HAL open science

Concentration results for approximate Bayesian computation without identifiability

Hien Nguyen, Trungtin Nguyen, Julyan Arbel, Florence Forbes

To cite this version:

Hien Nguyen, Trungtin Nguyen, Julyan Arbel, Florence Forbes. Concentration results for approximate Bayesian computation without identifiability. 2023. hal-03987197

HAL Id: hal-03987197

https://hal.science/hal-03987197

Preprint submitted on 14 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Concentration results for approximate Bayesian computation without identifiability

BY H. D. NGUYEN

School of Mathematics and Physics, Univ. of Queensland
St. Lucia 4067, Queensland, Australia.
h.nguyen7@uq.edu.au

T. T. NGUYEN, J. ARBEL, AND F. FORBES

Univ. Grenoble Alpes, Inria, CNRS
Grenoble INP, LJK 38000 Grenoble, France.
trung-tin.nguyen@inria.fr julyan.arbel@inria.fr florence.forbes@inria.fr

SUMMARY

We study the large sample behaviors of approximate Bayesian computation (ABC) posterior measures in situations when the data generating process is dependent on non-identifiable parameters. In particular, we establish the concentration of posterior measures on sets of arbitrarily measure that contain the equivalence set of the data generative parameter, when the sample size tends to infinity. Our theory also makes weak assumptions regarding the measurement of discrepancy between the data set and simulations, and in particular, does not require the use of summary statistics and is applicable to a broad class of kernelized ABC algorithms. We provide useful illustrations and demonstrations of our theory in practice, and offer a comprehensive assessment of the nature in which our findings complement other results in the literature.

Some key words: Approximate Bayesian computation; Posterior consistency; Pseudo-posterior measure; Non-identifiable; Concentration of mass; Large sample theory

1. INTRODUCTION

Approximate Bayesian computation (ABC) has become a leading paradigm for drawing inference when data generating processes do not possess known or tractable likelihood functions. Recent expositions regarding the history, varieties, and example applications of ABC can be found in the comprehensive volume of Sisson et al. (2018). Other reviews of the ABC literature can be found in Marin et al. (2012) and Beaumont (2019).

We shall place our investigation within the following general setting. Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space with element ω and expectation operator E. We let $\mathbf{X}_\infty = (X_i)_{i\in\mathbb{N}}$ be a sequence of random variables, where $X_i : \Omega \to \mathbb{X} \subset \mathbb{R}^d$ and $\mathbf{X}_n$ is a tuple containing the first n elements of $\mathbf{X}_\infty$ and has probability measure on the space $(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n))$ which is absolutely continuous with respect to $\mathfrak{m}$ (typically the counting or Lebesgue measure), defined by density $\mathrm{p}(\mathbf{x}_n|\theta_0)$, for some $\theta_0 \in \mathbb{T} \subset \mathbb{R}^q$ $(i, d, n, q \in \mathbb{N})$, which we shall call the *generative parameter*. We further let $\mathbf{Y}_\infty = (Y_i)_{i\in\mathbb{N}}$ be another sequence of random variables $Y_i : \Omega \to \mathbb{X}$, where the partial sequence $\mathbf{Y}_n$ also has a probability measure on $(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n))$ with density $\mathrm{p}(\mathbf{y}_n|\theta)$ and $\theta \in \mathbb{T}$. Throughout, $\mathbf{X}_\infty$ and $\mathbf{Y}_\infty$ are taken to be independent.

Let $w(\cdot,\cdot):\mathbb{R}_{\geq 0}\times\mathbb{R}_{>0}\to\mathbb{R}_{\geq 0}$, $(\delta,\epsilon)\mapsto w(\delta,\epsilon)$ be a *weighting* or *kernel* function, where we typically think of ϵ as a *threshold* parameter. For $\mathbf{x}_n,\mathbf{y}_n\in\mathbb{X}^n$, we then let $(\mathbf{x}_n,\mathbf{y}_n)\mapsto D(\mathbf{x}_n,\mathbf{y}_n)\in\mathbb{R}_{\geq 0}$ denote some notion of a *distance* or a *discrepancy* between $\mathbf{x}_n$ and $\mathbf{y}_n$ and further let θ be a random element on the probability space $(\mathbb{T},\mathcal{B}(\mathbb{T}),\Pi)$, where Π is absolutely continuous with respect to measure $\mathfrak{n}$, with density $\pi(\theta)$. We concern ourselves with studying the behavior of the *pseudo-posterior density*:

$$\pi_\epsilon(\theta|\mathbf{X}_n)=\frac{\pi(\theta)\,\mathcal{L}_\epsilon(\mathbf{X}_n|\theta)}{\mathcal{C}_\epsilon(\mathbf{X}_n)},\ \epsilon\in\mathbb{R}_{>0}, \tag{1}$$

where

$$\mathcal{L}_\epsilon(\mathbf{X}_n|\theta)=\int_{\mathbb{X}^n}w(D(\mathbf{X}_n,\mathbf{y}_n),\epsilon)\,\mathrm{p}(\mathbf{y}_n|\theta)\,\mathrm{d}\mathfrak{m}(\mathbf{y}_n)$$

and

$$\mathcal{C}_\epsilon(\mathbf{X}_n)=\int_{\mathbb{T}}\pi(\theta)\,\mathcal{L}_\epsilon(\mathbf{X}_n|\theta)\,\mathrm{d}\mathfrak{n}(\theta).$$

We shall also consider the *coarsened-posterior density*:

$$\bar{\pi}_{\epsilon,n}(\theta)=\frac{\pi(\theta)\,\mathrm{E}\{\mathcal{L}_\epsilon(\mathbf{X}_n|\theta)\}}{\bar{\mathcal{C}}_{\epsilon,n}},\ \epsilon\in\mathbb{R}_{>0}, \tag{2}$$

where

$$\bar{\mathcal{C}}_{\epsilon,n}=\int_{\mathbb{T}}\pi(\theta)\,\mathrm{E}\{\mathcal{L}_\epsilon(\mathbf{X}_n|\theta)\}\,\mathrm{d}\mathfrak{n}(\theta).$$

We note that the pseudo-posterior density of form (1) is often studied in the context of importance sampling ABC (Karabatsos & Leisen, 2018; Nguyen et al., 2020) or kernel ABC (Park et al., 2016; Li & Fearnhead, 2018b,a) algorithms. Our study of the coarsened-posterior density (2) can be viewed as a homage to and generalization of results from Miller & Dunson (2019), who consider the case where $\mathbf{X}_\infty=\mathbf{x}_\infty$, with probability 1, for some fixed sequence $\mathbf{x}_\infty$ (cf. Miller & Dunson, 2019, Eqn. 2.1). Of course, the usual accept/reject ABC algorithms (e.g. Marin et al. 2012, Algorithm 2) can be studied in this setting by taking $w(\delta,\epsilon)=[\![\delta<\epsilon]\!]$, where we use the *Iverson bracket* notation: $[\![\mathsf{A}]\!]=1$ if statement A is true, and $[\![\mathsf{A}]\!]=0$ otherwise. We shall refer to both (1) and (2) as *ABC posterior densities*, when it is immaterial to distinguish between them.

In recent works the asymptotic behaviors of (1) has been investigated by Bernton et al. (2019), Nguyen et al. (2020), and Legramanti et al. (2022), when $\epsilon\to 0$, for fixed n; Jiang (2018), and Nguyen et al. (2020), when $\epsilon>0$ is fixed and $n\to\infty$; and Frazier et al. (2018), Li & Fearnhead (2018b), Li & Fearnhead (2018a), Bernton et al. (2019), Frazier et al. (2020), and Legramanti et al. (2022), when $\epsilon\to 0$ and $n\to\infty$. Miller & Dunson (2019) provide analysis for the behavior of (2) when $\epsilon>0$ is fixed and $n\to\infty$, in the special case when $\mathbf{X}_\infty=\mathbf{x}_\infty$. Our current work extends upon the best known results when $\epsilon>0$ and $n\to\infty$. Namely, we prove broadly applicable results regarding the concentration of mass of the *pseudo-posterior* and *coarsened-posterior measures* $\Pi_{\epsilon,n}$ and $\bar{\Pi}_{\epsilon,n}$ (collectively, *ABC posterior measures*) as $n\to\infty$, in practical cases when the densities $\mathrm{p}(\mathbf{x}_n|\theta)$ are not identifiable with respect to $\theta\in\mathbb{T}$. We shall elaborate in the sequel.

2. Weak convergence of ABC posterior measures

Since we are operating on both the measure spaces $(\Omega, \mathcal{F}, \mathrm{P})$ and $(\mathbb{T}, \mathcal{B}(\mathbb{T}), \Pi)$, we shall use the convention of saying that events in the prior space are almost surely true (a.s.; with respect to ω), and that events in the latter space occur almost everywhere (a.e.; with respect to θ), when the events occur with probability 1 in the respective spaces.

Let $D_\infty(\cdot,\cdot) : \mathbb{T} \times \mathbb{T} \to \mathbb{R}_{\geq 0}$ and make the following assumptions:

A1 For each $\theta \in \mathbb{T}$, $D(\mathbf{X}_n, \mathbf{Y}_n) \xrightarrow[n\to\infty]{\text{a.s.}} D_\infty(\theta_0, \theta)$.

A2 For each $\epsilon > 0$, $\sup_{\delta \in \mathbb{R}_{\geq 0}} w(\delta, \epsilon) < \infty$.

A3 For each $\epsilon > 0$, $w(D_\infty(\theta_0, \cdot), \epsilon) : \mathbb{T} \to \mathbb{R}_{\geq 0}$ is continuous on a set $\tilde{\mathbb{T}} \subset \mathbb{T}$, such that $\Pi\left(\tilde{\mathbb{T}}\right) = 1$.

Theorem 1. *Under A1–A3, if*

$$\int_{\mathbb{T}} \pi(\theta)\, w(D_\infty(\theta_0, \theta), \epsilon)\, \mathrm{d}\mathfrak{n}(\theta) > 0,$$

then (i) for almost every $\theta \in \mathbb{T}$:

$$\bar{\pi}_{\epsilon,n}(\theta) \underset{n\to\infty}{\longrightarrow} \pi_\epsilon(\theta) = \frac{\pi(\theta)\, w(D_\infty(\theta_0, \theta), \epsilon)}{\int_{\mathbb{T}} \pi(\tau)\, w(D_\infty(\theta_0, \tau), \epsilon)\, \mathrm{d}\mathfrak{n}(\tau)},$$

and (ii) for almost every $\theta \in \mathbb{T}$:

$$\pi_\epsilon(\theta | \mathbf{X}_n) \xrightarrow[n\to\infty]{a.s.} \pi_\epsilon(\theta).$$

Let us first discuss the assumptions. A1 holds in a variety of settings, and is generally obtained by combining a strong law of large numbers and continuous mapping theorem. Notice that no assumption is made regarding the independence between elements of the sequences $\mathbf{X}_\infty$ and $\mathbf{Y}_\infty$. Further, no direct conditions are imposed on the limiting function D_∞, as it always appears composed with the weight function w, as in A3. The latter assumption requires continuity on a set of probability one with respect to the prior distribution, which is verified for the accept/reject kernel, for instance. Lastly, A2 is a simple requirement that the weight function w is bounded.

Remark 1. Theorem 1(i) can be viewed as a generalization of the large-sample results regarding the coarsened-posterior, proved in Miller & Dunson (2019, Sec. S3.1), extended to cases where $\mathbf{X}_\infty$ is a stochastic sequence with non-degenerate measure. Theorem 1(ii) makes the same conclusion as Jiang (2018, Thm. 1) and Nguyen et al. (2020, Thm. 2), although the proofs in the aforementioned works require some clarification, which we take the opportunity to make. All proofs of main results can be found in the Appendix of this text, in the Supplementary Materials.

