Approximate Bayesian computation via empirical likelihood

Kerrie L. Mengersen [*], **Pierre Pudlo** [†] [‡], **and Christian P. Robert** [§] [¶] [‖]

[*]Department of Statistics, QUT, Australia,[†]CBGP, INRA, Montpellier, France,[‡]Université Montpellier 2, I3M, Montpellier, France,[§]Université Paris Dauphine, CEREMADE, Paris, France,[¶]Institut Universitaire de France, and [‖]CREST, Paris, France

Submitted to Proceedings of the National Academy of Sciences of the United States of America

Approximate Bayesian computation (ABC) has now become an essential tool for the analysis of complex stochastic models when the likelihood function is unavailable. The well-established statistical method of empirical likelihood however provides another route to such settings that bypasses simulations from the model and the choices of the ABC parameters (summary statistics, distance, tolerance), while being provably convergent in the number of observations. Furthermore, avoiding model simulations leads to significant time savings in complex models, such as those used in population genetics. The ABC_{el} algorithm we develop in this paper also provides an evaluation of its own performance through an associated effective sample size. The method is illustrated using several examples, including estimation of standard and quantile distributions, and time series and population genetics models.

Bayesian statistics | likelihood-free methods | empirical likelihood | population genetics | robust statistics

Abbreviations: ABC, approximate Bayesian computation; DIY-ABC, Do-it-yourself ABC; AMIS, adaptive multiple importance sampling; IS, importance sampling; ABC_{el}, ABC with empirical likelihood; ESS, effective sample size; MLE, maximum likelihood estimator.

Bayesian statistical inference cannot easily operate when the likelihood function associated with the data is not completely known, i.e. cannot be computed in a manageable time, as is the case in most population genetic models (1, 2, 3). The fundamental reason for this difficulty with population genetics is that the statistical model associated with coalescent data needs to integrate over trees of high complexity. Similar computational problems occur on a regular basis in hidden Markov and other dynamic models (4). In those settings, traditional approximation tools based on stochastic simulation (5) are unavailable or unreliable. Indeed, the complexity of the latent structure defining the likelihood makes simulation of such structures too unstable to be trusted. Such settings call for alternative and often cruder approximations. The ABC methodology (1, 6) is a popular algorithm that achieves this by bypassing the computation of the likelihood function (see 7 and 8 for recent surveys).

The fast and polytomous development of the ABC algorithm is indicated by the very active literature in the domain, at both the methodological and the application levels. For instance, a whole new area of population genetic modelling (9, 8) has been explored thanks to the availability of such methods. However, there is a persistent reluctance to adopt ABC algorithms, found in both practitioners and theoreticians alike, as some consider the validation of the method is not steady enough (10, 11, 12). We propose in this paper to connect the ABC approach with a generic and convergent likelihood approximation called the empirical likelihood approach that validates the modified ABC technique as a convergent inference method when the number of observations grows to infinity. The empirical likelihood perspective, introduced by (13), is a robust statistical approach that does not require the specification of the likelihood function. While it does not appear to have been previously used in the ABC setting, this data analysis method is a natural tool to overcome the approximation effects of ABC algorithms. In this paper, we introduce the ABC_{el} algorithm and illustrate its performances on selected representative examples, comparing the outcome with the true posterior density whenever available, and with a standard ABC approximation (14) otherwise.

Statistical Methods

The ABC algorithm. The primary purpose of the ABC algorithm is to achieve the approximation of a simulation from the centrepiece of Bayesian inference, the posterior distribution $\pi(\boldsymbol{\theta}|\mathbf{y}) \propto \pi(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta})$ when it cannot be numerically computed but when the distributions corresponding to both the prior π and the likelihood f can be simulated by manageable computer devices. The original (6) ABC algorithm is as follows: given a sample $\mathbf{y}$ of observations from the sample space, a sample of parameters $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M)$ is produced by

Algorithm 1: ABC sampler
for $i = 1$ to M **do**
  repeat
    Generate $\boldsymbol{\theta}'$ from the prior distribution $\pi(\cdot)$
    Generate $\mathbf{z}$ from the likelihood $f(\cdot|\boldsymbol{\theta}')$
  until $\rho\{\eta(\mathbf{z}), \eta(\mathbf{y})\} \leq \epsilon$
  set $\boldsymbol{\theta}_i = \boldsymbol{\theta}'$,
end for

The parameters of the ABC algorithm are the summary statistic η, the distance $\rho\{\cdot, \cdot\}$ and the tolerance level $\epsilon > 0$. The basic justification of the ABC approximation is that, when using a sufficient statistic η, the distribution of the $\boldsymbol{\theta}_i$'s in the output of the algorithm converges to the genuine posterior distribution when ϵ goes to zero.

