mathbf{X^n}] > 1/2\}$, it follows that

$$P_0\big(|\hat{f}_n^\pi(x) - f_0(x)| > \epsilon|\mathbf{X^n}\big) \to 0. \tag{11}$$

The problem of estimating $f_0(0)$ under monotonicity constraints is another example of the effectiveness of penalization induced by integration on the parameters. Although we do not have a proof for inconsistency of the posterior of $f(0)$ or $f(L)$ when α satisfies (2a), we believe that due to the similarly to the maximum likelihood estimator, the posterior distribution is in this case not consistent.

The following Theorem gives an upper bound on the concentration rate of the posterior distribution under the point-wise loss.

Theorem 5. *Let f_0 be in $\mathcal{F}_L$ with $0 < L \leq \infty$ and Π be either a Type I or Type II prior satisfying **C2** or **C3** respectively with α satisfying **C1**, and let x be in $(0, L)$ such that f' exists in a neighbourhood of x and $f'(x) < 0$, then for C a positive constant*

$$\Pi\left(|f_P(x) - f_0(x)| > C\left(\frac{n}{\log(n)}\right)^{-2/9}|\mathbf{X^n}\right) \to 0. \tag{12}$$

when n goes to infinity.

Here the concentration rate is suboptimal. It is however the best rate that one can obtain using the usual approach by testing (see Hoffmann et al., 2013). Proving that the posterior concentrates at the rate $n^{-1/3}$ up to some power of $\log(n)$ would require some more refined control of the posterior distribution close to Bernstein-von Mises types of results, (see Castillo, 2013), which in the case of mixture models is very difficult to handle and beyond the scope of this paper.

Next, we derive from Theorem 4 the consistency of the posterior distribution for the sup-norm. This is particularly useful when considering confidence bands, as pointed out in Giné and Nickl (2010). Under similar assumptions as in Durot et al. (2012), we get the consistency of the posterior distribution for the sup-norm. Note that contrary to the setting in Durot et al. (2012), we do not restrict to sub-intervals of the support of the density. This is mainly due to the fact that the Bayesian approache is consistent at the boundaries of the support of f_0.

Theorem 6. *Let $f_0 \in \mathcal{F}_L$ with $0 < L < \infty$ be such that f_0' exists and $||f_0'||_\infty < \infty$ and for all $x \in [0, L]$, $f_0'(x) < 0$. Let also the prior Π be either a Type I or Type II prior satisfying **C2** or **C3** with α satisfying conditions (2b) and (2c) respectively. Then*

$$\Pi(\sup_{x\in[0,L]} |f_P(x) - f_0(x)| > \epsilon | X_n) \to 0. \tag{13}$$

Similar results as in Theorem 5 also hold for the concentration rate of the posterior distribution for the supremum over all subsets of the form (a, b) with $0 < a < b < L$ with the same rate.

3. Proofs

In this section we prove Theorems 1 to 6 given in Section 2. To prove Theorems 3–6, we need to construct tests that are adapted to the point-wise or supremum loss. The usual approach based on Le Cam (1986) cannot be applied in this case. We thus construct test based on the Maximum Likelihood Estimator.

3.1. Proof of Theorems 1 and 2

The proofs of Theorems 1 and 2 follow the general ideas of Ghosal et al. (2000). We first focus on a density on $\mathcal{F}_L$ with $L < \infty$ and extend these results to a monotone non-increasing density with support $\mathbb{R}^+$ that satisfy C4. We extended the approach used in Khazaei et al. (2010) to the concentration rate framework. More precisely, the proofs rely on the following Theorem which is adapted from Rivoirard and Rousseau (2012). To tackle the fact that the usual Kullback Leibler property is not satisfied in its usual sense, we consider truncated versions of the densities

$$f_n(\cdot) = \frac{f(\cdot)\mathbb{I}_{[0,\theta_n]}(\cdot)}{F(\theta_n)},\ f_{0,n}(\cdot) = \frac{f_0(\cdot)\mathbb{I}_{[0,\theta_n]}(\cdot)}{F_0(\theta_n)} \tag{14}$$

where θ_n is defined as

$$\theta_n = \inf\Big\{x, 1 - F_0(x) < \frac{\epsilon_n}{2n}\Big\}.$$

We then define the counterpart of the Kullback Leibler neighbourhoods

$$S_n(\epsilon_n, \theta_n) = \Bigg\{ f, KL(f_n, f_{0,n}) \le \epsilon_n^2, \\ \int f_{0,n}(x) \left(\log\left(\frac{f(x)}{f_0(x)}\right)\right)^2 dx \le \epsilon_n^2, \int_0^{\theta_n} f(x)dx \gtrsim 1 - \epsilon_n^2 \Bigg\}. \tag{15}$$

Theorem 7. *Let f_0 be the true density and let Π be a prior on $\mathcal{F}$ satisfying the following conditions: there exists a sequence (ϵ_n) such that $\epsilon_n \to 0$ and $n\epsilon_n^2 \to \infty$ and a constant $c > 0$, such that for any n there exist $\mathcal{F}_n \subset \mathcal{F}$ satisfying*

$$\Pi(\mathcal{F}_n^c) = o(\exp(-(c + 2)n\epsilon_n^2)).$$

For any $j \in \mathbb{N}$, $j > 0$, let $\mathcal{F}_{n,j} = \{f \in \mathcal{F}_n, j\epsilon_n < d(f, f_0) \leq (j+1)\epsilon_n\}$ and $N_{n,j}$ the Hellinger (or L_1) metric entropy of $\mathcal{F}_{n,j}$. There exists a $J_{0,n}$ such that for all $j \geq J_{0,n}$

$$N_{n,j} \leq (K-1)n\epsilon_n^2 j^2,$$

where K is an absolute constant.

Let $S_n(\epsilon_n, \theta_n)$ be defined as in (15) and let Π be such that

$$\Pi(S_n(\epsilon_n, \theta_n)) \geq \exp(-cn\epsilon_n^2). \tag{16}$$

We have:

$$\Pi(f : d(f_0, f) \leq J_{0,n}\epsilon_n | \mathbf{X^n}) = 1 + o_P(1).$$

The proof of this Theorem is postponed to Appendix B. We will thus prove that the conditions of Theorem 7 are satisfied in our case. Let f_0 be in $\mathcal{F}_L$. The following lemma states that (16) is satisfied.

Lemma 8. *Let Π be either a Type 1 or Type 2 prior on $\mathcal{F}_L$ as in Theorem 1 and let $S_n(\epsilon_n, \theta_n)$ be a set as in (15), then*

$$\Pi(S_n(\epsilon_n, \theta_n)) \gtrsim \exp\left\{C_1 \epsilon_n^{-1} \log(\epsilon_n)\right\}. \tag{17}$$

This lemma is proved in appendix A. The ϵ metric entropy of the set of bounded monotone non-increasing densities has been shown to be less than ϵ^{-1}, up to a constant (see Groeneboom (1986) or van der Vaart and Wellner (1996) for instance). As the prior puts mass on $\mathcal{F}_L$, on which $f(0)$ is not uniformly bounded, we consider an increasing sequence of sieves

$$\mathcal{F}_n = \left\{f \in \mathcal{F}_L, f(0) \leq M_n\right\}. \tag{18}$$

where $M_n = \exp\{cn^{1/3}\log(n)^{2/3}(t_2+1)^{-1}\}$ with t_2 as in the conditions C1. The following lemma shows that $\mathcal{F}_n$ covers most of the support of Π as n increases.

