A constraint on extensible quadrature rules

Art B. Owen
Stanford University

January 2015

Abstract

When the worst case integration error in a family of functions decays as $n^{-\alpha}$ for some $\alpha > 1$ and simple averages along an extensible sequence match that rate at a set of sample sizes $n_1 < n_2 < \cdots < \infty$, then these sample sizes must grow at least geometrically. More precisely, $n_{k+1}/n_k \geq \rho$ must hold for a value $1 < \rho < 2$ that increases with α. This result always rules out arithmetic sequences but never rules out sample size doubling. The same constraint holds in a root mean square setting.

1 Introduction

For both Monte Carlo (MC) and quasi-Monte Carlo (QMC) sampling, Sobol' (1998) recommends that the number n of sample points should be increased geometrically, not arithmetically. Specifically, he recommends using a sequence like n_1, $2n_1$, $4n_1$, etc., instead of n_1, $2n_1$, $3n_1$ and so on. For either MC or QMC, the estimate of an integral is a simple unweighted average of the integrand at n points.

In the case of Monte Carlo sampling with its slow $n^{-1/2}$ convergence rate, if n is too small to get a good answer then taking $n+k$ sample points for $k \ll n$ is unlikely to bring a meaningful improvement in accuracy. Sobol' (1993) studies the correlations among Monte Carlo estimates along sample sizes $n_k = 2^{k-1}n_1$.

In quasi-Monte Carlo sampling, much better convergence rates are sometimes possible, depending particularly on the smoothness and dimension of the problem space. Novak and Wozniakowski (2010) provide a comprehensive treatise on error rates for numerical integration. In favorable cases, a small change in n might make a meaningful reduction in the error bound. What we show here is that rate-optimal sample sizes are widely spaced in those favorable cases.

Sobol' (1998) showed that a better rate than $1/n$ cannot hold uniformly for all n. Hickernell et al. (2012) extended this finding to arithmetic sequences of sample sizes. They also showed that unequally weighted averages of function evaluations can attain a better than $1/n$ rate at all values of n. Suppose for instance that the first n sample points are partitioned into blocks of n_j points, for $j = 1, \ldots, J$ where the estimate from block j has error $O(n_j^{-\alpha})$. Then weighting those within block estimates proportionally to n_j^a, with $a \ge \alpha$, attains an error $O((J/n)^\alpha)$. One can commonly arrange $J = O(\log n)$ and then the error is $o(n^{-\alpha+\epsilon})$ for any $\epsilon > 0$.

It remains interesting to consider equal weight rules. For a complicated problem with weighted points from multiple spaces, keeping track of the weights becomes cumbersome. Also, there are quasi-Monte Carlo methods such as higher order digital nets (Dick, 2011), that are simultaneously rate optimal for more than one class of functions, each with its own rate. In such settings we might want to use the same weights for multiple integrands, but no single weighing might serve them all best. Finally, if the constraints we find here on equal weight rules are unpalatable for some given problem, it provides motivation to switch to an unequally weighted rule.

An outline of this note is as follows. Section 2 presents the insight from the appendix of Sobol' (1998) and the extension by Hickernell et al. (2012). If a QMC rule has worst case error $o(1/n)$ holding for all n, then the points $\boldsymbol{x}_i$ must have some very strange limit properties and the class of functions involved is odd enough that we could do very well using only one point $\boldsymbol{x}_n$ for very large n. A generalization of that argument shows that an $o(1/n)$ rate along an arithmetic sequence of sample sizes raises similar problems. Section 3 shows that if quadrature error in a class $\mathcal{F}$ of functions has a worst case lower bound $mn^{-\alpha}$ for $\alpha > 1$ and a specific sequence $\boldsymbol{x}_i$ is rate optimal at sample sizes $n_1 < n_2 < \cdots$, then necessarily $n_{k+1}/n_k \ge \rho$ for some constant $1 < \rho < 2$ depending on α and on how close the sequence comes to having the optimal constant. Section 4 considers root mean squared error for sequences of sample points incorporating some randomness. The same constraints hold in this setting as in the worst case setting.

2 Ruling out arithmetic sequences

It is not reasonable to expect a QMC rule to have errors of size $o(1/n)$ for all values of $n \ge N_0$ for some $N_0 > 0$. Intuitively, adding a single point makes a change of order $1/n$ to the estimated integral. Therefore two consecutive values of the integral estimate are ordinarily an order of magnitude farther apart from each other than they could be by the triangle inequality, with the true integral value making the third corner of the triangle. This idea is made precise in the appendix of Sobol' (1998), as we outline here.

