Multifidelity approximate Bayesian computation

Ruth Baker

@ruth_baker
Acknowledgements

• All the multifidelity ABC work was carried out by Dr Thomas Prescott.

• References:


  • Multifidelity approximate Bayesian computation with sequential Monte Carlo parameter sampling. arXiv (2020).
To understand the mechanisms driving collective cell motility, proliferation and death and their contributions to complex biological processes, such as those associated with development, disease and repair.

Goal: to interrogate multiplex quantitative data using validated and biologically realistic mathematical models.
Interdisciplinary methodology

Mathematical modelling
computational simulations

Data analysis and model testing

Experiments: wildtype and perturbation

improves

guides

improves

guides

improves

guides
Neural crest invasion

• Clinical need for a better understanding - failure results in significant morphological abnormalities.

• Model system - diverse cell invasion mechanisms.

Neural crest invasion: mathematics?

- Combinatorially intractable experimentally.
- Mechanistic models are the only solution.
- Translation from laboratory to clinic remains limited.
- Provide a bridge between *in vitro* and *in vivo*.

Neural crest invasion: recent work

- Population heterogeneity crucial for successful invasion.


Leader cell (can sense and respond to chemical cue)

Follower cell

Cells enter migratory domain

Distal target site

Chemical concentration
• Use scratch (wound healing) assays to explore the importance of cell-cell pushing in cell invasion.

How much biological detail do we need to include to faithfully recapitulate biological observations?

Example - *in silico* wound healing

\[ \delta(t) = 29.26 + 0.33t \]
\[ R^2 = 0.94 \]

- Consider a suite of agent-based models, with gradually increasing model complexity:

\[
\frac{d}{dt} x_i(t) = \sum_{i \neq j} F_{ij} + \xi_i
\]

Random movements

Position of cell i

Pairwise interactions

Example - *in silico* wound healing

- Carried out model comparison, using these quantitative data.
- Neglecting finite size effects, together with intercellular forces (e.g. cell pushing), significantly reduces our ability to mimic and predict cell invasion speeds and profiles.

The inverse problem

- Given quantitative data, can we estimate model parameters?

- Parameter inference using a Bayesian framework:

\[ P(\theta | D) \propto P(D | \theta) P(\theta) \]

- For the models we consider, the likelihood is intractable…
Approximate Bayesian computation

• Estimate the posterior using repeated forward simulation:

\[ \text{Algorithm 1 Rejection sampling ABC (ABC-RS)} \]

\textbf{Input:} Data \( y_{\text{obs}} \) and neighbourhood \( \Omega_\epsilon \); model \( f(\cdot \mid \theta) \); prior \( \pi \); sample index \( n = 0 \); stopping criterion \( S \).

\textbf{Output:} Weighted sample \( \{\theta_n, w_n\}_{n=1}^N \).

1: \textbf{repeat}
2: \quad \text{Increment } n \leftarrow n + 1.
3: \quad \text{Generate } \theta_n \sim \pi(\cdot).
4: \quad \text{Simulate } y_n \sim f(\cdot \mid \theta_n).
5: \quad \text{Set } w_n = \mathbb{I}(y_n \in \Omega_\epsilon).
6: \textbf{until } S = \text{true}.

\[ \theta_i \quad \rightarrow \quad \text{parameter space} \]

\[ y_i \quad \rightarrow \quad \text{data space} \]

\[ \Omega(\epsilon) \]

Ruth Baker
Approximate Bayesian computation

- Estimate the posterior using repeated forward simulation:

**Algorithm 1** Rejection sampling ABC (ABC-RS)

**Input:** Data $y_{\text{obs}}$ and neighbourhood $\Omega_\epsilon$; model $f(\cdot \mid \theta)$; prior $\pi$; sample index $n = 0$; stopping criterion $S$.

**Output:** Weighted sample $\{\theta_n, w_n\}_{n=1}^N$.

1: repeat
2: Increment $n \leftarrow n + 1$.
3: Generate $\theta_n \sim \pi(\cdot)$.
4: Simulate $y_n \sim f(\cdot \mid \theta_n)$.
5: Set $w_n = I(y_n \in \Omega_\epsilon)$.
6: until $S = \text{true}$.

