Adaptivity for ABC algorithms: the ABC-PMC scheme

Christian P. Robert
CEREMADE, Université Paris Dauphine, and CREST, INSEE
Mark A. Beaumont
School of Biological Sciences, University of Reading
Jean-Michel Marin
INRIA Saclay, Projet select, Université Paris-Sud and CREST, INSEE
Jean-Marie Cornuet
Department of Epidemiology and Public Health, Imperial College,
and Centre de Biologie et de Gestion des Populations (INRA)

Abstract. Sequential techniques can be adapted to the ABC algorithm to enhance its efficiency. For instance, when Sisson et al. (2007) introduced the ABC-PRC algorithm, the goal was to improve upon existing ABC-MCMC algorithms (Marjoram et al., 2003). While the ABC-PRC method is based upon the theoretical developments of Del Moral et al. (2006), the application to the setting of approximate Bayesian computation induces a bias in the approximation to the posterior distribution of interest, as we demonstrate in this paper via both theoretical reasoning and experimental results. It is however possible to devise an alternative version based on genuine importance sampling arguments that we call ABC-PMC in connection with the population Monte Carlo method introduced in Cappé et al. (2004). Furthermore, the ABC-PMC algorithm is simpler than the ABC-PRC algorithm in that it does not require any backward transition kernel and proposes an automatic scaling of the forward kernel. In this paper, we demonstrate the applicability of ABC-PMC and compare its performances with ABC-PRC.

Keywords: Approximate Bayesian computation, ABC-PRC, importance sampling, population Monte Carlo, sequential Monte Carlo.

1 Introduction

When the likelihood function is not available in a closed form, as in population Genetics, approximate Bayesian computational (ABC) methods have been introduced (Pritchard et al., 1999; Beaumont et al., 2002) as a rejection technique bypassing the computation of the likelihood function via a simulation from the corresponding distribution. Namely, if we observe $y \sim f(y|\theta)$ and if $\pi(\theta)$ is the prior distribution on the parameter θ, then the original ABC algorithm jointly simulates

$$\theta' \sim \pi(\theta) \qquad \text{and} \qquad x \sim f(x|\theta')$$

and accept the simulated θ' if and only if the auxiliary variable x is equal to the observed value, $x = y$. This algorithm is clearly legitimate in that the accepted θ''s are exactly distributed from the posterior. In the (more standard) event that y is a continuous

random variable, the ABC algorithm uses an approximation, replacing the strict equality $x = y$ with a tolerance zone, $\varrho(x, y) \leq \epsilon$, ϱ being a measure of discrepancy (like a distance between summary statistics) and ϵ a small enough number. The output is then distributed from the distribution with density proportional to

$$\pi(\theta)\, \mathbb{P}_\theta(\varrho(x, y) < \epsilon)$$

where $\mathbb{P}_\theta$ represents the distribution of x conditional on the value of θ. (This distribution is often summarised by $\pi(\theta|\varrho(x, y) < \epsilon)$, which is to be understood under the *marginal* distribution of $\varrho(x, y) < \epsilon$.) Improvements to this general scheme have this far been achieved in two ways: either by modifying the proposal distribution of the parameter θ to increase the density of x within the vicinity of y (Marjoram et al., 2003; Sisson et al., 2007); or by viewing the problem as one of conditional density estimation and developing techniques to allow for larger ϵ (Beaumont et al., 2002).

Sisson et al. (2007) have introduced a sequential Monte Carlo method called the ABC-PRC algorithm (where PRC stands for partial rejection control, as introduced in Liu (2001)). The simulation method is sequential (see, e.g., Robert and Casella, 2004, Chapter 14) in that simulated populations of N points (sometimes called *particles*) are generated at each iteration of the algorithm and that they are exploited to produce better proposals for a given target distribution. As demonstrated in, e.g., Douc et al. (2007), the reliance on earlier populations to build proposals is perfectly legitimate from a convergence point of view as long as an importance sampling perspective is adopted, and a progressive improvement in the choice of proposals is the appeal of using a sequence of samples rather than a single one, since Douc et al. (2007) establish that iterating the simulation of samples without modifying the proposal does not bring an improvement in the Kullback divergence between the target and the proposal distribution.

Marjoram et al. (2003) enjoys the same validity as the original ABC algorithm, namely that, if a Markov chain $(\theta^{(t)})$ is created via the transition function

$$\theta^{(t+1)} = \begin{cases} \theta' \sim K(\theta'|\theta^{(t)}) & \text{if } x \sim f(x|\theta') \text{ is such that } x = y \\ & \text{and } u \sim \mathcal{U}(0,1) \leq \frac{\pi(\theta')K(\theta^{(t)}|\theta')}{\pi(\theta^{(t)})K(\theta'|\theta^{(t)})}\,, \\ \theta^{(t)} & \text{otherwise,} \end{cases}$$

the stationary distribution of the Markov chain is the true posterior $\pi(\theta|y)$. Once again, in most situations, the distribution of y is absolutely continuous and the strict constraint $x = y$ is replaced with the approximation $\varrho(x, y) < \epsilon$.

