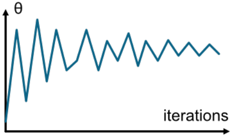
Coupled Markov chains with applications to Approximate Bayesian Computation for model based clustering

E. Bertoni, M. Caldarini, F. Di Filippo, G. Gabrielli, E. Musia

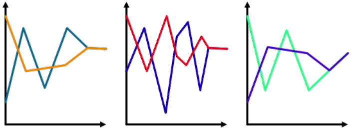
10 January 2022

A complex problem



⇓

Unbiased Markov chain Monte Carlo methods with couplings

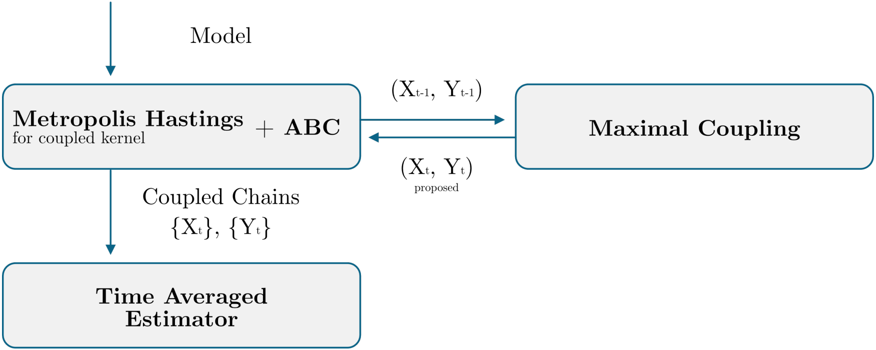


⇓

Approximate Bayesian Computation

The complete method: MCMC + Couplings + ABC

Implementation



Metropolis Hastings with couplings and ABC

1. Compute $s_{obs} = S(y_{obs})$;
2. generate $\mu_x^{(0)} \sim \pi(\mu)$ and $\mu_y^{(0)} \sim \pi(\mu)$ from prior density;
3. generate $\sigma_x^{2(0)} \sim \pi(\sigma^2)$ and $\mu_y^{2(0)} \sim \pi(\sigma^2)$ from prior density;
4. generate with a maximal coupling two samples of N observations such that $y_{1i} \sim \mathcal{N}(\mu_x^{(0)}, \sigma_x^{2(0)})$ and $y_{2j} \sim \mathcal{N}(\mu_y^{(0)}, \sigma_y^{2(0)})$;
5. compute $s_x^{(0)} = S(y_1)$ and $s_y^{(0)} = S(y_2)$;
6. until $K_h(||s_x^{(0)} - s_{obs}||) > 0$:
   - generate $\mu_x^{(0)} \sim \pi(\mu)$ and $\sigma_x^{2(0)} \sim \pi(\sigma^2)$ from prior densities;
   - generate a sample of N observations such that $y_{1i} \sim \mathcal{N}(\mu_x^{(0)}, \sigma_x^{2(0)})$;
   - compute $s_x^{(0)} = S(y_1)$;
7. until $K_h(||s_y^{(0)} - s_{obs}||) > 0$:
   - generate $\mu_y^{(0)} \sim \pi(\mu)$ and $\sigma_y^{2(0)} \sim \pi(\sigma^2)$ from prior densities;
   - generate a sample of N observations such that $y_{1i} \sim \mathcal{N}(\mu_y^{(0)}, \sigma_y^{2(0)})$;
   - compute $s_x^{(0)} = S(y_1)$;

Metropolis Hastings with couplings and ABC

8. for i = 1,...,N: given

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

- generate $[\theta_x^{(i)}, \theta_y^{(i)}]$ from a maximal coupling given $[\theta_x^{(i-1)}, \theta_y^{(i-1)}]$;
- generate from a maximal coupling two samples of N observations $y_1 \sim p(y|\theta_x^{(i)})$ and $y_2 \sim p(y|\theta_y^{(i)})$;
- compute $s_x^{(i)} = S(y_1)$ and $s_y^{(i)} = S(y_2)$;
- accept $\theta_x^{(i)}$ with probability

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

and accept $\theta_y^{(i)}$ with probability

$$\frac{K_h(||s_y^{(i)} - s_{obs}||)\pi(\theta_y^{(i)})}{K_h(||s_y^{(i-1)} - s_{obs}||)\pi(\theta_y^{(i-1)})}.$$

