
Supplementary material for “Modifications of the BIC for order selection in finite mixture models”

Hien Duy Nguyen · TrungTin Nguyen

Abstract This supplementary material contains complete proofs of the technical results and primary theorems.

S.1 Proofs

Proof of Proposition 1

Let P denote the probability measure with density f_0 with respect to $\mathfrak{m}$, and let $P_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$. Define the measure $\mathfrak{m}_0$ by

$$\mathfrak{m}_0(dx) = \mathbf{1}\{f_0(x) > 0\} \mathfrak{m}(dx).$$

Following van de Geer (2000), write $\bar{f} = (f + f_0)/2$ and

$$\bar{\mathcal{F}} = \{\bar{f} : f \in \mathcal{F}\}, \quad \bar{\mathcal{F}}^{1/2} = \{\bar{f}^{1/2} : \bar{f} \in \bar{\mathcal{F}}\}.$$

Let $\hat{f}_n : \Omega \rightarrow \mathcal{F}$ denote a maximum likelihood estimator:

$$\hat{f}_n \in \arg \max_{f \in \mathcal{F}} P_n \log f.$$

For any $0 < \delta \leq 1$, define the local class

$$\bar{\mathcal{F}}^{1/2}(\delta) = \left\{ \bar{f}^{1/2} : \bar{f} = (f + f_0)/2, f \in \mathcal{F}, \mathfrak{h}(f, f_0) \leq \delta \right\},$$

Hien Duy Nguyen

Department of Mathematics and Physical Science, La Trobe University, Melbourne, Australia

Institute of Mathematics for Industry, Kyushu University, Fukuoka, Japan

TrungTin Nguyen

ARC Centre of Excellence for the Mathematical Analysis of Cellular Systems

School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia

and the associated local bracketing entropy integral in $\mathcal{L}_2(\mathbf{m}_0)$:

$$J_{[]}(\delta, \bar{\mathcal{F}}^{1/2}(\delta), \mathcal{L}_2(\mathbf{m}_0)) = \delta \vee \int_0^\delta \sqrt{H_{[]} (u, \bar{\mathcal{F}}^{1/2}(\delta), \mathcal{L}_2(\mathbf{m}_0))} du. \quad (\text{S.1})$$

By van de Geer (2000, Cor. 7.5), if there exists a function $\Psi : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}$ such that

$$\Psi(\delta) \geq J_{[]}(\delta, \bar{\mathcal{F}}^{1/2}(\delta), \mathcal{L}_2(\mathbf{m}_0)),$$

where $\Psi(\delta)/\delta^2$ is non-increasing in δ , then there exists $c_1 > 0$ such that

$$\mathbb{P} \left(\int_{\mathbb{X}} \log \frac{\hat{f}_n}{f_0} dP_n \geq \delta^2 \right) \leq c_1 \exp \left\{ -\frac{n\delta^2}{c_1^2} \right\},$$

for every $\delta \geq \delta_n$, where δ_n satisfies $\sqrt{n}\delta_n^2 \geq c\Psi(\delta_n)$ for a universal constant $c > 0$.

It therefore suffices to show that our hypothesis $\Psi(\delta) \geq J(\delta)$, where J is defined in (3), implies

$$\Psi(\delta) \geq J_{[]}(\delta, \bar{\mathcal{F}}^{1/2}(\delta), \mathcal{L}_2(\mathbf{m}_0)).$$

We first relate the bracketing numbers of $\bar{\mathcal{F}}^{1/2}$ in $\mathcal{L}_2(\mathbf{m}_0)$ to those of $\mathcal{F}$ in $\mathcal{L}_1(\mathbf{m})$.

Lemma S.1 *If $\mathcal{F}$ is a space of probability densities on $\mathbb{X}$, then for each $\delta > 0$,*

$$N_{[]}(\delta, \bar{\mathcal{F}}^{1/2}, \mathcal{L}_2(\mathbf{m}_0)) \leq N_{[]} (2\delta^2, \mathcal{F}, \mathcal{L}_1(\mathbf{m})).$$

Proof Let $\delta > 0$ and let $([f_j^L, f_j^U])_{j \in [N]}$ be a $2\delta^2$ -bracketing of $\mathcal{F}$ with respect to the $\mathcal{L}_1(\mathbf{m})$ norm. Replacing bracket endpoints by their positive parts if necessary, we may assume that f_j^L and f_j^U are nonnegative; this preserves the bracket order for nonnegative densities and does not increase the $\mathcal{L}_1$ width. For each j , define $\bar{f}_j^L = (f_j^L + f_0)/2$ and $\bar{f}_j^U = (f_j^U + f_0)/2$. Then for $\bar{f} = (f + f_0)/2$, we have $\bar{f}_j^L \leq \bar{f} \leq \bar{f}_j^U$ whenever $f_j^L \leq f \leq f_j^U$.

Moreover, for $a, b \geq 0$, $(\sqrt{a} - \sqrt{b})^2 \leq |a - b|$. Hence,

$$\begin{aligned} \left\| (\bar{f}_j^U)^{1/2} - (\bar{f}_j^L)^{1/2} \right\|_{\mathcal{L}_2(\mathbf{m}_0)}^2 &= \int_{\mathbb{X}} \left\{ (\bar{f}_j^U)^{1/2} - (\bar{f}_j^L)^{1/2} \right\}^2 d\mathbf{m}_0 \\ &\leq \int_{\mathbb{X}} |\bar{f}_j^U - \bar{f}_j^L| d\mathbf{m}_0 \leq \int_{\mathbb{X}} |\bar{f}_j^U - \bar{f}_j^L| d\mathbf{m} \\ &= \frac{1}{2} \int_{\mathbb{X}} |f_j^U - f_j^L| d\mathbf{m} \leq \delta^2. \end{aligned}$$

Thus, $[(\bar{f}_j^L)^{1/2}, (\bar{f}_j^U)^{1/2}]$ is a δ -bracket for $\bar{\mathcal{F}}^{1/2}$ in $\mathcal{L}_2(\mathbf{m}_0)$, and the claim follows. $\square$

By the preceding lemma, for each $u > 0$ and $0 < \delta \leq 1$,

$$H_{\square} \left(u, \bar{\mathcal{F}}^{1/2}(\delta), \mathcal{L}_2(\mathbf{m}_0) \right) \leq H_{\square} \left(2u^2, \mathcal{F}(\delta), \mathcal{L}_1(\mathbf{m}) \right).$$

Since the bracketing entropy is non-increasing in its first argument, we have

$$H_{\square} \left(2u^2, \mathcal{F}(\delta), \mathcal{L}_1(\mathbf{m}) \right) \leq H_{\square} \left(u^2, \mathcal{F}(\delta), \mathcal{L}_1(\mathbf{m}) \right).$$

Therefore,

$$\begin{aligned} J_{\square} \left(\delta, \bar{\mathcal{F}}^{1/2}(\delta), \mathcal{L}_2(\mathbf{m}_0) \right) &= \delta \vee \int_0^{\delta} \sqrt{H_{\square} \left(u, \bar{\mathcal{F}}^{1/2}(\delta), \mathcal{L}_2(\mathbf{m}_0) \right)} du \\ &\leq \delta \vee \int_0^{\delta} \sqrt{H_{\square} \left(u^2, \mathcal{F}(\delta), \mathcal{L}_1(\mathbf{m}) \right)} du = J(\delta). \end{aligned}$$

Thus, if $\Psi(\delta) \geq J(\delta)$, then also $\Psi(\delta) \geq J_{\square} \left(\delta, \bar{\mathcal{F}}^{1/2}(\delta), \mathcal{L}_2(\mathbf{m}_0) \right)$, and Proposition 1 follows from van de Geer (2000, Cor. 7.5). $\square$

Proof of Lemma 3

For each $k \in [\bar{k}]$, let $\mathcal{N}_k^{\phi} = \left\{ -\log f : f \in \mathcal{M}_k^{\phi} \right\}$. We first show that $\mathcal{N}_k^{\phi}$ are GC classes. Indeed, if we write $\mathbb{S}_k = \left\{ (\pi_1, \dots, \pi_k) \in [0, 1]^k : \sum_{z=1}^k \pi_z = 1 \right\} \times \mathbb{T}^k \subset \mathbb{R}^{(m+1)k}$, then we can characterise $\mathcal{N}_k^{\phi}$ as

$$\mathcal{N}_k^{\phi} = \left\{ -\log f(\cdot; \psi_k) : \psi_k \in \mathbb{S}_k \right\},$$

where we index by the compact Euclidean space $\mathbb{S}_k$ instead of $\mathcal{M}_k^{\phi}$. Moreover, by A2 and A3 the map $\psi_k \mapsto -\log f(x; \psi_k)$ is continuous for each fixed x (since $f(x; \psi_k) > 0$), and $\mathbb{S}_k$ is compact. Hence, once an envelope in $\mathcal{L}_1(P)$ is available, the bracketing numbers $N_{\square} \left(\delta, \mathcal{N}_k^{\phi}, \mathcal{L}_1(P) \right)$ are finite for every $\delta > 0$ by the second part of Lemma 1. We can write $\ell_{k,n}(\psi_k) = -P_n \log f(\cdot; \psi_k)$ and $\ell_k(\psi_k) = -P \log f(\cdot; \psi_k)$, and by A1 and Lemma 1

$$\|P_n - P\|_{\mathcal{N}_k^{\phi}} = \max_{\psi_k \in \mathbb{S}_k} |\ell_{k,n}(\psi_k) - \ell_k(\psi_k)| \xrightarrow[n \rightarrow \infty]{\text{a.s.}} 0$$

if there exists a $G \in \mathcal{L}_1(P)$, such that $|\log f(x; \psi_k)| \leq G(x)$, for each $x \in \mathbb{X}$. We prove the following useful bound from Atienza et al. (2007).

