

# Approximate Bayesian Computation by Modelling Summary Statistics in a Quasi-likelihood Framework

Stefano Cabras<sup>\*</sup>, Maria Eugenia Castellanos Nueda<sup>†</sup>, and Erlis Ruli<sup>‡</sup>

**Abstract.** Approximate Bayesian Computation (ABC) is a useful class of methods for Bayesian inference when the likelihood function is computationally intractable. In practice, the basic ABC algorithm may be inefficient in the presence of discrepancy between prior and posterior. Therefore, more elaborate methods, such as ABC with the Markov chain Monte Carlo algorithm (ABC-MCMC), should be used. However, the elaboration of a proposal density for MCMC is a sensitive issue and very difficult in the ABC setting, where the likelihood is intractable. We discuss an automatic proposal distribution useful for ABC-MCMC algorithms. This proposal is inspired by the theory of quasi-likelihood (QL) functions and is obtained by modelling the distribution of the summary statistics as a function of the parameters. Essentially, given a real-valued vector of summary statistics, we reparametrize the model by means of a regression function of the statistics on parameters, obtained by sampling from the original model in a pilot-run simulation study. The QL theory is well established for a scalar parameter, and it is shown that when the conditional variance of the summary statistic is assumed constant, the QL has a closed-form normal density. This idea of constructing proposal distributions is extended to non constant variance and to real-valued parameter vectors. The method is illustrated by several examples and by an application to a real problem in population genetics.

**Keywords:** Estimating function, Likelihood-free methods, Markov chain Monte Carlo, Proposal distribution, Pseudo-likelihood.

## 1 Introduction

Many statistical applications in diverse fields such as biology, genetics and finance often involve stochastic models with analytically or computationally intractable likelihood functions. The rapidly growing literature on Approximate Bayesian Computation (ABC) has led to a set of methods which do not involve direct calculation of the likelihood, leading to Bayesian inference that is approximate in a sense that will be specified later.

ABC methods are becoming popular in genetics (Siegmund et al., 2008; Foll et al., 2008), epidemiology (Blum and Tran, 2010; Tanaka et al., 2006) and in population

---

<sup>\*</sup>Department of Mathematics, Università di Cagliari (Cagliari, Italy), [s.cabras@unica.it](mailto:s.cabras@unica.it) and Department of Statistics, Universidad Carlos III de Madrid (Madrid, Spain), [stefano.cabras@uc3m.es](mailto:stefano.cabras@uc3m.es)

<sup>†</sup>Department of Statistics and O.R., Universidad Rey Juan Carlos (Madrid, Spain), [maria.castellanos@urjc.es](mailto:maria.castellanos@urjc.es)

<sup>‡</sup>Department of Statistics, Università di Padova (Padova, Italy), [ruli@stat.unipd.it](mailto:ruli@stat.unipd.it)

biology (Ratmann et al., 2007; Hamilton et al., 2005; Cornuet et al., 2008) among other areas.

Formally, let  $\mathbf{y} = (y_1, \dots, y_n)$  be a random sample of size  $n$  drawn from a statistical model  $\pi(y | \boldsymbol{\theta})$  indexed by the parameter  $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$ . The likelihood for  $\boldsymbol{\theta}$ , corresponding to  $\pi(y | \boldsymbol{\theta})$  is  $L_N(\boldsymbol{\theta})$ , which is not available in closed expression. For a certain prior  $\pi(\boldsymbol{\theta})$ , the aim is to obtain the posterior distribution  $\pi_N(\boldsymbol{\theta} | \mathbf{y}) \propto L_N(\boldsymbol{\theta})\pi(\boldsymbol{\theta})$ , but as  $L_N(\boldsymbol{\theta})$  is inaccessible,  $\pi_N(\boldsymbol{\theta} | \mathbf{y})$  cannot be approximated by directly evaluating  $L_N(\boldsymbol{\theta})$ .

This difficulty may be overcome by using ABC methods. Specifically, let  $\mathbf{s} = s(y) \in \mathcal{S} \subseteq \mathbb{R}^p$  be a vector of observable summary statistics (e.g. mean, variance, quantiles etc.), which may not be sufficient, let  $\rho(\mathbf{s}, \mathbf{s}_{obs})$  be a metric distance between  $\mathbf{s}$  and its observed value  $\mathbf{s}_{obs}$  with  $\epsilon > 0$  the tolerance parameter. ABC methods approximate  $\pi_N(\boldsymbol{\theta} | \mathbf{y})$  by

$$\pi^\epsilon(\boldsymbol{\theta} | \mathbf{s}_{obs}) = \int_{\mathcal{S}} \pi^\epsilon(\boldsymbol{\theta}, \mathbf{s} | \mathbf{s}_{obs}) d\mathbf{s},$$

where  $\pi^\epsilon(\boldsymbol{\theta}, \mathbf{s} | \mathbf{s}_{obs}) \propto \pi(\boldsymbol{\theta})\pi(\mathbf{s} | \boldsymbol{\theta})\mathbb{I}_{\rho < \epsilon}$ , and  $\mathbb{I}_{\rho < \epsilon}$  is the indicator function for the event  $\{\mathbf{s} \in \mathcal{S} | \rho(\mathbf{s}_{obs}, \mathbf{s}) < \epsilon\}$ . They require a choice of  $\epsilon$  and for this purpose several authors (Bortot et al., 2007; Faisal et al., 2013; Ratmann et al., 2014; Barnes et al., 2012; Aeschbacher et al., 2012) suggest approaches where  $\epsilon$  is estimated as part of an extended model with respect to  $\pi(y | \boldsymbol{\theta})$ , while a recent approach based on diagnostic tools for ABC can be found in Prangle et al. (2013a). In this work another criterion for choosing  $\epsilon$  is discussed.

The basic version of the ABC algorithm relies on simulation by the mixture representation method consisting in generating, say,  $T$  values of  $\boldsymbol{\theta}$  from  $\pi(\boldsymbol{\theta})$  and using them to generate the corresponding  $T$  values of  $\mathbf{s}$  from  $\pi(y | \boldsymbol{\theta})$  at the simulated  $\boldsymbol{\theta}$ . We accept all values of  $\boldsymbol{\theta}$  such that  $\rho(\mathbf{s}, \mathbf{s}_{obs}) < \epsilon$ . For  $\epsilon \rightarrow 0$  the ABC method has been proven to return a consistent estimator of the posterior  $\pi(\boldsymbol{\theta} | \mathbf{s}_{obs})$  and under some assumptions it is also possible to provide the approximation error as shown in Biau et al. (2012). Moreover, if  $\mathbf{s}$  is sufficient and  $\epsilon \rightarrow 0$ , then  $\pi^\epsilon(\boldsymbol{\theta} | \mathbf{s}_{obs}) \rightarrow \pi_N(\boldsymbol{\theta} | \mathbf{y})$ . There is a certain agreement in that low dimensional, but informative summary statistics improve the accuracy of the ABC approximation (Blum et al., 2013).

One drawback of the basic ABC algorithm is that it can be extremely inefficient when the discrepancy between  $\pi(\boldsymbol{\theta})$  and  $\pi_N(\boldsymbol{\theta} | \mathbf{y})$  is relevant. Unfortunately, as  $L_N(\boldsymbol{\theta})$  is intractable, the discrepancy between  $\pi(\boldsymbol{\theta})$  and  $\pi_N(\boldsymbol{\theta} | \mathbf{y})$  is difficult to know a priori and not easy to assess. To deal with this issue, more advanced Monte Carlo methods, such as ABC-MCMC, originally developed in Marjoram et al. (2003) and further analyzed in several papers such as Beaumont et al. (2009a); Andrieu and Roberts (2009); Lee (2012), or Sequential Monte Carlo (SMC) methods (see, e.g., Beaumont et al., 2009b; Sisson et al., 2007) may be used. All these methods attempt to account for the observed data at the proposal stage. However, to accomplish this task, a proposal distribution or a perturbing kernel is required, which in practice is supplied by the analyst.

Another aspect of the method, which is of major concern in the ABC literature, is the choice of  $s$ , which should be informative for  $\theta$ . The same concern applies here, as we require  $s$  not to be ancillary with respect to  $\theta$ . Many suggestions can be found in the current literature. For instance, Fearnhead and Prangle (2012) propose considering the posterior mean of  $\theta$ , i.e.  $E_{\pi^\epsilon(\theta|\mathbf{s}_{obs})}(\theta)$ . The latter is estimated by means of a pilot-run simulation that depends on the specific observed sample. Moreover, Ruli et al. (2013) suggest choosing  $s$  as the score of the composite likelihood function obtained from  $\pi(y|\theta)$ .

The present work focuses on the study of a class of proposal distributions for ABC-MCMC, and a method for building proposal densities, which target the posterior  $\pi^\epsilon(\theta|\mathbf{s}_{obs})$ , is illustrated. Such proposal distributions depend on the model at hand and account for the observed data. These distributions for  $\theta$  are constructed in a way that leads to adopting a normal kernel on the space of  $s$ , and then consider a reparametrization from  $s$  to  $\theta$  by a suitable regression function  $f(\theta) = E_{\pi(y|\theta)}(s|\theta)$ . A recent approach, strongly connected with the use of such a regression function, can be found in Ratmann et al. (2014), where the  $f(\theta)$  is the binding function in indirect inference (Gourieroux et al., 1993).

We show that for scalar parameter problems and  $f(\theta)$  such proposal distributions arise from the class of quasi-likelihood functions (QL) of  $\theta$  (McCullagh, 1991) denoted by  $L_Q(\theta)$ . For multidimensional parameter problems, the QL is not tractable, but the idea can still be generalized to these contexts using asymptotic arguments. Indeed, for the vector of parameters  $\theta$ , we consider a multivariate normal kernel and a multivariate transformation from  $\mathbf{s}$  to  $\theta$ . For both the scalar and the multi parameter cases, these transformations are typically not available analytically and we estimate them in a pilot-run simulation. This pilot-run simulation is performed regardless of the specific observed sample and thus it can serve for routine analysis. This is an appealing feature of the proposed method as will be shown later by an application to a Genome Wide Association Study (GWAS).

Despite the fact that under some more elaborate requirements for the proposal distribution, later discussed, we end up in a proposal which is not the QL for  $\theta$ , we think that the connection of ABC and QL is important because  $L_N(\theta)$  is not available and the estimation theory of the QLs guarantees, asymptotically, that  $L_Q(\theta)$  targets  $L_N(\theta)$ .

The structure of the paper is as follows: Section 2 discusses  $L_Q(\theta)$ , which inspires the proposal distributions considered throughout the paper. These distributions will be embedded in the ABC-MCMC algorithm. The proposed ABC algorithm,  $ABC_{qt}$ , for scalar parameters is formally discussed in Section 3, while the generalization to  $p > 1$  is presented in Section 4. Section 5 illustrates the proposed method with some examples from the ABC literature and an important application to GWAS for population genetic isolates. Conclusions and further remarks are given in Section 6.

## 2 The two relevant tools: quasi-likelihood and the ABC-MCMC algorithm

The theory and use of estimating equations and that of the related quasi- and quasi-profile likelihood functions have received a good deal of attention in recent years; see, among others, Liang and Zeger (1995); Barndorff-Nielsen (1995); Desmond (1997); Heyde (1997); Adimari and Ventura (2002); Severini (2002); Wang and Hanfelt (2003); Jørgensen and Knudsen (2004); Bellio et al. (2008). In addition, Ventura et al. (2010); Lin (2006); Greco et al. (2008) discuss the use of QL functions in the Bayesian setting.

Let  $s = s(y) \in \mathbb{R}$  be a scalar summary statistic generated from  $\pi(s(y) | \theta)$  whose observed value is  $s_{obs}$ . We assume for convenience that the summary statistic lies on the real line, which can be easily achieved by suitable transformations (e.g. a log-transformation of the sample variance).

Moreover, suppose  $\theta$  is a scalar parameter, i.e.  $p = 1$  and let  $\Psi(s; \theta)$  be an unbiased estimating function of  $\theta$  based on  $s$ , i.e.  $E_{\pi(y|\theta)}\{\Psi(S; \theta)\} = 0$ .