In practice, however, the statistic η is insufficient and at the very best the approximation then converges to the genuine posterior $\pi(\boldsymbol{\theta}|\eta(\mathbf{y}))$ when ϵ goes to zero. This loss of information seems to be a necessary price to pay for the access to computable quantities. However, we demonstrate in this paper that it is actually far from necessary in that, when an empirical likelihood technique can be implemented, no reduction in information through the choice of a tolerance zone or of an insufficient summary statistic is required.

Reserved for Publication Footnotes

Empirical likelihood. Owen (13) developed empirical likelihood techniques as a robust alternative to classical likelihood approaches. He demonstrated that, for some categories of statistical models, this approach inherited the convergence properties of standard likelihood at a much lower cost in assumptions about the model. While the current ABC algorithms do require a fully defined and often complex (hence debatable) statistical model, we argue that they could take advantage of the approximation device provided by the empirical likelihood to overcome most of the calibration difficulties encountered by those earlier versions of the method.

Assume that the dataset $\mathbf{y}$ is composed of n independent replicates $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$ of some random vector Y with density f. Rather than defining the likelihood from the density f as in traditional likelihood approaches, the empirical likelihood method starts by defining parameters of interest, $\boldsymbol{\theta}$, as functionals of f, for instance as moments of f, and it then profiles a likelihood in a non-parametric manner. More precisely, given a set of constraints of the form

$$\mathbb{E}_F[h(Y,\theta)] = 0, \qquad [\mathbf{1}]$$

where the dimension of h is the number of constraints unequivocally defining $\boldsymbol{\theta}$, the empirical likelihood is defined as

$$L_{el}(\boldsymbol{\theta}|\mathbf{y}) = \max_{\mathbf{p}} \prod_{i=1}^{n} p_i \qquad [\mathbf{2}]$$

for $\mathbf{p}$ in the set $\{\mathbf{p} \in [0;1]^n,\ \sum p_i = 1,\ \sum_i p_i h(\mathbf{y}_i, \boldsymbol{\theta}) = 0\}$. For instance, in the one-dimensional case when $\theta = \mathbb{E}_f[Y]$, the empirical likelihood in θ is the maximum of the product $p_1 \cdots p_n$ under the constraint $p_1 y_1 + \ldots + p_n y_n = \theta$.

While the convergence of the empirical likelihood is well-established (13), the Bayesian use of empirical likelihoods has been little examined in the past, appart from a Monte Carlo study in (15), a probabilistic interpretation in (16).

ABC$_{\mathbf{el}}$. The most natural use of the empirical likelihood approximation is to act as if this representation was an exact likelihood. When incorporating this perspective into a raw sampler, this leads to the following algorithm:

Algorithm 2: Raw ABC$_{\mathbf{el}}$ sampler

for $i = 1$ to M **do**
  Generate $\boldsymbol{\theta}_i$ from the prior distribution $\pi(\cdot)$
  Set the weight $\omega_i = L_{el}(\boldsymbol{\theta}_i|\mathbf{y})$
end for

The output of this algorithm is therefore a sample of size M of parameters with associated weights. It can thus be used as an importance sampling output (5) and, in particular, the performance of the algorithm can be evaluated through the effective sample size

$$\text{ESS} = 1\Big/\sum_{i=1}^{M}\left\{\omega_i\Big/\sum_{j=1}^{M}\omega_j\right\}^2,$$

which approximates the size of an iid sample with the same variance as the original sample. As shown in (17), this quantity is always between 1 (corresponding to a very poor outcome) and M (corresponding to an iid perfect outcome).

Actually, any available classical algorithm that samples from a posterior distribution (e.g., Monte Carlo Markov chain, Population Monte Carlo, SMC algorithms, see (5)) may equally use the empirical likelihood as if it were the exact likelihood. For instance, to speed up the computation in the population genetics model introduced below, we resorted to the adaptive multiple importance sampling (AMIS) method proposed by (18) which is easy to parallelize on a multi-core computer. While the original target distribution is $\pi(\boldsymbol{\theta})L(\boldsymbol{\theta}|\mathbf{y})$ and the AMIS algorithm uses several (multivariate) Student's t distributions, denoted $t_3(\cdot|\mathbf{m}, \boldsymbol{\Sigma})$ (i.e., with three degrees of freedom, centered at a mean value $\mathbf{m}$ and with covariance matrix $\boldsymbol{\Sigma}$), as an importance sampling distribution, the algorithm can be adapted to the empirical likelihood in a straightforward manner:

Algorithm 3: ABC$_{\mathbf{el}}$-AMIS sampler

for $i = 1$ to M **do**
  Denote $q_1(\cdot)$ the prior distribution.
  Generate $\boldsymbol{\theta}_{1,i}$ from the prior distribution $q_1(\cdot)$
  Set $\omega_{1,i} = L_{el}(\boldsymbol{\theta}_{1,i}|\mathbf{y})$
end for
for $t = 2$ to T_M **do**
  Compute weighted mean $\mathbf{m}_t$ and weighted variance matrix $\boldsymbol{\Sigma}_t$ of the $\boldsymbol{\theta}_{s,i}$ $(1 \le s \le t-1,\ 1 \le i \le M)$.
  Denote $q_t(\cdot)$ the density of $t_3(\cdot|\mathbf{m}_t, \boldsymbol{\Sigma}_t)$.
  for $i = 1$ to M **do**
    Generate $\boldsymbol{\theta}_{t,i}$ from $q_t(\cdot)$.
    Set $\omega_{t,i} = \pi(\boldsymbol{\theta}_{t,i})L_{el}(\boldsymbol{\theta}_{t,i}|\mathbf{y})\big/\sum_{s=1}^{t-1} q_s(\boldsymbol{\theta}_{t,i})$
  end for
  for $r = 1$ to $t-1$ **do**
    for $i = 1$ to M **do**
      Update the weight of $\boldsymbol{\theta}_{r,i}$ as
      $\omega_{r,i} = \pi(\boldsymbol{\theta}_{t,i})L_{el}(\boldsymbol{\theta}_{r,i}|\mathbf{y})\big/\sum_{s=1}^{t-1} q_s(\boldsymbol{\theta}_{r,i})$
    end for
  end for
end for

The output of this algorithm is a weighted sample $\boldsymbol{\theta}_{t,i}$ of size $M \times T_M$.

Note that in contrast with the standard ABC algorithm, ABC$_{\text{el}}$ algorithms do not require simulations from the model, given that [**2**] provides a converging and non-parametric approximation of the likelihood function. Moreover, there is no need for calibrating the many tuning parameters found in standard ABC algorithms; in particular, the likelihood ratio acts as a natural distance and the use of importance weights produces an implicit and self-defined quantile on the original sample simulated from the prior. Notwithstanding these appealing qualities, we stress that the algorithm still requires calibration, in particular in the choice of the parameterisation of the distribution and of the corresponding constraints [**1**] in the empirical likelihood. Some examples of this are discussed below. We also stress that, from a Bayesian perspective, the pointwise mathematical validation of the method for a given sample size and even less for a given dataset is not available nor even meaningful.

Composite likelihood in population genetics. ABC was first introduced by population geneticists (2, 9, 6) interested in statistical inference about the evolutionary history of species, on the ground that no likelihood-based approach existed apart from very rudimentary and hence unrealistic situations. This approach has since been used in a number of biological studies (19, 20, 21), most of them including model choice. It is therefore crucial to obtain insights into the validity of such studies, particularly when they deal with issues of economical or ecological importance (see, e.g., (22)). This can be achieved in part by running a comparison using ABC$_{\text{el}}$. Furthermore, given the major gain in computing time for ABC$_{\text{el}}$, achieved by the absence of replications of the data, ABC$_{\text{el}}$ can be applied to more complex biological models.

The main difficulty when using the empirical likelihood in such settings is to derive a constraint [**1**] on the parameter of interest: in phylogeography, parameters like divergence

dates, effective population sizes, mutation rates, etc. cannot be expressed as moments of the distribution of the sample at a given locus. In particular, the datapoints are not iid. However, when considering microsatellite loci with the stepwise mutation model (23) and evolutionary scenarios composed of divergence, we can derive the pairwise composite scores whose zero is the pairwise maximum likelihood estimator. Composite likelihoods have been previously proved consistent for estimating recombination rates, introducing approximation of the dependency structure between nearby loci (24, 25, 26, 27).

