SUPPLEMENTARY MATERIAL TO "CONCENTRATION OF DISCREPANCY-BASED APPROXIMATE BAYESIAN COMPUTATION VIA RADEMACHER COMPLEXITY"

BY SIRIO LEGRAMANTI[1,*] DANIELE DURANTE[2,†], AND PIERRE ALQUIER[3,‡]

[1]Department of Economics, University of Bergamo [*]sirio.legramanti@unibg.it
[2]Department of Decision Sciences and Institute for Data Science and Analytics, Bocconi University, [†]daniele.durante@unibocconi.it
[3]Department of Information Systems, Decision Sciences and Statistics, ESSEC Business School, [‡]alquier@essec.edu

APPENDIX A: ADDITIONAL INTEGRAL PROBABILITY SEMIMETRICS

A.1. Two additional examples of IPS discrepancies. While MMD, Wasserstein-1 and summary-based distances provide the most notable examples of IPS discrepancies employed in ABC, two other relevant IPS instances are the total variation (TV) distance and the Kolmogorov–Smirnov (KS) distance, discussed below.

EXAMPLE A.1. (Total variation distance). *Although the total variation distance is not a common choice within discrepancy-based* ABC, *it still provides a notable example of* IPS, *obtained when $\mathfrak{F}$ is the class of measurable functions whose sup-norm is bounded by* 1; *i.e.* $\mathfrak{F} = \{f : ||f||_\infty \leq 1\}$.

EXAMPLE A.2. (Kolmogorov–Smirnov distance). *When $\mathcal{Y} = \mathbb{R}$ and $\mathfrak{F} = \{\mathbb{1}_{(-\infty,a]}\}_{a\in\mathbb{R}}$, then $\mathcal{D}_{\mathfrak{F}}$ is the Kolmogorov–Smirnov distance, which can also be written as $\mathcal{D}_{\mathfrak{F}}(\mu_1, \mu_2) = \sup_{x\in\mathcal{Y}} |F_1(x) - F_2(x)|$, where F_1 and F_2 are the cumulative distribution functions associated with μ_1 and μ_2, respectively.*

A.2. Validity of (III)–(IV) for the TV distance and Kolmogorov–Smirnov distance. Examples A.3 and A.4 verify the validity of assumptions (III) and (IV) under the TV distance and the Kolmogorov–Smirnov distance, respectively.

EXAMPLE A.3. (Total variation distance). *The* TV *distance satisfies (III) by definition, but in general not Assumption (IV), unless the cardinality $|\mathcal{Y}|$ of $\mathcal{Y}$ is finite. In fact, when $\mathcal{Y} = \mathbb{R}$ and $\mu \in \mathcal{P}(\mathcal{Y})$ is continuous, the probability that there exists an index $i \neq i'$ such that $x_i = x_{i'}$ is zero. Hence, with probability 1, for any vector $\epsilon_{1:n}$ of Rademacher variables there always exists a function f_ϵ from $\mathcal{Y}$ to $\{0; 1\}$ such that $f_\epsilon(x_i) = \mathbb{1}_{\{\epsilon_i=1\}}$. Therefore, $\sup_{f\in\mathfrak{F}} |(1/n)\sum_{i=1}^n \epsilon_i f(x_i)| \geq (1/n)\sum_{i=1}^n \mathbb{1}_{\{\epsilon_i=1\}}$, which implies that the Rademacher complexity $\mathfrak{R}_{\mu,n}(\mathfrak{F})$ is bounded below by $(1/n)\sum_{i=1}^n \mathbb{P}(\epsilon_i = 1) = 1/2$. Nonetheless, as mentioned above, the* TV *distance can still satisfy (IV) in specific contexts. For instance, leveraging the bound in Lemma 5.2 of Massart (2000), when the cardinality $|\mathcal{Y}|$ of $\mathcal{Y}$ is finite, there will be replicates in $[f(x_1), \dots, f(x_n)]$ whenever $n > |\mathcal{Y}|$. Hence, as $n \to \infty$, it will* *be impossible to find a function in $\mathfrak{F}$ which can interpolate any noise vector of Rademacher variables with $[f(x_1),\dots,f(x_n)]$, thus ensuring $\mathfrak{R}_n \to 0$.*

EXAMPLE A.4. (Kolmogorov–Smirnov distance). *The* KS *distance meets (III) by definition and, similarly to* MMD *with bounded kernels, also condition (IV) is satisfied without the need to impose additional constraints on the model μ_θ or on the data-generating process. More specifically, Assumption (IV) follows from the inequality $\mathfrak{R}_{\mu,n}(\mathfrak{F}) \leq 2[\log(n+1)/n]^{1/2}$ in Chapter 4.3.1 of Wainwright (2019). This is a consequence of the bounds on $\mathfrak{R}_{\mu,n}(\mathfrak{F})$ when $\mathfrak{F}$ is a class of b-uniformly bounded functions such that, for some $\nu \geq 1$, it holds* $\text{card}\{f(x_{1:n}) : f \in \mathfrak{F}\} \leq (n+1)^\nu$ *for any n and $x_{1:n}$ in $\mathcal{Y}^n$. When $\mathfrak{F} = \{1_{(-\infty,a]}\}_{a\in\mathbb{R}}$ each $x_{1:n}$ would divide the real line in at most $n+1$ intervals and every indicator function within $\mathfrak{F}$ will take value 1 for all $x_i \leq a$ and zero otherwise, meaning that* $\text{card}\{f(x_{1:n}) : f \in \mathfrak{F}\} \leq (n+1)$. *Therefore, by applying Equation (4.24) in Wainwright (2019), with $b=1$ and $\nu=1$, yields $\mathfrak{R}_{\mu,n}(\mathfrak{F}) \leq 2[\log(n+1)/n]^{1/2}$ for any $\mu \in \mathcal{P}(\mathcal{Y})$, which implies that Assumption (IV) is met. These derivations clarify the usefulness of the available techniques for upper bounding the Rademacher complexity (e.g., Wainwright, 2019, Chapter 4.3), leveraging, in this case, the notion of polynomial discrimination and the closely-related* VC *dimension.*

APPENDIX B: CONCENTRATION IN THE SPACE OF PARAMETERS

Theorem 3.3 in the main article is stated for neighborhoods within the space of distributions. Although such a perspective is in line with the overarching focus of current theory for discrepancy-based ABC (Jiang, Wu and Wong, 2018; Bernton et al., 2019; Nguyen et al., 2020; Frazier, 2020; Fujisawa et al., 2021), it shall be emphasized that similar results can be also derived in the space of parameters. To this end, it suffices to adapt Corollary 1 in Bernton et al. (2019) to our general framework, under the same additional assumptions, which are adapted below to the whole IPS class.

(V) The minimizer θ^* of $\mathcal{D}_{\mathfrak{F}}(\mu_\theta,\mu^*)$ exists and is well separated, meaning that for any $\delta > 0$ there is a $\delta' > 0$ such that $\inf_{\theta\in\Theta:d(\theta,\theta^*)>\delta} \mathcal{D}_{\mathfrak{F}}(\mu_\theta,\mu^*) > \mathcal{D}_{\mathfrak{F}}(\mu_{\theta^*},\mu^*) + \delta'$;

(VI) The parameters θ are identifiable, and there exist positive constants $K > 0$, $\nu > 0$ and an open neighborhood $U \subset \Theta$ of θ^* such that, for any $\theta \in U$, it holds that $d(\theta,\theta^*) \leq K\left[\mathcal{D}_{\mathfrak{F}}(\mu_\theta,\mu^*) - \varepsilon^*\right]^\nu$.

Assumptions (V) and (VI) essentially require that the parameters θ are identifiable, sufficiently well-separated, and that the distance $d(\cdot,\cdot)$ between parameter values has some reasonable correspondence with the discrepancy $\mathcal{D}_{\mathfrak{F}}(\cdot,\cdot)$ among the associated distributions. Although these two assumptions introduce a condition on the model, it shall be emphasized that (V) and (VI) are not specific to our framework (e.g., Frazier et al., 2018; Bernton et al., 2019; Frazier, 2020). On the contrary, these identifiability conditions are arguably customary and minimal requirements in parameter inference. Moreover, these two assumptions have been checked in Chérief-Abdellatif and Alquier (2022) for MMD and in Bernton et al. (2019) for Wasserstein distance, which are arguably the two most remarkable examples of IPS employed in the ABC context. Under (V) and (VI), it is possible to state Corollary B.1.

COROLLARY B.1. *Assume (I)–(IV) along with (V)–(VI), and that $\mathcal{D}_{\mathfrak{F}}$ denotes a discrepancy within the* IPS *class in Definition 2.1. Moreover, take $\bar{\varepsilon}_n \to 0$ as $n \to \infty$, with $n\bar{\varepsilon}_n^2 \to \infty$ and $\bar{\varepsilon}_n/\mathfrak{R}_n(\mathfrak{F}) \to \infty$. Then, the* ABC *posterior with threshold $\varepsilon_n = \varepsilon^* + \bar{\varepsilon}_n$ satisfies*

$$\pi_n^{(\varepsilon^*+\bar{\varepsilon}_n)}\left(\left\{\theta : d(\theta,\theta^*) > K\left[\frac{4\bar{\varepsilon}_n}{3} + 2\mathfrak{R}_n(\mathfrak{F}) + \left(\frac{2b^2}{n}\log\frac{n}{\bar{\varepsilon}_n^L}\right)^{1/2}\right]^{\nu}\right\}\right) \leq \frac{2\cdot 3^L}{c_\pi n},$$

with $\mathbb{P}_{y_{1:n}}$–probability going to 1 as $n \to \infty$.

As for Theorem 3.3, also Corollary B.1 holds more generally when replacing both $n/\bar{\varepsilon}_n^L$ and $c_\pi n$ with $M_n/\bar{\varepsilon}_n^L$ and $c_\pi M_n$, respectively, for any $M_n > 1$. The proof of Corollary B.1 follows directly from Theorem 3.3 and Assumptions (V)–(VI), thereby allowing to inherit the discussion after Theorem 3.3, also when the concentration is measured directly within the parameter space via $d(\theta,\theta^*)$. For instance, when $d(\cdot,\cdot)$ is the Euclidean distance and $\nu = 1$, this implies that whenever $\mathfrak{R}_n(\mathfrak{F}) = \mathcal{O}(n^{-1/2})$ the contraction rate will be in the order of $\mathcal{O}([\log(n)/n]^{1/2})$, which is the expected rate in parametric models.

APPENDIX C: EXTENSION TO NON-I.I.D. SETTINGS

Although the theoretical results in Sections 2–4 provide an improved understanding of the limiting properties of discrepancy-based ABC posteriors, the i.i.d. assumption in (I) rules out important settings which often require ABC. A remarkable case is that of time-dependent observations (e.g., Fearnhead and Prangle, 2012; Bernton et al., 2019; Nguyen et al., 2020; Drovandi and Frazier, 2022).

Section C.1 clarifies that the theory derived under i.i.d. assumptions in Section 2–4 can be naturally extended to these non-i.i.d. settings leveraging results for Rademacher complexity in β-mixing stochastic processes (Mohri and Rostamizadeh, 2008). Examples C.3–C.4 below show that such a class embraces several processes of direct practical interest. Extensions beyond this class, albeit relevant, are challenging even when the focus is on proving simpler, non-uniform, concentration results for a single discrepancy. Hence, these extensions are left for future research, that could be facilitated by the derivation of Rademacher complexity bounds for general processes beyond the β-mixing ones studied in Mohri and Rostamizadeh (2008).

