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,

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

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 $\mu_k$, we get

$$f(x, G_n(u)) - f(x, G_0) = \pi_0 \sum_{j=m_0}^{m} \pi_j(u) \left[ f(x, \theta_{j,n}(u)) - f(x, \theta_0) \right]$$

(A.1)

$$= \pi_0 \left[ \sum_{k=1}^{2d-1} \mu_k \varepsilon_n^k f^{(k)}(x, \theta_0) + r_n(x, u) \right],$$

with

$$r_n(x, u) = \sum_{j=m_0}^{m} \pi_j(u) \int_{\theta_0}^{\theta_{j,n}(u)} f^{(2d)}(x, \theta) \frac{(\theta_{j,n}(u) - \theta)^{2d-1}}{(2d-1)!} d\theta.$$
Since the moments $\mu_1, \ldots, \mu_{2d-2}$ that do not depend on $u$ but $\mu_{2d-1} = u$, substracting (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

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

Next, by expanding $f(X_{1,n}, \theta)$ around $\theta_0$, dividing by $f(X_{1,n}, G_n(0))$, taking expectations and applying Fubini Theorem to the remainder, we get

$$0 = \sum_{k=1}^{2d-1} \frac{h^k}{k!} E_{G_n(0)} \left[ \frac{f^{(k)}(X_{1,n}, \theta_0)}{f(X_{1,n}, G_n(0))} \right] + \int_{\theta_0}^{\theta_0 + h} \frac{(\theta_0 + h - \theta)^{2d-1}}{(2d-1)!} E_{G_n(0)} \left[ \frac{f^{(2d)}(X_{1,n}, \theta)}{f(X_{1,n}, G_n(0))} \right] d\theta;$$

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 $k \in [1, 2d-1]$

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

and in particular $E_{G_n(0)}(Z_{1,n}) = 0$. Dividing (A.1) by $f(x, G_n(0))$ and taking $x = X_{1,n}$ gives in addition $E_{G_n(0)}(R_{1,n}(u)) = 0.$
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 \( \Gamma_n = E_{G_n(0)} |Z_{1,n}|^2 \asymp 1 \) and \( E_{G_n(0)} |Z_{1,n}|^3 \lesssim 1 \). As a consequence, Lyapunov condition holds:

\[
\sum_{i=1}^{n} E_{G_n(0)} |Z_{i,n}|^3 \bigg/ \left( \sum_{i=1}^{n} E_{G_n(0)} |Z_{i,n}|^2 \right)^{3/2} \asymp 1 \quad \text{and} \quad E_{G_n(0)} |Z_{1,n}|^3 \lesssim \frac{1}{\sqrt{n}} \quad n \to \infty.
\]

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

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

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

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 \( G_n(0) \).

\[ \text{Convergence of } A_n(u) \]

We first show that

\[
A_n(u) = \sum_{i=1}^{n} Y_{i,n}(u) - u Z_n \sqrt{\frac{L^2}{n}} \to 0.
\]

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) \( E_{G_n(0)} \left| \sum_{i=1}^{n} R_{i,n}(u) \right|^2 = n E_{G_n(0)} |R_{1,n}(u)|^2 \xrightarrow{n \to \infty} 0. \)

To this end, we look at the remainder (A.2) for fixed \( u \): for any \( \theta \) in the integrand, any \( j \) and \( n \), we have \( |\theta_{j,n}(u) - \theta| \lesssim |h_j(u)| \varepsilon_n \). Using (15) for
$U = u$, we may then write
\[|r_n(x,u)| \leq \sum_{j=m_0}^{m} \pi_j(u) \int_{\theta_0 - h(u) \varepsilon_n}^{\theta_0 + h(u) \varepsilon_n} |f^{(2d)}(x,\theta)| \frac{h(u)^{2d-1}n^{-1/2}}{(2d-1)!} d\theta.\]
\[\leq n^{-1/2} \int_{|\theta - \theta_0| \leq \varepsilon_n} |f^{(2d)}(x,\theta)| d\theta.\]

Then, for $\alpha \in [1,4]$, using the convexity of $x \mapsto x^\alpha$ and Hölder’s inequality, we may write
\[|R_{1,n}(u)|^\alpha = \left| \frac{\pi_0}{f(\cdot, G_n(0))} \right|^\alpha n^{-\alpha/2} \int_{|\theta - \theta_0| \leq \varepsilon_n} |f^{(2d)}(\cdot,\theta)| \frac{f(\cdot, G_n(0))}{f(\cdot, G_n(0))} d\theta.\]

