MONASH University

MONASH BUSINESS SCHOOL

ISSN 1440-771X

Department of Econometrics and Business Statistics

http://business.monash.edu/econometrics-and-business-statistics/research/publications

Asymptotic Properties of Approximate Bayesian Computation

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

September 2017

Working Paper 12/17

AACSB ACCREDITED EFMD EQUIS ACCREDITED ASSOCIATION OF MBAS ACCREDITED

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 is becoming an accepted tool for statistical analysis in models with intractable likelihoods. With the initial focus being primarily on the practical import of this algorithm, exploration of its formal statistical properties has begun to attract more attention. In this paper we consider the asymptotic behaviour of the posterior distribution obtained by this method. We give general results on: (i) the rate at which the posterior concentrates on sets containing the true parameter (vector); (ii) the limiting shape of the posterior; and (iii) the asymptotic distribution of the ensuing posterior mean. These results hold under given rates for the tolerance used within the method, 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 are discussed.

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

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

1. INTRODUCTION

The use of approximate Bayesian computation methods in models with intractable likelihoods has gained increased momentum over recent years, extending beyond the original genetics applications. (See Marin et al., 2011, Sisson and Fan, 2011 and Robert, 2015, for recent reviews.) Whilst this approach initially appeared as a practical solution, attention has now shifted to the investigation of its formal statistical properties, especially in relation to the choice of summary statistics on which the technique typically relies; see, for example, Fearnhead and Prangle (2012), 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 denote these preprints by Creel et al. (2015), Li and Fearnhead (2016) and Martin et al. (2016).

This paper studies the large sample properties of both posterior distributions and posterior means obtained from approximate Bayesian computation algorithms. Under mild regularity conditions on the underlying summary statistics, 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 require asymptotic normality of the posterior, with the former result being attainable under weaker conditions on the tolerance than required for the latter. Validity of these results requires that the summaries converge toward a limit at a known rate, and that this limit, viewed as a mapping from parameters to summaries, be injective. These conditions have a close correspondence with those required for theoretical validity of indirect inference and related (frequentist) estimators (Gouriéroux et al., 1993; Gallant and Tauchen, 1996).

We focus on three aspects of asymptotic behaviour: posterior consistency, limiting posterior shape, and the asymptotic distribution of the posterior mean. This focus is broader than that of existing studies on the large sample properties of approximate Bayesian computation algorithms, in which the asymptotic properties of resulting point estimators have been the primary focus; see Creel et al. (2015), Jasra (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 summaries and the tolerance) required for concentration and those required for specific distributional results. Throughout the paper, 'posterior distribution' refers to the posterior distribution resulting from an approximate Bayesian computation algorithm.

2. PRELIMINARIES AND BACKGROUND

We observe data $y = (y_1, y_2, ..., y_T)^\intercal$, $T \geq 1$, drawn from the model $\{P_\theta : \theta \in \Theta\}$, where P_θ admits the corresponding conditional density $p(\cdot|\theta)$, and $\theta \in \Theta \subset \mathbb{R}^{k_\theta}$. Given a prior $p(\theta)$, the aim of the algorithms under study is to produce draws from an approximation to the posterior distribution $p(\theta|y) \propto p(y|\theta)p(\theta)$, in the case where both the parameters and pseudo-data (θ, z) can be easily simulated from $p(\theta)p(z|\theta)$, but where $p(z|\theta)$ 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

(1) Simulate θ^i, $i = 1, 2, ..., N$, from $p(\theta)$,
(2) Simulate $z^i = (z_1^i, z_2^i, ..., z_T^i)^\intercal$, $i = 1, 2, ..., N$, from the likelihood, $p(\cdot|\theta^i)$
(3) Select θ^i such that $d\{\eta(y), \eta(z^i)\} \leq \varepsilon$,where $\eta(\cdot)$ is a (vector) statistic, $d(\cdot, \cdot)$ is a distance function (or metric), and $\varepsilon > 0$ is the tolerance level.

Algorithm 1 thus samples θ and z from the joint posterior:

$$p_\varepsilon\{\theta, z|\eta(y)\} = p(\theta)p(z|\theta)\mathbb{1}_\varepsilon(z) \big/ \textstyle\int\int p(\theta)p(z|\theta)\mathbb{1}_\varepsilon(z)dzd\theta,$$

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

$$p_\varepsilon\{\theta|\eta(y)\} = \textstyle\int p_\varepsilon\{\theta, z|\eta(y)\}dz \tag{1}$$

approximates the exact posterior, $p(\theta|y)$, and draws of θ from $p_\varepsilon\{\theta, z|\eta(y)\}$ can be used to estimate features of $p(\theta|y)$. For example, using $p_\varepsilon\{\theta|\eta(y)\}$, one can estimate the posterior probability of the set $A \subset \Theta$ by calculating

$$\Pi_\varepsilon\{A|\eta(y)\} = \Pi\left[A|d\{\eta(y), \eta(z)\} \leq \varepsilon\right] = \int_A p_\varepsilon\{\theta|\eta(y)\}d\theta.$$

In practice however, models to which Algorithm 1 is applied are such that sufficiency is unavailable. Hence, the draws of θ can only be used to approximate $p_\varepsilon\{\theta|\eta(y)\}$, which differs from $p(\theta|y)$ (even as $\varepsilon \to 0$). Given the lack of accordance between $p_\varepsilon\{\theta|\eta(y)\}$ and $p(\theta|y)$, a means of assessing the behavior of p_ε in its own right, and of establishing whether or not p_ε behaves in a manner that is appropriate for statistical inference, is required. A reasonable means by which to gauge the statistical behavior of p_ε is asymptotic theory. Establishing the large sample behavior of p_ε, including point and interval estimates derived from p_ε, gives practitioners a set of guarantees on the reliability of approximate Bayesian computations. Moreover, by carefully studying the implications of these theoretical conclusions, we provide guidelines for producing approximate posteriors, via the implementation of Algorithm 1, that possess desirable statistical properties.

3. CONCENTRATION OF THE APPROXIMATE BAYESIAN COMPUTATION POSTERIOR

First, we set some notation used throughout the paper. Define $\mathcal{Z}$ as the space of simulated data; $\mathcal{B} = \{\eta(z) : z \in \mathcal{Z}\} \subset \mathbb{R}^{k_\eta}$ the range of the simulated summaries: $\eta(z) : \mathcal{Z} \to \mathcal{B}$; $d_1\{\cdot,\cdot\}$ a metric on Θ; $d_2\{\cdot,\cdot\}$ a metric on $\mathcal{B}$; $C > 0$ a generic constant; P_0 the measure generating y and P_θ the measure generating $z(\theta)$. We have $P_\theta = P_0$ for $\theta = \theta_0$, and denote $\theta_0 \in \text{Int}(\Theta)$ as the true parameter value. Let $\Pi(\theta)$ denote the prior measure with density $p(\theta)$.

For real-valued sequences $\{a_T\}_{T\geq 1}$ and $\{b_T\}_{T\geq 1}$, $a_T \lesssim b_T$ denotes $a_T \leq C b_T$ for some finite $C > 0$ and all T large, $a_T \asymp b_T$ denotes an equivalent order of magnitude, i.e., for some C, $a_T/b_T \to C$ as $T \to +\infty$, and $a_T \gg b_T$ indicates a larger order of magnitude. For x_T a random variable, $x_T = o_P(a_T)$ if $\lim_{T\to+\infty} \text{pr}(|x_T/a_T| \geq C) = 0$ for any $C > 0$ and $x_T = O_P(a_T)$ if for any $C \geq 0$ there exists a finite $M > 0$ such that $\text{pr}(|x_T/a_T| \geq M) \leq C$. The symbol $\|\cdot\|$ denotes the Euclidean norm.

At its most fundamental level, asymptotic validity of any Bayesian procedure requires Bayesian (or posterior) consistency. In our context Bayesian consistency equates with the following posterior concentration property: for any $\delta > 0$, as $T \to +\infty$

$$\Pi\left[d_1\{\theta,\theta_0\} > \delta | d_2\{\eta(y),\eta(z)\} \leq \varepsilon\right] = \int_{d_1\{\theta,\theta_0\}>\delta} p_\varepsilon\{\theta|\eta(y)\} d\theta = o_P(1). \quad (2)$$

This property is paramount in this setting since, for any $A \subset \Theta$, $\Pi\left[A | d_2\{\eta(y),\eta(z)\} \leq \varepsilon\right]$ differs from the exact posterior probability in a manner that can rarely be quantified. Without the reassurance of exact posterior inference, knowledge that the posterior obtained from Algorithm 1 will concentrate, in large samples, on the true value θ_0 generating the data becomes even more critical than if $p(\theta|y)$ were accessible.

Posterior concentration is related to the rate at which information about θ_0 accumulates in the sample. In Algorithm 1, information about θ_0 is not obtained from the intractable likelihood but through $d_2\{\eta(y),\eta(z)\}$. For a fixed k_η-dimensional summary $\eta(z)$, the posterior learns about θ_0 if $\eta(z)$ concentrates around some fixed value $b(\theta)$. Therefore, the amount of information Algorithm 1 provides about θ_0 depends on two factors: (1) the rate at which the observed and simulated summaries converge to well-defined limit counterparts $b(\theta_0)$ and $b(\theta)$; and (2) the rate at which information about θ_0 accumulates within the algorithm, governed by the rate at which ε goes to 0. To link both factors we consider ε as a T-dependent sequence $\varepsilon_T \to 0$ as $T \to +\infty$. We can now state the technical assumptions used to establish our first result, with a discussion of these assumptions to follow.

[A1] There exist a non-random map $b : \Theta \to \mathcal{B}$, and a function $\rho_T(u)$ with $\rho_T(u) \to 0$ as $T \to +\infty$ for all u and $\rho_T(u)$ monotone non-increasing in u (for any given T), such that for all $\theta \in \Theta$

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

with either of the following assumptions on $c(\cdot)$:
(i) There exist $c_0 < +\infty$ and $\delta > 0$ such that for all θ satisfying $d_2\{b(\theta), b(\theta_0)\} \leq \delta$ then $c(\theta) \leq c_0$.
(ii) There exists $a > 0$ such that $\int_\Theta c(\theta)^{1+a} d\Pi(\theta) < +\infty$.

[A2] There exists some $D > 0$ such that, for all $\xi > 0$ small enough, the prior probability satisfies

$$\Pi\left[d_2\{b(\theta), b(\theta_0)\} \leq \xi\right] \gtrsim \xi^D.$$

[A3] (i) The map b is continuous. (ii) The map b is injective and satisfies

$$\|\theta - \theta_0\| \leq L\|b(\theta) - b(\theta_0)\|^\alpha$$

on some open neighbourhood of θ_0 with $L > 0$ and $\alpha > 0$.

Remark 1: Assumptions [A1]-[A3] are applicable to a broad range of data structures, including, e.g., weakly dependent data. [A1] ensures that $\eta(z)$ concentrates on $b(\theta)$. This concentration is the engine behind posterior concentration and without [A1], or a similar assumption, Bayesian consistency will not occur. Assumption [A2] controls the degree of prior mass in a neighbourhood of θ_0 and is standard in Bayesian asymptotics. For ε_T small, the larger D, the smaller the amount of prior mass near θ_0. If Π is absolutely continuous with prior density $p(\theta)$ and if p is bounded, above and below, near θ_0, then $D = \dim(\theta) = k_\theta$. [A3] is an identification condition that is critical for obtaining posterior concentration around θ_0, where the injectivity of b depends on both the true structural model and the particular choice of η.

The following theorem details the behavior of $\Pi[\cdot|d_2\{\eta(y),\eta(z)\} \le \varepsilon_T]$ under [A1]-[A3].

THEOREM 1. *Assume that [A2] is satisfied. If [A1](i) holds with $\rho_T(\varepsilon_T) = o(1)$ or if [A1](ii) holds for constant a such that $\rho_T(\varepsilon_T) = o(\varepsilon_T^{D/(1+a)})$, then, for any M large enough,*

$$\Pi\left[d_2\{b(\theta), b(\theta_0)\} > 4\varepsilon_T/3 + \rho_T^{-1}(\varepsilon_T^D/M)|d_2\{\eta(y),\eta(z)\} \le \varepsilon_T\right] \lesssim 1/M. \tag{3}$$

Moreover, if [A3] holds then

$$\Pi\left[d_1\{\theta,\theta_0\} > L\{4\varepsilon_T/3 + \rho_T^{-1}(\varepsilon_T^D/M)\}^\alpha|d_2\{\eta(y),\eta(z)\} \le \varepsilon_T\right] \lesssim 1/M. \tag{4}$$

Equations (3) and (4) imply Bayesian consistency and allow us to deduce a posterior concentration rate, denoted generically by λ_T and depending on ε_T and the deviation control ρ_T on $d_2\{\eta(z), b(\theta)\}$. The posterior $\Pi\left[\cdot|d_2\{\eta(y),\eta(z)\} \le \varepsilon_T\right]$ concentrates at rate $\lambda_T \to 0$ if, for ε_T as in Theorem 1 and M sufficiently large,

$$\limsup_{T\to+\infty} \Pi\left[d_1\{\theta,\theta_0\} > \lambda_T M|d_2\{\eta(y),\eta(z)\} \le \varepsilon_T\right] = o_P(1).$$

To deduce this rate one need only solve, for a given deviation control function, $\lambda_T \asymp \varepsilon_T + \rho_T^{-1}(\varepsilon_T^D)$. To demonstrate this interplay between λ_T and ρ_T, consider the following two situations for [A1]:
(a) *Polynomial deviations:* There exist $v_T \to +\infty$ and $u_0, \kappa > 0$ such that

$$\rho_T(u) = 1/v_T^\kappa u^\kappa, \quad u \le u_0. \tag{5}$$

From (5) we have $\rho_T^{-1}(\varepsilon_T^D) = 1/v_T\varepsilon_T^{D/\kappa}$, so that equating (in order) ε_T and $\rho_T^{-1}(\varepsilon_T^D)$ yields $\varepsilon_T \asymp v_T^{-\kappa/(\kappa+D)}$. Choosing a tolerance ε_T of this order implies, in turn, that the posterior distribution of $b(\theta)$ concentrates at the rate

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

to $b(\theta_0)$. Under Assumption [A1](ii), the same rate can be achieved for all $a > 0$. Since $\varepsilon_T \asymp v_T^{-\kappa/(\kappa+D)}$ implies $\varepsilon_T \gg v_T^{-1}$, this ensures that the deviation control given in equation (5) satisfies Assumption [A1].
(b) *Exponential deviations:* there exist $h_\theta(\cdot) > 0$ and $v_T \to +\infty$ such that [A1] is satisfied with

$$\rho_T(u) = \exp\{-h_\theta(uv_T)\}, \tag{6}$$

and there exist finite $u_0, c, C > 0$ such that

$$\int_\Theta c(\theta)e^{-h_\theta(uv_T)}d\Pi(\theta) \le Ce^{-c(uv_T)^\tau}, \quad u \le u_0.$$

Hence if $c(\theta)$ is bounded from above and $h_\theta(u) \ge u^\tau$ for θ in a neighbourhood of the set $\{\theta; b(\theta) = b(\theta_0)\}$, then $\rho_T(u) \asymp e^{-c_0(uv_T)^\tau}$; thus, $\rho_T^{-1}(\varepsilon_T^D) \asymp \{\log(1/\varepsilon_T)\}^{1/\tau}/v_T$. Following similar arguments to those used in (a) above it follows that if we take $\varepsilon_T \asymp \{\log(v_T)\}^{1/\tau}/v_T$, the posterior distribution concentrates at the equivalent rate,

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

Under Assumption [A1](ii), the same rate can be achieved for all $a > 0$. Once again, $\varepsilon_T \asymp \{\log(v_T)\}^{1/\tau}/v_T \gg v_T^{-1}$ ensures that (6) will satisfy Assumption [A1].

