Evaluation of Tolerance Selection Strategies and Multifidelity Techniques in ABC Methods

Graziele Sena

November 11-14, 2024, Maceió, AL, Brazil

Contents

Introducing Our Problem

Some context

- Mathematical models are essential tools for understanding complex systems and predicting outcomes.
- Standard Bayesian inference:
  - Posterior simulation algorithms (MCMC, importance sampling, etc).
- Computational intractability of $L(y|\theta)$.
- Unable to numerically evaluate likelihood for any θ.

Approximate Bayesian Computation

Approximate Bayesian Computation (ABC) methods effectively approximate posterior distribution to make model calibration.

Contents

Objectives

This work focuses on the analysis of the tolerance techniques choice and the application of multifidelity process in ABC methodology.

- Compare three different **tolerance strategies** within ABC-SMC.
- Explore the integration of **multifidelity techniques** to reduce computational costs while maintaining or improving result accuracy.

Contents

In Bayesian framework:

$$p(\theta|y) \propto \mathcal{L}(y|\theta)\pi(\theta). \tag{1}$$

In ABC framework:

$$p(\theta|y) \propto p(\theta|\rho(y, y_{obs}) \leq \epsilon), \tag{2}$$

where,

- $\rho(.)$: some distance measure;
- y: simulated value;
- y_{obs}: observed value,
- ϵ: tolerance value.

How Does ABC Work?

Simple ABC Mechanism:

1. Sample θ^* from the prior distribution $p(\theta)$;
2. Simulate data set from the model, using parameter θ^*, to get D^*;
3. If D is "close enough" of D^*, accept θ^*; otherwise, reject,
4. Repeat until N particles (the parameter values or parameter sets) $\Theta^* = \{\theta_i^*; i = 1, \ldots, N\}$ are accepted.

Note: *"close enough" could be if $\|D - D^*\| \leq \epsilon$ for small ϵ.*

Note: *Computation increases exponentially as accuracy increases (i.e. as $\epsilon \to 0$).*

- **Sequential Monte Carlo Approximate Bayesian Computation** (ABC SMC):
  - The algorithm starts with a **higher tolerance level** and **gradually decreases it over iterations**.
  - Each particle (sample) has a **weight**, allowing the method to prioritize particles that better represent the posterior distribution in subsequent iterations.
  - **Kernel functions** are applied to **perturb particles** to create diversity.

ABC Basics

Choice of Tolerances

One fundamental point is the **choice of tolerances** ϵ for ABC

$$\|D - D^*\| \leq \epsilon \tag{3}$$

- Choosing an appropriate tolerance level is crucial for balancing accuracy and computational efficiency in ABC.
  - The choice of tolerance may be using predetermined vectors (trial and error).
  - Or from adaptive methods.

Another fundamental challenge in ABC methods is the need for a large number of simulations

- One way to overcome this is using models that can be simulated more cheaply through the multifidelity technique.
- Combines simulation with different levels of fidelity to improve the efficiency of the inference process.

Contents

Tolerance Strategies in ABC

The Choice of Tolerances

- The choice of tolerance values significantly impacts the computational efficiency of ABC methods.
  - Fixed.
  - Percentile-based.
  - Percentage-based.

Tolerance Choice Implemented

- Fixing values in advance:
  - Based in prior knowledge and **empirical tests** performed with the model.

Tolerance Choice Implemented

- Adaptive percentile and percentage selection:
  - Using the value corresponding to the percentile or percentage in the ordered distance vector (d^T_{t-1}) accepted in the previous iteration $t-1$.

Contents

Multifidelity Technique

in ABC context

- One way to improving the efficiency is the use of multifidelity models:
  - Combines less expensive models with more costly ones.

Numerical Methods

- **Euler Method**
- **Richardson Extrapolation**

Multifidelity Cost Comparative

in ABC-SMC with Multifidelity

By using this multifidelity strategy, the computational cost becomes proportional to the total number of Euler evaluations.

Evaluation of the Euler function with **step size h** → Additional evaluation with **step size h/2** → Number of samples + 2 x number of accepted particles

$$\text{Total Cost} = N_{h_{samples}} + 2 \cdot N_{\text{accepted}} \tag{4}$$

$$\text{Total Savings} = \left(N_{h_{samples}} + 2 \cdot N_{\text{accepted}}\right) - 2 \cdot N_{samples\frac{h}{2}} \tag{5}$$

Contents

Epidemiological Models

Susceptible - Infected - Recovery

- β: infection rate.
- γ: recovery rate.

SIR model ODE's

$$\begin{aligned} \frac{dS}{dt} &= -\beta S I \, , \\ \frac{dI}{dt} &= \beta S I - \gamma I \, , \\ \frac{dD}{dt} &= \gamma I \, . \end{aligned} \tag{6}$$

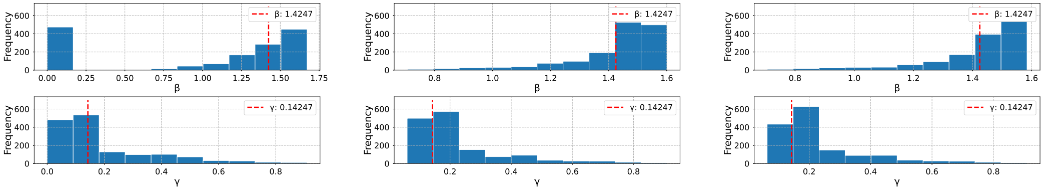
Contents

- Fixed.
- Percentile-based.
- Percentage-based.

