Marginal and joint space representations within ABC, and the issue of bias

Mark Briers
QinetiQ Ltd, Worcestershire, UK

Joint work with: Scott Sisson, Gareth Peters, and Yannan Fan (*UNSW, Australia*)
(Keith Copsey (QinetiQ) and Richard Everitt (Bristol))

Motivation

1. Specification of actual (marginal/joint) ABC target distribution did not explicitly appear within the literature

2. Must be possible to pose all ABC-SMC algorithms within a common framework (i.e. SMC samplers framework)

3. Biased/unbiased argument not resolved (within literature)

4. Algorithms with computational cost lower than O(N^2) and whose particle weights don't degenerate to zero

Contents

Background: Notation and Terminology

- Bayesian inference proceeds through the posterior distribution:

$$\pi(\theta|y) \propto \pi(y|\theta)\pi(\theta) \qquad \theta \in \Theta,\ y \in \mathcal{X}$$

- Increasing interest in scenarios where the likelihood function is intractable
- Standard simulation procedures based on repeated likelihood evaluations (eg MCMC, SMC) require modification
- Collection of methods utilised for posterior simulation in the presence of intractable likelihood has become known as "Approximate Bayesian Computation" (ABC)

Joint Space Representation

- ABC methods employ standard procedure of embedding target posterior within augmented model from which sampling is viable

$$\pi(\theta, x|y) \propto \pi(y|x,\theta)\pi(x|\theta)\pi(\theta),$$

  where $x \in \mathcal{X}$

- In ABC setting $x \sim \pi(x|\theta)$ is an artificial dataset distributed according to the likelihood

- $\pi(y|x,\theta)$ weights the intractable posterior, with high density in regions where summary statistics $T(x) \approx T(y)$ of x and y are similar

Weighting function

- Commonly utilised but "inefficient" form is:

$$\pi(y|x,\theta) \propto \begin{cases} 1 & \text{if } \rho(T(y),T(x)) \leq \epsilon \\ 0 & \text{else} \end{cases}$$

- Exclusively rewards those *T(x)* within an ϵ-tolerance of *T(y)* as measured by a distance function ρ

- Alternative forms of $\pi(y|x,\theta)$ have been evaluated, such as Gaussian and Epanechnikov kernels

Marginal Space Representation

- The target marginal distribution is:

$$\pi_{ABC}(\theta|y) \propto \int_{\mathcal{X}} \pi(y|x,\theta)\pi(x|\theta)\pi(\theta)dx$$

- If $\pi(y|x,\theta)$ for *T(y)*=*T(x)* and 0 elsewhere, and if *T* is ("exactly") sufficient, then $\pi_{ABC}(\theta|y) \equiv \pi(\theta|y)$ reduces to the true posterior

Marginal Space Representation: Monte Carlo Approximation

- The target marginal distribution can be approximated as follows:

$$
\begin{aligned}
\pi_{ABC}(\theta|y) &\propto \pi(\theta)\int_{\mathcal{X}} \pi(y|x,\theta)\pi(x|\theta)dx \\
&= \pi(\theta)\mathbb{E}_{\pi(x|\theta)}[\pi(y|x,\theta)] \\
&\approx \frac{\pi(\theta)}{S}\sum_{s=1}^{S}\pi(y|x^s,\theta),
\end{aligned}
$$

where $x^1,\ldots,x^S \sim \pi(x|\theta)$ are draws from the intractable likelihood

ABC-MCMC: Targeting the marginal space

- Sisson et al (2008) argue that ABC-MCMC (Marjoram et al (2003)) is explicitly targeting the marginal distribution $\pi_{ABC}(\theta|y)$
- Proof of detailed balance in Marjoram et al (2003) relates to marginal distribution $\pi_{ABC}(\theta|y)$

ABC-MCMC Algorithm $(S = 1)$

At stage t:

1. Generate $\theta \sim q(\theta_t, \theta)$ from a proposal distribution.
2. Generate $x \sim \pi(x|\theta)$ from the likelihood.
3. With probability $\min\left\{1, \frac{\pi(y|x,\theta)\pi(\theta)q(\theta,\theta_t)}{\pi(y|x_t,\theta_t)\pi(\theta_t)q(\theta_t,\theta)}\right\}$ accept $\theta_{t+1} = \theta$, $(x_{t+1} = x)$ otherwise set $\theta_{t+1} = \theta_t$, $(x_{t+1} = x_t)$.
4. Increment $t = t + 1$ and go to 1.

