FIDUCIAL MATCHING

FOR

THE APPROXIMATE POSTERIOR: F-ABC

Yannis G. Yatracos

Yau Mathematical Sciences Center

Tsinghua University

and

Beijing Institute of Mathematical Sciences

and Applications

January 9, 2022

Some key words: Approximate Bayesian Computation; Bayesian consistency; Black-Box data; Empirical measure; Fiducial (F)-ABC; F-ABC *for all*; Matching Support Probability for the Tolerance; Samplers

Summary

When sample, $\mathbf{X} = \mathbf{x}$, is observed from intractable c.d.f. F_θ, or a Black-Box with input θ, an Approximate Bayesian Computation (ABC) method provides approximate posterior, π_ϵ. θ^* is included in the support of π_ϵ when the Matching distance $\rho(S(\mathbf{x}^*), S(\mathbf{x})) \leq \epsilon; \mathbf{x}^*$ is a sample drawn from F_{θ^*}, θ^* is obtained from prior π, S is a summary statistic, $\epsilon > 0$. ABC concerns include: the use of *only one sample,* $\mathbf{x}^*$, for each θ^*; the choices of S, ρ and ϵ; $\pi_\epsilon(\theta^*)$, which is determined by arbitrary kernel, $K(\mathbf{x}, \mathbf{x}^*; \epsilon)$, creating visual π_ϵ-artifacts. The concerns are accommodated with the introduced Fiducial(F)-ABC *for all* (θ^* drawn): M $\mathbf{x}^*$ are drawn from F_{θ^*}; a universal S is used, the empirical measure indexed by sets *which activate its sufficiency* for exchangeable observations-vectors and have been neglected; a strong, probability distance ρ is used, inherently connected with S and Matching; light is thrown to ϵ's nature and value; π_ϵ is obtained from the proportions of $\mathbf{x}^*$ matching $\mathbf{x}$. F-ABC *for all* posterior is closer to Bayesian philosophy, which does not use θ^*-exclusions. Under *few, mild assumptions,* π_ϵ converges to the posterior, $\pi(\theta|\mathbf{x})$, when $\epsilon \downarrow 0$, and rates of concentration of $T(F_{\theta^*})$ to $T(F_\theta)$ are obtained when $n \uparrow \infty; T$ is a functional. Satisfactory F-ABC *for all* θ^* drawn posteriors are depicted for parametric and data generating models, including Tukey's (a, b, g, h)-model, a 5-parameter normal mixture and a time series model. Various advantages of the F-ABC *for all* are presented over coarsened posteriors for observations in $R^d, d \geq 1$.

1 Introduction

In Bayesian inference, central theme is the posterior model, $\pi(\theta^*|\mathbf{x})$, of stochastic parameter Θ given the observed sample $\mathbf{X} = \mathbf{x}; \theta^* (\in \mathbf{\Theta})$ is observed from the Θ-prior, π. An Approximate Bayesian Computation (ABC) method provides an approximate posterior for Θ when the sample's likelihood (the model) is intractable. Rubin (1984) described the first ABC method for $\mathbf{x}$ with cumulative distribution function (*c.d.f.*) F_θ : one sample $\mathbf{x}^*$ is drawn for each of several θ^*-values and the θ^* for which $\mathbf{x}^*$ and $\mathbf{x}$ "match" within $\epsilon(> 0)$ constitute Θ's approximate posterior, with weights $\pi_\epsilon(\theta^*)$. Since then, tools from model-based approaches are most often used to find "nearly sufficient" statistics for *Matching* $\mathbf{x}^*$ with $\mathbf{x}$, surprisingly neglecting the empirical measure, $\mu_{\mathbf{X}}$, *indexed by Borel sets* which are the ammunition to activate the sufficiency of $\mu_{\mathbf{X}}$ for exchangeable observations in $R^d, d \geq 1$. The use of Borel sets will dictate the corresponding matching distance to be used, as explained in section 3.

The basic ABC-rejection algorithm (Tavaré *et al.* 1997, Pritchard *et al.*, 1999) includes θ^* in the support of the approximate posterior π_ϵ, when for tolerance level ϵ either

$$\rho(\mathbf{X}^*, \mathbf{x}) \leq \epsilon, \text{ or} \tag{1}$$

$$\rho(S(\mathbf{X}^*), S(\mathbf{x})) \leq \epsilon; \tag{2}$$

ρ is generic matching distance, S is a summary statistic.

ABC concerns are presented, which are accommodated with the Fiducial(F)-ABC for all θ^* drawn[1] introduced herein: $M(> 1)$ $\mathbf{x}^*$-samples are observed for each θ^* and their proportion matching $\mathbf{x}$ is used to obtain $\pi_\epsilon(\theta^*)$, making the approach more *fiducial* (trustworthy) than ABC, where $M = 1$. F-ABC is *algorithmic* and can be used also for data from a Black-Box with input θ, without resort to tools dictated by parametric models,

[1] In brief, F-ABC *for all.*

adhering to the philosophy in Breiman (2001, Abstract) "If our goal as a field is to use data to solve problems, then we need to move away from exclusive dependence on data models and adopt a more diverse set of tools."

ABC Concerns

Robert (2017) provided a survey on recent ABC results, identifying three approximations causing concerns:

i) ABC degrades the data precision down to ϵ, replacing the event $\mathbf{X} = \mathbf{x}$ with (1),

ii) ABC substitutes for the likelihood a non-parametric approximation,

iii) ABC summarizes $\mathbf{x}$ by *an almost always insufficient* $S(\mathbf{x})$.

There are additional concerns and open questions in ABC:

a) The dimension and form of S, when the statistical nature of θ is unknown.

b) The choice of ρ, that is inherently related with S and Matching.

c) The choice of ϵ-value, ϵ's missing sampling interpretation and components and its dependence on the sample size n and the distance of F_θ and F_{θ^*}.

d) The "hard" inclusion-exclusion of θ^* in the support of π_ϵ using one sample $\mathbf{x}^*$ from F_{θ^*}.

e) The θ^*-weight $K(\frac{\mathbf{x}-\mathbf{x}^*}{\epsilon})$ used in π_ϵ, which often creates a K-dependent visual artifact[2].

f) The numerous, not easily verifiable, strong assumptions used in asymptotics.

Potential logical inconsistencies in ABC are not clarified:

g) Is non-selected θ^* included in the support of π_ϵ when it belongs in the convex hull of the selected?

h) For discrete Θ and with θ^* drawn *e.g.* twice, is θ^* included in the support of π_ϵ if only one of the simulated $\mathbf{x}_1^*, \mathbf{x}_2^*$ matches $\mathbf{x}$?

Affirmative answers to *g), h)* contradict the ABC-Algorithm in (1) and (2).

Concern *iii)* (Robert, 2017) confirms what was *naturally* expected: a plateau is *finally*

[2]Examples of smooth histograms' artifacts appear in Figure 1.

reached with the insistence on tools from the model-based approach, namely that a sufficient statistic is a set of estimates, S, providing information about θ, even when θ's statistical nature is unknown, F_θ is intractable and Neyman's Factorisation Criterion (*NFC*) cannot be used. Identifying S without *NFC* is like looking for a needle in a haystack. The implications of *iii)*, since Rubin's ABC outset in 1984, confirm Breiman (2001): "This commitment (to data models) has led to irrelevant theory, questionable conclusions, and has kept statisticians from working on a large range of interesting current problems."

Coarsened Approximate Posteriors and Additional Concerns

Concerns *iii)* and *a)* motivated recently new research directions in ABC, extending its scope but remaining model-centered: *I)* it is assumed the underlying data model depends *in reality* on parameter $\eta \in \mathbf{H}$ and belongs to $\mathcal{F}_{\mathbf{H}}$, a larger class of models than $\mathcal{F}_{\mathbf{\Theta}} = \{F_{\theta^*}, \theta^* \in \mathbf{\Theta}\}$, and *II)* the search for sufficient summary is *bypassed* in favor of the empirical *distribution*,

$$\hat{\mu}_n = n^{-1} \sum_{i=1}^{n} \delta_{X_i}; \tag{3}$$

δ_{x^*} is Dirac distribution with mass on $x^*(\in R^d), \mathbf{X} = \{X_1, \ldots, X_n\}$ (Miller and Dunson, 2019, Bernton *et al.*, 2019). Note that, if δ_{x^*} is Dirac *distribution* in the sense of Schwartz (1951), it is not an ordinary function since it is defined either as limit of functions or by its integral, and, *e.g.*, δ_x^2 is not defined (Schwartz, 1954). If δ_{X_i} were a Dirac *measure*, according to Dudley (1984, 10.3.1, Theorem), $\hat{\mu}_n$ should be evaluated at the Borel sets, $\mathcal{B}_d$, in R^d, to activate its sufficiency and have a statistical interpretation, but it is not. Thus, *I)* and *II)* led to a robust, *coarsened* (c) posterior (Miller and Dunson, 2019, Bernton *et al.*, 2019). However, robust approximate posteriors for $\mathcal{F}_{\mathbf{H}}$ may be sub-optimal for $\mathcal{F}_{\mathbf{\Theta}}$, and the role of $\pi(\theta)$ in the H-posterior is not clear, since it is not necessary that $\mathbf{\Theta} \subset \mathbf{H}$. Also, $\mathcal{F}_{\mathbf{H}}$-robust posteriors are not comparable with ABC posteriors for $\mathcal{F}_{\mathbf{\Theta}}$, as happens with the mean and the median of observations. The latter is confirmed indirectly by the authors, with adjective “coarsened” preceding “posterior”. Statements by these same authors follow, creating additional concerns.

Miller and Dunson (2019) write: “The main disadvantage of c-posteriors is that sometimes are less concentrated than one would like ...”[3] (in section 1), adding also that ρ-distances on densities, as relative entropy, Hellinger distance and various divergences “*may be undefined for empirical distributions*” (in section 3). Both statements and the definition of $\hat{\mu}_n$ in (3), reinforce raising the question: what information $\hat{\mu}_n$ carries for θ, F_θ and the underlying probability, P_θ, in $R^d, d > 1$?

A partial, indirect answer appeared in Bernton *et al.* (2019, Introduction, 2nd paragraph), “*We propose here to instead view data sets as empirical distributions* [4] and to rely on the Wasserstein (W_p) distance *between synthetic and observed data sets.*”, obtaining the WABC c-posterior and “*hoping*[5] *to avoid* the loss of information incurred by the use of summary statistics” (section 1.3, first paragraph); $p > 0$. In section 3.2 the authors write: “ ... the WABC distribution with a fixed ϵ does not converge to a Dirac mass, contrarily to the standard posterior. As argued in Miller and Dunson (2018), this can have some benefit in case of model misspecification: *the WABC posterior is less sensitive to perturbations of the data-generating process than the standard posterior.*” This statement reconfirms the coarsening of the WABC posterior and its difference from the posterior. In the last paragraph of section 3.3, it is added: “In high dimensions, the rate of convergence of the Wasserstein distance between *empirical measures*[6] is known to be slow (Talagrand, 1994).” and “Detailed analysis of WABC’s dependence on dimension is an interesting avenue of future research.”[7] In section 3, 2nd paragraph, it is written “We remark that the

[3]Confirms sub-optimality of c-posteriors.

[4]Thus, $\hat{\mu}_n$ is the data, $\mathbf{X}$, which is sufficient only in R.

[5]Confirms potential loss of information with WABC.

[6]The relation between empirical distribution and empirical measure was not provided.

[7]Confirms lack of large sample optimality results for WABC in high dimension.

assumptions underlying our results are typically hard to check in practice, ...". According to the authors, WABC is a c-posterior, thus we conclude, it shares its disadvantages.

Bernton *et al.* (2019) refer also to Fournier and Guillin (2015) and Weed and Bach (2017), for the upper bounds on the risk, $EW_p(\hat{\mu}_n, P_\theta)$, and for concentration inequalities when P_θ is defined either in R^d or on a compact metric space; $\hat{\mu}_n$ denotes in these papers the empirical measure indexed by sets. However, these elegant and deep mathematical results are not favorable to the use of $(\hat{\mu}_n, W_p)$ in ABC. The obtained bounds depend on p, d for the concentration inequalities and, in addition, to coefficient(s) from moment conditions for the risk bounds, but also on their relative orderings. In Weed and Bach (2017, Proposition 20), the Dvoretzky-Kiefer-Wolfowitz-Massart (D-K-W-M) rate, $e^{-2n\epsilon^2}$, remains valid for the probability that $W_p^p(\hat{\mu}_n, P_\theta)$ is larger than ϵ *augmented by its expectation;* $\epsilon > 0$. Compactness of P_θ's support and the moment conditions needed do not appear in Miller and Dunson (2019, assumptions in Theorem 5.3 and Corollaries 5.4 and 5.5) and in Bernton *et al.* (2019, Assumptions 1 and 2). However, similar or stronger results already hold with weaker assumptions for the empirical *c.d.f.*, $\hat{F}_{\mathbf{X}}$, and the empirical measure, $\mu_{\mathbf{X}}$, due to Glivenko-Cantelli Theorem and Large Deviations' inequalities.

Fiducial-ABC and Results

It is crystal clear that when θ's statistical nature is unknown, information about θ is obtained only from F_θ and P_θ, which become the parameters of interest. Main drawback of $\hat{\mu}_n$ in (3) is the inadequate information it provides for F_θ and P_θ unlike $\hat{F}_{\mathbf{X}}$ and $\mu_{\mathbf{X}}$ which are both evaluated on Borel sets and have statistical interpretations. *This information is valuable when matching* $\mathbf{x}$ *and* $\mathbf{x}^*$. Consequently, π_ϵ's coarsening near θ is also due to the reduced discriminating information, combined with the use of weak W-distance in (1) and (2).

The concerns led us to search for an alternative approach to ABC. *i)* and *ii)* seem unavoidable with intractable or unavailable continuous models. The previous paragraphs motivated the use of $\{\hat{F}_{\mathbf{X}}(x), x \in R\}$ and $\{\mu_{\mathbf{X}}(A), A \in \mathcal{B}_d\}$, since the latter is sufficient for *i.i.d.* and exchangeable data in $R^d, d \geq 1$; see, *e.g.,* Dudley (1984) and Lauritzen (2007). To match $\mathbf{x}$ with $\mathbf{x}^*$, the Kolmogorov distance $d_K(\hat{F}_{\mathbf{x}}, \hat{F}_{\mathbf{x}^*})$ is used when $d = 1$, and the Total Variation distance, TV, can be used when $d > 1$. Thus, $(\hat{F}_{\mathbf{X}}, d_K)$ and $(\mu_{\mathbf{X}}, TV)$ are natural candidates for (S, ρ), relaxing *iii), a), b)*. However, for $d > 1$, as explained in section 3, Wolfowitz's half-spaces, $\mathcal{V}$, which separate probabilities and are invariant under affine transformations and Vapnik-Cervonenkis subclass of Borel sets, $\mathcal{B}_d$, provide strong distance, $\tilde{\rho}$, used in applications. $\mathcal{V}$ separates probabilities since probability measures in $(R^d, \mathcal{B}_d)$ are equal ("match") *if and only if* they coincide on $\mathcal{V}$. $\tilde{\rho}(\mu_{\mathbf{X}}, \mu_{\mathbf{X}^*})$ measures the maximum "information loss" on the separating sets, $\mathcal{V}$. It is also seen in section 5.1, that for a Dirichlet prior, $DP(\alpha, G)$, on the models F_θ and $P_\theta, \theta \in \Theta$, $\hat{F}_{\mathbf{X}}$ and $\mu_{\mathbf{X}}$ are, respectively, approximate posterior means when $\alpha \neq 0$, and posterior means when $\alpha = 0$, one of the desired properties for Summary statistics in ABC (Fernhead and Prangle, 2012).

The Conditional Calibration framework (Rubin, 2019) and an observation in several models lead to Fiducial (F)-ABC matching with M $\mathbf{x}^*$ drawn from F_{θ^*}, making the ABC approach more trustworthy; usually, $50 \leq M \leq 200$. The matching support proportions of $\mathbf{x}^*$'s within the ϵ-tolerance, $p_{match}(\theta^*)$, provide $\pi_\epsilon(\theta^*), \theta^* \in \mathbf{\Theta}$. $p_{match}(\theta^*)$ estimates the $\mathbf{x}^*$-*matching support probability*, α, of event (2) that provides ϵ's sampling interpretation and value; $0 \leq \alpha \leq 1$. The motivating observation was that for several F_{θ^*}-models, $p_{match}(\theta^*)$ converges to 1 as θ^* converges to θ. The use of $p_{match}(\theta^*)$, reduces ϵ's "0-1" influence in the θ^*-selection, and allows to avoid the use of a kernel, thus providing a remedy for *c)-e)* and *g), h)*. When $M = 1$, F-ABC is nonparametric ABC. The $\hat{\theta}_{Match}$ maximizing $p_{match}(s), s \in \Theta$, is the Maximum Matching Support Probability Estimate (MMSPE, Yatracos, 2020). Its uniform rate of convergence in probability to θ for observations in $R^d, d \geq 1$, confirms the high concentration of the F-ABC *for all* approximate posterior around θ, as observed in

Examples.

In simulations from a normal model, nonparametric F-ABC used only *for the selected* θ^* in ABC with d_K, competes well against parametric ABC with a kernel, and improves most frequently the concentration of the ABC posterior. F-ABC *for all* posteriors are then depicted for the means of a bivariate normal with dependent components and for each parameter in Tukey's (a, b, g, h)-model, a 5-parameters normal mixture, a time series model and a quantile model. The "F-ABC *for all*" frequency histograms of posteriors are obtained using $p_{match}(\theta^*)$ for all θ^* drawn, and θ is most often in the modal neighborhood of $\hat{\theta}_{Match}$.

For the $\mathbf{X}^*$-matching support probability, α, with $\rho = d_K$ and real observations, an upper bound $\epsilon_{n,B}$ on $\epsilon = \epsilon_n$ is determined; $0 < \alpha < 1$. $\epsilon_{n,B}$ has two additive components: *A)* the observed or acceptable discrepancy between F_θ and the F_{θ^*}-models, and *B)* a component determined by a confidence related to α. From section 5.2 and for observations in R, the ϵ-value used in the Examples is in the interval $[n^{-.5}, 3n^{-.5}]$ with the coefficient in the upper bound, "3", possibly increased when $\theta \in R^k, k \geq 6$; ϵ can be also determined via α and the $\mathbf{x}^*$-Sampler used (section 4.1). An interval for ϵ can also be obtained for observations in $R^d, d > 1$, with an extension of Proposition 5.1 using either $\tilde{\rho}(\mu_{\mathbf{X}}, \mu_{\mathbf{X}^*})$ or its approximation and, respectively, Vapnik-Cervonenkis inequality and concentration inequalities with d_K. Under exchangeability on $F_\theta(\mathbf{y})$, the ABC and F-ABC posteriors with d_K-matching converge to $\pi(\theta|\mathbf{x})$ when ϵ converges to zero; n is fixed. For a continuous linear functional T on the space of *c.d.fs*, Bayesian consistency is established and the rate of concentration of $T(F_{\theta^*})$ around $T(F_\theta)$ depends on ϵ_n, the rate of concentration in probability of $\hat{F}_{\mathbf{X}}$ around F_θ, and T's modulus of continuity (section 6).