The QL for  $\theta$  based on  $\Psi(s; \theta)$  (McCullagh, 1991), is given by

$$L_Q(\theta) = \exp \left\{ \int_{c_0}^{\theta} A(t) \Psi(s; t) dt \right\}, \quad (1)$$

where  $A(\theta) = M(\theta)/\Omega(\theta)$ ,  $c_0$  is an arbitrary constant,

$$M(\theta) = -E \left\{ \frac{\partial \Psi(S; \theta)}{\partial \theta} \mid \theta \right\},$$

and

$$\Omega(\theta) = E\{\Psi(S; \theta)^2 \mid \theta\} = \text{Var}\{\Psi(S; \theta) \mid \theta\}.$$

When  $p = 1$ , a quasi likelihood for  $\theta$  is usually easy to derive, while for  $p > 1$  some difficulties arise. Moreover, as shown below, for a suitable estimating function and under  $\text{Var}\{\Psi(S; \theta) \mid \theta\}$  constant, (1) is a normal kernel.

The ABC-MCMC algorithm, proposed in Marjoram et al. (2003), summarized in Algorithm 1, evaluates  $L_N(\theta)$  indirectly via the indicator function  $\mathbb{I}_{\rho < \epsilon}$ , and uses the proposal density  $q(\theta^{(t)} \mid \theta^{(t-1)})$ .

Depending on how the proposal is defined, with Algorithm 1 we may implement the independent Metropolis Hastings (MH) or the Random Walk (RW) MH. SMC methods may also be considered as in Toni et al. (2009). Finally, the proposal  $q(\cdot)$  can also be viewed as an importance function for the implementation of an Importance Sampling (IS) simulation algorithm.

The proof that  $\pi^\epsilon(\theta \mid s_{obs})$  is the stationary distribution of Algorithm 1 is contained in Theorem 1 of Marjoram et al. (2003) and the rate of convergence depends on the choice of  $q(\cdot)$  and  $\epsilon$ . However, as  $L_N(\theta)$  is not tractable, it is not possible to further characterize the stochastic behavior of the induced chain. In fact, the theoretical conditions discussed in Mengersen and Tweedie (1996) and Atchadé and Perron (2007) may be

**Algorithm 1** The ABC-MCMC algorithm

- 
- 1: Set  $\epsilon > 0$ ,  $\theta^{(0)} = \theta_{init} \in \Theta$ ;
  - 2: **for**  $t = 1$  to  $T$  **do**
  - 3:   generate  $\theta^* \sim q(\theta^{(t)} | \theta^{(t-1)})$ ;
  - 4:   generate  $s \sim \pi(s(y) | \theta^*)$ ;
  - 5:   calculate  $\rho = \rho(s_{obs}, s)$ ;
  - 6:   with probability

$$\min \left\{ 1, \frac{\pi(\theta^*)q(\theta^{(t-1)} | \theta^*)}{\pi(\theta^{(t-1)})q(\theta^* | \theta^{(t-1)})} \mathbb{I}_{\rho < \epsilon} \right\}$$

accept  $\theta^*$  and set  $\theta^{(t)} = \theta^*$ , otherwise  $\theta^{(t)} = \theta^{(t-1)}$

- 7: **end for**
  - 8: **return**  $\theta^{(1)}, \dots, \theta^{(T)}$
- 

assessed if  $L_N(\theta)$  is available in a closed form expression. Notice that the approach can also be viewed as a pseudo marginal MH as we are working with an estimated likelihood when evaluating the indicator function  $\mathbb{I}_{\rho < \epsilon}$ , and results for convergence of such pseudo marginal algorithms can also be found in Andrieu and Roberts (2009).

### 3 The ABC<sub>ql</sub> for a scalar parameter

The main objective of this paper is to construct a proposal density centered on the bulk of the posterior distribution. In a setting where the likelihood cannot be computed explicitly, this issue could be cumbersome. For this purpose, we consider a QL derived from estimating functions based on  $s$ .

The following Proposition 1 provides the expression of  $L_Q(\theta)$  for a general statistic  $s$  and assuming  $\text{Var}_{\pi(y|\theta)}(S | \theta)$  is constant.

**Proposition 1.** Suppose  $p = 1$  and let  $f(\theta) = E_{\pi(y|\theta)}(S | \theta)$  be a bounded regression function under the sampling model  $\pi(y | \theta)$  for which the Jacobian  $|f'(\theta)| < \infty$  and that the conditional variance  $\text{Var}_{\pi(y|\theta)}(S | \theta) = \sigma_R^2$  is constant with respect to  $\theta$ .

Consider the following estimating function  $\Psi(s_{obs}; \theta) = s_{obs} - f(\theta)$ . In this case we have

$$L_Q(\theta) = \phi \left( \frac{f(\theta) - s_{obs}}{\sigma_R} \right), \quad (2)$$

where  $\phi(\cdot)$  is the density of the standard normal distribution.

**Proof.** Note that the estimating function is unbiased because  $E\{\Psi(s_{obs}; \theta) | \theta\} = E(S | \theta) - f(\theta) = 0$ . From the definition of  $L_Q(\theta)$  the following quantities are needed:

$$M(\theta) = -E \left( \frac{\partial \Psi}{\partial \theta} | \theta \right) = f'(\theta),$$

$$\begin{aligned}
\Omega(\theta) &= E\{\Psi(s_{obs}; \theta)^2 \mid \theta\} \\
&= \text{Var}(S - f(\theta) \mid \theta) \\
&= \text{Var}(S \mid \theta) \\
&= \sigma_R^2.
\end{aligned}$$

Then  $A(\theta) = f'(\theta)/\sigma_R^2$  and by (1) we have

$$\begin{aligned}
L_Q(\theta) &= \exp \left\{ \int_{c_0}^{\theta} \frac{f'(t)}{\sigma_R^2} (s_{obs} - f(t)) dt \right\} \\
&\propto \frac{1}{\sigma_R} \exp \left( -\frac{(f(\theta) - s_{obs})^2}{2\sigma_R^2} \right),
\end{aligned}$$

which is the kernel of the normal distribution centered at  $s_{obs}$  with variance  $\sigma_R^2$ .  $\square$

Expression (2) suggests that  $L_Q(\theta)$  is a normal density when working in the  $f(\theta)$  parametrization, and thus if one is able to make a change-of-variable from  $f(\theta)$  to  $\theta$ , it could be employed as a proposal for MCMC-ABC algorithms, leading to a broad class of ABC methods denoted  $\text{ABC}_{qt}$  algorithms.

Note that the constant variance assumption,  $\text{Var}_{\pi(y|\theta)}(S \mid \theta) = \sigma_R^2$ , leads to a closed-form proposal distribution, a normal density. However, as this assumption may be restrictive, we extend the idea of constructing proposal distributions based on  $L_Q(\theta)$ , but assuming a non constant variance,  $\sigma_R^2(\theta)$ , which can be estimated as well as  $f(\theta)$  (see Subsection 3.1). The theory of QL assures that also for non constant  $\sigma_R^2(\theta)$  there exists a corresponding  $L_Q(\theta)$  whose form is intractable and for this reason it cannot be used directly as a proposal density. Instead, our proposal distribution for a Random Walk Metropolis Hastings (RWMH) is based on a distribution of the form of  $L_Q(\theta)$  in (2) where  $\sigma_R^2(\theta)$  is non constant:

$$q^Q(\theta \mid \theta^{(t-1)}) = \phi \left( \frac{f(\theta) - f(\theta^{(t-1)})}{\sigma_R(\theta^{(t-1)})} \right) \mid f'(\theta) \mid. \quad (3)$$

Finally, this way of constructing proposals could also be extended to other types of distributions with heavy tails, such as the  $t$ -Student distribution.

On the other hand, ABC Importance Sampling (ABC-IS) can be implemented using  $q(\theta) = L_Q(\theta) \mid f'(\theta) \mid$  as the importance function from which it is possible to simulate. Assuming  $\sigma_R(\theta) = \sigma_R$ , it would be enough to simulate a sample  $z$  from a standard normal and then calculate  $f^{-1}(z\sigma_R + s_{obs})$  to have a draw of  $\theta$ . The ABC-IS is completed by the evaluation of the importance weights by means of  $q(\theta)$ . Also ABC-SMC can be used starting with a sample from  $q(\theta)$  and using steps S1-S3 from Toni et al. (2009). The computational requirements are almost the same as for the ABC-IS. In fact, for ABC-SMC it is necessary to calculate importance weights to be updated in the MC sequence.

---

**Algorithm 2** The ABC<sub>ql</sub> for  $p = 1$ 


---

**Require:**  $f$ ,  $f'(\theta)$ ,  $\sigma_R^2(\theta)$ , or their estimates  $(\widehat{f}, \widehat{f}'(\theta), \widehat{\sigma}_R^2(\theta))$ .
1: Set  $\epsilon > 0$  and  $\theta^{(0)} = f^{-1}(s_{obs})$ ;2: **for**  $t = 1$  to  $T$  **do**

3: generate

$$f^* \sim N(f(\theta^{(t-1)}), \sigma_R^2(\theta^{(t-1)}));$$

4: set  $\theta^* = \{\theta : f^{-1}(f^*) = \theta\}$ ;5: generate  $s \sim \pi(s(y) | \theta^*)$ ;6: calculate  $\rho = \rho(s_{obs}, s)$ ;7: calculate the derivative,  $f'(\theta)$ , of  $f(\theta)$ , at  $\theta^{(t-1)}$  and  $\theta^*$ ;

8: with probability

$$\min \left\{ 1, \frac{\pi(\theta^*) q^Q(\theta^{(t-1)} | \theta^*)}{\pi(\theta^{(t-1)}) q^Q(\theta^* | \theta^{(t-1)})} \mathbb{I}_{\rho < \epsilon} \right\}$$

accept  $\theta^*$  and set  $\theta^{(t)} = \theta^*$ , otherwise  $\theta^{(t)} = \theta^{(t-1)}$ 9: **end for**10: **return**  $\theta^{(1)}, \dots, \theta^{(T)}$ 


---

### 3.1 Estimation of $f(\theta)$ , $f'(\theta)$ and $\sigma_R^2(\theta)$

The function  $f(\theta)$  can be elicited, suggested by the model in Mengersen et al. (2013); Ratmann et al. (2007) or by theoretical arguments as in Heggland and Frigessi (2004). For instance, in the genetic model analyzed in Mengersen et al. (2013), where the constraint of the empirical likelihood plays the same role as  $\Psi(s_{obs}; \theta)$ ,  $f(\theta)$  is built upon the score function of the pairwise likelihood corresponding to the model. However, except in few specific situations,  $f(\theta)$ ,  $f'(\theta)$  and  $\sigma_R^2(\theta)$  are generally unknown, and we replace them in Algorithm 2 by estimates that can be obtained in a pilot-run simulation as stated in Algorithm 3. In the sequel, when referring to the ABC<sub>ql</sub> algorithms, our intention is always that  $f(\theta)$ ,  $f'(\theta)$  and  $\sigma_R^2(\theta)$  are unknown and replaced by an estimator with the sole purpose of providing the input as a proposal density for Algorithm 2 (or Algorithm 5). If  $f(\theta)$ ,  $f'(\theta)$  and  $\sigma_R^2(\theta)$  were known, the computational effort for the ABC<sub>ql</sub> algorithm would be reduced. In any case, the proof of the convergence of Algorithm 2 (or Algorithm 5) is discussed in the previous Section 2.

