

HILBERT SPACE ANALYSIS OF LATIN HYPERCUBE SAMPLING

PETER MATHÉ

(Communicated by David Sharp)

ABSTRACT. Latin Hypercube Sampling is a specific Monte Carlo estimator for numerical integration of functions on \mathbb{R}^d with respect to some product probability distribution function. Previous analysis established that Latin Hypercube Sampling is superior to independent sampling, at least asymptotically; especially, if the function to be integrated allows a good additive fit. We propose an explicit approach to Latin Hypercube Sampling, based on orthogonal projections in an appropriate Hilbert space, related to the ANOVA decomposition, which allows a rigorous error analysis. Moreover, we indicate why convergence cannot be uniformly superior to independent sampling on the class of square integrable functions. We establish a general condition under which uniformity can be achieved, thereby indicating the rôle of certain Sobolev spaces.

1. INTRODUCTION

Latin Hypercube Sampling is a specific Monte Carlo estimator for numerical integration of functions on \mathbb{R}^d . Thus we aim at approximating the integral $\int f(x)dF(x)$ against a given probability distribution function F by means of a Monte Carlo procedure, which is the sample mean $\vartheta_N(f) = \frac{1}{N} \sum_{j=1}^N f(X_j)$ of N randomly chosen points X_1, \dots, X_N in \mathbb{R}^d . For the moment we leave the way of choosing these sample points unspecified and introduce the error of (any) such randomized quadrature formula as

$$e(\vartheta_N, f) := \left(\mathbf{E} \left| \vartheta_N(f) - \int_{\mathbb{R}^d} f(x)dF(x) \right|^2 \right)^{1/2}.$$

We assume throughout that integrands belong to the Hilbert space of square integrable functions with respect to F . This will be the basic Hilbert space we shall be working in. It will be denoted by $L_2(dF)$ and its norm will be denoted by $\|\cdot\|_2$. For the above error criterion the standard Monte Carlo estimator, which is the sample mean of N independent identically distributed according to F points, denoted by τ_N for a moment, obeys $e(\tau_N, f)^2 = \text{Var}(f)/N$. In general, there will not be an exact expression for the error. Instead, the normalized error will converge to a

Received by the editors August 25, 1999.

2000 *Mathematics Subject Classification*. Primary 65C05; Secondary 62D05.

Key words and phrases. Latin Hypercube Sampling, stratified sampling, asymptotic variance.

©2000 American Mathematical Society

certain function $\Phi: L_2(dF) \rightarrow \mathbb{R}^+$, i.e.,

$$\lim_{N \rightarrow \infty} \sqrt{N}e(\vartheta_N, f) = \Phi(f).$$

For standard Monte Carlo we have

$$\Phi(f) = \sqrt{\text{Var}(f)} = \|f - \int f(x)dF(x)\|_2.$$

We note that Φ may thus be understood as measuring the distance of the integrand to the constant functions. If this is small, then the rate $N^{-1/2}$ is accompanied with a small constant. For different estimators than standard Monte Carlo, an appropriate function Φ will express different features. For Latin Hypercube Sampling, to be described below, the corresponding Φ_{LHS} measures the closeness of the integrands to the additive functions. Precisely,

$$\Phi_{LHS}(f) := \inf \left\{ \|f - h\|_2, h(x) = \sum_{l=1}^d h_l(x_l), x = (x_1, \dots, x_d) \in \mathbb{R}^d \right\},$$

where the infimum is taken over all linear combinations of univariate functions. This result goes back to Stein [20, Cor. 1] (see also [11, Thm. 1]). It will be transparent that this infimum is realized and the corresponding approximant is called *best additive fit*.

In general, Monte Carlo estimators will be applied to a variety of functions. If convergence of $\sqrt{N}e(\vartheta_N, f)$ to $\Phi(f)$ strongly depends on the integrand, then, given N , the value of $\Phi(f)/\sqrt{N}$ may not reflect the actual error appropriately. In contrast, if we knew that for a specific class B of integrands

$$\sup_{f \in B} \left| \sqrt{N}e(\vartheta_N, f) - \Phi(f) \right| \rightarrow 0,$$

then we would gain additional accuracy of our estimate. The standard Monte Carlo estimator fulfills this kind of uniformity in a trivial manner.

The aim of the present paper is the error asymptotics of Latin Hypercube Sampling. We will reprove Stein's result on the representation of Φ_{LHS} , stated as Theorem 1. Moreover we shall indicate that in general convergence is not uniform in the above sense. However, we provide instances where it is. This analysis is comprised in Corollary 1.

The paper is organized as follows. We begin studying stratified sampling in one dimension. The results proven there are not needed for proving the main results, but we think that they highlight some of the features, which might not be as clear in higher dimensions.

Then we state the main results in Section 4. In Section 5 we will prepare ourselves by identifying the best additive fit as orthogonal projection. We prove the main results in Section 6. Our proof of Theorem 1 differs from those given previously and enables us to obtain uniform convergence as indicated in Corollary 1. The final Section 7 is devoted to Latin Hypercube Sampling applied to vector-valued functions.

We present the material in such a way that it also surveys previous results. However, we use Hilbert space techniques throughout, since this seems to be the natural context.

2. DESCRIPTION OF LATIN HYPERCUBE SAMPLING

Latin Hypercube Sampling provides a specific Monte Carlo estimator for numerical integration of functions f on \mathbb{R}^d with respect to some product probability distribution function F , which can be written as

$$F(x) = \prod_{l=1}^d F_l(x_l), \quad x = (x_1, \dots, x_d) \in \mathbb{R}^d.$$

The intention of Latin Hypercube Sampling is to stratify each margin in order to improve performance. This method was proposed and first analyzed in [12, 7]. We refer to [20] for a detailed account.

Below, we assume that each of the F_l has a density with bounded support, which is bounded away from 0 there, and that for each N we are able to compute quantiles

$$x_i^l := \inf \left\{ u, \quad F_l(u) \geq \frac{i-1}{N} \right\}, \quad i = 1, \dots, N, \quad l = 1, \dots, d.$$

This results in a tiling of \mathbb{R}^d into hypercubes $\chi_i := \prod_{l=1}^d [x_i^l, x_{i+1}^l)$ (we let $x_{N+1}^l := \infty$) labelled by $\mathbf{i} \in \{1, \dots, N\}^d$, which we simply denote by N^d . By construction $\int_{\chi_i} dF(x) = N^{-d}$ for each $\mathbf{i} \in N^d$.