C.1. Convergence and concentration beyond i.i.d. settings. Let us assume again that $\mathcal{Y}$ is a metric space endowed with distance ρ. However, unlike the i.i.d. setting considered in Section 2, we now focus on the situation in which the observed data $y_{1:n} = (y_1,\dots,y_n) \in \mathcal{Y}^n$ are dependent and drawn from the joint distribution $\mu^{*(n)} \in \mathcal{P}(\mathcal{Y}^n)$, where $\mathcal{P}(\mathcal{Y}^n)$ is the space of probability measures on $\mathcal{Y}^n$. Under this more general framework, the i.i.d. case is recovered by assuming that $\mu^{*(n)}$ can be expressed as a product, i.e., $\mu^{*(n)} = \prod_{i=1}^n \mu^*$.

In the following, the above product structure is not imposed. Instead, we only assume that the marginal of $\mu^{*(n)}$ is constant, and denoted with μ^*. Such an assumption is met whenever $y_{1:n}$ is extracted from a stationary stochastic process $(y_t)_{t\in\mathbb{Z}}$, thus embracing a broader variety of applications of direct interest. Under these settings, a statistical model is defined as a collection of distributions in $\mathcal{P}(\mathcal{Y}^n)$, i.e., $\{\mu_\theta^{(n)} : \theta \in \Theta \subseteq \mathbb{R}^p\}$, with a constant marginal denoted by μ_θ. Notice that these assumptions of constant marginals μ^* and μ_θ are made also in the available concentration theory under non-i.i.d. settings (see e.g., Bernton et al., 2019; Nguyen et al., 2020) when requiring convergence of $\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \mu^*)$ and suitable concentration inequalities for $\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \mu_\theta)$. As a result, the settings we consider are not more restrictive than those addressed in discrepancy-specific theory. In fact, both Bernton et al. (2019) and Nguyen et al. (2020) explicitly refer to stationary processes when discussing the validity of the assumptions on $\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \mu^*)$ and $\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \mu_\theta)$ in non-i.i.d. contexts.

Given the above statistical model, a prior π on θ and a generic IPS discrepancy $\mathcal{D}_{\mathfrak{F}}$, the ABC posterior with threshold $\varepsilon_n \geq 0$ is defined as

$$\pi_n^{(\varepsilon_n)}(\theta) \propto \pi(\theta) \int_{\mathcal{Y}^n} \mathbb{1}\{\mathcal{D}_{\mathfrak{F}}(z_{1:n}, y_{1:n}) \leq \varepsilon_n\}\ \mu_\theta^{(n)}(dz_{1:n}).$$

This definition is the same as the one in Section 2, with the only difference that $\mu_\theta^n = \prod_{i=1}^n \mu_\theta$ is replaced by the joint $\mu_\theta^{(n)}$, since now the data are no more assumed to be independent.

In order to extend the convergence result in Corollary 3.2 together with the concentration statement in Theorem 3.3 to the above framework, we require an analog of Equation (2) in Lemma 2.6 for time-dependent data. This generalization can be derived leveraging results in Mohri and Rostamizadeh (2008) under the notion of β-mixing coefficients.

DEFINITION C.1 (β-mixing). *Consider the stationary sequence $(x_t)_{t\in\mathbb{Z}}$ of random variables, and let $\sigma_j^{j'}$ be the σ–algebra generated by the random variables x_k, $j \leq k \leq j'$, for any $j, j' \in \mathbb{Z} \cup \{-\infty, +\infty\}$. Then, for any integer $k > 0$, the β-mixing coefficient of $(x_t)_{t\in\mathbb{Z}}$ is defined as*

$$\beta(k) = \sup_{t\in\mathbb{Z}} \mathbb{E}[\sup_{A\in\sigma_{t+k}^{\infty}} |\mathbb{P}(A \mid \sigma_{-\infty}^{t}) - \mathbb{P}(A)|].$$

If $\beta(k) \to 0$ as $k \to \infty$, then the stochastic process $(x_t)_{t\in\mathbb{Z}}$ is said to be β-mixing.

Intuitively, $\beta(k)$ measures the dependence between the past (before t) and the future (after $t+k$) of the process. When such a dependence is weak, we expect that $\beta(k)$ will decay to 0 fast when $k \to \infty$. In the most extreme case, when the x_t's are i.i.d., we have $\beta(k) = 0$ for all $k > 0$. More generally, as clarified in Definition C.1, a process having $\beta(k) \to 0$ when $k \to \infty$ is named β-mixing. We refer the reader to Doukhan (1994) for an in-depth study of the main properties of β-mixing processes along with a more comprehensive discussion of relevant examples. The most remarkable ones will be also presented in the following.

Leveraging the notion of β-mixing coefficient, Lemma C.2 extends Lemma 2.6 to the dependent setting. The proof can be found in Appendix D and combines Proposition 2 and Lemma 2 in Mohri and Rostamizadeh (2008). For readability, let us also introduce the notation $s_n = \lfloor n/(2\lfloor\sqrt{n}\rfloor)\rfloor$. Note that $s_n \sim \sqrt{n}/2$ as $n \to \infty$, and thus $s_n \to \infty$.

LEMMA C.2. *Define $s_n = \lfloor n/(2\lfloor\sqrt{n}\rfloor)\rfloor$. Moreover, consider the stationary stochastic process $(x_t)_{t\in\mathbb{Z}}$ and denote with $\beta(k), k \in \mathbb{N}$, its β-mixing coefficients. Let $\mu^{(n)}$ be the joint* *distribution of a sample $x_{1:n}$ extracted from $(x_t)_{t\in\mathbb{Z}}$ and denote with $\mu = \mu^{(1)}$ its constant marginal. Then, for any b-uniformly bounded class $\mathfrak{F}$, any integer $n \geq 1$ and scalar $\delta \geq 0$,*

$$\text{(C.1)}\quad \mathbb{P}_{x_{1:n}}\left[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{x_{1:n}}, \mu) \leq 2\mathfrak{R}_{\mu,s_n}(\mathfrak{F}) + \frac{4b}{\sqrt{n}} + \delta\right] \geq 1 - 2\cdot\exp\left[-\frac{s_n\delta^2}{2b^2}\right] - 2s_n\beta(\lfloor\sqrt{n}\rfloor),$$

with $\mathfrak{R}_{\mu,s_n}(\mathfrak{F})$ the Rademacher complexity in Definition 2.5 for an i.i.d. sample of size s_n from μ.

Equation (C.1) extends (2) beyond i.i.d. settings. This extension provides a bound that still depends on the Rademacher complexity in Definition 2.5 for an i.i.d. sample — in this case from the common marginal μ of the process $(x_t)_{t\in\mathbb{Z}}$. As such, Assumption (IV) requires no modifications, and no additional validity checks relative to those discussed in Section 3.3. This suggests that the Rademacher complexity framework might also be leveraged to derive improved convergence and concentration results for discrepancy-based ABC posteriors in more general situations which do not necessarily meet Assumption (I). To prove these results we leverage again Assumptions (II), (III) and (IV), and replace (I) with condition (VII).

(VII) The data $y_{1:n}$ are from a β-mixing stochastic process $(y_t)_{t\in\mathbb{Z}}$ with mixing coefficients $\beta(k) \leq C_\beta e^{-\gamma k^\xi}$ for some $C_\beta, \gamma, \xi > 0$, common marginal μ, and generic joint $\mu^{*(n')}$ for a sample $y_{1:n'}$ from $(y_t)_{t\in\mathbb{Z}}$ for any $n' \in \mathbb{N}$. The same β-mixing conditions hold also for the process $(z_t)_{t\in\mathbb{Z}}$ associated with the synthetic data $z_{1:n}$ from the assumed model. In this case, the joint distribution for a generic sample $z_{1:n'}$ is $\mu_\theta^{(n')}$, $\theta \in \Theta$, and the common marginal is denoted by μ_θ. For simplicity and without loss of generality, we also assume that the constants C_β, γ, and ξ are the same for $(y_t)_{t\in\mathbb{Z}}$ and $(z_t)_{t\in\mathbb{Z}}$.

Assumption (VII) is clearly more general than (I). As discussed previously, it embraces several stochastic processes of substantial interest in practical applications, including those in Examples C.3–C.4 below; see Doukhan (1994) for additional examples and discussion.

EXAMPLE C.3 (Doeblin-recurrent Markov chains). *Let $(x_t)_{t\in\mathbb{Z}}$ be a Markov chain on $\mathcal{Y} \subset \mathbb{R}^d$ with transition kernel $P(\cdot,\cdot)$. Such a Markov chain is said to be Doeblin-recurrent if there exists a probability measure q, a constant $0 < c \leq 1$ and an integer $r > 0$ such that, for any measurable set A and any $x \in \mathbb{R}^d$, $P^r(x, A) \geq cq(A)$. When this is the case, $(x_t)_{t\in\mathbb{Z}}$ is β-mixing with $\beta(k) \leq 2(1-c)^{k/r}$; see e.g., Theorem 1 in page 88 of Doukhan (1994).*

EXAMPLE C.4 (Hidden Markov chains). *Assume $(x_t)_{t\in\mathbb{Z}}$ is a β-mixing stochastic process with coefficients $\beta_x(k), k \in \mathbb{N}$. If $\tilde{x}_t = F(x_t, \varepsilon_t)$ with ε_t i.i.d., then the β-mixing coefficients of $(\tilde{x}_t)_{t\in\mathbb{Z}}$ satisfy $\beta_{\tilde{x}}(k) = \beta_x(k)$. Therefore, $(\tilde{x}_t)_{t\in\mathbb{Z}}$ is also β-mixing and inherits the bounds on $\beta_x(k)$. These processes are often used in practice with $(x_t)_{t\in\mathbb{Z}}$ being a Markov chain. In this case $(\tilde{x}_t)_{t\in\mathbb{Z}}$ is called a Hidden Markov chain.*

Section 2.4.2 of Doukhan (1994) also provides conditions on F and on the i.i.d. sequence $(\varepsilon_t)_{t\in\mathbb{Z}}$ ensuring that a stationary process $(x_t)_{t\in\mathbb{Z}}$ satisfying $x_t = F(x_{t-1}, \ldots, x_{t-k}, \varepsilon_t)$ exists and is β-mixing. Lemma C.5 specializes such a result in the context of Gaussian AR(1) processes, which will be considered in the empirical study in Section C.2.

LEMMA C.5 (Gaussian AR(1) process). *Consider a generic sequence $(\varepsilon_t)_{t\in\mathbb{Z}}$ of i.i.d. random variables from a $\mathrm{N}(0,\sigma^2)$. Moreover, let $-1<\theta<1$ and $\psi\in\mathbb{R}$. Then the stationary solution to $x_t=\psi+\theta x_{t-1}+\varepsilon_t$ is β-mixing and has coefficients $\beta(k)\leq|\theta|^k/(2\sqrt{1-\theta^2})=(2\sqrt{1-\theta^2})^{-1}\exp(-k\log(1/|\theta|))$, $k\in\mathbb{N}$, thus meeting (VII).*

Notice that in the empirical study in Section C.2 the focus will be on inference for the AR parameter θ. Clearly, in this case it is not sufficient to focus on the marginal distribution of each x_t. Rather, one should leverage the bivariate distribution for the pairs $\tilde{x}_t:=(x_t,x_{t+1})$; see also Bernton et al. (2019) where such a strategy is named delay reconstruction. This procedure simply changes the focus to the bivariate stochastic process $(\tilde{x}_t)_{t\in\mathbb{Z}}$, but does not alter the mixing properties. In particular, if $\beta_x(k)$ and $\beta_{\tilde{x}}(k)$ are the mixing coefficients of $(x_t)_{t\in\mathbb{Z}}$ and $(\tilde{x}_t)_{t\in\mathbb{Z}}$, respectively, then from Definition C.1 we have $\beta_{\tilde{x}}(k)=\beta_x(k-1)$, for $k\geq 1$. Notice that identifiability is a key to ensure concentration in the space of parameters of interest as in Corollary B.1. This motivates further research, beyond the scope of this article, to derive delay reconstruction strategies ensuring identifiability in more complex processes.