More specifically, we are here approximating the intra-locus likelihood using a product over all pairs of genes in the sample at a given locus. Assuming that y_i^k denotes the allele of the i-th gene in the sample at the k-th locus, and that ϕ is the vector of parameters, then the so-called pairwise likelihood of the data at the k-th locus, namely $\mathbf{y}^k$, is defined by

$$\ell_2(\mathbf{y}^k|\phi) = \prod_{i<j} \ell_2(y_i^k, y_j^k|\phi)$$

and the corresponding pairwise score function is $\nabla_\phi \ell_2(\mathbf{y}^k|\phi)$. Pairwise score equations

$$\mathbb{E}_f[\nabla_\phi \ell_2(Y|\phi)] = 0$$

provide a constraint **[1]** in every way comparable to the score equations that give the maximum likelihood estimate and which is quite powerful for empirical likelihood derivations (28, pp. 48–50). Hence the empirical likelihood of the full dataset $\mathbf{y} = (\mathbf{y}^1, \ldots, \mathbf{y}^K)$ given ϕ is computed with **[2]** under the (multidimensional) constraint that

$$\sum_{k=1}^{K} p_k \nabla_\phi \log \ell_2(\mathbf{y}^k|\phi) = 0.$$

When the effective population size is identical over all populations of the demographical scenario, the time axis might be scaled so that coalescence of two genes in Kingman's genealogy occurs with rate $k(k-1)/2$ if there are k lineages. In this modified scale, mutations at a given locus arise with rate $\theta/2$ along the gene genealogy. Our mutation model is the simple stepwise mutation model of (23), i.e. the number of repeats of the mutated gene increases or decreases by one unit with equal probability. Given two microsatellite allelic states x_1 and x_2, their pairwise likelihood $\ell_2(x_1, x_2|\phi)$ depends only on the difference of the two states $x_1 - x_2$. If the two genes belong to individuals that lie in the same deme, then (see SI and (29))

$$\ell_2(x_1, x_2|\phi) = \frac{1}{\sqrt{1+2\theta}} \rho(\theta)^{|x_2 - x_1|},$$

where $\rho(\theta) = \theta/\left(1 + \theta + \sqrt{1+2\theta}\right)$. If the two genes belong to individuals that lie in two demes having diverged at time τ, then (29)

$$\ell_2(x_1, x_2|\phi) = \frac{\mathrm{e}^{-\tau\theta}}{\sqrt{1+2\theta}} \sum_{k=-\infty}^{+\infty} \rho(\theta)^{|k|} I_{|x_1 - x_2| - k}(\tau\theta)$$

where $I_\delta(z)$ denotes the δth-order modified Bessel function of the first kind evaluated at z. Computing the pairwise scores, i.e. partial derivatives of $\log \ell_2(x_1, x_2|\phi)$ from those equations, is straightforward, by recalling that

$$\frac{\mathrm{d}I_\delta(z)}{\mathrm{d}z} = (I_{\delta-1}(z) + I_{\delta+1}(z))/2\,.$$

Algorithm $\mathrm{ABC_{el}}$ is therefore directly available in this setting, and furthermore at a cost much lower than the one associated with the standard ABC algorithm.

Results

Normal distribution. Starting with the very simple example of a normal distribution with known variance (equal to one), we can check whether or not the empirical likelihood allows for a proper recovery of the true posterior distribution. Fig. S1 shows that a constraint based on the mean works well, as do the two constraints on mean and variance (Figure S2). On the other hand, using the three first moments in the definition of the empirical likelihood degrades the fit (Figure S3).

Quantile distributions. Quantile distributions are defined by a closed-form quantile function $F^-1(p;\theta)$, and generally have no closed form for the density function. They are of great interest because of their flexibility and the ease with which they can be simulated by a simple inversion of the uniform distribution. A range of methods, including ABC approaches (9), have been proposed for estimation; see supplementary material for further details. We focus here on the four-parameter g-and-k distribution, defined by its quantile function

$$Q(r; A, B, g, k) = A + B\left(1 + c\frac{1-\exp(-gz(r))}{1+\exp(-gz(r))}\right) \quad [3]$$
$$\times \left(1 + z(r)^2\right)^k z(r) \quad [4]$$

where $z(r)$ is the rth standard normal quantile; the parameters A, B, g and k represent location, scale, skewness and kurtosis, respectively and c measures the overall asymmetry (30, 31). We evaluated the $\mathrm{ABC_{el}}$ algorithm for estimation of this distribution using two values of $\theta = (A, B, g, k)$, two sets of priors and various combinations of n, M and p, where p is the number of percentiles used as constraints; details are in the supplementary material.

Figure 1 illustrates the true and fitted curves and a 95% credible region for the case with $n = 100, M = 5000$ and $p = 3$. The corresponding posterior means (standard deviations) for the parameters A, B, g, k were $3.08(0.14), 1.12(0.23), 1.79(0.25), 0.41(0.12)$, respectively. The choice of sample size and number of constraints did not substantively affect the accuracy of parameter estimates, but the precision was noticeably improved for the larger sample size; see Figures S4, S5, and S6.