Example 1.1. When

$$\theta \sim \mathcal{U}(-10.10)\,, \quad x|\theta \sim \frac{1}{2}\mathcal{N}(\theta, 1) + \frac{1}{2}\mathcal{N}(\theta, 1/100)\,,$$

as studied in Sisson et al. (2007), the posterior distribution associated with $y = 0$ is the normal mixture

$$\theta|y = 0 \sim \frac{1}{2}\mathcal{N}(0, 1) + \frac{1}{2}\mathcal{N}(0, 1/100)$$

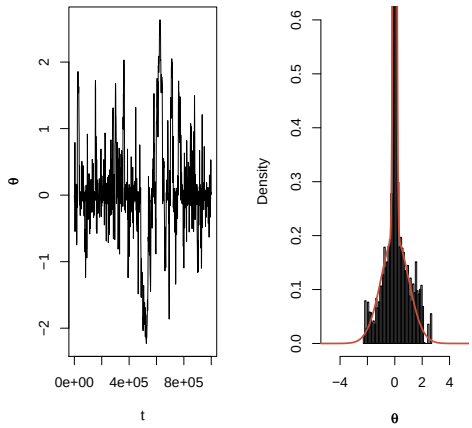

Figure 1: Output *(left)* and histogram (*(right)* of a sample produced by an ABC-MCMC algorithm, for the mixture model of Example 1.1, $T = 10^6$ iterations, $\epsilon = 0.025$ and a scale of $\tau = .15$. (*Note:* The exact posterior density plotted on top of the histogram is identical to the target of the simulation algorithm, $\pi(\theta|\varrho(x, y) < \epsilon)$.)

restricted to the set $[-10, 10]$. As in regular MCMC settings, the performances of the ABC-MCMC algorithm depend on the choice of the scale τ in the random walk proposal, $K(\theta'|\theta) = \tau^{-1}\varphi(\tau^{-1}(\theta-\theta'))$. However, even when $\tau = 0.15$ as in Sisson et al. (2007), the Markov chain mixes slowly, but still produces an acceptable fit over $T = 10^6$ iterations, as shown in Figure 1. We further note that, in this toy example, the true target is available as

$$\pi(\theta||x| < \epsilon) \propto \Phi(\epsilon - \theta) - \Phi(-\epsilon - \theta) + \Phi(0.1(\epsilon - \theta)) - \Phi(-0.1(\epsilon + \theta))\,.$$

It is therefore possible to check that, for the value $\epsilon = 0.025$, the true target is identical with the exact posterior density $\pi(\theta|y = 0)$ and we can thus clearly separate the issue of poor convergence of the algorithm (related with the choice of τ) from the issue of approximating the posterior density (related with the choice of ϵ). ◀

The ABC-PRC modification introduced in Sisson et al. (2007) consists in producing samples $(\theta_1^{(t)}, \ldots, \theta_N^{(t)})$ at each iteration $1 \leq t \leq T$ of the algorithm by using [except when $t = 1$ in which case a regular ABC step is implemented] Markov transition kernels K_t for the generation of the $\theta_i^{(t)}$'s, namely

$$\theta_i^{(t)} \sim K_t(\theta|\theta^\star)\,,$$

until $x \sim f(x|\theta_i^{(t)})$ is such that $\varrho(x, y) < \epsilon$, where $\theta^\star$ is selected at random among the previous $\theta_i^{(t-1)}$'s with probabilities $\omega_i^{(t-1)}$. The probability $\omega_i^{(t)}$ is derived by an

importance sampling argument as ($t > 1$)

$$\omega_i^{(t)} \propto \frac{\pi(\theta_i^{(t)}) L_{t-1}(\theta^\star|\theta_i^{(t)})}{\pi(\theta^\star) K_t(\theta_i^{(t)}|\theta^\star)}\,,$$

where L_{t-1} is an arbitrary transition kernel. (Sisson et al. (2007), suggest using $L_{t-1}(\theta'|\theta) = K_t(\theta|\theta')$ equal to a Gaussian kernel, which means that the weights are then all equal under a uniform prior π.) This importance ratio is inspired from Del Moral et al. (2006) who used a backward kernel L_{t-1} (or more exactly a sequence of backward kernels) in a sequential Monte Carlo algorithm to achieve unbiasedness in the marginal distribution of the current particle without computing this (unavailable) marginal.

