

Bayesian computation via empirical likelihood

Kerrie L. Mengersen^a, Pierre Pudlo^{b,c,d}, and Christian P. Robert^{e,f,g,1}

^aSchool of Mathematical Sciences, Queensland University of Technology, Brisbane, QLD 4001, Australia; ^bCentre de Biologie pour la Gestion des Populations, Institut National de la Recherche Agronomique, 34988 Montferrier-sur-Lez Cedex, France; ^cUniversité Montpellier 2, Institut de Mathématiques et de Modélisation de Montpellier, 34095 Montpellier Cedex 5, France; ^dInstitut de Biologie Computationnelle, Montpellier, France; ^eUniversité Paris Dauphine, Centre de Recherche en Mathématiques de la Décision, 75775 Paris Cedex 16, France; ^fInstitut Universitaire de France, Paris, France; and ^gCentre de Recherche en Statistique et Economie, 92245 Malakoff Cedex, France

Edited by Stephen E. Fienberg, Carnegie Mellon University, Pittsburgh, PA, and approved December 7, 2012 (received for review May 25, 2012)

Approximate Bayesian computation has become an essential tool for the analysis of complex stochastic models when the likelihood function is numerically unavailable. However, the well-established statistical method of empirical likelihood provides another route to such settings that bypasses simulations from the model and the choices of the approximate Bayesian computation parameters (summary statistics, distance, tolerance), while being convergent in the number of observations. Furthermore, bypassing model simulations may lead to significant time savings in complex models, for instance those found in population genetics. The Bayesian computation with empirical likelihood 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 distributions, time series, and population genetics models.

autoregressive models | Bayesian statistics | likelihood-free methods | coalescent model

Bayesian statistical inference cannot easily operate when the likelihood function associated with the data is not entirely known, or cannot be computed in a manageable time, as is the case in most population genetic models (1–3). 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 issues with the likelihood function often occur 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 approximate Bayesian computation (ABC) methodology (1, 6) is a popular solution that bypasses the computation of the likelihood function (surveys in refs. 7 and 8); ref. 9 validates a conditional version of ABC that applies to hierarchical Bayes models in a wide generality.

The fast and polytomous development of the ABC algorithm is indicated by the rising literature in the domain, at both the methodological and the application levels. For instance, a whole new area of population genetic modeling (8, 10) has been explored thanks to the availability of such methods. However, practitioners and theoreticians both show a reluctance in adopting ABC, with some doubt about the validity of the method (11–13). We propose in this paper to supplement the ABC approach with a generic and convergent likelihood approximation called the empirical likelihood that validates this Bayesian computational technique as a convergent inferential method when the number of observations grows to infinity. The empirical likelihood perspective, introduced by ref. 14, is a robust statistical approach that does not require the specification of the likelihood function. However, although it does not appear to have been used before in the settings that now rely on ABC, this data analysis method also is a broadly (albeit not universally) applicable and often fast approach, the approximation of which differs from the one found in ABC algorithms, even though both are rooted in nonparametric statistics. Therefore, this methodology can be used both as a solution per se and as a benchmark

against which to test the ABC output in many cases. This paper presents the Bayesian computation via empirical likelihood (BC_{el}) algorithm and illustrates its performances on selected representative examples, comparing the outcome with the true posterior density whenever available, and with an ABC approximation (15) otherwise.

Statistical Methods

ABC Algorithm. The primary purpose of the ABC algorithm is to approximate simulation from the centerpiece of Bayesian inference, the posterior distribution $\pi(\theta|\mathbf{y}) \propto \pi(\theta)f(\mathbf{y}|\theta)$, 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 $(\theta_1, \dots, \theta_M)$ is produced by

