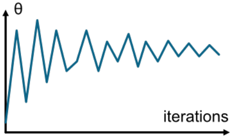
Coupled Markov chains with applications to Approximate Bayesian Computation for model based clustering

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

10 January 2022

A complex problem

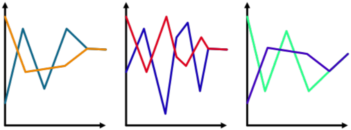


$\Downarrow$

Unbiased Markov chain Monte Carlo methods with couplings

$\Downarrow$

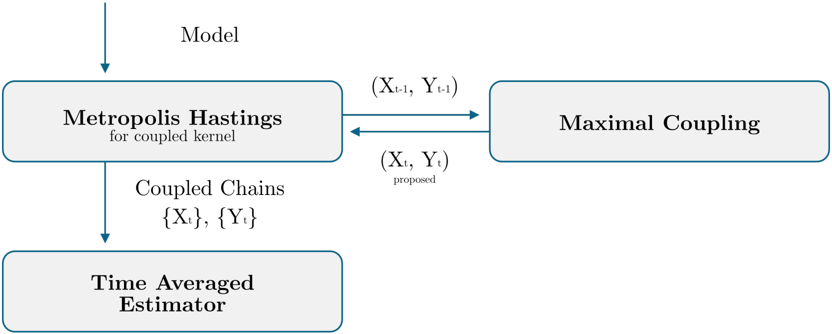
Approximate Bayesian Computation



Unbiased Markov chain Monte Carlo methods with couplings

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

Structure of the method



Time-averaged estimator

1. draw X_0 and Y_0 from an initial distribution π_0 and draw $X_1 \sim P(X_0, \cdot)$;
2. set $t = 1$: while $t < \max\{m, \tau\}$ and:
   a. draw $(X_{t+1}, Y_t) \sim \bar{P}\{(X_t, Y_{t-1}), \cdot\}$;
   b. set $t \leftarrow t + 1$;
3. compute the time-averaged estimator:

$$H_{k:m}(X, Y) = \frac{1}{m-k+1} \sum_{l=k}^{m} h(X_l) + \sum_{l=k+1}^{\tau-1} \min(1, \frac{l-k}{m-k+1})\{h(X_l) - h(Y_{l-1})\}.$$

Metropolis–Hasting algorithm for a coupled kernel

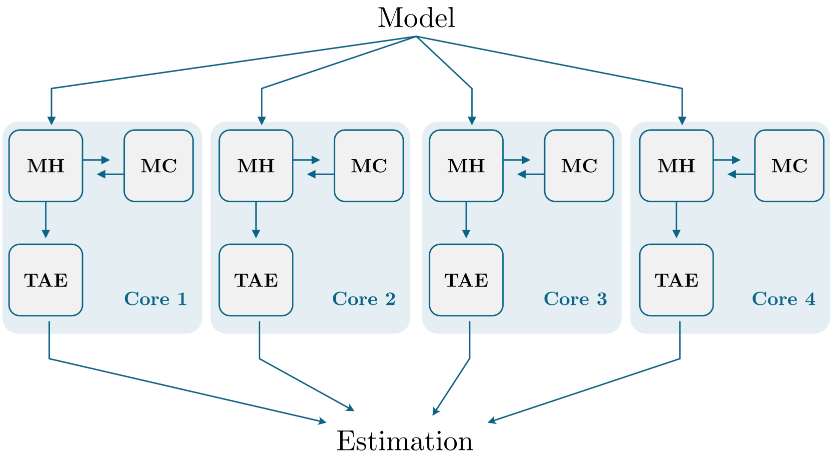
1. sample $(X^\star, Y^\star)|(X_t, Y_{t-1})$ from a maximal coupling of $q(X_t, \cdot)$ and $q(Y_{t-1}, \cdot)$;
2. sample $U \sim \mathcal{U}([0,1])$;
3. if
$$U \leq \min\left\{1, \frac{\pi(X^\star)q(X^\star, X_t)}{\pi(X_t)q(X_t, X^\star)}\right\}$$
then $X_{t+1} = X^\star$; otherwise $X_t = X_{t-1}$;
4. if
$$U \leq \min\left\{1, \frac{\pi(Y^\star)q(Y^\star, Y_t)}{\pi(Y_t)q(Y_t, Y^\star)}\right\}$$
then $Y_{t+1} = Y^\star$; otherwise $Y_t = Y_{t-1}$.

