Efficient learning in Approximate Bayesian Computation

Mohammed Sedki, Pierre Pudlo

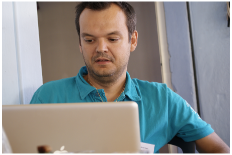
Université Montpellier 2

ABC in London: May 5th 2011

Efficient learning in Approximate Bayesian Computation

Mohammed Sedki, Pierre Pudlo

Université Montpellier 2



ABC in London: May 5th 2011

How to reduce computation time in ABC?

$$\pi_\varepsilon(\boldsymbol{\theta}|\mathbf{y}_{\text{obs}}) \propto \pi(\boldsymbol{\theta})\,\ell(\mathbf{z}|\boldsymbol{\theta})\,\mathbf{1}\{d(\mathbf{z},\mathbf{y}_{\text{obs}}) \leq \varepsilon\}$$

What is time consuming?

- simulations from the model

What is **inefficient** with acceptation-rejection algorithm?

- Sending $\boldsymbol{\theta}$'s everywhere with prior distribution
- Difficult to get a simulated $\mathbf{z}$ near the observed $\mathbf{y}_{\text{obs}}$

The idea

- Avoid the many rejected simulations when $\boldsymbol{\theta} \sim$ prior
- If parameter $\boldsymbol{\theta} \sim$ posterior: easier to have $d(\mathbf{z}, \mathbf{y}_{\text{obs}})$ small

$\Longrightarrow$ Introduce a temporal dimension (**Sequential techniques with T iterations**) to learn gradually the posterior

Sequential algorithms

Litterature

(1) ABC-Partial Rejection Control (PRC)
of Sisson, Fan and Tanaka (*PNAS* 2007, 2009)

(2) ABC-Population Monte Carlo (PMC)
of Beaumont, Cornuet, Marin and Robert (*Biometrika* 2009)

(3) Parallel sequential ABC
of Toni, Welch, Strelkowa, Ipsen and Stumpf (*JRSI*, 2009)

(4) ABC-Sequential Monte Carlo (SMC)
of Del Moral, Doucet and Jasra (2009)

(5) Drovandi and Pettitt (*Biometrics*, 2011)

Main difficulty: How to choose the tolerance thresholds

$$\varepsilon_1 \geq \cdots \geq \varepsilon_T$$

over T iterations?

None of them are really satisfactory!

ABC-Sequential Monte Carlo sampler

Assume: $\varepsilon_1 \geq \cdots \geq \varepsilon_T$ are fixed

At each iteration $1 \leq t \leq T$

- from a sample of $(\boldsymbol{\theta}_i^{(t)}, \mathbf{z}_i^{(t)})$ $(i = 1, \ldots N)$ distributed according to $\pi_{\varepsilon_t}(\cdot|\mathbf{y}_{\text{obs}})$

(1) pick one of them which satisfies $d(\mathbf{z}_i^{(t)}, \mathbf{y}_{\text{obs}}) \leq \varepsilon_{t+1}$

(2) move it according to a MCMC kernel $\pi_{\varepsilon_{t+1}}(\cdot|\mathbf{y}_{\text{obs}})$-invariant

- return to step (1) until we end with a new sample of size N: $(\boldsymbol{\theta}_i^{(t+1)}, \mathbf{z}_i^{(t+1)})$ $(i = 1, \ldots N)$ distributed according to $\pi_{\varepsilon_{t+1}}$

New adaptive scheme

- how to choose $\varepsilon_1 \geq \varepsilon_2 \geq \cdots \geq \varepsilon_T$?
- calibrated for time saving

ABC-Sequential Monte Carlo sampler

Assume: $\varepsilon_1 \geq \cdots \geq \varepsilon_T$ are fixed

At each iteration $1 \leq t \leq T$

- from a sample of $(\boldsymbol{\theta}_i^{(t)}, \mathbf{z}_i^{(t)})$ $(i = 1, \ldots N)$ distributed according to $\pi_{\varepsilon_t}(\cdot|\mathbf{y}_{\text{obs}})$

(1) pick one of them which satisfies $d(\mathbf{z}_i^{(t)}, \mathbf{y}_{\text{obs}}) \leq \varepsilon_{t+1}$

