Sequential Monte Carlo with Adaptive Weights for Approximate Bayesian Computation

Fernando V. Bonassi
Mike West
Department of Statistical Science, Duke University, Durham, NC

Abstract. Methods of Approximate Bayesian computation (ABC) are increasingly used for analysis of complex models. A major challenge for ABC is overcoming the often inherent problem of high rejection rates in the accept/reject methods based on prior:predictive sampling. A number of recent developments aim to address this with extensions based on sequential Monte Carlo (SMC) strategies. We build on this here, introducing an ABC SMC method that uses data-based adaptive weights. This easily implemented and computationally trivial extension of ABC SMC can very substantially improve acceptance rates, as is demonstrated in a series of examples with simulated and real data sets, including a currently topical example from dynamic modelling in systems biology applications.

Keywords: Adaptive simulation, approximate Bayesian computation (ABC), dynamic bionetwork models, importance sampling, mixture model emulators, sequential Monte Carlo (SMC)

1 Introduction

Methods of approximate Bayesian computation (ABC) are becoming increasingly exploited, especially for problems in which the likelihood function is analytically intractable or very expensive to compute (Pritchard et al. 1999; Marjoram et al. 2003). Recent applications run across areas such as evolutionary genetics (Beaumont et al. 2002), epidemiology (McKinley et al. 2009), astronomical model analysis (Cameron and Pettitt 2012), among others (Csilléry et al. 2010).

Vanilla ABC simulates parameter/data pairs \((\theta, x)\) from the prior distribution whose density is \(p(\theta)p(x|\theta)\), accepting \(\theta\) as an approximate posterior draw if its companion data \(x\) is “close enough” to the observed data \(x_{obs}\). If \(p(x, x_{obs})\) is the chosen measure of discrepancy, and \(\epsilon\) is a discrepancy threshold defining “close”, then accepted parameters are a sample from \(p(\theta|p(x, x_{obs}) < \epsilon)\). Often, it is not possible or efficient to use the complete data set. In that case, a reduced dimensional set of summary statistics \(S\) can be used so that the accepted draws sample from \(p(\theta|d(S, S_{obs}) < \epsilon)\). For simplicity of notation, the posterior will be denoted by \(p(\theta|d(x, x_{obs}) < \epsilon)\), with the understanding that the effective data \(x\) potentially represents such a summary of the original data.

A main issue is high rejection rates that result from: \(i\) the requirement that \(\epsilon\) be small to build faith that the approximate posterior is a good approximation to \(p(\theta|x_{obs})\), and \(ii\) the posterior \(p(\theta|x_{obs})\) may be concentrated in completely different regions of parameter space than the prior. To address this, modifications of vanilla ABC are

© 2004 International Society for Bayesian Analysis ba0001
Adaptive Weight ABC SMC

emerging, including regression adjustment strategies (e.g., Beaumont et al. 2002; Blum and François 2010; Bonassi et al. 2011), and automatic sampling schemes (Marjoram et al. 2003; Sisson et al. 2007) utilizing techniques such as sequential Monte Carlo (SMC). In the latter ABC methods, SMC is used in order to automatically, sequentially refine posterior approximations to be used to generate proposals for further steps. At each of a series of sequential steps indexed by $t$, these methods aim to generate draws from $p(\theta | \rho(x, x_{obs}) < \epsilon_t)$ where $\epsilon_t$ define a series of decreasing thresholds. Total acceptance rates can be significantly higher than with vanilla ABC scheme (Toni et al. 2009; Toni and Stumpf 2010).

The original version of the ABC SMC algorithm proposed in Sisson et al. (2007) was motivated by the SMC samplers methodology of Del Moral et al. (2006). Later, Beaumont et al. (2009) realized that this original method can result in biased samples relative to the true posterior, and this was followed by development of a corrected approach (Beaumont et al. 2009; Sisson et al. 2009; Toni et al. 2009). The general form of the algorithm, which relies fundamentally on sequential importance sampling, is shown in Figure 1.

