Electronic Journal of Statistics
Vol. 11 (2017) 4563–4595
ISSN: 1935-7524
DOI: 10.1214/17-EJS1311

Revisiting the Hodges-Lehmann estimator in a location mixture model: Is asymptotic normality good enough?

Fadoua Balabdaoui[1,2]

[1] *Université Paris-Dauphine, PSL Research University, CNRS, CEREMADE, 75016 Paris, France*

[2] *Seminar für Statistik, ETH Zürich, 8092, Zürich, Schweiz*
e-mail: fadoua.balabdaoui@stat.math.ethz.ch

Abstract: Location mixture models, resulting in shifting a common distribution with some probability, have been widely used to account for existence of clusters in the data. Assuming only symmetry of this common distribution allows for great flexibility, especially when the traditional normality assumption is violated. This semi-parametric model has been studied in several papers, where the mixture parameters are first estimated before constructing an estimator for the non-parametric component. The plug-in method suggested by Hunter et al. (2007) has the merit to be easily implementable and fast to compute. However, no result is available on the limit distribution of the obtained estimator, hindering for instance construction of asymptotic confidence intervals. In this paper, we give sufficient conditions on the symmetric distribution for asymptotic normality to hold. In case the symmetric distribution admits a log-concave density, our assumptions are automatically satisfied. The obtained result has to be used with caution in case the mixture location are too close or the mixing probability is close to 0 or 1. Three examples are considered where we show that the estimator is not to be advocated when the mixture components are not well separated.

Received December 2016.

1. Introduction

1.1. A brief overview

Consider the two-component location mixture model

$$F(x) = \pi\ G(x-\mu_1) + (1-\pi)\ G(x-\mu_2) \tag{1.1}$$

where $\pi \in [0,1]$, $-\infty < \mu_1 \leq \mu_2 < \infty$ and G is a symmetric distribution around zero, that is, $G(-x) = 1 - G(x_-)$ for $x \in \mathbb{R}$. This model is semi-parametric since the unknown parameters are the 3-dimensional vector (π, μ_1, μ_2) and the symmetric distribution G. It has been considered by several authors, e.g. Bordes et al. (2006), Hunter et al. (2007), Chee and Wang (2013), Butucea and Vandekerkhove (2014), and more recently Balabdaoui and Doss (2016). Whether

the goal is to estimate the mixed distribution or to classify new members in each of the existing clusters, (1.1) offers more flexibility than the Gaussian mixture model when G is not the distribution function of a normal variable. The simulation study carried out by Hunter et al. (2007) shows evidence that the estimators obtained under (1.1) outperforms the maximum likelihood estimator (MLE) under the Gaussian mixture model for the heavy-tailed distributions they considered. When testing for presence of mixing, the numerical results obtained in Balabdaoui and Doss (2016) for the asymptotic power also show the higher performance of the symmetric log-concave MLE when compared to the Gaussian one. In Chang and Walther (2007), the authors considered a more general mixture model where the mixture components are assumed to have log-concave densities. See also the clustering example of Cule et al. (2010) in a two-dimensional setting using the Wisconsin Breast cancer study. The aforementioned log-concave mixture models are of course more general than the location mixture model in (1.1) but suffer from being non-identifiable. See e.g. the counterexamples given in the Concluding Discussion section in Cule et al. (2010).

To estimate the distribution G, the focus in all the aforementioned papers on the mixture model in (1.1) is on first estimating the parameters of the mixture μ_1, μ_2 and π. In some way, these mixture parameters are viewed as nuisance parameters. In Butucea and Vandekerkhove (2014), an estimator of these parameters was constructed by converting symmetry of the density of G (assumed to exist) to the fact that the imaginary part of its Fourier transform has to be equal to 0. The authors showed that their estimator is consistent and established that it is asymptotically normal, under the assumption that the mixture model is identifiable. In fact, they used the identifiability conditions found by Hunter et al. (2007), which we now discuss. Hunter et al. (2007) considered the more flexible notion of 2-identifiability defined as follows. Consider the sets

$$\Theta = \{\theta = (\pi, \mu_1, \mu_2) : \pi \in [0,1], -\infty < \mu_1 < \mu_2 < \infty\}, \tag{1.2}$$
$$\mathcal{S} = \Big\{G : G \text{ is a symmetric c.d.f.}\Big\},$$

and

$$\mathcal{M} = \Big\{F : F = \pi\ G(x-\mu_1) + (1-\pi)\ G(x-\mu_2), (\pi, \mu_1, \mu_2, G) \in \Theta \times \mathcal{S}\Big\}.$$

Let φ be the application defined on $\Theta \times \mathcal{S}$ onto $\mathcal{M}$ such that $F = \varphi(G, \theta)$ is given by (1.1). An element $F \in \mathcal{M}$ is said to be 2-identifiable if $\varphi^{-1}(F)$ is a singleton in $\Theta \times \mathcal{S}$. See also Definition 1 in Hunter et al. (2007). The first goal of the authors was to determine the largest subset in Θ yielding 2-identifiable distribution functions $F \in \mathcal{M}$. They showed that this set is precisely

$$\Theta^* = \{\theta = (\pi, \mu_1, \mu_2) \in \Theta :\ \pi \in (0,1) \setminus \{1/2\}\}. \tag{1.3}$$

In other words, the only distributions which are not 2-identifiable are those which are themselves symmetric; see their Theorem 2. This result is proved by first showing that a necessary and sufficient condition for the vector parameter to yield a 2-identifiable mixed distribution given by (1.1) is that

$$\Delta(\pi, -\mu_1, -\mu_2) := \pi\delta_{-\mu_1} + (1-\pi)\delta_{-\mu_2}$$

is the unique distribution that gives a symmetric distribution around zero when convolved with

$$\Delta(\pi, \mu_1, \mu_2) := \pi\delta_{\mu_1} + (1-\pi)\delta_{\mu_2}$$

where δ_a is the Dirac distribution at some real a; see Theorem 1 of Hunter et al. (2007). In the sequel, $\theta^0 = (\pi^0, \mu_1^0, \mu_2^0) \in \Theta^*$ denotes the true mixture parameters, $F^0 = G^0 \star \Delta(\theta^0)$ the true mixed distribution and G^0 the true associated symmetric distribution, where $\star$ is the usual notation for the convolution operator. The previous preliminary result gives the following very useful identity, exploited in this paper at various places when deriving the asymptotics of the estimator of $(\pi^0, \mu_1^0, \mu_2^0)$:

$$\pi^0\Big(1 - F^0(\mu_1^0 - t) - F^0(\mu_1^0 + t)\Big) + (1-\pi^0)\Big(1 - F^0(\mu_2^0 - t) - F^0(\mu_2^0 + t)\Big) = 0 \tag{1.4}$$

for all $t \in \mathbb{R}$. This identity is satisfied only by $\pi = \pi^0, \mu_1 = \mu_1^0, \mu_2 = \mu_2^0$ and hence characterizes the true vector θ^0. To give the reader some background, the identity in (1.4) is an immediate consequence of two facts. The first one is

$$F^0 \star \Delta(\pi, -\mu_1, -\mu_2) = G^0 \star \big(\Delta(\pi^0, \mu_1^0, \mu_2^0) \star \Delta(\pi, -\mu_1, -\mu_2)\big)$$

for any $(\pi, \mu_1, \mu_2) \in \Theta^*$. The second one is that the distribution on the right-hand side is symmetric if and only if $\Delta(\pi^0, \mu_1^0, \mu_2^0) \star \Delta(\pi, -\mu_1, -\mu_2)$ is symmetric. As discussed above, the latter implies that $\pi = \pi^0, \mu_1 = \mu_1^0$ and $\mu_2 = \mu_2^0$ since θ^0 is the only vector in Θ^* satisfying this property. Therefore, we have that $F^0 \star \Delta(\pi^0, -\mu_1^0, -\mu_2^0) = (1 - F^0(-\cdot)) \star \Delta(\pi^0, \mu_1^0, \mu_2^0)$, which is exactly the identity given in (1.4). Then, it follows that

$$\begin{aligned}
\mathbb{D}^2(\theta^0) &= \int_{\mathbb{R}} \Big\{\pi^0\Big(1 - F^0(\mu_1^0 - t) - F^0(\mu_1^0 + t)\Big) \\
&\qquad + (1-\pi^0)\Big(1 - F^0(\mu_2^0 - t) - F^0(\mu_2^0 + t)\Big)\Big\}^2 dt \\
&= 0.
\end{aligned}$$

Note that the latter is nothing but saying that the L_2 distance between $F^0 \star \Delta(\pi^0, -\mu_1^0, -\mu_2^0)$ and $(1 - F^0(-\cdot)) \star \Delta(\pi^0, \mu_1^0, \mu_2^0)$ is equal to 0.

1.2. The plug-in estimator

Replacing θ^0 by an element θ to obtain $\mathbb{D}^2(\theta)$ and taking the empirical counterpart of the resulting expression, one can now construct the plug-in estimator of the mixture parameters through minimizing

$$\mathbb{D}_n^2(\theta) = \int_{\mathbb{R}} \Big\{\pi\Big(1 - \mathbb{F}_n(\mu_1 - t) - \mathbb{F}_n(\mu_1 + t)\Big)$$

$$+(1-\pi)\big(1-\mathbb{F}_n(\mu_2-t)-\mathbb{F}_n(\mu_2+t)\big)\Big\}^2 dt \qquad (1.5)$$

over Θ^*, where $\mathbb{F}_n$ is the empirical distribution. Let $\widehat{\theta}_n$ denote this estimator. To compute the symmetric log-concave MLE of the density of G^0, Balabdaoui and Doss (2016) replace first the vector of the unknown mixture parameters by $\widehat{\theta}_n$ and then maximize the log-likelihood over the class of symmetric and log-concave densities on $\mathbb{R}$. Choosing this estimator was motivated by the fact that it is very easy to implement and also fast to compute. Balabdaoui and Doss (2016) used the R-code of Hunter et al. (2007) made available on the web page of the first author. Moreover, Hunter et al. (2007) show, under some conditions, that the estimator convergences at the $\sqrt{n}$- rate. Assuming that these conditions hold, the fast rate of convergence of $\widehat{\theta}_n$ guarantees convergence of the nonparametric log-concave MLE of Balabdaoui and Doss (2016) to the true symmetric density at the (usual) $n^{2/5}$-rate in the L_1 distance. See Theorem 4.1 in Balabdaoui and Doss (2016).

1.3. Problem, motivation and limitations

in their Theorem 3 Hunter et al. (2007) show almost sure convergence of $\widehat{\theta}_n$ to the truth but no explicit proof on asymptotic normality was provided. Instead, Hunter et al. (2007) give in their Theorem 4 conditions under which $\widehat{\theta}_n$ converges at the $\sqrt{n}$-rate to a 3-dimensional centered Gaussian distribution with a dispersion matrix given by $J^{-1}\Sigma J^{-1}$. The matrix J is the Hessian matrix of $\theta \mapsto \mathbb{D}^2(\theta)$, assuming that it exists and is positive definite. The matrix Σ is the covariance matrix of a 3-dimensional vector playing the role of a gradient of $E[f_\theta(x,Y)]$ at θ^0 where f_θ is given below in (4.9); see also condition (iii) in Theorem 4 of Hunter et al. (2007). The three conditions given by Hunter et al. (2007) for Theorem 4 are connected to the theory of V-processes, since $\mathbb{D}_n^2(\theta)$ can be identified as such.

In a personal communication with David Hunter, the validity of the conditions of Theorem 4 has never been checked. Thus, the rate of convergence of $\widehat{\theta}_n$ and its limit are, up to now, open questions. Without this knowledge, it is not possible to construct asymptotic tests or confidence intervals for any of the mixture parameters. In this paper, we show that $\widehat{\theta}_n$ is indeed asymptotically normal under some sufficient conditions. The approach we have taken is based on the theory of empirical processes, which is well-suited for M-estimators. Although certain results from this theory can be used off the shelf, a substantial effort has been made to re-adapt them in the current setting. The fact is that the estimator $\widehat{\theta}_n$ minimizes a functional that is quadratic in $\mathbb{F}_n$ whereas most of the results in van der Vaart and Wellner (1996) are tailored for empirical processes that are linear in $\mathbb{F}_n$. Thus, we believe that some of the techniques developed in this article are of its interest in their own right, and may be applicable in some other M-estimation problems where the dependence on $\mathbb{F}_n$ is non-linear. Investigation of the weak convergence of $\widehat{\theta}_n$ allowed to have a clearer insight

into when this estimator can or cannot be used. The expression of the variance-covariance matrix obtained in Section 2 can be used to compute Monte-Carlo approximations of the asymptotic variances. The examples of Section 3 clearly indicate that the estimator is not to be advocated in case the mixture locations are two close to each other, or if the mixing probability is close to 0 or 1.

1.4. Organization of the paper

While thinking about the asymptotic behavior of $\widehat{\theta}_n$, we realized that existence of $\widehat{\theta}_n$ is yet to be confirmed. Although the estimator was defined in a very intuitive way, a formal proof of existence has not been provided in Hunter et al. (2007) nor in a separate note (personal communication with David Hunter). A formal proof that $\widehat{\theta}_n$ exists is given in Section 4.1. The structure of the remaining results is as follows: Section 2 is devoted to deriving the asymptotic distribution under some sufficient conditions on the density g^0 of the unknown symmetric distribution G^0. In Section 3 we give some examples to illustrate the theory. As mentioned above, the examples show that the estimator can be very inefficient when the mixture components are not well separated. In Section 4.2 we gather the proofs of the results yielding the $\sqrt{n}$-rate of convergence of $\widehat{\theta}_n$. The proof of the asymptotic distribution of the estimator along with some useful formulae are deferred to Supplement A (Balabdaoui, 2017).

2. Asymptotics of the (Hogdes-Lehmann) estimator

This section is devoted to the main subject of this paper: establishing the asymptotic distribution of the estimator $\widehat{\theta}_n$. The approach we chose to follow might look like a detour from V-processes, which may have been a more natural way to go as the functional to be minimized is a quadratic function of $\mathbb{F}_n$. Our attempts to check the conditions of Theorem 4 of Hunter et al. (2007) were not very successful. In the meantime, we realized that the theory of M-estimators can be applied in the current context. However, some effort was needed to be made to cast the problem into the theory of empirical processes. The proof is divided into four steps so that it is possible to apply the argmax continuous mapping theorem (or rather the argmin continuous mapping theorem here) for processes that converge weakly to a tight limit in $C([-K,K]^3)$, the space of continuous functions defined on the 3-dimensional compact $[-K,K]^3$ for some $K>0$. Our main reference is Theorem 3.2.2 of van der Vaart and Wellner (1996). In the sequel, the symmetric distribution G^0 is assumed to have a density g^0 with respect to Lebesgue measure. In the next section, we will show that $\widehat{\theta}_n$ converges to θ^0 at the parametric rate $\sqrt{n}$.

2.1. Deriving the $\sqrt{n}$-rate of convergence

The first step towards establishing the rate of convergence is showing consistency. The latter follows from Theorem 3 of Hunter et al. (2007).

