

Bayesian Indirect Inference using a Parametric Auxiliary Model

Dr Chris Drovandi
Queensland University of Technology, Australia
c.drovandi@qut.edu.au

Collaborators: Tony Pettitt and Anthony Lee

February 12, 2014



Intro to ABC (intractable likelihood)

- Bayesian inference is based on posterior distribution

$$p(\theta|y) \propto p(y|\theta)p(\theta).$$

- Thus Bayesian inference requires the likelihood function $p(y|\theta)$
- Many models have an **intractable likelihood** function
- One possible avenue is approximate Bayesian computation (**ABC**)

Intro to ABC (ABC target)

- ABC assumes that **simulation** from model is **straightforward**,
 $x \sim p(y|\theta)$
- Compares observed and simulated data on the basis of **summary statistics**, $s(\cdot)$
- Based on the following target distribution

$$p_\epsilon(\theta|y) \propto p(\theta) \int_x p(x|\theta) K_\epsilon(\|s(y) - s(x)\|) dx,$$

where ϵ is ABC tolerance and K is a kernel weighting function

- If $s(\cdot)$ is sufficient and $\epsilon \rightarrow 0$ then $p_\epsilon(\theta|y) \equiv p(\theta|y)$ (does not happen in practice)
- Thus **two sources of error**

Intro to ABC (choosing summary statistics)

- ABC has connections with kernel density estimation (Blum, 2009)
- Choice of summary statistics involve **trade-off between dimensionality and information loss**
- Non-parametric aspect wants summary as low-dimensional as possible
- But decreasing dimension means information loss
- **Choice of summary statistics most crucial to ABC approximation**

Intro to ABC (MCMC ABC)

- A number of algorithms to sample from ABC target (Rejection sampling (Beaumont 2002), MCMC (Marjoram et al 2003) and SMC (e.g. Sisson et al 2007 and Drovandi and Pettitt 2011))
- In this work used **MCMC ABC**. Involves the following steps:
- Propose new parameter $\theta^* \sim q(\cdot|\theta)$.
- Simulate a dataset, $x^* \sim p(y|\theta^*)$
- Accept with probability:

$$\alpha = \frac{p(\theta^*)K_\epsilon(\|s(y) - s(x^*)\|)q(\theta|\theta^*)}{p(\theta)K_\epsilon(\|s(y) - s(x)\|)q(\theta^*|\theta)}$$

- and repeat ...

Introduction

Approximate Bayesian Computation

- ABC becoming standard for dealing with intractable (**generative**) model $p(\mathbf{y}|\theta)$
- Choosing summary statistics most difficult aspect

Bayesian Indirect Inference (BII)

- Propose another (tractable) **auxiliary** model $p_A(\mathbf{y}|\phi)$
- Use the auxiliary model to build summary statistics or replace likelihood
- **Aim**: Review and Compare

BII Methods

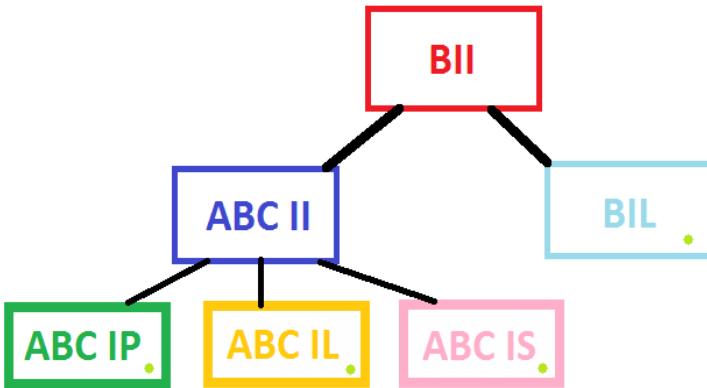


Figure: BII Methods

ABC II Methods

ABC target distribution

$$p_{\epsilon,n}(\theta|\mathbf{y}) \propto p(\theta) \int_{\mathbf{x}_n} p(\mathbf{x}_n|\theta) 1(\rho(s(\mathbf{x}_n), s(\mathbf{y})) \leq \epsilon) d\mathbf{x}_n,$$

where \mathbf{y} is observed data (assume N observations), \mathbf{x}_n is simulated data (of size nN) θ is parameter, $p(\theta)$ is prior, $p(\mathbf{x}_n|\theta)$ is likelihood, $s(\cdot)$ is summary statistic, $\rho(\cdot, \cdot)$ is discrepancy function