Now, Latin Hypercube Sampling consists in a specific way to pick N of these hypercubes and estimating $\int_{\chi_i} f(x)dF(x)$ by just picking one point, say X_i with

$$P(X_i \in C) = N^d F(C \cap \chi_i), \quad \text{for Borel sets } C \in \mathbb{R}^d.$$

We observe that then $\mathbf{E}f(X_i) = N^d \int_{\chi_i} f(x)dF(x)$. We finally take the sample mean as the estimator for $\int f(x)dF(x)$. More specifically, if $A \subset N^d$ denotes the chosen N hypercubes, then

$$\vartheta(f) := \frac{1}{N} \sum_{\mathbf{i} \in A} f(X_i).$$

Here randomness occurs when choosing A and the X_i , which is assumed to be done independently. With a slight abuse of notation we denote the estimator by $\vartheta_N(f)$, stressing its dependence of the sample size only.

To complete the construction of Latin Hypercube Sampling we need to specify the way of choosing some pattern A at random. This can be done by choosing permutations π_1, \dots, π_d of $\{1, \dots, N\}$ independently and uniformly. We then let

$$A := \{(\pi_1(k), \dots, \pi_d(k)), \quad k = 1, \dots, N\}.$$

By construction, each \mathbf{i} is equally likely to be in A , resulting in

$$p_i := P(\mathbf{i} \in A) = N/N^d = N^{-(d-1)}.$$

We further need to describe pairwise correlations. Therefore we note that a pair (\mathbf{i}, \mathbf{j}) may be chosen only if $i_l \neq j_l$, $l = 1, \dots, d$. We find it convenient to rewrite this as $\prod_{l=1}^d (1 - \delta(i_l, j_l)) = 1$, with $\delta(m, n) = 1$ if and only if $m = n$. With this notation pairwise correlation is

$$(1) \quad p_{ij} := P((\mathbf{i}, \mathbf{j}) \in A) = \frac{1}{(N(N-1))^{d-1}} \prod_{l=1}^d (1 - \delta(i_l, j_l)).$$

A more conventional way of describing the Latin Hypercube Sampling procedure is presented in [20, 16, 11]. But the enumeration of the X_i , $i \in A$ is very artificial and is averaged out taking the sample mean.

We close this section by showing that $\vartheta_N(f)$ is unbiased for integrable functions $f \in L_1(dF)$:

$$\begin{aligned} \mathbf{E}\vartheta_N(f) &= \mathbf{E} \frac{1}{N} \sum_{i \in A} f(X_i) = \frac{1}{N} \sum_{i \in N^d} \mathbf{E}(f(X_i)\mathbb{I}_A(i)) \\ &= \frac{1}{N} \sum_{i \in N^d} \mathbf{E}(f(X_i)\mathbb{I}_A(i)/\mathbb{I}_A(i))p_i \\ &= \frac{1}{N} \sum_{i \in N^d} N^d \int_{\mathcal{X}_i} f(x)dF(x)p_i \\ &= \sum_{i \in N^d} \int_{\mathcal{X}_i} f(x)dF(x) = \int f(x)dF(x). \end{aligned}$$

We did this in much detail, since this is a simple instance of calculations to be carried out more often below.

3. STRATIFIED SAMPLING ON $[0, 1]$

Stratification as a means of variance reduction is well established; we mention only [6, 4.3] and [8, 4.5].

Below we restrict to the special situation when F is the uniform distribution on $[0, 1]$ and we stratify equally spaced. To be precise, given N , let ξ_j , $j = 1, \dots, N$, be independent and uniformly distributed on intervals $\Delta_j := [(j-1)/N, j/N)$, respectively. These intervals generate a σ -algebra, denoted by \mathcal{F}_N . Conditional expectations with respect to \mathcal{F}_N can be calculated as

$$\mathbf{E}(g/\mathcal{F}_N) = \sum_{j=1}^N \left(N \int_{\Delta_j} g(x)dx \right) \mathbb{I}_{\Delta_j}(x);$$

hence

$$\|\mathbf{E}(g/\mathcal{F}_N)\|_2^2 = \sum_{j=1}^N \left(\sqrt{N} \int_{\Delta_j} g(x)dx \right)^2.$$

For a given function $g \in L_2([0, 1])$ the corresponding sample mean is denoted by $\vartheta_N(g) := \frac{1}{N} \sum_{j=1}^N g(\xi_j)$. We mention that this kind of stratification can be observed by Latin Hypercube Sampling in each margin. Therefore it seems natural to recourse to this situation more explicitly.

The main error properties of this stratification are summarized in

Proposition 1. 1. $\vartheta_N(g)$ is unbiased.

2. $\sqrt{N}e(\vartheta_N, g) = \|g - \mathbf{E}(g/\mathcal{F}_N)\|_2$.

3. For each square integrable function g we have

$$(2) \quad \lim_{N \rightarrow \infty} \sqrt{N}e(\vartheta_N, g) = 0,$$

while the uniform error obeys

$$(3) \quad \lim_{N \rightarrow \infty} \sqrt{N} \sup_{\|g\|_2 \leq 1} e(\vartheta_N, g) = 1.$$

