MONASH BUSINESS SCHOOL

ISSN 1440-771X

Department of Econometrics and Business Statistics

Asymptotic Properties of Approximate Bayesian Computation

D.T. Frazier, G.M. Martin, C.P. Robert and J. Rousseau

October 2016

Working Paper 18/16

Asymptotic Properties of Approximate Bayesian Computation

By D.T. FRAZIER, G.M. MARTIN
Department of Econometrics and Business Statistics, Monash University, Melbourne, Australia
david.frazier,gael.martin@monash.edu

C.P. ROBERT AND J. ROUSSEAU
Université Paris Dauphine PSL, CEREMADE CNRS. Paris, France
rousseau,xian@ceremade.dauphine.fr

SUMMARY

Approximate Bayesian computation (ABC) is becoming an accepted tool for statistical analysis in models with intractable likelihoods. With the initial focus being primarily on the practical import of ABC, exploration of its formal statistical properties has begun to attract more attention. In this paper we consider the asymptotic behaviour of the posterior obtained from ABC and the ensuing posterior mean. We give general results on: (i) the rate of concentration of the ABC posterior on sets containing the true parameter (vector); (ii) the limiting shape of the posterior; and (iii) the asymptotic distribution of the ABC posterior mean. These results hold under given rates for the tolerance used within ABC, mild regularity conditions on the summary statistics, and a condition linked to identification of the true parameters. Important implications of the theoretical results for practitioners of ABC are highlighted.

Some key words: asymptotic properties, Bayesian inference, Bernstein-von Mises theorem, consistency, likelihood-free methods

MSC2010 Subject Classification: 62F15, 62F12, 62C10

JEL Classification: C11, C15, C18

1. INTRODUCTION

The use of approximate Bayesian computation (ABC) methods in models with intractable likelihoods has gained increased momentum over recent years, extending beyond the original applications in the biological sciences. (See Marin et al., 2011, Sisson and Fan, 2011 and Robert, 2015, for recent reviews.) Whilst ABC evolved initially as a practical tool, attention has begun to shift to the investigation of its formal statistical properties, including as they relate to the choice of summary statistics on which the technique typically relies; see, for example, Fearnhead and Prangle (2012), the 2013 technical report of Gleim and Pigorsch (University of Bonn), Marin et al. (2014), Creel and Kristensen (2015), Drovandi et al. (2015), the 2015 preprint of Creel et al. (arxiv:1512.07385) and the 2016 preprints of Li and Fearnhead (arxiv:1506.03481) and Martin et al. (arxiv:1604.07949). Hereafter we will denote these preprints by Gleim and Pigorsch (2014), Creel et al. (2015), Li and Fearnhead (2016) and Martin et al. (2016).

In this paper we study the large sample properties of both posterior distributions and posterior means obtained from approximate Bayesian computation algorithms. Under mild regularity conditions on the summary statistics used in such algorithms, we characterize the rate of posterior concentration and show that the limiting shape of the posterior crucially depends on the interplay between the rate at which the summaries converge (in distribution) and the rate at which the tolerance used to accept parameter draws shrinks to zero. Critically, concentration around the truth and, hence, Bayesian consistency, places a less stringent condition on the speed with which the tolerance declines to zero than does asymptotic normality of the resulting posterior. Further, and in contrast to the textbook Bernstein-von Mises result, we show that asymptotic normality of the posterior mean does not necessarily require asymptotic normality of the posterior itself, with the former result being attainable under weaker conditions on the tolerance than required for the latter. Validity of all of these results requires that the simulated summaries converge toward some limiting counterpart at some known rate, and that this limit counterpart, viewed as a mapping from parameters to simulated summaries, be injective. These conditions have a close correspondence with those required for theoretical validity of indirect inference and related (frequentist) estimators (Gourieroux et al., 1993; Gallant and Tauchen, 1996; Heggland and Frigessi, 2004).

Our focus on three aspects of asymptotic behaviour, posterior consistency, limiting posterior shape, and the asymptotic distribution of the posterior mean, is much broader than that of existing studies on the large sample properties of approximate Bayesian computation algorithms, in which the asymptotic properties of point estimators derived from these algorithms have been the primary focus; see Creel et al. (2015), Jarsa (2015) and Li and Fearnhead (2016). Our approach allows for weaker conditions than those given in the aforementioned papers, permits a complete characterization of the limiting shape of the posterior, and distinguishes between the conditions (on both the summary statistics and the tolerance) required for concentration and the conditions required for specific distributional results. Throughout the paper the posterior distribution that is referred to is the posterior distribution resulting from an approximate Bayesian computation algorithm.

2. CONVERGENCE OF THE APPROXIMATE BAYESIAN COMPUTATION POSTERIOR

2·1. *Preliminaries and Background*

We observe data $\boldsymbol{y} = (y_1, y_2, ..., y_T)'$, $T \geq 1$, drawn from the model $\{P_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$, where $\boldsymbol{\theta} \mapsto P_{\boldsymbol{\theta}}$ admits the corresponding conditional density $p(\cdot|\boldsymbol{\theta})$, and $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^{d_\theta}$. Given a prior $p(\boldsymbol{\theta})$, the aim of the algorithms under study is to produce draws from an approximation to the posterior distribution

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}),$$

in the case where both the parameters and pseudo-data $(\boldsymbol{\theta}, \boldsymbol{z})$ can be easily simulated from $p(\boldsymbol{\theta})p(\boldsymbol{z}|\boldsymbol{\theta})$, but where $p(\boldsymbol{z}|\boldsymbol{\theta})$ is intractable. The simplest (accept/reject) form of the algorithm (Tavaré et al., 1997; Pritchard et al., 1999) is detailed in Algorithm 1.

Algorithm 1 Approximate Bayesian Computation algorithm

Simulate $\boldsymbol{\theta}^i$, $i = 1, 2, ..., N$, from $p(\boldsymbol{\theta})$,
Simulate $\boldsymbol{z}^i = (z^i_1, z^i_2, ..., z^i_T)'$, $i = 1, 2, ..., N$, from the likelihood, $p(\cdot|\boldsymbol{\theta}^i)$
Select $\boldsymbol{\theta}^i$ such that:

$$d\{\boldsymbol{\eta}(\boldsymbol{y}), \boldsymbol{\eta}(\boldsymbol{z}^i)\} \leq \varepsilon,$$

where $\boldsymbol{\eta}(\cdot)$ is a (vector) statistic, $d\{\cdot, \cdot\}$ is a distance function (or metric), and the tolerance level ε is chosen to be small.

Algorithm 1 thus samples $\boldsymbol{\theta}$ and $\boldsymbol{z}$ from the joint posterior:

$$p_\varepsilon\{\boldsymbol{\theta}, \boldsymbol{z}|\boldsymbol{\eta}(\boldsymbol{y})\} = \frac{p(\boldsymbol{\theta})p(\boldsymbol{z}|\boldsymbol{\theta})\mathbb{1}_\varepsilon[\boldsymbol{z}]}{\int_{\boldsymbol{\Theta}} \int_{\boldsymbol{z}} p(\boldsymbol{\theta})p(\boldsymbol{z}|\boldsymbol{\theta})\mathbb{1}_\varepsilon[\boldsymbol{z}]d\boldsymbol{z}d\boldsymbol{\theta}},$$

where $\mathbb{1}_\varepsilon[\boldsymbol{z}] = \mathbb{1}[d\{\boldsymbol{\eta}(\boldsymbol{y}), \boldsymbol{\eta}(\boldsymbol{z})\} \leq \varepsilon]$ is one if $d\{\boldsymbol{\eta}(\boldsymbol{y}), \boldsymbol{\eta}(\boldsymbol{z})\} \leq \varepsilon$ and zero else. Clearly, when $\boldsymbol{\eta}(\cdot)$ is sufficient and ε arbitrarily small,

$$p_\varepsilon\{\boldsymbol{\theta}|\boldsymbol{\eta}(\boldsymbol{y})\} = \int_{\boldsymbol{z}} p_\varepsilon\{\boldsymbol{\theta}, \boldsymbol{z}|\boldsymbol{\eta}(\boldsymbol{y})\}d\boldsymbol{z} \quad (1)$$

approximates the exact posterior, $p(\boldsymbol{\theta}|\boldsymbol{y})$, and draws of $\boldsymbol{\theta}$ from $p_\varepsilon\{\boldsymbol{\theta},\boldsymbol{z}|\boldsymbol{\eta}(\boldsymbol{y})\}$ can be used to estimate features of the true posterior. In practice however, the complexity of the models to which Algorithm 1 is applied implies, almost by definition, that sufficiency is unattainable. Hence, in the limit, as $\varepsilon \to 0$, the draws can be used only to approximate features of $p\{\boldsymbol{\theta}|\boldsymbol{\eta}(\boldsymbol{y})\}$. For a given total number of draws, N, reducing ε comes at a cost of reducing the probability of a draw being accepted and, hence, increasing the error associated with estimating $p\{\boldsymbol{\theta}|\boldsymbol{\eta}(\boldsymbol{y})\}$ using the accepted draws. The problem is exacerbated the larger is the dimension of $\boldsymbol{\eta}(\boldsymbol{y})$; see Blum (2010), Blum et al. (2013), Nott et al. (2014) and Biau et al. (2015). In practice ε tends to be chosen such that a certain (small) proportion of draws (from the total N) are selected, with attempts made to reduce the (kernel-based) estimation error via various post-sampling corrections (Beaumont et al., 2002; Blum, 2010; Blum and François, 2010). Other work gives emphasis to producing draws of $\boldsymbol{\theta}$ in such a way that $p\{\boldsymbol{\theta}|\boldsymbol{\eta}(\boldsymbol{y})\}$ is a better match to $p(\boldsymbol{\theta}|\boldsymbol{y})$ (Marjoram et al., 2003; Sisson et al., 2007; Beaumont et al., 2009; Toni et al., 2009; Wegmann et al., 2009); or to selecting a vector $\boldsymbol{\eta}(\cdot)$ that is informative in some well-defined sense (Joyce and Marjoram, 2008; Blum, 2010; Drovandi et al., 2011; Fearnhead and Prangle, 2012; Gleim and Pigorsch, 2013; Creel and Kristensen, 2015; Creel et al., 2015; Drovandi et al., 2015; Martin et al., 2016.)

The focus of this paper is on the asymptotic behaviour of $p_\varepsilon\{\boldsymbol{\theta}|\boldsymbol{\eta}(\boldsymbol{y})\}$ and related quantities, with results in this and all subsequent sections being founded on *simultaneous* asymptotic arguments related to T and ε.

2·2. *Main Result*

First, let us set some notation used throughout the remainder of the paper. Define $\mathcal{Z}$ as the space of simulated data; $\mathcal{B}=\{\boldsymbol{\eta}(\boldsymbol{z}):\boldsymbol{z}\in\mathcal{Z}\}$ the range of the simulated summaries - i.e. $\boldsymbol{\eta}(\boldsymbol{z}):\mathcal{Z}\to\mathcal{B}$; $d_1\{\cdot,\cdot\}$ a metric on $\boldsymbol{\Theta}$; $d_2\{\cdot,\cdot\}$ a metric on $\mathcal{B}$; $C>0$ a generic constant; P_0 the measure generating $\boldsymbol{y}$; and $P_{\boldsymbol{\theta}}$ the measure generating $\boldsymbol{z}(\boldsymbol{\theta})$. We have $P_{\boldsymbol{\theta}}=P_0$ for $\boldsymbol{\theta}=\boldsymbol{\theta}_0$, and denote $\boldsymbol{\theta}_0\in\text{Int}(\boldsymbol{\Theta})$ as the true parameter value. Let $\Pi(\boldsymbol{\theta})$ denote the prior measure with density $p(\boldsymbol{\theta})$.

For sequences $\{a_T\}$ and $\{b_T\}$, real valued, $a_T\lesssim b_T$ denotes $a_T\le Cb_T$, $a_T\asymp b_T$ denotes equivalent order of magnitude, $a_T\gg b_T$ indicates a larger order of magnitude and the symbols $o_P(a_T), O_P(b_T)$ have their usual meaning. We use $\|\cdot\|$ to denote the Euclidean norm, $\|\cdot\|_*$ a matrix norm and $|\cdot|$ the absolute value.

Unlike the standard notion of posterior concentration, analysing convergence of posteriors derived from Algorithm 1 requires the use of simultaneous asymptotics since $p_\varepsilon\{\cdot|\boldsymbol{\eta}(\boldsymbol{y})\}$ in (1) depends both on T, through $\boldsymbol{\eta}(\boldsymbol{y})$, and ε, via the selection criterion $d\{\cdot,\cdot\}\le\varepsilon$. In this way, for $A\subset\boldsymbol{\Theta}$, the posterior probability of A is most appropriately expressed as

$$\Pi\left(A|d_2\{\boldsymbol{\eta}(\boldsymbol{y}),\boldsymbol{\eta}(\boldsymbol{z})\}\le\varepsilon\right)=\Pi(A|\{\boldsymbol{\theta}:d_2\{\boldsymbol{\eta}(\boldsymbol{y}),\boldsymbol{\eta}(\boldsymbol{z})\}\le\varepsilon\})\equiv\int_A p_\varepsilon\{\boldsymbol{\theta}|\boldsymbol{\eta}(\boldsymbol{y})\}d\boldsymbol{\theta}.$$

With this notation, we say that the algorithm is Bayesian consistent (or posterior consistent) at $\boldsymbol{\theta}_0$ if for any $\delta>0$,

$$\Pi\left(d_1\{\boldsymbol{\theta},\boldsymbol{\theta}_0\}>\delta|d_2\{\boldsymbol{\eta}(\boldsymbol{y}),\boldsymbol{\eta}(\boldsymbol{z})\}\le\varepsilon\right)\to 0 \tag{2}$$

(in P_0-probability) as $T\to+\infty,\varepsilon\to0$. Bayesian consistency implies that sets containing $\boldsymbol{\theta}_0$ have posterior probability tending to one as $T\to+\infty$, with an implication of this result being the existence of a specific rate at which posterior probability concentrates on sets containing $\boldsymbol{\theta}_0$. We say that $\Pi\left(\cdot|d_2\{\boldsymbol{\eta}(\boldsymbol{y}),\boldsymbol{\eta}(\boldsymbol{z})\}\le\varepsilon\right)$ concentrates at rate $\lambda_T\to0$ if

$$\limsup_{\varepsilon\to0}\limsup_{T\to+\infty}\Pi\left(d_1\{\boldsymbol{\theta},\boldsymbol{\theta}_0\}>\lambda_T M|d_2\{\boldsymbol{\eta}(\boldsymbol{y}),\boldsymbol{\eta}(\boldsymbol{z})\}\le\varepsilon\right)\to0 \tag{3}$$

in P_0-probability when M goes to infinity.

In general terms, the posterior rate of concentration is related to the rate at which information about the true parameter vector accumulates in the sample. In such algorithms, information about $\boldsymbol{\theta}_0$ is not obtained from the likelihood, assumed intractable, but through $\boldsymbol{\eta}(\boldsymbol{y})$ and, in particular, through draws from $p(\boldsymbol{\theta})$ and $\boldsymbol{z}\sim P_{\boldsymbol{\theta}}$ that satisfy $d_2\{\boldsymbol{\eta}(\boldsymbol{y}),\boldsymbol{\eta}(\boldsymbol{z})\}\le\varepsilon$. In particular, when working with a fixed dimension for the summary statistic (vector) $\boldsymbol{\eta}(\boldsymbol{y})$, the posterior distribution learns (and concentrates) about $\boldsymbol{\theta}_0$ only if $\boldsymbol{\eta}(\boldsymbol{z})$ concentrates around some fixed value $\boldsymbol{b}(\boldsymbol{\theta})$ depending on $\boldsymbol{\theta}$. Therefore, the amount of information Algorithm 1 obtains about $\boldsymbol{\theta}_0$ depends on two factors: (a) the rate at which the observed (resp. simulated) summaries converge to a well-defined limit counterpart $\boldsymbol{b}(\boldsymbol{\theta}_0)$ (resp., $\boldsymbol{b}(\boldsymbol{\theta})$); (b) the rate at which ε goes to 0. The map $\boldsymbol{b} : \boldsymbol{\Theta} \to \mathcal{B}$ is non-random. To link both factors we consider ε as a T-dependent sequence $\varepsilon_T \to 0$ as $T \to +\infty$. Hence, by connecting the tolerance $\varepsilon_T = \varepsilon$ and T, we can now state the technical assumptions necessary for establishing the first main result of the paper.

[A1] The map $\boldsymbol{b} : \boldsymbol{\Theta} \to \mathcal{B}$ satisfies, for $\boldsymbol{y} \sim P_0$, $d_2\{\boldsymbol{\eta}(\boldsymbol{y}), \boldsymbol{b}(\boldsymbol{\theta}_0)\} = o_P(1)$.

