

Approximate Bayesian Computation

Davi Barreira

FGV - Escola de Matemática Aplicada

Table of contents

1. Objective & Motivation
2. Original ABC Algorithm

The objective of this presentation is to give an overview of the Approximate Bayesian Computation (ABC) algorithm through the replication of the paper **Approximate Bayesian computational methods** by Marin et al. (2012).

The paper talks about different variants of ABC by estimating the posterior of Moving Average models.

ABC methods are known as likelihood-free techniques, thus are a useful approach in problems that the likelihood is intractable, e.g., likelihood not available in closed form, or likelihood too expensive to calculate.

- Coalecent models in population genetics (Tavaré et al., 1997);
- Species dynamics (Jabot and Lohier, 2016);
- Real-world model of HIV transmission (McKinley et al., 2018).

Objective & Motivation

In some settings where we have latent variables, the likelihood is expressed as:

$$\ell(\boldsymbol{\theta} \mid \mathbf{y}) = \int \ell^*(\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{u}) d\mathbf{u}$$

Hence, \mathbf{y} is observed and \mathbf{u} is latent and $\boldsymbol{\theta}$ is the parameter of interest.

Original ABC Algorithm

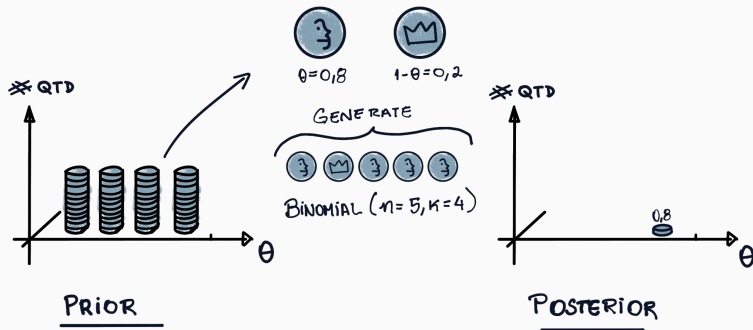
Rubin (1984) described the ABC algorithm as a thought experiment to explain how to sample from a posterior distribution. Tavaré et al. (1997) is usually considered the paper responsible for the proposing ABC for inferring the posterior distribution.

Algorithm 1: Original ABC method

```
for  $i=1$  to  $N$  do  
  repeat  
    Sample  $\theta' \sim \pi(\cdot)$   
    Generate  $\mathbf{z} \sim p(\cdot \mid \theta')$   
  until  $\mathbf{y} = \mathbf{z};$   
end
```

Original ABC Algorithm

Below we have an schematic drawing with an example of the ABC method for Beta/Binomial model.



The proof that the algorithm indeed results in an iid sample from the posterior is shown below. Let \mathbf{y} be the observed, $\boldsymbol{\theta}$ the parameter of interest and \mathbf{z} the generated samples.

$$p(\boldsymbol{\theta}_i) \propto \sum_{\mathbf{z} \in \mathbb{D}} \pi(\boldsymbol{\theta}_i) p(\mathbf{z} \mid \boldsymbol{\theta}_i) \mathbb{I}_{\mathbf{y}}(\mathbf{z}) = \pi(\boldsymbol{\theta}_i) p(\mathbf{y} \mid \boldsymbol{\theta}_i) \propto \pi(\boldsymbol{\theta}_i \mid \mathbf{y})$$

Original ABC Algorithm

Pritchard et al. (1999) extended the original algorithm to the case of continuous sample spaces.

Algorithm 2: ABC method for discrete and continuous distributions

```
for  $i=1$  to  $N$  do  
  repeat  
    Sample  $\theta' \sim \pi(\cdot)$   
    Generate  $\mathbf{z} \sim p(\cdot \mid \theta')$   
  until  $\rho[\eta(\mathbf{y}), \eta(\mathbf{z})] \leq \epsilon$ ;  
end
```

- η : function defining a statistic (e.g. the mean),
- ρ : a distance function,
- ϵ : acceptance tolerance.

Original ABC Algorithm

For this ABC algorithm, instead of the actual posterior, we get

$$\pi_{\epsilon}(\boldsymbol{\theta}, \mathbf{z} \mid \mathbf{y}) = \frac{\pi(\boldsymbol{\theta})p(\mathbf{z} \mid \boldsymbol{\theta})\mathbb{I}_{A_{\epsilon, \mathbf{y}}}(\mathbf{z})}{\int_{A_{\epsilon, \mathbf{y}} \times \boldsymbol{\theta}} \pi(\boldsymbol{\theta})p(\mathbf{z} \mid \boldsymbol{\theta})d\mathbf{z}d\boldsymbol{\theta}}$$

Where, $A_{\epsilon, \mathbf{y}} = \{\mathbf{z} \in \mathbb{D} \mid \rho[\eta(\mathbf{z}), \eta(\mathbf{y})] \leq \epsilon\}.$

Hence, for a tolerance (ϵ) "small enough", we expect a good approximation.

$$\pi_{\epsilon}(\boldsymbol{\theta} \mid \mathbf{y}) = \int \pi_{\epsilon}(\boldsymbol{\theta}, \mathbf{z} \mid \mathbf{y})d\mathbf{z} \approx \pi(\boldsymbol{\theta} \mid \mathbf{y})$$

- Jabot, F. and Lohier, T. (2016). Non-random correlation of species dynamics in tropical tree communities. *Oikos*, 125(12):1733–1742.
- Marin, J.-M., Pudlo, P., Robert, C. P., and Ryder, R. J. (2012). Approximate bayesian computational methods. *Statistics and Computing*, 22(6):1167–1180.
- McKinley, T. J., Vernon, I., Andrianakis, I., McCreesh, N., Oakley, J. E., Nsubuga, R. N., Goldstein, M., and White, R. G. (2018). Approximate bayesian computation and simulation-based inference for complex stochastic epidemic models. *Statist. Sci.*, 33(1):4–18.
- Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A., and Feldman, M. W. (1999). Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution*, 16(12):1791–1798.

- Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann. Statist.*, 12(4):1151–1172.
- Tavaré, S., Balding, D. J., Griffiths, R. C., and Donnelly, P. (1997). Inferring coalescence times from dna sequence data. *Genetics*, 145(2):505–518.