Maximal coupling

Set $p = \mathcal{N}(X_{t-1}, 1)$ and $q = \mathcal{N}(Y_{t-1}, 1)$, 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).

Parallelization



Study case

Model

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

$$\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$$
$$\mu_0 = 38, \quad \sigma_0^2 = 4$$

Dataset

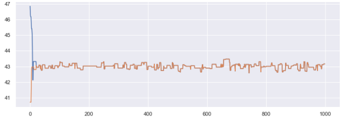
1000 samples generated from a Gaussian distribution:

$$Y_{obs} \sim \mathcal{N}(\mu_{obs}, \sigma_{obs}^2)$$
$$\mu_{obs} = 43, \quad \sigma_{obs}^2 = 5$$

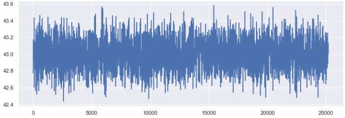
Results

$$\mathcal{N}(\mu_n, \sigma_n^2), \quad \mu_n = \frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma_{obs}^2}} \cdot \left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum y_{obs}}{\sigma_{obs}^2} \right) \simeq 42.99, \quad \sigma_n^2 = \frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma_{obs}^2}} \simeq 0.025$$

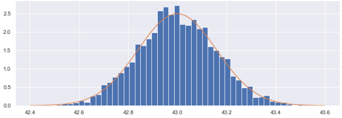
Coupled chains



Complete sampling



Sampling histogram



Time Averaged Estimators mean:

$$\mathbb{E}[H_{k:m}(X, Y)] = 42.9498$$

Approximate Bayesian Computation

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

ABC Metropolis Hastings

Inputs:

- a target posterior density $\pi(\theta|y_{obs}) \propto p(y_{obs}|\theta)\pi(\theta)$, consisting of a prior distribution $\pi(\theta)$ and a procedure of generating data under the model $p(y_{obs}|\theta)$;
- a Markov proposal density $g(\theta, \theta')$=$g(\theta'|\theta)$;
- an integer $N > 0$;
- a kernel function $K_h(u)$ and a scale parameter $h > 0$;
- a low dimensional vector of summary statistics $s = S(y)$.

Initialise:

repeat:

1. choose an initial parameter vector $\theta^{(0)}$ from the support of $\pi(\theta)$;
2. generate $y^{(0)} \sim p(y|\theta^{(0)})$ from the model and compute summary statistics $s^{(0)} = S(y^{(0)})$, until $K_h(\| s^{(0)} - s_{obs} \|) > 0$.

ABC Metropolis Hastings

Sampling for $i = 1, ..., N$:

1. generate candidate vector $\theta' \sim g(\theta^{(i-1)}, \theta)$ from the proposal density g;
2. generate $y' \sim p(y|\theta')$ from the model and compute summary statistics $s' = S(y')$;
3. with probability

$$\min\{1, \frac{K_h(\| s' - s_{obs} \|)\pi(\theta')g(\theta', \theta^{(i-1)})}{K_h(\| s^{(i-1)} - s_{obs} \|)\pi(\theta^{(i-1)})g(\theta^{(i-1)}, \theta')}\}$$

set $(\theta^{(i)}, s^{(i)}) = (\theta', s')$. Otherwise set $(\theta^{(i)}, s^{(i)}) = (\theta^{(i-1)}, s^{(i-1)})$.

Output:

- a set of correlated parameter vectors $\theta^{(1)}, ..., \theta^{(N)}$ from a Markov chain with stationary distribution $\pi_{ABC}(\theta|S_{obs})$.

Summary statistic:

Sample mean

Distance:

2-norm of the difference.

