

Optimal confidence for Monte Carlo integration of smooth functions

Robert J. Kunsch, and Daniel Rudolf

August 27, 2019

Abstract We study the information-based complexity of approximating integrals of smooth functions at absolute precision $\varepsilon > 0$ with confidence level $1 - \delta \in (0, 1)$ using function evaluations within randomized algorithms. The probabilistic error criterion is new in the context of integrating smooth functions. In previous research, Monte Carlo integration was studied in terms of the expected error (or the root mean squared error), for which linear methods achieve optimal rates of the error $e(n)$ in terms of the number n of function evaluations. In our context, usually methods that provide optimal confidence properties exhibit non-linear features. The optimal probabilistic error rate $e(n, \delta)$ for multivariate functions from classical isotropic Sobolev spaces $W_p^r(G)$ with sufficient smoothness on bounded Lipschitz domains $G \subset \mathbb{R}^d$ is determined. It turns out that the integrability index p has an effect on the influence of the uncertainty δ in the complexity. In the limiting case $p = 1$ we see that deterministic methods cannot be improved by randomization. In general, higher smoothness reduces the additional effort for diminishing the uncertainty. Finally, we add a discussion about this problem for function spaces with mixed smoothness.

Keywords Monte Carlo integration · Sobolev functions · information-based complexity · standard information · asymptotic error · confidence intervals

Mathematics Subject Classification (2010) 65C05 · 65D30 · 65J05

Acknowledgements The authors wish to express their gratitude to Erich Novak for many detailed hints and discussions during the work on this paper. We also wish to thank Glenn Byrenheidt, Stefan Heinrich, Lutz Kämmerer, David Krieg, and Mario Ullrich for their advice.

R.J. Kunsch (ORCID ID: <https://orcid.org/0000-0003-0835-9870>)
RWTH Aachen University, Lehrstuhl für Mathematik der Informationsverarbeitung, Pontdriesch 10, 52062 Aachen, Germany
E-mail: kunsch@mathc.rwth-aachen.de

D. Rudolf
Institute for Mathematical Stochastics, Universität Göttingen & Felix-Bernstein-Institute for Mathematical Statistics, Goldschmidtstr. 7, 37077 Göttingen, Germany
E-mail: daniel.rudolf@uni-goettingen.de

1 Introduction

We want to compute the integral

$$\text{INT } f = \int_G f(\mathbf{x}) \, d\mathbf{x} \quad (1)$$

of $f: G \rightarrow \mathbb{R}$ from the unit ball $\mathcal{B}_{\mathcal{W}}$ of a normed linear space \mathcal{W} of functions defined on a domain $G \subset \mathbb{R}^d$, using randomized algorithms that have access to function evaluations as the only source of additional information on f . The domain G must be a measurable set with non-zero Lebesgue measure, and the integrands f must be integrable, such that $\text{INT } f$ is well defined. The numerical discussion requires a few more assumptions on G , the typical domain we have in mind is the unit cube $G = [0, 1]^d$, but many results extend to arbitrary bounded and connected domains G with Lipschitz boundary, see Section A for details on the latter notion. The focus of this paper lies on the (ε, δ) -complexity $n_{\text{prob}}^{\text{MC}}(\varepsilon, \delta, \mathcal{W})$, that is, the minimal number n of function values needed by optimal randomized algorithms A_n in order to approximate the integral (1) such that

$$\mathbb{P}\{|A_n(f) - \text{INT } f| > \varepsilon\} \leq \delta \quad \text{for all } f \text{ with } \|f\|_{\mathcal{W}} \leq 1, \quad (2)$$

where $\|\cdot\|_{\mathcal{W}}$ is the norm of \mathcal{W} . A method with this property of guaranteeing a small (absolute) error $\varepsilon > 0$ with confidence $1 - \delta \in (0, 1)$ (or uncertainty δ) for inputs from the unit ball $\mathcal{B}_{\mathcal{W}}$ is called (ε, δ) -approximating in \mathcal{W} , see also [16]. We further consider the n -th minimal probabilistic Monte Carlo error at uncertainty δ , defined by

$$e_{\text{prob}}^{\text{MC}}(n, \delta, \mathcal{W}) := \inf \{\varepsilon > 0 \mid \exists (\varepsilon, \delta)\text{-approximating algorithm } A_n \text{ in } \mathcal{W}\}. \quad (3)$$

The probabilistic error criterion from above is less common in *information-based complexity* (IBC) where the standard notion of *Monte Carlo error* is some type of mean error. In general, for some $\ell \geq 1$, the ε -complexity $n_{\ell\text{-mean}}^{\text{MC}}(\varepsilon, \mathcal{W})$, with respect to the ℓ -mean error, is the minimal number n of function values required by an algorithm A_n such that

$$\left(\mathbb{E}|A_n(f) - \text{INT } f|^\ell\right)^{1/\ell} \leq \varepsilon \quad \text{for all } f \text{ with } \|f\|_{\mathcal{W}} \leq 1. \quad (4)$$

Accordingly, the n -th minimal ℓ -mean Monte Carlo error is given by

$$e_{\ell\text{-mean}}^{\text{MC}}(n, \mathcal{W}) := \inf_{A_n} \sup_{\|f\|_{\mathcal{W}} \leq 1} \left(\mathbb{E}|A_n(f) - \text{INT } f|^\ell\right)^{1/\ell}. \quad (5)$$

Here, as before, the infimum is taken over all randomized algorithms A_n that use at most n function values. Most frequently studied are the *root mean squared error*, that is, $e_{2\text{-mean}}^{\text{MC}}$, as well as the expected absolute error $e_{1\text{-mean}}^{\text{MC}}$. For more details on IBC we refer to the books [22, 23, 24, 27].

One might argue that for a randomized algorithm A_n using either error criterion (2) or error criterion (4) does not make any difference. However, this is not quite true. For example, using Markov's inequality it is always possible to construct (ε, δ) -approximating algorithms if for any $\varepsilon' > 0$ we know methods that achieve a mean error (4) smaller or equal ε' . That way, however, the cost estimates are not

optimal in terms of the δ -dependence, namely, they are polynomial rather than logarithmic. In some situations this can be fixed by using more advanced inequalities such as Hoeffding bounds. In other situations, commonly known algorithms may need to be modified which leads to more robust methods less prone to outliers. For this reason, the probabilistic error criterion is frequently used in statistics, see for example [8, 11, 12, 13]. It actually stems from statistics as it is equivalent to finding symmetric confidence intervals, namely, if A_n is (ε, δ) -approximating in \mathcal{W} , then $[A_n(f) - \varepsilon, A_n(f) + \varepsilon]$ is a confidence interval at level $1 - \delta$ for the quantity $\text{INT } f$, provided $\|f\|_{\mathcal{W}} \leq 1$. Furthermore, there are numerical problems that can be solved with respect to the probabilistic (ε, δ) -criterion but the mean error is unbounded, see [16]. In other words, the probabilistic criterion seems to yield a very general concept of solvability.

In Section 2 we provide two generic lower bounds for the n -th minimal probabilistic Monte Carlo error based on bump functions and discuss their application to particular function spaces. In Section 3 we study several approaches for deriving upper error bounds on Sobolev classes and discuss in which cases they lead to optimal rates. We mainly consider classical isotropic Sobolev spaces $W_p^r(G)$ on domains $G \subseteq \mathbb{R}^d$. For integer smoothness $r \in \mathbb{N}_0$ and integrability parameter $1 \leq p \leq \infty$, these spaces are given by

$$W_p^r(G) := \left\{ f \in L_p(G) \mid \|f\|_{W_p^r(G)} := \left(\sum_{\substack{\alpha \in \mathbb{N}_0^d \\ |\alpha|_1 \leq r}} \|D^\alpha f\|_{L_p(G)}^p \right)^{1/p} < \infty \right\}, \quad (6)$$

with the usual modification for $p = \infty$ and the weak derivative $D^\alpha f = \partial_{x_1}^{\alpha_1} \cdots \partial_{x_d}^{\alpha_d} f$ for multi-index $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_0^d$. Note that for $r = 0$ we obtain the Lebesgue spaces $L_p(G)$.

Our main result is for spaces $W_p^r(G)$ on bounded domains $G \subset \mathbb{R}^d$ with Lipschitz boundary (see Section A) and with sufficient smoothness, $rp > d$. In asymptotic notation (see definitions below) it states

$$e_{\text{prob}}^{\text{MC}}(n, \delta, W_p^r(G)) \asymp n^{-r/d} \min \left\{ 1, \left(\frac{\log \delta^{-1}}{n} \right)^{1-1/q} \right\} \quad (7)$$

with $q := \min\{p, 2\}$, or equivalently

$$n_{\text{prob}}^{\text{MC}}(\varepsilon, \delta, W_p^r(G)) \asymp \min \left\{ \varepsilon^{-d/r}, \varepsilon^{-1/\left(\frac{r}{d} + \frac{q-1}{q}\right)} (\log \delta^{-1})^{1/\left(\frac{q}{q-1} \cdot \frac{r}{d} + 1\right)} \right\}, \quad (8)$$

see Theorem 1 and Theorem 4. The condition $rp > d$ guarantees that the space $W_p^r(G)$ is compactly embedded in the space of continuous functions, see for instance [1]. Only then function evaluations are well defined and there exist deterministic integration methods, which in this case provide error bounds with rate $n^{-r/d}$. These worst case bounds come into play if we demand extremely high confidence $1 - \delta$ close to 1. It also turns out that for $p = 1$ the uncertainty δ does not play any role, which shows that deterministic methods are optimal in that case. In the power of n , we recover the well known gain of $1 - 1/p$ for $1 < p < 2$, and $1/2$ for $p \geq 2$, which Monte Carlo methods achieve compared to deterministic methods. The influence of the uncertainty δ grows with the gain in the error rate.

In terms of the complexity (8) we observe that the higher the smoothness r , the weaker the dependence on δ .

Weak asymptotic notation: For functions $e, f: \mathbb{N} \times (0, 1) \rightarrow \mathbb{R}$ we use the notation $e(n, \delta) \preceq f(n, \delta)$, meaning that there exist $n_0 \in \mathbb{N}$ and $\delta_0 \in (0, 1)$ such that $e(n, \delta) \leq cf(n, \delta)$ with some (possibly (d, r) -dependent) constant $c > 0$ for all $n \geq n_0$ and $\delta \in (0, \delta_0)$. Sometimes we add the restriction $n \geq \log \delta^{-1}$, then $e(n, \delta) \leq cf(n, \delta)$ is only meant to hold for $\delta \in (0, \delta_0)$ and $n \geq n_0 \log \delta^{-1}$. Similarly we denote asymptotics for complexity functions $n(\varepsilon, \delta)$, describing a behaviour for small $\varepsilon, \delta > 0$. Asymptotic equivalence $e(n, \delta) \asymp f(n, \delta)$ is a shorthand for $e(n, \delta) \preceq f(n, \delta) \preceq e(n, \delta)$. The notion $e(n, \delta) \prec f(n, \delta)$ means “ $e(n, \delta) \preceq f(n, \delta)$ but not $f(n, \delta) \preceq e(n, \delta)$ ”.

2 Lower bounds

We start with the lower bounds as these are easily obtained for the whole parameter range of the function spaces we consider. For now, the assumptions on the domain $G \subseteq \mathbb{R}^d$ are rather weak: In Section 2.1 we only need a positive Lebesgue measure, that is, $\lambda^d(G) > 0$. In Section 2.2 we additionally require that G has a non-empty interior. Both conditions are satisfied for bounded domains with Lipschitz boundary.

2.1 Auxiliary lemmas

Lower bounds shall hold for a very general class of algorithms. For simplicity, though, we mainly restrict to *non-adaptive* algorithms. A description of the more general class of *adaptive* algorithms, together with necessary modifications for the lower bound proofs, can be found in Section C in the appendix. As before, let \mathcal{W} be a space of functions defined on the domain G , equipped with a norm $\|\cdot\|_{\mathcal{W}}$. An abstract *non-adaptive Monte Carlo algorithm* defined for such functions is a family $A = (A^\omega)_{\omega \in \Omega}$ of mappings $A^\omega = \phi^\omega \circ \iota^\omega$, indexed with elements ω from a probability space $(\Omega, \Sigma, \mathbb{P})$. Here, $\iota^\omega: \mathcal{W} \rightarrow \mathbb{R}^{\tilde{n}(\omega)}$ is the so-called *information mapping* yielding information

$$\mathbf{y} = \iota^\omega(f) = (f(\mathbf{x}_1^\omega), \dots, f(\mathbf{x}_{\tilde{n}(\omega)}^\omega))$$

on the problem instance f , namely function values at *random nodes* $\mathbf{x}_i^\omega \in G$, and $\phi^\omega: \mathbb{R}^{\tilde{n}(\omega)} \rightarrow \mathbb{R}$ represents the generation of an output $A^\omega(f) = \phi^\omega(\mathbf{y})$ from that information. The number of used function values is a random variable, denoted by $\text{card}(A): \omega \mapsto \tilde{n}(\omega) \in \mathbb{N}_0$, which we call (*varying*) *cardinality* of A . We assume that the output $A(f): \omega \mapsto A^\omega(f) \in \mathbb{R}$ is measurable for all inputs f , and hence the (probabilistic) Monte Carlo error of the algorithm,

$$\begin{aligned} e(A, \delta, \mathcal{W}) &:= \sup_{\|f\|_{\mathcal{W}} \leq 1} \inf\{\varepsilon > 0 \mid \mathbb{P}\{|A(f) - \text{INT } f| > \varepsilon\} \leq \delta\} \\ &= \sup_{\|f\|_{\mathcal{W}} \leq 1} F_{\text{err}(f)}^{-1}(1 - \delta), \end{aligned} \tag{9}$$

is well defined. Here, $F_{\text{err}(f)}^{-1}$ is the *generalized inverse of the cumulative distribution function* of the real-valued random variable $\text{err}(f) := |A(f) - \text{INT } f|$.

