POLITECNICO
MILANO 1863

How approximate is ABC?

Tutor: Dott. Mario Beraha

January 8th, 2021

Marika Di Marcantonio,
Francesca Pietrobon,
Raffaele Saviello

ABC 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
  - $K_h(u)$ kernel function where $h > 0$ scale parameter
  - $s = S(y)$ summary statistic

- **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 $\frac{K_h(\|s-s_{obs}\|)\pi(\theta^{(i)})}{K*g(\theta^{(i)})}$,

     where $K \geq K_h(0)max_\theta \frac{\pi(\theta)}{g(\theta)}$, otherwise go to 1.

Practical difficulties

- Acceptance rate and tolerance
- Summary statistics
- Kernel
- Proposal

Dataset REACH

The Rotterdam Early Arthritis Cohort dataset contains **681 observations**.

The study's aim is to point out which of the following **12 factors** are associated with the development of rheumatoid arthritis:

ACCP	age	ESR	DC	stiffness	RF
gender	Sym	SJC	TJC	BCPH	BCPF

Model specification:

$$Y_i|\mathrm{X}_i, \beta \sim Bern(\phi(\mathrm{X}_i^T \beta)), \text{ i=1,...,n}$$
$$\beta_\gamma|\gamma \sim (1-\gamma_j)\delta_{\{0\}} + \gamma_j N(0, \sigma^2_{\beta_j})$$
$$\gamma_j|\theta_j \sim Bern(\theta_j), \text{ j=1,...,K}$$
$$\theta_j \sim \mathcal{U}(0,1)$$

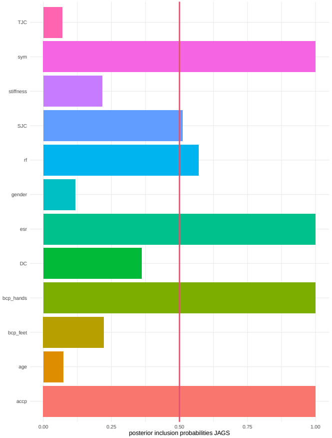
Practical difficulties

- **Acceptance rate and tolerance**
- Summary statistics
- Kernel
- Proposal

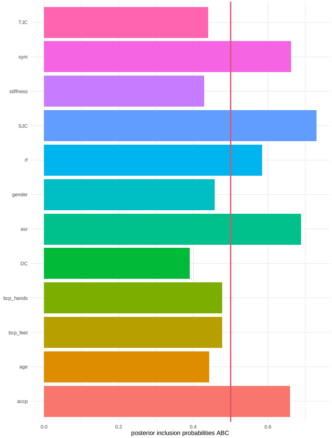
Acceptance Rate and Tolerance

- Simulate 200 000 observations
- Collect differences in the vector d $= \|y - y_{obs}\|$
- Set **tolerance** $= 1\%$ percentile of d
- **Acceptance rate** $\simeq 1\%$

Results



(a) MCMC using JAGS



(b) ABC rejection sampling

Figure: Comparison of covariates

Dataset GENUS

The dataset consists in **341 observations** divided in **12 genus** and, for each observation, we have:

BoW SVL BrW

Feature variable: $R = log\left(\frac{BrW/BoW}{1-BrW/BoW}\right)$

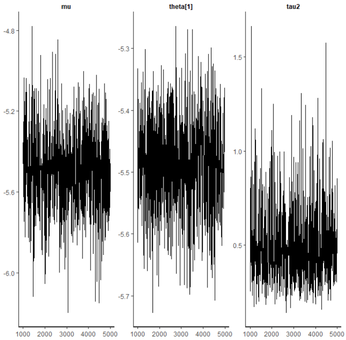
Hierarchical model:

$$\mathrm{R}_j = R_1, \ldots, R_{n_j} | \theta_j, \sigma^2 \sim N(\theta_j, \sigma^2), \text{ j=1,...,M=12 (genus)}$$
$$\theta_1, \ldots, \theta_M | \mu, \tau^2 \sim N(\mu, \tau^2)$$
$$(\mu, \tau^2) \sim N(m_0, s_0^2) * InvGamma(\alpha_0, \beta_0)$$
$$\sigma^2 \sim InvGamma(a_0, b_0)$$

