Nonparametric Bayesian estimation of densities under monotonicity constraint

S. Khazaei , J. Rousseau and F. Balabdaoui

May 26, 2011

Abstract

In this paper, we study consistency in nonparametric Bayesian estimation of a nonincreasing density on $\mathbb{R}^+$. Since such a density can be written as a mixture of uniform densities, it is natural to consider Bayesian procedures that are based on nonparametric mixture models. In this paper, we consider in particular as priors the Dirichlet mixture process and finite mixtures with unknown number of components. We show that for finite mixture priors, the Kullback-Leibler property is not satisfied, and we propose an alternative condition that still ensures consistency of the posterior distributions and which is satisfied in those models. A simulation study is provided at the end of the paper to illustrate the results.

Keywords: Nonparametric Bayesian inference, Consistency, entropy, Kullback-Leibler, k-monotone density, kernel mixture.

1 Introduction

In this paper we study the properties of Bayesian nonparametric estimation of a monotone nonincreasing density. Nonparametric estimation of a monotone nonincreasing density is a well known problem and has been considered both from theoretical and applied perspectives in the frequentist literature, see for instance the introduction in Balabdaoui and Wellner [3] for a review on the subject. In particular, monotone density estimation has applications in reliability, and serves as a preliminary analysis in survival analysis. Monotone nonincreasing densities on $\mathbb{R}^+$ have a mixture representation which allows for likelihood based inference, see for instance Balabdaoui and Wellner [3]. Alternative to the frequentist Maximum Likelihood approach is a Bayesian approach, also based on the likelihood. Since Williamson [26] and Lévy [14] (see also Gneiting [7]) it is known that a density f is monotone nonincreasing on $\mathbb{R}^+$ if and only if it can be written as a mixture of uniform densities; i.e.,

$$f(x) = \int_0^\infty \frac{1}{\theta} \mathbb{I}_{(0,\theta)}(x) dP(\theta) \tag{1.1}$$

where P is a mixing distribution on $\mathbb{R}^+$ and $\mathbb{I}_{(0,\theta)}(x)$ is the indicator function on the interval $(0, \theta)$. Note that if F is the cdf of f inversion of (1.1) yields the formula $P(] - \infty, x]) = F(x) - xf(x)$, which is valid for all continuity points x of P. Herefater, f will be denoted by f_P so that the dependence on the corresponding mixing distribution is made explicit.

Let $\mathcal{F}$ be the set of nonincreasing densities on $\mathbb{R}^+$. Characterizing monotone nonincreasing densities via (1.1) leads naturally to a mixture type prior on the class $\mathcal{F}$. In Bayesian nonparametric estimation of densities, many different types of priors have been used in the litterature for mixture models. For instance, Ferguson [6] and Lo [15] have introduced Dirichlet mixtures, Brunner and Lo[1989][] have considered Bayesian procedures for sampling from a unimodal density and Kottas and Krnjaic [2009][] extended the Bayesian semiparametric models using DPM for the error distribution, Petrone and Wasserman [19] studied, among others, the properties of Bernstein polynomials, Robert and Rousseau [21] and Rousseau [22] obtained consistency and rates of convergence for general mixtures of Betas, and Perron and Mengersen [20] and McVinish et al. [16] studied mixtures of triangulars, the well known family of Gaussian mixtures has been studied in particular by Ghosal and Van der Vaart (2001) [10] and Scricciolo (2001) [28]. A more general study on approximation properties of mixture models can be found in Wu and Ghosal [27].

We construct a prior on $\mathcal{F}$ by determining a prior on the mixing distribution P living on $\mathbb{R}^+$. In this paper, we are interested in two types of discrete distributions generating P:

- Dirichlet process priors: $P \sim \mathrm{DP}(H, \alpha)$, where H is a probability measure on $\mathbb{R}^+$ and α as a concentration parameter is a positive real number .
- Finite mixtures: Set $P = \sum_{j=1}^K w_j \delta_{\theta_j}$, with $K \in \mathbb{N}^*$, $w = (w_1, ..., w_K)$ is in the K dimensional simplex : $\{w = (w_1, ..., w_K), w_i \geq 0, \sum_i w_i = 1\}$, $0 \leq \theta_1 \leq \cdots \leq \theta_K$, and δ_{θ_j} is the Dirac distribution putting all its mass at θ_j. The prior on P is then defined by: $K \sim Q$ a probability distribution on $\mathbb{N}^*$, and conditionally on K, $(w_1, ..., w_K) \sim \pi_{w,K}$ and $\theta = (\theta_1, ..., \theta_K) \sim \pi_{\theta,K}$, where $\pi_{w,K}$ and $\pi_{\theta,K}$ are probabiliy distributions. A typical example is to take $\pi_{w,K}$ to be a Dirichlet distribution

and the θ_j's the order statistics of K independently distributed random variables with density α on $\mathbb{R}^+$.

Nonparametric Bayesian procedures seem to present the following paradox. On the one hand, they do not require precise information of the shape of the parameter (here function). On the other hand, they require the construction of a distribution on an infinite dimensional space, and this cannot be accomplished in a purely subjective way. Hence, as argued by Diaconis and Freedman [5] or illustrated by Lijoi et al. [25], strong consistency of the posterior distribution is a major issue in nonparametric Bayesian statistics.

Our aim in this article is to study the consistency of posterior distributions on $\mathcal{F}$ based on either of the two types of priors described above. More precisely let $X^n = (X_1, ..., X_n)$ be a sample of independently and identically distributed observations with a common probability distribution F_0 having a nonincreasing density f_0 with respect to Lebesgue measure. Let π be a probability measure on $\mathcal{F}$ and $\pi[.|X^n]$ denote the posterior distribution associated with π. We recall that strong consistency is satisfied when for all $\epsilon > 0$,

$$\pi\left[\{f \in \mathcal{F}; d(f_0, f) \leq \epsilon\}|X^n\right] \to 1, \quad P_0^\infty \quad \text{a.s.}, \tag{1.2}$$

where d denotes either the L_1 distance or the Hellinger distance between f_0 and f. Note that in the case of the nonincreasing densities, Lijoi et al. [25] have proved that strong consistency is equivalent to weak consistency, so that it is enough to prove the above convergence for weak neighbourhoods of f_0. Most results on posterior consistency (weak or strong) are based on the Kullback-Leibler property, i.e.

$$\pi\left[\{f \in \mathcal{F}; \mathrm{KL}(f_0, f) \leq \epsilon\}|X^n\right] > 0, \quad \forall \epsilon > 0, \tag{1.3}$$

where $\mathrm{KL}(f_1, f_2) = \int f_1 \log(f_1/f_2)dx$ is the Kullback-Leibler divergence between f_1 and f_2. If π satisfies the above condition, f_0 is said to be in the Kullback-Leibler support of π (hereafter calleds KL-support of π). General conditions on Bayesian mixture models are presented in Wu and Ghosal [27] to verify (1.3). In particular, they proved that Dirichlet process priors satisfy (1.3) in the framework of nonincreasing densities on $\mathbb{R}^+$ under mild conditions on the base measure and on f_0. The study of finite mixtures is more delicate since (1.3) is usually not valid. This point will be discussed in Section 2. To circumvent this problem, we prove a result (Theorem 2.1) where posterior weak consistency is established under conditions alternative to the usual KL condition in (1.3). This result has interest in its own right and could be used in other contexts. Hence it is provided in a separate section. Furthermore, we show consistency of the posterior distribution under Dirichlet process and finite mixture priors on [0,1] and on $\mathbb{R}^+$ in Sections 2.3.2 and 2.3.3 respectively.

