CREST
CENTER FOR RESEARCH
IN ECONOMICS AND STATISTICS

**Série des Documents de Travail**

**n° 2017-36**

# Negative association, ordering and convergence of resampling methods

**M. GERBER**[1]
**N. CHOPIN**[2]
**N. WHITELEY**[3]



[1] School of Mathematics, University of Bristol, UK. E-mail: mathieu.gerber@bristol.ac.uk
[2] CREST-ENSAE, France. E-mail : Nicolas.chopin@ensae.fr
[3] School of Mathematics, University of Bristol, UK. E-mail : nick.whiteley@bristol.ac.uk

# Negative association, ordering and convergence of resampling methods

Mathieu Gerber* Nicolas Chopin† Nick Whiteley*

We study convergence and convergence rates for resampling. Our first main result is a general consistency theorem based on the notion of negative association, which is applied to establish the almost-sure weak convergence of measures output from Kitagawa's (1996) stratified resampling method. Carpenter et al's (1999) systematic resampling method is similar in structure but can fail to converge depending on the order of the input samples. We introduce a new resampling algorithm based on a stochastic rounding technique of Srinivasan (2001), which shares some attractive properties of systematic resampling, but which exhibits negative association and therefore converges irrespective of the order of the input samples. We confirm a conjecture made by Kitagawa (1996) that ordering input samples by their states in $\mathbb{R}$ yields a faster rate of convergence; we establish that when particles are ordered using the Hilbert curve in $\mathbb{R}^d$, the variance of the resampling error is $o(N^{-(1+1/d)})$ under mild conditions, where $N$ is the number of particles. We use these results to establish asymptotic properties of particle algorithms based on resampling schemes that differ from multinomial resampling.

*Keywords*: Negative association, resampling, particle filtering

## 1. Introduction

A resampling scheme is a randomized procedure that takes as input random samples $X^n$ with nonnegative weights $W^n \geq 0$, $n = 1, \ldots, N$, such that $\sum_{n=1}^N W^n = 1$, and returns as an output resampled variables $X^{A^n}$, where $A^n$ is a random index in $\{1, \ldots, N\}$, such that, in some sense,

$$\frac{1}{N} \sum_{n=1}^N \delta(X^{A^n}) \approx \sum_{n=1}^N W^n \delta(X^n). \tag{1}$$

Here $\delta(x)$ denotes the Dirac measure at point $x$ (this slightly unconventional notation will make our equations more readable).

*School of Mathematics, University of Bristol, UK.
†CREST-ENSAE, France.

Resampling appears in various statistical procedures. The present work is primarily motivated by resampling within Sequential Monte Carlo methods such as particle filters, see e.g., Doucet et al. (2001) where it plays an important role in stabilizing Monte Carlo errors. Resampling also appears in survey sampling, where it is called 'unequal probability sampling' (see e.g. Tillé, 2006) and in the weighted bootstrap procedure (Barbe and Bertail, 1995).

Multinomial resampling is perhaps the simplest technique, where given the weights, the indices $A^n$ are generated conditionally independently from the finite distribution that assigns probability $W^n$ to outcome $n$. In particle filtering it is common practice to replace multinomial resampling with techniques which are computationally faster and empirically more accurate. However, these advanced resampling techniques are generally not straight-forward to analyse because they induce complicated dependence between output samples, and various aspects of their behaviour are still not understood.

Following definitions and an account of what is known about existing resampling techniques, our first main result, Theorem 1 in Section 2, is a general consistency result for resampling based on the notion of *negative association* (Joag-Dev and Proschan, 1983). An application of this theorem gives, to our knowledge, the first proof of almost-sure weak convergence of the random probability measures output from the stratified resamping method of Kitagawa (1996). A notable feature of Theorem 1 is that, although its assumptions allow the input samples to be ordered in an arbitrary and possibly random manner, its proof involves establishing a necessary and sufficient condition for almost-sure weak convergence involving the Hilbert space-filling curve. To do so we build on Gerber and Chopin (2015) who use the Hilbert curve to derive and analyse a quasi-Monte Carlo version of sequential Monte Carlo samplers.

The systematic resampling method of Carpenter et al. (1999), which involves a technique called universal sampling by Whitley (1994), is a very popular and computationally cheap resampling technique, with the property that the number of offspring of any sample with weight $W$ in a population of $N$ is with probability 1 either $\lfloor NW \rfloor$ or $\lfloor NW \rfloor + 1$. However, depending on the order of the input particles, the error variance for systematic resampling can fail to converge to zero as $N \to \infty$, see Cappé et al. (2005, Ex. 7.4.6) and L'Ecuyer and Lemieux (2000). We complement this insight by providing a counter-example to almost-sure weak convergence. We then introduce a new resampling method, called Srinivasan Sampling Process (SSP) resampling, which corrects this deficiency: it also has the property that offspring numbers are of the form either $\lfloor NW \rfloor$ or $\lfloor NW \rfloor + 1$, but it provably converges irrespective of the order of input particles, by another application of our Theorem 1.

Kitagawa (1996) conjectured that in the case that the state-space is $\mathbb{R}$, ordering the particles input to stratified resampling according to their states leads to faster convergence. In particular, he suggested that the integrated square error between empirical cdf's before and after resampling behaves as $\mathcal{O}(N^{-2})$, compared to the standard Monte Carlo rate $\mathcal{O}(N^{-1})$ in the un-ordered case. We confirm this conjecture by proving, under mild conditions, that for stratified resampling on state-space $\mathbb{R}^d$ with input particles ordered by their states using the Hilbert curve, the variance of the resampling error is $o(N^{-(1+1/d)})$. Kitagawa also examined the behaviour of a deterministic resampling

scheme; we identify the variant of it which is optimal in terms of the Kolmogorov metric when the state-space is $\mathbb{R}$. We also prove the almost-sure weak consistency of stratified and systematic when the particles are Hilbert-ordered.

Finally, we discuss the implications of our results on particle filtering. In particular, we show that particle estimates are consistent when resampling schemes such as e.g. SSP or stratified resampling are used. In addition, we show that the ordered version of stratified resampling dominates other resampling schemes in terms of asymptotic variance of particle estimates.

All the proofs are gathered in the supplementary materials.

## 2. Preliminaries

### 2.1. Notation and conventions

Let $\mathcal{X}$ be an open subset of $\mathbb{R}^d$, $\mathbb{X}$ its Borel $\sigma$-algebra, $\mathcal{P}(\mathcal{X})$ the set of probability measures on $(\mathcal{X}, \mathbb{X})$, $\mathcal{P}_b(\mathcal{X}) \subset \mathcal{P}(\mathcal{X})$ the subset of measures in $\mathcal{P}(\mathcal{X})$ which admit a continuous and bounded density with respect to $\lambda_d$, the Lebesgue measure on $\mathcal{X}$, and $\mathcal{P}_f(\mathcal{X}) \subset \mathcal{P}(\mathcal{X})$ the subset of measures in $\mathcal{P}(\mathcal{X})$ whose support is a finite set.

For integers $1 \leq a < b$, we will often use the index shorthands $z_{a:b} = (z_a, \ldots, z_b)$ and $z^{a:b} = (z^a, \ldots, z^b)$, and let $1 : N = \{1, \ldots, N\}$.

For any measurable mapping $\varphi$ from $(\mathcal{X}, \mathbb{X})$ to some measurable space $(\mathcal{Y}, \mathbb{Y})$ and a probability measure $\pi \in \mathcal{P}(\mathcal{X})$, we write $\pi_\varphi$ for the pushforward of $\pi$ by $\varphi$. The set of continuous and bounded functions on $\mathcal{X}$ is denoted by $\mathcal{C}_b(\mathcal{X})$ and we use the symbol "$\overset{\mathrm{w}}{\Rightarrow}$" to denote weak convergence; that is, for sequence $(\pi^N)_{N\geq 1}$ in $\mathcal{P}(\mathcal{X})$ and $\pi \in \mathcal{P}(\mathcal{X})$,

$$\pi^N \overset{\mathrm{w}}{\Rightarrow} \pi \quad \Leftrightarrow \quad \lim_{N\to+\infty} \pi^N(\varphi) = \pi(\varphi), \quad \forall \varphi \in \mathcal{C}_b(\mathcal{X}).$$

Throughout the paper we consider a fixed probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which all random variables are defined. With $\mathcal{B}([0,1]^{\mathbb{N}})$ denoting the Borel $\sigma$-algebra on $[0,1]^{\mathbb{N}}$, let $U = (U_1, U_2, \ldots)$ be a $([0,1]^{\mathbb{N}}, \mathcal{B}([0,1]^{\mathbb{N}}))$-valued random variable on $(\Omega, \mathcal{F}, \mathbb{P})$, such that $\mathbb{P}$ makes $(U_1, U_2, \ldots)$ independent of each other and all other random variables, and such that each $U_i$ is distributed uniformly on $[0,1]$.

We note that one can choose a countable subset of $\mathcal{C}_b(\mathcal{X})$ that completely determines weak convergence, hence for random measures $(\pi^N)_{N\geq 1}$, the event $\{\pi^N \overset{\mathrm{w}}{\Rightarrow} \pi\}$ is measurable.

For $\pi \in \mathcal{P}(\mathcal{X})$, we denote by $\pi(\varphi)$ the expectation $\int_{\mathcal{X}} \varphi(x)\pi(\mathrm{d}x)$, and for a random variable $Z = (Z_1, \ldots, Z_d)$ whose distribution is $\pi$ we denote by $F_\pi(a) = \mathbb{P}(Z_1 \leq a_1, \ldots, Z_d \leq a_d)$, $a = (a_1, \ldots, a_d)$, its CDF (cumulative distribution function), and, when $d = 1$, by $F_\pi^-$ its generalized inverse: $F_\pi^-(u) = \inf\{x : F_\pi(x) \geq u\}$.

For each $N \geq 1$ we consider a distinguished collection of random variables $\zeta^N = (X^{n,N}, W^{n,N})_{n=1}^N$, with each $X^{n,N}$ valued in $\mathcal{X}$, each $(W^{n,N})_{n=1}^N$ valued in $\mathbb{R}^+$, and such that $\mathbb{P}$-a.s., $\sum_{n=1}^N W^{n,N} = 1$. When no confusion may arise, we suppress dependence on

$N$ and write $\zeta^N = (X^n, W^n)_{n=1}^N$. We associate with $\zeta^N$ the random measure

$$\pi^N = \sum_{n=1}^N W^n \delta(x^n),$$

the (random) CDF

$$F_N(n) = \sum_{m=1}^N W^m \mathbb{1}(m \leq n), \quad n \in 1:N,$$

and its inverse is denoted $F_N^-$.

To lighten notation we shall write $\mathbb{P}_{\zeta^N}(\cdot)$, $\mathbb{E}_{\zeta^N}[\cdot]$, $\mathrm{Var}_{\zeta^N}[\cdot]$, $\mathrm{Cov}_{\zeta^N}[\cdot,\cdot]$ for conditional probability, expectation, variance and covariance given $\zeta^N$.

Let $\mathcal{Z}^N = \{(x,w) \in \mathcal{X}^N \times \mathbb{R}_+^N : \sum_{n=1}^N w_n = 1\}$ and define the disjoint union $\mathcal{Z} \coloneqq \bigcup_{N=1}^{+\infty} \mathcal{Z}^N$. So we may think of $\zeta^N$ as a random point in $\mathcal{Z}^N$, and hence $\mathcal{Z}$.

**Definition 1.** *$\mathcal{X} \subseteq \mathbb{R}^d$ is said to be cubifiable if there exist measurable sets $\mathcal{X}_i \subseteq \mathbb{R}$, $i = 1, \ldots, d$, such that*

1. *$\mathcal{X} = \times_{i=1}^d \mathcal{X}_i$;*
2. *For any $i \in 1:d$, there exists a $C^1$-diffeomorphism $\psi_i : \mathcal{X}_i \to (0,1)$ which is strictly increasing on $\mathcal{X}_i$.*

*We shall write $\psi(x) = (\psi_1(x_1), \ldots, \psi_d(x_d))$, $x = x_{1:d} \in \mathcal{X}$, the resulting $C^1$-diffeomorphism from $\mathcal{X}$ into $(0,1)^d$.*

We recall the reader that function $\psi : \mathcal{X} \to (0,1)^d$ is a $C^1$-diffeomorphism if it is a bijection and its inverse $\psi^{-1} : (0,1)^d \to \mathcal{X}$ is continuously differentiable. In what follows, for a cubifiable set $\mathcal{X}$ we denote by $\mathcal{D}(\mathcal{X})$ the set of all $C^1$-diffeomorphisms from $\mathcal{X}$ into $(0,1)^d$ that verify the conditions of Definition 1.

Cubifiable sets are essentially sets that can be written as $\mathcal{X} = \times_{i=1}^d (a_i, b_i)$ for some $a_i, b_i \in \mathbb{R} \cup \{-\infty, +\infty\}$. The point of these sets is to be able to work 'as if' $\mathcal{X} = (0,1)^d$. The hypercube $(0,1)^d$ will indeed play a key role below because the Hilbert space-filling curve, which is essential in this work, is defined on this hypercube.

Most of the results presented below assume that the limiting distribution $\pi$ admits a continuous and bounded density. Consequently, to work 'as if' $\mathcal{X} = (0,1)^d$ we will often assume that $\pi$ belongs to

$$\widetilde{\mathcal{P}}_b(\mathcal{X}) = \Big\{\pi \in \mathcal{P}_b(\mathcal{X}) : \exists \psi \in \mathcal{D}(\mathcal{X}) \text{ s.t. } \pi_\psi \in \mathcal{P}_b((0,1)^d)\Big\}.$$

The following result provides a sufficient condition to have $\pi \in \widetilde{\mathcal{P}}_b(\mathcal{X})$. We denote by $p_\pi$ the density (w.r.t $\lambda_d$) of $\pi \in \mathcal{P}_b(\mathcal{X})$ and, for $I \subset 1:d$, we write $x_I = (x_i,\, i \in I)$ and $x_{\setminus I} = (x_i,\, i \notin I)$.

**Lemma 1.** *Let $\mathcal{X}$ be a cubifiable set, $\delta > 0$ and $\pi \in \mathcal{P}_b(\mathcal{X})$ such that $\forall I \subseteq 1:d$, $\forall x_{\setminus I} \in \times_{i \notin I} \mathcal{X}_i$, $\sup_{x_I \in \times_{i \in I} \mathcal{X}_i} p_\pi(x) \prod_{i \in I} |x_i|^{1+\delta} \leq C$ for some $C < \infty$. Then $\pi \in \widetilde{\mathcal{P}}_b(\mathcal{X})$.*

Recall that $\sup_{x\in\mathbb{R}} p_\pi(x)|x| < +\infty$ for any $\pi \in \mathcal{P}_b(\mathbb{R})$. Therefore, as $\delta > 0$ is arbitrary in the lemma, very few extra conditions on the tails of $\pi \in \mathcal{P}_b(\mathbb{R})$ are needed in order to have $\pi \in \widetilde{\mathcal{P}}_b(\mathbb{R})$ when $d = 1$. When $d > 1$, assuming that $\pi \in \widetilde{\mathcal{P}}_b(\mathbb{R}^d)$ is more restrictive since the lemma requires some uniformity in the behaviour of tails. However, we note that members of $\widetilde{\mathcal{P}}_b(\mathbb{R}^d)$ may not have a first moment and therefore the sufficient condition of Lemma 1 appears to be quite weak.

## 2.2. Resampling schemes: definitions and properties

**Definition 2.** *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.*

Instances of the function $a_N^n$ are given below. We shall use the shorthands $\rho(z)$ for the random measure $\rho(U, z)$, $z \in \mathcal{Z}^N$, and $A^n$ for the random indices $a_N^n(U, \zeta^N)$. Introducing the quantities,

$$\#^n(u, z) = \text{card}\{i \in 1 : N \text{ s.t. } a_N^i(u, z) = n\}, \quad \Delta_{\rho,z}^n = \#^n(U, z) - NW^n, \tag{2}$$

a resampling scheme $\rho$ is said to be *unbiased* if, for any $N \geq 1$, $n \in 1 : N$ and $z \in \mathcal{Z}^N$,

$$\mathbb{E}[\Delta_{\rho,z}^n] = 0.$$

We now define the resampling schemes of primary interest in this work.

- **Multinomial resampling:** $\rho_{\text{multi}}$ such that

  $$a_N^n(u, \zeta^N) = F_N^-(u_n).$$

  In this case the $a_N^n(U, \zeta^N)$ are i.i.d. (independent and identically distributed) draws from the distribution which assigns probability $W^n$ to outcome $n$.

- **Stratified resampling:** $\rho_{\text{strat}}$ such that

  $$a_N^n(u, \zeta^N) = F_N^- \left( \frac{n - 1 + u_n}{N} \right).$$

- **Systematic resampling:** $\rho_{\text{syst}}$ such that

  $$a_N^n(u, \zeta^N) = F_N^- \left( \frac{n - 1 + u_1}{N} \right).$$

The following definition captures the notion of almost sure weak convergence of the random measures $(\pi^N)_{N\geq 1}$ which we shall study and is similar to condition (9) in Crisan and Doucet (2002).

**Definition 3.** *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 $(\zeta^N)_{N\geq 1}$ such that $\pi^N \overset{w}{\Rightarrow} \pi$, $\mathbb{P}$-a.s., one has*

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

Note that in this definition, whether or not $\pi^N \overset{\mathrm{w}}{\Rightarrow} \pi$ does not depend on the ordering of the components of each $\zeta^N = (X^{1:N}, W^{1:N})$.

It is well known that multinomial, stratified and systematic resampling are unbiased. An account of various properties of these methods can be found in Cappé et al. (2005, Sec. 7.4).

Crisan and Doucet (2002, Lemma 2) shows that multinomial resampling is $\mathcal{P}(\mathcal{X})$-consistent for any measurable set $\mathcal{X} \subseteq \mathbb{R}^d$.

It is easy to show (Stein, 1987; Douc et al., 2005) that stratified resampling dominates multinomial resampling in terms of variance, i.e.,

$$\mathrm{Var}\left[\rho_{\mathrm{strat}}(z)(\varphi)\right] \leq \mathrm{Var}\left[\rho_{\mathrm{multi}}(z)(\varphi)\right], \quad \forall z \in \mathcal{Z}$$

for any measurable $\varphi : \mathcal{X} \to \mathbb{R}$. Similar results are harder to derive for systematic resampling, owing to the strong dependencies between the resampled indices. However, it is known (Douc et al., 2005) that the variance of $\rho_{\mathrm{syst}}(\zeta^N)(\varphi)$ may not converge to 0 as $N \to +\infty$ (see also L'Ecuyer and Lemieux, 2000, for an explanation of this phenomenon).

# 3. Convergence of resampling schemes based on negative association

In this section we present a general consistency result for unbiased resampling schemes producing a set of random variables that satisfy a negative association condition (see Definition 4). More precisely, one of the hypotheses of Theorem 1 below is negative association of $(\#^n(U,z))_{n=1}^N$, where $\#^n(U,z)$ as in (2) is the number of times the resampling operation $\rho(z)$ selects $X^n$. Interestingly, this negative association hypothesis does not explicitly involve the resampled values $(X^{A^n})_{n=1}^N$ and, as shown below, is fulfilled by multinomial resampling, where given $\zeta^N$, $(X^{A^n})_{n=1}^N$ is a set of conditionally i.i.d. random variables, by stratified resampling, where $(X^{A^n})_{n=1}^N$ is a set of independent but not identically distributed random variables, and by a new SSP resampling scheme, introduced in the sequel, that produces conditionally dependent resampled values.

## 3.1. A general consistency result

Before stating the main result of this section we recall the definition of negatively associated (NA) random variables (Joag-Dev and Proschan, 1983).

**Definition 4.** *A collection of random variables $(Z^n)_{n=1}^N$ are negatively associated if, for every pair of disjoint subsets $I_1$ and $I_2$ of $\{1, \ldots N\}$,*

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

*for all coordinatewise non-decreasing functions $\varphi_1$ and $\varphi_2$, such that for $k \in \{1,2\}$, $\varphi_k : \mathbb{R}^{|I_k|} \to \mathbb{R}$ and such that the covariance is well-defined.*

**Theorem 1.** *Let $\mathcal{X}$ be a cubifiable set and $\rho$ be an unbiased resampling scheme such that the following conditions hold:*

*($H_1$) For any $N \geq 1$ and $z \in \mathcal{Z}^N$, the random variables $(\#^n(U,z))_{n=1}^N$ are negatively associated;*

*($H_2$) 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, $\rho$ is $\widetilde{\mathcal{P}}_b(\mathcal{X})$-consistent.*

The strategy of the proof is the following. In a first step, we show that there exists a sequence $(\sigma_N^*)_{N\geq 1}$, where each $\sigma_N^*$ is a permutation of $1:N$ depending on $z \in \mathcal{Z}^N$, such that $\rho$ is $\widetilde{\mathcal{P}}_b(\mathcal{X})$-consistent *if and only if*

$$\lim_{N\to+\infty} \max_{m\in 1:N} \Big| \sum_{n=1}^{m} \Delta_{\rho,z}^{\sigma_N^*(n)} \Big| = 0, \quad \mathbb{P}-a.s. \tag{3}$$

for any $z \in \mathcal{Z}^N$. The key tool to construct the sequence $(\sigma_N^*)_{N\geq 1}$ is the Hilbert space-filling curve; the details of which we postpone to Section 4, after applications of Theorem 1. In a second step, we show that the hypotheses $(H_1)$ and $(H_2)$ are sufficient to establish (3), via a maximal inequality for negatively associated random variables.

Before proceeding further we provide some insights about the conditions of Theorem 1. To this end we fix $N \geq 1$ and $z \in \mathcal{Z}^N$. First, because $\sum_{n=1}^N \#^n(U,z) = N$, $\mathbb{P}$-a.s., it is intuitively apparent that some random variables in the set $(\#^n(U,z))_{n=1}^N$ have to be negatively correlated in order for this constraint to hold. Condition $(H_1)$ may be understood as ruling out the existence of positively correlated random variables in this set. The rationale behind this condition is that positive correlation between offspring numbers must be introduced with care, otherwise the necessary condition (3) may not hold. The next proposition formalizes these ideas in the context of systematic resampling. Note that condition $(H_1)$ must hold for all $z$, and $(H_2)$ is uniform in $z$, and hence all permutations of the input particles.

**Proposition 1.** *The systematic resampling scheme $\rho_{\text{syst}}$ verifies $(H_2)$ with $r_N = 1$ but not $(H_1)$. In addition, there exist a cubifiable set $\mathcal{X} \subseteq \mathbb{R}^d$, a probability measure $\pi \in \widetilde{\mathcal{P}}_b(\mathcal{X})$ and a sequence $(\zeta^N)_{N\geq 1}$ such that $\pi^N \overset{w}{\Rightarrow} \pi$ but $\mathbb{P}\big(\rho_{\text{syst}}(\zeta^N) \overset{w}{\Rightarrow} \pi\big) < 1$.*

The first part of the proposition gives a known result that we recall to emphasize the fact that it is only condition $(H_1)$ of Theorem 1 which systematic resampling violates.

Note also that $r_N = 1$ is the smallest possible value of $r_N$ for which $(H_2)$ may hold. The proof of the existence part of Proposition 1 is constructive but somewhat technical.

Condition $(H_1)$ alone is not enough to guarantee the consistency of an unbiased resampling scheme. Indeed, if we consider for instance $\rho$ such that $\#^n(U, z) = N$ with probability $W^n$, it is easily checked $\rho$ is unbiased and such that condition $(H_1)$ is fulfilled while this resampling scheme is obviously not $\widetilde{\mathcal{P}}_b(\mathcal{X})$-consistent. This observation justifies the additional condition $(H_2)$, which is fulfilled for various popular resampling mechanism, see the next subsection.

## 3.2. Applications of Theorem 1

### 3.2.1. Multinomial resampling

As already mentioned, it is a known result that multinomial resampling is $\mathcal{P}(\mathcal{X})$-consistent for any measurable $\mathcal{X} \subseteq \mathbb{R}^d$ (Crisan and Doucet, 2002, Lemma 2). We next briefly explain how Theorem 1 may be applied to obtain a similar result.

**Corollary 1.** *Let $\mathcal{X}$ be a cubifiable set. Then, the resampling scheme $\rho_{\text{multi}}$ verifies conditions $(H_1)$ and $(H_2)$ of Theorem 1 and is therefore $\widetilde{\mathcal{P}}_b(\mathcal{X})$-consistent.*

To establish $(H_1)$, we may view the random variables $(\#^N(U, z))_{n=1}^N$ as occupancy numbers in a "Balls and Bins" problem; then $(H_1)$ is a direct consequence of Theorem 13 in Dubhashi and Ranjan (1998) (which establishes the NA of occupancy numbers).

For similar reasons, the conditions of Theorem 1 are also satisfied by the residual resampling scheme of Liu and Chen (1998).

### 3.2.2. Stratified resampling

To the best of our knowledge the next theorem is the first weak convergence result for Kitagawa's (1996) stratified resampling scheme.