- Common approach in II is to take $n > 1$
- Can be shown that ABC can be over-precise when $n > 1$
- Therefore take $n = 1$ for ABC II

ABC IP (Drovandi et al 2011)

- **Parameter** estimates of auxiliary model are summary statistics
- Simulate data from generative model $\mathbf{x} \sim p(\cdot|\boldsymbol{\theta})$
- Estimate auxiliary parameter

$$\hat{\phi}(\boldsymbol{\theta}, \mathbf{x}) = \arg \max_{\phi} p_A(\mathbf{x}|\phi).$$

- Compare with $\hat{\phi}(\mathbf{y})$

$$\rho(s(\mathbf{x}), s(\mathbf{y})) = \sqrt{(\hat{\phi}(\mathbf{x}) - \hat{\phi}(\mathbf{y}))^T I(\hat{\phi}(\mathbf{y})) (\hat{\phi}(\mathbf{x}) - \hat{\phi}(\mathbf{y}))}.$$

- Efficient weighting of summary statistics using observed Fisher information $I(\hat{\phi}(\mathbf{y}))$

ABC IL (Gleim and Pigorsch 2013)

- Same as ABC IP but uses **likelihood** in the discrepancy function
- Parameter estimates of auxiliary model are summary statistics
- Simulate data from generative model $\mathbf{x} \sim p(\cdot|\boldsymbol{\theta})$
- Estimate auxiliary parameter

$$\hat{\phi}(\boldsymbol{\theta}, \mathbf{x}) = \arg \max_{\phi} p_A(\mathbf{x}|\phi).$$

- Compare with $\hat{\phi}(\mathbf{y})$

$$\rho(s(\mathbf{x}), s(\mathbf{y})) = \log p_A(\mathbf{y}|\hat{\phi}(\mathbf{y})) - \log p_A(\mathbf{y}|\hat{\phi}(\mathbf{x})).$$

ABC IS (Gleim and Pigorsch 2013)

- Uses **scores** of auxiliary model as summary statistics

$$\mathbf{S}_A(\mathbf{y}, \phi) = \left(\frac{\partial \log p_A(\mathbf{y}|\phi)}{\partial \phi_1}, \dots, \frac{\partial \log p_A(\mathbf{y}|\phi)}{\partial \phi_{\dim(\phi)}} \right)^T,$$

- Simulate data from generative model $\mathbf{x} \sim p(\cdot|\theta)$
- Evaluate scores of auxiliary model based on simulated data and $\hat{\phi}(\mathbf{y})$, $\mathbf{S}_A(\mathbf{x}, \hat{\phi}(\mathbf{y}))$
- Note that $\mathbf{S}_A(\mathbf{y}, \hat{\phi}(\mathbf{y})) = \mathbf{0}$. Uses following discrepancy function

$$\rho(s(\mathbf{x}), s(\mathbf{y})) = \sqrt{\mathbf{S}_A(\mathbf{x}, \hat{\phi}(\mathbf{y}))^T \mathbf{I}(\hat{\phi}(\mathbf{y}))^{-1} \mathbf{S}_A(\mathbf{x}, \hat{\phi}(\mathbf{y}))}.$$

- Very fast when scores are analytic (no fitting of auxiliary model)

ABC II Assumptions

Assumption (ABC IP Assumptions)

The estimator of the auxiliary parameter, $\hat{\phi}(\theta, \mathbf{x})$, is unique for all θ with positive prior support.

Assumption (ABC IL Assumptions)

The auxiliary likelihood evaluated at the auxiliary estimate, $p_A(\mathbf{y}|\hat{\phi}(\mathbf{x}, \theta))$, is unique for all θ with positive prior support.

Assumption (ABC IS Assumptions)

The MLE of the auxiliary model fitted to the observed data, $\hat{\phi}(\mathbf{y})$, is an interior point of the parameter space of ϕ . The log-likelihood of the auxiliary model, $\log p_A(\cdot|\phi)$, is differentiable and the score, $\mathbf{S}_A(\mathbf{x}, \hat{\phi}(\mathbf{y}))$, is unique for \mathbf{x} that may be drawn from any θ that has positive prior support.

