Supplementary Material for "Scalable and Robust Bayesian Inference via the Median Posterior"

Stanislav Minsker[1] SMINSKER@MATH.DUKE.EDU
Sanvesh Srivastava[2,3] SS602@STAT.DUKE.EDU
Lizhen Lin[2] LIZHEN@STAT.DUKE.EDU
David B. Dunson[2] DUNSON@STAT.DUKE.EDU
Departments of Mathematics[1] and Statistical Science[2], Duke University, Durham, NC 27708
Statistical and Applied Mathematical Sciences Institute[3], 19 T.W. Alexander Dr, Research Triangle Park, NC 27709

1. Proof of Propositions 3.3

For all $z \in \mathbb{R}$, $|e^{iz} - 1 - iz| \le \frac{|z|^2}{2}$, implying that

$$\rho_k^2(\theta_1,\theta_2) = \|k(\cdot,\theta_1) - k(\cdot,\theta_2)\|_{\mathbb{H}}^2 = 2k(0) - 2k(\theta_1-\theta_2)$$
$$= 2\int_{\mathbb{R}^p} (1 - e^{i\langle x,\theta_1-\theta_2\rangle}) d\nu(x) \le \int_{\mathbb{R}^p} \langle x, \theta_1-\theta_2\rangle_{\mathbb{R}^p}^2 \, d\nu(x)$$
$$\le \|\theta_1-\theta_2\|_2^2 \int_{\mathbb{R}^p} \|x\|_2^2 d\nu(x),$$

hence $\|\theta_1-\theta_2\|_2 \ge \frac{\rho_k(\theta_1,\theta_2)}{\sqrt{\int_{\mathbb{R}^p} \|x\|_2^2 d\nu(x)}}$, which implies the result.

2. Proof of Theorem 3.4

Let

$$\mathcal{F}_L := \{f : \Theta \mapsto \mathbb{R} \text{ s.t. } \|f\|_L \le 1\},$$

where $\|f\|_L := \sup_{\theta_1\neq\theta_2} \frac{|f(\theta_1)-f(\theta_2)|}{\rho_k(\theta_1,\theta_2)}$ is the Lipschitz constant of f.

It is well-known ((Dudley, 2002), Theorem 11.8.2) that in this case $\|P-Q\|_{\mathcal{F}_L}$ is equal to the Wasserstein distance (also called the Kantorovich-Rubinstein distance)

$$d_{W_1}(P,Q) = \inf\Big\{\mathbb{E}\rho(\boldsymbol{X},\boldsymbol{Y}) : \mathcal{L}(\boldsymbol{X}) = P,\ \mathcal{L}(\boldsymbol{Y}) = Q\Big\}, \tag{1}$$

where $\mathcal{L}(\boldsymbol{Z})$ denotes the law of a random variable $\boldsymbol{Z}$ and the infimum on the right is taken over the set of all joint distributions of $(\boldsymbol{X},\boldsymbol{Y})$ with marginals P and Q.

Let $f \in \mathbb{H}$ - the RKHS associated to kernel k, and note that, due to the reproducing property and Cauchy-Schwarz inequality, we have

$$f(\theta_1) - f(\theta_2) = \langle f, k(\cdot,\theta_1) - k(\cdot,\theta_2)\rangle_{\mathbb{H}}$$
$$\le \|f\|_{\mathbb{H}} \|k(\cdot,\theta_1) - k(\cdot,\theta_2)\|_{\mathbb{H}} = \|f\|_{\mathbb{H}}\, \rho_k(\theta_1,\theta_2).$$

Therefore, $\mathcal{F}_k \subseteq \mathcal{F}_L$, hence $\|P-Q\|_{\mathcal{F}_k} \le \|P-Q\|_{\mathcal{F}_L}$. Hence, convergence with respect to $\|\cdot\|_{\mathcal{F}_L}$ implies convergence with respect to $\|\cdot\|_{\mathcal{F}_k}$.

By the definition of Wasserstein distance d_{W_1}, we have

$$d_{W_1}(\delta_0, \Pi_l(\cdot|\mathcal{X}_l)) = \int_\Theta \rho_k(\theta,\theta_0) d\Pi_l(\theta|\mathcal{X}_l). \tag{2}$$

Let C_1 be a large enough constant. Using (2) and assumption 3.1, it is easy to see that

$$d_{W_1}(P_0, \Pi_n(\cdot|X_1,\ldots,X_n))$$
$$\le \left(\frac{C_1}{\tilde{C}}\varepsilon_n\right)^{1/\gamma} + C_2 \int_{h(P_\theta,P_0)\ge C_1\varepsilon_n} d\Pi_n(\cdot|X_1,\ldots,X_n),$$

where $\tilde{C}$ is a constant in assumption 3.1 and $C_2 = \sup_{\theta_1,\theta_2} \rho_k(\theta_1,\theta_2) \le \frac{1}{\tilde{C}^{1/\gamma}}$ by assumption 3.1. s It remains to estimate the second term in the sum above: this is done exactly as in the proof of Theorem 2.1 in (Ghosal et al., 2000). Following these steps and using the assumption that $e^{-Kl\varepsilon_l^2/2} \le \varepsilon_l$, the result is easily deduced.

3. Proof of Corollary 3.5

Remark 3.1. The equation and theorem numbers in this proof refer to the main file.

It is enough to apply Theorem 2.1 with $\nu = 0$ to the independent random measures $\Pi_l(\cdot|G_j),\ j = 1,\ldots,m$. Note that the "weak concentration" assumption (3) is implied by (14).

References

Dudley, Richard M. *Real analysis and probability*, volume 74. Cambridge University Press, 2002.

Ghosal, Subhashis, Ghosh, Jayanta K, and Van Der Vaart, Aad W. Convergence rates of posterior distributions. *Annals of Statistics*, 28(2):500–531, 2000.