

Bayesian Modelling of Epidemic Events

Summary Statistic Selection for Approximate Bayesian Computation Methods

Student: Dominic Hutchinson, Supervisors: Dr. Daniel Lawson & Dr. Sam Tickle

University of Bristol, Department of Mathematics, Year: 2021, Credits: 40, Level: 7/M

1. Motivation

Many Bayesian inference questions for a model X with parameters θ are centred around calculating the posterior of the parameters given a realisation of the data. That is, the probability distribution for the value of the parameters, given a particular realisation of the mathematical model. Bayes’ Rule states the posterior as $(\theta|X)$ (1).

$$\mathbb{P}(\theta|X) = \frac{\mathbb{P}(X|\theta)\mathbb{P}(\theta)}{\mathbb{P}(X)} \propto \mathbb{P}(X|\theta)\mathbb{P}(\theta) \tag{1}$$

The likelihood function $\mathbb{P}(X|\theta)$ is the probability of observing a certain set of data under a certain configuration of the model. Direct computation of (1) is intractable in real world scenarios as the likelihood function is intractable. This motivates the need for methods which can estimate the parameter posteriors in scenarios where the likelihood function is unavailable.

2. Approximate Bayesian Computation

Approximate Bayesian Computation (ABC) methods, first presented in (Tavaré u.a., 1997), are a family of computational methods which can be used to approximate posteriors for the parameters of model where the likelihood is intractable. This is achieved by generating simulations from joint distribution (2) which can be used as simulations from the likelihood, rather than evaluating it explicitly.

$$\pi_{ABC}(\theta, s|s_{obs}) = K_{\varepsilon}(\|s - s_{obs}\|)\mathbb{P}(s|\theta)\pi_0(\theta) \tag{2}$$

where K_{ε} is an acceptance kernel with bandwidth ε , π_0 are the priors for the parameters, $s := s(x)$ is summary statistic values of simulated data x and $s_{obs} := s(x_{obs})$ are the summary statistic values of the observed data x_{obs} . These are all parameters of the ABC methods.

There are several flavours of ABC methods, namely: ABC-Rejection Sampling (Tavaré u.a., 1997); ABC-MCMC (Marjoram u.a., 2003); and, ABC-SMC (Sisson u.a., 2007). A general schema for ABC methods is provided in **Algorithm 1**.

Algorithm 1 - General ABC Schema

Require: Observed values x_{obs} ; Summary statistics $s(\cdot)$; Priors $\pi_0(\cdot)$; Theorised model $f(X|\cdot)$; Acceptance Kernel $K_{\varepsilon}(\cdot)$; Distance Measure $\|\cdot\|$.

- 1 Calculate summary statistic values $s_{obs} = s(x_{obs})$.
- 2 Until stopping condition reached:
 - I Sample a set of parameters $\tilde{\theta}$.
 - II Run the theorised model with sampled parameter $\tilde{x} = f_{\tilde{\theta}}(X|\tilde{\theta})$.
 - III Calculate summary statistic values $\tilde{s} = s(\tilde{x})$.
 - IV Accepted parameters $\tilde{\theta}$ with probability $K_{\varepsilon}(\|\tilde{s} - s_{obs}\|)$.
- 3 Return all accepted parameter sets $\hat{\Theta}$.

3. Summary Statistics

The success of ABC methods depends mainly on three user choices: summary statistics; distance measure; and, acceptance kernel. Of these, summary statistic selection is arguably the most important as the other two mainly effect the method’s rate of convergence. Whereas, choosing summary statistics which are uninformative while mean the method cannot converge on the true posterior.

$$\mathbb{P}(X|s(X)) = \mathbb{P}(X|s(X), \theta) \tag{3}$$

A statistic $s(\cdot)$ of a model X with model parameters θ is sufficient for parameters θ if it captures all the information which a sample of the model carries about said parameter (3). The identity function is a sufficient statistic for all models, but has high dimensions. In practice low-dimensional sufficient statistics are rare, and do not exist for non-Exponential family models (Pitman, 1936).

The ideal summary statistic is sufficient, so it retains maximum information, and of low-dimensions, so it is efficient to process and to avoid the “Curse of Dimensionality” which makes distinguishing between pairs of high-dimensional vectors near-impossible.

4. Minimising Entropy Approach

(Nunes u.a., 2010) present two algorithms which seek to select an optimal subset of summary statistics for a large set of hand-crafted summary statistics. The first, given in **Algorithm 2**, estimates the entropy of every subset of statistics using the k^{th} -nearest neighbour estimate of entropy. The subset with the lowest estimated entropy is returned as the recommended set of statistics as lower entropy values implies the distribution holds a greater amount of information.

The second algorithm extends the first by using the set of statistics returned by **Algorithm 2** to then generate a set of parameters $\hat{\Theta}_{ME}$ which are assumed to come from the true model. Each subset of statistics is evaluated again, this time calculating the MRSSE between the set of parameters accepted a given subset Θ_{acc} and $\hat{\Theta}_{ME}$.