Lemma 9. *Let $\mathcal{F}_n$ be defined by (18) and Π be either a Type 1 or Type 2 as in Theorem 1, then*

$$\Pi\big(\mathcal{F}_n^c\big) \lesssim e^{-cn^{1/3}\log(n)^{2/3}}.$$

Here again, the proof is postponed to appendix A. We now get an upper bound for the ϵ-metric entropy of the set $\mathcal{F}_n$. Recall that in Groeneboom (1985) it is proved that the L_1 metric entropy of monotone non-increasing densities on $[0, 1]$ bounded by M can be bounded from above by $C_0 \log(M)\epsilon_n^{-1}$. We cannot apply this result directly for the sets $\mathcal{F}_n$ as it would give a suboptimal control of the entropy to construct tests in a similar way as in Ghosal et al. (2000). In fact, the upper bound on the entropy of $\mathcal{F}_n$ is of the order of $e^{n\epsilon_n}$, while the usual conditions of Ghosal et al. (2000) requires an upper bound of the order $e^{n\epsilon_n^2}$. However as stated in Theorem 7 it is enough to bound the ϵ-metric entropy of the sets

$$\mathcal{F}_{n,j} = \{f \in \mathcal{F}_n, j\epsilon_n \leq d(f, f_0) \leq (j+1)\epsilon_n\},$$

for $j \in \mathbb{N}^*$. We can easily adapt the results of Groeneboom (1985) to positive monotone non-increasing functions on any interval $[a, b]$ and get the following lemma.

Lemma 10. *Let $\tilde{\mathcal{F}}$ be the set of positive monotone non-increasing functions on $[a, b]$ such that for all f in $\tilde{\mathcal{F}}$,$\int_a^b f \leq M_2$ and $f \leq M$, then*

$$N(\epsilon, \tilde{\mathcal{F}}, d) \lesssim \epsilon^{-1} \log(M+1)\Big((b-a) + 3M_2\Big).$$

The proof of this lemma is straightforward given the results of Groeneboom (1985) and is thus omitted. Let $x_{n,j} \in [0, L]$ such that $\epsilon_n/2 \leq x_{n,j} \leq \epsilon_n$. We denote for all f in $\mathcal{F}_{n,j}$ $f_{1,j} = f\mathbb{I}_{[0,x_{n,j})}$ and $f_{2,j} = f\mathbb{I}_{[x_{n,j},L]}$. Since for all f in $\mathcal{F}_{n,j}$ we have $\int_0^1 |f(x) - f_0(x)|dx \leq (j+1)\epsilon_n$ then

$$\int_0^{x_{n,j}} f(x)dx - \int_0^{x_{n,j}} f_0(x)dx \leq (j+1)\epsilon_n,$$

which implies that

$$x_{n,j} f(x_{n,j}) \leq x_{n,j} f_0(0) + (j+1)\epsilon_n,$$

which in turn gives

$$f(x_{n,j}) \leq f_0(0) + 2(j+1).$$

Recall that for all $f \in \mathcal{F}_n$ we have $f(0) \leq M_n$. Using lemma 10, we construct an $\epsilon_n/2$-net for the set $\mathcal{F}^1_{n,j} = \{f_{1,j}, f \in \mathcal{F}_{n,j}\}$ with N_1 points, and

$$\log(N_1) \lesssim \epsilon_n^{-1} \log(M_n + 1)\epsilon_n(j+2),$$

and thus deduce

$$\log(N_1) \leq C' n\epsilon_n^2 j^2 \tag{19}$$

Similarly, given that $f(x_{n,j}) \leq M + 2(j+1)$ we get an $\epsilon_n/2$-net for the set $\mathcal{F}^2_{n,j} = \{f_{2,j}, f \in \mathcal{F}_{n,j}\}$ with N_2 points and

$$\log(N_2) \leq \tilde{C}' n\epsilon_n^2 j^2. \tag{20}$$

This provides a ϵ_n-net for $\mathcal{F}_{n,j}$ with less than $N_1 \times N_2$ points. Given (19) and (20) the L_1 metric entropy of the sets $\mathcal{F}_{n,j}$ satisfies

$$\log(N(\mathcal{F}_{n,j}, \epsilon_n, L_1)) \lesssim n\epsilon_n^2 j^2. \tag{21}$$

The conditions of Theorem 7 are thus satisfied which concludes the proof of Theorem 1.

Extention to $\mathbb{R}^+$ Given that $f_0(x) \lesssim e^{-\beta x^\tau}$ when x goes to infinity, if θ_n is such that $\theta_n = \inf\{x, 1 - F_0(x) < \epsilon_n/(2n)\}$ then $\theta_n \lesssim (\log(n))^{1/\tau}$. Using similar arguments as before, Lemma 8 still holds under the exponential tail assumption. We now get an upper bound for the ϵ-metric entropy of $\mathcal{F}_{n,j}$. Here again, we split $\mathcal{F}_{n,j}$ into two parts. The construction of an $\epsilon_n/2$-net for $\mathcal{F}^1_{n,j}$ does not change and therefore (19) holds. Finally, let $\tilde{\mathcal{F}}^2_{n,j} = \{f \in \mathcal{F}^2_{n,j}, \forall x > \theta_n, f(x) = 0\}$. Given Lemma 10, we get for $c_1 > 0$ large enough an $\epsilon_n/(2c_1(j+1))$-net for $\tilde{\mathcal{F}}^2_{n,j}$ by considering $f^\star$ the restriction of f to $[x_{n,j}, \theta_n]$. We have

$$d(f, f^\star) \leq c_2(j+1)\epsilon_n,$$

where $d(\cdot,\cdot)$ is either the L_1 or Hellinger distance. Hence, for $c_1 > c_2$ an $\epsilon/2$-net for $\mathcal{F}^2_{n,j}$ with at most $e^{c_3 n\epsilon_n^2 j^2}$ points and thus

$$\log\left(N(\mathcal{F}^2_{n,j}, \epsilon_n, d)\right) \leq \tilde{C}'' n\epsilon_n^2 j^2.$$

We conclude using the same arguments as in the preceding section, and thus Theorem 2 is proved.

3.2. Proof of Theorems 3 and 5

To prove Theorems 3 and 5, we need to construct tests for all $x \in (0, L)$ of f_0 versus $|f_P(x) - f_0(x)| \geq \epsilon_n^{2/3}$ as the approach used in Ghosal et al. (2000) is not suited for the point-wise loss. As we have $\Pi(||f_P - f_0||_1 > \epsilon_n|\mathbf{X^n}) = o_{P_0}(1)$ we can consider functions f_P such that $||f_P - f_0||_1 \leq \epsilon_n$. We construct tests Φ_n such that

$$E_0^n(\Phi) = o(1), \quad \sup_{f, |f(x)-f_0(x)|>\epsilon_n} \mathrm{E}_f^n(1-\Phi) \leq e^{-Cn\epsilon_n^2}.$$