The accuracy and precision of the estimates were broadly comparable with the results obtained by (32) for the same distribution. Based on the whole experiment, the parameters A and B were well estimated in all cases, while the estimates of g and k were poorer for smaller values of n and M. For small n the estimates were more subject to the vagaries of sampling variation, whereas for small M they were subject to the influence of a smaller number of very large importance weights. However, given the speed of $\mathrm{ABC_{el}}$compared with the competing ABC algorithms, it is feasible to use even larger values of M than considered in this experiment, since there is no requirement to simulate new datasets at each iteration.

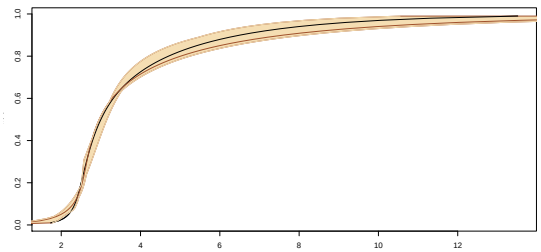

Fig. 1. True and fitted curves with a 95% credible region for a dataset of $n = 100$ observations from the g-and-k distribution, based on $M = 10^3$ simulations.

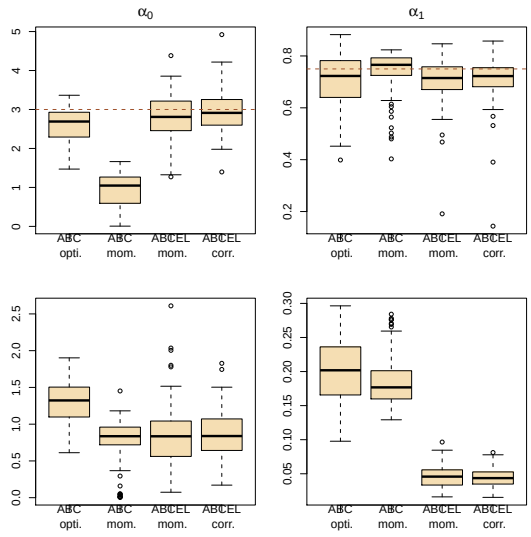

Fig. 2. Comparison of ABC evaluations of posterior expectations *(top, with true values in dashed lines)* and posterior variances *(bottom)* of the parameters (α_0, α_1) of the ARCH(1) model with 100 observations. The first two columns correspond to two choices of summary statistics for the standard ABC algorithm (least squares estimates and mean of the $\log y_t$'s plus autocorrelations of order 2 and 3, respectively). The last two columns correspond to two sets of constraints for the ABC_{el} version (first three moments and second moment plus autocorrelation of order 1 plus correlation with previous observation for the reconstituted ϵ_t's. All experiments are based on the same reference table of 10^4 simulations, with the tolerance ϵ chosen as the 1% quantile of the distances.

Moreover, this experiment is based on the very basic case of sampling from the prior; the results would be further improved by using an analogue of ABC_{el}-AMIS or alternative approaches similar to those proposed by (33) for traditional ABC.

Dynamic models. In dynamic models, the difficulty with empirical likelihood relates to the lack of independence in the observed data $(y_t)_{1\le t\le T}$. Indeed, those models can most often be represented as transforms of unobserved iid sequences $(\epsilon_t)_{1\le t\le T}$. The recovery of a converging empirical likelihood representation thus requires the reconstitution of the ϵ_t's as transforms of the data $\mathbf{y}$ and of the parameter θ. The independence between the ϵ_t's is then at least as important as moment conditions.

For instance, consider a standard and simple dynamic model, namely the ARCH(1) model:

$$y_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0,1)\,, \quad \sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2\,,$$

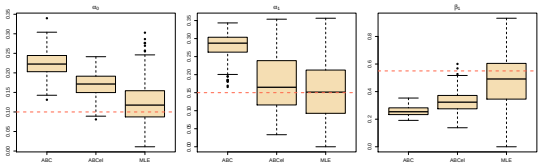

Fig. 3. Comparison of ABC evaluations of posterior expectations *(with true values in dashed lines)* of the parameters $(\alpha_0, \alpha_1, \beta_1)$ of the GARCH(1) model with 250 observations. The first row corresponds to an optimal ABC algorithm (using the MLE of the parameters as summary statistics and with the tolerance ϵ chosen as the 5% quantile of the distances), the second row corresponds to the ABC_{el} algorithm based on the constraints derived in (36), and the third row corresponds to the MLE derived by the R procedure **garch** when initialised at the true value of the paramters.

with a uniform prior over the simplex, i.e., $\alpha_0, \alpha_1 \ge 0$, $\alpha_0 + \alpha_1 \le 1$. While this model can be handled by other means, since the likelihood function is available, we will compare here the behaviour of standard and empirical likelihood ABC algorithms.

