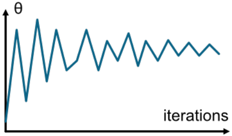
Coupled Markov chains with applications to Approximate Bayesian Computation for model based clustering

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

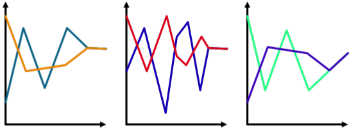
10 January 2022

A complex problem



⇓

Unbiased Markov chain Monte Carlo methods with couplings



⇓

Approximate Bayesian Computation

Unbiased Markov chain Monte Carlo methods with couplings

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

The road to parallelization: coupling of Markov chains

Faster MCMC $\Longrightarrow$ Parallelization

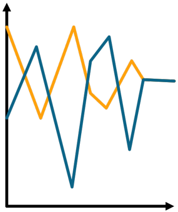
The road to parallelization: coupling of Markov chains

Faster MCMC $\Longrightarrow$ Parallelization $\Longleftrightarrow$ **Unbiased estimator**

The road to parallelization: coupling of Markov chains

Faster MCMC $\Longrightarrow$ Parallelization $\Longleftrightarrow$ **Unbiased estimator**

Exact estimations algorithms using **coupling of Markov chain**.



Rhee–Glynn estimator I

The goal is to estimate

$$\mathbb{E}_\pi[h(X)] = \int h(x)\pi(\mathrm{d}x).$$

The estimator we are going to construct is based on a coupled pair of Markov chains, $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq}$, which marginally start from π_0 and evolve accordingly to P.

Rhee–Glynn estimator II

We consider some assumptions:

1. as $t \to \infty$,
$$\mathbb{E}[h(X_t)] \to \mathbb{E}_\pi[h(X)];$$
and there exists $\eta > 0$ and $D < \infty$ such that $\mathbb{E}[|h(X_t)|^{2+\eta}] \leq D$ for all $t \geq 0$;

Rhee–Glynn estimator II

We consider some assumptions:

1. as $t \to \infty$,
$$\mathbb{E}[h(X_t)] \to \mathbb{E}_\pi[h(X)];$$
and there exists $\eta > 0$ and $D < \infty$ such that $\mathbb{E}[|h(X_t)|^{2+\eta}] \leq D$ for all $t \geq 0$;
2. the chains are such that the meeting time
$$\tau = \inf\{t \geq 1 : X_t = Y_{t-1}\}$$
satisfies $\mathbb{P}(\tau > t) \leq C\delta^t$ for all $t \geq 0$, for some constants $C < \infty$ and $\delta \in (0, 1)$;

Rhee–Glynn estimator II

We consider some assumptions:

1. as $t \to \infty$,
$$\mathbb{E}[h(X_t)] \to \mathbb{E}_\pi[h(X)];$$
and there exists $\eta > 0$ and $D < \infty$ such that $\mathbb{E}[|h(X_t)|^{2+\eta}] \leq D$ for all $t \geq 0$;
2. the chains are such that the meeting time
$$\tau = \inf\{t \geq 1 : X_t = Y_{t-1}\}$$
satisfies $\mathbb{P}(\tau > t) \leq C\delta^t$ for all $t \geq 0$, for some constants $C < \infty$ and $\delta \in (0, 1)$;
3. the chains stay together after meeting:
$$X_t = Y_{t-1} \text{ for all } t \geq \tau.$$

Rhee–Glynn estimator III

Thanks to the previous assumptions we can prove that:

$$\mathbb{E}_\pi[h(X)] = \mathbb{E}[h(X_k) + \sum_{t=k+1}^{\tau-1} \{h(X_t) - h(Y_{t-1})\}];$$

Rhee–Glynn estimator III

Thanks to the previous assumptions we can prove that:

$$\mathbb{E}_\pi[h(X)] = \mathbb{E}[h(X_k) + \sum_{t=k+1}^{\tau-1} \{h(X_t) - h(Y_{t-1})\}];$$

and we define the Rhee–Glynn estimator as:

$$H_k(X, Y) = h(X_k) + \sum_{t=k+1}^{\tau-1} \{h(X_t) - h(Y_{t-1})\}$$

which is **unbiased** by construction.

Time-averaged estimator I

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

Time-averaged estimator I

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

- $MCMC_{k:m}$ is the standard MCMC average;

Time-averaged estimator I

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

- $MCMC_{k:m}$ is the standard MCMC average;
- $BC_{k:m}$ is the bias correction;

Time-averaged estimator II

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:

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

Time-averaged estimator II

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 $(X_{t+1}, Y_t) \sim \bar{P}\{(X_t, Y_{t-1}), \cdot\}$; $\bar{P}$ must be evaluated before!
   b. set $t \leftarrow t + 1$;
3. compute the time-averaged estimator:

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

Time-averaged estimator III

Metropolis–Hasting algorithm allow us to calculate the coupled kernel $\bar{P}\{(X_t, Y_{t-1}), \cdot\}$:

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 $Y_t = Y_{t-1}$.

Time-averaged estimator III

Metropolis–Hasting algorithm allow us to calculate the coupled kernel $\bar{P}\{(X_t, Y_{t-1}), \cdot\}$:

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 $Y_t = Y_{t-1}$.

Approximate Bayesian Computation

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

POLITECNICO MILANO 1863

ABC rejection sampling algorithm

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 proposal density $g(\theta)$, with $g(\theta) > 0$ if $\pi(\theta|y_{obs}) > 0$;
- an integer $N > 0$;
- a kernel function $K_h(u)$ and a scale parameter $h > 0$;
- a low dimensional vector of summary statistics $s = S(y)$.

ABC rejection sampling algorithm

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

1. generate $\theta^{(i)} \sim g(\theta)$ from sampling density g;
2. generate $y \sim p(y|\theta^{(i)})$ from the likelihood;
3. compute summary statistic $s = S(y)$;
4. accept $\theta^{(i)}$ with probability $\frac{K_h(\|s - s_{obs}\|)\pi(\theta^{(i)})}{Kg(\theta^{(i)})}$, where $K \geq K_h(0) \max_\theta \frac{\pi(\theta)}{g(\theta)}$; else go to 1.

Output:

- a set of parameter vectors $\theta^{(1)}, ..., \theta^{(N)} \sim \pi_{ABC}(\theta|S_{obs})$.

Conclusions

Next steps

The next step will be the conclusion of the **separate multivariate implementation** of both solution to be tested on simulated data and the **parallelized multivariate implementation**.

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.