Denote $A_\epsilon^x := \{f, |f(x) - f_0(x)| > \epsilon\}$ that can be split into $A_\epsilon^{x,+} = \{f, f(x) - f_0(x) > \epsilon\}$ and $A_\epsilon^{x,-} = \{f, f(x) - f_0(x) < -\epsilon\}$ and denote $e_n = e_0\epsilon_n^{2/3}$ and $h_n = h_0 e_n$. Consider the tests

$$\begin{aligned}
\phi_n^+ &= \mathbb{I}\left\{n^{-1}\sum_{i=1}^n \mathbb{I}_{[x-h_n,x]}(X_i) - \int_{x-h_n}^x f_0(t)dt > c_n\right\} \\
\phi_n^- &= \mathbb{I}\left\{n^{-1}\sum_{i=1}^n \mathbb{I}_{[x,x+h_n]}(X_i) - \int_x^{x+h_n} f_0(t)dt < -c_n\right\}
\end{aligned}$$

We immediately get $\mathrm{E}_0^n(\max(\phi_n^+, \phi_n^-) = o(1)$. Note that if $f_P(x) > f_0(x) + e_n$ then

$$\begin{aligned}
\int_{x-h_n}^x f_P(t) - f_0(t)dt &\geq h_n(f_P(x) - f_0(x)) - \int_{x-h}^x f_0(t) - f_0(x)dt \\
&\geq h_n e_n - C_0 h^2
\end{aligned}$$

for some $C_0 > 0$ that only depends on f_0. Similarly if $f_P(x) < f_0(x) - e_n$ then for all $h > 0$

$$\int_x^{x+h} f_P(t) - f_0(t)dt \leq -he_n + C_0 h^2$$

We thus deduce for f_P such that $f_P(x) - f_0(x) > e_n$, we have

$$\begin{aligned} P_f(1-\phi_n^+) &\leq P_f\left(n^{-1}\sum_{i=1}^n \mathbb{I}_{[x-h_n,x]}(X_i) - \int_{x-h_n}^x f_P(t)dt \leq -h_n e_n + C_0 h^2 + c_n\right) \\ &\leq P_f\left(n^{-1}\sum_{i=1}^n \mathbb{I}_{[x-h,x]}(X_i) - \int_{x-h}^x f_P(t)dt \leq -h_0 e_n^2/2\right), \end{aligned}$$

if $c_n \leq e_n^2$ and $h_0 \leq 1/C_0$. Now note that for f_P such that $||f_P - f_0||_1 \leq \epsilon_n$

$$\begin{aligned} \int_{x-h_n}^x f_P &\geq -\int_0^\infty |f - f_0| + \int_{x-h_n}^x f_0 \\ &\geq -\epsilon_n + \int_{x-h_n}^x f_0 \\ &\geq -e_n + h_n f_0(x) \geq h_n f_0(x)/2. \end{aligned}$$

Moreover,

$$\int_{x-h_n}^x f_P \leq e_n + h_n f_0(x-h_n) \leq 2h_n f_0(x),$$

for n large enough and h small enough. We conclude that

$$\mathrm{Var}_{f_P}^n\left(n^{-1}\sum_{i=1}^n \mathbb{I}_{[x-h,x]}(X_i)\right) \leq 2hf_0(x).$$

Using Bernstein's inequality (e.g. van der Vaart and Wellner (1996) Lemma 2.2.9, p. 102) we get

$$P_f(1-\phi^+) \leq 2e^{-nh_n e_n^2/(2+e_n/3)}.$$

Similarly, we have

$$P_f(1-\phi_n^-) \leq 2e^{-nh_n e_n^2/(2+e_n/3)}.$$

Taking $\Phi_n = \max(\phi_n^+, \phi_n^-)$ we deduce

$$\begin{aligned} P_0(\Phi_n) &= o(1) \\ \sup_{f\in A_{e_n}^x} P_f(1-\Phi_n) &\leq e^{-Ch_0 e_n^3} \end{aligned}$$

We have

$$P_0(\Phi_n) = o(1)$$

$$\sup_{f \in A^x_{e_n}} P_f(1-\Phi_n) \leq e^{-Cne_0\epsilon_n^2}$$

Similarly to the proof of Theorem 7, following Khazaei et al. (2010), we get an exponentially small lower bound for D_n. More precisely, we get that

$$D_n \geq 2e^{-(c+2)n\epsilon_n^2}$$

with probability that goes to 1. Note that

$$\begin{aligned}\mathbf{E}_0^n\left(\frac{N_n}{D_n}\right) \leq \mathbf{E}_0^n(\Phi_n^x) + P_0^n(D_n \leq e^{-(c+2)n\epsilon_n^2}) \\ + \mathbf{E}_0^n(\Pi[\mathcal{F}_n^c|\mathbf{X^n}]) + e^{(c+2)n\epsilon_n^2}\int_{A_\epsilon \cap \mathcal{F}_n} \mathbf{E}_f^n(1-\Phi_n^x)d\Pi(f).\end{aligned} \tag{22}$$

Given the preceding results, we have

$$\mathbf{E}_0^n\left(\frac{N_n}{D_n}\right) \leq o(1) + e^{(c+2)n\epsilon_n^2}\sup_f \mathbf{E}_f^n(1-\Phi_n^x)$$

which ends the proof choosing e_0 large enough.

Consistency of a Bayesian estimator We consider in this section $\hat{f}_n^\pi(t)$, the Bayesian estimator associated with the absolute error loss, defined as the median of the posterior distribution. Consistency of the posterior mean, which is the most common Bayesian estimator is however not proved here but could nevertheless be an interesting result.

We first define $\hat{f}_n^\pi(t)$ such that

$$\hat{f}_n^\pi(t) = \inf\{x, \Pi[f_P(t) \leq x|\mathbf{X^n}] > 1/2\}. \tag{23}$$

In order to get consistency in probability we note that if $\hat{f}_n^\pi(t) - f_0(t) > \epsilon$ then

$$\Pi(f_P(t) > f_0(t) + \epsilon|\mathbf{X^n}) > 1/2.$$

And if $\hat{f}_n^\pi(t) - f_0(t) < -\epsilon$ then

$$\Pi(f_P(t) < f_0(t) - \epsilon|\mathbf{X^n}) > 1/2.$$

We deduce, using Markov's inequality and Theorem 3

$$\begin{aligned} P_0^n(\hat{f}_n^\pi(t) - f_0(t) > \epsilon) &\leq P_0^n(\Pi(f_P(t) > f_0(t) + \epsilon|\mathbf{X^n}) > 1/2) \\ &\leq 2\mathbf{E}_0^n(\Pi(f_P(t) > f_0(t) + \epsilon|\mathbf{X^n}) > 1/2) \\ &\leq o(1), \end{aligned}$$

and similarly

$$P_0^n(\hat{f}_n^\pi(t) - f_0(t) < -\epsilon) \leq o(1).$$

Thus, we have $P_0^n(|\hat{f}_n^\pi(t) - f_0(t)| > \epsilon) \to 0$ which gives the consistency in probability of $\hat{f}_n^\pi(t)$.

3.3. Proof of Theorem 4

The previous proof holds for all $x \in (0, L)$. We now need to prove the consistency of the posterior for $x = 0$ and $x = L$, when the prior satisfies conditions (2b) or (2c). We first consider the case $x = 0$, the case $x = L$ can be deduce with symmetric arguments.