The functions  $\widehat{f}(\theta)$  or  $\widehat{\sigma}_R^2(\theta)$  can be any estimator which provides smoothing regression functions, and  $\widehat{f}(\theta)$  is at least once differentiable. This implies that the main assumption for  $f(\theta)$  is to be monotone and once differentiable. We find it useful to use smoothing splines, for which the derivative,  $\widehat{f}'(\theta)$ , can be obtained analytically from splines coefficients. Other choices are possible and left to the convenience of the analyst that inspects the scatter diagram of points  $\left\{ \tilde{s}_m, \tilde{\theta}_m \right\}_1^M$  and provides a goodness-of-fit argument that justifies the choice. The inverse  $\widehat{f}^{-1}(f^*)$ , at some point  $f^*$ , can be obtained either analytically, with the bisection method on  $\widehat{f}(\theta) = f^*$  or by numerical

---

**Algorithm 3** Estimation of  $f(\theta)$ ,  $f'(\theta)$  and  $\sigma_R^2(\theta)$  for  $p = 1$

---

**Require:**  $M, \tilde{\Theta}$

- 1: consider  $M$  values  $\tilde{\theta} = (\tilde{\theta}_1, \dots, \tilde{\theta}_M)$  taken in a regular spaced grid of a suitable large subset  $\tilde{\Theta} \subseteq \Theta$ ;
  - 2: generate  $\tilde{s} = (\tilde{s}_1, \dots, \tilde{s}_M)$  where  $\tilde{s}_m \sim \pi(s(y) | \tilde{\theta}_m)$ ;
  - 3: regress  $\tilde{s}$  on  $\tilde{\theta}$  obtaining  $\hat{f}(\theta)$  and  $\hat{f}'(\theta)$ ;
  - 4: regress  $\left\{ \log \left( \hat{f}(\tilde{\theta}_m) - \tilde{s}_m \right)^2 \right\}_{m=1, \dots, M}$  on  $\tilde{\theta}$  obtaining  $\hat{\sigma}_R^2(\theta)$ .
  - 5: **return**  $\hat{f}(\theta)$ ,  $\hat{f}'(\theta)$  and  $\hat{\sigma}_R^2(\theta)$ .
- 

minimization of  $(\hat{f}(\theta) - f^*)^2$ , e.g., by a Newton-Raphson algorithm.

Some observations are appropriate.

- i*) Since we are able to simulate from  $\pi(y | \theta)$ , then  $\hat{f}(\theta)$  and  $\hat{\sigma}_R^2(\theta)$  can be practically estimated with a precision that depends on the available computational resources. Also more precision can be achieved by making the regular grid  $\tilde{\Theta}$  wider or by increasing the number of simulations,  $M$ . Values of  $M$  ranging from 100 to 1000 are enough in the examples discussed later for  $p = 1$ , while larger values are needed for  $p > 1$ , due to the curse of dimensionality as the computational effort increases exponentially with  $p$ .
- ii*) The range of  $\tilde{\Theta}$  should always be large enough to include the observed  $s_{obs}$  in order to gain precision around the bulk of the target posterior  $\pi^\epsilon(\theta | s_{obs})$ .
- iii*) The monotonicity assumption is a necessary condition for ABC as it states that there exists a relation between  $s$  and  $\theta$  through  $f$ , see e.g., Ratmann et al. (2007). The lack of monotonicity is not a fault of the proposed method, but instead it is an indication of the fact that  $s$  is not informative for  $\theta$  in some subset of the parameter space. This would be automatically recognized by the proposal as it would be essentially flat in such a region due to a small Jacobian.
- iv*) The function  $L_Q(\theta)$  can also be useful to fix  $\epsilon$ , because under the assumption that  $\pi_N(\theta | \mathbf{y})$  is  $L_N(\theta)$  dominated and that  $L_Q(\theta)$  was its approximation, then it would be enough to simulate  $\theta$  from  $L_Q(\theta)$  and  $s$  from  $S(y)|\theta$  obtaining thus the distribution for  $\rho(s_{obs}, s)$  and fixing  $\epsilon$  as its suitable quantile.
- v*) The computational cost for approximating  $\pi_N(\theta | \mathbf{y})$  with the proposed approach should include that for estimating  $\hat{f}(\theta)$ , approximating the inverse  $\hat{f}^{-1}(f^*)$  and its Jacobian  $\hat{f}'$ , where the latter is mainly important for  $p > 1$ , as for  $p = 1$  the derivative is obtained analytically. Such costs in the implemented examples are actually reduced by another interpolation of the inverse and Jacobian with splines, or its corresponding equivalent Generalized Additive Model (GAM) for  $p > 1$  (see the next section). This interpolation speeds up the MCMC because at each step we do not need to calculate its inverse and Jacobian, but just its



interpolation. Finally, because the most important cost is model simulations rather than regression estimation, we note that, in the analyzed examples, the number of simulated statistics  $M$  is no larger than 10% of the number of MCMC steps  $T$ . Such computational effort is enough to achieve the desired precision in formulating a proposal distribution.

## 4 The ABC<sub>ql</sub> for $p > 1$

Suppose  $p > 1$  and let  $\mathbf{s}_{obs}$  be the vector of the  $p$  observed statistics. In this case the theory for  $L_Q(\boldsymbol{\theta})$  is not well developed for finite samples; however, when  $p > 1$ ,  $L_Q(\boldsymbol{\theta})$  exists if and only if the matrix  $M(\boldsymbol{\theta})$  is symmetric. In the multi parameter case, we use the regular asymptotic argument for the likelihood (see e.g., Pace and Salvani, 1997, Ch. 4). That is, for  $n \rightarrow \infty$ , the Taylor expansion of the QL around its mode leads to the following multivariate normal QL:

$$L_Q(\boldsymbol{\theta}) = \Sigma_R^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(f(\boldsymbol{\theta}) - \mathbf{s}_{obs})^\top \Sigma_R^{-1}(f(\boldsymbol{\theta}) - \mathbf{s}_{obs})\right),$$

where  $f(\boldsymbol{\theta}) = E(\mathbf{S} \mid \boldsymbol{\theta})$  is a bounded monotone, and possibly non-linear regression function and  $\Sigma_R$  is the conditional covariance matrix of  $\mathbf{S} \mid \boldsymbol{\theta}$ . Following the same approach as for  $p = 1$ , we use  $L_Q(\boldsymbol{\theta})$  as a proposal distribution to be used in a MH scheme (see Algorithm 5) also considering a non constant covariance matrix  $\Sigma_R(\boldsymbol{\theta})$ , that is

$$q^Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t-1)}) = N_p\left(f(\boldsymbol{\theta}^{(t-1)}), \Sigma_R(\boldsymbol{\theta}^{(t-1)})\right) |J(\boldsymbol{\theta})|, \quad (4)$$

where  $N_p(\cdot, \cdot)$  denotes the  $p$ -variate normal distribution with mean  $f(\boldsymbol{\theta}^{(t-1)})$  and variance-covariance matrix  $\Sigma_R(\boldsymbol{\theta}^{(t-1)})$ , and  $J(\boldsymbol{\theta})$  is the Jacobian of the transformation  $f(\boldsymbol{\theta}) = E(\mathbf{S} \mid \boldsymbol{\theta})$ .

In the multi-parameter case we have a system of non-linear equations  $f(\boldsymbol{\theta}) = \mathbf{s}$  whose solution is  $\boldsymbol{\theta} = f^{-1}(\mathbf{s})$ . Such a solution and the calculation of the determinant of the Jacobian,  $|J(\boldsymbol{\theta})|$ , can be obtained using numerical methods for solving a non-linear system of equations and approximating the derivative of  $f(\boldsymbol{\theta})$  at  $\boldsymbol{\theta}$ . In the case of non constant covariance matrix, we use the proposal distribution with covariance

$$\Sigma_R(\boldsymbol{\theta}) = \begin{pmatrix} \sigma_{R1}^2(\boldsymbol{\theta}) & 0 & 0 \\ 0 & \ddots & \\ 0 & 0 & \sigma_{Rp}^2(\boldsymbol{\theta}) \end{pmatrix},$$

where  $\sigma_{R1}^2(\boldsymbol{\theta}), \dots, \sigma_{Rp}^2(\boldsymbol{\theta})$  are the conditional variance functions for each component of  $\mathbf{s}$  with respect to all  $p$  components of  $\boldsymbol{\theta}$ . Note that in this case we are forced to use a diagonal covariance matrix in order to guarantee that  $\Sigma_R(\boldsymbol{\theta})$  is positive definite. The correlation between the  $p$  parameters is then accounted for in the MCMC sampling. Algorithm 4 illustrates how to obtain estimates of  $f(\boldsymbol{\theta})$  and  $\Sigma_R(\boldsymbol{\theta})$  along with the calculation of the Jacobian corresponding to  $\hat{f}(\boldsymbol{\theta})$  for  $p > 1$ .

---

**Algorithm 4** Estimation of  $f(\boldsymbol{\theta})$ ,  $J(\boldsymbol{\theta})$  and  $\Sigma_R(\boldsymbol{\theta})$  for  $p > 1$

---

- 1: Consider the set of  $m = 1, \dots, M^p$  points  $\tilde{\boldsymbol{\theta}}_m = (\tilde{\theta}_{1m}, \dots, \tilde{\theta}_{pm})$ , each of  $p$  scalar coordinates over a regular lattice of  $\Theta_1 \times \dots \times \Theta_p$  and let  $\tilde{\boldsymbol{\theta}}$  be the  $M^p \times p$  matrix of all points;
  - 2: Generate  $\tilde{\mathbf{s}}_m \sim \pi(\mathbf{S}(y) \mid \tilde{\boldsymbol{\theta}}_m)$  and let  $\tilde{\mathbf{s}}$  be the  $M^p \times p$  matrix of all simulated statistics;
  - 3: **for all**  $j = 1, \dots, p$  **do** Regress column  $j$  of  $\tilde{\mathbf{s}}$ ,  $\tilde{s}_j$ , on  $\tilde{\boldsymbol{\theta}}_m$  obtaining  $\hat{f}_j(\boldsymbol{\theta})$  and regression residuals  $\mathbf{e}_j$ . Calculate  $\hat{J}(\boldsymbol{\theta})$  using Richardson's extrapolation (using R package `numDeriv`).
  - 4: **end for**
  - 5: Let  $\mathbf{e} = (\mathbf{e}_1, \dots, \mathbf{e}_p)$  be the  $M^p \times p$  matrix of regression residuals,
  - 6: **if**  $\Sigma_R(\boldsymbol{\theta}) = \Sigma_R$  constant **then**
  - 7:     calculate  $\hat{\Sigma}_R = M^{-1} \mathbf{e}^T \mathbf{e}$ ;
  - 8: **end if**
  - 9: **if**  $\Sigma_R(\boldsymbol{\theta})$  is non constant **then**
  - 10:     regress  $\log(\mathbf{e}_j^2)$  on  $\tilde{\boldsymbol{\theta}}_m$  to have  $\hat{\sigma}_{Rj}^2(\boldsymbol{\theta})$  for  $j = 1, \dots, p$  and obtain  $\hat{\Sigma}_R(\boldsymbol{\theta})$ .
  - 11: **end if**
  - 12: **return**  $\hat{f}(\boldsymbol{\theta}) = (\hat{f}_1(\boldsymbol{\theta}_1), \dots, \hat{f}_p(\boldsymbol{\theta}_p))$ ,  $\hat{J}(\boldsymbol{\theta})$  and  $\hat{\Sigma}_R(\boldsymbol{\theta})$ .
- 

Once we have estimated  $f(\boldsymbol{\theta})$  and  $\Sigma_R$  with  $\hat{f}(\boldsymbol{\theta})$  and  $\hat{\Sigma}_R$ , respectively, we can calculate the proposal  $q^Q(\cdot)$  in (4) and apply Algorithm 5. This is just Algorithm 2 in its multivariate version, where the distance function  $\rho : \mathbb{R}^p \rightarrow \mathbb{R}^+$  must consider the joint distance of all  $p$  coordinates of  $\mathbf{s}$  with respect to  $\mathbf{s}_{obs}$ . Also here  $L_Q(\boldsymbol{\theta})$  can be used to fix  $\epsilon$  in two ways. The first solution, which is the one adopted in this paper, consists of considering a common  $\epsilon$  for all  $p$  dimensions by characterizing the stochastic norm of  $\|\mathbf{s} - \mathbf{s}_{obs}\|$  and its quantiles. The second solution would be to consider different tolerance parameters, one for each of the  $p$  dimensions, by deriving the  $p$  marginals from the joint proposal and then, at each iteration  $t$ , updating each one of the  $p$  parameters separately.