---

Parameter space $\theta_i$ and data space $y_i$ with neighbourhood $\Omega(\epsilon)$.
Example - repressilator model

\[
\frac{\alpha_0 + \alpha f(p_j)}{\alpha_0 + \alpha f(p_j)} \rightarrow m_i \xrightarrow{1} 0 \\
m_i \xrightarrow{\beta} m_i + p_i \\
p_i \xrightarrow{\beta} 0
\]

for \((i, j) = (1, 3), (2, 1), \text{ and } (3, 2)\)

for \(i = 1, 2, 3\)

\[
f(p) = \frac{K_h^n}{K_h^n + p^n}
\]

\[
\text{for } i = 1, 2, 3
\]
Example - repressilator model

Step 1: sample from the prior

Step 2: simulate and measure distance from data

Step 3: accept if distance below threshold
Example - repressilator model

- Bottleneck - repeated simulation of the model.
- Trade off between simulation time and variance.
• Use a more intelligent exploration of parameter space.
  • Importance sampling, sequential Monte Carlo etc.

• Make the weight (accept / reject) less expensive to calculate.
  • Make simulations less expensive.

• Reduce model dimension, use a surrogate or approximate model, simulate over a shorter time interval, use coarser discretisation of space / time etc.

**Aim of this talk: to demonstrate that we can do both at once, whilst maintaining accuracy.**
Multifidelity ABC: the idea

Each model requires a distance function, data and acceptance threshold e.g.

\[ \Omega_\epsilon(d, y_{\text{obs}}) = \{y \in \mathcal{Y} \mid d(y, y_{\text{obs}}) < \epsilon \} \]
Multifidelity ABC: the problem

- How to combine the outputs of the two models, so that the result is unbiased weights?

- How can we make this process efficient?
Multifidelity ABC: the algorithm

• Attempt to make an “early decision” using the low-fidelity model, and “sometimes” check that decision using the high-fidelity model.

• Decision to check is made uniformly at random, with probability \( \alpha(\tilde{y}, \theta_n) \).

• Here, we will take a simple approach, assuming

\[
\alpha(\tilde{y}, \theta_n) = \eta_1 \mathbb{1}(\tilde{y} \in \tilde{\Omega}(\tilde{\epsilon})) + \eta_2 \mathbb{1}(\tilde{y} \notin \tilde{\Omega}(\tilde{\epsilon}))
\]

• We also assume, for simplicity,

\[
\tilde{\mathcal{Y}} = \mathcal{Y}, \quad \tilde{y}_{\text{obs}} = y_{\text{obs}}, \quad \tilde{d} = d, \quad \tilde{\epsilon} = \epsilon
\]

\[\implies \tilde{\Omega}_{\tilde{\epsilon}} = \Omega_{\epsilon}\]
### Algorithm 4 Rejection sampling multifidelity ABC (MF-ABC-RS)

**Input:** Data $y_{\text{obs}}$ and neighbourhood $\Omega_\epsilon$; prior $\pi$; models $\tilde{f}(\cdot \mid \theta)$, $f(\cdot \mid \tilde{y}, \theta)$; continuation probability function $\alpha = \alpha(\tilde{y}, \theta)$; sample index $n = 0$; stopping condition $S$.

**Output:** Weighted sample $\{\theta_n, w_n\}_{n=1}^N$.

1. repeat
2. Increment $n \leftarrow n + 1$.
3. Generate $\theta_n \sim \pi(\cdot)$.
4. Simulate $\tilde{y}_n \sim \tilde{f}(\cdot \mid \theta_n)$.
5. Set $w_n = \mathbb{I}(\tilde{y}_n \in \Omega_\epsilon)$.
6. Generate $u_n \sim \text{Uniform}(0, 1)$.
7. if $u_n < \alpha(\tilde{y}_n, \theta_n)$ then
   8. Simulate $y_n \sim f(\cdot \mid \tilde{y}_n, \theta_n)$.
   9. Update $w_n \leftarrow w_n + [\mathbb{I}(y \in \Omega_\epsilon) - w_n] / \alpha(\tilde{y}_n, \theta_n)$.
10. end if
11. until $S = \text{true}$.
For non-zero continuation probability, the weighted sample has the correct distribution, the ABC approximation to the posterior induced by the high-fidelity model.
Multifidelity ABC: efficiency

- Effective sample size:

\[ \text{ESS} = \frac{\left( \sum_n w_n \right)^2}{\sum_n w_n^2}. \]

- Observed efficiency - defined as the effective number of samples per time unit:

\[ \frac{\text{ESS}}{T_{\text{total}}} = \frac{\left( \sum_n w_n \right)^2}{\left( \sum_n w_n^2 \right) \left( \sum_n T_n \right)} \]

