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

$$\pi(\boldsymbol{\theta}|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}).$$

- What happens when likelihood $f(\mathbf{y}|\boldsymbol{\theta})$ unavailable?
- ABC instead uses model simulations and compares simulated with observed
- Naive Algorithm - Rejection Sampling
  - Sample $\boldsymbol{\theta} \sim \pi(\cdot)$
  - Simulate $\mathbf{x} \sim f(\cdot|\boldsymbol{\theta})$
  - Accept $\boldsymbol{\theta}$ if $\rho(\mathbf{y}, \mathbf{x}) \leq \epsilon$
- Repeat the above until we have N draws, $\boldsymbol{\theta}_\mathbf{1}, \ldots, \boldsymbol{\theta}_\mathbf{N}$

The Approximate Posterior

- Involves a joint 'approximate' posterior distribution

$$\pi(\boldsymbol{\theta}, \mathbf{x}|\mathbf{y}, \epsilon) \propto g(\mathbf{y}|\mathbf{x}, \epsilon) f(\mathbf{x}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta})$$

- $g(\mathbf{y}|\mathbf{x}, \epsilon)$ is a weighting function (Reeves and Pettitt, '05).
- How do we compare $\mathbf{y}$ and $\mathbf{x}$? Directly? But too high dimensional?
- Typically use statistics to summarise the data $\mathbf{S}(\cdot) = S_1(\cdot), \ldots, S_p(\cdot)$ (low dimensional)

$$\rho(\mathbf{y}, \mathbf{x}) = \|\mathbf{S}(\mathbf{y}) - \mathbf{S}(\mathbf{x})\|$$

- One popular choice is to set

$$g(\mathbf{y}|\mathbf{x}, \epsilon) = 1(\rho(\mathbf{y}, \mathbf{x}) \leq \epsilon)$$

- Choice of ϵ 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)

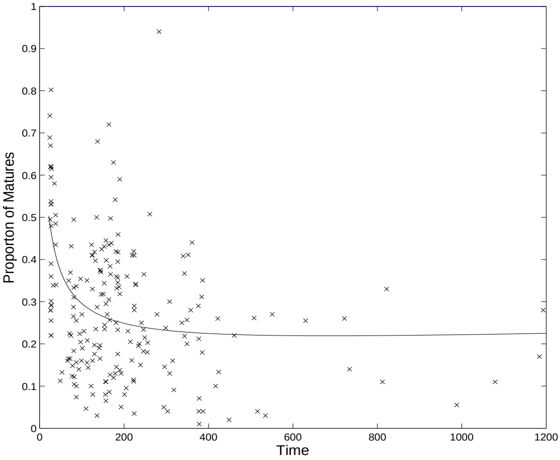
Advantage of indirect inference approach: knowledge that summary statistics are 'good' can be obtained prior to ABC analysis

ABC Algorithms

- Rejection Sampling (above) (Pritchard et al, 1999 and Beaumont et al, 2002)
  - Advantage: Simplicity, Independent draws
  - Disadvantage: Acceptance rate too low.
- MCMC ABC (Marjoram et al, 2003)
  - Carefully chosen proposal $q(\boldsymbol{\theta}^*, \mathbf{x}^*|\boldsymbol{\theta}, \mathbf{x}) = q(\boldsymbol{\theta}^*|\boldsymbol{\theta})f(\mathbf{x}^*|\boldsymbol{\theta}^*)$. Likelihoods Cancel
  - Advantage: Efficient relative to Rejection Sampling
  - Disadvantage: Still quite inefficient, can get stuck, multi-modal targets?
- SMC ABC (Sisson et al (2007), Toni et al (2009), Drovandi and Pettitt (2011), Beaumont et al (2009), Del Moral et al (2012))
- SMC Advantages: Very efficient, adaptive

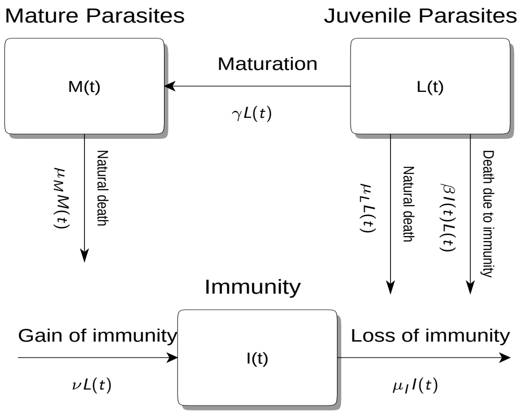
Motivating Application - Macroparasite Immunity

- 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 *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



Proporton of Matures
Time
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
200
400
600
800
1000
1200

Trivariate Markov Process of Riley et al (2003)



