University of Insubria

DIPARTIMENTO DI SCIENZA E ALTA TECNOLOGIA
PhD in Computer Science and Mathematics of Calculus

Report of (Markov Chain) Monte Carlo Simulation

On the Mathematical Foundations of Approximate Bayesian Computation

Teacher: *A. Mira* Student: *M. Tarsia*[*] Tutor: *D. Cassani* Advisor: *E. Mastrogiacomo*

Academic Year 2019–2020
14/11/2019

A mathematical frame for ABC

Underlying probability space: $(\Omega, \mathcal{A}, \mathbf{P})$. *Dimensions*: $d_{\mathcal{Y}}$, $d_{\mathcal{H}}$ and n in $\mathbb{N}^* \equiv \mathbb{N} \setminus \{0\}$. *Observations*: $y^{1:n}(\omega) \equiv y^{1:n} = (y^1, \dots, y^n)^\mathsf{T} \in \mathcal{Y}^n \subseteq \mathbb{R}^{d_{\mathcal{Y}} \cdot n}$, $\forall\, \omega \in \Omega$, where $\mathcal{Y} \subseteq \mathbb{R}^{d_{\mathcal{Y}}}$ has a metric $\rho_{\mathcal{Y}}$. *Parameters*: $\vartheta \in \mathcal{H}$ where $\mathcal{H} \subseteq \mathbb{R}^{d_{\mathcal{H}}}$ has a metric $\rho_{\mathcal{H}}$. *Prior*: $\pi \in \mathscr{P}(\mathcal{H})$. *Model*: $\{\mu_\vartheta^n\}_{\vartheta \in \mathcal{H}}$, family in $\mathscr{P}(\mathcal{Y}^n)$.

Notations. Let X be a topological space. We denote by $\mathcal{B}(X)$ the σ-algebra on X of the Borel subsets of X, by $\mathscr{P}(X)$ the class of the probability measures on $\mathcal{B}(X)$ and, given another topological space Y, by $\mathscr{B}(X, Y)$ the class of the Borel measurable functions from $X \equiv (X, \mathcal{B}(X))$ to $Y \equiv (Y, \mathcal{B}(Y))$.

Next, we'll write $\widetilde{\forall}\, \vartheta \in \mathcal{H}$ meaning $\forall\, \vartheta \in \mathcal{H}\ [\pi]$, i.e. (id est) for π-a.a. (almost all) $\vartheta \in \mathcal{H}$.

[A0 - a] The model $\{\mu_\vartheta^n\}_{\vartheta \in \mathcal{H}}$ is *generative* meaning that, $\widetilde{\forall}\, \vartheta \in \mathcal{H}$, it's possible to generate how many $z^{1:n} = (z^1, \dots, z^n)^\mathsf{T} \in \mathcal{Y}^n$ with $z^{1:n} \sim \mu_\vartheta^n$ we desire.

Pseudo-observations: $z^{1:n} \in \mathcal{Y}^n$ with $z^{1:n} \sim \mu_\vartheta^n$, $\widetilde{\forall}\, \vartheta \in \mathcal{H}$. *Deviation measure*: $\mathcal{D}$, pseudo-metric on $\mathcal{Y}^n$.

Notations. $\widetilde{\forall}\, \vartheta \in \mathcal{H}$, $\mathcal{Y}_\vartheta^n \coloneqq \{\, z^{1:n} \in \mathcal{Y}^n \mid z^{1:n} \sim \mu_\vartheta^n \,\} \subset \mathcal{Y}^n$ and, for any $\varepsilon \in \mathbb{R}_+ \equiv [0, +\infty[$,

$$D_\varepsilon^n \coloneqq \mathcal{D}(y^{1:n}, \cdot\,)^{-1}\big([0, \varepsilon]\big) \equiv \big\{\, z^{1:n} \in \mathcal{Y}^n \mid \mathcal{D}(y^{1:n}, z^{1:n}) \leq \varepsilon \,\big\} \in \mathcal{B}(\mathcal{Y}^n).$$

Remark. Although there may exist $y \neq z$ in $\mathcal{Y}^n$ s.t. (such that) $\mathcal{D}(y, z) = 0$, $\mathcal{D}$ remains non-negative, subadditive and componentwise continuous. Moreover, $D_\varepsilon^n \downarrow D_0^n \equiv \mathcal{D}(y^{1:n}, \cdot\,)^{-1}(0)$ as $\varepsilon \downarrow 0$.

[A0 - b] (under A0 - a) There exists $\varepsilon_0 > 0$ s.t., for any $\varepsilon \in]0, \varepsilon_0[$, the two following conditions hold.

1. The function $\vartheta \mapsto \mu_\vartheta^n[D_\varepsilon^n]$, to be seen as defined π-a.s. (almost surely), belongs to $\mathscr{B}(\mathcal{H}, [0, 1])$.
2. $\int_{\mathcal{H}} \mu_\vartheta^n[D_\varepsilon^n]\, \pi(\mathrm{d}\vartheta) > 0$ (i.e. $\neq 0$).

Remark. 2 of A0 - b is equivalent to having $\vartheta \mapsto \mu_\vartheta^n[D_\varepsilon^n]$, briefly $\mu_{(\cdot)}^n[D_\varepsilon^n]$, not π-a.s. identically zero.

► Regarding the whole continuation, we assume that <u>A0 - a and A0 - b</u> worth.

ABC thresholds: any $\varepsilon \in]0, \varepsilon_0[$. *ABC rejection algorithms*: hereunder.

(i) Choose $\varepsilon \in]0, \varepsilon_0[$. **(ii)** Draw $\vartheta \in \mathcal{H}$ by π and $z^{1:n} \in \mathcal{Y}_\vartheta^n$. **(iii)** Keep ϑ if, and only if, $z^{1:n} \in D_\varepsilon^n$.