To proceed, we require the notion of *almost sure weak convergence*, which is defined in Berti et al. (2006) and Grübel & Kabluchko (2016). Let $\mathscr{P}(\mathbb{T})$ be the space of probability measures on $\mathbb{T}$, and let $(\Pi_{\omega,n})_{n\in\mathbb{N}}$ be a sequence measures in $\mathscr{P}(\mathbb{T})$, indexed by $\omega \in \Omega$. We say that $\Pi_{\omega,n}$ almost surely weakly converges to $\Pi \in \mathscr{P}(\mathbb{T})$, if there exists a set $\tilde{\Omega} \in \mathcal{F}$ such that $\mathrm{P}\left(\tilde{\Omega}\right) = 1$ and for every $\omega \in \tilde{\Omega}$, $\Pi_{\omega,n}$ converges weakly to Π, as $n \to \infty$. By the usual definition of weak convergence, this is equivalent to the condition that

$$\int_{\mathbb{T}} f(\theta)\, \mathrm{d}\Pi_{\omega,n}(\theta) \underset{n\to\infty}{\longrightarrow} \int_{\mathbb{T}} f(\theta)\, \mathrm{d}\Pi(\theta),$$

for every bounded and continuous function $f : \mathbb{T} \to \mathbb{R}$, for each $\omega \in \tilde{\Omega}$.

COROLLARY 1. *Let the measures* $\Pi_{\epsilon,n}$ *and* $\bar{\Pi}_{\epsilon,n}$ *be defined by*

$$\Pi_{\epsilon,n}(\mathbb{A}) = \int_{\mathbb{A}} \pi_\epsilon(\theta|\mathbf{X}_n)\,\mathrm{d}\mathfrak{n}(\theta), \text{ and } \bar{\Pi}_{\epsilon,n}(\mathbb{A}) = \int_{\mathbb{A}} \bar{\pi}_{\epsilon,n}(\theta)\,\mathrm{d}\mathfrak{n}(\theta),$$

for $\mathbb{A} \in \mathcal{B}(\mathbb{T})$*, respectively, for each* $n \in \mathbb{N}$*. Then, under A1–A3, (i)* $\bar{\Pi}_{\epsilon,n}$ *converges weakly to* Π_ϵ*, and under A1–A3, (ii)* $\Pi_{\epsilon,n}$ *converges almost surely weakly to* Π_ϵ*, where* Π_ϵ *is defined by*

$$\Pi_\epsilon(\mathbb{A}) = \int_{\mathbb{A}} \pi_\epsilon(\theta)\,\mathrm{d}\mathfrak{n}(\theta).$$

3. CONCENTRATION OF MASS

We let $\mathbb{B}_\theta = \{\tau \in \mathbb{R}^q : \|\theta - \tau\|_2 < 1\}$ be the (open) unit ball in $\mathbb{R}^q$ with respect to the Euclidean norm $\|\cdot\|_2$, centered at $\theta \in \mathbb{R}^q$, and we note that we can *scale* $\mathbb{B}_0$ by a factor $\lambda > 0$ to obtain balls $\lambda\mathbb{B}_0 = \{\lambda b : b \in \mathbb{B}_0\}$ with any radius λ. Let $\mathbb{T}_0 \subset \mathbb{T}$ be some set of interest. We say that $\Theta_\lambda \subset \mathbb{T}$ is the set of the λ*-covering centres* for $\mathbb{T}_0$ if $\mathbb{T}_0 \subset \Theta_\lambda \oplus \lambda\mathbb{B}_0$, where $\mathbb{A} \oplus \mathbb{B} = \{a + b : a \in \mathbb{A}, b \in \mathbb{B}\}$ is the *Minkowski sum*. If Θ_λ is a finite set, then we will denote its *cardinality* by $|\Theta_\lambda|$. We shall denote the Lebesgue measure on $\mathbb{R}^q$ by Leb.

Let $\theta_0 \in \mathbb{T}$ and make the following assumptions:

B1 Let $D_\infty(\theta_0, \cdot) : \mathbb{T} \to \mathbb{R}_{\geq 0}$ and assume that the set of its zeroes $\Theta_0 = \{\theta \in \mathbb{T} : D_\infty(\theta_0, \theta) = 0\}$ is non-empty.

B2 For each $\lambda > 0$, there exists an $\epsilon > 0$, such that if $D_\infty(\theta_0, \theta) < \epsilon$, then $\inf_{\tau \in \Theta_0} \|\theta - \tau\|_2 < \lambda$.

B3 For each $\lambda > 0$, there exists a covering of Θ_0 with λ-covering centres $\Theta_\lambda \subset \mathbb{T}$, such that $\lambda^q |\Theta_\lambda| \underset{\lambda \to 0}{\longrightarrow} 0$.

B4 The weight function $w(\cdot, \cdot)$ can be decomposed as

$$w(\delta, \epsilon) = W(\delta, \epsilon)\,[\![\varphi(\delta) < \epsilon]\!],$$

where $W : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ and $\varphi : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is strictly increasing and bijective, with $\varphi(0) = 0$.

THEOREM 2. *Assume B1–B3. Then, (i) as* $\lambda \to 0$*,*

$$\mathrm{Leb}(\Theta_0 \oplus \lambda\mathbb{B}_0) \to 0.$$

Let $\int_{\mathbb{T}} \pi(\theta)\, w(D_\infty(\theta_0, \theta), \epsilon)\,\mathrm{d}\mathfrak{n}(\theta) > 0$*, for each* $\epsilon > 0$*. If we further assume A1–A3 and B4, then (ii) for every* $\lambda > 0$*, there exists an* $\epsilon > 0$*, such that*

$$\bar{\Pi}_{\epsilon,n}(\Theta_0 \oplus \lambda\mathbb{B}_0) \underset{n \to \infty}{\longrightarrow} 1,$$

and (iii) for every $\lambda > 0$*, there exists an* $\epsilon > 0$*, such that*

$$\Pi_{\epsilon,n}(\Theta_0 \oplus \lambda\mathbb{B}_0) \underset{n \to \infty}{\overset{a.s.}{\longrightarrow}} 1.$$

Let us now unpack the assumptions and conclusions of Theorem 2. Firstly, B1 simply assumes that the zeroes of $D_\infty(\theta_0, \theta) = 0$ exists. If $D_\infty : \mathbb{T} \times \mathbb{T} \to \mathbb{R}_{\geq 0}$ is a pseudometric on $\mathbb{T}$ (as per Richmond 2020, Sec. 11.1), then we may consider Θ_0 to be the equivalent class defined by the generative parameter: $\Theta_0 = [\theta_0]$, corresponding to the equivalence relationship: $D_\infty(\theta_0, \theta) = 0$. B2 is a primitive boundedness and an identification assumption. Here, if we again consider D_∞ to be a pseudometric, then we may consider the definition of a new metric over equivalent

classes $\mathbb{T}_\sim = \{[\theta] : \theta \in \mathbb{T}\}$ denoted by $D_\sim : \mathbb{T}_\sim \times \mathbb{T}_\sim \to \mathbb{R}_{\geq 0}$. Then B2 is satisfied by taking $D_\sim$ to be Hölder continous with respect to the *Euclidean Hausdorff distance* on sets $\|\cdot - \cdot\|_\sim : \mathbb{T}_\sim \times \mathbb{T}_\sim \to \mathbb{R}_{\geq 0}$ defined by

$$\|[\theta] - [\tau]\|_\sim = \max\left\{\sup_{t_1 \in [\theta]} \inf_{t_2 \in [\tau]} \|t_1 - t_2\|_2 , \sup_{t_1 \in [\tau]} \inf_{t_2 \in [\theta]} \|t_1 - t_2\|_2\right\},$$

in the sense that there exists an $L, K > 0$ such that $\|[\theta] - [\tau]\|_\sim \leq L D_\sim ([\theta], [\tau])^K$, for every $\theta, \tau \in \mathbb{T}$.

Another simplification of B2 can be obtained by taking $\Theta_0 = \{\theta_0\}$ to be a singleton. Then, B2 is implied by the condition that D_∞ is a metric and the identity function $\theta \mapsto \theta$ continuously maps the metric spaces $(\mathbb{T}, D_\infty)$ and $(\mathbb{T}, \|\cdot - \cdot\|_2)$, uniformly. This again is implied by the Lipschitz/Hölder condition that there exists an $L, K > 0$ such that $\|\theta - \tau\|_2 \leq L D_\infty (\theta, \tau)^K$, for every $\theta, \tau \in \mathbb{T}$.

Next, B3 states that there exists a set of covering centres Θ_λ of Θ_0 that does not grow too quickly, since we need the Lebesgue measure of the Euclidean balls covering Θ_0 to be small when λ is small, in order to establish Part (i) of the theorem. Here, the assumption is automatically fulfilled when Θ_0 is a finite set, which particularly holds true when Θ_0 is a singleton. Note, however that countability of Θ_0 is not a sufficient condition for B3.

Lastly, B4 implies that $w(D_\infty(\theta_0, \cdot), \epsilon)$ and its support are bounded, when taken together with A2. This holds for the classical *accept/reject kernel* $w(\delta, \epsilon) = [\![\delta < \epsilon]\!]$, but not for kernels with unbounded support, such as the Gaussian kernel $w(\delta, \epsilon) = \exp\left(-\delta^2/\epsilon\right)$, that is considered in Park et al. (2016) and Nguyen et al. (2020).

We can interpret Theorem 2 as follows. Part (i) states that under B1–B3, we can always cover the zero set Θ_0, of elements of $\mathbb{T}$ that are indistinguishable from θ_0, using D_∞, by Euclidean balls with radius λ, such that the total volume of the covering vanishes with respect to the Lebesgue measure, as $\lambda \to 0$. Then, Part (ii) states that if we further assume A1–A3 and B4, we have the fact that given any choice of $\lambda > 0$, we can pick an $\epsilon > 0$ such that the ABC posterior measure, $\bar{\Pi}_{\epsilon,n}$ or $\Pi_{\epsilon,n}$, of the covering of Θ_0 converges to full mass as $n \to \infty$. That is, regardless of how small the Lebesgue measure of our covering of Θ_0 is, we can always choose an $\epsilon > 0$ such that the ABC posterior always eventually concentrates its mass entirely within the covering. Part (iii) then makes the equivalent conclusion regarding the pseudo-posterior $\Pi_{\epsilon,n}$, in the almost sure sense.

The conclusions can be viewed as a kind of *posterior consistency*, as defined in Ghosal & Van der Vaart (2017, Ch. 6), where posterior consistency requires that the posterior measure concentrates on a point mass (with zero Lebesgue measure, in the continuous case), as $n \to \infty$, in some sense, in the case where Θ_0 is a singleton. Here, we can call our conclusion a *nearly posterior consistency* result, since we obtain the fact that the ABC posterior measures concentrate on a sets of negligible mass, instead, for any size of negligibility, and for potentially uncountable Θ_0.

Remark 2. Like Bernton et al. (2019), Frazier et al. (2020), and Legramanti et al. (2022), we can also permit misspecification between the density of the data generating process $\mathrm{p}(\mathbf{x}_n|\theta_0)$ and that of the simulated data $\mathrm{p}(\mathbf{y}_n|\theta)$, by allowing for the possibility that $D_\infty(\theta_0, \theta) > 0$, for all $\theta \in \mathbb{T}$. This can be achieved by replacing Θ_0 in B1 by $\Theta_* = \{\theta \in \mathbb{T} : D_\infty(\theta_0, \theta) = \epsilon_*\}$, where $\epsilon_* = \min_{\theta \in \mathbb{T}} D_\infty(\theta_0, \theta)$ (assuming that ϵ_* exists), and replacing B2 by the condition: for each $\lambda > 0$, there exists an $\epsilon > 0$, such that if $D_\infty(\theta_0, \theta) < \epsilon + \epsilon_*$, then $\inf_{\tau \in \Theta_*} \|\theta - \tau\|_2 < \lambda$.