Lemma S.2 *Under A2 and A3, for each $f(x; \psi_k) = \sum_{z=1}^k \pi_z \phi(x; \theta_z) \in \mathcal{M}_k^{\phi}$ and $x \in \mathbb{X}$,*

$$|\log f(x; \psi_k)| \leq \sum_{z=1}^k |\log \phi(x; \theta_z)| \leq k \max_{\theta \in \mathbb{T}} |\log \phi(x; \theta)|.$$

Proof Firstly observe that for every $x \in \mathbb{X}$,

$$\min_{z \in [k]} \phi(x; \theta_z) \leq f(x; \psi_k) = \sum_{z=1}^k \pi_z \phi(x; \theta_z) \leq \max_{z \in [k]} \phi(x; \theta_z)$$

and thus

$$\log \left\{ \min_{z \in [k]} \phi(x; \theta_z) \right\} \leq \log f(x; \psi_k) \leq \log \left\{ \max_{z \in [k]} \phi(x; \theta_z) \right\}.$$

Then,

$$\begin{aligned} |\log f(x; \psi_k)| &\leq \left| \log \left\{ \min_{z \in [k]} \phi(x; \theta_z) \right\} \right| \vee \left| \log \left\{ \max_{z \in [k]} \phi(x; \theta_z) \right\} \right| \\ &\leq \sum_{z=1}^k |\log \phi(x; \theta_z)| \leq k \max_{\theta \in \mathbb{T}} |\log \phi(x; \theta)|, \end{aligned}$$

as required, where $\max_{\theta \in \mathbb{T}} |\log \phi(x; \theta)|$ exists by A2 and A3. $\square$

Using Lemma S.2, we have $|\log f(x; \psi_k)| \leq k \max_{\theta \in \mathbb{T}} |\log \phi(x; \theta)|$. Then by A4, since we have $\max_{\theta \in \mathbb{T}} |\log \phi(x; \theta)| \leq G_1(x)$ with $G_1 \in \mathcal{L}_1(P)$, we can take $G(x) = kG_1(x)$, as required. The conclusion then follows by the fact that the minimum is Lipschitz with respect to the uniform norm (see, e.g., Hinderer et al. 2016, A.4.4), i.e.,

$$\left| \min_{\psi_k \in \mathbb{S}_k} \ell_{k,n}(\psi_k) - \min_{\psi_k \in \mathbb{S}_k} \ell_k(\psi_k) \right| \leq \max_{\psi_k \in \mathbb{S}_k} |\ell_{k,n}(\psi_k) - \ell_k(\psi_k)|,$$

where the minima exist by compactness of $\mathbb{S}_k$ and continuity of $\psi_k \mapsto \ell_{k,n}(\psi_k)$ and $\psi_k \mapsto \ell_k(\psi_k)$ (by A2–A4). $\square$

Proof of Lemma 4

For each $k \geq k_0$, we first bound $N_{[]}(\delta, \mathcal{M}_k^\phi, \mathcal{L}_1(\mathbf{m}))$. By A2 and A3, noting that each $f_k \in \mathcal{M}_k^\phi$ has the form $f_k(\cdot; \psi_k) = \sum_{z=1}^k \pi_z \phi(\cdot; \theta_z)$, Lemma 2 implies that if

$$\left| \sum_{z=1}^k \pi_z \phi(x; \theta_z) - \sum_{z=1}^k \pi_z^* \phi(x; \theta_z^*) \right| \leq G(x) \|\psi_k - \psi_k^*\|,$$

for some $G \in \mathcal{L}_1(\mathbf{m})$ and all $\psi_k, \psi_k^* \in \mathbb{S}_k$, then

$$N_{[]}(\delta, \mathcal{M}_k^\phi, \mathcal{L}_1(\mathbf{m})) \leq K \left(\frac{\|G\|_{\mathcal{L}_1(\mathbf{m})} \text{diam}(\mathbb{S}_k)}{\delta} \right)^{(m+1)k},$$

for some constant $K > 0$. Here $\text{diam}(\mathbb{S}_k) < \infty$ since $\mathbb{S}_k$ is compact.

Observe that

$$\begin{aligned}
& \left| \sum_{z=1}^k \pi_z \phi(x; \theta_z) - \sum_{z=1}^k \pi_z^* \phi(x; \theta_z^*) \right| \\
& \leq \sum_{z=1}^k |\pi_z \phi(x; \theta_z) - \pi_z^* \phi(x; \theta_z^*)| \\
& = \sum_{z=1}^k |\pi_z \{\phi(x; \theta_z) - \phi(x; \theta_z^*)\} + (\pi_z - \pi_z^*) \phi(x; \theta_z^*)| \\
& \leq \sum_{z=1}^k \{|\pi_z| |\phi(x; \theta_z) - \phi(x; \theta_z^*)| + |\phi(x; \theta_z^*)| |\pi_z - \pi_z^*|\} \\
& \leq \sum_{z=1}^k \left\{ |\phi(x; \theta_z) - \phi(x; \theta_z^*)| + \max_{\theta \in \mathbb{T}} |\phi(x; \theta)| |\pi_z - \pi_z^*| \right\}.
\end{aligned}$$

where we note that $|\pi_z| \leq 1$ and the maximum exists by A2 and A3. Then, by A5, we have L_ϕ , such that

$$|\phi(x; \theta_z) - \phi(x; \theta_z^*)| \leq L_\phi(x) \|\theta_z - \theta_z^*\|_1$$

and since there exists a $G_2 \in \mathcal{L}_1(\mathfrak{m})$, such that $L_\phi(x) + \max_{\theta \in \mathbb{T}} |\phi(x; \theta)| \leq G_2(x)$, for every $x \in \mathbb{X}$, it holds that

$$\begin{aligned}
& \left| \sum_{z=1}^k \pi_z \phi(x; \theta_z) - \sum_{z=1}^k \pi_z^* \phi(x; \theta_z^*) \right| \\
& \leq \sum_{z=1}^k \left\{ L_\phi(x) \|\theta_z - \theta_z^*\|_1 + \max_{\theta \in \mathbb{T}} |\phi(x; \theta)| |\pi_z - \pi_z^*| \right\} \\
& \leq G_2(x) \|\psi_k - \psi_k^*\|_1 \\
& \leq \sqrt{(m+1)k} G_2(x) \|\psi_k - \psi_k^*\|.
\end{aligned}$$

We can therefore take $G = \sqrt{(m+1)k} G_2$, as required.

Noting that $\mathcal{M}_k^\phi(\delta) = \left\{ f \in \mathcal{M}_k^\phi : \mathfrak{h}(f, f_0) \leq \delta \right\} \subset \mathcal{M}_k^\phi$, we can bound the entropy integral (3) by

$$\begin{aligned}
J(\delta) & \leq \delta \vee \int_0^\delta \sqrt{H_{[]} \left(u^2, \mathcal{M}_k^\phi, \mathcal{L}_1(\mathfrak{m}) \right)} du \\
& \leq \delta \vee \int_0^\delta \sqrt{\log \left\{ K \left(\frac{\|G\|_{\mathcal{L}_1(\mathfrak{m})} \text{diam}(\mathbb{S}_k)}{u^2} \right)^{(m+1)k} \right\}} du.
\end{aligned}$$

For $0 < u < 1$ the last square-root term is bounded by $C_1 + C_2 \sqrt{\log(1/u)}$ for suitable finite constants $C_1, C_2 > 0$ depending on k, G , and $\mathbb{S}_k$.

Jensen's inequality implies that

$$\begin{aligned} \int_0^\delta \sqrt{\log(1/u)} du &\leq \sqrt{\delta \int_0^\delta \log(1/u) du} \\ &= \sqrt{\delta \{\delta + \delta \log(1/\delta)\}} \leq \delta + \delta \sqrt{\log(1/\delta)}. \end{aligned}$$

Thus, we have

$$\begin{aligned} J(\delta) &\leq \delta \vee \left[(C_1 + C_2) \delta + C_2 \delta \sqrt{\log(1/\delta)} \right] \\ &\leq C_3 \delta + C_2 \delta \sqrt{\log(1/\delta)}, \end{aligned}$$

where $C_3 = 1 + C_1 + C_2$. Then, taking $\Psi(\delta) = C_3 \delta + C_2 \delta \sqrt{\log(1/\delta)}$, we observe $\Psi/\delta^2 = C_3 \delta^{-1} + C_2 \delta^{-1} \sqrt{\log(1/\delta)}$ is non-increasing, and thus Proposition 1 applies.

