arXiv: arXiv:0000.0000

PARAMETER RECOVERY IN TWO-COMPONENT CONTAMINATION MIXTURES: THE $\mathbb{L}^2$ STRATEGY

By Sébastien Gadat and Clément Marteau and Cathy Maugis-Rabusseau

Toulouse School of Economics, University of Toulouse I Capitole
Institut Camille Jordan, University of Lyon 1 Claude Bernard
Institut Mathématiques de Toulouse, INSA Toulouse

In this paper, we consider a parametric density contamination model. We work with a sample of i.i.d. data with a common density, $f^\star = (1-\lambda^\star)\phi + \lambda^\star\phi(.-\mu^\star)$, where the shape ϕ is assumed to be known. We establish the optimal rates of convergence for the estimation of the mixture parameters $(\lambda^\star, \mu^\star)$. In particular, we prove that the classical parametric rate $1/\sqrt{n}$ cannot be reached when at least one of these parameters is allowed to tend to 0 with n.

1. Introduction.

Because of their wide range of flexibility, finite mixtures are a popular tool to model the unknown distribution of heterogeneous data. They are found in several domains and have been at the core of several mathematical investigations. For a complete introduction to mixtures, we refer the reader to [MP00] and [FS06]. In most cases of interest, a sample $\mathcal{S}_n := (X_1, \ldots, X_n)$ of i.i.d. data is at our disposal, and each entry admits the probability density $f^\star$ w.r.t. the Lebesgue measure. For a finite mixture model, the density $f^\star$ is assumed to have the following shape:

$$f^\star = \sum_{k=1}^{K} \lambda_k \phi_k. \tag{1.1}$$

With such a representation, the population of interest can in some sense be decomposed into K different groups where each group k has a proportion λ_k and is distributed according to the density ϕ_k. For practical purposes, parametric models are often considered. In such cases, the densities ϕ_k are assumed to be known, at least up to some finite parameters, and the parameter estimation problem is often addressed using an EM-type algorithm [DLR77]. In contrast, with the impressive range of applications based on

AMS 2000 subject classifications: Primary 62G05, 62F15; secondary 62G20
Keywords and phrases: $\mathbb{L}^2$ contrast, parameter estimation, rate of convergence, two-component contamination mixture model

mixtures, theoretical issues related to mixture models are somewhat poorly understood.

Among the available theoretical results for mixtures, some of them are particularly linked to the density estimation problem. The works [GW00], [GvdV01] and [KRvdV10] develop a nonparametric Bayesian point of view, while exploiting both the approximation capacity of mixtures and their metric entropy size, first with Gaussian distributions and later with exponential power distributions. A Gaussian mixture estimator based on a non asymptotic penalized likelihood criterion is proposed in [MM11] and the adaptive properties of this estimator are investigated in [MRM13].

In the mixture models, the focus on the parameters themselves has received less theoretical attention because of their great mathematical difficulty despite their natural interest. It is indeed highly informative to obtain the estimation of the mixing distribution, and many applied works use this estimation for descriptive statistics. Among them, the unsupervised clustering with Bayesian interpretation is certainly one of the most widely used applications of mixtures (see, *e.g*, [MP00]). Given a dictionary of densities, [BTWB10] propose an estimation procedure based on the minimization of an $\mathbb{L}^2$ empirical criterion with a sparsity constraint, providing an estimation of the parameters of interest when the location parameters $\mu_k^\star$ (here $\phi_k = \phi(. - \mu_k^\star)$) are not too close to each other. [Che95] studied the estimation of the mixing distribution under a strong identifiability condition. As observed in the recent work of [HK15], the optimal rate depends on the knowledge of the number of components. [HN16a] show that the parameter estimation rates are slower for some weakly identifiable mixtures. Other extensions are available in [HN16b].

Finally, the EM algorithm (see, *e.g.*, [DLR77]) is a popular alternative for the analysis of the latent structures involved in the mixture models, but the analysis of the convergence rate of the final estimator is somewhat intricate. A first positive result about the *convergence* of this method is given in [Wu83] when the density is unimodal and certain smoothness conditions hold. However, when multimodality occurs, the behavior of the EM method remains mysterious. Some recent advances in the analysis of this famous method were brought by [BWY16], where a general result is given for a convergence of the sample-based EM towards the population one, up to initialization, Lipschitz and concavity conditions.

In this paper, we focus on the parameter estimation problem when the density of interest is a two-component contamination mixture:

$$f^\star = (1 - \lambda^\star)\phi + \lambda^\star \phi(. - \mu^\star), \tag{1.2}$$

where the density ϕ is *known* and the parameters $(\lambda^\star, \mu^\star)$ are to be estimated.
The estimation of the couple $(\lambda^\star, \mu^\star)$ has already been considered in the literature. In [BMV06], a slightly different model is considered where $f^\star = (1-\lambda^\star)\phi(.-\mu_1^\star)+\lambda^\star\phi(.-\mu_2^\star)$ and ϕ is assumed to be symmetric and unknown. Using a recurrence procedure based on an inversion formula, they propose an estimator for $\theta^\star = (\lambda^\star, \mu_1^\star, \mu_2^\star)$ and the function ϕ. In particular, the parameter $\lambda^\star$ is estimated at the 'classical' parametric rate $1/\sqrt{n}$, while the rate $n^{-1/4}$ is obtained for location parameters $(\mu_1^\star, \mu_2^\star)$. A similar problem is addressed in [BV14] where the rate $1/\sqrt{n}$ is reached for the estimation of the whole parameter $\theta^\star$. The estimation procedure is based on a computation of an empirical Fourier transform. In the setting considered here (i.e., when $f^\star$ is defined as in (1.2)), [CJL07] proposes an iterative procedure based on the empirical distribution function. In the so-called *sparse* setting where $\lambda \ll 1/\sqrt{n}$ and $\mu^\star \sim \sqrt{2r\log(n)}$ for some $r \in (0,1)$ as $n \to +\infty$, the authors derive rates of convergence for the estimation of $\lambda^\star$. In particular, they prove that the classical parametric rate cannot be attained in such a setting.
In all the aforementioned contributions except [CJL07], it is implicitly assumed that both location and proportion parameters are fixed with respect to n. The main aim of this paper is to fill this gap. We propose a procedure inspired by [BTWB10] and derive an estimator $(\hat\lambda, \hat\mu)$ for the couple $(\lambda^\star, \mu^\star)$. This estimator is based on the minimization of an $\mathbb{L}^2$ contrast instead of a usual maximum likelihood estimator of mixture parameters computed with an EM-type algorithm. Then, given a bound M s.t. $|\mu^\star| \leq M$ and under mild assumptions on the shape ϕ, we prove that:

$$\sup_{(\lambda^\star,\mu^\star)\in[0,1]\times[-M,M]} \mathbb{E}_{\lambda^\star,\mu^\star}[(\lambda^\star\mu^\star)^2(\hat\mu-\mu^\star)^2] \lesssim \frac{\log^2 n}{n}, \tag{1.3}$$

and

$$\sup_{\substack{(\lambda^\star,\mu^\star)\in[0,1]\times[-M,M]\\ \lambda^\star\{\mu^\star\}^2 \gtrsim n^{-1/2}}} \mathbb{E}_{\lambda^\star,\mu^\star}[\{\mu^\star\}^4(\hat\lambda-\lambda^\star)^2] \lesssim \frac{\log^2 n}{n}. \tag{1.4}$$

These results are completed by the corresponding lower bounds that ensure the optimality of (1.3) and (1.4). In particular, we can immediately deduce that the parametric rate of $1/\sqrt{n}$ is attained when $\lambda^\star$ and $\mu^\star$ are fixed, but is deteriorated as soon as these parameters are allowed to tend to 0 with n.

The paper is organized as follows. First, a preliminary oracle inequality for $\mathbb{L}^2$ density estimation is established in Section 2. On the basis of this result, some rates of convergence for the estimation of $(\lambda^\star, \mu^\star)$ are deduced (see Section 3.2) under some assumptions on ϕ presented in Section 3.1. Some lower bounds are provided in Section 4, first in a strong contamination model ($|\mu^\star| > m$ with m independent of n; see Section 4.1); and second, in a weak contamination model ($|\mu|$ can tend to 0 when $n \to +\infty$; see Section 4.2). Proofs of the upper bounds (resp. lower bounds) are given in Section 5 (resp. Section 6). Some simulations are presented in Section 7. Technical results are presented in Appendix A, whereas Appendix B is devoted to a needed refinement of the Cauchy-Schwarz inequality.

2. A preliminary result on $\mathbb{L}^2$ density estimation.

2.1. *Statistical setting and identifiability.* We recall that we have at our disposal an i.i.d. sample of size n denoted $\mathcal{S}_n := (X_1, \dots, X_n)$, where the distribution of each X_i is associated with a two-component contamination mixture model. More precisely, we assume that each X_i admits an unknown density $f^\star$ with respect to the Lebesgue measure on $\mathbb{R}$, which is given by:

$$f^\star = (1 - \lambda^\star)\phi + \lambda^\star \phi(. - \mu^\star). \tag{2.1}$$

In the following text, $\theta^\star = (\lambda^\star, \mu^\star)$ refers to the *parameters* of the two-component contamination mixture model. We assume that the density ϕ is a *known* function and that a real contamination of this baseline density ϕ occurs ($\lambda^\star > 0$). Finally, we assume that the unknown contamination shift $\mu^\star$ belongs to a bounded interval $[-M, M]$ where $M > 0$ is known.

Here and below, for any $\theta = (\lambda, \mu) \in (0, 1) \times \mathbb{R}$, we write:

$$f_\theta = f_{\lambda,\mu} = (1 - \lambda)\phi + \lambda\phi_\mu,$$

where ϕ_μ is defined according to the standard notation in location models:

$$\forall \mu \in \mathbb{R} \qquad \phi_\mu : x \longmapsto \phi(x - \mu).$$

In particular, as a slight abuse of notation, we write $f^\star = f_{\theta^\star} = f_{\lambda^\star,\mu^\star}$ and (when the meaning is clear following the context) $\hat{f} = f_{\hat{\theta}} = f_{\hat{\lambda},\hat{\mu}}$ for any estimator $\hat{\theta}$ of $\theta^\star$.

We aim to recover the unknown parameter $\theta^\star$ from the sample $\mathcal{S}_n$. This might be possible according to the next identifiability result, whose proof is given in Appendix A.

PROPOSITION 2.1. *Any two-component contamination mixture model is identifiable: $f_{\theta_1} = f_{\theta_2}$ if and only if $\theta_1 = \theta_2$.*

Such an identifiability result is well known in some more general cases up to additional assumptions on the baseline density ϕ (see, e.g., Theorem 2.1 of [BMV06] where the symmetry of ϕ is added to ensure the identifiability of the general mixture model without contamination). Here, the fact that one of the components of the mixture is constrained to be centered makes it possible to get rid of any additional assumption on ϕ. In particular, Proposition 2.1 holds as soon as ϕ is non-negative with $\int_{\mathbb{R}} \phi = 1$.

2.2. *Estimation strategy and oracle inequality on the $\mathbb{L}^2$ norms.* Our estimator will be built according to an optimal $\mathbb{L}^2$ density estimation constrained to the contamination models. For this purpose, we first define a grid over the possible values of λ and μ through:

$$\mathcal{M}_{\Lambda,\mathfrak{M}} := \{(\lambda,\mu) \,:\, \lambda \in \Lambda = \{\lambda_1,\ldots,\lambda_p\} \text{ and } \mu \in \mathfrak{M} = \{\mu_1,\ldots,\mu_q\}\},$$

where $\Lambda, \mathfrak{M}$ will depend on n to obtain good properties both from the statistical and approximation point of view. To obtain a good estimation of $f^\star$ and $\theta^\star$, we adopt a SURE approach (see, *e.g.*, [Ste81]) and choose an estimator that minimizes $\|f^\star - f_{\lambda,\mu}\|^2$ over the grid $\mathcal{M}_{\Lambda,\mathfrak{M}}$. Observing that:

$$\|f^\star - f_{\lambda,\mu}\|^2 - \|f^\star\|^2 = -2\langle f^\star, f_{\lambda,\mu}\rangle + \|f_{\lambda,\mu}\|^2,$$

and since $\|f^\star\|^2$ does not depend on (λ,μ), it is natural to introduce the following contrast function:

$$\forall (\lambda,\mu) \in \mathcal{M}_{\Lambda,\mathfrak{M}} \qquad \gamma_n(\lambda,\mu) := -\frac{2}{n}\sum_{i=1}^n f_{\lambda,\mu}(X_i) + \|f_{\lambda,\mu}\|^2,$$

leading to the estimator:

$$(\hat\lambda_n, \hat\mu_n) = \arg \min_{(\lambda,\mu)\in\mathcal{M}_{\Lambda,\mathfrak{M}}} \gamma_n(\lambda,\mu). \tag{2.2}$$

Our first main result, stated below, quantifies the performances of $\hat f$.

THEOREM 2.1. *Let $(\lambda^\star,\mu^\star) \in (0,1)\times\mathbb{R}$. Let $(\hat\lambda,\hat\mu)$ be the estimator defined in (2.2). Then, a positive constant $\mathcal{C}$ exists such that for all $0<\alpha<1$:*

$$\mathbb{E}\left[\|\hat f - f^\star\|^2\right] \le \left(\frac{1+\alpha}{1-\alpha}\right) \inf_{(\lambda,\mu)\in\mathcal{M}_{\Lambda,\mathfrak{M}}} \|f_{\lambda,\mu} - f^\star\|^2 + \frac{\mathcal{C}}{2\alpha}\frac{\log^2(|\mathcal{M}_{\Lambda,\mathfrak{M}}|)}{n}, \tag{2.3}$$

where $|\mathcal{M}_{\Lambda,\mathfrak{M}}|$ corresponds to the cardinal of the grid $\mathcal{M}_{\Lambda,\mathfrak{M}}$.

It is worth mentioning that the result above is almost assumption-free on the two-component contamination mixture model. Nevertheless, this result implicitly requires that the approximation term $\inf_{(\lambda,\mu)\in\mathcal{M}_{\Lambda,\mathfrak{M}}} \|f_{\lambda,\mu} - f^\star\|^2$ is comparable to the residual. In practice, this cannot be achieved unless we have an upper bound on the range for possible values of μ at our disposal. The proof of Theorem 2.1 is given in Section 5.1.
We stress that Theorem 2.1 is not the main interest of our work. It is a minimal requirement to further extend our analysis on the parameter estimation of the mixture models themselves. In particular, the following question now arises: *does the fact that $\hat{f}$ is a 'good' $\mathbb{L}^2$ estimator of $f^\star$ imply that the corresponding $\hat{\theta}$ provides a satisfying estimator of $\theta^\star$?* The positive answer to this question is the main contribution of our work and is described in the next section. In order to establish this result, some mild restrictions on the class of possible densities ϕ are required.

3. Estimation of the parameter $\boldsymbol{\theta^\star}$.

3.1. *Baseline assumptions.* We now introduce mild and sufficient assumptions for an optimal recovery of $\theta^\star$ from the oracle inequality (2.3) (in terms of convergence rates). In the following, we denote by $\mathcal{C}^k_p(\mathbb{R})$ the set of continuous functions that admits piecewise k continuous derivatives.

Assumption $(\mathbf{H}_{\mathcal{S}})$. *The density ϕ fulfills one of the two assumptions:*

$(\mathbf{H}_{\mathcal{S}\mathbf{1}})$ *ϕ is symmetric and belongs to $\mathcal{C}^3_p(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$*
or
$(\mathbf{H}_{\mathcal{S}\mathbf{2}})$ *ϕ belongs to $\mathcal{C}^3(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$.*

The set of admissible densities considered in Assumption $(\mathbf{H}_{\mathcal{S}})$ is very large, and contains many possible distributions (Gaussian, Cauchy and Laplace, to name a few). Note that if the density is not regular enough, the symmetry of ϕ is required.
Our second important assumption is concerned with a tight link that may exist between $\phi - \phi_\mu$ and μ itself. It requires a type of Lipschitz upper bound in the translation model.

Assumption $(\mathbf{H_{Lip}})$. *The density ϕ satisfies:*

$$\exists\, g \in \mathbb{L}^2(\mathbb{R}) \quad \forall x \in \mathbb{R} \quad \forall \mu \in [-M, M] \qquad |\phi(x) - \phi_\mu(x)| \leq |\mu| g(x), \tag{3.1}$$

and g satisfies the integrability condition:

$$\mathcal{J} := \int_{\mathbb{R}} g^2(x)\phi^{-1}(x)dx < +\infty.$$

This assumption will be of primary importance to obtain estimation results on the parameters of the mixture themselves. In particular, it will make it possible to derive a relationship between the $\mathbb{L}^2$ norm of $\phi - \phi_\mu$ and the size of $|\mu|$. Hence, under Assumption $(\mathbf{H_{Lip}})$, a good estimation of the density $f^\star$ for the $\mathbb{L}^2$ norm is assumed to yield a good estimation of the mixture parameters.

Assumption $(\mathbf{H_D})$. *The density ϕ satisfies:*

$$\mathcal{I}_\phi := \int \phi''(x)^2 \phi^{-1}(x) dx < +\infty.$$

Assumption $(\mathbf{H_D})$ will be needed for our lower bound results (see Section 4) but is not necessary to obtain good estimation properties. However, this assumption is very mild and is again satisfied for many probability distributions.

REMARK 3.1. *Instead of listing all the possible densities that both meet Assumptions* $(\mathbf{H}_\mathcal{S}), (\mathbf{H_{Lip}})$ *and* $(\mathbf{H_D})$*, we will show that* any log-concave *distribution ϕ written as:*

$$\phi = e^{-U} \quad \textit{with } U \textit{ convex such that} \quad |U'| + |U''| = o_{\pm\infty}(U),$$

satisfies these three conditions. The relationships between $(\mathbf{H}_\mathcal{S}), (\mathbf{H_{Lip}}), (\mathbf{H_D})$ *and the log-concave distributions are given in Appendix A.3.*

In the following text, we maintain a formalism that uses the three assumptions of Section 3.1 for the sake of generality.