[A2] There exists $\rho_T(u) \to 0$ as $T \to +\infty$ for all u, and $\rho_T(u)$ monotone non-increasing in u (for any given T), such that

$$P_{\boldsymbol{\theta}}\left(d_2\{\boldsymbol{\eta}(\boldsymbol{z}), \boldsymbol{b}(\boldsymbol{\theta})\} > u\right) \leq c(\boldsymbol{\theta})\rho_T(u), \quad \int_\Theta c(\boldsymbol{\theta}) d\Pi(\boldsymbol{\theta}) < +\infty$$

where $\boldsymbol{z} \sim P_{\boldsymbol{\theta}}$, and we assume either of the following assumptions on $c(\cdot)$:

(i) There exist $c_0 < +\infty$ and $\delta > 0$ such that for all $\boldsymbol{\theta}$ satisfying $d_2\{\boldsymbol{b}(\boldsymbol{\theta}_0), \boldsymbol{b}(\boldsymbol{\theta})\} \leq \delta$ then $c(\boldsymbol{\theta}) \leq c_0$.
(ii) There exists $a > 0$ such that

$$\int_\Theta c(\boldsymbol{\theta})^{1+a} d\Pi(\boldsymbol{\theta}) < +\infty.$$

[A3] There exists some $D > 0$ such that, for all $\varepsilon > 0$ small enough, the prior probability satisfies

$$\Pi\left[\{\boldsymbol{\theta} \in \boldsymbol{\Theta} : d_2\{\boldsymbol{b}(\boldsymbol{\theta}), \boldsymbol{b}(\boldsymbol{\theta}_0)\} \leq \varepsilon\}\right] \gtrsim \varepsilon^D.$$

[A4] **(i)** The map $\boldsymbol{b} : \boldsymbol{\Theta} \to \mathcal{B}$ is continuous. **(ii)** The map $\boldsymbol{b} : \boldsymbol{\Theta} \to \mathcal{B}$ is injective and satisfies

$$\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq L\|\boldsymbol{b}(\boldsymbol{\theta}) - \mathrm{b}(\boldsymbol{\theta}_0)\|^\alpha$$

on some open neighbourhood of $\boldsymbol{\theta}_0$ with $L > 0$ and $\alpha > 0$.

THEOREM 1. *Assume that **[A1]** and **[A3]** are satisfied. We then have the following result:*

*If **[A2](i)** holds, or if **[A2](ii)** holds with a such that $\rho_T(\varepsilon_T) = o(\varepsilon_T^{D/(1+a)})$, then, for M large enough,*

$$\Pi\left(d_2\{\boldsymbol{b}(\boldsymbol{\theta}), \boldsymbol{b}(\boldsymbol{\theta}_0)\} > 4\varepsilon_T/3 + \rho_T^{-1}(\varepsilon_T^D/M) | d_2\{\boldsymbol{\eta}(\boldsymbol{y}), \boldsymbol{\eta}(\boldsymbol{z})\} \leq \varepsilon_T\right) \lesssim 1/M. \tag{4}$$

*Moreover, if **[A4]** holds then*

$$\Pi\left(d_1\{\boldsymbol{\theta}, \boldsymbol{\theta}_0\} > L\{4\varepsilon_T/3 + \rho_T^{-1}(\varepsilon_T^D/M)\}^\alpha | d_2\{\boldsymbol{\eta}(\boldsymbol{y}), \boldsymbol{\eta}(\boldsymbol{z})\} \leq \varepsilon_T\right) \lesssim 1/M. \tag{5}$$

All statements apply for $T \to +\infty$ and $\varepsilon_T = o(1)$.

Not only does Theorem 1 imply posterior consistency but it also leads to a posterior concentration rate, denoted generically by λ_T, that depends on the deviation control $\rho_T(u)$ of $d_2\{\boldsymbol{\eta}(\boldsymbol{z}), \boldsymbol{b}(\boldsymbol{\theta})\}$. From (4) we see that the posterior concentration rate for $\boldsymbol{b}(\boldsymbol{\theta})$ is bounded from below by $O(\varepsilon_T)$. To understand the dependence on $\rho_T(\cdot)$, consider the following two standard situations for **[A2]**.

(a) *Polynomial deviations:* There exist $v_T \to +\infty$ and $u_0, \kappa > 0$ such that

$$\rho_T(u) = 1/v_T^\kappa u^\kappa, \quad u \leq u_0. \tag{6}$$

From (6) we have $\rho_T^{-1}(\varepsilon_T^D/M) = M^{1/\kappa}/(v_T \varepsilon_T^{D/\kappa})$, so that $\varepsilon_T \asymp M^{1/(\kappa+D)} v_T^{-\kappa/(\kappa+D)}$. This implies that $\rho_T(\varepsilon_T) \asymp \varepsilon_T^D/M$, which implies, in turn, that the posterior distribution of $\boldsymbol{b}(\boldsymbol{\theta})$ concentrates at a rate that is bounded above by

$$\lambda_T \asymp v_T^{-\kappa/(\kappa+D)} \tag{7}$$

to $\boldsymbol{b}(\boldsymbol{\theta}_0)$. In the case of Assumption **[A2](ii)**, the same rate can be achieved for all $a > 0$.

(b) *Exponential deviations:* there exists $h_\theta(\cdot) > 0$ such that

$$P_{\boldsymbol{\theta}}\left(d_2\{\boldsymbol{b}(\boldsymbol{\theta}), \boldsymbol{\eta}(\boldsymbol{z})\} > u\right) \leq c(\boldsymbol{\theta}) e^{-h_{\boldsymbol{\theta}}(u v_T)},$$

and there exist $c, C > 0$ such that

$$\int_\Theta c(\boldsymbol{\theta}) e^{-h_\theta(v_T u)} d\Pi(\theta) \leq C e^{-c(u v_T)^\tau}, \quad u \leq u_0.$$

Hence if $c(\boldsymbol{\theta})$ is bounded from above and $h_{\boldsymbol{\theta}}(u)$ is bounded from below by cu^τ for $\boldsymbol{\theta}$ in a neighbourhood of the set $\{\boldsymbol{\theta}; \boldsymbol{b}(\boldsymbol{\theta}) = \boldsymbol{b}(\boldsymbol{\theta}_0)\}$, then $\rho_T(u) \asymp e^{-c_0(u v_T)^\tau}$ and $\rho_T^{-1}(\varepsilon_T^D/M) \asymp \{\log(1/\varepsilon_T)\}^{1/\tau}/v_T$. This implies in particular that if $\varepsilon_T \asymp \log(M v_T)^{1/\tau}/v_T$, one obtains posterior concentration at a rate that is bounded above by

$$\lambda_T \asymp (\log v_T)^{1/\tau}/v_T. \tag{8}$$

For instance, if $\eta(\boldsymbol{z}) = T^{-1}\sum_{i=1}^T h(z_i)$ and if $\{h(z_i)\}_{i \leq T}$ are weakly dependent, say independent and identically distributed for simplicity and with finite k-th moment for $\boldsymbol{\theta}$ in the interior of Θ, then $b(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}\{h(Z)\}$. Furthermore, the Markov inequality implies that

$$P_{\boldsymbol{\theta}}\left(\|\eta(\boldsymbol{z}) - b(\boldsymbol{\theta})\| > u\right) \leq C_k E_{\boldsymbol{\theta}}\left(|h(Z)|^k\right)/(T^{1/2}u)^k,$$

and, with reference to (6), $v_T = T^{1/2}$ and $\kappa = k$. On the other hand if $|h(Z)|$ allows for an exponential moment : $E_{\boldsymbol{\theta}}\left(h^2(Z) e^{a_{\boldsymbol{\theta}}|h(Z)|}\right) \leq A_{\boldsymbol{\theta}} < +\infty$, then for $a_{\boldsymbol{\theta}} T^{1/2} \geq s > 0$

$$\begin{aligned} P_{\boldsymbol{\theta}}\left(\|\eta(\boldsymbol{z}) - b(\boldsymbol{\theta})\| > u\right) &\leq e^{-suT^{1/2}}\left[1 + \frac{s^2}{2T} E_{\boldsymbol{\theta}}\left\{h^2(Z) e^{s|h(Z)|/T^{1/2}}\right\}\right]^T \\ &\leq e^{-suT^{1/2} + s^2 A_{\boldsymbol{\theta}}/2} \leq e^{-u^2 T/(2A_{\boldsymbol{\theta}})}, \end{aligned}$$

choosing $s = uT^{1/2}/A_{\boldsymbol{\theta}} \leq a_{\boldsymbol{\theta}} T^{1/2}$ provided $u \leq a_{\boldsymbol{\theta}} A_{\boldsymbol{\theta}}$. Thus, with reference to the case of exponential deviations, $v_T = T^{1/2}$ and $h_{\boldsymbol{\theta}}(u v_T) = u^2 v_T^2/(2A_{\boldsymbol{\theta}})$. In both cases if the maps $\boldsymbol{\theta} \mapsto E_{\boldsymbol{\theta}}\left(|h(Z)|^k\right)$, $\boldsymbol{\theta} \mapsto a_{\boldsymbol{\theta}}$ and $\boldsymbol{\theta} \mapsto A_{\boldsymbol{\theta}}$ are continuous at $\boldsymbol{\theta}_0$ and positive, then **[A2](i)** and **[A2](ii)** are satisfied.

Remark 1: The role of Assumption **[A3]** in posterior concentration can be illustrated by analysing the role played by the constant D, which controls the degree of prior mass in a neighbourhood of $\boldsymbol{b}(\boldsymbol{\theta}_0)$. For ε_T small, the larger D is, the smaller is the amount of prior mass in the region of the truth and, with reference to case (a), the slower is the rate of posterior concentration in (7), for any $\kappa > 0$. Furthermore, if exponential deviations are assumed (in reference to case (b) above), the rate of posterior convergence in (8) is unaffected by the value of D.

Remark 2: Assumption **[A4]** is critical to the establishment of posterior concentration around $\boldsymbol{\theta}_0$, with satisfaction of the condition that $\boldsymbol{b}(\boldsymbol{\theta})$ be injective depending on both the true structural model and the particular choice of $\boldsymbol{\eta}(\boldsymbol{y})$. Further characterization of this condition, including the development of diagnostic methods for establishing its satisfaction, is the subject of on-going work by the authors.

3. Shape of the Asymptotic Posterior Distribution

In this section we analyse the limiting shape of the posterior measure. We again take ε_T to be sample-size dependent and we consider the shape of $\Pi\left(\cdot|d_2\{\boldsymbol{\eta}(\boldsymbol{y}), \boldsymbol{\eta}(\boldsymbol{z})\} \leq \varepsilon_T\right)$ for various relationships between ε_T and the rate at which summary statistics satisfy a central limit theorem. For notation's sake, in this and the following sections, we denote $\Pi\left(\cdot|d_2\{\boldsymbol{\eta}(\boldsymbol{y}), \boldsymbol{\eta}(\boldsymbol{z})\} \leq \varepsilon_T\right)$ as $\Pi_\varepsilon(\cdot|\boldsymbol{\eta}_0)$, where $\boldsymbol{\eta}_0 = \boldsymbol{\eta}(\boldsymbol{y})$.

We assume that there exists a sequence of (k, k) positive definite matrices $\boldsymbol{\Sigma}_T(\boldsymbol{\theta})$ such that for all $\boldsymbol{\theta}$ in a neighbourhood of $\boldsymbol{\theta}_0 \in \text{Int}(\boldsymbol{\Theta})$,

$$c_1\|\mathbf{D}_T\|_* \leq \|\boldsymbol{\Sigma}_T(\boldsymbol{\theta})\|_* \leq c_2\|\mathbf{D}_T\|_*, \quad \mathbf{D}_T = \text{diag}(d_T(1), \cdots, d_T(k)),$$

with $0 < c_1, c_2 < +\infty$, $d_T(j) \to +\infty$ for all j and the $d_T(j)$ possibly all distinct. Thus, we do not restrict ourselves to identical convergence rates for the components of the statistic $\boldsymbol{\eta}(\boldsymbol{z})$. For simplicity's sake we order the components so that

$$d_T(1) \leq \cdots \leq d_T(k). \tag{9}$$

For all j, we assume $\liminf_T d_T(j)\varepsilon_T = \limsup_T d_T(j)\varepsilon_T$. For any square matrix $\boldsymbol{A}$ of dimension k, if $k_1 \leq k$, $\boldsymbol{A}_{[k_1]}$ denotes the $k_1 \times k_1$ square upper sub-matrix of $\boldsymbol{A}$. Also, let $k_1 = \max\{j : \lim_T d_T(j)\varepsilon_T = 0\}$ and if, for all j, $\liminf_T d_T(j)\varepsilon_T > 0$ then $k_1 = 0$.

While Assumption **[A4]** requires the map $\boldsymbol{\theta} \mapsto \boldsymbol{b}(\boldsymbol{\theta})$ be injective (but not necessarily bijective) this section operates under the assumption that $\boldsymbol{b}(\cdot)$ is continuously differentiable at $\boldsymbol{\theta_0}$, and that the Jacobian $\boldsymbol{b}'(\boldsymbol{\theta}_0)$ has full rank $d_{\boldsymbol{\theta}}$. We also consider a local version of Assumption **[A2]** :

[A2′] There exist $\kappa > 1$ and $\delta > 0$ such that for all $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \delta$,

$$P_{\boldsymbol{\theta}}\left(\|\Sigma_T(\boldsymbol{\theta}_0)\{\boldsymbol{\eta}(\boldsymbol{z}) - \boldsymbol{b}(\boldsymbol{\theta})\}\| > u\right) \leq \frac{c_0}{u^\kappa}, \text{ for all } 0 < u \leq \delta d_T(k)$$

In addition to the above assumptions and Assumptions **[A1]-[A4]** in Section 2 (with posterior concentration of $\|\boldsymbol{b} - \boldsymbol{b}_0\|$ at rate $\lambda_T \gg 1/d_T(1)$ required), the following conditions are needed to establish the limiting shape of $\Pi_\varepsilon(\cdot|\boldsymbol{\eta}_0)$.

[A5] There exists $\delta > 0$ such that for all $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \delta$,

$$\boldsymbol{\Sigma}_T(\boldsymbol{\theta})\{\boldsymbol{\eta}(\boldsymbol{z}) - \boldsymbol{b}(\boldsymbol{\theta})\} \Rightarrow \mathcal{N}(0, I_k),$$

where I_k is the $k \times k$ identity matrix.

[A6] The sequence of functions $\boldsymbol{\theta} \mapsto \boldsymbol{\Sigma}_T(\boldsymbol{\theta})\mathbf{D}_T^{-1}$ is equicontinuous at $\boldsymbol{\theta}_0$.

The following condition will only be used in cases where at least one of the coordinates satisfies $d_T(j)\varepsilon_T = o(1)$.

[A7] For some positive δ and all $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \delta$, and for all ellipsoids

$$B_T = \left\{(t_1, \cdots, t_{k_1}) : \sum_{j=1}^{k_1} t_j^2/h_T(j)^2 \leq 1\right\},$$

with $h_T(j) \to 0$ for all $j \leq k_1$ and all $u \in \mathbb{R}^{k_1}$ fixed,

$$\begin{aligned} &\lim_T \frac{P_{\boldsymbol{\theta}}\left([\boldsymbol{\Sigma}_T(\boldsymbol{\theta})]_{[k_1]}\{\boldsymbol{\eta}(\boldsymbol{z}) - \boldsymbol{b}(\boldsymbol{\theta})\} - u \in B_T\right)}{\prod_{j=1}^{k_1} h_T(j)} = \varphi_{k_1}(u), \\ &\frac{P_{\boldsymbol{\theta}}\left([\boldsymbol{\Sigma}_T(\boldsymbol{\theta})]_{[k_1]}\{\boldsymbol{\eta}(\boldsymbol{z}) - \boldsymbol{b}(\boldsymbol{\theta})\} - u \in B_T\right)}{\prod_{j=1}^{k_1} h_T(j)} \leq H(u), \quad \int H(u)du < +\infty, \end{aligned} \tag{10}$$

for $\varphi_{k_1}(\cdot)$ the density of a k_1-dimensional normal random variate.

