Supplementary Materials
WASP: Scalable Bayes via barycenters of subset posteriors

Sanvesh Srivastava *[1,2], Volkan Cevher †[3], Quoc Tran-Dinh ‡[3] and David B. Dunson §[1]

[1]*Department of Statistical Science, Duke University, Durham, North Carolina, USA*
[2]*Statistical and Applied Mathematical Sciences Institute, Durham, North Carolina, USA*
[3]*Laboratory for Information and Inference Systems, École Polytechnique Fédérale de Lausanne, Switzerland*

1 Proof of Lemma 3.1

Proof Given $\epsilon > 0$, let $\mathcal{U}_\epsilon = \{\theta : \| \theta - \theta_0 \| < \epsilon\}$ be a neighborhood of θ_0. Since $\mathcal{P}_2(\boldsymbol{\Theta})$ includes probability measures that are parameterized by finite dimensional $\theta \in \boldsymbol{\Theta} \subset \mathbb{R}^D$, there exists a test function $\Phi_n = \Phi_n(X_{1:n})$ for testing $H_0 : \theta = \theta_0$ against $H : \theta \in \mathcal{U}_\epsilon^c$ and universal constants B and b such that

$$P_{\theta_0}(\Phi_n) \leqslant B \exp(-bn) \text{ and } \sup_{\theta \in \mathcal{U}_\epsilon^c} P_\theta(1 - \Phi_n) \leqslant B \exp(-bn); \tag{1}$$

see Le Cam (1986); Ghosal et al. (2000) for details. The definition of W_2 and $\mathcal{U}_\epsilon$ imply that

$$\begin{aligned}
W_2^2(\delta_{\theta_0}, \boldsymbol{\Pi}_n(\cdot|X_{1:n})) &= \int_{\boldsymbol{\Theta}} \| \theta - \theta_0 \|^2 d\, \boldsymbol{\Pi}_n(\theta \,|X_{1:n}) \\
&= \int_{\mathcal{U}_\epsilon \cap \boldsymbol{\Theta}_\epsilon} \| \theta - \theta_0 \|^2 d\, \boldsymbol{\Pi}_n(\theta \,|X_{1:n}) + \int_{\mathcal{U}_\epsilon^c \cap \boldsymbol{\Theta}_\epsilon} \| \theta - \theta_0 \|^2 d\, \boldsymbol{\Pi}_n(\theta \,|X_{1:n}) \\
&+ \int_{\boldsymbol{\Theta}_\epsilon^c} \| \theta - \theta_0 \|^2 d\, \boldsymbol{\Pi}_n(\theta \,|X_{1:n}) \\
&\overset{(i)}{\leqslant} \epsilon^2 + \int_{\mathcal{U}_\epsilon^c \cap \boldsymbol{\Theta}_\epsilon} \| \theta - \theta_0 \|^2 d\, \boldsymbol{\Pi}_n(\theta \,|X_{1:n}) + \int_{\boldsymbol{\Theta}_\epsilon^c} \| \theta - \theta_0 \|^2 d\, \boldsymbol{\Pi}_n(\theta \,|X_{1:n}) \\
&\overset{(ii)}{\leqslant} \epsilon^2 + 4M_\epsilon^2 \int_{\mathcal{U}_\epsilon^c \cap \boldsymbol{\Theta}_\epsilon} d\, \boldsymbol{\Pi}_n(\theta \,|X_{1:n}) + \int_{\boldsymbol{\Theta}_\epsilon^c} \| \theta - \theta_0 \|^2 d\, \boldsymbol{\Pi}_n(\theta \,|X_{1:n}) \\
&\leqslant \epsilon^2 + 4M_\epsilon^2\, \boldsymbol{\Pi}_n(\mathcal{U}_\epsilon^c \,|X_{1:n}) + \int_{\boldsymbol{\Theta}_\epsilon^c} \| \theta - \theta_0 \|^2 d\, \boldsymbol{\Pi}_n(\theta \,|X_{1:n}).
\end{aligned} \tag{2}$$

*ss602@stat.duke.edu
†volkan.cevher@epfl.ch
‡quoc.trandinh@epfl.ch
§dunson@duke.edu

The first ϵ^2 term in (i) follows from the definition of $\mathcal{U}_\epsilon$. Theorem 2.1 (b) implies that $\| \theta - \theta_0 \|^2 \leqslant 4M_\epsilon^2$ for $\theta \in \boldsymbol{\Theta}_\epsilon$ in (ii).

