



Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Robust approximate Bayesian inference

Erlis Ruli^{*}, Nicola Sartori, Laura Ventura

Department of Statistical Sciences, University of Padova, Italy



ARTICLE INFO

Article history:

Received 26 June 2018

Received in revised form 11 January 2019

Accepted 10 May 2019

Available online 4 June 2019

Keywords:

Influence function

Likelihood-free inference

 M -estimators

Quasi-likelihood

Robustness

Unbiased estimating function

ABSTRACT

We discuss an approach for deriving robust posterior distributions from M -estimating functions using Approximate Bayesian Computation (ABC) methods. In particular, we use M -estimating functions to construct suitable summary statistics in ABC algorithms. The theoretical properties of the robust posterior distributions are discussed. Special attention is given to the application of the method to linear mixed models. Simulation results and an application to a clinical study demonstrate the usefulness of the method. An R implementation is also provided in the `rrobustBLME` package.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

The normality assumption is the usual basis of many statistical analyses in several fields, such as medicine, health sciences, quality control and engineering statistics. Under this assumption, standard parametric estimation and testing procedures are simple and efficient. However, both from a frequentist or a Bayesian perspective, it is well known that these procedures are not robust when the normal distribution is just an approximate model or in the presence of outliers in the observed data. In these situations, robust statistical methods can be considered in order to produce statistical procedures that are stable with respect to small changes in the data or to small model departures; see [Huber and Ronchetti \(2009\)](#) for a review on robust methods.

The concept of robustness has been widely discussed in the frequentist literature; see, for instance, [Hampel et al. \(1986\)](#), [Tsou and Royall \(1995\)](#) and [Markatou et al. \(1998\)](#). Also Bayesian robustness with respect to model misspecification have attracted considerable attention. For instance, [Lazar \(2003\)](#), [Greco et al. \(2008\)](#), [Ventura et al. \(2010\)](#) and [Agostinelli and Greco \(2013\)](#) discuss approaches based on robust pseudo-likelihood functions, such as the empirical likelihood, as replacement of the genuine likelihood in Bayes' formula. [Lewis et al. \(2014\)](#) discuss an approach for building posterior distributions from robust M -estimators using constrained Markov Chain Monte Carlo (MCMC) methods. Recent approaches based on tilted likelihoods can be found in [Grünwald and van Ommen \(2017\)](#), [Watson and Holmes \(2016\)](#), [Miller and Dunson \(2018\)](#). Finally, approaches based on model embedding through heavy-tailed distributions are discussed by [Andrade and O'Hagan \(2006\)](#).

The aforementioned approaches may present some drawbacks. The empirical likelihood is not computable for small sample sizes and posterior distributions based on the quasi-likelihood can be easily obtained only for scalar parameters. The restricted likelihood approach of [Lewis et al. \(2014\)](#), as well as all the approaches based on estimating equations can be computationally cumbersome with some robust M -estimating functions (such as, for instance, those used in linear mixed

^{*} Corresponding author.

E-mail addresses: ruli@stat.unipd.it (E. Ruli), sartori@stat.unipd.it (N. Sartori), ventura@stat.unipd.it (L. Ventura).

effects models). The tilted and the weighted likelihood approaches refer to concepts of robustness that are not directly related to the one considered in this paper, which is based on the influence function (Hampel et al., 1986; Huber and Ronchetti, 2009). Finally, the idea of embedding the model in a larger structure has the cost of requiring the elicitation of a prior distribution for the extra parameters introduced. Moreover, the statistical procedures derived under an embedded model are not necessarily robust in a broad sense, since the larger model may still be too restricted.

Here we focus on the robustness approach based on the influence function and on the derivation of robust posterior distributions from robust M -estimating functions, i.e. estimating equations with bounded influence function (see, e.g., Huber and Ronchetti, 2009, Chap. 3). In particular, we propose an approach based on Approximate Bayesian Computation (ABC) methods (see, e.g., Beaumont et al., 2002) using robust M -estimating functions as summary statistics. The idea extends results of Ruli et al. (2016) on composite score functions to Bayesian robustness. The method is easy to implement and computationally efficient, even when the M -estimating functions are potentially cumbersome to evaluate. Theoretical properties, implementation details and simulation results are discussed.

The rest of the paper is structured as follows. Section 2 sets the background. Section 3 describes the proposed method and its properties. Section 4 investigates the properties of the proposed method in the context of linear mixed models through simulations and an application to a clinical study. Concluding remarks are given in Section 5.

2. Background on robust M -estimating functions

Let $y = (y_1, \dots, y_n)$ be a random sample of size n , having independent and identically distributed components, according to a distribution function $F_\theta = F(y; \theta)$, with $\theta \in \Theta \subseteq \mathbb{R}^d$, $d \geq 1$ and $y \in \mathcal{Y}$. Let $L(\theta)$ be the likelihood function based on model F_θ .

Furthermore, let

$$\Psi_\theta = \Psi(y; \theta) = \sum_{i=1}^n \psi(y_i; \theta) - c(\theta), \quad (1)$$

be an unbiased estimating function for θ , i.e. such that $E_\theta(\Psi(Y; \theta)) = 0$ for every θ . In (1), $\psi(\cdot)$ is a known function, $E_\theta(\cdot)$ is the expectation with respect to F_θ and the function $c(\cdot)$ is a consistency correction which ensures unbiasedness of the estimating function.

A general M -estimator (see, e.g., Hampel et al., 1986, Huber and Ronchetti, 2009) is defined as the root $\tilde{\theta}$ of the estimating equation $\Psi_\theta = 0$. The class of M -estimators is wide and includes a variety of well-known estimators. For example, it includes the maximum likelihood estimator (MLE), the maximum composite likelihood estimator (see, e.g., Ruli et al., 2016, and references therein) and the scoring rule estimator (see e.g. Dawid et al., 2016, and references therein). Under broad regularity conditions, assumed throughout this paper, an M -estimator is consistent and approximately normal with mean θ and variance

$$K(\theta) = H(\theta)^{-1}J(\theta)H(\theta)^{-T}, \quad (2)$$

where $H(\theta) = -E_\theta(\partial \Psi_\theta / \partial \theta^T)$ and $J(\theta) = E_\theta(\Psi_\theta \Psi_\theta^T)$ are the sensitivity and the variability matrices, respectively. The matrix $G(\theta) = K(\theta)^{-1}$ is known as the Godambe information and the form of $K(\theta)$ is due to the failure of the information identity since, in general, $H(\theta) \neq J(\theta)$.

The influence function (IF) of the estimator $\tilde{\theta}$ is $IF(x; \tilde{\theta}, F_\theta) \propto \psi(x; \theta)$ and it measures the effect on the estimator $\tilde{\theta}$ of an infinitesimal contamination at the point x , standardised by the mass of the contamination. A desirable robustness property for $\tilde{\theta}$ is that its IF is bounded (B-robustness), i.e. that $\psi(x; \theta)$ is bounded. Note that the IF of the MLE is proportional to the score function; therefore, in general, the MLE has unbounded IF , i.e. it is not B-robust.

3. Robust ABC inference

One possibility to perform robust Bayesian inference is to resort to a pseudo-posterior distribution of the form

$$\pi_R(\theta|y) \propto \pi(\theta)L_R(\theta), \quad (3)$$

where $\pi(\theta)$ is a prior distribution for θ and $L_R(\theta)$ is a pseudo-likelihood based on a robust Ψ_θ , such as the quasi- or the empirical likelihood. This approach has two main drawbacks: the empirical likelihood is not computable for very small sample sizes and for moderate sample sizes the corresponding posterior appears to have always heavy tails (see, e.g., Greco et al., 2008); moreover, the posterior distribution based on the quasi-likelihood can be easily obtained only for scalar parameters. A further limitation of this approach is related to computational cost, in the sense that it requires repeated evaluations of the consistency correction $c(\theta)$ in (1), which in practice is often cumbersome.