Leveraging Lemma C.2 along with the newly-introduced Assumption (VII), Proposition C.6 states convergence of the ABC posterior when $\varepsilon_n=\varepsilon$ is fixed and $n\to\infty$.

PROPOSITION C.6. *Under Assumptions (III), (IV) and (VII), for any $\varepsilon>\tilde{\varepsilon}$, it holds that*

$$\pi_n^{(\varepsilon)}(\theta)\to\pi(\theta\mid\mathcal{D}_{\mathfrak{F}}(\mu_\theta,\mu^*)\leq\varepsilon)\propto\pi(\theta)\mathbb{1}\left\{\mathcal{D}_{\mathfrak{F}}(\mu_\theta,\mu^*)\leq\varepsilon\right\},\tag{C.2}$$

almost surely with respect to $y_{1:n}\sim\mu^{*(n)}$, as $n\to\infty$.

According to Proposition C.6, replacing Assumption (I) with (VII), does not alter the uniform convergence properties of the ABC posterior originally stated in Corollary 3.2 under the i.i.d. assumption. This allows to inherit the discussion after Corollary 3.2 also beyond i.i.d. settings, while suggesting that similar extensions would be possible in the regime $\varepsilon_n\to\varepsilon^*$ and $n\to\infty$. These extensions are stated in Theorem C.7, which provides an important generalization of Theorem 3.3 beyond the i.i.d. case.

THEOREM C.7. *Let $\varepsilon_n=\varepsilon^*+\bar{\varepsilon}_n$, and assume (II), (III), (IV) and (VII). Then, if $\bar{\varepsilon}_n\to 0$ is such that $\sqrt{n}\bar{\varepsilon}_n^2\to\infty$ and $\bar{\varepsilon}_n/\mathfrak{R}_{s_n}(\mathfrak{F})\to\infty$, with $s_n=\lfloor n/(2\lfloor\sqrt{n}\rfloor)\rfloor$, we have*

$$\pi_n^{(\varepsilon^*+\bar{\varepsilon}_n)}\Big(\Big\{\theta:\mathcal{D}_{\mathfrak{F}}(\mu_\theta,\mu^*)>\varepsilon^*+\frac{4\bar{\varepsilon}_n}{3}+2\mathfrak{R}_{s_n}(\mathfrak{F})+\frac{4b}{\sqrt{n}}+\Big(\frac{2b^2}{s_n}\log\frac{n}{\bar{\varepsilon}_n^L}\Big)^{1/2}\Big\}\Big)\leq\frac{4\cdot 3^L}{c_\pi n},$$

with $\mathbb{P}_{y_{1:n}}$–probability going to 1 as $n\to\infty$, where $\mathfrak{R}_{s_n}=\sup_{\mu\in\mathcal{P}(\mathcal{Y})}\mathfrak{R}_{\mu,s_n}$.

As for Proposition C.6, also Theorem C.7 shows that informative concentration inequalities similar to those derived in Section 3.2, can be obtained beyond the i.i.d. setting. These results provide insights comparable to those in Theorem 3.3 with the only difference that in this case we require $\sqrt{n}\bar{\varepsilon}_n^2\to\infty$ rather than $n\bar{\varepsilon}_n^2\to\infty$ and the term $2b^2/n$ within the bound in Theorem 3.3 is now replaced by $2b^2/s_n$ with $s_n\sim\sqrt{n}/2$ as $n\to\infty$. This means that $\bar{\varepsilon}_n$ must shrink to zero with a rate at least $n^{1/4}$ slower than the one allowed in the i.i.d. setting.

This is an interesting result which clarifies that when moving beyond i.i.d. regimes concentration can still be achieved, although with a slower rate. Such a rate might be pessimistic in some models and we believe it may be improved under future refinements of Lemma C.2.

Notice that (VII) could be relaxed to include β-mixing processes whose coefficients $\beta(k)$ vanish to zero, but at a non-exponential rate, e.g., $\beta(k) \sim 1/(k+1)^{\xi}$ for some $\xi > 0$. In this case, we could still use Lemma C.2 to prove concentration, but with a smaller s_n, that would lead to even slower rates. However, we did not provide the most general result for the sake of readability. As for processes that are not β-mixing, we are not aware of results similar to Lemma C.2 in this context. This is an important direction for future research.

C.2. Illustrative simulation in non-i.i.d. settings. Let us illustrate the results in Section C.1 on a simple simulation study focusing on a contaminated Gaussian $\text{AR}(1)$ process. More specifically, the uncontaminated data are generated from the model $y^*_t = 0.5y^*_{t-1} + \varepsilon_t$ for $t = 1, \dots, 100$ with $\varepsilon_t \sim \text{N}(0, 1)$ independently, and initial state $y^*_0 \sim \text{N}(0, 1)$. Then, similarly to the simulation study in Section 5, these data are contaminated with a growing fraction $\alpha \in \{0.05, 0.10, 0.15\}$ of independent realizations from a $\text{N}(20, 1)$. As such, each observed data point y_t is either equal to y^*_t or to a sample from $\text{N}(20, 1)$, for $t = 1, \dots, 100$. For Bayesian inference, we assume an $\text{AR}(1)$ model $z_t = \theta z_{t-1} + \varepsilon_t$, with $\varepsilon_t \sim \text{N}(0, 1)$, and focus on learning θ via discrepancy-based ABC under a uniform prior on $[-1, 1]$ for θ.

Rejection ABC is implemented under the same settings and discrepancies considered in Section 5. However, as discussed in Section C.1, in this case we focus on distances among the empirical distributions of the $n = m = 100$ observed $(y_0, y_1), (y_1, y_2), \dots, (y_{99}, y_{100})$ and synthetic $(z_0, z_1), (z_1, z_2), \dots, (z_{99}, z_{100})$ pairs. This is consistent with the delay reconstruction strategy in Bernton et al. (2019) and is motivated by the fact that information on θ is in the bivariate distributions, rather than in the marginals. For the same reason, in implementing summary-based ABC we consider the sample covariance rather than the sample mean.

Table C.1 summarizes the concentration achieved by the different discrepancies analyzed under the aforementioned non-i.i.d. data generating process and model, at varying contamination $\alpha \in \{0.05, 0.10, 0.15\}$. The results are coherent with those displayed in Table 1 for the i.i.d. scenario and further clarify that discrepancies with guarantees of uniform convergence and concentration generally provide a robust choice, including in non-i.i.d. contexts.

TABLE C.1

Concentration and runtimes in seconds (for a single discrepancy evaluation) of ABC under MMD with Gaussian kernel, Wasserstein-1 distance, summary-based distance (covariance) and KL divergence for an $\text{AR}(1)$ Huber contamination model with $\alpha \in \{0.05, 0.10, 0.15\}$. $\text{MSE} = \hat{\mathbb{E}}_{\mu^*}[\hat{\mathbb{E}}_{\text{ABC}}(\theta - \theta_0)^2]$, $\theta_0 = 0.5$.

	MSE ($\alpha = 0.05$)	MSE ($\alpha = 0.10$)	MSE ($\alpha = 0.15$)	time
(IPS) MMD	**0.029**	**0.036**	**0.049**	$<$ 0.01”
(IPS) Wasserstein-1	0.043	0.091	0.180	$<$ 0.01”
(IPS) summary (covariance)	0.575	0.998	1.001	$<$ 0.01”
(non–IPS) KL	0.058	0.060	0.061	$<$ 0.01”

APPENDIX D: PROOFS OF THEOREMS, COROLLARIES AND PROPOSITIONS

PROOF OF THEOREM 3.1. Note that, by leveraging the first inequality in Lemma 2.6, we have $\mathbb{P}_{y_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \mu^*) > 2\mathfrak{R}_{\mu^*,n}(\mathfrak{F}) + \delta] \leq \exp(-n\delta^2/2b^2)$. Hence, setting $\delta = 1/n^{1/4}$, and recalling that $\mathfrak{R}_{\mu^*,n}(\mathfrak{F}) \leq \mathfrak{R}_n(\mathfrak{F})$, it follows $\mathbb{P}_{y_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \mu^*) > 2\mathfrak{R}_n(\mathfrak{F}) + 1/n^{1/4}] \leq \exp(-\sqrt{n}/2b^2)$; note that $\sum_{n\geq 0}\exp(-\sqrt{n}/2b^2) < \infty$. Therefore, if we define the event

$$\mathrm{E}_n = \{\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \mu^*) \leq 2\mathfrak{R}_n(\mathfrak{F}) + 1/n^{1/4}\}, \tag{D.1}$$

then $\mathbb{1}\{\mathrm{E}_n^c\} \to 0$ almost surely with respect to $y_{1:n} \overset{\text{i.i.d.}}{\sim} \mu^*$ as $n \to \infty$. Now, notice that

$$\pi_n^{(\varepsilon)}\{\theta : \mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) \leq \varepsilon\} = \pi_n^{(\varepsilon)}\{\theta : \mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) \leq \varepsilon\}\mathbb{1}\{\mathrm{E}_n\} + \pi_n^{(\varepsilon)}\{\theta : \mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) \leq \varepsilon\}\mathbb{1}\{\mathrm{E}_n^c\}.$$

Hence, in the following we focus on $\pi_n^{(\varepsilon)}\{\theta : \mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) \leq \varepsilon\}\mathbb{1}\{\mathrm{E}_n\}$. To this end, recall that

$$\begin{aligned}
\pi_n^{(\varepsilon)}(\theta) &\propto \pi(\theta)\int \mathbb{1}\left\{\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \hat{\mu}_{z_{1:n}}) \leq \varepsilon\right\}\mu_\theta^n(dz_{1:n}) \\
&= \pi(\theta)\int \mathbb{1}\left\{\mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) \leq \varepsilon + W_{\mathfrak{F}}(z_{1:n})\right\}\mu_\theta^n(dz_{1:n}) =: \pi(\theta)p_n(\theta),
\end{aligned}$$

where $W_{\mathfrak{F}}(z_{1:n}) = \mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) - \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \hat{\mu}_{z_{1:n}})$, whereas $p_n(\theta)$ denotes the probability of generating a sample $z_{1:n}$ from μ_θ^n which leads to accept the parameter value θ. Note that, by applying the triangle inequality twice, we have

$$-\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \mu_\theta) - \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \mu^*) \leq W_{\mathfrak{F}}(z_{1:n}) \leq \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \mu_\theta) + \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \mu^*),$$

and, hence, $|W_{\mathfrak{F}}(z_{1:n})| \leq \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \mu_\theta) + \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \mu^*)$. This implies that the quantity $p_n(\theta)$ can be bounded below and above as follows

$$\begin{aligned}
&\int \mathbb{1}\left\{\mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) \leq \varepsilon - \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \mu_\theta) - \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \mu^*)\right\}\mu_\theta^n(dz_{1:n}) \\
&\qquad \leq p_n(\theta) \leq \int \mathbb{1}\left\{\mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) \leq \varepsilon + \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \mu_\theta) + \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \mu^*)\right\}\mu_\theta^n(dz_{1:n}).
\end{aligned}$$