Finally, a simulation study was carried out with the goal of illustrating the theory for the Dirichlet process mixture priors. Description of the simulations and the obtained results can be found in Section 3.

2 Asymptotic properties of the posterior distribution

Let $\mathcal{M}$ be the set of probability distributions on $\mathbb{R}^+$. We denote by Π a probability measure on $\mathcal{M}$. Recall that f_P denotes the nonincreasing density with mixing distribution P (as defined in (1.1)). Let F_0, F and F_P be the cumulative distribution functions associated with f_0, f and f_P.

The posterior probability of any measurable set A of $\mathcal{M}$ given the observed sample X^n is given by

$$\Pi(A|X^n) = \frac{\int_A \prod_{i=1}^n f_P(X_i)d\Pi(P)}{\int_{\mathcal{M}} \prod_{i=1}^n f_P(X_i)d\Pi(P)}. \tag{2.1}$$

2.1 Remarks on the Kullback-Leibler condition

Wu and Ghosal [27] and Ghosh and Ramamorti [8] obtained some interesting results on weak consistency in Bayesian estimation of nonincreasing densities. They proved that if the weak support of Π is $\mathcal{M}$, then any continuous nonincreasing density f_0 satisfying

$$\int f_0(x)\,|\log f_0(x)|\,dx < \infty$$

is in the KL support of the prior, which implies weak consistency of the posterior at f_0. However, the condition on the weak support of Π can be very strong and quite difficult to prove. Thus, we propose an alternative set of conditions which can be, in some cases, easier to deal with.

As an example, consider the finite mixture type of priors described in Section 1. Recall that a realization under such a prior is written as

$$P = \sum_{j=1}^{K} w_j \delta_{\theta_j}.$$

Assume that the conditional probabilities $\pi_{w,K}$ and $\pi_{\theta,K}$ given K are absolutely continuous with respect to Lebesgue measure. Then if $f_0(x) > 0$ for all $x \in \mathbb{R}^+$, we have that

$$\Pi(\{P : \mathrm{KL}(f_0, f_P) = \infty\}) = 1. \tag{2.2}$$

Indeed, if $dP(\theta) = \sum_{j=1}^K w_j \delta_{\theta_j}(\theta)$ with $0 < \theta_1 < \ldots < \theta_K$, then for all $x > \theta_K, f_P(x) = 0$ and $\mathrm{KL}(f_0, f_P) = \infty$. Thus, Wu and Ghosal's result implies that such a prior does not have $\mathcal{M}$ as a weak support. Proving directly that a prior does not admit $\mathcal{M}$ as a weak support would have been a much more difficult task, as opposed to showing (2.2) and appealing to the result of Wu and Ghosal [27]. Next, we show that despite violation of the weak support condition, the posterior can still be consistent at f_0.

2.2 Consistency without the Kullback-Leibler property

Here, we give a general result that avoids the Kullback-Leibler property and still gives weak consistency of the posterior. It is not specific to the context of nonincreasing densities, and hence can be exploited in other situations. The result is given in Theorem 2.1 and will be applied to the case of nonincreasing densities.

Consider now a sequence $(\theta_{1n}, \theta_{2n})_n$ such that $(F_0(\theta_{2n}))_n$ and $(F_0(\theta_{1n}))_n$ converge to 1 and 0 respectively as n goes to infinity and denote

$$\begin{aligned} f_{0,n}(x) &= \frac{f_0(x)}{F_0(\theta_{2n}) - F_0(\theta_{1n})} \mathbb{1}_{\theta_{1n} \le x \le \theta_{2n}} \quad \text{and} \\ f_n(x) &= \frac{f(x)}{F(\theta_{2n}) - F(\theta_{1n})} \mathbb{1}_{\theta_{1n} \le x \le \theta_{2n}}. \end{aligned}$$

Set

$$\begin{aligned} &S_n(\epsilon, M) \\ &= \left\{ f; \mathrm{KL}(f_{0,n}, f_n) \le \epsilon; \int f_{0,n}(x) \left(\log\left(\frac{f_0(x)}{f(x)} \right) \right)^2 dx \le M, \int_{\theta_{1n}}^{\theta_{2n}} f(x)dx \ge 1 - \epsilon \right\}. \end{aligned}$$

Then, we have the following theorem:

Theorem 2.1. *Let π be a prior probability on the set of densities on $\mathbb{R}$ satisfying: there exists $c > 0$ such that for all $\epsilon > 0$, there exists $M > 0$*

$$\liminf_n e^{cn\epsilon} \pi\left[S_n(\epsilon, M)\right] > 0, \tag{2.3}$$

with $F_0(\theta_{2n})^n \to 1$ and $(1-F_0(\theta_{1n}))^n \to 1$ for n large enough.

Then for any weak neighbourhood U of f_0 we have that

$$\pi\left[U|X^n\right] \to 1$$

in probability.

Note that the above result is weaker in the sense that the convergence is not occuring almost surely. The condition on the prior mass of $S_n(\epsilon, M)$ is of a different nature than Le Cam's condition on posterior consistency; see Ghosh and Ramamoorthi [8]. Indeed, we do not require a bound from below for the prior mass of neighbourhoods with radii ϵ/n (in L_1), but we rather prove that we need not control the ratio f_0/f everywhere in x but only on compacts. This allows to accept nonparametric prior models where the approximating functions f have support smaller than that of f_0. An almost sure convergence could be obtained by considering neighbourhoods $\int f_{0,n}(x)\left[\log\left(\frac{f_0(x)}{f(x)}\right)\right]^{2p} dx$ with $p > 1$ and an extra condition on the closeness to 1 of $F_0(\theta_{2n}) - F_0(\theta_{1n})$. However we are mainly interested in convergence in probability and obtaining an almost sure convergence would only make the presentation more cumbersome. We now prove the above theorem.

