Consistency of Markov chain quasi Monte Carlo with multiple proposals

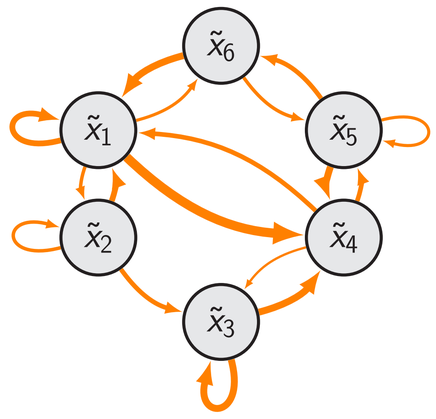
Tobias Schwedes, Ben Calderhead
Department of Mathematics, Imperial College London

Abstract

Markov chain Monte Carlo (MCMC) is a preferred sampling method due its wide applicability. However, it suffers from a slow scaling of $n^{-\frac{1}{2}}$, analogously to standard Monte Carlo. For the latter, scaling close to n^{-1} can be achieved by using more homogeneously distributed points in space than the usual pseudo-random numbers. With the same philosophy in mind we apply quasi-Monte Carlo to an MCMC method that uses multiple proposals. We show consistency and achieve an improved scaling of close to n^{-1} numerically.

Using multiple proposals in MCMC

Metropolis-Hastings (M-H) was generalised in (Cal14) such that it allows for parallelising a single chain by proposing multiple points via a kernel κ in parallel. In every MCMC iteration, a finite state Markov chain on the proposed points, i.e. $\tilde{x}_1, ..., \tilde{x}_{N+1}$, is constructed, as displayed in the figure on the right, from which samples are drawn.



In practice, generating proposals $x_{\setminus I} = x_{[1:N+1]\setminus\{I\}}$ from $\kappa(x_I, \cdot)$ is typically performed via applying its generalised inverse Ψ_{x_I} to a set of pseudo-random numbers $u_1, ..., u_{Nd} \in [0,1]$ such that $x_{\setminus I} = \Psi_{x_I}(u_1, ..., u_{Nd})$. The overall procedure is described in Algorithm 1.

Algorithm 1: Multiple proposal Metropolis-Hastings

Input: Initialise starting point $\tilde{x}_1 \in \Omega \subset \mathbb{R}^d$, number of proposals N, auxiliary variable $I = 1$ and counter $n = 0$, generate an pseudo-random sequence $u_1, u_2, ... \in (0,1]$;

1 **for** *each MCMC iteration* $i = 1, 2, ...$ **do**
2 Set $u^i = (u_{(i-1)Nd+1}, ..., u_{iNd}) \in (0,1]^{Nd}$, and sample $\tilde{x}_{\setminus I}$ conditioned on I, i.e., draw N new points from $\kappa(\tilde{x}_I, \cdot) = p(\tilde{x}_{\setminus I}|\tilde{x}_I)$ by the inverse $\Psi_{\tilde{x}_I}(u^i)$;
3 Compute the stationary distribution of $I|\tilde{x}_{1:N+1}$, i.e. $p(I = j|\tilde{x}_{1:N+1}) = \pi(\tilde{x}_j)\kappa(\tilde{x}_j, \tilde{x}_{\setminus j}) / \sum_k \pi(\tilde{x}_k)\kappa(\tilde{x}_k, \tilde{x}_{\setminus k})$, which can be done in parallel;
4 **for** $m = 1, ..., N$ **do**
5 Sample new I via $p(\cdot|\tilde{x}_{1:N+1})$;
6 Set new sample $x_{n+m} = \tilde{x}_I$;
7 **end**
8 Update counter $n = n + N$
9 **end**

Properties of multiple proposal M-H

- Satisfies detailed balance, thus updates leave the stationary distribution invariant
- The proposal kernel κ can be chosen freely, including formulations based on Langevin diffusion and Hamiltonian dynamics
- The likelihoods of the proposed states can be computed in parallel
- Due to the multiple alternatives to accept the next state, the acceptance rate is increased

Completely uniformly distributed numbers

The star discrepancy D_n^{*d} for a set of points $P = \{x_1, ..., x_n\} \subset (0,1]^d$ is defined by

$$D_n^*(P) = \sup_{b \in (0,1]^d} \left| \frac{\#\{x_i \in (0,b]\}}{n} - \mathrm{vol}\left((0,b]\right) \right|.$$

A sequence $(u_i)_{i \geq 1} \subset (0,1]$ is called c.u.d. if for any $d \geq 1$ the points $x_i^{(d)} = (u_i, \ldots, u_{i+d-1}) \in (0,1]^d$ fulfill $D_n^{*d}(x_1^{(d)}, \ldots, x_n^{(d)}) \to 0$ as $n \to \infty$.

