Bayesian Indirect Inference

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

Collaborators: Tony Pettitt

January 30, 2014
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)
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 $\epsilon$ 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
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
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 Pettit 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 ...
Approximate Bayesian Computation

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

Bayesian Indirect Inference (BII)

- Propose another (tractable) auxiliary model $p_A(y|\phi)$
- 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

\[ p_{\epsilon,n}(\theta|y) \propto p(\theta) \int_{x_n} p(x_n|\theta)1(\rho(s(x_n), s(y)) \leq \epsilon) \, dx_n, \]

where \( y \) is observed data (assume \( N \) observations), \( x_n \) is simulated data (of size \( nN \)), \( \theta \) is parameter, \( p(\theta) \) is prior, \( p(x_n|\theta) \) 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
Parameter estimates of auxiliary model are summary statistics

Simulate data from generative model $\mathbf{x} \sim p(\cdot|\theta)$

Estimate auxiliary parameter

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

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

$$\rho(s(\mathbf{x}), s(\mathbf{y})) = \sqrt{(\hat{\phi}(\mathbf{x}) - \hat{\phi}(\mathbf{y}))^T I(\hat{\phi}(\mathbf{y}))(\hat{\phi}(\mathbf{x}) - \hat{\phi}(\mathbf{y}))}.$$
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 $x \sim p(\cdot | \theta)$
- Estimate auxiliary parameter

$$\hat{\phi}(\theta, x) = \text{arg max}_\phi p_A(x|\phi).$$

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

$$\rho(s(x), s(y)) = \log p_A(y|\hat{\phi}(y)) - \log p_A(y|\hat{\phi}(x)).$$
Uses scores of auxiliary model as summary statistics

\[ S_A(y, \phi) = \left( \frac{\partial \log p_A(y | \phi)}{\partial \phi_1}, \ldots, \frac{\partial \log p_A(y | \phi)}{\partial \phi_{\text{dim}(\phi)}} \right)^T, \]

Simulate data from generative model \( x \sim p(\cdot | \theta) \)

Evaluate scores of auxiliary model based on simulated data and \( \hat{\phi}(y), S_A(x, \hat{\phi}(y)) \)

Note that \( S_A(y, \hat{\phi}(y)) = 0 \). Uses following discrepancy function

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

Very fast when scores are analytic (no fitting of auxiliary model)
ABC II Assumptions

Assumption (ABC IP Assumptions)

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

Assumption (ABC IL Assumptions)

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

Assumption (ABC IS Assumptions)

The MLE of the auxiliary model fitted to the observed data, \( \hat{\phi}(y) \), is an interior point of the parameter space of \( \phi \). The log-likelihood of the auxiliary model, \( \log p_A(\cdot|\phi) \), is differentiable and the score, \( S_A(x, \hat{\phi}(y)) \), is unique for \( x \) that may be drawn from any \( \theta \) that has positive prior support.
BIL (Gallant and McCulloogh 2009, Reeves and Pettitt 2005)

- Replaces true likelihood with auxiliary likelihood
- Simulate data from generative model $x_n \sim p(\cdot|\theta)$
- Estimate auxiliary parameter
  \[
  \hat{\phi}(\theta, x_n) = \arg\max_{\phi} p_A(x_n|\phi).
  \]
- Evaluate auxiliary likelihood $p_A(y|\hat{\phi}(\theta, x_n))$
- Not ABC. No summary statistics or ABC tolerance
- Has the following target distribution
  \[
  p_n(\theta|y) \propto p(\theta) \int_{x_n} p_A(y|\hat{\phi}(\theta, x_n))p(x_n|\theta)dx_n,
  \]
  which depends on $n$.  
- Theoretically behaves very different to ABC II
Assumption (BIL Assumptions)

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

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

The target distribution of BIL for \( n \to \infty \) is
\[
p_\infty(\theta|y) \propto p_A(y|\phi(\theta))p(\theta).
\]
Thus BIL targets the correct posterior distribution provided that \( p_A(y|\phi(\theta)) \propto p(y|\theta) \) as a function of \( \theta \).
What about the target for finite $n$?

