(2018), *pp.* 1–31

Asymptotic Properties of Approximate Bayesian Computation

BY D. T. FRAZIER, G. M. MARTIN

Department of Econometrics and Business Statistics, Monash University, Scenic Boulevard, Clayton Victoria 3800, Australia

david.frazier,gael.martin@monash.edu

C. P. ROBERT

Université Paris Dauphine, PSL Research University, 75775 Paris cedex 16, France

xian@ceremade.dauphine.fr

J. ROUSSEAU

Statistics Department, University of Oxford, 24-29 St Giles, OX1 3LB, U.K.

judith.rousseau@stats.ox.ac.uk

SUMMARY

Approximate Bayesian computation allows for statistical analysis in models with intractable likelihoods. In this paper we consider the asymptotic behaviour of the posterior distribution obtained by this method. We give general results on the rate at which the posterior distribution concentrates on sets containing the true parameter, its limiting shape, and the asymptotic distribution of the 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. Implications for practitioners are discussed.

Some key words: Approximate Bayesian computation; Asymptotics; Bernstein–von Mises theorem; Likelihood-free method; Posterior concentration.

1. INTRODUCTION

Interest in approximate Bayesian computation methods has begun to shift from its initial focus as a computational tool toward its validation as a statistical inference procedure; see, e.g., Fearnhead and Prangle (2012), Marin et al. (2014), Creel and Kristensen (2015), Drovandi et al. (2015), Creel et al. (arxiv:1512.07385), Martin et al. (arxiv:1604.07949) and Li and Fearnhead (2018a,b). Hereafter we denote these preprints by Creel et al. (2015), and Martin et al. (2016).

We study large sample properties of 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 posterior shape crucially depends on the interplay between the rate at which the summaries converge and the rate at which the tolerance used to select parameters shrinks to zero. Bayesian consistency places a less stringent condition on the speed with which the tolerance declines to zero than does asymptotic normality of the posterior distribution. Further, and in contrast to textbook Bernstein–von Mises results, asymptotic normality of the posterior mean does not require asymptotic normality of the posterior distribution, the former being attainable under weaker conditions on the tolerance than required for the latter. Validity of these results requires that the summaries converge toward a well-defined limit 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, see, e.g., Gouriéroux et al. (1993).

We focus on three aspects of asymptotic behaviour: posterior consistency, limiting posterior shape, and the asymptotic distribution of the posterior mean. Our 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) and Li and Fearnhead (2018a). Our approach allows both weaker conditions and a complete characterization of the limiting posterior shape. We distinguish between the conditions, on both the summaries and the tolerance, required for concentration and those required for distributional results. These results suggest how the tolerance in approximate Bayesian computation should be chosen to ensure posterior concentration, valid coverage levels for credible sets, and asymptotically normal and efficient point estimators.

2. PRELIMINARIES AND BACKGROUND

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

(1) Simulate θ^i $(i = 1, \dots, N)$ from $\pi(\theta)$,
(2) Simulate $z^i = (z^i_1, \dots, z^i_T)^\intercal$ $(i = 1, \dots, N)$ from the likelihood, $p(\cdot \mid \theta^i)$,
(3) Select θ^i such that $d\{\eta(y), \eta(z^i)\} \leq \varepsilon$, where $\eta(\cdot)$ is a statistic, $d(\cdot, \cdot)$ is a distance function, and $\varepsilon > 0$ is the tolerance level.

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

$$\pi_\varepsilon\{\theta, z \mid \eta(y)\} = \pi(\theta)p(z \mid \theta)\mathbb{1}_\varepsilon(z) \Big/ \int \pi(\theta)p(z \mid \theta)\mathbb{1}_\varepsilon(z)dzd\theta,$$

where $\mathbb{1}_\varepsilon(z) = \mathbb{1}[d\{\eta(y), \eta(z)\} \leq \varepsilon] = 1$ if $d\{\eta(y), \eta(z)\} \leq \varepsilon$, and zero otherwise. The approximate Bayesian computation posterior density is defined as

$$\pi_\varepsilon\{\theta \mid \eta(y)\} = \textstyle\int \pi_\varepsilon\{\theta, z \mid \eta(y)\}dz.$$

Below, we refer to $\pi_\varepsilon\{\theta \mid \eta(y)\}$ as the approximate posterior density. Likewise, the posterior probability of a set $A \subset \Theta$ associated with Algorithm 1 is

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

and we refer to $\Pi_\varepsilon\{\cdot \mid \eta(y)\}$ as the approximate posterior distribution. When $\eta(\cdot)$ is sufficient for the observed data y and ε is close to zero, $\pi_\varepsilon\{\theta \mid \eta(y)\}$ will be a good approximation to $\pi(\theta \mid y)$, and draws of θ from $\pi_\varepsilon\{\theta \mid \eta(y)\}$ can be used to estimate features of $\pi(\theta \mid y)$.

In practice $\eta(y)$ is rarely sufficient for y, and draws of θ can only be used to approximate $\pi\{\theta \mid \eta(y)\} = \lim_{\varepsilon \to 0} \pi_\varepsilon\{\theta \mid \eta(y)\}$. Given the general lack of sufficient statistics, we need to assess the behavior of the approximate posterior distribution $\Pi_\varepsilon\{\cdot \mid \eta(y)\}$, and to establish whether or not $\Pi_\varepsilon\{\cdot \mid \eta(y)\}$ behaves in a manner that is appropriate for statistical inference, with asymptotic theory being one obvious approach.

Establishing the large sample behavior of $\Pi_\varepsilon\{\cdot \mid \eta(y)\}$, including point and interval estimates derived from this distribution, gives practitioners guarantees on the reliability of approximate Bayesian computations. Furthermore, these results allow us to provide guidelines for choosing the tolerance ε so that $\Pi_\varepsilon\{\cdot \mid \eta(y)\}$ possesses desirable statistical properties.

Before presenting our results, we set notation used throughout the paper. Let $\mathcal{B} \subset \mathbb{R}^{k_\eta}$ denote the range of the simulated summaries $\eta(z)$. Let $d_1(\cdot,\cdot)$ be a metric on Θ and $d_2(\cdot,\cdot)$ a metric on $\mathcal{B}$. Take $\|\cdot\|$ to be the Euclidean norm. Throughout, C denotes a generic positive constant. 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 T large, $a_T \asymp b_T$ implies that $a_T \lesssim b_T \lesssim a_T$, 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 $C > 0$ there exists a finite $M > 0$ and a finite T such that $\text{pr}(|x_t/a_t| \geq M) \leq C$, for all $t > T$. All limits are taken as $T \to \infty$. When no confusion will result, $\lim_T$ replaces $\lim_{T\to\infty}$.

3. Concentration of the Approximate Bayesian Computation Posterior

We assume throughout that the model is correctly specified: for some θ_0 in the interior of Θ, we have $P_\theta = P_0$. Asymptotic validity of any Bayesian procedure requires posterior concentration, which is often referred to as Bayesian consistency. In our context, this equates to the following posterior concentration property: for any $\delta > 0$, and for some $\varepsilon > 0$,

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

This property is paramount since, for any $A \subset \Theta$, $\Pi\left[A \mid d_2\{\eta(y),\eta(z)\} \leq \varepsilon\right]$ will differ from the exact posterior probability. Without the guarantees of exact posterior inference, knowing that $\Pi_\varepsilon\{\cdot \mid \eta(y)\}$ will concentrate on θ_0 gives validity to its use as a means of expressing our uncertainty about θ.

Posterior concentration is related to the rate at which information about θ_0 accumulates in the sample. The amount of information Algorithm 1 provides depends on the rate at which the observed summaries $\eta(y)$ and the simulated summaries $\eta(z)$ converge to well-defined limit counterparts $b(\theta_0)$ and $b(\theta)$, and 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. These assumptions are applicable to a broad range of data structures, including weakly dependent data.

Assumption 1. There exist a non-random map $b : \Theta \to \mathcal{B}$, and a sequence of functions $\rho_T(u)$ that are monotone non-increasing in u for any T and satisfy $\rho_T(u) \to 0$ as $T \to \infty$. For fixed u, and for all $\theta \in \Theta$,

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

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

Assumption 2. There exists some $D > 0$ such that, for all $\xi > 0$ and some $C > 0$, the prior probability satisfies $\Pi\left[d_2\{b(\theta), b(\theta_0)\} \leq \xi\right] \geq C\xi^D$.

Assumption 3. (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. The convergence of $\eta(z)$ to $b(\theta)$ in Assumption 1 is the key to posterior concentration and without it, or a similar assumption, Bayesian consistency will not occur. The function $\rho_T(u)$ in Assumption 1 typically takes the form $\rho_T(u) = \rho(uv_T)$, for v_T a sequence such that $d_2\{\eta(z), b(\theta)\} = O_P(1/v_T)$, and where $\rho(uv_T)$ controls the tail behavior of $d_2\{\eta(z), b(\theta)\}$. The specific structure of $\rho(uv_T)$ will depend on what is assumed about the properties of the underlying summaries $\eta(z)$. In most cases, $\rho(uv_T)$ will have either a polynomial or exponential structure in uv_T, and thus satisfy one of the following rates.

(a) Polynomial: there exist a diverging positive sequence $\{v_T\}_{T\geq 1}$ and $u_0, \kappa > 0$ such that

$$P_\theta\left[d_2\{\eta(z), b(\theta)\} > u\right] \leq c(\theta)\rho_T(u), \quad \rho_T(u) = 1/(uv_T)^\kappa, \quad u \leq u_0, \tag{1}$$

where, for some $c_0 > 0$ and $\delta > 0$, $\int_\Theta c(\theta)d\Pi(\theta) < \infty$ and if $d_2\{b(\theta), b(\theta_0)\} \leq \delta$, then $c(\theta) \leq c_0$.
(b) Exponential: there exist $h_\theta(\cdot) > 0$ and $u_0 > 0$ such that

$$P_\theta\left[d_2\{\eta(z), b(\theta)\} > u\right] \leq c(\theta)\rho_T(u), \quad \rho_T(u) = \exp\{-h_\theta(uv_T)\}, \quad u \leq u_0, \tag{2}$$

where, for some $c, C > 0$, $\int_\Theta c(\theta)\exp\{-h_\theta(uv_T)\}d\Pi(\theta) \leq C\exp\{-c(uv_T)^\tau\}$.

To illustrate these cases for $\rho_T(\cdot)$, consider the summary statistics $\eta(z) = T^{-1}\sum_{i=1}^T g(z_i)$ where, for simplicity, $\{g(z_i)\}_{i\leq T}$ is independent and identically distributed, and $b(\theta) = E_\theta\{g(Z)\}$.

If $g(z_i) - b(\theta)$ has a finite moment of order κ, $\rho_T(u)$ will satisfy (1): from Markov's inequality,

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

With reference to (1), $\rho_T(u) = 1/(uv_T)^\kappa$, $v_T = T^{1/2}$ and $c(\theta) = C\mathbb{E}_\theta\{|g(Z)|^\kappa\} < \infty$. If the map $\theta \mapsto E_\theta\left\{|g(Z)|^\kappa\right\}$ is continuous at θ_0 and positive, Assumption 1 is satisfied.

If $\{g(z_i) - b(\theta)\}$ has a finite exponential moment, $\rho_T(u)$ will satisfy (2): from a version of the Bernstein inequality,

$$P_\theta\left\{\|\eta(z) - b(\theta)\| > u\right\} \leq \exp\left[-u^2T/\{2c(\theta)\}\right].$$

With reference to (2), $\rho_T(u) = \exp\{-h_\theta(uv_T)\}$, $h_\theta(uv_T) = u^2v_T^2/\{2c(\theta)\}$ and $v_T = T^{1/2}$. If the map $\theta \mapsto c(\theta)$ is continuous at θ_0 and positive, Assumption 1 is satisfied.