3.2. *Consistency rates on the parameters* $(\lambda^\star, \mu^\star)$. We now use our assumptions on ϕ to deduce some rates of convergence for the estimation of the couple $(\lambda^\star, \mu^\star)$ from the oracle inequality of Theorem 2.1. According to the assumption $\mu^\star \in [-M, M]$ for some given $M > 0$, we define the grid $\mathcal{M}_n = \mathcal{M}_{\Lambda,\mathfrak{M}}$ as:

$$\mathcal{M}_n = \Big\{ (\lambda, \mu) : \ \lambda = \frac{i}{\sqrt{n}}, \mu = \pm \frac{j}{\sqrt{n}} \quad \text{with} \quad i \in \{1, \ldots, \sqrt{n}\},\, j \in \{1, \ldots, M\sqrt{n}\} \Big\}, \tag{3.2}$$

so that the approximation term $\inf_{(\lambda,\mu)\in\mathcal{M}_n} \|f_{\lambda,\mu} - f^\star\|^2$ in Equation (2.3) can be made lower than n^{-1}, while keeping the size of $\log(|\mathcal{M}_n|)$ reasonable and of order $2\log(n)$. The next result, whose proof is given in Section 5.2, explicitly gives a non-asymptotic consistency rate of the estimation of $\mu^\star$ in terms of the sample size n, of the amount of contamination $\mu^\star$, and of the probability $\lambda^\star$ of this contamination itself.

THEOREM 3.1. *Let $(\hat{\lambda}_n, \hat{\mu}_n)$ be the estimator defined in (2.2) with $\mathcal{M}_n$ given in (3.2). If ϕ satisfies Assumptions $(\mathbf{H}_\mathcal{S})$ and $(\mathbf{H_{Lip}})$, a positive constant C_1 exists such that:*

$$\forall n \in \mathbb{N} \qquad \sup_{(\lambda^\star,\mu^\star)\in(0,1)\times[-M,M]} \mathbb{E}_{\lambda^\star,\mu^\star}\left[(\lambda^\star\mu^\star)^2(\hat{\mu}-\mu^\star)^2\right] \leq \frac{C_1 \log^2 n}{n}.$$

As an immediate consequence of Theorem 3.1, we can establish that for a *fixed* couple $(\lambda^\star, \mu^\star)$:

$$\mathbb{E}_{\lambda^\star,\mu^\star}\left[\left(\frac{\hat{\mu}}{\mu^\star}-1\right)^2\right] \leq \frac{C_1 \log^2 n}{n\{\lambda^\star\}^2\{\mu^\star\}^4}.$$

In particular, since $\mu^\star$ is allowed to tend to 0 with n, the estimator $\hat{\mu}$ will be consistent as soon as $\sqrt{n}\lambda^\star\{\mu^\star\}^2 \to +\infty$ as $n \to +\infty$. In a detection context, a two-component mixture distribution can be distinguished from that of a single component as soon as $\sqrt{n}\lambda^\star|\mu^\star| > \mathcal{C}$ for some positive constant $\mathcal{C}$ (see, e.g., [CJJ11] or [LMMR16]). Naturally, detection is "easier" than estimation in the sense that the first task requires weaker conditions on the parameters of interest than the second. Since the contamination level $\mu^\star$ is assumed to be upper bounded, it is worth observing that we implicitly require that $\lambda^\star \gg 1/\sqrt{n}$ as $n \to +\infty$.

Before checking the optimality of this result (see Section 4), we investigate the estimation of the contamination proportion $\lambda^\star$. According to the previous discussion, we will assume that $\lambda^\star\{\mu^\star\}^2$ is significantly larger than $n^{-1/2}\log^2 n$. This ensures that the contamination level $\mu^\star$ is consistently estimated. For this purpose, we introduce the set $\Theta_n(M, (\ell_n)_n, \overline{\lambda})$ indexed by a sequence $(\ell_n)_n$:

$$\Theta_n(M, (\ell_n)_n, \overline{\lambda}) := \left\{\theta = (\lambda, \mu) \, : \, \frac{\ell_n}{\mu^2\sqrt{n}} \leq \lambda \leq \overline{\lambda}, |\mu| \leq M\right\},$$

for some $\bar{\lambda} \in (0, 1)$.

THEOREM 3.2. *If ϕ satisfies Assumptions $(\mathbf{H}_\mathcal{S})$ and $(\mathbf{H_{Lip}})$ and the sequence $(\ell_n)_n$ is such that $\lim_{n\to+\infty} \frac{\ell_n}{\log n} = +\infty$, then a positive constant C_2 exists such that:*

$$\sup_{(\lambda^\star,\mu^\star)\in\Theta_n(M,(\ell_n)_n,\overline{\lambda})} \mathbb{E}_{\lambda^\star,\mu^\star}\left[\{\mu^\star\}^4(\hat{\lambda}-\lambda^\star)^2\right] \leq \frac{C_2 \log^2 n}{n}.$$

The proof is given in Section 5.3. Once again, we can immediately deduce from this bound that:

$$\mathbb{E}_{\lambda^\star,\mu^\star}\left[\left(\frac{\hat{\lambda}}{\lambda^\star}-1\right)^2\right] \leq \frac{C_2\log^2 n}{n\{\lambda^\star\}^2\{\mu^\star\}^4},$$

which only makes sense when $\sqrt{n}\lambda^\star\{\mu^\star\}^2 \to +\infty$ as $n \to +\infty$. We stress that in the particular case of fixed $\lambda^\star$ and $\mu^\star$ (w.r.t. n), these quantities can be estimated at the classical parametric rate of $1/\sqrt{n}$ (up to a logarithmic term).

4. Lower bounds.

We now derive some lower bounds on the estimation of $\lambda^\star$ and $\mu^\star$ and show that our previous results are *minimax optimal* with respect to the values of n, $\lambda^\star$ and $\mu^\star$ up to some $\log^2 n$ terms.

4.1. *Strong contamination model.*

For this purpose, we split our study into two cases and first consider the "standard" situation of a strong contamination, meaning that $\mu^\star$ is bounded from below by a constant independent on n: it translates the fact that the contamination is not negligible when $n \longrightarrow +\infty$. Let m and c be two positive constants, and:

$$\Theta_n(m,c) := \left\{\theta = (\lambda,\mu) : \frac{c}{\mu^2\sqrt{n}} \leq \lambda,\ m \leq |\mu|\right\}.$$

Note that this still allows a weak effect of contamination since $\lambda^\star$ can be on the order of $n^{-1/2}$. In this case, we obtain the next lower bound that matches (up to a log term) the upper bounds obtained in Theorems 3.1 and 3.2 as soon as Assumption ($\mathbf{H_D}$) is satisfied.

THEOREM 4.1. *Assume that ϕ satisfies ($\mathbf{H_{Lip}}$) and ($\mathbf{H_D}$). Consider two positive constants m and c such that $0 < \frac{c}{m^2\sqrt{n}} < 1$ so that $\Theta_n(m,c)$ is non empty. Then,*

(*i*) *a positive constant C_1 exists such that:*

$$\inf_{(\hat{\lambda},\hat{\mu})}\ \sup_{(\lambda,\mu)\in\Theta_n(m,c)} \mathbb{E}[\lambda^2(\hat{\mu}-\mu)^2] \geq \frac{C_1}{n}, \tag{4.1}$$

(*ii*) *a positive constant C_2 exists such that:*

$$\inf_{(\hat{\lambda},\hat{\mu})}\ \sup_{(\lambda,\mu)\in\Theta_n(m,c)} \mathbb{E}[(\hat{\lambda}-\lambda)^2] \geq \frac{C_2}{n}, \tag{4.2}$$

where the infimum is taken over all estimators $\hat{\theta} = (\hat{\lambda}, \hat{\mu})$ in (4.1) and (4.2).

Even though the proof relies on a Le Cam argument and leads to a n^{-1} rate, it clearly deserves a careful study for at least two reasons: the loss is asymmetric in (λ, μ) in $i)$ and the balance between λ, μ and n is unclear. We give the proof of this result in Section 6.2.

4.2. *Weak contamination model.* We now study the situation when the contamination $|\mu|$ is not yet bounded from below and can therefore tend to 0 as $n \longrightarrow +\infty$. Let $c > 0$, and:

$$\Theta_n(c) := \left\{ \theta = (\lambda, \mu) : \frac{c}{\mu^2 \sqrt{n}} \leq \lambda \right\}.$$

THEOREM 4.2. *Assume that ϕ satisfies $(\mathbf{H}_{\mathcal{S}\mathbf{1}})$ and $(\mathbf{H}_{\mathbf{D}})$. Then, $N > 0$ exists such that, for all $n > N$:*

(i) a positive constant C_1 exists such that:

$$\inf_{(\hat{\lambda},\hat{\mu})} \sup_{(\lambda,\mu)\in\Theta_n(c)} \mathbb{E}[\mu^4(\lambda - \hat{\lambda})^2] \geq \frac{C_1}{n}.$$

(ii) a positive constant C_2 exists such that:

$$\inf_{(\hat{\lambda},\hat{\mu})} \sup_{(\lambda,\mu)\in\Theta_n(c)} \mathbb{E}[\lambda^2\mu^2(\mu - \hat{\mu})^2] \geq \frac{C_2}{n}.$$

We emphasize that this last result is only true when dealing with a symmetric density function ϕ. We have not been able to extend it to the general situation induced by $(\mathbf{H}_{\mathcal{S}})$. Even if we believe that this result is still true in that case, our proof strategy cannot be used to extend our result to $(\mathbf{H}_{\mathcal{S}})$.

5. Proofs of the upper bounds.

5.1. *Preliminary oracle inequality.* We first establish a technical proposition that will be used to derive the proof of Theorem 2.1. For a given grid $\mathcal{M}_{\Lambda,\mathfrak{M}}$, we first introduce the theoretical minimizer of the $\mathbb{L}^2$-norm on this grid:

$$(\lambda_0, \mu_0) = \arg \min_{(\lambda,\mu)\in\mathcal{M}_{\Lambda,\mathfrak{M}}} \|f_{\lambda,\mu} - f^\star\|^2. \tag{5.1}$$

We then define $\mathcal{E}_n(\lambda, \mu)$ the empirical process indexed by $(\lambda, \mu) \in \mathcal{M}_{\Lambda,\mathfrak{M}}$ as:

$$\mathcal{E}_n(\lambda, \mu) = \frac{2}{n} \sum_{i=1}^{n} \left\{ f_{\lambda,\mu}(X_i) - f_{\lambda_0,\mu_0}(X_i) - [\langle f_{\lambda,\mu} - f_{\lambda_0,\mu_0}, f^\star \rangle] \right\}.$$

For all $(\lambda, \mu) \in \mathcal{M}_{\Lambda,\mathfrak{M}}$, the term $\mathcal{E}_n(\lambda, \mu)$ can be rewritten as:

$$\mathcal{E}_n(\lambda, \mu) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \mathbb{E}[Y_i]) \quad \text{where} \quad Y_i := 2[f_{\lambda,\mu}(X_i) - f_{\lambda_0,\mu_0}(X_i)]. \tag{5.2}$$

In particular, $\mathbb{E}[\mathcal{E}_n(\lambda, \mu)] = 0$ and:

$$\begin{aligned} \mathrm{Var}(Y_i) \leq \mathbb{E}[Y_i^2] &= 4\mathbb{E}[(f_{\lambda,\mu}(X_i) - f_{\lambda_0,\mu_0}(X_i))^2], \\ &= 4\int_{\mathbb{R}}[f_{\lambda,\mu}(x) - f_{\lambda_0,\mu_0}(x)]^2 f^\star(x)dx, \\ &\leq 4\|\phi\|_\infty \|f_{\lambda,\mu} - f_{\lambda_0,\mu_0}\|^2, \end{aligned}$$

since $\|f^\star\|_\infty \leq \|\phi\|_\infty$. We will use a normalized version of this process below, which naturally leads to the introduction of $\mathcal{G}_n(\lambda, \mu)$:

$$\forall (\lambda, \mu) \in \mathcal{M}_{\Lambda,\mathfrak{M}} \setminus \{(\lambda_0, \mu_0)\} \qquad \mathcal{G}_n(\lambda, \mu) = \frac{\mathcal{E}_n(\lambda, \mu)}{\|f_{\lambda,\mu} - f_{\lambda_0,\mu_0}\|}.$$

Our estimator $(\hat{\lambda}, \hat{\mu})$ defined in (2.2) satisfies the following useful property.

Lemma 5.1.

(i) *For any* (λ, μ) *such that* $\|f_{\lambda,\mu} - f_{\lambda_0,\mu_0}\| \geq n^{-1/2}$*:*

$$\forall s > 0 \qquad \mathbb{P}\left(|\mathcal{G}_n(\lambda, \mu)| > s\right) \leq \exp\left(-\frac{ns^2}{8\|\phi\|_\infty\left[1 + \frac{s\sqrt{n}}{3}\right]}\right). \tag{5.3}$$

(ii) *We can find* $C > 0$ *such that:*

$$\mathbb{E}\left[\mathcal{G}_n(\hat{\lambda}, \hat{\mu})^2 \mathbb{1}_{\mathcal{B}^c}\right] \leq \frac{C\log^2(|\mathcal{M}_{\Lambda,\mathfrak{M}}|)}{n}, \tag{5.4}$$

where $\mathcal{B}$ *is the event defined as* $\mathcal{B} = \left\{\|\hat{f} - f_{\lambda_0,\mu_0}\| \leq \frac{1}{\sqrt{n}}\right\}$.

Proof. In this proof, C refers to a constant that is independent of n, whose value may change from line to line.

Proof of (i): thanks to the Bennett inequality, we obtain for all $s > 0$:

$$\begin{aligned} &\mathbb{P}\left(|\mathcal{G}_n(\lambda, \mu)| > s\right) \\ &\leq \exp\left(-\frac{n^2 s^2 \|f_{\lambda,\mu} - f_{\lambda_0,\mu_0}\|^2}{8n\|\phi\|_\infty \|f_{\lambda,\mu} - f_{\lambda_0,\mu_0}\|^2 + 8n\|\phi\|_\infty s\|f_{\lambda,\mu} - f_{\lambda_0,\mu_0}\|/3}\right), \\ &= \exp\left(-\frac{ns^2}{8\|\phi\|_\infty\left[1 + s\|f_{\lambda,\mu} - f_{\lambda_0,\mu_0}\|^{-1}/3\right]}\right). \end{aligned}$$

Using the fact that $\|f_{\lambda,\mu} - f_{\lambda_0,\mu_0}\| \geq n^{-1/2}$, we obtain:

$$\mathbb{P}\left(|\mathcal{G}_n(\lambda,\mu)| > s\right) \leq \exp\left(-\frac{ns^2}{8\|\phi\|_\infty \left[1+\frac{s\sqrt{n}}{3}\right]}\right),$$

which is the desired Inequality (5.3).
<u>Proof of (ii)</u>: observe that for all $t > 0$,

$$\begin{aligned}
\mathbb{E}\left[\mathcal{G}_n^2(\hat\lambda,\hat\mu)\mathbb{1}_{\mathcal{B}^c}\right] &\leq t^2 + \mathbb{E}\left[\mathcal{G}_n^2(\hat\lambda,\hat\mu)\mathbb{1}_{\{|\mathcal{G}_n(\hat\lambda,\hat\mu)|>t\}}\mathbb{1}_{\mathcal{B}^c}\right],\\
&\leq t^2 + \mathbb{E}\left[\sup_{(\lambda,\mu):\|f_{\lambda,\mu}-f_{\lambda_0,\mu_0}\|\geq n^{-1/2}} \left\{\mathcal{G}_n^2(\lambda,\mu)\mathbb{1}_{\{|\mathcal{G}_n(\lambda,\mu)|>t\}}\right\}\right],\\
&\leq t^2 + \sum_{(\lambda,\mu):\|f_{\lambda,\mu}-f_{\lambda_0,\mu_0}\|\geq n^{-1/2}} \mathbb{E}\left[\mathcal{G}_n^2(\lambda,\mu)\mathbb{1}_{\{|\mathcal{G}_n(\lambda,\mu)|>t\}}\right]. \qquad (5.5)
\end{aligned}$$

Integrating by parts, we can remark that:

$$\mathbb{E}\left[\mathcal{G}_n^2(\lambda,\mu)\mathbb{1}_{\{|\mathcal{G}_n(\lambda,\mu)|>t\}}\right] = t^2\,\mathbb{P}(|\mathcal{G}_n(\lambda,\mu)|>t) + \int_{t^2}^{+\infty}\mathbb{P}(|\mathcal{G}_n(\lambda,\mu)|>\sqrt{x})dx.$$

Thus, if we choose $t = \left(\frac{16\|\phi\|_\infty \log(|\mathcal{M}_{\Lambda,\mathfrak{M}}|)}{3} \vee 3\right) n^{-1/2}$, then $t\sqrt{n}/3 \geq 1$, so that for any $s \geq t$ and for a fixed (λ,μ), (5.3) yields:

$$\begin{aligned}
&\mathbb{E}\left[\mathcal{G}_n^2(\lambda,\mu)\mathbb{1}_{\{|\mathcal{G}_n(\lambda,\mu)|>t\}}\right]\\
&\leq t^2\exp\left(-\log(|\mathcal{M}_{\Lambda,\mathfrak{M}}|)\right) + \int_{t^2}^{+\infty}\exp\left(-\frac{3\sqrt{nx}}{16\|\phi\|_\infty}\right)dx\\
&\leq C\frac{\log^2(|\mathcal{M}_{\Lambda,\mathfrak{M}}|)}{n}\times\frac{1}{|\mathcal{M}_{\Lambda,\mathfrak{M}}|} + 2\int_t^{+\infty} u\exp\left(-\frac{3\sqrt{n}u}{16\|\phi\|_\infty}\right)du,
\end{aligned}$$

for large enough C, where the last line comes from the size of t for the left-hand side, and from the change of variable $u = \sqrt{x}$ in the integral. The remaining integral may be integrated by parts, which in turn leads to:

$$\mathbb{E}\left[\mathcal{G}_n^2(\lambda,\mu)\mathbb{1}_{\{|\mathcal{G}_n(\lambda,\mu)|>t\}}\right] \leq C\frac{\log^2(|\mathcal{M}_{\Lambda,\mathfrak{M}}|)}{n}\times\frac{1}{|\mathcal{M}_{\Lambda,\mathfrak{M}}|}.$$

If we plug the above upper bound into (5.5), we then obtain that a sufficiently large constant C exists such that:

$$\mathbb{E}\left[\mathcal{G}_n^2(\hat\lambda,\hat\mu)\mathbb{1}_{\mathcal{B}^c}\right] \leq C\frac{\log^2(|\mathcal{M}_{\Lambda,\mathfrak{M}}|)}{n}\times\frac{|\mathcal{M}_{\Lambda,\mathfrak{M}}|}{|\mathcal{M}_{\Lambda,\mathfrak{M}}|}| = C\frac{\log^2(|\mathcal{M}_{\Lambda,\mathfrak{M}}|)}{n}.$$

□

We are now interested in the proof of the oracle inequality.