Proof. Plainly, for any $g \in L_2([0, 1])$ we have $\mathbf{E}g(\xi_j) = N \int_{(j-1)/N}^{j/N} g(x) dx$. Therefore $\vartheta_N(g)$ is easily seen to be unbiased. To prove the second assertion we only need to consider functions which integrate to 0. For those functions we obtain

$$\begin{aligned} Ne^2(\vartheta_N, g) &= \frac{1}{N} \sum_{k,l=1}^N \mathbf{E}g(\xi_k)g(\xi_l) \\ &= \sum_{k=1}^N \int_{\Delta_k} g^2(x) dx + N \sum_{k \neq l} \left(\int_{\Delta_k} g(x) dx \right) \left(\int_{\Delta_l} g(x) dx \right) \\ &= \|g\|_2^2 + N \left(\sum_{k=1}^N \int_{\Delta_k} g(x) dx \right)^2 - N \sum_{k=1}^N \left(\int_{\Delta_k} g(x) dx \right)^2 \\ &= \|g\|_2^2 - \sum_{k=1}^N \left(\sqrt{N} \int_{\Delta_k} g(x) dx \right)^2 \\ &= \|g\|_2^2 - \|\mathbf{E}(g/\mathcal{F}_N)\|_2^2 = \|g - \mathbf{E}(g/\mathcal{F}_N)\|_2^2. \end{aligned}$$

It is well known that for every $g \in L_2([0, 1])$ the conditional expectations $\mathbf{E}(g/\mathcal{F}_N)$ converge to g in $L_2([0, 1])$. We refer to [18, Prop. 9, Chapt. 6.4] for an explicit proof. This establishes (2).

However, $\|g - \mathbf{E}(g/\mathcal{F}_N)\|_2^2 \rightarrow 0$ does not hold uniformly for all g with norm bounded by some constant. To see this let $\Delta_k^{2N} = [(k-1)/2N, k/2N)$ be the partition corresponding to \mathcal{F}_{2N} . We consider the function

$$g^N(x) := \sum_{k=1}^{2N} (-1)^k \mathbb{I}_{\Delta_k^{2N}}(x), \quad x \in [0, 1].$$

By construction $\mathbf{E}(g^N/\mathcal{F}_N) = 0$, but $\|g^N\|_2 = 1$; hence $Ne(\vartheta_N, g^N)^2 = 1$. This establishes (3) and accomplishes the proof. \square

The asymptotics in (2) may be viewed as *super-convergence*. But, as (3) indicates, no additional speed can be monitored unless we have further information concerning the integrand, say some kind of smoothness. Let $H^1([0, 1])$ denote the the Sobolev space

$$H^1([0, 1]) := \{g \in L_2([0, 1]), g \text{ abs. cont.}, g' \text{ in } L_2([0, 1])\}.$$

We restrict considerations to the ball

$$B^1 := \{g \in H^1([0, 1]), \|g'\|_2 \leq 1\}.$$

On this class of functions stratified sampling provides an order optimal Monte Carlo integration procedure.

Proposition 2.

$$\sup_{g \in B^1} e(\vartheta_N, g) \leq \frac{1}{\pi N^{3/2}}.$$

Thus we obtain a uniform error bound with rate $N^{-3/2}$, which is known to be best possible (see [15, 2.2.9]). Typically, *ibid.*, this bound is achieved by separation of the main part (control variates) and we find it interesting that stratification does as well and is easier to carry out. We indicate the proof, since it is based on a specific version of the Poincaré inequality, exploited in more general form for Latin Hypercube Sampling, below.

Proof of Proposition 2. In view of Proposition 1 we need to show $\|g - \mathbf{E}(g/\mathcal{F}_N)\|_2 \leq (\pi N)^{-1}$ for $g \in B^1$, only. As already used above we have

$$(4) \quad \|g - \mathbf{E}(g/\mathcal{F}_N)\|_2^2 = \sum_{k=1}^N \int_{\Delta_k} \left| g(x) - N \int_{\Delta_k} g(y) dy \right|^2 dx.$$

The heart of further consideration is a special version of the Poincaré inequality, known as Wirtinger inequality (see [13, Chapt. 2, 64] for details), with best possible constant $1/\pi$. For any function $f \in B^1$ we can find $h \in L_2([0, 1])$ with $\|h\|_2 \leq 1$ with $f(x) = \int_0^x h(t) dt$. With this terminology the Wirtinger inequality reads

$$\|f - \int_0^1 f(x) dx\|_2 \leq \frac{1}{\pi} \|h\|_2.$$

An elegant argument with constant $1/\sqrt{8}$ is given in [19].

Scaling this to any subinterval Δ_k by letting $f(x) := g((k-1+x)/N)$ we obtain

$$\|g - N \int_{\Delta_k} g(x) dx\|_{\Delta_k} \leq \frac{1}{\pi N} \|h\|_{\Delta_k}.$$

Inserting this into (4) yields

$$\|g - \mathbf{E}(g/\mathcal{F}_N)\|_2 \leq \frac{1}{\pi N} \|h\|_2 \leq \frac{1}{\pi N},$$

whenever $g \in B^1$. The proof is complete. \square

4. MAIN RESULTS

Latin Hypercube Sampling, as described in Section 2, has a remarkable feature. As Stein [20] puts it: *LHS stratifies each marginal distribution of X_1, \dots, X_N as much as possible but otherwise picks the X_i 's randomly.* By doing this it performs, at least asymptotically, better than standard Monte Carlo Sampling, especially for functions which are close to being additive. This was expressed in terms of the function Φ_{LHS} introduced in the Introduction. Stein's result can be stated as

Theorem 1.

$$\lim_{N \rightarrow \infty} \sqrt{N} e(\vartheta_N, f) = \Phi_{LHS}(f) \text{ for each } f \in L_2(dF).$$

When proving this asymptotic by evaluating the squared error, in both references [20, 11] the authors pick the main contributions and claim that the remainder is of lower order. We aim at presenting a clear and complete proof, after establishing the best additive fit as an orthogonal projection in Section 5.

Moreover, we shall see that convergence is uniform on every relatively compact subset $B \in L_2(dF)$.

Corollary 1. 1.

$$\lim_{N \rightarrow \infty} \sup_{\|f\|_2 \leq 1} \left| \sqrt{N} e(\vartheta_N, f) - \Phi_{LHS}(f) \right| = 1.$$

2. If B is any relatively compact class of functions in $L_2(dF)$, then

$$(5) \quad \lim_{N \rightarrow \infty} \sup_{f \in B} \left| \sqrt{N} e(\vartheta_N, f) - \Phi_{LHS}(f) \right| = 0.$$

Remark 1. The second assertion in Corollary 1 is even a characterization. Precisely, if for some class B of functions

$$\lim_{N \rightarrow \infty} \sup_{f \in B} \left| \sqrt{N} e(\vartheta_N, f) - \Phi_{LHS}(f) \right| = 0,$$

then the class must be relatively compact in $L_2(dF)$. This can be drawn from the proof, in particular from Propositions 4 and 5.