To illustrate the above situations for [A1], we consider $\eta(z) = T^{-1}\sum_{i=1}^{T} g(z_i)$ and, for simplicity, let $\{g(z_i)\}_{i\leq T}$ be independent and identically distributed. First, consider the case of polynomial deviations and assume $g(z_i)$ has a finite moment of order κ for $\theta \in \text{Int}(\Theta)$. In this case $b(\theta) = E_\theta\{g(Z)\}$. Furthermore, the Markov inequality implies that

$$P_\theta\left\{\|\eta(z) - b(\theta)\| > u\right\} \leq CE_\theta\left\{|g(Z)|^\kappa\right\}/(T^{1/2}u)^\kappa,$$

and, with reference to equation (5), $v_T = T^{1/2}$. Then, if the map $\theta \mapsto E_\theta\left\{|g(Z)|^\kappa\right\}$ is continuous at θ_0 and positive, [A1](i) and (ii) (for all $a > 0$) are satisfied. On the other hand, if $|g(Z)|$ allows for an exponential moment we can consider exponential deviations: $E_\theta\left\{g(Z)^2 e^{a_\theta|g(Z)|}\right\} \leq c(\theta) < +\infty$, then for $a_\theta T^{1/2} \geq s > 0$,

$$P_\theta\left\{\|\eta(z) - b(\theta)\| > u\right\} \leq e^{-suT^{1/2}}\left[1 + \frac{s^2}{2T}E_\theta\left\{g(Z)^2 e^{s|g(Z)|/T^{1/2}}\right\}\right]^T$$
$$\leq e^{-suT^{1/2}+s^2c(\theta)/2} \leq e^{-u^2T/\{2c(\theta)\}},$$

choosing $s = uT^{1/2}/c(\theta) \leq a_\theta T^{1/2}$, provided $u \leq a_\theta c(\theta)$. Thus, with reference to (6), $v_T = T^{1/2}$ and $h_\theta(uv_T) = u^2v_T^2/\{2c(\theta)\}$. If the maps $\theta \mapsto a_\theta$ and $\theta \mapsto c(\theta)$ are continuous at θ_0 and positive, then [A1](i) and (ii) (for all $a > 0$) are satisfied.

Example 1: We now illustrate the conditions of Theorem 1 in a simple moving average model of order two:

$$y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2}, \tag{7}$$

where $\{e_t\}_{t=1}^T$ is a sequence of white noise random variables such that $E[e_t^{4+\delta}] < +\infty$ and some $\delta > 0$. Our prior for $\theta = (\theta_1, \theta_2)^\intercal$ is uniform over the following invertibility region,

$$-2 \leq \theta_1 \leq 2,\ \theta_1 + \theta_2 \geq -1, \theta_1 - \theta_2 \leq 1. \tag{8}$$

Following Marin et al. (2011), we choose as summary statistics for Algorithm 1 the sample autocovariances $\eta_j(y) = T^{-1}\sum_{t=1+j}^{T} y_t y_{t-j}$, for $j = 0, 1, 2$. For this choice the j-th component of $b(\theta)$ is $b_j(\theta) = E_\theta(z_t z_{t-j})$.

Now, take $d_2\{\eta(z), b(\theta)\} = \|\eta(z) - b(\theta)\|$. Under the moment condition for e_t given above, it can be shown that $V(\theta) = E[\{\eta(z) - b(\theta)\}\{\eta(z) - b(\theta)\}^\intercal]$ satisfies $\text{tr}\{V(\theta)\} < +\infty$ for all θ in (B2) and, by an application of Markov's inequality, we can conclude that

$$P_\theta\left[d_2\{\eta(z), b(\theta)\} > u\right] = P_\theta\left[d_2\{\eta(z), b(\theta)\}^2 > u^2\right] \leq \frac{\text{tr}\{V(\theta)\}}{u^2T} + o(1/T),$$

where the $o(1/T)$ term comes from the fact that there are finitely many non-zero covariance terms due to the m-dependent nature of the series, and with condition [A1] satisfied as a result. Given the structure of $b(\theta)$, the uniform prior $p(\theta)$ over (B2) automatically fulfills [A2] for θ_0 in this space. Lastly, we note that $\theta \mapsto b(\theta) = (1 + \theta_1^2 + \theta_2^2, (1 + \theta_2)\theta_1, \theta_2)^\intercal$ is an injective function and satisfies [A3].

Remark 2: The results of Theorem 1 can be visualised by fixing a particular value of θ, say $\tilde{\theta}$, and generating 'observed' data sets $\tilde{y}$ of increasing length, then running Algorithm 1 on these data sets. If the conditions of Theorem 1 are satisfied, the posterior density should concentrate on the value $\tilde{\theta}$ and become increasingly peaked as the sample size grows. This behavior is demonstrated in Section 2·1 of the Supplementary Material, through the moving average model of Example 1.

4. Shape of the Asymptotic Posterior Distribution

While posterior concentration states that $\Pi\left[d_1\{\theta, \theta_0\} > \delta | d_2\{\eta(y), \eta(z)\} \leq \varepsilon_T\right] = o_P(1)$ for an appropriate choice of ε_T, it does not indicate precisely how this mass accumulates, or the approximate amount of posterior mass within any neighbourhood of θ_0. This information is needed to obtain accurate expressions of uncertainty about point estimators of θ_0 and to ensure credible regions have proper frequentist coverage. To this end, we now analyse the limiting shape of the posterior measure. We consider the shape of $\Pi\left[\cdot|d_2\{\eta(y),\eta(z)\}\leq\varepsilon_T\right]$ 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 $\Pi\left[\cdot|d_2\{\eta(y),\eta(z)\}\leq\varepsilon_T\right]$ as $\Pi_\varepsilon(\cdot|\eta_0)$, where $\eta_0=\eta(y)$. Let $\|\cdot\|_*$ denote the spectral norm.

In addition to Assumption [A2] the following conditions are needed to establish the results of this section.

[A1′] Assumption [A1] holds and there exists a positive definite matrix $\Sigma_T(\theta_0)$, $c_0>0$, $\kappa>1$ and $\delta>0$ such that for all $\|\theta-\theta_0\|\leq\delta$, $P_\theta\left[\|\Sigma_T(\theta_0)\{\eta(z)-b(\theta)\}\|>u\right]\leq c_0u^{-\kappa}$ for all $0<u\leq\delta\|\Sigma_T(\theta_0)\|_*$.

[A3′] Assumption [A3] holds, the map b is continuously differentiable at θ_0 and the Jacobian $\nabla_\theta b(\theta_0)$ has full column rank k_θ.

[A4] For some $\delta>0$ and for all $\|\theta-\theta_0\|\leq\delta$, there exists a sequence of $(k_\eta\times k_\eta)$ positive definite matrices $\Sigma_T(\theta)$, with $k_\eta=\dim\{\eta(z)\}$, such that

$$\Sigma_T(\theta)\{\eta(z)-b(\theta)\}\Rightarrow\mathcal{N}(0,I_{k_\eta}),$$

where I_{k_η} is the $(k_\eta\times k_\eta)$ identity matrix.

[A5] There exists $v_T\to+\infty$ such that for all $\|\theta-\theta_0\|\leq\delta$, the sequence of functions $\theta\mapsto\Sigma_T(\theta)v_T^{-1}$ converges to some positive definite $A(\theta)$ and is equicontinuous at θ_0.

[A6] For some positive δ, all $\|\theta-\theta_0\|\leq\delta$, and for all ellipsoids $B_T=\left\{(t_1,\cdots,t_{k_\eta}):\sum_{j=1}^{k_\eta}t_j^2/h_T^2\leq 1\right\}$ and all $u\in\mathbb{R}^{k_\eta}$ fixed, for some $h_T\to 0$ as $T\to+\infty$,

$$\begin{aligned}\lim_T h_T^{-k_\eta}\,P_\theta\left[\Sigma_T(\theta)\{\eta(z)-b(\theta)\}-u\in B_T\right]&=\varphi_{k_\eta}(u),\\ h_T^{-k_\eta}\,P_\theta\left[\Sigma_T(\theta)\{\eta(z)-b(\theta)\}-u\in B_T\right]&\leq H(u),\qquad\int H(u)du<+\infty,\end{aligned}\tag{9}$$

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

Remark 3: [A1′] is a deviation control condition similar to [A1] but for $\Sigma_T(\theta_0)\{\eta(z)-b(\theta)\}$. Assumption [A4] is like a central limit theorem for $\{\eta(z)-b(\theta)\}$ and, as such, requires the existence of a positive-definite matrix $\Sigma_T(\theta)$. In simple cases, such as, for example, independent and identically distributed data with $\eta(z)=T^{-1}\sum_{i=1}^T g(z_i)$, $\Sigma_T(\theta)=v_TA_T(\theta)$ with $A_T(\theta)=A(\theta)+o_P(1)$ and $V(\theta)=E[\{g(Z)-b(\theta)\}\{g(Z)-b(\theta)\}^\intercal]=[A(\theta)^\intercal A(\theta)]^{-1}$. Assumptions [A3′] and [A6] are regularity conditions that ensure $\theta\mapsto b(\theta)$ and the variance-covariance matrix of $\{\eta(z)-b(\theta)\}$ are well-behaved, which allows the posterior behavior of a normalized version of $(\theta-\theta_0)$ to be governed by the posterior behavior of $\Sigma_T(\theta_0)\{b(\theta)-b(\theta_0)\}$. [A6] governs the pointwise convergence of a normalized version of the measure P_θ, therein dominated by $H(u)$. [A6] allows for the application of the dominated convergence theorem in Case (iii) of the following result, where $d_2\{\cdot,\cdot\}$ corresponds to the Euclidean distance:

THEOREM 2. *Under Assumptions [A1′], [A2] and [A3′]-[A5], with $\kappa>k_\theta$, the following hold:*
(i) If $\lim_T v_T\varepsilon_T=+\infty$, with probability approaching 1, the posterior distribution of $\varepsilon_T^{-1}(\theta-\theta_0)$ converges to the uniform distribution over the ellipsoid $\{x^\intercal B_0x\leq 1\}$ with $B_0=\nabla_\theta b(\theta_0)^\intercal\nabla_\theta b(\theta_0)$, meaning that for f continuous and bounded, with probability approaching 1,

$$\lim_{T\to+\infty}\int f\{\varepsilon_T^{-1}(\theta-\theta_0)\}d\Pi_\varepsilon(\theta|\eta_0)=\int_{u^\intercal B_0u\leq 1}f(u)du\Big/\int_{u^\intercal B_0u\leq 1}du.\tag{10}$$

(ii) If $\lim_T v_T\varepsilon_T=c>0$, there exists a non-Gaussian distribution on $\mathbb{R}^{k_\eta}$, Q_c, such that

$$\Pi_\varepsilon\left[\Sigma_T(\theta_0)\{b(\theta)-b(\theta_0)\}-Z_T^0\in B|\eta_0\right]\to Q_c(B),\tag{11}$$

where $Z_T^0 = \Sigma_T(\theta_0)\{\eta(y) - b(\theta_0)\}$, and for $\Phi_{k_\eta}(\cdot)$ the CDF of a k_η-dimensional standard normal random variable,

$$Q_c(B) \propto \int_B \Phi_{k_\eta}\left\{(Z-x)^\intercal A(\theta_0)^\intercal A(\theta_0)(Z-x) \le c^2\right\} dx.$$

(iii) If $\lim_T v_T\varepsilon_T = 0$ and Assumption [A6] holds then

$$\lim_{T\to+\infty} \Pi_\varepsilon\left[\Sigma_T(\theta_0)\{b(\theta)-b(\theta_0)\} - Z_T^0 \in B|\eta_0\right] = \Phi_{k_\eta}(B). \tag{12}$$

Remark 4: For a sufficiently regular η, Theorem 5 asserts that the crucial feature in determining the limiting shape of the posterior is the behaviour of $v_T\varepsilon_T$. If $\lim_T v_T\varepsilon_T > 0$, $\Pi_\varepsilon(\cdot|\eta_0)$ is not approximately Gaussian. In Case (i), which corresponds to a large tolerance ε_T, the posterior has nonstandard asymptotic behaviour. A heuristic argument is as follows. Under the assumptions of Theorem 5, if $v_T\varepsilon_T \gg 1$, $\|\eta(z)-\eta(y)\| \le \varepsilon_T$ is equivalent to the constraint $\|b(\theta)-b(\theta_0)\| \le \varepsilon_T\{1+o(1)\}$. Therefore, the probability $P_\theta\{\|\eta(z)-\eta(y)\| \le \varepsilon_T\}$ is itself equivalent to the indicator function on this constraint, which a Taylor series argument shows is also equivalent to $\|\nabla_\theta b(\theta_0)(\theta-\theta_0)\| \le \varepsilon_T\{1+o_P(1)\}$ by the regularity condition on b. Hence, asymptotically the posterior behaves like the prior distribution truncated over the above ellipsoid and by prior continuity this is equivalent to the uniform distribution over this ellipsoid. It is only when, in Case (iii), $\lim_T v_T\varepsilon_T = 0$ that a Bernstein-von Mises result is available. This behaviour has been also demonstrated, since the initial submission of our article, in a preprint by Li and Fearnhead (arXiv:1609.07135) where they observe that asymptotically the posterior distribution behaves like a convolution of a Gaussian distribution with variance of order $1/v_T^2$ and a uniform distribution over a ball of order ε_T, and depending on the order $v_T\varepsilon_T$ one of the two distributions dominates.

Remark 5: An immediate, and critical, consequence of Theorem 5 is that credible regions calculated from the posterior will coincide with frequentist confidence regions only if $\varepsilon_T = o(1/v_T)$. When $\varepsilon_T = O(1/v_T)$ credible regions have radius with the correct order of magnitude but miss the correct asymptotic coverage. Lastly, if $\varepsilon_T \gg v_T^{-1}$ the credible regions have frequentist coverage going to 1.

Remark 6: As with Theorem 1, the behavior of $\Pi(\cdot|\eta_0)$ described by Theorem 5 can be visualised. This is demonstrated in Section 2·2 of the Supplementary Material, through the model of Example 1. Formal verification of the conditions underpinning Theorem 2 is quite challenging, even in this case. Numerical results nevertheless highlight that for this particular choice of model and summaries a Bernstein-von Mises result holds, conditional on $\varepsilon_T = o(1/v_T)$, with $v_T = T^{1/2}$.

Remark 7: Assumption [A6], which is only required when $v_T\varepsilon_T = o(1)$, only applies to random variables $\eta(z)$ that are absolutely continuous with respect to the Lebesgue measure (or in the case of sums of random variables, to sums that are non-lattice; see Bhattacharya and Rao, 1986). For discrete $\eta(z)$, Assumption [A6] must be adapted for Theorem 5 to be satisfied. One such adaptation is given as [A6*]:

[A6*] There exist $\delta > 0$ and a countable set E_T such that for all $\|\theta-\theta_0\| < \delta$,

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

and

$$\sup_{\|\theta-\theta_0\|\le\delta} \sum_{x\in E_T} \left| p[\Sigma_T(\theta)\{x-b(\theta)\}|\theta] - v_T^{-k_\eta}|A(\theta_0)|^{-1/2}\varphi_{k_\eta}[\Sigma_T(\theta)\{x-b(\theta)\}] \right| = o(1).$$

Assumption [A6*] is satisfied, for instance, in the case when $\eta(z)$ is a sum of independent lattice random variables, as in the population genetics experiment detailed in Section 3.3 of Marin et al. (2014), which compares evolution scenarios of separated populations from a most recent common ancestor. Furthermore, this example is such that Assumptions [A1′], [A2] and [A3′]-[A5] also hold, which means that the conclusions of both Theorems 1 and 5 apply to this model with a large number of discrete summary statistics.

Remark 8: Theorem 5 generalises to the case where the components of $\eta(z)$ have different rates of convergence. The statement and proof of this more general result are deferred to Section 1 of the Supplementary Material.

5. ASYMPTOTIC DISTRIBUTION OF THE POSTERIOR MEAN

5·1. *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 5 is not a weaker, or stronger, result than the 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 this is not a necessary condition.