We propose an alternative method for computing posterior distributions based on robust M -estimating functions, extending the idea in Ruli et al. (2016). The method resorts to the ABC machinery (see, e.g., Beaumont et al., 2002) in which a standardised version of Ψ_θ , evaluated at a fixed value of θ , is used as a summary statistic. In Ruli et al. (2016) the composite score function is used as a model-based data reduction procedure for ABC in complex models. Here we generalise the approach to general unbiased robust estimating functions. In particular, let $\hat{\theta} = \hat{\theta}(y)$ be the M -estimate of

θ based on the observed sample y . Furthermore, let $B_R(\theta)$ be such that $J(\theta) = B_R(\theta)B_R(\theta)^\top$. The summary statistic in ABC is then the rescaled M -estimating function

$$\eta_R(y^*; \theta) = B_R(\theta)^{-1} \Psi(y^*; \theta), \quad (4)$$

evaluated at $\tilde{\theta}$, where y^* is a simulated sample. In the sequel we use the shorthand notation $\tilde{\eta}_R(y^*) = \eta_R(y^*; \tilde{\theta})$.

To generate posterior samples we propose to use the ABC-R algorithm with an MCMC kernel (Algorithm 1), which is similar to Algorithm 2 of Fearnhead and Prangle (2012); see also Marjoram et al. (2003). More specifically, the ABC-R algorithm (Algorithm 1) involves a kernel density $K_h(\cdot)$, which is governed by the bandwidth $h > 0$ and a proposal density $q(\cdot|\cdot)$; see Appendix for the implementation details.

Result: A Markov dependent sample $(\theta^{(1)}, \dots, \theta^{(m)})$ from $\pi_R^{ABC}(\theta|\tilde{\theta})$

Data: a starting value $\theta^{(0)}$, a proposal density $q(\cdot|\cdot)$

```

for  $i = 1 \rightarrow m$  do
  draw  $\theta^* \sim q(\cdot|\theta^{(i-1)})$ 
  draw  $y^* \sim F_{\theta^*}$ 
  draw  $u \sim U(0, 1)$ 
  if  $u \leq \frac{K_h(\tilde{\eta}_R(y^*))}{K_h(\tilde{\eta}_R(y^{(i-1)}))} \frac{\pi(\theta^*)q(\theta^{(i-1)}|\theta^*)}{\pi(\theta^{(i-1)})q(\theta^*|\theta^{(i-1)})}$  then
    | set  $(\theta^{(i)}, \tilde{\eta}_R^{(i)}) = (\theta^*, \tilde{\eta}_R(y^*))$ 
  else
    | set  $(\theta^{(i)}, \tilde{\eta}_R^{(i)}) = (\theta^{(i-1)}, \tilde{\eta}_R(y^{(i-1)}))$ 
  end
end

```

Algorithm 1: ABC-R algorithm with MCMC.

The proposed method gives Markov-dependent samples from the ABC-R posterior

$$\pi_R^{ABC}(\theta|\tilde{\theta}) = \frac{\int_{\mathcal{Y}^*} \pi(\theta) f(y^*; \theta) K_h(\tilde{\eta}_R(y^*)) dy^*}{\int_{\mathcal{Y}^* \times \Theta} \pi(\theta) f(y^*; \theta) K_h(\tilde{\eta}_R(y^*)) dy^* d\theta}. \quad (5)$$

While Algorithm 1 or the use of a kernel in (5) is not a new idea in the ABC literature, the novelty here is to incorporate in such machinery the robust summary statistic $\tilde{\eta}_R(y^*)$ in order to obtain a simulated sample from a robust posterior distribution. Using similar arguments to Soubeyrand et al. (2013), it can be shown that, for $h \rightarrow 0$, $\pi_R^{ABC}(\theta|\tilde{\theta})$ converges to $\pi(\theta|\tilde{\theta})$ pointwise (see also Blum, 2010), in the sense that $\pi_R^{ABC}(\theta|\tilde{\theta})$ and $\pi(\theta|\tilde{\theta})$ are equivalent for sufficiently small h . Since in general (4) does not give a sufficient summary statistic, then $\pi(\theta|\tilde{\theta})$ differs from $\pi(\theta|y)$ and information is lost by using (4) instead of y . However this difference pays off in terms of robustness in inference about θ .

Posteriors conditional on partial information have been extensively discussed in the literature. Soubeyrand and Haon-Lasportes (2015) study the properties of the ABC posterior when the summary statistic is the MLE or the pseudo-MLE derived from a simplified parametric model. An alternative version of the ABC-R algorithm could be based directly on $\tilde{\theta}$, used as the summary statistic and a, possibly rescaled, distance among the observed and the simulated value of the statistic. Apparently, these two versions of ABC, namely the one based on $\tilde{\theta}$ and that based on (4) seem to be treated in the literature as two separate approaches (see, e.g., Drovandi et al., 2015). However, both alternatives use essentially the same information, i.e. $\tilde{\theta}$, but through different distance metrics. In addition, for small tolerance levels, these two distances converge to zero, and both methods give a posterior distribution conditional on the same statistic $\tilde{\theta}$. Indeed, let $\tilde{\theta}$ be the summary statistic of the ABC posterior and let the corresponding tolerance threshold ϵ be sufficiently small and consider the random draw θ^* and its corresponding simulated summary statistics $\tilde{\theta}^*$ taken with the ABC algorithm. Then, by construction $\tilde{\theta}^*$ will be close to $\tilde{\theta}$. This implies that also $\tilde{\eta}_R(y^*) = \eta_R(y^*; \tilde{\theta})$ will be close to $\eta_R(y^*; \tilde{\theta}^*) = 0$, and hence θ^* is also a sample from the ABC-R posterior which uses the summary statistic $\tilde{\eta}_R$.

Nevertheless, the use of $\tilde{\theta}$ as summary statistic requires the solution of $\Psi_\theta = 0$ at each iteration of the algorithm, which could be computationally cumbersome. On the contrary, the proposed approach, besides sharing the same invariance properties stated by Ruli et al. (2016), i.e. invariance with respect to both monotonic transformation of the data and with respect to reparameterisations, has the advantage of avoiding computational problems related to the repeated evaluation of Ψ_θ as shown by the following lemma.

Lemma 3.1. *The ABC-R algorithm does not require repeated evaluations of the consistency correction $c(\theta)$ involved in Ψ_θ , as given by (1).*

Proof. Let $\tilde{\theta}$ be the solution of $\Psi_\theta = 0$, with Ψ_θ of the form (1). Then, for a given simulated y^* from F_{θ^*} , we have

$$\tilde{\eta}_R(y^*) = B_R(\tilde{\theta})^{-1}(\Psi(y^*; \tilde{\theta}) - \Psi(y; \tilde{\theta})) = \sum_{i=1}^n (\psi(y_i^*, \tilde{\theta}) - \psi(y_i, \tilde{\theta})).$$

This implies that $c(\theta)$ is computed only once, at $\tilde{\theta}$.

Theorem 3.1 shows that the proposed method gives a robust approximate posterior distribution with the correct curvature, even though Ψ_θ , unlike the full score function, does not satisfy the information identity. Here, correct curvature means that asymptotically the robust posterior distribution and its normal approximation have the same covariance matrix, which is the inverse of the Godambe information, i.e. $K(\theta)$.

Theorem 3.1. *The ABC-R algorithm with rescaled M-estimating function $\tilde{\eta}_R(y)$ as summary statistic, as $h \rightarrow 0$, leads to an approximate posterior distribution with the correct curvature and is also invariant to reparameterisations.*

Proof. The proof follows from Theorem 3.2 of Ruli et al. (2016), by substituting the composite estimating equation with the more general M-estimating function Ψ_θ .

The ABC-R algorithm delivers thus a robust approximate posterior distribution which does not need calibration. On the contrary, for (3) a calibration is typically required.

Theorem 3.2 shows that the proposed ABC posterior distribution is asymptotically normal.

