Approximate Bayesian Computation using Indirect Inference

Chris Drovandi
c.drovandi@qut.edu.au

Acknowledgement: Prof Tony Pettitt and Prof Malcolm Faddy

School of Mathematical Sciences, Queensland University of Technology

June 26, 2012
Approximate Bayesian Computation (ABC)

- Bayesian statistics involves inference based on the posterior distribution

\[ \pi(\theta | y) \propto f(y | \theta) \pi(\theta). \]

- What happens when likelihood \( f(y | \theta) \) unavailable?
- ABC instead uses model simulations and compares simulated with observed
- Naive Algorithm - Rejection Sampling
  - Sample \( \theta \sim \pi(\cdot) \)
  - Simulate \( x \sim f(\cdot | \theta) \)
  - Accept \( \theta \) if \( \rho(y, x) \leq \epsilon \)
- Repeat the above until we have \( N \) draws, \( \theta_1, \ldots, \theta_N \)
The Approximate Posterior

- Involves a joint ‘approximate’ posterior distribution
  \[ \pi(\theta, x|y, \epsilon) \propto g(y|x, \epsilon)f(x|\theta)\pi(\theta) \]

- \( g(y|x, \epsilon) \) is a weighting function (Reeves and Pettitt, '05).

- How do we compare \( y \) and \( x \)? Directly? But too high dimensional?

- Typically use statistics to summarise the data
  \( S(\cdot) = S_1(\cdot), \ldots, S_p(\cdot) \) (low dimensional)
  \[ \rho(y, x) = \|S(y) - S(x)\| \]

- One popular choice is to set
  \[ g(y|x, \epsilon) = 1(\rho(y, x) \leq \epsilon) \]

- Choice of \( \epsilon \) trade-off between accuracy and efficiency
The Approximation and Summary Statistic Choice

- Errors from insufficient summaries and $\epsilon > 0$ (see Fearnhead and Prangle (2012))
- Efficient algorithm gets low tolerance
- Thus summary statistic choice is vital for good approximation

Approaches for Summary Statistics

- Full data (Barthelme and Chopin (2011), White et al (2012))
- Data reduction techniques (Blum et al (2012))
- Choose subset out of larger set (Nunes and Balding (2010), Barnes et al (2012))
- Indirect inference (motivated by application in this talk)

Advantage of indirect inference approach: knowledge that summary statistics are ‘good’ can be obtained prior to ABC analysis
ABC Algorithms

  - Advantage: Simplicity, Independent draws
  - Disadvantage: Acceptance rate too low.

- **MCMC ABC** (Marjoram et al, 2003)
  - Carefully chosen proposal $q(\theta^*, x^* | \theta, x) = q(\theta^* | \theta) f(x^* | \theta^*)$.
  - Likelihoods Cancel
  - Advantage: Efficient relative to Rejection Sampling
  - Disadvantage: Still quite inefficient, can get stuck, multi-modal targets?


- **SMC Advantages**: Very efficient, adaptive
Motivating Application - Macroparasite Immunity

- Estimate parameters of a Markov process model explaining macroparasite population development with host immunity
- 212 hosts (cats) \(i = 1, \ldots, 212\). Each cat injected with \(l_i\) juvenile \textit{Brugia pahangi} larvae (approximately 100 or 200).
- At time \(t_i\) host is sacrificed and the number of matures are recorded
- Host assumed to develop an immunity
- Three variable problem: \(M(t)\) matures, \(L(t)\) juveniles, \(I(t)\) immunity.
- Only \(L(0)\) and \(M(t_i)\) is observed for each host
Proportion of Matures vs Time
The Model and Intractable Likelihood

- Deterministic form of the model

\[
\frac{dL}{dt} = -\mu_L L - \beta IL - \gamma L, \\
\frac{dM}{dt} = \gamma L - \mu_M M, \\
\frac{dI}{dt} = \nu L - \mu_I I,
\]

- $\mu_m, \gamma$ fixed. $\nu, \mu_L, \mu_I, \beta$ require estimation

- Likelihood-based inference appears intractable.
  - Could form complete likelihood with missing larvae and immunity (too much missing data)
  - Matrix exponential form of the likelihood (matrix too large)

- Simulation from the model straightforward using Gillespie’s algorithm (Gillespie, 1977)
Developing Summary Statistics

- Need to develop **summary statistics** that efficiently summarize data!
- **Covariates**: numbers of juveniles, sacrifice time.
- An approach based on **indirect inference** (Gouriéroux and Ronchetti, 1993)
  - Propose an auxiliary model $p_a(y|\theta_a)$ where parameter $\theta_a$ is easily estimable
  - Auxiliary model is flexible enough to provide a good description of the data
  - Simulate $x_\theta$ from target intractable likelihood $p(\cdot|\theta)$ and find $\hat{\theta}_a(x_\theta)$
  - Estimate $\theta$ using $\hat{\theta}_a(x_\theta)$ closest to $\hat{\theta}_a(y)$
- Estimates of parameters of the auxiliary model fitted to the data become the summary statistics in ABC.
- Compare and contrast models based on Beta Binomial (BB) and Binomial mixture (BM).
Summary statistics