This then provides a result in situations when the underlying data generating process is both unidentifiable and misspecified.

4. ILLUSTRATIVE EXAMPLES OF POSTERIOR CONCENTRATION

4.1. *Identifiable normal model*

Let us suppose that $\mathbf{X}_\infty$ is an independent and identically distributed (IID) sequence defined by $X_i \sim \mathrm{N}(\theta_0, 1)$ for each $i \in \mathbb{N}$, and that $\mathbf{Y}_\infty$ is independent of $\mathbf{X}_\infty$, with elements $Y_i \sim \mathrm{N}(\theta, 1)$, where $\theta_0, \theta \in \mathbb{T} = \mathbb{R}$. We will take $\pi(\theta) = \phi(\theta; 0, 1)$, where $\phi(\cdot; \mu, \sigma^2)$ is the normal density function with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$. We shall use the distance

$$D(\mathbf{X}_n, \mathbf{Y}_n) = \left| n^{-1} \sum_{i=1}^{n} X_i - n^{-1} \sum_{i=1}^{n} Y_i \right|. \tag{3}$$

The law of large numbers and continuous mapping implies that $D(\mathbf{X}_n, \mathbf{Y}_n) \xrightarrow[n\to\infty]{\text{a.s.}} D_\infty(\theta_0, \theta) = |\theta_0 - \theta|$. Provided that w satisfies A2 and A3, then Theorem 1 implies that the coarsened-posterior and pseudo-posterior densities

$$\bar{\pi}_{\epsilon,n}(\theta) = \frac{\phi(\theta; 0, 1)\, \mathrm{E}\left\{\int_{\mathbb{R}^n} w(D(\mathbf{X}_n, \mathbf{y}_n), \epsilon) \prod_{i=1}^{n} \phi(y_i; \theta, 1)\, \mathrm{d}\mathbf{y}_n\right\}}{\int_{\mathbb{R}} \phi(\tau; 0, 1)\, \mathrm{E}\left\{\int_{\mathbb{R}^n} w(D(\mathbf{X}_n, \mathbf{y}_n), \epsilon) \prod_{i=1}^{n} \phi(y_i; \tau, 1)\, \mathrm{d}\mathbf{y}_n\right\} \mathrm{d}\tau} \tag{4}$$

and

$$\pi_{\epsilon,n}(\theta) = \frac{\phi(\theta; 0, 1) \int_{\mathbb{R}^n} w(D(\mathbf{X}_n, \mathbf{y}_n), \epsilon) \prod_{i=1}^{n} \phi(y_i; \theta, 1)\, \mathrm{d}\mathbf{y}_n}{\int_{\mathbb{R}} \phi(\tau; 0, 1) \int_{\mathbb{R}^n} w(D(\mathbf{X}_n, \mathbf{y}_n), \epsilon) \prod_{i=1}^{n} \phi(y_i; \tau, 1)\, \mathrm{d}\mathbf{y}_n \mathrm{d}\tau} \tag{5}$$

converge to

$$\pi_\epsilon(\theta) = \frac{\phi(\theta; 0, 1)\, w(|\theta_0 - \theta|, \epsilon)}{\int_{\mathbb{R}} \phi(\tau; 0, 1)\, w(|\theta_0 - \tau|, \epsilon)\, \mathrm{d}\tau}, \tag{6}$$

where we write from now on $\mathrm{d}\mathfrak{m}$ and $\mathrm{d}\mathfrak{n}$ as d for the Lebesgue measure. To make the conclusions of Theorem 2, we require a choice of w that satisfies B4. Two possibilities are the venerable accept/reject kernel and the *triweight kernel*:

$$w(\delta, \epsilon) = \left\{1 - (\delta/\epsilon)^2\right\}^3 [\![\delta < \epsilon]\!],$$

which corresponds to a choice of $W(\delta, \epsilon) = \left\{1 - (\delta/\epsilon)^2\right\}^3$ and $\varphi(\delta) = \delta$. In either case, since Θ_0 is equal to the singleton $\{\theta_0\}$, it is then procedural to verify the remaining assumptions B1–B3 of Theorem 2, which then implies that the ABC posteriors can be made to concentrate on sets of arbitrarily small Lebesgue measure by making ϵ sufficiently small. We visualize the concentration of the limiting measure Π_ϵ in Figure 1, for both the cases of the accept/reject and triweight kernels.

4.2. *Finitely unidentifiable normal model*

We now instead suppose that $\mathbf{X}_\infty$ is an IID sequence defined by $X_i \sim \mathrm{N}(\theta_0^2, 1)$ for each $i \in \mathbb{N}$, and that $\mathbf{Y}_\infty$ is independent of $\mathbf{X}_\infty$, with elements $Y_i \sim \mathrm{N}(\theta^2, 1)$, where $\theta_0, \theta \in \mathbb{T} = \mathbb{R}$. We again use the distance (3), which almost surely converges to $D_\infty(\theta_0, \theta) = |\theta_0^2 - \theta^2|$.

For non-zero generative parameter θ_0, B1 holds with $\Theta_0 = \{-\theta_0, \theta_0\}$. B2 holds since $|\theta_0^2 - \theta^2| < \epsilon$ implies that $\min_{\tau \in \{-\theta_0, \theta_0\}} \{|\tau - \theta|\} < \sqrt{\epsilon} = \lambda$. The verification of the remaining

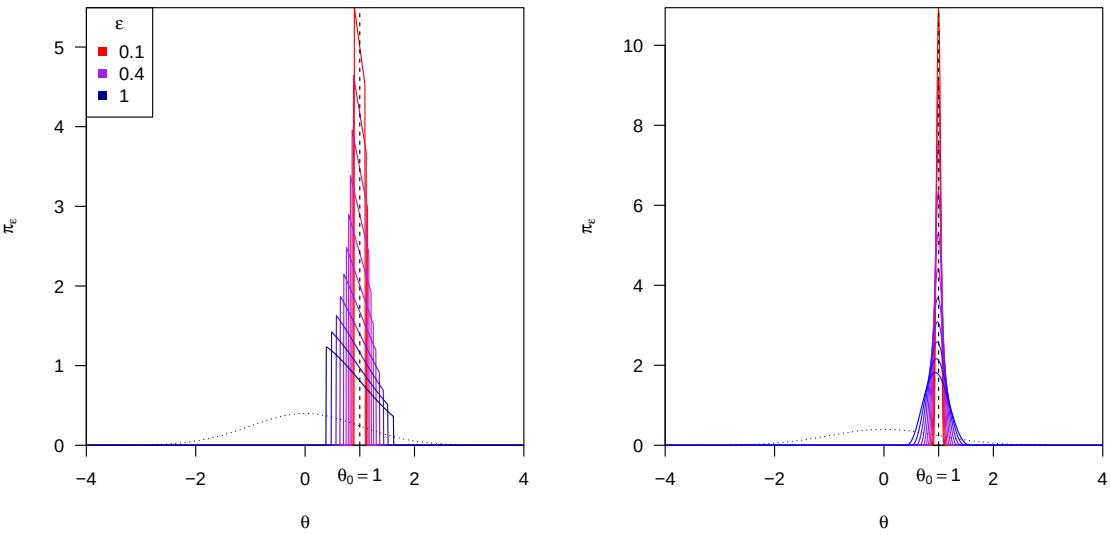

Fig. 1. Limiting posterior densities of form (6) using the accept/reject kernel (left) and triweight kernel (right), with $\theta_0 = 1$, are graphed for a range of values of 11 threshold values ϵ, between 0.1 and 1. The prior densities are drawn with dotted lines.

assumptions A1–A3 and B3–B4 for the case of w set to the accept/reject and triweight kernels follows from analogous arguments to those made in Section 4.1. As such, we have the fact that the ABC posterior measures concentrate mass on sets with arbitrarily small Lebesgue measure, for sufficiently small $\epsilon > 0$, and have respective limiting densities:

$$\pi_\epsilon(\theta) = \frac{\phi(\theta; 0, 1) \left[\!\left[\left|\theta_0^2 - \theta^2\right| < \epsilon \right]\!\right]}{\int_{\mathbb{R}} \phi(\tau; 0, 1) \left[\!\left[\left|\theta_0^2 - \tau^2\right| < \epsilon \right]\!\right] \mathrm{d}\tau}, \tag{7}$$

and

$$\pi_\epsilon(\theta) = \frac{\phi(\theta; 0, 1) \left\{ 1 - \left(\frac{\left|\theta_0^2 - \theta^2\right|}{\epsilon} \right)^2 \right\}^3 \left[\!\left[\left|\theta_0^2 - \theta^2\right| < \epsilon \right]\!\right]}{\int_{\mathbb{R}} \phi(\tau; 0, 1) \left\{ 1 - \left(\frac{\left|\theta_0^2 - \tau^2\right|}{\epsilon} \right)^2 \right\}^3 \left[\!\left[\left|\theta_0^2 - \tau^2\right| < \epsilon \right]\!\right] \mathrm{d}\tau}, \tag{8}$$

when w is taken to be either the accept/reject or triweight kernels. We plot the concentration of mass of the corresponding measures for various values of $\epsilon > 0$ in Figure 2.

4.3. *Infinitely unidentifiable normal model*

We further complicate the situation by supposing that $\mathbf{X}_\infty$ is an IID sequence defined by $X_i \sim \mathrm{N}(||\theta_0||_1, 1)$ for each $i \in \mathbb{N}$, and that $\mathbf{Y}_\infty$ is independent of $\mathbf{X}_\infty$, with IID elements $Y_i \sim \mathrm{N}(||\theta||_1, 1)$. In this example, $\theta_0 = (\theta_{01}, \theta_{02})$ and $\theta = (\theta_1, \theta_2)$ are elements in $\mathbb{T} = \mathbb{R}^2$ with $||\theta_0||_1 = |\theta_{01}| + |\theta_{02}|$ and $||\theta||_1 = |\theta_1| + |\theta_2|$ denoting their L_1-norms. We naturally use the prior $\pi(\theta) = \phi(\theta_1; 0, 1)\, \phi(\theta_2; 0, 1)$ and consider the distance (3), which almost surely converges to $D_\infty(\theta_0, \theta) = |||\theta_0||_1 - ||\theta||_1|$. B1 holds with

$$\Theta_0 = \{\tau = (\tau_1, \tau_2) : ||\tau||_1 = ||\theta_0||_1\}$$

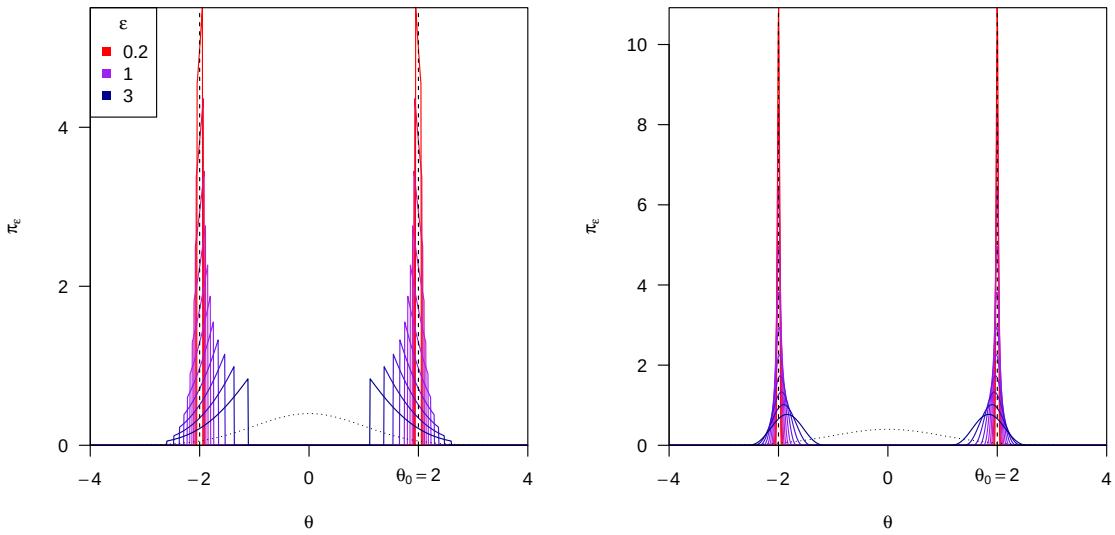

Fig. 2. Limiting posterior densities (7) with accept/reject kernel (left) and (8) with triweight kernel (right), with $\theta_0 = 2$, are graphed for a range of 11 threshold values ϵ, between 0.2 and 3. The prior densities are drawn with dotted lines.

