

**SUPPLEMENT TO 'STRONG IDENTIFIABILITY AND
OPTIMAL MINIMAX RATES FOR FINITE MIXTURE
ESTIMATION'**

BY PHILIPPE HEINRICH AND JONAS KAHN

*Université Lille 1
Laboratoire Paul Painlevé Bât. M2
Cité Scientifique
59655 Villeneuve d'Ascq, FRANCE*

APPENDIX A: TAYLOR EXPANSIONS AND L^P -CONVERGENCES IN
THE PROOF OF THEOREM 6.1

A.1. Proof of Step 1: the $Z_{i,n}$'s are centered. Recall that

$$Z_{i,n} = \pi_0 \frac{f^{(2d-1)}(X_{i,n}, \theta_0)}{f(X_{i,n}, G_n(0))}$$

and set for short $\theta_{j,n}(u) = \theta_0 + \varepsilon_n h_j(u)$ with $\varepsilon_n = n^{-1/(4d-2)}$. By Taylor expansion with remainder,

$$\begin{aligned} f(x, \theta_{j,n}(u)) - f(x, \theta_0) &= \sum_{k=1}^{2d-1} (\varepsilon_n h_j(u))^k f^{(k)}(x, \theta_0) \\ &\quad + \int_{\theta_0}^{\theta_{j,n}(u)} f^{(2d)}(x, \theta) \frac{(\theta_{j,n}(u) - \theta)^{2d-1}}{(2d-1)!} d\theta. \end{aligned}$$

By linearity of $G \mapsto f(\cdot, G)$, the definition of the mixing distributions $G_n(u)$ and G_0 , and the definition (14) of the moments μ_k , we get

$$\begin{aligned} f(x, G_n(u)) - f(x, G_0) &= \pi_0 \sum_{j=m_0}^{\mathbf{m}} \pi_j(u) [f(x, \theta_{j,n}(u)) - f(x, \theta_0)] \\ \text{(A.1)} \qquad \qquad \qquad &= \pi_0 \left[\sum_{k=1}^{2d-1} \mu_k \varepsilon_n^k f^{(k)}(x, \theta_0) + r_n(x, u) \right], \end{aligned}$$

with

$$r_n(x, u) = \sum_{j=m_0}^{\mathbf{m}} \pi_j(u) \int_{\theta_0}^{\theta_{j,n}(u)} f^{(2d)}(x, \theta) \frac{(\theta_{j,n}(u) - \theta)^{2d-1}}{(2d-1)!} d\theta.$$

Since the moments μ_1, \dots, μ_{2d-2} that do not depend on u but $\mu_{2d-1} = u$, subtracting (A.1) with $u = 0$ from (A.1) yields

$$f(x, G_n(u)) - f(x, G_n(0)) = \pi_0 \left[\frac{uf^{(2d-1)}(x, \theta_0)}{n^{1/2}} + r_n(x, u) - r_n(x, 0) \right].$$

Dividing by $f(x, G_n(0))$, we get with $x = X_{i,n}$,

$$Y_{i,n}(u) = \frac{uZ_{i,n}}{n^{1/2}} + R_{i,n}(u) - R_{i,n}(0)$$

with

$$Z_{i,n} = \pi_0 \frac{f^{(2d-1)}(X_{i,n}, \theta_0)}{f(X_{i,n}, G_n(0))} \quad \text{and} \quad R_{i,n}(u) = \pi_0 \frac{r_n(X_{i,n}, u)}{f(X_{i,n}, G_n(0))}.$$

Now, we can show that the i.i.d. vectors $(Y_{i,n}(u), Z_{i,n}, R_{i,n}(u))$ are centered under $G_n(0)$, for each fixed n and u . Indeed, first, we immediately have

$$\mathbb{E}_{G_n(0)} Y_{i,n}(u) = \int [f(x, G_n(u)) - f(x, G_n(0))] d\lambda(x) = 0.$$

Next, by expanding $f(X_{1,n}, \theta)$ around θ_0 , dividing by $f(X_{1,n}, G_n(0))$, taking expectations and applying Fubini Theorem to the remainder, we get

$$\begin{aligned} 0 &= \sum_{k=1}^{2d-1} \frac{h^k}{k!} \mathbb{E}_{G_n(0)} \left[\frac{f^{(k)}(X_{1,n}, \theta_0)}{f(X_{1,n}, G_n(0))} \right] \\ &\quad + \int_{\theta_0}^{\theta_0+h} \frac{(\theta_0 + h - \theta)^{2d-1}}{(2d-1)!} \mathbb{E}_{G_n(0)} \left[\frac{f^{(2d)}(X_{1,n}, \theta)}{f(X_{1,n}, G_n(0))} \right] d\theta; \end{aligned}$$

Proposition E.1 below ensures that each expectation exists, that Fubini Theorem is valid and that the remainder term is of order h^{2d} . Thus, we deduce iteratively that for $k \in \llbracket 1, 2d-1 \rrbracket$

$$\mathbb{E}_{G_n(0)} \left[\frac{f^{(k)}(X_{1,n}, \theta_0)}{f(X_{1,n}, G_n(0))} \right] = 0$$

and in particular $\mathbb{E}_{G_n(0)}(Z_{1,n}) = 0$. Dividing (A.1) by $f(x, G_n(0))$ and taking $x = X_{1,n}$ gives in addition $\mathbb{E}_{G_n(0)}(R_{1,n}(u)) = 0$.

