THE IMPORTANCE MARKOV CHAIN

Auteurs

Charly ANDRAL
Randal DOUC
Hugo MARIVAL
Christian ROBERT

① FRAMEWORK

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

Idea. Transform a sample $(\tilde{X}_k)_{k=1..K}$ from an *instrumental distribution* $\tilde{\pi}$ into a sample $(X_i)_{i=1..I}$ distributed under π.

Known techniques and drawbacks.
- *Rejection sampling* : need to know a constant M such that $\pi \leq M\tilde{\pi}$.
- *Importance sampling* : doesn't give a « real » sample but just a weighted sample used to compute estimates.

Markov chain approach. Instrumental and target samples do not have to be i.i.d. but can be generated by Markov chains targetting π and $\tilde{\pi}$ resp.

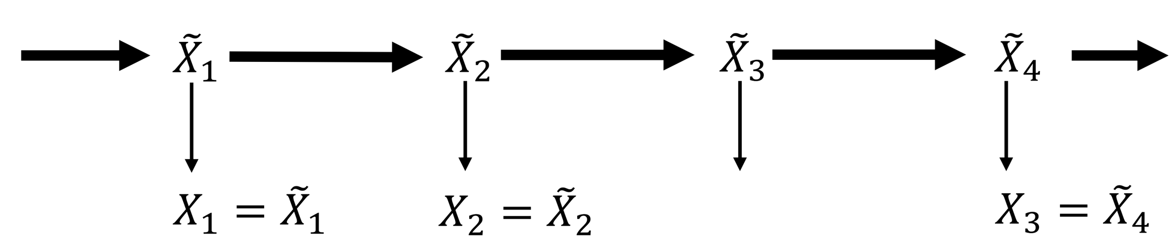
Notations. Q : Markov kernel targetting $\tilde{\pi}$, $\rho_\kappa = \kappa \frac{d\pi}{d\tilde{\pi}}$ density ratio, κ : tuning param.

② PRESENTATION OF THE IMC

REJECTION MC

Constraint. Constant M such that $\pi \leq M\tilde{\pi}$... Problem if such a const. M does not exist, is unknown or is too large leading to excessive rejection rate.