- $p_n$ and $p_\infty$ have same target provided that
  \[ E[p_A(y|\hat{\phi}(\theta, x_n))] = p_A(y|\phi(\theta)) \]  
  (Andrieu and Roberts 2009)

- In general $p_A(y|\hat{\phi}(\theta, x_n))$ is a biased, noisy estimate of $p_A(y|\phi(\theta))$

- 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 $n \to \infty$). 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 $n \to \infty$.
- Some Toy Examples Later...
Algorithm 1 MCMC ABC algorithm of Marjoram et al 2003.

1: Set $\theta^0$
2: for $i = 1$ to $T$ do
3: Draw $\theta^* \sim q(\cdot | \theta^{i-1})$
4: Simulate $x^* \sim p(\cdot | \theta^*)$
5: Compute $r = \frac{p(\theta^*)q(\theta^{i-1}|\theta^*)}{p(\theta^{i-1})q(\theta^*|\theta^{i-1})} 1(\rho(s(x^*), s(y)) \leq \epsilon)$
6: if 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 $\theta^0$
2. Simulate $x^*_n \sim p(\cdot | \theta^0)$
3. Compute $\phi^0 = \arg \max_\phi p_A(x^*_n | \phi)$
4. for $i = 1$ to $T$ do
   5. Draw $\theta^* \sim q(\cdot | \theta^{i-1})$
   6. Simulate $x^*_n \sim p(\cdot | \theta^*)$
   7. Compute $\hat{\phi}(x^*_n) = \arg \max_\phi p_A(x^*_n | \phi)$
   8. Compute $r = \frac{p_A(y|\hat{\phi}(x^*_n))\pi(\theta^*)q(\theta^{i-1}|\theta^*)}{p_A(y|\phi^{i-1})\pi(\theta^{i-1})q(\theta^*|\theta^{i-1})}$
   9. if uniform(0, 1) < $r$ then
      10. $\theta^i = \theta^*$
      11. $\phi^i = \hat{\phi}(x^*_n)$
   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 Poisson(\(\lambda\)) and auxiliary model Normal(\(\mu, \tau\))
- ABC II ‘gets lucky’ as \(\hat{\mu} = \bar{y}\), sufficient for \(\lambda\)
- BIL as \(n \to \infty\) approximates Poisson(\(\lambda\)) likelihood with 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 \((\mu, \sigma, \nu)\).
- 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

-1 -0.5 0 0.5 1
0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2
0 0.5 1 1.5 2 2.5 3

Chris Drovandi
University of Reading 2014
Quantile Distribution Example

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

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

(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)
Acc Prob: 1.5% for $n = 1$, 2.7% for $n = 2$ and 3.8% for $n = 4$
Quantile Distribution Example (Results)

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) $k$

ABC IS
ABC IS REG
ABC IL
BIL n=20
posterior

Chris Drovandi
University of Reading 2014
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.
Trivariate Markov Process of Riley et al. (2003)

M(t) \rightarrow \text{Mature Parasites} \quad \text{Juvenile Parasites} \quad L(t)

\gamma L(t) \quad \text{Maturation}

M(t) \downarrow \nu M(t) \quad L(t) \downarrow \nu L(t)

\mu M(t) \quad \beta I(t) L(t) \quad \mu L(t) \quad \mu M(t)

\mu L(t) \downarrow \beta I(t) L(t) \quad \mu L(t) \downarrow \beta I(t) L(t)

I(t) \downarrow \nu L(t) \quad I(t) \downarrow \mu I(t)

\nu L(t) \quad \mu I(t)

Gain of immunity \quad Loss of immunity

Immunity
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 \( p_i = \frac{\alpha_i}{\alpha_i + \beta_i} \) and \( \theta_i = \frac{1}{\alpha_i + \beta_i} \)
- 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 \( \phi = (\beta_0, \beta_1, \beta_2, \eta_{100}, \eta_{200}) \)
Macroparasite Immunity Results - Compare BII

(m) \( \nu \)

(n) \( \mu_I \)

(o) \( \mu_L \)

(p) \( \beta \)
Macroparasite Immunity Results - Regression Adjustment

(q) $\nu$

(r) $\mu_L$
Macroparasite Immunity Results - Increase $n$

(s) $\nu$

(t) $\mu_I$

(u) $\mu_L$

(v) $\beta$
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.
- 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.