(2) move it according to a MCMC kernel $\pi_{\varepsilon_{t+1}}(\cdot|\mathbf{y}_{\text{obs}})$-invariant

- return to step (1) until we end with a new sample of size N: $(\boldsymbol{\theta}_i^{(t+1)}, \mathbf{z}_i^{(t+1)})$ $(i = 1, \ldots N)$ distributed according to $\pi_{\varepsilon_{t+1}}$

New adaptive scheme

- how to choose $\varepsilon_1 \geq \varepsilon_2 \geq \cdots \geq \varepsilon_T$?
- calibrated for time saving

An ABC–Hastings-Metropolis $\pi_{\varepsilon_{t+1}}(\cdot|\mathbf{y}_{\text{obs}})$-invariant

Assume: prior = uniform on Θ_{prior}.

Let $\quad \boldsymbol{\theta}^t \sim \pi_{\varepsilon_{t+1}}(\cdot|\mathbf{y}_{\text{obs}}) \quad$ and

(R-W) Draw $\widetilde{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{\theta}^t, \Sigma), \quad \widetilde{\mathbf{z}}|\widetilde{\boldsymbol{\theta}} \sim \ell(\mathbf{z}|\widetilde{\boldsymbol{\theta}})$

(A-R) Set $\boldsymbol{\theta}^{t+1} = \begin{cases} \widetilde{\boldsymbol{\theta}} & \text{if } d(\widetilde{\mathbf{z}}, \mathbf{y}) \leq \varepsilon_{t+1}, \text{ and } \widetilde{\boldsymbol{\theta}} \text{ is in } \Theta_{\text{prior}} \\ \boldsymbol{\theta}^t & \text{otherwise.} \end{cases}$

Proposition (Majoram *et al.*, 2003)

Then, whatever Σ, $\boldsymbol{\theta}^{t+1} \sim \pi_{\varepsilon_{t+1}}(\cdot|\mathbf{y}_{\text{obs}})$

Notation

- average acceptance probability

$$\rho_{t+1} := \mathbb{P}_{\pi_{\varepsilon_{t+1}}(\cdot|\mathbf{y}_{\text{obs}})}\left(\boldsymbol{\theta}^t \neq \boldsymbol{\theta}^{t+1}\right)$$

Iteration t: from ε_t to ε_{t+1}

Input: $(\boldsymbol{\theta}_i^t, \mathbf{z}_i^t)$, $i = 1, \ldots, N$ distributed according to $\pi_{\varepsilon_t}(\cdot|\mathbf{y}_{\text{obs}})$

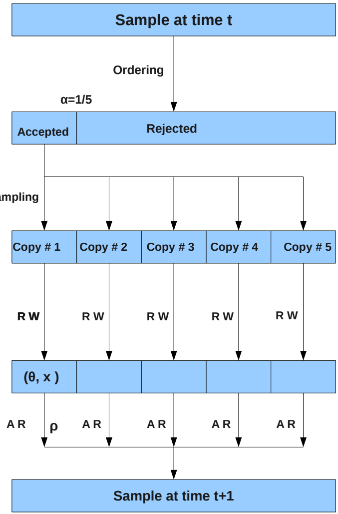
- **Order the sample**: $d(\mathbf{z}_1^t, \mathbf{y}_{\text{obs}}) \leq \cdots \leq d(\mathbf{z}_N^t, \mathbf{y}_{\text{obs}})$
- **Acception-Rejection**: a proportion $\alpha = \alpha_{t+1}$ is kept and set $\varepsilon_{t+1} = d(\mathbf{z}_{\alpha N}^t, \mathbf{y}_{\text{obs}})$
- **Copying**: duplicate to get a sample of size N
- **MCMC**: Apply one step of the Markov Chain and set $\hat{\rho}_{t+1}$ = proportion of accepted movements

Two pitfalls

α_{t+1} too small $\Longrightarrow$ $\hat{\rho}_{t+1} \approx 0$ $\Longrightarrow$ too many duplications

α_{t+1} too large $\Longrightarrow$ ε_{t+1} too large $\Longrightarrow$ too many iterations

Trade-off: α_{t+1} is adapted on the 1$^{\text{st}}$ copy s.t. $\alpha_{t+1} + \rho_{t+1} = 1$