Finally, in order to speed up the MCMC algorithm, especially for large  $p$ , it is worth noting that once  $\hat{f}(\boldsymbol{\theta})$  is estimated, its inverse and the Jacobian  $\hat{J}(\boldsymbol{\theta})$  can be further interpolated by means of their respective values calculated on the points of  $\boldsymbol{\theta}$  used for the pilot-run. With such interpolation, the inverse and Jacobians are calculated only on the points of the grid ( $M^p$  in total), which is much less computationally demanding than a calculation for all MCMC steps.

Many of the remarks outlined in *i*-*v*) hold in the multivariate case as well. In order to guarantee enough flexibility, and because we are mainly interested in predicting  $\mathbf{s}$ , we consider  $\hat{f}(\boldsymbol{\theta})$  and  $\hat{\sigma}_{R1}^2(\boldsymbol{\theta}), \dots, \hat{\sigma}_{Rp}^2(\boldsymbol{\theta})$ , to belong to the class of generalized additive regression models (Stone, 1985) in which each component of  $\boldsymbol{\theta}$  enters into the linear predictor by means of a smoothing spline as discussed, for instance, in Section 12.2 of Faraway (2006). The Jacobian of the non linear system,  $\hat{J}(\boldsymbol{\theta})$  which relates  $p$  summary statistics to the  $p$  parameters, is calculated using Richardson's extrapolation (implemented in the R package `numDeriv`). Finally, the inverse of the non linear system of equations at some point  $\mathbf{s}$  is obtained by Newton steps as detailed in Dennis and Schn-

---

**Algorithm 5** The  $ABC_{ql}$  for  $p > 1$ 


---

**Require:**  $f$ ,  $J(\theta)$  and  $\Sigma_R(\theta)$ , or their estimates  $(\hat{f}, \hat{J}(\theta)$  and  $\hat{\Sigma}_R(\theta)$ ).
1: Set  $\epsilon > 0$ ,  $\theta_0 = f^{-1}(\mathbf{s}_{obs})$ ;2: **for**  $t = 1$  to  $T$  **do**

3:   generate

$$f^* \sim N_p \left( f(\theta^{(t-1)}), \Sigma_R(\theta^{(t-1)}) \right);$$

4:   set  $\theta^* = \{ \theta : f^{-1}(f^*) = \theta \}$ ;5:   generate  $\mathbf{s} \sim \pi(\mathbf{s}(y) \mid \theta^*)$ ;6:   calculate  $\rho = \rho(\mathbf{s}_{obs}, \mathbf{s})$ ;7:   calculate the determinant of the Jacobian matrices  $J(\theta^{(t-1)})$  and  $J(\theta^*)$ ;

8:   with probability

$$\min \left\{ 1, \frac{\pi(\theta^*) q^Q(\theta^{(t-1)} \mid \theta^*)}{\pi(\theta^{(t-1)}) q^Q(\theta^* \mid \theta^{(t-1)})} \mathbb{I}_{\rho < \epsilon} \right\}$$

     accept  $\theta^*$  and set  $\theta^{(t)} = \theta^*$ , otherwise  $\theta^{(t)} = \theta^{(t-1)}$ 
9: **end for**10: **return**  $\theta^{(1)}, \dots, \theta^{(T)}$ 


---

abel (1996) (implemented in the R package `nleqslv`).

We acknowledge that the approach proposed here limits the number of observed statistics to be equal to the number of unknown parameters. This is in line with the general recommendation to keep the number of statistics, i.e. the number of estimating functions, equal to the number of parameters, as also discussed in Mengersen et al. (2013) and Ruli et al. (2013).

## 5 Examples

In this section we illustrate the proposed approach with four examples. The first is a coalescent model (Tavaré et al., 1997) with a scalar parameter of interest. The second example is a gamma model with two unknown parameters, and the third is the  $g$ -and- $k$  distribution (see, e.g., McVinish, 2012) with four unknown parameters. The last example is an application to a real dataset concerning GWAS, with three unknown parameters. In the first example we apply Algorithms 2 and 3, whereas for the other examples we apply Algorithms 4 and 5.

The coalescent model and four-parameter  $g$ -and- $k$  distribution are considered as benchmark examples. The example of the gamma model is useful in order to validate the procedure against a known  $L_N(\theta)$ . Finally, the example of GWAS is relevant for analyzing population genetic isolates for which the genealogy tree is known. Although these kinds of data are rare, they are exceedingly more powerful for detecting genes related to some phenotypes than the usual and more costly GWAS applied to larger

samples of open populations. In all the examples we have used RW-ABC-MCMC, although in Examples 1 to 3 we have used also the independent MH ABC algorithm, obtaining similar results. This is because the proposal distribution, constructed under the framework of QL functions, is located around the bulk of the posterior distribution,  $\pi^\epsilon(\boldsymbol{\theta} \mid \mathbf{s}_{obs})$ . Finally, the use of proper priors, in some examples, has the sole purpose of allowing comparison with existing methods. All examples, except the GWAS, have been implemented assuming that  $\hat{\sigma}_R^2(\boldsymbol{\theta})$  and  $\hat{\Sigma}_R(\boldsymbol{\theta})$  are non constant. In the sequel  $ABC_{ql}$  denotes the types of algorithms where the proposal for the MCMC has been approximated as explained above.

## 5.1 Coalescent model

In the following we consider an example from population genetics, namely the coalescent model analyzed in Tavaré et al. (1997) and Blum and François (2010), among others. Given a set of  $n$  DNA sequences, the aim is to estimate the effective mutation rate,  $\theta' > 0$ , under the infinitely-many-sites model. In this model, mutations occur at rate  $\theta'$  at DNA sites that have not been hit by mutation before. If a site is affected by a mutation, then it is said to be segregating in the sample. In this example, the summary statistic  $s' = y$  is the number of segregating sites. The generating mechanism for  $s'$  is the following:

- 1 Generate  $T_n$ , the length of the genealogical tree of the  $n$  sequences, where  $T_n = \sum_{j=2}^n jW_j$ , where  $W_j$  are independent Exponential random variables with mean  $2/j(j-1)$  such that  $T_n$  has mean  $\mu_{T_n} = 2 \sum_{j=1}^{n-1} 1/j$  and variance  $\sigma_{T_n}^2 = 4 \sum_{j=1}^{n-1} 1/j^2$ ;
- 2 Generate  $(S' \mid \theta', T_n) \sim \text{Poisson}(\theta' T_n / 2)$ .

Hence, the likelihood  $L_N(\theta')$  is given by the marginal density of  $(S' \mid \theta')$  with respect to  $T_n$ , which has a closed form only for  $n = 2$  as  $T_2 \sim \text{Exp}(1/2)$ . For large  $n$ , we approximate the inner integral in  $T_n$  by simulating  $10^5 W_j$  for  $j = 1, \dots, n$  and then obtaining the marginal density of  $(S' \mid \theta')$  by averaging over these  $10^5$  simulated values of  $T_n$ . This parametric approximation is denoted by  $\pi_{ap}(\theta' \mid s'_{obs})$  and relies on the partial knowledge of the likelihood as the marginal density is obtained using the Poisson likelihood.

In order to employ  $ABC_{ql}$ , we consider  $\theta = \log(\theta')$  and  $s = \log(s' + 1)$  instead of  $\theta'$  and  $s'$ , respectively. As an example, we consider  $n = 100$  and an observed value of  $s_{obs} = 2$ , which is a value likely to be obtained under  $\theta = 0$  (and hence  $\theta' = 1$ ). Figure 1 reports the result of the pilot-run study (top-left) with  $M = 1000$  with  $\hat{f}$ ,  $\hat{\sigma}_R^2(\boldsymbol{\theta})$  (top-right), the calculated Jacobian with the splines coefficients (bottom-left) and the approximated posterior (bottom-right) with  $\epsilon$  being the 10% quantile of the distribution of  $\rho(s, s_{obs})$  simulated from  $S \mid \theta$ , where  $\theta \sim L_Q(\boldsymbol{\theta})$ .

From Figure 1 we can see that the chosen summary statistic is not informative for small values of  $\theta$  because observing no segregating sites with  $n = 100$  samples may

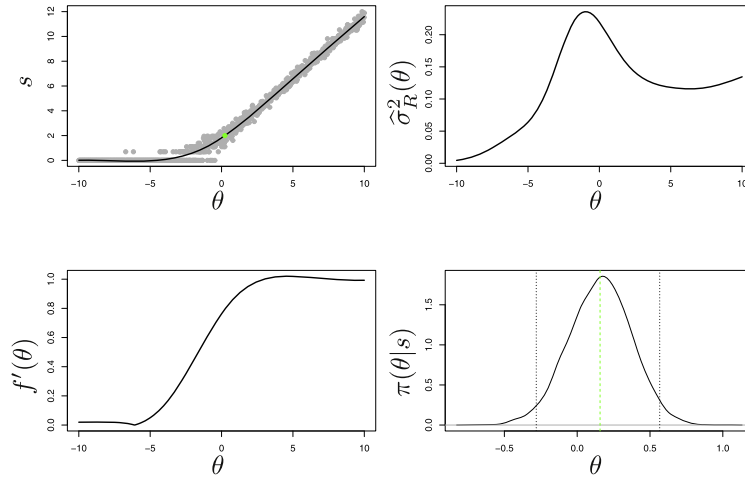


Figure 1: Example: coalescent model. (Top-left) Realizations of  $s$  for the pilot-run study along with the estimated  $\hat{f}$  and  $s_{obs} = 2$  (green). (Top-right) The estimated conditional variance of  $s|\theta$ ,  $\hat{\sigma}_R^2(\theta)$ . (Bottom-left) The Jacobian of  $\hat{f}(\theta)$  in the grid of the pilot-run. (Bottom-right) The approximated posterior  $\pi(\theta|s_{obs})$  with 95% credible interval and posterior mean (green).

occur for almost every mutation rate lower than  $e^{-5}$ . This is, of course, not a fault of the method, but of the chosen summary statistic and it may also occur in the standard original ABC approaches. Moreover, this is reflected by the proposed method as the estimated Jacobian is near 0 for values of  $\theta' < e^{-5}$ . It can also be seen that for larger values, the approximated Jacobian is nearly constant, suggesting that there exists a linear relation between  $s$  and  $\theta$ .

For  $\pi(\theta) = Exp(1)$  and  $n = 100$  we calculated, for each dataset simulated at different values of  $\theta \in (2, 3, \dots, 10)$ , the relative difference of quantiles of each posterior with respect to those obtained with the parametric approximation, as in Blum and François (2010). The relative difference is defined as  $(Q_p - Q_p^0)/Q_p^0$  where  $Q_p$  and  $Q_p^0$  are the  $p$ -th quantiles of the ABC posterior and that of the parametric approximation, respectively. Figure 2 shows the relative differences. We can clearly see that these differences are more robust with respect to  $\theta$  for the ABC<sub>qt</sub> rather than for the ABC, and this is due to the impact of the prior in the standard ABC algorithm. In fact, for  $\theta \rightarrow \infty$  the discrepancy between prior and posterior becomes important.