First, a natural empirical likelihood representation is based on the reconstituted ϵ_t's, defined as y_t/σ_t when the σ_t's are derived recursively. Figure 2 shows the result of an estimation of both parameters α_0 and α_1 when Algorihm ABC uses as summary statistics either the least square estimates of the parameters (directly obtained from the series (y_t^2)), which we label "optimal ABC" in connection with (34), or the mean of the series $\log(y_t^2)$ supplemented by the two first autocorrelations of the series (y_t^2)). The constraints in the empirical likelihood are either based on the three first moments of the reconstituted ϵ_t's or on the variance of those ϵ_t's complemented by both the correlation between the y_{t-1}'s and the ϵ_t's and the correlation between the ϵ_{t-1}'s and the ϵ_t's. As seen from this experiment, ABC_{el} does as well as the optimal ABC for the estimation of the parameters, but further brings a reduction in the variability of those estimates, thanks to the importance weights.

A much more complex dynamic model is the Garch(1, 1) model of (35) that can be formalised as the observation of

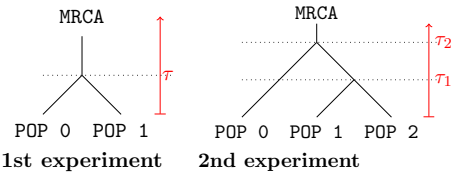

Fig. 4. Evolutionary scenarios of the two experiments in population genetics.

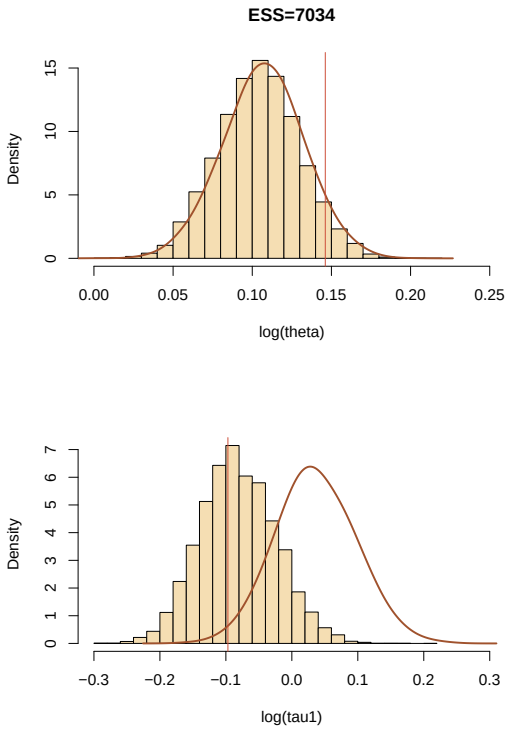

Fig. 5. Comparison of the original ABC (curve) with the histogram of the simulation from the ABC_{el}–AMIS sampler in the case of the population genetics model given in Scenario A, based on uniform priors on $(\log_{10}(\theta), \log_{10}(\tau))$ on $(-1, 1.5) \times (-1, 1)$ and 10^4 particles.

$y_t \sim \mathcal{N}(0, \sigma_t^2)$ when

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \qquad [5]$$

under the constraints $\alpha_0, \alpha_1, \beta_1 > 0$ and $\alpha_1 + \beta_1 < 1$, that is, $y_t = \sigma_t \epsilon_t$. (*Garch* stands for generalised autoregressive conditional heteroskedastic.) Given the constraints on the parameters, a natural prior is to choose an exponential distribution on α_0, for instance an exponential $\mathcal{E}xp(1)$ distribution, and a Dirichlet $D_3(1,1,1)$ on $(\alpha_1, \beta_1, 1-\alpha_1-\beta_1)$. A standard ABC approach requires the choice of summary statistics, which are necessarily non-sufficient since the model is a state-space model. Following (34), we use the maximum likelihood estimator as summary statistics, relying on the R function garch for its derivation despite its lack of stability (the true value of the parameters was used as initial value in the function). Since (36) derived natural score constraints for the empirical likelihood associated with this model, we used their constraints to build our ABC_{el} algorithm. Fig. 3 provides a comparison of both approaches with the MLE. It shows in particular that the ABC algorithm is unable to produce acceptable inference in this case, even in the most favourable case when it is initialised at a satisfactory maximum likelihood estimate (as shown by the bottom row). The ABC_{el} algorithm is performing better, even though it fails to catch the correct range of β_1.

Population genetics. We compare our proposal with the reliable ABC-based estimates given by (3). We set up two toy experiments that are designed to defeat ABC, using pseudo observed data. The two evolutionary scenarios are given in Figure 4. In all experiments, we only consider microsatellite loci and assume that the effective population size is identical over all populations of the scenario.