Proof of Theorem 2.1:

Note that by Theorem 4.2 of Ghosh and Ramamoorthi lemma 8.1 of Ghosal et al (2000) it is enough to prove that for all $\epsilon, \epsilon' > 0$,

$$P_0^n\left[e^{n\epsilon}D_n < \epsilon'\right] = o(1),$$

where

$$D_n = \int_{\mathcal{F}} \frac{f(X^n)}{f_0(X^n)} d\pi(f). \tag{2.4}$$

For $u > 0$, set

$$\Omega_n = \{(f, X^n); l_n(f) - l_n(f_0) > -3nu\}$$

where l_n is the log-likelihood, and

$$A_n = \{X^n, \forall i, \theta_{1n} \le X_i \le \theta_{2n}\}.$$

Then,

$$D_n \;\ge\; e^{-3nu}\int_{S_n(u,M)} 1\!\!1_{\Omega_n}(f)d\pi(f) = e^{-3nu}\pi[S_n(u, M) \cap \Omega_n].$$

Now choose $u(c+3) < \epsilon$. Since there exists $\rho > 0$ such that for n large enough $e^{cnu}\Pi\left[S_n(u,M)\right] > \rho$ we can write

$$\begin{aligned}
P_0^n\left[e^{n\epsilon}D_n < \epsilon'\right] &\leq P_0^n\left[e^{n(\epsilon-3u)}\Pi[S_n(u,M)\cap\Omega_n] < \epsilon'\right] \\
&\leq P_0^n\left[e^{n(\epsilon-(c+3)u)}\Pi[S_n(u,M)\cap\Omega_n] < \frac{\epsilon'}{\rho}\Pi\left[S_n(u,M)\right]\right] \\
&\leq P_0^n\left[\Pi[S_n(u,M)\cap\Omega_n^c] > \left(1-\frac{e^{-n(\epsilon-(c+3)u)}\epsilon'}{\rho}\right)\Pi\left[S_n(u,M)\right]\right] \\
&\leq \frac{2\int_{S_n(u,M)} P_0^n\left[\Omega_n^c(f)\right]d\Pi(f)}{\Pi\left[S_n(u,M)\right]}.
\end{aligned}$$

Moreover, for all $f \in S_n(u,M)$ we have that

$$\begin{aligned}
m_{n,u} &:= E_0^n\left[(l_n(f_0)-l_n(f))1\!\!1_{A_n}\right] \\
&= n(F_0(\theta_{2n})-F_0(\theta_{1n}))^{n-1}\int_{\theta_{1n}}^{\theta_{2n}} f_0(x)\log\left(\frac{f_0(x)}{f(x)}\right)dx \\
&= n(F_0(\theta_{2n})-F_0(\theta_{1n}))^n\left[\mathrm{KL}(f_{0n},f_n)+\log\left(\frac{F_0(\theta_{2n})-F_0(\theta_{1n})}{F(\theta_{2n})-F(\theta_{1n})}\right)\right] \\
&\leq n(F_0(\theta_{2n})-F_0(\theta_{1n}))^n\left[u-\log(1-u)\right] \\
&\leq 2nu(1+u)
\end{aligned}$$

and

$$\begin{aligned}
P_0^n\left[\Omega_n^c(f)\right] &= P_0^n\left[l_n(f)-l_n(f_0) < -3un\right] \\
&= P_0^n\left[\{l_n(f)-l_n(f_0) < -3un\}\cap A_n\right]+o(1) \\
&\leq \frac{E_0^n\left[(1\!\!1_{A_n}(l_n(f_0)-l_n(f))-m_{n,u})^2\right]}{n^2u^2(1-u)^2}+o(1).
\end{aligned}$$

Now, note that

$$\begin{aligned}
v_{n,u} &:= E_0^n\left[(1\!\!1_{A_n}(l_n(f_0)-l_n(f))-m_{n,u})^2\right] \\
&= n(F_0(\theta_{2n})-F_0(\theta_{1n}))^n\int_{\theta_{1n}}^{\theta_{2n}} f_{0n}\log^2\left(\frac{f_0(x)}{f(x)}\right)dx \\
&\quad +n(n-1)(F_0(\theta_{2n})-F_0(\theta_{1n}))^n\left(\int_{\theta_{1n}}^{\theta_{2n}} f_{0n}\log\left(\frac{f_0(x)}{f(x)}\right)dx\right)^2-m_{n,u}^2 \\
&\leq n(F_0(\theta_{2n})-F_0(\theta_{1n}))^n\int_{\theta_{1n}}^{\theta_{2n}} f_{0n}\log^2\left(\frac{f_0(x)}{f(x)}\right)dx+m_{n,u}^2((F_0(\theta_{2n})-F_0(\theta_{1n}))^{-n}-1) \\
&\leq nM+m_{n,u}^2((F_0(\theta_{2n})-F_0(\theta_{1n}))^{-n}-1).
\end{aligned}$$

We finally obtain that

$$P_0^n\left[\Omega_n^c(f)\right] \leq \frac{M}{nu^2(1-u)^2}+\frac{4(1+u)^2}{(1-u)^2}((F_0(\theta_{2n})-F_0(\theta_{1n}))^{-n}-1)=o(1)$$

and Theorem 2.1 is proved. □

2.3 Application to nonincreasing densities

In this section we apply Theorem 2.1 to the case of decreasing densities on $[0,1]$ and on $\mathbb{R}^+$. The basis for obtaining consistency of the posterior distribution is to construct a sequence of densities of the form f_P, with P a discrete distribution, that approximates f_0. The construction follows mainly the approach of Groenenboom [12], which we recall in the following section with a few changes for the sake of a better adaptation.

2.3.1 Approximative construction

The following constructive approximation is obtained on a compact interval $\mathbb{R}^+$, $[0,L]$ say. We therefore assume here that f_0 has support $[0,L]$, where L can be expressed as $L=\sup\{x\geq 0; f_0(x)>0\}$.