Remark 2. Assumption 2 controls the degree of prior mass in a neighbourhood of θ_0 and is standard in Bayesian asymptotics. For ξ small, the larger D, the smaller the amount of prior mass near θ_0. If the prior measure $\Pi(\theta)$ is absolutely continuous with prior density $\pi(\theta)$ and if π is bounded, above and below, near θ_0, then $D = \dim(\theta) = k_\theta$. Assumption 3 is an identification condition that is critical for obtaining posterior concentration around θ_0. Injectivity of b depends on both the true structural model and the particular choice of η. Without this identification condition posterior concentration at θ_0 cannot occur.

THEOREM 1. *If Assumptions* 1–2 *are satisfied, then, for M large enough, as $T \to \infty$ and $\varepsilon_T = o(1)$, with P_0 probability going to one,*

$$\Pi\left[d_2\{b(\theta), b(\theta_0)\} > \lambda_T \mid d_2\{\eta(y), \eta(z)\} \leq \varepsilon_T\right] \lesssim 1/M, \tag{3}$$

with $\lambda_T = 4\varepsilon_T/3 + \rho_T^{-1}(\varepsilon_T^D/M)$. Moreover, if Assumption 3 *also holds, as $T \to \infty$,*

$$\Pi\left[d_1(\theta, \theta_0) > L\lambda_T^\alpha \mid d_2\{\eta(y), \eta(z)\} \leq \varepsilon_T\right] \lesssim 1/M, \tag{4}$$

Since equations (3) and (4) hold for any M large enough, we can conclude that the posterior distribution behaves like an $o_P(1)$ random variable on sets that do not include θ_0, and Bayesian consistency of $\Pi_\varepsilon\{\cdot \mid \eta(y)\}$ follows. More generally, (3) and (4) give a posterior concentration rate, denoted by λ_T in Theorem 1, that depends on ε_T and on the underlying behavior of $\eta(z)$, as described by $\rho_T(u)$. We must consider the nature of this concentration rate in order to understand which choices for ε_T are appropriate under different assumptions on the summary statistics.

As mentioned above, the deviation control function $\rho_T(u)$ will often be of a polynomial (1) or exponential (2) form. Under these two assumptions, $\rho_T(u)$ has an explicit representation and the concentration rate λ_T can be obtained by solving the equation

$$\lambda_T = 4\varepsilon_T/3 + \rho_T^{-1}(\varepsilon_T^D/M).$$

(a) polynomial case: From equation (1), the deviation control function is $\rho_T(u) = 1/(uv_T)^\kappa$. To obtain the posterior concentration rate, we invert $\rho_T(u)$ to obtain $\rho_T^{-1}(\varepsilon_T^D) = 1/(\varepsilon_T^{D/\kappa}v_T)$, and then equate ε_T and $\rho_T^{-1}(\varepsilon_T^D)$, to obtain $\varepsilon_T \asymp v_T^{-\kappa/(\kappa+D)}$. This choice of ε_T implies concentration of the approximate posterior distribution at the rate

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

(b) exponential case: If the summary statistics admit an exponential moment, a faster rate of posterior concentration obtains. From equation (2), $\rho_T(u) = \exp\{-h_\theta(uv_T)\}$ and there exist finite $u_0, c, C > 0$ such that

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

Hence if $c(\theta)$ is bounded from above and if $h_\theta(u) \geq u^\tau$ for θ in a neighbourhood of θ_0, then $\rho_T(u) \asymp \exp\{-c_0(uv_T)^\tau\}$; thus, $\rho_T^{-1}(\varepsilon_T^D) \asymp (-\log \varepsilon_T)^{1/\tau}/v_T$. Following arguments similar to those used in (a) immediately above, if we take $\varepsilon_T \asymp (\log v_T)^{1/\tau}/v_T$, the approximate posterior distribution concentrates at the rate

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

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

$$y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} \ \ (t = 1, \ldots, T),$$

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,

$$-2 \leq \theta_1 \leq 2, \ \theta_1 + \theta_2 \geq -1, \theta_1 - \theta_2 \leq 1. \tag{5}$$

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 above, it follows that $V(\theta) = E[\{\eta(z) - b(\theta)\}\{\eta(z) - b(\theta)\}^\intercal]$ satisfies tr$\{V(\theta)\} < \infty$ for all θ in (5). By an application of Markov's inequality, we can conclude that

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

where the $o(1/T)$ term comes from the fact that there are finitely many non-zero covariance terms due to the m-dependence of the series, and Assumption 1 is satisfied. Given the structure of $b(\theta)$, the uniform prior $\pi(\theta)$ over (5) fulfills Assumption 2. Furthermore, $\theta \mapsto b(\theta) = (1 + \theta_1^2 + \theta_2^2, (1 + \theta_2)\theta_1, \theta_2)^\intercal$ is injective and satisfies Assumption 3. As noted in Remark 2, the injectivity of $\theta \mapsto b(\theta)$ is required for posterior concentration, and without it there is no guarantee that the posterior will concentrate on θ_0. Since the sufficient conditions for Theorem 1 are satisfied, approximate Bayesian computation based on this choice of statistics will yield an approximate posterior density that concentrates on θ_0.

Theorem 1 can also be visualized 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 approximate posterior density will become increasingly peaked at $\tilde{\theta}$ as T increases. Using Example 1, we demonstrate this behavior in the Supplementary Material.

4. Shape of the Asymptotic Posterior Distribution

4·1. *Assumptions and Theorem*

While posterior concentration states that $\Pi\left[d_1(\theta, \theta_0) > \delta \mid d_2\{\eta(y), \eta(z)\} \leq \varepsilon_T\right] = o_P(1)$ for an appropriate choice of ε_T, it does not indicate precisely how this mass accumulates, or the approximate amount of posterior probability 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 that credible regions have proper frequentist coverage. To this end, we now analyse the limiting shape of $\Pi\left[\cdot \mid d_2\{\eta(y), \eta(z)\} \leq \varepsilon_T\right]$ for various relationships between ε_T and the rate at which summary statistics satisfy a central limit theorem. In this and the following sections, we denote $\Pi\left[\cdot \mid d_2\{\eta(y), \eta(z)\} \leq \varepsilon_T\right]$ by $\Pi_\varepsilon\{\cdot \mid \eta(y)\}$. Let $\|\cdot\|_*$ denote the spectral norm.

In addition to Assumption 2 the following conditions are needed to establish the results of this section.

Assumption 4. Assumption 1 holds. There exists a sequence of positive definite matrices $\{\Sigma_T(\theta_0)\}_{T\geq 1}$, $c_0 > 0$, $\kappa > 1$ and $\delta > 0$ such that for all $\|\theta - \theta_0\| \leq \delta$, $P_\theta\left[\|\Sigma_T(\theta_0)\{\eta(z) - b(\theta)\}\| > u\right] \leq c_0 u^{-\kappa}$ for all $0 < u \leq \delta\|\Sigma_T(\theta_0)\|_*$, uniformly in T.

Assumption 5. Assumption 3 holds. The map $\theta \mapsto b(\theta)$ is continuously differentiable at θ_0 and the Jacobian $\nabla_\theta b(\theta_0)$ has full column rank k_θ.

Assumption 6. The value θ_0 is in the interior of Θ. 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 $\{\Sigma_T(\theta)\}_{T\geq 1}$, with $k_\eta = \dim\{\eta(z)\}$, such that for all open sets B

$$\sup_{|\theta-\theta_0|\leq\delta} \left|P_\theta\left[\Sigma_T(\theta)\{\eta(z) - b(\theta)\} \in B\right] - P\left\{\mathcal{N}(0, I_{k_\eta}) \in B\right\}\right| \to 0$$

in distribution as $T \to \infty$, where I_{k_η} is the $(k_\eta \times k_\eta)$ identity matrix.

Assumption 7. 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.

Assumption 8. For some positive δ, all $\|\theta - \theta_0\| \leq \delta$, all ellipsoids $B_T = \left\{(t_1, \ldots, t_{k_\eta}) : \sum_{j=1}^{k_\eta} t_j^2/h_T^2 \leq 1\right\}$ and all $u \in \mathbb{R}^{k_\eta}$ fixed, for all $h_T \to 0$, as $T \to \infty$,

$$\lim_T \sup_{|\theta-\theta_0|\leq\delta} \left|h_T^{-k_\eta} P_\theta\left[\Sigma_T(\theta)\{\eta(z) - b(\theta)\} - u \in B_T\right] - \varphi_{k_\eta}(u)\right| = 0,$$

$$h_T^{-k_\eta}\, P_\theta\left[\Sigma_T(\theta)\{\eta(z) - b(\theta)\} - u \in B_T\right] \leq H(u), \qquad \int H(u)du < \infty,$$

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

Remark 3. Assumption 4 is similar to Assumption 1 but for $\Sigma_T(\theta_0)\{\eta(z) - b(\theta)\}$. Assumption 6 is 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 independent and identically distributed data with $\eta(z) = T^{-1}\sum_{i=1}^T g(z_i)$, $\Sigma_T(\theta) = v_T A_T(\theta)$ 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 5 and 8 ensure that $\theta \mapsto b(\theta)$ and the 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 that of $\Sigma_T(\theta_0)\{b(\theta) - b(\theta_0)\}$. Assumption 8 governs the pointwise convergence of a normalized version of the measure P_θ, therein dominated by $H(u)$, and allows the application of the dominated convergence theorem in Case (iii) of the following result.

THEOREM 2. *Under Assumptions 2, 4–7, with $\kappa > k_\theta$, the following hold with probability going to 1. (i) If $\lim_T v_T\varepsilon_T = \infty$, the posterior distribution of $\varepsilon_T^{-1}(\theta - \theta_0)$ converges to the uniform distribution over the ellipsoid $\{w : w^\intercal B_0 w \leq 1\}$ with $B_0 = \nabla_\theta b(\theta_0)^\intercal \nabla_\theta b(\theta_0)$, meaning that for $f(\cdot)$ continuous and bounded,*

$$\int f\{\varepsilon_T^{-1}(\theta - \theta_0)\} d\Pi_\varepsilon\{\theta \mid \eta(y)\} \to \int_{u^\intercal B_0 u\leq 1} f(u)du \Big/ \int_{u^\intercal B_0 u\leq 1} du, \quad T \to \infty.$$

(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

$$\Pi_\varepsilon\left[\Sigma_T(\theta_0)\nabla_\theta b(\theta_0)(\theta - \theta_0) - \Sigma_T(\theta_0)\{\eta(y) - b(\theta_0)\} \in B \mid \eta(y)\right] \to Q_c(B), \quad T \to \infty.$$

In particular, $Q_c(B) \propto \int_B \int_{\mathbb{R}^{k_\eta}} \mathbb{1}\{(z - x)^\intercal A(\theta_0)^\intercal A(\theta_0)(z - x) \leq c\}\varphi_{k_\eta}(z) dz dx$.
(iii) If $\lim_T v_T\varepsilon_T = 0$ and Assumption 8 holds then,

$$\Pi_\varepsilon\left[\Sigma_T(\theta_0)\nabla_\theta b(\theta_0)(\theta - \theta_0) - \Sigma_T(\theta_0)\{\eta(y) - b(\theta_0)\} \in B \mid \eta(y)\right] \to \int_B \varphi_{k_\eta}(x)dx, \quad T \to \infty.$$

Remark 4. Theorem 2 generalizes 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 the Supplementary Material. Furthermore, as with Theorem 1, the behavior of $\Pi_\varepsilon\{\cdot \mid \eta(y)\}$ described by Theorem 2 can be visualized. This is demonstrated in the Supplementary Material. Formal verification of the conditions underpinning Theorem 2 is quite challenging, even in this case. Numerical results nevertheless suggest that for this particular choice of model and summaries a Bernstein–von Mises result holds, conditional on $\varepsilon_T = o(1/v_T)$, with $v_T = T^{1/2}$.