ABC posteriors: $\pi_{y^{1:n}}^\varepsilon \ll \pi$, $\forall\, \varepsilon \in]0, \varepsilon_0[$, whose density is proportional to $\mu_{(\cdot)}^n[D_\varepsilon^n]$: for any $B \in \mathcal{B}(H)$,

$$\pi_{y^{1:n}}^\varepsilon[B] = \frac{\int_B \mu_\vartheta^n[D_\varepsilon^n]\, \pi(\mathrm{d}\vartheta)}{\int_{\mathcal{H}} \mu_{\vartheta'}^n[D_\varepsilon^n]\, \pi(\mathrm{d}\vartheta')}.$$

[*]E-mail: `mtarsia1@uninsubria.it`. Web page: `https://www.uninsubria.it/hpp/marco.tarsia`.

Remark. Explicitly, $\forall\, \varepsilon \in \,]0, \varepsilon_0[$, $\pi^{\varepsilon}_{y^{1:n}} \in \mathscr{P}(H)$ is absolutely continuous w.r.t. (with respect to) π meaning that, for any $\mathcal{N} \in \mathcal{B}(\mathcal{H})$, if $\pi[\mathcal{N}] = 0$ then $\pi^{\varepsilon}_{y^{1:n}}[\mathcal{N}] = 0$ as well.

[A0 - c] For any $Y \in \mathcal{B}(\mathcal{Y}^n)$, $\mu^n_{(\,\cdot\,)}[Y] \in \mathscr{B}(\mathcal{H}, [0,1])$ (coherently w.r.t. A0 - b).

Model for the true posterior (under A0 - c): for any $Y \in \mathcal{B}(\mathcal{Y}^n)$ and $B \in \mathcal{B}(\mathcal{H})$ with $\pi[B] > 0$,

$$\mathrm{P}[Y|B] \doteq \frac{1}{\pi[B]} \int_B \mu^n_{\vartheta}[Y]\, \pi(\mathrm{d}\vartheta)$$

from which the corresponding posterior: for any $Y \in \mathcal{B}(\mathcal{Y}^n)$ and $B \in \mathcal{B}(\mathcal{H})$, whenever it makes sense,

$$\pi[B|Y] \equiv \frac{\int_B \mu^n_{\vartheta}[Y]\, \pi(\mathrm{d}\vartheta)}{\int_{\mathcal{H}} \mu^n_{\vartheta'}[Y]\, \pi(\mathrm{d}\vartheta')}$$

(through the Bayes' formula with π still as the prior). Therefore, the *true posterior* would be

$$\pi[\,\cdot\,|y^{1:n}] \coloneqq \pi\big[\,\cdot\,\big|\big\{y^{1:n}\big\}\big].$$

Remark. For any $\varepsilon \in \,]0, \varepsilon_0[$, $\pi^{\varepsilon}_{y^{1:n}}[\,\cdot\,] = \pi[\,\cdot\,|D^n_{\varepsilon}]$.

A convergence result for $\varepsilon \downarrow 0$

Notation. We'll denote by $\mathrm{m} \coloneqq \mathrm{m}^{d_{\mathcal{Y}}\cdot n}$ the Lebesgue measure on $\mathbb{R}^{d_{\mathcal{Y}}\cdot n}$ (on $\mathcal{B}(\mathbb{R}^{d_{\mathcal{Y}}\cdot n})$).

[A1] $\widetilde{\forall}\, \vartheta \in \mathcal{H}$, the two following conditions hold.

1. $\mu^n_{\vartheta} \ll \mathrm{m}$ with $f^n_{\vartheta} \doteq \mathrm{d}\mu^n_{\vartheta}/\mathrm{dm}$ s.t., $\widetilde{\forall}\, z^{1:n} \in \mathcal{Y}^n$ $[\mathrm{m}]$ for which it's defined, $f^n_{(\,\cdot\,)}(z^{1:n}) \in \mathscr{B}(\mathcal{H}, \mathbb{R}_+)$.

2. $f^n_{\vartheta}(\,\cdot\,)$ is continuous and $f^n_{(\,\cdot\,)}(y^{1:n})$ is not π-a.s. identically zero.

Remark. 1 of A1 implies A0 - c while 2 of A1 ensures that $\int_{\mathcal{H}} f^n_{\vartheta}(y^{1:n})\, \pi(\mathrm{d}\vartheta) > 0$ (eventually $+\infty$).

[A2] (under A1) There exist $\delta, \bar{\varepsilon} \in \,]0, +\infty[$ and $g \in L^1(\pi)$ with $g \geq \delta$ $[\pi]$ all s.t., $\widetilde{\forall}\, \vartheta \in \mathcal{H}$,

$$\delta \leq \sup_{z^{1:n} \in D^n_{\bar{\varepsilon}}} f^n_{\vartheta}(z^{1:n}) \leq g(\vartheta).$$

Remarks.

- A2 would imply 2 of A0 - b employing any $\varepsilon_0 \in \,]0, \bar{\varepsilon}]$ because, for any $\varepsilon \in \,]0, \bar{\varepsilon}[$, the function

$$\vartheta \mapsto \mu^n_{\vartheta}[D^n_{\varepsilon}] \equiv \int_{D^n_{\varepsilon}} f^n_{\vartheta}(z^{1:n})\, \mathrm{d}z^{1:n}$$

  cannot be π-a.s. identically zero. In particular, for any $\varepsilon \in \,]0, \bar{\varepsilon}[$, $\mathrm{m}[D^n_{\varepsilon}] > 0$ too.
- A2 implies that $f^n_{(\,\cdot\,)}(y^{1:n}) \in L^1(\pi)$ with $L^1(\pi)$-norm lower or equal than $\|g\|_1 \coloneqq \|g\|_{L^1(\pi)}$.
- Even the following generalization of A2 would work. $[\widetilde{\mathrm{A2}}]$ (under A1) There exist $g \in L^1(\pi)$ with $g > 0$ $[\pi]$ and $\tilde{\varepsilon} \in \,]0, +\infty[$ s.t., for any $\varepsilon \in \,]0, \tilde{\varepsilon}[$, there exists $\delta_{\varepsilon} \in \,]0, +\infty[$ s.t., $\widetilde{\forall}\, \vartheta \in \mathcal{H}$,

$$\delta_{\varepsilon} \leq \sup_{z^{1:n} \in D^n_{\varepsilon}} f^n_{\vartheta}(z^{1:n}) \leq g(\vartheta).$$

[A3] (under A1) $\widetilde{\forall}\, \vartheta \in \mathcal{H}$, $\mathcal{D}(y^{1:n}, \cdot\,)^{-1}(0) \subseteq f^n_{\vartheta}(\,\cdot\,)^{-1}\big(f^n_{\vartheta}(y^{1:n})\big)$.

