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

Gabrielli - Di Filippo - Caldarini - Musiari - Bertoni

November 9, 2021

Coupled
Markov
chains with
applications
to Approxi-
mate
Bayesian
Computa-
tion for
model
based
clustering

Gabrielli -
Di Filippo -
Caldarini -
Musiari -
Bertoni

A complex problem

In order to speed up computation time, P. Jacob, J. O'Leary and Y. Atchadé proposed in 2020 a way parallelize Monte Carlo algorithm using Markov chain couplings.

Coupled
Markov
chains with
applications
to Approxi-
mate
Bayesian
Computa-
tion for
model
based
clustering

Gabrielli -
Di Filippo -
Caldarini -
Musiari -
Bertoni

The road to parallelization

Standard Markov chain Monte Carlo methods are potentially biased for any fixed numbers of iterations.

The idea is to slit up the long Markov chain into many shorter coupled Markov chains; to allow this we need to start from an unbiased MCMC method.

P. Glynn and C. Rhee proposed in 2018 the class of exact estimations algorithms.

Coupled
Markov
chains with
applications
to Approxi-
mate
Bayesian
Computa-
tion for
model
based
clustering

Gabrielli -
Di Filippo -
Caldarini -
Musiari -
Bertoni

What is Markov chains coupling?

Markov chain coupling allow to reduce the convergence time.
(Completare con una spiegazione delle note)

Coupled
Markov
chains with
applications
to Approxi-
mate
Bayesian
Computa-
tion for
model
based
clustering

Gabrielli -
Di Filippo -
Caldarini -
Musiari -
Bertoni

Rhee-Glynn estimator

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.
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.

3.

Coupled
Markov
chains with
applications
to Approxi-
mate
Bayesian
Computa-
tion for
model
based
clustering

Gabrielli -
Di Filippo -
Caldarini -
Musiari -
Bertoni

Rhee-Glynn estimator

Thanks to the previous assumptions we can prove that:

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

$$\cdots$$

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

The Rhee-Glynn estimator is:

$$\cdots$$

which is unbiased by construction.

Coupled
Markov
chains with
applications
to Approxi-
mate
Bayesian
Computa-
tion for
model
based
clustering

Gabrielli -
Di Filippo -
Caldarini -
Musiari -
Bertoni

Time-averaged estimator

The Rhee-Glynn has to be make computationally feasible, the time-averaged estimator keeps the assumptions allowing the computation.

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

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

- $BC_{k:m}$ is the bias correction

- $k-1$ number of burn-in iterations;

- m is a fixed integer of number of maximum iterations,

- τ is a random variable representing the meeting time:

$$\tau = \inf\{t \geq 1 : X_t = Y_{t-1}\}$$

Coupled
Markov
chains with
applications
to Approxi-
mate
Bayesian
Computa-
tion for
model
based
clustering

Gabrielli -
Di Filippo -
Caldarini -
Musiari -
Bertoni

Application to Metropolis-Hasting algorithm

Both GS and MH satisfy the assumptions. Here how to adapt MH with coupling:

Step 1: Draw X_0 and Y_0 from an initial distribution π_0 and draw $X_1 \sim P(X_0, \cdot)$;

Step 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 $t \leftarrow t + 1$;

Step 3: compute

$$H_{k:m}(X, Y)$$

using the previous formula.

The following is the algorithm to calculate the coupled kernel $\bar{P}\{(X_t, Y_{t-1}), \cdot\}$ via MH:

Step 1: Sample $(X^\star, Y^\star)|(X_t, Y_{t-1})$ from a maximal coupling of $q(X_t, \cdot)$ and $q(Y_{t-1}, \cdot)$;

Step 2: sample $U \sim \mathcal{U}([0, 1])$;

Step 3: if

$$U \le \min\left\{1, \frac{\pi(X^\star)q(X^\star, X_t)}{\pi(X_t)q(X_t, X^\star)}\right\}$$

then $X_{t+1} = X^\star$; otherwise $X_t = X_{t-1}$.

Step 4: if

$$U \le \min\left\{1, \frac{\pi(Y^\star)q(Y^\star, Y_t)}{\pi(Y_t)q(Y_t, Y^\star)}\right\}$$

then $Y_{t+1} = Y^\star$; otherwise $Y_t = Y_{t-1}$.