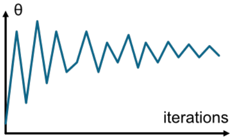

# Coupled Markov chains with applications to Approximate Bayesian Computation for model based clustering

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

15 February 2022

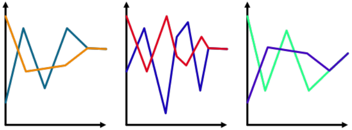
# A complex problem






⇓

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



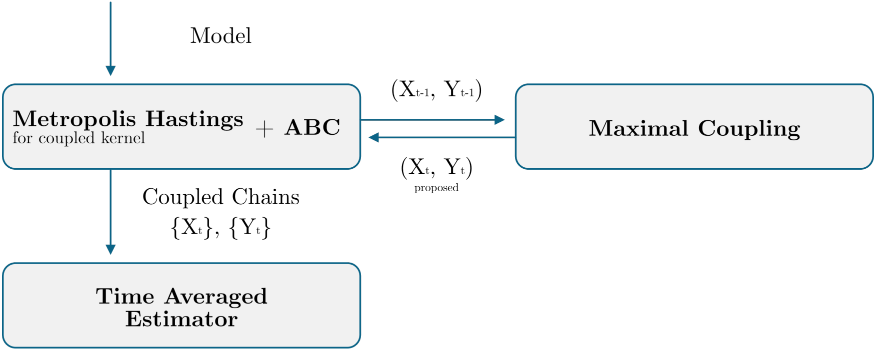

⇓

**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 $W^\star|Y \sim \mathcal{U}\{[0, q(Y)]\}$ until $W^\star > p(Y)$
  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

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

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

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

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

$$a = b = 1$$

# Metropolis Hastings with couplings and ABC



## Initilization

set $\boldsymbol{\theta} = (\mu, \sigma^2)$
for k=1,2:
until $K_h(||s_k^{(0)} - s_{obs}||) > 0$:

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

# Metropolis Hastings with couplings and ABC



## Iterations

for i = 1,...,M:
given $\boldsymbol{\theta}^{(i-1)} = (\mu^{(i-1)}, \sigma^{2(i-1)})$

- generate $[\boldsymbol{\theta}_1^{(i)}, \boldsymbol{\theta}_2^{(i)}]$ from a maximalcoupling given $[\boldsymbol{\theta}_1^{(i-1)}, \boldsymbol{\theta}_2^{(i-1)}]$;
- 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 $\boldsymbol{\theta}_k^{(i)}$ with probability

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

## Metropolis Hastings with couplings and ABC



### Output

for k= 1,2

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

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

## Study case



**Summary statistics**: Sample mean, Sample Variance

**Distance**: $L^2 - norm$

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

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

# Study case



**Summary statistics**: Sample mean, Sample Variance
**Distance**: $L^2 - norm$
**Kernel**: $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}, \quad K_h(u) = \frac{K(\frac{u}{h})}{h}$
**Proposal distribution**: $\begin{bmatrix} \mu' \\ log(\sigma^{2\prime}) \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} \mu^{(i)} \\ log(\sigma^{2(i)}) \end{bmatrix}, 0.1^2 \cdot \mathcal{I} \right)$

## Dataset

100 samples generated from a Gaussian distribution:

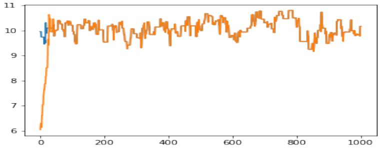
$$Y_{obs} \sim \mathcal{N}(\mu_{obs}, \sigma^2_{obs})$$

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

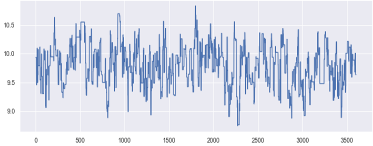
# Results for parameter $\mu$



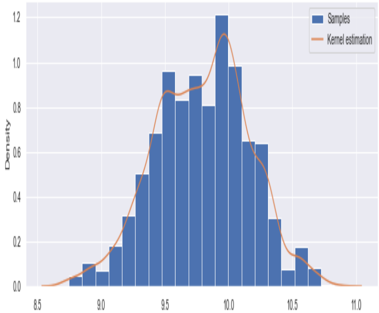
**Coupled chains from one processor**




**Samplings from all accepted chains**




**Sampling histogram with estimated distribution**

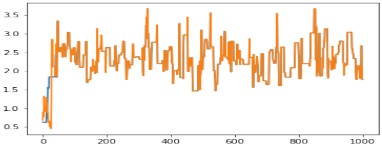



Time Averaged Estimators mean obtained: $\mathbb{E}[H_{k:m}(X_\mu, Y_\mu)] \simeq 9.8$

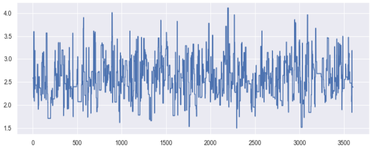
# Results for parameter $\sigma^2$



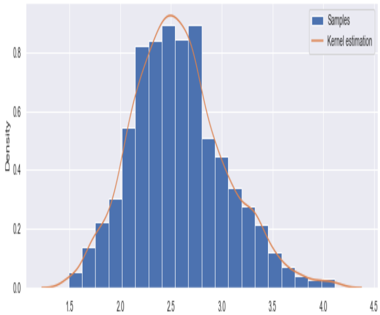
**Coupled chains from one processor**




**Samplings from all accepted chains**




**Sampling histogram with estimated distribution**




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

# 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.

## Prior distributions

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

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

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

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

## Study case



**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

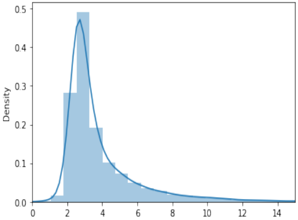
100 samples generated as follows:

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

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

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

**Quantile distribution:** $y_{obs}$



# Metropolis Hastings with couplings and ABC



## 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 $y_k \sim quantilefunction(z(r), \theta_k^0)$
- Compute $s_k^{(0)} = S(y_k)$

# Metropolis Hastings with couplings and ABC



## Iterations

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

- generate $\theta_1^{(i)}, \theta_2^{(i)}$ from 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 $\theta_k^{(i)} = \theta_k^{(i-1)}$;

# Metropolis Hastings with couplings and ABC



## Iterations

fo

### Couplings

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

1. $z^{(i,j)}(r) \sim \mathcal{N}(0,1)$
2. generate:

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

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

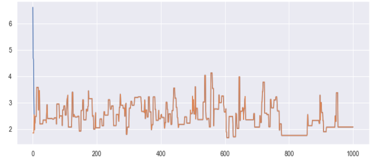
$Kh(\|s_k - s_{obs}\|)\pi(\theta_k)$

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

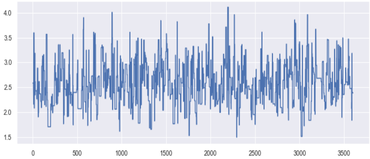
# Results for parameter a



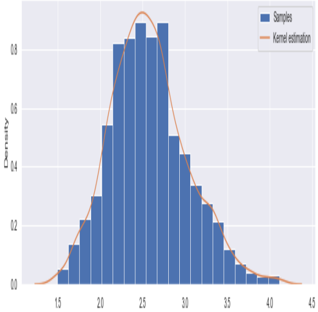
**Coupled chains from one core**




**Samplings from all accepted chains**




**Sampling histogram with estimated distribution**



# Bibliography