Applications of this ABC SMC algorithm have been presented in a variety of areas including population genetics (Beaumont et al. 2009), systems biology (Toni et al. 2009; Toni and Stumpf 2010; Liepe et al. 2010) and psychology (Turner and Van Zandt 2012). In terms of methodology development, there is increasing interest in extensions and improvements of this algorithm, as demonstrated in, for example, Lenormand et al. (2012); Filippi et al. (2012); Silk et al. (2012). Building on this momentum and open challenges to improving the methodology, our current focus is on the form of the ABC SMC of Figure 1. We note and comment on the different ABC SMC approach of Del Moral et al. (2011); this extends SMC samplers (Del Moral et al. 2006) and uses an MCMC kernel for propagation of particles. As a result, the computation of weights has linear complexity as a function of the number of particles, while a disadvantage is the implied increased variance of the resulting importance weights in comparison to the basic algorithm of Figure 1 that we begin with here.

One aspect of the algorithm of Figure 1 is that the computational demand in evaluating weights increases quadratically as a function of the number of particles. While this appears disadvantageous, in practice is is often just not a limiting factor. The reason for this is that in practical ABC applications, computation time is typically substantially dominated by the repeated data simulation steps; this is borne out in our own experiences discussed further below, and has been clearly indicated and discussed in other studies, including Beaumont et al. (2009); Filippi et al. (2012); Del Moral et al. (2011), for example.

2 ABC SMC with Adaptive Weights

The ABC SMC strategy above is intimately related to adaptive importance sampling using adaptively refined posterior approximations based on kernel mixtures as importance samplers. Originating as a method for direct posterior approximation (West
 Initialize threshold schedule $\epsilon_1 > \cdots > \epsilon_T$

2. Set $t = 1$
   
   For $i = 1, \ldots, N$
   
   - Simulate $\theta_i^{(1)} \sim p(\theta)$ and $x \sim p(x|\theta_i^{(1)})$ until $\rho(x, x_{obs}) < \epsilon_1$
   
   - Set $w_i = 1/N$

3. For $t = 2, \ldots, T$
   
   For $i = 1, \ldots, N$
   
   - Pick $\theta_i^*$ from the $\theta_j^{(t-1)}$’s with probabilities $w_j^{(t-1)}$
   
   - Draw $\theta_i^{(t)} \sim K_t(\theta_i^{(t)} | \theta_i^*)$ and $x \sim p(x|\theta_i^{(t)})$ until $\rho(x, x_{obs}) < \epsilon_t$
   
   - Compute new weights as
     
     $$w_i^{(t)} \propto \frac{p(\theta_i^{(t)})}{\sum_j w_j^{(t-1)} K_t(\theta_i^{(t)} | \theta_j^{(t-1)})}$$

   Normalize $w_i^{(t)}$ over $i = 1, \ldots, N$

Figure 1: ABC SMC algorithm (Beaumont et al. 2009; Sisson et al. 2009; Toni et al. 2009). Here $K_t(\cdot|\cdot)$ is conditional density that serves as a transition kernel to “move” sampled parameters and then appropriately weight accepted values. In contexts of real-valued parameters, for example, $K_t(\theta|\theta^*)$ might be taken as a multivariate normal or $T$ density centred at or near $\theta^*$, and whose shape and scales may decrease as $t$ increases.

Figure 2 shows the algorithmic description of this new ABC SMC with Adaptive
1. Initialize threshold schedule $\epsilon_1 > \cdots > \epsilon_T$

2. Set $t = 1$
   For $i = 1, \ldots, N$
   - Simulate $\theta_i^{(1)} \sim p(\theta)$ and $x \sim p(x|\theta_i^{(1)})$ until $\rho(x, x_{obs}) < \epsilon_1$
   - Set $w_i = 1/N$

3. $t = 2, \ldots, T$
   Compute data based weights $v_i^{(t-1)} \propto w_i^{(t-1)}K_{x,t}(x_{obs}|x_i^{(t-1)})$
   Normalize weights $v_i^{(t-1)}$
   For $i = 1, \ldots, N$
   - Pick $\theta^*_i$ from the $\theta_i^{(t-1)}$’s with probabilities $v_j^{(t-1)}$
   - Draw $\theta_i^{(t)} \sim K_{\theta,t}(\theta_i^{(t)}|\theta^*_t)$ and $x \sim p(x|\theta_i^{(t)})$ until $\rho(x, x_{obs}) < \epsilon_t$
   - Compute new weights as

   $$w_i^{(t)} \propto \frac{p(\theta_i^{(t)})}{\sum_j v_j^{(t-1)}K_{\theta,t}(\theta_i^{(t)}|\theta^*_t)}$$

   Normalize $w_i^{(t)}$ over $i = 1, \ldots, N$

Figure 2: ABC SMC with Adaptive Weights

Weights (ABC SMC AW). The inclusion of a new step where the weights are modified according to the respective values of $x$ adds computations. However, since computational time in ABC SMC algorithm is usually dominated by the extensive repetition of model simulations, the increased compute burden will often be negligible. Note also that original ABC SMC is a particular case when $K_{x,t}(\cdot|x)$ is uniform over the region of accepted values of $x$.