- Theoretical efficiency:

\[ \psi = \frac{\mathbb{E}(w)^2}{\mathbb{E}(w^2) \mathbb{E}(T)} \]
Theoretical efficiency can be written

\[ \psi(\eta_1, \eta_2) = \frac{\mathbb{E}(w)^2}{\mathbb{E}(w^2)\mathbb{E}(T)} = \frac{Z^2}{\phi(\eta_1, \eta_2)} \]

where

\[ \phi(\eta_1, \eta_2) = \left( W + \left( \frac{1}{\eta_1} - 1 \right) W_{fp} + \left( \frac{1}{\eta_2} - 1 \right) W_{fn} \right) \]

\[ \times (\bar{T}_{lo} + \eta_1 \bar{T}_{hi,p} + \eta_2 \bar{T}_{hi,n}) \]

probability of false positive

probability of false negative

average simulation time for low-fidelity model

average simulation time for high-fidelity model given low-fidelity models close/far
Multifidelity ABC: efficiency

- Derive analytical expressions for the optimal continuation probabilities, given estimates of these quantities.

- In practice: adapt the continuation probabilities “on the fly”, as samples are generated…
Multifidelity ABC: results

• Comparing results for a range of continuation probabilities:
Multifidelity ABC: conclusions

• Multifidelity ABC can provide time savings, through the combined use of high- and low-fidelity models.

• Can “learn” optimal continuation probabilities as the algorithm proceeds, separately controlling rates of checking early acceptance and early rejection.

• Rates of false positives and negatives can be reduced by generating the high-fidelity model output conditional on the low-fidelity model output.

• Enables smaller continuation probabilities and hence simulation cost.
Multifidelity ABC: the algorithm

**Algorithm 4** Rejection sampling multifidelity ABC (MF-ABC-RS)

**Input:** Data $y_{\text{obs}}$ and neighbourhood $\Omega_\epsilon$; prior $\pi$; models $\tilde{f}(\cdot \mid \theta)$, $f(\cdot \mid \tilde{y}, \theta)$; continuation probability function $\alpha = \alpha(\tilde{y}, \theta)$; sample index $n = 0$; stopping condition $S$.

**Output:** Weighted sample $\{\theta_n, w_n\}_{n=1}^N$.

1: repeat
2: Increment $n \leftarrow n + 1$.
3: Generate $\theta_n \sim \pi(\cdot)$.
4: Simulate $\tilde{y}_n \sim \tilde{f}(\cdot \mid \theta_n)$.
5: Set $w_n = \mathbb{1}(\tilde{y}_n \in \Omega_\epsilon)$.
6: Generate $u_n \sim \text{Uniform}(0, 1)$.
7: if $u_n < \alpha(\tilde{y}_n, \theta_n)$ then
8: Simulate $y_n \sim f(\cdot \mid \tilde{y}_n, \theta_n)$.
9: Update $w_n \leftarrow w_n + [\mathbb{1}(y \in \Omega_\epsilon) - w_n] / \alpha(\tilde{y}_n, \theta_n)$.
10: end if
11: until $S = \text{true}$. 
• Generating data from the high-fidelity model, conditional on the output of the low-fidelity model.

• Drive down the rates of false positives and false negatives.

• Approach heavily dependent on the choice of low-fidelity model.
Example: common noise input stream

\[ \hat{X}(t + \tau) = \hat{X}(t) + r \hat{X}(t) \tau + \sigma \hat{X}(t) \sqrt{\tau} \xi, \quad \hat{X}(0) = 1, \quad \xi \sim \mathcal{N}(0, 1). \]

Use the same Brownian path for generation of the high- and low-fidelity simulations:

Uncoupled paths

Coupled paths
Example: common noise input stream

\[ \hat{X}(t + \tau) = \hat{X}(t) + r\hat{X}(t)\tau + \sigma\hat{X}(t)\sqrt{\tau}\xi, \quad \hat{X}(0) = 1, \quad \xi \sim \mathcal{N}(0, 1). \]

Use the same Brownian path for generation of the high- and low-fidelity simulations:
Can we combine these multifidelity ideas with other ideas for increasing the efficiency of ABC?
ABC sequential Monte Carlo

- ABC-SMC uses a sequence of importance distributions to gradually increase accuracy of the posterior.