A.2. Proof of Step 2: Z_n is asymptotically $\mathcal{N}(0, 1)$. Recall that

$$Z_n = (n\Gamma_n)^{-1/2} \sum_{i=1}^n Z_{i,n} \quad \text{with} \quad Z_{i,n} = \pi_0 \frac{f^{(2d-1)}(X_{i,n}, \theta_0)}{f(X_{i,n}, G_n(0))}.$$

Proposition E.1 implies that $\Gamma_n = \mathbb{E}_{G_n(0)} |Z_{1,n}|^2 \asymp 1$ and $\mathbb{E}_{G_n(0)} |Z_{1,n}|^3 \asymp 1$. As a consequence, Lyapunov condition holds:

$$\frac{\sum_{i=1}^n \mathbb{E}_{G_n(0)} |Z_{i,n}|^3}{\left[\sum_{i=1}^n \mathbb{E}_{G_n(0)} |Z_{i,n}|^2\right]^{3/2}} = \frac{n \mathbb{E}_{G_n(0)} |Z_{1,n}|^3}{\left[n \mathbb{E}_{G_n(0)} |Z_{1,n}|^2\right]^{3/2}} \asymp \frac{1}{\sqrt{n}} \xrightarrow{n \rightarrow \infty} 0.$$

By ?, Theorem 27.3, Z_n is asymptotically $\mathcal{N}(0, 1)$.

A.3. Proof of Step 3: L^p -convergences of $A_n(u)$, $B_n(u)$, $C_n(u)$.

Recall the quantities:

$$(A.2) \quad r_n(x, u) = \sum_{j=m_0}^{\mathbf{m}} \pi_j(u) \int_{\theta_0}^{\theta_{j,n}(u)} f^{(2d)}(x, \theta) \frac{(\theta_{j,n}(u) - \theta)^{2d-1}}{(2d-1)!} d\theta,$$

$$(A.3) \quad Y_{i,n}(u) = un^{-1/2} Z_{i,n} + R_{i,n}(u) - R_{i,n}(0),$$

$$(A.4) \quad Z_n = n^{-1/2} \Gamma_n^{-1/2} \sum_{i=1}^n Z_{i,n}.$$

Recall also in the following computations that for each fixed n and u , the i.i.d. vectors $(Y_{i,n}(u), Z_{i,n}, R_{i,n}(u))$ are centered under $G_n(0)$.

Convergence of $A_n(u)$. We first show that

$$A_n(u) = \sum_{i=1}^n Y_{i,n}(u) - uZ_n \sqrt{\Gamma_n} \xrightarrow{L^2} 0.$$

From (A.3) and (A.4), we can write $A_n(u) = \sum_{i=1}^n (R_{i,n}(u) - R_{i,n}(0))$, so that it's enough to prove that for each u ,

$$(A.5) \quad \mathbb{E}_{G_n(0)} \left| \sum_{i=1}^n R_{i,n}(u) \right|^2 = n \mathbb{E}_{G_n(0)} |R_{1,n}(u)|^2 \xrightarrow{n \rightarrow \infty} 0.$$

To this end, we look at the remainder (A.2) for fixed u : for any θ in the integrand, any j and n , we have $|\theta_{j,n}(u) - \theta| \leq |h_j(u)| \varepsilon_n$. Using (15) for

$U = u$, we may then write

$$\begin{aligned} |r_n(x, u)| &\leq \sum_{j=m_0}^m \pi_j(u) \int_{\theta_0 - \mathbf{h}(u)\varepsilon_n}^{\theta_0 + \mathbf{h}(u)\varepsilon_n} \left| f^{(2d)}(x, \theta) \right| \frac{\mathbf{h}(u)^{2d-1} n^{-1/2}}{(2d-1)!} d\theta \\ &\stackrel{u}{\lesssim} n^{-1/2} \int_{|\theta - \theta_0| \stackrel{u}{\lesssim} \varepsilon_n} \left| f^{(2d)}(x, \theta) \right| d\theta. \end{aligned}$$

Then, for $\alpha \in \llbracket 1, 4 \rrbracket$, using the convexity of $x \mapsto x^\alpha$ and Hölder's inequality, we may write

$$\begin{aligned} |R_{1,n}(u)|^\alpha &= \left| \pi_0 \frac{r_n(\cdot, u)}{f(\cdot, G_n(0))} \right|^\alpha \\ &\stackrel{u}{\lesssim} n^{-\alpha/2} \left| \int_{|\theta - \theta_0| \stackrel{u}{\lesssim} \varepsilon_n} \frac{|f^{(2d)}(\cdot, \theta)|}{f(\cdot, G_n(0))} d\theta \right|^\alpha \\ &\stackrel{u}{\lesssim} n^{-\alpha/2} \varepsilon_n^{\alpha-1} \int_{|\theta - \theta_0| \stackrel{u}{\lesssim} \varepsilon_n} \left| \frac{f^{(2d)}(\cdot, \theta)}{f(\cdot, G_n(0))} \right|^\alpha d\theta. \end{aligned}$$

