Constructing ABC summary statistics: semi-automatic ABC

Paul Fearnhead and Dennis Prangle

5th May 2011

Background

- Good summary statistic choice crucial for efficiency of ABC methods
- *Semi-automatic ABC* is our summary statistic generation method
- Generic: can be used in any setting
- Semi-automatic: limited user input needed

Overview

- Theory
- Method
- Examples
- Summary / extensions

Part I

Theory

Notation

- Observed data $\mathbf{y}_{\text{obs}}$
- Generic data $\mathbf{y}$
- Parameters θ
- Prior $\pi(\theta)$
- Summary statistics $S(.)$
  - Giving $S(\mathbf{y})$ a vector of summaries
- ABC bandwidth $h \geq 0$

Example algorithm

ABC rejection sampling algorithm

For $i = 1, 2, \ldots, n$:

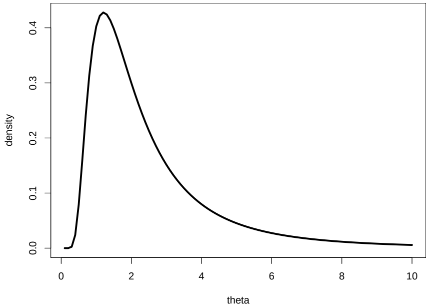
1. Sample θ_i from $\pi(\theta)$
2. Simulate data $\mathbf{y}_{\text{sim}}$ from model conditional on θ_i
3. If $||S(\mathbf{y}_{\text{sim}}) - S(\mathbf{y}_{\text{obs}})||_2 < h$ accept θ_i

Can generalise step 3 e.g. to be non-deterministic

ABC approximations

ABC has a hierarchy of approximations:

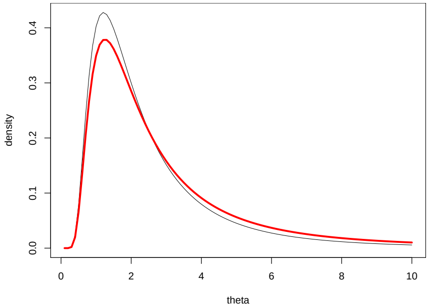
1. Posterior distribution: $\pi(\theta|\mathbf{y}_{\text{obs}})$
2. Posterior given observed summary statistics: $\pi(\theta|\mathbf{s}_{\text{obs}})$
    - where $\mathbf{s}_{\text{obs}} = S(\mathbf{y}_{\text{obs}})$
3. "ABC posterior", ABC algorithm target for $h > 0$
4. Monte Carlo approximation to ABC posterior



ABC approximations

ABC has a hierarchy of approximations:

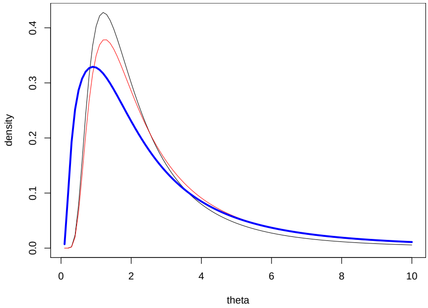
1. Posterior distribution: $\pi(\theta|\mathbf{y}_{\text{obs}})$
2. Posterior given observed summary statistics: $\pi(\theta|\mathbf{s}_{\text{obs}})$
   - where $\mathbf{s}_{\text{obs}} = S(\mathbf{y}_{\text{obs}})$
3. "ABC posterior", ABC algorithm target for $h > 0$
4. Monte Carlo approximation to ABC posterior



ABC approximations

ABC has a hierarchy of approximations:

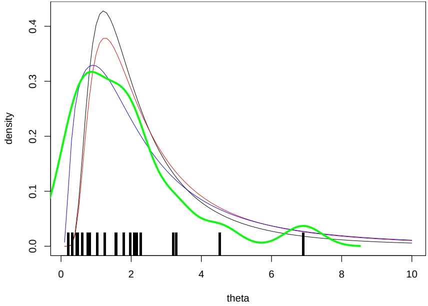
1. Posterior distribution: $\pi(\theta|\mathbf{y}_{\text{obs}})$
2. Posterior given observed summary statistics: $\pi(\theta|\mathbf{s}_{\text{obs}})$
    - where $\mathbf{s}_{\text{obs}} = S(\mathbf{y}_{\text{obs}})$
3. "ABC posterior", ABC algorithm target for $h > 0$
4. Monte Carlo approximation to ABC posterior