4·2. *Discussion of the Result*

Theorem 2 asserts that the crucial feature in determining the limiting shape of $\Pi_\varepsilon\{\cdot \mid \eta(y)\}$ is the behaviour of $v_T\varepsilon_T$. The implication of Theorem 2 is that only in the regime where $\lim_T v_T\varepsilon_T = 0$ will $100(1-\alpha)\%$ Bayesian credible regions calculated from $\Pi_\varepsilon\{\cdot \mid \eta(y)\}$ have frequentist coverage of $100(1-\alpha)\%$. If $\lim_T v_T\varepsilon_T = c > 0$, for c finite, $\Pi_\varepsilon\{\cdot \mid \eta(y)\}$ is not asymptotically Gaussian and credible regions will have incorrect magnitude, i.e., the coverage will not be at the nominal level. If $\lim_T v_T\varepsilon_T = \infty$, i.e., $\varepsilon_T \gg v_T^{-1}$, credible regions will have coverage that converges to 100%.

In Case (i), which corresponds to a large tolerance ε_T, the approximate posterior distribution has nonstandard asymptotic behaviour. In Case (i) $\Pi_\varepsilon\{\cdot \mid \eta(y)\}$ behaves like the prior distribution over $\{\theta : \|\nabla_\theta b(\theta_0)(\theta-\theta_0)\| \leq \varepsilon_T\{1+o_P(1)\}\}$, which, by prior continuity, implies that $\Pi_\varepsilon\{\cdot \mid \eta(y)\}$ is equivalent to a uniform distribution over this set. Li and Fearnhead (2018a) also establish this behaviour, and observe that asymptotically $\Pi_\varepsilon\{\cdot \mid \eta(y)\}$ behaves like a convolution of a Gaussian distribution, with variance of order $1/v_T^2$, and a uniform distribution over a ball of radius ε_T, and, where, depending on the order $v_T\varepsilon_T$, one distribution will dominate.

Assumption 8 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 8 must be adapted for Theorem 2 to be satisfied. One such adaptation is

Assumption 9. There exist $\delta > 0$ and a countable set E_T such that for all $\|\theta-\theta_0\| < \delta$, for all $x \in E_T$ such that $\operatorname{pr}\{\eta(z) = x\} > 0$, $\operatorname{pr}\{\eta(z) \in E_T\} = 1$ and

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

This is satisfied 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 satisfies Assumptions 2 and 4–7. Thus the conclusions of both Theorems 1 and 2 apply to this model.

5. ASYMPTOTIC DISTRIBUTION OF THE POSTERIOR MEAN

5·1. *Main Result*

The literature on the asymptotics of approximate Bayesian computation has so far focused primarily on asymptotic normality of the posterior mean. The posterior normality result in Theorem 2 is not weaker, or stronger, than the asymptotic normality of an approximate point estimator, as the results focus on different objects. However, existing proofs for asymptotic normality of the posterior mean all require asymptotic normality of the posterior distribution $\Pi_\varepsilon\{\cdot \mid \eta(y)\}$. We demonstrate that this is not a necessary condition.

For clarity we focus on the case of a scalar parameter θ and scalar summary $\eta(y)$, i.e., $k_\theta = k_\eta = 1$, but present an extension to the multivariate case in Section 5·2. This result requires a further assumption on the prior in addition to Assumption 2.

Assumption 10. The prior density $\pi(\theta)$ is such that (i) for θ_0 in the interior of Θ, $\pi(\theta_0) > 0$; (ii) the density function $\pi(\theta)$ is β- Hölder in a neighbourhood of θ_0: there exist $\delta, L > 0$ such that for all $|\theta -$

$\theta_0| \leq \delta$, and $\nabla_\theta^{(j)}\pi(\theta_0)$ the (j)-th derivative of $\pi(\theta_0)$,

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

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

THEOREM 3. *Let Assumptions* 2, 4–7*, with* $\kappa > \beta + 1$*, and* 10 *be satisfied. Furthermore, let* $\theta \mapsto b(\theta)$ *be* β*-Hölder in a neighbourhood of* θ_0*. Denoting* $E_{\Pi_\varepsilon}(\theta)$ *as the posterior mean of* θ*, the following characterisation holds with probability going to one:*
(i) If $\lim_T v_T\varepsilon_T = \infty$ *and* $v_T\varepsilon_T^{2\wedge(1+\beta)} = o(1)$*, then*

$$E_{\Pi_\varepsilon}\{v_T(\theta-\theta_0)\} \to \mathcal{N}[0, V(\theta_0)/\{\nabla_\theta b(\theta_0)\}^2], \tag{6}$$

in distribution as $T\to\infty$*, where* $V(\theta_0) = \lim_T \text{var}[v_T\{\eta(y) - b(\theta_0)\}]$*.*
(ii) If $\lim_T v_T\varepsilon_T = c \geq 0$*, and if when* $c = 0$ *Assumption* 8 *holds, then* (6) *also holds.*

There are two immediate consequences of Theorem 3: first, part (i) of Theorem 3 states that if one is only interested in obtaining accurate point estimators for θ_0, all we require is a tolerance ε_T satisfying $v_T\varepsilon_T^2 = o(1)$, which can significantly reduce the computational burden of approximate Bayesian computation; secondly, if one wants accurate point estimators of θ_0 and accurate expressions of the uncertainty associated with this point estimate, we require $\varepsilon_T = o(1/v_T)$. The first statement follows directly from part (i) of Theorem 3, while the second statement follows from part (ii) of Theorem 3 and recalling that, from Theorem 2, credible regions constructed from $\Pi_\varepsilon\{\cdot \mid \eta(y)\}$ will have proper frequentist coverage only if $\varepsilon_T = o(1/v_T)$. For $\varepsilon_T \asymp v_T^{-1}$ or $\varepsilon_T \gg v_T^{-1}$, the frequentist coverage of credible balls centered at $E_{\Pi_\varepsilon}(\theta)$ will not be equal to the nominal level.

As an intermediate step in the proof of Theorem 3, we demonstrate the following expansion for the posterior mean, where k denotes the integer part of $(\beta+1)/2$:

$$E_{\Pi_\varepsilon}(\theta-\theta_0) = \frac{\eta(y)-b(\theta_0)}{\nabla_\theta b(\theta_0)} + \sum_{j=1}^{\lfloor\beta\rfloor}\frac{\nabla_b^{(j)}b^{-1}(b_0)}{j!}\sum_{l=\lceil j/2\rceil}^{\lfloor(j+k)/2\rfloor}\frac{\varepsilon_T^{2l}\nabla_b^{(2l-j)}\pi(b_0)}{\pi(b_0)(2l-j)!} + O(\varepsilon_T^{1+\beta}) + o_P(1/v_T). \tag{7}$$

This 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 (7) 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

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

5·2. *Comparison with Existing Results*

Li and Fearnhead (2018a) analyse the asymptotic properties of the posterior mean and functions thereof. 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 (2018a) demonstrate asymptotic normality, with no bias, of the posterior mean if $\varepsilon_T = o(1/v_T^{3/5})$. Heuristically, the authors derive this result using an approximation of the posterior density $\pi_\varepsilon\{\theta \mid \eta(y)\}$, 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 (2018a) 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 8 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 8 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 (2018a). Our results also allow for an explicit representation of the bias that obtains for the posterior mean when $\liminf_T v_T\varepsilon_T^2 > 0$.

In further contrast to Li and Fearnhead (2018a), Theorem 2 completely characterizes the asymptotic behavior of the approximate posterior distribution 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 $\varepsilon_T = o(1/v_T^{1/2})$.

Li and Fearnhead (2018a) provide the interesting result that if $k_\eta > k_\theta \geq 1$ and $\varepsilon_T = o(1/v_T^{3/5})$, 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 asymptotic expansion for the posterior mean when $k_\eta > k_\theta \geq 1$.

THEOREM 4. *Let Assumptions* 2, 4–7 *and* 10 *be satisfied. Assume that* $v_T\varepsilon_T \to \infty$ *and* $v_T\varepsilon_T^2 = o(1)$. *Assume also that* $b(.)$ *and* $\pi(.)$ *are Lipschitz in a neighbourhood of* θ_0. *Then, for* $k_\eta > k_\theta \geq 1$,

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

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*

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

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* $\eta(y)$.

A consequence of Theorem 4 is that, for a fixed choice of summaries, the two-stage procedure advocated by Fearnhead and Prangle (2012) will not reduce the asymptotic variance over a point estimate produced via Algorithm 1. However, this two-stage procedure does reduce the Monte Carlo error inherent in estimating the approximate posterior distribution $\Pi_\varepsilon\{\cdot \mid \eta(y)\}$ by reducing the dimension of the statistics on which the matching in approximate Bayesian computation is based.

6. PRACTICAL IMPLICATIONS OF THE RESULTS

6·1. *General*

The approximate Bayesian computation approach in Algorithm 1 is typically not applied in practice. Instead, the acceptance step in Algorithm 1 is commonly replaced by the nearest-neighbour selection step and with $d_2\{\eta(z), \eta(y)\} = \|\eta(z) - \eta(y)\|$, see, e.g., Biau et al. (2015):

(3′) select all θ^i associated with the $\alpha = \delta/N$ smallest distances $\|\eta(z) - \eta(y)\|$ for some δ.

This nearest-neighbour version accepts draws of θ associated with an empirical quantile over the simulated distances $\|\eta(z) - \eta(y)\|$ and defines the acceptance probability for Algorithm 1. A key practical insight of our asymptotic results is that the acceptance probability, $\alpha_T = \text{pr}\{\|\eta(z) - \eta(y)\| \leq \varepsilon_T\}$, is only affected by the dimension of θ, as formalized in Corollary 1.

COROLLARY 1. *Under the conditions in Theorem* 2*:*
(i) If $\varepsilon_T \asymp v_T^{-1}$ *or* $\varepsilon_T = o(1/v_T)$, *then the acceptance rate associated with the threshold* ε_T *is*

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

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

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

This shows that choosing a tolerance $\varepsilon_T = o(1)$ is equivalent to choosing an $\alpha_T = o(1)$ quantile of $\|\eta(z) - \eta(y)\|$. 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 $\varepsilon_T = o(1/v_T)$, as required for the Bernstein–von Mises result in Theorem 2, the associated acceptance probability goes to zero at the faster rate, $\alpha_T = o(1/v_T^{k_\theta})$. 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 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 an approximate posterior distribution that concentrates; under the more stringent condition $\alpha_T = o(1/v_T^{k_\theta})$ the approximate posterior distribution both concentrates and is approximately Gaussian in large samples. These results give practitioners an understanding of what to expect from this procedure, and a means of detecting potential issues if this expected behaviour is not in evidence. Moreover, given that there is no direct link between $\Pi_\varepsilon\{\cdot \mid \eta(y)\}$ and the exact posterior distribution, these results give some understanding of the statistical properties that $\Pi_\varepsilon\{\cdot \mid \eta(y)\}$ should display when it is obtained from the popular nearest-neighbour version of the algorithm.

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 provides theoretical evidence of a curse-of-dimensionality encountered in these algorithms as the dimension of the parameters increases, with this being the first piece of work, to our knowledge, to link the dimension of θ to certain asymptotic properties for $\Pi_\varepsilon\{\cdot \mid \eta(y)\}$. This result provides theoretical justification for dimension reduction methods that process parameter dimensions individually and independent of the other 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).

While Corollary 1 demonstrates that the order of α_T is unaffected by the dimension of the summaries, α_T cannot 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. This approximation of α_T is a Monte Carlo estimate of a conditional expectation, and, as such, will be sensitive to the dimension of $\eta(\cdot)$ for any fixed number of Monte Carlo draws N; see Biau et al. (2015) for further discussion on this point. In addition, it can also be shown that if ε_T becomes much smaller than $1/v_T$, the dimension of $\eta(\cdot)$ will affect the behavior of Monte Carlo estimators for this acceptance probability. Specifically, when considering inference on θ_0 using the accept/reject approximate Bayesian computation algorithm, we require a sequence of Monte Carlo trials $N_T \to \infty$ as $T \to \infty$ that diverges faster the larger is k_η, the dimension of $\eta(\cdot)$. Such a feature highlights the lack of efficiency of the accept/reject approach when the sample size is large or if the dimension of the summaries is large. However, we note here that more efficient sampling approaches exist and could be applied in these settings. For example, Li and Fearnhead (2018a) consider an importance sampling approach to approximate Bayesian computation that yields acceptance rates satisfying $\alpha_T = O(1)$, so long as $\varepsilon_T = O(1/v_T)$. Therefore, in cases where the Monte Carlo error is likely to be large, these alternative sampling approaches should be employed.