THEOREM 2. *Assume that **[A1]-[A6]** and **[A2′]**, with $\kappa > d_\theta$, are satisfied, where for $\eta_1, \eta_2 \in \mathcal{B}$, $d_2\{\eta_1, \eta_2\} = \|\eta_1 - \eta_2\|$. The following results hold:*

(i) $\lim_T d_T(1)\varepsilon_T = +\infty$*: With probability approaching one, the posterior distribution of $\varepsilon_T^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$ converges to the uniform distribution over the ellipse $\{x^t B_0 x \leq 1\}$ with $B_0 = \boldsymbol{b}'(\boldsymbol{\theta})^t\boldsymbol{b}'(\boldsymbol{\theta})$. In other words, for all f continuous and bounded, with probability approaching one*

$$\int f\{\varepsilon_T^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\}d\Pi_\varepsilon(\boldsymbol{\theta}|\boldsymbol{\eta}_0) \to \frac{\int_{u^t B_0 u \leq 1} f(u)du}{\int_{u^t B_0 u \leq 1} du}. \tag{11}$$

(ii) *There exists $k_0 < k$ such that* $\lim_T d_T(1)\varepsilon_T = \lim_T d_T(k_0)\varepsilon_T = c$, $0 < c < +\infty$, *and* $\lim_T d_T(k_0 + 1)\varepsilon_T = +\infty$*: Assume that $\boldsymbol{\Sigma}_T(\boldsymbol{\theta})\boldsymbol{D}_T^{-1} \to \boldsymbol{A}(\boldsymbol{\theta})$ with $\boldsymbol{A}(\boldsymbol{\theta}_0)$ positive definite and that*

Leb $\left(\sum_{j=1}^{k_0}\left[\{\boldsymbol{b}'(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_{\mathbf{0}})\}_{[j]}\right]^2 \leq c\varepsilon_T^2\right) = +\infty$, *then*

$$\Pi_\varepsilon\left(\boldsymbol{\Sigma}_T(\boldsymbol{\theta}_0)\{\boldsymbol{b}(\boldsymbol{\theta})-\boldsymbol{b}(\boldsymbol{\theta}_0)\} - Z_T^0 \in B|\boldsymbol{\eta}_0\right) \to 0, \tag{12}$$

for all bounded measurable sets B.

(iii) *There exists* $k_1 < k$ *such that* $\lim_T d_T(k_1)\varepsilon_T = 0$ *and* $\lim_T d_T(k_1+1)\varepsilon_T = +\infty$*: Assume that* $\boldsymbol{\Sigma}_T(\boldsymbol{\theta})\boldsymbol{D}_T^{-1} \to \boldsymbol{A}(\boldsymbol{\theta})$ *with* $\boldsymbol{A}(\boldsymbol{\theta}_0)$ *positive definite, and that Assumption **[A7]** is satisfied, then* (11) *is satisfied.*

(iv) $\lim_T d_T(j)\varepsilon_T = c > 0$ *for all* $j \leq k$ *or Leb* $\left(\sum_{j=1}^{k_0}\left[\{\boldsymbol{b}'(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_{\mathbf{0}})\}_{[j]}\right]^2 \leq c\varepsilon_T^2\right) < +\infty$ *in case* ***(ii)****: There exists a non-Gaussian probability distribution on* $\mathbb{R}^k$*,* Q_c *which depends on c and is such that*

$$\Pi_\varepsilon\left(\boldsymbol{\Sigma}_T(\boldsymbol{\theta}_0)\{\boldsymbol{b}(\boldsymbol{\theta})-\boldsymbol{b}(\boldsymbol{\theta}_0)\} - Z_T^0 \in B|\boldsymbol{\eta}_0\right) \to Q_c(B). \tag{13}$$

More precisely,

$$Q_c(B) = \frac{\int_{x\in B} P_{Z\sim\mathcal{N}_k(0,I_k)}\left\{(Z-x)'\boldsymbol{A}(\boldsymbol{\theta}_0)'\boldsymbol{A}(\boldsymbol{\theta}_0)(Z-x) \leq c^2\right\}dx}{\int_{x\in\mathbb{R}^k} P_{Z\sim\mathcal{N}_k(0,I_k)}\left\{(Z-x)'\boldsymbol{A}(\boldsymbol{\theta}_0)'\boldsymbol{A}(\boldsymbol{\theta}_0)(Z-x) \leq c^2\right\}dx}.$$

(v) $\lim_T d_T(k)\varepsilon_T = 0$*: Assume that* $\boldsymbol{\Sigma}_T(\boldsymbol{\theta})\boldsymbol{D}_T^{-1} \to \boldsymbol{A}(\boldsymbol{\theta})$ *with* $\boldsymbol{A}(\boldsymbol{\theta}_0)$ *positive definite and that Assumption **[A7]**, i.e.,* (10)*, holds for* $k_1 = k$*, then*

$$\lim_T \Pi_\varepsilon\left(\boldsymbol{\Sigma}_T(\boldsymbol{\theta}_0)\{\boldsymbol{b}(\boldsymbol{\theta})-\boldsymbol{b}(\boldsymbol{\theta}_0)\} - Z_T^0 \in B|\boldsymbol{\eta}_0\right) = \Phi_k(B). \tag{14}$$

Remark 3: Theorem 2 asserts that the crucial feature for determining the limiting shape of the posterior is the behaviour of $d_T(j)\varepsilon_T$, for $j = 1, ..., k$. If $\varepsilon_T \to 0$ too slowly so that some (or all) of the components satisfy $\lim_T d_T(j)\varepsilon_T > 0$, as in cases **(i)**-**(iv)** above, then the posterior distribution is not asymptotically Gaussian. In particular, case **(i)**, which corresponds to a tolerance ε_T that is much too large, may seem surprising since the posterior converges towards a uniform distribution over an ellipse, which is nonstandard asymptotic behaviour. However, the proof of Theorem 1 shows that this behaviour results from the fast rate of convergence of $\boldsymbol{\eta}(\boldsymbol{z})$ towards $\boldsymbol{b}(\boldsymbol{\theta})$ under $P_{\boldsymbol{\theta}}$, relative to the size of the tolerance ε_T. A heuristic argument to demonstrate this fact follows: under the assumptions of Theorem 2, $\|\boldsymbol{\eta}(\boldsymbol{z})-\boldsymbol{\eta}(\boldsymbol{y})\| \leq \varepsilon_T$ is equivalent to the deterministic constraint $\|\boldsymbol{b}(\boldsymbol{\theta})-\boldsymbol{b}(\boldsymbol{\theta}_0)\| \leq \varepsilon_T[1+o(1)]$. Therefore, the probability $P_{\boldsymbol{\theta}}\left(\|\boldsymbol{\eta}(\boldsymbol{z})-\boldsymbol{\eta}(\boldsymbol{y})\| \leq \varepsilon_T\right)$ is itself equivalent to $\|\boldsymbol{b}(\boldsymbol{\theta})-\boldsymbol{b}(\boldsymbol{\theta}_0)\| \leq \varepsilon_T[1+o_P(1)]$, which is also equivalent to $\|\boldsymbol{b}'(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\| \leq \varepsilon_T[1+o_P(1)]$ by the regularity condition on $\boldsymbol{b}$. Hence, the posterior behaves like the prior distribution truncated over the above ellipsoid and prior continuity implies that this is equivalent to the uniform distribution over this ellipsoid.

If $\lim_T d_T(j)\varepsilon_T = c > 0$ for all $j \leq k$, such that $\varepsilon_T = O(1/d_T)$, the limiting form of the posterior is a perturbation of a Gaussian distribution. It is only when $\lim_T d_T(k)\varepsilon_T = 0$, as in case **(v)**, that a Bernstein-von Mises result is available.

Taking the case where $d_T(1) = \cdots = d_T(k) = d_T$ for the sake of simplicity, an important consequence of Theorem 2 is that resulting credible regions are only frequentist confidence regions if $\varepsilon_T = o(1/d_T)$. If $\varepsilon_T = O(1/d_T)$ then credible regions have radius with the correct order of magnitude but do not necessarily have the correct coverage asymptotically. Finally, even if ε_T is much larger in order than d_T^{-1} then the posterior leads to reasonable point estimators, for instance as seen in Theorem 3, part **(i)**, but the associated measures of uncertainty are meaningless.

Remark 4: Condition **[A7]** only applies to random variables $\boldsymbol{\eta}(\boldsymbol{z})$ that are absolutely continuous with respect to Lebesgue measure (or, in the case of sums of i.i.d random variables, to sums that are non-lattice; see Bhattacharya and Rao, 1986). The case of discrete $\boldsymbol{\eta}(\boldsymbol{z})$ requires an adaptation of condition **[A7]** that leads to the same conclusions in Theorem 2. For simplicity's sake we write this adaptation in the case where $d_T(1) = \cdots = d_T(k) = d_T$ and under the restriction that $d_T\varepsilon_T = o(1)$, so that we only need study case **(v)** in Theorem 2. Then **[A7]** can be replaced by:

[A7′] There exist $\delta > 0$ and a countable set E_T such that for all $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < \delta$,

$$P_{\boldsymbol{\theta}}\left(\boldsymbol{\eta}(\boldsymbol{z}) \in E_T\right) = 1; \text{ for all } \boldsymbol{t} \in E_T, P_{\boldsymbol{\theta}}\left(\boldsymbol{\eta}(\boldsymbol{z}) = \boldsymbol{t}\right) > 0.$$

and there exists a continuous and positive map $\boldsymbol{\theta} \mapsto \boldsymbol{A}(\boldsymbol{\theta})$ at $\boldsymbol{\theta}_0$ such that

$$\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| \le \delta} \sum_{t \in E_T} \left| p(\boldsymbol{\Sigma}_T(\boldsymbol{\theta})\{\boldsymbol{t} - \boldsymbol{b}(\boldsymbol{\theta})\}|\boldsymbol{\theta}) - d_T^{-k} \boldsymbol{A}(\boldsymbol{\theta}) \varphi[\boldsymbol{\Sigma}_T(\boldsymbol{\theta})\{\boldsymbol{t} - \boldsymbol{b}(\boldsymbol{\theta})\}] \right| = o(1).$$

Under this alternative condition, the conclusion of case **(v)** of Theorem 2 still holds.

Condition **[A7′]** is satisfied, for instance, in the case when $\boldsymbol{\eta}(\boldsymbol{z})$ is a sum of i.i.d. lattice random variables, as in the population genetic experiment detailed in Section 3.3 of Marin et al. (2014). Furthermore, this population genetic example is such that Assumptions **[A1]** - **[A6]** and **[A7′]** also hold, which means that the conclusions of both Theorems 1 and 2 apply to this model.

4. ASYMPTOTIC DISTRIBUTION OF THE POSTERIOR MEAN

4·1. *Main Result*

As noted above, the current literature on the asymptotics of approximate Bayesian computation has focused primarily on conditions guaranteeing asymptotic normality of the posterior mean (or functions thereof). To this end, it is important to stress that the posterior normality result in Theorem 2 is not a weaker, or stronger, result than that of asymptotic normality of an approximate point estimator; both results simply focus on different objects. That said, existing proofs of the asymptotic normality of the posterior mean all require asymptotic normality of the posterior. In this section, we demonstrate that asymptotic normality of the posterior is not a necessary condition for asymptotic normality of the posterior mean. To present the ideas in as transparent a manner as possible, we focus on the simple case of an unknown scalar parameter θ and known scalar summary $\eta(\boldsymbol{y})$. This result can be extended to the multivariate case at the cost of more involved notations, which we omit since the arguments required strongly mirror those of the univariate result. In addition to Assumptions **[A1]** to **[A7]**, we maintain the following assumptions on the prior.

[**A8**] The prior density $p(\cdot)$ satisfies the following: **(i)** For $\theta_0 \in \text{Int}(\Theta)$, $p(\theta_0) > 0$. **(ii)** The density function $\theta \to p(\theta)$ is β- Hölder in a neighbourhood of θ_0: There exist $\delta, L > 0$ such that for all $|\theta - \theta_0| \le \delta$,

$$\left| p(\theta) - \sum_{j=0}^{\lfloor \beta/2 \rfloor} (\theta - \theta_0)^j \frac{p^{(j)}(\theta_0)}{j!} \right| \le L|\theta - \theta_0|^\beta.$$

(iii) For $\Theta \subset \mathbb{R}$, we have $\int_\Theta |\theta|^\beta p(\theta) d\theta < \infty$.

THEOREM 3. *Assume that **[A1]** - **[A6]** and **[A2′]**, with $\kappa > \beta + 1$, together with **[A8]**, are satisfied. Assume also that $\theta \to b(\theta)$ is β-Hölder in a neighbourhood of θ_0 and that $b'(\theta_0) \neq 0$. Denoting $E_{\Pi_\varepsilon}[\theta]$ and $E_{\Pi_\varepsilon}[b]$ as the posterior means of θ and $b(\theta)$, respectively, we then have the following results:*

*If **(i)** :* $\liminf_T d_T \varepsilon_T = +\infty$ *then*

$$E_{\Pi_\varepsilon}[b - b_0] = \frac{Z_T^0}{d_T} + \sum_{j=1}^{k} \frac{p^{(2j-1)}(b_0)}{(2j-1)!(2j+1)} \varepsilon_T^{2j} + O(\varepsilon_T^{1+\beta}) + o_P(1/d_T),$$

where $k = \lfloor \beta/2 \rfloor$ *and, for* $b_0 = b(\theta_0)$, *and* $(b^{-1})^{(j)}(b_0)$ *the* j*-th derivative of the inverse of the map* $\theta \mapsto b(\theta)$,

$$E_{\Pi_\varepsilon}\left[\theta - \theta_0\right] = \frac{Z_T^0}{d_T b'(\theta_0)} + \sum_{j=1}^{\lfloor \beta \rfloor} \frac{(b^{-1})^{(j)}(b_0)}{j!} \sum_{l=\lceil j/2 \rceil}^{\lfloor (j+k)/2 \rfloor} \frac{\varepsilon_T^{2l} p^{(2l-j)}(b_0)}{p(b_0)(2l-j)!} + O(\varepsilon_T^{1+\beta}) + o_P(1/d_T) \quad (15)$$

Hence, if $\varepsilon_T^{2\wedge(1+\beta)} = o(1/d_T)$ *and* $\liminf_T d_T\varepsilon_T = +\infty$

$$E_{\Pi_\varepsilon}\left[\theta - \theta_0\right] = \frac{Z_T^0}{d_T b'(\theta_0)} + o_P(1/d_T), \quad E_{\Pi_\varepsilon}\left[d_T(\theta - \theta_0)\right] \Rightarrow \mathcal{N}(0, V(\theta_0)/\{b'(\theta_0)\}^2), \tag{16}$$

where $V(\theta_0) = \lim_T Var[d_T(\eta\{\boldsymbol{y}) - b(\theta_0)\}]$.

*If **(ii)**:* $\lim_T d_T\varepsilon_T = c \geq 0$ *and **[A7]** holds then* (16) *also holds.*

Remark 5: Equation (15) highlights a potential deviation from the expected asymptotic behaviour of the posterior mean $E_{\Pi_\varepsilon}(\theta)$, i.e., the behaviour corresponding to $\varepsilon_T = 0$. Indeed, the posterior mean is asymptotically normal for all values of ε_T, but is asymptotically unbiased only if the leading term in equation (15) is $Z_T^0/(d_T b'\{\theta_0)\}$, i.e., if and only if $d_T\varepsilon_T^2 = o(1)$ (assuming $\beta \geq 1$). When $d_T\varepsilon_T^2 \to +\infty$, this is not satisfied and the posterior mean has a bias that is equal to

$$\varepsilon_T^2\left(\frac{p'(b_0)}{3p(b_0)b'(\theta_0)} - \frac{b^{(2)}(\theta_0)}{2\{b'(\theta_0)\}^2}\right) + O(\varepsilon_T^4) + o_P(1/d_T),$$

provided $\beta \geq 3$. In particular, the "optimal" choice of summary statistics proposed by Fearnhead and Prangle (2010), namely, choosing $\eta(\boldsymbol{y})$ to be the true posterior mean $\hat{\theta} = E[\theta|\mathbf{y}]$, in which case $b(\theta) = \theta$, does not lead to asymptotic normality of the posterior mean around θ_0 when $d_T\varepsilon_T^2 \to +\infty$ if $p'(\theta_0) \neq 0$. However, we note that it is possible to de-bias the posterior mean when $d_T\varepsilon_T^2 \to +\infty$ and $d_T\varepsilon_T^4 = o(1)$ by choosing a prior that satisfies

$$p(\theta) \propto |b'(\theta)^{5/2}|, \text{ for all } \theta \in \Theta,$$

provided it is a proper prior distribution.

Remark 6: Part **(i)** of Theorem 3 demonstrates that asymptotic normality of the posterior mean does not require asymptotic normality of the posterior. However, part **(ii)** of Theorem 3 states that asymptotic normality of both the posterior and the posterior mean can be simultaneously achieved provided $\varepsilon_T = o(1/d_T)$.