**Corollary 2.** *Let $\mathcal{X}$ be a cubifiable set. Then, the resampling scheme $\rho_{\text{strat}}$ verifies conditions $(H_1)$ and $(H_2)$ of Theorem 1 and is therefore $\widetilde{\mathcal{P}}_b(\mathcal{X})$-consistent.*

It is worth noting that the conditions of Theorem 1 are also satisfied by the stratified version of the residual resampling scheme of Liu and Chen (1998), where the multinomial resampling part is replaced by a stratified resampling step. Denoting these two resampling schemes by $\rho_{\text{res/multi}}$ and $\rho_{\text{res/strat}}$ respectively, the stratified version of residual resampling has the interesting property that, for any measurable $\varphi : \mathcal{X} \to \mathbb{R}$ we have (see Douc et al., 2005, for the second inequality)

$$\operatorname{Var}\big[\rho_{\text{res/strat}}(z)(\varphi)\big] \leq \operatorname{Var}\big[\rho_{\text{res/multi}}(z)(\varphi)\big] \leq \operatorname{Var}\big[\rho_{\text{multi}}(z)(\varphi)\big], \quad \forall z \in \mathcal{Z}.$$

In addition, $\rho_{\text{res/strat}}$ has the advantage to be easier and slightly cheaper to implement than $\rho_{\text{res/multi}}$.

### 3.2.3. SSP resampling

To overcome the lack of consistency (in the sense of Definition 3) of systematic resampling we introduce below a new resampling scheme, named SSP (for Srinivasan Sampling Process) resampling, which both verifies the NA condition $(H_1)$ and shares the key property of systematic resampling concerning the numbers of offspring, namely that $|\Delta^n_{\rho_{\text{syst}},\xi^N}| \leq 1$ for all $n \in 1:N$, $\mathbb{P}$-a.s.

The underlying idea of SSP resampling is to see the resampling scheme as a rounding operation, where the vector of 'weights' $(NW^1, \ldots, NW^N)$ is $\mathbb{P}$-a.s. transformed into a point $(Y_1(U), \ldots, Y_N(U))$ in $\mathbb{N}^N$ satisfying the constraint $\sum_{n=1}^N Y_n(U) = N$.

Before proceeding further we recall the terminology that for $\xi \in \mathbb{R}^+$, the random variable $Y : \Omega \to \mathbb{N}$ is called a randomized rounding of $\xi$ if

$$\mathbb{P}\big(Y = \lfloor \xi \rfloor + 1\big) = \xi - \lfloor \xi \rfloor, \quad \mathbb{P}\big(Y = \lfloor \xi \rfloor\big) = 1 - \big(\xi - \lfloor \xi \rfloor\big).$$

Hence, any algorithmic technique for constructing randomized roundings that takes as input a vector $(\xi_1, \ldots, \xi_N) \in \mathbb{R}_+^N$ and returns $\mathbb{P}$-a.s. as output a vector $(Y_1(U), \ldots, Y_N(U)) \in \mathbb{N}^N$ verifying $\sum_{n=1}^N Y_n(U) = \sum_{n=1}^N \xi_n$ may be used to construct an unbiased resampling mechanism; systematic resampling is an example of a resampling scheme constructed in this way.

The SSP resampling scheme $\rho_{\text{ssp}} : [0,1]^{\mathbb{N}} \times \mathcal{Z} \to \mathcal{P}_f(\mathcal{X})$ is based on the Srinivasan's (2001) randomized rounding technique and is presented in Algorithm 1. To see that this latter indeed defines a randomized rounding process it suffices to note that step (2) leaves unchanged the expectation of the vector $(Y^n_{\text{ssp}}(U))_{n=1}^N$ while, by construction, each iteration of the algorithm leaves the quantity $\sum_{n=1}^N Y^n_{\text{ssp}}(U)$ unchanged with $\mathbb{P}$-probability one. By Dubhashi et al. (2007, Theorem 5.1; see also Kramer et al., 2011) the collection of random variables $(Y^n_{\text{ssp}}(U))_{n=1}^N$ produced by the SSP described in Algorithm 1 is NA. Together with Theorem 1, this result allows to readily show the consistency of $\rho_{\text{ssp}}$.

**Corollary 3.** *Let $\mathcal{X}$ be a cubifiable set. Then, the resampling scheme $\rho_{\text{ssp}}$ verifies conditions $(H_1)$ and $(H_2)$ of Theorem 1 and is therefore $\widetilde{\mathcal{P}}_b(\mathcal{X})$-consistent.*

Algorithm 1 has complexity $\mathcal{O}(N)$, like other standard resampling schemes. An open question is whether or not SSP resampling dominates multinomial resampling in terms of variance. See Section 5.5 for a numerical comparison.

The resampling scheme proposed in Crisan (2001) may also be interpreted as a randomized rounding technique. However, to the best of our knowledge, no convergence result exists for this resampling scheme.

## 4. Convergence of ordered resampling schemes

Kitagawa (1996, Appendix A) provided numerical results about the behaviour of stratified resampling in the case that $d = 1$ and the input particles are ordered according to their states. He conjectured that in this situation, the error of stratified resampling is of

**Algorithm 1** SSP resampling

**Inputs:** $u \in [0,1]^{\mathbb{N}}$ and $(\xi_1, \ldots, \xi_N) \in \mathbb{R}_+^N$ such that $\sum_{n=1}^N \xi_n \in \mathbb{N}$.
**Output**: $\big(Y_{\mathrm{ssp}}^1(u), \ldots, Y_{\mathrm{ssp}}^N(u)\big) \in \mathbb{N}^N$ such that $\sum_{n=1}^N Y_{\mathrm{ssp}}^n(u) = \sum_{n=1}^N \xi_n$.
Initialization: $\big(Y_{\mathrm{ssp}}^1(u), \ldots, Y_{\mathrm{ssp}}^N(u)\big) \leftarrow (\xi_1, \ldots, \xi_N)$, $(n, m, k) \leftarrow (1, 2, 1)$
Iterate the following steps until $\big(Y_{\mathrm{ssp}}^1(u), \ldots, Y_{\mathrm{ssp}}^N(u)\big) \in \mathbb{N}^N$:

**(1)** Let $\delta$ be the smallest number in $(0,1)$ such that at least one of $Y_{\mathrm{ssp}}^n(u)+\delta$ or $Y_{\mathrm{ssp}}^m(u)-\delta$ is an integer, and let $\epsilon$ be the smallest number in $(0,1)$ such that at least one of $Y_{\mathrm{ssp}}^n(u) - \epsilon$ or $Y_{\mathrm{ssp}}^m(u) + \epsilon$ is an integer.

**(2)** If $u_k \leq \epsilon/(\delta + \epsilon)$ set $(Y_{\mathrm{ssp}}^n(u), Y_{\mathrm{ssp}}^m(u)) \leftarrow (Y_{\mathrm{ssp}}^n(u) + \delta, Y_{\mathrm{ssp}}^m(u) - \delta)$; otherwise set $(Y_{\mathrm{ssp}}^n(u), Y_{\mathrm{ssp}}^m(u)) \leftarrow (Y_{\mathrm{ssp}}^n(u) - \epsilon, Y_{\mathrm{ssp}}^m(u) + \epsilon)$.

**(3)** Update $n$ and $m$ as follows:

1. If $\big(Y_{\mathrm{ssp}}^n(u), Y_{\mathrm{ssp}}^m(u)\big) \in \mathbb{N}^2$, $(n, m) \leftarrow (m+1, m+2)$;
2. If $Y_{\mathrm{ssp}}^n(u) \in \mathbb{N}$ and $Y_{\mathrm{ssp}}^m(u) \notin \mathbb{N}$ set $(n, m) \leftarrow (m, m+1)$;
3. if $Y_{\mathrm{ssp}}^n(u) \notin \mathbb{N}$ and $Y_{\mathrm{ssp}}^m(u) \in \mathbb{N}$ set $(n, m) \leftarrow (m, m+1)$.

**(4)** $k \leftarrow k + 1$

size $\mathcal{O}(N^{-2})$, compared to $\mathcal{O}(N^{-1})$ without the ordering. He also considered a deterministic resampling scheme, and found that in same $d = 1$ case and with ordered particles, it also exhibited $\mathcal{O}(N^{-2})$ convergence.

The purpose of this section is to provide a rigorous investigation of this topic. While Kitagawa (1996) measured the error introduced by a resampling scheme by the integrated square error between empirical CDF's before and after resampling, we compare below the probability measures before and after resampling by comparing their expectations for some test functions. Notably, we present in this section results on the convergence rate of the variance of stratified resampling when applied on ordered input particles. We first consider the case $d = 1$ and then the general $d \geq 1$ case in which particles input to resampling are ordered using the Hilbert space filling curve.

## 4.1. Ordered resampling schemes on univariate sets

In this subsection we present results for a univariate set $\mathcal{X}$, which is the set-up considered by Kitagawa (1996). The existence of a natural order in this context greatly facilitates the presentation and allows to derive more precise convergent results than in multivariate settings.

We denote below by $\rho_{\mathrm{strat}}^*$ the ordered stratified resampling scheme; that is, $\rho_{\mathrm{strat}}^* : [0,1]^{\mathbb{N}} \times \mathcal{Z} \to \mathcal{P}_f(\mathcal{X})$ is defined by

$$\rho_{\mathrm{strat}}^*(u, z) = \rho_{\mathrm{strat}}\big(u, (z_{\sigma_N^*(n)})_{n=1}^N\big), \quad (u, z) \in [0,1]^{\mathbb{N}} \times \mathcal{Z}^N$$

with $\sigma_N^*$ a permutation of $1:N$ such that $z_{\sigma_N^*(1)} \leq \cdots \leq z_{\sigma_N^*(N)}$. In words, $\rho^*_{\text{strat}}(\zeta^N)$ simply amounts to apply the stratified resampling scheme $\rho_{\text{strat}}$ on the ordered input point set $\zeta^{N,\sigma_N^*} := (X^{\sigma_N^*(n)}, W^{\sigma_N^*(n)})_{n=1}^N$. Notice that $\rho^*_{\text{strat}}(\zeta^N)$ is such that

$$X^{A^n} = F^-_{\pi^N}\left(\frac{n-1+U_n}{N}\right), \quad n \in 1:N; \tag{4}$$

that is, the resampled particles $(X^{A^n})_{n=1}^N$ are obtained by sampling from the empirical distribution $\pi^N$ using the stratified point set $((n-1+U_n)/N)_{n=1}^N$.

The following theorem shows that under mild conditions the variance induced by ordered stratified resampling converges faster than $N^{-1}$. In addition, it also provides condition under which one has a non-asymptotic bound of size $N^{-2}$ for this resampling method.

**Theorem 2.** *Let $\mathcal{X} \subseteq \mathbb{R}$ be a cubifiable set. Then, the following results hold:*

1. *Let $\pi \in \widetilde{\mathcal{P}}_b(\mathcal{X})$ have a strictly positive density, and $(\zeta^N)_{N\geq 1}$ be such that $\pi^N \overset{w}{\Rightarrow} \pi$, $\mathbb{P}$-a.s., and such that, $\lim_{N\to+\infty}\left(\max_{n\in 1:N} W^{n,N}\right) = 0$, $\mathbb{P}$-a.s. Then, for any $\varphi \in \mathcal{C}_b(\mathcal{X})$, $\text{Var}_{\zeta^N}\left[\rho^*_{\text{strat}}(\zeta^N)(\varphi)\right] = o(1/N)$, $\mathbb{P}$-a.s.*
2. *Let $\varphi : \mathcal{X} \to \mathbb{R}$ be a continuously differentiable function such that, for a $\delta > 0$, we have $\sup_{x\in\mathcal{X}} \frac{\mathrm{d}\varphi}{\mathrm{d}x}(x)|x|^{1+\delta} < +\infty$. Then, there exists a constant $C_\varphi < +\infty$ such that, for all $N \geq 1$,*

$$\text{Var}\left[\rho^*_{\text{strat}}(z)(\varphi)\right] \leq C_\varphi\, N^{-2}, \qquad \forall z \in \mathcal{Z}^N.$$

The second observation of Kitagawa (1996, p.23) is that deterministic resamplimg mechanisms may be used when applied to the ordered input particles $\zeta^{N,\sigma_N^*}$. In particular, he considered a resampling scheme defined by (4) but with the random variables $(U_n)_{n=1}^N$ replaced by a deterministic point in $\alpha \in (0,1)$. In the notation of this work, for $\alpha \in (0,1)$ Kitagawa (1996) considered the resampling scheme $\rho^*_\alpha$ defined by $\rho^*_\alpha(u,z) = \rho^*_{\text{strat}}(\alpha_\infty, z)$ with $\alpha_\infty$ the vector in $(0,1)^{\mathbb{N}}$ having $\alpha$ in all its entries. The consistency of this deterministic resampling mechanism trivially follows from Corollary 4 (see below) and the fact that

$$\|F_{\rho^*_\alpha(\zeta^N)} - F_{\pi^N}\|_\infty \leq \frac{1}{2N} + \left|\frac{\alpha - 1/2}{N}\right|. \tag{5}$$

Notice that the right-hand side of this expression is minimized for $\alpha = 0.5$. In fact, it is not difficult to check that the resampling scheme $\rho^*_{1/2}$ is optimal in the sense that it minimises $\|F_{\rho(\zeta^N)} - F_{\pi^N}\|_\infty$ among all resampling schemes $\rho$. One rationale for trying to minimize this quantity when considering deterministic resampling schemes is given by the generalized Koksma-Hlawka (Aistleitner and Dick, 2015, Theorem 1) which implies that

$$\left|\rho(\zeta^N)(\varphi) - \pi^N(\varphi)\right| \leq V(\varphi)\|F_{\rho(\zeta^N)} - F_{\pi^N}\|_\infty \tag{6}$$

with $V(\varphi)$ the variation of $\varphi$ in $\mathcal{X}$.

We end this subsection by noting that inequality (5) shows that systematic resampling is consistent when applied on the ordered input particles $\zeta^{N,\sigma_N^*}$.

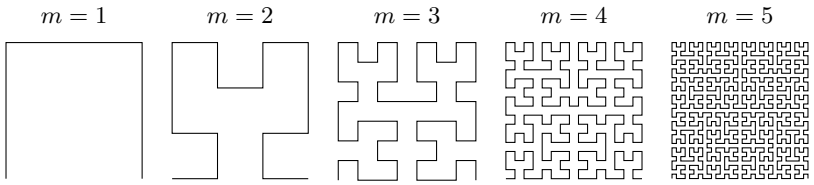


Figure 1: The Hilbert curve in dimension $d = 2$ is defined as the limit of this sequence. (source: Marc van Dongen)

## 4.2. Hilbert-ordered resampling schemes

In this subsection we generalize the results presented above to any dimension $d \geq 1$. The main challenge when $d > 1$ is to find an ordering of particles $\zeta^N$ which allows to improve upon the un-ordered version of the resampling scheme. Below we consider an ordering based on the Hilbert space filling curve.

### 4.2.1. Hilbert space filling curve and related definitions

For $\pi, \pi' \in \mathcal{P}(\mathcal{X})$, we use below the shorthand $\|\pi - \pi'\|_\star = \|F_\pi - F_{\pi'}\|_\infty$; note that the 'star' metric $\|\cdot\|_\star$ is the multivariate generalization of the Kolmogorov metric. The star discrepancy of the point set $u_{1:N}$ in $[0,1]^d$ is defined by

$$D_N^\star(u_{1:N}) = \left\| N^{-1} \sum_{i=1}^N \delta_{u_i} - \lambda_d \right\|_\star$$

The Hilbert curve is a space-filling curve, that is a continuous surjective function $H : [0,1] \to [0,1]^d$. It is defined as the limit of the sequence of functions depicted (for $d = 2$) in Figure 1, is Hölder continuous with exponent $1/d$ and is a measure-preserving mapping in the sense that $\lambda_d(H(I)) = \lambda_1(I)$ for any measurable set $I \in [0,1]$. This last property plays a crucial role in the derivation of the consistency results presented in the next subsection while the Hölder continuity of the Hilbert curve is central in our analysis of the variance of Hilbert-ordered stratified resampling (Section 4.2.3).

In what follows, we assume that $H(0) = (0, \ldots, 0)$. The Hilbert curve admits a one-to-one Borel measurable pseudo-inverse $h : [0,1]^d \to [0,1]$ such that $H(h(x)) = x$ for all $x \in [0,1]^d$, as shown in the next proposition.

**Proposition 2.** *There exists a one-to-one Borel measurable function $h : [0,1]^d \to [0,1]$ such that $H(h(x)) = x$ for all $x \in [0,1]^d$.*

For $d = 1$, we simply take $H(x) = h(x) = x$ for $x \in [0,1]$. Note that because $H(0) = (0, \ldots, 0)$ and $h$ is not surjective, $h((0,1)^d) \subsetneq (0,1)$. Although we refer to $H$ as *the* Hilbert curve in this work, there exist in fact several Hilbert curves.

For a cubifiable set $\mathcal{X}$ and dipheomorphism $\psi \in \mathcal{D}(\mathcal{X})$, we denote by $h_{\mathcal{X},\psi}$ the one-to-one mapping $x \mapsto h \circ \psi(x)$. To simplify the notation in what follows, we associate to a

cubifiable set $\mathcal{X}$ a dipheomorphism $\psi_{\mathcal{X}} \in \mathcal{D}(\mathcal{X})$ and use the shorthand $h_{\mathcal{X}} = h_{\mathcal{X},\psi_{\mathcal{X}}}$. In particular, when $\mathcal{X} = (0,1)^d$ we assume henceforth that $\psi_{\mathcal{X}}(x) = x$ for all $x \in \mathcal{X}$.

We now define $\sigma_N^*$ as a permutation of $1:N$ such that

$$h_{\mathcal{X}}(z_{\sigma_N^*(1)}) \leq \ldots \leq h_{\mathcal{X}}(z_{\sigma_N^*(N)})$$

and use it to extend the definition of the ordered stratified resampling scheme $\rho^*_{\text{strat}}$ introduced in the previous subsection to any $d \geq 1$; that is, for any $d \geq 1$ we define $\rho^*_{\text{strat}} : [0,1]^{\mathbb{N}} \times \mathcal{Z} \to \mathcal{P}_f(\mathcal{X})$ by

$$\rho^*_{\text{strat}}(u,z) = \rho_{\text{strat}}\big(u, (z_{\sigma_N^*(n)})_{n=1}^N\big), \quad (u,z) \in [0,1]^{\mathbb{N}} \times \mathcal{Z}^N.$$

The resampling scheme $\rho^*_{\text{strat}}(\zeta^N)$ is such that

$$X^{A^n} = \psi_{\mathcal{X}}^{-1} \circ H\left(F^-_{\pi^N_{h_{\mathcal{X}}}}\left(\frac{n-1+U_n}{N}\right)\right), \quad n \in 1:N \tag{7}$$

and thus $\rho^*_{\text{strat}}$ amounts to first sample from the empirical distribution $\pi^N_{h_{\mathcal{X}}}$ using the stratified point set $((n-1+U_n)/N)_{n=1}^N$ and then to 'project' the resulting sample in the original set $\mathcal{X}$ using the mapping $\psi_{\mathcal{X}}^{-1} \circ H$. Note that representation (7) of $\rho^*_{\text{strat}}$ extends the one given in (4) for $d = 1$ to any $d \geq 1$.

The ordered systematic resampling scheme $\rho^*_{\text{syst}}$ is defined in a similar way.

Although this is not apparent from the notation, when $d > 1$ the resampling schemes $\rho^*_{\text{strat}}$ and $\rho^*_{\text{syst}}$ depend on $\psi_{\mathcal{X}}$ through $\sigma_N^*$, and therefore different choices for $\psi_{\mathcal{X}}$ lead to different resampling mechanisms. Consequently, convergence results for these two resampling schemes will assume that the limiting distribution $\pi$ on $\mathcal{X}$ belongs to the subset $\mathcal{P}^*_b(\mathcal{X})$ of $\widetilde{\mathcal{P}}_b(\mathcal{X})$ defined by $\mathcal{P}^*_b(\mathcal{X}) = \{\pi \in \mathcal{P}_b(\mathcal{X}) : \pi_{\psi_{\mathcal{X}}} \in \mathcal{P}_b((0,1)^d)\}$.

To fix the ideas, when $\mathcal{X} = \mathbb{R}^d$ one can take for $\psi_{\mathcal{X}}$ the dipheomorphism $\psi(x) = (\tilde{\psi}(x_1), \ldots, \tilde{\psi}(x_d))$, with $\tilde{\psi} \in \mathcal{D}(\mathbb{R})$ defined by

$$\tilde{\psi}(x) = \frac{1}{2} + \frac{\sqrt{4+x^2}-2}{2x}\mathbb{1}_{\mathbb{R}\setminus\{0\}}(x), \quad x \in \mathbb{R}.$$

In this case, following Lemma 1, it is easily checked that $\pi \in \widetilde{\mathcal{P}}^*_b(\mathcal{X})$ when $\pi \in \mathcal{P}_b(\mathcal{X})$ is such that $\forall I \subseteq 1:d$, $\forall x_{\setminus I} \in \times_{i \notin I}\mathcal{X}_i$, $\sup_{x_I \in \times_{i\in I}\mathcal{X}_i} p_\pi(x) \prod_{i\in I} |x_i|^2 \leq C$ for some $C < \infty$.

### 4.2.2. Consistency

The following theorem provides a necessary and sufficient condition for the consistency of a resampling scheme.

**Theorem 3.** *Let $\mathcal{X}$ be a cubifiable set. Then, a resampling scheme $\rho$ 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 $\pi^N \overset{w}{\Rightarrow} \pi$, $\mathbb{P}$-a.s., we have*

$$\lim_{N\to\infty} \|\rho(\zeta^N)_{h_{\mathcal{X},\psi}} - \pi^N_{h_{\mathcal{X},\psi}}\|_\star = 0, \quad \mathbb{P}-a.s. \tag{8}$$

*for a $\psi \in \mathcal{D}(\mathcal{X})$ such that $\pi_\psi \in \mathcal{P}_b((0,1)^d)$.*

This result is a consequence of Theorem 9 (see Appendix A) that establishes the equivalence between the weak convergence and the convergence in the sense of star metric and shows that the Hilbert curve $H$ and its pseudo-inverse $h$ preserve these two modes of convergence.

A direct corollary of Theorem 3 is that any Hilbert-ordered resampling scheme satisfying the discrepancy condition in (10) below is consistent, and in particular the Hilbert-ordered versions of stratified and systematic resampling are consistent.

**Corollary 4.** *Let $\mathcal{X}$ be a cubifiable set. For each $N \geq 1$ and $n \in 1:N$, let $\phi_N^n : [0,1]^{\mathbb{N}} \to [0,1]$ be a measurable function and consider a resampling scheme of the form*

$$a_N^n(u, \zeta_N) = F_N^{\sigma_N^*,-}(\phi_N^n(u)) \tag{9}$$

*with $F_N^{\sigma_N^*,-}$ the inverse of the CDF $F_N^{\sigma_N^*}(n) = \sum_{m=1}^N W^{\sigma_N(m)} \mathbb{1}(m \leq n)$, $n \in 1:N$. Then, a sufficient condition for such a resampling scheme to be $\mathcal{P}_b^*(\mathcal{X})$-consistent is that*

$$\lim_{N \to +\infty} D_N^\star\big(\phi_N^{1:N}(U)\big) = 0, \quad \mathbb{P} - a.s. \tag{10}$$

*In particular, $\rho^*_{\text{strat}}$ and $\rho^*_{\text{syst}}$, which correspond respectively to $\phi_N^n(u) = (n - 1 + u_n)/N$ and $\phi_N^n(u) = (n - 1 + u_1)/N$, are $\mathcal{P}_b^*(\mathcal{X})$-consistent.*

### 4.2.3. Variance behaviour of Hilbert-ordered resampling

The main goal of this subsection is to study in detail the convergence rate of the error variance for Hilbert-ordered stratified resampling.

The next result generalizes the first part of Theorem 2 to any $d \geq 1$.

**Theorem 4.** *Let $\mathcal{X}$ be a cubifiable set, $\pi \in \mathcal{P}_b^*(\mathcal{X})$ have a strictly positive density, and let $(\zeta^N)_{N \geq 1}$ be such that $\pi^N \overset{w}{\Rightarrow} \pi$, $\mathbb{P}$-a.s., and such that,*

$$\lim_{N \to +\infty} \big( \max_{n \in 1:N} W^{n,N} \big) = 0, \quad \mathbb{P} - a.s.$$

*Then, for any $\varphi \in \mathcal{C}_b(\mathcal{X})$,*

$$\mathrm{Var}_{\zeta^N} \left[ \rho^*_{\text{strat}}(\pi^N)(\varphi) \right] = o(1/N), \quad \mathbb{P} - a.s.$$

Theorem 4 shows that under mild conditions Hilbert-ordered stratified resampling outperforms multinomial resampling asymptotically. The following result establishes its non-asymptotic behaviour under stronger assumptions on the test function $\varphi$.

