



HAL
open science

Adaptive approximate Bayesian computation for complex models

Maxime Lenormand, Franck Jabot, Guillaume Deffuant

► **To cite this version:**

Maxime Lenormand, Franck Jabot, Guillaume Deffuant. Adaptive approximate Bayesian computation for complex models. 2012. hal-00638484v3

HAL Id: hal-00638484

<https://hal.science/hal-00638484v3>

Preprint submitted on 31 Jul 2012 (v3), last revised 20 May 2015 (v4)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Adaptive approximate Bayesian computation for complex models

Maxime Lenormand · Franck Jabot · Guillaume Deffuant

Received: date / Accepted: date

Abstract We propose a new approximate Bayesian computation (ABC) algorithm that aims at minimizing the number of model runs for reaching a given quality of the posterior approximation. This algorithm automatically determines its sequence of tolerance levels and makes use of an easily interpretable stopping criterion. Moreover, it avoids the problem of particle duplication found when using a MCMC kernel. When applied to a toy example and to a complex social model, our algorithm is 2 to 8 times faster than the three main sequential ABC algorithms currently available.

Keywords Approximate Bayesian computation · Population Monte Carlo · Sequential Monte Carlo · Complex model

1 Introduction

Approximate Bayesian computation (ABC) techniques appear particularly relevant for calibrating stochastic models because their very principle includes stochasticity but they are applicable to any model. They generate a sample of model parameter values $(\theta_i)_{i=1,\dots,N}$ (often also called particles) from the prior distribution $\pi(\theta)$ and select the θ_i values leading to model outputs $x \sim f(x|\theta_i)$ satisfying a proximity criterion with the target data y ($\rho(x, y) \leq \varepsilon$, ρ expressing a distance, ε being a tolerance level). The selected sample of parameter values approximates the posterior distribution of parameters, leading to model outputs with the expected quality of approximation. However, in practice, running these techniques is very demanding computationally because sampling the whole space of parameters requires a number of simulations which grows exponentially with the number of parameters to identify. This tends to limit the application of these techniques to easily computable models (Beaumont 2010). In this paper, our goal is minimizing the number of model runs for reaching a given quality of posterior approximation, and thus to make the approach applicable to a larger set of models.

ABC is the subject of intense scientific researches and several improved versions of the original scheme are available, such as using local regressions to improve parameter inference (Beaumont et al. 2002; Blum and François 2010), automatically selecting informative summary statistics (Joyce and Marjoram 2008; Fearnhead and Prangle 2011), coupling to Markov chain Monte Carlo (Marjoram et al. 2003; Wegmann et al. 2009) or improving sequentially the posterior distributions with sequential Monte Carlo methods

M. Lenormand
IRSTEA, LISC, 24 avenue des Landais, 63172 AUBIERE, France
Tel.: +334-73-440617
E-mail: maxime.lenormand@irstea.fr

F. Jabot
IRSTEA, LISC, 24 avenue des Landais, 63172 AUBIERE, France
Tel.: +334-73-440733
E-mail: franck.jabot@irstea.fr

G. Deffuant
IRSTEA, LISC, 24 avenue des Landais, 63172 AUBIERE, France
Tel.: +334-73-440614
E-mail: guillaume.deffuant@irstea.fr

(Sisson et al. 2007; Toni et al. 2009; Beaumont et al. 2009). This last class of methods approximates progressively the posterior, using sequential samples $S^{(t)} = (\theta_i^{(t)})_{i=1,\dots,N}$ derived from sample $S^{(t-1)}$, and using a decreasing set of tolerance levels $\{\varepsilon_1, \dots, \varepsilon_T\}$. This strategy focuses the sampling effort in parts of the parameter space of high likelihood, avoiding to spend much computing time in systematically sampling the whole parameter space.

The first sequential method applied to ABC was proposed by Sisson et al. (2007) with the ABC-PRC (Partial Rejection Control). This method is based on a theoretical work of Del Moral et al. (2006) to ABC. However, Beaumont et al. (2009) has shown that this method leads to a bias in the approximation of the posterior. Beaumont et al. (2009); Toni et al. (2009) proposed a new algorithm, called Population Monte Carlo ABC in Beaumont et al. (2009) and hereafter called PMC. This algorithm, corrects the bias by affecting to each particle a weight corresponding to the inverse of its importance in the sample. It is particularly interesting in our perspective because it provides with a rigorous framework to the sequential sample idea, which seems a good way for minimizing the number of runs. In this approach, the problem is then defining the sequence of tolerance levels $\{\varepsilon_1, \dots, \varepsilon_T\}$. Drovandi and Pettitt (2011) and Del Moral et al. (2012) solve partly this problem by deriving the tolerance level at a given step from values of $\rho(x, y)$ of the previously selected sample. However, a difficulty remains: when to stop? If the final tolerance level ε_T is too large, the final posterior will be of bad quality. Inversely, a too small ε_T leads to a posterior that could have been obtained with less model runs.

Moreover, the MCMC kernel used in Drovandi and Pettitt (2011) and Del Moral et al. (2012) to sample new values $\theta_j^{(t)}$, despite its mathematical elegance, has a significant drawback in our view: it can lead to particle duplications. Indeed, each time the MCMC jumps from a particle to a new one which is not accepted, the initial particle is kept in the new sample of particles, hence when this occurs several times with the same initial particle, this particle appears several times in the new sample. As long as these duplications are very few compared with the size of the sample, their effect can be neglected, but it can easily happen that their number grows to a very significant part of the sample, then strongly deteriorating the quality of the posterior, as illustrated below. To solve this problem, Drovandi and Pettitt (2011) proposed to perform R MCMC jump trials instead of one, while Del Moral et al. (2012) proposed to resample the parameter values when too many are duplicated. Del Moral et al. (2012) also proposed to run the model M times for each particle, in order to decrease the variance of the acceptance ratio of the MCMC jump. However, these solutions increase the number of model runs, going against the initial benefit of using sequential samples.

In this paper, we propose a modification of the PMC algorithm that we call adaptive population Monte Carlo ABC (hereafter called APMC). This new algorithm determines by itself the sequence of tolerance levels as in Drovandi and Pettitt (2011) and Del Moral et al. (2012), and it also provides a stopping criterion. Furthermore, our approach avoids the problem of duplications. We prove that the computation of the weights associated to the particles in this algorithm lead to the intended posterior distribution and we also prove that the algorithm stops whatever the chosen value of the stopping parameter. We show that our algorithm, applied to a toy example and to an individual-based social model, requires significantly less simulations to reach a given quality level of the posterior distribution than the PMC algorithm of Beaumont et al. (2009), the replenishment SMC ABC algorithm of Drovandi and Pettitt (2011) (hereafter called RSMC) and the adaptive SMC ABC algorithm of Del Moral et al. (2012) (hereafter called SMC). These algorithms are detailed in Appendix A.