As before, consider the set A_ϵ^0 and split it in $A_\epsilon^{0,+}$ and $A_\epsilon^{0,-}$. Note that using the same test ϕ_n^- as before we easily get

$$\Pi(A_\epsilon^{0,-}|\mathbf{X^n}) = o_{P_0}(1).$$

We now consider $f_P \in A_\epsilon^{0,+}$. As before we can restrict ourselves to functions f_P such that $||f_P - f_0||_1 \leq \epsilon_n$. We thus have for $h = 2\epsilon_n/\epsilon$

$$\begin{aligned} f_P(0) - f_0(0) &\leq f_P(0) - f_P(h) + h^{-1} \int |f_0(t) - f_P(t)| dt \\ &\leq f_P(0) - f_P(h) + h^{-1}\epsilon_n \\ &= f_P(0) - f_P(h) + \epsilon/2. \end{aligned}$$

We now prove that the prior mass of the event $\{f_P(0) - f_P(h) > \epsilon/2\}$ is less that $e^{-(c+2)n\epsilon_n^2}$. Using Markov inequality we get

$$\Pi(f_P(0) - f_P(h) > \epsilon/2) \leq 2\epsilon^{-1} \int_0^h \frac{1}{\theta}\alpha(\theta) d\theta \leq e^{-a_2/h} \lesssim e^{-a_2 n\epsilon_n^2 \log(n)}.$$

Using the same control for D_n as in the proof of Theorem 7, and applying the usual method of Ghosal et al. (2000), we get the desired result.

3.4. Proof of Theorem 6

In this section, we prove that the posterior distribution is consistent in sup-norm. Here again, the main difficulty is to construct tests that are adapted to the considered loss. More precisely we construct a test Φ such that

$$E_0^n(\Phi) = o(1), \qquad \sup_{f, \sup_{[0,L]}|f - f_0| > \epsilon_n} \mathrm{E}_f^n(1 - \Phi) \leq e^{-Cn\epsilon_n^2}.$$

To do so, we consider a combination of the tests considered in the previous section noting that if the posterior distribution is consistent at the points of a sufficiently refined partition of $[0, L]$ then it is consistent for the sup-norm. Here again, we will only consider the case $L = 1$ without loss of generality. We first denote

$$B_\epsilon = \left\{ f, \sup_{[0,L]} \{|f(x) - f_0(x)| > \epsilon \right\}.$$

Let C_0' be a positive constant such that $||f_0'||_\infty \leq C_0'$ and let $(x_i)_i$ be the separation points of a $\epsilon/(8C_0')$ regular partition of $[0, 1]$ and $p = \mathrm{Card}\{(x_i)_i\}$.

Note that

$$B_\epsilon = \bigcup_{i=1}^{p} \{f, \sup_{[x_i, x_{i+1}]} \{|f(x) - f_0(x)| > \epsilon\}.$$

Recall that $A_\epsilon^x = \{f, |f(x) - f_0(x)| > \epsilon\}$. We consider the set $B_\epsilon \bigcap_{i=1}^p (A_{\epsilon/8}^{x_i})^c$. Given Theorem 3, we have that

$$\mathrm{E}_0^n \left(\Pi \left(\bigcup_{i=1}^{p} (A_{\epsilon/5}^{x_i}) \Big| \mathbf{X^n} \right) \right) = o(1).$$

If $f \in B_\epsilon$ we have for all $x \in [x_i, x_{i+1}]$,

$$|f(x) - f_0(x)| \leq |f(x) - f(x_i)| + |f(x_i) - f_0(x_i)| + |f_0(x_i) - f_0(x)|.$$

Given that f is monotone non-increasing, and given the hypotheses on f_0 we have

$$\begin{aligned} |f(x) - f(x_i)| &\leq |f(x_{i+1}) - f(x_i)| \\ &\leq |f(x_{i+1}) - f_0(x_{i+1})| + |f_0(x_{i+1}) - f_0(x_i)| + |f_0(x_i) - f(x_i)| \\ &\leq 3\epsilon/5, \end{aligned}$$

and for the same reasons

$$|f(x_i) - f_0(x_i)| + |f_0(x_i) - f_0(x)| \leq 2\epsilon/5,$$

which leads to

$$|f(x) - f_0(x)| \leq \epsilon,$$

and thus, taking the supremum over x, we get

$$\sup_{x \in [x_i, x_{i+1}]} |f(x) - f_0(x)| \leq \epsilon.$$

We then deduce

$$\Pi(B_\epsilon | \mathbf{X^n}) \leq \Pi \left(B_\epsilon \bigcap \left\{ \bigcap_{i=1}^{p} (A_{\epsilon/5}^{x_i})^c \right\} \right) + \Pi \left(\bigcup_{i=1}^{p} (A_{\epsilon/5}^{x_i}) \right) = o_{P_0}(1),$$

which gives the consistency of the posterior distribution in sup-norm.

4. Discussion

In this paper, we obtain an upper bound for the concentration rate of the posterior distribution under monotonicity constraints. This is of interest as in this model, the standard approach based on the seminal paper of Ghosal et al. (2000) cannot be applied directly. We prove that the concentration rate of the posterior is (up to a $\log(n)$ factor) the minimax estimation rate $(n/\log(n))^{-1/3}$ for standard losses such as L_1 or Hellinger.

We also prove that the posterior distribution is consistent for the point-wise loss at any point of the support and for the sup-norm loss. Studying asymptotic properties for these losses is difficult in general as the usual approach are well suited for losses that are related to the Hellinger metric. Obtaining more refined

results on the asymptotic behaviour of the posterior distribution will require refined control of the likelihood which in the case of non-parametric mixture models is a difficult task.

Aknowledgements

We are grateful to Judith Rousseau for helpful discussion and her assistance throughout. The author would also like to thank the referee and associate editor for their useful remarks which help to improve a former version of the paper. This work is partially funded by ANR Bandhits and Labex ECODEC.

Appendix A: Technical Lemmas

A.1. Proof of Lemma 8

To prove Lemma 8, we first construct stepwise constant functions such that these approximations are in the truncated Kullback Leibler neighbourhood of f_0. We then construct a set $\mathcal{N}$ included in $S_n(\epsilon_n, \theta_n)$ based on the considered piecewise constant approximation such that for Π a Type I or Type II prior $\Pi(\mathcal{N}) \geq e^{-Cn\epsilon_n^2}$.

We first construct a piecewise constant approximation of f_0 which is based on a sequential subdivision of the interval $[0, L]$ with more refined subdivisions where f_0 is less regular such that the number of points is less than ϵ_n^{-1} points.

This approximation is adapted from the proof of Theorem 2.5.7 in van der Vaart and Wellner (1996). We then identify a finite piecewise constant density by a mixture of uniforms for which the Hellinger distance between the piecewise constant approximation f_P of $f_0 \in \mathcal{F}$ and f_0 is less that ϵ_n and $||f_0/f_P||_\infty \leq M$. The following Lemma gives the form of a finite probability distribution P such that f_P is in the Kullback-Leibler neighbourhood of some $f \in \mathcal{F}$.

Lemma 11. *Let $f \in \mathcal{F}_L$ be such that $f(0) \leq M < +\infty$. For all $0 < \epsilon < 1$ there exists $m \lesssim L^{1/3}M^{1/3}\epsilon^{-1}$, $p = (p_1, \ldots, p_m) \in \mathfrak{S}_m$ and $x = (x_1, \ldots, x_m) \in [0, L]^m$ such that $P = \sum_{i=1}^m \delta_{x_i} p_i$ satisfies*

$$KL(f, f_P) \lesssim \epsilon^2, \ \int \left(\log\left(\frac{f}{f_P}\right)\right)^2 f \lesssim \epsilon^2, \tag{24}$$

where f_P is defined as in (1).

