Sequential quasi Monte Carlo

Mathieu Gerber

Université de Lausanne, Switzerland, and Centre de Recherche en Economie et Statistique, Paris, France

and Nicolas Chopin

Centre de Recherche en Economie et Statistique and Ecole Nationale de la Statistique et de l’Administration Economique, Paris, France

[Read before The Royal Statistical Society at a meeting organized by the Research Section on Wednesday, December 10th, 2014, Professor C. Jennison in the Chair]

Summary. We derive and study sequential quasi Monte Carlo (SQMC), a class of algorithms obtained by introducing QMC point sets in particle filtering. SQMC is related to, and may be seen as an extension of, the array-RQMC algorithm of L’Ecuyer and his colleagues. The complexity of SQMC is $O(N \log(N))$, where $N$ is the number of simulations at each iteration, and its error rate is smaller than the Monte Carlo rate $O_P(N^{-1/2})$. The only requirement to implement SQMC algorithms is the ability to write the simulation of particle $x^n_t$ given $x^n_{t-1}$ as a deterministic function of $x^n_{t-1}$ and a fixed number of uniform variates. We show that SQMC is amenable to the same extensions as standard SMC, such as forward smoothing, backward smoothing and unbiased likelihood evaluation. In particular, SQMC may replace SMC within a particle Markov chain Monte Carlo algorithm. We establish several convergence results. We provide numerical evidence that SQMC may significantly outperform SMC in practical scenarios.

Keywords: Array-randomized quasi Monte Carlo; Low discrepancy; Particle filtering; Quasi Monte Carlo; Randomized quasi Monte Carlo; Sequential Monte Carlo

1. Introduction

Sequential Monte Carlo (SMC) (also known as particle filtering) is a class of algorithms for computing recursively Monte Carlo approximations of a sequence of distributions $π_t(dx_t)$, $t ∈ \{0, T\}$. The initial motivation of SMC was the filtering of state space models (also known as hidden Markov models), i.e. given a latent Markov process $(x_t)$, observed imperfectly as for example $y_t = f(x_t) + ε_t$, recover at every time $t$ the distribution of $x_t$ given the data $y_0:t = (y_0, ..., y_t)$. The method’s popularity stems from the fact that it is the only realistic approach for filtering and related problems outside very specific cases (such as the linear Gaussian model). Recent research has further increased interest in SMC especially in statistics, in at least two directions. First, several papers (Neal, 2001; Chopin, 2002; Del Moral et al., 2006) have extended it to non-sequential problems, i.e., to sample from distribution $π$, one applies SMC to some artificial sequence $π_t$ that ends up at $π_T = π$. In certain cases, such an approach outperforms Markov chain Monte Carlo (MCMC) methods significantly. Second, the seminal paper of Andrieu et al. (2010) established that SMC may be used as a proposal mechanism within MCMC methods, leading to so-called particle MCMC (PMCMC) algorithms. Although not

Address for correspondence: Nicolas Chopin, Ecole Nationale de la Statistique et de l’Administration Economique, 3 avenue Pierre Larousse, 92245 Malakoff, Paris, France.
E-mail: nicolas.chopin@ensae.fr

© 2015 Royal Statistical Society
restricted to such problems, the PMCMC approach is the only possible approach for inference in state space models such that the transition kernel of \((x_t)\) may be sampled from, but does not admit a tractable density. Excitement about PMCMC is evidenced by the 30 papers or so that have appeared in the last 2 years on possible applications and extensions.

Informally, the error rate of SMC at iteration \(t\) is \(C_t N^{-1/2}\), where \(C_t\) is some function of \(t\). There has been much work on SMC error rates (e.g. Del Moral (2004)), but it seems fair to say that most of it has focused on the first factor \(C_t\), i.e. whether to establish that the error rate is bounded uniformly in time, \(C_t \leq C\), or to reduce \(C_t\) through more efficient algorithmic designs, such as better proposal kernels or resampling schemes.

In this work, we focus on the second factor \(N^{-1/2}\), i.e. we want the error rate to converge quicker relative to \(N\) than the standard Monte Carlo rate \(N^{-1/2}\). To do so, we adapt to the SMC context ideas borrowed from quasi Monte Carlo (QMC) methods, i.e. the idea of replacing random numbers by low discrepancy point sets.

The following subsections contain very brief introductions to SMC and QMC methods, with an exclusive focus on the concepts that are essential to follow this work. For a more extensive presentation of SMC sampling, the reader is referred to Doucet et al. (2001), Del Moral (2004) and Cappé et al. (2005), whereas for QMC and randomized QMC (RQMC) methods see chapter 5 of Glasserman (2004), chapters 5 and 6 of Lemieux (2009), and Dick and Pillichshammer (2010).

### 1.1. Introduction to sequential Monte Carlo

As already mentioned, the initial motivation of SMC is the sequential analysis of state space models, i.e. models for a Markov chain \((x_t)\) in \(\mathcal{X} \subseteq \mathbb{R}^d\),

\[
\begin{align*}
x_0 & \sim f_X^0(x_0), \\
x_t | x_{t-1} & \sim f^X(x_t | x_{t-1}),
\end{align*}
\]

which is observed only indirectly through some \(y_t\), with density \(y_t | x_t \sim f^Y(y_t | x_t)\).

This kind of model arises in many areas of science: in tracking, for instance, \(x_t\) may be the position of a ship (in two dimensions) or a plane (in three dimensions), and \(y_t\) may be a noisy angular observation (radar). In ecology, \(x_t\) would be the size of a population of bats in a cave, and \(y_t\) would be \(x_t\) plus noise, and so on.

The most standard inferential task for such models is that of filtering, i.e. to recover iteratively in time \(t\), \(p(x_t | y_{0:t})\), the distribution of \(x_t\), given the data collected up to time \(t\), \(y_{0:t} = (y_0, \ldots, y_t)\). One may also be interested in smoothing, \(p(x_{0:t} | y_{0:t})\), or likelihood evaluation, \(p(y_{0:t})\), notably when the model depends on a fixed parameter \(\theta\) which should be learnt from the data.

A simple Monte Carlo approach to filtering is sequential importance sampling: choose an initial distribution \(m_0(dx_0)\) and a sequence of Markov kernels \(m_t(x_{t-1}, dx_t)\), \(t \geq 1\). Then simulate \(N\) times iteratively from these \(m_t\), \(x_0^n \sim m_0(dx_0), x_1^n | x_0^n \sim m_1(x_1^n | x_0^n, dx_1), \ldots, x_t^n | x_{t-1}^n \sim m_t(x_t^n | x_{t-1}^n, dx_t)\), and reweight ‘particle’ (simulation) \(x_t^n\) as follows: \(w_0^n = G_0(x_0^n), w_t^n = w_{t-1}^n G_t(x_t^n | x_{t-1}^n)\), where the weight functions \(G_t\) are defined as

\[
\begin{align*}
G_0(x_0) &= \frac{f^Y(y_0 | x_0) f_X^0(x_0)}{m_0(x_0)}, \\
G_t(x_{t-1}, x_t) &= \frac{f^Y(y_t | x_t) f^X(x_t | x_{t-1})}{m_t(x_t | x_{t-1})},
\end{align*}
\]

and \(m_t(x_t | x_{t-1})\) in the denominator denotes the conditional probability density that is associated
Table 1. Algorithm 1: basic particle filter

| Time t = 0: |
|-------------|
| (a) generate \( x_0^n \sim m_0(dx_0) \) for all \( n \in 1:1 \); |
| (b) compute \( w_0^n = G_0(x_0^n) \) and \( W_0^n = w_0^n / \sum_{m=1}^{N} w_0^m \) for all \( n \in 1:1 \); |

From time \( t = 1 \) to time \( T \):

| Time t = 1 to T: |
|----------------|
| (a) generate \( a_0^n - \mathcal{M}(W_{1:N}^0) \) for all \( n \in 1:1 \), the multinomial distribution that produces outcome \( m \) with probability \( W_m^{t-1} \) (see algorithm 2 in Table 2); |
| (b) generate \( x_n^t \sim m_t(x_{t-1}^{a_{t-1}^n}, dx_t) \) for all \( n \in 1:1 \); |
| (c) compute \( w_n^t = G_t(x_n^t, x_{t-1}^{a_{t-1}^n}) \), and \( W_n^t = w_n^t / \sum_{m=1}^{N} w_t^m \) for all \( n \in 1:1 \); |

with kernel \( m_t(x_{t-1}, dx_t) \). Then it is easy to check that the weighted average

\[
\frac{\sum_{n=1}^{N} w_n^t \varphi(x_n^t)}{\sum_{m=1}^{N} w_t^m}
\]

is a consistent estimate of the filtering expectation \( \mathbb{E}[\varphi(x_t)|y_{0:t}] \), as \( N \to \infty \). However, it is well known that, even for carefully chosen proposal densities \( m_t \), sequential importance sampling quickly degenerates: as time progresses, increasingly more particles receive a negligible weight.

Surprisingly, there is a simple solution to this degeneracy problem: one may resample the particles, i.e. draw \( N \) times with replacement from the set of particles, with probabilities proportional to the weights \( w_t^0 \). In this way, particles with low weight become quickly discarded, whereas particles with large weight may have many children at the following iteration. Empirically, the effect of resampling is dramatic: the variance of filtering estimates typically remains stable over time, whereas without resampling it diverges exponentially fast.

The complexity of SMC is \( O(N) \). In particular, to implement the resampling step in \( O(N) \) time (step (a) at times \( t \geq 1 \) in algorithm 1), proceed as follows:

(a) generate \( u^{1:N} = \text{sort}(u^{1:N}) \), where the \( u^n \) are independent uniform variates (see page 214 of Devroye (1986) for a well-known algorithm to generate \( u^{1:N} \) directly in \( O(N) \) time, without any sorting);
(b) use the inverse transform method for discrete distributions, which is outlined in algorithm 2 in Table 2.

We shall reuse algorithm 2 in the sequential QMC (SQMC) method.

1.2. Introduction to quasi Monte Carlo

QMC is generally presented as a way to perform integration with respect to the (semiclosed) hypercube of dimension \( d \):
Table 2. Algorithm 2: resampling algorithm (inverse transform method)

| Input: u₁⁻¹ : N (such that 0 ≤ u₁ ≤ ... ≤ u N ≤ 1), W₁ : N (normalized weights) |
|-----------------------------------------------|
| Output: a₁⁻¹ : N (labels in 1 : N)             |
| s ← W₁, m ← 1                                |
| for n = 1 → N do                             |
|   while s < u m do                            |
|       m ← m + 1                              |
|       s ← s + W m                            |
|   end while                                   |
| a m ← m                                      |
| end for                                       |

\[
\frac{1}{N} \sum_{n=1}^{N} \varphi(u^n) \approx \int_{[0,1]^d} \varphi(u) \, du
\]

where the N vectors \( u^n \in [0, 1]^d \) must be chosen to have ‘low discrepancy’, i.e., informally, to be spread evenly over \([0, 1]^d\). (We respect the standard convention in the QMC literature to work with space \([0, 1]^d\), rather than \([0, 1]^d\), as it turns out to be technically more convenient.)

Formally, the general notion of discrepancy is defined as

\[
D(u^{1:N}; A) = \sup_{A \in \mathcal{A}} \left| \frac{1}{N} \sum_{n=1}^{N} 1(u^n \in A) - \lambda_d(A) \right|
\]

where \( \lambda_d(A) \) is the volume (Lebesgue measure on \( \mathbb{R}^d \)) of \( A \), and \( \mathcal{A} \) is a set of measurable sets. Two discrepancies are particularly useful in this work: the extreme discrepancy,

\[
D(u^{1:N}) = \sup_{[a, b]} \left| \frac{1}{N} \sum_{n=1}^{N} 1(u^n \in [a, b]) - \prod_{i=1}^{d} (b_i - a_i) \right|
\]

which is the discrepancy relative to the set \( A \) of \( d \)-dimensional intervals \( [a, b] := \Pi_{i=1}^{d} [a_i, b_i] \), \( 0 \leq a_i < b_i < 1 \), and the star discrepancy,

\[
D^*(u^{1:N}) = \sup_{[0, b]} \left| \frac{1}{N} \sum_{n=1}^{N} 1(u^n \in [0, b]) - \prod_{i=1}^{d} b_i \right|
\]

where again \( [0, b] = \Pi_{i=1}^{d} [0, b_i] \), \( 0 < b_i < 1 \). When \( d = 1 \), the star discrepancy is the Kolmogorov–Smirnov statistic for a uniformity test of the points \( u^n \).

These two discrepancies are related as follows (Niederreiter (1992), proposition 2.4):

\[
D^*(u^{1:N}) \leq D(u^{1:N}) \leq 2^d D^*(u^{1:N})
\]

The importance of the concept of discrepancy, and in particular of the star discrepancy, is highlighted by the Koksma–Hlawka inequality (see for example Kuipers and Niederreiter (1974) theorem 5.1):

\[
\left| \frac{1}{N} \sum_{n=1}^{N} \varphi(u^n) - \int_{[0,1]^d} \varphi(u) \, du \right| \leq V(\varphi) D^*(u^{1:N})
\]

which conveniently separates the effect of the smoothness of \( \varphi \) (as measured by \( V(\varphi) \), the total variation in the sense of Hardy and Krause; see chapter 2 of Niederreiter (1992) for a definition)
Quasi Monte Carlo

Fig. 1. QMC versus Monte Carlo sampling: $N = 256$ points sampled (a) independently and uniformly in $[0,1]^2$ and (b) QMC sequence (Sobol') in $[0,1]^2$ of the same length

and the effect of the discrepancy of the points $u^{1:N}$. The quantity $V(\varphi)$ is generally too difficult to compute in practice, and the Koksma–Hlawka inequality is used mainly to determine the asymptotic error rate (as $N \to \infty$), through the quantity $D^*(u^{1:N})$.

There are several methods to construct $u^{1:N}$ so that $D^*(u^{1:N}) = O(N^{-1+\varepsilon})$ for any $\varepsilon > 0$, which is of course better than the Monte Carlo rate $O_P(N^{-1/2})$. The best known rates are $O\{N^{-1}\log(N)^{d-1}\}$ for QMC point sets $u^{N,1:N}$ that are allowed to depend on $N$ (i.e. $u^{N,1:N}$ are not necessarily the $N$ first elements of $u^{N+1,1:N+1}$) and $O\{N^{-1}\log(N)^d\}$ for QMC sequences (i.e. $u^{1:N}$ are the first $N$ elements of a sequence $u^N$ which may be generated iteratively). For simplicity, we shall not distinguish further QMC point sets and QMC sequences, and we shall use the same notation $u^{1:N}$ in both cases (although our results will apply to both types of construction).

These asymptotic rates seem to indicate that the comparative performance of QMC over Monte Carlo sampling should deteriorate with $d$: for $d = 10$, $N^{-1}\log(N)^d \leq N^{-1/2}$ only for $N \geq 1.3 \times 10^{39}$. But, since these rates correspond to an upper bound for the error size, it is difficult to determine beforehand if and when QMC ‘breaks’ with the dimension. For instance, Glasserman (2004), page 327, gave a numerical example where QMC remains competitive relative to Monte Carlo sampling for $d \geq 150$ and $N \leq 10^5$.

Describing the different strategies to construct low discrepancy point sets is beyond the scope of this paper; see again Glasserman (2004), Lemieux (2009) and Dick and Pillichshammer (2010). Fig. 1 illustrates the greater regularity of a QMC point set over a set of random points.

1.3. Introduction to randomized quasi Monte Carlo

RQMC amounts to randomizing the points $u^{1:N}$ in such a way that

(a) they still have low discrepancy (with probability 1), and
(b) each $u^i \sim U([0,1]^d)$ marginally.

The simplest construction of such RQMC point sets is the random-shift method that was
proposed by Cranley and Patterson (1976) in which we take $u^n = v^n + w \mod 1$, where $w \sim \mathcal{U}([0,1]^d)$ and $v^{1:N}$ is a low discrepancy point set.

RQMC has two advantages over QMC. First, one then obtains an unbiased estimator of the integral of interest:

$$\mathbb{E}\left[ \frac{1}{N} \sum_{n=1}^{N} \varphi(u^n) \right] = \int_{[0,1]^d} \varphi(u) \, du,$$

which makes it possible to evaluate the approximation error through independent replications. We shall see that, in our context, this unbiasedness property will also be very convenient for another reason: to provide an unbiased estimate of the likelihood of the state space model considered.

Second, Owen (1997a, b, 1998) established that randomization may lead to better rates, in the following sense: under appropriate conditions, and for a certain type of randomization scheme known as nested scrambling, the mean-square error (MSE) of an RQMC estimator is $O(N^{-3+\varepsilon})$. The intuition behind this rather striking result is that randomization may lead to cancellation of certain error terms.

1.4. A note on array-randomized quasi Monte Carlo

Consider the following problem: we have a Markov chain in $\mathcal{X}$, whose evolution may be formulated as

$$x_t = \Gamma_t(x_{t-1}, u_t), \quad u_t \sim \mathcal{U}([0,1]^d), \quad t \geq 1, \quad x_0 \text{ is fixed},$$

and we wish to compute the expectation of $\Sigma_{t=1}^{T} \varphi_t(x_t)$, for certain functions $\varphi_t$.

From the two previous sections, we see that a simple approach to this problem would be to generate a QMC (or RQMC) point set $u^{1:N}$ in $[0,1]^d$, to transform $u'_n$ into $x'_n = \Gamma_t(x'_{n-1}, u'_n)$, and finally to return the corresponding empirical average $N^{-1} \Sigma_{t=1}^{T} \varphi_t(x'_n)$. The problem with this direct approach is that the dimension $dT$ of $u^{1:N}$ may be very large and, as we have seen, equidistribution properties of QMC point sets (as measured by the star discrepancy) deteriorate with the dimension.

An elegant alternative to this approach is the array-RQMC algorithm of L’Ecuyer et al. (2006); see also Lécot and Ogawa (2002), Lécot and Tuffin (2004) and L’Ecuyer et al. (2009). The main idea of this method is to replace the QMC point set in $[0,1]^d$ by $T$ QMC point sets $u^{1:N}_t$ in $[0,1]^d$. Then, $x^n_t$ is obtained as $x^n_t = \Gamma_t(x^n_{t-1}, u^n_t)$, where the ancestor $x^n_{t-1}$ of $x^n_t$ is chosen to be the $n$th ‘smallest’ point among the $x^n_{t-1}$s. Note that array-RQMC sampling therefore requires us to specify a total order for the state space $\mathcal{X}$; for instance one may define a certain $\omega : \mathcal{X} \rightarrow \mathbb{R}$ so that $\omega(x) \leq \omega(x')$ means that $x$ is ‘smaller’ than $x'$.

Array-RQMC was shown to have excellent empirical performance in L’Ecuyer et al. (2006, 2009), Lécot and Ogawa (2002) and Lécot and Tuffin (2004). However, it is currently lacking in terms of supporting theory (but see L’Ecuyer et al. (2008), for $d = 1$); in particular, it is not clear how to choose the order $\omega$, beside the obvious case where $\mathcal{X} \subset \mathbb{R}$. The SQMC algorithm that we develop in this paper may be seen as an extension of array-RQMC to particle filtering. In particular, it reuses the essential idea of generating one QMC point set at each step of the simulation process. As an added benefit, the convergence results that we obtain for the SQMC algorithm also apply to array-RQMC, provided that the state space is ordered through the Hilbert curve, as explained later.
1.5. Background, plan and notation

QMC is already very popular in finance for, for example, derivative pricing (Glasserman, 2004). However, and we may wonder why, it has not received more attention in statistics so far. The main reason seems to be the perceived difficulty of adapting it to non-independent simulation such as MCMC sampling; see, however, Chen et al. (2011) and references therein, in particular Tribble (2007), for exciting numerical and theoretical results in this direction which ought to change this perception.

Regarding SMC we are aware of two previous attempts to develop QMC versions of these algorithms: Ormoneit et al. (2001) and Fearnhead (2005); see also Guo and Wang (2006) who essentially proposed the same algorithm as Fearnhead (2005). Ormoneit et al. (2001) cast the SMC algorithm as a Monte Carlo algorithm in \( d(T + 1) \) dimensions, where \( d = \text{dim}(\mathcal{X}) \), and therefore requires the generation of a low discrepancy point set in \([0, 1)^{d(T+1)}\). But, as we have already explained, such an approach may not work well when \( d(T+1) \) is too large.

Our approach is closer to, and partly inspired by, the regularized particle filter of Fearnhead (2005), who, in the same spirit as array-RQMC, cast SMC as a sequence of \( T + 1 \) successive importance sampling steps of dimension \( d \). (He focused on the \( d = 1 \) case.) The main limitation of the regularized particle filter is that it has complexity \( O(N^2) \). This is because the importance sampling steps are defined with respect to a target which is a mixture of \( N \) components; hence the evaluation of a single importance weight costs \( O(N) \).

The SQMC algorithm that we develop in this paper has complexity \( O\{N \log(N)\} \) per time step. It is also based on a sequence of \( T + 1 \) importance sampling steps, but of dimension \( d + 1 \); the first component is used to determine which ancestor \( x_{m_t} \) should be assigned to particle \( x_n \). For \( d > 1 \), this requires us to ‘project’ the set of ancestors \( x_{1:t-1}^N \in \mathcal{X}^N \) into \([0, 1)^N\), by means of a space filling curve known as the Hilbert curve. The choice of this particular space filling curve is not only for computational convenience, but also because of its nice properties regarding conversion of discrepancy, as we shall explain. (One referee pointed out to us that the use of Hilbert curves in the context of array-RQMC has been suggested by Wächter and Keller (2008) but not implemented.)

The paper is organized as follows. Section 2 derives the general SQMC algorithm; first for \( d = 1 \); then for any \( d \) through the use of the Hilbert curve. Section 3 presents several convergence results; proofs of these results are in Appendix A. Section 4 shows how several standard extensions of the SMC method, such as forward smoothing, backward smoothing and PMCMC sampling, may be adapted to SQMC. Section 5 compares numerically SQMC with SMC methods. Section 6 concludes.

Most random variables in this work will be vectors in \( \mathbb{R}^d \), and will be denoted in bold, \( \mathbf{u} \) or \( \mathbf{x} \). In particular, \( \mathcal{X} \) will be an open set of \( \mathbb{R}^d \). The Lebesgue measure in dimension \( d \) is denoted by \( \lambda_d \). Let \( \mathcal{P}(\mathcal{X}) \) be the set of probability measures defined on \( \mathcal{X} \) dominated by \( \lambda_d \) (restricted to \( \mathcal{X} \)), and \( \pi(\varphi) \) be the expectation of function \( \varphi \) relative to \( \pi \in \mathcal{P}(\mathcal{X}) \). Let \( a:b \) be the set of integers \( \{a, \ldots, b\} \) for \( a \leq b \). We also use this notation for collections of random variables, e.g. \( x_{1:t}^{1:N} = (x_t^1, \ldots, x_t^N) \), \( x_{0:t} = (x_0, \ldots, x_t) \) and so on.

2. Sequential quasi Monte Carlo

The objective of this section is to construct the SQMC algorithm. For this, we discuss how to rewrite the SMC algorithm as a deterministic function of independent uniform variates \( u_{t}^{1:N} \), \( t = 0:T \), which then may be replaced by low discrepancy point sets.

2.1. Sequential Monte Carlo formalization

A closer inspection of our basic particle filter, algorithm 1, reveals that this algorithm is entirely...
determined by

(a) the sequence of proposal kernels \((m_t)_{t \geq 0}\) (which determine how particles are simulated) and

(b) the sequence of weight functions \((G_t)_{t \geq 0}\) (which determine how particles are weighted).

Our introduction to particle filtering focused on the specific expression (1) for \(G_t\), but useful SMC algorithms may be obtained by considering other weight functions; see for example the auxiliary particle filter of Pitt and Shephard (1999), as explained in Johansen and Doucet (2008), or the SMC algorithms for non-sequential problems that were mentioned in Section 1.

The exact expression and meaning of \(m_t\) and \(G_t\) will not play a particular role in the rest of the paper, so it is best to think of the SMC algorithm from now on as a generic algorithm, again based on a certain sequence \((m_t), m_0(dx_0)\) being an initial distribution, and \(m_t(x_{t-1}, dx_t)\) being a Markov kernel for \(t \geq 1\), and a certain sequence of functions, \(G_0 : \mathcal{X} \rightarrow \mathbb{R}^+, G_t : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}^+\), which produces the following consistent (as \(N \rightarrow \infty\)) estimators:

\[
\frac{1}{N} \sum_{n=1}^{N} \varphi(x^n_t) \rightarrow \tilde{Q}_t(\varphi),
\]

\[
\sum_{n=1}^{N} W^n_t \varphi(x^n_t) \rightarrow Q_t(\varphi).
\]

Here \(\varphi : \mathcal{X} \rightarrow \mathbb{R}\), and \(\tilde{Q}_t\) and \(Q_t\) are defined as follows:

\[
Z_t = \mathbb{E}\left[ G_0(x_0) \prod_{s=1}^{t} G_s(x_{s-1}, x_s) \right],
\]

\[
\tilde{Q}_t(\varphi) = \frac{1}{Z_{t-1}} \mathbb{E}\left[ \varphi(x_t) G_0(x_0) \prod_{s=1}^{t-1} G_s(x_{s-1}, x_s) \right],
\]

\[
Q_t(\varphi) = \frac{1}{Z_t} \mathbb{E}\left[ \varphi(x_t) G_0(x_0) \prod_{s=1}^{t} G_s(x_{s-1}, x_s) \right],
\]

with expectations taken with respect to the law of the non-homogeneous Markov chain \((x_t)\), e.g.

\[
Z_t = \int_{\mathcal{X}^{t+1}} \left\{ G_0(x_0) \prod_{s=1}^{t} G_s(x_{s-1}, x_s) \right\} m_0(dx_0) \prod_{s=1}^{t} m_s(x_{s-1}, dx_s),
\]

and with the conventions that \(Z_{-1} = 1\) and empty products equal 1, e.g. \(\tilde{Q}_0(\varphi) = m_0(\varphi)\).

For instance, for the standard filtering problem that was covered in Section 1, where \(G_t\) is set to expression (1), \(Q_t(\varphi)\) is the filtering expectation of \(\varphi\), i.e. \(\mathbb{E}[\varphi(x_t) | y_{0:t}]\), and \(\tilde{Q}_t(\varphi)\) is the predictive distribution of \(\varphi\), i.e. \(\mathbb{E}[\varphi(x_t) | y_{0:t-1}]\).

### 2.2. Towards sequential quasi Monte Carlo: sequential Monte Carlo as a sequence of importance sampling steps

The SQMC algorithm requires us to write any simulation as an explicit function of uniform variates. We therefore make the following assumption for our generic SMC sampler: to generate \(x^n_0 \sim m_0(dx_0)\), we compute \(x^n_0 = \Gamma_0(u^n_0)\) and, to generate \(x^n_t | x^n_{t-1}, dx_t\), we compute \(x^n_t = \Gamma_t(x^n_{t-1}, u^n_t)\), where \(u^n_t \sim \mathcal{U}([0,1]^d)\), and the functions \(\Gamma_t\) are easy to evaluate.

Initialization of algorithm 1 amounts to an importance sampling step, from \(m_0(dx_0)\) to
\( Q_0(dx_0) = m_0(dx_0) G_0(x_0)/Z_0, \) which produces the estimator

\[
\sum_{n=1}^{N} W^n_0 \varphi(x^n_0) = \frac{\sum_{n=1}^{N} G_0(x^n_0) \varphi(x^n_0)}{\sum_{m=1}^{N} G_0(x^m_0)}
\]
of \( Q_0(\varphi) \). To introduce QMC at this stage, we take \( x^n_0 = \Gamma_0(u^n_0) \) where \( u^n_0 \) is a low discrepancy point set in \([0, 1]^d\).

The key remark that underpins the SQMC algorithm is that iteration \( t \geq 1 \) of algorithm 1 also amounts to an importance sampling step, but this time from

\[
\tilde{Q}^N_t \{ d(\tilde{x}_{t-1}, x_t) \} = \frac{1}{Q^N_t} \tilde{Q}^N_t \{ d(\tilde{x}_{t-1}, x_t) \} G_t(\tilde{x}_{t-1}, x_t)
\]

where \( Q^N_t \) and \( \tilde{Q}^N_t \) are two random probability measures defined over \( \mathcal{X} \times \mathcal{X} \), a set of dimension \( 2d \). In particular, the generation of random variables \( a^n_{t-1}^1 \) and \( x^n_{t-1} \) in steps (a) and (b) of algorithm 1 is equivalent to sampling \( N \) times independently random variables \( (\tilde{x}^n_{t-1}, x^n_t) \) from \( Q^N_t \{ d(\tilde{x}_{t-1}, x_t) \} \), i.e. \( x^n_{t-1} = \tilde{x}^n_{t-1} \) (not to be mistaken for \( x^n_{t-1} \)), and \( x^n_t \sim m_t(\tilde{x}_{t-1}, dx_t) \).

On the basis of these remarks, the general idea behind SQMC is to replace at iteration \( t \) the \( N \) independent and identically distributed random numbers sampled from \( Q^N_t \{ d(\tilde{x}_{t-1}, x_t) \} \) by a low discrepancy point set relative to the same distribution.

When \( d = 1 \), this idea may be implemented as follows: generate a low discrepancy point set \( u^n_{1:N} \) in \([0, 1]^2\); let \( u^n_0 = (u^n_0, v^n_0) \); then set \( \tilde{x}^n_{t-1} = F^{-1}_N(u^n_0), \ x^n_t = \Gamma_t(x^n_{t-1}, v^n_t) \), where \( F^{-1}_N \) is the generalized inverse of the empirical cumulative distribution function (CDF)

\[
\hat{F}_N(x) = \sum_{n=1}^{N} W^n_{t-1} \delta\{x^n_{t-1} \leq x\}, \quad x \in \mathcal{X} \subset \mathbb{R}.
\]

It is easy to see that the most efficient way to compute \( \tilde{x}^n_{t-1} = F^{-1}_N(u^n_0) \) for all \( n \in 1:N \) is

(a) to sort the \( x^n_{t-1} \)'s, i.e. to find permutation \( \sigma \) such that \( x^{\sigma(1)}_{t-1} \leq \ldots \leq x^{\sigma(N)}_{t-1} \),
(b) to sort the \( u^n_{1:N} \) (call sort(\( u^n_{1:N} \)) the corresponding result),
(c) to obtain \( a^n_{t-1} \) as the output of algorithm 2, with inputs \( \text{sort}(u^n_{1:N}) \) and \( W^{\sigma(1:N)}_t \), and finally
(d) to set \( \tilde{x}^n_{t-1} = x^{\sigma(n)}_{t-1} \).

Algorithm 3 in Table 3 gives a pseudocode version of the SQMC algorithm for any \( d \), but note how step (b) at times \( t \geq 1 \) simplifies to what we have just described for \( d = 1 \).

When \( d > 1 \), the inverse transform method cannot be used to sample from the marginal distribution of \( \tilde{x}_{t-1} \) relative to \( Q^N_t \{ d(\tilde{x}_{t-1}, x_t) \} \), at least unless the \( x^n_{t-1} \) are ‘projected’ to the real line in some sense. This is the point of the Hilbert curve that is presented in the next section.

2.3. Hilbert space filling curve

The Hilbert curve is a continuous fractal map \( H : [0, 1] \to [0, 1]^d \), which ‘fills’ entirely \([0, 1]^d\). \( H \) is obtained as the limit of a sequence \((H_m), m \to \infty\), the first terms of which are depicted in Fig. 2.
At time $t = 0$:

(a) generate a QMC or an RQMC point set $u_{0;N}^{1}$ in $[0, 1)^d$, and compute $x_{0}^d = H_0(u_{0}^{1})$ for each $n \in 1:N$;
(b) compute $w_{0}^{n} = G_0(x_{0}^{n})$ and $W_{0}^{n} = w_{0}^{n}/\Sigma_{m=1}^{n}w_{0}^{m}$ for each $n \in 1:N$;

Iteratively, from time $t = 1$ to time $t = T$:

(a) generate a QMC or an RQMC point set $u_{t;N}^{1}$ in $[0, 1)^{d+1}$; let $u_{t}^{n} = (u_{t}^{n}, x_{t}^{n}) \in [0, 1) \times [0, 1)^d$;
(b) Hilbert sort—find permutation $\sigma_{t-1}$ such that $h \circ \psi(x_{t-1}^{\sigma_{t-1}(N)}) \leq \ldots \leq h \circ \psi(x_{t-1}^{\sigma_{t-1}(1)})$ if $d \geq 2$, or $x_{t-1}^{\sigma_{t-1}(1)} \leq \ldots \leq x_{t-1}^{\sigma_{t-1}(N)}$ if $d = 1$;
(c) find permutation $\tau$ such that $u_{t}^{\tau(1)} \leq \ldots \leq u_{t}^{\tau(N)}$, generate $u_{t;N}^{1}$ by using algorithm 2, with inputs $u_{t}^{\tau}, w_{t-1}^{n}$, and compute $x_{t}^{n} = H_{\tau}^{\sigma_{t-1}(u_{t}^{\tau(n)})}$ for each $n \in 1:N$;
(d) compute $w_{t}^{n} = G_{\tau}(x_{t-1}^{\sigma_{t-1}(u_{t}^{\tau(n)}))}$, and $W_{t}^{n} = w_{t}^{n}/\Sigma_{m=1}^{n}w_{t}^{m}$ for each $n \in 1:N$.

### Fig. 2. $H_{m}$ curve for $d = 2$ and $m = 1$ to $m = 6$ (source: Wikipedia)

The function $H$ admits a pseudoinverse $h : [0, 1)^{d} \to [0, 1]$, i.e. $H \circ h(x) = x$ for all $x \in [0, 1)^{d}$. $H$ is not a bijection because certain points $x \in [0, 1)^d$ have more than one preimage through $H$; however, the set of such points is of Lebesgue measure 0.

Informally, $H$ transforms $[0, 1]$ into $[0, 1)^d$, while preserving ‘locality’: if $x, x' \in [0, 1]$ are close, then $H(x)$ and $H(x')$ are close as well. We shall establish that $h$ also preserves discrepancy: a low discrepancy point set in $[0, 1)^d$ remains a low discrepancy point set in $[0, 1]$ when transformed through $h$. It is these properties that give to the Hilbert curve its appeal in the SQMC context (as opposed to other space filling curves, such as Z-ordering). We refer to Sagan (1994), Butz (1969) and Hamilton and Rau-Chaplin (2008) for how to compute $h$ in practice for any $d \geq 2$.

For $d = 1$, we simply set $H(x) = h(x) = x$ for $x \in [0, 1]$.