Let $\epsilon>0$ and construct f_ϵ in a manner similar to Groeneboom [12]: for $M\geq f_0(0)$, define $m\in\mathbb{N}^*$ such that $(1+\epsilon)^m-1=M$ (the value of M can be always adjusted such that such an m exists), and define for $i=1,...,m$

$$y_i=(1+\epsilon)^i-1,\quad \theta_i=\frac{L}{M}\left[(1+\epsilon)^i-1\right],\quad I_i=[\theta_{i-1},\theta_i),\quad l_i=\theta_i-\theta_{i-1}.$$

Following Groenboom's notation, set $\bar{f}_i=l_i^{-1}\int_{I_i}f_0(x)dx$. For all $i=1,...,m$, there exists a unique $j\in\{1,....,m\}$ such that $\bar{f}_i\in[y_{j-1},y_j)$. Now construct the stepwise function g_ϵ such that for $x\in I_i$ $g_\epsilon(x)=y_j$. Note that our approach differs from that of Groenenboom's in that we do not consider the closest value to $\bar{f}_i$ among $\{y_{j-1},y_j\}$. Then following Groenenboom [12], if $f_i=f_0(\theta_i)$ then

$$\begin{aligned}\int_0^L|f_0-g_\epsilon|(x)dx &\leq \epsilon\sum_{i=1}^m l_i(1+\bar{f}_i)+\sum_{i=1}^m l_i(f_{i-1}-f_i)\\ &\leq 2(L+1)\epsilon,\end{aligned}$$

and if $g=g_\epsilon/\int_0^L g_\epsilon(x)dx$ then

$$\int_0^L|f_0-g|(x)dx \leq 4(L+1)\epsilon. \tag{2.5}$$

We will now construct a discrete probability measure P on $[0, L]$ such that f_P is an approximation of g. Recall that g is piecewise constant, and set $g_j = g$ on I_j. Let

$$P(\theta) = \sum_{i=1}^{m} p_i \delta_{\theta_i}(\theta), \; p_m = L g_m, \quad p_i = \theta_i (g_i - g_{i+1}), \; i = m-1, ..., 1. \quad (2.6)$$

so that $f_P = g$ and $\int_0^L |f_P - f_0|(x)dx \leq 4(L+1)\epsilon$. Note that $f_P \geq g_m \geq y_1 = \epsilon$ implying $f_0/f_P \leq M/\epsilon$. Applying Lemma 8 in Ghosal and Van der Vaart [?]2007 together with Le Cam's inequality between Hellinger and the L_1 distance, we obtain that

$$\begin{aligned} \mathrm{KL}(f_0, f_P) &\leq C \int_0^L |f_0(x) - f_P(x)| dx \left(1 + \log\left(\left|\frac{f_0}{f_P}\right|_\infty\right)\right) \\ &\leq 2C'(L+1)\epsilon(1 + |\log \epsilon|). \end{aligned} \quad (2.7)$$

We now use the above construction to approximate the two families of discrete priors as described in Section 1. We first consider the case where f_0 is compactly supported on [0,1], or on some subinterval thereof.

2.3.2 Nonincreasing densities on $[0, 1]$

We have the following theorem.

Theorem 2.2. *Let f_0 be a monotone nonincreasing density on $[0, 1]$, such that $f_0(0) < \infty$. Consider a Dirichlet process prior $DP(H, \alpha)$ with $\alpha > 0$ and H be a positive probability density on $[0, 1]$ or a finite mixture model with*

$$Q(K) \geq e^{-CK\log(K)}, \quad \pi_{w,K} \geq K^{-K} c^K w_1^{a_1} ... w_K^{a_K}, \quad \theta_i \sim H \quad i.i.d,$$

for some positive constants $C, c, a_1, ..., a_K$. Then the posterior distribution is strongly consistent at f_0.

The proof is given in Appendix A. Note that consistency of the posterior distribution under a Dirichlet type of prior has already been proved by Wu and Ghosal [27] under the condition that f_0 is continuous. Hence, Theorem 2.2 extends their result to the case of nonincreasing densities f_0 admitting discontinuies. Consistency under a general finite mixture prior with unknown number of components is new, and presents a particular interest since the Kullback-Leibler support property is not satisfied in this case.

2.3.3 Nonincreasing Densities on $\mathbb{R}^+$

This section is an extension of the result of Wu and Ghosal [27] to the case where f_0 is not necessarily continuous and when the prior probability on the mixing distribution P does not have full support. In particular we consider both the Dirichlet process and the finite mixture priors; in the latter case the prior does not have full support (see Section 2.1).

Define the generalized inverse of f_0 as follows: for all $0 \leq u \leq f_0(0)$, $f_0^-(u) = \inf\{x \in \mathbb{R}^+; f_0(x) \leq u\}$, f_0^- is also nonincreasing. Since f_0 is noincreasing and integrable, recall that

$$\lim_{x \to +\infty} x f_0(x) = 0.$$

We have the following result:

Theorem 2.3. *Assume that $f_0(0) < +\infty$ and that f_0 is decreasing on $\mathbb{R}^+$. Assume also that*

$$\lim_{u \to \infty} (1 - F_0)(f_0^-(u))(\log u)^2 = 0 \tag{2.8}$$

then under the Dirichlet process prior or under finite mixture prios satisfying the same conditions as in Theorem 2.2, with $\mathbb{R}^+$ replacing $[0,1]$ and with the measure H satisfying $H(\theta) \geq \theta^{-a}$ for some $a > 0$, the posterior distribution is strongly consistent at f_0 in probability.

Condition (2.8) is in particular satisfied if $\int f_0(x)(\log f_0)^2(x)dx < +\infty$. Indeed, if f_0^- is bounded as u goes to 0, then f_0 is compactly supported and (2.8) is satisfied since $f_0^-(u)$ converges towards the upper bound of the support of f_0. Now, if we assume that f_0^- is not bounded, then integrability of $f_0(\log f_0)^2$ implies that $\int_{f_0^-(u)}^{\infty} f_0(x)(\log f_0(x))^2 dx = o(1)$. Choose $u < 1$. Then $(\log f_0(x))^2$ is increasing on $(f_0^-(u), +\infty)$ and condition (2.8) is verified.

Proof of Theorem 2.3:

The proof is based on constructing an approximation of f_0 as in Section 2.3.1, but this time on slices of $\mathbb{R}^+$. Let $\epsilon > 0$ be small and put $z_j = f_0^-(\epsilon^j)$. Assume that z_j goes to infinity with j (the case when z_j is bounded is much easier to handle). Define θ_n similarly to before, i.e. satisfying $F_n(\theta_n)^n \in (1-\epsilon, 1-\epsilon/2)$. Let J_n be the smallest j such that $z_j \geq \theta_n$. On each interval $[z_j, z_{j+1}]$, construct the function $g_j(x)$ to approximate $f_{0j}(x) = f_0(x)/\epsilon^j$ following the scheme of Section 2.3.1, for all $1 \leq j \leq J_n$. On $[0, z_1]$, consider a similar approximation scheme, associated to ϵ^2 instead of ϵ. For each $j \geq 1$

