Appendix to "Approximate Bayesian Computation with Kullback-Leibler Divergence as Data Discrepancy"

1 Technical Proof

1.1 Lévy's Upward Theorem

Theorem 0 (Lévy's Upward Theorem) *Denote by $\{Z_n\}_{n\geq 0}$ a collection of random variables, and by $\mathcal{F}_n$ a filtration on the same probability space. If $\sup_{n\geq 0}|Z_n|$ is integrable, $Z_n \to Z_\infty$ almost surely as $n \to \infty$ and $\mathcal{F}_n \uparrow \mathcal{F}_\infty$ then $\mathbb{E}[Z_n|\mathcal{F}_n] \to \mathbb{E}[Z_\infty|\mathcal{F}_\infty]$ both almost surely and in mean.*

1.2 Proof of Theorem 1

Let $Z_n = \mathbb{I}\left(\mathfrak{D}(\boldsymbol{X}, \boldsymbol{Y}) < \epsilon\right) = \mathbb{I}\left(\mathfrak{D}(\{X_i\}_{i=1}^n, \{Y_i\}_{i=1}^{m(n)}) < \epsilon\right)$ then $\sup_{n\to\infty}|Z_n| \geq 1$ and $Z_n \to Z_\infty = \mathbb{I}\left(\mathfrak{D}(p_{\theta^*}, p_\theta) < \epsilon\right)$. Let $\mathcal{F}_n = \sigma\text{-algebra}(X_1, \ldots, X_n)$ then $\mathcal{F}_n \uparrow \mathcal{F}_\infty = \cup_{k\geq 0}\mathcal{F}_k$. Using Lévy's Upward Theorem yields

$$\mathbb{E}_\theta\left[\mathbb{I}\left(\mathfrak{D}(\boldsymbol{X}, \boldsymbol{Y}) < \epsilon\right)|\boldsymbol{X}\right] \to \mathbb{I}\left(\mathfrak{D}(p_{\theta^*}, p_\theta) < \epsilon\right) \text{ a.s..}$$

Noting that

$$\mathbb{E}_\theta\left[\mathbb{I}\left(\mathfrak{D}(\boldsymbol{X}, \boldsymbol{Y}) < \epsilon\right)|\boldsymbol{X}\right] = \int \mathbb{I}\left(\mathfrak{D}(\boldsymbol{X}, \boldsymbol{y}) < \epsilon\right) p_\theta(\boldsymbol{y}) d\boldsymbol{y},$$

we have convergence of the numerator in (1)

$$\int \pi(\theta)\mathbb{I}\left(\mathfrak{D}(\boldsymbol{X}, \boldsymbol{y}) < \epsilon\right) p_\theta(\boldsymbol{y}) d\boldsymbol{y} \to \pi(\theta)\mathbb{I}\left(\mathfrak{D}(p_{\theta^*}, p_\theta) < \epsilon\right) \text{ a.s.,}$$

and convergence of the denominator in (1) by the dominated convergence theorem

$$\iint \pi(\theta')\mathbb{I}\left(\mathfrak{D}(\boldsymbol{X}, \boldsymbol{y}) < \epsilon\right) p_{\theta'}(\boldsymbol{y}) d\boldsymbol{y} d\theta' \to \int \pi(\theta')\mathbb{I}\left(\mathfrak{D}(p_{\theta^*}, p_{\theta'}) < \epsilon\right) d\theta'.$$

Combining two convergence results completes the proof.

1.3 Proof of Corollary 1

This is an immediate consequence of Theorem 1 in the main text and Theorem 2 in [**?**].

1.4 Proof of Corollary 2

As $n \to \infty$ and $m/n \to \alpha > 0$,

$$\begin{aligned}
\mathfrak{D}_{\text{CA}}(\boldsymbol{X}, \boldsymbol{Y}) &\to \frac{1}{1+\alpha}\int [1 - h(x)] p_{\theta^*}(x) dx + \frac{\alpha}{1+\alpha}\int h(y) p_\theta(y) dy \\
&= \frac{1}{1+\alpha} + \frac{1}{1+\alpha}\int h(x)[\alpha p_\theta(x) - p_{\theta^*}(x)] dx
\end{aligned}$$

If $h(x) = \mathbb{I}(\alpha p_\theta(x) \geq p_{\theta^*}(x))$,

$$\begin{aligned}
\text{RHS} &= \frac{1}{1+\alpha} + \frac{1}{1+\alpha}\int \max\left\{\frac{\alpha p_\theta(x)}{p_{\theta^*}(x)} - 1, 0\right\} p_{\theta^*}(x) dx \\
&= \frac{1}{1+\alpha}\int \left[1 + \max\left\{\frac{\alpha p_\theta(x)}{p_{\theta^*}(x)} - 1, 0\right\}\right] p_{\theta^*}(x) dx \\
&= \frac{1}{1+\alpha}\int \max\left\{\frac{\alpha p_\theta(x)}{p_{\theta^*}(x)}, 1\right\} p_{\theta^*}(x) dx \\
&= \frac{1}{1+\alpha}\int \left[\max\left\{\frac{\alpha p_\theta(x)}{p_{\theta^*}(x)}, 1\right\} - (\alpha \vee 1)\right] p_{\theta^*}(x) dx + \frac{\alpha \vee 1}{1+\alpha} \\
&= \mathfrak{D}_f(p_{\theta^*}||p_\theta) + c(\alpha)
\end{aligned}$$

1.5 Proof of Corollary 3

The auxiliary MLE asymptotically minimizes the KL divergence between p_θ and $p_A(\cdot|\phi)$.

$$\hat{\phi}(\boldsymbol{Y}) = \arg\max_{\phi\in\Phi}\prod_{i=1}^{m} p_A(Y_i|\phi) \to \hat{\phi}(\theta) = \arg\min_{\phi\in\Phi}\mathrm{KL}(p_\theta||p_A(\cdot|\phi))$$

Since the auxiliary model $\{p_A(\cdot|\phi) :\ \phi\in\Phi\}$ is bijective to the model $\{p_\theta :\ \theta\in\Theta\}$,

$$\min_{\phi\in\Phi}\mathrm{KL}(p_\theta||p_A(\cdot|\phi)) = 0, \quad p_A(\cdot|\hat{\phi}(\theta)) = p_\theta.$$

Then

$$\mathfrak{D}_{\mathrm{AL}}(\boldsymbol{X},\boldsymbol{Y}) \to \mathrm{KL}(p_\theta||p_A(\cdot|\hat{\phi}(\theta^*))) - \mathrm{KL}(p_\theta||p_A(\cdot|\hat{\phi}(\theta))) = \mathrm{KL}(p_\theta||p_{\theta^*}).$$