Approximate Bayesian Computation

Davi Barreira

FGV - Escola de Matemtica Aplicada

Table of contents

Objective & Motivation

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.

Objective & Motivation

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:

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

Hence, $\boldsymbol{y}$ is observed and $\boldsymbol{u}$ 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 infering the posterior distribution.

Algorithm 1: Original ABC method

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

Original ABC Algorithm

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

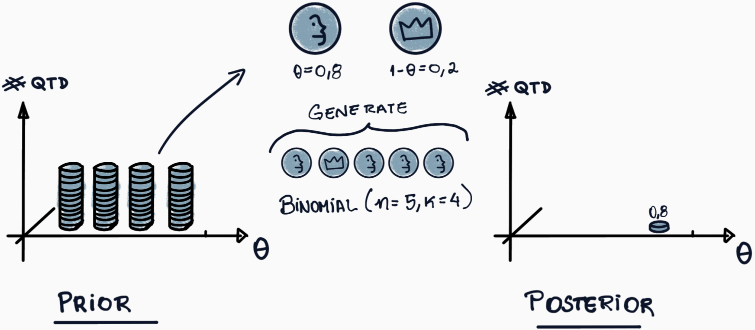

Figure 1: Schematic drawing of ABC method for Beta/Binomial model

Original ABC Algorithm

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

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

Original ABC Algorithm

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

Algorithm 2: ABC method for discrete and continuous distributions

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

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

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

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

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

Moving Average

We will use the Moving Average model, also denoted as MA(q), for assessing the performance of the ABC methods. The MA(q) process is a stochastic process defined by:

$$y_k = u_k + \sum_{i=1}^{q} \theta_i u_{k-i}$$

Where $(u_k)_{k \in \mathbb{Z}} \overset{iid}{\sim} N(0,1)$. For a $q = 2$, imposing the standard identifiability condition we obtain the following conditions:

$$-2 < \theta_1 < 2, \qquad \theta_1 + \theta_2 > -1, \qquad \theta_1 - \theta_2 < 1.$$

Hence, we use an uniform distribution over this triangular region as prior for $\boldsymbol{\theta}$. The likelihood of $\boldsymbol{y} \mid \boldsymbol{\theta}$ is more complex because of the need to integrate $\boldsymbol{u}$.

Moving Average

We generate a synthetic sample of length 100 using $(\theta_1, \theta_2) = (0.6, 0.2)$. For $q = 2$ we can also numerically calculate the real posterior and the marginal distributions.

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{y}) \propto \pi(\boldsymbol{\theta}) p(\boldsymbol{y} \mid \boldsymbol{\theta}), \qquad \boldsymbol{y} \mid \boldsymbol{\theta} \sim MVN(0, \Sigma)$$

$$\Sigma = \begin{bmatrix} 1+\theta_1^2+\theta_2^2 & \theta_1+\theta_2\theta_1 & \theta_2 & 0 & 0 & 0 & \dots & 0 \\ \theta_1+\theta_2\theta_1 & 1+\theta_1^2+\theta_2^2 & \theta_1+\theta_2\theta_1 & \theta_2 & 0 & 0 & \dots & 0 \\ \theta_2 & \theta_1+\theta_2\theta_1 & 1+\theta_1^2+\theta_2^2 & \theta_1+\theta_2\theta_1 & \theta_2 & 0 & \dots & 0 \\ 0 & \theta_2 & \theta_1+\theta_2\theta_1 & 1+\theta_1^2+\theta_2^2 & \theta_1+\theta_2\theta_1 & \theta_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \theta_2 & \theta_1+\theta_1\theta_2 & 1+\theta_1^2+\theta_2^2 \end{bmatrix}$$

Moving Average

For this model, the ABC algorithm consists of:

- Sample $\boldsymbol{\theta}^*$ from the uniform triangular prior using rejection sampling;
- For each $k \in \{-1, 0, 1, ..., 100\}$, sample $u_k \overset{iid}{\sim} N(0, 1)$.
- For each $k \in \{1, 2, ..., 100\}$, calculate $z_k = u_k + \sum_{i=1}^{2} \theta_i^* u_{k-i}$.