We note that in contrast to stratification in one dimension we do not gain additional speed of convergence (cf. Proposition 2), but the best we can hope for is improving the constant.

Since one point sets are compact, the asymptotics in (5) generalize Stein's result, Theorem 1.

Extending the analysis from Section 3, we consider explicitly the classical Sobolev space

$$H^1(dF) := \left\{ f \in L_2(dF), \text{ each } \frac{\partial f}{\partial x_l} \text{ ex. in gen. sense, and } \left\| \frac{\partial f}{\partial x_l} \right\|_2 < \infty \right\}.$$

This space occurs often in numerical analysis, since it is related to the Dirichlet problem. We mention [1, 2, 4, 9] for further information. We note explicitly that $H^1 \subset L_2(dF)$ is dense.

Of special interest will be the ball

$$B^1 := \left\{ f \in H^1(dF), \sum_{l=1}^d \left\| \frac{\partial f}{\partial x_l} \right\|_2^2 \leq 1 \right\}.$$

For this class of functions, relative compactness, hence uniformity of convergence, can be derived from the Poincaré inequality. We shall present a direct proof of uniform convergence on this class of integrands, as stated in (5). This is accomplished using the Poincaré inequality to establish a rate of convergence of certain conditional expectations (see Proposition 5), a result which is interesting in its own.

5. IDENTIFYING THE BEST ADDITIVE FIT

We start with the representation of the best additive fit h_a to a function f on \mathbb{R}^d as in [20],

$$h_a(x) := \sum_{l=1}^d g_l(x_l) - (d-1) \int f(x) dF(x),$$

where $g_l(x_l) := \int f(x) \prod_{j \neq l} dF_j(x_j)$, $l = 1, \dots, d$. In [20, Appendix B] it is proven that the best additive fit has this representation. It is also clear that $f \rightarrow h_a$ is linear and we aim at identifying this as an orthogonal projection in $L_2(dF)$. For this purpose we let \mathcal{F} denote the Borel σ -algebra on \mathbb{R}^d and we introduce

σ -algebras $\mathcal{F}_l := \sigma \left\{ \mathbb{R} \times \mathbb{R} \times \dots \times \overset{l}{A} \times \dots \times \mathbb{R}, A \in \mathcal{F}(\mathbb{R}) \right\}$. With this notation we have $g_l(x_l) = \mathbf{E}(f/\mathcal{F}_l)$, $l = 1, \dots, d$. By denoting $\mathcal{F}_0 := \{\emptyset, \mathbb{R}\}$, the trivial σ -algebra, thus $\int f(x)dF(x) = \mathbf{E}(f/\mathcal{F}_0)$ we identify

$$h_a = \sum_{l=1}^d \mathbf{E}(f/\mathcal{F}_l) - (d-1)\mathbf{E}(f/\mathcal{F}_0) = \mathbf{E}(f/\mathcal{F}_0) + \sum_{l=1}^d (\mathbf{E}(f/\mathcal{F}_l) - \mathbf{E}(f/\mathcal{F}_0)).$$

We now provide an explicit representation for the orthogonal projection. For each of the σ -algebras \mathcal{F}_l , $l = 0, \dots, d$, we introduce projections $P_l f := \mathbf{E}(f/\mathcal{F}_l)$ acting on $L_2(dF)$. Using the properties of conditional expectations this is readily seen to be the case (see [14, Prop. IV.3.1] for an explicit proof). Now we rewrite h_a as $Qf := P_0 f + \sum_{l=1}^d (P_l f - P_0 f)$. We have

Proposition 3. 1. Q is an orthogonal projection.
 2. For any $f \in L_2(dF)$ we have

$$\|Qf\|_2^2 = \|P_0 f\|_2^2 + \sum_{l=1}^d (\|P_l f\|_2^2 - \|P_0 f\|_2^2).$$

Proof. Since each P_l , $l = 0, \dots, d$, was self-adjoint, the same holds true for Q . It remains to prove $Q^2 = Q$. We note that for $k \neq l$

$$P_k P_l f = \mathbf{E}(\mathbf{E}(f/\mathcal{F}_l)/\mathcal{F}_k) = \int f(x)dF(x) = P_0 f,$$

since integration with respect to x_l which is missing in the inner integral is carried out in the outer integral. Therefore, for each $k = 0, \dots, d$, we have

$$P_k Qf = P_k P_0 f + \sum_{j=1}^d (P_k P_j f - P_k P_0 f) = P_k f$$

and consequently $Q^2 f = Qf$. The second assertion is as simple.

$$\begin{aligned} \|Qf\|_2^2 &= \langle Qf, f \rangle = \langle P_0 f, f \rangle + \sum_{l=1}^d (\langle P_l f, f \rangle - \langle P_0 f, f \rangle) \\ &= \|P_0 f\|^2 + \sum_{l=1}^d (\|P_l f\|_2^2 - \|P_0 f\|_2^2). \end{aligned}$$

The proof is complete. □

We mention without proof the representation $Q = I - \prod_{l=1}^d (I - P_l)$.

Using projection Q we can now rewrite $\Phi_{LHS}(f) = \|(I - Q)f\|_2$ and the assertion of Theorem 1 as

$$\lim_{N \rightarrow \infty} \sqrt{N} e(\vartheta_N, f) = \|(I - Q)f\|_2.$$

Since $(I - Q)(I - P_0) = I - Q$ we derive

$$\|(I - Q)f\|_2 = \|(I - Q)(I - P_0)f\|_2 \leq \|(I - P_0)f\|_2 = \sqrt{Var(f)}.$$

For this reason Latin Hypercube Sampling is superior to independent sampling, at least asymptotically, once we have proven Theorem 1.