**Theorem 5.** *Let $\mathcal{X}$ be a cubifiable set and $\varphi : \mathcal{X} \to \mathbb{R}$ be a measurable function such that there exist constants $C_{\varphi,\psi} < +\infty$ and $\gamma \in (0,1]$ verifying*

$$\left|\varphi \circ \psi_{\mathcal{X}}^{-1}(x) - \varphi \circ \psi_{\mathcal{X}}^{-1}(y)\right| \leq C_{\varphi,\psi_{\mathcal{X}}} \|x - y\|_2^\gamma, \quad \forall (x,y) \in (0,1)^d.$$

*Then, for any $N \geq 1$ we have*

$$\mathrm{Var}\left[\rho^*_{\text{strat}}(z)(\varphi)\right] \leq \left(2\sqrt{d+3}\right)^{2\gamma} \frac{C^2_{\varphi,\psi_{\mathcal{X}}}}{N^{1+\frac{\gamma}{d}}}, \quad \forall z \in \mathcal{Z}^N.$$

The key tool to establish this result is the generalized Koksma-Hlawka inequality of Aistleitner and Dick (2015, Theorem 1) that we already used in (6).

Note that, because of the use of the Hilbert curve in the resampling mechanism, the rate given in Theorem 5 cannot be improved by assuming differentiability on $\varphi$. This is true because the Hilbert curve is nowhere differentiable (see e.g. Zumbusch, 2003, Lemma 4.3, p.96). We also note that the rate reported in Theorem 5 for $\gamma = 1$ is in line with the one reported in He and Owen (2016), where in a similar set up a variance of size $\mathcal{O}(N^{-1-1/d})$ is found for a class of discontinuous functions having a Lipschitz component.

It should also be clear that the power $1/d$ appearing in the upper bound of Theorem 5 arises because the Hilbert curve is Hölder continuous with exponent $1/d$. This latter is 'optimal' in the sense that $1/d$ is the best possible Hölder exponent for measure-preserving mappings from $[0,1]$ onto $[0,1]^d$ (Jaffard and Nicolay, 2009, Lemma 6). For this reason it is seems hard to improve the upper bound of Theorem 5 by considering an alternative ordering of the particles.

An interesting property of Theorem 5 is that it holds for any $N \geq 1$ and requires no conditions on the weights and on the existence of a $\pi \in \mathcal{P}(\mathcal{X})$ such that $\pi^N \overset{\mathrm{W}}{\Rightarrow} \pi$. At the same time, this suggests that the rate of $N^{1+\gamma/d}$ is not optimal when a limiting distribution $\pi$ exists. Indeed, Theorem 5 does not take into account that, in the definition of $\rho^*_{\mathrm{strat}}(\pi^N)$ given in (7), the CDF $F_{\pi^N_{h_{\mathcal{X}}}}$ may converge to $F_{\pi_{h_{\mathcal{X}}}}$, the CDF of $\pi_{h_{\mathcal{X}}}$, which is potentially a 'smooth' function. This point is corrected in the next result.

**Theorem 6.** *Consider the set-up of Theorem 5, let $(\zeta^N)_{N\geq 1}$ and $\pi \in \mathcal{P}^*_b(\mathcal{X})$ be as in Theorem 4 and assume that*

$$\mathrm{Var}_{\zeta^N}\left[\frac{1}{N}\sum_{n=1}^{N} F^-_{\pi_{h_{\mathcal{X}}}}\Big(\frac{n-1+U_n}{N}\Big)\right] = o(N^{-2}), \quad \mathbb{P}-a.s. \tag{11}$$

*Then, for any measurable function $\varphi : \mathcal{X} \to \mathbb{R}$ satisfying the condition of Theorem 5, we have*

$$\mathrm{Var}_{\zeta^N}\left[\rho^*_{\mathrm{strat}}(\pi^N)(\varphi)\right] = o\big(N^{-(1+\frac{\gamma}{d})}\big), \quad \mathbb{P}-a.s. \tag{12}$$

*When there exists a constant $c > 0$ such that $c^{-1}\,\lambda_d(A) \leq \pi(A) \leq c\,\lambda_d(A)$ for all measurable set $A \subseteq \mathcal{X}$ condition (11) is verified.*

We note that the rate in (12) does not only depend on the underlying rate in (11) but also on the speed at which $\pi^N$ converges (in some sense) to $\pi$. More precisely, the rate in (12) depends on the rate at which the quantity $v_N := \|F^-_{\pi^N_{h_{\mathcal{X}}}}(u) - F^-_{\pi_{h_{\mathcal{X}}}}(u)\|_\infty$ converges to 0 as $N \to +\infty$. In particular, under the extra assumptions of the second part of the theorem, the rate in (12) becomes $\mathcal{O}\big(N^{-(1+\frac{2\gamma}{d})}\big)$ when $v_N = \mathcal{O}(1/N)$.

## 5. Implications for particle algorithms

We apply in this section our previous results to the study of particle algorithms.

## 5.1. Set-up

We consider a generic Feynman-Kac model, consisting of (a) a Markov chain, with initial distribution $\mu(\mathrm{d}x_0)$, Markov kernels $M_t : \mathcal{X} \to \mathcal{P}(\mathcal{X})$, $t \geq 1$, acting from $\mathcal{X}$ to itself; and (b) a sequence of measurable functions, $G_0 : \mathcal{X} \to \mathbb{R}_+$, $G_t : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ for $t \geq 1$. The corresponding Feynman-Kac distributions are defined as:

$$\mathbb{Q}_t(\mathrm{d}x_{0:t}) = \frac{1}{L_t}\mu(\mathrm{d}x_0)G_0(x_0)\prod_{s=1}^{t} M_t(x_{t-1}, \mathrm{d}x_t)G_s(x_{s-1}, x_s)$$

where

$$L_t = \int_{\mathcal{X}^{t+1}} \mu(\mathrm{d}x_0)G_0(x_0)\prod_{s=1}^{t} M_t(x_{t-1}, \mathrm{d}x_t)G_s(x_{s-1}, x_s),$$

assuming $L_t > 0$. In practice, we are usually interested in approximating the so-called filtering distributions, i.e. the marginal distributions $\pi_t(\mathrm{d}x_t) = \int_{x_{0:t-1}\in\mathcal{X}^t} \mathbb{Q}_t(\mathrm{d}x_{0:t})$. We also define $\ell_t = L_t/L_{t-1} = \left(\mathbb{Q}_{t-1}M_t\right)(G_t)$ and the operators, $V_0(\varphi) = \mathrm{Var}_\eta(\varphi)$, and for $t \geq 1$,

$$V_t(x_{t-1}, \varphi) = \mathrm{Var}_{M_t}(\varphi) = M_t\left(x_{t-1}, \{\varphi - M_t(\varphi)\}^2\right),$$

where $M_t(x_{t-1}, \varphi) := \int_{\mathcal{X}} \varphi(x_t)M_t(x_{t-1}, \mathrm{d}x_t)$, and $M_t(\varphi)$ is the function $x_{t-1} \to M_t(x_{t-1}, \varphi)$.

The subsequent results will rely on the following assumptions.

**(G)** Functions $G_t$ are continuous and upper bounded.

**(M)** The Markov kernels $M_t$ define a Feller process; i.e. $M_t(\varphi) \in \mathcal{C}_b(\mathcal{X})$ for all $\varphi \in \mathcal{C}_b(\mathcal{X})$.

A standard particle filter (Algorithm 2) generates at iteration $t$ a weighted sample, $(X_t^n, W_t^n)_{n=1}^N$, which approximates $\pi_t$ through the random measure: $\pi_t^N(\mathrm{d}x_t) = \sum_{n=1}^N W_t^n \delta(X_t^n)$.

**Algorithm 2** Standard PF

At time 0:

**(a)** Generate (for $n \in 1:N$) $X_0^n \sim \mu(\mathrm{d}x_0)$.

**(b)** Compute (for $n \in 1:N$) $w_0^n = G_0(X_0^n)$ and $W_0^n = w_0^n / \sum_{m=1}^N w_0^m$.

Recursively, for times $t = 1, \ldots, T$:

**(a)** Resample: for a given resampling scheme $\rho$, generate ancestor variables $A_t^{1:N}$, where $A_t^n = a_N^n(U_t, \zeta_{t-1}^N)$, $U_t \sim \mathbb{P}$, and $\zeta_{t-1}^N = (X_{t-1}^n, W_{t-1}^n)_{n=1}^N$ (as in Definition 3).

**(b)** Generate (for $n \in 1:N$) $X_t^n \sim M_t(X_{t-1}^{A_t^n}, \mathrm{d}x_t)$.

**(c)** Compute (for $n \in 1:N$) $w_t^n = G_t(X_{t-1}^{A_t^n}, X_t^n)$ and $W_t^n = w_t^n / \sum_{m=1}^N w_t^m$.

## 5.2. Consistency

We first state an almost sure weak convergence result for Algorithm 2 under the condition that $\rho$ is consistent for a suitable class of distributions (see Crisan, 2001, Theorem 2.3.2, p.23, for a proof).

**Proposition 3.** *Let $\mathcal{P}_0 \subseteq \mathcal{P}(\mathcal{X})$ and assume that the Feynman-Kac model defined by $(G_t)_{t\geq 0}$, $\mu$ and $(M_t)_{t\geq 1}$ is such that Assumptions (G) and (M) hold, and that $\pi_t \in \mathcal{P}_0$ for all $t \geq 0$. Then, for any $\mathcal{P}_0$-consistent resampling scheme $\rho : [0,1]^{\mathbb{N}} \times \mathcal{Z} \to \mathcal{P}_f(\mathcal{X})$ and $t \geq 0$, the particle approximation $\pi_t^N := \sum_{n=1}^N W_t^n \delta(X_t^n)$ of $\pi_t$ generated by Algorithm 2 is such that*

$$\pi_t^N \overset{w}{\Rightarrow} \pi_t, \quad \mathbb{P} - a.s. \tag{13}$$

As a corollary, when $\mathcal{X}$ is a cubifiable set and the assumptions of the proposition are satisfied with $\mathcal{P}_0 = \widetilde{\mathcal{P}}_b(\mathcal{X})$, this result shows that Algorithm 2 based on stratified and SSP resampling is consistent in the sense that (13) holds for any $t \geq 0$.

We recall that (13) implies that, for any $\varphi \in \mathcal{C}_b(\mathcal{X})$, $\pi_t^N(\varphi) \to \pi_t(\varphi)$, $\mathbb{P}$-a.s. When stratified resampling is used in Algorithm 2 we note that, because this resampling mechanism dominates multinomial resampling in term of variance (see Section 2.2), it also holds true that $\lim_{N\to+\infty} N\mathbb{E}\big[(\pi_t^N(\varphi) - \pi_t(\varphi))^2\big] < +\infty$ for any $\varphi \in \mathcal{C}_b(\mathcal{X})$. For unbounded measurable function $\varphi : \mathcal{X} \to \mathbb{R}$ such that $\pi_t(\varphi) < +\infty$, the results in Cappé et al. (2005, Chapter 9) imply that $\pi_t^N(\varphi) \to \pi_t(\varphi)$ in $\mathbb{P}$-probability. These facts are gathered in the next proposition.

**Proposition 4.** *Consider Algorithm 2 based on stratified resampling, assume that the assumptions of Proposition 3 hold and let $t \geq 0$. Then,*

1. *For any $\varphi \in \mathcal{C}_b(\mathcal{X})$, $\lim_{N\to+\infty} N\mathbb{E}\big[(\pi_t^N(\varphi) - \pi_t(\varphi))^2\big] < +\infty$;*
2. *For any measurable function $\varphi : \mathcal{X} \to \mathbb{R}$ such that $\pi_t(\varphi) < +\infty$, $\pi_t^N(\varphi) \to \pi_t(\varphi)$ in $\mathbb{P}$-probability.*

## 5.3. Central limit theorem

As shown in the previous section, the 'noise' introduced by the Hilbert ordered stratified resampling scheme $\rho^*_{\text{strat}}$ converges to zero faster than the usual $\mathcal{O}(N^{-1})$ Monte Carlo rate. The next result formalises the intuitive idea that, when Algorithm 2 is based on this resampling mechanism, the resampling step does not contribute to the asymptotic variance of the quantity $N^{1/2}\big\{\pi_t^N(\varphi) - \pi_t(\varphi)\big\}$. For sake of completeness, Theorem 7 also presents results for the multinomial resampling ($\rho_{\text{multi}}$) and residual reampling ($\rho_{\text{res/multi}}$) schemes for which a central limit theorem also exists (see Chopin, 2004; Künsch, 2005; Douc et al., 2005).

**Theorem 7.** *For Algorithm 2, assuming that $\mathcal{X}$ is a cubifiable set, $\mathcal{P}_0 = \mathcal{P}_b^*(\mathcal{X})$, $\rho \in \{\rho_{\text{multi}}, \rho_{\text{res/multi}}, \rho^\star_{\text{strat}}\}$ and that the Feynman-Kac model fulfils assumptions (G) and*

*(M), for any test function $\varphi \in \mathcal{C}_b(\mathcal{X})$, we have that (for any $t \geq 0$)*

$$N^{1/2}\left\{\sum_{n=1}^{N} W_t^n \varphi(X_t^n) - \pi_t(\varphi)\right\} \stackrel{w}{\Rightarrow} \mathcal{N}_d\left(0, \mathcal{V}_t\left[\varphi\right]\right)$$

*where the $\mathcal{V}_t(\varphi)$ are defined recursively as follows: $\widetilde{\mathcal{V}}_0\left[\varphi\right] = V_0(\varphi)$,*

$$\begin{aligned}
\mathcal{V}_t\left[\varphi\right] &= \frac{1}{\ell_t^2}\widetilde{\mathcal{V}}_t\left[G_t\left\{\varphi - \pi_t(\varphi)\right\}\right] \\
\widehat{\mathcal{V}}_t\left[\varphi\right] &= \mathcal{V}_t\left[\varphi\right] + R_t\left(\rho, \varphi\right) \\
\widetilde{\mathcal{V}}_{t+1}\left[\varphi\right] &= \widehat{\mathcal{V}}_t\left[M_{t+1}(\varphi)\right] + \pi_t\left[V_{t+1}(\varphi)\right]
\end{aligned}$$

*and*

$$0 = R_t(\rho_{\text{strat}^\star}, \varphi) \leq R_t(\rho_{\text{res/multi}}, \varphi) \leq R_t(\rho_{\text{multi}}, \varphi).$$

The proof is a simple combination of Theorem 4 and the proofs in the aforementioned papers (see the supplement).

An obvious corrolary of this theorem is that ordered stratified resampling dominates multinomial and residual resampling, in terms of the asymptotic variance of particle estimates generated by a particle filter. In fact, since the contribution of the resampling step is zero when ordered stratified resampling is used, this particular scheme may be declared as optimal (again, relative to the asymptotic variance for any test function).

## 5.4. A note on the auxiliary particle filter

The auxiliary particle filter (APF, Pitt and Shephard, 1999) is a variation on the standard particle filter, where the resampling weights are 'twisted' using some function $\eta_t : \mathcal{X} \to \mathbb{R}_{>0}$; that is, the resampling weight of ancestor $X_{t-1}^m$ is $\widetilde{W}_t^m \propto W_{t-1}^m \times \eta_{t-1}(X_{t-1}^m)$; $\sum_{n=1}^N \widetilde{W}_{t-1}^n = 1$. When a particle $X_t^n$ originates from ancestor $X_{t-1}^m$, i.e. $A_t^n = m$, it is assigned (un-normalised) weight $w_t^n = G_t(X_{t-1}^m, X_t^n) W_{t-1}^m / \widetilde{W}_{t-1}^m$, so as to correct for the discrepancy between the resampling weights and the actual weights.

Of particular interest is particle estimate

$$\ell_t^N = \frac{1}{N}\sum_{n=1}^{N} w_t^n = \frac{1}{N}\sum_{n=1}^{N} \frac{W_{t-1}^{A_t^n}}{\widetilde{W}_{t-1}^{A_t^n}} G_t(X_{t-1}^{A_t^n}, X_t^n)$$

of normalising constant $\ell_t$, and the cumulative product $L_t^N = \prod_{s=0}^t \ell_t^N$, which estimates $L_t = \prod_{s=0}^t \ell_t$. The latter quantity usually corresponds to the likelihood of the data observed up to time $t$ (for a certain model) and thus plays a central role in parameter estimation methods (e.g. particle Markov chain Monte Carlo, Andrieu et al., 2010).

**Theorem 8.** *Consider the APF Algorithm (as described above), a given FK model such that Assumptions (G), and (M) hold, and assume that functions $\eta_0$, …, $\eta_{t-2}$ are fixed. For $\rho = \rho_{\text{multi}}$, the function $\eta_{t-1}(x_{t-1}) = \sqrt{M_t(x_{t-1}, G_t^2)}$ minimises the variance of particle estimates $\ell_t^N$ and $L_t^N$.*

*For $\rho = \rho^\star_{\text{strat}}$, assuming in addition that $\mathcal{X}$ is compact, the quantities $N\text{Var}[\ell_t^N]$, $N\text{Var}[L_t^N]$ converge to a limit, which is minimal for $\eta_{t-1} = \eta^\star_{t-1}$, where $\eta^\star_{t-1}(x_{t-1}) = \sqrt{V_t(x_{t-1}, G_t)}$, among functions $\eta_{t-1} \in \mathcal{C}_b(\mathcal{X})$ that are positive almost everywhere. (In particular, $\eta^\star_{t-1}$ itself is assumed to be positive everywhere.)*

The usual recommendation (e.g. Johansen and Doucet, 2008) is to take $\eta_{t-1}(x_{t-1}) = M_t(x_{t-1}, G_t)$ (or some approximation of this quantity). This result shows however that the choice of the auxiliary function in the APF should actually depend on the resampling scheme. This point deserves further study, which we leave for future research.

## 5.5. Numerical experiments

We compare in this section the approximation $(\pi_t^N)_{t=0}^T$ of $(\pi_t)_{t=0}^T$ generated by Algorithm 2 under the resampling schemes $\rho_{\text{strat}}$ (stratified resampling), $\rho^*_{\text{strat}}$ (ordered stratified resampling) and $\rho_{\text{ssp}}$ (SSP resampling).

Following Guarniero et al. (2016), we consider the linear Gaussian state-space models where $X_0 \sim \mathcal{N}_d(0, I_d)$, and, for $t = 1, \ldots, T$,

$$\begin{aligned} X_t &= FX_{t-1} + V_t, & V_t &\sim \mathcal{N}_d(0, I_d), \\ Y_t &= X_t + W_t, & W_t &\sim \mathcal{N}_d(0, I_d), \end{aligned}$$

with $F = (\alpha^{|i-j|+1})_{i,j=1}^d$, $\alpha = 0.4$, $T = 500$ and $d = 5$. We focus on the problem of estimating the log-likelihood of the model, $\log p(y_{1:T})$, which is estimated from the output of Algorithm 2 by $\log L_T^N = \sum_{t=0}^T \log \ell_t^N$ (see Section 5.4).

We consider two Feynman-Kac models; a 'bootstrap' model, where $M_t(x_{t-1}, \mathrm{d}x_t)$ corresponds to the law of $X_t | X_{t-1} = x_{t-1}$, $G_t(x_{t-1}, x_t)$ is the probability density of $Y_t | X_t = x_t$; and a 'guided' model, where $M_t(x_{t-1}, \mathrm{d}x_t)$ is the Gaussian distribution $\mathcal{N}_d\left((y_t + Fx_{t-1})/2, I_d/2\right)$, $G_t(x_{t-1}, x_t)$ is the probability density of $\mathcal{N}_d(Fx_{t-1}, 2I_d)$ at point $y_t$. Both Feynman-Kac formalisms are such that $\pi_t$ is the filtering distribution at time $t$ of the model above. The point of the guided formalism is to reduce the variance of the weights (at each time $t$), and thus to reduce the variance of particle estimates.

Figure 2 shows the variance of the estimator $\log L_T^N$ obtained under the two above Feynman-Kac formalisms, as a function of $t \in 1 : T$, and for the resampling schemes $\rho_{\text{strat}}$, $\rho^*_{\text{strat}}$ and $\rho_{\text{ssp}}$. For each resampling scheme, the results of Figure 2 are based on 1 000 independent runs of Algorithm 2 with $N = 2^{13}$ particles.

As expected from the results of Section 4, the variance of $\log L_t^N$ is smaller with $\rho^*_{\text{strat}}$ than with $\rho_{\text{strat}}$; the relative gains are larger when the guided formalism is used (where the variances under $\rho_{\text{strat}}$ are about 40% higher than under $\rho^*_{\text{strat}}$). The results presented in Figure 2 suggest that $\rho_{\text{ssp}}$ is preferable to $\rho_{\text{strat}}$. This is particularly true with the guided formalism where the variances under $\rho_{\text{strat}}$ are about 20% higher than when $\rho_{\text{ssp}}$ is used. Lastly, the variances under SSP resampling are larger than under ordered stratified resampling but $\rho_{\text{ssp}}$ has the advantage to be faster. Indeed, SSP resampling requires $\mathcal{O}(N)$ operations against $\mathcal{O}(N \log N)$ for $\rho^*_{\text{strat}}$.

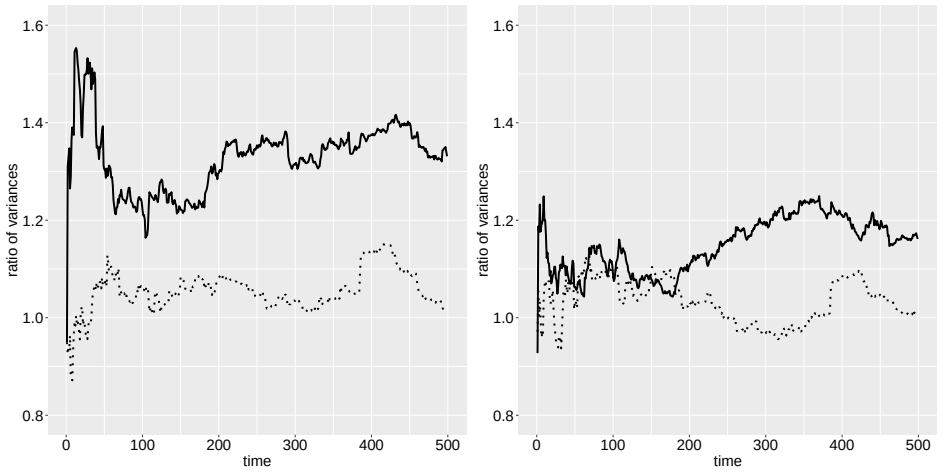


Figure 2: Estimation of the log-likelihood function as a function of $t$. The left (resp. right) plot gives the variance of SMC based on unordered stratified resampling divided by that of SMC based on Hilbert-ordered stratified resampling (resp. unordered SSP resampling). Continuous lines are for SMC based on the guided proposal while the dotted line is for the bootstrap particle filter. Results are based on 1 000 independent runs of the algorithms with $N = 2^{13}$ particles.

## 6. Conclusion

Resampling algorithms play a key role in various statistical methods, and notably in sequential Monte Carlo (SMC) since without the resampling operations these algorithms collapse exponentially fast as the number of iterations increases.

Multinomial resampling simply amounts to generate an i.i.d. sample from a discrete distribution and, for that reason, is arguably the best understood resampling mechanism. However, it is a common practice in the SMC literature to use other resampling mechanisms, such as stratified or systematic resampling, in order to reduce the computational cost and/or to reduce the approximation error. These "advanced" resampling schemes turn out to be much harder to analyse and understand, notably because their outputs depend on the order of the input particles, and thus lack of theoretical support. In this work we have provided a general consistency theorem for unbiased resampling schemes. In addition to a mild technical condition, this result shows that a sufficient condition for a resampling algorithm to be consistent is that it produces offspring numbers that are negatively associated.

Our results support the practice to abandon multinomial resampling for stratified resampling by providing strong theoretical guarantees for this resampling scheme, which has the remarkable property to be both cheaper and more accurate than multinomial resampling. For the same reasons, our results should encourage practitioners to abandon residual resampling for a version of this residual method where the multinomial resampling step is replaced by a stratified resampling step.

The systematic resampling scheme fails to produce offspring numbers that are negatively associated. As an alternative to it we have introduced the SSP resampling algorithm which (1) is similar to systematic resampling in term of offspring numbers and (2) verifies the conditions of our general consistency result. We also built an example suggesting that any general consistency results for systematic resampling would require to take into account the order of the input particles and have established its validity when they are ordered along the Hilbert curve.

Our practical recommendation is to prefer SSP resampling to systematic resampling since both have similar properties while only the former has been proven to be consistent. Systematic resampling has the advantage to be faster than SSP resampling but in most cases this gain is likely to be imperceptible. Our simulation study suggests that SSP resampling outperforms also stratified resampling in term of variance but no theoretical result exists to support this observation.

We have also derived various results showing that the variance of stratified resampling goes to zero faster than $N^{-1}$ when applied on an input point set ordered along the Hilbert curve, and notably a non-asymptotic bound of size $N^{-1-\frac{1}{d}}$. Unsurprisingly, when the dimension of the state-space is small and/or when a good proposal distribution is available, our simulation results show that ordering the particle before applying stratified resampling may lead to important variance reduction. These theoretical results on the variance of Hilbert ordered stratified resamplig are also of particular interest for sequential quasi-Monte Carlo (Gerber and Chopin, 2015), a quasi-Monte Carlo version of SMC, that converges at a faster but currently unknown rate.