2 Adaptive population Monte-Carlo approximate Bayesian computation

2.1 Overview of the APMC algorithm

The APMC algorithm follows the main principles of the sequential ABC, and defines on-line the tolerance level at each step like in Drovandi and Pettitt (2011) and Del Moral et al. (2012). For each tolerance level ε_t , it generates a sample $S^{(t)}$ of particles and computes their associated weights. This weighted sample approximates the posterior distribution, with an increasing approximation quality as ε_t decreases. We say that a parameter value $\theta_i^{(t)}$, satisfies the tolerance level ε_t , if when running the model we get $x \sim f(x|\theta_i^{(t)})$, such that its distance $\rho_i^{(t)} = \rho(x, y)$ to the target data y , is below ε_t . Suppose the APMC reached step $t - 1$, with a sample $S^{(t-1)}$ of $N_\alpha = \lfloor \alpha N \rfloor$ particles and their associated weights $(\theta_i^{(t-1)}, w_i^{(t-1)})_{i=1,\dots,N_\alpha}$, the main features of the APMC are (see Algorithm 5 for details):

- the algorithm generates $N - N_\alpha$ particles $(\theta_j^{(t-1)})_{j=N_\alpha+1,\dots,N}$ where $\theta_j^{(t-1)} \sim \mathcal{N}(\theta_j^*, \sigma_{(t-1)}^2)$, the seed θ_j^* is randomly drawn from the weighted set $(\theta_i^{(t-1)}, w_i^{(t-1)})_{i=1,\dots,N_\alpha}$ and the variance $\sigma_{(t-1)}^2$ of the Gaussian kernel $\mathcal{N}(\theta_j^*, \sigma_{(t-1)}^2)$ is twice the empirical variance of the weighted set $(\theta_i^{(t-1)}, w_i^{(t-1)})_{i=1,\dots,N_\alpha}$, following [Beaumont et al. \(2009\)](#).
- the weights $w_j^{(t-1)}$ of the new particles $(\theta_j^{(t-1)})_{j=N_\alpha+1,\dots,N}$ are computed so that these new particles can be combined with the sample $S^{(t-1)}$ of the previous step without causing a bias in the posterior distribution. These weights are given by Eq. 2 (see below).
- the algorithm concatenates the N_α previous particles $(\theta_i^{(t-1)})_{i=1,\dots,N_\alpha}$ with the $N - N_\alpha$ new particles $(\theta_j^{(t-1)})_{j=N_\alpha+1,\dots,N}$, together with their associated weights and distances to the data. This constitutes a new set noted $S_{temp}^{(t)} = (\theta_i^{(t)}, w_i^{(t)}, \rho_i^{(t)})_{i=1,\dots,N}$.
- the next tolerance level ε_t is determined as the first α -quantile of the $(\rho_i^{(t)})_{i=1,\dots,N}$.
- the new sample $S^{(t)} = (\theta_i^{(t)}, w_i^{(t)})_{i=1,\dots,N_\alpha}$ is then constituted from the N_α particles of $S_{temp}^{(t)}$ satisfying the tolerance level ε_t .
- if the proportion p_{acc} of particles satisfying the tolerance level ε_{t-1} among the $N - N_\alpha$ newly generated particles is below a chosen value $p_{acc,min}$, the algorithm stops, and its result is $(\theta_i^{(t)})_{i=1,\dots,N_\alpha}$ with their associated weights.

Note that in our algorithm, to get a number N_α of retained particles for the next step, the choice of ε_t is heavily constrained: it has to be at least equal to the first α -quantile of the $(\rho_i^{(t)})_{i=1,\dots,N}$ and smaller than the immediately superior $(\rho_i^{(t)})$ value. We chose to fix it to the first α -quantile for simplicity. This choice also ensures that the tolerance level decreases from one iteration to the next: in the worst case where $p_{acc} = 0$ (no newly simulated particles accepted), $\varepsilon_t = \varepsilon_{t-1}$. Our algorithm does not use a MCMC kernel and avoids duplicating particles. It requires a reweighting step in $O(N_\alpha^2)$ instead of $O(N_\alpha)$ in [Drovandi and Pettitt \(2011\)](#), but in our perspective, this computational cost is supposed negligible compared with the cost of running the model.

2.2 Weights correcting the kernel sampling bias

As pointed out by [Beaumont et al. \(2009\)](#), the newly generated particles $\theta_i^{(t)}$ in a sequential procedure are no more drawn from the prior distribution but from a specific probability density $d_i^{(t)}$ that depends on the particles selected at the previous step and on the chosen kernel. This introduces a bias in the procedure. This bias should be corrected by attributing a weight equal to $\pi(\theta_i^{(t)})/d_i^{(t)}$ to each newly generated particle $\theta_i^{(t)}$.

The density of probability $d_i^{(t)}$ to generate particle $\theta_i^{(t)}$ at step t is given by the sum of the probabilities to reach $\theta_i^{(t)}$ from one of the N_α particles of the previous step times their respective weights:

$$d_i^{(t)} = \sum_{j=1}^{N_\alpha} \frac{w_j^{(t-1)}}{\sum_{k=1}^{N_\alpha} w_k^{(t-1)}} \sigma_{t-1}^{-1} \varphi \left(\sigma_{t-1}^{-1} (\theta_i^{(t)} - \theta_j^{(t-1)}) \right) \quad (1)$$

where $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ is the kernel function.

This yields the expression of the weight $w_i^{(t)}$ to be attributed to the newly drawn particle $\theta_i^{(t)}$:

$$w_i^{(t)} = \frac{\pi(\theta_i^{(t)})}{\sum_{j=1}^{N_\alpha} \left(w_j^{(t-1)} / \sum_{k=1}^{N_\alpha} w_k^{(t-1)} \right) \sigma_{t-1}^{-1} \varphi \left(\sigma_{t-1}^{-1} (\theta_i^{(t)} - \theta_j^{(t-1)}) \right)} \quad (2)$$

This formula differs from the scheme of [Beaumont et al. \(2009\)](#) where the weights need only to be proportional to Eq. 2 at each step. Since we want to concatenate particles obtained at different steps of the algorithm (while [Beaumont et al. \(2009\)](#) generate the sample at step t from scratch), we need the scaling of weights to be consistent across the different steps of the algorithm. Using the weight of Eq. 2 guarantees the correction of the sampling bias throughout the APMC procedure and ensures that the N_α weighted particles $\theta_i^{(t)}$ produced at the t -th iteration follow the posterior distribution $\pi(\theta | \rho(x, y) < \varepsilon_t)$.