$f_{0j}(z_j) \leq 1$ and g_j is piecewise constant and the number of pieces is smaller than $m = \log(2)/\log(1+\epsilon)$. Since $\int_{z_j}^{z_{j+1}} f_{0j}(x)dx \leq z_{j+1} - z_j$, we have that

$$\int_{z_j}^{z_{j+1}} |g_j - f_{0j}| \leq 4\epsilon(z_{j+1} - z_j), \quad \text{and} \quad \int_{z_j}^{z_{j+1}} |g - f_0|(x)dx \leq 4\epsilon^{j+1}(z_{j+1} - z_j),$$

where $g(x) = g_j(x)\epsilon^j$ on (z_j, z_{j+1}). Hence,

$$\begin{aligned} \int_{z_1}^{z_{J_n}} |g - f_0|(x)dx &\leq 4\sum_{j=1}^{J_n-1} \epsilon^{j+1}(z_{j+1} - z_j) = 4\sum_{j=1}^{J_n-1} \epsilon^{j+1}(f_0^-(\epsilon^{j+1}) - f_0^-(\epsilon^j)) \\ &\leq 4\sum_{j=1}^{J_n-1} \int_{z_j}^{z_{j+1}} f_0(x)dx \leq 4\int_{z_1}^{\infty} f_0(x)dx = o(1) \end{aligned}$$

and by construction $f_0/g = f_{0j}/g_j \leq 1/\epsilon$ on $[z_j, z_{j+1}]$. Note that the construction of g on $[0, z_1]$ implies also that

$$\int_0^{z_1} |g - f_0|(x)dx \leq 4\epsilon^2 z_1,$$

where g is piecewise constant and the number of pieces on $[0, z_1]$ is bounded by $m_1 \leq \log(M+1)/\log(1+\epsilon^2)$. Since $xf_0(x)$ goes to zero as x goes to ∞, if ϵ is small enough, $z_1 \leq 1/(2\epsilon)$ and

$$\int_0^{z_1} |g - f_0|(x)dx \leq 2\epsilon$$

and we can normalize g such that the above properties remain valid. Note that $f_0/g(x) \leq M/\epsilon^2$ and define $f_{0n} = f_0 1\!\!1_{[0,\theta_n]}/F_0(\theta_n)$, then

$$\begin{aligned} \mathrm{KL}(f_{0n}, g) &\leq C\left(\epsilon + \int_{f_0^-(\epsilon)}^{\infty} f_0(x)dx\right)(1 + |\log\epsilon|) \\ \int_0^{\theta_n} f_0(x)\left(\log\left(\frac{f_0}{g}(x)\right)\right)^2 dx &\leq C\left(\epsilon + \int_{f_0^-(\epsilon)}^{\infty} f_0(x)dx\right)(1 + |\log\epsilon|)^2 \end{aligned}$$

Condition (2.8) implies that the right hand term of the first inequality above goes to zero and the second is bounded, when ϵ goes to 0. Moreover, condition (2.8) implies that we can define

$$u_n := \log(\epsilon^{J_n-1})^2(1 - F_0)(z_{J_n-1}) = o(1)$$

and by definition of J_n, $1 - F(z_{J_n-1}) \geq 1 - F_0(\theta_n) \geq c_1\epsilon/n$ so that,

$$(J_n - 1)^2 \log^2(\epsilon) \leq u_n(1 - F_0(\theta_n))^{-1} \leq u_n n/(c_1\epsilon).$$

Therefore, there exists $C > 0$ such that for $\epsilon > 0$ small enough

$$J_n - 1 \leq C u_n^{1/2} \epsilon^{-1/2} n^{1/2}.$$

Similarly to the construction on $[0,1]$, let P be such that $f_P = g$, i.e. $P(\theta) = \sum_{j=0}^{J_n} \sum_{i=1}^{m} w_{j,i} \delta_{\theta_{j,i}}(\theta)$ and the number of components is bounded by $J_n m_1 \leq u_n n^{1/2} \epsilon^{-5/2}$. Define $\mathcal{W}$ the set of P' satisfying the same conditions as in the proof of Theorem 2.2, i.e. if $U_{j,i} = (\theta_{j,i} - \epsilon^3, \theta_{j,i} + \epsilon^3)$, $j = 0, ..., J_n - 1$ and $i = 1, ..., m_1 - 1$ and $U_{j,m} = (z_{j+1}, z_{j+1} + \epsilon^3)$

$$|P'(U_{j,i}) - w_{j,i}| \leq \frac{\epsilon^3}{m_1 J_n}, \quad \text{with } U_0 = [0, \epsilon^{J_n}] \cap (\cup_i U_i)^c,$$

then $f_{0n}/f_{P'n} \leq 1/\epsilon$, by construction $\int_{\theta_n}^{\infty} f_{P'}(x)dx \leq \epsilon$ and

$$\begin{aligned}
\mathrm{KL}(f_{0n}, f_{P'}) &\leq C\left(|f_0 - f_P| + |f_{P'} - f_P|\right)(1 + |\log \epsilon|), \\
&\leq C\left(3\epsilon + \int_{f_0^-(\epsilon)}^{\infty} f_0(x)dx\right)(1 + |\log \epsilon|), \\
\int_0^{\theta_n} f_0(x)\left(\log\left(\frac{f_0}{g}(x)\right)\right)^2 dx &\leq C\left(3\epsilon + \int_{f_0^-(\epsilon)}^{\infty} f_0(x)dx\right)(1 + |\log \epsilon|)^2.
\end{aligned}$$

Thus, for any $\epsilon' > 0$, by choosing $\epsilon > 0$ small enough, $\mathcal{W} \subset S_n(\epsilon', 1)$. In the case of a Dirichlet process prior, Lemma 10 of [11] implies that the prior probability of $\mathcal{W}$ is bounded from below by

$$\begin{aligned}
&\mathrm{P}\left[\mathcal{D}(AH(U_0), ..., AH(U_{mJ_n})) \in (w_{j,i} \pm \epsilon^3/mJ_n, j = 0, ..., J_n - 1, i = 1, ..., m)\right] \\
\geq\ &\exp\left(c\epsilon^{-1}(\epsilon^{-1}\log(H(z_1)) + \sum_{j=1}^{J_n} \log(H(z_j))\right)
\end{aligned}$$

Using a similar argument as in the case z_1, $z_j \leq \epsilon^{-j}$ and $\log(H(z_j)) \geq aj\log(\epsilon)$, therefore

$$\begin{aligned}
&\mathrm{P}\left[\mathcal{D}(AH(U_0), ..., AH(U_{mJ_n})) \in (w_{j,i} \pm \epsilon^3/mJ_n, j = 0, ..., J_n - 1, i = 1, ..., m)\right] \\
\geq\ &\exp\left(-ca\epsilon^{-2}|\log \epsilon| J_n^2 \log n\right)
\end{aligned}$$

Since $J_n = o(\sqrt{n})$ condition (2.3) is verified. The same types of computations are applied to the finite model case, so that in both cases condition (2.3) is verified and Theorem 2.3 is proved. $\square$.