### 5.2 Gamma model with unknown shape and scale parameters

Let  $y \sim Gamma(\theta_1, \theta_2)$  with mean  $\exp(\theta_1 - \theta_2)$  and variance  $\exp(\theta_1 - 2\theta_2)$ , where  $\theta_1$  and  $1/\theta_2$  are the log of shape and scale, respectively. We have  $p = 2$ ,  $\theta \in \Theta = \Theta_1 \times \Theta_2 \equiv \mathbb{R}^2$  and consider the following two statistics  $\mathbf{s} = (s_1, s_2)$ , where  $s_1$  and  $s_2$  are the logarithms of the sample mean and the standard deviation, respectively. For  $\theta = (0, 0)$  we consider a

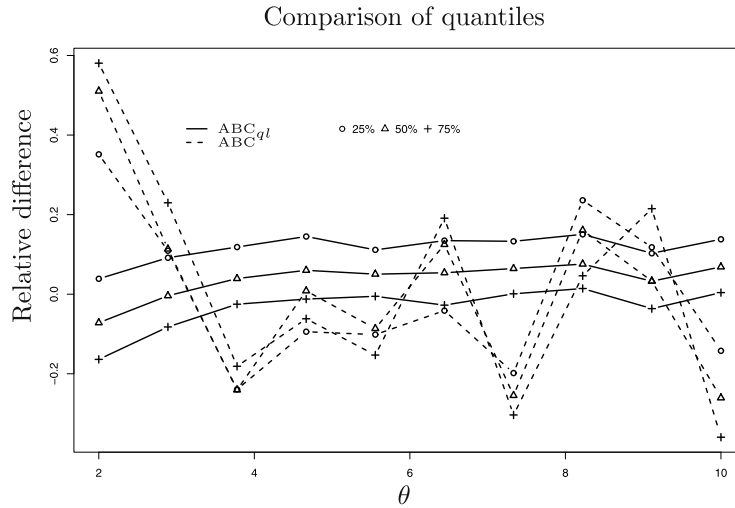


Figure 2: Example: coalescent model. Comparison of ABC and  $ABC_{ql}$  with the parametric approximation in terms of relative differences between quantiles of the parametric approximation and those of ABC and  $ABC_{ql}$ .

sample of size  $n = 10$  with  $\mathbf{s}_{obs} = (-0.12, -0.26)$  and estimate the posterior distribution under two independent standard normal priors,  $\pi(\boldsymbol{\theta}) = \pi(\theta_1) \times \pi(\theta_2) = N(0, 1) \times N(0, 1)$ . We consider the estimation of  $\hat{f}(\boldsymbol{\theta})$  with  $M = 100$ ,  $p = 2$  over a regular lattice of  $M^2 = 10^4$  points in  $(-2, 2) \times (-2, 2) \in \Theta_1 \times \Theta_2$ .

Figure 3 shows the conditional regression functions of each statistic against the parameters, which appear quite linear. Figure 4 illustrates the log of squared residuals and the estimation of the conditional variances  $\hat{\sigma}_{R1}^2(\boldsymbol{\theta})$  and  $\hat{\sigma}_{R2}^2(\boldsymbol{\theta})$  with respect to  $\boldsymbol{\theta}$ . Figures 5 and 6 illustrate the MCMC output. From the former we can deduce that the chain mixes well, and that marginal posteriors  $\pi^\epsilon(\theta_1 | \mathbf{s}_{obs})$  and  $\pi^\epsilon(\theta_2 | \mathbf{s}_{obs})$  are centered around the true values. From the latter, we see that the bivariate density  $\pi^\epsilon(\boldsymbol{\theta} | \mathbf{s}_{obs})$  obtained by the  $ABC_{ql}$ , and the true underlying posterior  $\pi_N(\boldsymbol{\theta} | y)$  are similar in terms of contour levels.

We notice that the output of the MCMC in Figure 6 is similar to that obtained using  $\Sigma_R$  constant (not reported here). This is presumably because, in this example, the use of logarithms on the scale of the summary statistics stabilizes their conditional variance with respect to  $\boldsymbol{\theta}$ . Another reason is because the chain moves in a radius of  $\mathbf{s}_{obs} < \epsilon$  where the variance functions  $\hat{\sigma}_{R1}^2(\boldsymbol{\theta})$ ,  $\hat{\sigma}_{R2}^2(\boldsymbol{\theta})$  are almost constant and do not differ significantly from the estimated diagonal of  $\Sigma_R$  when it is assumed constant.

### 5.3 Four-parameter $g$ -and- $k$ distribution

Distributions based on quantiles are of great interest because of their flexibility. However, although a stochastic representation is available, their density and hence the likelihood

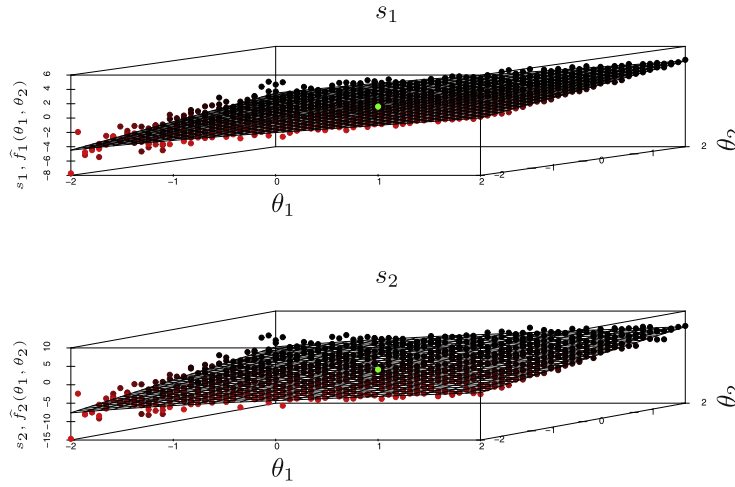


Figure 3: Example: gamma model. Conditional regression functions of each statistic against the two parameters. The green point is the observed value.

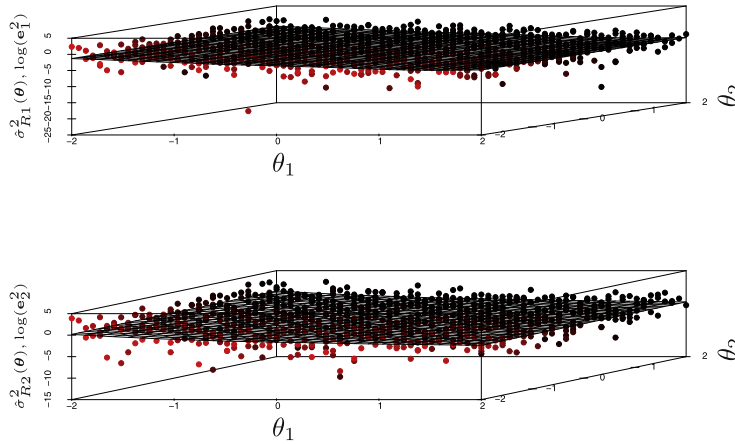


Figure 4: Example: gamma model. Conditional variance functions of each statistic against the two parameters.

$L_N(\theta)$  are not available in closed form and, in general, they are difficult to evaluate.

We focus on the four-parameter ( $p = 4$ ),  $g$ -and- $k$  distribution, which has the following stochastic representation

$$z \sim N(0, 1)$$

$$(y | z, \theta_1, \theta_2, \theta_3, \theta_4) = \theta_1 + \exp(\theta_2) \left( 1 + 0.8 \frac{1 - \exp(-\theta_3 z)}{1 + \exp(-\theta_3 z)} \right) (1 + z^2)^{\exp(\theta_4) - 1/2} .$$

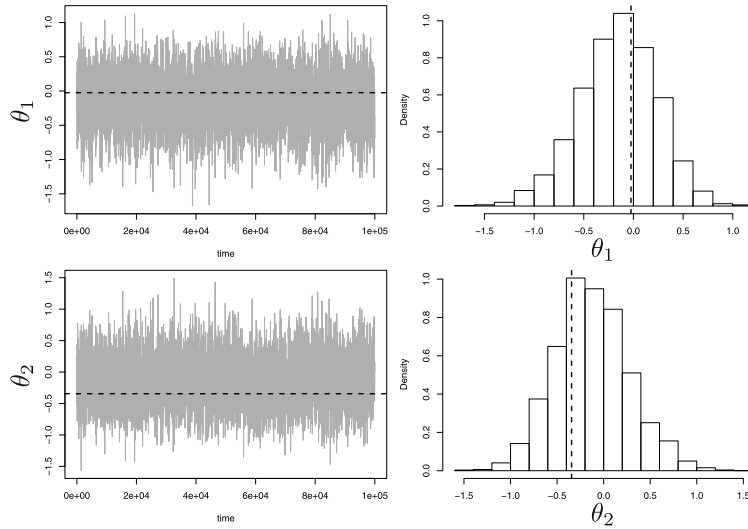


Figure 5: Example: gamma model. Marginal output of Algorithm 5, for  $\theta_1$  and  $\theta_2$  along with the histogram of the marginal posteriors  $\pi^\epsilon(\theta_1 | s_{obs})$  and  $\pi^\epsilon(\theta_2 | s_{obs})$ . Vertical dotted lines are the true values of  $\theta_1$  and  $\theta_2$  that are used to generate  $s_{obs}$ .

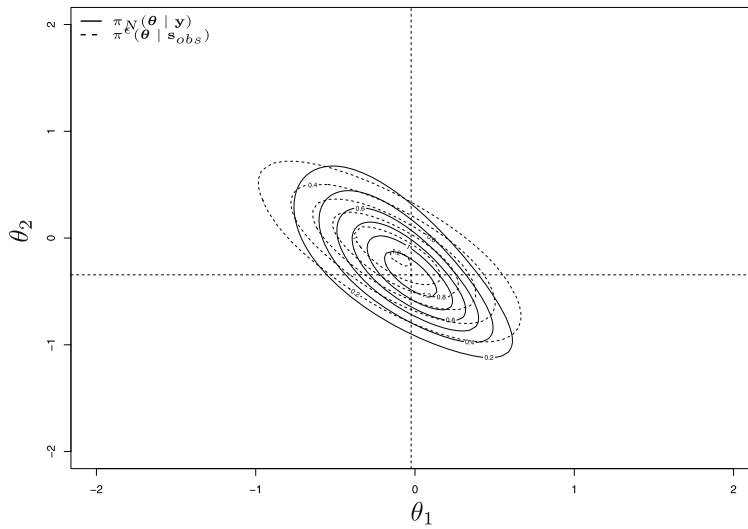


Figure 6: Example: gamma model. Contours for  $\pi^\epsilon(\boldsymbol{\theta} | s_{obs})$  (dashed) along with  $\pi_N(\boldsymbol{\theta} | \mathbf{y})$  (continuous). The posterior modes of  $\theta_1$  and  $\theta_2$  are represented by the cross point of the dashed lines, where the true value is  $\boldsymbol{\theta} = (0, 0)$ .



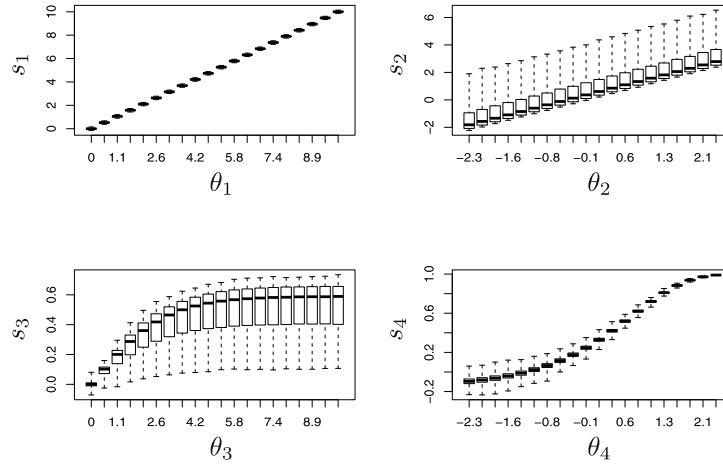


Figure 7: Example:  $g$ -and- $k$  distribution. Conditional distribution of  $s_i|\theta_i$ , for  $i = 1, \dots, 4$ . Each boxplot represents the marginal distribution of the statistic at the specified value of  $\theta_i$  in the horizontal axis.

The unknown parameters  $\theta_1, \exp(\theta_2), \theta_3$  and  $\exp(\theta_4) - 1/2$  represent location, scale, skewness and kurtosis, respectively (Haynes et al., 1997).

Such distributions have also been used for testing several ABC approaches as in Marjoram et al. (2003); McVinish (2012).

We consider the following four statistics all based on empirical quantiles  $q$  of  $y_q$ :

$$s = \left( y_{0.5}, \log(y_{0.75} - y_{0.25}), \frac{y_{0.75} + y_{0.25} - 2y_{0.5}}{y_{0.75} - y_{0.25}}, \log(y_{0.975} - y_{0.025}) \right) \in \mathbb{R}^4,$$

with the following meaning:  $s_1$  is the median,  $s_2$  is the log of the interquartile range and  $s_3$  is the skewness index described in Bowley (1937), a special case of the Hinkley (1975) index. Finally  $s_4$  is the log transformation of the kurtosis index described in Crow and Siddiqui (1967).

