Biometrika (2017), **xx**, x, *pp*. 1–18

Printed in Great Britain

Convergence of Regression Adjusted Approximate Bayesian Computation

BY WENTAO LI

School of Mathematics, Statistics and Physics, Newcastle University, Newcastle upon Tyne NE1 7RU, U.K.

wentao.li@newcastle.ac.uk

AND PAUL FEARNHEAD

Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, U.K.

p.fearnhead@lancaster.ac.uk

SUMMARY

We present asymptotic results for the regression-adjusted version of approximate Bayesian computation introduced by Beaumont et al. (2002). We show that for an appropriate choice of the bandwidth, regression adjustment will lead to a posterior that, asymptotically, correctly quantifies uncertainty. Furthermore, for such a choice of bandwidth we can implement an importance sampling algorithm to sample from the posterior whose acceptance probability tends to unity as the data sample size increases. This compares favourably to results for standard approximate Bayesian computation, where the only way to obtain a posterior that correctly quantifies uncertainty is to choose a much smaller bandwidth; one for which the acceptance probability tends to zero and hence for which Monte Carlo error will dominate.

Keywords: Approximate Bayesian computation; Importance sampling; Local-linear regression; Partial information.

1. INTRODUCTION

Modern statistical applications increasingly require the fitting of complex statistical models. Often these models are intractable in the sense that it is impossible to evaluate the likelihood function. This excludes standard implementation of likelihood-based methods, such as maximum likelihood estimation or Bayesian analysis. To overcome this problem there has been substantial interest in likelihood-free or simulation-based methods. These methods replace calculating the likelihood by simulation of pseudo datasets from the model. Inference can then be performed by comparing these pseudo datasets, simulated for a range of different parameter values, to the actual data.

Examples of such likelihood-free methods include simulated methods of moments (Duffie & Singleton, 1993), indirect inference (Gouriéroux & Ronchetti, 1993; Heggland & Frigessi, 2004), synthetic likelihood (Wood, 2010) and approximate Bayesian computation (Beaumont et al., 2002). Of these, approximate Bayesian computation methods are arguably the most common methods for Bayesian inference, and have been popular in population genetics (e.g. Beaumont et al., 2002; Cornuet et al., 2008), ecology (e.g. Beaumont, 2010) and systems biology (e.g.

Toni et al., 2009); more recently they have seen increased use in other application areas, such as econometrics (Calvet & Czellar, 2015) and epidemiology (Drovandi & Pettitt, 2011).

The idea of approximate Bayesian computation is to first summarize the data using low-dimensional summary statistics, such as sample means or autocovariances or suitable quantiles of the data. The posterior density given the summary statistics is then approximated as follows. Assume the data are $Y_{\text{obs}} = (y_{\text{obs},1}, \ldots, y_{\text{obs},n})$ and modelled as a draw from a parametric model with parameter $\theta \in \mathbb{R}^p$. Let $K(x)$ be a positive kernel, where $\max_x K(x) = 1$, and $\varepsilon > 0$ is the bandwidth. For a given d-dimensional summary statistic $s(Y)$, our model will define a density $f_n(s \mid \theta)$. We then define a joint density, $\pi_\varepsilon(\theta, s \mid s_{\text{obs}})$, for (θ, s) as

$$\frac{\pi(\theta) f_n(s \mid \theta) K\{\varepsilon^{-1}(s - s_{\text{obs}})\}}{\int_{\mathbb{R}^p \times \mathbb{R}^d} \pi(\theta) f_n(s \mid \theta) K\{\varepsilon^{-1}(s - s_{\text{obs}})\}\, d\theta ds}, \tag{1}$$

where $s_{\text{obs}} = s(Y_{\text{obs}})$. Our approximation to the posterior density is the marginal of this joint density,

$$\pi_\varepsilon(\theta \mid s_{\text{obs}}) = \int \pi_\varepsilon(\theta, s \mid s_{\text{obs}})\, ds. \tag{2}$$

We call $\pi_\varepsilon(\theta \mid s_{\text{obs}})$ the approximate Bayesian computation posterior density. For ease of exposition we will often shorten this to posterior density in the following. To distinguish it from the actual posterior given the summary we will always call this the true posterior.

The idea of approximate Bayesian computation is that we can sample from $\pi_\varepsilon(\theta \mid s_{\text{obs}})$ without needing to evaluate the likelihood function or $f_n(s \mid \theta)$. The simplest approach is via rejection sampling (Beaumont et al., 2002), which proceeds by simulating a parameter value and an associated summary statistic from $\pi(\theta) f_n(s \mid \theta)$. This pair is then accepted with probability $K\{\varepsilon^{-1}(s - s_{\text{obs}})\}$. The accepted pairs will be drawn from (1), and the accepted parameter values will be drawn from the posterior (2). Implementing this rejection sampler requires only the ability to simulate pseudo data sets from the model, and then to be able to calculate the summary statistics for those data sets.

Alternative algorithms for simulating from the posterior include adaptive or sequential importance sampling (Beaumont et al., 2009; Bonassi & West, 2015; Lenormand et al., 2013; Filippi et al., 2013) and Markov chain Monte Carlo approaches (Marjoram et al., 2003; Wegmann et al., 2009). These aim to propose parameter values in areas of high posterior probability, and thus can be substantially more efficient than rejection sampling. However, the computational efficiency of all these methods is limited by the probability of acceptance for data simulated with a parameter value that has high posterior probability.

This paper is concerned with the asymptotic properties of approximate Bayesian computation. It builds upon recent results by Li & Fearnhead (2015) and Frazier et al. (2016). They present results on the asymptotic behaviour of the posterior distribution and the posterior mean of approximate Bayesian computation as the amount of data, n, increases. Their results highlight the tension in approximate Bayesian computation between choices of the summary statistics and bandwidth that will lead to more accurate inferences, against choices that will reduce the computational cost or Monte Carlo error of algorithms for sampling from the posterior.

An informal summary of some of these earlier results is as follows. Assume a fixed dimensional summary statistic and that the true posterior variance given this summary decreases like $1/n$ as n increases. The theoretical results compare the posterior, or posterior mean, of approximate Bayesian computation, to the true posterior, or true posterior mean, given the summary of the data. The accuracy of using approximate Bayesian computation is governed by the choice of bandwidth, and this choice should depend on n. Li & Fearnhead (2015) shows that the optimal choice of this bandwidth will be $O(n^{-1/2})$. With this choice, estimates based on the posterior mean of approximate Bayesian computation can, asymptotically, be as accurate as estimates based on the true posterior mean given the summary. Furthermore the Monte Carlo error of an importance sampling algorithm with a good proposal distribution will only inflate the mean square error of the estimator by a constant factor of the form $1 + O(1/N)$, where N is the number of pseudo data sets. These results are similar to the asymptotic results of indirect inference, where Monte Carlo error for a Monte Carlo sample of size N also inflates the overall mean square error of estimators by a factor $1 + O(1/N)$ (Gouriéroux & Ronchetti, 1993). By comparison choosing a bandwidth which is $o(n^{-1/2})$ will lead to an acceptance probability that tends to zero as $n \to \infty$, and the Monte Carlo error of approximate Bayesian computation will blow up. Choosing a bandwidth that decays more slowly than $O(n^{-1/2})$ will also lead to a regime where the Monte Carlo error dominates, and can lead to a non-negligible bias in the posterior mean that inflates the error.

While the above results for a bandwidth that is $O(n^{-1/2})$ are positive in terms of point estimates, they are negative in terms of the calibration of the posterior. With such a bandwidth the posterior density of approximate Bayesian computation always over-inflates the parameter uncertainty: see Proposition 1 below and Theorem 2 of Frazier et al. (2016).

The aim of this paper is to show that a variant of approximate Bayesian computation can yield inference that is both accurate in terms of point estimation, with its posterior mean having the same frequentist asymptotic variance as the true posterior mean given the summaries, and calibrated, in the sense that its posterior variance equals this asymptotic variance, when the bandwidth converges to zero at a rate slower than $O(n^{-1/2})$. Furthermore, this means that the acceptance probability of a good approximate Bayesian computation algorithm will tend to unity as $n \to \infty$.

2. Notation and Set-up

We denote the data by $Y_{\text{obs}} = (y_{\text{obs},1}, \ldots, y_{\text{obs},n})$, where n is the sample size, and each observation, $y_{\text{obs},i}$, can be of arbitrary dimension. Assume the data are modelled as a draw from a parametric density, $f_n(y \mid \theta)$, and consider asymptotics as $n \to \infty$. This density depends on an unknown parameter $\theta \in \mathbb{R}^p$. Let $\mathscr{B}^p$ be the Borel sigma-field on $\mathbb{R}^p$. We will let θ_0 denote the true parameter value, and $\pi(\theta)$ the prior distribution for the parameter. Denote the support of $\pi(\theta)$ by $\mathcal{P}$. Assume that a fixed-dimensional summary statistic $s_n(Y)$ is chosen and its density under our model is $f_n(s \mid \theta)$. The shorthand S_n is used to denote the random variable with density $f_n(s \mid \theta)$. Often we will simplify notation and write s and S for s_n and S_n respectively. Let $N(x; \mu, \Sigma)$ be the normal density at x with mean μ and variance Σ. Let A^c be the complement of a set A with respect to the whole space. For a series x_n we write $x_n = \Theta(a_n)$ if there exist constants m and M such that $0 < m < |x_n/a_n| < M < \infty$ as $n \to \infty$. For a real function $g(x)$, denote its gradient function at $x = x_0$ by $D_x g(x_0)$. To simplify notation, D_θ is written as D. Hereafter ε is considered to depend on n, so the notation ε_n is used.

The conditions of the theoretical results are stated below.