Theorem 2.1. *Assume that g^0 admits a finite first moment. Then*

$$\widehat{\theta}_n \to \theta^0$$

almost surely as $n \to \infty$.

To be able to refine this result we will appeal for the theory of empirical processes. As we will show below, the centered and rescaled processs

$$\sqrt{n}\Big(\mathbb{D}_n^2(\theta) - \mathbb{D}^2(\theta) - (\mathbb{D}_n^2(\theta_0) - \mathbb{D}^2(\theta_0))\Big)$$

can be decomposed into processes whose maximal expectation over a neighborhood of θ_0 can be controlled using the theory of VC-classes. To pave the way to showing the main result of this section, we shall adopt the notation

$$\theta_t = \Big(\pi, \mu_1 - t, \mu_1 + t, \mu_2 - t, \mu_2 + t\Big)$$

for a given $t \in \mathbb{R}$. Also, we shall consider the collection of functions m_{θ_t} defined as

$$m_{\theta_t}(x) = \pi\big(\mathbb{1}_{x>\mu_1-t} - \mathbb{1}_{x\le\mu_1+t}\big) + (1-\pi)\big(\mathbb{1}_{x>\mu_2-t} - \mathbb{1}_{x\le\mu_2+t}\big)$$

for $\theta = (\pi, \mu_1, \mu_2) \in \Theta$. The symbols $\mathbb{P}_n$ and $\mathbb{P}$ will be used for the empirical and true measure respectively. The starting point here is to rewrite $\mathbb{D}_n^2(\theta)$ as

$$\begin{aligned}
\mathbb{D}_n^2(\theta) &= \int_{\mathbb{R}} (\mathbb{P}_n m_{\theta_t})^2\, dt \\
&= \int_{\mathbb{R}} \big(\mathbb{P}_n(m_{\theta_t} - m_{\theta_t^0}) + \mathbb{P}_n m_{\theta_t^0}\big)^2 dt \\
&= \int_{\mathbb{R}} \big(\mathbb{P}_n(m_{\theta_t} - m_{\theta_t^0})\big)^2 dt + 2\int_{\mathbb{R}} \mathbb{P}_n(m_{\theta_t} - m_{\theta_t^0})\ \mathbb{P}_n m_{\theta_t^0}\ dt \\
&\quad + \int_{\mathbb{R}} \Big(\mathbb{P}_n m_{\theta_t^0}\Big)^2\, dt
\end{aligned}$$

where the last term on the left side is equal to $\mathbb{D}_n(\theta^0)$. Developing $\mathbb{D}(\theta)$ in a similar way and using the fact that for all $t \in \mathbb{R}$

$$\begin{aligned}
\mathbb{P} m_{\theta_t^0} &= \pi^0\big(1 - F^0(\mu_1^0 - t) - F^0(\mu_1^0 + t)\big) \\
&\quad + (1-\pi^0)\big(1 - F^0(\mu_2^0 - t) - F^0(\mu_2^0 + t)\big) = 0
\end{aligned}$$

using (1.4) yields $\mathbb{D}^2(\theta) = \int_{\mathbb{R}} \big(\mathbb{P}(m_{\theta_t} - m_{\theta_t^0})\big)^2 dt$. Hence,

$$\begin{aligned}
&\mathbb{D}_n^2(\theta) - \mathbb{D}^2(\theta) - (\mathbb{D}_n^2(\theta^0) - \mathbb{D}^2(\theta^0)) \\
&\quad = \int_{\mathbb{R}} \Big[\big(\mathbb{P}_n(m_{\theta_t} - m_{\theta_t^0})\big)^2 - \big(\mathbb{P}(m_{\theta_t} - m_{\theta_t^0})\big)^2\Big] dt \\
&\qquad + 2\int_{\mathbb{R}} \mathbb{P}_n(m_{\theta_t} - m_{\theta_t^0})\ \mathbb{P}_n m_{\theta_t^0}\ dt
\end{aligned}$$

$$
\begin{aligned}
&= \int_{\mathbb{R}} \left[\left(\mathbb{P}_n(m_{\theta_t} - m_{\theta_t^0})\right)^2 - \left(\mathbb{P}(m_{\theta_t} - m_{\theta_t^0})\right)^2 \right] dt \\
&\quad + 2 \int_{\mathbb{R}} (\mathbb{P}_n - \mathbb{P})(m_{\theta_t} - m_{\theta_t^0}) \; (\mathbb{P}_n - \mathbb{P}) m_{\theta_t^0} \; dt \\
&\quad + 2 \int_{\mathbb{R}} \mathbb{P}(m_{\theta_t} - m_{\theta_t^0}) \; (\mathbb{P}_n - \mathbb{P}) m_{\theta_t^0} \; dt,
\end{aligned}
$$

using again (1.4)

$$
\begin{aligned}
&= \int_{\mathbb{R}} (\mathbb{P}_n - \mathbb{P})(m_{\theta_t} - m_{\theta_t^0}) \; (\mathbb{P}_n + \mathbb{P})(m_{\theta_t} - m_{\theta_t^0}) \; dt \\
&\quad + 2 \int_{\mathbb{R}} (\mathbb{P}_n - \mathbb{P})(m_{\theta_t} - m_{\theta_t^0}) \; (\mathbb{P}_n - \mathbb{P}) m_{\theta_t^0} \; dt \\
&\quad + 2 \int_{\mathbb{R}} \mathbb{P}(m_{\theta_t} - m_{\theta_t^0}) \; (\mathbb{P}_n - \mathbb{P}) m_{\theta_t^0} \; dt.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&\sqrt{n} \left| \mathbb{D}_n^2(\theta) - \mathbb{D}^2(\theta) - (\mathbb{D}_n^2(\theta^0) - \mathbb{D}^2(\theta^0)) \right| \\
&\quad \leq \sqrt{n} \int_{\mathbb{R}} (\mathbb{P}_n - \mathbb{P})^2 (m_{\theta_t} - m_{\theta_t^0}) dt \\
&\qquad + 2 \int_{\mathbb{R}} \sqrt{n} \left| (\mathbb{P}_n - \mathbb{P})(m_{\theta_t} - m_{\theta_t^0}) \right| \; \mathbb{P} \left| m_{\theta_t} - m_{\theta_t^0} \right| \; dt \\
&\qquad + 2 \int_{\mathbb{R}} \sqrt{n} \left| (\mathbb{P}_n - \mathbb{P})(m_{\theta_t} - m_{\theta_t^0}) \right| \; \left| (\mathbb{P}_n - \mathbb{P}) m_{\theta_t^0} \right| \; dt \\
&\qquad + 2 \int_{\mathbb{R}} \mathbb{P} \left| m_{\theta_t} - m_{\theta_t^0} \right| \; \sqrt{n} \left| (\mathbb{P}_n - \mathbb{P}) m_{\theta_t^0} \right| \; dt. \qquad (2.6)
\end{aligned}
$$

Let

$$
\|\theta_t - \theta_t^0\|_\infty = \|\theta - \theta^0\|_\infty = \max(|\pi - \pi^0|, |\mu_1 - \mu_1^0|, |\mu_2 - \mu_2^0|).
$$

Recall that G^0 is assumed to have a density g^0 with respect to Lebesgue measure. To derive the $\sqrt{n}$-rate of convergence of $\widehat{\theta}_n$, we will show the following intermediate theorem.

Theorem 2.2. *Suppose that $\int_{\mathbb{R}} (g^0(t))^{1/2} dt < \infty$. Then, there exists a constant $C > 0$ depending only on g^0 and $\mu_2^0 - \mu_1^0$ such that for $\delta > 0$ small enough we have that*

$$
E \sup_{\|\theta - \theta^0\|_\infty < \delta} \left[\sqrt{n} \left| \mathbb{D}_n^2(\theta) - \mathbb{D}^2(\theta) - (\mathbb{D}_n^2(\theta^0) - \mathbb{D}^2(\theta^0)) \right| \right] \leq C \left(\delta + \frac{\delta^{1/2}}{\sqrt{n}} \right) = \phi_n(\delta).
$$

The proof of the above bound will involve two main steps. In the first one, we shall control the maximal expectation of the process

$$
\mathbb{G}_n(m_{\theta_t} - m_{\theta_t^0}) := \sqrt{n}(\mathbb{P}_n - \mathbb{P})(m_{\theta_t} - m_{\theta_t^0})
$$

over the class

$$\mathcal{M}_{t,\delta} = \left\{ m_{\theta_t} - m_{\theta_t^0} : \|\theta_t - \theta_t^0\|_\infty = \|\theta - \theta^0\|_\infty < \delta \right\}$$

for a fixed $\delta > 0$. We will achieve this by showing that $\mathcal{M}_{t,\delta}$ can be embedded in the sum of VC-subgraphs. This will enable us to show that we have control on its entropy as if it were itself a VC-subgraph. The bound we will obtain for the supremum of the expectation of the process $\mathbb{G}_n(m_{\theta_t} - m_{\theta_t^0})$ will depend on $t \in \mathbb{R}$, the variable of integration. In the second step, and under some suitable assumptions on the true symmetric density g^0, we will show that this control is strong enough to allow for integrating the resulting supremum and hence obtain the desired bound given in Theorem 2.2. For a fixed $t \in \mathbb{R}$ let $F_{i,t},\ i = 1, \ldots, 6$, be the functions defined as

$$\begin{aligned}
F_{1,t}(x) &= \pi^0 \mathbb{1}_{x \in [\mu_1^0 - \delta + t, \mu_1^0 + \delta + t]} \\
F_{2,t}(x) &= \pi^0 \mathbb{1}_{x \in [\mu_1^0 - \delta - t, \mu_1^0 + \delta - t]} \\
F_{3,t}(x) &= (1 - \pi^0) \mathbb{1}_{x \in [\mu_2^0 - \delta + t, \mu_1^0 + \delta + t]} \\
F_{4,t}(x) &= (1 - \pi^0) \mathbb{1}_{x \in [\mu_2^0 - \delta - t, \mu_1^0 + \delta - t]} \\
F_{5,t}(x) &= \delta \mathbb{1}_{x \in [\mu_1^0 - \delta + t, \mu_2^0 + \delta + t]} \\
F_{6,t}(x) &= \delta \mathbb{1}_{x \in [\mu_1^0 - \delta - t, \mu_2^0 + \delta - t]}.
\end{aligned} \tag{2.7}$$

The following theorem gives the necessary ingredients for completing the first step of the proof as described above.

Theorem 2.3. *The following holds true:*

1. *The class of funtions*

$$\mathcal{I} = \left\{ c \mathbb{1}_{(a,b]}, a, b, c \in \mathbb{R} \right\}$$

   is a VC-subgraph of index 4. The same holds true if $(a, b]$ is replaced by $[a, b)$.

2. *Let $\mathcal{F}$ be a class of functions such that $\mathcal{F} = \mathcal{F}_1 + \ldots + \mathcal{F}_m$ for some integer $m > 0$ and $\mathcal{F}_i$ are VC-classes with envelopes F_i for $i = 1, \ldots, m$ respectively. Then, $\mathcal{F}$ is P-Donsker and there exists a constant $M > 0$ such that*

$$\begin{aligned}
E[\|\mathbb{G}_n\|_{\mathcal{F}}] &\leq m\ M \|F_1 + \ldots + F_m\|_{P,2}, \text{ and} \\
E[\|\mathbb{G}_n\|_{\mathcal{F}}^2] &\leq m^2\ M \|F_1 + \ldots + F_m\|_{P,2}^2
\end{aligned}$$

3. *There exists a constant $M > 0$ such that for all $t \in \mathbb{R}$*

$$\begin{aligned}
E[\|\mathbb{G}_n\|_{\mathcal{M}_{t,\delta}}] &\leq 6M \|F_{1,t} + \ldots + F_{6,t}\|_{P,2}, \text{ and} \\
E[\|\mathbb{G}_n\|_{\mathcal{M}_{t,\delta}}^2] &\leq 36M \|F_{1,t} + \ldots + F_{6,t}\|_{P,2}^2.
\end{aligned}$$

   where the functions $F_{i,t}, i = 1, \ldots, 6$ were defined in (2.7).

In the following proposition we give sufficient conditions for the first requirement for deriving the rate of convergence of an M-estimator using Theorem 3.2.5 of van der Vaart and Wellner (1996).

Proposition 2.4. *Assume that g^0 satisfies that*

- $\int_{\mathbb{R}} |t| g^0(t) dt < \infty$,
- *it admits a derivative, $(g^0)'$, almost everywhere that is bounded on $\mathbb{R}$.*

Then, there exists a small neighborhood of $\theta^0 = (\pi^0, \mu_1^0, \mu_2^0)$ and a constant $\kappa > 0$ such that for all $\theta = (\pi, \mu_1, \mu_2)$ in this neighborhood we have that

$$\mathbb{D}^2(\theta) - \mathbb{D}^2(\theta_0) \geq \kappa \, \|\theta - \theta^0\|_\infty.$$

Theorem 2.2 and Proposition 2.4 yield now the rate of convergence of the estimator of Hunter et al. (2007).