## Acknowledgements

We thank Anthony Lee for useful comments. Nicolas Chopin is partly supported by Labex Ecodec (anr-11-labx-0047).

# A. Convergent sequences of probability measures: star norm and transformations through the Hilbert curve and its inverse

The following theorem is the main tool for establishing Theorem 3.

**Theorem 9.** *Let $\mathcal{X}$ be a cubifiable set, $(\pi^N)_{N\geq 1}$ be a sequence in $\mathcal{P}(\mathcal{X})$, $\pi \in \widetilde{\mathcal{P}}_b(\mathcal{X})$ and $\psi \in \mathcal{D}(\mathcal{X})$ be such that $\pi_\psi \in \mathcal{P}_b((0,1)^d)$. Then, the following assertions are equivalent*

*(i)* $\pi^N \overset{w}{\Rightarrow} \pi$;

*(ii)* $\lim_{N\to+\infty} \|\pi^N - \pi\|_\star = 0$;

*(iii)* $\lim_{N\to+\infty} \|\pi^N_{h_{\mathcal{X},\psi}} - \pi_{h_{\mathcal{X},\psi}}\|_\star = 0$;

*(iv)* $\pi^N_{h_{\mathcal{X},\psi}} \overset{w}{\Rightarrow} \pi_{h_{\mathcal{X},\psi}}$.

Implications $(ii) \Rightarrow (iii)$ and $(iii) \Rightarrow (ii)$ respectively are due to Gerber and Chopin (2015, Theorem 3) and Schretter et al. (2016, Theorem 1). Implications $(ii) \Rightarrow (i)$ and $(iii) \Rightarrow (iv)$ are direct applications of the Portmanteau lemma (e.g. van der Vaart, 1998, Lemma 2.2, p.6). Implication $(i) \Rightarrow (ii)$ for $d = 1$ holds by Polyà's theorem (Pólya, 1920; see also Bickel and Millar, 1992, result (A.1)); note that Polyà's theorem only requires that $\pi \in \mathcal{P}(\mathcal{X})$ is such that $F_\pi$ is continuous. Implication $(i) \Rightarrow (ii)$ for $d > 1$ is new and proved following a similar argument as in Kuipers and Niederreiter (1974, Theorem 1.2, p.89) while implication $(iv) \Rightarrow (iii)$ is a consequence of Polyà's Theorem and of the continuity of $F_{\pi_{h_{\mathcal{X},\psi}}}$, which is established in the next lemma.

**Lemma 2.** *Let $\mathcal{X}$ be a cubifiable set, $\pi \in \widetilde{\mathcal{P}}_b(\mathcal{X})$ and $\psi \in \mathcal{D}(\mathcal{X})$ be such that $\pi_\psi \in \mathcal{P}_b((0,1)^d)$. Then, $\pi_{h_{\mathcal{X},\psi}}$ is a continuous probability measure on $(0,1)$.*

We also note the proofs of implications $(ii) \Rightarrow (iii)$ and $(iii) \Rightarrow (ii)$ in Gerber and Chopin (2015, 2017); Schretter et al. (2016) implicitly assume that the sequence $(\pi^N)_{N\geq 1}$ is such that (with $\mathcal{X} = (0,1)^d$)

$$\pi^N(\mathcal{H}_d) = 0, \quad \text{for all } N \text{ large enough}$$

where $\mathcal{H}_d$ is the set of points of $[0,1]^d$ that have more than pre-image through $H$. This point is corrected in the supplementary materials where a complete proof of Theorem 9 is provided.

## References


Aistleitner, C. and Dick, J. (2015). Functions of bounded variation, signed measures, and a general Koksma-Hlawja inequality. *Acta Arith.*, 167(2):143–171.

Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 72(3):269–342.

Barbe, P. and Bertail, P. (1995). *The weighted bootstrap*, volume 98 of *Lecture Notes in Statistics*. Springer-Verlag, New York.

Bickel, P. and Millar, P. (1992). Uniform convergence of probability measures on classes of functions. *Statistica Sinica*, pages 1–15.

Cappé, O., Moulines, E., and Rydén, T. (2005). *Inference in hidden Markov models*. Springer Series in Statistics. Springer, New York.

Carpenter, J., Clifford, P., and Fearnhead, P. (1999). Improved particle filter for nonlinear problems. *IEE Proc. Radar, Sonar Navigation*, 146(1):2–7.

Chopin, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.*, 32(6):2385–2411.

Crisan, D. (2001). Particles filters – a theoretical perspective. In Doucet, A., de Freitas, N., and Gordon, N. J., editors, *Sequential Monte Carlo methods in practice*, Stat. Eng. Inf. Sci., pages 17–41. Springer, New York.

Crisan, D. and Doucet, A. (2002). A survey of convergence results on particle filtering methods for practitioners. *IEEE Trans. Signal Process.*, 50(3):736–746.

Douc, R., Cappé, O., and Moulines, E. (2005). Comparison of resampling schemes for particle filtering. In *ISPA 2005. Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis*, pages 64–69. IEEE.

Doucet, A., de Freitas, N., and Gordon, N. J. (2001). *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, New York.

Dubhashi, D., Jonasson, J., and Ranjan, D. (2007). Positive influence and negative dependence. *Combinatorics, Probability and Computing*, 16(01):29–41.

Dubhashi, D. and Ranjan, D. (1998). Balls and bins: a study in negative dependence. *Random Structures Algorithms*, 13(2):99–124.

Gerber, M. and Chopin, N. (2015). Sequential quasi Monte Carlo. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 77(3):509–579.

Gerber, M. and Chopin, N. (2017). Convergence of sequential quasi-Monte Carlo smoothing algorithms. *Bernoulli*, 23(4B):2951–2987.

Guarniero, P., Johansen, A. M., and Lee, A. (2016). The iterated auxiliary particle filter. *Journal of the American Statistical Association*, (just-accepted).

He, Z. and Owen, A. B. (2016). Extensible grids: uniform sampling on a space filling curve. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 78(4):917–931.

Jaffard, S. and Nicolay, S. (2009). Pointwise smoothness of space-filling functions. *Applied and Computational Harmonic Analysis*, 26(2):181–199.

Joag-Dev, K. and Proschan, F. (1983). Negative association of random variables with applications. *The Annals of Statistics*, 11(1):286–295.

Johansen, A. M. and Doucet, A. (2008). A note on auxiliary particle filters. *Statist. Probab. Lett.*, 78(12):1498–1504.

Kitagawa, G. (1996). Monte Carlo filter and smoother for non-gaussian state space models. *Journal of Computational and Graphical Statistics*, 5(1):1–25.

Kramer, J. B., Cutler, J., and Radcliffe, A. (2011). Negative dependence and Srinivasan's sampling process. *Combinatorics, Probability and Computing*, 20(03):347–361.

Kuipers, L. and Niederreiter, H. (1974). *Uniform distribution of sequences*. Wiley-Interscience.

Künsch, H. R. (2005). Recursive Monte Carlo filters: algorithms and theoretical analysis. *Ann. Statist.*, 33(5):1983–2021.

L'Ecuyer, P. and Lemieux, C. (2000). Variance reduction via lattice rules. *Management Science*, 46(9):1214–1235.

Liu, J. S. and Chen, R. (1998). Sequential Monte Carlo methods for dynamic systems. *J. Amer. Statist. Assoc.*, 93(443):1032–1044.

Pitt, M. K. and Shephard, N. (1999). Filtering via simulation: auxiliary particle filters. *J. Amer. Statist. Assoc.*, 94(446):590–599.

Pólya, G. (1920). Über den zentralen grenzwertsatz der wahrscheinlichkeitsrechnung und das momentenproblem. *Mathematische Zeitschrift*, 8(3):171–181.

Schretter, C., He, Z., Gerber, M., Chopin, N., and Niederreiter, H. (2016). Van der Corput and golden ratio sequences along the Hilbert space-filling curve. In *Monte Carlo and Quasi-Monte Carlo Methods*, pages 531–544. Springer.

Srinivasan, A. (2001). Distributions on level-sets with applications to approximation algorithms. In *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*, pages 588–597. IEEE.

Stein, M. (1987). Large sample properties of simulations using Latin hypercube sampling. *Technometrics*, 29(2):143–151.

Tillé, Y. (2006). *Sampling algorithms*. Springer Series in Statistics. Springer, New York.

van der Vaart, A. W. (1998). *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

Whitley, D. (1994). A genetic algorithm tutorial. *Statistics and Computing*, 4(2):65–85.

Zumbusch, G. (2003). *Parallel multilevel methods*. Springer.

# Supplementary materials for "Negative association, ordering, and convergence of resampling methods"

Mathieu Gerber* Nicolas Chopin† Nick Whiteley*

## S1 Preliminaries

### S1.1 Additional notation and convention

For a signed measure $\pi$ with respect to $(\mathcal{X}, \mathbb{X})$, we define the extreme norm as

$$\|\pi\|_{\mathrm{E}} = \sup_{[a,b]} \left|\pi([a,b] \cap \mathcal{X})\right|,$$

and the star norm as

$$\|\pi\|_{\star} = \sup_{(-\infty,b]} \left|\pi((-\infty,b] \cap \mathcal{X})\right|.$$

In both cases, $a$ and $b$ are vectors in $\mathbb{R}^d$, and the supremums are with respect to multivariate intervals: $[a,b] = \prod_{i=1}^d [a_i, b_i]$, $(-\infty, b] = \prod_{i=1}^d (-\infty, b_i]$.

To avoid tedious repetition of the qualification "$\mathbb{P}$-a.s.", any inequalities involving these quantities are to be understood as holding $\mathbb{P}$-a.s. unless stated otherwise.

When $\mathcal{X}$ is a cubifiable set and $\pi \in \widetilde{\mathcal{P}}_b(\mathcal{X})$, and when dealing with a sequence $(\pi^N)_{n\geq 1}$ in $\mathcal{P}(\mathcal{X})$ that converges to $\pi$ weakly or in the sense of the star/extreme metric, we can assume without loss of generality in our computations that $\mathcal{X} = (0,1)^d$ and $\pi \in \mathcal{P}_b((0,1)^d)$. This is indeed true since

$$\|\pi^N - \pi\|_{\star} = \|\pi^N_{\psi} - \pi_{\psi}\|_{\star}, \quad \|\pi^N - \pi\|_{\mathrm{E}} = \|\pi^N_{\psi} - \pi_{\psi}\|_{\mathrm{E}}, \quad \pi^N \overset{\mathrm{w}}{\Rightarrow} \pi \Leftrightarrow \pi^N_{\psi} \overset{\mathrm{w}}{\Rightarrow} \pi_{\psi}$$

while, for a suitable $\psi \in \mathcal{D}(\mathcal{X})$, $\pi_{\psi} \in \mathcal{P}_b((0,1)^d)$.

Below we abuse notations as follows: $\pi(I) = \pi(I \cap \mathcal{X})$ for $\pi \in \mathcal{P}(\mathcal{X})$ and $I$ any $d-$dimensional interval (possibly not included in $\mathcal{X}$); e.g. for $\mathcal{X} = (0,1)^d$ and $a \in \mathbb{R}^d$, $\pi([0,a]) = \pi((0,a])$.

*School of Mathematics, University of Bristol, UK.
†CREST-ENSAE, France.

## S1.2 Hilbert space filling curve: Construction and properties

We start by stating some well-known properties of the Hilbert curve (see e.g. Zumbusch, 2003, Chapter 4). The presentation below is inspired by the one in He and Owen (2016).

For $m \geq 1$, let $\mathcal{I}_m^d = \{I_m^d(k)\}_{k=0}^{2^{md}-1}$, with $I_m^d(k) = [k2^{-md}, (k+1)2^{-md}]$, and $\mathcal{S}_m^d = \{S_m^d(k)\}_{k=0}^{2^{md}-1}$ be a collection of closed hyper-cubes of volume $2^{-md}$ that cover $[0,1]^d$. Then, one can define a sequence of mappings $H_m : \mathcal{I}_m^d \to \mathcal{S}_m^d$ such that:

1. $H_m$ is bijective; that is, $H_m(I_m^d(k)) \neq H_m(I_m^d(k'))$ for any $k \neq k'$;
2. The hyper-cubes $H_m(I_m^d(k))$ and $H_m(I_m^d(k+1))$ have one $(d-1)$-dimensional face in common (adjacency property);
3. If we split $I_m^d(k)$ into $2^d$ adjacent intervals $\{I_{m+1}^d(k_i)\}_{i=0}^{2^d-1}$ of length $2^{-(m+1)d}$, then $\cup_{i=0}^{2^d-1} H_{m+1}(I_{m+1}^d(k_i)) = H_m(I_m^d(k))$ (nesting property).

Then, the Hilbert curve is defined as follows. Let $x \in [0,1]$ and note that there exists a sequence $(I_m^d(k_m^x))_{m\geq 1}$ such that (i) $I_{m+1}^d(k_{m+1}^x) \subset I_m^d(k_m^x)$ and (ii) $\{x\} = \cap_{m\geq 1} I_m^d(k_m^x)$. Using the nesting property of $(H_m)_{m\geq 1}$, the set $\cap_{m\geq 1} H_m(I_m^d(k_m^x))$ contains a single point in $[0,1]^d$ and the Hilbert curve $H : [0,1] \to [0,1]^d$ is defined by

$$H(x) \in \cap_{m\geq 1} H_m(I_m^d(k_m^x)), \quad x \in [0,1]^d.$$

Note that Conditions 1-3 listed above do not uniquely define the sequence $(H_m)_{m\geq 1}$ and therefore, although we refer to $H$ as *the* Hilbert curve in this work, there exist in fact several Hilbert curves.

Function $H$ is not bijective. Indeed, if $x \in [0,1]^d$ has at least one dyadic coordinate, then for $m$ large enough there exist at least two distinct indices $k_m$ and $k'_m$ in $0 : 2^{dm} - 1$ such that $x \in S_m^d(k_m) \cap S_m^d(k'_m)$; recall that sets in $\mathcal{S}_m^d$ are closed. Since for $m$ large enough $x \in [0,1]^d$ belongs to more than one set in $\mathcal{S}_m^d$ if and only if $x$ has at least one dyadic coordinate, the set $\mathcal{H}_d \subset [0,1]^d$ of points in $[0,1]^d$ that have more than one pre-image through $H$ is such that $\#H^{-1}(x) \leq 2^d$ for all $x \in [0,1]^d$ and such that $\lambda_d(\mathcal{H}_d) = 0$. Lastly, it is easily checked that function $H$ is such that $\lambda_1(A) = \lambda_d(H(A))$ for any measurable set $A \subset [0,1]$ (bi-measure property) and that $\|H(x_1) - H(x_2)\|_\infty \leq C_d |x_1 - x_2|^{1/d}$ for some constant $C_d < +\infty$ and any $x_1, x_2 \in [0,1]$ (i.e. $H$ is Hölder with coefficient $1/d$).

We assume from henceforth that $H$ is such that $H(0) = (0, \ldots, 0)$ and, to simplify the notation, we use the convention that, for any $m \geq 1$, the sets in $\mathcal{S}_m^d$ are labelled so that

$$H(I_m^d(k)) = H_m(I_m^d(k)) = S_m^d(k), \quad \forall k \in 0 : 2^{dm} - 1, \ \forall m \geq 1. \tag{S.1}$$

Thanks to the above properties of the Hilbert curve, there exists a one-to-one Borel measurable function $h : [0,1]^d \to [0,1]$ such that $H(h(x)) = x$ for all $x \in [0,1]^d$, as shown in Proposition 2. For a cubifiable set $\mathcal{X}$ and a $\psi \in \mathcal{D}(\mathcal{X})$ we denote by $h_{\mathcal{X},\psi}$ the mapping $x \mapsto h \circ \psi(x)$. Note that, by construction $h_{\mathcal{X},\psi} : \mathcal{X} \to (0,1)$ is one-to-one and Borel measurable.

## S1.3 Some preliminary results

The following lemma is a direct extension of Niederreiter (1992, Lemma 2.5, p.15).

**Lemma S1.** *Let $\mathcal{X}$ be a cubifiable set, $\epsilon > 0$ and $z, \tilde{z} \in \mathcal{Z}^N$ be such that*

$$z = (x^n, W^n)_{n=1}^N, \quad \tilde{z} = (\tilde{x}^n, W^n)_{n=1}^N, \quad \max_{n \in 1:N} \|x^n - \tilde{x}^n\|_\infty \leq \epsilon.$$

*Let $\pi^N = \sum_{n=1}^N W^n \delta(x^n)$ and $\tilde{\pi}^N = \sum_{n=1}^N W^n \delta(\tilde{x}^n)$. Then, for any $\pi \in \widetilde{\mathcal{P}}_b(\mathcal{X})$, there exists a constant $c_\pi < +\infty$ (which depends only on $\pi$) such that*

$$\left| \|\tilde{\pi}^N - \pi\|_\star - \|\pi^N - \pi\|_\star \right| \leq c_\pi \, \epsilon$$

*and*

$$\left| \|\tilde{\pi}^N_{h_{\mathcal{X},\psi}} - \pi_{h_{\mathcal{X},\psi}}\|_\star - \|\pi^N_{h_{\mathcal{X},\psi}} - \pi_{h_{\mathcal{X},\psi}}\|_\star \right| \leq c_\pi \max_{n \in 1:N} \|h_{\mathcal{X},\psi}(x^n_{\pi^N}) - h_{\mathcal{X},\psi}(\tilde{x}^n)\|_\infty$$

*with $\psi \in \mathcal{D}(\mathcal{X})$ such that $\pi_\psi \in \mathcal{P}_b((0,1)^d)$.*

*Proof of Lemma S1.* Without loss of generality we assume that $\mathcal{X} = (0,1)^d$ and take $h_{\mathcal{X},\psi} = h$.

Let $B = [0, b] \in [0,1)^d$, $B^+ = [0, b + \epsilon] \cap [0,1)^d$ and $B^- = [0, b - \epsilon]$. If $\epsilon > b_i$ for at least one $i \in 1:d$, $B^- = \varnothing$. Then,

$$\pi^N(B^-) \leq \tilde{\pi}^N(B) \leq \pi^N(B^+). \tag{S.2}$$

By the definition of the star norm, we have

$$\left|\pi^N(B^+) - \pi(B^+)\right| \leq \|\pi^N - \pi\|_\star, \quad \left|\pi^N(B^-) - \pi(B^-)\right| \leq \|\pi^N - \pi\|_\star. \tag{S.3}$$

Combining (S.2) and (S.3) yields:

$$\begin{cases} -\big(\pi(B) - \pi(B^-)\big) - \|\pi^N - \pi\|_\star \leq \tilde{\pi}^N(B) - \pi(B) \\ \tilde{\pi}^N(B) - \pi(B) \leq (\pi(B^+) - \pi(B)) + \|\pi^N - \pi\|_\star. \end{cases} \tag{S.4}$$

Then, as $\pi$ admits a bounded density $p_\pi$, we have,

$$\begin{aligned} \pi(B) - \pi(B^-) &\leq \|p_\pi\|_\infty \lambda_d(B \setminus B^-) \leq \|p_\pi\|_\infty d\epsilon, \\ \pi(B^+) - \pi(B) &\leq \|p_\pi\|_\infty \lambda_d(B^+ \setminus B) \leq \|p_\pi\|_\infty d\epsilon. \end{aligned} \tag{S.5}$$

Therefore, combining (S.4) and (S.5), we obtain,

$$-\|p_\pi\|_\infty \, d\epsilon - \|\pi^N - \pi\|_\star \leq \tilde{\pi}^N(B) - \pi(B) \leq \|\pi^N - \pi\|_\star + \|p_\pi\|_\infty \, d\epsilon$$

and thus

$$\|\tilde{\pi}^N - \pi\|_\star \leq \|\pi^N - \pi\|_\star + \|p_\pi\|_\infty \, d\epsilon.$$

To complete the proof of the first part of the lemma it suffices to repeat the above computations while swapping the role of $\pi^N$ and $\tilde{\pi}^N$.

The second part of the lemma follows from similar computations, where $\epsilon$ is replaced by

$$\epsilon' := \max_{n\in 1:N} \|h(x^n) - h(\tilde{x}^n)\|_\infty$$

and where instead of (S.5) we have, by the definition of $\pi_h$ and by the bi-measure property of the Hilbert curve (see Section S1.2),

$$\begin{aligned}\pi_h(B) - \pi_h(B^-) &= \pi\big(H(B \setminus B^-)\big) \leq \|p_\pi\|_\infty \lambda_1(B \setminus B^-) \leq \|p_\pi\|_\infty \epsilon' \\ \pi(B^+) - \pi(B) &= \pi\big(H(B^+ \setminus)\big) \leq \|p_\pi\|_\infty \lambda_1(B^+ \setminus B) \leq \|p_\pi\|_\infty \epsilon'.\end{aligned}$$

□

The next lemma follows from the computations of Gerber and Chopin (2015, Theorem 3).

**Lemma S2.** *Let $\mu \in \mathcal{P}((0,1)^d)$ and $\pi \in \mathcal{P}_b((0,1)^d)$. Then,*

$$\|\mu - \pi\|_\star = \sup_{0\leq a\leq 1} \Big|\mu([0,a)) - \pi([0,a))\Big|, \ \|\mu - \pi\|_{\mathrm{E}} = \sup_{0\leq a<b\leq 1} \Big|\mu([a,b)) - \pi([a,b))\Big|$$

*and*

$$\|\mu_h - \pi_h\|_\star = \sup_{0\leq a\leq 1} \Big|\mu_h([0,a)) - \pi_h([0,a))\Big|, \ \|\mu_h - \pi_h\|_{\mathrm{E}} = \sup_{0\leq a<b\leq 1} \Big|\mu_h([a,b)) - \pi_h([a,b))\Big|.$$

*Proof of Lemma S2.* Below we only prove the second equality, the other ones being proved in a similar way.

Let $p$ be the density of $\pi$ with respect to $\lambda_d$. Let $\epsilon > 0$, $a \in [0,1]$, and $\delta_{a,\epsilon} \in [0, \epsilon/\|p\|_\infty]$, be small enough so that $\mu_h([0,a]) = \mu_h([0, a+\delta_{a,\epsilon}))$ and $a + \delta_{a,\epsilon} \leq 1$. (If $a = 1$ or $\mu$ is continuous, take $\delta_{a,\epsilon} = 0$.) Then,

$$|\mu_h([0,a)) - \pi_h([0,a))| \leq |\mu_h([0, a+\delta_{a,\epsilon})) - \pi_h([0, a+\delta_{a,\epsilon}))| + \pi_h([a, a+\delta_{a,\epsilon})) \quad \text{(S.6)}$$

and

$$|\mu_h([0,a]) - \pi_h([0,a])| \geq |\mu_h([0, a+\delta_{a,\epsilon})) - \pi_h([0, a+\delta_{a,\epsilon}))| - \pi_h([a, a+\delta_{a,\epsilon})). \quad \text{(S.7)}$$

By the bi-measure property of the Hilbert curve (see Section S1.2), the set $H((a, a+\delta_{a,\epsilon}))$ has Lebesgue measure $\delta_{a,\epsilon}$. Thus,

$$\pi_h([a, a+\delta_{a,\epsilon})) = \pi\big(H([a, a+\delta_{a,\epsilon}))\big) \leq \|p\|_\infty \delta_{a,\epsilon} \leq \epsilon.$$

Replacing $\pi_h((a, a+\delta_{a,\epsilon}))$ by $\epsilon$ in (S.6) and (S.7), and taking the supremum other $a$ yields

$$-\epsilon \leq \|\mu_h - \pi_h\|_\star - \sup_{0\leq a\leq 1} |\mu_h([0,a)) - \pi_h([0,a))| \leq \epsilon$$

implying that

$$\|\mu_h - \pi_h\|_\star = \sup_{0\leq a\leq 1} |\mu_h([0,a)) - \pi_h([0,a))|.$$

□

The next result notably follows from the computations of Gerber and Chopin (2015, Theorem 7).

**Lemma S3.** *Let $\mathcal{X}$ be a cubifiable set, $\pi \in \widetilde{\mathcal{P}}_b(\mathcal{X})$ be such that $\pi(x) = p(x)\lambda_d(\mathrm{d}x)$ for a strictly positive density $p : \mathcal{X} \to \mathbb{R}_+^*$ and $\psi \in \mathcal{D}(\mathcal{X})$ be such that $\pi_\psi \in \mathcal{P}_b((0,1)^d)$. Then,*

1. *$F_{\pi_{h_{\mathcal{X},\psi}}} : (0,1) \to [0,1]$ is strictly increasing.*
2. *For any sequence $(z^N)_{N\geq 1}$ such that, for all $N \geq 1$, $z^N \in \mathcal{Z}^N$ and such that*

$$\pi^N \overset{w}{\Rightarrow} \pi, \quad \lim_{N\to+\infty} \big( \max_{n\in 1:N} W^{n,N} \big) = 0,$$

   *we have $\lim_{n\to+\infty} \|F^-_{\pi^N_{h_{\mathcal{X},\psi}}} - F^-_{\pi_{h_{\mathcal{X},\psi}}}\|_\infty = 0$.*
3. *For any sequence $(F_N)_{N\geq 1}$ of continuous CDF such that $\lim_{N\to+\infty} \|F_N - F_{\pi_{h_{\mathcal{X},\psi}}}\|_\infty = 0$ we have $\lim_{n\to+\infty} \|F^-_N - F^-_{\pi_{h_{\mathcal{X},\psi}}}\|_\infty = 0$.*

*Proof of Lemma S3.* Without loss of generality we assume that $\mathcal{X} = (0,1)^d$ and take $h_{\mathcal{X},\psi} = h$. Let $(z^N)_{N\geq 1}$ and $\pi$ be as in the statement of the theorem.

