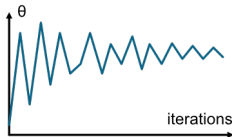




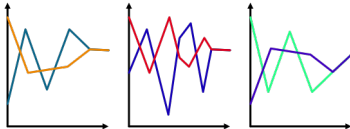
POLITECNICO
MILANO 1863

Coupled Markov chains with applications to Approximate Bayesian Computation for model based clustering

E. Bertoni, M. Caldarini, F. Di Filippo, G. Gabrielli, E. Musiari
15 February 2022



**Unbiased Markov chain
Monte Carlo methods with
couplings**



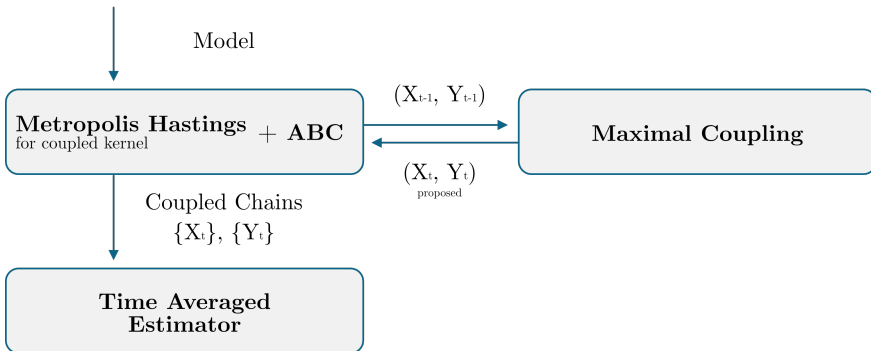
likelihood



**Approximate Bayesian
Computation**

The complete method: MCMC + Couplings + ABC

Implementation



Maximal coupling

Maximal coupling between two distributions p and q

- sample $X \sim p$;
- sample $W|X \sim \mathcal{U}\{[0, p(X)]\}$;
- if $W \leq q(X)$ then output (X, X) , otherwise:
 - ① sample $Y \sim q$;
 - ② sample $W^*|Y \sim \mathcal{U}\{[0, q(Y)]\}$ until $W^* > p(Y)$
 - ③ output (X, Y) .

Output

Distribution of a pair of random variables X, Y that maximizes $\mathbb{P}(X = Y)$ subject to the marginal constraints $X \sim p$ and $Y \sim q$

Metropolis Hastings with couplings and ABC

Model

$$Y_i | \mu, \sigma^2 \stackrel{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$$

$$\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$$

$$\sigma^2 \sim \text{InvGa}(a, b)$$

$$\mu_0 = 8, \quad \sigma_0^2 = 4$$

$$a = b = 1$$

Initialization

set $\theta = (\mu, \sigma^2)$

for $k=1,2$:

until $K_h(||s_k^{(0)} - s_{obs}||) > 0$:

- sample $\theta_k^{(0)} \sim \pi(\theta)$
- simulate n_{obs} observations $y_{ki} \sim \mathcal{N}(\mu^{(0)}, \sigma^{2(0)})$ from maximal coupling
- compute $s_k^{(0)} = S(y_k)$;

Iterations

for $i = 1, \dots, M$:

given $\theta^{(i-1)} = (\mu^{(i-1)}, \sigma^{2(i-1)})$

- generate $[\theta_1^{(i)}, \theta_2^{(i)}]$ from a maximal coupling given $[\theta_1^{(i-1)}, \theta_2^{(i-1)}]$;
- for $k=1,2$:
 - ① simulate n_{obs} observations $y_k^{(i)} \sim \mathcal{N}(\mu^{(i)}, \sigma^{2(i)})$ from maximal coupling;
 - ② compute $s_k^{(i)} = S(y_k)$;
 - ③ accept $\theta_k^{(i)}$ with probability

$$\frac{K_h(\|s_k^{(i)} - s_{obs}\|)\pi(\theta_k^{(i)})}{K_h(\|s_k^{(i-1)} - s_{obs}\|)\pi(\theta_k^{(i-1)})}$$

Output

for $k = 1, 2$

$$\mu_k^{(1)}, \dots, \mu_k^{(M)} \sim \pi_{ABC}(\mu | y_{obs});$$

$$\sigma_k^{2(1)}, \dots, \sigma_k^{2(M)} \sim \pi_{ABC}(\sigma | y_{obs}).$$

Summary statistics: Sample mean, Sample Variance

Distance: L^2 -norm

Kernel: $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$, $K_h(u) = \frac{K(\frac{u}{h})}{h}$

Proposal distribution: $\begin{bmatrix} \mu^* \\ \log(\sigma^{2*}) \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mu^{(i)} \\ \log(\sigma^{2(i)}) \end{bmatrix}, 0.1^2 \cdot \mathcal{I} \right)$

Summary statistics: Sample mean, Sample Variance

Distance: L^2 -norm

Kernel: $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$, $K_h(u) = \frac{K(\frac{u}{h})}{h}$