$\epsilon_{n,B}$'s components and the motivation for $p_{match}(\theta^*)$ in Propositions 5.1, 5.2 are presented for i.i.d. samples but hold also under weak-dependence, as well as for exchangeable samples, when a *D-K-W-M* type upper bound or a Large Deviations bound for empirical processes hold and is not necessarily exponential, *e.g.* the bound in linear time series by Chen and Wu (2018). The assumptions used for consistency and the concentration in Propositions 6.1 and 6.2 are *mild and fewer* than the assumptions in the ABC and c-ABC literature, relaxing *f).*

Related ABC Work

Lintusaari *et al* (2017) and Fearnhead (2018) provide accessible introductions to ABC presenting, respectively, recent developments and results on asymptotics. Tanaka *et al.* (2006, p. 1517 and Figure 4) indicate ϵ's choice is crucial for the sampler acceptance rates and the posterior densities. Fearnhead and Prangle (2012) show how to construct summary S to be used in (2), "which will enable inference about certain parameters of interest to be as accurate as possible" (in Summary). Biau *et al.* (2015), analyze ABC as a k-nearest neighbor method. Frazier *et al.* (2018) provide asymptotic theory for a posterior. Vihola and Franks (2020) suggest a balanced ϵ from a *range* of tolerances via Bayesian MCMC. Chaudhuri *et al.* (2020) propose a fast and easy-to-use ABC method based on empirical likelihood, a *natural* summary statistics. These authors use an algorithmic approach based on an information projection argument, refreshingly without kernel approximation of the summary statistic likelihood.

2 Nonparametric Fiducial ABC for all θ^*

Let $(\mathcal{Y}, \mathcal{C}_{\mathcal{Y}})$ denote space $\mathcal{Y}$ with σ-field $\mathcal{C}_{\mathcal{Y}}$. $\mathbf{X}$ is a sample of size n, obtained from the unknown θ-model with *c.d.f.* F_θ and density f_θ (or $f(\cdot|\theta)$) with respect to measure μ on $(\mathcal{X}, \mathcal{C}_{\mathcal{X}}); \theta \in \mathbf{\Theta}$. $\mathcal{X}$ is usually subset of R^d with the Borel σ-field, $\mathcal{B}_d, d \geq 1$. Let $\pi(\theta)$ be the assumed prior of Θ with respect to measure ν on $(\mathbf{\Theta}, \mathcal{C}_{\mathbf{\Theta}})$, with unknown posterior

$\pi(\theta|\mathbf{X}=\mathbf{x}); \theta \in \mathbf{\Theta}$. $\mathbf{X}^*$ is a sample of size n obtained from the sampler with model F_{θ^*}. $S(\mathbf{X})$ is a summary for $\mathbf{X}$, ρ measures the distance between $S(\mathbf{X})$ and $S(\mathbf{X}^*)$. $S(\mathbf{X})$ can be thought of as estimate of $T(F_\theta)$; T is generic functional of F_θ. $\mathbf{\Theta}$ is metrized with $d_{\mathbf{\Theta}}$ and generic $\tilde{d}$ and d_K are distances for *c.d.fs.* θ-identifiability is assumed, *i.e.*, $F_{\theta_1} = F_{\theta_2}$ implies $\theta_1 = \theta_2$. For $A \in \mathcal{C}_{\mathcal{Y}}, I_A(\mathbf{u}) = 1$ if $\mathbf{u} \in A$ and zero otherwise.

Definition 2.1 *For tolerance $\epsilon, \mathbf{X}$ and S, the $\mathbf{X}^*$-matching support probability α for θ^* is*

$$P[\rho(S(\mathbf{X}^*), S(\mathbf{X})) \leq \epsilon] = \alpha, \ 0 \leq \alpha \leq 1, \epsilon > 0. \tag{4}$$

For $\mathbf{\Theta}^* = \{\theta_1^*, \ldots, \theta_N^*\}$, the matching support probability is

$$\inf\{\alpha_i; i = 1, \ldots, N\}; \tag{5}$$

α_i is obtained from (4) for $\theta^* = \theta_i^*, i = 1, \ldots, N$. The observed $\mathbf{X} = \mathbf{x}$ can be used in (4).

The probability in (4) is not under one probability model as in confidence band calculations since $\mathbf{X}$ and $\mathbf{X}^*$ follow F_θ and F_{θ^*}, respectively. When $\mathbf{X} = \mathbf{x}$, ϵ is the α-quantile of $\rho(S(\mathbf{X}^*), S(\mathbf{x}))$ under F_{θ^*} and seeing density $f(\mathbf{x}|\theta^*)$ as "small probability" for small ϵ,

$$\pi(\theta^*|\mathbf{x}) \propto \pi(\theta^*) f(\mathbf{x}|\theta^*) \propto \pi(\theta^*) P_{\theta^*}[\rho(\mathbf{X}^*, \mathbf{x}) \leq \epsilon]. \tag{6}$$

A nonparametric estimate of this "small probability" is introduced in (7) with $S(\mathbf{x}), S(\mathbf{x}^*)$ instead of $\mathbf{x}, \mathbf{x}^*$. The α-value is omitted from the F-ABC notation, since it is determined along with ϵ in (4) .

F-ABC *Algorithm*

Obtain sample $\mathbf{X} = \mathbf{x}$ of size n from F_θ, select $\epsilon = \epsilon_n > 0$ and $\alpha = \alpha_n$ from $[0, 1]$.

1) Sample *i.i.d.* $\theta_1^*, \ldots, \theta_{N^*}^*$ from $\mathbf{\Theta}$ according to $\pi(\theta)$.

2) Repeat for each $\theta_i^*, \ i = 1, \ldots, N^*$:

a) Sample $\mathbf{X}_j^*$ with size n from $F_{\theta_i^*}, j = 1, \ldots, M$.

b) Compute the *matching support proportion*, $p_{match}(\theta_i^*)$, for the observed $\mathbf{x}_1^*, \ldots, \mathbf{x}_M^*$:

$$p_{match}(\theta_i^*) := p_{match}(\theta_i^*, \mathbf{x}) = \frac{Card(\{\mathbf{x}_i^* : \rho(S(\mathbf{x}_i^*), S(\mathbf{x})) \leq \epsilon_n,\ i = 1, \ldots, M\})}{M}. \tag{7}$$

c) θ^**-selection criterion: the* F-ABC *filter.* [8] Include θ_i^* in the domain of $\pi(\theta|\mathbf{x})$ when

$$p_{match}(\theta_i^*) \geq \alpha_n. \tag{8}$$

3) The selected θ^* in **2)** *c)* are

$$\mathbf{\Theta}_n^* = \{\theta_{sel,i}^*; i = 1, \ldots, N\},\ N \leq N^*. \tag{9}$$

Use $\{(\theta_{sel,i}^*, p_{match}(\theta_{sel,i}^*)); i = 1, \ldots, N\}$ to construct the F-ABC posterior for the selected.

If **2** *c)* is not used, $\mathbf{\Theta}_n^* = \{\theta_1^*, \ldots, \theta_{N^*}\}$, and $p_{match}(\theta_i^*), i = 1, \ldots, N^*$, are the weights in the F-ABC posterior for all θ^*, with

$$\pi_\epsilon(\theta_i^*) = \frac{p_{match}(\theta_i^*)}{\sum_{j=1}^{N^*} p_{match}(\theta_j^*)},\ i = 1, \ldots, N^*. \tag{10}$$

In simulations, *frequency histograms* are presented for π_ϵ.

Definition 2.2 *The matching support proportion for* $\mathbf{\Theta}_n^*$ *in (9) is* $\min\{p_{match}(\theta_{sel,i}^*); i = 1, \ldots, N\}$.

Remark 2.1 *An approach to compare parametric ABC with F-ABC: Observe that when* $M = 1$ *in* **2)***a) and* $\alpha_n = 1$ *in (8),* ρ_2-F-ABC *is* ρ_2-ABC. *To compare parametric* ρ_1-ABC *with* ρ_2-F-ABC, *start with* ρ_2-ABC, *use* M *additional* $\mathbf{x}^*$*-samples for the selected* θ^* *to obtain* $p_{match}(\theta^*)$ *for all* $(M+1)$ $\mathbf{x}^*$*-drawn, and proceed with* **3)** *to construct the* ρ_2-F-ABC *posterior for the selected* θ^**. When either* $\alpha_n = 0$ *in (8), or* **2)***c) is not used, all* θ^* *are selected for the posterior with their corresponding weights,* $p_{match}(\theta^*)$*, used to obtain* F-ABC *for all.*

[8] Optional. Not used in F-ABC for all θ^*. It is intended for users desiring to restrict more than ABC the approximate posterior.

Let

$$B_{\epsilon_n} = \{\mathbf{x}^* : \rho(S(\mathbf{x}^*), S(\mathbf{x})) \leq \epsilon_n\}. \tag{11}$$

Then, the F-ABC posterior of θ is

$$\pi_{f\text{-}abc}(\theta|B_{\epsilon_n}) = \frac{\pi(\theta) \cdot \int_{\mathcal{Y}} I_{B_{\epsilon_n}}(\mathbf{y}) f(\mathbf{y}|\theta)\mu(d\mathbf{y})}{\int_{\mathbf{\Theta}} \pi(s) \int_{\mathcal{Y}} I_{B_{\epsilon_n}}(\mathbf{y}) f(\mathbf{y}|s)\mu(d\mathbf{y})\ \nu(ds)}, = \frac{\pi(\theta) \cdot P_\theta^{(n)}(B_{\epsilon_n})}{\int_{\mathbf{\Theta}} \pi(s) \cdot P_s^{(n)}(B_{\epsilon_n})\nu(ds)}. \tag{12}$$

and for $H \in \mathcal{C}_{\mathbf{\Theta}}$, its F-ABC probability is

$$\Pi_{f\text{-}abc}(H|B_{\epsilon_n}) = \int_H \pi_{f\text{-}abc}(\theta|B_{\epsilon_n})\nu(d\theta) = \frac{\int_{\mathbf{\Theta}} \pi(\theta) \cdot P_\theta^{(n)}(H \cap B_{\epsilon_n})\nu(d\theta)}{\int_{\mathbf{\Theta}} \pi(s) \cdot P_s^{(n)}(B_{\epsilon_n})\nu(ds)}. \tag{13}$$

For ABC, π_{abc} and Π_{abc} are used instead.

4) *Determination of* ϵ_n, α_n : Sample several θ^*-values either from $\pi(\theta)$ or from a discretization of $\mathbf{\Theta}$ if it is known. Use one of them as base-value, θ_b^*, and obtain $\mathbf{x}$ generated by θ_b^*. *Select*, e.g., m θ^* at increasing *standardized* distance from θ_b^* taking into consideration its nature (if known) and obtain M $\mathbf{X}^*$-samples from each one of them and $\theta_b^*; 5 \leq m \leq 10$. Calculate $\rho(S(\mathbf{X}_i^*), S(\mathbf{x})), i = 1, \ldots, M$, and their empirical quantiles for each one of the selected θ^* and θ_b^*. For example, if θ^* is location parameter use $\theta_i^* = \theta_b^* \pm \sigma_i$; if θ^* is scale parameter, $\theta_i^* = c_i\theta_b^*, c_i \in [1 - \delta_1, 1 + \delta_2]; \sigma_i > 0, 0 < \delta_1 < 1, 0 < \delta_2 < 2, i = 1, \ldots, m$. Create a table similar to Table 1 in section 4.1. After examination of the empirical quantiles, decide on the ϵ_n to be used. Alternatively, using the results in section 5.2 for the proposed F-ABC and real observations, ϵ_n is used from $[n^{-.5}, 3n^{-.5}]$ for $\theta \in R^k, k \leq 5$.

3 Sufficient Summary and Matching Distance

Matching with sufficient summary, S, is preferred since $\pi(\theta|\mathbf{x}) = \pi(\theta|S(\mathbf{x}))$. Information for θ could be obtained via F_θ and P_θ which are unavailable. Thus, their sample counterparts, *i.e.* the empirical *c.d.f.*, $\hat{F}_{\mathbf{X}}$, and the empirical measure, $\mu_{\mathbf{X}}$, indexed by sets, are the tools to be used as summaries.

Definition 3.1 *For any n-size sample,* $\mathbf{Y} = \{Y_1, \ldots, Y_n\} = {}^9(Y_1, \ldots, Y_n)$, *of random vectors in* $R^d, n\hat{F}_{\mathbf{Y}}(y)$ *denotes the number of* Y_i*'s with all their components smaller or equal to the corresponding components of* y. $\hat{F}_{\mathbf{Y}}$ *is the empirical c.d.f. of* $\mathbf{Y}$.

The empirical measure, $\mu_{\mathbf{Y}}$, *of* $\mathbf{Y}$ *is*

$$\mu_{\mathbf{Y}}(A) = \frac{1}{n} \sum_{i=1}^{n} I_A(Y_i), \ A \in \mathcal{B}_d; \tag{14}$$

$I_A(y) = 1$ *if* $y \in A$ *and* 0 *otherwise,* $\mathcal{B}_d$ *are the Borel sets in* R^d.

When $\mathbf{X} = (X_1, \ldots, X_n) \in R^n, \hat{F}_{\mathbf{X}}$ is sufficient being equivalent to the order statistic. When $\mathbf{X} \in R^{nxd}, d > 1$, $\mu_{\mathbf{X}}$ in (14) evaluated on sets in $\mathcal{B}_b$ is sufficient when $X_1, \ldots, X_n$ are either *i.i.d* (Dudley, 1984, Theorem 10.1.3, p. 95) or *exchangeable* (de Finetti, 1931, and Hewitt and Savage, 1955, at least for compact sets in R^d). The results for exchangeable data appear in an accessible manner in Lauritzen (2007). Choices of distances for matching $\hat{F}_{\mathbf{X}}$ with $\hat{F}_{\mathbf{X}^*}$ and $\mu_{\mathbf{X}}$ with $\mu_{\mathbf{X}^*}$, are *naturally* the Kolmogorov distance, d_K, and the Total Variation distance, TV, respectively.

Definition 3.2 *For distribution functions* F, G *in* R^d, *with induced probabilities* P_F *and* P_G *in* $(R^d, \mathcal{B}_b)$, *the Kolmogorov and Total Variation distances are, respectively,*

$$d_K(F, G) = \sup\{|F(y) - G(y)|; y \in R^d\}, \tag{15}$$

$$TV(P_F, P_G) = \sup\{|P_F(A) - P_G(A)|; A \in \mathcal{B}_d\}, \tag{16}$$

$\mathcal{B}_d$ *are the Borel sets in* $R^d, d \geq 1$.

For good matching of $\mathbf{X}$ and $\mathbf{X}^*$ using $\hat{F}_{\mathbf{X}}, \hat{F}_{\mathbf{X}^*}$ and d_K, it is also required that $d_K(\hat{F}_{\mathbf{X}}, \hat{F}_{\mathbf{X}^*})$ approximates well $d_K(F_\theta, F_{\theta^*})$ with high probability. This holds for d_K since

$$|d_K(F_\theta, F_{\theta^*}) - d_K(\hat{F}_{\mathbf{X}}, \hat{F}_{\mathbf{X}^*})| \leq d_K(\hat{F}_{\mathbf{X}}, F_\theta) + d_K(\hat{F}_{\mathbf{X}^*}, F_{\theta^*}), \tag{17}$$

[9]Abuse of notation: the order in $\mathbf{Y}$ does not matter.

due to Glivenko-Cantelli Theorem and Concentration Inequalities, like *D-K-W-M*, which make the upper bound in (17) converge to 0 in probability.

Inequality (17) holds also using instead $\mu_{\mathbf{X}}, \mu_{\mathbf{X}^*}$ and TV, but the upper bound does not always converge to 0 in probability since $\mathcal{B}_d$ is not necessarily a *V-C* (Vapnik and Cervonenkis, 1971) class of sets. However, often, $TV(P_\theta, P_{\theta^*})$ is approximated as close as is wished for any θ, θ^*, by $\rho_{VC}(P_\theta, P_{\theta^*})$, with the supremum in (16) taken over a *V-C*-subclass of $\mathcal{B}_d$. Examples of such families of models can be found in Yatracos (1988), and include in particular those satisfying the Hoeffding-Wolfowitz condition on the sign changes for the densities' differences. $\rho_{VC}(\mu_{\mathbf{X}}, \mu_{\mathbf{X}^*})$ is used for matching and the size of its difference from $TV(P_\theta, P_\theta)$ converges to zero in Probability.

In the applications herein we use a *V-C* class for a supremum-type distance $\tilde{\rho}$ as in (16), with the Matching property that $\tilde{\rho}(P, Q) = 0$ implies $TV(P, Q) = 0$. Such class, $\mathcal{V}$, exists in R^d, consists of the *half-spaces* and the distance, $\tilde{\rho}$, was introduced by Wolfowitz (1954) and was used also by Beran and Millar (1986, p. 431) who present its properties: *a)* if $P(A) = Q(A)$ for each $A \in \mathcal{V}$, then P, Q agree also on $\mathcal{B}_d$ (Cramer and Wold, 1936), and *b)* $\mathcal{V}$ is a *V-C* class of index $(d + 1)$ (*e.g.*, Dudley, 1978). From *a)*, $\mathcal{V}$ is a class separating probabilities. The advantage with $\tilde{\rho}$ is that it can be approximated by d_K, as seen in (24).