Proposition 2.5. *Under the assumptions of Theorem 2.2 and Proposition 2.4 we have that*

$$\sqrt{n} \|\widehat{\theta}_n - \theta^0\|_\infty = O_p(1).$$

2.2. Deriving the asymptotic distribution

Fix $K > 0$ and let $h_1, h_2, h_3 \in [-K, K]$. In the following, we will write

$$\mu_{1,n,h_1}^0 = \mu_1^0 + \frac{h_1}{\sqrt{n}}, \quad \mu_{2,n,h_2}^0 = \mu_2^0 + \frac{h_2}{\sqrt{n}}, \quad \pi_{n,h_3}^0 = \pi^0 + \frac{h_3}{\sqrt{n}}$$

for $n \geq 1$. Write $h = (h_1, h_2, h_3)$ and consider the process

$$\begin{aligned}
&\mathbb{Q}_n(h) := n\mathbb{D}_n^2(\theta^0 + h/\sqrt{n}) \\
&= n \int_{\mathbb{R}} \Big\{ \pi_{n,h_3}^0 \left(1 - \mathbb{F}_n(\mu_{1,n,h_1}^0 - t) - \mathbb{F}_n(\mu_{1,n,h_1}^0 + t)\right) \\
&\qquad + (1 - \pi_{n,h_3}^0) \left(1 - \mathbb{F}_n(\mu_{2,n,h_2}^0 - t) - \mathbb{F}_n(\mu_{2,n,h_2}^0 + t)\right) \Big\}^2 dt \\
&= n \int_{\mathbb{R}} \Big\{ \pi^0 \left(1 - \mathbb{F}_n(\mu_{1,n,h_1}^0 - t) - \mathbb{F}_n(\mu_{1,n,h_1}^0 + t)\right) \\
&\qquad + (1 - \pi^0) \left(1 - \mathbb{F}_n(\mu_{2,n,h_2}^0 - t) - \mathbb{F}_n(\mu_{2,n,h_2}^0 + t)\right) \\
&\qquad + \frac{h_3}{\sqrt{n}} \Big(\mathbb{F}_n(\mu_{2,n,h_2}^0 - t) + \mathbb{F}_n(\mu_{2,n,h_2}^0 + t) \\
&\qquad\qquad - \mathbb{F}_n(\mu_{1,n,h_1}^0 - t) - \mathbb{F}_n(\mu_{1,n,h_1}^0 + t) \Big) \Big\}^2 dt \\
&= T_{n,1}(h_1, h_2) + T_{n,2}(h) + T_{n,3}(h)
\end{aligned}$$

where

$$\begin{aligned}
T_{n,1}(h_1, h_2) \quad = \quad & n \int_{\mathbb{R}} \Big\{ \pi^0 \left(1 - \mathbb{F}_n(\mu_{1,n,h_1}^0 - t) - \mathbb{F}_n(\mu_{1,n,h_1}^0 + t)\right) \\
& + (1 - \pi^0) \left(1 - \mathbb{F}_n(\mu_{2,n,h_2}^0 - t) - \mathbb{F}_n(\mu_{2,n,h_2}^0 + t)\right) \Big\}^2 dt,
\end{aligned}$$

$$
\begin{aligned}
T_{n,2}(h) \quad = \quad & 2h_3 \int_{\mathbb{R}} \sqrt{n}\Big\{\pi^0 \left(1 - \mathbb{F}_n(\mu^0_{1,n,h_1} - t) - \mathbb{F}_n(\mu^0_{1,n,h_1} + t)\right) \\
& +(1-\pi^0)\left(1 - \mathbb{F}_n(\mu^0_{2,n,h_2} - t) - \mathbb{F}_n(\mu^0_{2,n,h_2} + t)\right)\Big\} \\
& \times\Big(\mathbb{F}_n(\mu^0_{2,n,h_2} - t) + \mathbb{F}_n(\mu^0_{2,n,h_2} + t) \\
& \quad -\mathbb{F}_n(\mu^0_{1,n,h_1} - t) - \mathbb{F}_n(\mu^0_{1,n,h_1} + t)\Big)dt
\end{aligned}
$$

and

$$
\begin{aligned}
T_{n,3}(h) \quad & = h_3^2 \int_{\mathbb{R}} \Big(\mathbb{F}_n(\mu^0_{2,n,h_2} - t) + \mathbb{F}_n(\mu^0_{2,n,h_2} + t) \\
& \quad -\mathbb{F}_n(\mu^0_{1,n,h_1} - t) - \mathbb{F}_n(\mu^0_{1,n,h_1} + t)\Big)^2 dt.
\end{aligned}
$$

Theorem 2.6. *Let $K > 0$, $h = (h_1, h_2, h_3) \in [-K, K]^3$ and $\mathbb{U}$ be a standard Brownian bridge from $(0,0)$ to $(1,0)$. Also, let $s = \mu^0_2 - \mu^0_1$ and $\bar{g}_s(t) = (g^0(t-s) + g^0(t+s))/2$. We assume that*

- *g^0 is bounded on $\mathbb{R}$,*
- *g^0 changes direction of monotonicity only a finite number of times,*
- *g^0 admits a derivative $(g^0)'$ almost everywhere such that $(g^0)'$ is bounded on $\mathbb{R}$ and changes direction of monotonicity only a finite number of times,*
- *there exists $\alpha \in (0, 1/2)$ such that $\int_{\mathbb{R}} \left[G^0(t)(1-G^0(t)\right]^{1/2-\alpha} dt < \infty$ where G^0 is the CDF of g^0.*

Then, under the above assumptions

$$\mathbb{Q}_n \Rightarrow \mathbb{Q}^0$$

in $C([-K,K]^3)$, where

$$\mathbb{Q}^0(h) = h\mathbb{A}h^T + h\mathbb{V} + \mathbb{C}$$

with $\mathbb{A}$ a 3×3 matrix whose entries are given by

$$
\begin{aligned}
\mathbb{A}_{11} &= 4(\pi^0)^2 \int_{\mathbb{R}} \left(\pi^0 g^0(t) + (1-\pi^0)\bar{g}_s(t)\right)^2 dt, \\
\mathbb{A}_{22} &= 4(1-\pi^0)^2 \int_{\mathbb{R}} \left((1-\pi^0) g^0(t) + \pi^0 \bar{g}_s(t)\right)^2 dt, \\
\mathbb{A}_{33} &= \int_{\mathbb{R}} \left(G^0(t+s) - G^0(t-s)\right)^2 dt, \\
\mathbb{A}_{12} &= \mathbb{A}_{21} = 4\pi^0(1-\pi^0) \\
&\quad \times \int_{\mathbb{R}} \left(\pi^0 g^0(t) + (1-\pi^0)\bar{g}_s(t)\right)\left((1-\pi^0) g^0(t) + \pi^0 \bar{g}_s(t)\right) dt, \\
\mathbb{A}_{13} &= \mathbb{A}_{31}
\end{aligned}
$$

$$= -2\pi^0 \int_{\mathbb{R}} \Big(\pi^0 g^0(t) + (1-\pi^0)\bar{g}_s(t)\Big)\Big(G^0(t+s) - G^0(t-s)\Big)dt,$$

$$\mathbb{A}_{23} = \mathbb{A}_{32}$$

$$= -2(1-\pi^0) \int_{\mathbb{R}} \Big((1-\pi^0) g^0(t) + \pi^0\bar{g}_s(t)\Big)\Big(G^0(t+s) - G^0(t-s)\Big)dt,$$

$\mathbb{V}$ *a vector* $\mathbb{R}$ *whose components are distributed as*

$$\begin{aligned}
\mathbb{V}_1 \quad =_d \quad & 4\pi^0 \int_{\mathbb{R}} \Big(\pi^0 g^0(t) + (1-\pi^0) g_s^0(t)\Big) \times \\
& \Big(\pi^0\Big(\mathbb{U}(F^0(\mu_1^0 - t)) + \mathbb{U}(F^0(\mu_1^0 + t))\Big) \\
& +(1-\pi^0)\Big(\mathbb{U}(F^0(\mu_2^0 - t)) + \mathbb{U}(F^0(\mu_2^0 + t))\Big)\Big)dt, \\
\mathbb{V}_2 \quad =_d \quad & 4(1-\pi^0) \int_{\mathbb{R}} \Big((1-\pi^0) g^0(t) + \pi^0\bar{g}_s(t)\Big) \times \\
& \Big(\pi^0\Big(\mathbb{U}(F^0(\mu_1^0 - t)) + \mathbb{U}(F^0(\mu_1^0 + t))\Big) \\
& +(1-\pi^0)\Big(\mathbb{U}(F^0(\mu_2^0 - t)) + \mathbb{U}(F^0(\mu_2^0 + t))\Big)\Big)dt, \\
\mathbb{V}_3 \quad =_d \quad & -2 \int_{\mathbb{R}} \Big(\pi^0\Big(\mathbb{U}(F^0(\mu_1^0 - t)) + \mathbb{U}(F^0(\mu_1^0 + t))\Big) + \\
& (1-\pi^0)\Big(\mathbb{U}(F^0(\mu_2^0 - t)) + \mathbb{U}(F^0(\mu_2^0 + t))\Big)\Big) \\
& \times\Big(G^0(t+s) - G^0(t-s)\Big)dt,
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{C} \quad = \quad & {}_d \int_{\mathbb{R}} \Big\{\pi^0\Big(\mathbb{U}(F^0(\mu_1^0 - t)) + \mathbb{U}(F^0(\mu_1^0 + t))\Big) \\
& +(1-\pi^0)\Big(\mathbb{U}(F^0(\mu_2^0 - t)) + \mathbb{U}(F^0(\mu_2^0 - t))\Big)\Big\}^2 dt.
\end{aligned}$$

We can now state the main theorem of the paper.

Theorem 2.7. *Suppose that the conditions of Theorem 2.2 and Theorem 2.6 hold true. Also, suppose that the symmetric density* g^0 *satisfies* $\int_{\mathbb{R}} |t|^5 g^0(t)dt < \infty$. *Then, the matrix* $\mathbb{A}$ *defined above is definite positive and the process* $\mathbb{Q}^0$ *admits a unique minimizer* h^0 *given by the solution of the linear equation*

$$\mathbb{A}h^0 = -\frac{\mathbb{V}}{2}.$$

Furthermore, we have that

$$\sqrt{n}(\widehat{\theta}_n - \theta^0) \to_d h^0.$$

Remark 2.8. *The assumptions made above are satisfied by a variety of distributions. In particular, this is the case if we assume that the unknown symmetric density g^0 is log-concave on its support as done in* **?**. *Then, the density g^0 is bounded and changes direction of monotonicity only once since log-concavity implies unimodality. By Theorem 2.3 of Finner and Roters (1993), we know that G^0 and $1-G^0$ are both log-concave ion $\mathbb{R}$. Hence, Lemma 2 of Schoenberg (1951) can be applied to g^0, G^0 and $1-G^0$ and we can find constants $a>0, b>0$ such that $g^0(|t|) \leq a\exp(-b|t|)$ for all $t \in \mathbb{R}$, $G^0(t) \leq a\exp(-b|t|)$ for all $t \leq 0$ and $1-G_0(t) \leq a\exp(-bt)$ for all $t>0$. Therefore, we have that*

$$\int_{\mathbb{R}} (g^0(t))^{1/2} dt < \infty, \quad \int_{\mathbb{R}} [G^0(t)(1-G^0(t))]^{\gamma} dt < \infty, \quad \int_{\mathbb{R}} |t|^{\gamma} g^0(t) dt < \infty.$$

for any $\gamma > 0$, and in particular for $\gamma \in (0, 1/2)$.

Remark 2.9. *Recall that Hunter et al. (2007) have shown that the mixture model is identifiable whenever $\pi^0 \in (0,1) \setminus \{1/2\}$ and $\mu_1^0 < \mu_2^0$. Thus, it is not surprising that the matrix $\mathbb{A}$ is singular if $\pi^0 \in \{0, 1/2, 1\}$ or $\mu_1^0 = \mu_2^0$. To show this, it is sufficient to exhibit $h \neq (0,0,0)$ which satisfies $h\mathbb{A}h^T = 0$. As noted in Supplement A,*

$$h\mathbb{A}h^T = \int_{\mathbb{R}} \Big\{ \pi^0 \big(f(\mu_1^0 + t) + f(\mu_1^0 - t)\big) h_1 + (1-\pi^0)\big(f(\mu_2^0 + t) + f(\mu_2^0 - t)\big) h_2 \\ - \big(F(\mu_2^0 + t) - F(\mu_1^0 + t) + F(\mu_2^0 - t) - F(\mu_1^0 - t)\big) h_3 \Big\}^2 .$$

In case $\pi^0 = 0$ ($\pi^0 = 1$), then $h\mathbb{A}h^T = 0$ for $h = (1,0,0)$ ($h = (0,1,0)$). If $\pi^0 \notin \{0,1\}$ and $\mu_1^0 = \mu_2^0$, then $h\mathbb{A}h^T = 0$ for $(-(1-\pi^0), \pi^0, 0)$. Lastly, if $\pi^0 = 1/2$, then $f(\mu_1^0 + t) + f(\mu_1^0 - t) = f(\mu_2^0 + t) + f(\mu_2^0 - t) = g^0(t) + g_s^0(t)$ using the section on useful formulae in Supplement A. This implies that $h\mathbb{A}h^T = 0$ for $h = (-1, 1, 0)$.

Remark 2.10. *One may wonder whether the Hodges-Lehmann estimator is efficient. The question is obviously related to finding the efficient bound in this semi-parametric model. The tangent space with respect to the symmetric component g^0 is given by*

$$\Gamma = \Big\{ \frac{\pi^0 k(\cdot - \mu_1^0) + (1-\pi^0) k(\cdot - \mu_2^0)}{f^0}, \\ \text{such that } k \text{ is even and } \int_{\mathbb{R}} k(x) dx = 0 \Big\},$$

and

$$\Gamma^{\perp} = \Big\{ h : \text{there exist } c \in \mathbb{R} \text{ such that} \\ \pi^0 (h(\cdot + \mu_1^0) + h(- \cdot + \mu_1^0)) + (1-\pi^0)(h(\cdot + \mu_2^0) + h(- \cdot + \mu_2^0) = c \Big\}$$

the corresponding orthogonal space. To obtain the latter, one can follow the lines of the proof given by Ma and Yao (2015) (page 467) in the context a two-component mixture model where one component has a parametric form. In this

problem, one can relatively easily guess how projection of the score functions on the space $\Gamma^{\perp}$ and hence the efficient scores. Finding these efficient scores for the mixture parameters in our semi-parametric model is a much harder problem. The main difficulty stems from the fact that one needs to solve a nonstandard functional equation in the space of even functions that integrate to zero. Thus, finding the efficient bound on the variance and answering the question whether $\widehat{\theta}_n$ attains this bound asymptotically are questions that remain open.

2.3. Some comments on the asymptotic variance

It follows from the previous section that under the assumptions made on g^0 we have that

$$\sqrt{n}(\widehat{\theta}_n - \theta^0) \to_d \mathcal{N}(\mathbf{0}, \Sigma)$$

where

$$\Sigma = \frac{1}{4}\mathbb{A}^{-1}\Gamma\mathbb{A}^{-1} \tag{2.8}$$

and Γ is the 3×3 dispersion matrix of $\mathbb{V}$. The components of Γ can be related to the covariance of $\mathbb{U}$; i.e., $Cov(\mathbb{U}(x), \mathbb{U}(y)) = x \vee y - xy$ for $x, y \in [0, 1]$. However, the expressions obtained cannot be used to get explicit values. To give an example, we can write $\Gamma_{1,1}$, the variance of $\mathbb{V}_1$, as

$$\begin{aligned}\Gamma_{1,1} &= 4(\pi^0)^2 \int_{\mathbb{R}}\int_{\mathbb{R}} \Big(\pi^0 g^0(x) + (1-\pi^0)g_s^0(x)\Big)\Big(\pi^0 g^0(y) + (1-\pi^0)g_s^0(y)\Big) \\ &\qquad \times S(x,y)dxdy\end{aligned}$$

where $S = (\pi^0)^2 S_1 + \pi^0(1-\pi^0)S_2 + (1-\pi^0)^2 S_3$ with

$$\begin{aligned}S_1(x,y) &= 2\Big(F^0(\mu_1^0 - x) \wedge F^0(\mu_1^0 - y) - F^0(\mu_1^0 - x)F^0(\mu_1^0 - y) \\ &\quad + F^0(\mu_1^0 + x) \wedge F^0(\mu_1^0 - y) - F^0(\mu_1^0 + x)F^0(\mu_1^0 - y)\Big), \\ S_2(x,y) &= 4\Big(F^0(\mu_1^0 - x) \wedge F^0(\mu_2^0 - y) - F^0(\mu_1^0 - x)F^0(\mu_2^0 - y) \\ &\quad + F^0(\mu_1^0 + x) \wedge F^0(\mu_2^0 - y) - F^0(\mu_1^0 + x)F^0(\mu_2^0 - y)\Big), \text{ and} \\ S_3(x,y) &= 2\Big(F^0(\mu_2^0 - x) \wedge F^0(\mu_2^0 - y) - F^0(\mu_2^0 - x)F^0(\mu_2^0 - y) \\ &\quad + F^0(\mu_2^0 + x) \wedge F^0(\mu_2^0 - y) - F^0(\mu_2^0 + x)F^0(\mu_2^0 - y)\Big).\end{aligned}$$

From these expressions, we see that the terms $F^0(\mu_j^0 - x)F(\mu_j^0 - y),\ i, j = 1, 2$, and $F^0(\mu_1^0 - x)F(\mu_2^0 - y)$ are the main obstacle when trying to get a general formula for the integral defining $\Gamma_{1,1}$.