Theorem 3.2. *Assume the regularity assumptions of Soubeyrand and Haon-Lasportes (2015) and the usual regularity condition on M-estimators (Huber and Ronchetti, 2009, Chap. 4) are satisfied. Then, for $n \rightarrow \infty$ and $h \rightarrow 0$, the posterior $\pi_R^{ABC}(\theta|\tilde{\theta})$ is asymptotically equivalent to the density of the normal distribution with mean vector $\tilde{\theta}$ and covariance matrix $K(\tilde{\theta})$:*

$$\pi_R^{ABC}(\theta|\tilde{\theta}) \sim N_d(\tilde{\theta}, K(\tilde{\theta})). \tag{6}$$

Proof. The proof follows from Lemma 2 and Theorem 1 in Soubeyrand and Haon-Lasportes (2015) and from the asymptotic relation between the Wald-type statistic and the score-type statistic, i.e.

$$\eta_R(y; \theta)^T \eta_R(y; \theta) = \Psi_\theta^T J(\theta)^{-1} \Psi_\theta = (\tilde{\theta} - \theta)^T K(\theta)^{-1} (\tilde{\theta} - \theta) + o_p(1).$$

If $\psi(y; \theta)$ is bounded in y , i.e. if the estimator $\tilde{\theta}$ is B-robust, then the ABC-R posterior is resistant with respect to slight violations of model assumptions. More precisely, the following theorem shows that the ABC-R posterior inherits the robustness properties of the estimating equation.

Theorem 3.3. *If $\psi(y; \theta)$ is bounded in y , i.e. if the estimator $\tilde{\theta}$ is B-robust, then asymptotically the posterior mode, as well as other posterior summaries of $\pi_R^{ABC}(\theta|\tilde{\theta})$ have bounded IF.*

Proof. From Theorem 3.2, the asymptotic posterior mode of $\pi_R^{ABC}(\theta|\tilde{\theta})$ is $\tilde{\theta}$, which is B-robust. Moreover, following results in Greco et al. (2008), it can be shown that asymptotic posterior summaries have bounded IF if and only if the posterior mode has bounded IF.

Example. We consider an illustrative example in which we compare numerically the ABC-R posterior, with the classical posterior based on the assumed model and the pseudo-posterior (3) based on the empirical likelihood (Lazar, 2003; Greco et al., 2008). Scenarios with data simulated either from the assumed model or from a slightly misspecified model are considered.

Let F_θ be a location-scale distribution with location μ and scale $\sigma > 0$, and let $\theta = (\mu, \sigma)$. The Huber’s estimating function is a standard choice for robust estimation of location and scale parameters. The M-estimating function is $\Psi_\theta = (\Psi_\mu, \Psi_\sigma)$, with

$$\Psi_\mu = \sum_{i=1}^n \psi_{c_1}(z_i) \quad \text{and} \quad \Psi_\sigma = \sum_{i=1}^n (\psi_{c_2}(z_i)^2 - k(c_2)), \tag{7}$$

where $z_i = (y_i - \mu)/\sigma$, $i = 1, \dots, n$, $\psi_c(z) = \max[-c, \min(c, z)]$ is the Huber ψ -function, $c > 0$ is a scalar tuning constant which controls the desired degree of robustness of θ , and $k(\cdot)$ is a consistency correction term. Let F_θ be the normal distribution $N(\mu, \sigma^2)$ and assume μ and σ a priori independent with $\mu \sim N(0, 10^2)$ and $\sigma \sim \text{halfCauchy}(5)$, where $\text{halfCauchy}(a)$ is the halfCauchy distribution with scale parameter equal to a . We consider random samples of sizes $n = \{15, 30\}$ drawn from either the normal distribution with $\theta = (0, 1)$ and from a contaminated model $(1 - \delta)N(0, 1) + \delta N(0, \sigma_1^2)$, with $\sigma_1^2 > 0$. We set the contamination level equal to 10%, i.e. $\delta = 0.1$, and $\sigma_1^2 = 10$. Moreover, we fix $c_1 = 1.345$ and $c_2 = 2.07$, which imply that $\tilde{\mu}$ and $\tilde{\sigma}$ are, respectively, 5% and 10% less efficient than the corresponding MLE under the assumed model (see Huber and Ronchetti, 2009, Chap. 6).

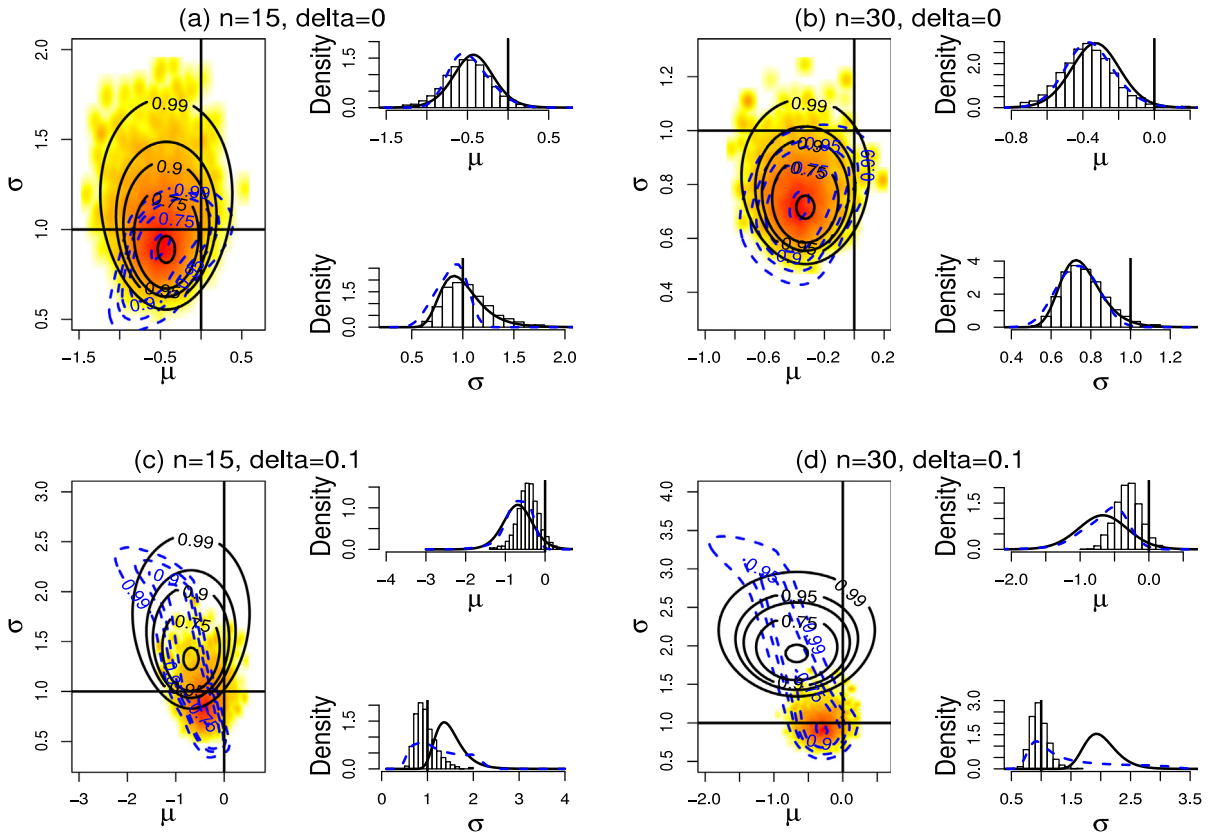


Fig. 1. First row: genuine (black solid), EL (blue dashed) and ABC-R posteriors (shaded image and histogram) for the normal model, when the data come from the central model $N(0, 1)$ with (a) $n = 15$ and (b) $n = 30$. Second row: genuine, EL and ABC-R posteriors for the normal model, when the data come from the contaminated model with $\delta = 0.1$, (c) $n = 15$ and (d) $n = 30$.

The genuine, e.g. the posterior based on the likelihood function of the normal model, and the pseudo-posterior (3) based on the empirical likelihood (EL) are computed by numerical integration. The ABC-R posterior is obtained using Algorithm 1. From the posterior distributions illustrated in Fig. 1 we note that, when the data come from the central model (panels (a)–(b)), i.e. for $\delta = 0$, all the posteriors are in reasonable agreement, even if the EL posterior behaves slightly worse, especially the marginal posterior of σ with $n = 15$. When the data are contaminated (panels (c)–(d)), the genuine posterior is less trustworthy as the bulk of the posterior drifts away from the true parameter value (vertical and horizontal straight lines). This is not the case however for the ABC-R posterior which remains centred around the true parameter value. We note that in the contaminated case, the ABC-R posterior is the one with smaller variability. This is due to the fact that the ABC-R posterior is not affected by the very outlying observations coming from the contamination component.