Taking expectations w.r.t. $G_n(0)$, we may use Fubini theorem in the estimation above. Since moreover $\theta$ in the integrand is between $\theta_0$ and $\theta_{j,n}(u)$ which converges to $\theta_0$, we may then apply Proposition E.1 so that
\[\mathbb{E}_{G_n(0)} \left| \frac{f^{(2d)}(\cdot,\theta)}{f(\cdot, G_n(0))} \right|^\alpha \leq 1\]
and thus
\[(A.6) \quad \mathbb{E}_{G_n(0)} |R_{1,n}(u)|^\alpha \leq n^{-\alpha/2} \varepsilon_n^{\alpha-1}.\]

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),
\[B'_n(u) = \sum_{i=1}^{n} [R_{i,n}(u) - R_{i,n}(0)]^2 + \frac{2u}{\sqrt{n}} \sum_{i=1}^{n} [R_{i,n}(u) - R_{i,n}(0)] Z_{i,u},\]
so that taking the $L^1$-norm and by the Cauchy-Schwarz inequality,
\[
\begin{align*}
    \mathbb{E}_{G_n(0)} \left| B_n'(u) \right| & \leq n \mathbb{E}_{G_n(0)} \left[ |R_{1,n}(u)|^2 + |R_{1,n}(0)|^2 \right] \\
    & + \sqrt{n \mathbb{E}_{G_n(0)} \left[ |R_{1,n}(u)|^2 + |R_{1,n}(0)|^2 \right]} \sqrt{\Gamma_n}
\end{align*}
\]
and, by (A.5) and the fact that $\Gamma_n = \mathbb{E}_{G_n(0)} Z_{1,n}^2 \approx 1$, the r.h.s. tends to 0.

Besides, we have
\[
\begin{align*}
    \mathbb{E}_{G_n(0)} \left| B_n''(u) \right|^2 & = \frac{u^2}{n^2} \mathbb{E}_{G_n(0)} \left[ \sum_{i=1}^{n} (Z_{i,n}^2 - \Gamma_n) \right]^2 \\
    & \approx n^{-1} \text{Var}_{G_n(0)}(Z_1^2)
\end{align*}
\]
which goes to zero since $\mathbb{E}_{G_n(0)} Z_{1,n}^4 \approx 1$ 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
\[
C_n(u) \approx n^{-3/2} \sum_{i=1}^{n} |Z_{i,n}|^3 + \sum_{i=1}^{n} |R_{i,n}(u)|^3 + \sum_{i=1}^{n} |R_{i,n}(0)|^3
\]
so that taking expectations
\[
\mathbb{E}_{G_n(0)} |C_n(u)| \approx n^{-1/2} \mathbb{E}_{G_n(0)} |Z_{1,n}|^3 + n \mathbb{E}_{G_n(0)} \left[ |R_{1,n}(u)|^3 + |R_{1,n}(0)|^3 \right].
\]

But each of the three terms in the r.h.s. tends to 0: the first one because of $\mathbb{E}_{G_n(0)} |Z_{1,n}|^3 \approx 1$ 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
\[
D_n(u) = \sum_{i=1}^{n} \log (1 + Y_{i,n}(u)) - \sum_{i=1}^{n} Y_{i,n}(u) + \frac{1}{2} \sum_{i=1}^{n} Y_{i,n}(u)^2.
\]

We show first that $D_n(u) = O(C_n(u))$ in probability. But this is a straightforward consequence of $|\log(1 + Y_{i,n}(u)) - Y_{i,n}(u)|^2 \leq |Y_{i,n}(u)|^3$ 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

\[
\log \left( \frac{f_{n,u}(X)}{f_{n,0}(X)} \right) - u Z_n \sqrt{\Gamma_n} + \frac{u^2}{2} \Gamma_n = A_n(u) + \frac{1}{2} B_n(u) + D_n(u)
\]

goes to 0 in probability.

APPENDIX B: AUXILIARY TRANSPORTATION CALCULUS

B.1. Two Lemmas.