Applying again Lemma 2.6 yields $\mathbb{P}_{z_{1:n}}\left[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \mu_\theta) > 2\mathfrak{R}_{\mu_\theta,n}(\mathfrak{F}) + \delta\right] \leq \exp(-n\delta^2/2b^2)$. Therefore, setting $\delta = 1/n^{1/4}$, and recalling that $\mathfrak{R}_{\mu_\theta,n}(\mathfrak{F}) \leq \mathfrak{R}_n(\mathfrak{F})$ and that we are on the event given in (D.1), it follows

$$\begin{aligned}
&-\exp(-\sqrt{n}/2b^2) + \mathbb{1}\{\mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) \leq \varepsilon - 4\mathfrak{R}_n(\mathfrak{F}) - 2/n^{1/4}\} \\
&\qquad \leq p_n(\theta) \leq \exp(-\sqrt{n}/2b^2) + \mathbb{1}\{\mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) \leq \varepsilon + 4\mathfrak{R}_n(\mathfrak{F}) + 2/n^{1/4}\}.
\end{aligned} \tag{D.2}$$

Now, notice that the acceptance probability is defined as $p_n = \int p_n(\theta)\pi(d\theta)$. Hence, integrating with respect to $\pi(\theta)$ in the above inequalities yields, for n large enough,

$$\pi\{\theta : \mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) \leq \varepsilon - c_{\mathfrak{F}}\} - e_n \leq p_n \leq \pi\{\theta : \mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) \leq \varepsilon + c_{\mathfrak{F}}\} + e_n,$$

where $c_{\mathfrak{F}} = 4\limsup \mathfrak{R}_n(\mathfrak{F})$, as in Equation (3), thus concluding the first part of the proof.

To proceed with the second part of the proof, notice that, by the definition of $\tilde{\varepsilon} = \inf\{\epsilon > 0 : \pi\{\theta : \mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) \leq \epsilon\} > 0\}$, the left part of the above inequality is bounded away from zero for n large enough, whenever $\varepsilon - c_{\mathfrak{F}} > \tilde{\varepsilon}$. This implies that also the acceptance probability p_n is strictly positive. As a consequence, for such n, it follows that

$$\pi_n^{(\varepsilon)}(\mathrm{A}) = \frac{\int p_n(\theta)\mathbb{1}_{\mathrm{A}}(\theta)\pi(d\theta)}{\int p_n(\theta)\pi(d\theta)} = \frac{\int p_n(\theta)\mathbb{1}_{\mathrm{A}}(\theta)\pi(d\theta)}{p_n},$$

is well-defined for any event A. Then, leveraging the upper bound in (D.2) yields

$$\begin{aligned}
\pi_n^{(\varepsilon)}\{\theta : \mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) > \varepsilon + 4\mathfrak{R}_n(\mathfrak{F})\} &= \frac{\int p_n(\theta)\mathbb{1}\left\{\mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) > \varepsilon + 4\mathfrak{R}_n(\mathfrak{F})\right\}\pi(d\theta)}{\int p_n(\theta)\pi(d\theta)} \\
\leq \frac{\int \mathbb{1}\left\{\mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) \leq \varepsilon + 4\mathfrak{R}_n(\mathfrak{F}) + 2/n^{1/4}\right\}\mathbb{1}\left\{\mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) > \varepsilon + 4\mathfrak{R}_n(\mathfrak{F})\right\}\pi(d\theta)}{p_n} \\
+ \frac{\int \exp(-\sqrt{n}/2b^2)\mathbb{1}\left\{\mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) > \varepsilon + 4\mathfrak{R}_n(\mathfrak{F})\right\}\pi(d\theta)}{p_n}.
\end{aligned}$$

To conclude the proof it is now necessary to control both terms. Note that we already proved that the denominator p_n is bounded away from zero for n large enough. Both numerators are bounded by 1, and going to 0 when $n \to \infty$. Thus, by the dominated convergence theorem, both summands in the above upper bound for $\pi_n^{(\varepsilon)}\{\theta : \mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) > \varepsilon + 4\mathfrak{R}_n(\mathfrak{F})\}$ go to zero. This implies, as a direct consequence, that $\pi_n^{(\varepsilon)}\{\theta : \mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) \leq \varepsilon + 4\mathfrak{R}_n(\mathfrak{F})\} \to 1$, almost surely with respect to $y_{1:n} \stackrel{\text{i.i.d.}}{\sim} \mu^*$ as $n \to \infty$, thereby concluding the proof. □

PROOF OF COROLLARY 3.2. Note that by combining Equation (2) in Lemma 2.6 with the result $\sum_{n>0} \exp[-n\delta^2/(2b^2)] < \infty$, the Borel–Cantelli Lemma implies that both $\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \mu_\theta)$ and $\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \mu^*)$ converge to 0 almost surely when $\mathfrak{R}_n(\mathfrak{F}) \to 0$ as $n \to \infty$. Hence, since

$$-\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \mu_\theta) - \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \mu^*) \leq \mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) - \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \hat{\mu}_{z_{1:n}}) \leq \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \mu_\theta) + \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \mu^*),$$

it follows that $\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \hat{\mu}_{y_{1:n}}) \to \mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*)$ almost surely as $n \to \infty$.

Combining this result with the proof of Theorem 1 in Jiang, Wu and Wong (2018) yields the statement of Corollary 3.2. Notice that, as discussed in Section 3.1, the limiting pseudo-posterior in Corollary 3.2 is well-defined only for those $\varepsilon > \tilde{\varepsilon}$, with $\tilde{\varepsilon}$ as in Theorem 3.1. □

PROOF OF THEOREM 3.3. Since Lemma 2.6 and $\mathfrak{R}_n(\mathfrak{F}) = \sup_{\mu \in \mathcal{P}(\mathcal{Y})} \mathfrak{R}_{\mu,n}(\mathfrak{F}) \geq \mathfrak{R}_{\mu,n}(\mathfrak{F})$ hold for every $\mu \in \mathcal{P}(\mathcal{Y})$, then, for every integer $n \geq 1$ and any scalar $\delta \geq 0$, Equation (2) implies $\mathbb{P}_{x_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{x_{1:n}}, \mu) \leq 2\mathfrak{R}_n(\mathfrak{F}) + \delta] \geq 1 - \exp(-n\delta^2/2b^2)$. Moreover, since this result holds for any $\delta \geq 0$, it follows that $\mathbb{P}_{x_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{x_{1:n}}, \mu) \leq 2\mathfrak{R}_n(\mathfrak{F}) + (c_1 - 2\mathfrak{R}_n(\mathfrak{F}))] \geq 1 - \exp[-n(c_1 - 2\mathfrak{R}_n(\mathfrak{F}))^2/2b^2]$, for any $c_1 \geq 2\mathfrak{R}_n(\mathfrak{F})$. Hence,

$$\mathbb{P}_{x_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{x_{1:n}}, \mu) \leq c_1] \geq 1 - \exp[-n(c_1 - 2\mathfrak{R}_n(\mathfrak{F}))^2/2b^2]. \tag{D.3}$$

Recalling the settings of Theorem 3.3, consider the sequence $\bar{\varepsilon}_n \to 0$ as $n \to \infty$, with $n\bar{\varepsilon}_n^2 \to \infty$ and $\bar{\varepsilon}_n/\mathfrak{R}_n(\mathfrak{F}) \to \infty$, which is possible by Assumption (IV). These regimes imply that $\bar{\varepsilon}_n$

goes to zero slower than $\mathfrak{R}_n(\mathfrak{F})$ and, hence, for n large enough, $\bar{\varepsilon}_n/3 > 2\mathfrak{R}_n(\mathfrak{F})$. Therefore, under Assumptions (I)–(III) it is now possible to apply (D.3) to $y_{1:n}$, by setting $c_1 = \bar{\varepsilon}_n/3$, which yields

$$\mathbb{P}_{y_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \mu^*) \leq \bar{\varepsilon}_n/3] \geq 1 - \exp[-n(\bar{\varepsilon}_n/3 - 2\mathfrak{R}_n(\mathfrak{F}))^2/2b^2].$$

Since $-n(\bar{\varepsilon}_n/3 - 2\mathfrak{R}_n(\mathfrak{F}))^2 = -n\bar{\varepsilon}_n^2[1/9 + 4(\mathfrak{R}_n(\mathfrak{F})/\bar{\varepsilon}_n)^2 - (4/3)\mathfrak{R}_n(\mathfrak{F})/\bar{\varepsilon}_n]$, it follows that $-n(\bar{\varepsilon}_n/3 - 2\mathfrak{R}_n(\mathfrak{F}))^2 \to -\infty$ when $n \to \infty$. From the above settings we also have that $n\bar{\varepsilon}_n^2 \to \infty$ and $\mathfrak{R}_n(\mathfrak{F})/\bar{\varepsilon}_n \to 0$, when $n \to \infty$. Therefore, as a consequence, we obtain $1 - \exp[-n(\bar{\varepsilon}_n/3 - 2\mathfrak{R}_n(\mathfrak{F}))^2/2b^2] \to 1$ as $n \to \infty$. Hence, in the rest of this proof, we will restrict to the event $\{\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \mu^*) \leq \bar{\varepsilon}_n/3\}$.