In Theorem Hunter et al. (2007), four sufficient conditions were given to ensure that the estimator $\widehat{\theta}_n$ has an asymptotic Gaussian distribution. The asymptotic variance given in the theorem is different from the one that we have found. When comparing the formula in (2.8) with the one in Hunter et al. (2007),

we see that the matrix $\mathbb{A}$ should play the role of J appearing in Condition (i) of Hunter et al. (2007). The latter should be equal to the second derivative of the function $\theta \mapsto \mathbb{D}^2(\theta)$ at the true value θ^0, under the additional assumption that is positive definite. From our calculations, we can see that the matrix $\mathbb{A}$ is not linked directly to the second derivative of $\theta \mapsto \mathbb{D}^2(\theta)$ but rather to the application $h \mapsto n\mathbb{D}^2(\theta^0 + hn^{-1/2})$. Condition (iii) of Hunter et al. (2007) assumes existence of a function Δ such that $E[\Delta(X)] = 0$, $E[\|\Delta\|^2] < \infty$, and where $\pi_1 f_\theta(x) - \pi_1 f_{\theta^0}(x) = (\theta - \theta^0)\Delta(x) + \|\theta - \theta^0\| r_\theta(x)$ with r_θ satisfying some uniform integrability condition in the neighborhood of θ^0. Here, f_θ is the same function defined in (4.9). Using the definition of π_1 we have $\pi_1 f_\theta(x) = E[f_\theta(x, X)] - \mathbb{D}^2(\theta)$. We know that $\mathbb{D}^2(\theta)$ takes the value 0 at θ^0 and has its gradient equal to 0 at the same value, and hence one should focus on $E[f_\theta(x, X)]$ to get $\Delta(X)$. Despite our efforts, it seems hard to connect the vector $\mathbb{V}$ given in Theorem 2.6 and the vector $\Delta(X)$ we obtain.

In the next section, we investigate numerically how the variances of the estimators $\widehat{\pi}_n, \widehat{\mu}_{1,n}, \widehat{\mu}_{2,n}$ behave as a function of the true span $\mu_2^0 - \mu_1^0$ and the mixing probability π^0.

3. Limitations of the asymptotic normality: Some examples

We consider the following three symmetric distributions

- $\mathcal{N}(0, 1)$: the standard Gaussian distribution
- $\mathcal{U}[-1, 1]$: the uniform distribution on $[-1, 1]$
- $\mathcal{L}(1)$: the double exponential (Laplace) on $\mathbb{R}$ with intensity equal to 1

for which we compute the asymptotic variance of the estimators $\widehat{\pi}_n, \widehat{\mu}_{1,n}, \widehat{\mu}_{2,n}$ for different ranges of μ_1^0, μ_2^0 and π^0. To be more precise, we compute Monte-Carlo estimates for these variances based the formula in (2.8) and Brownian bridge approximation. For the latter we sampled $N = 10^4$ of independent uniform random variables and computed $\widetilde{\mathbb{U}}_N(t) = \sqrt{N}(\widetilde{\mathbb{G}}_N(t) - t),\ t \in [0, 1]$ where $\widetilde{\mathbb{G}}_N$ is the empirical distribution of the obtained uniform sample. We estimated the variance matrix Γ of the vector $\mathbb{V}$ using B replications of $\widetilde{\mathbb{U}}_N$ and approximating the integrals defining $\mathbb{V}_j, j = 1, 2, 3$ by a finite sum over an equally spaced grid with chosen lower and upper endpoints and a given mesh. The matrix $\mathbb{A}$ can be explicitly computed for any θ^0 for $\mathcal{U}[-1, 1]$ and $\mathcal{L}(1)$. For the distribution $\mathcal{N}(0, 1)$ closed formulas can be found for the entries $\mathbb{A}_{11}, \mathbb{A}_{12}$ and $\mathbb{A}_{22}$, whereas the remaining ones can be approximated using the same discretization described above. The explicit formulas can be found in Supplement A.

In Table 1 we give the values of $n\times$ the empirical variances of the Hodges-Lehmann estimators of the mixture parameters π^0, μ_1^0 and μ_2^0 for the three symmetric (and log-concave) distributions used to illustrate the theory, and those of the corresponding theoretical asymptotic variances; i.e., the diagonal entries of the covariance matrix Σ given in (2.8). The main goal here is to obtain a numerical confirmation of our asymptotic result. The mixture parameters are chosen in a way that the mixture components are well-separated. The empirical variances are based on samples of size $n = 10000$ and 500 Monte Carlo repli-

cations, whereas an approximation of the asymptotic variances is obtained by taking $N = 10^5$ independent uniform random variables, a grid step to 0.01, and -60 and 60 as the upper and lower integration bounds and $B = 800$ replications of the vector $\mathbb{V}$. The obtained values are clearly close and we expect them to be much closer for larger sample sizes and finer approximation of the asymptotic covariance matrix.

Law	Variance	$(\pi^0, \mu_2^0) = (0.3, 3)$	$(\pi^0, \mu_2^0) = (0.2, 4)$	$(\pi^0, \mu_2^0) = (0.25, 5)$
$\mathcal{N}(0,1)$	Empirical	$(8.10, 3.35, 0.35)$	$(7.51, 1.65, 0.20)$	$(4.25, 1.64, 0.20)$
	Asymptotic	$(7.28, 3.04, 0.40)$	$(7.76, 1.60, 0.19)$	$(5.03, 1.59, 0.19)$
$\mathcal{U}[-1,1]$	Empirical	$(1.20, 0.47, 0.21)$	$(1.90, 0.37, 0.16)$	$(1.37, 0.46, 0.18)$
	Asymptotic	$(1.01, 0.48, 0.20)$	$(1.80, 0.47, 0.16)$	$(1.45, 0.42, 0.20)$
$\mathcal{L}(1)$	Empirical	$(8.15, 2.65, 0.48)$	$(10.84, 1.59, 0.23)$	$(6.95, 1.62, 0.24)$
	Asymptotic	$(9.10, 2.31, 0.43)$	$(11.99, 1.73, 0.27)$	$(6.99, 1.66, 0.24)$

TABLE 1

Values of $n\times$ the empirical variance of the Hodges-Lehmann estimators of π^0, μ_1^0 and μ_2^0 (Empirical) and those of the corresponding asymptotic variance (Asymptotic). To obtain the empirical variances the sample size was taken to be $n = 10000$ with 500 Monte Carlo replications. In computing the asymptotic variances, the Brownian bridge was approximated using $N = 10^5$ independent uniform random variables and the integral defining $\mathbb{V}$ is approximated using a grid step to 0.01, $[-60, 60]$ as the interval of integration and $B = 800$ Monte Carlo replications of $\mathbb{V}$ (see text for more details). The true mixture parameters are π^0 and μ_2^0 are as shown in the table, whereas $\mu_1^0 = 0$ in all the examples.

Let us now turn to another aspect of our asymptotic result. The plots in Figure 1 show the variance of the estimators of μ_1^0, μ_2^0 on the left and π^0 on the right as a function of $\pi^0 \in [0.01, 0.49] \cup [0.51, 0.99]$ for the fixed value $(\mu_1^0, \mu_2^0) = (0, 1)$ and in the case where the true symmetric density is Gaussian. The variances obtained are extremely large for small and large values of π^0, especially for the estimates of the mixture locations, reaching a maximal value of 1.85×10^6 for the estimate of μ_1^0 when the true mixing probability is $\pi^0 = 0.01$! The variances are also very large for the estimate of π^0 with a maximal value of 1600. A similar phenomenon can be seen in Figure 4 and 7 when the true density is $\mathcal{U}[-1, 1]$ and $\mathcal{L}(1)$ respectively for the same range of π^0 and the fixed value $(\mu_1^0, \mu_2^0) = (0, 0.5)$. The variances seem to be take somewhat smaller values, although still large, when the true span is increased; see Figure 2, 5 and 8. Although the results are disappointing, they are somehow expected as the estimation procedure fails to be efficient when the mixture components are not well separated. This is confirmed by the reasonable values obtained in Figure 3 and Figure 6. The variances are plotted as a function of the span $\mu_2^0 - \mu_1^0 \in [2, 10]$ for $\mathcal{N}(0, 1)$ and $\mu_2^0 - \mu_1^0 \in [2.5, 10]$ for $\mathcal{U}[-1, 1]$, and when the true mixing probability is $\pi^0 = 0.3$.

Although our numerical results are only shown for the Gaussian, uniform and double exponential distributions, they give a clear warning that the Hodges-Lehmann is not be used when it is believed that the mixture proportion is very small or large or that the components are not well separated. The difficulty lies of course in the fact that such an information is contained in the values of the parameters that we would like to estimate and hence cannot be known in advance.

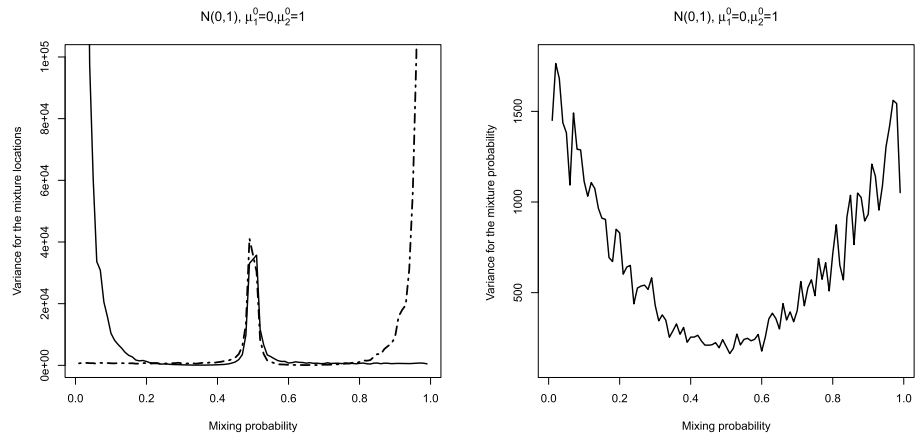

FIG 1. *Variance for the Hodges-Lehmann estimators of μ_1^0 in solid line and μ_2^0 in dashed line (left) and variance of the Hodges-Lehmann estimator for π^0 (right) as a function of π^0. Here, $\mu_1^0 = 0, \mu_2^0 = 1$ and the true symmetric distribution is $\mathcal{N}(0,1)$.*

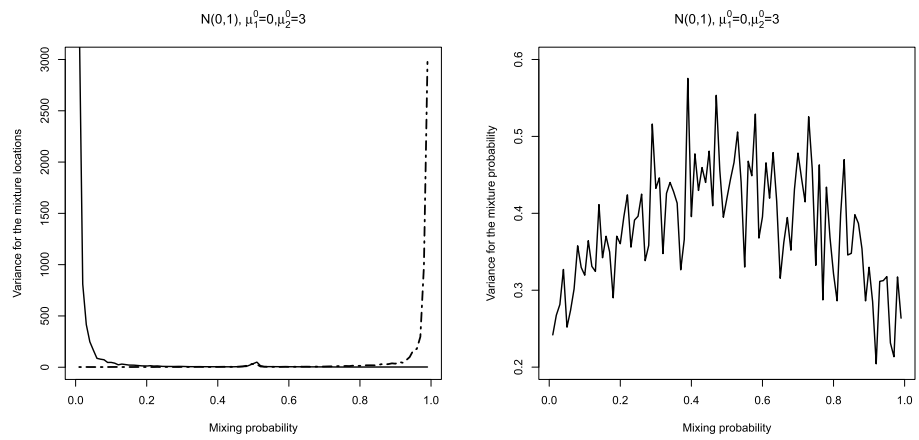

FIG 2. *Variance for the Hodges-Lehmann estimators of μ_1^0 in solid line and μ_2^0 in dashed line (left) and variance of the Hodges-Lehmann estimator for π^0 (right) as a function of π^0. Here, $\mu_1^0 = 0, \mu_2^0 = 3$ and the true symmetric distribution is $\mathcal{N}(0,1)$.*

4. Proofs

4.1. Proof of existence of the (Hogdes-Lehmann) estimator

Existence of $\widehat{\theta}_n = (\widehat{\pi}_n, \widehat{\mu}_{1,n}, \widehat{\mu}_{2,n})$ can be established using standard results from optimization. Hunter et al. (2007) established a very useful alternative representation of $\mathbb{D}_n(\theta)^2$ given in (4.10). For $\theta = (\pi, \mu_1, \mu_2) \in \Theta$ in (1.2), let f_θ be the function defined on $\mathbb{R} \times \mathbb{R}$ by

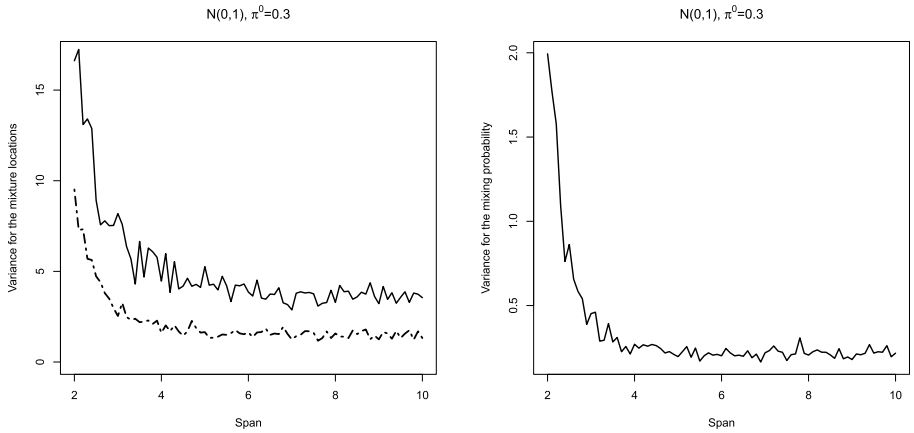

FIG 3. *Variance for the Hodges-Lehmann estimators of* μ_1^0 *in solid line and* μ_2^0 *in dashed line (left) and variance for the Hodges-Lehmann estimator* π^0 *(right) as a function of* $s = \mu_2^0 - \mu_1^0$. *Here,* $\pi^0 = 0.3$ *and the true symmetric distribution is* $\mathcal{N}(0,1)$.