We show that the second term in (16) goes to zero a.s. $[P_{\theta_0}^\infty]$.

$$\begin{aligned}\Pi_n(\mathcal{U}_\epsilon^c | X_{1:n}) &= \frac{\int_{\mathcal{U}_\epsilon^c} \prod_{i=1}^n p(X_i|\,\theta) d\,\Pi_n(\theta)}{\int_{\boldsymbol{\Theta}} \prod_{i=1}^n p(X_i|\,\theta) d\,\Pi_n(\theta)} = \frac{\int_{\mathcal{U}_\epsilon^c} \prod_{i=1}^n \frac{p(X_i|\,\theta)}{p(X_i|\,\theta_0)} d\,\Pi_n(\theta)}{\int_{\boldsymbol{\Theta}} \prod_{i=1}^n \frac{p(X_i|\,\theta)}{p(X_i|\,\theta_0)} d\,\Pi_n(\theta)} (\Phi_n + 1 - \Phi_n) \\ &\leqslant \Phi_n + \frac{(1-\Phi_n) \int_{\mathcal{U}_\epsilon^c} \prod_{i=1}^n \frac{p(X_i|\,\theta)}{p(X_i|\,\theta_0)} d\,\Pi_n(\theta)}{\int_{\boldsymbol{\Theta}} \prod_{i=1}^n \frac{p(X_i|\,\theta)}{p(X_i|\,\theta_0)} d\,\Pi_n(\theta)}. \end{aligned} \tag{3}$$

Using (15) and Markov's inequality, $\Phi_n \to 0$ a.s. $[P_{\theta_0}^\infty]$. To prove that $\Pi_n(\theta \in \mathcal{U}_\epsilon^c | X_{1:n}) \to 0$ a.s. $[P_{\theta_0}^\infty]$, it is sufficient to show that

(a) $P_{\theta_0}\left[(1-\Phi_n) \int_{\mathcal{U}_\epsilon^c} \prod_{i=1}^n \frac{p(X_i|\,\theta)}{p(X_i|\,\theta_0)} d\,\Pi_n(\theta)\right] \leqslant B \exp(-n\beta)$ for some β;

(b) $\exp(n\beta) \int_{\boldsymbol{\Theta}} \prod_{i=1}^n \frac{p(X_i|\,\theta)}{p(X_i|\,\theta_0)} d\,\Pi_n(\theta) \to \infty$ a.s. $[P_{\theta_0}^\infty]$ for every β.

To show (a), note that

$$\begin{aligned} &P_{\theta_0}\left[(1-\Phi_n) \int_{\mathcal{U}_\epsilon^c} \prod_{i=1}^n \frac{p(X_i|\,\theta)}{p(X_i|\,\theta_0)} d\,\Pi_n(\theta)\right] \\ &= \int (1-\Phi_n) \int_{\mathcal{U}_\epsilon^c} \prod_{i=1}^n \frac{p(X_i|\,\theta)}{p(X_i|\,\theta_0)} d\,\Pi_n(\theta) \left[\prod_{i=1}^n p(X_i|\,\theta_0)\right] d\nu^n(X_{1:n}) \\ &\overset{(i)}{=} \int_{\mathcal{U}_\epsilon^c} \int (1-\Phi_n) \prod_{i=1}^n \frac{p(X_i|\,\theta)}{p(X_i|\,\theta_0)} \left[\prod_{i=1}^n p(X_i|\,\theta_0) d\nu(X_i)\right] d\,\Pi_n(\theta) \\ &= \int_{\mathcal{U}_\epsilon^c} \left[\int (1-\Phi_n) \prod_{i=1}^n p(X_i|\,\theta) d\nu(X_i)\right] d\,\Pi_n(\theta) \\ &\overset{(ii)}{\leqslant} \sup_{\theta \in \mathcal{U}_\epsilon^c} P_\theta (1-\Phi_n)\, \Pi_n(\mathcal{U}_\epsilon^c) \leqslant B \exp(-bn). \end{aligned} \tag{4}$$

Fubini's theorem implies (i) and (ii) follows from (15). To show (b), given any $\beta > 0$, choose $\epsilon < \beta$ so that if $\mathsf{KL}(p(\cdot|\,\theta_0) \| p(\cdot|\,\theta)) < \epsilon$, then

$$\sum_{i=1}^n \log \frac{p(X_i|\,\theta)}{p(X_i|\,\theta_0)} + \beta n = n\left[\beta - n^{-1} \sum_{i=1}^n \log \frac{p(X_i|\,\theta_0)}{p(X_i|\,\theta)}\right] > n(\beta - \epsilon) \to \infty. \tag{5}$$

a.s. $[P_{\theta_0}^\infty]$. The null set N involved in (19) may depend on θ. To show that (19) is true for all $\theta \in \boldsymbol{\Theta}$, we define a product space $E = \{(\omega, \theta) : \exp(n\beta) \prod_{i=1}^n \frac{p(X_i|\,\theta)}{p(X_i|\,\theta_0)} \to \infty\}$, then (19) shows that $P_{\theta_0}^\infty(E_\theta) = 1$ for all $\theta \in \boldsymbol{\Theta}$. Another application of Fubini's theorem shows that there exists N in the probability space on which X_ns are defined such that $P_{\theta_0}^\infty(N) = 0$ and if $\omega \notin N$, then $\Pi_j(E^\omega) = 1$ for all j; therefore,