Proof. For a fixed ϵ, let f be in $\mathcal{F}_L$. Consider $\mathcal{P}_0$ the coarsest partition:

$$0 = x_0^0 < x_1^0 = L.$$

Then at the ith step, let $\mathcal{P}_i$ be the partition

$$0 = x_0^i < x_1^i < \cdots < x_{n_i}^i = L,$$

and define

$$\varepsilon_i = \max_j \left\{ (f(x_{j-1}^i) - f(x_j^i))(x_j^i - x_{j-1}^i)^{1/2} \right\}.$$

For each $j \geq 1$, if $(f(x^i_{j-1}) - f(x^i_j))(x^i_j - x^i_{j-1})^{1/2} \geq \frac{\varepsilon_i}{\sqrt{2}}$ we split the interval $[x_{j-1}, x_j]$ into two subsets of equal length. We then get a new partition $\mathcal{P}_{i+1}$. We continue the partitioning until the first k such that $\varepsilon_k^2 \leq \epsilon^3$. At each step i, let n_i be the number of intervals in $\mathcal{P}_i$, s_i the number of interval in $\mathcal{P}_i$ that have been divided to obtain $\mathcal{P}_{i+1}$, and $c = 1/\sqrt{2}$. Thus, it is clear that $\varepsilon_{i+1} \leq c\varepsilon_i$

$$\begin{aligned} s_i(c\varepsilon_i)^{2/3} &\leq \sum_j (f(x^i_{j-1}) - f(x^i_j))^{2/3}(x^i_j - x^i_{j-1})^{1/3} \\ &\leq \left(\sum_j f(x^i_{j-1}) - f(x^i_j)\right)^{2/3} \left(\sum_j x^i_j - x^i_{j-1}\right)^{1/3} \leq M^{2/3}L^{1/3}, \end{aligned}$$

using Hölder's inequality. We then deduce that

$$\begin{aligned} \sum_{j=1}^k n_j = k + \sum_{j=1}^k j s_{k-j} \leq 2\sum_{j=1}^k j s_{k-j} &\leq 2\sum_{j=1}^k j M^{2/3}L^{1/3}(c\varepsilon_{k-j})^{-2/3} \\ &\leq 2M^{2/3}L^{1/3}\varepsilon_k^{-2/3}2^{1/3}\sum_{j=1}^k j2^{-j/3} \\ &\leq K_0 M^{2/3}L^{1/3}\varepsilon_k^{-2/3}, \end{aligned}$$

where $K_0 = 2(1-2^{-2/3})^{-2}$. Thus

$$n_k \leq K_0 M^{2/3}L^{1/3}\epsilon^{-1}. \tag{25}$$

Now, for $f \in \mathcal{F}_L$, we prove that there exists a stepwise density with less than $K_0 M^{2/3}L^{1/3}\frac{1}{\epsilon}$ pieces such that

$$KL(f,h) \leq \epsilon^2 \text{ and } \int f \log(\frac{f_0}{f_P})^2(x)dx \lesssim \epsilon^2 \tag{26}$$

In order to simplify notations, we define

$$x_i = x^k_i, \quad l_i = x_i - x_{i-1}, \quad g_i = f(x_{i-1})^{1/2}.$$

We consider the partition constructed above associated with $f^{1/2}$, which is also a monotone non-increasing function that satisfies $f^{1/2}(0) \leq M^{1/2}$ (instead of M). We denote g the function defined as $g(x) = \sum \mathbb{I}_{[x_{i-1},x_i]}(x)g_i$

$$\begin{aligned} ||f^{1/2} - g||_2^2 = \int (f^{1/2} - g)^2(x)dx &= \sum_{i=1}^{n_k} \int_{I_i} (f^{1/2} - g)^2(x)dx \\ &\leq \sum_{i=1}^{n_k} \int_{I_i} (f^{1/2}(x^k_{i-1}) - f^{1/2}(x^k_i))^2 dx \\ &\leq \sum_{i=1}^{n_k} (x^k_i - x^k_{i-1})(f^{1/2}(x^k_{i-1}) - f^{1/2}(x^k_i))^2 \\ &\leq n_k\varepsilon_k^2 \leq L^{1/3}K_0 M^{1/3}\epsilon^2. \end{aligned}$$

We then define $h = \frac{g^2}{\int g^2}$ and and get an equivalent of $\int g^2$.

$$\begin{aligned}\int g^2 dx &= \int (g^2 - f)(x)dx + 1 \\ &= \int (g - \sqrt{f})(g + \sqrt{f})(x)dx + 1 \\ &= 1 + \mathcal{O}(\epsilon),\end{aligned}$$

and deduce that $(\int g^2)^{1/2} = 1 + \mathcal{O}(\varepsilon)$. Let H be the Hellinger distance

$$\begin{aligned}H(f,h) &= H\left(f, \frac{g^2}{\int g^2}\right) \\ &\leq H(f, g^2) + H(g^2, \frac{g^2}{\int g^2}) \\ &\leq L^{1/6} K_0 M^{1/6} \epsilon + \left(\int (g - \frac{g}{(\int g^2)^{1/2}})^2(x)dx\right)^{1/2} \lesssim \epsilon.\end{aligned}$$

Since $||f/h||_\infty = ||f/g^2||_\infty (\int g^2) \leq (\int g^2)$, together with the above bound on $H(f,h)$ and Lemma 8 from Ghosal and van der Vaart (2007), we obtain the required result.

Let P be a probability distribution defined by

$$P = \sum_{i=1}^{n_k} p_i \delta(x_i^k) \quad p_i = (h_{i-1} - h_i) x_i^k \quad p_{n_k} = h_{n_k} x_{n_k}^k = h_{n_k} L,$$

thus $f_P = h$ and given the previous result, lemma 11 is proved. □

Given Lemma 11, we now prove Lemma 8.

Proof of Lemma 8. We first consider the case where $\theta^{t_1} \lesssim \alpha(\theta) \lesssim \theta^{t_2}$ for small θ. For ϵ_n as in Theorem 1, define θ_n as

$$\theta_n = \inf\left\{x, 1 - F_0(x) < \frac{\epsilon_n}{2n}\right\}.$$

Note that F_0 is càdlàg, thus

$$F_0(\theta_n) \geq 1 - \epsilon_n/(2n) \text{ and } \forall y < \theta_n 1 - F_0(y) > \epsilon_n/(2n). \tag{27}$$

Using lemma 11 with $L = \theta_n$, we obtain that there exists a distribution $P = \sum_{i=1}^{n_k} \delta_{x_i} p_i$ such that

$$KL(f_{0,n}, f_P) \leq \epsilon_n^2, \text{ and } \int f_{0,n} \log\left(\frac{f_{0,n}}{f_P}\right)^2 \lesssim \epsilon_n^2.$$

Note that f_P has support $[0, \theta_n]$ and is such that $f_P(\theta_n) > 0$. Now, set $m = n_k$ and consider P' the mixing distribution associated with $\{m, x_1', \ldots, x_m',$