We first show the first part of the lemma. Let $0 \leq a < b \leq 1$, and let $k$, $m$ be integers such that $I^d_m(k) \subset [a,b]$ where $I^d_m(k) = [k2^{-md}, (k+1)2^{-md}]$ (as defined in Section S1.2). Then

$$F_{\pi_h}(b) - F_{\pi_h}(a) = \pi_h([a,b]) \geq \pi_h(I^d_m(k)) = \pi(S^d_m(k)) > 0$$

where $S^d_m(k) = H(I^d_m(k))$ is a hyper-cube of volume $2^{-m}$, again see Section S1.2. The last inequality comes from the fact that $\pi$ admits a positive density. Thus $F_{\pi_h}$ is increasing.

To establish the second part of the lemma we first show that, for all $u \in (0,1)$.

$$\lim_{N\to+\infty} |F^-_{\pi_h}(u) - F^-_{\pi^N_h}(u)| = 0. \tag{S.8}$$

This result is derived in the computations of Gerber and Chopin (2015, Theorem 7) but for sake of completness it is proved below.

Let $\epsilon > 0$ and $u \in (0,1)$. Because $F_{\pi_h}$ is continuous (Lemma 2) and strictly increasing, $F^{-1}_{\pi_h}$ is continuous and thus there exists a $\delta_{u,\epsilon} > 0$ such that,

$$|u' - u| \leq \delta_{u,\epsilon}, \quad \Longrightarrow \ |F^-_{\pi_h}(u') - F^-_{\pi_h}(u)| \leq \epsilon. \tag{S.9}$$

By assumption, for any $\delta_0 > 0$, there exists a $N_{\delta_0}$ such that, for all $N \geq N_{\delta_0}$,

$$\|F_{\pi^N_h} - F_{\pi_h}\|_\infty \leq \delta_0. \tag{S.10}$$

Let $x_N = F^-_{\pi^N_h}(u)$ and $u_N = F_{\pi_h}(x_N)$. Then, by (S.10),

$$|F_{\pi^N_h}(x_N) - F_{\pi_h}(x_N)| \leq \delta_0, \quad \forall N \geq N_{\delta_0}.$$

Let $r_N(u) = F_{\pi_h^N}(F^-_{\pi_h^N}(u)) - u$ so that

$$|F_{\pi_h^N}(x_N) - F_{\pi_h}(x_N)| = |u + r_N(u) - u_N| \leq \delta_0, \quad \forall N \geq N_{\delta_0}$$

Now note that $|r_N(u)| \leq \max_{n\in 1:N} W^{n,N}$ and thus, by assumption, for all $\delta' > 0$, there exists a $N_{\delta'}$ such that, a.s., $|r_N(u)| \leq \delta'$ for all $N \geq N_{\delta'}$. Let $\delta = \delta_0 + \delta'$ and set $N_\delta := N_{\delta_0} \vee N_{\delta'}$. Then, for $N \geq N_\delta$, we have $|u - u_N| \leq \delta$. By taking $\delta_0$ and $\delta'$ such that $\delta = \delta_{u,\epsilon}$, (S.9) implies that

$$|F^-_{\pi_h}(u) - F^-_{\pi_h}(u_N)| \leq \epsilon, \quad \forall N \geq N_{\delta_{u,\epsilon}}.$$

In addition, $F^-_{\pi_h}(u_N) = x_N = F^-_{\pi_h^N}(u)$ and therefore

$$|F^-_{\pi_h}(u) - F^-_{\pi_h^N}(u)| \leq \epsilon, \quad \forall N \geq N_{\delta_{u,\epsilon}}.$$

This shows (S.8).

Then, to show the second part of the theorem remark that, since $F^-_{\pi_h}$ is continuous on $(0, 1)$ and such that

$$\lim_{u\to 0} F^-_{\pi_h}(u) = 0, \quad \lim_{u\to 1} F^-_{\pi_h}(u) = 1,$$

the mapping $F^-_{\pi_h}$ can be extended to a continuous function $\tilde{F}^-_{\pi_h} : [0, 1] \to [0, 1]$ (Mytrofanov and Ravsky, 2012, Lemma 2) which is thus uniformly continuous on $[0, 1]$. Consequently, for any $\epsilon > 0$ there exists a $\delta_\epsilon > 0$ such that,

$$|u' - u| \leq \delta_\epsilon, \quad \implies |\tilde{F}^-_{\pi_h}(u') - \tilde{F}^-_{\pi_h}(u)| \leq \epsilon$$

and thus, by replacing $\delta_{u,\epsilon}$ by $\delta_\epsilon$ in the above computations, it follows that

$$|\tilde{F}^-_{\pi_h}(u) - F^-_{\pi_h^N}(u)| \leq \epsilon, \quad \forall N \geq N_{\delta_\epsilon}.$$

Since $N_{\delta_\epsilon}$ is independent of $u$,

$$\|\tilde{F}^-_{\pi_h} - F^-_{\pi_h^N}\|_\infty \leq \epsilon, \quad \forall N \geq N_{\delta_\epsilon}$$

and the proof is completed upon noting that

$$\|F^-_{\pi_h} - F^-_{\pi_h^N}\|_\infty \leq \|\tilde{F}^-_{\pi_h} - F^-_{\pi_h^N}\|_\infty, \quad \forall N \geq 1.$$

The last part of the lemma is obvious from the computations carried out to show the third part.

□

## S2 Proof of Section 2: Proof of Lemma 1

*Proof of Lemma 1.* We prove the result for the case $\mathcal{X} = \mathbb{R}^d$, its extension to any cubifiable sets being trivial.

Let $\pi \in \mathcal{P}_b(\mathbb{R}^d)$ be such that there exists a constant $C_\pi < +\infty$ such that, for any $I \subseteq 1:d$ and $x_{\setminus I} \in \mathbb{R}^{d-I}$, we have $\sup_{x_I \in \mathbb{R}^{|I|}} p_\pi(x) \prod_{i \in I} |x_i|^{1+\delta} \leq C_\pi$, and let $\psi \in \mathcal{D}(\mathbb{R}^d)$. Then, using the change of variable formula,

$$p_{\pi_\psi}(u) = p_\pi(\psi^{-1}(u)) \prod_{i=1}^d \frac{\mathrm{d}\psi_i(u_i)}{\mathrm{d}u_i}, \quad u \in (0,1)^d.$$

Hence, because $p_\pi$ is continuous and $\psi$ is a $C^1$-diffeomorphism, $p_{\pi_\psi}$ is continuous on $(0,1)^d$. We now show that for a suitable choice of $\psi \in \mathcal{D}(\mathbb{R}^d)$ the density $p_{\pi_\psi}$ is bounded.

To construct $\psi$, let $\alpha > 0$ and $\tilde{\psi} : \mathbb{R} \to (0,1)$ be such that

$$\tilde{\psi}^{-1}(u) = \frac{2u-1}{u^\alpha (1-u)^\alpha}, \quad u \in (0,1).$$

It is easily checked that $\lim_{u \to 0} \psi^{-1}(u) = -\infty$ and $\lim_{u \to 1} \psi^{-1}(u) = +\infty$. In addition,

$$\frac{\mathrm{d}\psi^{-1}(u)}{\mathrm{d}u} = \frac{2}{u^\alpha (1-u)^\alpha} + \frac{\alpha(2u-1)^2}{u^{\alpha+1}(1-u)^{\alpha+1}} > 0, \quad \forall u \in (0,1).$$

and thus $\tilde{\psi} \in \mathcal{D}(\mathbb{R})$.

Let $\psi \in \mathcal{D}(\mathbb{R}^d)$ be defined by $\psi(x) = (\tilde{\psi}(x_1), \ldots, \tilde{\psi}(x_d))$, $x \in \mathbb{R}^d$. Then, for $u \in (0,1)^d$, we have

$$p_{\pi_\psi}(u) = p_\pi\left(\frac{2u_1-1}{u_1^\alpha(1-u_1)^\alpha}, \ldots, \frac{2u_d-1}{u_d^\alpha(1-u_d)^\alpha}\right) \prod_{i=1}^d \left(\frac{2}{u_i^\alpha(1-u_i)^\alpha} + \alpha \frac{(2u_i-1)^2}{u_i^{\alpha+1}(1-u_i)^{\alpha+1}}\right).$$

Let $\underline{u} = \frac{1}{2} - \frac{1}{2}\left(\sqrt{1 - \frac{4\alpha}{4\alpha+2}}\right)$ and $\bar{u} = \frac{1}{2} + \frac{1}{2}\left(\sqrt{1 - \frac{4\alpha}{4\alpha+2}}\right)$ so that

$$\frac{2}{u_1^\alpha(1-u_1)^\alpha} \leq \alpha \frac{(2u_1-1)^2}{u_1^{\alpha+1}(1-u_1)^{\alpha+1}}, \quad \forall u_1 \in (0, \underline{u}] \cup [\bar{u}, 1)$$

while

$$\alpha \frac{(2u_1-1)^2}{u_1^{\alpha+1}(1-u_1)^{\alpha+1}} \leq \frac{2}{u_1^\alpha(1-u_1)^\alpha}, \quad \forall u_1 \in [\underline{u}, \bar{u}].$$

Next, let $\mathcal{B}_\alpha = \tilde{\psi}^{-1}([\underline{u}, \bar{u}])$, and, for $u \in (0,1)^d$, let $I_u = \{i \in 1:d : u_i \notin [\underline{u}, \bar{u}]\}$ and $x_i^{(u)} = \tilde{\psi}^{-1}(u_i)$, $i \in 1:d$. Then, for any $u \in (0,1)^d$ and $\alpha \geq 1$, and with the convention

that empty products equal one,

$$
\begin{aligned}
p_{\pi_\psi}(u) &\leq (2\alpha)^d p_\pi(x_1^{(u)}, \ldots, x_d^{(u)}) \prod_{i \notin I_u} \frac{2}{u_i^\alpha (1-u_i)^\alpha} \prod_{i \in I_u} \frac{\alpha(2u_i-1)^2}{u_i^{\alpha+1}(1-u_i)^{\alpha+1}} \\
&\leq \Big( \max_{u \in [\underline{u}, \bar{u}]} \frac{4\alpha}{u^\alpha (1-u)^\alpha} \Big)^d \max_{z_{\setminus I_u} \in \mathcal{B}_\alpha^{d-|I_u|}} p_\pi\Big(x_{I_u}^{(u)}, z_{\setminus I_u}\Big) \prod_{i \in I_u} |2u_i - 1|^{\frac{\alpha-1}{\alpha}} |x_i^{(u)}|^{\frac{1+\alpha}{\alpha}} \\
&= \Big( \max_{u \in [\underline{u}, \bar{u}]} \frac{4\alpha}{u^\alpha (1-u)^\alpha} \Big)^d p_\pi\Big(x_{I_u}^{(u)}, \tilde{x}_{\setminus I}^{(u)}\Big) \prod_{i \in I_u} |2u_i - 1|^{\frac{\alpha-1}{\alpha}} |x_i^{(u)}|^{\frac{1+\alpha}{\alpha}} \\
&\leq \Big( \max_{u \in [\underline{u}, \bar{u}]} \frac{4\alpha}{u^\alpha (1-u)^\alpha} \Big)^d p_\pi\Big(x_{I_u}^{(u)}, \tilde{x}_{\setminus I_u}^{(u)}\Big) \prod_{i \in I_u} |x_i^{(u)}|^{\frac{1+\alpha}{\alpha}}.
\end{aligned}
$$

for a $\tilde{x}_{\setminus I_u}^{(u)} \in \mathcal{B}_\alpha^{d-|I|}$ and where the equality holds because $\tilde{p}_\psi$ is continuous and $B_\alpha$ is compact. Then, the result follows by noting that $\alpha \geq 1$ is arbitrary and that, as $\alpha \to +\infty$, $(1+\alpha)/\alpha \to 1$.

□

# S3 Proofs of Section 3

## S3.1 Proof of Theorem 1

Before proving the result we recall the following maximal inequality.

**Theorem S1.** *(Shao, 2000, Theorem 3) Let $(Z^n)_{n=1}^N$ be a sequence of NA random variables with zero means and finite second moments. Let $B_N = \sum_{n=1}^N \mathbb{E}[(Z^n)^2]$. Then, for all $\epsilon > 0$ and $a > 0$,*

$$
\mathbb{P}\Big( \max_{m \in 1:N} \Big| \sum_{n=1}^m Z^n \Big| \geq \epsilon \Big) \leq 2\mathbb{P}\Big( \max_{n \in 1:N} |Z^n| > a \Big) + 4 \exp\Big( -\frac{\epsilon^2}{8B_N} \Big) + 4\Big( \frac{B_N}{4(\epsilon a + B_N)} \Big)^{\frac{\epsilon}{12a}}.
$$

*Proof of Theorem 1.* Let $\rho$ be an unbiased resampling scheme that satisfies $(H_1)$ and $(H_2)$, $\pi \in \widetilde{\mathcal{P}}_b(\mathcal{X})$, $\psi \in \mathcal{D}(\mathcal{X})$ be such that $\pi_\psi \in \mathcal{P}_b((0,1)^d)$ and $(\zeta^N)_{N \geq 1}$ be such that $\pi^N \overset{\mathrm{w}}{\Rightarrow} \pi$, $\mathbb{P}$-a.s.

By Theorem 3, for the required $\widetilde{\mathcal{P}}_b(\mathcal{X})$-consistency of $\rho$, it is necessary and sufficient that

$$
\lim_{N \to +\infty} \|\rho(\zeta^N)_{h_{\mathcal{X},\psi}} - \pi^N_{h_{\mathcal{X},\psi}}\|_\star = 0. \quad \mathbb{P} - a.s.
$$

For a given $N \geq 1$ and $z = (x^n, w^n)_{n=1}^N$ let $\sigma^*(z, \cdot)$ be a permutation of $1:N$ such that $h_{\mathcal{X},\psi}(x^{\sigma^*(z,1)}) \leq \cdots \leq h_{\mathcal{X},\psi}(x^{\sigma^*(z,N)})$, i.e., $\sigma^*$ sorts the points $x^n$ using the Hilbert curve.

Notice then that using the definition of $(\Delta^n_{\rho,z})_{n=1}^N$ given in (2), we have:

$$
\|\rho(\zeta^N)_{h_{\mathcal{X},\psi}} - \pi^N_{h_{\mathcal{X},\psi}}\|_\star = \frac{1}{N} \max_{m \in 1:N} \Big| \sum_{n=1}^m \Delta^{\sigma^*(\zeta^N, n)}_{\rho, \zeta^N} \Big|,
$$

so our goal in the following is to show that

$$\lim_{N\to+\infty}\frac{1}{N}\max_{m\in 1:N}\Big|\sum_{n=1}^{m}\Delta_{\rho,\zeta^N}^{\sigma^*(\zeta^N,n)}\Big|=0,\quad \mathbb{P}-a.s. \tag{S.11}$$

Under $(H_1)$, for any $N\geq 1$ and $z\in\mathcal{Z}^N$, the random variables $\big(\#^n(U,z)\big)_{n=1}^N$ are NA, using the definition of NA random variables and (2), the random variables $(\Delta_{\rho,z}^n)_{n=1}^N$ are NA too, as are also $(\Delta_{\rho,z}^{\sigma^*(z,n)})_{n=1}^N$. In addition, because $\rho$ is assumed to be unbiased, $\mathbb{E}[\Delta_{\rho,z}^{\sigma^*(z,n)}]=0$ for all $n\in 1:N$ and thus $(\Delta_{\rho,z}^{\sigma^*(z,n)})_{n=1}^N$ satisfy the assumptions of Theorem S1.

Let $t>0$ and $z\in\mathcal{Z}^N$. Then, using Theorem S1 with $a=r_N$ and $\epsilon=tN$, we have

$$\begin{aligned}\mathbb{P}\Big(\max_{m\in 1:N}\Big|\sum_{n=1}^{m}\Delta_{\rho,z}^{\sigma^*(z,n)})\Big|\geq tN\Big)&\leq 2\mathbb{P}\Big(\max_{n\in 1:N}\big|\Delta_{\rho,z}^n\big|>r_N\Big)\\&\quad+4\exp\bigg(-\frac{(Nt)^2}{8\sum_{n=1}^N\mathbb{E}\big[(\Delta_{\rho,z}^n)^2\big]}\bigg)+4\bigg(\frac{\sum_{n=1}^N\mathbb{E}\big[(\Delta_{\rho,z}^n)^2\big]}{4tNr_N+4\sum_{n=1}^N\mathbb{E}\big[(\Delta_{\rho,z}^n)^2\big])}\bigg)^{\frac{tN}{12r_N}}.\end{aligned}$$

Under $(H_2)$, $\sum_{n=1}^N\mathbb{E}\big[(\Delta_{\rho,z}^n)^2\big]\leq r_NN$ for $N$ large enough and thus, since for any $b\geq 0$, the mapping $x\mapsto x/(b+4x)$ is non-decreasing, we have

$$\bigg(\frac{\sum_{n=1}^N\mathbb{E}\big[(\Delta_{\rho,z}^n)^2\big]}{4tNr_N+4\sum_{n=1}^N\mathbb{E}\big[(\Delta_{\rho,z}^n)^2\big])}\bigg)^{\frac{tN}{12r_N}}\leq\bigg(\frac{r_NN}{4tNr_N+4r_NN}\bigg)^{\frac{tN}{12r_N}}\leq\bigg(\frac{1}{4}\bigg)^{\frac{tN}{12r_N}}.$$

The condition $\sum_{n=1}^N\mathbb{E}\big[(\Delta_{\rho,z}^n)^2\big]\leq r_NN$ also implies that

$$\exp\bigg(-\frac{(Nt)^2}{8\sum_{n=1}^N\mathbb{E}\big[(\Delta_{\rho,z}^n)^2\big]}\bigg)\leq\exp\bigg(-\frac{Nt^2}{8r_N}\bigg)$$

and thus

$$\mathbb{P}\Big(\max_{m\in 1:N}\Big|\sum_{n=1}^{m}\Delta_{\rho,z}^{\sigma^*(z,n)})\Big|\geq N\,t\Big)\leq 2\mathbb{P}\Big(\max_{n\in 1:N}\big|\Delta_{\rho,z}^n\big|>r_N\Big)+4\exp\bigg(-\frac{Nt^2}{8r_N}\bigg)+4\bigg(\frac{1}{4}\bigg)^{\frac{tN}{12r_N}}.$$

Since $r_N=o(N/\log N)$, we have $\sum_{N=1}^\infty\beta^{N/r_N}<+\infty$ for any $\beta\in(0,1)$; take e.g. $\beta=(1/4)^{\frac{t}{12}}$ and $\beta=\exp(-t^2/8)$,

$$\sum_{N=1}^\infty\sup_{z\in\mathcal{Z}^N}\mathbb{P}\Big(\max_{n\in 1:N}\big|\Delta_{\rho,z}^n\big|>r_N\Big)<+\infty,\quad\sum_{N=1}^\infty\exp\bigg(-\frac{Nt^2}{8r_N}\bigg)<+\infty,\quad\sum_{N=1}^\infty\bigg(\frac{1}{4}\bigg)^{\frac{tN}{12r_N}}<+\infty.$$

Using the tower property and the fact that $U$ is independent of $\zeta^N$,

$$\begin{aligned}\mathbb{P}\Big(\max_{m\in 1:N}\Big|\sum_{n=1}^{m}\Delta_{\rho,\zeta^N}^{\sigma^*(\zeta^N,n)})\Big|\geq N\,t\Big)&=\mathbb{E}\bigg[\mathbb{P}_{\zeta^N}\Big(\max_{m\in 1:N}\Big|\sum_{n=1}^{m}\Delta_{\rho,\zeta^N}^{\sigma^*(\zeta^N,n)})\Big|\geq N\,t\Big)\bigg]\\&\leq\sup_{z\in\mathcal{Z}^N}\mathbb{P}\Big(\max_{m\in 1:N}\Big|\sum_{n=1}^{m}\Delta_{\rho,z}^{\sigma^*(z,n)})\Big|\geq N\,t\Big),\end{aligned}$$

so we have proved:

$$\sum_{N=1}^{\infty} \mathbb{P}\Big( \max_{m\in 1:N} \Big| \sum_{n=1}^{m} \Delta_{\rho,\zeta^N}^{\sigma^*(\zeta^N,n)}) \Big| \geq N\, t \Big) < +\infty.$$

Therefore (S.11) holds by the Borel-Cantelli lemma and the proof is complete.

□

## S3.2 Proof of Proposition 1

*Proof of Proposition 1.* Let $\mathcal{X} = (0,1)$ be a cubifiable set. To construct $\pi \in \widetilde{\mathcal{P}}_b(\mathcal{X})$, let $p = 1/2$, $p' = 3$, $k = 3$, $a_1 = 2$, $a_2 = 5$, $b_1 = a_1 2^{-k}$, $b_2 = a_2 2^{-k}$ and $\epsilon = 2^{-k}$. Then,

$$p(1 + (a_1 - a_2 - 1)2^{-k}) + p'(a_2 - a_1 - 1)2^{-k} = 1 \tag{S.12}$$

and the function $p : (0,1) \to \mathbb{R}_+$ defined by

$$p(x) = \begin{cases} p, & x \in (0, b_1] \\ p + \frac{x-b_1}{\epsilon}(p' - p), & x \in (b_1, b_1 + \epsilon) \\ p', & x \in [b_1 + \epsilon, b_2 - \epsilon] \\ p' - \frac{x-(b_2-\epsilon)}{\epsilon}(p' - p), & x \in (b_2 - \epsilon, b_2) \\ p, & x \in [b_2, 1) \end{cases}$$

is a continuous and bounded probability density on $\mathcal{X}$ (w.r.t. to $\lambda_1$). Thus, $\pi(\mathrm{d}x) := p(x)\lambda_1(\mathrm{d}x)$ belongs to $\widetilde{\mathcal{P}}_b(\mathcal{X})$, as required. We now construct a sequence $(\zeta^N)_{N\geq 1}$ such that $\pi^N \overset{\mathrm{w}}{\Rightarrow} \pi$, $\mathbb{P}$-a.s.

Let $\big((\tilde{X}^{n,N})_{n=1}^N\big)_{N\geq 1}$ be a sequence of point sets in $\mathcal{X}$ such that, for all $m \geq 2$,

$$(\tilde{X}^{1,2^m}, \ldots, \tilde{X}^{2^m,2^m}) = \big(2^{-m}, 2^{1-m}, \ldots, 1 - 2^{-m}, v\big)$$

for some fixed non-dyadic number $v \in (b_2, 1)$. (This ensures that the $2^m$ points of the point set are all distinct.) For values of $N$ which are not powers of 2 we take for $(\tilde{X}^{n,N})_{n=1}^N$ a set of i.i.d. uniform random numbers in $\mathcal{X}$.

Next, for $N \geq 1$, let for $n = 1, \ldots, N$ $\tilde{W}^{n,N} = p(\tilde{X}^{n,N})/\sum_{m=1}^N p(\tilde{X}^{m,N})$ and define $\zeta^N = \big(\tilde{X}^{\sigma_N(n),N}, \tilde{W}^{\sigma_N(n),N}\big)_{n=1}^N$ where the sequence of permutations $(\sigma_N)_{N\geq 1}$ is defined below. Then, it is easily checked that $(\zeta^N)_{N\geq 1}$ is such that $\pi^N \overset{\mathrm{w}}{\Rightarrow} \pi$, $\mathbb{P}$-a.s., as required. We now construct a sequence of permutation $(\sigma_N)_{N\geq 1}$ for which $\mathbb{P}(\rho_{\mathrm{syst}}(\pi^N) \overset{\mathrm{w}}{\Rightarrow} \pi) < 1$. To follow the notation used throughout the paper we define $X^{n,N} = \tilde{X}^{\sigma_N(n),N}$ and $W^{n,N} = \tilde{W}^{\sigma_N(n),N}$ for $n \in 1:N$ so that $\zeta^N = \big(X^{n,N}, W^{n,N}\big)_{n=1}^N$.