To highlight the robustness properties of the ABC-R posterior, we consider a sensitivity analysis. A sample y of size $n = 31$ is taken from the central model and the aforementioned posteriors are computed from the contaminated data y^w given by the original data with the median observation $y_{(n+1)/2}$ replaced by $y_{(n+1)/2} + w$; w is a contamination scalar with possible values $\{-15, -14, \dots, 15\}$. The results of the sensitivity analysis, illustrated by means of violin plots in Fig. 2, highlight that the posterior median of the genuine posterior (panel (c)) is substantially driven by w . On the other hand, ABC-R and EL posteriors are robust. For all posteriors, the behaviour of the posterior median reflects the behaviour of the IF of the posterior mode. Furthermore, the variability of all posteriors is comparable for values of w close to 0. More generally, these plots confirm that the genuine and EL posteriors under contamination are much more dispersed than the ABC-R posterior.

4. Application to linear mixed models

Linear mixed models (LMM) are a popular choice when analysing data in the context of hierarchical, longitudinal or repeated measures. A general formulation is

$$y = X\alpha + \sum_{i=1}^{c-1} Z_i\beta_i + \varepsilon, \tag{8}$$

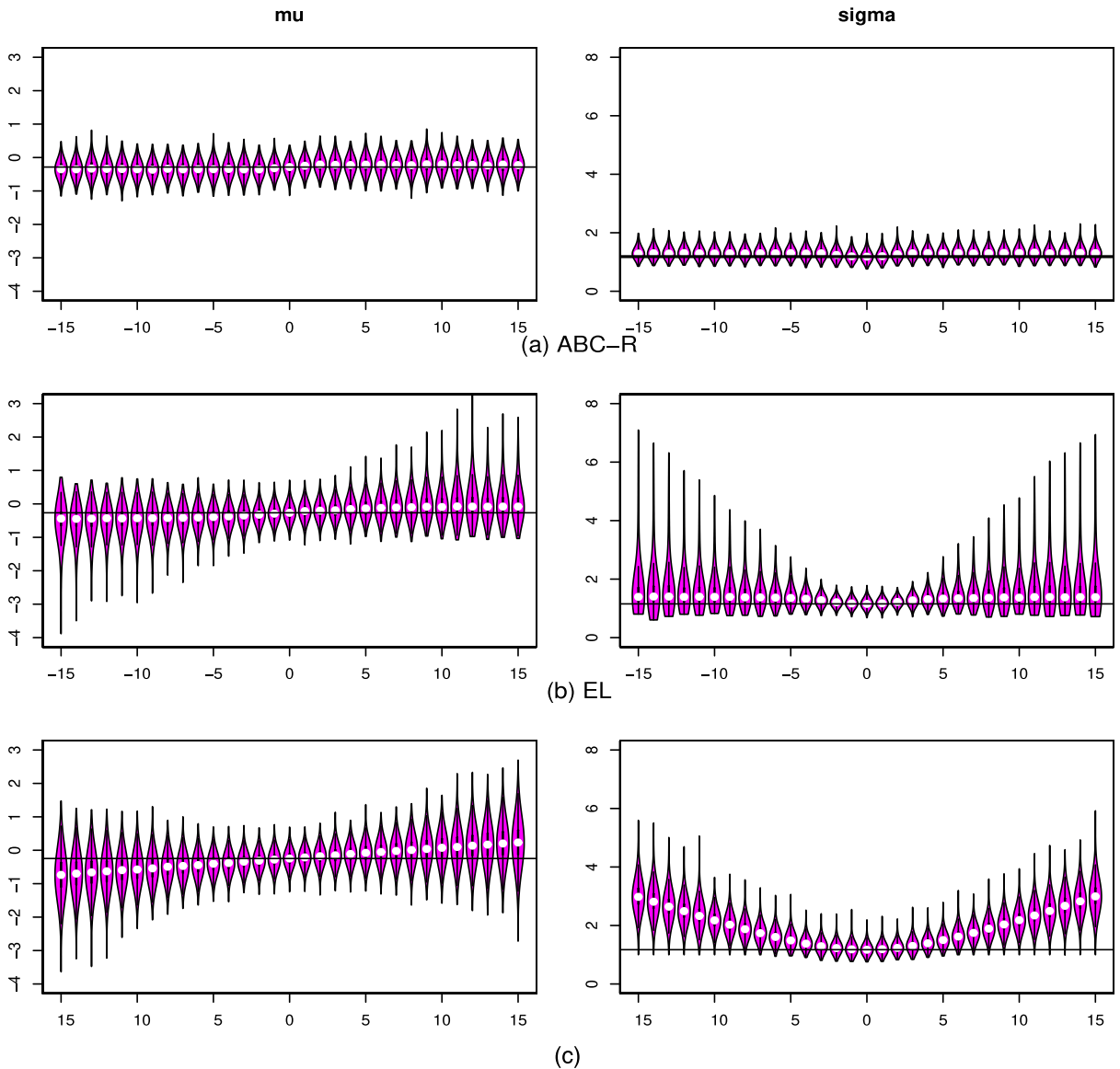


Fig. 2. Sensitivity analysis for marginal ABC-R (a), EL (b) and genuine (c) posteriors for μ (left columns) and σ (right) represented by means of violin plots. For each violin plot, the central circle represents the posterior median. The horizontal lines denote the corresponding posterior medians under y^w with $w = 0$.

where y is a n -dimensional vector of response observations, X and Z_i are known $n \times q$ and $n \times p_i$ design matrices, α is a q -vector of unknown fixed effects, the β_i are p_i -vectors of unobserved random effects ($1 \leq i \leq c - 1$) and ε is a vector of unobserved errors. The p_i levels of each random effect β_i are assumed to be independent with mean zero and variance σ_i^2 . Moreover, each random error ε_i is assumed to be independent with mean zero and variance σ_c^2 and $\beta_1, \dots, \beta_{c-1}$ and ε are assumed to be independent.

Here we focus on the classical normal LMM, which assumes that $\varepsilon \sim N_n(0_n, \sigma_c^2 I_n)$ and $\beta_i \sim N(0, \sigma_i^2)$, $i = 1, \dots, c - 1$. For a normal LMM, it follows that Y is multivariate normal with $E(Y) = X\alpha$ and $\text{var}(Y) = V = \sum_{i=1}^c \sigma_i^2 Z_i Z_i^T$, where $Z_c = I_n$. We assume that the set of $d = q + c$ unknown parameters $\theta = (\alpha, \sigma^2) = (\alpha, \sigma_1^2, \dots, \sigma_c^2)$ is identifiable. The validity and performance of this LMM requires strict adherence to the assumed model, which is usually chosen because it simplifies the analyses and not because it fits exactly the data at hand. The robust procedure discussed in this paper specifically takes into account the fact that the normal model is only approximate and then it produces statistical analyses that are stable with respect to outliers, deviations from the model or model misspecifications.

Although the n observations y are not independent, if the random effects are nested, then independent subgroups of observations can be found. Indeed, in many situations, y can be split into g independent groups of observations y_j ,

$j = 1, \dots, g$, and the log-likelihood is

$$\ell(\theta) = \log L(\theta) = -\frac{1}{2} \sum_{j=1}^g \{ \log |V_j| + (y_j - X_j \alpha)^T V_j^{-1} (y_j - X_j \alpha) \}, \tag{9}$$

where (y_1, \dots, y_g) and X and V are partitioned accordingly. Classical Bayesian inference for θ is based on $\pi(\theta|y) \propto L(\theta) \pi(\theta)$, where $\pi(\theta)$ is a prior distribution for θ . However, (9) can be very sensitive to model deviations (Richardson and Welsh, 1995; Richardson, 1997; Copt and Victoria-Feser, 2006); see also results of the simulation study in Section 4.1.

