THE IMPORTANCE MARKOV CHAIN

Hugo MARIVAL

Doctorant SAMOVAR/CITI

Auteurs : C.Andral, R.Douc, H.Marival, C.Robert

FRAMEWORK

Goal. Sample « according » to a *target distribution* π on $\mathbb{X}$ that does not work well with MCMC algorithms.

E.g. Slow space exploration for *multimodal distributions* which get stuck in a mode.

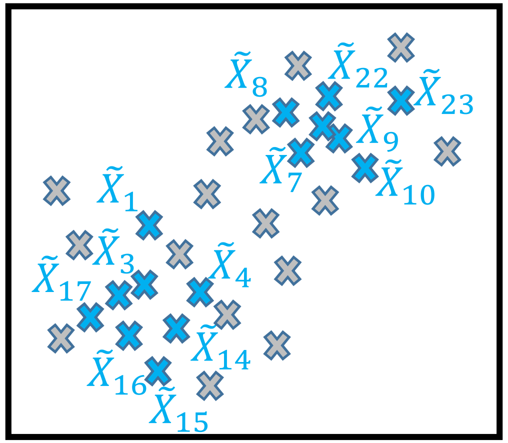
Idea. Sample from an instrumental distribution $\tilde{\pi}$ allowing better exploration, then transform the sample to get a new one distributed $\sim \pi$.

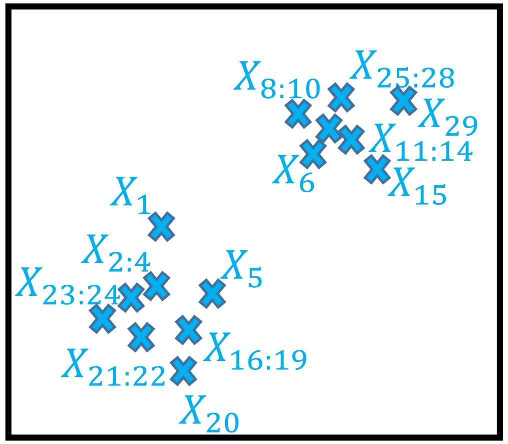
Which kernel ?
Markov chain output ?

$$\tilde{\pi} Q = \tilde{\pi} \qquad \pi {\color{red}P} = \pi$$

$$(\tilde{X}_k)_{k=1..K} \longrightarrow (X_i)_{i=1..I}$$

LLN ? CLT ?
Convergence speed ?





PLAN

1. Heuristics of the IMC algorithm
   i. Rejection
   ii. Rejection + MC
   iii. Rejection + MC + « discrete » IS

2. Proper IMC algorithm and main results
   i. Extended chain
   ii. Algorithm
   iii. Main results

3. Numerical experiments
   i. Improvements on multimodal target
   ii. Indep IMC vs. Indep MH

PRESENTATION AND HEURISTICS

REJECTION MC.

$\boxed{\pi \leq M\tilde{\pi} \Leftrightarrow \rho_\kappa \leq 1}$

⚠ M does not exist, too large, unknown !!

Algo. Perform the rejection algorithm on a sample $(\tilde{X}_k)_{k\in\mathbb{N}}$ generated by Q with acceptance function $\rho = \rho_M$.

Rejection kernel $\boldsymbol{S}$.

$$Sh(x) = \sum_{k=1}^{\infty} \mathbb{E}_x^Q \left[h(\tilde{X}_k)\rho(\tilde{X}_k) \prod_{i=1}^{k-1} (1-\rho(\tilde{X}_k)) \right]$$

$$\begin{tikzcd}
{} \arrow[r] & \tilde{X}_1 \arrow[r] \arrow[d] & \tilde{X}_2 \arrow[r] \arrow[d] & \tilde{X}_3 \arrow[r] \arrow[d] & \tilde{X}_4 \arrow[r] \arrow[d] & {} \\
& X_1 = \tilde{X}_1 & X_2 = \tilde{X}_2 & {} & X_3 = \tilde{X}_4 &
\end{tikzcd}$$

HEURISTICAL IMC.