Condition 1. *There exists some $\delta_0 > 0$, such that $\mathcal{P}_0 = \{\theta : |\theta - \theta_0| < \delta_0\} \subset \mathcal{P}$, $\pi(\theta) \in C^2(\mathcal{P}_0)$ and $\pi(\theta_0) > 0$.*

Condition 2. *The kernel satisfies (i) $\int vK(v)\,dv = 0$; (ii) $\int \prod_{k=1}^{l} v_{i_k} K(v)\,dv < \infty$ for any coordinates $(v_{i_1}, \ldots, v_{i_l})$ of v and $l \leq p + 6$; (iii) $K(v) \propto \overline{K}(\|v\|_\Lambda^2)$ where $\|v\|_\Lambda^2 = v^T \Lambda v$*

and Λ is a positive-definite matrix, and $K(v)$ is a decreasing function of $\|v\|_\Lambda$; (iv) $K(v) = O(e^{-c_1\|v\|^{\alpha_1}})$ for some $\alpha_1 > 0$ and $c_1 > 0$ as $\|v\| \to \infty$.

CONDITION 3. *There exists a sequence a_n, satisfying $a_n \to \infty$ as $n \to \infty$, a d-dimensional vector $s(\theta)$ and a $d \times d$ matrix $A(\theta)$, such that for all $\theta \in \mathcal{P}_0$,*

$$a_n\{S_n - s(\theta)\} \to N\{0, A(\theta)\}, \;\; n \to \infty,$$

in distribution. We also assume that $s_{\text{obs}} \to s(\theta_0)$ in probability. Furthermore, (i) $s(\theta)$ and $A(\theta) \in C^1(\mathcal{P}_0)$, and $A(\theta)$ is positive definite for any θ; (ii) for any $\delta > 0$ there exists $\delta' > 0$ such that $\|s(\theta) - s(\theta_0)\| > \delta'$ for all θ satisfying $\|\theta - \theta_0\| > \delta$; and (iii) $I(\theta) = Ds(\theta)^T A^{-1}(\theta) Ds(\theta)$ has full rank at $\theta = \theta_0$.

Let $\widetilde{f}_n(s \mid \theta) = N\{s; s(\theta), A(\theta)/a_n^2\}$ be the density of the normal approximation to S and introduce the standardized random variable $W_n(s) = a_n A(\theta)^{-1/2}\{S - s(\theta)\}$. We further let $f_{W_n}(w \mid \theta)$ and $\tilde{f}_{W_n}(w \mid \theta)$ be the densities for W_n under the true model for S and under our normal approximation to the model for S.

CONDITION 4. *There exists α_n satisfying $\alpha_n / a_n^{2/5} \to \infty$ and a density $r_{\max}(w)$ satisfying Condition 2 (ii)-(iii) where $K(v)$ is replaced with $r_{\max}(w)$, such that $\sup_{\theta \in \mathcal{P}_0} \alpha_n \left| f_{W_n}(w \mid \theta) - \widetilde{f}_{W_n}(w \mid \theta) \right| \leq c_3 r_{\max}(w)$ for some positive constant c_3.*

CONDITION 5. *The following statements hold: (i) $r_{\max}(w)$ satisfies Condition 2 (iv); and (ii) $\sup_{\theta \in \mathcal{P}_0^c} f_{W_n}(w \mid \theta) = O(e^{-c_2\|w\|^{\alpha_2}})$ as $\|w\| \to \infty$ for some positive constants c_2 and α_2, and $A(\theta)$ is bounded in $\mathcal{P}$.*

Conditions 1–5 are from Li & Fearnhead (2015). Condition 2 is a requirement for the kernel function and is satisfied by all commonly used kernels, such as any kernel with compact support or the Gaussian kernel. Condition 3 assumes a central limit theorem for the summary statistic with rate a_n, and, roughly speaking, requires the summary statistic to accumulate information when n. This is a natural assumption, since many common summary statistics are sample moments, proportions, quantiles and autocorrelations, for which a central limit theorem would apply. It is also possible to verify the asymptotic normality of auxiliary model-based or composite likelihood-based summary statistics (Drovandi et al., 2015; Ruli et al., 2016) by referring to the rich literature on asymptotic properties of quasi maximum-likelihood estimators (Varin et al., 2011) or quasi-posterior estimators (Chernozhukov & Hong, 2003). This assumption does not cover ancillary statistics, using the full data as a summary statistic, or statistics based on distances, such as an asymptotically chi-square distributed test statistic. Condition 4 assumes that, in a neighborhood of θ_0, $f_n(s \mid \theta)$ deviates from the leading term of its Edgeworth expansion by a rate $a_n^{-2/5}$. This is weaker than the standard requirement, $o(a_n^{-1})$, for the remainder from Edgeworth expansion. It also assumes that the deviation is uniform, which is not difficult to satisfy in a compact neighborhood. Condition 5 further assumes that $f_n(s \mid \theta)$ has exponentially decreasing tails with rate uniform in the support of $\pi(\theta)$. This implies that posterior moments from approximate Bayesian computation are dominated by integrals in the neighborhood of θ_0 and have leading terms with concise expressions. With Condition 5 weakened, the requirement of ε_n for the proper convergence to hold might depend on the specific tail behavior of $f_n(s \mid \theta)$.

Additionally, for the results regarding regression adjustment the following moments of the summary statistic are required to exist.

CONDITION 6. *The first two moments, $\int_{\mathbb{R}^d} s f_n(s \mid \theta)\, ds$ and $\int_{\mathbb{R}^d} ss^T f_n(s \mid \theta)\, ds$, exist.*

3. Asymptotics of approximate Bayesian computation

3·1. *Posterior*

First we consider the convergence of the posterior distribution of approximate Bayesian computation, denoted by $\Pi_\varepsilon(\theta \in A \mid s_{\rm obs})$ for $A \in \mathscr{B}^p$, as $n \to \infty$. The distribution function is a random function with the randomness due to $s_{\rm obs}$. We present two convergence results. One is the convergence of the posterior distribution function of a properly scaled and centered version of θ, see Proposition 1. The other is the convergence of the posterior mean, a result which comes from Li & Fearnhead (2015) but, for convenience, is repeated as Proposition 2.

The following proposition gives three different limiting forms for $\Pi_\varepsilon(\theta \in A \mid s_{\rm obs})$, corresponding to different rates for how the bandwidth decreases relative to the rate of the central limit theorem in Condition 3. We summarize these competing rates by defining $c_\varepsilon = \lim_{n\to\infty} a_n\varepsilon_n$.

Proposition 1. *Assume Conditions 1–5. Let θ_ε denote the posterior mean of approximate Bayesian computation. As $n \to \infty$, if $\varepsilon_n = o(a_n^{-3/5})$ then the following convergence holds, depending on the value of c_ε.*

(i) If $c_\varepsilon = 0$ then

$$\sup_{A\in\mathscr{B}^p}\left|\Pi_\varepsilon\{a_n(\theta-\theta_\varepsilon)\in A \mid s_{\rm obs}\} - \int_A \psi(t)\,dt\right| \to 0,$$

in probability, where

$$\psi(t) = N\{t; 0, I(\theta_0)^{-1}\}.$$

(ii) If $c_\varepsilon \in (0,\infty)$ then for any $A \in \mathscr{B}^p$,

$$\Pi_\varepsilon\{a_n(\theta-\theta_\varepsilon)\in A \mid s_{\rm obs}\} \to \int_A \psi(t)\,dt,$$

in distribution, where

$$\psi(t) \propto \int_{\mathbb{R}^d} N[t; c_\varepsilon\beta_0\{v - E_G(v)\}, I(\theta_0)^{-1}]G(v)\,dv, \quad \beta_0 = I(\theta_0)^{-1}Ds(\theta_0)^T A(\theta_0)^{-1},$$

and $G(v)$ is a random density of v, with mean $E_G(v)$, which depends on c_ε and $Z \sim N(0, I_d)$.

(iii) If $c_\varepsilon = \infty$ then

$$\sup_{A\in\mathscr{B}^p}\left|\Pi_\varepsilon\{\varepsilon_n^{-1}(\theta-\theta_\varepsilon)\in A \mid s_{\rm obs}\} - \int_A \psi(t)\,dt\right| \to 0,$$

in probability, where $\psi(t) \propto K\{Ds(\theta_0)t\}$.

For a similar result, under different assumptions, see Theorem 2 of Frazier et al. (2016). See also Soubeyrand & Haon-Lasportes (2015) for related convergence results for the true posterior given the summaries for some specific choices of summary statistics.

The explicit form of $G(v)$ is stated in the Supplementary Material. When we have the same number of summary statistics and parameters, $d = p$, the limiting distribution simplifies to

$$\psi(t) \propto \int_{\mathbb{R}^d} N\{Ds(\theta_0)t; c_\varepsilon v, A(\theta_0)\}K(v)\,dv.$$

The more complicated form in Proposition 1 (ii) above arises from the need to project the summary statistics onto the parameter space. The limiting distribution may depend on the value of the summary statistic, s_{obs}, in the space orthogonal to $Ds(\theta_0)^T A(\theta_0)^{-1/2}$. Hence the limit depends on a random quantity, Z, which can be interpreted as the noise in s_{obs}.

The main difference between the three convergence results is the form of the limiting density $\psi(t)$ for the scaled random variable $a_{n,\varepsilon}(\theta - \theta_\varepsilon)$, where $a_{n,\varepsilon} = a_n \mathbb{1}_{c_\varepsilon < \infty} + \varepsilon_n^{-1} \mathbb{1}_{c_\varepsilon = \infty}$. For case (i) the bandwidth is sufficiently small that the approximation in approximate Bayesian computation due to accepting summaries close to the observed summary is asymptotically negligible. The asymptotic posterior distribution is Gaussian, and equals the limit of the true posterior for θ given the summary. For case (iii) the bandwidth is sufficiently big that this approximation dominates and the asymptotic posterior distribution of approximate Bayesian computation is determined by the kernel. For case (ii) the approximation is of the same order as the uncertainty in θ, which leads to an asymptotic posterior distribution that is a convolution of a Gaussian distribution and the kernel. Since the limit distributions of cases (i) and (iii) are non-random in the space $L^1(\mathbb{R}^p)$, the weak convergence is strengthened to convergence in probability in $L^1(\mathbb{R}^p)$. See the proof in Appendix A.