From the pilot-run simulation shown in Figure 7, it is possible to notice that the relationship between statistics and parameters is not linear. In this example, we consider a sample simulated under the scenario discussed in Fearnhead and Prangle (2012) in which  $n = 10^4$  observations are generated with  $\theta = (3, 0, 2, -\log(2))$  and the uniform prior on  $[0, 10] \times [-\log(10), \log(10)] \times [0, 10] \times [-\log(10), \log(10)]$  is considered. Figure 8 reports the conditional distributions of the logarithm of squared residuals, indicating that the variance may be non constant. Therefore, for this model, we estimated with GAMs the conditional variance functions  $\hat{\sigma}_{R_j}^2(\theta)$ , for  $j = 1, \dots, 4$  as explained in Algorithm 4.

Figure 9 shows the output of the  $ABC_{qt}$  with the RW-ABC-MCMC algorithm for the four marginal posterior densities, for a given sample. We can see that the marginal

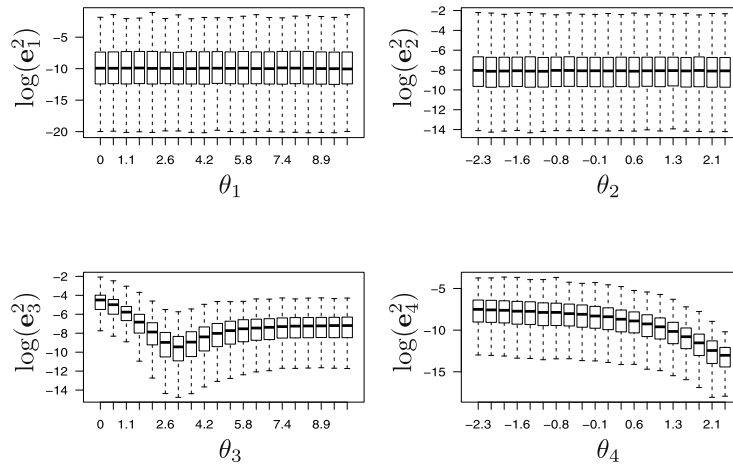


Figure 8: Example:  $g$ -and- $k$  distribution. Logarithms of squared residuals in the pilot-run. Each boxplot is the conditional distribution of  $\log(\mathbf{e}_j^2)|\theta_j$ ,  $j = 1, 2, 3, 4$ . Such values indicate that  $\Sigma_R(\boldsymbol{\theta})$  is not constant with respect to  $\boldsymbol{\theta}$ .

posterior distributions include, in their high posterior density interval, the true value of  $\boldsymbol{\theta}$  and the chains have good mixing.

Finally, the performance of the proposed method is assessed by a simulation study of 50 simulated datasets. As in Fearnhead and Prangle (2012), the 50 datasets are simulated from 50 different parameter values sampled from the prior. The expected quadratic errors for each marginal  $ABC_{ql}$  posterior are reported in Figure 10, where also the Mean Squared Error (MSE) of the Semi-Automatic ABC, taken from Fearnhead and Prangle (2012), is shown for comparison of the order of magnitudes.

From Figure 10, we can see that the expected quadratic error under  $ABC_{ql}$  is compatible with that reported for the Semi-Automatic ABC. Notice that our method uses a set of four observable summary statistics that differ from those ones used in Fearnhead and Prangle (2012).

## 5.4 GWAS for isolated populations with known genealogy

In this application, we address the problem of estimating DNA markers related to a certain phenotype such as, for instance, the presence of a certain disease. In this problem genotype is represented by a large set of DNA sequences known as Single-Nucleotide Polymorphisms or SNPs in the sequel. Such SNPs are usually observed in millions per individuals and thus fast and reliable statistical methods are needed in order to answer the scientific question as to which SNPs are mainly related to the disease. A dataset for a case/control study is usually collected on an open population where the degree of inbreeding, that is the mating of pairs who are closely related genetically, is unknown as it is usually negligible in open populations. Nonetheless, there exists certain evidence in the genetic literature that very valuable indications regarding SNPs/Disease

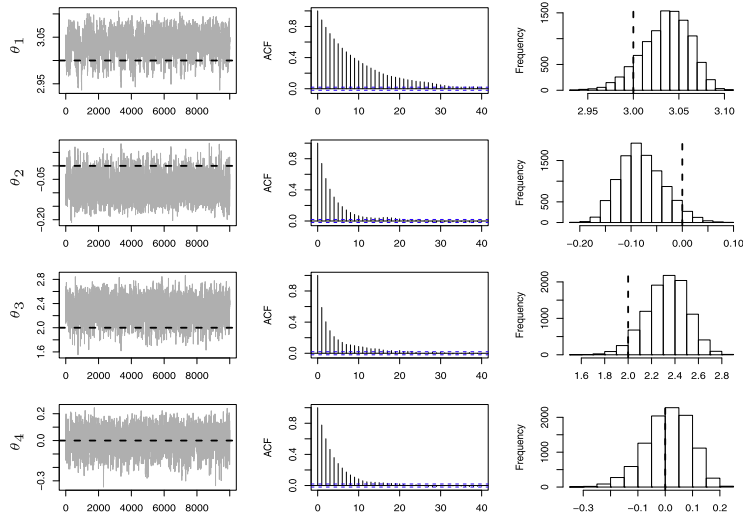


Figure 9: Example:  $g$ -and- $k$  distribution. The approximated marginal posterior  $\pi^\epsilon(\theta_i \mid \mathbf{s}_{obs})$ ,  $i = 1, 2, 3, 4$ .

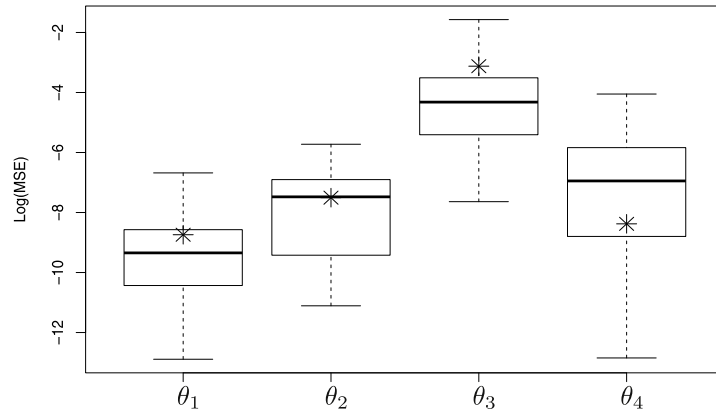


Figure 10: Example:  $g$ -and- $k$  distribution. The marginal distributions of the Log of MSE for each of the four parameters, in 50 replications of the considered simulation scenario, along with the Log of MSE of the Semi-Automatic ABC procedure reported in Fearnhead and Prangle (2012) (stars).

relationships may come from the study of isolate genetics, i.e. human samples for which inbreeding is also relevant and known. Such types of collected samples are very rare because there are very few genetic isolates in the world, and so the statistical methods to analyze them are not very well developed. One example of a genetic isolate for which data are also available is the Sardinian genetic isolate of the Ogliastra region, which is

situated in the center of the island of Sardinia (Cabras et al., 2011).

An example of such data may come from Figure 11 in which we have a population composed of 4 families, 18 individuals and 2 SNPs labeled as SNP1 and SNP2. From Figure 11 we have that ancestors have not been observed because have died (white); while offsprings are labelled as healthy (green) or affected (red) along with their SNP configurations.

Data in Figure 11 may be formally represented as follows: for individual  $i$  let  $Y_i \in \{0, 1\}$  represent the indicator of the phenotype, i.e.  $Y_i = 1$  if affected and  $Y_i = 0$  otherwise; assume  $X_i \in \{aA, Aa, aa, AA\}$  represents the genotype, e.g. the SNP configuration with three levels. For a genealogy composed of  $N$  individuals we have to model the corresponding pairs  $(Y_1, X_1), \dots, (Y_N, X_N)$ , where only  $n < N$  have been observed. At the phenotype level we assume the usual logit model  $(Y_i|X_i, \theta) \sim \text{Bernoulli}(p_i)$ , with  $p_i$  being the probability that individual  $i$  is affected.

If one considers the data in Figure 11 as  $n = 10$  independent observations and estimates, for instance, a logistic regression model of  $Y$  against  $X$ , or just considers the Fisher exact test among  $Y$  and  $X$ , one would end up finding no association. In particular, the Fisher exact test between  $Y$  and the first SNP has a  $p$ -value of 0.21 while with SNP2 it is exactly 1. While the latter  $p$ -value is reasonable as there is apparently no association between SNP2 and  $Y$ , the former is not, because all  $aa$  individuals are affected and all  $AA$  individuals are healthy and therefore there should be certain evidence of association between the first SNP and affection status. The shortcoming of this analysis is that it treats all individuals as independent and identically distributed, while it is clear that they are not.

On the other hand, the situation of highly dependent observations complicates the statistical model. In fact, the probabilistic model for the sample must take into account the genealogy that underlies the genetic variant transmission and also the model which relates the phenotype to the genotype. As the genotype is observed for the very last generations only, the configurations of the SNPs for the previous generations then constitute an enormous number of random latent variables. This makes it almost impossible to write the likelihood for the parameter relating the SNPs configuration with the phenotype, that is, the coefficients of a logistic regression between  $Y$  and  $X$ .

In this application we illustrate the  $ABC_{gl}$  method for the data in Figure 11 and we also consider a sample from the village of Talana (Ogliastra, Sardinia - Italy) that is affected by a reduced Mean Cell Volume (i.e. a  $MCV < 72$ ) disease for which it is known that there exists a genetic variant inside the Beta-Globin gene which determines it. The data, provided by the Centro Nazionale Ricerche of Italy, consists of  $N = 1997$  individuals, all in one tree, originating from two common ancestors. Only  $n = 49$  individuals of the later generations are observed, among whom only 5 are affected. Moreover, the proportion of affected, similar to the prevalence of the disease, is around 13%. There are 91 SNPs with three levels and we know that only one is inside the Beta-Globin gene.

In this analysis, we treat each SNP separately and set up a stochastic model for a single SNP. The overall analysis for all SNPs is made by the sequence of separated

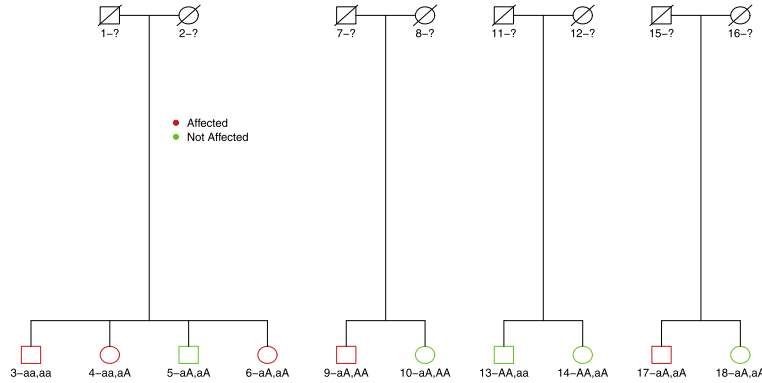


Figure 11: Example of a genealogy tree with 4 families, 18 individuals and 2 SNPs (SNP1,SNP2). Ancestors have not been observed (white); while the offspring are labelled as healthy (green) or affected (red).

analyses over all SNPs. In the linear predictor of the logistic regression, we consider as covariates the genotype for individual  $i$ :

$$\text{logit}(p_i) = \theta_1 \mathbb{I}_{X_i=\{aA,AA\}} + \theta_2 \mathbb{I}_{X_i=\{aa\}} + \theta_3 \mathbb{I}_{X_i=\{AA\}}.$$