ABC-MCMC: Targeting the joint space

- The ABC-MCMC algorithm proceeds exactly as before
- MH-ratio:

$$\frac{\pi(\theta_{t+1}, x_{t+1}|y)}{\pi(\theta_t, x_t|y)} \frac{q(\theta_t, x_t|\theta_{t+1}, x_{t+1})}{q(\theta_{t+1}, x_{t+1}|\theta_t, x_t)}$$

- Let $q(\theta_{t+1}, x_{t+1}|\theta_t, x_t) \triangleq \pi(x_{t+1}|\theta_{t+1})q(\theta_{t+1}|\theta_t)$

  Then MH-ratio:

$$\frac{\pi(\theta_{t+1})\pi(y|x_{t+1}\theta_{t+1})}{\pi(\theta_t)\pi(y|x_t, \theta_t)} \frac{q(\theta_t|\theta_{t+1})}{q(\theta_{t+1}|\theta_t)}$$

- This is the same expression as in the marginal ABC-MCMC algorithm (if *S*=1), with the same sequence of algorithmic operations!

SMC Samplers: Notation

- SMC samplers operates on an artificial sequence of target distributions
  - Defined to admit the desired target distribution as a marginal
- (IS) weight expression:

$$w_n\left(\theta_{1:n}\right) = \frac{\widetilde{\phi}_n(\theta_{1:n})}{\widetilde{\mu}_n\left(\theta_{1:n}\right)} = \frac{\phi_n(\theta_n)\prod_{k=1}^{n-1} L_k\left(\theta_{k+1}, \theta_k\right)}{\mu_1\left(\theta_1\right)\prod_{k=2}^{n} M_k\left(\theta_{k-1}, \theta_k\right)}.$$

$$w_n\left(\theta_{n-1}, \theta_n\right) = \frac{\phi_n(\theta_n) L_{n-1}\left(\theta_n, \theta_{n-1}\right)}{\phi_{n-1}(\theta_{n-1}) M_n\left(\theta_{n-1}, \theta_n\right)}.$$

ABC-SMC: targeting the marginal space

- Sequence of target distributions:

$$\phi_n(\theta_n) = \pi_{ABC,n}(\theta_n|y) \propto \int_{\mathcal{X}} \pi_n(y|x,\theta)\pi(x|\theta)\pi(\theta)dx$$

for $\epsilon_1 \geq \epsilon_2 \geq \ldots \geq \epsilon_n$.

- E.g.

$$\pi_n(y|x,\theta) \propto \begin{cases} 1 & \text{if } \rho(T(y),T(x)) \leq \epsilon_n \\ 0 & \text{else,} \end{cases}$$

ABC-SMC: targeting the marginal space

- Weighting function (Sisson et al, 2007):

$$
\begin{aligned}
w_t &\propto \frac{\pi_{ABC,n}(\theta_n|y)}{\pi_{ABC,n}(\theta_{n-1}|y)} \frac{L_{n-1}(\theta_{n-1}|\theta_n)}{K_n(\theta_n|\theta_n)} \\
&\approx \frac{\frac{\pi(\theta_n)}{B_n} \sum_{b=1}^{B_n} \pi_n(y|x_n^b, \theta_n)}{\frac{\pi(\theta_{n-1})}{B_{n-1}} \sum_{b=1}^{B_{n-1}} \pi_{n-1}(y|x_{n-1}^b, \theta_{n-1})} \frac{L_{n-1}(\theta_{n-1}|\theta_n)}{K_n(\theta_n|\theta_n)}
\end{aligned}
$$

where this expression follows by inserting the Monte Carlo approximation to the marginal distribution into the numerator and denomenator

ABC-SMC: targeting the marginal space

- Weighting function (Sisson et al, 2007):

$$
\begin{aligned}
w_t &\propto \frac{\pi_{ABC,n}(\theta_n|y)}{\pi_{ABC,n}(\theta_{n-1}|y)} \frac{L_{n-1}(\theta_{n-1}|\theta_n)}{K_n(\theta_n|\theta_n)} \\
&\approx \frac{\frac{\pi(\theta_n)}{B_n} \sum_{b=1}^{B_n} \pi_n(y|x_n^b, \theta_n)}{\frac{\pi(\theta_{n-1})}{B_{n-1}} \sum_{b=1}^{B_{n-1}} \pi_{n-1}(y|x_{n-1}^b, \theta_{n-1})} \frac{L_{n-1}(\theta_{n-1}|\theta_n)}{K_n(\theta_n|\theta_n)}
\end{aligned}
$$

where this expression follows by inserting the Monte Carlo approximation to the marginal distribution into the numerator and denomenator