$\exp(n\beta)\prod_{i=1}^{n}\frac{p(X_i|\theta)}{p(X_i|\theta_0)} \to \infty$ for all $\omega \notin E^{\omega}$, which in turn implies that

$$\exp(n\beta)\int\prod_{i=1}^{n}\frac{p(X_i|\theta)}{p(X_i|\theta_0)}d\Pi_n(\theta) \geqslant \int_{\mathcal{K}_\epsilon(\theta_0)}\exp\left(n\beta+\sum_{i=1}^{n}\log\frac{p(X_i|\theta)}{p(X_i|\theta_0)}\right)d\Pi_n(\theta) \to \infty \tag{6}$$

a.s. $[P_{\theta_0}^{\infty}]$ by Fatou's lemma, since $\liminf_{n\to\infty}\Pi_n(\mathcal{K}_\epsilon(\theta_0)) > 0\ \forall\epsilon > 0$.

We now follow a strategy similar to (17) to prove that the third term in (16) goes to zero a.s. $[P_{\theta_0}^{\infty}]$. Using (18) and (20), it is enough to show that $P_{\theta_0}\left[(1-\Phi_n)\int_{\boldsymbol{\Theta}_\epsilon^c}\|\theta-\theta_0\|^2\prod_{i=1}^{n}\frac{p(X_i|\theta)}{p(X_i|\theta_0)}d\Pi_n(\theta)\right] \leqslant B\exp(-n\beta)$ for some β. Following (18),

$$\begin{aligned}
&P_{\theta_0}\left[(1-\Phi_n)\int_{\boldsymbol{\Theta}_\epsilon^c}\|\theta-\theta_0\|^2\prod_{i=1}^{n}\frac{p(X_i|\theta)}{p(X_i|\theta_0)}d\Pi_n(\theta)\right] = \\
&= \int(1-\Phi_n)\int_{\boldsymbol{\Theta}_\epsilon^c}\|\theta-\theta_0\|^2\prod_{i=1}^{n}\frac{p(X_i|\theta)}{p(X_i|\theta_0)}d\Pi_n(\theta)\left[\prod_{i=1}^{n}p(X_i|\theta_0)\right]d\nu^n(X_{1:n}) \\
&\overset{(i)}{=} \int_{\boldsymbol{\Theta}_\epsilon^c}\|\theta-\theta_0\|^2\int(1-\Phi_n)\prod_{i=1}^{n}\frac{p(X_i|\theta)}{p(X_i|\theta_0)}\left[\prod_{i=1}^{n}p(X_i|\theta_0)d\nu(X_i)\right]d\Pi_n(\theta) \\
&= \int_{\boldsymbol{\Theta}_\epsilon^c}\|\theta-\theta_0\|^2\left[\int(1-\Phi_n)\prod_{i=1}^{n}p(X_i|\theta)d\nu(X_i)\right]d\Pi_n(\theta) \\
&\leqslant \sup_{\theta\in\mathcal{U}_\epsilon^c}P_\theta(1-\Phi_n)\int_{\boldsymbol{\Theta}_\epsilon^c}\|\theta-\theta_0\|^2 d\Pi_n(\theta) \leqslant B\exp(-bn)\epsilon^2 \to 0
\end{aligned} \tag{7}$$

a.s. $[P_{\theta_0}^{\infty}]$. The last inequality in (21) follows from Assumption 2.1 (a).