Results

Tolerances Strategies with SIR model

Figure: Comparison of histograms for the posterior distribution of β and γ for each tolerance strategy with $pop = 5$ and $h = 0.25$.



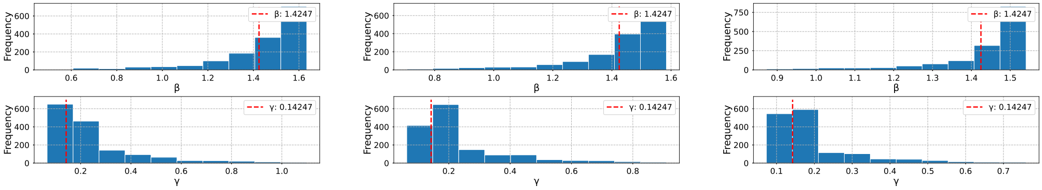
(a) Fixed tolerance method.   (b) Percentile tolerance method.   (c) Percentage tolerance method.

- We utilized our proposed approach with a **multifidelity** and **percentile-based** tolerance selection strategy.

Results

Multifidelity Techniques and Percentile-based Tolerance Strategy with SIR model

Figure: Comparison of the histograms of β and γ for percentile tolerance strategies with multifidelity techniques for $pop = 5$ and $h = 0.5, 0.25, 0.125$.

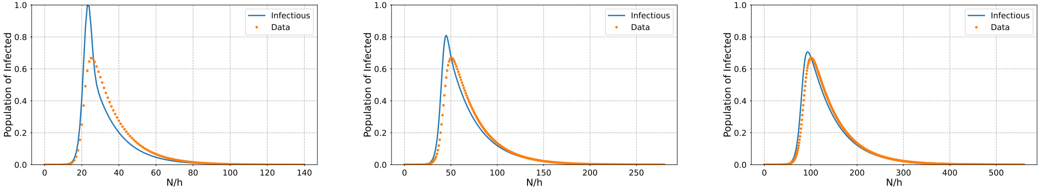


(a) $h = 0.5$.

(b) $h = 0.25$.

(c) $h = 0.125$.

Figure: Comparison of the infects population curve of SIR model for percentile tolerance strategies with multifidelity techniques for $pop = 5$, $N = 70$, $h = 0.5, 0.25, 0.125$, line in blue, and the observed data, orange dots.



(a) $h = 0.5$. (b) $h = 0.25$. (c) $h = 0.125$.

Strategies for $h = 0.5, 0.25, 0.125$

- Initially, all methods yield good results;

Table: Parameters with smallest distance with step size of $h = 0.5, 0.25$ and 0.125, and $pop = 5$.

	β	γ	dmin
$h = 0.5$	1.6333	0.1482	1.4247
$h = 0.25$	1.5832	0.1385	0.3901
$h = 0.125$	1.5176	0.1421	0.1610
$h = 0.167$	1.5516	0.1432	0.1868
$h = 0.083$	1.4855	0.1428	0.1310
$h = 0.042$	1.4538	0.1425	0.0972

- **Reference Parameter**:
  - Transmission rate: $\beta = 1.4247$
  - Recovery rate: $\gamma = 0.14286$

Table: Number of samples necessary for each population in $pop = 5$ round, with $h = 0.5, 0.25$ and 0.125.

	$pop1$	$pop2$	$pop3$	$pop4$	$pop5$
$h = 0.5$	1181	1121	892	923	915
$h = 0.25$	1770	925	848	1127	989
$h = 0.125$	2657	1117	926	1328	1157

Table: Number of samples necessary for each population in $pop = 5$ round, with $h = 0.5, 0.25$ and 0.125.

	$pop1$	$pop2$	$pop3$	$pop4$	$pop5$
$h = 0.5$	1181	1121	892	923	915
$h = 0.25$	1770	925	848	1127	989
$h = 0.125$	2657	1117	926	1328	1157

Table: Cost comparison between $h = 0.5$ and $h = 0.25$, and between $h = 0.25$ and $h = 0.125$, for Euler and multifidelity.

	$pop1$	$pop2$	$pop3$	$pop4$	$pop5$
$0.5/0.25$	-1759	-129	-204	-731	-463
$0.25/0.125$	-2944	-709	-404	-929	-725

Contents

8 Conclusions

Conclusions

- The adaptive strategies provide the best results.
- Initial analysis indicates that these strategies do not produce significant variations in the parameter estimates.
- The adjustment of the step size h present improvements.
- The implementation of multifidelity technique provides a more efficient calibration.

To be investigate is:

- Check how the other parameters are correlated.
- Application to complex and high-dimensional models.

LNCC

Thank you all for your attention!