TAILS ASSUMPTIONS AND POSTERIOR CONCENTRATION RATES FOR MIXTURES OF GAUSSIANS

ZACHARIE NAULET AND JUDITH ROUSSEAU

CEREMADE, Université Paris-Dauphine, France

Abstract. Nowadays in density estimation, posterior rates of convergence for location and location-scale mixtures of Gaussians are only known under light-tail assumptions; with better rates achieved by location mixtures. It is conjectured, but not proved, that the situation should be reversed under heavy tails assumptions. The conjecture is based on the feeling that there is no need to achieve a good order of approximation in regions with few data (say, in the tails), favoring location-scale mixtures which allow for spatially varying order of approximation. Here we test the previous argument on the Gaussian errors mean regression model with random design, for which the light tail assumption is not required for proofs. Although we cannot invalidate the conjecture due to the lack of lower bound, we find that even with heavy tails assumptions, location-scale mixtures apparently perform always worst than location mixtures. However, the proofs suggest to introduce hybrid location-scale mixtures that are find to outperform both location and location-scale mixtures, whatever the nature of the tails. Finally, we show that all tails assumptions can be released at the price of making the prior distribution covariate dependent.

1. Introduction

Nonparametric mixture models are highly popular in the Bayesian nonparametric literature, due to both their proven flexibility and relative easiness of implementation, see Hjort et al. (2010) for a review. They have been used in particular for density estimation, clustering and classification and recently nonparametric mixture models have also been proposed in nonlinear regression models, see for instance de Jonge and van Zanten (2010); Wolpert et al. (2011); Naulet and Barat (2015).

There is now a large literature on posterior concentration rates for nonparametric mixture models, initiated by Ghosal and Van Der Vaart (2001); Ghosal et al. (2007a) and improved by Kruijer et al. (2010); Shen et al. (2013); Scricciolo (2014) in the context of location mixtures of Gaussian distributions and studied by Canale and De Blasi (2013) in the context of location-scale Gaussian distributions and de Jonge and van Zanten (2010) in the case of location mixture models for nonlinear regression.

E-mail address: zacharie.naulet@dauphine.eu; rousseau@ceremade.dauphine.fr.
Date: June 24, 2018.
Location mixture of Gaussian densities can be written as
\[(1)\quad f_{\sigma,G}(x) = \int_{\mathbb{R}} \varphi_\sigma(x - \mu) dG(\mu),\]
while location-scale mixtures have the form
\[(2)\quad f_G(x) = \int_{\mathbb{R} \times \mathbb{R}^+} \varphi_\sigma(x - \mu) dG(\mu, \sigma).\]

These models are used in the Bayesian nonparametric literature to model smooth curves, typically probability densities, by putting a prior on the mixing distribution \(G\) (and on \(\sigma\) for location mixtures \((1)\)). The most popular prior distributions on \(G\) are either finite with unknown number of components, as in Kruijer et al. (2010) and the reknown Dirichlet Process (Ferguson (1973)) or some of its extensions. In both cases \(G\) is discrete almost surely.

In Kruijer et al. (2010) and later on in Shen et al. (2013); Scricciolo (2014) it was proved that location mixture of Gaussian distributions lead to adaptive (nearly) optimal posterior concentration rates (for \(L_1\) metrics) over collections of Hölder types functional classes, in the context of density estimation for independently and identically distributed random variables. Contrarily, in Canale and De Blasi (2013), suboptimal posterior concentration rates are derived and the authors obtain rates that are at best \(n^{-\beta/(2\beta+2)}\) up to a \(\log n\) term in place of \(n^{-\beta/(2\beta+1)}\). These results are obtained under strong assumptions on the tail of the true density \(f_0\), since it is assumed that \(f_0(x) \lesssim e^{-c|x|^\tau}\) when \(x\) goes to infinity, for some positive \(c, \tau\).

In Canale and De Blasi (2013), the authors suggest that location-scale mixtures might lead to suboptimal posterior concentration rates, for light tail distributions but might be more robust to tails, since the rate \(n^{-\beta/(2\beta+2)}\) is the minimax estimation rate for density estimation with regularity \(\beta\), under the \(L_2\) loss, see Reynaud-Bouret et al. (2011); Goldenshluger and Lepski (2014).

The question thus remains open as to how robust to tails mixtures of Gaussian distributions (either location or location-scale) are.

Interestingly in Bochkina and Rousseau (2016), much weaker tail constraints are necessary to achieve the minimax rate \(n^{-\beta/(2\beta+1)}\), for estimating densities on \(\mathbb{R}^+\) using mixtures of Gamma distributions. The authors merely require that \(F_0\) allows for a moment of order strictly greater than 2. However in Bochkina and Rousseau (2016) as well as in Kruijer et al. (2010); Shen et al. (2013); Scricciolo (2014), the smoothness functional classes are non standard and roughly correspond to requiring that the log-density is locally Hölder, which blurs the understanding of the robustness of Gaussian mixtures to tails. These smoothness conditions are required to ensure that the density \(f_0\) can be approximated by a mixture \(f_{\sigma,G}\) where \(G\) is a probability measure in terms of Kullback-divergence. Hence to better understand the ability of mixture models to capture heavy tails we study their use in nonparametric regression models:
\[(3)\quad Y_i = f(X_i) + \epsilon_i, \quad \epsilon_i \overset{i.i.d.}{\sim} N(0, s^2), \quad i = 1, \ldots, n, \quad X_1, \ldots, X_n \overset{i.i.d.}{\sim} Q_0, \quad f \in L^2(Q_0).\]

The parameter is \(f\) with prior distribution denoted by \(\Pi\). We assume that \(s\) is known, which is just a matter of convenience for proofs. All the results of
the paper can be translated to the case \( s \) unknown using the same methodology as Salomond (2013) or Naulet and Barat (2015). Our aim is to study posterior concentration rates in \( L^2(Q_0) \) around the true regression function \( f_0 \) defined by sequences \( \epsilon_n \) converging to zero with \( n \) and such that
\[
\Pi (d_n(f, f_0) \leq \epsilon_n \mid y^n, x^n) = 1 + o_p(1),
\]
under the model \( f_0 \), where \( d_n \) is the empirical \( \ell_2 \) distance of the covariates, defined as \( d_n(f, f_0)^2 := n^{-1} \sum_{i=1}^{n} |f(x_i) - f_0(x_i)|^2 \). By analogy to the case of density estimation of Reynaud-Bouret et al. (2011) and Goldenshluger and Lepski (2014) we assume that \( f_0 \in L^1 \) and belongs to a Hölder ball with smoothness \( \beta \). The tail condition are then on the design distribution and written as \( \int_{\mathbb{R}} |x|^p dQ_0(x) < \infty \) and our aim is to study the posterior concentration rate (4) for both location and location-scale mixtures.

We show in section 2, that in most cases location mixtures have a better posterior concentration rate than location-scale mixtures and unless \( p \) goes to infinity the posterior concentration rates is not as good as the usual \( n^{-\beta/(2\beta+1)} \). This rate is suboptimal for light tail design points, since in this case the minimax posterior concentration rate is given by \( n^{-\beta/(2\beta+1)} \). To improve on this rate we propose a new version of location-scale mixture models, which we call the hybrid location-scale mixture and we show that this nonparametric mixture model leads to better posterior concentration rates than the location mixture (and thus than the location-scale mixture). All these results are up to \( \log n \) terms. The results are summarized in table 1 which displays the value \( q \) defined by \( \epsilon_n^2 = n^{-q} \).

### Table 1. Summary of posterior rates of convergence for different types of mixtures.

|  | \( 0 < p < 2 \) | \( p \geq 2 \) |
|---|---|---|
| Location | \( p \leq 2\beta/(\beta + 1) \) | \( p \geq 2\beta/(\beta + 1) \) | \( p \leq 2\beta \) | \( p \geq 2\beta \) |
| Location-scale | \( 2\beta/\beta + 1 / 2\beta \) | \( 2\beta/3\beta + 1 / 2\beta \) | \( 2\beta/2\beta + 1 / 2\beta \) | \( 2\beta/2\beta + 1 + 2\beta/p \) |
| Hybrid | \( 2\beta/3\beta + 2 / 2\beta \) | \( 2\beta + 1 + 2\beta/p / p \) | \( 2\beta + 1 + 2\beta/p / p + 1 \) | \( 2\beta + 1 + 2\beta/p / 2\beta + 1 \) |

Although the results are presented in the regression model, we believe that similar phenomena should take place in the density estimation problem.

The main results with the description of the three types of prior models and the associated posterior concentration rates are presented in section 2. Proofs are presented in section 3 and some technical lemmas are proved in the appendix.

#### 1.1. Notations.

We call \( P_f(\cdot \mid X) \) the distribution of the random variable \( Y \mid X \) under the model (3), associated with the regression function \( f \). Given \( (X_1, \ldots, X_n) \), \( P_{f^n}(\cdot \mid X_1, \ldots, X_n) \) stands for the distribution of the random vector \( (Y_1, \ldots, Y_n) \) of independent random variables \( Y_j \sim P_f(\cdot \mid X_j) \). Also, for
any random variable \( X \) with distribution \( P \), and any function \( g \), \( Pg(X) \) denote the expectation of \( g(X) \).

For any \( \alpha > 0 \), we let \( \text{SGa}(\alpha) \) denote the symmetric Gamma distribution with parameter \( \alpha \); that is \( X \sim \text{SGa}(\alpha) \) has the distribution of the difference of two independent Gamma random variables with parameters \((\alpha, 1)\).

For any finite positive measure \( \alpha \) on the measurable space \((X, \mathcal{X})\), let \( \Pi_\alpha \) denote the symmetric Gamma process distribution with parameter \( \alpha \) (Wolpert et al., 2011; Naulet and Barat, 2015); that is, an \( M \sim \Pi_\alpha \) is a random signed measure on \((X, \mathcal{X})\) such that for any disjoint \( B_1, \ldots, B_k \in \mathcal{X} \), the random variables \( M(B_1), \ldots, M(B_k) \) are independent with distributions \( \text{SGa}(\alpha(B_i)) \), \( i = 1, \ldots, k \).

For any \( \beta > 0 \), we let \( \mathcal{C}^\beta \) denote the H"older space of order \( \beta \); that is the set of all functions \( f : \mathbb{R} \to \mathbb{R} \) that have bounded derivatives up to order \( m \), the largest integer smaller than \( \beta \), and such that the norm \( \|f\|_{C^\beta} := \sup_{0 < |x| < 1} \sup_{0 < |y| < 1} |f^{(k)}(x)| + \sup_{0 < |x - y| < 1} |f^{(m)}(x) - f^{(m)}(y)|/|x - y|^{\beta - m} \) is finite.

For \( 1 \leq p < \infty \) we let \( L^p \) be the space of function for which the norm \( \|f\|_p := \int |f(x)|^p \, dx \) is finite; and by \( L^\infty \) we mean the space of functions for which \( \|f\|_\infty := \sup_{x \in \mathbb{R}} |f(x)| \) is finite. For \( 0 \leq p, q \leq \infty \) and functions \( f \in L^p, g \in L^q \), we write \( f * g \) the convolution of \( f \) and \( g \), that is \( f * g(x) := \int f(x - y)g(y) \, dy \) for all \( x \in \mathbb{R} \). Moreover, we’ll use repeatedly Young’s inequality which state that \( \|f * g\|_r \leq \|f\|_p \|g\|_q \), with \( 1/p + 1/q = 1/r + 1 \).

If \( f \in L^1 \), then we define \( \hat{f} \) as the \((L^1)\) Fourier transform of \( f \); that is \( \hat{f}(\xi) := \int f(x)e^{-ix\xi} \, dx \) for all \( \xi \in \mathbb{R} \). Moreover, if \( \hat{f} \in L^1 \), then the inverse Fourier transform is well-defined and \( f(x) = (2\pi)^{-1} \int \hat{f}(\xi)e^{ix\xi} \, d\xi \). Also, we denote by \( \mathcal{S} \) the Schwartz space; that is the space of infinitely differentiable functions \( f : \mathbb{R} \to \mathbb{R} \) for which \( |x^rf^{(k)}(x)| < +\infty \) for all \( r > 0 \) and all \( k \in \mathbb{N} \). Then \( \mathcal{S} \subset L^1 \), and it is well known that the Fourier transform maps \( \mathcal{S} \) onto itself, thus the Fourier transform is always invertible on \( \mathcal{S} \). We note \( \|f\|_{r,k} = \sup_x \{|x|^r |f^{(k)}(x)|, x \in \mathbb{R} \} \) for any \( f \in \mathcal{S} \).

For two real numbers \( a, b \), the notation \( a \wedge b \) stand for the minimum of \( a \) and \( b \) whereas \( a \vee b \) stand for the maximum. Similarly, given two real valued functions \( f, g \) the function \( f \wedge g \) is the function which at \( x \) assigns the minimum of \( f(x) \) and \( g(x) \) and \( f \vee g \) has obvious definition. Throughout the paper \( C \) denotes a generic constant.

Inequalities up to a generic constant are denoted by \( \lesssim \) and \( \gtrsim \).

2. Posterior convergence rates for Symmetric Gamma mixtures

In this section we present the main results of the paper. We first present the three types of priors that are studied; i.e. location mixtures, location - scale mixtures and hybrid location-scale mixtures and for each of these families of priors we provide the associated posterior concentration rates.

Recall that we consider observations \((Y_i, X_i)_{i=1}^n\) independent and identically distributed according to model \((3)\) and we note \( y^n = (Y_1, \ldots, Y_n) \) and \( x^n = (X_1, \ldots, X_n) \). We denote the prior and the posterior distribution on \( f \) by \( \Pi(\cdot) \) and \( \Pi(\cdot | y^n, x^n) \) respectively.

2.1. Family of priors.
2.1.1. Location mixtures of Gaussians. A symmetric Gamma process location mixture of Gaussians prior $\Pi$ is the distribution of the random function $f(x) := \int \varphi((x - \mu)/\sigma) dM(\mu)$ where $\sigma \sim G_\sigma$ and $M \sim \Pi_\alpha$, with $\alpha$ a finite positive measure on $\mathbb{R}$, $G_\sigma$ a probability measure on $(0, \infty)$ and $\varphi(x) := e^{-x^2/2}$ for all $x \in \mathbb{R}$.

We restrict our discussion to priors for which the following conditions are verified. We assume that there are positive constants $a_1, a_2, a_3$ and $b_1, b_2, b_3, b_4$ such that $G_\sigma$ satisfies for $x \geq 1$

\begin{align*}
(5) & \quad G_\sigma(\sigma > x) \lesssim \exp(-a_1 x^{b_1}) \\
(6) & \quad G_\sigma(\sigma \leq 1/x) \lesssim \exp(-a_2 x^{b_2}) \\
(7) & \quad G_\sigma(x^{-1} \leq \sigma \leq x^{-1}(1 + t)) \gtrsim x^{b_3} t^{b_4} \exp(-a_3 x), \quad \forall t \in (0, 1).
\end{align*}

We let $\alpha := \alpha G_\mu$ for a positive constant $\alpha > 0$ and $G_\mu$ a probability distribution on $\mathbb{R}$. We assume that there are positive constants $b_5, b_6$ such that $G_\mu$ satisfies for all $x \in \mathbb{R}$

\begin{align*}
(8) & \quad G_\mu(|\mu - x| \leq t) \gtrsim t^{b_5} (1 + |x|)^{-b_6}, \quad \forall t \in (0, 1).
\end{align*}

The heavy tail condition on $G_\mu$ is required to not deteriorate the rate of convergence when $Q_0$ is heavy tailed.

Notice that equation (5) forbids the use of the classical inverse-Gamma distribution as prior distribution on $\sigma$ because of its heavy tail. In fact, it is always possible to weaken equation (5) to allow for Inverse-Gamma distribution (see Canale and De Blasi (2013); Naulet and Barat (2015)) but it complicates the proofs with no contribution to the subject of the paper. We found that among the usual distributions the inverse-Gaussian is more suitable for our purpose since it fulfills all the equations (5) to (7), as shown in proposition 1. We recall that the inverse-Gaussian distribution on $(0, \infty)$ with parameters $a > 0, b > 0$ has density with respect to Lebesgue measure

\[ f(x; a, b) := \left( \frac{b}{2\pi x^3} \right)^{1/2} \exp \left( -\frac{b(x-a)^2}{2a^2x} \right), \quad \forall x > 0, \]

and $f(x; a, b) = 0$ elsewhere.