Finally, Assumption (A2) implies that $\exists n_0$ such that $\forall n > n_0\ M_\epsilon^2\exp(-bn) < \epsilon^2$. Then combining (18), (20), and (21) with (16) yields

$$P_{\theta_0}^{\infty}\left[W_2^2(\delta_{\theta_0},\Pi_n(\cdot|X_{1:n})) \geqslant \epsilon_n^2\right] \leqslant \epsilon^2 + 8B\epsilon^2 + B\epsilon^2 = (9B+1)\epsilon^2 \to 0 \tag{8}$$

as $n\to\infty$; therefore, $\Pi_n(\cdot|X_{1:n})$ is strongly consistent at θ_0. ■

2 Proof of Proposition 3.1

The proof of this lemma follows from that of Lemma 3.1. Without loss of generality assume that all the subsets are of equal size and that $n = Km$. If $m \to \infty$, then the number of data in any subset $X_{[k]}$ goes to ∞; therefore, this proposition is proved by replacing n with m and $\Pi_n(\cdot|X_{1:n})$ with $\Pi_{k_n}(\cdot|X_{[k]})$ in the proof of Lemma 3.1.

3 Proof of Proposition 3.2

The proof of this proposition also follows from that of Lemma 3.1. Notice that $\Pi^{SA}_{k_n}(\cdot|X_{[k]})$ is equivalent to $\Pi_{k_n}(\cdot|\underbrace{X_{[k]},\dots,X_{[k]}}_{K\text{times}})$, the posterior distribution that has n data points; therefore, if $m \to \infty$, then $n = Km \to \infty$ and the proof of strong consistency of $\Pi^{SA}_{k_n}(\cdot|X_{[k]})$ at θ_0 follows from the proof of Lemma 3.1.

4 Proof of Theorem 3.3

For any set $\mathcal{U} \subset \mathbf{\Theta}$, the posterior probability $\overline{\Pi}_n(\mathcal{U}|X_{1:n})$

$$\overline{\Pi}_n(\mathcal{U}|X_{1:n}) = \lambda_1 \Pi^{SA}_{1_n}(\mathcal{U}|X_{[1]}) + \sum_{k=2}^{K} \lambda_k \widetilde{\Pi}^{SA}_{k_n}(\mathcal{U}|X_{[k]}), \tag{9}$$

where $\sum_{k=1}^{K} \lambda_k = 1$. Because $\{T^1_k\}_{k=1}^K$ are continuous measure preserving Borel maps $\mathbb{R}^D \to \mathbb{R}^D$, the definition of push-forward measures implies that

$$\widetilde{\Pi}^{SA}_{k_n}(\mathcal{U}|X_{[k]}) = \Pi^{SA}_{1_n}(\mathcal{U}|X_{[1]}) \tag{10}$$

for $k = 1, \dots, K$. Substituting (10) in (9) yields

$$\overline{\Pi}_n(\mathcal{U}|X_{1:n}) = \lambda_1 \Pi^{SA}_{1_n}(\mathcal{U}|X_{[1]}) + \sum_{k=2}^{K} \lambda_k \Pi^{SA}_{1_n}(\mathcal{U}|X_{[1]}). \tag{11}$$

Using Proposition 3.2, if $\epsilon > 0$ is given, then there exists n_1 such that

$$P^{\infty}_{\theta_0}\left[W_2^2(\delta_{\theta_0}, \Pi^{SA}_{1_n}(\cdot|X_{[1]})) < \epsilon^2\right] \geqslant 1 - \epsilon^2 \tag{12}$$

for all $n > n_1$. Using (11) and (12),

$$W_2^2(\delta_{\theta_0}, \Pi^{SA}_{1_n}(\cdot|X_{[1]})) < \epsilon^2 \implies W_2^2(\delta_{\theta_0}, \overline{\Pi}_n(\cdot|X_{1:n})) < \epsilon^2, \tag{13}$$

which in turn implies that

$$\begin{aligned} P^{\infty}_{\theta_0}\left[W_2^2(\delta_{\theta_0}, \overline{\Pi}_n(\cdot|X_{1:n})) < \epsilon^2\right] &\geqslant 1 - P^{\infty}_{\theta_0}\left[W_2^2(\delta_{\theta_0}, \Pi^{SA}_{1_n}(\cdot|X_{[1]}) > \epsilon^2)\right] \\ &\geqslant 1 - \epsilon^2 \to 1 \end{aligned} \tag{14}$$

a.s. $[P^{\infty}_{\theta_0}]$. This proves that $\overline{\Pi}_n(\cdot|X_{1:n})$ is strongly consistent at θ_0.