In the frequentist literature, there are two broad classes of estimators for robust estimation of Gaussian LMM: M -estimators (see, e.g., Richardson and Welsh, 1995; Richardson, 1997, and references therein) and S -estimators (Copt and Victoria-Feser, 2006). The latter are generally available for balanced designs whereas the former can be applied to a wide variety of situations; for instance it can deal with unbalanced designs and robustness with respect to the design matrix (Richardson, 1997). In this work we focus on M -estimators but it is worth stressing that the idea can be applied to S -estimators as well. Following Richardson and Welsh (1995), we focus on the system of M -estimating equations

$$X^T V^{-1/2} \psi_{c_1}(r) = 0, \tag{10}$$

$$\psi_{c_2}(r)^T V^{-1/2} Z_i Z_i^T V^{-1/2} \psi_{c_2}(r) - \text{tr}(CPZ_i Z_i^T) = 0, \quad i = 1, \dots, c, \tag{11}$$

where $r = V^{-1/2}(y - X\alpha)$ is the vector or scaled marginal residuals, $C = E_\theta [\psi_{c_2}(R)\psi_{c_2}(R)^T]$, with $R = V^{-1/2}(Y - X\alpha)$, $P = V^{-1} - V^{-1}X(X^T V^{-1}X)^{-1}X^T V^{-1}$ and $\text{tr}(\cdot)$ is the trace operator. The function $\text{tr}(CPZ_i Z_i^T)$ is a correction factor needed to ensure consistency at the Gaussian model for each $i = 1, \dots, c$. Eqs. (10)–(11) are called robust REML II estimating equations and are bounded versions of restricted likelihood equations. Richardson (1997) shows that the M -estimator based on (10)–(11) is asymptotically normal with mean equal to the true parameter θ and covariance matrix of the form (2). The ABC-R procedure in the normal LMM based on (10)–(11) will be studied by means of simulations in Section 4.1 and then applied to a dataset from a clinical study in Section 4.2.

4.1. Simulation study

Let us consider the two-component nested model

$$y_{ij} = \mu + \alpha_j + \beta_i + \varepsilon_{ij}, \tag{12}$$

where μ is the grand mean, α_j are the fixed effects, constrained such that $\sum_{j=1}^q \alpha_j = 0$, $\beta_i \sim N(0, \sigma_1^2)$ are the random effects and $\varepsilon_{ij} \sim N(0, \sigma_2^2)$ is the residual term, for $j = 1, \dots, q$ and $i = 1, \dots, g$. Model (12) is a particular case of (8) with $c = 2$, a single random effect β_1 with $p_1 = g$ levels and Z_1 the unit diagonal matrix. Moreover, the covariate is a categorical variable with q levels; hence the design matrix is given by $q - 1$ dummy variables.

We assess the properties of the proposed method via simulations with 500 Monte Carlo replications. For each Monte Carlo replication, the true values for (σ_1^2, σ_2^2) and for α are drawn uniformly in $(1, 10) \times (1, 10)$ and $(-5, 5)$, respectively. With these values, two datasets of size g are generated: one from the central model and one from the contaminated model $(1 - \delta)N(X_i^T \alpha, V_i) + \delta N(X_i^T \alpha, 15V_i)$, where X_i is the matrix of covariates for the i th unit, $\theta = (\alpha, \sigma_1^2, \sigma_2^2)$ and $\delta = 0.10$. We consider $q = \{3, 5, 7\}$ and $g = \{30, 50, 70\}$. The prior distributions are $\alpha \sim N_q(0, 10^2 I_q)$ and $(\sigma_1^2, \sigma_2^2) \sim \text{halfCauchy}(7) \times \text{halfCauchy}(7)$. For each scenario, we fit model (12) in the classical Bayesian way, using an adaptive random walk Metropolis–Hastings algorithm. The same model is fitted by the ABC-R method using the estimating equations (10)–(11). As in Richardson and Welsh (1995), we set $c_1 = 1.345$ and $c_2 = 2.07$ and we find $\tilde{\theta}$ solving (10)–(11) iteratively until convergence. The classical REML estimate, computed by the function `lmer` of the `lme4` package, is used as starting value. In our experiments, the convergence of the solution is quite rapid, i.e. $\tilde{\theta}$ stabilises within 10–15 iterations.

We assess the component-wise bias of the posterior median $\tilde{\theta}_m$ by the modulus of $\tilde{\theta}_m - \theta_0$ in logarithmic scale, where θ_0 is the true value. Moreover, the efficiency of the classical Bayesian estimator relative to the ABC-R estimator is assessed through the index MD_{MCMC}/MD_{ABC} , where $MD = \text{med}(|\tilde{\theta}_m - \theta_0|)$; see Richardson and Welsh (1995) and Copt and Victoria-Feser (2006). In addition, for each Monte Carlo replication we compute the Euclidean distance of $\tilde{\theta}_m$ from θ_0 , which can be considered as a global measure of bias. Contrary to Richardson and Welsh (1995), we consider a different θ_0 for each Monte Carlo replication. The bias and efficiency of the classical Bayesian posterior and of the ABC-R posterior for the 500 replications are illustrated in Figs. 3 and 4, respectively.

Under the central model, inference with the ABC-R and the classical Bayesian posteriors is roughly similar, i.e. both bias and efficiency compare equally well across the two methods. This holds both for the fixed effects α and for the variance components (σ_1^2, σ_2^2) . Under the contaminated model, we notice important differences among ABC-R and the classical Bayesian estimation. In particular, $\tilde{\theta}_m$ based on ABC-R is less biased, both globally and on a component by component basis, and more efficient. The gain in efficiency is particularly evident for the variance components.



Fig. 3. Bias of the ABC-R and classical (MCMC) Bayesian estimation of LMM under either the central (Full) or the contaminated model (Mix) for varying g and q . Rows refer to a parameter or combination of parameters (row all_par); columns within each cell refer to different values of q ; e.g. the last two rows (starting from top) have only two boxplots since α_6 and α_7 are available only with $q = 7$.

4.2. Effects of GRP94-based complexes on IL-10

The GRP94 dataset (Tramentozzi et al., 2016) concerns the measurement of glucose-regulated protein94 in plasma or other biological fluids and the study of its role as a tumour antigen, i.e. its ability to alter the production of immunoglobulines (IgGs) and inflammatory cytokines in the peripheral blood mononuclear cells (PBMCs) of tumour patients. The study involved 27 patients admitted to the division of General Surgery of the Civil Hospital of Padova for ablation of primary, solid cancer of the gastro-intestinal tract. For each patient, gender, age (expressed in years), type and stage of tumour (ordinal scales of four levels) are given. Patients' plasma and PBMCs were challenged with GRP94