6. ERROR ANALYSIS OF LATIN HYPERCUBE SAMPLING

Our presentation will benefit from various types of conditional expectations, all seen as orthogonal projections. We recall that Latin Hypercube Sampling is based on tiling χ_i , $i \in N^d$. These induce σ -algebras $\mathcal{F}_N := \sigma(\{\chi_i, i \in N^d\})$. Letting

$$f_N := \mathbf{E}(f/\mathcal{F}_N) = P_N f,$$

we obtain functions which are constant on each of the sub-cubes, labelled by i . We denote the corresponding values by f_i , thus we can write $f_N = \sum_{i \in N^d} f_i \chi_i$. In particular

$$\int_{\chi_i} f(x) dF(x) = \int_{\chi_i} f_N(x) dF(x) = N^{-d} f_i.$$

The key representation is provided in

Lemma 1. 1. For any $f \in L_2(dF)$ with $\int f(x) dF(x) = 0$ we have

$$(6) \quad e(\vartheta_N, f)^2 = \frac{1}{N} \|f\|_2^2 + (N/(N-1))^{d-1} \sum_{i,j \in N^d} \prod_{l=1}^d (1 - \delta(i_l, j_l)) (N^{-d} f_i) (N^{-d} f_j).$$

2. $Ne(\vartheta_N, f)^2 - Ne(\vartheta_N, f_N)^2 = \|f - f_N\|_2^2$.
3. $QP_N = P_N Q$.

Proof. Relation (6) is a direct consequence of pairwise correlation (1). We note that the double sum is the same for f and f_N . To prove the remaining assertions we note that, since each of the expressions does not change when switching from f to $f - \int f(x) dF(x)$, we may and do assume $\int f(x) dF(x) = 0$. Therefore, using (6) we obtain

$$Ne(\vartheta_N, f)^2 - Ne(\vartheta_N, f_N)^2 = \|f\|_2^2 - \|f_N\|_2^2 = \|f - f_N\|_2^2,$$

since f_N is the result of a projection. To prove the last assertion it is enough to verify $P_N P_l = P_l P_N$ for each $l = 1, \dots, d$. But this can be seen to equal, for $x \in \chi_i$,

$$(P_N P_l f)(x) = N \int_{\mathbb{R} \times \dots \times \Delta_{i_l} \times \dots \times \mathbb{R}} f(x) dF(x).$$

The proof is complete. \square

The main results, Theorem 1 and Corollary 1, are immediate consequences of the following three propositions:

Proposition 4. For each $f \in L_2(dF)$ the following bounds hold true:

$$\begin{aligned} |Ne(\vartheta_N, f)^2 - \|(I-Q)f\|_2^2| &\leq \|f - f_N\|_2^2 + |Ne(\vartheta_N, f_N)^2 - \|(I-Q)f_N\|_2^2|. \\ |Ne(\vartheta_N, f)^2 - \|(I-Q)f\|_2^2| &\geq \|Q(f - f_N)\|_2^2 - |Ne(\vartheta_N, f_N)^2 - \|(I-Q)f_N\|_2^2|. \end{aligned}$$

Proposition 5. 1. There is a constant $C > 0$, such that for $f \in B^1$, the ball in the Sobolev space $H^1(dF)$, we have

$$(7) \quad \|f - f_N\|_2 \leq \frac{C}{N} \left(\sum_{l=1}^d \left\| \frac{\partial f}{\partial x_l} \right\|_2^2 \right)^{1/2}.$$

2. $\lim_{N \rightarrow \infty} \|f - f_N\|_2 = 0$ for each $f \in L_2(dF)$.

3.

$$\sup_{\|f\|_2 \leq 1} \|Q(f - f_N)\|_2 = 1 \quad \text{for each } N,$$

while for each relatively compact subset $B \subset L_2(dF)$,

$$\lim_{N \rightarrow \infty} \sup_{f \in B} \|f - f_N\|_2 = 0.$$

Proposition 6.

$$\lim_{N \rightarrow \infty} \sup_{\|f\|_2 \leq 1} |Ne(\vartheta_N, f_N)^2 - \|(I-Q)f_N\|_2^2| = 0.$$

While Proposition 4 is a simple use of properties of projections and Proposition 5 gathers facts which are more or less folklore, the heart of our analysis is comprised in Proposition 6, which establishes uniform convergence of Latin Hypercube Sampling, whenever functions are piecewise constant.

We are not aware of any reference for convergence as in (7) or for comparable smoothness assumptions.

We turn to the proofs.

Proof of Proposition 4. In view of assertions 2 and 3 of Lemma 1 and using properties of orthogonal projections we derive

$$\begin{aligned} & Ne(\vartheta_N, f)^2 - \|(I-Q)f\|_2^2 \\ &= Ne(\vartheta_N, f)^2 - Ne(\vartheta_N, f_N)^2 + Ne(\vartheta_N, f_N)^2 - \|(I-Q)f_N\|_2^2 \\ &+ \|(I-Q)f_N\|_2^2 - \|(I-Q)f\|_2^2 \\ &= \|Q(f - f_N)\|_2^2 + Ne(\vartheta_N, f_N)^2 - \|(I-Q)f_N\|_2^2. \end{aligned}$$

Turning to absolute values completes the proof of the Proposition. □

Proof of Proposition 5. Assertion 2 is well known; we refer to [14, Chapt. IV.3]. However, we prefer to provide a direct proof of the first assertion, which we find interesting and which might not be known to a wider audience. We rely on the Poincaré inequality (see [9, Chapt. 11, § 4.2], [21, § 28]). A detailed study, including historical remarks, is presented in the recent monograph [13, Chapt. II, 64.]: There is a constant $C > 0$ such that

$$\|g - \int_{[0,1]^d} g(y)dy\|_2^2 \leq C^2 \sum_{l=1}^d \left\| \frac{\partial g}{\partial x_l} \right\|_2^2 \quad \text{for all } g \in H^1.$$

Scaling this to any sub-cube i , i.e., letting

$$g(x) := f(F_1^{-1}((i_1 - x_1)/N), \dots, F_d^{-1}((i_d - x_d)/N)), \quad x = (x_1, \dots, x_d) \in [0, 1]^d,$$

we infer

$$\|f - N^d \int_{\chi_i} f(y)dy\|_{\chi_i}^2 \leq C^2/N^2 \sum_{l=1}^d \left\| \frac{\partial f}{\partial x_l} \right\|_{\chi_i}^2,$$

which establishes estimate (7) after summing over i , thus uniform convergence on class B^1 .

