Survey methods and their use in Monte Carlo algorithms

Mathieu Gerber

University of Bristol

SMURF Workshop

August 22, 2018

Objectives of the talks

1. Present some examples where the problem of sampling from a finite population arises in Monte Carlo

   $\Longrightarrow$ special focus on sequential Monte Carlo methods.

2. Present some new results on resampling algorithms

   $\Longrightarrow$ Based on a joint work with Nicolas Chopin (CREST/ENSAE) and Nick Whiteley (Bristol)

3. Present some open questions

Monte Carlo basics

The goal of Monte Carlo methods is to approximate integrals of the form

$$I = \int_{\mathcal{X}} \varphi(x)\pi(\mathrm{d}x), \quad \mathcal{X} \subset \mathbb{R}^d \tag{1}$$

The basic observation underpinning Monte Carlo methods is that (1) is the expectation of φ under π, that is, $I = \mathbb{E}_\pi[\varphi]$

Therefore, if $\{(x^n, w^n)\}_{n=1}^N$ is a weighted point set such that

$$\pi^N := \sum_{n=1}^N w^n \delta_{x^n} \approx \pi$$

one can approximate $I = \mathbb{E}_\pi[\varphi]$ by $I^N = \mathbb{E}_{\pi^N}[\varphi]$.

Two main questions in Monte Carlo literature

1. How is the approximation error $\|I^N - I\|$ related to $\|\pi^N - \pi\|$?

   Some results:

   - $\mathbb{E}[(I^N - I)^p]^{1/p} \leq \frac{C_p}{N^{1/p}} \mathbb{E}_\pi[(\varphi - I)^p]^{1/p}$ if $x^n \overset{\text{iid}}{\sim} \pi$ and $w^n = N^{-1}$.
   - $|I^N - I| \leq \|\pi^N - \pi\|_\star$ if $V(\varphi) \leq 1$

2. How can we define π^N so that $\|\pi^N - \pi\|$ is small?

$\implies$ We focus on this second point in this talk.

Computing π^N

The simplest Monte Carlo scheme is Monte Carlo integration, where $x^n \overset{\text{i.i.d.}}{\sim} \pi$ and

$$\pi^N = \frac{1}{N}\sum_{n=1}^{N} \delta_{x^n}.$$

However, in most statistical problems we don't know how to sample i.i.d. from π and thus more complicated methods need to be used to compute a 'good' π^N.

Markov chain Monte Carlo (MCMC) and sequential Monte Carlo (SMC, or particle filters) are arguably the two most popular classes of Monte Carlo algorithms used in statistics.

$\Longrightarrow$ As will shall see, a crucial step of SMC amounts to sampling from a finite population.

SMC: Basic idea

In SMC, instead of trying to approximate π directly we start with the simpler problem of approximating π_0, a distribution chosen by the user and easy to sample from.

In a second step, we define a sequence $\{\pi_t\}_{t=0}^T$ such that $\pi_T = \pi$ and such that, in some sense, π_t is 'close' to π_{t-1}. For instance,

$$\pi_t \propto \pi_0^{(1-\rho_t)} \pi^{\rho_t}, \quad 0 = \rho_0 < \cdots < \rho_T = 1.$$

Then, the main idea underpinning SMC is that if π_0^N is a good approximation of π_0, and if π_1 is close to π_0, then it should be possible to use π_0^N to build a good approximation π_1^N of π_1.

By repeating this reasoning up to the terminal time T we end up with an approximation π_T^N of $\pi_T = \pi$.

Algorithmic description of SMC

Operations must be be performed for all $n \in 1:N$.
At time 0,

(a) Generate $x_0^n \sim \pi_0(\mathrm{d}x_0)$.