**Proposition 1.** The inverse-Gaussian distribution with parameters $b, a > 0$ satisfies equations (5) to (7) with $a_1 = b/(2a^2), b_1 = 1, a_2 = b/4, b_2 = 1, b_3 = 1, b_4 = 1$ and $a_3 = b/2$.

**Proof.** It suffices to write, for any $x \geq 1$

\[ G_\sigma(\sigma > x) \leq \left( \frac{b}{2\pi x^3} \right)^{1/2} \int_x^\infty \exp \left( -\frac{b(t-a)^2}{2a^2t} \right) dt \]

\[ \leq \left( \frac{b}{2\pi} \right)^{1/2} \exp \left( \frac{b}{a} - \frac{b}{2} \right) \int_x^\infty \exp \left( -\frac{bt}{2a^2} \right) dt. \]

Also, for any $x \geq 1$

\[ G_\sigma(\sigma \leq 1/x) \leq \left( \frac{b}{2\pi} \right)^{1/2} \int_0^{1/x} t^{-3/2} \exp \left( -\frac{b(t-a)^2}{2a^2t} \right) dt \]
Finally, for any $x \geq 1$ and $0 < t < 1$,
\[
G_\alpha \left( x^{-1} \leq \sigma \leq x^{-1}(1 + t) \right) \geq \left( \frac{b}{2\pi} \right)^{1/2} e^{b/a} \int_0^{1/x} t^{-3/2} e^{-b/(4t)} dt. \quad \square
\]

2.1.2. Location-scale mixtures of Gaussians. A symmetric Gamma process location-scale mixture of Gaussians prior $\Pi$ is the distribution of the random function $f(x) := \int \varphi((x - \mu)/\sigma) dM(\sigma, \mu)$ where $M \sim \Pi_\alpha$, with $\alpha$ a finite positive measure on $(0, \infty) \times \mathbb{R}$ and $\varphi(x) := e^{-x^2/2}$ for all $x \in \mathbb{R}$. We focus the attention of the reader on the fact that although we use the same notations (i.e., $\Pi, \alpha$) as the previous section, these are different distributions and in the sequel we pay attention as making the context clear enough to avoid confusions.

We restrict our discussion to priors for which $\alpha := \overline{\alpha} G_\sigma \times G_\mu$, with $\overline{\alpha} > 0$ and $G_\sigma, G_\mu$ satisfying the same assumptions as in section 2.1.1.

2.1.3. Hybrid location-scale mixtures of Gaussians. The proof of the results given in the two preceding sections suggests that neither location or location-scale mixtures can achieve the optimal rates, whatever the nature of the tails of $Q_0$. We show that we can get better upper bounds by introducing hybrid mixtures.

By a hybrid location-scale mixture of Gaussians, we mean the distribution $\Pi$ of the random function $f(x) := \int \varphi((x - \mu)/\sigma) dM(\sigma, \mu)$ where $M \sim \Pi_\alpha$, with $\alpha = \overline{\alpha} P_\sigma \times G_\mu$, $\overline{\alpha} > 0$, $P_\sigma \sim \Pi_\sigma$ and $G_\mu$ a probability measure satisfying equation (8). Here $\Pi_\sigma$ is a prior distribution on the space of probability measures (endowed with Borel $\sigma$-algebra). We now formulate conditions on $\Pi_\sigma$ that are the random analoguous to equations (5) and (6). For the same constants $a_1, a_2, b_1, b_2$ as in section 2.1.1, we consider the existence of positive constants $a_4, a_5$ such that $\Pi_\sigma$ satisfies for $x > 0$ large enough
\[
\Pi_\sigma \left( P_\sigma : P_\sigma(\sigma > x) \geq \exp(-a_1 x^{b_1}/2) \right) \lesssim \exp(-a_4 x^{b_1}),
\]
\[
\Pi_\sigma \left( P_\sigma : P_\sigma(\sigma < 1/x) \geq \exp(-a_2 x^{b_2}/2) \right) \lesssim \exp(-a_5 x^{b_2}).
\]
As a replacement of equation (7), we assume that for all $r \geq 1$ there are constants $a_6, b_7$ such that for any positive integer $J$ large enough
\[
\Pi_\sigma \left( \cap_{j=0}^J \left\{ P_\sigma : P_\sigma(2^{-j}, 2^{-j}(1 + 2^{-Jr}) \geq 2^{-j}) \right\} \right) \gtrsim \exp(-a_6 J^{b_7} 2^J).
\]

Equations (9) to (11) are rather restrictive and it is not clear a priori whether or not such distribution exists. For example, if $P_\sigma$ is chosen to be almost-surely an Inverse-Gaussian distribution with parameters $b, \mu$ then equation (11) is not satisfied. However, we now show that under conditions on the base measure, $\Pi_\sigma$ can be chosen as a Dirichlet Process, hereafter referred to as DP.

We recall that if $\Pi_\sigma$ is a Dirichlet Process distribution with base measure $\alpha_\sigma G(\cdot)$ on $(0, \infty)$ (Ferguson, 1973), then $P_\sigma \sim \Pi_\sigma$ is a random probability measure on $(0, \infty)$ such that for any Borel measurable partition $A_1, \ldots, A_k$
of \((0, \infty)\), the joint distribution \(P_{\sigma}(A_1), \ldots, P_{\sigma}(A_k)\) is the \(k\)-variate Dirichlet distribution with parameters \(\alpha_{\sigma}G(A_1), \ldots, \alpha_{\sigma}G(A_k)\).

**Proposition 2.** Let \(\alpha_{\sigma} > 0\), \(G_{\sigma}\) a probability measure on \((0, \infty)\) satisfying the same assumptions as in equations (5) to (7), and \(\Pi_{\sigma}\) be a Dirichlet Process with base measure \(\alpha_{\sigma}G_{\sigma}(\cdot)\). Then \(\Pi_{\sigma}\) satisfies equations (9) to (11) with constants 
\[a_4 = a_1, a_5 = a_2,\] a constant \(a_6 > 0\) eventually depending on \(r\), and \(b_7 = 0\).

**Proof.** We first prove equation (9). It follows from the definition of the DP that \(P_{\sigma}(x, \infty)\) has Beta distribution with parameters \(\alpha_{\sigma}G_{\sigma}(x, \infty)\) and \(\alpha_{\sigma}(1 - G_{\sigma}(x, \infty))\), then by Markov’s inequality
\[
\Pi_{\sigma}\left(P_{\sigma} : P_{\sigma}(x, \infty) \geq t\right) \leq \frac{G_{\sigma}(x, \infty)}{t}.
\]
Likewise, if \(t = \exp(-a_1 x^{b_1}/2)\) and \(G_{\sigma}\) satisfies equations (5) to (7), the conclusion follows. The same steps with \(G_{\sigma}(0, 1/x)\) give the proof of equation (10). It remains to prove equation (11). Let \(r \geq 1\) and define \(V_{j,r} := \{\sigma : 2^{-j} \leq \sigma \leq 2^{-j}(1 + 2^{-jr})\}\) for any integer \(0 \leq j \leq J\). For all \(r \geq 1\) the \(V_{j,r}\) ’s are disjoint. Set \(V_{c} := \cup_{j=0}^{J} V_{j,r}\). If \(\alpha_{\sigma}G_{\sigma}(V_{c}) \leq 1\) let \(V_{J+1,r} = V_{c}\) and \(M = 1\); otherwise split \(V_{c}\) into \(M > 1\) disjoint subsets \(V_{1,r}, \ldots, V_{M,r}\) such that \(\exp(-2^J) \leq \alpha_{\sigma}G_{\sigma}(V_{k,r}) \leq 1\) for all \(k = 1, \ldots, M\) and set \(V_{J+1,r} = V_{1,r}\), 
\[V_{J+2,r} = V_{2,r}, \ldots, V_{J+M,r} = V_{M,r}\] (since \(G_{\sigma}(0, \infty) = 1\) this can be done with a number \(M\) independent of \(J\)). For \(J\) large enough (so that \((J + M)2^{-J+1} < 1\)), acting as in Ghosal et al. (2000, lemma 6.1), it follows
\[
\Pi_{\sigma}\left(P_{\sigma} : P_{\sigma}[2^{-j}, 2^{-j}(1 + 2^{-jr})] \geq 2^{-j} \quad \forall 0 \leq j \leq J\right) \geq \frac{\Gamma(\alpha_{\sigma})2^{-j(J+M)}}{\prod_{j=0}^{J+M} \Gamma(\alpha_{\sigma}G_{\sigma}(V_{j+r}))}.
\]
Also, \(\alpha_{\sigma}G_{\sigma}(V_{j,r}) \leq 1\) implies \(\Gamma(\alpha_{\sigma}G_{\sigma}(V_{j,r})) \leq 1/(\alpha_{\sigma}G_{\sigma}(V_{j,r}))\), hence
\[
\Pi_{\sigma}\left(P_{\sigma} : P_{\sigma}[2^{-j}, 2^{-j}(1 + 2^{-jr})] \geq 2^{-j} \quad \forall 0 \leq j \leq J\right) \geq \frac{\Gamma(\alpha_{\sigma})\alpha_{\sigma}^{J+M+1}2^{-J(J+M)}}{\prod_{j=0}^{J+M} G_{\sigma}(V_{j+r})}.
\]
Since \(M\) does not depend on \(J\), one can find a constant \(C > 0\) such that
\[
\Pi_{\sigma}\left(P_{\sigma} : P_{\sigma}[2^{-j}, 2^{-j}(1 + 2^{-jr})] \geq 2^{-j} \quad \forall 0 \leq j \leq J\right) \geq \Gamma(\alpha_{\sigma}) \exp\left\{-CJ^2 + \sum_{j=0}^{J} \log G_{\sigma}(V_{j,r}) + \sum_{j=J+1}^{J+M} \log G_{\sigma}(V_{j+r})\right\}.
\]
By construction, the second sum in the rhs of the last equation is lower bounded by \(-M2^J\), whereas if \(G_{\sigma}\) satisfies equations (5) to (7), the first sum is lower bounded by \(-C’2^J\) for a constant \(C’ > 0\) eventually depending on \(r\). Then the proposition is proved. \(\square\)
2.2. Posterior concentration rates under the mixture priors. We let $\Pi(\cdot \mid y^n, x^n)$ denote the posterior distribution of $f \sim \Pi$ based on $n$ observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ modelled as in section 1. Let $(\epsilon_n)_{n \geq 1}$ be a sequence of positive numbers with $\lim_{n} \epsilon_n = 0$, and $d_n$ denote the empirical $L^2$ distance, that is $n d_n(f, g)^2 = \sum_{i=1}^{n} |f(X_i) - g(X_i)|^2$.

The following theorem is proved in Section 3.

**Theorem 1.** Consider the model (3), and assume that $f_0 \in L^1 \cap C^\beta$ and $Q_0, X^p < +\infty$. Then there exist a constant $C > 0$ and $t > 0$ depending only on $f_0$ and $Q_0$ such that

- If the prior $\Pi$ is the symmetric Gamma location mixture of Gaussians as defined in section 2.1.1
  
  $$\Pi \left( d_n(f, f_0)^2 > C n^{-2\beta/(3\beta+1)} (\log n)^t \mid y^n, x^n \right) = o_p(1)$$

  when $0 < p \leq 2$, and

  $$\Pi \left( d_n(f, f_0)^2 > C n^{-2\beta/(2\beta+1+2\beta/p)} (\log n)^t \mid y^n, x^n \right) = o_p(1)$$

  when $p > 2$.

- If the prior $\Pi$ is the symmetric Gamma location-scale mixture of Gaussians defined in section 2.1.2

  $$\Pi \left( d_n(f, f_0)^2 > C n^{-2\beta/(3\beta+2)} \wedge n^{-2\beta/(2\beta+1+2\beta/p)} (\log n)^t \mid y^n, x^n \right) = o_p(1)$$

  when $0 < p \leq 2$, and

  $$\Pi \left( d_n(f, f_0)^2 > C n^{-\beta/(\beta+1)} (\log n)^t \mid y^n, x^n \right) = o_p(1),$$

  when $p > 2\beta$.

- If the prior $\Pi$ is the hybrid symmetric Gamma location-scale mixture of Gaussians defined in section 2.1.3

  $$\Pi \left( d_n(f, f_0)^2 > C \left[ n^{-2\beta/(3\beta+2)} \wedge n^{-p/(p+1)} \right] (\log n)^t \mid y^n, x^n \right) = o_p(1),$$

  when $p \leq 2\beta$ or

  $$\Pi \left( d_n(f, f_0)^2 > C n^{-2\beta/(2\beta+1)} (\log n)^t \mid y^n, x^n \right) = o_p(1),$$

  when $p > 2\beta$.

The upper bounds on the rates in the previous paragraph are no longer valid when $p = 0$. Indeed the constant $C > 0$ depends on $p$ and might not be definite if $p = 0$; the reason is to be found in the fact that $C$ heavily depends on the ability of the prior to draw mixture component in regions of observed data, which remains concentrated near the origin when $p > 0$. In section 2.3, we overcome this issue by making the prior covariate dependent; this allows to derive rates under the assumption $p = 0$ (no tail assumption).
2.3. Relating the tail assumption: covariate dependent prior for location mixtures. Although the rates derived in section 3 do not depend on $p > 0$ when $p$ is small, the assumption $Q_0|X|^p < +\infty$ is crucial in proving the Kullback-Leibler condition. Indeed, this condition ensures that the covariates belong to a set $\mathcal{X}_n$ which is not too large, which allows us to bound from below the prior mass of Kullback-Leibler neighbourhoods of the true distribution. Surprisingly, it seems very difficult to get rid of this assumption under a fully Bayesian framework without fancy assumptions, while making the prior covariates dependent allows to drop all tail conditions on $Q_0$. Doing so, we can adapt to the tail behaviour of $Q_0$, as shown in the following theorem, which is an adaptation of the general theorems of Ghosal et al. (2007b). For convenience, in the sequel we drop out the superscript $n$ and we write $x$, $y$ for $x^n$, $y^n$, respectively. For $\epsilon > 0$ and any subset $A$ of a metric space equipped with metric $d$, we let $N(\epsilon, A, d)$ denote the $\epsilon$-covering number of $A$, i.e. $N(\epsilon, A, d)$ is the smallest number of balls of radius $\epsilon$ needed to cover $A$.

**Theorem 2.** Let $\Pi_x$ be a prior distribution that depends on the covariate vector $x$, $0 < c_2 < 1/4$ and $\epsilon_n \to 0$ with $n\epsilon_n^2 \to \infty$. Suppose that $\mathcal{F}_n \subseteq \mathcal{F}$ is such that $Q_0^n \Pi_x(\mathcal{F}_n) \lesssim \exp(-\frac{1}{2}(1+2c_2)n\epsilon_n^2)$ and $\log N(\epsilon_n/18, \mathcal{F}_n, d_n) \lesssim n\epsilon_n^2/4$ for $n$ large enough. If for any $x \in \mathbb{R}^n$ it holds $\Pi_x(f : d_n(f, f_0) \leq s\epsilon_n) \gtrsim \exp(-c_2n\epsilon_n^2)$, then for all $M > 0$ we have $\Pi_x(f : d_n(f, f_0) > M\epsilon_n \mid y, x) = o_p(1)$.

We apply theorem 2 to symmetric Gamma process location mixtures of Gaussians in the following way. Let $Q^\alpha_x$ denote the empirical measure of the covariate vector $x$. Given a a probability density function $g$, we let $G_x$ the probability measure which density is $z \mapsto \int g(z - x_i) dQ^\alpha_x(x)$.