Since $H^1 \subset L_2(dF)$ is dense and since the norms of P_N are uniformly bounded (by 1), we conclude point-wise convergence of f_N to f on all of $L_2(dF)$, and as a

consequence uniform convergence on any relatively compact subset, the last assertion in 3. Indeed, point-wise convergence and relative compactness imply for each $f \in B$ the estimate

$$\|f - f_N\|_2 \leq \|f - f^k\|_2 + \|f^k - f_N^k\|_2 + \|f_N^k - f_N\|_2 \leq 2\|f - f^k\|_2 + \|f^k - f_N^k\|_2$$

if $\{f^k, k = 1, \dots, M\}$ was an ε -net for B , and f^k was chosen with $\|f - f^k\|_2 \leq \varepsilon$.

It remains to prove the first assertion in 3. Fix any coordinate, say the first. Analogous to the proof of Proposition 1, let x_1^1, \dots, x_{2N}^1 be the partition generated by the distribution function F_1 and put $\Delta_k := [x_k^1, x_{k+1}^1)$. Consider the univariate function

$$f^N(x) := \sum_{k=1}^{2N} (-1)^k \mathbb{I}_{\Delta_k}(x_1), \quad x = (x_1, \dots, x_d).$$

By construction $f_N^N = \mathbf{E}(f^N / \mathcal{F}_N) = 0$. Since moreover $Qf = f, Qf^N = f^N, \|f^N\|_2 = 1$ we derive

$$\|Q(f^N - f_N^N)\|_2 = \|f^N - f_N^N\|_2 = \|f^N\|_2 = 1,$$

accomplishing the proof of the first assertion in 3. and of the proposition. □

Proof of Proposition 6. For the proof it will be convenient to further introduce σ -algebras and corresponding projections. Given a subset $I \subset \{1, \dots, d\}$ of cardinality $\#(I) = k$, we let $\Phi_I: \mathbb{R}^d \rightarrow \mathbb{R}^I$ be defined as

$$\Phi_I(x_1, \dots, x_d) = (x_j)_{j \in I}.$$

The pre-image $\Phi_I^{-1}(\mathcal{B}(\mathbb{R}^I))$ is a σ -algebra, called \mathcal{F}_I . The corresponding projection $P_I f := \mathbf{E}(f / \mathcal{F}_I)$ obeys

$$P_I f = \int f(x) \prod_{j \notin I} dF_j(x_j) =: \int f(x) dF_{-I}.$$

Above, we put $x = (x_I, x_{-I})$ and dF_I means $\prod_{l \in I} dF_l \dots$. We specify, in case $k = 0$, the projection $P_0 := P_\emptyset = \int_{\mathbb{R}^d} f(x) dF(x)$ and for subsets of cardinality one $P_l := P_{\{l\}}, l = 1, \dots, d$, which is consistent with the definition given in Section 5. These projections are close to the well-known ANOVA-decomposition (see [5, 17]).

It will henceforth be important to study the action of P_I on functions f_N . For this reason, given $\chi_i, \chi_j, i, j \in N^d$, we deduce for $\#(I) = k$

$$\begin{aligned} \langle P_I f_i \chi_i, f_j \chi_j \rangle &= N^{k-d} \int \left(\int f_i \chi_i(x) dF_{-I}(x_I) \right) f_j \chi_j(x_I, x_{-I}) dF_{-I}(x_{-I}) dF_I(x_I) \\ (8) \quad &= N^{k-2d} \prod_{l \in I} \delta(i_l, j_l) f_i f_j. \end{aligned}$$

As an immediate consequence

$$(9) \quad \|P_I f_N\|_2^2 = \langle P_I f_N, f_N \rangle = \sum_{i, j \in N^d} \langle P_I f_i \chi_i, f_j \chi_j \rangle = N^{k-2d} \sum_{i, j \in N^d} \prod_{l \in I} \delta(i_l, j_l) f_i f_j.$$

We assume $\int_{\mathbb{R}^d} f_N(x) dF(x) = 0$ and evaluate $e(\vartheta_N, f)^2 = \mathbf{E} |\vartheta_N(f_N)|^2$. Using (6) we derive

$$(10) \quad Ne(\vartheta_N, f_N)^2 = \|f_N\|_2^2 + (N/(N-1))^{d-1} N \sum_{i,j} \prod_{l=1}^d (1 - \delta(i_l, j_l)) N^{-2d} f_i f_j.$$

Taking into account

$$\prod_{l=1}^d (1 - \delta(i_l, j_l)) = \sum_{k=0}^d (-1)^k \sum_{I_k} \prod_{l \in I_k} \delta(i_l, j_l),$$

where \sum_{I_k} denotes the sum running over all subsets of N^d of cardinality k , we derive from (8) for each pair (i, j)

$$N^{-2d} \prod_{l=1}^d (1 - \delta(i_l, j_l)) f_i f_j = \sum_{k=0}^d (-1)^k N^{-k} \sum_{I_k} \langle P_{I_k} f_i \chi_i, f_j \chi_j \rangle;$$

hence

$$\sum_{i,j \in N^d} N^{-2d} \prod_{l=1}^d (1 - \delta(i_l, j_l)) f_i f_j = \sum_{k=0}^d (-1)^k N^{-k} \sum_{I_k} \|P_{I_k} f_N\|_2^2.$$

Using this we continue (10) by

$$\begin{aligned} Ne(\vartheta_N, f_N)^2 &= \|f_N\|_2^2 + (N/(N-1))^{d-1} N \sum_{k=0}^d (-1)^k N^{-k} \sum_{I_k} \|P_{I_k} f_N\|_2^2 \\ &= \|f_N\|_2^2 - (N/(N-1))^{d-1} \sum_{l=1}^d \|P_l f_N\|_2^2 \\ &\quad + (N/(N-1))^{d-1} \sum_{k=2}^d (-1)^k N^{-k+1} \sum_{I_k} \|P_{I_k} f_N\|_2^2. \end{aligned}$$