To present the ideas as clearly as possible, we focus on the case of a scalar parameter θ and scalar summary $\eta(y)$; i.e., in this section we take $k_\theta = k_\eta = 1$. An extension to the multivariate case is then presented in Section 5·2.

In addition to Assumptions [A1′], [A2] and [A3′]-[A6], we require a further assumption on the prior.

[A7] The prior density p is such that (i) For $\theta_0 \in \text{Int}(\Theta)$, $p(\theta_0) > 0$; (ii) The density function p is β-Hölder in a neighbourhood of θ_0: there exist $\delta, L > 0$ such that for all $|\theta - \theta_0| \leq \delta$, and $\nabla_\theta^{(j)} p(\theta_0)$ the (j)-th derivative of $p(\theta_0)$,

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

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

THEOREM 3. *Assumptions [A1′], [A2], [A3′]-[A5], with $\kappa > \beta + 1$ and [A7] are satisfied. Let $b = b(\theta)$. Assume that b is β-Hölder in a neighbourhood of θ_0 and that $\nabla_\theta b(\theta_0) \neq 0$. Denoting $E_{\Pi_\varepsilon}(\theta)$ and $E_{\Pi_\varepsilon}(b)$ as the posterior mean of θ and b, respectively, we then have the following characterisation:*
(i) If $\lim_T v_T \varepsilon_T = +\infty$ then, for $b_0 = b(\theta_0)$,

$$E_{\Pi_\varepsilon}(b - b_0) = \{\eta(y) - b(\theta_0)\} + \sum_{j=1}^{k} \frac{\nabla_\theta^{(2j-1)} p(b_0)}{(2j-1)!(2j+1)} \varepsilon_T^{2j} + O(\varepsilon_T^{1+\beta}) + o_P(1/v_T),$$

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

$$E_{\Pi_\varepsilon}(\theta - \theta_0) = \frac{\{\eta(y) - b(\theta_0)\}}{\nabla_\theta b(\theta_0)} + \sum_{j=1}^{\lfloor \beta \rfloor} \frac{\nabla_b^{(j)} b^{-1}(b_0)}{j!} \sum_{l=\lceil j/2 \rceil}^{\lfloor (j+k)/2 \rfloor} \frac{\varepsilon_T^{2l} \nabla_b^{(2l-j)} p(b_0)}{p(b_0)(2l-j)!} + O(\varepsilon_T^{1+\beta}) + o_P(1/v_T). \tag{13}$$

Hence, if $\varepsilon_T^{2\wedge(1+\beta)} = o(1/v_T)$,

$$E_{\Pi_\varepsilon}(\theta - \theta_0) = \frac{\{\eta(y) - b(\theta_0)\}}{\nabla_\theta b(\theta_0)} + o_P(1/v_T), \; E_{\Pi_\varepsilon}\{v_T(\theta - \theta_0)\} \Rightarrow \mathcal{N}(0, V(\theta_0)/\{\nabla_\theta b(\theta_0)\}^2), \tag{14}$$

where $V(\theta_0) = \lim_T var[v_T\{\eta(y) - b(\theta_0)\}]$.
(ii) If $\lim_T v_T \varepsilon_T = c \geq 0$, and when $c = 0$ Assumption [A6] holds, then (14) *also holds.*

Remark 9: Equation (13) highlights a potential deviation from the expected asymptotic behaviour of the posterior mean $E_{\Pi_\varepsilon}(\theta)$, i.e., the behaviour corresponding to $T \to +\infty$ and $\varepsilon_T \to 0$. Indeed, the posterior mean is asymptotically normal for all values of $\varepsilon_T = o(1)$, but is asymptotically unbiased only if the leading term in equation (13) is $[\nabla_\theta b(\theta_0)]^{-1}\{\eta(y) - b(\theta_0)\}$, which is satisfied under Case (ii) and in Case (i) if $v_T \varepsilon_T^2 = o(1)$ (given $\beta \geq 1$). However, in Case (i), if $\liminf_T v_T \varepsilon_T^2 > 0$, when $\beta \geq 3$, the posterior mean has a bias

$$\varepsilon_T^2 \left[\frac{\nabla_b p(b_0)}{3p(b_0)\nabla_\theta b(\theta_0)} - \frac{\nabla_\theta^{(2)} b(\theta_0)}{2\{\nabla_\theta b(\theta_0)\}^2}\right] + O(\varepsilon_T^4) + o_P(1/v_T).$$

Remark 10: 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 achieved if $\varepsilon_T = o(1/v_T)$. Therefore, an immediate consequence of Theorems 5 and 3 is that the posterior mean of $\Pi_\varepsilon(\cdot|\eta_0)$ is asymptotically Gaussian with zero bias provided $v_T\varepsilon_T^2 = o(1)$ but to obtain both an asymptotically Gaussian posterior mean and credible regions with asymptotically proper coverage we require $\varepsilon_T = o(1/v_T)$. If $v_T \asymp \varepsilon_T$ or $\varepsilon_T \gg v_T^{-1}$, then the frequentist coverage of credible balls centered at $E_{\Pi_\varepsilon}(\theta)$ is not equal to the nominal level. It goes to 1 in the latter case and to some constant greater than the nominal value in the former.

Remark 11: One consequence of Theorem 3 is that the 'two-stage' procedure advocated by Fearnhead and Prangle (2012) does not yield a reduction in asymptotic variance over a point estimate produced via Algorithm 1 using the 'first-stage' set of summaries. Consider a first-stage summary statistic $\eta(y)$ with dimension $k_\eta = k_\theta = 1$. When either $v_T\varepsilon_T \to +\infty$ or $\lim_T v_T\varepsilon_T = c \geq 0$, and $v_T\varepsilon_T^2 = o(1)$, such that $v_T\{\eta(y) - b(\theta)\} \Rightarrow N\{0, V(\theta_0)\}$, then $v_T E_{\Pi_\varepsilon}\{\theta - \theta_0\} \Rightarrow N(0, V(\theta_0)/\{\nabla_\theta b(\theta_0)\}^2)$. Then, using a regression-adjusted version of $\eta(y)$ as a summary statistic at the second stage, and applying Algorithm 1 with the same tolerance ε_T, leads to an equivalent asymptotic distribution for the ABC posterior mean. This comment remains valid when $k_\eta > k_\theta \geq 1$.

Remark 12: As in the earlier cases, the implications of Theorem 3 can be visualised. In particular, using the moving average example for illustration, the fact that asymptotic normality of the posterior mean does not require $= o(1/v_T)$ is highlighted in Section 2·3 of the Supplementary Material.

5·2. *Comparison with Existing Results*

Li and Fearnhead (2016) have, in parallel, analysed the asymptotic properties of the posterior mean or some function thereof. (See also Creel et al., 2015.) Under the assumption of a central limit theorem for the summary statistic 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, with no bias, of the posterior mean if $\varepsilon_T = o(v_T^{-3/5})$. Heuristically, the authors derive this result using an approximation of the posterior density $p_\varepsilon\{\theta|\eta(y)\}$, based on the Gaussian approximation of the density of $\eta(z)$ given θ and using properties of the maximum likelihood estimator conditional on $\eta(y)$. In contrast to our analysis, these authors allow the acceptance probability defining the algorithm to be an arbitrary density kernel in $\|\eta(y) - \eta(z)\|$. Consequently, their approach is more general than the accept/reject version considered in Theorem 3.

However, the conditions Li and Fearnhead (2016) require of $\eta(y)$ are stronger than ours. In particular, our results on asymptotic normality for the posterior mean only require weak convergence of $v_T\{\eta(z) - b(\theta)\}$ under P_θ, with polynomial deviations that need not be uniform in θ. These assumptions allow for the explicit treatment of models where the parameter space Θ is not compact. In addition, asymptotic normality of the posterior mean requires Assumption [A6] only if $\varepsilon_T = o(1/v_T)$. Hence if $\varepsilon_T \gg v_T^{-1}$, then only deviation bounds and weak convergence are required, which are much weaker than convergence of the densities. When $\varepsilon_T = o(1/v_T)$ then Assumption [A6] essentially implies local (in θ) convergence of the density of $v_T\{\eta(z) - b(\theta)\}$, but with no requirement on the rate of this convergence. This assumption is weaker than the uniform convergence required in Li and Fearnhead (2016). Our results also allow for an explicit representation of the bias that obtains for the posterior mean when $\liminf_T v_T\varepsilon_T^2 > 0$.

Li and Fearnhead (2016) also provide the interesting result that when $k_\eta > k_\theta \geq 1$ and if $\varepsilon_T = o(v_T^{-3/5})$, the posterior mean is asymptotically normal, and unbiased, but is not asymptotically efficient. To help shed light on this phenomenon, the following result gives an alternative to Theorem 3.1 of these authors and contains an explicit representation of the asymptotic expansion for the posterior mean when $k_\eta > k_\theta \geq 1$.

THEOREM 4. *Assumptions [A1′], [A2], [A3′]-[A5], with $\kappa > 2$, and [A7] are satisfied. Assume that $v_T\varepsilon_T \to +\infty$ and $v_T\varepsilon_T^2 = o(1)$. Assume also that $b(.)$ and $p(.)$ are Lipschitz in a neighbourhood of θ_0.*

Then

$$E_{\Pi_\varepsilon}\{v_T(\theta-\theta_0)\} = \{\nabla_\theta b(\theta_0)^\intercal \nabla_\theta b(\theta_0)\}^{-1}\nabla_\theta b(\theta_0)^\intercal v_T\{\eta(y)-b(\theta_0)\} + o_p(1).$$

In addition, if $\{\nabla_\theta b(\theta_0)^\intercal \nabla_\theta b(\theta_0)\}^{-1}\nabla_\theta b(\theta_0)^\intercal \neq \nabla_\theta b(\theta_0)^\intercal$*, the matrix*

$$var\left[\{\nabla_\theta b(\theta_0)^\intercal \nabla_\theta b(\theta_0)\}^{-1}\nabla_\theta b(\theta_0)^\intercal v_T\{\eta(y)-b(\theta_0)\}\right] - \{\nabla_\theta b(\theta_0)^\intercal V^{-1}(\theta_0)\nabla_\theta b(\theta_0)\}^{-1},$$

is positive semi-definite, where $\{\nabla_\theta b(\theta_0)^\intercal V^{-1}(\theta_0)\nabla_\theta b(\theta_0)\}^{-1}$ *is the optimal asymptotic variance achievable given* $\eta(y)$*.*

Finally, and in contrast to Li and Fearnhead (2016), our results in Theorem 5 completely characterize the asymptotic distribution of the posterior for all $\varepsilon_T = o(1)$ that admit posterior concentration. This general characterization allows us to demonstrate, via Theorem 3 part (i), that asymptotic normality and unbiasedness of the posterior mean remain achievable even if $\lim_T v_T\varepsilon_T = +\infty$, provided the tolerance satisfies $\varepsilon_T = o(v_T^{-1/2})$.

6. PRACTICAL IMPLICATIONS OF THE RESULTS

6·1. *General Implications*

Algorithm 1 does not reflect the way in which approximate Bayesian computation is typically applied in practice. The most common version replaces the acceptance step by a nearest-neighbour selection step, whereby draws of θ are retained only if they yield distances in the left tail of $\{\theta \in \Theta : d_2\{\eta(z), \eta(y)\}\}$; see, for example, Biau et al. (2015). More precisely, Step (3) in Algorithm 1 is substituted with

(3′) Select all θ^i associated with the $\alpha = \delta/N$ smallest distances $d_2\{\eta(z^i), \eta(y)\}$ for some δ.

This nearest-neighbour version corresponds to accepting draws of θ associated with an empirical quantile over the simulated distances $d_2\{\eta(z), \eta(y)\}$, which defines the acceptance probability for Algorithm 1. A key practical insight of this section is that, on the one hand, the order of magnitude of the acceptance probability, $\alpha_T = \text{pr}(\|\eta(z)-\eta(y)\| \leq \varepsilon_T)$, is only affected by the dimension of θ, as formalised in Corollary 1. On the other hand, if ε_T becomes much smaller than $1/v_T$, the dimension of $\eta(\cdot)$ impacts on our ability to consistently estimate this acceptance probability, as formalized in Corollary 2.

COROLLARY 1. *Under the conditions in Theorem* 5*:*
(i) If $\varepsilon_T \asymp v_T^{-1}$ *or* $\varepsilon_T = o(v_T^{-1})$*, then the acceptance rate associated with the threshold* ε_T *is*

$$\alpha_T = pr\left(\|\eta(z)-\eta(y)\| \leq \varepsilon_T\right) \asymp (v_T\varepsilon_T)^{k_\eta} \times v_T^{-k_\theta} \lesssim v_T^{-k_\theta}.$$

(ii) If $\varepsilon_T \gg v_T^{-1}$ *, then*

$$\alpha_T = pr\left(\|\eta(z)-\eta(y)\| \leq \varepsilon_T\right) \asymp \varepsilon_T^{k_\theta} \gg v_T^{-k_\theta}.$$

This shows that choosing a tolerance $\varepsilon_T = o(1)$ is equivalent to choosing an $\alpha_T = o(1)$ quantile of $\|\eta(z)-\eta(y)\|$ and, hence, theoretically rationalizes the nearest-neighbour version of Algorithm 1. It also demonstrates the role played by the dimension of θ on the rate at which α_T declines to zero. In Case (i), if $\varepsilon_T \asymp v_T^{-1}$, then $\alpha_T \asymp v_T^{-k_\theta}$. On the other hand, if $\varepsilon_T = o(v_T^{-1})$, as required for the Bernstein-von Mises result in Theorem 5, the associated acceptance probability goes to zero at the faster rate, $\alpha_T = o(v_T^{-k_\theta})$. In Case (ii), where $\varepsilon_T \gg v_T^{-1}$, it follows that $\alpha_T \gg v_T^{-k_\theta}$.

Linking ε_T and α_T as shown gives 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 occurs for large T: choosing $\alpha_T \gtrsim v_T^{-k_\theta}$ gives a posterior that concentrates; under the more stringent condition $\alpha_T = o(v_T^{-k_\theta})$ the posterior both concentrates *and* is approximately Gaussian in large samples. Such results are of critical importance as they give practitioners an understanding of what to expect from the procedure, and a means of detecting potential issues if this expected posterior behaviour is not produced when choosing a certain α_T quantile. Moreover, given that there is no direct link between the posterior $\Pi_\varepsilon(\cdot|\eta_0)$ and the exact posterior based on the full likelihood, this result at least gives researchers some understanding of the statistical properties that $\Pi_\varepsilon(\cdot|\eta_0)$ should display, for large T, when it is obtained from the popular nearest-neighbour version of the algorithm.

Furthermore, Corollary 1 demonstrates that to obtain reasonable statistical behavior, the rate at which α_T declines to zero must be faster the larger the dimension of θ, with the order of α_T unaffected by the dimension of η. This result thus provides theoretical evidence of a curse-of-dimensionality encountered in these algorithms as the dimension of the parameter of interest increases, with this being the first piece of work, to our knowledge, to link the dimension of θ to the acquisition of certain asymptotic properties for the posterior $\Pi_\varepsilon(\cdot|\eta_0)$. This finding also provides theoretical justification for dimension reduction methods that process parameter dimensions individually and independent of the other remaining dimensions; see, for example, the regression adjustment approaches of Beaumont et al. (2002), Blum (2010) and Fearnhead and Prangle (2012), 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.

