Approximate Bayesian Computation (ABC) for Stochastic Epidemic Models

Theo Kypraios & Phil O'Neill

https://www.maths.nottingham.ac.uk/plp/pmztk/

Modelling Process: Recap

- We have some observed data y;

- We formulate a model with parameters θ;

- We are interested in the posterior density $\pi(\theta|y)$.

Bayes' Theorem gives

$$\pi(\theta|y) = \frac{\pi(y|\theta)\pi(\theta)}{\pi(y)}$$

where

- $\pi(y|\theta)$ is the likelihood;
- $\pi(y)$ is a normalising constant;

and hence we write

$$\pi(\theta|y) \propto \pi(y|\theta)\pi(\theta).$$

Modelling Process: Recap

If the likelihood $\pi(y|\theta)$ is **intractable** then a solution is to introduce some *missing data* (x) as extra variables to produce a *tractable augmented* likelihood, denoted by $\pi(x, y|\theta)$ . . .

. . . and then use, for example, MCMC to sample from $\pi(\theta, x|y)$.

That is essentially what we did in Lectures 2 and 3 in which

- the removal times (r) where the observed data y;
- the infection and removal rates (β, γ) were the parameters θ;
- the infection times (i) were the missing data that we introduced to produce the augmented likelihood $\pi(x, y|\theta)$.

Outline

In this lecture we will look at an alternative method for sampling from the posterior distribution $\pi(\theta|y)$.

- Rejection sampling;

- Approximate Bayesian Computation (ABC);

  - Homogeneously mixing - final size data

  - 2-level mixing - final size data

  - ABC-MCMC

  - ABC-PMC

Rejection Sampling

Suppose we have discrete data y (eg final size), prior $\pi(\theta)$ for parameters θ (eg infection rate).

Consider the following algorithm:

ER1

1. Draw $\theta \sim \pi(\theta)$;
2. Accept θ with probability equal to $\pi(y|\theta)$;
3. Repeat until the required number of samples are drawn.

* One can show that the accepted values will be distributed according to the posterior distribution $\pi(\theta|y)$.

* Note that this algorithm requires that the likelihood, $\pi(y|\theta)$, can be evaluated for any θ (Step 2).

Consider the following algorithm:

ER2

1. Draw $\theta \sim \pi(\theta)$;
2. Draw $y' \sim \pi(\cdot|\theta)$, i.e. simulate data y' from the model;
3. Accept θ if $y = y'$;
4. Repeat until the required number of samples are drawn.

* Algorithm **ER2** is equivalent to algorithm **ER1**.

* The main difference is that the computation of the likelihood $\pi(y|\theta)$ (**ER1**, Step 2) is substituted by the **simulation of an event which occurs with that probability** (**ER2**, Steps 2 and 3).

* That means that the calculation of the likelihood is unnecessary as long as we can *simulate from our stochastic model*.

Consider some **final size data** y (i.e. total number of infectives out of the initial susceptible population of size N) from a Markov SIR model.

Suppose $N = 100$ and $y = 70$.

Let us also assume an Exponential infectious period of length 1 unit, $R_i - I_i \sim Exp(1)$

We wish to make inference for the infection rate θ, i.e. draw samples from $\pi(\theta|y)$.

- How can we draw samples from the posterior distribution $\pi(\theta|y)$?

- Shall we use MCMC? ER1? ER2?

Inference for Final Size Data

- Implementing an MCMC algorithm or the ER1 algorithm requires evaluation of the likelihood $\pi(y|\theta)$ for any θ.

- In other words, we need to able to compute how likely it is observe a final size of 70 out of a population of size 100 for any value $\theta \in (0, \infty)$.

- The *triangular equations* that were discussed in Lecture 4 could be used but unless Multiple Precision Arithmetic (MPA) is used, then calculations will break down.

- Hence, the likelihood $\pi(\theta|y)$ is *intractable*.

- However, we know how to simulate from an SIR model (Lecture 1).

Inference for Final Size Data

One option is to *augment* the observed data with *missing data*, treat them as additional parameters, and perform *data-augmentation* MCMC.

However, algorithm ER2 can be used instead:

ER2

1. Draw $\theta \sim \pi(\theta)$;
2. Draw $y' \sim \pi(\cdot|\theta)$, i.e. simulate data y' from the model;
3. Accept θ if $y = y'$;
4. Repeat until the required number of samples are drawn.

* Step 2 requires simulating from a Markov SIR model – we now how to do that (see Lecture 1).

* Very straightforward implementation $\rightarrow$ exact samples from $\pi(\theta|y)$!

