Improved Convergence of Regression Adjusted Approximate Bayesian Computation

Wentao Li and Paul Fearnhead
Lancaster University

Abstract

We present a number of surprisingly strong asymptotic results for the regression-adjusted version of Approximate Bayesian Computation (ABC) introduced by Beaumont et al. (2002). We show that for an appropriate choice of the bandwidth in ABC, using regression-adjustment will lead to an ABC posterior that, asymptotically, correctly quantifies uncertainty. Furthermore, for such a choice of bandwidth we can implement an importance sampling algorithm to sample from the ABC posterior whose acceptance probability tends to 1 as we increase the data sample size. This compares favorably to results for standard ABC, where the only way to obtain an ABC posterior that correctly quantifies uncertainty is to choose a much smaller bandwidth, for which the acceptance probability tends to 0 and hence for which Monte Carlo error will dominate.

Keywords: Approximate Bayesian computation; Asymptotics; Importance Sampling; Partial Information; Local-linear Regression.

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 prohibits standard implementation of likelihood-based methods, such as maximum likelihood estimation or a 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 the ability to simulate 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 the actual data.

Examples of such likelihood-free methods include simulate methods of moments (Duffie and Singleton, 1993), indirect inference (Gouriéroux and Ronchetti, 1993; Heggland and Frigessi, 2004), synthetic likelihood (Wood, 2010) and approximate Bayesian computation (Beaumont et al., 2002). Of these, approximate Bayesian computation (ABC) methods are arguably the most common methods for performing Bayesian inference. For a number of years, ABC methods have been popular in population genetics (Beaumont et al., 2002; Cornuet et al., 2008, e.g.), 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 and Czellar, 2015) and epidemiology (Drovandi and Pettitt, 2011).

The idea of ABC is to first summarise the data using low-dimensional summary statistics. The posterior density given the summary statistics is then approximated as follows. Assume the data is $\boldsymbol{Y}_{obs} = (y_{obs,1}, \cdots, y_{obs,n})$ and modelled as a draw from a parametric density $f_n(\boldsymbol{y}|\boldsymbol{\theta})$

with the parameter space $\mathbb{R}^p$. Let $K(\boldsymbol{x})$ be a kernel density, where $\max_{\boldsymbol{x}} K(\boldsymbol{x}) = 1$, and $\varepsilon > 0$ be the bandwidth. After choosing a d-dimension summary statistic $\boldsymbol{s}_n(\boldsymbol{Y})$, define a joint density on $(\boldsymbol{\theta}, \boldsymbol{s})$ as

$$\pi_\varepsilon(\boldsymbol{\theta}, \boldsymbol{s}|\boldsymbol{s}_{obs}) \triangleq \frac{\pi(\boldsymbol{\theta}) f_n(\boldsymbol{s}|\boldsymbol{\theta}) K(\varepsilon^{-1}(\boldsymbol{s} - \boldsymbol{s}_{obs}))}{\int_{\mathbb{R}^p \times \mathbb{R}^d} \pi(\boldsymbol{\theta}) f_n(\boldsymbol{s}|\boldsymbol{\theta}) K(\varepsilon^{-1}(\boldsymbol{s} - \boldsymbol{s}_{obs}))\, d\boldsymbol{\theta} d\boldsymbol{s}}, \tag{1}$$

where $\boldsymbol{s}_{obs} = \boldsymbol{s}_n(\boldsymbol{Y}_{obs})$. Our ABC approximation to the posterior is then the marginal of this joint density

$$\pi_\varepsilon(\boldsymbol{\theta}|\boldsymbol{s}_{obs}) = \int \pi_\varepsilon(\boldsymbol{\theta}, \boldsymbol{s}|\boldsymbol{s}_{obs}) \mathrm{d}\boldsymbol{s}.$$

We call $\pi_\varepsilon(\boldsymbol{\theta}|\boldsymbol{s}_{obs})$ the ABC posterior density.

The idea behind this way of approximating the posterior is that we can simulate from the ABC posterior density without needing to evaluate the likelihood function $f_n(\boldsymbol{s}|\boldsymbol{\theta})$. The simplest way of sampling from the ABC posterior is via rejection sampling (Beaumont et al., 2002). This proceeds by simulating a parameter value and an associated summary statistic from $\pi(\boldsymbol{\theta}) f_n(\boldsymbol{s}|\boldsymbol{\theta})$. This pair is then accepted with probability $K(\varepsilon^{-1}(\boldsymbol{s} - \boldsymbol{s}_{obs}))$. The accepted pairs will be draws from (1), and the accepted parameter values will be draws from the ABC posterior. Implementing this rejection sampler only needs the ability to simulate pseudo data sets from the model, and then to be able to calculate the summary statistics for those data sets.

There are a number of different algorithms for simulating from the ABC posterior. These include adaptive or sequential importance sampling (Beaumont et al., 2009; Bonassi et al., 2015; Lenormand et al., 2013; Filippi et al., 2013) and MCMC approaches (Marjoram et al., 2003; Wegmann et al., 2009; Meeds et al., 2015). 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 still limited by the probability of acceptance for data simulated with a parameter value that has high posterior probability. This has led to recent attempts to try and "optimise" the pseudo data sets simulated to substantially increase this acceptance probability (Meeds and Welling, 2015; Forneron and Ng, 2015).

This paper is concerned with the asymptotic properties of ABC. This is currently an active area of research, with a number of recent papers (Li and Fearnhead, 2015; Frazier et al., 2016; Zhong and Ghosh, 2016) presenting results on the asymptotic behaviour of the ABC posterior and the ABC posterior mean as the amount of data, n, increases. These results highlight the tension in ABC between choices of the summary statistics and bandwidth that will lead to more accurate inferences when using the ABC posterior, against choices that will reduce the computational cost or Monte Carlo error of algorithms for sampling from the ABC posterior.

An informal summary of some of these results is as follows. Assume we have a fixed dimensional summary statistic and assume that the posterior variance given this summary decreases like $1/n$ as n increases. The theoretical results compare the ABC posterior, or ABC posterior mean, to the posterior or posterior mean given the summary of the data. The accuracy of using ABC is then governed by the choice of bandwidth, and this choice should depend on n. Li and Fearnhead (2015) show that the optimal choice of this bandwidth will be $O(1/\sqrt{n})$. With this choice, estimates based on the ABC posterior mean will, asymptotically, be as accurate as estimates based on the 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. This constant factor is of the form $1 + O(1/N)$ where N are the number of pseudo data sets that are simulated. As such these results are similar to the asymptotic results of indirect inference (Gouriéroux and Ronchetti, 1993; Heggland and Frigessi, 2004). By comparison choosing a bandwidth which is $o(1/\sqrt{n})$ will lead to an acceptance probability that tends to 0 as $n \to \infty$, and the Monte Carlo error of ABC will blow-up. Choosing a bandwidth that is much bigger that $O(1/\sqrt{n})$ will also lead to a regime where the Monte Carlo error dominates, and can also lead to a non-negligible bias in the ABC posterior mean that also inflates the error.

Whilst the above results for using a bandwidth that is $O(1/\sqrt{n})$ are positive in terms of point estimates, they come with a negative result in terms of the calibration of the ABC posterior. With such a choice of bandwidth we will always have that the ABC posterior over-inflates the uncertainty in the parameter (see Proposition 3.1 below and Frazier et al., 2016). The aim of this paper is to show that a variation of ABC can lead to inference that is not only accurate in terms of point estimation, but also calibrated in the sense that the ABC posterior variance is the same as the (frequentist) asymptotic variance of the posterior mean. Furthermore this can be achieved in a way that the computational efficiency of ABC improves as we get more data, in the sense that the acceptance probability of a good ABC algorithm will tend to 1 as $n \to \infty$.

The variation of ABC that gives these surprisingly strong asymptotic results is the local-linear regression correction of Beaumont et al. (2002) (see also Marin et al., 2012). The idea is to simulate N pairs $(\boldsymbol{\theta}_i, \boldsymbol{s}_i)$ from the joint distribution (1). We then fit a linear model that predicts the parameter, $\boldsymbol{\theta}$, from the summary, $\boldsymbol{s}$. If we let $\widehat{\boldsymbol{\beta}}_\varepsilon$ denote the estimate regression coefficients from this linear model, then we replace our sampled parameter values with

$$\boldsymbol{\theta}_i - \widehat{\boldsymbol{\beta}}_\varepsilon(\boldsymbol{s}_i - \boldsymbol{s}_{obs}), \text{ for } i = 1, \ldots, N.$$

The intuition is that this will correct for any systemic bias in the $\boldsymbol{\theta}_i$ values caused by the discrepancy between $\boldsymbol{s}_i$ and $\boldsymbol{s}_{obs}$. It is expected to work well when the conditional mean $E[\boldsymbol{\theta}|\boldsymbol{s}]$ is close to a linear function (Beaumont et al., 2002) (see Blum and François, 2010, for a nonlinear extension). In Blum (2010) it is shown that the adjustment reduces the mean square error of posterior density estimation if we assume that the conditional mean $E[\boldsymbol{\theta}|\boldsymbol{s}]$ is linear and the residual $\boldsymbol{\theta} - E[\boldsymbol{\theta}|\boldsymbol{s}]$ does not depend on $\boldsymbol{s}$. In Nott et al. (2014), it is found that the first two moments of the regression adjusted sample can be formulated as the adjusted mean and variance in Bayes linear analysis (Goldstein and Wooff, 2007), and hence the adjusted ABC variance is smaller than the non-adjusted one. Our results are the first to look at the asymptotic properties of this method. They show that it has excellent asymptotic properties without requiring strong assumptions about the linearity of $E[\boldsymbol{\theta}|\boldsymbol{s}]$ or the homoscedasticity of the residuals. Furthermore they show that the reduction in the ABC posterior variance is by exactly the correct amount that the resulting ABC posterior distribution correctly captures the uncertainty in the parameter.

2 Notation and Set-up

As described above, we denote the data by $\boldsymbol{Y}_{obs} = (y_{obs,1}, \cdots, y_{obs,n})$, where n is the sample size, and each observation, $y_{obs,i}$, can be of arbitrary dimension. Assume the data is modelled as a draw from a parametric density, $f_n(\boldsymbol{y}|\boldsymbol{\theta})$, and we will be considering the asymptotics as $n \to \infty$. This density depends on an unknown parameter $\boldsymbol{\theta} \in \mathbb{R}^p$. Let $\mathscr{B}^p$ be the Borel sigma-field on $\mathbb{R}^d$. We will let $\boldsymbol{\theta}_0$ denote the true parameter value, and $\pi(\boldsymbol{\theta})$ the prior distribution for the parameter. Denote the support of $\pi(\boldsymbol{\theta})$ by $\mathcal{P}$. Assume a fixed dimension summary statistic $\boldsymbol{s}(\boldsymbol{Y})$ is chosen and its density, implied by $f_n(\boldsymbol{y}|\boldsymbol{\theta})$, is $f_n(\boldsymbol{s}|\boldsymbol{\theta})$. The shorthand $\boldsymbol{S}_n$ is used to denote the random variable with density $f_n(\boldsymbol{s}|\boldsymbol{\theta})$. For a set A, let A^c be its complement with respect to the whole space. For a series x_n, besides the limit notations $O(\cdot)$ and $o(\cdot)$, we use the notations, as $n \to \infty$, that $x_n = \Theta(\cdot)$ if there exists constants m and M such that $0 < m < |x_n/a_n| < M < \infty$.

The conditions of the theoretical results are stated below. (C1)-(C6) are from Li and Fearnhead (2015). (C2) states the requirements for the kernel function and is satisfied by all commonly used kernels in ABC. (C3) assumes a central limit theorem for the summary statistic with rate a_n. (C4) and (C5) assume that for $f_n(\boldsymbol{s}|\boldsymbol{\theta})$ as a density of $\boldsymbol{s}$, its tails are exponentially decreasing with uniform rates for $\boldsymbol{\theta} \in \mathcal{P}$. (C4) additionally assumes that in a neighborhood of $\boldsymbol{\theta}_0$, $f_n(\boldsymbol{s}|\boldsymbol{\theta})$ deviates from the leading term of its Edgeworth expansion by the rate $a_n^{-2/5}$, which is weaker than the standard requirement, $o(a_n^{-1})$, of the remainder from Edgeworth expansion.

(C1) There exists some $\delta_0 > 0$, such that $\mathcal{P}_0 \equiv \{\boldsymbol{\theta} : |\boldsymbol{\theta} - \boldsymbol{\theta}_0| < \delta_0\} \subset \mathcal{P}$.

(C2) (i) $\int \boldsymbol{v} K(\boldsymbol{v})\, d\boldsymbol{v} = 0$.

(ii) $\int \prod_{k=1}^{l} v_{i_k} K(\boldsymbol{v})\, d\boldsymbol{v} < \infty$ for any coordinates $(v_{i_1}, \cdots, v_{i_l})$ of $\boldsymbol{v}$ and $l \leq p + 6$.

(iii) $K(\boldsymbol{v}) = K(\|\boldsymbol{v}\|_\Lambda^2)$ where $\|\boldsymbol{v}\|_\Lambda^2 = \boldsymbol{v}^T \Lambda \boldsymbol{v}$ and Λ is a diagonal matrix, and $K(\boldsymbol{v})$ is a decreasing function of $\|\boldsymbol{v}\|_\Lambda$.

(iv) $K(\boldsymbol{v}) = O(e^{-c_1 \|\boldsymbol{v}\|^{\alpha_1}})$ for some $\alpha_1 > 0$ and $c_1 > 0$ as $\|\boldsymbol{v}\| \to \infty$.

(C3) There exists a sequence a_n, with $a_n \to \infty$ as $n \to \infty$, a d-dimensional vector $\boldsymbol{s}(\boldsymbol{\theta})$ and a $d \times d$ matrix $A(\boldsymbol{\theta})$, such that for all $\boldsymbol{\theta} \in \mathcal{P}_0$,

$$a_n(\boldsymbol{S}_n - \boldsymbol{s}(\boldsymbol{\theta})) \xrightarrow{\mathfrak{L}} N(0, A(\boldsymbol{\theta}));\ \text{as } n \to \infty.$$

Furthermore, that

(i) $\boldsymbol{s}(\boldsymbol{\theta})$ and $A(\boldsymbol{\theta}) \in C^1(\mathcal{P}_0)$, and $A(\boldsymbol{\theta})$ is positive definite for any $\boldsymbol{\theta}$;

(ii) $\boldsymbol{s}(\boldsymbol{\theta}) = \boldsymbol{s}(\boldsymbol{\theta}_0)$ if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$; and

(iii) $I(\boldsymbol{\theta}) \triangleq D\boldsymbol{s}(\boldsymbol{\theta})^T A^{-1}(\boldsymbol{\theta}) D\boldsymbol{s}(\boldsymbol{\theta})$ has full rank at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Define $\widetilde{f}_n(\boldsymbol{s}|\boldsymbol{\theta}) = N(\boldsymbol{s}; \boldsymbol{s}(\boldsymbol{\theta}), A(\boldsymbol{\theta})/a_n^2)$ and the standardization $W_n(\boldsymbol{s}) = a_n A(\boldsymbol{\theta})^{-1/2}(\boldsymbol{s} - \boldsymbol{s}(\boldsymbol{\theta}))$.

(C4) There exists α_n satisfying $\alpha_n / a_n^{2/5} \to \infty$ and a density $r_{max}(\boldsymbol{w})$ satisfying (C2) (ii)-(iv), such that $\sup_{\boldsymbol{\theta} \in \mathcal{P}_0} \alpha_n |f_{W_n}(\boldsymbol{w}|\boldsymbol{\theta}) - \widetilde{f}_{W_n}(\boldsymbol{w}|\boldsymbol{\theta})| \leq c_3 r_{max}(\boldsymbol{w})$ for some positive constant c_3.

(C5) $\sup_{\boldsymbol{\theta} \in \mathcal{P}_0^c} f_{W_n}(\boldsymbol{w}|\boldsymbol{\theta}) = O(e^{-c_2 \|\boldsymbol{w}\|^{\alpha_2}})$ as $\|\boldsymbol{w}\| \to \infty$ for some positive constants c_2 and α_2, and $A(\boldsymbol{\theta})$ is bounded in $\mathcal{P}$.

(C6) $\pi(\boldsymbol{\theta}) \in C^2(\mathcal{P}_0)$ and $\pi(\boldsymbol{\theta}_0) > 0$.

Additionally for the results regarding regression adjustment, it is required that the first two moments of the summary statistic exist.

(C7) $\int_{\mathbb{R}^d} \boldsymbol{s} f_n(\boldsymbol{s}|\boldsymbol{\theta})\, d\boldsymbol{s}$ and $\int_{\mathbb{R}^d} \boldsymbol{s}\boldsymbol{s}^T f_n(\boldsymbol{s}|\boldsymbol{\theta})\, d\boldsymbol{s}$ exist.

3 Asymptotics of ABC

In this section the asymptotic properties of ABC posterior and its local-linear adjusted variant are given. In Section 3.1, it is shown that the ABC posterior has the same limit as the posterior given the summary only when $\varepsilon = o(a_n^{-1})$, and over-inflates the posterior uncertainty for other choices of ε. This comes from comparing a new result, Propostion 3.1 on the convergence of the ABC posterior, with a result from Li and Fearnhead (2015) which quantifies the frequentist uncertainty of the point estimates based on ABC posterior mean. Recently Frazier et al. (2016, Theorem 2) also discusses the limits of ABC posterior, and gives results similar to Proposition 3.1. The main difference is that in Frazier et al. (2016), the ABC posterior uncertainty of $\boldsymbol{s}(\boldsymbol{\theta})$ is considered and the convergence is in a weaker sense and based on a different set of assumptions.

In Section 3.2, it is found that the local-linear regression adjustment with the optimal coefficients corrects the overinflation in ABC posterior uncertainty when ε is not $o(a_n^{-1})$. This correction holds up to $\varepsilon = o(a_n^{-3/5})$. The impact of using a finite sample estimate of the optimal coefficients is also discussed. In Section 3.3, we further show that it is possible to develop algorithms to sample from the ABC posterior when $\varepsilon = o(a_n^{-3/5})$ for which the acceptance rate tends to 1 as $\rightarrow \infty$. Hence, if using a local-linear regression adjustment we observe a gain in Monte Carlo efficiency for larger data sets.

The proofs of all results can be found in the appendix.

3.1 ABC Posterior

First we consider the convergence of ABC posterior distribution, denoted by $\Pi_\varepsilon(\boldsymbol{\theta} \in A|\boldsymbol{s}_{obs})$ for $A \in \mathscr{B}^p$, as $n \rightarrow \infty$. Note that the distribution function is a random variable in $\mathbb{R}$ with the randomness due to $\boldsymbol{s}_{obs}$. Similar to the classical Bernstein von-Mises theorem (Bickel and Doksum, 2015), the convergence of the posterior contains two parts. One is the convergence of the ABC posterior uncertainty, measured by the distribution function of a properly scaled and centered version of $\boldsymbol{\theta}$, and this is given in Proposition 3.1. The other is the convergence of ABC posterior mean. The latter result comes from Li and Fearnhead (2015), but is repeated for convenience as Proposition 3.2.

The following proposition gives three different limiting forms for $\Pi_\varepsilon(\boldsymbol{\theta} \in A|\boldsymbol{s}_{obs})$. These correspond to different rates for how the bandwidth decreases with n relative to the rate, a_n, of the central limit theorem in (C3). We summarise these competing rates by defining $c_\varepsilon = \lim_{n\rightarrow\infty} a_n\varepsilon_n$, and assume that the limit exists. Then for different values of c_ε we get the following limiting results.

Proposition 3.1. *Assume conditions (C1)-(C6). Let $\boldsymbol{\theta}_\varepsilon$ denote the ABC posterior mean. If $\varepsilon_n = o(a_n^{-3/5})$ then the following convergence holds depending on the vale of c_ε:*

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

$$\sup_{A\in\mathscr{B}^p}\left|\Pi_\varepsilon(a_n(\boldsymbol{\theta}-\boldsymbol{\theta}_\varepsilon) \in A|\boldsymbol{s}_{obs}) - \int_A \psi(\boldsymbol{t})\,d\boldsymbol{t}\right| \xrightarrow{P} 0,$$

where

$$\psi(\boldsymbol{t}) \propto N(\boldsymbol{t}; 0, I(\boldsymbol{\theta}_0)^{-1}).$$

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

$$\Pi_\varepsilon(a_n(\boldsymbol{\theta}-\boldsymbol{\theta}_\varepsilon) \in A|\boldsymbol{s}_{obs}) \xrightarrow{\mathfrak{L}} \int_A \psi(\boldsymbol{t})\,d\boldsymbol{t},$$

where

$$\psi(\boldsymbol{t}) \propto \int_{\mathbb{R}^d} N(\boldsymbol{t}; c_\varepsilon R(\boldsymbol{v}-E_G[\boldsymbol{v}]), I(\boldsymbol{\theta}_0)^{-1})G(\boldsymbol{v}; c_\varepsilon, \boldsymbol{Z})\,d\boldsymbol{v},$$

R is a rank-p $d \times p$ constant matrix, $\boldsymbol{Z} \sim N(0, I_d)$ and $G(\boldsymbol{v}; c_\varepsilon, \boldsymbol{Z})$ is a random density of $\boldsymbol{v}$ depending on c_ε and $\boldsymbol{Z}$.

