Adaptive Restore algorithm & Importance Monte Carlo

Christian P. Robert
U. Paris Dauphine & Warwick U.

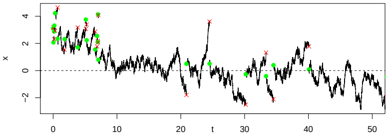
Workshop on Functional Inference and Machine Intelligence, U of Bristol

ERCall for applications

Ongoing 2023-2030 ERC funding for PhD and postdoctoral collaborations with

- Michael Jordan (Berkeley)
- Eric Moulines (Paris)
- Gareth O Roberts (Warwick)
- myself (Paris)

Part I: Adaptive Restore algorithm



Joint work with H McKimm, A Wang, M Pollock, and GO Roberts

`arXiv:2210.09901` – **Bernoulli**

MCMC regeneration

Reversibility attached with (sub-efficient) random walk behaviour

Recent achievements in non-reversible MCMC with PDMPs

[Bouchard-Côté et al., 2018; Bierkens et al., 2019]

Link with regeneration

[Nummelin, 1978; Mykland et al., X, 1995]

At regeneration times, Markov chain starts again: future independent of past

Scales poorly: regenerations recede exponentially with dimension

MCMC regeneration

Reversibility attached with (sub-efficient) random walk behaviour

Recent achievements in non-reversible MCMC with PDMPs

[Bouchard-Côté et al., 2018; Bierkens et al., 2019]

Link with regeneration

[Nummelin, 1978; Mykland et al., X, 1995]

At regeneration times, Markov chain starts again: future independent of past

Scales poorly: regenerations recede exponentially with dimension

PDMP: Marketing arguments

Current (?) default workhorse: reversible MCMC methods (incl. NUTS)

[Hoffman & Gelman, 2014]

Non-reversible MCMC algorithms based on piecewise deterministic Markov processes (aka PDMP) perform well empirically

PDMP: Marketing arguments

Non-reversible MCMC algorithms based on piecewise deterministic Markov processes (aka PDMP) perform well empirically

Quantitative convergence rates and variance now available

- Physics origins

  [Peters & De With, 2012; Krauth et al., 2009, 15, 16]

- geometric ergodicity for exponentially decaying tail target

  [Mesquita & Hespanha, 2010]

- ergodicity targets on the real line

  [Bierkens et al., 2016a,b]

Piecewise deterministic Markov process

PDMP sampler is a continuous-time, non-reversible, MCMC, method based on auxiliary variables

1. empirically state-of-the-art performances

   [Bouchard et al., 2017]

2. exact subsampling for big data

   [Bierkens et al., 2017]

3. geometric ergodicity for large class of distribution

   [Deligiannidis et al., 2017]

4. Ability to deal with intractable potential $\log \pi(x) = \int U_\omega(x)\mu(d\omega)$

   [Pakman et al., 2016]

Piecewise deterministic Markov process

Piecewise deterministic Markov process $\{z_t \in \mathcal{Z}\}_{t\in[0,\infty)}$, with three ingredients,

1. Deterministic dynamics: between events, deterministic evolution based on ODE

$$\mathrm{d}z_t/\mathrm{d}t = \Phi(z_t)$$

2. Event occurrence rate: $\lambda(t) = \lambda(z_t)$
3. Transition dynamics: At event time, τ, state prior to τ denoted by $z_{\tau-}$, and new state generated by $z_\tau \sim Q(\cdot|z_{\tau-})$.

[Davis, 1984, 1993]

Implementation hardships

Main difficulties of implementing PDMP come from

1. Computing the ODE flow Ψ: linear dynamic, quadratic dynamic
2. Simulating the inter-event time η_k: many techniques of superposition and thinning for Poisson processes

[Devroye, 1986]

Simulation by superposition plus thinning

Most implementations thru discrete-time schemes by sampling Bernoulli $B(\alpha(z))$
For

$$\Phi(z) = (x + v\varepsilon, v) \quad \text{and} \quad \alpha(z) = 1 \wedge \pi(x + v\varepsilon)/\pi(x)$$

sampling inter-event time for strictly convex $U(\cdot)$ obtained by solving

$$t^\star = \arg\min_t U(x + vt)$$

and additional randomization

- thinning: if there exists $\bar{\alpha}$ such that $\alpha(\Phi(z)) \geq \bar{\alpha}(x, k)$, accept-reject
- superposition and thinning: when $\alpha(z) = 1 \wedge \rho(\Phi(z))/\rho(z)$ and $\rho(\cdot) = \prod_i \rho_i(\cdot)$ then $\bar{\alpha}(z, k) = \prod_i \bar{\alpha}_i(z, k)$

[Bouchard et al., 2017]

Restore process

Take $\{Y_t\}_{t\geq 0}$ diffusion / jump process on $\mathbb{R}^d$ with infinitesimal generator L_Y and $Y_0 \sim \mu$