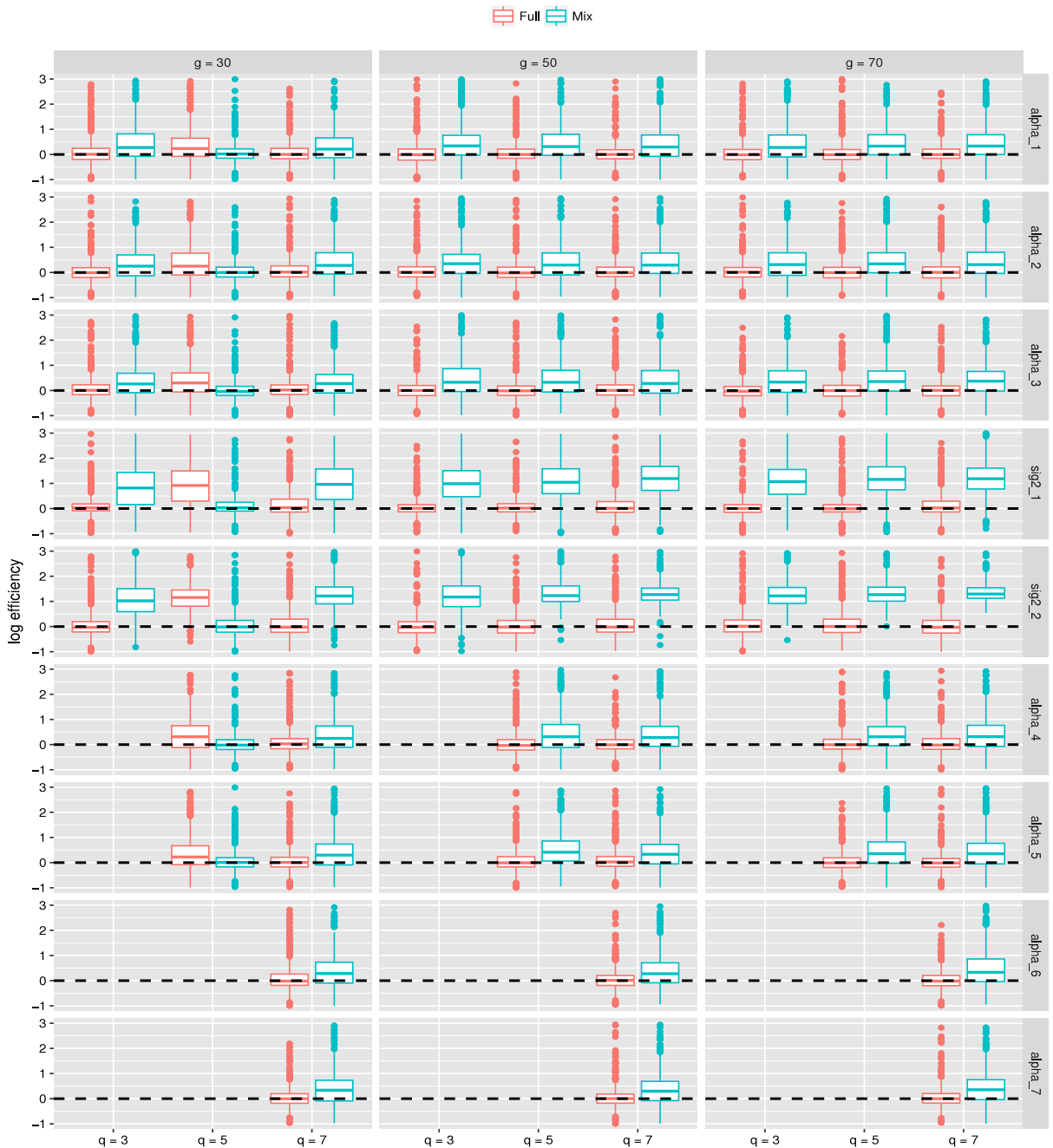


Fig. 4. Efficiency of the ABC-R compared to the classical Bayesian estimation of LMM under the central (Full) and the contaminated models (Mix) for varying g and q . Rows refer to a parameter and columns within each cell refer to different values of q ; e.g. the last two rows (starting from top) have only two boxplots since α_6 and α_7 are available only with $q = 7$.

complexes and the level of IgG and of the cytokines: interferon γ (IFN γ), interleukin 6 (IL-6), interleukin 10 (IL-10) and tumour necrosis factor α (TNF α) were measured. Owing to time and cost constraints, for patients IDs 17, 27 and 28 only IgG was measured. The following five treatments were considered: GRP94 at the dose of either 10 ng/ml or 100 ng/ml, GRP94 in complex with IgG (GRP94+IgG) at the doses 10 ng/ml or 100 ng/ml and IgG at the dose 100 ng/ml. Finally, baseline measurements of IgG and of the aforementioned cytokines were taken from untreated PMBCs. Although fresh patient's plasma and PMBCs are taken for each treatment and patient, the resulting measures are likely to be correlated since plasma and PMBCs are taken from the same patient. Hence, a LMM can be suitable for these data. Using paired Mann–Whitney tests, [Tramentozzi et al. \(2016\)](#) show that GRP94 in complex with IgG at the higher dose can significantly inhibit

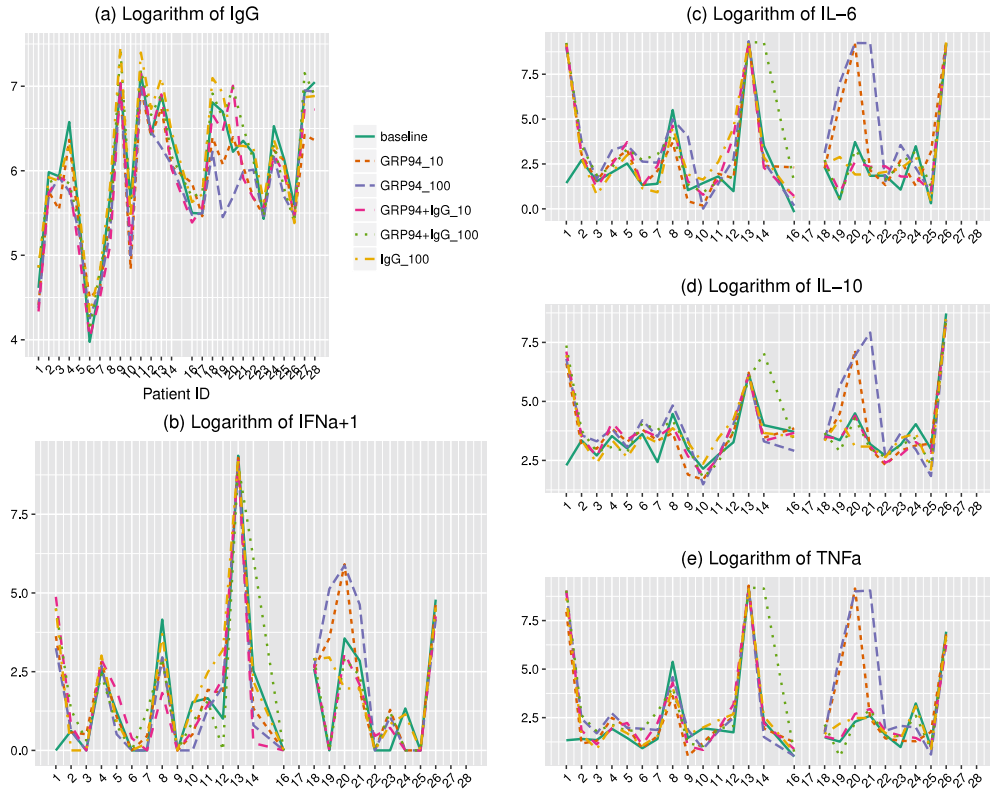


Fig. 5. Strip plots of IgG, IFN γ , IL-6, IL-10 and TNF α (in logarithmic scale) measured from PBMCs at baseline and after challenging with complexes of GRP94 and IgG. Values on the horizontal axis are (arbitrarily) ordered according to patient ID. Patient ID 15 was removed for clinical reasons and cytokines' measurements for patients with ID 17, 27 and 28 are missing.

the production of IgG, whereas GRP94 at both doses can stimulate the secretion of IL-6 and TNF α from PBMCs of cancer patients. In addition, some of the differences between treatments were significant for a specific gender; see [Tramontozzi et al. \(2016\)](#) for full details.

A feature of these data is the presence of extreme observations, both at baseline and challenged PMBCs-based measurements, as it can be seen from the strip plots in [Fig. 5](#). Such extreme observations induce high variability on the response measurements, especially for IFN γ , IL-6, IL-10 and TNF α . Hence, one must be cautious when fitting a LMM to such data.

We fit the two-component nested LMM (12) to the IL-10 with ABC-R using estimating equations (10)–(11). Since all measures are positive and some of them are highly skewed, a logarithmic transformation is used in order to alleviate distributional skewness. Furthermore, since [Tramontozzi et al. \(2016\)](#) highlight a possible gender effect (especially with respect to the cytokines) we also check for gender effects by including an interaction with gender. The model with interaction is

$$y_i = X_i^T \alpha + X_i^T \times w_i \gamma + \beta_i 1_6 + \varepsilon_i, \quad i = 1, \dots, 24, \tag{13}$$

where w_i is a dummy variable for gender, γ is the fixed effect of the treatment–gender interaction, and 1_6 is the unit vector of dimension 6. The interaction model (13) has 12 unknown fixed effects (α, γ).