**Corollary 1.** Then we let $\Pi_x$ be the distribution of the random function $f(x) := \int \varphi((x - \mu)/\sigma) dM(\mu)$, where $\sigma \sim G_\sigma$ and $M \sim \Pi_\alpha$ with $\alpha = \overline{\sigma}G_x$ for some $\overline{\sigma} > 0$. Assume that $G_\sigma$ satisfies equations (5) to (7) and that there exists a constant $b_8 > 0$ such that $\sup_{x \in \mathbb{R}^n} G_x(\mu : |\mu - s| \leq t) \lesssim t^{b_8}$ for all $0 < t, s \leq 1$. Then $\Pi_x(f : d_n(f, f_0) > M\epsilon_n \mid y, x) = o_p(1)$ with $\epsilon_n^2 \lesssim n^{-2\beta/(3\beta+1)}(\log n)^2-2\beta/(3\beta+1)$.

To prove corollary 1, note that neither the proof of lemma 4 or lemma 5 involve the base measure $\alpha$ (indeed, it only involves $\overline{\sigma}$); thus we can use the sieve $\mathcal{F}_n$ constructed in section 4.1.2. To apply theorem 2 it is then sufficient to prove that for all $x \in \mathbb{R}^n$

\[
(12) \quad \Pi_x(f : d_n(f, f_0) \leq s\epsilon_n) \gtrsim \exp(-c_2n\epsilon_n^2).
\]

This is done in lemma 1.

**Lemma 1.** Assume that there is a constant $b_8 > 0$ such that $\sup_{x \in \mathbb{R}^n} G_x(\mu : |\mu - s| \leq t) \lesssim t^{b_8}$ for all $0 < t, s \leq 1$. Also assume that $G_\sigma$ satisfies equations (5) to (7). Then equation (12) holds for the symmetric Gamma location mixture of Gaussians with base measure $\overline{\sigma}G_x$ if $\epsilon_n^2 \leq Cn^{-2\beta/(3\beta+1)}(\log n)^2-2\beta/(3\beta+1)$ for an appropriate constant $C > 0$.

The proof of lemma 1 is given in appendix B.
3. Proofs

To prove theorem 1 we follow the lines of Ghosal et al. (2000); Ghosal and Van Der Vaart (2001); Ghosal et al. (2007a). Namely we need to verify the following three conditions

- Kullback-Leibler condition: For a constant $0 < c_2 < 1/4$,
  \begin{equation}
  \Pi(\text{KL}(f_0, \epsilon_n)) \geq e^{-c_2 n \epsilon_n^2},
  \end{equation}
  where
  \[ \text{KL}(f_0, \epsilon_n) := \left\{ f : \frac{1}{2\sigma^2} \int |f_0(x) - f(x)|^2 \text{d}Q_0(x) \leq \epsilon_n^2 \right\}. \]

- Sieve condition: There exists $\mathcal{F}_n \subset \mathcal{F}$ such that
  \begin{equation}
  \Pi(\mathcal{F}_n^c) \leq e^{-\frac{1}{2}(1+2c_2)n\epsilon_n^2},
  \end{equation}

- Tests: Let $N(\epsilon_n/18, \mathcal{F}_n, d_n)$ be the logarithm of the covering number of $\mathcal{F}_n$ with radius $\epsilon_n/18$ in the $d_n(\cdot, \cdot)$ metric.
  \begin{equation}
  N(\epsilon_n/18, \mathcal{F}_n, d_n) \leq \frac{n\epsilon_n^2}{4}.
  \end{equation}

The Kullback-Leibler condition is proved by defining an approximation of $f$ by a discrete mixture under weak tail conditions. Although the general idea is close to Kruijer et al. (2010) or Scricciolo (2014), the construction remains quite different to be able to handle various tail behaviours. This is detailed in the following section.

3.1. Approximation theory. To describe the approximation of $f_0$ by a finite mixture, we first define a few notations.

Let $\hat{\chi}$ be a $C^\infty$ function that equals 1 on $[-1,1]$ and 0 outside $[-2,2]$. (think for instance as the convolution of $\mathbb{1}_{[-1,1]}$ with $x \mapsto \exp(-1/(1-x^2)) \mathbb{1}_{[-1,1]}(x)$). For any $\sigma > 0$ we use the shortened notation $\hat{\chi}_\sigma(\xi) := \hat{\chi}(2\sigma \xi)$. Define $\eta$ as the function which $L^1$ Fourier transform satisfies $\hat{\eta}(\xi) = \hat{\chi}(\xi)/\hat{\psi}(\xi)$ for all $\xi \in [-2,2]$ and $\hat{\eta}(\xi) = 0$ elsewhere. For two positive real numbers $h$ and $\sigma$, we define the kernel $K_{h,\sigma} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ such that

$$K_{h,\sigma}(x, y) := \frac{h}{\sigma} \sum_{k \in \mathbb{Z}} \varphi \left( \frac{x - h\sigma k}{\sigma} \right) \eta \left( \frac{y - h\sigma k}{\sigma} \right), \quad \forall (x, y) \in \mathbb{R} \times \mathbb{R}. $$

For a measurable function $f$ we introduce the operator associated with the kernel $K_{h,\sigma} f(x) := \int K_{h,\sigma}(x, y) f(y) \text{d}y$. The function $K_{h,\sigma} f$ will play the role of an approximation for the function $f$, and we will evaluate how this approximation becomes close to $f$ given $h$ and $\sigma$ sufficiently close to zero.

More precisely, we will prove that, when choosing $h$ appropriately, $f$ can be approximated by $K_{h,\sigma} \chi_\sigma \times f_0$ to the order $\sigma^{\beta}$. Moreover $K_{h,\sigma} \chi_\sigma \times f_0$ can be written as $\sum_{k \in \mathbb{Z}} u_k \varphi((x - \mu_k)/\sigma))$. In a second step we approximate $K_{h,\sigma} \chi_\sigma \times f_0$ by a truncated version of it, retaining only the $k$’s such that $|u_k|$ is large enough and $|\mu_k|$ not too large. In the case of location - scale and hybrid location - scale mixtures we consider a modification of this approximation to control better the number of components for which $\sigma$ needs to be small. We believe that these
constructions have interest in themselves. In particular they shed light on the relations between Gaussian mixtures and wavelet approximations.

These approximation properties are presented in the following two Lemmas which are proved in appendix A:

**Lemma 2.** There is $C > 0$ depending only on $\beta$ such that for any $f_0 \in L^1 \cap C^\beta$ and any $\sigma > 0$ we have $|\chi_\sigma * f_0(x) - f_0(x)| \leq C \| f_0 \|_{C^\beta}$ for all $x \in \mathbb{R}$.

**Lemma 3.** Let $f_\sigma := \chi_\sigma * f_0$ and $h \leq 1$. Then there is a universal constant $C > 0$ such that $|K_{h,\sigma} f_\sigma(x) - f_\sigma(x)| \leq C \| f_0 \|_{1} \sigma^{-1} e^{-4\sigma^2/h^2}$ for all $x \in \mathbb{R}$.

We now present the approximation schemes in the context of location mixtures.

### 3.2. Construction of the approximation under location mixtures

Let $0 < \sigma \leq 1$ and $h_\sigma \sqrt{\log \sigma^{-1}} : = 2\pi \sqrt{\beta + 1}$. Then combining the results of lemma 2 and lemma 3 we can conclude that $|K_{h_\sigma,\sigma} (\chi_\sigma * f_0)(x) - f_\sigma(x)| \lesssim \sigma^\beta$. Now we define the coefficients $u_k, k \in \mathbb{Z}$ so that

$$K_{h_\sigma,\sigma} (\chi_\sigma * f_0)(x) := \sum_{k \in \mathbb{Z}} u_k \varphi \left( \frac{x - \mu_k}{\sigma} \right), \quad \forall k \in \mathbb{Z},$$

where $\mu_k := h_\sigma \sigma k$ for all $k \in \mathbb{Z}$. Let define

$$\Lambda := \left\{ k \in \mathbb{Z} : |u_k| > \sigma^\beta, \quad |\mu_k| \leq \sigma^{-2\beta/p} + \sigma \sqrt{2(\beta + 1) \log \sigma^{-1}} \right\},$$

$$U_\sigma := \{ \sigma' : \sigma \leq \sigma' \leq \sigma(1 + \sigma^\beta) \},$$

and for all $k \in \Lambda$ we define $V_k := \{ \mu : |\mu - \mu_k| \leq \sigma^{\beta + 1} \}$ and $V = \cap_{k \in \Lambda} V_k$. We also denote

$$\mathcal{M}_\sigma := \left\{ M \text{ signed measure on } \mathbb{R} : |M(V_k) - u_k| \leq \sigma^\beta, \quad \forall k \in \Lambda : |M|(V^c) \leq \sigma^\beta \right\},$$

and for any $M \in \mathcal{M}_\sigma$, we write $f_{M,\sigma}(x) := \int \varphi((x - \mu)/\sigma) \, dM(\mu)$.

**Proposition 3.** For $\sigma > 0$ small enough, it holds $|\Lambda| \lesssim \sigma^{-(\beta + 1)} \wedge h_\sigma^{-1} \sigma^{-(2\beta/p + 1)}$. 

**Proof.** Because there is a separation of $h_\sigma \sigma$ between two consecutive $\mu_k$, it is clear that $|\Lambda| \leq 2h_\sigma^{-1} \sigma^{-(2\beta/p + 1)}$. Moreover, from proposition 9 we have the following estimate.

$$\| f_0 \|_{1} \sigma^{-1} \gtrsim \sum_{k \in \mathbb{Z}} |u_k| \gtrsim \sum_{k \in \Lambda} |u_k| \geq \sigma^\beta |\Lambda|. \quad \square$$

**Proposition 4.** For all $x \in \mathbb{R}$, all $\sigma > 0$ small enough and all $M \in \mathcal{M}_\sigma$ it holds $|f_{M,\sigma}(x) - f_0(x)| \lesssim h_\sigma^{-1}$. 

**Proof.** For any $M \in \mathcal{M}_\sigma$, we have that $|f_{M,\sigma}(x) - f_0(x)| \leq |f_{M,\sigma}(x)| + \| f_0 \|_{\infty}$. But, with $I \equiv I(x) := \{ k \in \mathbb{Z} : |x - \mu_k| \leq 2\sigma \}$,

$$f_{M,\sigma}(x) = \sum_{k \in \Lambda \cap I} \int_{V_k} \varphi \left( \frac{x - \mu}{\sigma} \right) \, dM(\mu)$$

$$+ \sum_{k \in \Lambda \cap I^c} \int_{V_k} \varphi \left( \frac{x - \mu}{\sigma} \right) \, dM(\mu) + \int_{V^c} \varphi \left( \frac{x - \mu}{\sigma} \right) \, dM(\mu).$$
Clearly the last term of this last expression is bounded above by \( \|\varphi\|_\infty \sigma^\beta \). For the second term, we have for any \( \mu \in V_k \) with \( k \in J^c \) that \( |x - \mu| \geq |x - \mu_k| - |\mu - \mu_k| \geq |x - \mu_k|/2 \). Then the second term of the rhs of equation (16) is bounded above by

\[
\sup_{k \in \Lambda \cap J^c} |M|(V_k) \sum_{k \in \mathbb{Z}} \varphi \left( \frac{x - h_k \sigma k}{\sigma} \right).
\]

Proceeding as in the proof of lemma 9, we deduce that the series in the last expression is bounded above by a constant times \( 1/h_\sigma \), whereas proposition 9 and Young’s inequality yields \( |M|(V_k) \leq |M(V_k) - u_k| + |u_k| \lesssim \sigma^\beta + \|\chi_\sigma \ast f_0\|_\infty \lesssim \sigma^\beta + \|\chi\|_1 \|f_0\|_\infty \). Therefore the second term of the rhs in equation (16) is bounded by a constant multiple of \( h_\sigma^{-1} \). Regarding the first term in equation (16), it is bounded by \( \|\varphi\|_\infty |J| \sup_{k \in \Lambda} |M|(V_k) \), which is in turn bounded by \( h_\sigma^{-1} \) times a constant. \( \square \)

**Proposition 5.** For all \( \sigma > 0 \) small enough, all \( x \in \mathbb{R} \) with \( |x| \leq \sigma^{2\beta/p} \) and all \( M \in \mathcal{M}_\sigma \) it holds \( |f_{M,\sigma}(x) - f_0(x)| \lesssim h_\sigma^{-2} \sigma^\beta \).

**Proof.** We define \( A_\sigma(\beta) := \sqrt{2 \log |\Lambda| + 2(\beta + 1) \log \sigma^{-1}} \). Then for any \( M \in \mathcal{M}_\sigma \), letting \( J := J(x) := \{ k \in \mathbb{Z} : |x - \mu_k| \leq 2\sigma A_\sigma(\beta) \} \), we may write

\[
f_{M,\sigma}(x) - K_{h_\sigma,\sigma}(\chi_\sigma \ast f_0)(x) = \sum_{k \in \Lambda \cap J} \int_{V_k} \left[ \varphi \left( \frac{x - \mu}{\sigma} \right) - \varphi \left( \frac{x - \mu_k}{\sigma} \right) \right] dM(\mu)
+ \sum_{k \in \Lambda \cap J^c} |M(V_k) - u_k| \varphi \left( \frac{x - \mu_k}{\sigma} \right) + \sum_{k \in \Lambda \cap J^c} \int_{V_k} \varphi \left( \frac{x - \mu}{\sigma} \right) dM(\mu)
- \sum_{k \in \Lambda \cap J^c} u_k \varphi \left( \frac{x - \mu_k}{\sigma} \right) - \sum_{k \in \Lambda} u_k \varphi \left( \frac{x - \mu_k}{\sigma} \right) + \int_{V^c} \varphi \left( \frac{x - \mu}{\sigma} \right) dM(\mu)
:= r_1(x) + r_2(x) + r_3(x) + r_4(x) + r_5(x) + r_6(x).
\]

With the same argument as in proposition 3, we deduce that \( |J| \leq 2h_\sigma^{-1} A_\sigma(\beta) \). The same proposition implies \( A_\sigma(\beta) \lesssim \sqrt{\log \sigma^{-1}} \). Recalling that \( |M|(V_k) \lesssim 1 + \|\chi\|_1 \|f_0\|_\infty \) for all \( k \in \Lambda \) and all \( M \in \mathcal{M}_\sigma \), it follows from proposition 11 that \( |r_1(x)| \lesssim A_\sigma(\beta) h_\sigma^{-1} \sigma^\beta \). From the definition of \( \mathcal{M}_\sigma \), it comes \( |r_2(x)| \leq \|\varphi\|_\infty |J| \sigma^\beta \leq 2\|\varphi\|_\infty A_\sigma(\beta) h_\sigma^{-1} \sigma^\beta \). Whenever \( k \in \Lambda \cap J^c \) and \( \mu \in V_k \), it holds \( |x - \mu| \geq |x - \mu_k| - |\mu - \mu_k| \geq \sigma A_\sigma(\beta) \). Therefore, \( |r_3(x)| \lesssim \varphi(\sigma A_\sigma(\beta)) |\Lambda| \lesssim \sigma^{\beta+1} \). With the same argument, proposition 9 and Young’s inequality we get \( |r_4(x)| \lesssim \|\chi_\sigma \ast f_0\|_\infty \varphi(2A_\sigma(\beta)) |\Lambda| \lesssim \|\chi\|_1 \|f_0\|_\infty \sigma^\beta \). Regarding \( r_5 \), we rewrite \( \Lambda^c = \Lambda_1^c \cup \Lambda_2^c \), with \( \Lambda_1^c := \{ k \in \mathbb{Z} : |u_k| \leq \sigma^\beta \} \) and \( \Lambda_2^c := \{ k \in \mathbb{Z} : |u_k| > \sigma^{-2\beta/p} + \sqrt{2(\beta + 1) \log \sigma^{-1}} \} \). Then,

\[
|r_5(x)| \leq \sum_{k \in \Lambda_1^c} |u_k| \varphi \left( \frac{x - \mu_k}{\sigma} \right) + \sum_{k \in \Lambda_2^c} |u_k| \varphi \left( \frac{x - \mu_k}{\sigma} \right)
\leq \sigma^\beta \sup_{x \in \mathbb{R}} \sum_{k \in \mathbb{Z}} \varphi \left( \frac{x - \mu_k}{\sigma} \right) + \sum_{k \in \Lambda_2^c} |u_k| \varphi \left( \frac{x - \mu_k}{\sigma} \right).
\]
The first term of the rhs of equation (18) is bounded by a multiple constant of $h_{\sigma}^{-1} \sigma^\beta$, with the same argument as in the proof of lemma 9. By definition of $\Lambda_0^i$, $|x - \mu_k| \geq \sigma \sqrt{2(\beta + 1) \log \sigma^{-1}}$ when $k \in \Lambda_0^i$ and $|x| \leq \sigma^{-2\beta/p}$. This implies, together with proposition 9 and Young's inequality, that the second term of the rhs of equation (18) is bounded by a constant multiple of $\sigma^\beta + 1 \sum_{k \in \mathbb{Z}} |u_k| \lesssim \|\chi_{\sigma} * f_0\|_1 \sigma^\beta \leq \|\chi_1\|_1 \|f_0\|_1 \sigma^\beta$ for all $|x| \leq \sigma^{-2\beta/p}$. Finally, we have the trivial bound $|r_0(x)| \leq \|\varphi\|_{\infty} |M|(V^c) \leq \|\varphi\|_{\infty} \sigma^\beta$.

\[ \square \]

3.3. Construction of the approximation under location-scale and hybrid location-scale mixtures. Let $\sigma_0 := 1$ and define recursively $\sigma_{j+1} := \sigma_j/2$ for any $j \geq 0$. Let $\Delta_0 := f_0 - \chi_{\sigma_0} * f_0$, and define recursively $\Delta_{j+1} := \Delta_j - \sigma_{j+1} * \Delta_j$, for any $j \geq 0$.