In this paper, we analyse the properties of the above ABC-PRC techniques and show in the following section via both theoretical and experimental arguments that this algorithm is biased. Moreover, we introduce a new algorithm called ABC-PRC in connection with the population Monte Carlo (PMC) method of Cappé et al. (2004). This correction is based on genuine importance sampling arguments and the section after next demonstrates its applicability as well as the improvement it brings compqared with ABC-PRC.

2 Bias in the ABC-PRC algorithm

2.1 Distribution of the ABC-PRC sample

In order to expose the difficulty in using the ABC-PRC weights as given in Sisson et al. (2007), we first consider the ideal and limiting case when $\epsilon = 0$. (In that case, we recall that both ABC and ABC-MCMC algorithms are correct samplers from $\pi(\theta|y)$.) This means we generate $\theta' \sim K_t(\theta|\theta^\star)$ and $x \sim f(x|\theta')$ until $x = y$. We now consider the distribution of the weighted $\theta_i^{(t)}$'s when those are simulated and weighted according to ABC-PRC. To evaluate the bias resulting from using the ABC-PRC weight in the second step, let us further assume without loss of generality that the $\theta_i^{(t-1)}$'s have been resampled using proper weights, i.e. that $\theta^\star \sim \pi(\theta|y)$. Then [denoting by $\theta^{(t-1)}$ the selected $\theta^\star$] the joint density of $(\theta^{(t-1)}, \theta^{(t)})$ is proportional to

$$\pi(\theta^{(t-1)}|y) K_t(\theta^{(t)}|\theta^{(t-1)}) f(y|\theta^{(t)})$$

with a marginalisation constant that only depends on y [and on the choice of K_t]. Therefore, if we use the weight ω_t proposed by Sisson et al. (2007) in PRC2.2, the weighted distribution of $\theta^{(t)}$ is such that, for an arbitrary integrable function $h(\theta)$, $\mathbb{E}[h(\theta^{(t)})\omega_t]$ is proportional to

$$\begin{aligned}
&\iint h(\theta^{(t)}) \frac{\pi(\theta^{(t)}) L_{t-1}(\theta^{(t-1)}|\theta^{(t)})}{\pi(\theta^{(t-1)}) K_t(\theta^{(t)}|\theta^{(t-1)})} \\
&\qquad \times \pi(\theta^{(t-1)}|y) K_t(\theta^{(t)}|\theta^{(t-1)}) f(y|\theta^{(t)}) \mathrm{d}\theta^{(t-1)} \mathrm{d}\theta^{(t)} \\
\propto \quad &\iint h(\theta^{(t)}) \frac{\pi(\theta^{(t)}) L_{t-1}(\theta^{(t-1)}|\theta^{(t)})}{\pi(\theta^{(t-1)}) K_t(\theta^{(t)}|\theta^{(t-1)})} \pi(\theta^{(t-1)}) f(y|\theta^{(t-1)})
\end{aligned}$$

$$\begin{aligned} & \times K_t(\theta^{(t)}|\theta^{(t-1)}) f(y|\theta^{(t)}) \mathrm{d}\theta^{(t-1)} \mathrm{d}\theta^{(t)} \\ \propto & \int h(\theta^{(t)}) \pi(\theta^{(t)}|y) \\ & \times \left\{ \int L_{t-1}(\theta^{(t-1)}|\theta^{(t)}) f(y|\theta^{(t-1)}) \mathrm{d}\theta^{(t-1)} \right\} \mathrm{d}\theta^{(t)} \end{aligned}$$

[with all proportionality terms being functions of y only]. If the weight was unbiased we should obtain

$$\mathbb{E}[h(\theta^{(t)})\omega_t] = \int h(\theta^{(t)}) \pi(\theta^{(t)}|y) \mathrm{d}\theta^{(t)} \,,$$

therefore we can conclude that there is a bias in the weight proposed by Sisson et al. (2007) unless

$$L_{t-1}(\theta^{(t-1)}|\theta^{(t)}) f(y|\theta^{(t-1)})$$

integrates to the same constant for all values of $\theta^{(t)}$. Apart from this special case—that is achievable when $L_{t-1}(\theta^{(t-1)}|\theta^{(t)}) = g(\theta^{(t-1)})$ but not in the random walk type proposal, i.e. when $L_{t-1}(\theta^{(t-1)}|\theta^{(t)}) = \varphi(\theta^{(t-1)} - \theta^{(t)})$—, the ABC-PRC weight is thus incorrect.

Paradoxically, the weight used in the ABC-PRC algorithm misses a $f(y|\theta^{(t-1)})$ term in its denominator, while the method is used when $f(y|\theta)$ is not available. This is exactly the difference between the weights used in Sisson et al. (2007) and those used in Del Moral et al. (2006), namely that, in the latter paper, the posterior $\pi(\theta^{(t-1)}|x)$ explicitly appears in the denominator instead of the prior. The accept-reject principle at the core of ABC allows for the replacement of the posterior by the prior in the numerator of the ratio, but not in the denominator.