ABC approximations

ABC has a hierarchy of approximations:

1. Posterior distribution: $\pi(\theta|\mathbf{y}_{\text{obs}})$
2. Posterior given observed summary statistics: $\pi(\theta|\mathbf{s}_{\text{obs}})$
    - where $\mathbf{s}_{\text{obs}} = S(\mathbf{y}_{\text{obs}})$
3. "ABC posterior", ABC algorithm target for $h > 0$
4. Monte Carlo approximation to ABC posterior



ABC Monte Carlo error result

- ABC Monte Carlo error increases in $\dim(S)$
- Asymptotic result for small h
- So if $\dim(S_1) >> \dim(S_2)$
- Then S_1 needs much bigger h to achieve reasonable Monte Carlo error...
- ...introducing ABC posterior error

ABC Monte Carlo error result: proof

- Proved for ABC rejection sampling ▸ Sketch
- Heuristic result for ABC MCMC and ABC importance sampling
- Should extend from importance sampling to ABC SMC algorithms

Summary statistics

Standard approach:

- Choose approximately sufficient $S(.)$
- Then $\pi(\theta|\mathbf{y}_{\text{obs}}) \approx \pi(\theta|\mathbf{s}_{\text{obs}})$

But:

- Typically such S must be high dimensional
- This causes large Monte Carlo error
- And we argue this approach not feasible in general

We focus on low dimensional S which is non-sufficient but still useful is some way

Summary statistics

- We focus on accurate point estimates
  - i.e. minimising expected squared bias of ABC posterior mean, given some true θ_0
- Result: an optimal choice of summary statistics is $S(\mathbf{y}) = E(\theta|\mathbf{y})$
  - i.e. parameter means conditional on data $\mathbf{y}$
- Satisfies low dimensionality requirement; one summary statistic for each parameter
  - (any fewer typically prevents identifiability)

Motivation for our method

- Shown that $E(\theta|\mathbf{y})$ desirable summary statistics
  - Accurate with low Monte Carlo error
- But not available in practice
- So our method is to construct a statistical estimate, $S(\mathbf{y}) \approx E(\theta|\mathbf{y})$
- Based on many simulated $(\theta, \mathbf{y})$ pairs

Further theory

- We argue this S is also useful for interval estimates for parameters
- Argument involves "calibration" property
- and "noisy ABC" algorithms
- More details in paper

Part II

Semi-automatic ABC method

Overview

1. (Optional) Initial ABC analysis using ad-hoc summary statistics
2. Choose a training region of parameter space
3. Simulate many $(\theta, \mathbf{y})$ values
   - Draw θ from prior truncated to training region
   - Then draw $\mathbf{y}$ from model conditional on θ
4. Estimate $\hat{\theta}(\mathbf{y}) \approx E(\theta|\mathbf{y})$
5. Use $S(\mathbf{y}) = \hat{\theta}(\mathbf{y})$ in ABC
   - Truncating prior to training region
   - $(\hat{\theta}(\mathbf{y})$ may have unexpected behaviour elsewhere)

Comparison to regression correction

Estimation

We focus on using linear regression for final step:

- Estimate each parameter θ_i as a linear combination of $f(\mathbf{y})$
- Where $f()$ a vector of appropriate functions of the data, the "explanatory variables"
  - e.g. raw data,
  - transformations,
  - approximate parameter estimates,
  - useful summaries
- Linear regression computationally cheap for large sets of simulated data
- Crude estimators in themselves
- But useful within ABC

User input

The user must:

- Perform initial ABC analysis
  - Only the rough support of the posterior needed
- Choose $f(\mathbf{y})$
  - Various diagnostics and tools available
  - Model insight also helpful
- Perform main ABC analysis