Let $< \cdot, \cdot >$ and $|| \cdot ||$ be, respectively, the inner-product and the Euclidean norm in R^d, U_d is the unit sphere in R^d,

$$U_d = \{u = (u_1, \ldots, u_d) \in R^d : ||u|| = 1\}. \tag{18}$$

Definition 3.3 *In* $(R^d, \mathcal{B}_d)$, *the half-space,* $A(a, t)$, *is*

$$A(a, t) = \{y \in R^d :< a, y >\leq t\}, \ t \in R, a \in U_d. \tag{19}$$

The class of half-spaces, $\mathcal{V}$, *is*

$$\mathcal{V} = \{A(a, t) : a \in U_d, t \in R\}. \tag{20}$$

Definition 3.4 *For probability measures* P, Q *in* $(R^d, \mathcal{B}_d)$ *and half-spaces* $\mathcal{V}$, *Wolfowitz's half-spaces distance is,*

$$\tilde{\rho}(P,Q) = \sup\{|P(A) - Q(A)|; A \in \mathcal{V}\} = \sup_{a \in U_d} \sup_{t \in R} |P(A(a,t)) - Q(A(a,t))|. \tag{21}$$

Observe that for $A = A(a,t)$ in (14), from (19)

$$I_{A(a,t)}(X_i) = 1 \iff < a, X_i > \leq t,$$

and using the notation

$$a \cdot \mathbf{X} = (< a, X_1 >, \ldots, < a, X_n >) \in R^n, \tag{22}$$

it follows that

$$\mu_{\mathbf{X}}(A(a,t)) = \frac{Card(< a, X_i > \leq t, \ i = 1, \ldots, n)}{n} = \hat{F}_{a \cdot \mathbf{X}}(t), \tag{23}$$

Since for $\mathbf{X} \in R^d, d > 1$, $\mu_{\mathbf{X}}$ is used for ϵ-matching $\mathbf{X}$ with $\mathbf{X}^*$,

$$\tilde{\rho}(\mu_{\mathbf{X}}, \mu_{\mathbf{X}^*}) = \sup_{a \in U_d} \sup_{t \in R} |\mu_{\mathbf{X}}(A(a,t)) - \mu_{\mathbf{X}^*}(A(a,t))| = \sup_{a \in U_d} d_K(\hat{F}_{a \cdot \mathbf{X}}, \hat{F}_{a \cdot \mathbf{X}^*}). \tag{24}$$

In practice, $\tilde{\rho}(\mu_{\mathbf{X}}, \mu_{\mathbf{X}^*})$ is approximated by

$$\tilde{\rho}_n(\mu_{\mathbf{X}}, \mu_{\mathbf{X}^*}) = \max_{a \in \{a_1, \ldots, a_{k_n}\} \subset U_d} \sup_{t \in R} |\mu_{\mathbf{X}}(A(a,t)) - \mu_{\mathbf{X}^*}(A(a,t))| = \max_{a \in \{a_1, \ldots, a_{k_n}\} \subset U_d} d_K(\hat{F}_{a \cdot \mathbf{X}}, \hat{F}_{a \cdot \mathbf{X}^*}). \tag{25}$$

where $a_1, \ldots, a_{k_n}$ are either a discretization of U_d or *i.i.d.* uniform in U_d, independent of $\mathbf{X}$ and $\mathbf{X}^*$. Beran and Millar (1986) showed already that if $a_1, \ldots, a_{k_n}$ are *i.i.d.* uniform on U_d and $k_n \uparrow \infty$ as $n \uparrow \infty$, then $\lim_{n \to \infty} \tilde{\rho}_n(P,Q) = \tilde{\rho}(P,Q)$ with probability 1. Note that the last equalities in (24) and (25) relate $\tilde{\rho}$ over all half-spaces with d_K-distance over all 1-dimensional projections of $\mathbf{X}, \mathbf{X}^*$. In the F-ABC Algorithm, with $d > 1$, $\mathbf{X}$ will match $\mathbf{X}^*$ when the last term in (25) is less than or equal to ϵ_n.

4 Implementation

In simulations, *w.l.o.g.* uniform prior, $\pi(\theta)$, is used over $\mathbf{\Theta}$, or over its discretization $\mathbf{\Theta}^*$. Frequency histograms for the F-ABC *for all* posteriors are presented, obtained using (10).

In section 4.1, a method to select ϵ_n is presented using simulations. In section 4.2, simulation comparisons are provided for ABC and F-ABC. The histograms are smoothed with the by default R-kernel in Figures 2 and 3. In Table 3, F-ABC for selected θ^* improves the concentration (MSE) of parametric ABC unlike Table 2. However, the F-ABC improvement holds in 48 out of 50 repetitions of the experiment.

In the remaining applications, preliminary approximate posteriors are obtained on $\mathbf{\Theta}$ that is subsequently restricted where $p_{match}(\theta^*)$ are positive; see, *e.g.* section 4.6. In section 4.3, Figure 5, ABC and F-ABC for all θ^* posteriors of means are obtained for a bivariate normal vector of correlated variables. In sections 4.4-4.7 posteriors are obtained for intractable models: Tukey's (a, b, g, h)-model, a normal mixture with parameters $(p, \mu_1, \sigma_1, \mu_2, \sigma_2)$, an autoregressive AR(1) model and a Quantile model.

4.1 ϵ_n and matching support probability α in practice

The goal is to implement the selection of ϵ_n and α_n in **4)** of section 2 when $\rho = d_K$. As illustration, Table 1 is provided for a sample of $n = 100$ normal random variables with mean θ and variance 1. With the notation in **4)** of section 2, $\theta_b^* = \theta = 0$ and $\mathbf{x}$ is obtained. $M = 500$ samples[10] are obtained for each θ_b^* and $\theta^* = .5, (.5), 4$ and d_K-distances are calculated; .5 corresponds to .5 standard deviation of the assumed location model. If $\epsilon = .63$ is used, with coordinates ($\theta^* = 1.5$, Quantile $= 95th$) in Table 1, it is expected that θ^* in the range $(-1.5, 1.5)$ are selected and the observed matching support probability

[10] $M = 500 > 200$ to increase table's accuracy, with execution time less than 15 seconds.

(Definition 2.2) will be (at least) .95. The dependence of ϵ and $\epsilon_{n,B}$ in the distance between F_θ and F_{θ^*} is observed in Table 1. The form of the obtained marginal posterior can lead to ϵ's fine tuning. The form of θ^* used to compare with θ_b^* will depend on the nature of the parameter. When θ_b^* is scale parameter, $\theta^* = c \cdot \theta_b^*$, *e.g.* with $c \in (0, 3]$. Alternatively, the upper bound $\epsilon_{n,B}$ for ϵ_n in section 5.2 can also be used and led us to choose in examples with real observations ϵ in $[n^{-.5}, 3n^{-.5}]$.

Empirical Quantiles of Kolmogorov distances between $\hat{F}_{\mathbf{x}}$ and $\hat{F}_{\mathbf{x}^*}$											
θ^*	MIN	25th	50th	60th	65th	70th	75th	80th	85th	90th	95th	MAX
0	0.04	0.07	0.09	0.1	0.1	0.11	0.11	0.12	0.12	0.13	0.14	0.19
0.5	0.12	0.2	0.23	0.24	0.25	0.25	0.26	0.27	0.28	0.29	0.3	0.39
1	0.25	0.38	0.41	0.42	0.42	0.43	0.44	0.44	0.45	0.46	0.48	0.55
1.5	0.47	0.55	0.57	0.58	0.59	0.59	0.6	0.61	0.61	0.62	0.63	0.69
2	0.6	0.68	0.71	0.71	0.72	0.72	0.73	0.73	0.74	0.75	0.76	0.79
2.5	0.72	0.8	0.82	0.83	0.83	0.83	0.84	0.84	0.85	0.86	0.87	0.91
3	0.82	0.89	0.9	0.91	0.91	0.91	0.92	0.92	0.92	0.93	0.93	0.95
3.5	0.89	0.94	0.95	0.96	0.96	0.96	0.96	0.96	0.97	0.97	0.97	0.99
4	0.94	0.97	0.98	0.98	0.98	0.99	0.99	0.99	0.99	0.99	1	1

Table 1: Potential ϵ_n-values the Quantiles, for matching support $\alpha, 0 < \alpha < 1$.

4.2 Comparison of parametric ABC with F-ABC

In simulations, we compare parametric ABC with F-ABC *for all* and F-ABC *for the selected* θ^*, neglecting **2)** *c)* of the F-ABC *Algorithm.* Remark 2.1 is followed. More precisely, we start ABC with d_K and ϵ and for the selected θ_i^* we draw M additional $\mathbf{x}^*$ to compute $p_{match}(\theta_i^*)$. The F-ABC posterior for these selected θ^* is obtained. The process is repeated for the non-selected θ^* in ABC and the F-ABC for all θ^* drawn posterior is obtained. Details follow.

An ABC example is used from Tavaré (2019, # 2, "A Normal example", p. 35). $X_1, \ldots, X_n$ are *i.i.d.* normal random variables, $\mathcal{N}(\theta, \sigma^2 = 1)$, denoted by $\mathbf{X}$. The prior for θ is uniform $U(a, b)$ with $a \to -\infty$ and $b \to \infty$. Attention is restricted to the sample mean, $\bar{X}_n$, since it is sufficient statistic. For fixed a, b the posterior $\pi(\theta|\bar{X}_n)$ is $\mathcal{N}(\theta, \frac{\sigma^2}{n})$ truncated in (a, b). For the ABC-simulations and a given ϵ^* it is assumed the observed $\bar{x}_n = 0$, θ^* is observed from $U(a, b)$ and is selected when $\rho(\bar{x}_n^*, \bar{x}_n = 0) = |\bar{x}_n^*| \le \epsilon^*$; $|\cdot|$ is absolute value. A flat, "0-1", kernel is used to select θ^*.

For nonparametric ABC with $d_K, \hat{F}_{\mathbf{x}}$ is used and ϵ is such that the number of selected θ^* from $U(-1, 1)$ does not differ much from that of the parametric ABC. The number of drawn θ^* is large, $N^* = 1,000$, such that the number of θ^* selected (N in Figures 2 and 3) is also large enough for determining the approximate posterior. Sample $\mathbf{X}_i^*$ is obtained from $\mathcal{N}(\theta_i^*, 1)$ and θ_i^* is selected if $d_K(\hat{F}_{\mathbf{x}}, \hat{F}_{\mathbf{x}_i^*}) \le \epsilon,\ i = 1, \ldots, N^*$. For F-ABC, $M = 200$ $\mathbf{X}^*$-samples of size n are drawn for each selected θ^*, but also for non-selected θ^*.

We used $n = 200, \theta = 0, a = -1, b = 1; \epsilon^* = .15, \epsilon = .12$ are both in $[n^{-.5}, 3n^{-.5}]$.

In Tables 2 and 3, simulation results are presented where the MSE of each method dominates the other. In Figures 2 and 3, frequency histograms are presented for ABC and F-ABC, and the corresponding density plots with Gaussian kernel. For the F-ABC approximate posteriors, the bandwidth was set at 0.05. Nonparametric F-ABC for selected θ^* is satisfactory compared with parametric ABC. We prefer F-ABC *for all* θ^*.

In several simulations, very frequently, the concentration (MSE) of the nonparametric F-ABC for the selected θ^* improves that of parametric ABC. To compare the MSE improvement with F-ABC for selected θ^*, 1000 MSE comparisons are made and the total number of times F-ABC improves ABC is recorded. The parameters are $\epsilon = .12, \epsilon^* = .15, n = 100, \theta = 0, a = -1, b = 1, N^* = 100, M = 100$. The process is repeated 50 times out of which 48 times F-ABC for selected θ^* improves the MSE of parametric ABC.

Concentration: Non Parametric ABC, F-ABC selected/drawn-Parametric ABC				
Nonparametric , $\epsilon = .12$				Parametric, $\epsilon^* = .15$
Parameter	ABC	F-ABC selected θ^*	F-ABC all drawn θ^*	ABC
Mean θ^*_{select}	- 0.0916	-0.0865	-0.0859	-0.0117
Variance θ^*_{select}	0.0182	0.0105	0.0274	0.0107
MSE θ^*_{select}	0.0266	0.018	0.0348	0.0108

Table 2: Mean, Variance and MSE of θ^*_{select}

Concentration: Non Parametric ABC, F-ABC selected/drawn-Parametric ABC				
Nonparametric , $\epsilon = .12$				Parametric, $\epsilon^* = .15$
Parameter	ABC	F-ABC selected θ^*	F-ABC all drawn θ^8	ABC
Mean θ^*_{select}	-0.00198	-0.00185	-0.00617	0.0112
Variance θ^*_{select}	0.0187	0.0111	0.0242	0.0138
MSE θ^*_{select}	0.0187	0.0111	0.0243	0.0139

Table 3: Mean, Variance and MSE of θ^*_{select}

4.3 ABC and F-ABC for all θ^* in R^2 with $\tilde{\rho}$-distance via d_K

Nonparametric ABC and F-ABC *for all* are implemented for the means of a bivariate normal with dependent components and $\mathbf{X} = (X_1, \ldots, X_n)$. d_K is used for $\mathbf{X}^*$-matching over 1-dimensional projections of $\mathbf{X}$ and $\mathbf{X}^*$, in order to approximate Wolfowitz's half-spaces distance, $\tilde{\rho}(\mu_{\mathbf{X}}, \mu_{\mathbf{X}^*})$, by $\tilde{\rho}_n(\mu_{\mathbf{X}}, \mu_{\mathbf{X}^*})$, as explained in section 3. Bivariate posteriors are depicted in Figure 4.

Using the notation in section 3, for $a, y \in R^2$, $< a, y >$ is the inner product of y and a,

$|| \cdot ||$ is Euclidean distance in R^2, $a \cdot \mathbf{X} = (< a, X_1 >, \ldots, < a, X_n >) \in R^n$,

$$\tilde{\rho}_n(\mu_{\mathbf{X}}, \mu_{\mathbf{X}^*}) = \max_{a \in \{a_1, \ldots, a_{k_n}\} \subset U_2} d_K(\hat{F}_{a \cdot \mathbf{X}}, \hat{F}_{a \cdot \mathbf{X}^*});$$

$a_1, \ldots, a_{k_n}$ are are *i.i.d.* uniform random vectors in $U_2 = \{u = (u_1, u_2) \in R^2 : ||u|| = 1\}$, independent of $\mathbf{X}$ and $\mathbf{X}^*$. Direction a used in $\tilde{\rho}_n$ may have form $(\cos(\phi), \sin(\phi))$, with ϕ uniform in $[0, \pi)$. In practice, ϕ is obtained from a discretization of $[0, \pi)$. $\tilde{\rho}_n$ approximates $\tilde{\rho}$ in (24) when $k_n \uparrow \infty$, but a moderately large $k_n = k$ is adequate.

A sample $\mathbf{x}$ of size $n = 50$ is observed from a bivariate normal with means $\theta = (0, 2)$, variances 1 and covariance .5. Assume the parameter space for θ is $\mathbf{\Theta} = [-1, 2]x[-2, 3] \subset R^2$. Instead of drawing θ^* randomly from $\mathbf{\Theta}$, a discretization $\mathbf{\Theta}^*$ of $\mathbf{\Theta}$ is used in order to observe the weights $p_{match(\theta^*)}$ along $\mathbf{\Theta}$. Using 15 equidistant θ_1^* and θ_2^*, respectively, in $[-1, 2]$ and $[-2, 3]$, obtain $\theta^* = (\theta_1^*, \theta_2^*)$ in $\mathbf{\Theta}^*, N = card(\mathbf{\Theta}^*) = 225$. Following Remark 2.1, to obtain $\tilde{\rho}_n$-ABC and $\tilde{\rho}_n$-F-ABC posteriors, one sample $\mathbf{X}^*$ is drawn initially for each θ^* in $\mathbf{\Theta}^*$. $k = 50$ a-directions are used in $\tilde{\rho}_n$, $\epsilon = .33$ in $[n^{-.5}, 3n^{-.5}]$ and 21 $\mathbf{X}^*$ match $\mathbf{X}$, thus selecting 21 θ^* from $\mathbf{\Theta}^*$. With F-ABC for all $\theta^* \in \mathbf{\Theta}^*$, without using **2***c)* in the F-ABC Algorithm, $M = 200$ independent copies of $\mathbf{X}^*$ are obtained for each $\theta^* \in \mathbf{\Theta}^*$. For the same 50 a-directions and the $M + 1$ matchings, $p_{match}(\theta^*)$ in (7) is calculated for $\rho = \tilde{\rho}_n$ and $\epsilon = .33$.

In Figure 4, the nonparametric ABC-posterior density (in green) and the F-ABC for all θ^* posterior histogram and density appear, created with R-functions *persp*, *hist*3*D* and *persp*3*D*, respectively. Comparison of the ABC and F-ABC densities indicates higher concentration in the latter near the means $(0, 2)$. In ABC (all green), the density's shape and the 0-values in the z-axis are due to the bivariate normal kernel used by default in R-function *kde*2*d* needed in *persp*. In F-ABC for all, no kernel is used in the histogram (in the middle). The matching proportions, $p_{match}(\theta^*)$, provide the z-values in Figure 4.

4.4 F-ABC for all with Tukey's (a, b, g, h)-model

Tukey's g-and-h model (see, e.g., Tukey, 1977) accommodates non-Gaussian data. The parameters, including location and scale are: $g(\in R)$ controlling skewness, $h(\geq 0)$ controlling tail heaviness, $a(\in R)$ for location and $b(> 0)$ for scale. Standard normal r.vs $Z_1, \ldots, Z_n$ are used to generate $\mathbf{X} = (X_1, \ldots, X_n)$,

$$X_i = a + b\frac{e^{gZ_i} - 1}{g} e^{.5hZ_i^2}, \; i = 1, \ldots, n. \tag{26}$$

The observed sample $\mathbf{X} = \mathbf{x}$ consists of $n = 20000$ i.i.d. r.vs[11] obtained from (26) with $a = 3, b = 4, g = 3.5, h = 2.5$. Parameter spaces are $\Theta_a = [2.5, 3.5], \Theta_b = [3.5, 4.5], \Theta_g = [3, 4], \Theta_h = [2, 3]$, and each interval is divided in 10 equal sub-intervals with the 11 end-points used to obtain for $\mathbf{\Theta} = \Theta_a x \Theta_b x \Theta_g x \Theta_h$ discretization $\mathbf{\Theta}^*$ with cardinality $N = 11^4$. $M = 50$ samples of size n are obtained using each element of $\mathbf{\Theta}^*$ with $\epsilon = .01$ in $[n^{-.5}, 3n^{-.5}]$. Smooth histograms for the posterior of each parameter are in Figure 5, and the corresponding histograms with weights the matching support proportions are in Figure 6.

The process was repeated with enlarged $\Theta_b = [3.2, 4.8]$ and discretization $\mathbf{\Theta}^*$ with cardinality $N = 21^4$ and $M = 100$. The maximum value of the weight $p_{match}(\theta^*)$ is .94, achieved at $\theta^* = (3, 4.08, 3.5, 2.5)$. Smooth histograms for the posterior of each parameter are in Figure 7, and the corresponding histograms with weights the matching support proportions in Figure 8.

4.5 F-ABC for all with a 5-parameters Normal mixture

The observed $\mathbf{X} = \mathbf{x}$, is realization of $n = 5000$ independent r.vs from a Normal mixture with two components, means $\mu_1 = 1, \mu_2 = 6$, standard deviations $\sigma_1 = 1, \sigma_2 = 1.5$ and

[11] The sample size increased in the remaining applications for more accurate posteriors.

weights, respectively, $p = p_1 = .3, p_2 = 1-p = .7$. Parameter space $\Theta_p = [0,1]$ is divided in 20 equal sub-intervals with the 21 end-points in its discretization and $\Theta_{\mu_1} = [.5, 1.5], \Theta_{\mu_2} = [5.5, 6.5], \Theta_{\sigma_1} = [.5, 1.5], \Theta_{\sigma_2} = [1, 2]$ are divided each in 10 equal sub-intervals with the 11 end-points used to obtain for $\mathbf{\Theta} = \Theta_p x \Theta_{\mu_1} x \Theta_{\sigma_1} x \Theta_{\mu_2} x \Theta_{\sigma_2}$ discretization $\mathbf{\Theta}^*$ with cardinality $N = 21x11^4$. $M = 50$ samples of size n are obtained for each element of $\mathbf{\Theta}^*$ and $\epsilon = .03$ in $[n^{-.5}, 3n^{-.5}]$. The F-ABC *for all* marginal densities are in Figure 9, using notation for the means $m1, m2$ and for the standard deviations $s1, s2$. The corresponding frequency histograms with weights the matching support proportions, $p_{match}(\theta^*)$, are in Figure 10.