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

$$\sup_{A\in\mathscr{B}^p}\left|\Pi_\varepsilon(\varepsilon_n^{-1}(\boldsymbol{\theta}-\boldsymbol{\theta}_\varepsilon)\in A|\boldsymbol{s}_{obs}) - \int_A \psi(\boldsymbol{t})\,d\boldsymbol{t}\right| \xrightarrow{P} 0,$$

where

$$\psi(\boldsymbol{t}) \propto K(D\boldsymbol{s}(\boldsymbol{\theta}_0)\boldsymbol{t}).$$

The explicit forms of R and $G(\boldsymbol{v}; c_\varepsilon, \boldsymbol{Z})$ are stated in the appendix. The main difference between the three different convergence results is the form of the limiting density $\psi(\boldsymbol{t})$ for the scaled random variable $a_{n,\varepsilon}(\boldsymbol{\theta}-\boldsymbol{\theta}_\varepsilon)$, where $a_{n,\varepsilon} = a_n\mathbb{1}_{c_\varepsilon<\infty} + \varepsilon_n^{-1}\mathbb{1}_{c_\varepsilon=\infty}$. For case (1) the bandwidth is sufficiently small that the ABC approximation due to accepting summaries "close" to the observed summary is asymptotically negligible. In this case we get an asymptotic ABC posterior distribution that is Gaussian, and is the same as the asymptotic limit of the true posterior for θ given the summary. For case (3) the bandwidth is sufficiently big that this ABC approximation dominates and the asymptotic ABC posterior distribution is determined by the kernel. For case (2) we have that the approximation is of the same order as the uncertainty in θ which leads to an asymptotic ABC posterior distribution that is a convolution of a Gaussian and the kernel.

Proposition 3.2. (Theorem 3.1 of Li and Fearnhead (2015)) *Assume the conditions of Proposition 3.1. If $\varepsilon_n = o(a_n^{-3/5})$, $a_n(\boldsymbol{\theta}_\varepsilon - \theta_0) \xrightarrow{\mathfrak{L}} N(0, I_{ABC}^{-1}(\boldsymbol{\theta}_0))$. If $\varepsilon_n = o(a_n^{-1})$ or $d = p$ or $A(\boldsymbol{\theta}_0)$ is diagonal, $I_{ABC}(\boldsymbol{\theta}_0) = I(\boldsymbol{\theta}_0)$. For other cases, $I_{ABC}(\boldsymbol{\theta}_0) \leq I(\boldsymbol{\theta}_0)$.*

Proposition 3.2 helps us to compare the frequentist variability in the ABC posterior mean with the asymptotic ABC posterior distribution given in Proposition 3.1. If $\varepsilon_n = o(a_n^{-1})$ then we see that the ABC posterior distribution is asymptotically normal with variance matrix $a_n^{-2}I(\boldsymbol{\theta}_0)^{-1}$, and the ABC posterior mean is also asymptotic 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. We will say that ε_n is negligible in this case.

For a larger ε_n which is the same order as a_n^{-1}, it can be seen that the ABC uncertainty has rate a_n^{-1}. However the the limit density, which is a convolution of the true limiting posterior given the summary with the kernel used in ABC, will overestimate the uncertainty by a constant factor. For ε_n that decreases slower than a_n^{-1}, the ABC 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, we see that it is much easier to get ABC 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 is the same as that of the parameter. However, accurately estimating the posterior variance, or getting the ABC posterior to accurately reflect the uncertainty in the parameter is much harder. As commented in the introduction, this is only possible for values of ε_n for which the acceptance probability in a standard ABC algorithm will go to 0 as n increases. In this case the computational cost of ABC will have to increase substantially with n.

3.2 Regression Adjusted ABC Posterior

The regression adjustment of Beaumont et al. (2002) involves post-processing the ABC output to try and improve the resulting approximation to the true posterior. After simulating the ABC sample $\{(\boldsymbol{\theta}_i, \boldsymbol{s}_i)\}_{i=1,\cdots,N}$, the posterior is approximated by the adjusted sample $\{\boldsymbol{\theta}_i - \widehat{\boldsymbol{\beta}}_\varepsilon(\boldsymbol{s}_i - \boldsymbol{s}_{obs})\}_{i=1,\cdots,N}$ where $\widehat{\boldsymbol{\beta}}_\varepsilon$ is the least square estimate of the coefficient matrix in the linear model

$$\boldsymbol{\theta}_i = \boldsymbol{\alpha} + \boldsymbol{\beta}(\boldsymbol{s}_i - \boldsymbol{s}_{obs}) + e_i,\ i = 1, \cdots, N,$$

where e_i are independent identically distributed error.

In Nott et al. (2014) it is noted that the first two moments of the regression adjusted ABC sample are Monte Carlo approximations of the adjusted mean and variance in Bayes linear analysis (Goldstein and Wooff, 2007). Suppose the unknown parameter, $\boldsymbol{\theta}$, follows a probability density $p(\boldsymbol{\theta})$ and the data, $\boldsymbol{s}$, follows the conditional density $p(\boldsymbol{s}|\boldsymbol{\theta})$. Given an observation $\boldsymbol{s}_{obs}$, define the linearly adjusted distribution of $\boldsymbol{\theta}$ to be the distribution of $\boldsymbol{a} + B\boldsymbol{s}_{obs} + \boldsymbol{e}$, where $\boldsymbol{e}$ is the residual of the model $\boldsymbol{\theta} = \boldsymbol{a} + B\boldsymbol{s} + \boldsymbol{e}$. If the linear estimate $\boldsymbol{a} + B\boldsymbol{s}$ is optimal, in the sense of minimising the square error loss $E[\|\boldsymbol{\theta} - \boldsymbol{a} - B\boldsymbol{s}\|^2]$, the linearly adjusted distribution is called the *regression adjusted distribution*, and is the distribution of

$$\boldsymbol{\theta} - B_{opt}(\boldsymbol{s} - \boldsymbol{s}_{obs}), \text{ where } (\boldsymbol{a}_{opt}, B_{opt}) = \text{argmin}_{\boldsymbol{a},B} E[\|\boldsymbol{\theta} - \boldsymbol{a} - B\boldsymbol{s}\|^2].$$

For the ABC posterior distribution, $(\boldsymbol{\theta}, \boldsymbol{s}) \sim \pi_\varepsilon(\boldsymbol{\theta}, \boldsymbol{s}|\boldsymbol{s}_{obs})$ and the regression adjusted ABC posterior distribution is the distribution of

$$\boldsymbol{\theta}^* \triangleq \boldsymbol{\theta} - \boldsymbol{\beta}_\varepsilon(\boldsymbol{s} - \boldsymbol{s}_{obs}),$$

where $(\boldsymbol{\alpha}_\varepsilon, \boldsymbol{\beta}_\varepsilon) = \underset{\boldsymbol{\alpha},\boldsymbol{\beta}}{\text{argmin}} E_\varepsilon[\|\boldsymbol{\theta} - \boldsymbol{\alpha} - \boldsymbol{\beta}(\boldsymbol{s} - \boldsymbol{s}_{obs})\|^2|\boldsymbol{s}_{obs}]$. Therefore the adjusted ABC sample is approximately from $\pi_\varepsilon(\boldsymbol{\theta}^*|\boldsymbol{s}_{obs})$, with $\boldsymbol{\beta}_\varepsilon$ replaced by the finite sample estimate. One feature of $\pi_\varepsilon(\boldsymbol{\theta}^*|\boldsymbol{s}_{obs})$ is that its variance is strictly smaller than the variance of $\pi_\varepsilon(\boldsymbol{\theta}|\boldsymbol{s}_{obs})$ as long as $\boldsymbol{s}$ is correlated with $\boldsymbol{\theta}$. The following result, which give the analogous results to those of Propositions 3.1 and 3.2 for regression adjusted ABC, shows that this reduction in variation is by the correct amount to make the resulting ABC posterior correctly quantify the posterior uncertainty.

Theorem 3.1. *Assume conditions (C1)-(C7). Denote the mean of $\pi_\varepsilon(\boldsymbol{\theta}^*|\boldsymbol{s}_{obs})$ by $\boldsymbol{\theta}^*_\varepsilon$. If $\varepsilon_n = o(a_n^{-3/5})$, the following convergence holds:*

$$\sup_{A \in \mathscr{B}^p} \left| \Pi_\varepsilon(a_n(\boldsymbol{\theta}^* - \boldsymbol{\theta}^*_\varepsilon) \in A|\boldsymbol{s}_{obs}) - \int_A N(\boldsymbol{t}; 0, I(\boldsymbol{\theta}_0)^{-1})\, d\boldsymbol{t} \right| \xrightarrow{P} 0,$$

$$\text{and } a_n(\boldsymbol{\theta}^*_\varepsilon - \boldsymbol{\theta}_0) \xrightarrow{\mathfrak{L}} N(0, I(\boldsymbol{\theta}_0)^{-1}).$$

Moreover, if $\boldsymbol{\beta}_\varepsilon$ is replaced by $\widetilde{\boldsymbol{\beta}}_\varepsilon$ satisfying $a_n\varepsilon_n(\widetilde{\boldsymbol{\beta}}_\varepsilon - \boldsymbol{\beta}_\varepsilon) = o_p(1)$, the above convergence still holds.

It can be seen that for the regression adjusted ABC posterior distribution, ε_n is negligible up to the order of $o(a_n^{-3/5})$. This is a slower rate than the rate at which the posterior contracts, which, as we will show in the next section, has important consequences in terms of the computational efficiency of ABC. The regression adjustment corrects both the additional noise of the ABC posterior mean when $d > p$ and the overestimated uncertainty of the ABC posterior. This correction comes from the removal of the first order bias caused by ε. In Blum (2010), it is shown that the regression adjustment reduces the bias of ABC posterior density when $E[\boldsymbol{\theta}|\boldsymbol{s}]$ is linear and the residuals $\boldsymbol{\theta} - E[\boldsymbol{\theta}|\boldsymbol{s}]$ are homoscedastic. Our results do not require these model assumptions, and suggest that the regression adjustment should be applied routinely with ABC provided the coefficients, $\boldsymbol{\beta}_\varepsilon$, can be estimated accurately.

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

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_\varepsilon &= \widehat{Cov}[\boldsymbol{s}, \boldsymbol{\theta}]\widehat{Var}[\boldsymbol{s}]^{-1} \\ &= \boldsymbol{\beta}_\varepsilon + \frac{1}{a_n\varepsilon_n}\widehat{Cov}\left[\frac{\boldsymbol{s} - \boldsymbol{s}_\varepsilon}{\varepsilon_n}, a_n(\boldsymbol{\theta}^* - \boldsymbol{\theta}^*_\varepsilon)\right]\widehat{Var}\left[\frac{\boldsymbol{s} - \boldsymbol{s}_\varepsilon}{\varepsilon_n}\right]^{-1},\end{aligned}$$

where $\widehat{Cov}$ and $\widehat{Var}$ are the sample covariance and variance matrices. Since $Cov[\boldsymbol{s}, \boldsymbol{\theta}^*] = 0$ and the distributions of $\boldsymbol{s} - \boldsymbol{s}_\varepsilon$ and $\boldsymbol{\theta}^* - \boldsymbol{\theta}^*_\varepsilon$ contract at rates ε_n and a_n^{-1} respectively, the error $\widehat{\boldsymbol{\beta}}_\varepsilon - \boldsymbol{\beta}_\varepsilon$ can be shown to have the rate $O_p((a_n\varepsilon_n)^{-1}\sqrt{N}^{-1})$ as $n \to \infty$ and $N \to \infty$. We omit the proof since it is tedious and similar to the proof of an asymptotic expansion of $\boldsymbol{\beta}_\varepsilon$ in the appendix. Thus, if N increases to infinity with n, $\widehat{\boldsymbol{\beta}}_\varepsilon - \boldsymbol{\beta}_\varepsilon$ will be $o_p((a_n\varepsilon_n)^{-1})$ and the convergence of Theorem 3.1 will hold instead.

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

Proposition 3.3. *Assume conditions (C1)-(C7). Consider $\boldsymbol{\theta}^* = \boldsymbol{\theta} - \widehat{\boldsymbol{\beta}}_\varepsilon(\boldsymbol{s} - \boldsymbol{s}_{obs})$. If $\varepsilon_n = o(a_n^{-3/5})$, as $n \to \infty$ and $N \to \infty$, it holds that for any $A \in \mathscr{B}^p$,*

$$\Pi_\varepsilon(a_n(\boldsymbol{\theta}^* - \boldsymbol{\theta}^*_\varepsilon) \in A | \boldsymbol{s}_{obs}) \xrightarrow{\mathfrak{L}} \int_A \psi(\boldsymbol{t})\, d\boldsymbol{t},$$

where

$$\psi(\boldsymbol{t}) \propto \int_{\mathbb{R}^d} N\left(\boldsymbol{t}; \frac{\tau}{\sqrt{N}} R(\boldsymbol{v} - E_G[\boldsymbol{v}]), I(\boldsymbol{\theta}_0)^{-1}\right) G\left(\boldsymbol{v}; \frac{\boldsymbol{\tau}}{\sqrt{N}}, \boldsymbol{Z}\right) d\boldsymbol{v},$$

and $\boldsymbol{\tau} = O_p(1)$.

In the above result the limiting distribution can be viewed as the convolution of the limiting distribution obtained when the optimal coefficients are used and that of a random variable that is $O_p(1/\sqrt{N})$.

3.3 Acceptance Rates when ε is Negligible

Finally we present results for the acceptance probability of ABC, the quantity that is central to the computational cost of importance sampling or MCMC based ABC algorithms. We will 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 $\sigma_n \boldsymbol{X} + \boldsymbol{\mu}_n$, where $\boldsymbol{X} \sim q(\cdot)$, $E[\boldsymbol{X}] = 0$ and σ_n and $\boldsymbol{\mu}_n$ are constants that can depend on n. The average acceptance probability would then be

$$p_{acc,q} \triangleq \int_{\mathcal{P}\times\mathbb{R}^d} q_n(\boldsymbol{\theta}) f_n(\boldsymbol{s}|\boldsymbol{\theta}) K(\varepsilon_n^{-1}(\boldsymbol{s} - \boldsymbol{s}_{obs}))\, d\boldsymbol{s} d\boldsymbol{\theta},$$

where $q_n(\boldsymbol{\theta})$ is the density of $\sigma_n \boldsymbol{X} + \boldsymbol{\mu}_n$.

We further assume that $\sigma_n(\boldsymbol{\mu}_n - \boldsymbol{\theta}_0) = O_p(1)$ which means $\boldsymbol{\theta}_0$ is in the coverage of $q_n(\boldsymbol{\theta})$. This is a natural requirement for any good proposal distribution. The prior distribution and $\boldsymbol{\theta}_0$ as a point mass are included in this proposal family. Also this condition would apply to an MCMC implementation of ABC after convergence, as at stationarity θ values would be proposed from the ABC posterior distribution.

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

Theorem 3.2. *Assume the conditions of Proposition 3.1 and $\varepsilon_n = o(1/\sqrt{a_n})$. Then the following convergence holds depending on the vale of c_ε and σ_n:*

(1) If $c_\varepsilon = 0$ *or* $\sigma_n/a_{n,\varepsilon}^{-1} \to \infty$ *then* $p_{acc,q} \xrightarrow{P} 0$.

(2) 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* $p_{acc,q} = \Theta_p(1)$.

(3) If $c_\varepsilon = \infty$ *and* $\sigma_n/a_{n,\varepsilon}^{-1} \to 0$*, then* $p_{acc,q} \xrightarrow{P} 1$.

The intuition for the above result is as follows. For the summary statistic $\boldsymbol{s}_n$ sampled with parameter value $\boldsymbol{\theta}$, the acceptance probability depends on

$$\frac{\boldsymbol{s}_n - \boldsymbol{s}_{obs}}{\varepsilon_n} = \frac{1}{\varepsilon_n}[(\boldsymbol{s}_n - \boldsymbol{s}(\boldsymbol{\theta})) + (\boldsymbol{s}(\boldsymbol{\theta}) - \boldsymbol{s}(\boldsymbol{\theta}_0)) + (\boldsymbol{s}(\boldsymbol{\theta}_0) - \boldsymbol{s}_{obs})], \tag{2}$$

where $\boldsymbol{s}(\boldsymbol{\theta})$ is the limit of $\boldsymbol{s}_n$ if $(C3)$ is satisfied. The distance between $\boldsymbol{s}_n$ and $\boldsymbol{s}_{obs}$ 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 $\boldsymbol{\theta}$, (2) will blow-up as $n \to \infty$ and hence $p_{acc,q}$ goes to 0. If ε_n decreases with rate slower than a_n^{-1}, (2) will go to 0 providing we have a proposal which ensures that the middle term is $o_p(\varepsilon_n)$, and hence $p_{acc,q}$ goes to 1.

Theorem 3.2 shows that for ABC without the regression adjustment, when ε_n is in the negligible regime $o(a_n^{-1})$, the acceptance rate will degenerate to 0 as $n \to \infty$ regardless of the choice of proposal density. This means that, if regression adjustment is not used, for an accurate approximation most sample generated in ABC will be rejected. On the other hand, with the regression adjustment, we can choose $\varepsilon_n = o(a_n^{-3/5})$, and its asymptotic effect will still be negligible. For such a choice, if our proposal density satisfies $\sigma_n = o(\varepsilon_n)$, the acceptance rate will go to 1 as $n \to \infty$. Equivalently, if our proposal density satisfies $\sigma_n = o(a_n^{-3/5})$ and ε_n is chosen by specifying the proportion of sample to be accepted, the choice of accepted proportion can go to 1 with the asymptotic effect of ε_n obtained negligible. This means that, with a good enough proposal density and the regression adjustment, accurate approximation can be achieved with accepting most of the simulated sample.

4 Numerical Example

We illustrate the above theoretical results on the g-and-k distribution (Haynes, 1998), and show the possible gain of computational efficiency from using the regression adjustment. This is a generalised family providing flexible distributional shapes including asymmetry, shorter and longer tails than the normal distribution, and is popular for testing ABC methods (e.g. Fearnhead and Prangle, 2012; Marin et al., 2013; Mengersen et al., 2013). It is defined by the quantile function,

$$F^{-1}(x; A, B, g, k) = A + B\left[1 + 0.8\frac{1-\exp\{-gz(x)\}}{1+\exp\{-gz(x)\}}\right]\{1+z(x)^2\}^k z(x),$$

where A and B are location and scale parameters, g and k 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 (A, B, g, k) has a uniform prior in $[0, 10]^4$ and multiple datasets are generated from the model with $(A, B, g, k) = (3, 1, 2, 0.5)$, which is the same setting as in Fearnhead and Prangle (2012). To illustrate the asymptotic behaviour of

ABC, data sets containing independent identically distributed observations are generated with size n ranging from 500 to $10,000$.

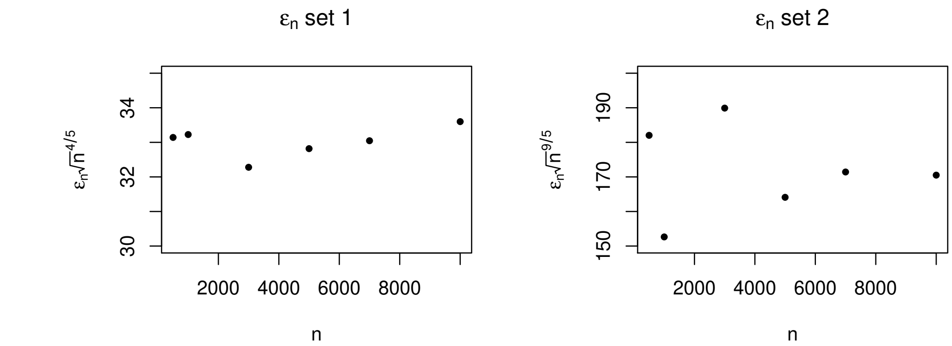
Consider estimating the posterior means, denoted by $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_p)$, and standard deviations given the summary statistic, denoted by $\boldsymbol{\sigma} = (\sigma_1, \cdots, \sigma_p)$, of the $p = 4$ parameter values using ABC approximation. The summary statistic is a set of evenly spaced quantiles of dimension 19. It has a moderate dimension in which case the acceptance probability of ABC method are usually small. The ABC bandwidth is selected so that it decreases to 0 faster or slower than $1/\sqrt{n}$, corresponding to $c_\varepsilon = \infty$ and 0 respectively, as n increases. The chosen bandwidths are roughly equal to $\sqrt{n}^{-4/5} r_1$ and $\sqrt{n}^{-9/5} r_2$ for some constants r_1 and r_2, which are shown in Figure 1(a). In every experiment, we use importance sampling to sample from the ABC posterior. The proposal distribution is a normal distribution with specified mean vector and variance matrix, and the sample size is large enough so that the Monte Carlo variability can be ignored.

First, for ε decreasing slower than $1/\sqrt{n}$, we compare the variability of the ABC means and the contraction speed of ABC standard deviations with and without regression adjustment, and the results are shown in Figure 1(b). In the left plot, both the variabilities of ABC means with and without adjustment are roughly proportional to the posterior standard deviation, and the latter is larger which reflects the extra variability caused by the higher dimension of the summary than that of the parameter. This result conforms to Proposition 3.2 and the mean convergence in Theorem 3.1. In the right plot, it can be seen that the ABC standard deviations without adjustment contract slower than the posterior standard deviation, indicating that it over-inflates the latter by a factor that diverges as n increases. With adjustment, ABC standard deviations contract at the correct speed and has relative error close to 0, hence correctly measures the posterior uncertainty. This conforms to Proposition 3.1 and Theorem 3.1.

