Recent advances in Approximate Bayesian Computation

Jean-Michel Marin

Université de Montpellier
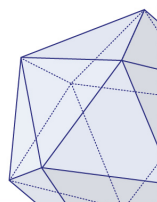
Institut Montpelliérain Alexander Grothendieck (IMAG)

Institut de Biologie Computationnelle (IBC)

Labex Numev

Likelihood-free rejection sampler

1) Set $i = 1$,

2) Generate $\boldsymbol{\theta}'$ from the prior distribution $\pi(\cdot)$,

3) Generate $\mathbf{z}$ from the likelihood $f(\cdot|\boldsymbol{\theta}')$,

4) If $\rho(\eta(\mathbf{z}), \eta(\mathbf{y})) \leq \epsilon$, set $\boldsymbol{\theta}_i = \boldsymbol{\theta}'$ and $i = i + 1$,

5) If $i \leq N$, return to 2).

ϵ reflects the tension between computability and accuracy:

- if $\epsilon \to \infty$, we get simulations from the prior,

- if $\epsilon \to 0$, we get simulations from the posterior.

$$\pi_\epsilon(\boldsymbol{\theta}|\mathbf{y}) = \frac{\int \pi(\boldsymbol{\theta}) f(\mathbf{z}|\boldsymbol{\theta}) \mathbb{I}_{A_{\epsilon,\mathbf{y}}}(\mathbf{z}) \mathrm{d}\mathbf{z}}{\int_{A_{\epsilon,\mathbf{y}} \times \Theta} \pi(\boldsymbol{\theta}) f(\mathbf{z}|\boldsymbol{\theta}) \mathrm{d}\mathbf{z} \mathrm{d}\boldsymbol{\theta}}$$

Curse of dimensionality

We have to summarize! If not the distance's values are too noisy.

Toy example: the simulated summary statistics $\eta(\mathbf{z}_1), \ldots, \eta(\mathbf{z}_N)$ and the observed one $\eta(\mathbf{y})$ are iid with uniform distribution on $[0,1]^d$

Let $d_\infty(d, N) = \mathbb{E}\left[\min_{i=1,\ldots,N} \left\|\eta(\mathbf{y}) - \eta(\mathbf{z}_i)\right\|_\infty\right]$

	$N = 100$	$N = 1,000$	$N = 10,000$	$N = 100,000$
$\delta_\infty(1, N)$	0.0025	0.00025	0.000025	0.0000025
$\delta_\infty(2, N)$	0.033	0.01	0.0033	0.001
$\delta_\infty(10, N)$	0.28	0.22	0.18	0.14
$\delta_\infty(200, N)$	0.48	0.48	0.47	0.46

Two views of the ABC approximation

$\Longrightarrow$ **Wilkinson (2013) SAGMB** shows that ABC is exact but for a different model to that intended.

$\Longrightarrow$ **Blum (2010) JASA** emphasizes that ABC is kernel smoothing approximation of the likelihood function

$$\pi_\epsilon(\boldsymbol{\theta}|\mathbf{y}) = \frac{\int \pi(\boldsymbol{\theta})f(\mathbf{z}|\boldsymbol{\theta})\mathbb{I}_{A_{\epsilon,\mathbf{y}}}(\mathbf{z})\mathrm{d}\mathbf{z}}{\int_{A_{\epsilon,\mathbf{y}}\times\Theta}\pi(\boldsymbol{\theta})f(\mathbf{z}|\boldsymbol{\theta})\mathrm{d}\mathbf{z}\mathrm{d}\boldsymbol{\theta}} = \frac{\pi(\boldsymbol{\theta})\int f(\mathbf{z}|\boldsymbol{\theta})K(\rho(\eta(\mathbf{z}),\eta(\mathbf{y})))\mathrm{d}\mathbf{z}}{\int \pi(\boldsymbol{\theta})f(\mathbf{z}|\boldsymbol{\theta})K(\rho(\eta(\mathbf{z}),\eta(\mathbf{y})))\mathrm{d}\mathbf{z}\mathrm{d}\boldsymbol{\theta}}\,.$$

These two interpretations allow for the possibility of more efficient ABC algorithms, generalized ABC schemes: **Wilkinson (2008) PhD Thesis** and **Fearnhead and Prangle (2012) JRSS B**.