$\boxed{\pi \nleq M\tilde{\pi} \Leftrightarrow \rho_\kappa \nleq 1}$

Idea. Use a *repetition kernel* $\tilde{R}$ to create a sample of repeated data points according to the density ratio :

$$\rho_\kappa(x) = \int n\tilde{R}(x, dn)$$

Remark. 0 repetitions corresponds to a rejection !

$$\begin{tikzcd}
{} \arrow[r] & \tilde{X}_1 \arrow[r] \arrow[d, "\tilde{N}_1 = 3"] & \tilde{X}_2 \arrow[r] \arrow[d, "\tilde{N}_2 = 1"] & \tilde{X}_3 \arrow[r] \arrow[d, "\tilde{N}_3 = 0"] & \tilde{X}_4 \arrow[r] \arrow[d, "\tilde{N}_4 = 2"] & {} \\
& X_1, X_2, X_3 & X_4 & {} & X_5, X_6 &
\end{tikzcd}$$

THE IMC ALGORITHM

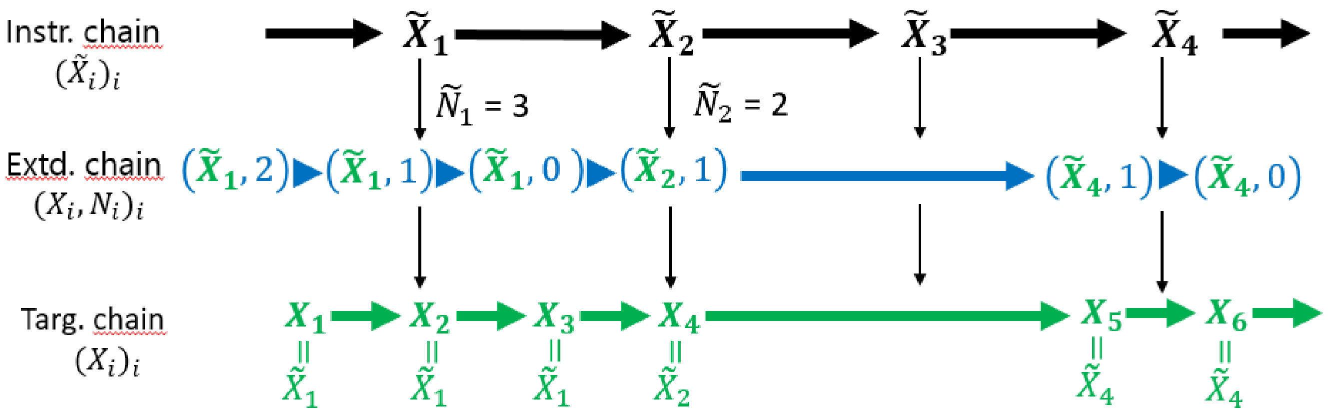
Goal. Create a MC from the ideas developed previously.

Idea. Create *extended chain* $(X_i, N_i)_{i\in\mathbb{N}}$ by adding remaining repetition counter N_i as second component and extract the sample from the first component at the end.

INIT. Set an arbitrary $\tilde{X}_0$ and $i = 0$.

REPEAT.

1. Draw $\tilde{X}_k \sim Q(\tilde{X}_{k-1}, \cdot)$ and $\tilde{N}_k \sim \tilde{R}(\tilde{X}_k, \cdot)$
2. Set $N_i = \tilde{N}_k$
3. While $N_i \geq 1$:
   a. $(X_i, N_i) \leftarrow (\tilde{X}_k, N_i - 1)$;
   b. $i \leftarrow i + 1$;



MAIN RESULTS

Result on P.

Hyp on $\boldsymbol{Q}$ and $\tilde{\boldsymbol{R}}$.

Invariant measure. P admits an invariant probability measure $\bar{\pi}$, which has π as its first marginal. Moreover, $\bar{\pi}$ is unique (under mild additional conditions).