The striking result of Corollary 1 is that, under the nearest-neighbour interpretation of Algorithm 1 and for an acceptance probability $\alpha_T = o(1)$, our asymptotic results are achievable if α_T decays to zero as some power of $v_T^{-k_\theta}$; i.e., the order of α_T is unaffected by the dimension of the summaries, k_η. However, α_T can not be accessed in practice and so the nearest-neighbour version of Algorithm 1 is implemented using a Monte Carlo approximation to α_T, which is based on the accepted draws of θ. It then makes sense to tie our results to this Monte Carlo approximation of α_T. To do so, we link the tolerance ε_T to the Monte Carlo error associated with estimating α_T, which depends on the number of draws in Algorithm 1, which we index explicitly by T and refer to as N_T hereafter. Specifically, for $\hat{\alpha}_T = \sum_{i=1}^{N_T} \mathbb{1}\left[d_2\{\eta(y), \eta(z)\} \le \varepsilon_T\right]/N_T$, Corollary 2 establishes conditions on N_T and ε_T under which $\hat{\alpha}_T = \alpha_T\{1 + o_p(1)\}$.

COROLLARY 2. *In the setting of Corollary* 1*, let N_T be the number of Monte Carlo draws in Algorithm* 1*. If either of the following is satisfied*
(i) $\varepsilon_T = o(v_T^{-1})$ and $(v_T\varepsilon_T)^{-k_\eta}\varepsilon_T^{-k_\theta} M \le N_T$ for M large enough;
(ii) $\varepsilon_T \gtrsim v_T^{-1}$ and $\varepsilon_T^{-k_\theta} M \le N_T$ for M large enough;
then the proportion of accepted draws, $\hat{\alpha}_T$, satisfies, with probability going to 1,

$$\hat{\alpha}_T = \alpha_T\{1 + o_p(1)\}.$$

Corollary 2 demonstrates that the large sample properties of approximate Bayesian posteriors discussed herein are achievable using the nearest-neighbour interpretation of Algorithm 1 and the estimated acceptance probability $\hat{\alpha}_T$. Furthermore, it is also critical to note that, depending on the choice of ε_T, consistency of $\hat{\alpha}_T$, for α_T, may require a larger number of Monte Carlo draws, N_T, the larger is k_η, the dimension of η. In particular, k_η impacts on N_T only if ε_T converges to zero fast enough to obtain an asymptotically Gaussian posterior. Hence, if one is not concerned with obtaining asymptotic Gaussianity and, thus, accurate frequentist coverage, it is best to choose $\varepsilon_T = \delta v_T^{-1/2}$ for some small constant $\delta > 0$. However, if a Bernstein-von Mises result is a goal, then the dimension k_η has a negative impact. Interpreted this way, Corollary 2 elucidates the tension between posterior accuracy and the computational cost required to obtain that accuracy. In addition, taken together, Corollaries 1 and 2 provide a theoretically sound rule for choosing ε_T (or α_T) and the number of Monte Carlo draws N_T that can easily be implemented in practice. Contrary to Li and Fearnhead (2016), we do not explore importance sampling approaches that would possibly allow for a non-vanishing acceptance rate, but study the basic Monte Carlo approximation. In particular, for the Monte Carlo error of the estimator of $E_{\Pi_\varepsilon}(\theta)$ to be of smaller order than the estimation error, which is the larger of ε_T and v_T^{-1}, one would need the number of simulations to satisfy $N_T \gg \alpha_T^{-1}$, which is satisfied under the conditions of Corollary 2.

Lastly, the results in this section suggest that the persistent opinion in the literature that ε_T in Algorithm 1 should always be taken "as small as the computing budget allows" is questionable. Once ε_T is chosen small enough to satisfy Case (iii) of Theorem 5, which leads to the most stringent requirement on the tolerance, $v_T\varepsilon_T = o(1)$, there may well be no gain in pushing ε_T (or, equivalently, α_T) closer to zero (as shown in Corollary 1) and thereby incurring an even larger computational cost (as shown in Corollary 2). We explore these ideas in the following section via a simple numerical illustration. The example adopted is sufficiently regular to ensure that the central limit theorem behind the Bernstein-von Mises result holds for a reasonably small sample size, and that the associated order condition on ε_T has practical content. This illustration highlights the fact that, despite the asymptotic foundations of the conclusions drawn above regarding the optimal choice of ε_T, those conclusions can be relevant even for sample sizes encountered in practice.

6·2. *Numerical Illustration of Quantile Choice*

Consider the simple example where we observe a sample $\{y_t\}_{t=1}^T$ from $y_t \sim \mathcal{N}(\mu, \sigma)$ with $T = 100$, and our goal is posterior inference on $\theta = (\mu, \sigma)^\intercal$. We use as summaries the sample mean and variance, $\bar{x}$ and s_T^2, which satisfy a central limit theorem at rate $T^{1/2}$. In order to guarantee approximate posterior normality, we choose an α_T quantile of the simulated distances according to $\alpha_T = o(1/T)$, because of the joint inference on μ and σ. For the purpose of this illustration, we will compare inference based on the nearest-neighbour version of 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 N of draws 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 $\widehat{p}_{\alpha_j}\{\mu|\eta(y)\}$, and the exact marginal posterior $p(\mu|y)$ as

$$RMSE_\mu(\alpha_j) = \left(\frac{1}{G} \sum_{g=1}^{G} \left[\widehat{p}^g_{\alpha_j}\{\mu|\eta(y)\} - p^g(\mu|y) \right]^2 \right)^{1/2}, \tag{15}$$

where $\widehat{p}^g$ is the ordinate of the density estimate from the nearest-neighbour version of 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 $\mu_0 = 1,\ \sigma_0 = 1$.

Before presenting the replication results, it is instructive to consider the graphical results of one particular run of the algorithm (for each of the α_j values). Figure 2 plots the resulting marginal posterior estimates and compares these with the exact (finite sample) marginal posteriors of μ and σ (respectively), the implication of the argument presented at the end of the previous section being that for large enough T, once ε_T reaches a certain threshold, decreasing the tolerance further will not necessarily result in more accurate estimates of these exact posteriors. This implication is in evidence in Figure 2: in the case of μ, there is a clear visual decline in the accuracy with which ABC estimates the exact marginal posterior when choosing quantiles smaller than α_2; whilst in the case of σ, the worst performing estimate is the one associated with the smallest value of α_j.

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 1 thus indicate that the larger and less computationally burdensome value of α_j yields on average a more accurate posterior estimate than that yielded by α_4. In brief, Table 1 paints a similar picture to that of Figure 2: 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 but a slight decline in accuracy beyond $\alpha_1 = 1/T^{1.1}$; for μ, estimates based on α_2 and α_3 are both more accurate that those based on α_4 and there is minimal gain in pushing the quantile below α_1.

These numerical results have important computational implications. To wit, and as we have 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 α_1, $N = 1.4e06$ for α_2, $N = 13.5e06$ for α_3, and $N = 41.0e06$ for

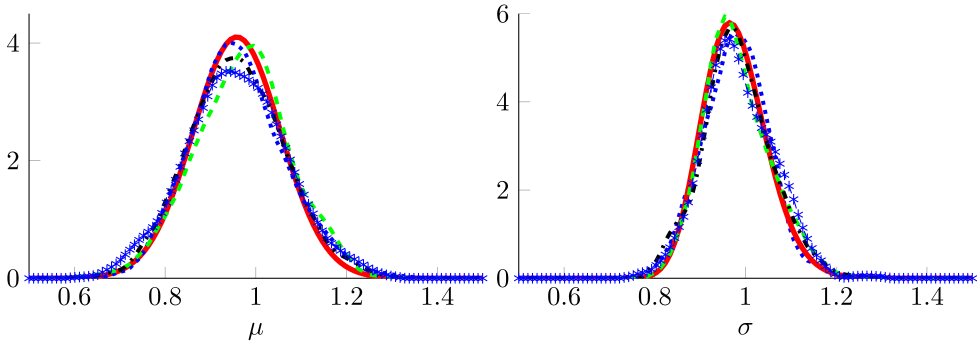


Figure 1. Exact marginal posteriors (—). Approximate Bayesian computation posteriors based on α_j: $\alpha_1 = 1/T^{1.1.}$ ($\cdots$); $\alpha_2 = 1/T^{3/2}$ (- - -); $\alpha_3 = 1/T^2$ (— · —); $\alpha_4 = 1/T^{5/2}$ —∗—

α_4. 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 5 has practical content, the same sort of qualitative results will continue to hold.

Table 1. *Ratio of the root mean square error for marginal posterior estimates based on the smallest quantile,* $\alpha_4 = 1/T^{5/2}$

	$\alpha_1 = 1/T^{1.1}$	$\alpha_2 = 1/T^{1.5}$	$\alpha_3 = 1/T^2$
$RMSE_\mu(\alpha_j)$	1.17	0.99	0.98
$RMSE_\sigma(\alpha_j)$	0.86	0.87	0.91

Acknowledgment

This research has been supported by Australian Research Council Discovery and l'Institut Universitaire de France grants. The third author is further affiliated with the University of Warwick. 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. We are also very grateful for the insightful and detailed comments of two anonymous referees and an associate editor on the initial draft of the paper.

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., 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.

[7] 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.

[8] 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.

[9] DROVANDI, C. C., PETTITT, A. N. and LEE, A. (2015). Bayesian Indirect Inference using a Parametric Auxiliary Model, *Statistical Science*, **30**, 72–95.

[10] 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.

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

[12] GOURIÉROUX, C., MONFORT, A. and RENAULT, E. (1993). Indirect Inference, *J. Applied Econom.*, **85**, S85–S118.

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

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

[15] MARIN, J-M., PUDLO, P., ROBERT, C.P. and RYDER, R. (2011). Approximate Bayesian Computation Methods, *Statistics and Computing*, **21**, 289–291.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] ROBERT, C.P. (2015). Approximate Bayesian Computation: a Survey on Recent Results, *Proceedings, MCQMC2014.*

[21] 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.

[22] TAVARÉ, S., BALDING, D.J., GRIFFITHS, R.C. and DONNELLY, P. (1997). Inferring Coalescence Times from DNA Sequence Data, *Genetics*, **145**, 505–518.

SUPPLEMENTARY MATERIALS

The following appendices contain proofs of all theoretical results, as well as numerical examples that illustrate the theoretical results.

A. PROOFS

A·1. *Proof of Theorem 1*

Let $\varepsilon_T > 0$ and assume that $y \in \Omega_\varepsilon = \{y : d_2\{\eta(y), b(\theta_0)\} \leq \varepsilon_T/3\}$. From Assumption [A1] and $\rho_T(\varepsilon_T/3) = o(1)$, $P_0\left(\Omega_\varepsilon\right) = 1 + o(1)$. Consider the joint event $A_\varepsilon(\delta') = \{(z,\theta) : d_2\{\eta(z), \eta(y)\} \leq \varepsilon_T\} \cap d_2\{b(\theta), b(\theta_0)\} > \delta'\}$. For all $(z,\theta) \in A_\varepsilon(\delta')$

$$\begin{aligned} d_2\{b(\theta), b(\theta_0)\} &\leq d_2\{\eta(z), \eta(y)\} + d_2\{b(\theta), \eta(z)\} + d_2\{b(\theta_0), \eta(y)\} \\ &\leq 4\varepsilon_T/3 + d_2\{b(\theta), \eta(z)\} \end{aligned}$$

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

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

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

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

and

$$\Pi\left(d_2\{b(\theta), b(\theta_0)\} > 4\varepsilon_T/3 + t_\varepsilon | d_2\{\eta(y), \eta(z)\} \leq \varepsilon_T\right) = \Pi_\varepsilon\left(d_2\{b(\theta), b(\theta_0)\} > 4\varepsilon_T/3 + t_\varepsilon | \eta_0\right)$$
$$\leq \frac{\int_\Theta P_\theta\left(d_2\{b(\theta), \eta(z)\} > t_\varepsilon\right) d\Pi(\theta)}{\int_\Theta P_\theta\left(d_2\{\eta(z), \eta(y)\} \leq \varepsilon_T\right) d\Pi(\theta)}. \quad \text{(A1)}$$

Moreover, since

$$d_2\{\eta(z), \eta(y)\} \leq d_2\{b(\theta), \eta(z)\} + d_2\{b(\theta_0), \eta(y)\} + d_2\{b(\theta), b(\theta_0)\} \leq \varepsilon_T/3 + \varepsilon_T/3 + d_2\{b(\theta), b(\theta_0)\}$$

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

$$\int_\Theta P_\theta\left(d_2\{\eta(z), \eta(y)\} \leq \varepsilon_T\right) d\Pi(\theta) \geq \int_{d_2\{b(\theta), b(\theta_0)\} \leq \varepsilon_T/3} P_\theta\left(d_2\{\eta(z), \eta(y)\} \leq \varepsilon_T/3\right) d\Pi(\theta)$$
$$\geq \Pi\left(d_2\{b(\theta), b(\theta_0)\} \leq \varepsilon_T/3\right) - \rho(\varepsilon_T/3) \int_{d_2\{b(\theta), b(\theta_0)\} \leq \varepsilon_T/3} c(\theta) d\Pi(\theta).$$

If part (i) of Assumption [A1] holds,

$$\int_{d_2\{b(\theta), b(\theta_0)\} \leq \varepsilon_T/3} c(\theta) d\Pi(\theta) \leq c_0 \Pi\left(d_2\{b(\theta), b(\theta_0)\} \leq \varepsilon_T/3\right)$$

and for ε_T small enough,

$$\int_\Theta P_\theta\left(d_2\{\eta(z), \eta(y)\} \leq \varepsilon_T\right) d\Pi(\theta) \geq \frac{\Pi\left(d_2\{b(\theta), b(\theta_0)\} \leq \varepsilon_T/3\right)}{2}, \quad \text{(A2)}$$

which, combined with (A1) and Assumption [A2], leads to

$$\Pi_\varepsilon\left(d_2\{b(\theta), b(\theta_0)\} > 4\varepsilon_T/3 + t_\varepsilon | \eta_0\right) \lesssim \rho_T(t_\varepsilon)\varepsilon_T^{-D} \lesssim \frac{1}{M} \quad \text{(A3)}$$

by choosing $t_\varepsilon = \rho_T^{-1}(\varepsilon_T^D/M)$ with M large enough. If part (ii) of Assumption [A1] holds, a Hölder inequality implies that

$$\int_{d_2\{b(\theta), b(\theta_0)\} \leq \varepsilon_T/3} c(\theta) d\Pi(\theta) \lesssim \Pi\left(d_2\{b(\theta), b(\theta_0)\} \leq \varepsilon_T/3\right)^{a/(1+a)}$$

and if ε_T satisfies

$$\rho_T(\varepsilon_T) = o\left(\varepsilon_T^{D/(1+a)}\right) = O\left(\Pi\left(d_2\{b(\theta), b(\theta_0)\} \leq \varepsilon_T/3\right)^{1/(1+a)}\right)$$

then (A3) remains valid.

A·2. *Proof of Theorem 2*

Herein we state and prove a generalization of Theorem 2 that allows for differing rates of convergence for $\eta(y)$. The result for Theorem 2 in the text is a direct consequence of this generalization.

In particular, in this section we assume that there exists a sequence of $(k_\eta \times k_\eta)$ positive definite matrices $\Sigma_T(\theta)$ such that for all θ in a neighbourhood of $\theta_0 \in \text{Int}(\Theta)$,

$$c_1 D_T \leq \Sigma_T(\theta) \leq c_2 D_T, \quad D_T = \text{diag}(v_T(1), \cdots, v_T(k)), \quad \text{(A4)}$$

with $0 < c_1, c_2 < +\infty$, $v_T(j) \to +\infty$ for all j and the $v_T(j)$ possibly all distinct and $A \leq B$ meaning that the matrix $B - A$ is semi-definite positive. Thus, in the presentation and proof of this generalization to Theorem 2 we do not restrict ourselves to identical convergence rates for the components of the statistic $\eta(z)$. For simplicity's sake we order the components so that

$$v_T(1) \leq \cdots \leq v_T(k_\eta). \quad \text{(A5)}$$

For all j, we assume $\lim_T v_T(j)\varepsilon_T = \limsup_T v_T(j)\varepsilon_T$. For any square matrix A of dimension k_η, if $q \leq k_\eta$, $A_{[q]}$ denotes the $(q \times q)$ square upper sub-matrix of A. Also, let $j_{\max} = \max\{j : \lim_T v_T(j)\varepsilon_T = 0\}$ and if, for all j, $\lim_T v_T(j)\varepsilon_T > 0$ then $j_{\max} = 0$.