A few words have to be added on the issue of function evaluations, since the Sobolev spaces we consider, see (6), are formally defined as spaces of equivalence classes of functions, identifying functions that coincide almost everywhere, that is, $f_1 \sim f_2$ iff $\lambda^d\{\mathbf{x} \in G: f_1(\mathbf{x}) \neq f_2(\mathbf{x})\} = 0$. The actual input of an algorithm, however, is always a particular function $f: G \rightarrow \mathbb{R}$ representing an equivalence class $[f] \in \mathcal{W}$, regardless of that we usually simply write $f \in \mathcal{W}$. If the space \mathcal{W} is embedded in the space of continuous functions, $\mathcal{W} \hookrightarrow C(G)$, there is a unique continuous representative for each equivalence class and we assume that we compute function values of this particular representative. However, if \mathcal{W} is not embedded in the space of continuous functions, $\mathcal{W} \not\hookrightarrow C(G)$, there exist equivalence classes $[f] \in \mathcal{W}$ without continuous representatives. In that case we only allow algorithms that satisfy for any two representatives f_1 and f_2 of the same equivalence class $[f] \in \mathcal{W}$ that the outputs $A(f_1)$ and $A(f_2)$ almost surely coincide. This can be assured if the distribution of each node $\mathbf{x}_i: \omega \mapsto \mathbf{x}_i^\omega$, conditioned on $\{\omega \in \Omega: i \leq \tilde{n}(\omega)\}$, possesses a Lebesgue density on G . Note that this is not the case for *deterministic algorithms* $Q_n = \phi \circ \iota$, with $\iota: \mathcal{W} \rightarrow \mathbb{R}^n$, $f \mapsto (f(\mathbf{x}_1), \dots, f(\mathbf{x}_n))$ for fixed nodes $\mathbf{x}_1, \dots, \mathbf{x}_n \in G$ and $\phi: \mathbb{R}^n \rightarrow \mathbb{R}$, which in the scenario $\mathcal{W} \not\hookrightarrow C(G)$ can be easily fooled if for each class $[f] \in \mathcal{W}$ we pick a representative that vanishes in the nodes $\mathbf{x}_1, \dots, \mathbf{x}_n$. In other words: deterministic information is useless there.

The constraining quantity for a complexity analysis is the expected cardinality

$$\overline{\text{card}}(A) := \mathbb{E} \text{card}(A)$$

of a method A . We are interested in bounds on the smallest $\varepsilon = e(A, \delta, \mathcal{W}) > 0$ if $\overline{\text{card}}(A) \leq \bar{n}$ for given $\bar{n} > 0$. In order to simplify the analysis, the main results in this section, see Lemma 1 and 2, are formulated and proven for non-adaptive algorithms A with *fixed cardinality* $\text{card}(A) \leq n$ for $n \in \mathbb{N}$. We indicate the dependence on $n \in \mathbb{N}$ by writing $A = A_n$ for such methods and repeat the definition of the n -th *minimal probabilistic Monte Carlo error at uncertainty* δ , given by

$$e_{\text{prob}}^{\text{MC}}(n, \delta, \mathcal{W}) := \inf_{A_n: \text{card}(A_n) \leq n} e(A_n, \delta, \mathcal{W}).$$

Restricting to fixed cardinality is no major shortcoming as can be seen by the following remark.

Remark 1 Define the minimal probabilistic error over general non-adaptive Monte Carlo algorithms, given the cardinality constraint $\bar{n} > 0$ and the uncertainty $\delta > 0$, as

$$\tilde{e}_{\text{prob}}^{\text{MC}}(\bar{n}, \delta, \mathcal{W}) := \inf_{A: \overline{\text{card}}(A) \leq \bar{n}} e(A, \delta, \mathcal{W}).$$

For a non-adaptive Monte Carlo algorithm A , with $n := \lfloor 2\overline{\text{card}}(A) \rfloor \in \mathbb{N}_0$, we have $\mathbb{P}\{\text{card}(A) \leq n\} \geq \frac{1}{2}$ by Markov's inequality. Since the cardinality of A does not depend on the input, we may construct a method A_n with cardinality less than n via rejection sampling. The idea is to consider realisations of independent copies $A^{(j)} \stackrel{\text{iid}}{\sim} A$ and check the number of required function evaluations until the corresponding cardinality function, say $\tilde{n}^{(j)}$, meets the condition $\tilde{n}^{(j)}(\omega) \leq n$. After that we start with computing function values of the input f at the nodes of the accepted realisation $A^{(j)}$ of A . (Such an approach will always be of interest if realisations with exceedingly large cardinality should be avoided.) This leads to

the fact that, if A is (ε, δ) -approximating in \mathcal{W} for some $\delta \in (0, 1/2)$, then A_n is at least $(\varepsilon, 2\delta)$ -approximating. In terms of the minimal probabilistic error we have

$$e_{\text{prob}}^{\text{MC}}(2n, 2\delta, \mathcal{W}) \leq \tilde{e}_{\text{prob}}^{\text{MC}}(n, \delta, \mathcal{W}) \leq e_{\text{prob}}^{\text{MC}}(n, \delta, \mathcal{W}). \quad (10)$$

If $e_{\text{prob}}^{\text{MC}}(n, \delta, \mathcal{W})$ behaves at most polynomially in n and δ^{-1} , then (10) implies asymptotic equivalence of the Monte Carlo error in the fixed cardinality setting and the varying cardinality setting. In Section 4, indeed, we consider (non-adaptive) algorithms with varying cardinality.

In the spirit of Bakhvalov [4], for proving lower bounds we switch to an *average input setting* with a discrete probability measure μ supported on functions f that belong to (equivalence classes $[f]$ within) the input set—which in our case is the unit ball $\mathcal{B}_{\mathcal{W}}$ of the space \mathcal{W} —and make use of the relation

$$\begin{aligned} \sup_{\|f\|_{\mathcal{W}} \leq 1} \mathbb{P}\{|A_n(f) - \text{INT } f| > \varepsilon\} &\geq \int_{\mathcal{B}_{\mathcal{W}}} \int_{\Omega} \mathbb{1}_{\{|A_n^\omega(f) - \text{INT } f| > \varepsilon\}} \, d\mathbb{P}(\omega) \, d\mu(f) \\ &= \int_{\Omega} \int_{\mathcal{B}_{\mathcal{W}}} \mathbb{1}_{\{|A_n^\omega(f) - \text{INT } f| > \varepsilon\}} \, d\mu(f) \, d\mathbb{P}(\omega) \\ &\geq \inf_{Q_n} \mu\{f : |Q_n(f) - \text{INT } f| > \varepsilon\}, \end{aligned} \quad (11)$$

where the infimum is taken over all *deterministic* integration methods Q_n that use n function values. (For fixed ω , the realisation A_n^ω of a given Monte Carlo algorithm can be regarded as a deterministic algorithm.) In the proof of the lower bounds we use the implication

$$\sup_{\|f\|_{\mathcal{W}} \leq 1} \mathbb{P}\{|A_n(f) - \text{INT } f| > \varepsilon\} > \delta \implies e(A_n, \delta, \mathcal{W}) \geq \varepsilon. \quad (12)$$

Depending on the integrability index p of the Sobolev classes we choose different probability measures μ in order to obtain appropriate lower bounds. Similarly to [20, Proposition 1 and 2 in Section 2.2.4] we have the following two generic lemmas, now for the probabilistic instead of the root mean squared error. The first one applies for integrability $2 \leq p \leq \infty$.

Lemma 1 *Let $n, N \in \mathbb{N}$ with $n \geq 17$, $N \geq 5n + 6$ and let $\gamma > 0$. Assume that there are functions $f_i : G \rightarrow \mathbb{R}$, for $i = 1, \dots, N$, satisfying the following conditions:*

1. *The sets $G_i := \{\mathbf{x} \in G : f_i(\mathbf{x}) \neq 0\}$, $i = 1, \dots, N$, are pairwise disjoint and $\text{INT } f_i = \gamma$ for all $i = 1, \dots, N$.*
2. *For all $\mathbf{s} = (s_i)_{i \in \{1, \dots, N\}} \in \{-1, 1\}^N$ the function $f_{\mathbf{s}} := \sum_{i=1}^N s_i f_i$ is an element of the input set $\mathcal{B}_{\mathcal{W}}$, that is, $\|f_{\mathbf{s}}\|_{\mathcal{W}} \leq 1$.*

Then, for all uncertainty levels $0 < \delta < 1/3$, we have

$$e_{\text{prob}}^{\text{MC}}(n, \delta, \mathcal{W}) \geq \gamma \min \left\{ n^{1/2} \sqrt{\log_4 \frac{1}{3\delta}}, n \right\}.$$

Proof Let μ be the uniform distribution on the finite set

$$\mathcal{F} := \left\{ f_{\mathbf{s}} = \sum_{i=1}^N s_i f_i : s_i \in \{-1, 1\} \right\} \subset \mathcal{B}_{\mathcal{W}},$$

and let $Q_n: \mathcal{F} \rightarrow \mathbb{R}$ be a deterministic algorithm using n function values. In the spirit of Bakhvalov (11) we aim to find lower bounds for the probability

$$\mu\{f: |Q_n(f) - \text{INT } f| > \varepsilon\} = \mathbb{P}\{|Q_n(f_{\mathbf{S}}) - \text{INT } f_{\mathbf{S}}| > \varepsilon\},$$

where $\mathbf{S} = (S_1, \dots, S_N)$ is a uniformly distributed random vector $\mathbf{S}: \Omega \rightarrow \{-1, 1\}^N$, and $(\Omega, \Sigma, \mathbb{P})$ is a suitable underlying probability space. (The shift to probabilities on random inputs instead of measuring sets with respect to μ is for notational convenience.) If $\mathbf{x} \in G_i$, then from computing $f_{\mathbf{S}}(\mathbf{x})$ we learn the corresponding sign $S_i = f_{\mathbf{S}}(\mathbf{x})/f_i(\mathbf{x})$. Without loss of generality, we may assume that the algorithm computes function values $y_i = f_{\mathbf{S}}(\mathbf{x}_i)$ with $\mathbf{x}_i \in G_i$ for $i = 1, \dots, n$. Hence, from such information we learn $\mathbf{S}_{[n]} := (S_1, \dots, S_n)$ for a given realisation of $f_{\mathbf{S}}$, while the remaining $k := N - n$ signs are still unknown. Thus, the conditional distribution of the random variable $\text{INT } f_{\mathbf{S}}$ given $\mathbf{S}_{[n]} = \mathbf{s}_{[n]} = (s_1, \dots, s_n) \in \{-1, 1\}^n$ coincides with the distribution of

$$\gamma \sum_{i=1}^n s_i + \gamma \sum_{i=n+1}^N S_i,$$

where S_{n+1}, \dots, S_N is an i.i.d. sequence of Rademacher random variables (that is, $\{-1, 1\}$ -uniformly distributed). The output $Q_n(f_{\mathbf{S}})$ may only depend on $\mathbf{s}_{[n]}$ and is thus constant under the conditional distribution. By X_k we denote the sum of k independent Rademacher distributed random variables, such that the distributions of X_k and $\sum_{i=n+1}^N S_i$ coincide. It follows that

$$\begin{aligned} \mu\{f: |Q_n(f) - \text{INT } f| > \varepsilon\} &\geq \inf_{\mathbf{s}_{[n]} \in \{-1, 1\}^n} \mathbb{P}(|Q_n(f_{\mathbf{S}}) - \text{INT } f_{\mathbf{S}}| > \varepsilon \mid \mathbf{S}_{[n]} = \mathbf{s}_{[n]}) \\ &\geq \inf_{a \in \mathbb{R}} \mathbb{P}\{\gamma |X_k - a| > \varepsilon\} \\ &= \inf_{a \in \mathbb{R}} 2^{-k} \sum_{j=0}^k \binom{k}{j} \mathbf{1}[\gamma |2j - k - a| > \varepsilon]. \end{aligned}$$

At most $k' := \lfloor \varepsilon/\gamma \rfloor + 1$ terms are removed from the binomial sum, optimally the central ones, so we have

$$\mu\{|Q_n(f) - \text{INT } f| > \varepsilon\} \geq 2^{-k} \left[\sum_{j=0}^{\lfloor \frac{k-k'}{2} \rfloor} \binom{k}{j} + \sum_{j=\lceil \frac{k+k'+1}{2} \rceil}^k \binom{k}{j} \right].$$