Two distance metrics are used. The raw distance between the series

$$\rho^2\{\boldsymbol{z}, \boldsymbol{y}\} = \sum_{k=1}^{n=100} (y_k - z_k)^2$$

And the sum of the quadratic distances between the first $q = 2$ autocovariances

$$\tau_j(\boldsymbol{x}) = \sum_{k=j+1}^{n=100} x_k x_{k-j}, \qquad \rho^2 = \sum_{j=0}^{q=2} (\tau_j(\boldsymbol{y}) - \tau_j(\boldsymbol{z}))^2$$

Moving Average

Below we present the results of running ABC for the MA(2) process.

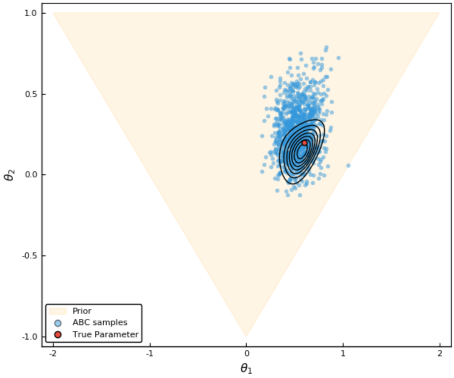

Figure 2: Comparison between the true posterior (*line in black*), with the samples produced using the ABC . The number of simulations is $N = 10^6$, and the threshold ϵ corresponds to the quantile of accepting 0.1%. The ρ used was the distance of the autocovariances.

Calibration of ABC

Summary Statistics(η). As the number of observations grow, using the raw distance between each observation becomes too prohibitive. The alternative is to try using summary statistics, if possible, sufficient statistics.

Fearnhead and Prangle (2010) created a way of constructing appropriate summary statistics for ABC in a semi-automatic manner.

Tolerance threshold. The standard practice is to use ϵ as a quantile of the simulated distances.

Calibration of ABC

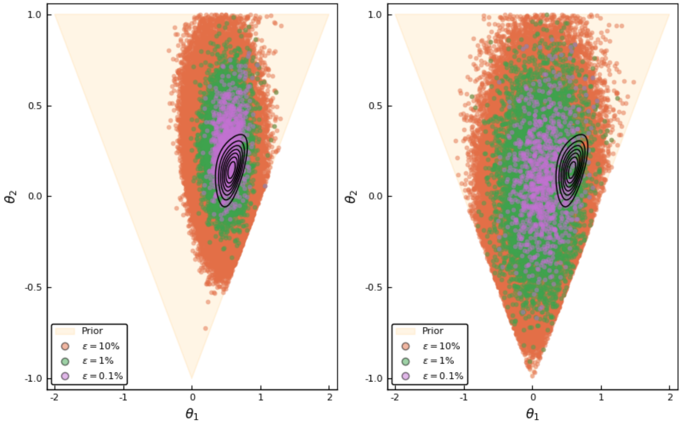

Figure 3: Comparison of ABC method when using autocovariance distance [1] (*left*) versus raw distance (*right*). The number of simulations is $N = 10^6$ and different thresholds ϵ are used.