- With B_N = 1 expression can be compared to (marginal) MCMC

ABC-SMC: targeting the marginal space

- Weighting function (Sisson et al, 2007):

$$
\begin{aligned}
w_t &\propto \frac{\pi_{ABC,n}(\theta_n|y)}{\pi_{ABC,n}(\theta_{n-1}|y)} \frac{L_{n-1}(\theta_{n-1}|\theta_n)}{K_n(\theta_n|\theta_n)} \\
&\approx \frac{\frac{\pi(\theta_n)}{B_n}\sum_{b=1}^{B_n} \pi_n(y|x_n^b, \theta_n)}{\frac{\pi(\theta_{n-1})}{B_{n-1}}\sum_{b=1}^{B_{n-1}} \pi_{n-1}(y|x_{n-1}^b, \theta_{n-1})} \frac{L_{n-1}(\theta_{n-1}|\theta_n)}{K_n(\theta_n|\theta_n)}
\end{aligned}
$$

  where this expression follows by inserting the Monte Carlo approximation to the marginal distribution into the numerator and denomenator

- With B_N = 1 expression can be compared to (marginal) MCMC

- *Biased algorithm: ratio of Monte Carlo estimates*

  - Bias reduces for large B_N but can be significant for B_N = 1

ABC-SMC: targeting the marginal space

- Sisson et al, 2009 (errata note for Sisson et al, 2007 paper)
- Directly evaluate the importance sampling distribution, hence avoiding the need for Monte Carlo approximation in the denominator of $\pi_{ABC,n}(\theta_n|y)$
- This corresponds to using (an approximation to) the optimal *L* kernel

$$\begin{aligned} w_n(\theta_n) &= \frac{\pi_{ABC,n}(\theta_n|y)}{\int_\Theta \pi_{n-1}(\theta_{n-1}|y)K(\theta_n|\theta_{n-1})d\theta_{n-1}} \\ &\approx \frac{\pi(\theta_n)}{\sum_{j=1}^N w_{n-1}(\theta_{n-1}^{(j)})k(\theta_n|\theta_{n-1})} \end{aligned}$$

- Unbiased algorithm, but now with O(N^2) computation

ABC-SMC: targeting the joint space

- Incremental weight on joint space given by:

$$w_n[(\theta_{n-1}, x_{n-1}), (\theta_n, x_n)] \propto \frac{\pi_n(\theta_n, x_n|y) L_{n-1}[(\theta_n, x_n), (\theta_{n-1}, x_{n-1})]}{\pi_{n-1}(\theta_{n-1}, x_{n-1}|y) M_n[(\theta_{n-1}, x_{n-1}), (\theta_n, x_n)]}$$

- As in ABC-MCMC, interpretation on joint space is through the use of "standard" SMC sampler methodology

- For general choice of L_{n-1}, a (generally intractable) normalising constant appears in weight expression which renders this model impractical

Some unbiased SMC-ABC algorithms

- ABC-SMC(MCMC)

$$w_n[(\theta_{n-1}, x_{n-1}), (\theta_n, x_n)] \propto \frac{\pi_n(\theta_{n-1}, x_{n-1}|y)}{\pi_{n-1}(\theta_{n-1}, x_{n-1}|y)} = \frac{\pi_n(y|x_{n-1}, \theta_{n-1})}{\pi_{n-1}(y|x_{n-1}, \theta_{n-1})},$$

- Note that it is possible for all weights to be assigned zero weight
  - Del Moral et al (2008) introduce automated ϵ_n selection scheme
  - Sisson et al (2007) include a PRC step
- ABC-SMC(PMC)
  - O(N^2) algorithm

Application

- Estimate: number and parameters of a CBRN release
- Variable number of releases of a given material, specified by:
  - Release location; Release times; Release amounts
- Grid of sensors located downwind of the releases
- Dispersion model for how the plumes propagate over time
  - This can be a complex stochastic process, varying dependent on atmospheric conditions (e.g. wind profiles), and the environment (e.g. buildings)
- Sensor model which relates the actual concentration to the measurements returned by the sensors
  - Again can be complex

ABC-SMC (RJMCMC)

- Use of Trans-dimensional Metropolis-Hastings kernel within ABC-SMC
- As usual, difficult to specify efficient transition distribution
  - Results in severe particle depletion