2.3 The stopping criterion

We stop the algorithm when the proportion of "accepted" particles (Eq. 3) among the $N - N_\alpha$ new particles is below a predetermined threshold $p_{acc_{min}}$. This choice of stopping rule ensures that additional simulations would only marginally change the posterior distribution. Note that this stopping criterion will be achieved even if $p_{acc_{min}} = 0$, this ensures that the algorithm converges. We present a formal proof of this assertion in Appendix B.

$$p_{acc}(t) = \frac{1}{N - N_\alpha} \sum_{k=N_\alpha+1}^N \mathbb{1}_{\rho_k^{(t-1)} < \varepsilon_{t-1}} \quad (3)$$

3 Experiments on a toy example

We consider four algorithms: APMC, PMC, the SMC and the RSMC. Their implementations in R ([R Development Core Team 2011](#)) are available ¹. We compare them on the toy example studied in [Sisson et al. \(2007\)](#) where $\pi(\theta) = \mathcal{U}_{[-10,10]}$ and $f(x|\theta) \sim \frac{1}{2}\phi(\theta, \frac{1}{100}) + \frac{1}{2}\phi(\theta, 1)$ where $\phi(\mu, \sigma^2)$ is the normal density of mean μ and variance σ^2 . In this example, we consider that $y = 0$ is observed, so that the posterior density of interest is proportional to $(\phi(0, \frac{1}{100}) + \phi(0, 1))\pi(\theta)$.

We structure the comparisons on two indicators: the number of simulations performed during the application of the algorithms, and the \mathbb{L}_2 distance between the exact posterior density and the histogram of particle values obtained with the algorithms. This \mathbb{L}_2 distance is computed on the 300-tuple obtained by dividing the support $[-10, 10]$ into 300 equally-sized bins.

We choose $N = 5000$ particles and a target tolerance level equal to 0.01. For the PMC algorithm we use a decreasing sequence of tolerance levels from $\varepsilon_1 = 2$ down to $\varepsilon_{11} = 0.01$. For the SMC algorithm, we use 3 different values for α : $\{0.9, 0.95, 0.99\}$ and $M = 1$ as in [Del Moral et al. \(2012\)](#). For the RSMC algorithm we use $\alpha = 0.5$ as in [Drovandi and Pettitt \(2011\)](#). To explore our algorithm, we test 9 different values for α : $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$, and 4 different values for $p_{acc_{min}}$: $\{0.01, 0.05, 0.1, 0.2\}$. In each case, we perform 50 times the algorithm, and compute the average and standard deviation of the two indicators: the total number of simulations and the \mathbb{L}_2 distance between the exact posterior density and the histogram of particle values. We used as kernel transition a normal distribution parameterized with twice the weighted variance of the previous sample, as in [Beaumont et al. \(2009\)](#).

We report below the effects of varying α and $p_{acc_{min}}$ on the performance of our algorithm, and compare it with the PMC, SMC and RSMC algorithms.

3.1 Particle duplication in SMC and RSMC

The number of distinct particles decreases during the course of the SMC algorithm whatever the value of α , as shown on Fig. 1a-b. The oscillations of the number of distinct particles are caused by the resampling step in the SMC algorithm (see [Del Moral et al. \(2012\)](#)), but they are not sufficient to counterbalance the overall decrease. This decrease deteriorates the posterior approximation as shown on Fig. 2. For the RSMC algorithm, the number of distinct particles is maintained at a reasonably high level (Fig. 1c), but this has a cost in terms of the number of required model runs (see Fig. 2). Note that the APMC and the PMC algorithms keep N distinct particles.

3.2 Influence of parameters on APMC

The values of α and $p_{acc_{min}}$ have an impact on the studied indicators. We find that smaller α and $p_{acc_{min}}$ improve the quality of the approximation (smaller \mathbb{L}_2 distance), and increase the total number of model runs, with $p_{acc_{min}}$ having the largest effect (Fig. 2). With a large α , the tolerance levels decrease slowly and there are numerous steps before the algorithm stops. In this toy example, our simulations show that all explored sets of $(\alpha, p_{acc_{min}})$ such that $p_{acc_{min}} < 0.1$ give good results for the criterion $Number\ of\ simulations \times \mathbb{L}_2^2$ (Fig. 3b). Large α provide slightly better results for small $p_{acc_{min}}$ while small α provide slightly better

¹ http://motive.cemagref.fr/people/maxime.lenormand/script_r_toyex

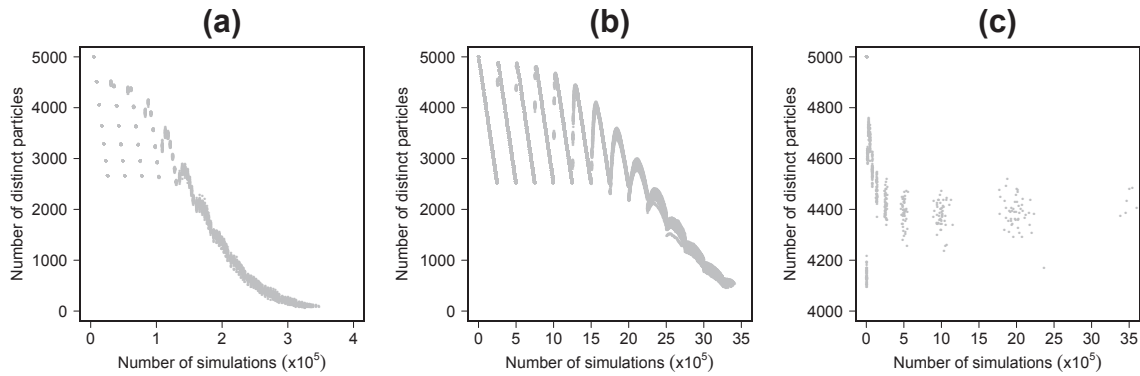


Fig. 1 Number of distinct particles in a sample of $N = 5000$ particles during the course of the SMC and RSMC algorithms applied to the toy example. (a) SMC with $\alpha = 0.9$ and $M = 1$; (b) SMC with $\alpha = 0.99$ and $M = 1$; (c) RSMC with $\alpha = 0.5$. In all three panels, the tolerance target is equal to 0.001.

results for large $p_{acc_{min}}$ (Fig. 3b). On this toy example it appears that intermediate values of α and $p_{acc_{min}}$ ($0.3 \leq \alpha \leq 0.7$ and $0.01 \leq p_{acc_{min}} \leq 0.05$), present a good compromise between number of model runs and the quality of the posterior approximation.