and B2 holds with $\epsilon = \lambda$ since $|||\theta_0||_1 - ||\theta||_1| < \epsilon$ implies that $\inf_{\tau \in \Theta_0} \{||\tau - \theta||_1\} < \epsilon = \lambda$ and $\|\tau - \theta\|_2 \leq \|\tau - \theta\|_1$ for each $\tau \in \Theta_0$. We also have to verify B3, which can be achieved by noting that Θ_0 is the contour of the L_1-ball of radius $||\theta_0||_1$ and thus consists of four line segments, each of length $a\sqrt{2}$, where $a = ||\theta_0||_1$. We can cover each of line segments with Euclidean balls of radius $\lambda > 0$ by placing a ball on each of the $a\sqrt{2}/(2\lambda)$ equidistant points along the line, of distance 2λ apart. Then, we can check B3 by evaluating $\lambda^q \left|\Theta_\lambda\right| = 4 \times \lambda^2 \left(a\sqrt{2}/(2\lambda)\right) = 2a\lambda\sqrt{2}$, which approaches zero as $\lambda \to 0$, as required. The verification of A1–A3 and B4 for concluding Theorem 1 Part (i) and Theorem 2 Part (ii) using w set to the accept/reject and triweight kernels is then procedural. We demonstrate concentration of mass by plotting the support of the limits of both of the ABC posterior measures, defined by (4) and (5), in Figure 3. For both kernels, the support satisfies $|||\theta||_1 - ||\theta_0||_1| \leq \epsilon$ or equivalently $||\theta_0||_1 - \epsilon \leq ||\theta||_1 \leq ||\theta_0||_1 + \epsilon$, which corresponds to the region between the L_1-balls of radii $||\theta_0||_1 + \epsilon$ and $||\theta_0||_1 - \epsilon$.

Remark 3. Besides our choices for w, one can make use of any number of common kernels with compact supports from the theory of density estimation, such as the triangular or Epanechnikov kernels (cf. Scott 2015, Ch. 6).

Remark 4. We have used the example of normal distributions for the measures of X_i, Y_i ($i \in \mathbb{N}$), and θ, in all of the examples above out of convenience. Of course, the same illustrations can be made if we replace all uses of the normal law $\mathrm{N}(\mu, \sigma^2)$, with density $\phi\left(\cdot; \mu, \sigma^2\right)$, by any generic location-scale law defined by density $\psi\left(\cdot; \mu, \sigma\right)$ with location and scale parameters $\mu \in \mathbb{R}$ and $\sigma > 0$, provided that the necessary integrals with respect to ψ exist.

5. NUMERICAL ESTIMATION OF POSTERIOR MEASURES

We have opted for examples with summary-based discrepancy functions D and for IID sequences $\mathbf{X}_\infty$ and $\mathbf{Y}_\infty$, for ease of understanding and simplicity of exposition. However, the

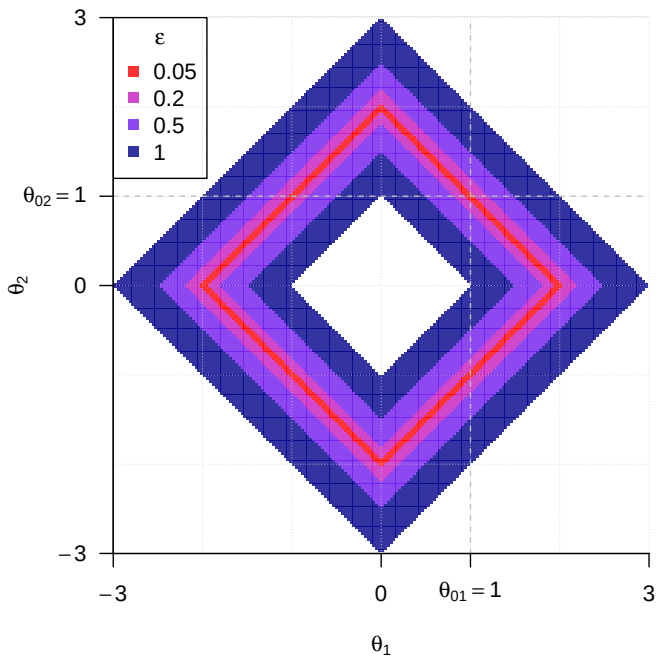

Fig. 3. Supports of limiting posterior measures from Section 4.3 for various threshold values ϵ, when $\theta_0 = (1, 1)$ and $||\theta_0||_1 = 2$.

main conclusions of Theorems 1 and 2 apply in much broader settings, as we will demonstrate below.

5.1. *Non-IID sequences from a first-order autoregressive model*

In this example, we consider $\mathbf{X}_\infty = (X_i)_{i\in\mathbb{N}}$, where X_i can be characterized as

$$X_i = |\theta_0|\, X_{i-1} + E_i,$$

where $\theta_0 \in (-1, 1) = \mathbb{T}$, $X_0 \sim \mathrm{N}\left(0, 1/\left\{1 - |\theta_0|^2\right\}\right)$, and $(E_i)_{i\in\mathbb{N}}$ is IID with $E_i \sim \mathrm{N}(0, 1)$. Similarly, we write $\mathbf{Y}_\infty = (Y_i)_{i\in\mathbb{N}}$, characterized by

$$Y_i = |\theta|\, Y_{i-1} + E'_i,$$

with $\theta \in \mathbb{T}$, where $Y_0 \sim \mathrm{N}\left(0, 1/\left\{1 - |\theta|^2\right\}\right)$ and $(E'_i)_{i\in\mathbb{N}}$ is IID with $E'_i \sim \mathrm{N}(0, 1)$. We shall use a uniform prior density $\pi(\theta) = [\![|\theta| < 1]\!]/2$. Via Hall & Heyde (1980, Thm. 6.6), we have the following strongly consistent estimators of $|\theta_0|$ and $|\theta|$ from $\mathbf{X}_n$ and $\mathbf{Y}_n$: $r_{0,n} = \sum_{i=2}^n X_i X_{i-1} / \sum_{i=2}^n X_i^2$ and $r_n = \sum_{i=2}^n Y_i Y_{i-1} / \sum_{i=2}^n Y_i^2$, respectively. Thus, the distance $D(\mathbf{X}_n, \mathbf{Y}_n) = |r_{0,n} - r_n|$ converges almost surely to $D_\infty(\theta_0, \theta) = ||\theta_0| - |\theta||$.

To estimate the ABC posterior measures, we sample $(\theta_k)_{k\in[m]}$ and compute $(D_k)_{k\in[m]}$ as described in Algorithm 1. For any given $\epsilon > 0$, we can then characterize the ABC posterior measures via their distributions functions, estimated by their respective empirical distribution functions. The larger the value of m, the more accurate the empirical distribution is to its target.

When we take w to be the accept/reject kernel, the empirical distribution function is given by:

$$F_{m,\epsilon}^{\mathrm{AR}}(\theta) = \left\{ \sum_{k=1}^{m} [\![D_k < \epsilon]\!] \right\}^{-1} \sum_{k=1}^{m} [\![D_k < \epsilon, \theta_k < \theta]\!] . \tag{9}$$

Similarly, when taking w to be the triweight kernel, we can estimate distribution functions that characterize the ABC posterior measures, for any given ϵ, via the empirical weighted distribution function:

$$F_{m,\epsilon}^{\mathrm{tri}}(\theta) = \left\{ \sum_{k=1}^{m} \left\{ 1 - \left(\frac{D_k}{\epsilon} \right)^2 \right\}^3 [\![D_k < \epsilon]\!] \right\}^{-1} \sum_{k=1}^{m} \left\{ 1 - \left(\frac{D_k}{\epsilon} \right)^2 \right\}^3 [\![D_k < \epsilon, \theta_k < \theta]\!] . \tag{10}$$

Figure 4 displays sample functions (9) and (10) from an experiment with $\theta_0 = 1/2$, $n \in \{100, 1000\}$, and $m = 10000$. We observe that the empirical distribution functions both concentrate around the estimates $\{-r_{0,n}, r_{0,n}\}$, which converge towards the zeroes of $D_\infty(\theta_0, \theta)$: $\Theta_0 = \{-\theta_0, \theta_0\} = \{-1/2, 1/2\}$. The support of the sample measures are supersets of the estimates $\{-r_{0,n}, r_{0,n}\}$ that decrease in size as ϵ decreases towards zero, as expected. We also observe that, for fixed ϵ the empirical distribution functions both converge to their limiting forms, as n increases, as predicted by Corollary 1.

Algorithm 1 Monte Carlo sampling for estimation of the pseudo-/coarsened-posterior measures.

Input: Data $\mathbf{X}_n$, discrepancy function D, number of Monte Carlo replications $m \geq 1$.
For $k \in [m]$:
  Sample θ_k from a measure with density $\pi(\theta)$;
  Generate $\mathbf{Y}_{n,k}$ from a measure with density $f(\mathbf{y}_n | \theta_k)$;
  Compute discrepancy $D_k = D(\mathbf{X}_n, \mathbf{Y}_{n,k})$.
Output: Discrepancies $\mathbf{D}_m = (D_k)_{k \in [m]}$; Parameters $\mathbf{\Theta}_m = (\theta_k)_{k \in [m]}$.

5.2. *Wasserstein distance*

For this example, we take $\mathbf{X}_\infty$ to be an IID sequence of random variables, where $X_i \sim \mathrm{N}\left(||\theta_0||_2^2, 1\right)$, for each $i \in \mathbb{N}$. We let $\theta_0 = (\theta_{01}, \theta_{02}) \in \mathbb{T} = [-2, 2]^2$ and endow $\mathbb{T}$ with the prior measure defined by the uniform density $\pi(\theta) = \left[\!\!\left[(\theta_1, \theta_2) \in [-2, 2]^2 \right]\!\!\right] / 16$. We then take $\mathbf{Y}_\infty$ to be an IID sequence, independent of $\mathbf{X}_\infty$, such that $Y_i \sim \mathrm{N}\left(||\theta||_2^2, 1\right)$, for each $i \in \mathbb{N}$.