The following technical properties of $H$ and $H_{m}$ will be useful later (but may be skipped on first reading). For $m \geq 0$, let $I_{m}^{d} = \{I_{m}^{d}(k)\}_{k=0}^{2^{md}-1}$ be the collection of consecutive closed intervals in $[0, 1]$ of equal size $2^{-md}$ and such that $\cup I_{m}^{d} = [0, 1]$. For $k \geq 0$, $S_{m}^{d}(k) = H_{m}(I_{m}^{d}(k))$ belongs to

| Table 3. Algorithm 3: SQMC algorithm |
|---------------------------------------|
| At time $t = 0$:
  - (a) generate a QMC or an RQMC point set $u_{0;N}^{1}$ in $[0, 1)^d$, and compute $x_{0}^d = H_0(u_{0}^{1})$ for each $n \in 1:N$;
  - (b) compute $w_{0}^{n} = G_0(x_{0}^{n})$ and $W_{0}^{n} = w_{0}^{n}/\Sigma_{m=1}^{n}w_{0}^{m}$ for each $n \in 1:N$;
| Iteratively, from time $t = 1$ to time $t = T$:
  - (a) generate a QMC or an RQMC point set $u_{t;N}^{1}$ in $[0, 1)^{d+1}$; let $u_{t}^{n} = (u_{t}^{n}, x_{t}^{n}) \in [0, 1) \times [0, 1)^d$;
  - (b) Hilbert sort—find permutation $\sigma_{t-1}$ such that $h \circ \psi(x_{t-1}^{\sigma_{t-1}(N)}) \leq \ldots \leq h \circ \psi(x_{t-1}^{\sigma_{t-1}(1)})$ if $d \geq 2$, or $x_{t-1}^{\sigma_{t-1}(1)} \leq \ldots \leq x_{t-1}^{\sigma_{t-1}(N)}$ if $d = 1$;
  - (c) find permutation $\tau$ such that $u_{t}^{\tau(1)} \leq \ldots \leq u_{t}^{\tau(N)}$, generate $u_{t;N}^{1}$ by using algorithm 2, with inputs $u_{t}^{\tau}, w_{t-1}^{n}$, and compute $x_{t}^{n} = H_{\tau}^{\sigma_{t-1}(u_{t}^{\tau(n)})}$ for each $n \in 1:N$;
  - (d) compute $w_{t}^{n} = G_{\tau}(x_{t-1}^{\sigma_{t-1}(u_{t}^{\tau(n)}))}$, and $W_{t}^{n} = w_{t}^{n}/\Sigma_{m=1}^{n}w_{t}^{m}$ for each $n \in 1:N$.

- $H_{m}$ curve for $d = 2$ and $m = 1$ to $m = 6$ (source: Wikipedia)
2.4. Sequential quasi Monte Carlo for \( d \geq 2 \)

Assume now \( d \geq 2 \), and consider the following change of variables at iteration \( t \geq 1 \):

\[
h^n_{t-1} = h \circ \psi(x^n_{t-1}) \in [0, 1],
\]

where \( h : [0, 1]^d \rightarrow [0, 1] \) is the inverse of the Hilbert curve defined in the previous section, and \( \psi : \mathcal{X} \rightarrow [0, 1]^d \) is some user-chosen bijection between \( \mathcal{X} \) and \( \psi(\mathcal{X}) \subset [0, 1]^d \). To preserve the low discrepancy property of \( x^n_{t-1} \) it is important to choose for \( \psi \) a mapping which is discrepancy preserving. This requires us to select \( \psi \) such that \( \psi(x) = (\psi_1(x_1), \ldots, \psi_d(x_d)) \) where the \( \psi_i \)'s are continuous and strictly monotone. But choosing such a \( \psi \) is trivial in most applications; for example apply the logistic transformation componentwise when \( \mathcal{X} = \mathbb{R}^d \) (see Section 5 for more details).

With this change of variables, we obtain particles \( h_{t-1}^{1:N} \) that lie in \([0, 1]\), and equation (5) becomes

\[
\mathbb{Q}^N_{t,h} \{ d(h_{t-1}, x_t) \} = \sum_{n=1}^{N} W^n_{t-1} \delta_{h^n_{t-1}}(d(h_{t-1}) m_t(x^n_{t-1}, dx_t).
\]

Sampling a low discrepancy sequence from \( \mathbb{Q}^N_{t,h} \{ d(h_{t-1}, dx_t) \) may then proceed exactly as for \( d = 1 \), i.e. use the inverse transform method to sample \( N \) points \( h_{t-1}^{1:N} \) from the marginal distribution \( \mathbb{Q}^N_{t,h} \{ d(h_{t-1}) \); then sample \( x^n_{t-1} \) conditionally on \( h_{t-1}^{1:N} \), with \( x^n_{t-1} = \psi^{-1} \circ h(h_{t-1}^{1:N}) \). The exact details of the corresponding operations are the same as for \( d = 1 \). We therefore obtain the general SQMC algorithm as described in algorithm 3.

To define the SQMC algorithm fully, we must choose a particular method to generate point sets \( u_t^{1:N} \) at each iteration. If QMC point sets are generated, we obtain a deterministic algorithm, whereas, if RQMC point sets are generated, we obtain a stochastic algorithm.

2.5. Complexity of sequential quasi Monte Carlo

The complexity of both steps (b) and (c) (for \( t \geq 1 \)) of the SQMC algorithm is \( O\{N \log(N)\} \), because they include a sort operation. The complexity of step (a) depends on the method chosen for generating the point sets \( u_t^{1:N} \). For instance, Hong and Hickernell (2003) proposed an \( O\{N \log(N)\} \) method that applies to most constructions of \((t, s)\) sequences (such as the Faure, the Sobol’, the Niederreiter or the Niederreiter–Xing sequences). The cost to randomize a QMC point set is only \( O(N) \) if one chooses the simple random shift approach, whereas nested scrambling methods for \((t, s)\)-sequences, which are such that all the results below hold, may be implemented at cost \( O\{N \log(N)\} \) (Owen, 1995; Hong and Hickernell, 2003).

To summarize, the overall complexity of SQMC sampling is \( O\{N \log(N)\} \), provided that the method to generate the point sets \( u_t^{1:N} \) is chosen appropriately.

3. Convergence study

We concentrate on two types of asymptotic results (as \( N \rightarrow \infty \)): consistency, and stochastic bounds, i.e. bounds on the MSE for the randomized SQMC algorithm (i.e. an SQMC algorithm
Based on RQMC point sets. We leave deterministic bounds of the error (for when deterministic QMC point sets are used) to future work. We find stochastic bounds more interesting, because

(a) results from Owen (1997a,b, 1998) suggest that we might obtain better convergence rates than for deterministic bounds and
(b) the randomized version of SQMC has more applications, as discussed in Section 4.

These results are specialized to the case where the simulation of $x^n_t$ at time $t$ is based on the inverse transform method, as explained in Section 3.1. Certain of our results require $\mathcal{X} = [0, 1]^d$, and $\psi$ is set to the identity function. (Recall that, to deal with certain QMC technicalities, we follow the standard practice of taking $X = [0, 1]^d$ rather than $X = [0, 1]$.) The fact that $X$ is bounded may not be such a strong restriction, as our results allow for unbounded test functions $\phi$; thus, one may accommodate for an unbounded state space (and expectations with respect to that space) through appropriate variable transforms.

We introduce the following extreme norm. For any signed measure $\mu$ over $\mathcal{X} = [0, 1]^d$,

$$
\|\mu\|_E = \sup_{B \in \mathcal{B}_{[0,1]^d}} |\mu(B)|,
$$

which generalizes the extreme discrepancy in the sense that

$$
\|S(x^{1:N}) - \lambda_d\|_E = D(x^{1:N})
$$

for any point set $x^{1:N}$ in $\mathcal{X}$, where $S$ is the operator that associates with $x^{1:N}$ its empirical distribution:

$$
x^{1:N} \in \mathcal{X}^N \mapsto S(x^{1:N}) = \frac{1}{N} \sum_{n=1}^{N} \delta_{x^n}.
$$

Our consistency results will be stated with this norm. Note that $\|\pi^N - \pi\|_E \to 0$ implies that $|\pi^N(\phi) - \pi(\phi)| \to 0$ for any continuous bounded function $\phi$, by the portmanteau lemma (van der Vaart (2007), lemma 2.2).

The next subsection explains how the inverse method may be used to generate $x^n_t$ given $x^n_{t-1}$. The two following subsections state preliminary results that should provide insights on the main ideas that underpin the proofs of our convergence results. Readers who are interested mostly in the main results may skip these subsections and go directly to Section 3.4 (consistency) and Section 3.5 (stochastic bounds).

This section will use the following standard notation: $\|\varphi\|_\infty$ for the supremum norm for functions $\varphi$, $L_2(\mathcal{X}, \mu)$ for the set of square integrable functions $\varphi : \mathcal{X} \to \mathbb{R}$ and $C_b(\mathcal{X})$ for the set of continuous bounded functions $\varphi : \mathcal{X} \to \mathbb{R}$.

### 3.1. Inverse transform method

We discuss here how to write the simulation of $x^n_t$ as $x^n_t = \Gamma_t(x^n_{t-1}, u^n_t)$, using the inverse transform method. Our convergence results are specialized to this particular $\Gamma_t$.

For a generic distribution $\pi \in \mathcal{P}(\mathcal{X})$, $\mathcal{X} \subset \mathbb{R}^d$, let $F_\pi$ be the Rosenblatt transformation (Rosenblatt, 1952) of $\pi$ defined through the following chain rule decomposition:

$$
F_\pi(x) = (u_1, \ldots, u_d)^T, \quad x = (x_1, \ldots, x_d)^T \in \mathcal{X},
$$

where, recursively, $u_1 = F_{\pi,1}(x_1)$, $F_{\pi,1}$ being the CDF of the marginal distribution of the first
component (relative to \(\pi\)) and, for \(i \geq 2\), \(u_i = F_{\pi,i}(x_i | x_{1:i-1})\), \(F_{\pi,i}(\cdot | x_{1:i-1})\) being the CDF of component \(x_i\), conditionally on \((x_1, \ldots, x_{i-1})\), again relative to \(\pi\). Similarly, we define the multivariate generalized inverse CDF (GICDF) \(F_{\pi}^{-1}\) through the following chain rule decomposition:

\[
F_{\pi}^{-1}(u) = (x_1, \ldots, x_d)^T, \quad u = (u_1, \ldots, u_d)^T \in [0, 1)^d,
\]

where, recursively, \(x_1 = F_{\pi,1}^{-1}(u_1)\), \(F_{\pi,1}^{-1}\) being the GICDF of the marginal distribution of the first component (relative to \(\pi\)) and, for \(i \geq 2\), \(x_i = F_{\pi,i}^{-1}(u_i | x_{1:i-1})\), \(F_{\pi,i}^{-1}(\cdot | x_{1:i-1})\) being the GICDF of component \(x_i\), conditionally on \((x_1, \ldots, x_{i-1})\), again relative to \(\pi\). Note that this function depends on the particular order of the components of \(\pi\). For some probability kernel \(K : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{X})\), define similarly \(F_K(x, \cdot)\) and \(F_K^{-1}(x, \cdot)\) as respectively the Rosenblatt transformation and the multivariate GICDF of distribution \(K(x, dx)\) for a fixed \(x\).

It is well known that taking \(\Gamma_0 = F_{m_0}^{-1}\) and \(\Gamma_t = F_{m_t}^{-1}\) lead to valid simulation algorithms, i.e. if \(x^n_0 = F_{m_0}^{-1}(u^n_0)\) and \(x^n_t = F_{m_t}^{-1}(x^n_{t-1}, u^n_t)\), then \(x^n_0 \sim m_0(dx_0)\) and \(x^n_t | x^n_{t-1} \sim m_t(x^n_{t-1}, dx_t)\) respectively, and when \(u^n_0\) and \(u^n_t\) are uniform random numbers.

### 3.2. Preliminary results: importance sampling

Since SQMC is based on importance sampling (e.g. initialization of algorithm 3), we need to establish the validity of importance sampling based on low discrepancy point sets; see Götz (2002) and Aistleitner and Dick (2015) for other results on QMC-based importance sampling.

**Theorem 1.** Let \(\pi\) and \(q\) be two probability measures on \([0, 1)^d\) such that the Radon–Nikodym derivative \(w(x) = \pi(dx) / q(dx)\) is continuous and bounded. Let \((x^{1:N})\) be a sequence of point sets in \([0, 1)^d\) such that \(\|S(x^{1:N}) - q\|_E \rightarrow 0\) as \(N \rightarrow \infty\), and define

\[
\pi^N = \sum_{n=1}^N W^n \delta_{x^n}, \quad W^n = \frac{w(x^n)}{\sum_{m=1}^N w(x^m)}.
\]

Then, \(\|\pi^N - \pi\|_E \rightarrow 0\) as \(N \rightarrow \infty\).

See Appendix A.1.2 for a proof.

Recall that in our notation we drop the dependence of point sets on \(N\), i.e. we write \((x^{1:N})\) rather than \((x^{N,1:N})\), although in full generality \(x^{1:N}\) may not necessarily be the \(N\) first points of a fixed sequence.

The next theorem gives the stochastic error rate when an RQMC point set is used.

**Theorem 2.** Consider the set-up of theorem 1. Let \((u^{1:N})\) be a sequence of random point sets in \([0, 1)^d\) such that \(u^n \sim U([0, 1)^d)\) marginally and, \(\forall \varphi \in L_2([0, 1)^d, \lambda_d)\),

\[
\text{var} \left\{ \frac{1}{N} \sum_{n=1}^N \varphi(u^n) \right\} = \mathcal{O}(r(N)),
\]

where \(r(N) \rightarrow 0\) as \(N \rightarrow \infty\). Let \(x^{1:N} = F_q^{-1}(u^{1:N})\) and assume that either one of the following two conditions is verified:

(a) \(F_q^{-1}\) is continuous and, for any \(\epsilon > 0\), there is an \(N_\epsilon \in \mathbb{N}\) such that, almost surely, \(D^*(u^{1:N}) \leq \epsilon, \forall N \geq N_\epsilon\);

(b) for any \(\epsilon > 0\) there is an \(N_\epsilon \in \mathbb{N}\) such that, almost surely,

\[
\|S(x^{1:N}) - q\|_E \leq \epsilon, \quad \forall N \geq N_\epsilon.
\]

Then, for all \(\varphi \in L_2(\mathcal{X}, \pi)\),
We motivated the use of the Hilbert curve as a way to transform back and forth between \([0, 1]^3\). Preliminary results: Hilbert curve and discrepancy

Section 3.1. More precisely, for a probability measure \(\pi\) the validity, in the QMC context, of the multivariate GICDF approach that was described in 1995; Matoˇsek, 1998; Hong and Hickernell, 2003) are such that \(r(N) = N^{-1}\). This result was established for \(N = \lambda b^m\) in Owen (1997a, 1998) and extended for an arbitrary \(N \) in Gerber (2014).

3.3. Preliminary results: Hilbert curve and discrepancy

We motivated the use of the Hilbert curve as a way to transform back and forth between \([0, 1]^d\) and \([0, 1]\) while preserving low discrepancy in some sense. This section formalizes this idea.

For a probability measure \(\pi\) on \([0, 1]^d\), we write \(\pi_h\) the image by \(h\) of \(\pi\). For a kernel \(K: [0, 1]^d \to \mathcal{P}(\mathcal{X})\), we write \(\pi_h \otimes K_h \{d(h_1, x_2)\}\) the image of \(\pi \otimes K\) by the mapping \((x_1, x_2) \in [0, 1]^d \times \mathcal{X} \mapsto (h(x_1), x_2)\), where \(\pi \otimes K\) denotes the joint probability measure \(\pi(dx_1) K(x_1, dx_2)\).

The following theorem is a technical result on the conversion of discrepancy through \(h\).

**Theorem 3.** Let \((\pi^N)\) be a sequence of probability measures on \([0, 1]^d\) such that \(\|\pi^N - \pi\|_E \to 0\), where \(\pi(dx) = \pi(x) \lambda_d(dx)\) admits a bounded probability density \(\pi(x)\). Then

\[
\|\pi_h^N - \pi_h\|_E \to 0, \quad \text{as } N \to \infty.
\]

See Appendix A.1.3 for a proof.

The following theorem is an extension of Hlawka and Mück (1972), Satz 2, which establishes the validity, in the QMC context, of the multivariate GICDF approach that was described in Section 3.1. More precisely, for a probability measure \(\pi\) on \([0, 1]^d\), Hlawka and Mück (1972), Satz 2, showed that \(\|S\{F^{-1}_\pi(u^{1:N})\} - \pi\|_E \leq c D^*(u^{1:N})^{1/d}\) (under some conditions on \(F_\pi\); see below).

**Theorem 4.** Let \(K: [0, 1]^{d_1} \to \mathcal{P}([0, 1]^{d_2})\) be a Markov kernel and assume that

(a) for a fixed \(x_1 \in [0, 1]^{d_1}\), the \(i\)th co-ordinate of \(F_K(x_1, x_2)\) is strictly increasing in \(x_{2i} \in [0, 1), i \in 1:d_2,\) and, viewed as a function of \(x_1\) and \(x_2\), \(F_K(x_1, x_2)\) is Lipschitz,

(b) \(\pi^N(dx) = \sum_{n=1}^{N} W^N_n \delta_{x^n}(dx)\), \(x^n \neq x^m\) \(\forall n \neq m \in 1:N\), and \(\max_{n \in 1:N} W^n \to 0\), and

(c) the sequence \((\pi^N)\) is such that \(\|\pi^N - \pi\|_E \to 0\) as \(N \to \infty\), where \(\pi(dx) = \pi(x) \lambda_{d_1}(dx)\) admits a strictly positive bounded density \(\pi\).

Let \((u^{1:N}), u^n = (u^n, v^n) \in [0, 1]^{1+d_2}\), be a sequence of point sets in \([0, 1]^{1+d_2}\) such that \(D^*(u^{1:N}) \to 0\) as \(N \to \infty\), and define \(P^N_h = (h^{1:N}, x_2^{1:N})\) where

\[
h^n = F^{-1}_\pi^N(u^n), \quad x_1^n = H(h^n), \quad x_2^n = F^K_1(x_1^n, v^n).
\]

Then

\[
\|S(P^N_h) - \pi^N_h \otimes K_h\|_E \to 0, \quad \text{as } N \to \infty.
\]

See Appendix A.2.2 for a proof.

Assumption (a) regarding the regularity of the vector-valued function \(F_K\) is the main assumption of theorem 4 and comes from Hlawka and Mück (1972), Satz 2. It is verified as soon as kernel \(K\) admits a density that is continuously differentiable on \([0, 1]^{d}\) (Hlawka and Mück...
(1972), page 232). Assumption 2 is a technical condition, which will always hold under the assumptions of our main results.

3.4. Consistency
We can now establish the consistency of SQMC; see Appendix A.3 for a proof of the following theorem. For convenience, let $F_{m_i}(x_{t-1}, x_t) = F_{m_0}(x_0)$ when $t = 0$.

Theorem 5. Consider the set-up of algorithm 3 where, for all $t \in [0, T]$, $(u_t^{1:N})$ is a (non-random) sequence of point sets in $[0, 1)^d_t$, with $d_0 = d$ and $d_t = d + 1$ for $t > 0$, such that $D^*(u_t^{1:N}) \to 0$ as $N \to \infty$. Assume that the following conditions hold for all $t \in [0, T]$

(a) the components of $x_t^{1:N}$ are pairwise distinct, $x_t^n \neq x_t^m$ for $n \neq m$;
(b) $G_t$ is continuous and bounded;
(c) $F_{m_i}(x_{t-1}, x_t)$ verifies assumption (a) of theorem 4;
(d) $Q_t(dx_t) = p_t(x_t) \lambda_d(dx_t)$ where $p_t(x_t)$ is a strictly positive bounded density.

Let $\bar{Q}_t(dx_t) = \sum_{n=1}^{N} W_t^n \delta_{x_t^n}(dx_t)$. Then, under assumptions (a)–(d), as $N \to \infty$,

$$\|\bar{Q}_t - Q_t\|_{E} \to 0, \quad \forall t \in [0, T].$$

Assumption (a) is stronger than necessary because for the result to hold it is enough that the number of identical particles does not grow too quickly as $N \to \infty$. This is a very weak restriction since assumption (a) holds almost surely when RQMC point sets are used, since then the particles are generated from a continuous GICDF. The assumption that the weight functions ($G_t$) are bounded is standard in the SMC literature (see for example Del Moral (2004)).

3.5. Stochastic bounds
Our second main result concerns stochastic bounds for the randomized version of the SQMC algorithm, i.e. SQMC based on randomized point sets $(u_t^n)$. See Appendix A.4 for a proof of the next result.

Theorem 6. Consider the set-up of algorithm 3 where $(u_t^{1:N})$, $t \in [0, T]$, are independent sequences of random point sets in $[0, 1)^d_t$, with $d_0 = d$ and $d_t = d + 1$ for $t > 0$, such that, for all $t \in [0, T]$, $u_t^n \sim U([0, 1)^d_t)$ marginally and

(a) for any $\epsilon > 0$, there is an $N_{\epsilon,t} > 0$ such that, almost surely, $D^*(u_t^{1:N}) \leq \epsilon$, $\forall N \geq N_{\epsilon,t}$, and
(b) for any function $\varphi \in L_2([0, 1)^d_t, \lambda_d_t)$, $\var{\{1/N \sum_{n=1}^{N} \varphi(u_t^n)\}} \leq C \sigma_{\varphi}^2 r(N)$ where $\sigma_{\varphi}^2 = \int\int (\varphi(u) - \varphi(v)dv)^2 du$, and where neither $C^*$ nor $r(N)$ depends on $\varphi$.

In addition, assume that the assumptions of theorem 5 are verified and that $F_{m_0}^{-1}$ is continuous. Let $\varphi \in L_2([0, 1)^d, \mathbb{Q}_t)$ for all $t \in [0, T]$. Then, $\forall t \in [0, T]$,

$$\mathbb{E}|\tilde{Q}_t^{N}(\varphi) - \mathbb{Q}_t(\varphi)| = \mathcal{O}(r(N)^{1/2}),$$

$$\var{\tilde{Q}_t^{N}(\varphi)} = \mathcal{O}(r(N)).$$

Note that the implicit constants in the expression above may depend on $\varphi$. Assumptions (a) and (b) are verified for $r(N) = N^{-1/2}$ if $u_t^{1:N}$ is the first $N$ points of a nested scrambled $(t,s)$-sequence in base $b \geq 2$. This result was established for $N = \lambda b^m$ in Owen (1997a, 1998) and can be extended to any pattern of $N$ by using Hickernell and Yue (2001), lemma 1. Consequently, for this construction of RQMC point sets, theorem 6 shows that the approximation error of the
SQMC algorithm goes to zero at least as fast as for the SMC algorithm. However, contrary to the $O(N^{-1})$ convergence rate of the SMC algorithm, this rate for the SQMC algorithm based on nested scrambled $(t, s)$-sequences is not exact but results from a worst case analysis. We can therefore expect to reach faster convergence on a smaller class of functions. The following result shows that it is indeed the case on the class of continuous and bounded functions; see Appendix A.4.4 for a proof.

**Theorem 7.** Consider the set-up of algorithm 3 where $(u_t^{1:N})$, $t \in 0:T$, are $(t, d_t)$-sequences in base $b \geq 2$, with $d_0 = d$ and $d_t = d + 1$ for $t > 0$, independently scrambled such that the results in Owen (1997a, 1998) hold. Let $N = \lambda b^m$, $1 \leq \lambda < b$, and assume that the following conditions hold:

(a) the assumptions of theorem 6 are verified;
(b) for $t \in 1:T$, $F_{m_t}^{-1}(x_{t-1}, x_t)$ is a continuous function of $x_{t-1}$.

Let $\varphi \in C_b(\mathcal{X})$. Then, $\forall t \in 0:T$,

$$E[\hat{Q}_t^N(\varphi) - Q_t(\varphi)] = o(N^{-1/2}),$$

$$\text{var}\{\hat{Q}_t^N(\varphi)\} = o(N^{-1}).$$

Thus, for SQMC based on the first $N = \lambda b^m$ points of nested scrambled $(t, s)$-sequences in base $b$, we obtain that the stochastic error of (the random version of) the SQMC algorithm converges faster than for the SMC algorithm. Note that we can relax the constraint on $N$ in theorem 7 by using Gerber (2014), corollary 2.

## 4. Extensions

### 4.1. Unbiased estimation of evidence, particle Markov chain Monte Carlo

Like SMC, the randomized version of the SQMC algorithm (i.e. the SQMC algorithm based on RQMC point sets) provides an unbiased estimator of the normalizing constant $Z_t$ of the Feynman–Kac model; see expression (2).

**Lemma 1.** Provided that $u_t^{1:N}$, $t \in 0:T$, are independent RQMC point sets in $[0, 1)^{d_t}$ (i.e. $u_t^n \sim U([0, 1)^{d_t})$ marginally), with $d_0 = d$ and $d_t = d + 1$ for $t > 0$, the quantity

$$Z_t^N = \left\{ \frac{1}{N} \sum_{n=1}^{N} G_0(x_0^n) \right\} \prod_{s=1}^{t} \left\{ \frac{1}{N} \sum_{n=1}^{N} G_s(x_{t-1}^n, x_t^n) \right\}$$

is an unbiased estimator of $Z_t$, $E[Z_t^N] = Z_t$.

We omit the proof, as it follows the same steps as for the SMC algorithm (Del Moral, 1996).

In a state space model parameterized by $\theta \in \Theta$, $Z_t = Z_t(\theta)$ is the marginal likelihood of the data up to time $t$. One may want to implement a Metropolis–Hastings sampler with respect to posterior density $\pi_T(\theta) \propto p(\theta) Z_T(\theta)$ for the full data set and for a prior distribution $p(\theta)$, but $Z_T(\theta)$ is typically intractable.

Andrieu et al. (2010) established that, by substituting $Z_T(\theta)$ with an unbiased estimate of $Z_T(\theta)$ in a Metropolis sampler, one obtains an exact MCMC algorithm, in the sense that the corresponding MCMC kernel leaves invariant $\pi_T(\theta)$. The algorithm so obtained is called the particle marginal Metropolis–Hastings (PMMH) algorithm. Andrieu et al. (2010) used the SMC method to obtain an unbiased estimate of $Z_T(\theta)$, i.e. at each iteration an SMC sampler is run to obtain that estimate. We shall call PMMH–SQMC the same algorithm, but with SQMC replacing SMC for the evaluation of an unbiased estimate of the likelihood.
The acceptance rate of the PMMH algorithm depends directly on the variability of the estimates of $Z_T(\theta)$. Since the point of (randomized) SQMC is to provide estimates with a lower variance than the SMC estimates (for a given $N$), we may expect that the PMMH–SQMC algorithm may require a smaller number of particles than the standard PMMH algorithm for satisfactory acceptance rates; see Section 5 for a numerical illustration of this.

4.2. Smoothing

Smoothing amounts to computing expectations $Q_t(\varphi)$ of functions $\varphi$ of the complete trajectory $x_{0:t}$; for example $Q_t(\varphi)$ is the expectation of $\varphi(x_{0:t})$ conditionally on data $y_{0:t}$ for a state space model with Markov process $(x_t)$ and observed process $(y_t)$. See Briers et al. (2010) for a general overview on SMC smoothing algorithms. This section discusses how to adapt certain of these algorithms to SQMC.

4.2.1. Forward smoothing

Forward smoothing amounts to carrying forward the complete trajectories of the particles, rather than simply keeping the last component $x_n^T$ (as in algorithm 1). A simple way to formalize forward smoothing is to introduce a path Feynman–Kac model, corresponding to the inhomogeneous Markov process $z_t = x_{0:t}$, and weight function (abusing the notation) $G_t(z_t) = G_t(x_t)$. Then forward smoothing amounts to algorithm 1 applied to this path Feynman–Kac model (substituting $x_t$ with $z_t = x_{0:t}$).

We may use the same remark to define an SQMC version of forward smoothing, i.e. simply apply SQMC to the same path Feynman–Kac model. The only required modification is that the Hilbert sort of step (b) at times $t \geq 1$ must now operate on some transformation of the vectors $z_n^t$, of dimension $(t + 1)d$, rather than vectors $x_n^t$ of dimension $d$ as in the original version.

Forward smoothing is sometimes used to approximate the smoothing expectation of additive functions, $\varphi(x_{0:t}) = \sum_{s=0}^{t} \tilde{\varphi}(x_s)$, such as the score function of certain models (e.g. Poyiadjis et al. (2011)). In that case, we may instead apply the SQMC method to the Feynman–Kac model corresponding to the inhomogeneous Markov process $z_t = (\sum_{s=0}^{t-1} \tilde{\varphi}(x_s), x_t)$. This means that, in practice, we may implement the Hilbert sort on a space of much lower dimension (i.e. the dimension of this new $z_t$), which is computationally more convenient.

4.2.2. Backward smoothing

Backward smoothing consists of two steps:

(a) a forward pass, where the SMC algorithm is run from time 0 to time $T$;

(b) a backward pass, where one constructs a trajectory $\tilde{x}_{0:T}$ recursively backwards in time, by selecting randomly each component $\tilde{x}_t$ out of the $N$ particle values $x_n^T$ that are generated during the forward pass.

An advantage of backward smoothing is that it is less prone to degenerate than forward smoothing. A drawback of backward smoothing is that generating a single trajectory costs $O(N)$; hence obtaining $N$ of them costs $O(N^2)$.

Backward smoothing for SQMC may be implemented in a similar way to SMC; see algorithm 4 in Table 4 for the backward pass that generates $N_B$ trajectories $\tilde{x}_{0:T}^{1:N_B}$ from the output of the SQMC algorithm. Note that backward smoothing requires that the Markov kernel $m_t(x_{t-1}, dx_t)$ admits a closed form density $m_t(x_t|x_{t-1})$ with respect to an appropriate dominat-
5. Numerical study

The objective of this section is to compare the performance of the SMC and SQMC methods. Our comparisons are either for the same number of particles $N$, or for the same amount of central processor unit (CPU) time to take into account the fact that SQMC has greater complexity than SMC. These comparisons will often be summarized through gain factors, which we define as ratios of MSEs (for a certain quantity) between SMC and SQMC estimates.

In the SQMC algorithm we generate $\tilde{u}_i^{1:N}$ as an Owen (1995) nested scrambled Sobol’ sequence by using the C++ library of T. Kollig and A. Keller (http://www.uni-kl.de/AG-Heinrich/SamplePack.html). Both the generation and the randomization of $(t, s)$-sequences in base 2 (such as the Sobol’ sequence) are very fast since logical operations can be used. To sort the particles according to their Hilbert index we use the C++ library of Chris Hamilton (http://web.cs.dal.ca/~chamilto/hilbert/index.html) to evaluate $H_m^{-1}\{\psi(x)\}$, $m \in \mathbb{N}$. Again, Hilbert computations are very fast as they are based on logical operations (see Hamilton and Rau-Chaplin (2008) for more details). In addition, thanks to the nesting property of the Hilbert curve (see Section 2.3) we only need to take $m$ sufficiently large such that different particles are mapped into different points of $[0, 1]$. Function $\Gamma_i$ is set to the inverse transform that was described in Section 3.1, and function $\psi$ to a componentwise (rescaled) logistic transform, i.e. $\psi(x) = (\psi_1(x_1), \ldots, \psi_d(x_d))$ with

$$\psi_i(x_i) = \left\{1 + \exp\left(-\frac{x_i - \tilde{\lambda}_i}{\tilde{\sigma}_i}\right)\right\}^{-1}, \quad i \in 1:d,$$

and where the constants $\tilde{\lambda}_i$ and $\tilde{\sigma}_i$ are used to solve numerical problems due to high values of $|x_i|$. For instance, when $(x_\iota)$ is a stationary process we chose $\tilde{\lambda}_i = \mu_i + 2\sigma_i$ and $\tilde{\sigma}_i = \mu_i - 2\sigma_i$ where $\mu_i$ and $\sigma_i$ are respectively the mean and the standard deviation of the stationary distribution of $(x_\iota)$.

The SMC algorithm is implemented by using systematic resampling (Carpenter et al., 1999) and all the random variables are generated by using standard methods (i.e. not using the multi-

| Table 4. Algorithm 4: backward step of SQMC backward smoothing |
|---------------------------------|
| **Input:** $x_0^{1:N}, \omega_0^{1:N}$ (output of the SQMC algorithm obtained after the Hilbert sort) |
| **Output:** $\tilde{w}_t^{1:N}$ (SQMC trajectories in $\mathcal{X}^{T+1}$) |
| Find permutation $\tau$ such that $\tilde{u}_0^{1:N} \leq \cdots \leq \tilde{u}_0^{N_B}$, generate $\tilde{w}_t^{1:N_B}$ by using algorithm 2, |
| with inputs $\tilde{u}_0^{1:N_B}$ and $W_T^{1:N}$, and set $\tilde{x}_T^n = \tilde{x}_T^w$ for all $n \in 1:N_B$; |
| for $t = T - 1 \rightarrow 0$ do |
| for $n \in 1:N_B$, set $\tilde{x}_T^n = \tilde{x}_T^w$ where $\tilde{a}_i^n = F_{\pi_i^{-1}}^t(\tilde{u}_t^{\pi_i(n)})$, $\pi_i^n = \sum_{m=1}^{N_B} \tilde{w}_m^t(\tilde{x}_t^n) \delta_m$ |
| and, for $m \in 1:N$, |
| $\tilde{w}_m^t(x_{t+1}) = W_t^{\pi_i(m)}(x_{t+1}) | x_t^n \rangle \bigg/ \left\{ \sum_{n=1}^N W_t^n m_t(x_{t+1}) | x_T^w \rangle \right\}$; |
| end for |
variates GICDF). The C and C++ code implementing both the SMC and the SQMC algorithm is available from https://bitbucket.org/mgerber/sqmc.

Even if theorems 6 and 7 are valid for any pattern of \( N \), choosing for \( N \) powers of 2 (with 2 the base of the Sobol' sequence) is both natural and optimal for QMC methods based on (scrambled) \((t,s)\)-sequences (see for example Owen (1997b), Hickernell and Yue (2001) and chapter 5 of Dick and Pillichshammer (2010)). Comparing the performance of the SQMC algorithm for different patterns of \( N \) is beyond the scope of this paper (see Gerber (2014) for a discussion of this point) and therefore we follow in this numerical study the standard approach in the QMC literature by considering values of \( N \) that are powers of 2. We nevertheless do one exception to this rule for the PMMH estimation on real data (Section 5.3) because doubling the number of particles to reduce the variance of the likelihood estimate that is used in the Metropolis–Hastings ratio may be very inefficient from a computational point of view. As we shall see, allowing \( N \) to differ from powers of the Sobol' base does not seem to alter the performance of the SQMC algorithm.