[1] in the rest of the slides we will only use the autocovariance distances

Calibration of ABC

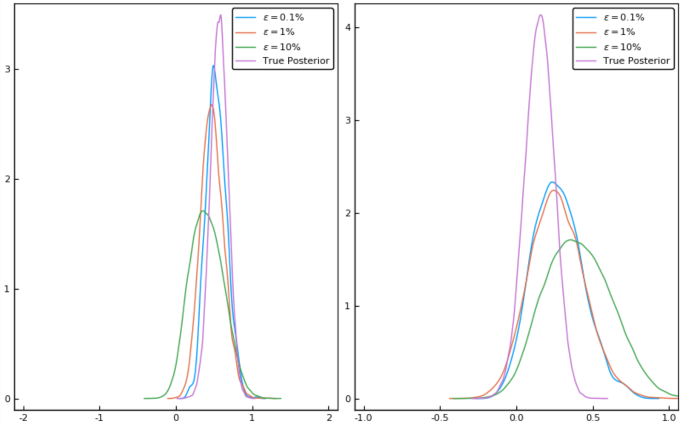

Figure 4: Comparison of ABC samples with the true posterior marginal distribution for θ_1 (*left*) and θ_2 (*right*).

Using non-informative priors is usually very inefficient, because it leads to lots of rejections. To tackle this problem, Marjoram et al. (2003) came up with MCMC-ABC.

Algorithm 3: MCMC-ABC

Use Algorithm 2 to get $(\boldsymbol{\theta}^{(0)}, \boldsymbol{z}^{(0)}$ from the target $\pi_\epsilon(\boldsymbol{\theta}, \boldsymbol{z} \mid \boldsymbol{y})$.
for *i=1 to N* **do**
  repeat
    Sample $\boldsymbol{\theta}'$ from the Markov kernel $q(\cdot \mid \theta^{(i-1)})$
    Generate $\boldsymbol{z} \sim p(\cdot \mid \boldsymbol{\theta}')$
    Sample $u \sim U[0,1]$
    if $u \leq \frac{\pi(\boldsymbol{\theta}')q(\boldsymbol{\theta}^{(i-1)}|)}{\pi(\boldsymbol{\theta}^{(i-1)})q(\boldsymbol{\theta}^{(i-1)}|)}$ *and* $\rho\{\eta(\boldsymbol{z}'), \eta(\boldsymbol{y})\} \leq \epsilon$ **then**
      Set $(\boldsymbol{\theta}^{(i)}, \boldsymbol{z}^{(i)}) = (\boldsymbol{\theta}', \boldsymbol{z}')$
    end
    else
      Set $(\boldsymbol{\theta}^{(i)}, \boldsymbol{z}^{(i)}) = (\boldsymbol{\theta}^{(i-1)}, \boldsymbol{z}^{(i-1)})$
    end
  until $\rho[\eta(\boldsymbol{y}), \eta(\boldsymbol{z})] \leq \epsilon$;
end

ABC Variations

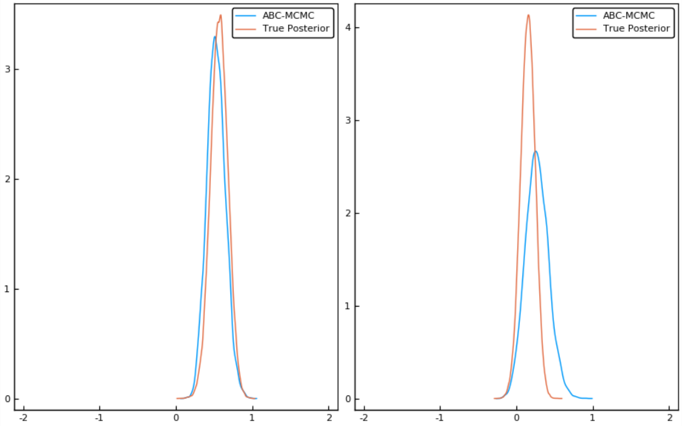

Figure 5: Comparison of ABC-MCMC samples with the true posterior marginal distribution for θ_1 (*left*) and θ_2 (*right*) using $\epsilon = 0.1\%$.