4.6 F-ABC for all with an AR(1) model

$\mathbf{X}$ is observed from an autoregressive AR(1) model, with X_1 having a normal distribution with mean 0 and variance $\sigma^2 = b^2/(1-a^2)$, and

$$X_t = aX_{t-1} + bZ_t, \qquad -1 < a < 1,\ b > 0; \tag{27}$$

Z_t is standard normal independent of $X_{t-1},\ t > 1$. X_t has the same distribution as $X_1, t > 1$. Parameters a, b are not identifiable due to the form of the variance. The vector (X_t, X_{t-1}) has stationary normal distribution with mean $(0,0)$, and covariance matrix $\Sigma(\theta), \theta = (a, b)$, with variances $b^2/(1-a)^2$, and covariance $ab^2/(1-a)^2$, and a, b are identifiable; see *e.g.* Bernton *et al.* (2019).

Model parameters $a = 0.5, b = 1$ are used to obtain $n = 1000$ X's from (27). In preliminary application of F-ABC *for all*, with $\epsilon = .08$ in $[n^{-.5}, 3n^{-.5}]$, the assumed parameter spaces $\Theta_a = [-.99, .99]$ and $\Theta_b = [0.5, 2]$, are divided each in 14 equal sub-intervals with the 15 end-points in each discretization to obtain for $\mathbf{\Theta} = \Theta_a x \Theta_b$ discretization $\mathbf{\Theta}^*$ with cardinality $N = 15^2$. $n = 999$ matching observations are obtained from a bivariate normal with means 0 and covariance $\Sigma(\theta^*)$ for each $\theta^* \in \mathbf{\Theta}^*$. The number of repeated samples for each θ^* is $M = 200$ and the number of projection directions used is $k = 60$. The posterior of (a, b) is concentrated in a neighborhood of $(0.5, 1)$.

For a more accurate posterior of (a, b), $n = 5000$ and $\epsilon = 0.03$ (in $[n^{-.5}, 3n^{-.5}]$) are used, and the parameter spaces are restricted to $\Theta_a = [0, .99)$ and $\Theta_b = [0.5, 1.5]$. For $\mathbf{\Theta} = \Theta_a x \Theta_b$, the discretization $\mathbf{\Theta}^*$ has cardinality $N = 25^2$. The maximum value of the weight $p_{match}(\theta^*)$ is .84 and is achieved at $\theta^* = (0.52, .98)$ and $\theta^{**} = (0.54, .98)$. In Figure 11, the F-ABC *for all* bivariate frequency histogram and its smooth histogram are depicted, and the corresponding marginals are in Figure 12. Bernton *et al.* (2019) use for matching bivariate observations $(x_{2k-1}, x_{2k}), k \geq 1$, and there is no indication for the mode of the approximate posterior.

4.7 F-ABC for all with a Quantile model

Observations X_t are obtained from a data-generating model borrowed from stochastic volatility models (Kim *et al.*, 1998),

$$X_t = b\epsilon_t e^{.5\eta_t}, \tag{28}$$

with the unobserved $\eta_t \sim N(0, a^2), \epsilon_t \sim N(0, 1)$. The parameter of interest is $\theta = (a, b)$.

The model parameters used to obtain $\mathbf{X}$ are $a = .8, b = .65$. For more accurate posterior of (a, b), we restricted the parameter space to $\Theta_a = \Theta_b = [.5, 1.5]$, divided Θ_a in 20 equal sub-intervals and Θ_b in 120 equal sub-intervals including the end-points in each discretization to obtain for $\mathbf{\Theta} = \Theta_a x \Theta_b$ discretization $\mathbf{\Theta}^*$ with cardinality $N = 21x121$. We used $M = 400, n = 10000$ and $\epsilon = 0.01$ in $[n^{-.5}, 3n^{-.5}]$. The maximum value of the weight $p_{match}(\theta^*)$ is .22 and was achieved at $\theta^* = (0.75, .64)$. The F-ABC *for all* posterior marginals appear in Figure 13.

5 The Matching tools: $\hat{F}_{\mathbf{X}}, \mu_{\mathbf{X}}, d_K, \tilde{\rho}, \tilde{\rho}_n, \epsilon, \alpha$ and $p_{match}(\theta^*)$

5.1 Pertinent properties of $\hat{F}_{\mathbf{X}}, \mu_{\mathbf{X}}, d_K, \tilde{\rho}$

$(\hat{F}_{\mathbf{X}}, d_K)$ and $(\mu_{\mathbf{X}}, \tilde{\rho})$ satisfy desired properties for summary statistics in ABC (Fearnhead and Prangle, 2012, Frazier *et al.*, 2018) with *binding function* $b(\theta)$, respectively, F_θ and the induced probability, P_θ. Assume a Dirichlet prior, $DP(\alpha, G)$, for $F_\theta, \theta \in \Theta$, then, see *e.g.* Walker *et al.*(1999),

$$E(F_\theta|\mathbf{X}) = \frac{n}{n+\alpha}\hat{F}_{\mathbf{X}} + \frac{\alpha}{n+\alpha}G.$$

Thus, for large n or when $\alpha = 0$, $E(F_\theta|\mathbf{X})$ is practically $\hat{F}_{\mathbf{X}}$, and the same holds for μ_n and P_θ. Also, *e.g.*, $F_{\theta_1} = F_{\theta_2}$ implies $\theta_1 = \theta_2$ due to identifiability, and if T is continuous with respect to d_K and a metric $d_{\mathbf{\Theta}}$ on $\mathbf{\Theta}$, it is expected that $T(\hat{F}_{\mathbf{X}})$ as estimate of $T(F_\theta)$ will inherit convergence properties of $\hat{F}_{\mathbf{X}}$ to F_θ. Similar results hold for $\mu_{\mathbf{X}}$ and P_θ, with $\mu_{\mathbf{X}}$ indexed by the class of half-spaces which is Vapnik-Cervonenkis class of index $(d+1)$, and Wolfowitz's half spaces distance $\tilde{\rho}$.

$d_K(\hat{F}_{\mathbf{x}^*}, \hat{F}_{\mathbf{x}})$ is not continuous function at $\mathbf{x}$ since it cannot be smaller than $\frac{1}{n}$ for *all* $\mathbf{x}^*$ at Euclidean distance $\delta > 0$ from $\mathbf{x}$. This makes d_K different from other ρ-distances used in ABC, (1), (2); see, *e.g.* Bernton *et al.* (2019, p. 39, proof of Proposition 3.1).

Lemma 5.1 *For any observed samples of size* n, $\mathbf{x}^* \neq \mathbf{x}_{\sigma(1:n)} \in R^d, d \geq 1$,

$$d_K(\hat{F}_{\mathbf{x}}, \hat{F}_{\mathbf{x}^*}) \geq \frac{1}{n}; \tag{29}$$

$\mathbf{x}_{\sigma(1:n)}$ *denotes a vector, permutation of the* $\mathbf{x}$ *components. Thus,*

$$d_K(\hat{F}_{\mathbf{x}}, \hat{F}_{\mathbf{x}^*}) = 0 \iff \mathbf{x}^* = \mathbf{x}_{\sigma(1:n)}. \tag{30}$$

5.2 On ϵ_n, α and $d_K, \tilde{\rho}, \tilde{\rho}_n$

For matching support probability α in (4), the F-ABC tolerance ϵ_n satisfies

$$P[d_K(\hat{F}_{\mathbf{X}^*}, \hat{F}_{\mathbf{X}}) > \epsilon_n] = 1 - \alpha, \ \ 0 \leq \alpha \leq 1. \tag{31}$$

An upper bound, $\epsilon_{n,B}$, on ϵ_n is obtained by equating in (31) an upper probability bound, $U(n, \epsilon_{n,B})$, with $1-\alpha$; see Lemma 7.1. Conditionally on $\mathbf{X} = \mathbf{x}$, $\epsilon_{n,B}(\mathbf{x})$ is similarly obtained under F_{θ^*}. The obtained bounds hold for observations in R. Similar results for $\epsilon_{n,B}$ hold with observations in $R^d, d > 1$, using either $\hat{F}_{\mathbf{X}}$ or $\mu_{\mathbf{X}}$ and are obtained as described after the Proof of Proposition 5.1, in Remark 7.1.

Proposition 5.1 *Let* $\mathbf{X}$ *be a sample of* n *random variables from cumulative distribution* F_θ, *with* θ *unknown, let* $\mathbf{X}^*$ *be a simulated* n*-size sample from a sampler used for* θ^* *and let* α *be the matching support probability for the tolerance* ϵ_n *in (31);* $0 \leq \alpha < 1$.

a) The upper bound for ϵ_n *is*

$$\epsilon_{n,B}(\theta, \theta^*) = d_K(F_\theta, F_{\theta^*}) + \sqrt{\frac{2}{n} \ln \frac{4}{1-\alpha}} \geq \sqrt{\frac{2}{n} \ln 4}. \tag{32}$$

b) Conditionally on $\mathbf{X} = \mathbf{x}$, *the upper bound for* ϵ_n *is*

$$\epsilon_{n,B}(\mathbf{x}, \theta^*) = d_K(\hat{F}_{\mathbf{x}}, F_{\theta^*}) + \sqrt{\frac{1}{2n} \ln \frac{2}{1-\alpha}} \geq \delta_n(\mathbf{x}, \theta^*) + \sqrt{\frac{1}{2n} \ln 2}. \tag{33}$$

In practice, $\min\{\epsilon_{n,B}(\theta, \theta^*), 1\}$ *and* $\min\{\epsilon_{n,B}(\mathbf{x}, \theta^*), 1\}$ *are used.*

(32) and (33) provide a structure for $\epsilon_{n,B}$. We preferred to use the lower bound in (32) since both summands do not depend on $\mathbf{x}$. This has led us to adopt after numerous simulations with observations in R, ϵ_n in $[n^{-.5}, 3n^{-.5}]$, modulo potential adjustments for the dimension of θ; note that $2 \ln 4$ in (32) is in $[1, 3]$. ϵ_n can be also determined via simulations; see Table 1, section 4.1, but it can be time consuming.

5.3 Motivation for $p_{match}(\theta^*)$

F-ABC is a nonparametric extension of ABC methods, with main differences already pesented in the Introduction. The use of $p_{match}(\theta^*)$ was motivated from the observation in several models that for the estimate $S(\mathbf{X})$ of $T(F_\theta)$ and $\tilde{d}, \rho$ generic distances:

$$\textit{when} \qquad d_{\mathbf{\Theta}}(\theta_1^*, \theta) \le d_{\mathbf{\Theta}}(\theta_2^*, \theta) \Rightarrow \tilde{d}(F_{\theta_1^*}, F_\theta) \le \tilde{d}(F_{\theta_2^*}, F_\theta) \tag{34}$$

$$\Rightarrow \ \forall\, \epsilon > 0, \qquad P_{\theta_2^*}[\rho(S(\mathbf{X}^*), T(F_\theta)) \le \epsilon] \le P_{\theta_1^*}[\rho(S(\mathbf{X}^*), T(F_\theta)) \le \epsilon]. \tag{35}$$

The implications in (34) and (35) hold often, *e.g.* for the normal model, with mean θ and variance 1, $d_{\mathbf{\Theta}} = \rho = |.|,\ \tilde{d} = d_K, S(\mathbf{X}) = \bar{X}_n,\ T(F_\theta) = \theta.$

In F-ABC in particular, with $\tilde{d} = \rho = d_K, T(F_\theta) = F_\theta, S(\mathbf{X}^*) = \hat{F}_{\mathbf{X}^*}$, (35) will also hold, at least for large n, when F_θ is replaced by $\hat{F}_{\mathbf{X}}$. Then, for families of *c.d.fs* in R with densities f_θ such that $f_{\theta_1^*}(x) - f_{\theta_2^*}(x)$ changes sign once, the upper bound in (35) increases to 1 with n if $\theta_{1,n}^*$ gets closer to θ (Yatracos, 2020, Propositions 7.2, 7.4 and Remark 7.2). The same holds in general, as Proposition 5.2 confirms when taking limits in (36) as $n \uparrow \infty$. The lower bound in (36) is also lower bound on the Probabilities in (35). Thus, it is expected the F-ABC posteriors concentrate near θ.

Proposition 5.2 *For n i.i.d. random vectors in R^d with c.d.f. F_{θ^*} and n large:*

$$P_{\theta^*}[d_K(F_{\mathbf{X}^*}, \hat{F}_{\mathbf{X}}) \le \epsilon_n] \ge 1 - C_1^*(d) \cdot \exp\{-n \cdot C_2^*(d) \cdot (\epsilon_n - d_K(F_{\theta^*}, F_\theta))^2\}; \tag{36}$$

$C_1^*(d),\ C_2^*(d)$ *are positive constants.*

$p_{match}(\theta^*)$ is also useful in the approximation of

$$E[h(\Theta)|\mathbf{X} = \mathbf{x}] = \int_{\mathbf{\Theta}} h(\theta)\pi(\theta|\mathbf{x})d\theta; \tag{37}$$

$\boldsymbol{\Theta} \subset R^k$. In F-ABC, (37) is approximated using $\boldsymbol{\Theta}_n^*$ in (9) which includes all θ^* drawn with F-ABC for all,

$$\int_{\boldsymbol{\Theta}} h(\theta)\pi(\theta|\mathbf{x})d\theta \approx \sum_{i=1}^{N^*} h(\theta_i^*)p_{match}(\theta_i^*). \tag{38}$$

6 Asymptotics

Under few, mild assumptions, results are obtained for Kolmogorov distance, d_K, when $\mathbf{X} \in R^{nxd}$, which hold also for the stronger distance, $\tilde{\rho}$, in (24).

In ABC, one question of interest is whether $\pi_{abc}(\theta|B_\epsilon)$ converges to $\pi(\theta|\mathbf{x})$ when $\mathbf{x}$ stays fixed and $\epsilon = \delta_m \downarrow 0$ as m increases.

Proposition 6.1 *Use the notation in section 2, for* ABC *and* F-ABC *with* $S(\mathbf{X}) = \hat{F}_{\mathbf{X}}, \rho = d_K$, *$n$ fixed and* B_{ϵ_n} *in (11). Under the exchangeability assumption, i.e.* $f(\mathbf{y}|\theta) = f(\mathbf{y}_{\sigma(1:n)}|\theta)$ *for any permutation* $\mathbf{y}_{\sigma(1:n)}$ *of* $\mathbf{y}$, *and with* ϵ_n *replaced by* $\delta_m \downarrow 0$ *as* m *increases,*

$$\lim_{m\to\infty} \pi_u(\theta|B_{\delta_m}) = \pi(\theta|\mathbf{x}), \qquad u = abc, f\text{-}abc. \tag{39}$$

For continuous $\mathbf{X}$, $(\mathcal{Y}, \mathcal{C}_{\mathcal{Y}})$ *is* R^{nxd} *with the Borel sets,* $\mathcal{B}$, *and* Θ *takes values in* $R^k, k \leq d$.

Another question of interest for ABC is whether the posterior $\pi_{abc}(\theta|B_{\epsilon_n})$ will place increasing probability mass around θ as n increases to infinity (Fearnhead, 2018), *i.e.* Bayesian consistency. Posterior concentration is proved for ABC and F-ABC, initially for fixed size ζ-neighborhood when $T(F_\theta)$ is the quantity of interest; T is a functional, $\zeta > 0$.

Proposition 6.2 *Use the notation in section 2 and let* $\mathcal{F}_{\boldsymbol{\Theta}} = \{F_\theta, \theta \in \boldsymbol{\Theta}\}$ *be subset of a metric space* $(\mathcal{F}, d_{\mathcal{F}})$ *of c.d.fs. Assume*

a) $d_{\mathcal{F}}(\hat{F}_{\mathbf{X}}, F_\theta) \leq \frac{o(k_n)}{k_n}, k_n \uparrow \infty$ *and* $P_\theta^{(n)}$*-probability* $\uparrow 1$, *as* n *increases, and* *b) T is a continuous functional on $\mathcal{F}$ with values in a metric space $(\mathcal{T}, d_{\mathcal{T}})$.*

Then, for ABC *and* F-ABC, $S(\mathbf{X}) = \hat{F}_{\mathbf{X}}, \rho = d_{\mathcal{F}}$ *and for any* $\zeta > 0$,

$$\lim_{n\to\infty} \Pi_u[\theta^* : d_{\mathcal{T}}(T(F_{\theta^*}), T(F_\theta)) \leq \zeta | B_{\epsilon_n}] = 1, \qquad u = abc, f\text{-}abc; \tag{40}$$

$$B_{\epsilon_n} = \{\mathbf{x}^* : d_{\mathcal{F}}(\hat{F}(\mathbf{x}^*), \hat{F}(\mathbf{x})) \leq \epsilon_n\}, \epsilon_n \downarrow 0 \text{ as } n \uparrow \infty. \tag{41}$$

Remark 6.1 *In Proposition 6.2, assumption a) holds for* i.i.d *samples with $d_{\mathcal{F}} = d_K$ and $k_n = \sqrt{n}$. Different k_n can be obtained under dependence via Large Deviations bounds. Special case of interest in b) when $T(F_\theta) = \theta$ and $d_{\mathcal{T}} = d_{\mathbf{\Theta}}$, the metric on $\mathbf{\Theta}$.*

To confirm Bayesian consistency for shrinking $d_{\mathcal{T}}$-neighborhoods of $T(F_\theta)$, let w be the modulus of continuity of T, *i.e.*

$$w(\tilde{\epsilon}) = \sup\{d_{\mathcal{T}}(T(F_\theta), T(F_\eta)) : d_{\mathcal{F}}(F_\theta, F_\eta) \leq \tilde{\epsilon}; \theta \in \mathbf{\Theta}, \eta \in \mathbf{\Theta}\}, \ \tilde{\epsilon} > 0. \tag{42}$$

Consistency was established for ζ-$d_{\mathcal{T}}$-neighborhood of $T(F_\theta)$ when (56) holds, *i.e.* when

$$\epsilon_n \leq \tilde{\epsilon} - \frac{2o(k_n)}{k_n},$$

thus it holds for the smallest $\tilde{\epsilon}$-value,

$$\tilde{\epsilon} = \epsilon_n + \frac{2o(k_n)}{k_n} \tag{43}$$

and since for ζ_n-$d_{\mathcal{T}}$-neighborhood of $T(F_\theta)$

$$\zeta_n = w(\tilde{\epsilon})$$

it follows that

$$\zeta_n = w(\epsilon_n + \frac{2o(k_n)}{k_n}) \geq w(\frac{2o(k_n)}{k_n}). \tag{44}$$

Lemma 6.1 *Under the assumptions of Proposition 6.2, the shortest $d_{\mathcal{T}}$-shrinking neighborhood of $T(F_\theta)$ for which Bayesian consistency holds has radius $w(\epsilon_n + \frac{2o(k_n)}{k_n}) \geq w(\frac{2o(k_n)}{k_n})$.*

Remark 6.2 *The rate of posterior concentration around $T(F_\theta)$ depends, as expected, on the rate in probability, k_n^{-1}, of the $d_{\mathcal{F}}$-concentration of $T(\hat{F}_{\mathbf{X}})$ around $T(F_\theta)$ which is not under the user's control, the tolerance ϵ_n and the modulus of continuity, w, of T. Similar conclusions in a different set-up have been obtained by Frazier* et al. *(2018).*

Acknowledgments

Many thanks are due to Professor Rudy Beran for communicating pertinent useful results in his 1986 paper with Professor Warry Millar.