Substituting Ψ into $\sqrt{n} \delta_n^2 \geq c \Psi(\delta_n)$ yields the inequality

$$\sqrt{n} \delta_n \geq c C_3 + c C_2 \sqrt{\log(1/\delta_n)} \implies \frac{\sqrt{n} \delta_n}{c C_3 + c C_2 \sqrt{\log(1/\delta_n)}} \geq 1$$

which is satisfied if we take $\delta_n = C_4 \sqrt{\log n/n}$, for a sufficiently large $C_4 > 0$ and all sufficiently large n , since

$$\frac{\sqrt{n} \delta_n}{c C_3 + c C_2 \sqrt{\log(1/\delta_n)}} = \frac{C_4 \sqrt{\log n}}{c C_3 + (c/\sqrt{2}) C_2 \sqrt{\log(n/C_4^2) - \log \log n}}.$$

Then, the conclusion of Proposition 1 implies that for $\delta = \delta_n$,

$$\mathbb{P} \left(\frac{n}{\log n} \left| P_n \log \hat{f}_{k,n} - P_n \log f_0 \right| \geq C_4^2 \right) \leq c_1 \exp \left(-\frac{C_4^2}{c_1^2} \log n \right)$$

holds for an MLE $\hat{f}_{k,n} \in \arg \max_{f \in \mathcal{M}_k^\phi} P_n \log f$, since $P_n \log \hat{f}_{k,n} \geq P_n \log f$, for all $f \in \mathcal{M}_k^\phi$. Thus, by definition of boundedness in probability

$$\frac{n}{\log n} \left\{ P_n \log \hat{f}_{k,n} - P_n \log f_0 \right\} = O_{\mathbb{P}}(1).$$

By writing $\hat{f}_{k,n} = f(\cdot; \hat{\psi}_{k,n})$ for $\hat{\psi}_{k,n} \in \arg \min_{\psi_k \in \mathbb{S}_k} \ell_{k,n}(\psi_k)$, we have for each $k, l \geq k_0$,

$$\begin{aligned} &\frac{n}{\log n} \left\{ \min_{\psi_k \in \mathbb{S}_k} \ell_{k,n}(\psi_k) - \min_{\psi_l \in \mathbb{S}_l} \ell_{l,n}(\psi_l) \right\} \\ &= \frac{n}{\log n} \left\{ -P_n \log \hat{f}_{k,n} + P_n \log \hat{f}_{l,n} \right\} \\ &= -\frac{n}{\log n} \left\{ P_n \log \hat{f}_{k,n} - P_n \log f_0 \right\} + \frac{n}{\log n} \left\{ P_n \log \hat{f}_{l,n} - P_n \log f_0 \right\} \\ &= O_{\mathbb{P}}(1) + O_{\mathbb{P}}(1) = O_{\mathbb{P}}(1), \end{aligned}$$

as required. $\square$

Proof of Proposition 2

Fix $k \in [\bar{k}]$. Under A1* and A2–A4, Lemma 3 holds exactly as stated, and hence

$$\hat{L}_{k,n} \xrightarrow[n \rightarrow \infty]{\text{a.s.}} L_k.$$

Since $\bar{k}$ is finite, taking the intersection over all $k \in [\bar{k}]$ yields

$$\max_{k \in [\bar{k}]} \left| \hat{L}_{k,n} - L_k \right| \xrightarrow[n \rightarrow \infty]{\text{a.s.}} 0.$$

By B1, for each fixed k , $\text{pen}_{k,n} \rightarrow 0$ and thus, from finiteness of $\bar{k}$, we also have

$$\max_{k \in [\bar{k}]} \text{pen}_{k,n} \xrightarrow[n \rightarrow \infty]{} 0.$$

Define the penalised empirical criterion as $C_{k,n} = \hat{L}_{k,n} + \text{pen}_{k,n}$ and write $C_k = L_k$. Then, it follows that

$$\max_{k \in [\bar{k}]} |C_{k,n} - C_k| \xrightarrow[n \rightarrow \infty]{\text{a.s.}} 0.$$

Let $C^* = \min_{k \in [\bar{k}]} C_k$. If $\mathcal{K}^* = [\bar{k}]$ then the result is trivial. Otherwise, define a positive gap

$$\gamma = \min_{k \notin \mathcal{K}^*} (C_k - C^*) > 0,$$

where the strict positivity arises because $\bar{k}$ is finite and $\mathcal{K}^* \neq [\bar{k}]$. Set $\epsilon = \gamma/4$. Then, it holds that with probability one, there exists an n_0 such that for all $n \geq n_0$,

$$\max_{k \in [\bar{k}]} |C_{k,n} - C_k| < \epsilon.$$

Fix $n \geq n_0$. If $k \notin \mathcal{K}^*$, then we have

$$C_{k,n} \geq C_k - \epsilon \geq (C^* + \gamma) - \epsilon = C^* + \frac{3}{4}\gamma.$$

On the other hand, if $k^* \in \mathcal{K}^*$, we have

$$C_{k^*,n} \leq C_{k^*} + \epsilon = C^* + \epsilon = C^* + \frac{1}{4}\gamma.$$

Thus, for every $n \geq n_0$,

$$\min_{k \notin \mathcal{K}^*} C_{k,n} > \min_{k^* \in \mathcal{K}^*} C_{k^*,n}$$

and therefore every minimiser of $k \mapsto C_{k,n}$ is in $\mathcal{K}^*$, thus $\hat{k}_n \in \mathcal{K}^*$. Hence $\mathbb{P}(\hat{k}_n \in \mathcal{K}^*) \rightarrow 1$. $\square$

Proof of Proposition 3

Fix an arbitrary $\text{pen} \in \mathcal{P}$. For each k and l satisfying $1 \leq k < l \leq \bar{k}$, define

$$\Delta_{k,l}(n) = \frac{n}{\log n} \{ \text{pen}_{l,n} - \text{pen}_{k,n} \}.$$

By B2, $\Delta_{k,l}(n) \rightarrow \infty$ as $n \rightarrow \infty$. Since the set of choices of indices k and l is finite, it holds that

$$\Delta_{\min}(n) = \min_{1 \leq k < l \leq \bar{k}} \Delta_{k,l}(n) \rightarrow \infty.$$

Indeed, if $\Delta_{\min}(n)$ did not diverge, there would exist a subsequence $(n_j)_j$ and a constant $M < \infty$, such that $\Delta_{\min}(n_j) \leq M$, for all j . For each j , choose a pair (k_j, l_j) that attains the minimum. By finiteness of the set of indices, some pair (k, l) occurs infinitely often along the sequence $(k_j, l_j)_j$, hence along a further subsequence, we would have $\Delta_{k,l}(n_j) \leq M$, which contradicts the fact that $\Delta_{k,l}(n) \rightarrow \infty$.

Set $b_n = 1$ for the finitely many n for which $\Delta_{\min}(n) \leq 1$ and set $b_n = \sqrt{\Delta_{\min}(n)}$ otherwise. Then $b_n \rightarrow \infty$. For each $k \in [\bar{k}]$, define

$$\widetilde{\text{pen}}_{k,n} = \text{pen}_{1,n} + \frac{\text{pen}_{k,n} - \text{pen}_{1,n}}{b_n}.$$

For all sufficiently large n , $\Delta_{\min}(n) > 0$, so $\text{pen}_{1,n} < \text{pen}_{k,n}$ for every $k \geq 2$. Hence $\widetilde{\text{pen}}_{k,n} \geq \text{pen}_{1,n} \geq 0$, for all k and all sufficiently large n ; changing finitely many values, if needed, preserves B1 and B2 and makes the array non-negative for all n . By B1 and finiteness of $\bar{k}$, we have $\widetilde{\text{pen}}_{k,n} \rightarrow 0$ for each fixed k . Thus $\widetilde{\text{pen}}$ satisfies B1. Next, for fixed k and l , with $1 \leq k < l \leq \bar{k}$,

$$\frac{n}{\log n} \{ \widetilde{\text{pen}}_{l,n} - \widetilde{\text{pen}}_{k,n} \} = \frac{1}{b_n} \Delta_{k,l}(n) \geq \frac{1}{b_n} \Delta_{\min}(n) \rightarrow \infty,$$

because the last expression equals $\sqrt{\Delta_{\min}(n)}$ eventually. Hence $\widetilde{\text{pen}}$ satisfies B2, and consequently, $\widetilde{\text{pen}} \in \mathcal{P}$.