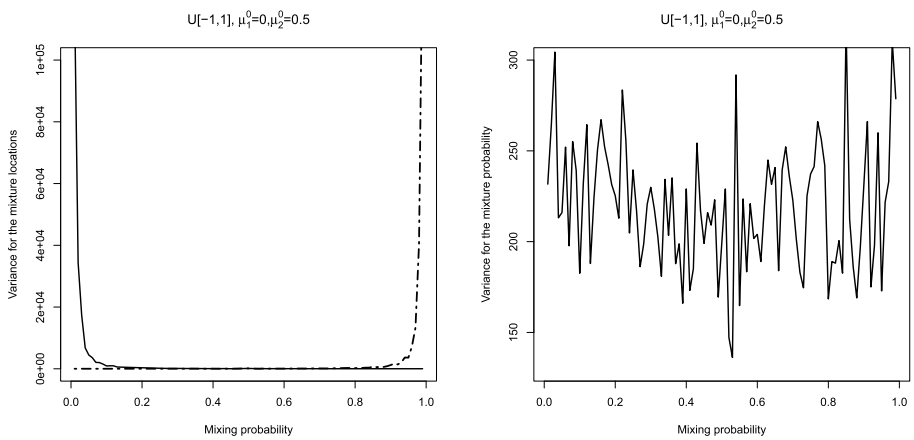

FIG 4. *Variance for the Hodges-Lehmann estimators of* μ_1^0 *in solid line and* μ_2^0 *in dashed line (left) and variance of the Hodges-Lehmann estimator for* π^0 *(right) as a function of* π^0. *Here,* $\mu_1^0 = 0, \mu_2^0 = 0.5$ *and the true symmetric distribution is* $\mathcal{U}(-1,1)$.

$$
\begin{aligned}
&f_\theta(x,y)\\
&= \pi^2\Big(|x+y-2\mu_1| - |x-y|\Big) + (1-\pi)^2\Big(|x+y-2\mu_2| - |x-y|\Big)\\
&+\pi(1-\pi)\Big(2|x+y-\mu_1-\mu_2| - |x-y+\mu_1-\mu_2| - |x-y+\mu_2-\mu_1|\Big)\\
&= \pi^2\psi_1(x,y,\mu_1,\mu_2) + (1-\pi)^2\psi_2(x,y,\mu_1,\mu_2) + \pi(1-\pi)\psi_3(x,y,\mu_1,\mu_2)\\
&= \Big(\psi_1(x,y,\mu_1,\mu_2) + \psi_2(x,y,\mu_1,\mu_2) - \psi_3(x,y,\mu_1,\mu_2)\Big)\pi^2
\end{aligned}
$$

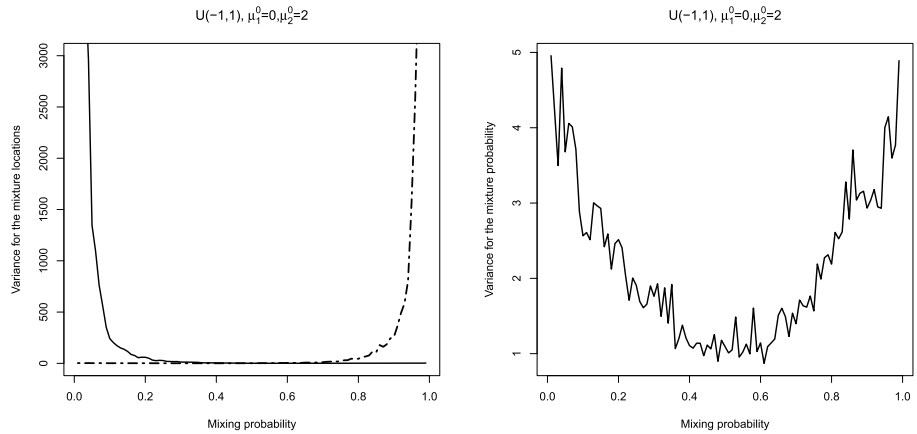

FIG 5. *Variance for the Hodges-Lehmann estimators of μ_1^0 in solid line and μ_2^0 in dashed line (left) and variance of the Hodges-Lehmann estimator for π^0 (right) as a function of π^0. Here, $\mu_1^0 = 0, \mu_2^0 = 2$ and the true symmetric distribution is $\mathcal{U}(-1, 1)$.*

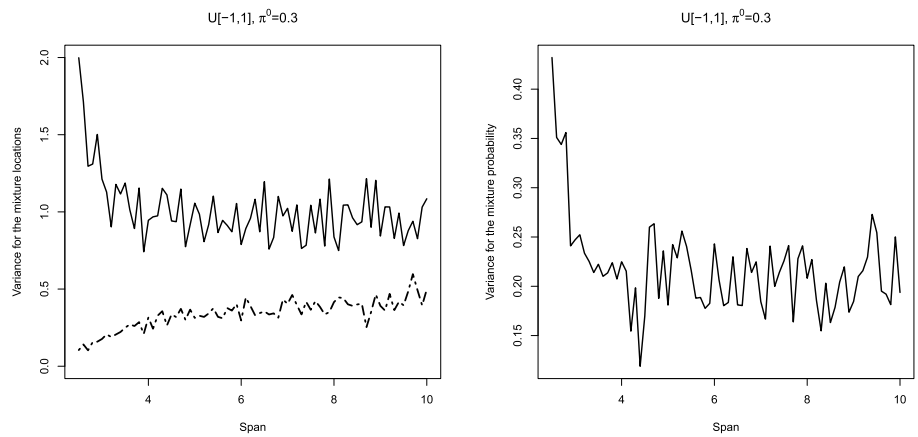

FIG 6. *Variance for the Hodges-Lehmann estimators of μ_1^0 in solid line and μ_2^0 in dashed line (left) and variance for the Hodges-Lehmann estimator π^0 (right) as a function of $s = \mu_2^0 - \mu_1^0$. Here, $\pi^0 = 0.3$ and the true symmetric distribution is $\mathcal{U}(-1, 1)$.*

$$+\big(-2\psi_2(x, y, \mu_1, \mu_2) + \psi_3(x, y, \mu_1, \mu_2)\big)\pi + \psi_2(x, y, \mu_1, \mu_2) \tag{4.9}$$

with

$$\begin{aligned}
&\psi_1(x, y), \mu_1, \mu_2) = |x + y - 2\mu_1| - |x - y|,\\
&\psi_2(x, y), \mu_1, \mu_2) = |x + y - 2\mu_2| - |x - y|,\\
&\psi_3(x, y), \mu_1, \mu_2) = 2|x + y - \mu_1 - \mu_2| - |x - y + \mu_1 - \mu_2| - |x - y + \mu_2 - \mu_1|.
\end{aligned}$$

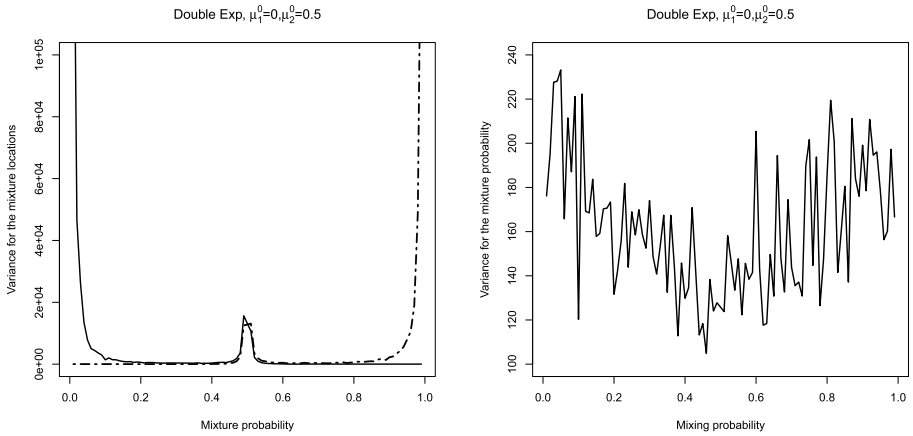

FIG 7. *Variance for the Hodges-Lehmann estimators of μ_1^0 in solid line and μ_2^0 in dashed line (left) and variance of the Hodges-Lehmann estimator for π^0 (right) as a function of π^0. Here, $\mu_1^0 = 0, \mu_2^0 = 0.5$ and the true symmetric distribution is $\mathcal{L}(1)$.*

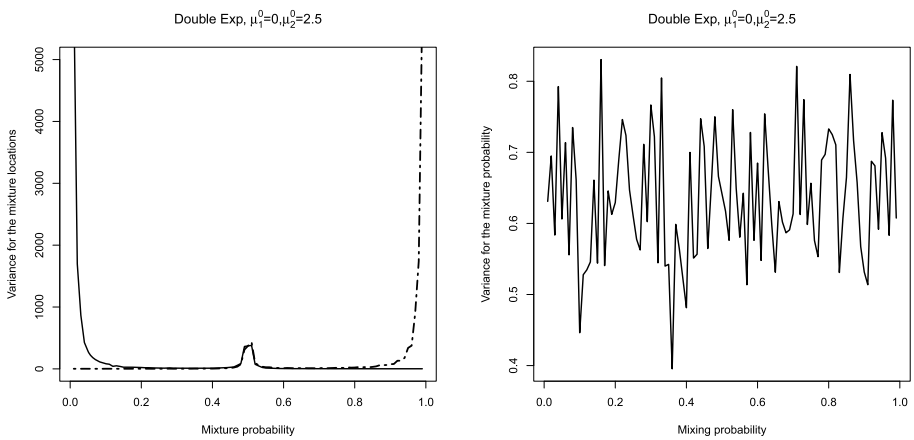

FIG 8. *Variance for the Hodges-Lehmann estimators of μ_1^0 in solid line and μ_2^0 in dashed line (left) and variance of the Hodges-Lehmann estimator for π^0 (right) as a function of π^0. Here, $\mu_1^0 = 0, \mu_2^0 = 2.5$ and the true symmetric distribution is $\mathcal{L}(1)$.*

Now, formula (15) of Hunter et al. (2007) yields

$$\begin{aligned}\mathbb{D}_n(\theta)^2 &= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n} f_\theta(X_i, X_j) \\ &= \Big[\bar{\psi}_1(\mu_1,\mu_2) + \bar{\psi}_2(\mu_1,\mu_2) - \bar{\psi}_3(\mu_1,\mu_2)\Big]\pi^2 \\ &\quad + \Big[-2\bar{\psi}_2(\mu_1,\mu_2) + \bar{\psi}_3(\mu_1,\mu_2)\Big]\pi + \bar{\psi}_2(\mu_1,\mu_2) \qquad (4.10)\end{aligned}$$

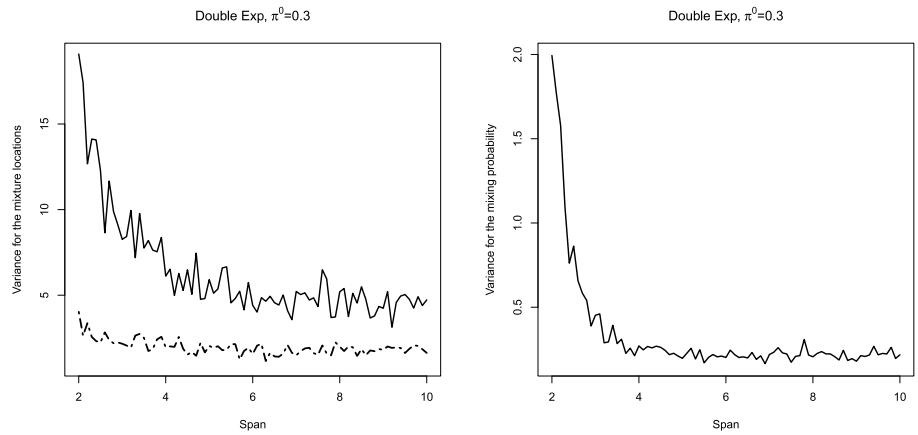

FIG 9. *Variance for the Hodges-Lehmann estimators of μ_1^0 in solid line and μ_2^0 in dashed line (left) and variance for the Hodges-Lehmann estimator π^0 (right) as a function of $s = \mu_2^0 - \mu_1^0$. Here, $\pi^0 = 0.3$ and the true symmetric distribution is $\mathcal{L}(1)$.*

where $\bar{\psi}_k(\mu_1, \mu_2) = n^{-2} \sum_{j=1}^n \sum_{i=1}^n \psi_k(X_i, X_j, \mu_1, \mu_2)$ for $k \in \{1, 2, 3\}$. To have a simpler framework, one can extend the space of minimization to the closed set $\widetilde{\Theta} = [0, 1] \times \{(\mu_1, \mu_2) \in \mathbb{R}^2 : \mu_1 \leq \mu_2\}$. For fixed μ_1 and μ_2, the function in π defined (4.10) is a parabola. Hence, straightforward calculations show that the optimal value of $\mathbb{D}^2(\theta)$ in π for the fixed μ_1 and μ_2 is given by

$$\mathbb{D}_n(\widetilde{\pi}, \mu_1, \mu_2)^2 = \frac{4\bar{\psi}_1(\mu_1, \mu_2)\bar{\psi}_2(\mu_1, \mu_2) - [\bar{\psi}_3(\mu_1, \mu_2)]^2}{4\big[\bar{\psi}_1(\mu_1, \mu_2) + \bar{\psi}_2(\mu_1, \mu_2) - \bar{\psi}_3(\mu_1, \mu_2)\big]}$$

whenever $\bar{\psi}_1(\mu_1, \mu_2) + \bar{\psi}_2(\mu_1, \mu_2) - \bar{\psi}_3(\mu_1, \mu_2) > 0$, and

$$\widetilde{\pi}(\mu_1, \mu_2) = \frac{2\bar{\psi}_2(\mu_1, \mu_2) - \bar{\psi}_3(\mu_1, \mu_2)}{2\Big[\bar{\psi}_1(\mu_1, \mu_2) + \bar{\psi}_2(\mu_1, \mu_2) - \bar{\psi}_3(\mu_1, \mu_2)\Big]} \in [0, 1].$$

The goal now is to show that letting $\mu_1 \to -\infty$ and $\mu_2 \to \infty$ does not make the target function $\theta \mapsto \mathbb{D}_n(\theta)^2$ any smaller. There are two cases to consider:

- there exists $c > 0$ such that $|\mu_1 + \mu_2| \leq c$
- $\mu_1 + \mu_2 \to -\infty$ or $\mu_1 + \mu_2 \to \infty$.