Let $m \geq k$ and, with the shorthand $N_m = 2^m$, remark that, with $\mathbb{P}$-probability one

(as $\zeta^{N_m}$ is non-random for any $m \geq 1$)

$$\sum_{n=1}^{N_m} \mathbb{1}\big(X^{n,N_m} \in (0, b_1]\big) = a_1 2^{m-k}, \quad \sum_{n=1}^{N_m} \mathbb{1}\big(X^{n,N_m} \in [b_2, 1)\big) = 2^m - a_2\, 2^{m-k} + 1,$$
$$\sum_{n=1}^{N_m} \mathbb{1}\big(X^{n,N_m} \in (b_1, b_1 + \epsilon)\big) = \sum_{n=1}^{N_m} \mathbb{1}\big(X^{n,N_m} \in (b_2 - \epsilon, b_2)\big) = 2^{m-k} - 1,$$
$$\sum_{n=1}^{N_m} \mathbb{1}\big(X^{n,N_m} \in [b_1 + \epsilon, b_2 - \epsilon]\big) = (a_2 - a_1 - 2)2^{m-k} + 1$$
$$\sum_{n=1}^{N_m} p(X^{n,N_m}) \mathbb{1}\big(X^{n,N_m} \in (b_1, b_1 + \epsilon)\big) = \sum_{n=1}^{N_m} p(X^{n,N_m}) \mathbb{1}\big(X^{n,N_m} \in (b_2 - \epsilon, b_2)\big)$$
$$= \big(2^{m-k} - 1\big)\frac{p' - p}{2}.$$

and thus, taking $a = a_2 - a_1$, we have

$$\frac{1}{N_m} \sum_{n=1}^{N_m} p(X^{n,N_m}) = p\big(1 - a2^{-k} + 2^{-m}\big) + p'\big((a-2)2^{-k} + 2^{-m}\big) + (p' - p)\big(2^{-k} - 2^{-m}\big)$$
$$= p\big(1 - (a+1)2^{-k} + 2^{-m+1}\big) + p'(a-1)2^{-k} = p\, 2^{-m+1} + 1$$

using (S.12). Since $p = 0.5$, for $m \geq k$ and $n$ such that $X^{n,N_m} \in (0, b_1] \cup [b_2, 1)$,

$$N_m W^{n,N_m} = \frac{N_m}{2\,(N_m + 1)} = \frac{1}{2} - \frac{1}{2(N_m + 1)}. \tag{S.13}$$

Let $P_m$ denote the number of points $X^{n,N_m}$ in $(0, b_1]$; $P_m = a_1 2^{m-k} = N_m/4$. Note that $P_m \leq 2^m - a_2 2^{m-k} + 1$, i.e. it is possible to pair each point in $(0, b_1]$ with a different point in $[b_2, 1)$.

We take $\sigma_{N_m}$ to be a permutation that alternates between points in $(0, b_1)$ and points in $[b_2, 1]$; $\sigma_{N_m}(1 : N_m) = \Big(1, N_m, 2, N_m - 1, 3, \ldots, P_m, N_m - P_m + 1, \ldots\Big)$; the remaining components are arbitary.

We now show that for this sequence $(\zeta^N)_{N \geq 1}$ and probability measure $\pi \in \widetilde{\mathcal{P}}_b(\mathcal{X})$ we have $\mathbb{P}(\rho_{\text{syst}}(\pi^N) \overset{\text{w}}{\Rightarrow} \pi) \leq 3/4$.

Given (S.13) and given how systematic resampling operates, all the points in $(0, b_1]$ (resp. in $[b_2, 1]$) will get exactly one (resp. 0) off-spring as soon as:

$$u_1 < \frac{1}{2} - \frac{1}{2N_m + 1} - \frac{P_m - 1}{N_m + 1} = \frac{1}{4}$$

where $u_1$ is the first component of $u$.

Then, with $\mathbb{P}$-probability at least $1/4$, we have, for any $m \geq k$ and using (S.13),

$$\|\rho_{\text{syst}}(\zeta^{N_m}) - \pi^N\|_\star = \frac{1}{N_m} \max_{i \in 1:N_m} \left| \sum_{n=1}^{i} \Delta^{\sigma^*(\zeta^{N_m}, n)}_{\rho_{\text{syst}}, \zeta^{N_m}} \right| \geq \frac{1}{N_m} \left| \sum_{n=1}^{P_m} \Delta^{\sigma^*(\zeta^{N_m}, n)}_{\rho_{\text{syst}}, \zeta^{N_m}} \right|$$
$$= \frac{P_m}{N_m} \left| \frac{1}{2} + \frac{1}{2(1+N_m)} \right| \geq \frac{1}{8}.$$

where $\sigma^\star$ is a permutation that orders the points. This shows that $\mathbb{P}\big(\lim_{N \to +\infty} \|\rho_{\text{syst}}(\pi^N) - \pi\|_\star = 0\big) \leq 3/4$ and thus, by Theorem 3, $\mathbb{P}(\rho_{\text{syst}}(\pi^N) \overset{\text{w}}{\Rightarrow} \pi) \leq 3/4$. The proof is complete. □

## S3.3 Proof of Corollaries 1

*Proof of Corollary 1.* Assumption $(H_1)$ may be deduced from the proof of Corollary 2 (see below). We show below that $(H_2)$ holds for $r_N = \max(\sqrt{3N \log(N)/2}), 1) = o(N/\log N)$.

Let $N \geq 2$ and $z \in \mathcal{Z}^N$. For any $n \in 1:N$, $\Delta^n_{\rho_{\text{multi}}, z} = \sum_{i=1}^N (B^i_n - W^n)$ where $B^1_n, \ldots, B^N_n$ are i.i.d. random variables in $\{0,1\}$ such that $\mathbb{E}[B^i_n] = W^n$. Therefore, by Hoeffding's inequality,

$$\mathbb{P}\Big(\big|\Delta^n_{\rho_{\text{multi}}, z}\big| > \epsilon\, N\Big) \leq 2e^{-2\epsilon^2 N}, \quad \forall \epsilon > 0,\ \forall n \in 1:N.$$

Applying this result with $\epsilon = \sqrt{3 \log N/(2N)}$, we have

$$\mathbb{P}\Big(\max_{n \in 1:N} \big|\Delta^n_{\rho_{\text{multi}}, z}\big| > r_N\Big) \leq \sum_{n=1}^N \mathbb{P}\Big(\big|\Delta^n_{\rho_{\text{multi}}, z}\big| > r_N\Big) \leq 2\, N e^{-3 \log N} = \frac{2}{N^2}.$$

To conclude the proof it remains to show that $\sum_{n=1}^N \mathbb{E}\big[(\Delta^n_{\rho_{\text{multi}}, z})^2\big] \leq r_N\, N$. To this end, remark that

$$\mathbb{E}\big[(\Delta^n_{\rho_{\text{multi}}, z})^2\big] = \text{Var}\Big(\sum_{i=1}^N B^i_n\Big) = NW^n(1 - W^n), \quad \forall n \in 1:N$$

and thus $\sum_{n=1}^N \mathbb{E}\big[(\Delta^n_{\rho_{\text{multi}}, z})^2\big] \leq N \leq r_N\, N$ as required.

□

## S3.4 Proof of Corollary 2

*Proof of Corollary 2.* Let $N \geq 1$ and $z \in \mathcal{Z}^N$. Then, as $|\Delta_{\rho_{\text{strat}}, z}| \leq 2$, $\mathbb{P}$-a.s., it follows that

$$\sum_{n=1}^N \mathbb{E}\big[(\Delta^n_{\rho_{\sigma,\text{strat}}, z})^2\big] \leq 4\, N, \quad \mathbb{P}\Big(\max_{n \in 1:N} \big|\Delta^n_{\rho_{\text{strat}}, z}\big| > 4\Big) = 0$$

and thus $(H_2)$ holds for sequence $r_N = 4$.

To show $(H_1)$ we define

$$A_n = \big\{u \in [0,1] : F_N^-(u) = n\big\},\ V^n = \frac{n-1+U_n}{N},\ p_m^n := \mathbb{P}(V^n \in A_m), \quad n, m \in 1:N.$$

The collection of sets $(A_n)_{n=1}^N$ form a partition of $[0,1]$. Consequently, we can see $(A_n)_{n=1}^N$ as bins and the collection of independent random variables $(V^m)_{m=1}^N$ as balls in a "Ball and Bins" problem (see, e.g. Dubhashi and Ranjan, 1998), where for $n, m \in 1:N$, the probability that ball $n$ falls into bin $m$ is given by $p_m^n \in [0,1]$. The collection of random variables $(\#^n(U,z))_{n=1}^N$ can therefore be interpreted as occupancy numbers for the "Ball and Bins" problem we just described; that is, $\#^n(U,z)$ is the number of balls that fall into bin $n$. By Dubhashi and Ranjan (1998, Theorem 13), occupancy numbers in "Ball and Bins" problems are NA and thus $(\#^n(U,z))_{n=1}^N$ satisfies $(H_1)$. □

## S3.5 Proof of Corollary 3

*Proof of Corollary 3.* Let $N \geq 1$ and $z \in \mathcal{Z}^N$. Then, following a similar argument as in the proof of Corollary 2, it is easily checked that $|\Delta_{\rho_{\text{ssp}},z}| \leq 1$, $\mathbb{P}$-a.s. so that $(H_2)$ is verified for sequence $r_N = 1$. Lastly, $(H_1)$ is verified as well because $\rho_{\text{ssp}}$ is based on what is called a linear SSP process in Kramer et al. (2011) and thus, by Dubhashi et al. (2007, Theorem 5.1), the collection of random variables $(\#^n(U,z))_{n=1}^N$ is NA. □

## S3.6 Proofs for Section 4

### S3.6.1 Proof of Theorem 2

*Proof of Theorem 2.* The first part of the theorem is a particular case of Theorem 4 while the second part is direct consequence of Theorem 6 and of the computations used in the proof of Lemma 1. □

### S3.6.2 Proof of Proposition 2

*Proof of Proposition 2.* Remind first that, for any $x \in [0,1]^d$, the set $H^{-1}(x) \subset [0,1]$ contains at most $2^d$ elements. Then, let $h : [0,1]^d \to [0,1]$ be the mapping defined by

$$h(x) = \min H^{-1}(x), \quad x \in [0,1]^d.$$

By construction $H(h(x)) = x$ for all $x \in [0,1]^d$ and $h$ is one-to-one. Thus, to establish the proposition it remains to show that $h$ is a Borel measurable function.

To see this, remark that the mapping $h$ is such that, for all $m \geq 1$,

$$h^{-1}(I_m^d(0)) = S_m^d(0), \quad h^{-1}(I_m^d(k)) = S_m^d(k) \setminus \cup_{i=0}^{k-1} S_m^d(i), \quad k \in 1:2^{md}-1 \tag{S.14}$$

where, for $k \geq 1$, the set $S_m^d(k) \setminus \cup_{i=0}^{k-1} S_m^d(i)$ is obtained by removing the edges that $S_m^d(k)$ has in common with the closed hyper-cubes $S_m^d(i)$, $i = 0, \ldots, k-1$. To show that

(S.14) indeed holds, let $m \geq 1$ and note that, using the definition of $h$ and (S.1), we have

$$\begin{aligned} h^{-1}(I_m^d(0)) &= \{x \in [0,1]^d : h(x) \in I_m^d(0)\} \\ &= \{x \in [0,1]^d : H^{-1}(x) \cap I_m^d(0) \neq \varnothing\} \\ &= S_m^d(0) \end{aligned}$$

while, for $k \in 1 : 2^{md} - 1$,

$$\begin{aligned} h^{-1}(I_m^d(k)) &= \{x \in [0,1]^d : h(x) \in I_m^d(k)\} \\ &= \{x \in [0,1]^d : H^{-1}(x) \cap I_m^d(k) \neq \varnothing\} \setminus \{x \in [0,1]^d : H^{-1}(x) \cap \big( \cup_{p=0}^{k-1} I_m^d(p)\big) \neq \varnothing\} \\ &= S_m^d(k) \setminus \Big( \cup_{p=0}^{k-1} S_m^d(p)\Big). \end{aligned}$$

Next, let $\mathcal{B}([0,1])$ be the Borel $\sigma$-algebra on $[0,1]$ and $I \in \mathcal{B}([0,1])$. Then, because dyadic numbers are dense in $[0,1]$, there exists a sequence of closed dyadic intervals $(I_{m_n}^d(k_n))_{n\geq 1}$ such that $I = \cup_{n\geq 1} I_{m_n}^d(k_n)$, and thus

$$h^{-1}(I) = \cup_{n\geq 1} h^{-1}(I_{m_n}^d(k_n)).$$

By (S.14), the set $h^{-1}(I_{m_n}^d(k_n)) \subset [0,1]^d$ is an hypercube (which may be either open, or closed, or neither closed nor open) and is therefore a Borel set of $[0,1]^d$. This completes the proof. □

### S3.6.3 Proof of Theorem 3

*Proof of Theorem 3.* Since $\mathcal{X}$ is cubifiable and we consider probability measures in $\widetilde{\mathcal{P}}_b(\mathcal{X})$, we can assume without loss of generality that $\mathcal{X} = (0,1)^d$ and take $h_{\mathcal{X},\psi} = h$.

Let $\pi \in \widetilde{\mathcal{P}}_b(\mathcal{X})$ and $(\pi^N)_{N\geq 1}$ be a sequence such that $\pi^N \in \mathcal{P}_f^N(\mathcal{X})$ and $\pi^N \overset{\mathrm{w}}{\Rightarrow} \pi$. Note that, by Theorem 9, $\pi^N \overset{\mathrm{w}}{\Rightarrow} \pi$ implies that

$$\lim_{N\to\infty} \|\pi_h^N - \pi_h\|_\star = 0. \tag{S.15}$$

To establish the "if" part simply note that, by (S.15) and under the sufficient condition provided in the statement of the theorem,

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

so that the result follows from Theorem 9.

To establish the "only if" part assume that $\rho(\zeta^N) \overset{\mathrm{w}}{\Rightarrow} \pi$, $\mathbb{P}$-a.s. By Theorem 9, $\lim_{N\to\infty} \|\rho(\zeta^N)_h - \pi_h\|_\star = 0$, $\mathbb{P}$-a.s. and therefore, by (S.15) and the triangle inequality,

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

This completes the proof. □

### S3.6.4 Proof of Theorem 4

*Proof of Theorem 4.* Since $\mathcal{X}$ is cubifiable and $\pi \in \widetilde{\mathcal{P}}_b^*(\mathcal{X})$, we can assume without loss of generality that $\mathcal{X} = (0,1)^d$ and thus $h = h_{\mathcal{X}}$. Let $\varphi \in \mathcal{C}_b(\mathcal{X})$ and denote $\varphi_H = \varphi \circ H \in \mathcal{C}_b((0,1))$. Then, for any $N \geq 1$, we have:

$$
\begin{aligned}
&N\mathrm{Var}_{\zeta^N}\left[\rho^*_{\mathrm{strat}}(\zeta^N)(\varphi)\right] \\
&= \pi^N(\varphi^2) - N\sum_{n=1}^{N}\Big(\int_{\frac{n-1}{N}}^{\frac{n}{N}} \varphi_H \circ F^-_{\pi_h^N}(u)\mathrm{d}u\Big)^2 \\
&= \pi^N(\varphi^2) - N\sum_{n=1}^{N}\Big(\int_{\frac{n-1}{N}}^{\frac{n}{N}} \varphi_H \circ F^-_{\pi_h}(u)\mathrm{d}u + \int_{\frac{n-1}{N}}^{\frac{n}{N}} \big(\varphi_H \circ F^-_{\pi_h^N} - \varphi_H \circ F^-_{\pi_h}\big)(u)\mathrm{d}u\Big)^2 \\
&= \pi^N(\varphi^2) - N\sum_{n=1}^{N}\Big(\int_{\frac{n-1}{N}}^{\frac{n}{N}} \varphi_H \circ F^-_{\pi_h}(u)\mathrm{d}u\Big)^2 \\
&- N\sum_{n=1}^{N}\Big(\int_{\frac{n-1}{N}}^{\frac{n}{N}} \big(\varphi_H \circ F^-_{\pi_h^N} - \varphi_H \circ F^-_{\pi_h}\big)(u)\mathrm{d}u\Big)^2 \\
&- 2N\sum_{n=1}^{N}\Big(\int_{\frac{n-1}{N}}^{\frac{n}{N}} \varphi_H \circ F^-_{\pi_h}(u)\mathrm{d}u\Big)\Big(\int_{\frac{n-1}{N}}^{\frac{n}{N}} \big(\varphi_H \circ F^-_{\pi_h^N} - \varphi_H \circ F^-_{\pi_h}\big)(u)\mathrm{d}u\Big). \qquad \text{(S.16)}
\end{aligned}
$$

For the first term, under the assumptions of the theorem

$$
\lim_{N\to+\infty} \pi^N(\varphi^2) = \pi(\varphi^2), \quad \mathbb{P}-a.s. \qquad \text{(S.17)}
$$

For the second term, since $F_{\pi_h}$ is continuous (Lemma 2) and strictly increasing on $[0,1]$ (Lemme S3), $F^-_{\pi_h}$ is a continuous function on $[0,1]$. Hence, the function $\varphi_H \circ F^-_{\pi_h}$ belongs to $\mathcal{C}_b([0,1])$ and is Riemann integrable. Consequently,

$$
N\sum_{n=1}^{N}\Big(\int_{\frac{n-1}{N}}^{\frac{n}{N}} \varphi_H \circ F^-_{\pi_h}(u)\mathrm{d}u\Big)^2 = \frac{1}{N}\sum_{n=1}^{N}\{\varphi_H \circ F^-_{\pi_h}(u_n)\}^2 \to \pi(\varphi^2) \qquad \text{(S.18)}
$$

for some $u_i \in [n-1/N, n/N]$ (mean value theorem).

For the fourth term,

$$
\begin{aligned}
&N\Big|\sum_{n=1}^{N}\Big(\int_{\frac{n-1}{N}}^{\frac{n}{N}} \varphi_H \circ F^-_{\pi_h}(u)\mathrm{d}u\Big)\Big(\int_{\frac{n-1}{N}}^{\frac{n}{N}} \big(\varphi_H \circ F^-_{\pi_h^N} - \varphi_H \circ F^-_{\pi_h}\big)(u)\mathrm{d}u\Big)\Big| \\
&\qquad \leq \|\varphi\|_\infty \int_0^1 \big|\varphi_H \circ F^-_{\pi_h^N}(u) - \varphi_H \circ F^-_{\pi_h}(u)\big|\mathrm{d}u. \qquad \text{(S.19)}
\end{aligned}
$$

By Lemma S3, on an event of $\mathbb{P}$-probability 1, $\lim_{N\to+\infty} |F^-_{\pi_h^N}(u) - F^-_{\pi_h}(u)| = 0$ for all $u \in [0,1]$. Therefore, using the fact that $\varphi_H$ is continuous and bounded, we have, by the dominated convergence theorem,

$$
\lim_{N\to+\infty} \int_0^1 \big|\varphi_H \circ F^-_{\pi_h^N}(u) - \varphi_H \circ F^-_{\pi_h}(u)\big|\mathrm{d}u = 0, \quad \mathbb{P}-a.s. \qquad \text{(S.20)}
$$

and, similarly for the third term:

$$\begin{aligned}\lim_{N\to+\infty} N\sum_{n=1}^{N}\Big(\int_{\frac{n-1}{N}}^{\frac{n}{N}} \big(\varphi_H\circ F^-_{\pi_h^N} - \varphi_H\circ F^-_{\pi_h}\big)(u)\mathrm{d}u\Big)^2 \\ \leq 2\|\varphi\|_\infty \lim_{N\to+\infty}\int_0^1 \big|\varphi_H\circ F^-_{\pi_h^N}(u) - \varphi_H\circ F^-_{\pi_h}(u)\big|\mathrm{d}u \\ = 0.\end{aligned} \tag{S.21}$$

Putting (S.16)-(S.21) together shows that $N\mathrm{Var}\left[\rho^*_{\mathrm{strat}}(\pi^N)(\varphi)\right] \to 0$, $\mathbb{P}$-a.s. as required. □

*Proof of Corrolary 4.* Let $N \geq 1$, $z \in \mathcal{Z}^N$, and $u_{1:N} \in [0,1]^N$. Then, for any $a \in [0,1]$,

$$\begin{aligned}\left|\frac{1}{N}\sum_{n=1}^{N}\mathbb{1}\big(F^-_{\pi^N_{h_\mathcal{X}}}(u_n) \leq a\big) - \pi^N_{h_\mathcal{X}}([0,a])\right| &= \left|\frac{1}{N}\sum_{n=1}^{N}\mathbb{1}\Big(u_n \leq F_{\pi^N_{h_\mathcal{X}}}(a)\Big) - F_{\pi^N_{h_\mathcal{X}}}(a)\Big)\right| \\ &\leq D^\star_N(u_{1:N}).\end{aligned}$$

Thus, the condition (10) implies the condition (8) of Theorem 3. For the second part of the corollary, we simply use the well-known fact (Niederreiter, 1992, Theorem 2.6, p.15) that

$$D^\star_N\big(u_{1:N}\big) = \frac{1}{2N} + \max_{1\leq n\leq N}\left|u_n - \frac{n-1/2}{N}\right| \tag{S.22}$$

which implies that for the instances of $\phi^n_N$ corresponding to stratified resampling and systematic resampling given in the statement of the corollary, $D^\star_N(\phi^n_N(U)) \leq 1/N$, $\mathbb{P}$-a.s. □

### S3.6.5 Proof of Theorem 5

We first recall the following result due to Aistleitner and Dick (2015) that will play a key role in the poof of Theorems 5 and 6.

**Theorem S2.** *(Aistleitner and Dick, 2015, Theorem 1) Let $\varphi : [0,1]^d \to \mathbb{R}$ be a measurable function, $\pi \in \mathcal{P}([0,1]^d)$ and $(x_n)_{n=1}^N$ be a set of $N \geq 1$ points in $[0,1]^d$. Then,*

$$\left|\frac{1}{N}\sum_{n=1}^{N}\varphi(x_n) - \int_{[0,1]^d}\varphi(x)\pi(\mathrm{d}x)\right| \leq V(\varphi)\left\|N^{-1}\sum_{n=1}^{N}\delta(x_n) - \pi\right\|_\star$$

*where $V(\varphi)$ is the variation of $\varphi$ in the sense of Hardy and Krause.*

*Proof of Theorem 5.* Since $\mathcal{X}$ is cubifiable and $\pi \in \widetilde{\mathcal{P}}^*_b(\mathcal{X})$, we can assume without loss of generality that $\mathcal{X} = (0,1)^d$ and thus $h_\mathcal{X} = h$.

Let $N \geq 1$, $z \in \mathcal{Z}^N$ and, for $n \in 1:N$,

$$V^n = \frac{n-1+U_n}{N}, \quad \hat{X}^n = H\circ F^-_{\pi^N_h}(V^n), \quad \bar{x}^n = H\Big(\mathbb{E}_{\zeta^N}\big[F^-_{\pi^N_h}(V^n)\big]\Big).$$

Let $\varphi_I : [0,1] \to [0,1]$ be such that $\varphi_I(x) = x$, $\forall x \in [0,1]$, and remark that the variation of $\varphi_I$ is $V(\varphi_I) = |\varphi_I(1) - \varphi_I(0)| = 1$. Therefore, by Theorem S2,

$$\begin{aligned}\mathrm{Var}\left[\frac{1}{N}\sum_{n=1}^{N} F^-_{\pi_h^N}(V^n)\right] \leq V(\varphi_I)^2 \mathbb{E}\big[\|\rho^*_{\mathrm{strat}}(\zeta^N)_h - \pi_h^N\|_\star^2\big] \leq \mathbb{E}\big[D_N^\star\big((V^n)_{n=1}^N\big)^2\big] \\ \leq \frac{1}{N^2}\end{aligned} \tag{S.23}$$

where the second inequality comes from equation (10) in the proof of Corollary 4 and the last one is due to (S.22) and the definition of $(V^n)_{n=1}^N$.