Recall: $N = 100, y = 70, \gamma = 1$. We also assume that a priori $\theta \sim Exp(1)$

ER2

1. theta = rexp(n=1, rate = 1) # Draw from the prior
2. y.prime = simSIR.Markov(100, beta = theta, gamma = 1)
3. Accept theta if y.prime = 70
4. Repeat until the required number of samples are drawn.

There is **no approximation** up to this stage – ER2 will produce exact draws from the desired posterior distribution.

What if $N = 1000$ and $y = 782$? Will it take long to hit the target?

Consider the following algorithm:

ABC

1. Draw $\theta \sim \pi(\theta)$;
2. Draw $y' \sim \pi(\cdot|\theta)$, i.e. simulate data y' from the model;
3. Accept θ if $|y - y'| \leq h$;
4. Repeat until the required number of samples are drawn.

* In Step 3 we no longer require an exact match between the observed data y the simulated y' – Instead we allow for some discrepancy; large or small will depend on the value of h.

* The ABC algorithm above produces samples from an *approximation* to the posterior distribution $\pi(\theta|y)$.

ABC – Some Remarks

- Of course, if $h = 0$, then ABC2 is equivalent to ER2.

- Increasing h will considerably improve the acceptance rate of the ABC algorithm.

- However, there will now be accepted values of θ for which it was previously (using ER2) unlikely to generate data such that $y = y'$.

- The hope is that if h remains small, then the resulting estimate of the posterior will be still be close to the true posterior.

- Naturally, there is trade off between efficiency and accuracy.

- The marginal likelihood, $\pi(y)$, is derived as a by-product. Can be used for model choice (Bayes Factors etc).

- Step 3 can be more general; for example we can replace

    3. Accept θ if $|y - y'| \leq h$;

  with

    3. Accept θ if $||y - y'|| \leq h$;

  where $|| \cdot ||$ is some distance measure, such as the Euclidean distance.

- If we want a Gamma (2,2) infectious period instead of an Exp(1), we replace

    2. y.prime = simSIR.Markov(100, beta = theta, gamma = 1)

  with

    2. y.prime = simSIR.Non.Markov.Gamma(100, beta = theta, gamma = 2,
       delta = 2)

ABC for Final Size Data: Recap

Input: observed data y, stochastic model $\pi(\cdot|\theta)$, prior distributions for model parameters $\pi(\theta)$, distance measure $||\cdot||$, tolerance h.

ABC

1. Draw $\theta \sim \pi(\theta)$;
2. Draw $y' \sim \pi(\cdot|\theta)$, i.e. simulate data y' from the model;
3. Accept θ if $||y - y'|| \leq h$;
4. Repeat until the required number of samples are drawn.

Output: samples from (an approximation of) the posterior distribution $\pi(\theta|y)$.

ABC for 2-Level Mixing Models

We consider epidemic models in populations which mix at two levels, "global" and "local" (Lecture 4).

More precisely, we consider models in which

- each infectious individual i has a 'global' rate λ_G for infecting each other individual in the population . . .

- . . . plus a 'local' rate λ_L, typically much larger, of infecting each other individual among a set of neighbours, $N(i)$.

Our main concern is the case where the population is partitioned into local groups or households (or farms/yards etc).

The Table below presents some Asian influenza epidemic household data taken from Longini & Koopman (1982). There is a community which consists of **42 households each of them of size 3**.

Number of cases	Number of households
0	29
1	9
2	2
3	2

Task: Assuming a constant infectious period of 1 unit, infer the infection rates λ_L and λ_G.

EBC for 2-level Mixing Models

ER2

1. Draw $\theta \sim \pi(\theta)$;
2. Draw $\boldsymbol{y}' \sim \pi(\cdot|\theta)$, i.e. simulate data $\boldsymbol{y}'$ from the model;
3. Accept θ if $\boldsymbol{y} = \boldsymbol{y}'$;
4. Repeat until the required number of samples are drawn.

The total final size is equal to 47 but the actual data are

$$\boldsymbol{y} = (y_0, y_1, y_2, y_3) = (29, 9, 2, 2)$$

corresponding to 29 households where nobody got infected, 9 households where only 1 individual got infected, 2 households where 2 individuals got infected etc.

Step 3 now is more ambitious than before since we need to hit the target of the total final size but also the exact number of households with finali size 0,1,2 and 3.

ABC for 2-level Mixing Models

Doing exact inference seems very challenging and therefore, we seek for alternative, but **approximate ways** to draw inference for $\boldsymbol{\theta} = (\lambda_L, \lambda_G)$.

ABC

1. Draw $\theta \sim \pi(\theta)$;
2. Draw $\boldsymbol{y}' \sim \pi(\cdot|\theta)$, i.e. simulate data $\boldsymbol{y}'$ from the model;
3. Accept θ if $||\boldsymbol{y} - \boldsymbol{y}'|| \leq h$;
4. Repeat until the required number of samples are drawn.