We consider the first case. Let us denote $T_n = n^{-2} \sum_{j=1}^n \sum_{i=1}^n (X_i + X_j) = 2\bar{X}_n$, $D_n = n^{-2} \sum_{j=1}^n \sum_{i=1}^n |X_i - X_j|$ and $S_n(\mu_1, \mu_2) = n^{-2} \sum_{j=1}^n \sum_{i=1}^n |X_i + X_j - (\mu_1 + \mu_2)|$. Then, for $|\mu_1|$ and μ_2 large enough it is easy to see that

$$\begin{aligned} \bar{\psi}_1(\mu_1, \mu_2) &= -2\mu_1 + T_n - D_n, \\ \bar{\psi}_2(\mu_1, \mu_2) &= 2\mu_2 - T_n - D_n, \\ \bar{\psi}_3(\mu_1, \mu_2) &= -2(\mu_2 - \mu_1) + 2S_n(\mu_1, \mu_2). \end{aligned} \tag{4.11}$$

Hence,

$$\widetilde{\pi}(\mu_1,\mu_2) = \frac{3\mu_2 - \mu_1 - (T_n + D_n^+) - S_n(\mu_1,\mu_2)}{4(\mu_2-\mu_1) - 2D_n^+ - 2S_n(\mu_1,\mu_2)}.$$

Note that $\widetilde{\pi}(\mu_1,\mu_2) > 0$ for $|\mu_1|$ and μ_2 large enough. Also, $1-\widetilde{\pi}(\mu_1,\mu_2)$ has the same sign as $\mu_2 - 3\mu_1 + T_n - D_n - S_n(\mu_1,\mu_2)$ and hence $\widetilde{\pi}(\mu_1,\mu_2) < 1$ for $|\mu_1|$ and μ_2 large enough. After some algebra, it follows that

$$\begin{aligned}
&\mathbb{D}_n(\widetilde{\pi},\mu_1,\mu_2)^2 \\
&= \frac{1}{4(\mu_2-\mu_1) - 2D_n - 2S_n(\mu_1,\mu_2)} \\
&\quad \times\Big(D_n^2 - T_n^2 + 2(\mu_1+\mu_2)T_n - (\mu_1+\mu_2)^2 - S_n(\mu_1,\mu_2)^2 \\
&\qquad +2\,(\mu_2-\mu_1)(S_n(\mu_1,\mu_2) - D_n)\Big) \\
&\to \frac{1}{2}\,(S_n(\mu_1,\mu_2) - D_n) \\
&= \frac{1}{2}\left[\frac{1}{n^2}\sum_{j=1}^n\sum_{i=1}^n |X_i + X_j - \mu_1 - \mu_2| - \frac{1}{n^2}\sum_{j=1}^n\sum_{i=1}^n |X_i - X_j|\right]
\end{aligned}$$

as $|\mu_1|,\mu_2 \to \infty$. To get the convergence to the above limit, we used the fact that $|\mu_1+\mu_2| \le c$ for some $c>0$ as it is the assumption in this first case. But note that if we take μ_1 and μ_2 such that $-\mu_1,\mu_2 \in [|X_{(n)}|, |X_{(n)}| + c]$, where $X_{(n)} = \max_{1\le i\le n} X_i$, then $\mu_1+\mu_2 \in [-c,c]$ and still have $\bar{\psi}_k(\mu_1,\mu_2)$ given by the expressions in (4.11) for $j=1,2,3$. Now, taking $\pi = 1/2$ yields

$$\begin{aligned}
\mathbb{D}_n^2(1/2,\mu_1,\mu_2) &= \frac{1}{4}\Big(\bar{\psi}_1(\mu_1,\mu_2) + \bar{\psi}_2(\mu_1,\mu_2) + \bar{\psi}_3(\mu_1,\mu_2)\Big) \\
&= \frac{1}{2}(S_n(\mu_1,\mu_2) - D_n) \\
&= \frac{1}{2}\left[\frac{1}{n^2}\sum_{j=1}^n\sum_{i=1}^n |X_i + X_j - \mu_1 - \mu_2| - \frac{1}{n^2}\sum_{j=1}^n\sum_{i=1}^n |X_i - X_j|\right].
\end{aligned}$$

Thus, we have found a point in the compact set $[0,1]\times\big[-|X_{(n)}|, |X_{(n)}| + c\big]^2$, yielding the same value of the target function $\theta \mapsto \mathbb{D}_n^2(\theta)$ as the limit of $\mathbb{D}_n^2(\widetilde{\pi}(\mu_1,\mu_2),\mu_1,\mu_2)$ when $|\mu_1|,|\mu_2| \to \infty$ under the additional constraint that $|\mu_1+\mu_2| \le c$. Now, we consider the second case and assume that $\mu_1+\mu_2 = \mu_2 - |\mu_1| \to \infty$ as the other case can be handled similarly. Note that only the expression of $\bar{\psi}_3(\mu_1,\mu_2)$ differs in this case as we have that

$$\bar{\psi}_3(\mu_1,\mu_2) = 4\mu_1 - 2T_n.$$

Some algebra yields

$$\widetilde{\pi}(\mu_1,\mu_2) = \frac{2(\mu_2-\mu_1) - D_n}{2\Big(\mu_2 - 3\mu_1 - D_n + T_n\Big)}$$

which can be shown to be in $(0,1)$ for $|\mu_1|, \mu_2$ large enough. More algebra yields

$$\begin{aligned}
&\mathbb{D}_n(\widetilde{\pi}, \mu_1, \mu_2)^2 \\
&= \frac{-4\mu_1\big(\mu_2+\mu_1-T_n\big) + 2\big(\mu_2+\mu_1\big)T_n - D_n(\mu_2-\mu_1) + D_n^2 - 2T_n^2}{-6\mu_1 + 2\mu_2 - 2D_n + 2T_n} \\
&\to \infty
\end{aligned}$$

as $|\mu_1| \to \infty, \mu_2 \to \infty, \mu_1+\mu_2 \to \infty$. To see this, we use the facts that $\mu_1 = -|\mu_1|$ and

$$\frac{\mu_2 - |\mu_1|}{|\mu_1|\Big(\mu_2 - |\mu_1| - T_n\Big)} = \frac{1}{|\mu_1|}\frac{1}{1 - T_n/(\mu_2 - |\mu_1|)} \quad \to \quad 0$$

and

$$\frac{\mu_2}{|\mu_1|\Big(\mu_2 - |\mu_1| - T_n\Big)} \quad = \quad \frac{\mu_2 - |\mu_1|}{|\mu_1|\Big(\mu_2 - |\mu_1| - T_n\Big)} + \frac{1}{\mu_2 - |\mu_1| - T_n} \to 0$$

so that the claimed limit follows after dividing both the numerator and denominator by $4|\mu_1|\Big(\mu_2 - |\mu_1| - T_n\Big)$.

We conclude that the minimization problem can be performed on the compact set $[0,1] \times \Big\{(\mu_1,\mu_2) \in \mathbb{R}^2 : \{-D \le \mu_1 \le \mu_2 \le D\}\Big\}$ for some constant $D > 0$. Combining this with continuity of $\mathbb{D}_n$ (since f_θ is continuous) gives existence of $\widehat{\theta}_n$. Almost sure convergence of $\widehat{\theta}_n$ to the true parameter (see Theorem 2.1 below) will ensure that $\widehat{\theta}_n \in \Theta^*$ with probability one for n large enough. □

4.2. Proofs of the $\sqrt{n}$-rate of convergence of $\widehat{\theta}_n$

We start with showing Theorem 2.3.

Proof of Theorem 2.3. The proof of (1) goes along the lines of the hint given in van der Vaart and Wellner (1996) for Problem 20, page 153. Take three points (x_1,t_1), (x_2,t_2), and $(x_3,y_3) \in \mathbb{R}^2$ such that $x_1 \le x_2 \le x_3$. Then, the class of all subgraphs $\big\{c\mathbb{1}_{(a,b]}, a, b, c \in \mathbb{R}\big\}$ pick out the point (x_2,t_2) unless $t_2 > \max(t_1,t_3)$. Now, take four points (x_1,t_1), (x_2,t_2), (x_3,t_3) and $(x_4,t_4) \in \mathbb{R}^2$ such that $x_1 \le x_2 \le x_3 \le x_4$. We want to show that these points are not shattered by the considered class of subgraphs. Suppose they are shattered. Then, this implies that every subset of three points with nondecreasing x_i is also shattered. Hence, we should have $t_2 > \max(t_1,t_3)$, $t_3 > \max(t_2,t_4)$ (and also $t_2 > \max(t_1,t_4)$ and $t_3 > \max(t_1,t_4)$), which is impossible. We conclude that the considered class is a VC-subgraph of index 4.

To show (2), let us fix $\epsilon > 0$. Note that $\widetilde{F} = F_1 + \ldots + F_m$ is an envelope for each of the classes $\mathcal{F}_j,\ j = 1,\ldots,m$. Consider the covering numbers $N_j = \mathcal{N}(\epsilon\|\widetilde{F}\|_{Q,2}, \mathcal{F}_j, L_2(Q))$ for some probability measure Q. An element $f \in \mathcal{F}$ can

be written as $f = f_1 + \ldots + f_m$ for $(f_1, \ldots, f_m) \in \mathcal{F}_1 \times \ldots \times \mathcal{F}_m$. Let $(f_{i_1}, \ldots, f_{i_m})$ be the m-tuple such that $\|f_j - f_{i_j}\| \leq \epsilon \|\widetilde{F}\|_{Q,2}$ for $j = 1, \ldots, m$. Then,

$$\|f - \sum_{j=1}^{m} f_{i_j}\|_{Q,2} \leq \; m \; \|\widetilde{F}\|_{Q,2} \; \epsilon.$$

Hence, $N(\epsilon \; m \; \|\widetilde{F}\|_{Q,2}, \mathcal{F}, L_2(Q)) \leq \prod_{j=1}^{m} N_j$. Using Theorem 2.6.7 of van der Vaart and Wellner (1996), it follows that there exist universal constant $K_j > 0, j = 1, \ldots, m$ such that for any probability measure Q such that $\|F_j\|_{Q,2} > 0$ for $j = 1, \ldots, m$

$$N\big(\epsilon \;\; m \; \|\tilde{F}\|_{Q,2}, \mathcal{F}, L_2(Q)\big) \leq \prod_{j=1}^{m} K_j V_j \; (16e)^{\sum_{j=1}^{m} V_j} \left(\frac{1}{\epsilon}\right)^{2(\sum_{j=1}^{m} V_j - m)}$$

where $V_j \geq 1$ is an upper bound for the VC-index of the class $\mathcal{F}_j$. Now, note that $m\widetilde{F} = m(F_1 + F_2 + \ldots + F_m) \geq (F_1 + F_2 + \ldots + F_m)$ is also an envelope for $\mathcal{F}$. Using the notation of van der Vaart and Wellner (1996), set

$$J(\eta, \mathcal{F}) = \sup_{Q} \int_0^{\eta} \sqrt{1 + \log N(\epsilon m \|\widetilde{F}\|_{Q,2}, \mathcal{F}, L_2(Q))} d\epsilon$$

where the supremum is taken over all probability measures Q such that $\|F_j\|_{Q,2} > 0$. Set $K = \log(\prod_{j=1}^{m} K_j V_j \; (16e)^{\sum_{j=1}^{m} V_j})$ and $\gamma = 2(\sum_{j=1}^{m} V_j - m)$. Then, by the calculations above we have that

$$\begin{aligned} J(\eta, \mathcal{F}) \quad &\leq \quad \int_0^{\eta} \sqrt{1 + K + \gamma \log(1/\epsilon)} d\epsilon \\ &\leq \quad (1 + K) \vee \gamma \int_{1/\eta}^{\infty} \frac{\sqrt{1 + \log(x)}}{x^2} dx < \infty \end{aligned}$$

for any $\eta > 0$, in particular for $\eta = 1$. By Theorem 2.14.1 of van der Vaart and Wellner (1996), we conclude that there exists a universal constant $\tilde{M} > 0$ such that

$$E[\|\mathbb{G}_n\|_{\mathcal{F}}] \leq m \; \tilde{M} J(1, \mathcal{F}) \|\widetilde{F}\|_{P,2}, \text{ and } E[\|\mathbb{G}_n\|_{\mathcal{F}}^2] \leq m^2 \; \tilde{M}^2 J(1, \mathcal{F})^2 \|\widetilde{F}\|_{P,2}^2$$

and the result follows by taking $M = \max(\tilde{M}^2 J(1, \mathcal{F})^2, 1)$. Now, we show (3). Let $m_{\theta_t} - m_{\theta^0}$ be an element in $\mathcal{M}_{t,\delta}$. We can write

$$\begin{aligned} & (m_{\theta_t} - m_{\theta^0})(x) \\ & = \pi\big(1 - \mathbb{1}_{x \leq \mu_1 + t} - \mathbb{1}_{x \leq \mu_1 - t}\big) + (1 - \pi)\big(1 - \mathbb{1}_{x \leq \mu_2 + t} - \mathbb{1}_{x \leq \mu_2 - t}\big) \\ & - \pi^0\big(1 - \mathbb{1}_{x \leq \mu_1^0 + t} - \mathbb{1}_{x \leq \mu_1^0 - t}\big) + (1 - \pi^0)\big(1 - \mathbb{1}_{x \leq \mu_2^0 + t} - \mathbb{1}_{x \leq \mu_2^0 - t}\big) \\ = \quad & -\pi\big(\mathbb{1}_{x \leq \mu_1 + t} + \mathbb{1}_{x \leq \mu_1 - t}\big) - (1 - \pi)\big(\mathbb{1}_{x \leq \mu_2 + t} + \mathbb{1}_{x \leq \mu_2 - t}\big) \\ & + \pi^0\big(\mathbb{1}_{x \leq \mu_1^0 + t} + \mathbb{1}_{x \leq \mu_1^0 - t}\big) + (1 - \pi^0)\big(\mathbb{1}_{x \leq \mu_2^0 + t} + \mathbb{1}_{x \leq \mu_2^0 - t}\big) \\ = \quad & \pi^0\big(\mathbb{1}_{x \leq \mu_1^0 + t} - \mathbb{1}_{x \leq \mu_1 + t}\big) + \pi^0\big(\mathbb{1}_{x \leq \mu_1^0 - t} - \mathbb{1}_{x \leq \mu_1 - t}\big) \end{aligned}$$

$$
\begin{aligned}
& +(1-\pi^0)\big(\mathbb{1}_{x\le\mu_2^0+t}-\mathbb{1}_{x\le\mu_2+t}\big)+(1-\pi^0)\big(\mathbb{1}_{x\le\mu_2^0-t}-\mathbb{1}_{x\le\mu_2-t}\big)\\
& +(\pi-\pi^0)\big(\mathbb{1}_{x\le\mu_2+t}-\mathbb{1}_{x\le\mu_1+t}\big)+(\pi-\pi^0)\big(\mathbb{1}_{x\le\mu_2-t}-\mathbb{1}_{x\le\mu_1-t}\big)\\
= \quad & g_{1,t}(x)+\ldots+g_{6,t}(x)
\end{aligned}
$$

Then, it is clear that

$$
\mathcal{M}_{\delta,t}\subset\sum_{j=1}^{6}\mathcal{F}_j
$$

with $\mathcal{F}_j=\mathcal{I}$ for $j=1,\ldots,6$, where $\mathcal{I}$ is the class defined in the first statement of the theorem. On the other hand, we have that $\|g_{j,t}\|_\infty\le F_{j,t}, j=1,\ldots,6$ defined in (2.7). The claim follows from 2. □

Proof of Theorem 2.2. Fix $\delta\in(0,1]$. Before returning to the inequality in (2.6), we will find upper bounds for $\|F_{t,j}\|$ for $j=1,\ldots,6$. We have that

$$
\begin{aligned}
\|F_{t,1}\|_{P,2}^2 &= (\pi^0)^2\int_{\mu_1^0-\delta+t}^{\mu_1^0+\delta+t} f^0(x)dx\\
&= (\pi^0)^2\left\{\pi^0\int_{-\delta+t}^{\delta+t} g^0(x)dx+(1-\pi^0)\int_{\mu_1^0-\mu_2^0-\delta+t}^{\mu_1^0-\mu_2^0+\delta+t} g^0(x)dx\right\}\\
&\le \int_{t-\delta}^{t+\delta} g^0(x)dx+\int_{\mu_2^0-\mu_1^0-\delta-t}^{\mu_2^0-\mu_1^0+\delta-t} g^0(x)dx
\end{aligned}
$$

using symmetry of g^0 and $\max(\pi^0,1-\pi^0)\le 1$. Using the inequality $\sqrt{u+v}\le\sqrt{u}+\sqrt{v}$ for all $u,v\ge 0$, we have that

$$
\|F_{t,1}\|_{P,2}\le\left(\int_{t-\delta}^{t+\delta} g^0(x)dx\right)^{1/2}+\left(\int_{\mu_2^0-\mu_1^0-\delta-t}^{\mu_2^0-\mu_1^0+\delta-t} g^0(x)dx\right)^{1/2}.
$$

