Journal of Complexity 53 (2019) 40–67

Contents lists available at ScienceDirect

Journal of Complexity

journal homepage: www.elsevier.com/locate/jco

Solvable integration problems and optimal sample size selection☆

Robert J. Kunsch [a,*], Erich Novak [b], Daniel Rudolf [c]

[a] Institut für Mathematik, Universität Osnabrück, Albrechtstr. 28a, 49076 Osnabrück, Germany
[b] Mathematisches Institut, Universität Jena, Ernst-Abbe-Platz 2, 07743 Jena, Germany
[c] Institute for Mathematical Stochastics, Universität Göttingen & Felix-Bernstein-Institute for Mathematical Statistics, Goldschmidtstraße 7, 37077 Göttingen, Germany

ARTICLE INFO

Article history:
Received 6 June 2018
Received in revised form 27 September 2018
Accepted 23 October 2018
Available online 1 November 2018

Keywords:
Monte Carlo integration
Information-based complexity
Adaptive sample size
Confidence intervals
Variance estimation

ABSTRACT

We compute the integral of a function or the expectation of a random variable with minimal cost and use, for our new algorithm and for upper bounds of the complexity, i.i.d. samples. Under certain assumptions it is possible to select a sample size based on a variance estimation, or – more generally – based on an estimation of a (central absolute) p-moment. That way one can guarantee a small absolute error with high probability, the problem is thus called solvable. The expected cost of the method depends on the p-moment of the random variable, which can be arbitrarily large.

In order to prove the optimality of our algorithm we also provide lower bounds. These bounds apply not only to methods based on i.i.d. samples but also to general randomized algorithms. They show that – up to constants – the cost of the algorithm is optimal in terms of accuracy, confidence level, and norm of the particular input random variable. Since the considered classes of random variables or integrands are very large, the worst case cost would be infinite. Nevertheless one can define adaptive stopping rules such that for each input the expected cost is finite.

We contrast these positive results with examples of integration problems that are not solvable.

☆ Communicated by H. Wozniakowski.

* Corresponding author.
E-mail addresses: robert.kunsch@uni-osnabrueck.de (R.J. Kunsch), erich.novak@uni-jena.de (E. Novak), daniel.rudolf@uni-goettingen.de (D. Rudolf).

https://doi.org/10.1016/j.jco.2018.10.007
0885-064X/

1. Introduction

We want to compute the integral

$$\mathrm{INT}(f) = \int_G f(\mathbf{x})\,\mathrm{d}\pi(\mathbf{x}) \tag{1}$$

of a π-integrable function $f : G \to \mathbb{R}$ where π is a probability measure on the domain G, or the expectation

$$\mathbb{E}\,Y = \int_\Omega Y(\omega)\,\mathrm{d}\mathbb{P}(\omega) \tag{2}$$

of a random variable Y mapping from a probability space $(\Omega, \Sigma, \mathbb{P})$ to $\mathbb{R}$. We allow randomized algorithms and want to achieve this goal up to some absolute error tolerance $\varepsilon > 0$, based on finitely many function values or i.i.d. samples, respectively.

Ideally, we would like to do it for an arbitrary L_1-function f or an arbitrary integrable random variable Y. But there is no algorithm for this task, even if we assume C^∞-smoothness of the integrand, see Theorem 3.15. Hence, we have to shrink the set of possible inputs. In this paper we define solvable integration problems in a very natural way: INT is *solvable* for a class $\mathcal{F}$ of integrands if, for every $\varepsilon, \delta > 0$, there is an algorithm A such that

$$\mathbb{P}\{|A(f) - \mathrm{INT}(f)| \le \varepsilon\} \ge 1 - \delta, \tag{3}$$

for all $f \in \mathcal{F}$. The cost of A, that is, the number of used function evaluations may depend on f but is almost surely finite. If (3) holds for $\varepsilon, \delta > 0$, then the method A is called (ε, δ)-*approximating*. Analogously we define solvability for $\mathbb{E}$ and classes $\mathcal{Y}$ of random variables. With this concept we can allow quite large sets $\mathcal{F}$. Observe that we deviate from the existing literature on information-based complexity (IBC), see for example the books of Novak and Woźniakowski [29–31], where "solvable" means that we have a uniform cost bound and usually only bounded sets $\mathcal{F}$ are considered.

In this paper we mainly discuss input classes of the form

$$\begin{aligned} \mathcal{F}_{p,q,K} = \mathcal{F}_{p,q,K}(\pi) &:= \{f \in L_q(\pi) \mid \|f - \mathrm{INT}(f)\|_q \le K\,\|f - \mathrm{INT}(f)\|_p\}, \\ \mathcal{Y}_{p,q,K} &:= \{Y \in L_q \mid \|Y - \mathbb{E}\,Y\|_q \le K\,\|Y - \mathbb{E}\,Y\|_p\}, \end{aligned} \tag{4}$$

where $1 \le p < q \le \infty$ and $K > 1$. The assumption, in fact, is a cone condition on (central) norms, for $f \in \mathcal{F}_{p,q,K}$ (resp. $Y \in \mathcal{Y}_{p,q,K}$) any scaled and shifted version $af + c$ (resp. $aY + c$) with $a, c \in \mathbb{R}$ is still in the input class. These cone-shaped classes give a theoretical foundation to the idea of adaptively determining a sample size for i.i.d. sampling on ground of a variance estimation, or – more generally – a p-moment estimation. The case $p = 2$ and $q = 4$ (bounded kurtosis) was studied by Hickernell, Jiang, Liu, and Owen [13].

We stress that such problems cannot be solved in the sense that the expected error (or root mean squared error) is bounded for the whole class, see Theorem 3.5. Hence, the understanding of "solvable" as "small error with high probability" is essential here.

We provide a new algorithm $A^{\varepsilon,\delta}_{p,q,K}$ that solves (2) on classes $\mathcal{Y}_{p,q,K}$ (respectively, (1) on $\mathcal{F}_{p,q,K}$) and prove an upper bound of the cost. To measure the cost, we consider the random variable $n(\omega, Y)$ that determines the sample size of the algorithm, i.e. the number of copies of Y used by the algorithm. Then the *expected cost* for some input Y is given by $\bar{n}(A^{\varepsilon,\delta}_{p,q,K}, Y) := \mathbb{E}\,n(\,\cdot\,, Y)$. Our main results are:

- For arbitrary error thresholds $\varepsilon > 0$ and uncertainty levels $\delta \in (0, 1/2)$ we construct an (ε, δ)-approximating algorithm $A^{\varepsilon,\delta}_{p,q,K}$ such that for $Y \in \mathcal{Y}_{p,q,K}$

$$\bar{n}(A^{\varepsilon,\delta}_{p,q,K}, Y) \le C_q\,K^{pq/(q-p)} \left(1 + \left(\frac{\|Y - \mathbb{E}\,Y\|_1}{\varepsilon}\right)^{\max\{1+1/(q-1),\,2\}}\right) \log \delta^{-1} \tag{5}$$

  where $C_q > 0$, see Theorem 2.17 in Section 2.
- In addition to that, in Section 3 we show matching lower bounds holding for *all* algorithms, which we explain carefully in the following.

Irrespective of the specified accuracy ε and the input random variable, the cost bound (5) is at least some constant times $K^{pq/(q-p)} \log \delta^{-1}$. This fixed cost comes from estimating $\|Y - \mathbb{E}\, Y\|_1$ within our particular algorithm and can be interpreted as the price we need to pay for not knowing the statistical dispersion of Y. Theorem 3.7 shows that such constant effort is inevitable for all (ε, δ)-approximating algorithms, with a lower bound that reproduces the dependence on K and δ. The cost bound (5) also reveals the adaption of our algorithm to the input, the lower bounds of Theorems 3.9 and 3.10 exhibit the same dependence on ε, δ, and the norm. In doing so, we obtain (up to constants) tight upper and lower bounds for the (ε, δ)-complexity of computing expected values (2).

For classes of inputs like (4), problems (1) and (2) are closely related. However, the class of (randomized) algorithms for the integration problem (1), as considered in numerical analysis or complexity theory, is usually broader than the class of algorithms that are used in statistics for the expectation problem (2). For the numerical problem INT we consider all randomized algorithms that use finitely many function values for any input $f \in \mathcal{F}$, whereas for the statistical problem $\mathbb{E}$ the kind of available information is restricted: We may use only information from realizations of independent random variables $Y_1, Y_2, \ldots, Y_n$, which all have the same distribution as the *basis experiment* $Y \in \mathcal{Y}$, where $n = n(\omega, Y)$ is an almost surely finite random variable given by a stopping rule that depends on the already observed experiments. On the other hand, methods designed for computing expected values can be used for integration problems by building Monte Carlo experiments $Y_i := f(\mathbf{X}_i)$ with i.i.d. sampling nodes $\mathbf{X}_1, \mathbf{X}_2, \ldots \overset{\text{iid}}{\sim} \pi$. Usually, the interpretation of computing expected values as a statistical problem means that no deeper knowledge on the probability space on which Y is defined is taken into consideration. One can design algorithms that work for simulated samples as well as real-live data, all of which are provided by the user. In contrast, the numerical interpretation opens the possibility of including knowledge on the local structure of integrands f from the input class, typically smoothness assumptions, which could make simple i.i.d. sampling inferior to more sophisticated methods. Such locality, however, is not present in the definition of the classes $\mathcal{F}_{p,q,K}$. For the wider class of general randomized algorithms that are feasible for the integration problem (1), lower bound proofs are more technical and postponed to Section 3.4. Actually, we reproduce the lower bound results for i.i.d.-based algorithms, see in particular Corollary 3.14.

The new results for complexity on cones (4) are different from classical statements made in numerical analysis, where the cost of an algorithm is given by the cost for the worst input from a class of inputs (usually a bounded set). Here, the cost is infinite for the whole class $\mathcal{Y}_{p,q,K}$, but it is finite for any single input Y.

Historical remarks.

Varying sample size has been considered in statistical testing since the 1940s, so-called *sequential tests* have been promoted by Wald 1945 [37] and 1948 [38]. Later books on sequential methods include Siegmund 1985 [34, Chapter VII]. We also refer to the literature given in these sources. Most of the early works deal with quite restricted assumptions on the class of input random variables, e.g. normality in Stein 1945 [35]. Other results under more general assumptions, e.g. finite (but unknown) variance, are of asymptotic nature, see Chow and Robbins 1965 [6]. To our knowledge, Hickernell et al. 2013 [13] were the first who by introducing the cones $\mathcal{Y}_{2,4,K}$ provided general conditions on distributions such that the mean of a random variable with unknown variance can be estimated to within a given absolute error threshold with guaranteed confidence. We improve the algorithm and the upper bound and also prove matching lower bounds.

Studies on optimally adapted sample size selection can be found in papers of Dagum et al. [8], Gajek, Niemiro, and Pokarowski [10], and Huber [15]. These authors study the complexity of approximating the expected value $\mathbb{E}\, Y > 0$ of a random variable $0 \le Y \le 1$ to within a prescribed relative error, in detail $|A(Y) - \mathbb{E}\, Y|/\mathbb{E}\, Y \le \varepsilon$, with high confidence $1-\delta$. In such a situation, the cost is roughly inversely proportional to the target value $\mathbb{E}\, Y$, hence increasing as the latter approaches 0. The stopping rule is rather simple: The sample size n is chosen as the first number such that $Y_1 + \ldots + Y_n \ge r$ for some given threshold $r > 0$, the output is then r/n. Here, $r \asymp \varepsilon^{-2} \log \delta^{-1}$.

Outlook.

This paper is in the tradition of IBC, see [27,36]. A recent survey on the complexity of the integration problem is [28]. Let F_1, F_2 and R be normed linear spaces. For a (linear) solution operator $S : F_2 \to R$, in IBC one usually assumes that the input f is from a ball $\mathcal{B}_2 \subset F_2$. In the underlying work we assume that f is from a cone $\mathcal{F}_{p,q,K}$. The idea to take "cones, not balls" was suggested and studied by Hickernell and his colleagues, see [7] and [13]. Interesting cones could be of the form

$$\mathcal{F} = \{f \in F_2 \mid \|f\|_{F_2} \le K \, \|f\|_{F_1}\}, \tag{6}$$

where $\|\cdot\|_{F_i}$ denotes the norm of F_i for $i = 1,2$, and F_2 is embedded into F_1 with $\|\cdot\|_{F_1} \le C \, \|\cdot\|_{F_2}$ and $K > 1/C > 0$. For numerical problems of this type one can ask: Is there a deterministic or randomized algorithm that terminates (with finite cost, depending on f) for every input $f \in \mathcal{F}$ with a (probabilistic) ε-approximation with respect to the error measured in the norm $\|\cdot\|_R$ of the space R? The worst case cost might be infinite for any succeeding algorithm, still one could examine whether the cost that an algorithm A exhibits for an input f is close to the minimal cost that any algorithm which succeeds on $\mathcal{F}$ would possess for this particular input f. In our setting, we partially answer this question with the concept of the minimal fixed cost, see Theorem 3.7, but there we lack a dependence on the particular input. By intersecting $\mathcal{F}$ with a ball $\mathcal{B} \subset F_2$, though, we can study at least the classical worst case complexity on $\mathcal{F} \cap \mathcal{B}$ to get an idea on how the cost must depend on the size of the input, compare Theorems 3.9 and 3.10. Such an analysis seems to be appropriate for algorithms on the cone $\mathcal{F}$ that use an estimator for $\|f\|_{F_1}$ with an error estimate proportional to $\|f\|_{F_2}$. Actually, a main ingredient of our algorithm is the approximation of the (centered) L_p-norm of a function or random variable in $F_2 = L_q$. The problem of estimating the L_p-norm was recently studied by Heinrich [12] for functions from a Sobolev space F_2. The cone assumption in (6) then implies an upper bound for the stronger norm $\|f\|_{F_2}$, and optimal algorithms for the problem S on the ball $\mathcal{B}_2 \subset F_2$ can be scaled accordingly to conclude the new method.

2. A moment-adapted algorithm and its analysis

2.1. The general structure of the algorithm

The aim is to compute the expected value $\mathbb{E}\,Y$ for random variables Y from the cone $\mathcal{Y}_{p,q,K}$, see (4). The special case $p = 2$ and $q = 4$ has been introduced by Hickernell et al. 2013 [13] together with an algorithm that solves the problem. We modify their method, that way providing an algorithmic solution to the whole parameter range, $1 \le p < q \le \infty$, and obtain optimal cost rates in ε and δ.

Similarly to Hickernell et al. [13], see also Stein [35], we consider a two-stage algorithm. In the first step we estimate the central absolute p-moment of Y, based on which we choose the number of samples that we take for the second step in order to compute a good approximation of the mean $\mathbb{E}\,Y$. In both stages a median of independent estimators is used.

Algorithm 2.1 ($A^{k',s,\eta}_{k,p,m}$). *Parameters:* odd natural numbers k and k', integer $m \ge 2$; real numbers $p, s \ge 1$, and $\eta > 1$.

Input: Generator for i.i.d. copies $Y_1, Y_2, \ldots$ of a random variable Y.

Proceed as follows:

1. Realize k independent p-moment estimations using $n_1 := k\,m$ observations,

$$\widehat{R}^{(\ell)}_m := \frac{1}{m} \sum_{i=(\ell-1)m+1}^{\ell m} |Y_i - \widehat{M}^{(\ell)}_m|^p, \qquad \text{where} \quad \widehat{M}^{(\ell)}_m := \frac{1}{m} \sum_{i=(\ell-1)m+1}^{\ell m} Y_i,$$

for $\ell = 1, \ldots, k$, and determine their median (which is well defined since k is odd),

$$\widehat{R}_{k,m} := \operatorname{med}(\widehat{R}^{(1)}_m, \ldots, \widehat{R}^{(k)}_m).$$

2. Realize k' independent mean estimations, using $n_2 := k'\,m'$ additional observations, where we choose $m' := \max\{\lceil \eta\,(\widehat{R}_{k,m})^s \rceil,\ 1\}$,

$$\widetilde{M}_{m'}^{(\ell)} := \frac{1}{m'} \sum_{i=n_1+(\ell-1)m'+1}^{n_1+\ell m'} Y_i,$$

here $\ell = 1, \dots, k'$. Finally, output the median (which is well defined since k' is odd),

$$A_{k,p,m}^{k',s,\eta}(Y) = \widetilde{M}_{k',m'} := \mathrm{med}(\widetilde{M}_{m'}^{(1)}, \dots, \widetilde{M}_{m'}^{(k')}).$$

The overall cost is $n(\omega, Y) = n_1 + n_2 = k\,m + k'\ \max\{\lceil \eta\,(\widehat{R}_{k,m}(\omega))^s \rceil,\ 1\}$.