**Lemma B.1.** Assume without loss of generality that \( G_0 = \sum_{i=1}^{m} \pi_i \delta_{\theta_i,0} \), with all \( \pi_{i,0} \geq \rho_0 > 0 \) and all \( \theta_{i+1,0} - \theta_{i,0} \geq \epsilon_0 > 0 \). Then for \( \epsilon = \epsilon_0 \rho_0 / 8 \),

\[
G = \sum_{i=1}^{m} \pi_i \delta_{\theta_i} \quad \implies \forall i, \pi_i > \frac{\rho_0}{2} \quad \text{and} \quad |\theta_i - \theta_{i,0}| < \frac{\epsilon_0}{4},
\]

for a suitable numbering of the components of \( G \). As a consequence, the support points \( \theta_i \) of \( G \) are \( \epsilon_0 / 2 \)-separated.

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

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

Set \( J_{i,0} = \{ j \in [1, m] : |\theta_j - \theta_{i,0}| < \epsilon_0 / 4 \} \). Then for each \( \theta_{i,0} \),

\[
W(G, G_0) \geq \frac{\epsilon_0}{4} \sum_{j \notin J_{i,0}} \Pi(\theta_j, \theta_{i,0})
\]

\[
= \frac{\epsilon_0}{4} \left[ \pi_{i,0} - \sum_{j \in J_{i,0}} \Pi(\theta_j, \theta_{i,0}) \right]
\]

\[
\geq \frac{\epsilon_0}{4} \left[ \rho_0 - \sum_{j \in J_{i,0}} \Pi(\theta_j, \theta_{i,0}) \right].
\]

Thus, if \( W(G, G_0) < \rho_0 \epsilon_0 / 8 \), then we must have for each \( i \),

\[
\sum_{j \in J_{i,0}} \Pi(\theta_j, \theta_{i,0}) > \frac{\rho_0}{2}
\]

\[
(B.2)
\]
and each $J_{i,0}$ is non empty. Furthermore, the (disjoint) $J_{i,0}$'s, $i \in [1, m]$, 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 $\pi_i \geq \Pi(\theta_i, \theta_{i,0}) > \rho_0/2$. 

**Lemma B.2.** Recall that $J_r = [1, m + m']$. Consider the probability measures $\pi_\bullet$ and $\pi'_\bullet$ on $J_r$ defined by

$$
\pi_j = \sum_{j \in J \cap [1, m]} \varpi_j \quad \text{and} \quad \pi'_j = -\sum_{j \in J \cap [m+1, m+m']} \varpi_j.
$$

There is a probability measure $\Pi$ on $J_r \times J_r$ with marginal measures $\pi_\bullet$ and $\pi'_\bullet$ 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:
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 $\pi_{J_2} \land \pi'_{J_2}$ and $\Pi(K_2, K_2)$ with $\pi_{K_2} \land \pi'_{K_2}$ so that the off-diagonal weights $\Pi(J_2, K_2)$ and $\Pi(K_2, J_2)$ are determined by the marginals 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:
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.

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,

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

Select distinct mixing distributions \( G_n \) and \( G_n' \) in \( \mathcal{G}_{0 < m_0} \) 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
\[ \Sigma_n := \sum_{j=1}^{m} \left[ |\theta_j - \theta'_j| + |\pi_j - \pi'_j| \right]. \]

We shall actually prove that
\[ (B.3) \quad W_1(G_n, G'_n) \lesssim \Sigma_n \lesssim \|F(\cdot, G_n) - F(\cdot, G'_n)\|_\infty. \]

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 \( \theta'_j \) 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 \( \rho_0 \) be the smallest weight of \( G_0 \). By Lemma B.1, we can choose \( \varepsilon > 0 \) such that for any \( G \in \mathcal{G}_{\leq \rho_0} \) with \( W_1(G, G_0) < \varepsilon \), the weights of \( G \) are more than a half of \( \rho_0 \) and the support points of \( G \) are \( \varepsilon_0 \)-separated for some \( \varepsilon_0 > 0 \). Since the \( G_n \) and \( G'_n \) are in \( \mathcal{G}_{\leq \rho_0} \) with \( W_1(G_n, G_0) \vee W_1(G'_n, G_0) < \varepsilon \), the weights \( \pi_j \) are thus at least \( \rho_0/2 \) and each vector \( (\theta'_j)_{1 \leq j \leq m} \) is \( \varepsilon_0 \)-separated. So that by Proposition 2.3 and (B.4),
\[ \|T_n(\cdot)\|_\infty \geq \sum_{j=1}^{m} \left[ |\pi_j - \pi'_j| + \frac{\rho_0}{2} |\theta_j - \theta'_j| \right] \]
and (B.3) follows.

