POLITECNICO
MILANO 1863

How approximate is ABC?

Tutor: Dott. Mario Beraha

November 20th, 2020

Marika Di Marcantonio,
Francesca Pietrobon,
Raffaele Saviello

Complexity in Bayesian Inference

In Bayesian inference, knowledge about θ is contained in the posterior distribution and prior beliefs are updated by observing data y_{obs} through the likelihood function $p(y_{obs}|\theta)$ of the model.

$$\pi(\theta|y_{obs}) = \frac{p(y_{obs}|\theta)\pi(\theta)}{\int_{\Theta} p(y_{obs}|\theta)\pi(\theta)d\theta}$$

Numerical methods are needed to proceed
when the posterior distribution isn't available in closed form

Numerical methods

- **Monte Carlo approach** (e.g. MCMC Metropolis-Hastings)

  $\theta^{(1)}, \theta^{(2)}, \dots \sim \pi(\theta|y_{obs})$ s.t. $\pi(\theta|y_{obs}) \approx \frac{1}{N}\sum_{i=1}^{N} \delta_{\theta^{(i)}}(\theta)$

  1. Numerical evaluation of the likelihood function could be computationally prohibitive
  2. The need to repeatedly evaluate the posterior distribution to draw samples makes the implementation of simulation techniques impractical

- **Approximate Bayesian analysis**

  Approximation to the model at the expense of some error

ABC

Likelihood-Free Method

Likelihood-based analysis that proceeds **without direct numerical evaluation** of the likelihood function.

Approximate Bayesian Computation

Particular case of *likelihood-free* method that produces an **approximation to the posterior distribution** resulting from the imperfect matching of data $\|y - y_{obs}\|$ or summary statistics $\|s - s_{obs}\|$

ABC has emerged as an effective and intuitively way of performing an approximate Bayesian analysis

Standard Rejection Sampling Algorithm

- **Inputs**
  - $N > 0$ integer
  - $\pi(\theta|y_{obs}) \propto p(y_{obs}|\theta)\pi(\theta)$ target posterior density
  - $g(\theta)$ proposal density
- **Sampling**

  For i=1:N

  1. Generate $\theta^{(i)} \sim g(\theta)$
  2. Accept $\theta^{(i)}$ with probability $\frac{\pi(\theta^{(i)}|y_{obs})}{Kg(\theta^{(i)})} \propto p(y_{obs}|\theta^{(i)})$ otherwise go to 1

  where $K \geq max_{\theta} \frac{\pi(\theta|y_{obs})}{g(\theta)}$

Likelihood-Free Rejection Sampling Algorithm

- **Inputs**
  - $N > 0$ integer
  - $\pi(\theta|y_{obs}) \propto p(y_{obs}|\theta)\pi(\theta)$ target posterior density
  - $g(\theta)$ proposal density
- **Sampling**

  For i=1:N

  1. Generate $\theta^{(i)} \sim g(\theta)$
  2. Generate $y \sim p(y|\theta^{(i)})$
  3. If $y = y_{obs}$ then accept $\theta^{(i)}$ with probability $\frac{\pi(\theta^{(i)})}{Kg(\theta^{(i)})}$ otherwise go to 1

  where $K \geq max_\theta \frac{\pi(\theta)}{g(\theta)}$

Likelihood-Free Rejection Sampling Algorithm (2)

- **Inputs**
  - $N > 0$ integer
  - $\pi(\theta|y_{obs}) \propto p(y_{obs}|\theta)\pi(\theta)$ target posterior density
  - $g(\theta)$ proposal density

- **Sampling**

  For i=1:N

  1. Generate $\theta^{(i)} \sim g(\theta)$
  2. Generate $y \sim p(y|\theta^{(i)})$
  3. If $\|y - y_{obs}\| \leq h$ then accept $\theta^{(i)}$ with probability $\frac{\pi(\theta^{(i)})}{Kg(\theta^{(i)})}$ otherwise go to 1

  where $K \geq max_\theta \frac{\pi(\theta)}{g(\theta)}$

ABC Rejection Sampling Algorithm

- **Inputs**
  - $N > 0$ integer
  - $\pi(\theta|y_{obs}) \propto p(y_{obs}|\theta)\pi(\theta)$ target posterior density
  - $g(\theta)$ sampling density
  - $K_h(u)$ kernel function and $h > 0$ scale parameter
- **Sampling**

  For i=1:N
  1. Generate $\theta^{(i)} \sim g(\theta)$
  2. Generate $y \sim p(y|\theta^{(i)})$
  3. Accept $\theta^{(i)}$ with probability $\frac{K_h(\|y-y_{obs}\|)\pi(\theta^{(i)})}{Kg(\theta^{(i)})}$ otherwise go to 1

  where $K \geq K_h(0) max_\theta \frac{\pi(\theta)}{g(\theta)}$

ABC Rejection Sampling Algorithm (2)

- **Inputs**
  - $N > 0$ integer
  - $\pi(\theta|y_{obs}) \propto p(y_{obs}|\theta)\pi(\theta)$ target posterior density
  - $g(\theta)$ sampling density
  - $K_h(u)$ kernel function and $h > 0$ scale parameter
  - $s = S(y)$ summary statistics