We employ Lemma 3 twice, namely, for odd k with

$$t = \lceil (k' - 1)/2 \rceil \quad \text{and} \quad t = \lceil k'/2 \rceil \leq \varepsilon/(2\gamma) + 1$$

as well as for even k with

$$t = \lceil k'/2 \rceil \quad \text{and} \quad t = \lceil (k' + 1)/2 \rceil \leq \varepsilon/(2\gamma) + 3/2.$$

In order to match the conditions of Lemma 3, we restrict to $\varepsilon/\gamma \leq (k - 6)/4$. Hence, under that assumption we have

$$\mu\{|Q_n(f) - \text{INT } f| > \varepsilon\} \geq \frac{1}{1 + 2/\sqrt{\pi}} \exp\left(-\frac{4(\log 2)(\varepsilon/\gamma + 2)^2}{k}\right). \quad (13)$$

Note that $k = N - n \geq (5n + 6) - n = 4n + 6$, so then the condition $\varepsilon/\gamma \leq n$ is sufficient for (13) to hold. The right-hand side of (13) can be further simplified via $(\varepsilon/\gamma + 2)^2 \leq 2(\varepsilon^2/\gamma^2 + 4)$, exploiting $k > 4n$ and also $n \geq 17$. For $0 < \varepsilon \leq \gamma n$ this leads to

$$\mu\{|Q_n(f) - \text{INT } f| > \varepsilon\} > \frac{2^{-8/n}}{1 + 2/\sqrt{\pi}} 2^{-2\varepsilon^2/(n\gamma^2)} > \frac{1}{3} 4^{-\varepsilon^2/(n\gamma^2)}. \quad (14)$$

By Bakhvalov's trick (11) this is a lower bound for the worst case uncertainty $\sup_{\|f\|_{\mathcal{W}} \leq 1} \mathbb{P}\{|A_n(f) - \text{INT } f| > \varepsilon\}$, holding for any Monte Carlo algorithm A_n . Regarding the right-hand side of (14) as a given δ , the implication (12) finally provides the assertion. Pay attention that for too small δ , namely $0 < \delta < \frac{1}{3} 4^{-n}$, isolating ε in (14) is misleading to delusive error bounds exceeding γn which, however, violates the conditions on ε . In this case we can only conclude that $\varepsilon = \gamma n$ is a lower bound for $e_{\text{prob}}^{\text{MC}}(n, \delta, \mathcal{W})$. \square

The following result will be useful for integrability $1 < p < 2$.

Lemma 2 *Let $n, N, M \in \mathbb{N}$ with $N \geq 4n$ and $M \leq N$, and let $\gamma > 0$. Assume that there are functions $f_i: G \rightarrow \mathbb{R}$, for $i = 1, \dots, N$, satisfying:*

1. *The sets $G_i := \{\mathbf{x} \in G: f_i(\mathbf{x}) \neq 0\}$, $i = 1, \dots, N$, are pairwise disjoint, and for all $i = 1, \dots, N$ we have $\text{INT } f_i = \gamma$.*
2. *For all $I \subset \{1, \dots, N\}$ with $\#I = M$ and for all $\mathbf{s} = (s_i)_{i \in I} \in \{-1, 1\}^M$, the function $f_{I, \mathbf{s}} := \sum_{i \in I} s_i f_i$ is an element of the input set $\mathcal{B}_{\mathcal{W}}$, that is, $\|f_{I, \mathbf{s}}\|_{\mathcal{W}} \leq 1$.*

Then, for all $0 < \delta < \frac{1}{2} 2^{-\lceil M/2 \rceil}$, we have

$$e_{\text{prob}}^{\text{MC}}(n, \delta, \mathcal{W}) \geq \frac{1}{2} \gamma M.$$

Proof Let μ be the uniform distribution on the finite set

$$\mathcal{F} := \left\{ f_{I, \mathbf{s}} = \sum_{i \in I} s_i f_i: I \subset \{1, \dots, N\} \text{ with } \#I = M, s_i \in \{-1, 1\} \right\} \subset \mathcal{B}_{\mathcal{W}},$$

and let $Q_n: \mathcal{F} \rightarrow \mathbb{R}$ be a deterministic algorithm using n function values. In the spirit of Bakhvalov (11) we aim to find lower bounds for the probability

$$\mu\{f: |Q_n(f) - \text{INT } f| > \varepsilon\} = \mathbb{P}\{|Q_n(f_{I, \mathbf{s}}) - \text{INT } f_{I, \mathbf{s}}| > \varepsilon\},$$

with a random set $I \subset \{1, \dots, N\}$ of cardinality $\#I = M$ chosen uniformly at random, and an independent uniformly distributed random vector $\mathbf{S} = (S_1, \dots, S_N)$ in $\{-1, 1\}^N$. If $\mathbf{x} \in G_i$, then from computing $f_{I, \mathbf{s}}(\mathbf{x})$ we learn whether $i \in I$ (which is equivalent to $f_{I, \mathbf{s}}(\mathbf{x}) \neq 0$), and if so, we also learn the corresponding sign $S_i = f_{\mathbf{S}}(\mathbf{x})/f_i(\mathbf{x})$. Without loss of generality, we may assume that the algorithm computes function values $y_i = f_{I, \mathbf{s}}(\mathbf{x}_i)$ with $\mathbf{x}_i \in G_i$ for $i = 1, \dots, n$. Let $m(I) := \#\{i \in I: y_i \neq 0\}$ denote the number of detected subdomains G_i where the function $f_{I, \mathbf{s}}$ is non-zero. The random variable $m(I)$ is distributed according to a hypergeometric distribution with population of size N containing $M \leq N$ items of interest and admitting $n < N$ draws without replacement. The expected value is

$$\mathbb{E} m(I) = \frac{n}{N} M \leq \frac{1}{4} M,$$

and using Markov's inequality we conclude

$$\mathbb{P}\{m(I) \leq \frac{1}{2}M\} \geq \frac{1}{2}. \quad (15)$$

Given the information $(f_{I,\mathbf{S}}(\mathbf{x}_1), \dots, f_{I,\mathbf{S}}(\mathbf{x}_n)) = \mathbf{y}$ with $m(I) = m$, there are still $k := M - m$ unknown signs S_i on subdomains G_i where the function does not vanish, namely for $i \in \{n+1, \dots, N\} \cap I$. Therefore, the conditional distributions of INT $f_{I,\mathbf{S}}$ given $I \cap \{1, \dots, n\} = J$ with $\#J = m$ and $(S_i)_{i \in J} = (s_i)_{i \in J} \in \{-1, 1\}^m$ coincides with the distribution of

$$\gamma \sum_{i \in J} s_i + \gamma \sum_{i \in I \setminus J} S_i.$$

By X_k we denote the sum of k independent Rademacher distributed random variables, such that the distributions of X_k and $\sum_{i \in I \setminus J} S_i$ coincide. Similarly to the proof of Lemma 1, the uncertainty, now conditioned $m(I) = m$, can be bounded by a binomial sum (recall $k = M - m$),

$$\mathbb{P}(|Q_n(f_{I,\mathbf{S}}) - \text{INT } f_{I,\mathbf{S}}| > \varepsilon \mid m(I) = m) \geq \inf_{a \in \mathbb{R}} 2^{-k} \sum_{j=0}^k \binom{k}{j} \mathbb{1}[\gamma|2j - k - a| > \varepsilon].$$

For $0 < \varepsilon < \frac{1}{2}M\gamma$, up to $k' := \lceil \varepsilon/\gamma \rceil + 1 \leq \lceil \frac{1}{2}M \rceil$ terms are removed, such that

$$\mathbb{P}(|Q_n(f_{I,\mathbf{S}}) - \text{INT } f_{I,\mathbf{S}}| > \varepsilon \mid m(I) = m) \geq 2^{-k} \left[\sum_{j=0}^{\lfloor \frac{k-k'}{2} \rfloor} \binom{k}{j} + \sum_{j=\lceil \frac{k+k'+1}{2} \rceil}^k \binom{k}{j} \right].$$

Note that $m \leq \frac{1}{2}M$ implies $k = M - m \geq \lceil \frac{1}{2}M \rceil \geq k'$, so Lemma 4 can be used under the condition formulated in (15) to bound this from below by $2^{-k'}$. Hence,

$$\begin{aligned} \mu\{|Q_n(f) - \text{INT } f| > \varepsilon\} &\geq \sum_{m=0}^{\lceil M/2 \rceil} \mathbb{P}(|Q_n(f_{I,\mathbf{S}}) - \text{INT } f_{I,\mathbf{S}}| > \varepsilon \mid m(I) = m) \cdot \mathbb{P}\{m(I) = m\} \\ &\stackrel{\text{Lemma 4}}{\geq} 2^{-k'} \cdot \mathbb{P}\{m(I) \leq \frac{1}{2}M\} \stackrel{(15)}{\geq} \frac{1}{2} 2^{-\lceil M/2 \rceil}. \end{aligned}$$

By Bakhvalov's trick (11) this is a lower bound for the worst case uncertainty $\sup_{\|f\|_{\mathcal{W}} \leq 1} \mathbb{P}\{|A_n(f) - \text{INT } f| > \varepsilon\}$, and for $0 < \delta < \frac{1}{2} 2^{-\lceil M/2 \rceil}$ the implication (12), taking the supremum over $0 < \varepsilon < \frac{1}{2}M\gamma$, proves the assertion. \square

2.2 Lower bounds for Sobolev classes

The norms of classical (integer smoothness) Sobolev spaces $W_p^r(G)$ with *integrability index* p possess the property that for any decomposition of the support of

a function $f \in W_p^r(G)$ into subdomains G_1, \dots, G_M that are pairwise essentially disjoint, that is $\lambda^d(G_i \cap G_j) = 0$ for $i \neq j$, we have

$$\|f\|_{W_p^r(G)} = \left(\sum_{i=1}^M \|f\|_{W_p^r(G_i)}^p \right)^{1/p} \leq M^{1/p} \max_{i=1, \dots, M} \|f\|_{W_p^r(G_i)}, \quad (16)$$

with the usual modification for $p = \infty$. The *smoothness* has an effect on scalings, namely for functions $\varphi: \mathbb{R}^d \rightarrow \mathbb{R}$ and $\psi(\mathbf{x}) := \varphi(m\mathbf{x} - \mathbf{i})$, with $m > 0$ and $\mathbf{i} \in \mathbb{R}^d$, we have the following well known relation between the derivatives,

$$\|D^\alpha \psi\|_{L_p(\mathbb{R}^d)} = m^{|\alpha|_1 - d/p} \|D^\alpha \varphi\|_{L_p(\mathbb{R}^d)}, \quad \text{for } \alpha \in \mathbb{N}_0^d.$$

For $m \geq 1$ this leads to the scaling property

$$\|\psi\|_{W_p^r(\mathbb{R}^d)} \leq m^{r-d/p} \|\varphi\|_{W_p^r(\mathbb{R}^d)}. \quad (17)$$

If $\text{supp } \varphi, \text{supp } \psi \subseteq G$, then this relation holds also for the norms of the restricted space $W_p^r(G)$.

Theorem 1 *Let $G \subset \mathbb{R}^d$ be a measurable domain with non-empty interior. Further, let $r \in \mathbb{N}_0$, $1 \leq p \leq \infty$ and define $q := \min\{p, 2\}$. Then we have the asymptotic lower bound*

$$e_{\text{prob}}^{\text{MC}}(n, \delta, W_p^r(G)) \succeq \min \left\{ n^{-r/d}, n^{-(r/d+1-1/q)} (\log \delta^{-1})^{1-1/q} \right\}.$$

Proof Since the interior of G is non-empty, there exists a cubic subdomain. We restrict to functions that are supported within that rectangular subdomain, and by scaling, without loss of generality, we may assume $G = [0, 1]^d$.

Let $\varphi: \mathbb{R}^d \rightarrow \mathbb{R}$ be a sufficiently smooth function supported on $[0, 1]^d$ with $\|\varphi\|_{W_p^r([0, 1]^d)} \leq 1$ and $\gamma_0 := \text{INT } \varphi > 0$. We call φ *bump function*. For $m \in \mathbb{N}$ we split $[0, 1]^d$ into $N = m^d$ subcubes and equip each subcube with a scaled, shifted bump function $\psi_{\mathbf{i}}(\mathbf{x}) := \varphi(m\mathbf{x} - \mathbf{i})$, indexed by $\mathbf{i} \in [m]^d := \{0, \dots, m-1\}^d$.