The idea of approximating the joint distribution of parameters and data together is also present in Bonassi et al. (2011), where mixture modeling is applied to the joint distribution and then used as a form of nonlinear regression adjustment. The smoothing step in ABC SMC AW can be seen as an automatic simplified version of that approach, about which we say more in the concluding Section 6.

3 Theoretical Aspects

We discuss some of the structure of ABC SMC AW to provide insight as to why it can be expected to lead to improved acceptance rates.

First, note that the final output of ABC SMC methods is a sample from the target distribution $p(\theta|d(x, x_{obs}) < \epsilon_T)$. This provides the practitioner with the convenient
option of using the output as a final approximation or as an input to be refined by using a preferred regression adjustment technique, such as local linear regression (Beaumont et al. 2002) or mixture modeling (Bonassi et al. 2011). However, in the context of intermediate ABC SMC steps, more refined approximations have the potential to improve efficiency; this can be automatically achieved by use of kernel smoothing techniques. At each intermediate step \( t \), this motivates the following approximation for the joint density \( p(x, \theta) \), locally around \( x = x_{\text{obs}} \):

\[
\hat{p}_t(x, \theta | A) = \int \int p(\theta)p(\tilde{x}|\tilde{\theta}, A)K_\theta(\theta|\tilde{\theta})K_x(x|\tilde{x})d\tilde{x}d\tilde{\theta}
\]  

(1)

where \( A = \{ x : d(x, x_{\text{obs}}) < \epsilon_t \} \) represents the acceptance event at step \( t \), and \( K_\theta \) and \( K_x \) are the kernel functions as described in Section 2. This defines an implicit posterior emulator, namely

\[
\hat{p}_t(\theta | x_{\text{obs}}) \propto \int \int p(\theta)p(\tilde{x}|\tilde{\theta}, A)K_\theta(\theta|\tilde{\theta})K_x(x_{\text{obs}}|\tilde{x})d\tilde{x}d\tilde{\theta}.
\]  

(2)

We now explore equation (2) to ensure that: (A) it defines a valid importance sampler with target \( p(\theta)p(x, x_{\text{obs}}) < \epsilon_t \) at each step \( t \), and (B) the resulting expected acceptance rates are higher than for regular ABC SMC. We address these two points in turn.

**A:** At step \( t \), note that \( \theta \) is sampled from the distribution having density \( g(\theta) = \sum_j v_j^{(t-1)}K_{t,\theta}(\theta|\theta_j^{(t-1)}) \). Given this value, the extra step then simulates the acceptance event \( A \) from \( pr(A | \theta) \), where \( pr(A | \theta) = \int_A p(x, \theta)dx \), resulting in the overall proposal

\[
g(\theta|A) \propto g(\theta)pr(A | \theta).
\]  

(3)

Finally, to sample from \( p(\theta|A) = p(\theta|d(x, x_{\text{obs}}) < \epsilon_t) \) based on this proposal, the importance sampling weights are

\[
w(\theta) \propto \frac{p(\theta)pr(A | \theta)}{g(\theta)pr(A | \theta)} = \frac{p(\theta)}{g(\theta)},
\]  

(4)

which have exactly the same form as in Figure 2.

**B:** At step \( t \), equation (1) describes the approximation for the joint density of \( (x, \theta) \) locally around \( x = x_{\text{obs}} \). Based on this representation, it is possible to show that the proposal distribution implicitly defined in ABC SMC AW results in higher prior predictive density over the acceptance region for the next SMC step. This is seen as follows. For simplicity of notation, use \( p_0(x, \theta) \) to refer to \( \hat{p}_t(x, \theta) \) at the current step \( t \). The proposal density in ABC SMC is \( p_0(\theta) \), whereas that for ABC SMC AW is \( p_0(\theta|x_{\text{obs}}) \). These two densities induce marginal prior predictive densities \( p_0(x) \) and \( p_1(x) \), respectively. Integration of a prior predictive density over the acceptance region \( A_{t+1} = \{ x : \rho(x, x_{\text{obs}}) < \epsilon_{t+1} \} \) yields the corresponding acceptance probability for the next SMC step, namely \( pr(A_{t+1}) \). We now show that \( pr(A_{t+1}) > pr_0(A_{t+1}) \) so that
Adaptive Weight ABC SMC