The vector of coefficients  $\theta$  are usually interpreted as the log of the odds ratio for the probabilities of being affected given a SNP configuration. In order to account for the fact that the sample  $(y_1, x_1), \dots, (y_n, x_n)$  is not i.i.d. we include the transmission model for the genetic variants. Specifically, let  $X_{i_1}$  and  $X_{i_2}$  be the SNP configuration for the ancestors of individual  $i$ . Then the probabilistic model for the transmission of a genotype variant is assumed to be regulated by the usual Mendelian inheritance model of transmission, where the ancestor individuals are assumed to be known, according to the genealogical tree. This law is also known as the law of independent assortment, segregation or dominance, see for instance Levitan (1988). Therefore, if individual  $i$  is a descendent in the tree

$$(X_i | X_{i_1}, X_{i_2}) \sim \text{Mendel's law},$$

while if  $i$  is a founder or her/his ancestors are not in the tree, then we assume the following prior distribution for configuration of ancestors:

$$X_i \sim \text{Trinomial}(1/3, 1/3, 1/3).$$

The summary statistics, calculated only for the  $n$  observations, are the observed log-odds of the proportion of affected among all individuals that have a certain configuration. Specifically, let  $\#\{\omega\}$  count the number of occurrences of type  $\omega$ ,

$$s_k = \log \left( \frac{1 + \#\{Y = 1 | X = k\}}{2 + \#\{X = k\}} \right),$$

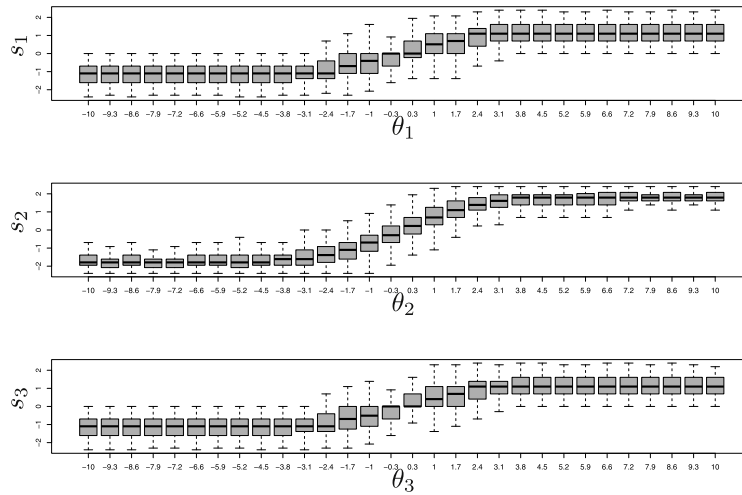


Figure 12: Pilot-run for the genealogy tree in Figure 11 with  $M = 30^3$  points. Each boxplot is the conditional distribution of  $s_k|\theta_k$ ,  $k = 1, 2, 3$ .

where  $k = 1, 2, 3$  corresponds to  $X = aa$ ,  $X = \{aA, Aa\}$  and  $X = AA$ , respectively. Note that we add one individual in the numerator and two in the denominator in order to guarantee that  $s_1, s_2$  and  $s_3$  are always defined. This, of course, constitutes a limitation for very small samples from which, however, it would be difficult to estimate very strong signals for a SNP that is a risk or protection factor.

In order to run  $ABC_{qt}$ , we performed a pilot-run simulation with  $M = 30^3$  points on a regular grid of log odds ratios between -10 and 10. This pilot-run study depends only on the genealogy tree and not on the observed genotypes or phenotypes. In the case of GWAS analysis, this provides a saving in computational efforts, as in general there are many genes to be analysed.

The results for the pilot-run study are summarized in Figure 12 where we can see that the chosen statistics are quite informative around the null hypothesis of no association. For very large signals, e.g.  $|\theta_k| > 3$ , the summary statistics are very weakly informative. This is not the fault of the summary statistics, but it is due to the small observed sample, as is typical in genetic isolates.

Figure 13 illustrates the conditional distributions of the logarithm of squared residuals with respect to  $\theta_1, \theta_2$  and  $\theta_3$ , which could be used to estimate the variance function  $\Sigma_R(\theta)$ . However, in this case we find it reasonable to assume that  $\Sigma_R(\theta)$  is constant and we estimate it with  $\hat{\Sigma}_R$  as explained in Algorithm 4.

We complete the algorithm by using the Euclidean distance  $\rho(\mathbf{s}, \mathbf{s}_{obs})$ , between  $\mathbf{s}$  and its observed value  $\mathbf{s}_{obs}$  further weighted by a term that takes into account the simulated configurations for the SNP. This term makes the distance tend to 0 when there are

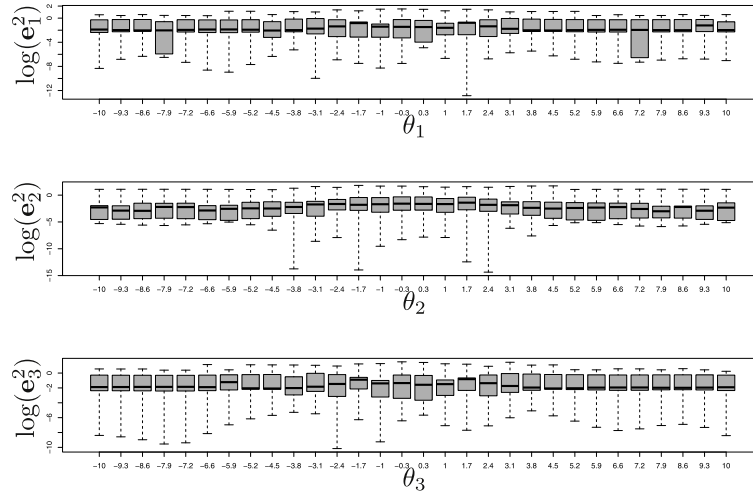


Figure 13: Logarithms of squared residuals in the pilot-run for the genealogy tree in Figure 11. Each boxplot is the conditional distribution of  $\log(\mathbf{e}_k^2)|\theta_k$ ,  $k = 1, 2, 3$ . Such values indicate that  $\Sigma_R(\boldsymbol{\theta})$  is constant with respect to  $\boldsymbol{\theta}$ .

many matches between simulated and observed configurations,

$$1 - \frac{\#\{x_{obs} = x_{sim}\}}{n}.$$

The tolerance parameter  $\epsilon$  has been fixed in order to obtain an acceptance probability in the RW-ABC-MCMC algorithm around 30%. The output of the chain for data in Figure 11 is represented in Figure 14.

From Figure 14 we can see that SNP1, which has the largest signal, exhibits values of  $\theta$  with the largest posterior mean and the largest uncertainty. Moreover, the approximated marginal posteriors for  $\theta_1$  and  $\theta_3$  for SNP1 are very skewed. For SNP2 where there is no signal, posterior distributions are centered around 0. These results are also reflected by the logarithm of the Bayes Factors (BFs) for  $(\theta > 0 | \mathbf{s}_{obs})$  against  $(\theta < 0 | \mathbf{s}_{obs})$ ,  $\Pr(\theta > 0 | \mathbf{s}_{obs}) / \Pr(\theta < 0 | \mathbf{s}_{obs})$  which is defined as long as the posterior of  $\theta | \mathbf{s}_{obs}$  exists. In fact, there is substantial evidence for the configurations of SNP1 to be risk or protective factors, but not for SNP2.

We repeated the above analysis for the Talana data and found that the SNP inside the Beta-Globin (**rs11036238**) is among the first three SNPs, out of 91, with the highest posterior mean (in absolute value) as shown in Figure 15. Those SNPs also have the largest Bayes Factors. However, the greater uncertainty for the first three SNPs in Figure 15 is due to the fact that with only  $n = 49$  observed individuals we cannot be very precise in estimating very large signals as discussed above.

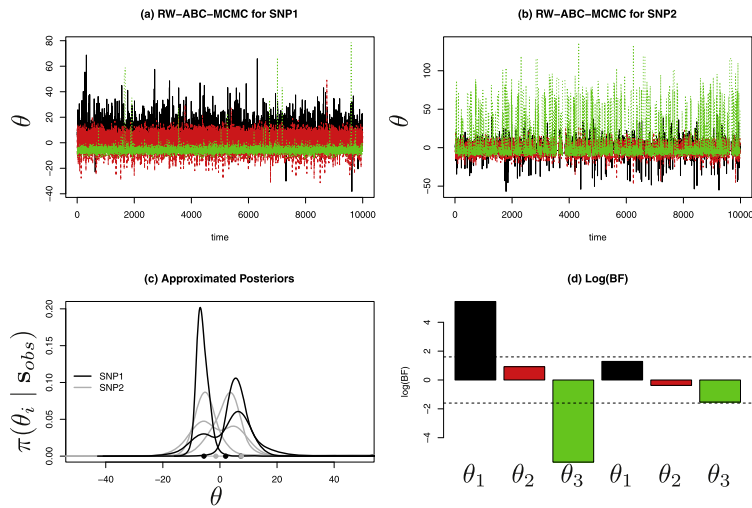


Figure 14: Output of the analysis for data in Figure 11. Chain output for RW-ABC-MCMC for the two SNPs (a,b) and corresponding density estimation (c) with posterior means represented by dots. Logarithm of Bayes Factor for  $(\theta > 0 \mid \mathbf{s}_{obs})$  against  $(\theta < 0 \mid \mathbf{s}_{obs})$  along with the reference lines at  $\pm 1.6$  (d).

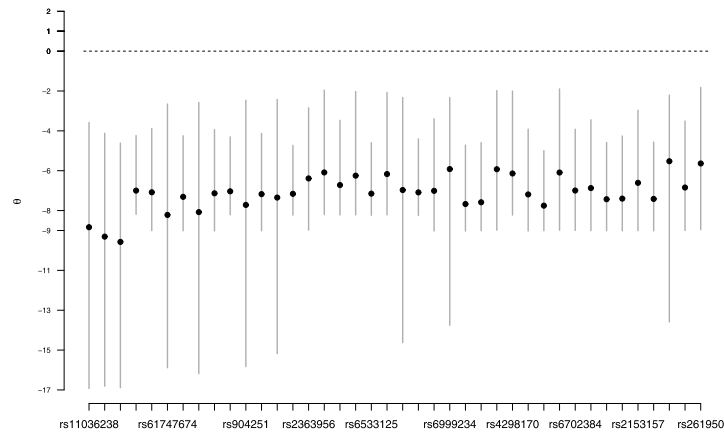


Figure 15: For data from the village of Talana we report, for 40 different SNPs, 95% credible intervals for those  $\theta$ s with the largest posterior mean (in absolute) value (dot). The posterior is approximated with the RW-ABC-MCMC.

## 6 Conclusions

Recently, the idea of using simulation from the model to approximate the distribution of summary statistics seems to be proliferating in the ABC literature (see, e.g. Prangle et al. (2013b); Wood (2010); Ratmann et al. (2014)). In this paper we also used this type



of approach. In particular, we simulate from the model varying the parameters in a grid to approximate the distribution of summary statistics as a function of the parameters in order to build a suitable proposal for ABC-MCMC. Such a proposal distribution can be implemented in a RW fashion or can be used as an independence kernel, although we focused mainly on RW type MCMC algorithms.

In scalar parameter problems with conditional constant variance of summary statistics with respect to the parameters, we showed by using the definition of quasi-likelihood (McCullagh, 1991) that this proposal is a normal kernel in the auxiliary space,  $f(\theta)$ . In multiparameter problems or when the variance of the regression function cannot be assumed to be constant, the theory of quasi-likelihoods only suggests a form for the proposal. In fact, analogously to the scalar parameter case with constant variance, we propose using a multivariate normal kernel in the auxiliary space.

A key point for the success of our method is that the summary statistics must vary when changing the parameter values. Moreover, there must be a one-to-one relation between them and again, the choice of  $\mathbf{s}$  is critical, as an ancillary statistic is useless for gathering information about  $\theta$ . Such a non ancillarity assumption is usually required over the whole parameter space  $\Theta$  and it may happen that there could be parts of the parameter space where  $\mathbf{s}$  is locally ancillary. This happens, for instance, in the coalescent model for low mutation rates and also in the application to GWAS. Such ancillarity, however, is properly accounted for in the discussed proposal density for ABC-MCMC. Another problem with our approach may lie in the asymptotic argument for  $p > 1$  which may not hold in some applications when  $L_Q(\theta)$  is irregular.

The proposed  $ABC_{ql}$  seems to perform quite well in the above examples compared to other available methods; it is straightforward to apply and its implementation does not require more than just basic notions of regression analysis. We discussed two possible ABC-MCMC algorithms, the RW-MH with or without constant regression variance. Although we focus mainly on the non constant variance assumption, we found that the original independent MH with constant regression variance leads to satisfactory results for the discussed examples. This is because we are only implementing a proposal distribution for the ABC-MCMC and also because the estimation theory of QL functions allows for a good approximation of  $L_N(\theta)$  which is reflected in the proposal.