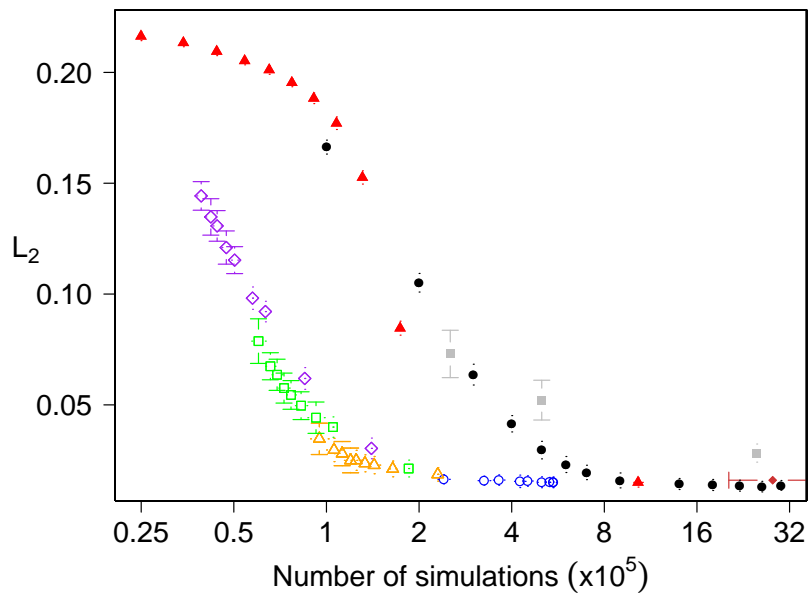


Fig. 2 Posterior quality (\mathbb{L}_2) versus computing cost (number of simulations) averaged over 50 replicates. Vertical and horizontal bars represent the standard deviations among replicates. Algorithm parameters used for APMC: α in $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ and $p_{acc_{min}}$ in $\{0.01, 0.05, 0.1, 0.2\}$. Blue circles are used for $p_{acc_{min}} = 0.01$, orange triangles for $p_{acc_{min}} = 0.05$, green squares for $p_{acc_{min}} = 0.1$, and purple diamonds for $p_{acc_{min}} = 0.2$. PMC: red plain triangles for a sequence of tolerance levels from $\varepsilon_1 = 2$ down to $\varepsilon_{11} = 0.01$. SMC: grey plain square for α in $\{0.9, 0.95, 0.99\}$ (from left to right), $M = 1$ and a ε target equal to 0.01. RSMC: brown plain diamond for $\alpha = 0.5$ and a ε target equal to 0.01. Results obtained with a standard rejection-based ABC algorithm are depicted with black plain circles.

3.3 Comparing performances

Whatever the value of α and $p_{acc_{min}}$, the APMC algorithm always yields better results than the other three algorithms. It requires between 2 and 8 times less simulations to reach a given posterior quality \mathbb{L}_2 (Fig. 2). Furthermore, good approximate posterior distributions are very quickly obtained (Fig. 2). The

compromise between simulation speed and convergence level can also be illustrated using the criterion $\text{Number of simulations} \times \mathbb{L}_2^2$ (Glynn and Whitt 1992). This criterion is smaller for the APMC algorithm (Fig. 3a).

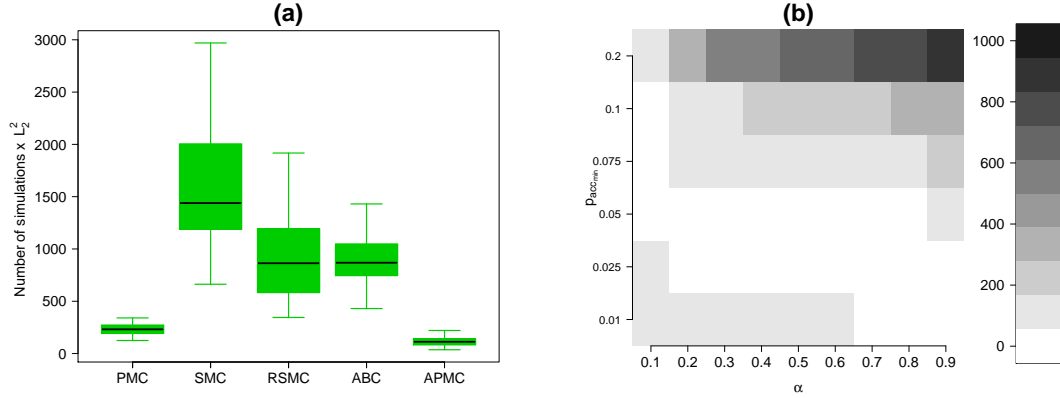


Fig. 3 (a) Boxplot of the criterion “squared \mathbb{L}_2 distance times the number of simulations” for the different ABC algorithms. APMC: for α in $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ and $p_{acc_{min}} = 0.01$; SMC: for α in $\{0.9, 0.95, 0.99\}$, $M = 1$ and a ε target equal to 0.01; RSMC: for $\alpha = 0.5$ and a ε target equal to 0.01; ABC: for a ε target equal to 0.01; PMC: for a sequence of tolerance levels from $\varepsilon_1 = 2$ to $\varepsilon_{11} = 0.01$. (b) Criterion “squared \mathbb{L}_2 distance times the number of simulations” in the APMC algorithm for the different values of α and $p_{acc_{min}}$. Each cell depicts the average of the criterion over the 50 performed replicates of the APMC.

4 Application to the model SimVillages

In this section, we check if our algorithm still performs better than the PMC, the RSMC and the SMC when applied to an individual-based social model developed during the European project PRIMA². The aim of the model is to simulate the effect of a scenario of job creation (or destruction) on the evolution of the population and activities in a network of municipalities.

4.1 Model and data

The model simulates the dynamics of virtual individuals living in 7 interconnected villages in a rural area of Auvergne (a region of Central France). A single run of the model SimVillages with seven rural municipalities takes about 1.4 seconds on a desktop machine (PC Intel 2.83 GHz). The dynamics include demographic change (aging, marriage, divorce, births and deaths), activity change (change of jobs, unemployment, inactivity, retirement), and movings from one municipality to another or outside of the set. The model also includes a dynamics of creation / destruction of jobs of proximity services, derived from the size of the local population. More details on the model can be found in Huet and Deffuant (2011). The individuals (about 3000) are initially generated using the 1990 census data of the National Institute of Statistics and Economic Studies (INSEE), some of them are given a job type and a location for this job (in a municipality of the set or outside), they are organised in households living in a municipality of the set. The model dynamics is mostly data driven, but four parameters cannot be directly derived from the available data. They are noted θ_p for $1 \leq p \leq 4$, described in Table 1.