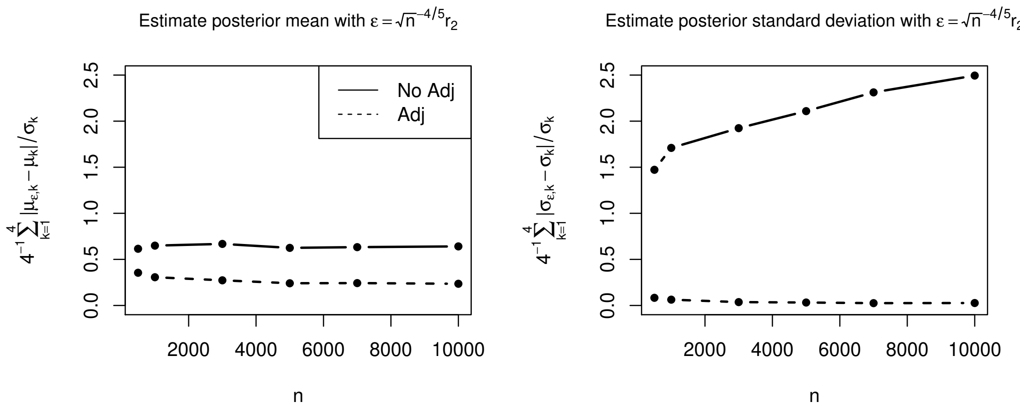
Second, the error of estimating the optimal regression coefficients $\boldsymbol{\beta}_\varepsilon$ by the simulated sample is investigated in Figure 1(c). It can be seen that the error is roughly proportional to $(\sqrt{n}\varepsilon_n)^{-1}$, since after scaled by $\sqrt{n}\varepsilon_n$, there is no obvious increasing or decreasing trend in the error. This is the same as the rate indicated by the arguments for Proposition 3.3.

Third, for the two bandwidth sets, we compare their resulted probability of accepting the sample proposed from good proposal densities, and the results are in Figure 2. For both cases, the proposal densities have the means equal to $\boldsymbol{\mu}$. When using the larger bandwidth set, the proposal density has the standard deviations twice the values of $\boldsymbol{\sigma}$, corresponding to the condition of Theorem 3.2(3), and the acceptance probability goes to 1 as n increases. When using the smaller bandwidth set, the proposal density has the standard deviations equal to $\boldsymbol{\sigma}$. This proposal can be expected to give the highest acceptance probability among all sensible choices of proposal since it basically covers the same areas as the posterior itself. It can be seen that in this case, the acceptance probability goes to 0, therefore indicating that the bandwidth in the regime of $c_\varepsilon = 0$ is too small for a fixed acceptance probability. This conforms to the results in Theorem 3.2.

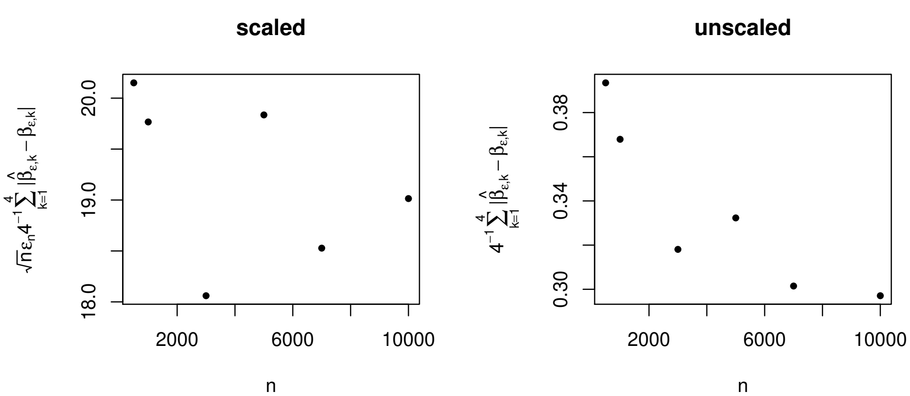
Finally, we show the computational efficiency in ABC approximation that can be gained from using the regression adjustment. This is illustrated by choosing the ABC bandwidth via fixing the proportion of sample to be accepted, and comparing the accepted proportions needed to achieve certain approximation accuracy for estimates with and without the adjustment. Higher proportion means more sample simulated can be kept for inference. The accuracy is measured



(a) Two sets of ε_n are tested, with rates $O_p(\sqrt{n}^{-4/5})$ and $O_p(\sqrt{n}^{-9/5})$ respectively.



(b) Verification of Proposition 3.1, 3.2 and Theorem 3.1



(c) Verification of Proposition 3.3

Figure 1: For each n, 50 data sets are generated, and each result in (a)-(c) is the average over 50 data sets. In each set of (a), ε_n is the average of εs used for the data sets with size n. (b) reports the mean of the scaled error of $\boldsymbol{\mu}_\varepsilon$ and the mean of relative error of $\boldsymbol{\sigma}_\varepsilon$. The size of sample forming the estimates is 9×10^4. (c) reports the scaled and unscaled mean error of $\widehat{\boldsymbol{\beta}}_\varepsilon$. The optimal coefficients are estimated using sample size 9×10^6.

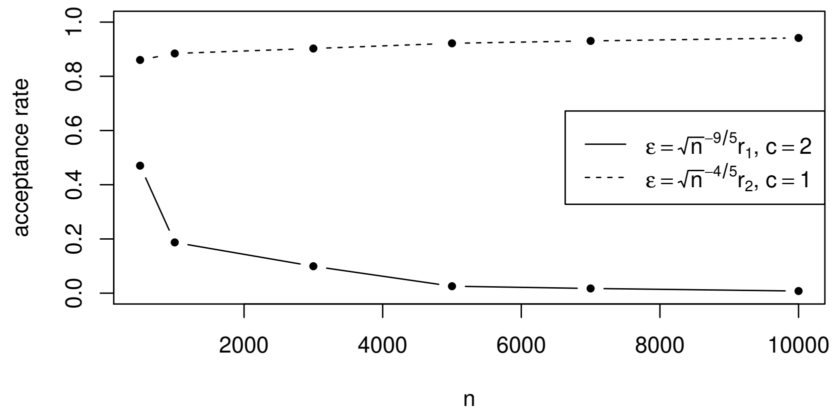


Figure 2: Verification of Theorem 3.2: plots of acceptance rate of ABC for the two regimes for ε. The proposal density has standard deviations $c\boldsymbol{\sigma}$. For each n, the average over 50 data sets is reported.

by the average relative errors of estimating $\boldsymbol{\mu}$ or $\boldsymbol{\sigma}$,

$$RE_\mu = \frac{1}{p}\sum_{k=1}^{p}\frac{|\widehat{\mu}_k - \mu_k|}{\mu_k} \text{ and } RE_\sigma = \frac{1}{p}\sum_{k=1}^{p}\frac{|\widehat{\sigma}_k - \sigma_k|}{\sigma_k},$$

for estimators $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\sigma}}$. The covariance matrix of the proposal density is selected to inflate the posterior covariance matrix by a constant factor c^2, and the mean vector is 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 are in Figure 3.

In Figure 3, it can be seen that the acceptance rate needed for regression adjusted estimate is higher than that for the unadjusted estimate in almost all the cases. For estimating the posterior mean, the improvement is not obvious, and full acceptance can be observed for both estimates. For estimating the posterior standard deviations, the improvement is much more significant. To achieve each level of accuracy, the acceptance rates of the unadjusted estimates are all close to 0. In contrast, those of the regression-adjusted estimates are higher by up to 2 orders of magnitude which means the Monte Carlo sample size needed to achieve the same accuracy can be reduced by around 2 orders of magnitude.

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., and Robert, C. P. (2009). Adaptive approximate Bayesian computation. *Biometrika*, 96(4):983–990.

Beaumont, M. A., Zhang, W., and Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, 162:2025–2035.

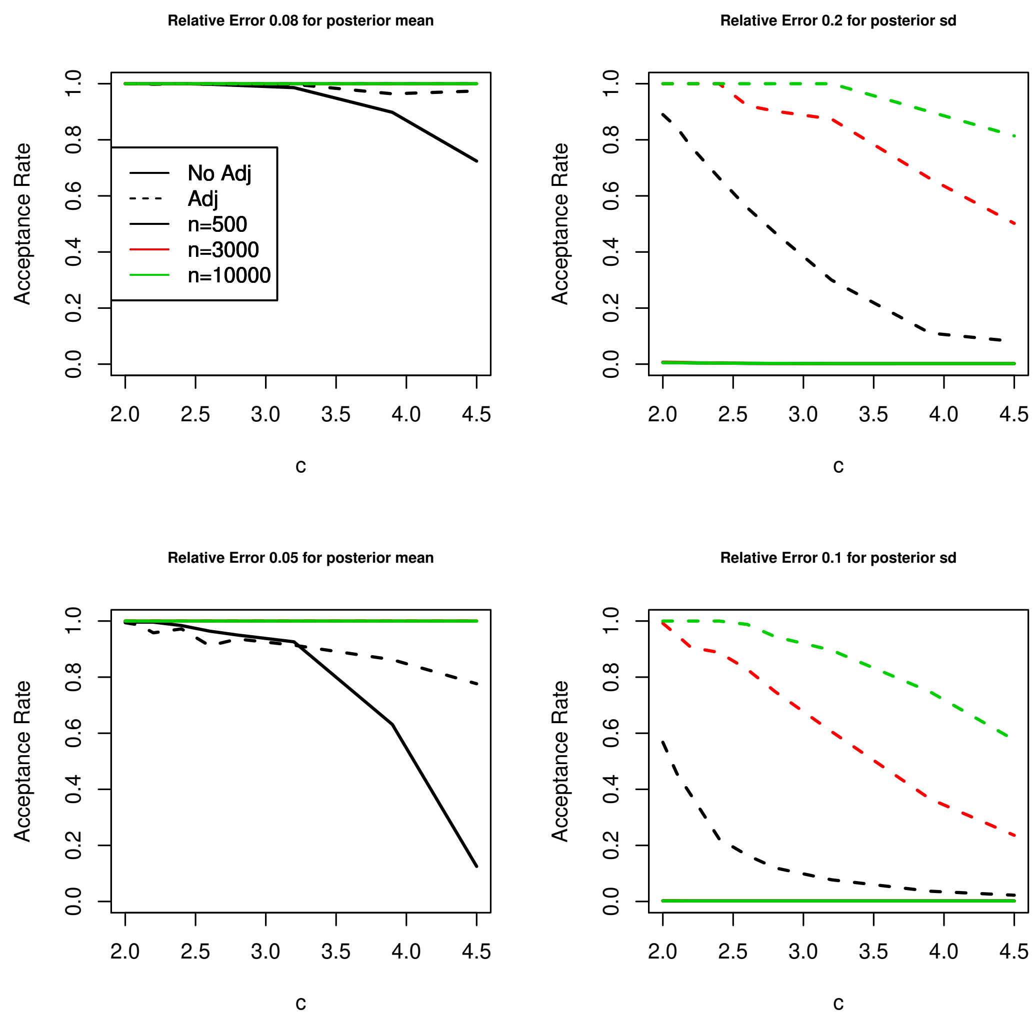


Figure 3: Acceptance rates required for different degrees of accuracy of ABC and different variances of the proposal distribution (which are proportional to c). In each plot we show results for standard ABC (full-line) and regression adjusted ABC (dashed-line) and for different values of n: $n = 500$ (black), $n = 3,000$ (red) and $n = 10,000$ (green). Results are for a relative error of 0.08 and 0.05 in the posterior mean (top-left and bottom-left respectively) and for a relative error of 0.2 and 0.1 in the posterior standard deviation (top-right and bottom-right respectively).

Bickel, P. J. and Doksum, K. A. (2015). *Mathematical statistics: Basic Ideas and Selected Topics*, volume 1. CRC Press, second edition.

Blum, M. G. (2010). Approximate Bayesian computation: a nonparametric perspective. *Journal of the American Statistical Association*, 105(491).

Blum, M. G. and François, O. (2010). Non-linear regression models for approximate Bayesian computation. *Statistics and Computing*, 20(1):63–73.

Bonassi, F. V., West, M., et al. (2015). Sequential Monte Carlo with adaptive weights for approximate Bayesian computation. *Bayesian Analysis*, 10(1):171–187.

Calvet, L. E. and Czellar, V. (2015). Accurate methods for approximate Bayesian computation filtering. *Journal of Financial Econometrics*, 13(4):798–838.

Cornuet, J.-M., Santos, F., Beaumont, M. A., Robert, C. P., Marin, J.-M., Balding, D. J., Guillemaud, T., and Estoup, A. (2008). Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics*, 24(23):2713–2719.

Drovandi, C. C. and Pettitt, A. N. (2011). Estimation of parameters for macroparasite population evolution using approximate Bayesian computation. *Biometrics*, 67(1):225–233.

Duffie, D. and Singleton, K. J. (1993). Simulated moments estimation of Markov models of asset prices. *Econometrica*, 61(4):929–952.

Fearnhead, P. and Prangle, D. (2012). Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3):419–474.

Filippi, S., Barnes, C. P., Cornebise, J., and Stumpf, M. P. (2013). On optimality of kernels for approximate Bayesian computation using sequential monte carlo. *Statistical applications in genetics and molecular biology*, 12(1):87–107.

Forneron, J.-J. and Ng, S. (2015). A likelihood-free reverse sampler of the posterior distribution. *arXiv preprint arXiv:1506.04017*.

Frazier, D. T., Martin, G. M., Robert, C. P., and Rousseau, J. (2016). Asymptotic properties of approximate Bayesian computation. *arXiv:1607.06903*.

Ghosal, S., the, J. K., and Samanta, T. (1995). On convergence of posterior distributions. *The Annals of Statistics*, pages 2145–2152.

Ghosh, J. K., Ghosal, S., and Samanta, T. (1994). Stability and convergence of the posterior in non-regular problems. In *Statistical Decision Theory and Related Topics V*, pages 183–199. Springer.

Goldstein, M. and Wooff, D. (2007). *Bayes linear statistics, theory and methods*, volume 716. John Wiley & Sons.

Gouriéroux, C. and Ronchetti, E. (1993). Indirect inference. *Journal of Applied Econometrics*, 8:s85–s118.

Haynes, M. (1998). *Flexible distributions and statistical models in ranking and selection procedures, with applications.* PhD thesis, Queensland University of Technology, Brisbane.

Heggland, K. and Frigessi, A. (2004). Estimating functions in indirect inference. *Journal of the Royal Statistical Society, series B*, 66:447–462.

Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical estimation: asymptotic theory.* Springer-Verlag, New York.

Lenormand, M., Jabot, F., and Deffuant, G. (2013). Adaptive approximate Bayesian computation for complex models. *Computational Statistics*, 28(6):2777–2796.

Li, W. and Fearnhead, P. (2015). On the asymptotic efficiency of ABC estimators. *arXiv preprint arXiv:1506.03481*.

Marin, J.-M., Pillai, N. S., Robert, C. P., and Rousseau, J. (2013). Relevant statistics for

Bayesian model choice. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Marin, J.-M., Pudlo, P., Robert, C. P., and Ryder, R. J. (2012). Approximate Bayesian computational methods. *Statistics and Computing*, 22(6):1167–1180.

Marjoram, P., Molitor, J., Plagnol, V., and Tavare, S. (2003). Markov chain Monte Carlo without likelihoods. *PNAS*, 100:15324–15328.

Meeds, E., Leenders, R., and Welling, M. (2015). Hamiltonian ABC. *arXiv:1503.01916*.

Meeds, E. and Welling, M. (2015). Optimization Monte Carlo: Efficient and embarrassingly parallel likelihood-free inference. In *Advances in Neural Information Processing Systems*, pages 2080–2088.

Mengersen, K. L., Pudlo, P., and Robert, C. P. (2013). Bayesian computation via empirical likelihood. *Proceedings of the National Academy of Sciences*, 110(4):1321–1326.

Nott, D. J., Fan, Y., Marshall, L., and Sisson, S. (2014). Approximate Bayesian computation and Bayes' linear analysis: toward high-dimensional ABC. *Journal of Computational and Graphical Statistics*, 23(1):65–86.

Toni, T., Welch, D., Strelkowa, N., Ipsen, A., and Stumpf, M. P. (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, 6(31):187–202.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.

Wegmann, D., Leuenberger, C., and Excoffier, L. (2009). Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics*.

Wood, S. N. (2010). Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, 466(7310):1102–1104.

Zhong, X. and Ghosh, M. (2016). Approximate Bayesian computation via sufficient dimension reduction. *arXiv:1608.05173*.

Appendix A

Technical lemmas needed for the main results are presented. Throughout the appendix the data are considered to be random. First some notations regarding the approximate likelihood and posterior in ABC are given. Consider some fixed $\delta < \delta_0$, divide $\mathbb{R}^p$ into $B_\delta = \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < \delta\}$ and B_δ^c. Recall that $\widetilde{f}_n(\boldsymbol{s}|\boldsymbol{\theta})$ denotes the normal density with mean $\boldsymbol{s}(\boldsymbol{\theta})$ and covariance matrix $A(\boldsymbol{\theta})/a_n^2$. Let $\widetilde{f}_\varepsilon(\boldsymbol{s}_{obs}|\boldsymbol{\theta}) = \int_{\mathbb{R}^d} \widetilde{f}_n(\boldsymbol{s}_{obs} + \varepsilon_n \boldsymbol{v}|\boldsymbol{\theta}) K(\boldsymbol{v})\, d\boldsymbol{v}$, $\widetilde{\pi}_\varepsilon(\boldsymbol{\theta}|\boldsymbol{s}_{obs})$ be the density proportional to $\pi(\boldsymbol{\theta})\widetilde{f}_\varepsilon(\boldsymbol{s}_{obs}|\boldsymbol{\theta})$, $\widetilde{\Pi}_\varepsilon(\boldsymbol{\theta} \in A|\boldsymbol{s}_{obs})$ be the distribution function with density $\widetilde{\pi}_\varepsilon(\boldsymbol{\theta}|\boldsymbol{s}_{obs})$, and its mean value be $\widetilde{\boldsymbol{\theta}}_\varepsilon$.

To begin with, the following lemma states that Π_ε and $\widetilde{\Pi}_\varepsilon$ are asymptotically the same.

Lemma 1. *Assume the conditions (C1)-(C6). If $\varepsilon_n = o(a_n^{-1/2})$, then*

(a) $\Pi_\varepsilon(\boldsymbol{\theta} \in B_\delta^c|\boldsymbol{s}_{obs})$ and $\widetilde{\Pi}_\varepsilon(\boldsymbol{\theta} \in B_\delta^c|\boldsymbol{s}_{obs})$ are $o_p(1)$,

(b) $sup_{A\in\mathscr{B}^p} \left|\Pi_\varepsilon(\boldsymbol{\theta} \in A_\theta \cap B_\delta|\boldsymbol{s}_{obs}) - \widetilde{\Pi}_\varepsilon(\boldsymbol{\theta} \in A_\theta \cap B_\delta|\boldsymbol{s}_{obs})\right| = o_p(1)$, and

(c) $a_{n,\varepsilon}(\boldsymbol{\theta}_\varepsilon - \widetilde{\boldsymbol{\theta}}_\varepsilon) = o_p(1)$, where $\widetilde{\boldsymbol{\theta}}_\varepsilon$ is the mean of $\widetilde{\Pi}_\varepsilon(\cdot|\boldsymbol{s}_{obs})$.

The following lemma gives an expansion of the optimal linear coefficient matrix $\boldsymbol{\beta}_\varepsilon$.

Lemma 2. *Assume the conditions (C1)-(C7). Then if $\varepsilon_n = o(a_n^{-3/5})$,*

$$\boldsymbol{\beta}_\varepsilon = \begin{cases} \boldsymbol{\beta}_0 + O_p(a_n^{-7/5}\varepsilon_n^{-1}), & c_\varepsilon = 0, \\ \boldsymbol{\beta}_0 + o_p(a_n^{-2/5}), & c_\varepsilon > 0, \end{cases}$$

$$\text{where } \boldsymbol{\beta}_0 = \left[D\boldsymbol{s}(\boldsymbol{\theta}_0)^T A(\boldsymbol{\theta}_0)^{-1} D\boldsymbol{s}(\boldsymbol{\theta}_0)\right]^{-1} D\boldsymbol{s}(\boldsymbol{\theta}_0)^T A(\boldsymbol{\theta}_0)^{-1}.$$

Note that when ε_n is very small, although $\boldsymbol{\beta}_\varepsilon$ could go to infinity, the overall effect of $\boldsymbol{\beta}_\varepsilon(\boldsymbol{s} - \boldsymbol{s}_{obs})$ is still small and the remaining results are the same as other cases.

Consider an approximate of the ABC posterior probability function $\Pi_\varepsilon(\boldsymbol{\theta}^* \in A|\boldsymbol{s}_{obs})$ of $\boldsymbol{\theta}^*$, $\widehat{\Pi}_\varepsilon(\boldsymbol{\theta}^* \in A|\boldsymbol{s}_{obs})$, defined as the posterior probability with the prior density $\pi_{B_\delta}(\boldsymbol{\theta}^*)$ and the likelihood function $\widetilde{f}_\varepsilon^*(\boldsymbol{s}_{obs}|\boldsymbol{\theta}^*)$ as the following,

$$\pi_{B_\delta}(\boldsymbol{\theta}^*) \triangleq \frac{\pi(\boldsymbol{\theta}^*)\mathbb{1}_{\boldsymbol{\theta}^* \in B_\delta}}{\int_{B_\delta} \pi(\boldsymbol{\theta}^*)\, d\boldsymbol{\theta}^*} \text{ and } \widetilde{f}_\varepsilon^*(\boldsymbol{s}_{obs}|\boldsymbol{\theta}^*) \triangleq \frac{\int_{\|\boldsymbol{\beta}_\varepsilon \varepsilon_n \boldsymbol{v}\| \le M_\delta} \pi(\boldsymbol{\theta}^* + \boldsymbol{\beta}_\varepsilon \varepsilon_n \boldsymbol{v}) \widetilde{f}_n(\boldsymbol{s}_{obs} + \varepsilon \boldsymbol{v}|\boldsymbol{\theta}^* + \boldsymbol{\beta}_\varepsilon \varepsilon_n \boldsymbol{v}) K(\boldsymbol{v})\, d\boldsymbol{v}}{\int_{\|\boldsymbol{\beta}_\varepsilon \varepsilon_n \boldsymbol{v}\| \le M_\delta} \pi(\boldsymbol{\theta}^* + \boldsymbol{\beta}_\varepsilon \varepsilon_n \boldsymbol{v}) K(\boldsymbol{v})\, d\boldsymbol{v}},$$

where M_δ is some constant larger than δ. It is defined by modifying the normal counterpart of the ABC posterior probability, $\widetilde{\Pi}_\varepsilon(\boldsymbol{\theta}^* \in A|\boldsymbol{s}_{obs})$. This can be seen by noting that $\widetilde{\Pi}_\varepsilon(\boldsymbol{\theta}^* \in A|\boldsymbol{s}_{obs})$ can be written as a posterior probability of $\boldsymbol{\theta}^*$, the prior of which is $\int_{\mathbb{R}^d} \pi(\boldsymbol{\theta}^* + \boldsymbol{\beta}_\varepsilon \varepsilon_n \boldsymbol{v}) K(\boldsymbol{v})\, d\boldsymbol{v}$ and the likelihood of which is $\widetilde{f}_\varepsilon^*(\boldsymbol{s}_{obs}|\boldsymbol{\theta}^*)$ with the integrated areas being $\mathbb{R}^d$. With such modifications, the prior is fixed as n changes, so conforming with the setting in Ghosal et al. (1995), and the integrated areas in the likelihood are compact, useful for technical reasons.