7 Appendix

Proof of Lemma 5.1: The smaller d_K-distance between $\hat{F}_{\mathbf{x}}$ and $\hat{F}_{\mathbf{x}^*}$ occurs when $\mathbf{x}, \mathbf{x}^*$ differ by a small $\delta > 0$ in one coordinate of one observation and their distance is $\frac{1}{n}$. $\square$

Lemma 7.1 *Let $\mathbf{X} = \mathbf{x}, \mathbf{X}^* = \mathbf{x}^*$ and let $U(n,\epsilon)$ be positive function defined for positive integers n and $\epsilon > 0, 0 \leq \alpha \leq 1$, such that*

$$1 - \alpha = P[d_K(\hat{F}_{\mathbf{x}}, \hat{F}_{\mathbf{x}^*}) > \epsilon] \leq U(n, \epsilon). \tag{45}$$

Let $\epsilon_B : U(n, \epsilon_B) = 1 - \alpha$. Then $\epsilon_B \geq \epsilon$.

Proof of Lemma 7.1: Since $U(n, \epsilon_B) = 1 - \alpha$,

$$P[d_K(\hat{F}_{\mathbf{x}}, \hat{F}_{\mathbf{x}^*}) > \epsilon_B] \leq U(n, \epsilon_B) = 1 - \alpha = P[d_K(\hat{F}_{\mathbf{x}}, \hat{F}_{\mathbf{x}^*}) > \epsilon]$$

which implies $\epsilon_B \geq \epsilon$. $\square$

Theorem 7.1 *(Dvoretzky, Kiefer and Wolfowitz, 1956, and Massart, 1990, providing the tight constant) Let $\hat{F}_{\mathbf{Y}}$ denote the empirical c.d.f of the size n sample $\mathbf{Y}$ of i.i.d. random variables obtained from cumulative distribution F. Then, for any $\epsilon > 0$,*

$$P[d_K(\hat{F}_{\mathbf{Y}}, F) > \epsilon] \leq U_{DKWM} = 2e^{-2n\epsilon^2} \tag{46}$$

Proof of Proposition 5.1: *a)*

$$P[d_K(\hat{F}_{\mathbf{X}^*}, \hat{F}_{\mathbf{X}}) > \epsilon_n] \le P[d_K(\hat{F}_{\mathbf{X}^*}, F_{\theta^*}) + d_K(F_{\theta^*}, F_\theta) + d_K(F_\theta, \hat{F}_{\mathbf{X}}) > \epsilon_n]$$

$$\le P[d_K(\hat{F}_{\mathbf{X}^*}, F_{\theta^*}) > \frac{\epsilon_n - d_K(F_{\theta^*}, F_\theta)}{2}] + P[d_K(\hat{F}_{\mathbf{X}}, F_\theta) > \frac{\epsilon_n - d_K(F_{\theta^*}, F_\theta)}{2}]$$

$$\le 4\exp\{-\frac{n}{2}(\epsilon_n - d_K(F_{\theta^*}, F_\theta))^2\}$$

The right side of the last inequality, obtained from (46), is made equal to $1-\alpha$,

$$4\exp\{-\frac{n}{2}(\epsilon_{n,B} - d_K(F_{\theta^*}, F_\theta))^2\} = 1 - a \iff \epsilon_{n,B} = d_K(F_{\theta^*}, F_\theta) + \sqrt{\frac{2}{n}\ln\frac{4}{1-\alpha}}.$$

b) $P[d_K(\hat{F}_{\mathbf{X}^*}, \hat{F}_{\mathbf{x}}) > \epsilon_n] \le P[d_K(\hat{F}_{\mathbf{X}^*}, F_{\theta^*}) + d_K(F_{\theta^*}, \hat{F}_{\mathbf{x}}) > \epsilon_n] \le 2\exp\{-2n(\epsilon_n - d_K(F_{\theta^*}, \hat{F}_{\mathbf{x}}))^2\}$ obtaining with matching support probability α,

$$\epsilon_{n,B}(\mathbf{x}) = d_K(F_{\theta^*}, \hat{F}_{\mathbf{x}}) + \sqrt{\frac{1}{2n}\ln\frac{2}{1-\alpha}}. \qquad \square$$

Generalizations of (46) in R^d have been obtained, at least, by Kiefer and Wolfowitz (1958), Kiefer (1961) and Devroye (1977); $d > 1$. The differences in upper bound U in (46) are in the multiplicative constant, in the exponent of the exponential and on the sample size for which the exponential bound holds which may also depend on ϵ. The constants used are not determined except for Devroye (1977).

For example, following the Proof in Proposition 5.1 *b)*, conditionally on $\mathbf{X} = \mathbf{x}$:

i) Using Kiefer and Wolfowitz (1958), with the upper bound in (46) $U_{KW} = C_1(d)e^{-C_2(d)n\epsilon^2}$,

$$\epsilon_{n,B}(\mathbf{x}, \theta^*) = d_K(\hat{F}_{\mathbf{x}}, F_{\theta^*}) + \sqrt{\frac{1}{nC_2(d)}\ln\frac{C_1(d)}{1-\alpha}}.$$

ii) Using Kiefer (1961), with the upper bound in (46) $U_K = C_3(b,d)e^{-(2-b)n\epsilon^2}$, for every $b \in (0,2)$,

$$\epsilon_{n,B}(\mathbf{x}, \theta^*) = d_K(\hat{F}_{\mathbf{x}}, F_{\theta^*}) + \sqrt{\frac{1}{n(2-b)}\ln\frac{C_3(b,d)}{1-\alpha}}.$$

iii) Using Devroye (1977), with the upper bound in (46) $U_{De} = 2e^2(2n)^d e^{-2n\epsilon^2}$ valid for $n\epsilon^2 \ge d^2$,

$$\epsilon_{n,B}(\mathbf{x}, \theta^*) = d_K(\hat{F}_{\mathbf{x}}, F_{\theta^*}) + \sqrt{\frac{1}{2n}[\ln\frac{2}{1-\alpha} + 2 + d\ln(2n)]}.$$

Remark 7.1 *A lower bound for $\epsilon_{n,B}$ as those in Proposition 5.1 can be obtained, for $\tilde{\rho}_n(\mu_{\mathbf{X}}, \mu_{\mathbf{X}^*})$ in (25) using k_n and one of U_{KW}, U_K, U_{De}, and for $\tilde{\rho}(\mu_{\mathbf{X}}, \mu_{\mathbf{X}^*})$ using Vapnik-Cervonenkis inequality. In the latter, the lower bound has form $C_d\sqrt{\frac{\log n}{n}}$, and can be used to provide an interval for ϵ_n.*

Proof of Proposition 5.2: Follows along the first three lines in the proof of Proposition 5.1 *a)*, with the exponential upper bound obtained using the U_{KW} in *i)* (Kiefer and Wolfowitz, 1958), with $C_1^*(d), C_2^*(d)$ the adjustments of $C_1(d), C_2(d)$. □

Proof of Proposition 6.1: The arguments used for ABC hold for F-ABC.

a) $\mathcal{Y}$ discrete: The ABC posterior with $\rho = d_K$ in (12) is

$$\pi_{abc}(\theta|B_{\delta_m}) = \frac{\pi(\theta) \cdot \int_{\mathcal{Y}} I_{B_{\delta_m}}(\mathbf{y}^*) f(\mathbf{y}^*|\theta)\mu(d\mathbf{y}^*)}{\int_{\mathbf{\Theta}} \pi(s) \int_{\mathcal{Y}} I_{B_{\delta_m}}(\mathbf{y}^*) f(\mathbf{y}^*|s)\mu(d\mathbf{y}^*)\ \nu(ds)}.$$

With integral denoting sum, it is enough to prove that the integral in the numerator of $\pi_{abc}(\theta|B_{\delta_m})$ is proportional to $f(\mathbf{x}|\theta)$.

For $A \in \mathcal{C}_{\mathcal{Y}}$, let

$$Q_\theta(A) = \int_A f(\mathbf{y}^*|\theta)\mu(d\mathbf{y}^*),\ A \in \mathcal{A}.$$

Q_θ is a probability measure on $\mathcal{C}_{\mathcal{Y}}$.

Since n and $\mathbf{x}$ are fixed, for $\delta_k \geq \frac{1}{n} > \delta_{k+1}$

$$B_{\delta_1} \supseteq B_{\delta_2} \supseteq \ldots \supseteq B_{\delta_k} \tag{47}$$

and from Lemma 5.1 for $m > k, B_{\delta_m} = \{\mathbf{x}_{\sigma(1:n)}\}$. Therefore,

$$\lim_{m\to\infty} B_{\delta_m} = \cap_{m=1}^{\infty} B_{\delta_m} = \{\mathbf{x}_{\sigma(1:n)}\} \tag{48}$$

and

$$\lim_{m\to\infty} \int_{\mathcal{Y}} I_{B_{\delta_m}}(\mathbf{y}^*) f(\mathbf{y}^*|\theta)\mu(d\mathbf{y}^*) \ = \lim_{m\to\infty} Q_\theta(B_{\delta m}) = Q_\theta(\cap_{m=1}^{\infty} B_{\delta_m}) = f(\mathbf{x}|\theta)\mu(\{\mathbf{x}_{\sigma(1:n)}\}), \tag{49}$$

with the last equality due to exchangeability of $f(\mathbf{x}|\theta)$.

b) $\mathcal{Y}$ continuous: Then, the right side of (49) vanishes, since $\mu(\{\mathbf{x}_{\sigma(1:n)}\}) = 0$. A different approach is used, via the notion of regular conditional probability.

When $\mathcal{Y}$ is a Euclidean space R^{nxd} with Borel σ-field, $\mathcal{B}_d$, and Θ takes values in $R^k, k \leq d$, the integral in the numerator of $\pi_{abc}(\theta|B_{\delta_m})$,

$$\int_{\mathcal{Y}} I_{B_{\delta_m}}(\mathbf{y}^*) f(\mathbf{y}^*|\theta)\mu(d\mathbf{y}^*)$$

is a regular conditional probability, $P[\mathbf{X}^* \in B|\Theta = \theta], B = B_{\delta_m}$ (Breiman, 1992, Chapter 4, p. 79, Theorem 4.34), *i.e.*, with θ fixed, it is a probability for $B \in \mathcal{B}_d$ and with fixed B it is a version of the conditional density, $\theta \in \mathbf{\Theta}$. Thus, for fixed θ, from (48),

$$\lim_{m\to\infty} P[\mathbf{X}^* \in B_{\delta_m}|\Theta = \theta] = P[\{\mathbf{x}_{\sigma(1:n)}\}|\Theta = \theta]$$

and due to exchangeability is proportional to $f(\mathbf{x}|\theta)$ *a.s.* . $\square$

Proof of Proposition 6.2: The arguments used for ABC hold for F-ABC. For the probability in (40), using (13) for ABC with

$$H = \{\theta^* : d_{\mathcal{T}}(T(F_{\theta^*}), T(F_\theta)) \leq \zeta\}, \tag{50}$$

$$\Pi_{abc}(H|B_{\epsilon_n}) = \frac{\int_{\mathbf{\Theta}} I_H(\theta^*)\pi(\theta^*) \cdot \int_{\mathcal{Y}} I_{B_{\epsilon_n}}(\mathbf{y}^*) f(\mathbf{y}^*|\theta^*)\mu(d\mathbf{y}^*)\nu(d\theta^*)}{\int_{\mathbf{\Theta}} \pi(s) \int_{\mathcal{Y}} I_{B_{\epsilon_n}}(\mathbf{y}^*) f(\mathbf{y}^*|s)\mu(d\mathbf{y}^*)\ \nu(ds)} = \frac{\int_{\mathbf{\Theta}} \pi(\theta^*) \cdot P_{\theta^*}^{(n)}(H \cap B_{\epsilon_n})\nu(d\theta)}{\int_{\mathbf{\Theta}} \pi(s) \cdot P_s^{(n)}(B_{\epsilon_n})\nu(ds)}. \tag{51}$$

$P_{\theta^*}^{(n)}(H \cap B_{\epsilon_n})$ in the numerators of (51) will be bounded below using continuity of T and triangular inequality.

Since T is continuous, for $\zeta > 0$ there is $\tilde{\epsilon} > 0$ such that if

$$d_{\mathcal{F}}(F_{\theta^*}, F_\theta) \leq \tilde{\epsilon} \text{ then } d_{\mathcal{T}}(T(F_{\theta^*}), T(F_\theta)) \leq \zeta,$$

and then from (41), (50)

$$P_{\theta^*}^{(n)}(H \cap B_{\epsilon_n}) \geq P_{\theta^*}^{(n)}\left[\{d_{\mathcal{F}}(F_{\theta^*}, F_\theta) \leq \tilde{\epsilon}\} \cap B_{\epsilon_n}\right]. \tag{52}$$

Since

$$d_{\mathcal{F}}(F_{\theta^*}, F_\theta) \leq d_{\mathcal{F}}(F_{\theta^*}, \hat{F}_{\mathbf{x}^*}) + d_{\mathcal{F}}(\hat{F}_{\mathbf{x}^*}, \hat{F}_{\mathbf{x}}) + d_{\mathcal{F}}(\hat{F}_{\mathbf{x}}, F_\theta) \tag{53}$$

if

$$d_{\mathcal{F}}(F_{\theta^*}, \hat{F}_{\mathbf{x}^*}) + d_{\mathcal{F}}(\hat{F}_{\mathbf{x}^*}, \hat{F}_{\mathbf{x}}) + d_{\mathcal{F}}(\hat{F}_{\mathbf{x}}, F_\theta) \leq \tilde{\epsilon} \text{ then } d_{\mathcal{F}}(F_{\theta^*}, F_\theta) \leq \tilde{\epsilon}$$

and therefore, for the right side of (52)

$$P_{\theta^*}^{(n)}[\{d_{\mathcal{F}}(F_{\theta^*}, F_\theta) \leq \tilde{\epsilon}\} \cap B_{\epsilon_n}) \geq P_{\theta^*}^{(n)}[\{d_{\mathcal{F}}(F_{\theta^*}, \hat{F}_{\mathbf{x}^*}) + d_{\mathcal{F}}(\hat{F}_{\mathbf{x}^*}, \hat{F}_{\mathbf{x}}) + d_{\mathcal{F}}(\hat{F}_{\mathbf{x}}, F_\theta) \leq \tilde{\epsilon}\} \cap B_{\epsilon_n}]. \tag{54}$$

From the assumptions,

$$d_{\mathcal{F}}(F_{\theta^*}, \hat{F}_{\mathbf{X}^*}) \leq \frac{o(k_n)}{k_n} \text{ and } d_{\mathcal{F}}(F_\theta, \hat{F}_{\mathbf{X}}) \leq \frac{o(k_n)}{k_n}$$

with $P_{\theta^*}^{(n)}$ and $P_\theta^{(n)}$ probabilities converging to one, respectively, and assuming $\mathbf{x}^*, \mathbf{x}$ are in these subsets the right side of (54)

$$P_{\theta^*}^{(n)}[\{d_{\mathcal{F}}(F_{\theta^*}, \hat{F}_{\mathbf{x}^*}) + d_{\mathcal{F}}(\hat{F}_{\mathbf{x}^*}, \hat{F}_{\mathbf{x}}) + d_{\mathcal{F}}(\hat{F}_{\mathbf{x}}, F_\theta) \leq \tilde{\epsilon}\} \cap B_{\epsilon_n}) \geq P_{\theta^*}^{(n)}[\{d_{\mathcal{F}}(\hat{F}_{\mathbf{x}^*}, \hat{F}_{\mathbf{x}}) \leq \tilde{\epsilon} - 2\frac{o(k_n)}{k_n}\} \cap B_{\epsilon_n}]. \tag{55}$$

For $\epsilon_n \downarrow 0$ as n increases, eventually

$$\epsilon_n \leq \tilde{\epsilon} - \frac{2o(k_n)}{k_n}, \tag{56}$$

and the right side of (55)

$$P_{\theta^*}^{(n)}[\{d_{\mathcal{F}}(\hat{F}_{\mathbf{x}^*}, \hat{F}_{\mathbf{x}}) \leq \tilde{\epsilon} - 2\frac{o(k_n)}{k_n}\} \cap B_{\epsilon_n}] = P_{\theta^*}^{(n)}[B_{\epsilon_n}]. \tag{57}$$

(40) follows from (52), (54)-(57) since, when taking the limit in (51) as n increases to infinity, for large n numerator and denominator coincide. $\square$.

References

[1] Beran, R. and Millar, P. W. (1986) Confidence Sets for a Multivariate Distribution. *Ann. Statist.* **14**, 431-443.

[2] Bernton, E. , Jacob, P. E., Gerber, M. and Robert, C. P. (2019) Approximate Bayesian computation with the Wasserstein distance.*JRSS B*, **81**, 235-269. arXiv:1905.03747v1

[3] Biau, G., Cérou, F. and Guyader, A. (2015) New insights into approximate Bayesian computation. *Annales de l' IHP (Probab. Stat.)* **51**, 376-403.

[4] Breiman, L. (2001) Statistical Modeling: The Two Cultures. *Stat. Science* **16**, 3, 199-231.

[5] Breiman, L. (1992) *Probability* Classics in Applied Mathematics, SIAM.

[6] Chaudhuri, S., Ghosh, S., Nott, D. and Pham, K. C. (2020) On a Variational Approximation based Empirical Likelihood ABC method. Preprint.

[7] Chen, L. and Wu, W. B. (2018) Concentration inequalities for empirical processes of linear time series. *J. Machine Learning Research* **18**, 1-46.