We may expect the two following situations to be challenging for SQMC:

(a) small \( N \) (because our results are asymptotic) and
(b) large \( d \) (because of the usual deterioration of QMC with respect of the dimension, and also because of the Hilbert sort step).

Thus we consider examples of varying dimensions (from 1 to 10), and we shall also make \( N \) vary within a large range (between \( 2^4 \) and \( 2^{17} \)).

5.1. Example 1: a non-linear and non-stationary univariate model

We consider the following popular toy example (Gordon et al., 1993; Kitagawa, 1996):

\[
\begin{align*}
    y_t &= x_t^2/a + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}_1(0, 1), \quad t \geq 0, \\
    x_t &= b_1 x_{t-1} + b_2 \frac{x_{t-1}}{1 + x_{t-1}^2} + b_3 \cos(b_4 t) + \sigma \nu_t, \quad \nu_t \sim \mathcal{N}_1(0, 1), \quad t > 0
\end{align*}
\]

and \( x_0 \sim \mathcal{N}_d(0, 2) \), where \( \mathcal{N}_d(\mu, \Sigma) \) denotes the \( d \)-dimensional Gaussian distribution with mean \( \mu \) and covariance matrix \( \Sigma \). We generate observations from 100 time steps of the model, with the parameters set as in Gordon et al. (1993): \( a = 20, b = (0.5, 25, 8, 1.2), \sigma^2 = 10 \) and \( x_0 = 0.1 \). Inference in this model is non-trivial because the observation \( y_t \) does not allow us to identify the sign of \( x_t \), and because the weight function \( G_t(x_t) \) is bimodal if \( y_t > 0 \) (with modes at \( \pm (20y_t)^{1/2} \)). In addition, we expect this model to be challenging for SQMC because of the high non-linearity of the Markov transition \( m_t(x_{t-1}, dx_t) \).

All the results that are presented below are based on 500 independent runs of the SMC and SQMC algorithms. Fig. 3 presents results concerning the estimation of the log-likelihood functions evaluated at the true value of the parameters. Figs 3(a) and 3(b) show that, compared with SMC, the SQMC algorithm yields faster convergence of both the mean and the variance of the estimates.

These better consistency properties of the SQMC algorithm are also illustrated in Fig. 3(c) where we have reported for each \( N \) the range in which lie the 500 estimates of the log-likelihood. From this plot we see that quickly the SQMC estimates stay in a very tiny interval whereas, in contrast, the SMC estimates are much more dispersed, even for large values of \( N \).

Fig. 3(d) shows the MSE of SQMC and SMC as a function of CPU time. We see that the gain of SQMC over SMC not only increases with \( N \), as predicted by the theory, but also with
the CPU time, which is of more practical interest. In contrast, in this particular case (log-likelihood evaluation for this univariate model), when \( N \) is small the reduction in MSE brought by SQMC does not compensate its greater running time. Nevertheless, we observe that the SQMC outperforms the SMC algorithm very quickly, i.e. as soon as the CPU time is larger than or equal to \( 10^{-1.5} \approx 0.03 \) s.

In Fig. 4(a) we have reported the gain factor for the estimation of \( \mathbb{E}[x_t|y_0:t] \) as a function of \( t \) and for various values of \( N \). From this plot we observe both significant and increasing gain of the SQMC over the SMC method.
Fig. 4 compares SQMC and SMC backward smoothing for the estimation of $E[x_t|y_{0:T}]$ as a function of $t$ and for $N \in \{27, 29\}$. As for the filtering problem, the SQMC algorithm significantly outperforms the SMC algorithm with gain factors that increase with the number of particles.

### 5.2. Example 2: Multivariate Stochastic Volatility Model

We consider the following multivariate stochastic volatility (SV) model proposed by Chan et al. (2006):

\[
\begin{align*}
    y_t &= S_t^{1/2} \epsilon_t, \\
    x_t &= \mu + \Phi (x_{t-1} - \mu) + \Psi^{1/2} \nu_t, \\
\end{align*}
\]

where $S_t = \text{diag}(\exp(x_{t1}), \ldots, \exp(x_{td}))$, $\Phi$ and $\Psi$ are diagonal matrices and $(\epsilon_t, \nu_t) \sim N_{2d}(0_{2d}, C)$, with $C$ a correlation matrix and $0_{2d} = (0, \ldots, 0) \in \mathbb{R}^{2d}$.

To study the relative performance of the SQMC over the SMC method as the dimension $d$ of the hidden process increases we perform simulations for $d \in \{1, 2, 4, 10\}$. The parameters that we use for the simulations are the same as in Chan et al. (2006): $\phi_{ii} = 0.9$, $\mu_i = -9$ and $\psi_{ii} = 0.1$ for all $i = 1, \ldots, d$ and

\[
C = \begin{pmatrix}
    0.61_d & 0.4I_d & -0.11_d - 0.2I_d \\
    -0.11_d - 0.2I_d & 0.81_d + 0.2I_d
\end{pmatrix}
\]

where $I_d$ is the $d$-dimensional identity matrix, and $1_d$ is the $d \times d$ matrix having 1 in all its entries. Note that the errors terms $\epsilon_t$ and $\nu_t$ are correlated so the weight function $G_t$ now depends both on $x_{t-1}$ and on $x_t$. The prior distribution for $x_0$ is the stationary distribution of the process $(x_t)$ and we take $T = 399$.

Figs 5(a)–5(c) present results for the estimation of the log-likelihood (evaluated at the true value of the parameters and for the complete data set $y_{0:T}$), for $d \in \{1, 2, 4\}$. We see that the gain factor increases quickly with $N$ and, more importantly, the MSE of the SQMC estimate
Fig. 5. Log-likelihood estimation of SV model (7) (the graphs are obtained from 200 independent SQMC (———) and SMC (---) runs): MSE as a function of CPU time for (a) $d=1$, (b) $d=2$ and (c) $d=4$, and (d) gain factor as a function of $N$ for $d=1, 2, 4, 10$

converges faster than that of the SMC estimate even as a function of CPU time. In fact, except for a very small interval for the four-dimensional model, the SQMC algorithm always outperforms the SMC algorithm in terms of MSE for the same CPU effort. We note the particularly impressive values of the gain factor that we obtain for $d=1$ when $N$ is large: around $4.2 \times 10^4$ for $N=2^{17}$. Fig. 5(d) plots the gain factors as a function of $N$, for the same values of $d$, plus $d=10$. The improvement that is brought by SQMC decreases with the dimension and, in fact, for $d=10$, the gain factor is essentially 1 for the values of $N$ considered; yet for $d=4$ we still observe some notable improvement, e.g. a gain factor of 10 for $N \approx 10^5$. We now focus on $d=1, 2, 4$. 

![Image](image_url)
Fig. 6. Log-likelihood estimation of the SV model (7)—gain factor as a function of $t$, obtained from 200 independent SQMC and SMC runs: (a) univariate model ($\cdots$, $N = 2^5$; $\ldots$, $N = 2^{10}$; $\ldots$, $N = 2^{13}$); (b) univariate model ($\cdots$, $N = 2^{14}$; $\ldots$, $N = 2^{15}$; $\ldots$, $N = 2^{16}$; $\ldots$, $N = 2^{17}$); (c) bivariate model ($\cdots$, $N = 2^5$; $\cdots$, $N = 2^{10}$; $\cdots$, $N = 2^{13}$); (d) bivariate model ($\cdots$, $N = 2^{14}$; $\cdots$, $N = 2^{15}$; $\ldots$, $N = 2^{16}$; $\ldots$, $N = 2^{17}$); (e) four-dimensional model ($\cdots$, $N = 2^5$; $\cdots$, $N = 2^{10}$; $\ldots$, $N = 2^{13}$); (f) four-dimensional model ($\cdots$, $N = 2^{14}$; $\cdots$, $N = 2^{15}$; $\ldots$, $N = 2^{16}$; $\ldots$, $N = 2^{17}$).
Fig. 7. Filtering of the multivariate SV model (7)—gain factor as a function of $t$ for the estimation of $\mathbb{E}[x_{1t}|y_{0:t}]$, obtained from 200 independent SQMC and SMC runs ($\cdots$, $N = 2^5$; $\cdots$, $N = 2^{10}$; $\cdots$, $N = 2^{17}$); (a) bivariate model; (b) four-dimensional model

Fig. 6 represents the evolution with respect to $t$ of the MSE for the partial log-likelihood of data $y_{0:t}$ up to time $t$; gain factors are reported for various values of $N$. As we can see from these graphs, the performance of the SQMC algorithm does not seem to depreciate with $t$.

Finally, Fig. 7 shows that SQMC also gives impressive gains when $d > 1$ concerning the estimation of the filtering expectation $\mathbb{E}[x_{1t}|y_{0:t}]$ of the first component of $x_t$.

5.3. Application: Bayesian estimation of multivariate stochastic volatility by using the particle marginal Metropolis–Hastings algorithm on real data

To compare the SMC with the SQMC method when used as a way to approximate the likelihood within a PMMH algorithm, as described in Section 4.1, we turn our attention to the Bayesian estimation of the multivariate SV model (7), for $d = 2$. As in Chen et al. (2006), we take the prior

$\phi_{ii} \sim U(0, 1), \quad 1/\psi_{ii} \sim \text{gamma}(10 \exp(-10), 10 \exp(-3)) \quad i = 1, \ldots, d,$

where $\phi_{ii}$ and $\psi_{ii}$ denote respectively the diagonal elements of $\Phi$ and $\Psi$, and a flat prior for $\mu$. In addition, we assume that $C$ is uniformly distributed on the space of correlation matrices which are such that the errors terms $\epsilon_t$ and $\nu_t$ are independent (no leverage effects). To sample from the posterior distribution of the parameters we use a Gaussian random-walk Metropolis–Hastings algorithm with covariance matrix $\Sigma$ calibrated so that the acceptance probability of the algorithm becomes, as $N \to \infty$, close to 25%. The matrix $\Sigma$, as well as the starting point of the Markov chain, is calibrated by using a pilot run of the algorithm with $\Sigma = 0.011^2 I_8$ and starting at the value of the parameters that we used above for the simulations. To compare the PMMH–SQMC algorithm with the PMMH–SMC algorithm, we run the two algorithms during $10^5$ iterations and for values of $N$ ranging from 10 to 200, where $N$ increases from 10 to 100 by increments of 10 and then by increments of 50.

We consider the following data set: the two series are the mean-corrected daily returns on the Nasdaq and Standard and Poor’s 500 indices for the period ranging from January 3rd, 2012, to October 21st, 2013, so that the data set contains 452 observations.

Fig. 8 shows the Metropolis–Hastings acceptance rate and the effective sample sizes (see
Robert and Casella (2004), section 12.3.5, for a definition) for the PMMH–SQMC algorithm and for the standard PMMH algorithm. We first observe that the acceptance rate of the PMMH–SQMC algorithm increases very quickly with \( N \). Indeed, it is already 20% for only 30 particles whereas for the same number of particles the acceptance rate for the standard PMMH algorithm is approximatively 6.5%. As far as the acceptance rate is concerned, there is no significant gain in taking \( N > 60 \) for the PMMH–SQMC algorithm whereas for the plain Monte Carlo algorithm the acceptance rate is only about 20% for \( N = 200 \) and therefore much smaller than the target of 25%. Looking at the results for the effective sample sizes, we see that the same conclusions hold. More precisely, for the PMMH–SQMC algorithm, the effective sample sizes increase with \( N \) much faster than for the PMMH–SMC algorithm. Indeed, for \( N \in 10 : 50 \), the effective sample size for the former is between 2.18 and 14.94 times larger than for the PMMH–SMC algorithm.

### 5.4. Example 3: neural decoding

Neural decoding models are used for brain–machine interfaces to make inference about an organism’s environment from its neural activity. More precisely, we consider the problem of decoding a set of environment variables \( \mathbf{p}_t \in \mathbb{R}^2 \), from the firing ensemble of \( d_y \) neurons. The latent vector \( \mathbf{p}_t \) may be interpreted as two-dimensional hand kinetics for motor cortical decoding (see Koyama et al. (2010) and references therein for more details about neural decoding models). Denoting \( \dot{\mathbf{p}}_t \), the vector of velocities, the neural decoding model that we consider is given by (Koyama et al., 2010)

\[
\begin{align*}
\mathbf{y}_{ti} | \mathbf{x}_{0:t} & \sim \mathcal{P} \{ \Delta \exp (\alpha_i + \beta_i^T \mathbf{x}_t) \}, \quad i \in 1 : d_y, \quad t \geq 0, \\
\mathbf{x}_t & = \Phi \mathbf{x}_{t-1} + \Psi \epsilon_t, \quad \epsilon_t \sim \mathcal{N}_2 (\mathbf{0}_2, \sigma^2 \mathbf{I}_2), \quad t > 0,
\end{align*}
\]

and \( \mathbf{x}_0 \sim \mathcal{N}_4 (\mathbf{0}_4, \mathbf{I}_4) \), where \( \mathbf{x}_t = (\mathbf{p}_t, \dot{\mathbf{p}}_t) \), the \( \mathbf{y}_{t|s} \) are conditionally independent, \( \mathcal{P} (\lambda) \) denotes the Poisson distribution with parameter \( \lambda \), \( \Delta \) is the duration of the interval over which spikes are counted at each time step and
Fig. 9. Log-likelihood estimation of the neural decoding model (8) (the graphs are obtained from 200 independent SQMC and SMC runs): (a) ratio of the SMC and the SQMC MSE; (b) SQMC (—) and SMC (–––) estimates

\[
\Phi = \begin{pmatrix} I_2 & \Delta I_2 \\ 0_2 & I_2 \end{pmatrix},
\]

\[
\Psi^T = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.
\]

Realistic values for the parameters (see Koyama et al. (2010)) that we shall take in our simulations are \(d_y = 10\), \(T = 23\), \(\Delta = 0.03\), \(\sigma^2 = 0.019\), \(\alpha_i \sim \text{IID} \mathcal{N}(2.5, 1)\) and \(\beta_i \sim \mathcal{U}([0, 1/d])\).

One important aspect of this model is that the dimension of the noise term \(\epsilon_t\) is lower than the dimension of \(x_t\). As a result, two components of \(x_t\) are deterministic functions of \(x_{t-1}\). Many tracking problems have a similar structure.

This requires us to adapt the SQMC algorithm slightly as follows: we sample jointly the ancestor variables \(a_t^{1:N}\) and the new velocities \(\dot{p}_n^t\) as in steps (b) and (c) of algorithm 3; then we obtain the new \(p_n^t\) as \(p_n^t = p_n^{t-1} + \dot{p}_n^t\), i.e. the deterministic linear transformation of \(p_n^{t-1}\) and \(\dot{p}_n^{t-1}\) defined by the model. In this case the dimension of the point set \(u_t^{1:N}\) is 3 for \(t > 0\); we could say that \(d = 2\) in this case, even if the dimension of \(x_t\) itself is 4.

Figs 9 and 10 present respectively results for the estimation of the log-likelihood (evaluated at the true value of the parameters) and for the estimation of the filtering expectation \(E[x_t|y_{0:t}]\) for \(i \in 1:d\). Concerning the log-likelihood estimation we observe a fast increase in the gain factor after about \(2^{11}\) particles with a maximum close to 21 when \(N\) is very large. The gain of the SQMC algorithm compensates its longer running time after only about 0.17 s. Important and increasing (in \(N\)) gains are also observed for the estimation of the filtering expectations.

6. Conclusion and future work

The main message of the paper is that SMC algorithm users should be strongly encouraged to switch to the SQMC method, as it is typically much more accurate (produces estimates with
Fig. 10. Filtering of the neural decoding model (8)—ratio of the SMC and SQMC MSE for the estimation of $E[x_{kt} \mid y_{0:t}]$ as a function of $t$, obtained from 200 independent SQMC and SMC runs ($\cdots \cdots \cdot N = 2^6, \cdots \cdots \cdot N = 2^{17}$): (a) $k = 1$; (b) $k = 2$; (c) $k = 3$; (d) $k = 4$

smaller errors) than SMC sampling. We add the word ‘typically’ to recall that our asymptotic analysis, by construction, proves only that the SQMC error is smaller than the SMC error for $N$ sufficiently large. But our range of numerical examples, which are representative of real world filtering problems, makes us optimistic that in most practical cases the SQMC algorithm should outperform the SMC algorithm even for moderate values of $N$.

The main price to pay to switch to SQMC is that users should spend some time thinking about how to write the simulation of $x^n_t$ given $x^n_{t-1}$ as $x^n_t = \Gamma_t(x^n_{t-1}, u^n_t)$, where $u^n_t \sim \mathcal{U}(0, 1)$ and $\Gamma_t$ is a deterministic function that is easy to evaluate. Fortunately, this is often straightforward. In fact, there are many models of interest where $x^n_t$ given $x^n_{t-1}$ is linear and Gaussian. Since this case is already implemented in our program, adapting it to such a model should be just a matter of changing a few lines of code (to evaluate the probability density of $y_t$ given $x_t$).

Regarding future work, the most pressing tasks seem
(a) to refine the convergence rate of the SQMC error and
(b) to establish that it does not degenerate over time (in the spirit of time uniform estimates for SMC; see page 244 of Del Moral (2004)).

Regarding the former, He and Owen (2014) make the interesting conjecture that the MSE of SQMC estimates converges at rate $O(N^{-1-2/d})$. This would explain why the relative performance of the SQMC method decreases with the dimension. Fortunately, the majority of the state space models of interest in signal processing, finance or other fields are such that $d \leq 6$. A notable exception is geophysical data assimilation (in for example meteorology or oceanography) for which $d$ can be very large, but for such large dimensional problems the SMC algorithm seems to perform too poorly for practical use anyway (Bocquet et al., 2010).

Finally, it is also our hope that this paper will help QMC to garner wider recognition in Bayesian computation and related fields. Granted, QMC is more technical than standard Monte Carlo methods, and there is perhaps something specific about particle filtering that makes the introduction of QMC so effective. Yet we can only think that the full potential of QMC in statistics remains underexplored.

Acknowledgements

We thank the referees, Christophe Andrieu, Simon Barthelmé, Arnaud Doucet, Paul Fearnhead, Simon Lacoste-Julien and Art Owen for excellent remarks that helped us to improve the paper greatly. The second author is partially supported by a grant from the French National Research Agency as part of the ‘Investissements d’Avenir’ programme (ANR-11-LABEX-0047).

Appendix A: Proofs

A.1. Importance sampling: theorems 1 and 2

A.1.1. Preliminary calculation

Let $\tilde{q}(dx) = S(x_1^N)(dx) = N^{-1}\sum_{n=1}^{N}\delta_{x_n}(dx)$, and, as a preliminary calculation, take $\varphi \in L_2([0, 1]^d, \lambda_\varphi)$ and

$$|\pi^N(\varphi) - \pi(\varphi)| = \left| \frac{N^{-1}\sum_{n=1}^{N} w(x_n) \varphi(x_n)}{N^{-1}\sum_{n=1}^{N} w(x_n)} - \pi(\varphi) \right|$$

$$\leq \left| \frac{N^{-1}\sum_{n=1}^{N} w(x_n) \varphi(x_n)}{N^{-1}\sum_{n=1}^{N} w(x_n)} - N^{-1} \sum_{n=1}^{N} w(x_n) \varphi(x_n) \right|$$

$$+ \left| N^{-1} \sum_{n=1}^{N} w(x_n) \varphi(x_n) - q(w\varphi) \right|$$

$$\leq \frac{N^{-1} \sum_{n=1}^{N} w(x_n) |\varphi(x_n)|}{N^{-1}\sum_{n=1}^{N} w(x_n)} |q(w) - \tilde{q}(w)|$$

$$+ |\tilde{q}(w\varphi) - q(w\varphi)| .$$

(9)

We shall use this inequality in the two following proofs.
A.1.2. Proof of theorem 1

Take \( \varphi \equiv 1_B \) for \( B \in \mathcal{B}_{[0,1]^d} \) in inequality (9). Consider the first term in the inequality. The ratio is bounded by 1, and (since \( w \) is bounded) \( |q(w) - \hat{q}(w)| \to 0 \) by the portmanteau lemma (van der Vaart (2007), lemma 2.2). Now consider the second term.

We follow essentially the same steps as in van der Vaart (2007), lemma 2.2. Without loss of generality we assume that \( q(dx) \) is a continuous probability measure (the same argument as in van der Vaart (2007) is used for the general case).

Let \( \epsilon > 0 \) and take \( J \in \mathcal{B}_{[0,1]^d} \) such that \( q(J^c) \leq \epsilon \). Since \( J \) is compact, \( w(\cdot) \) is uniformly continuous on \( J \).

Let \( \eta > 0 \) be such that \( \|x - y\| \leq \eta \Rightarrow |w(x) - w(y)| \leq \epsilon, \forall (x,y) \in J^c \). Let \( \{J_k\}_{k=1}^m \) be a split of \( J \) into a finite collection of \( m \) closed hyperrectangles with radius (at most) \( \eta \). Let

\[
g(x) = \sum_{k=1}^m w(x_k) \mathbb{1}_{J_k}(x)
\]

and note that \( |w(x) - g(x)| \leq 2\epsilon, \forall x \in J \). Thus

\[
\left| \int_B w(x) \{\hat{q}(dx) - q(dx)\} \right| \leq \int_B \{w(x) - g(x)\} \hat{q}(dx) + \int_B g(x) \{\hat{q}(dx) - q(dx)\} + \int_B \{w(x) - g(x)\} q(dx)
\]

where for the first term we have

\[
\left| \int_B \{w(x) - g(x)\} \hat{q}(dx) \right| \leq \int_{B \cap J} \{w(x) - g(x)\} \hat{q}(dx) + \int_{J \cap B} w(x) \hat{q}(dx) \\
\leq 2^d \epsilon + \|w\|_{\infty} \hat{q}(J^c) \\
\leq \epsilon(2^d + 2\|w\|_{\infty})
\]

as \( \hat{q}(J^c) \) converges to \( q(J^c) \), and thus \( \hat{q}(J^c) \leq 2\epsilon \) for \( N \) sufficiently large; and for the second term

\[
\left| \int_B g(x) \{\hat{q}(dx) - q(dx)\} \right| \leq \int_{B \cap J} g(x) \{\hat{q}(dx) - q(dx)\} + \int_{J \cap B} g(x) \{\hat{q}(dx) - q(dx)\} \\
\leq \|\hat{q}(dx) - q(dx)\|_{E} \sum_{k=1}^m w(x_k).
\]

Finally, for the third term:

\[
\left| \int_B \{w(x) - g(x)\} q(dx) \right| \leq \int_{B \cap J} \{w(x) - g(x)\} q(dx) + \int_{J \cap B} w(x) q(dx) \\
\leq \epsilon(2^d + \|w\|_{\infty}).
\]

Putting inequalities (10)–(12) together shows that, for all \( B \in \mathcal{B}_{[0,1]^d} \),

\[
\left| \int_B w(x) \{\hat{q}(dx) - q(dx)\} \right| \leq \epsilon(2^{d+1} + 3\|w\|_{\infty}) + \|\hat{q}(dx) - q(dx)\|_{E} \sum_{k=1}^m w(x_k) \\
\leq \epsilon(2^{d+2} + 3\|w\|_{\infty})
\]

for \( N \) sufficiently large (as \( \|\hat{q}(dx) - q(dx)\|_{E} \to 0 \), which concludes the proof of theorem 1.

A.1.3. Proof of theorem 2

We prove first \( L_1 \)-convergence (the first part of theorem 2). We start again from inequality (9), but for any \( \varphi \in L_1([0,1]^d, \lambda_d) \). For the second term, by Jensen’s inequality

\[
\mathbb{E} |\hat{q}(w\varphi) - q(w\varphi)| \leq \text{var} \{\hat{q}(w\varphi)\}^{1/2} = O\{r(N)^{1/2}\}
\]

by assumption. For the first term, using the Cauchy–Schwarz inequality \( \mathbb{E}||CD|| \leq (\mathbb{E}|C|^2 \mathbb{E}|D|^2)^{1/2} \) with
\[
C = \frac{N^{-1} \sum_{n=1}^{N} w(x^n) |\varphi(x^n)|}{N^{-1} \sum_{n=1}^{N} w(x^n)},
\]

\[
D = q(w) - \hat{q}(w),
\]

we have \(\mathbb{E}[D^2]^{1/2} = O(r(N)^{1/2})\), and what remains to prove is that \(\mathbb{E}[C^2] = O(1)\).

From inequality (13), and under assumption (b), we see that there exists \(N\) such that with probability \(1\) \(N^{-1} \sum_{n=1}^{N} w(x^n) \geq \frac{1}{2}\) as soon as \(N \geq N_\epsilon\). Under assumption (a), a bound similar to inequality (13) is easily obtained by replacing \(x^{1:N}\) with \(u^{1:N}\) and observing that \(w \circ F_q^{-1}\) is continuous and bounded. Thus, for \(N\) sufficiently large,

\[
\mathbb{E}[C^2] \leq 4 \mathbb{E}\left[ \left\{ N^{-1} \sum_{n=1}^{N} w(x^n) |\varphi(x^n)| \right\}^2 \right] \leq O(r(N)) + 4\pi(|\varphi|^2) = O(1).
\]

We now prove \(L_2\)-convergence (the second part of theorem 2):

\[
\text{var}\{\pi^N(\varphi)\} \leq \text{var}\{\pi^N(\varphi) - \hat{q}(w\varphi)\}^{1/2} + \text{var}\{\hat{q}(w\varphi)\}^{1/2},
\]

with \(\text{var}\{\hat{q}(w\varphi)\} = O\{r(N)\}\) by assumption, and, for the first term,

\[
\mathbb{E}\{\pi^N(\varphi) - \hat{q}(w\varphi)\}^2 = \mathbb{E}\left[ \left( \sum_{n=1}^{N} \{W_n - N^{-1} w(x^n)\} \varphi(x^n) \right)^2 \right]
\]

\[
= \mathbb{E}\left[ \left( 1 - N^{-1} \sum_{n=1}^{N} w(x^n) \right) \left\{ \sum_{n=1}^{N} W_n \varphi(x^n) \right\}^2 \right]
\]

\[
\leq 4 \mathbb{E}\left[ (1 - \hat{q}(w))^2 \hat{q}(w\varphi)^2 \right]
\]

for \(N\) sufficiently large, using the same argument as above (as \(\hat{q}(w) \rightarrow 1\)). Then

\[
\mathbb{E}\{1 - \hat{q}(w)\}^2 \hat{q}(w\varphi)^2 \leq \mathbb{E}\{1 - \hat{q}(w)\}^2 \{\hat{q}(w\varphi) - \pi(\varphi)\}^2 - \pi(\varphi)^2 \mathbb{E}\{1 - \hat{q}(w)\}^2
\]

\[
+ 2\pi(\varphi) \mathbb{E}[|\hat{q}(w\varphi)| \{1 - \hat{q}(w)\}^2]
\]

where, for the second term, \(\mathbb{E}\{1 - \hat{q}(w)\}^2 = \text{var}\{\hat{q}(w)\} = O\{r(N)\}\), for the first term

\[
\mathbb{E}\{1 - \hat{q}(w)\}^2 \{\hat{q}(w\varphi) - \pi(\varphi)\}^2 \leq \mathbb{E}\{1 - \hat{q}(w)\}^2 \{\hat{q}(w\varphi) - q(w\varphi)\}^2
\]

\[
= (1 + \|w\|_\infty)^2 \text{var}\{\hat{q}(w\varphi)\}
\]

and finally for the third term

\[
\mathbb{E}[|\hat{q}(w\varphi)| \{1 - \hat{q}(w)\}^2] \leq \mathbb{E}[|\hat{q}(w\varphi) - q(w\varphi)| \{1 - \hat{q}(w)\}^2 + |q(w\varphi)| \text{var}\{\hat{q}(w)\}]
\]

with

\[
\mathbb{E}[|\hat{q}(w\varphi) - q(w\varphi)| \{1 - \hat{q}(w)\}^2] \leq (1 + \|w\|_\infty) \text{var}\{\hat{q}(w\varphi)\}^{1/2} \text{var}\{\hat{q}(w)\}^{1/2}
\]

\[
= O\{r(N)\},
\]

which concludes the proof.

For subsequent uses (see the proof of theorem 6), we note that these computations imply, for \(N\) sufficiently large,

\[
\text{var}\{\pi^N(\varphi)\} \leq 2(1 + \|w\|_\infty) \text{var}\{\hat{q}(w\varphi)\}^{1/2} + \{1 + 2|\pi(\varphi)|\} \text{var}\{\hat{q}(w)\}^{1/2},
\]

\[
|\pi^N(\varphi) - \pi(\varphi)| \leq \text{var}\{\hat{q}(w\varphi)\}^{1/2} + 2 \text{var}\{\hat{q}(w)\}^{1/2} \text{var}\{\hat{q}(w\varphi)\} + \pi(|\varphi|)^2)^{1/2}.
\]

(14)
A.2. Hilbert curve and discrepancy: theorems 3 and 4

The proofs in this section rely on the properties of the Hilbert curve that were laid out in Section 2.3 and the corresponding notation.

A.2.1. Proof of theorem 3

We first show that \( \| \pi_h^N - \pi_h \|_E = \sup_{0 \leq a < b \leq 1} | \pi_h^N ([a, b]) - \pi_h ([a, b]) | \). Because \( \pi_h \) is a continuous probability measure on \([0, 1]\), the result is obvious if \( \pi_h^N \) is continuous as well. Let \( 0 \leq a < b < 1 \) be such that \( b \) is a discontinuity point of \( F_h \) and let \( \delta > 0 \) be sufficiently small that \( \pi_h^N ([a, b]) = \pi_h^N ([a, b + \delta]) \) and \( b + \delta < 1 \). Then,

\[
| \pi_h^N ([a, b]) - \pi_h ([a, b]) | - | \pi_h^N ([a, b + \delta]) - \pi_h ([a, b + \delta]) | \leq \pi_h ([b, b + \delta]).
\]

By the bimeasure property of the Hilbert curve, the set \( H([b, b + \delta]) \) has Lebesgue measure \( \delta \) in \([0, 1]^d\) and, therefore, \( \pi_h ([b, b + \delta]) = \pi (H([b, b + \delta])) \leq \| \pi \|_\infty \delta \) where \( \| \pi \|_\infty < \infty \) by assumption. Hence, for all \( \epsilon > 0 \) sufficiently small,

\[
\| \pi_h^N - \pi_h \|_E - \sup_{0 \leq a < b \leq 1} | \pi_h^N ([a, b]) - \pi_h ([a, b]) | \leq \epsilon.
\]

To prove theorem 3 note that the above computations imply that

\[
\| \pi_h^N - \pi_h \|_E \leq 2 \sup_{b \in (0, 1)} | \pi_h^N ([0, b]) - \pi_h ([0, b]) |.
\]

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

\[
\| \pi_h^N - \pi_h \|_E \leq 2 \sup_{b \in (0, 1)} | \pi_h^N ([0, b]) - \pi_h ([0, b]) |.
\]

with \( J = H(I) \). Since \( I \) is the union of \( k^* \) intervals in \( T_m^d \), \( J \) is the union of \( k^* \) hypercubes in \( S_m^d \), and therefore (using a similar argument to that above and Neiderreiter (1992), proposition 2.4),

\[
| \pi_h^N (J) - \pi_h (J) | \leq c | \pi_h - \pi_h |_E \leq 2^{dn} r(N)
\]

for a constant \( c \) and where \( r(N) = \| \pi_h - \pi_h \|_E \).

For the second term of equation (16), by the properties of the Hilbert curve,

\[
| \pi_h^N (2^{-dn} k, b) - \pi_h (2^{-dn} k, b) | \leq \pi_h [ I^d_m (k) ] + \pi_h \{ I^d_m (0) \} = \pi_h \{ S_m^d (k) \} + \pi_h \{ S_m^d (0) \} \leq 2 \pi \{ S_m^d (0) \} + r(N) = O \{ 2^{-dn} \} \}
\]

where the last inequality comes from the fact that \( \pi (x) \) is a bounded density.

In the case \( b < 2^{-dn} \), similar computations show that

\[
| \pi_h^N (J) - \pi_h (I) | \leq \pi_h [ I^d_m (0) ] + \pi_h \{ I^d_m (0) \} = O \{ 2^{-dn} \} \}
\]

To conclude, we choose \( m \) so that \( 2^{-dn} = O \{ r(N)^{1/2} \} \), which gives

\[
\sup_{b \in (0, 1)} | \pi_h^N ([0, b]) - \pi_h ([0, b]) | = O \{ r(N)^{1/2} \}.
\]

Finally, since replacing \([0, b] \) by \([0, a] \) changes nothing to the proof of the result above, we may conclude that \( \sup_{I \in \mathcal{B}_{[0,1]}} | \pi_h^N (I) - \pi_h (I) | = O \{ r(N)^{1/2} \} \).
A.2.2. Proof of theorem 4

A.2.2.1. Preliminary computations. The proof of theorem 4 is based on Hlawka and Mück (1972), Satz 2. Compared with their method the main technical difficulty comes from the fact that the Rosenblatt transformation $F_{p_h^N \otimes K_h}$ is not continuous because $\pi_h^N$ is a weighted sum of Dirac measures. To control the ‘jumps’ of the inverse Rosenblatt transformation $F_{p_h^N \otimes K_h}^{-1}$ that are introduced by the discontinuity of $\pi_h^N$, we first prove the following lemma.

Lemma 2. Consider the set-up of theorem 4. For $n \in 1:N$, let $h_n^1 = H(x_n^1)$ and assume that the points $h_n^1$ are labelled so that $n < m \Rightarrow h_n^1 < h_m^1$. (The inequality is strict because, by assumption (b) of theorem 4, the points $x_1^N$ are distinct.) Without loss of generality, assume that $h_1^1 > 0$ and let $h_0^1 = 0$. Then, as $N \to \infty$,

$$\max_{n \in 1:N} |h_n^1 - h_0^1| \to 0.$$ 

To prove lemma 2, let $J_N = [h_n^1 - 1, h_n^1]$ where $|h_n^1 - h_{n-1}^1| = \max_{n \in 1:N} |h_n^1 - h_0^1|$. Since $J_N$ contains at most two points, we have

$$\pi_h(J_N) \leq \pi_h^N(J_N) + r_2(N) \leq 2r_1(N) + r_2(N)$$

where $r_1(N) = \max_{n \in 1:N} W_n^c$ and $r_2(N) = \|\pi_h^N - \pi_h\|_E$; note that $r_1(N) \to 0$ by assumption (b) of theorem 4 and $r_2(N) \to 0$ by assumption (c) of theorem 4. Therefore, $\pi_h(J_N) \to 0$ as $N \to \infty$.