- **Sampling**

  For i=1:N

  1. Generate $\theta^{(i)} \sim g(\theta)$
  2. Generate $y \sim p(y|\theta^{(i)})$
  3. Compute summary statistic $s = S(y)$
  4. Accept $\theta^{(i)}$ with probability $\frac{K_h(\|s-s_{obs}\|)\pi(\theta^{(i)})}{Kg(\theta^{(i)})}$ otherwise go to 1

  where $K \geq K_h(0)max_\theta \frac{\pi(\theta)}{g(\theta)}$

Example: ABC with summary statistics

- **Model**
  $y = (y_1, ..., y_n)^T$, with $y_i \sim Poisson(\lambda), \lambda \sim Gamma(\alpha, \beta)$, and $\lambda | y \sim Gamma(\alpha + n\bar{y}, \beta + n)$
  .
- **Example Hypotesis**
  - $\alpha = \beta = 1$;
  - $\bar{y}$ sufficient for the model;
  - $K_h(u)$ uniform kernel in $[-h, h]$, with $||u||$ euclidean distance;
  - s_{obs} as $\{\bar{y}, v, (\bar{y}, v)^T\}$ with v sample standard deviation;
  - $y_{obs} = (0, 0, 0, 0, 5)$, $(\bar{y}_{obs}, v^2_{obs}) = (1, 5)$ (N.B. inconsistent with the theoretical Poisson model)
- **Aim**
  Show how well $\pi(\lambda | s_{obs})$ can approximate $\pi(\lambda | y_{obs})$.

Example: Posterior approximation with $h = 0$

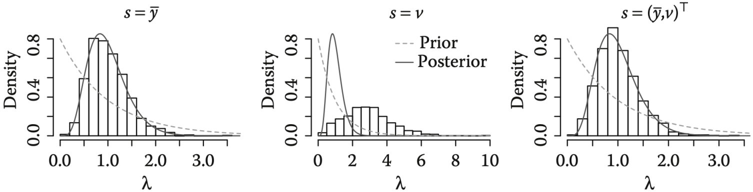


Figure: Chapman & Hall/CRC, "Handbook of Approximate Bayesian Computation", pg.29.

In this particular scenario we obtain:

1. $\bar{y}$ is sufficient, so $\pi(\lambda|\bar{y}_{obs}) = \pi(\lambda|y_{obs})$;
2. v_{obs} is consistent with a Poisson model with $\bar{y}$ larger than the one observed, so there's a shift in the mean;
3. $\bar{y}$ is sufficient, so $\pi(\lambda|\bar{y}_{obs}, v_{obs}) = \pi(\lambda|\bar{y}_{obs})$;

Example: Posterior Approximation with $h = 0.3$

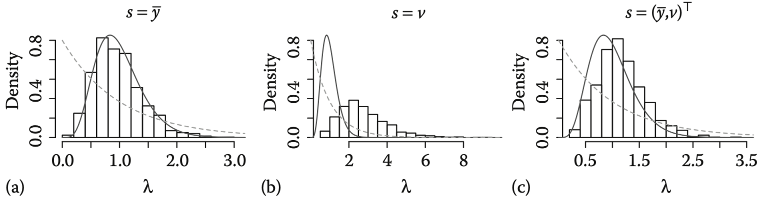


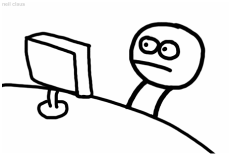
Figure: Chapman & Hall/CRC, "Handbook of Approximate Bayesian Computation", pg.29

1. No significant variation for the first two results;
2. Shift to the right in the approximated posterior given by the inconsistency of $(\bar{y}_{obs}, v_{obs})$ with the Poisson model

Example: Conclusions

- Either in an exact ($h = 0$) or in an approximated ($h > 0$) scenario, the use of a sufficient statistics can lead to a very accurate approximation of the posterior distribution;
- Include progressively more summary statistics into s_{obs} until the ABC posterior approximation does not change appreciably will not provide the most accurate posterior approximation , especially when observed data are not consistent with the theoretical distribution;

Problem: what if there are no suitable sufficient statistics available?

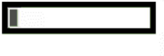


ABC approximations

The challenge in implementing ABC analysis is to **reduce the impact of the approximation** while **restricting the required computation** to acceptable levels.

1. Assume models are approximations to real data-generation process
2. Use summary statistics rather than full datasets
3. Weight summary statistics within a region of the observed ones
$$\pi_{ABC}(\theta|s_{obs}) = \pi(\theta) \int K_h(||s - s_{obs}||)p(s|\theta)ds$$
4. Reduce Monte Carlo error
by using larger numbers of samples from the posterior
or by reducing the variability of importance weights

FURTHER DEVELOPMENTS



- How to compute posterior distribution in ABC?
- When is ABC convenient?
- Practical difficulties

REFERENCES

Chapman & Hall, 2019. *Handbook of Approximate Bayesian Computation*, CRC Press.