If $2 \leq p \leq \infty$, we choose $m := \lceil (5n+6)^{1/d} \rceil$ and $f_{\mathbf{i}} := m^{-r} \psi_{\mathbf{i}}$. Hence, by (16) and (17) with $M = N$ we have

$$\left\| \sum_{\mathbf{i} \in [m]^d} s_{\mathbf{i}} f_{\mathbf{i}} \right\|_{W_p^r([0, 1]^d)} \leq 1 \quad \text{for arbitrary } s_{\mathbf{i}} \in \{\pm 1\}. \quad (18)$$

Then $\gamma = \text{INT } f_{\mathbf{i}} = m^{-r-d} \gamma_0$. Restricting to $n \geq 17$ and $0 < \delta \leq 1/3$, we can apply Lemma 1 and obtain

$$\begin{aligned} e_{\text{prob}}^{\text{MC}}(n, \delta, W_p^r([0, 1]^d)) &\geq \gamma_0 m^{-r-d} \min \left\{ n^{1/2} \sqrt{\log_4 \frac{1}{3\delta}}, n \right\} \\ &\succeq \min \{ n^{-r/d-1/2} \sqrt{\log \delta^{-1}}, n^{-r/d} \}. \end{aligned}$$

If $1 \leq p < 2$, we restrict to $0 \leq \delta < 1/4$ and choose $m := \lceil (4n)^{1/d} \rceil$. In the case $2^{-2n-1} \leq \delta$, we take $M = 2 \lceil \log_2(4\delta)^{-1} \rceil \leq 2 \log_2(2\delta)^{-1}$, and easily see that

$M \leq 4n \leq N$ is fulfilled. Here, put $f_{\mathbf{i}} := m^{-r} (N/M)^{1/p} \psi_{\mathbf{i}}$, and note that by (16) and (17) we have

$$\left\| \sum_{\mathbf{i} \in I} s_{\mathbf{i}} f_{\mathbf{i}} \right\|_{W_p^r([0,1]^d)} \leq 1, \quad \text{for arbitrary } s_{\mathbf{i}} \in \{\pm 1\} \text{ and } I \subseteq [m]^d \text{ with } \#I = M.$$

We have $\gamma = \text{INT } f_{\mathbf{i}} = \gamma_0 m^{-r-d+d/p} M^{-1/p}$, and Lemma 2 implies

$$e_{\text{prob}}^{\text{MC}}(n, \delta, W_p^r([0,1]^d)) \geq \frac{1}{2} \gamma_0 m^{-r-d+d/p} M^{1-1/p} \succeq n^{-(r/d+1-1/p)} (\log \delta^{-1})^{1-1/p}.$$

For small $\delta \in (0, 2^{-2n-1})$, however, we may just choose $M = 4n$, and similarly we obtain

$$e_{\text{prob}}^{\text{MC}}(n, \delta, W_p^r([0,1]^d)) \succeq n^{-r/d},$$

which finishes the proof. \square

Remark 2 (Lower bounds for non-integer smoothness) There are several approaches to generalize Sobolev spaces for non-integer smoothness $r > 0$. For example the *Slobodeckii space* $W_p^r(G)$ is given as the set of functions with finite norm

$$\|f\|_{W_p^r(G)} := \left(\|f\|_{W_p^{\lfloor r \rfloor}(G)}^p + \sum_{|\alpha|_1 = \lfloor r \rfloor} \int_G \int_G \frac{|D^\alpha f(\mathbf{x}) - D^\alpha f(\mathbf{z})|^p}{|\mathbf{x} - \mathbf{z}|^{d+(r-\lfloor r \rfloor)p}} d\mathbf{x} d\mathbf{z} \right)^{1/p},$$

where $1 \leq p < \infty$, see for instance the book of Triebel [28, p. 36]. For such spaces it is not clear whether inequality (16) still holds. However, the strategy of proof of the previous theorem works in this setting as well but is more technical. Namely, by introducing an additional constant in order to take the non-local nature of fractional derivatives into account, one can also construct fooling functions composed of bumps on disjoint subcubes with random sign.

Let us consider an easier example: Namely, classes of Hölder continuous functions with fractional smoothness $0 < \beta \leq 1$ (and integrability parameter $p = \infty$) given by

$$C^\beta([0,1]^d) := \left\{ f : [0,1]^d \rightarrow \mathbb{R} \mid |f|_{C^\beta} := \sup_{\mathbf{x}, \mathbf{z} \in [0,1]^d, \mathbf{x} \neq \mathbf{z}} \frac{|f(\mathbf{x}) - f(\mathbf{z})|}{|\mathbf{x} - \mathbf{z}|_\infty^\beta} < \infty \right\}. \quad (19)$$

With the choice $f_{\mathbf{i}} := \frac{1}{2} m^{-\beta} \psi_{\mathbf{i}}$, within the proof above, one can ensure (18), but with the norm $\|f\|_{C^\beta} := \|f\|_\infty + |f|_{C^\beta}$ instead. Thus, loosing just a factor 1/2, we still have the same weak asymptotic lower bound as should be expected from generalizing the integer smoothness case, namely

$$e_{\text{prob}}^{\text{MC}}(n, \delta, C^\beta([0,1]^d)) \succeq n^{-\beta/d} \min \left\{ 1, \sqrt{\frac{\log \delta^{-1}}{n}} \right\}. \quad (20)$$

This fits very well to the upper bounds of Theorem 5.

3 Upper Bounds

3.1 Probability amplification

One of the most elementary methods of ‘probability amplification’ is the so-called ‘median trick’, see Alon et al. [2] and Jerrum et al. [14]. The following proposition is a minor modification of [19, Proposition 2.1, in particular (2.6)] from Niemi and Pokarowski, now adapted to the language of algorithms and IBC.

As in Section 2.1, we consider a general function space \mathcal{W} equipped with a norm $\|\cdot\|_{\mathcal{W}}$ and take its unit ball $\mathcal{B}_{\mathcal{W}}$ as input set.

Proposition 1 (Median trick) *Let $\varepsilon > 0$, $\alpha \in (0, 1/2)$ and let A_m be an arbitrary Monte Carlo algorithm satisfying*

$$\sup_{\|f\|_{\mathcal{W}} \leq 1} \mathbb{P}\{|A_m(f) - \text{INT } f| > \varepsilon\} \leq \alpha.$$

For an odd $k \in \mathbb{N}$, define

$$A_{m,k}(f) := \text{med} \left\{ A_m^{(1)}(f), \dots, A_m^{(k)}(f) \right\}$$

as the empirical median of k independent realisations of $A_m(f)$. Then

$$\sup_{\|f\|_{\mathcal{W}} \leq 1} \mathbb{P}\{|A_{m,k}(f) - \text{INT } f| > \varepsilon\} \leq \frac{1}{2} (4\alpha(1-\alpha))^{k/2} < 2^{k-1} \alpha^{k/2}.$$

The previous proposition can be used to derive upper bounds for the probabilistic (ε, δ) -complexity $n_{\text{prob}}^{\text{MC}}(\varepsilon, \delta, \mathcal{W})$ in terms of the ℓ -mean error complexity $n_{\ell\text{-mean}}^{\text{MC}}(\varepsilon, \mathcal{W})$, compare (5).

Theorem 2 *Let $\varepsilon > 0$, $\delta \in (0, 1/2)$ and let $\ell \geq 1$. Then for the (ε, δ) -complexity holds*

$$n_{\text{prob}}^{\text{MC}}(\varepsilon, \delta, \mathcal{W}) \leq 2 \log_2 \delta^{-1} \cdot n_{\ell\text{-mean}}^{\text{MC}}\left(8^{-1/\ell} \varepsilon, \mathcal{W}\right).$$

In particular, if $e_{\ell\text{-mean}}^{\text{MC}}(n, \mathcal{W}) \leq n^{-\varrho}$ for some $\varrho > 0$, then we have

$$e_{\text{prob}}^{\text{MC}}(n, \delta, \mathcal{W}) \leq \left(\frac{\log \delta^{-1}}{n}\right)^{\varrho} \quad \text{for } n \geq \log \delta^{-1}.$$

Proof Without loss of generality, we assume that $n_{\ell\text{-mean}}^{\text{MC}}(8^{-1/\ell} \varepsilon, \mathcal{W}) < \infty$, otherwise the claimed inequality is trivial. Hence, there is an $m \in \mathbb{N}$ such that $e_{\ell\text{-mean}}^{\text{MC}}(m, \mathcal{W}) < 8^{-1/\ell} \varepsilon$. This implies that there is a Monte Carlo algorithm A_m such that

$$e_{\ell\text{-mean}}^{\text{MC}}(A_m, \mathcal{W}) := \sup_{\|f\|_{\mathcal{W}} \leq 1} \left(\mathbb{E} \|A_m(f) - \text{INT } f\|^{\ell}\right)^{1/\ell} \leq 8^{-1/\ell} \varepsilon.$$

Thus, for any $f \in \mathcal{B}_{\mathcal{W}}$, by Markov’s inequality we have

$$\mathbb{P}\{|A_m(f) - \text{INT } f| > \varepsilon\} \leq \left(\frac{e_{\ell\text{-mean}}^{\text{MC}}(A_m, \mathcal{W})}{\varepsilon}\right)^{\ell} \leq \frac{1}{8}.$$

Now we apply Proposition 1 with $\alpha = 1/8$, and k chosen as the smallest odd natural number that satisfies $k \geq 2 \log_2(2\delta)^{-1}$. Note that $k \leq 2 \log_2 \delta^{-1}$ for $0 < \delta \leq 1/2$. Then we obtain the desired complexity bound

$$n_{\text{prob}}^{\text{MC}}(\varepsilon, \delta, \mathcal{W}) \leq k \cdot n_{\ell\text{-mean}}^{\text{MC}}(8^{-1/\ell} \varepsilon, \mathcal{W}) \leq 2 \log_2 \delta^{-1} \cdot n_{\ell\text{-mean}}^{\text{MC}}(8^{-1/\ell} \varepsilon, \mathcal{W}).$$

In terms of the error quantities, for fixed m and odd $k \geq 2 \log_2(2\delta)^{-1}$ we can state

$$e_{\text{prob}}^{\text{MC}}(km, \delta, \mathcal{W}) \leq 8^{1/\ell} e_{\ell\text{-mean}}^{\text{MC}}(m, \mathcal{W}).$$

Assuming $e_{\ell\text{-mean}}^{\text{MC}}(m, \mathcal{W}) \leq C m^{-\varrho}$ for $m \geq m_0 \in \mathbb{N}$ with a suitable constant C , and given a total amount of n function values to use with $n \geq 2m_0 \log_2 \delta^{-1}$, put $m := \lfloor n/(2 \log_2 \delta^{-1}) \rfloor$. We conclude

$$\begin{aligned} e_{\text{prob}}^{\text{MC}}(n, \delta, \mathcal{W}) &\leq 8^{1/\ell} \cdot C \cdot \left[\frac{n}{2 \log_2 \delta^{-1}} \right]^{-\varrho} \\ &\leq 8^{1/\ell} \cdot C \cdot \left(\frac{2}{\log 2} \left(1 + \frac{1}{m_0} \right) \right)^{\varrho} \cdot \left(\frac{\log \delta^{-1}}{n} \right)^{\varrho}. \quad \square \end{aligned}$$

Remark 3 (Integrating L_p -functions) The lower bounds of Theorem 1 match the upper bounds from Theorem 2, iff the rate of convergence is related to the integrability index by $\varrho = 1 - 1/q$ where $q := \min\{p, 2\}$. This is only the case for smoothness $r = 0$, i.e., when L_p -balls are the considered input sets. In that case,

$$e_{\text{prob}}^{\text{MC}}(n, \delta, L_p) \asymp \left(\frac{\log \delta^{-1}}{n} \right)^{1-1/q}, \quad \text{for } n \succeq \log \delta^{-1}. \quad (21)$$

Here we used estimates of the q -mean error for the standard i.i.d.-based Monte Carlo method applied to L_p -functions which, for example, can be found in [3, Theorem 2], [9, Proposition 5.4], [20, Sect. 2.2.8, Proposition 3], as well as [26, Proof of Theorem 1].

Remark 4 (Alternatives to the median trick) Catoni [5] proposes an alternative scheme based on random samples of L_p -functions, that suppresses outliers and (in contrast to the median trick) is symmetric in the sense that permuting the sample data does not change the result. Compare also Huber [13] where this approach is combined with the median trick.

3.2 Separation of the main part

Separation of the main part, also known as *control variates*, is a well established technique of variance reduction which uses the approximation of functions with respect to an L_q -norm in order to exploit the smoothness of the given input set.