To measure the distance between partial sequences $\mathbf{X}_n$ and $\mathbf{Y}_n$, we use the sample 1-Wasserstein distance

$$D(\mathbf{X}_n, \mathbf{Y}_n) = n^{-1} \sum_{i=1}^{n} \left| X_{(i)} - Y_{(i)} \right| ,$$

where $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ and $Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(n)}$ are the order statistics of $\mathbf{X}_n$ and $\mathbf{Y}_n$, respectively (cf. Peyré & Cuturi 2019, Rem. 2.28). In this case, the 1-Wasserstein distance between measures on $\mathbb{R}$ is just the L_1 distance between the distribution functions of X_1 and Y_1, F_{θ_0} and F_θ (say) (Peyré & Cuturi, 2019, Rem. 2.30). The Glivenko–Cantelli theorem implies that $D(\mathbf{X}_n, \mathbf{Y}_n)$ converges almost surely to $D_\infty(\theta_0, \theta) = \int_{-\infty}^{\infty} |F_{\theta_0}(x) - F_\theta(x)| \, \mathrm{d}x$. To verify B2, we use the fact that the 1-Wasserstein distance between normal distributions $\mathrm{N}(\mu_1, \sigma^2)$ and $\mathrm{N}(\mu_2, \sigma^2)$ is $|\mu_1 - \mu_2|$, for $\mu_1, \mu_2 \in \mathbb{R}$ (cf. Chafai & Malrieu, 2010, Example. 2.5).

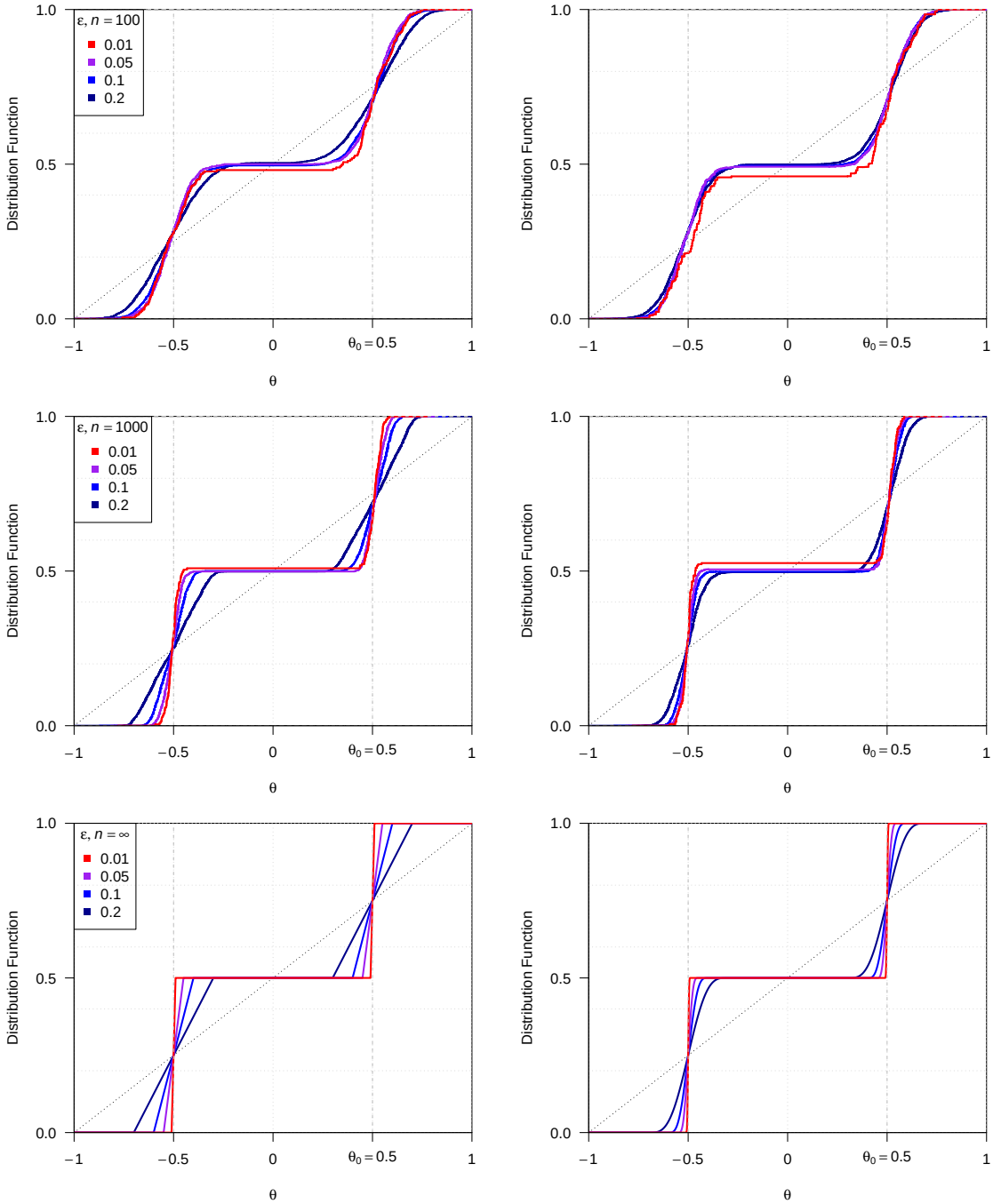

Fig. 4. Empirical distribution function $F^{\mathrm{AR}}_{m,\epsilon}$ (9) (left) and weighted empirical distribution function $F^{\mathrm{tri}}_{m,\epsilon}$ (10) (right) for the experiment described in Section 5.1, with $\theta_0 = 1/2$, $m = 10000$, and sample sizes equal to $n = 100$ (top row), $n = 1000$ (middle row), and $n = \infty$ (bottom row; limiting measures). Dotted diagonal lines represent the distribution function of the prior measure.

We again use Algorithm 1 to obtain estimates of the ABC posteriors using the accept/reject kernel and the triweight kernel in place of w. Simulations are carried out with $\theta_{01} = \theta_{02} = 1/\sqrt{2}$, $n \in \{100, 1000\}$, and $m = 10000$. Interpreting the expression $\theta_k < \theta$ as meaning $\theta_{k1} < \theta_1$ and $\theta_{k2} < \theta_2$ (for $\theta_k, \theta \in \mathbb{R}^2$), we provide representations of the obtained estimated empirical distribution functions of form (9) and (10) via their weighted point masses in Figure 5, along with an illustration of the supports of the limiting ABC posterior measures for various values of ϵ. We observe, as predicted, that the estimated ABC posterior measures concentrate on supersets of the zeroes of D_∞, which can be characterized by the contour of the L_2-ball of radius $\|\theta_0\|_2 = 1$: $\Theta_0 = \left\{\theta = (\theta_1, \theta_2) \in \mathbb{R}^2 : \|\theta\|_2 = \|\theta_0\|_2\right\}$, where the sets get smaller in volume as ϵ decreases, and where smaller weights $w\left(D_k, \epsilon\right)$ are assigned for larger deviations from Θ_0. As expected from Corollary 1, we observe that the supports of the empirical representations converge to those of the limiting measures, as n increases, for fixed values of ϵ.

To complement these numerical results, we provide an additional example in the Appendix, motivated by a sound source localization application from Forbes et al. (2022). Further, we note that when $\mathbf{X}_\infty$ and $\mathbf{Y}_\infty$ arise from general location-scale distributions instead of normal distributions, we can use the formula of Gelbrich (1990, Cor. 2.4), together with the Wasserstein norm equivalence result of Garling (2018, Cor. 21.2.4), in order to verify B2.

6. DISCUSSION

Investigations regarding the asymptotics of ABC algorithms have seen ongoing progress and our work is only one piece of a larger tapestry of findings that provide theoretical results and guarantees for such techniques. As such, we find that a discussion of the placement of our work in proximity to other treatments is useful and imperative. Firstly, we note that our work follows the progress of Jiang (2018), Miller & Dunson (2019), and Nguyen et al. (2020), who focus on the use of limit theorems to derive asymptotics regarding the ABC posteriors $\bar{\Pi}_{n,\epsilon}$ and $\Pi_{n,\epsilon}$. This differs from the approaches of Frazier et al. (2018), Bernton et al. (2019), Frazier et al. (2020), and Legramanti et al. (2022), who rely on concentration of probability inequalities in order to provide rates at which the posterior measure *converges in probability*. However, although we do not provide in-probability rates, our results establish the stronger modes of L_1 and almost sure convergence of posterior objects instead.

It is also notable that Theorem 2 only provides nearly posterior consistency results, rather than true posterior consistency, as discussed in Section 3, which would require a result that guarantees that $\bar{\Pi}_{n,\epsilon}$ or $\Pi_{n,\epsilon}$ converges to a point mass, in some mode of convergence, when ϵ is taken to be a decreasing function of n, as $n \to \infty$. Such a result is in fact guaranteed by Frazier et al. (2018, Thm. 1), Bernton et al. (2019, Cor. 1), and Legramanti et al. (2022, Cor. 1), in probability. We note that both Frazier et al. (2018, Thm. 1), Bernton et al. (2019, Cor. 1), and Legramanti et al. (2022, Cor. 1) make Hölder continuity assumptions that are stronger than our uniform continuity assumption when interpreting B2 in the identifiable case, and obviously much stronger than our assumptions in non-identifiable cases. In particular, in comparison to Frazier et al. (2018, Thm. 1), we do not require that D be a distance between summary statistics. This is also true when comparing with the results of Li & Fearnhead (2018b,a).

Next, unlike Frazier et al. (2018), Bernton et al. (2019), Frazier et al. (2020), and Legramanti et al. (2022), we provide asymptotic results for weights w other than the accept/reject kernel. This is a feature that is shared with Li & Fearnhead (2018b,a). However, our structural assumptions A3 and B4 are fairly mild in comparison to the location-scale form requirements and the existence of higher order moments of Li & Fearnhead (2018b,a). However, the additional restrictions together with stronger assumptions regarding the concentration of measure and weak convergence of