The main task for the subsequent sections is to determine appropriate values for the parameters k, k', m, s, and η, such that the algorithm $A_{k,p,m}^{k',s,\eta}$ is (ε, δ)-approximating on the class $\mathcal{Y}_{p,q,K}$, that is,

$$\mathbb{P}\left\{ |A_{k,p,m}^{k',s,\eta}(Y) - \mathbb{E}\,Y| \le \varepsilon \right\} \ge 1 - \delta$$

for all $Y \in \mathcal{Y}_{p,q,K}$ and $\varepsilon, \delta > 0$. Hence the problem (2) is solvable on $\mathcal{Y}_{p,q,K}$, see Theorem 2.15. The second task is then to control the expected cost of the algorithm for any fixed input Y, and this is where k might need to be larger than k', see Lemma 2.16. We are not confined to estimating the p-moment for $Y \in \mathcal{Y}_{p,q,K}$, indeed, any $\widetilde{p} \in [1, q)$ is possible. We provide all the tools to understand these variants and briefly comment on possible strategies, see Remark 2.18. For simplicity, however, we focus on an algorithm $A_{\varepsilon,\delta}^{p,q,K}$ that does the job via an estimate of the first central moment (i.e. $\widetilde{p} = 1$), though the cost bound involves probably suboptimal K-dependent constants for the second stage then.

The described algorithm uses a method of 'probability amplification', namely the 'median trick', see Alon et al. [1] and Jerrum et al. [17]. The intention is to build an estimator $\widehat{\theta}$ for an unknown quantity $\theta \in \mathbb{R}$ such that with high probability the outcome lies within a certain interval $I \subset \mathbb{R}$. This interval depends on θ and represents the deviation of $\widehat{\theta}$ from θ that we are willing to tolerate. The following proposition is a minor modification of [26, Proposition 2.1, see also (2.6)] from Niemiro and Pokarowski.

Proposition 2.2 (*Median Trick*). *Let $(\widehat{\theta}_\ell)_{\ell \in \mathbb{N}}$ be a sequence of real-valued i.i.d. random variables, and let $I \subset \mathbb{R}$ be an interval such that*

$$\mathbb{P}\{\widehat{\theta}_\ell \in I\} \ge 1 - \alpha,$$

where $0 < \alpha < 1/2$. For an odd natural number k, define the median of k independent estimators, $\widehat{\theta} := \mathrm{med}(\widehat{\theta}_1, \dots, \widehat{\theta}_k)$. Then

$$\mathbb{P}\{\widehat{\theta} \in I\} \ge 1 - \frac{1}{2}(4\alpha(1-\alpha))^{k/2}.$$

Remark 2.3. Already Nemirovsky and Yudin 1983 [25, Proposition on p. 244] suggest a similar probability amplification scheme. Recently, Devroye et al. [9] and Huber [16] worked on improvements of the median-of-means. Finally, let us mention that several proposals for a generalization of the median-of-means for regression problems have been published, see [14,18,19,21–24].

2.2. Some interpolation results

The condition that defines the cone $\mathcal{Y}_{p,q,K}$ has implications on the relation between other central absolute moments for exponents $1 \le \widetilde{p} < \widetilde{q} \le q$.

Lemma 2.4. *Let $Z \in L_q$ be a random variable and $1 \le p < r < q \le \infty$ where $\frac{1}{r} = \frac{1-\lambda}{p} + \frac{\lambda}{q}$ for some $0 < \lambda < 1$. Then*

$$\|Z\|_p \le \|Z\|_r \le \|Z\|_p^{1-\lambda} \|Z\|_q^{\lambda}.$$

Proof. The first inequality is well-known as a consequence of Jensen's inequality. For the second inequality we use Hölder's inequality, without loss of generality $Z \geq 0$, and

$$\|Z\|_r^r = \mathbb{E}\left[Z^{(1-\lambda)r}\, Z^{\lambda r}\right] \leq \|Z^{(1-\lambda)r}\|_{\frac{p}{(1-\lambda)r}}\, \|Z^{\lambda r}\|_{\frac{q}{\lambda r}} = \|Z\|_p^{(1-\lambda)r} \|Z\|_q^{\lambda r},$$

finishes the proof. □

As an immediate consequence we obtain the following inclusion properties:

Corollary 2.5. *Let $1 \leq p < r < q \leq \infty$ where $\frac{1}{r} = \frac{1-\lambda}{p} + \frac{\lambda}{q}$. Then*

1. $\mathcal{Y}_{p,q,K} \subset \mathcal{Y}_{r,q,K} \subset \mathcal{Y}_{p,q,K'}$ *with* $K' := K^{\frac{1}{1-\lambda}} = K^{\frac{r(q-p)}{p(q-r)}} > K$,
2. $\mathcal{Y}_{p,q,K} \subset \mathcal{Y}_{p,r,K'}$ *with* $K' := K^{\lambda} = K^{\frac{q(r-p)}{r(q-p)}} < K$.

In particular,

- *we have* $\mathcal{Y}_{p,q,K} \subset \mathcal{Y}_{1,q,K'}$ *with* $K' := K^{\frac{p(q-1)}{q-p}}$ *(that is $K' = K^p$ for $q = \infty$),*
- *for $q \geq 2$ we have* $\mathcal{Y}_{p,q,K} \subset \mathcal{Y}_{1,2,K'}$ *with* $K' := K^{\frac{pq}{2(q-p)}}$ *(that is $K' = K^{p/2}$ in the case $q = \infty$).*

Proof. Apply Lemma 2.4 to $Z := Y - \mathbb{E}\, Y$. □

Remark 2.6. A special case of the above results shows how the cone that has been considered in Hickernell et al. [13] is embedded in a larger cone of L_2-functions, we have $\mathcal{Y}_{2,4,K} \subset \mathcal{Y}_{1,2,K^2}$, such that it is sufficient to provide algorithms that work for the larger cone.

We use an extension of a special case of the Marcinkiewicz–Zygmund inequality, see [32] and [2, Theorem 2] as well as [27, Sect. 2.2.8, Proposition 3]. We provide a proof of this inequality by using the same technique as in [11, Proposition 5.4] and [33]. This leads to slightly better constants than those in the literature.

Lemma 2.7. *Let $1 < q \leq 4$. Then for any mean zero random variable $Z \in L_q$ and independent copies $Z_1, \ldots, Z_m$ of Z, that is, $Z_i \sim Z$ for $i = 1, \ldots, m$, we have*

$$\left\| \frac{1}{m} \sum_{i=1}^{m} Z_i \right\|_q \leq \begin{cases} 2^{2/q-1}\, m^{-(1-1/q)}\, \|Z\|_q & \text{if } 1 < q \leq 2, \\ (2\sqrt{3})^{1-2/q}\, m^{-1/2}\, \|Z\|_q & \text{if } 2 < q \leq 4. \end{cases}$$

Proof. We apply the Riesz–Thorin interpolation theorem to an operator TP, where

$$P : L_q([0,1]) \to L_q([0,1]), \quad f \mapsto f - \mathrm{INT}(f)$$

projects onto the subspace of mean zero functions,

$$L_q^0([0,1]) := \{f \in L_q([0,1]) \mid \mathrm{INT}(f) = 0\},$$

and

$$T : L_q([0,1]) \to L_q([0,1]^m), \quad f(x) \mapsto \frac{1}{m} \sum_{i=1}^{m} f(x_i).$$

Note that random variables $Z \in L_q$ correspond to functions $f \in L_q([0,1])$ and mean zero random variables correspond to functions from the space $L_q^0([0,1])$. We are interested in the quantity

$$M_q := \sup_{\substack{\|f\|_q \leq 1 \\ \mathrm{INT}(f)=0}} \|Tf\|_q = \|T\|_{L_q^0 \to L_q} \leq \sup_{\|f\|_q \leq 1} \|TPf\|_q = \|TP\|_{L_q \to L_q}.$$

Let $1 \leq p < r < q \leq \infty$ with $\frac{1}{r} = \frac{1-\lambda}{p} + \frac{\lambda}{q}$, where $0 < \lambda < 1$. Then the Riesz–Thorin theorem states

$$\|TP\|_{L_r \to L_r} \leq \|TP\|_{L_p \to L_p}^{1-\lambda} \|TP\|_{L_q \to L_q}^{\lambda}, \tag{7}$$

see [4, Chapter 4: Corollary 1.8] for details and a more general formulation. Knowing the special cases

$$\|P\|_{L_q\to L_q} = \begin{cases} 2 & \text{if } q \in \{1,\infty\}, \\ 1 & \text{if } q = 2, \end{cases}$$

we can apply the Riesz–Thorin theorem also to P (similarly to (7)) and obtain

$$\|P\|_{L_q\to L_q} \le 2^{|2/q-1|} \qquad \text{for } 1 \le q \le \infty\,. \tag{8}$$

It is an immediate consequence of the triangle inequality that

$$M_1 = \|T\|_{L_1^0\to L_1} = 1\,. \tag{9}$$

There is a direct way to estimate M_q for even $q \in \mathbb{N}$. We switch back to random variables. Let X, X_i for $i = 1,\dots,m$ be i.i.d. random variables uniformly distributed in $[0,1]$ so that $Z_i := f(X_i) \sim f(X)$ are i.i.d. mean zero random variables for $f \in L_q^0([0,1])$. Then

$$\mathbb{E}\left|\frac{1}{m}\sum_{i=1}^m Z_i\right|^q = m^{-q} \sum_{\mathbf{i}\in\{1,\dots,m\}^q} \mathbb{E}\prod_{k=1}^q Z_{i_k}\,. \tag{10}$$

Observe that $\mathbb{E}\prod_{k=1}^q Z_{i_k}$ vanishes if there is at least one $i \in \{1,\dots,m\}$ that occurs only once in $\mathbf{i} = (i_1,\dots,i_q)$. (This is due to independence of the Z_i and $\mathbb{E}\,Z_i = 0$.) For all the other terms a generalized version of Hölder's inequality yields

$$\mathbb{E}\prod_{k=1}^q Z_{i_k} \le \|Z\|_q^q.$$

It remains to count the terms that do not vanish, for the two relevant special cases we obtain:

$q = 2$: There are m terms of the shape Z_i^2.

$q = 4$: There are m terms of the shape Z_i^4, and $3\,m(m-1)$ terms of the shape $Z_i^2 Z_j^2$ with distinct indices $i \ne j$. Altogether these are less than $3\,m^2$ terms.

Plugging this into (10) and taking the qth root we have

$$M_2 = m^{-1/2}\,, \qquad \text{and} \qquad M_4 \le 3^{1/4}\,m^{-1/2}\,. \tag{11}$$

We interpolate the operator norm of TP between the special values for which by (8) together with (9) and (11) we have

$$\|TP\|_{L_1\to L_1} = 2\,, \quad \|TP\|_{L_2\to L_2} = m^{-1/2}\,, \quad \text{and} \quad \|TP\|_{L_4\to L_4} \le \sqrt{2}\,3^{1/4}\,m^{-1/2}.$$

Hence, by the Riesz–Thorin theorem as stated in (7), for $p = 1 < r < q = 2$ with $\lambda = 1 - 2/r$, and for $p = 2 < r < q = 4$ with $\lambda = 2 - 4/r$, we obtain the desired upper bounds on M_q for $1 < q \le 4$. □

2.3. Estimating central absolute moments

We consider estimators for the (central absolute) p-moment $\varrho^p := \mathbb{E}\,|Y - \mathbb{E}\,Y|^p$ of a random variable $Y \in \mathcal{Y}_{p,q,K}$, based on m independent copies $Y_1,\dots,Y_m$ of Y and their residuals,

$$\widehat{R}_m := \frac{1}{m}\sum_{i=1}^m |Y_i - \widehat{M}_m|^p\,, \qquad \text{where} \quad \widehat{M}_m := \frac{1}{m}\sum_{i=1}^m Y_i\,. \tag{12}$$

The first moment and the corresponding estimator, that is for $p = 1$, is also known as 'mean absolute deviation', 'average absolute deviation', or 'mean deviation from the mean'. This and other *measures of scale* like 'median absolute deviation from the median' constitute alternatives to the widely known standard deviation as an empirical quantification of the dispersion of a random variable.

Lemma 2.8. *Let $Y \in \mathcal{Y}_{p,q,K}$ and $\varrho := \|Y - \mathbb{E}\,Y\|_p$.*

(a) If $1 = p < q \le 2$, then

$$\|\widehat{R}_m - \varrho\|_q \;\le\; \frac{2^{2/q-1}\,(1+2K)}{m^{1-1/q}}\,\varrho \;\le\; \frac{6\,K}{m^{1-1/q}}\,\varrho.$$

(b) Let $1 \le p \le 2$, and $p < q \le 2p$. Then, with $r := q/p$, we have

$$\|\widehat{R}_m - \varrho^p\|_r \le \frac{26\,K^p}{m^{1-1/r}}\,\varrho^p.$$

Proof. To shorten the notation, define $a := \mathbb{E}\,Y$.

We start with proving (a). Via the triangle inequality, we separate the inaccuracy of the empirical mean $\widehat{M}_m$ from the deviation of a single observation Y_i,

$$|Y_i - a| - |\widehat{M}_m - a| \;\le\; |Y_i - \widehat{M}_m| \;\le\; |Y_i - a| + |\widehat{M}_m - a|.$$

Hence, for the estimator (12) we can write

$$\widehat{R}_m = \underbrace{\frac{1}{m}\sum_{i=1}^{m} |Y_i - a|}_{=:\widehat{\varrho}_m} + r_m$$

with bounded remainder $|r_m| \le |\widehat{M}_m - a|$. For the L_q-norm this gives

$$\|\widehat{R}_m - \varrho\|_q \;\le\; \|\widehat{\varrho}_m - \varrho\|_q + \|r_m\|_q \;\le\; \|\widehat{\varrho}_m - \varrho\|_q + \|\widehat{M}_m - a\|_q\,.$$

By Lemma 2.7 we conclude

$$\begin{aligned}\|\widehat{R}_m - \varrho\|_q \;&\le\; \frac{2^{2/q-1}}{m^{1-1/q}}\,\big(\|\,|Y-a| - \varrho\|_q + \|Y-a\|_q\big)\\ &\le\; \frac{2^{2/q-1}}{m^{1-1/q}}\,\big(\varrho + 2\,\|Y-a\|_q\big) \;\le\; \frac{2^{2/q-1}\,(1+2K)}{m^{1-1/q}}\,\varrho\,.\end{aligned}$$

This finishes the proof of (a).