(b) Compute $w_0^n = N^{-1}$

Recursively, for $t = 1:T$,

(a) Generate $A_{t-1}^{1:N} \sim \rho\big(\{x_{t-1}^n, w_{t-1}^n\}_{n=1}^N\big)$ **[resampling]**

(b) Generate $x_t^n \sim m_t(x_{t-1}^{A_{t-1}^n}, \mathrm{d}x_t)$ **[mutation]**

(c) Compute

$$w_t^n = \frac{w_t(x_t^n)}{\sum_{m=1}^N w_t(x_t^m)}, \quad w_t(x) = \frac{\pi_t(\mathrm{d}x)}{\pi_{t-1}(\mathrm{d}x)}$$

Output: An approximation $\pi_t^N = \sum_{n=1}^N w_t^n \delta_{x_t^n}$ of π_t for all $t \geq 0$ such that (hopefully)

$$\pi_t^N \Rightarrow \pi_t, \quad \text{as} \ \ N \to +\infty, \quad a.s.$$

Resampling schemes: Informal definition

A resampling scheme ρ takes as an input a weighted sample $\left\{(x^n, w^n)\right\}_{n=1}^N$ and returns as an output $\{A^n\}_{n=1}^N$, a set of random indices in $\{1, \ldots, N\}$.

Resampling steps play a central role in SMC since they

1. Prevent the particle system $\left\{(x_t^n, w_t^n)\right\}_{n=1}^N$ to collapse (i.e. to end up with a particle system where one particle has weight equal to one).
2. Prevent the particle approximation π_t^N of π_t to deteriorate as t increases (time uniform bounds)

On the other hand, a good resampling scheme should be such that

$$\frac{1}{N} \sum_{n=1}^{N} \delta_{x^{A^n}} \approx \sum_{n=1}^{N} w^n \delta_{x^n}$$

to minimize the noise introduced by the resampling operation at the current iteration t.

Most commonly used resampling methods

- **Multinomial resampling:**

$$A^n = F_N^{-}(u^n), \quad n = 1, \dots, N, \quad F_N(x) = \sum_{n=1}^{N} w^n \mathbb{I}(n \leq x)$$

  where $\{u^n\}_{n=1}^N$ are i.i.d. $\mathcal{U}(0,1)$ random variables.

- **Stratified resampling**:

$$A^n = F_N^{-}\left(\frac{n-1+u^n}{N}\right), \quad n = 1, \dots, N$$

  where $\{u^n\}_{n=1}^N$ are i.i.d. $\mathcal{U}(0,1)$ random variables.

- **Systematic resampling**:

$$A^n = F_N^{-}\left(\frac{n-1+u}{N}\right), \quad n = 1, \dots, N$$

  where $u \sim \mathcal{U}(0,1)$.

Numerical comparison of resampling schemes

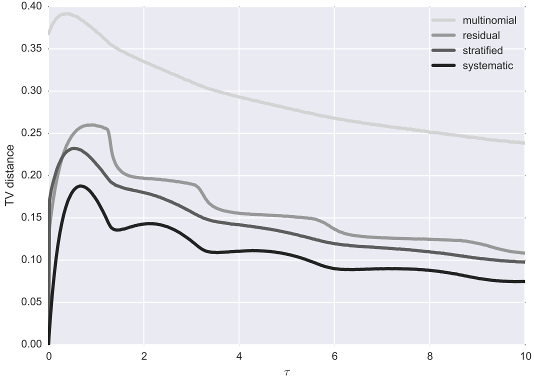


Figure 1 : TV distance between empirical distributions of weighted particles, and resampled particles as a function of τ; particles are $\mathcal{N}(0,1)$ random variables, weight function is $w(x) = \exp(-\tau x^2/2)$.

Main issues

Multinomial resampling is easy to understand/analyse since

$$x^{A^n} \overset{\text{iid}}{\sim} \sum_{n=1}^{N} w^n \delta_{x^n}.$$

But little is known about the properties of stratified and systematic resampling.

Indeed, despite the popularity these two resampling mechanisms most results on particle filtering assume that multinomial resampling is used.

In particular, it is not even known whether or not particle filters are still weakly convergent (i.e. $\pi_t^N \Rightarrow \pi_t$ as $N \to +\infty$) when stratified or systematic resampling are used instead of multinomial resampling.