ABC Variations

Another variation of ABC is called *Noisy* ABC, that was proposed by Wilkinson (2013). The original ABC algorithm can be thought as a rejection algorithm using a uniform kernel ($\mathbb{I}_{A_{\epsilon,\boldsymbol{y}}(\boldsymbol{z})}$). The *Noisy* version generalizes this, allowing the use of different kernels, hence:

$$\pi_\epsilon(\boldsymbol{\theta}, \boldsymbol{z} \mid \boldsymbol{y}) = \frac{\pi(\boldsymbol{\theta}) p(\boldsymbol{z} \mid \boldsymbol{\theta}) K_\epsilon(\boldsymbol{y} - \boldsymbol{z})}{\int \pi(\boldsymbol{\theta}) p(\boldsymbol{z} \mid \boldsymbol{\theta}) K_\epsilon(\boldsymbol{y} - \boldsymbol{z}) d\boldsymbol{z} d\boldsymbol{\theta}}$$

Now, instead of accepting if $\rho\{\eta(\boldsymbol{y}), \eta(\boldsymbol{z})\} \leq \epsilon$, we accept with probability $\frac{K_\epsilon(\boldsymbol{y}-\boldsymbol{z})}{max\{K_\epsilon(\boldsymbol{y}-\boldsymbol{z}\})}$.

ABC Variations

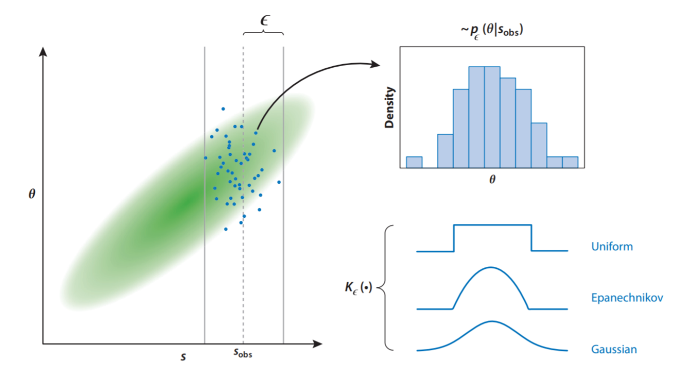

Figure 6: Illustration of *Noisy* ABC rejection kernels, where s is the statistic from the ABC sampler and s_{obs} is the observed value from the data. Figure from Beaumont (2019).

ABC Variations

Sequential techniques are also used with ABC to enhance the efficiency of the algorithms. A popular method in this regard is the ABC-PMC (ABC population Monte Carlo) by Beaumont et al. (2009). It estimates the scale of the random walk step from the previous simulations and uses a sequence of tolerance thresholds ($\epsilon_1 \geq ... \geq \epsilon_T$) to approximate the distribution.

A recent work by Simola et al. (2019) propose a method for adaptively selecting this sequence of tolerances that improves computationla efficiency and defines a stopping rule, thus assisting in automating the termination of the sampling procedure.

ABC Variations

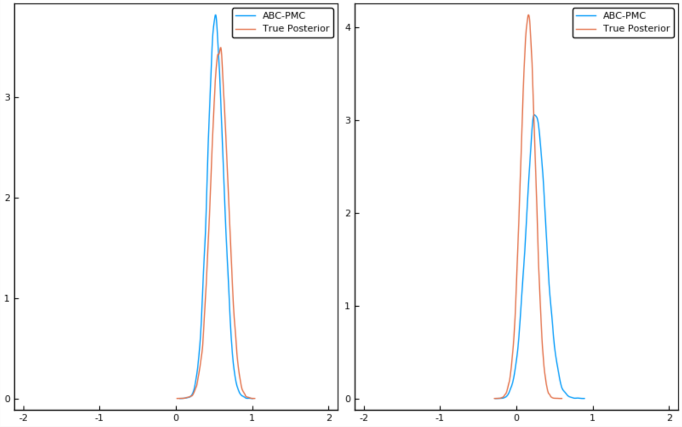

Figure 7: Comparison of ABC-PMC samples with the true posterior marginal distribution for θ_1 (*left*) and θ_2 (*right*) using $\epsilon = 0.1\%$.