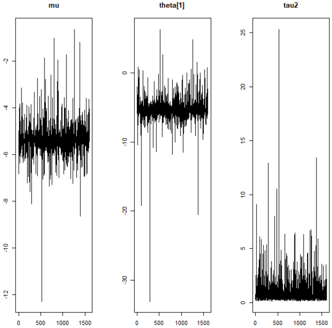
Practical difficulties

- Acceptance rate and tolerance
- **Summary statistics**
- Kernel
- Proposal

Comparison of parameters

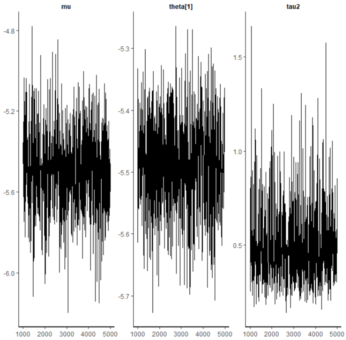


(a) MCMC using STAN

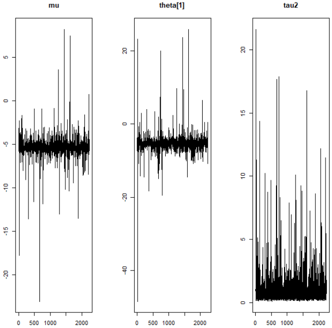


(b) ABC Rejection Sampling using the **mean** as summary statistic

Comparison of parameters



(c) MCMC using STAN



(d) ABC Rejection Sampling using the **median** as summary statistic

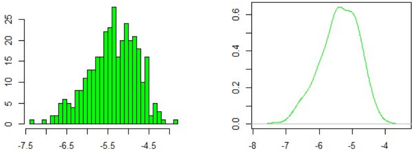


Figure: True value of R

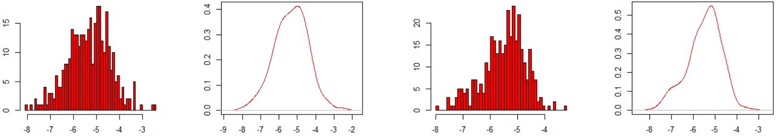


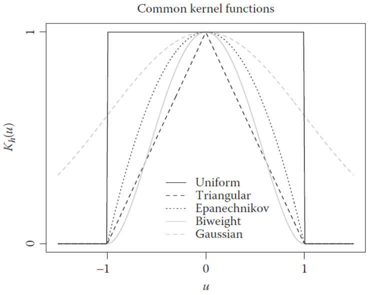
Figure: Estimated values of R: **mean** and **median** respectively

Practical difficulties

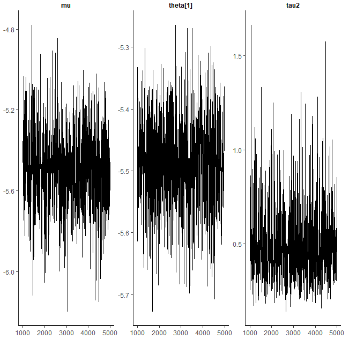
- Acceptance rate and tolerance
- Summary statistics
- **Kernel**
- Proposal