BIL (Gallant and McCulloch 2009, Reeves and Pettitt 2005)

- Replaces true likelihood with auxiliary likelihood
- Simulate data from generative model $\mathbf{x}_n \sim p(\cdot|\boldsymbol{\theta})$
- Estimate auxiliary parameter

$$\hat{\phi}(\boldsymbol{\theta}, \mathbf{x}_n) = \arg \max_{\phi} p_A(\mathbf{x}_n|\phi).$$

- Evaluate auxiliary likelihood $p_A(\mathbf{y}|\hat{\phi}(\boldsymbol{\theta}, \mathbf{x}_n))$
- Not ABC. No summary statistics or ABC tolerance
- Has the following target distribution

$$p_n(\boldsymbol{\theta}|\mathbf{y}) \propto p(\boldsymbol{\theta}) \int_{\mathbf{x}_n} p_A(\mathbf{y}|\hat{\phi}(\boldsymbol{\theta}, \mathbf{x}_n))p(\mathbf{x}_n|\boldsymbol{\theta})d\mathbf{x}_n,$$

which depends on n .

- Theoretically behaves very different to ABC II

BIL Theory

Assumption (*BIL Assumptions*)

As $n \rightarrow \infty$, $p_A(\mathbf{y}|\hat{\phi}(\mathbf{x}_n, \boldsymbol{\theta}))$ converges in probability to $p_A(\mathbf{y}|\phi(\boldsymbol{\theta}))$ for all $\boldsymbol{\theta}$ with positive prior support.

Result (*BIL target as $n \rightarrow \infty$*)

The target distribution of BIL for $n \rightarrow \infty$ is $p_\infty(\boldsymbol{\theta}|\mathbf{y}) \propto p_A(\mathbf{y}|\phi(\boldsymbol{\theta}))p(\boldsymbol{\theta})$. Thus BIL targets the correct posterior distribution provided that $p_A(\mathbf{y}|\phi(\boldsymbol{\theta})) \propto p(\mathbf{y}|\boldsymbol{\theta})$ as a function of $\boldsymbol{\theta}$.

BII Theory

- What about the target for finite n ?
- p_n and p_∞ have same target provided that $E[p_A(\mathbf{y}|\hat{\phi}(\boldsymbol{\theta}, \mathbf{x}_n))] = p_A(\mathbf{y}|\phi(\boldsymbol{\theta}))$ (Andrieu and Roberts 2009)
- In general $p_A(\mathbf{y}|\hat{\phi}(\boldsymbol{\theta}, \mathbf{x}_n))$ is a biased, noisy estimate of $p_A(\mathbf{y}|\phi(\boldsymbol{\theta}))$
- Can anticipate better approximations for BII for n large
- Contrasts with ABC II which needs $n = 1$

Synthetic Likelihood as a BIL Method

- Reduces data to set of summary statistics. Target distribution:

$$p(\theta|s(\mathbf{y})) \propto p(s(\mathbf{y})|\theta)p(\theta).$$

- But $p(s(\mathbf{y})|\theta)$ is unknown
- Synthetic likelihood approach Wood (2010) assumes a parametric model, $p_A(s(\mathbf{y})|\phi(\theta))$. Thus a BIL method.
- Wood takes p_A to be multivariate normal $N(\mu(\theta), \Sigma(\theta))$
- For some θ , simulate n replicates of size N , $\mathbf{x}_n = (s(\mathbf{x}_1^{(1)}), \dots, s(\mathbf{x}_1^{(n)}))$, iid sample from $p(s(\mathbf{y})|\theta)$. Fit auxiliary model to this dataset:

$$\mu(\mathbf{x}_n, \theta) = \frac{1}{n} \sum_{i=1}^n s(\mathbf{x}_1^{(i)}),$$

$$\Sigma(\mathbf{x}_n, \theta) = \frac{1}{n} \sum_{i=1}^n (s(\mathbf{x}_1^{(i)}) - \mu(\mathbf{x}_n, \theta))(s(\mathbf{x}_1^{(i)}) - \mu(\mathbf{x}_n, \theta))^T.$$