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

Same as previous: - DA RIVEDERE

Model

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

$$\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$$
$$\mu_0 = 38, \quad \sigma_0^2 = 4$$

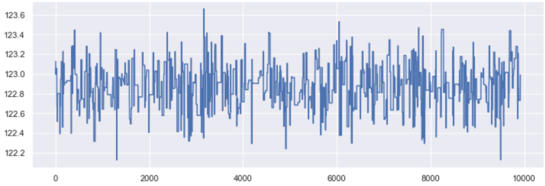
Dataset

1000 samples generated from a Gaussian distribution:

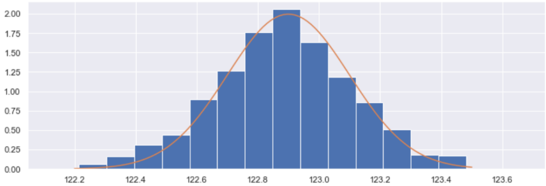
$$Y_{obs} \sim \mathcal{N}(\mu_{obs}, \sigma^2_{obs})$$
$$\mu_{obs} = 43, \quad \sigma^2_{obs} = 5$$

Results

Sampling

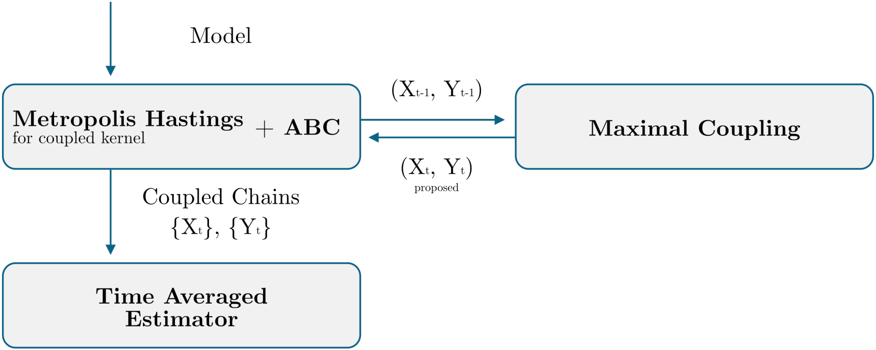


Sampling histogram with real distribution



The complete method: MCMC + Couplings + ABC

Implementation



Metropolis Hastings with couplings and ABC

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

Metropolis Hastings with couplings and ABC

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

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

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

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

Metropolis Hastings with couplings and ABC

As output we get two sets of parameter vectors:

$$\theta_x^{(1)}, ..., \theta_x^{(N)} \sim \pi_{ABC}(\theta | y_{obs});$$

$$\theta_y^{(1)}, ..., \theta_y^{(N)} \sim \pi_{ABC}(\theta | y_{obs}).$$

Study case

Summary statistic:

Sample mean

Distance:

2-norm of the difference.

Kernel:

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

DA SISTEMARE

Model

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

$$\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$$
$$\mu_0 = 38, \quad \sigma_0^2 = 4$$

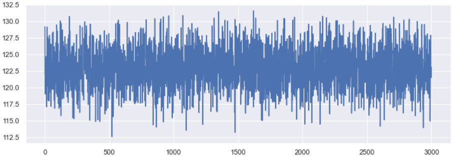
Dataset

1000 samples generated from a Gaussian distribution:

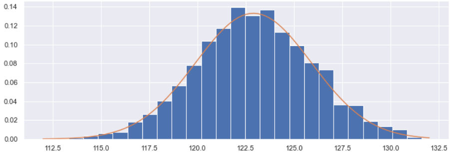
$$Y_{obs} \sim \mathcal{N}(\mu_{obs}, \sigma^2_{obs})$$
$$\mu_{obs} = 43, \quad \sigma^2_{obs} = 5$$

Results

Sampling



Sampling histogram with real distribution



Conclusions

Next steps

The next step will be the conclusion of the **multivariate implementation** the MCMC with couplings and approximate bayesian computation.

Further steps will be testing on more complex data.

Bibliography

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.