We turn to (b). First note that, for $x, y \in \mathbb{R}$ and $p > 1$, we have

$$|x|^p - p\,|x|^{p-1}|y| \;\le\; |x-y|^p \;\le\; |x|^p + p\,|2x|^{p-1}|y| + |2y|^p\,, \tag{13}$$

where the second inequality can be proven by distinguishing the cases $|y| < |x|$ and $|y| \le |x|$ with subcases $|x| \le |x-y|$ and $|x| > |x-y|$. We use (13) with $x = Y_i - a$ and $y = \widehat{M}_m - a$. Hence

$$\widehat{R}_m \;=\; \underbrace{\frac{1}{m}\sum_{i=1}^{m} |Y_i - a|^p}_{=:\widehat{\varrho}_m^{\,p}} + r_m,$$

with the remainder r_m bounded by

$$|r_m| \;\le\; \frac{2^{p-1}\,p}{m}\sum_{i=1}^{m} |Y_i - a|^{p-1}\,|\widehat{M}_m - a| + 2^p\,|\widehat{M}_m - a|^p.$$

By the triangle inequality we obtain the norm estimate

$$\begin{aligned}\|\widehat{R}_m - \varrho^p\|_r \;&\le\; \|\widehat{\varrho}_m^{\,p} - \varrho^p\|_r + \|r_m\|_r\\ &\le\; \|\widehat{\varrho}_m^{\,p} - \varrho^p\|_r + \frac{2^{p-1}\,p}{m}\sum_{i=1}^{m} \big\|\,|Y_i - a|^{p-1}\,|\widehat{M}_m - a|\big\|_r + 2^p\,\big\||\widehat{M}_m - a|^p\big\|_r\,.\end{aligned}$$

We aim to control the norm of the product term. For L_q-integrable non-negative random variables X, Z, from Hölder's inequality, and recalling $pr = q$, we have

$$\mathbb{E}\, X^{r(p-1)}\, Z^r \;\le\; \|X^{r(p-1)}\|_{p/(p-1)}\, \|Z^r\|_p \;\le\; \|X\|_q^{r(p-1)}\, \|Z\|_q^r.$$

Taking $X = |Y_i - a|$ and $Z = |\widehat{M}_m - a|$, this leads to

$$\|\widehat{R}_m - \varrho^p\|_r \;\le\; \|\widehat{\varrho}_m^p - \varrho^p\|_r + 2^{p-1}\, p\, \|Y-a\|_q^{p-1}\, \|\widehat{M}_m - a\|_q + 2^p\, \|\widehat{M}_m - a\|_q^p\,.$$

Finally, Lemma 2.7 gives

$$\begin{aligned}
&\|\widehat{R}_m - \varrho^p\|_r \\
&\le\; 2^{2/r-1}\, m^{-(1-1/r)}\, \||Y-a|^p - \varrho^p\|_r \\
&+\; \|Y-a\|_q^p \begin{cases} 2^{p+2/q-2}\, p\, m^{-(1-1/q)} + 2^{2/r}\, m^{-p(1-1/q)} & \text{if } q \le 2, \\ 2^{p-1}\, p\, (2\sqrt{3})^{1-2/q}\, m^{-1/2} + 2^p\, (2\sqrt{3})^{p(1-2/q)}\, m^{-p/2} & \text{if } 2 < q \le 4, \end{cases} \\
&\le\; (2\,\varrho^p + 24\, \|Y-a\|_q^p)\, m^{-(1-1/r)} \;\le\; \frac{26\,K^p}{m^{1-1/r}}\, \varrho^p\,,
\end{aligned}$$

which finishes the proof. □

Combining Markov's inequality with the above result, we obtain probabilistic bounds for the error of the empirical p-moment estimator.

Lemma 2.9. *Let $Y \in \mathcal{Y}_{p,q,K}$ with $\varrho := \|Y - \mathbb{E}\,Y\|_p$, and let $0 < \alpha < 1$.*

(a) If $1 = p < q \le 2$, then

$$\mathbb{P}\left\{ |\widehat{R}_m - \varrho| \le \frac{2^{2/q-1}\,(1+2K)}{\alpha^{1/q}\, m^{1-1/q}}\, \varrho \right\} \;\ge\; 1 - \alpha.$$

(b) In the case $1 \le p \le 2$, and $p < q \le 2p$, we have

$$\mathbb{P}\left\{ |\widehat{R}_m - \varrho^p| \le \frac{26\,K^p}{\alpha^{p/q}\, m^{1-p/q}}\, \varrho^p \right\} \;\ge\; 1 - \alpha.$$

From this, via Proposition 2.2, we derive the following probabilistic guarantees for the modified median-of-empirical-moments estimator $\widehat{R}_{k,m}$ from Algorithm 2.1 which exhibit enhanced confidence levels.

Proposition 2.10. *Let $Y \in \mathcal{Y}_{p,q,K}$ with $\varrho := \|Y - \mathbb{E}\,Y\|_p$, and let $0 < \alpha < 1/2$.*

(a) If $1 = p < q \le 2$, then for the first moment estimator from Algorithm 2.1 we have

$$\mathbb{P}\left\{ |\widehat{R}_{k,m} - \varrho| \le \frac{2^{2/q-1}\,(1+2K)}{\alpha^{1/q}\, m^{1-1/q}}\, \varrho \right\} \;\ge\; 1 - \frac{1}{2}(4\alpha(1-\alpha))^{k/2}.$$

(b) If $1 \le p \le 2$ and $p < q \le 2p$, then for the p-moment estimator from Algorithm 2.1 we have

$$\mathbb{P}\left\{ |\widehat{R}_{k,m} - \varrho^p| \le \frac{26\,K^p}{\alpha^{p/q}\, m^{1-p/q}}\, \varrho^p \right\} \;\ge\; 1 - \frac{1}{2}(4\alpha(1-\alpha))^{k/2}.$$

Remark 2.11 (*Choice of the Parameters for the First Stage*)**.** We need to control how badly we underestimate the p-moment of $Y \in \mathcal{Y}_{p,q,K}$ by the estimator $\widehat{R}_{k,m}$. We consider either $1 = p < q \le 2$ or $1 \le p \le 2$ and $p < q \le 2p$. For an overall confidence level $1 - \delta$ of the error of the algorithm we need a p-moment estimate that satisfies

$$\mathbb{P}\left\{ \widehat{R}_{k,m} \ge (1-\gamma)\,\varrho^p \right\} \ge 1 - \delta_1 \tag{14}$$

for some $0 < \gamma < 1$ and $0 < \delta_1 < \delta$. This situation can be established by the use of Proposition 2.10:

- Fix some $0 < \alpha_1 < 1/2$.
- Choose an odd $k \geq 2\log(2\delta_1)^{-1}/\log(4\alpha_1(1-\alpha_1))^{-1}$, in order to guarantee the correct confidence level. For $\alpha_1 = 1/4$ and $\delta_1 = \delta/2$, this simplifies to $k \geq 2\log\delta^{-1}/\log\frac{4}{3}$.
- Choose

$$m \geq \begin{cases} \frac{2^{1/(q-1)-1}\,(1+2K)^{1+1/(q-1)}}{\gamma^{1+1/(q-1)}\,\alpha_1^{1/(q-1)}} & \text{if } p = 1 < q \leq 2, \\ \frac{26^{1+p/(q-p)}\,K^{pq/(q-p)}}{\gamma^{1+p/(q-p)}\,\alpha_1^{p/(q-p)}} & \text{if } 1 < p \leq 2 \text{ and } p < q \leq 2p, \end{cases}$$

  in order to describe the correct event. With $\alpha_1 = 1/4$ and $\gamma = 1/2$, and with further simplifications, it is sufficient to choose

$$m \geq \begin{cases} 3 \cdot 48^{1/(q-1)}\,K^{1+1/(q-1)} & \text{if } p = 1 < q \leq 2, \\ 52 \cdot 208^{p/(q-p)}\,K^{pq/(q-p)} & \text{if } 1 < p \leq 2 \text{ and } p < q \leq 2p. \end{cases}$$

The cost for estimating the p-moment is determined by the product $n_1 = k\,m$. One can optimize the choice of α_1 numerically in order to minimize this product. Such optimization depends on the ratio q/p. The optimal α_1 approaches $1/2$ as the ratio q/p approaches 1. However, the choice $\alpha_1 = 1/4$ is sufficient. The parameters γ and δ_1 affect the relative precision and the uncertainty at which we estimate the p-moment. This has consequences for the sample size needed in the second stage of the algorithm, see Section 2.4.

2.4. Determining the sample size

Given $1 \leq p < q \leq \infty$, we put $\widetilde{q} := \min\{q, 2\}$. For $Y \in \mathcal{Y}_{p,q,K}$ we define $\tau := \|Y - \mathbb{E}\,Y\|_{\widetilde{q}}$, which is the norm of interest when determining the required sample size for estimating the mean. We start with considerations for mean estimators assuming that τ is known. Then, fixing $m' \in \mathbb{N}$, for

$$\widehat{M}_{m'} := \frac{1}{m'}\sum_{i=1}^{m'} Y_i\,, \qquad Y_i \overset{\text{iid}}{\sim} Y\,,$$

Markov's inequality and Lemma 2.7 immediately imply the following.

Lemma 2.12. *Let $Y \in L_{\widetilde{q}}$ with $1 < \widetilde{q} \leq 2$ and $\tau := \|Y - \mathbb{E}\,Y\|_{\widetilde{q}}$. Then we have for $0 < \alpha < 1$ that*

$$\mathbb{P}\left\{|\widehat{M}_{m'} - \mathbb{E}\,Y| \leq \frac{2^{2/\widetilde{q}-1}\,\tau}{\alpha^{1/\widetilde{q}}\,(m')^{1-1/\widetilde{q}}}\right\} \geq 1 - \alpha.$$

As a consequence, for the median-of-means estimator

$$\widehat{M}_{k',m'} := \operatorname{med}(\widehat{M}_{m'}^{(1)}, \ldots, \widehat{M}_{m'}^{(k')}) \qquad \text{where} \quad \widehat{M}_{m'}^{(\ell)} := \frac{1}{m'}\sum_{i=(\ell-1)m'+1}^{\ell m'} Y_i,$$

with fixed odd number k', via Proposition 2.2 we obtain probabilistic error bounds with enhanced confidence level. A similar result for the median-of-means estimator for $Y \in L_q$ with $1 < q < 2$ is obtained by Bubeck et al. in [5, Lemma 2]. The constants we get turn out to be slightly better.

Proposition 2.13. *Let $Y \in L_{\widetilde{q}}$ with $1 < \widetilde{q} \leq 2$ and $\tau := \|Y - \mathbb{E}\,Y\|_{\widetilde{q}}$. Then we have for $0 < \alpha < 1/2$ that*

$$\mathbb{P}\left\{|\widehat{M}_{k',m'} - \mathbb{E}\,Y| \leq \frac{2^{2/\widetilde{q}-1}\,\tau}{\alpha^{1/\widetilde{q}}\,(m')^{1-1/\widetilde{q}}}\right\} \geq 1 - \frac{1}{2}(4\alpha(1-\alpha))^{k'/2}.$$

If τ is known, then any choice

$$m' \geq \frac{2^{1/(\widetilde{q}-1)-1}}{\alpha^{1/(\widetilde{q}-1)}}\left(\frac{\tau}{\varepsilon}\right)^{1+1/(\widetilde{q}-1)} \tag{15}$$

would give probabilistic bounds for an absolute error $\varepsilon > 0$, where the confidence level is determined by α and k'. Unfortunately, we do not know τ but rely on estimates of the central L_p-norm $\varrho = \|Y - \mathbb{E}\,Y\|_p$ for $Y \in \mathcal{Y}_{p,q,K}$. In the case $p = \widetilde{q} = 2$, this is directly the norm of interest $\varrho = \tau$. If however $1 \le p < 2$, we exploit the cone condition

$$\tau = \|Y - \mathbb{E}\,Y\|_{\widetilde{q}} \le K' \|Y - \mathbb{E}\,Y\|_p = K'\varrho \quad \text{where } K' := \begin{cases} K & \text{if } q \le 2, \\ K^{\frac{q(2-p)}{2(q-p)}} & \text{if } q > 2, \end{cases}$$

see Corollary 2.5 for the inclusion $\mathcal{Y}_{p,q,K} \subset \mathcal{Y}_{p,2,K'}$ in the case $q > 2 = \widetilde{q}$. Remark 2.11 provides parameter settings for the p-moment estimator $\widehat{R}_{k,m}$ such that

$$\mathbb{P}\Big\{\varrho^p \le \widehat{R}_{k,m}/(1-\gamma)\Big\} \ge 1 - \delta_1.$$

If for the median-of-means estimator $\widetilde{M}_{k',m'}$ of Algorithm 2.1 we have that

$$\mathbb{P}\left(|\widetilde{M}_{k',m'} - \mathbb{E}\,Y| \le \varepsilon \;\|\; \varrho^p \le \widehat{R}_{k,m}/(1-\gamma)\right) \ge 1 - \delta_2\,, \tag{16}$$

with $\delta_1 + \delta_2 = \delta$, then

$$\mathbb{P}\{|A^{k',s,\eta}_{k,p,m}(Y) - \mathbb{E}\,Y| \le \varepsilon\} \ge (1-\delta_1)(1-\delta_2) \ge 1 - (\delta_1 + \delta_2) = 1 - \delta.$$

Hoping that $\varrho^p \le \widehat{R}_{k,m}/(1-\gamma)$ holds indeed, the parameter m' in $\widetilde{M}_{k',m'}$ can be chosen as in (15), just replacing τ by

$$K' \cdot (\widehat{R}_{k,m}/(1-\gamma))^{1/p} \ge \tau\,. \tag{17}$$

Hence, m' is not a fixed number anymore but depends on the p-moment estimate. Still, since the conditional distribution of $\widetilde{M}_{k',m'}$ after the p-moment estimation is equal to the distribution of $\widetilde{M}_{k',m'}$ with the respective fixed m', Proposition 2.13 may be used for the conditional case in order to obtain (16). The consequential choice of the parameters for the second stage of Algorithm 2.1 is summarized in the following remark.

Remark 2.14 (*Choice of the Parameters for the Second Stage*)**.** Let $Y \in \mathcal{Y}_{p,q,K}$, with $1 \le p \le 2$ and $p < q \le 2p$. From Remark 2.11 we know how to choose m and k within the first stage of $A^{k',s,\eta}_{k,p,m}$, see Algorithm 2.1. Now, with Proposition 2.13 and the considerations (15)–(17), we can choose s, k' and η depending on $\varepsilon > 0$, uncertainty $0 < \delta_2 < \delta$ and p, q as follows:

- Fix some $0 < \alpha_2 < 1/2$.
- Choose an odd number $k' \ge 2\log(2\delta_2)^{-1}/\log(4\alpha_2(1-\alpha_2))^{-1}$, in order to guarantee the conditional confidence level. For $\alpha_2 = 1/4$ and $\delta_2 = \delta/2$, this simplifies to $k' \ge 2\log\delta^{-1}/\log\frac{4}{3}$.
- Choose m' according to (15) and (17) with α_2 instead of α. For $\alpha_2 = 1/4$ and $\gamma = 1/2$, it is sufficient to choose

$$s := \tfrac{1}{p}\max\left\{1 + \tfrac{1}{q-1}, 2\right\}$$
$$\eta := \begin{cases} 16^{1/(q-1)}\, K^{1+1/(q-1)}\, \varepsilon^{-(1+1/(q-1))} & \text{for } q \le 2, \\ 16\, K^{q(2-p)/(q-p)}\, \varepsilon^{-2} & \text{for } 2 \le q. \end{cases}$$

This gives algorithms $A^{k',s,\eta}_{k,p,m}$ for approximating the mean $\mathbb{E}\,Y$ of random variables $Y \in \mathcal{Y}_{p,q,K}$ from cones with very specific parameter ranges, namely $1 \le p \le 2$ and $p < q \le 2p$. These methods are also suitable for general cones $\mathcal{Y}_{p,q,K}$ since by Corollary 2.5 they are contained in specific cones, in particular $\mathcal{Y}_{p,q,K} \subset \mathcal{Y}_{1,\widetilde{q},K'}$ with $\widetilde{q} := \min\{q, 2\}$ and some K'. In other words, it is always feasible to rely on a first moment estimate. The following theorem summarizes the solvability result:

Theorem 2.15. *The problem of computing expected values (2) is solvable for all input classes $\mathcal{Y}_{p,q,K}$, $1 \le p < q \le \infty$, see (4). In detail, for any $\varepsilon, \delta > 0$, the algorithm $A^{k',s,\eta}_{k,1,m}$ (see Algorithm 2.1), with odd natural numbers k, $k' \ge 2\frac{\log\delta^{-1}}{\log(4/3)}$, and*

- *for* $1 \le p < q \le 2$ *with*

$$m := \lceil 3 \cdot 48^{1/(q-1)} K^{pq/(q-p)} \rceil ,$$
$$s := 1 + \tfrac{1}{q-1} ,$$
$$\eta := 16^{1/(q-1)} K^{pq/(q-p)} \varepsilon^{-(1+1/(q-1))} ;$$

- *for* $1 \le p < q$ *and* $2 \le q$, *with*

$$m := \lceil 144 K^{pq/(q-p)} \rceil ,$$
$$s := 2 ,$$
$$\eta := 16 K^{pq/(q-p)} \varepsilon^{-2} ,$$

is (ε, δ)*-approximating on* $\mathcal{Y}_{p,q,K}$.

2.5. Estimates on the cost

Obviously, the cost $n(\omega, Y)$ of the considered algorithms is a random variable. The quantity of interest is the expected cost

$$\bar{n}(A^{k',s,\eta}_{k,p,m}, Y) := \mathbb{E}\, n(\,\cdot\,, Y) \le km + k'(1 + \eta\ \mathbb{E}\, \widehat{R}^s_{k,m}).$$

The following lemma is essential, since it provides an upper bound on $\mathbb{E}\, \widehat{R}^s_{k,m}$.

