Bayesian Modelling of Epidemic Processes

D. Hutchinson

March 31, 2021

Abstract

Contents

1   Introduction

What is a model? A (simple) mathematical formulation of a process which incorporates parameters of interest and likely some stochastic processes. Models need to be computational tractable (i.e. fairly simple)

"All models are wrong, some are useful".

What to use models for? check intuition, explanation & prediction.

What is "posterior estimation"?

The problem - Posterior estimation when likelihood is intractable. "Likelihood-free" estimation. (Classical example of determining most recent common ancestor of two DNA strands. Likelihood is intractable due to number of branches growing factorially. (Burr and Skurikhin (2013))

Motivation

Bayes Rule? Describe each component & why is likelihood intractable?

Why now? More, better data. Greater computational power.

What can posterior be used for?

Motivating Examples

DNA mutation (Marjoram and Tavaré (2006))

History

Traditional parameter estimation methods - "Maximum Likelihood".

Neutrality testing - (Hypothesis testing), compare results against a null hypothesis for a parameter value.

Successful Applications of these Methods

2   Bayesian Modelling

Bayes' Rule
vs. Frequentist modelling
Stochastic vs deterministic models
Consistency

3 Approximate Bayesian Computation

What are "Simulation Methods"? (Possible now due to greater computational power and quality/quantity of data available) Two versions: allow for running models with stochastic components and testing variety of results; or, simulation for statistical inference and parameter estimation (vary parameters in simulation and compare results to observed data).

Simulation methods are useful here as it is easier to simulate from a distribution than to calculate it.

ABC extends many Bayesian simulation methods to only require approximate matching of observed and simulated data, rather than exact. (ABC allows for more simplified models?). Trade-offs?

In general, we never know if our calculated posterior is actually close to the true posterior.

ABC simulates from the likelihood, rather than explicitly determine the likelihood.

ABC has two uses: calibrating model, comparing models.

What is the computational effiency?

How to tune tolerance ε? Run tests showing performance with different tolerances (plots of posteriors when using different tolerances).

Curse of dimensionality (regression methods can counter this)

Accept-Reject-ABC

Perform poorly when prior and posterior and very different (especially when overlap is small).

Worth tracking "acceptance rate" to test algorithms performance. Don't want too high or too low, time v quality.

Variations - Sample until N accepted (how to determine width of acceptance); Sample M times and keep best N (how to determine best); Sample M times, keep all and weight parameters by distance of simulated data from observed data (weighted linear-regression).

Accept-reject algorithms for MLE (rather than Bayesian). (Use mode of posterior distribution as MLE or use simulation to approximate likelihood).

Using a uniform kernel can be interpreted as sampling from a model which has uniformly distributed additive error.

Independent samples

Acceptance-rate analogous to evidence.

User choices: summary stats; threshold; distance measure.

Acceptance-rate vs tolerance plot?

re

3.1 Importance Sampling

Use when have a better knowledge about parameter value distribution. (May be good for RONA as we are given good R rate estimates?)

Sample parameters θ from an "Importance distribution" $\xi(\theta)$ rather than a "prior" $\pi(\theta)$. Weight each accepted parameter θ as $\frac{\pi(\theta)}{\xi(\theta)}$.

Less variance between observed and simulated summary statistic values.

Could use rejection algorithm to determine a good importance distribution.

Independent samples

`https://bookdown.org/rdpeng/advstatcomp/importance-sampling.html`

ABC-MCMC, -SMC & -PMC

MCMC, SMC are search processes

ABC-MCMC=Marjoram *et al.* (2003) (Markov Chain Monte Carlo).

ABC-SMC=Sisson *et al.* (2007) (Sequential Monte Carlo).

ABC-PMC=Beaumont *et al.* (2009) (Population Monte-Carlo).

Good approach for large data sets, and when prior & posterior are likely to be very different.

Sequences of dependent samples

Adaptive approach to ABC-SMC (Moral *et al.* (2012))

When does MCMC converge?

MCMC w/o summary stats converges to the true posterior $(\theta|D)$.

What is the burn-in period like? How good/important is mixing?

Advantage of ABC-MCMC over ABC-rejection? Fewer simulations required to get n accepted samples (for a given tolerance ε).

Acceptance-rate vs tolerance plot?

Stationary distribution of the MCMC is an estimate of the posterior?

How to choose perturbation kernels (Filippi *et al.* (2012)).

Regression Adjustment

Beaumont et al - Local Linear Regressions (LOCL)

Blum and Francois' - Nonlinear Conditional heteroscedastic regressions (NCH). (Uses neural networks)

Review

Which algorithm to use in different scenarios - complexity of model, amount of data available.

4 Summary Statistic Selection

In this chapter I motivate the research into summary statistic selection and discuss features which are desirable in a summary statistic. I then describe five methods for summary statistic selection methods: three which use handcrafted summary statistics; and two which generate their own; Finally, I use a toy example of an SIR model to compare these methods.

Motivation

The study of summary statistics has relevance beyond ABC methods, largely due to the recent "Big-Data Revolution" which has seen the rate at which data can be collected and stored significantly outpace improvements in computational power. This has motivated research into effective methods to reduce the size of datasets so that more computationally intensive algorithms can be used to analyse the data.

A summary statistic S is a statistic which reduces the dimensionality of some observed data, in a deterministic fashion, whilst retaining as much information about the observed data as possible. Reducing the dimensionality of data is desirable as it reduces the computational requirements to analyse the data.

$$s : \mathbb{R}^m \to \mathbb{R}^p \text{ with } m > p$$

Ideally, a summary statistic would compress the observed data without any information loss (A property known as "sufficiency"). However, sufficient summary statistics are very rare in practice and we often have to trade-off information retention against dimensionality reduction.