The general idea of the construction is that $|\Delta_j| \lesssim \sigma_j^\beta$, as shown in proposition 10 in appendix, and that similarly to wavelet decomposition, we approximate a function $f_0$ by Hölder $\beta$ by

\[ f_1 := K_0(\chi_{\sigma_0} * f_0) + \sum_{j=1}^J K_j(\chi_{\sigma_j} * \Delta_{j-1}). \]

where $J \geq 1$ is a large enough integer, $h_J \sqrt{J} := 2\pi/\sqrt{J \log 2}$, and $K_j := K_{h_J, \sigma_j}$. By induction, we get that $\Delta_j = \Delta_0 - \sum_{j=1}^j \chi_{\sigma_{j+1}} * \Delta_j$. It follows,

\[ f_1 - f_0 = K_0(\chi_{\sigma_0} * f_0) - \chi_{\sigma_0} * f_0 + \sum_{j=1}^J K_j(\chi_{\sigma_j} * \Delta_{j-1}) \]

\[ = \Delta_j + K_0(\chi_{\sigma_0} * f_0) - \chi_{\sigma_0} * f_0 + \sum_{j=1}^J [K_j(\chi_{\sigma_j} * \Delta_{j-1}) - \chi_{\sigma_j} * \Delta_{j-1}] . \]

Therefore, from lemma 3 and proposition 10 and Young's inequality, the error of approximating $f_0$ by $f_1$ is

\[ |f_1(x) - f_0(x)| \]

\[ \leq |\Delta_j| + |K_0(\chi_{\sigma_0} * f_0) - \chi_{\sigma_0} * f_0| + \sum_{j=1}^{J} |K_j(\chi_{\sigma_j} * \Delta_{j-1}) - \chi_{\sigma_j} * \Delta_{j-1}| \]

\[ \lesssim \|f_0\|_{C^\beta} \sigma_j^\beta + \|\chi_{\sigma_0} * f_0\|_1 \sigma_0^{-1} e^{-4\pi^2/h_j^2} + e^{-4\pi^2/h_j^2} \sum_{j=1}^{J} \|\chi_{\sigma_j} * \Delta_{j-1}\| \sigma_j^{-1} \]

\[ \lesssim \|f_0\|_{C^\beta} \sigma_j^\beta + \|f\|_1 e^{-4\pi^2/h_j^2} + \|f_0\|_1 e^{-4\pi^2/h_j^2} \sum_{j=1}^{J} 2^j \]

\[ \lesssim \|f_0\|_{C^\beta} \sigma_j^\beta + \|f_0\|_1 (1 + 2^J) e^{-4\pi^2/h_j^2} \lesssim \sigma_j^\beta. \]

The reason for considering different scale parameters in the construction, is to deal with fat tail, the heuristic being that in the tail we do not require as precise an approximation as in the center. In particular small values of $j$ will be used to estimate the function far off in the tails. To formalize this, we define $\zeta_j := 2^{(J-j)(2\beta/p)}$, and $A_j := [-\zeta_j, \zeta_j]$, for all $j = 0, \ldots J$. We also define
\( I_j = [-1, 1] \), and for all \( j = 0, \ldots, J - 1 \) we set \( I_j := A_j \setminus A_{j+1} \). Notice that by definition of \( K_j \), we can write,

\[
K_0(\chi \sigma_0 \ast f_0)(x) := \sum_{k \in \mathbb{Z}} u_{0k} \varphi((x - h_j \sigma_0 k)/\sigma_0)
\]

\[
K_j(\chi \sigma_j \ast \Delta_{j-1})(x) := \sum_{k \in \mathbb{Z}} u_{jk} \varphi((x - h_j \sigma_j k)/\sigma_j), \quad \forall j \geq 1.
\]

To ease notation, we define \( \mu_{jk} := h_j \sigma_j k \) for all \( j \geq 0 \) and all \( k \in \mathbb{Z} \). In the sequel we shall need the following subset of indexes,

\[
\Lambda := \left\{ (j, k) \in \{0, \ldots, J\} \times \mathbb{Z} : |u_{jk}| > \sigma_j^\beta, \quad |\mu_{jk}| \leq \zeta_j + \sqrt{2(\beta + 1) \log \sigma_j^{-1}} \right\}.
\]

We prove below that we can approximate \( f_1 \) by a finite mixture corresponding to retaining only the components associated to indices in \( \Lambda \) and that we can bound the cardinality of \( \Lambda \) by \( O(J \log J \sigma_j^{-2\beta/p}) \).

To any \((j, k) \in \Lambda\) we associate \( U_j := \{ \sigma : \sigma_j \leq \sigma \leq \sigma_j(1 + \sigma_j^\beta) \}, V_j := \{ \mu : |\mu - \mu_{jk}| \leq \sigma_j \sigma_j^\beta \} \) and \( W_{jk} := U_j \times V_j \). We denote by \( \mathcal{M} \) the set of signed measures \( M \) on \((0, \infty) \times \mathbb{R}\) such that \( |M(W_{jk}) - u_{jk}| \leq \sigma_j^\beta \) for all \((j, k) \in \Lambda\), and \( |M(W^c)| \leq \sigma_j^\beta \), where \( W^c \) is the relative complement of the union of all \( W_{jk} \) for \((j, k) \in \Lambda\).

For any \( M \in \mathcal{M} \), we write

\[
f_M(x) := \int \varphi((x - \mu)/\sigma) \, dM(\sigma, \mu).
\]

In proposition 6 we control the cardinality of \( \Lambda \) while in proposition 8 we control the error between \( f_M \) and \( f_1 \) on the decreasing sequence of intervals \([-\zeta_j, \zeta_j]\). Proposition 7 provides a crude uniform upper bound on \( f_M \) and \( f_0 \).

**Proposition 6.** There is a constant \( C > 0 \) depending only on \( f_0 \) and \( Q_0 \) such that \( |\Lambda| \leq C[\sigma_j^{-(\beta+1)} \wedge (J \log J) \sigma_j^{-2\beta/p}] \) if \( p \leq 2\beta \), and \( |\Lambda| \leq C(J \log J) \sigma_j^{-1} \) if \( p > 2\beta \).

**Proof.** First notice that because of propositions 9 and 10, we always have the bound

\[
4\|f_0\|_1 \sigma_j^{-1} \geq 2\|f_0\|_1 \sum_{j=0}^J \sigma_j^{-1} \geq \sum_{j=0}^J \sum_{k \in \mathbb{Z}} |u_{jk}| \geq \sum_{(j, k) \in \Lambda} |u_{jk}| \geq \sigma_j^\beta |\Lambda|.
\]

If \( p \leq 2\beta \), we define \( B := \sqrt{2(\beta + 1) \log 2} \), so that \( \sqrt{2(\beta + 1) \log \sigma_j^{-1}} = B \sqrt{J} \).

Now consider those indexes \( j \) with \( \zeta_j \leq B \sqrt{J} \). An elementary computation shows that there are at most \( \lesssim \log J \) such indexes. Therefore, recalling that there is a separation of \( h_j \sigma_j \) between two consecutive \( \mu_{jk} \) and that there are at most \( J \) indexes \( j \) with \( \zeta_j > B \sqrt{J} \)

\[
|\Lambda| \lesssim \sum_{j=0}^J \frac{4\zeta_j}{h_j \sigma_j} + \log J \times \frac{2B \sqrt{J}}{h_j \sigma_j}
\]

\[
\leq 4h_j^{-1} \sigma_j^{-2\beta/p} \sum_{j=0}^J 2^{-j(2\beta/p - 1)} + 2B(\sqrt{J} \log J)h_j^{-1} \sigma_j^{-1}.
\]
Because \( h_J \sqrt{J} \lesssim 1 \) by definition, and because \( p \leq 2\beta \), the result follows from the last equation and equation (19). If \( p > 2\beta \), the reasoning is the same as in the first part, but we can rewrite in this situation the equation (20) as

\[
|\Lambda| \leq 4h_J^{-1} \sigma_J^{-1} \sum_{j=0}^{J} 2^{(j-J)(1-\frac{2\beta}{p})} + 2B(\sqrt{J} \log J) h_J^{-1} \sigma_J^{-1}.
\]

Since \( p > 2\beta \), the conclusion is immediate. \( \square \)

**Proposition 7.** For all \( x \in \mathbb{R} \), all \( J > 0 \) large enough and all \( M \in \mathcal{M} \), it holds \( |f_M(x) - f_0(x)| \lesssim J^{3/2} \).

**Proof.** Let \( \mathcal{I} \equiv \mathcal{I}(x) := \{(j, k) \in \{0, \ldots, J\} \times \mathbb{Z} : |x - \mu_{jk}| \leq 2\sigma_j \} \). Then the proof is almost identical to proposition 4. It suffices to notice that

- \(|M|(W_{jk}) \leq |M| (W_{jk}) - u_{jk} + |u_{jk}| \) is always bounded above by a constant, because of the definition of \( M \), of propositions 9 and 10.
- \( |x - \mu_j|/\sigma \geq (1/4)|x - \mu_k|/\sigma_j \) whenever \((\sigma, \mu) \in W_{jk} \) and \((j, k) \in \Lambda \cap \mathcal{I}^c \), as soon as \( J \) is large enough.
- \(|\mathcal{I}| \leq 5Jh_J^{-1} \) for \( J \geq 1 \). \( \square \)

**Proposition 8.** If \( f_0 \in \mathcal{C}_\beta \), for all \( J > 0 \) large enough, all \( 0 \leq j \leq J \), all \( x \in [-\zeta_j, \zeta_j] \) and all \( M \in \mathcal{M} \), it holds \( |f_M(x) - f_0(x)| \lesssim J^{3/2} \sigma_j^3 \).

The proof of proposition 8 is given in appendix C.

4. **Proof of theorem 1**

As mentioned earlier, the proof of theorem 1 boils down to verifying conditions (13), (14) and (15) for the three types of priors.

4.1. **Case of the location mixture.**

4.1.1. **Kullback-Leibler condition for location mixtures.** In this Section we verify condition (13) in the case of the location mixture prior, using the results of section 3.2

By Chebychev inequality, we have \( Q_0[-\sigma^{-2\beta/p}, \sigma^{2\beta/p}] \leq \sigma^{2\beta} Q_0|X|^p \). Then by bringing together results from propositions 4 and 5, we can find a constant \( C > 0 \) such that for all \( M \in \mathcal{M}_\sigma \)

\[
\int |f_{M, \sigma}(x) - f_0(x)|^2 \, dQ_0(x) \leq \sup_{|x| > \sigma^{-2\beta/p}} |f_{M, \sigma}(x) - f_0(x)|^2 Q_0[-\sigma^{-2\beta/p}, \sigma^{2\beta/p}] + \sup_{|x| \leq \sigma^{-2\beta/p}} |f_{M, \sigma}(x) - f_0(x)|^2 \leq C \sigma^{2\beta} (\log \sigma^{-1})^2.
\]

By equation (7), we have \( G_\sigma(U_\sigma) \gtrsim \sigma^{-b_3} \sigma^{b_4} \exp(-a_3/\sigma) \). Moreover, there is a separation of \( h_\sigma \sigma \) between two consecutive \( \mu_k \) and \( h_\sigma \sigma \ll \sigma \), thus all the \( V_k \) with \( k \in \Lambda \) are disjoint. By assumptions on \( G_\mu \) (see equation (8)), \( \alpha_k := \overline{\alpha} G_\mu(V_k) \gtrsim \sigma^{b_n(\beta+1)(1+|\mu_k|^{-b_k})} \) for all \( k \in \Lambda \). We also define \( \alpha^c := \alpha(V^c) \). For \( \sigma \) small enough, there is a constant \( C' > 0 \) not depending on \( \sigma \) such that \( \alpha^c > C' \). Moreover, since \( \alpha \) has finite variation we can assume without loss of generality that \( C' \leq \alpha^c \leq 1 \), otherwise we split \( V^c \) into disjoint parts, each of
them having $\alpha$-measure smaller than one. With $\epsilon_n^2 := C \sigma^{2\beta}(\log \sigma^{-1})^2$, using that $\Gamma(\alpha) \leq 2\alpha^{\alpha-1}$ for $\alpha \leq 1$, it follows the lower bound

$$\Pi(\text{KL}(f_0, \epsilon_n)) \geq G_\sigma(U_\sigma) \Pi(\mathcal{M}_\sigma) \gtrsim \sigma^{-b_3 + b_4 \beta} e^{-a_3 \sigma^{-1}} \frac{\sigma^\beta}{3c\Gamma(\alpha^2)} \prod_{k \in \Lambda} \left( \frac{\sigma^{\beta} e^{-2|u_k|}}{3c\Gamma(\alpha_k)} \right)$$

$$\gtrsim \exp \left\{ -K |\Lambda| \log \sigma^{-1} - a_3 \sigma^{-1} - 2 \sum_{k \in \Lambda} |u_k| - \sum_{k \in \Lambda} \log \frac{1}{\alpha_k} \right\}$$

$$\gtrsim \exp \left\{ -K |\Lambda| \log \sigma^{-1} - K \sigma^{-1} - \sum_{k \in \Lambda} \log \frac{1}{\alpha_k} \right\},$$

for a generic constant $K > 0$. From the definition of $\alpha_k$, it holds

$$\sum_{k \in \Lambda} \log \frac{1}{\alpha_k} \lesssim |\Lambda| \log \sigma^{-1} + \sum_{k \in \Lambda} \log (1 + |\mu_k|),$$

when $\sigma$ is small enough. Also,

$$\sum_{k \in \Lambda} \log (1 + |\mu_k|) = \sum_{k \in \Lambda} \log (1 + |\mu_k|) \mathbb{1}\{|\mu_k| \leq 1\} + \sum_{k \in \Lambda} \log (1 + |\mu_k|) \mathbb{1}\{|\mu_k| > 1\}$$

$$\leq |\{k \in \Lambda : |\mu_k| \leq 1\}| + |\Lambda| \log 2 + \sum_{k \in \Lambda} \log |\mu_k|$$

$$\leq 2 h^{-1}_\sigma \sigma^{-1} + 4 |\Lambda| \frac{2\beta}{p} \log \sigma^{-1} \lesssim |\Lambda| \log \sigma^{-1} + \sigma^{-1}$$

Because $|\Lambda| > \sigma^{-1}$ for $\sigma$ small enough, it follows from all of the above the existence of a constant $K' > 0$, depending only on $f$, $\varphi$ and $\Pi$, such that

$$\Pi(\text{KL}(f_0, \epsilon_n)) \geq \exp \left\{ -K'|\Lambda| \log \sigma^{-1} \right\}.$$

Then for an appropriate constant $C'''' > 0$, as a consequence of proposition 3, we can have $\Pi(\text{KL}(f_0, \epsilon_n)) \geq e^{-C''''n^2}$ if

$$\epsilon_n^2 = \begin{cases} C''''n^{-2\beta/(3\beta + 1)} (\log n)^{2-2\beta/(3\beta + 1)} & 0 < p \leq 2, \\ C''''n^{-2\beta/(2\beta + 1 + 2\beta/p)} (\log n)^{2-3\beta/(2\beta + 1 + 2\beta/p)} & p > 2. \end{cases}$$

4.1.2. **Sieve construction for location mixtures.** We construct the following sequence of subsets of $\mathcal{F}$, also called a sieve. With the notation $f_{M,\sigma}(x) := \int \varphi((x - \mu)/\sigma) dM(\mu)$,

$$\mathcal{F}_n(H, \epsilon) := \left\{ f = f_{M,\sigma} : \sum_{i=1}^\infty |u_i|/n \leq M, \sum_{i=1}^\infty |u_i| \mathbb{1}\{|u_i| \leq n^{-1}\} \leq \epsilon \right\}.$$

The next two lemmas show that $\mathcal{F}_n(H, \epsilon)$ defined as above satisfies all the condition stated in equations (14) and (15) if $H$ and $\delta$ are chosen small enough.