Lemma 2.16. *Let* $Y \in \mathcal{Y}_{p,q,K}$ *with* $1 \le p \le 2$ *and* $p < q \le 2p$. *Let* $\varrho = \|Y - \mathbb{E}\, Y\|_p$ *and* $s = \frac{1}{p} \max\{1 + 1/(q-1), 2\}$. *For* $0 < \gamma < 1$ *and* $\alpha_1 = 1/4$ *choose* m *such that (via Lemma 2.9) for the p-moment estimators* $\widehat{R}^{(\ell)}_m$ *from Algorithm 2.1 we have*

$$\mathbb{P}\left\{ |\widehat{R}^{(\ell)}_m - \varrho^p| \le \gamma \varrho^p \right\} \ge 1 - \alpha_1 . \tag{18}$$

Assume that $k \ge 4ps/q$, *with* k *being odd. Then the median of p-moment estimators,* $\widehat{R}_{k,m}$, *satisfies*

$$\mathbb{E}\, \widehat{R}^s_{k,m} \le (1 + 3\gamma)^s \varrho^{ps}.$$

Proof. Lemma 2.9 and Proposition 2.10 are stated for $0 < \alpha < 1/2$, so our assumption (18) implies

$$\mathbb{P}\{|\widehat{R}_{k,m} - \varrho^p| \le c\, \alpha^{-p/q} \varrho^p\} \ge 1 - \tfrac{1}{2}(4\alpha(1-\alpha))^{k/2},$$

where $c = \alpha_1^{p/q} \gamma = 4^{-p/q} \gamma$. We use the simplified estimate

$$\mathbb{P}\left\{ |\widehat{R}_{k,m} - \varrho^p|^s > c^s \alpha^{-ps/q} \varrho^{ps} \right\} \le 2^{k-1} \alpha^{k/2}.$$

Suppose $\mathbb{E}\, \widehat{R}^s_{k,m}$ is finite, then by the triangle inequality applied to the $\|\cdot\|_s$-norm, we obtain

$$\mathbb{E}\, \widehat{R}^s_{k,m} \le \left(\varrho^p + \|\widehat{R}_{k,m} - \varrho^p\|_s\right)^s .$$

The norm therein can be estimated by

$$\mathbb{E}\, |\widehat{R}_{k,m} - \varrho^p|^s \le t_0 + \int_{t_0}^\infty \mathbb{P}\{|\widehat{R}_{k,m} - \varrho^p|^s > t\}\, \mathrm{d}t.$$

This leads to $\alpha = c^{q/p} \varrho^q\, t^{-q/(ps)}$. We choose $t_0 = \alpha_1^{-ps/q} c^s \varrho^{ps} = 4^{ps/q} c^s \varrho^{ps}$, and we use

$$\int_{t_0}^\infty t^{-kq/(2ps)}\, \mathrm{d}t = \left(\frac{kq}{2ps} - 1\right)^{-1} t_0^{-kq/(2ps)+1},$$

which holds for $kq/(2ps) > 1$, that is $k > 2ps/q$, only. Then

$$\int_{t_0}^\infty \mathbb{P}\{|\widehat{R}_{k,m} - \varrho^p|^s > t\}\, \mathrm{d}t$$

$$\leq 2^{k-1}\, c^{kq/(2p)}\, \varrho^{kq/2} \left(\frac{kq}{2ps} - 1\right)^{-1} (4^{ps/q}\, c^s\, \varrho^{ps})^{-kq/(2ps)+1}$$
$$= \left(\frac{kq}{2ps} - 1\right)^{-1} 2^{2ps/q-1}\, c^s\, \varrho^{ps}\,.$$

Altogether we obtain

$$\mathbb{E}\,\widehat{R}_{k,m}^s \leq \left(1 + \left(1 + \left(\frac{kq}{2ps} - 1\right)^{-1} \frac{1}{2}\right)^{1/s} 2^{2p/q}\, c\right)^s \varrho^{ps}\,.$$

Taking $k \geq 4ps/q$, and with $c \leq 2^{-p/q}\,\gamma$ as well as $p/q < 1$, this further simplifies to $\mathbb{E}\,\widehat{R}_{k,m}^s \leq \left(1 + 2\,(3/2)^{1/s}\,\gamma\right)^s \varrho^{ps}$. We have $s \geq 1$ such that $1/s \leq 1$, which leads to the assertion. □

With the additional assumption on the median trick that $k \geq 4ps/q$, we can estimate the expected cost for fixed input. For the particular setting from Theorem 2.15, where we use $\mathcal{Y}_{p,q,K} \subset \mathcal{Y}_{1,\widetilde{q},K'}$ with $\widetilde{q} := \min\{q, 2\}$, we get the following upper bound on the cost.

Theorem 2.17. *For computing $\mathbb{E}\,Y$ on $\mathcal{Y}_{p,q,K}$ with $1 \leq p < q \leq \infty$, see (4), consider $A_{k,1,m}^{k',s,\eta}$, see Algorithm 2.1. For any $\varepsilon > 0$ and $0 < \delta < 1/2$ choose m, s, and η as in Theorem 2.15, where k and k' are the least odd natural numbers such that $k, k' \geq 2\log\delta^{-1}/\log\frac{4}{3}$, and in addition $k \geq \max\{4/(q-1), 4\}$. Then, $A_{p,q,K}^{\varepsilon,\delta} := A_{k,1,m}^{k',s,\eta}$ is (ε, δ)-approximating on $\mathcal{Y}_{p,q,K}$ and*

$$\bar{n}(A_{p,q,K}^{\varepsilon,\delta}, Y) \leq C_q\, K^{pq/(q-p)} \left(1 + \left(\frac{\|Y - \mathbb{E}\,Y\|_1}{\varepsilon}\right)^{\max\{1+1/(q-1),2\}}\right) \log\delta^{-1}$$

where $Y \in \mathcal{Y}_{p,q,K}$, and $C_q > 0$ is a q-dependent constant.

Note that in the above result, $K > 1$ and $\delta < 1/2$ ensure that additive constants can be absorbed by C_q and are not lost by writing $K^{pq/(q-p)}$ and $\log\delta^{-1}$ as factors.

Remark 2.18 (*Balancing the Cost Between the Two Stages*)**.** There is some arbitrariness about how much effort we put on each of the two stages for $Y \in \mathcal{Y}_{p,q,K}$. The more accurately we estimate $\|Y - \mathbb{E}\,Y\|_1$, that is with relative accuracy γ tending to 0, the more accurately we can choose the sample size n_2 for the second stage. In return, being parsimonious in the first stage, which means $\gamma \to 1$, we likely overshoot the sample size for the second stage. We thus suggest the following compromise: Choose $\gamma \in (0, 1)$ such that the cost of each of the two stages exceeds the cost of the respective limiting cases by roughly the same factor.

The algorithm for which we give cost bounds in Theorem 2.17 uses an estimate for the first moment. With this strategy, for any input set $\mathcal{Y}_{p,q,K}$, the cost for the first stage possesses the optimal dependence on K, compare the lower bounds in Theorem 3.7. But since actually a higher moment is of interest for the sample size selection in the second stage, namely $\mathbb{E}\,|Y - \mathbb{E}\,Y|^{\widetilde{q}}$, with $\widetilde{q} := \min\{q, 2\}$, compare Proposition 2.13, the inflation parameter η has to be proportional to $K^{pq/(q-p)}$. For some instances $Y \in \mathcal{Y}_{p,q,K}$ with a far smaller ratio between the different moments we thus actually overshoot the sample size by this factor. It might be hence beneficial for the adaption to estimate higher moments within the first stage:

- For $\mathcal{Y}_{p,q,K}$ with $1 \leq p < q \leq 2$, estimate the p-moment $\mathbb{E}\,|Y - \mathbb{E}\,Y|^p$. The cost for the first stage has still the optimal dependence on K, the inflation factor η for the second stage is only proportional to $K^{q/(q-1)}$.
- For input sets $\mathcal{Y}_{p,q,K}$ with $2 < q$, estimate the variance. That way we directly estimate the moment that is decisive for the sample size in stage two. The cost for stage one, though, will have a suboptimal dependence on K if $p < 2$.

Remark 2.19. Hickernell et al. consider in [13] the expectation problem on $\mathcal{Y}_{2,4,K}$. They provide an (ε, δ)-approximating two-stage algorithm on this special class of inputs and that way prove solvability. The main difference to our Algorithm 2.1 is that they do not use the median trick, which is tantamount to the parameter choice $k = k' = 1$. The cost of their algorithm then depends polynomially on the uncertainty δ instead of the optimal logarithmic dependence $\log \delta^{-1}$ which we have.

For the special setting $\mathcal{Y}_{2,4,K}$, the constants of our complexity bounds can be slightly improved. In detail, there is an unbiased empirical variance estimator

$$\widehat{V}_m := \frac{1}{m-1} \sum_{i=1}^{m} (Y_i - \widehat{M}_m)^2, \qquad m \geq 2,$$

which shall replace $\widehat{R}_m$, see (12), and its variance can be precisely estimated since Y possesses a fourth moment. Then, Lemma 1 in Hickernell et al. [13] can be used instead of Lemma 2.9 in our paper for the confidence level α_1 before probability amplification. In combination with the median trick this enables us to construct algorithms with more favorable cost bounds.

3. Lower bounds

3.1. Algorithms, solvability, and complexity

We start with a formal definition of algorithms that are admissible for the computation of $\mathbb{E}\, Y$ (or other quantities associated to a random variable Y, e.g. different moments), as well as the larger class of algorithms admissible for numerical integration of a function f. Recall that $(\Omega, \Sigma, \mathbb{P})$ is a probability space. In the context of algorithms that take random variables as input, let $\mathcal{Y}$ be a subset of all real-valued random variables on $(\Omega, \Sigma, \mathbb{P})$. For any input $Y \in \mathcal{Y}$ we assume that we have a sequence of i.i.d. copies $Y_1, Y_2, \ldots \sim Y$ defined on $(\Omega, \Sigma, \mathbb{P})$. The algorithms may depend on an additional random variable U with values in a measurable space $(\mathbb{U}, \Sigma_U)$ which influences how the i.i.d. sample is used. This random variable is assumed to be independent of the i.i.d. sample $(Y_i)_{i\in\mathbb{N}}$. In contrast, for algorithms that take functions as input, the randomness is purely created by the algorithm design, whereas the input class $\mathcal{F} = \mathcal{F}(\pi)$ is a subset of all measurable functions $f\colon G \to \mathbb{R}$ on the probability space (G, Σ_G, π) and algorithms may interrogate function values $f(\mathbf{x})$ at any point $\mathbf{x} \in G$.

Definition 3.1 (*Abstract Algorithms*)**.** Assume that we have

- an auxiliary random variable U with values in a measurable space $(\mathbb{U}, \Sigma_U)$;
- a measurable output function $\phi\colon \mathbb{U} \times \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$; and
- a measurable function $n\colon \mathbb{U} \times \mathbb{R}^{\mathbb{N}} \to \mathbb{N}_0 \cup \{\infty\}$ (called *stopping rule*) for which the value of $\mathbb{1}[n(u, (y_j)_{j\in\mathbb{N}}) \geq i]$ only depends on $(u, y_1, \ldots, y_{i-1})$.

Based on this we define two types of algorithms:

1. An *i.i.d.-based algorithm* A on $\mathcal{Y}$ is a mapping $A : \Omega \times \mathcal{Y} \to \mathbb{R}$ given by
$$A(\omega, Y) := \phi(U(\omega), Y_1(\omega), \ldots, Y_{n(\omega,Y)}(\omega), 0, 0, \ldots),$$
where $(Y_i)_{i\in\mathbb{N}}$ is a sequence of i.i.d. copies of $Y \in \mathcal{Y}$ and independent of U; further, the *stopping time*
$$n(\omega, Y) := n(U(\omega), (Y_i(\omega))_{i\in\mathbb{N}})$$
is assumed to be $\mathbb{P}$-almost surely finite for all $Y \in \mathcal{Y}$.
2. A *general randomized algorithm* Q on $\mathcal{F}$ is a mapping $Q : \Omega \times \mathcal{F} \to \mathbb{R}$ given by
$$Q(\omega, f) := \phi(U(\omega), N(\omega, f))$$
with a measurable information map $N : \Omega \times \mathcal{F} \to \mathbb{R}^{\mathbb{N}}$, $N(\omega, f) = (y_i)_{i\in\mathbb{N}}$, satisfying
$$y_i = \begin{cases} f(\mathbf{X}_i(\omega, y_1, \ldots, y_{i-1})) & \text{for } n(U(\omega), (y_j)_{j\in\mathbb{N}}) \geq i, \\ 0 & \text{for } n(U(\omega), (y_j)_{j\in\mathbb{N}}) < i, \end{cases}$$

for a sequence $(\mathbf{X}_i)_{i\in\mathbb{N}}$ of measurable functions $\mathbf{X}_i : \Omega \times \mathbb{R}^{i-1} \to G$ (called *sampling nodes*)[1]; further, the *stopping time* (also called *cardinality function*)

$$n(\omega, f) := n(U(\omega), N(\omega, f))$$

is assumed to be $\mathbb{P}$-almost surely finite for all $f \in \mathcal{F}$.

Thus, the application of an algorithm to an input, that is $A(Y) := A(\cdot, Y)$ or $Q(f) := Q(\cdot, f)$, respectively, is a random variable on $(\Omega, \Sigma, \mathbb{P})$.

Every i.i.d.-based algorithm provides a general randomized algorithm: For the integration problem (1) on a class $\mathcal{F} = \mathcal{F}(\pi)$ of π-measurable real-valued functions, let $\mathbf{X}$ be a π-distributed random variable and define the input class $\mathcal{Y} := \{f(\mathbf{X}) \mid f \in \mathcal{F}\}$. For every $f \in \mathcal{F}$ and an i.i.d. sequence of π-distributed sampling nodes $(\mathbf{X}_i)_{i\in\mathbb{N}}$ independent of U, observe that $(f(\mathbf{X}_i))_{i\in\mathbb{N}}$ is an i.i.d. sequence of copies of $f(\mathbf{X}) \in \mathcal{Y}$, thus applicable within any i.i.d.-based algorithm A. That way, $A(\omega, f(\mathbf{X}))$ can be interpreted as a general randomized algorithm $Q(\omega, f)$ with restricted type of information. Lower bound proofs are less technical for the smaller class of i.i.d.-based algorithms, see Sections 3.2 and 3.3, but can be extended to general integration methods, see Section 3.4. Let us mention here that we deviate from the "standard" technique of proving lower bounds for general randomized algorithms in IBC via an average case setting, see for example [29, Lemma 4.37]. Such a direct approach for proving lower bounds for the integration with "small error at high confidence" is developed in [20] with application to Sobolev classes.

Before proceeding with our results, we formally introduce the concepts of solvability and complexity for the expectation problem (2) on a class $\mathcal{Y}$ and, analogously, for the integration problem (1) on a class $\mathcal{F}$.

Definition 3.2 (*Solvability*)**.** The problem (2) of computing an expected value is *solvable* on a class of random variables $\mathcal{Y}$, iff for any $\varepsilon > 0$, and $\delta \in (0, 1)$, there is an i.i.d.-based algorithm A such that

$$\mathbb{P}\{|A(Y) - \mathbb{E}\, Y| \le \varepsilon\} \ge 1 - \delta,$$

for all $Y \in \mathcal{Y}$. In that case A is called (ε, δ)-*approximating (on* $\mathcal{Y}$*)*.

The integration problem (1) is *solvable* on $\mathcal{F}$ iff for any $\varepsilon > 0$, and $\delta \in (0, 1)$, there is a general randomized algorithm Q such that

$$\mathbb{P}\{|Q(f) - \mathrm{INT}(f)| \le \varepsilon\} \ge 1 - \delta,$$

for all $f \in \mathcal{F}$. Here Q is also called (ε, δ)-*approximating (on* $\mathcal{F}$*)*.

Remark 3.3 (*Other Notions of Solvability*)**.** If a problem is even solvable in the sense that for any error threshold $\varepsilon > 0$ we can find an algorithm A that guarantees a bound for the *expected error* $\mathbb{E}\,|A(Y) - \mathbb{E}\, Y| \le \varepsilon$ (or the *mean squared error* $\mathbb{E}\,|A(Y) - \mathbb{E}\, Y|^2 \le \varepsilon^2$), using Markov's inequality we can also guarantee solvability in the probabilistic sense of the definition above. The reverse statement is *not* true. In fact, Theorem 3.5 shows that the notions differ for certain problems.