$\implies$ I will now present some results that contribute to fill these gaps.

Resampling schemes: Formal definition

Definition

A resampling scheme is a mapping $\rho : [0,1]^{\mathbb{N}} \times \mathcal{Z} \to \mathcal{P}_f(\mathcal{X})$ such that, for any $N \geq 1$ and $z = \{x^n, w^n\}_{n=1}^N \in \mathcal{Z}^N$,

$$\rho(u, z) = \frac{1}{N} \sum_{n=1}^{N} \delta(x^{a_N^n(u,z)}),$$

where for each n, $a_N^n : [0,1]^{\mathbb{N}} \times \mathcal{Z}^N \to 1 : N$ is a certain measurable function.

Notation:

1. $\mathcal{P}(\mathcal{X})$ is the set probability measures on $\mathcal{X}$.
2. $\mathcal{P}_f(\mathcal{X})$ is the set of discrete probability measures on $\mathcal{X}$.
3. $\mathcal{Z} := \bigcup_{N=1}^{+\infty} \mathcal{Z}^N$ with $\mathcal{Z}^N = \left\{(x, w) \in \mathcal{X}^N \times \mathbb{R}_+^N : \sum_{n=1}^N w_n = 1\right\}$.

Consistent resampling schemes

We consider in this work that a resampling scheme is consistent if it is weak-convergence-preserving.

Definition

Let $\mathcal{P}_0 \subseteq \mathcal{P}(\mathcal{X})$. Then, we say that a resampling scheme $\rho : [0,1]^{\mathbb{N}} \times \mathcal{Z} \to \mathcal{Z}$ is $\mathcal{P}_0$-consistent if, for any $\pi \in \mathcal{P}_0$ and random sequence $(\zeta^N)_{N\geq 1}$ such that $\pi^N \Rightarrow \pi$, $\mathbb{P}$-a.s., one has

$$\rho(\zeta^N) \Rightarrow \pi, \quad \mathbb{P} - \text{a.s.}$$

Remarks:

1. All the random variables are defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.
2. ζ^N is a r.v. that takes its value in $\mathcal{Z}^N$ and $\pi^N \in \mathcal{P}_f(\mathcal{X})$ is the corresponding probability measure.
3. It is well known that multinomial resampling is $\mathcal{P}(\mathcal{X})$-consistent.

A general consistency result: Preliminaries

- To simplify the presentation we assume henceforth that $\mathcal{X} = \mathbb{R}^d$.

- The random variables $\{Z^n\}_{n=1}^N$ are negatively associated (NA) if, for every pair of disjoint subsets I_1 and I_2 of $\{1, \dots N\}$,

$$\mathrm{Cov}\Big(\varphi_1\big(Z^n, n \in I_1\big),\, \varphi_2\big(Z^n, n \in I_2\big)\Big) \leq 0$$

  for all coordinatewise non-decreasing functions φ_1 and φ_2, such that for $k \in \{1, 2\}$, $\varphi_k : \mathbb{R}^{|I_k|} \to \mathbb{R}$ and such that the covariance is well-defined.

- Let $\widetilde{\mathcal{P}}_b(\mathcal{X}) \subset \mathcal{P}(\mathcal{X})$ be a set of probabilities densities with "not too thick tails" (see paper for exact definition).

  $\Longrightarrow$ The condition on the tail is weak as it does not impose that $\pi \in \widetilde{\mathcal{P}}_b(\mathcal{X})$ has a finite first moment.