Remark. Extensively, for any $z^{1:n} \in \mathcal{Y}^n$, if $\mathcal{D}(y^{1:n}, z^{1:n}) = 0$ then, $\widetilde{\forall}\, \vartheta \in \mathcal{H}$, $f^n_{\vartheta}(z^{1:n}) = f^n_{\vartheta}(y^{1:n})$. Hence, if $\mathcal{D}(y^{1:n}, \cdot\,)^{-1}(0) = \big\{y^{1:n}\big\}$, which happens when $\mathcal{D}$ is an actual metric, then A3 trivially holds.

Proposition. *Under assumptions* A1, A2 *and* A3*, the three following conditions hold.*

a. *The ABC rejection algorithm and the ABC posterior are well defined for any* $\varepsilon \in \,]0, \varepsilon_0 \vee \bar{\varepsilon}[$.

b. *The true posterior* $\pi[\,\cdot\,|y^{1:n}]$ *makes sense and it takes the following expression*: *for any* $B \in \mathcal{B}(\mathcal{H})$,

$$\pi[B|y^{1:n}] = \frac{\int_B f^n_\vartheta(y^{1:n})\,\pi(\mathrm{d}\vartheta)}{\int_\mathcal{H} f^n_{\vartheta'}(y^{1:n})\,\pi(\mathrm{d}\vartheta')}$$

c. *The ABC posterior strongly converges to the true posterior as* $\varepsilon \downarrow 0$: *for any* $B \in \mathcal{B}(\mathcal{H})$,

$$\pi^\varepsilon_{y^{1:n}}[B] \to \pi[B|y^{1:n}] \quad \textit{as } \varepsilon \downarrow 0.$$

Proof (**b** *and* **c**). The thesis essentially matches with the fact that, for any $B \in \mathcal{B}(\mathcal{H})$,

$$\int_B \frac{1}{\mathrm{m}[\mathrm{D}^n_\varepsilon]}\,\mu^n_\vartheta[\mathrm{D}^n_\varepsilon]\,\pi(\mathrm{d}\vartheta) \to \int_B f^n_\vartheta(y^{1:n})\,\pi(\mathrm{d}\vartheta) \quad \text{as } \varepsilon \downarrow 0$$

as a consequence of the classical Lebesgue's dominated convergence theorem. Indeed, on one side, pointwise convergence: $\widetilde{\forall}\,\vartheta \in \mathcal{H}$, $\mathrm{m}[\mathrm{D}^n_\varepsilon]^{-1}\mu^n_\vartheta[\mathrm{D}^n_\varepsilon] \equiv \mathrm{m}[\mathrm{D}^n_\varepsilon]^{-1}\int_{D^n_\varepsilon} f^n_\vartheta(z^{1:n})\,\mathrm{d}z^{1:n} \to f^n_\vartheta(y^{1:n})$ as $\varepsilon \downarrow 0$ due to the basic integral mean value theorem leaning on the continuity of $f^n_\vartheta(\,\cdot\,)$ and A3. On the other side, dominance: $\widetilde{\forall}\,\vartheta \in \mathcal{H}$ and $\forall\,\varepsilon \in \,]0, \bar{\varepsilon}[$, $0 \le \mathrm{m}[\mathrm{D}^n_\varepsilon]^{-1}\mu^n_\vartheta[\mathrm{D}^n_\varepsilon] \le g(\vartheta)$ from A2, and $g \in L^1(\pi)$. □

Optimal transport theory in ABC

Let's visualize $(\mathcal{Y}, \rho_\mathcal{Y})$ as a separable and complete metric space in such a way that it is also a Radon space, i.e. any element in $\mathscr{P}(\mathcal{Y})$ is a Radon probability measure (outer regular on Borel subsets and inner regular on open subsets), and let's choose an unit cost function $c\colon \mathcal{Y} \times \mathcal{Y} \to [0, +\infty]$ which is lower semicontinuous (thus Borel measurable) and a parameter $p \in [1, +\infty[$ of summability.

Notation. We'll denote by $\mathscr{P}_p(\mathcal{Y})$ the subclass of $\mathscr{P}(\mathcal{Y})$ whose elements have finite p-th moment.

Kantorovich's formulation. For any $\mu, \nu \in \mathscr{P}_p(\mathcal{Y})$, let's consider the subclass $\Gamma(\mu,\nu)$ of $\mathscr{P}(\mathcal{Y}\times\mathcal{Y})$ whose elements γ are the couplings with marginals μ and ν. Then the Kantorovich's formulation of the optimal transport problem related to $(\mathcal{Y}, \rho_\mathcal{Y})$, c and p is

$$\mathcal{K}(\mu,\nu) \doteq \inf_{\gamma\in\Gamma(\mu,\nu)} \int_{\mathcal{Y}\times\mathcal{Y}} c(y,y')\,\mathrm{d}\gamma(y,y').$$

It can be shown that there exists a minimizer $\gamma^* \in \Gamma(\mu,\nu)$ for such a problem which could be determined by means of gradient descent algorithms.

Example. For $c = (\rho_\mathcal{Y})^p$, $\mathcal{K}$ coincides with the p-power of the Wasserstein or Kantorovich-Rubinstein distance: in symbols, $\mathcal{K} = \mathcal{W}_p{}^p$.

Monge's formulation. For any $\mu, \nu \in \mathscr{P}_p(\mathcal{Y})$, let's consider the subclass $\mathsf{T}(\mu,\nu)$ of $\mathscr{B}(\mathcal{Y}) := \mathscr{B}(\mathcal{Y},\mathcal{Y})$ whose elements T are such that $T_\#\mu = \nu$.