The Model and Intractable Likelihood

- Deterministic form of the model

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

- μ_m, γ fixed. ν, μ_L, μ_I, β 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 $p_a(\mathbf{y}|\boldsymbol{\theta}_\mathbf{a})$ where parameter $\boldsymbol{\theta}_\mathbf{a}$ is easily estimable
  - Auxiliary model is flexible enough to provide a good description of the data
  - Simulate $\mathbf{x}_{\boldsymbol{\theta}}$ from target intractable likelihood $p(\cdot|\boldsymbol{\theta})$ and find $\hat{\boldsymbol{\theta}}_\mathbf{a}(\mathbf{x}_{\boldsymbol{\theta}})$
  - Estimate $\boldsymbol{\theta}$ using $\hat{\boldsymbol{\theta}}_\mathbf{a}(\mathbf{x}_{\boldsymbol{\theta}})$ closest to $\hat{\boldsymbol{\theta}}_\mathbf{a}(\mathbf{y})$
- 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, $\boldsymbol{\theta}_a^{\mathbf{x}}$, and observations y, $\hat{\boldsymbol{\theta}}_a^{\mathbf{y}}$, with the Mahalanobis distance

$$\rho(\mathbf{y},\mathbf{x}) = \rho(\hat{\boldsymbol{\theta}}_a^{\mathbf{y}},\boldsymbol{\theta}_a^{\mathbf{x}}) = \sqrt{(\hat{\boldsymbol{\theta}}_a^{\mathbf{y}} - \boldsymbol{\theta}_a^{\mathbf{x}})^T \mathbf{S}^{-1}(\hat{\boldsymbol{\theta}}_a^{\mathbf{y}} - \boldsymbol{\theta}_a^{\mathbf{x}})},$$

- 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

$$\text{logit}(p_i) = \beta_0 + \beta_1\log(t_i) + \beta_2(\log(t_i))^2,$$

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

- Five parameters

Auxiliary Binomial Mixture model

- An auxiliary model based on a three component Binomial mixture.
- The ith observation has the distribution

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

  where $w_3 = 1 - w_1 - w_2$.
- Reparameterise the θ_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, $\mathbf{\Theta} = (w_1, w_2, \gamma_0^1, \gamma_0^2, \gamma_0^3, \gamma_1)$.

ABC Summary Statistics from auxiliary models

- 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 ✓

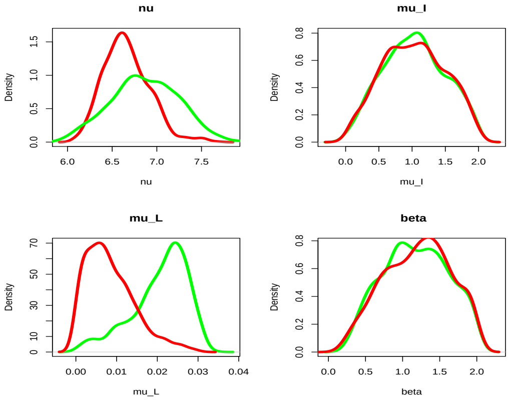
SMC Algorithm Settings and Posterior Results

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:
  - ν: U(0,1)
  - μ_L: U(0,1)
  - μ_I: U(0,2),
  - β: U(0,2),

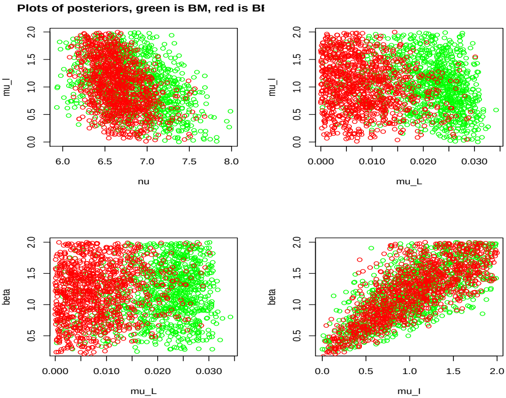
Posterior Densities

Posteriors Beta Binomial vs Binomial mixture



Bivariate Posteriors

Posteriors Beta Binomial vs Binomial mixture

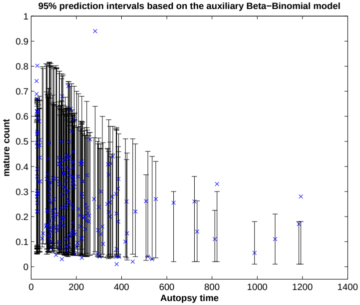


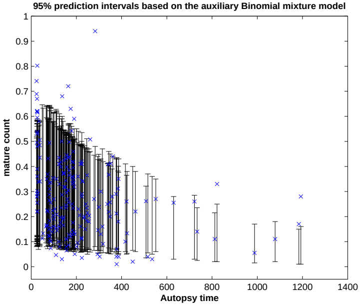
Posterior Density Differences

- For ν BB posterior is more concentrated than BM, similar modes
- For μ_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

- Drovandi et al (2011). JRSS C 60, 317-337.
- Drovandi and Pettitt (2011). Biometrics 67, 225-233.
- Riley et al (2003). J. Theoret. Biol. 225, 419430.
- Heggland and Frigessi (2004). JRSS B 66, 447-462.
- Nunes and Balding (2010). Stat. Appl. Genet. Mol. Biol. 9(1).