Algorithm 1: ABC sampler.

```
for  $i = 1$  to  $M$  do
  repeat
    Generate  $\theta'$  from the prior distribution  $\pi(\cdot)$ 
    Generate  $\mathbf{z}$  from the likelihood  $f(\cdot|\theta')$ 
    until  $\rho\{\eta(\mathbf{z}), \eta(\mathbf{y})\} \leq \epsilon$ 
    set  $\theta_i = \theta'$ 
  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 θ_i 's in the output of the algorithm converges to the genuine posterior distribution when ϵ goes to zero (16).

In practice, however, the statistic η is nonsufficient and at best the approximation then converges to the genuine posterior $\pi(\theta|\eta(\mathbf{y}))$ when ϵ goes to zero. This loss of information seems to be a necessary price to pay for the access to computable quantities. Furthermore, as argued below, it can be evaluated against the empirical likelihood approximation when the latter is available. Indeed, this approach does not require an information reduction through the choice of a tolerance zone or of a nonsufficient summary statistic.

Empirical Likelihood. Owen (14) 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 (as detailed in *SI Text*). Whereas ABC algorithms do require a fully defined and often complex (hence debatable) statistical model, we argue that one should take advantage of the approximation device provided by the empirical likelihood to overcome most of the calibration difficulties encountered by ABC, at least as a convenient benchmark against which to test ABC solutions.

Assume that the dataset \mathbf{y} is composed of n independent replicates $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$ of some random vector Y with density f . Rather than defining the likelihood from the density f as usual, the empirical likelihood method starts by defining parameters of interest θ as functionals of f , for instance as

Author contributions: C.P.R. designed research; K.L.M., P.P., and C.P.R. performed research; K.L.M., P.P., and C.P.R. analyzed data; and K.L.M., P.P., and C.P.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence should be addressed. E-mail: Christian.Robert@ceremade.dauphine.fr.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1208827110/-DCSupplemental.

moments of f , and it then profiles a nonparametric likelihood. More precisely, given a set of constraints of the form

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

where the dimension of h sets the number of constraints unequivocally defining θ , the empirical likelihood is defined as

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

for \mathbf{p} in the set $\{\mathbf{p} \in [0, 1]^n, \sum p_i = 1, \sum_i p_i h(\mathbf{y}_i, \theta) = 0\}$. 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 $p_1 y_1 + \cdots + p_n y_n = \theta$. (Solving 2 is done using the R package “emplik” developed by ref. 17 and based on the Newton–Lagrange algorithm.) When the data are not independent and identically distributed (iid), an underlying iid structure may sometimes be exploited, as illustrated in *Dynamic Model*. However, this is not always the case, meaning that the empirical likelihood method remains out of reach in some complex cases when ABC can still be implemented. Furthermore, as pointed out in *SI Text* by a quote from Owen (18), the validation of the approach depends on a choice of the set of constraints that ensures convergence.

Whereas the convergence of the empirical likelihood is well-established (*SI Text* and ref. 18), the Bayesian use of empirical likelihoods has been little examined in the past, apart from a Monte Carlo study in ref. 19, and a probabilistic one in ref. 20.

BC_{el}. The most natural use of the empirical likelihood approximation is to act as if this representation was an exact likelihood, as in ref. 19. Incorporating this perspective into a basic sampler leads to the following algorithm:

Algorithm 2: Basic BC_{el} sampler.

```

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

```

The output of BC_{el} is a sample of size M of parameters with associated weights, which operate as an importance sampling output (5). Thus, the performance of the algorithm can be evaluated through the effective sample size (ESS):

$$ESS = 1 / \left(\frac{1}{M} \sum_{i=1}^M \left\{ \omega_i / \sum_{j=1}^M \omega_j \right\}^2 \right),$$

which approximates the size of an iid sample with the same variance as the original sample. As shown in ref. 21, this quantity is always between 1 (corresponding to a very poor outcome) and M (corresponding to an iid perfect outcome).

Any algorithm that samples from a posterior distribution (e.g., Markov chain Monte Carlo, population Monte Carlo, sequential Monte Carlo algorithms, ref. 5) may instead use the empirical likelihood as a proxy to 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, ref. 22), which is easy to parallelize on a multicore computer. Although the original target distribution is $\pi(\theta)L(\theta|\mathbf{y})$ and the AMIS algorithm uses several (multivariate) Student’s t distributions denoted $t_3(\cdot|\mathbf{m}, \Sigma)$ (i.e., with three degrees of freedom, centered at mean \mathbf{m} , and with covariance matrix Σ) as an importance sampling distribution, the algorithm can be adapted to the empirical likelihood in a straightforward manner:

Algorithm 3: BC_{el}–AMIS sampler.