Regardless of whether one uses a more efficient sampling procedure than the simple accept/reject approach, Corollary 1 demonstrates that taking a tolerance sequence as small as possible will not necessarily yield more accurate results. That is, Corollary 1 questions the persistent opinion that the tolerance in Algorithm 1 should always be taken as small as the computing budget allows. Once ε_T is chosen small enough to satisfy Case (iii) of Theorem 2, 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 any closer to zero, especially since pushing ε_T closer to zero can drastically increase the required computational burden. In the following section we numerically demonstrate this result in a simple example. In particular, we demonstrate that for a choice of tolerance ε_T that admits a Bernstein–von Mises result, there is no gain in taking a tolerance that is smaller than this value, while the computational cost associated with such a choice, for a fixed level of Monte Carlo error, drastically increases.

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 $T = 100$. 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 asymptotic normality of the approximate posterior distribution, we must 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, $\alpha_1 = 1/T^{1.1}$, $\alpha_2 = 1/T^{3/2}$, $\alpha_3 = 1/T^2$ and $\alpha_4 = 1/T^{5/2}$.

Draws for (μ, σ) are simulated on $[0.5, 1.5] \times [0.5, 1.5]$ according to independent uniforms $\mathcal{U}[0.5, 1.5]$. The number of draws, N, is chosen so that we retain 250 accepted draws for each of the different choices $(\alpha_1, \ldots, \alpha_4)$. The exact finite sample marginal posterior densities of μ and σ are produced by numerically evaluating the likelihood function, normalizing over the support of the prior and marginalising with respect to each parameter. Given the sufficiency of $(\bar{x},\ s_T^2)$, the exact marginal posteriors densities for μ and σ are equal to those based directly on the summaries themselves. Hence, we are able to assess the impact of the choice of α, per se, on the ability of the nearest-neighbour version of Algorithm 1 to replicate the exact marginal posteriors.

We summarize the accuracy of the resulting approximate posterior density estimates, across the four quantile choices, using root mean squared error. In particular, over fifty simulated replications, and in the case of the parameter μ, we estimate the root mean squared error between the marginal posterior density obtained from Algorithm 1 using α_j, and denoted by $\widehat{\pi}_{\alpha_j}\{\mu \mid \eta(y)\}$, and the exact marginal posterior density, $\pi(\mu \mid y)$, using

$$\text{RMSE}_\mu(\alpha_j) = \left[\frac{1}{G}\sum_{g=1}^{G}\left\{\widehat{\pi}_{\alpha_j}^{g}\{\mu \mid \eta(y)\} - \pi^{g}(\mu \mid y)\right\}^2\right]^{1/2}. \tag{8}$$

The term $\widehat{\pi}_{\alpha_j}^{g}$ is the ordinate of the density estimate from the nearest-neighbour version of Algorithm 1 and π^g is the ordinate of the exact posterior density, at the g-th grid point upon which the density is estimated. $\text{RMSE}_\sigma(\alpha_j)$ is computed analogously. The value of $\text{RMSE}_\mu(\alpha_j)$ is averaged over fifty replications to account for sampling variability. For each replication, we fix $T = 100$ and generate observations using 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 1 plots the resulting marginal posterior estimates and compares these with the exact finite sample marginal posterior densities of μ and σ. At the end of Section 6·1, we argued 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 posterior densities. This implication is evident in Fig. 1: in the case of μ, there is a clear visual decline in the accuracy with which approximate Bayesian computation estimates the exact marginal posterior densities when choosing quantiles smaller than α_2; whilst in the case of σ, the worst performing estimate is that associated with the smallest value of α_j.

The results in Table 1 report average root mean squared error, relative to the average value associated with $\alpha_4 = 1/T^{5/2}$. Values smaller than one indicate that the larger, and less computationally burdensome, value of α_j yields a more accurate estimate than that obtained using α_4. In brief, Table 1 paints a similar picture to that of Fig. 1: for σ, the estimates based on α_j, $j = 1, 2, 3$, are all more accurate than those based on α_4; for μ, estimates based on α_2 and α_3 are both more accurate that those based on α_4.

These numerical results have important implications for implementation of approximate Bayesian computation. In particular, to keep the level of Monte Carlo error constant across the α_j quantile choices, as we have done in this simulation setting via the retention of 250 draws, this requires taking: $N = 210e03$ for α_1, $N = 1.4e06$ for α_2, $N = 13.5e06$ for α_3, and $N = 41.0e06$ for α_4. That is, the computational burden associated with decreasing the quantile in the manner indicated increases dramatically: approxi-

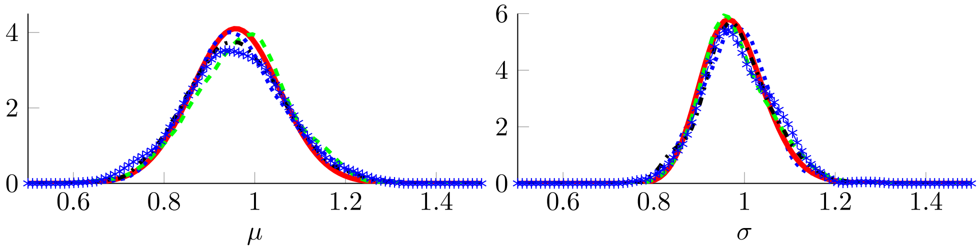

Figure 1. Comparison of exact and approximate posterior densities for various tolerances. Exact marginal posterior densities (—). Approximate Bayesian computation posterior densities based on $\alpha_1 = 1/T^{1.1.}$ ($\cdots$); $\alpha_2 = 1/T^{3/2}$ (- - -); $\alpha_3 = 1/T^2$ (— · —); $\alpha_4 = 1/T^{5/2}$ —∗—

mate posterior densities based on α_4 for example require a value of N that is three orders of magnitude greater than those based on α_1, but this increase in computational burden yields no, or minimal, gain in accuracy. The extension of such explorations to more scenarios is beyond the scope of this paper; however, we speculate that, with due consideration given to the properties of both the true data generating process and the chosen summary statistics and, hence, of the sample sizes for which Theorem 2 has practical content, similar qualitative results will continue to hold.

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

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

AVG-RMSE is the ratio of the average root mean square errors as defined in (8).

ACKNOWLEDGMENT

This research was supported by Australian Research Council Discovery and l'Institut Universitaire de France grants. The third author is further affiliated with the University of Warwick and the fourth with the Université Paris Dauphine. We are grateful to Wentao Li and Paul Fearnhead for spotting a mistake in a previous version of Theorem 3 and to the editorial team for its help.

BIBLIOGRAPHY

BEAUMONT, M. A., CORNUET, J-M., MARIN, J-M., & ROBERT, C. P. (2009). Adaptive Approximate Bayesian Computation. *Biometrika* **96**, 983–990.

BEAUMONT, M. A., ZHANG, W., & BALDING, D. J. (2002). Approximate Bayesian Computation in Population Genetics. *Genetics* **162**, 2025–2035.

BHATTACHARYA, R. N., AND RAO, R. R. (1986). *Normal Approximation and Asymptotic Expansions*. Volume 64 of Classics in Applied Mathematics. Philadelphia: SIAM.

BIAU, G., CÉROU, F., & GUYADER, A. (2015). New insights into Approximate Bayesian Computation. *Ann. Inst. H. Poincaré Probab. Statist.* **51**, 376–403.

BLUM, M. G. B., NUNES, M. A., PRANGLE, D., & SISSON, S. A. (2013). A Comparative Review of Dimension Reduction Methods in Approximate Bayesian Computation, *Statistical Science* **28**, 189–208.

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.

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

FEARNHEAD, P., AND PRANGLE, D. (2012). Constructing Summary Statistics for Approximate Bayesian Computation: Semi-automatic Approximate Bayesian Computation (with Discussion). *J. R. Statist. Soc.* B **74**, 419–474.

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

LI, W., AND FEARNHEAD, P., (2018a). On the Asymptotic Efficiency of Approximate Bayesian Computation Estimators. *Biometrika*, doi: 10.1093/biomet/asx078.

LI, W., AND FEARNHEAD, P., (2018b). Convergence of Regression Adjusted Approximate Bayesian Computation. *Biometrika*, doi: 10.1093/biomet/asx081.

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

MARIN, J-M., PILLAI, N., ROBERT, C. P., & ROUSSEAU, J. (2014). Relevant Statistics for Bayesian model choice. *J. R. Statist. Soc.* B **76**, 833–859.

MARJORAM, P., MOLITOR, J., PLAGNOL, V., & TAVARÉ, S. (2003). Markov Chain Monte Carlo Without Likelihoods. *Proceedings of the National Academy of Science USA* **100**, 15324–15328.

NOTT D., FAN, Y., MARSHALL, L., & SISSON, S. (2014). Approximate Bayesian Computation and Bayes Linear Analysis: Towards High-dimensional ABC. *J. Comput. Graphical Statist.* **23**, 65–86.

PRITCHARD, J. K., SEILSTAD, M. T., PEREZ-LEZAUN, A., & FELDMAN, M. W. (1999). Population Growth of Human Y Chromosomes: A Study of Y Chromosome Microsatellites. *Molecular Biology and Evolution* **16**, 1791–1798.

SISSON S. AND FAN, Y. (2011). Likelihood-free Markov Chain Monte Carlo. *Handbook of Markov Chain Monte Carlo* (Eds. S. Brooks, A. Gelman, G. L. Jones, X-L. Meng). Boca Raton: Chapman and Hall/CRC Press.

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

Asymptotic Properties of Approximate Bayesian Computation: Supplementary Material

SUMMARY

This supplementary material contains proofs of Theorems 1–4 and Corollary 1 in the paper. In addition, we illustrate the implications of Theorems 1–3 in the paper with a series of simulated examples based on the moving average model of Example 1.

7. PROOFS

7·1. *Proof of Theorem 1*

Let $\varepsilon_T > 0$, where, by assumption $\varepsilon_T = o(1)$, and assume that $y \in \Omega_\varepsilon = \{y : d_2\{\eta(y), b(\theta_0)\} \leq \varepsilon_T/3\}$. From assumption 1 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 $(z,\theta) \in A_\varepsilon(\delta')$

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

Hence $(z,\theta) \in A_\varepsilon(\delta')$ 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

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

and

$$\Pi[d_2\{b(\theta),b(\theta_0)\} > 4\varepsilon_T/3 + t_\varepsilon \mid d_2\{\eta(y),\eta(z)\} \le \varepsilon_T] = \Pi_\varepsilon\left[d_2\{b(\theta),b(\theta_0)\} > 4\varepsilon_T/3 + t_\varepsilon \mid \eta(y)\right]$$
$$\le \int_\Theta P_\theta\left[d_2\{b(\theta),\eta(z)\} > t_\varepsilon\right] d\Pi(\theta) \Big/ \int_\Theta P_\theta\left[d_2\{\eta(z),\eta(y)\} \le \varepsilon_T\right] d\Pi(\theta). \quad (9)$$

Moreover, since

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

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

$$\int_\Theta P_\theta\left[d_2\{\eta(z),\eta(y)\} \le \varepsilon_T\right] d\Pi(\theta) \ge \int_{d_2\{b(\theta),b(\theta_0)\}\le\varepsilon_T/3} P_\theta\left[d_2\{\eta(z),b(\theta)\} \le \varepsilon_T/3\right] d\Pi(\theta)$$
$$\ge \Pi\left[d_2\{b(\theta),b(\theta_0)\} \le \varepsilon_T/3\right] - \rho_T(\varepsilon_T/3)\int_{d_2\{b(\theta),b(\theta_0)\}\le\varepsilon_T/3} c(\theta)d\Pi(\theta).$$

If part (i) of assumption 1 holds,

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

and for ε_T small enough, or for T large enough, so that $\rho_T(\varepsilon_T/3)$ is small,

$$\int_\Theta P_\theta\left[d_2\{\eta(z),\eta(y)\} \le \varepsilon_T\right] d\Pi(\theta) \ge \Pi\left[d_2\{b(\theta),b(\theta_0)\} \le \varepsilon_T/3\right]/2,$$

which, combined with (9) and assumption 2, leads to

$$\Pi\left[d_2\{b(\theta),b(\theta_0)\} > 4\varepsilon_T/3 + t_\varepsilon \mid d_2\{\eta(z),\eta(y)\} \le \varepsilon_T\right] \lesssim \rho_T(t_\varepsilon)\varepsilon_T^{-D} \lesssim 1/M \quad (10)$$

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

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

and if ε_T satisfies

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

then (10) remains valid.