Assume now that $\max_{n \in 1:N} |h_n^1 - h_{n-1}^1| \not= 0$. Then, this means that there is an $\epsilon \in (0, 1)$ such that, for all $N > 1$, there is an $N^* \geq N$ for which $\lambda_1(J_{N^*}) \geq \epsilon$. Assume first that $J_{N^*} \subseteq [0, 1 - \epsilon/2]$. In that case, we have $\pi_h(J_{N^*}) \geq c_i$ for a constant $c_i > 0$. Indeed, by the continuity of the Hilbert curve, the set $H([0, 1 - \epsilon/2])$ is compact and, therefore, $\forall x \in H([0, 1 - \epsilon/2]), \pi(x) \geq \pi(\epsilon)$ for a constant $\pi(\epsilon) > 0$ because the density $\pi(x)$ is continuous and strictly positive. Therefore, if $J_{N^*} \subseteq [0, 1 - \epsilon/2]$, we have

$$\pi_h(J_{N^*}) = \pi \{H(J_{N^*})\} \geq \pi(\epsilon) \lambda_1(J_{N^*}) = \pi(\epsilon) \lambda_1(J_{N^*}) \geq \epsilon \pi(\epsilon)$$

where the second equality uses the bimeasure property of the Hilbert curve.

Assume now that $J_{N^*} \not\subseteq [0, 1 - \epsilon/2]$. Write $J_{N^*} = [a_{N^*}, b_{N^*}]$ and note that, since $\lambda_1(J_{N^*}) \geq \epsilon$, we have $a_{N^*} < 1 - \epsilon$ and therefore

$$\pi_h(J_{N^*}) = \pi_h \left( [a_{N^*}, 1 - \epsilon/2] \right) + \pi_h \left( (1 - \epsilon/2, b_{N^*}] \right) \geq \left( 1 - \epsilon/2, a_{N^*} \right) \pi(\epsilon) \geq \epsilon \pi(\epsilon).$$

Thus, this shows that, if $\max_{n \in 1:N} |h_n^1 - h_{n-1}^1| \not= 0$, then there is an $\epsilon \in (0, 1)$ such that $\limsup_{N \to \infty} \pi_h(J_N) \geq (\epsilon \pi(\epsilon))/2 > 0$. This contradicts the fact that $\pi_h(J_N) \to 0$ as $N \to \infty$ and the proof is complete.

A.2.2.2. Proof of theorem 4. We use the shorthand $\alpha_N(B) = S(U_1^N)(B)$ for any set $B \subseteq [0, 1]^{1+d_2}$. We have

$$\|S(P_h^N) - \pi_h^N \otimes K_h\|_E = \sup_{B \subseteq B_{[0,1]^{1+d_2}}} |\alpha_N(E^N(B)) - \lambda_{1+d_2}(E^N(B))|$$

where

$$B_{[0,1]^{1+d_2}} = \{B = [a, b] \subseteq B_{[0,1]^{1+d_2}} : \min_{n \in 1:N} h(x_n^0) \leq F_{\pi_h^N}(b_1) \leq \max_{n \in 1:N} h(x_n^0)\},$$

and where, for an arbitrary set $B_{[0,1]^{1+d_2}}$, we use the shorthand $E^N(B)$ for the set

$$\{(a_1, b_2) \in [0, 1]^{1+d_2} : F_{\pi_h^N}(a_1) \leq u_1 \leq F_{\pi_h^N}(b_1), u_2 \in F_{K_h}(F_{\pi_h^N}^{-1}(u_1), [a', b']).\}$$

Let $\mathcal{P}$ be a partition of $[0, 1]^{1+d_2}$ in $L_{1+d_2}$ congruent hyperrectanges $W$ of size $L_d \times L_d \times \ldots \times L_d$ where $L \geq 1$ is an arbitrary integer. Let $B = [a_1, b_1] \times [a', b'] \subseteq B_{[0,1]^{1+d_2}}$, $U_1$ the set of the elements of $\mathcal{P}$ that are strictly in $E^N(B)$, $U_2$ the set of elements $W \in \mathcal{P}$ such that $W \cap \partial(E^N(B)) \not= \emptyset$, $U_1 = U_1 \cup U_1$, $U_2 = U_2 \cup U_2$ and $U'_1 = E^N(B) \setminus U_1$ so that

$$\alpha_N(E^N(B)) - \lambda_{1+d_2}(E^N(B)) = \alpha_N(U_1) - \lambda_{1+d_2}(U_1) + \alpha_N(U'_1) - \lambda_{1+d_2}(U'_1).$$
To bound $\alpha_N(U'_1) - \lambda_1 + d_2(U'_1)$, note that we can cover $U'_1$ with sets in $U_2$; hence
\[
\alpha_N(U'_1) - \lambda_1 + d_2(U'_1) \leq \alpha_N(U_2)
\]
and
\[
\alpha_N(U'_1) - \lambda_1 + d_2(U'_1) \geq -\lambda_1 + d_2(U_2)
\]
so, by the definition of $D(u^{1,N})$,
\[
|\alpha_N(U'_1) - \lambda_1 + d_2(U'_1)| \leq |\alpha_N(U_2) - \lambda_1 + d_2(U_2)| + \lambda_1 + d_2(U_2) \\
\leq \#U_2 \{D(u^{1,N}) + L^{-(d_1+d_2)}\}.
\]
We therefore have
\[
|\alpha_N\{E^N(B)\} - \lambda_1 + d_2\{E^N(B)\}| \leq |\alpha_N(U_1) - \lambda_1 + d_2(U_1)| + \#U_2 \{D(u^{1,N}) + L^{-(d_1+d_2)}\} \\
\leq L^{d_1+d_2}D(u^{1,N}) + \#U_2 \{D(u^{1,N}) + L^{-(d_1+d_2)}\}.
\]

The rest of the proof is dedicated to bounding $\#U_2$, the number of hyperrectangles in $\mathcal{P}$ that are required to cover $\partial\{E^N(B)\}$. To that effect, first note that, using the continuity of $F_K$ and the fact that $B$ and $E^N(B)$ are closed sets, we can easily show that $E^N(\partial(B)) \subset \partial\{E^N(B)\}$. Let $\#U_2^{(1)}$ and $\#U_2^{(2)}$ be respectively the number of hyperrectangles in $\mathcal{P}$ that we need to cover $E^N(\partial(B))$ and to cover $\partial(B) := \partial\{E^N(B)\} \setminus E^N(\partial(B))$. Hence, $\#U_2 \leq \#U_2^{(1)} + \#U_2^{(2)}$ and we now bound $\#U_2^{(2)}$, $i \in 1:2$.

To bound $\#U_2^{(2)}$ we first cover $\partial(B)$ with hyperrectangles belonging to a partition $\mathcal{P}'$ of the set $[0,1]^{1+d_2}$. We construct $\mathcal{P}'$ as a partition of the set $[0,1]^{1+d_2}$ into hyperrectangles $W'$ of size $L'^{-d_1} \times L'^{-1} \times \ldots \times L'^{-1}$ such that, for all points $(h_1, x_2)$ and $(h'_1, x'_2)$ in $W'$, we have
\[
\|F_k(h_1, x_2) - F_k(h'_1, x'_2)\|_\infty = \|F_k(H(h_1), x_2) - F_k(H(h'_1), x'_2)\|_\infty \leq L'^{-1}
\]
and
\[
|F_{s_2}(h_1) - F_{s_2}(h'_1)| \leq L'^{-d_1}.
\]

Let $L' = 2^m$ for an integer $m \geq 0$, so that $h_1$ and $h'_1$ are in the same interval $I^d_{m_1}(k) \in I_{m_1}^d$, and $H(h_1)$ and $H(h'_1)$ belong to the same hypercube in $S_{m_1}^d$. Let $L_K$ be the Lipschitz constant of $F_K$; then
\[
\|F_k\{H(h_1), x_2\} - F_k\{H(h'_1), x'_2\}\|_\infty \leq C_k \{\|x_2 - x'_2\|_\infty \lor \|H(h_1) - H(h'_1)\|_\infty\} \\
\leq C_K L'^{-1}
\]
and condition (17) is verified as soon as $L' \geq C_K L$. Let us now look at condition (18). We have
\[
|F_{s_2}(h_1) - F_{s_2}(h'_1)| \leq 2\|F_{s_2} - F_{s_2}\|_\infty + |F_{s_2}(h_1) - F_{s_2}(h'_1)| \\
\leq 2r_2(N) + |F_{s_2}(h_1) - F_{s_2}(h'_1)|
\]
where, as in the proof of lemma 2, $r_2(N) = \|\pi^N - \pi\|_E$. Since $h_1$ and $h'_1$ are in the same interval $I^d_{m_1}(k) \in I_{m_1}^d$,
\[
|F_{s_2}(h_1) - F_{s_2}(h'_1)| \leq \pi_2\{I^d_{m_1}(k)\} = \pi\{S_{m_1}^d(k)\} \leq \|\pi\|_\infty / L'^{d_1}
\]
as $\pi$ is bounded. To obtain both condition (17) and condition (18), we can take $L' = 2^m$ to be the smallest power of 2 such that $L' \geq k_N L$ where
\[
k_N = C_K + \left\{\frac{\|\pi\|_\infty}{1 - 2r_2(N)L'^{d_1}}\right\}^{1/d_1}
\]
which implies that we assume from now on that $L'^{-d_1} \geq 4r_2(N)$ for $N$ sufficiently large.

Let $R \in \partial B$ be a $d_2$-dimensional face of $B$ and let $\mathcal{R}$ be the set of hyperrectangles $W' \in \mathcal{P}'$ such that $R \cap W' \neq \emptyset$. Note that $\#\mathcal{R} \leq L'^{d_1+d_2-1} \leq (2k_N L'^{d_1+d_2-1})$. For each $W' \in \mathcal{R}$, take a point $r_{W'} = (r_{W'}^1, r_{W'}^2) \in R \cap W'$ and define
\[
r_{W'} = (r_{W'}^1, r_{W'}^2) = F_{k_N \otimes K_0}(r_{W'}) \in E^N(R).
\]
Let $\mathcal{R}$ be the collection of hyperrectangles $\bar{W}$ of size $4L'^{-d_1} \times 2L'^{-1} \times \ldots \times 2L'^{-1}$ and having point $r_{W'}$, $W' \in \mathcal{R}$, as middle point.
For an arbitrary \( u = (u_1, u_2) \in E^N(R) \), let \( h_1 = a_1 \lor F_{-1}^{-1}(u_1) \) and \( x_2 = F_{-1}^{-1}(h_1, u_2) \). Since \( x = (h_1, x_2) \in R \), \( x \) is in one hyperrectangle \( W \in R \). Hence, using conditions (17) and (18),
\[
|u_1 - r_i^W| \leq |F_{x_2}^{-1}(h_1) - F_{x_2}^{-1}(r_1^W)| + |u_1 - F_{x_2}^{-1}(h_1)| \leq L^{-d_i} + r_1(N),
\]
where, as in the proof of lemma 2, \( r_1(N) = \max_{a \in R} W_n, \) and
\[
\|u_2 - r_2^W\|_\infty = \|F_{x_2}(h_1, x_2) - F_{x_2}(r_1^W, r_2^W)\|_\infty \leq L^{-1}.
\]
Assume from now on that \( L^{-d_i} \geq r_1(N) + 4r_2(N) \). Then, this shows that \( u \) belongs to the hyperrectangle \( \tilde{W} \in R \) with centre \( \tilde{r}^W \) so \( E^N(R) \) is covered by at most \( \#R = \#R \leq (2k_0L)^{d_1 + d_2 - 1} \) hyperrectangles \( \tilde{W} \in R \).
To go back to the initial partition of \([0, 1)^{d_1 + d_2} \) with hyperrectangles in \( P \), note that every hyperrectangle in \( R \) is covered by at most \( e^* \) hyperrectangles in \( P \) for a constant \( e^* \). Finally, since the set \( \partial B \) is made of the union of \( 2(d_1 + 1) d_2 \)-dimensional faces of \( B \), we have
\[
\#U_2^{(1)} \leq c_N L^{d_1 + d_2 - 1}
\]
where \( c_N = 2(d_1 + 1)(2k_0L)^{d_1 + d_2 - 1} e^* \).

We now consider the problem of bounding \( \#U_2^{(2)} \), the number of hyperrectangles in \( P \) that we need to cover the set \( P(B) = \partial(E^N(B)) \setminus E^N(\partial(B)) \). Note that \( P(B) \) contains the boundaries of the set \( E^N(B) \) that are due to the discontinuities of \( F_{x_2}^{-1} \) at \( a \).
To that effect, we show that there is a finite collection \( \{D_{m,s}\}_{m=1}^{N} \) of sets in \( B^N \) such that, for any \( u = (u_1, u_2) \in P(B) \), there is an \( m^* \in 1 : k \) and a point \( \tilde{u} = (\tilde{u}_1, \tilde{u}_2) \in E^N(\partial(D_{m,s})) \) which verifies that \( \tilde{u}_1 = u_1 \) and \( \|u_2 - \tilde{u}_2\|_\infty \leq C_3 r(N)^{1/d_1} \) for a constant \( C_3 \) and where \( r(N) = \max_{a \in R} W_n \); that \( r_1(N) \rightarrow 0 \) as \( N \rightarrow \infty \) by lemma 1. Hence, by taking \( L \) sufficiently small (i.e. such that \( L^{-d} \geq C_3 r(N)^{1/d_1} \)), we have \( \#U_2^{(2)} \leq \sum_{m=1}^{N} \#U_2^{(1)} \) where \( \#U_2^{(2)} \) is the number of hyperrectangles in \( P \) that we need to cover \( E^N(\partial(D_{m,s})) \). Then, because the bound that we derived above for the number of these hyperrectangles required to cover \( E^N(\partial(B)) \) is uniform in \( B \in B^N \), we can conclude by using inequality (19) that \( U_2^{(2)} \leq k c_0 L^{d_1 + d_2 - 1} \).

To construct the collection \( \{D_{m,s}\}_{m=1}^{N} \), let \( u = (u_1, u_2) \in P(B) \), i.e. \( u_1 = F_{x_2}(h_i^1) \) for an \( n^* \in 1 : N \) and \( u_2 = F_{x_2}(h_i^2, x^* ) \) with \( x^* \in (a, b)^c \). By the definition of the boundary of a set, for any \( \epsilon > 0 \) there is a \( v = (v_1, v_2) \notin E^N(B) \) such that \( \|u - v\|_\infty \leq \epsilon \). Let \( \epsilon > 0 \) and assume that the point \( v = (u_1 - v_1, u_2) \) verifies this condition, i.e. \( u_2 \notin F_{x_2}(h_i^2, [a, b]) \), \( n^* > 1 \). (The case \( u_1 = (u_1 + \epsilon, u_2) \) is treated in a similar way: just replace \( n^* \) by \( n^* + 1 \) in what follows.)

We now show that there is a set \( B_2 \in B^N \) and a point \( \tilde{u} = (u_1, u_2) \in E^N(\partial(B)) \) such that \( \|u_2 - \tilde{u}_2\|_\infty \leq C_3 r(N)^{1/d_1} \) for a constant \( C_3 \). We consider the set \( B_2 = \{a_1, b_1\} \times [a^N, b^N] \) where \( a^N < b^N \in [0, 1)^{d_2} \). To construct \([a^N, b^N]\), we write \( F_i(h_1, x_{1:i-1}, x_i) \) the \( i \)th co-ordinate of \( F_{x_2}(h_1, x) \) (with the natural convention \( F_i(h_1, x_{1:i-1}, x_i) = F_i(h_1, x_{1:i}) \) when \( i = 1 \)).

Let \( i^* \) be the smallest index \( i \in 1 : d_2 \) such that \( u_2 \notin F_i(h^*_{-1}, x_{1:i-1}, x_i) \) \( \forall x \in [a, b] \). Then, for \( i \in 1 : i^* \), set \( \tilde{u}_2 = u_2 \) and \( \tilde{x}_i = x_i^* \), whereas, for \( i \in 1 : i^* \), we set \( a_i^N = a_i^* \) and \( b_i^N = b_i^* \).

To choose \( u_2^* \) and \( \tilde{x}_i^* \) we proceed as follows: if \( F_i(h^*_{-1}, x_{1:i-1}, a_i^*) < u_2^* \) we take \( u_2^* = F_i(h_i^1, x_{1:i-1}, a_i^*) \) and \( \tilde{x}_i^* = a_i^* \) so that, noting \( C_H \) the Hölder constant of \( H \),
\[
0 \leq u_2^* - u_2^* \leq F_i(h_i^1, x_{1:i-1}, a_i^*) - F_i(h_i^1, x_{1:i-1}, b_i^*) \leq C_K C_H r_3(N)^{1/d_1}
\]
as required; if \( F_i(h^*_{-1}, x_{1:i-1}, a_i^*) > u_2^* \), we take \( u_2^* = F_i(h_i^1, x_{1:i-1}, a_i^*) \) and \( \tilde{x}_i^* = a_i^* \) so that
\[
0 \leq u_2^* - u_2^* \leq F_i(h_i^1, x_{1:i-1}, a_i^*) - F_i(h_i^1, x_{1:i-1}, a_i^*) \leq C_K C_H r_3(N)^{1/d_1}
\]
as required.

Then, for \( i \in i^* + 1 : d_2 \), take \( u_2^* = u_2^* \) and \( a_i^N = 0 \). Finally, to construct the right boundaries \( b_i^N, i \in i^* + 1 : d_2 \), we define
\[
\tilde{u}_2^* = \max\{\sup\{v \in F_i(h_i^1, [0, b_i^1])\}\}, \quad i = 1, \ldots, d_2.
\]
Note that \( u_2^* \in (0, 1) \) for all \( i \in 1 : d_2 \). Indeed, the continuity of \( F_i \) and the fact that \( [0, b_i^1] \) is compact imply that
\[
u_i^N := \sup\{v \in F_i(h_i^1, [0, b_i^1])\} \in F_i(h_i^1, [0, b_i^1]).
\]
Then, since \( b_i^1 \in (0, 1) \) and \( F_i \) is strictly increasing with respect to its \( i \)th co-ordinate on \([0, 1) \), we indeed have \( v_i^N \in (0, 1) \) for all \( n \in 1 : N \).
The right boundaries $b_i^N, i \in \{d_2\}$ are then defined recursively as follows:

$$b_i^N = \inf\{c \in [0, 1], g_i(c) \geq u_i^*, i = i^* + 1, \ldots, d_2, \}$$

where

$$g_i(c) = \min_{(h_1, x_{i-1}) \in [a_{i-1}, b_i^N] \times [a_{i-1}, b_i^N]} \tilde{F}_i(h_1, x_{i-1}, c),$$

with $\tilde{F}_i(\cdot)$ the continuous extension of $F_i(\cdot)$ on $[0, 1]^{i+1}$. (Such an extension exists because $F_i$ is Lipschitz continuous.) Because $\tilde{F}_i(h_1, x_{i-1}, c)$ is continuous in $(h_1, x_{i-1}, c)$ and $[a_{i-1}, b_i^N] \times [a_{i-1}, b_i^N] \times [0, 1]$ is compact, the function $g_i$ is continuous on $[0, 1]$ with $g_i(0) = 0$ and $g_i(1) = 1$. Therefore, as $u_i^* \in (0, 1)$, we indeed have $b_i^N \in (0, 1)$ for all $i \in \{d_2\}$, as required.

To show that $a = (u_1, u_2) \in E^{N} [\partial(B^N)]$, note that, by the construction of $b_i^N$, we have, for all $i \in \{d_2\}$,

$$F_i(h_1, x_{i-1}, b_i^N) \geq u_i^* \geq u_2, \quad \forall (h_1, x_{i-1}) \in [a_{i-1}, b_i^N] \times [a_{i-1}, b_i^N].$$

Because the continuity of $F_i(h_1, x_{i-1}, c)$ is continuous in $(h_1, x_{i-1}, c)$ and $[a_{i-1}, b_i^N] \times [a_{i-1}, b_i^N]$ is compact, the function $g_i$ is continuous on $[0, 1]$ with $g_i(0) = 0$ and $g_i(1) = 1$. Therefore, as $u_i^* \in (0, 1)$, we indeed have $b_i^N \in (0, 1)$ for all $i \in \{d_2\}$, as required.

To prove this result, let $B_1 \times B_2 \subseteq B_{[0, 1]}^{d_1+d_2}, B_2 = [a_2, b_2], a_2$.

$$\int_{B_1 \times B_2} (\pi^N \otimes K - \pi \otimes K)(\mathrm{d}x_1, \mathrm{d}x_2) = \int_{B_1} K(x_1, B_2)(\pi^N - \pi)(\mathrm{d}x_1) = \int_{B_1} \lambda_{d_1} \{F_k(x_1, B_2)\}(\pi^N - \pi)(\mathrm{d}x_1).$$
The function $x_1 \to \lambda_{d_2} \{ F_k(x_1, B_2) \}$ is continuous and bounded and therefore we proceed as in the proof of theorem 1. But, since $\lambda_{d_2} \{ F_k(x_1, B_2) \}$ depends on $(a_2, b_2)$ and we want to take the supremum over $a_2, b_2 \in (0, 1)^{d_2}$, we need to make sure that, on a compact set $J$, for any $\varepsilon > 0$ we can find $\eta > 0$ which does not depend on $(a_2, b_2)$ such that, for $x_1, x'_1 \in J$,

$$\|x_1 - x'_1\|_\infty \leq \eta \Rightarrow |\lambda_{d_2} \{ F_k(x_1, B_2) \} - \lambda_{d_2} \{ F_k(x'_1, B_2) \}| \leq \varepsilon.$$  

To see that this is true, note that $\partial \{ F_k(x_1, B_2) \} = F_k(x_1, \partial B_2)$. Hence, for any point $c \in \partial F_k(x_1, B_2)$ there is a $p \in \partial B_2$ such that $c = F_k(x_1, p)$ and therefore, by the Hölder property of $F_k$, we have

$$\|x_1 - x'_1\|_\infty \leq \eta \Rightarrow \|c - c'\|_\infty \leq C_k \eta^\kappa, \quad c' = F_k(x'_1, p) \in \partial F_k(x'_1, B_2),$$

where $C_k$ and $\kappa$ are respectively the Hölder constant and the Hölder exponent of $F_k$. But, since $\|w\|$ is continuous and bounded (by assumption (b) and the continuity of the Hilbert curve) and therefore, by theorem 1, assumption (c), the Hölder property of $F_k$, we have

$$\|x_1 - x'_1\|_\infty \leq \eta \Rightarrow |\lambda_{d_2} \{ F_k(x_1, B_2) \} - \lambda_{d_2} \{ F_k(x'_1, B_2) \}| \leq \varepsilon.$$  

Let $\tilde{F}_k$ be the continuous extension of $F_k$ on $[0, 1)^{d_1 + d_2}$ (which exists because $F_k$ is Hölder continuous on $[0, 1)^{d_1 + d_2}$). Let $w > 0$, $x \in [0, 1)^{d_1}$ and $a \leq b$, $(a, b) \in [0, 1)^{2d_2}$. Then, define

$$A^+(w, x, a, b) = \{ u \in [0, 1)^{d_2} : \exists p \in \partial[a, b] \text{ such that } \|u - \tilde{F}_k(x, p)\|_\infty \leq C_k w^\kappa \}$$

and, noting $\tilde{F}_k(x_i, x_j)$ the $i$th component of $\tilde{F}_k(x_1, x_2)$, $i \in 1 : d_2$,

$$A^-(w, x, a, b) = \{ u \in \tilde{F}_k(x_i, x_j), i \in 1 : d_2 \} \text{ and } f : [0, 1)^{d_1} \times [0, 1)^{d_1} \times [0, 1)^{2d_2} \to [0, 1]$$

be the mapping

$$(w, x, a, b) \in [0, 1)^{d_1} \times [0, 1)^{d_1} \times [0, 1)^{2d_2} \mapsto f(w, x, a, b) = \lambda_{d_2} \{ A^+(w, x, a, b) \} - \lambda_{d_2} \{ A^-(w, x, a, b) \}.$$  

Note that for a fixed $w$ the function $f(w, \cdot)$ is continuous on $[0, 1)^{d_1} \times B^*$ (as $\tilde{F}_k$ is continuous). Therefore, for all $x_1$ and $x'_1$ in $J$ such that $\|x_1 - x'_1\| \leq \eta$, we have

$$|\lambda_{d_2} \{ F_k(x_1, B_2) \} - \lambda_{d_2} \{ F_k(x'_1, B_2) \}| \leq m(\eta).$$

with

$$m(\eta) := \max_{(x, a, b) \in J \times B'} f(\eta, x, a, b).$$

Because $f$ is continuous and $J \times B^*$ is compact, $m(\eta)$ is continuous so, for any $\varepsilon > 0$, there is an $\eta > 0$ (that depends only on $m(\cdot)$ and is therefore independent of $B_2$) such that $m(\eta) \leq \varepsilon$. This concludes the proof of lemma 3.

We now prove theorem 5. By the result of Hlawka and Mück (1972), Satz 2, and assumption (c), $(x_i^{N})_N$ is such that $\|S(x_i^{N}) - m_0\|_E = O(1)$. In addition, the importance weight function $Q_0(x_0)/m_0(x_0) = G_0(x_0)/m_0(G_0)$ is continuous and bounded by assumption (b). Therefore, $\|Q_0 - Q_0\|_E = O(1)$ by theorem 1.

Assume that the result is true at time $t \geq 0$ and let $P_{t+1,h}^N = (h_i^{N}, x_{t+1}^N)$ where $h_i^N = h(x_i^{(a_i, b_i)})$. Then, the result is true at time $t + 1$ if

$$\|S(P_{t+1,h}^N)_J - Q_{t,h} \otimes m_{t+1,h}\|_E = O(1).$$

To see that, let $G_{t,h}(h_{t-1}, x_t) = G_t \{ H(h_{t-1}, x_t) \}$ and $\Psi_{t+1}$ be the Boltzmann–Gibbs transformation associated with $G_{t+h} \{ h \}$ (see Del Moral (2004), definition 2.3.3). Then, the importance weight function

$$\Psi_{t+1}(Q_{t,h} \otimes m_{t+1}) - Q_{t,h} \otimes m_{t+1} = G_{t+1,h}(h, x_{t+1})$$

is continuous and bounded (by assumption (b) and the continuity of the Hilbert curve) and therefore theorem 1 implies that $\|Q_{t+1,h} - Q_{t,h}\|_E = O(1)$ if condition (20) is verified.

To show condition (20), note that

$$\|S(P_{t+1,h}^N) - Q_{t,h} \otimes m_{t+1,h}\|_E \leq \|S(P_{t+1,h}^N) - \tilde{Q}_{t+1,h}^N\|_E + \|Q_{t+1,h}^N - \tilde{Q}_{t,h} \otimes m_{t+1,h}\|_E.$$  

By the inductive hypothesis, $\|\tilde{Q}_t^N - Q_t\|_E = O(1)$ so, by theorem 3, assumption (c), the Hölder property of the Hilbert curve and lemma 3,

$$\|Q_{t+1,h}^N - \tilde{Q}_{t,h} \otimes m_{t+1,h}\|_E = \|Q_{t,h} \otimes m_{t+1,h} - \tilde{Q}_{t,h} \otimes m_{t+1,h}\|_E = O(1).$$
Finally, note that
\[ W_t^n \leq \frac{\|G_t\|_\infty}{S(P_{t,h}^N)(G_{t,h})} = O(1) \]

because \( S(P_{t,h}^N)(G_{t,h}) = O(N^{-1}) \) by the inductive hypothesis and the fact that \( G_{t,h} \) is continuous and bounded (by assumption (b) and the continuity of the Hilbert curve). Together with the inductive hypothesis and assumptions (a), (c) and (d), this implies that all the assumptions of theorem 4 are verified and therefore \( \|S(P_{t+1,h}^N) - \mathcal{Q}_{t+1,h}\|_E = O(1) \) as required.

A.4. Stochastic bounds
A.4.1. Set-up of the proof of theorem 6
The result is proved by induction. By assumption (b) of theorem 5, the weight function \( Q_0(dx_0)/m_0(dx_0) = G_0(x_0)/m_0(G_0) \) is continuous and bounded. Therefore, the continuity of \( F_{m_0}^{-1} \), the assumptions on \((u_i^{(N)})\) (assumptions (a) and (b)) and theorem 2 give the result at time \( t = 0 \).

Assume that the result is true at time \( t \geq 0 \) and let \( \bar{\mathcal{Q}}_{t+1} = \mathcal{Q}_{t+1}^{(N)}(\varphi) \) where \( \varphi : [0, 1]^d \rightarrow \mathbb{R} \) verifies the conditions of theorem 6. As mentioned previously, iteration \( t + 1 \) of the SQMC algorithm is a QMC importance sampling step from the proposal distribution \( \mathcal{Q}_{t+1,h} \) to the target \( w_{t+1,h}(h, x_{t+1}) \mathcal{Q}_{t+1,h}(d(h, x_{t+1})) \) where

\[ w_{t+1,h}(h, x_{t+1}) := \frac{G_{t+1,h}(h, x_{t+1})}{C_{t+1}^{(N)}} \]

with \( C_{t+1}^{(N)} = \mathcal{Q}_{t+1,h}(G_{t+1,h}) \) and \( G_{t+1,h} \) as in the proof of theorem 5. To bound \( \text{var}(\bar{\mathcal{I}}_{t+1}^{(N)}) \) and \( \mathbb{E}[\bar{\mathcal{I}}_{t+1}^{(N)}(\varphi)] \) we therefore naturally want to use expressions (14) and (15) derived in the proof of theorem 2. To that effect, we need to show that, for \( N \) sufficiently large and almost surely, the assumptions given in theorem 2 on the weight function and on the point set at hand (assumption (b) of theorem 2) are satisfied.

To see that the conditions on the weight function are fulfilled, note first that \( w_{t+1,h}^{(N)} \) is continuous by assumption (b) of theorem 5 and by the continuity of the Hilbert curve. To show that \( w_{t+1,h}^{(N)} \) is almost surely bounded for \( N \) sufficiently large, first note that, by assumption (a), it is clear from the proofs of theorem 3 and of theorem 5 that, for all \( \epsilon > 0 \) and for all \( t \geq 0 \), there is an \( N_{\epsilon,t}^{(N)} \) such that, almost surely,

\[ \|\mathcal{Q}_{t,h} - \mathcal{Q}_{t,h}\|_E \leq \epsilon, \quad \forall N \geq N_{\epsilon,t}^{(N)}. \]

In addition, under the assumptions of theorem 6, \((C_{t+1}^{(N)})^{-1} \) is almost surely bounded above and below away from 0, for \( N \) sufficiently large. Indeed, by lemma 3 and using the H"older property of the Hilbert curve,

\[ \|\mathcal{Q}_{t+1,h}^{(N)} - \mathcal{Q}_{t,h} \otimes m_{t+1,h}\|_E = O(1) \]

and, in particular, under the conditions of theorem 6, for any \( \delta > 0 \), we have, almost surely,

\[ \|\mathcal{Q}_{t+1,h}^{(N)} - \mathcal{Q}_{t,h} \otimes m_{t+1,h}\| \leq \delta \]

(21)

for \( N \) sufficiently large (see the proof of lemma 3 and the proof of theorem 1). Writing \( C_{t+1} = \mathcal{Q}_{t,h} \otimes m_{t+1,h}(G_{t+1}) \), this observation, together with the fact that

\[ |C_{t+1}^{N} - C_{t+1}| = |\mathcal{Q}_{t+1,h}^{(N)}(G_{t+1,h}) - \mathcal{Q}_{t,h} \otimes m_{t+1,h}(G_{t+1})| \]

where \( G_{t+1,h} \) is continuous and bounded (by assumption (b) of theorem 5 and the continuity of the Hilbert curve), implies that, almost surely, \( C_{t+1} + \delta \geq C_{t+1}^{N} \geq C_{t+1} - \delta := c_\delta > 0 \) for \( N \) sufficiently large (the computations are as in the proof of theorem 1). Hence, almost surely, \( \|w_{t+1,h}^{(N)}\|_\infty \leq c_\delta^{-1}\|G_{t+1}\|_\infty \), for \( N \) sufficiently large.

Finally, to show that the point set \( P_{t+1,h}^{(N)} \) (which is defined as in the proof of theorem 5) verifies assumption (b) of theorem 2, note that, from theorem 5 and under the assumptions of theorem 6, for any \( \epsilon > 0 \) there is an \( N_\epsilon > 0 \) such that, almost surely, \( \|S(P_{t+1,h}^{N}) - \mathcal{Q}_{t,h} \otimes m_{t+1,h}\|_E \leq \epsilon \) for all \( N \geq N_\epsilon \). Together with result (21), this shows that, as required, for any \( \epsilon > 0 \) we have, almost surely and for \( N \) sufficiently large,

\[ \|S(P_{t+1,h}^{N}) - \mathcal{Q}_{t+1,h}\|_E \leq \epsilon. \]
A.4.2. Proof of theorem 6: L₂-convergence

Using expression (14) given in the proof of theorem 2, we have for N sufficiently large

\[
\text{var}(\hat{I}_{t+1}^N) \leq [2(1 + c_{\theta}^{-1} ||G_{t+1}||_\infty) \text{var}(S(P_{t+1,h}^N)(w_{t+1,h}^N))]^{1/2} + 1/2 |E[S(P_{t+1,h}^N)(\varphi w_{t+1,h}^N)]| \text{var}(S(P_{t+1,h}^N)(\varphi w_{t+1,h}^N))^{1/2}.]
\]  

(22)

We first bound \(\text{var}(S(P_{t+1,h}^N)(w_{t+1,h}^N))\). Let \(\mathcal{F}_t^N\) be the \(\sigma\)-algebra that is generated by the point set \((h_{1:N}^L, x_{1:t}^N)\). Then, by assumption (b),

\[
\text{var}(S(P_{t+1,h}^N)(w_{t+1,h}^N)|\mathcal{F}_t^N) \leq C^* r(N)^2 \sigma^2_N
\]

with \(C^*\) as in the statement of theorem 6 and \(\sigma^2_N \leq ||w_{t+1,h}^N||_\infty \leq c_{\theta}^{-1} ||G_{t+1}||_\infty\) almost surely and for \(N\) sufficiently large. Therefore, since \(E[S(P_{t+1,h}^N)(w_{t+1,h}^N)|\mathcal{F}_t^N] = 1\), we have

\[
\text{var}(S(P_{t+1,h}^N)(w_{t+1,h}^N)) = O\{r(N)\}.
\]  

(23)

Next, we need to bound \(\text{var}(S(P_{t+1,h}^N)(\varphi w_{t+1,h}^N))\). Note that

\[
\mathcal{Q}_{t+1}^N \{(C_{t+1})^{-2} \varphi^2 G_{t+1}^2 \} \leq \frac{1}{(C_{t+1})^2} ||G_{t+1}||^2_\infty \mathcal{Q}_{t}^N \{m_{t+1}(\varphi^2)\},
\]

where the last factor is almost surely finite for all \(N\). Indeed, since \(\varphi \in L_2(\mathcal{X}, \mathcal{Q}_{t+1})\), \(m_{t+1}(\varphi^2)(x_t)\) is finite for almost all \(x_t \in \mathcal{X}\) and the integral with respect to \(\mathcal{Q}_{t}^N\) is a finite sum. Hence, for all \(N, \varphi \in L_2(\mathcal{X}, \mathcal{Q}_{t+1}^N, m_{t+1})\) almost surely so that, by assumption (b), we have almost surely

\[
\text{var}(S(P_{t+1,h}^N)(\varphi w_{t+1,h}^N)|\mathcal{F}_t^N) \leq C^* r(N)^2 \sigma^2_{N, \varphi}
\]

where, with probability 1 and for \(N\) sufficiently large, \(\sigma^2_{N, \varphi} \leq c_{\theta}^{-1} ||G_{t+1}||^2_\infty \mathcal{Q}_{t+1}^N\{\varphi^2)\}.\) We now need to show that \(E[\mathcal{Q}_{t+1}^N(\varphi^2)]\) is bounded.