Proposal distribution: $\begin{bmatrix} \mu^* \\ \log(\sigma^{2*}) \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mu^{(i)} \\ \log(\sigma^{2(i)}) \end{bmatrix}, 0.1^2 \cdot \mathcal{I} \right)$

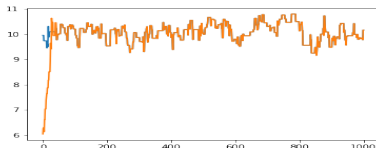
Dataset

100 samples generated from a Gaussian distribution:

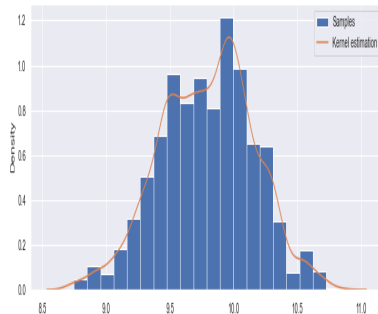
$$Y_{obs} \sim \mathcal{N}(\mu_{obs}, \sigma_{obs}^2)$$

$$\mu_{obs} = 10, \sigma_{obs}^2 = 3$$

Coupled chains from one processor



Sampling histogram with estimated distribution

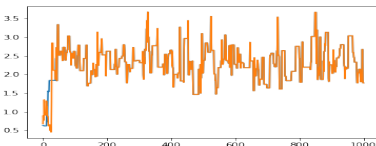


Samplings from all accepted chains

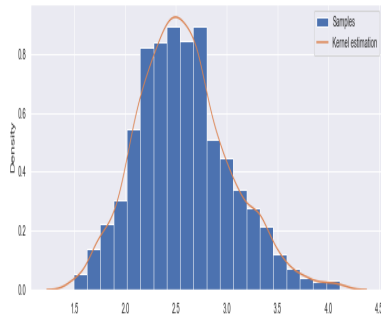


Time Averaged Estimators mean obtained: $\mathbb{E}[H_{k:m}(X_\mu, Y_\mu)] \simeq 9.8$

Coupled chains from one processor



Sampling histogram with estimated distribution



Samplings from all accepted chains



Time Averaged Estimators mean obtained: $\mathbb{E}[H_{k:m}(X_{\sigma^2}, Y_{\sigma^2})] \simeq 2.56$

Numerical experiment: g-and-k distribution

Quantile function

$$r \in (0, 1) \mapsto a + b \left(1 + 0.8 \left(\frac{1 - \exp(-g \cdot z(r))}{1 + \exp(-g \cdot z(r))} \right) \right) (1 + z(r)^2)^k \cdot z(r)$$

where $z(r)$ is the r -th quantile of the standard Normal distribution.

Prior distributions

$$a \sim \mathcal{U}([0, 10])$$

$$b \sim \mathcal{U}([0, 10])$$

$$g \sim \mathcal{U}([0, 10])$$

$$k \sim \mathcal{U}([0, 10])$$

Summary statistics: 10 quantiles

Distance: L^2 – norm

Kernel:

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}, \quad K_h(u) = \frac{K(\frac{u}{h})}{h}$$

Dataset

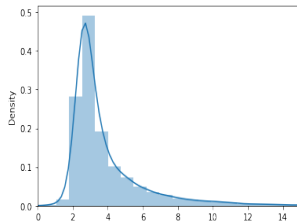
100 samples generated as follows:

$$a_{obs} = 3, b_{obs} = 1, g_{obs} = 2, k_{obs} = 0.5$$

$$y_{obs} \sim \text{quantilefunction}(z(r), \theta_{obs})$$

$$z(r) \sim \mathcal{N}(0, 1)$$

Quantile distribution: y_{obs}



Initialization

given $\theta = (a, b, g, k)$

for $k=1,2$:

Untill $K_h(\|s_k^{(0)} - s_{obs}\|) > 0$:

- Generate $\theta_k^{(0)} \sim \mathcal{U}([0, 10]^4)$ from prior density.
- Generate $z(r) \sim \mathcal{N}(0, 1)$
- Generate a sample of n_{obs} observations such that $y_k \sim \text{quantilefunction}(z(r), \theta_k^0)$
- Compute $s_k^{(0)} = S(y_k)$

Iterations

for i in $(1, \dots, M)$:

- generate $\theta_1^{(i)}, \theta_2^{(i)}$ from $\text{maximalcoupling}(\theta_1^{(i-1)}, \theta_2^{(i-1)})$
- simulate n_{obs} observations $y_1^{(i)}, y_2^{(i)}$ from maximal coupling
- Compute the summaries $s_k^{(i)} = S(y_k^{(i)})$
- Accept $\theta_k^{(i)}$ with probability

$$\frac{Kh(\|s_k^{(i)} - s_{obs}\|)\pi(\theta_k^{(i)})}{Kh(\|s_k^{(i-1)} - s_{obs}\|)\pi(\theta_k^{(i-1)})}$$

otherwise $\theta_k^{(i)} = \theta_k^{(i-1)}$;