As in this case there is no extra-experimental information, we assume vague priors. In particular, $\alpha_j \sim N(0, 100)$ and $\gamma_j \sim N(0, 100)$, for $j = 1, \dots, 6$. For the variance components, following [Gelman \(2006\)](#), we assume $\sigma_1^2 \sim \text{halfCauchy}(7)$ and $\sigma_2^2 \sim \text{halfCauchy}(7)$ in both models. However, we note that one of the features of the proposed method is the simultaneous ability to have robustness to possible model misspecification and to include prior information on model parameters, if available.

ABC-R posterior samples are drawn using [Algorithm 1](#). For comparison purposes, we fit also a classical Bayesian LMM with the aforementioned prior and an adaptive random walk Metropolis–Hastings algorithm is used for sampling from this posterior. [Fig. 6](#) compares the ABC-R and the classical posterior for a subset of the fixed effects of models (12) and (13) by means of kernel density estimations. The parameters shown are those referring to the treatments based on GRP94 at the dose of 10 ng/ml (GRP94_10), GRP94 at the dose of 100 ng/ml (GRP94_100) and GRP94 in complex with IgG at the dose of 100 ng/ml (GRP94+IgG_100), which according to [Tramontozzi et al. \(2016\)](#) are the most prominent. The

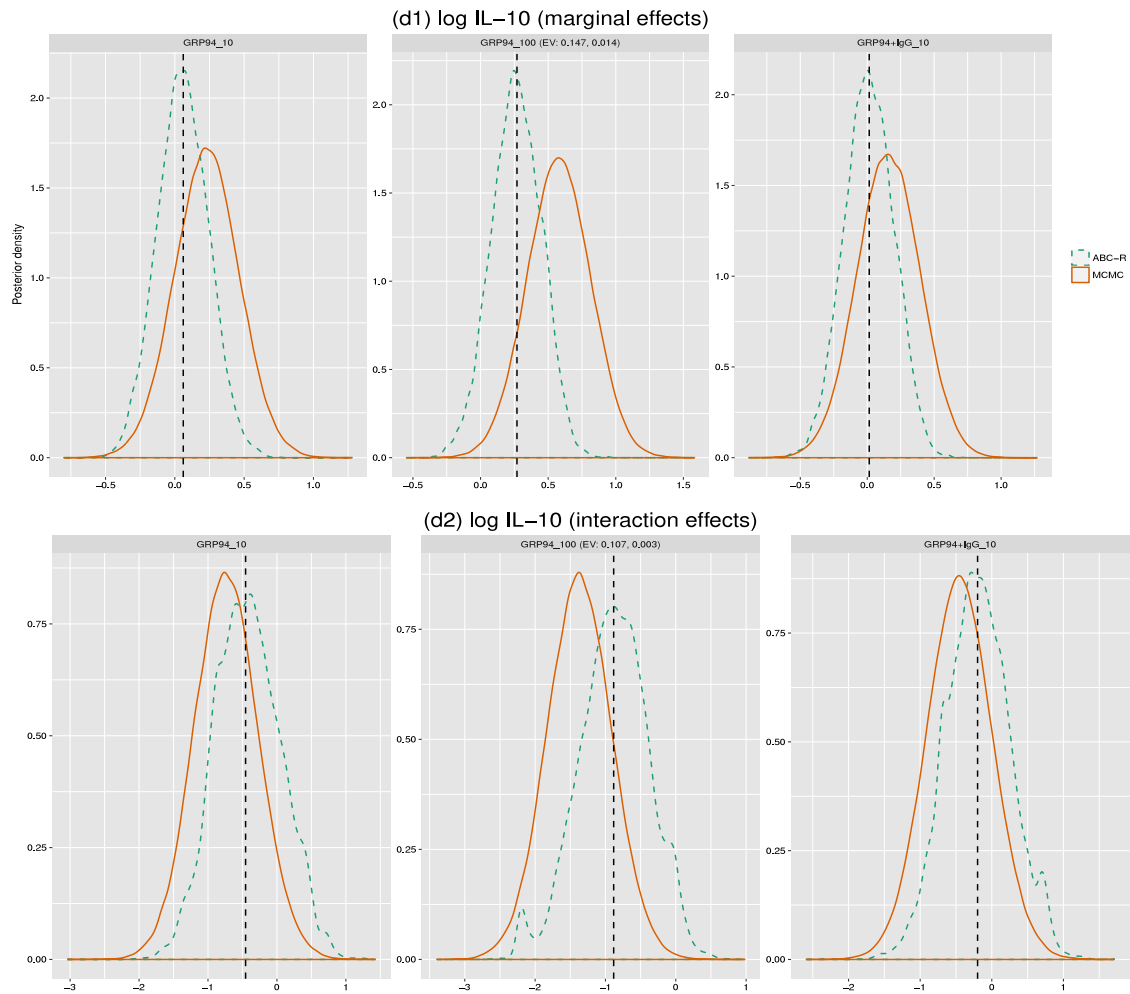


Fig. 6. Comparison of robust (ABC-R) and full (MCMC) posterior distributions of the fixed effects of the LMM without interaction with gender (12) and with interaction (13), fitted to the log IL-10. The first row refers to the posterior of the effects of the treatments against the baseline without interaction; the second refers to the posterior considering interactions of the treatments with gender (with baseline and female being the reference categories). Numbers within parenthesis refer to the FBST evidence in favour of H_0 that the parameter is equal to zero, computed under the Full Bayesian Significance Testing (FBST) setting of Pereira et al. (2008); inside the parenthesis, the first (last) value from left refers to the ABC-R (classical) posterior. Dashed vertical lines correspond to components of θ .

first row (d1) illustrates the marginal posteriors of the parameters of (12) (with baseline being the reference category). The second row (d2) shows the marginal posteriors of the parameters of (13) (with baseline and female being the reference categories). Numbers within parenthesis in the plot subtitles give the evidence in favour of the null hypothesis H_0 that the parameter is equal to zero, computed under the Full Bayesian Significance Testing (FBST) setting of Pereira et al. (2008); inside the parenthesis, the first (last) value from left refers to the ABC-R (classical) posterior.

The FBST in favour of H_0 has been proposed by Pereira and Stern (1999) as an intuitive measure of evidence, defined as the posterior probability related to the less probable points of the parametric space. It favours H_0 whenever it is large and it is based on a specific loss function and thus the decision made under this procedure is the action that minimises the corresponding posterior risk (Pereira et al., 2008). The FBST solves the drawback of the usual Bayesian procedure for testing based on the Bayes factor (BF), that is, when the null hypothesis is precise and improper or vague priors are assumed, the BF can be undetermined and it can lead to the so-called Jeffreys–Lindley paradox.

There is a high posterior probability that the effect of GRP94_100 with or without interaction with gender is different from the baseline, since the evidence of H_0 is rather low under the classical Bayesian LMM. However, such effects vanish under the robust ABC-R procedure. This is an indication to the fact that the classical LMM posterior in the case of log IL-10 is likely to be driven by few extreme observations.

5. Discussion

Currently, the only available approach for obtaining posterior distributions explicitly using robust unbiased estimating functions is through pseudo-likelihood methods such as the empirical or the quasi-likelihood (Greco et al., 2008). Bissiri

et al. (2016) show how robust posterior distribution can be based on generic loss functions, in some special cases derived from robust estimating equations. In this work, we present an alternative approach that directly incorporates robust estimating functions into approximate Bayesian computation techniques. With respect to available approaches based on pseudo-likelihoods, our method can be computationally faster when the evaluation of the estimating function is expensive.

Motivated by the GRP94 dataset, we focused on two-component nested LMM, but more complex models can be fitted since the estimating equations (10)–(11) are very general (see Richardson, 1997). For instance, it is possible to deal with models with multiple random effects or even with robustness with respect to the design matrix. An R implementation of the proposed method is provided in the `robustBLME` package (Ruli et al., 2018).

The proposed method can be applied to any unbiased robust estimating equations, such as S -estimating equations. The study of the proposed approach with S -estimating in the proposed approach is left for future work.