$p'_1 \dots, p'_m\}$ with $\sum_{i=1}^m p'_i = 1$. Define for $1 \leq i \leq m-1$ the set $U_i = [0 \vee (x_i - \epsilon_n^3/M, x_i + \epsilon_n^3/M]$ and $U_m = (\theta_n, \theta_n + \epsilon_n(L-\theta_n) \wedge \epsilon_n^3/M]$. Construct P' such that $x'_i \in U_i$ and $|P'(U_i) - p_i| \leq \epsilon^2 m^{-1}$. We get

$$\forall t \in [0, \theta_n] \; f'_P(t) > \frac{p'_m}{x'_m}.$$

Given that $x'_m \in U_m$, we get $x'_m \leq \theta_n + \epsilon_n(L-\theta_n) \wedge \epsilon_n^3/M \lesssim \theta_n$ for n large enough. Note also that $p'_m \geq p_m - \epsilon_n^2 m^{-1}$. Given the construction of Lemma 11, we deduce

$$p_m \geq \frac{f_0(x_{i-1})}{1 + \mathcal{O}(\epsilon_n)} \gtrsim f_0(x_{i-1}),$$

for n large enough. Furthermore, given (27)

$$\forall z < \theta_n, \; f_0(z)(L-z) \geq \int_z^L f_0(t)dt \geq \frac{\epsilon_n}{2n},$$

thus

$$\forall t \in [0, \theta_n] \; f'_P(t) \gtrsim \frac{\frac{\epsilon_n}{2n} - \epsilon_n^2 m^{-1}}{\theta_n} \gtrsim \frac{\epsilon_n}{n},$$

and deduce that $||f_0/f_{P'}||_\infty \lesssim \frac{n}{\epsilon_n}$ Lemma 8 from Ghosal and van der Vaart (2007) gives us that

$$\begin{aligned}\int_0^{\theta_n} f_0(x) \log\left(\frac{f_0}{f_{P'}}\right)(x)dx &\lesssim \left(\epsilon_n^2 + H^2(f_P, f_{P'})\right)(1 + |\log(\epsilon_n/n)|) \\ &\lesssim \left(\epsilon_n^2 + |f_P - f_{P'}|_1\right)(1 + |\log(\epsilon_n/n)|).\end{aligned}$$

Given the mixture representation (1) of f_0 and f_P, we get

$$\begin{aligned}&\left(\epsilon_n^2 + |f_P - f_{P'}|_1\right)(1 + \log(n)) \\ &\quad \lesssim \left(\epsilon_n^2 + \int_0^{\theta_n} \Big| \sum \Big(\frac{p_i}{x_i} - \frac{p'_i}{x'_i}\Big) \mathbb{I}_{x \leq x_i} + \sum \frac{p_i}{x_i}(\mathbb{I}_{x \leq x_i} - \mathbb{I}_{x \leq x'_i}) \Big| dx\right)(1 + \log(n)) \\ &\quad \lesssim \left(\epsilon_n^2 + \sum |\frac{x_i}{x'_i} - 1| p'_i + \sum |p'_i - p_i| + \sum \frac{p_i}{x_i}|x'_i - x_i|\right)(1 + |\log(n)|) \\ &\quad \lesssim \epsilon_n^2(1 + |\log(n)|).\end{aligned}$$

Generally speaking, denoting $U_0 = [0,1] \cap (\cup_{i=1}^m U_i)^c$ and $\mathcal{N} = \{P', |P'(U_i) - p_i| \leq \epsilon_n^2 m^{-1}\}$ we obtain that for all $P' \in \mathcal{N}$

$$\int_0^{\theta_n} f_0(x) \log\left(\frac{f_0}{f_{P'}}\right)(x)dx \lesssim \epsilon_n^2(1 + |\log(n)|),$$

and similarly

$$\int_0^{\theta_n} f_0(x) \log\left(\frac{f_0}{f_{P'}}\right)^2(x)dx \lesssim \epsilon_n^2(1 + |\log(n)|)^2,$$

for ϵ_n small enough. Note also that for all $P' \in \mathcal{N}$ and n large enough, as before we get

$$\int_{\theta_n}^{L} f_{P'}(x)dx \lesssim \frac{\epsilon_n}{n}.$$

We now derive a control on k, the number of steps until $\varepsilon_k \leq \epsilon_n^{3/2}$ in the construction of Lemma 11. At step $k-1$, we have $\varepsilon_{k-1} \geq \epsilon_n^{3/2}$. It is clear that for all j, $\varepsilon_j \leq 2^{-1/2}\varepsilon_{j-1}$, thus

$$M^{1/2}L^{1/2}2^{-(k-1)/2} \geq \varepsilon_{k-1} \geq \epsilon_n^{3/2}$$
$$\log(M^{1/2}L^{1/2}) - (k-1)\frac{\log(2)}{2} \geq \frac{3}{2}\log(\epsilon_n).$$

Finally, we have

$$k \leq \frac{2}{\log(2)}(\log(M^{1/2}L^{1/2}) - \frac{3}{2}\log(\epsilon_n)) + 1. \tag{28}$$

We can then get a lower bound for $\Pi[\mathcal{N}]$ and, given that for ϵ_n small enough and n large enough, we have

$$\mathcal{N} \subset S_n(\epsilon_n, \theta_n),$$

we can deduce a lower bound for $\Pi(S_n(\epsilon_n, \theta_n))$. For the Type 1 prior, we have similarly to Ghosal et al. (2000)

$$\begin{aligned}\Pi[\mathcal{N}] &= \Pr(\mathcal{D}(A\alpha(U_0), \ldots, A\alpha(U_{n_k})) \in [p_i \pm \epsilon_n^2/n_k]) \\ &\geq \frac{\Gamma(A)}{\prod_i \Gamma(A\alpha(U_i))} \prod_j \int_{(p_i-\epsilon_n^2/n_k)\wedge 0}^{(p_i+\epsilon_n^2/n_k)} x_j^{A\alpha(U_j)-1} dx_j.\end{aligned}$$

Given condition C1, we have

$$\alpha(U_i) \geq \int_{U_i} \alpha_0 \theta^{t_1} d\theta,$$

thus

$$\alpha(U_i) \geq 2\epsilon_n^3 \alpha_0 {x_i}^{t_1}.$$

For n large enough and ϵ sufficiently small we have as in Lemma 6.1 of Ghosal et al. (2000)

$$\Pi(\mathcal{N}) \gtrsim \exp\{C_1 n_k \log(\epsilon)\}.$$

Note that given (25), $n_k \lesssim \epsilon_n^{-1}$ which gives the desired result. For the Type 2 prior, we write

$$\mathcal{N}' = \left\{ P' = \sum_{j=1}^{n_k} p'_j \delta_{x'_j}, |p'_j - p_j| \leq \epsilon^2/n_k, |x'_j - x_j| \leq \epsilon_n^3 \right\} \subset S_n(\epsilon_n, \theta_n),$$

we then deduce a lower bound for $\Pi[S_n(\epsilon_n, \theta_n)]$

$$\begin{aligned}
\Pi[\mathcal{N}'] &\geq Q(K = n_k) \prod_{j=1}^{n_k} n_k^{-n_k} c^{n_k} \int_{\max(0, p_i - \epsilon^2/n_k)}^{p_i + \epsilon^2/n_k} w_j^{a_j} \, dw_j \prod_{j=1}^{n_k} \alpha(U_i) \\
&\geq \exp\Big\{ - c n_k \log n_k + \sum \log(\alpha(U_i)) \\
&\qquad + n_k(\log(c) - \log(n_k)) + \sum a_j \log(2\epsilon^2/n_k) \Big\} \\
&\gtrsim \exp\left\{ C_1' \epsilon^{-1} \log(\epsilon) \right\}.
\end{aligned}$$

We now consider the case where $e^{-a_1/\theta} \leq \alpha(\theta) \leq e^{-a_2/\theta}$ if θ is close to 0 and $\sup_{x \in [0,\delta]} |f_0'(x)| \leq C_0$. We have that for n large enough and $C > 0$, a constant depending on f_0, $f_0(0) - f_0(\epsilon_n) \leq C\epsilon_n$. Following Lemma 11, we can construct a piecewise constant approximation of f_0 on $[\delta, L]$. On $[0, \delta]$, consider the regular partition with $\lfloor \epsilon_n^{-1} \rfloor$ points and the piecewise constant approximation of f_0 defined as before (i.e. $f_i = f_0(x_{i-1})$). Again, this approximation can be identified with a measure P. Given the assumptions on f_0 we immediately get that $KL(f_0, f_P) \lesssim \epsilon_n^2$.