3 Simulation study

We now present some simulations, in the case of a Dirichlet mixture. Our prior on $\mathcal{F}$ is defined by: $P \sim DP(H, \alpha)$, where H is base measure with $\alpha > 0$ as scale parameter. The use of an inverse Gamma distribution allows to have explicit full conditional distributions. Recall that the Dirchlet process mixture model can be express using the following hierarchical representation:

$$\begin{aligned} X_i|\theta_i &\sim \mathcal{U}_{[0,\theta_i]}, \quad \text{independently} && (3.1)\\ \theta_i|P &\sim P = IG(a,b) \quad \text{i.i.d} && (3.2)\\ P &\sim DP(H,\alpha). && (3.3)\end{aligned}$$

Using the Sethuraman [24] representation of the Dirichlet process,

$$P = \sum_{j=1}^{\infty} p_j \delta_{Z_j}, \quad \text{where} \quad Z_j \sim H, \quad j = 1, 2, ... \tag{3.4}$$

and

$$p_1 = V_1, p_j = V_j \prod_{i<j}(1 - V_i), \quad V_j \sim Beta(1, \alpha) \quad \text{i.i.d.}$$

we use the retrospective MCMC sampling algorithm proposed by Papaspillopoulos and Roberts [18] . When considering the model in (3.1) and (3.4), there is a number of quantities of which we may want to provide posterior inference. These include the allocation variables $K = (K_1, ..., K_n)$, defined by $K_i = j$ if and only if $\theta_i = Z_j$, the number of clusters in the population, the weights p_j and the density f_P. In our case, we are primarily interesed in f_P.

We now introduce some notations. Define $s_j = \sum_{i=1}^{n} 1_{\{K_i=j\}}$ for $j = 1, 2, ...$ the number of observations in the jth class. Furthermore, let

$$J^{(a)} = \{j \in \mathbb{N} : s_j > 0\}, \quad J^{(d)} = \{j \in \mathbb{N} : s_j = 0\} = \mathbb{N} \setminus J^{(a)}$$

so that $J^{(a)}$ and $J^{(d)}$ are the sets of all "alive" and "dead" components respectively.

From Proposition 1 of Papaspillopoulos and Roberts [18], we have that Z and V are independent conditionally on X and K with

$$Z_j|X^n, K \sim \begin{cases} \frac{b^a}{\Gamma(a)}(\frac{1}{Z_j})^{a+1} e^{-\frac{b}{Z_j}}, & \text{for} \quad j \in J^{(d)} \\ \prod_{\{i:K_i=j\}} \frac{b^a}{\Gamma(a)}(\frac{1}{Z_j})^{a+2} e^{-\frac{b}{Z_j}}, & \text{for} \quad j \in J^{(a)}. \end{cases} \tag{3.5}$$

Conditionally on K, the random variable V is independent of (X, Z) and its conditional distribution is given by

$$V_j|K \sim Beta(s_j + 1, n - \sum_{i=1}^{j} s_j + \alpha) \quad \text{for all } j = 1, 2, ... \tag{3.6}$$

Also, conditionally on (X^n, V, Z) we have that

$$P(K_i = j|X^n, V, Z) \propto p_j \frac{1\!\!1_{X_i \leq Z_j}}{Z_j} \quad \text{for all } j = 1, 2, ... \tag{3.7}$$

where p_j is defined by (3.4).

We obtain the conditional distribution of (V, Z) given K and X. Hence, the (Z_j, V_j) are independent and for each j their conditional distribution is given by

$$P(V_j, Z_j|K, X^n) \tag{3.8}$$
$$= \begin{cases} \frac{b^a}{\Gamma(a)}(\frac{1}{Z_j})^{a+1} e^{-\frac{b}{Z_j}} \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)}(1-v)^(1-\alpha), & \text{for} \quad j \in J^{(d)} \\ \prod_{\{i:K_i=j\}} \frac{b^a}{\Gamma(a)}(\frac{1}{Z_j})^{a+2} e^{-\frac{b}{Z_j}} Beta(s_j + 1, n - \sum_{i=1}^{j} s_j + \alpha), & \text{for} \quad j \in J^{(a)}. \end{cases}$$

Note that given a realisation $(V_j, Z_j), j \in \mathbb{N}$ from the posterior distribution we can compute the corresponding $f_P(x)$ from:

$$\begin{aligned} f_P(x) &= \sum_{j=1}^{max\{K\}} \frac{p_j}{Z_j} \mathbb{I}_{(x \leqslant Z_j)} + \sum_{j=max\{K\}+1}^{\infty} \frac{p_j}{Z_j} \mathbb{I}_{(x \leqslant Z_j)} \\ &\overset{d}{=} \sum_{j=1}^{max\{K\}} \frac{p_j}{Z_j} \mathbb{I}_{(x \leqslant Z_j)} + \tilde{f}_P(x) \prod_{j=1}^{max\{K\}} (1 - V_j) \end{aligned} \tag{3.9}$$

where $\tilde{f_P}$ is sampled from the prior, using Guglielmi and Tweedie [13] and where the second equality is an equality in distribution.

Our aim is to compute the posterior mean $\widehat{f}(x) = E^{\Pi}(f_P(x)|X^n)$ as a Bayesian estimate of f_P, and evaluate the corresponding loss $L1$, $d_1(\widehat{f}, f_0) = \int |\widehat{f}(x) - f_0|(x)dx$, and posterior risks $E^{\Pi}[d(f_P, f_0)|X^n]$.
To this end, let us consider the grid on $\mathbb{R}^+$ defined by $G_{\mathbb{R}^+} = \{F_0^{-1}(\frac{g}{G});\ g = 1, ..., G\}$, where G is a large integer. For a given density f computed on $G_{\mathbb{R}^+}$ we approximate $d(f, f_0)$ by

$$\widehat{d}(f, f_0) = \frac{1}{G} \sum_{g=1}^{G} \left| \frac{f(x_g)}{f_0(x_g)} - 1 \right|.$$

Algorithm
Initialisation of K^0, Z^0 **and** V^0
for $(t \in \{0, ..., T-1\})$ **do**
  Generate (Z^t, V^t) **given** X^n **and** K^t **from (??) and (??) respectively**
  Generate K^t **given** (X^n, Z^t, V^t) **from (3.6)**
  Generate $\tilde{f}_P(x_g)$ **from the prior**
  Generate $(f^t(x_g))_{g=1}^G$ **using (3.8)**
end for
Compute the estimator $\frac{1}{T}\sum_{t=1}^T f^t(x_g)$**,** $g = 1, ..., G$
Compute the estimated loss $\widehat{d}(f^t, f_0)$ **end.**

Table 1: Pseudo-code of retrospective MCMC algorithm to compute $\widehat{f}$.