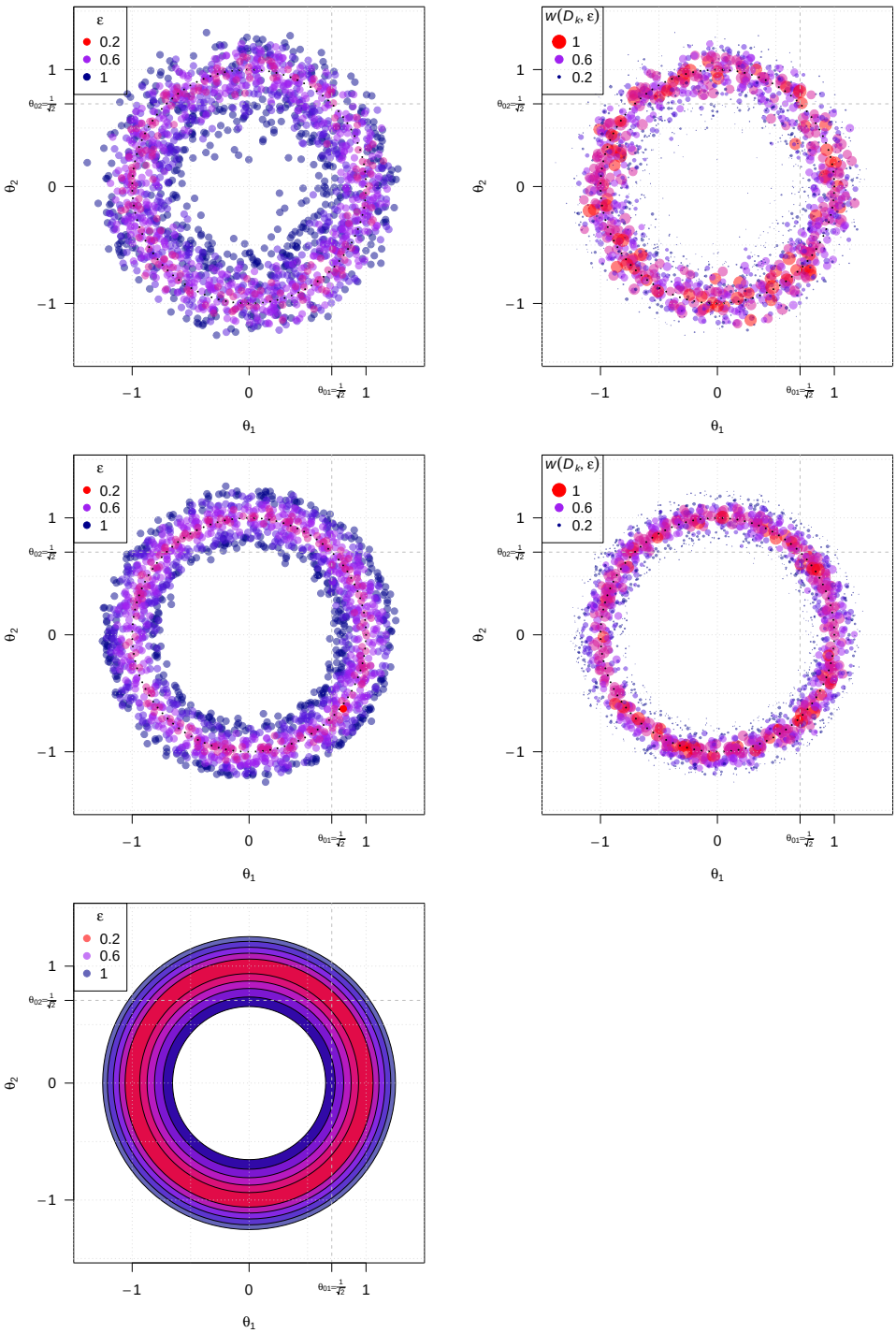

Fig. 5. Point mass representations of empirical distribution function $F^{\mathrm{AR}}_{m,\epsilon}$ (9) (left) and weighted empirical distribution function $F^{\mathrm{tri}}_{m,\epsilon}$ (10) (right) for the experiment described in Section 5.2, with $\theta_0 = \left(1/\sqrt{2}, 1/\sqrt{2}\right)$, $m = 10000$, sample size values: $n = 100$ (top row), $n = 1000$ (middle row), and $n = \infty$ (bottom row; support of limiting measures). Dotted lines represent the set Θ_0. Left: point masses for different threshold values $\epsilon > 0$. Right: point masses for $\epsilon = 1$, with the sizes of points indicating the weights $w\left(D_k, \epsilon\right)$ (as a fraction of the maximum observed weight).

summaries of $\mathbf{X}_n$ and $\mathbf{Y}_n$, and concentration of the prior measure Π permits Frazier et al. (2018); Li & Fearnhead (2018b,a), and Frazier et al. (2020) to establish weak and strong convergence of $\Pi_{n,\epsilon}$ to measures with explicit forms.

In summary, we have provided a set of results that permit the establishment of nearly posterior consistency of ABC posterior measures in the L_1 and almost sure sense, in scenarios that lack identifiability and when discrepancies are possibly not Hölder continuous, and when the sequences $\mathbf{X}_\infty$ and $\mathbf{Y}_\infty$ are non-IID. Our results provide valuable complements to the existing literature that deliver theoretical guarantees to situations that are not covered by previous works.

REFERENCES

BEAUMONT, M. A. (2019). Approximate Bayesian computation. *Annual Review of Statistics and its Application* **6**, 379–403.

BERNTON, E., JACOB, P. E., GERBER, M. & ROBERT, C. P. (2019). Approximate Bayesian computation with the Wasserstein distance. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **81**, 235–269.

BERTI, P., PRATELLI, L. & RIGO, P. (2006). Almost sure weak convergence of random probability measures. *Stochastics and Stochastics Reports* **78**, 91–97.

CHAFAI, D. & MALRIEU, F. (2010). On fine properties of mixtures with respect to concentration of measure and Sobolev type inequalities. *Annales de l'IHP Probabilités et Statistiques* **46**, 72–96.

FORBES, F., NGUYEN, H. D., NGUYEN, T. & ARBEL, J. (2022). Summary statistics and discrepancy measures for approximate Bayesian computation via surrogate posteriors. *Statistics and Computing* **32**.

FRAZIER, D. T., MARTIN, G. M., ROBERT, C. P. & ROUSSEAU, J. (2018). Asymptotic properties of approximate Bayesian computation. *Biometrika* **105**, 593–607.

FRAZIER, D. T., ROBERT, C. P. & ROUSSEAU, J. (2020). Model misspecification in approximate Bayesian computation: consequences and diagnostics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **82**, 421–444.

GARLING, D. J. (2018). *Analysis on Polish spaces and an introduction to optimal transportation*. Cambridge University Press.

GELBRICH, M. (1990). On a formula for the L2 Wasserstein metric between measures on Euclidean and Hilbert spaces. *Mathematische Nachrichten* **147**, 185–203.

GHOSAL, S. & VAN DER VAART, A. (2017). *Fundamentals of nonparametric Bayesian inference*, vol. 44. Cambridge University Press.

GRÜBEL, R. & KABLUCHKO, Z. (2016). A functional central limit theorem for branching random walks, almost sure weak convergence and applications to random trees. *The Annals of Applied Probability* **26**, 3659–3698.

HALL, P. & HEYDE, C. C. (1980). *Martingale limit theory and its application*. Academic Press.

JIANG, B. (2018). Approximate Bayesian computation with Kullback-Leibler divergence as data discrepancy. In *International Conference on Artificial Intelligence and Statistics*. PMLR.

KARABATSOS, G. & LEISEN, F. (2018). An approximate likelihood perspective on ABC methods. *Statistics Surveys* **12**, 66–104.

LEGRAMANTI, S., DURANTE, D. & ALQUIER, P. (2022). Concentration and robustness of discrepancy-based ABC via Rademacher complexity. *arXiv preprint arXiv:2206.06991* .

LI, W. & FEARNHEAD, P. (2018a). Convergence of regression-adjusted approximate Bayesian computation. *Biometrika* **105**, 301–318.

LI, W. & FEARNHEAD, P. (2018b). On the asymptotic efficiency of approximate Bayesian computation estimators. *Biometrika* **105**, 285–299.

MARIN, J.-M., PUDLO, P., ROBERT, C. P. & RYDER, R. J. (2012). Approximate Bayesian computational methods. *Statistics and Computing* **22**, 1167–1180.

MILLER, J. W. & DUNSON, D. B. (2019). Robust Bayesian inference via coarsening. *Journal of the American Statistical Association* **114**, 1113–1125.

NGUYEN, H. D., ARBEL, J., LÜ, H. & FORBES, F. (2020). Approximate Bayesian computation via the energy statistic. *IEEE Access* **8**, 131683–131698.

PARK, M., JITKRITTUM, W. & SEJDINOVIC, D. (2016). K2-ABC: Approximate Bayesian computation with kernel embeddings. In *Artificial Intelligence and Statistics*. PMLR.

PEYRÉ, G. & CUTURI, M. (2019). Computational optimal transport: with applications to data science. *Foundations and Trends in Machine Learning* **11**, 355–607.

RICHMOND, T. (2020). *General Topology*. An Introduction. De Gruyter.

SCOTT, D. W. (2015). *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons.

SISSON, S. A., FAN, Y. & BEAUMONT, M. (2018). *Handbook of approximate Bayesian computation*. CRC Press.

Supplementary Materials for "Concentration results for approximate Bayesian computation without identifiability"

BY H. D. NGUYEN

School of Mathematics and Physics, Univ. of Queensland St. Lucia 4067, Queensland Australia.

h.nguyen7@uq.edu.au

T. T. NGUYEN, J. ARBEL, AND F. FORBES

Univ. Grenoble Alpes, Inria, CNRS
Grenoble INP, LJK 38000 Grenoble, France.

trung-tin.nguyen@inria.fr julyan.arbel@inria.fr florence.forbes@inria.fr

SUMMARY

This supplement contains an Appendix comprising proofs of the main results from the primary document. We also include external technical results required within the proofs. An additional numerical illustration is also reported in the final section.

APPENDIX

Technical results

LEMMA 1. *Let $(\mathcal{F}_n)_{n\in\mathbb{N}}$ be a filtration of $\mathcal{F}$, and let $\mathcal{F}_\infty = \sigma\left(\bigcup_{n\in\mathbb{N}} \mathcal{F}_n\right)$. Let $(U_n)_{n\in\mathbb{N}}$ be a sequence of random variables, such that $U_n \xrightarrow[n\to\infty]{a.s.} U$, where $|U_n| \leq V$, for each $n \in \mathbb{N}$, such that $E(V) < \infty$. Then, $\lim_{n\to\infty} \mathrm{E}\left(U_n|\mathcal{F}_n\right) = \mathrm{E}\left(U|\mathcal{F}_\infty\right)$, almost surely and in L_1.*

Remark 1. Lemma 1 is often called Hunt's Lemma and can be found in Dellacherie & Meyer (1980) and Spataru (2013, Thm. 29.32).

Proofs of main results

Proof of Theorem 1

We prove this result via Lemma 1. Start by setting

$$U_n = w\left(D\left(\mathbf{X}_n, \mathbf{Y}_n\right), \epsilon\right),$$

where both $\mathbf{X}_n$ and $\mathbf{Y}_n$ are functions of $\omega \in \Omega$. By A2, we can set the (deterministic) constant $v = \sup_\delta w\left(\delta, \epsilon\right) < \infty$, so that $|U_n| \leq v < \infty$. By A1

$$D\left(\mathbf{X}_n, \mathbf{Y}_n\right) \xrightarrow[n\to\infty]{\text{a.s.}} D_\infty\left(\theta_0, \theta\right)$$

and $w\left(\cdot, \epsilon\right)$ is almost surely continuous at $D_\infty\left(\theta_0, \theta\right)$ by A3, so by continuous mapping:

$$U_n \to U = w\left(D_\infty\left(\theta_0, \theta\right), \epsilon\right).$$

Now set $\mathcal{F}_n = \sigma\left(\mathbf{X}_n\right)$, which approaches $\mathcal{F}_\infty = \bigcup_{n\in\mathbb{N}} \mathcal{F}_n$. Lemma 1 applied directly implies that

$$\mathrm{E}\left\{w\left(D\left(\mathbf{X}_n, \mathbf{Y}_n\right), \epsilon\right)|\mathcal{F}_n\right\} \to \mathrm{E}\left\{w\left(D_\infty\left(\theta_0, \theta\right), \epsilon\right)|\mathcal{F}_\infty\right\},$$

almost surely and in L_1, as $n \to \infty$. Note that the left-hand side reduces to

$$\begin{aligned}\mathrm{E}\left\{w\left(D\left(\mathbf{X}_n, \mathbf{Y}_n\right), \epsilon\right) | \mathcal{F}_n\right\} &= \mathrm{E}\left\{w\left(D\left(\mathbf{X}_n, \mathbf{Y}_n\right), \epsilon\right) | X_1, \ldots, X_n\right\} \\ &= \int_{\mathbb{X}^n} w\left(D\left(\mathbf{X}_n, \mathbf{y}_n\right), \epsilon\right) \mathrm{p}\left(\mathbf{y}_n | \theta\right) \mathrm{d}\mathfrak{m}(\mathbf{y}_n) \\ &= \mathcal{L}_\epsilon\left(\mathbf{X}_n | \theta\right)\end{aligned}$$

and that the right-hand side reduces to a constant, we have

$$\mathcal{L}_\epsilon\left(\mathbf{X}_n | \theta\right) \underset{n\to\infty}{\overset{\text{a.s., } \mathrm{L}_1}{\longrightarrow}} w\left(D_\infty\left(\theta_0, \theta\right), \epsilon\right). \tag{1}$$