Adaptive scheme

Calibration

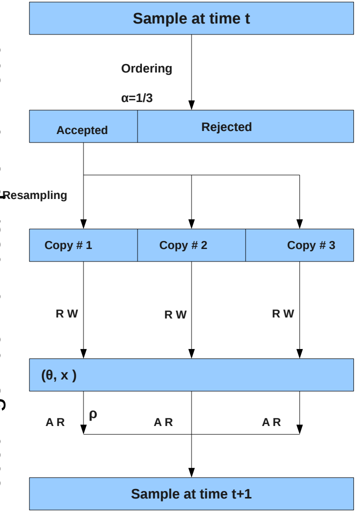
Increase α **from** $1/L$ **to** 1 **by** $1/L$

Compute on copy $\sharp$ 1

- $\varepsilon_{t+1} = d(\mathbf{z}^{(t)}_{[\alpha N]}, \mathbf{y}_{\text{obs}})$
- proposed $(\widetilde{\boldsymbol{\theta}}_i, \mathbf{z}_i)$'s
- ρ_{t+1} =proportion of pairs that have moved during MCMC

Until $\alpha + \rho_{t+1} \geq 1$.

Adaptive scheme



Calibration

Increase α **from** $1/L$ **to** 1 **by** $1/L$

Compute on copy $\sharp$ 1

- $\varepsilon_{t+1} = d(\mathbf{z}^{(t)}_{[\alpha N]}, \mathbf{y}_{\text{obs}})$
- proposed $(\widetilde{\boldsymbol{\theta}}_i, \mathbf{z}_i)$'s
- ρ_{t+1} =proportion of pairs that have moved during MCMC

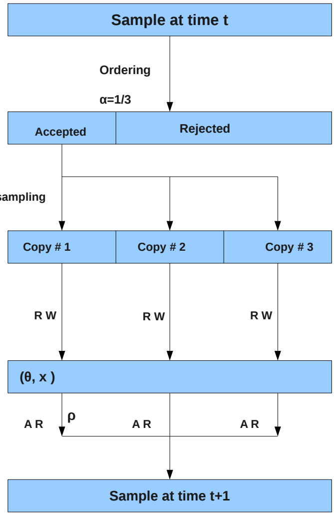
Until $\alpha + \rho_{t+1} \geq 1$**.**

When α increases,

- Old copy $\sharp$1 is nested into the new one
- Many of the proposed $(\widetilde{\boldsymbol{\theta}}_i, \mathbf{z}_i)$

are already computed

Adaptive scheme



Calibration

Increase α **from** $1/L$ **to** 1 **by** $1/L$

Compute on copy $\sharp$ 1

- $\varepsilon_{t+1} = d(\mathbf{z}^{(t)}_{[\alpha N]}, \mathbf{y}_{\text{obs}})$
- proposed $(\widetilde{\boldsymbol{\theta}}_i, \mathbf{z}_i)$'s
- ρ_{t+1} =proportion of pairs that have moved during MCMC

Until $\alpha + \rho_{t+1} \geq 1$.

When α increases,

- Old copy $\sharp$1 is nested into the new one
- Many of the proposed $(\widetilde{\boldsymbol{\theta}}_i, \mathbf{z}_i)$

are already computed

At the end

Apply MCMC on the other copies

Initialization and stopping rule

Get a first rough approximation of the posterior

Draw many pairs $(\boldsymbol{\theta}_i, \mathbf{z}_i)$ from $\pi(\boldsymbol{\theta})\ell(\mathbf{z}|\boldsymbol{\theta})$
until var(kept) $\ll$ var(prior)

where var(kept) $=$ variance of the N closest to $\mathbf{y}_{\text{obs}}$

Initialization and stopping rule

Get a first rough approximation of the posterior

Draw many pairs $(\boldsymbol{\theta}_i, \mathbf{z}_i)$ from $\pi(\boldsymbol{\theta})\ell(\mathbf{z}|\boldsymbol{\theta})$
until $\quad \text{var(kept)} \ll \text{var(prior)}$

where $\quad \text{var(kept)} = \text{variance of the } N \text{ closest to } \mathbf{y}_{\text{obs}}$

Warning

When it is impossible, prior $\approx$ posterior
$\longrightarrow$ stop there and do not run the sequential algorithm!