To establish this, we prove that, for all \(t \geq 0\) and for \(N\) sufficiently large, we have, \(\forall f \in L_1(\mathcal{X}, \mathcal{Q}_{t}^N \otimes m_{t+1})\),

\[
E[\mathcal{Q}_{t+1}^N(f)] \leq c_{t+1} \mathcal{Q}_{t}^N \otimes m_{t+1}(|f|)
\]  

(24)

for constant \(c_{t+1}\).

Equation (24) is true for \(t = 0\). Indeed, let \(f \in L_1(\mathcal{X}, \mathcal{Q}_{0} \otimes m_{1})\) and note that, under the conditions of theorem 6, almost surely and for \(N\) sufficiently large, \(\{S(x_{0:N}^L)(G_0)\}^{-1} \leq \tilde{c}_0 < \infty\) for a constant \(\tilde{c}_0\). Hence, for \(N\) sufficiently large (with the convention \(G_t(x_{t-1}, x_t) = G_0(x_t)\) if \(t = 0\)),

\[
E[\mathcal{Q}_{t+1}^N(f)] = E\left[\{S(x_{0:N}^L)(G_t)\}^{-1} \frac{1}{N} \sum_{n=1}^{N} G_0(x_n^L) \int_{\mathcal{X}} f(x_n^L, x_t) \delta_{n} \otimes m_{t+1}(|f|)\right]
\]

where \(c_0 = \tilde{c}_0 m_0(G_0)\). Assume that inequality (24) is true for \(t \geq 0\) and note that, under the conditions of theorem 6, almost surely and for \(N\) sufficiently large, \(\{S(P_{t+1}^N)(G_t)\}^{-1} \leq \tilde{c}_t < \infty\) for a constant \(\tilde{c}_t\). Then, for \(N\) sufficiently large (with the convention \(G_t(x_{t-1}, x_t) = G_0(x_t)\) if \(t = 0\)),

\[
E[\mathcal{Q}_{t+1}^N(f)] = E\left[\{S(P_{t+1}^N)(G_t)\}^{-1} \frac{1}{N} \sum_{n=1}^{N} G_t(x_{t-1}^L, x_t^L) \int_{\mathcal{X}} f(x_n^L, x_t^L) \delta_{n} \otimes m_{t+1}(|f|)\right]
\]

\[
= \tilde{c}_t E\left[\int_{\mathcal{X}^3} G_t(x_{t-1}, x_t) |f(x_t, x_{t+1})| \mathcal{Q}_{t-1}^N \otimes m_{t} \otimes m_{t+1}(dx_{t-1:t+1})\right]
\]

\[
= \tilde{c}_t c_{t-1} \int_{\mathcal{X}^3} G_t(x_{t-1}, x_t) |f(x_t, x_{t+1})| \mathcal{Q}_{t-1}^N \otimes m_{t} \otimes m_{t+1}(dx_{t-1:t+1})
\]

\[
= c_t \int_{\mathcal{X}^3} |f(x_t, x_{t+1})| \Psi_t(x_{t-1} \otimes m_{t}) \otimes m_{t+1}(dx_{t-1:t+1})
\]
\[
\mathcal{X}^2 = c_t \int_{\mathcal{X}^2} |f(x_t, x_{t+1})| \mathcal{Q}_t \otimes m_{t+1}(dx_{t+1})
\]
\[
= c_t \mathcal{Q}_t \otimes m_{t+1}(|f|)
\]

with \(c_t = c_{t-1} \tilde{c}_t \{ \mathcal{Q}_{t-1} \otimes m_t(G_t) \} \) and \(\tilde{c}_t\) be the Boltzmann–Gibbs transformation associated with \(G_t\) (see Del Moral (2004), definition 2.3.3) and where the second inequality uses the inductive hypothesis and the fact that the mapping

\[
(x_{t-1}, x_t) \mapsto G_t(x_{t-1}, x_t)m_{t+1}(|f|)(x_t)
\]
begins to \(L_1(\mathcal{X}^2, \mathcal{Q}_{t-1} \otimes m_t)\). This shows inequality (24) and therefore, for \(N\) sufficiently large, \(\mathbb{E}[\sigma_{N,t}^2] \leq c\) for a constant \(c\) so \(\mathbb{E}[\text{var}\{\mathcal{S}(P_{t+1, h}^N)(\varphi w_{t+1, h}^N)|\mathcal{F}_{t}^N\}] = O\{r(N)\}\). In addition

\[
\mathbb{E}[\mathcal{S}(P_{t+1, h}^N)(\varphi w_{t+1, h}^N)|\mathcal{F}_{t}^N] = \frac{\mathcal{Q}_t^N \{ m_{t+1}(\varphi G_{t+1}) \}}{C_{t+1}^N}
\]
\[
= \frac{\mathcal{Q}_t \{ m_{t+1}(\varphi G_{t+1}) \}}{C_{t+1}^N} + \left( \frac{\hat{\mathcal{Q}}_t - \mathcal{Q}_t}{C_{t+1}^N} \right) \{ m_{t+1}(\varphi G_{t+1}) \}
\]
where \(\mathcal{Q}_t \{ m_{t+1}(\varphi G_{t+1}) \} < \infty\) because \(\varphi \in L_2(\mathcal{X}, \mathcal{Q}_{t+1})\). Since

\[
\frac{\mathcal{Q}_t \{ m_{t+1}(\varphi G_{t+1}) \}}{C_{t+1}^N} = C_{t+1}^N \mathcal{Q}_t \{ m_{t+1}(\varphi G_{t+1}) \} + \frac{C_{t+1} - C_{t+1}^N}{C_{t+1}^N} \mathcal{Q}_t \{ m_{t+1}(\varphi G_{t+1}) \},
\]
we therefore have, for \(N\) sufficiently large,

\[
\text{var} \left( \frac{\mathcal{Q}_t \{ m_{t+1}(\varphi G_{t+1}) \}}{C_{t+1}^N} \right) = \text{var} \left( \frac{C_{t+1} - C_{t+1}^N}{C_{t+1}^N} \right)
\]
\[
\leq \left( \frac{C_{t+1} - C_{t+1}^N}{C_{t+1}^N} \right)^2 \mathbb{E}[\text{var}\{\hat{\mathcal{Q}}_t - \mathcal{Q}_t \}{ m_{t+1}(G_{t+1})}\}].
\]

Since \(\|G_{t+1}\|_\infty < \infty\), \(m_{t+1}(G_{t+1})\) is bounded and the inductive hypothesis implies that the term on the right-hand side of the inequality sign is \(O\{r(N)\}\). In addition, for all \(N\) sufficiently large,

\[
\text{var} \left[ \frac{1}{C_{t+1}^N} (\hat{\mathcal{Q}}_t - \mathcal{Q}_t) \{ m_{t+1}(\varphi G_{t+1}) \} \right] \leq C_{t+1}^2 \mathbb{E}[\text{var}\{\hat{\mathcal{Q}}_t - \mathcal{Q}_t \}{ m_{t+1}(\varphi G_{t+1})}\}^2].
\]

Since \(\{ m_{t+1}(\varphi G_{t+1})\} \leq \|G_{t+1}\|_{\infty} m_{t+1}(\varphi^2 G_{t+1})(x)\), we have

\[
\mathcal{Q}_t \{ m_{t+1}(\varphi G_{t+1}) \} \leq \|G_{t+1}\|_{\infty} C_{t+1} \mathcal{Q}_t \{ \varphi^2 \} < \infty
\]
by assumption. Therefore, \(m_{t+1}(\varphi G_{t+1}) \in L_2(\mathcal{X}, \mathcal{Q}_t)\) so, by the inductive hypothesis,

\[
\text{var}(\mathbb{E}[\mathcal{S}(P_{t+1, h}^N)(\varphi w_{t+1, h}^N)|\mathcal{F}_{t}^N]) = O\{r(N)\}.
\]

Hence,

\[
\text{var}(\mathbb{E}[\mathcal{S}(P_{t+1, h}^N)(\varphi w_{t+1, h}^N)]) = O\{r(N)\}.
\]

The last term of inequality (22) that we need to control is

\[
\mathbb{E}[\mathcal{S}(P_{t+1, h}^N)(\varphi w_{t+1, h}^N)] = \mathbb{E}\left[ \frac{1}{C_{t+1}^N} \hat{\mathcal{Q}}_t \{ m_{t+1}(\varphi G_{t+1}) \} \right].
\]

Since we saw that \(m_{t+1}(\varphi G_{t+1}) \in L_2(\mathcal{X}, \mathcal{Q}_t)\), we have, for \(N\) sufficiently large,
We saw above that the first term on the right-hand side is zero. We need to establish only that, for any function \( \varphi \in \mathcal{C}_0 \), the proof of theorem 6 where we do not use the inductive hypothesis. Inspection of this proof shows that

\[
\mathbb{E}[S(P_{t+1,h}^N)(\varphi w_{t+1,h}^N)] = \mathbb{E}\left[ \frac{1}{C_{t+1}^N} \right] \mathbb{Q}_{t} \{ m_{t+1}(\varphi G_{t+1}) \} + \mathbb{E}\left[ \frac{1}{C_{t+1}^N} (\hat{Q}_t^N - \mathbb{Q}_t) \{ m_{t+1}(\varphi G_{t+1}) \} \right]
\]

\[
\leq c \mathbb{Q}_t \{ m_{t+1}(\varphi G_{t+1}) \} + O\{r(N)^{1/2}\}
\]

using previous computations.

Combining expressions (22), (23), (25) and (26), we obtain \( \text{var}(I_{t+1}^N) = O\{r(N)\} \).

### A.4.3. Proof of theorem 6: \( L_1 \)-convergence

Let \( I_{t+1}^N = \mathbb{Q}_{t+1}(\varphi) \) and \( I_{t+1}^N = \mathbb{Q}_{t+1}^N(\varphi) \) so that

\[
\mathbb{E}[I_{t+1}^N - I_{t+1}] \leq \mathbb{E}[I_{t+1}^N - I_{t+1}^N] + \mathbb{E}[I_{t+1}^N - I_{t+1}].
\]

Then, using expression (15) in the proof of theorem 2, we have, for \( N \) sufficiently large,

\[
\mathbb{E}[I_{t+1}^N - I_{t+1}] \leq \text{var}\{S(P_{t+1,h}^N)(\varphi w_{t+1,h}^N)\}^{1/2} + 2\text{var}\{S(P_{t+1,h}^N)(\varphi w_{t+1,h}^N)\}
\]

\[
+ \mathbb{E}[S(P_{t+1,h}^N)(\varphi w_{t+1,h}^N)^2]^{1/2} \text{var}\{S(P_{t+1,h}^N)(\varphi w_{t+1,h}^N)\}^{1/2}
\]

\[
= O\{r(N)^{1/2}\}
\]

from above computations. In addition,

\[
\mathbb{E}[I_{t+1}^N - I_{t+1}] = \mathbb{E}\left[ \left( \frac{\hat{Q}_t^N}{C_{t+1}^N} - \mathbb{Q}_t \right) \{ m_{t+1}(\varphi G_{t+1}) \} \right]
\]

\[
\leq \mathbb{E}\left[ \frac{1}{C_{t+1}^N} (\hat{Q}_t^N - \mathbb{Q}_t) \{ m_{t+1}(\varphi G_{t+1}) \} \right]
\]

\[
+ \mathbb{E}\left[ \frac{C_{t+1} - C_{t+1}^N}{C_{t+1}^N C_{t+1}} \hat{Q}_t^N \{ m_{t+1}(\varphi G_{t+1}) \} \right]
\]

By the inductive hypothesis and the above computations, the first term after the inequality sign is \( O\{r(N)^{1/2}\} \). In addition, for \( N \) sufficiently large, the second term after the inequality sign is bounded by

\[
\mathbb{E}\left[ \frac{C_{t+1} - C_{t+1}^N}{C_{t+1}^N C_{t+1}} \hat{Q}_t^N \{ m_{t+1}(\varphi G_{t+1}) \} \right] \leq \frac{\delta}{c_t C_{t+1}} \mathbb{E}[\hat{Q}_t^N - \mathbb{Q}_t] \{ m_{t+1}(\varphi G_{t+1}) \]}

\[
+ \frac{\mathbb{E}[\hat{Q}_t] \{ m_{t+1}(\varphi G_{t+1}) \}}{c_t C_{t+1}} \mathbb{E}[\hat{Q}_t^N - \mathbb{Q}_t] \{ m_{t+1}(G_{t+1}) \]
\]

We saw above that the first term on the right-hand side is \( O\{r(N)^{1/2}\} \). In addition, \( m_{t+1}(G_{t+1}) \) belongs to \( L_2(\mathcal{X}, \mathbb{Q}_t) \) because \( \|G_{t+1}\|_\infty < \infty \). Hence, by the inductive hypothesis, the second term after the inequality sign is also \( O\{r(N)^{1/2}\} \) and the proof is complete.

### A.4.4. Proof of theorem 7

To avoid confusion between the \( t \) of the time index and the \( t \) of the \((t,s)\)-sequence we replace the latter by \( \tilde{t} \) in what follows.

The result is true at time \( t = 0 \) by theorem 2. To obtain the result for \( t \geq 1 \) we need to modify the steps in the proof of theorem 6 where we do not use the inductive hypothesis. Inspection of this proof shows that we need to establish only that, for any function \( \varphi \in \mathcal{C}_0([0, 1)^{t-1}) \), we have

\[
\mathbb{E}[\text{var}\{S(P_{t+1,h}^N)(\varphi)\}] = O(N^{-1}).
\]

Let \( N = \lambda b^m \). Then, from the proof of Owen (1998), theorem 1, and using the same notation as in Owen (1998) (note in particular the new meaning for the symbol \( u \)), we have

\[
\text{var}\{S(P_{t+1,h}^N)(\varphi)\} \leq \frac{C}{N} \sum_{|u| > 0} \sum_{|v| > m - |u|} \sigma_{N,u,v}^2.
\]
for a constant \(c\), where \(|u|\) is the cardinality of \(u \subseteq \{1, \ldots, d+1\}\), \(\kappa\) is a vector of \(|u|\) non-negative integers \(k_j, j \in u\), and \(|\kappa| = \sum_{j \in u} k_j\). Note that \(\kappa\) depends implicitly on \(u\). The \(\sigma^2_{\kappa,u,v}\) are such that

\[
\sigma^2_{N,u,v} = \hat{Q}^N_{t+1,b}(\varphi^2) - \hat{Q}^N_{t+1,b}(\varphi)^2 = \sum_{|u|>0} \sum_{\kappa} \sigma^2_{\kappa,u,v},
\]

with \(\sigma^2_{N,u,v} = \int_{[0,1]} \nu_{N,u,v}(x)^2 dx\) and

\[
\nu_{N,u,v}(x) = \sum_{\gamma \in \gamma(u)} \sum_{\tau \in \tau(u)} \langle \varphi \circ F^{-1}_{Q_{\kappa,u,v}}(\kappa,u,v), \psi_{u,v} \rangle \psi_{u,v}(x)
\]

where \(\langle f_1, f_2 \rangle = \int f_1(x) f_2(x) dx\), \(\psi_{u,v}\) is bounded and all the sums in the definition of \(\nu_{N,u,v}(x)\) are finite (see Owen (1977a) for more details).

Similarly, let

\[
\sigma^2 = \hat{Q}^N_{t+1,b}(\varphi^2) - \hat{Q}^N_{t+1,b}(\varphi)^2 = \sum_{|u|>0} \sum_{\kappa} \sigma^2_{\kappa,u,v},
\]

where \(\sigma^2_{u,v} = \int_{[0,1]} \nu_{u,v}(x)^2 dx\) and with

\[
\nu_{u,v}(x) = \sum_{\tau \in \tau(u)} \sum_{\kappa \in \kappa(u)} \langle \varphi \circ F^{-1}_{Q_{\kappa,u,v}}(\kappa,u,v), \psi_{u,v} \rangle \psi_{u,v}(x).
\]

We first want to establish that \(\|\sigma^2_{N,u,v} - \sigma^2_{u,v}\| = O(1)\) almost surely. Note that

\[
\|\nu_{N,u,v} - \nu_{u,v}\| \leq c \sum_{\gamma \in \gamma(u)} \sum_{\tau \in \tau(u)} \|\varphi \circ F^{-1}_{Q_{\kappa,u,v}}(\kappa,u,v) - \varphi \circ F^{-1}_{Q_{\kappa,u,v}}(\kappa,u,v)\|
\]

for a constant \(c > 0\). To show that the term inside the absolute value sign is almost surely \(O(1)\), assume that, for all \(\hat{u} \in [0, 1]\),

\[
|F^{-1}_{Q_{\kappa,u,v}}(\hat{u}) - F^{-1}_{Q_{\kappa,u,v}}(\hat{u})| = O(1) \quad \text{almost surely.}
\]

Using the continuity of \(\varphi\) and the continuity of the Hilbert curve \(H\), and the fact that \(F^{-1}_{m_{t+1}}(x_t, x_{t+1})\) is a continuous function of \(x_t\) (assumption (b)), we have for any \((h_t, x_{t+1}) \in [0, 1)\)

\[
|\varphi \circ F^{-1}_{Q_{\kappa,u,v}}(h_t, x_{t+1}) - \varphi \circ F^{-1}_{Q_{\kappa,u,v}}(h_t, x_{t+1})| = O(1), \quad \text{almost surely,}
\]

and therefore, since \(\varphi\) and \(\psi_{u,v}\) are bounded, we have, by the dominated convergence theorem,

\[
|\langle \varphi \circ F^{-1}_{Q_{\kappa,u,v}}, \psi_{u,v} \rangle - \langle \varphi \circ F^{-1}_{Q_{\kappa,u,v}}, \psi_{u,v} \rangle| \to 0, \quad \text{almost surely.}
\]

We now establish that, for all \(\hat{u} \in [0, 1]\),

\[
|F^{-1}_{Q_{\kappa,u,v}}(\hat{u}) - F^{-1}_{Q_{\kappa,u,v}}(\hat{u})| \to 0 \quad \text{almost surely.}
\]

The proof of this result is inspired by Barvínok et al. (1991), theorem 2.

First, note that because \(p_i(x_i) > 0\) for all \(x_i \in [0, 1)^d\) (assumption (d) of theorem 5) the function \(F_{Q_{\kappa,u,v}}\) is continuous and strictly increasing on \([0, 1]\) (see the proof of lemma 2). Let \(\epsilon > 0\) and \(\hat{u} \in [0, 1]\). Then, by the continuity of \(F_{Q_{\kappa,u,v}}\), there is a \(\delta_{\hat{u}, \epsilon} > 0\) such that

\[
|\hat{u} - \hat{u}| \leq \delta_{\hat{u}, \epsilon} \Rightarrow |F^{-1}_{Q_{\kappa,u,v}}(\hat{u}) - F^{-1}_{Q_{\kappa,u,v}}(\hat{u})| \leq \epsilon.
\]

(27)

In the proof of theorem 6 we saw that, for any \(\delta_0 > 0\), there is an \(N_0\) such that, for all \(N \geq N_0\),

\[
\|F_{Q_{\kappa,u,v}} - F_{Q_{\kappa,u,v}}\| \leq \delta_0, \quad \text{almost surely.}
\]

(28)

Let \(x_N = F^{-1}_{Q_{\kappa,u,v}}(\hat{u}_1)\) and \(u_N = F_{Q_{\kappa,u,v}}(x_N)\). Then, by result (28),
Hence, using Fatou’s lemma, let

\[ \| \delta \| \leq \| \sigma \|, \quad \forall N \geq N_0, \text{ almost surely.} \]

Now note that \( r_N(\tilde{u}_1) = E_{\tilde{Q}_{\tilde{u}_1}} \{ F_{\tilde{Q}_{\tilde{u}_1}}^{-1} (\tilde{u}_1) \} - \tilde{u}_1 \) so that

\[ |F_{\tilde{Q}_{\tilde{u}_1}}^{-1}(x_N) - F_{\tilde{Q}_{\tilde{u}_1}}^{-1}(x_N)| = |\tilde{u}_1 + r_N(\tilde{u}_1) - u_N| \leq \delta_0, \quad \forall N \geq N_0, \text{ almost surely.} \]

Consequently, \( \| \nu_{u,N} - \nu_{u,N} \|_1 \leq C(1) \) almost surely so, by the dominated convergence theorem, \( \sigma^2_{u,N} \to \sigma^2_{u,N} \) almost surely. Also, because \( \varphi \) is continuous and bounded, \( \sigma^2_{u,N} \to \sigma^2 \) almost surely by theorem 5 and the portmanteau lemma (van der Vaart (2007), lemma 2.2). To simplify the notation in what follows, let \( \tilde{\sigma}^2 = \sum_{|\kappa| = |u|} \sigma^2_{u,N} \) and \( \tilde{\sigma}^2_{u,N} = \sum_{|\kappa| = |u|} \sigma^2_{u,N} \) (note that \( l \) depends implicitly on \( u \)), and note that, for \( m \geq l + d + 1, \)

\[ \sum_{|\kappa| = |u|} \sum_{m-l-|\kappa|} \tilde{\sigma}^2_{N,u,l} = \sigma^2_{N} - \sum_{|\kappa| = |u|} \sum_{l=0}^{m-1} \| l \leq m - l - |\kappa| \| \tilde{\sigma}^2_{N,u,l}. \]

By Fubini’s theorem,

\[ \mathbb{E} \left[ \sum_{|\kappa| = |u|} \sum_{m-l-|\kappa|} \tilde{\sigma}^2_{N,u,l} \right] = \mathbb{E} \left[ \sigma^2_{N} \right] - \sum_{|\kappa| = |u|} \sum_{l=0}^{m-1} \| l \leq m - l - |\kappa| \| \mathbb{E} \left[ \tilde{\sigma}^2_{N,u,l} \right] \]

where, by the dominated convergence theorem, \( \mathbb{E} \left[ \sigma^2_{N} \right] \to \sigma^2 \). In addition, since in the definition of \( \sigma^2_{N,u,l} \) and \( \tilde{\sigma}^2_{N,u,l} \) the notation \( \sum_{|\kappa| = |u|} \) denotes a finite sum, we have, for any \( u \) and \( l, \tilde{\sigma}^2_{N,u,l} \to \sigma^2_{u,l} \) almost surely and therefore, by the dominated convergence theorem, \( \mathbb{E} \left[ \tilde{\sigma}^2_{N,u,l} \right] \to \sigma^2_{u,l} \) (because \( \sigma^2_{N} \) is bounded by \( \| \varphi \|_{\infty}^2 \)). Hence, using Fatou’s lemma,

\[ 0 \leq \limsup_{m \to \infty} \mathbb{E} \left[ \sigma^2_{N} - \sum_{|\kappa| = |u|} \sum_{l=0}^{m-1} \| l \leq m - l - |\kappa| \| \tilde{\sigma}^2_{N,u,l} \right] \]

\[ \leq \limsup_{m \to \infty} \mathbb{E} \left[ \sigma^2_{N} \right] + \limsup_{m \to \infty} \left\{ - \sum_{|\kappa| = |u|} \sum_{l=0}^{m-1} \| l \leq m - l - |\kappa| \| \mathbb{E} \left[ \tilde{\sigma}^2_{N,u,l} \right] \right\} \]

\[ \leq \sigma^2 - \sum_{|\kappa| = |u|} \sum_{l=0}^{m-1} \| l \leq m - l - |\kappa| \| \mathbb{E} \left[ \tilde{\sigma}^2_{N,u,l} \right] \]

\[ = \sigma^2 - \sum_{|\kappa| = |u|} \sum_{l=0}^{m-1} \tilde{\sigma}^2_{u,l} = 0, \]

since the indicator functions converge to 1.

References

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

Andreu, C., Doucet, A. and Holenstein, R. (2010) Particle Markov chain Monte Carlo methods (with discussion). *J. R. Statist. Soc. B.*, 72, 269–342.
Barvinek, E., Daler, I. and Francu, J. (1991) Convergence of sequences of inverse functions. *Arch. Math.,* 27, 201–204.

Bocquet, M., Pires, C. A. and Wu, L. (2010) Beyond Gaussian statistical modeling in geophysical data assimilation. *Monthly Weath. Rev.,* 138, 2997–3023.

Briers, M., Doucet, A. and Maskell, S. (2010) Smoothing algorithms for state-space models. *Ann. Inst. Statist. Math.,* 62, 61–89.

Butz, A. R. (1969) Convergence with Hilbert’s space filling curve. *J. Computul Syst. Sci.,* 3, 128–146.

Cappé, O., Moulines, E. and Rydén, T. (2005) *Inference in Hidden Markov Models.* New York: Springer.

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

Chan, D., Kohn, R. and Kirby, C. (2006) Multivariate stochastic volatility models with correlated errors. *Econometr. Rev.,* 25, 245–274.

Chen, S., Dick, J. and Owen, A. B. (2011) Consistency of Markov chain quasi-Monte Carlo on continuous state spaces. *Ann. Statist.,* 39, 673–701.

Chopin, N. (2002) A sequential particle filter for static models. *Biometrika, 89,* 539–552.

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

Cranley, R. and Patterson, T. (1976) Randomization of number theoretic methods for multiple integration. *SIAM J. Numer. Anal.*, 13, 904–914.

Del Moral, P. (1996) Non-linear filtering: interacting particle resolution. *Markov Process. Reltd Flds,* 2, 555–581.

Del Moral, P. (2004) *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications.* New York: Springer.

Del Moral, P., Doucet, A. and Jasra, A. (2006) Sequential Monte Carlo samplers. *J. R. Statist. Soc. B,* 68, 411–436.

Del Moral, P. and Guionnet, A. (1999) Central limit theorem for nonlinear filtering and interacting particle systems. *Ann. Appl. Probab.,* 9, 275–297.

Devroye, L. (1986) *Non-uniform Random Variate Generation.* New York: Springer.

Dick, J. and Pillichshammer, F. (2010) *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration.* Cambridge: Cambridge University Press.

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

Fearnhead, P. (2005) Using random quasi-Monte Carlo within particle filters, with application to financial time series. *J. Computul Graph. Statist.,* 14, 751–769.

Gerber, M. (2014) On integration methods based on scrambled nets of arbitrary size. *Preprint arXiv:1408.2773.*

Glasserman, P. (2004) *Monte Carlo Methods in Financial Engineering.* New York: Springer.

Gordon, N. J., Salmond, D. J. and Smith, A. F. M. (1993) Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc. F,* 140, 107–113.

Götz, M. (2002) Discrepancy and the error in integration. *Monatsh. Math.,* 136, 99–121.

Guo, D. and Wang, X. (2006) Quasi-Monte Carlo filtering in nonlinear dynamic systems. *IEEE Trans. Signl Process.,* 54, 2084–2098.

Hamilton, C. H. and Rau-Chaplin, A. (2008) Compact Hilbert indices for multidimensional data. In *Proc. 1st Int. Conf. Complex, Intelligent and Software Intensive Systems.* New York: Institute of Electrical and Electronics Engineers.

He, Z. and Owen, A. B. (2014) Extensible grids: uniform sampling on a space-filling curve. *Preprint arXiv: 1406.4549.*

Hickernell, F. J. and Yue, R.-X. (2001) The mean square discrepancy of scrambled (t, s)-sequences. *SIAM J. Numer. Anal.,* 38, 1089–1112.

Hlawka, E. and Mück, R. (1972) Über eine Transformation von gleichverteilten Folgen II. *Computing,* 9, 127–138.

Hong, H. S. and Hickernell, F. J. (2003) Algorithm 823: Implementing scrambled digital sequences. *ACM Trans. Math. Softwr.,* 29, 95–109.

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

Kitagawa, G. (1996) Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Computul Graph. Statist.,* 5, 1–25.

Koyama, S., Castellanos Pérez-Bolde, L., Shalizi, C. R. and Kass, R. E. (2010) Approximate methods for state-space models. *J. Am. Statist. Ass.,* 105, 170–180.

Kuipers, L. and Niederreiter, H. (1974) *Uniform Distribution of Sequences.* New York: Wiley-Interscience.

Künsch, H. R. (2005) Approximate methods for state-space models. *Ann. Statist.,* 33, 1983–2021.

Lécot, C. and Ogawa, S. (2002) Quasirandom walk methods. In *Monte Carlo and Quasi-Monte Carlo Methods 2000,* pp. 63–85. New York: Springer.

Lécot, C. and Tuffin, B. (2004) Quasi-Monte Carlo methods for estimating transient measures of discrete time Markov chains. In *Monte Carlo and Quasi-Monte Carlo Methods 2002,* pp. 329–343. New York: Springer.
functions which have been recently considered by Murray mechanism needs to be smooth and it is helpful if it can be expressed directly in terms of shocks which are
This involves generating $N$

Gaussian then this can be useful in determining the optimal choice of $k$

Andrieu and Roberts (2009) and Andrieu et al

More generally Gerber and Chopin introduce a mapping from the dimension of $k$

The states in this case will be ordered according to the corresponding index $k$

552 Discussion on the Paper by Gerber and Chopin

L’Ecuyer, P., Lécot, C. and L’Archevêque-Gaudet, A. (2009) On array-RQMC for Markov chains: mapping alternatives and convergence rates. In Monte Carlo and Quasi-Monte Carlo Methods 2008, pp. 485–500. Berlin: Springer.

L’Ecuyer, P., Lécot, C. and Tuffin, B. (2006) Randomized quasi-Monte Carlo simulation of Markov chains with an ordered state space. In Monte Carlo and Quasi-Monte Carlo Methods 2004, pp. 331–342. Berlin: Springer.

L’Ecuyer, P., Lécot, C. and Tuffin, B. (2008) A randomized quasi-monte carlo simulation method for markov chains. Ops Res., 56, 958–975.

Lemieux, C. (2009) Monte Carlo and Quasi-Monte Carlo Sampling. New York: Springer.

Matošek, J. (1998) On the L2-discrepancy for anchored boxes. J. Complex., 14, 527–536.

Neal, R. M. (2001) Annealed importance sampling. Statist. Comput., 11, 125–139.

Niederreiter, H. (1992) Random Number Generation and Quasi-Monte Carlo Methods. Philadelphia: Society for Industrial and Applied Mathematics.

Ormoneit, D., Lemieux, C. and Fleet, D. J. (2001) Lattice particle filters. In Proc. 17th Conf. Uncertainty in Artificial Intelligence, pp. 395–402. San Francisco: Morgan Kaufmann.

Owen, A. B. (1997a) Monte Carlo variance of scrambled net quadrature. SIAM J. Numer. Anal., 34, 1884–1910.

Owen, A. B. (1997b) Scramble net variance for integrals of smooth functions. Ann. Statist., 25, 1541–1562.

Owen, A. B. (1998) Scrambling Sobol’ and Niederreiter-Xing points. J. Complex., 14, 466–489.

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

Poyiadjis, G., Doucet, A. and Singh, S. S. (2011) Particle approximations of the score and observed information matrix in state space models with application to parameter estimation. Biometrika, 98, 65–80.

Robert, C. P. and Casella, G. (2004) Monte Carlo Statistical Methods, 2nd edn. New York: Springer.

Rosenblatt, M. (1952) Remarks on a multivariate transformation. Ann. Math. Statist., 23, 470–472.

Sagan, H. (1992) Space-filling Curves. New York: Springer.

Tribble, S. D. (2007) Markov chain Monte Carlo algorithms using completely uniformly distributed driving sequences. PhD Thesis. Stanford University, Stanford.

van der Vaart, A. W. (2007) Asymptotic Statistics. Cambridge: Cambridge University Press.

Wächter, C. and Keller, A. (2008) Efficient simultaneous simulation of Markov chains. In Monte Carlo and Quasi-Monte Carlo Methods 2006, pp. 669–684. New York: Springer.

Discussion on the paper by Gerber and Chopin

Michael Pitt (University of Warwick, Coventry)

It is a pleasure to congratulate the authors on a novel paper which should provoke interest within the statistical community. The paper addresses how quasi Monte Carlo methods can be used in conjunction with sequential Monte Carlo (SMC) algorithms for a wide range of models. Detailed descriptions are provided for the inexpensive approach which relies on the Hilbert space filling curve. Impressive numerical results are also provided, for a variety of applications, when the dimension of the state, $d$, is low.

The main concern is the ‘proposal’ of equation (5), page 517:

$$Q(x_{i-1}^k, x_i) = W_{i-1}^k m_{i} (x_{i-1}^k, x_i).$$

This involves generating $N$ uniform distributions $u$, of dimension $d + 1$ where the first is simply for the univariate index $k$. The remaining $d$ uniform distributions are for going from $x_{i-1}^k$ to $x_i$. This transition mechanism needs to be smooth and it is helpful if it can be expressed directly in terms of shocks which are easily derived (by inversion) from the uniform variates. Typically this will cover the non-linear transition functions which have been recently considered by Murray et al. (2013) and Hall et al. (2014).

The key issue is that, whereas the index $k$ is univariate, the state to which it is attached, $x_{i-1}$, is not. The one-dimensional state case, of Section 2.2, is straightforward as the empirical distribution function may be directly inverted. The states in this case will be ordered according to the corresponding index $k$. More generally Gerber and Chopin introduce a mapping from the dimension of $x_{i-1}$ to the dimension of $k$. This is the pseudoinverse (from the Hilbert curve), $h: [0, 1]^d \rightarrow [0, 1]$. This has attractive properties and is relatively straightforward and inexpensive to implement. This allows the inversion that is associated with this one-dimensional object to be carried out. Crucially, closeness in this lower dimension guarantees closeness of the points in the higher dimensional space.

