Bayesian Indirect Inference

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

Collaborators: Tony Pettitt

January 30, 2014

Intro to ABC (intractable likelihood)

- Bayesian inference in 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 $\epsilon \to 0$ 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 $p(\boldsymbol{y}|\boldsymbol{\theta})$
- Choosing summary statistics most difficult aspect

Bayesian Indirect Inference (BII)

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

BII Methods

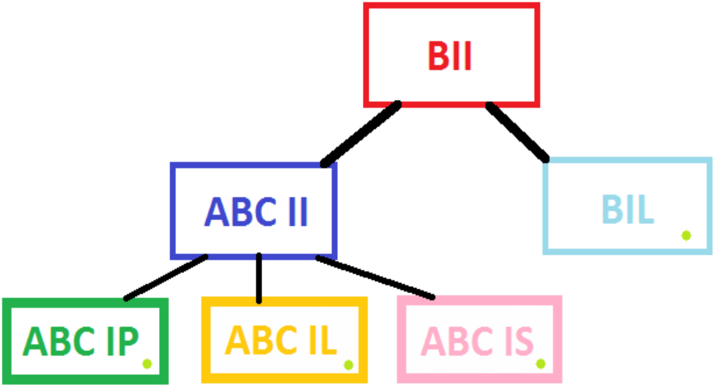


Figure: BII Methods

ABC II Methods

ABC target distribution

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

where $\boldsymbol{y}$ is observed data (assume N observations), $\boldsymbol{x_n}$ is simulated data (of size nN) $\boldsymbol{\theta}$ is parameter, $p(\boldsymbol{\theta})$ is prior, $p(\boldsymbol{x_n}|\boldsymbol{\theta})$ 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 $\boldsymbol{x} \sim p(\cdot|\boldsymbol{\theta})$
- Estimate auxiliary parameter

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

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

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

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

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 $\boldsymbol{x} \sim p(\cdot|\boldsymbol{\theta})$
- Estimate auxiliary parameter

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

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

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

ABC IS (Gleim and Pigorsch 2013)

- Uses scores of auxiliary model as summary statistics

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

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

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

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

ABC II Assumptions

Assumption (*ABC IP Assumptions*)

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

Assumption (*ABC IL Assumptions*)

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

Assumption (*ABC IS Assumptions*)

The MLE of the auxiliary model fitted to the observed data, $\hat{\phi}(\boldsymbol{y})$, 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, $\boldsymbol{S_A}(\boldsymbol{x}, \hat{\phi}(\boldsymbol{y}))$, is unique for $\boldsymbol{x}$ that may be drawn from any $\boldsymbol{\theta}$ that has positive prior support.

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

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

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

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

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

  which depends on n.
- Theoretically behaves very different to ABC II

BIL Theory

Assumption (*BIL Assumptions*)

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

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

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

BIL Theory

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

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.

Set $\boldsymbol{\theta}^{\mathbf{0}}$
for $i = 1$ **to** T **do**
  Draw $\boldsymbol{\theta}^* \sim q(\cdot|\boldsymbol{\theta}^{\boldsymbol{i-1}})$
  Simulate $\boldsymbol{x}^* \sim p(\cdot|\boldsymbol{\theta}^*)$
  Compute $r = \frac{p(\boldsymbol{\theta}^*)q(\boldsymbol{\theta}^{\boldsymbol{i-1}}|\boldsymbol{\theta}^*)}{p(\boldsymbol{\theta}^{\boldsymbol{i-1}})q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{\boldsymbol{i-1}})} 1(\rho(s(\boldsymbol{x}^*), s(\boldsymbol{y})) \leq \epsilon)$
  if uniform$(0,1) < r$ **then**
    $\boldsymbol{\theta}^{\boldsymbol{i}} = \boldsymbol{\theta}^*$
  else
    $\boldsymbol{\theta}^{\boldsymbol{i}} = \boldsymbol{\theta}^{\boldsymbol{i-1}}$
  end if
end for

Sampling from BIL Target - MCMC BIL

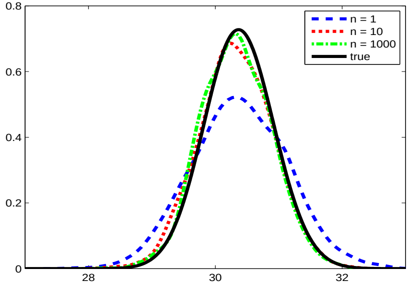
Find increase in acceptance probability with increase in n

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

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

Toy Example 1 (Drovandi and Pettitt 2013)

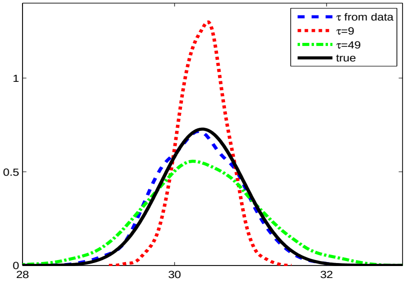
- True model Poisson(λ) and auxiliary model Normal(μ, τ)
- ABC II 'gets lucky' as $\hat{\mu} = \bar{y}$, sufficient for λ
- BIL as $n \rightarrow \infty$ approximates Poisson(λ) likelihood with Normal(λ, λ) likelihood. Not exact.