In order to illustrate the practical effect of this bias in the weight of Sisson et al. (2007), we first consider a toy situation based on a discrete distribution, namely the Beta-binomial case.

Example 2.1. Here, $f(y|\theta)$ is the density of a binomial $\mathcal{B}(n, \theta)$ distribution and we choose $\pi(\theta)$ to be the constant density of a $\mathcal{U}(0, 1)$ distribution. Since the support of $f(y|\theta)$ is finite, we can implement the ideal exact ABC algorithm (i.e. accepting only when the simulated x is identical to the observed y) to produce a sample simulated from the true posterior, which is then equal to a $\mathcal{B}e(y + 1, n - y + 1)$ distribution (see, e.g., Robert, 2001).

When implementing the ABC-PRC algorithm of Sisson et al. (2007), the initial importance sampling distribution can be chosen to be equal to the prior distribution, i.e. $\mu_1 = \pi$. Then, the first sample $\theta_i^{(1)}$ is exactly distributed from the true posterior $\mathcal{B}e(y+1, n-y+1)$ since the importance weights are equal to 1 and since the acceptance step has probability $f(y|\theta)$ to occur. For the following ABC-PRC steps ($t \geq 2$), we use for K_t the random walk proposal of Sisson et al. (2007), except that we first operate a logistic change of variables to account for the fact that the θ's are restricted to vary between 0 and 1. This means that $K_t(\theta|\theta^\star)$ is a normal distribution on the logistic

transform of θ:

$$K_t(\theta|\theta^\star) = \frac{1}{\sqrt{2\pi}\sigma_t} \exp\left\{-(\log\{\theta/(1-\theta)\} - \log\{\theta^\star/(1-\theta^\star)\})^2/2\sigma_t^2\right\} \frac{1}{\theta(1-\theta)},$$

the final fraction in the above being the Jacobian due to the change of variable. In order to reproduce the adaptive features of the ABC-PRC algorithm, we also use a sequence σ_t of standard deviations based on twice the (weighted) empirical variance of the previous sample of $\eta_i^{(t-1)} = \log\{\theta_i^{(t-1)}/(1-\theta_i^{(t-1)})\}$. (The factor 2 corresponds to the optimal choice of scale in terms of Kullback-Leibler divergence for a random walk.) Following the recommendation in Sisson et al. (2007), we take the backward kernel $L_{t-1}(\theta|\theta^\star)$ to be the same normal distribution on the logit scale, which means that

$$\omega_i^{(t)} = \frac{\pi(\theta_i^{(t)})L_{t-1}(\theta^\star|\theta_i^{(t)})}{\pi(\theta^\star)K_t(\theta_i^{(t)}|\theta^\star)} = \frac{L_{t-1}(\theta^\star|\theta_i^{(t)})}{K_t(\theta_i^{(t)}|\theta^\star)} = \frac{\theta_i^{(t)}(1-\theta_i^{(t)})}{\theta^\star(1-\theta^\star)},$$

equal to the ratio of the Jacobians, except for $t = 1$. (Note that the ratio is naturally invariant by a change of variables and that it is thus the same whether it is expressed in terms of the $\theta^{(t)}$'s or in terms of the $\eta^{(t)}$'s.)

Figure 2 monitors the histograms of the consecutive samples produced by ABC-PRC against the graph of the true posterior distribution $\mathcal{B}e(y+1, n-y+1)$ when $y = 3$ and $n = 7$. As clearly shown by this figure, the fit after the (exact) first step deteriorates along iterations. ◀

Quite obviously, the bias does not vanish along iterations since there is a factor similar to

$$\int L_{t-1}(\theta^{(t-1)}|\theta^{(t)}) f(y|\theta^{(t-1)}) \mathrm{d}\theta^{(t-1)}$$

appearing at each iteration. Figure 2 shows that the effect of this cumulative bias tends to level off with iterations, rather than increasing with T as it would if only an additional $f(y|\theta)$ factor would appear at each iteration.

2.2 Bias in the continuous case

In a continuous environment with the additional approximation due to the tolerance zone $\varrho(x, y) < \epsilon$, there is no particular reason for the situation to improve, even though we can see that the bias in the weights generaly decreases as ϵ increases. The following example illustrates this point in the case of the mixture example of Sisson et al. (2007).