```

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

```

for $i = 1$ to M do

```

  Update the weight of  $\theta_{r,i}$  as
   $\omega_{r,i} = \pi(\theta_{r,i})L_{el}(\theta_{r,i}|\mathbf{y}) / \sum_{s=1}^{t-1} q_s(\theta_{r,i})$ 
end for
end for
end for

```

The output is thus a weighted sample $\theta_{t,i}$ of size MT_M .

In contrast with ABC, BC_{el} algorithms do not usually require simulations from the sampling model, given that 2 provides a converging and nonparametric approximation of the likelihood function. This feature thus induces very significant improvements in computing time when the production of pseudodatasets is time-consuming, because solving 1 is usually immediate. This is for instance the case in population genetics and *Time Gains in Population Genetic Models* provides an illustration of a huge improvement in comparison with ABC in two experiments described below. However, the improvement in speed may vanish in cases when producing an iid structure connected with the constraint 1 requires simulations from the sampling model, as illustrated by a counterexample for point processes in *SI Text*, BC_{el}, and ABC, then breaking even in terms of computing time. Even though the computing time required by BC_{el} is customarily negligible compared with ABC (or does not induce any extra time as in the point process counterexample), we further caution against opposing both approaches solely based on computing times, because they differ in the approximations they provide to a genuine Bayesian analysis and thus should be used in conjunction.

Using empirical likelihoods means there is no calibration of the many tuning parameters of ABC algorithms; most significantly, the likelihood ratio acts as a natural distance and importance weights produce an implicit and self-defined quantile on the original sample simulated from the prior. Notwithstanding these appealing qualities, BC_{el} still requires calibration, in particular in the choices of the parametrization of the sampling distribution and of the corresponding constraints 1 defining the empirical likelihood. Some examples are discussed below. The BC_{el}–AMIS sampler also implies computing values of the prior density, up to a constant, which may be a hindrance in peculiar cases.

Composite Likelihood in Population Genetics. ABC was introduced by population geneticists (2, 6, 10) interested in statistical inference about the evolutionary history of species, as no likelihood-based approach existed apart from very rudimentary and hence unrealistic situations. This approach has been used in a number of biological studies (23–25), most of them including model choice. It is therefore crucial to obtain insights into the validity of such studies, particularly when they have economic, biological, or ecological consequences (e.g., ref. 26). This can be achieved in part by running a comparison using BC_{el}. Furthermore, given the major gain in computing time due to the absence of replications of the data, BC_{el} can be applied to more complex biological models.

The main caveat when using the empirical likelihood in such settings is selecting 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 sampling distribution at a given locus. In particular, the data are not iid. However, when considering microsatellite loci with the stepwise mutation model (27) and evolutionary scenarios composed of divergence, we can derive the pairwise composite scores whose zero is the pairwise maximum likelihood estimator (MLE). Composite likelihoods have been proved consistent for estimating recombination rates, introducing an approximation of the dependency structure between nearby loci (28–31). (Also, ref. 32 gives composite likelihoods used in a likelihood-free setting.)

More specifically, we are approximating the intralocus likelihood by 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 $\nabla_{\phi} \log \ell_2(\mathbf{y}^k|\phi)$. Pairwise score equations

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