7·2. *Generalization of Theorem 2 and its Proof*

We obtain a generalization of Theorem 2 that allows differing rates of convergence for $\eta(y)$. We assume here 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 θ_0, where θ_0 is in the interior of Θ,

$$c_1 D_T \le \Sigma_T(\theta) \le c_2 D_T, \quad D_T = \mathrm{diag}\{v_T(1),\dots,v_T(k)\}, \quad (11)$$

with $0 < c_1, c_2 < \infty$, $v_T(j) \to \infty$ for all j and the $v_T(j)$ are possibly all distinct. For square matrices A, B, $A \le B$ means that the matrix $B - A$ is positive semi-definite. Thus, this generalization of Theorem 2 does not require identical convergence rates for the components of the statistic $\eta(z)$. For simplicity, we order the components so that

$$v_T(1) \le \ldots \le v_T(k_\eta). \quad (12)$$

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

In addition to assumption 2 in Section 3 of the text, the following conditions are needed to establish this generalization of Theorem 2.

Assumption 11. Assumption 1 holds and the sequence of positive definite matrices $\{\Sigma_T(\theta_0)\}_{T\geq 1}$ in (11) exists. For $\kappa > 1$ and $\delta > 0$, such that for all $\|\theta-\theta_0\| \leq \delta$, $P_\theta\left[\|\Sigma_T(\theta_0)\{\eta(z)-b(\theta)\}\| > u\right] \leq c_0/u^\kappa$ for all $0 < u \leq \delta v_T(1)$ and $c_0 < \infty$.

Assumption 12. Assumption 3 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.

Assumption 13. Given the sequence of $k_\eta \times k_\eta$ positive definite matrices $\Sigma_T(\theta)$ defined in (11), for some $\delta > 0$ and all $\|\theta-\theta_0\| \leq \delta$, the convergence in distribution: for all open sets B

$$\sup_{|\theta-\theta_0|\leq\delta}\left|P_\theta\left[\Sigma_T(\theta)\{\eta(z)-b(\theta)\}\in B\right] - P\left\{\mathcal{N}(0, I_{k_\eta})\in B\right\}\right| \longrightarrow 0$$

holds, where I_{k_η} is the $k_\eta \times k_\eta$ identity matrix.

Assumption 14. For all $\|\theta-\theta_0\| \leq \delta$, the sequence of functions $\theta \mapsto \Sigma_T(\theta)D_T^{-1}$ converges to some positive definite matrix $A(\theta)$ and is equicontinuous at θ_0.

Assumption 15. For some positive δ and all $\|\theta-\theta_0\| \leq \delta$, and for all ellipsoids

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

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

$$\lim_{T\to\infty}\sup_{|\theta-\theta_0|\leq\delta}\left|\frac{P_\theta\left[\{\Sigma_T(\theta)\}_{[j_{\max}]}\{\eta(z)-b(\theta)\}-u\in B_T\right]}{\prod_{j=1}^{j_{\max}} h_T(j)} - \varphi_{j_{\max}}(u)\right| = 0,$$

$$\frac{P_\theta\left[\{\Sigma_T(\theta)\}_{[j_{\max}]}\{\eta(z)-b(\theta)\}-u\in B_T\right]}{\prod_{j=1}^{j_{\max}} h_T(j)} \leq H(u), \qquad \int H(u)du < \infty,$$

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

THEOREM 6. *Assume that Assumptions 2, 11, with $\kappa > k_\theta$, and 12–14, are satisfied. The following results hold with P_0 probability approaching one:*

(i) if $\lim_{T\to\infty} v_T(1)\varepsilon_T = \infty$, the posterior distribution of $\varepsilon_T^{-1}(\theta-\theta_0)$ converges to the uniform distribution over the ellipse $\{w : w^\intercal B_0 w \leq 1\}$ with $B_0 = \{\nabla_\theta b(\theta_0)^\intercal \nabla_\theta b(\theta_0)\}$. Hence, for all $f(\cdot)$ continuous and bounded,

$$\int f\{\varepsilon_T^{-1}(\theta-\theta_0)\}d\Pi_\varepsilon\{\theta \mid \eta(y)\} \to \int_{u^\intercal B_0 u\leq 1} f(u)du \Big/ \int_{u^\intercal B_0 u\leq 1} du. \tag{13}$$

(ii) if there exists $k_0 < k_\eta$ such that $\lim_{T\to\infty} v_T(1)\varepsilon_T = \lim_{T\to\infty} v_T(k_0)\varepsilon_T = c$, $0 < c < \infty$, and $\lim_{T\to\infty} v_T(k_0+1)\varepsilon_T = \infty$, assuming

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

then

$$\Pi_\varepsilon\left[\Sigma_T(\theta_0)\{b(\theta)-b(\theta_0)\} - \Sigma_T(\theta_0)\{\eta(y)-b(\theta_0)\} \in B \mid \eta(y)\right] \to 0,$$

for all bounded measurable sets B, where Leb(.) *denotes the Lebesgue measure.*

(iii) if there exists $j_{\max} < k_\eta$ *such that* $\lim_{T\to\infty} v_T(j_{\max})\varepsilon_T = 0$ *and* $\lim_{T\to\infty} v_T(j_{\max}+1)\varepsilon_T = \infty$*, if assumption* 15 *is satisfied and* κ *is such that*

$$\left\{\prod_{j=1}^{j_{\max}} v_T(j)\right\}^{-1/(\kappa+j_{\max})} v_T(j_{\max}+1)^{-\kappa/(\kappa+j_{\max})} = o(\varepsilon_T),$$

then (13) *is satisfied.*

(iv) if $\lim_{T\to\infty} v_T(j)\varepsilon_T = c > 0$ *for all* $j \le k_\eta$ *or if case (ii) holds with*

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

then there exists a non-Gaussian probability distribution on $\mathbb{R}^{k_\eta}$*,* Q_c*, such that*

$$\Pi_\varepsilon\left[\Sigma_T(\theta_0)\{b(\theta)-b(\theta_0)\} - \Sigma_T(\theta_0)\{\eta(y)-b(\theta_0)\} \in B \mid \eta(y)\right] \to Q_c(B).$$

In particular,

$$Q_c(B) \propto \int_B \int_{\mathbb{R}^{k_\eta}} \mathbb{1}_{(z-x)^\intercal A(\theta_0)^\intercal A(\theta_0)(z-x)} \varphi_{k_\eta}(z)dzdx.$$

(v) if $\lim_{T\to\infty} v_T(k_\eta)\varepsilon_T = 0$ *and under assumption* 15 *holding for* $j_{\max} = k_\eta$*, then, for* $\Phi_{k_\eta}(\cdot)$ *the cumulative distribution function of the* k_η*-dimensional standard normal*

$$\lim_{T\to\infty} \Pi_\varepsilon\left[\Sigma_T(\theta_0)\{b(\theta)-b(\theta_0)\} - \Sigma_T(\theta_0)\{\eta(y)-b(\theta_0)\} \in B \mid \eta(y)\right] = \Phi_{k_\eta}(B).$$

Proof. We work with $b(\theta)$ instead of θ as the parameter, with injectivity of $\theta \mapsto b(\theta)$ required to restate all results in terms of θ. For mathematical convenience, we demonstrate this result in the case where $d_2(\eta_1,\eta_2) = \|\eta_1-\eta_2\|$, however, it holds for any metric d_2 by the equivalence of all metrics on $\mathcal{B}$.

We control the approximate Bayesian computation posterior expectation of non-negative and bounded functions $f_T(\theta-\theta_0)$ by

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

where the second equality uses the posterior concentration of $\|\theta-\theta_0\|$ at the rate $\lambda_T \gg 1/v_T(1)$. For $b_0 = b(\theta_0)$, define

$$Z_T^0 = \Sigma_T(\theta_0)\{\eta(y)-b_0\}, \quad Z_T = \Sigma_T(\theta_0)\{\eta(z)-b(\theta)\},$$

with

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

For fixed θ,

$$\begin{aligned}
&\|\Sigma_T^{-1}(\theta_0)\left[\Sigma_T(\theta_0)\{\eta(z)-b(\theta)\} - \Sigma_T(\theta_0)\{b(\theta)-b_0\}\right]\| \\
&\qquad \asymp \|D_T^{-1}\left[\Sigma_T(\theta_0)\{\eta(z)-b(\theta)\} - \Sigma_T(\theta_0)\{b(\theta)-b_0\}\right]\|
\end{aligned}$$

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) : We have $\lim_{T\to\infty} v_T(1)\varepsilon_T = \infty$. 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 $\Omega_{n,0}(M) = \{\|Z_T^0\| \leq M/2\}$, which has probability smaller than ϵ by choosing M large enough, we have that

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

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

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

as soon as $\|x(\theta)\| \leq 1 - M/\{v_T(1)\varepsilon_T\}$ with M as above. This, combined with the continuity of $\pi(\cdot)$ at θ_0 and assumption 12, implies that

$$\begin{aligned} &\int f\{\varepsilon_T^{-1}(\theta - \theta_0)\} d\Pi_\varepsilon\left\{\theta \mid \eta(y)\right\} \\ &= \frac{\int_{\|\theta-\theta_0\| \leq \lambda_T} f\{\varepsilon_T^{-1}(\theta - \theta_0)\} P_\theta\left\{\|\Sigma_T^{-1}(\theta_0)(Z_T - Z_T^0) + \varepsilon_T x(\theta)\| \leq \varepsilon_T\right\} d\theta}{\int_{\|\theta-\theta_0\| \leq \lambda_T} P_\theta\left\{\|\Sigma_T^{-1}(\theta_0)(Z_T - Z_T^0) + \varepsilon_T x(\theta)\| \leq \varepsilon_T\right\} d\theta}\{1 + o(1)\} + o_P(1) \\ &= \frac{\int_{\|x(\theta)\| \leq 1 - M/\{v_T(1)\varepsilon_T\}} f\{\varepsilon_T^{-1}(\theta - \theta_0)\} d\theta}{\int_{\|x(\theta)\| \leq 1 - M/\{v_T(1)\varepsilon_T\}} d\theta}\{1 + o(1)\} \\ &+ \frac{\int_{\|\theta-\theta_0\| \leq \lambda_T} 1\!\!1_{\|x(\theta)\| > 1 - M/\{v_T(1)\varepsilon_T\}} f\{\varepsilon_T^{-1}(\theta - \theta_0)\} P_\theta\left\{\|\Sigma_T^{-1}(\theta_0)(Z_T - Z_T^0) + \varepsilon_T x(\theta)\| \leq \varepsilon_T\right\} d\theta}{\int_{\|x(\theta)\| \leq 1 - M/\{v_T(1)\varepsilon_T\}} d\theta} \end{aligned} \tag{14}$$

The first term is approximately equal to

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

and the regularity of the function $\theta \mapsto b(\theta)$ 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

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

The second integral ratio in the right hand side of (14) converges to 0. It can be split into an integral over $1 + M/\{v_T(1)\varepsilon_T\} \geq \|x(\theta)\| \geq 1 - M/\{v_T(1)\varepsilon_T\}$ and another over $1 + M/\{v_T(1)\varepsilon_T\} \leq \|x(\theta)\|$. The first part N_2 is bounded as

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

Since

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

the second part of the second term, N_3, which is the integral over $\|x(\theta)\| > 1 + M/\{v_T(1)\varepsilon_T\}$, is bounded by

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

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 $\{w : w^\intercal B_0 w \leq 1\}$ and (i) is proved.