**APPENDIX C: CONSTRUCTION OF THE SCALING SEQUENCES**

From compacity of the real subset \( \Theta \), we may pass (finitely) many times to subsequences of the \( \theta_j \) (\( n \) is skipped from notations) for \( j \in \left[ 1, m + m' \right] \) to obtain the following properties:
1. All the $\vartheta_j$ 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 $P$ be the set of all pairs $(j, j')$ and $P_0$ the set of pairs of multiplicity. Next, define a binary relation between pairs outside $P_0$:

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

The binary relation $\prec$ defines a total order by property 4. Furthermore, we get an equivalence relation outside $P_0$ by

$$(j, j') \sim (j'', j''') \iff (j, j') \prec (j'', j''')$$

so that we can split $P \setminus P_0$ into say $R$ equivalence classes $P_1 < \ldots < P_R$. For each $s \in [1, R - 1]$, choose a pair $(j, j')$ in $P_s$ and set $\varepsilon_s = |\vartheta_j - \vartheta_{j'}|$. Set also $\varepsilon_0 \equiv 0$. If there is a pair in $P_R$ 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 $P_R$, set $\varepsilon_R = |\vartheta_j - \vartheta_{j'}|$ and $\varepsilon_{R+1} \equiv 1$ and define $S = R + 1$. To guarantee the order

$$\varepsilon_0 < \varepsilon_1 < \ldots < \varepsilon_S$$

we may select once more subsequences of the $\varepsilon_s$. 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 $\theta_1, \ldots, \theta_j$ all distinct. Write

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

Define for each $(i, \ell) \in I$ 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

$$A(\theta_1, \ldots, \theta_j) = [a_{1,0} \cdots a_{1,d_1-1} \cdots a_{j,0} \cdots a_{j,d_j-1}].$$

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


Proof. Set for short \( A = A(\theta_1, \ldots, \theta_j) \). Let \( \Lambda = (\lambda_i, \ell)_{(i, \ell) \in I} \) be a vector such that \( AA = 0 \). Proving the lemma is equivalent to proving that \( \Lambda = 0 \).

Note that for each \( p \in [0, d - 1] \)
\[
(A\Lambda)_p = \sum_{(i, \ell) \in I} \lambda_i, \ell a_{i, \ell}[p] = \sum_{(i, \ell) \in I} \lambda_i, \ell \frac{\theta_i^{p-\ell}}{(p-\ell)!} 1_{p \geq \ell} = 0,
\]
so that for any \((d-1)\)-degree polynomial \( P(x) = \sum_{p=0}^{d-1} c_p x^p / p! \), we have
\[
(c_0, \ldots, c_{d-1}) A\Lambda = \sum_{p=0}^{d-1} c_p (A\Lambda)_p = \sum_{(i, \ell) \in I} \lambda_i, \ell P(\ell)(\theta_i) = 0.
\]
Consider any \( i \in [1, j] \). Choosing polynomials
\[
P(x) = (x - \theta_i)^k \prod_{i' \neq i} (x - \theta_i')^{d_{i'}}
\]
successively from \( k = d_i - 1 \) down to \( k = 0 \) in \((D.2)\) yields \( \lambda_{i,k} = 0 \) for \( k = 0, \ldots, d_i - 1 \) and we are done. \( \square \)