We use our algorithm to identify the distribution of the four parameters for which the simulations, initialized with the 1990 census data, satisfy matching criteria with the data of the 1999 and 2006 census. The set of summary statistics $\{S_m\}_{1 \leq m \leq M}$ and the associated discrepancy measure used ρ_m are described

² Prototypical policy Impacts on Multifunctional Activities in rural municipalities - EU 7th Framework Research Programme; 2008-2011; <https://prima.cemagref.fr/the-project>

in Table 2. We note S_m the simulated summary statistics and S'_m the observed statistics. The eight summary statistics are normalized (variance equalization) and they are combined using the infinity norm (Eq. 4):

$$\|(\rho_m(S_m, S'_m))_{1 \leq m \leq M}\|_\infty = \sup_{1 \leq m \leq M} \rho_m(S_m, S'_m) \quad (4)$$

We first generate a sample of length N from the prior $\mathcal{U}_{[a,b]}$, where $[a, b]$ is available for each parameter in Table 1, with a Latin hypercube (Carnell 2009) and we select the best N_α particles. To move the particles, we use as kernel transition a multivariate normal distribution parameterized with twice the weighted variance-covariance matrix of the previous sample (Filippi et al. 2011).

As in the section 3, we perform a parameter study and compare APMC with its three competitors. For APMC, α varies in $(\{0.3, 0.5, 0.7\})$ and $p_{acc_{min}}$ in $(\{0.01, 0.05, 0.1, 0.2\})$, and we set $N_\alpha = 5000$ particles. For the PMC, SMC and RSMC we also set $N = 5000$ particles and a tolerance level target equal to 1.4. The tolerance value $\varepsilon = 1.4$ corresponds to the average final tolerance value we obtain with APMC for $p_{acc_{min}} = 0.01$. Note that otherwise this final tolerance is difficult to set properly and a worse choice for this value would have lead to worse performances of these algorithms. For the PMC algorithm, we use the decreasing sequence of tolerance levels $\{3, 2.5, 2, 1.7, 1.4\}$. For the SMC algorithm, we use 3 different values for the couple (α, M) : $\{(0.9, 1), (0.99, 1), (0.9, 15)\}$. For the RSMC algorithm we use $\alpha = 0.5$, as in Drovandi and Pettitt (2011). For each algorithm and parameter setting, we perform 5 replicates.

We approximated posterior density (unknown in this case) with the original rejection-based ABC algorithm, starting with $N = 10,000,000$, selecting 7890 particles below the tolerance level $\varepsilon = 1.4$.

To compute the \mathbb{L}_2 distance between posterior densities, we divided each parameter support into 4 equally sized bins, leading to a grid of $4^4 = 256$ cells, and we computed on this grid the sum of the squared differences between histogram values.

Table 1 SimVillages parameter descriptions

Parameters	Description	Range
θ_1	Average number of children per woman	[0, 4]
θ_2	Probability to accept a new residence for a household	[0, 1]
θ_3	Probability to make couple for two individuals	[0, 1]
θ_4	Probability to split for a couple in a year	[0, 0.5]

Table 2 Summary statistic descriptions

Summary statistic	Description	Measure of discrepancy
S_1	Number of inhabitants in 1999	\mathbb{L}_1 distance
S_2	Age distribution in 1999	χ^2 distance
S_3	Household type distribution in 1999	χ^2 distance
S_4	Net migration in 1999	\mathbb{L}_1 distance
S_5	Number of inhabitants in 2006	\mathbb{L}_1 distance
S_6	Age distribution in 2006	χ^2 distance
S_7	Household type distribution in 2006	χ^2 distance
S_8	Net migration in 2006	\mathbb{L}_1 distance

4.2 Study of APMC result

APMC yields a unimodal approximate posterior distribution for the model SimVillages (Fig. 4). Interestingly, parameters θ_1 and θ_4 are slightly correlated (Fig. 4c). This is logical since they have contradictory effects on the number of child in the population. What is less straightforward is that we are able to partly tease apart these two effects with the available census data, since we get a peak in the approximate posterior distribution instead of a ridge.

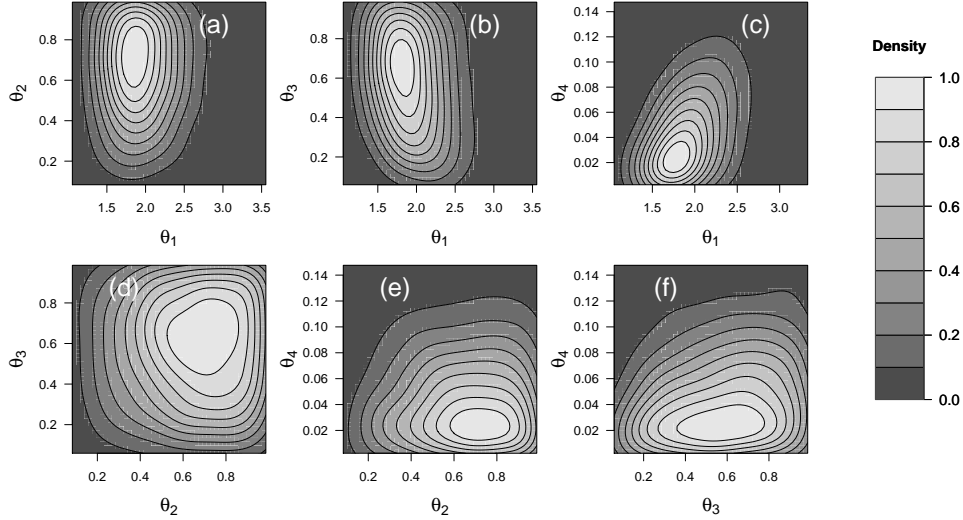


Fig. 4 Contour plot of the bivariate joint densities of θ_i and θ_j obtained with our algorithm, and with $\alpha = 0.5$ and $p_{acc_{min}} = 0.01$; (a) θ_1 and θ_2 ; (b) θ_1 and θ_3 ; (c) θ_1 and θ_4 ; (d) θ_2 and θ_3 ; (e) θ_2 and θ_4 ; (f) θ_3 and θ_4 .

4.3 Influence of parameters on APMC

As for the toy example, we find that the intermediate values of $(\alpha, p_{acc_{min}})$ that we used lead to similar results (Fig. 5c). In practice, we therefore recommend to use $\alpha = 0.5$ and $p_{acc_{min}}$ between 0.01 and 0.05 depending on the wished level of convergence.

4.4 Comparing performances

APMC requires between 2 and 7 times less simulations to reach a given posterior quality than the other algorithms \mathbb{L}_2 (Fig. 5a). Again, the gain in simulation number is progressive during the course of the algorithm. The *Number of simulations* $\times \mathbb{L}_2^2$ criterion is again smaller for the APMC algorithm (Fig. 5b).