- For a sequence of thresholds $\epsilon_1 > \epsilon_2 > \cdots > \epsilon_T$:
  - for $t = 1, \ldots, T - 1$:
    - generate $\left\{ w_i^{(t)}, \theta_i^{(t)} \right\}_{i=1}^{N_t}$ using importance distribution $\hat{q}_t$ and $\Omega(\epsilon_t)$;
    - define the next importance distribution, $\hat{q}_{t+1}(\theta)$, proportional to
      \[
      q_{t+1}(\theta) = \begin{cases} 
      \frac{\sum_{i=1}^{N_t} w_i^{(t)} K_t(\theta | \theta_i^{(t)})}{\sum_{m=1}^{N_t} w_m^{(t)}} & \pi(\theta) > 0 \\
      0 & \text{else;}
      \end{cases}
      \]
    - generate $\left\{ w_i^{(T)}, \theta_i^{(T)} \right\}_{i=1}^{N_T}$ using importance distribution $\hat{q}_T$ and $\Omega(\epsilon_T)$. 

Towards SMC: importance sampling

- First, need to integrate importance sampling into the multifidelity ABC framework:

**Algorithm 5** Multifidelity ABC importance sampling (MF-ABC-IS)

**Input:** Data $y_{\text{obs}}$ and neighbourhood $\Omega_e$; prior $\pi$; models $\tilde{f}(\cdot \mid \theta)$, $f(\cdot \mid \tilde{y}, \theta)$; continuation probability function $\alpha = \alpha(\tilde{y}, \theta)$; sample index $n = 0$; importance distribution $\hat{q}$ proportional to $\alpha(\tilde{y}, \theta)$, target probability $S$.

**Output**:

1: repeat
2: Input
3: Generate $u_n \sim \text{Uniform}(0, 1)$.
4: if $u_n < \alpha(\tilde{y}_n, \theta_n)$ then
5:     Simulate $y_n \sim f(\cdot \mid \tilde{y}_n, \theta_n)$.
6:     Update $w_n \leftarrow w_n + \left[ \mathbb{I}(y_n \in \Omega_e) - w_n \right] / \alpha(\tilde{y}_n, \theta_n)$.
7: end if
8: Update $w_n \leftarrow [\pi(\theta_n)/q(\theta_n)] w_n$.
9: until $S = \text{true}$.

For SMC: how do we sample from the importance distribution, given the weights that result from multifidelity ABC can be negative?
Towards multifidelity SMC-ABC

- Use defensive importance sampling, first defining a new (non-negative) importance distribution.

- Estimate continuation probabilities for each generation “on the fly”, using information from the previous generations.
MF ABC SMC in action

- Kuramoto oscillator network:
  \[ \dot{\phi}_i = \omega_i + \frac{K}{M} \sum_{j=1}^{M} \sin (\phi_j - \phi_i) \]

- Low-fidelity model - based on tracking Daido order parameters:
  \[ Z_n(t) = \frac{1}{M} \sum_{j=1}^{M} \exp(i n \phi_j) \]

  assume \( Z_n(t) = Z_1(t)^n \)

  to get
  \[ \dot{\tilde{R}} = \left( \frac{K}{2} - \gamma \right) \tilde{R} - \frac{K}{2} \tilde{R}^3 \] (magnitude)
  \[ \dot{\tilde{\Phi}} = \omega_0 \] (phase)
MF ABC SMC in action

- Typical simulation output:

Kuramot parameter: magnitude

Kuramot parameter: phase
MF ABC SMC in action

Efficiencies: ESS of last generation to total simulation time

Empirical posterior means of parameters

Ruth Baker
Towards multifidelity SMC-ABC

- Stopping criterion at each generation: ESS \geq 400.
Towards multifidelity SMC-ABC

Continuation probabilities by generation

- Probability of requiring high-fidelity model simulation given low-fidelity model

- Probability of requiring high-fidelity model simulation given low-fidelity model: close
Demonstrated that it is possible to incorporate both multifidelity and SMC approaches in generating the ABC posterior.

- Can choose the sequence of acceptance thresholds adaptively, e.g. to maintain efficiency across generations.

Many open questions remain, including:

- How best to design and estimation continuation probabilities?
- How to use multiple low-fidelity models?
- How to choose optimal perturbation kernels?
Acknowledgements

Oxford
Tom Prescott
Chris Lester
Casper Beentjes
Daniel Wilson
Andrew Parker
Rasa Giniunaite
Philip Maini
David Kay
Mike Giles

Further afield
Mat Simpson (QUT)
Alex Matsuaka (QUT)
Kit Yates (Bath)
Louise Dyson (Warwick)
Linus Schumacher (Edinburgh)
Paul Kulesa (Stowers Institute)
Rebecca McLennan (Stowers Institute)