PROOF OF THEOREM 2.1. The best approximation term (λ_0, μ_0) over the grid $\mathcal{M}_{\Lambda,\mathfrak{M}}$ is defined in (5.1) and the event $\mathcal{B} = \left\{\|\hat{f} - f_{\lambda_0,\mu_0}\| \leq \sqrt{\frac{1}{n}}\right\}$ is introduced in Proposition 5.1. On the event $\mathcal{B}$, the situation is easy using the Young inequality $2ab \leq \alpha a^2 + \alpha^{-1} b^2$ so that for all $\alpha > 0$,

$$\begin{aligned}
\mathbb{E}\left[\|\hat{f} - f^\star\|^2 \mathbb{1}_{\mathcal{B}}\right] &\leq (1+\alpha)\|f_{\lambda_0,\mu_0} - f^\star\|^2 + (1+\alpha^{-1})\mathbb{E}\left[\|\hat{f} - f_{\lambda_0,\mu_0}\|^2 \mathbb{1}_{\mathcal{B}}\right], \\
&\leq (1+\alpha)\|f_{\lambda_0,\mu_0} - f^\star\|^2 + \frac{1+\alpha^{-1}}{n}. \qquad (5.6)
\end{aligned}$$

We provide below a similar control on the event $\mathcal{B}^c$. First, observe that according to the definition of $(\hat{\lambda}, \hat{\mu})$, for all $(\lambda, \mu) \in \mathcal{M}_{\Lambda,\mathfrak{M}}$, we have:

$$\begin{aligned}
& \gamma_n(\hat{\lambda}, \hat{\mu}) + \|f^\star\|^2 \leq \gamma_n(\lambda, \mu) + \|f^\star\|^2, \\
\Leftrightarrow \quad & \|\hat{f} - f^\star\|^2 \leq \|f_{\lambda,\mu} - f^\star\|^2 + 2\left[\frac{1}{n}\sum_{i=1}^n \hat{f}(X_i) - \langle \hat{f}, f^\star\rangle\right] \\
& \qquad -2\left[\frac{1}{n}\sum_{i=1}^n f_{\lambda,\mu}(X_i) - \langle f_{\lambda,\mu}, f^\star\rangle\right].
\end{aligned}$$

This inequality being true for $(\lambda, \mu) = (\lambda_0, \mu_0)$, we obtain:

$$\|\hat{f} - f^\star\|^2 \mathbb{1}_{\mathcal{B}^c} \leq \|f_{\lambda_0,\mu_0} - f^\star\|^2 + \mathcal{E}_n(\hat{\lambda}, \hat{\mu})\mathbb{1}_{\mathcal{B}^c}.$$

This implies that for all $0 < \alpha < 1$:

$$\begin{aligned}
& \|\hat{f} - f^\star\|^2 \mathbb{1}_{\mathcal{B}^c} \leq \|f_{\lambda_0,\mu_0} - f^\star\|^2 + \|\hat{f} - f_{\lambda_0,\mu_0}\| \frac{\mathcal{E}_n(\hat{\lambda}, \hat{\mu})}{\|\hat{f} - f_{\lambda_0,\mu_0}\|}\mathbb{1}_{\mathcal{B}^c}, \\
\Rightarrow \quad & \|\hat{f} - f^\star\|^2 \mathbb{1}_{\mathcal{B}^c} \leq \|f_{\lambda_0,\mu_0} - f^\star\|^2 + \frac{\alpha}{2}\|\hat{f} - f_{\lambda_0,\mu_0}\|^2 \mathbb{1}_{\mathcal{B}^c} + \frac{1}{2\alpha}\mathcal{G}_n^2(\hat{\lambda}, \hat{\mu})\mathbb{1}_{\mathcal{B}^c}.
\end{aligned}$$

Using $\|u+v\|^2 \leq 2\|u\|^2 + 2\|v\|^2$, we then deduce that:

$$\|\hat{f} - f^\star\|^2 \mathbb{1}_{\mathcal{B}^c} \leq \frac{(1+\alpha)}{(1-\alpha)}\|f_{\lambda_0,\mu_0} - f^\star\|^2 + \frac{1}{2\alpha}\mathcal{G}_n^2(\hat{\lambda}, \hat{\mu})\mathbb{1}_{\mathcal{B}^c}. \qquad (5.7)$$

We can conclude the proof taking (5.4) in (5.7), and (5.6) together. $\square$

5.2. *Proof of Theorem 3.1.* We aim to apply the oracle inequality established in Theorem 2.1. First, we need an upper bound on the approximation term given by $\|f_{\lambda_0,\mu_0} - f^\star\|^2$ when (λ_0, μ_0) belongs to our grid $\mathcal{M}_n$. We can observe that for all $(\lambda, \mu) \in (0,1) \times \mathbb{R}$,

$$\begin{aligned} \|f_{\lambda,\mu} - f^\star\|^2 &= \|(1-\lambda)\phi + \lambda\phi_\mu - (1-\lambda^\star)\phi - \lambda^\star\phi_{\mu^\star}\|^2 \\ &= \|(\lambda^\star - \lambda)\{\phi - \phi_\mu\} + \lambda^\star\{\phi_\mu - \phi_{\mu^\star}\}\|^2 \\ &\leq 2(\lambda^\star - \lambda)^2\|\phi - \phi_\mu\|^2 + 2\{\lambda^\star\}^2\|\phi_\mu - \phi_{\mu^\star}\|^2. \end{aligned} \tag{5.8}$$

Using Proposition A.1, we can find two positive constants $\overline{\kappa}$ and $\underline{\kappa}$ such that:

$$\forall(\mu, \tilde{\mu}) \in \mathbb{R}^2 \qquad \underline{\kappa}(\mu - \tilde{\mu})^2 \leq \|\phi_\mu - \phi_{\tilde{\mu}}\|^2 \leq \overline{\kappa}(\mu - \tilde{\mu})^2, \tag{5.9}$$

which in turn implies that:

$$\|f_{\lambda,\mu} - f^\star\|^2 \leq 8\|\phi\|^2(\lambda^\star - \lambda)^2 + 2\overline{\kappa}\,\{\lambda^\star\}^2(\mu - \mu^\star)^2.$$

In particular, the definition of $\mathcal{M}_n$ given in (3.2) makes it possible to find a constant $C > 0$ such that:

$$\|f_{\lambda_0,\mu_0} - f^\star\|^2 = \inf_{(\lambda,\mu)\in\mathcal{M}_n} \|f_{\lambda,\mu} - f^\star\|^2 \leq \frac{C}{n}. \tag{5.10}$$

At the same time, observe that (5.8) leads to:

$$\begin{aligned} \|\hat{f} - f^\star\|^2 &= (\lambda^\star - \hat{\lambda})^2\|\phi - \phi_{\hat{\mu}}\|^2 + \{\lambda^\star\}^2\|\phi_{\hat{\mu}} - \phi_{\mu^\star}\|^2 \\ &\qquad + 2(\lambda^\star - \hat{\lambda})\lambda^\star\langle\phi - \phi_{\hat{\mu}}, \phi_{\hat{\mu}} - \phi_{\mu^\star}\rangle. \end{aligned}$$

Using Proposition B.2 with $a = \hat{\mu}$ and $b = \mu^\star - \hat{\mu}$ and (5.9), a positive constant c exists such that:

$$\begin{aligned} &\|\hat{f} - f^\star\|^2 \\ &\quad \geq (\lambda^\star - \hat{\lambda})^2\|\phi - \phi_{\hat{\mu}}\|^2 + \{\lambda^\star\}^2\|\phi_{\hat{\mu}} - \phi_{\mu^\star}\|^2 \\ &\qquad -2\left|\lambda^\star - \hat{\lambda}\right|\lambda^\star\|\phi - \phi_{\hat{\mu}}\|\|\phi_{\hat{\mu}} - \phi_{\mu^\star}\|\left(1 - c\|\phi - \phi_{\mu^\star}\|^2\right) \\ &\quad \geq (\lambda^\star - \hat{\lambda})^2\|\phi - \phi_{\hat{\mu}}\|^2 + \{\lambda^\star\}^2\|\phi_{\hat{\mu}} - \phi_{\mu^\star}\|^2 \\ &\qquad -\left[(\lambda^\star - \hat{\lambda})^2\|\phi - \phi_{\hat{\mu}}\|^2 + \{\lambda^\star\}^2\|\phi_{\hat{\mu}} - \phi_{\mu^\star}\|^2\right]\left(1 - c\|\phi - \phi_{\mu^\star}\|^2\right) \\ &\quad \geq c(\lambda^\star - \hat{\lambda})^2\|\phi - \phi_{\hat{\mu}}\|^2\|\phi - \phi_{\mu^\star}\|^2 + c\{\lambda^\star\}^2\|\phi_{\hat{\mu}} - \phi_{\mu^\star}\|^2\|\phi - \phi_{\mu^\star}\|^2 \\ &\quad \geq c\underline{\kappa}^2(\lambda^\star - \hat{\lambda})^2\{\hat{\mu}\}^2\{\mu^\star\}^2 + c\underline{\kappa}^2\{\lambda^\star\}^2\{\mu^\star\}^2(\hat{\mu} - \mu^\star)^2. \end{aligned}$$

We see here the central role of the refinement of the Cauchy-Schwarz inequality (see Appendix B at the end of the paper) to obtain a tractable bound that involves the parameters of the mixture themselves, from the bound on the $\mathbb{L}^2$-norm of $\hat{f} - f^\star$. We now use the oracle inequality on $\|\hat{f} - f^\star\|^2$ to deduce that a constant $C > 0$ exists such that:

$$\mathbb{E}\left[(\lambda^\star - \hat{\lambda})^2\{\hat{\mu}\}^2\{\mu^\star\}^2 + \{\lambda^\star\}^2\{\mu^\star\}^2(\hat{\mu} - \mu^\star)^2\right] \leq \frac{C\log^2 n}{n}. \tag{5.11}$$

In particular, we immediately deduce from (5.11) that:

$$\mathbb{E}\left[\{\lambda^\star\}^2\{\mu^\star\}^2(\hat{\mu} - \mu^\star)^2\right] \leq \frac{C\log^2 n}{n}.$$

This result being uniform in $(\lambda^\star, \mu^\star)$, we obtain the proof of Theorem 3.1.

□

Unfortunately, we cannot directly use a similar approach for the estimation of $\lambda^\star$. Indeed, we have to first ensure that $\hat{\mu}$ is close to $\mu^\star$ with a large enough probability.

5.3. *Proof of Theorem 3.2.* Let $\mathcal{B}$ and $\mathcal{D}$ be the events respectively defined as:

$$\mathcal{B} = \left\{\|\hat{f} - f_{\lambda_0,\mu_0}\| \leq \sqrt{\frac{1}{n}}\right\} \tag{5.12}$$

and

$$\mathcal{D} = \left\{|\mathcal{G}_n(\hat{\lambda}, \hat{\mu})| \leq \frac{16\|\phi\|_\infty \log(n|\mathcal{M}_n|)}{3\sqrt{n}}\right\}. \tag{5.13}$$

Below, the control of the quadratic risk of $\hat{\mu}$ will be investigated according to the partition $\mathcal{B}, \mathcal{B}^c \cap \mathcal{D}$ and $\mathcal{B}^c \cap \mathcal{D}^c$.

Control of the risk on $\mathcal{B}$. Equation (5.6) together with (5.10) indicates that:

$$\|\hat{f} - f^\star\|^2 \; \mathbb{1}_{\mathcal{B}} \leq \frac{C}{n}.$$

Then, Equation (5.11) implies that:

$$\left(\frac{\hat{\mu}}{\mu^\star} - 1\right)^2 \mathbb{1}_{\mathcal{B}} \leq \frac{C}{n\{\lambda^\star\}^2\{\mu^\star\}^4} \leq \frac{C}{\ell_n^2}. \tag{5.14}$$

Control of the risk on $\mathcal{B}^c\cap\mathcal{D}$. On the set $\mathcal{B}^c\cap\mathcal{D}$, we apply Inequality (5.7), which yields:

$$\begin{aligned}\|\hat{f}-f^\star\|^2\ \mathbb{1}_{\mathcal{B}^c\cap\mathcal{D}} &\leq \frac{(1+\alpha)}{(1-\alpha)}\|f_{\lambda_0,\mu_0}-f^\star\|^2+\frac{1}{2\alpha}|\mathcal{G}_n(\hat{\lambda},\hat{\mu})|^2\ \mathbb{1}_{\mathcal{B}^c\cap\mathcal{D}}\\ &\leq C\frac{\log^2(n|\mathcal{M}_n|)}{n}\end{aligned}$$

for some positive constant C. Since the size of $|\mathcal{M}_{\Lambda_n,\mathfrak{M}_n}|$ is a polynomial of n, we can find a constant C such that Equation (5.11) leads to:

$$\left(\frac{\hat{\mu}}{\mu^\star}-1\right)^2\mathbb{1}_{\mathcal{B}^c\cap\mathcal{D}}\leq C\frac{\log^2 n}{n\{\lambda^\star\}^2\{\mu^\star\}^4}\leq C\frac{\log^2 n}{\ell_n^2}. \tag{5.15}$$

Since we assume that $(\lambda^\star,\mu^\star)\in\Theta_n(M,(\ell_n)_n,\overline{\lambda})$ with $\ell_n/\log n\longrightarrow+\infty$ when $n\longrightarrow+\infty$, Equations (5.14) and (5.15) imply that for large enough n,

$$\left(\frac{\hat{\mu}}{\mu^\star}-1\right)^2[\mathbb{1}_{\mathcal{B}}+\mathbb{1}_{\mathcal{B}^c\cap\mathcal{D}}]\leq\frac{1}{4}.$$

Remark that for positive x and y: $|x/y-1|\leq\frac{1}{2}$ implies that $y\leq 2x$. Applying this simple remark to the former inequality yields:

$$\{\mu^\star\}^2\left[\mathbb{1}_{\mathcal{B}}+\mathbb{1}_{\mathcal{B}^c\cap\mathcal{D}}\right]\leq 4\{\hat{\mu}\}^2\left[\mathbb{1}_{\mathcal{B}}+\mathbb{1}_{\mathcal{B}^c\cap\mathcal{D}}\right]. \tag{5.16}$$

Control of the risk on $\mathcal{B}^c\cap\mathcal{D}^c$. Applying (5.3) we can check that:

$$\mathbb{P}(\mathcal{B}^c\cap\mathcal{D}^c)\leq\mathbb{P}(\mathcal{D}^c)\leq\frac{C}{n}$$

for some positive constant C.

Synthesis. Using (5.16), a large enough N exists such that for $n\geq N$:

$$\begin{aligned}&\mathbb{E}[(\hat{\lambda}-\lambda^\star)^2\{\mu^\star\}^4]\\ &\quad=\ \mathbb{E}[(\hat{\lambda}-\lambda^\star)^2\{\mu^\star\}^4(\mathbb{1}_{\mathcal{B}}+\mathbb{1}_{\mathcal{B}^c\cap\mathcal{D}})]+\mathbb{E}[(\hat{\lambda}-\lambda^\star)^2\{\mu^\star\}^4\mathbb{1}_{\mathcal{B}^c\cap\mathcal{D}^c}],\\ &\quad\leq\ 4\mathbb{E}[(\hat{\lambda}-\lambda^\star)^2\{\mu^\star\}^2\{\hat{\mu}\}^2]+M^4\mathbb{P}(\mathcal{D}^c),\\ &\quad\leq\ \frac{C\log^2(n)}{n},\end{aligned}$$

for some constant $C>0$, according to (5.11). This result being uniform in $(\lambda^\star,\mu^\star)$, we obtain the proof of Theorem 3.2.

□

6. Proofs of the lower bounds.

6.1. *Asymmetric risk.* We begin by a useful lemma, which is a generalization of the Le Cam method for proving lower bounds if the loss involved in the statistical model is not symmetric, meaning that $\rho(\theta_1, \theta_2)$ is generally not equal to $\rho(\theta_2, \theta_1)$, but still satisfies a weak triangle inequality. Hence, the Le Cam Lemma requires a small modification in the spirit of the remark of [Yu97] (Example 2, Section 3).
In the sequel, $d_{\mathrm{TV}}(P, Q)$ and $\mathrm{KL}(P, Q)$ denote the total variation distance and the Kullback-Leibler divergence between two measures, P and Q, respectively.