ABC as BIL with Non-Parametric Auxiliary Model

- ABC is recovered as a BIL method by selecting a non-parametric auxiliary likelihood
- Full data (dABC): simulate n datasets $\mathbf{x}_n = (\mathbf{x}_1^{(1)}, \dots, \mathbf{x}_1^{(n)})$ and estimate likelihood

$$\frac{1}{n} \sum_{i=1}^n K_{\epsilon}(\rho(\mathbf{y}, \mathbf{x}_1^{(i)})),$$

- Data reduced to summary statistic (sABC):

$$\frac{1}{n} \sum_{i=1}^n K_{\epsilon}(\rho(\mathbf{s}(\mathbf{y}), \mathbf{s}(\mathbf{x}_1^{(i)}))).$$

- ABC II is a special sABC method where summary statistics come from a parametric auxiliary model



Theoretical comparison of BIL and ABC II

- BIL exact when true model contained within auxiliary model (as $n \rightarrow \infty$). Suggests to choose flexible auxiliary model
- Even when true model is special case, ABC II statistics not sufficient in general
- When ABC II have sufficient statistics, BIL not exact in general, even when $n \rightarrow \infty$
- Some Toy Examples Later...

Sampling from ABC II Target - MCMC ABC

Algorithm 1 MCMC ABC algorithm of Marjoram et al 2003.

- 1: Set θ^0
 - 2: **for** $i = 1$ **to** T **do**
 - 3: Draw $\theta^* \sim q(\cdot | \theta^{i-1})$
 - 4: Simulate $\mathbf{x}^* \sim p(\cdot | \theta^*)$
 - 5: Compute $r = \frac{p(\theta^*)q(\theta^{i-1} | \theta^*)}{p(\theta^{i-1})q(\theta^* | \theta^{i-1})} 1(\rho(s(\mathbf{x}^*), s(\mathbf{y})) \leq \epsilon)$
 - 6: **if** $\text{uniform}(0, 1) < r$ **then**
 - 7: $\theta^i = \theta^*$
 - 8: **else**
 - 9: $\theta^i = \theta^{i-1}$
 - 10: **end if**
 - 11: **end for**
-

Sampling from BIL Target - MCMC BIL

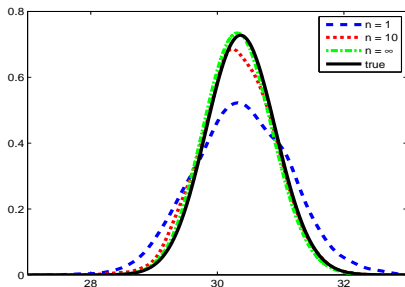
Find increase in acceptance probability with increase in n

Algorithm 1 MCMC BIL algorithm (see also Gallant and McCulloch 2009).

- 1: Set θ^0
 - 2: Simulate $\mathbf{x}_n^* \sim p(\cdot | \theta^0)$
 - 3: Compute $\hat{\phi}^0 = \arg \max_{\phi} p_A(\mathbf{x}_n^* | \phi)$
 - 4: **for** $i = 1$ **to** T **do**
 - 5: Draw $\theta^* \sim q(\cdot | \theta^{i-1})$
 - 6: Simulate $\mathbf{x}_n^* \sim p(\cdot | \theta^*)$
 - 7: Compute $\hat{\phi}(\mathbf{x}_n^*) = \arg \max_{\phi} p_A(\mathbf{x}_n^* | \phi)$
 - 8: Compute $r = \frac{p_A(\mathbf{y} | \hat{\phi}(\mathbf{x}_n^*)) \pi(\theta^*) q(\theta^{i-1} | \theta^*)}{p_A(\mathbf{y} | \hat{\phi}^{i-1}) \pi(\theta^{i-1}) q(\theta^* | \theta^{i-1})}$
 - 9: **if** $\text{uniform}(0, 1) < r$ **then**
 - 10: $\theta^i = \theta^*$
 - 11: $\phi^i = \hat{\phi}(\mathbf{x}_n^*)$
 - 12: **else**
 - 13: $\theta^i = \theta^{i-1}$
 - 14: $\phi^i = \phi^{i-1}$
 - 15: **end if**
 - 16: **end for**
-

Toy Example 1 (Drovandi and Pettitt 2013)

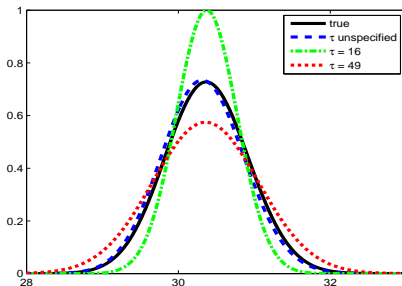
- True model $\text{Poisson}(\lambda)$ and auxiliary model $\text{Normal}(\mu, \tau)$
- ABC II 'gets lucky' as $\hat{\mu} = \bar{y}$, sufficient for λ
- BIL as $n \rightarrow \infty$ approximates $\text{Poisson}(\lambda)$ likelihood with $\text{Normal}(\lambda, \lambda)$ likelihood. Not exact.