The approach can clearly be used in the pseudomarginal Metropolis–Hastings algorithm; see Beaumont (2003), Andrieu and Roberts (2009) and Andrieu et al. (2010). If the estimator of the log-likelihood is Gaussian then this can be useful in determining the optimal choice of $N$, the number of particles, as shown.
Fig. 11. Histogram of the log-likelihood estimator (over 3000 replications for various values of the size of the data $T$ and various values of $N$, the number of samples for the estimator): (a) $T = 100$, $N = 18$; (b) $T = 316$, $N = 31$; (c) $T = 1000$, $N = 56$; (d) $T = 3162$, $N = 101$

by Pitt et al. (2012) and Doucet et al. (2015). The issue has also been examined by Sherlock et al. (2015). A central limit theorem exists for the log-likelihood estimator arising from standard SMC estimators; see Bérand et al. (2014). A natural issue is therefore to explore the existence of a central limit theorem for the randomized sequential quasi Monte Carlo estimator. To do this we consider an extremely simple probit example of the following form where the likelihood estimator $\hat{L}$ is given through $\hat{L} = \sum_{t=1}^{T} \log(\hat{p}_t)$ and $\hat{p}_t = \sum_{k=1}^{N} I(x^k_t < 0)/N$, $x^k_t \sim \mathcal{N}(\mu, 1)$.

In this case we can easily use randomized quasi Monte Carlo methods as we employ stratified methods to generate the value of $x^k_t$ as $x^k_t = \mu + \Phi^{-1}(u^k_t)$, where $u^k_t = (k + u_t)/N$, with $u_t \sim U(0, 1)$. For examples of stratified sampling in particle filtering see Carpenter et al. (1999). We consider varying $N$ with $T$ as $N \propto \sqrt{T}$ targeting the standard deviation of the log-likelihood estimator, at a fixed central parameter ordinate, as $\sigma = 0.55$. We display the distribution (over 3000 replications) with the assumed Gaussian distribution $\mathcal{N}(-\sigma^2/2, \sigma^2)$ in Fig. 11. It is clear, for this simple example, that the Gaussian approximation is good for moderately large $T$. In this case, it is reasonable to assume that the variance is of the form

$$\sigma^2 = T \tilde{\gamma}^2 / N^\alpha,$$

with $\alpha \in [1, 2]$, where $\tilde{\gamma}^2$ is $O(1)$. Under some additional assumptions, including that the error in the log-likelihood estimator is Gaussian, we wish to minimize the overall computing time $CT$, given as

$$CT = N \text{IF}(\sigma),$$

where $\text{IF}(\sigma)$ is the integrated auto-correlation time associated with a function of the parameter and $N$ can be expressed as a function of $\sigma$ by using equation (29). The precise construction of $\text{IF}(\sigma)$ may be found in Pitt et al. (2012) and Doucet et al. (2015). The results for $\alpha = 2$ (when $d = 1$) are shown in
Fig. 12. Theoretical computing time against the standard deviation of the likelihood estimator $\sigma$

Fig. 12 with CT minimized around $\sigma = 0.55$. Most importantly, for these methods it appears that $N$ must rise with $\sqrt{T}$, rather than $T$ as for standard pseudomarginal methods. This means that the methods, for low dimensional problems, may be competitive with simulated maximum likelihood methods. The big advantage of pseudomarginal methods, in contrast with simulated maximum likelihood, is that they do not require continuity of the likelihood estimator as a function of the parameters, when the random number seed is fixed.

A minor point is that, in Fig. 3, the mean-squared error results do not look as impressive as I would expect although the bias and the variance of the estimators are far better for sequential quasi Monte Carlo than for SMC. I am not quite sure why this is so.

Gerber and Chopin have demonstrated that, for all aspects of estimation associated with SMC, the careful use of quasi Monte Carlo methods offers huge potential. It gives me great pleasure to propose the vote of thanks.

Arnaud Doucet (University of Oxford)

In seconding the vote of thanks I congratulate Gerber and Chopin on a paper that proposes the first efficient combination of (randomized) quasi Monte Carlo (QMC) and sequential Monte Carlo (SMC) methods.

Over the past 15 years, there have been various attempts to combine QMC with SMC. For example, an early contribution by Ormoneit et al. (2001) used a shifted lattice rule in $dT$ dimensions where $d$ is the state dimension and $T$ the time horizon. However, their resampling step did not appear to rely on QMC and they needed to introduce a random permutation of indices at each time step to limit bias. Although they reported some empirical gains, they did not present theoretical results establishing the benefits of this approach over standard SMC methods. It is not a surprise as efficient methods for simulating Markov chains by using QMC only appeared subsequently (L’Ecuyer et al., 2006) and have been mostly limited to the case where $d = 1$ as they require an ordering of the state space.

The key insight of Gerber and Chopin is the introduction of the Hilbert curve to define an ordering when $d > 1$. The Hilbert curve has been previously used in a Markov chain Monte Carlo setting (Skilling, 2004), where it is used to define a generic slice sampler for arbitrary target distributions on $\mathbb{R}^d$, but it is a construction that most statisticians are not familiar with. Remarkably, the Hilbert curve appears to have all the ‘right’ properties, i.e. conservation of low discrepancy between $[0, 1]^d$ and $[0,1]$, which Gerber
and Chopin have leveraged to establish that their sequential QMC (SQMC) method converges at a faster rate than standard SMC algorithms. It would be interesting to have access to sharper upper bounds on $|Q_N(\varphi) - Q_t(\varphi)|$ and $\sqrt{\text{var}Q_N(\varphi)}$ of the form $\alpha(d)/N^{\beta(d)}$. The empirical results of the authors suggest that $\beta(d)$ is a decreasing function of $d$ with $\lim \beta(d) = \frac{1}{2}$ but that impressive gains are possible for moderate values of $d$, i.e. $d < 10$ in Fig. 5. This unfortunately appears to preclude the use of SQMC to sample high dimensional target distributions.

I would like the authors to comment on the use of their approach to compute smoothed additive functionals, i.e. $\varphi(x_0) = \sum_{n=1}^N \hat{\varphi}(x_n)$. In this case, the standard SMC estimates of computational complexity $O(N)$ satisfy typically $|E\{\hat{Q}_N(\varphi)\} - Q_u(\varphi)| = O(t/N)$ and $\text{var}Q_N(\varphi) = O(t^2/N)$ whereas the forward–backward smoothing SMC estimates of complexity $O(N^2)$ satisfy $|E\{\hat{Q}_N(\varphi)\} - Q_u(\varphi)| = O(t/N)$ and $\text{var}Q_N(\varphi) = O(t/N)$; see Kantas et al. (2015) and Poyiadjis et al. (2011). For a fixed complexity, these mean-squared errors are thus of the same order but the mean-squared error of the standard SMC is variance dominated whereas it is bias dominated for the forward–backward smoothing SMC; see Fig. 1 in Kantas et al. (2015) for an illustration. However, a significant advantage of the forward–backward procedure is that its computational complexity can often be reduced to $O(N)$ by using rejection sampling (Douc et al., 2011). In the context of SQMC, what can be said about the mean-squared errors of the procedures in Sections 4.2.1 and 4.2.2? What happens in terms of rate of convergence if we use a standard rejection sampling procedure as in Douc et al. (2011) in Section 4.2.2? If this lowers the rate of convergence, would it be possible to develop a low discrepancy rejection sampling procedure instead?

The vote of thanks was passed by acclamation.

Chris J. Oates, Daniel Simpson and Mark Girolami (University of Warwick, Coventry)

This paper is timely for highlighting the benefits of quasi Monte Carlo (QMC) in contemporary computational statistical methodology. Below we address the question of whether there is scope to reduce the error of QMC estimators further. The analysis of QMC used by Gerber and Chopin is rooted in the Koksma–Hlawka inequality

$$\left| \frac{1}{N} \sum_{n=1}^N \varphi(u^n) - \int_{[0,1]^d} \varphi(u) \, du \right| \leq V(\varphi) \, D^*(u^{1:N})$$

where $\varphi:[0,1]^d \to \mathbb{R}$ is a test function of interest, $u^{1:N}$ is a point set (or sequence), $V(\varphi)$ is the (Hardy–Krause) total variation and $D^*(u^{1:N})$ is the (star) discrepancy term that is the target of the QMC innovation. Our discussion explores the potential to tackle the rate constant $V(\varphi)$ simultaneously in conjunction with the use of QMC methods to tackle $D^*(u^{1:N})$. This direction has received considerably less attention due to typical analytic intractability of the rate constant. Hickernell et al. (2005) showed that classical control variate strategies from Monte Carlo sampling are typically not well suited to QMC, since the total variation is only weakly related to the Monte Carlo variance that is the target of classical variance reduction techniques. Below we hint towards a general strategy to reduce QMC error that targets the rate constant directly.

Following recent work on ‘control functionals’ by Oates et al. (2015), we consider evaluation of $\varphi$ on two sets $u^{1:N}$ and $v^{1:N}$ at a computational cost (asymptotically) equivalent to evaluating $\varphi$ on one such set. The first set $u^{1:N}$ is used to compute an arithmetic mean

$$I_{CF} = \frac{1}{N} \sum_{n=1}^N \hat{\varphi}_N(u^n)$$

based on a surrogate function $\hat{\varphi}_N:[0,1]^d \to \mathbb{R}$. This surrogate function is itself estimated from the second set $v^{1:N}$, in a preliminary step. In situations where $\hat{\varphi}_N$ can be made to satisfy

(a) $\int \hat{\varphi}_N(u) \, du = \int \varphi(u) \, du$ for all $N \in \mathbb{N}$ and

(b) $V(\hat{\varphi}_N) \to 0$ as $N \to \infty$,

then the control functional estimator $I_{CF}$ is unbiased (in an appropriate sense) and has asymptotically zero error relative to the standard QMC estimator. Oates and Girolami (2015) provide an explicit implementation of this strategy in the more general reproducing kernel Hilbert space formulation of QMC methodology (Dick et al., 2013).

As a simple example, we note that, for differentiable $\varphi$ with sufficiently regular partial derivatives, a basic implementation produces a total variation $V(\varphi_N)$ that vanishes at a rate $O(N^{-1/d})$. Thus control
Fig. 13. Convergence of the control functional QMC estimators (Δ) with respect to the more standard alternatives MC (○) and QMC (□): this example considers the (one-dimensional) test function $\varphi(x) = \sin(2\pi x) + 4x$; the usual MC rate is $O(N^{-1/2})$ and the usual QMC rate is $O(N^{-1+\varepsilon})$ (for any $\varepsilon > 0$); our approach has a rate that is $O(N^{-2+\varepsilon})$ (here we present the randomized QMC case; for $u^{1:N}$ we used a scrambled Halton sequence of length $N$ with a uniform random shift modulo 1 and for $v^{1:N}$ we used a uniform grid on the unit interval; error bars denote ±1 standard deviation; for a fair comparison, the same number of function evaluations for $\varphi$ was used in evaluating each estimator; it is striking how much the estimation error can be reduced by using control functionals).

Functional QMC estimators are asymptotically superior to standard (randomized) QMC estimators under appropriate regularity conditions. Preliminary empirical results strongly support our theoretical analysis; an example is given in Fig. 13.

Given the gains in accuracy that are provided by QMC, it is surely a priority to establish complementary methodology that targets the rate constant governing the practical performance of these algorithms. Control functionals provide one (explicit) route to achieve this goal. The combination of control functionals with the sequential QMC approach of Gerber and Chopin should provide a highly effective approach to estimation.

M. Pollock, A. M. Johansen, K. Łatuszyński and G. O. Roberts (University of Warwick, Coventry)
We congratulate Gerber and Chopin on an excellent paper. It has inspired us to consider ways to incorporate
Discussion on the Paper by Gerber and Chopin

quasi Monte Carlo within sequential Monte Carlo schemes in settings in which the transition density of the latent process is intractable and pseudomarginal methods are deployed. In particular, consider filtering for partially observed (jump) diffusions (e.g. Fearnhead et al. (2008), Pollock (2013) and Pollock et al. (2014)), in which (in the simplest setting) the latent process is a diffusion satisfying the stochastic differential equation

$$dX_t = \alpha(X_t)dt + dB_t, \quad X_0 = x, \quad t \in [0, T].$$

(30)

In this setting the data comprise partial observations (arising at a finite collection of time points) of the latent process; the extension to noisy observations is trivial. The transition density of the latent process (under certain regularity conditions) can be shown (Dacunha-Castelle and Florens-Zmirou, 1986) to have the following form between any two time points \(a\) and \(b\), where \(0 < a < b < T\), with \(\mathcal{W}_{t_a,t_b}^{x_a,x_b}\) denoting the law of a Brownian bridge between \(x_a\) and \(x_b\) over \([a, b]\):

$$p_{b-a}(x_b|x_a) = \mathcal{N}(x_b; x_a, b-a) \exp\left( \int_{t_a}^{t_b} \alpha(u) \, du \right) \mathbb{E}_{\mathcal{W}_{t_a,t_b}^{x_a,x_b}} \left[ \exp\left\{ -\int_a^b \frac{\alpha^2(W_t) + \alpha'(W_t)}{2} \, dt \right\} \right].$$

(31)

To propagate particles between consecutive observation times, \(a\) and \(b\), one could simulate from the proposal \(\hat{p}_{b-a}(x_b|x_a)\), perhaps by rejection sampling, and modify the weight of each particle by a factor corresponding to an unbiased estimate of \(\psi(W)\) (where \(W \sim \mathcal{W}_{t_a,t_b}^{x_a,x_b}\)). Supposing that

$$\forall \, t \in [0, T] \quad \frac{\alpha^2(W_t) + \alpha'(W_t)}{2} \in [L, U],$$

and letting \(\kappa \sim \text{Poi}\{ (U-L)(b-a) \}\) and \((\xi_1, \ldots) \sim \text{IID} \, U[a, b]\), we have the representation

$$\psi(W) = \exp\{-L(b-a)\} \mathbb{E}\left[ \mathbb{E}\left[ \prod_{i=1}^{\kappa} \frac{2U - \alpha^2(W_{t_i}) - \alpha'(W_{t_i})}{2(U-L)} \bigg| \kappa, W \right] \bigg| W \right].$$

(32)

An unbiased estimate of \(\psi(W)\), using a finite dimensional realization of the sample path, is obtained by sampling \(\kappa \sim \text{Poi}\{ (U-L)(b-a) \}\) and \((\xi_1, \ldots, \xi_\kappa) \sim \text{IID} \, U[a, b]\) and employing a simple Monte Carlo approximation (Beskos et al., 2006).

This scheme uses an unbiased estimator constructed by simulating \(\kappa\) and then using a \(\kappa\)-dimensional uniform random variable to approximate the inner expectation. Finding a lower variance unbiased estimator of \(\psi(W)\) is desirable, and one would like to exploit randomized quasi Monte Carlo (RQMC). However, as the dimension of the random variable is random, it is not straightforward to do this directly. One could instead sample \(\kappa\) in the usual manner and approximate the inner expectation conditionally using an RQMC point set. There is clearly a computational cost associated with such an RQMC method, which will only be appropriate for problems in which the variance of the simple Monte Carlo estimator of \(\psi(W)\) is large and \(\kappa\) is typically small.

Axel Finke, Andreas L. Hetland, Anthony Lee and Adam M. Johansen (University of Warwick, Coventry)

We congratulate Gerber and Chopin on this extremely interesting paper, the intricate construction of which employs many disparate ideas including those of quasi Monte Carlo (QMC) and the Hilbert curve for stratification which also interested Steigleder and McCool (2003), allowing their benefits to be realized in a sequential Monte Carlo (SMC) context.

As alluded to by the authors, (sequential) QMC (SQMC) can improve the performance of pseudomarginal methods which are most often associated with Markov chain Monte Carlo algorithms. Here, we identify a pseudomarginal SMC algorithm for which this holds.

Assume that we wish to perform SMC-based filtering for a state space model in which the \(d\)-dimensional state vector \(X_t\) (taking values in \(\mathbb{R}^d\)) admits a decomposition

$$X_t = [X_t^{(1)}, X_t^{(2)}],$$

where \(X_t^{(1)}\) takes values \(\mathbb{R}^{d^{(1)}}\), i.e. \(d^{(1)} + d^{(2)} = d\). If most of the variation in the importance weights can be attributed to the second component, it is beneficial to integrate \(X_t^{(2)}\) out analytically and to perform SMC only on the marginal space.
Discussion on the Paper by Gerber and Chopin

Unfortunately, such marginal (i.e. ‘Rao–Blackwellized’) particle filters (Andrieu and Doucet, 2002) can rarely be implemented since the marginal importance weights involve intractable integrals. Instead, $X_t^{(2)}$ can be integrated out empirically by using the pseudomarginal framework: for each of the $N^{(1)}$ upper level $X_t^{(1)}$-particles, we perform a step in a lower level SMC algorithm with $N^{(2)}$ lower level $X_t^{(2)}$-particles. From these, we obtain a sampling approximation of the intractable marginal importance weights, $\hat{W}_t$.

This algorithm was introduced by Johansen et al. (2012). It can exploit the structure of distributed architectures while remaining exact. Indeed, for $N^{(2)} = 1$, it reduces to a standard SMC algorithm on the joint space; as $N^{(2)} \to \infty$, it mimics the intractable marginal algorithm.

As demonstrated in Fig. 14, variance reductions can be brought about by employing SQMC at the upper or lower level. Deploying SQMC seems to be most useful at the level which accounts for most of the variation in the importance weights. Note, however, that, conditionally on $X_t^{(1)}$, the random weights $\hat{W}_t$ cannot usually be sampled via their generalized inverse cumulative distribution function as the latter is intractable.

Finally, splitting the state space in this manner and performing SQMC only on subregions might be useful to alleviate the deterioration of the relative performance of SQMC in higher dimensions.

Frank Critchley (The Open University, Milton Keynes)
Especially as a relative layman in its subject matter, it is a pleasure to congratulate Gerber and Chopin on such a clearly written and stimulating paper, and I thank them for it. I have one central question and one comment.

Well-known parameterization effects in Markov chain Monte Carlo sampling prompt one central question: to which forms of reparameterization is sequential quasi Monte Carlo methodology invariant (or suitably equivariant), and to which forms is it not so? In the latter case, an important supplementary question arises: is an optimal choice possible? Especially pertinent here are the geometric properties of the Hilbert curve, Fig. 2 indicating that it is equivariant under rotation, invariant to reflection of the first axis but not the second, and changes under permutation of axis labels.

Finally, I share the authors’ enthusiasm for establishing more fully the range of statistical problems for which quasi Monte Carlo ideas have the potential to deliver clear added value. To do this will, it seems, require further clarification about what, essentially, makes them ‘tick’.

Dao Nguyen (University of Michigan, Ann Arbor)
I congratulate Gerber and Chopin for this very stimulating paper. They have come up with a novel approach, adapting to the sequential Monte Carlo context ideas borrowed from quasi Monte Carlo (QMC). Their method is much more accurate, based on extending array-randomized QMC to par-
Table 5. Average acceptance rate for $N = 20, \ldots, 200$

| Method       | Results for the following values of $T$: |       |       |
|--------------|-----------------------------------------|-------|-------|
|              | $T = 1000$                              | $T = 10000$ | $T = 100000$ |
| PMMH–SQMC    | 0.011                                   | 0.0016 | 0.00019 |
| PMMH–SMC     | 0.011                                   | 0.0017 | 0.00018 |

†PMMH: pseudomarginal Metropolis–Hastings.

ticle filtering with slightly computational complexity. In addition, they establish several interesting theoretical results with regular assumptions. However, some issues that I would like to raise are as follows.

Recall that purposes of randomized QMC approaches are variance reductions and error estimations (Tuffin, 2008); their drawbacks are the sacrifices of computation speeds and convergence rates. In addition, the deterioration of QMC with rather low dimension is much more severe than noted in the paper. For a very small $d$ (e.g. $d = 3$), QMC is only better than Monte Carlo when $N^{-1} \log(N)^d \leq N^{-1}/2 \rightarrow N \geq \log(N)^6$. By simple calculation, this implies that $N \geq (6 \log(6))^{6/6} \approx 1543781$, which virtually means that the theory does not support most of the practical implementation. In these regards, I recommend the need for a guideline for $N$, in the absence of theoretical support.

Regarding implementation, I did some experiments using the same setting for the multivariate stochastic volatility (SV) model. Code was provided by Mathieu Gerber (https://bitbucket.org/mgerber/sqmc). Unfortunately, I fail to understand the rather low acceptance rate results as in Table 5. Furthermore, the implementation appears computationally intensive for large-scale problems, and it is likely to be resolved by parallel computing. There have been some prior attempts to parallelize both SMC and QMC algorithms (Cvetanovska and Stojanovski, 2012; McAlinn et al., 2012; Murray, 2012; Zhou, 2013). Hence, I suggest that the parallelization of the SQMC algorithm deserves further study.

SQMC is inherently a particle filter; this implies that any problem of SMC will most probably be encountered by SQMC. It is well known that resampling at every step leads to degeneracy unless the size of the particle population increases exponentially with $t$ (Arulampalam et al., 2002). Therefore it would be interesting to investigate the possibilities of degeneracy in SQMC and methods to improve it. Moreover, it would be worth relaxing some rather strong conditions such as the bounded transition density or bounded observation density in SQMC to make it of a more practical interest. In addition, SMC is replaced by SQMC in a pseudomarginal approach. Sherlock et al. (2013) reported that the pseudomarginal algorithm is optimally efficient when the acceptance rate is approximately 7%. Therefore, I wonder what the optimal acceptance rate for pseudomarginal Metropolis–Hastings SQMC would be.

Overall, the approach proposed by the authors looks promising and worthy of further exploration.

The following contributions were received in writing after the meeting.

**Julyan Arbel** (Collegio Carlo Alberto, Moncalieri) and **Igor Prünster** (University of Torino and Collegio Carlo Alberto, Moncalieri)

We congratulate Gerber and Chopin for a stimulating paper. We would like to comment on the potential implication of sequential quasi Monte Carlo in Bayesian non-parametrics.

Non-parametric mixtures are commonly used for density estimation and can be thought of as an extension of finite mixture models when the number of clusters is unknown. The setting is as follows: observations $y_{1:t}$ are spread out into $k_t$ clusters; the cluster labels, or allocation variables, are denoted by $x_{1:t}$, and are interpreted as the states of $y_{1:t}$ in the context of sequential Monte Carlo (SMC). Note that the states are discrete and elements of $\{1, \ldots, k_t\}$. The $n_t$ observations allocated to the same cluster $j$ are independent and identically distributed from a given parametric kernel $K(\cdot; \theta_j)$. This model is static in the sense that data are usually not recorded sequentially. However, Pólya-urn-type schemes essentially formulate this model sequentially by artificially thinking of observation $y_t$ as occurring at time $t$. So the non-parametric mixture model can be cast as an SMC sampler as follows (see Liu (1996) and Fearnhead (2004)):
\[ y_i | x_i, \theta \sim K(y_i; \theta_{x_i}), \]
\[ x_i | x_{i-1} \sim f^X(x_i | x_{i-1}). \]

The predictive distribution \( f^X \) is particularly simple for the Dirichlet process (Ferguson, 1973; Blackwell and MacQueen, 1973). Denote by \( \alpha \) and \( G_0 \) the precision parameter and the base measure of the Dirichlet process. Then the kernel for the states is given by the following probability mass function on \( \{1, \ldots, k_{i-1} + 1\} \):

\[
m_t(x_{i-1}, x_i = j) = p_{t,j} \propto \left\{
\begin{array}{ll}
K_{i-1,j}(y_i | x_{i-1}) & \text{if } j \leq k_{i-1}, \\
K_0(y_i) & \text{if } j = k_{i-1} + 1
\end{array}
\right.
\]

(33)

where \( K_{i} \) and \( K_0 \) involve integrals of \( K \) and \( G_0 \) which can be calculated in conjugate cases and approximated otherwise. Note that \( m_t(x_{i-1}, x_i) \) is also available in closed form expressions for broader classes of random probability measures including the two-parameter Poisson–Dirichlet process (Pitman and Yor, 1997) and normalized random measures with independent increments (Regazzini et al., 2003; James et al., 2009).

As stressed by Gerber and Chopin, the key ingredient of sequential quasi Monte Carlo is the deterministic transform \( \Gamma_i \). A possible choice for expression (33) is

\[
\Gamma_i(x_{i-1}, u_i) = \min \left\{ j \in \{1, \ldots, k_{i-1} + 1\} : \sum_{j=1}^{j} p_{t,j} > u_i \right\}
\]

(34)

where \( u_i \sim \mathcal{U}((0, 1)) \). The case of discrete proposal \( m_t(x_{i-1}, x_i) \) does not seem to be discussed in the paper. Would the authors expect a gain in efficiency by considering a quasi Monte Carlo alternative like equation (34) instead of standard SMC?

Le Bao (Pennsylvania State University, State College)

In the paper, Gerber and Chopin suggest the use of quasi Monte Carlo (QMC) samples within each iteration of the sequential Monte Carlo sampling algorithm. As shown in both a convergence study and numerical results, this novel sampling strategy results in a better Monte Carlo estimate of a certain quantity, i.e. the filtering expectation \( E[\psi(x_i) | y_{0:t}] \).

As they point out, sequential QMC (SQMC) is based on importance sampling, the key idea of which is replacing the hard-to-sample distribution by distributions that are easy to sample from and adjusting the sampling distribution and target distribution by importance weights. In Bayesian models, the resulting Monte Carlo samples are often used to approximate the posterior distribution. Besides obtaining a consistent estimator, one may also be interested in checking how well SQMC samples represent the conditional distribution of \( x_i | y_{0:t} \) and ask the following questions. Does it cover all high density regions? Does it provide sufficient unique samples within each high density region? Raifyer and Bao (2010) discussed the coverage and sampling efficiency for incremental mixture importance sampling, which was not designed for sequential Monte Carlo problems. It would be helpful if the authors can comment on those properties when SQMC is applied for Bayesian inference.

My second point is regarding step (a) of algorithm 3 when \( t > 0 \). Low discrepancy points \( (\mu^t, v^t) \in [0, 1) \times [0, 1)^d \) are generated by QMC or randomized QMC, where \( \mu^t \) is used for resampling \( x_{t-1} \), and \( v^t \) is used to generate \( x_t \). I am wondering whether it introduces any interaction between the resampling stages and the new point generation process, or is it equivalent to generating \( \mu^t \) and \( v^t \) separately? Intuitively, it makes sense to let the weights from the last iteration, \( W_{t-1}^{n_t} \), contribute in the \( x_t \) generation process, especially when the weights are extremely imbalanced. I look forward to hearing some insight or further explanation from the authors.

Finally, I thank the authors for bringing the QMC method to our attention as it seems to be sufficiently generic to be adapted in other samplers.

Alexandros Beskos (University College London) and Ajay Jasra (National University of Singapore)

We congratulate Gerber and Chopin on a very important contribution to the literature on sequential Monte Carlo (SMC) methods. Our comment is on the potential extension of the methodology to the case where the dimension of \( x_t, d \), is very large; as they remarked, the current methodology may not work well as \( d \) grows. In Beskos et al. (2014) we have co-developed an SMC methodology (called the space–time particle filter) which is designed for a class of state space models and can have polynomial complexity in \( d \) (for some stability). Denoting \( x_t(1:i) \) and \( x_t(i) \) respectively the first \( i \) dimensions of \( x_t \) and the \( i \)th dimension of \( x_t, i \in \{1, \ldots, d\} \), if we can write (taking \( x_0 \) fixed) for each \( t \geq 1 \)

\[
f^Y(y_t | x_t) f^X(x_t | x_{t-1}) = \prod_{j=1}^{d} \alpha_t, j \{y_t, x_{t-1}, x_t(1 : j - 1)\}
\]
for some ‘closely related’ positive functions \{\alpha_t, j\}_{t \geq 1, j \in \{1, \ldots, d\}}, then it can be possible to construct SMC methods which are robust to the effect of large \(d\); see Beskos et al. (2014) for details and for example Rebeschini and Van Handel (2015) for models which have such structure. The basic idea behind the approach in Beskos et al. (2014) is that SMC can work well with regard to the time parameter (e.g. Del Moral (2013)), and the above structure allows for a particle filter also to run in the space direction. In the context of sequential quasi Monte Carlo we believe that one could attempt to use the ideas of Beskos et al. (2014) in conjunction with those in the paper, if one can write the simulation of \(x_t(i)\) in the form

\[x_t(i) = \Gamma_t \{x_{t-1}, x_t(i-1), u_t(i)\}, u_t(i) \sim U([0, 1]).\]

We expect that such a method could be stable in high dimensions, i.e. with polynomial cost in \(d\) for stability, especially, if sequential quasi Monte Carlo can be shown to work well for large \(t\), with \(d\) fixed.

**Daniela Castro** (Pontificia Universidad Católica de Chile, Santiago) and **Emilio Porcu** (Universidad Técnica Federico Santa María, Valparaíso)

**Introduction**

Gerber and Chopin present a novel approach to overcome the difficulties arising by introducing quasi-Monte-Carlo (QMC) point sets in sequential importance sampling. We believe that the paper will have a significant influence on disciplines where strong computational statistical tools are needed. We congratulate them for producing such a rigorous and detailed paper. Our comments are based on the curiosity for applying alternative methods to deal with the degeneracy of the specific particle filtering, as well as on applications in our own research interest.

**Alternatives to the Hilbert filling curve method**

Our starting point is algorithm 1 in Table 1, where the authors illustrate a basic particle filter. To avoid the degeneracy problem being intrinsic to the sequential importance sampling, a selection stage is included. This stage makes the introduction of QMC into sequential Monte Carlo (SMC) a non-easy task for \(d \geq 1\). To overcome this problem, they introduce Hilbert filling curves. In particular, this method makes use of the pseudoinverse, coupled with an appropriate bijection, to reduce the high dimension problem to a considerably simpler one, defined over the compact set \([0,1]\), and where we can proceed as for \(d = 1\), where the procedure can be straightforwardly implemented.

Other solutions proposed to deal with the degeneracy on the path space are adaptive SMC (see for instance Cornebise et al. (2008)) and SMC samplers with partial rejection control, introduced by Liu (2001). Adaptive SMC methods can be used, for example, to adapt the way the particles are sampled and weighted, whereas partial rejection control can be used to reject a particle mutation when its incremental importance weight is below a certain threshold. It might be interesting to study the applicability of these two methods as an alternative to the Hilbert filling curve mechanism proposed by the authors.

**Small sample properties**

As the authors mention, the approach proposed is closer to the regularized particle filter of Fearnhead (2005). The main apprehension with this method is that its computational complexity is quadratic in the number of particles. Nevertheless, in the examples considered in Fearnhead (2005), there is some significant gain in efficiency in practical use. Aware of the challenges when using small \(N\), the authors also provide examples in this direction, comparing the performance of sequential QMC with SMC. From this point of view, it might be timely to compare the gains in efficiency of sequential QMC versus the regularized particle filtering.

**Applications**

Our main curiosity is related to the adaptation of the proposed procedure in a context of vector Hilbert valued fields, indexed by some covariance operator. For such fields, preferential sampling is key for simulation through spectral techniques. In particular, Shinozuka and Deodatis (1996) showed how to include a preferential sampling routine in a spectral approach to simulate a vector Gaussian field with a given spectrum.

**D. S. Coad** (Queen Mary, University of London) and **R. C. Weng** (National Chengchi University, Taipei City)

We congratulate Gerber and Chopin on this elegant paper, which develops a simulation method that often produces more accurate estimates than sequential Monte Carlo sampling. To apply the method, it is necessary to write the current state of a particle given its previous state as a deterministic function of the
Discussion on the Paper by Gerber and Chopin

latter and a fixed number of uniform variates. Several convergence results are established. We feel that the method has considerable practical potential.

Sequential Monte Carlo sampling and sequential quasi Monte Carlo attempt to select a set of random points to represent the posterior distribution. As an alternative to the stochastic approach, there are algorithms that use a deterministic set of points. For instance, the ‘unscented’ Kalman filter introduced by Julier and Uhlmann (1997) picks a minimal set of sample points to estimate the mean and covariance of the state variable. All these Bayesian filtering techniques have been shown to be successful in certain applications. It would be interesting to compare the stochastic approach with the deterministic one. Typically, the deterministic approach is successful when the distribution is unimodal, whereas the stochastic approach works for more flexible distributions. However, the deterministic approach is more efficient. The choice of method would depend on the application.

In Section 4, unbiased estimation of the normalizing constant in the Feynman–Kac model is discussed. Of course, this is just one such model where estimation of this constant is of interest. Another is the Ising model, which is a special case of the Potts model. To avoid estimating the constant in the former case, Møller et al. (2006) constructed the proposal distribution so that the normalizing constant cancels from the Metropolis–Hastings ratio. We can also use Stein’s (1986) identity, which enables asymptotic expansions to be obtained for posterior expectations. Although this identity has been found to be useful in several areas, it appears not to have been fully exploited. For example, we are currently applying it to some recursive Bayesian estimation problems, which include those involving the Kalman filter. Using the expectation equations in Stein’s identity is a deterministic approach as well.

Richard G. Everitt (University of Reading)

I congratulate Gerber and Chopin on their interesting paper and thank them for bringing quasi Monte Carlo methods to the attention of the computational statistics community. My contribution considers how the methodological techniques in this paper may, or may not, be applied to more general settings. In my view, two of the main points to consider here are the limitation of the methodology to spaces of moderate dimension and the necessity to formulate the Markov chain in terms of the Rosenblatt transform \( \Gamma \) of variables on \([0, 1]^d\).

Regarding the dimension, the authors note that the method may not work well in dimensions above \( d = 6 \). Although there are many state space models that fall into this category, there are also many that do not. In addition, a natural extension of the sequential quasi Monte Carlo method would be its use in place of a sequential Monte Carlo (SMC) sampler (Del Moral et al., 2006). However, in most cases where an SMC sampler might be used, the dimension is higher than \( d = 6 \). One way of addressing this might exist in models where there are conditional independence relationships between the variables in the state space, both within and between time slices. Briers et al. (2005) described a method for exploiting such conditional independences in the context of SMC so that the sampling is decomposed into local simulations of smaller dimension. However, their method is based on the approximate technique of loopy belief propagation. Do the authors see any way that their method, or other quasi Monte Carlo techniques, might be applied in higher dimensional settings (via exploiting conditional independence or otherwise)?

Axel Gandy and F. Din-Houn Lau (Imperial College London)

We congratulate Gerber and Chopin on a very interesting and stimulating paper. In this note, we briefly discuss one potential modification of the approach. The proposed sequential quasi Monte Carlo (QMC) algorithm collapses all the information about the geometry of the particles into a one-dimensional ordering via the Hilbert curve (on which a one-dimensional component of the QMC sequence is then used).

Can more of the geometry of the current particles be retained by using a different projection and using more than just one dimension of the QMC sampler? One approach to do so would be to replace the method of selecting existing particles in sequential QMC by the following scheme. Suppose that \( \mathcal{X} = \mathbb{R}^d \) is the space the particles reside in.

\( \text{Step 1:} \) compute a weighted principal components analysis of the particles to find the first two principal components.