LEMMA 6.1. *Let $(\mathbb{P}_\theta)_{\theta \in \Theta}$ be a family of measures indexed by Θ and assume that $\rho : (\theta_1, \theta_2) \in \Theta^2 \mapsto \rho(\theta_1, \theta_2) \in \mathbb{R}^+$ satisfies the weak triangle inequality:*

$$\forall (\theta_1, \theta_2, \theta_3) \in \Theta^3, \ \rho(\theta_1, \theta_3) + \rho(\theta_2, \theta_3) \geq \rho(\theta_1, \theta_2) \wedge \rho(\theta_2, \theta_1). \tag{6.1}$$

Let $\Phi : \mathbb{R}^+ \to \mathbb{R}^+$ be a non-decreasing function. Let $\delta > 0$ and $(\theta_1, \theta_2) \in \Theta^2$ such that $\rho(\theta_1, \theta_2) \wedge \rho(\theta_2, \theta_1) \geq 2\delta$. Then,

$$\begin{aligned}
\inf_{\hat\theta} \sup_{\theta \in \Theta} \mathbb{E}\left[\Phi(\rho(\theta, \hat\theta))\right] &\geq \frac{\Phi(\delta)}{2}\left\{1 - d_{TV}(\mathbb{P}_{\theta_1}^{\otimes n}, \mathbb{P}_{\theta_2}^{\otimes n})\right\}, \\
&\geq \frac{\Phi(\delta)}{2}\left\{1 - \sqrt{\frac{n}{2} KL(\mathbb{P}_{\theta_1}, \mathbb{P}_{\theta_2})}\right\},
\end{aligned}$$

where the infimum is taken over all estimators $\hat\theta$.

PROOF. First, we observe that:

$$\mathbb{E}[\Phi(\rho(\theta, \hat\theta))] \geq \Phi(\delta)\mathbb{P}(\rho(\theta, \hat\theta) \geq \delta),$$

since Φ is a non-decreasing function. Let $\mathcal{V} = \{1, 2\}$ and $\Psi(\hat\theta) = \operatorname*{argmin}_{v \in \mathcal{V}} \rho(\theta_v, \hat\theta)$. We can show that $\rho(\theta_v, \hat\theta) < \delta$ implies that $\Psi(\hat\theta) = v$. According to Condition (6.1), we have:

$$\rho(\theta_v, \hat\theta) \geq \rho(\theta_v, \theta_{v'}) \wedge \rho(\theta_{v'}, \theta_v) - \rho(\theta_{v'}, \hat\theta) > 2\delta - \rho(\theta_{v'}, \hat\theta).$$

Now, if $\rho(\theta_v, \hat\theta) < \delta$, then $\delta > 2\delta - \rho(\theta_{v'}, \hat\theta)$, so that $\rho(\theta_{v'}, \hat\theta) > \delta$, which is necessarily larger than $\rho(\theta_v, \hat\theta)$. Hence, we obtain $\Psi(\hat\theta) = v$.
Equivalently, for $v \in \{1, 2\}$, we have $\Psi(\hat\theta) \neq v \Longrightarrow \rho(\theta_v, \hat\theta) > \rho(\theta_{v'}, \hat\theta)$ since:

$$2\delta \leq \rho(\theta_v, \theta_{v'}) \wedge \rho(\theta_{v'}, \theta_v) \leq \rho(\theta_v, \hat\theta) + \rho(\theta_{v'}, \hat\theta) \leq 2\rho(\theta_v, \hat\theta).$$

The rest of the proof proceeds from the standard Le Cam argument: Φ is non decreasing so that:

$$\begin{aligned}\sup_{\theta\in\Theta}\mathbb{E}[\Phi(\rho(\theta,\hat\theta))] &\geq \Phi(\delta)\sup_{\theta\in\Theta}\mathbb{P}(\rho(\theta,\hat\theta)\geq\delta)\\ &\geq \frac{\Phi(\delta)}{2}\{\mathbb{P}(\rho(\theta_1,\hat\theta)\geq\delta)+\mathbb{P}(\rho(\theta_2,\hat\theta)\geq\delta)\}\\ &\geq \frac{\Phi(\delta)}{2}\{\mathbb{P}_{\theta_1}^{\otimes n}(\Psi(\hat\theta)\neq 1)+\mathbb{P}_{\theta_2}^{\otimes n}(\Psi(\hat\theta)\neq 2)\}.\end{aligned}$$

Taking an infimum over all tests Ψ (see, *e.g.*, [LCY00]) we obtain:

$$\begin{aligned}\inf_{\hat\theta}\sup_{\theta\in\Theta}\mathbb{E}[\Phi(\rho(\theta,\hat\theta))] &\geq \frac{\Phi(\delta)}{2}\inf_{\Psi}\{\mathbb{P}_{\theta_1}^{\otimes n}(\Psi\neq 1)+\mathbb{P}_{\theta_2}^{\otimes n}(\Psi\neq 2)\}\\ &\geq \frac{\Phi(\delta)}{2}\left\{1-d_{\mathrm{TV}}(\mathbb{P}_{\theta_1}^{\otimes n},\mathbb{P}_{\theta_2}^{\otimes n})\right\}.\end{aligned}$$

Pinsker's inequality:

$$d_{\mathrm{TV}}(\mathbb{P}_{\theta_1}^{\otimes n},\mathbb{P}_{\theta_2}^{\otimes n})\leq\sqrt{\frac{1}{2}\mathrm{KL}(\mathbb{P}_{\theta_1}^{\otimes n},\mathbb{P}_{\theta_2}^{\otimes n})}=\sqrt{\frac{n}{2}\mathrm{KL}(\mathbb{P}_{\theta_1},\mathbb{P}_{\theta_2})}$$

ends the proof. □

6.2. *Lower bound for the strong contamination model.* We now study the lower bounds in the first regime, namely when μ is lower bounded by a constant m that is independent of n.
Proof of Theorem 4.1

Point (i). We apply Lemma 6.1 with $\Phi(t)=t^2$ and the loss function ρ defined as:

$$\forall(\theta_1,\theta_2)\in\Theta_n(m,c)^2\qquad \rho(\theta_1,\theta_2)=\lambda_1|\mu_1-\mu_2|.$$

Remark that ρ satisfies the weak triangle inequality (6.1). Indeed, for all $(\theta_1,\theta_2,\theta_3)\in\Theta_n(m,c)^3$, we have:

$$\begin{aligned}\rho(\theta_1,\theta_3)+\rho(\theta_2,\theta_3) &= \lambda_1|\mu_1-\mu_3|+\lambda_2|\mu_2-\mu_3|\\ &\geq \min(\lambda_1,\lambda_2)|\mu_1-\mu_2|\\ &\geq \rho(\theta_1,\theta_2)\wedge\rho(\theta_2,\theta_1).\end{aligned}$$

We introduce the subset

$$\Theta_n(m,M,c,\bar\lambda):=\left\{\theta=(\lambda,\mu):\frac{c}{\mu^2\sqrt{n}}\leq\lambda\leq\bar\lambda,\ m\leq|\mu|\leq M\right\}$$

where $0 < m < M$ and $0 < \frac{c}{m^2\sqrt{n}} < \bar{\lambda} < 1$. Then, $\Theta_n(m, M, c, \overline{\lambda}) \subset \Theta_n(m, c)$. We consider $\theta_1 = (\lambda, \mu_1)$ and $\theta_2 = (\lambda, \mu_2)$; their values will be chosen later to ensure that $(\theta_1, \theta_2) \in \Theta_n(m, M, c, \overline{\lambda})^2$. According to Lemma 6.1 applied with $\delta = \frac{\lambda|\mu_1-\mu_2|}{2}$, we can write:

$$\inf_{\hat{\theta}} \sup_{\theta\in\Theta_n(m,c)} \mathbb{E}[\lambda^2(\hat{\mu}-\mu)^2] \geq \inf_{\hat{\theta}} \sup_{\theta\in\Theta_n(m,M,c,\overline{\lambda})} \mathbb{E}[\lambda^2(\hat{\mu}-\mu)^2] \tag{6.2}$$
$$\geq \frac{\delta^2}{2}\left\{1-\sqrt{\frac{n}{2}\mathrm{KL}(\mathbb{P}_{\theta_1},\mathbb{P}_{\theta_2})}\right\}.$$

We can compute the Kullback-Leibler divergence between the two mixtures $\mathbb{P}_{\theta_1}$ and $\mathbb{P}_{\theta_2}$: if $f_1 = (1-\lambda)\phi + \lambda\phi_{\mu_1}$ (resp. $f_2 = (1-\lambda)\phi + \lambda\phi_{\mu_2}$) is the density of $\mathbb{P}_{\theta_1}$ (resp. $\mathbb{P}_{\theta_2}$) w.r.t. the Lebesgue measure, we have:

$$\begin{aligned}
\mathrm{KL}(\mathbb{P}_{\theta_1},\mathbb{P}_{\theta_2}) &= \int \log\left[\frac{f_1(x)}{f_2(x)}\right] f_1(x)dx \\
&= \int \log\left[1+\frac{f_1(x)-f_2(x)}{f_2(x)}\right] f_1(x)dx \\
&\leq \int \frac{f_1(x)-f_2(x)}{f_2(x)} f_1(x)dx,
\end{aligned}$$

where we used the inequality $\log(1+t) \leq t$. If we once again write $f_1 = f_2 + f_1 - f_2$, we obtain:

$$\begin{aligned}
\mathrm{KL}(\mathbb{P}_{\theta_1},\mathbb{P}_{\theta_2}) &\leq \int \frac{f_1(x)-f_2(x)}{f_2(x)}\left[f_2(x)+f_1(x)-f_2(x)\right]dx \\
&= \int \frac{[f_1(x)-f_2(x)]^2}{f_2(x)}dx \\
&\leq \lambda^2 \int \frac{[\phi_{\mu_1}(x)-\phi_{\mu_2}(x)]^2}{(1-\lambda)\phi(x)+\lambda\phi_{\mu_2}(x)}dx
\end{aligned}$$

since $f_2(x) \geq (1-\lambda)\phi(x)$ and $f_1(x) - f_2(x) = \lambda[\phi_{\mu_1}(x) - \phi_{\mu_2}(x)]$. On the basis of Assumption ($\mathbf{H_{Lip}}$), we know that $|\phi_{\mu_1} - \phi_{\mu_2}| \leq |\mu_1 - \mu_2|g$ and we obtain:

$$\mathrm{KL}(\mathbb{P}_{\theta_1},\mathbb{P}_{\theta_2}) \leq \frac{\lambda^2(\mu_1-\mu_2)^2\mathcal{J}}{1-\bar{\lambda}}, \tag{6.3}$$

where $\mathcal{J} := \|g\phi^{-1/2}\|^2$ is the constant involved in ($\mathbf{H_{Lip}}$).

We now choose λ, μ_1 and μ_2 so that we obtain the largest possible value in (6.2), while satisfying the constraints given in $\Theta_n(m, M, c, \overline{\lambda})$. Without loss of generality, we set $\mu_1 < \mu_2$ and we need to find a choice of these parameters such that $m \leq \mu_1 < \mu_2 \leq M$ and $\frac{c}{\mu_1^2\sqrt{n}} \leq \lambda \leq \bar{\lambda}$. We set:

$$\mu_1 = m \qquad \text{and} \qquad \lambda = \frac{c}{m^2\sqrt{n}} < \bar{\lambda}.$$

For a given $\epsilon > 0$, we choose μ_2 such that $\frac{n}{2}\mathrm{KL}(\mathbb{P}_{\theta_1}, \mathbb{P}_{\theta_2}) \leq 1 - \epsilon$. Using (6.3), we arrive at the calibration:

$$\mu_2 - \mu_1 = \sqrt{\frac{2(1-\bar{\lambda})(1-\epsilon)}{\lambda^2 \mathcal{J} n}}.$$

It remains to check that $\mu_2 \leq M$. From our choice of λ and μ_1, we see that:

$$\mu_2 = m\left[1 + \sqrt{\frac{2(1-\bar{\lambda})m^2}{c^2\mathcal{J}}(1-\epsilon)}\right] \leq m\left[1 + \sqrt{\frac{2m^2(1-\epsilon)}{c^2\mathcal{J}}}\right]$$

which can be made smaller than M if $1 - \epsilon \leq \frac{c^2\mathcal{J}(M-m)^2}{2m^4}$. If we plug these choices of λ, μ_1 and μ_2 into (6.2), we obtain:

$$\begin{aligned}
\inf_{\hat{\theta}} \sup_{\theta \in \Theta_n(m,M,c,\bar{\lambda})} \mathbb{E}[(\hat{\mu} - \mu)^2] &\geq \lambda^2 \times \frac{2(1-\bar{\lambda})(1-\epsilon)}{8\lambda^2\mathcal{J}n} \times \left[1 - \sqrt{1-\epsilon}\right] \\
&\geq \frac{(1-\bar{\lambda})(1-\epsilon)\epsilon}{8\mathcal{J}n},
\end{aligned}$$

which is the desired lower bound of the minimax risk (4.1).

Point (ii). We keep the same Φ and define $\rho(\theta_1, \theta_2) = |\lambda_1 - \lambda_2| = \rho(\theta_2, \theta_1)$. We consider $\theta_1 = (\lambda_1, \mu)$ and $\theta_2 = (\lambda_2, \mu)$ such that $|\lambda_1 - \lambda_2| = \frac{\epsilon}{\sqrt{n}}$ and

$$\frac{c}{m^2\sqrt{n}} = \lambda_1 < \lambda_2 \leq \bar{\lambda},$$

μ and ϵ have to be chosen hereafter. Since $\lambda_2 = \lambda_1 + \frac{\epsilon}{\sqrt{n}} \leq \bar{\lambda}$, we must choose ϵ such that:

$$\epsilon \leq \bar{\lambda}\sqrt{n} - \frac{c}{m^2}, \tag{6.4}$$

which is possible since we assumed that $\frac{c}{m^2\sqrt{n}} < \bar{\lambda}$. From Lemma 6.1,

$$\begin{aligned}
\inf_{\hat{\theta}} \sup_{\theta \in \Theta_n(m,c)} \mathbb{E}[(\lambda - \hat{\lambda})^2] &\geq \inf_{\hat{\theta}} \sup_{\theta \in \Theta_n(m,M,c,\bar{\lambda})} \mathbb{E}[(\lambda - \hat{\lambda})^2] \\
&\geq \frac{\epsilon^2}{2n}\left\{1 - \sqrt{\frac{n}{2}\mathrm{KL}(\mathbb{P}_{\theta_1}, \mathbb{P}_{\theta_2})}\right\}.
\end{aligned}$$

We can upper bound the Kullback-Leibler divergence as:

$$\begin{aligned}
\mathrm{KL}(\mathbb{P}_{\theta_1}, \mathbb{P}_{\theta_2}) &\leq \int \log\left[f_1(x) - f_2(x)\right]^2 f_2(x)^{-1} dx \\
&\leq (\lambda_1 - \lambda_2)^2 \int \left[\phi_\mu(x) - \phi(x)\right]^2 f_2(x)^{-1} dx \\
&\leq \frac{(\lambda_1 - \lambda_2)^2 \mu^2}{1 - \bar{\lambda}} \int g(x)^2 \phi(x)^{-1} dx \\
&\leq \frac{\mu^2 \epsilon^2 \mathcal{J}}{(1 - \bar{\lambda}) n}.
\end{aligned}$$

By choosing:

$$\mu = \frac{m + M}{2} \qquad \text{and} \qquad \epsilon \leq \sqrt{\frac{2(1 - \bar{\lambda})}{\mathcal{J}(m + M)^2}}, \tag{6.5}$$

we obtain $\frac{n}{2}\mathrm{KL}(\mathbb{P}_{\theta_1}, \mathbb{P}_{\theta_2}) \leq \frac{1}{4}$. Considering the minimal admissible value of ϵ in (6.4) and (6.5) now leads to a choice of the parameters θ_1 and θ_2 such that:

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta_n(m,c)} \mathbb{E}[(\lambda - \hat{\lambda})^2] \geq \frac{\epsilon^2}{4n}.$$

This last inequality is the second lower bound (4.2). □

6.3. *Lower bound for the weak contamination model.*

PROOF OF THEOREM 4.2

Point (i). We consider $\Phi(t) = t^2$ and the loss function ρ defined as:

$$\rho(\theta_1, \theta_2) = \mu_1^2 |\lambda_1 - \lambda_2|.$$

Note that ρ satisfies (6.1) since $\forall (\theta_1, \theta_2, \theta_3) \in \Theta_n(c)^3$,

$$\begin{aligned}
\rho(\theta_1, \theta_3) + \rho(\theta_2, \theta_3) &= \mu_1^2 |\lambda_1 - \lambda_3| + \mu_2^2 |\lambda_2 - \lambda_3| \\
&\geq \min(\mu_1^2, \mu_2^2) |\lambda_1 - \lambda_2| \\
&\geq \rho(\theta_1, \theta_2) \wedge \rho(\theta_2, \theta_1).
\end{aligned}$$

To obtain a convenient lower bound, we need to use Lemma 6.1 and find a couple of parameters (θ_1, θ_2) that belongs to the admissible set and such that $\mathrm{KL}(\mathbb{P}_{\theta_1}, \mathbb{P}_{\theta_2})$ is small enough. In particular, the proximity between $\mathbb{P}_{\theta_1}$ and $\mathbb{P}_{\theta_2}$ will be obtained by a careful matching of the first moments of the two distributions, which is a good method for obtaining efficient lower bounds in mixture models (see, *e.g.*, [BG14] or [HK15]). We give an example of this method below. First, remark that:

$$\mathrm{KL}(\mathbb{P}_{\theta_1}, \mathbb{P}_{\theta_2}) = \int \log\left[\frac{f_1(x)}{f_2(x)}\right] f_1(x)dx.$$

