Approximate Bayesian Computations via Sufficient Dimension Reduction

Harris Quach

November 28, 2018

Approximate Bayesian Computation: Context

- Observed sample of size n, $y_{obs} \in \mathbb{R}^n \sim f_\theta$, prior $\pi(\theta)$; we want draws from posterior $\pi(\theta|y_{obs}) \propto \pi(\theta) f(y_{obs}|\theta)$
- Problem: $f(y|\theta)$ is intractable - computationally expensive, no analytic form, etc. **But** we can simulate from the f_θ.
- Idea: To sample from posterior, find θ that generate simulations y_{sim} matching y_{obs}, i.e. $y_{sim} = y_{obs}$
- Matching y_{sim} to y_{obs} difficult if n is large, especially if y is continuous.

ABC: Role of Sufficiency

- Easier if we have a lower-dimensional summary statistic $\varphi = \varphi(y)$; ideally, $\varphi(y)$ is a sufficient statistic; $\varphi(y)$ being informative on θ also works.
- Easier if instead of matching, we settle for close enough: $\rho(\varphi(y_{sim}), \varphi(y_{obs})) < \varepsilon$ for ρ a metric, $\varepsilon > 0$

Algorithm 1: ABC

Given: proposal $g(\theta)$; a summary statistic $\varphi(\cdot)$; a metric ρ with some tolerance ε; your acceptance rule as a function of closeness

1. Draw $\theta_{sim} \sim g(\theta)$ for $sim = 1, ..., S$
2. Draw $y_{sim} \in \mathbb{R}^n \sim f_{\theta_{sim}}$ for $sim = 1, ..., S$
3. Accept θ_{sim} according to your rule depending on closeness, e.g $\rho(\varphi(y_{sim}), \varphi(y_{obs})) < \varepsilon$

Result: S Draws from a posterior that approximates $\pi(\theta|y_{obs})$

$$\pi(\theta|y_{obs}) \approx \pi(\theta|\varphi_{obs}) \approx \pi_{ABC}(\theta|\varphi_{obs})$$

Sufficient Dimension Reduction: Context

- But what if we have no idea about φ?
- We have $\theta \in \mathbb{R}^B$, $Y \in \mathbb{R}^{n \times B}$
- Objective: Find an transformation $\varphi(Y)$ such that

$$\theta \perp\!\!\!\perp Y|\varphi(Y), \quad \textit{or} \quad P(\theta|Y) = P(\theta|\varphi(Y))$$

- Finding an informative summary $\varphi = \varphi(y)$ such that $\pi(\theta|y) = \pi(\theta|\varphi(Y))$ is a sufficent dimension reduction problem!

Sufficient Dimension Reduction: Heuristics and Methods

- Being informative means $\varphi(Y)$ explains variation in θ
- Work with matrices like $\Lambda_{sdr} = E($"Variation between θ and Y" $)$;
- The SDR methods we consider are will generally produce estimated functions of the form:

$$\hat{\varphi}(Y) = eigen.vec_d(\hat{\Lambda}_{sdr})(``\widehat{VCov_Y}")^{-1}``Y"$$

Estimating Sufficient Statistic via Simulation

Algorithm 2: Estimating Sufficient Statistic for ABC via Simulation

Given: proposal $g(\theta)$;

Draw $\theta_{sim} \sim g(\theta)$ for $sim = 1, ..., B$

For each θ_{sim}: Draw $y_{sim}^{(r)} \in \mathbb{R}^n \sim f_{\theta_{sim}}$ for $r = 1, ..., R$

begin For each $r = 1, ..., R$:

  Let $\theta^{(r)} = (\theta_1^r, ..., \theta_B^r) \in \mathbb{R}^B$ and $Y^{(r)} = (y_1^r, ..., y_B^r) \in \mathbb{R}^{n \times B}$

  Estimate $\hat{\Lambda}_{sdr}^{(r)}$

end

Construct $\hat{\Lambda}_{sdr} = \frac{1}{R} \sum_{r=1}^{R} \hat{\Lambda}_{sdr}^{(r)}$; and

$\hat{\varphi}(y) = eigen.vec_d(\hat{\Lambda}_{sdr})(\text{“}\widehat{VCov_Y}\text{”})^{-1}\text{“}y\text{”}$

Output: $\hat{\varphi}$: an estimated sufficient statistic for θ

Example: AR(1) in Ghosh & Zhong (2016)

- For $t = 1, ..., n,\ n = 100,\ \sigma = 0.5,\ B = 1000$;

$$y_{t+1} = \theta y_t + \eta_t, \quad \eta_t \sim N(0, \sigma^2), \quad \theta \sim Unif(-1, 1) \implies \theta | y, \sigma \sim N\left(\frac{\sum_2^n y_t y_{t-1}}{\sum_2^n y_{t-1}^2}, \frac{\sigma^2}{\sum_2^n y_{t-1}^2}\right)$$

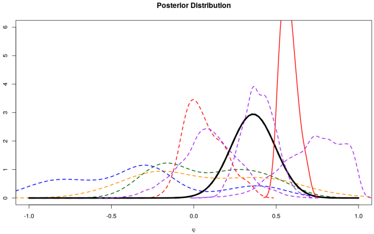
Algorithm 3: ABC with SDR for AR(1)

Given: proposal $g(\theta)$; estimated summary statistic $\hat{\varphi}(\cdot)$; a metric ρ with some tolerance ε; your acceptance rule as a function of closeness

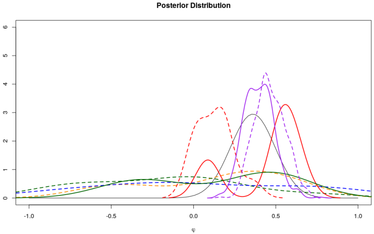
1. Draw $\theta_{sim} \sim g(\theta)$ for $sim = 1, ..., S$
2. Accept θ_{sim} according to your rule depending on closeness, e.g $\rho(\hat{\varphi}(y_{sim}), \hat{\varphi}(y_{obs})) < \varepsilon$

Output: S Draws from $\pi_{ABC}(\theta|\hat{\varphi}_{obs}) \approx \pi(\theta|\varphi(y_{obs})) \approx \pi(\theta|y_{obs})$

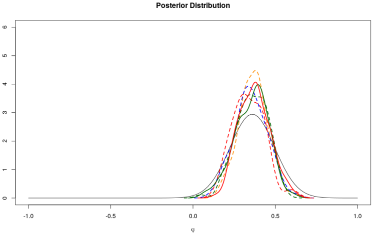
True Posterior vs ABC via (DR, SIR, SAVE, IHT dashed), **GSIR**, **GSAVE**



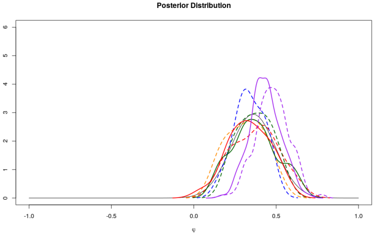
$g = \pi$; $R = 30$ no rep; $B = 1000$, $S = 1000$



$g = \pi$; $R = 200$; $S = 1000$; $MCMC - TruncNorm$



$g = TruncNorm$; $R = 100$; $S = 1000$



$g = TruncNorm$; $R = 200$; $S = 1000$

Conclusion

- Averaging to estimate Λ_{sdr} enables use of SDR by speeding up computation;
- Repeated drawing for each θ_{sim} improves effectiveness of non-linear SDR;
- SDR provides an automated way to construct useful summary statistics for ABC;
- Need to develop better SDR implementation within an MCMC ABC framework;