\( \text{Step 2:} \) sort the particles according to the second principal component and divide the second principal component into 10 non-overlapping windows of (roughly) equal weight.

\( \text{Step 3:} \) within each window, sort particles according to the first principal component.

\( \text{Step 4:} \) For \( n = 1, \ldots, N \),
Discussion on the Paper by Gerber and Chopin

(a) on the basis of the total window weights, use the first QMC component to select a window and
(b) within the window selected, use the second QMC component to sample a particle by using the
inversion method based on the first principal component.

Fig. 15 illustrates the method proposed. Intuitively, partitioning using the windows groups ‘local’ parti-
cles in the second principal component. Using a window might better preserve the ‘locality’ of the particles
than would the Hilbert curve. We then proceed to sample from the ‘finer’ first principal component.

This approach has an order $N \log(N)$ complexity. Extensions using overlapping windows are possible and
probably desirable; issues to be taken into account include setting appropriate weights of the overlapping
windows, whether to use wraparound windows at the boundaries and preserving the overall complexity
of the procedure (which might be possible by using tree-based data structures). Furthermore, principal
component analysis is only one example of a method that can be used to project the dimension of the
particle space to a lower dimensional space. Evaluating this and similar approaches, in particular to
check whether they indeed lead to improvements of the original sequential QMC approach should be an
interesting topic of research.

Zhijian He (Tsinghua University, Beijing) and Art B. Owen (Stanford University)

We congratulate Gerber and Chopin for a very interesting paper with much promise for applications.
Sequential quasi Monte Carlo (SQMC) is similar to array-randomized quasi Monte Carlo (RQMC)
(L’Ecuyer et al., 2008), in using $T$ sets of $N$ points in $[0, 1]^d$ instead of one point set of dimension $dT$.
The trick is in connecting the $N$ output states generated by time $t$ to $N$ QMC vectors used to generate step
$t+1$. The Hilbert curve yields the missing ‘sorting hat’ making it possible to prove consistency and even
a convergence rate, despite the lack of smoothness in matching output to input.

We looked (He and Owen, 2014) into the related, much simpler problem of piping a one-dimensional
RQMC point set through the Hilbert function to obtain a $d$-dimensional point set. We showed that the
mean-squared error rate is $O(n^{-1-2/d})$ for Lipshitz continuous integrands of dimension $d \geq 3$. Dimension
has an adverse effect and we predict the same for SQMC. Although that rate is unpleasant for large $d$, it
is known to be the best possible (Novak, 1988). QMC often involves tricks to reduce effective dimension
(Caflisch et al., 1997). Some of those methods might pay off for SQMC.

We point out one escape route from having to write the simulation as an explicit function of uniform
variables. Usually the problem is how to handle acceptance–rejection sampling. That can be done with
uniform variables but requires an indefinitely large number of them. One can use RQMC for the first

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig15.png}
\caption{Illustration of the method suggested: windows in the second principal component}
\end{figure}
few attempts (maybe just one) and then paste in independent uniform random numbers for any needed follow-up sampling. Hörmann et al. (2004) included many acceptance–rejection schemes with acceptance rate above 1 − ϵ for any ϵ > 0. That ϵ might spoil the convergence rate but still leave us with a big reduction in the error constant.

We were concerned that the choice of ψ could be critical. Suppose for instance that the outputs at step t have a much heavier distribution than logistic. Then the Hilbert curve will encounter points only in or near the 2d corners of [0, 1]d. Conversely, if those outputs have a much smaller variance than the logistic distribution the curve will find those points mostly near the centre (1, 1, ..., 1) of the cube. We have observed just such anomalies in a two-dimensional stochastic volatility example, but, interestingly, the anomalies had no material effect on the variance of our estimates.

Ajay Jasra, Daniel Paulin and Alexandre H. Thiery (National University of Singapore)

We congratulate Gerber and Chopin for their interesting paper that shows that the power of quasi Monte Carlo methods can be harnessed even in sequential Monte Carlo (SMC) type settings. We especially like theorem 7, which shows that the asymptotic variance of the sequential quasi Monte Carlo estimate is of o(N−1), which compares favourably with the O(N−1) variance of standard SMC.

We have a few comments on this theorem. The theorem uses the scrambling scheme defined in Owen (1997a, b). This was improved in Owen (2008), yielding a slightly better quasi Monte Carlo variance bound of O{N−3−d/3 log(N)d−1} for sufficiently smooth functions. Since these rates are substantially better than O(N−1), it would be interesting to investigate whether such rates also hold for SMC methods, under appropriate conditions. Indeed, in some of the simulation examples, the error rate seems to be of O(N−2).

Theorem 2 of Owen (2008) (see also corollary 1 of Gerber (2014) has shown that the variance of the randomized quasi Monte Carlo estimate cannot be worse than a small absolute constant times the variance of the standard Monte Carlo estimate. As suggested by the simulation results, it would be interesting to see whether such a bound could also be shown for sequential quasi Monte Carlo estimates.

Hans R. Künsch (Eidgenössische Technische Hochschule Zürich)

Gerber and Chopin present convincing arguments for the use of randomized quasi Monte Carlo (QMC) methods in particle filters. Systematic resampling (Carpenter et al., 1999), which is commonly used, is already a randomized QMC technique, but the new contribution of this paper for the resampling step is the Hilbert sorting of the particles. This leads to a resampling method which utilizes the geometric location of the particles in the state space. That this brings advantages can be understood intuitively: if |σ(t)(n) − σ(t)(k)| = 1, then not only the multiplicities of x^n_t and x^k_t individually but also the multiplicity of the pair (x^n_t, x^k_t) is less than 1 away from the expected value. Hence undersampling of one of these two particles is likely to be compensated by oversampling of the other, which is desirable if the two particles are close in state space. However, Euclidean distance of the ψ(x^n_t) in [0, 1]^d is maybe not the best choice, and Hilbert sorting can pull two particles ψ(x^n_t) and ψ(x^k_t) which are close far apart.

It is not clear to me whether the larger gains come from using QMC in the resampling or in the propagation step. If it is the former, then it might be worthwhile to consider other resampling methods which utilize the location of particles in state space. One such method has been proposed by Reich and Cotter (2013): they set P(d^n_t = k) = P_{nk} where the matrix P minimizes ∑n,k P_{nk}∥x^n_t − x^k_t∥2 subject to ∑k P_{nk} = 1 and ∑n P_{nk} = NW^2. This is a linear programming problem whose solution has at most 2N − 1 non-zero probabilities P_{nk}. This method allows us to use other norms than the Euclidean. Its computational complexity is O(N^2) because one must compute all pairwise distances, but there may be approximations which have lower complexity because the exact values of large distances do not matter.

Simon Lacoste-Julien (Institut National de Recherche en Informatique et en Automatique and École Normale Supérieure, Paris) and Fredrik Lindsten (University of Cambridge)

Gerber and Chopin present an elegant theory for justifying the use of quasi Monte Carlo (QMC) point sets in sequential Monte Carlo (SMC). We hope that the impressive empirical results obtained will stimulate an interest in alternative sampling approaches for SMC. One such alternative uses an optimization algorithm to obtain samples. Bach et al. (2012) showed how the Frank–Wolfe (FW) algorithm can be used to obtain adaptive quadrature rules with a potentially faster rate of convergence than Monte Carlo integration. In Lacoste-Julien et al. (2015), we investigated this approach in a particle filter; at time t, we run FW quadratic to approximate the empirical mixture distribution 1/N ∑i=1,...,N Q[d(ξ−1,i, ξ,i)] by a weighted set of N points (corresponding to steps (a)–(c) in algorithm 3 of the paper). These points are then reweighted by evaluating the likelihood in the usual way. By optimizing the position of the particles, we can obtain better accuracy than random or QMC sampling.
The optimization objective in FW quadrature is the maximum mean discrepancy (MMD) between the target distribution \( Q \) and the empirical distribution \( Q_N \) of the weighted point set: MMD(\( Q, Q_N \)) := \( \sup_{\varphi \in \mathcal{H}} |\mathbb{E}(\varphi) - \mathbb{E}(\varphi)| \), where \( \mathcal{H} \) is a reproducing kernel Hilbert space (RKHS) specified by a user-chosen kernel. We can reuse the convergence theory from the FW algorithm to deduce rates on MMD(\( Q, Q_N \)). In Lacoste-Julien et al. (2015), we showed consistency in MMD of using FW quadrature in a particle filter and obtained an \( O(N^{-1}) \) rate of convergence on the restricted class of functions in a finite dimensional RKHS (compare with the \( o(N^{-1/2}) \) rate given in theorem 7).

Compared with QMC, the computational overhead of FW quadrature is quite substantial. However, in scenarios where the computational bottleneck is the evaluation of the likelihood, it can be worthwhile to spend more computational effort to optimize the position of the particles before weight assignment. To illustrate the usefulness of FW quadrature, and how it compares with the QMC approach in the paper, we ran some simulations to sample from a Gaussian mixture in \( d = 2 \) dimensions, with \( K = 100 \) components (mimicking the mixture model if any step is crucial. If \( l = 1 \), just sort the states and points by their (increasing) first co-ordinate. If \( l > 1 \), one may map the states \( x_t \in [0, 1)^d \rightarrow [0, 1) \); then proceed as for \( l = 1 \). Gerber and Chopin do this, using a Hilbert curve mapping. They use QMC points in \( d \) dimensions, but their points are sorted (implicitly) by a hidden additional co-ordinate. The mapping \( [0, 1)^d \rightarrow [0, 1) \) is not essential. For example, if \( l = 2 \), one can sort the states and the points in \( \sqrt{N} \) groups of size \( \sqrt{N} \) by their first co-ordinate; then sort each group by the second co-ordinate. This extends to \( l > 2 \). With this multivariate sort, L’Ecuyer

**Pierre L’Ecuyer (Université de Montréal)**

Gerber and Chopin combine sequential Monte Carlo with randomized quasi Monte Carlo (RQMC) to accelerate convergence. Their proof of \( o(n^{-1/2}) \) convergence rate is remarkable. They apply RQMC as in the array RQMC method discussed below, for which convergence rate theory is still thin despite impressive empirical performance.

Array RQMC simulates an array of \( N \) dependent realizations of a Markov chain, using RQMC so that the distribution of the \( N \) states has low discrepancy \( D_t \) with respect to its theoretical distribution, at any step \( t \). If \( x_t = \pi_t(x_{t-1}, u_t) \), where \( x_{t-1} \sim \mathcal{U}([0, 1]^d) \), \( u_t \sim \mathcal{U}([0, 1]^d) \), and \( \mathcal{U} \) means uniform, array-RQMC approximates the \( (l+d) \)-dimensional integral \( \mathbb{E}[D_t] \), using \( N \) RQMC points. It matches each state to a point whose first \( l \) co-ordinates are near that point, and it uses the next \( d \) co-ordinates for \( u_t \).

The matching step is crucial. If \( l = 1 \), just sort the states and points by their (increasing) first co-ordinate. If \( l > 1 \), one may map the states \( x_t \in [0, 1)^d \rightarrow [0, 1) \); then proceed as for \( l = 1 \). Gerber and Chopin do this, using a Hilbert curve mapping. They use RQMC points in \( d \) dimensions, but their points are sorted (implicitly) by a hidden additional co-ordinate. The mapping \( [0, 1)^d \rightarrow [0, 1) \) is not essential. For example, if \( l = 2 \), one can sort the states and the points in \( \sqrt{N} \) groups of size \( \sqrt{N} \) by their first co-ordinate; then sort each group by the second co-ordinate. This extends to \( l > 2 \). With this multivariate sort, L’Ecuyer
et al. (2009) observed an $O(n^{-2})$ variance when pricing an Asian option with array-RQMC. It would be interesting to see what rate the Hilbert curve mapping can achieve for this example.

When $\mathcal{X} \subseteq \mathbb{R}^d$ instead of $[0, 1)^d$ Gerber and Chopin map $x_i \rightarrow [0, 1)^d$ by using a logistic transformation. The choice of transformation and its parameters may have a significant influence on the overall convergence. For the multivariate sort, no transformation is needed: it works directly in $\mathbb{R}^d$. The Hilbert sort could also be adapted to work directly in $\mathbb{R}^d$.

The sequential Monte Carlo part uses importance sampling and particle resampling: a form of splitting (Kahn and Harris, 1949; L’Ecuyer et al., 2007) with fixed effort ($N$ remains constant). Combinations of array-RQMC with several variants of splitting and with importance sampling were examined and tested by L’Ecuyer et al. (2007). Fixed effort better fits array-RQMC than splitting variants that require a random $N$, and stratification easily improves the multinomial resampling process.

Array-RQMC was previously developed to estimate the expectation of a function of the state: not in the context of filtering for parameter estimation and for estimating posterior densities. But the same ideas apply. L’Ecuyer et al. (2008) proved an $O(n^{-3/2})$ convergence rate for the variance for $l=1$ and special types of RQMC points. Faster rates have been observed empirically in examples. Proving those rates in a general setting (under appropriate conditions) is challenging and deserves further research effort.

**Christiane Lemieux (University of Waterloo)**

I congratulate Gerber and Chopin on their work, which represents a big leap in the challenging topic of combining quasi Monte Carlo (QMC) methods with sequential Monte Carlo (SMC) methods. Before this work, the question of how to reassign particles to paths in the resampling step of a particle filter had been a major obstacle towards a successful combination of the two approaches, amenable to theoretical results concerning the convergence rate of its error. The authors’ use of a Hilbert space filling curve cleverly circumvents this obstacle through its nice interaction with QMC sampling, and also provides a systematic way to implement the reassignment of particles. I expect that this work will be followed by several papers pushing further the use of QMC methods within SMC algorithms, now that this important obstacle has been lifted by the authors.

I agree with the authors that trying to improve the convergence rate is a priority for future work. Here I wish to discuss another point, which stems from the observation that the algorithm proposed seems to work much better in very low dimensions. Hence I wonder whether this method would benefit from techniques designed to reduce the effective dimension of a problem, which are well developed for integration problems based on the multinormal distribution (Caflisch et al., 1997; Wang and Sloan, 2011). One may think that, with the small values of $d$ considered here, these techniques may not be as useful as they have been on the financial problems with $d > 100$ for which they were designed. However, the examples in the paper suggest that the problem must be effectively not much more than two dimensional at each time step for the sequential QMC sampler to improve significantly on its SMC counterpart. From this point of view, even a modest reduction in the effective dimension could make a difference.

To explore this idea further, let $g_{t,\text{ld}}(u_1, \ldots, u_d; x_{t-1})$ and $g_{t,\text{pc}}(u_1, \ldots, u_d; x_{t-1})$ be the transformations used to generate $x_t$ given $x_{t-1}$ using a Cholesky and a principal components decomposition respectively. Inspired by Wang and Sloan (2011), let us assess each method’s ability to use as efficiently as possible its first $k$ inputs via

$$
\delta(k, \cdot) = 1 - \frac{E_{t-1} \| x_t - g_t(u_1, \ldots, u_k, 0, \ldots, 0; x_{t-1}) \|^2}{E_{t-1} \| x_t - E_{t-1}(x_t) \|^2}.
$$

for $k \in \{1, \ldots, d\}$. When $\delta(k, \cdot)$ is close to 1, it means that the first $k$ variables $u_1, \ldots, u_k \in [0, 1)$ explain most of the variability of $x_t$, and thus we can say that the effective dimension of each sampling step is about $k$. Numerical comparisons are given in Table 6.

**Fredrik Lindsten and Sumeetpal S. Singh (University of Cambridge)**

We thank Gerber and Chopin for an interesting and inspirational paper. We would like to comment on the proposed sequential quasi Monte Carlo (SQMC) backward smoother. A key part of the SQMC formalism is to reformulate simulation from the state transition kernel $m_t(x_{t-1}, dx_t)$ as an explicit function of uniform variates: $x_t = \Gamma_t(x_{t-1}, v_t)$, where $v_t$ is $d$ dimensional uniform. This is a very useful representation since, indeed, many models are expressed directly in terms of $\Gamma_t$ (though $v_t$ is not necessarily uniform), whereas the kernel $m_t$ may be intractable. However, whereas the forward-in-time sequential Monte Carlo
Although this expression is also intractable, we can collapse over the artificial state component forward particle genealogy). For simulating jointly the ancestor particle at time $j$ given by the weights Gaussian). The target distribution for the backward sampler, operating on the augmented state, is then

$$\tilde{x}_t \sim \frac{1}{m_t} \prod_{i=1}^d \gamma_i(x_i | x_{i-1}^\prime).$$

We thus obtain a backward smoother which requires evaluation of only $d$.

Comparison of the standard method based on Cholesky decomposition versus principal components to decompose the correlation matrix of $v_t$ in example 2 of the paper, which is given by the lower right entry of $C$.

| $k$ | Results for $d=2$ | Results for $d=4$ | Results for $d=10$ |
|-----|------------------|------------------|------------------|
|     | $\delta(k, \text{std})$ | $\delta(k, \text{pc})$ | $\delta(k, \text{std})$ | $\delta(k, \text{pc})$ | $\delta(k, \text{std})$ | $\delta(k, \text{pc})$ |
| 1   | 0.82             | 0.90             | 0.73             | 0.85             | 0.68             | 0.82             |
| 2   | 1                | 0.86             | 0.90             | 0.77             | 0.84             |
| 3   | —                | 0.93             | 0.95             | 0.82             | 0.86             |
| 4   | —                | 1                | 1                | 0.85             | 0.88             |
| 5   | —                | —                | —                | 0.88             | 0.90             |
| 9   | —                | —                | —                | 0.98             | 0.98             |

†The code made available by the authors suggests they have used the standard method based on Cholesky decomposition. The results suggest that, although principal components do not reduce the effective dimension by large amounts, they certainly cause the problem to become closer to being one dimensional. For example, when $d=10$, we have $\delta(1, \text{pc}) = 0.82$ whereas $\delta(1, \text{std}) = 0.68$. These two numbers are respectively 0.85 and 0.73 when $d=4$. It would be interesting to see how the use of principal components within the sequential QMC algorithm translates into actual gains over the SMC method.

and SQMC algorithms make no explicit reference to $m_t$, the backward smoother requires evaluation of the density of $m_t$, which limits its applicability.

One generic approach to backward sampling for models with intractable transitions is to use (a variant of) approximate Bayesian computation. Specifically, we can introduce an augmented state vector $(x_t, \tilde{x}_t)$ as

$$x_t = \tilde{x}_t + \epsilon n_t,$$

$$\tilde{x}_t = \Gamma_t(x_{t-1}, v_t),$$

where $\epsilon$ is a parameter controlling the approximation error and $n_t \sim p_n$ is an ‘artificial’ noise term (e.g. Gaussian). The target distribution for the backward sampler, operating on the augmented state, is then given by the weights

$$\tilde{W}_{t-1}^j(x_t, \tilde{x}_t) \propto W_{t-1}^j p_n \{ \epsilon^{-1}(x_t - \tilde{x}_t) \} m_t(x_t | x_{t-1}^j).$$

Although this expression is also intractable, we can collapse over the artificial state component $\tilde{x}_t$ by simulating jointly the ancestor particle at time $t-1$ and $\tilde{x}_t$. The implicated joint target distribution cannot be simulated from exactly, but we can, however, replace an exact draw with a Markov chain Monte Carlo move initialized at the original particle genealogy. The connection to approximate Bayesian computation is perhaps most easily seen if we make use of the following Markov chain Monte Carlo procedure (reminiscent of conditional importance sampling or particle Gibbs sampling).

Let $\tilde{x}_t$ be the current backward sample and let $a_t$ be its ancestor index at time $t-1$ (obtained from the forward particle genealogy). For $j \in 1:N \setminus a_t$, simulate or generate uniform distributions $\tilde{v}_t^j$ and compute $x_t^j = \Gamma_t(x_{t-1}^j, \tilde{v}_t^j)$. Set $\tilde{x}_t^j = \tilde{x}_t$. Augment the backward trajectory by simulating $\tilde{x}_{t-1}^j$ with $\tilde{P}(\tilde{x}_{t-1} = x_{t-1}^j) = W_{t-1}^j(\tilde{x}_t) \propto W_{t-1}^j p_n \{ \epsilon^{-1}(x_t - \tilde{x}_t) \}$ (see algorithm 4 of the paper).

We thus obtain a backward smoother which requires evaluation of only $\Gamma_t$, the (user-chosen) density $p_n$, and simulation or generation of uniform distributions. We have implemented this approach, albeit using standard Monte Carlo sampling and not QMC, for the three-dimensional Lorenz 63 model with very promising empirical performance.

Han Liu (Princeton University) and Yong Zeng (University of Missouri, Kansas City)

We congratulate Gerber and Chopin for making an important improvement of sequential Monte Carlo
(SMC) by incorporating quasi Monte Carlo to improve the accuracy. Here we would like to make two comments. The first is about a possible limitation of sequential quasi Monte Carlo (SQMC), along with a recent possible solution. The second is about a potential application of SQMC on high frequency financial data analysis.

First, SMC suffers from sample impoverishment. Namely, the number of particles required can grow dramatically with the state dimension. Compared with SMC, SQMC replaces uniform random numbers by low discrepancy point sets and may have the same limitation. One type of approaches to mitigate this difficulty focuses on finding better ways to generate samples such as Pitt and Shephard (1999) mentioned in this paper. Here, we would like to point out another approach, named implicit particle filtering, proposed by Chorin and Tu (2009). A target probability density function commonly has one or more high probability regions. To keep tractable the number of particles needed for a reasonable approximation of the probability density. Chorin and Tu (2009) proposed to concentrate the particles on the high probability regions of the probability density, even when the state space has large dimension. Particle-by-particle minimizations are used to determine the high probability regions, and data-dependent algebraic equations are solved to produce a high probability sample from the target density. More discussions and applications can be found in Chorin et al. (2010) and Morzfeld et al. (2012).

Second, we would like to point out potential applications of SQMC in estimating a general model of partially observed Markov processes with marked point process observations, proposed for modelling financial ultrahigh frequency transactions price data in Hu et al. (2014). The model has an intuitive random-arrival-time state space representation. The state process is a continuous time multivariate Markov process, including multivariate diffusion, stochastic volatility, jump and Levy processes. The observation times are governed by a non-explosive point process and the noise is described by a generic transition probability distribution. Other observable economic or market factors are permitted to influence the state process, the observation times, the noise and the observations. The normalized filtering equation is derived and numerically solved via Markov chain approximation methods to obtain approximate posterior state process, the observation times, the noise and the observations. The normalized filtering equation is derived and numerically solved via Markov chain approximation methods to obtain approximate posterior distributions for parameters and the intrinsic value process. Particle Markov chain Monte Carlo sampling (Andrieu et al., 2010) via SQMC provides an alternative approach for the Bayes estimation if realtime updating is not required.

Jorge Mateu (University Jaume I, Castellón)

Gerber and Chopin are to be congratulated on a valuable contribution and thought-provoking paper in the context of Monte Carlo sampling, deriving sequential quasi Monte Carlo (SQMC) algorithms. My first comment is to emphasize the wide variety of fields of statistics that could benefit from this enhancement of Monte Carlo strategies for sampling. I would like to focus my comments on two particular possible applications of the SQMC algorithm, as extensions of the examples in Sections 5.2 and 5.3 in the paper.

When modelling spatially distributed Poisson or binomial responses $Y_i$, for example, if $Y_i$ measures disease prevalence or species richness in region $i$, we may use a hierarchical generalized linear model in which, conditionally on a latent Gaussian field $Z$, the $Y_i$ have independent distributions from the exponential family, with link function $g$ such that $g(E[Y_i|Z]) = \beta_i^T b + Z_i$, for vectors $\beta_i$ of possible explanatory variables. In this context, the auto-correlation structure of the underlying normal field affects the spatial correlation of the response variable. Sampling in this context is highly complicated by using a non-Bayesian likelihood approach, and usually classical Monte Carlo approaches are slow and with convergence problems. We wonder whether adapting SQMC algorithms to this context will bring better results.

In a related context, simulation-based techniques for optimal design were proposed in Muller (1999) and Muller et al. (2004) for problems where neither maximization of the criterion nor the integration to evaluate the design criterion, e.g. the expected utility, can be performed analytically. In a Bayesian approach, approximation of the integral by Monte Carlo techniques is feasible by sampling from the joint distribution of parameters and data. If the utility is an easy-to-evaluate function of the parameters and the data, a simple strategy would be to approximate the criterion at each design point and to determine the optimal design on the basis of these approximations. However, there are certain difficulties with this approach if the utility is a complex functional of the posterior distribution, i.e. in most Bayesian design problems. In particular, if for instance the likelihood is not available analytically, standard Markov chain Monte Carlo techniques cannot be applied, but approximate Bayesian computation (ABC) methods can be used for posterior inference. Again we need here to sample from these posterior distributions, and the ABC methods provide a way in this direction. We pose the question whether an adaptation of SQMC algorithms can be useful alternatives to ABC methods under the Bayesian context.
Omiros Papaspiliopoulos (Institució Catalana de Ricerca e Estudis Avancats–Universitat Pompeu Fabra, Barcelona)

I congratulate Gerber and Chopin for a thought-provoking paper. My contribution simply presents a narrative of the proposed methodology with the aim of understanding why quasi Monte Carlo (QMC) has found limited application in Bayesian computation thus far, and the potential of the methodology for large $d$.

I focus on the setting of Section 1.4 and take as given that generating low discrepancy uniform distributions in $dT$ dimensions is impractical when $T$ is large.

A simpler but relevant problem is that of estimating expectations of functions $\phi_i(x_i)$. Then, QMC can be used to draw samples directly from the marginal distribution of $x_i$. If the marginal density of $x_i$ is computable then either by a direct method or by importance sampling QMC can be used; in the latter low discrepancy sequences can be used for generating from the proposal. I believe that the approach of Fearnhead (2005) can fit in this framework, where the marginal density is not analytically computable but can be approximated by using a particle filter. The cost of this computation is $O(N)$ and as a result the cost of introducing QMC in the algorithm is $O(N^2)$. In this framework no paths are simulated; instead QMC is used to sample the marginal distributions.

An alternative approach is to generate Markov chain paths by using QMC. Low discrepancy sequences can be generated for each time point $t$, but the problem becomes how to pair a seed $u^t_r$ at time $t$ with one of the simulated values $x^t_{d-1}$ at time $t−1$, to create ‘low discrepancy paths’. When $d=1$ the solution is to pair the ranked (in magnitude) parents to the ranked seeds. For $d>1$ neither the parent values nor the seeds can be ranked.

The authors propose a methodology that addresses problems with both of these approaches. First, using a standard sequential Monte Carlo technique they propose generating paths to avoid the $O(N^2)$ cost associated with sampling from (approximations to) the marginal distributions in state space models. Second, for $d>1$ they order parents according to their projected value by using the Hilbert space filling curve and relate parent values at time $t−1$ to seeds generated at time $t$ via a one-dimensional projection of the seeds used for sampling the ancestors. A further set of low discrepancy seeds $v^t_j$ is used to generate the value of $x^t_j$, once a parent has been assigned.

The empirical results suggest that this strategy for small $d$ succeeds in producing ‘low discrepancy paths’. I am wondering whether the more modest success for larger $d$ is because the pairing between the parents at time $t−1$ and the seeds generated at time $t$, $u^t$, and $v^t$, is only via the one-dimensional $u^t$.

Christian P. Robert (Université Paris-Dauphine and University of Warwick, Coventry)

Quasi Monte Carlo (QMC) methods are ignored by the statistical community, despite regular attempts by well-respected researchers to induce broader use. Gerber and Chopin are to be congratulated and will hopefully contribute to a wider diffusion of QMC methods.

At a purely mathematical level, that randomized low discrepancy sequences produce both unbiased estimators and error rates of order $O\{N^{−1}\log(N)^{d−1}\}$ at cost $O\{N \log(N)\}$ implies that randomized QMC methods should always replace regular Monte Carlo methods. Since this is not so, we may wonder why. The major difficulty with QMC is its requirement to express Monte Carlo estimators as deterministic transforms of a fixed number of uniform distributions. This is a strong deterrent to the dissemination of QMC alternatives, as moving to a random number of baseline uniform distributions does not seem achievable in full generality. It is thus no surprise that the proposal appears in connection with particle filters. Indeed, this field centres on dynamic importance sampling and hence enjoys a local independent and identically distributedness that relates better to QMC integrators than single-chain Markov chain Monte Carlo algorithms. For instance, each particular resampling step consists in multinomial simulation and hence could be turned into QMC. Actually, lower variance solutions that amount to QMC, like systematic and residual sampling (Gundersen and Jensen, 1987; Fearnhead, 1998; Chopin, 2004), have been proposed in the particle literature.

Here, the authors apply QMC to the particles themselves, using a global low discrepancy sequence for the resampling and the move steps: as the cost of using the Hilbert curve grows quickly with dimension, two sequences in lieu of one would bring a significant cost reduction. Moreover, they still assume the deterministic transform representation

$$x^t_i = \Gamma_t (x^t_{d-1}, u^t_i),$$

which is a stumbling block for those contemplating switching to QMC, as illustrated in the paper by using only normal baseline distributions. In a sequential setting with unknown parameters $\theta$, the transform $\Gamma$
changes each time $\theta$ is modified, which impacts computing cost when the inverse cumulative distribution function is not available. Since simulating $\theta$ is unlikely to benefit from QMC improvements, what is the relevance of the overall scheme? The most variable item in the Gibbs sampling steps expands its inefficiency to the joint kernel.

In connection with lemma 1 and the sequential QMC approximation of the evidence, Rao–Blackwellization using all proposed moves be considered: is this easier and significant within QMC since the gain may be of a lower order? Similarly, the trick at the end of Section 4.2.1, using sequential QMC on a single instead of $t + 1$ vectors, is unclear, but is there a connection with the particle learning literature (Lopes et al., 2010; Carvalho et al., 2010).

I am looking forward to the aftermath of this paper that will expose whether QMC is bound to become the reference in simulation computational statistics or to remain a niche activity away from mainstream simulation.

Daniel Rudolf (Universität Jena)

I congratulate Gerber and Chopin for this extraordinary contribution and I am very happy that now quasi Monte Carlo approaches will achieve wider recognition also in statistics. The numerical experiments in the paper are very promising and I believe that the theoretic results are not yet as far developed as the experiments indicate. I would like to comment on the consistency result of Section 3.4. For me the main challenge would be to derive an explicit error bound instead of the asymptotic statement from theorem 5. To achieve such a result an explicit estimate of the importance sampling procedure (see theorem 1 in Section 3.2) might be useful. As the authors have already pointed out, results in that direction can be found in Aistleitner and Dick (2015).

However, this is certainly not enough to obtain an explicit error estimate. Since this is a difficult question we might start with asking whether for all $T \in \mathbb{N}$ and all $N \in \mathbb{N}$ there is a deterministic sequence of point sets $\mathcal{U}_{T,N} = (U_{i,N})_{0 \leq i \leq T}$ with

$$U_{i,N} = \{u_i^1, \ldots, u_i^N\} \subset [0, 1)^{d+1},$$

Table 7. Algorithm 5: likelihood-free population QMC sampler

| Input: observations $y$, decreasing sequence $(\epsilon_1, \ldots, \epsilon_N)$, prior $\pi$ with cumulative distribution function $F_\pi$ |
|---|
| At iteration $t = 1$: |
| set $j = 1$; |
| for $i = 1$ to $N$ do |
| repeat |
| generate $u_j$ as the $j$th element in a one-dimensional randomized QMC sequence; |
| set $\theta_i^{(1)} = F^{-1}_\pi (u_j)$ and $\eta \sim f(\eta | \theta_i^{(1)});$ |
| increment $j$ |
| until $\rho(\eta(y), \eta(y)) \leq \epsilon_1$; |
| set $\omega_i^{(1)} = 1/N$; |
| end for |
| Take $\Sigma_1$ as twice the empirical variance of the $\theta_i^{(1)}$s |
| for $t = 2$ to $T$ do |
| set $j = 1$; |
| for $i = 1$ to $N$ do |
| repeat |
| generate $(u_j, v_j)$ as the $j$th element in a two-dimensional randomized QMC sequence; |
| select an ancestor $\theta^*_j$ from the $\theta_i^{(1)}$ using $v_j$ as for sequential QMC; |
| generate $\theta_i^{(0)}$ using $\theta^*_j$ and $u_j$ as for sequential QMC; |
| increment $j$ |
| until $\rho(\eta(z), \eta(y)) \leq \epsilon_1$; |
| set $\omega_i^{(0)} \propto \pi(\theta_i^{(0)}) / \sum_{j=1}^N \omega_j^{(0)}$ and $\varphi \{ \Sigma_{i-1}^{1/2} (\theta_i^{(0)} - \theta_j^{(t-1)}) \};$ |
| end for |
| Take $\Sigma_t$ as twice the weighted variance of the $\theta_i^{(0)}$s |
| end for |
such that the approximation scheme $\hat{Q}_t^N$, given by algorithm 3, using $U_{t,N}$ satisfies

$$||\hat{Q}_t^N - Q_t||_E \leq r(N), \quad \forall t \in \{0, 1, \ldots, T\},$$

with an explicit known function $r: \mathbb{N} \rightarrow \mathbb{R}_+$ which goes to zero as $N$ increases. For example, one such function could be $r(N) = C_1/\sqrt{N}$ where the constant $C_1$ might even be independent of the dimension. Or, as for quasi Monte Carlo approaches very common, $r(N) = C_2/N^{1-\epsilon}$ for any $\epsilon$, but here $C_2$ scales probably not very well with respect to the dimension.

Let me finally mention that such types of question are also asked in other contexts and can sometimes be answered with probabilistic arguments; see for example Heinrich et al. (2001). This strategy is also related to the considerations in Blumer et al. (1989) and Vapnik and Cervonenkis (1971).

Robin J. Ryder (Université Paris-Dauphine)

One of the reasons why Gerber and Chopin are so successful in their new methodology is the ability to know in advance the number of uniform variables which are needed for one step of the algorithm. This allows greater efficiency and, to use the terminology at the end of Section 1.2, the use of quasi Monte Carlo (QMC) point sets (fixed length) instead of QMC sequences (unbounded length). Nonetheless, as noted by the authors, the use of QMC sequences is at least theoretically possible.

When reading the paper, I was immediately curious about the possibility of adapting this methodology in an approximate Bayesian computation (ABC) setting. In an attempt to develop a naive ABC population QMC algorithm, it is natural to use QMC sequences instead of QMC point sets, as shown in algorithm 5 in Table 7, which is adapted from Marin et al. (2012).

I tried this algorithm on a toy example: $y_i \sim N(x_i, \sigma^2)$ where $(x_i)$ is a Markov chain taking values in $\{-2, 2\}$, with probability of switching at any iteration equal to $\theta$, and attempted to estimate $\theta$. For the generation of QMC sequences, I used the randtoolbox R package (Dutang and Savicky, 2013).