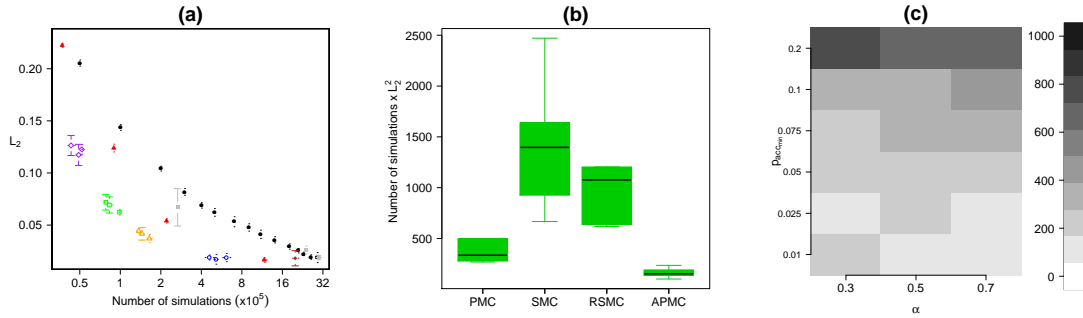


Fig. 5 (a) Posterior quality (\mathbb{L}_2) versus computing cost (number of simulations) averaged over 5 replicates. Vertical and horizontal bars represent the standard deviations among replicates. Algorithm parameters used for APMC: α in $\{0.3, 0.5, 0.7\}$ and $p_{acc_{min}}$ in $\{0.01, 0.05, 0.1, 0.2\}$. Blue circles are used for $p_{acc_{min}} = 0.01$, orange triangles for $p_{acc_{min}} = 0.05$, green squares for $p_{acc_{min}} = 0.1$, and purple diamonds for $p_{acc_{min}} = 0.2$. PMC: red plain triangles for a sequence of tolerance levels from $\varepsilon_1 = 3$ to $\varepsilon_5 = 1.4$. SMC: grey plain square for (α, M) in $\{(0.9, 1), (0.99, 1)\}$, grey star for $(\alpha, M) = (0.9, 15)$ and a ε target equal to 1.4. RSMC: brown plain diamond for $\alpha = 0.5$ and a ε target equal to 1.4. Results obtained with a standard rejection-based ABC algorithm are depicted with black plain circles. (b) Boxplot of the criterion “squared \mathbb{L}_2 distance times the number of simulations” for the different algorithms. APMC: for α in $\{0.3, 0.5, 0.7\}$ and $p_{acc_{min}} = 0.01$; SMC: for (α, M) in $\{(0.9, 1), (0.99, 1), (0.9, 15)\}$ and a ε target equal to 0.01; RSMC: for $\alpha = 0.5$ and a ε target equal to 0.01; ABC: for a ε target equal to 1.4; PMC: for a sequence of tolerance levels from $\varepsilon_1 = 3$ to $\varepsilon_5 = 1.4$. (c) Criterion “squared \mathbb{L}_2 distance times the number of simulations” in the APMC algorithm for the different values of α and $p_{acc_{min}}$. Each cell depicts the average of the criterion over the 5 performed replicates of the APMC.

5 Discussion

The good performances of APMC should of course be confirmed on other examples. Nevertheless we argue that they are due to the main assets of our approach:

- We choose an appropriate reweighting process instead of a MCMC kernel, which corrects the sampling bias without duplicating particles;
- We define an easy to interpret stopping criterion that automatically defines the number of sequential steps.

Therefore, we can have some confidence in the good performances of APMC on other examples.

In the future, it would be interesting to evaluate this algorithm on models involving a larger number of parameters and/or multi-modal posterior distributions. Moreover, APMC could benefit from other improvements, in particular by performing a semi-automatic selection of informative summary statistics after the first ABC step (Joyce and Marjoram 2008; Fearnhead and Prangle 2011) and by using local regressions for post-processing the final posterior distribution (Beaumont et al. 2002; Blum and François 2010). We did not perform such combinations in the present contribution, so that our algorithm is directly comparable with the three other sequential algorithms we looked at. However, they would be straightforward, because the different improvements concern different steps of the ABC procedure.

Appendix A: Description of the algorithms

Algorithm 1 Likelihood-free rejection sampler (ABC)

```

Given  $N$  the number of particles
for  $i = 1$  to  $N$  do
  repeat
    Generate  $\theta^* \sim \pi(\theta)$ 
    Simulate  $x \sim f(x|\theta^*)$ 
  until  $\rho(S(x), S(y)) < \varepsilon$ 
  Set  $\theta_i = \theta^*$ 
end for

```

Algorithm 2 Population Monte Carlo Approximate Bayesian Computation (PMC)

```

Given  $N$  the number of particles and a decreasing sequence of tolerance level  $\varepsilon_1 \geq \dots \geq \varepsilon_T$ ,
For  $t = 1$ ,
for  $i = 1$  to  $N$  do
  repeat
    Simulate  $\theta_i^{(1)} \sim \pi(\theta)$  and  $x \sim f(x|\theta_i^{(1)})$ 
  until  $\rho(S(x), S(y)) < \varepsilon_1$ 
  Set  $w_i^{(1)} = \frac{1}{N}$ 
end for
Take  $\sigma_2^2$  as twice the weighted empirical variance of  $(\theta_i^{(1)})_{1 \leq i \leq N}$ 
for  $t = 2$  to  $T$  do
  for  $i = 1$  to  $N$  do
    repeat
      Sample  $\theta_j^*$  from  $\theta_j^{(t-1)}$  with probabilities  $w_j^{(t-1)}$ 
      Generate  $\theta_i^{(t)} | \theta_j^* \sim \mathcal{N}(\theta_j^*, \sigma_i^2)$  and  $x \sim f(x|\theta_i^{(t)})$ 
    until  $\rho(S(x), S(y)) < \varepsilon_t$ 
    Set  $w_i^{(t)} \propto \frac{\pi(\theta_i^{(t)})}{\sum_{j=1}^N w_j^{(t-1)} \sigma_i^{-1} \varphi(\sigma_i^{-1}(\theta_i^{(t)} - \theta_j^{(t-1)}))}$ 
  end for
  Take  $\sigma_{t+1}^2$  as twice the weighted empirical variance of  $(\theta_i^{(t)})_{1 \leq i \leq N}$ 
end for
Where  $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ 

```

Algorithm 3 Sequential Monte Carlo Approximate Bayesian Computation Replenishment (RSMC)

Given $N, \varepsilon_1, \varepsilon_T, c, \alpha \in [0, 1]$ and $N_\alpha = \lfloor \alpha N \rfloor$,

for $i = 1$ to N **do**

repeat

 Simulate $\theta_i \sim \pi(\theta)$ and $x \sim f(x|\theta_i)$

$\rho_i = \rho(S(x), S(y))$

until $\rho_i \leq \varepsilon_1$

end for

Sort (θ_i, ρ_i) by ρ_i

Set $\varepsilon_{MAX} = \rho_N$

while $\varepsilon_{MAX} > \varepsilon_T$ **do**

 Remove the N_α particles with largest ρ

 Set $\varepsilon_{NEXT} = \rho_{N-N_\alpha}$

 Set $i_{acc} = 0$

 Compute the parameters of the proposal *MCMC* $q(\cdot, \cdot)$ with the $N - N_\alpha$ particles.