Taking expectations w.r.t. $G_n(0)$, we may use Fubini theorem in the estimation above. Since moreover θ in the integrand is between θ_0 and $\theta_{j,n}(u)$ which converges to θ_0 , we may then apply Proposition E.1 so that

$$\mathbb{E}_{G_n(0)} \left| \frac{f^{(2d)}(\cdot, \theta)}{f(\cdot, G_n(0))} \right|^\alpha \lesssim 1$$

and thus

$$(A.6) \quad \mathbb{E}_{G_n(0)} |R_{1,n}(u)|^\alpha \stackrel{u}{\lesssim} n^{-\alpha/2} \varepsilon_n^{\alpha-1}.$$

Take $\alpha = 2$ to obtain the bound $n^{-1-1/(4d-2)}$ and thus (A.5).

Convergence of $B_n(u)$. Write $B_n(u) = B'_n(u) + B''_n(u)$ with

$$B'_n(u) = \sum_{i=1}^n Y_{i,n}(u)^2 - \frac{u^2}{n} \sum_{i=1}^n Z_{i,n}^2 \quad \text{and} \quad B''_n(u) = \frac{u^2}{n} \sum_{i=1}^n Z_{i,n}^2 - u^2 \Gamma_n.$$

We show that $B'_n(u) \xrightarrow{L^1} 0$ and $B''_n(u) \xrightarrow{L^1} 0$.

From (A.3),

$$B'_n(u) = \sum_{i=1}^n [R_{i,n}(u) - R_{i,n}(0)]^2 + \frac{2u}{\sqrt{n}} \sum_{i=1}^n [R_{i,n}(u) - R_{i,n}(0)] Z_{i,n},$$

so that taking the L^1 -norm and by the Cauchy-Schwarz inequality,

$$\begin{aligned} \mathbb{E}_{G_n(0)} |B'_n(u)| &\underset{u}{\lesssim} n \mathbb{E}_{G_n(0)} [|R_{1,n}(u)|^2 + |R_{1,n}(0)|^2] \\ &\quad + \sqrt{n \mathbb{E}_{G_n(0)} [|R_{1,n}(u)|^2 + |R_{1,n}(0)|^2]} \sqrt{\Gamma_n} \end{aligned}$$

and, by (A.5) and the fact that $\Gamma_n = \mathbb{E}_{G_n(0)} Z_{1,n}^2 \underset{u}{\lesssim} 1$, the r.h.s. tends to 0. Besides, we have

$$\mathbb{E}_{G_n(0)} |B''_n(u)|^2 = \frac{u^2}{n^2} \mathbb{E}_{G_n(0)} \left| \sum_{i=1}^n (Z_{i,n}^2 - \Gamma_n) \right| \underset{u}{\lesssim} n^{-1} \text{Var}_{G_n(0)}(Z_{1,n}^2)$$

which goes to zero since $\mathbb{E}_{G_n(0)} Z_{1,n}^4 \underset{u}{\lesssim} 1$ by Proposition E.1.

Convergence of $C_n(u)$. Finally we show that

$$C_n(u) = \sum_{i=1}^n |Y_{i,n}(u)|^3 \xrightarrow{L^1} 0.$$

A rough bound of (A.3) gives

$$C_n(u) \underset{u}{\lesssim} n^{-3/2} \sum_{i=1}^n |Z_{i,n}|^3 + \sum_{i=1}^n |R_{i,n}(u)|^3 + \sum_{i=1}^n |R_{i,n}(0)|^3$$

so that taking expectations

$$\mathbb{E}_{G_n(0)} |C_n(u)| \underset{u}{\lesssim} n^{-1/2} \mathbb{E}_{G_n(0)} |Z_{1,n}|^3 + n \mathbb{E}_{G_n(0)} [|R_{1,n}(u)|^3 + |R_{1,n}(0)|^3].$$

But each of the three terms in the r.h.s. tends to 0: the first one because of $\mathbb{E}_{G_n(0)} |Z_{1,n}|^3 \underset{u}{\lesssim} 1$ by Proposition E.1, the second and the third ones because of (A.6) for $\alpha = 3$. Thus $C_n(u)$ converges to 0 in L^1 .

LAN property. The above convergences entail the desired LAN property. Set

$$D_n(u) = \sum_{i=1}^n \text{Log}(1 + Y_{i,n}(u)) - \sum_{i=1}^n Y_{i,n}(u) + \frac{1}{2} \sum_{i=1}^n Y_{i,n}(u)^2.$$

We show first that $D_n(u) = O(C_n(u))$ in probability. But this is a straightforward consequence of $|\text{Log}(1 + Y_{i,n}(u)) - Y_{i,n}(u) + Y_{i,n}(u)^2/2| \leq |Y_{i,n}(u)|^3$ for $|Y_{i,n}(u)| \leq 2/3$ together with $C_n(u) \rightarrow 0$ in L^1 so that we get

$$|D_n(u)| \leq C_n(u)$$

with probability going to one.