Denote with $\mathbb{P}_{\theta,z_{1:n}}$ the joint distribution of $\theta \sim \pi$ and $z_{1:n}$ i.i.d. from μ_θ. By definition of conditional probability, for any c_2, including $c_2 > 2\mathfrak{R}_n(\mathfrak{F})$, it follows that

$$\text{(D.4)}\quad \begin{aligned} &\pi_n^{(\varepsilon^*+\bar{\varepsilon}_n)}(\{\theta : \mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) > \varepsilon^* + 4\bar{\varepsilon}_n/3 + c_2\}) \\ &\qquad = \frac{\mathbb{P}_{\theta,z_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) > \varepsilon^* + 4\bar{\varepsilon}_n/3 + c_2, \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \hat{\mu}_{y_{1:n}}) \leq \varepsilon^* + \bar{\varepsilon}_n]}{\mathbb{P}_{\theta,z_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \hat{\mu}_{y_{1:n}}) \leq \varepsilon^* + \bar{\varepsilon}_n]}. \end{aligned}$$

To derive an upper bound for the above ratio, we first identify an upper bound for its numerator. In addressing this goal, we leverage the triangle inequality $\mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) \leq \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \mu_\theta) + \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \hat{\mu}_{y_{1:n}}) + \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \mu^*)$, since $\mathcal{D}_{\mathfrak{F}}$ is a semimetric, and the previously-proved result that the event $\{\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \mu^*) \leq \bar{\varepsilon}_n/3\}$ has $\mathbb{P}_{y_{1:n}}$–probability going to 1, thereby obtaining

$$\begin{aligned} &\mathbb{P}_{\theta,z_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) > \varepsilon^* + 4\bar{\varepsilon}_n/3 + c_2, \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \hat{\mu}_{y_{1:n}}) \leq \varepsilon^* + \bar{\varepsilon}_n] \\ &\qquad \leq \mathbb{P}_{\theta,z_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \hat{\mu}_{y_{1:n}}) + \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \mu_\theta) + \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \mu^*) > \varepsilon^* + 4\bar{\varepsilon}_n/3 + c_2, \\ &\qquad\qquad \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \hat{\mu}_{y_{1:n}}) \leq \varepsilon^* + \bar{\varepsilon}_n] \\ &\qquad \leq \mathbb{P}_{\theta,z_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \mu_\theta) + \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \mu^*) > \bar{\varepsilon}_n/3 + c_2] \leq \mathbb{P}_{\theta,z_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \mu_\theta) > c_2]. \end{aligned}$$

Rewriting $\mathbb{P}_{\theta,z_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \mu_\theta) > c_2]$ as $\int_{\theta\in\Theta} \mathbb{P}_{z_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \mu_\theta) > c_2 \mid \theta]\,\pi(\mathrm{d}\theta)$ and applying (D.3) to $z_{1:n}$ yields

$$\begin{aligned} &\int_{\theta\in\Theta} \mathbb{P}_{z_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \mu_\theta) > c_2|\theta]\pi(\mathrm{d}\theta) = \int_{\theta\in\Theta} (1 - \mathbb{P}_{z_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \mu_\theta) \leq c_2|\theta])\pi(\mathrm{d}\theta) \\ &\qquad \leq \int_{\theta\in\Theta} \exp[-n(c_2 - 2\mathfrak{R}_n(\mathfrak{F}))^2/2b^2]\pi(\mathrm{d}\theta) = \exp[-n(c_2 - 2\mathfrak{R}_n(\mathfrak{F}))^2/2b^2]. \end{aligned}$$

Hence, the numerator of the ratio in Equation (D.4) can be upper bounded by $\exp[-n(c_2 - 2\mathfrak{R}_n(\mathfrak{F}))^2/2b^2]$ for any $c_2 > 2\mathfrak{R}_n(\mathfrak{F})$. As for the denominator, defining the event $\mathrm{E}_n := \{\theta \in \Theta : \mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) \leq \varepsilon^* + \bar{\varepsilon}_n/3\}$ and applying again the triangle inequality, we have that

$$\mathbb{P}_{\theta,z_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \hat{\mu}_{y_{1:n}}) \leq \varepsilon^* + \bar{\varepsilon}_n] \geq \int_{\mathrm{E}_n} \mathbb{P}_{z_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \hat{\mu}_{y_{1:n}}) \leq \varepsilon^* + \bar{\varepsilon}_n \mid \theta]\,\pi(\mathrm{d}\theta)$$

$$\geq \int_{\mathrm{E}_n} \mathbb{P}_{z_{1:n}} [\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \mu^*) + \mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) + \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \mu_\theta) \leq \varepsilon^* + \bar{\varepsilon}_n \mid \theta] \, \pi(\mathrm{d}\theta)$$

$$\geq \int_{\mathrm{E}_n} \mathbb{P}_{z_{1:n}} [\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \mu_\theta) \leq \bar{\varepsilon}_n/3 \mid \theta] \, \pi(\mathrm{d}\theta),$$

where the last inequality follows directly from the fact that it is possible to restrict to the event $\{\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \mu^*) \leq \bar{\varepsilon}_n/3\}$, and that we are integrating over $\mathrm{E}_n := \{\theta \in \Theta : \mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) \leq \varepsilon^* + \bar{\varepsilon}_n/3\}$. Applying again (D.3) to $z_{1:n}$, with $c_1 = \bar{\varepsilon}_n/3 > 2\mathfrak{R}_n(\mathfrak{F})$, the last term of the above inequality can be further lower bounded by

$$\int_{\mathrm{E}_n} (1 - \exp[-n(\bar{\varepsilon}_n/3 - 2\mathfrak{R}_n(\mathfrak{F}))^2/2b^2])\pi(\mathrm{d}\theta)$$
$$= \pi(\mathrm{E}_n)(1 - \exp[-n(\bar{\varepsilon}_n/3 - 2\mathfrak{R}_n(\mathfrak{F}))^2/2b^2]),$$

with $\pi(\mathrm{E}_n) \geq c_\pi(\bar{\varepsilon}_n/3)^L$ by (II), and, as shown before, $1 - \exp[-n(\bar{\varepsilon}_n/3 - 2\mathfrak{R}_n(\mathfrak{F}))^2/2b^2] \to 1$, when $n \to \infty$, which implies, for n large enough, $1 - \exp[-n(\bar{\varepsilon}_n/3 - 2\mathfrak{R}_n(\mathfrak{F}))^2/2b^2] > 1/2$. Leveraging both results, the denominator in (D.4) is lower bounded by $(c_\pi/2)\,(\bar{\varepsilon}_n/3)^L$. Let us now combine the upper and lower bounds derived, respectively, for the numerator and the denominator of the ratio in (D.4), to obtain

$$\text{(D.5)} \quad \pi_n^{(\varepsilon^*+\bar{\varepsilon}_n)}(\{\theta \in \Theta : \mathcal{D}_{\mathfrak{F}}(\mu_\theta,\mu^*) > \varepsilon^* + 4\bar{\varepsilon}_n/3 + c_2\}) \leq \frac{\exp[-n(c_2 - 2\mathfrak{R}_n(\mathfrak{F}))^2/2b^2]}{(c_\pi/2)(\bar{\varepsilon}_n/3)^L},$$

with $\mathbb{P}_{y_{1:n}}$–probability going to 1 as $n \to \infty$. To conclude the proof it suffices to replace c_2 in (D.5) with $2\mathfrak{R}_n(\mathfrak{F}) + \sqrt{(2b^2/n)\log(M_n/\bar{\varepsilon}_n^L)}$, which is never lower than $2\mathfrak{R}_n(\mathfrak{F})$. Finally, setting $M_n = n$ yields the statement of Theorem 3.3. □

PROOF OF COROLLARY B.1. Corollary B.1 follows by replacing the bounds in the proof of Corollary 1 by Bernton et al. (2019) with the newly-derived ones in Theorem 3.3. □

PROOF OF COROLLARY 4.1. Recall that in the case of MMD with kernels bounded by 1 we have $\mathfrak{R}_n(\mathfrak{F}) \leq n^{-1/2}$. Hence, regarding the upper and lower bounds on p_n in (3) it holds

$$\pi\{\theta : \mathcal{D}_{\mathrm{MMD}}(\mu_\theta, \mu^*) \leq \varepsilon - c_{\mathfrak{F}}\} \geq \pi\{\theta : \mathcal{D}_{\mathrm{MMD}}(\mu_\theta, \mu^*) \leq \varepsilon - 4/\sqrt{n}\},$$
$$\pi\{\theta : \mathcal{D}_{\mathrm{MMD}}(\mu_\theta, \mu^*) \leq \varepsilon + c_{\mathfrak{F}}\} \leq \pi\{\theta : \mathcal{D}_{\mathrm{MMD}}(\mu_\theta, \mu^*) \leq \varepsilon + 4/\sqrt{n}\}.$$

Combining the above inequalities with the result in (3), and taking the limit for $n \to \infty$, proves the first part of the statement. The second part is a direct application of Corollary 3.2 to the case of MMD with bounded kernels, after noticing that the aforementioned inequality $\mathfrak{R}_n(\mathfrak{F}) \leq n^{-1/2}$ implies $\mathfrak{R}_n(\mathfrak{F}) \to 0$ as $n \to \infty$. □

PROOF OF COROLLARY 4.2. To prove Corollary 4.2, it suffices to plug $\bar{\varepsilon}_n = [(\log n)/n]^{\frac{1}{2}}$ and $b = 1$ into the statement of Theorem 3.3, and then upper-bound the resulting radius via the inequalities $\mathfrak{R}_n(\mathfrak{F}) \leq n^{-1/2}$ and $\log n \geq 1$. The latter holds for any $n \geq 3$ and hence for $n \to \infty$. □

PROOF OF PROPOSITION 4.3. We first show that, under $(A1)$–$(A3)$, Assumptions 1 and 2 made in Bernton et al. (2019) are satisfied under MMD when $f_n(\bar{\varepsilon}_n) = 1/(n\bar{\varepsilon}_n^2)$ and $c(\theta) = \mathbb{E}_z\left[k(z,z)\right]$, with $z \sim \mu_\theta$.

Consistent with the above goal, first recall that, by standard properties of MMD,

$$\mathcal{D}^2_{\text{MMD}}(\mu_1,\mu_2) = \mathbb{E}_{x_1,x_1'}\left[k(x_1,x_1')\right] - 2\mathbb{E}_{x_1,x_2}\left[k(x_1,x_2)\right] + \mathbb{E}_{x_2,x_2'}\left[k(x_2,x_2')\right], \tag{D.6}$$

with $x_1, x_1' \sim \mu_1$ and $x_2, x_2' \sim \mu_2$, all independently; see e.g., Chérief-Abdellatif and Alquier (2022). Since $k(x,x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}$ (see e.g., Muandet et al., 2017), the above result implies that

$$\begin{aligned} &\mathcal{D}^2_{\text{MMD}}(\mu_1,\mu_2) \\ &= \mathbb{E}_{x_1,x_1'}\left[\langle \phi(x_1), \phi(x_1') \rangle_{\mathcal{H}}\right] - 2\mathbb{E}_{x_1,x_2}\left[\langle \phi(x_1), \phi(x_2) \rangle_{\mathcal{H}}\right] + \mathbb{E}_{x_2,x_2'}\left[\langle \phi(x_2), \phi(x_2') \rangle_{\mathcal{H}}\right] \\ &= ||\mathbb{E}_{x_1}[\phi(x_1)]||^2_{\mathcal{H}} - 2\left\langle \mathbb{E}_{x_1}[\phi(x_1)], \mathbb{E}_{x_2}[\phi(x_2)]\right\rangle_{\mathcal{H}} + ||\mathbb{E}_{x_2}[\phi(x_2)]||^2_{\mathcal{H}} \\ &= \|\mathbb{E}_{x_1}\left[\phi(x_1)\right] - \mathbb{E}_{x_2}\left[\phi(x_2)\right]\|^2_{\mathcal{H}}. \end{aligned} \tag{D.7}$$