Acceptance probabilities: $n = 1$ 46%, $n = 10$ 67%, $n = 100$ 72%
and $n = 1000$ 73% for increasing n

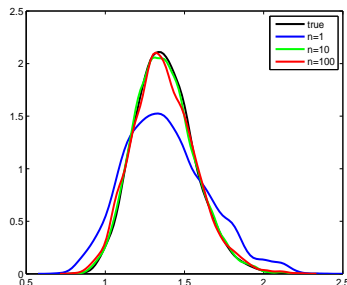
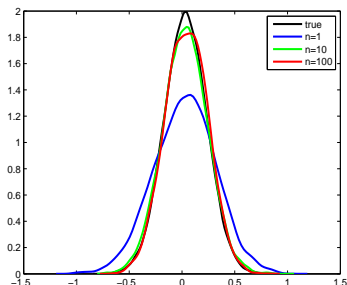
Toy Example 1

- Results for BIL when auxiliary model mis-specified
- Normal($\mu, 9$) (underdispersed) Normal($\mu, 49$) (overdispersed)
- ABC II will still work by chance

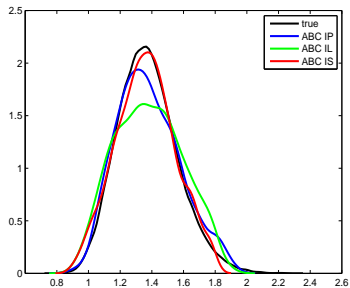
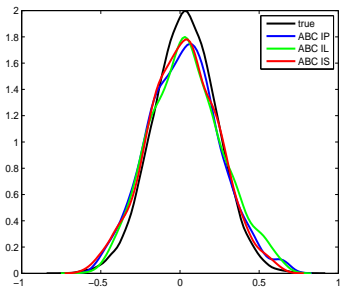


Toy Example 2

- True model t -distribution($\mu, \sigma, \nu = 1$) and auxiliary model t -distribution(μ, σ, ν)
- Here BII exact as $n \rightarrow \infty$ since true is special case of auxiliary
- ABC II do not produce sufficient statistics (full set of order statistics are minimal sufficient)



Toy Example 3 cont



Quantile Distribution Example

- Model of Interest: g -and- k quantile distribution
- Defined in terms of its quantile function:

$$Q(z(p); \theta) = a + b \left(1 + c \frac{1 - \exp(-gz(p))}{1 + \exp(-gz(p))} \right) (1 + z(p)^2)^k z(p). \quad (1)$$

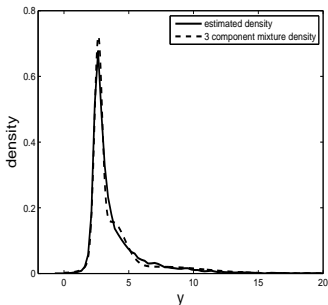
p - quantile, $z(p)$ - standard normal quantile, $\theta = (a, b, g, k)$,
 $c = 0.8$ (see Rayner and MacGillivray 2003).

- Numerical likelihood evaluation possible
- Simulation easier via inversion method

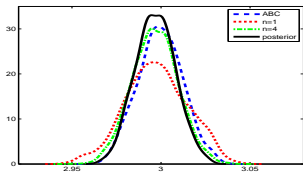
Data consists of 10000 independent draws with $a = 3$, $b = 1$,
 $g = 2$ and $k = 0.5$

Quantile Distribution Example (Cont...)