Main result

Theorem ($\mathcal{X} = \mathbb{R}^d$ to simplify)

Let $\rho : [0,1]^{\mathbb{N}} \times \mathcal{Z} \to \mathcal{P}_f(\mathcal{X})$ be an unbiased resampling scheme such that:

(H1) *For any $N \geq 1$ and $z \in \mathcal{Z}^N$ the collection of random variables $\{N_{\rho,z}^n\}_{n=1}^N$ is negatively associated;*

(H2) *There exists a sequence $(r_N)_{N\geq 1}$ of non-negative real numbers such that $r_N = o(N/\log N)$, and, for N large enough,*

$$\sup_{z\in\mathcal{Z}^N} \sum_{n=1}^N \mathbb{E}\big[(\Delta_{\rho,z}^n)^2\big] \leq r_N\, N, \qquad \sum_{N=1}^\infty \sup_{z\in\mathcal{Z}^N} \mathbb{P}\Big(\max_{n\in 1:N} \big|\Delta_{\rho,z}^n\big| > r_N\Big) < +\infty.$$

Then, ρ is $\widetilde{\mathcal{P}}_b(\mathcal{X})$-consistent.

Notation: $N_{\rho,z}^n = \sum_{m=1}^N \mathbb{I}(A^m = n)$ and $\Delta_{\rho,z}^n = N_{\rho,z}^n - NW^n$.

Definition: ρ is unbiased if $\mathbb{E}\big[\Delta_{\rho,z}^n\big] = 0$ for all n and $z \in \mathcal{Z}$.

Applications

From the previous theorem we deduce the following corollary:

Corollary

- *Multinomial resampling is* $\widetilde{\mathcal{P}}_b(\mathcal{X})$*-consistent (not new);*
- *Residual resampling is* $\widetilde{\mathcal{P}}_b(\mathcal{X})$*-consistent (not new);*
- *Stratified resampling is* $\widetilde{\mathcal{P}}_b(\mathcal{X})$*-consistent (new);*
- *Residual/Stratified resampling is* $\widetilde{\mathcal{P}}_b(\mathcal{X})$*-consistent (new);*
- *SSP/Pivotal resampling is* $\widetilde{\mathcal{P}}_b(\mathcal{X})$*-consistent (new).*

Remark: The above result prove the consistency of resampling schemes based on conditional Poisson sampling, Sampford sampling or Pareto sampling.

Applications

Our general consistency result can be easily extended to the case where $M_N \leq N$ elements are (re)sampled, provided that e.g. $N/M_N \to K$ (for some $K < +\infty$).

$\Longrightarrow$ **Application for survey sampling:** Almost sure weak consistency of the Horvitz-Thomson "distribution" measure

$$\frac{1}{N}\sum_{n=1}^{M_N}\frac{1}{\pi_{N,A^n}}\delta_{Y_{A^n}}, \quad \pi_{N,n} = \frac{M_N G(Z_n)}{\sum_{m=1}^{N} G(Z_m)}$$

under various sampling methods (and suitable assumptions on the link function G and $(Z_n)_{n\geq 1}$).

Strategy of the proof (assume $\mathcal{X} = (0,1)^d$ to simplify)

1. In a first step we show that a resampling scheme ρ is $\widetilde{\mathcal{P}}_b(\mathcal{X})$-consistent if and only if, for any $\pi \in \widetilde{\mathcal{P}}_b(\mathcal{X})$ and sequence $(\zeta^N)_{N\geq 1}$ such that $\pi^N \implies \pi$, $\mathbb{P}$-a.s., we have

$$\lim_{N\to\infty} \|\rho(\zeta^N)_h - \pi_h^N\|_\star = 0, \quad \mathbb{P} - a.s. \tag{2}$$

   with $h : \mathcal{X} \to (0,1)$ a measurable pseudo-inverse of the Hilbert space filling curve.

2. In a second step, noting that, for every $z = (x^n, w^n)_{n=1}^N \in \mathcal{Z}^N$

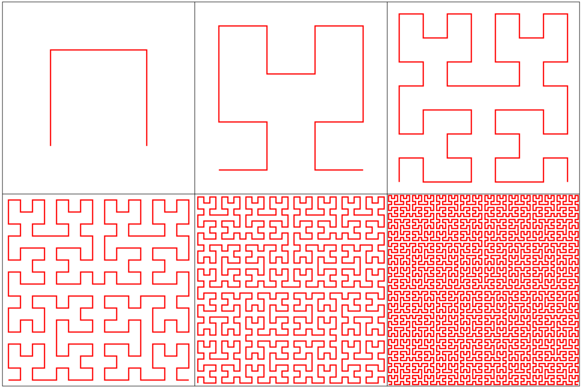
$$\|\rho(z)_h - \pi_h^N\|_\star = \max_{m\in 1:N} \Big| \sum_{n=1}^{m} \Delta_{\rho,z}^{\sigma_h(n)} \Big|, \quad h(x^{\sigma_h(1)}) \leq \cdots \leq h(x^{\sigma_h(N)})$$