5 Proof of Theorem 3.4

Let $\mathcal{U}_n = \{\theta : \|\theta - \theta_0\| < \epsilon_n\}$ be a neighborhood of θ_0. Using Assumptions (A1) – (A4) and noticing that $\mathcal{P}_2(\boldsymbol{\Theta})$ includes probability measures that are parameterized by finite dimensional $\theta \in \boldsymbol{\Theta} \subset \mathbb{R}^D$, there exists a test function $\Phi_n = \Phi_n(X_{1:n})$ for testing $H_0 : \theta = \theta_0$ against $H : \theta \in \mathcal{U}_n^c$ and universal constants B and b such that

$$P_{\theta_0}(\Phi_n) \leqslant B\exp(-bn) \text{ and } \sup_{\theta \in \mathcal{U}_n^c} P_\theta(1 - \Phi_n) \leqslant B\exp(-bn); \tag{15}$$

see Le Cam (1986, Chapter 16) for details. The definition of W_2 and $\mathcal{U}_n$ imply that

$$\begin{aligned}
W_2^2(\delta_{\theta_0}, \boldsymbol{\Pi}_n(\cdot|X_{1:n})) &= \int_{\boldsymbol{\Theta}} \|\theta - \theta_0\|^2 d\boldsymbol{\Pi}_n(\theta\,|X_{1:n}) \\
&= \int_{\mathcal{U}_n \cap \boldsymbol{\Theta}_n} \|\theta - \theta_0\|^2 d\boldsymbol{\Pi}_n(\theta\,|X_{1:n}) + \int_{\mathcal{U}_n^c \cap \boldsymbol{\Theta}_n} \|\theta - \theta_0\|^2 d\boldsymbol{\Pi}_n(\theta\,|X_{1:n}) \\
&\quad + \int_{\boldsymbol{\Theta}_n^c} \|\theta - \theta_0\|^2 d\boldsymbol{\Pi}_n(\theta\,|X_{1:n}) \\
&\overset{\text{(i)}}{\leqslant} \epsilon_n^2 + \int_{\mathcal{U}_n^c \cap \boldsymbol{\Theta}_n} \|\theta - \theta_0\|^2 d\boldsymbol{\Pi}_n(\theta\,|X_{1:n}) + \int_{\boldsymbol{\Theta}_n^c} \|\theta - \theta_0\|^2 d\boldsymbol{\Pi}_n(\theta\,|X_{1:n}) \\
&\overset{\text{(ii)}}{\leqslant} \epsilon_n^2 + 4M_n^2 \int_{\mathcal{U}_n^c \cap \boldsymbol{\Theta}_n} d\boldsymbol{\Pi}_n(\theta\,|X_{1:n}) + \int_{\boldsymbol{\Theta}_n^c} \|\theta - \theta_0\|^2 d\boldsymbol{\Pi}_n(\theta\,|X_{1:n}) \\
&\leqslant \epsilon_n^2 + 4M_n^2\, \boldsymbol{\Pi}_n(\theta \in \mathcal{U}_n^c\,|X_{1:n}) + \int_{\boldsymbol{\Theta}_n^c} \|\theta - \theta_0\|^2 d\boldsymbol{\Pi}_n(\theta\,|X_{1:n}).
\end{aligned} \tag{16}$$

The first ϵ_n^2 term in (i) follows from the definition of $\mathcal{U}_n$. Assumption (A3) implies that $\|\theta - \theta_0\|^2 \leqslant 4M_n^2$ for $\theta \in \boldsymbol{\Theta}_n$ in (ii).

We show that the second term in (16) goes to zero a.s. $[P_{\theta_0}^\infty]$.

$$\begin{aligned}
\boldsymbol{\Pi}_n(\theta \in \mathcal{U}_n^c\,|X_{1:n}) &= \frac{\int_{\mathcal{U}_n^c} \prod_{i=1}^n p(X_i|\,\theta) d\boldsymbol{\Pi}_n(\theta)}{\int_{\boldsymbol{\Theta}} \prod_{i=1}^n p(X_i|\,\theta) d\boldsymbol{\Pi}_n(\theta)} = \frac{\int_{\mathcal{U}_n^c} \prod_{i=1}^n \frac{p(X_i|\,\theta)}{p(X_i|\,\theta_0)} d\boldsymbol{\Pi}_n(\theta)}{\int_{\boldsymbol{\Theta}} \prod_{i=1}^n \frac{p(X_i|\,\theta)}{p(X_i|\,\theta_0)} d\boldsymbol{\Pi}_n(\theta)} (\Phi_n + 1 - \Phi_n) \\
&\leqslant \Phi_n + \frac{(1 - \Phi_n) \int_{\mathcal{U}_n^c} \prod_{i=1}^n \frac{p(X_i|\,\theta)}{p(X_i|\,\theta_0)} d\boldsymbol{\Pi}_n(\theta)}{\int_{\boldsymbol{\Theta}} \prod_{i=1}^n \frac{p(X_i|\,\theta)}{p(X_i|\,\theta_0)} d\boldsymbol{\Pi}_n(\theta)}.
\end{aligned} \tag{17}$$

