A. Mira - thanks to C. Drovandi

a university for the real world®

QUT

Section 11

Outline

Introduction to approximate Bayesian computation (ABC)

Toy example (Courtesy of Dennis Prangle)

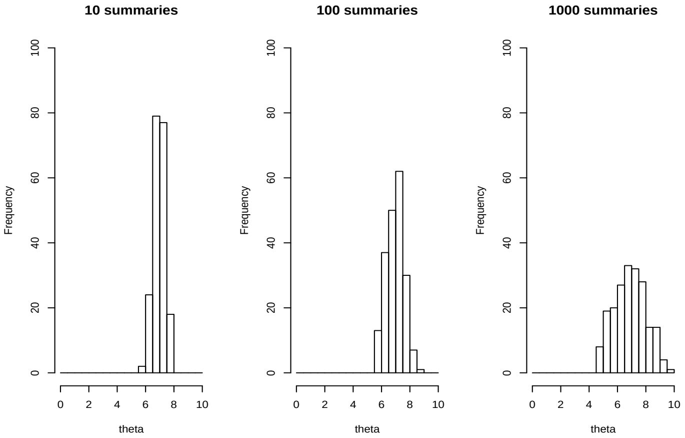
Let θ be a scalar parameter

Suppose $y_i = \begin{cases} \theta + e_i & \text{for } i = 1, 2, \ldots, 10 \\ e_i & \text{for } i = 11, 12, \ldots, d \end{cases}$

where $e_i \sim N(0, 1)$ (iid)

First 10 components are informative about θ (signal)

Others are uninformative (noise)


10 summaries
100 summaries
1000 summaries
Frequency
theta
0
20
40
60
80
100
0
2
4
6
8
10

Intuitive explanation

In each case the same information about θ is being used

But, as d increases, close matches are increasingly difficult due to the *noise* matching well

So less is learnt about θ

More generally there is a curse of dimensionality issue. As there are more dimensions for the data to vary in, it is increasingly hard to get close matches

Information vs Dimensionality

It is now clear that we should not necessarily compare full datasets, especially if they are large, due to the curse of dimensionality.

To allow for easier matching, it is necessary to reduce the full dataset to a set of summary statistics, $S(y)$

We know that $S(y)$ is not sufficient, but we reduce the effect of the ϵ error.

Ultimately, choice of summary statistic is a trade-off between dimensionality and information loss.

Some Approaches for Choosing Summary Statistics

Subset Selection - choose the best subset out of a large (ad-hoc) set

Auxiliary model - find an (tractable) alternative model that fits the data well and use that to create summary statistics

Semi-automatic ABC - use (estimates) of the posterior means as the summary statistics

Subset Selection

Start from a set of data features: potential summaries

Try different subsets of summary statistics and see which gives the best results

Choose based on some measure of ABC posterior concentration (criterion). e.g. standard deviation or entropy

Easy to implement, especially with ABC rejection

Fairly robust given that you start with a large number of summaries

Might be very expensive to enumerate all possible subsets

Not obvious how to choose the criterion

Summaries from Auxiliary Models

In indirect inference, estimation of θ is facilitated by specifying an alternative (auxiliary) parametric model with parameter ϕ.

One could use the parameter estimates (e.g. MLE) of this auxiliary model fitted to the observed data, $\phi(y)$, as a summary statistic for ABC. The simulated summary statistic comes from fitting the auxiliary model to data simulated from the 'true' model, $\phi(x)$.

Can be expensive as requires fitting the auxiliary model to each dataset simulated from the 'true' model.

Alternatively can use the score of the auxiliary model as summary statistics

$$S_A(y,\phi) = \left(\frac{\partial \log p_A(y|\phi)}{\partial \phi_1}, \cdots, \frac{\partial \log p_A(y|\phi)}{\partial \phi_{\dim(\phi)}} \right)^T,$$

Need a parameter value for ϕ. Choose $\phi(y)$ for example.

Massive improvement in computation time if analytic expression for the score as only need to fit the auxiliary model to observed data once before ABC analysis.