for $j = 1$ to N_α **do**

 Simulate $\theta_{N-N_\alpha+j} \sim (\theta_i)_{1 \leq i \leq N-N_\alpha}$

for $k = 1$ to R **do**

 Generate $\theta^* \sim q(\theta^*, \theta_{N-N_\alpha+j})$ et $x^* \sim f(x^*|\theta^*)$

 Generate $u \sim \mathcal{U}_{[0,1]}$

if $u \leq 1 \wedge \frac{\pi(\theta^*)q(\theta_{N-N_\alpha+j}, \theta^*)}{\pi(\theta_{N-N_\alpha+j})q(\theta^*, \theta_{N-N_\alpha+j})} \mathbb{1}_{\rho(S(x^*), S(y)) \leq \varepsilon_{NEXT}}$ **then**

 Set $\theta_{N-N_\alpha+j} = \theta^*$

 Set $\rho_{N-N_\alpha+j} = \rho(S(x^*), S(y))$

$i_{acc} \leftarrow i_{acc} + 1$

end if

end for

end for

 Set $p_{acc} = \frac{i_{acc}}{RN_\alpha}$

 Set $R = \frac{\log(c)}{\log(1 - p_{acc})}$

end while

Algorithm 4 Adaptive Sequential Monte Carlo Approximate Bayesian Computation (SMC)

Given $N, M, \alpha \in [0, 1], \varepsilon_0 = \infty, \varepsilon$ and N_T ,

For $t = 0$,

for $i = 1$ to N **do**

 Simulate $\theta_i^{(0)} \sim \pi(\theta)$

for $k = 1$ to M **do**

 Simulate $X_{(i,k)}^{(0)} \sim f(\cdot | \theta_i^{(0)})$

end for

 Set $W_i^{(0)} = \frac{1}{N}$

end for

We have $ESS((W_i^{(0)}), \varepsilon_0) = N$ where $ESS((W_i^{(0)}), \varepsilon_0) = \left(\sum_{i=1}^N (W_i^{(0)})^2\right)^{-1}$

Set $t = 1$

while $\varepsilon_{t-1} > \varepsilon$ **do**

 Determine ε_t resolving $ESS((W_i^{(t)}), \varepsilon_t) = \alpha ESS((W_i^{(t-1)}), \varepsilon_{t-1})$ where $W_i^{(t)} \propto W_i^{(t-1)} \frac{\sum_{k=1}^M \mathbb{1}_{A_{\varepsilon_{t-1}, y}}(X_{(i,k)}^{(t-1)})}{\sum_{k=1}^M \mathbb{1}_{A_{\varepsilon_{t-1}, y}}(X_{(i,k)}^{(t-1)})}$ et $A_{\varepsilon, y} =$

$\{x | \rho(S(x), S(y)) < \varepsilon\}$

if $\varepsilon_t < \varepsilon$ **then**

$\varepsilon_t = \varepsilon$

end if

if $ESS((W_i^{(t)}), \varepsilon_t) < N_T$ **then**

for $i = 1$ to N **do**

 Simulate $(\theta_{(i)}^{(t-1)}, X_{(i,1:M)}^{(t-1)})$ in $(\theta_{(j)}^{(t-1)}, X_{(j,1:M)}^{(t-1)})$ with probabilities $W_j^{(t-1)}, 1 \leq j \leq N$

 Set $W_i^{(t)} = \frac{1}{N}$

end for

end if

for $t = 1$ to N **do**

if $W_j^{(t)} > 0$ **then**

 Generate $\theta^* \sim K(\theta^* | \theta_{(i)}^{(t-1)})$

for $k = 1$ to M **do**

 Simulate $X_{(*,k)} \sim f(\cdot | \theta^*)$

end for

 Generate $u < \mathcal{U}_{[0,1]}$

if $u \leq 1 \wedge \frac{\sum_{k=1}^M \mathbb{1}_{A_{\varepsilon_t, y}}(X_{(*,k)}) \pi(\theta^*) K_t(\theta_{(i)}^{(t-1)} | \theta^*)}{\sum_{k=1}^M \mathbb{1}_{A_{\varepsilon_t, y}}(X_{(i,k)}^{(t-1)}) \pi(\theta_{(i)}^{(t-1)}) K_t(\theta^* | \theta_{(i)}^{(t-1)})}$ **then**

 Set $(\theta_{(i)}^{(t)}, X_{(i,1:M)}^{(t)}) = (\theta^*, X_{(*,1:M)})$

else

 Set $(\theta_{(i)}^{(t)}, X_{(i,1:M)}^{(t)}) = (\theta_{(i)}^{(t-1)}, X_{(i,1:M)}^{(t-1)})$

end if

end if

end for

end while

Algorithm 5 Adaptive Population Monte Carlo Approximate Bayesian Computation

Given N , $N_\alpha = \lfloor \alpha N \rfloor$ the number of particles to keep at each iteration among the N particles ($\alpha \in [0, 1]$) and $p_{acc_{min}}$ the minimal acceptance rate.

for $t = 1$ **do**

for $i = 1$ to N **do**

 Simulate $\theta_i^{(0)} \sim \pi(\theta)$ and $x \sim f(x|\theta_i^{(0)})$

 Set $\rho_i^{(0)} = \rho(S(x), S(y))$

 Set $w_i^{(0)} = 1$

end for

 Let $\varepsilon_1 = Q_{\rho^{(0)}}(\alpha)$ the first α -quantile of $\rho^{(0)}$ where $\rho^{(0)} = \{\rho_i^{(0)}\}_{1 \leq i \leq N}$

 Let $\{(\theta_i^{(1)}, w_i^{(1)}, \rho_i^{(1)})\} = \{(\theta_i^{(0)}, w_i^{(0)}, \rho_i^{(0)}) | \rho_i^{(0)} \leq \varepsilon_1, 1 \leq i \leq N\}$

 Take σ_1^2 as twice the weighted empirical variance of $\{(\theta_i^{(1)}, w_i^{(1)})\}_{1 \leq i \leq N_\alpha}$

 Set $p_{acc} = 1$

$t \leftarrow t + 1$

end for

while $p_{acc} > p_{acc_{min}}$ **do**

for $i = N_\alpha + 1$ to N **do**

 Pick θ_j^* from $\theta_j^{(t-1)}$ with probability $\frac{w_j^{(t-1)}}{\sum_{k=1}^{N_\alpha} w_k^{(t-1)}}$, $1 \leq j \leq N_\alpha$

 Generate $\theta_i^{(t-1)} | \theta_i^* \sim \mathcal{N}(\theta_i^*, \sigma_{(t-1)}^2)$ and $x \sim f(x|\theta_i^{(t-1)})$