References

[Cal14] Calderhead, Ben: A general construction for parallelizing Metropolis- Hastings algorithms. In: *Proceedings of the National Academy of Sciences* 111 (2014), Nr. 49, S. 17408–17413

[CDO11] Chen, S ; Dick, Josef ; Owen, Art B.: Consistency of Markov chain quasi-Monte Carlo on continuous state spaces. In: *The Annals of Statistics* (2011), S. 673–701

[CMNO12] Chen, Su ; Matsumoto, Makoto ; Nishimura, Takuji ; Owen, Art B.: New inputs and methods for Markov chain quasi-Monte Carlo. In: *Monte Carlo and quasi-Monte Carlo methods 2010*. Springer, 2012, S. 313–327

C.u.d. numbers in practice

The c.u.d. construction used in our simulations was introduced in (CMNO12) and relies on a linear-feedback shift register with a transition mechanism based on primitive polynomials over the Galois field $GF(2)$. Referring to Fig. 1, the resulting sequence is more homogeneously distributed than pseudo-random numbers.

(a) c.u.d. (b) pseudo-random

Figure : Segments of c.u.d. and pseudo-random finite sequences in $(0,1]$.

Regularity conditions

Coupling: For any MCMC iteration in Algorithm 1, let $\phi(x_I, (u_1, ..., u_{Nd})) = (x_1, ..., x_N)$ denote the innovation operator that produces the N new samples. Let $\mathcal{C} \subset (0,1]^{Nd}$ have positive Jordan measure. If for any $u \in \mathcal{C}$ it holds $\phi(x, u) = \phi(x', u)\ \forall\ x, x' \in \Omega$, then $\mathcal{C}$ is called a *coupling region*.

Integrability: Let x_n be a sample from the ith iteration. Any $g : [0,1]^{iNd} \to \mathbb{R}$, defined by $g(u^1, \ldots, u^i) = f(x_n(u^1, \ldots, u^i))$, where $f : \mathbb{R}^d \to \mathbb{R}$ is bounded and continuous, is Riemann integrable.

Consistency theorem

Let us assume the above regularity and that the Markov chain defined by Algorithm 1 is irreducible and Harris recurrent with stationary distribution π. If the sequence of pseudo-random numbers is replaced by c.u.d. numbers, then $(x_n)_{n \geq 1}$ still consistently samples π.

Numerical results

Example problem: Estimate mean and distribution of a one-dimensional Gaussian

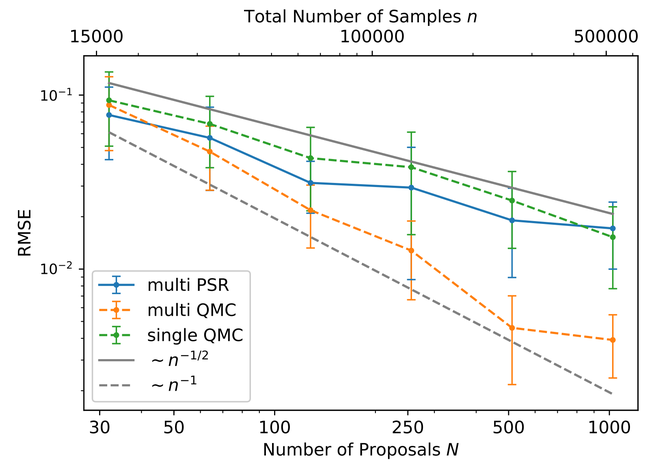


Figure : RMSE of the sample mean based on M-H with independent proposals using pseudo-random (PSR) vs. c.u.d. (QMC) numbers as driving sequences, resp., for increasing proposal numbers and sample sizes. The results are based on 25 MCMC simulations

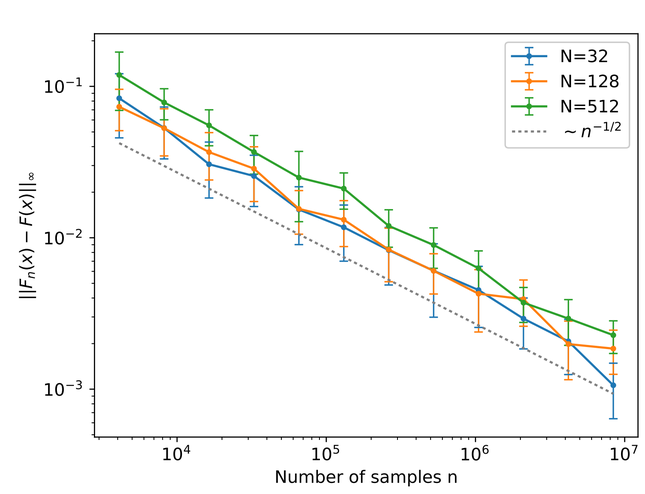


Figure : Maximum-norm error for the empirical distribution function of samples from the N proposal M-H with random walk proposals, using c.u.d. numbers as driving sequence, for increasing sample sizes. The results are based on 25 MCMC simulations

Future research

- Performance in higher dimensions
- Consistency proof for larger classes of proposal kernels

t.schwedes14@imperial.ac.uk