Let $\mu = \mu(f) = \int f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$ and $\hat\mu_n = \hat\mu_n(f) = (1/n)\sum_{i=1}^n f(\boldsymbol{x}_i)$. The integral is over $[0,1]^d$ and $\boldsymbol{x}_i \in [0,1]^d$ for $i \ge 1$. Let

$$\eta_n = \eta_n(f) = \frac{1}{n}\sum_{i=1}^n f(\boldsymbol{x}_i) - \mu$$

and suppose that the points $\boldsymbol{x}_i$ are chosen such

$$\sup_{f\in\mathcal{F}} |\eta_n(f)| \le \epsilon(n)$$

where $\epsilon(n) = o(1/n)$, and $\mathcal{F}$ is a class of integrands. Sobol' considered $\epsilon(n) = O(n^{-\alpha})$ for some $\alpha > 1$. The classes $\mathcal{F}$ that we study are usually balls with respect to a seminorm, such as the standard deviation in Monte Carlo and the total variation in the sense of Hardy and Krause, for quasi-Monte Carlo.

Sobol' observed that

$$\begin{aligned}|f(\boldsymbol{x}_{n+1}) - \mu| &= |(n+1)\eta_{n+1} - n\eta_n| \\ &\le (n+1)\epsilon(n+1) + n\epsilon(n) \\ &\to 0\end{aligned}$$

as $n \to \infty$. As a result $\lim_{n\to\infty} f(\boldsymbol{x}_i) = \mu(f)$ for all $f \in \mathcal{F}$. The set $\mathcal{F}$ cannot be very rich in this case. As Sobol' noted, for $d = 1$, if $\mathcal{F}$ contains x it cannot also contain x^2.

If we had such a sequence $\boldsymbol{x}_i$ and a class $\mathcal{F}$ we might simply estimate μ by $f(\boldsymbol{x}_n)$ for one extremely large n, perhaps the largest one for which we can compute $\boldsymbol{x}_n$. Alternatively, when $f \in \mathcal{F}$ are all known to be integrable and anti-symmetric functions on $[0,1]^d$ we can take $\boldsymbol{x}_1 = (1/2, \dots, 1/2)$ and have a zero error. This is the most favorable case for antithetic sampling (Hammersley and Morton, 1956). MC and QMC methods are ordinarily designed for more general purpose use, and so such special settings are of limited importance.

Hickernell et al. (2012) extend Sobol's argument to show that we should not expect $\epsilon(nk) = o(1/n)$ as $n \to \infty$ for any integer $k \ge 1$. We would then have a class of functions $\mathcal{F}$ with

$$\lim_{n\to\infty} \frac{1}{k}\sum_{i=1}^{k} f(\boldsymbol{x}_{nk+i}) \to \mu(f)$$

for all $f \in \mathcal{F}$. That is a very limited class, and once again, we could solve the problem uniformly over that class simply by taking k points $\boldsymbol{x}_{nk+1}, \dots \boldsymbol{x}_{(n+1)k}$ for some very large n.

3 Geometric spacing for the worst case setting

The class $\mathcal{F}$ of real-valued functions on $[0,1]^d$ has a superlinear worst case lower bound if

$$\sup_{f\in\mathcal{F}} \left| \frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{x}_i) - \mu(f) \right| > mn^{-\alpha} \tag{1}$$

holds for some $\alpha > 1$, $m > 0$, all $n \ge 1$ and all $\boldsymbol{x}_i \in [0,1]^d$. There is a uniformly rate optimal sequence for this class, if for some $\boldsymbol{x}_i \in [0,1]^d$ and a sequence of sample sizes $n_1 < n_2 < \cdots$,

$$\sup_{f\in\mathcal{F}} \left| \frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{x}_i) - \mu(f) \right| \le Mn^{-\alpha}, \quad \forall n \in \{n_1, n_2, \dots\} \tag{2}$$

holds, where $m \le M < \infty$.