Algorithm 2 - Minimum Entropy Summary Statistic Selection

Adapted from (Nunes u.a., 2010).

require: Observations values x_{obs} ; Set of Summary Statistics S ; Number of Simulations m ; Priors $\pi_0(\cdot)$; Theorised model $f(X|\cdot)$; Mapping $f(\cdot)$; Distance Measure $\|\cdot\|$.

for $S' \in 2^S$ **do**

$\Theta \leftarrow$ Parameter sets accepted from ABC-Rejection Sampling using S' .

$\hat{H}_{S'} \leftarrow \hat{H}(\Theta)$

$S_{ME}^* \leftarrow \operatorname{argmin}_{S' \in 2^S} \hat{H}_{S'}$

return S_{ME}^*

5. Semi-Automatic ABC

(Fearnhead u.a., 2011) presents an algorithm which removes the need for the user to specify a set of hand-crafted summary statistics and rather using regression to generate its own summary statistics. A pilot run of ABC-Rejection Sampling is run to acquire a set of accepted parameters $\hat{\Theta}$ and their associated observations $X_{\hat{\Theta}}$. Regression is then performed between these two sets of values to generate summary statistics. **Algorithm 3** provides a specification of this algorithm which uses linear regression.

Algorithm 3 - Semi-Automatic ABC

Adapted from (Fearnhead u.a., 2011).

require: Observations values x_{obs} ; Set of Summary Statistics S ; Number of Simulations m ; Priors $\pi_0(\cdot)$; Theorised model $f(X|\cdot)$; Mapping $f(\cdot)$; Distance Measure $\|\cdot\|$.

$f_{\theta} \leftarrow$ Posterior from pilot run of an ABC-method using x_{obs} and S

$\hat{\Theta} \leftarrow m$ simulations from f_{θ}

$X_{\hat{\theta}} \leftarrow X \left(\hat{\theta} \right)$ for each $\hat{\theta} \in \hat{\Theta}$; $\hat{X} \leftarrow \{X_{\hat{\theta}_1}, \dots, X_{\hat{\theta}_m}\}$

$F \leftarrow f(\hat{X})$; $\tilde{F} \leftarrow F$ with a preceding column of 1s

for $i = 1, \dots, \rho$ **do**

$A_i \leftarrow i^{th}$ element of each set in $\hat{\Theta}$

$(\alpha^{(i)}, \beta^{(i)}) \leftarrow (\tilde{F}^T \tilde{F}^{-1}) \tilde{F}^T A_i$

$s_i(x) := \beta^{(i)} x$

return $\{s_1, \dots, s_{\rho}\}$

6. Results

| Algorithm | Statistics | Dimensions | ABC-SMC MSE |
|--------------------|--|------------|-------------|
| Control | Identity Function | 90 | 121,777 |
| Joyce-Marjoram | [Final Susceptible Population] | 1 | 101,730,336 |
| Minimum Entropy | [Mean Infectious Population, Mean Removed Population] | 2 | 1,131,712 |
| 2-Step ME | [Peak Infectious Population Size, Mean Infectious Population, Mean Removed Population] | 3 | 228,150 |
| Semi-Automatic ABC | N/A | 2 | 643,255 |

Table: Mean Square Error of models fitted using estimated posterior means from Adaptive ABC-SMC running for 5,000 simulations and target acceptance rate $\alpha = 0.9$ when using the recommended summary statistics from each algorithm on a standard SIR model over 30 time-periods with population size 100,000 and parameters $\beta = 1$ & $\gamma = 1$.

None of the selection algorithms produce summary statistics which perform as well as the identity function. However, 2-Step Minimum Entropy and Semi-Automatic ABC produce statistics which are not significantly worse despite having 3 or 2 dimensions, respectively.

References

Pitman, E. J. G.(1936): *Sufficient statistics and intrinsic accuracy*567-579.

Tavaré, Simon / Balding, David J. / Griffiths, R. C. / Donnelly, Peter(1997): *Inferring Coalescence Times From DNA Sequence Data*, 2: 505–518.

Marjoram, Paul / Molitor, John / Plagnol, Vincent / Tavaré, Simon(2003): *Markov chain Monte Carlo without likelihoods*, 26: 15324–15328.

Sisson, S. A. / Fan, Y. / Tanaka, Mark M.(2007): *Sequential Monte Carlo without likelihoods*, 6: 1760–1765.

Nunes, Matthew A / Balding, David J(2010): *On Optimal Selection of Summary Statistics for Approximate Bayesian Computation*, 1: .

Fearnhead, Paul / Prangle, Dennis(2011): *Constructing Summary Statistics for Approximate Bayesian Computation: Semi-automatic ABC*.