Part III

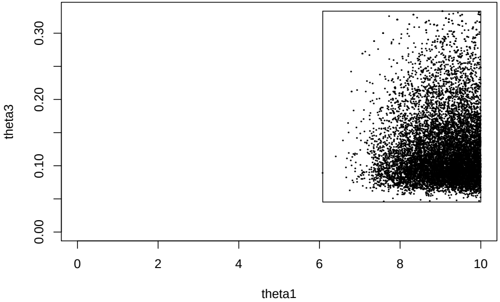
Examples

Partially observed queue

- Single server queue (initially empty)
- Service times uniform on $[\theta_1, \theta_2]$
- Inter-arrival times exponential with rate θ_3
- Only inter-departure times $y_1, y_2, \ldots, y_{50}$ observed
- Priors:
  - $\theta_1, \theta_2 - \theta_1$ both uniform on $[0, 10]$
  - θ_3 uniform on $[0, 1/3]$
- ABC analysis of Blum and François (2010) used 20 quantiles of data as summary statistics

Partially observed queue

- Step 1: We used Blum and François's summary statistics in an initial (MCMC) ABC analysis
- Step 2: Training region chosen to be hypercube containing all output points i.e.



Partially observed queue

- Step 3: We sample N parameter vectors from prior restricted to training region

Iteration	θ_1	θ_2	θ_3
1	9.25	13.34	0.24
2	9.91	15.21	0.08
3	7.99	14.26	0.07
4	8.67	11.16	0.15
5	6.58	10.46	0.08
⋮	⋮	⋮	⋮

Partially observed queue

- For each θ we simulate data $\mathbf{y}$ from the model
- And record explanatory variables $f(\mathbf{y})$
  - We chose ordered inter-departure times

Iteration	θ_1	θ_2	θ_3	$y^{(1)}$	$y^{(2)}$	$\dots$
1	9.97	16.04	0.19	10.16	10.30	$\dots$
2	9.36	11.61	0.12	9.39	9.43	$\dots$
3	6.84	14.34	0.15	6.96	7.08	$\dots$
4	7.78	12.15	0.05	7.97	8.20	$\dots$
5	7.04	16.81	0.12	7.16	7.18	$\dots$
6	6.79	10.96	0.07	7.01	7.02	$\dots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Partially observed queue

- Step 4: Linear regression produced

$$\hat{\theta}_1(\mathbf{y}) = 0.016 + 1.039y^{(1)} + 0.003y^{(2)} + \dots$$
$$\hat{\theta}_2(\mathbf{y}) = 1.138 - 0.945y^{(1)} + 0.022y^{(2)} + \dots$$
$$\hat{\theta}_3(\mathbf{y}) = 0.324 + 0.003y^{(1)} - 0.001y^{(2)} + \dots$$

- Step 5: We used $S(\mathbf{y}) = (\hat{\theta}_1(\mathbf{y}), \hat{\theta}_2(\mathbf{y}), \hat{\theta}_3(\mathbf{y}))$ within (MCMC) ABC

Partially observed queue

Mean squared bias over 50 datasets with varying parameters

	θ_1	θ_2	θ_3
Blum/François summary statistics	1.1	2.2	0.0013
Semi-automatic ABC	0.022	1.1	0.0013

- (n.b both analyses used same overall number of simulated data sets)
- Our method requires little user input here

Partially observed queue

Mean squared bias over 50 datasets with varying parameters

	θ_1	θ_2	θ_3
Blum/François summary statistics	1.1	2.2	0.0013
...with regression correction	0.022	1.1	0.0013
Semi-automatic ABC	0.022	1.1	0.0013