Remark. Here $T_\#\mu$ stands for the push-forward or image measure $\mu^T \equiv T(\mu)$ of μ through T: that is, the element in $\mathscr{P}(\mathcal{Y})$ defined by $T_\#\mu[A] \doteq \mu[T^{-1}(A)]$, $A \in \mathcal{B}(\mathcal{Y})$.

Then, at least when μ and ν are both atomic (not diffuse) or otherwise when μ is not atomic (diffuse), the Monge's formulation of the optimal transport problem related to $(\mathcal{Y}, \rho_\mathcal{Y})$, c and p is

$$\mathcal{M}(\mu,\nu) \doteq \inf_{T\in\mathsf{T}(\mu,\nu)} \int_\mathcal{Y} c\big(y, T(y)\big)\,\mu(\mathrm{d}y).$$

Example. Let's assume that $d_{\mathcal{Y}} = 1$ and $\mathcal{Y} = \mathbb{R}$ with $\rho_{\mathcal{Y}}$ equal to the usual Euclidean metric.

Notation. For any $\eta \in \mathscr{P}(\mathbb{R})$, we'll denote by F_η and F_η^{-1} the cumulative distribution function of η and the quantile function of η respectively.

If there exists a function $\varphi \colon \mathbb{R} \to \mathbb{R}$ which is convex and such that, for any $y, y' \in \mathbb{R}$, $c(y, y') = \varphi(y - y')$ then, for any $\mu, \nu \in \mathscr{P}_p(\mathbb{R})$ with μ not atomic, the function $T^* := F_\nu^{-1} \circ F_\mu \in \mathsf{T}(\mu, \nu)$ is an optimal transport map w.r.t. the Monge's formulation and the following identity holds:

$$\mathcal{M}(\mu, \nu) \equiv \int_{-\infty}^{+\infty} \varphi\big(y - T^*(y)\big)\, \mu(\mathrm{d}y) = \int_0^1 \varphi\big(F_\mu^{-1}(t) - F_\nu^{-1}(t)\big)\, \mathrm{d}t.$$

Moreover, if φ is strictly convex, then such a T^* is the unique optimal transport map.

Radon's metric. For any $\mu, \nu \in \mathscr{P}_p(\mathcal{Y})$,

$$\rho_{\mathcal{R}}(\mu, \nu) \doteq \sup_{h \in C^0(\mathcal{Y}, [-1,1])} \int_{\mathcal{Y}} h(y)\, (\mu - \nu)(\mathrm{d}y)$$

defines a metric on $\mathscr{P}_p(\mathcal{Y})$ whose notion of convergence corresponds with the total variation convergence.

Remark. The space $(\mathcal{Y}, \rho_{\mathcal{Y}})$ is a Hausdorff, namely T_2, and locally compact as a topological space.

Some lower bounds for $n \to +\infty$

Notation. $\forall\, n \in \mathbb{N}^*$, we'll write $\widetilde{\forall}\, y^{1:n} \in \mathcal{Y}^n$ meaning to vary of $y^{1:n}(\omega) \equiv y^{1:n}$ in $\mathcal{Y}^n$ for $\mathbf{P}$-a.a. $\omega \in \Omega$.

Deviation measure of distributions: once and for all, $\forall\, n \in \mathbb{N}^*$, $\widetilde{\forall}\, y^{1:n} \in \mathcal{Y}^n$, $\widetilde{\forall}\, \vartheta \in \mathcal{H}$ and $\forall\, z^{1:n} \in \mathcal{Y}^n_\vartheta$, we univocally associate an element in $\mathscr{P}(\mathcal{Y})$, possibly in $\mathscr{P}_p(\mathcal{Y})$ ($p \geq 1$), to both of $y^{1:n}$ and $z^{1:n}$, let be

$$\mu_n \equiv \mu_{y^{1:n}} \text{ to } y^{1:n} \quad \text{and} \quad \mu_{\vartheta,n} \equiv \mu_{\vartheta, z^{1:n}} \text{ to } z^{1:n}$$

and we select a pseudo-distance $\mathcal{T}$ on $\mathscr{P}(\mathcal{Y})$, possibly on $\mathscr{P}_p(\mathcal{Y})$.

Example. $\mu_n \equiv \widehat{\mu}_n := \sum_{k=1}^n \delta_{y^k}$ and $\mu_{\vartheta,n} \equiv \widehat{\mu}_{\vartheta,n} := \sum_{k=1}^n \delta_{z^k}$ where, for any $x \in \mathcal{Y}$ and $B \in \mathcal{B}(\mathcal{Y})$,

$$\delta_x[B] \equiv \mathbb{1}_B(x) \doteq \begin{cases} 1, & \text{if } x \in B, \\ 0, & \text{if } x \notin B. \end{cases}$$

Remark. The space $(\mathscr{P}(\mathcal{Y}), \mathcal{T})$, possibly $(\mathscr{P}_p(\mathcal{Y}), \mathcal{T})$, may not be of Hausdorff as a topological space.

[B0] $\forall\, n \in \mathbb{N}^*$ and $\widetilde{\forall}\, \vartheta \in \mathcal{H}$, the three following conditions hold.

1. $\mathcal{Y}^n_\vartheta \in \mathcal{B}(\mathcal{Y}^n)$.

2. $\widetilde{\forall}\, y^{1:n} \in \mathcal{Y}^n$, the function $z^{1:n} \mapsto \mathcal{T}(\mu_n, \mu_{\vartheta,n})$ belongs to $\mathscr{B}(\mathcal{Y}^n_\vartheta, \mathbb{R}_+)$.