By virtue of assertion 2 from Proposition 3 we have for functions f_N with $P_0 f_N = 0$

$$\|(I-Q)f_N\|_2^2 = \|f_N\|_2^2 - \sum_{l=1}^d \|P_l f_N\|_2^2.$$

This leads to

$$\begin{aligned} N\mathbf{E} |\vartheta_N(f_N)|^2 - \|(I-Q)f_N\|_2^2 &= - \left((N/(N-1))^{d-1} - 1 \right) \sum_{l=1}^d \|P_l f_N\|_2^2 \\ &\quad + (N/(N-1))^{d-1} \sum_{k=2}^d (-1)^k N^{-k} \sum_{I_k} \|P_{I_k} f_N\|_2^2. \end{aligned}$$

Finally we arrive at

$$\begin{aligned} & \sup_{\|f\|_2 \leq 1} |Ne(\vartheta_N, f_N)^2 - \|(I-Q)f_N\|_2^2| \\ & \leq d \left((N/(N-1))^{d-1} - 1 \right) + (N/(N-1))^{d-1} \sum_{k=2}^d \binom{d}{k} N^{-k+1} \\ & \leq \frac{d(2^d + d)}{N-1} + \frac{2^{2d-1}}{N}, \end{aligned}$$

which completes the proof. \square

7. INTEGRATION OF VECTOR-VALUED FUNCTIONS

The previous analysis may easily be extended to the integration of vector-valued functions, a case which also attracted attention (see [11]). Here we assume that X with norm denoted by $\|\cdot\|$ is a separable real Banach space, endowed with the Borel σ -algebra. We denote by $L_2(dF, X)$ the Banach space of (equivalence classes of) Borel measurable functions having finite strong second moment

$$\|f\|_{L_2(dF, X)} := \left(\int_{\mathbb{R}^d} \|f(x)\|^2 dF(x) \right)^{1/2}.$$

The theory of integration of vector-valued functions is outlined in full detail in [3].

As before Latin Hypercube Sampling may be used as Monte Carlo integration for functions from $L_2(dF, X)$ by letting $\vartheta_N(f) := \frac{1}{N} \sum_{i \in A} f(X_i)$, now taking values in X . The particular case $X = \mathbb{R}^p$ was studied in [11].

Again we agree to consider as error criterion

$$e(\vartheta_N, f) := \left(\mathbf{E} \|\vartheta_N(f) - \int f(x) dF(x)\|^2 \right)^{1/2}.$$

In general we cannot hope to give explicit representation of the error asymptotics; even for independent sampling the rate of convergence of $\vartheta_N(f)$ to $\int f(x) dF(x)$ depends on geometric properties of the Banach space X , in particular the type is important (see [10, Chapt. 9.2] for further information).

But if the Banach space itself is a Hilbert space with scalar product $\langle \cdot, \cdot \rangle$, then we may easily extend Theorem 1, since then $L_2(dF, X)$ turns out to be a Hilbert space with scalar product

$$\langle f, g \rangle_{L_2(dF, X)} := \int_{\mathbb{R}^d} \langle f(x), g(x) \rangle dF(x).$$

In an analogous manner we define

$$P_l(f) := \mathbf{E}(f/\mathcal{F}_l), \quad l = 1, \dots, d$$

and correspondingly Q , now being orthogonal projections in $L_2(dF, X)$. As an analog to Theorem 1 we state

Theorem 2. *Let X be a separable Hilbert space. For all functions $f \in L_2(dF, X)$ the following error asymptotics hold true:*

$$\lim_{N \rightarrow \infty} Ne(\vartheta_N, f)^2 = \|(I-Q)f\|_{L_2(dF, X)}^2.$$

Proof. There is no need to outline the proof in detail. This can be carried out exactly as in Section 6 by replacing products of function values, e.g., $f(X_i)f(X_j)$ by respective scalar products $\langle f(X_i), f(X_j) \rangle$. Thus we obtain complete analogs of Propositions 4 and 6. We only need to establish $\|f - f_N\|_{L_2(dF, X)} \rightarrow 0$. The same proof as for the real-valued case extends to this more general situation (see [22, Chapt. II, Thm. 4.1]). Thus f_N converges to f in $L_2(dF, X)$. \square

We mention that there is a corresponding version of Corollary 1.

It is a different task, again when considering the vector-valued case, to ask for properties of the covariance structure of $\vartheta_N(f)$, a problem not occurring in the real-valued case. Again, a simple instance of this, X being \mathbb{R}^p , equipped with Euclidean norm, was studied in [11]. But, for this specific kind of problems, X need not carry a Hilbert structure.

To be precise, having fixed X let X' be the dual Banach space. Since $\vartheta_N(f)$ is a random element in X , having finite strong second moment we may consider its covariance $C_N^{lhs} : X' \times X' \rightarrow \mathbb{R}$, defined by

$$(11) \quad C_N^{lhs}(a, b) := \mathbf{E} \langle \vartheta_N(f) - \int f dF, a \rangle \langle \vartheta_N(f) - \int f dF, b \rangle, \quad a, b \in X'.$$

We refer to [22, Chapt. 3] for the general framework. We recall in particular that each covariance function is bilinear and symmetric. As a consequence its values on $X' \times X'$ are completely determined by those on the diagonal, since for any symmetric bilinear form $B : X' \times X' \rightarrow \mathbb{R}$ we have

$$(12) \quad B(a, b) = 1/4 (B(a + b, a + b) - B(a - b, a - b)).$$

This will enable a simple proof of

Theorem 3. *For any pair of $a, b \in X'$ we have*

$$(13) \quad \lim_{N \rightarrow \infty} NC_N^{lhs}(a, b) = \int_{\mathbb{R}^d} (I - Q)(\langle f(x), a \rangle)(I - Q)(\langle f(x), b \rangle) dF(x).$$

Proof. By our preceding discussion we only need to show

$$NC_N^{lhs}(a, a) \rightarrow \int_{\mathbb{R}^d} |(I - Q)(\langle f(x), a \rangle)|^2 dF(x), \quad a \in X'.$$