Regeneration rate κ with associated tour length

$$\tau = \inf\left\{t \geq 0 : \int_0^t \kappa(Y_s)ds \geq \xi\right\} \text{ with } \xi \sim \text{Exp}(1)$$

$(\{Y_t^{(i)}\}_{t\geq 0}, \tau^{(i)})_{i=0}^{\infty}$ iid realisations inducing regeneration times

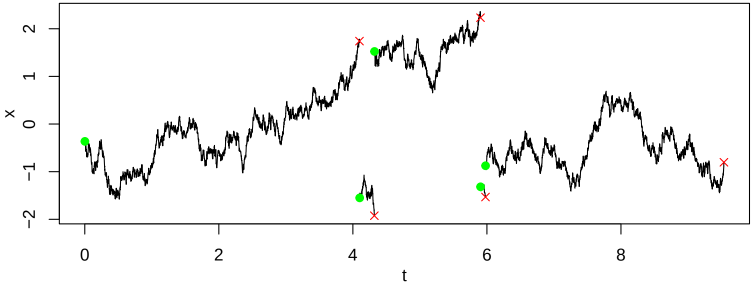
$$T_j = \sum_{i=0}^{j-1} \tau^{(i)}$$

Restore process $\{X_t\}_{t\geq 0}$ given by:

$$X_t = \sum_{i=0}^{\infty} \mathbb{I}_{[T_i, T_{i+1})}(t) Y_{t-T_i}^{(i)}$$

[Wang et al., 2021]

Restore process



Path of five tours of Brownian Restore with $\pi \equiv \mathcal{N}(0, 1^2)$, $\mu \equiv \mathcal{N}(0, 2^2)$ and C such that $\min_{x \in \mathbb{R}} \kappa(x) = 0$, with $\mathcal{K} = 200$ and $\Lambda_0 = 1000$. First and last output states shown by green dots and red crosses

Stationarity of Restore

Infinitesimal generator of $\{X_t\}_{t \geq 0}$

$$L_X f(x) = L_Y f(x) + \kappa(x) \int [f(y) - f(x)] \mu(y) dy$$

with adjoint $L_Y^\dagger$

Regeneration rate κ chosen as

$$\kappa(x) = L_Y^\dagger \pi(x) \big/ \pi(x) + C\mu(x) \big/ \pi(x)$$

Implies $\{X_t\}_{t \geq 0}$ is π-invariant

$$\int_{\mathbb{R}^d} L_X f(x) \pi(x) dx = 0$$

Restore sampler

Rewrite

$$\kappa(x) = \frac{L_Y^\dagger \pi(x)}{\pi(x)} + C\frac{\mu(x)}{\pi(x)} = \tilde{\kappa}(x) + C\frac{\mu(x)}{\pi(x)}$$

with

- $\tilde{\kappa}$ partial regeneration rate
- $C > 0$ regeneration constant and
- $C\mu$ regeneration measure, large enough for $\kappa(\cdot) > 0$

Resulting Monte Carlo method called Restore Sampler

[Wang et al., 2021]

WARWICK THE UNIVERSITY OF WARWICK DAUPHINE UNIVERSITÉ PARIS | PSL erc European Research Council

Restore sampler convergence

Given π-invariance of $\{X_t\}_{t\geq 0}$, Monte Carlo validation follows:

$$\mathbb{E}_\pi[f] = \mathbb{E}_{X_0\sim\mu}\Big[\int_0^{\tau^{(0)}} f(X_s)dx\Big]\Big/\mathbb{E}_{X_0\sim\mu}[\tau^{(0)}] \tag{1}$$

and a.s. convergence of ergodic averages:

$$\frac{1}{t}\int_0^t f(X_s)ds \to \mathbb{E}_\pi[f] \tag{2}$$

For iid $Z_i := \int_{T_i}^{T_{i+1}} f(X_s)ds$, CLT

$$\sqrt{n}\left(\int_0^{T_n} f(X_s)dx/T_n - \mathbb{E}_\pi[f]\right) \to \mathcal{N}(0, \sigma_f^2) \tag{3}$$

where

$$\sigma_f^2 := \mathbb{E}_{X_0\sim\mu}\left[\left(Z_0 - \tau^{(0)}\mathbb{E}_\pi[f]\right)^2\right]\Big/\left(\mathbb{E}_{X_0\sim\mu}[\tau^{(0)}]\right)^2$$

Restore sampler convergence

Given π-invariance of $\{X_t\}_{t\geq 0}$, Monte Carlo validation follows:

$$\mathbb{E}_\pi[f] = \mathbb{E}_{X_0\sim\mu}\Big[\int_0^{\tau^{(0)}} f(X_s)dx\Big] \Big/ \mathbb{E}_{X_0\sim\mu}[\tau^{(0)}] \tag{1}$$

and a.s. convergence of ergodic averages:

$$\frac{1}{t}\int_0^t f(X_s)ds \to \mathbb{E}_\pi[f] \tag{2}$$

Estimator variance depends on expected tour length: prefer μ favouring long tours