Finally, for all sufficiently large n , $b_n > 1$ and $\text{pen}_{1,n} < \text{pen}_{k,n}$ for every $k \geq 2$, so

$$\widetilde{\text{pen}}_{k,n} < \text{pen}_{k,n}, \quad k = 2, \dots, \bar{k},$$

while $\widetilde{\text{pen}}_{1,n} = \text{pen}_{1,n}$. Thus $\widetilde{\text{pen}} \preceq \text{pen}$ and the domination is strict in at least one coordinate eventually. Since $\text{pen} \in \mathcal{P}$ was arbitrary, no element of $\mathcal{P}$ can be minimal. $\square$

Proof of Proposition 5

We start by proving that $\mathcal{R}_n^* \gtrsim n^{-1}$. Since $\mathcal{M}_1 \subset \mathcal{F}$, we have

$$\mathcal{R}_n^* \gtrsim \inf_{\tilde{f}_n} \sup_{\mu \in [-b, b]} \mathbb{E}_{\phi(\cdot; \mu)} \left\{ \mathfrak{h}^2 \left(\tilde{f}_n, \phi(\cdot; \mu) \right) \right\}.$$

Fix $\Delta > 0$ and set $\mu_{\pm} = \pm \Delta / \sqrt{n}$, and observe that for large n , $\mu_{\pm} \in [-b, b]$. Let $f_{\pm} = \phi(\cdot; \mu_{\pm})$. It holds that

$$\int_{-\infty}^{\infty} \sqrt{f_{-}(x) f_{+}(x)} dx = \exp \left\{ -\frac{(\mu_{+} - \mu_{-})^2}{8} \right\} = \exp \left\{ -\frac{\Delta^2}{2n} \right\},$$

hence

$$\mathfrak{h}^2(f_{-}, f_{+}) = 1 - \int_{-\infty}^{\infty} \sqrt{f_{-}(x) f_{+}(x)} dx = 1 - \exp \left\{ -\frac{\Delta^2}{2n} \right\} \geq \frac{\Delta^2}{4n},$$

since $1 - \exp(-u) \geq u/2$ for all $u \in [0, 1]$, and $\Delta^2 < 2n$ for sufficiently large n . Next, we apply Le Cam's two point method (cf. Wainwright 2019, Sec. 15.2) to get

$$\begin{aligned} & \max \left\{ \mathbb{E}_{f_{-}} \mathfrak{h}^2(\tilde{f}_n, f_{-}), \mathbb{E}_{f_{+}} \mathfrak{h}^2(\tilde{f}_n, f_{+}) \right\} \\ & \geq c_1 \mathfrak{h}^2(f_{-}, f_{+}) \{1 - \|f_{-}^{\otimes n} - f_{+}^{\otimes n}\|_{\text{TV}}\}. \end{aligned}$$

where $f^{\otimes n}$ is the density of $\mathbf{X}_n$ consisting of IID replicates of X with density f , and $c_1 > 0$ is an absolute constant. Next, observe that

$$\begin{aligned} \mathfrak{h}^2(f_{-}^{\otimes n}, f_{+}^{\otimes n}) &= 1 - \int_{\mathbb{R}^n} \sqrt{f_{-}^{\otimes n}(\mathbf{x}_n) f_{+}^{\otimes n}(\mathbf{x}_n)} d\mathbf{x}_n \\ &= 1 - \prod_{i=1}^n \int_{-\infty}^{\infty} \sqrt{f_{-}(x_i) f_{+}(x_i)} dx_i. \end{aligned}$$

Thus,

$$\begin{aligned} \mathfrak{h}^2(f_{-}^{\otimes n}, f_{+}^{\otimes n}) &= 1 - \left\{ \int_{-\infty}^{\infty} \sqrt{f_{-}(x) f_{+}(x)} dx \right\}^n \\ &= 1 - \left(\exp \left\{ -\frac{\Delta^2}{2n} \right\} \right)^n = 1 - \exp \left\{ -\frac{\Delta^2}{2} \right\}. \end{aligned}$$

Using the comparison inequality $\|f_{-}^{\otimes n} - f_{+}^{\otimes n}\|_{\text{TV}} \leq \sqrt{2} \mathfrak{h}(f_{-}^{\otimes n}, f_{+}^{\otimes n})$, we get

$$\|f_{-}^{\otimes n} - f_{+}^{\otimes n}\|_{\text{TV}} \leq \sqrt{2} \sqrt{1 - \exp \{-\Delta^2/2\}}.$$

Choosing Δ sufficiently small so that $1 - \|f_{-}^{\otimes n} - f_{+}^{\otimes n}\|_{\text{TV}} \geq c_2 > 0$, we have

$$\max \left\{ \mathbb{E}_{f_{-}} \mathfrak{h}^2(\tilde{f}_n, f_{-}), \mathbb{E}_{f_{+}} \mathfrak{h}^2(\tilde{f}_n, f_{+}) \right\} \geq c_1 c_2 \frac{\Delta^2}{4n} = \frac{c}{n}.$$

Since $\mu_{\pm} \in [-b, b]$, this lower bound holds for every measurable estimator $\tilde{f}_n$ after taking the supremum over $f_0 \in \mathcal{F}$. Taking the infimum over all $\tilde{f}_n$ gives $\mathcal{R}_n^* \geq c/n$, proving the first inequality. The second inequality is immediate by definition of $\mathcal{R}_n^*$.

To proceed, we require Massart (2007, Thm. 7.11) which we reproduce in the context of this work. We first note that for any nonnegative $f, g \in \mathcal{L}_1(\mathfrak{m})$, we will write

$$\mathfrak{h}(f, g) = \left\{ \frac{1}{2} \int_{\mathbb{X}} \left(f^{1/2} - g^{1/2} \right)^2 d\mathfrak{m} \right\}^{1/2}.$$

If f and g are probability densities, then $\mathfrak{h}(f, g)$ is the usual Hellinger distance. We write $H_{[\cdot]}(\epsilon, \mathcal{F}, \mathfrak{h})$ for the $\mathfrak{h}$ -bracketing entropy of $\mathcal{F}$, where brackets $[f^L, f^U]$ may have endpoints $f^L, f^U \in \mathcal{L}_1(\mathfrak{m})$ that are nonnegative but not necessarily densities, and $\mathfrak{h}(f^L, f^U) \leq \epsilon$.

Theorem S.1 *Let $\mathbf{X}_n$ consist of IID replicates with common density f_0 with respect to $\mathfrak{m}$ on $\mathbb{X}$. Let $\{\mathcal{F}_k\}_{k \in \mathcal{K}}$ be an at most countable collection of models, where each $\mathcal{F}_k$ is a set of densities with respect to $\mathfrak{m}$. For each $k \in \mathcal{K}$ let $\hat{f}_{k,n} \in \arg \min_{f \in \mathcal{F}_k} \ell_n(f)$ be an MLE on $\mathcal{F}_k$. Assume the following:*

1. *For each $k \in \mathcal{K}$ there exists a countable subset $\mathcal{F}_k^\circ \subset \mathcal{F}_k$ such that for every $f \in \mathcal{F}_k$ there exists $(f_j)_{j \geq 1} \subset \mathcal{F}_k^\circ$ with $\log f_j(x) \rightarrow \log f(x)$ for every $x \in \mathbb{X}$.*
2. *For each $k \in \mathcal{K}$ there exists a function $\Psi_k : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}$ such that Ψ_k is nondecreasing and $\xi \mapsto \Psi_k(\xi)/\xi$ is nonincreasing on $(0, \infty)$, and such that for every $\delta > 0$ and every $f \in \mathcal{F}_k$, writing the δ -Hellinger ball*

$$\mathcal{F}_k(f, \delta) = \{g \in \mathcal{F}_k : \mathfrak{h}(g, f) \leq \delta\},$$

one has

$$\int_0^\delta \sqrt{H_{[\cdot]}(u, \mathcal{F}_k(f, \delta), \mathfrak{h})} du \leq \Psi_k(\delta).$$

3. *The sequence $(\rho_k)_{k \in \mathcal{K}}$ consists of nonnegative weights such that*

$$\Upsilon = \sum_{k \in \mathcal{K}} e^{-\rho_k} < \infty.$$

For each $k \in \mathcal{K}$ let $\xi_k > 0$ be the unique solution of

$$\Psi_k(\xi) = \sqrt{n} \xi^2.$$

Let $\text{pen}_{k,n} \geq 0$ be penalties and define the penalised selector

$$\hat{k}_n = \min \arg \min_{k \in \mathcal{K}} \left\{ \ell_n(\hat{f}_{k,n}) + \text{pen}_{k,n} \right\}, \hat{f}_n = \hat{f}_{\hat{k}_n,n}.$$

Then there exist absolute constants $\kappa > 0$ and $C > 0$ such that if, for all $k \in \mathcal{K}$,

$$\text{pen}_{k,n} \geq \kappa \left(\xi_k^2 + \frac{\rho_k}{n} \right),$$

a minimiser $\hat{k}_n$ exists and moreover

$$E_{f_0} \{ \mathfrak{h}^2(f_0, \hat{f}_n) \} \leq C \left[\inf_{k \in \mathcal{K}} \left\{ \mathfrak{R}(f_0 \| \mathcal{F}_k) + \text{pen}_{k,n} \right\} + \frac{\Upsilon}{n} \right],$$

where $\mathfrak{R}(f_0 \| \mathcal{F}_k) = \inf_{f \in \mathcal{F}_k} \mathfrak{R}(f_0 \| f)$.

To make use of Theorem S.1, we need the following lemmas.