[8] Cramér, H. and Wold, H. (1936) Some theorems on distribution functions.*J. London Math. Soc.* **11**, 290-294.

[9] de Finetti, B. (1931) Funzione caratteristica di un fenomeno aleatorio. *Atti della R. Academia Nazionale dei Lincei, Serie 6. Memorie, Classe di Scienze Fisiche, Mathematice e Naturale,* **4**, 251-299.

[10] Devroye, L. P. (1977) A Uniform Bound for the Deviation of Empirical Distribution Functions. *J. Multiv. Anal.* **7**, 594-597.

[11] Dudley, R. M. (1984) A course on empirical processes. École d' Été de Probabilités de St. Flour. *Lecture Notes in Math.* **1097**, 2-142, Springer Verlag, New York.

[12] Dudley, R. M. (1978) Central limit theorem for empirical measures. *Ann. Prob.* **6**, 899-929.

[13] Dvoretzky, A., Kiefer, J. and and Wolfowitz, J. (1956) Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Stat.* **27**, 642-669

[14] Fearnhead, P. (2018) Asymptotics of ABC. *Handbook of Approximate Bayesian Computation*, Editor:Routledge Handbooks Online.

[15] Fearnhead, P. and Prangle, D. (2012) Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation. *J. R. Statist. Soc. B* **74**, 419-474.

[16] Fournier, N. and Guillin, A. (2015) On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields,* **162**, 707-738.

[17] Frazier, D. T., Martin, G. M., Robert, C. P. and Rousseau, J. (2018) Asymptotic properties of approximate Bayesian Computation. *Biometrika* **105**, 593-607.

[18] Hewitt, E. and Savage, L. J. (1955) Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.* **80**, 470-501.

[19] Kiefer, J. (1961) On Large Deviations of the Empiric D. F. of Vector Chance Variables and a Law of the Iterated logarithm. *Pacific J. of Mathematics* **11**, 649-660.

[20] Kiefer, J. and Wolfowitz, J. (1958) On the deviations of the empiric distribution function of vector chance variables. *Trans. Amer. Math. Soc.* **87**, 173-186.

[21] Kim, S., Shephard, N. and Chib, S. (1998) Stochastic volatility: likelihood inference and comparison with ARCH models. *The Review of Economic Studies*, **65**, 361-393.

[22] Lauritzen, S. (2007) Exchangeability and de Finetti's Theorem. Lecture Notes, University of Oxford,
http://www.stats.ox.ac.uk/∼steffen/teaching/grad/definetti.pdf

[23] Lintusaari, J., Gutmann, M. U., Dutta, R., Kaski, S. and Corander, J. (2017) Fundamentals and Recent Developments in Approximate Bayesian Computation. *Syst. Biol.* **66**, e66-e82.

[24] Massart, P. (1990) The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Ann. Prob.* **18**, 1269-1283

[25] Miller, J. W. and Dunson, D. B. (2019) Robust Bayesian inference via coarsening. *J. Am. Stat. Assoc.* **114**, 1113-1125

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

[27] Robert C.P. (2016) Approximate Bayesian Computation: A Survey on Recent Results. In: Cools R., Nuyens D. (eds) Monte Carlo and Quasi-Monte Carlo Methods. Springer Proceedings in Mathematics & Statistics, vol 163. Springer, Cham

[28] Rubin, D. B. (2019) Conditional Calibration and the Sage Statistician. *Survey Methodology* **45**, 187-198.

[29] Rubin, D. B. (1984) Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician. *Ann. Statist.* **12**, pp. 213-244.

[30] Schwartz, L. (1954) Sur l'impossibilité de la multiplication des distributions. *C. R. Acad. Sci. Paris* **239**, 47-48.

[31] Schwartz, L. (1951) Théorie des distributions. **1-2**, Hermann.

[32] Tanaka, M.M., Francis, A. R., Luciani, F. and Sisson, S. A. (2006) Using Approximate Bayesian Computation to Estimate Tuberculosis Transmission Parameters From Genotype Data. *Genetics*, **173**, 1511–1520.

[33] Tavaré, S. (2019). An introduction to Approximate Bayesian Computation. *Summer Program, Herbert and Florence Irving Institute for Cancer Dynamics.* https://cancerdynamics.columbia.edu/content/summer-program

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

[35] Tukey, J. W. (1977) Modern techniques in data analysis. NSF-sponsored regional research conference at Southeastern Massachusetts University, North Dartmouth, MA.

[36] Vapnik, V. N. and Cervonenkis, A. Y. (1971) On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Appl.*, **16**, 264-280.

[37] Vihola, M. and Franks, J. (2020) On the use of approximate Bayesian computation Markov chain Monte Carlo with inflated tolerance and post correction. *Biometrika* https://doi.org/10.1093/biomet/asz078

[38] Walker, S. G., Damien, P., Laud, P. W. and Smith, A. F. M. (1999) Bayesian nonparametric inference for random distributions and related functions. *J. R. Statist. Soc. B* **61**, 3, 485-527.

[39] Weed, J. and Bach, F. (2017) Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *arXiv:17007.00087*

[40] Wolfowitz, J. (1954) Generalization of the theorem of Glivenko-Cantelli. *Ann. Math. Statist.* **25** , 131-138.

[41] Yatracos, Y. G. (2021) Fiducial Matching for the Approximate Posterior: F-ABC. DOI: 10.13140/RG.2.2.20775.06568

[42] Yatracos, Y. G. (2020) Matching Estimation for Data Generating Experiments. DOI: 10.13140/RG.2.2.30964.58245

[43] Yatracos, Y. G. (1988) A note on L_1 consistent estimation. *Canadian J. Statist.* **16**, **3**, 283-292.

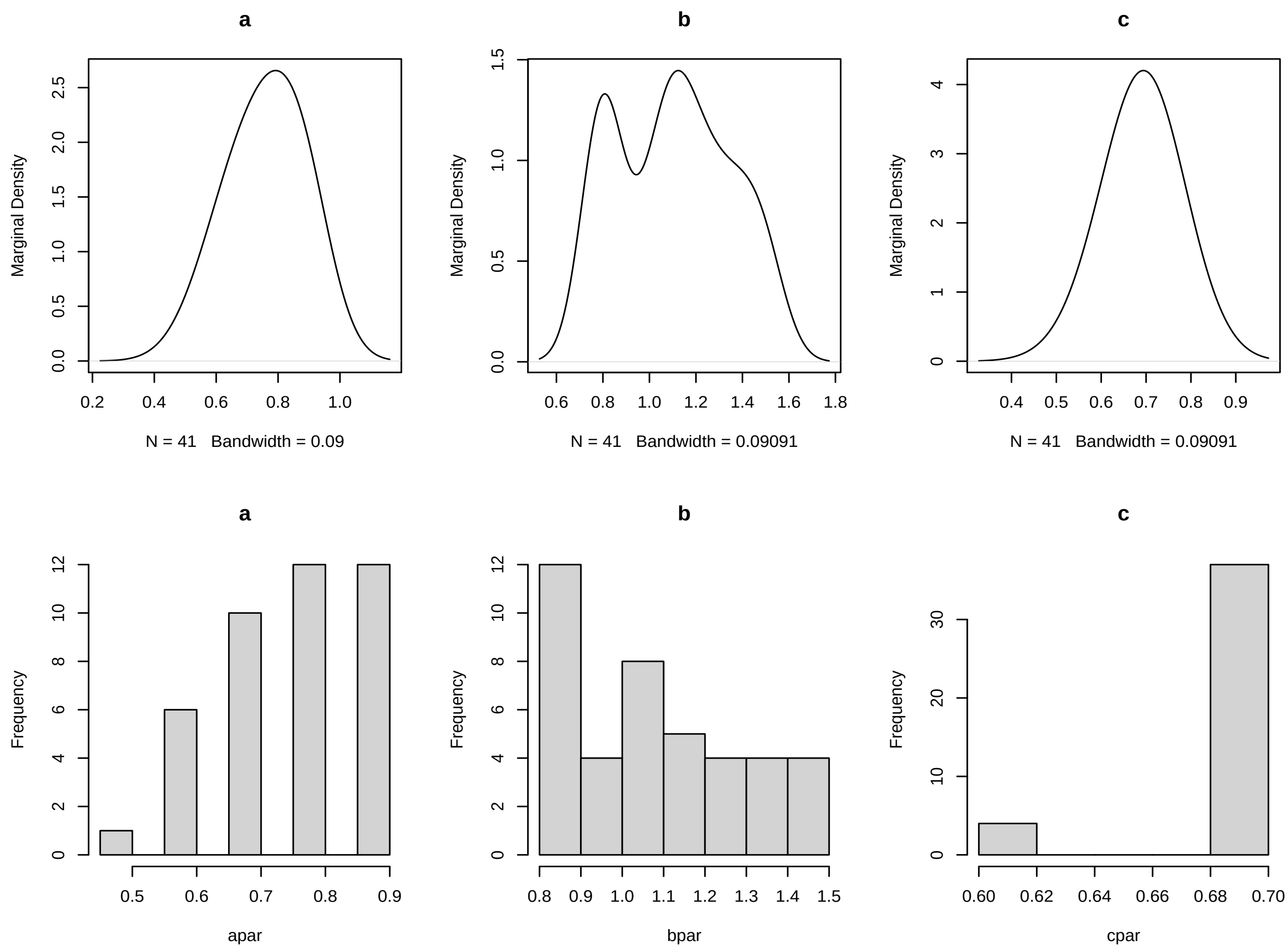

Figure 1: Artifacts in smooth histograms

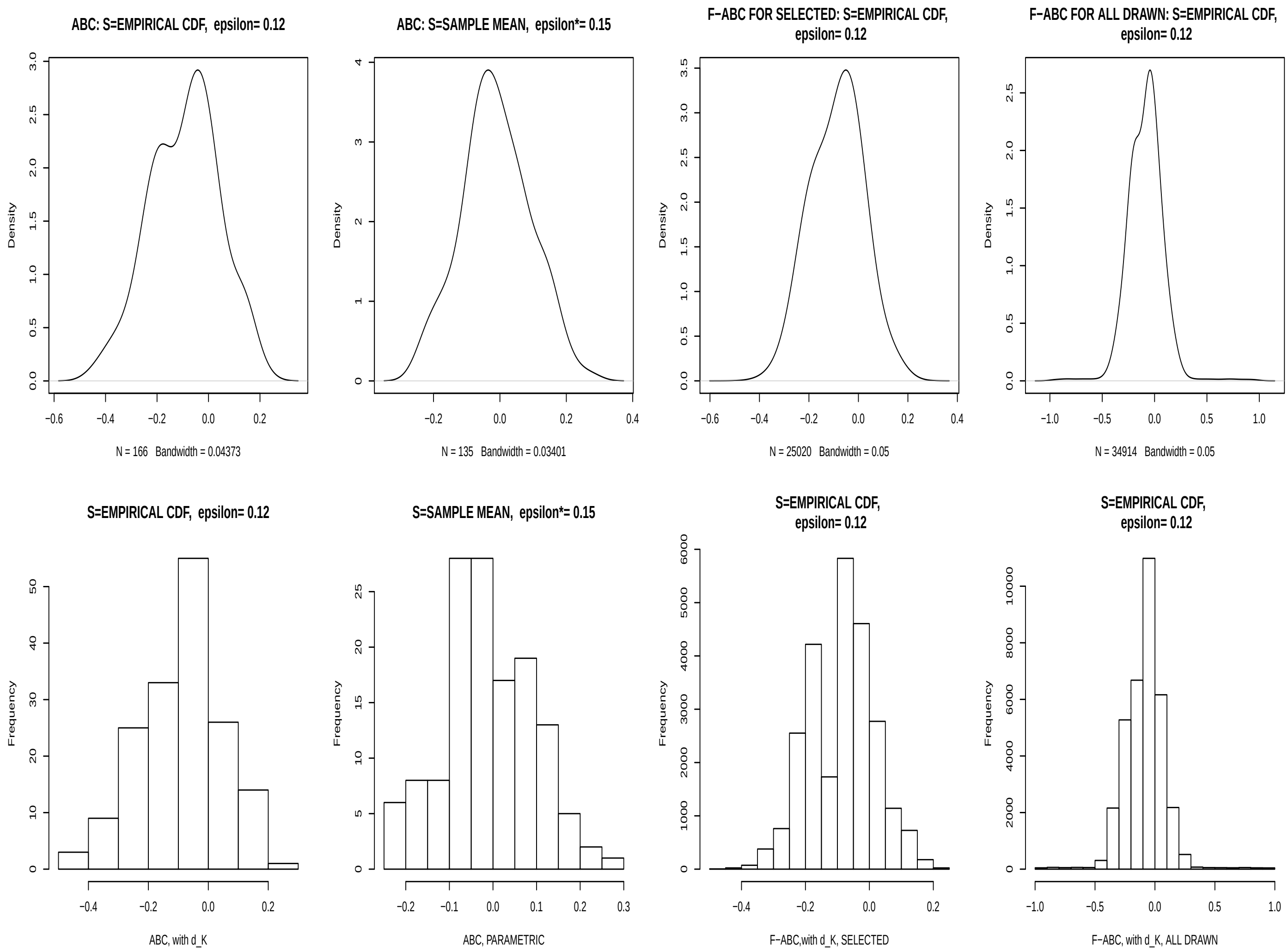

Figure 2: ABC and F-ABC posterior densities and histograms for the mean θ of a normal with variance 1, and the unknown mean $\theta = 0$, #1. Observe in the last histogram the higher concentration of FABC for all around $\theta = 0$

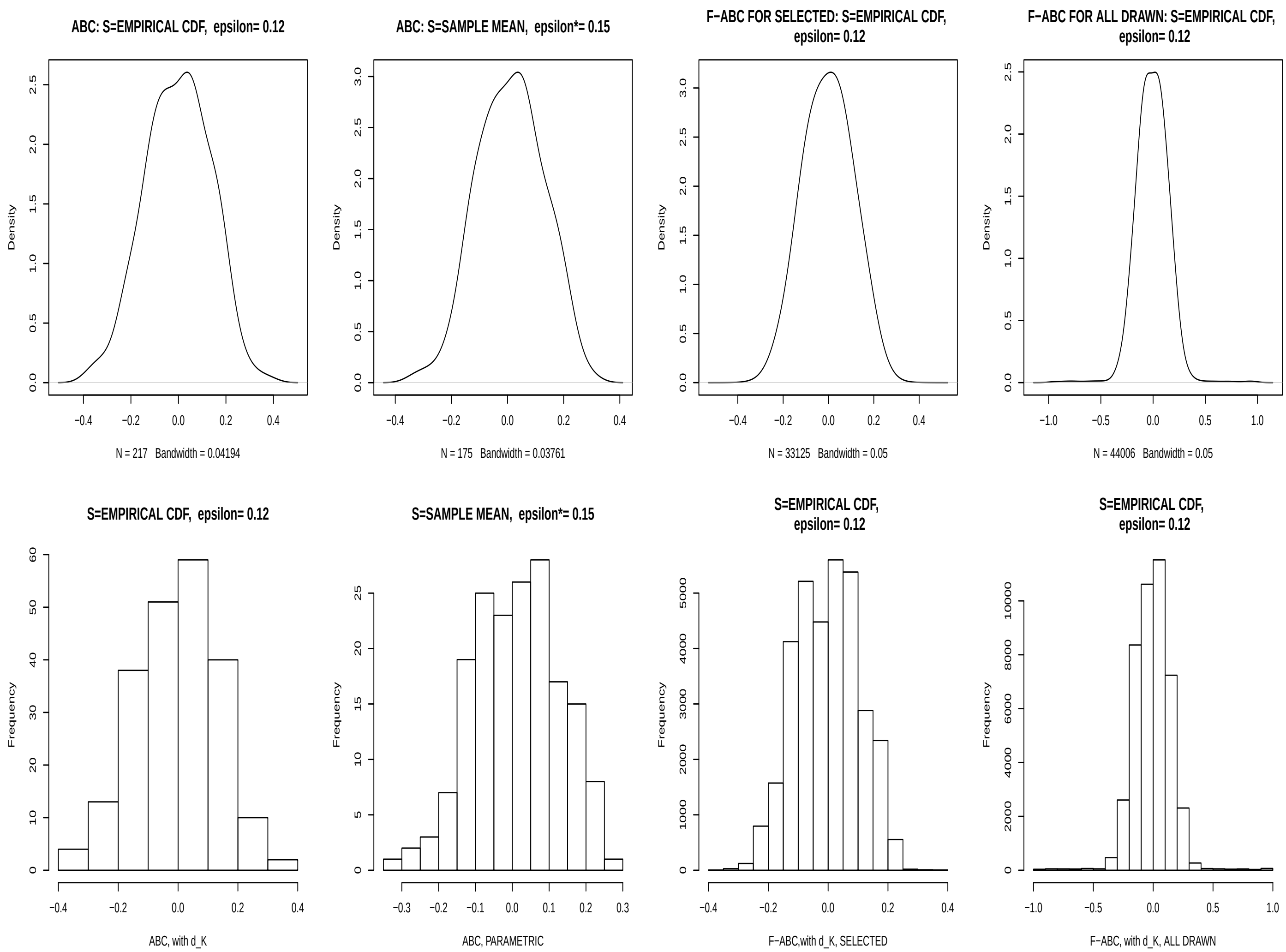

Figure 3: ABC and F-ABC posterior densities and histograms for the mean θ of a normal with variance 1, and the unknown mean $\theta = 0$, #2. Observe in the last histogram the higher concentration of FABC for all around $\theta = 0$.