3. $\widetilde{\forall}\, y^{1:n} \in \mathcal{Y}^n$ and $\forall\, \varepsilon \in \,]0, \varepsilon_0[$,

$$\mu^n_\vartheta[D^n_\varepsilon] \geq \mu^n_\vartheta\big[\{\, z^{1:n} \in \mathcal{Y}^n_\vartheta \mid \mathcal{T}(\mu_n, \mu_{\vartheta,n}) \leq \varepsilon \,\}\big].$$

Remark. 3 of B0 holds if, $\forall\, n \in \mathbb{N}^*$, $\widetilde{\forall}\, y^{1:n} \in \mathcal{Y}^n$, $\widetilde{\forall}\, \vartheta \in \mathcal{H}$ and $\forall\, z^{1:n} \in \mathcal{Y}^n_\vartheta$, $\mathcal{D}(y^{1:n}, z^{1:n}) \leq \mathcal{T}(\mu_n, \mu_{\vartheta,n})$.

[B1] (under B0) There exists unique $\mu_\star \in \mathscr{P}(\mathcal{Y})$, possibly in $\mathscr{P}_p(\mathcal{Y})$, s.t. the following occurs.

1. For any $n \in \mathbb{N}^*$, $\omega \mapsto \mathcal{T}(\mu_n, \mu_\star)$ is $\mathcal{A}$-measurable as a function from Ω to $\mathbb{R}_+$.

2. $\mathcal{T}(\mu_n, \mu_\star) \to 0$, $\mathbf{P}$-a.s., as $n \to +\infty$.

Remark. B1 implies that, for any $\xi > 0$, $\mathbf{P}\big[\{\, \omega \in \Omega \mid \mathcal{T}(\mu_n, \mu_\star) > \xi \,\}\big] \to 0$ as $n \to +\infty$.

[B2] (under B1) $\widetilde{\forall}\,\vartheta \in \mathcal{H}$, there exists unique $\mu_\vartheta \in \mathscr{P}(\mathcal{Y})$, possibly in $\mathscr{P}_p(\mathcal{Y})$, s.t. the following occurs.

1. The function $\vartheta \mapsto \mathcal{T}(\mu_\vartheta, \mu_\star)$ belongs to $\mathscr{B}(\mathcal{H}, \mathbb{R}_+)$.

2. $\forall\, n \in \mathbb{N}^*$ and $\widetilde{\forall}\,\vartheta \in \mathcal{H}$, the function $z^{1:n} \mapsto \mathcal{T}(\mu_{\vartheta,n}, \mu_\vartheta)$ belongs to $\mathscr{B}(\mathcal{Y}_\vartheta^n, \mathbb{R}_+)$.

3. There exists $\tau \in [0,1[$ such that, $\widetilde{\forall}\,\vartheta \in \mathcal{H}$ and $\forall\,\varepsilon > 0$,

$$\limsup_n \mu_\vartheta^n\big[\{\, z^{1:n} \in \mathcal{Y}_\vartheta^n \mid \mathcal{T}(\mu_{\vartheta,n}, \mu_\vartheta) > \varepsilon \,\}\big] \le \tau.$$

4. There exist $\sigma \in [0,\tau]$ and $\varepsilon_1 > 0$ such that, $\widetilde{\forall}\,\vartheta \in \mathcal{H}$ and $\forall\,\varepsilon \in \,]0,\varepsilon_1[$,

$$\liminf_n \mu_\vartheta^n\big[\{\, z^{1:n} \in \mathcal{Y}_\vartheta^n \mid \mathcal{T}(\mu_{\vartheta,n}, \mu_\vartheta) > \varepsilon \,\}\big] \ge \sigma.$$

Remark. 3 of B2, without specific requests on ε, is equivalent to any version of that in which upper bounds for ε are imposed. Furthermore if, $\widetilde{\forall}\,\vartheta \in \mathcal{H}$ and $\forall\,\varepsilon > 0$, $\mu_\vartheta^n[\mathcal{T}(\mu_{\vartheta,n}, \mu_\vartheta) > \varepsilon] \to 0$ as $n \to +\infty$ (shortly put), then any $\tau \in [0,1[$ satisfies 3 of B2 while only $\sigma = 0$ but any $\varepsilon_1 > 0$ fulfill 4 of B2.

[B3] (under 1 and 2 of B2) There exists $\vartheta_\star \in \mathcal{H}$ which minimizes $\vartheta \mapsto \mathcal{T}(\mu_\vartheta, \mu_\star)$ over $\mathcal{H}$: simbolically,

$$\vartheta_\star \in \arg\min_{\mathcal{H}} \mathcal{T}\big(\mu_{(\,\cdot\,)}, \mu_\star\big).$$

Notations. We'll denote $\varepsilon_\star \doteq \mathcal{T}(\mu_{\vartheta_\star}, \mu_\star) \equiv \min_{\mathcal{H}} \mathcal{T}\big(\mu_{(\,\cdot\,)}, \mu_\star\big) \ge 0$ and, $\widetilde{\forall}\,\vartheta \in \mathcal{H}$, $\mathcal{T}_\vartheta \coloneqq \mathcal{T}(\mu_\vartheta, \mu_\star) \ge \varepsilon_\star$.

[B4] (under B3) There exist a neighborhood $U_\star \subset \mathcal{H}$ of $\vartheta_\star$, a connected neighborhood $I_0 \subset \mathbb{R}_+$ of zero and a strictly increasing function $\psi \colon I_0 \to \mathbb{R}_+$ all s.t., $\widetilde{\forall}\,\vartheta \in U_\star$,

$$\mathcal{T}_\vartheta - \varepsilon_\star \le \psi\big(\rho_{\mathcal{H}}(\vartheta, \vartheta_\star)\big).$$

Notations. We'll write "for any $(y^{1:n})_n$" meaning to vary of $(y^{1:n}(\omega))_n \equiv (y^{1:n})_n$, with $y^{1:n}(\omega) \equiv y^{1:n}$ in $\mathcal{Y}^n$ for any $n \in \mathbb{N}^*$, w.r.t. a $\omega \in \Omega$. Lastly, for any $\varepsilon > 0$, we'll denote by ε^- any element of $]0,\varepsilon]$.