Lemma S.3 Let $\mathcal{F}_1 = \{ \phi(\cdot; \mu) : \mu \in [-b, b] \}$ and

$$\mathcal{F}_2 = \left\{ f(\cdot; \theta) = \pi \phi(\cdot; \mu_1) + (1 - \pi) \phi(\cdot; \mu_2) : \theta = (\pi, \mu_1, \mu_2) \in \mathbb{T}_2 \right\},$$

where

$$\mathbb{T}_2 = [\underline{\pi}, 1 - \underline{\pi}] \times \{ (\mu_1, \mu_2) \in [-b, b]^2 : \mu_1 \leq \mu_2 \},$$

for some fixed $\underline{\pi} \in (0, 1/2)$, and write $d_1 = 1$ and $d_2 = 3$ for the parameter dimensions. For $k \in \{1, 2\}$ and $f \in \mathcal{F}_k$, define the δ -Hellinger ball

$$\mathcal{F}_k(f, \delta) = \{ g \in \mathcal{F}_k : \mathfrak{h}(g, f) \leq \delta \}.$$

Then, for each $k \in \{1, 2\}$, there exist constants $C_k < \infty$ and $D_k < \infty$ depending only on $(b, \underline{\pi})$ such that for every $f \in \mathcal{F}_k$, every $\delta \in (0, 1]$, and every $u \in (0, \delta]$,

$$H_{[\cdot]}(u, \mathcal{F}_k(f, \delta), \mathfrak{h}) \leq D_k \log \left(\frac{C_k \delta}{u} \right).$$

In particular, defining

$$B_k = \int_0^1 \sqrt{\log \left(\frac{C_k}{t} \right)} dt, \quad \Psi_k(\delta) = B_k \sqrt{D_k} \delta,$$

one has, for all $f \in \mathcal{F}_k$ and all $\delta \in (0, 1]$,

$$\int_0^\delta \sqrt{H_{[\cdot]}(u, \mathcal{F}_k(f, \delta), \mathfrak{h})} du \leq \Psi_k(\delta).$$

Proof Apply Gassiat & Van Handel (2012, Thm. III.4) to one-dimensional Gaussian location mixtures with location set $[-b, b]$ and at most two components, taking the centre of the Hellinger ball to be the density f in the present notation. That theorem gives a local bracketing bound for the square-root likelihood-ratio class $\{(g/f)^{1/2} : g \in \mathcal{F}_k, \mathfrak{h}(g, f) \leq \delta\}$ of the form

$$N_{[\cdot]} \left(\eta, \{(g/f)^{1/2} : g \in \mathcal{F}_k, \mathfrak{h}(g, f) \leq \delta\}, \mathcal{L}_2(f\mathfrak{m}) \right) \leq \left(\frac{C\delta}{\eta} \right)^{D_k},$$

for $0 < \eta \leq \delta \leq 1$, where D_k is finite and the constant C depends only on the compact location set, the maximum order, and the centre f . Because the centres considered here range over the compact families $\mathcal{F}_1$ and $\mathcal{F}_2$ with mixing weights in $[\underline{\pi}, 1 - \underline{\pi}]$, the same proof gives a uniform constant depending only on $(b, \underline{\pi})$.

If $[l, u]$ is such a bracket for $(g/f)^{1/2}$ in $\mathcal{L}_2(f\mathbf{m})$, then, after replacing l by $l \vee 0$, the bracket $[fl^2, fu^2]$ contains g . Moreover, on a local Hellinger ball with $\delta \leq 1$, the Hellinger width of this density bracket is bounded by a universal constant times $\|u - l\|_{\mathcal{L}_2(f\mathbf{m})}$. Adjusting constants therefore yields

$$H_{[\cdot]}(u, \mathcal{F}_k(f, \delta), \mathfrak{h}) \leq D_k \log \left(\frac{C_k \delta}{u} \right),$$

for $0 < u \leq \delta \leq 1$, uniformly in $f \in \mathcal{F}_k$, with C_k and D_k depending only on (b, π) . Integrating this bound and making the substitution $u = \delta t$ gives

$$\int_0^\delta \sqrt{H_{[\cdot]}(u, \mathcal{F}_k(f, \delta), \mathfrak{h})} du \leq \delta \sqrt{D_k} \int_0^1 \sqrt{\log \left(\frac{C_k}{t} \right)} dt = B_k \sqrt{D_k} \delta.$$

□

Lemma S.4 *Assume the setting of Lemma S.3. For $k \in \{1, 2\}$ define*

$$\Psi_k(\delta) = B_k \sqrt{D_k} \delta,$$

for $\delta > 0$, where $B_k < \infty$ and $D_k < \infty$ are as in Lemma S.3. Then Ψ_k is nondecreasing on $(0, \infty)$ and $\delta \mapsto \Psi_k(\delta)/\delta$ is nonincreasing on $(0, \infty)$. Moreover, the equation $\Psi_k(\xi) = \sqrt{n} \xi^2$ has the unique solution

$$\xi_k = \frac{B_k \sqrt{D_k}}{\sqrt{n}}.$$

In particular, since $D_k < \infty$ is fixed, one has $\xi_k^2 = O(n^{-1})$.

Proof The function $\Psi_k(\delta) = B_k \sqrt{D_k} \delta$ is linear in δ , hence nondecreasing, and $\Psi_k(\delta)/\delta = B_k \sqrt{D_k}$ is constant, hence nonincreasing. For $\xi > 0$, $\Psi_k(\xi) = \sqrt{n} \xi^2$ is equivalent to $B_k \sqrt{D_k} = \sqrt{n} \xi$, which directly yields the unique solution $\xi_k = B_k \sqrt{D_k} / \sqrt{n}$. □

To prove the final inequality, we apply Theorem S.1 with $\rho_1 = \rho_2 = 1$ and

$$\Upsilon = \sum_{k \in \{1, 2\}} \exp(-\rho_k) = 2/e.$$

Condition 1 of Theorem S.1 holds for $\mathcal{F}_1$ and $\mathcal{F}_2$ by taking parameters in countable dense subsets of the compact parameter spaces, since the Gaussian log-densities are pointwise continuous in the parameters. By Lemma S.3, Condition 2 of Theorem S.1 holds with $\Psi_k(\delta) = B_k \sqrt{D_k} \delta$. By Lemma S.4, the fixed point equation $\Psi_k(\xi) = \sqrt{n} \xi^2$ has the unique solution $\xi_k = B_k \sqrt{D_k} / \sqrt{n}$, so $\xi_k^2 = B_k^2 D_k / n$.

Define an AIC-type penalty $\text{pen}_{k,n}^{\text{AIC}} = ad_k/n$ with $a > 0$ chosen so that

$$\frac{ad_k}{n} \geq \kappa \left(\frac{B_k^2 D_k}{n} + \frac{1}{n} \right), \quad k \in \{1, 2\},$$

where $\kappa > 0$ is the absolute constant in Theorem S.1. Such a choice is possible since $d_k \in \{1, 3\}$ is fixed and $B_k, D_k < \infty$ depend only on b and $\underline{\pi}$. Recall that $\hat{k}_n^{\text{AIC}}$ is the corresponding penalised selector and $\hat{f}_n^{\text{AIC}} = \hat{f}_{\hat{k}_n^{\text{AIC}}, n}$.

Then Theorem S.1 yields, for every $f_0 \in \mathcal{F}$,

$$\mathbb{E}_{f_0} \{ \mathfrak{h}^2(f_0, \hat{f}_n^{\text{AIC}}) \} \leq C \left[\inf_{k \in \{1, 2\}} \left\{ \mathfrak{R}(f_0 \| \mathcal{F}_k) + \text{pen}_{k, n}^{\text{AIC}} \right\} + \frac{\mathcal{Y}}{n} \right],$$

where $C > 0$ is the absolute constant in Theorem S.1. Now, fix $f_0 \in \mathcal{F} = \mathcal{F}_1 \cup \mathcal{F}_2$ and let $k_0 \in \{1, 2\}$ be such that $f_0 \in \mathcal{F}_{k_0}$. Then $\mathfrak{R}(f_0 \| \mathcal{F}_{k_0}) = 0$, and hence

$$\mathbb{E}_{f_0} \{ \mathfrak{h}^2(f_0, \hat{f}_n^{\text{AIC}}) \} \leq C \left(\frac{ad_{k_0}}{n} + \frac{\mathcal{Y}}{n} \right) \leq \frac{C'}{n},$$

with $C' = C(ad_2 + \mathcal{Y})$ and $d_2 = 3$. Taking the supremum over $f_0 \in \mathcal{F}$ gives

$$\sup_{f_0 \in \mathcal{F}} \mathbb{E}_{f_0} \{ \mathfrak{h}^2(\hat{f}_n^{\text{AIC}}, f_0) \} \leq \frac{C'}{n},$$

as required. $\square$

Proof of Proposition 6

Recall that $\Lambda_{k, n} = \sup_{f \in \mathcal{F}_k} P_n \log f = -\inf_{f \in \mathcal{F}_k} \ell_n(f)$. Then $\hat{k}_n = 2$ implies

$$\Lambda_{2, n} - \Lambda_{1, n} \geq \Delta_n. \quad (\text{S.2})$$

For $\delta > 0$, we define the symmetric two-component density

$$x \mapsto f_\delta(x) = \frac{1}{2} \phi(x; -\delta) + \frac{1}{2} \phi(x; +\delta),$$

which belongs to $\mathcal{F}_2$ provided that $\delta \leq b$ and $\underline{\pi} \leq 1/2 \leq 1 - \underline{\pi}$. Let $f_* = \phi(\cdot; 0) \in \mathcal{F}_1$. We require the following result.