By assumption, g^0 changes direction of monotonicity only a finite number of times. This implies that for $\delta>0$ small enough

$$
g^0(x)\le\max\Big(g^0(t-\delta),g^0(t+\delta)\Big)\le g^0(t-\delta)+g^0(t+\delta),
$$

for all $x\in[t-\delta,t+\delta]$. To see this, use the fact that the length of the interval $[t-\delta,t+\delta]$ is 2δ and hence should be included in a bigger interval on which g^0 is either increasing or decreasing. In the first case, $g^0(x)\le g^0(t+\delta)$ wheres in the second $g^0(x)\le g^0(t-\delta)$. Hence, using again the inequality $\sqrt{u+v}\le\sqrt{u}+\sqrt{v}$ and integrability of $g_0^{1/2}$ we have that

$$
\begin{aligned}
\int_{\mathbb{R}}\left(\int_{t-\delta}^{t+\delta} g^0(x)dx\right)^{1/2}dt &\le (2\delta)^{1/2}\left(\int_{\mathbb{R}} g^0(t-\delta)^{1/2}dt+\int_{\mathbb{R}}(g^0(t+\delta)^{1/2}dt\right)\\
&= 2^{3/2}\delta^{1/2}\int_{\mathbb{R}}(g^0(x))^{1/2}dx.
\end{aligned}
$$

Similarly, we get

$$\int_{\mathbb{R}} \left(\int_{\mu_2^0-\mu_1^0-\delta-t}^{\mu_2^0-\mu_1^0+\delta-t} g^0(x)dx \right)^{1/2} \quad \leq \quad 2^{3/2}\delta^{1/2} \int_{\mathbb{R}} (g^0(x))^{1/2}dx.$$

The calculations above imply that

$$\int_{\mathbb{R}} \|F_{t,1}\|_{P,2}^j dt \leq D^j \delta^{j/2}, j = 1, 2$$

where $D > 0$ is a constant depending only on g^0.

In a very similar manner it can be shown that the same bounds apply for $\int_{\mathbb{R}} \|F_{t,i}\|_{P,2}^j dt$ for $i = 2, 3, 4$ and $j = 1, 2$. Now, we turn to $F_{t,5}$ and $F_{t,6}$. We have that

$$\begin{aligned} \|F_{t,5}\|_{P,2}^2 &= \delta^2 \int_{\mu_1^0-\delta+t}^{\mu_2^0+\delta+t} f^0(x)dx \\ &\leq \delta^2 \left(\int_{-\delta+t}^{\mu_2^0-\mu_1^0+\delta+t} g^0(x)dx + \int_{\mu_1^0-\mu_2^0-\delta+t}^{\delta+t} g^0(x)dx \right). \quad (4.12) \end{aligned}$$

The assumption that g^0 changes direction of monotonicity only a finite number of time implies that we can find $-\infty < A < B < \infty$ such that g^0 is increasing on $(-\infty, A)$ and decreasing on (B, ∞). Furthermore, we know also that there exists $M > 0$ such that $g^0 \leq M$ on $[A, B]$. Hence,

$$\begin{aligned} &\int_{\mathbb{R}} \left(\int_{-\delta+t}^{\mu_2^0-\mu_1^0+\delta+t} g^0(x)dx \right)^{1/2} \\ &= \int_{-\infty}^{A-\delta-(\mu_2^0-\mu_1^0)} \left(\int_{-\delta+t}^{\mu_2^0-\mu_1^0+\delta+t} g^0(x)dx \right)^{1/2} dt \\ &\quad + \int_{A-\delta-(\mu_2^0-\mu_1^0)}^{B+\delta} \left(\int_{-\delta+t}^{\mu_2^0-\mu_1^0+\delta+t} g^0(x)dx \right)^{1/2} dt \\ &+ \int_{B+\delta}^{\infty} \left(\int_{-\delta+t}^{\mu_2^0-\mu_1^0+\delta+t} g^0(x)dx \right)^{1/2} dt \\ &\leq (\mu_2^0 - \mu_1^0 + 2\delta)^{1/2} \int_{-\infty}^{A-\delta-(\mu_2^0-\mu_1^0)} (g^0(\mu_2^0 - \mu_1^0 + \delta + t))^{1/2} dt \\ &\quad + (B - A + 2\delta + \mu_2^0 - \mu_1^0)(\mu_2^0 - \mu_1^0 + 2\delta)^{1/2} M^{1/2} \\ &\quad + (\mu_2^0 - \mu_1^0 + 2\delta)^{1/2} \int_{B+\delta}^{\infty} (g^0(-\delta + t))^{1/2} dt \\ &\leq (\mu_2^0 - \mu_1^0 + 2\delta)^{1/2} \int_{-\infty}^{A} (g^0(x))^{1/2} dx \end{aligned}$$

$$
\begin{aligned}
&+(B-A+2\delta+\mu_2^0-\mu_1^0)(\mu_2^0-\mu_1^0+2\delta)^{1/2}M^{1/2}\\
&+(\mu_2^0-\mu_1^0+2\delta)^{1/2}\int_B^\infty (g^0(x))^{1/2}dx\\
\le\ & \tilde{D}_1
\end{aligned}
$$

where $\tilde{D}_1>0$ depends only on g^0 and $\mu_2^0-\mu_1^0$. As the second term in (4.12) can be handled similarly it follows that there exists a constant $\tilde{D}>0$ depending on g^0 and $\mu_2^0-\mu_1^0$ only such that

$$\int_{\mathbb{R}} \|F_{t,5}\|_{P,2}^j dt \le \tilde{D}^j\delta^j$$

for $j=1,2$. A similar bound applies for $\int_{\mathbb{R}}\|F_{t,6}\|_{P,2}^j dt$.

Now, we turn to the inequality in (2.6). It follows from the calculations above, that for any $t\in\mathbb{R}$ the class $\mathcal{M}_{t,\delta}$ is VC of index at most 4 and with a finite envelope. It follows from Theorem 2.3, the c_r-inequality and the bounds obtained above that there exists a constant $C_1>0$ depending only on g^0 such that

$$\int_{\mathbb{R}} E[\|\sqrt{n}(\mathbb{P}_n-\mathbb{P})\|_{\mathcal{M}_{t,\delta}}^2]dt = \int_{\mathbb{R}} E[\|\mathbb{G}_n\|_{\mathcal{M}_{t,\delta}}^2]dt \le C_1\delta$$

for all $\delta\in(0,1]$. This in turn implies a bound on the maximal expectation of the first term in the right side of (2.6):

$$
\begin{aligned}
\sqrt{n}\int_{\mathbb{R}} E\left[\sup_{m_{\theta_t}-m_{\theta_t^0}\in\mathcal{M}_{\delta,t}}(\mathbb{P}_n-\mathbb{P})^2(m_{\theta_t}-m_{\theta_t^0})\right]dt &= \frac{1}{\sqrt{n}}\int_{\mathbb{R}} E[\|\mathbb{G}_n\|_{\mathcal{M}_{t,\delta}}^2]dt\\
&\le \frac{C_1\delta}{\sqrt{n}}. \qquad (4.13)
\end{aligned}
$$

To tackle the second term, note that boundedness of g^0 implies that there exists a constant $C'>0$ depending only on $g^0(0)$ and $\mu_2^0-\mu_1^0$ such that

$$\mathbb{P}|m_{\theta_t}-m_{\theta_t^0}|\le C'\delta$$

for all $\delta\in(0,1]$ and $t\in\mathbb{R}$. Indeed, we have that

$$\mathbb{P}|m_{\theta_t}-m_{\theta_t^0}|\le \|F_{t,1}\|_{P,1}+\ldots+\|F_{t,6}\|_{P,1}$$

where similar calculations to the those developed above show that

$$
\begin{aligned}
\|F_{t,1}\|_{P,1} &\le \int_{t-\delta}^{t+\delta} g^0(x)dx + \int_{\mu_2^0-\mu_1^0-\delta-t}^{\mu_2^0-\mu_1^0+\delta-t} g^0(x)dx \qquad (4.14)\\
&\le 4\delta g^0(0)
\end{aligned}
$$

and a similar bound applies for $\|F_{t,2}\|_{P,1}, \|F_{t,3}\|_{P,1}$ and $\|F_{t,4}\|_{P,1}$. Also,

$$\|F_{t,5}\|_{P,1} \le \delta\left(\int_{t-\delta}^{\mu_2^0-\mu_1^0+\delta+t} g^0(x)dx + \int_{\mu_1^0-\mu_2^0+t-\delta}^{t+\delta} g^0(x)dx\right), \quad (4.15)$$

$$\leq \quad 2\delta(\mu_2^0 - \mu_1^0 + 2\delta)g^0(0)$$

and a similar bound applies for $\|F_{t,6}\|_{P,1}$. It follows that there exists $C' > 0$ depending only on g^0 and $\mu_2^0 - \mu_1^0$ such that

$$\begin{aligned}
&\int_{\mathbb{R}} E\left[\sqrt{n} \sup_{m_{\theta_t} - m_{\theta_t^0} \in \mathcal{M}_{\delta,t}} \left|(\mathbb{P}_n - \mathbb{P})(m_{\theta_t} - m_{\theta_t^0})\right|\right] \mathbb{P}\left|m_{\theta_t} - m_{\theta_t^0}\right| dt \\
&\leq C'\delta \int_{\mathbb{R}} E[\|\mathbb{G}_n\|_{\mathcal{M}_{t,\delta}}] dt \\
&\leq C_2 \delta^{3/2}. \qquad (4.16)
\end{aligned}$$

To bound the maximal expectation of the third term in (2.6), we use the Cauchy-Schwarz inequality:

$$\begin{aligned}
&E\left[\int_{\mathbb{R}} \sqrt{n} \sup_{m_{\theta_t} - m_{\theta_t^0} \in \mathcal{M}_{\delta,t}} \left|(\mathbb{P}_n - \mathbb{P})(m_{\theta_t} - m_{\theta_t^0})\right| \left|(\mathbb{P}_n - \mathbb{P})m_{\theta_t^0}\right| dt\right] \\
&\leq \int_{\mathbb{R}} E[\|\mathbb{G}_n\|^2_{\mathcal{M}_{t,\delta}}]^{1/2} \; E\left[\left((\mathbb{P}_n - \mathbb{P})m_{\theta_t^0}\right)^2\right]^{1/2} dt \\
&= \frac{1}{\sqrt{n}} \int_{\mathbb{R}} E[\|\mathbb{G}_n\|^2_{\mathcal{M}_{t,\delta}}]^{1/2} \;\; (\mathbb{P}m^2_{\theta_t^0})^{1/2} dt, \text{ since } \mathbb{P}m_{\theta_t^0} = 0 \text{ by } (1.4) \\
&\leq \frac{2C_1^{1/2}\delta^{1/2}}{\sqrt{n}} = \frac{C_3\delta^{1/2}}{\sqrt{n}}
\end{aligned}$$

using the fact that $m^2_{\theta_t^0} \leq 4$ and where C_1 is the same constant in (4.13). Finally, the fourth term can be bounded using the known inequality $E[|X|] \leq E[|X^2|]^{1/2}$ for any random variable X combined with the already used fact that $\mathbb{P}m^2_{\theta_t^0} \leq 4$ and that there exists a constant $\tilde{C}' > 0$ depending only on g^0 and $\mu_2^0 - \mu_1^0$ such that

$$\int_{\mathbb{R}} \|F_{t,j}\|_{P,1} dt \leq \tilde{C}'\delta.$$

To see this for $j = 1$ and 5 for example, we can use the inequalities in (4.14) and (4.15) combined with monotonicity of g^0 as done above. The same holds for $j = 2, 3, 4$ and 6. Thus, we obtain

$$\begin{aligned}
&E\left[\int_{\mathbb{R}} \left|\sqrt{n}(\mathbb{P}_n - \mathbb{P})m_{\theta_t^0}\right| \mathbb{P}|m_{\theta_t} - m_{\theta_t^0}| dt\right] \\
&\leq \int_{\mathbb{R}} E[(\sqrt{n}(\mathbb{P}_n - \mathbb{P})m_{\theta_t^0})^2]^{1/2} \mathbb{P}|m_{\theta_t} - m_{\theta_t^0}| dt \\
&\leq 2\int_{\mathbb{R}} \sum_{j=1}^{6} \|F_{t,j}\|_{P,1} dt
\end{aligned}$$

$$\leq C_4\delta, \text{ with } C_4 = 2\tilde{C}'.$$

for a constant $C_4 > 0$ depending only on g^0. Using the fact that $\delta^{3/2} \leq \delta$ and $\delta \leq \delta^{1/2}$ for all $\delta \in (0,1]$ completes the proof of the theorem. □

Proof of Proposition 2.4. It follows from Theorem 1 of Hunter et al. (2007) that θ^0 is the only element $\theta \in \Theta = \{(\pi, u_1, u_2) : \pi \in [0,1], -\infty < u_1 < u_2 < \infty\}$ such that $\mathbb{D}(\theta) = 0$. Therefore, θ^0 is the unique minimizer of $\mathbb{D}$ over the previous set. Now the inequality of the proposition is guaranteed if the map $\theta \mapsto \mathbb{D}^2(\theta)$ is twice continuously differentiable on small neighborhood of θ^0. In the sequel we will write F, f, G and g for F^0, f^0, G^0 and g^0 respectively. Also, the integrand in $\mathbb{D}^2(\theta)$ will be denoted by

$$\begin{aligned}\psi(\pi, \mu_1, \mu_2, t) \quad = \quad & \Big\{\pi(1 - F(\mu_1 - t) - F(\mu_1 + t)) \\ & +(1-\pi)(1 - F(\mu_2 - t) - F(\mu_2 + t))\Big\}^2.\end{aligned}$$