Since ϕ is a piecewise $\mathcal{C}^3$ function on $\mathbb{R}$, considering a shift $\mu = o(1)$, we can write a third order Taylor expansion:

$$\forall x \in \mathbb{R} \quad \phi_\mu(x) = \phi(x) - \mu\phi'(x) + \frac{\mu^2}{2}\phi''(x) - \frac{\mu^3}{6}\phi^{(3)}(\xi_{x,\mu}),$$

where $\xi_{x,\mu}$ belongs to the interval defined by x and $x-\mu$. In particular, assuming that $\phi^{(3)}$ is bounded on $\mathbb{R}$ leads to:

$$\forall x \in \mathbb{R} \quad \phi_\mu(x) = \phi(x) - \mu\phi'(x) + \frac{\mu^2}{2}\phi''(x) + o(\mu^2).$$

This Taylor expansion permits us to write, for small values of μ_1:

$$\begin{aligned}
\log[f_1(x)] &= \log[(1-\lambda_1)\phi(x) + \lambda_1\phi_{\mu_1}(x)] \\
&= \log\left[(1-\lambda_1)\phi(x) + \lambda_1\phi(x) - \lambda_1\mu_1\phi'(x) + \frac{1}{2}\lambda_1\mu_1^2\phi''(x) + o(\mu_1^2)\right] \\
&= \log[\phi(x)] + \log\left[1 - \lambda_1\mu_1\frac{\phi'(x)}{\phi(x)} + \frac{1}{2}\lambda_1\mu_1^2\frac{\phi''(x)}{\phi(x)} + o(\mu_1^2)\right] \\
&= \log[\phi(x)] - \lambda_1\mu_1\frac{\phi'(x)}{\phi(x)} + \frac{1}{2}\lambda_1\mu_1^2\frac{\phi''(x)}{\phi(x)} - \frac{1}{2}\lambda_1^2\mu_1^2\left(\frac{\phi'(x)}{\phi(x)}\right)^2 + o(\mu_1^2).
\end{aligned}$$

In the same way, for small values of μ_2:

$$\begin{aligned}
\log[f_2(x)] &= \log[(1-\lambda_2)\phi(x) + \lambda_2\phi_{\mu_2}(x)] \\
&= \log[\phi(x)] - \lambda_2\mu_2\frac{\phi'(x)}{\phi(x)} + \frac{1}{2}\lambda_2\mu_2^2\frac{\phi''(x)}{\phi(x)} - \frac{1}{2}\lambda_2^2\mu_2^2\left(\frac{\phi'(x)}{\phi(x)}\right)^2 + o(\mu_2^2).
\end{aligned}$$

We thus obtain:

$$\begin{aligned}
\log[f_1(x)] - \log[f_2(x)] &= (\lambda_2\mu_2 - \lambda_1\mu_1)\frac{\phi'(x)}{\phi(x)} + \frac{1}{2}(\lambda_1\mu_1^2 - \lambda_2\mu_2^2)\frac{\phi''(x)}{\phi(x)} \\
&\quad + \frac{1}{2}(\lambda_2^2\mu_2^2 - \lambda_1^2\mu_1^2)\left(\frac{\phi'(x)}{\phi(x)}\right)^2 + o(\mu_1^2) + o(\mu_2^2).
\end{aligned}$$

In particular, we observe that the term above can be considered as a "second order term" if θ_1 and θ_2 are chosen such that $\lambda_1\mu_1 = \lambda_2\mu_2$, which corresponds to the first moment of $\mathbb{P}_{\theta_1}$ and $\mathbb{P}_{\theta_2}$. If $\lambda_1\mu_1 = \lambda_2\mu_2$, we obtain:

$$\log[f_1(x)] - \log[f_2(x)] = \frac{1}{2}(\lambda_1\mu_1^2 - \lambda_2\mu_2^2)\frac{\phi''(x)}{\phi(x)} + o(\mu_1^2) + o(\mu_2^2).$$

We deduce that:

$$
\begin{aligned}
&\mathrm{KL}(\mathbb{P}_{\theta_1}, \mathbb{P}_{\theta_2}) \\
&= \int \left[\frac{1}{2}(\lambda_1\mu_1^2 - \lambda_2\mu_2^2)\frac{\phi''(x)}{\phi(x)} + o(\mu_1^2) + o(\mu_2^2)\right] f_1(x)dx \\
&= \frac{1}{2}(\lambda_1\mu_1^2 - \lambda_2\mu_2^2)\left[(1-\lambda_1)\int \phi''(x)dx + \lambda_1 \int \frac{\phi''(x)\phi(x-\mu_1)}{\phi(x)}dx\right] + o(\mu_1^2) + o(\mu_2^2).
\end{aligned}
$$

The smoothness and symmetry of ϕ leads to $\int \phi''(x)dx = 0$. We deduce that:

$$
\begin{aligned}
&\int \frac{\phi''(x)\phi(x-\mu_1)}{\phi(x)}dx \\
&\quad= \int \frac{\phi''(x)}{\phi(x)}[\phi(x) - \mu_1\phi'(x) + \frac{1}{2}\phi''(x) + o(\mu_1^2)]dx \\
&\quad= \int \phi''(x)dx - \mu_1 \int \frac{\phi''(x)\phi'(x)}{\phi(x)}dx + \frac{1}{2}\mu_1^2 \int \frac{\phi''(x)}{\phi(x)}dx + o(\mu_1^2)dx \\
&\quad= \frac{1}{2}\mu_1^2\mathcal{I}_\phi + \underset{n\to+\infty}{o}(\mu_1^2),
\end{aligned}
$$

where the last line comes from the fact that ϕ satisfies $(\mathbf{H_D})$ and that $x \mapsto \phi''(x)\phi'(x)/\phi(x)$ is an odd function. Finally, since $\lambda_1\mu_1 = \lambda_2\mu_2$, we deduce that:

$$
\begin{aligned}
\mathrm{KL}(\mathbb{P}_{\theta_1}, \mathbb{P}_{\theta_2}) &= \frac{1}{4}(\lambda_1\mu_1^2 - \lambda_2\mu_2^2)\lambda_1\mu_1^2\mathcal{I}_\phi + o(\mu_1^4) \\
&= \frac{1}{4}\left(1 - \frac{\lambda_1}{\lambda_2}\right)\lambda_1^2\mu_1^4\mathcal{I}_\phi + o(\mu_1^4). \qquad (6.6)
\end{aligned}
$$

Next, let $\bar{\lambda} \in (0,1)$. Choosing $\lambda_2 = \frac{\bar{\lambda}}{2} < \bar{\lambda}$ and $\lambda_1 = \frac{1}{\alpha}\lambda_2$ with $\alpha = \frac{1+\sqrt{5}}{2}$, we have:

$$\left(1 - \frac{\lambda_1}{\lambda_2}\right)\lambda_1^2 = (\lambda_1 - \lambda_2)^2.$$

Thus,

$$\mathrm{KL}(\mathbb{P}_{\theta_1}, \mathbb{P}_{\theta_2}) = \frac{1}{4}(\lambda_2 - \lambda_1)^2\mu_1^4\mathcal{I}_\phi + o(\mu_1^4).$$

In order to apply Lemma 6.1, let $\delta > 0$ such that $2\delta = \rho(\theta_1, \theta_2) \wedge \rho(\theta_2, \theta_1)$. According to our constraint $\lambda_1\mu_1 = \lambda_2\mu_2$ and $\lambda_2 = \alpha\lambda_1 > \lambda_1$, we observe that $\mu_2 < \mu_1$ so that:

$$2\delta = \mu_2^2|\lambda_1 - \lambda_2|.$$

We deduce that:

$$|\lambda_1 - \lambda_2|\mu_1^2 = |\lambda_1 - \lambda_2|\left(\frac{\lambda_2}{\lambda_1}\right)^2\mu_2^2 = 2\delta\alpha^2$$

and

$$\mu_1^2 = \left(\frac{\lambda_2}{\lambda_1}\right)^2 \mu_2^2 = \alpha^2 \frac{4\alpha}{(\alpha-1)\bar{\lambda}}\delta.$$

Thus,

$$\mathrm{KL}(\mathbb{P}_{\theta_1}, \mathbb{P}_{\theta_2}) = \delta^2 \alpha^4 \mathcal{I}_\phi + o(\delta^2),$$

and according to Lemma 6.1, we obtain:

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta_n(c)} \mathbb{E}[\mu^4(\lambda - \hat{\lambda})^2] \geq \frac{\delta^2}{2}\left\{1 - \sqrt{\frac{n}{2}\delta^2\left[\alpha^4 \mathcal{I}_\phi + o(1)\right]}\right\}.$$

The choice of δ is determined by the right brackets that should be non-negative. We can choose:

$$\delta = \left[2n\alpha^4 \mathcal{I}_\phi\right]^{-\frac{1}{2}},$$

so that $\frac{n}{2}\delta^2\left[\alpha^4 \mathcal{I}_\phi + o(1)\right] = \frac{1}{4}(1+o(1))$. Thus, an integer N exists such that:

$$\forall n \geq N \qquad \inf_{\hat{\theta}} \sup_{\theta \in \Theta_n(c)} \mathbb{E}[\mu^4(\lambda - \hat{\lambda})^2] \geq \frac{\delta^2}{6} = \frac{1}{12\alpha^4 \mathcal{I}_\phi n}.$$

This ends the proof of the first point.

Point (ii). We define in this case the loss function $\rho(\theta_1, \theta_2) = \lambda_1|\mu_1||\mu_1 - \mu_2|$ and $\Phi(t) = t^2$. The function ρ satisfies the weak triangle inequality (6.1) since $\forall(\theta_1, \theta_2, \theta_3) \in \Theta_n(c)^3$:

$$\begin{aligned} \rho(\theta_1, \theta_3) + \rho(\theta_2, \theta_3) &= \lambda_1|\mu_1||\mu_1 - \mu_3| + \lambda_2|\mu_2||\mu_2 - \mu_3| \\ &\geq \min(\lambda_1|\mu_1|, \lambda_2|\mu_2|)|\mu_1 - \mu_2| \\ &\geq \rho(\theta_1, \theta_2) \wedge \rho(\theta_2, \theta_1). \end{aligned}$$

The proof follows the same lines as the ones of (i) and our starting point is once again the Kullback-Leibler divergence asymptotics given in Equation (6.6). Our baseline relationship $\lambda_1\mu_1 = \lambda_2\mu_2$ is still necessary and we obtain:

$$\mathrm{KL}(\mathbb{P}_{\theta_1}, \mathbb{P}_{\theta_2}) = \frac{\mathcal{I}_\phi}{4}\left(1 - \frac{\mu_2}{\mu_1}\right)\lambda_1^2 \mu_1^4 + o(\mu_1^4).$$

We choose $\mu_1 = 2\mu_2$ so that $\lambda_2 = 2\lambda_1$ and:

$$\rho(\theta_1, \theta_2) \wedge \rho(\theta_2, \theta_1) = \lambda_1\mu_1|\mu_1 - \mu_2| = \frac{1}{2}\lambda_1\mu_1^2 := 2\delta.$$

The coefficients λ_1 and λ_2 can be made explicit, e.g., $\lambda_1 = \bar{\lambda}/2$ and $\lambda_2 = \bar{\lambda}$. This choice implies that $\mu_1 = 2\sqrt{2\delta/\bar{\lambda}}$. These settings can be used in the result of Lemma 6.1 and we obtain:

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta_n(c)} \mathbb{E}[\lambda^2 \mu^2 (\mu - \hat{\mu})^2] \geq \frac{\delta^2}{2} \left\{ 1 - \sqrt{\frac{n\delta^2}{2} \left[2\mathcal{I}_\phi + o(1) \right]} \right\}.$$

We can obtain an efficient lower bound by choosing:

$$\delta_n := \frac{1}{2\sqrt{n\mathcal{I}_\phi}},$$

which implies, of course, that $\mu_1 = o(1)$ and $\mu_2 = o(1)$. According to this choice, an integer N exists such that $\forall n \geq N$:

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta_n(c)} \mathbb{E}[\lambda^2 \mu^2 (\mu - \hat{\mu})^2] \quad \geq \quad \frac{1}{8n\mathcal{I}_\phi} \times (1 - \frac{1}{2})/2 = \frac{1}{32n\mathcal{I}_\phi}.$$

This ends the proof of the second point. $\square$

7. Simulation study.

Distributions. In this section, we assess the performance of the $\mathbb{L}^2$-estimator given in (2.2) on four particular cases of baseline density ϕ. We study the following features:

- Standard Gaussian case with $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$.
- Non-smooth distribution with the Laplace density $\phi(x) = \frac{1}{2} e^{-|x|}$.
- Heavy tailed distribution with the Cauchy density: $\phi(x) = \frac{1}{\pi(1+x^2)}$.
- Asymmetry with the skew Gaussian density: $\phi(x) = 2\psi(x)\Psi(\alpha x)$, where ψ and Ψ, respectively, denote the density and the cumulative function of the standard Gaussian distribution and where α is the asymmetry parameter different from 0 (in the simulations, we fix $\alpha = 10$). This example of asymmetric distributions has been introduced in the recent work on mixture models of [Lin09].

REMARK 7.1. *An easy consequence of Proposition A.2 is that the log-concave Gaussian and Laplace distributions satisfy assumptions* $(\mathbf{H}_\mathcal{S})$, $(\mathbf{H}_{\mathbf{Lip}})$ *and* $(\mathbf{H}_\mathbf{D})$ *so that Theorems 4.1 and 4.2 apply to these situations.*
Concerning the Cauchy distribution case, we can explicitly compute $\phi - \phi_\mu$*:*

$$|\phi(x) - \phi_\mu(x)| = |\mu| \frac{|2x - \mu|}{\pi[1 + (x-\mu)^2][1 + x^2]} \leq C\phi(x)|\mu|,$$

for a large enough constant C. Hence, the assumptions $(\mathbf{H}_{\mathcal{S}}), (\mathbf{H}_{\mathbf{Lip}})$ and $(\mathbf{H}_{\mathbf{D}})$ are satisfied with $g = C\phi$ for the Cauchy distribution.
The skew Gaussian density ϕ satisfies:

$$|\phi(x) - \phi_\mu(x)| \leq 2\psi(x) \left|\Psi(\alpha x) - \Psi(\alpha(x-\mu))\right| + 2\Psi(\alpha(x-\mu)) \left|\psi(x) - \psi(x-\mu)\right|.$$

If we define g as $g(x) := 4 \sup_{[x-M;x+M]} \psi(t) \times \sup_{[x-M;x+M]} \Psi(\alpha t)$, we can check that $(\mathbf{H}_{\mathcal{S}}), (\mathbf{H}_{\mathbf{Lip}})$ and $(\mathbf{H}_{\mathbf{D}})$ are satisfied. In particular, the integrability condition $(\mathbf{H}_{\mathbf{D}})$ is satisfied for large x because $\Psi(\alpha x) \longrightarrow 1$ when $x \longrightarrow +\infty$. Conversely, if $x \longrightarrow -\infty$, we have:

$$\begin{aligned} g^2(x)\phi^{-1}(x) &\lesssim \left[\psi^{-1}(x)\Psi^{-1}(\alpha x)\right] \sup_{[x-M;x+M]} \psi^2(t) \times \sup_{[x-M;x+M]} \Psi^2(\alpha t) \\ &\lesssim \left[\alpha x e^{x^2/2} e^{\alpha^2 x^2/2}\right] e^{-(x-M)^2} \times e^{-\alpha^2(x-M)^2} [\alpha(x-M)]^{-2} \\ &\lesssim e^{-(x-2M)^2/4} e^{-\alpha^2(x-2M)^2/4}, \end{aligned}$$

which leads to the integrability condition around $-\infty$.

Our estimator requires the calculation of the contrast γ_n and, in particular, the value of the $\mathbb{L}^2$ norm:

$$\|f_{\lambda,\mu}\|^2 = \left[\lambda^2 + (1-\lambda)^2\right] \|\phi\|^2 + 2\lambda(1-\lambda)\langle \phi, \phi_\mu \rangle,$$

that involves the value of inner product $\langle \phi, \phi_\mu \rangle$ for any value of the location parameter $\mu \in [-M, M]$. In the first three examples of distributions, a closed formula exists:

- Gaussian density: $\langle \phi, \phi_\mu \rangle = (4\pi)^{-\frac{1}{2}} \exp\left[-\frac{1}{4}\mu^2\right]$
- Laplace density: $\langle \phi, \phi_\mu \rangle = \frac{1}{4} e^{-|\mu|}(1 + |\mu|)$
- Cauchy density: $\langle \phi, \phi_\mu \rangle = \frac{2}{\pi(4+\mu^2)}$

Unfortunately, such a formula is not available (to our knowledge) for the skew Gaussian density: there is no analytical expression of $\langle \phi, \phi_\mu \rangle$. Instead, we used a Monte-Carlo procedure to evaluate this quantity for each value of μ in our grid $\mathcal{M}_n$ given in (3.2). To obtain a sufficient approximation of these inner products, we used a number of Monte-Carlo iterations T_{MC} each time of the order $T_{MC} \propto n^2$ (where n will be the sample size used for our estimation problem).

Statistical setting. We have worked with a fixed value of $\lambda^\star = \frac{1}{4}$ while $\mu^\star$ is allowed to vary with n. Below, we used the following relationship between $\mu^\star$ and n:

$$\mu^\star = \sqrt{\frac{1}{\lambda^\star n^\nu}} \qquad \text{with} \quad \nu = \frac{\alpha}{24}, \ \alpha \in \{1, \ldots, 24\}.$$

For each value of the parameter $\mu^\star$, we used 10^3 Monte-Carlo simulations to obtain reliable results, while the grid size is determined by fixing the maximal value of the unknown $|\mu^\star|$ as $M = 10$. Finally, we sampled a set of $n = 5000$ observations each time.