- For each model
- Compare the auxiliary estimates for simulated data $x$, $\theta^x_a$, and observations $y$, $\hat{\theta}^y_a$, with the Mahalanobis distance

$$
\rho(y, x) = \rho(\hat{\theta}^y_a, \theta^x_a) = \sqrt{(\hat{\theta}^y_a - \theta^x_a)^T S^{-1}(\hat{\theta}^y_a - \theta^x_a)},
$$

- where $S$ is the covariance matrix for the MLE
Auxiliary Beta-Binomial model

- The data show too much variation for Binomial
- A Beta-Binomial model has an extra parameter to capture dispersion

\[ p(m_i|\alpha_i, \beta_i) = \binom{l_i}{m_i} \frac{B(m_i + \alpha_i, l_i - m_i + \beta_i)}{B(\alpha_i, \beta_i)} \]

- Useful reparameterisation \( p_i = \alpha_i / (\alpha_i + \beta_i) \) and \( \theta_i = 1 / (\alpha_i + \beta_i) \)
- Relate the proportion and over dispersion parameters to time, \( t_i \), and initial larvae, \( l_i \), covariates

\[ \logit(p_i) = \beta_0 + \beta_1 \log(t_i) + \beta_2(\log(t_i))^2, \]

\[ \log(\theta_i) = \begin{cases} \eta_{100}, & \text{if } l_i \approx 100 \\ \eta_{200}, & \text{if } l_i \approx 200 \end{cases} \]

- Five parameters
An auxiliary model based on a three component Binomial mixture.

The $i$th observation has the distribution

$$p(m_i|\Theta) = \binom{l_i}{m_i} \sum_{k=1}^{3} w_k(\theta_i^k)^{m_i}(1 - \theta_i^k)^{l_i-m_i},$$

where $w_3 = 1 - w_1 - w_2$.

Reparameterise the $\theta_i^k$, $\text{logit}(\theta_i^k) = \gamma_0^k + \gamma_1 \log(t_i)$, so that each component has the same slope but a different intercept.

Six parameters, $\Theta = (w_1, w_2, \gamma_0^1, \gamma_0^2, \gamma_0^3, \gamma_1)$. 
How to choose between different auxiliary models?
- use data analytical tools for the original data set
- Try each model/summary statistic (Nunes and Balding, 2010)
- Former is far less computer intensive but does it give best ABC approximation?

Fit to data
- Relative measure: Beta-Binomial better than Binomial mixture by 170 on log likelihood scale and 1 less parameter
- Absolute measure: Generalized Pearson statistics
  - Beta-Binomial, 213.2 on 207 d.o.f.
  - Binomial mixture, 278.9 on 206 d.o.f.
- Plots suggest Binomial mixture does not capture variability of data
- Residual analysis: Beta-binomial ✓
SMC Algorithm Settings and Posterior Results

Algorithm Settings

- Take $N = 1000$ particles, start with 60% acceptance, discard half each iteration, finish with 3% acceptance for MCMC kernel.
- Repeat MCMC kernel to get about 99% acceptance at each iteration.
- Fixed values $\gamma = 0.04$, $\mu_M = 0.0015$
- Prior choices:
  - $\nu$: $U(0,1)$
  - $\mu_L$: $U(0,1)$
  - $\mu_I$: $U(0,2)$
  - $\beta$: $U(0,2)$
Posterior Densities

Postiors Beta Binomial vs Binomial mixture
Bivariate Posteriors

Posterior Beta Binomial vs Binomial mixture

Plots of posteriors, green is BM, red is BI

Chris Drovandi  ISBA 2012
For $\nu$ BB posterior is more concentrated than BM, similar modes.

For $\mu_L$ similar concentration but very different modes, 0.0055 vs 0.0244.

Using the Nunes-Balding criterion for summary statistics would prefer BB to BM.

But what are the predictions from each ABC fitted model?
ABC Fit to data

95 percent prediction interval from the stochastic model using ABC modal estimates: with auxiliary model Beta-Binomial (left) and the Binomial mixture (right) model.

ABC model fit based on Beta-Binomial auxiliary model captures variability of data
Summary of Method

Advantages

- Can be applied in non-iid settings
- Appropriate of statistics assessed via data analytic techniques
- Statistics are complicated functions of data

Disadvantages

- Likelihood maximisation for each simulated data drawn
Reference