PROPOSITION 2. *(Theorem 3.1 of Li & Fearnhead, 2015) Assume conditions of Proposition 1. As $n \to \infty$, if $\varepsilon_n = o(a_n^{-3/5})$, $a_n(\theta_\varepsilon - \theta_0) \to N\{0, I_{ABC}^{-1}(\theta_0)\}$ in distribution. If $\varepsilon_n = o(a_n^{-1})$ or $d = p$ or the covariance matrix of the kernel is proportional to $A(\theta_0)$ then $I_{ABC}(\theta_0) = I(\theta_0)$. For other cases, $I(\theta_0) - I_{ABC}(\theta_0)$ is semi-positive definite.*

Proposition 2 helps us to compare the frequentist variability in the posterior mean of approximate Bayesian computation with the asymptotic posterior distribution given in Proposition 1. If $\varepsilon_n = o(a_n^{-1})$ then the posterior distribution is asymptotically normal with variance matrix $a_n^{-2} I(\theta_0)^{-1}$, and the posterior mean is also asymptotically normal with the same variance matrix. These results are identical to those we would get for the true posterior and posterior mean given the summary.

For an ε_n which is the same order as a_n^{-1}, the uncertainty in approximate Bayesian computation has rate a_n^{-1}. However the limiting posterior distribution, which is a convolution of the true limiting posterior given the summary with the kernel, will overestimate the uncertainty by a constant factor. If ε_n decreases slower than a_n^{-1}, the posterior contracts at a rate ε_n, and thus will over-estimate the actual uncertainty by a factor that diverges as $n \to 0$.

In summary, it is much easier to get approximate Bayesian computation to accurately estimate the posterior mean. This is possible with ε_n as large as $o(a_n^{-3/5})$ if the dimension of the summary statistic equals that of the parameter. However, accurately estimating the posterior variance, or getting the posterior to accurately reflect the uncertainty in the parameter, is much harder. As commented in Section 1, this is only possible for values of ε_n for which the acceptance probability in a standard algorithm will go to zero as n increases. In this case the Monte Carlo sample size, and hence the computational cost, of approximate Bayesian computation will have to increase substantially with n.

As one application of our theoretical results, consider observations that are independent and identically distributed from a parametric density $f(\cdot \mid \theta)$. One approach to construct the summary statistics is to use the score vector of some tractable approximating auxiliary model evaluated at the maximum auxiliary likelihood estimator (Drovandi et al., 2015). Ruli et al. (2016) constructs an auxiliary model from a composite likelihood, so the auxiliary likelihood for a single observation is $\prod_{i \in \mathscr{I}} f(y \in A_i \mid \theta)$ where $\{A_i : i \in \mathscr{I}\}$ is a set of marginal or conditional events for y. Denote the auxiliary score vector for a single observation by $cl_\theta(\cdot \mid \theta)$ and the maximum auxiliary likelihood estimator for our data set by $\widehat{\theta}_{\text{cl}}$. Then the summary statistic, s, for any pseudo data set $\{y_1, \ldots, y_n\}$ is $\sum_{j=1}^n cl_\theta(y_j \mid \widehat{\theta}_{\text{cl}})/n$.

For $y \sim f(\cdot \mid \theta)$, assume the first two moments of $cl_\theta(y \mid \theta_0)$ exist and $cl_\theta(y \mid \theta)$ is differentiable at θ. Let $H(\theta) = E_\theta\{\partial cl_\theta(y \mid \theta_0)/\partial\theta\}$ and $J(\theta) = \text{var}_\theta\{cl_\theta(y \mid \theta_0)\}$. Then if $\widehat{\theta}_{\text{cl}}$ is consistent for θ_0, Condition 3 is satisfied with

$$n^{1/2}[S - E_\theta\{cl_\theta(Y \mid \theta_0)\}] \to N\{0, J(\theta)\}, \;\; n \to \infty,$$

in distribution, and with $I(\theta_0) = H(\theta_0)^T J(\theta_0)^{-1} H(\theta_0)$.

Our results show that the posterior mean of approximate Bayesian computation, using $\varepsilon_n = O(n^{-1/2})$, will have asymptotic variance $I(\theta_0)^{-1}/n$. This is identical to the asymptotic variance of the maximum composite likelihood estimator (Varin et al., 2011). Furthermore, the posterior variance will overestimate this just by a constant factor. As we show below, using the regression correction of Beaumont et al. (2002) will correct this overestimation and produce a posterior that correctly quantifies the uncertainty in our estimates.

An alternative approach to construct an approximate posterior using composite likelihood is to use the product of the prior and the composite likelihood. In general, this leads to a poorly calibrated posterior density which substantially underestimates uncertainty (Ribatet et al., 2012). Adjustment of the composite likelihood is needed to obtain calibration, but this involves estimation of the curvature and the variance of the composite score (Pauli et al., 2011). Empirical evidence that approximate Bayesian computation more accurately quantifies uncertainty than alternative composite-based posteriors is given in Ruli et al. (2016).

3·2. *Regression Adjusted Approximate Bayesian Computation*

The regression adjustment of Beaumont et al. (2002) involves post-processing the output of approximate Bayesian computation to try to improve the resulting approximation to the true posterior. Below we will denote a sample from the posterior of approximate Bayesian computation by $\{(\theta_i, s_i)\}_{i=1,\ldots,N}$. Under the regression adjustment, we obtain a new posterior sample by using $\{\theta_i - \widehat{\beta}_\varepsilon(s_i - s_{\text{obs}})\}_{i=1,\ldots,N}$ where $\widehat{\beta}_\varepsilon$ is the least square estimate of the coefficient matrix in the linear model

$$\theta_i = \alpha + \beta(s_i - s_{\text{obs}}) + e_i, \quad i = 1, \ldots, N,$$

where e_i are independent identically distributed errors.

We can view the adjusted sample as follows. Define a constant, α_ε, and a vector β_ε as

$$(\alpha_\varepsilon, \beta_\varepsilon) = \underset{\alpha,\beta}{\arg\min}\, E_\varepsilon[\|\theta - \alpha - \beta(s - s_{\text{obs}})\|^2 \mid s_{\text{obs}}],$$

where expectation is with respect the joint posterior distribution of (θ, s) given by approximate Bayesian computation. Then the ideal adjusted posterior is the distribution of $\theta^* = \theta - \beta_\varepsilon(s - s_{\text{obs}})$ where $(\theta, s) \sim \pi_\varepsilon(\theta, s)$. The density of θ^* is

$$\pi^*_\varepsilon(\theta^* \mid s_{\text{obs}}) = \int_{\mathbb{R}^d} \pi_\varepsilon\{\theta^* + \beta_\varepsilon(s - s_{\text{obs}}), s \mid s_{\text{obs}}\}\, ds$$

and the sample we get from regression-adjusted approximate Bayesian computation is a draw from $\pi^*_\varepsilon(\theta^* \mid s_{\text{obs}})$ but with β_ε replaced by its estimator.

The variance of $\pi^*_\varepsilon(\theta^* \mid s_{\text{obs}})$ is strictly smaller than that of $\pi_\varepsilon(\theta \mid s_{\text{obs}})$ provided s is correlated with θ. The following results, which are analogous to Propositions 1 and 2, show that this reduction in variation is by the correct amount to make the resulting adjusted posterior correctly quantify the posterior uncertainty.

THEOREM 1. *Assume Conditions 1–6. Denote the mean of* $\pi^*_\varepsilon(\theta^* \mid s_{\rm obs})$ *by* θ^*_ε. *As* $n \to \infty$, *if* $\varepsilon_n = o(a_n^{-3/5})$,

$$\sup_{A \in \mathscr{B}^p} \left| \Pi_\varepsilon\{a_n(\theta^* - \theta^*_\varepsilon) \in A \mid s_{\rm obs}\} - \int_A N\{t; 0, I(\theta_0)^{-1}\}\, dt \right| \to 0,$$

in probability, and

$$a_n(\theta^*_\varepsilon - \theta_0) \to N\{0, I(\theta_0)^{-1}\},$$

in distribution. Moreover, if β_ε *is replaced by* $\widetilde{\beta}_\varepsilon$ *satisfying* $a_n\varepsilon_n(\widetilde{\beta}_\varepsilon - \beta_\varepsilon) = o_p(1)$, *the above results still hold.*

The limit of the regression adjusted posterior distribution is the true posterior given the summary provided ε_n is $o(a_n^{-3/5})$. This is a slower rate than that at which the posterior contracts, which, as we will show in the next section, has important consequences in terms of the computational efficiency of approximate Bayesian computation. The regression adjustment corrects both the additional noise of the posterior mean when $d > p$ and the overestimated uncertainty of the posterior. This correction comes from the removal of the first order bias caused by ε. Blum (2010) shows that the regression adjustment reduces the bias of approximate Bayesian computation when $E(\theta \mid s)$ is linear and the residuals $\theta - E(\theta \mid s)$ are homoscedastic. Our results do not require these assumptions, and suggest that the regression adjustment should be applied routinely with approximate Bayesian computation provided the coefficients β_ε can be estimated accurately.

With the simulated sample, β_ε is estimated by $\widehat{\beta}_\varepsilon$. The accuracy of $\widehat{\beta}_\varepsilon$ can be seen by the following decomposition,

$$\begin{aligned}\widehat{\beta}_\varepsilon &= \mathrm{cov}_N(s, \theta)\mathrm{var}_N(s)^{-1} \\ &= \beta_\varepsilon + \frac{1}{a_n\varepsilon_n}\mathrm{cov}_N\left\{\frac{s - s_\varepsilon}{\varepsilon_n}, a_n(\theta^* - \theta^*_\varepsilon)\right\}\mathrm{var}_N\left\{\frac{s - s_\varepsilon}{\varepsilon_n}\right\}^{-1},\end{aligned}$$

where cov_N and var_N are the sample covariance and variance matrices, and s_ϵ is the sample mean. Since $\mathrm{cov}(s, \theta^*) = 0$ and the distributions of $s - s_\varepsilon$ and $\theta^* - \theta^*_\varepsilon$ contract at rates ε_n and a_n^{-1} respectively, the error $\widehat{\beta}_\varepsilon - \beta_\varepsilon$ can be shown to have the rate $O_p\{(a_n\varepsilon_n)^{-1}N^{-1/2}\}$ as $n \to \infty$ and $N \to \infty$. We omit the proof, since it is tedious and similar to the proof of the asymptotic expansion of β_ε in Lemma 4. Thus, if N increases to infinity with n, $\widehat{\beta}_\varepsilon - \beta_\varepsilon$ will be $o_p\{(a_n\varepsilon_n)^{-1}\}$ and the convergence of Theorem 1 will hold instead.