Besides solvability of the problem, for an (ε, δ)-approximating algorithm A (resp. Q) with $Y \in \mathcal{Y}$ (resp. $f \in \mathcal{F}$) we are interested in the cost, given by the sample size $n(\omega, Y)$ (resp. $n(\omega, f)$). Note that $n(\cdot, Y)$ is a random variable that depends on Y, for example, it might depend on the p-moment of Y. For the complexity analysis we aim to control the *expected cardinality*,

$$\bar{n}(A, Y) := \mathbb{E}\, n(\cdot, Y) \qquad (\text{resp. } \bar{n}(Q, f) := \mathbb{E}\, n(\cdot, f)), \tag{19}$$

so that running the algorithm multiple times for several random variables of comparable difficulty does not take an uncontrollably long computing time.

[1] The y_i are well defined since by the definition of a stopping rule the outcome of the statement $n(U(\omega), (y_j)_{j\in\mathbb{N}}) \ge i$ depends only on $U(\omega)$ and $y_1, \ldots, y_{i-1}$.

For classical input sets $\mathcal{Y}$ (resp. $\mathcal{F}$) such as L_q-balls, one usually considers the worst case cost

$$\bar{n}^{\mathrm{wor}}(A,\mathcal{Y}) := \sup_{Y\in\mathcal{Y}} \bar{n}(A,Y) \qquad (\text{resp. } \bar{n}^{\mathrm{wor}}(Q,\mathcal{F}) := \sup_{f\in\mathcal{F}} \bar{n}(Q,f)). \tag{20}$$

For $\varepsilon,\delta>0$ the *minimal worst case cost* (or simply the *worst case complexity* of the problem) is given by

$$\bar{n}^{\mathrm{wor}}(\varepsilon,\delta,\mathcal{Y}) := \inf\{n>0 \mid \exists\ (\varepsilon,\delta)\text{-approximating } A \text{ with } \bar{n}^{\mathrm{wor}}(A,\mathcal{Y})\le n\},$$
$$(\text{resp. } \bar{n}^{\mathrm{wor}}(\varepsilon,\delta,\mathcal{F}) := \inf\{n>0 \mid \exists\ (\varepsilon,\delta)\text{-approximating } Q \text{ with } \bar{n}^{\mathrm{wor}}(Q,\mathcal{F})\le n\}).$$

However, for the cone-shaped input sets $\mathcal{Y}=\mathcal{Y}_{p,q,K}$ we are interested in, this quantity is infinite which can be seen via an appropriate lower bound proof, compare Theorems 3.9 and 3.10 and consider the limit of intersections of $\mathcal{Y}_{p,q,K}$ with bigger and bigger L_q-balls. Hence, we need an alternative concept of complexity. We take two perspectives here.

First, we consider the *fixed cost* of an i.i.d.-based algorithm A (resp. a general randomized algorithm Q), that is,

$$\bar{n}^{\mathrm{fix}}(A,\mathcal{Y}) := \inf_{Y\in\mathcal{Y}} \bar{n}(A,Y), \qquad (\text{resp. } \bar{n}^{\mathrm{fix}}(Q,\mathcal{F}) := \inf_{f\in\mathcal{F}} \bar{n}(Q,f)). \tag{21}$$

For $\mathcal{Y}=\mathcal{Y}_{p,q,K}$ and Algorithm 2.1 this is essentially the cost for estimating the p-moment. Given any problem, we define its *minimal fixed cost* by

$$\bar{n}^{\mathrm{fix}}(\varepsilon,\delta,\mathcal{Y}) := \inf\{n>0 \mid \exists\ (\varepsilon,\delta)\text{-approximating } A \text{ with } \bar{n}^{\mathrm{fix}}(A,\mathcal{Y})\le n\},$$
$$(\text{resp. } \bar{n}^{\mathrm{fix}}(\varepsilon,\delta,\mathcal{F}) := \inf\{n>0 \mid \exists\ (\varepsilon,\delta)\text{-approximating } Q \text{ with } \bar{n}^{\mathrm{fix}}(Q,\mathcal{F})\le n\}),$$

which is the fixed cost that *all* (ε,δ)-approximating algorithms exhibit, not only those that start with a p-moment estimation, see Theorem 3.7 for $\mathcal{Y}=\mathcal{Y}_{p,q,K}$.

Second, we intersect the cone shaped input sets $\mathcal{Y}_{p,q,K}$ with (semi-)balls of radius $\tau>0$ with respect to the centered $L_{\widetilde{q}}$-norm, $\widetilde{q}:=\min\{q,2\}$, that is,

$$\tau\mathcal{B}_{\widetilde{q}} := \{Y\in L_{\widetilde{q}} \mid \|Y-\mathbb{E}\,Y\|_{\widetilde{q}}\le\tau\},$$
$$(\text{resp. } \tau\mathcal{B}_{\widetilde{q}} := \{f\in L_{\widetilde{q}} \mid \|f-\mathrm{INT}(f)\|_{\widetilde{q}}\le\tau\},)$$

and consider the worst case complexity,

$$\bar{n}^{\mathrm{wor}}(\varepsilon,\delta,\mathcal{Y}_{p,q,K}\cap\tau\mathcal{B}_{\widetilde{q}}), \qquad (\text{resp. } \bar{n}^{\mathrm{wor}}(\varepsilon,\delta,\mathcal{F}_{p,q,K}\cap\tau\mathcal{B}_{\widetilde{q}}),)$$

which is the worst case cost of optimal algorithms that are adapted to this reduced problem. With $\|Y-\mathbb{E}\,Y\|_1\le\|Y-\mathbb{E}\,Y\|_{\widetilde{q}}\le\tau$, our particular algorithm $A^{p,q,K}_{\varepsilon,\delta}$, see Theorem 2.17, provides an upper bound for this quantity. Lower bounds are given in Theorems 3.9 and 3.10.

3.2. Unsolvability results and fixed cost bounds

We start with the statement that higher integrability alone is not a sufficient assumption, the corresponding problem is unsolvable.

Theorem 3.4 (*Unsolvability for the Whole Space L_q*)**.** *The problem (2) of computing the expectation via i.i.d.-based algorithms is unsolvable for the class $\mathcal{Y}_q=L_q$ of q-integrable random variables, where $q\in[1,\infty]$.*

Proof. Let A be an arbitrary i.i.d.-based algorithm for $\mathcal{Y}_q$, further let $\varepsilon>0$ and $\delta\in(0,1)$. By definition, the i.i.d.-based algorithm A terminates almost surely for any input, in particular, for the zero random variable $0\in\mathcal{Y}_q$. Hence, for the stopping time $n(\omega,0)$ of A with input 0 we have

$$\mathbb{P}\{n(\cdot\,,0)>n_0\}\xrightarrow[n_0\to\infty]{}0.$$

Consider Bernoulli random variables Z with probabilities $\mathbb{P}\{Z=0\}=1-p$ and $\mathbb{P}\{Z=1/p\}=p\in(0,1)$, hence $\mathbb{E}\,Z=1$. For any $a>0$ we have $aZ\in\mathcal{Y}_q$ since it is a discrete random variable. We choose $a>2\varepsilon$ such that $|A(Y)-a|\le\varepsilon$ and $|A(Y)|\le\varepsilon$ are disjoint events. For i.i.d. copies $Z_1,Z_2,\ldots\sim Z$ we have

$$\mathbb{P}\{Z_1=\cdots=Z_{n_0}=0\}=(1-p)^{n_0}\xrightarrow[p\to 0]{}1.$$

Now, consider the failing probability for a random variable aZ,

$$\begin{aligned}
&\mathbb{P}\{|A(aZ)-a|>\varepsilon\}\\
&\quad\ge\ \mathbb{P}\{n(\,\cdot\,,aZ)\le n_0 \text{ and } Z_1=\cdots=Z_{n_0}=0\}\\
&\qquad\cdot\mathbb{P}(|A(aZ)-a|>\varepsilon\mid n(\,\cdot\,,aZ)\le n_0 \text{ and } Z_1=\cdots=Z_{n_0}=0)\\
&\quad\ge\ \mathbb{P}\{n(\,\cdot\,,0)\le n_0\}\cdot\mathbb{P}\{Z_1=\cdots=Z_{n_0}=0\}\\
&\qquad\cdot\mathbb{P}(|A(0)|\le\varepsilon\mid n(\,\cdot\,,0)\le n_0)\\
&\quad=\ \mathbb{P}\{|A(0)|\le\varepsilon \text{ and } n(\,\cdot\,,0)\le n_0\}\cdot\mathbb{P}\{Z_1=\cdots=Z_{n_0}=0\}\\
&\quad\ge\ (\mathbb{P}\{|A(0)|\le\varepsilon\}-\mathbb{P}\{n(\,\cdot\,,0)>n_0\})\cdot(1-p)^{n_0}\\
&\quad\xrightarrow[p\to 0]{}\ \mathbb{P}\{|A(0)|\le\varepsilon\}-\mathbb{P}\{n(\,\cdot\,,0)>n_0\}\\
&\quad\xrightarrow[n_0\to\infty]{}\ \mathbb{P}\{|A(0)|\le\varepsilon\}\,.
\end{aligned}$$

Hence, for any admitted uncertainty $\delta\in(0,1/2)$, if $\mathbb{P}\{|A(0)|\le\varepsilon\}\ge 1-\delta>1/2$, there are random variables of type aZ for which $\mathbb{P}\{|A(aZ)-a|>\varepsilon\}>1/2>\delta$. This shows that the algorithm cannot be (ε,δ)-approximating. □

We now turn to more restricted input classes. First, we stress that for the cone-shaped input classes $\mathcal{Y}_{p,q,K}$ the worst case *expected error* is unbounded for any algorithm. (A fortiori, the same does hold for the mean squared error.)

Theorem 3.5 (*Impossibility of Finite Worst Case Expected Error*)**.** *For any i.i.d.-based algorithm A on $\mathcal{Y}_{p,q,K}$, with $1\le p<q\le\infty$ and $K\ge 1$, we have*

$$\sup_{Y\in\mathcal{Y}_{p,q,K}}\mathbb{E}\,|A(Y)-\mathbb{E}\,Y|=\infty.$$

Proof. Consider a random variable Z with $\mathbb{P}\{Z=0\}=\mathbb{P}\{Z=2\}=1/2$ and observe that $\mathbb{E}\,Z=1$ as well as $\|Z-1\|_p=1$ for all $1\le p\le\infty$. Consequently, for any $a\in\mathbb{R}$ we have $aZ\in\mathcal{Y}_{p,q,K}$ for all $1\le p<q\le\infty$ and $K\ge 1$. We reduce the problem to distinguishing aZ from the zero random variable $0\in\mathcal{Y}_{p,q,K}$,

$$\begin{aligned}
\sup_{Y\in\mathcal{Y}_{p,q,K}}\mathbb{E}\,|A(Y)-\mathbb{E}\,Y|&\ge\max\{\mathbb{E}\,|A(0)|,\mathbb{E}\,|A(aZ)-a|\}\\
&\ge\frac{1}{2}\,\mathbb{E}\,|A(0)-A(aZ)+a|\,.
\end{aligned}$$

By definition, the i.i.d.-based algorithm A terminates almost surely for any input, in particular, for the zero random variable $0\in\mathcal{Y}_{p,q,K}$. Hence, for the stopping time $n(\omega,0)$ of A with input 0, there exists an $n_0\in\mathbb{N}$ such that $\mathbb{P}\{n(\,\cdot\,,0)\le n_0\}>0$. Furthermore,

$$\begin{aligned}
\mathbb{E}\,|A(0)-A(aZ)+a|&\ge\ |a|\cdot\mathbb{P}\{A(0)=A(aZ)\}\\
&\ge\ |a|\cdot\mathbb{P}\left\{n(\,\cdot\,,0)\le n_0\ \text{ and }\ Z_1=\cdots=Z_{n_0}=0\right\}\\
&=\ |a|\cdot\mathbb{P}\{n(\,\cdot\,,0)\le n_0\}\cdot\mathbb{P}\{Z_1=\cdots=Z_{n_0}=0\}.
\end{aligned}$$

Finally, the assertion follows by the fact that $\mathbb{P}\{Z_1=\cdots=Z_{n_0}=0\}=2^{-n_0}$ and

$$\sup_{Y\in\mathcal{Y}_{p,q,K}}\mathbb{E}\,|A(Y)-\mathbb{E}\,Y|\ \ge\ \frac{|a|}{2^{n_0+1}}\,\mathbb{P}\{n(\,\cdot\,,0)\le n_0\}\xrightarrow[|a|\to\infty]{}\infty.\quad\square$$

The previous theorem tells us that certain standard worst case error notions such as the expected absolute error or the mean squared error are not suitable for cone-shaped input classes $\mathcal{Y}_{p,q,K}$. This is the reason for us to focus on the concept of "small error with high probability". We start with an auxiliary result.

Lemma 3.6. *For a Bernoulli random variable Y with $\mathbb{P}\{Y=0\} = 1-a$ and $\mathbb{P}\{Y=1\} = a \in (0, 1/2]$ satisfying*

$$a \geq K^{-pq/(q-p)} \qquad \text{or} \qquad a = \tfrac{1}{2}\,, \tag{22}$$

we have $Y \in \mathcal{Y}_{p,q,K}$ for $1 \leq p < q \leq \infty$ and $K \geq 1$.

Proof. Clearly, $\mathbb{E}\, Y = a$, such that we need to study

$$\kappa_{p,q}(a) := \frac{\|Y-a\|_q}{\|Y-a\|_p}\,, \qquad a \in (0, 1/2].$$

The case $a = 1/2$ is explained by $\kappa_{p,q}(1/2) = 1$. For $a \in (0, 1/2]$ we have

$$\kappa_{p,\infty}(a) = \frac{1-a}{\big((1-a)a^p + a(1-a)^p\big)^{1/p}} \leq a^{-1/p}.$$

For $p < q < \infty$ we interpolate the L_q-norm, see Lemma 2.4,

$$\kappa_{p,q}(a) \leq \big(\kappa_{p,p}(a)\big)^{1-\lambda} \big(\kappa_{p,\infty}(a)\big)^{\lambda},$$

where $\lambda = (q-p)/q$, and $\kappa_{p,p}(a) = 1$ by definition. Hence $\kappa_{p,q}(a) \leq a^{-(q-p)/(pq)}$. By inversion we obtain the sufficient condition on a such that $\kappa_{p,q}(a) \leq K$, which is the cone condition $Y \in \mathcal{Y}_{p,q,K}$. □

The next theorem shows that – up to a q-dependent constant – the fixed cost for the first stage of Algorithm 2.1 is really needed, compare the term in Theorem 2.17 that is independent of the error threshold ε.

Theorem 3.7 (*Fixed Cost*). *For any $\varepsilon > 0$ and $\delta \in (0, 1)$, we have the following fixed cost lower bound for i.i.d.-based algorithms on $\mathcal{Y}_{p,q,K}$ with $1 \leq p < q \leq \infty$ and $K > 1$,*

$$\bar{n}^{\,\mathrm{fix}}(\varepsilon, \delta, \mathcal{Y}_{p,q,K}) \;\geq\; \frac{1}{2\log 2}\, K^{pq/(q-p)}\, \log \delta^{-1}.$$

Proof. Let $Y \in \mathcal{Y}_{p,q,K}$ and A be an (ε, δ)-approximating i.i.d.-based sampling algorithm with stopping time $n(\omega, Y)$. To shorten the notation let $\bar{n} := \bar{n}(A, Y)$ be the expected cardinality for Y.

Consider a Bernoulli random variable Z with probabilities $\mathbb{P}\{Z=1\} = a$ and $\mathbb{P}\{Z=0\} = 1-a$, where $a = \mathbb{E}\, Z \in (0, 1/2]$ is chosen such that

$$\|Z - \mathbb{E}\, Z\|_q < K\, \|Z - \mathbb{E}\, Z\|_p.$$

For $c > 0$ let $Y' := Y + c\, Z$. Observe that for sufficiently large c we have

$$\frac{\|Y' - \mathbb{E}\, Y'\|_q}{\|Y' - \mathbb{E}\, Y'\|_p} \overset{\Delta\text{-ineq.}}{\leq} \frac{c\, \|Z - \mathbb{E}\, Z\|_q + \|Y - \mathbb{E}\, Y\|_q}{c\, \|Z - \mathbb{E}\, Z\|_p - \|Y - \mathbb{E}\, Y\|_p} \xrightarrow[c\to\infty]{} \frac{\|Z - \mathbb{E}\, Z\|_q}{\|Z - \mathbb{E}\, Z\|_p} < K,$$

hence, there is a number $c_0 > 0$ such that $Y' \in \mathcal{Y}_{p,q,K}$ for $c > c_0$.