Obvious question:

- What distance measure $|| \cdot ||$ should we use?

- Compare the total final size and ignore the group structure (not very clever):

$$||\boldsymbol{y} - \boldsymbol{y}'|| = \left| \sum_{i=0}^{3} y_i - \sum_{i=0}^{3} y_i' \right|$$

- Compare the final sizes in each groups::

$$||\boldsymbol{y} - \boldsymbol{y}'|| = \sum_{i=0}^{3} |y_i - y_i'|$$

- . . .

Data on influenza outbreak in Tecumseh @ Michigan, Winter 1980-1981 are shown below:

No. infected	Susceptibles per household						
	1	2	3	4	5	6	7
0	44	62	47	38	9	3	2
1	10	13	8	11	5	3	0
2		9	2	7	3	0	0
3			3	5	1	0	0
4				1	0	0	0
5					1	0	0
6						0	0
7							0
Total	54	84	60	62	19	6	2

Data: Influenza Outbreak

- Households are not all the same size.

- ER2 will be very inefficient in this case.

- Data $\boldsymbol{y}$ consist of the number of households of size j with final size i denoted by

$$\boldsymbol{y} = \{y_{ij}\}, \text{ for } i, = 1, \ldots, 7 \quad j = 1, \ldots, 7.$$

- A sensible distance measure could be

$$\|\boldsymbol{y} - \boldsymbol{y}'\| = \sum_{i=1}^{7} \sum_{j=1}^{7} \left| y_{ij} - y'_{ij} \right|$$

Data: Equine Influenza

Baguelin *et. al* (2009) implemented an ABC-type algorithm to draw inference for a two-level mixing model on some data on equine influenza on horses.

Data look like this:

Yard	N_i	d_i
1	60	18
2	83	78
3	103	80
4	141	78
5	36	6
6	25	19
7	19	16
8	190	139
9	32	30
10	41	21

- The number of initially susceptible horses in each yard (N_i) is different across yards.

- A natural distance measure could be :

$$||\boldsymbol{y} - \boldsymbol{y}'|| = \sum_{i=1}^{n} \left| \frac{d_i - d_i'}{N_i} \right|$$

in which look at discrepancies of the proportions of the number of infectives per yard (d_i/N_i) rather than discrepancies at the absolute numbers (d_i).

ABC Challenges

ABC algorithms are:

- fairly intuitive;

- easy to implement;

- amenable to routine analysis of data.

However, (at least the vanilla) ABC algorithms can be very inefficient.

Why?

- Simulating from (complex) models can be expensive;

- If the prior is vague, it will take ages to hit the target.

ABC Challenges

ABC algorithms are:

- fairly intuitive;

- easy to implement;

- amenable to routine analysis of data.

However, (at least the vanilla) ABC algorithms can be very inefficient.

Why?

- Simulating from (complex) models can be expensive;

- **If the prior is vague, it will take ages to hit the target.**

ABC within MCMC

- To overcome issues caused by a low acceptance probability when the prior and posterior distributions are rather different Marjoram *et al.* developed the **ABC-MCMC** algorithm.

- ABC-MCMC embeds the simulation step of ABC into a standard Metropolis–Hastings algorithm.

- ABC-MCMC is very similar to the standard MCMC algorithm:

MCMC

1. Propose $\theta' \sim q(\cdot|\theta)$
2. Accept θ' with probability

$$\min \left(1, \frac{\pi(y|\theta')}{\pi(y|\theta)} \frac{\pi(\theta')}{\pi(\theta)} \frac{q(\theta|\theta')}{q(\theta'|\theta)} \right)$$

3. Repeat until required posterior samples are obtained.

ABC-MCMC

ABC-MCMC

1. Propose $\theta' \sim q(\cdot|\theta)$;
2. Simulate K datasets, $y_1, y_2, \ldots, y_K$, from the model using the parameter θ and calculate

$$r(y, \theta') = \frac{1}{K} \sum_{k=1}^{K} \mathbb{I}(||y - y_k|| \leq h)$$

3. Accept θ' with probability

$$\min \left(1, \frac{r(y, \theta')}{r(y, \theta)} \frac{\pi(\theta')}{\pi(\theta)} \frac{q(\theta|\theta')}{q(\theta'|\theta)} \right)$$

4. Repeat until required posterior samples are obtained.

Essentially, the ratio $\pi(y|\theta')/\pi(y|\theta)$ is approximated by the ratio of proportion of K simulations that "match the data".

Intuition says that the larger the K the better; it turns out that there is no **much** improvement for large K.