Example 2.2. (Example 1.1 continued) When considering the mixture setting of Example 1.1, Figure 3 shows the output of ten consecutive iterations of the ABC-PRC algorithm, using a decreasing sequence of ϵ_t's, from $\epsilon_1 = 2$ downto $\epsilon_{10} = 0.01$, and

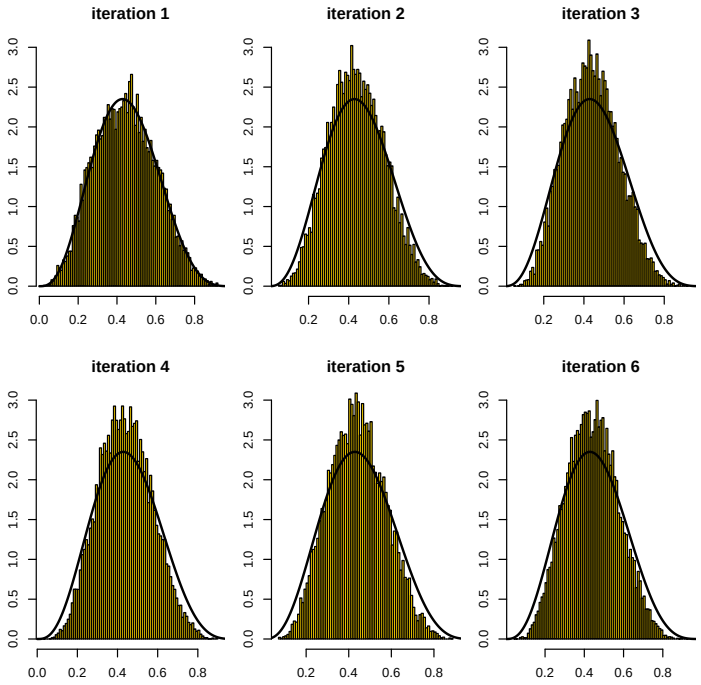

Figure 2: Histograms of samples of θ's produced by ABC-PRC in the Beta-binomial case, compared with the true posterior density $\mathcal{B}e(y+1, n-y+1)$ when $y = 3$ and $n = 7$, based on $M = 10^4$ simulations and $T = 6$ iterations, and started with a uniform proposal μ_1 equal to the prior distribution.

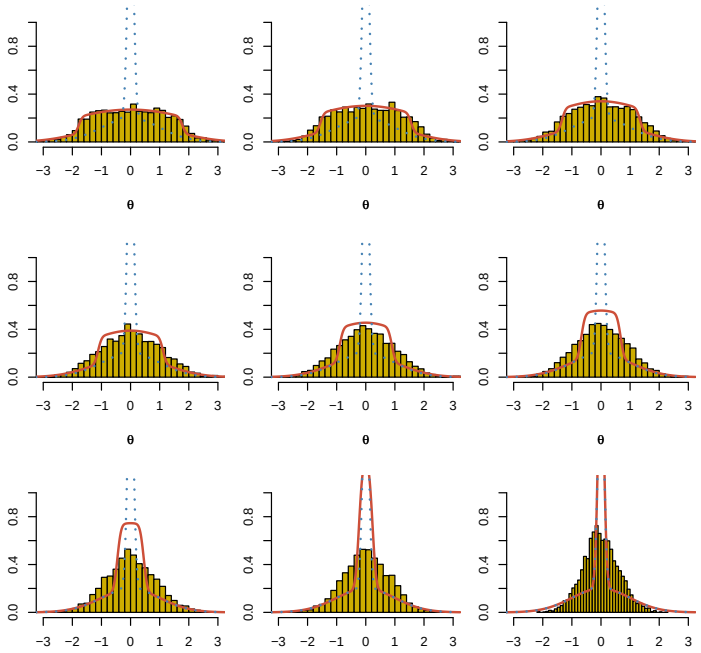

Figure 3: Histograms of the last nine weighted samples produced by ten consecutive iteration of ABC-PRC for the mixture target of Example 2.2, with a sequence of ϵ_t's, from $\epsilon_1 = 2$ downto $\epsilon_{10} = 0.01$ and a constant scale $\tau = .15$ in the random walk, based on $M = 5 \times 10^3$ simulations. (*Note:* The exact posterior density is plotted on top of the histogram with a dotted blue curve, while the target of the simulation algorithm, $\pi(\theta|\varrho(x, y) < \epsilon)$, is represented with brown full lines.)

a scale in the Gaussian random walk equal to $\tau = 0.15$. (Since τ is not explicitely specified in the original paper, we chose it equal to the scale used for the ABC-MCMC illustration. Note that an adaptive scale as the one adopted in the following section does exhibit worse biases.). As shown by this graph, using the ABC-PRC algorithm leads to a bias in terms of the target when ϵ is small, a somehow surprising contrast with the good fit produced on Figure 2 in Sisson et al. (2007), even though the values of the ϵ_t's we used are smaller. For the final values of ϵ_t, the output does not concentrate enough around the mode and misses the tails of the target, while larger values of ϵ_t in ABC-PRC produces a better fit (but is, obviously, farther from the true posterior.) ◀

3 Correction via importance sampling: ABC-PMC

3.1 Population Monte Carlo

Since the missing factor in the importance weight of Sisson et al. (2007) is related with the unknown likelihood $f(x|\theta)$, it would appear that a resolution of the problem would require an estimation of the likelihood based on earlier samples. This is however not the case in that a standard importance sampling perspective allows for a more direct approach, in a spirit similar to the generic Population Monte Carlo algorithm of Cappé et al. (2004).