Alternatively we can get an idea of the additional error for large N from the following proposition.

PROPOSITION 3. *Assume Conditions 1–6. Consider* $\theta^* = \theta - \widehat{\beta}_\varepsilon(s - s_{\rm obs})$. *As* $n \to \infty$, *if* $\varepsilon_n = o(a_n^{-3/5})$ *and* N *is large enough, for any* $A \in \mathscr{B}^p$,

$$\Pi_\varepsilon\{a_n(\theta^* - \theta^*_\varepsilon) \in A \mid s_{\rm obs}\} \to \int_A \psi(t)\, dt,$$

in distribution, where

$$\psi(t) \propto \int_{\mathbb{R}^d} N\left[t; \frac{\eta}{N^{1/2}}\{v - E_G(v)\}, I(\theta_0)^{-1}\right] G(v)\, dv,$$

when $c_\varepsilon < \infty$,

$$\psi(t) \propto \int_{\mathbb{R}^p} N\left\{t; \frac{\eta}{N^{1/2}} Ds(\theta_0)t', I(\theta_0)^{-1}\right\} K\{Ds(\theta_0)t'\}\, dt',$$

when $c_\varepsilon = \infty$, *and* $\eta = O_p(1)$.

The limiting distribution here can be viewed as the convolution of the limiting distribution obtained when the optimal coefficients are used and that of a random variable, which relates to the error in our estimate of β_ε, and that is $O_p(N^{-1/2})$.

3·3. *Acceptance Rates when* ε *is Negligible*

Finally we present results for the acceptance probability of approximate Bayesian computation, the quantity that is central to the computational cost of importance sampling or Markov chain Monte Carlo-based algorithms. We consider a set-up where we propose the parameter value from a location-scale family. That is, we can write the proposal density as the density of a random variable, $\sigma_n X + \mu_n$, where $X \sim q(\cdot)$, $E(X) = 0$ and σ_n and μ_n are constants that can depend on n. The average acceptance probability $p_{\text{acc},q}$ would then be

$$\int_{\mathcal{P}\times\mathbb{R}^d} q_n(\theta) f_n(s \mid \theta) K\{\varepsilon_n^{-1}(s - s_{\text{obs}})\}\, ds d\theta,$$

where $q_n(\theta)$ is the density of $\sigma_n X + \mu_n$. This covers the proposal distribution in fundamental sampling algorithms, including the random-walk Metropolis algorithm and importance sampling with unimodal proposal distribution, and serves as the building block for many advanced algorithms where the proposal distribution is a mixture of distributions from location-scale families, such as iterative importance sampling-type algorithms.

We further assume that $\sigma_n(\mu_n - \theta_0) = O_p(1)$, which means θ_0 is in the coverage of $q_n(\theta)$. This is a natural requirement for any good proposal distribution. The prior distribution and θ_0 as a point mass are included in this proposal family. This condition would also apply to many Markov chain Monte Carlo implementations of approximate Bayesian computation after convergence.

As above, define $a_{n,\varepsilon} = a_n \mathbb{1}_{c_\varepsilon<\infty} + \varepsilon_n^{-1}\mathbb{1}_{c_\varepsilon=\infty}$ to be the smaller of a_n and ε_n^{-1}. Asymptotic results for $p_{\text{acc},q}$ when σ_n has the same rate as $a_{n,\varepsilon}^{-1}$ are given in Li & Fearnhead (2015). Here we extend those results to other regimes.

THEOREM 2. *Assume the conditions of Proposition 1. As* $n \to \infty$, *if* $\varepsilon_n = o(a_n^{-1/2})$*: (i) if* $c_\varepsilon = 0$ *or* $\sigma_n/a_{n,\varepsilon}^{-1} \to \infty$, *then* $p_{\text{acc},q} \to 0$ *in probability; (ii) if* $c_\varepsilon \in (0, \infty)$ *and* $\sigma_n/a_{n,\varepsilon}^{-1} \to r_1 \in [0, \infty)$, *or* $c_\varepsilon = \infty$ *and* $\sigma_n/a_{n,\varepsilon}^{-1} \to r_1 \in (0, \infty)$, *then* $p_{\text{acc},q} = \Theta_p(1)$*; (iii) if* $c_\varepsilon = \infty$ *and* $\sigma_n/a_{n,\varepsilon}^{-1} \to 0$, *then* $p_{\text{acc},q} \to 1$ *in probability.*

The proof of Theorem 2 can be found in the Supplementary Material. The underlying intuition is as follows. For the summary statistic, s, sampled with parameter value θ, the acceptance probability depends on

$$\frac{s - s_{\text{obs}}}{\varepsilon_n} = \frac{1}{\varepsilon_n}[\{s - s(\theta)\} + \{s(\theta) - s(\theta_0)\} + \{s(\theta_0) - s_{\text{obs}}\}], \tag{3}$$

where $s(\theta)$ is the limit of s in Condition 3. The distance between s and s_{obs} is at least $O_p(a_n^{-1})$, since the first and third bracketed terms are $O_p(a_n^{-1})$. If $\varepsilon_n = o(a_n^{-1})$ then, regardless of the value of θ, (3) will blow up as $n \to \infty$ and hence $p_{\text{acc},q}$ goes to 0. If ε_n decreases with a rate slower than a_n^{-1}, (3) will go to zero providing we have a proposal which ensures that the middle term is $o_p(\varepsilon_n)$, and hence $p_{\text{acc},q}$ goes to unity.

Theorem 1 shows that, without the regression adjustment, approximate Bayesian computation requires ε_n to be $o(a_n^{-1})$ if its posterior is to converge to the true posterior given the summary. In this case Theorem 2 shows that the acceptance rate will degenerate to zero as $n \to \infty$ regardless of the choice of $q(\cdot)$. On the other hand, with the regression adjustment, we can choose $\varepsilon_n = o(a_n^{-3/5})$ and still have convergence to the true posterior given the summary. For such a choice, if our proposal density satisfies $\sigma_n = o(\varepsilon_n)$, the acceptance rate will go to unity as $n \to \infty$.

4. Numerical Example

Here we illustrate the gain of computational efficiency from using the regression adjustment on the g-and-k distribution, a popular model for testing approximate Bayesian computation methods (e.g., Fearnhead & Prangle, 2012; Marin et al., 2014). The data are independent and identically distributed from a distribution defined by its quantile function,

$$F^{-1}(x;\alpha,\beta,\gamma,\kappa) = \alpha + \beta\left[1 + 0.8\frac{1-\exp\{-\gamma z(x)\}}{1+\exp\{-\gamma z(x)\}}\right]\{1+z(x)^2\}^{\kappa} z(x), \quad x \in [0,1],$$

where α and β are location and scale parameters, γ and κ are related to the skewness and kurtosis of the distribution, and $z(x)$ is the corresponding quantile of a standard normal distribution. No closed form is available for the density but simulating from the model is straightforward by transforming realisations from the standard normal distribution.

In the following we assume the parameter vector $(\alpha, \beta, \gamma, \kappa)$ has a uniform prior in $[0, 10]^4$ and multiple datasets are generated from the model with $(\alpha, \beta, \gamma, \kappa) = (3, 1, 2, 0.5)$. To illustrate the asymptotic behaviour of approximate Bayesian computation, 50 data sets are generated for each of a set of values of n ranging from 500 to $10,000$. Consider estimating the posterior means, denoted by $\mu = (\mu_1, \ldots, \mu_4)$, and standard deviations, denoted by $\sigma = (\sigma_1, \ldots, \sigma_4)$, of the parameters. The summary statistic is a set of evenly spaced quantiles of dimension 19.

The bandwidth is chosen via fixing the proportion of the Monte Carlo sample to be accepted, and the accepted proportions needed to achieve certain approximation accuracy for estimates with and without the adjustment are compared. A higher proportion means more simulated parameter values can be kept for inference. The accuracy is measured by the average relative errors of estimating μ or σ,

$$\mathrm{RE}_\mu = \frac{1}{4}\sum_{k=1}^{4}\frac{|\widehat{\mu}_k - \mu_k|}{\mu_k}, \quad \mathrm{RE}_\sigma = \frac{1}{4}\sum_{k=1}^{4}\frac{|\widehat{\sigma}_k - \sigma_k|}{\sigma_k},$$

for estimators $\widehat{\mu} = (\widehat{\mu}_1, \ldots, \widehat{\mu}_4)$ and $\widehat{\sigma} = (\widehat{\sigma}_1, \ldots, \widehat{\sigma}_4)$. The proposal distribution is normal, with the covariance matrix selected to inflate the posterior covariance matrix by a constant factor c^2 and the mean vector selected to differ from the posterior mean by half of the posterior standard deviation, which avoids the case that the posterior mean can be estimated trivially. We consider a series of increasing c in order to investigate the impact of the proposal distribution getting worse.

The results in Figure 1 show that the required acceptance rate for the regression adjusted estimates is higher than that for the unadjusted estimates in almost all cases. For estimating the posterior mean, the improvement is small. For estimating the posterior standard deviations, the improvement is much larger. To achieve each level of accuracy, the acceptance rates of the unadjusted estimates are all close to zero. Those of the regression-adjusted estimates are higher by up to 2 orders of magnitude, so the Monte Carlo sample size needed to achieve the same accuracy can be reduced correspondingly.