The proof of Theorem 1 below makes use of some basic facts about fixed-point iterations. Let $g(x)$ be a continuous function on the interval $[a, b]$ taking values in $[a, b]$. Then g has at least one fixed point $x_* \in [a, b]$, with $g(x_*) = x_*$. If also, g has Lipschitz constant $\lambda < 1$ for all $x \in [a, b]$, then the fixed point x_* is unique. Now consider the fixed point iteration $x_{n+1} = g(x_n)$. Under the Lipschitz condition, x_n converges to x_* from any $x_1 \in [a, b]$. These facts are consequences of Kelley (1999, Theorem 4.2.1). When g has derivative g' with $0 < g'(x) < 1$ on $[a, b]$, then the convergence to x_* is monotone: if $x_1 < x_*$ then $x_n < x_{n+1} < x_*$ for all $n \ge 1$, or if $x_1 > x_*$, then $x_n > x_{n+1} > x_*$ for all $n \ge 1$ (Ackleh et al., 2009, Remark 2.6).

Theorem 1. *Let $\mathcal{F}$ have a worst case lower bound given by* (1) *with $\alpha > 1$ and $0 < m \le M < \infty$. Suppose that there also exists a uniformly rate optimal sequence $\boldsymbol{x}_i$ satisfying* (2)*. If $\rho = \rho_k = n_{k+1}/n_k$, then*

$$\rho \ge 1 + \left[\frac{m}{M}(1 + \rho^{1-\alpha})^{-1}\right]^{1/(\alpha-1)} > 1 + \Bigl(\frac{m}{2M}\Bigr)^{1/(\alpha-1)}. \tag{3}$$

Proof. From the lower bound (1), there is an $f \in \mathcal{F}$ with

$$\begin{aligned} m(n_{k+1} - n_k)^{-\alpha} &\le \left|\frac{1}{n_{k+1} - n_k}\sum_{i=n_k+1}^{n_{k+1}} (f(\boldsymbol{x}_i) - \mu)\right| \\ &= (n_{k+1} - n_k)^{-1}\left|\sum_{i=1}^{n_{k+1}} (f(\boldsymbol{x}_i) - \mu) - \sum_{i=1}^{n_k}(f(\boldsymbol{x}_i) - \mu)\right| \\ &\le M(n_{k+1} - n_k)^{-1}(n_{k+1}^{1-\alpha} + n_k^{1-\alpha}), \end{aligned} \tag{4}$$

where we have applied upper bounds from (2). Writing $\rho = n_{k+1}/n_k$ and rearranging (4), yields the first inequality in (3).

Next we define $g(\rho) = 1 + ((m/M)(1 + \rho^{1-\alpha})^{-1})^{1/(\alpha-1)}$, the middle quantity in (3). It has derivative

$$g'(\rho) = \Bigl(\frac{m}{M}\Bigr)^{1/(\alpha-1)}\bigl(1 + \rho^{1-\alpha}\bigr)^{-1/(\alpha-1)-1}\rho^{-\alpha}$$

which is positive for $\rho \in [1, 2]$. Therefore $1 < g(1) \le g(\rho) \le g(2) < 2$ and so g maps $[1, 2]$ into $[1, 2]$. Because $g' \le \lambda \equiv (m/M)^{1/(\alpha-1)}2^{-1/(\alpha-1)-1}$ the Lipschitz constant for g is at most $\lambda < 1$. Thus g has a unique fixed point $\rho_* \in (1, 2)$ with $g(\rho_*) = \rho_*$. If ρ satisfies the first inequality in (3), then $\rho \ge \rho_*$. By the monotone convergence of this iteration, $\rho_* > g(1) = 1 + (m/(2M))^{1/(\alpha-1)}$. □

A rate optimal sequence n_k must be spread out at least geometrically, with a ratio $n_{k+1}/n_k \ge \rho$ where ρ is now the smallest number satisfying the first inequality in (3). This ρ depends on m/M and on α. By definition $\rho > 1$. Also we can find by inspection that $\rho = 2$ satisfies the first inequality in (3) and so Theorem 1 never rules out a doubling of the sample size. The lower bound on the critical extension factor n_{k+1}/n_k always satisfies $1 < \rho < 2$.