The following lemma, similar to Lemma 1, says that the ABC posterior probability of $\boldsymbol{\theta}^*$ and $\widehat{\Pi}_\varepsilon(\boldsymbol{\theta}^* \in A|\boldsymbol{s}_{obs})$ are asymptotically the same. Recall that $\boldsymbol{\theta}_\varepsilon^*$ is the mean of $\pi_\varepsilon(\boldsymbol{\theta}^*|\boldsymbol{s}_{obs})$.

Lemma 3. *Assume the conditions (C1)-(C7). If $\varepsilon_n = o(a_n^{-3/5})$,*

(a) $\Pi_\varepsilon(\boldsymbol{\theta}^* \in B_\delta^c|\boldsymbol{s}_{obs})$ and $\widehat{\Pi}_\varepsilon(\boldsymbol{\theta}^* \in B_\delta^c|\boldsymbol{s}_{obs})$ are $o_p(1)$,

(b) $\sup_{A \in \mathscr{B}^p} \left|\Pi_\varepsilon(\boldsymbol{\theta}^* \in A \cap B_\delta|\boldsymbol{s}_{obs}) - \widehat{\Pi}_\varepsilon(\boldsymbol{\theta}^* \in A \cap B_\delta|\boldsymbol{s}_{obs})\right| = o_p(1)$, and

(c) $a_n(\boldsymbol{\theta}_\varepsilon^* - \widehat{\boldsymbol{\theta}}_\varepsilon^*) = o_p(1)$, where $\widehat{\boldsymbol{\theta}}_\varepsilon^*$ is the mean of $\widehat{\Pi}_\varepsilon(\cdot|\boldsymbol{s}_{obs})$.

Appendix B

The proofs of the main results are presented here. Let $\boldsymbol{W}_{obs} = a_n A(\boldsymbol{\theta}_0)^{-1/2}(\boldsymbol{s}_{obs} - \boldsymbol{s}(\boldsymbol{\theta}_0))$, and by (C3), $\boldsymbol{W}_{obs} \xrightarrow{\mathfrak{L}} \boldsymbol{Z}$ where $\boldsymbol{Z} \sim N(0, I_d)$.

Proof of Proposition 3.1. The proof proceeds as follows. First the convergence of $\widetilde{\Pi}_\varepsilon$ of the properly scaled and centered $\boldsymbol{\theta}$ is considered. Then its limit is shown to be $\psi(\boldsymbol{t})$. Finally, we show that the convergence of $\widetilde{\Pi}_\varepsilon$ implies the convergence of Π_ε.

First, since $\widetilde{\Pi}_\varepsilon(\boldsymbol{\theta} \in A|\boldsymbol{s}_{obs})$ is the posterior probability with prior $\pi(\boldsymbol{\theta})$ and likelihood $\widetilde{f}_\varepsilon(\boldsymbol{s}_{obs}|\boldsymbol{\theta})$, the posterior convergence results in Ghosal et al. (1995, Proposition 2) and Ghosh et al. (1994, Remark 2.4) can be used. More specifically, let $Z_n(\boldsymbol{t}) = \widetilde{f}_\varepsilon(\boldsymbol{s}_{obs}|\boldsymbol{\theta}_0 + a_{n,\varepsilon}^{-1}\boldsymbol{t})/\widetilde{f}_\varepsilon(\boldsymbol{s}_{obs}|\boldsymbol{\theta}_0)$ be the ratio of the ABC likelihood. If $Z_n(\boldsymbol{t})$ satisfies the set of conditions from Ibragimov and Has'minskii (1981, Theorem 10.2), given in the supplementary material as (IH) and one of which stated below,

(IH3) the finite-dimensional distributions of $\{Z_n(\boldsymbol{t}) : \boldsymbol{t} \in \mathbb{R}^p\}$ converge to those of a stochastic process $\{Z(\boldsymbol{t}) : \boldsymbol{t} \in \mathbb{R}^p\}$, then by Ghosh et al. (1994, Remark 2.4) the following would hold,

$$a_{n,\varepsilon}(\widetilde{\boldsymbol{\theta}}_\varepsilon - \boldsymbol{\theta}_0) \xrightarrow{\mathfrak{L}} \boldsymbol{\tau} \text{ and } \widetilde{\Pi}_\varepsilon(a_{n,\varepsilon}(\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}}_\varepsilon) \in A|\boldsymbol{s}_{obs}) \xrightarrow{\mathfrak{L}} \int_A \xi(\boldsymbol{t}+\boldsymbol{\tau})\,d\boldsymbol{t},$$

where $\boldsymbol{\tau} = \int \boldsymbol{t}\xi(\boldsymbol{t})\,d\boldsymbol{t}$, $\xi(\boldsymbol{t}) = Z(\boldsymbol{t})/\int Z(\boldsymbol{t})d\boldsymbol{t}$. Furthermore, if $\xi(\boldsymbol{t}+\boldsymbol{\tau})$ is nonrandom, by Ghosal et al. (1995, Theorem 1),

$$\sup_{A\in\mathscr{B}^p}\left|\widetilde{\Pi}_\varepsilon(a_{n,\varepsilon}(\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}}_\varepsilon) \in A|\boldsymbol{s}_{obs}) - \int_A \xi(\boldsymbol{t}+\boldsymbol{\tau})\,d\boldsymbol{t}\right| \xrightarrow{P} 0. \tag{3}$$

Here we give the proof of (IH3) and those of the others have been given in the first version of Li and Fearnhead (2015) [1]. In order to obtain the weak convergence in (IH3), the weak convergence of $\widetilde{f}_\varepsilon(\boldsymbol{s}_{obs}|\boldsymbol{\theta}_0 + a_{n,\varepsilon}^{-1}\boldsymbol{t})$ for fixed $\boldsymbol{t}$ is needed. For $\widetilde{f}_n(\boldsymbol{s}_{obs} + \varepsilon_n\boldsymbol{v}|\boldsymbol{\theta}_0 + a_{n,\varepsilon}^{-1}\boldsymbol{t})$ in the integrands in $\widetilde{f}_\varepsilon(\boldsymbol{s}_{obs}|\boldsymbol{\theta}_0 + a_{n,\varepsilon}^{-1}\boldsymbol{t})$, by Taylor expansion around $a_{n,\varepsilon}^{-1}\boldsymbol{t} = 0$, we have

$$\widetilde{f}_n(\boldsymbol{s}_{obs} + \varepsilon_n\boldsymbol{v}|\boldsymbol{\theta}_0 + a_{n,\varepsilon}^{-1}\boldsymbol{t}) = N(D\boldsymbol{s}(\boldsymbol{\theta}_0+\xi_1)\boldsymbol{t}; A(\boldsymbol{\theta}_0)^{1/2}\boldsymbol{W}_{obs} + a_n\varepsilon_n\boldsymbol{v}, A(\boldsymbol{\theta}_0 + a_{n,\varepsilon}^{-1}\boldsymbol{t})),$$

where ξ_1 is a p-dimension function of $\boldsymbol{t}$ satisfying $\|\xi_1\| \le \|a_{n,\varepsilon}^{-1}\boldsymbol{t}\|$. Consider the following decomposition. For a rank-p $d\times p$ matrix A, a rand-d $d\times d$ matrix B and a d-dimension vector c, let $P = A^TA$ and by matrix algebra we have

$$N(A\boldsymbol{t}; B\boldsymbol{v}+c, I_d) = N(\boldsymbol{t}; (A^TA)^{-1}A^T(\boldsymbol{c}+B\boldsymbol{v}), (A^TA)^{-1})g(\boldsymbol{v}; A, B, \boldsymbol{c}), \tag{4}$$
$$\text{where } g(\boldsymbol{v}; A, B, \boldsymbol{c}) = \frac{1}{(2\pi)^{(d-p)/2}}\exp\{-\frac{1}{2}(\boldsymbol{c}+B\boldsymbol{v})^T(I - A(A^TA)^{-1}A^T)(\boldsymbol{c}+B\boldsymbol{v})\}.$$

Denote the vector $(a_{n,\varepsilon}^{-1}\boldsymbol{t}^T, \xi_1^T)^T$ by V_1 and $(\boldsymbol{W}_{obs}^T, a_n\varepsilon_n)^T$ by V_2. Then by the decomposition (4), $\widetilde{f}_n(\boldsymbol{s}_{obs} + \varepsilon_n\boldsymbol{v}|\boldsymbol{\theta}_0 + a_{n,\varepsilon}^{-1}\boldsymbol{t})$ can be written as

$$\widetilde{f}_n(\boldsymbol{s}_{obs} + \varepsilon_n\boldsymbol{v}|\boldsymbol{\theta}_0 + a_{n,\varepsilon}^{-1}\boldsymbol{t}) = N(\boldsymbol{t}; \boldsymbol{\beta}_\varepsilon'(\xi_1)A(\boldsymbol{\theta}_0)^{1/2}\boldsymbol{W}_{obs} - a_n\varepsilon_n\boldsymbol{\beta}_\varepsilon'(\xi_1)\boldsymbol{v}, I(\boldsymbol{\theta}_0; V_1)^{-1})g'(\boldsymbol{v}; V_1, V_2), \tag{5}$$

where $\boldsymbol{\beta}_\varepsilon'(\boldsymbol{x}) = [D\boldsymbol{s}(\boldsymbol{\theta}_0+\boldsymbol{x})^TA(\boldsymbol{\theta}_0 + a_{n,\varepsilon}^{-1}\boldsymbol{t})^{-1}D\boldsymbol{s}(\boldsymbol{\theta}_0+\boldsymbol{x})]^{-1}D\boldsymbol{s}(\boldsymbol{\theta}_0+\boldsymbol{x})^TA(\boldsymbol{\theta}_0 + a_{n,\varepsilon}^{-1}\boldsymbol{t})^{-1}$, $I(\boldsymbol{\theta}_0; V_1) = D\boldsymbol{s}(\boldsymbol{\theta}_0+\xi_1)^TA(\boldsymbol{\theta}_0 + a_{n,\varepsilon}^{-1}\boldsymbol{t})^{-1}D\boldsymbol{s}(\boldsymbol{\theta}_0+\xi_1)$ and $g'(\boldsymbol{v}; V_1, V_2)$ corresponds to $g(\boldsymbol{v}; A, B, \boldsymbol{c})$ in (4). The notations $I(\boldsymbol{\theta}_0; V_1)$ and $g'(\boldsymbol{v}; V_1, V_2)$ mean that the functions depend on n through V_1 and/or V_2. Since $|g'(\boldsymbol{v}; V_1, V_2)|$ is bounded by a constant and $I(\boldsymbol{\theta}_0; V_1)$ and $g'(\boldsymbol{v}; V_1, V_2)$ are continuous as functions of V_1 and/or V_2, by dominated convergence theorem, $\widetilde{f}_\varepsilon(\boldsymbol{s}_{obs}|\boldsymbol{\theta}_0 + a_{n,\varepsilon}^{-1}\boldsymbol{t})$ is continuous as a function of V_1, V_2 and $\boldsymbol{\beta}_\varepsilon'(\xi_1)$. Then the limit of $\widetilde{f}_\varepsilon(\boldsymbol{s}_{obs}|\boldsymbol{\theta}_0 + a_{n,\varepsilon}^{-1}\boldsymbol{t})$ depends on whether $a_n\varepsilon_n$ is bounded.

If $c_\varepsilon < \infty$, since $\boldsymbol{\beta}_\varepsilon'(\xi_1) \to \boldsymbol{\beta}_0$ where $\boldsymbol{\beta}_0$ is defined in Lemma 2, $V_1 \to 0$ and $V_2 \xrightarrow{\mathfrak{L}} (\boldsymbol{Z}^T, c_\varepsilon)^T$ where $\boldsymbol{Z} \sim N(0, I_d)$, by the continuous mapping theorem (Van der Vaart, 2000) we have

$$\widetilde{f}_\varepsilon(\boldsymbol{s}_{obs}|\boldsymbol{\theta}_0 + a_{n,\varepsilon}^{-1}\boldsymbol{t}) \xrightarrow{\mathfrak{L}} N(\boldsymbol{t}; \boldsymbol{\beta}_0A(\boldsymbol{\theta}_0)^{1/2}\boldsymbol{Z} - c_\varepsilon\boldsymbol{\beta}_0\boldsymbol{v}, I(\boldsymbol{\theta}_0)^{-1})g'(\boldsymbol{v})K(\boldsymbol{v}),$$

where $g'(\boldsymbol{v})$ is $g'(\boldsymbol{v}; V_1, V_2)$ replacing V_1 and V_2 with their limits. Then by continuous mapping theorem again, for one-dimension $Z_n(\boldsymbol{t})$,

$$Z_n(\boldsymbol{t}) \xrightarrow{\mathfrak{L}} Z(\boldsymbol{t}) \triangleq \frac{\int N(\boldsymbol{t}; \boldsymbol{\beta}_0A(\boldsymbol{\theta}_0)^{1/2}\boldsymbol{Z} - c_\varepsilon\boldsymbol{\beta}_0\boldsymbol{v}, I(\boldsymbol{\theta}_0)^{-1})g'(\boldsymbol{v})K(\boldsymbol{v})\,d\boldsymbol{v}}{\int N(0; \boldsymbol{\beta}_0A(\boldsymbol{\theta}_0)^{1/2}\boldsymbol{Z} - c_\varepsilon\boldsymbol{\beta}_0\boldsymbol{v}, I(\boldsymbol{\theta}_0)^{-1})g'(\boldsymbol{v})K(\boldsymbol{v})\,d\boldsymbol{v}}.$$

[1] Available at https://arxiv.org/pdf/1506.03481v1.pdf

If $c_\varepsilon = \infty$, the above convergence does not hold and we need a different expression of $\widetilde{f}_n(\boldsymbol{s}_{obs} + \varepsilon_n \boldsymbol{v}|\boldsymbol{\theta}_0 + a_{n,\varepsilon}^{-1}\boldsymbol{t})$. Consider the transformation $\boldsymbol{v}' = a_n(\boldsymbol{s}_{obs} - \boldsymbol{s}(\boldsymbol{\theta}_0 + \varepsilon_n \boldsymbol{v})) + a_n \varepsilon_n \boldsymbol{v}$ and we have

$$\widetilde{f}_n(\boldsymbol{s}_{obs} + \varepsilon_n \boldsymbol{v}|\boldsymbol{\theta}_0 + a_{n,\varepsilon}^{-1}\boldsymbol{t})K(\boldsymbol{v}) = N(\boldsymbol{v}'; 0, A(\boldsymbol{\theta}_0 + \varepsilon_n \boldsymbol{t}))K((a_n\varepsilon_n)^{-1}[\boldsymbol{v}' - A(\boldsymbol{\theta}_0)^{1/2}\boldsymbol{W}_{obs}] + D\boldsymbol{s}(\boldsymbol{\theta}_0 + \varepsilon_n \xi_2)^T \boldsymbol{t}),$$

where ξ_2 is a p-dimension function of $\boldsymbol{t}$ satisfying $\|\xi_2\| \leq \|\varepsilon_n \boldsymbol{t}\|$. Let $V_1' = (\varepsilon_n \boldsymbol{t}^T, \xi_2^T)^T$. The by continuous mapping theorem we have

$$\widetilde{f}_n(\boldsymbol{s}_{obs} + \varepsilon_n \boldsymbol{v}|\boldsymbol{\theta}_0 + a_{n,\varepsilon}^{-1}\boldsymbol{t})K(\boldsymbol{v}) \xrightarrow{P} N(\boldsymbol{v}'; 0, A(\boldsymbol{\theta}_0))K(D\boldsymbol{s}(\boldsymbol{\theta}_0)^T \boldsymbol{t})$$

and

$$Z_n(\boldsymbol{t}) \xrightarrow{\mathfrak{L}} Z(\boldsymbol{t}) \triangleq K(D\boldsymbol{s}(\boldsymbol{\theta}_0)^T \boldsymbol{t}).$$

The arguments can be easily generalised to multi-dimension case. Therefore (IH3) holds.

Now we show that $\xi(\boldsymbol{t} + \boldsymbol{\tau}) = \psi(\boldsymbol{t})$. The derivation is mainly algebraic using (4). Then let $A = A(\boldsymbol{\theta}_0)^{-1/2} D\boldsymbol{s}(\boldsymbol{\theta}_0)$, $B = c_\varepsilon A(\boldsymbol{\theta}_0)^{-1/2}$, $\boldsymbol{c} = \boldsymbol{Z}$, we have the following expressions of $\boldsymbol{\tau}$,

$$\boldsymbol{\tau} = \begin{cases} Z_{\boldsymbol{s}}, & \text{when } c_\varepsilon = 0, \\ Z_{\boldsymbol{s}} + c_\varepsilon R E_G[\boldsymbol{v}] & \text{when } c_\varepsilon \in (0, \infty), \\ 0, & \text{when } c_\varepsilon = \infty, \end{cases}$$

where $R = (A^T A)^{-1} A^T B$, $G(\boldsymbol{v}; c_\varepsilon, \boldsymbol{Z}) = g(\boldsymbol{v}; A, B, \boldsymbol{c})K(\boldsymbol{v}) / \int g(\boldsymbol{v}; A, B, \boldsymbol{c})K(\boldsymbol{v})\, d\boldsymbol{v}$ and $Z_{\boldsymbol{s}} = (A^T A)^{-1} A^T \boldsymbol{Z}$. Then plugging in $\boldsymbol{\tau}$, $\xi(\boldsymbol{t} + \boldsymbol{\tau}) = \psi(\boldsymbol{t})$ holds.

Finally, we show that

$$\sup_{A \in \mathscr{B}^p} \left| \Pi_\varepsilon(a_{n,\varepsilon}(\boldsymbol{\theta} - \boldsymbol{\theta}_\varepsilon) \in A|\boldsymbol{s}_{obs}) - \widetilde{\Pi}_\varepsilon(a_{n,\varepsilon}(\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}}_\varepsilon) \in A|\boldsymbol{s}_{obs}) \right| = o_p(1), \tag{6}$$

by which the proposition holds. According to Lemma 1 and (3), it is sufficient to show that

$$\sup_{A \in \mathscr{B}^p} \left| \int_{\{\boldsymbol{t}:\boldsymbol{t} + a_{n,\varepsilon}(\boldsymbol{\theta}_\varepsilon - \widetilde{\boldsymbol{\theta}}_\varepsilon) \in A\}} \psi(\boldsymbol{t})\, d\boldsymbol{t} - \int_A \psi(\boldsymbol{t})\, d\boldsymbol{t} \right| = o_p(1).$$

Let $\boldsymbol{\lambda}_n = a_{n,\varepsilon}(\boldsymbol{\theta}_\varepsilon - \widetilde{\boldsymbol{\theta}}_\varepsilon)$. By transforming $\boldsymbol{t}$ to $\boldsymbol{t} + \boldsymbol{\lambda}_n$, the LHS of the above is upper bounded by $\int_{\mathbb{R}^p} |\psi(\boldsymbol{t} - \boldsymbol{\lambda}_n) - \psi(\boldsymbol{t})|\, d\boldsymbol{t}$. Since this upper bound is a continuous function of $\boldsymbol{\lambda}_n$, by continuous mapping theorem, it is $o_p(1)$ by Lemma 1(c). Therefore the proposition holds. □

Proof of Theorem 3.1. In this proof, first the convergence of $\widehat{\Pi}_\varepsilon$ is obtained, which then implies the convergence of Π_ε.

Since $\widehat{\Pi}_\varepsilon$ is a posterior probability of $\boldsymbol{\theta}^*$ with prior $\pi_{B_\delta}(\boldsymbol{\theta}^*)$ and likelihood function $\widetilde{f}_\varepsilon^*(\boldsymbol{s}_{obs}|\boldsymbol{\theta}^*)$, Ghosal et al. (1995, Theorem 1) can be applied to obtain results similar to (3), by verifying the conditions similar to (IH) for the likelihood ratio $Z_n^*(\boldsymbol{t})$, i.e. $\widetilde{f}_\varepsilon^*(\boldsymbol{s}_{obs}|\boldsymbol{\theta}_0 + a_n^{-1}\boldsymbol{t}) / \widetilde{f}_\varepsilon^*(\boldsymbol{s}_{obs}|\boldsymbol{\theta}_0)$. The proof of the following condition is given here,

(IH3*) the finite-dimensional distributions of $\{Z_n^*(\boldsymbol{t}) : \boldsymbol{t} \in \mathbb{R}^p\}$ converge to those of a stochastic process $\{Z^*(\boldsymbol{t}) : \boldsymbol{t} \in \mathbb{R}^p\}$.