ABC SMC AW improves acceptance rates over regular ABC SMC. Our proof relies on the assumption that acceptance probability \( pr_0(A_{t+1}|\Theta) \) is positively correlated with \( p_0(x_{\text{obs}}|\Theta) \) with respect to \( \theta \sim p_0(\theta) \). This is a reasonable assumption to make in regular cases, including the limiting case that \( \epsilon_{t+1} \to 0 \) when the correlation tends to 1.

Under ABC SMC we have

\[
pr_0(A_{t+1}) = \int pr_0(A_{t+1}|\theta)p_0(\theta)d\theta = E(pr_0(A_{t+1}|\Theta)),
\]

where the expectation is with respect to \( p_0(\cdot) \). The corresponding value under ABC SMC AW is

\[
pr_1(A_{t+1}) = \int pr_0(A_{t+1}|\theta)p_0(\theta|x_{\text{obs}})d\theta
\]

\[
= \int pr_0(A_{t+1}|\theta)p_0(x_{\text{obs}}|\theta)p_0(\theta)d\theta = \frac{E(pr_0(A_{t+1}|\Theta)p_0(x_{\text{obs}}|\Theta))}{p_0(x_{\text{obs}})}
\]

\[
> \frac{E(pr_0(A_{t+1}|\Theta))E(p_0(x_{\text{obs}}|\Theta))}{p_0(x_{\text{obs}})} = pr_0(A_{t+1}),
\]

assuming \( \text{Cov}(pr_0(A_{t+1}|\Theta)p_0(x_{\text{obs}}|\Theta)) > 0 \).

4 Illustrative Example: Normal Mixtures

4.1 A Standard Example

A simple example taken from previous studies of ABC SMC (Sisson et al. 2007; Beaumont et al. 2009) concerns scalar data \( x|\theta \sim 0.5N(\theta, 1) + 0.5N(\theta, 0.01) \) and prior \( \theta \sim \mathcal{U}(-10, 10) \). With observed value \( x_{\text{obs}} = 0 \), the target posterior is \( \theta|x_{\text{obs}} \sim 0.5N(0, 1) + 0.5N(0, 1/100) \) truncated to \((-10, 10)\).

We follow details in Sisson et al. (2007) with discrepancy measure \( \rho(x, x_{\text{obs}}) = |x - x_{\text{obs}}| \) and threshold schedule \( \epsilon_{1:3} = (2, 0.5, 0.025) \), and we use normal kernels \( K_{\theta,t} \) and \( K_{x,t} \) with standard deviations (or bandwidth parameters) \( h_\theta \) and \( h_x \), respectively. Following standard recommendations in West (1993b) and Scott and Sain (2005), the bandwidths \( h_k \), for \( k \in \{x, \theta\} \), are set at \( h_k = \hat{\sigma}_k/N^{1/6} \) where \( \hat{\sigma}_k \) is the standard deviation, which is computed based on the values of the particles and their respective weights. This standard rule-of-thumb specification is an asymptotic approximation to the optimal bandwidth choice based on the mean integrated squared error of the product kernel density estimate (Scott 1992).

We ran ABC SMC and ABC SMC AW to obtain samples of \( N = 5,000 \) particles in each case. Table 1 shows the average number of simulation steps per accepted particle; the adaptive weight modification requires about 30\% fewer simulations. As show in Figure 3, the resulting posterior approximations are quite accurate. More importantly from the viewpoint of the methodology here, the two strategies give very closely similar results, with the AW variant substantially improving the computational efficiency in
terms of acceptance rates. The suggested equivalence of accuracy of the methods is also confirmed in Figure 4, which displays a comparison of Monte Carlo estimates based on 50 repeated runs of both methods using \( N = 1,000 \) particles. The plots in the figure do not indicate any significant difference in the distribution of Monte Carlo estimates produced by both methods.

\[
\begin{array}{cccc}
 t & \epsilon_t & \text{ABC SMC} & \text{ABC SMC AW} \\
 1 & 2 & 5.01 & 4.96 \\
 2 & 0.5 & 4.33 & 2.38 \\
 3 & 0.025 & 39.71 & 27.22 \\
 \hline
\text{Total} & & 49.05 & 34.56
\end{array}
\]

Table 1: Normal mixture example: Average number of simulation steps per accepted particle.