Using (15) and Markov's inequality, $\Phi_n \to 0$ a.s. $[P_{\theta_0}^\infty]$. To prove that $\boldsymbol{\Pi}_n(\theta \in \mathcal{U}_n^c\,|X_{1:n}) \to 0$ a.s. $[P_{\theta_0}^\infty]$, it is sufficient to show that

(a) $P_{\theta_0}\left[(1 - \Phi_n) \int_{\mathcal{U}_n^c} \prod_{i=1}^n \frac{p(X_i|\,\theta)}{p(X_i|\,\theta_0)} d\boldsymbol{\Pi}_n(\theta)\right] \leqslant B\exp(-n\beta)$ for some β;

(b) $\exp(n\beta) \int_{\boldsymbol{\Theta}} \prod_{i=1}^n \frac{p(X_i|\,\theta)}{p(X_i|\,\theta_0)} d\boldsymbol{\Pi}_n(\theta) \to \infty$ a.s. $[P_{\theta_0}^\infty]$ for every β.

To show (a), note that

$$
\begin{aligned}
& P_{\theta_0}\left[(1-\Phi_n)\int_{\mathcal{U}_n^c}\prod_{i=1}^{n}\frac{p(X_i\mid\theta)}{p(X_i\mid\theta_0)}d\Pi_n(\theta)\right] \\
&= \int(1-\Phi_n)\int_{\mathcal{U}_n^c}\prod_{i=1}^{n}\frac{p(X_i\mid\theta)}{p(X_i\mid\theta_0)}d\Pi_n(\theta)\left[\prod_{i=1}^{n}p(X_i\mid\theta_0)\right]d\nu^n(X_{1:n}) \\
&\overset{(i)}{=} \int_{\mathcal{U}_n^c}\int(1-\Phi_n)\prod_{i=1}^{n}\frac{p(X_i\mid\theta)}{p(X_i\mid\theta_0)}\left[\prod_{i=1}^{n}p(X_i\mid\theta_0)d\nu(X_i)\right]d\Pi_n(\theta) \\
&= \int_{\mathcal{U}_n^c}\left[\int(1-\Phi_n)\prod_{i=1}^{n}p(X_i\mid\theta)d\nu(X_i)\right]d\Pi_n(\theta) \\
&\overset{(ii)}{\leqslant} \sup_{\theta\in\mathcal{U}_n^c}P_\theta(\Phi_n)\,\Pi_n(\mathcal{U}_n^c)\leqslant B\exp(-bn).
\end{aligned}
\tag{18}
$$

Fubini's theorem implies (i) and (ii) follows from (15). To show (b), given any $\beta>0$, choose $\epsilon<\beta$ so that if $\mathrm{KL}(p(\cdot\mid\theta_0)\|p(\cdot\mid\theta))<\epsilon$, then

$$
\sum_{i=1}^{n}\log\frac{p(X_i\mid\theta)}{p(X_i\mid\theta_0)}+\beta n=n\left[\beta-n^{-1}\sum_{i=1}^{n}\log\frac{p(X_i\mid\theta_0)}{p(X_i\mid\theta)}\right]>n(\beta-\epsilon)\to\infty. \tag{19}
$$

a.s. $[P_{\theta_0}^\infty]$. The null set N involved in (19) may depend on θ. To show that (19) is true for all $\theta\in\boldsymbol{\Theta}$, we define a product space $E=\{(\omega,\theta):\exp(n\beta)\prod_{i=1}^{n}\frac{p(X_i\mid\theta)}{p(X_i\mid\theta_0)}\to\infty\}$, then (19) shows that $P_{\theta_0}^\infty(E_\theta)=1$ for all $\theta\in\boldsymbol{\Theta}$. Another application of Fubini's theorem shows that there exists N in the probability space on which X_ns are defined such that $P_{\theta_0}^\infty(N)=0$ and if $\omega\notin N$, then $\Pi_j(E^\omega)=1$ for all j; therefore, $\exp(n\beta)\prod_{i=1}^{n}\frac{p(X_i\mid\theta)}{p(X_i\mid\theta_0)}\to\infty$ for all $\omega\notin E^\omega$, which in turn implies that

$$
\exp(n\beta)\int\prod_{i=1}^{n}\frac{p(X_i\mid\theta)}{p(X_i\mid\theta_0)}d\Pi_n(\theta)\geqslant\int_{\mathcal{K}_\epsilon(\theta_0)}\exp\left(n\beta+\sum_{i=1}^{n}\log\frac{p(X_i\mid\theta)}{p(X_i\mid\theta_0)}\right)d\Pi_n(\theta)\to\infty \tag{20}
$$

a.s. $[P_{\theta_0}^\infty]$ by Fatou's lemma, since $\liminf_{n\to\infty}\Pi_n(\mathcal{K}_\epsilon(\theta_0))>0\ \forall\epsilon>0$.