In addition to [A2] in Section 3 of the text, the following conditions (reproduced from the main paper here for the convenience of the reader) are needed to establish the generalization of Theorem 2, with Assumptions [A4′]–[A6′] being modifications of Assumptions [A4]–[A6], respectively, that cater for the extension to varying rates of convergence for $\eta(z)$.

[A1′] Assumption [A1] holds andthere exist a positive definite matrix $\Sigma_T(\theta_0)$, $\kappa > 1$ and $\delta > 0$ such that for all $\|\theta - \theta_0\| \leq \delta$, $P_\theta\left(\|\Sigma_T(\theta_0)\{\eta(z) - b(\theta)\}\| > u\right) \leq \frac{c_0}{u^\kappa}$ for all $0 < u \leq \delta v_T(1)$ and $c_0 < +\infty$.

[A3′] Assumption [A3] holds, the function $b(\cdot)$ is continuously differentiable at θ_0 and the Jacobian $\nabla_\theta b(\theta_0)$ has full column rank k_θ.

[A4′] Given the sequence of $(k_\eta \times k_\eta)$ positive definite matrices $\Sigma_T(\theta)$ defined in (A4), for some $\delta > 0$ and all $\|\theta - \theta_0\| \leq \delta$,

$$\Sigma_T(\theta)\{\eta(z) - b(\theta)\} \Rightarrow \mathcal{N}(0, I_{k_\eta}),$$

where I_{k_η} is the $(k_\eta \times k_\eta)$ identity matrix.

[A5′] For all $\|\theta - \theta_0\| \leq \delta$, the sequence of functions $\theta \mapsto \Sigma_T(\theta) D_T^{-1}$ converges to some positive definite $A(\theta)$ and is equicontinuous at θ_0.

[A6′] For some positive δ and all $\|\theta - \theta_0\| \leq \delta$, and for all ellipsoids

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

with $\lim_T h_T(j) = 0$, for all $j \leq j_{\max}$ and all $u \in \mathbb{R}^{j_{\max}}$ fixed,

$$\begin{aligned} \lim_T \frac{P_\theta\left[\{\Sigma_T(\theta)\}_{[j_{\max}]}\{\eta(z) - b(\theta)\} - u \in B_T\right]}{\prod_{j=1}^{j_{\max}} h_T(j)} &= \varphi_{j_{\max}}(u), \\ \frac{P_\theta\left[\{\Sigma_T(\theta)\}_{[j_{\max}]}\{\eta(z) - b(\theta)\} - u \in B_T\right]}{\prod_{j=1}^{j_{\max}} h_T(j)} &\leq H(u), \quad \int H(u) du < +\infty, \end{aligned} \tag{A6}$$

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

THEOREM 5. *Assume that [A1′], with $\kappa > k_\theta$, [A2] and [A3′] -[A5′], 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) $\lim_T v_T(1)\varepsilon_T = +\infty$: With probability approaching one, the posterior distribution of $\varepsilon_T^{-1}(\theta - \theta_0)$ converges to the uniform distribution over the ellipse $\{x^\intercal B_0 x \leq 1\}$ with $B_0 = \nabla_\theta b(\theta_0)^\intercal \nabla_\theta b(\theta_0)$. In other words, for all f continuous and bounded, with probability approaching one

$$\int f\{\varepsilon_T^{-1}(\theta - \theta_0)\} d\Pi_\varepsilon(\theta|\eta_0) \to \frac{\int_{u^\intercal B_0 u \leq 1} f(u) du}{\int_{u^\intercal B_0 u \leq 1} du}. \tag{A7}$$

(ii) There exists $k_0 < k_\eta$ such that $\lim_T v_T(1)\varepsilon_T = \lim_T v_T(k_0)\varepsilon_T = c$, $0 < c < +\infty$, and $\lim_T v_T(k_0 + 1)\varepsilon_T = +\infty$: Assume Leb $\left(\sum_{j=1}^{k_0}\left[\{\nabla_\theta b(\theta_0)(\theta - \theta_0)\}_{[j]}\right]^2 \leq c\varepsilon_T^2\right) = +\infty$, then

$$\Pi_\varepsilon\left[\Sigma_T(\theta_0)\{b(\theta) - b(\theta_0)\} - Z_T^0 \in B|\eta_0\right] \to 0, \tag{A8}$$

for all bounded measurable sets B.

(iii) There exists $j_{\max} < k_\eta$ such that $\lim_T v_T(j_{\max})\varepsilon_T = 0$ and $\lim_T v_T(j_{\max} + 1)\varepsilon_T = +\infty$: Assume that Assumption [A6′] is satisfied, then (A7) *is satisfied.*

(iv) $\lim_T v_T(j)\varepsilon_T = c > 0$ *for all* $j \le k_\eta$ *or, and with reference to case* (ii), *Leb* $\left(\sum_{j=1}^{k_0}\left[\{\nabla_\theta b(\theta_0)(\theta-\theta_0)\}_{[j]}\right]^2 \le c\varepsilon_T^2\right) < +\infty$*: There exists a non-Gaussian probability distribution on* $\mathbb{R}^{k_\eta}$, Q_c *that depends on* c *and is such that*

$$\Pi_\varepsilon\left[\Sigma_T(\theta_0)\{b(\theta)-b(\theta_0)\} - Z_T^0 \in B|\eta_0\right] \to Q_c(B). \tag{A9}$$

More precisely, for $\Phi_{k_\eta}(\cdot)$ *the CDF of a* k_η*-dimensional standard normal random variable*

$$Q_c(B) \propto \int_B \Phi_{k_\eta}\left\{(Z-x)^\intercal A(\theta_0)^\intercal A(\theta_0)(Z-x) \le c^2\right\} dx$$

(v) $\lim_T v_T(k_\eta)\varepsilon_T = 0$*: Assume that Assumption [A6′] holds for* $j_{\max} = k_\eta$*, then*

$$\lim_{T\to+\infty} \Pi_\varepsilon\left[\Sigma_T(\theta_0)\{b(\theta)-b(\theta_0)\} - Z_T^0 \in B|\eta_0\right] = \Phi_{k_\eta}(B). \tag{A10}$$

Proof. We work with b instead of θ as the parameter, with injectivity of $\theta \mapsto b(\theta)$ required to re-state all results in terms of θ. Set $\Sigma_T(\theta_0)\{\eta(y)-b_0\} = Z_T^0$, for $b_0 = b(\theta_0)$, and $\eta_0 = \eta(y)$. We control the approximate Bayesian computation posterior expectation of non-negative and bounded functions $f_T(\theta - \theta_0)$:

$$\begin{aligned} E_{\Pi_\varepsilon}\{f_T(\theta-\theta_0)|\eta_0\} &= \int f_T(\theta-\theta_0) d\Pi_\varepsilon(\theta|\eta_0) \\ &= \int f_T(\theta-\theta_0)\mathbb{1}_{\|\theta-\theta_0\|\le\lambda_T} d\Pi_\varepsilon(\theta|\eta_0) + o_P(1) \\ &= \frac{\int_{\|\theta-\theta_0\|\le\lambda_T} p(\theta) f_T(\theta-\theta_0) P_\theta(\|\eta(z)-\eta(y)\|\le\varepsilon_T)\, d\theta}{\int_{\|\theta-\theta_0\|\le\lambda_T} p(\theta) P_\theta(\|\eta(z)-\eta(y)\|\le\varepsilon_T)\, d\theta} + o_P(1) \end{aligned}$$

where the second equality uses the posterior concentration of $\|\theta-\theta_0\|$ at the rate $\lambda_T \gg 1/v_T(1)$. Now,

$$\begin{aligned} \Sigma_T(\theta_0)\{\eta(z)-\eta(y)\} &= \Sigma_T(\theta_0)\{\eta(z)-b(\theta)\} + \Sigma_T(\theta_0)\{b(\theta)-b_0\} - \Sigma_T(\theta_0)\{\eta(y)-b_0\} \\ &= \Sigma_T(\theta_0)\{\eta(z)-b(\theta)\} + \Sigma_T(\theta_0)\{b(\theta)-b_0\} - Z_T^0. \end{aligned}$$

Set $Z_T = \Sigma_T(\theta_0)\{\eta(z)-b(\theta)\}$ and $Z_T^0 = \Sigma_T(\theta_0)\{\eta(y)-b(\theta_0)\}$. For fixed θ,

$$\|\Sigma_T^{-1}(\theta_0)\left[\Sigma_T(\theta_0)\{\eta(z)-b(\theta)\} - x\right]\| \asymp \|D_T^{-1}\left[\Sigma_T(\theta_0)\{\eta(z)-b(\theta)\} - x\right]\|$$

and

$$\Sigma_T(\theta_0)\{b(\theta)-b_0\} - Z_T^0 = \Sigma_T(\theta_0)\nabla_\theta b(\theta_0)(\theta-\theta_0)[1+o(1)] - Z_T^0 \in B.$$

Case (i) : $\lim_T v_T(1)\varepsilon_T = +\infty$. Consider $x(\theta) = \varepsilon_T^{-1}(b(\theta)-b_0)$ and $f_T(\theta-\theta_0) = f\{\varepsilon_T^{-1}(\theta-\theta_0)\}$, where $f(\cdot)$ is a non-negative, continuous and bounded function. On the event $\Omega_{n,0}(M) = \{\|Z_T^0\| \le M/2\}$ which has probability smaller than ϵ by choosing M large enough, we have that

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

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

$$\begin{aligned} P_\theta\left(\|\Sigma_T^{-1}(\theta_0)(Z_T-Z_T^0) + \varepsilon_T x\| \le \varepsilon_T\right) &\ge P_\theta\left\{\|\Sigma_T^{-1}(\theta_0)(Z_T-Z_T^0)\| \le \varepsilon_T(1-\|x\|)\right\} \\ &\ge P_\theta\left\{\|Z_T-Z_T^0\| \le v_T(1)\varepsilon_T(1-\|x\|)\right\} \ge 1-\epsilon \end{aligned}$$

as soon as $\|x\| \leq 1 - M/\{v_T(1)\varepsilon_T\}$ with M as above. This, combined with the continuity of $p(\cdot)$ at θ_0 and condition [A3′], implies that

$$
\begin{aligned}
&\int f\{\varepsilon_T^{-1}(\theta-\theta_0)\}d\Pi_\varepsilon(\theta|\eta_0) \\
&= \frac{\int_{\|\theta-\theta_0\|\leq\lambda_T} f\{\varepsilon_T^{-1}(\theta-\theta_0)\}P_\theta\left(\|\Sigma_T^{-1}(\theta_0)(Z_T-Z_T^0)+\varepsilon_T x\|\leq\varepsilon_T\right)d\theta}{\int_{\|\theta-\theta_0\|\leq\lambda_T} P_\theta\left(\|\Sigma_T^{-1}(\theta_0)(Z_T-Z_T^0)+\varepsilon_T x\|\leq\varepsilon_T\right)d\theta}\{1+o(1)\}+o_P(1) \\
&= \frac{\int_{\|x(\theta)\|\leq 1-M/\{v_T(1)\varepsilon_T\}} f\{\varepsilon_T^{-1}(\theta-\theta_0)\}d\theta}{\int_{\|x(\theta)\|\leq 1-M/\{v_T(1)\varepsilon_T\}} d\theta}\{1+o(1)\} \\
&+ \frac{\int_{\|\theta-\theta_0\|\leq\lambda_T} 1\!\!1_{\|x(\theta)\|>1-M/\{v_T(1)\varepsilon_T\}} f\{\varepsilon_T^{-1}(\theta-\theta_0)\}P_\theta\left(\|\Sigma_T^{-1}(\theta_0)(Z_T-Z_T^0)+\varepsilon_T x\|\leq\varepsilon_T\right)d\theta}{\int_{\|x(\theta)\|\leq 1-M/\{v_T(1)\varepsilon_T\}} d\theta}+o_P(1).
\end{aligned}
\tag{A11}
$$

The first term is approximately equal to

$$
N_1 = \frac{\int_{\|b(\varepsilon_T u+\theta_0)-b_0\|\leq 1} f(u)du}{\int_{\|b(\varepsilon_T u+\theta_0)-b_0\|\leq 1} du}
$$

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

$$
\int_{\|b(\varepsilon_T u+\theta_0)-b_0\|\leq\varepsilon_T} du = \int_{\|\nabla_\theta b(\theta_0)u\|\leq 1} du + o(1) = \int_{u^\intercal B_0 u\leq 1} du + o(1)
$$

with $B_0 = \nabla_\theta b(\theta_0)^\intercal \nabla_\theta b(\theta_0)$. This leads to

$$
N_1 = \int_{u^\intercal B_0 u\leq 1} f(u)du \Big/ \int_{u^\intercal B_0 u\leq 1} du.
$$

We now show that the second term of the right hand side of (A11) converges to 0. We split it into an integral over $1 + M\{v_T(1)\varepsilon_T\}^{-1} \geq \|x(\theta)\| \geq 1 - M\{v_T(1)\varepsilon_T\}^{-1}$ and $1 + M\{v_T(1)\varepsilon_T\}^{-1} \leq \|x(\theta)\|$. This leads, for the first part, to an upper bound

$$
N_2 \leq \frac{\|f\|_\infty \int_{1+M/\{v_T(1)\varepsilon_T\}\geq\|x(\theta)\|>1-M/\{v_T(1)\varepsilon_T\}} d\theta}{\int_{\|x(\theta)\|\leq 1-M/\{v_T(1)\varepsilon_T\}} d\theta} \lesssim \{v_T(1)\varepsilon_T\}^{-1} = o(1)
$$

Finally, for the third integral over $\|x(\theta)\| > 1 + M\{v_T(1)\varepsilon_T\}^{-1}$ we have

$$
\begin{aligned}
P_\theta\left(\|\Sigma_T^{-1}(\theta_0)(Z_T-Z_T^0)+\varepsilon_T x(\theta)\|\leq\varepsilon_T\right) &\leq P_\theta\left(\|\Sigma_T^{-1}(\theta_0)(Z_T-Z_T^0)\|\geq\varepsilon_T\|x(\theta)\|-\varepsilon_T\right) \\
&\leq P_\theta\left(\|Z_T-Z_T^0\|\geq v_T(1)\varepsilon_T(\|x(\theta)\|-1)\right) \leq c_0\{v_T(1)\varepsilon_T(\|x(\theta)\|-1)\}^{-\kappa},
\end{aligned}
$$

which leads to

$$
\begin{aligned}
N_3 &= \frac{\int_{\|\theta-\theta_0\|\leq\lambda_T} 1\!\!1_{\|x(\theta)\|>1+M/\{v_T(1)\varepsilon_T\}} f\{\varepsilon_T^{-1}(\theta-\theta_0)\}P_\theta\left(\|\Sigma_T^{-1}(\theta_0)(Z_T-Z_T^0)+\varepsilon_T x(\theta)\|\leq\varepsilon_T\right)d\theta}{\int_{\|x(\theta)\|\leq 1-M/\{v_T(1)\varepsilon_T\}} d\theta} \\
&\lesssim M^{-\kappa}\varepsilon_T^{-k_\theta}\int_{2\geq\|x\|>1+M/\{v_T(1)\varepsilon_T\}} d\theta + 2^\kappa\varepsilon_T^{-k_\theta}\int_{2\leq\|x(\theta)\|}\{v_T(1)\varepsilon_T\|x(\theta)\|\}^{-\kappa}d\theta \\
&\lesssim M^{-\kappa} + \varepsilon_T^{-k_\theta}\int_{c_1\varepsilon_T\leq\|\theta-\theta_0\|}\{v_T(1)\|\nabla_\theta b(\theta_0)(\theta-\theta_0)\|\}^{-\kappa}d\theta \lesssim M^{-\kappa}
\end{aligned}
$$

provided $\kappa > 1$. Since M can be chosen arbitrarily large, putting N_1, N_2 and N_3 together, we obtain that the approximate Bayesian computation posterior distribution of $\varepsilon_T^{-1}(\theta-\theta_0)$ is asymptotically uniform over the ellipsoid $\{x^\intercal B_0 x \leq 1\}$ and (i) is proved.