The algorithm thus becomes: For large $T > 0$ the posterior mean, $\widehat{f}$, is approximated by

$$\tilde{f}(x) = \frac{1}{T}\sum_{t=1}^{T} f_{P^t}(x), \quad x \in G_{\mathbb{R}+},$$

the distance between $\hat{f}$ and f_0 is approximated by $\widehat{d}(\widehat{f}, f_0) = \widehat{d}(\tilde{f}, f_0)$ and the posterior risk by

$$\widehat{\mathbb{E}^{\pi}}[d(f, f_0)|X^n] = \frac{1}{T}\sum_{t=1}^{T} \widehat{d}(f^t, f_0).$$

We have designed our simulation study with $f_0(x) = e^{-x}$.

In the following table, we give estimates of the frequentist expectation of the posterior risk $E_0[\mathbb{E}^{\pi}[d(f_0, f_P)|X^n]]$ under different sample sizes to illustrate the convergence of the posterior distribution. Following Gadja *et al.* [17], we have use importance sampling approximations to compute the above expectation. More precisely for a given sample size n we have simulated M i.i.d samples of size n distributed according to f_0, say $x_{(1)}, ..., x_{(M)}$, where $x_{(j)} = (x_{1,(j)}, ..., x_{n,(j)})$. We have run the above MCMC algorithm to compute the posterior distribution given $x_{(1)}$, $P^{\Pi}[.|x_{(1)}]$, we have then approximated $E^{\Pi}[d(f, f_0)|x_{(j)}]$ for $j = 2, ..., M$ by

$$\frac{\sum_{t=1}^{T} \widehat{d}(f^t, f_0) w(t, x_{(j)}, x_{(1)})}{\sum_{t=1}^{T} w(t, x_{(j)}, x_{(1)})}, \quad w(t, x_{(j)}, x_{(1)}) = \frac{\prod_{i=1}^{n} f^t(x_{i,(j)})}{\prod_{i=1}^{n} f^t(x_{i,(1)})}.$$

Figure 1: posterior means and 95% credible intervals of the density. Left: $n = 100$, right: $n = 1000$

A definite decrease can be observed, at a rate which seems to be slightly slower than $n^{-1/3}$.

Figure 1 shows the posterior mean estimates of the density, for two sample sizes : $n = 100$ and $n = 1000$, together with the true density (f_0, in dotted lines) and the pointwise 95 % credible intervals. The improvement due to the increase of n is quite significant.

Number of observation	$n = 100$	$n = 500$	$n = 1000$	$n = 5000$
Posterior mean	0.3530083	0.1944597	0.1246980	0.1027860

Table 2: Estimation of posterior expectation of L_1-distance between f_P and f_0.

Acknowledgements

This work was supported by the 2007–2010 ANR-07-BLAN-0237-01 “SP Bayes" grant. This work is part of the first author’s PhD thesis.

A Proof of Theorem 2.2

To prove Theorem 2.2, we use Theorem 2.1. Let $\epsilon > 0$ and $\theta_n = 1 - \frac{\epsilon}{2nf_0(0)}$ then $1 - F_0(\theta_n) \leq f_0(\theta_n)\epsilon/(2nf_0(0)) \leq \epsilon/(2n)$ and $F_0(\theta_n)^n \geq 1 - \epsilon$ for n large enough. We now construct f_P such that

$$\mathrm{KL}(f_{0n}, f_{Pn}) \leq \epsilon'$$

for $\epsilon' > 0$, where f_{0n} and f_{Pn} are the restrictions of f_0 and f_P respectively on $[0, \theta_n]$. Consider the construction of Section 2.3.1, with $L = \theta_n$ then applying (2.7) we obtain that

$$\mathrm{KL}(f_{0n}, f_{Pn}) \leq C'\epsilon(L+1)(1 + |\log \epsilon|) \leq 2C'\epsilon(1 + |\log \epsilon|)$$

choosing ϵ small enough implies that $\mathrm{KL}(f_{0n}, f_{Pn}) \leq \epsilon'$. Moreover, using Lemma 8 of Ghosal and Van der Vaart [11],

$$\int f_{0n} \log^2 (f_{0n}/f_{Pn}) \leq C''\epsilon(1 + |\log \epsilon|)^2 \leq 1$$

if ϵ is chosen small enough. We also have that by construction f_P has support $[0, \theta_n]$ so that $\int_0^{\theta_n} f_P(u)du = 1$.

Let P' be the mixing distribution associated with $\{m, \theta'_1, ..., \theta'_{m-1}, w'_1, w'_2, ..., w'_m\}$ with $\sum_{j=1}^{m-1} w'_j = 1$. Recall that $|\theta_j - \theta_{j+1}| \geq \epsilon/(2M)$. Choose $0 < \delta < 1$ and define $U_j = (\theta_j - \epsilon^3, \theta_j + \epsilon^3)$ if $j \leq m-1$ and $U_m = (\theta_n, \theta_n + \epsilon(1-\theta_n) \wedge \epsilon^3)$. We construct P' such that: $\theta'_j \in U_j$ and $|w_j - w'_j| \leq \epsilon^2/m$. Then $w'_m \geq \epsilon/2$ and

$$\frac{f_{0n}(x)}{f_{P'n}(x)} \leq \frac{2f_0(0)}{\epsilon} \leq \frac{2M}{\epsilon}.$$

By definition, if ϵ is small enough $U_j \cap U_i = \emptyset$ if $i \neq j$ and $|\theta'_j - \theta_j| \geq \epsilon\theta_j$. Thus there exists $C_1 > 0$ such that

$$\begin{aligned}
&\int_0^{\theta_n} f_0 \log\left(\frac{f_0}{f_{P'}}\right)(x)dx \\
&\leq C_1\left(\epsilon + |f_P - f_{P'}|_1\right)(1 + |\log \epsilon|) \\
&\leq C_1\left(\epsilon + \left[\max_{j\leq m}\left|\sum_{i\geq j}\left(\frac{w'_i}{\theta'_i} - \frac{w_i}{\theta_i}\right)\right| + (1+\epsilon)\sum_{j=1}^{m-1}\frac{w_j}{\theta_j}\,|\theta_j - \theta'_j|\right]\right)(1 + |\log \epsilon|) \\
&\leq C_1\left(2\epsilon + (1+\epsilon)\epsilon^2\right)(1 + |\log \epsilon|).
\end{aligned}$$

Generally speaking, denote by $U_0 = [0,1] \cap \left(\cup_{j=1}^m U_j\right)^c$ and by $\mathcal{W} = \{P'; |P'(U_j) - w_j| \leq \epsilon^2/m, j = 1, ..., m\}$, if $P' \in \mathcal{W}$, we also obtain

$$\int_0^{\theta_n} f_0 \log\left(\frac{f_0}{f_{P'}}\right)(x)dx \leq C_1\left(2\epsilon + (1+\epsilon)\epsilon^2\right)(1 + |\log \epsilon|),$$

and similarly

$$\int_0^{\theta_n} f_0 \left(\log\left(\frac{f_0}{f_{P'}}\right)\right)^2(x)dx \leq C'_1\left(2\epsilon + (1+\epsilon)\epsilon^2\right)(1 + |\log \epsilon|)^2 \leq 1$$

if ϵ is small enough. Note also that for all $P' \in \mathcal{W}$,

$$\int_{\theta_n}^1 f_{P'}(x)dx \leq f_{P'}(\theta_n)(1-\theta_n) \leq \frac{\epsilon}{n}.$$