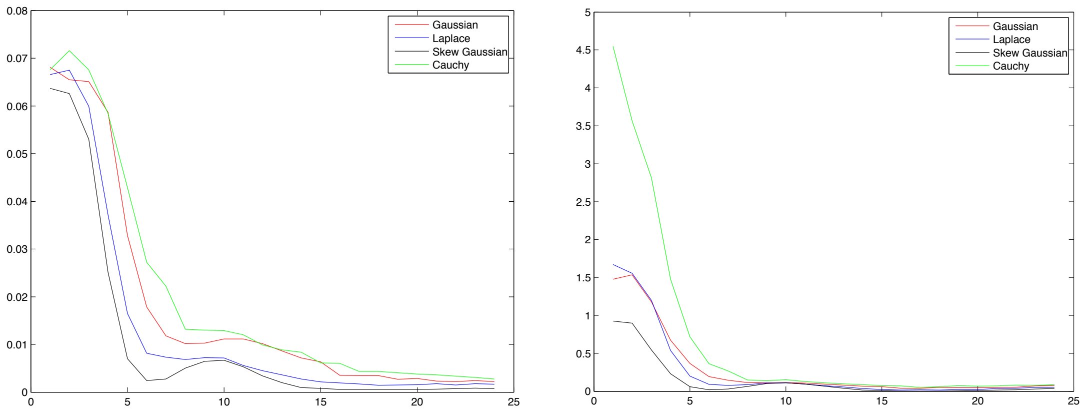

FIG 1. *Mean square error of estimating $\lambda^\star$ (left) and $\mu^\star$ (right) for the 24 values of ν in descending order.*

In Fig. 1, for each case of the mixture model, we represent the evolution of the mean square error for the estimation of $\lambda^\star$ and of $\mu^\star$ when ν varies between 1/24 and 1:

$$\nu \longmapsto \mathrm{MSE}(\lambda) = \frac{1}{10^3} \sum_{j=1}^{10^3} (\hat{\lambda}_j - \lambda^\star)^2$$

and

$$\nu \longmapsto \mathrm{MSE}(\mu) = \frac{1}{10^3} \sum_{j=1}^{10^3} (\hat{\mu}_j - \mu^\star)^2.$$

As pointed out in Fig. 1, the estimation of $\lambda^\star$ and $\mu^\star$ performs quite well as soon as ν is lower than 1/2 but becomes completely inconsistent when $\nu > 1/2$, even if we use a sample size of 5000 observations.

We also represent the violin plot of these estimations indicating the same behavior in each particular case (Gaussian and Laplace in Fig. 2; Cauchy and skew Gaussian in Fig. 3).

Again, a similar conclusion holds: the estimators derived from (2.2) exhibit a low bias and variance when ν is chosen small enough (lower than 1/2, which corresponds to values greater than 12 in the horizontal axes of Figs.

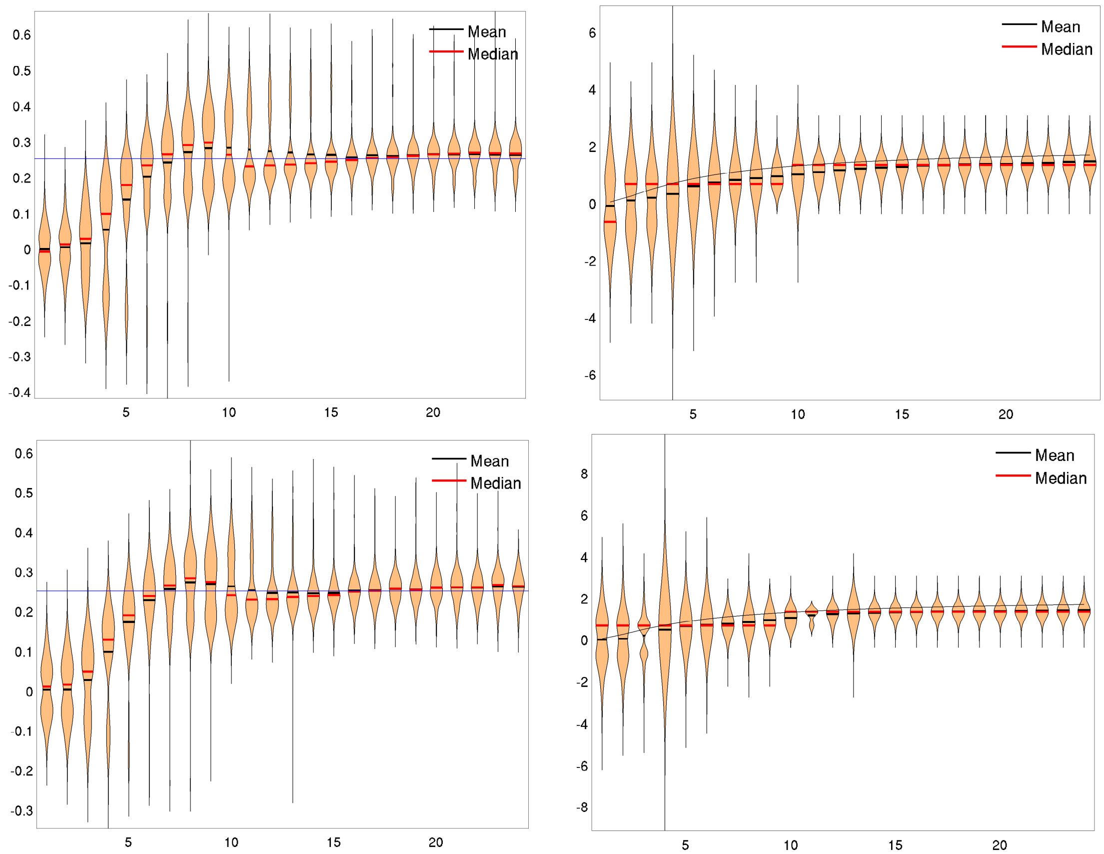

FIG 2. *Evaluation of $\lambda^\star$ (on the left) and $\mu^\star$ (on the right) for our estimators when* ***Gaussian*** *mixtures (top) and* ***Laplace*** *mixtures (bottom) are considered, for the 24 values of ν in descending order.*

2-3). In contrast, the estimation is seriously damaged for values of ν greater than 1/2 (which corresponds to values lower than 11 in the horizontal axes of Figs. 2-3). Finally, it should be noted that the shape of the density ϕ does not seem to have a big influence on the estimation ability, even though the Cauchy distribution settings may be seen as the most difficult problem (as represented by the green MSE in Fig. 1).

APPENDIX A: TECHNICAL RESULTS

A.1. Identifiability result.

PROOF OF PROPOSITION 2.1. We assume that two parameters $\theta_1 = (\lambda_1, \mu_1)$ and $\theta_2 = (\lambda_2, \mu_2)$ exist such that $f_{\theta_1} = f_{\theta_2}$. In that case, consider the Fourier

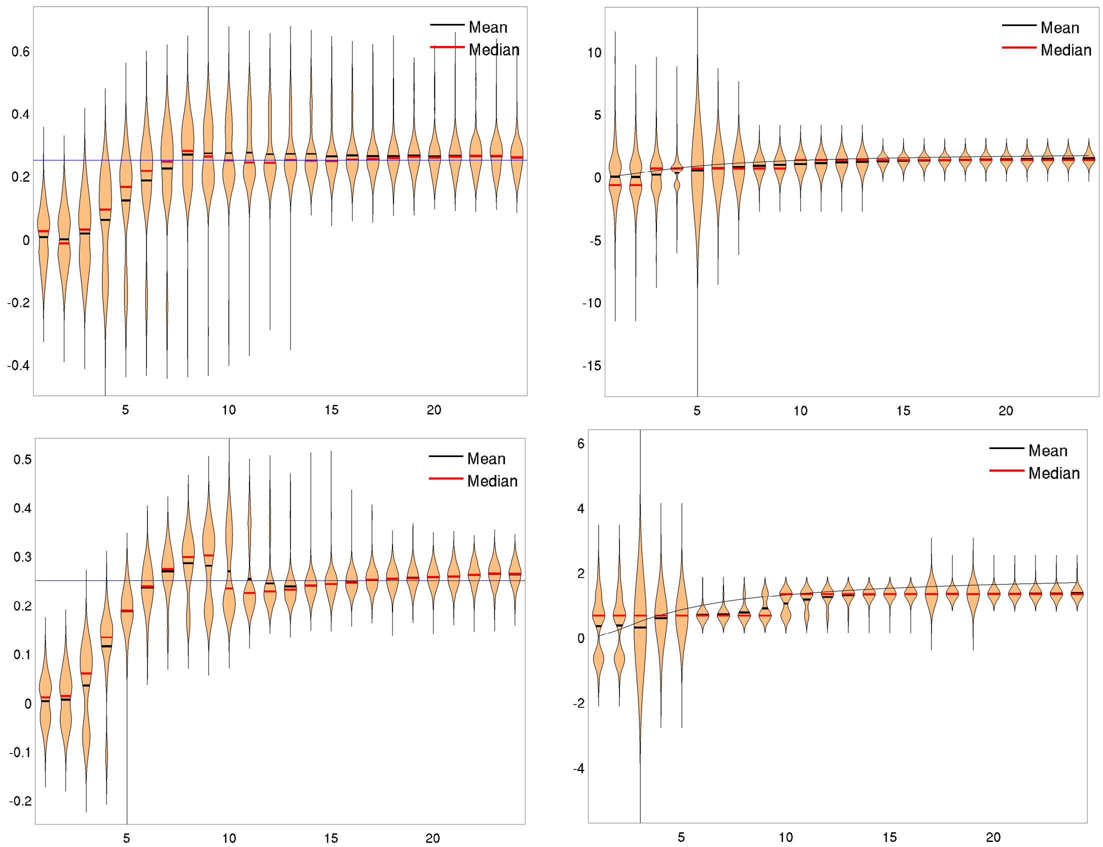

FIG 3. *Evaluation of $\lambda^\star$ (on the left) and $\mu^\star$ (on the right) for our estimators when* ***Cauchy*** *mixtures (top) and* ***skew Gaussian*** *mixtures (bottom) are considered, for the 24 values of ν in descending order.*

transform of X whose density is f_{θ_1}. This Fourier transform is given by

$$\varphi_X(\xi) = \mathbb{E}[e^{\mathrm{i}\xi X}] = \left[(1-\lambda_1) + \lambda_1 e^{\mathrm{i}\xi\mu_1}\right]\hat{\phi}(\xi),$$

where $\hat{\phi}$ is the Fourier transform of ϕ and $\mathfrak{i}$ is the complex number such that $\mathfrak{i}^2 = -1$. Since $f_{\theta_1} = f_{\theta_2}$, we then deduce that:

$$\forall \xi \in \mathbb{R} \qquad \left[(1-\lambda_1) + \lambda_1 e^{\mathrm{i}\xi\mu_1}\right]\hat{\phi}(\xi) = \left[(1-\lambda_2) + \lambda_2 e^{\mathrm{i}\xi\mu_2}\right]\hat{\phi}(\xi).$$

Since $\phi \in \mathbb{L}^1(\mathbb{R})$, $\hat{\phi}$ is continuous and cannot be zero everywhere. Thus, we can find an open interval I such that $\hat{\phi}(\xi) \neq 0$ in I and the Lebesgue measure of I is strictly positive. Hence,

$$\forall \xi \in I \quad (1-\lambda_1) + \lambda_1 e^{\mathrm{i}\xi\mu_1} = (1-\lambda_2) + \lambda_2 e^{\mathrm{i}\xi\mu_2},$$

and from the analytical property of the exponential map, we deduce that:

$$\forall \xi \in I \qquad (1-\lambda_1)+\lambda_1[\cos(\xi\mu_1)+\mathfrak{i}\sin(\xi\mu_1)] = (1-\lambda_2)+\lambda_2[\cos(\xi\mu_2)+\mathfrak{i}\sin(\xi\mu_2)]$$

Identifying now the imaginary parts yields:

$$\forall \xi \in I \qquad \lambda_1 \sin(\xi\mu_1) = \lambda_2 \sin(\xi\mu_2)$$

It is classical that the family of functions $(\xi \mapsto \sin(\alpha_1\xi), \xi \mapsto \sin(\alpha_2\xi))$ is linearly independent if and only if $|\alpha_1| \neq |\alpha_2|$. We can deduce that, necessarily, $\mu_1 = \pm\mu_2$ and, since λ_1 and λ_2 are positive, $\mu_1 = \mu_2$. □

A.2. Connection between $\|\phi - \phi_\mu\|$ and $|\mu|$.

PROPOSITION A.1. *Let any $M > 0$ be given and assume that ϕ satisfies $(\mathbf{H}_\mathcal{S})$ and $(\mathbf{H_{Lip}})$, then two constants $0 < \underline{\kappa} < \overline{\kappa} < +\infty$ exist such that:*

$$\forall (\mu, \tilde{\mu}) \in [-M, M]^2 \qquad \underline{\kappa}(\mu - \tilde{\mu})^2 \leq \|\phi_\mu - \phi_{\tilde{\mu}}\|^2 \leq \overline{\kappa}(\mu - \tilde{\mu})^2. \tag{A.1}$$

PROOF. We prove the upper and lower bounds separately. According to the shift invariance of the $\mathbb{L}^2$ norm, we only establish these inequalities when $\tilde{\mu} = 0$. Using $(\mathbf{H_{Lip}})$, the upper bound simply derives from:

$$\|\phi - \phi_\mu\|^2 = \int_\mathbb{R} [\phi(x) - \phi(x-\mu)]^2 \, dx \leq \int_\mathbb{R} |\mu|^2 g^2(x) dx = \mu^2 \|g\|^2,$$

which is the desired inequality if we choose $\overline{\kappa} = \|g\|^2$. Concerning the lower bound, we have:

$$\frac{\|\phi(.) - \phi(. - \mu)\|^2}{\mu^2} = \int_\mathbb{R} \left[\frac{\phi(x) - \phi(x-\mu)}{\mu}\right]^2 dx.$$

Inequality (3.1) brought by Assumption $(\mathbf{H_{Lip}})$ makes it possible to apply the Lebesgue convergence theorem, which implies:

$$\begin{aligned}
\lim_{\mu \longrightarrow 0} \frac{\|\phi(.) - \phi(. - \mu)\|^2}{\mu^2} &= \int_\mathbb{R} \lim_{\mu \to 0} \left[\frac{\phi(x) - \phi(x-\mu)}{\mu}\right]^2 dx, \\
&= \|\phi'\|^2 > 0.
\end{aligned}$$

Indeed, ϕ being piecewise differentiable ($\phi \in \mathcal{C}^1_p(\mathbb{R})$), $\frac{\phi(x)-\phi(x-\mu)}{\mu} \longrightarrow \phi'(x)$ almost surely when $\mu \longrightarrow 0$.
Now, ϕ is continuous and $\psi : \mu \longrightarrow \frac{\|\phi - \phi_\mu\|^2}{\mu^2} \in \mathcal{C}^0([-M, M], \mathbb{R})$ from the Lebesgue convergence theorem. This continuous map ψ attains its lower

bound on $[-M, M]$ and the identifiability result of Proposition 2.1 implies that this lower bound is positive. This leads to the existence of $\underline{\kappa} > 0$ such that:

$$\|\phi - \phi_\mu\|^2 \geq \underline{\kappa}|\mu|^2.$$

□

A.3. Log-concave distributions. In this section, we establish that most of the log-concave real distributions satisfy the assumptions $(\mathbf{H}_{\mathcal{S}})$, $(\mathbf{H}_{\mathbf{Lip}})$ and $(\mathbf{H}_{\mathbf{D}})$. For this purpose, we introduce the associated class of probability measures:

$$\mathcal{LC} := \left\{\phi = e^{-U} : U \text{ is convex}, U \in \mathcal{C}_p^3(\mathbb{R}) \text{ and } |U'| + |U''| = o_{\pm\infty}(U)\right\}.$$

The set of possible densities is rich. For example, it contains Laplace and Gaussian distributions. However, the set $\mathcal{LC}$ does not capture the situation where $U(x) = e^{|x|}$ or $U(x) = e^{x^2}$ since U exhibits variations that are too great for large values of x.

Proposition A.2. *Assume that μ varies in $[-M, M]$ and that $\phi \in \mathcal{LC}$. Let $\varepsilon \in (0, M)$. If we set:*

$$g(x) := \sqrt{\frac{1}{\varepsilon} \int_{x-M}^{x} \phi'^2} \vee \sqrt{\frac{1}{\varepsilon} \int_{x}^{x+M} \phi'^2} \vee \sup_{u \in [x-\varepsilon, x+\varepsilon]} |\phi'(u)|,$$

$(\mathbf{H}_{\mathbf{Lip}})$ and $(\mathbf{H}_{\mathbf{D}})$ hold:

i) $\forall \mu \in [-M, M] \quad \forall x \in \mathbb{R} \qquad |\phi(x) - \phi_\mu(x)| \leq |\mu|\, g(x).$

ii) $g\phi^{-1/2} \in \mathbb{L}^2(\mathbb{R})$

iii) $\phi''\phi^{-1/2} \in \mathbb{L}^2(\mathbb{R})$

Proof. We provide a proof in the case when $\phi \in \mathcal{C}^2$. This proof can be extended when $\phi \in \mathcal{C}_p^2$ according to some small modifications that are left to the reader.

Proof of *(i)*: Remark first that $\forall \mu \in [-M, M]$:

$$\forall x \in \mathbb{R} \qquad |\phi(x) - \phi_\mu(x)| = \left|\int_{x-\mu}^{x} \phi'(u)du\right| \leq \sqrt{|\mu|}\sqrt{\int_{(x-\mu)\wedge x}^{(x-\mu)\vee x} \phi'^2},$$

where the last upper bound comes from the Cauchy-Schwarz inequality. Let $\varepsilon \in (0, M)$. If $|\mu| \in [\varepsilon, M]$, we obtain that:

$$|\phi(x) - \phi_\mu(x)| \leq |\mu| \left(\sqrt{\frac{g_1(x)}{\varepsilon}} \vee \sqrt{\frac{g_2(x)}{\varepsilon}}\right),$$

where

$$g_1(x) = \int_{x-M}^{x} \phi'(u)^2 du \qquad \text{and} \qquad g_2(x) = \int_{x}^{x+M} \phi'(u)^2 du.$$

If $|\mu| \in [0, \varepsilon)$,

$$|\phi(x) - \phi_\mu(x)| \leq |\mu| \sup_{u \in [x-\varepsilon, x+\varepsilon]} |\phi'(u)| := |\mu| g_{3,\varepsilon}(x).$$

Proof of (ii): In order to prove that $g\phi^{-1/2} \in \mathbb{L}^2(\mathbb{R})$, we separately prove that $g_1\phi^{-1}, g_2\phi^{-1}$ and $g_{3,\varepsilon}^2\phi^{-1}$ belong to $\mathbb{L}^1(\mathbb{R})$.