provide a constraint 1 in every way comparable to the score equations that give the maximum likelihood estimator and which is quite powerful for empirical likelihood derivations (ref. 18), pp 48–50). Hence the empirical likelihood of the full dataset $\mathbf{y} = (\mathbf{y}^1, \dots, \mathbf{y}^K)$ given ϕ is computed with 2 under the (multidimensional) constraint that

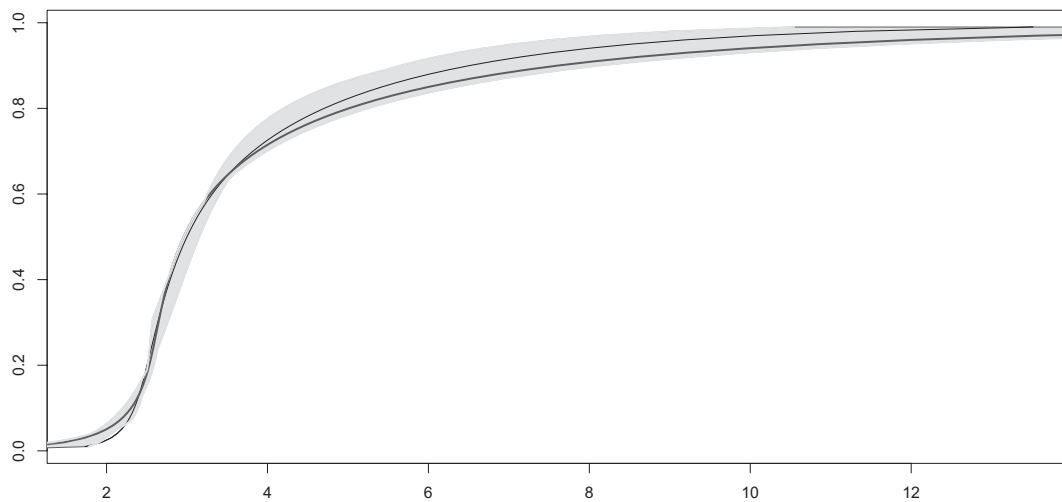


Fig. 1. True (black) and fitted (brown) cdf functions with a pointwise 95% credible (shaded gray) region centered on the fitted cumulative distribution function for a dataset of $n = 100$ observations from the g -and- k distribution, based on $M = 10^3$ simulations of BC_{el} .

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

When the effective population size is identical over all populations of the demographic scenario, the time axis may 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 ref. 27; i.e., the number of repeats of the mutated gene increases or decreases by 1 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 states $x_1 - x_2$. If both genes belong to individuals that lie in the same deme, then (SI Text and ref. 33)

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

where $\rho(\theta) = \theta / (1 + \theta + \sqrt{1 + 2\theta})$. If the two genes belong to individuals from demes having diverged at time τ , then (33)

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

where $I_{\delta}(z)$ denotes the δ th-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 virtue of recurrence properties of the Bessel functions. Algorithm BC_{el} is therefore directly available in this setting, and furthermore at a cost much lower than the one associated with ABC algorithms (Table S1).

Results

Normal Distribution. Starting with the benchmark of a normal distribution with known variance (equal to 1), we can check that the empirical likelihood allows for a proper recovery of the true posterior distribution on the mean. Fig. S1 shows that a constraint 1 based on the mean works well, as do the two constraints on mean and second central moment, $\mathbb{E}[(X - \theta)^2] = 0$ (Fig. S2). On the other hand, using the first three central moments in the empirical likelihood may degrade the fit (three cases in Fig. S3). Whereas this poor fit is not signaled by the ESS (which is often larger than in Figs. S1 and S2, because of the growing disconnection between the approximation and the true likelihood and hence a more uniform range of the weights), a parallel run of the method with different collections of constraints does detect the discrepancy. This illustrates the variability of the empirical likelihood approximation, as well as its sensitivity to the choice of

defining constraints. Although a drawback of the method, this variability can be tested and evaluated by comparing outcomes, due to often limited computing costs. This toy experiment also supports the generic recommendation (18) to keep the number of constraints and parameters equal.

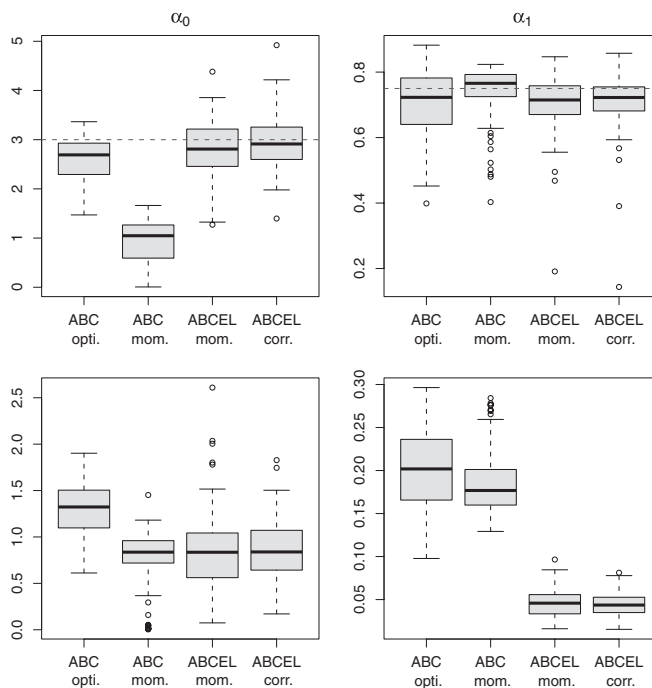


Fig. 2. (Upper) Comparison of ABC evaluations of posterior expectations, with true values given in dashed lines. (Lower) Posterior variances of the parameters (α_0, α_1) of the ARCH (1) model with 100 observations. The first two columns correspond to two choices of summary statistics for the ABC algorithm (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 BC_{el} alternative (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.

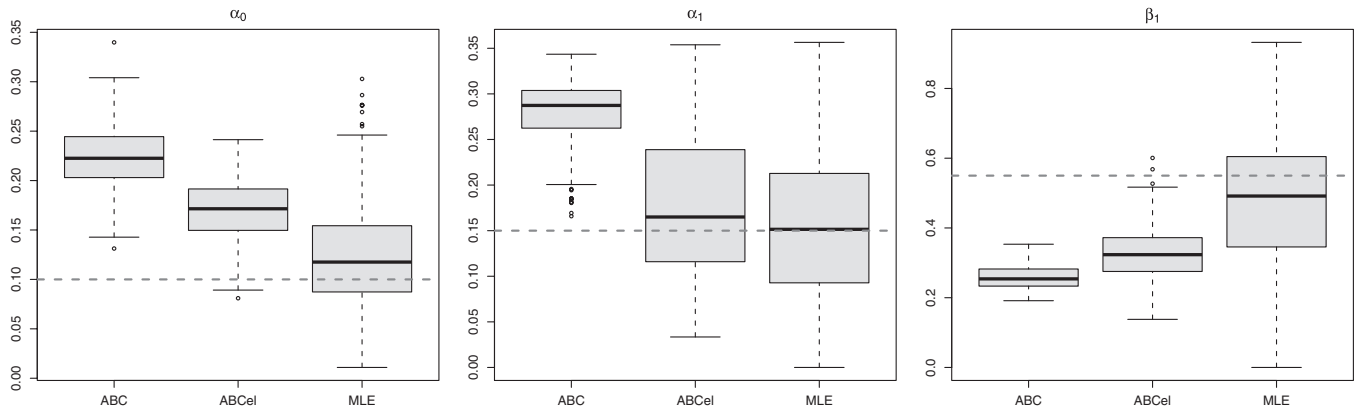


Fig. 3. Comparison of evaluations of posterior expectations, with true values in dashed lines, of the parameters $(\alpha_0, \alpha_1, \beta_1)$ of the GARCH (1) model with 250 observations. The first row corresponds to an optimal ABC algorithm (using the MLE as summary statistic and with the tolerance ϵ chosen as the 5% quantile of the distances), the second row corresponds to the BC_{el} algorithm based on the constraints derived in ref. 40, and the third row corresponds to the MLE derived by the R procedure GARCH initialized at the true parameter value.