Outline

A - Efficient algorithms

B - Regression adjustment

C - Summary statistics

D - Softwares

A - Efficient algorithms

Rejection sampling can perform poorly if the tolerance level ϵ is small.

Consequently various sequential Monte Carlo algorithms have been constructed as an alternative

Sisson et al. (2007) PNAS

Beaumont, Cornuet, Marin and Robert (2009) Biometrika

Del Moral et al. (2012) Statistics and Computing

Marin, Pudlo and Sedki (2012) IEEE Proceedings of WSC

Filippi et al. (2013) SAGMB

Sequential Monte Carlo samplers

The key idea is to decompose the difficult problem of sampling from $\pi_\epsilon(\boldsymbol{\theta}, \mathbf{z}|\mathbf{y})$ into a series of simpler subproblems.

The algorithm begins at time 0 sampling from $\pi_{\epsilon_0}(\boldsymbol{\theta}, \mathbf{z}|\mathbf{y})$ with large ϵ_0, then simulating from an increasing difficult sequence of target distribution $\pi_{\epsilon_t}(\boldsymbol{\theta}, \mathbf{z}|\mathbf{y})$, that is $\epsilon_t < \epsilon_{t-1}$.

Likelihood-free MCMC sampler

Likelihood-free MCMC sampler **Majoram et al. (2003) PNAS**

1) Use the likelihood free rejection sampling to get a realization $(\boldsymbol{\theta}^{(0)}, \mathbf{z}^{(0)})$ from the ABC target distribution $\pi_\epsilon(\boldsymbol{\theta}, \mathbf{z}|\mathbf{y})$,

2) Set $t = 1$,

3) Generate $\boldsymbol{\theta}'$ from the Markov kernel $q\left(\cdot|\boldsymbol{\theta}^{(t-1)}\right)$,

4) Generate $\mathbf{z}'$ from the likelihood $f(\cdot|\boldsymbol{\theta}')$,

5) Generate u from $\mathcal{U}_{[0,1]}$,

6) If $u \leq \frac{\pi(\boldsymbol{\theta}')q(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{\theta}')}{\pi(\boldsymbol{\theta}^{(t-1)}q(\boldsymbol{\theta}'|\boldsymbol{\theta}^{(t-1)}))}\mathbb{I}_{A_{\epsilon,\mathbf{y}}}(\mathbf{z}')$,
set $(\boldsymbol{\theta}^{(t)}, \mathbf{z}^{(t)}) = (\boldsymbol{\theta}', \mathbf{z}')$ else $(\boldsymbol{\theta}^{(t)}, \mathbf{z}^{(t)})) = (\boldsymbol{\theta}^{(t-1)}, \mathbf{z}^{(t-1)})$,

7) Set $t = t + 1$,

8) If $t \leq N$ return to **3)**.

The acceptance probability does not involve the calculation of the likelihood. Indeed,

$$\frac{\pi_\epsilon(\boldsymbol{\theta}', \mathbf{z}'|\mathbf{y})}{\pi_\epsilon(\boldsymbol{\theta}^{(t-1)}, \mathbf{z}^{(t-1)}|\mathbf{y})} \times \frac{q(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{\theta}')f(\mathbf{z}^{(t-1)}|\boldsymbol{\theta}^{(t-1)})}{q(\boldsymbol{\theta}'|\boldsymbol{\theta}^{(t-1)})f(\mathbf{z}'|\boldsymbol{\theta}')},$$

$$= \frac{\pi(\boldsymbol{\theta}')\,\cancel{f(\mathbf{z}'|\boldsymbol{\theta}')}\,\mathbb{I}_{A_{\epsilon,\mathbf{y}}}(\mathbf{z}')}{\pi(\boldsymbol{\theta}^{(t-1)})\,\cancel{f(\mathbf{z}^{(t-1)}|\boldsymbol{\theta}^{(t-1)})}\,\cancel{\mathbb{I}_{A_{\epsilon,\mathbf{y}}}(\mathbf{z}^{(t-1)})}} \times \frac{q(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{\theta}')\,\cancel{f(\mathbf{z}^{(t-1)}|\boldsymbol{\theta}^{(t-1)})}}{q(\boldsymbol{\theta}'|\boldsymbol{\theta}^{(t-1)})\,\cancel{f(\mathbf{z}'|\boldsymbol{\theta}')}},$$