   we show that the hypotheses (H1) and (H2) are sufficient to establish (2), via a maximal inequality for negatively associated random variables due to Shao (2000).

The Hilbert space filling curve

The Hilbert space filling curve $H : [0,1] \to [0,1]^d$ is a continuous and surjective mapping.

It is defined as the limit of a sequence $(H_m)_{m\geq 1}$



First six elements of the sequence $(H_m)_{m\geq 1}$ for $d = 2$ (source: Wikipedia)

What about systematic sampling?

We show that systematic resampling is not consistent in the sense of the above definition, i.e. there exist a continuous probability measure π and a random sequence $(\zeta^N)_{N\geq 1}$ such that $\pi^N \Rightarrow \pi$, $\mathbb{P}$-a.s., but

$$\mathbb{P}\big(\rho_{\text{syst}}(\zeta^N) \Rightarrow \pi\big) < 1.$$

Open question 1: Is systematic resampling consistent when applied on $\{(x^{\sigma_N(n)}, W^{\sigma_N(n)})\}_{n=1}^N$, where σ_N is a random permutation of $1:N$?

Open question 2: More generally, is a resampling scheme ρ is such that $\sup_{z,n}|\Delta^n_{\rho,z}| \leq C$ consistent when applied on $\{(x^{\sigma_N(n)}, W^{\sigma_N(n)})\}_{n=1}^N$?

$\Longrightarrow$ Use of deterministic resampling mechanism?

Digression: Variance of systematic sampling

Let $h(x^1) \leq \cdots \leq h(x^N)$ and, for σ a permutation of $1:N$ and $n \in 1:N$ let

$$F_\sigma(n) = \sum_{i=1}^{n} W^{\sigma(i)}, \quad a^n_{\sigma,N} = \{NF_{\sigma,N}(\sigma^{-1}(n)-1)\}, \quad b^n_{\sigma,N} = \{NF_{\sigma,N}(\sigma^{-1}(n))\}$$

and, assuming that $\min_n W^n > 0$,

$$c^n_{\sigma,N} = \begin{cases} 1, & b^n_{\sigma,N} > a^n_{\sigma,N} \\ -1 & b^n_{\sigma,N} < a^n_{\sigma,N} \end{cases}, \quad I^n_{\sigma,N} = \begin{cases} [a^n_{\sigma,N}, b^n_{\sigma,N}), & c^n_{\sigma,N} = 1 \\ [b^n_{\sigma,N}, a^n_{\sigma,N}), & c^n_{\sigma,N} = -1. \end{cases}$$

Then, following L'Ecuyer and Lemieux (2000), under systematic resampling

$$\begin{aligned}
&\mathrm{Var}\Big(\frac{1}{N}\sum_{i=1}^{N}\varphi\big(x^{A^n}\big)\Big) \\
&\quad = \frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N}\varphi(x^n)\varphi(x^m)c^n_{\sigma,N}c^m_{\sigma,N}\big(\lambda_1(I^n_{\sigma,N}\cap I^m_{\sigma,N}) - \lambda_1(I^n_{\sigma,N})\lambda_1(I^m_{\sigma,N})\big)
\end{aligned}$$

Hilbert ordered stratified resampling: Variance

As mentioned above, a good resampling scheme is such that

$$\mathrm{Var}\big(\rho(z)(\varphi)\big)$$

is small (for some class of functions) and all $z \in \mathcal{Z}^N$.