Q targets $\tilde{\pi}$ + mild addional conditions

SLLN. For any $\xi \in M_1(\mathbb{X})$ and h such that $\pi(h) < \infty$,

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} h(X_k) = \pi(h), \qquad \mathbb{P}_\xi^P - a.s.$$

SLLN for Q

CLT. For any $\xi \in M_1(\mathbb{X})$ and h such that $\pi(h) < \infty$, there exists $\sigma^2(h) > 0$ and $\chi \in M_1(\mathbb{X})$ s.t.

$$n^{-1/2} \sum_{i=1}^{n} (h(X_i) \ - \pi h) \rightsquigarrow^{\mathcal{L}} \mathcal{N}\big(0, \sigma^2(h)\big), \qquad \mathbb{P}_\chi^P - law.$$

1. Q admits a solution to the Poisson equation associated to $\rho_\kappa h_0$ where $h_0 = h - \pi(h)$.
2. Mild assumptions on $\tilde{R}$.

Geom ergod. There exist constants $\delta, \beta_r > 1$ and $\zeta < \infty$ such that for all $\xi' \in M_1(\mathbb{X} \times \mathbb{N})$,

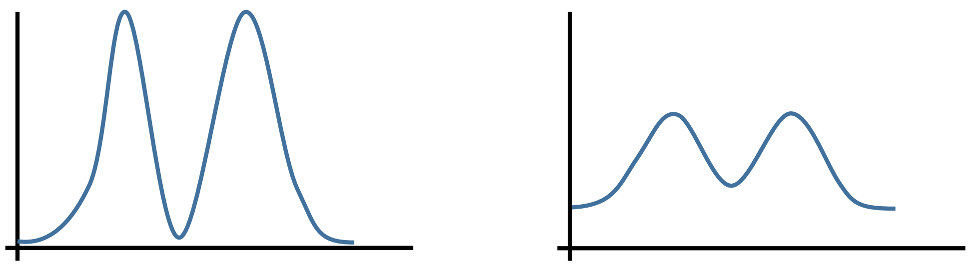
$$\sum_{k=1}^{\infty} \delta^k \, d_{TV}\big(\xi P^k, \bar{\pi}\big) \le \zeta \int_{\mathbb{X}\times\mathbb{N}} \beta_r^n \, V(x) \xi'(dxdn)$$

1. Small set for Q
2. Drift condition on that set for a function V

NUMERICAL EXPERIMENTS

A. Imrovements on multimodal target.

Idea. Sample smoothened version of multimodal distribution with Q then transform sample into original multimodal target distrib sample with IMC.



Question. Benefits of targetting smoothened distribution vs original ?

Setup. We target $\pi = \sum_{i=1}^{6} \mathcal{N}(\mu_i, I_5)$ an unnormalized gauss. mix., with $\mu_i \sim \mathcal{N}(0{,}10^2 I_5)$

- $\tilde{\pi} = \pi^{\beta}$ for $\beta \in (0{,}1)$
- Q is a No-U-turn Sampler (NUTS) and $\tilde{R}$ is a shifted Bernouilli kernel

We estimate the MSE of π by running 200 chains for each $\beta \in \{0{,}004, 0{,}01, 0{,}04, 0{,}1{,}1\}$

β	0.004	0.010	0.040	0.100	1.000
MSE	16.150	6.123	**0.544**	17.863	33.982

NUMERICAL EXPERIMENTS

B. Indep IMC vs Indep MH

Setup. We target $\pi = \frac{1}{2}\mathcal{N}(0,4) + \frac{1}{2}\mathcal{N}(0,4)$ using :

- $\tilde{\pi} = \mathcal{N}(0,4)$
- $Q(x,\cdot) = \tilde{\pi}(\cdot)$ for all $x \in \mathbb{X}$
- $\tilde{R}$ a shifted Bernouilli kernel

	X	X^2	X^3	X^4
MH	6.20e-03	2.33e-02	1.49e+00	1.57e+01
IMC	3.49e-03	9.74e-03	8.40e-01	7.18e+00

Table. MSE for the first 4 moments for 10^4 repetitions of chains of length 10^4.