Proposition. *Under assumptions* B0, B1, 1, 2 *and* 3 *of* B2 *and* B3*, the following occurs so far as*

$$\varepsilon_\star < \varepsilon_0$$

for any $\varepsilon \in \,]0, \varepsilon_0 - \varepsilon_\star[$, $(y^{1:n})_n$ with $n \equiv n_\varepsilon$ large enough and with probability $\mathbf{P}$ going to 1 as $n \to +\infty$.

a. $\pi_{y^{1:n}}^{\varepsilon_\star+\varepsilon}\big[\mathcal{T}_{(\,\cdot\,)} \ge \varepsilon_\star + \varepsilon^-/3\big] \ge (1-\tau)\,\pi\big[\varepsilon_\star + \varepsilon^-/3 \le \mathcal{T}_{(\,\cdot\,)} \le \varepsilon_\star + \varepsilon/3\big]$.

b. $\pi_{y^{1:n}}^{\varepsilon_\star+\varepsilon}\big[\mathcal{H} \setminus \arg\min_{\mathcal{H}} \mathcal{T}_{(\,\cdot\,)}\big] \ge (1-\tau)\,\pi\big[\varepsilon_\star < \mathcal{T}_{(\,\cdot\,)} \le \varepsilon_\star + \varepsilon/3\big]$.

c. *Under assumption* 4 *of* B2*, let's suppose that in* 3 *of* B0 *the equality holds and that $\varepsilon_\star < \varepsilon_1/2$. Then, for any $\varepsilon \in \,]0, \varepsilon_0 - \varepsilon_\star[$ even more enough small,*

$$\lambda_\varepsilon \coloneqq (1-\sigma)\,\pi\big[\mathcal{T}_{(\,\cdot\,)} \le \varepsilon_\star + 5\varepsilon/3\big] + \tau\pi\big[\mathcal{T}_{(\,\cdot\,)} > \varepsilon_\star + 5\varepsilon/3\big] > 0$$

and

$$\pi_{y^{1:n}}^{\varepsilon_\star+\varepsilon}\big[\mathcal{T}_{(\,\cdot\,)} \ge \varepsilon_\star + \varepsilon^-/3\big] \ge \frac{1-\tau}{\lambda_\varepsilon}\,\pi\big[\varepsilon_\star + \varepsilon^-/3 \le \mathcal{T}_{(\,\cdot\,)} \le \varepsilon_\star + \varepsilon/3\big].$$

d. *Under assumption* B4*, for any $\zeta \in I_0 \setminus \{0\}$ and $r > 0$ small enough,*

$$\pi_{y^{1:n}}^{\varepsilon_\star+\varepsilon}\big[\rho_{\mathcal{H}}(\,\bullet\,, \vartheta_\star) \ge r\big] \ge \pi_{y^{1:n}}^{\varepsilon_\star+\varepsilon}\big[\mathcal{T}_{(\,\cdot\,)} \ge \varepsilon_\star + \psi(\zeta)\big]$$

for which lower bounds of a *and eventually* c *hold if also ζ is small enough.*

Proof. First of all, by virtue of the classical Fatou's lemma (and of A0 - b) it's simple to realize that, for any $\varepsilon \in]0, \varepsilon_0 - \varepsilon_\star[$ and $(y^{1:n})_n$, $\int_{\mathcal{H}} \limsup_n \, \mu_\vartheta^n[D^n_{\varepsilon_\star+\varepsilon}] \, \pi(\mathrm{d}\vartheta) > 0$ and, for any $\zeta > 0$,

$$\liminf_n \pi^{\varepsilon_\star+\varepsilon}_{y^{1:n}} \big[\mathcal{T}_{(\,.\,)} \geq \varepsilon_\star + \zeta\big] \geq \frac{\int_{\{\mathcal{T}_{(\,.\,)} \geq \varepsilon_\star+\zeta\}} \liminf_n \mu_\vartheta^n[D^n_{\varepsilon_\star+\varepsilon}] \, \pi(\mathrm{d}\vartheta)}{\int_{\mathcal{H}} \limsup_n \mu_{\vartheta'}^n[D^n_{\varepsilon_\star+\varepsilon}] \, \pi(\mathrm{d}\vartheta')}.$$

a. Since $\int_{\mathcal{H}} \limsup_n \, \mu_\vartheta^n[D^n_{\varepsilon_\star+\varepsilon}] \, \pi(\mathrm{d}\vartheta) \leq 1$, the thesis will be obtained once has been demonstrated that, for any $\varepsilon \in]0, \varepsilon_0 - \varepsilon_\star[$, $\zeta \in]0, \varepsilon/3]$ and $\mathbf{P}$-a.a. $(y^{1:n})_n$ with $n \equiv n_\varepsilon$ large enough to have $\mathcal{T}(\mu_n, \mu_\star) \leq \varepsilon/3$,

$$\int_{\{\mathcal{T}_{(\,.\,)} \geq \varepsilon_\star+\zeta\}} \liminf_n \mu_\vartheta^n[D^n_{\varepsilon_\star+\varepsilon}] \, \pi(\mathrm{d}\vartheta) \geq (1-\tau) \, \pi\big[\varepsilon_\star + \zeta \leq \mathcal{T}_{(\,.\,)} \leq \varepsilon_\star + \varepsilon/3\big].$$

Indeed B0 applies and, thanks to the triangle inequality which holds for $\mathcal{T}$, it's easy to verify that, $\forall\, n \in \mathbb{N}^*$, $\widetilde{\forall}\, y^{1:n} \in \mathcal{Y}^n$, $\widetilde{\forall}\, \vartheta \in \mathcal{H}$ and $\forall\, \varepsilon \in]0, \varepsilon_0 - \varepsilon_\star[$, if $\mathcal{T}_\vartheta \leq \varepsilon_\star + \varepsilon/3$ and $\mathcal{T}(\mu_n, \mu_\star) \leq \varepsilon/3$, then

$$\big\{\, z^{1:n} \in \mathcal{Y}^n_\vartheta \;\big|\; \mathcal{T}(\mu_{\vartheta,n}, \mu_\vartheta) \leq \varepsilon/3 \big\} \subset \big\{\, z^{1:n} \in \mathcal{Y}^n_\vartheta \;\big|\; \mathcal{T}(\mu_{\vartheta,n}, \mu_n) \leq \varepsilon_\star + \varepsilon \big\}$$