In the first experiment, we consider two populations having diverged at time τ in the past, see Figure 4. Our pseudo observed datasets are made of thirty diploid individuals per population genotyped at hundred independent loci. We compare the marginal posterior distributions of the two unknown parameters θ and τ computed with the original ABC method (using the DIY-ABC software of (37)) and with the ABC_{el}-AMIS sampler. In this case, results are improved when the θ-component of the composite scores, namely $\partial_\theta \log \ell_2(\mathcal{D}|\phi)$, is restricted the sum over all pairs of genes lying in the same population. Otherwise, ABC_{el} systematically under-estimates θ. This means that the information regarding θ in the part of the (unobserved) gene genealogy that links both populations is either too noisy, either becoming to much stressed or not retrived by the pairwise composite approximation. Figure 5 shows the typical discrepency between both results: ABC and ABC_{el} agree on the mutation rate θ, but the ABC_{el} estimation of τ is more accurate, see also Table 1.

In the second experiment, we consider three populations, see Figure 4: the last two populations diverged at time τ_1 and their common ancestral population diverged from the first population at time τ_2. The sample is made of thirty diploid individuals per population genotyped at hundred independant loci. In constrast to the the first experiment, all components of the composite scores are computed here by summing over all pairs of genes whatever the population they belong to. The results given in Table 1 shows that ABC and ABC_{el} mainly agree on both parameters θ and τ_1, but ABC_{el} is slightly more accurate than genuine ABC on τ_2.

It should be noticed also that, on a six core CPU, ABC_{el} takes about one minute (per dataset) to provide the results described above, whereas computing the many simulated datasets (preliminary to any ABC analysis) on such large datasets requires several hours of computation.

Discussion

Since its introduction by (1) and (6), ABC has been extensively used in several areas involving complex likelihoods, primarily in population genetics, both for point estimation and testing of hypotheses. The experience gained for a long time on summary statistics in population genetics has helped ABC to become an efficient algorithm for parameter estimation.

In population genetics, when the dataset is composed of large sets of markers, the summary statistics proposed in DIY-ABC (which are means of some quantitative statistics over all hundred loci) lose some information, while ABC_{el} manages to find much more information, more specifically to estimate the dates of divergence on large datasets.

When compared with ABC, the significant time savings provided by ABC_{el} can certainly open new doors in population genetics. The statistical study of large datasets or complex models might be considered. For instance, genuine ABC development is severely hindered by the time spent to simulate a dataset when modelling isolation by distance in a continuously distributed population, or when studying a large set of SNP markers even on much simpler evolutionary scenarios.

ACKNOWLEDGMENTS. The last two authors are grateful to Jean-Marie Cornuet for his help and availability. Their work has been partly supported by the Agence Nationale de la Recherche (ANR) through the 2009–2012 project Emile. The third author is grateful to Patrice Bertail, Brunero Liseo, and Art Owen for useful discussions.

References

1. Tavaré S, Balding D, Griffith R, Donnelly P (1997) Inferring coalescence times from DNA sequence data. *Genetics* 145:505–518.
2. Beaumont M, Zhang W, Balding D (2002) Approximate Bayesian computation in population genetics. *Genetics* 162:2025–2035.
3. Cornuet JM, et al. (2008) Inferring population history with DIYABC: a user-friendly approach to Approximate Bayesian Computation. *Bioinformatics* 24:2713–2719.
4. Cappé O, Moulines E, Rydén T (2004) *Hidden Markov Models* (Springer-Verlag, New York).
5. Robert C, Casella G (2004) *Monte Carlo Statistical Methods* (Springer-Verlag, New York), second edition.
6. Pritchard J, Seielstad M, Perez-Lezaun A, Feldman M (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution* 16:1791–1798.
7. Beaumont M (2010) Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics* 41:379–406.
8. Lopes J, Beaumont M (2010) ABC: a useful Bayesian tool for the analysis of population data. *Infection, Genetics and Evolution* 10:825–832.
9. Marjoram P, Molitor J, Plagnol V, Tavaré S (2003) Markov chain Monte Carlo without likelihoods. *Proc. Nat. Acad. Sci. USA* 100:15324–15328.
10. Templeton A (2010) Coherent and incoherent inference in phylogeography and human evolution. *Proc. Nat. Acad. Sci. USA* 107(14):6376–6381.
11. Berger J, Fienberg S, Raftery A, Robert C (2010) Incoherent phylogeographic inference. *Proc. Nat. Acad. Sci. USA* 107:E57.
12. Robert C, Cornuet JM, Marin JM, Pillai N (2011) Lack of confidence in approximate Bayesian computation model choice. *Proc. Natl Acad. Sci USA* 108:15112–15117.
13. Owen AB (1988) Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* 75:237–249.
14. Marin J, Pudlo P, Robert C, Ryder R (2012) Approximate Bayesian computational methods. *Statistics and Computing* (To appear.).
15. Lazar NA (2003) Bayesian empirical likelihood. *Biometrika* 90:319–326.