There is a certain chance that A cannot distinguish Y from Y', namely, if $Z_i = 0$ for all generated i.i.d. samples $Y_i' = Y_i + c\, Z_i$. Via Jensen's inequality and by exploiting the fact that the function $x \mapsto (1-a)^x$ is convex, we have

$$\begin{aligned}
\mathbb{P}\{Z_1 = \cdots = Z_{n(\,\cdot\,,Y)} = 0\} &= \sum_{n=0}^{\infty} \mathbb{P}\{n(\,\cdot\,, Y) = n \;\text{ and }\; Z_1 = \cdots = Z_n = 0\} \\
&= \sum_{n=0}^{\infty} \mathbb{P}\{n(\,\cdot\,, Y) = n\}\, (1-a)^n \geq (1-a)^{\bar{n}}\,.
\end{aligned}$$

The algorithm A returns only finite values, thus, for any $\theta \in (0, (1-a)^{\bar{n}})$ we find a value $b \in \mathbb{R}$ such that

$$\mathbb{P}\{A(Y) \le b \ \text{ and } \ Z_1 = \cdots = Z_{n(\cdot,Y)} = 0\} \ge (1-a)^{\bar{n}} - \theta.$$

We want to compute $\mathbb{E}\, Y' = \mathbb{E}\, Y + c\, a$, which can be arbitrarily large when $c \to \infty$. Now, choose c such that $\mathbb{E}\, Y' - b > \varepsilon$ for an arbitrary error threshold $\varepsilon > 0$. Then

$$\begin{aligned}\mathbb{P}\left\{|A(Y') - \mathbb{E}\, Y'| > \varepsilon\right\} &\ge \mathbb{P}\{A(Y) \le b \ \text{ and } \ Z_1 = \cdots = Z_{n(\cdot,Y)} = 0\} \\ &\ge (1-a)^{\bar{n}} - \theta.\end{aligned}$$

Letting $\theta \to 0$ leads to $\mathbb{P}\left\{|A(Y') - \mathbb{E}\, Y'| > \varepsilon\right\} \ge (1-a)^{\bar{n}}$ such that the uncertainty δ is at least $(1-a)^{\bar{n}}$. In other words,

$$\bar{n} \ \ge \ \frac{\log \delta^{-1}}{\log(1-a)^{-1}} \ \ge \ \frac{1}{2\log 2}\, a^{-1} \ \log \delta^{-1}.$$

Here, we used $\log(1-a)^{-1} \le 2(\log 2)\, a$ for $a \in (0, 1/2]$. In particular with $a = 1/2$ this is

$$\bar{n} \ \ge \ \frac{\log \delta^{-1}}{\log 2}\,. \tag{23}$$

For $K \ge 2^{(q-p)/(pq)}$ we put $a = K^{-pq/(q-p)}$, see Lemma 3.6, which gives

$$\bar{n} \ \ge \ \frac{1}{2\log 2}\, K^{pq/(q-p)} \ \log \delta^{-1}\,. \tag{24}$$

Trivially, this holds true for smaller $K > 1$ as well, since in these cases inequality (24) is weaker than (23). □

3.3. Worst case complexity for bounded central norm

In this section we provide lower bounds of the worst case complexity for i.i.d.-based algorithms on cone-shaped input classes intersected with (semi-)balls. We start with an auxiliary result due to Wald, see [37, Theorem on p. 156], which provides a lower bound for the complexity of statistical tests.

Proposition 3.8 (*Wald 1945*). *Let $Y^{(1)}$ and $Y^{(2)}$ be two discrete random variables that only take values in a finite set $D \subset \mathbb{R}$, and consider the probability ratio as a function*

$$r(y) := \frac{\mathbb{P}\{Y^{(1)} = y\}}{\mathbb{P}\{Y^{(2)} = y\}} \qquad \text{for } y \in D.$$

Let T be an i.i.d.-based algorithm satisfying for $0 < \delta < 1/2$ that

$$\mathbb{P}\{T(Y^{(i)}) \ne i\} \le \delta\,, \qquad \text{for } i = 1,2.$$

(T is a statistical test for hypotheses 1 vs. 2 with confidence $1-\delta$.) Then

$$\bar{n}(T, Y^{(i)}) \ge \frac{(1-2\delta)\log\frac{1-\delta}{\delta}}{|\mathbb{E}\log r(Y^{(i)})|}, \qquad \text{for } i = 1,2.$$

We start with lower bounds for cones of random variables with bounded variance, that is $\mathcal{Y}_{p,q,K}$ with $q \ge 2$. We consider the worst case complexity when $\mathcal{Y}_{p,q,K}$ is intersected with a ball of radius $\sigma > 0$ with respect to the *centered* L_2-norm (i.e. the standard deviation, which is a semi-norm),

$$\sigma \mathcal{B}_2 := \{Y \in L_2 \mid \|Y - \mathbb{E}\, Y\|_2 \le \sigma\}.$$

In this case finding worst case lower bounds reduces to distinguishing two Bernoulli random variables with antithetic probability of success.

Theorem 3.9 (*Worst Case Complexity for Bounded Variance*). *Let $1 \le p < q \le \infty$ with $q \ge 2$ and let $K > 1$ as well as $\sigma > 0$. Then*

$$\bar{n}^{\mathrm{wor}}(\varepsilon, \delta, \mathcal{Y}_{p,q,K} \cap \sigma \mathcal{B}_2) \ge \frac{1}{4 \log 3} \left(\frac{\sigma}{\varepsilon}\right)^2 \log\left(\tfrac{4}{3}\delta\right)^{-1}$$

for any $0 < \varepsilon \le \min\{1 - 2/(K+1),\ 1/2\}\,\sigma$, and $0 < \delta \le 1/4$.

Proof. We consider shifted and scaled Bernoulli random variables $Y^{(1)}, Y^{(2)}$ that only take values $\pm\sigma$ with positive probability. For $i = 1,2$, obviously, $Y^{(i)} \in \sigma \mathcal{B}_2$ since

$$\|Y^{(i)} - \mathbb{E}\, Y^{(i)}\|_2 \le \|Y^{(i)}\|_2 = \sigma\,.$$

Further, by the triangle inequality we have

$$\frac{\|Y^{(i)} - \mathbb{E}\, Y^{(i)}\|_q}{\|Y^{(i)} - \mathbb{E}\, Y^{(i)}\|_p} \le \frac{\|Y^{(i)}\|_q + |\mathbb{E}\, Y^{(i)}|}{\|Y^{(i)}\|_p - |\mathbb{E}\, Y^{(i)}|} = \frac{2}{1 - |\mathbb{E}\, Y^{(i)}|/\sigma} - 1.$$

Hence, $|\mathbb{E}\, Y^{(i)}| \le (1 - 2/(K+1))\,\sigma$ ensures $Y^{(i)} \in \mathcal{Y}_{p,q,K}$. Introducing a parameter $\alpha \in (0,\ 1 - 2/(K+1)]$, we further specify $Y^{(1)}$ and $Y^{(2)}$ by

$$\mathbb{P}\{Y^{(1)} = +\sigma\} = \mathbb{P}\{Y^{(2)} = -\sigma\} = \frac{1+\alpha}{2}\,,$$
$$\mathbb{P}\{Y^{(1)} = -\sigma\} = \mathbb{P}\{Y^{(2)} = +\sigma\} = \frac{1-\alpha}{2}\,.$$

Hence

$$\mathbb{E}\, Y^{(1)} = -\mathbb{E}\, Y^{(2)} = \alpha\,\sigma\,.$$

Now let $\varepsilon \in (0, \alpha\sigma)$ and note that from any (ε, δ)-approximating algorithm A on $\mathcal{Y}_{p,q,K} \cap \sigma \mathcal{B}_2$ we can build a statistical test T distinguishing $Y^{(1)}$ from $Y^{(2)}$ by

$$T(Y^{(i)}) := \begin{cases} 1 & A(Y^{(i)}) > 0\,, \\ 2 & A(Y^{(i)}) \le 0\,. \end{cases}$$

Obviously, the expected cardinality of A and T is the same and the failure probability does not exceed $\delta < \frac{1}{2}$. Then, from Proposition 3.8 we obtain

$$\bar{n}(A, Y^{(i)}) \ge \frac{(1 - 2\delta) \log \frac{1-\delta}{\delta}}{\alpha \log \frac{1+\alpha}{1-\alpha}}.$$

Restricting to $0 < \alpha \le 1/2$ we may exploit convexity of $\log \frac{1+\alpha}{1-\alpha}$, replacing it by $2(\log 3)\,\alpha$. We also restrict to $0 < \delta \le 1/4$ and obtain

$$\bar{n}(A, Y^{(i)}) \ge \frac{1}{4 \log 3}\, \alpha^{-2} \log\left(\tfrac{4}{3}\delta\right)^{-1}.$$

The limit $\alpha \to \varepsilon/\sigma$ and continuity in ε gives the desired complexity bound for the range of error thresholds $0 < \varepsilon \le \min\{1 - 2/(K+1),\ 1/2\}\,\sigma$. □

For classes $\mathcal{Y}_{p,q,K}$ with less integrability, $1 \le p < q \le 2$, we intersect the cone with a ball of radius $\tau > 0$ with respect to the centered L_q-norm,

$$\tau \mathcal{B}_q := \{Y \in L_q \mid \|Y - \mathbb{E}\, Y\|_q \le \tau\},$$

and we obtain the following worst case lower bound.

Theorem 3.10 (*Worst Case Complexity for Bounded Central L_q-norm, $1 < q \le 2$*). *Let $1 \le p < q \le 2$ and $K > 1$ as well as $\tau > 0$. Then there is a constant $c_{q,K} > 0$, depending only on K and q, such that*

$$\bar{n}^{\mathrm{wor}}(\varepsilon, \delta, \mathcal{Y}_{p,q,K} \cap \tau \mathcal{B}_q) \ge c_{q,K} \left(\frac{\tau}{\varepsilon}\right)^{1+1/(q-1)} \log\left(\tfrac{4}{3}\delta\right)^{-1}$$

for any $0 < \varepsilon \le \frac{1}{6}(1 - 1/K)\,\tau$ and $0 < \delta \le 1/4$.

Proof. We consider random variables $Y^{(1)}, Y^{(2)}$ with

$$\begin{aligned}\mathbb{P}\{Y^{(i)} = \tau'\} \;=\; \mathbb{P}\{Y^{(i)} = -\tau'\} \;&=\; \frac{1-\beta}{2}\,,\\ \mathbb{P}\{Y^{(i)} = 0\} \;&=\; \beta\,(1-\gamma)\,,\\ \mathbb{P}\{|Y^{(i)}| = \gamma^{-1/q}\,\tau'\} \;&=\; \beta\,\gamma\,,\end{aligned}$$

for $i = 1,2$, where $0 < \tau' < \tau$ and $0 < \beta, \gamma < 1$. Now we choose β and τ such that $Y^{(i)} \in \mathcal{Y}_{p,q,K} \cap \tau\mathcal{B}_q$. Observe that

$$\begin{aligned}&|\mathbb{E}\,Y^{(i)}| \le \beta\,\gamma^{1-1/q}\,\tau' \le \beta\,\tau',\\ &(1-\beta)\,\tau' \le (1-\beta)^{1/p}\,\tau' \le \|Y^{(i)}\|_p \le \|Y^{(i)}\|_q = \tau'.\end{aligned}$$

Using the triangle inequality, for $0 < \beta < 1/2$ we obtain

$$\frac{\|Y^{(i)} - \mathbb{E}\,Y^{(i)}\|_q}{\|Y^{(i)} - \mathbb{E}\,Y^{(i)}\|_p} \;\le\; \frac{\|Y^{(i)}\|_q + |\mathbb{E}\,Y^{(i)}|}{\|Y^{(i)}\|_p - |\mathbb{E}\,Y^{(i)}|} \;\le\; \frac{1+\beta}{1-2\beta},$$

such that if we put

$$\beta := \frac{1}{2}\left(1 - \frac{3}{1+2K}\right),$$

this ensures $Y^{(i)} \in \mathcal{Y}_{p,q,K}$. Furthermore,

$$\|Y^{(i)} - \mathbb{E}\,Y^{(i)}\|_q \;\le\; \|Y^{(i)}\|_q + |\mathbb{E}\,Y^{(i)}| \;\le\; (1+\beta)\,\tau',$$

so choosing $\tau' := \tau/(1+\beta)$ guarantees $Y^{(i)} \in \tau\mathcal{B}_q$. We specify $Y^{(1)}$ and $Y^{(2)}$ by

$$\begin{aligned}\mathbb{P}\{Y^{(1)} = \gamma^{-1/q}\tau'\} \;=\; \mathbb{P}\{Y^{(2)} = -\gamma^{-1/q}\tau'\} \;&=\; \frac{3}{4}\,\beta\,\gamma\,,\\ \mathbb{P}\{Y^{(1)} = -\gamma^{-1/q}\tau'\} \;=\; \mathbb{P}\{Y^{(2)} = \gamma^{-1/q}\tau'\} \;&=\; \frac{1}{4}\,\beta\,\gamma\,.\end{aligned}$$

Their respective expected values are then given by

$$\mathbb{E}\,Y^{(1)} \;=\; -\,\mathbb{E}\,Y^{(2)} \;=\; \frac{1}{2}\,\beta\,\gamma^{1-1/q}\,\tau'.$$

In order to guarantee that $\frac{1}{2}\beta\,\gamma^{1-1/q}\,\tau' > \varepsilon$, we choose $\gamma > \left(2\varepsilon/(\beta\tau')\right)^{1+1/(q-1)}$. Since we need $\gamma < 1$, this is only possible for $\varepsilon \in (0, \frac{1}{2}\beta\tau')$. As in the proof of Theorem 3.9, cost bounds for (ε,δ)-approximating algorithms A on $\mathcal{Y}_{p,q,K} \cap \tau\mathcal{B}_q$ are derived from the cost bounds of Proposition 3.8 for distinguishing $Y^{(1)}$ from $Y^{(2)}$,

$$\bar{n}(A, Y^{(1)}) \;\ge\; \frac{2(1-2\delta)\log\frac{1-\delta}{\delta}}{\beta\,\gamma\,\log 3}.$$

By letting $\gamma \to (2\varepsilon(1+\beta)/(\beta\tau))^{1+1/(q-1)}$, we obtain

$$\bar{n}^{\mathrm{wor}}(\varepsilon,\delta,\mathcal{Y}_{p,q,K}\cap\tau\mathcal{B}_q) \;\ge\; c_{q,K}\left(\frac{\tau}{\varepsilon}\right)^{1+1/(q-1)}\log\left(\tfrac{4}{3}\delta\right)^{-1}$$

with $c_{q,K} := \frac{1}{\beta\log 3}\left(\frac{\beta}{2(1+\beta)}\right)^{1+1/(q-1)}$, and the assertion is proven. □

Remark 3.11 (*Dependence on K*)**.** The value of the lower bound in Theorem 3.9 does not depend on K, and in Theorem 3.10 the prefactor $c_{q,K}$ approaches a q-dependent constant as K grows to infinity. In contrast, our upper bounds in Theorem 2.17 do have a K-dependence, but as pointed out in Remark 2.18, in cases of cones $\mathcal{Y}_{p,q,K}$ with $q > 2$ we can directly estimate the variance (which is the moment of interest) and no K-dependence occurs in the scale dependent part of the cost bound,

that is, the part which depends on the norm of the input and the accuracy ε. In that respect, the lower bounds seem to be optimal.