Remark 7: The results of Theorem 3 entirely rest on the injectivity of the map $\theta \mapsto b(\theta)$. In particular, for $\eta_0 = \eta(\boldsymbol{y})$, Theorem 3 relies on the fact that if the posterior distribution of θ is

$$\Pi_\varepsilon(\theta \in B|\eta_0) = \frac{\int_B P_\theta\left(d\{\eta(\boldsymbol{z}), \eta(\boldsymbol{y})\} \leq \varepsilon_T\right) d\Pi(\theta)}{\int_\Theta P_\theta\left(d\{\eta(\boldsymbol{z}), \eta(\boldsymbol{y})\} \leq \varepsilon_T\right) d\Pi(\theta)},$$

by the injectivity of $\theta \mapsto b(\theta)$, the posterior distribution of $b(\theta) = b$ follows as

$$\Pi_\varepsilon(b \in A|\eta_0) = \frac{\int_A P_b\left(d\{\eta(\boldsymbol{z}), \eta(\boldsymbol{y})\} \leq \varepsilon_T\right) d\tilde{\Pi}(b)}{\int_{b(\Theta)} P_b\left(d\{\eta(\boldsymbol{z}), \eta(\boldsymbol{y})\} \leq \varepsilon_T\right) d\tilde{\Pi}(b)},$$

where $\tilde{\Pi}$ is the transformed measure of the prior Π by b. Hence, the results of Theorem 3 can be obtained by analysing the posterior mean

$$E_{\Pi_\varepsilon}[d_T(b - b_0)] = \frac{\int_{b(\Theta)} d_T(b - b_0) P_b\left(d\{\eta(\boldsymbol{z}), \eta(\boldsymbol{y})\} \leq \varepsilon_T\right) d\tilde{\Pi}(b)}{\int_{b(\Theta)} P_b\left(d\{\eta(\boldsymbol{z}), \eta(\boldsymbol{y})\} \leq \varepsilon_T\right) d\tilde{\Pi}(b)}. \tag{17}$$

From (17), the result for $E_{\Pi_\varepsilon}[d_T(\theta - \theta_0)]$ follows from the delta theorem.

Remark 8: All results derived in this and the previous sections, as well as all relevant remarks made, can be shown to remain applicable in the case where $\boldsymbol{\eta}(\cdot)$ is taken as a vector of estimating equations derived from an auxiliary model (see, e.g., Drovandi et al., 2015; Martin et al., 2016), provided conditions corresponding to those in **[A1]** to **[A8]** are satisfied. Details are omitted for the sake of brevity.

4·2. *Comparison with Existing Results*

Li and Fearnhead (2016) have, in parallel, analyzed the asymptotic properties of the posterior mean, or some function thereof. (See also Creel et al., 2015.) Under a central limit theorem assumption for the summary statistics, and further regularity assumptions on the convergence of the density of the summary statistics to this normal limit, including the existence of an Edgeworth expansion with exponential controls on the tails, Li and Fearnhead (2016) demonstrate asymptotic normality of the posterior mean if $\varepsilon_T = o(d_T^{-3/5})$. Heuristically, the authors derive this result using an approximation of the posterior density $p_\varepsilon\{\theta|\eta(\boldsymbol{y})\}$, based on the Gaussian approximation of the density of $\eta(\boldsymbol{z})$ given θ, and using properties of the maximum likelihood estimator conditional on $\eta(\boldsymbol{y})$. By using an arbitrary kernel (as in Fearnhead and Prangle, 2012) $K_\varepsilon(\|\eta(\boldsymbol{y}) - \eta(\boldsymbol{z})\|)$, rather than the uniform version herein, the results of Li and Fearnhead (2016) are more general than those presented in our Theorem 3. However, their conditions on the summary statistics $\eta(\boldsymbol{y})$ are significantly stronger than ours.

In particular, when $\lim_T d_T\varepsilon_T = +\infty$, our results on asymptotic normality for the posterior mean only require weak convergence of $d_T\{\eta(\boldsymbol{z}) - b(\theta)\}$ under P_θ, with polynomial deviations that need not be uniform in θ. These weaker assumptions also allow for the explicit treatment of models where the parameter space Θ is not compact. In the case where $d_T\varepsilon_T = O(1)$, asymptotic normality of the posterior mean requires Assumption **[A7]**, which essentially boils down to local (in θ) convergence of the density of $d_T\{\eta(\boldsymbol{z}) - b(\theta)\}$, but without any requirement on the rate of this convergence; with this assumption still being weaker than the uniform convergence required in Li and Fearnhead (2016). Our results also allow for an explicit representation of the bias that exists in the posterior mean if the tolerance does not satisfy $\varepsilon_T = o(d_T^{-1/2})$.

Moreover, and in contrast to Li and Fearnhead (2016), our results in Theorem 2 completely characterize the asymptotic distribution of the posterior for all $\varepsilon_T = o(1)$ that admit posterior concentration. In part, this general characterization allows us to demonstrate, via Theorem 3 **(i)**, that asymptotic normality and efficiency (in the sense discussed in Li and Fearnhead, 2016) of the posterior mean remains achievable even if $\lim_T d_T\varepsilon_T = +\infty$, provided the tolerance satisfies $\varepsilon_T = o(d_T^{-1/2})$. Furthermore, this general treatment allows us to conclude that asymptotic normality of the posterior mean is achievable for essentially all $\varepsilon_T = o(1)$ but with a bias term if $d_T\varepsilon_T^2 \neq o(1)$. However, provided one is willing to consider the use of bias reduction methods outlined in Remark 5, asymptotic normality of the posterior mean can still be achieved under the weaker requirement that $\varepsilon_T^4 = o(1/d_T)$.

The primary reason for the difference in assumptions and results between the approach used herein and Li and Fearnhead (2016) is that our approach is not restricted to treating the posterior density (or distribution) in the same fashion as the likelihood of $\eta(\boldsymbol{z})$. Instead, we make assumptions directly on the probability distribution itself through assumptions on $P_\theta\left(\|\eta(\boldsymbol{z}) - \eta(\boldsymbol{y})\| \leq \varepsilon_T\right)$.

5. PRACTICAL IMPLICATIONS OF THE RESULTS

5·1. *General Implications*

It is common practice in approximate Bayesian computation algorithms to define the tolerance ε_T implicitly by considering an empirical quantile over the prior simulated distances $d_2\{\boldsymbol{\eta}(\boldsymbol{z}), \boldsymbol{\eta}(\boldsymbol{y})\}$. The implications of Theorems 1 and 2 for the choice of ε_T can be extended straightforwardly to this scenario and, as we will show, several relevant practical insights into the application of this algorithm ensue. With reference to the rates in equation (9), consider, for simplicity's sake, the case where $d_T(1) = \cdots = d_T(k) = d_T$ and the distances are Euclidean, with $k \geq d_{\boldsymbol{\theta}}$. Then, let us consider two cases regarding the tolerance ε_T: one where $\varepsilon_T = o(1/d_T)$, as required for the Bernstein-von Mises result in Theorem 2; and a second one where $\varepsilon_T \gtrsim d_T^{-1}$, as required for Theorem 1. In the first case, using the same types of computations as in the proof of Theorem 2 in the Appendix, we have, for $Z_T = \boldsymbol{\Sigma}_T(\boldsymbol{\theta}_0)\{\boldsymbol{\eta}\{\boldsymbol{z}) - \boldsymbol{b}(\boldsymbol{\theta})\}$,

$$\text{Pr}\left(\|\boldsymbol{\eta}(\boldsymbol{z}) - \boldsymbol{\eta}(\boldsymbol{y})\| \le \varepsilon_T\right) = \int_{\boldsymbol{\Theta}} P_{\boldsymbol{\theta}}\left(\|Z_T - Z_T^0 - d_T\{\boldsymbol{b}(\boldsymbol{\theta}) - \boldsymbol{b}(\boldsymbol{\theta}_0)\}\| \le \varepsilon_T d_T\right) p(\boldsymbol{\theta}) d\boldsymbol{\theta}$$
$$\asymp (\varepsilon_T d_T)^k \int_{\boldsymbol{\Theta}} \varphi\{Z_T^0 + d_T \boldsymbol{b}'(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\} d\boldsymbol{\theta} \asymp \varepsilon_T^k d_T^{k-d_{\boldsymbol{\theta}}};$$

while, in the second case ($\varepsilon_T \gtrsim d_T^{-1}$) we have

$$\text{Pr}\left[\|\boldsymbol{\eta}(\boldsymbol{z}) - \boldsymbol{\eta}(\boldsymbol{y})\| \le \varepsilon_T\right] \asymp \int_{\Theta} \varphi\{Z_T^0 + d_T \boldsymbol{b}'(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\} d\boldsymbol{\theta} \asymp \varepsilon_T^{d_{\boldsymbol{\theta}}}.$$

In particular, choosing a tolerance, $\varepsilon_T = o(1)$, is equivalent to choosing an $\alpha_T = o(1)$ quantile of $\|\boldsymbol{\eta}(\boldsymbol{z}) - \boldsymbol{\eta}(\boldsymbol{y})\|$. With regard to both cases above, choosing $\alpha_T = o(d_T^{-d_{\boldsymbol{\theta}}})$ induces a tolerance $\varepsilon_T \asymp (\alpha_T d_T^{d_{\boldsymbol{\theta}}})^{1/k} d_T^{-1}$, while taking $o(1) = \alpha_T \gtrsim d_T^{-d_{\boldsymbol{\theta}}}$ induces the tolerance $\varepsilon_T \asymp \alpha_T^{1/d_{\boldsymbol{\theta}}}$.

There are several consequences of this result. First, by linking the order of ε_T with the quantile of the distances $\|\boldsymbol{\eta}(\boldsymbol{z}) - \boldsymbol{\eta}(\boldsymbol{y})\|$, we theoretically rationalize the oft-used practice of only selecting draws of $\boldsymbol{\theta}$ that yield distances in the left tail of $\{\boldsymbol{\theta} \in \boldsymbol{\Theta} : \|\boldsymbol{\eta}(\boldsymbol{z}) - \boldsymbol{\eta}(\boldsymbol{y})\|\}$. See, for example, Biau et al. (2015). Secondly, this construction allows us to explicitly express the impact of $d_{\boldsymbol{\theta}} = \dim(\boldsymbol{\Theta})$ on the choice of tolerance: everything else unchanged, the larger $d_{\boldsymbol{\theta}}$ is, the smaller is the tolerance ε_T (and associated quantile level, α_T) required to achieve a given level of accuracy for the selected draws in Algorithm 1. This result thus gives theoretical evidence of the so-called curse-of-dimensionality encountered in these algorithms as the dimension of the parameters of interest increases, this issue being distinct from that concerning the dimension of the summary statistics. While the former of these problems has been acknowledged heuristically elsewhere, to our knowledge this is the first piece of work to pinpoint the underlying cause. In addition, this finding gives theoretical justification for the commonly used dimension reduction methods that treat parameter dimensions individually and independent of the other remaining dimensions; see, for example, the regression adjustment approaches of Beaumont et al. (2002) and Blum (2010), and the integrated (auxiliary) likelihood approach of Martin et al. (2016), all of which treat parameters one-at-a-time in the hope of obtaining more accurate marginal posterior inference.

A third implication of the above result is that, by linking ε_T and α_T as shown, one has a means of choosing the α_T quantile of the simulations (or equivalently the tolerance ε_T) in such a way that a particular type of posterior behaviour is expected to be evident, at least for large T. That is, if $\varepsilon_T \gtrsim d_T^{-1}$ the posterior can be expected to concentrate, but if one is willing to impose the more stringent condition $\varepsilon_T = o(1/d_T)$ the posterior will both concentrate *and* be approximately Gaussian in large samples. Such results are important as they give to practitioners an understanding of what to expect from the procedure, and a means of detecting potential issues if this expected posterior behaviour is not in evidence when choosing a certain α_T quantile (or tolerance ε_T).

Furthermore, whilst the general consensus is that, everything else unchanged, choosing ε_T smaller yields more accurate results for larger computing budgets, these theoretical results demonstrate that what really matters for accurate inference based on these algorithms is to choose ε_T small enough to enable the (limiting) shape of the posterior to be determined by the (asymptotic behaviour of the) summaries $\boldsymbol{\eta}(\boldsymbol{y})$, rather than being influenced by the choice of ε_T itself. To this end, we can interpret the results of Theorem 2 as follows: if $\varepsilon_T = o(1/d_T)$, choosing a smaller tolerance $\tilde{\varepsilon}_T \ll \varepsilon_T$ and re-running Algorithm 1 will not significantly alter the shape of the posterior (for N sufficiently large so that Monte Carlo error is minimal, and for large enough T), with the result holding equivalently for choosing some quantile level such that $\tilde{\alpha}_T \ll \alpha_T$. Stated another way, once ε_T has reached the $o(1/d_T)$ threshold, decreasing the tolerance further, at the cost of more expensive numerical computations, will not necessarily yield a more accurate posterior estimate where, by ‘accurate’ we mean here a posterior that is in close accordance with the asymptotic Gaussian distribution that obtains theoretically, under regularity.

This latter result thus contradicts some persistent opinion on such algorithms that the tolerance should always be taken “as small as the computing budget allows.” Theorem 2 states that ε_T should indeed be ‘small’ but, simultaneously, it gives an upper bound on the gain in accuracy one can hope to achieve by taking ε_T small. To demonstrate this idea, consider the following numerical illustration. The example adopted is sufficiently regular to ensure that the central limit theorem that underpins the Bernstein-von Mises result holds for a reasonably small sample size, and that the associated order condition on ε_T has some practical content. That is, the illustration highlights the fact that, despite the asymptotic foundations of the conclusions drawn above regarding the optimal choice of the tolerance ε_T, or equivalently the quantile α_T, those conclusions can be relevant even in the moderately sized samples often encountered in practice.

5·2. *Numerical Illustration of Quantile Choice*

Consider the simple example where we observe a sample $\{y_t\}_{t=1}^T$ from $y_t \sim_{i.i.d.} N(\mu, \sigma)$ and $T = 100$, and our goal is posterior inference on $\theta = (\mu, \sigma)'$. We use as our summaries the sample mean and variance, $\bar{x}$ and s_T^2. These summaries satisfy a central limit theorem at rate $T^{1/2}$, and so if we wish to guarantee approximate posterior normality, we choose an α_T quantile of the simulated distances according to $\alpha_T = o(1/T)$, since we wish to conduct joint inference on μ and σ. For the purpose of this illustration, we will compare inference based on Algorithm 1 using four different choices of α_T, where we drop the subscript T for notational simplicity: $\alpha_1 = 1/T^{1.1}$, $\alpha_2 = 1/T^{3/2}$, $\alpha_3 = 1/T^2$ and $\alpha_4 = 1/T^{5/2}$.

Draws for (μ, σ) are simulated on $[0.5, 1.5] \times [0.5, 1.5]$ according to independent uniforms $\mathcal{U}[0.5, 1.5]$. The number of simulation draws N is chosen so that we retain 250 accepted draws for each of the different choices $(\alpha_1, ..., \alpha_4)$. The exact (finite sample) marginal posteriors of μ and σ are produced by numerically evaluating the likelihood function, normalizing over the support of the prior and marginalizing with respect to each parameter. Given the sufficiency of $\bar{x}$, s_T^2, the exact marginal posteriors for μ and σ are equal to those based directly on the summaries themselves.

We summarize the accuracy of the resulting posterior estimates, across these four quantile choices, by computing the average, over 50 replications, of the root mean squared error of the estimates of the exact posteriors for each parameter. For example, in the case of the parameter μ, we define the root mean squared error between the marginal posterior obtained from Algorithm 1 using α_j and denoted by $\widehat{p}_{\alpha_j}\{\mu|\boldsymbol{\eta}(\boldsymbol{y})\}$, and the exact marginal posterior $p(\mu|\boldsymbol{y})$ as

$$RMSE_\mu(\alpha_j) = \left(\frac{1}{G} \sum_{g=1}^{G} \left[\widehat{p}_{\alpha_j}^g\{\mu|\boldsymbol{\eta}(\boldsymbol{y})\} - p^g(\mu|\boldsymbol{y}) \right]^2 \right)^{1/2}, \tag{18}$$

where $\widehat{p}^g$ is the ordinate of the density estimate from Algorithm 1 and p^g the ordinate of the exact posterior density, at the g-th grid point upon which the density is estimated. The root mean squared error for the σ marginal is computed analogously. Across the 50 replications we fix $T = 100$ and generate observations according to the parameter values $\mu_0 = 1,\ \sigma_0 = 1$.

The results in Table 1 report average root mean squared errors, each as a ratio to the value associated with $\alpha_4 = 1/T^{5/2}$. Values smaller than one thus indicate that the larger (and, hence, less computationally burdensome) value of α_j yields (on average) a more accurate posterior estimate than that yielded by α_4. In brief, we see that for σ, the estimates based on α_j, $j = 1, 2, 3$ are all more accurate (on average) than those based on α_4, with there being no gain in accuracy (in fact, we observe a slight decline) beyond $\alpha_1 = 1/T^{1.1}$. In the case of μ, estimates based on α_2 and α_3 are both more accurate that those based on α_4 and with there being minimal gain in pushing the quantile below $\alpha_1 = 1/T^{1.1}$.