But

$$\begin{aligned} C_N^{lhs}(a, a) &= \mathbf{E} \left| \langle \vartheta_N(f) - \int f(x) dF(x), a \rangle \right|^2 \\ &= \mathbf{E} \left| \vartheta_N(\langle f, a \rangle) - \int \langle f(x), a \rangle dF(x) \right|^2 \\ &= e(\vartheta_N, \langle f, a \rangle)^2. \end{aligned}$$

Now Theorem 1 yields

$$\lim_{N \rightarrow \infty} NC_N^{lhs}(a, a) = \|(I - Q)\langle f, a \rangle\|_2^2 = \int |(I - Q)\langle f(x), a \rangle|^2 dF(x).$$

Thus, the convergence as stated in Theorem 3 holds true for $a = b$, hence for all $a, b \in X'$ by (12), which completes the proof. \square

As in [11, Cor. 1] we may now compare the covariance structure of Latin Hypercube Sampling and of independent sampling, denoted by $C_N^{iid}(a, b)$ henceforth. It is well known and easy to check, using the same reasoning as in Theorem 3, that

$$NC_N^{iid}(a, b) = \int_{\mathbb{R}^d} (\mathbf{I} - P_0)(\langle f(x), a \rangle)(\mathbf{I} - P_0)(\langle f(x), b \rangle) dF(x).$$

As a consequence

$$\begin{aligned} \lim_{N \rightarrow \infty} N(C_N^{iid}(a, a) - C_N^{lhs}(a, a)) &= \|(\mathbf{I} - P_0)\langle f, a \rangle\|_2^2 - \|(\mathbf{I} - Q)\langle f, a \rangle\|_2^2 \\ &= \|Q(\mathbf{I} - P_0)\langle f, a \rangle\|_2^2 \geq 0, \end{aligned}$$

since Q was an orthogonal projection and $P_0 = QP_0$. This is the natural generalization of Corollary 1 in [11].

REFERENCES

- [1] Robert A. Adams. *Sobolev spaces*. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London, 1975. Pure and Applied Mathematics, Vol. 65. MR **56**:9247
- [2] E. B. Davies. *Heat kernels and spectral theory*. Cambridge University Press, Cambridge, 1990. MR **92a**:35035
- [3] J. Diestel and J. J. Uhl, Jr. *Vector measures*. American Mathematical Society, Providence, R.I., 1977. With a foreword by B. J. Pettis, Mathematical Surveys, No. 15. MR **56**:12216
- [4] R. E. Edwards. *Functional Analysis. Theory and Applications*. Holt, Rinehart & Winston, New York, Chicago, 1965. MR **36**:4308
- [5] B. Efron and C. Stein. The jackknife estimate of variance. *Ann. Statist.*, 9(3):586–596, 1981. MR **82k**:62074
- [6] George S. Fishman. *Monte Carlo. Concepts, algorithms, and applications*, Springer-Verlag, New York, 1996. MR **97g**:65019
- [7] R. L. Iman and W. J. Conover. Small sample sensitivity analysis techniques for computer models, with an application to risk assessment. *Comm. Statist., Part A: Theory and Methods*, 9:1749–1874, 1980. MR **81m**:65012
- [8] Malvin H. Kalos and Paula A. Whitlock. *Monte Carlo methods. Vol. I*. John Wiley & Sons Inc., New York, 1986. Basics. MR **88e**:65009
- [9] L. V. Kantorovich and G. P. Akilov. *Функциональный анализ*. “Nauka”, Moscow, third edition, 1984. MR **86m**:46001
- [10] Michel Ledoux and Michel Talagrand. *Probability in Banach spaces. Isoperimetry and processes*, Springer-Verlag, Berlin, 1991. MR **93c**:60001
- [11] W.-L. Loh. On Latin Hypercube Sampling. *Ann. Statist.*, 24:2058–2080, 1996. MR **98k**:62008
- [12] M. D. McKay, W. J. Conover, and R. J. Beckman. A comparison of three methods for selecting values of output variables in the analysis of output from a computer code. *Technometrics*, 21:239–245, 1979. MR **80k**:65012
- [13] D. S. Mitrinović, J. E. Pečarić, and A. M. Fink. *Classical and new inequalities in analysis*. Kluwer Academic Publishers Group, Dordrecht, 1993. MR **94c**:00004
- [14] J. Neveu. *Bases Mathématiques du Calcul des Probabilités*. Maison et Cie, 1964. MR **33**:6659
- [15] Erich Novak. *Deterministic and stochastic error bounds in numerical analysis*. Lect. Notes Math. 1349. Springer-Verlag, Berlin, 1988. MR **90a**:65004
- [16] A. B. Owen. A Central Limit Theorem for Latin Hypercube Sampling. *J. R. Statist. Soc. B*, 54:541–551, 1992. MR **93j**:62047
- [17] Art B. Owen. Monte Carlo variance of scrambled net quadrature. *SIAM J. Numer. Anal.*, 34(5):1884–1910, 1997. MR **98h**:65006
- [18] H. L. Royden. *Real analysis*. Macmillan Publishing Company, New York, third edition, 1988. MR **90g**:00004
- [19] L. Saloff-Coste. Lectures on finite Markov chains. In *Lectures on probability theory and statistics (Saint-Flour, 1996)*, pages 301–413. Springer, Berlin, 1997. MR **99b**:60119
- [20] M. Stein. Large sample properties of simulations using Latin Hypercube Sampling. *Technometrics*, 29:143–151, 1987. MR **88e**:62055; correction MR **91j**:62023

- [21] H. Triebel. *Höhere Analysis*. VEB Deutscher Verlag der Wissenschaften, Berlin, 1972. Hochschulbücher für Mathematik, Band 76. MR **50**:12511
- [22] N.N. Vakhania, V. I. Tarieladze, and S.A. Chobanjan. *Probability Distributions on Banach Spaces*. D. Reidel, Dordrecht, Boston, Lancaster, Tokyo, 1987. MR **97k**:60007

WEIERSTRASS-INSTITUTE FOR APPLIED ANALYSIS AND STOCHASTICS, MOHRENSTRASSE 39,
D- 10117 BERLIN, GERMANY
E-mail address: `mathe@wias-berlin.de`