We now follow a strategy similar to (17) to prove that the third term in (16) goes to zero a.s. $[P_{\theta_0}^\infty]$. Using (18) and (20), it is enough to show that $P_{\theta_0}\left[(1-\Phi_n)\int_{\boldsymbol{\Theta}_n^c}\|\theta-\theta_0\|^2\prod_{i=1}^{n}\frac{p(X_i\mid\theta)}{p(X_i\mid\theta_0)}d\Pi_n(\theta)\right]\leqslant$

$B\exp(-n\beta)$ for some β. Following (18),

$$
\begin{aligned}
&P_{\theta_0}\left[(1-\Phi_n)\int_{\Theta_n^c}\|\theta-\theta_0\|^2\prod_{i=1}^{n}\frac{p(X_i|\theta)}{p(X_i|\theta_0)}d\Pi_n(\theta)\right]= \\
&=\int(1-\Phi_n)\int_{\Theta_n^c}\|\theta-\theta_0\|^2\prod_{i=1}^{n}\frac{p(X_i|\theta)}{p(X_i|\theta_0)}d\Pi_n(\theta)\left[\prod_{i=1}^{n}p(X_i|\theta_0)\right]d\nu^n(X_{1:n}) \\
&\overset{(i)}{=}\int_{\Theta_n^c}\|\theta-\theta_0\|^2\int(1-\Phi_n)\prod_{i=1}^{n}\frac{p(X_i|\theta)}{p(X_i|\theta_0)}\left[\prod_{i=1}^{n}p(X_i|\theta_0)d\nu(X_i)\right]d\Pi_n(\theta) \\
&=\int_{\Theta_n^c}\|\theta-\theta_0\|^2\left[\int(1-\Phi_n)\prod_{i=1}^{n}p(X_i|\theta)d\nu(X_i)\right]d\Pi_n(\theta) \\
&\leqslant\sup_{\theta\in\mathcal{U}_n^c}P_\theta(\Phi_n)\int_{\Theta_n^c}\|\theta-\theta_0\|^2d\Pi_n(\theta)\leqslant B\exp(-bn)\epsilon_n^2\to 0
\end{aligned}
\tag{21}
$$

a.s. $[P_{\theta_0}^\infty]$. The last inequality in (21) follows from Theorem 2.1 and Prohorov's theorem.

Finally, Assumption (A2) implies that $\exists n_0$ such that $\forall n>n_0$ $M_n^2\exp(-bn)<\epsilon_n^2$. Then combining (18), (20), and (21) with (16) yields

$$P_{\theta_0}^\infty\left[W_2^2(\delta_{\theta_0},\Pi_n(\cdot|X_{1:n}))\geqslant\epsilon_n^2\right]\leqslant\epsilon_n^2+8B\epsilon_n^2+B\epsilon_n^2=(9B+1)\epsilon_n^2\to 0 \tag{22}$$

as $n\to\infty$; therefore, $\Pi_n(\cdot|X_{1:n})$ is strongly consistent at θ_0.

6 Optimization algorithms

Problem (17) – (18) is a linear program (LP) with a special structure. We reformulate this problem in a standard setting. Since $\sum_{k=1}^K N_k=N$, the LP consists of $n_x:=N(\sum_{k=1}^K N_k+1)=N(N+1)$ variables and $m:=KN+\sum_{k=1}^K N_k+1=KN+(N+1)$ linear equality constraints, where the $N+1$ last constraints come from $N+1$ simplex constraints.

First, let $\mathrm{col}_j(\mathbf{Z})$ be the column-wise operator that takes the j-th column of matrix $\mathbf{Z}$, and $[\mathbf{z}_1,\cdots,\mathbf{z}_s]$ be the column-wise concatenation of s vectors $\mathbf{z}_j$. We introduce the new variable $\tilde{\mathbf{t}}_{kj}:=N_k\mathrm{col}_j(\mathbf{T}_k)$ and define

$$
\begin{aligned}
\mathbf{x}\ &:=[\mathbf{a},\tilde{\mathbf{t}}_{11},\cdots,\tilde{\mathbf{t}}_{1N_1},\cdots,\tilde{\mathbf{t}}_{K1},\cdots,\tilde{\mathbf{t}}_{KN_K}] \\
&:=[\mathbf{x}_{[0]},\mathbf{x}_{[1]},\cdots,\mathbf{x}_{[N_1]},\mathbf{x}_{[N_1+1]},\cdots,\mathbf{x}_{[N]}].
\end{aligned}
\tag{23}
$$

Here, vector $\mathbf{x}_{[0]}\equiv\mathbf{a}$ and each vector $\mathbf{x}_{[l]}$ is a sub-vector in $\mathbb{R}^N$ and is generated from the columns of matrix $\mathbf{T}_k$ for $l=1,\ldots,N$.