Initialization and stopping rule

Get a first rough approximation of the posterior

Draw many pairs $(\boldsymbol{\theta}_i, \mathbf{z}_i)$ from $\pi(\boldsymbol{\theta})\ell(\mathbf{z}|\boldsymbol{\theta})$
until var(kept) $\ll$ var(prior)

where var(kept) $=$ variance of the N closest to $\mathbf{y}_{\text{obs}}$

Warning

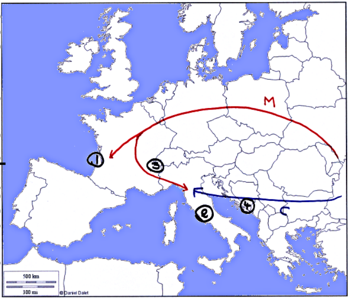
When it is impossible, prior $\approx$ posterior
$\longrightarrow$ stop there and do not run the sequential algorithm!

Stop rule of the sequential algorithm

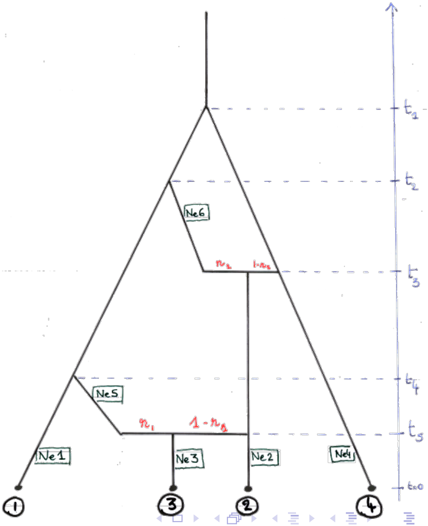
stop at time T when
average acceptance probability in H-M: $\rho_T \leq 0.1$

Illustration in population genetics

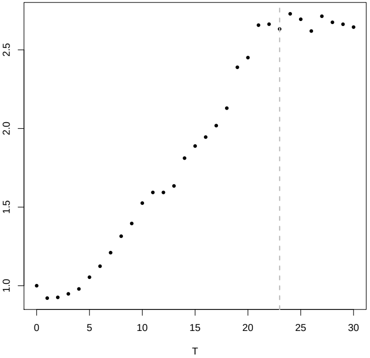
Invasion of Europa by the honeybee



Hence, a coalescence process on each branch of the following scenario:



Efficiency on the illustration



$$\text{Time factor} = \frac{\text{Numb. of simu. in classical ABC}}{\text{Numb. of simu in our proposal}}$$

with

- equal final tolerance threshold
- equal (effective) sample size

The End

The End

Any questions ?

Calibration of Σ in MCMC

Remember: prior = uniform on Θ_{prior}.

Let $\boldsymbol{\theta}_j^t \sim \pi_{\varepsilon_{t+1}}(\cdot|\mathbf{y}_{\text{obs}})$ and

(R-W) Draw $\widetilde{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{\theta}_j^t, \Sigma_t)$, $\widetilde{\mathbf{z}}|\widetilde{\boldsymbol{\theta}} \sim \ell(\mathbf{z}|\widetilde{\boldsymbol{\theta}})$

(A-R) Set $\boldsymbol{\theta}_j^{t+1} = \begin{cases} \widetilde{\boldsymbol{\theta}} & \text{if } d(\widetilde{\mathbf{z}}, \mathbf{y}) \le \varepsilon_{t+1}, \text{ and } \widetilde{\boldsymbol{\theta}} \text{ is in } \Theta_{\text{prior}} \\ \boldsymbol{\theta}^t & \text{otherwise.} \end{cases}$

(1) Compute Σ_{prior}, variance of prior distr.

(2) Find β such that $\mathbb{P}_{\boldsymbol{\theta}\sim\pi}\left(\mathcal{N}(\boldsymbol{\theta}, \beta\Sigma_{prior}) \in \Theta_{\text{prior}}\right) \approx 0.6$

Then, $\Sigma_t = \beta \times \text{Var}\,(\boldsymbol{\theta}_j^t)$