and thus

$$\int_{\{\mathcal{T}_{(\,.\,)} \geq \varepsilon_\star+\zeta\}} \liminf_n \mu_\vartheta^n[D^n_{\varepsilon_\star+\varepsilon}] \, \pi(\mathrm{d}\vartheta) \geq \int_{\{\varepsilon_\star+\zeta \leq \mathcal{T}_{(\,.\,)} \leq \varepsilon_\star+\varepsilon/3\}} \liminf_n \mu_\vartheta^n[\mathcal{T}(\mu_{\vartheta,n}, \mu_\vartheta) \leq \varepsilon/3] \, \pi(\mathrm{d}\vartheta)$$

from which we conclude considering that by 3 of B2, $\widetilde{\forall}\, \vartheta \in \mathcal{H}$ and $\forall\, \varepsilon > 0$,

$$\liminf_n \mu_\vartheta^n[\mathcal{T}(\mu_{\vartheta,n}, \mu_\vartheta) \leq \varepsilon/3] \equiv 1 - \limsup_n \mu_\vartheta^n[\mathcal{T}(\mu_{\vartheta,n}, \mu_\vartheta) > \varepsilon/3] \geq 1 - \tau.$$

b. That's a corollary of the previous result: for any $\varepsilon \in]0, \varepsilon_0 - \varepsilon_\star[$ and $(y^{1:n})_n$ with $n \equiv n_\varepsilon$ large enough,

$$\begin{aligned}
\pi^{\varepsilon_\star+\varepsilon}_{y^{1:n}} \big[\mathcal{H} \setminus \arg\min\nolimits_{\mathcal{H}} \mathcal{T}_{(\,.\,)}\big] &\equiv \pi^{\varepsilon_\star+\varepsilon}_{y^{1:n}} \big[\mathcal{T}_{(\,.\,)} > \varepsilon_\star\big] \\
&= \pi^{\varepsilon_\star+\varepsilon}_{y^{1:n}} \big[\textstyle\bigcup_{\zeta \downarrow 0} \big\{\mathcal{T}_{(\,.\,)} \geq \varepsilon_\star + \zeta\big\}\big] \\
&= \sup_{0<\zeta\leq\varepsilon/3} \pi^{\varepsilon_\star+\varepsilon}_{y^{1:n}} \big[\mathcal{T}_{(\,.\,)} \geq \varepsilon_\star + \zeta\big] \\
&\geq (1-\tau) \sup_{0<\zeta\leq\varepsilon/3} \pi\big[\varepsilon_\star + \zeta \leq \mathcal{T}_{(\,.\,)} \leq \varepsilon_\star + \varepsilon/3\big] \\
&= (1-\tau) \, \pi\big[\varepsilon_\star < \mathcal{T}_{(\,.\,)} \leq \varepsilon_\star + \varepsilon/3\big].
\end{aligned}$$

c. It would be sufficient to show that, for any $\varepsilon \in]0, \varepsilon_0 - \varepsilon_\star[$ even more enough small and $\mathbf{P}$-a.a. $(y^{1:n})_n$,

$$\int_{\mathcal{H}} \limsup_n \mu_\vartheta^n[D^n_{\varepsilon_\star+\varepsilon}] \, \pi(\mathrm{d}\vartheta) \leq \lambda_\varepsilon \equiv (1-\sigma) \, \pi\big[\mathcal{T}_{(\,.\,)} \leq \varepsilon_\star + 5\varepsilon/3\big] + \tau \pi\big[\mathcal{T}_{(\,.\,)} > \varepsilon_\star + 5\varepsilon/3\big]$$

(bearing in mind the procedure of the first proof). Indeed, thanks again to the subadditivity of $\mathcal{T}$, $\forall\, n \in \mathbb{N}^*$, $\widetilde{\forall}\, y^{1:n} \in \mathcal{Y}^n$, $\widetilde{\forall}\, \vartheta \in \mathcal{H}$ and $\forall\, \varepsilon \in]0, \varepsilon_0 - \varepsilon_\star[$, if $\mathcal{T}_\vartheta > \varepsilon_\star + 5\varepsilon/3$ and $\mathcal{T}(\mu_n, \mu_\star) \leq \varepsilon/3$, then

$$\big\{\, z^{1:n} \in \mathcal{Y}^n_\vartheta \;\big|\; \mathcal{T}(\mu_{\vartheta,n}, \mu_n) \leq \varepsilon_\star + \varepsilon \big\} \subset \big\{\, z^{1:n} \in \mathcal{Y}^n_\vartheta \;\big|\; \varepsilon/3 < \mathcal{T}(\mu_{\vartheta,n}, \mu_\vartheta) \leq \mathcal{T}_\vartheta + \varepsilon_\star + 4\varepsilon/3 \big\}$$

so, using the hypothesis $\mu_\vartheta^n[D^n_{\varepsilon_\star+\varepsilon}] = \mu_\vartheta^n[\mathcal{T}(\mu_{\vartheta,n}, \mu_n) \leq \varepsilon_\star + \varepsilon]$ (i.e. $\leq$), and breaking the integral,

$$\begin{aligned}
\int_{\mathcal{H}} \limsup_n \mu_\vartheta^n[D^n_{\varepsilon_\star+\varepsilon}] \, \pi(\mathrm{d}\vartheta) \leq &\int_{\{\mathcal{T}_{(\,.\,)} \leq \varepsilon_\star+5\varepsilon/3\}} \limsup_n \mu_\vartheta^n[\mathcal{T}(\mu_{\vartheta,n}, \mu_\vartheta) \leq 2\varepsilon_\star + 3\varepsilon] \, \pi(\mathrm{d}\vartheta) \\
&+ \int_{\{\mathcal{T}_{(\,.\,)} > \varepsilon_\star+5\varepsilon/3\}} \limsup_n \mu_\vartheta^n[\mathcal{T}(\mu_{\vartheta,n}, \mu_\vartheta) > \varepsilon/3] \, \pi(\mathrm{d}\vartheta)
\end{aligned}$$