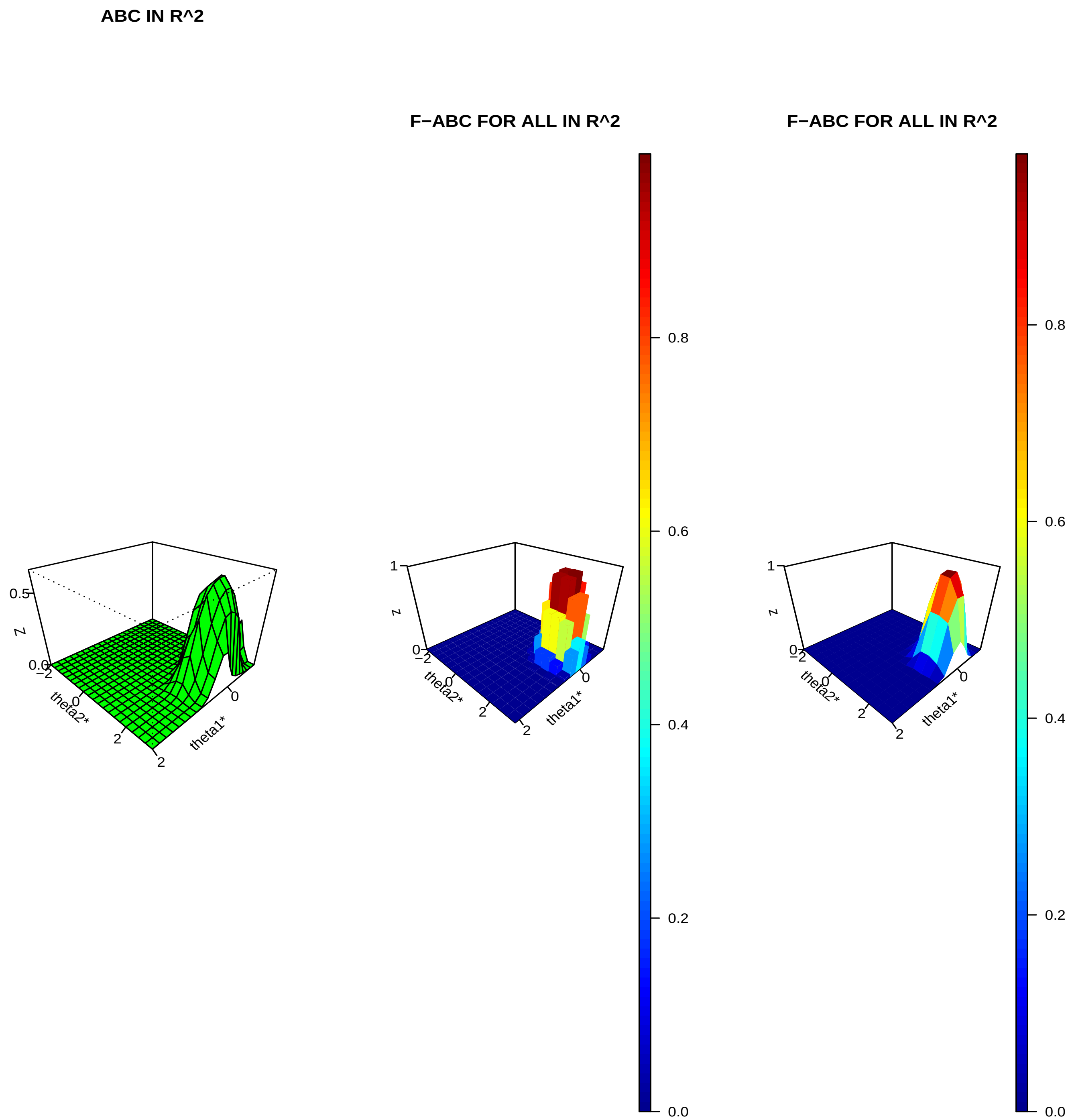

Figure 4: ABC with d_K and F-ABC for all θ^* drawn in a Bivariate normal with dependent components and unknown means $\theta_1 = 0, \theta_2 = 2$.

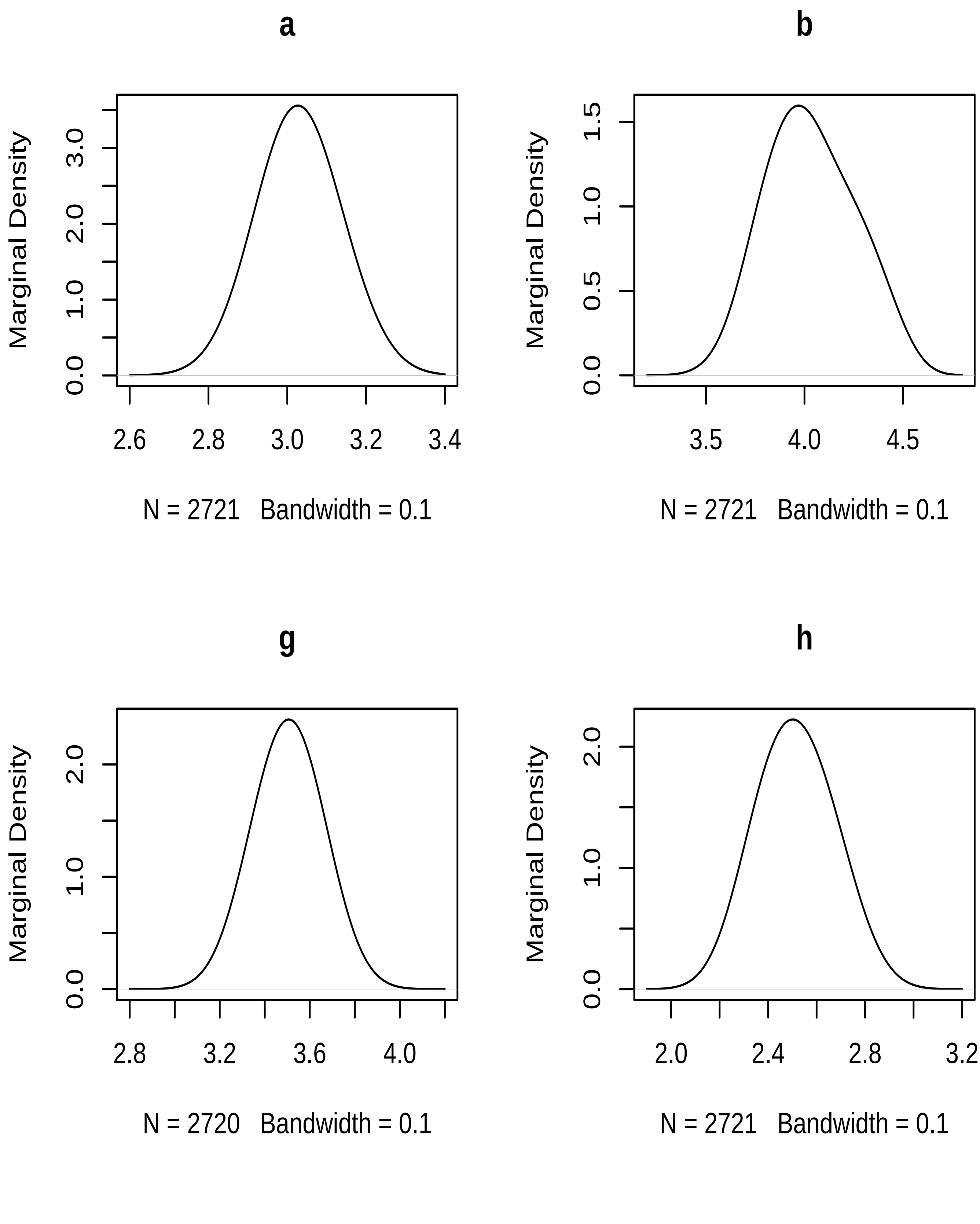

Figure 5: F-ABC for all θ^* drawn in Tukey's (a,b,g,h)-model with parameters $a = 3$, $b = 4, g = 3.5, h = 2.5$.

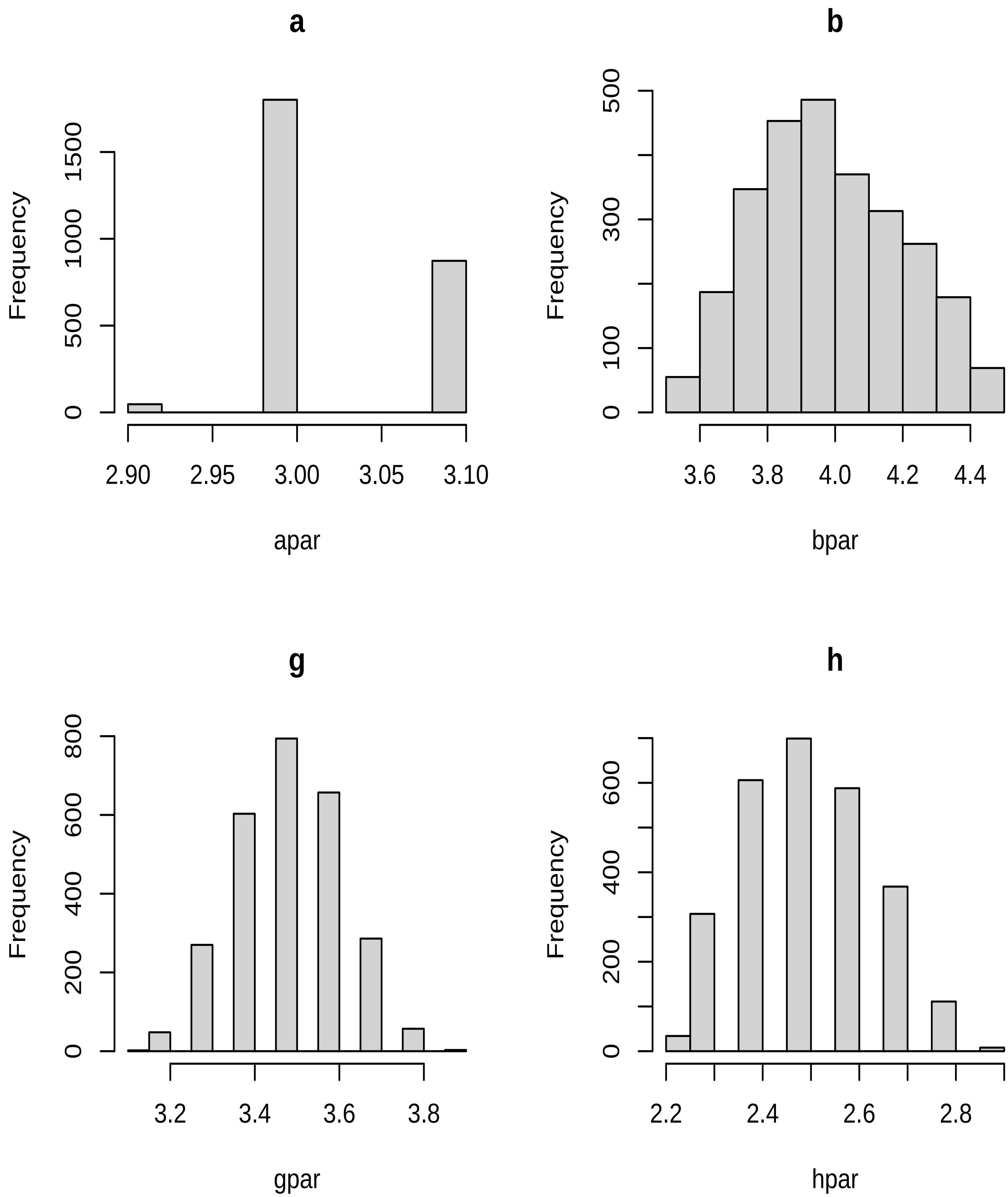

Figure 6: F-ABC for all θ^* drawn in Tukey's (a,b,g,h)-model with parameters $a = 3$, $b = 4, g = 3.5, h = 2.5$.

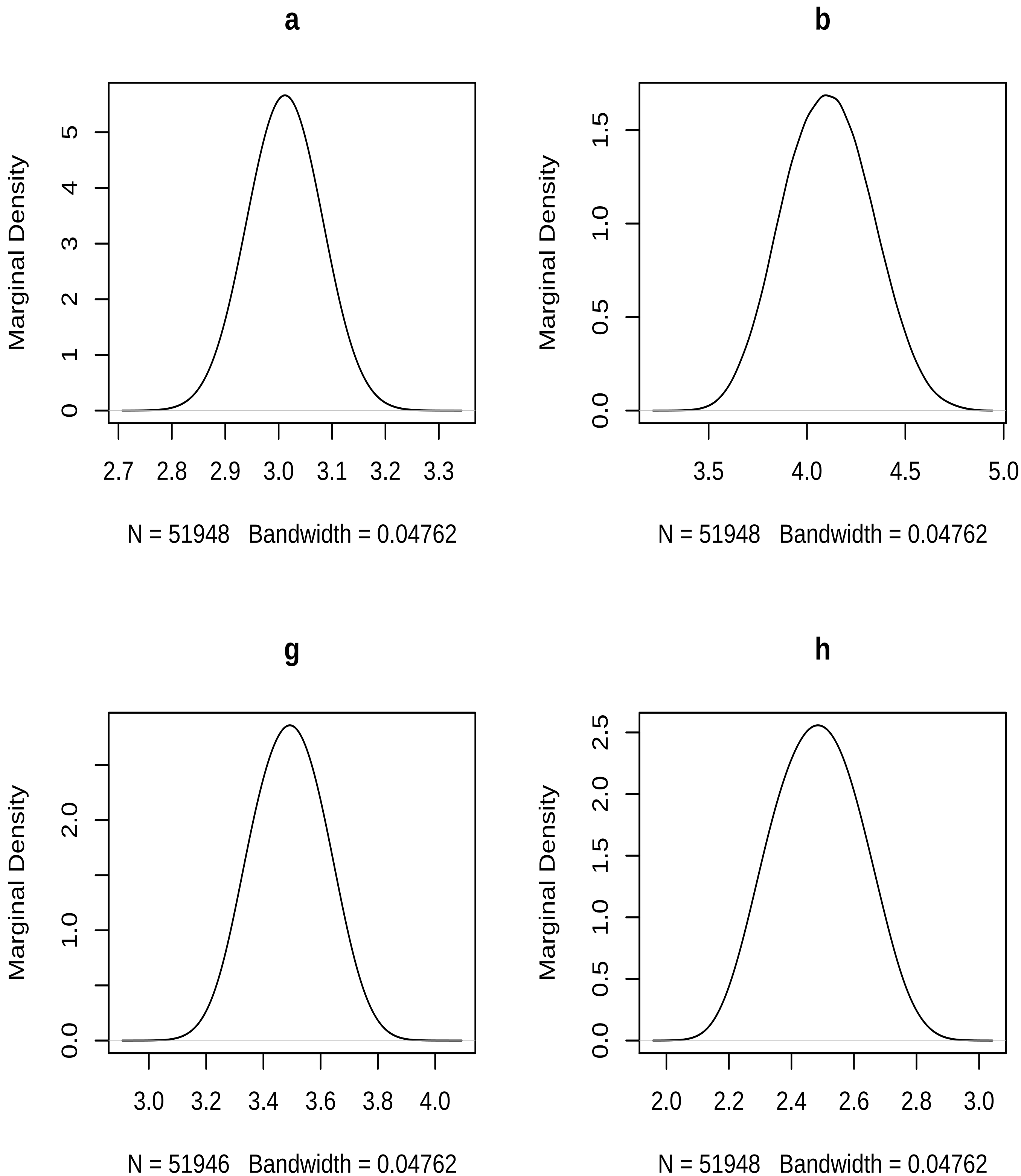

Figure 7: F-ABC for all θ^* drawn in Tukey's (a,b,g,h)-model with parameters $a = 3$, $b = 4, g = 3.5, h = 2.5$. Finer dicretization Θ_b^* with enlarged Θ_b.

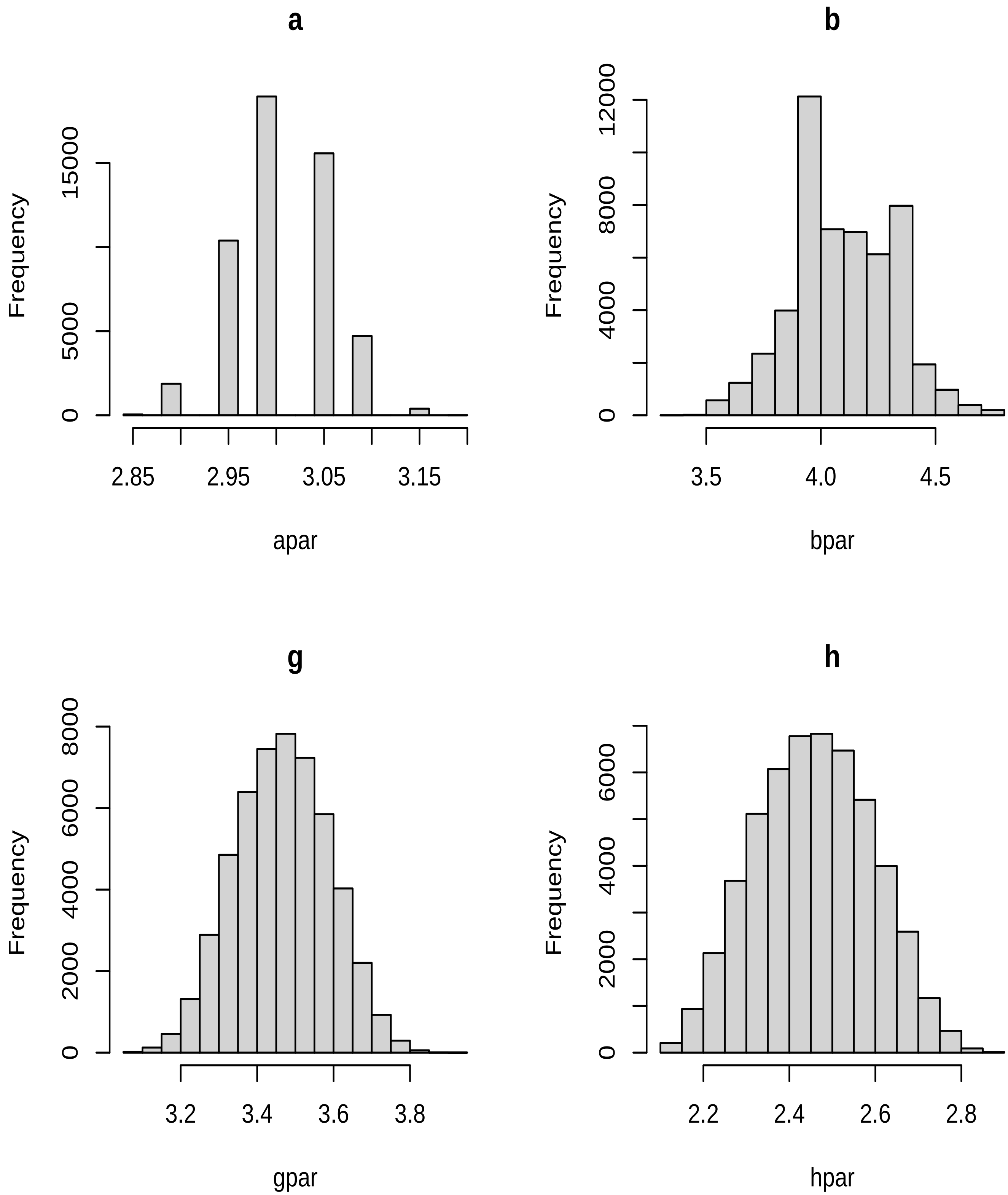

Figure 8: F-ABC for all θ^* drawn in Tukey's (a,b,g,h)-model with parameters $a = 3$, $b = 4, g = 3.5, h = 2.5$. Finer dicretization Θ^* with enlarged Θ_b.