Case (ii) : We have $\infty > \lim_{T\to\infty} v_T(1)\varepsilon_T = c > 0$ and $\lim_{T\to\infty} v_T(k_\eta)\varepsilon_T = \infty$. We consider

$$f_T(\theta - \theta_0) = \mathbb{1}_{\Sigma_T(\theta_0)\{b(\theta)-b_0\}-Z_T^0\in B}.$$

With an obvious abuse of notation, we let

$$x = \Sigma_T(\theta_0)\{b(\theta) - b_0\} - Z_T^0.$$

We choose k_0 such that, for all $j \leq k_0$, $\lim_{T\to\infty} v_T(j)\varepsilon_T = c$ and for all $j > k_0$, $\lim_{T\to\infty} v_T(j)\varepsilon_T = \infty$. 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. Then,

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

where $\tilde{Z}_T = A_T^{-1}(\theta_0)Z_T \longrightarrow \mathcal{N}\{0, A(\theta_0)I_{k_\eta}A(\theta_0)^\intercal\}$ 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 $M_T\{v_T(k_0+1)\varepsilon_T\}^{-2} = o(1)$,

$$\begin{aligned} P_\theta\{\|D_T^{-1}(\tilde{Z}_T - x_T)\| \leq \varepsilon_T\} &\leq P_\theta\left[\sum_{j=1}^{k_0}\{\tilde{Z}_T(j) - x_T(j)\}^2 \leq v_T(1)^2\varepsilon_T^2\right] \\ &\geq P_\theta\left(\sum_{j=1}^{k_0}\{\tilde{Z}_T(j) - x_T(j)\}^2 \leq v_T(1)^2\varepsilon_T^2\left[1 - M_T\{v_T(k_0+1)\varepsilon_T\}^{-2}\right]\right) \\ &\quad - P_\theta\left[\sum_{j=k_0+1}^{k}\{\tilde{Z}_T(j) - x_T(j)\}^2 > M_T^{-1}\{\varepsilon_T v_T(k_0+1)\}^{-2}\right] \\ &\geq P_\theta\left(\sum_{j=1}^{k_0}\{\tilde{Z}_T(j) - x_T(j)\}^2 \leq v_T(1)^2\varepsilon_T^2\left[1 - M_T\{v_T(k_0+1)\varepsilon_T\}^{-2}\right]\right) - o(1). \end{aligned} \tag{15}$$

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

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

then as in case (i) we can bound

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

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

Case (iii) : We have $\lim_{T\to\infty} v_T(1)\varepsilon_T = 0$ and $\lim_{T\to\infty} v_T(k_\eta)\varepsilon_T = \infty$. Again we consider

$$f_T(\theta-\theta_0) = 1\!\!1_{\Sigma_T(\theta_0)\{b(\theta)-b_0\}-Z_T^0\in B}$$

and $x(\theta) = \Sigma_T(\theta_0)\{b(\theta)-b_0\} - Z_T^0$. As in the computations producing (15), and under assumption 15, we have

$$\begin{aligned}P_\theta\left\{\|D_T^{-1}(\tilde Z_T - x_T)\| \leq \varepsilon_T\right\} &\leq P_\theta\left[\sum_{j=1}^{j_{\max}} v_T(j)^{-2}\{\tilde Z_T(j) - x_T(j)\}^2 \leq \varepsilon_T^2\right]\\ &\geq P_\theta\left[\sum_{j=1}^{j_{\max}} v_T(j)^{-2}\{\tilde Z_T(j) - x_T(j)\}^2 \leq \varepsilon_T^2/2\right] - P_\theta\left[\sum_{j\geq j_{\max}}\{\tilde Z_T(j)-x_T(j)\}^2 > \epsilon_T^2 v_T(j_{\max}+1)^2/2\right]\\ &\geq \varphi_{j_{\max}}(x_{[k_1]})\{1+o(1)\}\prod_{j=1}^{j_{\max}}\{v_T(j)\varepsilon_T\} - c_0\{\epsilon_T v_T(j_{\max}+1)/2\}^{-\kappa}\end{aligned}$$

uniformly when $\|\theta-\theta_0\| < \lambda_T$ where $\varphi_{j_{\max}}$ is the zero mean Gaussian density in $j_{\max}$ dimensions, with covariance $\{A(\theta_0)^2\}_{[j_{\max}]}$. Since $\{\epsilon_T v_T(j_{\max}+1)/2\}^{-\kappa} = o\left[\prod_{j=1}^{j_{\max}}\{v_T(j)\varepsilon_T\}\right]$ this implies, as in case (ii), that with probability going to one

$$\limsup_{T\to\infty}\Pi_\varepsilon\left[\Sigma_T(\theta_0)\{b(\theta)-b_0\} - Z_T^0\in B\mid\eta(y)\right] \lesssim \frac{\int_{A(\theta_0)B}\varphi_{j_{\max}}(x_{[j_{\max}]})dx}{\int_{\|x\|\leq M}\varphi_{j_{\max}}(x_{[j_{\max}]})dx} \lesssim M^{-(k_\eta - j_{\max})}$$

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

Case (iv) : We have $\lim_{T\to\infty} v_T(j)\varepsilon_T = c > 0$ for all $j \leq k_\eta$. To prove equation (13) in the text, we use the computation of case (ii) with $k_0 = k_\eta$, so that (15) implies that for all x

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

uniformly in $\|\theta-\theta_0\|\leq\delta$ where $Z_\infty\sim\mathcal{N}(0, I_{k_\eta})$.

We set $u = \{\nabla_\theta b(\theta_0)^\intercal \Sigma_T(\theta_0)^\intercal \Sigma_T(\theta_0)\nabla b(\theta_0)\}^{1/2}(\theta-\theta_0)$. When $\|\theta-\theta^*\|\leq\lambda_T$, $x(\theta) = \Sigma_T(\theta_0)\nabla b(\theta_0)(\theta-\theta_0)(1+o(1)) = x(u)\{1+o(1)\}$, we can write $x(\theta) = x(u)$ and $\|x(u)\| = \|u\|$.

Choosing M and T large enough, by the dominated convergence theorem,

$$\Pi_\varepsilon\left[\Sigma_T(\theta_0)\{b(\theta)-b_0\}\,-\,Z_T^0\in B\mid\eta(y)\right]\le\frac{\int_{x(u)\in B}P\left\{\|A^{-1}(\theta_0)Z_\infty-A^{-1}(\theta_0)x(u)\|^2\le c^2\right\}du+o_p(1)}{\int_{\|u\|\le M}P\left\{\|A^{-1}(\theta_0)Z_\infty-A^{-1}(\theta_0)x(u)\|^2\le c^2\right\}du+o_p(1)}$$

$$\ge\frac{\int_{x(u)\in B}P\left\{\|A^{-1}(\theta_0)Z_\infty-A^{-1}(\theta_0)x(u)\|^2\le c^2\right\}du+o_p(1)}{\int_{\|u\|\le M}P\left\{\|A^{-1}(\theta_0)Z_\infty-A^{-1}(\theta_0)x(u)\|^2\le c^2\right\}du+o_p(1)}$$

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

$$\int_{\|u\|\le M}P\left\{\|\tilde Z_\infty-A^{-1}(\theta_0)x(u)\|^2\le c^2\right\}du\to\int_{u\in\mathbb{R}^{k_\theta}}P\left\{\|\tilde Z_\infty-A^{-1}(\theta_0)x(u)\|^2\le c^2\right\}du<\infty,$$

the result follows.

Case (v) : We have $\lim_{T\to\infty}v_T(k)\varepsilon_T=0$. Take $\Sigma_T(\theta_0)=A_T(\theta_0)D_T$. For some $\delta>0$ and all $\|\theta-\theta_0\|\le\delta$,

$$P_\theta\left[\left\|D_T^{-1}\{A_T^{-1}(\theta_0)Z_T-A_T^{-1}(\theta_0)x\}\right\|\le\varepsilon_T\right]=P_\theta\left[\left\{A_T^{-1}(\theta_0)Z_T-A_T^{-1}(\theta_0)x\right\}\in B_T\right]+o(1).$$

From both assertions of assumption 15 and by the dominated convergence theorem, the above implies for $j_{\max}=k_\eta$ that

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

Likewise, similar arguments yield

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

Together, these two equivalences yield the result in case (v). □

7·3. *Proof of Theorem 3*

Case (i) : Define $b=b(\theta)$ and $b_0=b(\theta_0)$ and, with a slight abuse of notation, in this proof we let $Z_T^0=v_T\{\eta(y)-b_0\}$ and $x=v_T(b-b_0)-Z_T^0$. We approximate the ratio

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

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

$$\begin{aligned}N_T&=\int xP_x\left\{\mid\eta(z)-\eta(y)\mid\le\varepsilon_T\right\}\pi\{b_0+(x+Z_T^0)/v_T\}dx\\&=\int_{|x|\le v_T\varepsilon_T-M}xP_x\left(\mid Z_T+x\mid\le v_T\varepsilon_T\right)\pi\{b_0+(x+Z_T^0)/v_T\}dx\\&\quad+\int_{|x|\ge v_T\varepsilon_T-M}xP_x\left(\mid Z_T+x\mid\le v_T\varepsilon_T\right)\pi\{b_0+(x+Z_T^0)/v_T\}dx,\end{aligned}\tag{16}$$

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 (16). If $0\le x\le v_T\varepsilon_T-M$, then

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

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

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

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

$$\begin{aligned} N_1 &= \int_{|x| \leq v_T\varepsilon_T - M} x P_x\left(\mid Z_T + x \mid \leq v_T\varepsilon_T\right) \pi\{b_0 + (x + Z_T^0)/v_T\} dx \\ &= \int_{|x| \leq v_T\varepsilon_T - M} x \pi\{b_0 + (x + Z_T^0)/v_T\} dx + O(M^{-\kappa+2}) \,. \end{aligned}$$

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

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

where $k = \lfloor \beta/2 \rfloor$. We split the second integral of (16) over $v_T\varepsilon_T - M \leq \mid x \mid \leq v_T\varepsilon_T + M$ and over $\mid x \mid \geq v_T\varepsilon_T + M$. We treat the latter as before: with probability going to one,