Summarizing, we get that

$$\text{Log} \left(\frac{f_{n,u}(X)}{f_{n,0}(X)} \right) - uZ_n\sqrt{\Gamma_n} + \frac{u^2}{2}\Gamma_n = A_n(u) + \frac{1}{2}B_n(u) + D_n(u)$$

goes to 0 in probability.

APPENDIX B: AUXILIARY TRANSPORTATION CALCULUS

B.1. Two Lemmas.

LEMMA B.1. *Assume without loss of generality that $G_0 = \sum_{i=1}^m \pi_{i,0} \delta_{\theta_{i,0}}$, with all $\pi_{i,0} \geq \rho_0 > 0$ and all $\theta_{i+1,0} - \theta_{i,0} \geq \varepsilon_0 > 0$. Then for $\varepsilon = \varepsilon_0 \rho_0 / 8$,*

$$\left. \begin{array}{l} G = \sum_{i=1}^m \pi_i \delta_{\theta_i} \\ W_1(G, G_0) < \varepsilon \end{array} \right\} \implies \forall i, \pi_i > \frac{\rho_0}{2} \text{ and } |\theta_i - \theta_{i,0}| < \frac{\varepsilon_0}{4},$$

for a suitable numbering of the components of G . As a consequence, the support points θ_i of G are $\varepsilon_0/2$ -separated.

PROOF. By the very definition of W_1 (not the dual form), there is a probability measure Π on $\Theta \times \Theta$ with marginals $G = \Pi(\cdot \times \Theta)$ and $G_0 = \Pi(\Theta \times \cdot)$ such that, with $\Pi(\theta_j, \theta_{i,0}) := \Pi(\{\theta_j\} \times \{\theta_{i,0}\})$,

$$(B.1) \quad W_1(G, G_0) = \sum_{i,j=1}^m |\theta_j - \theta_{i,0}| \Pi(\theta_j, \theta_{i,0}).$$

Set $J_{i,0} = \{j \in \llbracket 1, m \rrbracket : |\theta_j - \theta_{i,0}| < \varepsilon_0/4\}$. Then for each $\theta_{i,0}$,

$$\begin{aligned} W(G, G_0) &\geq \frac{\varepsilon_0}{4} \sum_{j \notin J_{i,0}} \Pi(\theta_j, \theta_{i,0}) \\ &= \frac{\varepsilon_0}{4} \left[\pi_{i,0} - \sum_{j \in J_{i,0}} \Pi(\theta_j, \theta_{i,0}) \right] \\ &\geq \frac{\varepsilon_0}{4} \left[\rho_0 - \sum_{j \in J_{i,0}} \Pi(\theta_j, \theta_{i,0}) \right]. \end{aligned}$$

Thus, if $W(G, G_0) < \rho_0 \varepsilon_0 / 8$, then we must have for each i ,

$$(B.2) \quad \sum_{j \in J_{i,0}} \Pi(\theta_j, \theta_{i,0}) > \frac{\rho_0}{2}$$

and each $J_{i,0}$ is non empty. Furthermore, the (disjoint) $J_{i,0}$'s, $i \in \llbracket 1, m \rrbracket$, are all singletons ; otherwise there would be at least one $J_{i,0}$ empty, since G_0 has exactly m support points and G at most m . Considering a suitable numbering for the components of G , we can thus write $J_{i,0} = \{i\}$ so that $|\theta_i - \theta_{i,0}| < \varepsilon_0/4$ for each i and (B.2) yields $\pi_i \geq \Pi(\theta_i, \theta_{i,0}) > \rho_0/2$. \square

LEMMA B.2. *Recall that $J_r = \llbracket 1, m + m' \rrbracket$. Consider the probability measures π_\bullet and π'_\bullet on J_r defined by*

$$\pi_J = \sum_{j \in J \cap \llbracket 1, m \rrbracket} \varpi_j \quad \text{and} \quad \pi'_J = - \sum_{j \in J \cap \llbracket m+1, m+m' \rrbracket} \varpi_j.$$

There is a probability measure Π on $J_r \times J_r$ with marginal measures π_\bullet and π'_\bullet such that

$$\forall J \in \mathcal{T}, \quad \Pi(J, J) := \Pi(J \times J) = \pi_J \wedge \pi'_J.$$

PROOF. We proceed by induction on the generations of children, filling successively the products $J \times J$ with $\pi_J \wedge \pi'_J$.