**Lemma 4.** Let $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ be arbitrary and $d_n$ be the empirical $L^2$-distance associated with $x$. Then for any $n^{-1/2} < \epsilon_n \leq 1$, $0 < H \leq 1$ and $n$ sufficiently large there is a constant $C > 0$ not depending on $n$ such that

$$\log N(\epsilon_n, \mathcal{F}_n(H, \epsilon_n), d_n) \leq CHn \epsilon_n^2.$$
Proof. We write $F_n \equiv F_n(H, \epsilon_n)$ to ease notations. The proof is based on arguments from Shen et al. (2013), it uses the fact that the covering number $N(\epsilon_n, F_n, d_n)$ is the minimal cardinality of an $\epsilon_n$-net over $(F_n, d_n)$. We recall that $(F_n, d_n)$ has $\epsilon_n$-net $F_n,\epsilon$, if for any $f \in F_n$ we have $m \in F_n,\epsilon$ such that $d_n(f, m) < \epsilon_n$. Let $S_n := \cup_{i=1}^n \{ x : |x - x_i| \leq n^{1/b_1} \sqrt{6 \log n} \}$, $R_n := \{ \mu \in \mathbb{R} : \mu = k/n^{3/2+1/b_2}, k \in \mathbb{Z}, \mu \in S_n \}$ and,

$$F_n,\epsilon := \left\{ f = \sum_{i \in \mathcal{I}} u_i \varphi \left( \frac{x - \mu_i}{\sigma} \right) : \begin{array}{l}
|\mathcal{I}| \leq H n \epsilon_n^2 / \log n, n^{-1/b_2} \leq \sigma \leq n^{1/b_1} \\
\forall i \in \mathcal{I} : |u_i| \leq n, \mu_i \in R_n \\
u_i = k n^{-3/2} H^{-1}, k \in \mathbb{Z}, \sigma = k n^{3/2+1/b_2}, k \in \mathbb{N},
\end{array} \right\}.$$

We claim that there is a constant $\delta > 0$ such that $F_n,\epsilon$ is a $\delta \epsilon$-net over $(F_n, d_n)$. Indeed, let $f \in F_n$ be arbitrary, so that $f = \sum_{i=1}^\infty u_i \varphi((\cdot - \mu_i)/\sigma)$. We define $\mathcal{J} := \mathbb{N} \cup \{ \infty \}, \mathcal{K} := \{ i : |u_i| > n^{-1} \}$, and $\mathcal{L} := \{ i : \mu_i \in S_n \}$. Now choose $\mathcal{I} = \mathcal{J} \cap \mathcal{K} \cap \mathcal{L}$, and notice that $|\mathcal{I}| \leq |\mathcal{K}| \leq H n \epsilon_n^2 / \log n$. Hence we can pick a $m \in F_n,\epsilon$ with $m(x) = \sum_{i \in \mathcal{I}} u_i ' \varphi((x - \mu_i')/\sigma')$. Moreover, for any $j = 1, \ldots, n$

$$|f(x_j) - m(x_j)| \leq \sum_{\mathcal{J} \cap \mathcal{K} \cap \mathcal{L}^c} |u_i| \varphi((x_j - \mu_i)/\sigma) + \sum_{\mathcal{J} \cap \mathcal{K} \cap \mathcal{L}^c} |u_i| \varphi((x_j - \mu_i)/\sigma) \\
+ \sum_{i \in \mathcal{I}} |u_i| \varphi((x_j - \mu_i)/\sigma) - \varphi((x_j - \mu_i)/\sigma') \\
+ \sum_{i \in \mathcal{I}} |u_i - u_i'| \varphi((x_j - \mu_i')/\sigma').$$

The fourth term in the rhs of the last equation is bounded above by $\epsilon_n$. Regarding the third term, for any $i \in \mathcal{L}^c$ we have $|x_j - \mu_i|/\sigma > \sqrt{6 \log n}$ for all $j = 1, \ldots, n$. Then the third term is bounded by $|\mathcal{K}| \varphi(\sqrt{6 \log n}) \leq H n \epsilon_n^2 n^{-2} / \log n \leq \epsilon_n$. Since we can always choose $m \in F_n,\epsilon$ with $|u_i - u_i'| \leq n^{-3/2} H^{-1}$ for all $i \in \mathcal{I}$, $|\mu_i - \mu_i'| \leq n^{-3/2-1/b_2}$ for all $i \in \mathcal{I}$, and $|\sigma - \sigma'| \leq n^{-3/2-1/b_2}$, it follows from proposition 11

$$|f(x_j) - m(x_j)| \leq 2 \epsilon_n + \sum_{i \in \mathcal{I}} |u_i - u_i'| + \sum_{i \in \mathcal{I}} |u_i| \varphi((x_j - \mu_i)/\sigma) - \varphi((x_j - \mu_i)/\sigma') \\
\leq 2 \epsilon_n + \sum_{i \in \mathcal{I}} |u_i - u_i'| + 4 \sum_{i \in \mathcal{I}} |u_i| |\sigma_i - \sigma_i'| / \sigma_i + \sum_{i \in \mathcal{I}} |u_i| \left| \frac{\mu_i - \mu_i'}{\sigma_i} \right| \leq 8 \epsilon_n,$$

for all $j = 1, \ldots, n$. Therefore $d_n(f, m) \leq 8 \epsilon_n$, and the claim is proved with $\delta := 8$. To finish the proof, it suffices to compute the cardinality of $F_n,\epsilon$.

A straightforward computation shows that $|R_n| \leq n^{5/2+1/b_1+1/b_2} \sqrt{6 \log n} \leq n^{4/1+b_1+1/b_2}$ for all $n \geq 1$, then

$$\log N(c_3 \epsilon_n, F_n, d_n) \leq |\mathcal{I}| \log \left( \frac{n}{n^{3/2}} \times n^{4+1/b_1+1/b_2} \right) + \log \left( \frac{n^{1/b_1}}{n^{-3/2-1/b_2}} \right) \leq H \left( \frac{11}{2} + \frac{2}{b_1} + \frac{2}{b_2} \right) n \epsilon_n^2,$$
where the last line holds when $n$ becomes large enough. Then the lemma is proved with $C := (11/2 + 2/b_1 + 2/b_2)/64$. □

**Lemma 5.** Assume that there is $n_0 \in \mathbb{N}$, and $0 < \gamma_1 \leq \gamma_2 < 1$ such that $n^{-\gamma_2/2} \leq \epsilon_n \leq n^{-\gamma_1/2}$ for all $n \geq n_0$. Then $\Pi(\mathcal{F}_n(H, \epsilon_n)) \leq \exp\left(-\frac{H}{4}(1 - \gamma_2)n\epsilon_n^2\right)$ for all $n \geq n_0$.

**Proof.** We use the fact that $M \sim \Pi_\alpha$ is almost surely purely-atomic (Kingman, 1992). Then from the definition of $\mathcal{F}_n$ it follows

$$
\Pi(\mathcal{F}_n^c) \leq G_\sigma(\sigma \leq n^{-1/b_2}) + G_\sigma(\sigma > n^{1/b_1}) + \Pi_\alpha\left(\sum_{i=1}^{\infty} |u_i| > n\right) \\
+ \Pi_\alpha\left(\sum_{i=1}^{\infty} |u_i| \mathbb{1}\{|u_i| \leq n^{-1}\} > \epsilon_n\right) \\
+ \Pi_\alpha\left(|\{i : |u_i| > n^{-1}\}| > Hn\epsilon_n^2/\log n\right).
$$

We bound each of the term as follows. By assumption $G_\sigma(\sigma \leq n^{-1/b_2}) \lesssim e^{-a_2 n}$ and $G_\sigma(\sigma > n^{1/b_1}) \lesssim e^{-a_2 n}$. Notice that $\sum_{i=1}^{\infty} |u_i| = |M|$, where $|M|$ denote the total variation of the measure $M$. Since by definition we have $M \overset{d}{=} M_1 - M_2$, with $M_1, M_2$ independent Gamma random measures with same base measure $\alpha(\cdot)$, it follows that $|Q|$ has the distribution of a Gamma random variable with shape parameter $2\alpha$. Then by Markov’s inequality,

$$
\Pi_\alpha\left(\sum_{i=1}^{\infty} |u_i| > n\right) = \Pi_\alpha\left(e^{1/2|M|} > e^{2n}\right) \leq 2^{2\alpha}e^{-\frac{1}{2}n}.
$$

Also, by the superposition theorem (Kingman, 1992, section 2), for any $M \sim \Pi_\alpha$ we have $M \overset{d}{=} M_3 + M_4$, where $M_3$ and $M_4$ are independent random measures with total variation $|M_3|$ and $|M_4|$ having Laplace transforms (for all $t \in \mathbb{R}$ for which the integrals in the expressions converge)

$$
Ee^{t|M_3|} := \exp\left\{\frac{2\alpha}{\Gamma(1/\gamma)}\int_{1/n}^{\infty} (e^{tx} - 1)x^{-1}e^{-x} \, dx\right\},
$$

$$
Ee^{t|M_4|} := \exp\left\{\frac{2\alpha}{\Gamma(1/\gamma)}\int_{0}^{1/n} (e^{tx} - 1)x^{-1}e^{-x} \, dx\right\}.
$$

$M_3$ and $M_4$ are almost-surely purely atomic, $M_3$ has only jumps greater than $1/n$ (almost surely) which number is distributed according to a Poisson distribution with intensity $2\alpha E_1(1/n)$, where $E_1$ denotes the exponential integral $E_1$ function: $E_1(x) = \int_{x}^{\infty} \frac{e^{-t}}{t} \, dt$. Likewise, $M_4$ has only jumps smaller or equal to $1/n$ (almost-surely) which number is almost-surely infinite. Recalling that $E_1(x) = \gamma + \log(1/x) + o(1)$ for $x$ small, it holds $2\alpha \gamma \leq 2\alpha E_1(1/n) \leq 6\alpha \log n \leq x_n$ for $n$ sufficiently large, with $x_n := Hn\epsilon_n^2/\log n$. Thus using Chernoff’s bound on Poisson distribution, we get

$$
\Pi_\alpha\left(|\{i : |u_i| > n^{-1}\}| > Hn\epsilon_n^2/\log n\right) \leq e^{-2\alpha E_1(1/n)}\left(e^{2\alpha E_1(1/n)}\right)^{x_n} \\
\leq \exp\left\{-\frac{1}{2}x_n \log x_n\right\}.
$$
But, $\log x_n = \log n + \log H - 2 \log \epsilon_n^{-1} - \log \log n \geq (1 - \gamma_2) \log n + \log H - \log \log n \geq \frac{1}{2}(1 - \gamma_2) \log n$ for large $n$. Therefore, as $n \to \infty$

$$\Pi_\alpha \left( \{ i : |u_i| > n^{-1} \} \right) < H ne_n^2 / \log n \leq \exp \left\{ -\frac{H}{4}(1 - \gamma_2)n \epsilon_n^2 \right\}.$$  

Finally, we use again Markov’s inequality to get

$$\Pi_\alpha \left( \sum_{i=1}^{\infty} |u_i| \mid \{ |u_i| \leq n^{-1} \} > \epsilon_n \right) = \Pi_\alpha \left( e^{n \epsilon_n |M_i|} > e^{n \epsilon_n^2} \right) \leq e^{-n \epsilon_n^2} \exp \left\{ 2\pi \int_0^{1/n} (e^{n \epsilon_n x} - 1)x^{-1} e^{-x} dx \right\}.$$  

But for $x \in (0, 1/n)$, we have $e^{n \epsilon_n x} - 1 \leq n(e^{n \epsilon_n \delta_n} - 1)x$, thus the integral in the previous expression is bounded by $2\pi(e^{-n} - 1)$, which is in turn bounded by $2\pi(e - 1)$ because $\epsilon_n \leq 1$ if $n \geq n_0$.  

4.2. Case of the location-scale mixture.

4.2.1. Kullback-Leibler condition. By Chebychev inequality, we have $Q_0[-\zeta_j, \zeta_j] \leq \zeta_j^p Q_0|X|^p$. Therefore, bringing together results from propositions 7 and 8,

$$\int |f_M(x) - f_0(x)|^2 \, dQ_0(x)$$

$$= \sum_{j=0}^J \int_{I_j} |f_M(x) - f_0(x)|^2 \, dQ_0(x) + \int_{A_0}^J |f_M(x) - f(x)|^2 \, dQ_0(x)$$

$$\leq J^3 \sum_{j=0}^J \sigma_j^{2\beta} Q_0(I_j) + J^3 Q_0(A_0^c).$$

Then we can find a constant $C > 0$ such that $\int |f_M(x) - f_0(x)|^2 \, dQ_0(x) \leq CJ^3 \sigma_j^{2\beta}$ for all $M \in \mathcal{M}$ and $J$ large enough.