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

Finally, we study the second integral term for N_T in (16) over $v_T\varepsilon_T - M \leq \mid x \mid \leq v_T\varepsilon_T + M$. Using the assumption that $\pi(\cdot)$ is Hölder we obtain that

$$\begin{aligned} \mid N_2 \mid &= \left| \int_{v_T\varepsilon_T - M}^{v_T\varepsilon_T + M} x P_x\left(\mid Z_T + x \mid \leq v_T\varepsilon_T\right) \pi\{b_0 + (x + Z_T^0)/v_T\} dx \right. \\ &\quad \left. + \int_{-v_T\varepsilon_T - M}^{-v_T\varepsilon_T + M} x P_x\left(\mid Z_T + x \mid \leq v_T\varepsilon_T\right) \pi\{b_0 + (x + Z_T^0)/v_T\} dx \right| \\ &\leq \pi(b_0) \left| \int_{v_T\varepsilon_T - M}^{v_T\varepsilon_T + M} x P_x\left(\mid Z_T + x \mid \leq v_T\varepsilon_T\right) dx + \int_{-v_T\varepsilon_T - M}^{-v_T\varepsilon_T + M} x P_x\left(\mid Z_T + x \mid \leq v_T\varepsilon_T\right) dx \right| \\ &\quad + L \cdot M \varepsilon_T^{1+\beta\wedge 1} v_T^{\beta\wedge 1} + o_P(1) \\ &\lesssim \left| v_T\varepsilon_T \int_{-M}^{M} \left\{P_y\left(Z_T \leq -y\right) - P_y\left(Z_T \geq -y\right)\right\} dy \right| \\ &\quad + \left| v_T\varepsilon_T \int_{-M}^{M} y \left\{P_y\left(Z_T \leq -y\right) + P_y\left(Z_T \geq -y\right)\right\} dy \right| + O(M \varepsilon_T^{1+\beta\wedge 1} v_T^{\beta\wedge 1}) + o_P(1), \end{aligned}$$

with M fixed but arbitrarily large. By the dominated convergence theorem and the locally uniform Gaussian limit of Z_T, for any arbitrarily large, but fixed M,

$$\int_{-M}^{M} \{P_y(Z_T \leq -y) - P_y(Z_T \geq -y)\}\, dy = Mo(1)$$

and

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

This implies that

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

Therefore, regrouping all terms, and since $\varepsilon_T^{1+\beta\wedge 1} v_T^{\beta\wedge 1} = o(v_T\varepsilon_T)$ for all $\beta > 0$ and $\varepsilon_T = o(1)$, we obtain the representation

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

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

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

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

$$\frac{N_T}{D_T} = v_T \sum_{j=1}^{k} \frac{\nabla_b^{(2j-1)}\pi(b_0)}{\pi(b_0)(2j-1)!(2j+1)} \varepsilon_T^{2j} + o_P(1) + O(\varepsilon_T^{1+\beta} v_T). \tag{17}$$

Using the definition of N_T/D_T, dividing (17) by v_T, and rearranging terms yields

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

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

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

provided $\kappa > 1$. To control the term over $|b(\theta) - b_0| \leq \delta$, 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 $\kappa > j+1$,

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

This implies, in particular, that

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

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

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

assuming $\beta \geq 3$.

Case (ii) : Recall that $b = b(\theta)$, $b_0 = b(\theta_0)$, and define

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

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

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

which can be rewritten as

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

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

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

By injectivity of the map $\theta \mapsto b(\theta)$ in assumption 3 and assumption 6, 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 $V_0 = V(\theta_0) = \lim_{T\to\infty}\text{var}[v_T\{\eta(y)-b_0\}]$. Using arguments that mirror those in the proof of Theorem 2 part (v), by assumption 15 and the dominated convergence theorem

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

where the second equality follows from assumption 8 and the dominated convergence theorem. The result 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 assumptions 8 and 10,

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

If $h_T \to c > 0$, then

$$\begin{aligned}\frac{N_T}{h_T} &= \pi(b_0)\int x\cdot \mathrm{pr}\left\{|\ \mathcal{N}(0,1) + x/V_0^{1/2}\ |\leq c/V_0^{1/2}\right\}dx \\ &\quad + \int x\left[P_x^*(|\ Z_T + x\ |\leq h_T) - \mathrm{pr}\{|\ \mathcal{N}(0,1) + x/V_0^{1/2}\ |\leq c/V_0^{1/2}\}\right]\pi\{b_0 + (x + Z_T^0)/v_T\}dx \\ &\quad + o_P(1). \end{aligned} \tag{18}$$

The result follows if

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

respectively, $P_x^*(|\ Z_T + x\ |\leq h_T) - \mathrm{pr}\{|\ \mathcal{N}(0,1) + x/V_0^{1/2}\ |\leq c/V_0^{1/2}\} = o(1)$, for which a sufficient condition is that

$$\int |\ x\ |\left|P_x^*(|\ Z_T + x\ |\leq h_T) - \varphi(x/V_0^{1/2})\right|\pi\{b_0 + (x + Z_T^0)/v_T\}dx = o_P(1), \tag{19}$$

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

To show that the integral in (19) 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): When $|\ x\ |\leq M$, the following equivalences are satisfied:

$$\begin{aligned}&\sup_{x:|x|\leq M} |\ \pi\{b_0 + (x + Z_T^0)/v_T\} - \pi(b_0)\ |= o_P(1),\\ &\sup_{|\theta-\theta^*|\leq 1/v_T} |\ P_\theta^*(|\ Z_T + x\ |\leq h_T) - \varphi(x/V_0^{1/2})\ |= o_P(1).\end{aligned}$$

The first equation is satisfied by assumption 10 and the fact that by assumption 6 $Z_T^0/v_T = o_P(1)$. The second term follows from assumption 10 . We can now conclude that equation (19) is $o_P(1)$ over $|\ x\ |\leq M$, using the dominated convergence theorem.

The same holds for the first term in equation (18), without requiring assumption 10.

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

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

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 11,

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

which in turns implies that

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

The same computation can be conducted in case (18).

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

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

By assumption 4, for some $\kappa > 2$ we can bound N_3 as follows:

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

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

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

where assumptions 10 and 11 ensure $\int c(b)|b|\pi(b)db < \infty$. The same computation can be conducted in case (18).

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

7·4. *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 $G_0 = \nabla_\theta b(\theta_0)$, $b_0 = b(\theta_0)$ and let $Z_T^0 = v_T\{\eta(y) - b_0\}$ and

$$x(\theta) = v_T(\theta - \theta_0) - (G_0^\intercal G_0)^{-1} G_0^\intercal Z_T^0.$$

We show that $E_{\Pi_\epsilon}\{x(\theta)\} = o_p(1)$. We write

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

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

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

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 space that is orthogonal to G_0. 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 term proportional to

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

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

$$D_T \gtrsim \int_{\|G_0x(\theta)\|\leq v_T\varepsilon_T - M} P_\theta\left\{\|\eta(z)-\eta(y)\|\leq\varepsilon_T\right\}d\theta \gtrsim \varepsilon_T^{k_\theta}.$$

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 $o_p(\varepsilon_T^{k_\theta})$, with

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

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 $\mathcal{S} = \{u \in \mathbb{R}^{k_\theta} : \|u\| = 1\}$, noting that $\theta = \theta(u, r)$

$$\begin{aligned}I_2 &= v_T^{-k_\theta}(G_0^\intercal G_0)^{-1/2}\int_{v_T\varepsilon_T-M}^{v_T\varepsilon_T+M} r^{k_\theta}\int_{u\in\mathcal{S}} uP_\theta\left(\|Z_T - P_{G_0}^{\perp}Z_T^0\|^2 + r^2 - 2r\langle\Gamma_0 Z_T, u\rangle \le v_T^2\varepsilon_T^2\right)dudr\\
&\quad + O\left(v_T\varepsilon_T^{2+k_\theta}\right)\\
&= v_T^{-k_\theta}(G_0^\intercal G_0)^{-1/2}\int_{-M}^{M}(v_T\varepsilon_T + r)^{k_\theta}\times\\
&\quad\int_{u\in\mathcal{S}} uP_\theta\left(\|Z_T - P_{G_0}^{\perp}Z_T^0\|^2 - 2r\langle\Gamma_0 Z_T, u\rangle - 2\varepsilon_T v_T\langle\Gamma_0 Z_T, u\rangle \le -r^2 - 2rv_T\varepsilon_T\right)dudr\\
&\quad + O\left(v_T\varepsilon_T^{2+k_\theta}\right),\end{aligned}$$

where du denotes the Lebesgue measure on $\mathcal{S}$. Moreover, we have

$$\begin{aligned}&P_\theta\left(\|Z_T - P_{G_0}^{\perp}Z_T^0\|^2 - 2r\langle\Gamma_0 Z_T, u\rangle - 2\varepsilon_T v_T\langle\Gamma_0 Z_T, u\rangle \le -r^2 - 2rv_T\varepsilon_T\right)\\
&= P_\theta\left\{\langle\Gamma_0 Z_T, u\rangle \ge \frac{r\varepsilon_T v_T}{r + \varepsilon_T v_T} + \frac{\|Z_T - P_{G_0}^{\perp}Z_T^0\|^2 + r^2}{2(v_T\varepsilon_T + r)}\right\}\end{aligned}$$

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

$$\begin{aligned}&P_\theta\left(\|Z_T - P_{G_0}^{\perp}Z_T^0\|^2 \ge a_T\right) \lesssim c_0 a_T^{-\kappa/2},\\
&|P_\theta\left(\langle\Gamma_0 Z_T, u\rangle \ge r\right) - P_\theta\left\{\langle\Gamma_0 Z_T, u\rangle \ge r - 2a_T/(v_T\varepsilon_T)\right\}| = o(1),\end{aligned}$$

with for all r and u, $P_\theta\left(\langle\Gamma_0 Z_T, u\rangle \ge r\right) = \{1 - \Phi(r/\|\Gamma_0 A(\theta_0)^{1/2}\|)\} + o(1)$, uniformly over $\|\theta - \theta_0\| \le \delta$ and $A(\theta_0)$ as in the proof of Theorem 2. 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

$$I_2 = \varepsilon_T^{k_\theta}(G_0^\intercal G_0)^{-1/2}\int_{-M}^{M}\int_{u\in\mathcal{S}} u\left[1 - \Phi\{r/\|\Gamma_0 A(\theta_0)^{1/2}\|\}\right]dudr + o(\varepsilon_T^{k_\theta}) = o(\varepsilon_T^{k_\theta}),$$

which completes the proof.

7·5. *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 6, case (v), in this Supplementary Material, we have, for $Z_T = \Sigma_T(\theta_0)\{\eta(z) - b(\theta)\}$,

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

In the case where $\varepsilon_T \gtrsim v_T^{-1}$, then the computations in cases (ii) and (iii) in the proof of Theorem 6 imply

$$\alpha_T = P_\theta\left\{\|\eta(z) - \eta(y)\| \le \varepsilon_T\right\} \asymp \int_\Theta \varphi\{Z_T^0 + \nabla_\theta b(\theta_0)v_T(\theta - \theta_0)\}d\theta \asymp \varepsilon_T^{k_\theta}.$$

8. Illustrative Example

In this section we illustrate the implications of Theorems 1–3 in the moving model of order two that was introduced in Example 1. Consider observations from the data generating process

$$y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} \quad (t = 1, \ldots, T), \tag{20}$$

where $e_t \sim \mathcal{N}(0,1)$ is independently and identically distributed. Our prior belief for $\theta = (\theta_1, \theta_2)^\intercal$ is uniform over the invertibility region

$$\{(\theta_1,\theta_2)^\intercal : \, -2 \le \theta_1 \le 2, \; \theta_1 + \theta_2 \ge -1, \theta_1 - \theta_2 \le 1\}\,. \tag{21}$$

We follow Marin *et al.* (2011) and 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 $\eta(y) = \{\eta_0(y), \eta_1(y), \eta_2(y)\}^\intercal$. The binding function $b(\theta)$ then has the simple analytical form:

$$\theta \mapsto b(\theta) = \begin{bmatrix} E_\theta(z_t^2) \\ E_\theta(z_t z_{t-1}) \\ E_\theta(z_t z_{t-2}) \end{bmatrix} \equiv \begin{pmatrix} 1+\theta_1^2+\theta_2^2 \\ \theta_1 + \theta_1\theta_2 \\ \theta_2 \end{pmatrix}.$$

The following subsections demonstrate the implications of the limit results in the main text within the confines of the above example. By simultaneously shifting the sample size T and the tolerance parameter ε_T we can graphically illustrate Theorems 1–3.

Each demonstration considers minor variants of the following general simulation design: the true parameter vector generating the observed data is fixed at $\theta_0 = (\theta_{1,0}, \theta_{2,0})^\intercal = (0.6, 0.2)^\intercal$; for a given sample size, T, of 500, 1000, 50000, observed data, $y = (y_1, \ldots, y_T)^\intercal$, is generated from the process in equation (20); the posterior density is estimated via Algorithm 1 with the tolerance chosen to be a particular order of T, and using $N = 50,000$ Monte Carlo draws taken from uniform priors satisfying (21). In these examples we take $d_2\{\eta(z), \eta(y)\} = \|\eta(z) - \eta(y)\|$.