and finally, by 4 of B2, $\widetilde{\forall}\, \vartheta \in \mathcal{H}$ and $\forall\, \varepsilon \in]0, \varepsilon_0 - \varepsilon_\star[$ small enough to have also $2\varepsilon_\star + 3\varepsilon < \varepsilon_1$,

$$\limsup_n \mu_\vartheta^n[\mathcal{T}(\mu_{\vartheta,n}, \mu_\vartheta) \leq 2\varepsilon_\star + 3\varepsilon] \equiv 1 - \liminf_n \mu_\vartheta^n[\mathcal{T}(\mu_{\vartheta,n}, \mu_\vartheta) > 2\varepsilon_\star + 3\varepsilon] \leq 1 - \sigma.$$

d. Let's fix any $r > 0$ small enough to have $r \leq \zeta$ and $\{\rho_{\mathcal{H}}(\,\cdot\,,\vartheta_\star) < r\} \subseteq U_\star$. Then B4 guarantees that there exists $\mathcal{N} \in \mathcal{B}(\mathcal{H})$ with $\pi[\mathcal{N}] = 0$ such that, for any $\vartheta \in \mathcal{H} \setminus \mathcal{N}$ with $\rho_{\mathcal{H}}(\vartheta,\vartheta_\star) < r$,

$$\mathcal{T}_\vartheta - \varepsilon_\star \leq \psi\big(\rho_{\mathcal{H}}(\vartheta,\vartheta_\star)\big) < \psi(r) \leq \psi(\zeta)$$

and thus $\{\rho_{\mathcal{H}}(\,\cdot\,,\vartheta_\star) < r\} \setminus \mathcal{N} \subset \big\{\mathcal{T}_{(\,\cdot\,)} < \varepsilon_\star + \psi(\zeta)\big\}$ or, equivalently,

$$\{\rho_{\mathcal{H}}(\,\cdot\,,\vartheta_\star) \geq r\} \cup \mathcal{N} \supset \big\{\mathcal{T}_{(\,\cdot\,)} \geq \varepsilon_\star + \psi(\zeta)\big\}$$

hence the thesis ($\forall\, n \in \mathbb{N}^*$, $\forall\, y^{1:n} \in \mathcal{Y}^n$ and $\forall\, \varepsilon \in\,]0, \varepsilon_0 - \varepsilon_\star[$, $\pi^{\varepsilon_\star+\varepsilon}_{y^{1:n}}[\mathcal{N}] = 0$ as well). □

Remark. Let's discuss how a condition consistent with A2 as the following could interact.

[A2′] (under A1) There exist $\delta, \varepsilon' \in\,]0, +\infty[$ and $g \in L^1(\pi)$ with $g \geq \delta\ [\,\pi\,]$ all s.t., $\widetilde{\forall}\ \vartheta \in \mathcal{H}$ and $\forall\ (z^{1:n})_n$ with $z^{1:n} \in D^n_{\varepsilon'}$ for any $n \in \mathbb{N}^*$,

$$\delta \leq \liminf\nolimits_n f^n_\vartheta(z^{1:n}) \quad \text{and} \quad \limsup\nolimits_n f^n_\vartheta(z^{1:n}) \leq g(\vartheta).$$

Proposition. *Under assumptions* B0, B1, 1 *and* 2 *of* B2, B3, A1 *and* A2′*, the following occurs so far as* $\varepsilon_\star < \varepsilon_0 \wedge \varepsilon'$ *and for any* $\varepsilon \in\,]0, \varepsilon_0 \wedge \varepsilon' - \varepsilon_\star[$ *and* $\mathbf{P}$*-a.a.* $(y^{1:n})_n$.

a. *For any* $\zeta > 0$, $\pi^{\varepsilon_\star+\varepsilon}_{y^{1:n}}\big[\mathcal{T}_{(\,\cdot\,)} \geq \varepsilon_\star + \zeta\big] \geq \dfrac{\delta}{\|g\|_1}\,\pi\big[\mathcal{T}_{(\,\cdot\,)} \geq \varepsilon_\star + \zeta\big]$.

b. $\pi^{\varepsilon_\star+\varepsilon}_{y^{1:n}}\big[\mathcal{H} \setminus \arg\min_{\mathcal{H}} \mathcal{T}_{(\,\cdot\,)}\big] \geq \dfrac{\delta}{\|g\|_1}\,\pi\big[\mathcal{H} \setminus \arg\min_{\mathcal{H}} \mathcal{T}_{(\,\cdot\,)}\big]$.

Proof. For any $\varepsilon \in\,]0, \varepsilon_0 \wedge \varepsilon' - \varepsilon_\star[$, $\mathbf{P}$-a.a. $(y^{1:n})_n$ and $\zeta > 0$,

$$\pi^{\varepsilon_\star+\varepsilon}_{y^{1:n}}\big[\mathcal{T}_{(\,\cdot\,)} \geq \varepsilon_\star + \zeta\big] = \frac{\int_{\{\mathcal{T}_{(\,\cdot\,)} \geq \varepsilon_\star+\zeta\}} \pi(\mathrm{d}\vartheta) \int_{D^n_{\varepsilon_\star+\varepsilon}} f^n_\vartheta(z^{1:n})\,\mathrm{d}z^{1:n}}{\int_{\mathcal{H}} \pi(\mathrm{d}\vartheta') \int_{D^n_{\varepsilon_\star+\varepsilon}} f^n_{\vartheta'}(\bar{z}^{1:n})\,\mathrm{d}\bar{z}^{1:n}}.$$

a. We use directly A2′ provided we reduce further ε (note the cancellation of both the terms $\mathrm{m}[D^n_{\varepsilon_\star+\varepsilon}]$).
b. That's an elementary consequence of the above result in a way already seen previously. □

Starting reference

[1] E. Bernton, P. E. Jacob, M. Gerber, C. P. Robert. *Approximate Bayesian computation with the Wasserstein distance.* J. R. Statist. Soc. B (2019). 81, Part 2, pp. 235–269.