By equation (7), we have $G, \mu_j \geq \sigma_j^{-b_j} \sigma_j^{b_j \beta} \exp(-a_3/\sigma_j)$ for all $j = 0, \ldots, J$. Moreover, there is a separation of $h, j, \sigma_j$ between two consecutive $\mu_j$ and $h, j, \sigma_j$ such that all the $W_{jk}$ with $(j, k) \in \Lambda$ are disjoint. By equation (8), we have $\alpha_{jk} := \alpha_{\mu_j} G_{\mu_j}(V_{jk}) \geq \sigma_j^{b_j(\beta+1)+b_j} \exp(-a_3/\sigma_j)(1 + \mu_j)_{-b_0}$ for all $(j, k) \in \Lambda$. We also define $\alpha^c := \alpha(W^C)$. For $J$ large enough, there is a constant $C' > 0$ not depending on $J$ such that $\alpha^c > C'$. Moreover, since $\alpha$ has finite variation we can assume without loss of generality that $C' \leq \alpha^c \leq 1$, otherwise we split $W^c$ into disjoint parts, each of them having $\alpha$-measure smaller than one. With $\epsilon_n^2 := CJ^3 \sigma_j^{2\beta}$, using that $\Gamma(\alpha) \leq 2\alpha^{\alpha-1}$ for $\alpha \leq 1$ and $\mathcal{M} \subset KL(f_0, \epsilon_n)$, it follows the lower bound

$$\Pi(\mathcal{KL}(f_0, \epsilon_n)) \geq \frac{\sigma_j^{2\beta}}{3e \Gamma(\alpha^c)} \prod_{(j, k) \in \Lambda} \left( \frac{\sigma_j^{2\beta} e^{-2|u_{jk}|}}{3e \Gamma(\alpha_{jk})} \right)$$

(21)
\[ \geq \frac{\sigma_j^\beta}{3e\Gamma(\alpha^e)} \prod_{(j,k) \in \Lambda} \exp \left\{ -2|u_{jk}| - \beta \log \sigma_j^{-1} + \log \frac{1}{6e} + (\alpha_{jk} - 1) \log \alpha_{jk} \right\} \]

\[ \geq \exp \left\{ -KJ|\Lambda| - 2\sum_{(j,k) \in \Lambda} |u_{jk}| - \sum_{(j,k) \in \Lambda} \log \alpha_{jk}^{-1} \right\}, \]

for a constant \( K > 0 \) depending only on \( C \) and \( \beta \). We now evaluate the sums involved in the rhs of equation (21). As before, be have that \( \sum_{(j,k) \in \Lambda} |u_{jk}| \leq 4\|f_0\|_1\sigma_j^{-1} \) (see for instance the proof of proposition 8). Act as in section 4.1.1 to find that

\[ \sum_{(j,k) \in \Lambda} \log \alpha_{jk}^{-1} \leq J|\Lambda| + J^{3/2}\sigma_j^{-1} + |\Lambda|\sigma_j^{-1}. \]

The term proportional to \( |\Lambda|\sigma_j^{-1} \) is entirely responsible for the bad rates in location-scale mixtures, and the aim of the hybridation of next section is to get rid of it. For a constant \( K' > 0 \),

\[ \Pi(\text{KL}(f_0, \epsilon_n)) \geq \exp \left\{ -K'|\Lambda|\sigma_j^{-1} \right\}. \]

Then for an appropriate constant \( C' > 0 \) we can have \( \Pi(\text{KL}(f_0, \epsilon_n)) \geq e^{-c_2n\epsilon_n^2} \) if

\[ \epsilon_n^2 = \begin{cases} C'[n^{-2\beta/(3\beta+2)}(\log n)^t_1 \land n^{-2\beta/(2\beta+1+2\beta/p)}(\log n)^{t_2}], & p \leq 2\beta, \\ C'n^{-\beta/(\beta+1)}(\log n)^{t_3}, & p > 2\beta, \end{cases} \]

where \( t_1 := 4-8\beta/(3\beta+2) \), \( t_2 := 4-4\beta/(2\beta+1+2\beta/p) \) and \( t_3 := 4-2\beta/(\beta+1) \).

4.2.2. Sieve construction. Using the notation \( f_M(x) := \int \varphi((x-\mu)/\sigma) dM(\sigma, \mu) \), we construct the following sieve.

(22) \( F_n(H, \epsilon) := \left\{ f = f_M, M = \sum_{i=1}^\infty u_i \delta_{\sigma_i, \mu_i}, \sum_{i=1}^\infty |u_i| \leq \epsilon, \sum_{i=1}^\infty |u_i| \mathbb{1}\{|u_i| \leq n^{-1}\} \leq Hn/\log n, \sum_{i=1}^\infty |u_i| \mathbb{1}\{|u_i| \leq \epsilon\} \leq \epsilon, \sum_{i=1}^\infty |u_i| \mathbb{1}\{|u_i| > \epsilon\} \leq \epsilon \right\} \).

Lemma 6. Let \( x = (x_1, \ldots, x_n) \in \mathbb{R}^n \) be arbitrary and \( d_n \) be the empirical \( L^2 \)-distance associated with \( x \). Then for any \( n^{-1/2} < \epsilon_n \leq 1 \) \( 0 < H < 1 \) and \( n \) sufficiently large there is a constant \( C > 0 \) not depending on \( n \) such that \( \log N(\epsilon_n, F_n(H, \epsilon_n), d_n) \leq C Hn\epsilon_n^2 \).

The proof is almost identical to lemma 4, with the same constant \( C > 0 \).

Lemma 7. Assume that there is \( n_0 \in \mathbb{N} \), and \( 0 < \gamma_1 < \gamma_2 < 1 \) such that \( n^{-\gamma_1/2} \leq \epsilon_n \leq n^{-\gamma_2/2} \) for all \( n \geq n_0 \). Then \( \Pi(F_n(H, \epsilon_n)^c) \leq \exp(-\frac{H}{4}(1 - \gamma_2)n\epsilon_n^2) \) for all \( n \geq n_0 \).

Proof. We first write the estimate

\[ \Pi(F_n^c) \leq \Pi_\alpha \left( \sum_{i=1}^\infty |u_i| > n \right) + \Pi_\alpha \left( \sum_{i=1}^\infty |u_i| \mathbb{1}\{|u_i| \leq \delta\} > \epsilon_n \right) + \Pi_\alpha \left( \sum_{i=1}^\infty |u_i| \mathbb{1}\{|u_i| > \epsilon_n\} \right) \]

\[ + \Pi_\alpha \left( \sum_{i=1}^\infty |u_i| \mathbb{1}\{|u_i| > \delta, n^{-1/2} < \sigma_i \leq n^{1/2}\} > \epsilon_n \right) + \Pi_\alpha \left( \sum_{i=1}^\infty |u_i| \mathbb{1}\{|u_i| > \delta, n^{-1/2} < \sigma_i \leq n^{1/2}\} \right) > Hn\epsilon_n^2/\log n. \]

The first three terms in the rhs above obey the same bounds as in the proof of lemma 5, using the same arguments. The last two terms are bounded using the
same trick, thus we simply bound the last term and left the other to the reader. Notice that the random variable $U := \sum_{i=1}^{\infty} |u_i| \mathbb{1}\{\sigma_i > n^{1/b_1}\}$ has Gamma distribution with parameters $2\alpha(A_n), 1$, with $A_n := \{(\sigma, \mu) : \sigma > n^{1/b_1}\}$. For $n$ large, by assumptions on $P_\sigma$, it holds $\alpha(A_n) \ll \epsilon_n$. Then by Chebychev inequality, for $n$ large enough

$$\Pi_\alpha \left( \sum_{i=1}^{\infty} |u_i| \mathbb{1}\{\sigma_i > n^{1/b_1}\} > \epsilon_n \right) \leq \Pr(U - EU > \epsilon_n - EU) \leq \Pr(U - EU > \epsilon_n/2) \leq 16\epsilon_n^{-2} \alpha(A_n)^2.$$  

The conclusion follows from the assumptions on $G_\sigma$ which imply $\alpha(A_n) = \overline{\alpha}G_\sigma(\sigma > n^{1/b_1}) \leq \exp(-a_1 n)$. □

4.3. **Hybrid location-scale mixtures.** Obviously, given the definition of hybrid mixtures (see section 4.3), most of the proof is redundant with the location-scale case, and in the sequel we deal only with the parts that differ.

4.3.1. **Kullback-leibler condition.** Let $M \equiv \mathcal{M}(\beta, J, f, \Lambda)$ be the set of signed measures constructed in section 3.3. For any integer $J > 0$ let $\Omega_J$ be the event

$$\Omega_J := \left\{ P_\sigma : P_\sigma|_{2^{-J}, 2^{-J}(1 + 2^{-J})} \geq 2^{-J} \quad \forall 0 \leq j \leq J \right\}.$$  

Then with arguments and constant $C > 0$ from section 4.2.1, letting $c_n^2 := CJ^4 \sigma_j^{2\beta}$, we have

$$\Pi(\KL(f_0, \epsilon_n)) \geq \Pi(M) \geq \Pi(M | \Omega_J) \Pi_\sigma(\Omega_J).$$  

But by equation (11) we have $\Pi_\sigma(\Omega_J) \geq \exp(-a_6 J^\beta 2^J)$ and on $\Omega_J$ it holds $\alpha(W_{jk}) = \overline{\sigma}P_\sigma(U_j)G_\mu(V_{jk}) \geq \overline{\sigma} \frac{1}{2^J} G_\mu(V_{jk})$ for all $(j, k) \in \Lambda$. Then act as in equation (21) to find a constant $K > 0$ such that (recalling that $\sigma_j = 2^{-J}$)

$$\Pi(\KL(f_0, \epsilon_n)) \geq \exp \left\{ -K(j^{\beta}) \sqrt{J^{1/2}} \sigma_j^{-1} - KJ|\Lambda| \right\}.$$  

Because of proposition 6 we can have $\Pi(\KL(f_0, \epsilon_n)) \geq e^{-c_2 n^2}$ if for an appropriate constant $C' > 0$

$$\epsilon_n^2 = \begin{cases} C'[n^{-2\beta/(3\beta+1)}(\log n)^{-6\beta/(3\beta+1)} \wedge n^{-p/(p+1)}(\log n)^{-4-p/(p+1)}] & p \leq 2\beta, \\ C'[n^{-2\beta/(2\beta+1)}(\log n)^{-2-3\beta/(2\beta+1)}] & p > 2\beta. \end{cases}$$

4.3.2. **Sieve construction.** We use the same sieve $F_n(H, \epsilon)$ as in equation (22). The definition of $F_n(j, \epsilon)$ is independent of $\Pi$ thus the conclusion of lemma 4 holds for hybrid location-scale mixtures. It remains to show that $\Pi(\mathcal{F}_n(H, \epsilon^n)) \leq \exp(-2c_2 n^2)$, which is the object of the next lemma.

**Lemma 8.** Assume that there is $n_0 \in \mathbb{N}$, and $0 < \gamma_1 \leq \gamma_2 < 1$ such that $n^{-\gamma/2} \leq \epsilon_n \leq n^{-\gamma_1/2}$ for all $n \geq n_0$. Then there is a constant $\gamma_2 < \gamma < 1$ such that $\Pi(\mathcal{F}_n(H, \epsilon^n)) \lesssim \exp(-\frac{H}{4}(1 - \gamma)n^{\gamma_2})$ for all $n \geq n_0$.

**Proof.** We proceed as in the proof of lemma 7. Following the same steps, we deduce that it is sufficient to prove that

$$\Pi_\alpha \left( \sum_{i=1}^{\infty} |u_i| \mathbb{1}\{\sigma_i > n^{1/b_1}\} > \epsilon_n \right) \lesssim e^{-2c_2 n},$$

$$\Pi_\alpha \left( \sum_{i=1}^{\infty} |u_i| \mathbb{1}\{\sigma_i \leq n^{-1/b_2}\} > \epsilon_n \right) \lesssim e^{-2c_2 n}.$$
Since the proofs are almost identical for the two previous conditions, we only prove the first and left the second to the reader. Notice that by equation (23) we have
\[ \Pi_\alpha \left( \sum_{i=1}^{\infty} |u_i| \mathbb{1}\{\sigma_i > n^{1/b_1}\} > \epsilon_n \right) \leq 16\pi c^{-2} P_\sigma(\sigma > n^{1/b_1})^2. \]

Letting \( \Omega := \{ P_\sigma : P_\sigma(\sigma > n^{1/b_1}) < \exp(-a_1 n/2) \} \), with a slight abuse of notation, it follows from equation (9)
\[ \Pi_\alpha \left( \sum_{i=1}^{\infty} |u_i| \mathbb{1}\{\sigma_i > n^{1/b_1}\} > \epsilon_n \right) \leq \Pi_\alpha \left( \sum_{i=1}^{\infty} |u_i| \mathbb{1}\{\sigma_i > n^{1/b_1}\} > \epsilon_n \right| \Omega \right) + \Pi_\sigma(\Omega^c) \]
\[ \lesssim \epsilon_n^2 \exp(-a_1 n) + \exp(-a_4 n). \]

5. Proof of Theorem 2

The proof follows the same lines as Ghosal et al. (2007b) with additional cares. The first step consists on rewriting expectation of the posterior distribution as follows. Let \( (\phi_n(\cdot \mid \cdot))_{n \geq 0} \) be a sequence of test functions such that for \( n \) large enough
\[ Q_0^n [P_0^n[\phi_n(y \mid x) \mid x]] \lesssim N(\epsilon/18, F_n, d_n) \exp \left( -\frac{n \epsilon_n^2}{2} \right), \]
\[ \sup_{\{f : d_n(f, f_0) \geq 17\epsilon_n/18\} \cap F_n} Q_0^n [P_f^n[1 - \phi_n(y \mid x) \mid x] \lesssim 2^n \exp \left( -\frac{n \epsilon_n^2}{2} \right). \]

The existence of such test functions is standard and follows for instance from Birgë (2006, proposition 4), or Ghosal and van der Vaart (2007, section 7.7). From here, we bound the posterior distribution in a standard fashion,
\[ Q_0^n [P_0^n[\Pi_x(\{ f : d_n(f, f_0) > \epsilon_n \} \mid y, x) \mid x]] \leq Q_0^n[P_0^n[\Pi_x(F_n^c) \mid y, x) \mid x] + Q_0^n[P_0^n[\Pi(\{ f : d_n(f, f_0) > \epsilon_n \} \cap F_n \mid y, x) \mid x]]. \]

So that,
\[ Q_0^n [P_0^n[\Pi_x(\{ f : d_n(f, f_0) > \epsilon_n \} \mid y, x) \mid x]] \leq Q_0^n[P_0^n[\Pi_x(F_n^c) \mid y, x) \mid x] + Q_0^n[P_0^n[\phi_n(y \mid x)\Pi_x(\{ f : d_n(f, f_0) > \epsilon_n \} \cap F_n \mid y, x) \mid x]] + Q_0^n[P_0^n[(1 - \phi_n(y \mid x))\Pi_x(\{ f : d_n(f, f_0) > \epsilon_n \} \cap F_n \mid y, x) \mid x]]. \]

Now, to any \( x \in \mathbb{R}^n \), we associate the event
\[ E_n(x) := \left\{ y \in \mathbb{R}^n : \int \prod_{i=1}^{n} p_{f_0}(x_i, y_i) / p_{f}(x_i, y_i) d\Pi_x(f) \geq \exp \left( -(1 + 4c_2)^n \epsilon_n^2 / 4 \right) \right\}. \]

Consider the first term of the rhs of equation (24). We can rewrite,
\[ Q_0^n[P_0^n[\Pi_x(F_n^c) \mid y, x) \mid x] \]
\[
\leq e^{\frac{\chi}{4} (4c_2 + 1) \epsilon_n^2} \int_{\mathbb{R}^n} \int_{E_n(x)} \prod_{i=1}^n \frac{p_f(x_i, y_i)}{p_{f_0}(x_i, y_i)} d\Pi_x(f) dP_0^n(y | x) dQ_0^n(x)
\]
\[
+ \int_{\mathbb{R}^n} \int_{E_n(x)^c} dP_0^n(y | x) dQ_0^n(x)
\]
\[
= e^{\frac{\chi}{4} (4c_2 + 1) \epsilon_n^2} \int_{\mathbb{R}^n} \int_{E_n(x)} dP_0^n(y | x) dQ_0^n(x)
\]
\[
+ \int_{\mathbb{R}^n} \int_{E_n(x)^c} dP_0^n(y | x) dQ_0^n(x)
\]
\[
\leq e^{\frac{\chi}{4} (4c_2 + 1) \epsilon_n^2} \int_{\mathbb{R}^n} \Pi_x(F_n^c) dQ_0^n(x) + \int_{\mathbb{R}^n} \int_{E_n(x)^c} dP_0^n(y | x) dQ_0^n(x),
\]
where the third line follows from Fubini’s theorem. The same reasoning applies to the other terms of equation (24), using the test functions introduced above and \( 0 < c_2 < 1/4 \). Hence the theorem is proved if we show that
\[
\int_{\mathbb{R}^n} \int_{E_n(x)} dP_0^n(y | x) dQ_0^n(x) = o(1).
\]
But under the condition of the theorem, Ghosal et al. (2007b, Lemma 10) implies that
\[
P_0^n \left( \int_{\mathbb{R}^n} \prod_{i=1}^n \frac{p_f(x_i, Y_i)}{p_{f_0}(x_i, Y_i)} d\Pi_x(f) < \exp \left( -\frac{1}{4} (1 + 4c_2) \epsilon_n^2 \right) | x \right) = o(1).
\]

APPENDIX A. PROOFS OF LEMMAS 2 AND 3 AND SOME TECHNICAL RESULTS ON THE KERNELS

A.1. Proof of Lemma 2. Clearly, \( \|\chi \ast f\|_1 \leq \|\chi\|_1 \|f\|_1 \) by Young’s inequality, so that \( \chi \ast f \in L^1 \) and \( (\chi \ast f)^\wedge (\xi) = \hat{\chi}(\xi) \hat{f}(\xi) \), showing that the support of the Fourier transform of \( \chi \ast f \) is included in \( [-1/\sigma, 1/\sigma] \). Moreover, using again Young’s inequality we get that \( \|\chi \ast f\|_\infty \leq \|\chi\|_1 \|f\|_\infty \), thus \( \chi \ast f \in L^\infty \).