Within this section we assume that $G \subset \mathbb{R}^d$ is a bounded Lipschitz domain, see [21] for details. Let \mathcal{W} be a normed linear space consisting of continuous functions on G with corresponding unit ball $\mathcal{B}_{\mathcal{W}}$. Moreover, \mathcal{W} shall be continuously embedded in the space of continuous functions, $\mathcal{W} \hookrightarrow C(G)$, hence, for any fixed $x \in G$, the function evaluation operator, $\text{ev}_x: \mathcal{W} \rightarrow \mathbb{R}$, $f \mapsto f(x)$, is a continuous

functional. For $q \geq 1$ let $L_q(G)$ be the Lebesgue space equipped with the norm $\|\cdot\|_{L_q(G)}$. Within the approximation step we only consider linear methods

$$A_n : \mathcal{W} \rightarrow L_q(G), \quad f \mapsto g := \sum_{i=1}^n f(\mathbf{x}_i) g_i, \quad (22)$$

with nodes $\mathbf{x}_i \in G$ and functions $g_i \in L_q(G)$. The minimal L_q -approximation error of such methods is denoted by

$$e^{\det}(n, \mathcal{W} \hookrightarrow L_q) := \inf_{A_n} \sup_{\|f\|_{\mathcal{W}} \leq 1} \|A_n(f) - f\|_{L_q(G)}, \quad (23)$$

and the ε -complexity $n^{\det}(\varepsilon, \mathcal{W} \hookrightarrow L_q)$ is the minimal number of function evaluations needed in order to achieve an L_q -approximation error smaller than ε . The idea is to apply a Monte Carlo integration method $M_n : L_q \rightarrow \mathbb{R}$ to the difference $f - g$ between approximating and original function, while the integral of $g = A_n(f)$ is considered to be known. (Typically, the functions g_i are piecewise polynomials or other simple functions, thus INT g_i can be determined exactly.) This approach leads to the following theorem.

Theorem 3 (Separation of the main part) *For any $n \in \mathbb{N}$ we have*

$$e_{\text{prob}}^{\text{MC}}(2n, \delta, \mathcal{W}) \leq e^{\det}(n, \mathcal{W} \hookrightarrow L_q) \cdot e_{\text{prob}}^{\text{MC}}(n, \delta, L_q).$$

Proof Let A_n be a linear approximation method, see (22), such that for some $\alpha > 0$ it is guaranteed for all $f \in \mathcal{B}_{\mathcal{W}}$ that

$$\|A_n(f) - f\|_{L_q(G)} \leq \alpha.$$

Further, let $M_n = (M_n^\omega)_{\omega \in \Omega}$ be a Monte Carlo method that approximates INT h for inputs $h \in L_q$ by using certain random nodes. With this we define a new Monte Carlo method $Q_{2n} = (Q_{2n}^\omega)_{\omega \in \Omega}$ (a randomized quadrature rule) as follows:

1. Compute the approximation $g := A_n(f) \in L_q(G)$, using function values $f(\mathbf{x}_i)$ at nodes $\mathbf{x}_1, \dots, \mathbf{x}_n \in G$.
2. Return

$$Q_{2n}^\omega(f) := \text{INT } g + \alpha M_n^\omega \left(\alpha^{-1} \cdot (f - g) \right)$$

where M_n evaluates $\alpha^{-1} \cdot (f - g)$ at random nodes $\mathbf{X}_{n+1}^\omega, \dots, \mathbf{X}_{2n}^\omega \in G$. (At this point there are no specific assumptions about the distribution of the random nodes and the way they are used within M_n . Further, if M_n is homogeneous, that is, $M_n^\omega(\lambda f) = \lambda M_n^\omega(f)$ for any $\lambda \in \mathbb{R}$, then α cancels out.)

Indeed, the information $\mathbf{y} = (f(\mathbf{x}_1), \dots, f(\mathbf{x}_n), f(\mathbf{X}_{n+1}^\omega), \dots, f(\mathbf{X}_{2n}^\omega))$ suffices to execute the algorithm, namely,

$$\text{INT } g = \sum_{i=1}^n f(\mathbf{x}_i) \text{INT } g_i, \quad \text{and}$$

$$[\alpha^{-1} \cdot (f - g)](\mathbf{X}_{n+j}^\omega) = \alpha^{-1} \cdot \left(f(\mathbf{X}_{n+j}^\omega) - \sum_{i=1}^n f(\mathbf{x}_i) g_i(\mathbf{X}_{n+j}^\omega) \right) \quad \text{for } j = 1, \dots, n.$$

By writing $\varepsilon = \alpha\varepsilon'$, the uncertainty of the algorithm can be traced back to the uncertainty of M_n . Namely, if M_n is (ε', δ) -approximating in $L_q(G)$, then

$$\begin{aligned} \mathbb{P}\{|Q_{2n}(f) - \text{INT } f| > \varepsilon\} &= \mathbb{P}\left\{|M_n(\alpha^{-1} \cdot (f - g)) - \text{INT}(\alpha^{-1} \cdot (f - g))| > \varepsilon'\right\} \\ &\leq \delta. \end{aligned}$$

With (almost) optimal methods A_n and M_n^ω , α can get arbitrarily close to the minimal approximation error $e^{\text{det}}(n, \mathcal{W} \hookrightarrow L_q)$, and ε' arbitrarily close to $e_{\text{prob}}^{\text{MC}}(n, \delta, L_q)$, while keeping the uncertainty bounded by δ , thus ε approaches the stated error bound. \square

As long as function evaluations are continuous, it suffices to work with deterministic approximation methods of the form (22). For isotropic Sobolev spaces $W_p^r(G)$ on bounded Lipschitz domains $G \subset \mathbb{R}^d$, this is the case iff $rp > d$. In this setting it is well known that, with $q := \min\{p, 2\}$,

$$e^{\text{det}}(n, W_p^r(G) \hookrightarrow L_q(G)) \asymp n^{-r/d}, \quad \text{if } rp > d. \quad (24)$$

For $G = [0, 1]^d$, this result can be achieved with piecewise polynomial interpolation, see for instance Heinrich [9, Proposition 5.1], technical details of approximation methods are contained in Ciarlet [6]. For the general case of bounded Lipschitz domains $G \subset \mathbb{R}^d$, see Novak and Triebel [21, Theorem 23]. From this we conclude optimal upper bounds.

Theorem 4 *Let $G \subset \mathbb{R}^d$ be a bounded Lipschitz domain. Let $r \in \mathbb{N}$ and $1 \leq p \leq \infty$ with $rp > d$ and write $q := \min\{2, p\}$. Then we have the asymptotic rate*

$$e_{\text{prob}}^{\text{MC}}(n, \delta, W_p^r(G)) \asymp n^{-r/d} \min\left\{1, \left(\frac{\log \delta^{-1}}{n}\right)^{1-1/q}\right\}.$$

Proof The lower bounds follow from Theorem 1.

For $n \geq \log \delta^{-1}$, we combine (21) with (24) via Theorem 3 and obtain

$$e_{\text{prob}}^{\text{MC}}(n, \delta, W_p^r(G)) \leq n^{-r/d} \left(\frac{\log \delta^{-1}}{n}\right)^{1-1/q}.$$

For $n < \log \delta^{-1}$ we rely on deterministic quadrature. This problem is easier than approximation in the sense that if $g \in L_q(G)$ is an approximation of f , then $|\text{INT } f - \text{INT } g| \leq (\lambda^d(G))^{1-1/q} \cdot \|f - g\|_{L_q(G)}$. Hence,

$$\begin{aligned} e^{\text{det}}(n, W_p^r(G)) &:= \inf_{A_n} \sup_{\|f\|_{\mathcal{W}} \leq 1} |A_n(f) - \text{INT } f| \\ &\leq e^{\text{det}}(n, W_p^r(G) \hookrightarrow L_q(G)) \asymp n^{-r/d}. \end{aligned}$$

See also Novak [20, 1.3.12] for a direct derivation on $G = [0, 1]^d$. \square

Remark 5 (Lower smoothness) The condition $rp > d$ is necessary to guarantee that the evaluation of functions on $W_p^r(G)$ for $1 < p \leq \infty$ is well-defined. (For $p = 1$, the condition $r = d$ is also sufficient, but then deterministic methods already provide the optimal error rates.) In the cases $1 < p < \infty$ with $rp \leq d$ one can still

use separation of the main part, but with a randomized approximation scheme, see Heinrich [10] for the case $G = [0, 1]^d$. That way, for any $1 \leq p \leq \infty$ and general $r \in \mathbb{N}$ we have

$$e_{1\text{-mean}}^{\text{MC}}(n, W_p^r([0, 1]^d)) \asymp n^{-(r/d+1-1/q)},$$

with $q := \min\{p, 2\}$. Probability amplification, see Theorem 2, yields

$$e_{\text{prob}}^{\text{MC}}(n, \delta, W_p^r([0, 1]^d)) \leq \left(\frac{\log \delta^{-1}}{n} \right)^{r/d+1-1/q}, \quad \text{for } n \geq \log \delta^{-1}.$$

The power of $\log \delta^{-1}$ in this upper bound exceeds the power of the lower bound by r/d which gets close to 1 for $p \rightarrow 1$. We conjecture that the influence of δ is smaller at least if one is close to the regime where functions are continuous.

Instead of deterministic algorithms of the form (22) one might also consider general randomized methods for the approximation of functions in \mathcal{W} . For those let $e_{\text{prob}}^{\text{MC}}(n, \delta, \mathcal{W} \hookrightarrow L_q)$ be the smallest $\varepsilon > 0$, such that there exists a general randomized approximation algorithm satisfying

$$\mathbb{P}\{\|A_n(f) - f\|_{L_q} > \varepsilon\} \leq \delta \quad \text{for all } f \text{ with } \|f\|_{\mathcal{W}} \leq 1.$$

Similarly to Theorem 3 one can show that

$$e_{\text{prob}}^{\text{MC}}(2n, \delta, \mathcal{W}) \leq e_{\text{prob}}^{\text{MC}}(n, \delta/2, \mathcal{W} \hookrightarrow L_q) \cdot e_{\text{prob}}^{\text{MC}}(n, \delta/2, L_q). \quad (25)$$

Such an approach, however, seems to rely on complicated algorithms, since non-linearity might be inevitable in order to suppress outliers. There might be easier implementable Monte Carlo methods for integration that achieve a better order of convergence without relying on the approximation of functions. Such methods are needed in spaces of mixed smoothness, see the discussion in Section 4.

Anyway, studying Sobolev embeddings $W_p^r(G) \hookrightarrow L_q(G)$ in terms of approximation with high confidence within the regime $d(q-p) < rqp \leq dq$ for general integrability parameters $1 \leq p, q < \infty$, is an interesting problem on its own, compare Heinrich [10].

3.3 Stratified sampling

Let us introduce stratified sampling for the approximation of INT f for integrable functions f defined on $G = [0, 1]^d$. For $m \in \mathbb{N}$ we split the unit cube $[0, 1]^d$ into m^d subcubes given by

$$G_{\mathbf{i}} = \prod_{j=1}^d \left[\frac{i_j}{m}, \frac{i_j + 1}{m} \right),$$

with $\mathbf{i} \in [m]^d := \{0, \dots, m-1\}^d$ and $\mathbf{i} = (i_1, \dots, i_d)$. Let $(\mathbf{X}_{\mathbf{i}})_{\mathbf{i} \in [m]^d}$ be a sequence of independent random variables with $\mathbf{X}_{\mathbf{i}}$ uniformly distributed in $G_{\mathbf{i}}$. Then, stratified sampling is given by

$$S_m^d(f) := \frac{1}{m^d} \sum_{\mathbf{i} \in [m]^d} f(\mathbf{X}_{\mathbf{i}}), \quad (26)$$

which uses m^d function evaluations of f . Compared to the separation of the main part, stratified sampling is easier to implement. We show that in some cases it provides optimal results in terms of the (ε, δ) -complexity. Since the algorithm of stratified sampling does not depend on δ (compare the median trick), we obtain a universal method in terms of the uncertainty.

We use Hoeffding's inequality which is, for the convenience of the reader, stated in the following proposition.

Proposition 2 (Hoeffding's inequality) *Let $n \in \mathbb{N}$ and let Y_1, \dots, Y_n be independent real-valued random variables such that each of it is supported on an interval of finite length $b_i > 0$, that is, $\text{ess sup } Y_i - \text{ess inf } Y_i \leq b_i$ for all $i = 1, \dots, n$. Then, for $S_n := \frac{1}{n} \sum_{i=1}^n Y_i$ and all $\varepsilon > 0$ we have*

$$\mathbb{P}\{|S_n - \mathbb{E} S_n| > \varepsilon\} \leq 2 \exp\left(-\frac{2n^2 \varepsilon^2}{\sum_{i=1}^n b_i^2}\right).$$

First, we consider Hölder classes $C^\beta([0, 1]^d)$ with smoothness $\beta \in (0, 1]$, see (19). Compare also [4] for the result in terms of the root mean squared error.

Theorem 5 *For the classes of Hölder-continuous functions on $[0, 1]^d$, stratified sampling achieves the optimal rate of convergence, such that*

$$e_{\text{prob}}^{\text{MC}}\left(n, \delta, C^\beta([0, 1]^d)\right) \asymp n^{-\beta/d} \min\left\{1, \sqrt{\frac{\log \delta^{-1}}{n}}\right\}.$$

Proof Concerning the lower bounds, see (20) in Remark 2.