There are two points to note from this simple experiment.

(a) The QMC version systematically outperforms the basic population ABC sampler, with a variance typically about a 30th for an equal number of particles.

(b) The computational time of the QMC version explodes in a non-linear fashion when the number of particles increases, as shown in Fig. 17. This is intriguing and might be due to the cost of generating a QMC sequence, at least in the implementation that I used.

There is certainly much more work to be done to adapt these new ideas to an ABC setting. In particular, there is one situation where QMC point sets could certainly be useful: in ABC, one typically accepts a proposed value $\theta^*$ if that value leads to simulated data $z$ which are close to the observed data $y$, in the sense that $d(y, z) < \epsilon$ for some pseudometric $d$ and threshold $\epsilon$. The threshold $\epsilon$ is not always chosen in advance but is sometimes a quantile from a fixed number of attempts: there must surely be situations where QMC could be used to improve the estimator for such flavours of ABC. I was surprised not to find any investigation of this idea in the literature.

Donatello Telesca and Weng Kee Wong (University of California at Los Angeles)

We congratulate Gerber and Chopin on an interesting and thought-provoking paper. The authors present a solid theoretical foundation of their work and help to clarify the role of quasi Monte Carlo (QMC) methods in the solution of sequential problems (hence the term sequential QMC). The work also provides an impressive and elegant solution to tackle the high computational burden that is commonly seen in such problems. As such, results in the paper are likely to have a strong influence on the design of particle filters.

We focus our discussion on potential applications of sequential QMC to solve Bayesian or frequentist inferential problems in biostatistics, where complex stochastic systems are often used as a schematic description of biomedical or biobehavioural processes. In particular, we consider inferential issues for large data sets, where standard Markov chain Monte Carlo or bootstrap methods may provide unfeasible solutions. In this setting, applications of particle filters (Chopin, 2002; Del Moral et al., 2006) can provide practitioners with a promising operational direction.

We seek a clarification on the potential extension of the work to ‘big data’ problems. Using the authors’ notation, the natural structure of static inference problems may best be described by having a large $d$ and a potentially smaller $T$. Here, $d$ is the dimension of the set of the parameters of interest and $T$ could be conceived as the number of data slices.

The authors noted that, as $d$ increases, QMC methods become seemingly less appealing in this context because low discrepancy point sets are not easily distinguished from standard i.i.d. points. However,
our reading is that the authors appeared to have handled this issue by ingeniously combining QMC with the Markov structure of the problem to define antithetic behaviour in the evolution of the particle system.

We concur with the authors’ remark that the full potential of QMC sampling in statistics remains unexplored. A potentially interesting direction is to explore whether the construction of array-randomized QMC sampling extends easily to Markov processes beyond Markov chains. In particular, we suspect that hierarchical systems or, more generally, directed acyclic graphs that encompass a wide family of useful statistical models could benefit from the use of QMC methods. We appreciate the authors’ thoughts and experience on such extensions.

Howell Tong (London School of Economics and Political Science)

At the heart of modern Monte-Carlo-based statistics in their multifarious guises is the tacit assumption that there is an inexhaustible supply of random numbers that meet the user’s demand, e.g. they are independently and uniformly distributed. Now, often pseudorandom numbers are used instead for their convenience and speed and their generators can be viewed as deterministic non-linear dynamical systems. Although many modern pseudorandom-number generators are capable of generating pseudorandom...
numbers with very long cycle period, say in the billions, chaos literature suggests the existence of ‘pseudo-
cycles’, in that the trajectory does not produce an exact replica but a slightly blurred version of itself. In
the chaos literature, recurrence plots are used. (See, for example, Chan and Tong (2001).) What is worry-
ing is that pseudocycles are not well understood. For example, we might not know how soon a pseudo-
cycle kicks in, although it seems most likely to be much sooner than the exact cycle. I am, of course,
aware of the existence of various tests that are designed to test departure from randomness of pseudo-
random numbers. However, are we sure that we can safely rule out the possibility that these tests may
sometimes fail to detect pseudocycles? von Neumann (1951) said ‘Anyone who considers arithmetical meth-
ods of producing random digits is, of course, in a state of sin’. It seems to me relevant to know how big the
sin is.

Ruochen Wu and Oliver Linton (University of Cambridge)
In this stimulating paper, Gerber and Chopin introduced a new sampling technique, sequential quasi
Monte Carlo (SQMC), which is a sequential version of quasi Monte Carlo (QMC). Similarly to QMC, the
SQMC method employs a low discrepancy point set instead of random draws. The generation of sequences
is based on importance sampling, and the Hilbert space filling curve is used to allow sampling when the
dimension \( d \geq 1 \).

The authors compared the mean-squared error (MSE) of the SQMC and the SMC as functions of the
central processor unit time, which demonstrated that the SQMC algorithm has faster converging MSE
than the SMC algorithm. The gain factors when \( d = 1, 2, 4 \) were also shown to have remained over the
time periods.

From the application point of view, the new SQMC sampling could be applied to decision problems
with a latent Markov process. In such problems, the current utility \( u(s_t, a_t) \) depends on the current state
\( s_t \) of a decision maker and the action \( a_t \) she takes. However, such utility can be observed only with an
error term, as the decision is made by the decision maker who does not typically report their valuation
functions. Attempts to make estimation of such latent Markov chains have employed various sampling
techniques (MacRae, 1977; Rust, 1997; Kearns et al., 2002; Chang et al., 2013). As encouraging results
have been obtained in this paper with SQMC compared with sequential Monte Carlo (SMC), the new
method could be expected to give faster convergence and smaller MSE in estimating such models.

One point that may cause concern is in Fig. 5(d), which shows the gain factor of the SQMC algorithm
as a function of the number of particles, \( N \), when the dimension \( d \) is fixed at 1, 2, 4 and 10. The results
demonstrate that the gain of the SQMC in comparison with the SMC algorithm deteriorates when the
dimension becomes higher, which is expected. However, what may limit the application of SQMC to
empirical analyses is the fact that the gain factor is around 1 regardless of \( N \), when \( d \) is 10, which is not
yet very high. This indicates that the advantage of SQMC regarding the MSE diminishes relatively rapidly
when the dimension increases. It would be nice if the authors could explore, and cope with, the potential
reasons in future researches, e.g. the way in which the low discrepancy point set is generated.

The authors replied later, in writing, as follows.

We are very grateful to all our colleagues, in particular the two main discussants, for their warm and
insightful comments. Our response will address, in turn

(a) potential applications of sequential quasi Monte Carlo (SQMC),
(b) possible ways to improve SQMC,
(c) theoretical challenges and
(d) other points.

Potential applications
Several discussants seem enthusiastic about applying SQMC to their problems and ask us how well SQMC
may perform in these cases. We offer below educated guesses, which may of course be disproved by
numerical experiments.

An application that we find most promising is the Rao–Blackwellized approach of Finke and his col-
lagues, where the state variable is split in two parts, \( X_t = (X_t^{(1)}, X_t^{(2)}) \), a local particle filter is run for \( X_t^{(2)},
\) conditionally on \( X_t^{(1)} \), and a global filter is run ‘pseudomarginally’ for \( X_t^{(1)} \), but with global weights ap-
proximated by local filters. Since the improvement in performance brought by SQMC seems to deteriorate
with the dimension, it certainly makes plenty of sense in some settings to replace straight SQMC with such
a hybrid strategy, where SQMC would be run locally for only a small number of (important) components
of the state variable.
Random-walk particle filtering for diffusion models (Pollock and his colleagues) is another exciting application: there, the weights are a function of $\kappa$ uniform distributions, where $\kappa$ is also random. As noted by Pollock and his colleagues, this is an added difficulty, but which could be approached in several ways. For instance, one may generate a Sobol’ sequence of dimension $\kappa_{\text{max}} + 1$, where the first component is used to choose $\kappa$, and, conditionally on $\kappa$, components from 2 to $\kappa + 1$ are used for the uniform distributions. (For $\kappa > \kappa_{\text{max}}$, one may use standard random numbers, in the spirit of He and Owen’s rejection suggestion; see below). Note also that Sobol’ sequences are nested, i.e. the first $\kappa'$ components of a Sobol’ sequence of dimension $\kappa_{\text{max}} + 1$ is a Sobol’ sequence of dimension $\kappa'$. See also Caflisch et al. (1997) for an efficient way to generate Brownian bridges in the quasi Monte Carlo (QMC) framework (as mentioned by He and Owen, and Lemieux).

Another interesting application is sequential approximate Bayesian computation (Ryder; see also Mateu) for likelihood-free inference. In sequential approximate Bayesian computation applications,

- the dimension of the parameter space $\Theta$ is often small and
- the mutation kernel may be a simple Gaussian random walk;

both aspects seem to indicate that SQMC might perform well. In contrast, likelihood-free problems are often such that generating artificial data sets from the model is not amenable to QMC, so one may need to apply a hybrid strategy, with QMC points for $\theta$, and random points for the artificial data sets, as proposed by Ryder.

The three above discussions (Finke and his colleagues, Pollock and his colleagues, and Ryder) suggest combining randomized (R)QMC with Monte Carlo, i.e. to use QMC for the $d_1 < d$ most ‘important’ components, and random-number generation for the remaining components. This strategy, which is known as hybrid Monte Carlo, was studied by Ökten et al. (2006), who showed that it may be more efficient than both plain Monte Carlo and plain (R)QMC in high dimensional problems. Theoretically, Ökten et al. (2006), corollary 6, showed that the star discrepancy of a hybrid Monte Carlo point set of size $N$ goes to 0 as $N \to \infty$ with probability 1. Thus, as far as consistency is concerned, mixing QMC point sets and uniform random sampling is a valid way to choose the $u_{1,N}$’s that are used as inputs of SQMC.

Another question concerns the construction of a QMC version of sequential Monte Carlo (SMC) samplers (Chopin, 2002; Del Moral et al., 2006), i.e. SMC algorithms used for non-sequential problems. However, and as pointed out by Everitt (see also Telesca and Wong), SMC samplers do not seem to be an ideal candidate for a QMC treatment for the moment, as they are typically applied in situations where $d$ is large (typically $d \gg 10$) and involve Metropolis–Hastings steps which impose non-trivial transformations of the QMC point sets. We did some experiments on a logistic regression with 20 covariates and observed a very modest improvement. That said, there may be ways to improve the performance of SQMC for such problems, as elaborated in the following section. Inference on Dirichlet processes (Arbel and Prünster) is another interesting application, but with the added difficulty that $\kappa$, is discrete and non-Markov in these settings.

Finally, Oates and his colleagues mention a nice way to combine (non-sequential) QMC with control functionals (a functional generalization of control variates). More generally, we hope to see a renewed interest among statisticians for QMC and RQMC, outside the filtering problem that we have addressed in this paper. See also the contribution by Robert.

**Possible ways to improve sequential quasi Monte Carlo**

Several comments concern the effect of the Hilbert sort step on performance, and in particular on the choice of $\psi$, the function that maps $X$ into $[0, 1]^d$. He and Owen, and L’Ecuyer, worry about situations where $\psi$ sends all the points to the centre, or instead near the edges. We believe (and have observed numerically, like Owen and He) that $\psi$ does not have much influence on performance. We believe that it stems from the fractal nature of the Hilbert curve, which makes it possible to zoom indefinitely into any part of $[0, 1]^d$ (up to numerical error, obviously).

L’Ecuyer, Gandy and Lau, and Papaspiliopoulos suggest that the quick deterioration of SQMC as $d$ increases may be due to the Hilbert resampling step. Indeed, as the Hölder exponent of the Hilbert curve is $1/d$, transformations through this space filling curve become quickly non-smooth as $d$ increases, reducing the performance of SQMC.

To overcome this limitation, Gandy and Lau suggest an alternative way to order ancestors, based on the two first principal components of these ancestors. This is a nice idea, which may actually be combined with Hilbert sorting: compute the first $k \ll d$ principal components of the ancestors, and sort them according to the Hilbert curve. It might be difficult to establish any formal result for such an approach, but this is definitely worth trying in cases where $d$ may be large. However, it is worth noting that this procedure will preserve the unbiased property of SQMC likelihood estimates and thus may be safely used inside.
particle Markov chain Monte Carlo algorithms. L’Ecuyer claims that the mapping \([0, 1]^d \rightarrow [0, 1]\) is not essential and that a multivariate sort can be used instead. This may be true in practice, but it is worth reminding that the resampling interpretation of the ordering step is the key ingredient for the theoretical analysis of SQMC. As explained in Section 1.4, this also allows us to establish the validity of (a version of) array-RQMC in dimension \(d > 1\) for which a supporting theory was lacking for almost 20 years.

This comment seems also related to some extent to the insightful point of Critchley, i.e. the effect of the parameterization of \(x_t\) on the performance of the algorithm. Contrary to standard Monte Carlo sampling, SQMC is not parameterization invariant; for instance, if the Rosenblatt transform is used, then swapping the two first components of vector \(x_t\) leads to a (slightly) different algorithm. We are not very clear yet on how to choose the parameterization to achieve best performance, but we refer to the nice proposal of Lemieux, who, for our multivariate stochastic volatility example, shows that a certain linear transformation, based on a principal component decomposition, allows us to reduce the effective dimension and hence may improve the performance of SQMC. This simple idea is probably the first to try to improve the performance of SQMC for large \(d\).

Beskos and Jasra discuss a class of state space models where SQMC may work well when \(d \gg 1\). Following this idea it is worth noting that we can expect better gains in settings where the Markov transition \(m_t(x_{t-1}, dx_t)\) is such that \(x_t\) can be partitioned into groups of components that are, conditionally on \(x_{t-1}\), independent. Also, instead of generating new particles by using the Markov transition one can use a proposal of the form \(q(x_{t-1}, dx_t) = \phi_{\sup t}^{1/d} q(\chi_{t-1}, dx_t)\) which is ‘close’ (in some sense) to the probability distribution proportional to \(G_t(x_{t-1}, x_t) m_t(x_{t-1}, dx_t)\). However, this will change the regularity of the potential function \(G_t\), and thus it is difficult to predict in full generality the overall efficiency of this procedure.

A last interesting comment about the applicability of SQMC when the dimension \(d\) is large is made by Liu and Zeng, who propose to combine implicit particle filtering and SQMC. The idea of implicit particle filtering is to force the particles \(x_{1:N}^{t}\), generated at time \(t\), to be in high probability regions of \(Q_{t}^{N}(dx_t)\). The expected gain of implicit particle filtering is therefore to require fewer particles than SMC to be efficient. From a practical point of view, mixing these two filtering techniques seems easy and is certainly worth trying.

**Theoretical developments**

Several comments are about the theoretical properties of SQMC. Doucet and Jasra and his colleagues ask about the existence of an explicit convergence rate for the stochastic version of SQMC. In particular, future research should establish its dependence on the dimension \(d\), and whether the fast convergence rates of quadratures based on scrambled nets for smooth functions (Owen, 1997b, 1998) may be recovered in the SQMC context. Pitt wonders whether there is a central limit theorem (CLT) for the randomized version of SQMC. The existence of a CLT for quadratures based on scramble nets (Loh, 2003) makes us optimistic about the possibility of establishing a CLT for SQMC when such nets are used as inputs. However, the CLT of Loh (2003) is based on some strong assumptions about the smoothness of the integrand and hence its adaptation to SQMC is likely to be non-trivial because of the resampling steps.

A related question is the convergence rate of the SQMC estimate of the likelihood function and the optimal calibration of \(N\) in particle marginal Metropolis–Hastings (PMMH)–SQMC (see Pitt, and also Nguyen). Since the SQMC variance decreases faster with \(N\) than the Monte Carlo variance, it makes sense to suppose that the acceptance rate that one should target with PMMH–SQMC should be higher than for PMMH–SMC. The experiment conducted by Pitt seems to go in this direction.

Another important theoretical question, posed by Rudolf, concerns the existence of a function \(r(N)\) such that

\[
\sup_{t \geq 0} \| Q_{t}^{N} - Q_{t} \| \leq r(N), \quad r(N) \rightarrow 0 \text{ as } N \rightarrow \infty.
\]

Indeed, the existence of a time uniform bound for the extreme metric is related to one of the key characteristics of SMC, i.e. its stability over time. However, we think that the existence of such a bound is (as for SMC) related to the forgetting properties of the Feynman–Kac model and not directly to the choice of the QMC point sets \(u_{1:N}, t \geq 0\). Nevertheless, the construction of good QMC point sets for SQMC is an open and interesting problem.

Doucet wonders about the time behaviour of SQMC estimates of smoothing expectations of additive functions. Some preliminary computations suggest that, for a fixed \(N\), the variance is, as for SMC, of order \(O(r^2)\). However, one can expect to obtain an overall \(O(1/N)\) convergence rate for the variance and, in some settings, the forward estimation of the smoothing expectation of additive functions may be efficient provided that \(T\) is not too big. More work in this direction is needed.

The theoretical study of QMC forward–backward smoothing (see the comment by Doucet) is an
Discussion on the Paper by Gerber and Chopin

on-going work but its convergence rate remains an open problem (we have only established the validity of the procedure, which is already far from trivial). The idea of using rejection sampling to reduce the computational cost seems promising; see the comment of He and Owen for a possible way to implement it. It is also worth exploring in future research whether the procedure proposed by Lindsten and Singh in models where the Markov transition is intractable could be adapted to SQMC.

Other points

Both Künsch and Robert mention that systematic resampling (a standard resampling strategy in particle filtering) is implicitly a QMC procedure, because it is based on a low discrepancy point set of dimension 1. Künsch wonders whether the benefits brought by QMC come mostly from the selection stage (in our case, the choice of ancestors based on the Hilbert curve), or from the mutation stage (where \( x_t \) is generated, conditionally on ancestor \( x_{t-1} \)); see also Le Bao. Our intuition is that it is important to have both simultaneously. Paul Fearnhead (personal communication) did some numerical experiments on using QMC only at the mutation step, but he obtained disappointing results. Alternatively, if QMC is used only at the selection stage, we believe that the performance would be on a par with systematic resampling (on unordered ancestors).

Tong makes a valid point on the pitfalls of pseudorandom generators. If we use only deterministic QMC sequences in SQMC, these pitfalls are avoided, but we still prefer the randomized version, as it has more applications.

Coad and Wang mention the ‘unscented’ Kalman filter (Julier and Uhlmann, 1997) as an alternative to particle filtering, and therefore SQMC. Although deterministic approximations certainly have their use in some applications, there is also a need for exact (in the limit) methods, if only to assess how accurate such deterministic approximations are.

Lacoste-Julien and Lindsten mention an interesting algorithm based on Frank–Wolfe optimization. As for SQMC, the objective is to choose more cleverly the particles that are generated at any time period to beat the Monte Carlo convergence rate. For the algorithm that was discussed by Lindsten and Lacoste-Julien the convergence rate is \( O(1/N) \) (compared with \( O(1/\sqrt{N}) \) for SQMC. However, it costs \( O(N^2) \) and, if we understand correctly, is restricted to linear Gaussian transitions. Given its slow running time, this method may be an interesting alternative to SQMC when the evaluation of the potential functions is very costly and one wants to use a small number of particles.

As recalled by Castro and Porcu the regularized particle filter (RPF) of Fearnhead (2005) is another filtering algorithm which beats the Monte Carlo error rate, at the cost of \( O(N^2) \). We compared the RPF with SQMC on a univariate stochastic volatility model, and we observed that the faster convergence rate of the RPF does not compensate its longer running time, even for small values of \( N \). However, as for the filter based on Frank–Wolfe optimization mentioned by Lindsten and Lacoste-Julien, the RPF may be competitive when the evaluation of the potential functions is very expensive.

Conclusion

To conclude, we thank again all the discussants for their enthusiastic and stimulating comments, and we strongly hope that this paper will renew statisticians’ interest in QMC, whether in sequential problems or otherwise, as we find this piece of methodology to be fascinating and full of potential for statistical problems.

References in the discussion

Aistleitner, Ch. and Dick, J. (2015) Functions of bounded variation, signed measures, and a general Koksma–Hlawka inequality. *Acta Arithm.*, **167**, 143–171.

Andrieu, C. and Doucet, A. (2002) Particle filtering for partially observed Gaussian state space models. *J. R. Statist. B*, **64**, 827–836.

Andrieu, C., Doucet, A. and Holenstein, R. (2010) Particle Markov chain Monte Carlo methods (with discussion). *J. R. Statist. Soc. B*, **72**, 269–342.

Andrieu, C. and Roberts, G. O. (2009) The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.*, **37**, 697–725.

Arulampalam, M. S., Maskell, S., Gordon, N. and Clapp, T. (2002) A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Trans. Signal Process.*, **50**, 174–188.

Bach, F., Lacoste-Julien, S. and Obozinski, G. (2012) On the equivalence between herding and conditional gradient algorithms. In *Proc. 29th Int. Conf. Machine Learning* (eds J. Langford and J. Pineau), pp. 1359–1366. New York: Omnipress.

Beaumont, M. (2003) Estimation of population growth or decline in genetically monitored population. *Genetics*, **164**, 1139.
Berard, J., Del Moral, P. and Doucet, A. (2014) A lognormal central limit theorem for particle approximations of normalizing constants. *Electron. J. Probab.*, 19, 1–28.

Beskos, A., Crisan, D., Jasra, A., Kamatani, K. and Zhou, Y. (2014) A stable particle filter in high-dimensions. (Available from arxiv.org/abs/1412.3501.)

Beskos, A., Papaspiliopoulos, O., Roberts G. O. and Fearnhead, P. (2006) Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *J. R. Statist. Soc. B.*, 68, 333–382.

Blackwell, D. and MacQueen, J. B. (1973) Ferguson distributions via Polya urn schemes. *Ann. Statist.*, 1, 353–355.

Blumer, A., Ehrenfeucht, A., Haussler D. and Warmuth, M. K. (1989) Learnability and the Vapnik-Chervonenkis dimension. *J. Ass. Comput. Mach.*, 36, 929–965.

Briers, M., Doucet, A. and Singh, S. S. (2005) Sequential auxiliary particle belief propagation. *7th Int. Conf. Information Fusion*.

Caflisch, R. E., Morokoff, W. J. and Owen, A. B. (1997) Valuation of mortgage backed securities using Brownian bridges to reduce effective dimension. *J. Computul Finan.*, 1, 27–46.

Carpenter, J., Clifford, P. and Fearnhead, P. (1999) Improved particle filter for non-linear problems. *IEE Proc. F.*, 146, 2–7.

Carvalho, C. M., Johannes, M. M., Lopes, H. F. and Polson, N. S. (2010) Particle learning and smoothing. *Statist. Sci.*, 25, 88–106.

Chan, K.-S. and Tong, H. (2001) *Chaos: a Statistical Perspective*. New York: Springer.

Chang, H. S., Hu, J., Fu, M. C. and Marcus, S. I. (2013) *Super-samples from kernel herding*. In *Learning and Uncertainty in Artificial Intelligence, Catalina Island* (eds P. Grünwald and P. Spirtes), pp. 109–116. Corvallis: Association for Uncertainty in Artificial Intelligence Press.

Chopin, N. (2002) A sequential particle filter for static models. *Biometrika*, 89, 539–552.

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

Chorin, A., Morzfeld, M. and Tu, X (2010) Implicit particle filters for data assimilation. *Communs Appl. Math. Computul Sci.*, 5, 221–240.

Chorin, A. and Tu, X. (2009) Implicit sampling for particle filters. *Proc. Natn. Acad. Sci. USA.*, 106, 17249–17254.

Cornebise, J., Moulines, É. and Olsson, J. (2008) Adaptive methods for sequential importance sampling with application to state space models. *Statist. Comput.*, 18, 461–480.

Cvetanovska, V. and Stojanovski, T. (2012) Using high performance computing and monte carlo simulation for pricing american options. *Preprint arXiv:1205.0106*.

Dacunha-Castelle, D. and Florens-Zmirou, D. (1986) Estimation of the coefficients of a diffusion from discrete observations. *Stochastics*, 19, 263–284.

Del Moral, P. (2013) *Mean Field Simulation for Monte Carlo Integration*. Boca Raton: Chapman and Hall.

Del Moral, P., Doucet, A. and Jasra, A. (2006) Sequential Monte Carlo samplers. *J. R. Statist. Soc. B.*, 68, 411–436.

Dick, J., Kuo, F. Y. and Sloan, I. H. (2013) High-dimensional integration: the Quasi-Monte Carlo way. *Acta Numer.*, 22, 133–288.

Douc, R., Garivier, A., Moulines, E. and Olsson, J. (2011) Sequential Monte Carlo smoothing for general state space hidden Markov models. *Ann. Appl. Probab.*, 21, 2109–2145.

Doucet, A., Pitt, M., Deligiannidis, G. and Kohn, R. (2015) Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator. *Biometrika*, to be published, doi 10.1093/biomet/asu075.

Dutang, C. and Savicky, P. (2013) *randtoolbox: generating and testing random numbers*. Available from CRAN.R-project.org.

Gundersen, H. J. G. and Jensen, E. B. (1987) The efficiency of systematic sampling in sterology and its prediction. *J. Microsc.*, 147, 229–263.

Hall, J., Pitt, M. K. and Kohn, R. (2014) Bayesian inference for nonlinear structural time series models. *J. Econometr.*, 179, 99–111.

He, Z. and Owen, A. B. (2014) Extensible grids: uniform sampling on a space-filling curve. *Technical Report*. Stanford University, Stanford.

Heinrich, S., Novak, E., Wasilkowski, G. W. and Woźniakowski, H. (2001) The inverse of the star-discrepancy depends linearly on the dimension. *Acta Arith.*, 96, 279–302.
Hickernell, F., Lemieux, C. and Owen, A. (2005) Control variates for Quasi-Monte Carlo. *Statist. Sci.*, 20, 1–31.

Hörmann, W., Leydold, J. and Derflinger, G. (2004) *Automatic Nonuniform Random Variate Generation*. Berlin: Springer.

Hu, X., Kuipers, D. and Zeng, Y. (2014) Bayesian inference via filtering equations for ultra-high frequency data (i): model and estimation. To be published.

James, L. F., Lijoi, A. and Prünster, I. (2009) Posterior analysis for normalized random measures with independent increments. *Scand. J. Statist.*, 36, 76–97.

Johansen, A. M., Whiteley, N. and Doucet, A. (2012) Exact approximation of Rao-Blackwellised particle filters. *Syst. Identifiern*, 16, 488–493.

Julier, S. J. and Uhlmann, J. K. (1997) A new extension of the Kalman filter to non-linear system. In *Proc. AeroSense: 11th Int. Symp. Aerospace/Defense Sensing, Simulation and Controls*. Orlando: International Society for Optical Engineering.

Kahn, H. and Harris, T. E. (1949) Estimation of particle transmission by random sampling. In *Monte Carlo Method*, pp. 27–30. Gaithersburg: National Bureau of Standards.

Kantas, N., Doucet, A., Singh, S. H., Maciejowski, J. and Chopin, N. (2014) On particle methods for parameter estimation in state-space models. *Ann. Statist. Sci.*, to be published.

Kearns, M., Mansour, Y. and Ng, A. Y. (2002) A sparse sampling algorithm for near-optimal planning in large Markov decision processes. *Mach. Learn.*, 49, 193–208.

Lacoste-Julien, S., Lindsten, F. and Bach, F. (2015) Sequential kernel herding: Frank-Wolfe optimization for particle filtering. In *Proc. 18th Int. Conf. Artificial Intelligence and Statistics* (eds G. Lebanon and S. V. N. Vishwanathan), pp. 544–552. JMLR.

L'Ecuyer, P., Demers, V. and Tuffin, B. (2007) Rare-events, splitting, and quasi-Monte Carlo. *ACM Trans. Modling Comput. Simuli*, 17, article 9.

L’Ecuyer, P., Lécot, C. and L’Archevèque-Gaudet, A. (2009) On array-RQMC for Markov chains: mapping alternatives and convergence rates. In *Monte Carlo and Quasi-Monte Carlo Methods 2008*, pp. 485–500. Berlin: Springer.

L’Ecuyer, P., Lécot, C. and Tuffin, B. (2006) Randomized quasi-Monte Carlo simulation of Markov chains with an ordered state space. In *Monte Carlo and Quasi-Monte Carlo Methods 2004*, pp. 331–342. New York: Springer.

L’Ecuyer, P., Lécot, C. and Tuffin, B. (2008) A randomized quasi-Monte Carlo simulation method for Markov chains. *Oups Res.*, 56, 958–975.

Liu, J. S. (1996) Nonparametric hierarchical Bayes via sequential imputation. *Ann. Statist.*, 24, 910–930.

Liu, J. S. (2001) *Monte Carlo Strategies in Scientific Computing*. New York: Springer.

Loh, W.-L. (2003) On the asymptotic distribution of scrambled net quadrature. *Ann. Statist.*, 31, 1282–1324.

Lopes, H. F., Carvalho, C. M., Johannes, M. S. and Polson, N. S. (2010) Particle learning for sequential Bayesian computation (with discussion). In *Bayesian Statistics 9* (eds J. M. Bernardo, J. O. Berger, M. J. Bayarri, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West), pp. 175–196. Oxford: Oxford University Press.

MacRae, E. C. (1977) Estimation of time-varying Markov processes with aggregate data. *Econometrica*, 45, 183–198.

Marin, J. M., Pudlo, P., Robert, C. P. and Ryder, R. J. (2012) Approximate Bayesian computational methods. *Statist. Comput.*, 22, 1167–1180.

McElinn, K., Katsura, H. and Nakatsuma, T. (2012) Fully parallel particle learning for gpgpus and other parallel devices. *Preprint arXiv*:1212.1639.

Møller, J., Pettitt, A. N., Reeves, R. and Berthelsen, K. K. (2006) An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. *Biometrika*, 93, 451–458.

Morrison, R. (2012) On the asymptotic distribution of scrambled net quadrature. *Ann. Statist.*, 31, 1282–1324.

Morzfeld, M., Tu, X., Atkins, E. and Chorin, A. J. (2012) A random map implementation of implicit filters. *J. Computnl Phys.*, 231, 2049–2066.

Muller, P. (1999) Simulation based optimal design. In *Bayesian Statistics 6* (eds J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), pp. 459–474. New York: Oxford University Press.

Muller, P., Sanso, B. and De Iorio, M. (2004) Optimal Bayesian design by inhomogeneous Markov chain simulation. *J. Am. Statist. Ass.*, 99, 788–798.

Murray, L. (2012) Gpu acceleration of the particle filter: the Metropolis resampler. *Preprint arXiv*:1202.6163.

Murray, L., Jones, E. and Parslow, J. (2013) On disturbance state-space models and the particle marginal Metropolis-Hastings sampler. *Soc. Industrl Appl. Math.*, 1, 494–521.

von Neumann, J. (1951) Various techniques used in connection with random digits. In *J. Soc. Industrl Appl. Math.*, 494–521.

Novak, E. (1988) *Deterministic and Stochastic Error Bounds in Numerical Analysis*. Berlin: Springer.
Ormoneit, D., Lemieux, C. and Fleet, D. J. (2001) Lattice particle filters. In Proc. 17th Conf. Uncertainty in Artificial Intelligence, pp. 395–402. San Francisco: Morgan Kaufmann.

Owen, A. B. (1997a) Scrambled net variance for integrals of smooth functions. Ann. Statist., 25, 1541–1562.

Owen, A. B. (1997b) Monte Carlo variance of scrambled net quadrature. SIAM J. Numer. Anal., 34, 1884–1910.

Owen, A. B. (1998) Scrambling Sobol’ and Niederreiter-Xing points. J. Complexity, 14, 466–489.

Owen, A. B. (2008) Local antithetic sampling with scrambled nets. Ann. Statist., 36, 2319–2343.

Pitman, J. and Yor, M. (1997) The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. Ann. Probab., 25, 855–900.

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

Pitt, M. K., Silva, R., Giordani, P. and Kohn, R. (2012) On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. J. Econometr., 171, 134–151.

Pollock, M. (2013) Some Monte Carlo methods for jump diffusions. PhD Thesis. Department of Statistics, University of Warwick, Coventry.

Pollock, M., Johansen, A. and Roberts, G. (2014) On the exact and ε-strong simulation of (jump) diffusions. Bernoulli, to be published.

Poyiadjis, G., Doucet, A. and Singh, S. S. (2011) Particle approximations of the score and observed information matrix in state space models with applications to parameter estimation. Biometrika, 98, 65–80.

Raftery, A. E. and Bao, L. (2010) Estimating and projecting trends in HIV/AIDS generalized epidemics using incremental mixture importance sampling. Biometrics, 66, 1162–1173.

Rebeschini, P. and Van Handel, R. (2015) Can local particle filters beat the curse of dimensionality? Ann. Appl. Probab., to be published.

Regazzini, E., Lijoi, A. and Prünster, I. (2003) Distributional results for means of normalized random measures with independent increments. Ann. Statist., 31, 560–585.

Reich, S. and Cotter, C. J. (2013) Ensemble filter techniques for intermittent data assimilation. In Large Scale Inverse Problems: Computational Methods and Applications in the Earth Sciences (eds M. Cullen, M. A. Freitag, S. Kindermann and R. Scheichl), pp. 91–134. Berlin: de Gruyter.

Rust, J. (1997) Using randomization to break the curse of dimensionality. Econometrica, 65, 487–516.

Sherlock, C., Thiery, A., Roberts, G. and Rosenthal, J. (2015) On the efficiency of pseudo-marginal random walk Metropolis algorithms. Ann. Statist., 43, 238–275.

Shinozuka, M. and Deodatis, G. (1996) Simulation of multi-dimensional Gaussian stochastic field by spectral representation. Appl. Mech. Rev., 49, 29–53.

Skilling, J. (2004) Using the Hilbert curve. In Bayesian Inference Maximum Entropy Methods in Sciences and Engineering: Proc. 23rd Int. Wrkshp Bayesian Inference and Maximum Entropy Methods in Science and Engineering, American Institute of Physics Publishing.

Steigleder, M. and McCool, M. D. (2003) Generalized stratified sampling using the Hilbert curve. J. Graph. Tools, 8, 41–47.

Stein, C. (1986) Approximate Computation of Expectations. Hayward: Institute of Mathematical Statistics.

Tuffin, B. (2008) Randomization of quasi-Monte Carlo methods for error estimation: survey and normal approximation, Monte Carlo methods and application. Monte Carlo Meth. Appl., 10, 617–628.

Vapnik, V. and Cervonenkis, A. (1971) On the uniform convergence of relative frequencies events to their probabilities, Theor. Probab. Appl., 16, 264–280.

Wang, X. and Sloan, I. H. (2011) Quasi-Monte Carlo methods in financial engineering: an equivalence principle and dimension reduction. Ops Res., 59, 80–95.

Zhou, Y. (2013) vsmc: parallel sequential monte carlo in c++. Preprint arXiv:1306.5583.