Because \( \hat{\chi} \) is \( C^\infty \) and compactly supported, for any integer \( q \geq 0 \) we have
\[
(iu)^q \chi(u) = (2\pi)^{-1} \int \hat{\chi}^{(q)}(\xi) e^{iu\xi} d\xi.
\]
Clearly \( \hat{\chi} \) is Schwartz, hence by Fourier inversion we have that
\[
\int u^q \chi(u) e^{-i\xi u} du = (-i)^q \hat{\chi}^{(q)}(\xi), \quad \forall \xi \in \mathbb{R}.
\]
But, by construction \( \hat{\chi}(0) = 1 \), and for any \( q \geq 1 \) we have \( \hat{\chi}^{(q)}(0) = 0 \). It follows that \( \int \chi(u) du = 1 \), and \( \int u^q \chi(u) du = 0 \) for any \( q \geq 1 \). Whence, letting \( m \) be the largest integer smaller than \( \beta \), and using Taylor’s formula with exact remainder term
\[
\chi \ast f(x) - f(x)
\]
\[
\int \chi(y) [f(x - y) - f(x)] \, dy = \int \chi(y) [f(x - \sigma y) - f(x)] \, dy
\]
\[
= \sum_{k=1}^{m} \frac{(-1)^k \sigma^k}{k!} \int u^k \chi(u) \, du
\]
\[
+ \int \chi(y) \int_{0}^{1} (1 - u)^{m-1} \left[ f^{(m)}(x - u\sigma y) - f^{(m)}(x) \right] \, du \, dy
\]
\[
= \int \chi(y) \int_{0}^{1} (1 - u)^{m-1} \left[ f^{(m)}(x - u\sigma y) - f^{(m)}(x) \right] \, du \, dy.
\]

Therefore, because \( f \in C^\beta \),
\[
|\chi \ast f(x) - f(x)|
\]
\[
\leq \sigma^m \int |y^m \chi(y)| \int_{0}^{1} \frac{(1 - u)^{m-1}}{(m-1)!} |f^{(m)}(x - u\sigma y) - f^{(m)}(x)| \, du \, dy
\]
\[
\leq \|f\|_{C^\beta} \sigma^m \int |y^m \chi(y)| \, dy \int_{0}^{1} \frac{(1 - u)^{m-1}}{(m-1)!} u^{\beta - m} \, du.
\]

**A.2. Proof of lemma 3.** We mostly follow the proof of Hangelbroek and Ron (2010, proposition 1). Writing,
\[
K_{h,\sigma} f_\sigma(x) = \int \frac{h}{\sigma} \sum_{k \in \mathbb{Z}} \varphi \left( \frac{x - h\sigma k}{\sigma} \right) \eta \left( \frac{y - h\sigma k}{\sigma} \right) \, f_\sigma(y) \, dy
\]
\[
= \frac{h}{\sigma} \sum_{k \in \mathbb{Z}} \varphi \left( \frac{x - h\sigma k}{\sigma} \right) \int \eta \left( \frac{y - h\sigma k}{\sigma} \right) \, f_\sigma(y) \, dy
\]
\[
= \frac{h}{2\pi} \sum_{k \in \mathbb{Z}} \varphi \left( \frac{x - h\sigma k}{\sigma} \right) \int \hat{\eta}(\sigma \xi) \hat{f}_\sigma(\xi) e^{i\xi h\sigma k} \, d\xi
\]
\[
= \int \hat{\eta}(\sigma \xi) \hat{f}_\sigma(\xi) \frac{h}{2\pi} \sum_{k \in \mathbb{Z}} \varphi \left( \frac{x - h\sigma k}{\sigma} \right) e^{i\xi h\sigma k} \, d\xi.
\]

Then we can invoke the *Poisson summation formula* (Härdle et al., 1998, theorem 4.1), which is obviously valid for \( \varphi \), and
\[
\sum_{k \in \mathbb{Z}} \varphi \left( \frac{x - h\sigma k}{\sigma} \right) e^{i\xi h\sigma k} = \frac{1}{h} \sum_{m \in \mathbb{Z}} \hat{\varphi} \left( \sigma \xi + \frac{2\pi m}{h} \right) e^{i(\sigma \xi + \frac{2\pi m}{h})x/\sigma}.
\]

Therefore, recalling that \( \hat{f}_\sigma \) is supported on \([-1/\sigma, 1/\sigma]\) and \( \hat{\chi} \) equals 1 on \([-1, 1]\),
\[
K_{h,\sigma} f_\sigma(x) = \frac{1}{2\pi} \int \hat{\chi}(\sigma \xi) \hat{f}_\sigma(\xi) \sum_{m \in \mathbb{Z}} \frac{\hat{\varphi}(\sigma \xi + \frac{2\pi m}{h})}{\hat{\varphi}(\sigma \xi)} e^{i(\sigma \xi + \frac{2\pi m}{h})x/\sigma} \, d\xi
\]
\[
= f_\sigma(x) + \frac{1}{2\pi} \sum_{m \in \mathbb{Z} \setminus \{0\}} \int \hat{f}_\sigma(\xi) \frac{\hat{\varphi}(\sigma \xi + \frac{2\pi m}{h})}{\hat{\varphi}(\sigma \xi)} e^{i(\sigma \xi + \frac{2\pi m}{h})x/\sigma} \, d\xi.
\]
It follows that,
\[ |K_{h,\sigma}f_\sigma(x) - f_\sigma(x)| \leq \frac{1}{2\pi} \|\hat{f}_\sigma\|_1 \sup_{\xi \in [-1,1]} \sum_{m \in \mathbb{Z}\setminus\{0\}} \left| \frac{\hat{\varphi}(\xi + 2\pi m/h)}{\hat{\varphi}(\xi)} \right|. \]

Now, \( \|\hat{f}_\sigma\|_1 \leq 2\sigma^{-1}\|\hat{f}_\sigma\|_\infty \leq 2\sigma^{-1}\|f_\sigma\|_1 \leq 2\sigma^{-1}\|f\|_1 \), which is finite because of lemma 2. Recalling that by assumption \( \hat{\varphi} \) is Gaussian, it follows for all \( \xi \in [-1,1] \) and all \( h \leq 1 \),
\[ \sum_{m \in \mathbb{Z}\setminus\{0\}} \left| \frac{\hat{\varphi}(\xi + 2\pi m/h)}{\hat{\varphi}(\xi)} \right| \leq \exp \left\{ -\frac{1}{2}(\xi + 2\pi m/h)^2 + \frac{1}{2}\xi^2 \right\} \leq e^{-1/2} \sum_{m \in \mathbb{Z}\setminus\{0\}} e^{-4\pi^2 m^2/h^2} \leq 4e^{-1/2}e^{-4\pi^2/h^2}. \]

Then the lemma is proved with \( C := 8e^{-1/2} \).

A.3. Some other technical results on \( K_{h,\sigma} \).

**Lemma 9.** There is a universal constant \( C > 0 \) such that for all \( x \in \mathbb{R} \), all \( 0 < h \leq 1 \) and all \( \sigma > 0 \), \( \sum_{k \in \mathbb{Z}} |\eta((x - h\sigma k)/\sigma)| \leq Ch^{-1} \). Moreover, \( \eta \in \mathcal{S} \).

**Proof.** We first prove that \( \eta \in \mathcal{S} \). Obviously \( \hat{\eta} \in \mathcal{S} \), and therefore so is \( \hat{\eta} \). Since the Fourier transform and the inverse Fourier transform are continuous mapping of \( \mathcal{S} \) onto itself, it is immediate that \( \eta \in \mathcal{S} \).

We finish the proof by remarking that \( x \mapsto \sum_{k \in \mathbb{Z}} |\eta((x - h\sigma k)/\sigma)| \) is periodic with period \( h\sigma \), hence it suffices to check that it is bounded for \( x \in [0,h\sigma] \). If \( x \in [0,h\sigma] \), then \( |x - h\sigma k| \geq |h\sigma k|/2 \) for any \( |k| \geq 2 \), so that
\[
\sum_{k \in \mathbb{Z}} |\eta((x - h\sigma k)/\sigma)| \leq 3 \sup_{u \in \mathbb{R}} |\eta(u)| + \sum_{|k| \geq 2} |\eta((x - h\sigma k)/\sigma)| \leq 3\|\eta\|_{0,0} + \|\eta\|_{2,0} \sum_{|k| \geq 2} (1 + |hk|/2)^{-2} \leq 3\|\eta\|_{0,0} + 4\|\eta\|_{2,0}/h,
\]
which concludes the proof of the first assertion with \( C := 3\|\eta\|_{0,0} + 4\|\eta\|_{2,0} \), because of the assumption \( h \leq 1 \). \( \square \)

The following Lemma gives some control on the coefficients of \( f \) on \( \eta \).

**Proposition 9.** Let \( 0 < h \leq 1 \) and \( a_k(f) := (h/\sigma) \int \eta((y - h\sigma k)/\sigma)f(y) dy \).
Then there are universal constants \( C, C' > 0 \), depending only on \( \varphi \), such that \( \sum_{k \in \mathbb{Z}} |a_k(f)| \leq C\|f\|_1\sigma^{-1} \), and for all \( k \in \mathbb{Z} \), \( |a_k(f)| \leq C'\|f\|_\infty \).

**Proof.** For the first assertion of the proposition, we write,
\[
\sum_{k \in \mathbb{Z}} |a_k(f)| \leq \frac{h}{\sigma} \sum_{k \in \mathbb{Z}} \int |f(y)| |\eta((y - h\sigma k)/\sigma)| dy \leq \sigma^{-1}\|f\|_1 \sup_{y \in \mathbb{R}} h \sum_{k \in \mathbb{Z}} |\eta((y - h\sigma k)/\sigma)|,
\]
and the conclusion follows from lemma 9. The proof of the second assertion is simpler. Indeed,

\[ |a_k(f)| \leq \frac{h}{\sigma} \int |f(y)||\eta((y - h\sigma k)/\sigma)| \, dy \leq h\|f\|_{\infty} \int |\eta(u)| \, du, \]

where the last integral is bounded because \( \eta \in S \) by lemma 9. \( \square \)

**Appendix B. Proof of Lemma 1**

Let \( x \in \mathbb{R}^n \) arbitrary, \( \sigma > 0 \) and \( h_\sigma \sqrt{\log \sigma^{-1}} := 2\pi \sqrt{\beta + 1} \). Recall that from lemmas 2 and 3 we have \( \|K_{h_\sigma \sigma}(\chi_\sigma * f_0) - f_0\|_\infty \lesssim \sigma^\beta \), where \( K_{h_\sigma \sigma}(\chi_\sigma * f_0)(z) := \sum_{k \in \mathbb{Z}} u_k \varphi((z - h_\sigma \sigma k)/\sigma) \). Define \( S_n(x) := \bigcup_{i=1}^n \{ z \in \mathbb{R} : |z - x_i| \leq \sigma \sqrt{2(\beta + 1) \log \sigma^{-1}} \} \)

\[ \Lambda(x) := \{ k \in \mathbb{Z} : |u_k| > \sigma^\beta, \quad h_\sigma \sigma k \in S_n(x) \}. \]

Also define \( U_\sigma := \{ \sigma' : \sigma \leq \sigma' \leq \sigma(1 + \sigma^3) \} \), and for all \( k \in \Lambda(x) \) define \( V_k := \{ \mu : |\mu - h_\sigma \sigma k| \leq \sigma^\beta + 1 \} \). We denote by \( \mathcal{M}_\sigma \) the set of signed measures \( M \) on \( \mathbb{R} \) such that \( |M(V_k) - u_k| \leq \sigma^\beta \) for all \( k \in \Lambda(x) \) and \( |M(V) \leq \sigma^2 \), where \( V \) is the relative complement of the union of all \( V_k \) for \( k \in \Lambda(x) \). For any \( M \in \mathcal{M}_\sigma \), we write \( f_{M, \sigma}(z) := \int \varphi((z - \mu)/\sigma) \, dM(\mu) \). Act as in proposition 5 to find that \( d_\sigma(f, f_0) \leq C_{h_\sigma}^2 \sigma^\beta \) for any \( M \in \mathcal{M}_\sigma \), with a constant \( C > 0 \) not depending on \( x \). By construction of \( S_n(x) \), for all \( k \in \Lambda(x) \) there is at least one \( x_i \) such that \( |h_\sigma \sigma k - x_i| \leq \sigma \sqrt{2(\beta + 1) \log \sigma^{-1}} \). Then for any \( k \in \Lambda(x) \), by definition of \( G_x \)

\[ \bar{\sigma} G_x(V_k) \geq n^{-1} \int_{h_\sigma \sigma k - \sigma^\beta + 1}^{h_\sigma \sigma k + \sigma^\beta + 1} g(z - x_i) \, dz \geq a_1 n^{-1} \sigma^{a_2(\beta + 1)} \]

Remark that \( |\Lambda(x)| \lesssim \sigma^{-(\beta + 1)} \) independently of \( x \) (see proposition 3) and letting \( \epsilon_n = C' h_\sigma^{-2} \sigma^\beta \) we can mimic the steps of section 4.1.1 to find that

\[ \Pi_x(f : d_\sigma(f, f_0) \leq \epsilon_n) \gtrsim \exp \{ -C''|\Lambda(x)| \log \sigma^{-1} - C'''|\Lambda(x)| \log n \} \]

\[ \gtrsim \exp(-c_2 n \epsilon_n^2), \]

for a constant \( C'' > 0 \) not depending on \( x \) and \( \epsilon_n^2 \) defined in the lemma.

**Appendix C. Some technical results on the construction of the approximation in the case of location-scale mixtures**

**Proposition 10.** Let \( f_0 \in C_\beta \). For any \( j \geq 0 \), we have \( |\Delta_j(x)| \leq C \|f_0\|_{C\beta} \sigma_j^\beta \), with the same constant \( C > 0 \) as in lemma 2. Moreover, \( \|\Delta_j\|_1 \leq 2 \|f_0\|_1 \) for all \( j \geq 0 \).