The proof for the others can be seen in the supplementary material. The verification of (IH3*) basically follows the lines of verifying (IH3). Let $\boldsymbol{\theta}(\boldsymbol{t},\boldsymbol{v}) = \boldsymbol{\theta}_0 + a_n^{-1}\boldsymbol{t} + \boldsymbol{\beta}_\varepsilon\varepsilon_n\boldsymbol{v}$. For $\widetilde{f}_n(\boldsymbol{s}_{obs} + \varepsilon_n\boldsymbol{v}|\boldsymbol{\theta}(\boldsymbol{t},\boldsymbol{v}))$ in the integrands of $\widetilde{f}^*_\varepsilon(\boldsymbol{s}_{obs}|\boldsymbol{\theta}_0 + a_n^{-1}\boldsymbol{t})$, by Taylor expansion around $a_n^{-1}\boldsymbol{t}+\boldsymbol{\beta}_\varepsilon\varepsilon_n\boldsymbol{v} = 0$, we have

$$\widetilde{f}_n(\boldsymbol{s}_{obs} + \varepsilon_n\boldsymbol{v}|\boldsymbol{\theta}(\boldsymbol{t},\boldsymbol{v})) = N(D\boldsymbol{s}(\boldsymbol{\theta}_0 + \xi_3)\boldsymbol{t}; A(\boldsymbol{\theta}_0)^{1/2}\boldsymbol{W}_{obs} + a_n\varepsilon_n[I_d - D\boldsymbol{s}(\boldsymbol{\theta}_0 + \xi_3)\boldsymbol{\beta}_\varepsilon]\boldsymbol{v}, A(\boldsymbol{\theta}(\boldsymbol{t},\boldsymbol{v}))),$$

where ξ_3 is a function of $\boldsymbol{t}$ and $\boldsymbol{v}$ satisfying $\|\xi_3\| \leq \|a_n^{-1}\boldsymbol{t}+\boldsymbol{\beta}_\varepsilon\varepsilon_n\boldsymbol{v}\|$. Denote the vector $(a_n^{-1}\boldsymbol{t}^T, \xi_3, \boldsymbol{\beta}_\varepsilon^{(v)}\varepsilon_n)^T$ by V_3 and $(\boldsymbol{W}_{obs}^T, a_n\varepsilon_n, \boldsymbol{\beta}_\varepsilon^{(v)})^T$ by V_4 where $\boldsymbol{\beta}_\varepsilon^{(v)}$ is the vector of all elements of $\boldsymbol{\beta}_\varepsilon$. Then by (4), $\widetilde{f}_n(\boldsymbol{s}_{obs} + \varepsilon_n\boldsymbol{v}|\boldsymbol{\theta}(\boldsymbol{t},\boldsymbol{v}))$ can be written as

$$\widetilde{f}_n(\boldsymbol{s}_{obs} + \varepsilon_n\boldsymbol{v}|\boldsymbol{\theta}(\boldsymbol{t},\boldsymbol{v})) = N(\boldsymbol{t}; \boldsymbol{\beta}_\varepsilon'(\xi_3)A(\boldsymbol{\theta}_0)^{1/2}\boldsymbol{W}_{obs} - a_n\varepsilon_n(\boldsymbol{\beta}_\varepsilon'(\xi_3) - \boldsymbol{\beta}_\varepsilon)\boldsymbol{v}, I(\boldsymbol{\theta}_0; V_3, \boldsymbol{v})^{-1})g'(\boldsymbol{v}; V_3, V_4), \tag{7}$$

where $I(\boldsymbol{\theta}_0; V_1, \boldsymbol{v}) = D\boldsymbol{s}(\boldsymbol{\theta}_0+\xi_3)^T A(\boldsymbol{\theta}(\boldsymbol{t},\boldsymbol{v}))^{-1} D\boldsymbol{s}(\boldsymbol{\theta}_0+\xi_3)$ and $g'(\boldsymbol{v}; V_3, V_4)$ corresponds to $g(\boldsymbol{v}; A, B, \boldsymbol{c})$ in (4). Then by continuous mapping theorem, since $a_n\varepsilon_n(\boldsymbol{\beta}_\varepsilon'(\xi_3) - \boldsymbol{\beta}_\varepsilon) = o_p(1)$, $\boldsymbol{\beta}_\varepsilon'(\xi_3) \to \boldsymbol{\beta}_0$, $V_3 \xrightarrow{P} 0$ and $V_2 \xrightarrow{\mathfrak{L}} (\boldsymbol{Z}^T, c_\varepsilon, \boldsymbol{\beta}_0^{(v)})^T$ where $\boldsymbol{Z} \sim N(0, I_d)$, we have

$$\widetilde{f}^*_\varepsilon(\boldsymbol{s}_{obs}|\boldsymbol{\theta}_0 + a_n^{-1}\boldsymbol{t}) \xrightarrow{\mathfrak{L}} N(\boldsymbol{t}; \boldsymbol{\beta}_0 A(\boldsymbol{\theta}_0)^{1/2}\boldsymbol{Z}, I(\boldsymbol{\theta}_0)) \int_{\mathbb{R}^d} g'(\boldsymbol{v})K(\boldsymbol{v})\, d\boldsymbol{v},$$

where $g'(\boldsymbol{v})$ is $g'(\boldsymbol{v}; V_3, V_4)$ replacing V_3 and V_4 with their weak convergence limits. Here whether c_ε is finite or not does not affect the convergence of $g'(\boldsymbol{v})$. Therefore for one-dimension $Z_n^*(\boldsymbol{t})$,

$$Z_n^*(\boldsymbol{t}) \xrightarrow{\mathfrak{L}} Z^*(\boldsymbol{t}) \triangleq \frac{N(\boldsymbol{t}; \boldsymbol{\beta}_0 A(\boldsymbol{\theta}_0)^{1/2}\boldsymbol{Z}, I(\boldsymbol{\theta}_0)^{-1})}{N(0; \boldsymbol{\beta}_0 A(\boldsymbol{\theta}_0)^{1/2}\boldsymbol{Z}, I(\boldsymbol{\theta}_0)^{-1})}.$$

The arguments can be easily generalised to multi-dimension case. Therefore (IH3*) holds.

Let $\xi^*(\boldsymbol{t}) = Z^*(\boldsymbol{t})/\int Z^*(\boldsymbol{t})\, d\boldsymbol{t}$ and $\boldsymbol{\tau}^* = \int \boldsymbol{t}\xi^*(\boldsymbol{t})\, d\boldsymbol{t}$. Obviously $\xi^*(\boldsymbol{t}+\boldsymbol{\tau}^*)$ is equal to $N(\boldsymbol{t}; 0, I(\boldsymbol{\theta}_0)^{-1})$. Then by Ghosal et al. (1995, Theorem 1), similar to (3), it holds that

$$\sup_{A\in\mathscr{B}^p} \left| \widehat{\Pi}_\varepsilon(a_n(\boldsymbol{\theta}^* - \widehat{\boldsymbol{\theta}}^*_\varepsilon) \in A|\boldsymbol{s}_{obs}) - \int_A N(\boldsymbol{t}; 0, I(\boldsymbol{\theta}_0)^{-1})\, d\boldsymbol{t} \right| \xrightarrow{P} 0.$$

Similar to the arguments for (6), by Lemma 3, it holds that

$$\sup_{A\in\mathscr{B}^p} \left| \Pi_\varepsilon(a_n(\boldsymbol{\theta}^* - \boldsymbol{\theta}^*_\varepsilon) \in A|\boldsymbol{s}_{obs}) - \widehat{\Pi}_\varepsilon(a_n(\boldsymbol{\theta}^* - \widehat{\boldsymbol{\theta}}^*_\varepsilon) \in A|\boldsymbol{s}_{obs}) \right| = o_p(1).$$

Therefore the first convergence in the theorem holds. With (IH*) hold, the second convergence holds by Ibragimov and Has'minskii (1981, Theorem 10.2).

If $\boldsymbol{\beta}_\varepsilon$ is replaced by a $p \times d$ matrix $\widehat{\boldsymbol{\beta}}_\varepsilon$ satisfying $a_n\varepsilon_n(\widehat{\boldsymbol{\beta}}_\varepsilon - \boldsymbol{\beta}_\varepsilon) = o_p(1)$, the verifications of (IH1*) and (IH2*) are not affected. Since the only requirement for $\boldsymbol{\beta}_\varepsilon$ in the above is $a_n\varepsilon_n(\boldsymbol{\beta}_\varepsilon' - \boldsymbol{\beta}_\varepsilon) = o_p(1)$, the above arguments will hold for $\widehat{\boldsymbol{\beta}}_\varepsilon$. Therefore the theorem holds. □

Proof of Proposition 3.3. With $\widehat{\boldsymbol{\beta}}_\varepsilon$ replacing $\boldsymbol{\beta}_\varepsilon$, since $\widehat{\boldsymbol{\beta}}_\varepsilon - \boldsymbol{\beta}_\varepsilon = O_p((a_n\varepsilon_n)^{-1}\sqrt{N}^{-1})$, in (7), $a_n\varepsilon_n(\boldsymbol{\beta}_\varepsilon' - \boldsymbol{\beta}_\varepsilon) = O_p(\sqrt{N}^{-1})$. The verifications of (IH1*) and (IH2*) are not affected. Then by letting $\boldsymbol{\tau} = \sqrt{N}a_n\varepsilon_n(\boldsymbol{\beta}_\varepsilon' - \boldsymbol{\beta}_\varepsilon)$ and replacing $a_n\varepsilon_n\boldsymbol{\beta}_\varepsilon'(\xi_1)$ in (5) with $\boldsymbol{\tau}/\sqrt{N}$, it can be seen that (7) is almost the same as (5). Then the remaining proof simple follows the same line as in the proof of Proposition 3.1. □

Proof of Theorem 3.2. The integrand of $p_{acc,q}$ is similar to that of the normalising constant $\int_{\mathbb{R}^p}\int_{\mathbb{R}^d}\pi(\boldsymbol{\theta})f_n(\boldsymbol{s}|\boldsymbol{\theta})K(\varepsilon_n^{-1}(\boldsymbol{s}-\boldsymbol{s}_{obs}))\,d\boldsymbol{s}d\boldsymbol{\theta}$ of $\pi_\varepsilon(\boldsymbol{\theta}|\boldsymbol{s}_{obs})$. The decomposition of this normalising constant has been derived in Li and Fearnhead (2015) and restated in the supplementary materials. Following the same arguments, $p_{acc,q}$ can be decomposed as $\varepsilon_n^d\int_{B_\delta}q_n(\boldsymbol{\theta})\widetilde{f}_\varepsilon(\boldsymbol{s}_{obs}|\boldsymbol{\theta})\,d\boldsymbol{\theta}(1+o_p(1))$. Consider the transformation $\boldsymbol{t}=\boldsymbol{t}(\boldsymbol{\theta})\triangleq a_{n,\varepsilon}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)$. Let $r_{n,\varepsilon}=\sigma_n/a_{n,\varepsilon}^{-1}$ and $\boldsymbol{c}_\mu=\sigma_n(\boldsymbol{\mu}_n-\boldsymbol{\theta}_0)$. Then plugging the expression of $q_n(\boldsymbol{\theta})$ into the above decomposition of $p_{acc,q}$ gives

$$p_{acc,q}=\varepsilon_n^d\int_{t(B_\delta)}(r_{n,\varepsilon})^{-p}q(r_{n,\varepsilon}^{-1}\boldsymbol{t}-\boldsymbol{c}_\mu)\widetilde{f}_\varepsilon(\boldsymbol{s}_{obs}|\boldsymbol{\theta}_0+a_{n,\varepsilon}^{-1}\boldsymbol{t})\,d\boldsymbol{t}(1+o_p(1)).$$

Li and Fearnhead (2015, Lemma 10) gives an expansion of $\widetilde{f}_\varepsilon(\boldsymbol{s}_{obs}|\boldsymbol{\theta}_0+a_{n,\varepsilon}^{-1}\boldsymbol{t})$, with which $p_{acc,q}$ can be expanded as

$$p_{acc,q}=(a_{n,\varepsilon}\varepsilon_n)^d\int_{t(B_\delta)\times\mathbb{R}^d}(r_{n,\varepsilon})^{-p}q(r_{n,\varepsilon}^{-1}\boldsymbol{t}-\boldsymbol{c}_\mu)g_n(\boldsymbol{t},\boldsymbol{v})\,d\boldsymbol{v}d\boldsymbol{t}[1+O_p(a_n^2\varepsilon_n^4)], \tag{8}$$

$$\text{where } g_n(\boldsymbol{t},\boldsymbol{v})=\begin{cases} N\Big(D\boldsymbol{s}(\boldsymbol{\theta}_0)\boldsymbol{t};a_n\varepsilon_n\boldsymbol{v}+A(\boldsymbol{\theta}_0)^{1/2}\boldsymbol{W}_{obs},A(\boldsymbol{\theta}_0)\Big)K(\boldsymbol{v}), & \text{if } c_\varepsilon\in[0,\infty),\\ N\Big(D\boldsymbol{s}(\boldsymbol{\theta}_0)\boldsymbol{t};\boldsymbol{v}+\frac{1}{a_n\varepsilon_n}A(\boldsymbol{\theta}_0)^{1/2}\boldsymbol{W}_{obs},\frac{1}{a_n^2\varepsilon_n^2}A(\boldsymbol{\theta}_0)\Big)K(\boldsymbol{v}), & \text{if } c_\varepsilon=\infty,\end{cases}$$

$$\text{and }\int_{t(B_\delta)\times\mathbb{R}^d}g_n(\boldsymbol{t},\boldsymbol{v})\,d\boldsymbol{v}d\boldsymbol{t}<M\text{ in probability for some positive constant } M.$$

The remainder term $O_p(a_n^2\varepsilon_n^4)=o_p(1)$ since $\varepsilon_n=o(1/\sqrt{a_n})$. It can be seen that the convergence rate of $p_{acc,q}$ depends on the integral in (8). Denote this integral by $Q_{n,\varepsilon}$. When $c_\varepsilon\in[0,\infty)$, note that the normal density in $g_n(\boldsymbol{t},\boldsymbol{v})$ is bounded by some positive constant c_N in probability. Then (1) holds by $a_{n,\varepsilon}\varepsilon_n=a_n\varepsilon_n\to0$ and the following two inequalities,

$$Q_{n,\varepsilon}\le\begin{cases} c_N, & \text{if } c_\varepsilon\in[0,\infty),\\ (r_{n,\varepsilon})^{-p}\sup_{\mathbb{R}^p}q(\boldsymbol{x})M, & \text{if } r_{n,\varepsilon}\to r>0,\end{cases}\quad\text{in probability.}$$

For the integrand of $Q_{n,\varepsilon}$, consider the non-random versions that $\boldsymbol{c}_\mu$ and $\boldsymbol{W}_{obs}$ are replaced by constant vectors $\boldsymbol{c}$ and $\boldsymbol{w}$, and denote the corresponding $g_n(\boldsymbol{t},\boldsymbol{v})$ by $g_n(\boldsymbol{t},\boldsymbol{v}|\boldsymbol{w})$. For any $\boldsymbol{t}\in t(B_\delta)$ and $\boldsymbol{v}\in\mathbb{R}^d$, we have the following convergence,

$$(r_{n,\varepsilon})^{-p}q(r_{n,\varepsilon}^{-1}\boldsymbol{t}-\boldsymbol{c})\begin{cases}=\Theta(1), & \text{if } r_{n,\varepsilon}\to r>0,\\ \to\delta(\boldsymbol{t}), & \text{if } r_{n,\varepsilon}\to0,\end{cases}\text{ and } g_n(\boldsymbol{t},\boldsymbol{v}|\boldsymbol{w})\begin{cases}=\Theta(1), & \text{if } c_\varepsilon<\infty,\\ \to\delta(\boldsymbol{v}-D\boldsymbol{s}(\boldsymbol{\theta}_0)\boldsymbol{t})K(\boldsymbol{v}), & \text{if } c_\varepsilon=\infty,\end{cases}$$

where $\delta(\cdot)$ is the Dirac delta function. For the random versions, the above results still hold with Θ replace by Θ_p and $\to\delta(\cdot)$ by $\xrightarrow{P}\delta(\cdot)$. Then for (2), by Fatou's lemma, $p_{acc,q}$ is bounded away from 0 in probability and hence $\Theta_p(1)$. For (3), $a_{n,\varepsilon}\varepsilon_n=1$ and $Q_{n,\varepsilon}\xrightarrow{P}K(0)=1$ by Fatou's lemma. Therefore $p_{acc,q}\xrightarrow{P}1$. ∎

Supplementary Material for "Improved Convergence of Regression Adjusted Approximate Bayesian Computation"

First some limit notations and conventions are given. For a series x_n, besides the limit notations $O(\cdot)$ and $o(\cdot)$, we use the notations, as $n \to \infty$, that $x_n = \Theta(\cdot)$ if there exists constants m and M such that $0 < m < |x_n/a_n| < M < \infty$, and $x_n = \Omega(a_n)$ if $|x_n/a_n| \to \infty$. For two sets A and B, the sum of integrals $\int_A f(\boldsymbol{x})\,d\boldsymbol{x} + \int_B f(\boldsymbol{x})\,d\boldsymbol{x}$ is written as $(\int_A + \int_B) f(\boldsymbol{x})\,d\boldsymbol{x}$. Denote any polynomial of the elements of $\boldsymbol{v}$ up to the order of l by $P_l(\boldsymbol{v})$. We say a square matrix A is bounded if there exists constants $\lambda_{min}(A)$ and $\lambda_{max}(A)$ such that $\lambda_{min}(A) \leq A \leq \lambda_{max}(A)$ and a rectangular matrix B is bounded if $B^T B$ is bounded. Denote the identity matrix with dimension d by I_d. Use the convention that $\boldsymbol{v}^2 = \boldsymbol{v}\boldsymbol{v}^T$ for the vector $\boldsymbol{v}$.

The following basic asymptotic results will be used throughout.

Lemma 4. *(i) For a series of random variables Z_n, if $Z_n \xrightarrow{\mathfrak{L}} Z$, $Z_n = O_p(1)$; (ii) For a series of continuous function $g_n(\boldsymbol{x})$, if $g_n(\boldsymbol{x}) = O(1)$ (or $\Theta(1)$) for any $\boldsymbol{x}$, then $g_n(Z_n) = O_p(1)$ (or $\Theta_p(1)$).*

Some notations regarding the approximate likelihood and posterior in ABC are given. For $A \subset \mathbb{R}^p$ and the scalar function $h(\boldsymbol{\theta}, \boldsymbol{s})$, let $\pi_A(h) = \int_A \int_{\mathbb{R}^d} h(\boldsymbol{\theta}, \boldsymbol{s})\pi(\boldsymbol{\theta}) f_n(\boldsymbol{s}|\boldsymbol{\theta}) K(\varepsilon_n^{-1}(\boldsymbol{s} - \boldsymbol{s}_{obs}))\varepsilon_n^{-d}\,d\boldsymbol{s}d\boldsymbol{\theta}$ and $\widetilde{\pi}_A(h) = \int_A \int_{\mathbb{R}^d} h(\boldsymbol{\theta}, \boldsymbol{s})\pi(\boldsymbol{\theta}) \widetilde{f}_n(\boldsymbol{s}|\boldsymbol{\theta}) K(\varepsilon_n^{-1}(\boldsymbol{s} - \boldsymbol{s}_{obs}))\varepsilon_n^{-d}\,d\boldsymbol{s}d\boldsymbol{\theta}$. With these notation, the ABC probability function $\Pi_\varepsilon(\boldsymbol{\theta} \in A|\boldsymbol{s}_{obs}) = \pi_A(1)/\pi_{\mathcal{P}}(1)$ and its normal counterpart $\widetilde{\Pi}_\varepsilon(\boldsymbol{\theta} \in A|\boldsymbol{s}_{obs}) = \widetilde{\pi}_A(1)/\widetilde{\pi}_{\mathcal{P}}(1)$. Notations from Li and Fearnhead (2016) will also be used.

Before proceeding the proof, we state several results from Li and Fearnhead (2015) that will be used throughout.

Lemma 5. *Assume conditions (C1)-(C6). Then the following results hold.*

(i) For any $\delta < \delta_0$, $\pi_{B_\delta^c}(1)$ and $\widetilde{\pi}_{B_\delta^c}(1)$ are $O_p(e^{-a_{n,\varepsilon}^{\alpha_\delta} c_\delta})$ for some positive constants c_δ and α_δ depending on δ.

(ii) $\pi_{B_\delta}(1) = \widetilde{\pi}_{B_\delta}(1)(1 + O_p(\alpha_n^{-1}))$ and $sup_{A\subset B_\delta} |\pi_A(1) - \widetilde{\pi}_A(1)| / \widetilde{\pi}_{B_\delta}(1) = O_p(\alpha_n^{-1})$.

(iii) If $\varepsilon_n = o(a_n^{-1/2})$, $\widetilde{\pi}_{B_\delta}(1)$ and $\pi_{B_\delta}(1)$ are $\Theta_p(a_{n,\varepsilon}^{d-p})$. Therefore, $\widetilde{\pi}_{\mathcal{P}}(1)$ and $\pi_{\mathcal{P}}(1)$ are $\Theta_p(a_{n,\varepsilon}^{d-p})$.

(iv) If $\varepsilon_n = o(a_n^{-1/2})$, $\boldsymbol{\theta}_\varepsilon = \widetilde{\boldsymbol{\theta}}_\varepsilon + o_p(a_{n,\varepsilon}^{-1})$. If $\varepsilon_n = o(a_n^{-3/5})$, $\boldsymbol{\theta}_\varepsilon = \widetilde{\boldsymbol{\theta}}_\varepsilon + o_p(a_n^{-1})$.