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 (10), has been proposed for estimation (*SI Text*). We focus here on the four-parameter g -and- k distribution, defined by its quantile function, denoted $Q(r; A, B, g, k)$ and equal to

$$A + B \left(1 + c \frac{1 - \exp(-gz(r))}{1 + \exp(-gz(r))} \right) \left(1 + z(r)^2 \right)^k z(r),$$

where $z(r)$ is the r th standard normal quantile; the parameters A, B, g , and k represent location, scale, skewness, and kurtosis, respectively, and c measures the overall asymmetry (34, 35). We evaluated the BC_{el} algorithm for estimating 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 in *SI Text*).

Fig. 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 (SDs) for the parameters A, B, g , and k were 3.08(0.14), 1.12(0.23), 1.79(0.25), and 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 (*Figs. S4–S6*).

The accuracy and precision of the estimates were broadly comparable with the results obtained by ref. 36 for the same distribution. Based on the whole experiment, the parameters A and B were well estimated in all cases, whereas 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 BC_{el} compared with competing ABC algorithms, it is feasible to use even larger values of M than considered in this experiment, because there is no requirement to simulate new datasets at each iteration. Moreover, this experiment is based on the very basic case of sampling from the prior; the results would be further improved by using an analog of BC_{el} -AMIS or alternative approaches similar to those proposed by ref. 37 for ABC.

Dynamic Models. In dynamic models, the difficulty with empirical likelihood stems from the dependence in the data $(y_t)_{1 \leq t \leq T}$. However, these models can be represented as transforms of unobserved iid sequences $(\epsilon_t)_{1 \leq t \leq T}$. 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 . Independence between the ϵ_t 's is then at least as important as moment conditions. (This implies equivalent computing times for ABC and BC_{el} .)

For instance, consider a simple dynamic model, namely the autoregressive conditionally heteroskedastic 1 [Arch(1)] model:

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

with a uniform prior over the simplex, i.e., $\alpha_0, \alpha_1 \geq 0, \alpha_0 + \alpha_1 \leq 1$. Whereas this model can be handled by other means, because the likelihood function is available we will compare here the behavior of ABC and BC_{el} 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. Fig. 2 shows the result of estimating both parameters α_0 and α_1 when Algorithm ABC uses as summary statistics either the least-square estimates of the parameters [derived from the series (y_t^2)], which we label “optimal ABC” in connection with ref. 38, or the mean of the series $\log(y_t^2)$ supplemented by the first two autocorrelations of the series (y_t^2) . The constraints in the empirical likelihood are either based on the first three moments of the reconstituted ϵ_t 's or on the variance of those ϵ_t 's complemented by both the correlations between the y_{t-1} 's and the ϵ_t 's and between the ϵ_{t-1} 's and the ϵ_t 's. As seen from this experiment, BC_{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 generalized ARCH (GARCH)(1, 1) model of ref. 39 that can be formalized as the observation of $y_t \sim \mathcal{N}(0, \sigma_t^2)$ when

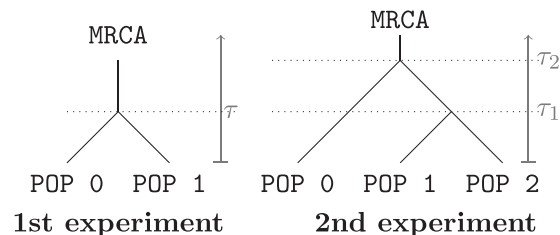


Fig. 4. Evolutionary scenarios of the two experiments in population genetics.

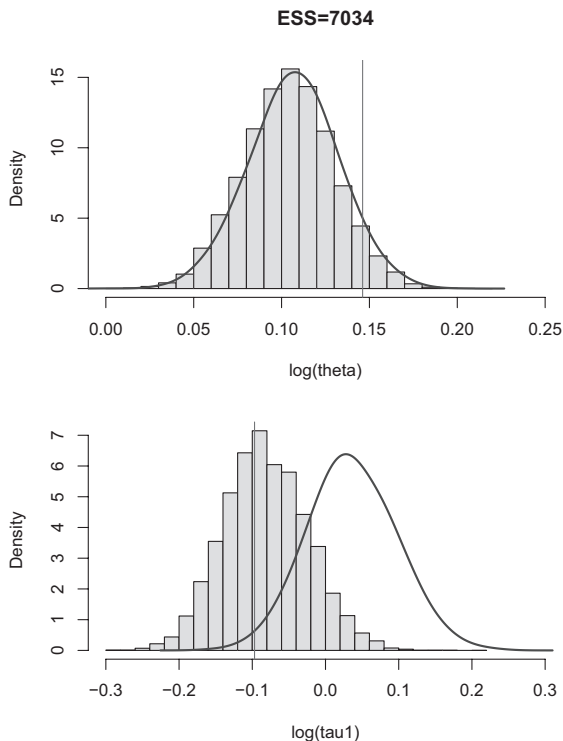


Fig. 5. Comparison of the original ABC (Curve) with the histogram of the simulation from the BC_{e1} -AMIS sampler in the case of the population genetics model given in scenario A, based on uniform priors on $(\log_{10}(\theta), \log_{10}(\tau))$ on $(-1, 1.5) \times (-1, 1)$ and 10^4 particles.

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad [3]$$