Corollary D.2. Let \( A(\theta_1, \ldots, \theta_j) \) be as in \((D.1)\). Let \( \varepsilon > 0 \) and define the set of \( \varepsilon \)-separated vectors in \( \Theta^j \) by
\[
\Theta^j_\varepsilon = \{ (\theta_i)_{1 \leq i \leq j} : \forall i \neq i', |\theta_i - \theta_{i'}| \geq \varepsilon \}.
\]
For any vector \( \Lambda \in \mathbb{R}^d \) and any vector \((\theta_i)_{1 \leq i \leq j} \in \Theta^j_\varepsilon \),
\[
\|A(\theta_1, \ldots, \theta_j)\Lambda\| = \|A\Lambda\|.
\]
Proof. Note that the norm \( \|A(\theta_1, \ldots, \theta_j)\Lambda\| \) is a continuous function of \(((\theta_1, \ldots, \theta_j), \Lambda)\) on the compact space \( \Theta^j_\varepsilon \times S(0, 1) \) where \( S(0, 1) \) is the \( d \)-dimensional unit sphere. Its infimum and supremum are attained on \( \Theta^j_\varepsilon \times S(0, 1) \), say at \(((\theta_{i*})_{1 \leq i \leq j}, \Lambda_*)\) and \(((\theta_{i*}')_{1 \leq i \leq j}, \Lambda_*')\). Now, by Lemma D.1, \( c_\varepsilon(\varepsilon) = \|A(\theta_{i*1}, \ldots, \theta_{i*j})\Lambda_*\| \) and \( c_\varepsilon(\varepsilon) = \|A(\theta_{i*1}', \ldots, \theta_{i*j}')\Lambda_*'\| \) are positive so that \( c_\varepsilon(\varepsilon) \|\Lambda\| \leq \|A(\theta_1, \ldots, \theta_j)\Lambda\| \leq c_\varepsilon(\varepsilon) \|\Lambda\| \) for every \( \Lambda \) and every \((\theta_i)_{1 \leq i \leq j} \) in \( \Theta^j_\varepsilon \). \( \square \)
APPENDIX E: ($P, \alpha$)-SMOOTHNESS

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

**Proposition E.1.** Assume that the family $\{f(\cdot, \theta), \theta \in \Theta\}$ is ($p, \alpha$)-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 $j \in [m_0, m]$, $\theta_{j,n} \xrightarrow{n \to \infty} \theta_0$,
- For all $n$ large enough, $\sum_{j=m_0}^{m} \pi_{j,n} \geq \rho_0$.

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

$$E_G \left| \frac{f(p(\cdot, \theta'))}{f(\cdot, G_n)} \right|^\alpha \lessapprox 1_{\theta_0, \rho_0}$$

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

$$E_G \left| \frac{f(p(\cdot, \theta_0)}{f(\cdot, G)} \right|^\alpha \gtrapprox 1.$$  

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

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

Recall that $E_G$ is the expectation w.r.t. $f(x, G)d\lambda(x)$ and, likewise, $E_\theta$ 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 $E_\theta$ by $E_G$ in the former expression. Since the function $1/y^\alpha$ is convex on positive reals, by Jensen inequality, setting $A = \sum_{j=m_0}^{m} \pi_{j,n}$,

$$\sum_{j=m_0}^{m} \pi_{j,n} \left| \frac{f(p(x, \theta'))}{f(x, \theta_{j,n})} \right|^\alpha \geq \left| \frac{f(p(x, \theta'))}{\sum_{j=m_0}^{m} \pi_{j,n} f(x, \theta_{j,n})} \right|^\alpha \geq A^\alpha \left| \frac{f(p(x, \theta'))}{f(x, G_n)} \right|^\alpha,$$
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 $(p, \alpha)$-smoothness. It is a simple consequence of rewriting:

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

By assumption, there is a set $B$ of measure $\lambda(B) = M > 0$ on which the function $|f^{(p)}(x, \theta)|$ 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

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

$\square$

**E.2. $(p, \alpha)$-smoothness of exponential families.** Given our definition of $(p, \alpha)$-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 $\Theta$ be such that its $\varepsilon$-neighbourhood $\Theta \oplus B(0, \varepsilon)$ is included in $\Theta_0$. Then $\{f(\cdot, \theta), \theta \in \Theta\}$ is $(p, \alpha)$-smooth for any $p$ and $\alpha$.

**Proof.** Indeed,

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

imsart-aos ver. 2014/10/16 file: SuppAOSv18.tex date: July 7, 2017
so that we get from (4)

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

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 \( \Theta_0 \) for \( (\theta' - \theta'') < \varepsilon/\alpha \), we obtain the finiteness of \( E_{p,\alpha}(\theta, \theta', \theta'') \). Continuity is clear.

REFERENCES


E-mail: philippe.heinrich@math.univ-lille1.fr  E-mail: jonas.kahn@math.univ-lille1.fr