Proof. (i) is from Li and Fearnhead (2015, Lemma 3), (ii) is from Li and Fearnhead (2015, equation 7), (iii) is from Li and Fearnhead (2015, Lemma 6) and (iv) is from Li and Fearnhead (2015, Lemma 6 and Theorem 3.1). □

Proof of Lemma 1. (a)-(c) hold immediately by Lemma 5. □

Condition (IH) for Proposition 3.1. The posterior convergence results in Ghosal et al. (1995, Proposition 2) and Ghosh et al. (1994, Remark 2.4) assume the following set of conditions from Ibragimov and Has'minskii (1981, Theorem 10.2),

(IH1) For some $M > 0$, $m > 0$ and $\alpha > 0$,

$$E_{\boldsymbol{\theta}_0}[Z_n^{1/2}(\boldsymbol{t}_1) - Z_n^{1/2}(\boldsymbol{t}_2)]^2 \leq M(1 + R^m)\|\boldsymbol{t}_1 - \boldsymbol{t}_2\|^\alpha,$$

for all $\boldsymbol{t}_1, \boldsymbol{t}_2 \in U_n$ satisfying $\|\boldsymbol{t}_1\| \leq R$ and $\|\boldsymbol{t}_2\| \leq R$, where $U_n = \{a_{n,\varepsilon}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) : \boldsymbol{\theta} \in B_\delta\}$.

(IH2) For all $\boldsymbol{t} \in U_n$,

$$E_{\boldsymbol{\theta}_0} Z_n^{1/2}(\boldsymbol{t}) \leq exp\{-g_n(\|\boldsymbol{t}\|)\},$$

where $\{g_n\}$ is a sequence of real-value functions on $[0, \infty)$ satisfying the following: (a) for a fixed $n \geq 1$, $g_n(y) \uparrow \infty$ as $y \uparrow \infty$; (b) for any $N > 0$,

$$\lim_{y\to\infty, n\to\infty} y^N \exp\{-g_n(y)\} = 0;$$

(IH3) The finite-dimensional distributions of $\{Z_n(\boldsymbol{t}) : \boldsymbol{t} \in \mathbb{R}^p\}$ converge to those of a stochastic process $\{Z(\boldsymbol{t}) : \boldsymbol{t} \in \mathbb{R}^p\}$,

where the expectations are taken with respect to $\boldsymbol{s}_{obs}$.

To prove the lemmas for Theorem 3.1, some notations regarding the regression adjusted ABC posterior, similar to those defined previously, are needed. Consider the transformations $\boldsymbol{t}(\boldsymbol{\theta}) = a_{n,\varepsilon}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$ and $\boldsymbol{v}(\boldsymbol{s}) = \varepsilon^{-1}(\boldsymbol{s} - \boldsymbol{s}_{obs})$. For $A \subset \mathbb{R}^p$ and the scalar function $h(\boldsymbol{t}, \boldsymbol{v})$ in $\mathbb{R}^p \times \mathbb{R}^d$, let $\pi_{A,tv}(h) = \int_{t(A)} \int_{\mathbb{R}^d} h(\boldsymbol{t}, \boldsymbol{v})\pi(\boldsymbol{\theta}_0 + a_{n,\varepsilon}^{-1}\boldsymbol{t}) f_n(\boldsymbol{s}_{obs} + \varepsilon_n \boldsymbol{v}|\boldsymbol{\theta}_0 + a_{n,\varepsilon}^{-1}\boldsymbol{t})K(\boldsymbol{v})\, d\boldsymbol{v} d\boldsymbol{t}$ and $\widetilde{\pi}_{A,,tv}(h) = \int_{t(A)} \int_{\mathbb{R}^d} h(\boldsymbol{t}, \boldsymbol{v})\pi(\boldsymbol{\theta}_0 + a_{n,\varepsilon}^{-1}\boldsymbol{t}) \widetilde{f}_n(\boldsymbol{s}_{obs} + \varepsilon_n \boldsymbol{v}|\boldsymbol{\theta}_0 + a_{n,\varepsilon}^{-1}\boldsymbol{t})K(\boldsymbol{v})\, d\boldsymbol{v} d\boldsymbol{t}$ where $t(A)$ is the transformed A under $\boldsymbol{t}(\boldsymbol{\theta})$.

Proof of Lemma 2. From its definition, $\boldsymbol{\beta}_\varepsilon = Cov_\varepsilon[\boldsymbol{\theta}, \boldsymbol{s}]Var_\varepsilon[\boldsymbol{s}]^{-1}$. To evaluate the covariance matrices, we need to evaluate $\pi_{\mathbb{R}^p}((\boldsymbol{\theta} - \boldsymbol{\theta}_0)^{k_1}(\boldsymbol{s} - \boldsymbol{s}_{obs})^{k_2})/\pi_{\mathbb{R}^p}(1)$ for $(k_1, k_2) = (0,0), (1,0), (1,1), (0,1)$ and $(0,2)$.

First of all, we show that the integration in B_δ^c is ignorable. For any $\delta < \delta_0$, let $M_\delta = \min(M_1, \delta)$, it can be seen that $\pi_{B_\delta^c}((\boldsymbol{\theta} - \boldsymbol{\theta}_0)^{k_1}(\boldsymbol{s} - \boldsymbol{s}_{obs})^{k_2}) = O_p(e^{-a_{n,\varepsilon}^{\alpha_\delta} c_\delta})$ for some positive constants c_δ and α_δ by noting that

$$\begin{aligned}
&\pi_{B_\delta^c}((\boldsymbol{\theta} - \boldsymbol{\theta}_0)^{k_1}(\boldsymbol{s} - \boldsymbol{s}_{obs})^{k_2}) \\
\leq& \int_{B_\delta^c} \boldsymbol{\theta}^{k_1}\pi(\boldsymbol{\theta})\, d\boldsymbol{\theta} \int_{\mathbb{R}^d} \boldsymbol{s}^{k_2} K(\frac{\boldsymbol{s} - \boldsymbol{s}_{obs}}{\varepsilon_n})\varepsilon_n^{-d}\, d\boldsymbol{s} \left[\sup_{\boldsymbol{\theta}\in\mathcal{P}_0^c} \sup_{\|\boldsymbol{s}-\boldsymbol{s}_{obs}\|\leq M_\delta/2} f_n(\boldsymbol{s}|\boldsymbol{\theta}) + \sup_{\boldsymbol{\theta}\in B_\delta^c\setminus\mathcal{P}_0^c} \sup_{\|\boldsymbol{s}-\boldsymbol{s}_{obs}\|\leq M_\delta/2} f_n(\boldsymbol{s}|\boldsymbol{\theta})\right] \\
&+ \int_{B_\delta^c} \boldsymbol{\theta}^{k_1}\pi(\boldsymbol{\theta})\, d\boldsymbol{\theta} \int_{\mathbb{R}^d} \boldsymbol{s}^{k_2} f_n(\boldsymbol{s}|\boldsymbol{\theta})\, d\boldsymbol{s} K(\varepsilon_n^{-1} M_\delta/2)\varepsilon_n^{-d}
\end{aligned}$$

and following the arguments in the proof of Li and Fearnhead (2015, Lemma 3).

Then for the integration in B_δ, using the transformation $\boldsymbol{t}(\boldsymbol{\theta})$ and $\boldsymbol{v}(\boldsymbol{s})$ and Lemma 5,

$$\begin{aligned}
&\frac{\pi_{B_\delta}((\boldsymbol{\theta} - \boldsymbol{\theta}_0)^{k_1}(\boldsymbol{s} - \boldsymbol{s}_{obs})^{k_2})}{\pi_{B_\delta}(1)} \\
=& a_{n,\varepsilon}^{-k_1}\varepsilon_n^{k_2}\left[\frac{\widetilde{\pi}_{B_\delta,tv}(\boldsymbol{t}^{k_1}\boldsymbol{v}^{k_2})}{\widetilde{\pi}_{B_\delta,tv}(1)} + \alpha_n^{-1}\frac{\int_{t(B_\delta)}\int \boldsymbol{t}^{k_1}\boldsymbol{v}^{k_2}\pi(\boldsymbol{\theta}_0 + a_{n,\varepsilon}^{-1}\boldsymbol{t}) r_n(\boldsymbol{s}_{obs} + \varepsilon_n\boldsymbol{v}|\boldsymbol{\theta}_0 + a_{n,\varepsilon}^{-1}\boldsymbol{t})K(\boldsymbol{v})\, d\boldsymbol{v}d\boldsymbol{t}}{\widetilde{\pi}_{B_\delta,tv}(1)}\right](1 + O_p(\alpha_n^{-1})),
\end{aligned}$$

where $r_n(s|\boldsymbol{\theta})$ is the scaled remainder $\alpha_n[f_n(\boldsymbol{s}|\boldsymbol{\theta}) - \widetilde{f}_n(\boldsymbol{s}|\boldsymbol{\theta})]$. In the above, the remainder term in the square brackets is $O_p(\alpha_n^{-1})$ by Li and Fearnhead (2015, Lemma 7). Therefore we have

$$\frac{\pi_{B_\delta}((\boldsymbol{\theta}-\boldsymbol{\theta}_0)^{k_1}(\boldsymbol{s}-\boldsymbol{s}_{obs})^{k_2})}{\pi_{B_\delta}(1)} = a_{n,\varepsilon}^{-k_1}\varepsilon_n^{k_2}\left[\frac{\widetilde{\pi}_{B_\delta,tv}(\boldsymbol{t}^{k_1}\boldsymbol{v}^{k_2})}{\widetilde{\pi}_{B_\delta,tv}(1)} + O_p(\alpha_n^{-1})\right].$$

In the proof of Li and Fearnhead (2015, Theorem 3.1), $\widetilde{\pi}_{B_\delta,tv}(\boldsymbol{t})/\widetilde{\pi}_{B_\delta,tv}(1)$ is evaluated based on an expansion of $\widetilde{f}_n(\boldsymbol{s}_{obs}+\varepsilon_n\boldsymbol{v}|\boldsymbol{\theta}_0+a_{n,\varepsilon}^{-1}\boldsymbol{t})K(\boldsymbol{v})$ with $g_n(\boldsymbol{t},\boldsymbol{v})$, given in the appendix of this paper, as the leading term. Here $\widetilde{\pi}_{B_\delta,tv}(\boldsymbol{t}^{k_1}\boldsymbol{v}^{k_2})/\widetilde{\pi}_{B_\delta,tv}(1)$ can be evaluated similarly, and we have

$$\frac{\widetilde{\pi}_{B_\delta,tv}(\boldsymbol{t}^{k_1}\boldsymbol{v}^{k_2})}{\widetilde{\pi}_{B_\delta,tv}(1)} = O_p(a_{n,\varepsilon}^{-1}) + O_p(a_n^2\varepsilon_n^4) + \begin{cases} k_n^{-1}P^{-1}A^T\boldsymbol{W}_{obs} + P^{-1}A^TB_nE_{g_n^{(2)}}[\boldsymbol{v}], & (k_1,k_2)=(1,0), \\ k_n^{-1}P^{-1}A^T\boldsymbol{W}_{obs}E_{g_n^{(2)}}[\boldsymbol{v}] + P^{-1}A^TB_nE_{g_n^{(2)}}[\boldsymbol{v}\boldsymbol{v}^T], & (k_1,k_2)=(1,1), \\ E_{g_n^{(2)}}[\boldsymbol{v}], & (k_1,k_2)=(0,1), \\ E_{g_n^{(2)}}[\boldsymbol{v}\boldsymbol{v}^T], & (k_1,k_2)=(0,2), \end{cases}$$

$$\text{where } A = A(\boldsymbol{\theta}_0)^{-1/2}DS(\boldsymbol{\theta}_0),\ P = A^TA,\ B_n = \begin{cases} a_n\varepsilon_nA(\boldsymbol{\theta}_0)^{-1/2}, & c_\varepsilon<\infty \\ A(\boldsymbol{\theta}_0)^{-1/2}, & c_\varepsilon=\infty \end{cases},\ k_n = \begin{cases} 1, & c_\varepsilon<\infty \\ a_n\varepsilon_n, & c_\varepsilon=\infty \end{cases},$$

$$\text{and } g_n^{(2)}(\boldsymbol{v}) \propto k_n^{d-p}\exp\{-k_n^2(B_n\boldsymbol{v}+\boldsymbol{W}_{obs})^T(I-AP^{-1}A^T)(B_n\boldsymbol{v}+\boldsymbol{W}_{obs})/2\}K(\boldsymbol{v}).$$

Note that $E_{g_n^{(2)}}[\boldsymbol{v}\boldsymbol{v}^T] = \Theta_p(1)$. Then since $\alpha_n^{-1} = o(a_n^{-2/5})$, by algebra we have $Cov_\varepsilon(\boldsymbol{\theta},\boldsymbol{s}) = \varepsilon_n^2P^{-1}A^TA(\boldsymbol{\theta}_0)^{-1/2}Var_{g_n^{(2)}}[\boldsymbol{v}] + o_p(a_n^{-7/5}\varepsilon_n)$ and $Var_\varepsilon[\boldsymbol{s}] = \varepsilon_n^2Var_{g_n^{(2)}}[\boldsymbol{v}](1+o_p(a_n^{-2/5}))$. Then $\boldsymbol{\beta}_\varepsilon = \boldsymbol{\beta}_0 + o_p(a_n^{-2/5}) + O_p(a_n^{-7/5}\varepsilon_n^{-1})$ and the lemma holds. □

For $A\subset\mathbb{R}^p$ and $B\subset\mathbb{R}^d$, let $\pi(A,B) = \int_A\int_B \pi(\boldsymbol{\theta})f_n(\boldsymbol{s}|\boldsymbol{\theta})K(\varepsilon_n^{-1}(\boldsymbol{s}-\boldsymbol{s}_{obs}))\varepsilon_n^{-d}\,d\boldsymbol{s}d\boldsymbol{\theta}$ and $\widetilde{\pi}(A,B) = \int_A\int_B \pi(\boldsymbol{\theta})\widetilde{f}_n(\boldsymbol{s}|\boldsymbol{\theta})K(\varepsilon_n^{-1}(\boldsymbol{s}-\boldsymbol{s}_{obs}))\varepsilon_n^{-d}\,d\boldsymbol{s}d\boldsymbol{\theta}$. Let $\widetilde{\pi}_\varepsilon(\boldsymbol{\theta},\boldsymbol{s}|\boldsymbol{s}_{obs})$ be the density proportional to $\pi(\boldsymbol{\theta})\widetilde{f}(\boldsymbol{s}|\boldsymbol{\theta})K(\varepsilon_n^{-1}(\boldsymbol{s}-\boldsymbol{s}_{obs}))\varepsilon_n^{-d}$ and let $\widetilde{\boldsymbol{\theta}}_\varepsilon^*$ be the mean value of $\boldsymbol{\theta}^*$ under $\widetilde{\pi}_\varepsilon(\boldsymbol{\theta},\boldsymbol{s}|\boldsymbol{s}_{obs})$. Denote the mean values of $\boldsymbol{s}$ of $\pi_\varepsilon(\boldsymbol{\theta},\boldsymbol{s}|\boldsymbol{s}_{obs})$ and $\widetilde{\pi}_\varepsilon(\boldsymbol{\theta},\boldsymbol{s}|\boldsymbol{s}_{obs})$ by $\boldsymbol{s}_\varepsilon$ and $\widetilde{\boldsymbol{s}}_\varepsilon$ respectively.

Proof of Lemma 3. For (a), write $\Pi_\varepsilon(\boldsymbol{\theta}^*\in B_\delta^c|\boldsymbol{s}_{obs})$ as $\pi(\mathbb{R}^p,\{\boldsymbol{s}:\boldsymbol{\theta}^*(\boldsymbol{\theta},\boldsymbol{s})\in B_\delta^c\})/\pi(\mathbb{R}^p,\mathbb{R}^d)$, and by Lemma 5, we have $\pi(\mathbb{R}^p,\mathbb{R}^d) = \pi_{\mathcal{P}}(1) = \Theta_p(a_{n,\varepsilon}^{d-p})$. By the triangle inequality,

$$\pi(\mathbb{R}^p,\{\boldsymbol{s}:\boldsymbol{\theta}^*(\boldsymbol{\theta},\boldsymbol{s})\in B_\delta^c\}) \le \pi(B_{\delta/2}^c,\mathbb{R}^d) + \pi(B_{\delta/2},\{\boldsymbol{s}:\|\boldsymbol{\beta}_\varepsilon(\boldsymbol{s}-\boldsymbol{s}_{obs})\|\ge\frac{\delta}{2}\}).$$

By Lemma 5, the first term in the RHS is $\pi_{B_{\delta/2}^c}(1)$ and has the order $o_p(1)$. Then it remains to show that the order of the second term is $o_p(1)$. The derivation depends on whether $\boldsymbol{\beta}_\varepsilon$ goes to ∞ or is bounded as $n\to\infty$. When $\varepsilon_n = \Omega(a_n^{-7/5})$ or $\Theta(a_n^{-7/5})$, $\boldsymbol{\beta}_\varepsilon-\boldsymbol{\beta}_0 = o_p(1)$ and so the coordinates of $\boldsymbol{\beta}_\varepsilon$ is upper bounded. Denote a constant upper bound of coordinates of $\boldsymbol{\beta}_\varepsilon$ by β_{sup}. Then obviously

$$\pi(B_{\delta/2},\{\boldsymbol{s}:\|\boldsymbol{\beta}_\varepsilon(\boldsymbol{s}-\boldsymbol{s}_{obs})\|\ge\frac{\delta}{2}\}) \le K(\varepsilon^{-1}\frac{\delta}{2\beta_{sup}})\varepsilon_n^{-d},$$

and by Condition (C2)(iv), it is $o_p(1)$. When $\varepsilon_n = o(a_n^{-7/5})$, $\boldsymbol{\beta}_\varepsilon$ is unbounded and β_{sup} does not exist. To see that the result is the same as before, first note that $sup_{\boldsymbol{s}\in\mathbb{R}^d}\int_{B_{\delta/2}}\pi(\boldsymbol{\theta})f_n(\boldsymbol{s}|\boldsymbol{\theta})\,d\boldsymbol{\theta}$ is bounded as $n\to\infty$. This can be shown by decomposing $f_n(\boldsymbol{s}|\boldsymbol{\theta})$ according to Condition

(C4), and then integrating the leading term, the normal limit of $f_n(\boldsymbol{s}|\boldsymbol{\theta})$, and the remainder term, bounded by a transformed $r_{max}(\cdot)$, with respect to $\boldsymbol{\theta}$. Then up to a constant factor, $\pi(B_{\delta/2}, \{\boldsymbol{s} : \|\boldsymbol{\beta}_\varepsilon(\boldsymbol{s}-\boldsymbol{s}_{obs})\| \geq \frac{\delta}{2}\})$ is bounded by

$$\int_{\|\boldsymbol{\beta}_\varepsilon(\boldsymbol{s}-\boldsymbol{s}_{obs})\|\geq\frac{\delta}{2}} K(\frac{\boldsymbol{s}-\boldsymbol{s}_{obs}}{\varepsilon_n})\varepsilon_n^{-d}\, d\boldsymbol{s} = \int_{\|\boldsymbol{\beta}_\varepsilon\varepsilon_n\boldsymbol{v}\|\geq\frac{\delta}{2}} K(\boldsymbol{v})\, d\boldsymbol{v}$$

which is $o_p(1)$ since $\boldsymbol{\beta}_\varepsilon\varepsilon_n = O_p(a_n^{-7/5})$. Therefore $\Pi_\varepsilon(\boldsymbol{\theta}^* \in B_\delta^c|\boldsymbol{s}_{obs}) = o_p(1)$. For $\widehat{\Pi}_\varepsilon(\boldsymbol{\theta}^* \in B_\delta^c|\boldsymbol{s}_{obs})$, since the support of its prior is B_δ, it is also $o_p(1)$. Therefore (a) holds.

For (b) and (c), first we prove the similar results for $\widetilde{\Pi}_\varepsilon$ and $\widetilde{\boldsymbol{\theta}}_\varepsilon^*$, denoted by (b') and (c'), then show that the impact of changing the integration areas from $\mathbb{R}^d$ is negligible, and finally show that (b') and (c') imply that (b) and (c) hold.

For (b'), we have

$$\begin{aligned}
&\sup_{A\in\mathscr{B}^p}\left|\Pi_\varepsilon(\boldsymbol{\theta}^* \in A_\theta\cap B_\delta|\boldsymbol{s}_{obs}) - \widetilde{\Pi}_\varepsilon(\boldsymbol{\theta}^* \in A_\theta\cap B_\delta|\boldsymbol{s}_{obs})\right| \\
=&\frac{\sup_{A\in\mathscr{B}^p}|\pi(\mathbb{R}^p,\{\boldsymbol{s}:\boldsymbol{\theta}^*(\boldsymbol{\theta},\boldsymbol{s})\in A_\theta\cap B_\delta\}) - \widetilde{\pi}(\mathbb{R}^p,\{\boldsymbol{s}:\boldsymbol{\theta}^*(\boldsymbol{\theta},\boldsymbol{s})\in A_\theta\cap B_\delta\})|}{\widetilde{\pi}_{B_\delta}(1)} + o_p(1) \\
\leq&\frac{\sup_{A\in\mathscr{B}^p}|\pi(B_\delta,\{\boldsymbol{s}:\boldsymbol{\theta}^*(\boldsymbol{\theta},\boldsymbol{s})\in A_\theta\cap B_\delta\}) - \widetilde{\pi}(B_\delta,\{\boldsymbol{s}:\boldsymbol{\theta}^*(\boldsymbol{\theta},\boldsymbol{s})\in A_\theta\cap B_\delta\})|}{\widetilde{\pi}_{B_\delta}(1)} + \frac{\pi_{B_\delta^c}(1)+\widetilde{\pi}_{B_\delta^c}(1)}{\widetilde{\pi}_{B_\delta}(1)} + o_p(1) \\
\leq&\frac{\int_{B_\delta}\int_{\mathbb{R}^d}\pi(\boldsymbol{\theta})|f(\boldsymbol{s}|\boldsymbol{\theta}) - \widetilde{f}(\boldsymbol{s}|\boldsymbol{\theta})|K(\varepsilon_n^{-1}(\boldsymbol{s}-\boldsymbol{s}_{obs}))\varepsilon_n^{-d}\,d\boldsymbol{s}d\boldsymbol{\theta}}{\widetilde{\pi}_{B_\delta}(1)} + o_p(1).
\end{aligned}$$

Then by Li and Fearnhead (2015, Lemma 7), (b') holds.