A central result in the main text is that the choice of ε_T drives the large sample behavior of the approximate posterior distribution and its mean. To highlight this fact, our numerical experiments will use different choices for the tolerance. In particular, and with reference to the illustration of Theorem 2, in the main text, the choices of ε_T are $\{1/T^{0.4}, 1/T^{0.5}, 1/T^{0.55}\}$. In this example, we have that $v_T = T^{0.5}$ and the three tolerance choices represent respectively cases (i), (ii) and (iii) of Theorem 2. Our use of different tolerances relays the distinction between what condition on ε_T is required for posterior concentration, and what is required to yield asymptotic normality of the posterior measure. With regard to Theorem 3, the choice $\varepsilon_T = 1/T^{0.4}$ highlights that asymptotic normality of the posterior mean can be achieved despite a lack of Gaussianity for the posterior measure itself.

8·1. *Theorem 1*

Theorem 1 implies that under regularity, as $T \to \infty$, the posterior measure $\Pi_\varepsilon\{\cdot \mid \eta(y)\}$ concentrates on sets containing θ_0, namely $\Pi_\varepsilon\{d_1(\theta, \theta) \le \delta \mid \eta(y)\} = 1 + o_P(1)$ for all $\delta > 0$, provided $\varepsilon_T = o(1)$. To demonstrate this concentration result, we take $\varepsilon_T = 1/T^{0.4}$ and run Algorithm 1, taking $N = 50,000$ draws from the prior. The results are presented in Fig. 2. To keep the Monte Carlo error at a constant level, for each sample size we retain 100 simulated values of θ that lead to realizations of $\|\eta(y) - \eta(z)\|$ below the tolerance, in agreement with the nearest-neighbor interpretation of algorithm 1.

Figure 2 shows that the posterior measure $\Pi_\varepsilon\{\cdot \mid \eta(y)\}$ is concentrating on $\theta_0 = (0.6, 0.2)^\intercal$ as T increases. The results in Fig. 2 reflect the fact that a tolerance proportional to $\varepsilon_T = 1/T^{0.4}$ will be small enough to yield posterior concentration. However, and with reference to Theorem 2, this tolerance may or may not yield asymptotic normality and, hence, correct asymptotic coverage of credible intervals. We explore this issue in the following section.

8·2. *Theorem 2*

Theorem 2 states that the shape of the approximate 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 measure will have a non-Gaussian limiting shape.

This result can be visualized by considering two alternative values for the tolerance: $\varepsilon_T = 1/T^{0.4}$ and $\varepsilon_T = 1/T^{0.55}$. Figure 3 and Fig. 4 display the resulting approximate posterior density estimates using these two tolerance rules for sample size $T = 500$ and $T = 1000$, respectively. Figure 3 demonstrates that at $T = 500$ neither posterior density, for θ_1 or θ_2 and across both tolerance rules, has a shape that is particularly Gaussian. However, at $T = 1000$, and for both θ_1 and θ_2, Fig. 4 demonstrates that the

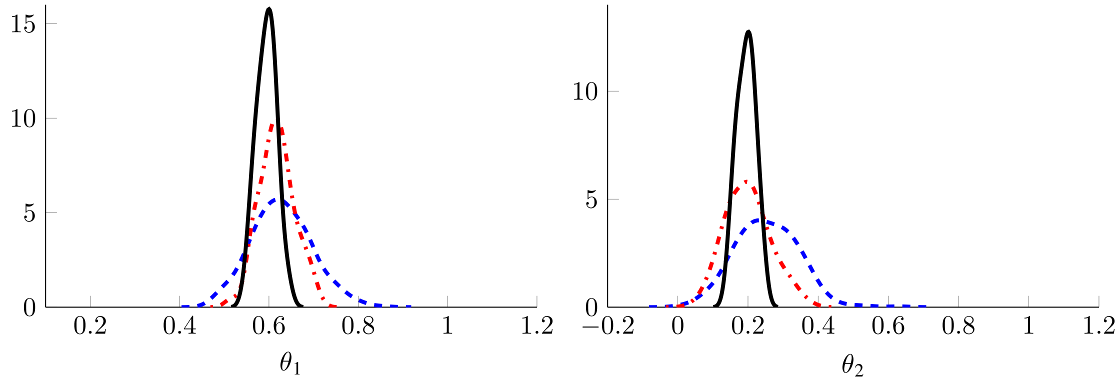

Figure 2. Posterior concentration demonstration. Estimated approximate posterior distributions across sample sizes T=500 (- - -); T=1000 (– · –); T=5000 (—).

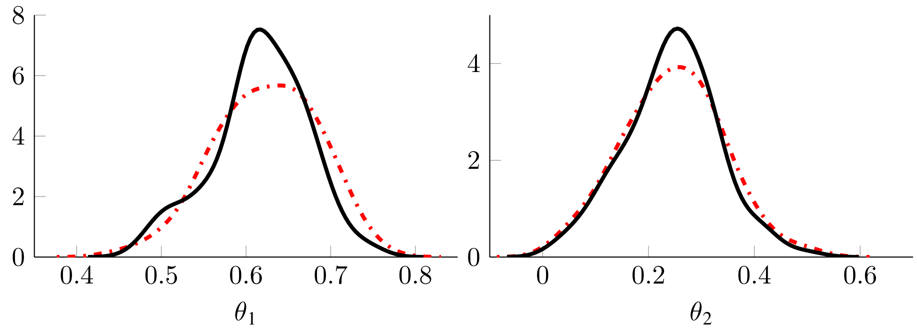

Figure 3. Comparison of two tolerance rules for ε_T : $\varepsilon_T = 1/T^{0.4}$ (– · –); $\varepsilon_T = 1/T^{0.55}$ (—); The sample size is $T = 500$.

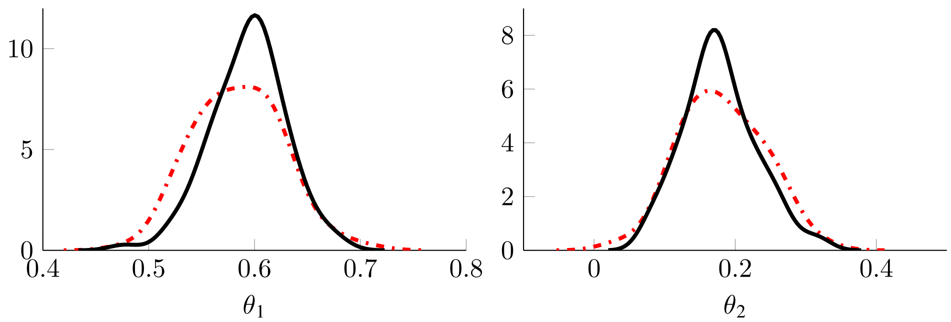

Figure 4. Same information as Fig. 3 but for $T = 1000$.

posterior densities based on the tolerance $\varepsilon_T = 1/T^{0.55}$, which satisfies the conditions for the Bernstein–von Mises result, appear to be approximately Gaussian. In contrast, the approximate posterior densities constructed from $\varepsilon_T = 1/T^{0.4}$ display non-Gaussian features.

From a practical perspective, the key result of Theorem 2 is that for credible regions built from $\Pi_\varepsilon\{\cdot \mid \eta(y)\}$ to have asymptotically correct frequentist coverage, it must be that $\varepsilon_T = o(1/v_T)$, where v_T is such that $\|\eta(z) - b(\theta)\| = O_P(1/v_T)$. In the moving average model example, $v_T = T^{0.5}$ and Theorem 2 implies that choosing a tolerance $\varepsilon_T = 1/T^{0.4}$, which corresponds to case (i) of Theorem 2, will yield credible sets whose coverage converges to one asymptotically; choosing a tolerance of $\varepsilon_T = 1/T^{.55}$, which corresponds to case (iii) of Theorem 2, will lead to asymptotically correct coverage rates; a tolerance of $\varepsilon_T = 1/T^{0.5}$ will yield coverage that is asymptotically of the correct magnitude, in that the coverage will not be zero or one, but will in general differ from the nominal level.

To demonstrate this point we generate 1000 observed artificial data sets with sample sizes $T = 500$ and $T = 1000$, and for each data set we run Algorithm 1 for all three alternative values of ε_T. For a given sample, and a given tolerance, we produce the approximate Bayesian computation posterior density in the manner described above and compute the 95% credible intervals for θ_1 and θ_2. The average length and the Monte Carlo coverage rate, across the 1000 replications, is then recorded in Table 2 for each scenario. The average length of the credible regions is clearly 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 $\varepsilon_T = 1/T^{0.55}$, which provides numerical support for the theoretical results.

Table 2. *Gaussianity of the approximate posterior distributions: 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	94.50

Width stands for average length and Cov. for Monte Carlo coverage rate.

8·3. *Theorem 3*

The key result of Theorem 3 is that even when $\Pi_\varepsilon\{\cdot \mid \eta(y)\}$ 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 $\Pi_\varepsilon\{\cdot \mid \eta(y)\}$ 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 (20) and $\theta_0 = (0.6, 0.2)^\intercal$, and produce 1000 posterior densities based on the tolerance ε_T being one of $\{1/T^{0.4}, 1/T^{0.5}, 1/T^{0.55}\}$. 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 Fig. 5. Figure 6 contains the results for $T = 1000$.

Figure 5 demonstrates that the standardized 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 Fig. 6.

[*Received March* 2018. *Revised March* 2018]

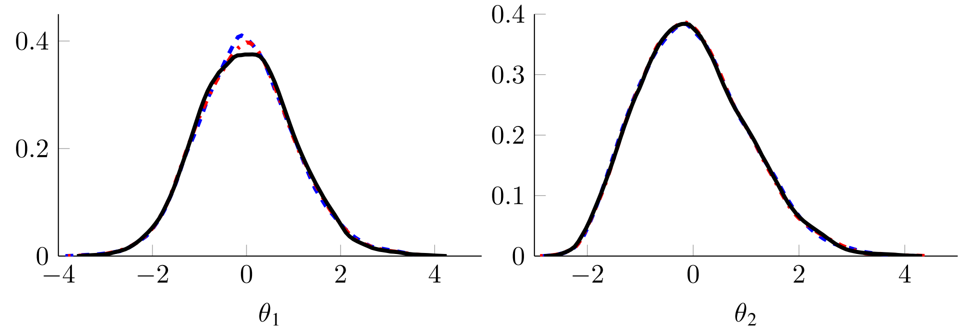

Figure 5. Comparison of different tolerance rules for ε_T: $\varepsilon_T = 1/T^{0.4}$ (- - -); $\varepsilon_T = 1/T^{0.5}$ (– · –); $\varepsilon_T = 1/T^{0.55}$ (—); The sample size is $T = 500$.

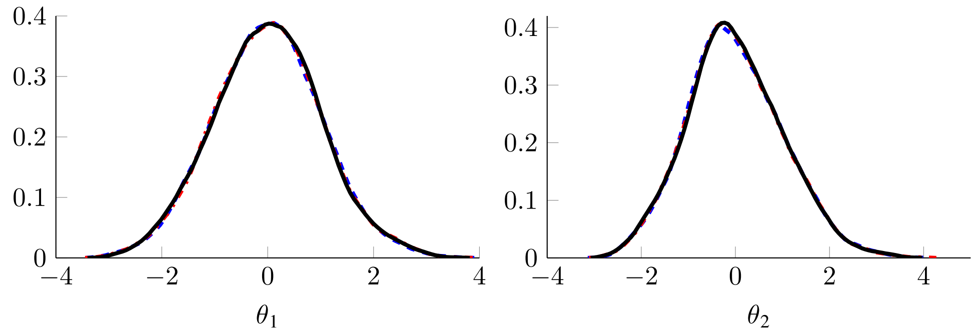

Figure 6. Same information as Fig. 5 but for $T = 1000$.