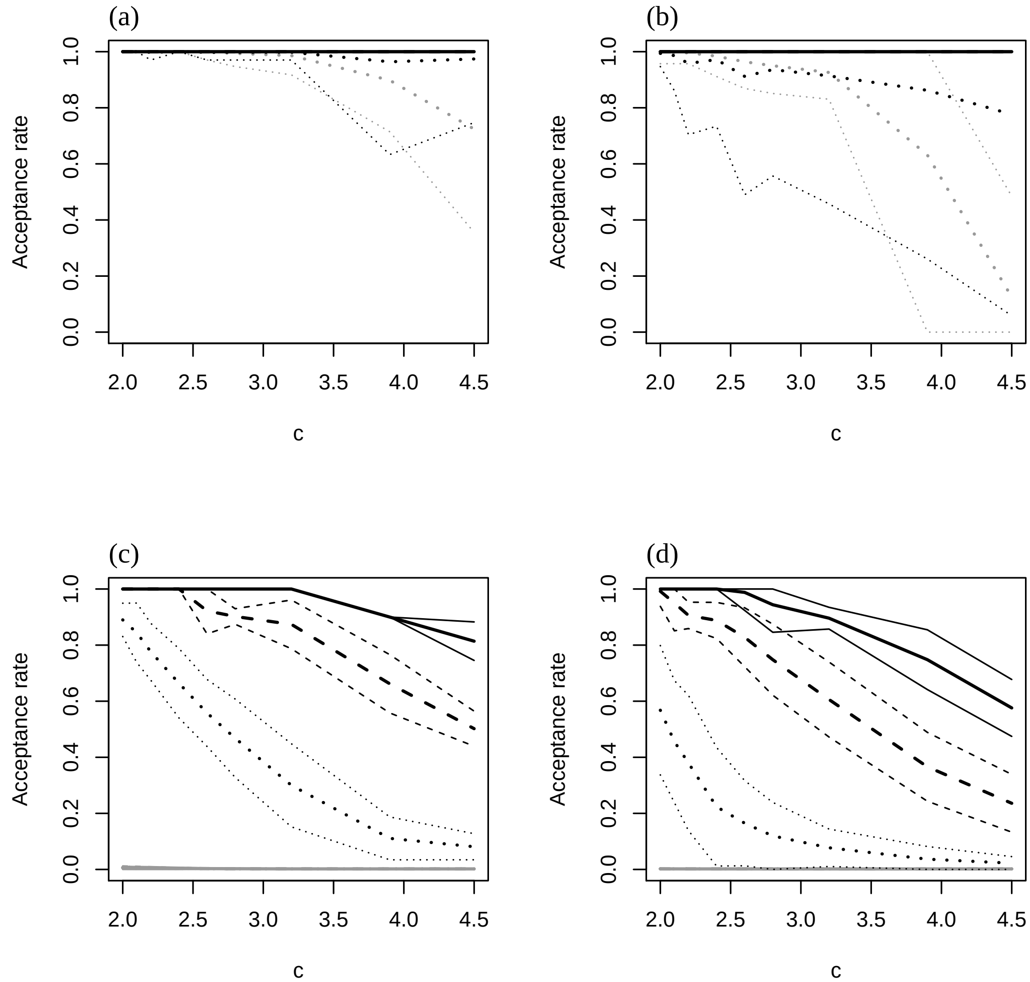

Fig. 1: Acceptance rates required for different degrees of accuracy of approximate Bayesian computation and different variances of the proposal distribution (which are proportional to c). In each plot we show results for standard (grey-line) and regression adjusted (black-line) approximate Bayesian computation and for different values of n: $n = 500$ (dotted), $n = 3,000$ (dashed) and $n = 10,000$ (solid). The averages over 50 data sets (thick) and their 95% confidence intervals (thin) are reported. Results are for a relative error of 0.08 and 0.05 in the posterior mean, in (a) and (b) respectively, and for a relative error of 0.2 and 0.1 in the posterior standard deviation, in (c) and (d) respectively.

5. Discussion

One way to implement approximate Bayesian computation so that the acceptance probability tends to 1 as n increases is to use importance sampling with a suitable proposal from a location-scale family. The key difficulty with finding a suitable proposal is to ensure that the location parameter is close to the true parameter, where close means the distance is $O(\varepsilon_n)$. This can be achieved by having a preliminary analysis of the data, and using the point estimate of the parameter from this preliminary analysis as the location parameter (Beaumont et al., 2009; Li & Fearnhead, 2015).

Acknowledgment

This work was funded by the Engineering and Physical Sciences Research Council, under the i-like programme grant.

Supplementary Material

Proofs of lemmas and Theorem 2 are included in the online supplementary material.

Appendix

Proof of Result from Section 3·1

Throughout the data are considered to be random. For any integer $l > 0$ and a set $A \subset \mathbb{R}^l$, we use the convention that $cA + x$ denotes the set $\{ct + x : t \in A\}$ for $c \in \mathbb{R}$ and $x \in \mathbb{R}^l$. For a non-negative function $h(x)$, integrable in $\mathbb{R}^l$, denote the normalised function $h(x)/\int_{\mathbb{R}^l} h(x)\,dx$ by $h(x)^{(\text{norm})}$. For a vector x, denote a general polynomial of elements of x with degree up to l by $P_l(x)$. For any fixed $\delta < \delta_0$, let B_δ be the neighborhood $\{\theta : \|\theta - \theta_0\| < \delta\}$. Let $\pi_\delta(\theta)$ be $\pi(\theta)$ truncated in B_δ. Recall that $\widetilde{f}_n(s \mid \theta)$ denotes the normal density with mean $s(\theta)$ and covariance matrix $A(\theta)/a_n^2$. Let $t(\theta) = a_{n,\varepsilon}(\theta - \theta_0)$ and $v(s) = \varepsilon_n^{-1}(s - s_{\text{obs}})$, rescaled versions θ and s. For any $A \in \mathscr{B}^p$, let $t(A)$ be the set $\{\phi : \phi = t(\theta) \text{ for some } \theta \in A\}$.

Define $\widetilde{\Pi}_\varepsilon(\theta \in A \mid s_{\text{obs}})$ to be the normal counterpart of $\Pi_\varepsilon(\theta \in A \mid s_{\text{obs}})$ with truncated prior, obtained by replacing $\pi(\theta)$ and $f_n(s \mid \theta)$ in Π_ε by $\pi_\delta(\theta)$ and $\widetilde{f}_n(s \mid \theta)$. So let

$$\widetilde{\pi}_\varepsilon(\theta, s \mid s_{\text{obs}}) = \frac{\pi_\delta(\theta)\widetilde{f}_n(s \mid \theta)K\{\varepsilon_n^{-1}(s - s_{\text{obs}})\}}{\int_{B_\delta}\int_{\mathbb{R}^d} \pi_\delta(\theta)\widetilde{f}_n(s \mid \theta)K\{\varepsilon_n^{-1}(s - s_{\text{obs}})\}\,d\theta ds},$$

$\widetilde{\pi}_\varepsilon(\theta \mid s_{\text{obs}}) = \int_{\mathbb{R}^d} \widetilde{\pi}_\varepsilon(\theta, s \mid s_{\text{obs}})\,ds$ and $\widetilde{\Pi}_\varepsilon(\theta \in A \mid s_{\text{obs}})$ be the distribution function with density $\widetilde{\pi}_\varepsilon(\theta \mid s_{\text{obs}})$. Denote the mean of $\widetilde{\Pi}_\varepsilon$ by $\widetilde{\theta}_\varepsilon$. Let $W_{\text{obs}} = a_n A(\theta_0)^{-1/2}\{s_{\text{obs}} - s(\theta_0)\}$ and $\beta_0 = I(\theta_0)^{-1}Ds(\theta_0)^T A(\theta_0)^{-1}$. By Condition 3, $W_{\text{obs}} \to Z$ in distribution as $n \to \infty$, where $Z \sim N(0, I_d)$.

Since the approximate Bayesian computation likelihood within $\widetilde{\Pi}_\varepsilon$ is an incorrect model for s_{obs}, standard posterior convergence results do not apply. However, if we condition on the value of the summary, s, then the distribution of θ is just the true posterior given s. Thus we can express the posterior from approximate Bayesian computation as a continuous mixture of these true posteriors. Let $\widetilde{\pi}_{\varepsilon,tv}(t, v) = a_{n,\varepsilon}^{-d}\pi_\delta(\theta_0 + a_{n,\varepsilon}^{-1}t)\widetilde{f}_n(s_{\text{obs}} + \varepsilon_n v \mid \theta_0 + a_{n,\varepsilon}^{-1}t)K(v)$. For any $A \in \mathscr{B}^p$, we rewrite $\widetilde{\Pi}_\varepsilon$ as,

$$\widetilde{\Pi}_\varepsilon(\theta \in A \mid s_{\text{obs}}) = \int_{\mathbb{R}^d}\int_{t(B_\delta)} \widetilde{\Pi}(\theta \in A \mid s_{\text{obs}} + \varepsilon_n v)\widetilde{\pi}_{\varepsilon,tv}(t, v)^{(\text{norm})}\,dtdv, \qquad (4)$$

where $\widetilde{\Pi}(\theta \in A \mid s)$ is the posterior distribution with prior $\pi_\delta(\theta)$ and likelihood $\widetilde{f}_n(s \mid \theta)$.

Using results from Kleijn & van der Vaart (2012), the leading term of $\widetilde{\Pi}(\theta \in A \mid s_{\text{obs}} + \varepsilon_n v)$ can be obtained and is stated in the following lemma.

LEMMA 1. *Assume Conditions 3 and 4. If* $\varepsilon_n = O(a_n^{-1})$, *for any fixed* $v \in \mathbb{R}^d$ *and small enough* δ,

$$\sup_{A \in \mathscr{B}^p} \left| \widetilde{\Pi}\{a_n(\theta - \theta_0) \in A \mid s_{\text{obs}} + \varepsilon_n v\} - \int_A N[t; \beta_0\{A(\theta_0)^{1/2} W_{\text{obs}} + c_\varepsilon v\}, I(\theta_0)^{-1}]\, dt \right| \to 0,$$

in probability as $n \to \infty$.

This leading term is Gaussian, but with a mean that depends on v. Thus asymptotically, the posterior of approximate Bayesian computation is the distribution of the sum of a Gaussian random variable and $\beta_0 c_\varepsilon V$, where V has density proportional to $\int \widetilde{\pi}_{\varepsilon,tv}(t, v)\, dt$.