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

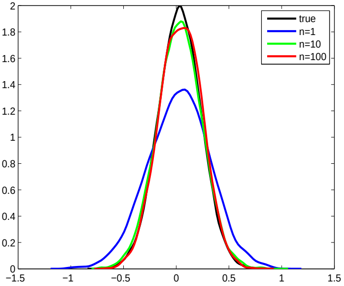
Toy Example 1

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

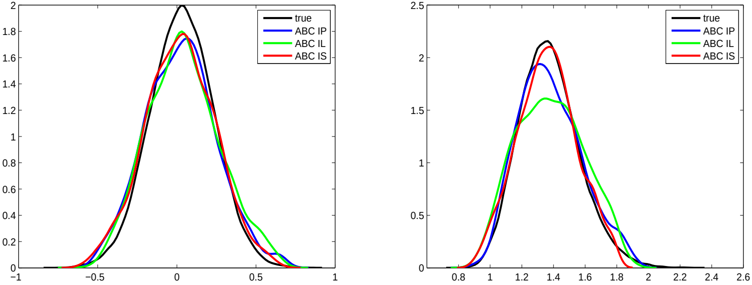


Toy Example 2

- True model t-distribution($\mu, \sigma, \nu = 1$) and auxiliary model t-distribution(μ, σ, ν)
- Here BIL exact as $n \to \infty$ 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:

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

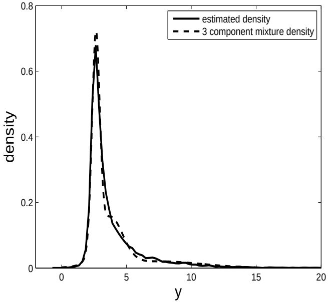
p - quantile, $z(p)$ - standard normal quantile, $\boldsymbol{\theta} = (a, b, g, k)$, $c = 0.8$ (see Rayner and MacGillivray 2003).

- Numerical likelihood evaluation possible
- Simulation easier via inversion method

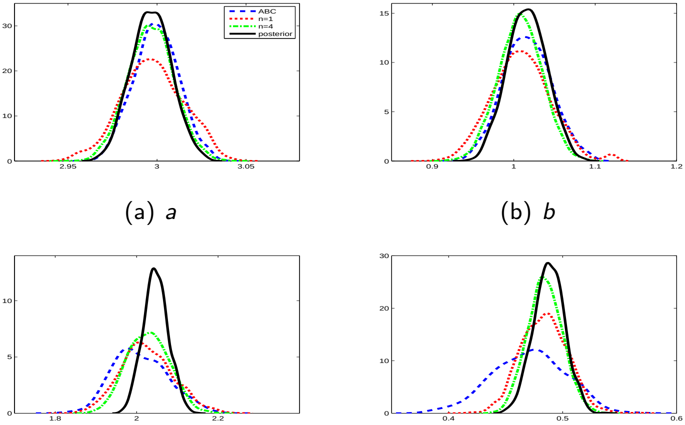
Data consists of 10000 independent draws with $a = 3$, $b = 1$, $g = 2$ and $k = 0.5$

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)

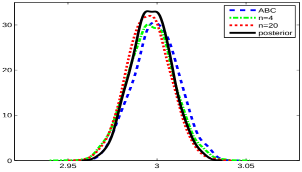


(a) a (b) b

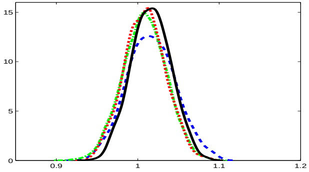
(c) g (d) k

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

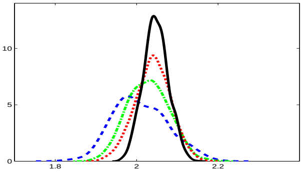
Quantile Distribution Example (Results)



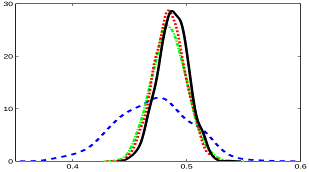
(e) a



(f) b



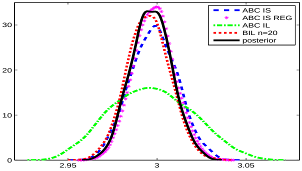
(g) g



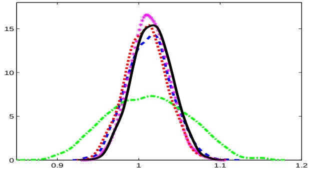
(h) k

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

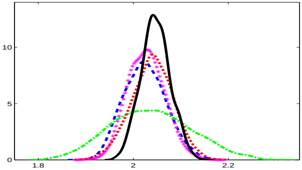
Quantile Distribution Example (Results)