under the constraints $\alpha_0, \alpha_1, \beta_1 > 0$ and $\alpha_1 + \beta_1 < 1$, that is, $y_t = \sigma_t \epsilon_t$. Given the constraints on the parameters, a natural priority is to choose an exponential distribution on α_0 , for instance an exponential $\text{Exp}(1)$ distribution, and a Dirichlet $D_3(1, 1, 1)$ on $(\alpha_1, \beta_1, 1 - \alpha_1 - \beta_1)$. An ABC approach requires the choice of summary statistics, which are necessarily nonsufficient because the model is a state-space model. Following ref. 38, we use the MLE as summary statistics, relying on the R function `garch` for its derivation despite its lack of stability. Because ref. 40 derived natural score constraints for the empirical likelihood associated with this model, we used their constraints to build our BC_{e1} 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 favorable case when it is initialized at a satisfactory maximum likelihood estimate (as shown by the bottom row). The BC_{e1} algorithm is performing better, even though it fails to catch the correct range of β_1 .

Another type of non-iid model relying on the superposition of an unknown number of gamma point processes and processed in ref. 41 through a (non-Bayesian) alternative to ABC is discussed in *SI Text* as an additional illustration of the possibilities of the empirical likelihood perspective for complex models, offering a free benchmark for evaluating the ABC outcome. Fig. S7 shows a clear improvement brought by BC_{e1} over the corresponding ABC outcome.

Population Genetics. We compare our proposal with the reliable ABC-based estimates given by ref. 3. We set up two toy experiments that are designed to defeat ABC, using pseudoobserved data. The two evolutionary scenarios are given in Fig. 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 which diverged at time τ in the past (Fig. 4, *Left*). Our pseudoobserved datasets are made of 30 diploid individuals per population genotyped at 100 independent loci. We compare the marginal posterior distributions of the unknown parameters θ and τ computed with the ABC method [using the Do It Yourself ABC (DIYABC) software of ref. 42] and with the BC_{e1} -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 to the sum over all pairs of genes lying in the same population. Otherwise, as can be checked via a quick simulation experiment, BC_{e1} systematically underestimates θ . Fig. 5 shows the typical discrepancy between both results: ABC and BC_{e1} agree on the mutation rate θ , but the BC_{e1} estimation of τ is more accurate (Table 1).

In the second experiment, we consider three populations (Fig. 4, *Right*): 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 comprises 30 diploid individuals per population genotyped at 100 independent loci. In contrast with the first experiment, all components of the composite scores are computed here by summing over all pairs of genes whatever the population to which they belong. The results given in Table 1 show that ABC and BC_{e1} mainly agree on both parameters θ and τ_1 , but BC_{e1} is slightly more accurate than ABC on τ_2 .

Table S1 gives a comparison of the computing times for both algorithms, showing the difference of magnitudes between them. This is due to the simulation of the simulated datasets for ABC: Although this difference should not be overinterpreted, it signals a potential for self-assessment and testing that is missing for ABC methods.

Table 1. Comparison of the original ABC and ABC_{e1} on 100 Monte Carlo replicates

Experiment	Root-mean-square error of posterior mean		Median absolute deviation of posterior median		Coverage of the credible interval with probability 0.8	
	ABC	ABC_{e1}	ABC	ABC_{e1}	ABC	ABC_{e1}
First experiment						
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						
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