To make this argument rigorous, and to find the distribution of this sum of random variables we need to introduce several functions that relate to the limit of $\widetilde{\pi}_{\varepsilon,tv}(t', v)$. For a rank-$p$ $d \times p$ matrix A, a rank-d $d \times d$ matrix B and a d-dimensional vector c, define $g(v; A, B, c) = \exp[-(c + Bv)^T\{I - A(A^TA)^{-1}A^T\}(c + Bv)/2]/(2\pi)^{(d-p)/2}$. Let

$$g_n(t, v) = \begin{cases} N\left\{Ds(\theta_0)t; a_n\varepsilon_n v + A(\theta_0)^{1/2}W_{\text{obs}}, A(\theta_0)\right\} K(v), & c_\varepsilon < \infty, \\ N\left\{Ds(\theta_0)t; v + \frac{1}{a_n\varepsilon_n}A(\theta_0)^{1/2}W_{\text{obs}}, \frac{1}{a_n^2\varepsilon_n^2}A(\theta_0)\right\} K(v), & c_\varepsilon = \infty, \end{cases}$$

$G_n(v)$ be $g\{v; A(\theta_0)^{-1/2}Ds(\theta_0), a_n\varepsilon_n A(\theta_0)^{-1/2}, W_{\text{obs}}\}K(v)$, and $E_{G_n}(\cdot)$ be the expectation under the density $G_n(v)^{(\text{norm})}$. In both cases it is straightforward to show that $\int_{\mathbb{R}^p} g_n(t, v)\, dt = |A(\theta_0)|^{-1/2}G_n(v)$. Additionally, for the case $c_\varepsilon = \infty$, define $v'(v, t) = A(\theta_0)^{1/2}W_{\text{obs}} + a_n\varepsilon_n v - a_n\varepsilon_n Ds(\theta_0)t$ and

$$g_n'(t, v') = N\{v'; 0, A(\theta_0)\}K\left\{Ds(\theta_0)t + \frac{1}{a_n\varepsilon_n}v' - \frac{1}{a_n\varepsilon_n}A(\theta_0)^{1/2}W_{\text{obs}}\right\}.$$

Then with the transformation $v' = v'(v, t)$, $g_n(t, v)dv = g_n'(t, v')dv'$. Let

$$g(t, v) = \begin{cases} N\{Ds(\theta_0)t; c_\varepsilon v + A(\theta_0)^{1/2}Z, A(\theta_0)\}K(v), & c_\varepsilon < \infty, \\ K\{Ds(\theta_0)t\}N\{v; 0, A(\theta_0)\}, & c_\varepsilon = \infty, \end{cases}$$

$G(v)$ be $g\{v; A(\theta_0)^{-1/2}Ds(\theta_0), c_\varepsilon A(\theta_0)^{-1/2}, Z\}K(v)$ and $E_G(\cdot)$ be the expectation under the density $G(v)^{(\text{norm})}$. When $c_\varepsilon < \infty$, $\int_{\mathbb{R}^p} g(t, v)\, dt = |A(\theta_0)|^{-1/2}G(v)$.

Expansions of $\int_{\mathbb{R}^d}\int_{t(B_\delta)} \widetilde{\pi}_{\varepsilon,tv}(t, v)dtdv$ are given in the following lemma.

LEMMA 2. *Assume Conditions 1–3. If* $\varepsilon_n = o(a_n^{-1/2})$, *then* $\int_{\mathbb{R}^d}\int_{t(B_\delta)} |\widetilde{\pi}_{\varepsilon,tv}(t, v) - \pi(\theta_0)g_n(t, v)|\, dtdv \to 0$ *in probability and* $\int_{\mathbb{R}^d}\int_{t(B_\delta)} g_n(t, v)\, dtdv = \Theta_p(1)$, *as* $n \to \infty$. *Furthermore, for* $l \leq 6$, $\int_{\mathbb{R}^d}\int_{t(B_\delta)} P_l(v)g_n(t, v)\, dtdv$ *converges to* $|A(\theta_0)|^{-1/2}\int_{\mathbb{R}^d} P_l(v)G(v)\, dv$ *in distribution when* $c_\varepsilon < \infty$ *and converges to* $\int_{\mathbb{R}^p} P_l\{Ds(\theta_0)t\}K\{Ds(\theta_0)t\}\, dt$ *in probability when* $c_\varepsilon = \infty$, *as* $n \to \infty$.

The following lemma states that Π_ε and $\widetilde{\Pi}_\varepsilon$ are asymptotically the same and gives an expansion of $\widetilde{\theta}_\varepsilon$.

LEMMA 3. *Assume Conditions 1–4. If* $\varepsilon_n = o(a_n^{-1/2})$, *then*

(a) for any $\delta < \delta_0$, $\Pi_\varepsilon(\theta \in B_\delta^c \mid s_{\text{obs}})$ *and* $\widetilde{\Pi}_\varepsilon(\theta \in B_\delta^c \mid s_{\text{obs}})$ *are* $o_p(1)$;

(b) there exists a $\delta < \delta_0$ *such that* $\sup_{A \in \mathscr{B}^p} \left|\Pi_\varepsilon(\theta \in A \cap B_\delta \mid s_{\text{obs}}) - \widetilde{\Pi}_\varepsilon(\theta \in A \cap B_\delta \mid s_{\text{obs}})\right| = o_p(1)$;

(c) if, in addition, Condition 5 holds, then $a_{n,\varepsilon}(\theta_\varepsilon - \widetilde{\theta}_\varepsilon) = o_p(1)$, *and* $\widetilde{\theta}_\varepsilon = \theta_0 + a_n^{-1}\beta_0 A(\theta_0)^{1/2}W_{\rm obs} + \varepsilon_n\beta_0 E_{G_n}(v) + r_{n,1}$ *where the remainder* $r_{n,1} = o_p(a_n^{-1})$.

Proof of Proposition 1. Lemma 10 in the Supplementary Material shows that $\Pi_\varepsilon\{a_{n,\varepsilon}(\theta - \theta_\varepsilon) \in A \mid s_{\rm obs}\}$ and $\widetilde{\Pi}_\varepsilon\{a_{n,\varepsilon}(\theta - \widetilde{\theta}_\varepsilon) \in A \mid s_{\rm obs}\}$ have the same limit, in distribution when $c_\varepsilon \in (0,\infty)$ and in total variation form when $c_\varepsilon = 0$ or ∞. Therefore it is sufficient to only consider the convergence of $\widetilde{\Pi}_\varepsilon$ of the properly scaled and centered θ.

When $a_n\varepsilon_n \to c_\varepsilon < \infty$, according to (4),

$$\widetilde{\Pi}_\varepsilon\{a_n(\theta - \widetilde{\theta}_\varepsilon) \in A \mid s_{\rm obs}\} = \int_{\mathbb{R}^d}\int_{t(B_\delta)} \widetilde{\Pi}\{a_n(\theta - \theta_0) \in A + a_n(\widetilde{\theta}_\varepsilon - \theta_0) \mid s_{\rm obs} + \varepsilon_n v\}\widetilde{\pi}_{\varepsilon,tv}(t',v)^{(\rm norm)}\, dt'dv.$$

By Lemma 1 and Lemma 3(c), we have

$$\sup_{A\in\mathscr{B}^p}\left|\widetilde{\Pi}\{a_n(\theta - \theta_0) \in A + a_n(\widetilde{\theta}_\varepsilon - \theta_0) \mid s_{\rm obs} + \varepsilon_n v\} - \int_A N\{t; \mu_n(v), I(\theta_0)^{-1}\}\, dt\right| = o_p(1),$$

where $\mu_n(v) = \beta_0\{c_\varepsilon v - a_n\varepsilon_n E_{G_n}(v)\} - a_n r_{n,1}$. Then with Lemma 2, the leading term of $\widetilde{\Pi}_\varepsilon\{a_n(\theta - \widetilde{\theta}_\varepsilon) \in A \mid s_{\rm obs}\}$ equals

$$\sup_{A\in\mathscr{B}^p}\left|\widetilde{\Pi}_\varepsilon\{a_n(\theta - \widetilde{\theta}_\varepsilon) \in A \mid s_{\rm obs}\} - \int_{\mathbb{R}^d}\int_{t(B_\delta)}\int_A N\{t; \mu_n(v), I(\theta_0)^{-1}\}g_n(t',v)^{(\rm norm)}\, dtdt'dv\right| = o_p(1). \tag{5}$$

The numerator of the leading term of (5) is in the following form,

$$\int_{\mathbb{R}^d}\int_{t(B_\delta)}\int_A N\{t; c_\varepsilon\beta_0 v + x_3, I(\theta_0)^{-1}\}N\{Ds(\theta_0)t'; x_1 v + x_2, A(\theta_0)\}K(v)\, dtdt'dv,$$

where $x_1 \in \mathbb{R}$, $x_2 \in \mathbb{R}^d$ and $x_3 \in \mathbb{R}^p$. This is continuous by Lemma 9 in the Supplementary Material. Then since $E_{G_n}(v) \to E_G(v)$ in distribution as $n \to \infty$ by Lemma 2, we have

$$\begin{aligned}&\int_{\mathbb{R}^d}\int_{t(B_\delta)}\int_A N\{t; \mu_n(v), I(\theta_0)^{-1}\}g_n(t',v)\, dtdt'dv\\ &\quad\to \int_{\mathbb{R}^d}\int_{\mathbb{R}^p}\int_A N[t; c_\varepsilon\beta_0\{v - E_G(v)\}, I(\theta_0)^{-1}]g(t',v)\, dtdt'dv,\end{aligned}$$

in distribution as $n \to \infty$. Putting the above results together, it holds that

$$\widetilde{\Pi}_\varepsilon\{a_n(\theta - \widetilde{\theta}_\varepsilon) \in A \mid s_{\rm obs}\} \to \int_A\int_{\mathbb{R}^p} N[t; c_\varepsilon\beta_0\{v - E_G(v)\}, I(\theta_0)^{-1}]G(v)^{(\rm norm)}\, dvdt,$$

in distribution, and statement (ii) of the proposition holds.

When $c_\varepsilon = 0$, since $\mu_n(v)$ does not depend on v, (5) becomes

$$\sup_{A\in\mathscr{B}^p}\left|\widetilde{\Pi}_\varepsilon\{a_n(\theta - \widetilde{\theta}_\varepsilon) \in A \mid s_{\rm obs}\} - \int_A N\{t; -a_n\varepsilon_n\beta_0 E_{G_n}(v) - a_n r_{n,1}, I(\theta_0)^{-1}\}\, dt\right| = o_p(1),$$

and by the continuous mapping theorem (van der Vaart, 2000),

$$\int_{\mathbb{R}^p}\left|N\{t; -a_n\varepsilon_n\beta_0 E_{G_n}(v) - a_n r_{n,1}, I(\theta_0)^{-1}\} - N\{t; 0, I(\theta_0)^{-1}\}\right|\, dt = o_p(1).$$

Therefore $\sup_{A\in\mathscr{B}^p}\left|\widetilde{\Pi}_\varepsilon\{a_n(\theta - \widetilde{\theta}_\varepsilon) \in A \mid s_{\rm obs}\} - \int_A N\{t; 0, I(\theta_0)^{-1}\}\, dt\right| = o_p(1)$, and statement (i) of the proposition holds.