Next, let $\varphi : \mathcal{X} \to \mathbb{R}$ and $C_{\varphi,\psi_\mathcal{X}} \in (0,+\infty)$ be as in the statement of the theorem. Then, denoting by $\tilde{C}_d \in (0,+\infty)$ the Hölder constant of the Hilbert curve $H$ for the $\|\cdot\|_2$ norm, we have

$$\begin{aligned}
\mathrm{Var}[\rho^*_{\mathrm{strat}}(z)(\varphi)] &= \frac{1}{N^2}\sum_{n=1}^{N} \mathrm{Var}\Big[\varphi(\hat{X}^n) - \varphi(\bar{x}^n)\Big] \\
&\leq \frac{1}{N^2}\sum_{n=1}^{N} \mathbb{E}\Big[\big(\varphi(\hat{X}^n) - \varphi(\bar{x}^n)\big)^2\Big] \\
&\leq \frac{C^2_{\varphi,\psi_\mathcal{X}}}{N^2}\sum_{n=1}^{N} \mathbb{E}\Big[\big\|\hat{X}^n - \bar{x}^n\big\|_2^{2\gamma}\Big] \\
&\leq \frac{C^2_{\varphi,\psi_\mathcal{X}}\,\tilde{C}_d^{2\gamma}}{N^2}\sum_{n=1}^{N} \mathbb{E}\Big[\Big(F^-_{\pi_h^N}(V^n) - \mathbb{E}\big[F^-_{\pi_h^N}(V^n)\big]\Big)^{\frac{2\gamma}{d}}\Big] && \text{(S.24)} \\
&\leq \frac{C^2_{\varphi,\psi_\mathcal{X}}\,\tilde{C}_d^{2\gamma}}{N^2}\sum_{n=1}^{N} \mathbb{E}\Big[\Big(F^-_{\pi_h^N}(V^n) - \mathbb{E}\big[F^-_{\pi_h^N}(V^n)\big]\Big)^{2}\Big]^{\frac{\gamma}{d}} && \text{(S.25)} \\
&\leq \frac{C^2_{\varphi,\psi_\mathcal{X}}\,\tilde{C}_d^{2\gamma}}{N}\left(\frac{1}{N}\sum_{n=1}^{N} \mathbb{E}\Big[\Big(F^-_{\pi_h^N}(V^n) - \mathbb{E}\big[F^-_{\pi_h^N}(V^n)\big]\Big)^{2}\Big]\right)^{\frac{\gamma}{d}} && \text{(S.26)} \\
&= \frac{C^2_{\varphi,\psi_\mathcal{X}}\,\tilde{C}_d^{2\gamma}}{N}\Big(\frac{1}{N}\sum_{n=1}^{N} \mathrm{Var}\big[F^-_{\pi_h^N}(V^n)\big]\Big)^{\frac{\gamma}{d}} \\
&\leq C^2_{\varphi,\psi_\mathcal{X}}\,\tilde{C}_d^{2\gamma} N^{-1-\frac{\gamma}{d}} && \text{(S.27)}
\end{aligned}$$

where (S.24) is due to the Hölder continuity of the Hilbert curve, (S.25) and (S.26) are due to Jensen's inequality and the fact that, for any $\alpha \in (0,1)$, the function $x^\alpha$ is concave on $\mathbb{R}_+$ while (S.27) comes from (S.23). The result follows from the fact that $\tilde{C}_d \leq 2\sqrt{d+3}$ (see e.g. the proof of Zumbusch, 2003, Lemma 4.3, pp 97-99). □

### S3.6.6 Proof of Theorem 6

*Proof of Theorem 6.* Using the same notation and computations as in the proof of Theorem 5 we note that, to establish the result, it is enough to show below that

$$\mathrm{Var}_{\zeta^N}\left[\frac{1}{N}\sum_{n=1}^{N} F^{-}_{\pi^N_h}(V^n)\right] = o(N^{-2}), \quad \mathbb{P}-a.s.$$

Let $N \geq 1$ and define

$$X^n_h = h(X^n), \quad a_n = F_{\pi^N_h}(X^{n-1}_h), \quad b_n = F_{\pi^N_h}(X^n_h), \quad n \in 1:N$$

with the convention $F_{\pi^N_h}(X^n_h) = 0$ when $n = 0$.

Next, let $\alpha > 0$ and $F_{\pi^N_h,\alpha} : [0, X^N_h] \to [0,1]$ be the mapping defined by

$$F_{\pi^N_h,\alpha}(z) = a_n + (b_n - a_n)\Big(\frac{z - X^{n-1}_h}{X^n_h - X^{n-1}_h}\Big)^{\frac{1}{\alpha}}, \quad z \in [X^{n-1}_h, X^n_h], \;\; n \in 1:N$$

with the convention $X^n_h = 0$ when $n = 0$. Let $F^{-}_{\pi^N_h,\alpha} : [0,1] \to [0, X^N_h]$ be the inverse of $F_{\pi^N_h,\alpha}$, which is defined by

$$F^{-}_{\pi^N_h,\alpha}(u) = X^{n-1}_h + \big(X^n_h - X^{n-1}_h\big)\left(\frac{u - a_n}{b_n - a_n}\right)^{\alpha}, \quad u \in [a_n, b_n], \; n \in 1:N.$$

For any $\alpha > 0$, the function $F_{\pi^N_h,\alpha}$ is continuous and strictly increasing on $[0, X^N_h]$ and such that $F_{\pi^N_h,\alpha}(X^n_h) = F_{\pi^N_h}(X^n_h)$ for all $n \in 1:N$. This last property implies that

$$\|F_{\pi^N_h,\alpha} - F_{\pi^N_h}\|_\infty \leq \max_{n\in 1:N} W^{n,N}. \tag{S.28}$$

As preliminary computations, remark that

$$F^{-}_{\pi^N_h}(u) - F^{-}_{\pi^N_h,\alpha}(u) = (X^n_h - X^{n-1}_h)\left[1 - \left(\frac{u - a_n}{b_n - a_n}\right)^{\alpha}\right], \quad u \in (a_n, b_n], \; n \in 1:N$$

and thus

$$\int_{a_n}^{b_n} \big(F^{-}_{\pi^N_h}(u) - F^{-}_{\pi^N_h,\alpha}(u)\big)^2 \mathrm{d}u = (X^n_h - X^{n-1}_h)^2 (b_n - a_n)\frac{2\alpha^2}{(\alpha+1)(2\alpha+1)}. \tag{S.29}$$

Lastly, let $\alpha_N > 0$ be such that

$$\frac{2\alpha_N^2}{(\alpha_N+1)(2\alpha_N+1)} = \frac{1}{N^2}$$

and, to simplify the notation, we use the shorthand $\widetilde{F}_{\pi^N_h} = F_{\pi^N_h,\alpha_N}$ and $\widetilde{F}^{-}_{\pi^N_h} = F^{-}_{\pi^N_h,\alpha_N}$ in what follows.

By construction $\widetilde{F}_{\pi_h^N}$ is continuous and strictly increasing on $[0, X_h^N]$ and thus $\widetilde{F}^-_{\pi_h^N}$ is continuous on $[0,1]$. In addition, by (S.28), Lemma S3 and the assumptions of the theorem,

$$\lim_{N\to+\infty} \|\widetilde{F}^-_{\pi_h^N} - F^-_{\pi_h}\|_\infty = 0, \quad \mathbb{P}-a.s. \tag{S.30}$$

Let $g_N : [0,1] \to [0,1]$ be defined by

$$g_N(u) = \widetilde{F}^-_{\pi_h^N}(u) - F^-_{\pi_h}(u), \quad u \in [0,1].$$

Then,

$$\begin{aligned}\mathrm{Var}_{\zeta^N}\left[\frac{1}{N}\sum_{n=1}^N F^-_{\pi_h^N}(V^n)\right] \le \Bigg(&\mathrm{Var}_{\zeta^N}\left[\frac{1}{N}\sum_{n=1}^N F^-_{\pi_h}(V^n)\right]^{1/2} + \mathrm{Var}_{\zeta^N}\left[\frac{1}{N}\sum_{n=1}^N g_N(V^n)\right]^{1/2} \\ &+ \mathrm{Var}_{\zeta^N}\left[\frac{1}{N}\sum_{n=1}^N \left(\widetilde{F}^-_{\pi_h^N}(V^n) - F^-_{\pi_h^N}(V^n)\right)\right]^{1/2}\Bigg)^2\end{aligned}$$

where, by assumption, the first variance is of order $o(N^{-2})$, $\mathbb{P}$-a.s. In addition, for the last variance, we have, using the properties of the random variables $(V^n)_{n=1}^N$ and (S.29),

$$\begin{aligned}\mathrm{Var}_{\zeta^N}\left[\frac{1}{N}\sum_{n=1}^N \left(\widetilde{F}^-_{\pi_h^N}(V^n) - F^-_{\pi_h^N}(V^n)\right)\right] &= \frac{1}{N^2}\sum_{n=1}^N \mathrm{Var}_{\zeta^N}\left[\widetilde{F}^-_{\pi_h^N}(V^n) - F^-_{\pi_h^N}(V^n)\right] \\ &\le \frac{1}{N^2}\sum_{n=1}^N \mathbb{E}_{\zeta^N}\left[\widetilde{F}^-_{\pi_h^N}(V^n) - F^-_{\pi_h^N}(V^n)\right]^2 \\ &= \frac{1}{N}\int_0^1 \left(\widetilde{F}^-_{\pi_h^N}(u) - F^-_{\pi_h^N}(u)\right)^2 \mathrm{d}u \\ &= \frac{1}{N}\sum_{n=1}^N \int_{a_n}^{b_n} \left(F^-_{\pi_h^N}(u) - \widetilde{F}^-_{\pi_h^N}(u)\right)^2 \mathrm{d}u \\ &\le \frac{1}{N}\frac{2\alpha_N^2}{(\alpha_N+1)(2\alpha_N+1)} \\ &= \frac{1}{N^3}.\end{aligned}$$

Hence, to prove the theorem it remains to show that

$$\mathrm{Var}_{\zeta^N}\left[\frac{1}{N}\sum_{n=1}^N g_N(V^n)\right] = o(N^{-2}), \quad \mathbb{P}-a.s. \tag{S.31}$$

To establish this result note that, using the properties of $(V^n)_{n=1}^N$ and the mean value theorem (which can be used because $g_N$ is continuous, see Lemmas 2 and S3),

$$\mathrm{Var}_{\zeta^N}\Big[\sum_{n=1}^N g_N(V^n)\Big] = N\Big(\int_0^1 g_N(u)^2 \mathrm{d}u - \frac{1}{N}\sum_{n=1}^N g_N(v_n)^2\Big)$$

for some $v_n \in [(n-1)/N, n/N]$.

By (S.22), we have $D_N^*((v_n)_{n=1}^N) \leq N^{-1}$ and thus, by Theorem S2,

$$\mathrm{Var}_{\zeta^N}\Big[\sum_{n=1}^N g_N(V^n)\Big] \leq N\, D_N^*\big((v_n)_{n=1}^N\big)\, V(g_N^2) \leq V(g_N^2) \tag{S.32}$$

with $V(g_N^2)$ the variation of the function $g_N^2$ on $[0,1]$.

To control this quantity recall that $V(g_N^2) = \sup_{P\in \mathrm{P}} S_{g_N^2}(P)$ where P is the set of all partitions of $[0,1]$ and where, for a partition $P = (z_i)_{i=0}^{M_P} \in \mathrm{P}$ of size $M_p$,

$$S_{g_N^2}(P) := \sum_{i=1}^{M_P} |g_N(z_i)^2 - g_N(z_{i-1})^2|.$$

Next, remark that for any $P = (z_i)_{i=0}^{M_P} \in \mathrm{P}$ we have

$$\begin{aligned} S_{g_N^2}(P) &= \sum_{i=1}^{M_P} |g_N(z_i)^2 - g_N(z_{i-1})^2| \\ &\leq \sum_{i=1}^{M_P} |g_N(z_i)||g_N(z^i) - g_N(z_{i-1})| + \sum_{i=1}^{M_P} |g_N(z_{i-1})||g_N(z_i) - g_N(z_{i-1})| \\ &\leq 2\|g_N\|_\infty \sum_{i=1}^{M_P} |g_N(z_i) - g_N(z_{i-1})| \end{aligned}$$

so that $V(g_N^2) \leq 2\|g_N\|_\infty V(g_N) \leq 4\|g_N\|_\infty$ where the last inequity uses the fact that

$$V(g_N) = V(\widetilde{F}^-_{\pi_h^N} - F^-_{\pi_h}) \leq V(\widetilde{F}^-_{\pi_h^N}) + V(F^-_{\pi_h}) = |\widetilde{F}^-_{\pi_h^N}(1) - \widetilde{F}^-_{\pi_h^N}(0)| + |F^-_{\pi_h}(1) - F^-_{\pi_h}(0)| \leq 2.$$

Using (S.30),

$$\lim_{N\to+\infty} \|g_N\|_\infty = \lim_{n\to+\infty} \|\widetilde{F}^-_{\pi_h^N} - F^-_{\pi_h}\|_\infty = 0, \quad \mathbb{P}-a.s.$$

and thus, together with (S.32), it follows that

$$\lim_{N\to+\infty} \mathrm{Var}_{\zeta^N}\Big[\sum_{n=1}^N g_N(V^n)\Big] = o(1), \quad \mathbb{P}-a.s.$$

showing (S.31). This completes the proof of the first part of the theorem.

We now prove the second part of the theorem. Recall that, by the bi-measure property of Hilbert curve, for any $0 < a < b < 1$ we have

$$F_{\pi_h}(b) - F_{\pi_h}(a) = \pi_h\big((a,b]\big) = \pi\big(H((a,b])\big),$$

where $\lambda_d(H(a,b]) = (b-a)$. Therefore, because by assumption there exists a constant $c < +\infty$ such that $c^{-1}\lambda_d(A) \leq \pi(A) \leq c\lambda_d(A)$ for all measurable sets $A \subseteq \mathcal{X}$, if follows that

$$\frac{1}{c}|b-a| \leq \left|F_{\pi_h}(b) - F_{\pi_h}(a)\right| \leq c\,|b-a|, \quad \forall a, b \in (0,1).$$

Therefore, $F_{\pi_h}$ is bi-Lipschitz on $(0,1)$ and thus $F^-_{\pi_h}$ is Lipschitz on $(0,1)$. Using this last property of $F_{\pi_h}$ it is readily checked that the rate in (11) is $\mathcal{O}(N^{-3})$ and the result follows.

□

## S3.7 Proofs of Section 5

### S3.7.1 Proof of Theorem 7

We prove below this more general result.

**Theorem S3.** *For Algorithm 2, assuming that $\mathcal{X}$ is a cubifiable set, $\mathcal{P}_0 = \mathcal{P}_b^*(\mathcal{X})$, $\rho \in \{\rho_{\text{multi}}, \rho_{\text{res/multi}}, \rho^\star_{\text{strat}}\}$ and that the Feynman-Kac model fulfils assumptions (G) and (M), for any test function $\varphi \in \mathcal{C}_b(\mathcal{X})$, we have that (for any $t \geq 0$)*

$$N^{1/2}\left\{\frac{1}{N}\sum_{n=1}^{N}\varphi(X_t^n) - (\pi_{t-1}M_t)(\varphi)\right\} \overset{w}{\Rightarrow} \mathcal{N}_d\left(0, \widetilde{\mathcal{V}}_t[\varphi]\right) \tag{S.33}$$

$$N^{1/2}\left\{\sum_{n=1}^{N}W_t^n\varphi(X_t^n) - \pi_t(\varphi)\right\} \overset{w}{\Rightarrow} \mathcal{N}_d\left(0, \mathcal{V}_t[\varphi]\right) \tag{S.34}$$

$$N^{1/2}\left\{\frac{1}{N}\sum_{n=1}^{N}\varphi(X_t^{A_{t+1}^n}) - \pi_t(\varphi)\right\} \overset{w}{\Rightarrow} \mathcal{N}_d\left(0, \hat{\mathcal{V}}_t[\varphi]\right), \quad t \geq 0 \tag{S.35}$$

*where the $\mathcal{V}_t(\varphi)$ are defined recursively as follows: $\widetilde{\mathcal{V}}_0[\varphi] = V_0(\varphi)$,*

$$\begin{aligned}
\mathcal{V}_t[\varphi] &= \frac{1}{\ell_t^2}\widetilde{\mathcal{V}}_t[G_t\{\varphi - \pi_t(\varphi)\}] \\
\widehat{\mathcal{V}}_t[\varphi] &= \mathcal{V}_t[\varphi] + R_t(\rho, \varphi) \\
\widetilde{\mathcal{V}}_{t+1}[\varphi] &= \widehat{\mathcal{V}}_t[M_{t+1}(\varphi)] + \pi_t[V_{t+1}(\varphi)]
\end{aligned}$$

*and*

$$0 = R_t(\rho_{\text{strat}^\star}, \varphi) \leq R_t(\rho_{\text{res/multi}}, \varphi) \leq R_t(\rho_{\text{multi}}, \varphi).$$

*Proof of Theorem S3.* There is nothing to prove for multinomial and residual resampling, and, for $\rho = \rho^\star_{\text{strat}}$, it is enough to prove that (S.34) $\Rightarrow$ (S.35) for all $t \geq 0$, as (S.35) $\Rightarrow$ (S.33) $\Rightarrow$ (S.34) have already been established in e.g. Chopin (2004). Note in addition that Assumptions (M) and (V) ensure that the operators $\mathcal{V}$, $\widehat{\mathcal{V}}$ and $\widetilde{\mathcal{V}}$ map $\mathcal{C}_b(\mathcal{X})$ into itself.

Assuming (S.34),

$$N^{1/2}\left\{\sum_{n=1}^{N} W_t^n \varphi(X_t^n) - \pi_t(\varphi)\right\} \stackrel{\mathrm{w}}{\Rightarrow} \mathcal{N}_d\left(0, \mathcal{V}_t\left[\varphi\right]\right)$$

we have

$$N^{1/2}\left\{\frac{1}{N}\sum_{n=1}^{N} \varphi(X_t^{A_{t+1}^n}) - \pi_t(\varphi)\right\} = N^{1/2}\left\{\frac{1}{N}\sum_{n=1}^{N} \varphi(X_t^{A_{t+1}^n}) - \sum_{n=1}^{N} W_t^n \varphi(X_t^n)\right\} + N^{1/2}\left\{\sum_{n=1}^{N} W_t^n \varphi(X_t^n) - \pi_t(\varphi)\right\}.$$

and the result is proven provided we can apply Theorem 4 to the first term: then this term converges to 0 in $L^2$, and thus in probability (by Chebyshev's inequality), and we can apply Slutsky's theorem to conclude.

To apply Theorem 4, we need to establish that, $\mathbb{P}-a.s.$, $\max_n W_t^n \to 0$ as $N \to +\infty$. By assumption (G), there exists a $\gamma_t < +\infty$ such that

$$W_t^n = \frac{G_t(X_t^n)}{\sum_{m=1}^N G_t(X_t^m)} \leq \frac{\gamma_t}{N}\left(\frac{1}{N}\sum_{m=1}^{N} G_t(X_{t-1}^{A_t^m}, X_t^m)\right)^{-1}$$

and we know that

$$\left\{\frac{1}{N}\sum_{m=1}^{N} G_t(X_{t-1}^{A_t^m}, X_t^m) - \left(\pi_{t-1}M_t\right)\left(G_t\right)\right\} \to 0 \quad \mathbb{P}-a.s.$$

where $\left(\pi_{t-1}M_t\right)\left(G_t\right) = \ell_t > 0$. Thus $\mathbb{P}\left(\max_{n\in 1:N} W_t^n \to 0\right) = 1$. □

### S3.7.2 Proof of Theorem 8

*Proof of Theorem 8.* Let $\mathcal{F}_{t-1} = \sigma(X_0^{1:N}, \ldots, U_{t-1}, X_{t-1}^{1:N})$, then, for any unbiased scheme:

$$\mathbb{E}\left[\ell_t^N | \mathcal{F}_{t-1}\right] = \sum_{n=1}^{N} W_{t-1}^n M_t(X_{t-1}^n, G_t)$$

which does not depend on $\eta_{t-1}$. Thus we wish to minimise the expectation of $\mathrm{Var}\left[\ell_t^N | \mathcal{F}_{t-1}\right]$. (The same remark applies to $L_t^N$, as $\mathbb{E}\left[L_t^N | \mathcal{F}_{t-1}\right] = L_{t-1}^N \mathbb{E}\left[\ell_t^N | \mathcal{F}_{t-1}\right]$. For simplicity, we work with $\ell_t^N$ from now on.) Under multinomial resampling, the $(A_t^n, X_t^n)$'s are IID conditional on $\mathcal{F}_{t-1}$, thus (for any $n$)

$$\mathrm{Var}\left[\ell_t^N | \mathcal{F}_{t-1}\right] = N^{-1}\mathrm{Var}\left[w_t^n | \mathcal{F}_{t-1}\right]$$

and this quantity is minimal when

$$\mathbb{E}\left[\left(w_t^n\right)^2 | \mathcal{F}_{t-1}\right] = \sum_{n=1}^{N} \frac{\left(W_{t-1}^n\right)^2}{\widetilde{W}_{t-1}^n} M_t(X_{t-1}^n, G_t^2)$$

is minimal. Using constrained optimisation (the constraint being $\sum_{n=1}^N \widetilde{W}_{t-1}^n = 1$), it is easy to see that this quantity is minimal (with respect to the $N$ twisted weights) when

$$\widetilde{W}_{t-1}^n \propto W_{t-1}^n \sqrt{M_t(X_{t-1}^n, G_t^2)}$$

and thus taking $\eta_{t-1}(x_{t-1}) = M_t(x_{t-1}, G_t^2)$ gives a lower bound for the corresponding expectation.

We now take $\rho = \rho^\star_{\text{strat}}$; the $(A_t^n, X_t^n)$'s are not conditionally IID any more, but

$$\text{Var}\left[\ell_t^N | \mathcal{F}_{t-1}\right] = \text{Var}\left\{\mathbb{E}\left[\ell_t^N | \hat{\mathcal{F}}_{t-1}\right] | \mathcal{F}_{t-1}\right\} + \mathbb{E}\left\{\text{Var}\left[\ell_t^N | \hat{\mathcal{F}}_{t-1}\right] | \mathcal{F}_{t-1}\right\}$$

where $\hat{\mathcal{F}}_{t-1} = \sigma(X_{t-1}^{1:N}, A_t^{1:N})$. For the first term:

$$\mathbb{E}\big[\ell_t^N | \hat{\mathcal{F}}_{t-1}\big] = \frac{1}{N}\sum_{n=1}^N \frac{W_{t-1}^{A_t^n}}{\widetilde{W}_{t-1}^{A_t^n}} M_t(X_{t-1}^{A_t^n}, G_t) = \frac{\sum_{n=1}^N \widetilde{w}_{t-1}^n}{\sum_{n=1}^N w_{t-1}^n} \times \frac{1}{N}\sum_{n=1}^N \psi(X_{t-1}^{A_t^n})$$

where $\psi(x_{t-1}) = M_t(x_{t-1}, G_t)/\eta_{t-1}(x_{t-1})$. Since $\mathcal{X}$ is compact and $\eta_{t-1}$ takes values in $\mathbb{R}_{>0}$, $\psi \in \mathcal{C}_b(\mathcal{X})$, and we can apply Theorem 4:

$$\text{MSE}\left[\frac{1}{N}\sum_{n=1}^N \psi(X_{t-1}^{A_t^n})\right] = o(N^{-1})$$

whith a constant that depends only on $\psi$. Hence, the first term is at a $o(N^{-1/2})$ $L^2-$distance of quantity

$$\frac{\sum_{n=1}^N \widetilde{w}_{t-1}^n}{\sum_{n=1}^N w_{t-1}^n} \times \sum_{n=1}^N \widetilde{W}_{t-1}^n \psi(X_{t-1}^n) = \frac{\sum_{n=1}^N \widetilde{w}_{t-1}^n \psi(X_{t-1}^n)}{\sum_{n=1}^N w_{t-1}^n} = \frac{\sum_{n=1}^N w_{t-1}^n M_t(X_{t-1}^n, G_t)}{\sum_{n=1}^N w_{t-1}^n}$$

which (a) does not depend on $\eta_{t-1}$; (b) converges at rate $\mathcal{O}_P(N^{-1/2})$. Hence the part that depends on $\eta_{t-1}$ becomes negligible when $N \to +\infty$.

Now for the second term:

$$\text{Var}\left[\ell_t^N | \hat{\mathcal{F}}_{t-1}\right] = \frac{1}{N^2}\sum_{n=1}^N \left(\frac{W_{t-1}^{A_t^n}}{\widetilde{W}_{t-1}^{A_t^n}}\right)^2 \text{Var}_{M_t}(X_{t-1}^{A_t^n}, G_t)$$

the expectation of which is

$$\mathbb{E}\left\{\text{Var}\left[\ell_t^N | \hat{\mathcal{F}}_{t-1}\right] | \mathcal{F}_{t-1}\right\} = \frac{1}{N}\sum_{n=1}^N \frac{\left(W_{t-1}^n\right)^2}{\widetilde{W}_{t-1}^n} \text{Var}_{M_t}(X_{t-1}^n, G_t)$$

which is mininal (under the constraint that $\sum_{n=1}^N \widetilde{W}_{t-1}^n = 1$) when $\widetilde{W}_{t-1}^n \propto W_{t-1}^n \sqrt{\text{Var}_{M_t}(X_{t-1}^n, G_t)}$, hence the following function minimises the second term:

$$\eta_{t-1}(x_{t-1}) = \sqrt{\text{Var}_{M_t}(X_{t-1}^n, G_t)}.$$

□

# S4 Proofs of Appendix A

## S4.1 Proof of Lemma 2

*Proof of Lemma 2.* Without loss of generality we assume that $\mathcal{X} = (0,1)^d$ and take $h_{\mathcal{X},\psi} = h$. Let $(z^N)_{N\geq 1}$ and $\pi$ be as in the statement of the lemma and take $\epsilon > 0$, $a \in [0,1]$, and $\gamma > 0$ small enough so that the ball $B$ of centre $H(a)$ and radius $\gamma$ is such that $\pi(B) \leq \epsilon$. Since $H$ is continuous, there exists $\delta > 0$ such that $|b-a| \leq \delta$ implies that $\|H(b) - H(a)\| \leq \gamma$, and thus $H(b) \in B$. For any such $b$,

$$|F_{\pi_h}(b) - F_{\pi_h}(a)| \leq \pi(B) \leq \epsilon$$

and the result follows. □

## S4.2 Proof of Theorem 9

We prove Theorem 9 by a succession of lemmas. Lemma S4 shows the implication $(ii) \Rightarrow (iii)$, Lemma S5 shows the implication $(iii) \Rightarrow (ii)$ while Lemma S6 shows the remaining equivalences.