Metropolis Hastings with couplings and ABC

As output we get two sets of parameter vectors:

$$\mu_x^{(1)}, ..., \mu_x^{(N)} \sim \pi_{ABC}(\mu|y_{obs});$$
$$\mu_y^{(1)}, ..., \mu_y^{(N)} \sim \pi_{ABC}(\mu|y_{obs}).$$

$$\sigma_x^{2(1)}, ..., \sigma_x^{2(N)} \sim \pi_{ABC}(\sigma^2|y_{obs});$$
$$\sigma_y^{2(1)}, ..., \sigma_y^{2(N)} \sim \pi_{ABC}(\sigma^2|y_{obs}).$$

Maximal coupling of both parameters

Set $p = \mathcal{N}(X_{t-1}, 0.1^2 * \mathbb{I})$ and $q = \mathcal{N}(Y_{t-1}, 0.1^2 * \mathbb{I})$, then:

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

Summary statistic:

Sample mean, Sample Variance

Distance:

$$L^2 - \textit{norm of the difference of } S(y) \textit{ and } s_{obs}$$

Kernel:

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

Model

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

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

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

$$\pi(\mu,\sigma) = \pi(\mu) * \pi(\sigma)$$

$$\mu_0 = 34, \quad \sigma_0^2 = 3$$

$$a = b = 1$$

Dataset

100 samples generated from a Gaussian distribution:

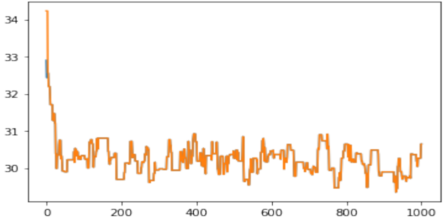
$$Y_{obs} \sim \mathcal{N}(\mu_{obs}, \sigma^2_{obs})$$

$$\mu_{obs} = 30, \sigma^2_{obs} = 2$$

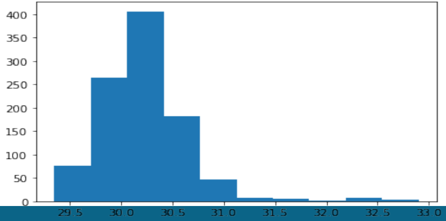
Results

$E[\mu|y] = 30,19$

Coupled chains



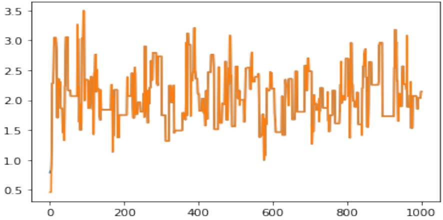
Sampling histogram with real distribution



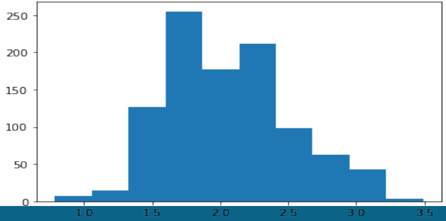
Results

$E[\sigma^2|y] = 2,06$

Coupled chains



Sampling histogram with real distribution



Model

QUANTILE FUNCTION:

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

where

$$z(r)$$

is the r-th quantile of the standard Normal distribution.

Model

Prior distribution: $a \sim \mathcal{U}([0,10])$
$b \sim \mathcal{U}([0,10])$
$g \sim \mathcal{U}([0,10])$
$k \sim \mathcal{U}([0,10])$

Study case

Summary statistic:

10 quantile + minimum

Distance:

$$L^2 - \textit{norm of the difference of } S(y) \textit{ and } s_{obs}$$

Kernel:

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

Study case

Dataset

100 samples generated from the quantile distribution:

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

$$y_{obs} = a_{obs} + b_{obs}(1 + 0.8\left(\frac{1 - exp(-g_{obs} \cdot z(r))}{1 + exp(-g_{obs} \cdot z(r))}\right)(1 + z(r)^2)^k \cdot z(r)$$

and

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

ABC ALGORITHM

INITIALIZATION:
$\theta^{(0)} = (a^{(0)}, b^{(0)}, g^{(0)}, k^{(0)})$
$\theta^{(0)} \sim \mathcal{U}([0, 10]^4)$
for j in (1,...,n):

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

compute the summaries statistics: $s^{(0)} = S(y^{(0)})$
Untill $Kh(\|s^{(0)} - s_{obs}\|) > 0$

- Generate $\theta^{(0)} \sim \mathcal{U}([0, 10]^4)$ from prior density.
- Generate $z^{(0)}(r) \sim \mathcal{N}(0, 1)$
- Generate a sample of 1000 observations such that $y \sim quantilefunction(z^{(0)}(r), \theta^{(0)})$
- Compute $s^{(0)} = S(y)$

ABC ALGORITHM

for i in (1,...,M):