For more complex models, with many parameters, it becomes difficult for a single summary statistic to accurately summarise the observed data, thus, in practice it is common to apply a set of summary statistics $\{s_1, \ldots, s_k\}$ to the same dataset with each targeting a different aspect of the model. As long as the sum of the dimensions of the outputs from the summary statistics in the set is less than the cardinality of the observed data, then using a set of summary statistics still produces effective dimensionality reduction.

$$m > \sum_{i=1}^{k} p_i \text{ where } s_i : \mathbb{R}^m \to \mathbb{R}^{p_i}$$

The success of ABC methods depends mainly on three user choices: choice of summary statistic; choice of distance measure; and choice of acceptance kernel. Of these, summary statistic choice is arguably the most important as the other two mainly effect the rate at which the algorithm converges on the posterior mean. Whereas, choosing summary statistics which are uninformative can lead to the parameter posteriors returned by the algorithm being drastically different from the true parameter posteriors. This is trivial to realise if you consider a scenario where $s(x) = c$, for some constant $c \in \mathbb{R}$, is used as the sole summary statistic as this would result in the returned posteriors being the same as the priors supplied to the algorithm.

In practice, the quality of the posteriors returned from an ABC method is limited by the amount of computational time which is dedicated to running the algorithm. For some problems, such as .... , it is reasonable to dedicate the majority of your computing time on summary statistic selection, rather than on model fitting, as it is clear that the more computationally efficient ABC methods (e.g. ABC-Rejection Sampling) will be sufficient to fit the model, given a good choice of summary statistics.

Current Thinking

Traditionally, summary statistics for ABC methods are chosen manually using expert, domain-specific knowledge. Utilising this expert knowledge is desirable as these statistics will incentivise exploring regions of the parameter space which have been scientifically shown to be relevant to the given problem and thus more likely to contain the true parameter values (Similarly, these statistics will disincentivise exploring regions which have been shown to not be of interest). Often these hand-crafted statistics are highly interpretable which is important when using ABC methods to fit models which will influence real-world decisions as you will likely need to convince stakeholders that your results are reasonable.

However, relying on expert knowledge to choose summary statistics limits the scenarios where ABC methods can be applied to only those where there has already been significant research. And, leads to statistics being chosen due to their prevalence in a field rather than their suitability to ABC methods. Moreover, the use of handcrafted summary statistics means that any misunderstandings or errors in a field will be encoded into the model fitting process, possibly leading to misspecification.

When using a set of summary statistics, expert knowledge is generally not sufficient to determine how best to weight each summary statistic. Some of the methods I describe below allow can be used to automate the process of determining these weights by specifying multiple versions of the same summary statistic but with each version having a different weight.

4.1 Properties of Summary Statistics

Unbiased, efficient, consistent, (All covered by "sufficient").

Sufficiency

4.2 Methods for Summary Statistic Selection

When thinking about summary statistic selection it is useful to consider the summary statistics as a feature of your theorised model. This makes the process of selecting summary statistics analogous to model selection, with each combination of summary statistics being considered is a different model. This is the motivation behind most summary statistic selection methods.

Approximate Sufficient Subset Joyce and Marjoram (2008)

Minimising Entropy Nunes and Balding (2010)

Two-Step Minimum Entropy Nunes and Balding (2010)

Semi-Automatic ABC Fearnhead and Prangle (2011)

Non-Linear Projection

Toy Example

4.3 Model Selection

Theorems which state when a model is misspecified that bayesian inference will put mass on the distributions "closest to the ground truth" rely on strong regularity conditions. Grünwald and van Ommen (2018)

Introduce learning rate (SafeBayes) Grünwald and van Ommen (2018)

5 ABC and Epidemic Events

References

Beaumont, M. A., Cornuet, J.-M., Marin, J.-M. and Robert, C. P. (2009). Adaptive approximate Bayesian computation. *Biometrika* **96**(4), 983–990.

Burr, T. and Skurikhin, A. (2013). Selecting summary statistics in approximate bayesian computation for calibrating stochastic models. *BioMed research international* **2013**, 210646.

Fearnhead, P. and Prangle, D. (2011). Constructing summary statistics for approximate bayesian computation: Semi-automatic abc .

Filippi, S., Barnes, C., Cornebise, J. and Stumpf, M. P. H. (2012). On optimality of kernels for approximate bayesian computation using sequential monte carlo .

Grünwald, P. and van Ommen, T. (2018). Inconsistency of bayesian inference for misspecified linear models, and a proposal for repairing it .

Joyce, P. and Marjoram, P. (2008). Approximately Sufficient Statistics and Bayesian Computation. *Statistical Applications in Genetics and Molecular Biology* **7**(1), 1–18.

Marjoram, P. and Tavaré, S. (2006). Modern computational approaches for analysing molecular genetic variation data. *Nat Rev Genet* **7**, 759–770.

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.

Moral, P., Doucet, A. and Jasra, A. (2012). An adaptive sequential monte carlo method for approximate bayesian computation **22**(5), 1009–1020.

Nunes, M. and Balding, D. (2010). On optimal selection of summary statistics for approximate bayesian computation. *Statistical Applications in Genetics and Molecular Biology* **9**(1).

Sisson, S. A., Fan, Y. and Tanaka, M. M. (2007). Sequential monte carlo without likelihoods. *Proceedings of the National Academy of Sciences* **104**(6), 1760–1765.