[Wang et al., 2020]

Minimal regeneration

Minimal regeneration measure, $C^+\mu^+$ corresponding to smallest possible rate

$$\kappa^+(x) := \tilde{\kappa}(x) \vee 0 = \tilde{\kappa}(x) + C^+\mu^+(x) \Big/ \pi(x)$$

leading to

$$\mu^+(x) = \frac{1}{C^+}[0 \vee -\tilde{\kappa}(x)]\pi(x)$$

Frequent regeneration not necessarily detrimental, except when when μ is not well-aligned to π, leading to wasted computation

"Minimal Restore" maximizes expected tour length / minimizes asymptotic variance

[Wang et al., 2021]

Constant approximation

When target π lacks its normalizing constant Z

$$\pi(x) = \tilde{\pi}(x)/Z,$$

take energy $U(x) := -\log \pi(x) = \log Z - \log \tilde{\pi}(x)$
E.g., when $\{Y_t\}_{t\geq 0}$ Brownian, $\tilde{\kappa}$ function of $\nabla U(x)$ and $\Delta U(x)$

$$\tilde{\kappa}(x) := \frac{1}{2}\left(\|\nabla U(x)\|^2 - \Delta U(x)\right)$$

Constant approximation

When target π lacks its normalizing constant Z

$$\pi(x) = \tilde{\pi}(x)/Z,$$

take energy $U(x) := -\log \pi(x) = \log Z - \log \tilde{\pi}(x)$
In regeneration rate, Z absorbed into C

$$\kappa(x) = \tilde{\kappa}(x) + C\frac{\mu(x)}{(\tilde{\pi}(x)/Z)} = \tilde{\kappa}(x) + CZ\frac{\mu(x)}{\tilde{\pi}(x)} = \tilde{\kappa}(x) + \tilde{C}\frac{\mu(x)}{\tilde{\pi}(x)}$$

where $\tilde{C} = CZ$ set by user. Since

$$C = 1/\mathbb{E}_{\mu}[\tau],$$

using n tours with simulation time T,

$$Z \approx \tilde{C}T/n$$

Adapting Restore

Adaptive Restore process defined by enriching underlying continuous-time Markov process with regenerations at rate κ^+ from distribution μ_t at time t

Convergence of (μ_t, π_t) to (μ^+, π): a.s. convergence of stochastic approximation algorithms for discrete-time processes on compact spaces

[Benaïm et al., 2018; McKimm et al., 2024]

Adapting Restore

Initial regeneration distribution μ_0 and updates by addition of point masses

$$\mu_t(x) = \begin{cases} \mu_0(x), & \text{if } N(t) = 0, \\ \frac{t}{a+t}\frac{1}{N(t)}\sum_{i=1}^{N(t)} \delta_{X_{\zeta_i}}(x) + \frac{a}{a+t}\mu_0(x), & \text{if } N(t) > 0, \end{cases}$$

where $a > 0$ constant and ζ_i arrival times of inhomogeneous Poisson process $(N(t) : t \geq 0)$ with rate $\kappa^-(X_t)$

$$\kappa^-(x) := [0 \vee -\tilde{\kappa}(x)]$$

Poisson process simulated by Poisson thinning, under (strong) assumption

$$K^- := \sup_{x \in \mathcal{X}} \kappa^-(x) > 0$$

Adaptive Brownian Restore Algorithm

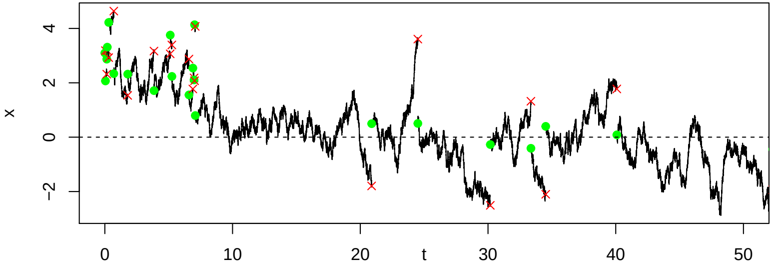
$t \leftarrow 0, E \leftarrow \emptyset, i \leftarrow 0, X \sim \mu_0$.