ABC within Sequential Monte Carlo (SMC)

- We can also embed ABC simulation steps in Sequential Monte Carlo (SMC) algorithms.

- The idea is to sample from a sequence of approximate ABC posteriors under successively lower acceptance tolerances.

- The sample from one iteration, known as *particles*, is used to help choose which θ values to simulate from in the next.

- SMC methods for ABC have a number of potential advantages over MCMC and vanilla ABC.

- Unlike vanilla ABC they concentrate on simulating datasets from θ regions with relatively high acceptance probabilities, avoiding wasting computational resources.

- Unlike MCMC, they can adapt tuning choices, such as acceptance tolerances, during their operation

ABC-PMC

There are many ways via which we can embed ABC within SMC.

Here we focus on the algorithm proposed by Toni *et al.* (2009).

Effectively it performs repeated importance sampling, also known as population Monte Carlo (PMC).

Ingredients:

observed data y, tolerances $\epsilon_1 > \ldots > \epsilon_M$, distance measure $|| \cdot ||$, number of particles N, kernel $K(\cdot|\cdot)$, parameters governing $\pi(\theta)$.

Output:

weighted samples from (an approximation to) $\pi(\theta|y)$

ABC-PMC

Algorithm 1 ABC-PMC

1: Let $t = 1$.

2: Repeat the following steps until there are N acceptances.

    (a) Sample θ^* from the importance density $q_t(\theta)$ given in (1) below.

    (b) If $\pi(\theta^*) = 0$ reject and return to (a).

    (c) Simulate dataset y^* from the model using parameters θ^*.

    (c) Accept θ^* if $||y - y^*|| \leq \varepsilon_t$

    Denote by $\theta_1^t, \theta_2^t, \ldots, \theta_N^t$ the accepted parameter values.

3: Let

$$w_i^t = \pi(\theta_i^t)/q_t(\theta_i^t)$$

    for $i = 1, 2, \ldots, N$.

4: Increment $t = t + 1$.

5: Repeat Steps 2 and 3 until $t = M$.

6: Return final accepted sample and weights.

The importance density in Step 2a) is given by

$$q_t(\theta) = \begin{cases} \pi(\theta) & \text{if } t = 1, \\ \sum_{i=1}^{N} w_i^{t-1} K_t(\theta | \theta_i^{t-1}) / \sum_{i=1}^{N} w_i^{t-1} & \text{otherwise.} \end{cases} \quad (1)$$

In other words, in the first iteration $q_t(\theta)$ is the prior.

For the subsequent iterations, weighted samples from the previous particle population are drawn and then perturbed using the kernel K_t.

The choice of the kernel is arbitrary.

A good choice is

$$K(\theta^t|\theta^{t-1}) = \phi\left(\theta^{t-1}, 2 \cdot \Sigma^{t-1}\right)$$

where

- $\phi(\cdot, \cdot)$ is the density of a (multivariate) Normal distribution;

- Σ^{t-1} is the empirical covariance matrix of the particle population at time $t-1$, $\{\theta_i^{t-1}\}_{1 \leq i \leq N}$, calculated using weights $\{w_i^{t-1}\}_{1 \leq i \leq N}$.

The choice of the *tolerance ladder* $\epsilon_1 > \ldots > \epsilon_M$ is arbitrary.

However, the efficiency of the algorithm depends (among other things) on $\epsilon_1 > \ldots > \epsilon_M$.

One approach is to determine the sequence of tolerances adaptively during the algorithm.

To do so one selects an initial tolerance, such as $\epsilon_1 = \infty$, and then selects ϵ_{t+1} between steps 3 and 4.

The value used is the α quantile of $d_1^t, d_2^t, \ldots, d_N^t$, where these are the $||y - y^*||$ distances from the accepted simulations at time t, and $0 < \alpha < 1$ is a tuning parameter.

Summary

- Key ideas relating to drawing (approximate) Bayesian inference for stochastic epidemic models using ABC methods;

- Broadly speaking, the methods are very flexible and can be used for more complex models;

- However, as with any application of ABC (or even MCMC methods), the existence of an algorithm does not necessarily imply that is either efficient or practical.

- ABC methods are very well suited for cases where the models of interest have a relatively few parameters but the likelihood of the observed data is either unavailable or difficult to compute.

Further Considerations

- ABC is a very active area of research;

- Temporal data $\rightarrow$ summary statistics: how?

- Distance metrics;

- Computational aspects;

Mathematical Biosciences
Volume 287, May 2017, Pages 42-53

A tutorial introduction to Bayesian inference for
stochastic epidemic models using Approximate
Bayesian Computation

Theodore Kypraios [a], Peter Neal [b], Dennis Prangle [c]

⊞ Show more

Code: https://github.com/kypraios/epiABC