These numerical results clearly have important computational implications. To wit, and as we have been done in this study, the retention of 250 draws (and, hence, the maintenance of a given level of Monte Carlo accuracy) requires taking: $N = 210e03$ for $\alpha_1 = 1/T^{1.1}$, $N = 1.4e06$ for $\alpha_2 = 1/T^{3/2}$, $N = 13.5e06$ for $\alpha_3 = 1/T^2$ and $N = 41.0e06$ for $\alpha_4 = 1/T^{5/2}$. That is, the computational burden associated with decreasing the quantile in the manner indicated increases drastically: posteriors based on α_4 (for example) require a value of N that is three orders of magnitude greater than those based on α_1, but this increase in computational burden yields no, or minimal, gain in accuracy! The extension of such explorations to more scenarios is beyond the scope of this paper; however, we speculate that, with due consideration given to the properties of both the true data generating process and the chosen summary statistics and, hence, of the sample sizes for which Theorem 2 has practical content, the same sort of qualitative results will continue to hold.

Table 1. *Root mean square error of the estimated marginal posterior over 50 runs of the approximate Bayesian computation algorithm and across four different quantile choices; recorded as a ratio to the root mean square error for estimates based on the smallest quantile, $\alpha_4 = 1/T^{5/2}$. Numbers above one signify worse performance, and numbers below one signify better performance. For all replications the sample size is held fixed at $T = 100$ and each posterior estimate is based on 250 retained draws.*

	$\alpha_1 = 1/T^{1.1}$	$\alpha_2 = 1/T^{1.5}$	$\alpha_3 = 1/T^2$
$RMSE_\mu(\alpha_j)$	1.1733	0.9883	0.9793
$RMSE_\sigma(\alpha_j)$	0.8628	0.8663	0.9057

ACKNOWLEDGEMENT

This research has been supported by Australian Research Council Discovery and l'Institut Universitaire de France grants. We thank the participants at CFE-CMStatistics 2015 and MCMSki 2016 for very constructive comments on an earlier draft of the paper. A previous version of Theorem 3 contained an error, which was brought to our attention by Wentao Li and Paul Fearnhead. We are grateful to them for spotting this error. The third author is further affiliated with the University of Warwick.

BIBLIOGRAPHY

[1] Beaumont, M.A., Cornuet, J-M., Marin, J-M. and Robert, C.P. (2009). Adaptive Approximate Bayesian Computation, *Biometrika*, **96**, 983-990.

[2] Beaumont, M.A., Zhang, W. and Balding, D.J. (2002). Approximate Bayesian Computation in Population Genetics, *Genetics*, **162**, 2025-2035.

[3] Bhattacharya, R. N. and Rao, R. R. (1986) *Normal Approximation and Asymptotic Expansions* Wiley Series in Probability and Mathematical Statistics

[4] Biau, G., Cérou, F. and Guyader, A. (2015). New insights into Approximate Bayesian Computation, *Ann. Inst. H. Poincaré Probab. Statist.*, **51**, 376–403.

[5] Blum, M.G.B. (2010). Approximate Bayesian Computation: a Nonparametric Perspective, *J. Amer. Statist. Assoc.*, **105**, 1178-1187.

[6] Blum, M.G.B. and François, O. (2010). Non-linear Regression Models for Approximate Bayesian Computation, *Statistics and Computing*, **20**, 63–73.

[7] Blum, M.G.B., Nunes, M.A., Prangle, D. and Sisson, S.A. (2013). A Comparative Review of Dimension Reduction Methods in Approximate Bayesian Computation, *Statistical Science*, **28**, 189–208.

[8] Creel, M. and Kristensen, D. (2015). ABC of SV: Limited Information Likelihood Inference in Stochastic Volatility Jump-Diffusion Models, *J. Empir. Finance,* **31**, 85-108.

[9] Drovandi, C.C., Pettitt, A.N. and Faddy, M.J. (2011). Approximate Bayesian Computation Using Indirect Inference, *J. R. Statis Soc. C*, **60**, 1-21.

[10] Drovandi, C. C., Pettitt, A. N. and Lee, A. (2015). Bayesian Indirect Inference using a Parametric Auxiliary Model, *Statistical Science*, **30**, 72–95.

[11] Fearnhead, P. and Prangle, D. (2012). Constructing Summary Statistics for Approximate Bayesian Computation: Semi-automatic Approximate Bayesian Computation, *J. R. Statist. Soc. B*, **74**, 419–474.

[12] Gallant, A.R. and Tauchen, G. (1996). Which Moments to Match, *Econometric Theory*, **12**, 657–681.

[13] Gouriéroux, C., Monfort, A. and Renault, E. (1993). Indirect Inference, *J. Applied Econom.*, **85**, S85–S118.

[14] Heggland, K. and Frigessi, A. (2004). Estimating Functions in Indirect Inference, *J. R. Statist. Soc. B*, **66**, 447-462.

[15] Jasra, A. (2015). Approximate Bayesian Computation for a Class of Time Series Models, *International Statistical Review* , **83**, 405-435.

[16] Joyce, P. and Marjoram, P. (2008). Approximately Sufficient Statistics and Bayesian Computation, *Statistical applications in Genetics and Molecular Biology*, **7**, 1–16.

[17] Marin, J-M., Pudlo, P., Robert C. and Ryder, R. (2011). Approximate Bayesian Computation Methods, *Statistics and Computing*, **21**, 289–291.

[18] Marin, J-M., Pillai, N., Robert, C.P. and Rousseau, J. (2014). Relevant statistics for Bayesian model choice, *J. R. Statist. Soc. B*, **76**, 833–859.

[19] Marjoram, P., Molitor, J., Plagonal, V. and Tavaré, S. (2003). Markov Chain Monte Carlo Without Likelihoods, *Proceedings of the National Academy of Science USA*, **100**, 15324–15328.

[20] Nott D., Fan, Y., Marshall, L. and Sisson, S. (2014). Approximate Bayesian Computation and Bayes Linear Analysis: Towards High-dimensional ABC, *J. Comput. Graphical Statist.*, **23**, 65–86.

[21] Pritchard, J.K., Seilstad, M.T., Perez-Lezaun, A. and Feldman, M.W. (1999). Population Growth of Human Y Chromosomes: A Study of Y Chromosome Microsatellites, *Molecular Biology and Evolution*, **16**, 1791–1798.

[22] Robert, C.P. (2015). Approximate Bayesian Computation: a Survey on Recent Results, *Proceedings, MC-QMC2014.*

[23] Sisson S. and Fan, Y. (2011). Likelihood-free Markov Chain Monte Carlo, In *Handbook of Markov Chain Monte Carlo* (Eds. Brooks, Gelman, Jones, Meng). Chapman and Hall/CRC Press.

[24] Sisson, S., Fan, Y. and Tanaka, M. (2007). Sequential Monte Carlo without Likelihoods, *Proceedings of the National Academy of Sciences, USA*, **104**, 1760–1765.

[25] Tavaré, S., Balding, D.J., Griffiths, R.C. and Donnelly, P. (1997). Inferring Coalescence Times from DNA Sequence Data, *Genetics*, **145**, 505–518.

[26] Toni, T., Welch, D., Strelkowa, N., Ipsen, A. and Stumpf, M.P.H. (2009). Approximate Bayesian Computation Scheme for Parameter Inference and Model Selection in Dynamical Systems, *JRSS (Interface)* , **6**, 187-202.

[27] Wegmann, D., Leuenberger, C. and Excoffier, L. (2009). Efficient Approximate Bayesian Computation Coupled with Markov chain Monte Carlo with Likelihood, *Genetics*, **182**, 1207–1218.

A. SUPPLEMENTARY MATERIAL

A·1. *Proof of Theorem 1*

Proof of Theorem 1. Let $\varepsilon_T > 0$ and assume that $\boldsymbol{y} \in \Omega_\varepsilon = \{\boldsymbol{y}; d_2\{\boldsymbol{\eta}(\boldsymbol{y}), \boldsymbol{b}(\boldsymbol{\theta}_0)\} \leq \varepsilon_T/3\}$. From Assumption $[\mathbf{A1}]$ $P_0\left(\Omega_\varepsilon\right) = 1 + o(1)$. Consider the joint event $A_\varepsilon(\delta') = \{(\boldsymbol{z}, \boldsymbol{\theta}) : d_2\{\boldsymbol{\eta}(\boldsymbol{z}), \boldsymbol{\eta}(\boldsymbol{y})\} \leq \varepsilon_T\} \cap d_2\{\boldsymbol{b}(\boldsymbol{\theta}), \boldsymbol{b}(\boldsymbol{\theta}_0)\} > \delta'\}$. We have, that for all $(\boldsymbol{z}, \boldsymbol{\theta}) \in A_\varepsilon(\delta')$

$$\begin{aligned} d_2\{\boldsymbol{b}(\boldsymbol{\theta}), \boldsymbol{b}(\boldsymbol{\theta}_0)\} &\leq d_2\{\boldsymbol{\eta}(\boldsymbol{z}), \boldsymbol{\eta}(\boldsymbol{y})\} + d_2\{\boldsymbol{b}(\boldsymbol{\theta}), \boldsymbol{\eta}(\boldsymbol{z})\} + d_2\{\boldsymbol{b}(\boldsymbol{\theta}_0), \boldsymbol{\eta}(\boldsymbol{y})\} \\ &\leq 4\varepsilon_T/3 + d_2\{\boldsymbol{b}(\boldsymbol{\theta}), \boldsymbol{\eta}(\boldsymbol{z})\} \end{aligned}$$

so that $(\boldsymbol{z}, \boldsymbol{\theta}) \in A_\varepsilon(\delta')$ implies that

$$d_2\{\boldsymbol{b}(\boldsymbol{\theta}), \boldsymbol{\eta}(\boldsymbol{z})\} > \delta' - 4\varepsilon_T/3$$

and choosing $\delta' \geq 4\varepsilon_T/3 + t_\varepsilon$ leads to

$$P\left(A_\varepsilon(\delta')\right) \leq \int_\Theta P_{\boldsymbol{\theta}}\left(d_2\{\boldsymbol{b}(\boldsymbol{\theta}), \boldsymbol{\eta}(\boldsymbol{z})\} > t_\varepsilon\right) d\Pi(\boldsymbol{\theta}),$$

and

$$\begin{aligned} \Pi\left(d_2\{\boldsymbol{b}(\boldsymbol{\theta}), \boldsymbol{b}(\boldsymbol{\theta}_0)\} > 4\varepsilon_T/3 + t_\varepsilon | d_2\{\boldsymbol{\eta}(\boldsymbol{y}), \boldsymbol{\eta}(\boldsymbol{z})\} \leq \varepsilon_T\right) &:= \Pi_\varepsilon\left(d_2\{\boldsymbol{b}(\boldsymbol{\theta}), \boldsymbol{b}(\boldsymbol{\theta}_0)\} > 4\varepsilon_T/3 + t_\varepsilon | \boldsymbol{\eta}_0\right) \\ &\leq \frac{\int_\Theta P_\theta\left(d_2\{\boldsymbol{b}(\boldsymbol{\theta}), \boldsymbol{\eta}(\boldsymbol{z})\} > t_\varepsilon\right) d\Pi(\boldsymbol{\theta})}{\int_\Theta P_\theta\left(d_2\{\boldsymbol{\eta}(\boldsymbol{z}), \boldsymbol{\eta}(\boldsymbol{y})\} \leq \varepsilon_T\right) d\Pi(\boldsymbol{\theta})}. \end{aligned} \tag{A1}$$

Moreover, since

$$d_2\{\boldsymbol{\eta}(\boldsymbol{z}), \boldsymbol{\eta}(\boldsymbol{y})\} \leq d_2\{\boldsymbol{b}(\boldsymbol{\theta}), \boldsymbol{\eta}(\boldsymbol{z})\} + d_2\{\boldsymbol{b}(\boldsymbol{\theta}_0), \boldsymbol{\eta}(\boldsymbol{y})\} + d_2\{\boldsymbol{b}(\boldsymbol{\theta}), \boldsymbol{b}(\boldsymbol{\theta}_0)\} \leq \varepsilon_T/3 + \varepsilon_T/3 + d_2\{\boldsymbol{b}(\boldsymbol{\theta}), \boldsymbol{b}(\boldsymbol{\theta}_0)\}$$

provided $d_2\{\boldsymbol{b}(\boldsymbol{\theta}), \boldsymbol{\eta}(\boldsymbol{z})\} \leq \varepsilon_T/3$, then

$$\begin{aligned} \int_\Theta P_{\boldsymbol{\theta}}\left(d_2\{\boldsymbol{\eta}(\boldsymbol{z}), \boldsymbol{\eta}(\boldsymbol{y})\} \leq \varepsilon_T\right) d\Pi(\boldsymbol{\theta}) &\geq \int_{d_2\{\boldsymbol{b}(\boldsymbol{\theta}), \boldsymbol{b}(\boldsymbol{\theta}_0)\} \leq \varepsilon_T/3} P_{\boldsymbol{\theta}}\left(d_2\{\boldsymbol{\eta}(\boldsymbol{z}), \boldsymbol{\eta}(\boldsymbol{y})\} \leq \varepsilon_T/3\right) d\Pi(\boldsymbol{\theta}) \\ &\geq \Pi\left(d_2\{\boldsymbol{b}(\boldsymbol{\theta}), \boldsymbol{b}(\boldsymbol{\theta}_0)\} \leq \varepsilon_T/3\right) - \rho_T(\varepsilon_T/3) \int_{d_2\{\boldsymbol{b}(\boldsymbol{\theta}), \boldsymbol{b}(\boldsymbol{\theta}_0)\} \leq \varepsilon_T/3} c(\boldsymbol{\theta}) d\Pi(\boldsymbol{\theta}). \end{aligned}$$

If part **(i)** of Assumption **[A2]** holds,

$$\int_{d_2\{\boldsymbol{b}(\boldsymbol{\theta}),\boldsymbol{b}(\boldsymbol{\theta}_0)\}\leq\varepsilon_T/3} c(\boldsymbol{\theta})d\Pi(\boldsymbol{\theta}) \leq c_0\Pi\left(d_2\{\boldsymbol{b}(\boldsymbol{\theta}),\boldsymbol{b}(\boldsymbol{\theta}_0)\}\leq\varepsilon_T/3\right)$$

and for ε_T small enough,

$$\int_\Theta P_{\boldsymbol{\theta}}\left(d_2\{\boldsymbol{\eta}(\boldsymbol{z}),\boldsymbol{\eta}(\boldsymbol{y})\}\leq\varepsilon_T\right)d\Pi(\boldsymbol{\theta}) \geq \frac{\Pi\left(d_2\{\boldsymbol{b}(\boldsymbol{\theta}),\boldsymbol{b}(\boldsymbol{\theta}_0)\}\leq\varepsilon_T/3\right)}{2}, \tag{A2}$$

which, combined with (A1) and Assumption [**A3**], leads to

$$\Pi_\varepsilon\left(d_2\{\boldsymbol{b}(\boldsymbol{\theta}),\boldsymbol{b}(\boldsymbol{\theta}_0)\}>4\varepsilon_T/3+t_\varepsilon|\boldsymbol{\eta}_0\right)\lesssim\rho_T(t_\varepsilon)\varepsilon_T^{-D}\lesssim\frac{1}{M} \tag{A3}$$

by choosing $t_\varepsilon=\rho_T^{-1}(\varepsilon_T^D/M)$ with M large enough. If part **(ii)** of Assumption **[A2]** holds, a Hölder inequality implies that

$$\int_{d_2\{\boldsymbol{b}(\boldsymbol{\theta}),\boldsymbol{b}(\boldsymbol{\theta}_0)\}\leq\varepsilon_T/3} c(\boldsymbol{\theta})d\Pi(\boldsymbol{\theta}) \lesssim \Pi\left(d_2\{\boldsymbol{b}(\boldsymbol{\theta}),\boldsymbol{b}(\boldsymbol{\theta}_0)\}\leq\varepsilon_T/3\right)^{a/(1+a)}$$

and if ε_T satisfies

$$\rho_T(\varepsilon_T)=o(\varepsilon_T^{D/(1+a)})=O(\Pi\left(d_2\{\boldsymbol{b}(\boldsymbol{\theta}),\boldsymbol{b}(\boldsymbol{\theta}_0)\}\leq\varepsilon_T/3\right)^{1/(1+a)})$$