16. Schennach SM (2005) Bayesian exponentially tilted empirical likelihood. *Biometrika* 92:31–46.
17. Liu J (2001) *Monte Carlo Strategies in Scientific Computing* (Springer-Verlag, New York).
18. Cornuet JM, Marin JM, Mira A, Robert C (2012) Adaptive multiple importance sampling. *Scandinavian Journal of Statistics* (To appear.).
19. Estoup A, Beaumont M, Sennedot F, Moritz C, Cornuet J (2004) Genetic analysis of complex demographic scenarios: spatially expanding populations of the cane toad, Bufo Marinus. *Evolution* 58:2021–2036.
20. Estoup A, Clegg S (2003) Bayesian inferences on the recent island colonization history by the bird Zosterops lateralis lateralis. *Mol. Ecol.* 12:657–674.
21. Fagundes N, et al. (2007) Statistical evaluation of alternative models of human evolution. *Proc. Nat. Acad. Sci. USA* 104:17614–17619.
22. Lombaert E, et al. (2010) Bridgehead effect in the worldwide invasion of the biocontrol Harlequin Ladybird. *PloS ONE* 5:e9743.
23. Ohta H, Kimura M (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genet. Res.* 22.
24. Hudson RR (2001) Two-locus sampling distributions and their application. *Genetics* 159:1805–1817.
25. Kim Y, Stephan W (2002) Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* 160:765–777.
26. McVean G, Awadalla P, Fearnhead P (2002) A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* 160:1231–1241.
27. Fearnhead P (2003) Consistency of estimators of the population-scaled recombination rate. *Theoretical Population Biology* 64:67–79.
28. Owen AB (2001) *Empirical Likelihood* (Chapman & Hall).
29. Wilson IJ, Balding DJ (1998) Genealogical inference from microsatellite data. *Genetics* 150:499–510.
30. Haynes M, MacGillivray H, Mengersen K (1997) Robustness of ranking and selection rules using generalised g-and$k$ distributions. *J. Stat. Plan. Inference* 65:45–66.
31. Gilchrist W (2000) *Statistical Modelling with Quantile Functions* (Chapman and Hall).
32. Allingham D, King R, Mengersen K (2009) Bayesian estimation of quantile distributions. *Statistics and Computing* 19:189–201.
33. Drovandi C, Pettitt A (2011) Likelihood-free bayesian estimation of multivariate quantile distributions. *Computational Statistics and Data Analysis* 55:2541–2556.
34. Fearnhead P, Prangle D (2012) Semi-automatic approximate Bayesian computation. *J. Royal Statist. Society Series B* To appear, with discussion.
35. Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31:307–327.
36. Chan N, Ling S (2006) Empirical likelihood for GARCH models. *Econometric Theory* 22:403.
37. Cornuet J, Ravigné V, Estoup A (2010) Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics* 11:401.

Table 1. Comparison of the original ABC and ABC_{el} on 100 Monte Carlo replicates. We use two point estimates of the parameters: (1) posterior mean and (2) posterior median, and measured the error between the estimation and the "true" value used to generate the observation with (1) the root mean square error in the case of the posterior mean and (2) the median absolute deviation in the case of the posterior median. We also compare credibility intervals (of probability 0.8) through the proportion of Monte Carlo replicates in which the "true" value fall into this interval.

First experiment						
	Root Mean Square Error of posterior mean		Median Absolute Deviation of posterior median		Coverage of the credibility interval of probability 0.8	
	ABC	ABC_{el}	ABC	ABC_{el}	ABC	ABC_{el}
θ	0.0971	0.0949	0.071	0.059	0.68	0.81
τ	0.315	0.117	0.272	0.077	1.0	0.80

Second experiment						
	Root Mean Square Error of posterior mean		Median Absolute Deviation of posterior median		Coverage of the credibility interval of probability 0.8	
	ABC	ABC_{el}	ABC	ABC_{el}	ABC	ABC_{el}
θ	0.0593	0.0794	0.0484	0.0528	0.79	0.76
τ_1	0.472	0.483	0.320	0.280	0.88	0.76
τ_2	29.6	4.76	4.13	3.36	0.89	0.79