As $K > 1$ approaches 1, though, the lower bounds only hold for small $\varepsilon > 0$. This is not reflected in the upper bound of Theorem 2.17, likely because when constructing our algorithm we ignored the fact that the set $\mathcal{Y}_{p,q,K} \cap \tau \mathcal{B}_{\tilde{q}}$ might be much smaller than the set $\tau \mathcal{B}_{\tilde{q}}$, for $\tau > 0$ and $\tilde{q} := \min\{q, 2\}$. Indeed, for small K the distributions of random variables contained in $\mathcal{Y}_{p,q,K}$ tend to be bimodal, so observing samples from both modes could already enable for a rough approximation of the expectation by returning a value right between the two modes. To get a better intuition we consider the degenerate case $K = 1$, where $\mathcal{Y}_{p,q,K}$, independently from p and q, contains only random variables that either take one constant value, or two distinct values with probability $1/2$ each. Then, for any $\delta \in (0, 1)$ there is an i.i.d.-based algorithm that provides an exact solution with probability at least $(1 - \delta)$. In detail, put $n := \lceil \log_2 \delta^{-1} \rceil + 1$ and define

$$A(Y) := \tfrac{1}{2} \left(\min\{Y_1, \dots, Y_n\} + \max\{Y_1, \dots, Y_n\}\right).$$

If Y has two distinct values, it is with probability $2^{-(n-1)}$ that we only observe one of these. If, however, we observe both values, the output will perfectly match the expected value. Note that in this setting the fixed cost lower bound of Theorem 3.7 still applies.

The prefactor $c_{q,K}$ in Theorem 3.10 tends to zero for small K. We suspect that this is a shortcoming of the lower bound proof.

3.4. Extension to general integration methods

Here we consider the integration problem defined in (1). Intuitively, as long as the measure π has no atoms on G, it seems reasonable to restrict to i.i.d.-based algorithms. (If, however, π has atoms, then there is a non-zero probability that we compute the same function value twice when using i.i.d. samples.) We are going to show that the previously obtained lower bounds for i.i.d.-based algorithms also hold for general randomized algorithms. The key observation is that L_p-norms do not reflect any kind of local structure of functions.

We impose the following weak standing assumption concerning the existence of suitable partitions, which for $G \subseteq \mathbb{R}^d$, equipped with the Borel σ-algebra, is always satisfied whenever π has no atoms.

Assumption 3.12. For every $N \in \mathbb{N}$ we assume that there are pairwise disjoint measurable sets $G_{1,N}, \dots, G_{N,N} \subset G$ with $\bigcup_{j=1}^{N} G_{j,N} = G$ possessing equal measure $\pi(G_{j,N}) = 1/N$ for all $j = 1, \dots, N$.

While lower bounds for i.i.d. based algorithms were obtained by lower bounds for distinguishing random variables $Y^{(1)}$ and $Y^{(2)}$ with expected values at a distance exceeding 2ε, now the idea is to distinguish between two collections $\mathcal{F}^{(1)}$ and $\mathcal{F}^{(2)}$ of functions with distinct integral values,

$$\mathrm{INT}(f) = \begin{cases} a_1 & \text{if } f \in \mathcal{F}^{(1)}, \\ a_2 & \text{if } f \in \mathcal{F}^{(2)}, \end{cases}$$

at distance $|a_2 - a_1| > 2\varepsilon$. Such collections are constructed as follows.

We consider only functions that take at most finitely many distinct values, each with a certain probability. For this let $D \subset \mathbb{R}$ with $\#D < \infty$ and

$$p_z^{(i)} \in [0, 1] \cap \mathbb{Q}, \qquad \text{for } z \in D \text{ and } i = 1,2,$$

satisfying

$$\sum_{z \in D} p_z^{(i)} = 1, \qquad \text{for } i = 1,2.$$

Consider the (unbounded) set of all common denominators of these probabilities, namely,

$$\mathbb{D}_{\mathbf{p}} := \{N \in \mathbb{N} \mid N \cdot p_z^{(i)} \in \mathbb{N}_0 \text{ for all } z \in D \text{ and } i = 1, 2\}.$$

Then, for $N \in \mathbb{D}_{\mathbf{p}}$ and $i = 1,2$, the sets

$$\mathcal{F}_N^{(i)} := \Big\{ f : G \to \mathbb{R} \,|\, f \text{ is constant on each } G_{j,N}, \text{ where } j = 1, \dots, N; \\ \text{and } \pi\{\mathbf{x}\colon f(\mathbf{x}) = z\} = p_z^{(i)} \text{ for } z \in D \Big\}$$

are non-empty. In detail, for $f \in \mathcal{F}_N^{(i)}$, the number of subdomains $G_{j,N}$ with $f(G_{j,N}) = z$ is exactly $N \cdot p_z^{(i)} \in \mathbb{N}_0$. Furthermore, for any function $f \in \mathcal{F}_N^{(i)}$ and any permutation Π on $\{1, \dots, N\}$, the function f_Π with $f_\Pi(G_{j,N}) := f(G_{\Pi(j),N})$ is also contained in $\mathcal{F}_N^{(i)}$. Now, for $i = 1,2$ we define

$$\mathcal{F}^{(i)} := \bigcup_{N \in \mathbb{D}_{\mathbf{p}}} \mathcal{F}_N^{(i)}.$$

This leads to the integral value

$$\mathrm{INT}(f) = \sum_{z \in D} p_z^{(i)} z =: a_i$$

for $f \in \mathcal{F}^{(i)}$.

Proposition 3.13. *Let $\mathbf{X}$ be a π-distributed random variable and let $\mathcal{F}^{(i)}$ with $i = 1,2$ be defined as above. Then*

$$\inf_Q \max_{i=1,2} \sup_{f \in \mathcal{F}^{(i)}} \mathbb{P}\{Q(f) \neq a_i\} = \inf_A \max_{i=1,2} \sup_{f \in \mathcal{F}^{(i)}} \mathbb{P}\{A(f(\mathbf{X})) \neq a_i\},$$

where the infimum on the left-hand side runs over all general randomized algorithms Q with

$$\sup_{f \in \mathcal{F}^{(i)}} \bar{n}(Q, f) \le \bar{n}_i \qquad \text{for } i = 1, 2, \tag{25}$$

and on the right-hand side over all i.i.d.-based algorithms A with the same cost constraint. In particular, this implies that no general randomized algorithm is better than the best possible i.i.d.-based algorithm in distinguishing $\mathcal{F}^{(1)}$ from $\mathcal{F}^{(2)}$.

Proof. For $N \in \mathbb{D}_{\mathbf{p}}$ let $\mu_N^{(i)}$ be the uniform distribution on $\mathcal{F}_N^{(i)}$, for $i = 1,2$. Then, for any general randomized algorithm Q defined on $\mathcal{F}^{(1)} \cup \mathcal{F}^{(2)}$ we have

$$\bar{n}_i \ge \sup_{f \in \mathcal{F}^{(i)}} \bar{n}(Q, f) \ge \int_{\mathcal{F}_N^{(i)}} \bar{n}(Q, f) \, \mathrm{d}\mu_N^{(i)}(f), \tag{26}$$

and

$$\sup_{f \in \mathcal{F}^{(i)}} \mathbb{P}\{Q(f) \neq a_i\} \ge \int_{\mathcal{F}_N^{(i)}} \mathbb{P}\{Q(f) \neq a_i\} \, \mathrm{d}\mu_N^{(i)}(f) \\ \overset{\text{Fubini}}{=} (\mathbb{P} \otimes \mu_N^{(i)})\{(\omega, f)\colon Q(\omega, f) \neq a_i\}. \tag{27}$$

We now consider general randomized algorithms Q that minimize the maximal average uncertainty

$$\max_{i=1,2} (\mathbb{P} \otimes \mu_N^{(i)})\{(\omega, f)\colon Q(\omega, f) \neq a_i\} \tag{28}$$

while satisfying the constraint (25). (Replacing a randomized worst case setting by an average case setting for lower bound proofs is a trick that has already been used by Bakhvalov 1959 [3].) When using n function evaluations, the function is known on at most n of the subdomains of $G_{1,N}, \dots, G_{N,N}$. Without loss of generality, we may assume that the algorithm indeed gains knowledge on n different subdomains. (If not, then a modified method with improved information mapping that avoids computing two function values on the same subdomain while preserving the knowledge of the original information map, can still return the same output or even improve the algorithm.) Due to symmetry of the probability measure $\mu_N^{(i)}$, the conditional distribution of function values on the remaining $N - n$

subdomains is independent of the choice of the n subdomains where the function has been evaluated. Hence, for simplicity, we assume that the algorithm uses one function value within each of the first n subdomains $G_{1,N}, \dots, G_{n,N}$. According to Definition 3.1, such a method can be described via a random variable U, a stopping rule $n(u, y_1, y_2, \dots)$, and an output function $\phi(u, y_1, y_2, \dots)$, such that

$$Q(\omega, f) = \phi(U(\omega), f(\mathbf{x}_1), \dots, f(\mathbf{x}_{n(\omega,f)}), 0, 0, \dots),$$

with stopping time given by $n(\omega, f) := n(U(\omega), f(\mathbf{x}_1), \dots, f(\mathbf{x}_N), 0, 0, \dots)$, and $\mathbf{x}_j \in G_{j,N}$ for $j = 1, \dots, N$. In other words, the infimum over all general randomized algorithms Q for the quantity (28) can be reduced to an infimum over stopping rules $n(u, y_1, y_2, \dots)$ and output functions ϕ while the information of function values is predetermined.

These algorithmic components n and ϕ can also be used with an i.i.d. sample $(f(\mathbf{X}_j))_{j\in\mathbb{N}}$, where $(\mathbf{X}_j)_{j\in\mathbb{N}}$ is a sequence of independent, π-distributed random variables that are independent of U. By this we obtain an i.i.d.-based algorithm

$$A(\omega, f(\mathbf{X})) := \phi(U, f(\mathbf{X}_1(\omega)), \dots, f(\mathbf{X}_{n(\omega,f(\mathbf{X}))}(\omega)), 0, 0, \dots),$$

where $n(\omega, f(\mathbf{X})) := n(U(\omega), f(\mathbf{X}_1(\omega)), f(\mathbf{X}_2(\omega)), \dots)$. For any such A we have

$$(\mathbb{P} \otimes \mu_N^{(i)})\{(\omega, f) : A(\omega, f(\mathbf{X})) \neq a_i\} = \sup_{f\in\mathcal{F}^{(i)}} \mathbb{P}\{A(f(\mathbf{X})) \neq a_i\},$$

since $\mathbb{P}\{A(f(\mathbf{X})) \neq a_i\} = \mathbb{P}\{A(g(\mathbf{X})) \neq a_i\}$ for all representing functions $f, g \in \mathcal{F}^{(i)}$ of the same subclass.

For the following considerations we restrict to methods with some absolute cardinality limit $n(\omega, f) \le M \le N$. If a general randomized algorithm Q occasionally uses more than M samples, by Markov's inequality, in the $\mu_N^{(i)}$-average setting this happens only with average probability

$$(\mathbb{P} \otimes \mu_N^{(i)})\{(\omega, f) \mid n(\omega, f) > M\} \le \frac{\bar{n}_i}{M}. \tag{29}$$

We can define another general randomized algorithm Q' based on Q by simply stopping Q whenever it would compute yet another function value after already knowing M function values, and returning an arbitrary output then. The uncertainty increases at most by $\bar{n}_i/M$ (and vanishes in the limit $M \to \infty$ which we perform later on), whereas the cost is even reduced.

Now we compare the probabilities at which function values are observed by a general randomized algorithm Q_N in the $\mathbb{P} \otimes \mu_N^{(i)}$ setting and by an i.i.d.-based algorithm A, where both methods are constructed from the same stopping rule n and output function ϕ. (We write Q_N instead of Q to make clear that the method is based on the previously described optimal information with function evaluations at nodes $\mathbf{x}_k \in G_{k,N}$ for the $\mu_N^{(i)}$-setting while the stopping rule $n(u, y_1, y_2, \dots)$ and the output function ϕ remain unchanged.) If $n(\omega, f)$ is absolutely bounded by $M \le N$, it suffices to consider the probability of different outcomes of M function evaluations. For an i.i.d. sample we have

$$\mathbb{P}\{f(\mathbf{X}_1) = z_1, \dots, f(\mathbf{X}_M) = z_M\} = \prod_{j=1}^{M} p_{z_j}^{(i)} \qquad \text{for } f \in \mathcal{F}^{(i)} \text{ and } z_1, \dots, z_M \in D, \tag{30}$$

and in the $\mu_N^{(i)}$-average setting we have

$$\begin{aligned} \prod_{j=1}^{M} \frac{N \cdot p_{z_j}^{(i)} - M}{N} &\le \mu_N^{(i)}\{f(\mathbf{x}_1) = z_1, \dots, f(\mathbf{x}_M) = z_M\} \\ &= \prod_{j=1}^{M} \frac{N \cdot p_{z_j}^{(i)} - \#\{\ell = 1, \dots, j-1 \colon z_\ell = z_j\}}{N - (j-1)} \\ &\le \prod_{j=1}^{M} \frac{N \cdot p_{z_j}^{(i)}}{N - M}, \end{aligned} \tag{31}$$

where the lower bound holds if $M \le N \cdot p_z^{(i)}$ for all $z \in D$ and $i = 1,2$. This is true for large N, and in the limit of $N \to \infty$ while M is fixed, the probability of (31) approaches (30). Determining the uncertainty probability is now a discrete problem, namely, by the independence of U from the i.i.d.-sample we have

$$\mathbb{P}\{A(f(\mathbf{X})) \neq a_i\} = \sum_{(z_1,\dots,z_M)\in D^M} \mathbb{P}\{\phi(U,z_1,\dots,z_M,0,\dots) \neq a_i\} \cdot \mathbb{P}\{f(\mathbf{X}_1) = z_1,\dots,f(\mathbf{X}_M) = z_M\}$$

for $f \in \mathcal{F}^{(i)}$. On the other hand, within the $\mu_N^{(i)}$-average setting we have

$$(\mathbb{P}\otimes\mu_N^{(i)})\{(\omega,f)\colon Q_N(\omega,f) \neq a_i\} = \sum_{(z_1,\dots,z_M)\in D^M} \mathbb{P}\{\phi(U,z_1,\dots,z_M,0,\dots) \neq a_i\} \cdot \mu_N^{(i)}\{f(\mathbf{x}_1) = z_1,\dots,f(\mathbf{x}_M) = z_M\}\,.$$

Here, we wrote $\phi(U,z_1,\dots,z_M,0,\dots)$, which in fact only depend on $z_1,\dots,z_{n(U,\mathbf{z})}$ where $n(U,\mathbf{z}) := n(U,z_1,\dots,z_M,0,\dots)$. Hence, for any $\eta > 0$ there is an $N_0 \in \mathbb{D}_{\mathbf{p}}$ such that for all $N \ge N_0 \ge M$, for all stopping rules $n(u,y_1,y_2,\dots) \le M$, and for all output functions ϕ, the corresponding methods Q_N and A satisfy for $f \in \mathcal{F}^{(i)}$ that

$$\mathbb{P}\{A(f(\mathbf{X})) \neq a_i\} \le \eta + (\mathbb{P}\otimes\mu_N^{(i)})\{(\omega,f)\colon Q_N(\omega,f) \neq a_i\}\,. \tag{32}$$

Similarly, for any $\nu > 0$ and sufficiently large N we obtain

$$\begin{aligned}\bar{n}(A,f(\mathbf{X})) &= \sum_{(z_1,\dots,z_M)\in D^M} \mathbb{E}\, n(U,z_1,\dots,z_M,0,\dots)\cdot\mathbb{P}\{f(\mathbf{X}_i) = z_i,\ i=1,\dots,M\}\\ &\le \nu + \sum_{(z_1,\dots,z_M)\in D^M} \mathbb{E}\, n(U,z_1,\dots,z_M,0,\dots)\cdot\mu_N^{(i)}\{f(\mathbf{x}_i) = z_i,\ i=1,\dots,M\}\\ &= \nu + \int_{\mathcal{F}_N^{(i)}} \bar{n}(Q_N,f)\,\mathrm{d}\mu_N^{(i)}(f) \overset{(26)}{\le} \nu + \bar{n}_i\,.\end{aligned}$$

Hence, A does not satisfy the constraint (25), but $\nu \to 0$ as $N \to \infty$. The violation of the constraint can be fixed by defining a modified method A' that, independently from all random variables used in A, decides with probability

$$\zeta := \max_{i=1,2} \frac{\nu}{\nu + \bar{n}_i} \tag{33}$$

that no information at all shall be collected and simply returns an arbitrary value, whereas with probability $1-\zeta$ the original algorithm A is executed. That way the uncertainty of the method A' only rises by ζ, which vanishes in the limit $\nu \to 0$.