Case $x \longrightarrow +\infty$. Since U is convex, U' is an increasing map, and for large enough x, U' is positive (otherwise U would not diverge to $+\infty$ and ϕ would not be in $\mathbb{L}^1(\mathbb{R})$). Then, $x_0 > 0$ exists such that:

$$\begin{aligned}
\forall x \geq x_0 \qquad g_1(x)\phi^{-1}(x) &= e^{U(x)} \int_{x-M}^{x} U'(u)^2 e^{-2U(u)} du \\
&\leq U'(x) e^{U(x)} \int_{x-M}^{x} U'(u) e^{-2U(u)} du \\
&\leq U'(x) e^{U(x)} \frac{e^{-2U(x-M)} - e^{-2U(x)}}{2} \\
&\leq \frac{U'(x)}{2} e^{-2U(x-M)+U(x)}.
\end{aligned}$$

The mean value theorem leads to:

$$\forall x \geq x_0 \quad \exists \xi \in [x-M, x] \qquad U(x-M) = U(x) - MU'(\xi) \geq U(x) - MU'(x).$$

Consequently, we obtain:

$$\forall x \geq x_0 \qquad g_1(x)\phi^{-1}(x) \leq \frac{U'(x)}{2} e^{-U(x)+2MU'(x)}.$$

The density $\phi \in \mathcal{LC}$ and we can find K large enough such that:

$$\forall |x| \geq K \qquad -U(x) + 2MU'(x) \leq -(1-\eta)U(x)$$

For such an x, we have $g_1(x)\phi^{-1}(x) \leq \frac{U'(x)}{2} e^{-(1-\eta)U(x)} \in \mathbb{L}^1(\mathbb{R})$.
Concerning $g_2(x)\phi(x)^{-1}$, we can now use a closed argument to obtain:

$$\begin{aligned}
g_2(x)\phi^{-1}(x) &\leq \frac{U'(x+M)}{2} e^{-U(x)} \\
&\leq \frac{U'(x+M)}{2} e^{-U(x+M)+MU'(x+M)} \\
&\leq \frac{U'(x+M)}{2} e^{-(1-\eta)U(x+M)}.
\end{aligned}$$

Hence, $g_2(x)\phi^{-1} \in \mathbb{L}^1(\mathbb{R})$. We now consider $g_{3,\varepsilon}^2\phi^{-1}$:

$$g_{3,\varepsilon}^2(x)\phi^{-1}(x) = \sup_{u\in[x-\varepsilon,x+\varepsilon]} U'(u)^2 e^{-2U(u)+U(x)}.$$

If $u \in [x-\varepsilon, x]$, the mean value theorem leads to:

$$\begin{aligned} U(u) &= U(x) - (x-u)U'(\xi) \text{ with } \xi \in]u, x[\\ &\geq U(x) - \varepsilon U'(x). \end{aligned}$$

Then,

$$U'(u)^2 e^{-2U(u)+U(x)} \leq U'(u)^2 e^{-U(x)+2\varepsilon U'(x)} \leq U'(x)^2 e^{-U(x)+2\varepsilon U'(x)}.$$

Using the fact that $|U''| + |U'| = o_{\pm\infty}(U)$, we can find a positive constant $C > 0$, a parameter $\eta \in (0,1)$ and for K large enough such that $\forall x \geq K$:

$$U'(u)^2 e^{-2U(u)+U(x)} \leq CU'(x)e^{-(1-\eta)U(x)}. \tag{A.2}$$

If $u \in [x, x+\varepsilon]$, $U(u) \geq U(x+\varepsilon) - \varepsilon U'(x+\varepsilon)$ and $U(x+\varepsilon) = U(x) + \varepsilon U'(\tilde{\xi}) \geq U(x)$ since $U'(\tilde{\xi}) \geq 0$ for all $x \geq x_0$. Thus,

$$\begin{aligned} -2U(u) + U(x) &\leq -2U(x+\varepsilon) + 2\varepsilon U'(x+\varepsilon) + U(x) \\ &\leq -U(x+\varepsilon) + 2\varepsilon U'(x+\varepsilon). \end{aligned}$$

Then, in the same way, we can find a positive constant $C > 0$, a parameter $\eta \in (0,1)$ and a large enough K such that $\forall x \geq K$,

$$U'(u)^2 e^{-2U(u)+U(x)} \leq CU'(x+\varepsilon)e^{-(1-\eta)U(x+\varepsilon)}. \tag{A.3}$$

Thus, (A.2) and (A.3) imply that $g_{3,\varepsilon}^2\phi^{-1} \in \mathbb{L}^1(\mathbb{R}_+)$. As a maximum of three functions in $\mathbb{L}^1(\mathbb{R})$, we deduce that $g^2\phi^{-1} \in \mathbb{L}^1(\mathbb{R}_+)$.

Case $x \longrightarrow -\infty$. A careful inspection of the proof above allows us to use similar inequalities to show that $g_1\phi^{-1}$, $g_2\phi^{-1}$ and $g_{3,\varepsilon}\phi^{-1}$ belong to $\mathbb{L}^2(\mathbb{R})$.

Proof of (iii): A direct computation shows that, almost surely:

$$\phi''^2\phi^{-1} = [U'' - U'^2]^2 e^{-U} \leq 2U''^2 e^{-U} + 2U'^4 e^{-U}$$

Again, using the fact that $|U''| + |U'| = o_{\pm\infty}(U)$, we can find a positive constant $C > 0$, a parameter $\eta \in (0,1)$ and for K large enough such that $\forall x \geq K$:

$$\begin{aligned} U''^2(x)e^{-U(x)} &\leq CU''(x)e^{-(1-\eta)U(x)} \\ &\leq C(U'(x)e^{-(1-\eta)U(x)})' + C(1-\eta)U'(x)^2 e^{-(1-\eta)U(x)} \\ &\leq C(U'(x)e^{-(1-\eta)U(x)})' + C^2(1-\eta)U'(x)e^{-(1-\eta)^2U(x)}, \end{aligned}$$

which is integrable when $x \longrightarrow +\infty$. A similar argument leads to $U'^4 e^{-U} \leq CU' e^{-(1-\eta)U}$. We can repeat the same argument when $x \longrightarrow -\infty$ with an adaptation of the sign of $U'(x)$. We can conclude that $\phi''^2\phi^{-1} \in \mathbb{L}^1(\mathbb{R})$. □

APPENDIX B: REFINEMENT OF A CAUCHY-SCHWARZ INEQUALITY

In this section, without loss of generality, we normalize the density ϕ to 1, meaning (with a slight abuse of notation) that:

$$\forall \mu \in \mathbb{R} \qquad \|\phi_\mu\| = 1.$$

In what follows, we assume that ϕ satisfies $(\mathbf{H}_\mathcal{S})$ and $(\mathbf{H}_{\mathbf{Lip}})$. In particular, these conditions imply the "asymptotic decorrelation" of the location model.

PROPOSITION B.1. *Assume that ϕ satisfies $(\mathbf{H}_\mathcal{S})$, then:*

$$\lim_{|a| \longrightarrow +\infty} \langle \phi, \phi_a \rangle = 0.$$

PROOF. The continuity of ϕ implies that ϕ is bounded by a constant K on $\mathbb{R}$ and that:

$$\lim_{|x| \longrightarrow +\infty} \phi(x) = 0,$$

which in turns implies that:

$$\lim_{|a| \longrightarrow +\infty} \langle \phi, \phi_a \rangle = \lim_{|a| \longrightarrow +\infty} \int \phi(x-a)\phi(x)dx = 0,$$

from the Lebesgue dominated convergence theorem. □

B.1. Main inequality. We are interested in the next property, which can be viewed as a refinement of the Cauchy-Schwarz inequality. Its proof relies on somewhat technical second- and third-order expansions that are given in Section B.2 that includes several technical lemmas, and on the following ratio:

$$R(a,b) = \frac{|\langle \phi - \phi_a, \phi_{a+b} - \phi_a \rangle|}{\|\phi - \phi_a\| \, \|\phi_{a+b} - \phi_a\|} := \frac{|N(a,b)|}{D(a,b)}. \tag{B.1}$$

According to Lemma B.1, the function $(a,b) \mapsto R(a,b)$ defines a continuous map as soon as $a \neq 0$ and $b \neq 0$.

PROPOSITION B.2. *If ϕ satisfies $(\mathbf{H}_\mathcal{S})$ and $(\mathbf{H}_{\mathbf{Lip}})$, then a constant $c > 0$ exists such that $\forall (a,b) \in \mathbb{R}^2$:*

$$|\langle \phi - \phi_a, \phi_{a+b} - \phi_a \rangle| \leq \|\phi - \phi_a\| \, \|\phi_{a+b} - \phi_a\| \left(1 - c\,\|\phi - \phi_{a+b}\|^2\right). \tag{B.2}$$

PROOF. The proof relies on a partition of $\mathbb{R}^2$. Around the diagonal $a+b=0$, Lemmas B.2 (far from the origin) and B.3 (near the origin) show that a couple (ϵ, c_ϵ) exists such that:

$$|a+b| \leq \epsilon \Longrightarrow R(a,b) \leq 1 - c_\epsilon \|\phi_{a+b} - \phi\|^2.$$

Now, outside the diagonal, Lemma B.4 shows that a constant $\tilde{c}_\epsilon$ exists such that:

$$|a+b| \geq \epsilon \Longrightarrow R(a,b) \leq 1 - \tilde{c}_\epsilon.$$

Since$\|\phi_{a+b} - \phi\|^2 \leq 2$, it also implies that:

$$|a+b| \geq \epsilon \Longrightarrow R(a,b) \leq 1 - \frac{\tilde{c}_\epsilon}{2} \|\phi_{a+b} - \phi\|^2.$$

This concludes the proof. □

B.2. Technical lemmas.

B.2.1. *Properties of the location model $(\phi_a)_{a\in\mathbb{R}}$.* In the following text, we will have to compute several Taylor's expansions that involve $(\phi_a)_{a\in\mathbb{R}}$ and its successive derivatives.

PROPOSITION B.3. *If the density ϕ satisfies $(\mathbf{H_{Lip}})$ and $(\mathbf{H}_\mathcal{S})$, then:*

(i) $\langle \phi, \phi' \rangle = 0$
(ii) $\langle \phi', \phi'' \rangle = 0$.
(iii) *For any $a \in \mathbb{R}^\star$, ϕ' and $\phi - \phi_a$ are not proportional.*

PROOF. Item *(i)* If ϕ is $\mathcal{C}^2$, then the conclusion is immediate. Otherwise, ϕ is piecewise $\mathcal{C}^3$, and we can find a finite set of points $a_{-p} = -\infty < a_{-(p-1)} < \ldots < a_{-1} \leq a_0 = 0 \leq a_1 < \ldots a_p = +\infty$, with $a_{-j} = -a_j$, such that ϕ is $\mathcal{C}^1$ on each segment $[a_j, a_{j+1}]$. Integrating by part on each segment, we have:

$$\langle \phi, \phi' \rangle = \sum_{j=-p}^{p-1} \int_{a_j}^{a_{j+1}} \phi'(x)\phi(x)dx = \frac{1}{2} \sum_{j=-p}^{p-1} [\phi^2(x)]_{a_j}^{a_{j+1}}.$$

Since ϕ is continuous and $\lim_{\pm\infty} \phi = 0$, we deduce that $\langle \phi, \phi' \rangle = 0$.
Item *(ii)* We use the same argument and obtain that:

$$\langle \phi', \phi'' \rangle = \frac{1}{2} \sum_{j=-p}^{-1} [\phi'^2(x)]_{a_j}^{a_{j+1}} + \frac{1}{2} \sum_{j=0}^{p-1} [\phi'^2(x)]_{a_j}^{a_{j+1}}.$$

The function ϕ being symmetric, we have:

$$\sum_{j=-p}^{-1} [\phi'^2(x)]_{a_j}^{a_{j+1}} = -\sum_{j=0}^{p-1} [\phi'^2(x)]_{a_j}^{a_{j+1}},$$

so that $\langle \phi', \phi'' \rangle = 0$.
Item (iii) We assume that:

$$\exists \lambda \in \mathbb{R} \quad \forall x \in \mathbb{R} \qquad \phi'(x) = \lambda[\phi(x) - \phi(x-a)] \tag{B.3}$$

If $\lambda \neq 0$, it implies that ϕ' is continuous everywhere (since ϕ_b and ϕ are continuous). Considering $x^* \in \arg\max \phi$, we use (B.3) to obtain:

$$\phi'(x^*) = 0 \Longleftrightarrow \phi(x^*) = \phi(x^* + a).$$

In particular, we cannot have $\lim_{x \longrightarrow +\infty} \phi(x) = 0$, and $\phi \notin \mathbb{L}^2(\mathbb{R})$. We deduce that, necessarily, $\lambda = 0$ and $a = 0$, which is a contradiction. □

B.2.2. *Properties of the ratio R.*

Lemma B.1. *The function R defined in (B.1) is a continuous function on $\mathbb{R}^2$ with:*

$$\forall u \in \mathbb{R}^\star \qquad R(0,u) = R(u,0) = \frac{|\langle \phi', \phi_u - \phi \rangle|}{\|\phi'\| \|\phi - \phi_u\|}.$$

Moreover, R is bounded from above by 1 and:

$$R(a,b) = 1 \Longleftrightarrow a + b = 0.$$

Proof. We first study the continuity of R and consider two cases.

• When $b \neq 0$ is fixed and $a \longrightarrow 0$, the assumption $(\mathbf{H_{Lip}})$ implies that $|\phi(x-a) - \phi(x)| \leq |a| g(x)$ with $g \in \mathbb{L}^2(\mathbb{R})$. We can apply the Lebesgue Theorem and obtain, when $a \longrightarrow 0$,

$$\begin{aligned} N(a,b) &= \int [\phi(x) - \phi_a(x)][\phi_{a+b}(x) - \phi_a(x)] dx \\ &\sim a \int \phi'(x)[\phi_b(x) - \phi(x)] dx \quad \text{when } a \to 0. \end{aligned}$$

A similar computation shows that, when $a \to 0$,

$$D(a,b) \sim a \sqrt{\int \phi'(x)^2 dx} \sqrt{\int [\phi(x) - \phi_b(x)]^2 dx}.$$

Hence, R has a limit when $a \longrightarrow 0$ and $b \neq 0$ is fixed. For the sake of convenience, we keep the notation $R(0,b)$ to refer to this limit and the Cauchy-Schwarz inequality shows that:

$$R(0,b) := \lim_{a \longrightarrow 0} R(a,b) = \frac{|\langle \phi', \phi_b - \phi \rangle|}{\|\phi'\| \|\phi_b - \phi\|} \leq 1.$$

For symmetry reasons in a and b, the same results hold for $a \longmapsto R(a,0)$.

- The situation is easier on $(0,0)$: the Lebesgue Theorem yields:

$$|N(a,b)| \underset{(0,0)}{\sim} |ab| \int \phi'(x)^2 dx \quad \text{and} \quad D(a,b) \underset{(0,0)}{\sim} |ab| \int \phi'(x)^2 dx,$$

meaning that R can also be extended to a continuous map in the neighborhood of $(0,0)$ with $R(0,0) = 1$.

As pointed out above, the Cauchy-Schwarz inequality implies that $\|R\|_\infty \leq 1$. Assume now that $R(a,b) = 1$, if $a+b \neq 0$ and $a \neq 0$. Then the equality in the Cauchy-Schwarz inequality implies that, necessarily, $\phi - \phi_a$ and $\phi_{a+b} - \phi_a$ are proportionnal, which is impossible from the identifiability of the model. Assume now that $a+b \neq 0$ and $a = 0$. The equality $R(0,b)$ is possible if and only if ϕ' is proportional to $\phi_b - \phi$, and (iii) of Proposition B.3 shows that in that case, $b = 0$, which is a contradiction since $a = 0$ and $a+b \neq 0$. To sum up, R can be extended to a continuous map on $\mathbb{R}^2$, which is stricly lower than 1 outside the diagonal $a+b=0$. □

The next lemma concerns the behavior of R around the diagonal $a+b=0$ when a or b are not close to 0.