Figure 3: Normal mixture example: Approximate (lines) and exact (shaded) posterior densities.

### 4.2 Multivariate Mixture Examples

To more aggressively explore performance, we have run studies on multivariate versions of the normal mixture example. Generally, take \( p \)-dimensional data \( x | \theta \sim 0.5N_p(\theta, I_p) + 0.5N_p(\theta, 0.01I_p) \), where \( x = (x_1, \ldots, x_p)' \) and \( \theta = (\theta_1, \ldots, \theta_p)' \). The prior has \( \theta \) uniformly distributed on \([ -10, 10 ]^p \).

For the method implementation we use normal kernels \( K_{\theta,t} \) and \( K_{x,t} \) with diagonal variance matrices and with scalar bandwidths defined via the standard rule-of-thumb (Silverman 1986; Scott and Sain 2005) as used in the example of previous sec-
Adaptive Weight ABC SMC

Figure 4: Monte Carlo estimates for the normal mixture example: Box-plots of estimated posterior moments (as annotated) based on 50 repeated runs of SMC ABC and SMC ABC AW.

That is, for each scalar dimension $k$ of parameters and data, if $\hat{\sigma}_k$ denotes the standard deviation in that dimension, then $h_k = \hat{\sigma}_k N^{-1/(d+4)}$ where $N$ is the number of particles and $d$ is the total dimension (parameters and data). The discrepancy used is $\rho(x, x_{\text{obs}}) = \sum_k (x_k - x_{k,\text{obs}})^2$.

Given the observation $x_{\text{obs}} = (0, \ldots, 0)'$, we ran ABC SMC and ABC SMC AW to obtain samples of $N = 5,000$ particles for different cases of dimension $p$. In each case, the threshold schedule was defined to keep correspondence with the schedule used in the example of Section 4.1. This correspondence was achieved by selecting threshold values that result in the same percentiles of the discrepancy distribution of prior:model simulations. The comparison results for repeat simulations for all cases with $p \in \{1, 2, \ldots, 50\}$ show substantial reduction in the number of required model simulations induced by the adaptive weight modification. Across this range of dimensions, the reductions seen run from about 30 – 70%, confirming that the adaptive weighting strategy can be effective in higher-dimensional problems, even with $p$ as high as 50 or more.

5 Comparison in Applications

Two studies demonstrate the improvements achievable using adaptive weights in a couple of interesting applied contexts, one using real data and the other synthetic data (so that “truth” is known). In each we use the same algorithm setup as described in Section 4.2. For the threshold schedule, we follow Beaumont et al. (2002) with a pilot study of prior:model simulations to identify $\epsilon_T$ as a very low percentile of the distribution of the discrepancies, and then define a schedule $\epsilon_{1:T}$ to gradually reduce to that level. See also Bonassi et al. (2011) for more insight into this approach and variants. Specific values of the $\epsilon_t$ are noted in the following sections.
5.1 Toggle Switch Model in Dynamic Bionetworks

A real data example from systems biology comes from studies of dynamic cellular networks based on measures of expression of genes (network nodes) at one or more “snapshots” in time. The model, context and flow cytometry data set come from Bonassi et al. (2011). This is data from experiments on bacterial cells using an engineered gene circuit with design related to the toggle switch model (Gardner et al. 2000). The model describes the dynamical behavior of a network with two genes \( u \) and \( v \) with doubly repressive interactions. In discrete time the specific model form is

\[
\begin{align*}
 u_{c,t+h} &= u_{c,t} + h\alpha_u/(1 + v_{c,t+1}^{\beta_u}) - h(1 + 0.03u_{c,t}) + h0.5\xi_{c,u,t}, \\
 v_{c,t+h} &= v_{c,t} + h\alpha_v/(1 + u_{c,t+1}^{\beta_v}) - h(1 + 0.03v_{c,t}) + h0.5\xi_{c,v,t},
\end{align*}
\]

over time \( t \), where \( c \) indexes bacterial cells, \( h \) is a small time step, the \( \xi_{c,.t} \) are independent standard normals, and for some specified initial values \( u_{c,0}, v_{c,0} \).