Furthermore, when simulation from  $\pi(y | \theta)$  is costly another alternative to ABC could be using the  $L_Q(\theta)$  as a surrogate of  $L_N(\theta)$  as in Cabras et al. (2014). For other surrogate pseudo-likelihoods used in the ABC context, see also Mengersen et al. (2013); Pauli et al. (2011) among others.

## References

- Adimari, G. and Ventura, L. (2002). “Quasi-Profile Log Likelihoods for Unbiased Estimating Functions.” *Annals of the Institute of Statistical Mathematics*, 54: 235–244. [414](#)
- Aeschbacher, S., Beaumont, M. A., and Futschik, A. (2012). “A novel approach for choosing summary statistics in approximate Bayesian computation.” *Genetics*, 192(3): 1027–1047. [412](#)

- Andrieu, C. and Roberts, G. (2009). “The pseudo-marginal approach for efficient Monte Carlo computations.” *The Annals of Statistics*, 37(2): 697–725. [412](#), [415](#)
- Atchadé, Y. F. and Perron, F. (2007). “On the geometric ergodicity of Metropolis-Hastings algorithms.” *Statistics*, 41(1): 77–84. [414](#)
- Barndorff-Nielsen, O. (1995). “Quasi profile and directed likelihoods from estimating functions.” *Annals of the Institute of Statistical Mathematics*, 47: 461–464. [414](#)
- Barnes, C. P., Filippi, S., Stumpf, M. P., and Thorne, T. (2012). “Considerate approaches to constructing summary statistics for ABC model selection.” *Statistics and Computing*, 22(6): 1181–1197. [412](#)
- Beaumont, M., Cornuet, J. M., Marin, J. M., and Robert, C. P. (2009a). “Adaptive approximate Bayesian computation.” *Biometrika*, 96(4): 983–990. [412](#)
- Beaumont, M. A., Cornuet, J.-M., Marin, J.-M., and Robert, C. P. (2009b). “Adaptive approximate Bayesian computation.” *Biometrika*, 96: 983–990. [412](#)
- Bellio, R., Greco, L., and Ventura, L. (2008). “Modified quasi-profile likelihoods from estimating functions.” *Journal of Statistical Planning and Inference*, 138: 3059–3068. [414](#)
- Biau, G., Cérou, F., and Guyader, A. (2012). “New Insights into Approximate Bayesian Computation.” *arXiv preprint*, arXiv:1207.6461. [412](#)
- Blum, M. G. B. and François, O. (2010). “Non-linear regression models for Approximate Bayesian Computation.” *Statistics and Computing*, 20: 63–73. [422](#), [423](#)
- Blum, M. G. B., Nunes, M., Prangle, D., and Sisson, S. A. (2013). “A comparative review of dimension reduction methods in approximate Bayesian computation.” *Statistical Science*, 28(2): 135–281. [412](#)
- Blum, M. G. B. and Tran, V. C. (2010). “HIV with contact tracing: a case study in approximate Bayesian computation.” *Biostatistics*, 11: 644–660. [411](#)
- Bortot, P., Coles, S. G., and Sisson, S. A. (2007). “Inference for stereological extremes.” *Journal of the American Statistical Association*, 102: 84–92. [412](#)
- Bowley, A. L. (1937). *Elements of statistics*, volume 8 of *Studies in economics and political science*. London: P. S. King & Son, Ltd.; New York, C. Scribner’s Sons, 6th edition. [427](#)
- Cabras, S., Castellanos, M. E., Biino, G., Persico, I., Sassu, A., Casula, L., del Giacco, S., Bertolino, F., Pirastu, M., and Pirastu, N. (2011). “A strategy analysis for genetic association studies with known inbreeding.” *BMC Genetics*, 12:63. [430](#)
- Cabras, S., Castellanos, M. E., and Ruli, E. (2014). “A Quasi likelihood approximation of posterior distributions for likelihood-intractable complex models.” *Metron* (in press). [435](#)
- 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: 2713–2719. [412](#)

- Crow, E. L. and Siddiqui, M. M. (1967). “Robust Estimation of Location.” *Journal of the American Statistical Association*, 62: 353–389. [427](#)
- Dennis, J. E. J. and Schnabel, R. B. (1996). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Siam. [420](#)
- Desmond, A. F. (1997). “Optimal estimating functions, quasi-likelihood and statistical modelling.” *Journal of Statistical Planning and Inference*, 60: 77–104. [414](#)
- Faisal, M., Futschik, A., and Hussain, I. (2013). “A new approach to choose acceptance cutoff for approximate Bayesian computation.” *Journal of Applied Statistics*, 40(4): 862–869. [412](#)
- Faraway, J. J. (2006). *Extending the Linear Model with R*. Springer, New York. [420](#)
- Fearnhead, P. and Prangle, D. (2012). “Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation.” *Journal of Royal Statistical Society: Series B*, 74: 419–474. [413](#), [427](#), [428](#), [429](#)
- Foll, M., Beaumont, M. A., and Gaggiotti, O. (2008). “An approximate Bayesian computation approach to overcome biases that arise when using amplified fragment length polymorphism markers to study population structure.” *Genetics*, 179: 927–939. [411](#)
- Gourieroux, C., Monfort, A., and Renault, E. (1993). “Indirect Inference.” *Journal of Applied Econometrics*, 8: S85–S118. [413](#)
- Greco, L., Racugno, W., and Ventura, L. (2008). “Robust likelihood functions in Bayesian inference.” *Journal of Statistical Planning and Inference*, 138: 1258–1270. [414](#)
- Hamilton, G., Currat, M., Ray, N., Heckel, G., Beaumont, M., and Excoffier, L. (2005). “Bayesian estimation of recent migration rates after a spatial expansion.” *Genetics*, 170: 409–417. [412](#)
- Haynes, M. A., MacGillivray, H. L., and Mengersen, K. L. (1997). “Robustness of ranking and selection rules using generalised g-and-k distributions.” *Journal of Statistical Planning and Inference*, 65: 45–66. [427](#)
- Heggland, K. and Frigessi, A. (2004). “Estimating functions in indirect inference.” *Journal of the Royal Statistical Society: Series B*, 66: 447–462. [417](#)
- Heyde, C. C. (1997). *Quasi-likelihood and its application: a general approach to optimal parameter estimation*. Springer Verlag. [414](#)
- Hinkley, D. V. (1975). “On Power Transformations to Symmetry.” *Biometrika*, 62: 101–111. [427](#)
- Jørgensen, B. and Knudsen, S. J. (2004). “Parameter orthogonality and bias adjustment for estimating functions.” *Scandinavian Journal of Statistics*, 31: 93–114. [414](#)
- Lee, A. (2012). “On the choice of MCMC kernels for approximate Bayesian computation with SMC samplers.” In *Proceedings of the 2012 Winter Simulation Conference (WSC)*, 1–12. IEEE. [412](#)

- Levitan, M. (1988). *Textbook of human genetics, third edition*. Oxford University Press. 431
- Liang, K. Y. and Zeger, S. L. (1995). “Inference based on estimating functions in the presence of nuisance parameters.” *Statistical Science*, 10: 158–173. 414
- Lin, L. (2006). “Quasi Bayesian likelihood.” *Statistical Methodology*, 3: 444–455. 414
- Marjoram, P., Molitor, J., Plagnol, V., and Tavaré, S. (2003). “Markov chain Monte Carlo without likelihoods.” *Proceedings of the National Academy of Sciences of the United States of America*, 100: 15324–8. 412, 414, 427
- McCullagh, P. (1991). “Quasi-likelihood and estimating functions.” In Hinkley, D., Reid, N., and Snell, E. (eds.), *Statistical Theory and Modelling*, 265–286. Chapman and Hall: London. 413, 414, 435
- McVinish, R. (2012). “Improving ABC for quantile distributions.” *Statistics and Computing*, 22: 1199–1207. 421, 427
- Mengersen, K. L., Pudlo, P., and Robert, C. P. (2013). “Approximate Bayesian computation via empirical likelihood.” *Proceedings of the National Academy of Sciences of the United States of America*, 110(4): 1321–1326. 417, 421, 435
- Mengersen, K. L. and Tweedie, R. L. (1996). “Rates of Convergence of the Hastings and Metropolis Algorithms.” *The Annals of Statistics*, 24: 101–121. 414
- Pace, L. and Salvan, A. (1997). *Principles of Statistical Inference*. Singapore: World Scientific. 419
- Pauli, F., Racugno, W., and Ventura, L. (2011). “Bayesian composite marginal likelihoods.” *Statistica Sinica*, 21: 149–164. 435
- Prangle, D., Blum, M. G. B., Popovic, G., and Sisson, S. A. (2013a). “Diagnostic tools of approximate Bayesian computation using the coverage property.” *arXiv preprint*, arXiv:1301.3166. 412
- Prangle, D., Fearnhead, P., Cox, M. P., Biggs, P. J., and French, N. P. (2013b). “Semi-automatic selection of summary statistics for ABC model choice.” *arXiv preprint*, arXiv:1302.5624v1. 434
- Ratmann, O., Camacho, A., Meijer, A., and Donker, G. (2014). “Statistical modelling of summary values leads to accurate Approximate Bayesian Computations.” *arXiv preprint*, arXiv:1305.4283. 412, 413, 434
- Ratmann, O., Jørgensen, O., Hinkley, T., Stumpf, M., Richardson, S., and Wiuf, C. (2007). “Using likelihood-free inference to compare evolutionary dynamics of the protein networks of *H. pylori* and *P. falciparum*.” *PLoS Computational Biology*, 3: 2266–2276. 412, 417, 418
- Ruli, E., Sartori, N., and Ventura, L. (2013). “Approximate Bayesian Computation with composite score functions.” *arXiv preprint*, arXiv:1311.7286v1. 413, 421
- Severini, T. (2002). “Modified estimating functions.” *Biometrika*, 89: 333–343. 414

- Siegmund, K. D., Marjoram, P., and Shibata, D. (2008). “Modeling DNA methylation in a population of cancer cells.” *Statistical applications in genetics and molecular biology*, 7: 1–21. [411](#)
- Sisson, S., Fan, Y., and Tanaka, M. (2007). “Sequential Monte Carlo without likelihoods.” *Proceedings of the National Academy of Sciences of the United States of America*, 104: 1760–1765. [412](#)
- Stone, C. (1985). “Additive regression and other nonparametric models.” *Annals of Statistics*, 13: 689–705. [420](#)
- 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. [411](#)
- Tavaré, S., Balding, D. J., Griffiths, R. C., and Donnelly, P. (1997). “Inferring coalescence times from DNA sequence data.” *Genetics*, 145: 505–18. [421](#), [422](#)
- Toni, T., Welch, D., Strelkowa, N., Ipsen, A., and Stumpf, M. P. H. (2009). “Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems.” *Journal of the Royal Society Interface*, 6: 187–202. [414](#), [416](#)
- Ventura, L., Cabras, S., and Racugno, W. (2010). “Default prior distributions from quasi- and quasi-profile likelihoods.” *Journal of Statistical Planning and Inference*, 140: 2937–2942. [414](#)
- Wang, M. and Hanfelt, J. J. (2003). “Adjusted profile estimating function.” *Biometrika*, 90: 845–858. [414](#)
- Wood, S. N. (2010). “Statistical inference for noisy nonlinear ecological dynamic systems.” *Nature*, 466: 1102–1104. [434](#)

### Acknowledgments

María Eugenia Castellanos was partially funded by Ministerio de Ciencia e Innovación grant MTM2013-42323. Stefano Cabras has been partially funded by Ministerio de Ciencia e Innovación grant MTM2013-42323, ECO2012-38442, RYC-2012-11455 and together with Erlis Ruli were partially funded by Ministero dell’Istruzione, dell’Università e della Ricerca of Italy. All authors have been partially financially supported by Regione Autonoma della Sardegna under grant CRP-59903.

The authors also thank Mario Pirastu and Maria Pina Concas for providing genealogy tree data.