From a practical perspective we recommend to fit both classical and robust LMMs and compare their posteriors, say by FSBT. If the differences are mild then the posterior is probably not impacted by outliers so the classical LMM can be safely used. On the contrary, if there are important differences between them, then it is likely that the LMM posterior is driven by outliers and therefore the robust posterior would be a safer choice.

Acknowledgements

This work was partially supported by University of Padova, Italy (Progetti di Ricerca di Ateneo 2015, CPDA153257) and by the Italian Ministry of Education under the PRIN 2015 grant (2015EASZFS_003).

Appendix. Computational details

Provided simulation from F_{θ} is fast, the main demanding requirement of the proposed method is essentially the computation of the observed $\tilde{\theta}$ and the scaling matrix $B_R(\theta)$ evaluated at $\tilde{\theta}$. Given that, for large sample sizes,

$$\eta_R(y; \theta) \sim N_d(0_d, I_d),$$

where 0_d is a d -vector of zeros and I_d is the identity matrix of order d , it is reasonable to replace $K_h(\cdot)$ with the multivariate normal density centred at zero and with covariance matrix hI_d . In order to choose the bandwidth h we consider several pilot runs of the ABC-R algorithm for a grid of h values, and select the value of h that delivers approximately 0.1% acceptance ratio (as done, for instance, by Fearnhead and Prangle, 2012).

Contrary to other ABC-MCMC algorithms in which the proposal requires pilot runs (see, Cabras et al., 2015, for building proposal distributions in ABC-MCMC), in our case a scaling matrix for the proposal $q(\cdot|\cdot)$ can be readily build, almost effortlessly, by using the usual sandwich formula (2) evaluated at $\tilde{\theta}$ (see also Ruli et al., 2016). Even in cases in which $H(\theta)$ and $J(\theta)$ are not analytically available, they can be straightforwardly estimated via simulation. Indeed, in our experience, 100–500 samples from the model $F_{\tilde{\theta}}$, give estimates with reasonably low Monte Carlo variability (see also Cattelan and Sartori, 2015). Throughout the examples considered we use the multivariate t -density with 5 degrees of freedom as the proposal density $q(\cdot|\cdot)$ and the ABC-R is always started from $\tilde{\theta}$. In the ABC algorithm, we fix the tolerance threshold in order to give a pre-specified but small acceptance ratio, as frequently done in the ABC literature. In our experimentations we found that an acceptance value of 0.1% gives satisfactory results.

References

- Agostinelli, C., Greco, L., 2013. A weighted strategy to handle likelihood uncertainty in Bayesian inference. *Comput. Statist.* 28, 319–339.
- Andrade, J.A.A., O'Hagan, A., 2006. Bayesian robustness modeling using regularly varying distributions. *Bayesian Anal.* 1, 169–188.
- Beaumont, M.A., Zhang, W., Balding, D.J., 2002. Approximate Bayesian computation in population genetics. *Genetics* 162, 2025–2035.
- Bissiri, P.G., Holmes, C.C., Walker, S.G., 2016. A general framework for updating belief distributions. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 78, 1103–1130.
- Blum, M.G.B., 2010. Approximate Bayesian computation: a nonparametric perspective. *J. Amer. Statist. Assoc.* 105, 1178–1187.
- Cabras, S., Castellanos Nueda, M.E., Ruli, E., 2015. Approximate Bayesian computation by modelling summary statistics in a quasi-likelihood framework. *Bayesian Anal.* 10, 411–439.
- Cattelan, M., Sartori, N., 2015. Empirical and simulated adjustments of composite likelihood ratio statistics. *J. Stat. Comput. Simul.* 86, 1056–1067.
- Copt, S., Victoria-Feser, M.P., 2006. High-breakdown inference for mixed linear models. *J. Amer. Statist. Assoc.* 101, 292–300.
- Dawid, A.P., Musio, M., Ventura, L., 2016. Minimum scoring rule inference. *Scand. J. Stat.* 43, 123–138.
- Drovandi, C.C., Pettitt, A.N., Faddy, M.J., 2015. Bayesian indirect inference using a parametric auxiliary model. *Statist. Sci.* 30, 72–95.
- Fearnhead, P., Prangle, D., 2012. Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation (with Discussion). *J. R. Stat. Soc. Ser. B Stat. Methodol.* 74, 419–474.
- Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal.* 1 (3), 515–534.
- Greco, L., Racugno, W., Ventura, L., 2008. Robust likelihood functions in Bayesian inference. *J. Statist. Plann. Inference* 138, 1258–1270.
- Grünwald, P., van Ommen, T., 2017. Inconsistency of Bayesian inference for misspecified linear models, and a proposal for repairing it. *Bayesian Anal.* 12 (4), 1069–1103.
- Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A., 1986. *Robust Statistics. The Approach Based on Influence Functions*. John Wiley & Sons, Chichester, UK.
- Huber, P.J., Ronchetti, E.M., 2009. *Robust Statistics*. John Wiley & Sons, Hoboken, New Jersey.
- Lazar, N.A., 2003. Bayesian empirical likelihood. *Biometrika* 90, 319–326.
- Lewis, J.R., MacEachern, S.N., Lee, Y., 2014. Bayesian Restricted Likelihood. Technical report No. 878, The Ohio State University, USA.
- Marjoram, P., Molitor, J., Plagnol, V., Tavaré, S., 2003. Markov chain Monte Carlo without likelihoods. *Proc. Natl. Acad. Sci.* 100, 15324–15328.

- Markatou, M., Basu, A., Lindsay, B.G., 1998. Weighted likelihood equations with bootstrap root search. *J. Amer. Statist. Assoc.* 93, 740–750.
- Miller, J.W., Dunson, D.B., 2018. Robust Bayesian inference via coarsening. *J. Amer. Statist. Assoc.* <http://dx.doi.org/10.1080/01621459.2018.1469995>.
- Pereira, C.A.d.B., Stern, J.M., 1999. Evidence and Credibility: a full Bayesian test of precise hypothesis. *Entropy* 1, 14–115.
- Pereira, C.A.d.B., Stern, J.M., Wechsler, S., 2008. Can a significance test be genuinely Bayesian? *Bayesian Anal.* 3, 79–100.
- Richardson, A.M., 1997. Bounded influence estimation in the mixed linear model. *J. Amer. Statist. Assoc.* 92, 154–161.
- Richardson, A.M., Welsh, A.H., 1995. Robust restricted maximum likelihood in mixed linear models. *Biometrics* 51, 1429–1439.
- Ruli, E., Sartori, N., Ventura, L., 2016. Approximate Bayesian computation with composite score functions. *Stat. Comput.* 26, 679–692.
- Ruli, E., Sartori, N., Ventura, L., 2018. RobustBLME: Robust bayesian linear mixed-effects models using ABC. <https://cran.r-project.org/package=robustBLME>, R package version 0.1.3.
- Soubeyrand, S., Carpentier, F., Guiton, F., Klein, E.K., 2013. Approximate Bayesian computation with functional statistics. *Stat. Appl. Genet. Mol. Biol.* 12, 17–37.
- Soubeyrand, S., Haon-Lasportes, E., 2015. Weak convergence of posteriors conditional on maximum pseudo-likelihood estimates and implications in ABC. *Statist. Probab. Lett.* 107, 84–92.
- Tramentozzi, E., Ruli, E., Angriman, I., Bardini, R., Campora, M., Guzzardo, V., Zamarchi, R., Rossi, E., Rugge, M., Finotti, P., 2016. Grp94 in complexes with IgG is a soluble diagnostic marker of gastrointestinal tumors and displays immune-stimulating activity on peripheral blood immune cells. *Oncotarget* 7, 72923–72940.
- Tsou, T.-S., Royall, R.M., 1995. Robust likelihoods. *J. Amer. Statist. Assoc.* 90, 316–320.
- Ventura, L., Cabras, S., Racugno, W., 2010. Default prior distributions from quasi- and quasi-profile likelihoods. *J. Statist. Plann. Inference* 43, 2937–2942.
- Watson, J., Holmes, C., 2016. Approximate models and robust decisions. *Statist. Sci.* 31, 465–489.