We use two point estimates of the parameters (*i*) posterior mean and (*ii*) posterior median, and measured the error between the estimation and the "true" value used to generate the observation with (*i*) the root-mean-square error in the case of the posterior mean and (*ii*) the median absolute deviation in the case of the posterior median. We also compare credible intervals (with probability 0.8) through the proportion of Monte Carlo replicates in which the true value falls into this interval.

Discussion

Compared with ABC methods, the (often) significant time savings provided by BC_{el} due to the lack of pseudosample simulation may open wider ranges for processing models involving complex likelihoods. For instance, in population genetics, ABC is severely hindered by the time spent simulating a dataset when modeling isolation by distance in a continuously distributed population, or when studying a large set of SNP markers even on quite simple evolution scenarios. Moreover, when the dataset is composed of large sets of markers, the summary statistics proposed in ABC (in DIYABC, these are averages of some quantitative statistics over all loci) ignores some (statistical) information, whereas BC_{el} manages to recover most of it, more specifically to estimate divergence on large datasets. Improvements in accuracy of estimation and computational efficiency are also possible in other contexts as illustrated in the range of examples given above.

Even when BC_{el} requires the same computing time as ABC, it uses the outcome in a very different perspective and provides a benchmark likelihood that helps in evaluating the pertinence of the ABC approximation, as illustrated in the gamma point process of SI.

We acknowledge a caveat of the empirical likelihood: it requires a careful choice of the constraint **1**. Those pivotal quantities have to be connected to the parameter in an identifying way, which may require complex manipulations as in the gamma process case or may even be impossible. However, repeated experimentation is often available, as illustrated by the normal example and the population genetic experiments (where we computed the composite score on both a restricted set of pairs and all pairs of genes). Checking for the accuracy of the approximation means that a constraint in BC_{el} should be tested on simulated datasets in controlled experiments where the true parameters are known, although much less than in ABC runs. Then we can test coverage of credibility intervals, and measure the error of various point estimates based on the output of the scheme.

ACKNOWLEDGMENTS. P.P. and C.P.R. thank Jean-Marie Cornuet for his help and availability. C.R.P. is grateful to Patrice Bertail, Chris Drovandi, Brunero Liseo, and Art Owen for useful discussions. Comments and suggestions from the Editor and Reviewers greatly contributed to improve both the presentation and scope of the paper. P.P. and C.P.R. were partly supported by the Agence Nationale de la Recherche through the 2009–2012 Project Emile.

- Tavaré S, Balding DJ, Griffiths RC, Donnelly P (1997) Inferring coalescence times from DNA sequence data. *Genetics* 145(2):505–518.
- Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics* 162(4):2025–2035.
- Cornuet JM, et al. (2008) Inferring population history with DIY ABC: A user-friendly approach to approximate Bayesian computation. *Bioinformatics* 24(23):2713–2719.
- Cappé O, Moulines E, Rydén T (2004) *Hidden Markov Models* (Springer, New York).
- Robert C, Casella G (2004) *Monte Carlo Statistical Methods* (Springer, New York), 2nd Ed.
- Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW (1999) Population growth of human Y chromosomes: A study of Y chromosome microsatellites. *Mol Biol Evol* 16(12):1791–1798.
- Beaumont M (2010) Approximate Bayesian computation in evolution and ecology. *Annu Rev Ecol Syst* 41:379–406.
- Lopes JS, Beaumont MA (2010) ABC: A useful Bayesian tool for the analysis of population data. *Infect Genet Evol* 10(6):826–833.
- Bazin E, Dawson KJ, Beaumont MA (2010) Likelihood-free inference of population structure and local adaptation in a Bayesian hierarchical model. *Genetics* 185(2):587–602.