For the upper bounds, we start with the case $n = m^d$ with $m \in \mathbb{N}$ and employ S_m^d . Obviously, this method is unbiased, i.e., $\mathbb{E} S_m^d(f) = \text{INT } f$. By Hölder continuity, the random variables $Y_{\mathbf{i}} = f(\mathbf{X}_{\mathbf{i}})$ are each supported on intervals of length

$$b_{\mathbf{i}} := \sup_{G_{\mathbf{i}}} f - \inf_{G_{\mathbf{i}}} f \leq m^{-\beta}$$

for $|f|_{C^\beta} \leq 1$. This implies the worst case bound $|S_m^d(f) - \text{INT } f| \leq m^{-\beta} = n^{-\beta/d}$. Further, Hoeffding's inequality (Proposition 2) leads to

$$\mathbb{P}\{|S_m^d(f) - \text{INT } f| > \varepsilon\} \leq 2 \exp\left(-\frac{2m^{2d} \varepsilon^2}{m^d \cdot m^{-2\beta}}\right) = 2 \exp\left(-2m^{d+2\beta} \varepsilon^2\right).$$

This is guaranteed to be at most δ for

$$\varepsilon = \frac{1}{\sqrt{2}} m^{-(\beta+d/2)} \sqrt{\log \frac{2}{\delta}} = \frac{1}{\sqrt{2}} n^{-(\beta/d+1/2)} \sqrt{\log \frac{2}{\delta}}.$$

Given a total number $n \in \mathbb{N}$ of function values to use, we choose $m := \lfloor n^{1/d} \rfloor$. Employing the method S_m^d , see (26), we actually only use $m^d \leq n$ function values. For $n \geq 2^d$ we have $m \geq \frac{1}{2} n^{1/d}$, hence,

$$e_{\text{prob}}^{\text{MC}}\left(n, \delta, C^\beta([0, 1]^d)\right) \leq \min\left\{2^\beta \cdot n^{-\beta/d}, 2^{\beta+(d-1)/2} \cdot n^{-(\beta/d+1/2)} \sqrt{\log \frac{2}{\delta}}\right\}. \quad \square$$

Now we consider the isotropic Sobolev classes $W_p^1([0, 1]^d)$ of smoothness 1. For Hoeffding's inequality to be applicable, we need $W_p^1([0, 1]^d) \hookrightarrow L_\infty([0, 1]^d)$ which is the case for $p > d$. For $p \geq 2$, the asymptotic upper error bound of stratified sampling derived in the proof of the next theorem perfectly matches the lower bounds from Theorem 1.

Theorem 6 *Let $p > d$ and write $q := \min\{p, 2\}$. Then stratified sampling leads to*

$$e_{\text{prob}}^{\text{MC}}(n, \delta, W_p^1([0, 1]^d)) \leq n^{-1/d} \min\left\{1, n^{-(1-1/q)} \sqrt{\log \delta^{-1}}\right\}.$$

Proof We start with the one-dimensional case considering the method S_n^1 , see (26). Hence, the unit interval $[0, 1]$ is split into intervals G_0, \dots, G_{n-1} of length n^{-1} . For $x_1 < x_2$ from $[0, 1]$ we have

$$|f(x_2) - f(x_1)| = \left| \int_{[x_1, x_2]} f'(x) dx \right| \leq \int_{[x_1, x_2]} |f'(x)| dx.$$

Hence, on the i -th interval G_i , we bound $b_i := \text{ess sup}_{G_i} f - \text{ess inf}_{G_i} f$ by

$$b_i \leq \int_{G_i} |f'(x)| dx \leq n^{-1} \left(n \int_{G_i} |f'(x)|^q dx \right)^{1/q} = n^{-(1-1/q)} \|f'\|_{L_q(G_i)},$$

where we used Jensen's inequality. Furthermore

$$\|f'\|_{L_q([0,1])} = \left(\sum_{i \in [n]} \|f'\|_{L_q(G_i)}^q \right)^{1/q} \geq n^{1-1/q} \left(\sum_{i \in [n]} b_i^q \right)^{1/q} \geq n^{1-1/q} \left(\sum_{i \in [n]} b_i^2 \right)^{1/2},$$

exploiting $q \leq 2$ in the last inequality. Applying Hoeffding's inequality (Proposition 2) for f with $\|f\|_{W_p^1([0,1])} \leq 1$, and using $\|f'\|_{L_q([0,1])} \leq \|f\|_{W_p^1([0,1])}$, we obtain

$$\mathbb{P}\{|S_n^1(f) - \text{INT } f| > \varepsilon\} \leq 2 \exp\left(-2 n^{4-2/q} \varepsilon^2\right).$$

This is guaranteed to be at most δ for

$$\varepsilon = \frac{1}{\sqrt{2}} n^{-(2-1/q)} \sqrt{\log \frac{2}{\delta}},$$

which shows the probabilistic part of the assertion for $d = 1$. The worst case part of the bound follows from

$$\begin{aligned} |S_n^1(f) - \text{INT } f| &\leq \frac{1}{n} \sum_{i \in [n]} b_i \leq \frac{1}{n} \sum_{i \in [n]} \int_{G_i} |f'(x)| dx = n^{-1} \|f'\|_{L_1(G)} \\ &\leq n^{-1} \|f\|_{W_p^1(G)} \leq n^{-1}. \end{aligned}$$

In higher dimension, $d \geq 2$, splitting $[0, 1]^d$ into m^d subcubes $G_{\mathbf{i}}$ with $\mathbf{i} \in [m]^d$, we exploit the embedding $W_p^1([0, 1]^d) \hookrightarrow L_\infty([0, 1]^d)$. Namely, incorporating scaling, we bound the spread of function values within $G_{\mathbf{i}}$ by

$$b_{\mathbf{i}} := \text{ess sup}_{G_{\mathbf{i}}} f - \text{ess inf}_{G_{\mathbf{i}}} f \leq C m^{d/p-1} \|f\|_{W_p^1(G_{\mathbf{i}})}, \quad (27)$$

with some constant $C > 0$ depending only on p and d . From this, with $p > d \geq 2$, we conclude

$$\left(\sum_{\mathbf{i} \in [m]^d} b_{\mathbf{i}}^2 \right)^{1/2} \leq m^{d(1/2-1/p)} \left(\sum_{\mathbf{i} \in [m]^d} b_{\mathbf{i}}^p \right)^{1/p} \leq C m^{-(1-d/2)} \|f\|_{W_p^1([0,1]^d)},$$

compare (16). Applying Hoeffding's inequality similarly to the one-dimensional case, we obtain

$$e_{\text{prob}}^{\text{MC}} \left(S_m^d, \delta, W_p^1([0,1]^d) \right) \leq \frac{C}{\sqrt{2}} m^{-(1+d/2)} \sqrt{\log \frac{2}{\delta}}.$$

Further, applying Jensen's inequality leads to a worst case bound

$$|S_m^d(f) - \text{INT } f| \leq \frac{1}{m^d} \sum_{\mathbf{i} \in [m]^d} b_{\mathbf{i}} \leq \left(\frac{1}{m^d} \sum_{\mathbf{i} \in [m]^d} b_{\mathbf{i}}^p \right)^{1/p} \leq C m^{-1} \|f\|_{W_p^1([0,1]^d)}.$$

If a total number $n \in \mathbb{N}$ of function values to use is given, choosing $m := \lfloor n^{1/d} \rfloor$, similarly to the proof of Theorem 5, we obtain the right order. \square

Remark 6 The one-dimensional problem contains cases with small integrability $1 < p < 2$ for which we do not obtain the optimal δ -dependence. It is not known to us whether this is a deficiency of the method or of the proof. Anyway, to build an optimal algorithm in that case, we may use separation of the main part, which is equally simple since f may be approximated on G_1 by just one function value.

In the case of discontinuous functions, $p < d$, it remains challenging to find methods that detect and discourage outliers within stratified sampling. One idea might be to take several function values out of each subcube. This could improve also on the above mentioned case $d = 1$ and $1 < p < 2$. Any result in that direction might offer reasonable alternatives to control variates, where the case of small smoothness is also open.

4 Challenges in mixed smoothness spaces

In the recent years spaces of dominating mixed smoothness gained a lot of interest in the study of high-dimensional problems. For a survey on this topic we refer to the paper of Dũng, Temlyakov, and Ullrich [7].

For integer smoothness $r \in \mathbb{N}$ and integrability $1 \leq p \leq \infty$, on domains $G \subset \mathbb{R}^d$, Sobolev spaces of dominating mixed smoothness can be defined by

$$\mathbf{W}_p^{\text{mix}, r}(G) := \left\{ f \in L_p(G) \mid \|f\|_{\mathbf{W}_p^{\text{mix}, r}(G)} := \left(\sum_{\substack{\boldsymbol{\alpha} \in \mathbb{N}_0^d \\ |\boldsymbol{\alpha}|_{\infty} \leq r}} \|D^{\boldsymbol{\alpha}} f\|_{L_p(G)}^p \right)^{1/p} \leq \infty \right\}.$$

Lower bounds for the integration problem can be shown by scaling bump functions $\varphi: [0,1]^d \rightarrow \mathbb{R}$ in one coordinate, that is, for $m \in \mathbb{N}$ we define functions

$\psi_i(\mathbf{x}) := \varphi(mx_1 - i, x_2, \dots, x_d)$, where $i \in \{0, \dots, m-1\} = [m]$. By using those, similarly to Theorem 1 one can obtain

$$e_{\text{prob}}^{\text{MC}}(n, \delta, \mathbf{W}_p^{\text{mix},r}([0,1]^d)) \succeq \min \left\{ n^{-r}, n^{-(r+1-1/q)} (\log \delta^{-1})^{1-1/q} \right\}. \quad (28)$$

When talking about upper bounds it is useful to note that the integration problem is as difficult for the non-periodic spaces $\mathbf{W}_p^{\text{mix},r}(G)$ as for the zero-boundary space $\mathring{\mathbf{W}}_p^{\text{mix},r}(G) := \{f \in \mathbf{W}_p^{\text{mix},r}(\mathbb{R}^d) \mid \text{supp } f \subseteq G\}$. Namely, the integral of any function $f \in \mathbf{W}_p^{\text{mix},r}([0,1]^d)$, via a change of variables, can be traced back to the integral of a function $h := |\det \Phi'| \cdot (f \circ \Phi) \in \mathring{\mathbf{W}}_p^{\text{mix},r}([0,1]^d)$ with zero boundary condition, where $\Phi: [0,1]^d \rightarrow [0,1]^d$ is a smooth bijection. That way we only lose a constant, see Nguyen, Ullrich, and Ullrich [18]. Let us mention that our lower bounds are based on bump functions with zero boundary, so the lower bounds hold with the same constants.

The optimal order of convergence in terms of the root mean squared error was determined by Ullrich [29], namely

$$e_{2\text{-mean}}^{\text{MC}}(n, \mathbf{W}_p^{\text{mix},r}([0,1]^d)) \asymp n^{-(r+1-1/q)}, \quad \text{for } r \geq \max\{1/p - 1/2, 0\},$$

where $q = \min\{p, 2\}$. The result is based on a randomly shifted and dilated Frolov rule, developed by Krieg and Novak [15], given by

$$Q_{B,\mathbf{v}}(f) := \frac{1}{|\det B|} \sum_{\substack{\mathbf{m} \in \mathbb{Z}^d, \text{ with} \\ B^{-\top}(\mathbf{m}+\mathbf{v}) \in [0,1]^d}} f\left(B^{-\top}(\mathbf{m}+\mathbf{v})\right), \quad (29)$$

where $f \in \mathring{\mathbf{W}}_p^{\text{mix},r}([0,1]^d)$, which is of course only evaluated inside $[0,1]^d$. Here, $B = \text{diag}(\mathbf{u})B_n$ with dilation random variable \mathbf{u} and independent shift random variable \mathbf{v} distributed according to the uniform distribution in $[1/2, 3/2]^d$ and $[0,1]^d$, respectively, as well as a suitable generator matrix $B_n = n^{1/d}B_1$. ‘Suitable’ means $\det B_n = n$ and $\prod_{j=1}^d |(B_1 \mathbf{m})_j| \geq c > 0$ for all $\mathbf{m} \in \mathbb{Z}^d \setminus \{0\}$. In particular, the expected number of function evaluations is n . Via Theorem 2 one can build a method by independent repetition which provides

$$e_{\text{prob}}^{\text{MC}}(n, \delta, \mathbf{W}_p^{\text{mix},r}([0,1]^d)) \preceq \left(\frac{\log \delta^{-1}}{n} \right)^{r+1-1/q} \quad \text{for } n \succeq \log \delta^{-1}, \quad (30)$$

with $q := \min\{p, 2\}$. Unfortunately, we do not achieve the optimal dependence on δ . The algorithm (29) does not possess desirable confidence guarantees, as the following one-dimensional counter example shows. This is not surprising as the number of random parameters is fixed by the dimension and thus we do not expect to observe concentration phenomena for $n \rightarrow \infty$, which is in contrast to stratified sampling.