- (n.b both analyses used same overall number of simulated data sets)
- Our method requires little user input here
- Semi-automatic ABC and regression correction both successful

g-and-k distribution

- The g-and-k distribution has a flexible shape and small number of parameters
- No closed form density but simulation easy
  - (MLEs are available numerically)
- Parameters (A, B, g, k) related to location, scale, skewness, kurtosis
- Consider inference given $10,000$ iid draws
  - Uniform prior on $[0, 10]^4$
- ABC analysis by Allingham et al (2009) used 10,000 order statistics as S

g-and-k distribution

- Training region based on analysis using 10,000 summary statistics
- Many $(\theta, \mathbf{y})$ values simulated
- Choice of $f(\mathbf{y})$ bigger issue than previous example
  - Use subset of order statistics?
  - Include powers of order statistics?
- We fit linear regressions for a range of different $f(\mathbf{y})$ choices
- Then compared BIC values
- We chose 100 evenly spaced order statistics and their first 4 powers

g-and-k distribution

Mean squared bias over 50 datasets

	A	B	g	k
10,000 summary statistics	0.0059	0.0013	3.85	0.00063
...with regression correction	0.00040	0.0017	0.28	0.00051
Semi-automatic ABC	0.00016	0.00056	0.044	0.00023
MLE	0.00016	0.00055	0.0013	0.00014

- n.b. same number of simulated data sets in all analyses
- Good semi-automatic ABC results despite poor training region for g
- Choice of $f(\mathbf{y})$ more involved but some tools available

Part IV

Conclusion

Summary

- Results:
  - Monte Carlo error increases with $\dim(S)$
  - $S(\mathbf{y}) = E(\theta|\mathbf{y})$ ideal for ABC point estimation
- Motivates an estimate of $E(\theta|\mathbf{y})$ as $S(\mathbf{y})$
- Semi-automatic ABC is our method to produce such an estimate
- Effectiveness shown in two examples
- More details:
  - Our paper `www.arxiv.org/abs/1004.1112`
  - Supplementary material in `www.maths.lancs.ac.uk/~prangle/thesis_DP.pdf`

Extensions

- Nuisance parameters
- Alternatives to linear regression
- ABC for state space models
  - Application of noisy ABC
  - More details in paper
  - Including simple systems biology example
- Extension to model choice

Model choice

- ABC model choice not robust to choice of summary statistics (Robert et al 2011)
- But using many statistics for approximate sufficiency has Monte Carlo error problems
- Could adapt semi-automatic ABC to choose summary statistics for model choice
  - View model choice as inference for a categorical parameter
  - and use e.g. logistic regression
- Would need to extend our theory to show that results robust and have a useful practical interpretation

Part V

Appendix

ABC Monte Carlo error

- Consider ABC rejection sampling
- Monte Carlo error depends on acceptance rate α
- (i.e. variance of an ABC estimator $\propto 1/\alpha$)

$$\alpha = \int \pi(\theta)\pi(\mathbf{y}|\theta)\mathbb{I}[||S(\mathbf{y}) - S(\mathbf{y}_{\text{obs}})||_2 < h] d\mathbf{y} d\theta$$

- α also crucial to Monte Carlo error in other ABC algorithms

ABC Monte Carlo error asymptotics

$$\alpha = \int \pi(\theta)\pi(\mathbf{y}|\theta)\mathbb{I}\left[||S(\mathbf{y}) - S(\mathbf{y}_{\text{obs}})||_2 < h\right] \mathrm{d}\mathbf{y}\mathrm{d}\theta$$

Rearrange to

$$\alpha = \int \pi(\mathbf{s})\mathbb{I}\left[||\mathbf{s} - S(\mathbf{y}_{\text{obs}})||_2 < h\right] \mathrm{d}\mathbf{s}$$

Using Taylor expansion, for small h

$$\alpha \propto h^d \pi(S(\mathbf{y}_{\text{obs}})) + \text{smaller terms}$$

where d is dimension of summary statistics

Back

Regression correction

1. Initial ABC analysis
2. Estimate $\hat{\theta}(\mathbf{y}) \approx E(\theta|\mathbf{y})$
   - Based on accepted $(\theta, \mathbf{y})$ values
3. Use $\hat{\theta}(\mathbf{y})$ to adjust ABC results

Comparison

- Both methods use a similar regression step
- But use results for different purposes
- Regression correction can be applied to semi-automatic ABC results

Back