while $i < n$ **do**

- $\tilde{\tau} \sim \text{Exp}(K^+), s \sim \text{Exp}(\Lambda_0), \tilde{\zeta} \sim \text{Exp}(K^-)$.
- **if** $\tilde{\tau} < s$ and $\tilde{\tau} < \tilde{\zeta}$ **then**
  - $X \sim \mathcal{N}(X, \tilde{\tau}), t \leftarrow t + \tilde{\tau}, u \sim \mathcal{U}[0,1]$.
  - **if** $u < \kappa^+(X)/K^+$ **then**
    - **if** $|E| = 0$ **then**
      - $X \sim \mu_0$.
    - **else**
      - $X \sim \mathcal{U}(E)$ with probability $t/(a+t)$, else $X \sim \mu_0$.
    - **end**
    - $i \leftarrow i + 1$.
  - **end**
- **else if** $s < \tilde{\tau}$ and $s < \tilde{\zeta}$ **then**
  - $X \sim \mathcal{N}(X, s), t \leftarrow t + s$, record X, t, i.
- **else**
  - $X \sim \mathcal{N}(X, \tilde{\zeta}), t \leftarrow t + \tilde{\zeta}, u \sim \mathcal{U}[0,1]$.
  - If $u < \kappa^-(X)/K^-$ then $E \leftarrow E \cup \{X\}$.
- **end**

end

ABRA...cadabra



Path of an Adaptive Brownian Restore process with $\pi \equiv \mathcal{N}(0,1)$, $\mu_0 \equiv \mathcal{N}(2,1)$, $a = 10$. Green dots and red crosses as first and last output states of each tour.

Calibrating initial regeneration and parameters

- μ_0 as initial approximate of π, e.g., $\mu_0 \equiv \mathcal{N}_d(0, I)$ with π pre-transformed
- trade-off in choosing discrete measure dominance time: smaller choices of a for faster convergence versus larger values of a for more regenerations from μ_0, hence better exploration (range of a: between 10^3 and 10^4)
- $\mathcal{K}^+$ and K^- based on quantiles of $\tilde{\kappa}$, from preliminary MCMC runs
- or, assuming π close to Gaussian, initial guess of K^- is $d/2$ and initial estimate for $\mathcal{K}^+$ based on χ^2 approximation

Final remarks

- Adaptive Restore benefits from global moves, for targets hard to approximate with a parametric distribution, with large moves across the space
- Use of minimal regeneration rate makes simulation computationally feasible and more likely in areas where π has significant mass
- In comparison with wandom walk Metropolis, ABRA can be slow but with higher efficiency for targets π with skewed tails

Part II: Importance Monte Carlo

Joint work with C Andral, R Douc and H Marival

arXiv:2207.08271 – **Stoch Proc & Appli**

Rejection Markov chain

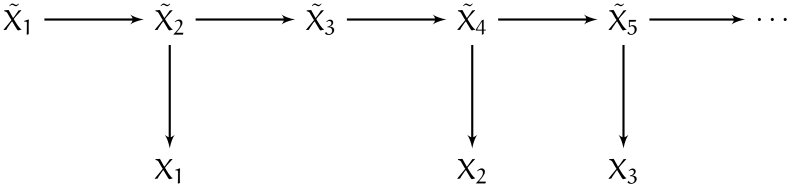
Mixing MCMC and rejection sampling :

1. Draw Markov chain $(\tilde{X}_i)_{1 \leq i \leq m}$ from kernel Q targeting $\tilde{\pi}$
2. With probability $\rho(\tilde{X}_i) \in (0, 1)$ (e.g. $\rho = \pi/(M\tilde{\pi})$), accept proposal
3. Resulting in new Markov chain $(X_i)_{1 \leq i \leq n}$, $(n \leq m)$

Rejection Markov chain

Mixing MCMC and rejection sampling :

1. Draw Markov chain $(\tilde{X}_i)_{1\leq i\leq m}$ from kernel Q targeting $\tilde{\pi}$
2. With probability $\rho(\tilde{X}_i) \in (0,1)$ (e.g. $\rho = \pi/(M\tilde{\pi})$), accept proposal
3. Resulting in new Markov chain $(X_i)_{1\leq i\leq n}$, $(n \leq m)$



Rejection Markov Chain (cont'd)

Corresponding to kernel S for (X_i)

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

Nice (recursive) equation

$$Sh(x) = Q(\rho h)(x) + Q((1-\rho)Sh)(x)$$

Rejection Markov Chain (cont'd)

Corresponding to kernel S for (X_i)

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

Nice (recursive) equation

$$Sh(x) = Q(\rho h)(x) + Q((1-\rho)Sh)(x)$$

Invariant measure for S

If $\tilde{\pi}Q = \tilde{\pi}$, then S is $\rho \cdot \tilde{\pi}$ invariant

$$\int \rho(x)\tilde{\pi}(dx)S(x, dy) = \rho(y)\tilde{\pi}(dy)$$

Direct consequence : if $\rho = \frac{\pi}{M\tilde{\pi}}$, S is π-invariant

Invariant measure for S

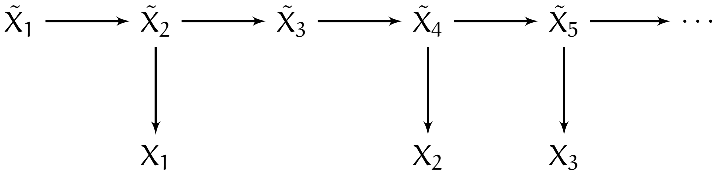
If $\tilde{\pi}Q = \tilde{\pi}$, then S is $\rho \cdot \tilde{\pi}$ invariant