Post-processing of ABC

Local linear regression was proposed by Beaumont et al. (2002) as a way improve the result of simulations without the need to use restrictively low threshold values. The idea is then to use a weighted least squares regression of $\boldsymbol{\theta}$ on $(\eta(\boldsymbol{y}) - \eta(\boldsymbol{z}))$, with weights according to a chosen kernel.

$$\boldsymbol{\theta}^* = \boldsymbol{\theta} - (\eta(\boldsymbol{y}) - \eta(\boldsymbol{z}))^T \hat{\beta}$$

Post-processing of ABC

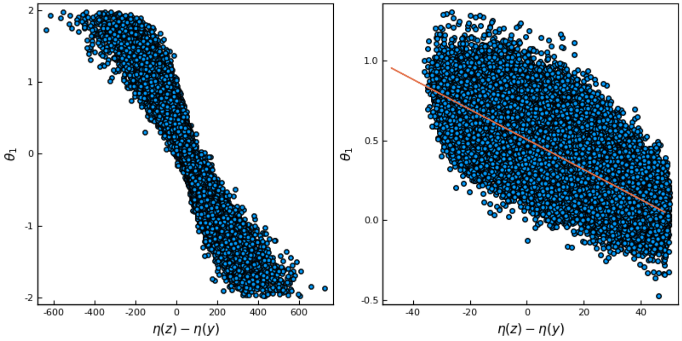

Figure 8: Scatter plots of simulated θ_1 and $(\eta(\boldsymbol{y}) - \eta(\boldsymbol{z}))$ for autocovariance with *lag* $= 1$. On the *left* there are all the $N = 10^6$ simulations, while in the *right* only the accepted samples for $\epsilon = 10\%$ with the regression line.

Post-processing of ABC

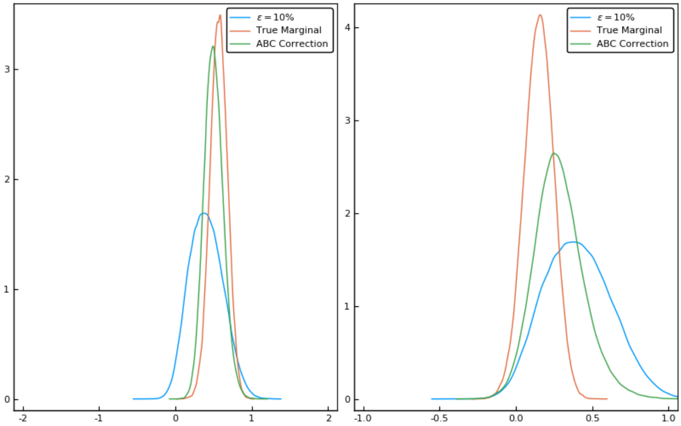

Figure 9: Comparison of ABC samples corrected through local linear regression versus the true marginal posterior distribution for θ_1 (*left*) and θ_2 (*right*) using $\epsilon = 10\%$.

Model Choice

The estimation of the posterior using ABC is straightfoward.

Algorithm 4: ABC Model Choice

for *i=1 to N* **do**
  repeat
    Sample $m \sim \pi(\mathcal{M} = m)$
    Sample $\boldsymbol{\theta}_m \sim \pi_m(\boldsymbol{\theta}_m)$
    Generate $\boldsymbol{z} \sim p_m(\cdot \mid \boldsymbol{\theta}_m)$
  until $\rho[\eta(\boldsymbol{y}), \eta(\boldsymbol{z})] \leq \epsilon$;
  Set $m^{(i)} = m$ and $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}_m$
end