Example 1 We consider the integration problem in a one dimensional setting on $\mathring{\mathbf{W}}_2^{\text{mix},r}([0,1])$. The random Frolov rule that uses n function values on average is determined by

$$Q_n(f) := \frac{1}{un} \sum_{\substack{m \in \mathbb{Z}, \text{ with} \\ (m+v)/(un) \in [0,1]}} f\left(\frac{m+v}{un}\right),$$

with independent random variables u and v uniformly distributed in $[1/2, 3/2]$ and $[0, 1]$, respectively. Let $\varphi \in \mathring{\mathbf{W}}_2^{\text{mix},r}([0, 1])$ with integral $\gamma_0 := \int_0^1 \varphi dx > 0$ and norm $\|\varphi\|_{\mathbf{W}_2^{\text{mix},r}} \leq 1$. For the function

$$f_n(x) := (2n)^{-r} \sum_{\substack{k=0, \dots, n-1, \text{ with} \\ (2nx-2k) \in [0,1]}} \varphi(2nx-2k)$$

observe that $\|f_n\|_{\mathbf{W}_2^{\text{mix},r}} \leq 1$ and $\int_0^1 f_n dx = \gamma_0/2^{r+1} \cdot n^{-r}$. Furthermore, the algorithm returns 0 if all the function values are computed inside $\bigcup_{k=0}^{n-1} \left[\frac{2k+1}{2n}, \frac{k+1}{n} \right]$, which is where f_n vanishes. This happens if $\frac{v}{un} \in \left[\frac{1}{2n}, \frac{1}{n} \right]$ and $\frac{n-1+v}{un} \in \left[\frac{2n-1}{2n}, 1 \right]$, in particular for shifts $v \in \left[\frac{1}{2}, \frac{3}{4} \right]$ and dilations $u \in \left[1 - \frac{1}{4n}, 1 \right]$. This means, with probability exceeding $\delta_n := \frac{1}{16n}$ the error is $\gamma_0/2^{r+1} \cdot n^{-r}$, hence,

$$e_{\text{prob}}^{\text{MC}}(Q_n, \delta_n, \mathring{\mathbf{W}}_2^{\text{mix},r}([0, 1])) \succeq n^{-r} \succ n^{-(r+1/2)} \sqrt{\log \delta_n^{-1}} \asymp n^{-(r+1/2)} \sqrt{\log n}.$$

This reveals a significant gap to the general lower bound (28). \square

Separation of the main part does not provide the optimal rate in n , but the dependence on δ can be reduced. Since we may restrict to the integration problem for functions with zero boundary condition, $\mathring{\mathbf{W}}_p^{\text{mix},r}([0, 1])$, we may apply results for the approximation of periodic functions, denoted by $\widetilde{\mathbf{W}}_p^{\text{mix},r}([0, 1]^d)$. Namely,

$$e^{\text{det}}(n, \widetilde{\mathbf{W}}_p^{\text{mix},r}([0, 1]^d) \hookrightarrow L_p) \preceq n^{-r} (\log n)^{(r+1/2)(d-1)}, \quad \text{for } 1 < p < \infty,$$

which can be found in [7, (5.11)]. Applying Theorem 3, for $n \succeq \log \delta^{-1}$ we conclude that

$$e_{\text{prob}}^{\text{MC}}(n, \delta, \mathbf{W}_p^{\text{mix},r}([0, 1]^d)) \preceq n^{-(r+1-1/q)} (\log n)^{(r+1/2)(d-1)} (\log \delta^{-1})^{1-1/q}, \quad (31)$$

where $q := \min\{p, 2\}$. Here the δ -dependence is optimal, but the rate in n is affected by logarithmic terms.

Finally, deterministic quadrature is known to achieve

$$e^{\text{det}}(n, \mathbf{W}_p^{\text{mix},r}([0, 1]^d)) \asymp n^{-r} (\log n)^{(d-1)/2}, \quad \text{for } 1 < p < \infty, \quad (32)$$

see [7, Theorem 8.14]. This catches the case $n \prec \log \delta^{-1}$.

It remains a challenging open problem to find randomized integration methods that have the right dependence on the uncertainty while fully exploiting the smoothness.

A Domains with Lipschitz boundary

From Novak and Triebel [21], we cite the following definition of a Lipschitz domain.

Definition 1 A *bounded Lipschitz domain* in \mathbb{R}^d , where $d \geq 2$, is a non-empty bounded open connected set $G \subseteq \mathbb{R}^d$ such that its boundary ∂G can be covered by finitely many open balls $B_1, \dots, B_J \subseteq \mathbb{R}^d$, each centered at ∂G , such that

$$B_j \cap G = B_j \cap G_j \quad \text{for } j = 1, \dots, J,$$

where G_j are suitable rotations of sets

$$G'_j := \left\{ (\mathbf{x}', x_d) \in \mathbb{R}^{d-1} \times \mathbb{R} \mid x_d > h_j(\mathbf{x}') \right\} \subseteq \mathbb{R}^d$$

with Lipschitz continuous functions $h_j: \mathbb{R}^{d-1} \rightarrow \mathbb{R}$.

The boundary of G'_j is the graph of h_j , that is, $\partial G'_j = \{(\mathbf{x}', h_j(\mathbf{x}')) \mid \mathbf{x}' \in \mathbb{R}^{d-1}\}$. (The inclusion \supseteq is obvious, further, the strict epigraph G'_j of h_j is an open set due to the Lipschitz continuity of h_j , and the same holds for the strict hypograph $\{(\mathbf{x}', x_d) \in \mathbb{R}^{d-1} \times \mathbb{R} \mid x_d < h_j(\mathbf{x}')\}$.) In detail, if $\mathbf{x} = (\mathbf{x}', x_d) \in \mathbb{R}^{d-1} \times \mathbb{R}$ with $x_d \neq h_j(\mathbf{x}')$, then the Euclidean ball around \mathbf{x} with radius $r := \frac{|h_j(\mathbf{x}') - x_d|}{2 \cdot \max\{1, L_j\}}$, where $L_j \geq 0$ is the Lipschitz constant of h_j , does not intersect with the graph of h_j .) By the integrability of h_j on bounded domains $[-b, b]^{d-1}$, $b \in \mathbb{N}$, we have $\lambda^d(\partial G_j) = \lambda^d(\partial G'_j) = 0$ for $j = 1, \dots, J$. Therefore, the boundary $\partial G = \bigcup_{j=1}^J (B_j \cap \partial G_j)$ of a bounded Lipschitz domain G has Lebesgue measure $\lambda^d(\partial G) = 0$. Hence, Sobolev spaces defined on the closure \bar{G} coincide with the Sobolev spaces defined on G . For instance, the standard domain $[0, 1]^d$ is closed, but since its interior $(0, 1)^d$ is a Lipschitz domain, the closed unit cube $[0, 1]^d$ very well belongs to the realm of admissible domains, its boundary is Lipschitz.

Definition 2 A bounded measurable set $G \subseteq \mathbb{R}^d$ is said to have a *Lipschitz boundary* if its interior $\text{int}(G)$ is a bounded Lipschitz domain and $\partial G = \partial \text{int}(G)$.

B Estimates on binomial sums

In Section 2 we need the following two inequalities about binomial sums. The first lemma is a minor extension of [17, Proposition 7.3.2], holding also for odd k , and with slightly improved constants.

Lemma 3 For all $k \in \mathbb{N}$ and $t \in \mathbb{N}_0$ we have

$$\begin{aligned} 2^{-k} \sum_{j=0}^{\lfloor k/2 \rfloor - t} \binom{k}{j} &= 2^{-k} \sum_{j=\lceil k/2 \rceil + t}^k \binom{k}{j} \\ &\geq \frac{1}{2 + 4/\sqrt{\pi}} \begin{cases} \exp\left(-\frac{16(\log 2)t^2}{k}\right) & \text{for odd } k \text{ and } t \in [0, \frac{k+3}{8}], \\ \exp\left(-\frac{16(\log 2)(t-1/2)^2}{k}\right) & \text{for even } k \text{ and } t \in [0, \frac{k+6}{8}]. \end{cases} \end{aligned}$$

Proof First, recall that $\binom{k}{\lfloor k/2 \rfloor} < 2^k / \sqrt{\pi \lceil k/2 \rceil}$, which for even k follows from Stirling's approximation, $\sqrt{2\pi} n^{n+1/2} \exp(-n) \exp\left(\frac{1}{12n+1}\right) < n! < \sqrt{2\pi} n^{n+1/2} \exp(-n) \exp\left(\frac{1}{12n}\right)$, see [25], and for odd k can be derived from $k+1$ via Pascal's rule. Hence,

$$2^{-k} \sum_{j=0}^{\lfloor k/2 \rfloor - t} \binom{k}{j} \geq \frac{1}{2} - 2^{-k} t \binom{k}{\lfloor k/2 \rfloor} > \frac{1}{2} - \frac{t}{\sqrt{\pi \lceil k/2 \rceil}}.$$

For $0 \leq t \leq \sqrt{\lceil k/2 \rceil} / (1 + 2/\sqrt{\pi})$, this gives the absolute lower bound $\frac{1}{2 + 4/\sqrt{\pi}}$.

For larger t we follow the approach of [17, Proposition 7.3.2]. Basic estimates yield

$$\begin{aligned} 2^{-k} \sum_{j=0}^{\lfloor k/2 \rfloor - t} \binom{k}{j} &\geq 2^{-k} \sum_{j=\lceil k/2 \rceil - 2t + 1}^{\lfloor k/2 \rfloor - t} \binom{k}{j} \\ &\geq 2^{-k} t \binom{k}{\lceil k/2 \rceil - 2t + 1} \\ &= 2^{-k} t \binom{k}{\lfloor k/2 \rfloor} \prod_{i=1}^{2t-1} \frac{\lceil k/2 \rceil - 2t + 1 + i}{\lceil k/2 \rceil + i} \\ &\geq 2^{-k} t \binom{k}{\lfloor k/2 \rfloor} \left(\frac{\lceil k/2 \rceil - 2t + 2}{\lceil k/2 \rceil + 1} \right)^{2t-1}. \end{aligned}$$

Next, we use $1 - x \geq \exp(-2(\log 2)x)$ for $0 \leq x \leq 1/2$. For odd k we set $x = 2t/(\lceil k/2 \rceil + 1)$ and for even k we set $x = (2t - 1)/(k/2 + 1)$. This is where $t \leq (k+6)/8$ for even k , and $t \leq (k+3)/8$ for odd k , comes into play. Finally, we use $\binom{k}{\lceil k/2 \rceil} \geq 2^k/(2\sqrt{\lceil k/2 \rceil})$, and obtain

$$2^{-k} \sum_{j=0}^{\lceil k/2 \rceil - t} \binom{k}{j} \geq \frac{t}{2\sqrt{\lceil k/2 \rceil}} \begin{cases} \exp\left[-\frac{8(\log 2)t(t-1/2)}{\lceil k/2 \rceil + 1}\right] & \text{for odd } k, \\ \exp\left[-\frac{8(\log 2)(t-1/2)^2}{k/2 + 1}\right] & \text{for even } k. \end{cases}$$

For $t \geq \sqrt{\lceil k/2 \rceil}/(1 + 2/\sqrt{\pi})$, the prefactor simplifies as stated in the claimed inequality. \square

Lemma 4 For all $k, k' \in \mathbb{N}_0$ with $k \geq k'$ we have

$$2^{-k} \left[\sum_{j=0}^{\lfloor \frac{k-k'}{2} \rfloor} \binom{k}{j} + \sum_{j=\lfloor \frac{k+k'+1}{2} \rfloor}^k \binom{k}{j} \right] \geq 2^{-k'}.$$

Proof The proof follows by induction over $k \geq k'$. A speciality here is that in the induction step we assume the statement for k and prove it for $k+2$, which is sufficient when the base case is verified for $k = k'$ and $k = k' + 1$.

For $k = k'$ and $k = k' + 1$ we have $2^{-k'} \binom{k}{0}$ and $2^{-(k'+1)} [\binom{k'}{0} + \binom{k'+1}{1}]$, respectively, which proves the inequality. (We even have equality.)

For the induction step from k to $k+2$ where $k \geq k'$, via Pascal's rule, as well as using $\binom{k}{\lfloor \frac{k+2-k'}{2} \rfloor} \geq \binom{k}{\lfloor \frac{k-k'}{2} \rfloor}$, we obtain

$$\sum_{j=0}^{\lfloor \frac{k+2-k'}{2} \rfloor} \binom{k+2}{j} = 4 \sum_{j=0}^{\lfloor \frac{k-k'}{2} \rfloor - 1} \binom{k}{j} + 3 \binom{k}{\lfloor \frac{k-k'}{2} \rfloor} + \binom{k}{\lfloor \frac{k+2-k'}{2} \rfloor} \geq 4 \sum_{j=0}^{\lfloor \frac{k-k'}{2} \rfloor} \binom{k}{j}.$$

Similarly, with $\binom{k}{\lfloor \frac{k+k'+1}{2} \rfloor} \geq \binom{k}{\lfloor \frac{k+k'+3}{2} \rfloor}$, one can show

$$\sum_{j=\lfloor \frac{k+k'+3}{2} \rfloor}^{k+2} \binom{k+2}{j} \geq 4 \sum_{j=\lfloor \frac{k+k'+1}{2} \rfloor}^k \binom{k}{j}.$$

Now, by the induction hypothesis the assertion is proven. \square

C Lower bounds for adaptive methods

We supplement the discussion of Section 2.1 by a brief theoretical treatment of *adaptive* methods without repeating notions that have already been explained there. All upper bounds within this paper rely on non-adaptive methods, most with fixed cardinality, some with varying cardinality. For lower bounds, it is desirable to extend them to as broad classes of algorithms as possible, hopefully showing that additional features such as adaptivity are not helpful, thus sticking to simple methods is justified.