$$= \frac{\pi(\boldsymbol{\theta}')q(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{\theta}')}{\pi(\boldsymbol{\theta}^{(t-1)}q(\boldsymbol{\theta}'|\boldsymbol{\theta}^{(t-1)})}\mathbb{I}_{A_{\epsilon,\mathbf{y}}}(\mathbf{z}')\,.$$

For more details and method in the field of MCMC without likelihood, one can see the review paper of **Sisson and Fan (2010) Handbook of Markov Chain Monte Carlo**.

B - Regression adjustments

An alternative to the likelihood-free rejection sampler, proposed by **Beaumont et al. (2002) Genetics**, uses post-hoc adjustment of the parameter values to try to weaken the effect of the discrepancy between accepted $\eta(\mathbf{z}_i)$ and $\eta(\mathbf{y})$.

Denote $\mathbb{E}(\theta|\eta(\mathbf{y})) = m(\eta(\mathbf{y}))$, use a quite large tolerance and, for the accepted simulations, suppose that

$$\theta_i = m(\eta(\mathbf{z}_i)) + e_i\,,$$

where e_i is a centered noise with variance that does not depend on i.

The methodology

1) Estimate $m(\cdot)$ by $\hat{m}(\cdot)$ using a local linear regression model;

2) Calculate the residuals $\hat{e}_i = \theta_i - \hat{m}(\eta(\mathbf{z}_i))$;

3) Deduce the adjusted values

$$\theta_i^* = \hat{m}(\eta(\mathbf{y})) + \hat{e}_i = \theta_i + (\hat{m}(\eta(\mathbf{y})) - \hat{m}(\eta(\mathbf{z}_i))) \ .$$

This method allows us to use a larger tolerance values and can substantially improve posterior accuracy with less computation.

Blum and Francois (2010) Statistics and Computing proposed a nonlinear heteroscedastic model and used feed-forward neural networks for both the conditional mean and variance.

C - Summary statistics

Suppose we are given a candidate set $\eta(\mathbf{y}) = (\eta_1(\mathbf{y}), \ldots, \eta_p(\mathbf{y}))$ of summaries from which to choose, different strategies:

- best subset selection

- projection

- regularisation techniques

Best subset selection

Introduce a criterion and use a search algorithm (for instance forward selection)

- **Joyce and Marjoram (2008) SAGMB**, τ-sufficiency
- **Nunes and Balding (2010) SAGMB**, entropy

Projection

Fearnhead and Prangle (2012) JRSS B introduce
semi-automatic ABC:

(i) use a pilot run of ABC to determine a region of non-negligible posterior mass;

(ii) simulate sets of parameter values and data;

(iii) use the simulated sets of parameter values and data to estimate the summary statistics; and

(iv) run ABC with this choice of summary statistics.

Blum and François (2010) use feed-forward neural network
regression models.

Regularisation techniques

Blum, Nunes, Prangle and Fearnhead (2013) Statistical Science uses ridge regression.

D - Sofwares

abc R package several ABC algorithms for performing parameter estimation and model selection

http://onlinelibrary.wiley.com/doi/10.1111/j.2041-210X.2011.00179.x/epdf

abctools R package tuning ABC analyses

https://journal.r-project.org/archive/2015-2/nunes-prangle.pdf

abcrf R package ABC via random forests

EasyABC R package several algorithms for performing efficient ABC sampling schemes, including 4 sequential sampling schemes and 3 MCMC schemes

DIY-ABC software performs parameter estimation and model selection for population genetics models

ABC-SysBio python package parameter inference and model selection for dynamical systems

ABCtoolbox programs various ABC algorithms including rejection sampling, MCMC without likelihood, a particle-based sampler, and ABC-GLM

msBayes software package consisting of several C and R programs

PopABC software package for inference of the pattern of demographic divergence, coalescent simulation, bayesian model choice