To prove (i), we note that A3 implies that (1) holds for almost every $\theta \in \mathbb{T}$. The L_1 convergence then implies that

$$\mathrm{E}\left\{\mathcal{L}_\epsilon\left(\mathbf{X}_n | \theta\right)\right\} \underset{n\to\infty}{\overset{\text{a.e.}}{\longrightarrow}} w\left(D_\infty\left(\theta_0, \theta\right), \epsilon\right) \text{ and } \pi\left(\theta\right) \mathrm{E}\left\{\mathcal{L}_\epsilon\left(\mathbf{X}_n | \theta\right)\right\} \underset{n\to\infty}{\overset{\text{a.e.}}{\longrightarrow}} \pi\left(\theta\right) w\left(D_\infty\left(\theta_0, \theta\right), \epsilon\right).$$

Applying the dominated convergence theorem, upon noting that $\mathrm{E}\left\{\mathcal{L}_\epsilon\left(\mathbf{X}_n | \theta\right)\right\} \leq \sup_\delta w\left(\delta, \epsilon\right)$ and that $\pi\left(\theta\right)$ is a density function, we then obtain

$$\bar{\mathcal{C}}_{\epsilon,n} \underset{n\to\infty}{\longrightarrow} \int_{\mathbb{T}} \pi\left(\theta\right) w\left(D_\infty\left(\theta_0, \theta\right), \epsilon\right) \mathrm{d}\mathfrak{n}\left(\theta\right),$$

which completes the proof.

To prove (ii), we again use A3 and (1) to now obtain the fact that for almost every $\theta \in \mathbb{T}$

$$\pi\left(\theta\right) \mathcal{L}_\epsilon\left(\mathbf{X}_n | \theta\right) \underset{n\to\infty}{\longrightarrow} \pi\left(\theta\right) w\left(D_\infty\left(\theta_0, \theta\right), \epsilon\right), \tag{2}$$

almost surely. We now apply Lemma 1 in an alternative manner. That is, using the same filtration $\left(\mathcal{F}_n\right)_{n\in\mathbb{N}}$, we write the normalization term $\mathcal{C}_\epsilon\left(\mathbf{X}_n\right)$ as the conditional expectation

$$\mathrm{E}_{(\omega,\theta)}\left\{w\left(D\left(\mathbf{X}_n, \mathbf{Y}_n\right), \epsilon\right) | \mathcal{F}_n\right\} = \int_{\mathbb{T}} \int_{\mathbb{X}^n} w\left(D\left(\mathbf{X}_n, \mathbf{y}_n\right), \epsilon\right) \mathrm{p}\left(\mathbf{y}_n | \theta\right) \pi\left(\theta\right) \mathrm{d}\mathfrak{m}\left(\mathbf{y}_n\right) \mathrm{d}\mathfrak{n}\left(\theta\right)$$

with respect to the product measure defined by the joint density $\mathrm{p}\left(\mathbf{y}_n | \theta\right) \pi\left(\theta\right)$. By the same argument as used to obtain (1), Lemma 1 implies that

$$\mathrm{E}_{(\omega,\theta)}\left\{w\left(D\left(\mathbf{X}_n, \mathbf{Y}_n\right), \epsilon\right) | \mathcal{F}_n\right\} \to \mathrm{E}_{(\omega,\theta)}\left\{w\left(D\left(\theta_0, \theta\right), \epsilon\right) | \mathcal{F}_\infty\right\}$$

almost surely and in L_1, as $n \to \infty$. By Tonelli's theorem and by definition of $\mathrm{E}_{(\omega,\theta)}$, we have

$$\begin{aligned}\mathrm{E}_{(\omega,\theta)}\left\{w\left(D\left(\theta_0, \theta\right), \epsilon\right) | \mathcal{F}_\infty\right\} &= \int_\Omega \int_{\mathbb{T}} w\left(D\left(\theta_0, \theta\right), \epsilon\right) \pi\left(\theta\right) \mathrm{d}\mathfrak{n}\left(\theta\right) \mathrm{dP}\left(\omega\right) \\ &= \int_{\mathbb{T}} \pi\left(\theta\right) w\left(D\left(\theta_0, \theta\right), \epsilon\right) \mathrm{d}\mathfrak{n}\left(\theta\right)\end{aligned}$$

since $w\left(D\left(\theta_0, \theta\right), \epsilon\right)$ is bounded and does not depend on ω. Thus,

$$\mathcal{C}_\epsilon\left(\mathbf{X}_n\right) \underset{n\to\infty}{\overset{\text{a.s.}}{\longrightarrow}} \int_{\mathbb{T}} \pi\left(\theta\right) w\left(D\left(\theta_0, \theta\right), \epsilon\right) \mathrm{d}\mathfrak{n}\left(\theta\right). \tag{3}$$

We then obtain the desired result via an application of Slutsky's theorem for ratios.

Proof of Corollary 1

To prove Part (i), it is sufficient to check that

$$\int_{\mathbb{T}} f(\theta)\,\pi(\theta)\,\mathrm{E}\left\{\mathcal{L}_\epsilon\left(\mathbf{X}_n|\theta\right)\right\}\mathrm{d}\mathfrak{n}(\theta) \underset{n\to\infty}{\longrightarrow} \int_{\mathbb{T}} f(\theta)\,\pi(\theta)\,w\left(D_\infty\left(\theta_0,\theta\right),\epsilon\right)\mathrm{d}\mathfrak{n}(\theta),$$

for each bounded continuous function $f:\mathbb{T}\to\mathbb{R}$, which follows from the proof of Theorem 1 and an application of the dominated convergence theorem. Similarly, to prove Part (ii), it suffices to show that

$$\int_{\mathbb{T}} f(\theta)\,\pi(\theta)\,\mathcal{L}_\epsilon\left(\mathbf{X}_n|\theta\right)\mathrm{d}\mathfrak{n}(\theta) \underset{n\to\infty}{\overset{\text{a.s.}}{\longrightarrow}} \int_{\mathbb{T}} f(\theta)\,\pi(\theta)\,w\left(D_\infty\left(\theta_0,\theta\right),\epsilon\right)\mathrm{d}\mathfrak{n}(\theta),$$

which follows via the same argument used to obtain (3), with the additional fact that $f(\theta)$ is bounded and thus has finite expectation with respect to the product measure $\mathrm{P}\times\Pi$.

Proof of Theorem 2

We begin our proof of Part (i) by defining the *Minkowski sum* of sets $\mathbb{A},\mathbb{B}\subset\mathbb{T}$ as $\mathbb{A}\oplus\mathbb{B}=\{a+b: a\in\mathbb{A}, b\in\mathbb{B}\}$. By B3, we have λ-covering centres $\Theta_\lambda\subset\mathbb{T}$ of Θ_0 (as defined by B1 and B2). This then implies that $\Theta_0\subset\Theta_\lambda\oplus\lambda\mathbb{B}_0$. By the monotonicity of Minkoswki addition, $\Theta_0\oplus\lambda\mathbb{B}_0\subset\Theta_\lambda\oplus\lambda\mathbb{B}_0\oplus\lambda\mathbb{B}_0=\Theta_\lambda\oplus 2\lambda\mathbb{B}_0$, where we use the fact that if $\mathbb{A}$ is convex, then $\alpha\mathbb{A}+\beta\mathbb{A}=(\alpha+\beta)\,\mathbb{A}$ (cf. Schneider 2013, Ch. 3). Now, using the fact that the Lebesgue measure of $\lambda\mathbb{B}_0$ is $\pi^{q/2}\lambda^q/\Gamma(q/2+1)$ and by monotonicity, we obtain the fact that

$$\mathrm{Leb}\left(\Theta_\lambda\oplus\lambda\mathbb{B}_0\right)\le\mathrm{Leb}\left(\Theta_\lambda\oplus 2\lambda\mathbb{B}_0\right)=\left|\Theta_\lambda\right|\frac{\pi^{q/2}\left(2\lambda\right)^q}{\Gamma\left(q/2+1\right)}\underset{\lambda\to 0}{\longrightarrow}0,$$

since $\left|\Theta_\lambda\right|\lambda^q\underset{\lambda\to 0}{\longrightarrow}0$.

To prove Parts (ii) and (iii), notice that for each $\epsilon>0$, $w\left(D_\infty\left(\theta_0,\theta\right),\epsilon\right)$ has support on $\left\{\theta: D_\infty\left(\theta_0,\theta\right)<\varphi^{-1}\left(\epsilon\right)\right\}$, by B4. But by B1 and B2, for each $\lambda>0$, there exists a $\epsilon>0$, such that if $D_\infty\left(\theta_0,\theta\right)<\varphi^{-1}\left(\epsilon\right)$, then $\min_{\tau\in\Theta_0}\|\theta-\tau\|<\lambda$, and thus

$$\left\{\theta: D_\infty\left(\theta_0,\theta\right)<\varphi^{-1}\left(\epsilon\right)\right\}\subset\Theta_0\oplus\lambda\mathbb{B}_0,$$

by Part (i), and

$$\begin{aligned}\int_{(\Theta_0\oplus\lambda\mathbb{B}_0)\cap\mathbb{T}}\pi(\theta)\,w\left(D_\infty\left(\theta_0,\theta\right),\epsilon\right)\mathrm{d}\mathfrak{n}(\theta)&\ge\int_{\{\varphi(D_\infty(\theta_0,\theta))<\epsilon\}\cap\mathbb{T}}\pi(\theta)\,w\left(D_\infty\left(\theta_0,\theta\right),\epsilon\right)\mathrm{d}\mathfrak{n}(\theta)\\&=\int_{\mathbb{T}}\pi(\theta)\,w\left(D_\infty\left(\theta_0,\theta\right),\epsilon\right)\mathrm{d}\mathfrak{n}(\theta),\end{aligned}$$

where the equality holds by the fact that $w\left(D_\infty\left(\theta_0,\theta\right),\epsilon\right)$ is zero when $\varphi\left(D_\infty\left(\theta_0,\theta\right)\right)\ge\epsilon$. Thus,

$$\frac{\int_{(\Theta_0\oplus\lambda\mathbb{B}_0)\cap\mathbb{T}}\pi(\theta)\,w\left(D_\infty\left(\theta_0,\theta\right),\epsilon\right)\mathrm{d}\mathfrak{n}(\theta)}{\int_{\mathbb{T}}\pi(\theta)\,w\left(D_\infty\left(\theta_0,\theta\right),\epsilon\right)\mathrm{d}\mathfrak{n}(\theta)}=1$$

since $(\Theta_0 \oplus \lambda\mathbb{B}_0) \cap \mathbb{T} \subset \mathbb{T}$. Next, under A1–A3 we follow the proof of Corollary 1 Part (i), and conclude via the dominated convergence theorem that