The observable data set is a sample from the marginal distribution of levels of just one of the two genes at one point in time, viz. \( y = \{y_c, c = 1:2,000\} \), where \( y_c \) is a noisy measurements of \( u_\tau \) at a given time point \( \tau \). The measurement error process is, cell-by-cell, given by

\[
y_c = u_{c,\tau} + \mu + \mu\eta_c/u_{c,\tau}^{\gamma}, \quad c = 1, \ldots, 2,000,
\]

where the \( \eta_c \) are standard normals, independent over cells and independent of the stochastic terms \( \xi_{c,.t} \) in the state evolution model of equation (5). The study of Bonassi et al. (2011) uses \( h = 1, \tau = 300 \) and initial state \( u_{c,0} = v_{c,0} = 10 \), adopted here. The full set of 7 model parameters is \( \theta = (\alpha_u, \alpha_v; \beta_u, \beta_v; \mu, \sigma, \gamma) \).

The data is shown in Figure 5. Following Bonassi et al. (2011), the sample data set is reduced to a set of summary “reference signatures” that define the effective data \( x \) we aim to condition on. The details of the dimension reduction from the full, original sample \( y \) to \( x \) are not of primary interest here, though are of course critically important for the application area; readers interested in the specifics of the applied context and the precise definition of \( x \) can consult Bonassi et al. (2011). Here we simply start with the reduced, 11-dimensional data summary \( x \), giving supplementary code and data that provide the relevant details and precise definition of \( x \). We do note that, in addition to dimension reduction, the specific definition of \( x \) in Bonassi et al. (2011) has the advantage of producing an orthogonal projection of transformed raw data so that the sample elements of \( x \) are uncorrelated, which makes our use of diagonal variance matrices in the kernels \( K_{x,t} \) particularly apt.

The prior for \( \theta \) is comprised of independent uniforms on real-valued transformed parameters, defined by finite ranges for each based on substantive biochemical background information. The priors are also taken from the prior study in Bonassi et al. (2011) and are indicated by the ranges of the plots in Figure 5. We summarize analyses with \( N = 2,000 \) particles and \( \epsilon_{1.5} = (2500, 750, 250, 150, 75) \) where the final tolerance level \( \epsilon_T = 75 \) corresponds to an approximate \( 10^{-4} \) quantile of simulated prior discrepancies. Table 2 shows the average number of prior:datagen steps per acceptance for both ABC SMC and ABC SMC AW.
Figure 5: Toggle switch study: Data are shown in the lower right frame (sample size $C = 2,000$ cells). The other frames represent approximate posterior marginal densities for each of the 7 toggle switch model parameters, as annotated, to compare results from ABC SMC (red) and ABC SMC AW (blue).
Evidently, SMC ABC AW leads to a reduction of roughly 30% simulation steps, a practically very material gain in efficiency. Combined with this, Figure 5 shows that the approximate posteriors from ABC SMC AW are practically the same as those from ABC SMC, the differences being easily attributable to Monte Carlo variation.

The toggle switch example is also used to illustrate some aspects of computation time of ABC SMC in a real application. We implemented the methods discussed by using a PC Intel 3.33 GHz. The total processing time for the ABC SMC algorithm was 70,543 seconds, while for the version with adaptive weights the time was 49,332. From that comparison we can observe the same gain in efficiency as observed in the number of required simulation steps, confirming the point that the computation of the adaptive weights adds negligible processing time to the ABC SMC algorithm.

We also implemented the method of Del Moral et al. (2011) to illustrate of the performance difference that is induced by the computation of the weights based on linear complexity in the number of particles. This implementation was based on the same settings discussed before, i.e., with the same number of particles, threshold schedule and normal kernel. The remaining element to be defined in that method is the number \( M \) of synthetic data sets generated for each particle, and here we use \( M = 31 \) so that the method relies on roughly the same overall number of simulations steps that were used in the ABC SMC implementation discussed before (without adaptive weights). Given this setup, the implementation of the method of Del Moral et al. (2011) was only 0.8% faster than the previous ABC SMC implementation. This result confirms the point discussed in Section 1, that the linear complexity in the computation of the weights does not necessarily result in significant gain in computation time for practical applications of ABC SMC. Lenormand et al. (2012) presents a more complete comparison of these methods under different scenarios, which also suggest that the algorithm with linear complexity in the computation of the weights does not necessarily result in better overall computational efficiency. That study also takes into account the accuracy of the methods in the overall comparison.