Consider the same sets $\mathcal{N}$ as before, with the same partitions $U_1, \ldots, U_n$. Using similar computations as in Lemma 6.1 of Ghosal et al. (2000) we get that

$$\Pi(\mathcal{N}) \geq \exp\left\{ C_1(n_k + \epsilon_n^{-1}) \log(\epsilon_n) + \sum \log(\alpha(U_i)) \right\}$$

For the U_i included in $[\delta, L]$ we have $\alpha(U_i) \gtrsim \epsilon_n^{3/2}$. For the U_i included in $[0, \delta]$ we have $\alpha(U_i) \gtrsim \epsilon_n \exp\{-a/(i\epsilon_n)\}$, which gives

$$\sum \alpha(U_i) \lesssim -\epsilon_n^{-1} \log(n)$$

We conclude the proof using similar arguments as before. □

A.2. Proof of Lemma 3.1

The proof of Lemma 3.1 is straightforward and comes directly from C1 and C2.

Proof. Recall that given (1), $f(0) = \int_{[0,1]} \frac{1}{\theta} dP(\theta)$. Then

$$\Pi\left[\int_0^1 \frac{1}{\theta} dP(\theta) \geq M_n \right] = \Pi\left[\int_0^{2M_n^{-1}} \frac{1}{\theta} dP(\theta) + \int_{2M_n^{-1}}^1 \frac{1}{\theta} dP(\theta) \geq M_n \right].$$

Note that

$$\int_{2M_n^{-1}}^1 \frac{1}{\theta} dP(\theta) \leq M_n/2 \int_{2M_n^{-1}}^1 dP(\theta) \leq M_n/2.$$

Thus the set $\{P, \int_0^{2M_n^{-1}} \theta^{-1} dP(\theta) \geq M_n/2\}$ contains $\mathcal{F}_n^c$ and

$$\Pi[\mathcal{F}_n^c] \leq \Pi\left[\int_0^{2M_n^{-1}} \frac{1}{\theta} dP(\theta) > M_n/2 \right]$$

$$\leq \quad 2M_n^{-1} E\left[\int_0^{2M_n^{-1}} \frac{1}{\theta} dP(\theta)\right],$$

using Markov inequality. Then for a Type 1 prior when n large enough

$$\begin{aligned}\Pi[\mathcal{F}_n^c] &\leq 2M_n^{-1}\int_0^{2M_n^{-1}} \frac{1}{\theta}\alpha(\theta)d\theta \\ &\leq 2M_n^{-1}\int_0^{2M_n^{-1}} \theta^{t_2-1}d\theta = \frac{(2M_n^{-1})^{t_2+1}}{t_2} = Ce^{-cn^{1/3}\log(n)^{2/3}}.\end{aligned}$$

For a Type 2 prior, we have that

$$\begin{aligned}\Pi[\mathcal{F}_n^c] &\leq \sum_{h=1}^{\infty} Q(K=k)\pi_k\left[\min_{j\leq k} x_j \leq M_n^{-1}\right] \\ &\leq \left(\sum_{h=1}^{\infty} kQ(K=k)\right)\alpha([0,M_n^{-1}]) \\ &\leq C'e^{-cn^{1/3}\log(n)^{2/3}}.\end{aligned}$$

□

Appendix B: Adaptation of Theorem 4 of Rivoirard and Rousseau (2012)

This Theorem is a slight modification of Theorem 2.9 of Ghosal et al. (2000). The main difference lies in the handling of the denominator D_n in

$$\Pi(f : d(f_0,f) \geq J_{0,n}\epsilon_n|\mathbf{X^n}) = \frac{\int_{d(f,f_0)\geq J_{0,n}\epsilon_n}\prod_{i=1}^n \frac{f(X_i)}{f_0(X_i)}d\Pi(f)}{\int\prod_{i=1}^n \frac{f(X_i)}{f_0(X_i)}d\pi(f)} = \frac{N_n}{D_n},$$

as in general, it require a lower bound on the prior mass of Kullback Leibler neighborhood of f_0. Here we prove that under condition (16) we have for some constants $c, C > 0$

$$P_0^n(D_n < ce^{-Cn\epsilon_n^2}) = o(1).$$

Let $l_n(f)$ be the log likelihood associated with f and define $\Omega_n = \{(f,\mathbf{X^n}), l_n(f) - l_n(f_0) > -C_1 n\epsilon_n^2\}$ for some constant $C_1 > 0$. Define also $A_n = \{\mathbf{X^n}, \forall i X_i \leq \theta_n\}$. We thus have

$$D_n \geq e^{-C_1 n\epsilon_n^2}\int_{S_n(\epsilon_n,\theta_n)} \mathbb{I}_{\Omega_n} d\Pi(f) = e^{-C_1 n\epsilon_n^2}\Pi(S_n(\epsilon_n,\theta_n)\cap\Omega_n).$$

Note that given (16) we have that there exists $\rho > 0$ such that for n large enough $e^{-C_2 n\epsilon_n^2}\Pi(S_n(\epsilon_n,\theta_n) > \rho$. We now write

$$\begin{aligned}&P_0^n(D_n < e^{-Cn\epsilon_n^2}) \\ &\quad \leq P_0^n\left(e^{(C-C_1)n\epsilon_n^2}\Pi(S_n(\epsilon_n,\theta_n)\cap\Omega_n) < c\right)\end{aligned}$$

$$
\begin{aligned}
&\leq P_0^n\left(e^{(C-C_1-C_2)n\epsilon_n^2}\Pi(S_n(\epsilon_n,\theta_n)\cap\Omega_n<\frac{c}{\rho}\Pi(S_n(\epsilon_n,\theta_n)\right)\\
&\leq P_0^n\left(\Pi(S_n(\epsilon_n,\theta_n)\cap\Omega_n^c)>\left(1-e^{-(C-C_1-C_2)n\epsilon_n^2}\frac{c}{\rho}\right)\Pi(S_n(\epsilon_n,\theta_n))\right)\\
&\leq \frac{2\int_{S_n(\epsilon_n,\theta_n)}P_0^n(\Omega_n^c)d\Pi(f)}{\Pi(S_n(\epsilon_n,\theta_n))}.
\end{aligned}
$$