We show that Hilbert ordered stratified resampling is such that

- For any $(\zeta^N)_{N\geq 1}$ such that $\pi^N \Rightarrow \pi$ and any continuous and bounded φ,

$$N\mathrm{Var}\big(\rho(\zeta^N)(\varphi)\big) \to 0.$$

- For any $N \geq 1$, $z \in \mathcal{Z}^N$

$$\mathrm{Var}\big(\rho(\zeta^N)(\varphi)\big) \leq C_{\varphi,d} N^{-1-1/d}$$

  for sufficiently 'smooth' φ (e.g. φ Lipschitz continuous if $\mathcal{X} = (0,1)^d$).

$\implies$ Improvements compared to multinomial resampling, but the gains decrease quickly with d.

Hilbert ordered stratified resampling

It is easy to see that for $d = 1$ and for ordered stratified/systematic resampling

$$\Big\| \frac{1}{N} \sum_{n=1}^{N} \delta_{x^{A^n}} - \sum_{n=1}^{N} W^n \delta_{x^n} \Big\|_\star \leq \frac{1}{N}, \quad a.s.$$

(the upper bound is almost optimal).

Remark: For multinomial resampling,

$$\Big\| \frac{1}{N} \sum_{n=1}^{N} \delta_{x^{A^n}} - \sum_{n=1}^{N} W^n \delta_{x^n} \Big\|_\star = \mathcal{O}(N^{-1/2} \log\log N), \quad a.s.$$

An interesting open question

For $d \geq 1$, the best we can hope for with ordered stratified resampling is

$$\left\| \frac{1}{N} \sum_{n=1}^{N} \delta_{x^{A^n}} - \sum_{n=1}^{N} W^n \delta_{x^n} \right\|_\star \leq \frac{C}{N^{1/d}}, \quad a.s.$$

However, its is known (Aistleitner and Dick, 2014, Theorem 1) that for every $M \geq 1$ there exists a point set $\hat{x}^{1:M}$ such that

$$\left\| \frac{1}{M} \sum_{m=1}^{M} \delta_{\hat{x}^m} - \sum_{n=1}^{N} W^n \delta_{x^n} \right\|_\star \leq 63\sqrt{d} \frac{\left(2 + \log_2(M)\right)^{\frac{3d+1}{2}}}{M} \tag{3}$$

Open question 3: How can we construct such a point set $\hat{x}^{1:M}$ (note that elements of $\hat{x}^{1:M}$ don't need to be elements of $x^{1:N}$).

Open question 4: Can we achieve (3) with a sampling/resampling algorithm, i.e. so that elements of $\hat{x}^{1:M}$ are elements of $x^{1:N}$?

$\Longrightarrow$ What is the best we can do with a sampling mechanism?

Extensible resampling schemes

As already mentioned, stratified and systematic resampling usually outperform multinomial resampling in practice.

In particular, the variance $\text{Var}\big(N^{-1}\sum_{n=1}^{N}\varphi(x^{A^n})\big)$ is always smaller with stratified/pivotal (thanks Guillaume!) resampling than with multinomial resampling

However, and contrary to multinomial resampling, stratified/systematic/ pivotal resampling **are not extensible**.

Open question 5: Does there exist an extensible resampling scheme for which $\text{Var}\big(N^{-1}\sum_{n=1}^{N}\varphi(x^{A^n})\big)$ is always smaller than with multinomial resampling?

$\Longrightarrow$ useful to increase the number of particles in SMC whenever needed.

Digression: Chairman assignment problem

It is known that, for $N > 1$,

$$\sup_{\{W^n\}_{n=1}^N} \inf_{(a^m)_{m\geq 1}} \sup_{n\in 1:N,\, m\in 1:M} \left|\sum_{m=1}^{M} \mathbb{1}(a^m = n) - MW^n\right| = 1 - \frac{1}{2(N-1)}$$

Tijdeman (1979) provides an algorithm that, for given $\{W^n\}_{n=1}^N$, generates a sequence $(a_m)_{m\geq 1}$ for which

$$D(\{W^n\}_{n=1}^N) := \sup_{n\in 1:N,\, m\in 1:M} \left|\sum_{m=1}^{M} \mathbb{1}(a^m = n) - MW^n\right| \leq 1 - \frac{1}{2(N-1)}$$

and thus has the optimal worst case behaviour; that is,

$$\sup_{\{W^n\}_{n=1}^N} D(\{W^n\}_{n=1}^N) = 1 - \frac{1}{2(N-1)}.$$

Open question 6: Can this result be used to provide an extensible (re)sampling algorithm with good properties?