<table>
<thead>
<tr>
<th>( t )</th>
<th>( \epsilon_t )</th>
<th>ABC SMC</th>
<th>ABC SMC AW</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2500</td>
<td>12.3</td>
<td>12.2</td>
</tr>
<tr>
<td>2</td>
<td>750</td>
<td>16.1</td>
<td>10.6</td>
</tr>
<tr>
<td>3</td>
<td>250</td>
<td>24.3</td>
<td>20.6</td>
</tr>
<tr>
<td>4</td>
<td>150</td>
<td>42.1</td>
<td>28.2</td>
</tr>
<tr>
<td>5</td>
<td>75</td>
<td>96.4</td>
<td>61.8</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>191.3</td>
<td>133.4</td>
</tr>
</tbody>
</table>

Table 2: Toggle switch study: Average number of simulations per accepted particle.

5.2 Queuing System

The second study is of a queuing system previously discussed in Heggland and Frigessi (2004) and addressed using ABC by Blum and François (2010) and Fearnhead and
Prangle (2012). The context is a single server, first-come-first-serve queue (M/G/1). Three parameters $\theta = (\theta_1, \theta_2, \theta_3)'$ determine the distributions of service and inter-arrival times; service times are uniform on $[\theta_1, \theta_2]$ while inter-arrival times are exponential with rate $\theta_3$. In the notation of Heggland and Frigessi (2004), $W_r$ is the inter-arrival time of the $r$th customer and $U_r$ the corresponding service time. The inter-departure time process $\{Y_r, r = 1, 2, \ldots \}$ is then

$$ Y_r = \begin{cases} U_r, & \text{if } \sum_{i=1}^r W_i \leq \sum_{i=1}^{r-1} Y_i, \\ U_r + \sum_{i=1}^r W_i - \sum_{i=1}^{r-1} Y_i, & \text{if } \sum_{i=1}^r W_i > \sum_{i=1}^{r-1} Y_i. \end{cases} \quad (7) $$

Inter-arrival times are unobserved; the observed data from $R$ customers is $x = \{Y_1, \ldots, Y_R\}$ where $R$ is the number of customers.

We generated a synthetic data set following the specification in Blum and François (2010); here $R = 50$ with “true” parameters $\theta = (1, 5, 0.2)'$. Summary statistics are taken as 3 equidistant quantiles together with the minimum and maximum values of the inter-departure times. The prior has $(\theta_1, \theta_2 - \theta_1, \theta_3)$ uniformly distributed on $[0, 10]^3$.

Analysis used $N = 1,000$ particles and discrepancy schedule $\epsilon_{1:5} = (200, 100, 10, 2, 1)$. The final level $\epsilon_T = 1$ corresponds to a value close to the $10^{-4}$ quantile of simulated prior discrepancies. Figure 6 shows close agreement between the estimated marginal posterior densities under ABC SMC and ABC SMC AW, and, incidentally, that they support regions containing the true parameters underlying this synthetic data set. Figure 7 displays the comparison of Monte Carlo estimates based on 50 repeat runs of the two methods using $N = 1,000$ particles. The plots in the figure do not suggest any significant difference in the distribution of Monte Carlo estimates that result.

In order to study the gain in efficiency derived from the use of adaptive weights, we repeated this analysis 100 times, each repeat involving a new prior:data simulation. Table 3 shows summaries of the numbers of data generation steps necessary for the ABC SMC and ABC SMC AW algorithms. We see that we realized an average reduction of about 58% in the number of simulation steps by using adaptive weighting.