$$\begin{aligned}\int_{(\Theta_0\oplus\lambda\mathbb{B}_0)\cap\mathbb{T}} \pi(\theta)\,\mathrm{E}\left\{\mathcal{L}_\epsilon\left(\mathbf{X}_n|\theta\right)\right\}\mathrm{d}\mathfrak{n}(\theta) &= \int_{\mathbb{T}} [\![\theta\in\Theta_0\oplus\lambda\mathbb{B}_0]\!]\,\pi(\theta)\,\mathrm{E}\left\{\mathcal{L}_\epsilon\left(\mathbf{X}_n|\theta\right)\right\}\mathrm{d}\mathfrak{n}(\theta)\\ &\underset{n\to\infty}{\longrightarrow} \int_{\mathbb{T}} [\![\theta\in\Theta_0\oplus\lambda\mathbb{B}_0]\!]\,\pi(\theta)\,w\left(D_\infty\left(\theta_0,\theta\right),\epsilon\right)\mathrm{d}\mathfrak{n}(\theta)\\ &= \int_{(\Theta_0\oplus\lambda\mathbb{B}_0)\cap\mathbb{T}} \pi(\theta)\,w\left(D_\infty\left(\theta_0,\theta\right),\epsilon\right)\mathrm{d}\mathfrak{n}(\theta),\end{aligned}$$

as required to show Part (ii). Similarly, Part (iii) follows from A1–A3, via the dominated converge theorem on an almost sure set, to obtain

$$\begin{aligned}\int_{(\Theta_0\oplus\lambda\mathbb{B}_0)\cap\mathbb{T}} \pi(\theta)\,\mathcal{L}_\epsilon\left(\mathbf{X}_n|\theta\right)\mathrm{d}\mathfrak{n}(\theta) &= \int_{\mathbb{T}} [\![\theta\in\Theta_0\oplus\lambda\mathbb{B}_0]\!]\,\pi(\theta)\,\mathcal{L}_\epsilon\left(\mathbf{X}_n|\theta\right)\mathrm{d}\mathfrak{n}(\theta)\\ &\underset{n\to\infty}{\overset{\text{a.s.}}{\longrightarrow}} \int_{\mathbb{T}} [\![\theta\in\Theta_0\oplus\lambda\mathbb{B}_0]\!]\,\pi(\theta)\,w\left(D_\infty\left(\theta_0,\theta\right),\epsilon\right)\mathrm{d}\mathfrak{n}(\theta)\\ &= \int_{(\Theta_0\oplus\lambda\mathbb{B}_0)\cap\mathbb{T}} \pi(\theta)\,w\left(D_\infty\left(\theta_0,\theta\right),\epsilon\right)\mathrm{d}\mathfrak{n}(\theta),\end{aligned}$$

as in Corollary 1 Part (ii).

Binaural sound source localisation

The following example is a simplified version of the synthetic sound source localization example of Forbes et al. (2022). Here, we take $\mathbf{X}_\infty$ to be an IID sequence of random variables, with $X_i \sim \mathrm{N}\left(\mathrm{ITD}(\theta_0), 1/4\right)$, for each $i \in \mathbb{N}$, and $\theta_0 \in \mathbb{R}^2$ is interpreted as the 2D position of a sound source, assumed to be captured only through the noisy measurements of so-called interaural time differences (ITDs). In a binaural setting with two microphones, the ITD is defined as the difference between the time of arrival to the first and second microphone and given by $\mathrm{ITD}(\theta_0) = |\|\theta_0-\mu_1\|_2 - \|\theta_0-\mu_2\|_2|$ for a source in θ_0 and microphones at positions $\mu_1, \mu_2 \in \mathbb{R}^2$. ITD measurements only allow to determine a pair of hyperbolas on which the source maybe. We let $\theta_0 = (\theta_{01}, \theta_{02}) \in \mathbb{T} = [-2,2]^2$ and endow $\mathbb{T}$ with the prior measure defined by the uniform density $\pi(\theta) = \left[\!\!\left[(\theta_1,\theta_2) \in [-2,2]^2\right]\!\!\right]/16$. We then take $\mathbf{Y}_\infty$ to be an IID sequence, independent of $\mathbf{X}_\infty$, such that $Y_i \sim \mathrm{N}\left(\mathrm{ITD}(\theta), 1/4\right)$, for each $i \in \mathbb{N}$.

We proceed in the same manner as in Section 5.2 of the main text. That is, to measure the distance between partial sequences $\mathbf{X}_n$ and $\mathbf{Y}_n$, we use the sample 1-Wasserstein distance

$$D\left(\mathbf{X}_n, \mathbf{Y}_n\right) = n^{-1}\sum_{i=1}^{n}\left|X_{(i)} - Y_{(i)}\right|,$$

where $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ and $Y_{(1)} \le Y_{(2)} \le \cdots \le Y_{(n)}$ are the order statistics of $\mathbf{X}_n$ and $\mathbf{Y}_n$, respectively. Again, the 1-Wasserstein distance between measures on $\mathbb{R}$ is just the L_1 distance between the distribution functions of X_1 and Y_1, F_{θ_0} and F_θ, and we have the convergence of $D\left(\mathbf{X}_n, \mathbf{Y}_n\right)$, almost surely, to $D_\infty\left(\theta_0,\theta\right) = \int_{-\infty}^{\infty}\left|F_{\theta_0}(x) - F_\theta(x)\right|\mathrm{d}x$, with B2 verified again via the form of the 1-Wasserstein distance: $|\mathrm{ITD}(\theta_0) - \mathrm{ITD}(\theta)|$ (cf. Chafai & Malrieu, 2010, Example. 2.5).

We again use Algorithm 1 to obtain estimates of the ABC posteriors using the accept/reject kernel and the triweight kernel in place of w. Simulations are carried out with $\theta_{01} = \theta_{02} = 1$, $n \in \{100, 1000, 10000\}$, and $m = 10000$. We provide representations of the obtained estimated

empirical distribution functions via their weighted point masses in Figure 1, along with visual indicators of the supports of the limiting ABC posterior measures for various values of ϵ. We observe, as predicted, that the estimated ABC posterior measures concentrate on supersets of the zeroes of D_∞ which consists of two hyperbolas as shown in Figure 1. We visualize the boundaries of the support sets for different levels of ϵ via patterned lines on the plots. Observe that smaller values of ϵ correspond to sets with elements that sit closer to the zeroes of D_∞. We also observe that smaller weights $w\left(D_k, \epsilon\right)$ are assigned for larger deviations from the zeroes of D_∞. Furthermore, as expected, we observe that the supports of the empirical representations converge to those of the limiting measures, as n increases, for each fixed value of ϵ.

REFERENCES

CHAFAI, D. & MALRIEU, F. (2010). On fine properties of mixtures with respect to concentration of measure and Sobolev type inequalities. *Annales de l'IHP Probabilités et Statistiques* **46**, 72–96.

DELLACHERIE, C. & MEYER, P.-A. (1980). *Probabilities and Potential, B: Theory of Martingales*. Elsevier.

FORBES, F., NGUYEN, H. D., NGUYEN, T. & ARBEL, J. (2022). Summary statistics and discrepancy measures for approximate Bayesian computation via surrogate posteriors. *Statistics and Computing* **32**.

SCHNEIDER, R. (2013). *Convex Bodies: The BrunnMinkowski Theory*. Encyclopedia of Mathematics and its Applications. Cambridge: Cambridge University Press, 2nd ed.

SPATARU, A. (2013). *Analysis and Probability*. Oxford: Elsevier.

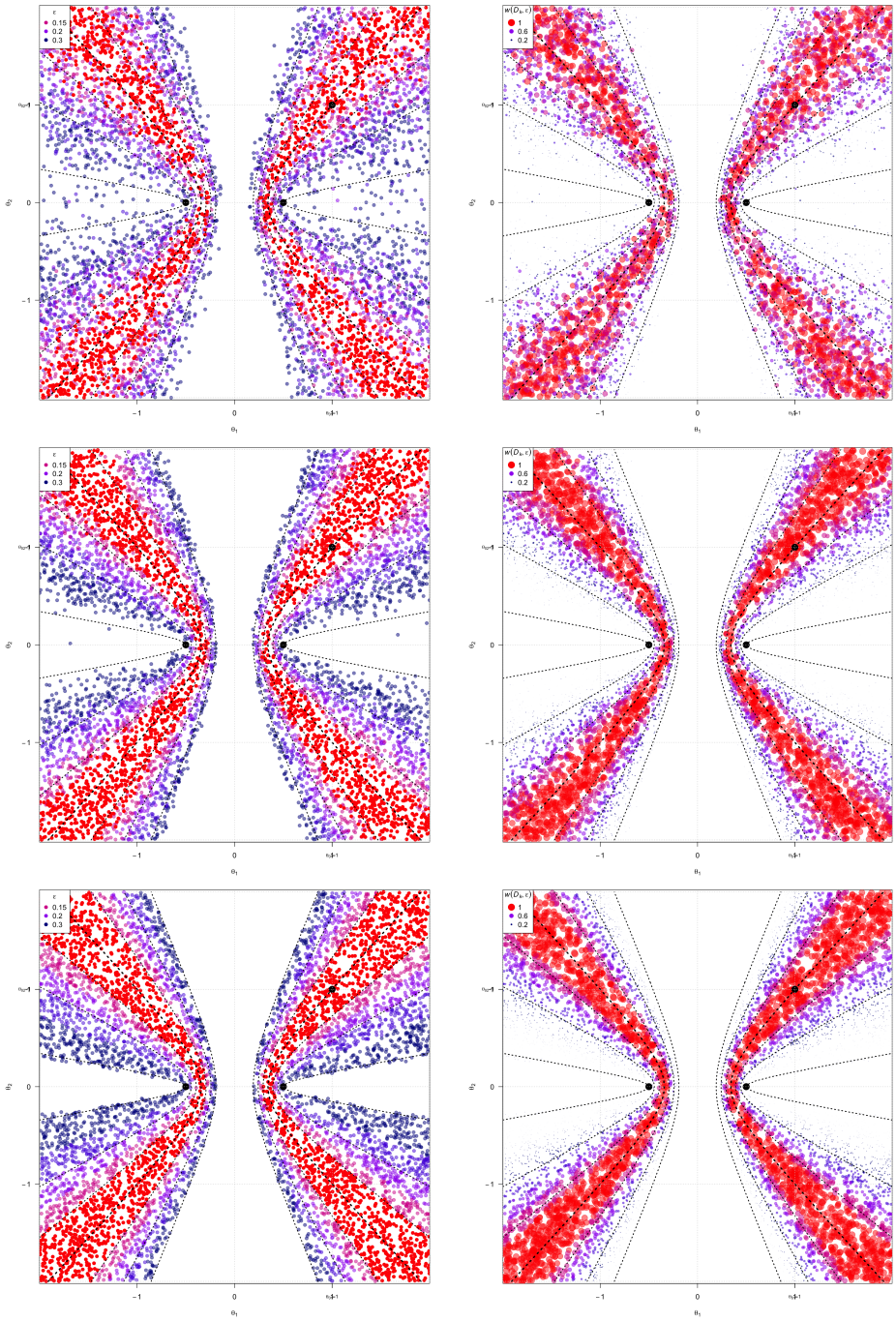

Fig. 1. Point mass representations of empirical distribution function $F^{\mathrm{AR}}_{m,\epsilon}$ (9) (left) and weighted empirical distribution function $F^{\mathrm{tri}}_{m,\epsilon}$ (10) (right) for the experiment described in the section labeled: Binaural sound source localisation, with $\theta_0 = (1, 1)$, $m = 10000$, sample size values: $n = 100$ (top row), $n = 1000$ (middle row), and $n = 10000$ (bottom row). Black dots represent the locations of the two microphones on the horizontal axis (μ_1 and μ_2), and another dot represents the source θ_0 on the thick dotted hyperbola that visualizes the set Θ_0. Additional patterned lines indicate the boundaries of the support of the limiting ABC posteriors for various fixed values of ϵ. Left: point masses for different threshold values $\epsilon > 0$ from 0.15 to 0.3. Right: point masses for $\epsilon = 0.3$, with the sizes of points indicating the weights $w(D_k, \epsilon)$ (as a fraction of the maximum observed weight).