Taking into account the way the t-th iteration sample of ABC-PRC is produced, it is indeed natural to modify the importance weight associated with an accepted simulation $\theta_i^{(t)}$ as

$$\omega_i^{(t)} \propto \pi(\theta_i^{(t)}) \Big/ \hat{\pi}_t(\theta_i^{(t)}) \,,$$

where

$$\hat{\pi}_t(\theta^{(t)}) = \sum_{j=1}^{N} \omega_j^{(t-1)} K_t(\theta^{(t)}|\theta_j^{(t-1)}) \,.$$

is the distribution used to generate the $\theta_i^{(t)}$'s. It is then straightforward to check that, whatever the distribution $\tilde{\pi}(\theta^{(t-1)})$ of the $\theta_j^{(t-1)}$'s is, the above weight corrects for the choice of the importance distribution since

$$\mathbb{E}[\omega^{(t)} h(\theta^{(t)})] \propto \int h(\theta^{(t)}) \, \frac{\pi(\theta^{(t)})}{\hat{\pi}(\theta^{(t)})} \, \hat{\pi}(\theta^{(t)}) f(\theta^{(t)}|y) \tilde{\pi}(\theta^{(t-1)}) \, \mathrm{d}\theta^{(t)} \, \mathrm{d}\theta^{(t-1)} \,.$$

This is essentially the proof for the unbiasedness of the population Monte Carlo method of Cappé et al. (2004) and the fact that the kernel K_t depends on the earlier simulations $(\theta^{(t-1)})_t$ [for instance by adjusting the variance of the random walk on those simulations] does not jeopardise the validity of the method. In addition, it must be noted that Douc et al. (2007) have proved that the kernel K_t *must be* modified at each iteration for the iterations to make sense, i.e. for those iterations to bring an asymptotic improvement on the Kullback-Leibler divergence between the proposal $\hat{\pi}(\theta^{(t)})$ and the (fixed) target $\pi(\theta|y)$: if for instance the variance of the random walk does not change from one iteration to the next, the approximation of the target $\pi(\theta|y)$ by $\hat{\pi}(\theta^{(t)})$ does not change either and it is more profitable (from a variance point of view) to increase the number of points at the second iteration.

When considering the special case of componentwise independent random walk proposals, i.e. when

$$K_t(\theta_k^{(t)}|\theta_k^{(t-1)}) = \tau_k^{-1} \varphi\{\tau_k^{-1}(\theta_k^{(t)} - \theta_k^{(t-1)})\}$$

for each component k of the parameter vector $\theta_k^{(t)}$, the (asymptotically) optimal choice of the scale factor τ_k can be found for each iteration. Indeed, when using a Kullback-

Leibler measure of divergence between the target and the proposal,

$$\mathbb{E}\left[\log\left\{\pi(\theta^{(t)}|y)\Big/\prod_k \tau_k^{-1}\varphi\{\tau_k^{-1}(\theta_k^{(t)}-\theta_k^{(t-1)})\}\, f(y|\theta^{(t)})\right\}\right]$$

where the expectation $\mathbb{E}$ is taken under the product distribution

$$(\theta^{(t)},\theta^{(t-1)}) \sim \pi(\theta^{(t)}|y)\times\pi(\theta^{(t-1)}|y)\,,$$

the minimisation of the Kullback divergence leads to maximise component-wise

$$\mathbb{E}[\log \tau_k^{-1}\varphi\{\tau_k^{-1}(\theta_k^{(t)}-\theta_k^{(t-1)})\}]$$

under the product distribution $\pi(\theta^{(t)}|y)\times\pi(\theta^{(t-1)}|y)$. As already mentioned above, the optimal scale is then to choose τ_k^2 equal to $\mathbb{E}[(\theta_k^{(t)}-\theta_k^{(t-1)})^2]$, that is,

$$\tau_k^2 = 2\mathrm{var}(\theta_k|y)\,,$$

under the posterior distribution. The implementation of this updating scheme on the scale is obviously straightforward.