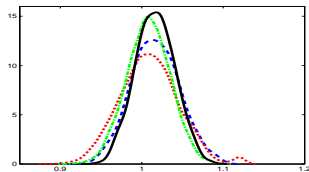
- Auxiliary model is a 3-component normal mixture model
- Flexible and fits data well
- But breaks assumption of ABC IP (parameter estimates not unique)



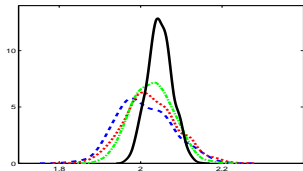
Quantile Distribution Example (Results)



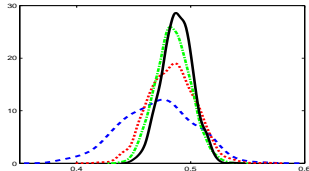
(a) a



(b) b



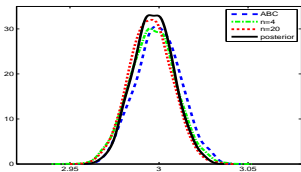
(c) g



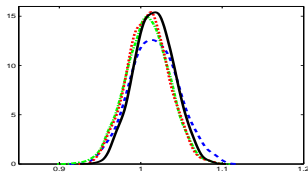
(d) k

Acc Prob: 1.5% for $n = 1$, 2.7% for $n = 2$ and 3.8% for $n = 4$

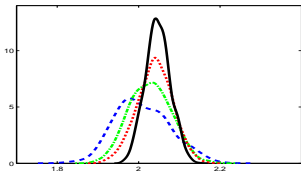
Quantile Distribution Example (Results)



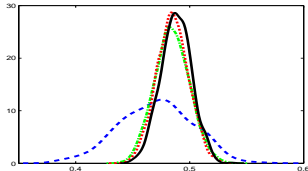
(e) a



(f) b



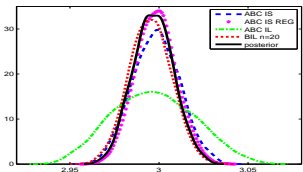
(g) g



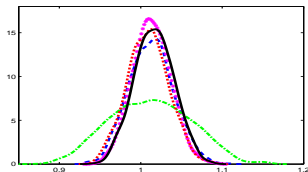
(h) k

MCMC acc prob: 6.5% for $n = 10$ and 8.4% for $n = 20$

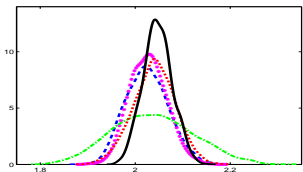
Quantile Distribution Example (Results)



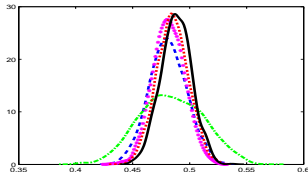
(i) *a*



(j) *b*



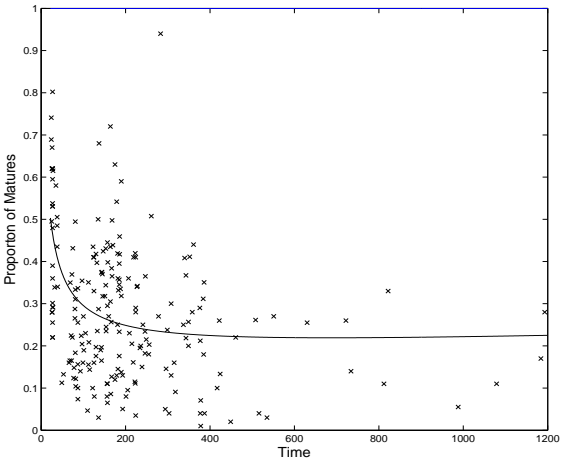
(k) *g*



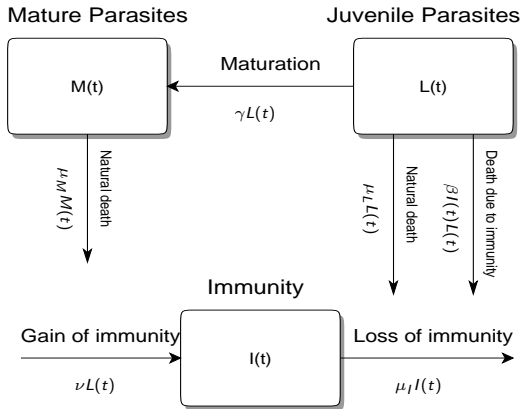
(l) *k*

Macroparasite Immunity Example

- Estimate parameters of a Markov process model explaining **macroparasite population development with host immunity**
- 212 hosts (cats) $i = 1, \dots, 212$. Each cat injected with I_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
- **No tractable likelihood**