Case (ii) : $+\infty > \lim_T v_T(1)\varepsilon_T = c > 0$ and $\lim_T v_T(k_\eta)\varepsilon_T = +\infty$. We consider $f_T(\theta - \theta_0) = \mathbb{1}[\Sigma_T(\theta_0)\{b(\theta) - b_0\} - Z_T^0 \in B]$ and $x(\theta) = \Sigma_T(\theta_0)\{b(\theta) - b_0\} - Z_T^0$.

Set k_0 such that for all $j \leq k_0$, $\lim_T v_T(j)\varepsilon_T = c$ and for all $j > k_0$, $\lim_T v_T(j)\varepsilon_T = +\infty$. We write $\Sigma_T(\theta_0) = A_T(\theta_0)D_T$, so that $A_T(\theta_0) \to A(\theta_0)$ as $T \to +\infty$, where $A(\theta_0)$ is positive definite and symmetric.

$$
\begin{aligned}
P_\theta\left(\|\Sigma_T^{-1}(\theta_0)\left[\Sigma_T(\theta_0)\{\eta(z) - b(\theta)\} - x\right]\| \leq \varepsilon_T\right) &= P_\theta\left(\|D_T^{-1}A_T^{-1}(\theta_0)\left(Z_T - x\right)\| \leq \varepsilon_T\right) \\
&= P_\theta\left(\|D_T^{-1}(\tilde{Z}_T - x_T)\| \leq \varepsilon_T\right)
\end{aligned}
$$

where $\tilde{Z}_T = A_T^{-1}(\theta_0)Z_T \Rightarrow \mathcal{N}(0, A(\theta_0)I_{k_\eta}A(\theta_0)^\intercal)$ and $x_T = A_T^{-1}(\theta_0)x = A^{-1}(\theta_0)x + o_P(1)$.

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

$$
\begin{aligned}
&P_\theta\left(\|D_T^{-1}(\tilde{Z}_T - x_T)\| \leq \varepsilon_T\right) \leq P_\theta\left[\sum_{j=1}^{k_0}\{\tilde{Z}_T(j) - x_T(j)\}^2 \leq v_T(1)^2\varepsilon_T^2\right] \\
&\geq P_\theta\left[\sum_{j=1}^{k_0}\{\tilde{Z}_T(j) - x_T(j)\}^2 \leq v_T(1)^2\varepsilon_T^2[1 - M_T\{v_T(k_0+1)\varepsilon_T\}^{-2}]\right] \\
&\quad - P_\theta\left[\sum_{j=k_0+1}^{k}\{\tilde{Z}_T(j) - x_T(j)\}^2 > M_T^{-1}\{\varepsilon_T v_T(k_0+1)\}^{-2}\right] \\
&\geq P_\theta\left(\sum_{j=1}^{k_0}\{\tilde{Z}_T(j) - x_T(j)\}^2 \leq v_T(1)^2\varepsilon_T^2[1 - M_T^{-1}\{v_T(k_0+1)\varepsilon_T\}^{-2}]\right) - o(1)
\end{aligned}
\tag{A12}
$$

This implies that for all x and all $\|\theta - \theta_0\| \leq \lambda_T$

$$
P_\theta\left(\|D_T^{-1}(\tilde{Z}_T - x_T)\| \leq \varepsilon_T\right) = P_\theta\left(\sum_{j=1}^{k_0}\left[\left\{A^{-1}(\theta_0)Z_T\right\}(j) - \{A^{-1}(\theta_0)x\}(j)\right]^2 \leq c\right) + o(1)
$$

Since $A^{-1}(\theta_0)x = D_T\nabla_\theta b(\theta_0)(\theta - \theta_0) - A^{-1}(\theta_0)Z_T^0$, if

$$
Leb\left(\sum_{j=1}^{k_0}\left[\{\nabla_\theta b(\theta_0)(\theta - \theta_0)\}_{[j]}\right]^2 \leq c\varepsilon_T^2\right) = +\infty\,,
$$

then as in case (i) we can bound

$$
\Pi_\varepsilon\left\{\Sigma_T(\theta_0)(b - b_0) - Z_T^0 \in B|\eta^0\right\} \leq \frac{\int_{A^{-1}(\theta_0)x\in B} P_\theta\left(\sum_{j=1}^{k_0}\left[\left\{A^{-1}(\theta_0)Z_T\right\}(j) - z(j)\right]^2 \leq c\right)d\theta}{\int_{|\theta|\leq M} P_\theta\left(\sum_{j=1}^{k_0}\left[\{A^{-1}(\theta_0)Z_T\}(j) - z(j)\right]^2 \leq c\right)d\theta} + o_P(1)
$$

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

Case (iii) : $\lim_T v_T(1)\varepsilon_T = 0$ and $\lim_T v_T(k_\eta)\varepsilon_T = +\infty$. Again we consider $f_T(\theta - \theta_0) = \mathbb{1}[\Sigma_T(\theta_0)\{b(\theta) - b_0\} - Z_T^0 \in B]$ and $x(\theta) = \Sigma_T(\theta_0)\{b(\theta) - b_0\} - Z_T^0$. As in the computations leading to (A12), we have: setting $e_T = M_T(v_T(k_0+1)\varepsilon_T^{-1})^{-2}$, under Assumption [A6′],

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

when $\|\theta-\theta_0\| < \lambda_T$ where $\varphi_{j_{\max}}$ is the centered Gaussian density of the $j_{\max}$ dimensional vector, with covariance $\{A(\theta_0)^2\}_{[j_{\max}]}$. This implies as in case (ii) that, with probability going to 1

$$\limsup_T \Pi_\varepsilon\left\{\Sigma_T(\theta_0)(b-b_0) - Z_T^0 \in B|\eta_0\right\} \le \frac{\int_{A(\theta_0)B}\varphi_{j_{\max}}(x_{[j_{\max}]})dx}{\int_{|x|\le M}\varphi_{j_{\max}}(x_{[j_{\max}]})dx} \le M^{-(k_\eta - j_{\max})}$$

and choosing M arbitrary large leads to equation (11) in the text.

Case (iv) : If $\lim_T v_T(j)\varepsilon_T = c > 0$ for all $j \le k_\eta$. To prove equation (13) in the text, we use the computation of case (ii) with $k_0 = k_\eta$, so that (A12) implies that

$$\begin{aligned}
P_\theta\left(\|D_T^{-1}(\tilde{Z}_T - x_T)\| \le \varepsilon_T\right) &= P_\theta\left(\|\tilde{Z}_T - x_T\|^2 \le v_T(1)^2\varepsilon_T^2\right) \\
&= P\left(\|A^{-1}(\theta_0)Z_T - A^{-1}(\theta_0)x\|^2 \le c^2\right) + o(1)
\end{aligned}$$

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

$$\begin{aligned}
\Pi_\varepsilon\left\{\Sigma_T(\theta_0)(b-b_0) - Z_T^0 \in B|\eta^0\right\} &\le \frac{\int_{x\in B} P_\theta\left(\|A^{-1}(\theta_0)Z_T - A^{-1}(\theta_0)x\|^2 \le c^2\right)dx}{\int_{|x|\le M} P_\theta\left(\|A^{-1}(\theta_0)Z_T - A^{-1}(\theta_0)x\|^2 \le c^2\right)dx} \\
&\ge \frac{\int_{x\in B} P_\theta\left(\|A^{-1}(\theta_0)Z_T - A^{-1}(\theta_0)x\|^2 \le c^2\right)dx}{\int_{|x|\le M} P_\theta\left(\|A^{-1}(\theta_0)Z_T - A^{-1}(\theta_0)x\|^2 \le c^2\right)dx + \delta} + o_P(1)
\end{aligned}$$

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

$$\int_{|x|\le M} P_\theta\left(\|\tilde{Z}_T - A^{-1}(\theta_0)x\|^2 \le c^2\right)dx \to \int_{x\in\mathbb{R}^{k_\eta}} P_\theta\left(\|\tilde{Z}_T - A^{-1}(\theta_0)x\|^2 \le c^2\right)dx < +\infty,$$

the result follows.

Case (v) : $\lim_T v_T(k)\varepsilon_T = 0$. Take $\Sigma_T(\theta_0) = A_T(\theta_0)D_T$. For some $\delta > 0$ and all $\|\theta-\theta_0\| \le \delta$,

$$\begin{aligned}
P_\theta\left(\left\|D_T^{-1}\{A_T^{-1}(\theta_0)Z_T - A_T^{-1}(\theta_0)x\}\right\| \le \varepsilon_T\right) &= P_\theta\left(\left\|D_T^{-1}\{A^{-1}(\theta_0)Z_T - A^{-1}(\theta_0)x\}\right\| \le \varepsilon_T\right) + o(1) \\
&= P_\theta\left(\left\|A^{-1}(\theta_0)Z_T - A^{-1}(\theta_0)x\right\|^2 \le v_T^2(k)\varepsilon_T^2\right) + o(1) \\
&= P_\theta\left[\left\{A^{-1}(\theta_0)Z_T - A^{-1}(\theta_0)x\right\} \in B_T\right] + o(1).
\end{aligned}$$

From both assertions of Assumption [A6′] and the dominated convergence theorem, the above implies (for $j_{\max} = k_\eta$)

$$\frac{1}{\prod_{j=1}^{k_\eta}\varepsilon_T v_T(j)}\int P_\theta\left[\left\{A^{-1}(\theta_0)Z_T - A^{-1}(\theta_0)x\right\} \in B_T\right]dx = \int \varphi_{k_\eta}(x)dx + o(1) = 1 + o(1)\,.$$

Likewise, similar arguments yield

$$\frac{1}{\prod_{j=1}^{k_\eta}\varepsilon_T v_T(j)}\int 1\!\!1_{x\in B}P_\theta\left[\left\{A^{-1}(\theta_0)Z_T - A^{-1}(\theta_0)x\right\}\in B_T\right]dx = \int 1\!\!1_{x\in B}\varphi_{k_\eta}(x)dx + o(1)$$
$$= \Phi_{k_\eta}(B) + o(1).$$

Together, these two equivalences yield the desired result. □

A·3. *Proof of Theorem 3*

Proof of Theorem 3, Case (i). $v_T\varepsilon_T \to +\infty$ Defining $b = b(\theta)$, $b_0 = b(\theta_0)$ and $x = v_T(b - b_0) - Z_T^0$. With a slight abuse of notation, in this proof we let $Z_T^0 = v_T\{\eta(y) - b(\theta_0)\}$. We approximate the ratio

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

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

$$\begin{aligned}N_T &= \int xP_x\left(|\eta(z)-\eta(y)|\le\varepsilon_T\right)p\{b_0+(x+Z_T^0)/v_T\}dx\\ &= \int_{|x|\le v_T\varepsilon_T - M} xP_x\left(|Z_T+x|\le v_T\varepsilon_T\right)p\{b_0+(x+Z_T^0)/v_T\}dx \\ &\quad+ \int_{|x|\ge v_T\varepsilon_T - M} xP_x\left(|Z_T+x|\le v_T\varepsilon_T\right)p\{b_0+(x+Z_T^0)/v_T\}dx,\end{aligned} \tag{A13}$$

where the condition $\lim_T v_T\varepsilon_T = +\infty$ 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 (A13). If $0 \le x \le v_T\varepsilon_T - M$, then

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

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

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

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

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

A Taylor expansion of $p\{b_0 + (x+Z_T^0)/v_T\}$ around $\gamma_0 = b_0 + Z_T^0/v_T$ then leads to, for $\nabla_b^j p(\theta)$ the j-th derivative of $p(b)$ (with respect to b),

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

where $k=\lfloor\beta/2\rfloor$. We split the second integral of (A13) into $v_T\varepsilon_T - M \le |x| \le v_T\varepsilon_T + M$ and $|x| \ge v_T\varepsilon_T + M$. We treat the latter as before: with probability going to 1,

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

Finally, we study the second integral term for N_T in (A13) over $v_T\varepsilon_T - M \le |x| \le v_T\varepsilon_T + M$. Using the assumption that $p(\cdot)$ is Hölder we obtain that

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

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,

$$\int_{-M}^{M}\left\{P_y\left(Z_T\le -y\right) - P_y\left(Z_T\ge -y\right)\right\}dy = Mo(1)$$

and

$$\int_{-M}^{M} y\left\{P_y\left(Z_T\le -y\right) + P_y\left(Z_T\ge -y\right)\right\}dy = \int_{-M}^{M} y\left\{1+o(1)\right\}dy = M^2o(1).$$

This implies that

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

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

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

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

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

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

$$\frac{N_T}{D_T} = v_T\sum_{j=1}^{k}\frac{\nabla_b^{(2j-1)}p(b_0)}{p(b_0)(2j-1)!(2j+1)}\varepsilon_T^{2j} + o_P(1) + O(\varepsilon_T^{1+\beta}v_T) \tag{A14}$$

Using the definition of N_T/D_T, dividing (A14) by v_T, and rearranging terms yields

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

To obtain the posterior mean of θ, we write

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

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

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

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

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

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

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

This implies, in particular, that

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

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

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

and $E_{\Pi_\varepsilon}\{v_T(\theta - \theta_0)\} \Rightarrow \mathcal{N}(0, V(\theta_0)/\{\nabla_\theta b(\theta_0)\}^2)$, while if $v_T \varepsilon_T^2 \to +\infty$

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

assuming $\beta \geq 3$. □

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

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

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

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

which can be rewritten as

$$E_{\Pi_\varepsilon}\{v_T(b - b_0)\} - Z_T^0 = \int \frac{x P_x(|\eta(y) - \eta(z)| \leq \varepsilon_T) p[b(\theta_0) + \{x + Z_T^0\}/v_T] dx}{\int P_x(|\eta(y) - \eta(z)| \leq \varepsilon_T) p\{b(\theta_0) + (x + Z_T^0)/v_T\} dx}$$

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

$$E_{\Pi_\varepsilon}[v_T\{b - b_0\}] - Z_T^0 = \int \frac{x P_x(|Z_T + x| \leq v_T\varepsilon_T) p\{b(\theta_0) + (x + Z_T^0)/v_T\} dx}{\int P_x(|Z_T + x| \leq v_T\varepsilon_T) p\{b(\theta_0) + (x + Z_T^0)/v_T\} dx} = \frac{N_T}{D_T}.$$

By injectivity of the map $\theta \mapsto b(\theta)$ (Assumptions [A3]) and Assumption [A4], the result follows when $E_{\Pi_\varepsilon}\{v_T(b - b_0)\} - Z_T^0 = o_P(1)$.