then (A3) remains valid. □

A·2. *Proof of Theorem* 2

Proof of Theorem 2. We work with $\boldsymbol{b}$ instead of $\boldsymbol{\theta}$ as the parameter, with injectivity of $\boldsymbol{\theta}\mapsto\boldsymbol{b}(\boldsymbol{\theta})$ required to re-state all results in terms of $\boldsymbol{\theta}$. Set $\boldsymbol{\Sigma}_T(\boldsymbol{\theta}_0)(\boldsymbol{\eta}(\boldsymbol{y})-\boldsymbol{b}_0)=Z_T^0$, for $\boldsymbol{b}_0=\boldsymbol{b}(\boldsymbol{\theta}_0)$, and $\boldsymbol{\eta}_0=\boldsymbol{\eta}(\boldsymbol{y})$. We control the ABC posterior expectation of non-negative and bounded functions $f_T(\boldsymbol{\theta}-\boldsymbol{\theta}_0)$:

$$\begin{aligned}E_{\Pi_\varepsilon}\left[f_T(\boldsymbol{\theta}-\boldsymbol{\theta}_0)|\boldsymbol{\eta}_0\right]&=\int f_T(\boldsymbol{\theta}-\boldsymbol{\theta}_0)d\Pi_\varepsilon\left(\theta|\boldsymbol{\eta}_0\right)\\&=\int f_T(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\mathbb{1}_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\leq\lambda_T}d\Pi_\varepsilon\left(\theta|\boldsymbol{\eta}_0\right)+o_P(1)\\&=\frac{\int_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\leq\lambda_T}p(\boldsymbol{\theta})f_T(\boldsymbol{\theta}-\boldsymbol{\theta}_0)P_{\boldsymbol{\theta}}\left(\|\boldsymbol{\eta}(\boldsymbol{z})-\boldsymbol{\eta}(\boldsymbol{y})\|\leq\varepsilon_T\right)d\boldsymbol{\theta}}{\int_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\leq\lambda_T}p(\boldsymbol{\theta})P_{\boldsymbol{\theta}}\left(\|\boldsymbol{\eta}(\boldsymbol{z})-\boldsymbol{\eta}(\boldsymbol{y})\|\leq\varepsilon_T\right)d\boldsymbol{\theta}}+o_P(1)\end{aligned}$$

where the second equality uses the posterior concentration of $\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|$ at the rate $\lambda_T\gg\varepsilon_T\vee 1/d_T(1)$. Now,

$$\begin{aligned}\boldsymbol{\Sigma}_T(\boldsymbol{\theta}_0)\{\boldsymbol{\eta}(\boldsymbol{z})-\boldsymbol{\eta}(\boldsymbol{y})\}&=\boldsymbol{\Sigma}_T(\boldsymbol{\theta}_0)\{\boldsymbol{\eta}(\boldsymbol{z})-\boldsymbol{b}(\boldsymbol{\theta})\}+\boldsymbol{\Sigma}_T(\boldsymbol{\theta}_0)\{\boldsymbol{b}(\boldsymbol{\theta})-\boldsymbol{b}_0\}-\boldsymbol{\Sigma}_T(\boldsymbol{\theta}_0)\{\boldsymbol{\eta}(\boldsymbol{y})-\boldsymbol{b}_0\}\\&=\boldsymbol{\Sigma}_T(\boldsymbol{\theta}_0)\{\boldsymbol{\eta}(\boldsymbol{z})-\boldsymbol{b}(\boldsymbol{\theta})\}+\boldsymbol{\Sigma}_T(\boldsymbol{\theta}_0)\{\boldsymbol{b}(\boldsymbol{\theta})-\boldsymbol{b}_0\}-Z_T^0.\end{aligned}$$

Set $Z_T=\boldsymbol{\Sigma}_T(\boldsymbol{\theta}_0)\{\boldsymbol{\eta}(\boldsymbol{z})-\boldsymbol{b}(\boldsymbol{\theta})\}$ and $Z_T^0=\boldsymbol{\Sigma}_T(\boldsymbol{\theta}_0)\{\boldsymbol{\eta}(\boldsymbol{y})-\boldsymbol{b}(\boldsymbol{\theta}_0)\}$. For fixed $\boldsymbol{\theta}$,

$$\|\boldsymbol{\Sigma}_T^{-1}(\boldsymbol{\theta}_0)\left[\boldsymbol{\Sigma}_T(\boldsymbol{\theta}_0)\{\boldsymbol{\eta}(\boldsymbol{z})-\boldsymbol{b}(\boldsymbol{\theta})\}-x\right]\|\asymp\|\boldsymbol{D}_T^{-1}\left[\boldsymbol{\Sigma}_T(\boldsymbol{\theta}_0)\{\boldsymbol{\eta}(\boldsymbol{z})-\boldsymbol{b}(\boldsymbol{\theta})\}-x\right]\|$$

and that

$$\boldsymbol{\Sigma}_T(\boldsymbol{\theta_0})\{\boldsymbol{b}(\boldsymbol{\theta})-\boldsymbol{b_0}\}-Z_T^0=\boldsymbol{\Sigma}_T(\boldsymbol{\theta_0})\boldsymbol{b}'(\boldsymbol{\theta_0})(\boldsymbol{\theta}-\boldsymbol{\theta_0})(1+o(1))-Z_T^0\in B$$

Case (i) : $\liminf_T d_T(1)\varepsilon_T=+\infty$. Consider $x(\theta)=\varepsilon_T^{-1}(\boldsymbol{b}(\boldsymbol{\theta})-\boldsymbol{b_0})$ and $f_T(\boldsymbol{\theta}-\boldsymbol{\theta}_0)=f\{\varepsilon_T^{-1}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\}$, where f is a nonnegative, continuous and bounded function. On the event $\Omega_{n,0}(M)=\{\|Z_T\|^0\leq M/2\}$ which has probability smaller than ϵ by choosing M large enough, we have that

$$P_{\boldsymbol{\theta}}\left(\|Z_T-Z_T^0\|\leq M\right)\geq P_{\boldsymbol{\theta}}\left(\|Z_T\|\leq M/2\right)\geq 1-\frac{c(\theta)}{M^\kappa}\geq 1-\frac{c_0}{M^\kappa}\geq 1-\epsilon$$

for all $\|\theta - \theta_0\| \le \lambda_T$. Since, $\boldsymbol{\eta}(\boldsymbol{z}) - \boldsymbol{\eta}(\boldsymbol{y}) = \boldsymbol{\Sigma}_T^{-1}(\boldsymbol{\theta}_0)(Z_T - Z_T^0) + \varepsilon_T x$, we have that on $\Omega_{n,0}$,

$$\begin{aligned} P_{\boldsymbol{\theta}}\left(\|\boldsymbol{\Sigma}_T^{-1}(\boldsymbol{\theta}_0)(Z_T - Z_T^0) + \varepsilon_T x\| \le \varepsilon_T\right) &\ge P_{\boldsymbol{\theta}}\left(\|\boldsymbol{\Sigma}_T^{-1}(\boldsymbol{\theta}_0)(Z_T - Z_T^0)\| \le \varepsilon_T(1 - \|x\|)\right) \\ &\ge P_{\boldsymbol{\theta}}\left(\|Z_T - Z_T^0\| \le d_T(1)\varepsilon_T(1 - \|x\|)\right) \ge 1 - \epsilon \end{aligned}$$

as soon as $\|x\| \le 1 - M/d_T(1)\varepsilon_T$ with M as above. This, combined with the continuity of $p(.)$ at θ_0 and condition **A4**, implies that

$$\begin{aligned} &\int f\{\varepsilon_T^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\} d\Pi_\varepsilon\left(\theta|\boldsymbol{\eta}_0\right) \\ &= \frac{\int_{\|\theta-\theta_0\| \le \lambda_T} f\{\varepsilon_T^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\} P_{\boldsymbol{\theta}}\left(\|\boldsymbol{\Sigma}_T^{-1}(\boldsymbol{\theta}_0)(Z_T - Z_T^0) + \varepsilon_T x\| \le \varepsilon_T\right) d\boldsymbol{\theta}}{\int_{\|\theta-\theta_0\| \le \lambda_T} P_{\boldsymbol{\theta}}\left(\|\boldsymbol{\Sigma}_T^{-1}(\boldsymbol{\theta}_0)(Z_T - Z_T^0) + \varepsilon_T x\| \le \varepsilon_T\right) d\boldsymbol{\theta}}(1 + o(1)) + o_P(1) \\ &= \frac{\int_{\|x(\theta)\| \le 1 - M/d_T(1)\varepsilon_T} f\{\varepsilon_T^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\} d\theta}{\int_{\|x(\theta)\| \le 1 - M/d_T(1)\varepsilon_T} d\theta}(1 + o(1)) \\ &+ \frac{\int_{\|\theta-\theta_0\| \le \lambda_T} 1\!\!1_{\|x(\theta)\| > 1 - M/(d_T(1)\varepsilon_T)} f\{\varepsilon_T^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\} P_{\boldsymbol{\theta}}\left(\|\boldsymbol{\Sigma}_T^{-1}(\boldsymbol{\theta}_0)(Z_T - Z_T^0) + \varepsilon_T x\| \le \varepsilon_T\right) d\boldsymbol{\theta}}{\int_{\|x(\theta)\| \le 1 - M/d_T(1)\varepsilon_T} d\boldsymbol{\theta}} + o_p(1). \end{aligned} \tag{A4}$$

The first term is approximately equal to

$$N_1 = \frac{\int_{\|\mathbf{b}(\varepsilon_T u + \theta_0) - \mathbf{b}_0\| \le 1} f(u) du}{\int_{\|\mathbf{b}(\varepsilon_T u + \theta_0) - \mathbf{b}_0\| \le 1} du}$$

and the regularity of the function $\boldsymbol{\theta} \to \boldsymbol{b}(\boldsymbol{\theta})$ implies that

$$\int_{\|\mathbf{b}(\varepsilon_T u + \theta_0) - \mathbf{b}_0\| \le \varepsilon_T} du = \int_{\|\boldsymbol{b}'(\boldsymbol{\theta_0})u\| \le 1} du + o(1) = \int_{u^t B_0 u \le 1} du + o(1)$$

with $B_0 = \boldsymbol{b}'(\boldsymbol{\theta_0})^t \boldsymbol{b}'(\boldsymbol{\theta_0})$. This leads to

$$N_1 = \frac{\int_{u^t B_0 u \le 1} f(u) du}{\int_{u^t B_0 u \le 1} du}.$$

We now show that the second term of the right hand side of (A4) converges to 0. We split it into an integral over $1 + M\{d_T(1)\varepsilon_T\}^{-1} \ge \|x(\theta)\| \ge 1 - M\{d_T(1)\varepsilon_T\}^{-1}$ and $1 + M\{d_T(1)\varepsilon_T\}^{-1} \le \|x(\theta)\|$. This leads, for the first part to an upper bound

$$N_2 \le \frac{\|f\|_\infty \int_{1 + M/(d_T(1)\varepsilon_T) \ge \|x(\theta)\| > 1 - M/(d_T(1)\varepsilon_T)} d\boldsymbol{\theta}}{\int_{\|x(\theta)\| \le 1 - M/d_T(1)\varepsilon_T} d\boldsymbol{\theta}} \lesssim \{d_T(1)\varepsilon_T\}^{-1} = o(1)$$

Finally for the third integral, if $\|x(\theta)\| > 1 + M\{d_T(1)\varepsilon_T\}^{-1}$ then

$$\begin{aligned} P_{\boldsymbol{\theta}}\left(\|\boldsymbol{\Sigma}_T^{-1}(\boldsymbol{\theta}_0)(Z_T - Z_T^0) + \varepsilon_T x\| \le \varepsilon_T\right) &\le P_{\boldsymbol{\theta}}\left(\|\boldsymbol{\Sigma}_T^{-1}(\boldsymbol{\theta}_0)(Z_T - Z_T^0)\| \ge \varepsilon_T\|x\| - \varepsilon_T\right) \\ \le P_{\boldsymbol{\theta}}\left(\|Z_T - Z_T^0\| \ge d_T(1)\varepsilon_T(\|x\| - 1)\right) &\le \frac{c_0}{(d_T(1)\varepsilon_T(\|x\| - 1))^\kappa}, \end{aligned}$$

which leads to

$$N_3 = \frac{\int_{\|\theta-\theta_0\|\leq\lambda_T} 1\!\!1_{\|x(\theta)\|>1+M/(d_T(1)\varepsilon_T)} f\{\varepsilon_T^{-1}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\} P_{\boldsymbol{\theta}}\left(\|\boldsymbol{\Sigma}_T^{-1}(\boldsymbol{\theta}_0)(Z_T - Z_T^0) + \varepsilon_T x\| \leq \varepsilon_T\right) d\boldsymbol{\theta}}{\int_{\|x(\theta)\|\leq 1-M/d_T(1)\varepsilon_T} d\boldsymbol{\theta}}$$

$$\lesssim M^{-\kappa}\varepsilon_T^{-d_\theta} \int_{2\geq\|x\|>1+M/d_T(1)\varepsilon_T} d\boldsymbol{\theta} + 2^\kappa \varepsilon_T^{-d_\theta} \int_{2\leq\|x(\boldsymbol{\theta})\|} \{d_T(1)\varepsilon_T\|x(\boldsymbol{\theta})\|\}^{-\kappa} d\boldsymbol{\theta}$$

$$\lesssim M^{-\kappa} + \varepsilon_T^{-d_\theta} \int_{c_1\varepsilon_T\leq\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|} \{d_T(1)\|\boldsymbol{b}'(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\|\}^{-\kappa} d\boldsymbol{\theta} \lesssim M^{-\kappa}$$

as soon as $\kappa > d_\theta$. Since M can be chosen arbitrarily large, putting together N_1, N_2 and N_3, we obtain that the ABC posterior distribution of $\varepsilon_T^{-1}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)$ is asymptotically uniform over the ellipse $\{x^t B_0 x \leq 1\}$ and **(i)** is proved.

Case (ii) : $+\infty > \lim_T d_T(1)\varepsilon_T = c > 0$ and $\lim_T d_T(k)\varepsilon_T = +\infty$. We consider $f_T(\boldsymbol{\theta}-\boldsymbol{\theta}_0) = 1\!\!1[\boldsymbol{\Sigma}_T(\boldsymbol{\theta}_0)\{\boldsymbol{b}(\boldsymbol{\theta})-\boldsymbol{b}_0\} - Z_T^0 \in B]$ and $x(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_T(\boldsymbol{\theta}_0)\{\boldsymbol{b}(\boldsymbol{\theta})-\boldsymbol{b}_0\} - Z_T^0$.

Set k_0 such that for all $j \leq k_0$, $\lim_T d_T(j)\varepsilon_T = c$ and for all $j > k_0$, $\lim_T d_T(j)\varepsilon_T = +\infty$. We write $\boldsymbol{\Sigma}_T(\boldsymbol{\theta}_0) = \boldsymbol{A}_T(\boldsymbol{\theta}_0)\boldsymbol{D}_T$, so that $\boldsymbol{A}_T(\boldsymbol{\theta}_0) \to_p \boldsymbol{A}(\boldsymbol{\theta}_0)$ as T goes to infinity, where $\boldsymbol{A}(\boldsymbol{\theta}_0)$ is symmetric.

$$P_{\boldsymbol{\theta}}\left(\|\boldsymbol{\Sigma}_T^{-1}(\boldsymbol{\theta}_0)\left(\boldsymbol{\Sigma}_T(\boldsymbol{\theta}_0)(\boldsymbol{\eta}(\boldsymbol{z})-\boldsymbol{b}(\boldsymbol{\theta})) - x\right)\| \leq \varepsilon_T\right) = P_{\boldsymbol{\theta}}\left(\|\boldsymbol{D}_T^{-1}\boldsymbol{A}_T^{-1}(\boldsymbol{\theta}_0)\left(Z_T - x\right)\| \leq \varepsilon_T\right)$$
$$= P_{\boldsymbol{\theta}}\left(\|\boldsymbol{D}_T^{-1}(\tilde{Z}_T - x_T)\| \leq \varepsilon_T\right)$$

where $\tilde{Z}_T = \boldsymbol{A}_T^{-1}(\boldsymbol{\theta}_0)Z_T \Rightarrow \mathcal{N}(0, \boldsymbol{A}(\boldsymbol{\theta})I_k\boldsymbol{A}(\boldsymbol{\theta})')$ and $x_T = \boldsymbol{A}_T^{-1}(\boldsymbol{\theta}_0)x = \boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)x + o(1)$.