When $c_\varepsilon = \infty$, Lemma 1 cannot be applied to the posterior distribution within (4) directly. With transformation $v'' = a_n\varepsilon_n v$, $\widetilde{\Pi}_\varepsilon\{\theta \in A \mid s_{\text{obs}}\}$ equals,

$$\int_{\mathbb{R}^d}\int_{t(B_\delta)} \widetilde{\Pi}(\theta \in A \mid s_{\text{obs}} + a_n^{-1}v'')\widetilde{\pi}_{\varepsilon,tv}(t', a_n^{-1}\varepsilon_n^{-1}v'')^{(\text{norm})}\, dt'dv'',$$

which implies that

$$\widetilde{\Pi}_\varepsilon\{\varepsilon_n^{-1}(\theta - \widetilde{\theta}_\varepsilon) \in A \mid s_{\text{obs}}\} =$$
$$\int_{\mathbb{R}^d}\int_{t(B_\delta)} \widetilde{\Pi}\{a_n(\theta-\theta_0) \in a_n\varepsilon_n A + a_n(\widetilde{\theta}_\varepsilon - \theta_0) \mid s_{\text{obs}} + a_n^{-1}v''\}\widetilde{\pi}_{\varepsilon,tv}(t', a_n^{-1}\varepsilon_n^{-1}v'')^{(\text{norm})}\, dt'dv''.$$

Then Lemma 1 can be applied. Using Lemma 2 and transforming v'' back to v we have

$$\sup_{A\in\mathscr{B}^p}\Big|\widetilde{\Pi}_\varepsilon\{\varepsilon_n^{-1}(\theta - \widetilde{\theta}_\varepsilon) \in A \mid s_{\text{obs}}\}$$
$$- \int_{\mathbb{R}^d}\int_{t(B_\delta)}\int_A (a_n\varepsilon_n)^p N\{a_n\varepsilon_n t; \mu_n'(v), I(\theta_0)^{-1}\}g_n(t', v)^{(\text{norm})}\, dtdt'dv\Big| = o_p(1), \qquad (6)$$

where $\mu_n'(v) = \beta_0\{A(\theta_0)^{1/2}W_{\text{obs}} + a_n\varepsilon_n v\} - a_n(\widetilde{\theta}_\varepsilon - \theta_0)$.

Let $t''(t,t') = a_n\varepsilon_n(t - t') + a_n(\widetilde{\theta}_\varepsilon - \theta_0)$ and $t''(A, B_\delta)$ be the set $\{t''(t,t') : t \in A, t' \in t(B_\delta)\}$. With transformations $v' = v'(v,t')$ and $t'' = t''(t,t')$, since $\beta_0 Ds(\theta_0) = I_p$, we have

$$\int_{\mathbb{R}^d}\int_{t(B_\delta)}\int_A (a_n\varepsilon_n)^p N\{a_n\varepsilon_n t; \mu_n'(v), I(\theta_0)^{-1}\}g_n(t', v)\, dtdt'dv$$
$$= \int_{\mathbb{R}^d}\int_{t(B_\delta)}\int_A (a_n\varepsilon_n)^p N\{a_n\varepsilon_n(t-t'); \beta_0 v' - a_n(\widetilde{\theta}_\varepsilon - \theta_0), I(\theta_0)^{-1}\}g_n'(t', v')\, dtdt'dv'$$
$$= \int_{\mathbb{R}^d}\int_{t''(A,B_\delta)}\int_A N\{t''; \beta_0 v' - a_n(\widetilde{\theta}_\varepsilon - \theta_0), I(\theta_0)^{-1}\}g_n'\left[t - \frac{1}{a_n\varepsilon_n}\{t'' - a_n(\widetilde{\theta}_\varepsilon - \theta_0)\}, v'\right] dtdt''dv'.$$

The idea now is that as $n\to\infty$, $a_n\varepsilon_n \to \infty$, so the g_n' term in the integral will tend to $g_n'(t, v')$. Then by integrating first with respect to t'' and then with respect to v, we get the required result.

To make this argument rigorous, consider the following function,

$$\int_{\mathbb{R}^d}\int_{t''(A,B_\delta)}\int_A N\{t''; \beta_0 v', I(\theta_0)^{-1}\}N\{v'; 0, A(\theta_0)\}$$
$$\times \left|K\{Ds(\theta_0)t + x_1 v' - x_2 t'' + x_3\} - K\{Ds(\theta_0)t\}\right|\, dtdt''dv',$$

where $x_1 \in \mathbb{R}$, $x_2 \in \mathbb{R}$ and $x_3 \in \mathbb{R}^d$. This is continuous by Lemma 9 in the Supplementary Material, so by the continuous mapping theorem,

$$\sup_{A\in\mathscr{B}^p}\left|\int_{\mathbb{R}^d}\int_{t''(A,B_\delta)}\int_A N\{t''; \beta_0 v', I(\theta_0)^{-1}\}g_n'\left(t - \frac{1}{a_n\varepsilon_n}t'', v'\right) dtdt''dv' - \int_A K\{Ds(\theta_0)t\}\, dt\right| = o_p(1).$$

Then using Lemma 2,

$$\sup_{A\in\mathscr{B}^p}\left|\widetilde{\Pi}_\varepsilon\{\varepsilon_n^{-1}(\theta - \widetilde{\theta}_\varepsilon) \in A \mid s_{\text{obs}}\} - \int_A K\{Ds(\theta_0)t\}\, dt \Big/ \int_{\mathbb{R}^p} K\{Ds(\theta_0)t\}\, dt\right| = o_p(1).$$

Therefore statement (iii) of the proposition holds.

Proof of Result from Section 3·2

The intuition behind this result is that, as shown above, the joint posterior of (θ, s) under approximate Bayesian computation can be viewed as a marginal distribution for s times a conditional for θ given s. The latter is just the true posterior for θ given s, and this posterior converges to a Gaussian limit that depends on s only through its mean. Regression adjustment works because it corrects for the dependence of this mean on s, so if we work with the regression adjusted parameter θ^* then the conditional distribution for θ^* given s will tend to a Gaussian limit whose mean is the same for all s.

The following lemma gives an expansion of the optimal linear coefficient matrix β_ε. The order of the remainder leads to successful removal of the first order bias in the posterior distribution of approximate Bayesian computation.

LEMMA 4. *Assume Conditions 1-6. Then if $\varepsilon_n = o(a_n^{-3/5})$, $a_n\varepsilon_n(\beta_\varepsilon - \beta_0) = o_p(1)$.*

The following lemma, similar to Lemma 3, says that the approximate Bayesian computation posterior distribution of θ^* is asymptotically the same as $\widetilde{\Pi}_\varepsilon$. Recall that θ_ε^* is the mean of $\pi_\varepsilon^*(\theta^* \mid s_{\text{obs}})$. Let $\widetilde{\pi}_\varepsilon^*(\theta^* \mid s_{\text{obs}}) = \int_{\mathbb{R}^d} \widetilde{\pi}_\varepsilon\{\theta^* + \beta_\varepsilon(s - s_{\text{obs}}), s \mid s_{\text{obs}}\}\, ds$ and $\widetilde{\theta}_\varepsilon^*$ be the mean of $\widetilde{\pi}_\varepsilon^*(\theta^* \mid s_{\text{obs}})$.

LEMMA 5. *Assume Conditions 1-6. If $\varepsilon_n = o(a_n^{-3/5})$, then*
(a) for any $\delta < \delta_0$, $\Pi_\varepsilon(\theta^* \in B_\delta^c \mid s_{\text{obs}})$ and $\widetilde{\Pi}_\varepsilon(\theta^* \in B_\delta^c \mid s_{\text{obs}})$ are $o_p(1)$;
(b) there exists $\delta < \delta_0$ such that $\sup_{A\in\mathscr{B}^p}\left|\Pi_\varepsilon(\theta^* \in A\cap B_\delta \mid s_{\text{obs}}) - \widetilde{\Pi}_\varepsilon(\theta^* \in A\cap B_\delta \mid s_{\text{obs}})\right| = o_p(1)$;
(c) $a_n(\theta_\varepsilon^* - \widetilde{\theta}_\varepsilon^*) = o_p(1)$, and $\widetilde{\theta}_\varepsilon^* = \theta_0 + a_n^{-1}\beta_0 A(\theta_0)^{1/2}W_{\text{obs}} + \varepsilon_n(\beta_0 - \beta_\varepsilon)E_{G_n}(v) + r_{n,2}$ where the remainder $r_{n,2} = o_p(a_n^{-1})$.