Lemma B.2. *For any $\eta > 0$, we can find $\epsilon > 0$ such that:*

$$\forall |a| \geq \eta \quad \forall |h| \leq \epsilon \qquad R(a, -a+h) \leq 1 - c_\eta \|\phi_h - \phi\|^2.$$

Proof. To establish the desired inequality, remark that:

$$\begin{aligned}
& R(a,-a+h) \leq 1 - c\|\phi_h - \phi\|^2 \\
& \iff \quad N(a,-a+h) \leq D(a,-a+h) - c\|\phi_h - \phi\|^2 D(a,-a+h) \\
\text{(B.4)} \quad & \iff \quad D(a,-a+h) - N(a,-a+h) > c\|\phi_h - \phi\|^2 D(a,-a+h).
\end{aligned}$$

We use a Taylor expansion when $h = o(1)$ and compute:

$$\begin{aligned}
N(a,b) &= N(a,-a+h) = \langle \phi - \phi_a, \phi_h - \phi_a \rangle \\
&= \|\phi - \phi_a\|^2 - \langle \phi', \phi - \phi_a \rangle h + \frac{\langle \phi'', \phi - \phi_a \rangle}{2} h^2 + o(h^2).
\end{aligned}$$

In the meantime, we have:

$$
\begin{aligned}
&D(a,b) = D(a,-a+h)\\
&= \|\phi-\phi_a\|\sqrt{\|\phi-\phi_a\|^2 - 2h\langle\phi',\phi-\phi_a\rangle + h^2\|\phi'\|^2 + h^2\langle\phi'',\phi-\phi_a\rangle + o(h^2)}\\
&= \|\phi-\phi_a\|^2\sqrt{1-\frac{2\langle\phi',\phi-\phi_a\rangle}{\|\phi-\phi_a\|^2}h + \frac{\|\phi'\|^2+\langle\phi'',\phi-\phi_a\rangle}{\|\phi-\phi_a\|^2}h^2 + o(h^2)}\\
&= \|\phi-\phi_a\|^2 - \langle\phi',\phi-\phi_a\rangle h\\
&\quad + \left(\frac{\|\phi'\|^2}{2} + \frac{\langle\phi'',\phi-\phi_a\rangle}{2} - \frac{\langle\phi',\phi-\phi_a\rangle^2}{2\|\phi-\phi_a\|^2}\right)h^2 + o(h^2).
\end{aligned}
$$

Consequently, we obtain:

$$
\begin{aligned}
&D(a,-a+h) - N(a,-a+h)\\
&= \frac{h^2}{2\|\phi-\phi_a\|^2}\underbrace{\left[\|\phi'\|^2\|\phi-\phi_a\|^2 - \langle\phi',\phi-\phi_a\rangle^2\right]}_{:=\psi(a)} + o(h^2).
\end{aligned}
\tag{B.5}
$$

Note that ψ is a continuous function of a from the Lebesgue Theorem and as pointed out above, it is impossible for ϕ' to be proportional to $\phi-\phi_a$, which in turn implies with the Cauchy-Schwarz inequality that

$$\forall a \neq 0 \qquad \psi(a) > 0.$$

We can study the limit of ψ when $|a| \longrightarrow +\infty$. From (i) of Proposition B.3:

$$|\langle\phi',\phi-\phi_a\rangle| \leq |\langle\phi',\phi\rangle| + |\langle\phi',\phi_a\rangle| \leq \|\phi'\|\|\phi_a\| = \|\phi'\|.$$

Moreover,

$$\lim_{|a|\longrightarrow+\infty} \|\phi'\|^2\|\phi-\phi_a\|^2 = 2\|\phi'\|^2.$$

Hence, we can find A large enough such that:

$$|a| \geq A \Longrightarrow \psi(a) \geq \frac{\|\phi'\|^2}{2} > 0,$$

and we conclude that:

$$\min_{|a|\geq\eta} \psi(a) = m_\eta > 0. \tag{B.6}$$

Equations (B.5) and (B.6) show that an ϵ exists such that:

$$\forall |h| \leq \epsilon \qquad D(a,-a+h) - N(a,-a+h) \geq \frac{m_\eta}{4\|\phi-\phi_a\|^2}h^2.$$

Since $\|\phi-\phi_a\|^2$ and D are upper bounded by 2, we deduce that:

$$\forall |h|\leq \epsilon \qquad (D-N)(a,-a+h)\geq \frac{m_\eta}{8}h^2 \geq \frac{m_\eta}{16}D(a,-a+h)h^2.$$

This inequality associated with $\|\phi_h-\phi\|^2 = h^2\|\phi'\|^2 + o(h^2)$ leads to the desired inequality (B.4) with $c = \frac{m_\eta}{32\|\phi'\|^2}$ for h small enough. □

The next lemma concerns the behavior of R around the origin $(0,0)$.

LEMMA B.3. *Two constants $(\eta, c_\eta)\in\mathbb{R}_+^2$ exist such that:*

$$|a|\vee|b|\leq \eta \Longrightarrow R(a,b)\leq 1-c_\eta\|\phi_{a+b}-\phi\|^2$$

PROOF. • We first study R around the origin when $a\neq 0$ and $b\neq 0$. In that case, we can use $\langle\phi',\phi^{(3)}\rangle = -\|\phi''\|^2$ and (ii) of Proposition B.3 to obtain:

$$\begin{aligned}
|N(a,b)| &= \left|\langle a\phi' - \frac{a^2}{2}\phi'' + \frac{a^3}{6}\phi^{(3)}, -b\phi' + \frac{(a+b)^2-a^2}{2}\phi'' - \frac{(a+b)^3-a^3}{6}\phi^{(3)}\rangle\right| \\
&= \left|-ab\|\phi'\|^2 + \left(\frac{a^2b^2}{4}+\frac{ab^3}{6}+\frac{a^3b}{6}\right)\|\phi''\|^2 + o(a^2)+o(b^2)\right| \\
&= |ab|\left[\|\phi'\|^2 - \|\phi''\|^2\left(\frac{a^2+b^2}{6}+\frac{ab}{4}\right) + o(a^2)+o(b^2)\right]
\end{aligned}$$

Since $D(a,b) = \|\phi-\phi_a\|\|\phi-\phi_b\|$, we compute in a first step:

$$\begin{aligned}
\|\phi-\phi_a\| &= \left\|a\phi' - \frac{a^2}{2}\phi'' + \frac{a^3}{6}\phi^{(3)} + o(a^2)\right\| \\
&= \left[a^2\|\phi'\|^2 + \frac{a^4}{4}\|\phi''\|^2 + \frac{a^4}{3}\langle\phi',\phi^{(3)}\rangle + o(a^2)\right]^{1/2} \\
&= \left[a^2\|\phi'\|^2 + \frac{a^4}{4}\|\phi''\|^2 - \frac{a^4}{3}\langle\phi'',\phi''\rangle + o(a^2)\right]^{1/2} \\
&= \left[a^2\|\phi'\|^2 - \frac{a^4}{12}\|\phi''\|^2 + o(a^2)\right]^{1/2}
\end{aligned}$$

In a second step, we obtain the expansion of D as:

$$\begin{aligned}
D(a,b) &= \|\phi-\phi_a\|\|\phi-\phi_b\| \\
&= \left[a^2\|\phi'\|^2 - \frac{a^4}{12}\|\phi''\|^2 + o(a^2)\right]^{1/2}\left[b^2\|\phi'\|^2 - \frac{b^4}{12}\|\phi''\|^2 + o(b^2)\right]^{1/2} \\
&= |ab|\|\phi'\|^2 - \|\phi''\|^2|ab|\frac{a^2+b^2}{24} + o(a^2+b^2)
\end{aligned}$$

Hence, we see that:

$$\begin{aligned} D(a,b) - |N(a,b)| &\geq |ab|\|\phi''\|^2 \left[-\frac{a^2+b^2}{24} + \frac{a^2+b^2}{6} + \frac{ab}{4} \right] + o(a^2+b^2) \\ &= \frac{3\|\phi''\|^2}{24} |ab|(a+b)^2 + o(a^2+b^2). \end{aligned}$$

Using the argument in Equation (B.4) again, we can check that:

$$c\|\phi_{a+b} - \phi\|^2 D(a,b) \sim c \underbrace{\|\phi'\|^2 (a+b)^2}_{\|\phi_{a+b}-\phi\|^2} \times \underbrace{|ab|\|\phi'\|^2}_{D(a,b)} = c|ab|(a+b)^2\|\phi'\|^4,$$

which means that if $c < \frac{3\|\phi''\|^2}{24\|\phi'\|^4}$, then (B.4) holds for small enough a and b.
• We now study the situation when $a = 0$, that involves the function $R(0,.)$ defined in Lemma B.1. We have:

$$\begin{aligned} \left| \langle \phi', \phi_b - \phi \rangle \right| &= \left| \int \phi' \left[-b\phi' + \frac{b^2}{2}\phi'' - \frac{b^3}{6}\phi^{(3)} \right] + o(b^3) \right| \\ &= |b|\|\phi'\|^2 - \frac{|b|^3}{6}\|\phi''\|^2 + o(b^3), \end{aligned}$$

where we applied (ii) of Proposition B.3 and an integration by parts $\int \phi' \phi^{(3)} = -\|\phi''\|^2$. At the same time, we have:

$$\begin{aligned} \|\phi'\|\|\phi - \phi_b\| &= \|\phi'\| \left(b^2\|\phi'\|^2 - \frac{b^4}{12}\|\phi''\|^2 + o(b^2) \right)^{1/2} \\ &= |b|\|\phi'\|^2 - \frac{|b|^3}{24}\|\phi''\|^2 + o(b^3). \end{aligned}$$

Using the same argument, we obtain:

$$D(0,b) - |N(0,b)| = \frac{|b|^3}{8}\|\phi''\|^2,$$

although for any constant c:

$$c\, D(0,b)\|\phi_b - \phi\|^2 \sim c|b|\|\phi'\|^2 \times b^2\|\phi'\|^2 = c|b|^3\|\phi'\|^4.$$

Again, if $c < \frac{\|\phi''\|^2}{8\|\phi'\|^4}$, then (B.4) holds for small enough b, which ends the proof of the Lemma. □

The remaining lemma studies the behavior of R outside the diagonal.

LEMMA B.4. *For any $\epsilon > 0$, a constant c_ϵ exists such that:*

$$|a+b| \geq \epsilon \Longrightarrow R(a,b) \leq 1 - c_\epsilon.$$

PROOF. Consider the function $\varphi : h \longmapsto |\langle \phi, \phi_h \rangle| = \langle \phi, \phi_h \rangle$, the last equality resulting from the positivity of ϕ and ϕ_h. The dominated convergence theorem shows that φ is continuous and the Cauchy-Schwarz inequality implies that φ is a bounded function whose values belong to $[0,1]$. From the identifiability result of Proposition 2.1, we then have:

$$\varphi(h) = 1 \Longleftrightarrow h = 0.$$

Finally, Proposition B.1 implies that $\lim_{|h| \longmapsto +\infty} \varphi(h) = 0$. Taken together, these elements show that for any $\epsilon > 0$, φ attains its upper bound on $B(0,\epsilon)^c$. It yields:

$$\forall \epsilon > 0 \quad \exists \eta_\epsilon > 0 \qquad \sup_{|h| \geq \epsilon} \varphi(h) \leq 1 - \eta_\epsilon. \tag{B.7}$$

• We first consider the case where $|a| \wedge |b| \longrightarrow +\infty$ with $\epsilon \leq |a+b|$. In that case, if we denote $h = a+b$ and use $\lim_{|a| \longrightarrow +\infty} \langle \phi, \phi_a \rangle = 0$, then we can find M_ϵ large enough such that:

$$|a| \wedge |b| \geq M_\epsilon \qquad \Longrightarrow$$
$$\frac{|N(a,b)|}{D(a,b)} = \frac{|1 + \langle \phi, \phi_h \rangle - \langle \phi, \phi_a \rangle - \langle \phi, \phi_b \rangle|}{\|\phi - \phi_a\| \|\phi - \phi_b\|} \leq \frac{1 + \sup_{\epsilon \leq |h|} \varphi(h)}{2} \times \frac{1 - \frac{\eta_\epsilon}{3}}{1 - \frac{\eta_\epsilon}{2}} \leq 1 - \frac{\eta_\epsilon}{3},$$

where η_ϵ is defined in (B.7).

• We now consider the case where $|a| \longrightarrow +\infty$ although $|b|$ remains bounded by M_ϵ, so that $b \in B(0, M_\epsilon) \setminus \{0\}$. In that case, we compute:

$$N(a,b) = \left| \langle \phi, \phi_{a+b} \rangle - \langle \phi, \phi_a \rangle - \langle \phi, \phi_b \rangle + \|\phi_a\|^2 \right| \longrightarrow 1 - \langle \phi, \phi_b \rangle \quad \text{if} \quad |a| \longrightarrow +\infty.$$

At the same time, we also consider D and check that:

$$D(a,b) = \|\phi - \phi_a\| \|\phi_{a+b} - \phi_a\| \longrightarrow 2\sqrt{1 - \langle \phi, \phi_b \rangle} \quad \text{when} \quad |a| \longrightarrow +\infty.$$

We then obtain:

$$\lim_{|a| \longrightarrow +\infty} R(a,b) = \frac{\sqrt{1 - \langle \phi, \phi_b \rangle}}{2} \leq \frac{1}{2}.$$

Hence, we can find a constant A_ϵ sufficiently large such that:

$$\forall |a| \geq A_\epsilon \quad \forall b \in B(0, M_\epsilon) \quad R(a,b) \leq \frac{3}{4}.$$

• If a and b now belong to the compact set:

$$\mathcal{E}_\epsilon := \left\{(a,b) \in \mathbb{R}^2 \,:\, |a| \leq A_\epsilon,\, |b| \leq M_\epsilon,\, |a+b| \geq \epsilon\right\},$$

we know that R is a continuous function on $\mathcal{E}_{\epsilon,A,M}$ and attains its upper bound, which is strictly lower than 1 by the Cauchy-Schwarz inequality. Consequently,

$$\exists \tilde{\eta}_\epsilon > 0\, \forall (a,b) \in \mathcal{E}_\epsilon \qquad R(a,b) \leq 1 - \tilde{\eta}_\epsilon.$$

Taking all the bounds obtained outside of the diagonal together, we obtain the lemma with $c_\epsilon = (\tilde{\eta}_\epsilon \wedge \eta_\epsilon/3 \wedge 1/4)$. □

ACKNOWLEDGMENTS

This work was partially supported by the French Agence Nationale de la Recherche (ANR- 13-JS01-0001-01, project MixStatSeq).

REFERENCES

[BG14] D. Bontemps and S. Gadat. Bayesian methods for the shape invariant model. *Electron. J. Statist.*, 8(1):1522–1568, 2014.

[BMV06] L. Bordes, S. Mottelet, and P. Vandekerkhove. Semiparametric estimation of a two-component mixture model. *Ann. Statist.*, 34(3):1204–1232, 2006.

[BTWB10] F. Bunea, A. B. Tsybakov, M. H. Wegkamp, and A. Barbu. Spades and mixture models. *Ann. Statist.*, 38(4):2525–2558, 2010.

[BV14] C. Butucea and P. Vandekerkhove. Semiparametric mixtures of symmetric distributions. *Scand. J. Stat.*, 41(1):227–239, 2014.

[BWY16] S. Balakrishnan, M. Wainwright, and B. Yu. Statistical guarantees for the EM algorithm: From population to sample-based analysis. *The Annals of Statistics, to appear*, 2016.

[Che95] J. H. Chen. Optimal rate of convergence for finite mixture models. *Ann. Statist.*, 23(1):221–233, 1995.

[CJJ11] T. T. Cai, X. J. Jeng, and J. Jin. Optimal detection of heterogeneous and heteroscedastic mixtures. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 73(5):629–662, 2011.

[CJL07] T. T. Cai, J. Jin, and M. G. Low. Estimation and confidence sets for sparse normal mixtures. *Ann. Statist.*, 35(6):2421–2449, 2007.

[DLR77] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B*, 39(1):1–38, 1977. With discussion.

[FS06] S. Frühwirth-Schnatter. *Finite mixture and Markov switching models.* Springer Series in Statistics. Springer, New York, 2006.

[GvdV01] S. Ghosal and A. W. van der Vaart. Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Ann. Statist.*, 29(5):1233–1263, 2001.

[GW00] C. R. Genovese and L. Wasserman. Rates of convergence for the Gaussian mixture sieve. *Ann. Statist.*, 28(4):1105–1127, 2000.

[HK15] P. Heinrich and J. Kahn. Optimal rates for finite mixture estimation. *Preprint*, 2015.

[HN16a] N. Ho and X. Nguyen. Convergence rates of parameter estimation for some weakly identifiable finite mixtures. *Annals of statistics, to appear*, 2016.

[HN16b] N. Ho and X. Nguyen. On strong identifiability and convergence rates of parameter estimation in finite mixtures. *Electron. J. Statist.*, 10(1):271–307, 2016.

[KRvdV10] W. Kruijer, J. Rousseau, and A. W. van der Vaart. Adaptive Bayesian density estimation with location-scale mixtures. *Electron. J. Stat.*, 4:1225–1257, 2010.

[LCY00] L. Le Cam and G. Yang. *Asymptotics in Statistics: Some Basic Concepts*. Springer series in statistics. Springer Verlag, New-York, 2000.

[Lin09] T. Lin. Maximum likelihood estimation for multivariate skew normal mixture models. *Journal of Multivariate Analysis*, 100:257–265, 2009.

[LMMR16] B. Laurent, C. Marteau, and C. Maugis-Rabusseau. Non asymptotic detection of two component mixtures with unknown means. *Bernoulli*, 22:242–274, 2016.

[MM11] C. Maugis and B. Michel. A non asymptotic penalized criterion for Gaussian mixture model selection. *ESAIM Probab. Stat.*, 15:41–68, 2011.

[MP00] G. McLachlan and D. Peel. *Finite Mixture Models*. Wiley series in Probability and Statistics, 2000.

[MRM13] C. Maugis-Rabusseau and B. Michel. Adaptive density estimation for clustering with Gaussian mixtures. *ESAIM Probab. Stat.*, 17:698–724, 2013.

[Ste81] C. Stein. Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, 9:1135–1151, 1981.

[Wu83] C. F. J. Wu. On the convergence properties of the EM algorithm. *Ann. Statist.*, 11:95–103, 1983.

[Yu97] *Festschrift for Lucien Le Cam*, chapter Assouad, Fano, and Le Cam. Springer Verlag, 1997.

Toulouse School of Economics
Université Toulouse 1 - Capitole.
21 allées de Brienne
31000 Toulouse, France.
E-mail: sebastien.gadat@math.univ-toulouse.fr

Institut Camille Jordan
Université Lyon 1 - Claude Bernard
43 boulevard du 11 novembre 1918
69622 Villeurbanne, France.
E-mail: clement.marteau@math.univ-lyon1.fr

Institut Mathématiques de Toulouse
INSA de Toulouse,
135 avenue de Rangueil,
31077 Toulouse, France.
E-mail: cathy.maugis@insa-toulouse.fr