Leveraging Equations (D.6)–(D.7) and basic Markov inequalities, for any $\bar{\varepsilon}_n \geq 0$, it holds

$$\begin{aligned} &\mathbb{P}_{y_{1:n}}\left[\mathcal{D}_{\text{MMD}}(\hat{\mu}_{y_{1:n}},\mu^*) > \bar{\varepsilon}_n\right] \leq (1/\bar{\varepsilon}_n^2)\mathbb{E}_{y_{1:n}}\left[\mathcal{D}^2_{\text{MMD}}(\hat{\mu}_{y_{1:n}},\mu^*)\right] \\ &= (1/\bar{\varepsilon}_n^2)\mathbb{E}_{y_{1:n}}[\|(1/n)\textstyle\sum_{i=1}^n \phi(y_i) - \mathbb{E}_y\left[\phi(y)\right]\|^2_{\mathcal{H}}] \leq [1/(n^2\bar{\varepsilon}_n^2)]\textstyle\sum_{i=1}^n \mathbb{E}_{y_i}[\|\phi(y_i)\|^2_{\mathcal{H}}] \\ &\leq [1/(n\bar{\varepsilon}_n^2)]\mathbb{E}_{y_1}[\|\phi(y_1)\|^2_{\mathcal{H}}] = [1/(n\bar{\varepsilon}_n^2)]\mathbb{E}_{y_1}\left[k(y_1,y_1)\right] = [1/(n\bar{\varepsilon}_n^2)]\mathbb{E}_y\left[k(y,y)\right], \end{aligned}$$

with $y \sim \mu^*$. Since $[1/(n\bar{\varepsilon}_n^2)]\mathbb{E}_y\left[k(y,y)\right] \to 0$ as $n \to \infty$ by condition $(A1)$, we have that $\mathcal{D}_{\text{MMD}}(\hat{\mu}_{y_{1:n}},\mu^*) \to 0$ in $\mathbb{P}_{y_{1:n}}$–probability as $n \to \infty$, thus meeting Assumption 1 in Bernton et al. (2019). Moreover, as a direct consequence of the above derivations,

$$\mathbb{P}_{z_{1:n}}\left[\mathcal{D}_{\text{MMD}}(\hat{\mu}_{z_{1:n}},\mu_\theta) > \bar{\varepsilon}_n\right] \leq [1/(n\bar{\varepsilon}_n^2)]\mathbb{E}_z\left[k(z,z)\right].$$

Thus, setting $1/(n\bar{\varepsilon}_n^2) = f_n(\bar{\varepsilon}_n)$ and $\mathbb{E}_z\left[k(z,z)\right] = c(\theta)$, with $z \sim \mu_\theta$, ensures that

$$\mathbb{P}_{z_{1:n}}\left[\mathcal{D}_{\text{MMD}}(\hat{\mu}_{z_{1:n}},\mu_\theta] > \bar{\varepsilon}_n\right] \leq c(\theta)f_n(\bar{\varepsilon}_n),$$

with $f_n(u) = 1/(nu^2)$ strictly decreasing in u for any fixed n, and $f_n(u) \to 0$ as $n \to \infty$, for fixed u. Moreover, by Assumptions $(A2)$–$(A3)$, $c(\theta) = \mathbb{E}_z\left[k(z,z)\right]$ is π-integrable and there exist a $\delta_0 > 0$ and a $c_0 > 0$ such that $c(\theta) < c_0$ for any θ satisfying $(\mathbb{E}_{z,z'}\left[k(z,z')\right] - 2\mathbb{E}_{z,y}\left[k(y,z)\right] + \mathbb{E}_{y,y'}\left[k(y,y')\right])^{1/2} = \mathcal{D}_{\text{MMD}}(\mu_\theta,\mu^*) \leq \varepsilon^* + \delta_0$. This ensures that Assumption 2 in Bernton et al. (2019) holds.

Finally, note that Assumption 3 in Bernton et al. (2019), is verified by our Assumption (II). Therefore, under the assumptions in Proposition 4.3 it is possible to apply Proposition 3 in Bernton et al. (2019) with $f_n(\bar{\varepsilon}_n) = 1/(n\bar{\varepsilon}_n^2)$, $c(\theta) = \mathbb{E}_z\left[k(z,z)\right]$ and $R = M_n$, which yields the concentration result in Proposition 4.3. □

PROOF OF PROPOSITION 4.4. Assumptions (A1') and (A2') imply that the norm $|||x|||_{\psi_1}$ (see Equation (14) in Lei (2020) for a definition) is uniformly bounded when $x \sim \mu_\theta$ and $x \sim \mu^*$. Thus, we can use (15) in Lei (2020) to obtain

$$\mathbb{P}_{z_{1:n}}(\mathcal{D}_{\text{wass}}(\mu_\theta, \hat{\mu}_{z_{1:n}}) > u) \leq \exp[-c'n(u - c_1 n^{-1/\max(d,3)})_+^2] =: f_n(u)$$

and, similarly, $\mathbb{P}_{y_{1:n}}(\mathcal{D}_{\text{wass}}(\mu^*, \hat{\mu}_{y_{1:n}}) > u) \leq f_n(u)$, where c' depends only on the constant c in (A1') and (A2'), and c_1 is a universal constant. Notice that (15) in Lei (2020) requires $d > 2$. If $d = 1$, we can define $x' = (x, 0, 0) \in \mathbb{R}^3$ and apply the result in $\mathbb{R}^3$ (we can proceed similarly if $d = 2$). This is why Proposition 4.4 is stated with $\max(d, 3)$. The above bounds ensure that Assumptions 1–2 in Bernton et al. (2019) are met. Moreover, condition (II) verifies Assumption 3 in Bernton et al. (2019). Thus, we can apply Proposition 3 of Bernton et al. (2019) with $f_n(\cdot)$ defined as above and under the vanishing conditions on $\bar{\varepsilon}_n$ in Proposition 4.4. This implies that when $n \to \infty$, and $\bar{\varepsilon}_n \to 0$ such that $f_n(\bar{\varepsilon}_n) \to 0$, for some $C \in (0, \infty)$ and any $M_n \in (0, \infty)$, with $\mathbb{P}_{y_{1:n}}$–probability going to 1 as $n \to \infty$, it holds

$$\pi_n^{(\varepsilon^* + \bar{\varepsilon}_n)}(\{\theta : \mathcal{D}_{\text{wass}}(\mu_\theta, \mu^*) > \varepsilon^* + 4\bar{\varepsilon}_n/3 + f_n^{-1}(\bar{\varepsilon}_n^L/M_n)\}) \leq C/M_n.$$

Recall that our restriction $n^{-1/\max(d,3)} \ll \bar{\varepsilon}_n$, together with $n\bar{\varepsilon}_n^2 \to \infty$, imply that for n large enough $f_n(\cdot)$ is invertible and $f_n(\bar{\varepsilon}_n) \to 0$. On such a range, we have that $f_n^{-1}(\bar{\varepsilon}_n^L/M_n) = [(1/(c'n)) \log(M_n/\bar{\varepsilon}_n^L)]^{1/2} + c_1 n^{-1/\max(d,3)}$, which concludes the proof. $\square$

PROOF OF LEMMA C.2. Let $(x_t)_{t \in \mathbb{Z}}$ be a stochastic process with β-mixing coefficients $\beta(k)$, for $k \in \mathbb{N}$. Moreover, denote with $\mu^{(n)}$ and μ the joint distribution of a sample $x_{1:n}$ from $(x_t)_{t \in \mathbb{Z}}$ and its constant marginal $\mu = \mu^{(1)}$, respectively. Then, by combining Proposition 2 and Lemma 2 in Mohri and Rostamizadeh (2008) under our notation, we have that for any b-uniformly bounded class $\mathfrak{F}$, any integer $n \geq 1$ and any scalar $\delta \geq 0$, the inequality

$$\mathbb{P}_{x_{1:n}}\left[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{x_{1:n}}, \mu) > 2\mathfrak{R}_{\mu, n/(2K)}(\mathfrak{F}) + \delta\right] \leq 2\exp(-n\delta^2/Kb^2) + 2(n/2K - 1)\beta(K),$$

holds for every integer $K > 0$ such that $n/(2K) \in \mathbb{N}$, where $\mathfrak{R}_{\mu, n/(2K)}$ is the Rademacher complexity based an i.i.d. sample of size $n/(2K)$ from μ; see Definition 2.5. The above concentration inequality also implies

$$\begin{aligned}
&\mathbb{P}_{x_{1:n}}\left[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{x_{1:n}}, \mu) \leq 2\mathfrak{R}_{\mu, n/(2K)}(\mathfrak{F}) + \delta\right] \\
&\qquad \geq 1 - 2\exp(-n\delta^2/Kb^2) - 2(n/2K - 1)\beta(K) \\
&\qquad\qquad \geq 1 - 2\exp[-n\delta^2/(4Kb^2)] - 2(n/2K)\beta(K).
\end{aligned} \tag{D.8}$$

Notice that, in order to ensure that both $2\exp(-n\delta^2/4Kb^2)$ and $2(n/2K)\beta(K)$ vanish to zero (under Assumption (VII) for $\beta(K)$), it is tempting to apply (D.8) with $K = n^\alpha$ for some $0 < \alpha < 1$. Unfortunately, there is no reason for such a K to be an integer. A solution, would be to let $K = \lfloor n^\alpha \rfloor$, but in this case $n/(2K)$ might not be an integer. To address such issues, it is necessary to consider a careful modification of (D.8). To this end write the Euclidean division $n = n' + r$ where $n' = 2Kh \leq n$, $h = \lfloor n/(2K) \rfloor$ and $0 \leq r < 2K$. Then, under the common marginal assumption and recalling also Proposition 1 in Mohri and Rostamizadeh (2008) together with the triangle inequality and the fact that the functions f within $\mathfrak{F}$ are b-uniformly bounded, we have that

$$\begin{aligned}
\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{x_{1:n}}, \mu) &= \sup_{f \in \mathfrak{F}} \left| \frac{1}{n} \sum\nolimits_{i=1}^{n} [f(x_i) - \mathbb{E}_\mu f(x)] \right| \\
&\leq \sup_{f \in \mathfrak{F}} \left| \frac{1}{n} \sum\nolimits_{i=1}^{n'} [f(x_i) - \mathbb{E}_\mu f(x)] \right| + \frac{1}{n} \sup_{f \in \mathfrak{F}} \left| \sum\nolimits_{i=n'+1}^{n'+r} [f(x_i) - \mathbb{E}_\mu f(x)] \right| \\
&\leq \sup_{f \in \mathfrak{F}} \left| \frac{1}{n'} \sum\nolimits_{i=1}^{n'} [f(x_i) - \mathbb{E}_\mu f(x)] \right| + \frac{2br}{n} \leq \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{x_{1:n'}}, \mu) + \frac{4bK}{n},
\end{aligned}$$

where the last inequality follows directly from the definition of $\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{x_{1:n'}}, \mu)$ together with the fact that $0 \leq r < 2K$. Applying now (D.8) to $\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{x_{1:n'}}, \mu)$ yields

$$\mathbb{P}_{x_{1:n}} \left[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{x_{1:n'}}, \mu) + 4bK/n \leq 2\mathfrak{R}_{\mu,h}(\mathfrak{F}) + \delta + 4bK/n \right] \geq 1 - 2\exp(-h\delta^2/2b^2) - 2h\beta(K).$$

Therefore, since $\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{x_{1:n}}, \mu) \leq \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{x_{1:n'}}, \mu) + 4bK/n$ we also have that

$$\mathbb{P}_{x_{1:n}} [\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{x_{1:n}}, \mu) \leq 2\mathfrak{R}_{\mu,h}(\mathfrak{F}) + \delta + 4bK/n] \geq 1 - 2\exp(-h\delta^2/2b^2) - 2h\beta(K).$$

To conclude the proof, notice that to prove convergence and concentration of the ABC posterior it will be sufficient to let $K = \lfloor \sqrt{n} \rfloor$. Therefore, by replacing $K = \lfloor \sqrt{n} \rfloor$ in the above inequality and within the expression for $h = \lfloor n/(2K) \rfloor$ we have

$$\mathbb{P}_{x_{1:n}} \left[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{x_{1:n}}, \mu) \leq 2\mathfrak{R}_{\mu,s_n}(\mathfrak{F}) + \frac{4b}{\sqrt{n}} + \delta \right] \geq 1 - 2\exp(-s_n\delta^2/2b^2) - 2s_n\beta(\lfloor \sqrt{n} \rfloor)$$

where $s_n = \lfloor n/(2\lfloor \sqrt{n} \rfloor) \rfloor$ and the term $4b/\sqrt{n}$ follows directly from the fact that $\lfloor \sqrt{n} \rfloor / n \leq \sqrt{n}/n = 1/\sqrt{n}$. $\square$

PROOF OF LEMMA C.5. The proof follows the arguments used in Chapter 2 of Doukhan (1994) to study general Markov chains. In particular, when $(x_t)_{t \in \mathbb{Z}}$ is a stationary Markov chain with invariant distribution $\pi(\cdot)$ and transition kernel $P(\cdot, \cdot)$, a result proven in Davydov (1974) and recalled in page 87–88 of Doukhan (1994) gives $\beta(k) = \mathbb{E}_{x \sim \pi} \| P^k(x, \cdot) - \pi(\cdot) \|_{\mathrm{TV}}$. When $-1 < \theta < 1$, standard results for the AR(1) model in Lemma C.5 lead to the invariant distribution $\pi = \mathrm{N}(\psi/(1-\theta), \sigma^2/(1-\theta^2))$.