$$\int \rho(x)\tilde{\pi}(dx)S(x, dy) = \rho(y)\tilde{\pi}(dy)$$

Direct consequence : if $\rho = \dfrac{\pi}{M\tilde{\pi}}$, S is π-invariant

From rejection to importance

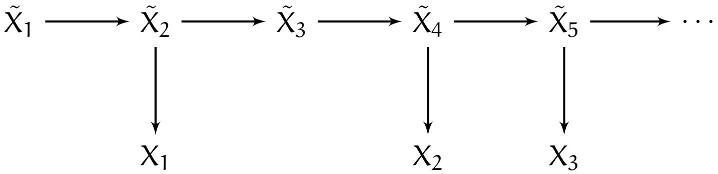
From



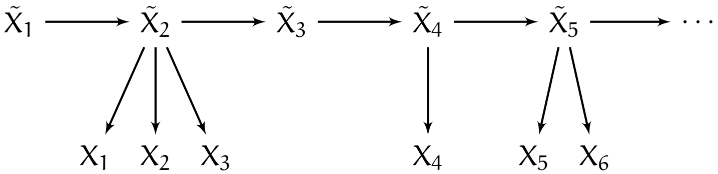
to

From rejection to importance

From



to



Importance Markov chain

- Case when ρ may take values above 1
- Define $\rho_\kappa(x) := \kappa\frac{\pi(x)}{\tilde{\pi}(x)}$ for free parameter κ
- Allow to repeat elements of the chain
- Accept $\tilde{N}_i \sim \tilde{R}(\tilde{X}_i, .)$ copies of $\tilde{X}_i$, with $\tilde{R}_i$ kernel in $\mathbb{N}$ s.t. $\mathbb{E}[\tilde{N}_i|\tilde{X}_i] = \rho_\kappa(\tilde{X}_i)$, e.g.

$$\tilde{N}_i = \left\lfloor \rho_\kappa(\tilde{X}_i) \right\rfloor + \mathcal{B}e\left(\left\{\rho_\kappa(\tilde{X}_i)\right\}\right)$$

Importance Markov chain

- Case when ρ may take values above 1
- Define $\rho_\kappa(x) := \kappa\frac{\pi(x)}{\tilde{\pi}(x)}$ for free parameter κ
- Allow to repeat elements of the chain
- Accept $\tilde{N}_i \sim \tilde{R}(\tilde{X}_i, .)$ copies of $\tilde{X}_i$, with $\tilde{R}_i$ kernel in $\mathbb{N}$ s.t. $\mathbb{E}[\tilde{N}_i|\tilde{X}_i] = \rho_\kappa(\tilde{X}_i)$, e.g.

$$\tilde{N}_i = \left\lfloor \rho_\kappa(\tilde{X}_i) \right\rfloor + \mathcal{B}e\left(\left\{\rho_\kappa(\tilde{X}_i)\right\}\right)$$

Related works

(i) Self-regenerative chain: independent draws ($Q(x,.) = \tilde{\pi}$) and $N_i = V \cdot S$ copies where $S \sim \mathcal{G}eo(q)$, $V \sim \mathcal{B}e(p)$

[Sahu and Zhigljavsky 2003; Gåsemyr 2002]

(ii) Proposals in a semi-Markov approach

[Malefaki and Iliopoulos 2008]

(iii) Dynamic Weighting Monte Carlo and *correctly weighted* joint density:

$$\int w f(x, w) dw \propto \pi(x)$$

[Wong and Liang 1997; Liu, Liang, and Wong 2001]

IMC algorithm

Algorithm 1: Importance Markov chain (IMC)

$\ell \leftarrow 0$

Set an arbitrary $\tilde{X}_0$

for $k \leftarrow 1$ to n **do**

  Draw $\tilde{X}_k \sim Q(\tilde{X}_{k-1}, \cdot)$ and $\tilde{N}_k \sim \tilde{R}(\tilde{X}_k, \cdot)$

  Set $N_\ell = \tilde{N}_k$

  while $N_\ell \geq 1$ **do**

    Set $(X_\ell, N_\ell) \leftarrow (\tilde{X}_k, N_\ell - 1)$

    Set $\ell \leftarrow \ell + 1$

  end

end

Dynamics of the chain

If $N_k > 0$:

$$X_k \longrightarrow X_{k+1} = X_k$$

$$N_k \neq 0 \longrightarrow N_{k+1} = N_k - 1$$

If $N_k = 0$:

$$X_k \longrightarrow X_{k+1} \sim Q(X_k, .)$$

$$\downarrow$$

$$N_k = 0 \qquad N_{k+1} \sim R(X_{k+1}, .)$$

Formalisation of IMC

Define an extended Markov chain (X_i, N_i) where X_i and $\tilde{X}_i$ in same space, and $N_i \in \mathbb{N}$ (counting number of repetitions)

