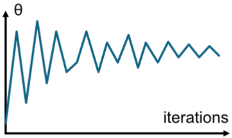
Coupled Markov chains with applications to Approximate Bayesian Computation for model based clustering

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

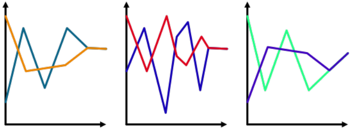
10 January 2022

A complex problem



⇓

Unbiased Markov chain Monte Carlo methods with couplings



⇓

Approximate Bayesian Computation

Approximate Bayesian Computation

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 $K_h(\| s^{(0)} - s_{obs} \|) > 0$.

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 $K_h(\| s^{(0)} - s_{obs} \|) > 0$.

- a kernel function $K_h(u)$ and a scale parameter $h > 0$:

$$\pi(\theta, y|y_{obs}) \propto \mathbb{1}(\| y - y_{obs} \| \leq h) p(y|\theta)\pi(\theta)$$

$$\Downarrow$$

$$\pi_{ABC}(\theta, y|y_{obs}) \propto K_h(u) p(y|\theta)\pi(\theta)$$

Where K is a standard smoothing kernel function and:

$$K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right), \quad \text{with} \quad u = \| y - y_{obs} \|$$

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

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 $K_h(\| s^{(0)} - s_{obs} \|) > 0$.

- a low dimensional vector of summary statistics $s = S(y)$:

$$K_h(\| y - y_{obs} \|)$$

$$\Downarrow$$

$$K_h(\| S(y) - S(y_{obs}) \|)$$

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 $K_h(\| s^{(0)} - s_{obs} \|) > 0$.

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, vector of 9 quantiles

Distance:

2-norm of the difference

Kernel:

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

Model

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

$$\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$$
$$\mu_0 = 8, \quad \sigma_0^2 = 4$$

Dataset

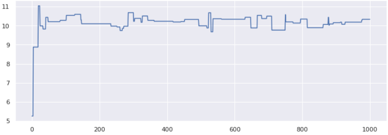
100 samples generated from a Gaussian distribution:

$$Y_{obs} \sim \mathcal{N}(\mu_{obs}, \sigma^2_{obs})$$
$$\mu_{obs} = 10, \quad \sigma^2_{obs} = 3$$

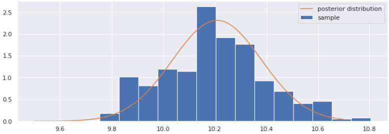
Posterior distribution:

$$\mathcal{N}(\mu_n, \sigma_n^2), \mu_n = \frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma_{obs}^2}} \cdot \left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum y_{obs}}{\sigma_{obs}^2} \right) \simeq 10.151, \sigma_n^2 = \frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma_{obs}^2}} \simeq 0.0298$$

Sampling

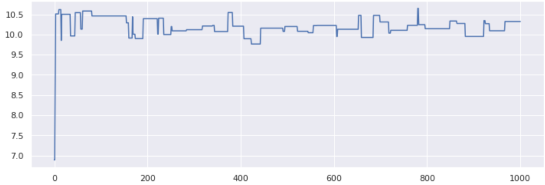


Sampling histogram with real distribution

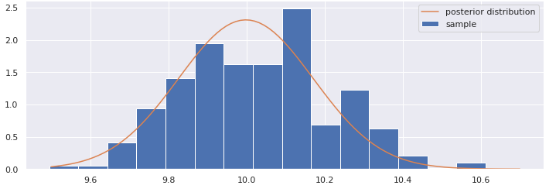


The same model using as summary statistic a vector of 9 quantiles:

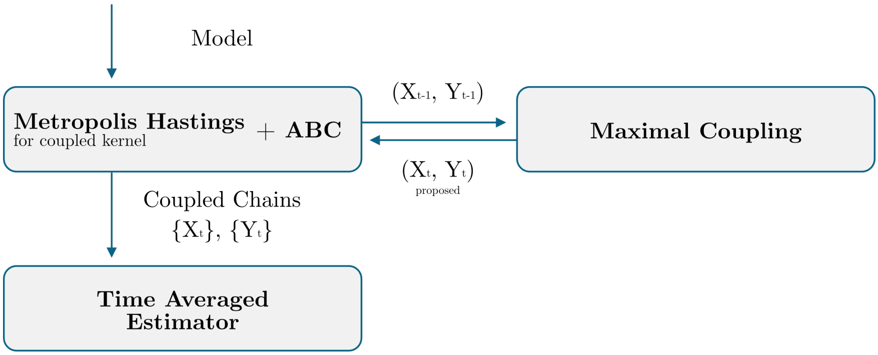
Sampling



Sampling histogram with real distribution



The complete method: MCMC + Couplings + ABC


Model
Metropolis Hastings
for coupled kernel
+ ABC
(X_{t-1}, Y_{t-1})
Maximal Coupling
(X_t, Y_t)
proposed
Coupled Chains
$\{X_t\}, \{Y_t\}$
Time Averaged
Estimator

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 $y_{1i} \sim \mathcal{N}(\theta_x^{(0)}, \sigma_{obs}^2)$ and $y_{2j} \sim \mathcal{N}(\theta_y^{(0)}, \sigma_{obs}^2)$;
4. compute $s_x^{(0)} = S(y_1)$ and $s_y^{(0)} = S(y_2)$;
5. until $K_h(||s_x^{(0)} - s_{obs}||) > 0$:
    - generate $\theta_x^{(0)} \sim \pi(\mu)$ from prior density;
    - generate a sample of N observations such that $y_{1i} \sim \mathcal{N}(\theta_x^{(0)}, \sigma_{obs}^2)$;
    - compute $s_x^{(0)} = S(y_1)$;
6. until $K_h(||s_y^{(0)} - s_{obs}||) > 0$:
    - generate $\theta_y^{(0)} \sim \pi(\mu)$ from prior density;
    - generate a sample of N observations such that $y_{2j} \sim \mathcal{N}(\theta_y^{(0)}, \sigma_{obs}^2)$;
    - compute $s_y^{(0)} = S(y_2)$;

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

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

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

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

As output we get two sets of parameter vectors:

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

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

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

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

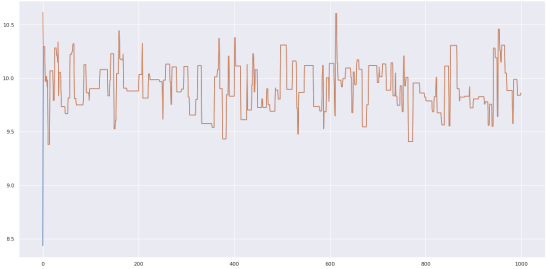
$$\mu \sim \mathcal{N}(\mu_0,\sigma_0^2)$$
$$\sigma^2 \sim InvGa(a,b)$$
$$\pi(\mu,\sigma) = \pi(\mu) * \pi(\sigma)$$
$$\mu_0 = 34, \quad \sigma_0^2 = 3$$
$$\text{a=b=1}$$

Dataset

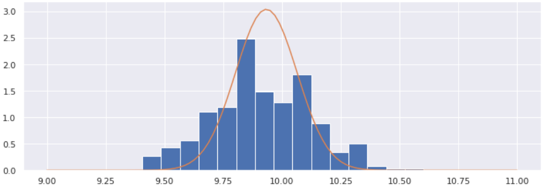
100 samples generated from a Gaussian distribution:

$$Y_{obs} \sim \mathcal{N}(\mu_{obs},\sigma_{obs}^2)$$
$$\mu_{obs} = 30, \quad \sigma_{obs}^2 = 2$$

Coupled chains



Sampling histogram with real distribution



Conclusions

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

Further steps will be implementing the version with unknown variance and testing on more complex data.
Finally, making comparisons with a standard MCMC algorithm.

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.