Now, to reformulate the objective function, we introduce $\tilde{\mathbf{m}}_{kj}:=N_k^{-1}\mathrm{col}_j(\mathbf{M}_k)$ and a new vector $\mathbf{c}$ as follows:

$$
\begin{aligned}
\mathbf{c}\ &:=[\mathbf{0}^N,\tilde{\mathbf{m}}_{11},\cdots,\tilde{\mathbf{m}}_{1N_1},\tilde{\mathbf{m}}_{21},\cdots,\tilde{\mathbf{m}}_{2N_2}, \\
&\qquad\cdots,\tilde{\mathbf{m}}_{K1},\cdots,\tilde{\mathbf{m}}_{KN_K}] \\
&:=[\mathbf{c}_{[0]},\mathbf{c}_{[1]},\cdots,\mathbf{c}_{[N_1]},\mathbf{c}_{[N_1+1]},\cdots,\mathbf{c}_{[N]}].
\end{aligned}
\tag{24}
$$

Next, let $\mathbb{I}_N$ be the identity matrix in $\mathbb{R}^N$, and we introduce a matrix $\mathbf{A}$ as:

$$\begin{aligned}\mathbf{A} := & \begin{bmatrix} -\mathbb{I}_N & \mathbf{E}_{N_1} & \mathbf{0}^{N_2} & \cdots & \mathbf{0}^{N_K} \\ -\mathbb{I}_N & \mathbf{0}^{N_1} & \mathbf{E}_{N_2} & \cdots & \mathbf{0}^{N_K} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -\mathbb{I}_N & \mathbf{0}^{N_1} & \mathbf{0}^{N_2} & \cdots & \mathbf{E}_{N_K} \end{bmatrix}_{KN\times N(N+1)} \\ = & [\mathbf{A}_0, \mathbf{A}_1, \cdots, \mathbf{A}_N],\end{aligned} \tag{25}$$

where $\mathbf{E}_i := [N_1^{-1}\mathbb{I}_N, N_2^{-1}\mathbb{I}_N, \cdots, N_i^{-1}\mathbb{I}_N]$ for $i = 1, \ldots, N_k$ and $k = 1, \ldots, K$.

Finally, using the definitions above, we can simply reformulate the LP problem (17) – (18) into the following compact form:

$$f^* := \begin{cases} \min\limits_{\mathbf{x}\in\mathbb{R}^{n_x}} & \left\{ f(\mathbf{x}) := \mathbf{c}^T\mathbf{x} = \sum\limits_{l=0}^{N} \mathbf{c}_{[l]}^T \mathbf{x}_{[l]} \right\} \\ \text{s.t.} & \mathbf{A}\mathbf{x} = \mathbf{0}^{KN}, \\ & \mathbf{x}_{[l]} \in \Delta_N, \ l = 0, \ldots, N, \end{cases} \tag{26}$$

where Δ_N is the standard simplex in $\mathbb{R}^N$, i.e., $\Delta_N := \{\mathbf{u} \in \mathbb{R}^N_+ : \mathbf{1}^T\mathbf{u} = 1\}$.

From the reconstruction of $\mathbf{A}$, we can easily show that the sparsity of $\mathbf{A}$ is $s := \frac{N+K}{KN(N+1)}$. Hence, if we also count for the simplex constraints, then the overall sparsity of (26) is $s_{LP} := \frac{2N+K+1}{KN(N+1)}$, which is very sparse when N is large. Due to the simplex constraints, problem (26) always admits an optimal solution.

Although (26) is a linear program, but it is large-scale when N is large. By exploiting the sparsity of this problem, one can solve it efficiently by using off-the-shelf centralized LP solvers such as CPLEX or Gurobi. Alternatively, we can also exploit specific structure of (26) to develop appropriate decomposition methods that can be scaled naturally to sufficiently large dimension and can be implemented in a parallel or distribution fashion.

References

Ghosal, S., J. K. Ghosh, and A. W. Van Der Vaart (2000). Convergence rates of posterior distributions. *Annals of Statistics 28*(2), 500–531.

Le Cam, L. (1986). Asymptotic methods in statistical decision theory.