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

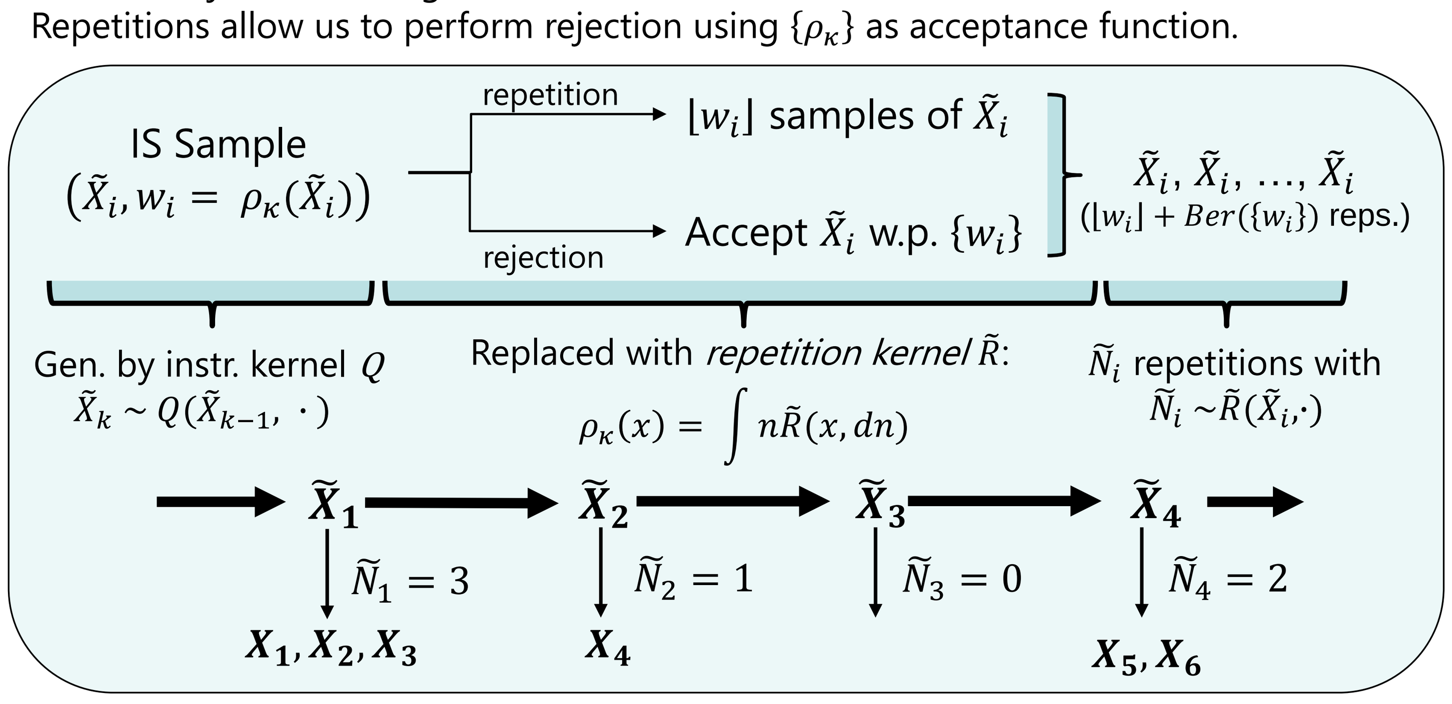


Output. A *rejection* chain $(X_i)_{i=1..I}$, generated by the submarkovian *rejection kernel* S defined, for $h \in F_+(\mathbb{X})$, by:

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

HEURISTICAL IMC

Idea. Remove constraint $\pi \not\leq M\tilde{\pi}$ i.e. allowing $\rho_\kappa \geq 1$ using repetitions in conjunction with the Rejection MC algorithm.
Repetitions allow us to perform rejection using $\{\rho_\kappa\}$ as acceptance function.



Choice of $\tilde{R}$. The optimal choice for $\tilde{R}$ in terms of variance reduction is the *shifted Bernouilli* :

$$\tilde{R}_{opt}(x,\cdot) = (1 - \{\rho_\kappa(x)\})\delta_{\lfloor \rho_\kappa(x) \rfloor} + \{\rho_\kappa(x)\}\delta_{\lfloor \rho_\kappa(x) \rfloor + 1}$$

A **MH algorithm** with proposal $A(x, dy)$ and acceptance rate $\alpha(x,y)$ targetting π can be obtained from the IMC by taking $\tilde{R}(x,\cdot) = Geom(p(x))$ where $p(x) = \int_{\mathbb{X}} \alpha(x,y)A(x,dy)$.

③ THE IMC ALGORITHM

EXTENDED MARKOV CHAIN

Setup. Add remaining repetition counter as second component and extract the sample from the first component at the end.

Property. Extended chain $(X_i, N_i)_{i\in\mathbb{N}}$ is markovian with transition kernel :

$$Ph(x,n) = \mathbb{1}_{\{n\geq 1\}} h(x, n-1) + \mathbb{1}_{\{n=0\}} \sum_{n'\geq 1} \int_{\mathbb{X}} S(x, dx') \frac{\tilde{R}(x', n'+1)}{\rho_{\tilde{R}}(x')} h(x', n')$$

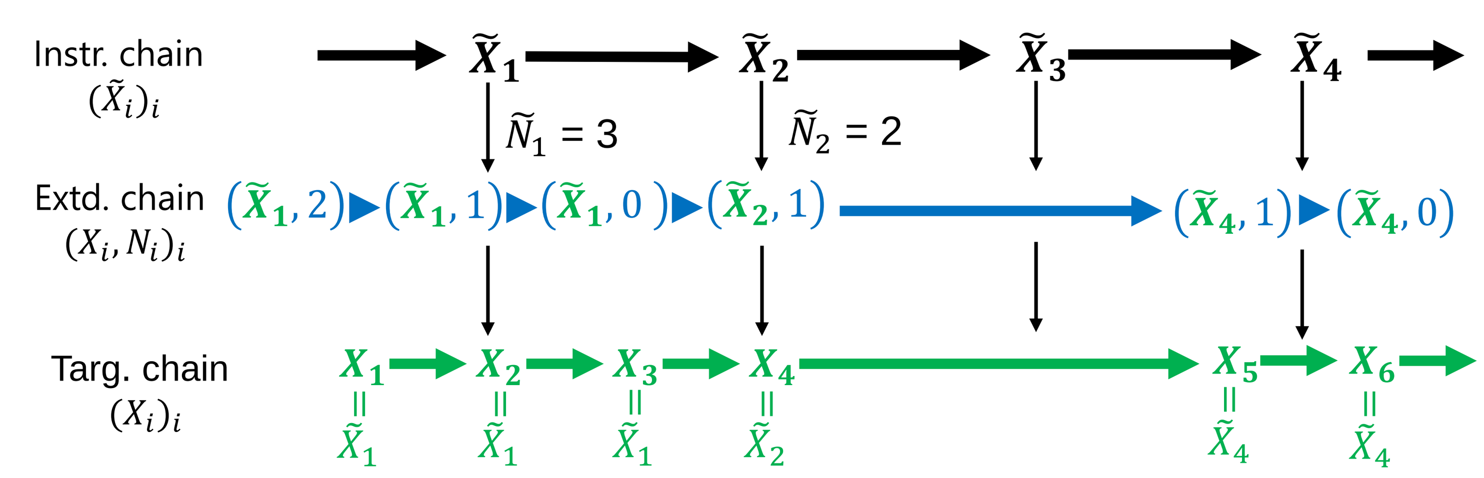
where $\rho_{\tilde{R}}(x') = \tilde{R}(x, [1{:}\infty))$.