Iterations

for i in $(1, \dots, M)$:

Couplings

for j in $(1, \dots, n_{obs})$:

- ① $z^{(i,j)}(r) \sim \mathcal{N}(0, 1)$

- ② generate:

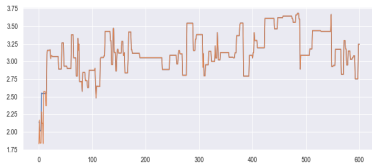
$$y_1^{(ij)} \sim \text{quantilefunction}(z^{(i,j)}(r), \theta_1^{(i-1)})$$

$$y_2^{(ij)} \sim \text{quantilefunction}(z^{(i,j)}(r), \theta_2^{(i-1)})$$

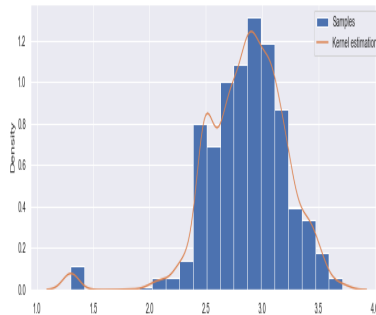
$$\kappa n(\|S_k^{(i)} - S_{obs}\|) \pi(\theta_k^{(i)})$$

otherwise $\theta_k^{(i)} = \theta_k^{(i-1)}$;

Coupled chains from one core



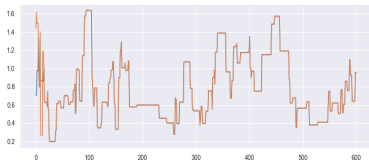
Samplings from all accepted chains



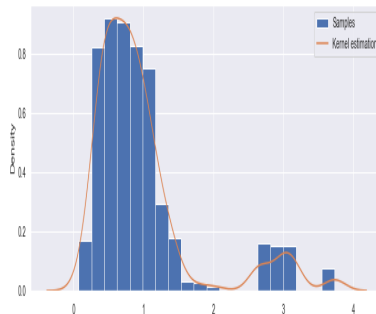
Sampling histogram with estimated distribution



Coupled chains from one core



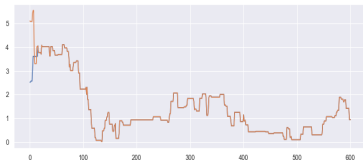
**Sampling histogram with
estimated distribution**



Samplings from all accepted chains



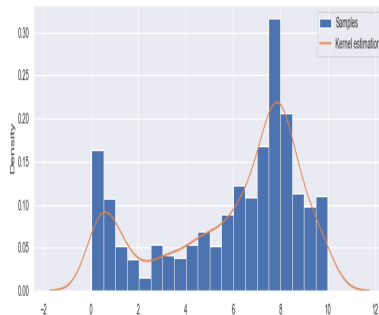
Coupled chains from one core



Samplings from all accepted chains

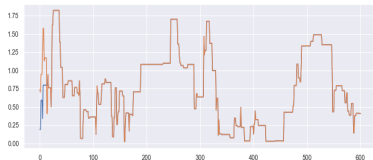


**Sampling histogram with
estimated distribution**

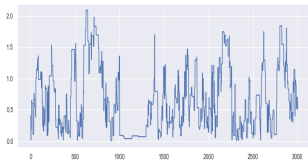


Results for parameter g

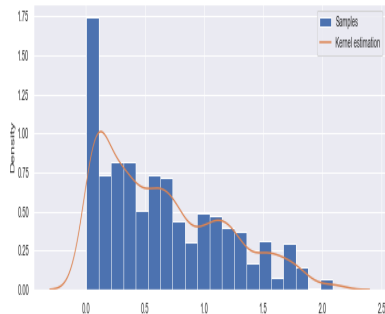
Coupled chains from one core



Samplings from all accepted chains



**Sampling histogram with
estimated distribution**



Pierre Jacob, John O'Leary, and Yves Atchadé.

Unbiased markov chain monte carlo with couplings.

Journal of the Royal Statistical Society: Series B (Statistical Methodology), 82, 08 2017.

Peter W. Glynn and Chang han Rhee.

Exact estimation for markov chain equilibrium expectations, 2014.

Jeffrey S. Rosenthal.

Faithful couplings of markov chains: Now equals forever.

Advances in Applied Mathematics, 18(3):372–381, 1997.

Dylan Cordaro.

Markov chain and coupling from the past.

2017.

Jinming Zhang.

Markov chains, mixing times and coupling methods with an application in social learning.

2020.

S. A. Sisson, Y. Fan, and M. A. Beaumont.

Overview of approximate bayesian computation, 2018.

Y. Fan and S. A. Sisson.

Abc samplers, 2018.

Dennis Prangle.

Summary statistics in approximate bayesian computation, 2015.