- Marjoram P, Molitor J, Plagnol V, Tavaré S (2003) Markov chain Monte Carlo without likelihoods. *Proc Natl Acad Sci USA* 100(26):15324–15328.
- Templeton AR (2010) Coherent and incoherent inference in phylogeography and human evolution. *Proc Natl Acad Sci USA* 107(14):6376–6381.
- Berger JO, Fienberg SE, Raftery AE, Robert CP (2010) Incoherent phylogeographic inference. *Proc Natl Acad Sci USA* 107(41):E157–author reply E158.
- Robert CP, Cornuet JM, Marin JM, Pillai NS (2011) Lack of confidence in approximate Bayesian computation model choice. *Proc Natl Acad Sci USA* 108(37):15112–15117.
- Owen AB (1988) Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* 75(2):237–249.
- Marin J, Pudlo P, Robert C, Ryder R (2012) Approximate Bayesian computational methods. *Stat Comput* 22(6):1167–1180.
- Biau G, Cérou F, Guyader A (2012) *New Insights into Approximate Bayesian Computation*. Tech Rep HAL 00721164. Available at <http://hal.inria.fr/hal-00721164/en>.
- Zhou M (2012) emplik: Empirical likelihood ratio for censored/truncated data. *R package version 0.9-8-2*.
- Owen AB (2001) *Empirical Likelihood* (Chapman & Hall, London).
- Lazar NA (2003) Bayesian empirical likelihood. *Biometrika* 90:319–326.
- Schennach SM (2005) Bayesian exponentially tilted empirical likelihood. *Biometrika* 92(1):31–46.
- Liu J (2001) *Monte Carlo Strategies in Scientific Computing* (Springer, New York).
- Cornuet JM, Marin JM, Mira A, Robert C (2012) Adaptive multiple importance sampling. *Scand J Stat* 39(4):798–812.
- Estoup A, Beaumont M, Sennedot F, Moritz C, Cornuet JM (2004) Genetic analysis of complex demographic scenarios: Spatially expanding populations of the cane toad, *Bufo marinus*. *Evolution* 58(9):2021–2036.
- Estoup A, Clegg SM (2003) Bayesian inferences on the recent island colonization history by the bird *Zosterops lateralis lateralis*. *Mol Ecol* 12(3):657–674.
- Fagundes NJ, et al. (2007) Statistical evaluation of alternative models of human evolution. *Proc Natl Acad Sci USA* 104(45):17614–17619.
- Lombaert E, et al. (2010) Bridgehead effect in the worldwide invasion of the bio-control harlequin ladybird. *PLoS ONE* 5(3):e9743.
- 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(2):201–204, and reprinted in (2007) 89(5–6):367–370.
- Hudson RR (2001) Two-locus sampling distributions and their application. *Genetics* 159(4):1805–1817.
- Kim Y, Stephan W (2002) Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* 160(2):765–777.
- McVean G, Awadalla P, Fearnhead P (2002) A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* 160(3):1231–1241.
- Fearnhead P (2003) Consistency of estimators of the population-scaled recombination rate. *Theor Popul Biol* 64(1):67–79.
- Barthelmé S, Chopin N (2012) Expectation-propagation for likelihood-free inference. Tech Rep arXiv:1107.5959v2. Available at <http://arxiv.org/pdf/1107.5959>.
- Wilson LJ, Balding DJ (1998) Genealogical inference from microsatellite data. *Genetics* 150(1):499–510.
- Haynes M, MacGillivray H, Mengersen K (1997) Robustness of ranking and selection rules using generalized g - and k distributions. *J Stat Plann Inference* 65(1):45–66.
- Gilchrist W (2000) *Statistical Modelling with Quantile Functions* (Chapman & Hall, London).
- Allingham D, King R, Mengersen K (2009) Bayesian estimation of quantile distributions. *Stat Comput* 19(2):189–201.
- Drovandi C, Pettitt A (2011) Likelihood-free Bayesian estimation of multivariate quantile distributions. *Comput Stat Data Anal* 55(9):2541–2556.
- Fearnhead P, Prangle D (2012) Semi-automatic approximate Bayesian computation. *J R Stat Soc, B* 74(3):419–474.
- Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. *J Econom* 31(3):307–327.
- Chan N, Ling S (2006) Empirical likelihood for GARCH models. *Econometric Theory* 22(3):403–428.
- Cox D, Kartsonaki C (2012) The fitting of complex parametric models. *Biometrika* 99(3):741–747.
- Cornuet JM, Ravnigné V, Estoup A (2010) Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinf* 11:401.