1. given $\theta^{(i-1)} = (a^{(i-1)}, b^{(i-1)}, g^{(i-1)}, k^{(i-1)})$
   $\theta^{(i)} \sim \mathcal{N}(\theta^{(i-1)}, I)$
   for j in (1,...,n):
   $z^{(i,j)}(r) \sim \mathcal{N}(0,1)$
   $y^{(ij)} \sim quantilefunction(z^{(i,j)}(r), \theta^{(i-1)})$
2. Compute the summaries $s^{(i)} = S(y^{(i)})$
3. Accept $\theta^{(i)}$ with probability $\frac{Kh(\|s^{(i)} - s_{obs}\|)\pi(\theta^{(i)})}{Kh(\|s^{(i-1)} - s_{obs}\|)\pi(\theta^{(i-1)})}$

ABC + COUPLED MARKOV CHAIN for parameter a

parameters b,g,k are fixed equal to the observation parameters

$a_1^{(0)} \sim \mathcal{U}([0, 10])$

$a_2^{(0)} \sim \mathcal{U}([0, 10])$

for j in (1,...,n_{obs}):

- $z^{(0,j)}(r) \sim \mathcal{N}(0, 1)$
- $y_1^{(0,j)} \sim quantilefunction(z^{(0,j)}(r))$
- compute $s_1^{(0)} = S(y_1^{(0)})$

for j in (1,...,n_{obs}):

- $z^{(0,j)}(r) \sim \mathcal{N}(0, 1)$
- $y_2^{(0,j)} \sim quantilefunction(z^{(0,j)}(r))$
- compute the summaries statistics: $s_2^{(0)} = S(y_2^{(0)})$

ABC + COUPLED MARKOV CHAIN for parameter a

Untill $Kh(\|s_1^{(0)} - s_{obs}\|) > 0$:

1. Generate $a_1^{(0)} \sim \mathcal{U}([0,10])$ from prior density.
2. Generate $z_1^{(0)}(r) \sim \mathcal{N}(0,1)$
3. Generate a sample of 1000 observations such that $y_1 \sim quantilefunction(z_1^{(0)}(r), a_1^0)$
4. Compute $s_1^{(0)} = S(y_1)$

Untill $Kh(\|s_2^{(0)} - s_{obs}\|) > 0$:

1. Generate $a_2^{(0)} \sim \mathcal{U}([0,10])$ from prior density.
2. Generate $z_2^{(0)}(r) \sim \mathcal{N}(0,1)$
3. Generate a sample of 1000 observations such that $y_2 \sim quantilefunction(z_2^{(0)}(r), a_2^{(0)})$
4. Compute $s_2^{(0)} = S(y_2)$

ABC + COUPLED MARKOV CHAINS for parameter a

for i in (1,...,M):

1. generate $a_1^{(i)}, a_2^{(i)}$ from maximalcoupling$(a_1^{(i-1)}, a_2^{(i-1)})$

for j in (1,...,n):

- $z^{(i,j)}(r) \sim \mathcal{N}(0,1)$
- generate: $y_1^{(ij)} \sim quantilefunction(z^{(i,j)}(r), \theta^{(i-1)})$
  $y_2^{(ij)} \sim quantilefunction(z^{(i,j)}(r), \theta^{(i-1)})$

2. Compute the summaries $s_1^{(i)} = S(y_1^{(i)})$ and $s_2^{(i)} = S(y_2^{(i)})$
3. acceptance:
   - Accept $a_1^{(i)}$ with probability $\frac{Kh(\|s_1^{(i)} - s_{obs}\|)\pi(a_1^{(i)})}{Kh(\|s_1^{(i-1)} - s_{obs}\|)\pi(a_1^{(i-1)})}$ otherwise $a_1^{(i)} = a_1^{(i-1)}$
   - Accept $a_2^{(i)}$ with probability $\frac{Kh(\|s_2^{(i)} - s_{obs}\|)\pi(a_2^{(i)})}{Kh(\|s_2^{(i-1)} - s_{obs}\|)\pi(a_2^{(i-1)})}$ otherwise $a_2^{(i)} = a_2^{(i-1)}$

Maximal coupling

given $(a_1^{(i-1)},\ a_2^{(i-1)})$

set $p(\cdot|a_1^{(i-1)}) = \mathcal{N}(a_1^{(i-1)}, 1)$ and $q(\cdot|a_2^{(i-1)}) = \mathcal{N}(a_2^{(i-1)}, 1)$