For (c'), we have $a_n(\boldsymbol{\theta}_\varepsilon^* - \widetilde{\boldsymbol{\theta}}_\varepsilon^*) = a_n(\boldsymbol{\theta}_\varepsilon - \widetilde{\boldsymbol{\theta}}_\varepsilon) + a_n(\boldsymbol{s}_\varepsilon - \widetilde{\boldsymbol{s}}_\varepsilon)$. By Lemma 5, $a_n(\boldsymbol{\theta}_\varepsilon - \widetilde{\boldsymbol{\theta}}_\varepsilon) = o_p(1)$. For $a_n(\boldsymbol{s}_\varepsilon - \widetilde{\boldsymbol{s}}_\varepsilon)$, similar to the arguments of the proof of Lemma 2,

$$\begin{aligned}
\boldsymbol{s}_\varepsilon - \boldsymbol{s}_{obs} &= \varepsilon_n\left[\frac{\widetilde{\pi}^*_{B_\delta}(\boldsymbol{v})}{\widetilde{\pi}^*_{B_\delta}(1)} + O_p(\alpha_n^{-1})\right](1+O_p(e^{-a_{n,\varepsilon}^{\alpha_\delta}c_\delta})) \\
\text{and } \widetilde{\boldsymbol{s}}_\varepsilon - \boldsymbol{s}_{obs} &= \varepsilon_n\frac{\widetilde{\pi}^*_{B_\delta}(\boldsymbol{v})}{\widetilde{\pi}^*_{B_\delta}(1)}(1+O_p(e^{-a_{n,\varepsilon}^{\alpha_\delta}c_\delta})).
\end{aligned}$$

Then $a_n(\boldsymbol{s}_\varepsilon - \widetilde{\boldsymbol{s}}_\varepsilon) = O_p(\alpha_n^{-1}a_n\varepsilon_n)$ which is $o_p(1)$ if $\varepsilon_n = o(a_n^{-3/5})$.

For any constant $M_\delta > \delta$, (b') and (c') still hold when the integrated areas of $\boldsymbol{s}$ are changed from $\mathbb{R}^d$ to $\{\boldsymbol{s} : \|\boldsymbol{\beta}_\varepsilon(\boldsymbol{s}-\boldsymbol{s}_{obs})\| \leq M_\delta\}$. This can be seen by the following decomposition,

$$\widetilde{\Pi}_\varepsilon(\boldsymbol{\theta}^* \in A\cap B_\delta|\boldsymbol{s}_{obs}) = \frac{\pi(\{\boldsymbol{\theta}:\boldsymbol{\theta}^*(\boldsymbol{\theta},\boldsymbol{s})\in A\cap B_\delta\},\{\boldsymbol{s}:\|\boldsymbol{\beta}_\varepsilon(\boldsymbol{s}-\boldsymbol{s}_{obs})\|\leq M_\delta\})}{\pi(\{\boldsymbol{\theta}:\boldsymbol{\theta}^*(\boldsymbol{\theta},\boldsymbol{s})\in B_\delta\},\{\boldsymbol{s}:\|\boldsymbol{\beta}_\varepsilon(\boldsymbol{s}-\boldsymbol{s}_{obs})\|\leq M_\delta\})}(1+o_p(1)),$$

which holds because $\pi(\{\boldsymbol{\theta}:\boldsymbol{\theta}^*(\boldsymbol{\theta},\boldsymbol{s})\in B_\delta\},\{\boldsymbol{s}:\|\boldsymbol{\beta}_\varepsilon(\boldsymbol{s}-\boldsymbol{s}_{obs})\| > M_\delta\})/\pi(\mathbb{R}^p,\mathbb{R}^d) \leq \pi(B^c_{M_\delta-\delta},\mathbb{R}^d)/\pi(\mathbb{R}^p,\mathbb{R}^d) = o_p(1)$.

To see that (b') and (c') imply (b) and (c), since $\widetilde{\Pi}_\varepsilon$ and $\widehat{\Pi}_\varepsilon$ only differs by the prior, it is sufficient to show that the relative difference between the prior densities goes to 0 in the rate of $o(a_n^{-1})$ uniformly , i.e.

$$a_n\sup_{\boldsymbol{\theta}^*\in B_\delta}\left|\frac{\pi(\boldsymbol{\theta}^*) - \pi^{(K)}(\boldsymbol{\theta}^*;\varepsilon_n)}{\pi^{(K)}(\boldsymbol{\theta}^*;\varepsilon_n)}\right| = o(1), \tag{9}$$

where $\pi^{(K)}(\boldsymbol{\theta}^*;\varepsilon_n) = \int_{\mathbb{R}^d} \pi(\boldsymbol{\theta}^* + \boldsymbol{\beta}_\varepsilon \varepsilon_n \boldsymbol{v}) K(\boldsymbol{v})\, d\boldsymbol{v}$. For the numerator in (9), by dividing $\mathbb{R}^d$ into $\{\boldsymbol{v} : \|\varepsilon_n \boldsymbol{v}\| \leq M\}$ for some $M > 0$ and its complement and the Taylor expansion to the square term, we have

$$\begin{aligned} & a_n \sup_{\boldsymbol{\theta}^* \in B_\delta} \left| \pi^{(K)}(\boldsymbol{\theta}^*;\varepsilon_n) - \pi(\boldsymbol{\theta}^*) \right| \\ \leq & a_n \varepsilon_n^2 \sup_{\boldsymbol{\theta}^* \in B_\delta} \left| \int_{\|\varepsilon_n \boldsymbol{v}\| \leq M} \boldsymbol{v}^T \boldsymbol{\beta}_\varepsilon^T H\pi(\boldsymbol{\theta}^* + \boldsymbol{\beta}_\varepsilon \varepsilon_n' \boldsymbol{v}) \boldsymbol{\beta}_\varepsilon \boldsymbol{v} K(\boldsymbol{v})\, d\boldsymbol{v} \right| \\ & + a_n \sup_{\boldsymbol{\theta}^* \in B_\delta} \int_{\|\varepsilon_n \boldsymbol{v}\| > M} |\pi(\boldsymbol{\theta}^* + \boldsymbol{\beta}_\varepsilon \varepsilon_n \boldsymbol{v}) - \pi(\boldsymbol{\theta}^*)|\, K(\boldsymbol{v})\, d\boldsymbol{v}, \end{aligned}$$

where $|\varepsilon_n'| < \varepsilon_n$. Since $H\pi(\boldsymbol{\theta}^* + \boldsymbol{\beta}_\varepsilon \varepsilon_n' \boldsymbol{v})$ is bounded for $\boldsymbol{\theta}^* \in B_\delta$ and $\|\varepsilon_n \boldsymbol{v}\| \leq M$ when δ and M are small enough, the first term in the above is $O(a_n \varepsilon_n^2)$. Since $\pi(\boldsymbol{\theta})$ is bounded for $\boldsymbol{\theta} \in \mathbb{R}^p$, the second term decreases in the same rate as $a_n \int_{\|\boldsymbol{v}\| > \varepsilon_n^{-1} M} K(\boldsymbol{v})\, d\boldsymbol{v}$. By Chebyshev's inequality, the rate is $O(a_n \varepsilon_n^2)$ which is $o_p(1)$ with $\varepsilon_n = o(a_n^{-3/5})$.

For the denominator in (9), for small enough δ such that $\inf_{\boldsymbol{\theta}^* \in B_\delta} \pi(\boldsymbol{\theta}^*) > 0$, $\pi^{(K)}(\boldsymbol{\theta}^*;\varepsilon_n)$ is obviously lower bounded by some positive constant. Therefore

$$\sup_{\boldsymbol{\theta}^* \in B_\delta} \left| \frac{\pi(\boldsymbol{\theta}^*) - \pi^{(K)}(\boldsymbol{\theta}^*;\varepsilon_n)}{\pi^{(K)}(\boldsymbol{\theta}^*;\varepsilon_n)} \right| = O(\varepsilon_n^2) = o(a_n^{-1}),$$

and (b') and (c') imply (b) and (c). □

Consider conditions (IH*) for $Z_n^*(\boldsymbol{t})$ by replacing $Z_n(\boldsymbol{t})$ and U_n in (IH) with $Z_n^*(\boldsymbol{t})$ and $U_n^* = \{a_n(\boldsymbol{\theta} - \boldsymbol{\theta}_0) : \boldsymbol{\theta} \in B_\delta\}$.

Proof of (IH1*) and (IH2*) for Theorem 3.1. Denote the set $\{\boldsymbol{v} : \|\boldsymbol{\beta}_\varepsilon \varepsilon_n \boldsymbol{v}\| \leq M_\delta\}$ by $V_{M_\delta,n}$. Throughout this proof, consider small enough δ and $M_\delta > \delta$, such that for $\boldsymbol{\theta} \in B_{M_\delta+\delta}$, $A(\boldsymbol{\theta}) \in (A_{min}, A_{max})$ for finite square matrices A_{min}, A_{max}, $\inf_{\boldsymbol{\theta} \in B_{\delta+M_\delta}} \pi(\boldsymbol{\theta}) > 0$, $\sup_{\boldsymbol{\theta} \in B_{\delta+M_\delta}} D\pi(\boldsymbol{\theta}) < \infty$ and $D\boldsymbol{s}(\boldsymbol{\theta}) D\boldsymbol{s}(\boldsymbol{\theta})^T \in (m_{D\boldsymbol{s}}^2 I_d, M_{D\boldsymbol{s}}^2 I_d)$ for constants $m_{D\boldsymbol{s}}, M_{D\boldsymbol{s}}$.

For (IH1*), noting that the expectation is taken with respect to $\boldsymbol{s}_{obs} \sim \widetilde{f}_\varepsilon^*(\cdot|\boldsymbol{\theta}_0)$, we have the alternative expression

$$E_{\boldsymbol{\theta}_0}[Z_n^{*1/2}(\boldsymbol{t}_1) - Z_n^{*1/2}(\boldsymbol{t}_2)]^2 = 2[1 - \int_{\mathbb{R}^d} \widetilde{f}_\varepsilon^*(\boldsymbol{s}|\boldsymbol{\theta}_0 + a_n^{-1}\boldsymbol{t}_1)^{1/2} \widetilde{f}_\varepsilon^*(\boldsymbol{s}|\boldsymbol{\theta}_0 + a_n^{-1}\boldsymbol{t}_2)^{1/2}\, d\boldsymbol{s}].$$

Let $\boldsymbol{\theta}(\boldsymbol{t},\boldsymbol{v}) = \boldsymbol{\theta}_0 + a_n^{-1}\boldsymbol{t} + \boldsymbol{\beta}_\varepsilon \varepsilon_n \boldsymbol{v}$. Then by Cauchy-Schwartz inequality, the integral in the above expression is lower bounded as

$$\begin{aligned} & \int_{\mathbb{R}^d} \widetilde{f}_\varepsilon^*(\boldsymbol{s}|\boldsymbol{\theta}_0 + a_n^{-1}\boldsymbol{t}_1)^{1/2} \widetilde{f}_\varepsilon^*(\boldsymbol{s}|\boldsymbol{\theta}_0 + a_n^{-1}\boldsymbol{t}_2)^{1/2}\, d\boldsymbol{s} \\ \geq & \int_{V_{M_\delta,n}} \int_{\mathbb{R}^d} \widetilde{f}_n(\boldsymbol{s} + \varepsilon\boldsymbol{v}|\boldsymbol{\theta}(\boldsymbol{t}_1,\boldsymbol{v}))^{1/2} \widetilde{f}_n(\boldsymbol{s} + \varepsilon\boldsymbol{v}|\boldsymbol{\theta}(\boldsymbol{t}_2,\boldsymbol{v}))^{1/2}\, d\boldsymbol{s} \Pi_{\pi,K}(\boldsymbol{t}_1,\boldsymbol{t}_2,\boldsymbol{v})\, d\boldsymbol{v}, \end{aligned} \tag{10}$$

$$\text{where } \Pi_{\pi,K}(\boldsymbol{t}_1,\boldsymbol{t}_2,\boldsymbol{v}) = \frac{\pi(\boldsymbol{\theta}(\boldsymbol{t}_1,\boldsymbol{v}))^{1/2} \pi(\boldsymbol{\theta}(\boldsymbol{t}_2,\boldsymbol{v}))^{1/2} K(\boldsymbol{v})}{[\int_{V_{M_\delta,n}} \pi(\boldsymbol{\theta}(\boldsymbol{t}_1,\boldsymbol{v})) K(\boldsymbol{v})\, d\boldsymbol{v}]^{1/2} [\int_{V_{M_\delta,n}} \pi(\boldsymbol{\theta}(\boldsymbol{t}_2,\boldsymbol{v})) K(\boldsymbol{v})\, d\boldsymbol{v}]^{1/2}}.$$

For any $\boldsymbol{t} \in U_n^*$ and $\boldsymbol{v} \in V_{M_\delta,n}$, $\boldsymbol{\theta}(\boldsymbol{t},\boldsymbol{v}) \in B_{M_\delta+\delta}$. For bounding purpose, we can assume that $A(\boldsymbol{\theta}(\boldsymbol{t},\boldsymbol{v})) \equiv I_d$ without loss of generality, because $\widetilde{f}_n(\boldsymbol{s} + \varepsilon\boldsymbol{v}|\boldsymbol{\theta}(\boldsymbol{t},\boldsymbol{v}))$ is lower bounded as

$$\widetilde{f}_n(\boldsymbol{s}+\varepsilon\boldsymbol{v}|\boldsymbol{\theta}(\boldsymbol{t}_1,\boldsymbol{v})) \geq \frac{a_n^d}{(2\pi)^{d/2}|A_{max}|^{1/2}}\exp\{-\frac{a_n^2}{2}\|A_{max}^{1/2}(\boldsymbol{s}+\varepsilon\boldsymbol{v}-s(\boldsymbol{\theta}(\boldsymbol{t},\boldsymbol{v})))\|^2\}.$$

Note that

$$\int_{\mathbb{R}^d}\widetilde{f}_n(\boldsymbol{s}+\varepsilon\boldsymbol{v}|\boldsymbol{\theta}(\boldsymbol{t}_1,\boldsymbol{v}))^{1/2}\widetilde{f}_n(\boldsymbol{s}+\varepsilon\boldsymbol{v}|\boldsymbol{\theta}(\boldsymbol{t}_2,\boldsymbol{v}))^{1/2}\,d\boldsymbol{s} = \exp\{-\frac{a_n^2}{8}\|\boldsymbol{s}(\boldsymbol{\theta}(\boldsymbol{t}_1,\boldsymbol{v}))-\boldsymbol{s}(\boldsymbol{\theta}(\boldsymbol{t}_2,\boldsymbol{v}))\|^2\}.$$

By plugging the above into (10), we have the following upper bound,

$$\begin{aligned} E_{\boldsymbol{\theta}_0}[Z_n^{*1/2}(\boldsymbol{t}_1)-Z_n^{*1/2}(\boldsymbol{t}_2)]^2 \leq\, & 2\int_{V_{M_\delta,n}}[1-\exp\{-\frac{a_n^2}{8}\|\boldsymbol{s}(\boldsymbol{\theta}(\boldsymbol{t}_1,\boldsymbol{v}))-\boldsymbol{s}(\boldsymbol{\theta}(\boldsymbol{t}_2,\boldsymbol{v}))\|^2\}]\Pi_{\pi,K}(\boldsymbol{t}_1,\boldsymbol{t}_2,\boldsymbol{v})\,d\boldsymbol{v} \\ & + 2(1-\int_{V_{M_\delta,n}}\Pi_{\pi,K}(\boldsymbol{t}_1,\boldsymbol{t}_2,\boldsymbol{v})\,d\boldsymbol{v}). \end{aligned} \tag{11}$$

For the first term in the above, from the fact $1-e^{-x}\leq x$ for $x>0$ and the inequality $\Pi_{\pi,K}(\boldsymbol{t}_1,\boldsymbol{t}_2,\boldsymbol{v}) \leq K(\boldsymbol{v})\sup_{\boldsymbol{\theta}\in B_{M_\delta+\delta}}\pi(\boldsymbol{\theta})/\inf_{\boldsymbol{\theta}\in B_{M_\delta+\delta}}\pi(\boldsymbol{\theta})$, it is upper bounded by

$$\int_{V_{M_\delta,n}} a_n^2\|\boldsymbol{s}(\boldsymbol{\theta}(\boldsymbol{t}_1,\boldsymbol{v}))-\boldsymbol{s}(\boldsymbol{\theta}(\boldsymbol{t}_2,\boldsymbol{v}))\|^2K(\boldsymbol{v})\,d\boldsymbol{v}\frac{\pi_{\sup/\inf}}{8},\ \text{where}\ \pi_{\sup/\inf}=\frac{\sup_{\boldsymbol{\theta}\in B_{M_\delta+\delta}}\pi(\boldsymbol{\theta})}{\inf_{\boldsymbol{\theta}\in B_{M_\delta+\delta}}\pi(\boldsymbol{\theta})}.$$

Then applying Taylor expansion at $\boldsymbol{t}_1-\boldsymbol{t}_2=0$, the first term of (11) is upper bounded by

$$\frac{M_{D\boldsymbol{s}}^2\pi_{\sup/\inf}}{32}\|\boldsymbol{t}_1-\boldsymbol{t}_2\|^2.$$

For the second term of (11), we have

$$\begin{aligned} & |1-\int_{V_{M_\delta,n}}\Pi_{\pi,K}(\boldsymbol{t}_1,\boldsymbol{t}_2,\boldsymbol{v})\,d\boldsymbol{v}| \\ \leq & \frac{|\int_{V_{M_\delta,n}}\pi(\boldsymbol{\theta}(\boldsymbol{t}_1,\boldsymbol{v}))^{1/2}\pi(\boldsymbol{\theta}(\boldsymbol{t}_2,\boldsymbol{v}))^{1/2}K(\boldsymbol{v})\,d\boldsymbol{v}-[\int_{V_{M_\delta,n}}\pi(\boldsymbol{\theta}(\boldsymbol{t}_1,\boldsymbol{v}))K(\boldsymbol{v})\,d\boldsymbol{v}]^{1/2}[\int_{V_{M_\delta,n}}\pi(\boldsymbol{\theta}(\boldsymbol{t}_2,\boldsymbol{v}))K(\boldsymbol{v})\,d\boldsymbol{v}]^{1/2}|}{\inf_{\boldsymbol{\theta}\in B_{M_\delta+\delta}}\pi(\boldsymbol{\theta})\int_{V_{M_\delta,n}}K(\boldsymbol{v})\,d\boldsymbol{v}}. \end{aligned}$$

By applying Taylor expansion at $\boldsymbol{t}_1-\boldsymbol{t}_2=0$ on the numerator to the third order, it can be seen that the first and second order are equal to 0. Then we have

$$|1-\int_{V_{M_\delta,n}}\Pi_{\pi,K}(\boldsymbol{t}_1,\boldsymbol{t}_2,\boldsymbol{v})\,d\boldsymbol{v}| \leq c_4a_n^{-2}\|\boldsymbol{t}_1-\boldsymbol{t}_2\|^2$$

for some positive constant c_4. Therefore combining these two upper bounds, (IH1*) holds.

For (IH2*), by applying Jensen's inequality twice, $E_{\boldsymbol{\theta}_0}Z_n^{1/2}(\boldsymbol{t})$ is upper bounded as the following,

$$\begin{aligned} E_{\boldsymbol{\theta}_0}Z_n^{1/2}(\boldsymbol{t}) &= \int_{\mathbb{R}^d}\widetilde{f}_\varepsilon^*(\boldsymbol{s}|\boldsymbol{\theta}_0+a_n^{-1}\boldsymbol{t})^{1/2}\frac{\widetilde{f}_\varepsilon^*(\boldsymbol{s}|\boldsymbol{\theta}_0)^{1/2}}{\int_{\mathbb{R}^d}\widetilde{f}_\varepsilon^*(\boldsymbol{s}|\boldsymbol{\theta}_0)^{1/2}\,d\boldsymbol{s}}\,d\boldsymbol{s}\cdot\int_{\mathbb{R}^d}\widetilde{f}_\varepsilon^*(\boldsymbol{s}|\boldsymbol{\theta}_0)^{1/2}\,d\boldsymbol{s} \\ &\leq [\int_{\mathbb{R}^d}\widetilde{f}_\varepsilon^*(\boldsymbol{s}|\boldsymbol{\theta}_0+a_n^{-1}\boldsymbol{t})\widetilde{f}_\varepsilon^*(\boldsymbol{s}|\boldsymbol{\theta}_0)^{1/2}\,d\boldsymbol{s}]^{1/2}[\int_{\mathbb{R}^d}\widetilde{f}_\varepsilon^*(\boldsymbol{s}|\boldsymbol{\theta}_0)^{1/2}\,d\boldsymbol{s}]^{1/2} \\ &\leq [a_n^{-d}\int_{\mathbb{R}^d}\widetilde{f}_\varepsilon^*(\boldsymbol{s}|\boldsymbol{\theta}_0+a_n^{-1}\boldsymbol{t})\widetilde{f}_\varepsilon^*(\boldsymbol{s}|\boldsymbol{\theta}_0)\,d\boldsymbol{s}]^{1/4}[a_n^{d/2}\int_{\mathbb{R}^d}\widetilde{f}_\varepsilon^*(\boldsymbol{s}|\boldsymbol{\theta}_0)^{1/2}\,d\boldsymbol{s}]^{1/2}. \end{aligned} \tag{12}$$

Similar to (IH1*), for bounding purpose assume that $A(\boldsymbol{\theta}) \equiv I$ for $\boldsymbol{\theta} \in B_{M_\delta+\delta}$ without loss of generality. In the following, the upper bounds for the two integrals in (12) are provided.