As for $P^k(x, \cdot)$, notice that, under such an AR(1) model with starting point x, we can write

$$\begin{aligned}
x_k &= \theta x_{k-1} + \psi + \varepsilon_k = \theta(\theta x_{k-2} + \psi + \varepsilon_{k-1}) + \psi + \varepsilon_k = \ldots \\
&= \theta^k x + \psi \sum\nolimits_{l=0}^{k-1} \theta^l + \sum\nolimits_{l=0}^{k-1} \theta^l \varepsilon_{k-l}.
\end{aligned}$$

Therefore, $P^k(x, \cdot) = \mathrm{N}(\theta^k x + \psi \sum_{l=0}^{k-1} \theta^l, \sigma^2 \sum_{l=0}^{k-1} \theta^{2l})$.

Moreover, notice that, by direct application of standard properties of finite power series, we have $\sum_{l=0}^{k-1}\theta^l=(1-\theta^k)/(1-\theta)$ and $\sum_{l=0}^{k-1}\theta^{2l}=(1-\theta^{2k})/(1-\theta^2)$. Since our goal is to derive an upper bound for $\beta(k)$ and provided that the KL divergence among Gaussian densities is available in closed form, let us first consider the Pinsker's inequality

$$\|P^k(x,\cdot)-\pi(\cdot)\|_{\mathrm{TV}}\leq[\mathcal{D}_{\mathrm{KL}}(P^k(x,\cdot),\pi(\cdot))/2]^{1/2}$$

where $\mathcal{D}_{\mathrm{KL}}$ stands for the KL divergence. Since both $P^k(x,\cdot)$ and $\pi(\cdot)$ are Gaussian, then

$$\begin{aligned}
&\mathcal{D}_{\mathrm{KL}}(P^k(x,\cdot),\pi(\cdot))\\
&=\frac{1}{2}\left[\frac{\sigma^2(1-\theta^{2k})}{\sigma^2}-1+\frac{[\theta^k x+\psi(1-\theta^k)/(1-\theta)-\psi/(1-\theta)]^2}{\sigma^2/(1-\theta^2)}+\log\frac{\sigma^2}{\sigma^2(1-\theta^{2k})}\right]\\
&=\frac{1}{2}\left[(1-\theta^{2k})-1+\frac{\theta^{2k}[x-\psi/(1-\theta)]^2}{\sigma^2/(1-\theta^2)}+\log\frac{1}{1-\theta^{2k}}\right]\\
&=\frac{1}{2}\left[-\theta^{2k}+\frac{\theta^{2k}[x-\psi/(1-\theta)]^2}{\sigma^2/(1-\theta^2)}+\log\left(1+\frac{\theta^{2k}}{1-\theta^{2k}}\right)\right]\\
&\leq\frac{1}{2}\left[-\theta^{2k}+\frac{\theta^{2k}[x-\psi/(1-\theta)]^2}{\sigma^2/(1-\theta^2)}+\frac{\theta^{2k}}{1-\theta^{2k}}\right]\leq\frac{\theta^{2k}}{2}\left[-1+\frac{[x-\psi/(1-\theta)]^2}{\sigma^2/(1-\theta^2)}+\frac{1}{1-\theta^2}\right].
\end{aligned}$$

Therefore, by leveraging the above result, together with standard properties of the expectation, we have

$$\begin{aligned}
\beta(k)&\leq\mathbb{E}_{x\sim\pi}[\mathcal{D}_{\mathrm{KL}}(P^k(x,\cdot),\pi(\cdot))/2]^{1/2}\leq[\mathbb{E}_{x\sim\pi}\mathcal{D}_{\mathrm{KL}}(P^k(x,\cdot),\pi(\cdot))/2]^{1/2}\\
&\leq\left[\frac{\theta^{2k}}{4}\left[-1+\frac{\mathbb{E}_{x\sim\pi}[x-\psi/(1-\theta)]^2}{\sigma^2/(1-\theta^2)}+\frac{1}{1-\theta^2}\right]\right]^{1/2}=\sqrt{\frac{\theta^{2k}}{4(1-\theta^2)}}=\frac{|\theta|^k}{2\sqrt{1-\theta^2}},
\end{aligned}$$

which concludes the proof. □

PROOF OF PROPOSITION C.6. Under Assumption (VII), for any fixed $\delta>0$, we have that $\sum_{n>0}[2\exp(-s_n\delta^2/2b^2)+2s_nC_\beta\exp(-\gamma\lfloor\sqrt{n}\rfloor^\xi)]<\infty$. Therefore, combining Lemma C.2 with Assumption (IV), both $\mathcal{D}_{\mathfrak{F}}(\hat\mu_{z_{1:n}},\mu_\theta)$ and $\mathcal{D}_{\mathfrak{F}}(\hat\mu_{y_{1:n}},\mu^*)$ converge to 0 almost surely as $n\to\infty$, by the Borel–Cantelli Lemma. As a result, since

$$-\mathcal{D}_{\mathfrak{F}}(\hat\mu_{z_{1:n}},\mu_\theta)-\mathcal{D}_{\mathfrak{F}}(\hat\mu_{y_{1:n}},\mu^*)\leq\mathcal{D}_{\mathfrak{F}}(\mu_\theta,\mu^*)-\mathcal{D}_{\mathfrak{F}}(\hat\mu_{y_{1:n}},\hat\mu_{z_{1:n}})\leq\mathcal{D}_{\mathfrak{F}}(\hat\mu_{z_{1:n}},\mu_\theta)+\mathcal{D}_{\mathfrak{F}}(\hat\mu_{y_{1:n}},\mu^*),$$

it holds that $\mathcal{D}_{\mathfrak{F}}(\hat\mu_{z_{1:n}},\hat\mu_{y_{1:n}})\to\mathcal{D}_{\mathfrak{F}}(\mu_\theta,\mu^*)$ almost surely as $n\to\infty$. To conclude it suffices to apply again the proof of Theorem 1 in Jiang, Wu and Wong (2018); see also the proof of Corollary 3.2. □

PROOF OF THEOREM C.7. To prove Theorem C.7 we will follow the same line of reasoning as in the proof of Theorem 3.3. However, in this case we leverage Lemma C.2 instead of Lemma 2.6. To this end, letting $\delta=c_1-2\mathfrak{R}_{s_n}(\mathfrak{F})-4b/\sqrt{n}$ with $c_1\geq2\mathfrak{R}_{s_n}(\mathfrak{F})+4b/\sqrt{n}$

and $\mathfrak{R}_{s_n} = \sup_{\mu \in \mathcal{P}(\mathcal{Y})} \mathfrak{R}_{\mu,s_n}$, we obtain, under Assumption (VII), Equation (D.9) below, instead of (D.3).

$$\text{(D.9)}\quad \mathbb{P}_{x_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{x_{1:n}}, \mu) \le c_1] \\ \ge 1 - 2\exp[-s_n(c_1 - 2\mathfrak{R}_{s_n}(\mathfrak{F}) - 4b/\sqrt{n})^2/2b^2] - 2s_n C_\beta \exp(-\gamma\lfloor\sqrt{n}\rfloor^{\xi}).$$

As in Theorem 3.3, let $c_1 = \bar{\varepsilon}_n/3$ and notice that, by the settings of Theorem C.7, for n large enough $\bar{\varepsilon}_n/3 > 2\mathfrak{R}_{s_n}(\mathfrak{F}) + 4b/\sqrt{n}$. Therefore, applying (D.9) to $y_{1:n}$, with $c_1 = \bar{\varepsilon}_n/3$, leads to the following upper bound

$$\mathbb{P}_{y_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \mu^*) \le \bar{\varepsilon}_n/3] \\ \ge 1 - 2\exp[-s_n(\bar{\varepsilon}_n/3 - 2\mathfrak{R}_{s_n}(\mathfrak{F}) - 4b/\sqrt{n})^2/2b^2] - 2s_n C_\beta \exp(-\gamma\lfloor\sqrt{n}\rfloor^{\xi}).$$

Recall that, under the settings of Theorem C.7, we have that $\sqrt{n}\bar{\varepsilon}_n^2 \to \infty$ and $\bar{\varepsilon}_n/\mathfrak{R}_{s_n}(\mathfrak{F}) \to \infty$ and, therefore,

$$s_n(\bar{\varepsilon}_n/3 - 2\mathfrak{R}_{s_n}(\mathfrak{F}) - 4b/\sqrt{n})^2 \sim s_n\bar{\varepsilon}_n^2/9 \sim \sqrt{n}\bar{\varepsilon}_n^2 \to \infty.$$

Combining this result with the fact that $2s_n C_\beta \exp(-\gamma\lfloor\sqrt{n}\rfloor^{\xi}) \to 0$ as $n \to \infty$, it follows that the lower bound for $\mathbb{P}_{y_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \mu^*) \le \bar{\varepsilon}_n/3]$ goes to 1 as $n \to \infty$. Hence, in the remaining part of the proof, we will restrict to the event $\{\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \mu^*) \le \bar{\varepsilon}_n/3\}$.