We then have for $M_T \to +\infty$ such that $M_T d_T(k_0+1)^{-2}\varepsilon_T^2 = o(1)$.

$$\begin{aligned} P_{\boldsymbol{\theta}}\left(\|D_T^{-1}(\tilde{Z}_T - x_T)\| \leq \varepsilon_T\right) &\leq P_{\boldsymbol{\theta}}\left(\sum_{j=1}^{k_0}[\tilde{Z}_T(j) - x_T(j)]^2 \leq d_T(1)^2\varepsilon_T^2\right) \\ &\geq P_{\boldsymbol{\theta}}\left(\sum_{j=1}^{k_0}[\tilde{Z}_T(j) - x_T(j)]^2 \leq d_T(1)^2\varepsilon_T^2[1 - M_T\{d_T(k_0+1)\varepsilon_T^{-1}\}^{-2}]\right) \\ &\quad - P_{\boldsymbol{\theta}}\left(\sum_{j=k_0+1}^{k}[\tilde{Z}_T(j) - x_T(j)]^2 > M_T^{-1}[\varepsilon_T d_T(k_0+1)]^2\right) \\ &\geq P_{\boldsymbol{\theta}}\left(\sum_{j=1}^{k_0}(\tilde{Z}_T(j) - x_T(j))^2 \leq d_T(1)^2\varepsilon_T^2[1 - M_T^{-1}\{d_T(k_0+1)\varepsilon_T^{-1}\}^2]\right) - o(1) \end{aligned} \tag{A5}$$

This implies that for all x and all $\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| \leq \lambda_T$

$$P_{\boldsymbol{\theta}}\left(\|\boldsymbol{D}_T^{-1}(\tilde{Z}_T - x_T)\| \leq \varepsilon_T\right) = P_{\boldsymbol{\theta}}\left(\sum_{j=1}^{k_0}\left[\left\{\boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)Z_T\right\}(j) - \{\boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)x\}(j)\right]^2 \leq c\right) + o(1)$$

Since $\boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)x = \boldsymbol{D}_T\,\boldsymbol{b}'(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0) - \boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)Z_T^0$, if $Leb\left[\sum_{j=1}^{k_0}\left\{[\boldsymbol{b}'(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta_0})]_{[j]}\right\}^2 \leq c\varepsilon_T^2\right] = +\infty$, then as in the case **(i)** we can bound

$$\Pi_\varepsilon\left(\boldsymbol{\Sigma}_T(\boldsymbol{\theta}_0)(\boldsymbol{b}-\boldsymbol{b}_0) - Z_T^0 \in B|\boldsymbol{\eta}^0\right) \leq \frac{\int_{\boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)x\in B} P_{\boldsymbol{\theta}}\left(\sum_{j=1}^{k_0}\left[\left\{\boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)Z_T\right\}(j) - z(j)\right]^2 \leq c\right) d\boldsymbol{\theta}}{\int_{|\boldsymbol{\theta}|\leq M} P_{\boldsymbol{\theta}}\left(\sum_{j=1}^{k_0}[\{\boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)Z_T\}(j) - z(j)]^2 \leq c\right) d\boldsymbol{\theta}} + o_P(1)$$

which goes to zero when M goes to infinity. Since M can be chosen arbitrarily large, (12) is proven.

Case (iii) : $\lim_T d_T(1)\varepsilon_T = 0$ and $\lim_T d_T(k)\varepsilon_T = +\infty$. Again we consider $f_T(\boldsymbol{\theta} - \boldsymbol{\theta}_0) = \mathbb{1}[\boldsymbol{\Sigma}_T(\boldsymbol{\theta}_0)\{\boldsymbol{b}(\boldsymbol{\theta}) - \boldsymbol{b}_0\} - Z_T^0 \in B]$ and $x(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_T(\boldsymbol{\theta}_0)\{\boldsymbol{b}(\boldsymbol{\theta}) - \boldsymbol{b}_0\} - Z_T^0$. As in the computations leading to (A5), we have: setting $e_T = M_T(d_T(k_0+1)\varepsilon_T^{-1})^{-2}$, under Assumption **[A7]**,

$$
\begin{aligned}
P_{\boldsymbol{\theta}}\left(\|\boldsymbol{D}_T^{-1}(\tilde{Z}_T - x_T)\| \le \varepsilon_T\right) &\le P_{\boldsymbol{\theta}}\left(\sum_{j=1}^{k_1} d_T(j)^{-2}[\tilde{Z}_T(j) - x_T(j)]^2 \le \varepsilon_T^2\right) \\
&\ge P_{\boldsymbol{\theta}}\left(\sum_{j=1}^{k_1} d_T(j)^{-2}[\tilde{Z}_T(j) - x_T(j)]^2 \le \varepsilon_T^2(1-e_T)\right) \\
&= \varphi_{k_1}(x_{[k_1]})(1+o(1))\prod_{j=1}^{k_1}(d_T(j)\varepsilon_T),
\end{aligned}
$$

when $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < \lambda_T$ where φ_{k_1} is the centred Gaussian density of the k_1 dimensional vector, with covariance $(\boldsymbol{A}(\boldsymbol{\theta}_0)^2)_{[k_1]}$. This implies as in case **(ii)** that, with probability going to 1

$$
\limsup_T \Pi_\varepsilon\left(\boldsymbol{\Sigma}_T(\boldsymbol{\theta}_0)(\boldsymbol{b} - \boldsymbol{b}_0) - Z_T^0 \in B|\boldsymbol{\eta}_0\right) \le \frac{\int_{\boldsymbol{A}(\boldsymbol{\theta}_0)B}\varphi_{k_1}(x_{[k_1]})dx}{\int_{|x|\le M}\varphi_{k_1}(x_{[k_1]})dx} \le M^{-(k-k_1)}
$$

and choosing M arbitrary large leads to (11).

Case (iv) : If $\lim_T d_T(j)\varepsilon_T = c > 0$ for all $j \le k$. To prove (13), we use the computation of case **(ii)** with $k_0 = k$, so that (A5) implies that

$$
\begin{aligned}
P_{\boldsymbol{\theta}}\left(\|\boldsymbol{D}_T^{-1}(\tilde{Z}_T - x_T)\| \le \varepsilon_T\right) &= P_{\boldsymbol{\theta}}\left(\|\tilde{Z}_T - x_T\|^2 \le d_T(1)^2\varepsilon_T^2\right) \\
&= P\left(\|\boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)Z_T - \boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)x\|^2 \le c^2\right) + o(1)
\end{aligned}
$$

and for all $\delta > 0$, choosing M large enough, and when T is large enough

$$
\begin{aligned}
\Pi_\varepsilon\left(\boldsymbol{\Sigma}_T(\boldsymbol{\theta}_0)(\boldsymbol{b} - \boldsymbol{b}_0) - Z_T^0 \in B|\boldsymbol{\eta}^0\right) &\le \frac{\int_{x\in B} P_{\boldsymbol{\theta}}\left(\|\boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)Z_T - \boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)x\|^2 \le c^2\right)dx}{\int_{|x|\le M} P_{\boldsymbol{\theta}}\left(\|\boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)Z_T - \boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)x\|^2 \le c^2\right)dx} \\
&\ge \frac{\int_{x\in B} P_{\boldsymbol{\theta}}\left(\|\boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)Z_T - \boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)x\|^2 \le c^2\right)dx}{\int_{|x|\le M} P_{\boldsymbol{\theta}}\left(\|\boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)Z_T - \boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)x\|^2 \le c^2\right)dx + \delta} + o_P(1)
\end{aligned}
$$

Since M can be chosen arbitrarily large and since when M goes to infinity,

$$
\int_{|x|\le M} P_{\boldsymbol{\theta}}\left(\|\tilde{Z}_T - \boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)x\|^2 \le c^2\right)dx \to \int_{x\in\mathbb{R}^k} P_{\boldsymbol{\theta}}\left(\|\tilde{Z}_T - \boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)x\|^2 \le c^2\right)dx < +\infty,
$$

(13) is proved.

Case (v) : $\lim_T d_T(k)\varepsilon_T = 0$. Take $\boldsymbol{\Sigma}_T(\boldsymbol{\theta}_0) = \boldsymbol{A}_T(\boldsymbol{\theta}_0)\boldsymbol{D}_T$. For some $\delta > 0$ and all $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \le \delta$,

$$
\begin{aligned}
P_\theta\left(\left\|\boldsymbol{D}_T^{-1}[\boldsymbol{A}_T^{-1}(\boldsymbol{\theta}_0)Z_T - \boldsymbol{A}_T^{-1}(\boldsymbol{\theta}_0)x]\right\| \le \varepsilon_T\right) &= P_\theta\left(\left\|\boldsymbol{D}_T^{-1}[\boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)Z_T - \boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)x]\right\| \le \varepsilon_T\right) + o(1) \\
&= P_{\boldsymbol{\theta}}\left(\left\|\boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)Z_T - \boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)x\right\|^2 \le d_T^2(k)\varepsilon_T^2\right) + o(1) \\
&= P_{\boldsymbol{\theta}}\left(\left\{\boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)Z_T - \boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)x\right\} \in B_T\right) + o(1).
\end{aligned}
$$

From both assertions of Assumption **[A7]** and the Dominated Convergence Theorem, the above implies (for $k_1 = k$)

$$
\begin{aligned}
\frac{1}{\prod_{j=1}^k \varepsilon_T d_T(k)}\int_{x\in\mathcal{L}_T} P_{\boldsymbol{\theta}}\left(\left\{\boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)Z_T - \boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)x\right\} \in B_T\right) &= \int_{x\in\mathcal{L}_T}\varphi_k(x)dx + o(1) \\
&= 1 + o(1)
\end{aligned}
$$

Likewise, similar arguments yield

$$\frac{1}{\prod_{j=1}^{k}\varepsilon_T d_T(k)}\int_{x\in\mathcal{L}_T} 1\!\!1_{x\in B} P_{\boldsymbol{\theta}}\left(\left\{\boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)Z_T - \boldsymbol{A}^{-1}(\boldsymbol{\theta}_0)x\right\}\in B_T\right) = \int_{x\in\mathcal{L}_T} 1\!\!1_{x\in B}\varphi_k(x)dx + o(1)$$
$$= \Phi(B) + o(1).$$

Together, these two yield the desired result. □

A·3. *Proof of Theorem* 3

Proof of Theorem 3*, Case (i)* $d_T\varepsilon_T \to +\infty$. Defining $b = b(\theta)$, $b_0 = b(\theta_0)$ and $x = d_T(b - b_0) - Z_T^0$ with $Z_T^0 = d_T\{\eta(\boldsymbol{y}) - b(\theta_0)\}$, we will approximate the ratio

$$E_{\Pi_\varepsilon}[d_T(b-b_0)] - Z_T^0 = \frac{N_T}{D_T} = \frac{\int xP_x\left(|\eta(\boldsymbol{z})-\eta(\boldsymbol{y})|\le\varepsilon_T\right)p[b_0 + \{x+Z_T^0\}/d_T]dx}{\int P_x\left(|\eta(\boldsymbol{z})-\eta(\boldsymbol{y})|\le\varepsilon_T\right)p[b_0 + \{x+Z_T^0\}/d_T]dx}$$

We first approximate the numerator N_T: $d_T\{\eta(\boldsymbol{z}) - \eta(\boldsymbol{y})\} = d_T\{\eta(\boldsymbol{z}) - b\} + x$ and that $b = b_0 + (x + Z_T^0)/d_T$. Denote $Z_T = d_T\{\eta(\boldsymbol{z}) - b\}$, then

$$\begin{aligned} N_T &= \int xP_x\left(|\eta(\boldsymbol{z})-\eta(\boldsymbol{y})|\le\varepsilon_T\right)p[b_0 + \{x+Z_T^0\}/d_T]dx \\ &= \int_{|x|\le d_T\varepsilon_T - M} xP_x\left(|Z_T + x|\le d_T\varepsilon_T\right)p[b_0 + \{x+Z_T^0\}/d_T]dx \\ &+ \int_{|x|\ge d_T\varepsilon_T - M} xP_x\left(|Z_T + x|\le d_T\varepsilon_T\right)p[b_0 + \{x+Z_T^0\}/d_T]dx, \end{aligned} \tag{A6}$$

where the condition $\lim_T d_T\varepsilon_T = +\infty$ is used in the representation of the real line over which the integral defining N_T is specified.

We start by studying the first integral term in (A6). If $0 \le x \le d_T\varepsilon_T - M$ then

$$\begin{aligned} 1 \ge P_x\left(|Z_T + x|\le d_T\varepsilon_T\right) &= 1 - P_x\left(Z_T > d_T\varepsilon_T - x\right) - P_x\left(Z_T < -d_T\varepsilon_T - x\right) \\ &\ge 1 - 2(d_T\varepsilon_T - x)^{-\kappa}. \end{aligned}$$

Using a similar argument for $x \le 0$, we obtain, for all $|x| \le d_T\varepsilon_T - M$,

$$1 - 2(d_T\varepsilon_T - |x|)^{-\kappa} \le P_x\left(|Z_T + x|\le d_T\varepsilon_T\right) \le 1$$

and choosing M large enough implies that if $\kappa > 2$,

$$\begin{aligned} N_1 &= \int_{|x|\le d_T\varepsilon_T - M} xP_x\left(|Z_T + x|\le d_T\varepsilon_T\right)p[b_0 + \{x+Z_T^0\}/d_T]dx \\ &= \int_{|x|\le d_T\varepsilon_T - M} xp[b_0 + \{x+Z_T^0\}/d_T]dx + O(M^{-\kappa+2}) \end{aligned}$$

A Taylor expansion of $p[b_0 + \{x+Z_T^0\}/d_T]$ around $\gamma_0 = b_0 + Z_T^0/d_T$ then leads to

$$\begin{aligned} N_1 &= 2\sum_{j=1}^{k}\frac{p^{(2j-1)}(\gamma_0)}{(2j-1)!(2j+1)d_T^{2j-1}}(\varepsilon_T d_T)^{2j+1} + O(M^{-\kappa+2}) + o_P(1) + O(\varepsilon_T^{2+\beta}d_T^2) \\ &= 2d_T^2\sum_{j=1}^{k}\frac{p^{(2j-1)}(\gamma_0)}{(2j-1)!(2j+1)}\varepsilon_T^{2j+1} + O(M^{-\kappa+2}) + o_P(1) + O(\varepsilon_T^{2+\beta}d_T^2) \end{aligned}$$

We split the second integral of (A6) into $d_T\varepsilon_T - M \leq |x| \leq d_T\varepsilon_T + M$ and $|x| \geq d_T\varepsilon_T + M$. We treat the latter as before: with probability going to 1,

$$
\begin{aligned}
|N_3| &\leq \int_{|x|\geq d_T\varepsilon_T+M} |x| P_x\left(|Z_T + x| \leq d_T\varepsilon_T\right) p[b_0 + \{x + Z_T^0\}/d_T]dx \\
&\leq \int_{|x|\geq d_T\varepsilon_T+M} \frac{|x| c[b_0 + \{x + Z_T^0\}/d_T]}{(|x| - d_T\varepsilon_T)^\kappa} p[b_0 + \{x + Z_T^0\}/d_T]dx \\
&\leq c_0\|p\|_\infty \int_{d_T\varepsilon_T+M\leq|x|\leq\delta d_T} \frac{|x|}{(|x| - d_T\varepsilon_T)^\kappa} dx + \frac{d_T}{(\delta d_T)^{\kappa-1}} \int c(\theta)d\Pi(\theta) \lesssim M^{-\kappa+2} + O(d_T^{-\kappa+2})\,.
\end{aligned}
$$

Finally, we study the second integral term for N_T in (A6) over $d_T\varepsilon_T - M \leq |x| \leq d_T\varepsilon_T + M$. Using the assumption that $p(\cdot)$ is Hölder we obtain that

$$
\begin{aligned}
|N_2| &= \left| \int_{d_T\varepsilon_T-M}^{d_T\varepsilon_T+M} x P_x\left(|Z_T + x| \leq d_T\varepsilon_T\right) p[b_0 + \{x + Z_T^0\}/d_T]dx \right. \\
&\quad \left. + \int_{-d_T\varepsilon_T-M}^{-d_T\varepsilon_T+M} x P_x\left(|Z_T + x| \leq d_T\varepsilon_T\right) p[b_0 + \{x + Z_T^0\}/d_T]dx \right| \\
&\leq p(b_0) \left| \int_{d_T\varepsilon_T-M}^{d_T\varepsilon_T+M} x P_x\left(|Z_T + x| \leq d_T\varepsilon_T\right) dx + \int_{-d_T\varepsilon_T-M}^{-d_T\varepsilon_T+M} x P_x\left(|Z_T + x| \leq d_T\varepsilon_T\right) dx \right| \\
&\quad + LM\varepsilon_T^{1+\beta\wedge1} d_T^{\beta\wedge1} + o_P(1) \\
&\lesssim \left| d_T\varepsilon_T \int_{-M}^{M} \left[P_y\left(Z_T \leq -y\right) - P_y\left(Z_T \geq -y\right)\right] dy \right| \\
&\quad + \left| d_T\varepsilon_T \int_{-M}^{M} y\left[P_y\left(Z_T \leq -y\right) + P_y\left(Z_T \geq -y\right)\right] dy \right| + O(M\varepsilon_T^{1+\beta\wedge1} d_T^{\beta\wedge1}) + o_P(1),
\end{aligned}
$$

with M fixed but arbitrarily large. By the Dominated Convergence Theorem and the Gaussian limit of Z_T, for any arbitrarily large, but fixed M,

$$
\int_{-M}^{M} \left[P_y\left(Z_T \leq -y\right) - P_y\left(Z_T \geq -y\right)\right] dy = Mo(1)
$$

and

$$
\int_{-M}^{M} y\left[P_y\left(Z_T \leq -y\right) + P_y\left(Z_T \geq -y\right)\right] dy = \int_{-M}^{M} y\left(1 + o(1)\right) dy = M^2 o(1).
$$