Lemma S.5 *There exist constants $\delta_0 \in (0, 1]$ and $0 < c_0 \leq C_0 < \infty$, depending only on b and $\underline{\pi}$, such that for all $0 < \delta \leq \delta_0$,*

$$c_0 \delta^4 \leq \mathfrak{h}^2(f_\delta, \mathcal{F}_1) \leq C_0 \delta^4, \quad c_0 \delta^4 \leq \mathfrak{R}(f_\delta \| f_*) \leq C_0 \delta^4,$$

where $f_\delta(x) = (1/2) \phi(x; -\delta) + (1/2) \phi(x; +\delta)$ and $f_* = \phi(\cdot; 0)$.

Proof Throughout, $X \sim \text{N}(0, 1)$ and $\mathbb{E}_*$ denotes expectation under the measure defined by f_* . Using $\phi(x; \pm\delta) = \phi(x; 0) \exp(\pm\delta x - \delta^2/2)$, we have, for each $x \in \mathbb{R}$,

$$\frac{f_\delta(x)}{f_*(x)} = \exp\left(-\frac{\delta^2}{2}\right) \cosh(\delta x). \quad (\text{S.3})$$

Let

$$r_\delta(x) = \exp\left(-\frac{\delta^2}{2}\right) \cosh(\delta x).$$

A Taylor expansion at $\delta = 0$ gives, for each fixed x ,

$$r_\delta(x) = 1 + \frac{\delta^2}{2}(x^2 - 1) + \delta^4\left(\frac{x^4}{24} - \frac{x^2}{4} + \frac{1}{8}\right) + O(\delta^6(1 + x^6)). \quad (\text{S.4})$$

as $\delta \downarrow 0$. To pass to integrals under f_* we will use domination. For $\delta_0 \in (0, 1]$ fixed, (S.3) implies, for all $|\delta| \leq \delta_0$,

$$0 < r_\delta(x) \leq \exp(\delta_0^2/2) \cosh(\delta_0 x) \leq \exp(\delta_0^2/2) \exp(\delta_0|x|),$$

so in particular $r_\delta(X)$ and $\log r_\delta(X)$ have finite moments of all orders under f_* , since $E_* \exp(\lambda|X|) < \infty$ for every $\lambda > 0$. The functions of (δ, x) that appear below are analytic in δ near zero, and their derivatives up to the required orders are bounded, for $|\delta| \leq \delta_0$, by a polynomial in $|x|$ times $\exp(\delta_0|x|)$. This dominating function is integrable under f_* , so the Taylor expansions below may be integrated term by term.

Write

$$\mathfrak{h}^2(f_\delta, f_*) = \frac{1}{2} \int f_*(x) \left(\sqrt{r_\delta(x)} - 1 \right)^2 dx = \frac{1}{2} E_* \left[\left(\sqrt{r_\delta(X)} - 1 \right)^2 \right].$$

From (S.4) we have $r_\delta(x) = 1 + \delta^2 a(x) + \delta^4 b(x) + O(\delta^6(1 + x^6))$ with $a(x) = (x^2 - 1)/2$ and $b(x) = x^4/24 - x^2/4 + 1/8$. Using $\sqrt{1+u} = 1 + \frac{1}{2}u - \frac{1}{8}u^2 + O(u^3)$ as $u \rightarrow 0$ and substituting $u = \delta^2 a(x) + \delta^4 b(x) + O(\delta^6(1 + x^6))$, we obtain the pointwise expansion

$$\sqrt{r_\delta(x)} - 1 = \frac{\delta^2}{2} a(x) + O(\delta^4(1 + x^4)) = \frac{\delta^2}{4} (x^2 - 1) + O(\delta^4(1 + x^4)),$$

as $\delta \downarrow 0$, for each fixed x . Squaring yields

$$\left(\sqrt{r_\delta(x)} - 1 \right)^2 = \frac{\delta^4}{16} (x^2 - 1)^2 + O(\delta^6(1 + x^6)).$$

The remainder term is dominated by a constant multiple of $\delta^6(1 + |x|^6)$ for $|\delta| \leq \delta_0$, and $(1 + |X|^6)$ is integrable under f_* . Hence, by the dominated convergence theorem,

$$\mathfrak{h}^2(f_\delta, f_*) = \frac{1}{2} \left\{ \frac{\delta^4}{16} E_*[(X^2 - 1)^2] + O(\delta^6) \right\} = \frac{\delta^4}{32} E_*[(X^2 - 1)^2] + O(\delta^6).$$

Since $E_*[(X^2 - 1)^2] = 2$ for $X \sim N(0, 1)$, we obtain

$$\mathfrak{h}^2(f_\delta, f_*) = \frac{\delta^4}{16} + O(\delta^6). \quad (\text{S.5})$$

as $\delta \downarrow 0$, and thus $\mathfrak{h}^2(f_\delta, f_*) \asymp \delta^4$ for sufficiently small δ .

Recall that $\mathcal{F}_1 = \{\phi(\cdot; \mu) : \mu \in [-b, b]\}$. Since $\mathfrak{h}$ is continuous in μ and $[-b, b]$ is compact, the infimum over μ is attained. We prove the lower bound by working in the Hilbert space $\mathcal{L}_2(f_* \mathbf{m})$. Set

$$q_\delta(x) = \sqrt{\frac{f_\delta(x)}{f_*(x)}} - 1, \quad p_\mu(x) = \sqrt{\frac{\phi(x; \mu)}{f_*(x)}} - 1.$$

Then

$$2\mathfrak{h}^2(f_\delta, \phi(\cdot; \mu)) = \|q_\delta - p_\mu\|_{\mathcal{L}_2(f_* \mathbf{m})}^2.$$

Let E and O denote the closed subspaces of even and odd functions in $\mathcal{L}_2(f_* \mathbf{m})$, and let Π_E and Π_O be the corresponding orthogonal projections. The expansion above gives the uniform $\mathcal{L}_2(f_* \mathbf{m})$ expansion

$$q_\delta = \frac{\delta^2}{4}(x^2 - 1) + O_2(\delta^4),$$

where $O_2(a_\delta)$ denotes a term whose $\mathcal{L}_2(f_* \mathbf{m})$ norm is bounded by a constant multiple of a_δ . Also, for $|\mu|$ small,

$$p_\mu = \frac{\mu}{2}x + \mu^2 \left(\frac{x^2}{8} - \frac{1}{4} \right) + O_2(|\mu|^3),$$

uniformly over $|\mu|$ in a sufficiently small neighbourhood of zero. Therefore there exist constants $\eta > 0$, $c_1 > 0$, and $C_1 < \infty$ such that, whenever $|\mu| \leq \eta$,

$$\|\Pi_O p_\mu\|_{\mathcal{L}_2(f_* \mathbf{m})} \geq c_1 |\mu|, \quad \|\Pi_E p_\mu\|_{\mathcal{L}_2(f_* \mathbf{m})} \leq C_1 \mu^2.$$

Since q_δ is even, if $M\delta^2 \leq |\mu| \leq \eta$, then

$$\|q_\delta - p_\mu\|_{\mathcal{L}_2(f_* \mathbf{m})}^2 \geq \|\Pi_O p_\mu\|_{\mathcal{L}_2(f_* \mathbf{m})}^2 \geq c_1^2 M^2 \delta^4.$$

On the other hand, if $|\mu| < M\delta^2$, then

$$\|\Pi_E(q_\delta - p_\mu)\|_{\mathcal{L}_2(f_* \mathbf{m})} \geq \frac{\delta^2}{4} \|x^2 - 1\|_{\mathcal{L}_2(f_* \mathbf{m})} - O(\delta^4) - C_1 M^2 \delta^4,$$

which is at least a positive constant times δ^2 for all sufficiently small δ . Hence the squared distance is bounded below by a positive constant times δ^4 in this region as well.

It remains to consider $|\mu| \geq \eta$. By compactness of $\{\mu \in [-b, b] : |\mu| \geq \eta\}$ and continuity of Hellinger distance,

$$\inf_{|\mu| \geq \eta} \mathfrak{h}^2(f_*, \phi(\cdot; \mu)) > 0.$$

Since $\mathfrak{h}(f_\delta, f_*) \rightarrow 0$, the same infimum with f_δ in place of f_* remains bounded away from zero for all sufficiently small δ , and hence is also bounded below by a constant multiple of δ^4 . Combining the three regions gives

$$\inf_{\mu \in [-b, b]} \mathfrak{h}^2(f_\delta, \phi(\cdot; \mu)) \geq c_0 \delta^4$$

for all sufficiently small δ , for some $c_0 > 0$. The upper bound $\mathfrak{h}^2(f_\delta, \mathcal{F}_1) \leq \mathfrak{h}^2(f_\delta, f_*) \leq C_0 \delta^4$ follows from (S.5). This proves the two-sided bound for $\mathfrak{h}^2(f_\delta, \mathcal{F}_1)$.