First step. Consider the first generation of children, denoted by J_1, J_2, J_3 etc. Take J_1 and $K_1 = J_r \setminus J_1$ and fill $\Pi(J_1, J_1)$ with $\pi_{J_1} \wedge \pi'_{J_1}$ and $\Pi(K_1, K_1)$ with $\pi_{K_1} \wedge \pi'_{K_1}$. The off-diagonal weights $\Pi(J_1, K_1)$ and $\Pi(K_1, J_1)$ (at least one of them is zero) are then determined by the marginals weights on J_1 and K_1 . Here is an illustration:

	J_1	K_1	
J_1	$\pi_{J_1} \wedge \pi'_{J_1}$	$\Pi(J_1, K_1)$	π_{J_1}
K_1	$\Pi(K_1, J_1)$	$\pi_{K_1} \wedge \pi'_{K_1}$	π_{K_1}
	π'_{J_1}	π'_{K_1}	1

Second step. Repeat the first step with J_2 and $K_2 := K_1 \setminus J_2$ instead of J_1 and K_1 : fill $\Pi(J_2, J_2)$ with $\pi_{J_2} \wedge \pi'_{J_2}$ and $\Pi(K_2, K_2)$ with $\pi_{K_2} \wedge \pi'_{K_2}$ so that the off-diagonal weights $\Pi(J_2, K_2)$ and $\Pi(K_2, J_2)$ are determined by the marginal weights of the previous step. Consequently, the remaining weights $\Pi(J_2, J_1)$, $\Pi(K_2, J_1)$, $\Pi(J_1, J_2)$ and $\Pi(J_1, K_2)$ are completed by the initial marginal weights:

	J_1	J_2	K_2	
J_1	$\pi_{J_1} \wedge \pi'_{J_1}$			π_{J_1}
J_2		$\pi_{J_2} \wedge \pi'_{J_2}$	$\Pi(J_2, K_2)$	π_{J_2}
K_2		$\Pi(K_2, J_2)$	$\pi_{K_2} \wedge \pi'_{K_2}$	π_{K_2}
	π'_{J_1}	π'_{J_2}	π'_{K_2}	1

Third step. Repeat the previous procedure with J_3 and $K_3 := K_2 \setminus J_3$ instead of J_2 and K_2 and so on until all the children of the first generation are exhausted.

Next steps. Repeat the previous three steps for the children of J_1 , considering J_1 instead of J_r , next for the children of J_2 and so on until all generations are exhausted. \square

B.2. Proof of (20). Let $G_0 \in \mathcal{G}_{m_0}$. We have to prove that, for some $\varepsilon > 0$ that will be chosen later,

$$L := \inf_{\substack{G \neq G' \in \mathcal{G}_{\leq m_0} \\ W_1(G, G_0) < \varepsilon \\ W_1(G', G_0) < \varepsilon}} \frac{\|F(\cdot, G) - F(\cdot, G')\|_\infty}{W_1(G, G')} > 0.$$

Select distinct mixing distributions G_n and G'_n in $\mathcal{G}_{\leq m_0}$ with $W_1(G_n, G_0) < \varepsilon$ and $W_1(G'_n, G_0) < \varepsilon$ such that $\|F(\cdot, G_n) - F(\cdot, G'_n)\|_\infty / W_1(G_n, G'_n)$ converges to L .

Write $G_n = \sum_{j=1}^m \pi_j \delta_{\theta_j}$, $G'_n = \sum_{j=1}^m \pi'_j \delta_{\theta'_j}$ (keeping n implicit in weights

and support-points) and set

$$\Sigma_n := \sum_{j=1}^m [|\theta_j - \theta'_j| + |\pi_j - \pi'_j|].$$

We shall actually prove that

$$(B.3) \quad W_1(G_n, G'_n) \asymp \Sigma_n \asymp \|F(\cdot, G_n) - F(\cdot, G'_n)\|_\infty.$$

The l.h.s. of (B.3) follows from the straightforward inequality

$$W_1(G_n, G'_n) = \sup_{|f|_{\text{Lip}} \leq 1} \int_{\Theta} f d(G_n - G'_n) \leq \sum_{j=1}^m |\theta_j - \theta'_j| + \text{Diam}(\Theta) |\pi_j - \pi'_j|.$$

To prove the r.h.s. of (B.3) we start by a Taylor expansion of $F(x, \theta_j)$ around θ'_j and the use of Assumption B(1) to get, uniformly in x ,

$$\left| F(x, \theta_j) - F(x, \theta'_j) - (\theta_j - \theta'_j) F^{(1)}(x, \theta'_j) \right| \leq o(\theta_j - \theta'_j)$$

so that by linearity $F(x, G_n) - F(x, G'_n) = T_n(x) + o(\Sigma_n)$ with

$$(B.4) \quad T_n(x) = \sum_{j=1}^m (\pi_j - \pi'_j) F(x, \theta'_j) + \pi_j (\theta_j - \theta'_j) F^{(1)}(x, \theta'_j).$$

Let ρ_0 be the smallest weight of G_0 . By Lemma B.1, we can choose $\varepsilon > 0$ such that for any $G \in \mathcal{G}_{\leq m_0}$ with $W_1(G, G_0) < \varepsilon$, the weights of G are more than a half of ρ_0 and the support points of G are ε_0 -separated for some $\varepsilon_0 > 0$. Since the G_n and G'_n are in $\mathcal{G}_{\leq m_0}$ with $W_1(G_n, G_0) \vee W_1(G'_n, G_0) < \varepsilon$, the weights π_j are thus at least $\rho_0/2$ and each vector $(\theta'_j)_{1 \leq j \leq m}$ is ε_0 -separated. So that by Proposition 2.3 and (B.4),

$$\|T_n(\cdot)\|_\infty \underset{\varepsilon_0}{\gtrsim} \sum_{j=1}^m \left[|\pi_j - \pi'_j| + \frac{\rho_0}{2} |\theta_j - \theta'_j| \right]$$

and (B.3) follows.