This implies that

$$
N_2 \lesssim M^2 o(d_T\varepsilon_T) + M\varepsilon_T^{1+\beta\wedge1} d_T^{\beta\wedge1} + o_P(1)
$$

where the $o(\cdot)$ holds as T goes to infinity. Therefore, regrouping all terms, and since $\varepsilon_T^{1+\beta\wedge1} d_T^{\beta\wedge1} = o(d_T\varepsilon_T)$ for all $\beta > 0$ and $\varepsilon_T = o(1)$, we obtain the representation

$$
N_T = 2d_T^2 \sum_{j=1}^{k} \frac{p^{(2j-1)}(\gamma_0)}{(2j-1)!(2j+1)} \varepsilon_T^{2j+1} + M^2 o(d_T\varepsilon_T) + O(M^{-\kappa+2}) + O(d_T^{-\kappa+2}) + O(\varepsilon_T^{2+\beta} d_T^2) + o_P(1)\,.
$$

We now study the denominator in a similar manner. This leads to

$$\begin{aligned} D_T &= \int P_x\left(|\eta(\boldsymbol{z}) - \eta(\boldsymbol{y})| \leq \varepsilon_T\right) p[b_0 + \{x + Z_T^0\}/d_T] dx \\ &= \int_{|x| \leq d_T \varepsilon_T - M} p[b_0 + \{x + Z_T^0\}/d_T](1 + o(1)) dx + O(1) \\ &= 2p(b_0) d_T \varepsilon_T (1 + o_P(1)). \end{aligned}$$

Combining D_T and N_T we obtain, $\varepsilon_T = o(1)$,

$$\frac{N_T}{D_T} = d_T \sum_{j=1}^{k} \frac{p^{(2j-1)}(b_0)}{p(b_0)(2j-1)!(2j+1)} \varepsilon_T^{2j} + o_P(1) + O(\varepsilon_T^{1+\beta} d_T) \tag{A7}$$

Using the definition of N_T/D_T, dividing (A7) by d_T, and rearranging terms yields

$$E_{\Pi_\varepsilon}[b - b_0] = \frac{Z_T^0}{d_T} + \sum_{j=1}^{k} \frac{p^{(2j-1)}(b_0)}{p(b_0)(2j-1)!(2j+1)} \varepsilon_T^{2j} + O(\varepsilon_T^{1+\beta}) + o_P(1/d_T),$$

To obtain the posterior mean of θ, we write

$$\theta = b^{-1}[b(\theta)] = \theta_0 + \sum_{j=1}^{\lfloor \beta \rfloor} \frac{\{b(\theta) - b_0\}^j}{j!} (b^{-1})^{(j)}(b_0) + R(\theta)$$

where $|R(\theta)| \leq L|b(\theta) - b_0|^\beta$ provided $|b(\theta) - b_0| \leq \delta$. We compute the ABC mean of θ by splitting the range of integration in $|b(\theta) - b_0| \leq \delta$ and $|b(\theta) - b_0| > \delta$. A Cauchy-Schwarz inequality leads to

$$\begin{aligned} E_{\Pi_\varepsilon}\left[|\theta - \theta_0| \mathbb{1}_{|b(\theta) - b_0| > \delta}\right] &= \frac{1}{2\varepsilon_T d_T p(b_0)(1 + o_P(1))} \int_{|b(\theta) - b_0| > \delta} |\theta - \theta_0| P_\theta\left(|\eta(\boldsymbol{z}) - \eta(\boldsymbol{y})| \leq \varepsilon_T\right) p(\theta) d\theta \\ &\leq \frac{2^\kappa \left(\int_\Theta (\theta - \theta_0)^2 p(\theta) d\theta\right)^{1/2} \left(\int_\Theta c(\theta)^2 p(\theta) d\theta\right)^{1/2}}{d_T^\kappa \delta^\kappa} (1 + o_P(1)) \\ &= o_P(1/d_T) \end{aligned}$$

provided $\kappa > 1$. To control the former term, we use computations similar to earlier ones so that

$$E_{\Pi_\varepsilon}\left[\{\theta - \theta_0\} \mathbb{1}_{|b(\theta) - b_0| \leq \delta}\right] = \sum_{j=1}^{\lfloor \beta \rfloor} \frac{(b^{-1})^{(j)}(b_0)}{j!} E_{\Pi_\varepsilon}\left[\{b(\theta) - b_0\}^j\right] + o_P(1/d_T),$$

where, for $j \geq 2$ and $\kappa > j + 1$,

$$\begin{aligned} E_{\Pi_\varepsilon}\left[\{b(\theta) - b_0\}^j\right] &= \frac{1}{d_T^j} \frac{\int_{|x| \leq \varepsilon_T d_T - M} x^j p[b_0 + \{x + Z_T^0\}/d_T] dx}{2\varepsilon_T d_T p(b_0)} + o_P(1/d_T) \\ &= \sum_{l=0}^{k} \frac{p^{(l)}(b_0)}{2\varepsilon_T d_T^{j+l+1} p(b_0) l!} \int_{|x| \leq \varepsilon_T d_T - M} x^{j+l} dx + o_P(1/d_T) + O(\varepsilon_T^{1+\beta}) \\ &= \sum_{l=\lceil j/2 \rceil}^{\lfloor (j+k)/2 \rfloor} \frac{\varepsilon_T^{2l} p^{(2l-j)}(b_0)}{p(b_0)(2l-j)!} + o_P(1/d_T) + O(\varepsilon_T^{1+\beta}) \end{aligned}$$

This implies, in particular, that

$$E_{\Pi_\varepsilon}[\theta - \theta_0] = \frac{Z_T^0 (b^{-1})^{(1)}(b_0)}{d_T} + \sum_{j=1}^{\lfloor \beta \rfloor} \frac{(b^{-1})^{(j)}(b_0)}{j!} \sum_{l=\lceil j/2 \rceil}^{\lfloor (j+k)/2 \rfloor} \frac{\varepsilon_T^{2l} p^{(2l-j)}(b_0)}{p(b_0)(2l-j)!} + o_P(1/d_T) + O(\varepsilon_T^{1+\beta})$$

Hence, if $\varepsilon_T^2 = o(1/d_T)$ and $\beta \geq 1$,

$$E_{\Pi_\varepsilon}[\theta - \theta_0] = \frac{Z_T^0}{d_T b'(\theta_0)} + o_P(1/d_T)$$

and $E_{\Pi_\varepsilon}[d_T\{\theta - \theta_0\}] \Rightarrow \mathcal{N}(0, V_0/(b'(\theta_0))^2)$, while if $d_T\varepsilon_T^2 \to +\infty$

$$E_{\Pi_\varepsilon}[\theta - \theta_0] = \varepsilon_T^2 \left[\frac{p'(b_0)}{3p(b_0)b'(\theta_0)} - \frac{b^{(2)}(\theta_0)}{2(b'(\theta_0))^2}\right] + O(\varepsilon_T^4) + o_P(1/d_T),$$

assuming $\beta \geq 3$. □

Proof of Theorem 3, *Case (ii)* $d_T\varepsilon_T \to c$, $c \geq 0$. Recall that $b = b(\theta)$ and define

$$E_{\Pi_\varepsilon}[b] = \int \frac{bP_b\left(|\eta(\boldsymbol{y}) - \eta(\boldsymbol{z}))| \leq \varepsilon_T\right)p(b)db}{\int P_b\left(|\eta(\boldsymbol{y}) - \eta(\boldsymbol{z})| \leq \varepsilon_T\right)p(b)db}.$$

Considering the change of variables $b \mapsto x = d_T(b - b_0) - Z_T^0$ and using the above equation we have

$$E_{\Pi_\varepsilon}[b] = \int \frac{(b_0 + (x + Z_T^0)/d_T)P_x\left(|\eta(\boldsymbol{y}) - \eta(\boldsymbol{z})| \leq \varepsilon_T\right)p[b_0 + \{x + Z_T^0\}/d_T]dx}{\int P_x\left(|\eta(\boldsymbol{y}) - \eta(\boldsymbol{z})| \leq \varepsilon_T\right)p[b_0 + \{x + Z_T^0\}/d_T]dx},$$

which can be rewritten as

$$E_{\Pi_\varepsilon}[d_T\{b - b_0\}] - Z_T^0 = \int \frac{xP_x\left(|\eta(\boldsymbol{y}) - \eta(\boldsymbol{z})| \leq \varepsilon_T\right)p[b_0 + \{x + Z_T^0\}/d_T]dx}{\int P_x\left(|\eta(\boldsymbol{y}) - \eta(\boldsymbol{z})| \leq \varepsilon_T\right)p[b_0 + \{x + Z_T^0\}/d_T]dx}$$

Recalling that $d_T(\eta(\boldsymbol{z}) - \eta(\boldsymbol{y})) = d_T(\eta(\boldsymbol{z}) - b) + d_T(b - b_0) - Z_T^0 = Z_T + x$ we have

$$E_{\Pi_\varepsilon}[d_T\{b - b_0\}] - Z_T^0 = \int \frac{xP_x\left(|Z_T + x| \leq d_T\varepsilon_T\right)p[b_0 + \{x + Z_T^0\}/d_T]dx}{\int P_x\left(|Z_T + x| \leq d_T\varepsilon_T\right)p[b_0 + \{x + Z_T^0\}/d_T]dx} = \frac{N_T}{D_T}.$$

By injectivity of the map $\theta \mapsto b(\theta)$ (Assumptions [**A4**]) and [**A5**]), the result follows when $E_{\Pi_\varepsilon}[d_T\{b - b_0\}] - Z_T^0 = o_P(1)$.

Consider first the denominator. Defining $h_T = d_T\varepsilon_T$ and using arguments that mirror those in the proof of Theorem 2 part (v), by Assumptions [**A7**] and the DCT

$$\frac{D_T}{p(b_0)h_T} = h_T^{-1}\int P_x(|Z_T + x| \leq h_T)dx + o_P(1) = \int \varphi[x/A(\theta^0)]dx + o_P(1) = 1 + o_P(1),$$

where the second equality follows from Assumption [**A7**] and the DCT. The result now follows if $N_T/h_T = o_P(1)$. To this end, define $P_x^*(|Z_T + x| \leq h_T) = P_x(|Z_T + x| \leq h_T)/h_T$ and, if $h_T = o(1)$ by [**A7**(i)] and [**A8**],

$$\begin{aligned}\frac{N_T}{h_T} &= \int xP_x^*(|Z_T + x| \leq h_T)p[b_0 + \{x + Z_T^0\}/d_T]dx \\ &= p(b_0)\int x\varphi[x/A(\theta^0)]dx + \int x\left\{P_x^*(|Z_T + x| \leq h_T) - \varphi[x/A(\theta^0)]\right\} \\ &\quad p[b_0 + \{x + Z_T^0\}/d_T]dx + o_P(1).\end{aligned}$$

If $h_T \to c > 0$ then

$$\begin{aligned}\frac{N_T}{h_T} &= p(b_0)\int xP\{|\mathcal{N}(0,1) + x/A(\theta^0)| \leq c/A(\theta^0)\}dx \\ &\quad + \int x\left[P_x^*(|Z_T + x| \leq h_T) - P\{|\mathcal{N}(0,1) + x/A(\theta^0)| \leq c/A(\theta^0)\}\right]p[b_0\{x + Z_T^0\}/d_T]dx + o_P(1).\end{aligned} \tag{A8}$$

The result now follows if $\int x \left\{P_x^*(|Z_T + x| \le h_T) - \varphi[x/A(\theta^0)]\right\} p[b_0 + \{x + Z_T^0\}/d_T]dx = o_P(1)$ (resp. $P_x^*(|Z_T + x| \le h_T) - P\{|\mathcal{N}(0,1) + x/A(\theta^0)| \le c/A(\theta^0)\} = o(1)$) for which a sufficient condition is that

$$\int |x| \left|P_x^*(|Z_T + x| \le h_T) - \varphi[x/A(\theta^0)]\right| p[b_0 + \{x + Z_T^0\}/d_T]dx = o_P(1), \tag{A9}$$

or the equivalent in the case $h_T \to c > 0$.

To show that the integral in (A9) is $o_P(1)$ we break the region of integration into three areas: (i) $|x| \le M$; (ii) $M \le |x| \le \delta d_T$; (iii) $|x| \ge \delta d_T$.

Area (i): Over $|x| \le M$, the following equivalences are satisfied:

$$\sup_{x:|x|\le M} |p[b_0 + \{x + Z_T^0\}/d_T] - p(b_0)| = o_P(1)$$
$$\sup_{x:|x|\le M} |P^*(|Z_T + x| \le h_T) - \varphi[x/A(\theta^0)]| = o_P(1).$$

The first equation is satisfied by **[A8]** and the fact that by **[A5]** $Z_T^0/d_T = o_P(1)$. The second term follows from **[A7]** and DCT. We can now conclude that equation (A9) is $o_P(1)$ over $|x| \le M$.

The same holds for (A8), without the need of Assumption **[A7]**.

Area (ii): Over $M \le |x| \le \delta d_T$ the integral of the second term is finite and can be made arbitrarily small for large M enough. Therefore, it suffices to show that

$$\int_{M\le|x|\le\delta d_T} |x|P^*(|Z_T + x| \le h_T)p[b_0 + \{x + Z_T^0\}/d_T]dx$$

if finite.

When $|x| > M$, $|Z_T + x| \le h_T$ implies that $|Z_T| > |x|/2$ since $h_T = O(1)$. Hence using Assumption **[A2]**,

$$|x|P_x^*(|Z_T + x| \le h_T) \le |x|P_x^*(|Z_T| > |x|/2) \le c_0 \frac{|x|}{|x|^\kappa}$$

which in turns implies that

$$\int_{M\le|x|\le\delta d_T} P^*(|Z_T + x| \le h_T)p[b_0 + \{x + Z_T^0\}/d_T]dx \le C \int_{M\le|x|\le\delta d_T} \frac{1}{|x|^{\kappa-1}}dx \le M^{-\kappa+2}$$

The same computation can be conducted in the case (A8).

Area (iii): Over $|x| \ge \delta d_T$ the second term is again negligible for δd_T large. Our focus then becomes

$$N_3 = \frac{1}{h_T} \int_{|x|\ge\delta d_T} |x|P_x^*(|Z_T + x| \le h_T)p[b_0 + \{x + Z_T^0\}/d_T]dx.$$

For some $\kappa > 2$ we can bound N_3 as follows:

$$\begin{aligned}
N_3 =& \frac{1}{h_T} \int_{|x|\ge\delta d_T} |x|P_x(|x + Z_T| \le h_T)p[b_0 + \{x + Z_T^0\}/d_T]dx \\
\le& \frac{1}{h_T} \int_{|x|\ge\delta d_T} \frac{|x|c(b_0 + (x + Z_T^0)/d_T)}{(1 + |x| - h_T)^\kappa} p[b_0 + \{x + Z_T^0\}/d_T]dx \\
\lesssim& \frac{d_T^2}{h_T} \int_{|b-\eta(\boldsymbol{y})|\ge\delta} \frac{c(b)|b - \eta(\boldsymbol{y})|}{[1 + d_T|b - \eta(\boldsymbol{y})| - h_T]^\kappa} p(b)db
\end{aligned}$$

Since $\eta(\boldsymbol{y}) = b_0 + O_P(1/d_T)$ we have, for T large,

$$N_3 \lesssim \frac{d_T^2}{h_T} \int_{|b-b_0|\ge\delta/2} \frac{c(b)|b|p(b)}{(1 + d_T\delta - h_T)^\kappa}db \lesssim \frac{d_T^2}{h_T} \left[\int c(b)|b|p(b)db\right] O(d_T^{-k}) \lesssim O(d_T^{1-\kappa}\varepsilon_T) = o(1),$$

where **[A2]** and **[A8]** ensure $\int c(b)|b|p(b)db < \infty$. The same computation can be conducted in the case (A8).

Combining the results for the three areas we can conclude that $N_T/D_T = o_P(1)$ and the result follows. □