- Need to resolve this issue, or look at alternative methodology

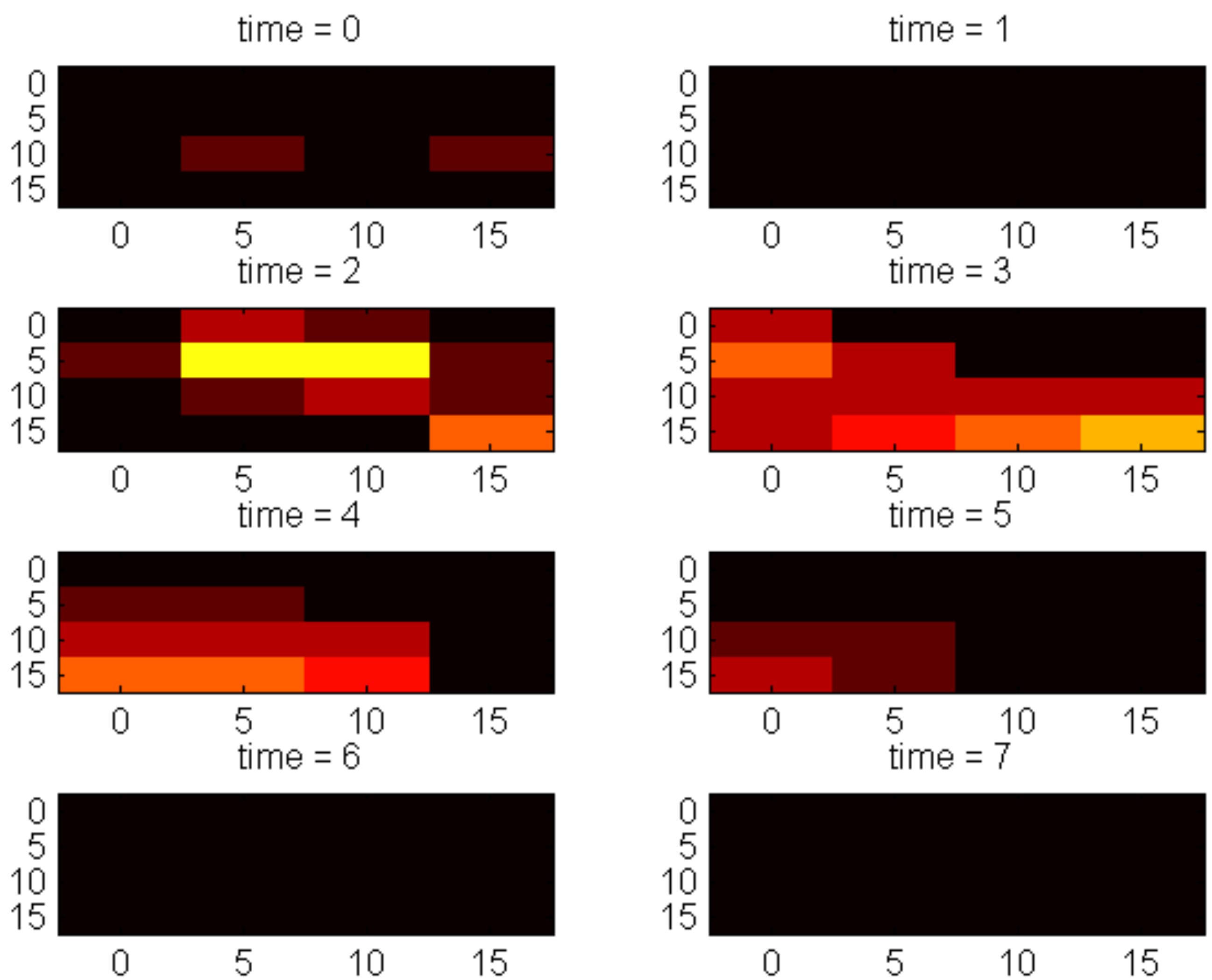
Three releases

- Simulated three-release scenario

Release	Time	x-coord	Mass
1st	0	(40,15)	250
2nd	0	(30,5)	300
3rd	0.8	(40,15)	200

Ground-truth release details

Cross wind of (-11,0)



Sensor measurements (4-by-4 arrays)

Results

Release	MAP number of releases	Time	x-coord	y-coord	Mass
Ground-truth	*3*	*0* *0* *0.8*	*40* *30* *40*	*15* *5* *15*	*250* *300* *200*
ABC-SMC	Fixed = 3	-0.43 -0.21 0.1	45.0 32.0 45.9	15.7 5.2 16.0	205.7 263.0 198.7
Trans-dimensional ABC-SMC	3	-0.97 -0.03 -0.36	53.4 30.4 49.6	16.7 5.3 15.2	302.9 261.6 162.5

- Although errors exist in the values, once effects from early or late release predictions are taken into account one can clearly link the estimates to the ground-truth values.
- Hence one can obtain reasonable estimates, even when the number of releases is unknown.

Conclusions

1. *Specification of actual (marginal/joint) ABC target distribution did not explicitly appear within the literature*
   Marginal and joint space representations exposed (explicitly)

2. *Must be possible to pose all ABC-SMC algorithms within a common framework (i.e. SMC samplers framework)*
   Clarity on use of ABC-SMC for both marginal and joint space representations

3. *Biased/unbiased argument not resolved (within literature)*
   Bias argument now (hopefully) resolved

4. *Algorithms with computational cost lower than O(N^2) and whose particle weights don't degenerate to zero*
   Only one SMC algorithm that has cost lower than O(N^2)?

- Novel application of ABC-SMC(RJ-MCMC) presented

- Questions?