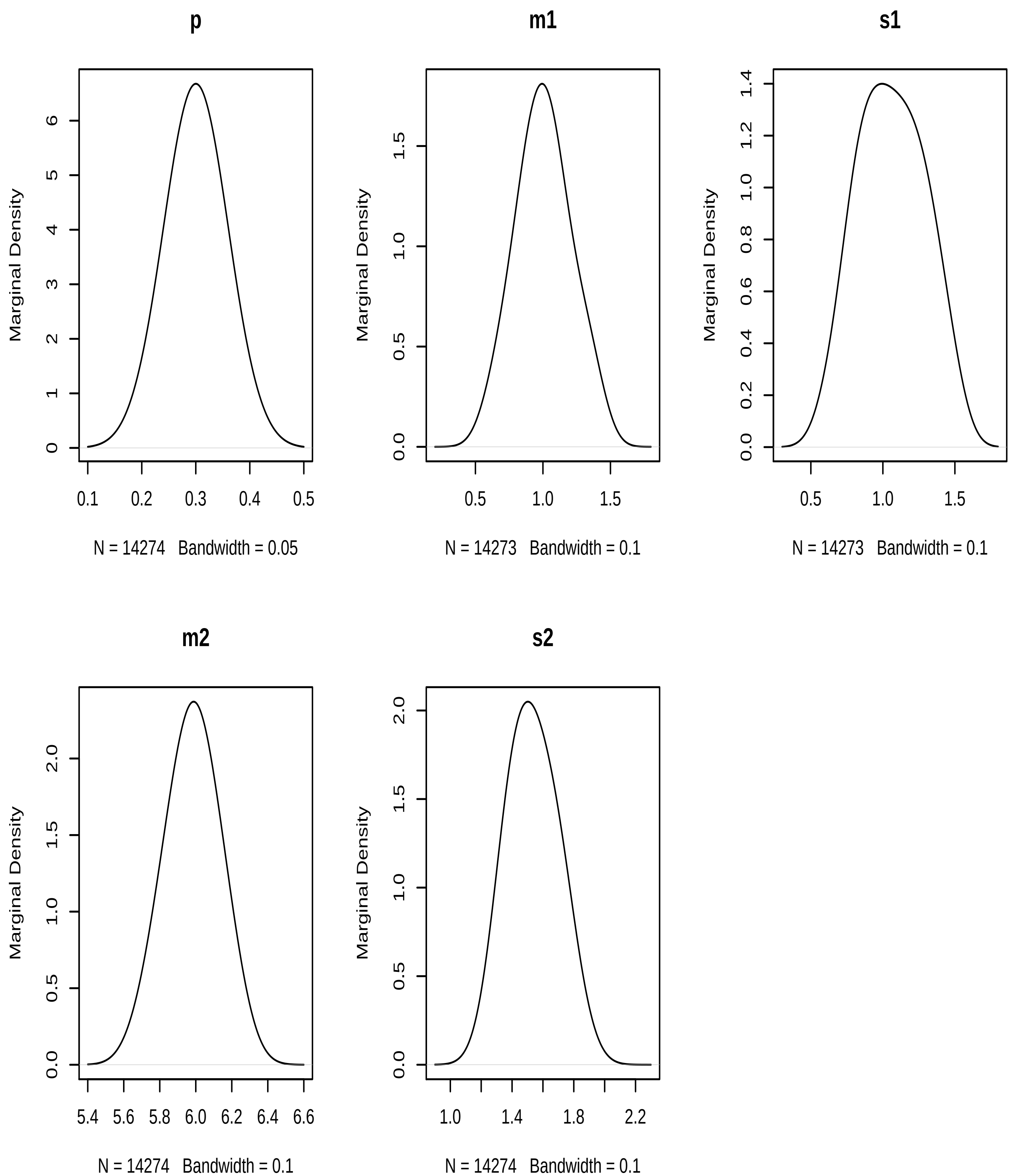

Figure 9: F-ABC for all θ^* drawn in a Normal mixture with parameters $p = .3$, $\mu_1 = m1 = 1, \sigma_1 = s1 = 1, \mu_2 = m2 = 6, \sigma_2 = s2 = 1.5$.

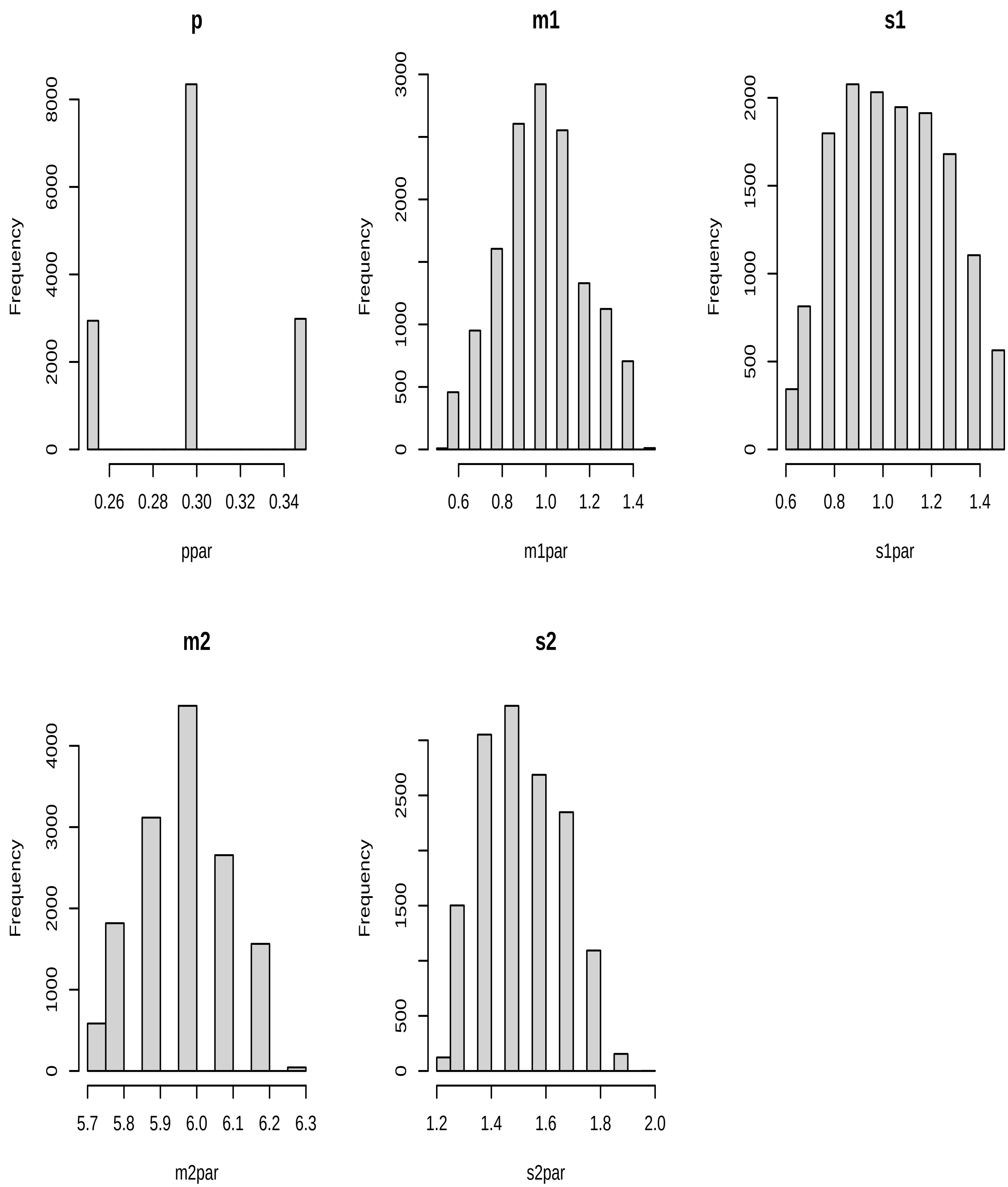

Figure 10: F-ABC for all θ^* drawn in a Normal mixture with parameters, $p = .3$, $\mu_1 = m1 = 1, \sigma_1 = s1 = 1, \mu_2 = m2 = 6, \sigma_2 = s2 = 1.5$.

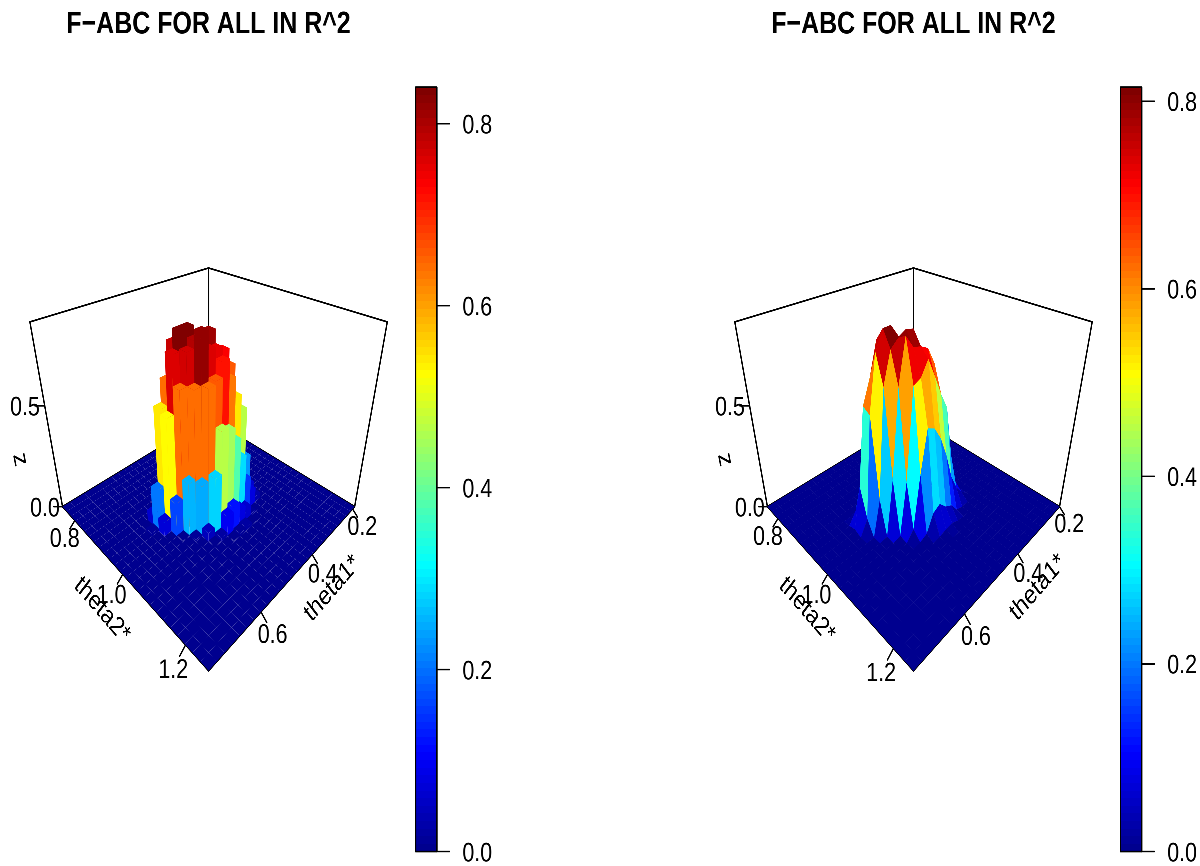

Figure 11: F-ABC for all θ^* drawn in a Time series AR(1) model, θ_1=a=.5, θ_2=b=1

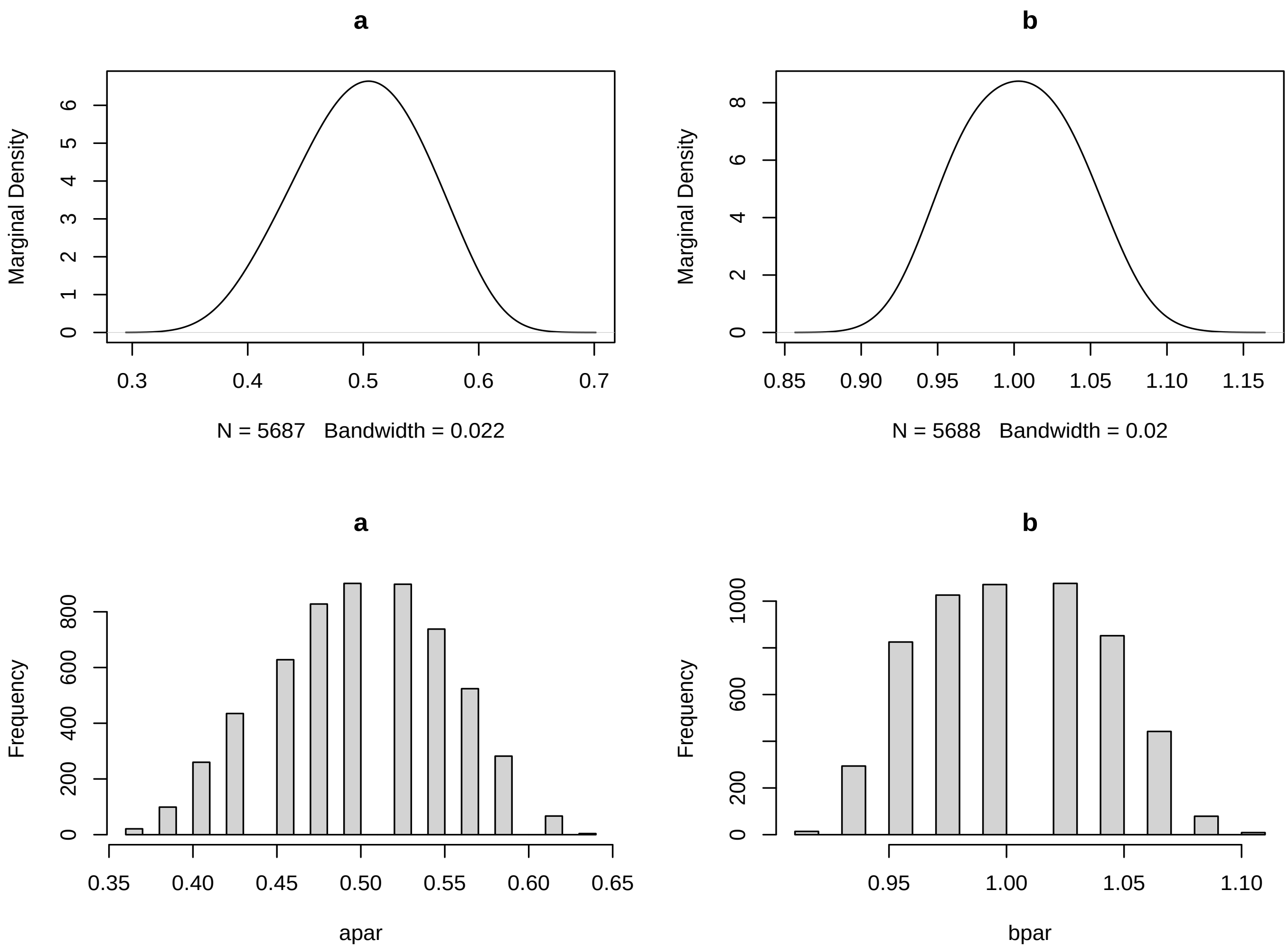

Figure 12: F-ABC for all θ^* drawn in a Time series AR(1) model, a=.5, b=1

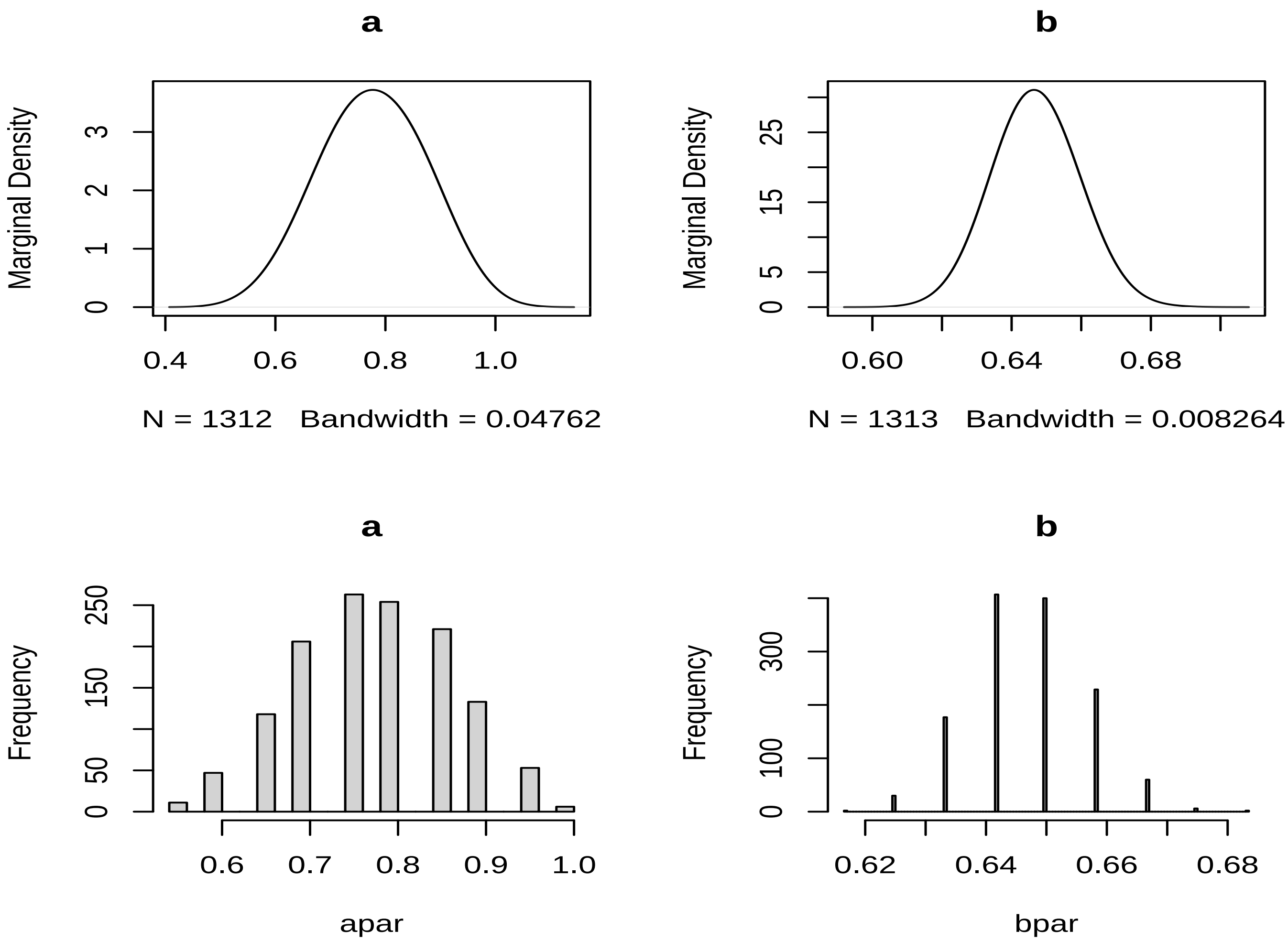

Figure 13: F-ABC for all θ^* drawn in a Quantile model, a=.8, b=.65