The corresponding ABC-PMC scheme is then as follows:

ABC-PMC algorithm
Given a decreasing sequence of approximation levels $\epsilon_1,\ldots,\epsilon_T$,

1. At iteration $t=1$,

   For $i=1,...,N$
   Simulate $\theta_i^{(1)}\sim\pi(\theta)$ and $x\sim f(x|\theta_i^{(1)})$ until $\varrho(x,y)<\epsilon_1$
   Set $\omega_i^{(1)}=1/N$

   Take σ_2^2 as twice the empirical variance of the $\theta_i^{(1)}$'s

2. At iteration $2\le t\le T$,

   For $i=1,...,N$, repeat

   Pick $\theta_i^\star$ from the $\theta_j^{(t-1)}$'s with probabilities $\omega_j^{(t-1)}$
   generate $\theta_i^{(t)}|\theta_i^\star\sim\mathcal{N}(\theta_i^\star,\sigma_t^2)$ and $x\sim f(x|\theta_i^{(t)})$

   until $\varrho(x,y)<\epsilon_t$
   Set $\omega_i^{(t)}\propto\pi(\theta_i^{(t)})/\sum_{j=1}^N\omega_j^{(t-1)}\varphi\left(\sigma_t^{-1}\left\{\theta_i^{(t)}-\theta_j^{(t-1)})\right\}\right)$

   Take σ_{t+1}^2 as twice the weighted empirical variance of the $\theta_i^{(t)}$'s

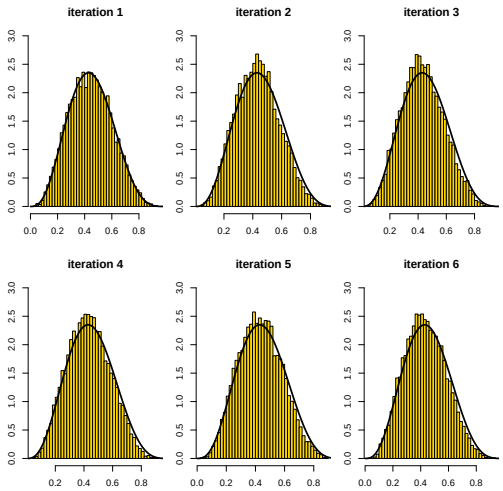

Figure 4: Histograms of the weighted samples produced by six consecutive iteration of ABC-PMC for the Beta-binomial target of Example 2.1, with a adaptive scale τ_t in the random walk, based on $M = 10^4$ simulations.

3.2 Illustrations

Example 3.1. (Example 2.1 continued) In the case of the beta-binomial model, we modify the sampling weights from iteration 2 onwards according to our scheme, namely

$$\begin{aligned}\omega_i^{(t)} &\propto \pi(\theta_i^{(t)})/\hat{\pi}(\theta_i^{(t)}) \\ &\propto \theta_i^{(t)}(1-\theta_i^{(t)}) \Big/ \sum_{j=1}^{N} \omega_j^{(t-1)} \varphi\left(\sigma_t^{-1}\left\{\text{logit}(\theta_i^{(t)})\right.\right. \\ &\qquad \left.\left. -\text{logit}(\theta_j^{(t-1)})\right\}\right) ,\end{aligned}$$

where σ_t^2 is twice the weighted variance of the $\text{logit}(\theta_j^{(t-1)})$'s. Figure 4 shows the outcome of the ABC-PMC scheme, with a much better fit of the [true] posterior distribution compared with ABC-PRC and a constant behaviour along iterations. We also note that the range of the importance weights remains quite limited (with a ratio from 1 to 4) and that the variance τ_t stabilises around 1 within a few iterations. ◀

Example 3.2. (Example 2.2 continued) For the mixture model, using the ABC-PMC algorithm with the corrected weights leads to a recovery of the target, whether using a fixed $\tau = 0.15$ or a sequence of adaptive τ_t's based on the variance of the previous sample following the ABC-PMC algorithm, as shown on Figures 5 and 6, respectively. The difference between both is actually difficult to spot, the estimated variance being again more stable. This means that $T = 10$ iterations are not necessary in that setting and that a faster decrease to $\epsilon = 0.01$ would also give a good fit. ◀

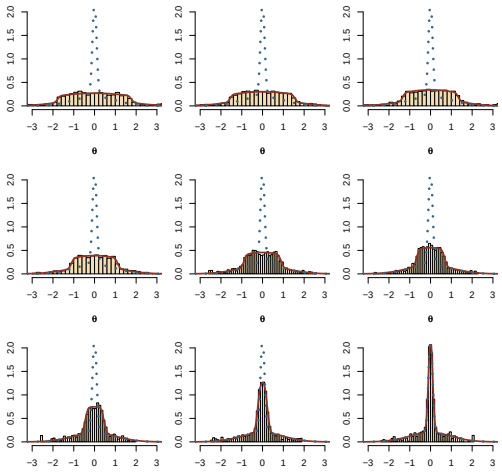

Figure 5: Histograms of the nine last weighted samples produced by ten consecutive iteration of ABC-PMC for the mixture target of Example 2.2, with a sequence of ϵ_t's, from $\epsilon_1 = 2$ downto $\epsilon_{10} = 0.01$ and a constant scale $\tau = .15$ in the random walk, based on $M = 5 \times 10^3$ simulations. (*Note:* The exact posterior density is plotted on top of the histogram with a dotted blue curve, while the target of the simulation algorithm, $\pi(\theta|\varrho(x, y) < \epsilon)$, is represented with brown full lines.)