Trivariate Markov Process of Riley et al (2003)



Auxiliary Beta-Binomial model

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

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

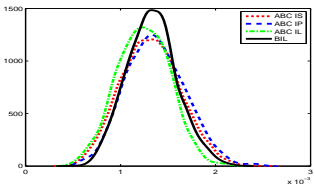
- Useful reparameterisation $p_i = \alpha_i / (\alpha_i + \beta_i)$ and $\theta_i = 1 / (\alpha_i + \beta_i)$
- Relate the proportion and over dispersion parameters to time, t_i , and initial larvae, l_i , covariates

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

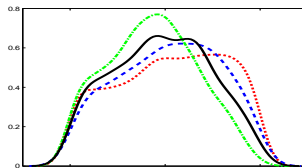
$$\log(\theta_i) = \begin{cases} \eta_{100}, & \text{if } l_i \approx 100 \\ \eta_{200}, & \text{if } l_i \approx 200 \end{cases},$$

- Five parameters $\phi = (\beta_0, \beta_1, \beta_2, \eta_{100}, \eta_{200})$

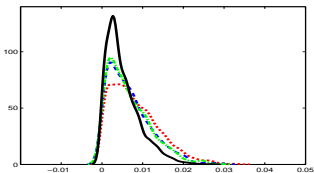
Macroparasite Immunity Results - Compare BII



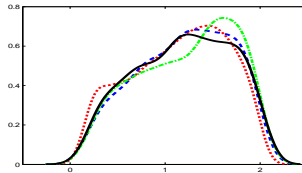
(m) ν



(n) μ_I

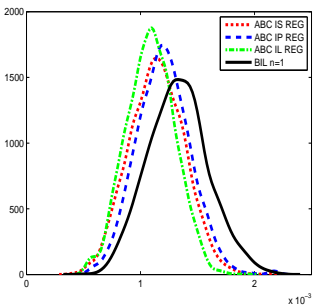


(o) μ_L

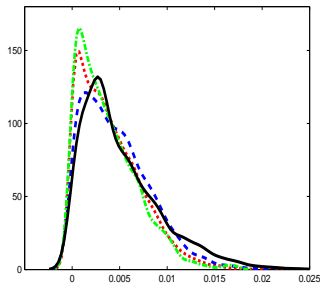


(p) β

Macroparasite Immunity Results - Regression Adjustment

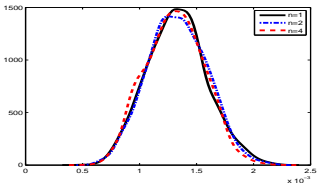


(q) ν

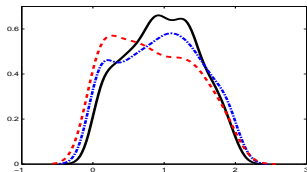


(r) μ_L

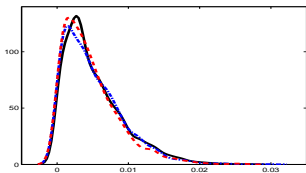
Macroparasite Immunity Results - Increase n



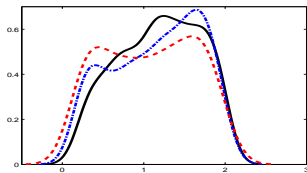
(s) ν



(t) μ_I



(u) μ_L



(v) β

Discussion

- BIL very different to ABC II theoretically
- BIL needs to have a good auxiliary model. ABC II might be useful if auxiliary model mis-specified
- ABC II more flexible; can incorporate other summary statistics

Key References

- Drovandi et al. (2011). Approximate Bayesian Computation using Indirect Inference. JRSS C.
- Drovandi and Pettitt (2013). Bayesian Indirect Inference. <http://eprints.qut.edu.au/63767/>
- Gleim and Pigorsch (2013). Approximate Bayesian Computation with Indirect Summary Statistics.
- Gallant and McCulloch (2009). On the determination of general scientific models with application to asset pricing. JASA.
- Reeves and Pettitt (2005). A theoretical framework for approximate Bayesian computation. 20th IWSM.