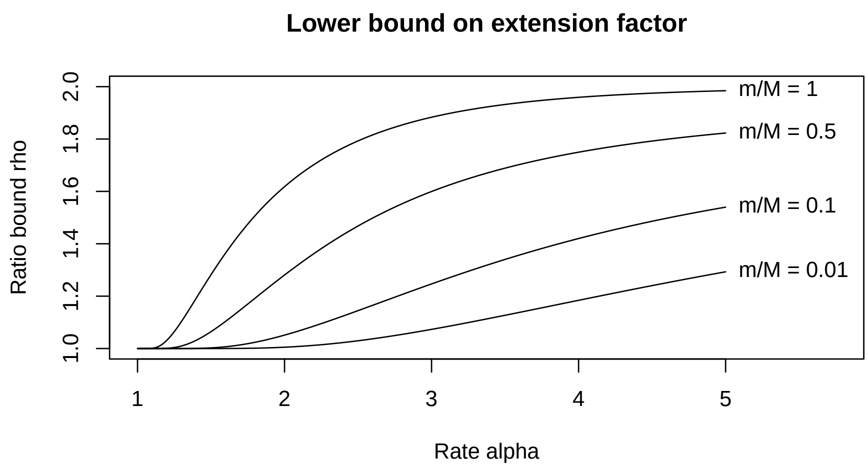

Figure 1: For worst case rates $n^{-\alpha}$ this figure shows a lower bound on the extension factor n_{k+1}/n_k in a uniformly rate-optimal integration sequence.

Figure 1 shows this extension bound as a function of α for various levels of the ratio m/M. The values there were computed via Brent's algorithm (Brent, 1973). The case $m/M = 1$ is of special interest. It describes a rate-optimal rule that also attains the optimal constant. That case has the highest bound on the extension factor.

If the rate $n^{-\alpha}$ is generalized to $n^{-\alpha}\log(n)^{\beta}$ for $\alpha > 1$ and $\beta > 0$, then a geometric spacing is still necessary. For any $1 < \gamma < \alpha$ and large enough n_k it is necessary to have $\rho_k = n_{k+1}/n_k \geq 1 + (m/2M)^{1/(1-\gamma)}$.

4 The root mean square error setting

The class $\mathcal{F}$ of real-valued functions on $[0,1]^d$ has a superlinear root mean square lower bound if

$$\sup_{f\in\mathcal{F}} \mathbb{E}\biggl(\biggl|\frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{x}_i) - \mu(f)\biggr|^2\biggr)^{1/2} > mn^{-\alpha} \tag{5}$$

holds for some $\alpha > 1$, $m > 0$, all $n \geq 1$ and any random $\boldsymbol{x}_i \in [0,1]^d$. Here $\mathbb{E}(\cdot)$ denotes expectation with respect to the randomness in $\boldsymbol{x}_i$. The sequence of random $\boldsymbol{x}_i \in [0,1]^d$ is uniformly rate-optimal for this class if there is a sequence of sample sizes $n_1 < n_2 < \cdots$, for which

$$\sup_{f\in\mathcal{F}} \mathbb{E}\biggl(\biggl|\frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{x}_i) - \mu(f)\biggr|\biggr)^{1/2} \leq Mn^{-\alpha}, \quad \forall n \in \{n_1, n_2, \dots\} \tag{6}$$

holds, where $m \le M < \infty$.

The same theorem holds for the root mean square error case as holds for the worst case. It is not necessary to assume that any of $f(\boldsymbol{x}_i)$ are unbiased or to make any assumption about the correlation structure among the $f(\boldsymbol{x}_i)$. We only need to square one of the identities in Theorem 1 and then use standard moment inequalities from probability theory.

Theorem 2. *Let $\mathcal{F}$ have a root mean square lower bound given by* (5) *with $\alpha > 1$ and $0 < m \le M < \infty$. Suppose that there also exists a uniformly rate optimal sequence of random $\boldsymbol{x}_i$ satisfying* (6). *If $\rho = \rho_k = n_{k+1}/n_k$, then*

$$\rho \ge 1 + \left[\frac{m}{M}(1+\rho^{1-\alpha})^{-1}\right]^{1/(\alpha-1)} \ge 1 + \Big(\frac{m}{2M}\Big)^{1/(\alpha-1)}. \qquad (7)$$

Proof. To shorten some expressions, let $\Delta_k = n_{k+1} - n_k$ and recall that $\eta_n = (1/n)\sum_{i=1}^n (f(\boldsymbol{x}_i) - \mu(f))$. We begin with the identity,

$$\frac{1}{\Delta_k}\sum_{i=n_k+1}^{n_{k+1}} (f(\boldsymbol{x}_i) - \mu) = \frac{1}{\Delta_k}\big(n_{k+1}\eta_{n_{k+1}} - n_k\eta_{n_k}\big). \qquad (8)$$