An abstract *adaptive Monte Carlo algorithm* for functions from \mathcal{W} is a family $A = (A^\omega)_{\omega \in \Omega}$ of mappings $A^\omega = \phi^\omega \circ \iota^\omega$, where $\iota^\omega : \mathcal{W} \rightarrow c_{00} := \bigcup_{n \in \mathbb{N}_0} \mathbb{R}^n$ returns a terminating sequence $\mathbf{y} = \iota^\omega(f) \in \mathbb{R}^{\tilde{n}(\omega, f)}$ of function values at *sequentially selected nodes*, with

$$y_1 = f(\mathbf{x}_1^\omega), \quad \text{and} \quad y_i = f(\mathbf{x}_i^\omega(y_1, \dots, y_{i-1})), \quad \text{for } i = 2, \dots, \tilde{n}(\omega, f).$$

Here, the input-dependent random cardinality $\text{card}(A, f) : \omega \mapsto \tilde{n}(\omega, f)$ is determined by a *stopping rule*, that is, the decision whether or not to stop after n function evaluations only depends on the randomness and the function values collected up to that point, in detail,

$\mathbb{1}[\bar{n}(\omega, f) \leq i] = T^\omega(y_1, \dots, y_i)$ with a so-called *termination function* $T^\omega : c_{00} \rightarrow \{0, 1\}$. Taking the classical worst case perspective with respect to the input class, we aim to study the power of algorithms with a (possibly non-integer) bound $\bar{n} > 0$ on the worst expected cardinality,

$$\overline{\text{card}}(A, \mathcal{W}) := \sup_{\|f\|_{\mathcal{W}} \leq 1} \mathbb{E} \text{card}(A, f), \quad (33)$$

i.e., the quantity of interest is

$$\tilde{e}_{\text{prob}}^{\text{MC-ada}}(\bar{n}, \delta, \mathcal{W}) := \inf_{A: \overline{\text{card}}(A, \mathcal{W}) \leq \bar{n}} e(A, \delta, \mathcal{W}). \quad (34)$$

As in Section 2.1, we consider a discrete probability measure μ on the set of inputs. For fixed $\omega \in \Omega$, we define the μ -average cardinality as

$$\text{card}(A^\omega, \mu) := \int_{\mathcal{B}_{\mathcal{W}}} \text{card}(A^\omega, f) d\mu(f). \quad (35)$$

A Fubini argument shows that $\overline{\text{card}}(A, \mathcal{W}) \geq \mathbb{E} \text{card}(A, \mu)$, and, provided $\overline{\text{card}}(A, \mathcal{W}) \leq \bar{n}$, Markov's inequality implies

$$\mathbb{P}\{\text{card}(A, \mu) \leq 2\bar{n}\} \geq \frac{1}{2}. \quad (36)$$

Analogously to (11), we perform a trick in the spirit of Bakhvalov [4] and obtain

$$\sup_{\|f\|_{\mathcal{W}} \leq 1} \mathbb{P}\{|A(f) - \text{INT } f| > \varepsilon\} \geq \frac{1}{2} \inf_{Q: \overline{\text{card}}(Q, \mu) \leq 2\bar{n}} \mu\{f: |Q(f) - \text{INT } f| > \varepsilon\}, \quad (37)$$

where the infimum is taken over deterministic algorithms Q (for each fixed $\omega \in \Omega$, we may consider A^ω to be deterministic).

How does Lemma 1 change if we allow for adaptive algorithms? Roughly speaking, our lower bounds on $e_{\text{prob}}^{\text{MC}}(n, \delta, \mathcal{W})$, that is, with non-adaptive methods with fixed cardinality, hold for $\tilde{e}_{\text{prob}}^{\text{MC-ada}}(\frac{n}{4}, \frac{\delta}{4}, \mathcal{W})$ in the adaptive setting with varying cardinality.

Lemma 5 *Let $\bar{n} \geq 17/4$ and $N \in \mathbb{N}$ with $N \geq 20\bar{n} + 6$, and let $\gamma > 0$. Assume that there are functions $f_i : G \rightarrow \mathbb{R}$, for $i = 1, \dots, N$, satisfying conditions 1. and 2. as in Lemma 1. Then, for any uncertainty level $0 < \delta < 1/12$, we have*

$$\tilde{e}_{\text{prob}}^{\text{MC-ada}}(\bar{n}, \delta, \mathcal{W}) \geq \gamma \min \left\{ 2\bar{n}^{1/2} \sqrt{\log_4 \frac{1}{12\delta}}, 4\bar{n} \right\}$$

Proof (idea) The μ -average input setting is as in Lemma 1, that is, we plug random function $f_{\mathbf{S}}$ into deterministic algorithms Q . Assuming $\text{card}(Q, \mu) \leq 2\bar{n}$, by Markov's inequality it follows

$$\mathbb{P}\{\text{card}(Q, f_{\mathbf{S}}) > 4\bar{n}\} \leq \frac{1}{2}.$$

Conditioning on $\text{card}(Q, f_{\mathbf{S}}) \leq 4\bar{n}$, thanks to the symmetry of the measure μ , also for adaptive methods Q , we end up with an estimate on binomial sums, compare proof of Lemma 1. \square

The changes for Lemma 2 are similar, our lower bounds on $e_{\text{prob}}^{\text{MC}}(n, \delta, \mathcal{W})$ constitute lower bounds for $\tilde{e}_{\text{prob}}^{\text{MC-ada}}(\frac{n}{9}, \frac{\delta}{3}, \mathcal{W})$ as well.

Lemma 6 *Let $\bar{n} > 0$ and $N, M \in \mathbb{N}$ with $N \geq \lceil 36\bar{n} \rceil$ and $M \leq N$, and let $\gamma > 0$. Assume that there are functions $f_i : G \rightarrow \mathbb{R}$, for $i = 1, \dots, N$, satisfying conditions 1. and 2. as in Lemma 2, Then, for any uncertainty level $0 < \delta < \frac{1}{6} 2^{-\lceil M/2 \rceil}$, we have*

$$\tilde{e}_{\text{prob}}^{\text{MC-ada}}(\bar{n}, \delta, \mathcal{W}) \geq \frac{1}{2} \gamma M.$$

Proof (idea) The μ -average setting is as in Lemma 2, that is, we plug random functions $f_{I,\mathbf{S}}$ into deterministic algorithms Q . Assuming $\text{card}(Q, \mu) \leq 2\bar{n}$, by Markov's inequality it follows

$$\mathbb{P}\{\text{card}(Q, f_{I,\mathbf{S}}) > 6\bar{n}\} \leq \frac{1}{3}. \quad (38)$$

We aim to bound the expected number $m(I, \mathbf{S})$ of subdomains detected by Q where the function $f_{I,\mathbf{S}}$ does not vanish. Let Q' be an algorithm that always uses $n' = \lfloor 6\bar{n} \rfloor$ function values from n' different subdomains G_i , and that computes the same function values $f_{I,\mathbf{S}}(\mathbf{x}_i)$ as Q for $i = 1, \dots, \min\{\text{card}(Q, f_{I,\mathbf{S}}), n'\}$. Let $m'(I, \mathbf{S})$ denote the expected number of subdomains detected by Q' where the function $f_{I,\mathbf{S}}$ does not vanish. For this quantity we have

$$\mathbb{E} m'(I, \mathbf{S}) = \frac{n'}{N} M \leq \frac{1}{6} M,$$

and by Markov's inequality we conclude

$$\mathbb{P}\{m'(I, \mathbf{S}) > \frac{1}{2} M\} \leq \frac{1}{3}. \quad (39)$$

Combining (38) and (39), we bound

$$\mathbb{P}\{m(I, \mathbf{S}) \leq \frac{1}{2} M\} \geq \mathbb{P}\left\{\text{card}(Q, f_{I,\mathbf{S}}) \leq n' \text{ and } m'(I, \mathbf{S}) \leq \frac{1}{2} M\right\} \geq 1 - \frac{1}{3} - \frac{1}{3} = \frac{1}{3}.$$

The remaining part of the proof follows the lines of the proof of Lemma 2 after (15). \square

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. R.A. ADAMS, J.J.F. FOURNIER. *Sobolev Spaces*, second edition. Academic Press, 2002.
2. N. ALON, Y. MATIAS, M. SZEGEDY. The space complexity of approximating the frequency moments. *STOC '96 Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, 20–29, 1996.
3. B. VON BAHR, C. ESSEEN. Inequalities for the r th absolute moment of a sum of random variables, $1 \leq r \leq 2$. *Annals of Mathematical Statistics*, 36:299–303, 1965.
4. N. S. BAKHVALOV. On the approximate calculation of multiple integrals. *Vestnik MGU, Ser. Math. Mech. Astron. Phys. Chem.* 4:3–18, 1959, in Russian. English translation: *Journal of Complexity*, 31(4):502–516, 2015.
5. O. CATONI. Challenging the empirical mean and empirical variance: A deviation study. *Annales de l'Institut Henri Poincaré – Probabilités et Statistiques*, 48(4):1148–1185, 2012.
6. P.G. CIARLET. *The Finite Element Method for Elliptic Problems*. SIAM, 1978.
7. D. DÜNG, V.N. TEMLYAKOV, T. ULLRICH. *Hyperbolic Cross Approximation*. Advanced Courses in Mathematics. CRM Barcelona. Birkhuser/Springer, to appear.
8. L. GAJEK, W. NIEMIRO, P. POKAROWSKI. Optimal Monte Carlo integration with fixed relative precision. *Journal of Complexity*, 29:4–26, 2013.
9. S. HEINRICH. Random approximation in numerical analysis. *Proceedings of the Conference "Functional Analysis" Essen*, pp. 123–171, 1994.
10. S. HEINRICH. Randomized approximation of Sobolev embeddings. *Proceedings of the MCQMC 2006*, 445–459, 2008.
11. F.J. HICKERNELL, L. JIANG, Y.W. LIU, A. OWEN. Guaranteed conservative fixed width confidence intervals via Monte Carlo sampling. *Proceedings of the MCQMC 2012*, Springer, 2013.
12. M. HUBER. A Bernoulli mean estimate with known relative error distribution. *Random Structures and Algorithms*, 50:173–182, 2017.
13. M. HUBER. An optimal (ϵ, δ) -approximation scheme for the mean of random variables with bounded relative variance. Available on arXiv:1706.01478v1 [stat.CO], 2017.

14. M. JERRUM, L. VALIANT, V. VAZIRANI. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Sciences*, 43:169–188, 1986.
15. D. KRIEG, E. NOVAK. A universal algorithm for univariate integration. *Foundations of Computational Mathematics*, 17:895–916, 2017.
16. R.J. KUNSCH, E. NOVAK, D. RUDOLF. Solvable integration problems and optimal sample size selection. *Journal of Complexity*, 53:40–67, 2019.
17. J. MATOUSEK, J. VONDRAK. *The Probabilistic Method*, manuscript.
18. V.K. NGUYEN, M. ULLRICH, T. ULLRICH. Change of variable in spaces of mixed smoothness and numerical integration of multivariate functions on the unit cube. *Constructive Approximation*, 46:69–108, 2017.
19. W. NIEMIRO, P. POKAROWSKI. Fixed precision MCMC estimation by median of product of averages. *Journal of Applied Probability*, 46:309–329, 2009.
20. E. NOVAK. Deterministic and Stochastic Error Bounds in Numerical Analysis. *Lecture Notes in Mathematics* 1349, Springer-Verlag, Berlin, 1988.
21. E. NOVAK, H. TRIEBEL. Function spaces in Lipschitz domains and optimal rates of convergence for sampling. *Constructive Approximation*, 23:325–350, 2006.
22. E. NOVAK, H. WOŹNIAKOWSKI. *Tractability of Multivariate Problems*, Volume I, Linear Information. European Mathematical Society, 2008.
23. E. NOVAK, H. WOŹNIAKOWSKI. *Tractability of Multivariate Problems*, Volume II, Standard Information for Functionals. European Mathematical Society, 2010.
24. E. NOVAK, H. WOŹNIAKOWSKI. *Tractability of Multivariate Problems*, Volume III, Standard Information for Operators. European Mathematical Society, 2012.
25. H. ROBBINS. A remark on Stirling’s formula. *The American Mathematical Monthly*, 62(1):26–29, 1955.
26. D. RUDOLF, N. SCHWEIZER. Error bounds of MCMC for functions with unbounded stationary variance. *Statistics & Probability Letters*, 99:6–12, 2015.
27. J.F. TRAUB, G.W. WASILKOWSKI, H. WOŹNIAKOWSKI. *Information-Based Complexity*. Academic Press, 1988.
28. H. TRIEBEL. *Theory of Function Spaces*. Birkhäuser, 1983.
29. M. ULLRICH. A Monte Carlo method for integration of multivariate smooth functions. *SIAM Journal on Numerical Analysis*, 55(3):1188–1200, 2017.