Consider first the denominator. Define $h_T = v_T\varepsilon_T$ and $V_0 = V(\theta_0) \equiv \lim_T \text{var}[v_T\{\eta(y) - b(\theta_0)\}]$. Using arguments that mirror those in the proof of Theorem 2 part (v), by Assumption [A6′] and the dominated convergence theorem

$$\frac{D_T}{p(b_0)h_T} = h_T^{-1} \int P_x(|Z_T + x| \leq h_T) dx + o_P(1) = \int \varphi\{x/V_0^{1/2}\} dx + o_P(1) = 1 + o_P(1),$$

where the second equality follows from Assumption [A6] and the dominated convergence theorem. 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 [A6] and [A7],

$$\begin{aligned}\frac{N_T}{h_T} &= \int x P_x^*(|Z_T + x| \leq h_T) p\{b_0 + (x + Z_T^0)/v_T\} dx \\ &= p(b_0) \int x \varphi\{x/V_0^{1/2}\} dx + \int x \left\{ P_x^*(|Z_T + x| \leq h_T) - \varphi\{x/V_0^{1/2}\} \right\} \\ &\quad \times p\{b_0 + (x + Z_T^0)/v_T\} dx + o_P(1).\end{aligned}$$

If $h_T \to c > 0$, then

$$\begin{aligned}\frac{N_T}{h_T} &= p(b_0) \int x P\{|\mathcal{N}(0,1) + x/V_0^{1/2}| \leq c/V_0^{1/2}\} dx \\ &\quad + \int x \left[P_x^*(|Z_T + x| \leq h_T) - P\{|\mathcal{N}(0,1) + x/V_0^{1/2}| \leq c/V_0^{1/2}\} \right] p\{b_0(x + Z_T^0)/v_T\} dx + o_P(1).\end{aligned} \tag{A15}$$

The result now follows if $\int x \left[P_x^*(|Z_T + x| \leq h_T) - \varphi\{x/V_0^{1/2}\}\right] p\{b_0 + (x + Z_T^0)/v_T\}dx = o_P(1)$, respectively, $P_x^*(|Z_T + x| \leq h_T) - P\{|\mathcal{N}(0,1) + x/V_0^{1/2}| \leq c/V_0^{1/2}\} = o(1)$, for which a sufficient condition is that

$$\int |x| \left|P_x^*(|Z_T + x| \leq h_T) - \varphi\{x/V_0^{1/2}\}\right| p\{b_0 + (x + Z_T^0)/v_T\}dx = o_P(1), \tag{A16}$$

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

To show that the integral in (A16) is $o_P(1)$ we break the region of integration into three areas: (i) $|x| \leq M$; (ii) $M \leq |x| \leq \delta v_T$; (iii) $|x| \geq \delta v_T$.

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

$$\sup_{x:|x|\leq M} |p\{b_0 + (x + Z_T^0)/v_T\} - p(b_0)| = o_P(1)$$
$$\sup_{x:|x|\leq M} |P^*(|Z_T + x| \leq h_T) - \varphi\{x/V_0^{1/2}\}| = o_P(1).$$

The first equation is satisfied by [A7] and the fact that by [A4] $Z_T^0/v_T = o_P(1)$. The second term follows from [A7] and the dominated convergence theorem. We can now conclude that equation (A16) is $o_P(1)$ over $|x| \leq M$.

The same holds for the first term in equation (A15), without requiring [A7].

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

$$\int_{M\leq|x|\leq\delta v_T} |x| P_x^*(|Z_T + x| \leq h_T) p\{b_0 + (x + Z_T^0)/v_T\}dx$$

is finite.

When $|x| > M$, $|Z_T + x| \leq h_T$ implies that $|Z_T| > |x|/2$ since $h_T = O(1)$. Hence, using Assumption [A1′],

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

which in turns implies that

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

The same computation can be conducted in the case (A15).

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

$$N_3 = \frac{1}{h_T} \int_{|x|\geq\delta v_T} |x| P_x^*(|Z_T + x| \leq h_T) p\{b_0 + (x + Z_T^0)/v_T\}dx.$$

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

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

Since $\eta(y) = b_0 + O_P(1/v_T)$ we have, for T large,

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

where [A1′] and [A7] ensure $\int c(b)|b|p(b)db<\infty$. The same computation can be conducted in the case (A15).

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

A·4. *Proof of Theorem 4*

The proof follows the same lines as the proof of Theorem 3, with some extra technicalities due to the multivariate nature of θ. Define $G_0=\nabla_\theta b(\theta_0)$ and let $x(\theta)=v_T(\theta-\theta_0)-(G_0^\intercal G_0)^{-1}G_0^\intercal Z_T^0$, where $Z_T^0=v_T\{\eta(y)-b(\theta_0)\}$. We show that $E_{\Pi_\epsilon}\{x(\theta)\}=o_p(1)$. We write

$$E_{\Pi_\epsilon}\{x(\theta)\}=\frac{\int_\Theta x(\theta)P_\theta\left(\|\eta(z)-\eta(y)\|\leq\varepsilon_T\right)p(\theta)d\theta}{\int_\Theta P_\theta\left(\|\eta(z)-\eta(y)\|\leq\varepsilon_T\right)p(\theta)d\theta}=\frac{N_T}{D_T},$$

and study the numerator and denominator separately. Since for all $\epsilon>0$ there exists $M_\epsilon>0$ such that, for all $M>M_\epsilon$, $P_{\theta_0}(\|Z_T^0\|>M/2)<\epsilon$, we can restrict ourselves to the event $\|Z_T^0\|\leq M/2$ for some M large.

We first study the numerator N_T and we split Θ into $\{\|G_0x(\theta)\|\leq v_T\varepsilon_T-M\}$, $\{v_T\varepsilon_T-M\leq\|G_0x(\theta)\|\leq v_T\varepsilon_T+M\}$ and $\{\|G_0x(\theta)\|>v_T\varepsilon_T+M\}$. The first integral is equal to

$$\begin{aligned}I_1&=p(\theta_0)\int_{\|G_0x(\theta)\|\leq v_T\varepsilon_T-M}\{x(\theta)+O(v_T\varepsilon_T^2)\}P_\theta\left(\|\eta(z)-\eta(y)\|\leq\varepsilon_T\right)d\theta\\&=p(\theta_0)\int_{\|G_0x(\theta)\|\leq v_T\varepsilon_T-M}\{x(\theta)+O(v_T\varepsilon_T^2)\}d\theta\\&-p(\theta_0)\int_{\|G_0x(\theta)\|\leq v_T\varepsilon_T-M}\{x(\theta)+O(v_T\varepsilon_T^2)\}P_\theta\left(\|\eta(z)-\eta(y)\|>\varepsilon_T\right)d\theta\end{aligned}$$

The first term in I_1 can be made arbitrarily small for M large enough. For the second term in I_1, we note

$$\begin{aligned}v_T\varepsilon_T<\|v_T\{\eta(z)-\eta(y)\}\|&=\|Z_T-Z_T^0+v_TG_0(\theta-\theta_0)\|+O(\|\theta-\theta_0\|^2)\\&=\|Z_T-P_{G_0}^\perp Z_T^0+G_0x(\theta)\|+O(\|\theta-\theta_0\|^2)\\&\leq\|Z_T\|+\|P_{G_0}^\perp Z_T^0\|+\|G_0x(\theta)\|+O(\|\theta-\theta_0\|^2)\\&\leq\|Z_T\|+M/2+\|G_0x(\theta)\|+O(\|\theta-\theta_0\|^2)\end{aligned}$$

where $P_{G_0}^\perp$ is the orthogonal projection onto the vector space that is orthogonal to G_0 and $Z_T=\{\eta(z)-b(\theta)\}$. Therefore, if $\|G_0x(\theta)\|\leq v_T\varepsilon_T-M$, then

$$M/2\leq v_T\varepsilon_T-M/2-\|G_0x(\theta)\|\leq\|Z_T\|.$$

Hence, the second term of the right hand side of I_1 is bounded by a constant times

$$\begin{aligned}&\int_{\|G_0x(\theta)\|\leq v_T\varepsilon_T-M}2\|G_0x(\theta)\|P_\theta\left(\|Z_T\|>\varepsilon_Tv_T-M/2-\|G_0x(\theta)\|\right)d\theta\\&\lesssim\int_{\|G_0x(\theta)\|\leq v_T\varepsilon_T-M}\frac{\|G_0x(\theta)\|}{(v_T\varepsilon_T-M/2-\|G_0x(\theta)\|)^\kappa}d\theta\\&\lesssim v_T^{-k_\theta}\int_0^{v_T\varepsilon_T-M}\frac{r^{k_\theta}}{(v_T\varepsilon_T-M/2-r)^\kappa}dr\lesssim\varepsilon_T^{k_\theta}M^{-\kappa}\end{aligned}$$

The integral over $\{\|G_0x(\theta)\|>v_T\varepsilon_T+M\}$, I_3, is treated similarly. This leads to $\|I_1+I_3\|\leq M^{-\kappa}\varepsilon_T^{k_\theta}$.

Likewise, using similar arguments we can show

$$D_T \gtrsim \int_{\|G_0 x(\theta)\|\leq v_T\varepsilon_T - M} P_\theta\left(\|\eta(z)-\eta(y)\|\leq \varepsilon_T\right)d\theta \gtrsim \varepsilon_T^{k_\theta}.$$

All that remains is to prove that the second integral I_2, the integral over $\{v_T\varepsilon_T - M \leq \|G_0x(\theta)\| \leq v_T\varepsilon_T + M\}$, is $o_p(\varepsilon_T^{k_\theta})$.

$$I_2 = \int_{v_T\varepsilon_T - M\leq\|G_0x(\theta)\|\leq v_T\varepsilon_T+M}\{x(\theta)+O(v_T\varepsilon_T^2)\}P_\theta\left(\|\eta(z)-\eta(y)\|\leq\varepsilon_T\right)d\theta.$$

Since

$$v_T^2\|\eta(z)-\eta(y)\|^2 = \|Z_T - P_{G_0}^{\perp}Z_T - G_0x(\theta)\|^2 = \|Z_T - P_{G_0}^{\perp}Z_T^0\|^2 + \|G_0x(\theta)\|^2 - 2\langle Z_T, G_0x(\theta)\rangle,$$

where $\langle\cdot,\cdot\rangle$ is the inner product, setting $u=(G_0^{\intercal}G_0)^{1/2}x(\theta)\|G_0x(\theta)\|^{-1}$, $r=\|G_0x(\theta)\|$, $\Gamma_0 = (G_0^{\intercal}G_0)^{-1/2}G_0^{\intercal}$, then, for $\mathcal{S}=\{u\in\mathbb{R}^{k_\theta};\|u\|=1\}$,

$$\begin{aligned}
I_2 &= v_T^{-k_\theta}(G_0^{\intercal}G_0)^{-1/2}\int_{v_T\varepsilon_T-M}^{v_T\varepsilon_T+M} r^{k_\theta}\int_{u\in\mathcal{S}} uP_\theta\left(\|Z_T-P_{G_0}^{\perp}Z_T^0\|^2+r^2-2r\langle\Gamma_0Z_T,u\rangle\leq v_T^2\varepsilon_T^2\right)d\lambda(u)dr \\
&\quad + O\left(v_T\varepsilon_T^{2+k_\theta}\right) \\
&= v_T^{-k_\theta}(G_0^{\intercal}G_0)^{-1/2}\int_{-M}^{M}(v_T\varepsilon_T+r)^{k_\theta}\times \\
&\quad \int_{u\in\mathcal{S}} uP_\theta\left(\|Z_T-P_{G_0}^{\perp}Z_T^0\|^2-2r\langle\Gamma_0Z_T,u\rangle-2\varepsilon_Tv_T\langle\Gamma_0Z_T,u\rangle\leq -r^2-2rv_T\varepsilon_T\right)d\lambda(u)dr \\
&\quad + O\left(v_T\varepsilon_T^{2+k_\theta}\right),
\end{aligned}$$

where $\lambda(.)$ is the Lebesgue measure on $\mathcal{S}$. Moreover, we have

$$\begin{aligned}
&P_\theta\left(\|Z_T-P_{G_0}^{\perp}Z_T^0\|^2-2r\langle\Gamma_0Z_T,u\rangle-2\varepsilon_Tv_T\langle\Gamma_0Z_T,u\rangle\leq -r^2-2rv_T\varepsilon_T\right) \\
&= P_\theta\left(\langle\Gamma_0Z_T,u\rangle\geq\frac{r\varepsilon_Tv_T}{r+\varepsilon_Tv_T}+\frac{\|Z_T-P_{G_0}^{\perp}Z_T^0\|^2+r^2}{2(v_T\varepsilon_T+r)}\right)
\end{aligned}$$

and for any $a_T > M$ with $a_T = o(v_T\varepsilon_T)$,

$$\begin{aligned}
&P_\theta\left(\|Z_T-P_{G_0}^{\perp}Z_T^0\|^2\geq a_T\right)\lesssim c_0a_T^{-\kappa/2}, \\
&|P_\theta\left\{\langle\Gamma_0Z_T,u\rangle\geq r\right\}-P_\theta\left\{\langle\Gamma_0Z_T,u\rangle\geq r-2a_T/(v_T\varepsilon_T)\right\}| = o(1),
\end{aligned}$$

with for all r and u, $P_\theta\{\langle\Gamma_0Z_T,u\rangle\geq r\} = [1-\Phi\{r/\|\Gamma_0A(\theta_0)^{1/2}\|\}]+o(1)$. Since for all $r\in[-M,M]$, $(v_T\varepsilon_T+r)^{k_\theta} = (v_T\varepsilon_T)^{k_\theta}+O(M(v_T\varepsilon_T)^{k_\theta-1})$, the dominated convergence theorem implies

$$I_2 = \varepsilon_T^{k_\theta}(G_0^{\intercal}G_0)^{-1/2}\int_{-M}^{M}\int_{u\in\mathcal{S}}u[1-\Phi\{r/\|\Gamma_0A(\theta_0)^{1/2}\|\}]d\lambda(u)dr + o(\varepsilon_T^{k_\theta}) = o(\varepsilon_T^{k_\theta})$$

which completes the proof.

A·5. *Proof of Corollary 1*

Consider first the case where $\varepsilon_T = o(v_T^{-1})$. Using the same types of computations as in the proof of Theorem 5 in this Supplementary Material, we have, for $Z_T = \Sigma_T(\theta_0)\{\eta(z)-b(\theta)\}$,

$$\begin{aligned}
\alpha_T &= \int_\Theta P_\theta\left(\|Z_T-Z_T^0-v_T\{b(\theta)-b(\theta_0)\}\|\leq\varepsilon_Tv_T\right)p(\theta)d\theta \\
&\asymp (\varepsilon_Tv_T)^{k_\eta}\int_\Theta\varphi\{Z_T^0+v_T\nabla_\theta b(\theta_0)(\theta-\theta_0)\}d\theta \asymp \varepsilon_T^{k_\eta}v_T^{k_\eta-k_\theta}\lesssim v_T^{-k_\theta}.
\end{aligned}$$

In the case where $\varepsilon_T \gtrsim v_T^{-1}$, then the proof of Theorem 5 implies that

$$\alpha_T = \text{pr}\left[\|\eta(z) - \eta(y)\| \leq \varepsilon_T\right] \asymp \int_\Theta \varphi\{Z_T^0 + v_T b'(\theta_0)(\theta - \theta_0)\} d\theta \asymp \varepsilon_T^{k_\theta}.$$

A·6. *Proof of Corollary 2*

Proof. The proof is a direct consequence of Corollary 1 and a central limit theorem based on the independent draws from Algorithm 1, which allows us to deduce that $\hat{\alpha}_T = \alpha_T + o_p(\alpha_T)$. The result then follows from algebra. □

B. ILLUSTRATIVE EXAMPLE

In this section we demonstrate the implications of Theorems 1–3 using a particular model, namely, the moving average model of order two introduced in Example 1 and discussed in Remarks 2, 6 and 13. We first remind the reader of this model before using it to illustrate the results of each theorem in the three subsequent sections.