APPENDIX C: CONSTRUCTION OF THE SCALING SEQUENCES

From compactity of the real subset Θ , we may pass (finitely) many times to subsequences of the ϑ_j (n is skipped from notations) for $j \in \llbracket 1, m + m' \rrbracket$ to obtain the following properties:

1. All the ϑ_j converge,
2. All the sequences $|\vartheta_j - \vartheta_{j'}|$ are monotone (in n),
3. Either $|\vartheta_j - \vartheta_{j'}| > 0$ for all n , or $|\vartheta_j - \vartheta_{j'}| \equiv 0$ and in that case (j, j') is called “pair of multiplicity” ,
4. The ratios $\frac{|\vartheta_{j''} - \vartheta_{j'''}}{|\vartheta_j - \vartheta_{j'}|}$ are monotone when (j, j') is not a pair of multiplicity.

We shall then define an equivalence relation between pairs. Let \mathcal{P} be the set of all pairs (j, j') and \mathcal{P}_0 the set of pairs of multiplicity. Next, define a binary relation between pairs outside \mathcal{P}_0 :

$$(j, j') \prec (j'', j''') \iff \sup_n \frac{|\vartheta_{j''} - \vartheta_{j'''}|}{|\vartheta_j - \vartheta_{j'}|} < \infty.$$

The binary relation \prec defines a total order by property 4. Furthermore, we get an equivalence relation outside \mathcal{P}_0 by

$$(j, j') \sim (j'', j''') \iff (j, j') \prec (j'', j''') \text{ and } (j'', j''') \prec (j', j')$$

so that we can split $\mathcal{P} \setminus \mathcal{P}_0$ into say R equivalence classes $\mathcal{P}_1 \prec \dots \prec \mathcal{P}_R$. For each $s \in \llbracket 1, R-1 \rrbracket$, choose a pair (j, j') in \mathcal{P}_s and set $\varepsilon_s = |\vartheta_j - \vartheta_{j'}|$. Set also $\varepsilon_0 \equiv 0$. If there is a pair in \mathcal{P}_R which is exactly of order one i.e $|\vartheta_j - \vartheta_{j'}| \asymp 1$, set $\varepsilon_R \equiv 1$ and define $S = R$. Otherwise, choose as before a pair (j, j') in \mathcal{P}_R , set $\varepsilon_R = |\vartheta_j - \vartheta_{j'}|$ and $\varepsilon_{R+1} \equiv 1$ and define $S = R+1$. To guarantee the order

$$\varepsilon_0 < \varepsilon_1 < \dots < \varepsilon_S$$

we may select once more subsequences of the ε_s . Thus, Lemma 7.1 is proved.

APPENDIX D: MATRIX TOOL AND SEPARATION PROPERTY

LEMMA D.1. *Let j, d_i and d be positive integers such that $\sum_{i=1}^j d_i = d$. Consider numbers $\theta_1, \dots, \theta_j$ all distinct. Write*

$$\mathcal{I} = \{(i, \ell) \in \mathbb{N} : 1 \leq i \leq j, 0 \leq \ell < d_i\}.$$

Define for each $(i, \ell) \in \mathcal{I}$ a d -dimensional column vector as follows:

$$a_{i,\ell}[p] = \frac{\theta_i^{p-\ell}}{(p-\ell)!} \mathbf{1}_{p \geq \ell}, \quad 0 \leq p < d,$$

and stack these vectors in a $d \times d$ matrix

$$(D.1) \quad A(\theta_1, \dots, \theta_j) = [a_{1,0} | \dots | a_{1,d_1-1} | \dots | a_{j,0} | \dots | a_{j,d_j-1}].$$

Then, the rank of $A(\theta_1, \dots, \theta_j)$ is d .