Associated kernel

$$\begin{aligned} Ph(x,n) &= \mathbb{I}_{\{n\geq 1\}} h(x,n-1) \\ &+ \mathbb{I}_{\{n=0\}} \sum_{n'=0}^{\infty} \int_{X} S(x,dx')R(x',n')h(x',n') \end{aligned}$$

where

$$\rho_{\tilde{R}}(x) = \tilde{R}(x,[1,\infty)) \in [0,1]$$

$$R(x,n) := \tilde{R}(x,n+1)/\rho_{\tilde{R}}(x)$$

Formalisation of IMC

Define an extended Markov chain (X_i, N_i) where X_i and $\tilde{X}_i$ in same space, and $N_i \in \mathbb{N}$ (counting number of repetitions)

Associated kernel

$$\begin{aligned} Ph(x,n) &= \mathbb{I}_{\{n\geq 1\}} h(x, n-1) \\ &\qquad + \mathbb{I}_{\{n=0\}} \sum_{n'=0}^{\infty} \int_X S(x, dx') R(x', n') h(x', n') \end{aligned}$$

where

$$\begin{aligned} \rho_{\tilde{R}}(x) &= \tilde{R}(x, [1, \infty)) \in [0, 1] \\ R(x, n) &:= \tilde{R}(x, n+1)/\rho_{\tilde{R}}(x) \end{aligned}$$

Invariant for P

Measure $\bar{\pi}$ on $\mathsf{X} \times \mathbb{N}$:

$$\overline{\pi}(h) = \kappa^{-1} \sum_{n=1}^{\infty} \int_{\mathsf{X}} \tilde{\pi}(dx) \tilde{R}(x, n) \sum_{k=0}^{n-1} h(x, k)$$

such that

1. If $\sum_{n=0}^{\infty} \tilde{R}(x, n) n = \rho_{\kappa}(x)$, marginal of $\overline{\pi}$ on the first component equal to π
2. if $\tilde{\pi} Q = \tilde{\pi}$, Markov kernel $\overline{\pi}$-invariant

Invariant for P

Measure $\bar{\pi}$ on $\mathsf{X} \times \mathbb{N}$:

$$\overline{\pi}(h) = \kappa^{-1} \sum_{n=1}^{\infty} \int_{\mathsf{X}} \tilde{\pi}(dx) \tilde{R}(x, n) \sum_{k=0}^{n-1} h(x, k)$$

such that

1. If $\sum_{n=0}^{\infty} \tilde{R}(x, n) n = \rho_\kappa(x)$, marginal of $\overline{\pi}$ on the first component equal to π
2. if $\tilde{\pi} Q = \tilde{\pi}$, Markov kernel $\overline{\pi}$-invariant

Strong Law of Large Numbers

Theorem 1

If for every $\xi \in \mathsf{M}_1(\mathsf{X})$ *and* $\tilde{\pi}$ *integrable function* g,

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

then, for every $\bar{\pi}$ *integrable function* g,

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{K_n-1} g(X_k) N_k = \bar{\pi}(g), \quad \mathbb{P}_\xi^P - as$$

Central limit theorem

Theorem 2

Under conditions on Q, $\tilde{\pi}$, *take* $h : X \to \mathbb{R}$ *as solution of a Poisson equation for* Q, *then*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(h(X_i) - \pi(h) \right) \overset{\mathbb{P}_X^P-law}{\rightsquigarrow} \mathcal{N}(0, \sigma^2(h)),$$

where

$$\sigma^2(h) = \kappa\tilde{\sigma}^2(\rho h_0) + \kappa^{-1}\hat{\sigma}^2(h_0, \kappa),$$

$\tilde{\sigma}^2(\rho h_0)$ *is the variance obtained with* Q,
$\hat{\sigma}^2(h_0, \kappa) := \int_X h_0^2(x)\mathbb{V}\mathrm{ar}_x^{\tilde{R}}[N]\tilde{\pi}(dx)$,
$\mathbb{V}\mathrm{ar}_x^{\tilde{R}}[N] := \int_{\mathbb{N}} \tilde{R}(x, dn)n^2 - \left(\int_{\mathbb{N}} \tilde{R}(x, dn)n\right)^2$.