Proof of Theorem 1. Similar to the proof of Proposition 1, it is sufficient to only consider the convergence of $\widetilde{\Pi}_\varepsilon$ of the properly scaled and centered θ^*. Similar to (4)

$$\widetilde{\Pi}_\varepsilon(\theta^* \in A \mid s_{\text{obs}}) = \begin{cases} \int_{\mathbb{R}^d}\int_{t(B_\delta)} \widetilde{\Pi}(\theta \in A + \varepsilon_n\beta_\varepsilon v \mid s_{\text{obs}} + \varepsilon_n v)\widetilde{\pi}_{\varepsilon,tv}(t', v)^{(\text{norm})}\, dt'dv, & c_\varepsilon < \infty, \\ \int_{\mathbb{R}^d}\int_{t(B_\delta)} \widetilde{\Pi}(\theta \in A + \varepsilon_n\beta_\varepsilon v \mid s_{\text{obs}} + a_n^{-1}v)\widetilde{\pi}_{\varepsilon,tv}(t', a_n^{-1}\varepsilon_n^{-1}v)^{(\text{norm})}\, dt'dv, & c_\varepsilon = \infty. \end{cases}$$

Similar to (5), by Lemma 1 and Lemma 5(c), we have

$$\sup_{A\in\mathscr{B}^p}\left|\widetilde{\Pi}_\varepsilon\{a_n(\theta^* - \widetilde{\theta}_\varepsilon^*) \in A \mid s_{\text{obs}}\} - \int_{\mathbb{R}^d}\int_{t(B_\delta)}\int_A N\{t; \mu_n^*(v), I(\theta_0)^{-1}\}g_n(t', v)^{(\text{norm})}\, dtdt'dv\right| = o_p(1), \tag{7}$$

where $\mu_n^*(v) = \{a_n\varepsilon_n(\beta_0 - \beta_\varepsilon) + (c_\varepsilon - a_n\varepsilon_n)\beta_0\mathbb{1}_{c_\varepsilon<\infty}\}\{v - E_{G_n}(v)\} + a_n r_{n,2}$. Since $a_n\varepsilon_n(\beta_\varepsilon - \beta_0) = o_p(1)$, by Lemma 9 in the Supplementary Material and the continuous mapping theorem,

$$\int_{\mathbb{R}^d}\int_{t(B_\delta)}\int_{\mathbb{R}^p}\left|N\{t; \mu_n^*(v), I(\theta_0)^{-1}\} - N\{t; 0, I(\theta_0)^{-1}\}\right| g_n(t', v)\, dtdt'dv = o_p(1).$$

Then we have

$$\sup_{A\in\mathscr{B}^p}\left|\widetilde{\Pi}_\varepsilon\{a_n(\theta^* - \widetilde{\theta}_\varepsilon^*) \in A \mid s_{\text{obs}}\} - \int_A N\{t; 0, I(\theta_0)^{-1}\}\, dt\right| = o_p(1),$$

and the first convergence in the theorem holds.

The second convergence in the theorem holds by Lemma 5(c).

Since the only requirement for β_ε in the above is $a_n\varepsilon_n(\beta_\varepsilon' - \beta_\varepsilon) = o_p(1)$, the above arguments will hold if β_ε is replaced by a $p \times d$ matrix $\widehat{\beta}_\varepsilon$ satisfying $a_n\varepsilon_n(\widehat{\beta}_\varepsilon - \beta_\varepsilon) = o_p(1)$. □

Proof of Proposition 3. Consider θ^* where β_ε is replaced by $\widehat{\beta}_\varepsilon$. Let $\eta_n = N^{1/2}a_n\varepsilon_n(\widehat{\beta}_\varepsilon - \beta_\varepsilon)$. Since $\widehat{\beta}_\varepsilon - \beta_\varepsilon = O_p\{(a_n\varepsilon_n)^{-1}N^{-1/2}\}$ as $n \to \infty$, $\eta_n = O_p(1)$ as $n \to \infty$ and let its limit be η. In this case, if we replace $\mu_n^*(v)$ in (7) with $\mu_n^*(v) + N^{-1/2}\eta_n\{v - E_{G_n}(v)\}$, denoted by $\widehat{\mu}_n^*(v)$, the equation still holds. Denote this equation by $(7')$. Limits of the leading term in $(7')$ can be obtained by arguments similar as those for (5).

When $c_\varepsilon < \infty$, since for fixed v, $\widehat{\mu}_n^*(v)$ converges to $N^{-1/2}\eta\{v - E_G(v)\}$ in distribution, by following the same line we have

$$\widetilde{\Pi}_\varepsilon\{a_n(\theta^* - \widetilde{\theta}_\varepsilon^*) \in A \mid s_{\text{obs}}\} \to \int_A \int_{\mathbb{R}^p} N[t; N^{-1/2}\eta\{v - E_G(v)\}, I(\theta_0)^{-1}]G(v)^{(\text{norm})}\, dvdt,$$

in distribution.

When $c_\varepsilon = \infty$, by Lemma 2 we have $E_{G_n}(v) \to 0$ in probability, and $\int_{\mathbb{R}^d}\int_{t(B_\delta)} g_n(t', v)\, dt'dv \to \int_{\mathbb{R}^p} K\{Ds(\theta_0)t\}\, dt$ in probability. Then with transformation $v' = v'(v, t')$, for fixed v,

$$\begin{aligned}\widehat{\mu}_n^*(v) &= \{a_n\varepsilon_n(\beta_0 - \beta_\varepsilon) + N^{-1/2}\eta_n\}\left\{Ds(\theta_0)t' + \frac{1}{a_n\varepsilon_n}v' - \frac{1}{a_n\varepsilon_n}A(\theta_0)^{1/2}W_{\text{obs}} - E_{G_n(v)}\right\}\\ &\to N^{-1/2}\eta Ds(\theta_0)t',\end{aligned}$$

in distribution. Recall that $g_n(t', v)\, dv = g_n'(t', v')\, dv'$. Then by Lemma 9 in the Supplementary Material and the continuous mapping theorem,

$$\begin{aligned}&\int_{\mathbb{R}^d}\int_{t(B_\delta)}\int_A N\{t; \widehat{\mu}_n^*(v), I(\theta_0)^{-1}\}g_n'(t', v')\, dtdt'dv\\ \to &\int_{\mathbb{R}^d}\int_{t(B_\delta)}\int_A N\{t; N^{-1/2}\eta Ds(\theta_0)t', I(\theta_0)^{-1}\}K\{Ds(\theta_0)t'\}\, dtdt'dv,\end{aligned}$$

in distribution. Therefore by $(7')$ and the above convergence results,

$$\widetilde{\Pi}_\varepsilon\{a_n(\theta^* - \widetilde{\theta}_\varepsilon^*) \in A \mid s_{\text{obs}}\} \to \int_A \int_{\mathbb{R}^p} N\{t; N^{-1/2}\eta Ds(\theta_0)t', I(\theta_0)^{-1}\}K\{Ds(\theta_0)t'\}^{(\text{norm})}\, dt'dt,$$

in distribution. □

References

Beaumont, M. A. (2010). Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics* **41**, 379–406.

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.

Blum, M. G. (2010). Approximate Bayesian computation: a nonparametric perspective. *Journal of the American Statistical Association* **105**, 1178–1187.

Bonassi, F. V. & West, M. (2015). Sequential Monte Carlo with adaptive weights for approximate Bayesian computation. *Bayesian Analysis* **10**, 171–187.

Calvet, L. E. & Czellar, V. (2015). Accurate methods for approximate Bayesian computation filtering. *Journal of Financial Econometrics* **13**, 798–838.

CHERNOZHUKOV, V. & HONG, H. (2003). An MCMC approach to classical estimation. *Journal of Econometrics* **115**, 293–346.

CORNUET, J.-M., SANTOS, F., BEAUMONT, M. A., ROBERT, C. P., MARIN, J.-M., BALDING, D. J., GUILLEMAUD, T. & ESTOUP, A. (2008). Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* **24**, 2713–2719.

DROVANDI, C. C. & PETTITT, A. N. (2011). Estimation of parameters for macroparasite population evolution using approximate Bayesian computation. *Biometrics* **67**, 225–233.

DROVANDI, C. C., PETTITT, A. N. & LEE, A. (2015). Bayesian indirect inference using a parametric auxiliary model. *Statistical Science* **30**, 72–95.

DUFFIE, D. & SINGLETON, K. J. (1993). Simulated moments estimation of Markov models of asset prices. *Econometrica* **61**, 929–952.

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

FILIPPI, S., BARNES, C. P., CORNEBISE, J. & STUMPF, M. P. (2013). On optimality of kernels for approximate Bayesian computation using sequential Monte Carlo. *Statistical Applications in Genetics and Molecular Biology* **12**, 87–107.

FRAZIER, D. T., MARTIN, G. M., ROBERT, C. P. & ROUSSEAU, J. (2016). Asymptotic Properties of Approximate Bayesian Computation. *arXiv.1607.06903* .

GOURIÉROUX, C. & RONCHETTI, E. (1993). Indirect inference. *Journal of Applied Econometrics* **8**, s85–s118.

HEGGLAND, K. & FRIGESSI, A. (2004). Estimating functions in indirect inference. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **66**, 447–462.

KLEIJN, B. & VAN DER VAART, A. (2012). The Bernstein-von-Mises theorem under misspecification. *Electronic Journal of Statistics* **6**, 354–381.

LENORMAND, M., JABOT, F. & DEFFUANT, G. (2013). Adaptive approximate Bayesian computation for complex models. *Computational Statistics* **28**, 2777–2796.

LI, W. & FEARNHEAD, P. (2015). On the asymptotic efficiency of approximate Bayesian computation estimators. *arxiv.1506.03481* .

MARIN, J.-M., PILLAI, N. S., ROBERT, C. P. & ROUSSEAU, J. (2014). Relevant statistics for Bayesian model choice. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76**, 833–859.

MARJORAM, P., MOLITOR, J., PLAGNOL, V. & TAVARE, S. (2003). Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences* **100**, 15324–15328.

PAULI, F., RACUGNO, W. & VENTURA, L. (2011). Bayesian composite marginal likelihoods. *Statistica Sinica* **21**, 149–164.

RIBATET, M., COOLEY, D. & DAVISON, A. C. (2012). Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica* **22**, 813–845.

RULI, E., SARTORI, N. & VENTURA, L. (2016). Approximate Bayesian computation with composite score functions. *Statistics and Computing* **26**, 679–692.

SOUBEYRAND, S. & HAON-LASPORTES, E. (2015). Weak convergence of posteriors conditional on maximum pseudo-likelihood estimates and implications in ABC. *Statistics & Probability Letters* **107**, 84–92.

TONI, T., WELCH, D., STRELKOWA, N., IPSEN, A. & STUMPF, M. P. (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface* **6**, 187–202.

VAN DER VAART, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press.

VARIN, C., REID, N. & FIRTH, D. (2011). An overview of composite likelihood methods. *Statistica Sinica* **21**, 5–42.

WEGMANN, D., LEUENBERGER, C. & EXCOFFIER, L. (2009). Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics* **182**, 1207–1218.

WOOD, S. N. (2010). Statistical inference for noisy nonlinear ecological dynamic systems. *Nature* **466**, 1102–1104.