We have T observations from the data generating process

$$y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2}, \tag{B1}$$

where, for simplicity, we consider independent $e_t \sim \mathcal{N}(0, 1)$. Our prior belief for $\theta = (\theta_1, \theta_2)^\intercal$ is uniform over the invertibility region

$$\{(\theta_1, \theta_2)^\intercal : \ -2 \leq \theta_1 \leq 2, \ \theta_1 + \theta_2 \geq -1, \theta_1 - \theta_2 \leq 1\}. \tag{B2}$$

Following Marin *et al.* (2011), we choose as summary statistics for Algorithm 1 the sample autocovariances $\eta_j(y) = \frac{1}{T}\sum_{t=1+j}^T y_t y_{t-j}$, for $j = 0, 1, 2$ so that $\eta(y) = [\eta_0(y), \eta_1(y), \eta_2(y)]^\intercal$. From the definition of $\eta(z)$ and the process in (B1) the j-th component of $b(\theta)$, $j = 0, 1, 2$, is $b_j(\theta) = E_\theta(z_t z_{t-j})$ and $b(\theta)$ has the simple analytical form:

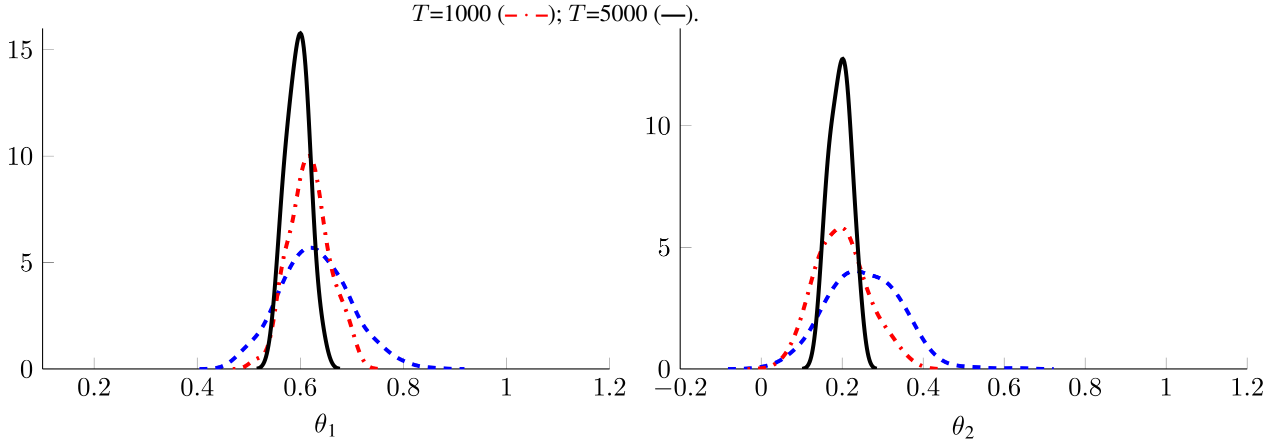
$$\theta \mapsto b(\theta) = \begin{pmatrix} 1 + \theta_1^2 + \theta_2^2 \\ \theta_1 + \theta_1\theta_2 \\ \theta_2 \end{pmatrix}.$$

The following subsections demonstrate the implications of the limit results in the main text within the confines of the above example, using 'observed' data y simulated from (B1). By simultaneously shifting the sample size T and the T-dependent tolerance parameter ε_T, we are able to demonstrate graphically Theorems 1–3 in a coherent manner that illustrates the importance of these results to practitioners, and which stands as a warning in regards to the choice of the tolerance in approximate Bayesian computation applications.

For each of the three demonstrations, which correspond to Theorems 1-3 in the text, we consider minor variants of the following general simulation design: the true parameter vector generating the simulated 'observed' data is fixed at $\theta_0 = (\theta_{1,0}, \theta_{2,0})^\intercal = (0.6, 0.2)^\intercal$; for a given sample length $T \in \{500, 1000, 50000\}$, $y = (y_1, y_2, ..., y_T)^\intercal$, is generated from the process in equation (B1); for each given sample of size T, the posterior is then estimated via Algorithm 1 (with the tolerance chosen to be a particular order of T) using $N = 50,000$ Monte Carlo draws taken from uniform priors satisfying (B2). For every demonstration, we take $d_1\{\theta, \theta_0\} = \|\theta - \theta_0\|$ and $d_2\{\eta(z), \eta(y)\} = \|\eta(z) - \eta(y)\|$.

As all three theorems illustrate, the choice of ε_T is critical in determining the large sample behavior of the approximate Bayesian computation posterior distribution and the approximate Bayesian computation posterior mean. To highlight this fact, we illustrate the results in a range of numerical experiments using different choices for the tolerances. In particular, and with reference to the illustration of Theorem 2, the choices $\varepsilon_T \in \{1/T^{0.4}, 1/T^{0.5}, 1/T^{0.55}\}$ (where $v_T = T^{0.5}$ for this example) correspond to the distinct Cases (i), (ii) and (iii), and the use of all three tolerances serves to highlight the distinction between what condition on ε_T is required for posterior concentration, and what is required to yield asymptotic normality of the posterior. With regard to Theorem 3, the choice, $\varepsilon_T = 1/T^{0.4}$, is used to highlight Case (i), whereby

Figure 2. Posterior concentration demonstration. Estimated posterior results according to sample size: T=500 (- - -); T=1000 (– · –); T=5000 (—).



asymptotic normality of the posterior mean can be achieved despite a lack of Gaussianity for the posterior itself, whilst $\varepsilon_T \in \{1/T^{0.5}, 1/T^{0.55}\}$ correspond to Case (ii).

B·1. *Theorem 1*

Theorem 1 implies that under regularity, as $T \to +\infty$, the posterior measure $\Pi_\varepsilon(\cdot|\eta_0)$, with $\eta_0 = \eta(y)$, concentrates on sets containing θ_0; i.e., $\Pi_\varepsilon(\|\theta - \theta_0\| \leq \delta|\eta_0) = 1 + o_P(1)$ for all $\delta > 0$. For this to be satisfied, it must be that $\lim_T \varepsilon_T = 0$. We adopt $\varepsilon_T = 1/T^{0.4}$ for this demonstration, running Algorithm 1 with this choice of tolerance, and taking $N = 50,000$ draws from the prior. The results are presented in Figure 2. Note that in order to fix a level of Monte Carlo error, for each sample size we retain 100 simulated values of θ that lead to realizations of $\|\eta(y) - \eta(z)\|$ less than this tolerance. In this way, we are using the nearest-neighbor interpretation of Algorithm 1.

From Figure 2 it is clear that the posterior measure $\Pi_\varepsilon(\cdot|\eta_0)$ is concentrating on $\theta_0 = (0.6, 0.2)^\intercal$. However, the corresponding p-value associated with the KS-test for the scaled and centered posterior distributions of θ_1 and θ_2 would lead to rejection (at the 5% level) of the null hypothesis that these posteriors are standard normal for each of the considered sample sizes. The results in Figure 2 reflect the fact that a tolerance proportional to $\varepsilon_T = 1/T^{0.4}$ is small enough to yield posterior concentration but not small enough to yield a posterior with asymptotically Gaussian shape. We elaborate on this point in the next subsection.

B·2. *Theorem 2*

Theorem 2 states that the shape of the (standardized) posterior measure is determined in large part by the speed at which ε_T goes to 0. If this convergence is too slow, then the posterior will have a non-standard asymptotic shape, in the sense that it will not be approximately Gaussian in large samples. The illustration in the previous section highlights the lack of asymptotic Gaussianity for the posterior measure under Case (i) of Theorem 2, whereby $\lim_T v_T\varepsilon_T = +\infty$. Now, consider the two alternative values for the tolerance, $\varepsilon_T \in \{1/T^{0.5}, 1/T^{0.55}\}$, which respectively fit with Case (ii) ($\lim_T v_T\varepsilon_T = c > 0$) and Case (iii) ($\lim_T v_T\varepsilon_T = 0$), with only the latter case yielding asymptotic Gaussianity of the posterior (the Bernstein-von Mises result). These results are displayed in Figures 3 and 4 for $T = 500$ and $T = 1000$ (respectively).

Figure 3 demonstrate that at $T = 500$, and for $\varepsilon_T = 1/T^{0.5}$, neither posterior for θ_1 or θ_2 has a shape that is particularly Gaussian. The p-value associated with the KS-test for the scaled and centered posterior distributions in each panel would lead to a rejection (at the 5% level) of the null hypothesis that these

Figure 3. Comparison of different tolerance rules for ε_T : $\varepsilon_T = 1/T^{0.5}$ (– · –); $\varepsilon_T = 1/T^{0.55}$ (—); The sample size is $T = 500$.

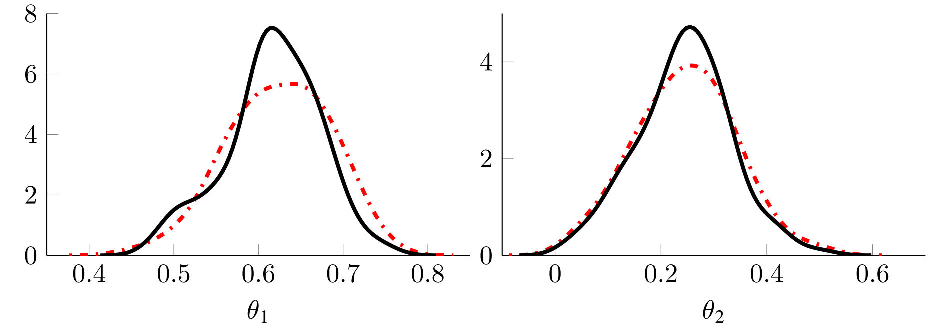
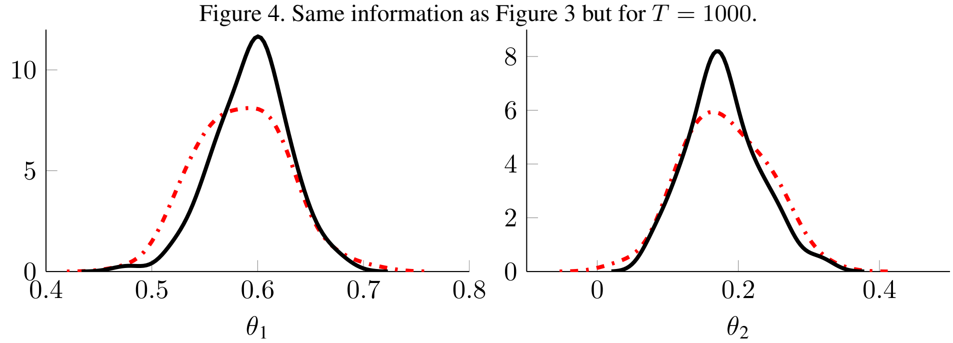


Figure 4. Same information as Figure 3 but for $T = 1000$.



posteriors are standard normal. However, for θ_2 and the tolerance $\varepsilon_T = 1/T^{0.55}$, the p-value associated with the corresponding KS-test of Gaussianity would not lead us to a rejection of the null hypothesis.

At $T = 1000$, and for both θ_1 and θ_2, Figure 4 demonstrates that the posteriors based on the tolerance $\varepsilon_T = 1/T^{0.55}$ (which satisfies the conditions for the Bernstein-von Mises result) are approximately Gaussian. Indeed, the p-value associated with the KS-test for the scaled and centered posterior distributions in each of the two panels would not lead to a rejection (at the 5% level) of the null hypothesis that these posteriors are standard normal. However, the corresponding KS-test would lead to a rejection for the posteriors calculated from the tolerance $\varepsilon_T = 1/T^{0.5}$.

A key practical insight from Theorem 2 is that for the posterior $\Pi_\varepsilon(\cdot|\eta_0)$ to be asymptotically Gaussian, and thus for credible regions built from $\Pi_\varepsilon(\cdot|\eta_0)$ to have asymptotically correct frequentist coverage, it must be that $\varepsilon_T = o(1/v_T)$, where v_T is such that $\|\eta(z) - b(\theta)\| = O_P(1/v_T)$ and $v_T = \sqrt{T}$ in this example. To demonstrate this point we generate 1000 'observed' artificial data sets (of sample sizes T=500 and T=1000) and for each data set we run Algorithm 1 for each of the three alternative values of ε_T. For a given sample, and a given tolerance, we produce the approximate Bayesian computation posterior in the manner described above and compute the 95% credible intervals for θ_1 and θ_2. The average length (Width) and the Monte Carlo coverage rate (Cov.), across the 1000 replications, is then recorded in Table 2 for each scenario. It is clear that the average length of the credible regions is larger, and the Monte Carlo coverage further from the nominal value of 95%, the further is the tolerance from the value required to produce asymptotic Gaussianity, namely $\varepsilon_T = 1/T^{0.55}$, providing numerical support for the theoretical results.

Table 2. *The tolerances are* $\varepsilon_1 = 1/T^{0.4}$, $\varepsilon_2 = 1/T^{0.5}$ *and* $\varepsilon_3 = 1/T^{0.55}$.

	Width			Cov		.
T=500	ε_1	ε_2	ε_3	ε_1	ε_2	ε_3
θ_1	0.2602	0.2294	0.2198	96.30	95.60	95.60
θ_2	0.3212	0.3108	0.3086	98.30	97.00	96.00
T=1000						
θ_1	0.1823	0.1573	0.1484	96.80	96.20	95.50
θ_2	0.2366	0.2244	0.2219	96.60	94.30	94.00

B·3. *Theorem 3*

The key result of Theorem 3 is that even if $\Pi_\varepsilon(\cdot|\eta_0)$ is not asymptotically Gaussian, the posterior mean associated with Algorithm 1, $\hat{\theta} = E_{\Pi_\varepsilon}(\theta)$, can still be asymptotically Gaussian, and asymptotically unbiased so long as $\lim_T v_T\varepsilon_T^2 = 0$. However, as proven in Theorem 2, the corresponding confidence regions and uncertainty measures built from $\Pi_\varepsilon(\cdot|\eta_0)$ will only be an adequate reflection on the actual uncertainty associated with $\hat{\theta}$ if $\varepsilon_T = o(1/v_T)$.

In this section we once again generate 1000 'observed' data sets of a given sample size ($T = 500$ and $T = 1000$) according to equation (B1) and $\theta_0 = (0.6, 0.2)^\intercal$, and produce 1000 posteriors based on $\varepsilon_T \in \{1/T^{0.4}, 1/T^{0.5}, 1/T^{0.55}\}$ in the manner described above. For each of the three values of ε_T, and for a sample size of $T = 500$, we record the posterior mean across the 1000 replications and plot the relevant empirical densities in Figure 5. (Figure 6 contains the results for $T = 1000$).

From Figure 5, we see that the (standardised) Monte Carlo sampling distribution of $\hat{\theta} = E_{\Pi_\varepsilon}[\theta]$, over the 1000 replications, and for each of the three values of ε_T, is approximately Gaussian for both parameters and centered at zero. This accords with the theoretical results, which only require that $\lim_T \varepsilon_T = 0$, for asymptotic Gaussianity, and $\lim_T v_T\varepsilon_T^2 = 0$, for zero asymptotic bias, a condition that is satisfied for each of the three tolerance values. This result is also in evidence for $T = 1000$, as can be seen in Figure 6. Lastly, for both $T = 500$ and $T = 1000$, the p-value associated with the KS-test of the null hypothesis that the scaled and centered sampling distribution of the posterior means is standard normal are larger than 5% for both θ_1 and θ_2 and across the three tolerance levels.

Figure 5. Comparison of different tolerance rules for ε_T: $\varepsilon_T = 1/T^{0.4}$ (- - -); $\varepsilon_T = 1/T^{0.5}$ (– · –); $\varepsilon_T = 1/T^{0.55}$ (—); The sample size is $T = 500$.

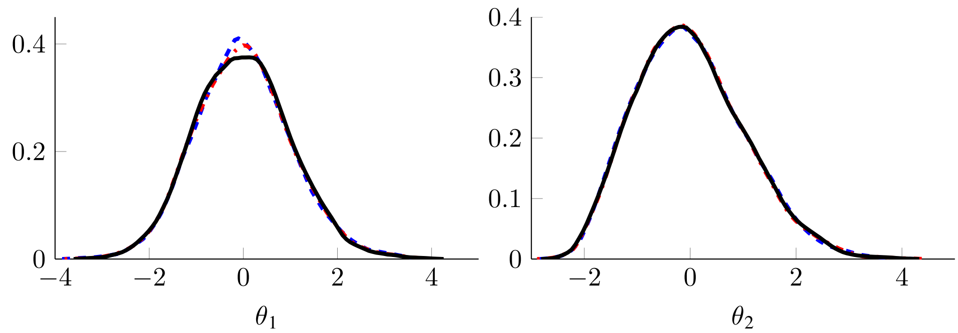


Figure 6. Same information as Figure 5 but for $T = 1000$.