To summarize, since i.i.d.-based algorithms are special general randomized algorithms, for sufficiently large N we have

$$\begin{aligned}\inf_A \max_{i=1,2} \sup_{f\in\mathcal{F}^{(i)}} \mathbb{P}\{A(f(\mathbf{X})) \neq a_i\} &\overset{(27)}{\ge} \inf_Q \max_{i=1,2} (\mathbb{P}\otimes\mu_N^{(i)})\{(\omega,f)\colon Q(\omega,f) \neq a_i\}\\ &\overset{(29),(32),(33)}{\ge} \inf_A \max_{i=1,2} \sup_{f\in\mathcal{F}^{(i)}} \mathbb{P}\{A(f(\mathbf{X})) \neq a_i\} - \frac{\bar{n}_i}{M} - \eta - \zeta\,,\end{aligned}$$

where the infimum runs over all methods satisfying the cardinality constraint (25). Fixing M and letting $N \to \infty$, η and ζ vanish. Finally, the limit $M \to \infty$ proves the assertion. □

An immediate consequence of the previous proposition is that the different types of lower bounds for i.i.d.-based algorithms on $\mathcal{Y}_{p,q,K}$ carry over to lower bounds for general randomized algorithms on $\mathcal{F}_{p,q,K}(\pi)$.

Corollary 3.14. *Let $1 \le p < q \le \infty$, $K > 1$ and $\tau > 0$. Then, for the integration problem (1),*

1. *on $\mathcal{F} = L_q(\pi)$ we obtain unsolvability, see corresponding Theorem 3.4;*

and on cone-shaped input classes $\mathcal{F}_{p,q,K}(\pi)$ we obtain

2. *impossibility for any general randomized algorithm Q to guarantee a finite worst case expected error, that is,*

$$\sup_{f\in\mathcal{F}_{p,q,K}(\pi)} \mathbb{E}\,|Q(f) - \mathrm{INT}(f)| \;=\; \infty,$$

   see corresponding Theorem 3.5;
3. *the following lower bound for the fixed cost of general randomized algorithms,*

$$\bar{n}^{\mathrm{fix}}(\varepsilon,\delta,\mathcal{F}_{p,q,K}(\pi)) \;\geq\; \frac{1}{2\log 2}\, K^{pq/(q-p)}\, \log \delta^{-1}$$

   for any $\varepsilon > 0$ and $\delta \in (0,1)$, see corresponding Theorem 3.7;
4. *the following lower bound for the worst case complexity for bounded norm,*

$$\bar{n}^{\mathrm{wor}}(\varepsilon,\delta,\mathcal{F}_{p,q,K}(\pi)\cap \tau\mathcal{B}_{\widetilde{q}}) \;\geq\; c_{q,K}\,\left(\frac{\tau}{\varepsilon}\right)^{1+1/(\widetilde{q}-1)}\,\log\left(\tfrac{4}{3}\delta\right)^{-1}$$

   for any $0 < \varepsilon < \frac{1}{6}(1-1/K)\,\tau$ and $0 < \delta < 1/4$, with $\widetilde{q} := \min\{q,2\}$, and a constant $c_{q,K} > 0$ depending only on q and K, see corresponding Theorems 3.9 and 3.10.

Proof. In all cases, the random variables have to be modified before building corresponding function classes, now possessing rational weights for the different values such that the cone condition within $\mathcal{Y}_{p,q,K}$ is still satisfied. Since $\mathbb{Q}$ is dense in $\mathbb{R}$, we obtain the same lower bounds. Cost bounds at given confidence level (as to be found in the cited theorems) are equivalent to uncertainty bounds at given cost budget (as considered in Proposition 3.13), namely due to monotonicity and continuity of the connection between the two quantities, inversion does the job.

In Theorems 3.9 and 3.10 we are interested in the worst case cost, so the cost constraint we impose when applying Proposition 3.13 will be with $\bar{n}_1 = \bar{n}_2$.

For generalizing the unsolvability results Theorems 3.4 and 3.5, the cardinality constraint (25) in Proposition 3.13 must be replaced by a constraint of type

$$\mathbb{P}\{\bar{n}(A,f) > M\} \leq \nu_M \qquad \text{for all } f \in \mathcal{F}^{(1)} \cup \mathcal{F}^{(2)}, \tag{34}$$

with a monotonously decaying sequence $(\nu_M)_{M\in\mathbb{N}_0}$ converging to 0. Consequently, within the proof of Proposition 3.13, the estimate (29) has to be adjusted accordingly, but all following arguments with the limit $M \to \infty$ will work the same. In detail, $\mathcal{F}^{(1)}$ only consists of the zero function. Since an algorithm that shall distinguish the zero function from a class $\mathcal{F}^{(2)}$ consisting of non-zero functions can stop when observing a non-zero function value, we may assume $\nu_M := \mathbb{P}\{\bar{n}(A,0) > M\}$, so the constraint is actually no restriction.

For generalizing Theorem 3.7 we need to extend Proposition 3.13 to the task of distinguishing finitely many function classes $\mathcal{F}^{(0)},\dots,\mathcal{F}^{(m)}$ with integral values at separated distances $> 2\varepsilon$, the proof works similar. By taking the limit $m \to \infty$, we may even consider countably many function classes $\mathcal{F}^{(i)}$, $i \in \mathbb{N}_0$. Now for the detailed transcription of the proof of Theorem 3.7: We take any function $f^* \in \mathcal{F}_{p,q,K}$ (which corresponds to $Y \in \mathcal{Y}_{p,q,K}$, the input for which we eventually derive a cost bound); further we consider function classes $\mathcal{F}^{(i)}$ with $i \in \mathbb{N}_0$ (which correspond to random variables of type $c_i Z$), such that $f := f^* + \widetilde{f} \in \mathcal{F}_{p,q,K}$ for $\widetilde{f} \in \mathcal{F}^{(i)}$. The set $\mathcal{F}^{(0)}$ only contains the zero function, for all other indices $i \in \mathbb{N}$ we specify for $\widetilde{f} \in \mathcal{F}^{(i)}$ that

$$\pi\{\widetilde{f}(\mathbf{x}) = 0\} = 1 - a\,, \qquad \text{and} \qquad \pi\{\widetilde{f}(\mathbf{x}) = c_0 + 3\,i\varepsilon/a\} = a,$$

with $c_0 > 0$ and $a \in \mathbb{Q}$ as in the proof of Theorem 3.7. Consequently,

$$\mathrm{INT}(\widetilde{f}) = \begin{cases} 0 & \text{for } \widetilde{f} \in \mathcal{F}^{(0)}, \\ ac_0 + 3\,i\varepsilon & \text{for } \widetilde{f} \in \mathcal{F}^{(i)} \text{ with } i \in \mathbb{N}. \end{cases}$$

The task of determining $\mathrm{INT}(f) = \mathrm{INT}(f^* + \widetilde{f})$, not knowing the index $i \in \mathbb{N}_0$ for which $\widetilde{f} \in \mathcal{F}^{(i)}$, can be traced back to determining $\mathrm{INT}(\widetilde{f})$, since for any general algorithm Q acting on $\bigcup_{i=0}^{\infty} \mathcal{F}^{(i)}$, by

$$Q'(f) := Q(f - f^*) + \mathrm{INT}(f^*)$$

another algorithm acting on $f^* + \bigcup_{i=0}^{\infty} \mathcal{F}^{(i)}$ and exhibiting the same reliability is defined. This also holds for algorithms that use i.i.d. samples, because in the integration setting the points $\mathbf{X}_i \sim \pi$ are user-generated and one may use $f^*(\mathbf{X}_i)$ within Q'. Since it suffices to observe one function value $\widetilde{f}(\mathbf{x}_j) = z_i \neq 0$ in order to know that $\widetilde{f} \in \mathcal{F}^{(i)}$ with $i \neq 0$, the cost $\bar{n}_0 = \bar{n}(Q, f^*) = \bar{n}^{\mathrm{wor}}(Q', \mathcal{F}^{(0)})$ is an upper bound for the worst case cost $\bar{n}_i = \bar{n}^{\mathrm{wor}}(Q', \mathcal{F}^{(i)})$ in the other cases. □

We can formulate an even stronger unsolvability result that holds for the small space of smooth functions.

Theorem 3.15. *The integration problem (1) for the class of infinitely differentiable functions,*

$$\mathcal{F} = C^{\infty}([0, 1]),$$

with respect to the Lebesgue measure on $[0, 1]$*, is not solvable.*

Proof. Split $[0, 1]$ into N intervals of equal length, $G_i := [(i-1)/N, i/N)$, with $i = 1, \ldots, N$. It is well known that there exist C^{∞}-functions f_i on $\mathbb{R}$ that are supported on the closures of the intervals $\overline{G_i}$ and positive on the interior of G_i. Scaling does not affect the property of belonging to the class $\mathcal{F} = C^{\infty}([0, 1])$, so $\mathrm{INT}(f_i)$ can be arbitrarily large. Similarly to the proof of Proposition 3.13, we aim to distinguish the zero function from a $\mu_N^{(2)}$-average input where $\mu_N^{(2)}$ is the uniform distribution on the set $\mathcal{F}_N^{(2)} := \{f_1, \ldots, f_N\}$. Analogously to the proof of Theorem 3.4, the $\mu_N^{(2)}$-probability of observing a non-zero function value after n_0 function evaluations vanishes in the limit $N \to \infty$. In the same manner we then obtain that for any $\varepsilon > 0$ and $\delta \in (0, 1/2)$ there is no (ε, δ)-approximating algorithm for the whole class $C^{\infty}([0, 1])$. □

Acknowledgments

We thank Fred Hickernell, Mark Huber, and Henryk Woźniakowski for many valuable discussions over several years and we also thank the referees. Parts of the work have been done at the ESI in Vienna during the programme "Tractability of High Dimensional Problems and Discrepancy" in October 2017. Daniel Rudolf gratefully acknowledges support of the Felix-Bernstein-Institute for Mathematical Statistics in the Biosciences (Volkswagen Foundation), Germany and the Campus laboratory AIMS, Germany.

References

[1] N. Alon, Y. Matias, M. Szegedy, The space complexity of approximating the frequency moments, in: STOC '96 Proceedings of The Twenty-Eighth Annual ACM Symposium on Theory of Computing, 1996, pp. 20–29.
[2] B. von Bahr, C. Esseen, Inequalities for the rth absolute moment of a sum of random variables, $1 \leq r \leq 2$, Ann. Math. Stat. 36 (1965) 299–303.
[3] N.S. Bakhvalov, On the approximate calculation of multiple integrals, Vestnik MGU, Ser. Math. Mech. Astron. Phys. Chem. 4 (1959) 3–18, (in Russian), English translation: J. Complexity, 31(4):502-516, 2015.
[4] C. Bennett, R. Sharpley, Interpolation of Operators, in: Pure and Applied Mathematics, vol. 129, Academic Press Inc, Boston, MA, 1988.
[5] S. Bubeck, N. Cesa-Bianchi, G. Lugosi, Bandits with heavy tail, IEEE Trans. Inform. Theory 59 (11) (2013) 7711–7717.
[6] Y.S. Chow, H. Robbins, On the asymptotic theory of fixed-width sequential confidence intervals for the mean, Ann. Math. Stat. 36 (2) (1965) 457–462.
[7] N. Clancy, Y. Ding, C. Hamilton, F.J. Hickernell, Y. Zhang, The cost of deterministic, adaptive, automatic algorithms: cones, not balls, J. Complexity 30 (1) (2014) 21–45.
[8] P. Dagum, R. Karp, M. Luby, S. Ross, An optimal algorithm for MC estimation, SIAM J. Comput. 19 (5) (2000) 1484–1496.
[9] L. Devroye, M. Lerasle, G. Lugosi, R.I. Oliveira, Sub-gaussian mean estimators, Ann. Statist. 44 (2016) 2695–2725.
[10] L. Gajek, W. Niemiro, P. Pokarowski, Optimal Monte Carlo integration with fixed relative precision, J. Complexity 29 (2013) 4–26.

[11] S. Heinrich, Random approximation in numerical analysis, in: Proceedings of the Conference Functional Analysis Essen, 1994, pp. 123–171.
[12] S. Heinrich, On the complexity of computing the L_q norm, J. Complexity 49 (2018) 1–26.
[13] F.J. Hickernell, L. Jiang, Y.W. Liu, A. Owen, Guaranteed conservative fixed width confidence intervals via Monte Carlo sampling, in: Proceedings of the MCQMC 2012, Springer, 2013.
[14] D. Hsu, S. Sabato, Loss minimization and parameter estimation with heavy tails, J. Mach. Learn. Res. 17 (2016) 1–40.
[15] M. Huber, A Bernoulli mean estimate with known relative error distribution, Random Structures Algorithms 50 (2017) 173–182.
[16] M. Huber, An optimal (ϵ, δ)-approximation scheme for the mean of random variables with bounded relative variance, 2017, Available on arXiv:1706.01478v1 [stat.CO].
[17] M. Jerrum, L. Valiant, V. Vazirani, Random generation of combinatorial structures from a uniform distribution, Theoret. Comput. Sci. 43 (1986) 169–188.
[18] E. Joly, G. Lugosi, Robust estimation of U-statistics, Stochastic Process. Appl. 126 (12) (2016) 3760–3773.
[19] E. Joly, G. Lugosi, R.I. Oliveira, On the estimation of the mean of a random vector, Electron. J. Stat. 11 (2017) 440–451.
[20] R.J. Kunsch, D. Rudolf, Optimal confidence for Monte Carlo integration of smooth functions, 2018, Available on arXiv:1809.09890 [math.NA].
[21] M. Lerasle, R.I. Oliveira, Robust empirical mean estimators, 2012, Available on arXiv:1112.3914v1 [math.ST].
[22] G. Lugosi, S. Mendelson, Risk minimization by median-of-means tournaments, 2016, Available on arXiv:1608.00757v1 [math.ST].
[23] G. Lugosi, S. Mendelson, Regularization, sparse recovery, and median-of-means tournaments, 2017, Available on arXiv:1701.04112v1 [math.ST].
[24] S. Minsker, Geometric median and robust estimation in Banach spaces, Bernoulli 21 (4) (2015) 2308–2335.
[25] A.S. Nemirovsky, D.B. Yudin, Problem Complexity and Method Efficiency in Optimization, John Wiley & Sons, 1983, English translation, originally in Russian, 1979.
[26] W. Niemiro, P. Pokarowski, Fixed precision MCMC estimation by median of product of averages, J. Appl. Probab. 46 (2009) 309–329.
[27] E. Novak, Deterministic and Stochastic Error Bounds in Numerical Analysis, in: Lecture Notes in Mathematics, vol. 1349, Springer-Verlag, Berlin, 1988.
[28] E. Novak, Some results on the complexity of numerical integration, in: Ronald Cools, Dirk Nuyens (Eds.), Monte Carlo and Quasi-Monte Carlo Methods, MCQMC, Leuven, Belgium, Springer, 2016.
[29] E. Novak, H. Woźniakowski, Tractability of Multivariate Problems, in: Linear Information, vol. I, European Mathematical Society, 2008.
[30] E. Novak, H. Woźniakowski, Tractability of Multivariate Problems, in: Standard Information for Functionals, vol. II, European Mathematical Society, 2010.
[31] E. Novak, H. Woźniakowski, Tractability of Multivariate Problems, in: Standard Information for Operators, vol. III, European Mathematical Society, 2012.
[32] Y. Ren, H. Liang, On the best constant in Marcinkiewicz–Zygmund inequality, Statist. Probab. Lett. 53 (2001) 227–233.
[33] D. Rudolf, N. Schweizer, Error bounds of MCMC for functions with unbounded stationary variance, Statist. Probab. Lett. 99 (2015) 6–12.
[34] D. Siegmund, Sequential Analysis: Tests and Confidence Intervals, Springer-Verlag, 1985.
[35] C. Stein, A two-sample test for a linear hypothesis whose power is independent of the variance, Ann. Math. Stat. 16 (1945) 243–258.
[36] J.F. Traub, G.W. Wasilkowski, H. Woźniakowski, Information-Based Complexity, Academic Press, 1988.
[37] A. Wald, Sequential tests of statistical hypotheses, Ann. Math. Stat. 16 (1945) 117–186.
[38] A. Wald, Sequential Analysis, John Wiley, 1948.