 Set $\rho_i^{(t-1)} = \rho(S(x), S(y))$

 Set $w_i^{(t-1)} = \frac{\pi(\theta_i^{(t-1)})}{\sum_{j=1}^{N_\alpha} (w_j^{(t-1)} / \sum_{k=1}^{N_\alpha} w_k^{(t-1)}) \sigma_{t-1}^{-1} \varphi(\sigma_{t-1}^{-1} (\theta_i^{(t-1)} - \theta_j^{(t-1)}))}$

end for

 Set $p_{acc} = \frac{1}{N - N_\alpha} \sum_{k=N_\alpha+1}^N \mathbb{1}_{\rho_i^{(t-1)} < \varepsilon_{t-1}}$

 Let $\varepsilon_t = Q_{\rho^{(t-1)}}(\alpha)$ where $\rho^{(t-1)} = \{\rho_i^{(t-1)}\}_{1 \leq i \leq N}$

 Let $\{(\theta_i^{(t)}, w_i^{(t)}, \rho_i^{(t)})\} = \{(\theta_i^{(t-1)}, w_i^{(t-1)}, \rho_i^{(t-1)}) | \rho_i^{(t-1)} \leq \varepsilon_t, 1 \leq i \leq N\}$

 Take σ_t^2 as twice the weighted empirical variance of $\{(\theta_i^{(t)}, w_i^{(t)})\}_{1 \leq i \leq N_\alpha}$

$t \leftarrow t + 1$

end while

Where $\forall u \in [0, 1]$ and $X = \{x_1, \dots, x_n\}$, $Q_X(u) = \inf\{x \in X | F_X(x) \geq u\}$ and $F_X(x) = \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{x_k \leq x}$.

Where $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

Appendix B: Proof that the algorithm stops

We know that there exists $\varepsilon_\infty > 0$ such that $\varepsilon_t \xrightarrow[t \rightarrow +\infty]{} \varepsilon_\infty$ because, by construction of the algorithm (ε_t) is a positive decreasing sequence and it is bounded by 0.

For each $\theta \in \Theta$, we consider the distance $(\rho(x, y) | \theta)$ as a random variable $\rho(\theta)$. Let $f_{\rho(\theta)}$ be the probability density function of $\rho(\theta)$.

The probability $\mathbb{P}[\rho(\theta) \geq \varepsilon_t]$ that the drawn distance associated to parameter θ is higher than the current tolerance ε_t satisfies:

$$\begin{aligned} \mathbb{P}[\rho(\theta) \geq \varepsilon_t] &= 1 - \mathbb{P}[(\rho(\theta) < \varepsilon_t)] \\ &= 1 - \int_{\varepsilon_\infty}^{\varepsilon_t} f_{\rho(\theta)}(x) dx \end{aligned}$$

We define:

$$\mathbb{P}_{max} = \sup_{\theta \in \Theta} \left\{ \sup_{x \in \mathbb{R}^+} \{f_{\rho(\theta)}(x)\} \right\}$$

We have:

$$\mathbb{P}[\rho(\theta) \geq \varepsilon_t] \geq 1 - \mathbb{P}_{max}(\varepsilon_t - \varepsilon_\infty)$$

The $N - N_\alpha$ particles are independent and identically distributed from π_{t+1} the density defined by the algorithm, hence the probability $\mathbb{P}[p_{acc}(t+1) = 0]$ that no particle is accepted at step $t+1$ is such that:

$$\mathbb{P}[p_{acc}(t+1) = 0] \geq (1 - \mathbb{P}_{max}(\varepsilon_t - \varepsilon_\infty))^{N - N_\alpha}$$

If $\mathbb{P}_{max} < +\infty$, because $\varepsilon_t - \varepsilon_\infty \xrightarrow[t \rightarrow +\infty]{} 0$, we have:

$$\mathbb{P}[p_{acc}(t+1) = 0] \xrightarrow[t \rightarrow +\infty]{} 1$$

We can conclude that $p_{acc}(t)$ converges in probability towards 0 if $\mathbb{P}_{max} < +\infty$. This ensures that the algorithm stops, whatever the chosen value of p_{accmin} .

Acknowledgement

This publication has been funded by the Prototypical policy impacts on multifunctional activities in rural municipalities collaborative project, European Union 7th Framework Programme (ENV 2007-1), contract no. 212345. The work of the first author has been funded by the Auvergne region.

References

- Beaumont, M. A. (2010). *Approximate Bayesian computation in evolution and ecology*, volume 41 of *Annual Review of Ecology, Evolution, and Systematics*.
- Beaumont, M. A., Cornuet, J., Marin, J., and Robert, C. P. (2009). Adaptive approximate Bayesian computation. *Biometrika*, 96(4):983–990.
- Beaumont, M. A., Zhang, W., and Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, 162(4):2025–2035.
- Blum, M. G. B. and François, O. (2010). Non-linear regression models for approximate Bayesian computation. *Statistics and Computing*, 20(1):63–73.
- Carnell, R. (2009). lhs: Latin hypercube samples. *R package version 0.5*.
- Del Moral, P., Doucet, A., and Jasra, A. (2006). Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 68(3):411–436.
- Del Moral, P., Doucet, A., and Jasra, A. (2012). An adaptive sequential Monte Carlo method for approximate Bayesian computation. *Statistics and Computing*. Article in Press.
- Drovandi, C. C. and Pettitt, A. N. (2011). Estimation of parameters for macroparasite population evolution using approximate Bayesian computation. *Biometrics*, 67(1):225–233.
- Fearnhead, P. and Prangle, D. (2011). Constructing summary statistics for approximate Bayesian computation: Semi-automatic ABC. Technical Report 1004.1112, arXiv.org.
- Filippi, S., Barnes, C., and Stumpf, M. P. H. (2011). On optimal kernel in ABC-SMC.
- Glynn, P. and Whitt, W. (1992). The asymptotic efficiency of simulation estimators. *Oper. Res.*, 40(3):505–520.
- Huet, S. and Deffuant, G. (2011). Common framework for the microsimulation model in prima project. Technical report, Cemagref LISC.
- Joyce, P. and Marjoram, P. (2008). Approximately sufficient statistics and Bayesian computation. *Statistical Applications in Genetics and Molecular Biology*, 7(1).
- Marjoram, P., Molitor, J., Plagnol, V., and Tavaré, S. (2003). Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America*, 100(26):15324–15328.
- R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Sisson, S. A., Fan, Y., and Tanaka, M. M. (2007). Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America*, 104(6):1760–1765.
- Toni, T., Welch, D., Strelkowa, N., Ipsen, A., and Stumpf, M. P. H. (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, 6:187.
- Wegmann, D., Leuenberger, C., and Excoffier, L. (2009). Efficient approximate bayesian computation coupled with markov chain monte carlo without likelihood. *Genetics*, 182(4):1207–1218.