Next, by (S.3),

$$\mathfrak{K}(f_\delta \| f_*) = \int_{-\infty}^{\infty} f_\delta(x) \log \left(\frac{f_\delta(x)}{f_*(x)} \right) dx = \mathbb{E}_{f_\delta} [\log r_\delta(X)].$$

Using $\log r_\delta(x) = -\delta^2/2 + \log \cosh(\delta x)$ and the Taylor expansion $\log \cosh(u) = u^2/2 - u^4/12 + O(u^6)$ (for fixed u), we obtain for each fixed x ,

$$\log r_\delta(x) = \frac{\delta^2}{2}(x^2 - 1) - \frac{\delta^4}{12}x^4 + O(\delta^6 x^6).$$

As above, domination holds for $|\delta| \leq \delta_0$, and therefore we may integrate term-by-term to obtain

$$\mathfrak{K}(f_\delta \| f_*) = \frac{\delta^4}{4} + O(\delta^6),$$

as $\delta \downarrow 0$, and hence $\mathfrak{K}(f_\delta \| f_*) \asymp \delta^4$ for sufficiently small δ . Collecting the bounds we have the conclusion of the lemma, upon choosing δ_0 sufficiently small and adjusting constants. $\square$

Now define

$$\delta_n = \left(\frac{\Delta_n}{8C_0} \right)^{1/4}$$

and let $f_n = f_{\delta_n}$. Since $\Delta_n \rightarrow 0$, we have $\delta_n \rightarrow 0$, and for all n large enough, $\delta_n \leq \delta_0$ and $\delta_n \leq b$, so that $f_n \in \mathcal{F}_2$, and thus Lemma S.5 applies. In particular,

$$\mathfrak{h}^2(f_n, \mathcal{F}_1) \geq c_0 \delta_n^4 = \frac{c_0}{8C_0} \Delta_n, \quad \mathfrak{K}(f_n \| f_*) \leq C_0 \delta_n^4 = \frac{\Delta_n}{8}. \quad (\text{S.6})$$

In the following step, we work under the law f_n . We keep P_n for the empirical measure $P_n g = n^{-1} \sum_{i=1}^n g(X_i)$, and we write P_{f_n} and E_{f_n} for probability and expectation under f_n . Recall that $\hat{k}_n = 2$ implies $\Lambda_{2,n} - \Lambda_{1,n} \geq \Delta_n$, and since $f_* \in \mathcal{F}_1$ we have $\Lambda_{1,n} \geq P_n \log f_*$. Hence

$$\{\hat{k}_n = 2\} \subseteq \{\Lambda_{2,n} - P_n \log f_* \geq \Delta_n\}. \quad (\text{S.7})$$

We decompose

$$\Lambda_{2,n} - P_n \log f_* = R_n + P_n \log \left(\frac{f_n}{f_*} \right), \quad (\text{S.8})$$

where $R_n = \Lambda_{2,n} - P_n \log f_n \geq 0$ since $\Lambda_{2,n} = \sup_{f \in \mathcal{F}_2} P_n \log f \geq P_n \log f_n$, and $f_n \in \mathcal{F}_2$. We control the two terms on the right-hand side of (S.8), separately.

Firstly, under f_n the model $\mathcal{F}_2$ is correctly specified with a finite-dimensional parameter on the compact set $\mathbb{T}_2$. Moreover, $f_n = f_{\delta_n}$ has component means $\mu_1 = -\delta_n$ and $\mu_2 = \delta_n$, so its distance to the boundary $\{\mu_1 = \mu_2\}$ is $|\mu_2 - \mu_1| = 2\delta_n$. Since $n\Delta_n \rightarrow \infty$, we have $n^2\Delta_n \rightarrow \infty$, and therefore

$$\sqrt{n} \delta_n = \left(\frac{n^2 \Delta_n}{8C_0} \right)^{1/4} \rightarrow \infty.$$

Thus the true parameter is asymptotically separated from $\{\mu_1 = \mu_2\}$ on the local $n^{-1/2}$ scale. To control R_n , let $\hat{f}_{2,n} \in \arg \max_{f \in \mathcal{F}_2} P_n \log f$, so that $\Lambda_{2,n} = P_n \log \hat{f}_{2,n}$ and hence

$$R_n = \Lambda_{2,n} - P_n \log f_n = P_n \log \left(\frac{\hat{f}_{2,n}}{f_n} \right) \geq 0.$$

By Lemma S.3 (with $k = 2$), the local Hellinger bracketing entropy satisfies

$$H_{[]} (u, \mathcal{F}_2(f_n, \delta), \mathfrak{h}) \leq d_2 \log \left(\frac{C_2 \delta}{u} \right),$$

for $0 < u \leq \delta \leq 1$, and therefore one may take $\Psi_2(\delta) = B_2 \sqrt{d_2} \delta$ in the notation of van de Geer (2000, Thm. 7.4). Let $\xi_2 = B_2 \sqrt{d_2} / \sqrt{n}$ be the corresponding fixed point from Lemma S.4. Then van de Geer (2000, Thm. 7.4) implies that there exists a universal constant $c > 0$ such that one may take $\delta_{2,n} = c \xi_2$, and van de Geer (2000, Cor. 7.5) yields a constant $c_1 < \infty$ (depending only on b and $\underline{\pi}$ through the preceding entropy bounds) such that, for all $\delta \geq \delta_{2,n}$,

$$\mathbb{P}_{f_n} (R_n \geq \delta^2) \leq c_1 \exp \left(-\frac{n\delta^2}{c_1^2} \right).$$

Since $n\Delta_n \rightarrow \infty$ and $\delta_{2,n}^2 \asymp n^{-1}$, we have $\eta\Delta_n \geq \delta_{2,n}^2$ for all sufficiently large n and every fixed $\eta > 0$. Therefore, for any fixed $\eta > 0$, taking $\delta = \sqrt{\eta\Delta_n}$ yields

$$\mathbb{P}_{f_n} (R_n \geq \eta\Delta_n) \leq c_1 \exp \left(-\frac{\eta n\Delta_n}{c_1^2} \right) \rightarrow 0,$$

and hence $R_n = o_{\mathbb{P}_{f_n}}(\Delta_n)$.

Next, let

$$Z_{i,n} = \log \left(\frac{f_n(X_i)}{f_*(X_i)} \right),$$

for $(X_i)_{i=1}^n$ IID with common PDF f_n . Then $P_n \log(f_n/f_*) = n^{-1} \sum_{i=1}^n Z_{i,n}$ and $\mathbb{E}_{f_n}(Z_{1,n}) = \mathfrak{K}(f_n \| f_*) \leq \Delta_n/8$ by (S.6). Furthermore, by the Taylor expansion of $\log r_\delta(x)$ used in Lemma S.5 (with $\delta = \delta_n$) and the same domination argument, one has $\text{Var}_{f_n}(Z_{1,n}) \lesssim \delta_n^4 \lesssim \Delta_n$ for all sufficiently large n . Consequently,

$$\text{Var}_{f_n} \left\{ P_n \log \left(\frac{f_n}{f_*} \right) \right\} = \frac{1}{n} \text{Var}_{f_n}(Z_{1,n}) \lesssim \frac{\Delta_n}{n}.$$

By Chebyshev's inequality,

$$\begin{aligned} & \mathbb{P}_{f_n} \left(P_n \log \left(\frac{f_n}{f_*} \right) \geq \frac{\Delta_n}{2} \right) \\ & \leq \mathbb{P}_{f_n} \left(P_n \log \left(\frac{f_n}{f_*} \right) - \mathbb{E}_{f_n}(Z_{1,n}) \geq \frac{3\Delta_n}{8} \right) \\ & \lesssim \frac{1}{n\Delta_n} \rightarrow 0, \end{aligned}$$

since $n\Delta_n \rightarrow \infty$. Combining (S.7) and (S.8), we obtain

$$\mathbb{P}_{f_n}(\hat{k}_n = 2) \leq \mathbb{P}_{f_n} \left(R_n \geq \frac{\Delta_n}{2} \right) + \mathbb{P}_{f_n} \left(P_n \log \left(\frac{f_n}{f_*} \right) \geq \frac{\Delta_n}{2} \right) \rightarrow 0,$$

and therefore $\mathbb{P}_{f_n}(\hat{k}_n = 1) \rightarrow 1$.

On the event $\{\hat{k}_n = 1\}$, we have $\hat{f}_n \in \mathcal{F}_1$, and therefore

$$\mathfrak{h}^2(\hat{f}_n, f_n) \geq \inf_{g \in \mathcal{F}_1} \mathfrak{h}^2(g, f_n) = \mathfrak{h}^2(f_n, \mathcal{F}_1).$$

Taking expectations under f_n and using (S.6) yields

$$\mathbb{E}_{f_n} \{\mathfrak{h}^2(\hat{f}_n, f_n)\} \geq \mathfrak{h}^2(f_n, \mathcal{F}_1) \mathbb{P}_{f_n}(\hat{k}_n = 1) \geq \frac{c_0}{8C_0} \Delta_n (1 + o(1)).$$

Since $f_n \in \mathcal{F}$, we conclude that

$$\sup_{f_0 \in \mathcal{F}} \mathbb{E}_{f_0} \{\mathfrak{h}^2(\hat{f}_n, f_0)\} \geq \mathbb{E}_{f_n} \{\mathfrak{h}^2(\hat{f}_n, f_n)\} \geq c \Delta_n,$$

for all n large enough, with $c = c_0/(16C_0)$, since $1 + o(1) \geq 1/2$. This completes the proof. $\square$