For all $f\in S_n(\epsilon_n,\theta_n)$ we compute

$$
\begin{aligned}
m_n &= \mathrm{E}_0^n(l_n(f_0)-l_n(f)\mathbb{I}_{A_n})\\
&= nF_0(\theta_n)^{n-1}\int_0^{\theta_n} f_0\log\left(\frac{f_0(x)}{f(x)}\right)dx\\
&= nF_0(\theta_n)^n\left(KL(f_{0,n},f_n)+\log\left(\frac{F_0(\theta_n)}{F(\theta_n)}\right)\right)\\
&\leq C_3 n\epsilon_n^2,
\end{aligned}
$$

and

$$
\begin{aligned}
P_0^n(\Omega_n^c) &= P_0^n(l_n(f)-l_n(f_0)<-C_1n\epsilon_n^2)\\
&= P_0^n(\{l_n(f)-l_n(f_0)<-C_1n\epsilon_n^2\}\cap A_n)+o(1)\\
&\leq P_0^n(\{l_n(f_0)-l_n(f)-m_n>(C_1-C_3)n\epsilon_n^2\}\cap A_n)+o(1)\\
&\leq \frac{\mathrm{E}_0^n\left(\{l_n(f_0)-l_n(f)-m_n\}\mathbb{I}_{A_n}\right)^2}{(C_1-C_3)^2(n\epsilon_n^2)^2}+o(1).
\end{aligned}
$$

We then compute for C_5 and C_6 some fixed constants

$$
\begin{aligned}
v_n &= \mathrm{E}_0^n(\{l_n(f_0)-l_n(f)-m_n\}\mathbb{I}_{A_n})^2\\
&= (F_0(\theta_n))^{n-1}\left(n\int_0^{\theta_n} f_0\log^2\left(\frac{f_0(x)}{f(x)}\right)dx\right.\\
&\quad\left.+n(n-1)\left(\int_0^{\theta_n} f_{0,n}\log\left(\frac{f_0(x)}{f(x)}\right)dx\right)^2-m_n^2\right)\\
&= (F_0(\theta_n))^{n-1}\left(n\int_0^{\theta_n} f_0\log^2\left(\frac{f_0(x)}{f(x)}\right)dx+\frac{n-1}{n}F_0(\theta_n)^{-2n+2}m_n^2-m_n^2\right)\\
&\leq nF_0(\theta_n)^n\int_0^{\theta_n} f_{0,n}\log^2\left(\frac{f_0(x)}{f(x)}\right)dx+\frac{n-1}{n}m_n^2F_0(\theta_n)^{n-1}(F_0(\theta_n)^{-2n+2}-1)\\
&\leq C_5n\epsilon_n^2+C_6(n\epsilon_n^2)^2\epsilon_n.
\end{aligned}
$$

We finally obtain that for all $f\in S_n(\epsilon_n,\theta_n)$, $P_0^n(\Omega_n^c)=o(1)$. We end the proof using similar arguments as in Ghosal et al. (2000).

References

Balabdaoui, F. and Wellner, J. A. (2007). Estimation of a k-monotone density: limit distribution theory and the spline connection. *Ann. Statist.*, 35(6):2536–2564. MR2382657

Brunner, L. J. and Lo, A. Y. (1989). Bayes methods for a symmetric unimodal density and its mode. *Ann. Statist.*, 17(4):1550–1566. MR1026299

Castillo, I. (2013). On Bayesian supremum norm contraction rates. *arXiv preprint arXiv:1304.1761*.

Donoho, D. L. and Liu, R. C. (1991). Geometrizing rates of convergence, ii. *The Annals of Statistics*, pages 633–667. MR1105839

Durot, C., Kulikov, V. N., and Lopuhaä, H. P. (2012). The limit distribution of the L_∞-error of Grenander-type estimators. *Ann. Statist.*, 40(3):1578–1608. MR3015036

Ferguson, T. S. (1983). Bayesian density estimation by mixtures of normal distributions. In *Recent Advances in Statistics*, pages 287–302. Academic Press, New York. MR0736538

Ghosal, S., Ghosh, J. K., and van der Vaart, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.*, 28(2):500–531. MR1790007

Ghosal, S. and van der Vaart, A. (2007). Posterior convergence rates of Dirichlet mixtures at smooth densities. *Ann. Statist.*, 35(2):697–723. MR2336864

Ghosal, S. and van der Vaart, A. W. (2001). Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Ann. Statist.*, 29(5):1233–1263. MR1873329

Giné, E. and Nickl, R. (2010). Confidence bands in density estimation. *Ann. Statist.*, 38(2):1122–1170. MR2604707

Grenander, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuarietidskr.*, 39:125–153 (1957). MR0093415

Groeneboom, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Vol. II (Berkeley, Calif., 1983)*, Wadsworth Statist./Probab. Ser., pages 539–555. Wadsworth, Belmont, CA. MR0822052

Groeneboom, P. (1986). Some current developments in density estimation. In *Mathematics and Computer Science (Amsterdam, 1983)*, volume 1 of *CWI Monogr.*, pages 163–192. North-Holland, Amsterdam. MR0873578

Hoffmann, M., Rousseau, J., and Schmidt-Hieber, J. (2013). On adaptive posterior concentration rates. *arXiv preprint arXiv:1305.5270*.

Khazaei, S., Rousseau, J., and Balabdaoui, F. (2010). Bayesian Nonparametric Inference of decreasing densities. In *42èmes Journées de Statistique*. Marseille, France.

Kruijer, W., Rousseau, J., Van Der Vaart, A., et al. (2010). Adaptive bayesian density estimation with location-scale mixtures. Technical report.

Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer Series in Statistics. Springer-Verlag, New York. MR0856411

Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates. I. Density estimates. *Ann. Statist.*, 12(1):351–357. MR0733519

Prakasa Rao, B. L. S. (1970). Estimation for distributions with monotone failure rate. *Ann. Math. Statist.*, 41:507–519. MR0260133

Reynaud-Bouret, P., Rivoirard, V., and Tuleau-Malot, C. (2011). Adaptive density estimation: a curse of support? *J. Statist. Plann. Inference*, 141(1):115–139. MR2719482

Rivoirard, V. and Rousseau, J. (2012). Bernstein-von Mises theorem for linear functionals of the density. *Ann. Statist.*, 40(3):1489–1523. MR3015033

Robertson, T., Wright, F. T., and Dykstra, R. L. (1988). *Order Restricted Statistical Inference*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Ltd., Chichester. MR0961262

Rousseau, J. (2010). Rates of convergence for the posterior distributions of mixtures of betas and adaptive nonparametric estimation of the density. *Ann. Statist.*, 38(1):146–180. MR2589319

Rousseau, J. and Mengersen, K. (2011). Asymptotic behaviour of the posterior distribution in overfitted mixture models. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 73(5):689–710. MR2867454

Shen, W., Tokdar, S. T., and Ghosal, S. (2013). Adaptive Bayesian multivariate density estimation with Dirichlet mixtures. *Biometrika*, 100(3):623–640. MR3094441

Sun, J. and Woodroofe, M. (1996). Adaptive smoothing for a penalized NPMLE of a non-increasing density. *J. Statist. Plann. Inference*, 52(2):143–159. MR1392133

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and Empirical Processes*. Springer Series in Statistics. Springer-Verlag, New York. With applications to statistics. MR1385671

Williamson, R. E. (1956). Multiply monotone functions and their Laplace transforms. *Duke Math. J.*, 23:189–207. MR0077581

Wu, Y. and Ghosal, S. (2008). Kullback Leibler property of kernel mixture priors in Bayesian density estimation. *Electron. J. Stat.*, 2:298–331. MR2399197