The expected square of the left side of (8) is no smaller than $m^2/\Delta_k^{2\alpha}$. The expected square of the right side of (8) is

$$\begin{aligned}
&\frac{1}{\Delta_k^2}\Big[n_{k+1}^2\mathbb{E}(\eta_{n_{k+1}}^2) - 2n_{k+1}n_k\mathbb{E}(\eta_{n_{k+1}}\eta_{n_k}) + n_k^2\mathbb{E}(\eta_{n_k}^2)\Big]\\
&\le \frac{1}{\Delta_k^2}\Big[n_{k+1}^2\mathbb{E}(\eta_{n_{k+1}}^2) + 2n_{k+1}n_k\sqrt{\mathbb{E}(\eta_{n_{k+1}}^2)\mathbb{E}(\eta_{n_k}^2)} + n_k^2\mathbb{E}(\eta_{n_k}^2)\Big]\\
&\le \frac{M^2}{\Delta_k^2}\Big[n_{k+1}^{2(1-\alpha)} + 2n_{k+1}^{1-\alpha}n_k^{1-\alpha} + n_k^{2(1-\alpha)}\Big].
\end{aligned}$$

As a result,

$$\frac{m^2}{\Delta_k^{2\alpha}} \le \frac{M^2}{\Delta_k^2}\Big[n_{k+1}^{1-\alpha} + n_k^{1-\alpha}\Big]^2. \qquad (9)$$

Taking the square root of both sides of (9) and rearranging, yields (4), from which the theorem follows just as it did for the worst case analysis. □

5 Discussion

We have found a constraint on the spacings of a rate-optimal equal-weight extensible MC or QMC sequence. The constraint only applies when the convergence is better than $O(1/n)$.

We can use that constraint in reverse as follows. Suppose that a rate-optimal sequence n_k includes sample sizes with $n_{k+1}/n_k = \rho \in (1,2)$ and attains the error rate $O(n^{-\alpha})$ for $\alpha > 1$. Then from (3) we obtain

$$\frac{M}{m} \ge \frac{(1+\rho^{1-\alpha})^{-1/(\alpha-1)}}{\rho - 1}.$$

As we approach an arithmetic progression by letting $\rho \downarrow 1$, the inefficiency in the constant factor increases without bound.

The constraint only applies to rate-optimal sequences. In particular it does not apply to an extensible sequence that may be inefficient by a logarithmic factor.

Acknowledgments

I thank Alex Kreinin and Sergei Kucherenko for pointing out to me the elegant argument in the Appendix of Sobol' (1998), and Erich Novak for sharing his slides from MCQMC 2014. This work was supported by grants DMS-0906056 and DMS-1407397 of the U.S. National Science Foundation.

References

Ackleh, A. S., Allen, E. J., Kearfott, R. B., and Seshaiyer, P. (2009). *Classical and modern numerical analysis: Theory, methods and practice*. CRC Press, Boca Raton, FL.

Brent, R. (1973). *Algorithms for Minimization without Derivatives*. Prentice-Hall, Englewood Cliffs, NJ.

Dick, J. (2011). Higher order scrambled digital nets achieve the optimal rate of the root mean square error for smooth integrands. *The Annals of Statistics*, 39(3):1372–1398.

Hammersley, J. M. and Morton, K. W. (1956). A new Monte Carlo technique: antithetic variates. *Mathematical proceedings of the Cambridge philosophical society*, 52(3):449–475.

Hickernell, F. J., Kritzer, P., Kuo, F. Y., and Nuyens, D. (2012). Weighted compound integration rules with higher order convergence for all n. *Numerical Algorithms*, 59(2):161–183.

Kelley, C. T. (1999). *Iterative methods for optimization*, volume 18. Siam, Philadelphia.

Novak, E. and Wozniakowski, H. (2010). *Tractability of multivariate problems*, volume II: standard information for functionals. European Mathematical Society, Zurich.

Sobol', I. M. (1993). Asymmetric convergence of approximations of the Monte Carlo method. *Computational Mathematics and Mathematical Physics*, 33(10):1391–1396.

Sobol', I. M. (1998). On quasi-Monte Carlo integrations. *Mathematics and Computers in Simulation*, 47:103–112.