Central limit theorem

Theorem 2

Under conditions on Q, $\tilde{\pi}$, *take* $h : X \to \mathbb{R}$ *as solution of a Poisson equation for* Q, *then*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(h(X_i) - \pi(h) \right) \overset{\mathbb{P}_X^P - law}{\rightsquigarrow} \mathcal{N}(0, \sigma^2(h)),$$

where

$$\sigma^2(h) = \kappa \tilde{\sigma}^2(\rho h_0) + \kappa^{-1} \hat{\sigma}^2(h_0, \kappa),$$

$\tilde{\sigma}^2(\rho h_0)$ *variance coming from instrumental chain,* $\hat{\sigma}^2(h_0, \kappa)$ *variance from random number of repetitions*

choosing $\tilde{R}$

$\hat{\sigma}$ depends on the variance of $\tilde{R}$

$$\mathbb{V}\mathrm{ar}_x^{\tilde{R}}[N] := \int_{\mathbb{N}} \tilde{R}(x, dn)n^2 - \left(\int_{\mathbb{N}} \tilde{R}(x, dn)n\right)^2$$

For N integer-value random variable such that $\mathbb{E}[N] = \rho < \infty$,

$$\mathbb{V}\mathrm{ar}(N) \geq \{\rho\}(1 - \{\rho\})$$

LoBound met by $N = \lfloor \rho \rfloor + S$, where $S \sim \mathrm{Ber}(\{\rho\})$ (used as "shifted Bernoulli" kernel default)

Geometric ergodicity

Theorem 3

Under assumptions on Q *and* $\tilde{R}$, P *has unique invariant probability measure* $\bar{\pi}$ *and there exist constants* $\delta, \beta_r > 1$, $\zeta < \infty$, *such that for all* $\xi \in \mathsf{M}_1(\mathsf{X} \times \mathbb{N})$,

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

WARWICK THE UNIVERSITY OF WARWICK
DAUPHINE UNIVERSITÉ PARIS | PSL
erc European Research Council

Pseudo-marginal version

- Cases when the density π not available but replaced by (unbiased) estimate, leading to pseudo-marginal method.

  [Andrieu and Roberts 2009]

- Extension: Importance Markov chain "pseudo-marginal compatible" when unbiased estimate available
  draw $\hat{\pi}(X_i)$ (with expectation $\pi(x)$) and

$$N_i \sim \lfloor \hat{\pi}/\tilde{\pi}(X_i) \rfloor + \mathcal{B}e\left(\{\hat{\pi}/\tilde{\pi}(X_i)\}\right)$$

  by enlarging the chain structure

- ... and even with unbiased estimate of $\tilde{\pi}$
- resulting in higher variance

Pseudo-marginal version

- Cases when the density π not available but replaced by (unbiased) estimate, leading to pseudo-marginal method.
  [Andrieu and Roberts 2009]

- Extension: Importance Markov chain "pseudo-marginal compatible" when unbiased estimate available
  draw $\hat{\pi}(X_i)$ (with expectation $\pi(x)$) and

$$N_i \sim \lfloor \hat{\pi}/\tilde{\pi}(X_i) \rfloor + \mathcal{B}e\left(\{\hat{\pi}/\tilde{\pi}(X_i)\}\right)$$

  by enlarging the chain structure
- ... and even with unbiased estimate of $\tilde{\pi}$
- resulting in higher variance

IMC in practice

Several factors to choose :

- auxiliary distribution $\tilde{\pi}$
- kernel Q
- value of κ

About κ

Several things to note

- κ is arbitrary if π and $\tilde{\pi}$ are unnormalized.
- the length of the final chain (X_i) grows linearly in κ for a fixed length n chain $(\tilde{X}_i)$
- hence, automatic tuning κ achieved by setting length of chain
- For $\text{ESS}_\kappa := (\sum_{i=1}^n \tilde{N}_i)^2 / \sum_{i=1}^n \tilde{N}_i^2$ and usual IS ESS: $\text{ESS}_{\text{IS}} := \left(\sum_{i=1}^n \rho(\tilde{X}_i)\right)^2 \Big/ \sum_{i=1}^n \rho(\tilde{X}_i)^2$

$$\text{ESS}_\kappa \underset{\kappa\longrightarrow\infty}{\longrightarrow} \text{ESS}_{\text{IS}}$$

Warning: notion of ESS not accounting for convergence of Markov chain

About κ

Several things to note

- κ is arbitrary if π and $\tilde{\pi}$ are unnormalized.
- the length of the final chain (X_i) grows linearly in κ for a fixed length n chain $(\tilde{X}_i)$
- hence, automatic tuning κ achieved by setting length of chain
- For $\text{ESS}_\kappa := (\sum_{i=1}^n \tilde{N}_i)^2 / \sum_{i=1}^n \tilde{N}_i^2$ and usual IS ESS: $\text{ESS}_{\text{IS}} := \left(\sum_{i=1}^n \rho(\tilde{X}_i)\right)^2 \Big/ \sum_{i=1}^n \rho(\tilde{X}_i)^2$