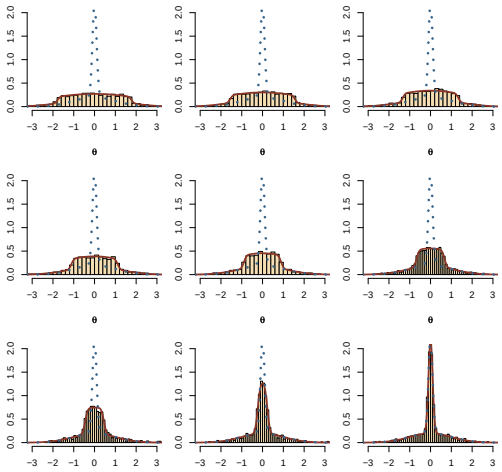

Figure 6: Histograms of the nine last weighted samples produced by ten consecutive iteration of ABC-PMC for the mixture target of Example 2.2, with a sequence of ϵ_t's, from $\epsilon_1 = 2$ downto $\epsilon_{10} = 0.01$ and an adaptive scale τ_t in the random walk, based on $M = 5 \times 10^3$ simulations. (*Note:* The exact posterior density is plotted on top of the histogram with a dotted blue curve, while the target of the simulation algorithm, $\pi(\theta|\varrho(x, y) < \epsilon)$, is represented with brown full lines.)

4 Conclusion

While the ABC-PRC algorithm relies on biased weights due to an inappropriate translation of the sequential scheme of Del Moral et al. (2006), with a visible impact on the quality of the approximation, we have shown in this paper that the same Markov transition kernels [and thus the same computing power] can be used to produce an unbiased scheme.

The new ABC-PMC scheme is based on an importance argument that does not require a backward kernel as in Sisson et al. (2007). We have thus established that the adaptive schemes of Douc et al. (2007) and Cappé et al. (2007) are also appropriate in this setting, towards a better fit of the proposal kernel K_t to the target $\pi(\theta|\varrho(x,y) < \epsilon)$. An important remark associated with this work is that the number of iterations T can be controlled via the modifications in the parameters of K_t, a stopping rule being that the iterations should stop when those parameters have settled, while the more fundamental issue of selecting a sequence of ϵ_t's towards a proper approximation of the true posterior can rely on the stabilisation of the estimators of some quantities of interest associated with this posterior.

Acknowledgements

The authors' research is partly supported by the Agence Nationale de la Recherche (ANR, 212, rue de Bercy 75012 Paris) through the 2005 project ANR-05-BLAN-0196-01 Misgepop. Parts of this paper were written during CPR's visit to the Isaac Newton Institute in Cambridge whose peaceful and stimulating environment was deeply appreciated.

5 References

Beaumont, M., W. Zhang, and D. Balding. 2002. Approximate Bayesian Computation in Population Genetics. *Genetics* 162: 2025–2035.

Cappé, O., R. Douc, A. Guillin, J.-M. Marin, and C. Robert. 2007. Adaptive importance sampling in general mixture classes. *Statist. Comput.* (To appear, arXiv:0710.4242).

Cappé, O., A. Guillin, J.-M. Marin, and C. Robert. 2004. Population Monte Carlo. *J. Comput. Graph. Statist.* 13(4): 907–929.

Del Moral, P., A. Doucet, and A. Jasra. 2006. Sequential Monte Carlo samplers. *J. Royal Statist. Society Series B* 68(3): 411–436.

Douc, R., A. Guillin, J.-M. Marin, and C. Robert. 2007. Convergence of adaptive mixtures of importance sampling schemes. *Ann. Statist.* 35(1). ArXiv:0708.0711.

Liu, J. 2001. *Monte Carlo Strategies in Scientific Computing*. Springer-Verlag, New York.

Marjoram, P., J. Molitor, V. Plagnol, and S. Tavaré. 2003. Markov chain Monte Carlo without likelihoods. *Proc Natl Acad Sci U S A* 100(26): 15324–15328.

Pritchard, J. K., M. T. Seielstad, A. Perez-Lezaun, and M. W. Feldman. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol. Biol. Evol.* 16: 1791–1798.

Robert, C. 2001. *The Bayesian Choice*. 2nd ed. Springer-Verlag, New York.

Robert, C. and G. Casella. 2004. *Monte Carlo Statistical Methods*. 2nd ed. Springer-Verlag, New York.

Sisson, S. A., Y. Fan, and M. Tanaka. 2007. Sequential Monte Carlo without likelihoods. *Proc. Natl. Acad. Sci. USA* 104: 1760–1765.