**Lemma S4.** *Let $\mathcal{X}$ be a cubifiable set, $\pi \in \widetilde{\mathcal{P}}_b(\mathcal{X})$, $\psi \in \mathcal{D}(\mathcal{X})$ be such that $\pi_\psi \in \mathcal{P}_b((0,1)^d)$ and $(\pi^N)_{N\geq 1}$ be a sequence in $\mathcal{P}(\mathcal{X})$ such that $\lim_{N\to+\infty} \|\pi^N - \pi\|_\star = 0$. Then,*

$$\lim_{N\to+\infty} \|\pi^N_{h_{\mathcal{X},\psi}} - \pi_{h_{\mathcal{X},\psi}}\|_\star = 0.$$

*Proof of Lemma S4.* Without loss of generality we assume that $\mathcal{X} = (0,1)^d$ and take $h_{\mathcal{X},\psi} = h$.

We first assume that

$$\pi^N_h(I^d_m(k)) = \pi^N(S^d_m(k)), \quad \text{for all } k \in 0 : 2^{md} - 1 \text{ and } m \geq 1 \text{ large enough.} \tag{S.36}$$

In this case, the result follows from Gerber and Chopin (2015, Theorem 3) but for sake of completeness the whole argument is presented below.

Let $I = [0,b]$, $b \in (0,1)$, and $m \in \mathbb{N}$ (which may depend on $N$) and assume first that $b \geq 2^{-dm}$, so that $I^d_m(0) \subseteq I$. Take $\tilde{I} = [0, k^*2^{-dm}]$, where $k^* \leq (2^{dm}-1)$ is the largest integer such that $k^*2^{-dm} \leq b$. Then

$$\begin{aligned}
\left|\pi^N_h(I) - \pi_h(I)\right| &\leq \left|F_{\pi^N_h}(k^*2^{-dm}) - F_{\pi_h}(k^*2^{-dm})\right| \\
&\quad + \left|\pi^N_h(I) - F_{\pi^N_h}(k^*2^{-dm}) - \left\{\pi_h(I) - F_{\pi_h}(k^*2^{-dm})\right\}\right| \\
&= \left|\pi^N(J) - \pi(J)\right| + \left|\pi^N_h\big((k^*2^{-dm}, b]\big) - \pi_h\big((k^*2^{-dm}, b]\big)\right|
\end{aligned} \tag{S.37}$$

with $J = H(\tilde{I})$. Note that the last equality holds by the definition of $\pi_h$ and by (S.36).

Next, since $\tilde{I}$ is the union of $k^*$ intervals in $\mathcal{I}^d_m$, $J$ is the union of $k^* \leq 2^{md}$ closed hypercubes in $\mathcal{S}^d_m$, and thus,

$$\left|\pi^N(J) - \pi(J)\right| \leq k^*\|\pi^N - \pi\|_{\mathrm{E}} \leq 2^{dm}\|\pi^N - \pi\|_{\mathrm{E}}.$$

For the second term of (S.37), we have under (S.36)

$$\begin{aligned}\left|\pi_h^N\left((k2^{-dm},b]\right)-\pi_h\left((k2^{-dm},b]\right)\right| &\le \pi_h^N\big(I_m^d(k)\big)+\pi_h\big(I_m^d(k)\big)\\ &\le \pi^N\big(S_m^d(k)\big)+\pi\big(S_m^d(k)\big)\\ &\le 2\pi\big(S_m^d(k)\big)+\|\pi^N-\pi\|_{\mathrm{E}}\\ &= \mathcal{O}\left(2^{-dm}\vee\|\pi^N-\pi\|_{\mathrm{E}}\right)\end{aligned}$$

where the last inequality comes from the fact that $\pi$ has a bounded bounded density with respect to $\lambda_d$.

In case $b<2^{-dm}$, similar computations show that

$$\left|\pi_h^N(I)-\pi_h(I)\right|\le\pi_h^N(I_m^d(0))+\pi_h(I_m^d(0))=\mathcal{O}\left(2^{-dm}\vee\|\pi^N-\pi\|_{\mathrm{E}}\right).$$

To conclude the proof under (S.36), remark that

$$\begin{aligned}\|\pi^N-\pi\|_{\mathrm{E}} &= \sup_{0\le a<b\le 1}\left|\pi^N([a,b))-\pi([a,b))\right|\\ &\le 2^d\sup_{0<b\le 1}\left|\pi^N([0,b))-\pi([0,b))\right|\\ &= 2^d\|\pi^N-\pi\|_\star\end{aligned}$$

where the two equalities are due to Lemma S2, first part, and the inequality to Niederreiter (1992, Proposition 2.4, p.15). Hence, under the assumptions of the lemma, $\|\pi^N-\pi\|_{\mathrm{E}}=o(1)$ and thus, choosing $m$ so that $2^{-dm}=\mathcal{O}(\|\pi^N-\pi\|_{\mathrm{E}}^{1/2})$ gives

$$\|\pi_h^N-\pi_h\|_\star=\mathcal{O}\left(\|\pi^N-\pi\|_{\mathrm{E}}^{1/2}\right)=o(1).$$

This shows the results under (S.36).

Assume now that (S.36) does not hold. To facilitate the presentation we assume below that $\pi^N(\mathrm{d}x)=\sum_{n=1}^N W^{n,N}\delta(x^{n,N})$ for a $z^N=(x^{n,N},W^{n,N})_{n=1}^N\in\mathcal{Z}^N$; that is, that $(\pi^N)_{N\ge1}$ is a sequence in $\mathcal{P}_f(\mathcal{X})$. Then, because $x\in\mathcal{H}_d$ if and only if $x$ has at least one dyadic coordinate, for any $\epsilon>0$ there exists a sequence $(\tilde z^N)_{N\ge1}$ such that (S.36) holds and such that, for all $N\ge1$,

$$\tilde z^N=(\tilde x^{n,N},W^{n,N}),\quad \max_{n\in1:N}\|x^{n,N}-\tilde x^{n,N}\|_\infty\le\epsilon.$$

Then, by Lemma S1, first part,

$$\left|\|\tilde\pi^N-\pi\|_\star-\|\pi^N-\pi\|_\star\right|\le c_\pi\,\epsilon$$

for a constant $c_\pi<+\infty$ which depends only on $\pi$.

Under the assumptions of the theorem we therefore have $\lim_{N\to+\infty}\|\tilde\pi^N-\pi\|_\star=0$ and thus, from above,

$$\lim_{N\to+\infty}\|\tilde\pi_h^N-\pi_h\|_\star=0. \tag{S.38}$$

To conclude the proof we choose the sequence $(\tilde{z}^N)_{N\geq 1}$ such that, for all $N \geq 1$,

$$\max_{n\in 1:N} |h(x^{n,N}) - h(\tilde{x}^{n,N})| \leq (\epsilon/C_d)^d$$

with $C_d < +\infty$ the Hölder constant of the Hilbert curve for the $\|\cdot\|_\infty$ norm. Such a sequence indeed exists because, since $H(h(x)) = x$ for all in $x \in [0,1]^d$ and by the Hölder property of the Hilbert curve,

$$\begin{aligned} \|x^{n,N} - \tilde{x}^{n,N}\|_\infty &= \|H(h(x^{n,N})) - H(h(\tilde{x}^{n,N}))\|_\infty \\ &\leq C_d |h(x^{n,N}) - h(\tilde{x}^{n,N})|^{1/d} \leq \epsilon. \end{aligned} \tag{S.39}$$

Then, by Lemma S1, second part,

$$\left| \|\tilde{\pi}_h^N - \pi_h\|_\star - \|\pi_h^N - \pi_h\|_\star \right| \leq c_\pi \, (\epsilon/C_d)^d$$

with $c_\pi + \infty$ depending only on $\pi$. Hence, since $\epsilon > 0$ is arbitrary and using (S.38),

$$\lim_{N\to+\infty} \|\pi_h^N - \pi_h\|_\star = 0.$$

This completes the proof of the lemma.

□

**Lemma S5.** *Let $\mathcal{X}$ be a cubifiable set, $\pi \in \widetilde{\mathcal{P}}_b(\mathcal{X})$, $\psi \in \mathcal{D}(\mathcal{X})$ be such that $\pi_\psi \in \mathcal{P}_b((0,1)^d)$ and $(\pi^N)_{N\geq 1}$ be a sequence in $\mathcal{P}(\mathcal{X})$ such that $\lim_{N\to+\infty} \|\pi^N_{h_{\mathcal{X},\psi}} - \pi_{h_{\mathcal{X},\psi}}\|_\star = 0$. Then,*

$$\lim_{N\to+\infty} \|\pi^N - \pi\|_\star = 0.$$

*Proof.* Without loss of generality we assume that $\mathcal{X} = (0,1)^d$ and take $h_{\mathcal{X},\psi} = h$.

We first assume (S.36). In this case, the result follows from similar computations as in Schretter et al. (2016, Theorem 1) but for sake of completeness the whole argument is reproduced below.

Let $m \geq 0$ be an arbitrary integer and $a \in [0,1)^d$ be such that $S_m^d(0) \subseteq B := [0,a]$. Let $\mathcal{S}_m^B = \{W \in \mathcal{S}_m^d : W \subseteq B\}$, $\tilde{B} = \cup\mathcal{S}_m^B$ and $\mathcal{D}_m^B = \{W \in \mathcal{S}_m^d : (B \setminus \tilde{B}) \cap W \neq \varnothing\}$. Then, let $\tilde{\mathcal{D}}_m^B$ be the set of $\#\mathcal{D}_m^B$ disjoint subsets of $[0,1]^d$ such that

1. $\forall \tilde{W} \in \tilde{\mathcal{D}}_m^B$, $\exists\, W \in \mathcal{D}_m^B \mid \tilde{W} \subseteq W$,
2. $\cup\tilde{\mathcal{D}}_m^B = \mathcal{D}_m^B$,
3. $\tilde{B} \cap \{\cup\tilde{\mathcal{D}}_m^B\} = \varnothing$.

Note that $\tilde{\mathcal{D}}_m^B$ is obtained by removing boundaries of the elements in $\mathcal{D}_m^B$ such that the above conditions 2 and 3 are satisfied. Then, we have

$$|\pi^N(B) - \pi(B)| \leq |\pi^N(\tilde{B}) - \pi(\tilde{B})| + \sum_{\tilde{W}\in\tilde{\mathcal{D}}_m^B} |\pi^N(\tilde{W} \cap B) - \pi(\tilde{W} \cap B)| \tag{S.40}$$

where, under (S.36),

$$|\pi^N(\tilde{B}) - \pi(\tilde{B})| = |\pi_h^N\big(h(\tilde{B})\big) - \pi_h(h(\tilde{B}))| \leq 2^{md}\|\pi_h^N - \pi_h\|_{\mathrm{E}}.$$

For the second term of (S.40), take $\tilde{W} \in \tilde{\mathcal{D}}_m^B$ and note that $\tilde{W} \subseteq S_m^d(k)$ for a $k \in \{0, \ldots, 2^{dm} - 1\}$. Then, with $p$ the bounded density of $\pi$ with respect to the Lebesgue measure,

$$\begin{aligned}
|\pi^N(\tilde{W} \cap B) - \pi(\tilde{W} \cap B)| &\leq \pi^N(S_m^d(k))) + \pi(S_m^d(k)) \\
&\leq \pi_h^N(I_m^d(k)) + \|p\|_\infty \lambda_d(S_m^d(k)) \\
&\leq 2\|p\|_\infty \lambda_d(S_m^d(k)) + \|\pi_h^N - \pi_h\|_{\mathrm{E}} \\
&= 2\|p\|_\infty 2^{-dm} + \|\pi_h^N - \pi_h\|_{\mathrm{E}}.
\end{aligned}$$

Thus,

$$\sum_{\tilde{W} \in \tilde{\mathcal{D}}_m^B} |\pi^N(\tilde{W} \cap B) - \pi\tilde{W} \cap B)| \leq 2\|p\|_\infty d\, 2^{-m} + d\, 2^{m(d-1)}\|\pi_h^N - \pi_h\|_{\mathrm{E}}$$

since $\#\tilde{\mathcal{D}}_m^B = \#\mathcal{D}_m^B \leq d\, 2^{m(d-1)}$ (Schretter et al., 2016).

Hence, for all $a \in [0,1)^d$ such that $S_m^d(0) \subseteq [0, a]$ we have

$$|\pi^N([0,a]) - \pi([0,a])| \leq 2\|p\|_\infty d\, 2^{-m} + \|\pi_h^N - \pi_h\|_{\mathrm{E}}\big(1 + 5d\, 2^{md}\big)\,.$$

Finally, if $a \in [0,1)^d$ is such that $S_m^d(0) \not\subseteq [0, a]$, we proceed exactly as above, but now $\tilde{B}$ is empty and therefore the first term in (S.40) disappears.

To conclude the proof under (S.36) remark that

$$\begin{aligned}
\|\pi_h^N - \pi_h\|_{\mathrm{E}} &= \sup_{0 \leq a < b \leq 1} \Big|\pi_h^N([a,b)) - \pi_h([a,b))\Big| \\
&\leq 2 \sup_{0 < b \leq 1} \Big|\pi_h^N([0,b)) - \pi_h([0,b))\Big| \\
&= 2\|\pi_h^N - \pi_h\|_\star
\end{aligned}$$

where the two equalities are due to Lemma S2, second part, and the inequality to Niederreiter (1992, Proposition 2.4, p.15). Hence, $\|\pi_h^N - \pi_h\|_{\mathrm{E}} = o(1)$ under the assumptions of the lemma. We then choose $m$ such that $2^{-m} \sim 2^{md}\,\|\pi_h^N - \pi_h\|_{\mathrm{E}}$, which implies

$$\lim_{N \to +\infty} \|\pi^N - \pi\|_\star = 0$$

as required

If (S.36) does not hold the result follows using a similar argument as in the proof of lemma S4. To facilitate the presentation we assume below that $\pi^N(\mathrm{d}x) = \sum_{n=1}^N W^{n,N}\delta(x^{n,N})$ for a $z^N = (x^{n,N}, W^{n,N})_{n=1}^N \in \mathcal{Z}^N$; that is, that $(\pi^N)_{N \geq 1}$ is a sequence in $\mathcal{P}_f(\mathcal{X})$. Let $\epsilon > 0$ and choose a sequence $(\tilde{z}^N)_{N \geq 1}$ such that (S.36) holds and such that, for all $N \geq 1$,

$$\tilde{z}^N = (\tilde{x}^{n,N}, W^{n,N}), \quad \max_{n \in 1:N} \|x^{n,N} - \tilde{x}^{n,N}\|_\infty \leq \epsilon$$

and

$$\max_{n\in 1:N} |h(x^{n,N}) - h(\tilde{x}^{n,N})| \leq (\epsilon/C_d)^d$$

for a constant $C_d < \infty$; note that such a sequence $(\tilde{z}^N)_{N\geq 1}$ exists by (S.39). Then, by Lemme S1 and under the assumptions of the lemma, $\lim_{N\to+\infty} \|\tilde{\pi}_h^N - \pi_h\|_\star = 0$ and thus, from above, $\lim_{N\to+\infty} \|\tilde{\pi}^N - \pi\|_\star = 0$. Using again Lemma S1 we conclude that $\lim_{N\to+\infty} \|\pi^N - \pi\|_\star = 0$ and the proof is complete. □

**Lemma S6.** *Let $\mathcal{X}$ a cubifiable set, $\pi \in \widetilde{\mathcal{P}}_b(\mathcal{X})$, $\psi \in \mathcal{D}(\mathcal{X})$ be such that $\pi_\psi \in \mathcal{P}_b((0,1)^d)$ and $(\pi^N)_{N\geq 1}$ a sequence in $\mathcal{P}(\mathcal{X})$. Then,*

$$\pi^N \overset{w}{\Rightarrow} \pi \Leftrightarrow \lim_{N\to+\infty} \|\pi^N - \pi\|_\star = 0$$

*and*

$$\pi^N_{h_{\mathcal{X},\psi}} \overset{w}{\Rightarrow} \pi_{h_{\mathcal{X},\psi}} \Leftrightarrow \lim_{N\to+\infty} \|\pi^N_{h_{\mathcal{X},\psi}} - \pi_{h_{\mathcal{X},\psi}}\|_{\mathrm{E}} = 0$$

In order to prove Lemma S6 we need Lemma S7 below, which is a straightforward consequence of e.g. van der Vaart (1998, Lemma 2.2, p.6).

**Lemma S7.** *Let $\pi \in \mathcal{P}(\mathcal{X})$ and $(\pi^N)_{N\geq 1}$ be a sequence in $\mathcal{P}(\mathcal{X})$. Then, the following statements are equivalent:*

$$\lim_{N\to+\infty} |\pi^N(\varphi) - \pi(\varphi)| = 0, \qquad \forall \varphi \in \mathcal{C}_b(\mathcal{X}) \tag{a}$$

$$\lim_{N\to+\infty} |F_{\pi^N}(a) - F_\pi(a)| = 0, \qquad \forall a \in \mathcal{X} \tag{b}$$

*In addition, if $\pi \in \mathcal{P}_b(\mathcal{X})$, statements (a) and (b) are equivalent to*

$$\lim_{N\to+\infty} |\pi^N(B) - \pi(B)| = 0, \qquad \textit{for all Borel set } B \in \mathbb{X}. \tag{c}$$

*Proof of Lemma S6.* Without loss of generality we assume that $\mathcal{X} = (0,1)^d$ and take $h_{\mathcal{X},\psi} = h$. Let $\pi$ and $(\pi^N)_{N\geq 1}$ be as in the statement of the lemma and $p : \mathcal{X} \to \mathbb{R}_+$ be the continuous and bounded density of $\pi$.

The implications "⇐" are direct consequences of (b)⇒(a) in Lemma S7.

We first show the implication "⇒" for the first part of the lemma. To this end, we follow a similar argument as in Kuipers and Niederreiter (1974, Theorem 1.2, p.89). Note first that $\pi^N \overset{\mathrm{w}}{\Rightarrow} \pi$ implies (c) in Lemma S7, hence $\pi^N \overset{\mathrm{w}}{\Rightarrow} \pi$ implies that

$$|\pi^N\big([a,b]\big) - \pi\big([a,b]\big)| \to 0, \quad \forall [a,b] \subset (0,1)^d. \tag{S.41}$$

For a fixed $\epsilon > 0$, let $m_\epsilon \geq 2$ be the smallest positive integer such that

$$\left(\frac{2\|p\|_\infty}{m_\epsilon}\Big(2 + \frac{1}{m_\epsilon}\Big)\right) \vee \left(\frac{1}{m_\epsilon} + \|p\|_\infty\Big(\frac{2}{m_\epsilon}\Big)^d\Big(1 + \frac{1}{m_\epsilon}\Big)\right) \leq \epsilon \tag{S.42}$$

and let $B_k = \prod_{i=1}^d [k_i/m_\epsilon, (k_i+1)/m_\epsilon]$ with $k_i \in 0 : (m_\epsilon - 1)$ for all $i \in 1 : d$. Then, using (S.41), there exists a $N_\epsilon \geq 1$ such that, for all $N \geq N_\epsilon$ and $k \in 0 : (m_\epsilon - 1)$, we have

$$\pi(B_k)\Big(1 - \frac{1}{m_\epsilon}\Big) \leq \pi^N(B_k) \leq \pi(B_k)\Big(1 + \frac{1}{m_\epsilon}\Big). \tag{S.43}$$

Let $J = [a, b] \subset (0, 1)^d$. There exist hypercubes $J_1$ and $J_2$, which are (possibly empty) finite unions of hypercubes $B_k$, such that

$$J_1 \subseteq J \subseteq J_2, \quad \lambda_d(J - J_1) \leq (2/m_\epsilon)^d, \quad \lambda_d(J_2 - J) \leq (2/m_\epsilon)^d.$$

This implies that $\pi(J) - \pi(J_1) \leq \|p\|_\infty (2/m_\epsilon)^d$ and $\pi(J_2) - \pi(J) \leq \|p\|_\infty (2/m_\epsilon)^d$.

Assume first that that $J$ is such that $J_1 \neq \varnothing$. Then, we have for all $N \geq N_\epsilon$

$$\pi(J_1)\Big(1 - \frac{1}{m_\epsilon}\Big) \leq \pi^N(J_1) \leq \pi^N(J) \leq \pi^N(J_2) \leq \pi(J_2)\Big(1 + \frac{1}{m_\epsilon}\Big)$$

thus

$$\Big(\pi(J) - \|p\|_\infty \Big(\frac{2}{m_\epsilon}\Big)^d\Big)\Big(1 - \frac{1}{m_\epsilon}\Big) \leq \pi^N(J) \leq \Big(\pi(J) + \|p\|_\infty \Big(\frac{2}{m_\epsilon}\Big)^d\Big)\Big(1 + \frac{1}{m_\epsilon}\Big)$$

and since $\pi(J) \leq 1$

$$-\frac{1}{m_\epsilon} - \|p\|_\infty \Big(\frac{2}{m_\epsilon}\Big)^d \Big(1 + \frac{1}{m_\epsilon}\Big) \leq \pi^N(J) - \pi(J) \leq \frac{1}{m_\epsilon} + \|p\|_\infty \Big(\frac{2}{m_\epsilon}\Big)^d \Big(1 + \frac{1}{m_\epsilon}\Big). \tag{S.44}$$

Consider now the case $J_1 = \varnothing$ and define $J_2$ as above. Then, $\lambda_d(J_2) \leq 2/m_\epsilon$ and thus, for all $N \geq N_\epsilon$,

$$|\pi^N(J) - \pi(J)| \leq \pi^N(J_2) + \pi(J_2) \leq \pi(J_2)\Big(2 + \frac{1}{m_\epsilon}\Big) \leq \frac{2\|p\|_\infty}{m_\epsilon}\Big(2 + \frac{1}{m_\epsilon}\Big). \tag{S.45}$$

Therefore, combining (S.42), (S.44) and (S.45), we have that, for all $N \geq N_\epsilon$ and using (S.43),

$$\sup_{[a,b] \subset (0,1)} |\pi^N\big([a, b]\big) - \pi\big([a, b]\big)| \leq \epsilon,$$

which concludes the proof of the first part of the lemma.

The implication "$\Rightarrow$" in the second part of the lemma is due to the continuity of $F_{\pi_h}$ (Lemma 2) and to the Polyà's Theorem (Pólya, 1920; see also Bickel and Millar, 1992, result (A.1)). Alternatively, we can establish this implication following the same computation as per above. To do so, take $d = 1$ and $m_\epsilon = 2^{\tilde{m}_\epsilon}$ for some $\tilde{m}_\epsilon \geq 1$. Then, as $\pi_h$ is a continuous probability measure under the assumptions of the lemma (Lemma 2), $\pi_h(B_k) = \pi(S_{\tilde{m}_\epsilon}(k)) \leq \|p\|_\infty / m_\epsilon$ for all $k$ and, by part (c) of Lemma S7, when $\pi_h^N \overset{\text{w}}{\Rightarrow} \pi_h$ we have

$$|\pi_h^N\big([a, b]\big) - \pi_h\big([a, b]\big)| \to 0, \quad \forall [a, b] \subset (0, 1).$$

which replaces (S.41). □

## References


Aistleitner, C. and Dick, J. (2015). Functions of bounded variation, signed measures, and a general Koksma-Hlawja inequality. *Acta Arith.*, 167(2):143–171.

Bickel, P. and Millar, P. (1992). Uniform convergence of probability measures on classes of functions. *Statistica Sinica*, pages 1–15.

Chopin, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.*, 32(6):2385–2411.

Dubhashi, D., Jonasson, J., and Ranjan, D. (2007). Positive influence and negative dependence. *Combinatorics, Probability and Computing*, 16(01):29–41.

Dubhashi, D. and Ranjan, D. (1998). Balls and bins: a study in negative dependence. *Random Structures Algorithms*, 13(2):99–124.

Gerber, M. and Chopin, N. (2015). Sequential quasi Monte Carlo. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 77(3):509–579.

He, Z. and Owen, A. B. (2016). Extensible grids: uniform sampling on a space filling curve. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 78(4):917–931.

Kramer, J. B., Cutler, J., and Radcliffe, A. (2011). Negative dependence and Srinivasan's sampling process. *Combinatorics, Probability and Computing*, 20(03):347–361.

Kuipers, L. and Niederreiter, H. (1974). *Uniform distribution of sequences*. Wiley-Interscience.

Mytrofanov, M. A. and Ravsky, A. V. (2012). Approximation of continuous functions on Fréchet spaces. *Journal of Mathematical Sciences*, 185(6):792–799.

Niederreiter, H. (1992). *Random number generation and quasi-Monte Carlo methods*, volume 63 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Pólya, G. (1920). Über den zentralen grenzwertsatz der wahrscheinlichkeitsrechnung und das momentenproblem. *Mathematische Zeitschrift*, 8(3):171–181.

Schretter, C., He, Z., Gerber, M., Chopin, N., and Niederreiter, H. (2016). Van der Corput and golden ratio sequences along the Hilbert space-filling curve. In *Monte Carlo and Quasi-Monte Carlo Methods*, pages 531–544. Springer.

Shao, Q.-M. (2000). A comparison theorem on moment inequalities between negatively associated and independent random variables. *J. Theoret. Probab.*, 13(2):343–356.

van der Vaart, A. W. (1998). *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

Zumbusch, G. (2003). *Parallel multilevel methods*. Springer.