Kernel

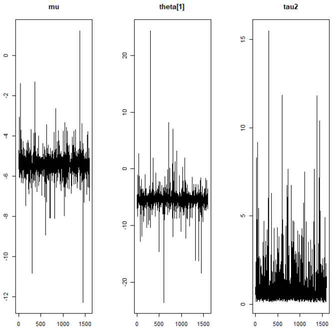
Kernel	K(u)
Uniform	$\frac{1}{2}\mathbb{1}_{\lvert u\rvert\leq 1}$
Triangular	$(1-\lvert u\rvert)\mathbb{1}_{\lvert u\rvert\leq 1}$
Epanechnikov	$\frac{3}{4}(1-u^2)\mathbb{1}_{\lvert u\rvert\leq 1}$
Biweight	$\frac{15}{16}(1-u^2)^3\mathbb{1}_{\lvert u\rvert\leq 1}$
Gaussian	$\frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}u^2}$



Comparison of parameters

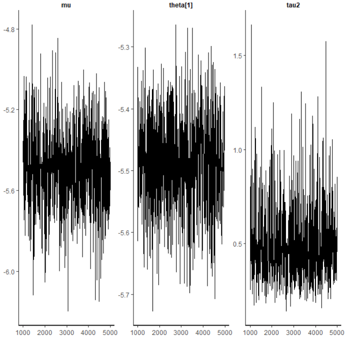


(a) MCMC using STAN

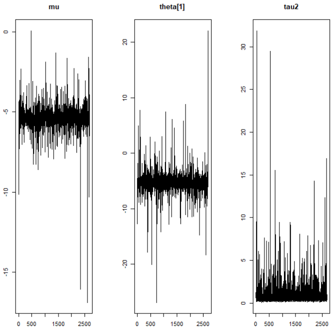


(b) ABC Rejection Sampling using the **mean** as summary statistic **and uniform kernel**

Comparison of parameters



(c) MCMC using STAN



(d) ABC Rejection Sampling using the **mean** as summary statistic **and gaussian kernel**

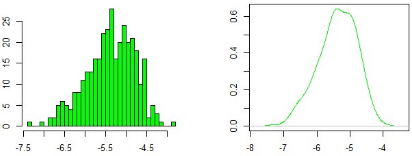


Figure: True value of R

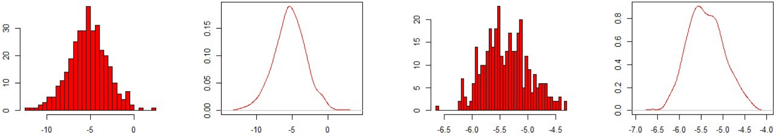


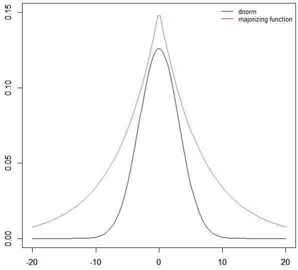
Figure: Estimated values of R: mean with **uniform** and **gaussian kernel** respectively

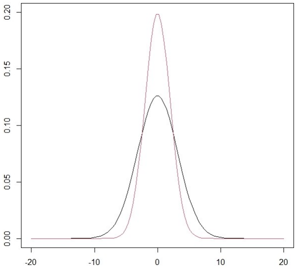
Practical difficulties

- Acceptance rate and tolerance
- Summary statistics
- Kernel
- **Proposal**

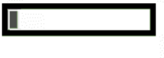
Proposal

1. **Majorizing function**
   It can be obtained by simply rescaling a proposal density
2. **Importance weights density**
   The proposal density is large in parameter ranges that are preferred by the data.

1. 

2. 

FURTHER DEVELOPMENTS



ABC MCMC
ABC SMC

REFERENCES

Chapman & Hall, 2019. *Handbook of Approximate Bayesian Computation*, CRC Press.
Jackman, 2009. *Bayesian Analysis for the Social Sciences*, WILEY.
Alsingy, Wandeltyzx, Feeneyy, 2018. *Optimal proposals for ABC*, ISBA.
Simola. *Developments in ABC and Statistical Applications in Astrostatistics*, UniPd.

Corradin. *TA11, TA12 "Bayesian Statistics" course*, PoliMi.