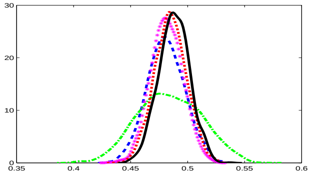
(i) a



(j) b



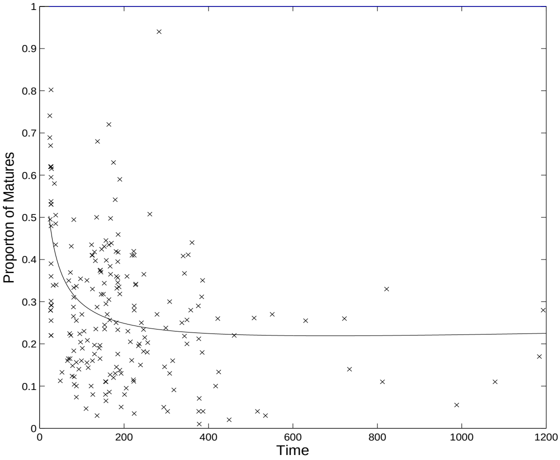
(k) g



(l) k

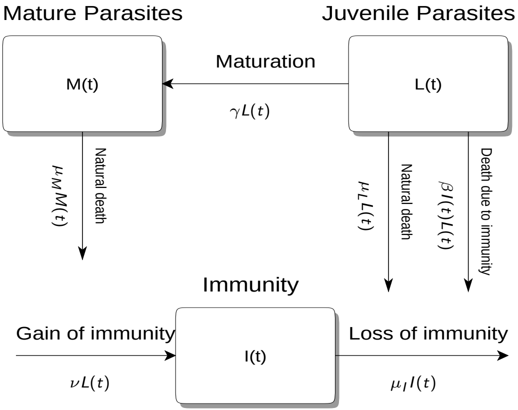
Macroparasite Immunity Example

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



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)



Auxiliary Beta-Binomial model

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

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

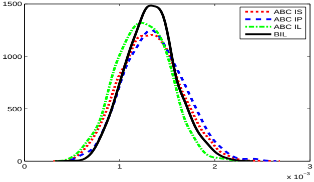
- Useful reparameterisation $p_i = \alpha_i/(\alpha_i+\beta_i)$ and $\theta_i = 1/(\alpha_i+\beta_i)$
- Relate the proportion and over dispersion parameters to time, t_i, and initial larvae, l_i, covariates

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

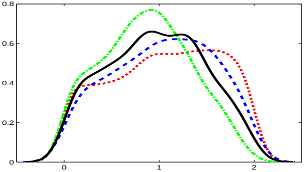
$$\log(\theta_i) = \begin{cases} \eta_{100}, & \text{if } l_i \approx 100 \\ \eta_{200}, & \text{if } l_i \approx 200 \end{cases},$$

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

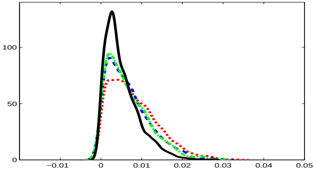
Macroparasite Immunity Results - Compare BII



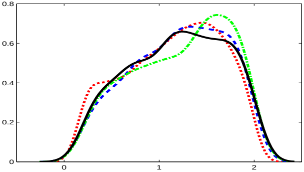
(m) ν



(n) μ_I

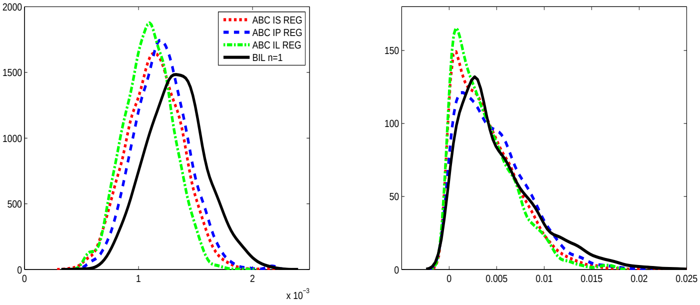


(o) μ_L



(p) β

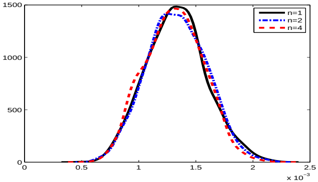
Macroparasite Immunity Results - Regression Adjustment



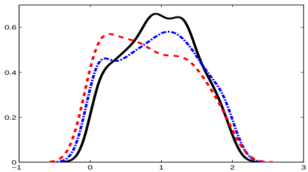
(q) ν

(r) μ_L

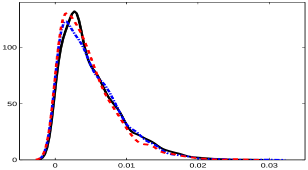
Macroparasite Immunity Results - Increase n



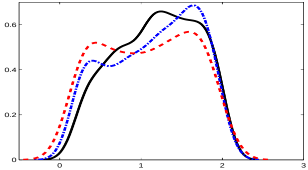
(s) ν



(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 is 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 McCullogh (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.