<table>
<thead>
<tr>
<th>$t$</th>
<th>$\epsilon_t$</th>
<th>min</th>
<th>mean</th>
<th>max</th>
<th>ABC SMC AW</th>
<th>ABC SMC AW</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>200</td>
<td>1.0</td>
<td>1.3</td>
<td>3.1</td>
<td>1.0</td>
<td>1.3</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>1.2</td>
<td>1.4</td>
<td>2.1</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>3.3</td>
<td>12.7</td>
<td>760.5</td>
<td>1.2</td>
<td>2.4</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2.4</td>
<td>9.0</td>
<td>134.1</td>
<td>1.4</td>
<td>3.9</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1.9</td>
<td>6.9</td>
<td>107.5</td>
<td>1.1</td>
<td>4.5</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>31.3</td>
<td>13.1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Queuing system analysis: Summaries of 100 replicate synthetic data analyses, showing average numbers of simulation steps per accepted particle.
Figure 6: Queuing system analysis: Summaries of analyses of simulated data set where the 3 frames represent margins for each of the model parameters as annotated, showing results from ABC SMC (blue) and ABC SMC AW (red). The true known parameter values underlying the synthetic data are marked as red triangles on the horizontal axes.
Figure 7: Monte Carlo estimates for the queueing system example: Box-plots of estimated posterior moments (as annotated) based on 50 repeated runs of SMC ABC and SMC ABC AW.
6 Additional Discussion

This paper has introduced ABC SMC AW, shown that it is theoretically expected to improve the effectiveness of ABC SMC based on adaptive, data-based weights, and demonstrated some of the practical potential in two interesting model contexts from related literature. In other studies we have observed similar, practically very material improvements in terms of higher acceptance rates, and are utilizing the new approach in other applications as a routine. It is simple to implement, requiring only a minor extension of standard ABC SMC code. Further, the computational overheads adaptive weighting generates are—in anything but trivial models—typically quite negligible relative to the main expense of forward simulations of prior predictive distributions. We also note that the adaptive weighting idea has the potential to be integrated into other ABC SMC extensions, even though this integration requires careful study of potential computational and theoretical implications.

The basic idea of adaptive weights links closely to adaptive importance sampling and direct posterior approximations based on mixtures of kernel forms. As noted in Section 2, the local smoothing for nonlinear regression adjustment in ABC of Bonassi et al. (2011) uses multivariate normal mixtures in related ways as “local” posterior approximations in regions defined by the threshold setting. There the mixture modelling, used to define posterior emulators, is based on large-scale Bayesian nonparametric models that can have many mixture components and so flexibly adapt to the shapes of local posterior contours. Incidentally, mixture fitting is computationally effective based on GPU parallelized code (Suchard et al. 2010; Cron and West 2011). This connection suggests a more general adaptive weighting strategy that uses a joint kernel that is not of product form, with the potential to customize the weighting of sampled parameters further. That is, in the AW algorithm of Figure 2, the kernel \( K_{\theta,t}(\theta|\theta^*) \) would be modified to have shape characteristics that depend also on the locale in which the kernel location \( \theta^* \) sits. Building on the connections with multivariate normal mixtures suggests specific ways in which these modifications could be developed, and this is under investigation. At this point, however, it is unclear just how beneficial this will be in practice. Such extensions will require substantial additional computational overheads to identify and compute local kernel functions, which may more than offset the gains in efficiency the extensions can be expected to generate. The gains in acceptance rates already achieved by the very simple (to code and run) adaptive weighting method of this paper can already be very substantial, as our examples highlight.

References


About the Authors

Fernando Bonassi graduated with his PhD in Statistical Science from Duke University in 2013. His research interests include Bayesian computation, especially using ABC and SMC methods, Bayesian mixture modelling, foundations of Bayesian inference, decision analysis, applications in dynamic systems in areas including immunology and gene network studies, among other topics. Bonassi holds BS and MS degrees in Statistics from the Universidade de São Paulo where he studied at the Instituto de Matemática e Estatística (IME-USP, Brazil), as well as the MS in Statistical Science from Duke University.

Mike West is the Arts & Sciences Professor of Statistical Science in the Department of Statistical Science at Duke University. His research is in several areas of Bayesian statistical modelling and computation, focusing currently on methodology development and applications of complex stochastic modelling in high-dimensional problems, large-scale computation, financial econometrics and societally relevant “big data” issues in global financial systems, dynamic networks, advanced imaging and statistical modelling in systems biology applications, and other areas. West is an expert in time series analysis, forecasting, and computational statistics, and has published widely in interdisciplinary applications in signal processing, finance, econometrics, climatology, public health, genomics, immunology, neurophysiology, systems biology and other areas.

Acknowledgments

This work was supported in part by grants from the U.S. National Science Foundation (DMS-1106516) and National Institutes of Health (P50-GM081883 and RC1-AI086032). Any opinions, findings and conclusions or recommendations expressed in this work are those of the authors and do not necessarily reflect the views of the NSF or NIH.