PROOF. Set for short $A = A(\theta_1, \dots, \theta_j)$. Let $\Lambda = (\lambda_{i,\ell})_{(i,\ell) \in \mathcal{I}}$ be a vector such that $A\Lambda = 0$. Proving the lemma is equivalent to proving that $\Lambda = 0$. Note that for each $p \in \llbracket 0, d-1 \rrbracket$

$$(A\Lambda)_p = \sum_{(i,\ell) \in \mathcal{I}} \lambda_{i,\ell} a_{i,\ell}[p] = \sum_{(i,\ell) \in \mathcal{I}} \lambda_{i,\ell} \frac{\theta_i^{p-\ell}}{(p-\ell)!} \mathbf{1}_{p \geq \ell} = 0,$$

so that for any $(d-1)$ -degree polynomial $P(x) = \sum_{p=0}^{d-1} c_p x^p/p!$, we have

$$(D.2) \quad (c_0, \dots, c_{d-1})A\Lambda = \sum_{p=0}^{d-1} c_p (A\Lambda)_p = \sum_{(i,\ell) \in \mathcal{I}} \lambda_{i,\ell} P^{(\ell)}(\theta_i) = 0.$$

Consider any $i \in \llbracket 1, j \rrbracket$. Choosing polynomials

$$P(x) = (x - \theta_i)^k \prod_{\substack{i'=1 \\ i' \neq i}}^j (x - \theta_{i'})^{d_{i'}}$$

successively from $k = d_i - 1$ down to $k = 0$ in (D.2) yields $\lambda_{i,k} = 0$ for $k = 0, \dots, d_i - 1$ and we are done. \square

COROLLARY D.2. *Let $A(\theta_1, \dots, \theta_j)$ be as in (D.1). Let $\varepsilon > 0$ and define the set of ε -separated vectors in Θ^j by*

$$\Theta_\varepsilon^j = \{(\theta_i)_{1 \leq i \leq j} : \forall i \neq i', \quad |\theta_i - \theta_{i'}| \geq \varepsilon\}.$$

For any vector $\Lambda \in \mathbb{R}^d$ and any vector $(\theta_i)_{1 \leq i \leq j} \in \Theta_\varepsilon^j$,

$$\|A(\theta_1, \dots, \theta_j)\Lambda\| \underset{\varepsilon, d}{\asymp} \|\Lambda\|.$$

PROOF. Note that the norm $\|A(\theta_1, \dots, \theta_j)\Lambda\|$ is a continuous function of $((\theta_1, \dots, \theta_j), \Lambda)$ on the compact space $\Theta_\varepsilon^j \times S(0, 1)$ where $S(0, 1)$ is the d -dimensional unit sphere. Its infimum and supremum are attained on $\Theta_\varepsilon^j \times S(0, 1)$, say at $((\theta_{*i})_{1 \leq i \leq j}, \Lambda_*)$ and $((\theta_i^*)_{1 \leq i \leq j}, \Lambda^*)$. Now, by Lemma D.1, $c_*(\varepsilon) = \|A(\theta_{*1}, \dots, \theta_{*j})\Lambda_*\|$ and $c^*(\varepsilon) = \|A(\theta_1^*, \dots, \theta_j^*)\Lambda^*\|$ are positive so that $c_*(\varepsilon) \|\Lambda\| \leq \|A(\theta_1, \dots, \theta_j)\Lambda\| \leq c^*(\varepsilon) \|\Lambda\|$ for every Λ and every $(\theta_i)_{1 \leq i \leq j}$ in Θ_ε^j . \square

APPENDIX E: (P, α) -SMOOTHNESS

E.1. Inherited smoothness for mixing distributions. Being (p, α) -smooth ensures finiteness of similar integrals when some θ_j are replaced with mixing distributions with components close to the θ_j :

PROPOSITION E.1. *Assume that the family $\{f(\cdot, \theta), \theta \in \Theta\}$ is (p, α) -smooth with $\alpha \geq 1$ and let $\varepsilon > 0$ as in (5). Let also $\rho_0 > 0$, $\theta_0 \in \Theta$ and positive integers $m \geq m_0$. Define mixing distributions*

$$G_n = \sum_{j=1}^m \pi_{j,n} \delta_{\theta_{j,n}}$$

such that

- For all $j \in \llbracket m_0, m \rrbracket$, $\theta_{j,n} \xrightarrow{n \rightarrow \infty} \theta_0$,
- For all n large enough, $\sum_{j=m_0}^m \pi_{j,n} \geq \rho_0$.

Then for any θ' satisfying $|\theta' - \theta_0| < \varepsilon/2$, for any mixing distribution G :

$$(E.1) \quad \mathbb{E}_G \left| \frac{f^{(p)}(\cdot, \theta')}{f(\cdot, G_n)} \right|_{\theta_0, \rho_0}^\alpha \leq 1$$

for n large enough. If, in addition, the function $x \mapsto |f^{(p)}(x, \theta_0)|$ has nonzero integral under λ , then for any mixing distribution G ,

$$(E.2) \quad \mathbb{E}_G \left| \frac{f^{(p)}(\cdot, \theta_0)}{f(\cdot, G)} \right|_{\theta_0}^\alpha \geq 1.$$

PROOF. For large n and all $j \in \llbracket m_0, m \rrbracket$, we have $|\theta_{j,n} - \theta'| < \varepsilon$ for all θ' such that $|\theta' - \theta_0| < \varepsilon/2$. For all such (j, n) and all θ , by (5) and compactness and continuity, there is a finite C such that

$$E_{p,\alpha}(\theta, \theta', \theta_{j,n}) = \mathbb{E}_\theta \left| \frac{f^{(p)}(\cdot, \theta')}{f(\cdot, \theta_{j,n})} \right|^\alpha \leq C.$$