For all $\epsilon' > 0$ there exists $\epsilon > 0$ such that for all n large enough

$$\mathcal{W} = \{P'; |P'(U_j) - w_j| \leq \epsilon^2/m, j = 1, ..., m\} \subset S_n(\epsilon', M).$$

In the case of the Dirichlet process prior

$$\begin{aligned}
\Pi\left[\mathcal{W}\right] &= \Pr\left[\mathcal{D}(\alpha H(U_0), ..., \alpha H(U_m)) \in (w_j \pm \epsilon^2/m, j = 0, ..., m)\right], \quad w_0 = 0 \\
&\geq c_\epsilon(1 - \theta_n) \\
&\geq c_\epsilon \epsilon/(2n),
\end{aligned}$$

for some $c_\epsilon > 0$, which achieves the proof of the consistency of the posterior in the case of a Dirichlet prior. In the case of a finite mixture prior, we write

$$\mathcal{W} = \{P'(\theta) = \sum_{j=1}^{m} w'_j \delta_{\theta'_j}(\theta), |w'_j - w_j| \leq \epsilon^2/m, |\theta'_j - \theta_j| \leq \epsilon^3\}$$

and similarly to before,

$$\Pi(\mathcal{W}) \geq c'_\epsilon(1 - \theta_n), \quad c'_\epsilon > 0$$

so that Theorem 2.2 is proved. □

References

[1] ALSTON, C. LOW CHOY, S. AND KUHNERT, P., MCVINISH, R. (2005). Bayesian Model Comparison: Review and Discussion. *International Statistical Institute,* 55th session.

[2] BARRON, A., SCHERVISH, M. J. and WASSERMAN, L. (1999). The consistency of posterior distributions in nonparametric problems. *Ann. statist.,* **27** 536-561.

[3] BALABDAOUI, F. and WELLNER, J. A. (2007). Estimation of a k-monotone density: limit distribution theory and the Spline connection, *Ann. of Statist.* **35**, 2536-2564.

[4] BALABDAOUI, F. and WELLNER, J. A. (2010). Estimation of a k-monotone density: characterizations, consistency and minimax lower bounds, *Statistica Neerlandica* **64**, 45-70.

[5] DIACONIS, P. and FREEDMAN, D. (1986). On inconsistent bayes estimates of location . *Ann. Statist.,* **14** 67-87.

[6] FERGUSSON, T. S. (1983). Bayesian density estimation by mixtures of normal distributions. *Recent Advances in statistics* (Razvi, M. Rustagi, J. and Siegmund, D. Eds.) 287-302. Academic Press, New York.

[7] GNEITING, T. (1999). Radial positive definite functions generated by Eu- clid's hat. *J. Multivariate Analysis* **69**, 88 - 119.

[8] GHOSH, J. K. AND RAMAMROOTHI R. V. (2003). *Bayesian Nonparametrics.* Springer, New York.

[9] GHOSAL, S., GHOSH, J. K. and RAMAMOORTHI, R.V. (1999). Posterior consistency of dirichlet mixtures in density estimation. *Ann. Statist.,* **27** 143-158.

[10] GHOSAL, S. and VAN DER VAART, A. (2001). Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Ann. Statist.* **29** 1233-1263.

[11] GHOSAL, S. and VAN DER VAART, A. (2007) Convergence rates of posterior distributions for noniid observations. *Ann. of Statist.* **35** 192–223

[12] GRONENBOOM, P. (1986). *Some current Development in Density Estimation.* In Mathematics and Computer Science, CWI Monograph 1 (Eds. J. W. de Bakker, M. Hazewinkel, J. K. Lenstra), Elsevier, Amsterdam (1986), 163-192.

[13] Guglielmi A, Tweedie R. L. (2001). Markov chain Monte Carlo estimation of the law of the mean of a Dirichlet process. *Bernoulli* **7**, 573-592.

[14] Lévy, P. (1962). Extensions d'un théorème de D. Dugué et M. Girault. Z. *Wahrscheinlichkeitstheorie* **1**, 159 - 173.

[15] Lo, A.Y. (1984). On a class of Bayesian nonparametric estimates I: density estimates. *Ann. Statist.,* **1** 38-58.

[16] McVinish, R., Rousseau J., and Mengerson K. (2008): Bayesian Goodness-of-Fit Testing with Mixtures of Triangular Distributions. *Scandinavian journal of statistics*

[17] D. Gadja, C. Guihenneuc, J. Rousseau, K. Mengersen and D. Nur (2010): Use in practice of importance sampling for repeated MCMC for Poisson models. *Elect. Journ. Statist.* To appear.

[18] Papaspiliopoulos, O. and Roberts G.O. (2008). Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models. *Biometrika* **95** 169-186.

[19] Petrone, S. and Wasserman, L.(2002). Consistency of Bernstein polynomial posteriors. *J. Roy. Statist. Soc., B* **64** 79-100.

[20] Perron F. and Mengersen, K. (2001). Bayesian Nonparametric Modelling using Mixtures of Triangular Distributions, *Biometrics*, **57**, 2, 518-529.

[21] Robert, C., Rousseau, J. (2003). Bayesian Goodness of Fit -A Mixture Approach to Bayesian Goodness of Fit. (revised version).

[22] Rousseau, J. (2010). Rates of convergence for the posterior distributions of mixture of Betas and adaptive nonparametric estimation of the densities. *Ann. Statist.*. **38**, p.146-180.

[23] Schwartz, L. (1965). On Bayes Procedures. *Z. Wahrscheinlichkeitstheorie. and Verw. Gebiete.* **4** 10-26.

[24] Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* **4** 639-650.

[25] Walker, S. G., Lijoi, A. and Pruenster, I. (2004). Contributions to the understanding of Bayesian consistency. *preprint*.

[26] Williamson, R. E. (1956). Multiply monotone functions and their Laplace transforms. *Duke Math.J* **23**, 189 - 207.

[27] Wu, Y. Ghosal, S. (2008). Kullback Leibler property of kernel mixture priors in Bayesian density estimation. *Electronic Journal of Statistics*, **2** 298-331.

[28] Scricciolo, C. (2001) Convergence rates of posterior distributions for Dirichlet mixtures of normal densities. Working paper 2001.21 Univ. Padova