MCMC basics

MCMC algorithms can be used to sample a trajectory $\{x_n\}_{t=1}^{N}$ of a Markov chain $(x_t)_{t\geq 1}$ having π as invariant distribution.

Then, $I = \mathbb{E}_\pi[\varphi]$ is estimated by

$$\frac{1}{N}\sum_{n=1}^{N}\varphi(x^n) \tag{4}$$

Problems:

1. The number of distinct values in the set $\{x_n\}_{t=1}^{N}$ is (much) smaller than N.
2. The random variables $(x_t)_{t\geq 1}$ are autocorrelated.

Consequently, N needs to be very large for the estimate (4) to give a good approximation of I

$\implies$ large memory cost and computing (4) may be costly when evaluating φ is expensive.

MCMC thinning: Goal

Construct a point set $\{\hat{x}^m\}_{m=1}^M$, with $M \ll N$, such that:

$$\hat{\pi}^M = \frac{1}{M}\sum_{m=1}^{M}\delta_{\hat{x}^m} \approx \frac{1}{N}\sum_{n=1}^{N}\delta_{x^n} =: \pi^N.$$

$\Longrightarrow$ Related problem: Construction of coresets in Big Data settings.

Main question: If we want $\{\hat{x}^m\}_{m=1}^M$ to be a sample from $\{x^n\}_{n=1}^N$, how should we define the inclusion probabilities in a meaningful way?

$\Longrightarrow$ Can a simple sampling procedure be used to construct a good set $\{\hat{x}^m\}_{m=1}^M$?

$\Longrightarrow$ use of reservoir method for online point selection?

MCMC thinning: First idea

Discard $k-1$ out of k observations, with k chosen so that $\text{cor}(X_t, X_{t+k})$ is small.

This idea has been proved (for $d = 1$) to improve statistical efficiency if k is well-chosen and the cost of evaluating φ is large enough (Owen, 2017).

$\Longrightarrow$ The main limitation is that the optimal thinning (i.e. k) depends on the particular function φ and is hard (impossible?) to establish in practice.

MCMC thinning: Second approach

Choose $\{\hat{x}^m\}_{m=1}^M$ such that

$$(\hat{x}^1, \dots, \hat{x}^M) \in \underset{(z_1,\dots,z_M)}{\arg\min}\, W_p\Big(\frac{1}{M}\sum_{m=1}^{M} \delta_{z_m}, \pi^N\Big)., \quad p \geq 1 \tag{5}$$

Then, on the one hand (Weed and Bach, 2017)

$$\mathbb{E}\big[W_p(\hat{\pi}^M, \pi^N)\big] = \mathcal{O}(M^{-\frac{1}{2p}})$$

while, on the other hand (Dudley, 1968)

$$\mathbb{E}\big[W_p(\pi^N, \pi)\big] = \mathcal{O}(N^{-\frac{1}{d}})$$

(for continuous π)

Remark: This approach has been proposed in Claici and Salomon (2018) to build coresets in Big Data setting.

Problem: Solving (5) for $p = 2$ is doable (see Claici and Salomon, 2018) but is very expensive.

Conclusion

Sampling from a finite population is a crucial step of SMC

It remains some open questions of interest for the SMC community, notably the study of resampling mechanisms under random ordering

$\Longrightarrow$ validity of systematic resampling? validity of deterministic resampling? CLT for SMC estimates based on stratified/pivotal resampling?

In MCMC/Big Data set-up, the (streaming) sampling problem of interest is largely unsolved.