Interpretation. While the counter is nonzero, decrease it by one. When it reaches zero :
- Use transition kernel S ensuring transitions from one *accepted* point to another
- Use *conditionned* transition kernel $R(x,n) = \frac{\tilde{R}(x, n+1)}{\rho_{\tilde{R}}(x)}$ taking into account that we only draw accepted points i.e. $\tilde{N} \geq 1$.

ALGORITHM AND MAIN RESULTS

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 $N_i \geq 1$: $(X_i, N_i) \leftarrow (\tilde{X}_k, N_i - 1)$; $i \leftarrow i + 1$;



Proposition [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).

Theorem [SLLN]. If Q admits a SLLN starting from any intital distribution, i.e. for any $\xi \in M_1(\mathbb{X})$ and g measurable function such that $\tilde{\pi}(g) < \infty$,

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} g(\tilde{X}_k) = \tilde{\pi}(g), \qquad \mathbb{P}_\xi^Q - a.s.$$

Then P also admits a SLLN starting from any initial distribution on $\mathbb{X} \times \mathbb{N}$.

Theorem [CLT]. Assume that Q admits a solution to the Poisson equation associated to $\rho_\kappa(h - \pi(h))$. Then, under mild additional assumptions on $\tilde{R}$, the kernel P admits a CLT for h, i.e. there exists a constant $\sigma^2(h) > 0$ and a distribution χ, easily expressed using ξ, and corresponding to the distribution of the first accepted couple, such that,

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

Theorem [Geometric ergodicity]. Assume a specific set is a small set for Q and a drift condition on that set for a function V. Then there exist constants $\delta, \beta_r > 1$ and $\zeta < \infty$ such that for all $\xi' \in M_1(\mathbb{X} \times \mathbb{N})$,

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

④ NUMERICAL EXPERIMENTS

IMPROVED SAMPLER

Question. Can IMC improve the effictiveness of a sampler targetting a multimodal distribution by using it on an instrumental smoothened version of the latter distrib. ?

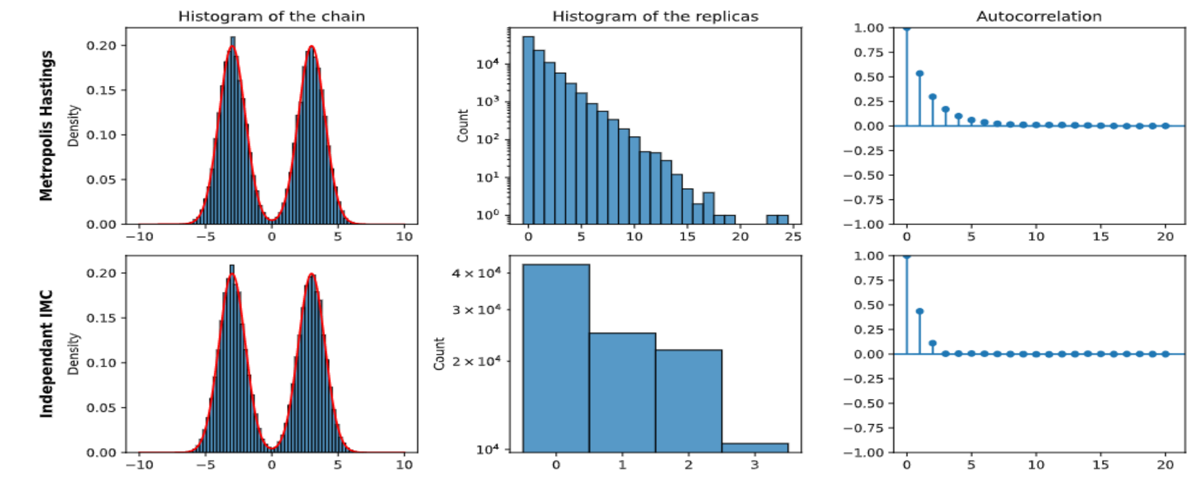
Setup. We target $\pi = \sum_{i=1}^n \mathcal{N}(\mu_i, I_d)$ an unnormalized gauss. mix., with $\mu_i \sim \mathcal{N}(0, 10^2 I_d)$
- $d = 5, n = 6$
- $\tilde{\pi} = \pi^\beta$ for $\beta \in (0,1)$ and $(x,\cdot) = \tilde{\pi}(\cdot)$ for all $x \in \mathbb{X}$
- Q is a No-U-turn Sampler (NUTS) and $\tilde{R}$ 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

INDEP. IMC VS. INDEP. MH ALGORITHM

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
- $N_{iterations} = 100{,}000$



Analysis. Both histograms of the chain look similar, but autocorrelation suggests better mixing for IMC. Also fewer max repetitions for IMC (reminder : geometric distrib. for MH).

	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.

Journée doctorants SAMOVAR
23 Mars 2023

Contact hugo.marival@telecom-sudparis.eu