Approximate Bayesian Computation

Chris Drovandi, Charisse Farr

October 24, 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?
  - can't use analytic Bayesian inference, Bayesian sampling schemes or Variational Bayes techniques
  - these techniques all require knowledge of, or evaluation of, the likelihood function

- Approximate Bayesian Computation (ABC) techniques provide a means for drawing samples from a (approximate) posterior distribution without evaluating the likelihood function.
  - computer simulation models simulate the measurements that could be received for a given set of parameters
- Potential samples from the posterior distribution are proposed. However, when accepting, rejecting and weighting these samples, rather than calculating the likelihood function, the algorithms compare simulated data with the actual measured data.
- ABC uses model simulations and compares simulated with observed

ABC Samplers

- three of the most common forms of ABC samplers are
  - ABC Rejection Sampling
  - ABC MCMC Sampling
  - ABC Population Monte Carlo Sampling

ABC Rejection Sampling Algorithm I

- Sample $\boldsymbol{\theta} \sim \pi(\cdot)$
    - sample a point $\boldsymbol{\theta}$ from the prior $\pi(\cdot)$
- Simulate $\mathbf{x} \sim f(\cdot|\boldsymbol{\theta})$
    - simulate $\mathbf{x}$ from the measurement model using the sampled parameters $\mathbf{x}$
- Accept $\boldsymbol{\theta}$ if $\rho(\mathbf{y}, \mathbf{x}) \leq \epsilon$
    - accept $\boldsymbol{\theta}$ if the distance between the actual measurement and the simulated measurement $\rho(\mathbf{y}, \mathbf{x})$ is less than or equal to ϵ, the ABC distance tolerance
- Repeat the above until we have N draws, $\boldsymbol{\theta}_\mathbf{1}, \ldots, \boldsymbol{\theta}_\mathbf{N}$

The Choice of ϵ

- The quality of the approximation increases as $\epsilon > 0$ decreases
  - using a very small value of ϵ with continuous data is likely lead to a large number of samples being rejected, making the procedure computationally expensive
- The choice of ϵ is a trade-off between accuracy of the approximation, and the computational effort.

Summary Statistic

- To reduce the computational expense of the procedure, we use instead statistics to summarise the data, rather than the full data. So $\mathbf{S}(\cdot) = S_1(\cdot), \ldots, S_p(\cdot)$ (low dimensional)

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

- The effect of the approximation
  - Errors from insufficient summaries and $\epsilon > 0$

ABC Algorithms

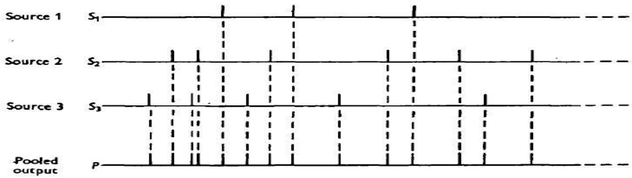
- Advantages
  - Simplicity
  - Independent draws
  - Easy to implement
- Disadvantages
  - Acceptance rate too low
  - High rejection rate if the prior and posterior distributions are not similar

Poisson Distribution Example

- Poisson Distribution
  - True data: $\lambda = 2, n = 1000$
- Investigate three sets of summary statistics
  - sample mean (minimial sufficient)
  - sample variance (insufficient, low dimensional)
  - order statistics (sufficient, high dimensional)
- Prior Simulation
  - T=10^4
  - Gamma prior
  -
  - ϵ selected by keeping best 100 simulations

Gamma Renewal Process Example

- N independent stochastic process with Gamma inter-arrival times
- First observation $Y_1 = \min(X_1, ..., X_N)$ where X_i are iid from Gamma(α,β).
- Take the process that has Y_1. Simulate its next time. Y_2 is the new minimum out of the N values. Repeat...



Intractable Likelihood

Gamma Renewal Process Example (continued...)

- Data: 100 observations with $\alpha = 0.3$, $\beta = 0.5$ and $N = 5$
- Priors: $\alpha \sim G(1.3, 1)$, $\beta \sim G(1.5, 1)$ and N discrete uniform $N \in \{1, 2, \ldots, 20\}$
- Compare 2 sets of summaries
  - All data (sufficient but high-dimensional)
  - Sample mean and variance of sqrt of diff of data (insufficient but low dimensional) (why?)
- Algorithm Settings:
  - 10^5 prior simulations
  - Keep the best 1000

Gamma Renewal Process Example (Results)

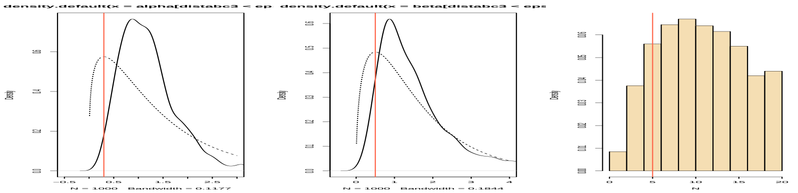


Figure : Summary Stats: All Data

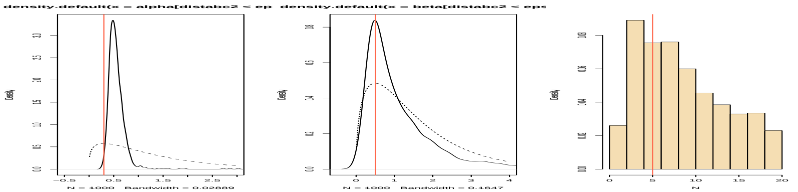


Figure : Summary Stats: sample mean and variance of sqrt of differenced

Closing Remarks

- Rejection sampling no good for vague priors
- Can embed ABC within MCMC (Marjoram et al (2003)) or SMC (Sisson et al 2007, Drovandi+Pettitt 2011) to reduce tolerances
- Biggest issue is (automatic) selection of summary statistics
  - Choose best subset out of large collection of summaries (e.g. Nunes+Balding 2010)
  - Use (estimates) of posterior means as summaries (Fearnhead+Prangle 2012)
  - Indirect Inference (Drovandi et al 2011)