Let $\mathbb{P}_{\theta,z_{1:n}}$ corresponds to the joint distribution of $\theta \sim \pi$ and $z_{1:n}$ from $\mu_\theta^{(n)}$. Then, as a direct consequence of the definition of conditional probability, for every positive c_2, including $c_2 > 2\mathfrak{R}_{s_n}(\mathfrak{F}) + 4b/\sqrt{n}$, it follows that

$$\text{(D.10)}\quad \pi_n^{(\varepsilon^*+\bar{\varepsilon}_n)}(\{\theta : \mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) > \varepsilon^* + 4\bar{\varepsilon}_n/3 + c_2\}) \\ = \frac{\mathbb{P}_{\theta,z_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) > \varepsilon^* + 4\bar{\varepsilon}_n/3 + c_2, \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \hat{\mu}_{y_{1:n}}) \le \varepsilon^* + \bar{\varepsilon}_n]}{\mathbb{P}_{\theta,z_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \hat{\mu}_{y_{1:n}}) \le \varepsilon^* + \bar{\varepsilon}_n]}.$$

To upper bound the ratio in (D.10), let us first derive an upper bound for the numerator. To this end, consider the triangle inequality $\mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) \le \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \mu_\theta) + \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \hat{\mu}_{y_{1:n}}) + \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \mu^*)$ (recall that $\mathcal{D}_{\mathfrak{F}}$ is a semimetric), along with the previously-proved result that the event $\{\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \mu^*) \le \bar{\varepsilon}_n/3\}$ has $\mathbb{P}_{y_{1:n}}$–probability going to 1. Hence, for n large enough we have

$$\mathbb{P}_{\theta,z_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) > \varepsilon^* + 4\bar{\varepsilon}_n/3 + c_2, \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \hat{\mu}_{y_{1:n}}) \le \varepsilon^* + \bar{\varepsilon}_n] \\ \le \mathbb{P}_{\theta,z_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \hat{\mu}_{y_{1:n}}) + \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \mu_\theta) + \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \mu^*) > \varepsilon^* + 4\bar{\varepsilon}_n/3 + c_2, \\ \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \hat{\mu}_{y_{1:n}}) \le \varepsilon^* + \bar{\varepsilon}_n] \\ \le \mathbb{P}_{\theta,z_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \mu_\theta) + \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}}, \mu^*) > \bar{\varepsilon}_n/3 + c_2] \le \mathbb{P}_{\theta,z_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \mu_\theta) > c_2],$$

where $\mathbb{P}_{\theta,z_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \mu_\theta) > c_2] = \int_{\theta\in\Theta} \mathbb{P}_{z_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}}, \mu_\theta) > c_2 \mid \theta]\, \pi(\mathrm{d}\theta)$.

Therefore, leveraging the above result and applying (D.9) to $z_{1:n}$ yields,

$$\int_{\theta\in\Theta} \mathbb{P}_{z_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}},\mu_\theta) > c_2 \mid \theta]\pi(\mathrm{d}\theta) = \int_{\theta\in\Theta}(1-\mathbb{P}_{z_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}},\mu_\theta) \leq c_2 \mid \theta])\pi(\mathrm{d}\theta)$$
$$\leq 2\exp[-s_n(c_2 - 2\mathfrak{R}_{s_n}(\mathfrak{F}) - 4b/\sqrt{n})^2/2b^2] + 2s_n C_\beta \exp(-\gamma\lfloor\sqrt{n}\rfloor^\xi).$$

This controls the numerator in (D.10). As for the denominator, defining the event $\mathrm{E}_n := \{\theta\in\Theta : \mathcal{D}_{\mathfrak{F}}(\mu_\theta,\mu^*) \leq \varepsilon^* + \bar{\varepsilon}_n/3\}$ and applying again the triangle inequality, we have that

$$\mathbb{P}_{\theta,z_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}},\hat{\mu}_{y_{1:n}}) \leq \varepsilon^* + \bar{\varepsilon}_n] \geq \int_{\mathrm{E}_n} \mathbb{P}_{z_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}},\hat{\mu}_{y_{1:n}}) \leq \varepsilon^* + \bar{\varepsilon}_n \mid \theta]\,\pi(\mathrm{d}\theta)$$
$$\geq \int_{\mathrm{E}_n} \mathbb{P}_{z_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}},\mu^*) + \mathcal{D}_{\mathfrak{F}}(\mu_\theta,\mu^*) + \mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}},\mu_\theta) \leq \varepsilon^* + \bar{\varepsilon}_n \mid \theta]\,\pi(\mathrm{d}\theta)$$
$$\geq \int_{\mathrm{E}_n} \mathbb{P}_{z_{1:n}}[\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{z_{1:n}},\mu_\theta) \leq \bar{\varepsilon}_n/3 \mid \theta]\,\pi(\mathrm{d}\theta).$$

The last inequality follows from that fact that we can restrict to the event $\{\mathcal{D}_{\mathfrak{F}}(\hat{\mu}_{y_{1:n}},\mu^*) \leq \bar{\varepsilon}_n/3\}$, and that we are integrating over $\mathrm{E}_n := \{\theta\in\Theta : \mathcal{D}_{\mathfrak{F}}(\mu_\theta,\mu^*) \leq \varepsilon^* + \bar{\varepsilon}_n/3\}$. Let us now apply again (D.9) to $z_{1:n}$, with $c_1 = \bar{\varepsilon}_n/3$, to further lower bound the last term of the above inequality by

$$\int_{\mathrm{E}_n} [1 - 2\exp[-s_n(\bar{\varepsilon}_n/3 - 2\mathfrak{R}_{s_n}(\mathfrak{F}) - 4b/\sqrt{n})^2/2b^2] - 2s_n C_\beta \exp(-\gamma\lfloor\sqrt{n}\rfloor^\xi)]\pi(\mathrm{d}\theta)$$
$$= \pi(\mathrm{E}_n)[1 - 2\exp[-s_n(\bar{\varepsilon}_n/3 - 2\mathfrak{R}_{s_n}(\mathfrak{F}) - 4b/\sqrt{n})^2/2b^2] - 2s_n C_\beta \exp(-\gamma\lfloor\sqrt{n}\rfloor^\xi)].$$

Note that, by Assumption (II), $\pi(\mathrm{E}_n) \geq c_\pi(\bar{\varepsilon}_n/3)^L$. Moreover, as shown before, the quantity $1 - 2\exp[-s_n(\bar{\varepsilon}_n/3 - 2\mathfrak{R}_{s_n}(\mathfrak{F}) - 4b/\sqrt{n})^2/2b^2] - 2s_n C_\beta \exp(-\gamma\lfloor\sqrt{n}\rfloor^\xi)$ goes to 1 when $n\to\infty$, which also implies that, for a large enough n,

$$1 - 2\exp[-s_n(\bar{\varepsilon}_n/3 - 2\mathfrak{R}_{s_n}(\mathfrak{F}) - 4b/\sqrt{n})^2/2b^2] - 2s_n C_\beta \exp(-\gamma\lfloor\sqrt{n}\rfloor^\xi) > 1/2.$$

Therefore, leveraging both results, the denominator in (D.10) can be lower bounded by the term $(c_\pi/2)\,(\bar{\varepsilon}_n/3)^L$.

To proceed with the proof, let us combine the upper and lower bounds derived, respectively, for the numerator and the denominator of the ratio in (D.10). This yields, for any integer K,

$$\text{(D.11)}\quad \pi_n^{(\varepsilon^*+\bar{\varepsilon}_n)}(\{\theta : \mathcal{D}_{\mathfrak{F}}(\mu_\theta,\mu^*) > \varepsilon^* + 4\bar{\varepsilon}_n/3 + c_2\})$$
$$\leq \frac{2\exp[-s_n(c_2 - 2\mathfrak{R}_{s_n}(\mathfrak{F}) - 4b/\sqrt{n})^2/2b^2] + 2s_n C_\beta \exp(-\gamma\lfloor\sqrt{n}\rfloor^\xi)}{(c_\pi/2)(\bar{\varepsilon}_n/3)^L},$$

with $\mathbb{P}_{y_{1:n}}$–probability going to 1 as $n\to\infty$.

Replacing c_2 in (D.11) with $2\mathfrak{R}_{s_n}(\mathfrak{F}) + \sqrt{(2b^2/s_n)\log(n/\bar{\varepsilon}_n^L)} + 4b/\sqrt{n}$, gives

$$\pi_n^{(\varepsilon^*+\bar{\varepsilon}_n)}\Big(\Big\{\theta : \mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) > \varepsilon^* + \frac{4\bar{\varepsilon}_n}{3} + 2\mathfrak{R}_{s_n}(\mathfrak{F}) + \frac{4b}{\sqrt{n}} + \Big(\frac{2b^2}{s_n}\log\frac{n}{\bar{\varepsilon}_n^L}\Big)^{1/2}\Big\}\Big) \leq \frac{4\cdot 3^L}{nc_\pi}\Big(1 + \frac{ns_nC_\beta\exp(-\gamma\lfloor\sqrt{n}\rfloor^\xi)}{\bar{\varepsilon}_n^L}\Big).$$

To conclude, note that

$$ns_nC_\beta\exp(-\gamma\lfloor\sqrt{n}\rfloor^\xi)/\bar{\varepsilon}_n^L = n^{1+L/2}s_nC_\beta\exp(-\gamma\lfloor\sqrt{n}\rfloor^\xi)/((n\bar{\varepsilon}_n^2)^{L/2}),$$

where the numerator goes to 0 and the denominator goes to ∞ when $n \to \infty$, under the setting of Theorem C.7. Therefore, with $\mathbb{P}_{y_{1:n}}$–probability going to 1 as $n \to \infty$, we have

$$\pi_n^{(\varepsilon^*+\bar{\varepsilon}_n)}\Big(\Big\{\theta : \mathcal{D}_{\mathfrak{F}}(\mu_\theta, \mu^*) > \varepsilon^* + \frac{4\bar{\varepsilon}_n}{3} + 2\mathfrak{R}_{s_n}(\mathfrak{F}) + \frac{4b}{\sqrt{n}} + \Big(\frac{2b^2}{s_n}\log\frac{n}{\bar{\varepsilon}_n^L}\Big)^{1/2}\Big\}\Big) \leq \frac{4\cdot 3^L}{nc_\pi},$$

concluding the proof. □

REFERENCES

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

CHÉRIEF-ABDELLATIF, B.-E. and ALQUIER, P. (2022). Finite sample properties of parametric MMD estimation: Robustness to misspecification and dependence. *Bernoulli* **28** 181 – 213.

DAVYDOV, Y. A. (1974). Mixing conditions for Markov chains. *Theory of Probability & Its Applications* **18** 312–328.

DOUKHAN, P. (1994). *Mixing: Properties and Examples* **85**. Springer Science & Business Media.

DROVANDI, C. and FRAZIER, D. T. (2022). A comparison of likelihood-free methods with and without summary statistics. *Statistics and Computing* **32** 1–23.

FEARNHEAD, P. and PRANGLE, D. (2012). Constructing summary statistics for approximate Bayesian computation: Semi-automatic approximate Bayesian computation. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **74** 419–474.

FRAZIER, D. T. (2020). Robust and efficient approximate Bayesian computation: A minimum distance approach. *arXiv preprint arXiv:2006.14126*.

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

FUJISAWA, M., TESHIMA, T., SATO, I. and SUGIYAMA, M. (2021). γ-ABC: Outlier-robust approximate Bayesian computation based on a robust divergence estimator. In *International Conference on Artificial Intelligence and Statistics* 1783–1791. PMLR.

JIANG, B., WU, T.-Y. and WONG, W. H. (2018). Approximate Bayesian computation with Kullback-Leibler divergence as data discrepancy. In *International Conference on Artificial Intelligence and Statistics* 1711–1721. PMLR.

LEI, J. (2020). Convergence and concentration of empirical measures under Wasserstein distance in unbounded functional spaces. *Bernoulli* **26** 767–798.

MASSART, P. (2000). Some applications of concentration inequalities to statistics. In *Annales de la Faculté des sciences de Toulouse: Mathématiques* **9** 245–303.

MOHRI, M. and ROSTAMIZADEH, A. (2008). Rademacher complexity bounds for non-i.i.d. processes. In *Advances in Neural Information Processing Systems* **21** 1097–1104.

MUANDET, K., FUKUMIZU, K., SRIPERUMBUDUR, B. and SCHÖLKOPF, B. (2017). Kernel mean embedding of distributions: A review and beyond. *Foundations and Trends in Machine Learning* **10** 1–141.

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

WAINWRIGHT, M. J. (2019). *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge University Press.