Proof of Proposition 7

Recall that $\Lambda_{k,n} = \sup_{f \in \mathcal{M}_k} P_n \log f = -\inf_{f \in \mathcal{M}_k} \ell_n(f)$, so that $\hat{k}_n = 2$ implies $\Lambda_{2,n} - \Lambda_{1,n} \geq \Delta_n$, as in (S.2). Assume by contradiction that $\hat{k}_n$ is consistent but $\Delta_n \not\rightarrow 0$. If there exist $\epsilon > 0$ and a subsequence $(n_m)_m$ such that $\Delta_{n_m} \leq -\epsilon$ for all m , then, since $\mathcal{M}_1 \subset \mathcal{M}_2$, we have $\Lambda_{2,n_m} - \Lambda_{1,n_m} \geq 0 > \Delta_{n_m}$ for every m . Hence $\hat{k}_{n_m} = 2$ for every m , contradicting pointwise consistency at $f_* \in \mathcal{M}_1$. Thus, after passing to a subsequence if necessary, there exist $\epsilon > 0$ and a subsequence $(n_m)_m$ such that $\Delta_{n_m} \geq \epsilon$ for all m . For $\delta > 0$, let f_δ and f_* be as in Lemma S.5. Since $f_* \in \mathcal{M}_1$, Lemma S.5 gives $\mathfrak{R}(f_\delta \| f_*) \rightarrow 0$ as $\delta \downarrow 0$, hence we may fix $\delta \in (0, b]$ such that

$$D_\delta = \inf_{g \in \mathcal{M}_1} \mathfrak{R}(f_\delta \| g) \leq \mathfrak{R}(f_\delta \| f_*) \leq \frac{\epsilon}{4}.$$

Write $\mathbb{P}_\delta$ and $\mathbb{E}_\delta$ for probability and expectation under the law with density f_δ . Because $\mathcal{M}_1$ and $\mathcal{M}_2$ are compact parametric families and $f \mapsto \log f$ admits an integrable envelope under f_δ , the classes $\{\log f : f \in \mathcal{M}_k\}$ are $\mathbb{P}_\delta$ -GC, via Lemma 1. Consequently,

$$\sup_{f \in \mathcal{M}_k} |P_n \log f - \mathbb{E}_\delta \log f| \rightarrow 0,$$

in $\mathbb{P}_\delta$ -probability, for each $k \in \{1, 2\}$, and therefore

$$\Lambda_{k,n} \rightarrow \Lambda_k = \sup_{f \in \mathcal{M}_k} \mathbb{E}_\delta \log f$$

in $\mathbb{P}_\delta$ -probability, for each $k \in \{1, 2\}$, as well. Since $f_\delta \in \mathcal{M}_2$, we have $\Lambda_2 = \mathbb{E}_\delta \log f_\delta$ and

$$\Lambda_1 = \sup_{g \in \mathcal{M}_1} \mathbb{E}_\delta \log g = \mathbb{E}_\delta \log f_\delta - \inf_{g \in \mathcal{M}_1} \mathfrak{R}(f_\delta \| g) = \mathbb{E}_\delta \log f_\delta - D_\delta,$$

so $\Lambda_2 - \Lambda_1 = D_\delta \leq \epsilon/4$. Hence $\Lambda_{2,n} - \Lambda_{1,n} \rightarrow D_\delta$ in P_δ -probability, and thus

$$P_\delta \left(\Lambda_{2,n_m} - \Lambda_{1,n_m} < \frac{\epsilon}{2} \right) \rightarrow 1.$$

On this event we have $\Lambda_{2,n_m} - \Lambda_{1,n_m} < \Delta_{n_m}$, so $\hat{k}_{n_m} \neq 2$, hence $\hat{k}_{n_m} = 1$. Therefore $P_\delta(\hat{k}_{n_m} = 1) \rightarrow 1$, contradicting the assumed pointwise consistency at $f_\delta \in \mathcal{M}_2 \setminus \mathcal{M}_1$. This proves that $\Delta_n \rightarrow 0$.

Next, assume by contradiction that $\hat{k}_n$ is consistent but $n\Delta_n \not\rightarrow \infty$. Then there exist $M < \infty$ and a subsequence $(n_m)_m$ such that $n_m\Delta_{n_m} \leq M$ for all m . We now work under $f_* \in \mathcal{M}_1$ and write P_* and E_* for probability and expectation under f_* . Let

$$S_n = P_n(X^2 - 1) = \frac{1}{n} \sum_{i=1}^n (X_i^2 - 1),$$

and $\delta_n = \min \left\{ [S_n]_+^{1/2}, b \right\}$, and note that $f_{\delta_n} \in \mathcal{M}_2$ for all n . Since $\Lambda_{2,n} = \sup_{f \in \mathcal{M}_2} P_n \log f \geq P_n \log f_{\delta_n}$, we have

$$\Lambda_{2,n} - \Lambda_{1,n} \geq P_n \log \left(\frac{f_{\delta_n}}{f_*} \right) - \{ \Lambda_{1,n} - P_n \log f_* \}. \quad (\text{S.9})$$

We seek a lower bound for the first term. Using (S.3) and the expansion of $\log r_\delta(x)$ in the proof of Lemma S.5, there exists $\delta_0 \in (0, 1]$ such that for all $|\delta| \leq \delta_0$,

$$\log \left(\frac{f_\delta(x)}{f_*(x)} \right) = \frac{\delta^2}{2}(x^2 - 1) - \frac{\delta^4}{4} - \frac{\delta^4}{12}(x^4 - 3) + O(\delta^6 x^6).$$

Under f_* we have $S_n = O_{P_*}(n^{-1/2})$, hence $\delta_n = O_{P_*}(n^{-1/4})$ and thus $P_*(\delta_n \leq \delta_0) \rightarrow 1$. On the event $\{\delta_n \leq \delta_0\}$, the preceding expansion and the moment bounds under f_* imply

$$P_n \log \left(\frac{f_{\delta_n}}{f_*} \right) = \frac{\delta_n^2}{2} S_n - \frac{\delta_n^4}{4} + o_{P_*}(n^{-1}) = \frac{1}{4} [S_n]_+^2 + o_{P_*}(n^{-1}),$$

and therefore

$$nP_n \log \left(\frac{f_{\delta_n}}{f_*} \right) = \frac{1}{4} [\sqrt{n} S_n]_+^2 + o_{P_*}(1). \quad (\text{S.10})$$

Since $\sqrt{n} S_n$ converges weakly to $N(0, 2)$ under f_* , it follows from (S.10) that for each fixed $A > 0$,

$$\liminf_{n \rightarrow \infty} P_* \left(nP_n \log \left(\frac{f_{\delta_n}}{f_*} \right) \geq A \right) = \rho_A > 0. \quad (\text{S.11})$$

Next, we control the term $\Lambda_{1,n} - P_n \log f_*$ directly. Writing $\bar{X}_n = P_n X$, the normal location likelihood gives

$$\Lambda_{1,n} - P_n \log f_* = \sup_{\mu \in [-b, b]} \left\{ \mu \bar{X}_n - \frac{1}{2} \mu^2 \right\}.$$

On the event $|\bar{X}_n| \leq b$, this supremum is $\bar{X}_n^2/2$. The multivariate central limit theorem gives

$$(\sqrt{n} \bar{X}_n, \sqrt{n} S_n) \rightsquigarrow (Z_1, Z_2),$$

where $Z_1 \sim N(0, 1)$, $Z_2 \sim N(0, 2)$, and Z_1 and Z_2 are independent, since $E_*[X(X^2 - 1)] = 0$. Choose constants $q > 0$ and $t > 0$ such that $t^2/4 - q^2/2 > M + 1$. The event $\{|Z_1| < q, Z_2 > t\}$ has positive probability. Combining (S.10) with the preceding likelihood identity and the joint central limit theorem, we obtain

$$\liminf_{m \rightarrow \infty} P_*(n_m(\Lambda_{2,n_m} - \Lambda_{1,n_m}) > M + 1) > 0.$$

Since $n_m \Delta_{n_m} \leq M$, this implies $\liminf_{m \rightarrow \infty} P_*(\hat{k}_{n_m} = 2) > 0$. This contradicts the assumed pointwise consistency at $f_* \in \mathcal{M}_1$. Therefore $n \Delta_n \rightarrow \infty$, and the proof is complete. $\square$

References

- Atienza, N., Garcia-Heras, J., Munoz-Pichardo, J., & Villa, R. (2007). On the consistency of mle in finite mixture models of exponential families. *Journal of Statistical Planning and Inference*, 137, 496–505.
- Gassiat, E. & Van Handel, R. (2012). Consistent order estimation and minimal penalties. *IEEE Transactions on Information Theory*, 59, 1115–1128.
- Hinderer, K., Rieder, U., & Stieglitz, M. (2016). *Dynamic Optimization*. Springer.
- Massart, P. (2007). *Concentration Inequalities and Model Selection: École d'Été de Probabilités de Saint-Flour XXXIII - 2003*, volume 1896 of *Lecture Notes in Mathematics*. Berlin, Heidelberg: Springer. eBook ISBN: 978-3-540-48503-2.
- van de Geer, S. (2000). *Empirical Processes in M-estimation*, volume 6. Cambridge university press.
- Wainwright, M. J. (2019). *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Number 48 in Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press. Hardback ISBN: 9781108498029.