Summaries from Auxiliary Models

If auxiliary model fits data well gives a ‘good’ feeling about summaries derived from that model. This can all be done prior to the ABC analysis.

Control curse-of-dimensionality by choosing parsimonious auxiliary model.

Semi-automatic ABC

Fearnhead and Prangle (2012) JRSS B (FP hereafter) suggest using the posterior means as summary statistics for ABC.

Optimal in the sense that posterior means minimise quadratic loss

$$L(\theta_0, \hat{\theta}; A) = (\theta_0 - \hat{\theta})^Y A(\theta_0 - \hat{\theta}),$$

where θ_0 is true value and $\hat{\theta}$ is posterior mean.

But we don't have access to the posterior mean.

Idea is to estimate posterior mean using regression.

Simplest FP approach

Perform regression on each parameter θ_i, $i = 1, \dots, p$, where p is number of parameters.

Draw large collection of parameter values, say M, from the prior $\theta_i^j \sim \pi(\theta)$.

For each parameter sample, simulate a dataset, x_i^j from the model and calculate a set of summary statistic (possibly high dimensional) values, $s_i^j = f(x_i^j)$.

Perform a multiple linear regression:

$$\theta_i^j = \beta_0 + f(x_i^j)^T \beta + \epsilon_j, \;\; j = 1, \dots, M.$$

$\beta_0 + f(x_i^j)^T \beta$ is a model for $E[\theta | x_i^j]$.

Denote estimates of regression coefficients ($\hat{\beta}_0$ and $\hat{\beta}$).

The 'observed' summary statistic, $S(y) = \hat{\beta}_0 + f(y)^T \hat{\beta}$ is thus an estimate of the posterior mean. In ABC, the simulated summary statistic for some simulated dataset x is $S(x) = \hat{\beta}_0 + f(x)^T \hat{\beta}$.

Note we can omit $\hat{\beta}_0$ as it will cancel in $S(y) - S(x)$.

Extensions

‘Pilot run’ with some $S(y)$ to reduce the parameter space over which to perform the regression.

Use model choice criterion to choose ‘best’ regression model.

Normal linear regression not a necessity.

Advantages and Disadvantages

Advantages:

- $f(y)$ may be very high dimensional.
- Fights the curse of dimensionality: One summary statistic per parameter.
- Has theoretical support in terms of point estimation.

Disadvantages:

- Still not a sufficient statistic!
- May not be easy to find suitable regression model.

Choosing $\rho(y, x)$

Another difficult problem in ABC in choosing how to compare summary statistics

L_1 distance. Sum of absolute differences between each component of summary statistics.

L_2 distance. Euclidean distance.

Mahalanobis distance

$$\rho(y, x) = (S(y) - S(x))^T W (S(y) - S(x)),$$

where W is a 'weighting' matrix that takes into account the scale of summaries and correlations between summaries. Not easy to determine appropriate W.

Discrepancy functions via Auxiliary Model Approach

The weighting matrix can be based on the estimated covariance matrix of auxiliary parameter estimates, $J(\phi_y)$, obtained from fitting auxiliary model to the data.

Direct comparison of auxiliary parameters

$$\rho(y, x) = (\phi_x - \phi_y)^T J(\phi_y)(\phi_x - \phi_y)$$

Indirect comparison of auxiliary parameters through likelihood function of auxiliary model

$$\rho(y, x) = \log p_A(y|\phi_y) - \log p_A(y|\phi_x)$$

Comparison through score of auxiliary model

$$\rho(y, x) = S_A(x, \phi_y)^T J(\phi_y)^{-1} S_A(x, \phi_y)$$

Current Research

STILL NO UNIVERSALLY ACCEPTED METHOD FOR OBTAINING SUMMARY STATISTICS. AT THIS STAGE, THE OPTIMAL CHOICE IS PROBLEM DEPENDENT.

STILL NO UNIVERSALLY ACCEPTED WAY TO COMPARE SUMMARY STATISTICS.