The posterior probability $\pi(\mathcal{M} = m \mid \boldsymbol{y})$ if the acceptance frequency from model m,

$$\frac{1}{N} \sum_{i=1}^{N} \mathbb{I}_{m^{(i)} = m}$$

Post-processing of ABC

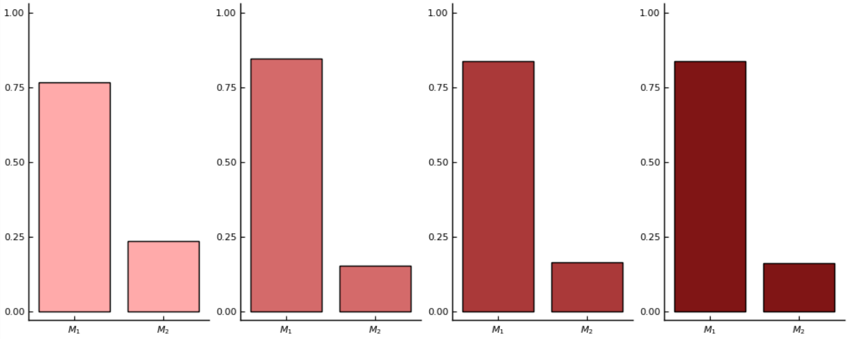

Figure 10: Barplots of evolution of Bayes factor approximations in terms of visits to models MA(1) and MA(2) using ABC using thresholds of 10, 1, 0.1 , 0.01 % on autocovariance distance. The true model is MA(2) and the true Bayes factor is 0.952.

Post-processing of ABC

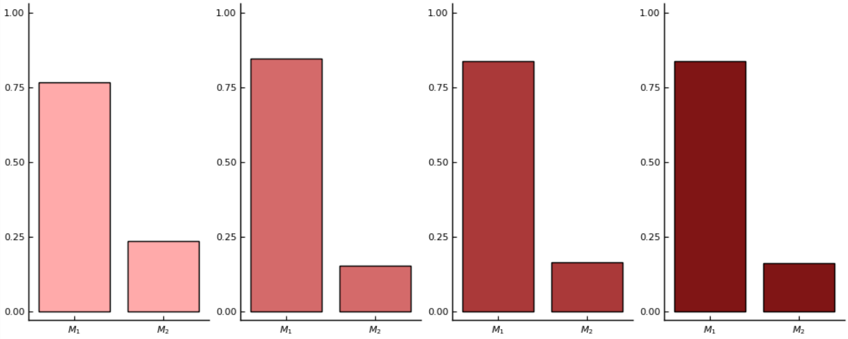

Figure 11: Barplots of evolution of Bayes factor approximations in terms of visits to models MA(1) and MA(2) using ABC using thresholds of 10, 1, 0.1 , 0.01 % on autocovariance distance. The true model is MA(1) and the true Bayes factor is 0.943.

References i

Beaumont, M., Zhang, W., and Balding, D. (2002). Approximate bayesian computation in population genetics. *Genetics*, 162(4):2025–2035. ID number: ISI:000180502300043.

Beaumont, M. A. (2019). Approximate bayesian computation. *Annual Review of Statistics and Its Application*, 6(1):379–403.

Beaumont, M. A., Cornuet, J.-M., Marin, J.-M., and Robert, C. P. (2009). Adaptive approximate bayesian computation. *Biometrika*, 96(4):983990.

Fearnhead, P. and Prangle, D. (2010). Constructing summary statistics for approximate bayesian computation: Semi-automatic abc.

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.

Marjoram, P., Molitor, J., Plagnol, V., and Tavaré, S. (2003). Markov chain monte carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 100(26):15324–15328.

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.

References iii

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann. Statist.*, 12(4):1151–1172.

Simola, U., Cisewski-Kehe, J., Gutmann, M. U., and Corander, J. (2019). Adaptive approximate bayesian computation tolerance selection.

Tavaré, S., Balding, D. J., Griffiths, R. C., and Donnelly, P. (1997). Inferring coalescence times from dna sequence data. *Genetics*, 145(2):505–518.

Wilkinson, R. D. (2013). Approximate bayesian computation (abc) gives exact results under the assumption of model error. *Statistical Applications in Genetics and Molecular Biology*, 12(2).