1. $a_1' \sim p(\cdot|a_1^{(i-1)})$
2. if $w_1 \sim \mathcal{U}((0, p(a'|a_1^{(i-1)}))) < q(a'|a_2^{(i-1)})$: $a_1^{(i)} = a_1'$ *and* $a_2^{(i)} = a_1'$

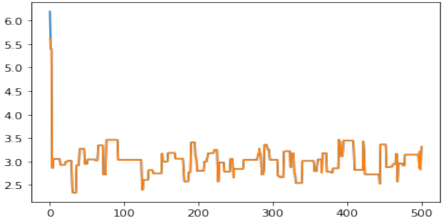
otherwise until $w_2 \sim \mathcal{U}((0, q(a'|a_2^{(i-1)})) > p(a'|a_1^{(i-1)})$:

1. generate $a_2' \sim q(\cdot|a_2^{(i-1)})$
2. sample $l \sim \mathcal{U}((0, 1))$

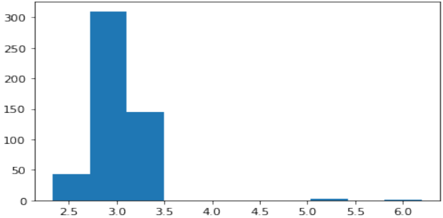
and then set: $a_1^{(i)} = a_1'$ *and* $a_2^{(i)} = a_2'$

Results: parameter a

Coupled chains

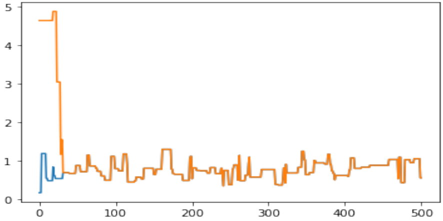


Sampling histogram with real distribution

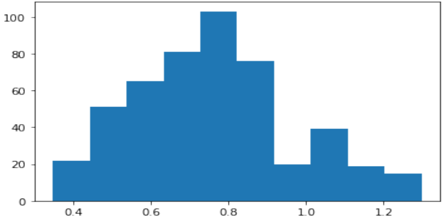


Results: parameter b

Coupled chains

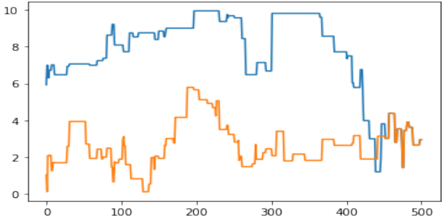


Sampling histogram

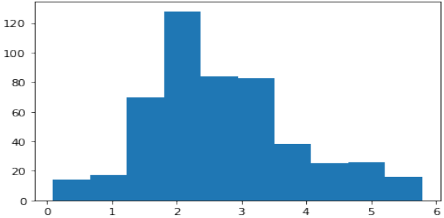


Results: parameter g

Coupled chains

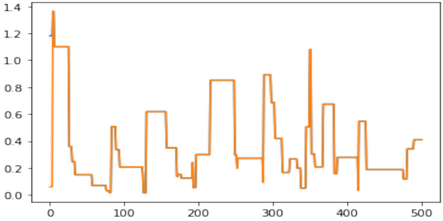


Sampling histogram

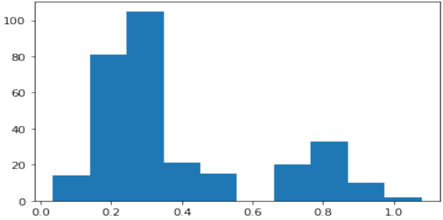


Results: parameter k

Coupled chains



Sampling histogram



given $\theta = (a, b, g, k)$
$\theta_1^{(0)} \sim \mathcal{U}([0, 10]^4)$
$\theta_2^{(0)} \sim \mathcal{U}([0, 10]^4)$
for j in (1,...,n_{obs}):

- $z^{(0,j)}(r) \sim \mathcal{N}(0, 1)$
- $y_1^{(0,j)} \sim quantilefunction(z^{(0,j)}(r), \theta_1^{(0)})$
- compute $s_1^{(0)} = S(y_1^{(0)})$

for j in (1,...,n_{obs}):

- $z^{(0,j)}(r) \sim \mathcal{N}(0, 1)$
- $y_2^{(0,j)} \sim quantilefunction(z^{(0,j)}(r), theta_2^{(0)})$
- compute the summaries statistics: $s_2^{(0)} = S(y_2^{(0)})$

ABC + COUPLED MARKOV CHAIN for parameter a

Untill $Kh(\|s_1^{(0)} - s_{obs}\|) > 0$:

1. Generate $\theta_1^{(0)} \sim \mathcal{U}([0,10]^4)$ from prior density.
2. Generate $z_1^{(0)}(r) \sim \mathcal{N}(0,1)$
3. Generate a sample of 1000 observations such that $y_1 \sim quantilefunction(z_1^{(0)}(r), \theta_1^0)$
4. Compute $s_1^{(0)} = S(y_1)$

Untill $Kh(\|s_2^{(0)} - s_{obs}\|) > 0$:

1. Generate $\theta_2^{(0)} \sim \mathcal{U}([0,10]^4)$ from prior density.
2. Generate $z_2^{(0)}(r) \sim \mathcal{N}(0,1)$
3. Generate a sample of 1000 observations such that $y_2 \sim quantilefunction(z_2^{(0)}(r), theta_2^{(0)})$
4. Compute $s_2^{(0)} = S(y_2)$

ABC + COUPLED MARKOV CHAINS

for i in (1,...,M):

1. generate $\theta_1^{(i)}, \theta_2^{(i)}$ from maximalcoupling$(\theta_1^{(i-1)}, \theta_2^{(i-1)})$

for j in (1,...,n):

- $z^{(i,j)}(r) \sim \mathcal{N}(0,1)$
- generate: $y_1^{(ij)} \sim quantilefunction(z^{(i,j)}(r), \theta_1^{(i-1)})$
  $y_2^{(ij)} \sim quantilefunction(z^{(i,j)}(r), \theta_1^{(i-1)})$

2. Compute the summaries $s_1^{(i)} = S(y_1^{(i)})$ and $s_2^{(i)} = S(y_2^{(i)})$
3. acceptance:
   - Accept $a_1^{(i)}$ with probability $\frac{Kh(\|s_1^{(i)} - s_{obs}\|)\pi(\theta_1^{(i)})}{Kh(\|s_1^{(i-1)} - s_{obs}\|)\pi(\theta_1^{(i-1)})}$ otherwise $a_1^{(i)} = \theta_1^{(i-1)}$
   - Accept $a_2^{(i)}$ with probability $\frac{Kh(\|s_2^{(i)} - s_{obs}\|)\pi(\theta_2^{(i)})}{Kh(\|s_2^{(i-1)} - s_{obs}\|)\pi(\theta_2^{(i-1)})}$ otherwise $\theta_2^{(i)} = \theta_2^{(i-1)}$

Maximal coupling

given $(\theta_1^{(i-1)},\ \theta_2^{(i-1)})$

set $p(\cdot|\theta_1^{(i-1)}) = \mathcal{N}(\theta_1^{(i-1)}, 0.1^2\mathbb{I})$ and $q(\cdot|\theta_2^{(i-1)}) = \mathcal{N}(\theta_2^{(i-1)}, 0.1^2\mathbb{I})$

1. $\theta_1' \sim p(\cdot|\theta_1^{(i-1)})$
2. if $w_1 \sim \mathcal{U}((0, p(\theta'|\theta_1^{(i-1)}))) < q(\theta'|\theta_2^{(i-1)}) : \theta_1^{(i)} = \theta_1'$ *and* $\theta_2^{(i)} = \theta_1'$

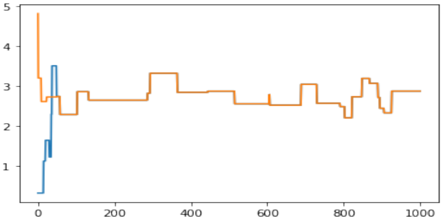
otherwise until $w_2 \sim \mathcal{U}((0, q(\theta'|\theta_2^{(i-1)})) > p(\theta'|\theta_1^{(i-1)})$:

- generate $\theta_2' \sim q(\cdot|\theta_2^{(i-1)})$

and then set: $\theta_1^{(i)} = \theta_1'$ *and* $\theta_2^{(i)} = \theta_2'$

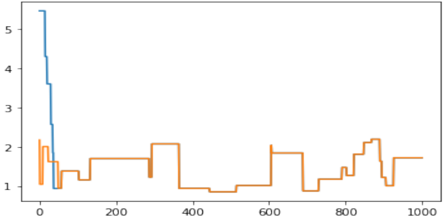
Results: parameter a

Coupled chains



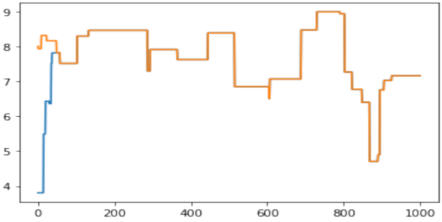
Results: parameter b

Coupled chains



Results: parameter g

Coupled chains



Results: parameter k

Coupled chains