For the second multiplier in (12), in $\widetilde{f}_\varepsilon^*(\boldsymbol{s}|\boldsymbol{\theta}_0)$, the argument of the normal density is $a_n[\boldsymbol{s} - \boldsymbol{s}(\boldsymbol{\theta}_0 + \boldsymbol{\beta}_\varepsilon \varepsilon_n \boldsymbol{v})] + a_n\varepsilon_n \boldsymbol{v}$ which can be expanded as

$$\boldsymbol{v}' + a_n\varepsilon_n D\boldsymbol{s}^\perp(\boldsymbol{v})\boldsymbol{v},$$
$$\text{where } \boldsymbol{v}' = a_n[\boldsymbol{s} - \boldsymbol{s}(\boldsymbol{\theta}_0)] \text{ and } D\boldsymbol{s}^\perp(\boldsymbol{v}) = I - D\boldsymbol{s}(\boldsymbol{\theta}_0 + \boldsymbol{\theta}_0 + \boldsymbol{\beta}_\varepsilon \varepsilon_n' \boldsymbol{v})\boldsymbol{\beta}_\varepsilon.$$

Then we have

$$a_n^{d/2}\int_{\mathbb{R}^d} \widetilde{f}_\varepsilon^*(\boldsymbol{s}|\boldsymbol{\theta}_0)^{1/2}\, d\boldsymbol{s} \le \pi_{\frac{\sup}{\inf}} \int_{\mathbb{R}^d}[\int_{V_{M_\delta,n}} N(\boldsymbol{v}' + a_n\varepsilon_n D\boldsymbol{s}^\perp(\boldsymbol{v})\boldsymbol{v}; 0, I_d)K(\boldsymbol{v})\, d\boldsymbol{v}]^{1/2}\, d\boldsymbol{v}'.$$

By dividing $\mathbb{R}^d$ into $R_1 = \{\boldsymbol{v} : \|a_n\varepsilon_n D\boldsymbol{s}^\perp(\boldsymbol{v})\boldsymbol{v}\| \le \|\boldsymbol{T}\|/2\}$ and R_1^c, the integral in the RHS of the above is smaller than

$$\int_{\mathbb{R}^d}[N(\frac{\boldsymbol{T}}{2}; 0, I_d)\int_{V_{M_\delta,n}\cap R_1} K(\boldsymbol{v})\, d\boldsymbol{v} + \frac{1}{(a_n\varepsilon_n)^d}K(\frac{\|\boldsymbol{T}\|}{2a_n\varepsilon_n\sqrt{M_{D\boldsymbol{s}^\perp}}})\int_{V_{M_\delta,n}\cap R_1^c} N(\frac{1}{2}m_{D\boldsymbol{s}^\perp}\boldsymbol{v}, 0, \frac{1}{a_n\varepsilon_n}I_d)\, d\boldsymbol{v}]^{1/2}\, d\boldsymbol{T}$$

where $M_{D\boldsymbol{s}^\perp}$ and $m_{D\boldsymbol{s}^\perp}$ are the bounds of $D\boldsymbol{s}^\perp(\boldsymbol{v})$ for $\boldsymbol{v} \in V_{M_\delta,n}$. The above is $\Theta_p(1)$ by the fact that $(a+b)^\gamma \le a^\gamma + b^\gamma$ for $\gamma < 1$. Therefore the second multiplier in the RHS of (12) is $\Theta_p(1)$.

For the first multiplier in (12), by integrating out $\boldsymbol{s}$, we have

$$\begin{aligned} &a_n^{-d}\int_{\mathbb{R}^d} \widetilde{f}_\varepsilon^*(\boldsymbol{s}|\boldsymbol{\theta}_0 + a_n^{-1}\boldsymbol{t}_2)\widetilde{f}_\varepsilon^*(\boldsymbol{s}|\boldsymbol{\theta}_0)\, d\boldsymbol{s} \\ \le&\pi^2_{\sup/\inf}\int_{V_{M_\delta,n}\times V_{M_\delta,n}} N(a_n[\boldsymbol{s}(\boldsymbol{\theta}(\boldsymbol{t},\boldsymbol{v})) - \boldsymbol{s}(\boldsymbol{\theta}(0,\boldsymbol{w}))]; a_n\varepsilon_n(\boldsymbol{v}-\boldsymbol{w}), 2I_d)K(\boldsymbol{v})K(\boldsymbol{w})\, d\boldsymbol{w}d\boldsymbol{v}. \end{aligned} \tag{13}$$

and this bound needs to be evaluated differently for $c_\varepsilon < \infty$ and $c_\varepsilon = \infty$.

When $c_\varepsilon < \infty$, since $\boldsymbol{\beta}_\varepsilon a_n\varepsilon_n$ is $O_p(1)$, denote its upper bound in probability by $M_{\boldsymbol{\beta}}$. Consider dividing $\mathbb{R}^d$ into $R_2 = \{(\boldsymbol{v},\boldsymbol{w}) : c_5\|\boldsymbol{v}-\boldsymbol{w}\| \le \|\boldsymbol{t}\|\}$, where $c_5 = (M_{\boldsymbol{\beta}} + 4m_{D\boldsymbol{s}}^{-1}c_\varepsilon)$, and R_2^c. Since $a_n[\boldsymbol{s}(\boldsymbol{\theta}(\boldsymbol{t},\boldsymbol{v})) - \boldsymbol{s}(\boldsymbol{\theta}(0,\boldsymbol{w}))] \ge m_{D\boldsymbol{s}}\|\boldsymbol{t} + \boldsymbol{\beta}_\varepsilon a_n\varepsilon_n(\boldsymbol{v}-\boldsymbol{w})\|$ by Taylor expansion at $\boldsymbol{\theta}(\boldsymbol{t},\boldsymbol{v}) - \boldsymbol{\theta}(0,\boldsymbol{w}) = 0$, in R_2 it holds that $\|a_n[\boldsymbol{s}(\boldsymbol{\theta}(\boldsymbol{t},\boldsymbol{v})) - \boldsymbol{s}(\boldsymbol{\theta}(0,\boldsymbol{w}))]\|/2 \ge \|a_n\varepsilon_n(\boldsymbol{v}-\boldsymbol{w})\|$ for large enough n. Then

$$\begin{aligned} &\int_{(V_{M_\delta,n}\times V_{M_\delta,n})\cap R_2} N(a_n[\boldsymbol{s}(\boldsymbol{\theta}(\boldsymbol{t},\boldsymbol{v})) - \boldsymbol{s}(\boldsymbol{\theta}(0,\boldsymbol{w}))]; a_n\varepsilon_n(\boldsymbol{v}-\boldsymbol{w}), 2I_d)K(\boldsymbol{v})K(\boldsymbol{w})\, d\boldsymbol{w}d\boldsymbol{v} \\ \le&\int_{(V_{M_\delta,n}\times V_{M_\delta,n})\cap R_2} \frac{1}{(4\pi)^{d/2}}\exp\{-\frac{a_n^2}{8}\|\boldsymbol{s}(\boldsymbol{\theta}(\boldsymbol{t},\boldsymbol{v})) - \boldsymbol{s}(\boldsymbol{\theta}(0,\boldsymbol{w}))\|^2\}K(\boldsymbol{v})K(\boldsymbol{w})\, d\boldsymbol{w}d\boldsymbol{v}. \\ \le&\frac{1}{(4\pi)^{d/2}}\exp\{-\frac{1}{8}m_{D\boldsymbol{s}}^2(1 - M_{\boldsymbol{\beta}}c_5^{-1})^2\|\boldsymbol{t}\|^2\}, \end{aligned}$$

where the last inequality holds by noting that in R_2, $a_n\|\boldsymbol{s}(\boldsymbol{\theta}(\boldsymbol{t},\boldsymbol{v})) - \boldsymbol{s}(\boldsymbol{\theta}(0,\boldsymbol{w}))\| \ge m_{D\boldsymbol{s}}(1 - M_{\boldsymbol{\beta}}c_5^{-1})\|\boldsymbol{t}\|$. In R_2^c, since either $\|\boldsymbol{v}\|$ or $\|\boldsymbol{w}\|$ is larger than or equal to $c_5^{-1}\|\boldsymbol{t}\|/2$,

$$\begin{aligned} &\int_{(V_{M_\delta,n}\times V_{M_\delta,n})\cap R_2} N(a_n[\boldsymbol{s}(\boldsymbol{\theta}(\boldsymbol{t},\boldsymbol{v})) - \boldsymbol{s}(\boldsymbol{\theta}(0,\boldsymbol{w}))]; a_n\varepsilon_n(\boldsymbol{v}-\boldsymbol{w}), 2I_d)K(\boldsymbol{v})K(\boldsymbol{w})\, d\boldsymbol{w}d\boldsymbol{v} \\ \le&(4\pi)^{-d/2}\int_{R_2} K(\boldsymbol{v})K(\boldsymbol{w})\, d\boldsymbol{w}d\boldsymbol{v} \\ \le&2(4\pi)^{-d/2}\int_{\|\boldsymbol{v}\|\ge c_5^{-1}\|\boldsymbol{t}\|/2} K(\boldsymbol{v})\, d\boldsymbol{v}. \end{aligned}$$

Therefore the first multiplier in (12) satisfies

$$
\begin{aligned}
&a_n^{-d}\int_{\mathbb{R}^d}\widetilde{f}_{\varepsilon}^{*}(\boldsymbol{s}|\boldsymbol{\theta}_0+a_n^{-1}\boldsymbol{t}_2)\widetilde{f}_{\varepsilon}^{*}(\boldsymbol{s}|\boldsymbol{\theta}_0)\,d\boldsymbol{s}\\
\leq&\pi_{\sup/\inf}^2\left[\frac{1}{(4\pi)^{d/2}}\exp\{-\frac{1}{8}m_{D\boldsymbol{s}}^2(1-M_{\boldsymbol{\beta}}c_5^{-1})^2\|\boldsymbol{t}\|^2\}+2(4\pi)^{-d/2}\int_{\|\boldsymbol{v}\|\geq c_5^{-1}\|\boldsymbol{t}\|/2}K(\boldsymbol{v})\,d\boldsymbol{v}\right].
\end{aligned}
$$

Since it can be shown that $\int_{\|\boldsymbol{v}\|\geq x}K(\boldsymbol{v})\,d\boldsymbol{v}$ is $O(e^{-c_6x^{\alpha}})$ for some positive constants c_6 and α, $(IH2)$ holds when $c_{\varepsilon}<\infty$.

When $c_{\varepsilon}=\infty$, denote the constant upper bound of $\boldsymbol{\beta}_{\varepsilon}$ by $\boldsymbol{\beta}_{\sup}$. By Taylor expansion at $\boldsymbol{\theta}(\boldsymbol{t},\boldsymbol{v})-\boldsymbol{\theta}(0,\boldsymbol{w})=0$, the RHS of (13) is equal to

$$
\pi_{\sup/\inf}^2\int_{V_{M_{\delta},n}\times V_{M_{\delta},n}}\frac{1}{(4\pi)^{d/2}}\exp\{-\frac{1}{4}\|D(\boldsymbol{t},\boldsymbol{v}-\boldsymbol{w})\boldsymbol{t}-D^{\perp}(\boldsymbol{t},\boldsymbol{v}-\boldsymbol{w})a_n\varepsilon_n(\boldsymbol{v}-\boldsymbol{w})\|^2\}K(\boldsymbol{v})K(\boldsymbol{w})\,d\boldsymbol{w}d\boldsymbol{v},\tag{14}
$$

where $D(\boldsymbol{t},\boldsymbol{v}-\boldsymbol{w})=D\boldsymbol{s}(\boldsymbol{\theta}_0+\xi_1(\boldsymbol{t},\boldsymbol{v}-\boldsymbol{w}))$, $D^{\perp}(\boldsymbol{t},\boldsymbol{v}-\boldsymbol{w})=I_d-D(\boldsymbol{t},\boldsymbol{v}-\boldsymbol{w})\boldsymbol{\beta}_{\varepsilon}$ and $\|\xi_1(\boldsymbol{t},\boldsymbol{v}-\boldsymbol{w})\|\leq\|a_n^{-1}\boldsymbol{t}+\boldsymbol{\beta}_{\varepsilon}\varepsilon_n(\boldsymbol{v}-\boldsymbol{w})\|$. Let $P_D(\boldsymbol{t},\boldsymbol{v}-\boldsymbol{w})=D(\boldsymbol{t},\boldsymbol{v}-\boldsymbol{w})^TD(\boldsymbol{t},\boldsymbol{v}-\boldsymbol{w})$ and $P_{D^{\perp}}(\boldsymbol{t},\boldsymbol{v}-\boldsymbol{w})=D^{\perp}(\boldsymbol{t},\boldsymbol{v}-\boldsymbol{w})^TD^{\perp}(\boldsymbol{t},\boldsymbol{v}-\boldsymbol{w})$. For simplicity, we omit the arguments of D, $D^{\perp}$, P_D , $P_{D^{\perp}}$ and ξ_1. For small enough δ, note that they are all bounded.

To investigate the limit behaviour when $\|\boldsymbol{t}\|\to\infty$ required by $(IH2)$, consider two rates of $\|\boldsymbol{t}\|$. When $\|\boldsymbol{t}\|\to\infty$ with the rate faster than Qa_n, let $R_3=\{(\boldsymbol{v},\boldsymbol{w}):\|D^{\perp}a_n\varepsilon_n(\boldsymbol{v}-\boldsymbol{w})\|\leq\|D\boldsymbol{t}\|/2\}$. For $(\boldsymbol{v},\boldsymbol{w})\in R_3^c$, since D and $D^{\perp}$ are bounded, there exists $c_4>0$ such that

$$
\|\boldsymbol{v}-\boldsymbol{w}\|\geq\frac{c_4\|\boldsymbol{t}\|}{a_n\varepsilon_n}=\frac{c_4}{a_n^{3/5}\varepsilon_n}\left(\frac{\|\boldsymbol{t}\|}{a_n}\right)^{2/5}\|\boldsymbol{t}\|^{3/5}.
$$

Then for large enough n, $\|\boldsymbol{v}-\boldsymbol{w}\|\geq c_4Q^{2/5}\|\boldsymbol{t}\|^{3/5}$ which implies either $\boldsymbol{v}$ or $\boldsymbol{w}$ is larger than $c_4Q^{2/5}\|\boldsymbol{t}\|^{3/5}/2$. Then similar to the arguments when $c_{\varepsilon}<\infty$, we have

$$
\begin{aligned}
&a_n^{-d}\int_{\mathbb{R}^d}\widetilde{f}_{\varepsilon}^{*}(\boldsymbol{s}|\boldsymbol{\theta}_0+a_n^{-1}\boldsymbol{t}_2)\widetilde{f}_{\varepsilon}^{*}(\boldsymbol{s}|\boldsymbol{\theta}_0)\,d\boldsymbol{s}\\
\leq&\pi_{\frac{\sup}{\inf}}^2\left[\frac{1}{(4\pi)^{d/2}}\exp\{-\frac{1}{8}m_D^2\|\boldsymbol{t}\|^2\}+2(4\pi)^{-d/2}\int_{\|\boldsymbol{v}\|\geq c_4Q^{2/5}\|\boldsymbol{t}\|^{3/5}/2}K(\boldsymbol{v})\,d\boldsymbol{v}\right],
\end{aligned}\tag{15}
$$

where m_D^2 is the lower bound of DD^T.

Now consider $\|\boldsymbol{t}\|\to\infty$ with the rate slower than Qa_n. In (14), by matrix algebra we have

$$
\|D\boldsymbol{t}-D^{\perp}a_n\varepsilon_n(\boldsymbol{v}-\boldsymbol{w})\|^2\geq\|P_D^{-1/2}(P_D\boldsymbol{t}-D^TD^{\perp}a_n\varepsilon_n(\boldsymbol{v}-\boldsymbol{w}))\|^2.
$$

Let $R_4=\{(\boldsymbol{v},\boldsymbol{w}):\|D^TD^{\perp}a_n\varepsilon_n(\boldsymbol{v}-\boldsymbol{w})\|\leq\|P_D\boldsymbol{t}\|/2\}$ and denote the lower bound of P_D by m_{P_D}. For $(\boldsymbol{v},\boldsymbol{w})\in R_4^c$, to obtain an appropriate lower bound for $\|\boldsymbol{v}-\boldsymbol{w}\|$, first note that the leading term of Taylor expansion of $D^TD^{\perp}$ at $\xi_1=0$ is linear in ξ_1, since

$$
D^TD^{\perp}=D\boldsymbol{s}(\boldsymbol{\theta}_0+\xi_1)^T-D\boldsymbol{s}(\boldsymbol{\theta}_0+\xi_1)^TD\boldsymbol{s}(\boldsymbol{\theta}_0+\xi_1)\boldsymbol{\beta}_{\varepsilon}
$$

has 0 constant term, and the coefficient matrices are uniformly bounded in n. Denote the upper bound of the coefficient matrices by $M_{D,\xi}$. Then since $\|\xi_1\|\leq\|a_n^{-1}\boldsymbol{t}+\boldsymbol{\beta}_{\varepsilon}\varepsilon_n(\boldsymbol{v}-\boldsymbol{w})\|$, we have

$$
\begin{aligned}
\|D^TD^{\perp}a_n\varepsilon_n(\boldsymbol{v}-\boldsymbol{w})\|&\leq2M_{D,\xi}\|a_n^{-1}\boldsymbol{t}+\boldsymbol{\beta}_{\varepsilon}\varepsilon_n(\boldsymbol{v}-\boldsymbol{w})\|\|a_n\varepsilon_n(\boldsymbol{v}-\boldsymbol{w})\|\\
&\leq2M_{D,\xi}(\boldsymbol{\beta}_{\sup}a_n\varepsilon_n^2[\|\boldsymbol{v}-\boldsymbol{w}\|+\frac{1}{2\boldsymbol{\beta}_{\sup}}\frac{\|\boldsymbol{t}\|}{a_n\varepsilon_n}]^2-\frac{1}{4\boldsymbol{\beta}_{\sup}}\frac{\|\boldsymbol{t}\|^2}{a_n}).
\end{aligned}
$$

Together with the definition of R_4^c, we have

$$\begin{aligned}\|\boldsymbol{v}-\boldsymbol{w}\| &\geq \sqrt{\frac{1}{4\boldsymbol{\beta}_{\sup}}\frac{\|\boldsymbol{t}\|^2}{a_n^2\varepsilon_n^2}+\frac{1}{2\boldsymbol{\beta}_{\sup}M_{D,\xi}m_{P_D}}\frac{\|\boldsymbol{t}\|}{a_n\varepsilon_n^2}}-\frac{1}{2\boldsymbol{\beta}_{\sup}}\frac{\|\boldsymbol{t}\|}{a_n\varepsilon_n}.\\ &\geq \sqrt{\frac{\|\boldsymbol{t}\|}{2\boldsymbol{\beta}_{\sup}a_n\varepsilon_n^2}\left(\frac{1}{M_{D,\xi}}+\frac{\|\boldsymbol{t}\|}{2\boldsymbol{\beta}_{\sup}a_n}\right)}-\sqrt{\frac{\|\boldsymbol{t}\|}{2\boldsymbol{\beta}_{\sup}a_n\varepsilon_n^2}}\sqrt{\frac{Q}{2\boldsymbol{\beta}_{\sup}}}\\ &\geq c_5\frac{\|\boldsymbol{t}\|^{1/2}}{\sqrt{a_n}\varepsilon_n},\\ \text{where } c_5 &= \left(\frac{1}{\sqrt{M_{D,\xi}}}-\sqrt{\frac{Q}{2\boldsymbol{\beta}_{\sup}}}\right)\sqrt{\frac{1}{2\boldsymbol{\beta}_{\sup}}},\end{aligned}$$

and the second inequality holds by $\|\boldsymbol{t}\| \leq Qa_n$. Note that $\sqrt{a_n}\varepsilon_n \to 0$. Then with similar arguments, we have

$$\begin{aligned}&a_n^{-d}\int_{\mathbb{R}^d}\widetilde{f}_\varepsilon^*(\boldsymbol{s}|\boldsymbol{\theta}_0+a_n^{-1}\boldsymbol{t}_2)\widetilde{f}_\varepsilon^*(\boldsymbol{s}|\boldsymbol{\theta}_0)\,d\boldsymbol{s}\\ \leq&\pi^2_{\frac{\sup}{\inf}}\left[\frac{1}{(4\pi)^{d/2}}\exp\{-\frac{1}{8}m_{P_D}\|\boldsymbol{t}\|^2\}+2(4\pi)^{-d/2}\int_{\|\boldsymbol{v}\|\geq c_5\|\boldsymbol{t}\|^{1/2}/(\sqrt{a_n}\varepsilon_n)}K(\boldsymbol{v})\,d\boldsymbol{v}\right].\end{aligned}$$

Combining with (15), (IH2*) holds when $c_\varepsilon = \infty$. □