**Proof.** Notice that \( \|\Delta_{j+1}\|_1 \leq \|\Delta_j\|_1 + \|\chi_{\sigma_{j+1}} * \Delta_j\| \leq (1 + \|\chi_1\|) \|\Delta_j\|_1 \), by Young’s inequality. Since \( f_0 \in L^1 \), this implies \( \Delta_j \in L^1 \) for all \( j \geq 0 \). Since \( \Delta_{j+1}(\xi) = \hat{\Delta}_j(\xi) - \hat{\chi}_{\sigma_{j+1}}(\xi) \hat{\Delta}_j(\xi) \), we get \( \hat{\Delta}_j(\xi) = \hat{f}_0(\xi) \prod_{l=1}^{j} (1 - \chi_{\sigma_l}(\xi)) \), by induction. Because \( \sigma_{j+1} = \sigma_j/2 \), and by construction of \( \chi_\sigma \) we have \( \hat{\chi}_{\sigma_m}(\xi) \hat{\chi}_{\sigma_l}(\xi) = \hat{\chi}_{\sigma_m}(\xi) \) for any \( m > l \), hence the last equation can be rewritten as \( \Delta_j(\xi) = f_0(\xi)(1 - \hat{\chi}_{\sigma_j}(\xi)) \). Then we deduce that \( \Delta_j = f_0 - \chi_{\sigma_j} * f_0 \). By
lemma 2, this implies that $|\Delta_j(x)| \leq C\|f_0\|_{L^2}\sigma_j^\beta$. From the same estimate, it is clear that $\|\Delta_j\| \leq \|f_0\|_1 + \|\chi_{\sigma_j} * f_0\| \leq 2\|f_0\|_1$. \hfill \Box

C.1. Proof of proposition 8. Let define $A(\beta, J) := (2 \log |\Lambda| + 2\beta \log \sigma_j^{-1})^{1/2}$ and $J \equiv J(x) := \{(j, k) \in \{0, \ldots, J\} \times \mathbb{Z} : |x - \mu_{jk}| \leq 4A(\beta, J)\sigma_j\}$. For any $M \in \mathcal{M}$ we can write

$$f_M(x) - f_0(x) = \sum_{(j,k)\in \Lambda \cap J} \int_{W_{jk}} \left[ \varphi \left( \frac{x - \mu}{\sigma} \right) - \varphi \left( \frac{x - \mu_{jk}}{\sigma_j} \right) \right] dM(\sigma, \mu)$$

$$+ \sum_{(j,k)\in \Lambda \cap J^c} \left[ M(W_{jk}) - u_{jk} \right] \varphi \left( \frac{x - \mu_{jk}}{\sigma_j} \right) + \sum_{(j,k)\in \Lambda \cap J^c} \int_{W_{jk}} \varphi \left( \frac{x - \mu}{\sigma} \right) dM(\sigma, \mu)$$

$$- \sum_{(j,k)\in \Lambda \cap J^c} u_{jk} \varphi \left( \frac{x - \mu_{jk}}{\sigma_j} \right) - \sum_{(j,k)\notin \Lambda} u_{jk} \varphi \left( \frac{x - \mu_{jk}}{\sigma_j} \right)$$

$$+ \int_{W^c} \varphi \left( \frac{x - \mu}{\sigma} \right) dM(\sigma, \mu)$$

$$:= r_1(x) + r_2(x) + r_3(x) + r_4(x) + r_5(x) + r_6(x).$$

The proof follows similar steps as the proof of proposition 5. From the definition of $A(\beta, J)$ and proposition 6, we deduce that $A(\beta, J) \lesssim \sqrt{J}$ for $J$ large enough. Also, there is a separation of $h_j\sigma_j$ between two consecutive $\mu_{jk}$. Then there are no more than $2A(\beta, J)\sigma_j/(h_j\sigma_j) = 2A(\beta, J)h_j^{-1}$ distinct values of $\mu_{jk}$ in an interval of length $2A(\beta, J)\sigma_j$. Thus the bound $|\Lambda \cap J| \leq 2(J + 1)A(\beta, J) \lesssim J^{3/2}$ holds. It follows from proposition 11 that $|r_1(x)| \lesssim |\Lambda \cap J|\sigma_j^\beta \lesssim J^{3/2}\sigma_j^\beta$.

Obviously, $|r_2(x)| \leq \|\varphi\|_\infty |\Lambda \cap J|\sigma_j^\beta \lesssim J^{3/2}\sigma_j^\beta$. Whenever $(j,k) \in \Lambda \cap J^c$ and $(\sigma, \mu) \in W_{jk}$, choosing $J$ large enough so that $1/2 \leq \sigma_j/\sigma \leq 2$ and $|\mu - \mu_{jk}| \leq \sigma_j A(\beta, J)/2$, it holds $|x - \mu| \geq A(\beta, J)\sigma$. Therefore, $|r_3(x)| \lesssim \varphi(A(\beta, J)\Lambda) \lesssim \sigma_j^\beta$. With the same reasoning we get $|r_4(x)| \lesssim \|f\|_\infty \sigma_j^\beta$. Regarding $r_6$, we have the obvious bound $|r_6(x)| \leq \|\varphi\|_\infty \sigma_j^\beta$. The $r_5$ term is more subtle and constitutes the remainder of the proof.

Let $\Lambda_1^j := \{(j,k) \in \{0, \ldots, J\} \times \mathbb{Z} : |u_{jk}| \leq \sigma_j^\beta\}$ and $\mathcal{K}_j := \{k \in \mathbb{Z} : |\mu_{jk}| > \zeta_j + \sqrt{2(\beta + 1) \log \sigma_j^{-1}}\}$. Assuming that $x \in [-\zeta_q, \zeta_q]$ for some $0 \leq q \leq J$, we can bound $r_5(x)$ as follows,

$$|r_5(x)| \leq \sum_{(j,k)\in \Lambda_1^j} |u_{jk}| \varphi \left( \frac{x - \mu_{jk}}{\sigma_j} \right)$$

$$+ \sum_{j \leq q} \sum_{k \in \mathcal{K}_j} u_{jk} \varphi \left( \frac{x - \mu_{jk}}{\sigma_j} \right) + \sum_{j > q} \sum_{k \in \mathcal{K}_j} u_{jk} \varphi \left( \frac{x - \mu_{jk}}{\sigma_j} \right),$$

where the third term of the rhs does not exist if $q = J$. The first term of the rhs of equation (26) is bounded by $\sigma_j^\beta \sup_{x \in \mathbb{R}} \sum_{j=0}^J \sum_{k \in \mathbb{Z}} \varphi((x - \mu_{jk})/\sigma)$, which is in turn bounded by a constant multiple of $J^{3/2}\sigma_j^\beta$ (see for instance the proof of
lemma 9). Because of propositions 9 and 10, when \( x \in [-\zeta_q, \zeta_q] \) we always have

\[
\sum_{j\leq q} \sum_{k \in \mathcal{K}_j} |u_{jk}| \varphi \left( \frac{x - \mu_{jk}}{\sigma_j} \right) \leq \sup_{j \leq q} \varphi \left( \frac{x - \mu_{jk}}{\sigma_j} \right) \sum_{j \leq J} \sum_{k \in \mathcal{K}_j} |u_{jk}| \\
\leq \sigma_j^{\beta+1} \sum_{j \leq J} 2\|f_0\|_1 \sigma_j^{-1} \leq 4\|f_0\|_1 \sigma_j^{\beta}.
\]

Regarding the second term of the rhs of equation (26), we introduce the sets of indexes \( \mathcal{L}_j \equiv \mathcal{L}_j(x) := \{ k \in \mathcal{K}_j : |x - \mu_{jk}| \leq \sigma_j \sqrt{2(\beta + 1) \log \sigma_j^{-1}} \} \). Then, we can split again the sum as

\[
\sum_{j > q} \sum_{k \in \mathcal{K}_j} u_{jk} \varphi \left( \frac{x - \mu_{jk}}{\sigma_j} \right) = \\
\sum_{j > q} \sum_{k \in \mathcal{L}_j} u_{jk} \varphi \left( \frac{x - \mu_{jk}}{\sigma_j} \right) + \sum_{j > q} \sum_{k \notin \mathcal{L}_j} u_{jk} \varphi \left( \frac{x - \mu_{jk}}{\sigma_j} \right).
\]

With exactly the same reasoning as before, we get that the first sum of the rhs of the last expression is bounded above by \( 4\|f\|_1 \sigma_j^{\beta} \). Concerning the second term, for any \( j \geq 1 \) we get from propositions 9 and 10, together with the definition of \( u_{jk} \), that \( |u_{jk}| \lesssim \|f\|_1 \sigma_j^{\beta} \). Since there is \( h_j \sigma_j \) separation between two consecutive \( \mu_{jk} \), we deduce that \( |\mathcal{L}_j| \lesssim 2h_j^{-1} \sqrt{2(\beta + 1) \log \sigma_j^{-1}} \). Therefore, for \( J \) large enough and \( x \in [-\zeta_q, \zeta_q] \) with \( 0 \leq q \leq J \),

\[
|r_5(x)| \lesssim \|f_0\|_1 \sigma_j^{\beta} + \|f_0\|_1 \sqrt{2(\beta + 1) \log \sigma_j^{-1}} \sum_{j > q} \sigma_j^{\beta} \lesssim \sqrt{J} \sigma_q^{\beta}.
\]

The conclusion of the proposition follows by combining all the preceding points.

**APPENDIX D. ELEMENTARY RESULTS**

**Proposition 11.** Let \( \varphi(x) = \exp(-x^2/2) \). Then, for all \( \mu_1, \mu_2 \in \mathbb{R} \), and all \( \sigma_1, \sigma_2 > 0 \) with \( 1/2 \leq \sigma_1/\sigma_2 \leq 2 \),

\[
\sup_{x \in \mathbb{R}} \left| \varphi \left( \frac{x - \mu_1}{\sigma_1} \right) - \varphi \left( \frac{x - \mu_2}{\sigma_2} \right) \right| \leq 4 \frac{|\sigma_1 - \sigma_2|}{\sigma_1 \vee \sigma_2} + \frac{|\mu_1 - \mu_2|}{\sigma_1 \vee \sigma_2}.
\]

**Proof.** Without loss of generality we can assume that \( \sigma_1 \leq \sigma_2 \). Using the triangle inequality, we first write

\[
(27) \quad \left| \varphi \left( \frac{x - \mu_1}{\sigma_1} \right) - \varphi \left( \frac{x - \mu_2}{\sigma_2} \right) \right| \\
\leq \left| \varphi \left( \frac{x - \mu_1}{\sigma_1} \right) - \varphi \left( \frac{x - \mu_2}{\sigma_1} \right) \right| + \left| \varphi \left( \frac{x - \mu_2}{\sigma_1} \right) - \varphi \left( \frac{x - \mu_2}{\sigma_2} \right) \right| \\
\leq \sup_{u \in \mathbb{R}} \left| \varphi \left( u + \frac{\mu_1 - \mu_2}{\sigma_1} \right) - \varphi(u) \right| + \sup_{u \in \mathbb{R}} \left| \varphi \left( \frac{\sigma_1}{\sigma_2} u \right) - \varphi(u) \right|.
\]
The first term of the rhs of equation (27) is obviously bounded by $|\mu_1 - \mu_2|/\sigma_1$. Regarding the second term of the rhs of equation (27),

$$\left| \varphi \left( \frac{\sigma_1}{\sigma_2} u \right) - \varphi(u) \right| \leq \left| \frac{\sigma_1}{\sigma_2} - 1 \right| \left( \frac{\sigma_1}{\sigma_2} \vee 1 \right)^2 \sup_x x^2 \varphi(x),$$

which terminates the proof. □

**Proposition 12.** Let $X \sim \text{SGa}(\alpha, 1)$, with $0 < \alpha \leq 1$. Then for any $x \in \mathbb{R}$ and any $0 < \delta \leq 1/2$ we have $\Pr\{|X - x| \leq \delta\} \geq \frac{\delta e^{-2|x|}}{3e \Gamma(\alpha)}$.

**Proof.** Assume for instance that $x \geq 0$. Recalling that $X$ is distributed as the difference of two independent $\text{Ga}(\alpha, 1)$ distributed random variables, it follows

$$\Pr\{|X - x| \leq \delta\} \geq \frac{1}{\Gamma(\alpha)} \int_0^\infty y^{\alpha-1} e^{-y} \frac{1}{\Gamma(\alpha)} \int_{x+y}^{x+y+\delta} z^{\alpha-1} e^{-z} \, dz \, dy.$$  

Because $\alpha \leq 1$, the mapping $z \mapsto z^{\alpha-1} e^{-z}$ is monotonically decreasing on $\mathbb{R}^+$, then the last integral in the rhs of the previous equation is lower bounded by

$$\delta (x + y + \delta)^{\alpha-1} e^{-(x+y+\delta)} \geq \delta e^{-2(x+y+\delta)}. $$

Then

$$\Pr\{|X - x| \leq \delta\} \geq \frac{\delta e^{-2(x+\delta)}}{\Gamma(\alpha)^2} \int_0^\infty y^{\alpha-1} e^{-3y} \, dy \geq \frac{3^{-\alpha} e^{-2(x+\delta)}}{\Gamma(\alpha)} \frac{1}{\delta} \geq \frac{\delta e^{-2|x|}}{3e \Gamma(\alpha)}.$$  

The proof when $x < 0$ is obvious. □

**References**

Lucien Birgé. Model selection via testing: an alternative to (penalized) maximum likelihood estimators. In *Annales de l’IHP Probabilités et statistiques*, volume 42, pages 273–325, 2006.

N Bochkina and J Rousseau. Adaptive density estimation based on a mixture of Gammas. *ArXiv e-prints*, May 2016.

Antonio Canale and Pierpaolo De Blasi. Posterior consistency of nonparametric location-scale mixtures for multivariate density estimation. *arXiv preprint arXiv:1306.2671*, 2013.

R. de Jonge and J. H. van Zanten. Adaptive nonparametric Bayesian inference using location-scale mixture priors. *Ann. Statist.*, 38:3300–3320, 2010.

Thomas S Ferguson. A bayesian analysis of some nonparametric problems. *The annals of statistics*, pages 209–230, 1973.

Subhashis Ghosal and Aad van der Vaart. Convergence rates of posterior distributions for noniid observations. *Ann. Statist.*, 35(1):192–223, 02 2007. doi: 10.1214/009053606000001172. URL http://dx.doi.org/10.1214/009053606000001172.

Subhashis Ghosal and Aad W Van Der Vaart. Entropies and rates of convergence for maximum likelihood and bayes estimation for mixtures of normal densities. *Annals of Statistics*, pages 1233–1263, 2001.

Subhashis Ghosal, Jayanta K Ghosh, and Aad W Van Der Vaart. Convergence rates of posterior distributions. *Annals of Statistics*, 28(2):500–531, 2000.

Subhashis Ghosal, Aad Van Der Vaart, et al. Posterior convergence rates of dirichlet mixtures at smooth densities. *The Annals of Statistics*, 35(2):697–723, 2007a.
Subhashis Ghosal, Aad Van Der Vaart, et al. Convergence rates of posterior distributions for noniid observations. *The Annals of Statistics*, 35(1):192–223, 2007b.

A. Goldenshluger and O. Lepski. On adaptive minimax density estimation on $r^d$. *Probability Theory and Related Fields*, 159(3):479–543, 2014. ISSN 1432-2064. doi: 10.1007/s00440-013-0512-1. URL http://dx.doi.org/10.1007/s00440-013-0512-1.

Thomas Hangelbroek and Amos Ron. Nonlinear approximation using gaussian kernels. *Journal of Functional Analysis*, 259(1):203–219, 2010.

Wolfgang Härdle, Gerard Kerkyacharian, Dominique Picard, and Alexander Tsybakov. Wavelets. In *Wavelets, Approximation, and Statistical Applications*, pages 1–16. Springer, 1998.

N. L. Hjort, C. Holmes, P. Müller, and S. G. Walker. *Bayesian Nonparametrics*. Cambridge University Press, Cambridge, UK, 2010.

John Frank Charles Kingman. *Poisson processes*, volume 3. Oxford university press, 1992.

W. Kruijer, J. Rousseau, and A. van der Vaart. Adaptive Bayesian density estimation with location-scale mixtures. *Electron. J. Stat.*, 4:1225–1257, 2010.

Zacharie Naulet and Eric Barat. Some aspects of symmetric gamma process mixtures. *arXiv preprint arXiv:1504.00476*, 2015.

P. Reynaud-Bouret, V. Rivoirard, and C. Tuleau-Malot. Adaptive density estimation: a curse of support? *Journal of Statistical Planning and Inference*, 141:115–139, 2011.

Jean-Bernard Salomond. Bayesian testing for embedded hypotheses with application to shape constrains. *arXiv preprint arXiv:1303.6466*, 2013.

C. Scricciolo. Adaptive Bayesian density estimation in $L^p$-metrics with Pitman-Yor or normalized inverse-Gaussian process kernel mixtures. *Bayesian Analysis*, 9:475–520, 2014.

Weining Shen, Surya T Tokdar, and Subhashis Ghosal. Adaptive bayesian multivariate density estimation with dirichlet mixtures. *Biometrika*, 100(3):623–640, 2013.

Robert L Wolpert, Merlise A Clyde, and Chong Tu. Stochastic expansions using continuous dictionaries: Lévy adaptive regression kernels. *The Annals of Statistics*, pages 1916–1962, 2011.