Recall that \mathbb{E}_G is the expectation w.r.t. $f(x, G)d\lambda(x)$ and, likewise, \mathbb{E}_θ is the expectation w.r.t. $f(x, \theta)d\lambda(x)$. Since $f(x, G)$ is a convex combination of some $f(x, \theta)$, we may replace \mathbb{E}_θ by \mathbb{E}_G in the former expression. Since the function $1/y^\alpha$ is convex on positive reals, by Jensen inequality, setting $A = \sum_{j=m_0}^m \pi_{j,n}$,

$$\sum_{j=m_0}^m \frac{\pi_{j,n}}{A} \left| \frac{f^{(p)}(x, \theta')}{f(x, \theta_{j,n})} \right|^\alpha \geq \left| \frac{f^{(p)}(x, \theta')}{\sum_{j=m_0}^m \frac{\pi_{j,n}}{A} f(x, \theta_{j,n})} \right|^\alpha \geq A^\alpha \left| \frac{f^{(p)}(x, \theta')}{f(x, G_n)} \right|^\alpha,$$

and taking expectations with respect to G we obtain the upper bound:

$$\mathbb{E}_G \left| \frac{f^{(p)}(\cdot, \theta')}{f(\cdot, G_n)} \right|^\alpha \leq \frac{C}{A^\alpha} \leq \frac{C}{\rho_0^\alpha}.$$

The lower bound does not depend on (p, α) -smoothness. It is a simple consequence of rewriting:

$$\mathbb{E}_G \left| \frac{f^{(p)}(\cdot, \theta_0)}{f(\cdot, G)} \right|^\alpha = \int \frac{|f^{(p)}(x, \theta_0)|^\alpha}{f(x, G)^{\alpha-1}} d\lambda(x).$$

By assumption, there is a set B of measure $\lambda(B) = M > 0$ on which the function $|f^{(p)}(x, \theta_0)|$ is more than some $\delta > 0$. Now, for M small enough, the set $B \cap \{f(x, G) \leq 2/M\}$ is of measure at least $M/2$ and thus

$$\int \frac{|f^{(p)}(x, \theta_0)|^\alpha}{f(x, G)^{\alpha-1}} d\lambda(x) \geq \left[\frac{M}{2} \right]^\alpha \delta^\alpha.$$

□

E.2. (p, α) -smoothness of exponential families. Given our definition of (p, α) -smoothness, it only makes sense to consider one-parameter one-dimensional families. However, generalisation to higher dimensions should be easy.

PROPOSITION E.2. *Consider an exponential family with natural parameter $\theta \in \Theta_0 \subset \mathbb{R}$, with $g \in C^\infty$ and with a sufficient one-dimensional statistic $T(x)$, so that*

$$f(x, \theta) = h(x)g(\theta) \exp(\theta T(x)).$$

Let Θ be such that its ε -neighbourhood $\Theta \oplus B(0, \varepsilon)$ is included in Θ_0 . Then $\{f(\cdot, \theta), \theta \in \Theta\}$ is (p, α) -smooth for any p and α .

PROOF. Indeed,

$$\begin{aligned} f^{(p)}(x, \theta') &= h(x)e^{\theta' T(x)} \left[\sum_{k=0}^p \binom{p}{k} g^{(k)}(\theta') T^{p-k}(x) \right] \\ \frac{f^{(p)}(x, \theta')}{f(x, \theta'')} &= \frac{e^{(\theta' - \theta'') T(x)}}{g(\theta'')} \left[\sum_{k=0}^p \binom{p}{k} g^{(k)}(\theta') T^{p-k}(x) \right] \\ \left| \frac{f^{(p)}(x, \theta')}{f(x, \theta'')} \right|^\alpha &= \frac{e^{\alpha(\theta' - \theta'') T(x)}}{g^\alpha(\theta'')} \left| \sum_{k=0}^p \binom{p}{k} g^{(k)}(\theta') T^{p-k}(x) \right|^\alpha \end{aligned}$$

so that we get from (4)

$$E_{p,\alpha}(\theta, \theta', \theta'') = \frac{g(\theta) \mathbb{E}_{\theta + \alpha(\theta' - \theta'')} \left| \sum_{k=0}^p \binom{p}{k} g^{(k)}(\theta') T^{p-k}(\cdot) \right|^\alpha}{g^\alpha(\theta'') g(\theta + \alpha(\theta' - \theta''))}.$$

Since all the moments of the sufficient statistic $T(x)$ are finite under a distribution in the exponential family, and since $\theta + \alpha\theta' - \alpha\theta''$ is in Θ_0 for $(\theta' - \theta'') < \varepsilon/\alpha$, we obtain the finiteness of $E_{p,\alpha}(\theta, \theta', \theta'')$. Continuity is clear. \square

REFERENCES

BILLINGSLEY, P. (1995). *Probability and Measure. Wiley Series in Probability and Statistics*. Wiley, New York.

E-MAIL: philippe.heinrich@math.univ-lille1.fr

E-MAIL: jonas.kahn@math.univ-lille1.fr