Also, let $\pi \in (\pi^0 - \delta, \pi^0 + \delta), \mu_1 \in (\mu_1^0 - \delta, \mu_1^0 + \delta), \mu_2 \in (\mu_2^0 - \delta, \mu_2^0 + \delta)$ for some $\delta > 0$ smal enough such that $\delta \leq (\mu_2^0 - \mu_1^0)/2$. Now,

$$\begin{aligned}&\frac{\partial \psi}{\partial \pi} \\ &= 2\Big\{F(\mu_2 - t) + F(\mu_2 + t) - F(\mu_1 - t) - F(\mu_1 + t)\Big\} \\ &\times \Big\{\pi(1 - F(\mu_1 - t) - F(\mu_1 + t)) + (1-\pi)(1 - F(\mu_2 - t) - F(\mu_2 + t))\Big\}. \\ \left|\frac{\partial \psi}{\partial \pi}\right| &\leq 4\Big\{|F(\mu_2 - t) - F(\mu_1 - t)| + |F(\mu_2 + t) - F(\mu_1 + t)|\Big\} \\ &\leq 4\Big\{F(\mu_2^0 + \delta - t) - F(\mu_1^0 - \delta - t) + F(\mu_2^0 + \delta + t) - F(\mu_1^0 - \delta + t)\Big\}\end{aligned}$$

which is integrable since $\int_0^\infty (1 - F(x))dx < \infty$ and $\int_{-\infty}^0 F(t)dt < \infty$, an easy consequence of integrability of $X \sim F$. By the Lebesgue dominated convergence theorem, it follows that $\mathbb{D}^2$ admits a continuous first partial derivative with respect to π in $(\pi^0 - \delta, \pi^0 + \delta)$ and

$$\frac{\partial \mathbb{D}^2}{\partial \pi} = \int_{\mathbb{R}} \frac{\partial \psi}{\partial \pi} dt.$$

Computing the second partial derivative yields

$$\begin{aligned}\frac{\partial^2 \psi}{\partial \pi^2} \quad &= \quad 2\Big\{F(\mu_2 - t) + F(\mu_2 + t) - F(\mu_1 - t) - F(\mu_1 + t)\Big\}^2 \\ &\leq \quad 4\Big\{(F(\mu_2 - t) - F(\mu_1 - t))^2 + (F(\mu_2 + t) - F(\mu_1 + t))^2\Big\} \\ &\leq \quad 4\Big\{F(\mu_2^0 + \delta - t) - F(\mu_1^0 - \delta - t) + F(\mu_2^0 + \delta + t) - F(\mu_1^0 - \delta + t)\Big\}.\end{aligned}$$

Similar arguments show that $\mathbb{D}^2$ admits a continuous second partial derivative with respect to π in $(\pi^0 - \delta, \pi^0 + \delta)$ with

$$\frac{\partial^2 \mathbb{D}^2}{\partial \pi^2} = \int_{\mathbb{R}} 2\Big\{F(\mu_2 - t) + F(\mu_2 + t) - F(\mu_1 - t) - F(\mu_1 + t)\Big\}^2 dt.$$

Now, the partial derivative of ψ with respect to μ_1 is given by

$$\begin{aligned}
&\frac{\partial \psi}{\partial \mu_1} \\
&= -2\pi\Big(f(\mu_1 - t) + f(\mu_1 + t)\Big) \\
&\quad \times\Big\{\pi(1 - F(\mu_1 - t) - F(\mu_1 + t)) + (1 - \pi)(1 - F(\mu_2 - t) - F(\mu_2 + t))\Big\}
\end{aligned}$$

which implies that

$$\left|\frac{\partial \psi}{\partial \mu_1}\right| \quad \leq \quad 4\Big(f(\mu_1 - t) + f(\mu_1 + t)\Big)$$

where

$$f(\mu_1 - t) = \pi^0 g(\mu_1 - \mu_1^0 - t) + (1 - \pi^0) g(\mu_1 - \mu_2^0 - t),$$

and

$$f(\mu_1 + t) = \pi^0 g(\mu_1 - \mu_1^0 + t) + (1 - \pi^0) g(\mu_1 - \mu_2^0 + t).$$

If we focus on the function $t \mapsto g(\mu_1 - \mu_1^0 - t)$, we see that $-t - \delta \leq \mu_1 - \mu_1^0 - t \leq \delta - t$. Hence using the assumption that g changes monotonicity only a finite number of times, we have that for δ small enough,

$$g(\mu_1 - \mu_1^0 - t) \leq \max\Big(g(-t - \delta), g(-t + \delta)\Big) \leq g(-t - \delta) + g(-t + \delta)$$

which is integrable. We can in a similar manner bound the remaining functions. We conclude that $f(\mu_1 - t) + f(\mu_1 + t)$ is bounded above by a nonnegative and integrable function. It follows that $\mathbb{D}^2$ admits a continuous first partial derivative with respect to μ_1 in $(\mu_1^0 - \delta, \mu_1^0 + \delta)$. Furthermore,

$$\begin{aligned}
\frac{\partial^2 \psi}{\partial \mu_1^2} &= -2\pi\Big(f'(\mu_1 - t) + f'(\mu_1 + t)\Big) \\
&\quad \times \Big\{\pi(1 - F(\mu_1 - t) - F(\mu_1 + t)) + (1 - \pi)(1 - F(\mu_2 - t) - F(\mu_2 + t))\Big\} \\
&\quad + 4\pi^2\Big(f(\mu_1 - t) + f(\mu_1 + t)\Big)^2.
\end{aligned}$$

Now, there exists $M > 0$ such that

$$\Big| \; - 2\pi\Big(f'(\mu_1 - t) + f'(\mu_1 + t)\Big)$$

$$
\begin{aligned}
&\times \Big\{\pi(1-F(\mu_1-t)-F(\mu_1+t))+(1-\pi)(1-F(\mu_2-t)-F(\mu_2+t))\Big\}\Big| \\
&\le M\Big\{\Big(1-F(\mu_1^0-\delta-t)+F(\mu_1^0+\delta+t)\Big)\mathbb{1}_{t\le 0} \\
&\quad+\Big(1-F(\mu_1^0-\delta+t)+F(\mu_1^0+\delta-t)\Big)\mathbb{1}_{t>0}\Big\} \\
&\quad+M\Big\{\Big(1-F(\mu_2^0-\delta-t)+F(\mu_2^0+\delta+t)\Big)\mathbb{1}_{t\le 0} \\
&\quad+\Big(1-F(\mu_2^0-\delta+t)+F(\mu_2^0+\delta-t)\Big)\mathbb{1}_{t>0}\Big\}.
\end{aligned}
$$

Above we used the fact that f' is bounded. The function on the right side is integrable since any real a we have that $\int_0^\infty(1-F(a+|t|))dt<\infty$ and $\int_{\mathbb{R}}F(a-|t|)dt<\infty$, a consequence of integrability of $X\sim F$ (implied by integrability of $Y\sim G$). Also, we have that

$$
4\pi^2\Big(f(\mu_1-t)+f(\mu_1+t)\Big)^2\le 8f(0)\Big(f(\mu_1-t)+f(\mu_1+t)\Big)
$$

which we have already shown to be bounded above by an integrable function. We conclude that $\mathbb{D}^2$ admits a continuous second partial derivative with respect to μ_1 in $(\mu_1^0-\delta,\mu_1^0+\delta)$ with

$$
\begin{aligned}
&\frac{\partial^2\mathbb{D}^2}{\partial(\mu_1^0)^2} \\
&=\int_{\mathbb{R}}2\pi\Big(f'(\mu_1-t)+f'(\mu_1+t)\Big) \\
&\times\Big\{\pi(1-F(\mu_1-t)-F(\mu_1+t))+(1-\pi)(1-F(\mu_2-t)-F(\mu_2+t))\Big\}dt \\
&\quad+\ 4\pi^2\int_{\mathbb{R}}\Big(f(\mu_1-t)+f(\mu_1+t)\Big)^2dt.
\end{aligned}
$$

Similar arguments can be used to show that $\mathbb{D}^2$ admits a continuous second partial derivative with respect to μ_2 in $(\mu_2^0-\delta,\mu_2^0+\delta)$ with

$$
\begin{aligned}
&\frac{\partial^2\mathbb{D}^2}{\partial(\mu_2^0)^2} \\
&=\int_{\mathbb{R}}2(1-\pi)\Big(f'(\mu_2-t)+f'(\mu_2+t)\Big) \\
&\times\Big\{\pi(1-F(\mu_1-t)-F(\mu_1+t))+(1-\pi)(1-F(\mu_2-t)-F(\mu_2+t))\Big\}dt \\
&\quad+4(1-\pi)^2\ \int_{\mathbb{R}}\Big(f(\mu_1-t)+f(\mu_1+t)\Big)^2dt.
\end{aligned}
$$

We compute the crossed partial derivative of ψ with respect to π and μ_1:

$$
\begin{aligned}
&\frac{\partial^2\psi}{\partial\mu_1\partial\pi} \\
&=-2\Big\{f(\mu_1-t)+f(\mu_1+t)\Big\}
\end{aligned}
$$

$$\times\Big(\pi(1-F(\mu_1-t)-F(\mu_1+t))+(1-\pi)(1-F(\mu_2-t)-F(\mu_2+t))\Big)$$
$$-2\pi\Big\{f(\mu_1-t)+f(\mu_1+t)\Big\}$$
$$\times\Big(F(\mu_2-t)+F(\mu_2+t)-F(\mu_1-t)-F(\mu_1+t)\Big).$$

so that

$$\left|\frac{\partial^2\psi}{\partial\mu_1\partial\pi}\right|\leq 8\Big\{f(\mu_1-t)+f(\mu_1+t)\Big\}$$

and we have already shown above that the function on the right side is bounded by an integrable function. Similar arguments can be used for μ_2. Finally, we calculate

$$\frac{\partial^2\psi}{\partial\mu_2\partial\mu_1}=2\pi(1-\pi)\Big(f(\mu_1-t)+f(\mu_1+t)\Big)\Big(f(\mu_2-t)+f(\mu_2+t)\Big)$$

which can be bounded by an integrable function by bounding each one of the terms $f(\mu_1-t)+f(\mu_1+t)$ and $f(\mu_2-t)+f(\mu_2+t)$. We conclude that $\mathbb{D}^2$ is twice continuously differentiable on a small neighborhood of θ^0. Since θ^0 is the unique minimum of $\mathbb{D}^2$, it follows that the quadratic form

$$\begin{aligned}
&Q(h_1,h_2,h_3)\\
&=\frac{\partial^2\mathbb{D}^2}{\partial\pi^2}(\theta^0)h_1^2+2\frac{\partial^2\mathbb{D}^2}{\partial\pi\partial\mu_1}(\theta^0)h_1h_2+2\frac{\partial^2\mathbb{D}^2}{\partial\pi\partial\mu_2}(\theta^0)h_1h_3+2\frac{\partial^2\mathbb{D}^2}{\partial\mu_1\partial\mu_2}(\theta^0)h_2h_3\\
&\quad+\frac{\partial^2\mathbb{D}^2}{\partial(\mu_1)^2}(\theta^0)h_2^2+\frac{\partial^2\mathbb{D}^2}{\partial(\mu_2)^2}(\theta^0)h_3^2\\
&=hBh^T>0,
\end{aligned}$$

with $h=(h_1,h_2,h_3)$ such that h_1,h_2,h_3 are small enough and $h\neq(0,0,0)$. Here B is the 3×3 matrix of second partial derivatives of $\mathbb{D}^2$ evaluated at θ^0. This implies that B has to be positive definite. Let λ_0 be the smallest eigenvalue of B with $\lambda_0>0$. Then, there exists a small neighborhood of θ^0 such that

$$\begin{aligned}
\mathbb{D}^2(\theta)-\mathbb{D}^2(\theta^0) &= (\theta-\theta^0)B(\theta-\theta_0)^T+o(\|\theta-\theta_0\|_2^2)\\
&\quad\text{where } \|\cdot\|_2 \text{ denotes the } L_2\text{-norm in } \mathbb{R}^3\\
&\geq \lambda_0\|\theta-\theta^0\|_2+o(\|\theta-\theta^0\|_2^2)\\
&\geq \frac{\lambda_0}{2}\|\theta-\theta^0\|_2^2\geq\frac{\lambda_0}{2}\|\theta-\theta^0\|_\infty^2.
\end{aligned}$$

Taking $\kappa=\lambda_0/2$ completes the proof. □

Proof of Proposition 2.5. Using the notation of van der Vaart and Wellner (1996), we write $\mathbb{M}=-\mathbb{D}^2$ and $\mathbb{M}_n=-\mathbb{D}_n^2$. Then, the first two requirements of their Theorem 3.2.5 are clearly satisfied. Furthermore, we have

$$\frac{\phi_n(\delta)}{\delta}=1+\frac{\delta^{-1/2}}{\sqrt{n}}$$

is decreasing (and so the power α in their theorem is equal to 1). Also, if $r_n = \sqrt{n}/2$, then

$$\begin{aligned} r_n^2 \phi_n\left(\frac{1}{r_n}\right) &= \frac{n}{4}\left(\frac{2}{\sqrt{n}} + \frac{1}{n^{3/4}}\right) \\ &\leq \frac{n}{4}\frac{4}{\sqrt{n}}, \quad \text{for } n \text{ large enough} \\ &= \sqrt{n} \end{aligned}$$

and the result follows. □

Supplementary Material

Supplement A: Proofs of Theorem 2.6 and 2.7 and some useful formulae

(doi: 10.1214/17-EJS1311SUPP; .pdf). In this supplementary file we provide proofs of Theorem 2.6 and 2.7 describing the weak limiting distribution of the estimator of Hunter et al. (2007). Some formulae used in the derivation of this limit are also given.

References

Balabdaoui, F. (2017). Supplement A: Proofs of Theorem 2.6 and 2.7 and some useful formulae. *Electronic Journal of Statistics*. doi: http://dx.doi.org/10.1214/17-EJS1311SUPP.

Balabdaoui, F. and Doss, C. (2016). Inference for a two-component mixture of symmetric distributions under log-concavity. *Bernoulli* to appear. MR3706787

Bordes, L., Mottelet, S. and Vandekerkhove, P. (2006). Semiparametric estimation of a two-component mixture model. *Ann. Statist.* **34** 1204–1232. MR2278356

Butucea, C. and Vandekerkhove, P. (2014). Semiparametric mixtures of symmetric distributions. *The Scandinavian Journal of Statistics* **41** 227–239. MR3181141

Chang, G. T. and Walther, G. (2007). Clustering with mixtures of log-concave distributions. *Computational Statistics & Data Analysis* **51** 6242–6251. MR2408591

Chee, C.-S. and Wang, Y. (2013). Estimation of finite mixtures with symmetric components. *Stat. Comput.* **23** 233–249. URL http://dx.doi.org/10.1007/s11222-011-9305-5 MR3016941

Cule, M., Samworth, R. and Stewart, M. (2010). Maximum likelihood estimation of a multidimensional log-concave density. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** 545–607. MR2758237

Finner, H. and Roters, M. (1993). Distribution functions and log-concavity. *Comm. Statist. Theory Methods* **22** 2381–2396. URL http://dx.doi.org/10.1080/03610929308831156 MR1238175

Hunter, D. R., Wang, S. and Hettmansperger, T. P. (2007). Inference for mixtures of symmetric distributions. *Ann. Statist.* **35** 224–251. URL http://dx.doi.org/10.1214/009053606000001118 MR2332275

Ma, Y. and Yao, W. (2015). Flexible estimation of a semiparametric two-component mixture model with one parametric component. *Electron. J. Stat.* **9** 444–474. URL http://dx.doi.org/10.1214/15-EJS1008 MR3326131

Schoenberg, I. J. (1951). On Pólya frequency functions. I. The totally positive functions and their Laplace transforms. *J. Analyse Math.* **1** 331–374. MR0047732

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