$$\text{ESS}_\kappa \xrightarrow[\kappa \to \infty]{} \text{ESS}_{\text{IS}}$$

Warning: notion of ESS not accounting for convergence of Markov chain

About κ

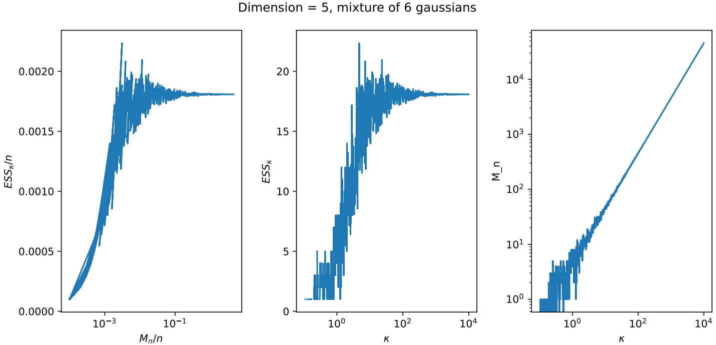
Several things to note

- κ is arbitrary if π and $\tilde{\pi}$ are unnormalized.
- the length of the final chain (X_i) grows linearly in κ for a fixed length n chain $(\tilde{X}_i)$
- hence, automatic tuning κ achieved by setting length of chain
- For $\text{ESS}_\kappa := (\sum_{i=1}^n \tilde{N}_i)^2 / \sum_{i=1}^n \tilde{N}_i^2$ and usual IS ESS: $\text{ESS}_{\text{IS}} := \left(\sum_{i=1}^n \rho(\tilde{X}_i)\right)^2 \Big/ \sum_{i=1}^n \rho(\tilde{X}_i)^2$

$$\text{ESS}_\kappa \underset{\kappa \to \infty}{\longrightarrow} \text{ESS}_{\text{IS}}$$

Warning: notion of ESS not accounting for convergence of Markov chain

Impact of κ



Impact of κ on ESS and Markov chain length

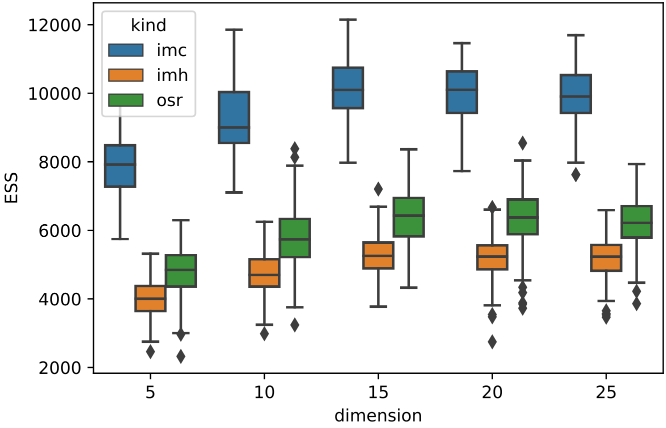
Independent IMC and normalizing flows

- Target: π a d-dimensional distribution, with 2^d modes, concentrated around the sphere

$$\pi(x) \propto \exp\left(-\frac{1}{2}\left(\frac{\|x\|-2}{0.1}\right)^2 - \sum_{i=1}^{d} \log\left(e^{-\frac{1}{2}(\frac{x_i+3}{0.6})^2} + e^{-\frac{1}{2}(\frac{x_i-3}{0.6})^2}\right)\right)$$

- Instrumental density and kernel: a normalizing flow T is trained to approximate π: $Q(x,.) = \tilde{\pi}(.) = T_\sharp \mathcal{N}(0,1)$
- Comparison with Metropolis–Hastings and Self Regenerative MCMC

[Sahu and Zhigljavsky 2003; Gabrié et al 2022]



kind
imc
imh
osr
ESS
12000
10000
8000
6000
4000
2000
5
10
15
20
25
dimension

Conclusion

- Versatile framework that applies to many different kernels Q and auxiliary distributions $\tilde{\pi}$
- Extensions: adaptive version – multiple auxiliary chains – delayed acceptance

References I

Andrieu, Christophe and Gareth O. Roberts (Apr. 2009). "The Pseudo-Marginal Approach for Efficient Monte Carlo Computations". In: The Annals of Statistics 37.2, pp. 697–725.

Gåsemyr, Jørund (2002). Markov Chain Monte Carlo Algorithms with Independent Proposal Distribution and Their Relationship to Importance Sampling and Rejection Sampling. Tech. rep.

Liu, Jun S, Faming Liang, and Wing Hung Wong (June 2001). "A Theory for Dynamic Weighting in Monte Carlo Computation". In: Journal of the American Statistical Association 96.454, pp. 561–573.

Malefaki, Sonia and George Iliopoulos (Apr. 2008). "On Convergence of Properly Weighted Samples to the Target Distribution". In: Journal of Statistical Planning and Inference 138.4, pp. 1210–1225.

Sahu, Sujit K. and Anatoly A. Zhigljavsky (June 2003). "Self-Regenerative Markov Chain Monte Carlo with Adaptation". In: Bernoulli 9.3, pp. 395–422.

Wong, Wing Hung and Faming Liang (Dec. 1997). "Dynamic Weighting in Monte Carlo and Optimization". In: Proceedings of the National Academy of Sciences 94.26, pp. 14220–14224.