

Super-polynomial accuracy of median of means

Zexin Pan & Art B. Owen
Stanford University

Annotated I

These are slides that I presented at MCQMC 2022 in Linz after Zexin Pan was unable to attend. That's a pity. He would have interacted well with much that took place.

The project began with the question about what would happen if base 2 generator matrices were chosen completely at random. The matrices could be bad (e.g., all zeros) but they could also be good (e.g., Niederreiter-Xing) and as it turns out good choices are much more common than bad ones. Better than random generator matrices is to use a random Matousek scramble. Those also are mostly good but every once in a while bad. If we take the median of a modest number of replicates the result can be much more accurate than taking the mean.

Annotated II

We choose some example functions. For instance $x \exp(x)$ is analytic as our theory requires. It is not highly oscillatory which would make it less likely that we see the asymptotic behavior in a modest sample size. It is a ‘positive control’. However, it is not symmetric or anti-symmetric or a polynomial or trigonometric polynomial or specially sensitive to dyadic numbers. Those are things that might make it artificially easy.

We also use $\max(0, (x - 1/3))$. This is a harder function than our theory assumes because it is not smooth. Also the value $1/3$ has a slowly converging base 2 expansion which we think is a disadvantage for Sobol’ points.

Zexin Pan



- Sobol' Gains are powers of 2
- Where are the logs?
There are functions in BVHK with error $> c \log(n)^{(d-1)/2-\epsilon}/n$ infinitely often
E. Novak \rightarrow Traub, Wasilkowski, Wozniakowski (1988) \rightarrow unpublished Trojan
- What if Sobol' had used completely random generator matrices?

Idea in brief

Usual randomized QMC:

- Randomize via [Matousek \(1998\)](#)
- Replicate R times

$$\hat{\mu}_r = \frac{1}{n} \sum_{i=0}^{n-1} f(\mathbf{x}_i^{(r)}), \quad r = 1, \dots, R$$

$$\hat{\mu} = \frac{1}{R} \sum_{r=1}^R \hat{\mu}_r$$

Our proposal

$$\hat{\mu} = \mathbf{median}(\hat{\mu}_1, \dots, \hat{\mu}_R)$$

Because

Most of the $\hat{\mu}_r$ are very good

A few are very bad outliers

Median of means

Classic method in theoretical computer science

Jerrum, Valiant, Vazirani (1986), Lecué & Lerasle (2020)

Uses in information based complexity

Kunsch, Novak, Rudolf (2019)

Recent uses in (R)QMC

Goda, L'Ecuyer (2022) [Random lattices]

Hofstadler, Rudolf (2022) [Laws of large numbers]

Gobet, Lerasle, Métivier (2022) [Robust RQMC]

Superpolynomial accuracy

For $d = 1$ and $(0, m, 1)$ -net in base $b = 2$

Analytic f

error like $O(n^{-c \log_2(n)})$

$c < 3 \log(2)/\pi^2 \approx 0.21$

$f^{(\alpha)}$ is Hölder- λ

error like $O(n^{-\alpha-\lambda+\epsilon})$

P & O (2021) [arXiv:2111.12676](https://arxiv.org/abs/2111.12676)

Conjecture

Analytic f

Error like $O(n^{-c \log_2(n)/d})$ for $d \geq 1$

Work in progress.

Algorithm for $d = 1$

$$\mathbb{N}_0 \longrightarrow \{0, 1\}^m \longrightarrow [0, 1)$$

$$i = \sum_{k=1}^m i_k 2^{k-1} \quad \vec{i} = (i_1, i_2, \dots, i_m)^\top \in \{0, 1\}^m$$

$$a = \sum_{k=1}^{\infty} a_k 2^{-k} \quad \vec{a} = (a_1, a_2, \dots, a_m)^\top \in \{0, 1\}^m$$

Unscrambled net

$a_0, a_1, \dots, a_{2^m-1}$ with

$$\vec{a}_i = C \vec{i} \pmod{2}$$

$$C \in \{0, 1\}^{m \times m}$$

C has full rank over \mathbb{Z}_2

$C = I_m$ for van Der Corput

Linearly scrambled net

$$\vec{x}_i = M\vec{a}_i = MC\vec{i} \text{ mod } 2$$

$$M = \begin{pmatrix} 1 & & & & & & \\ u & 1 & & & & & \\ u & u & 1 & & & & \\ u & u & u & 1 & & & \\ \vdots & \vdots & \vdots & \vdots & \ddots & & \\ u & u & u & u & \cdots & 1 & \\ u & u & u & u & \cdots & u & \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \\ u & u & u & u & \cdots & u & \end{pmatrix} \in \{0, 1\}^{E \times m} \quad E \geq m$$

$$u \sim \text{iid } \mathbf{U}\{0, 1\}$$

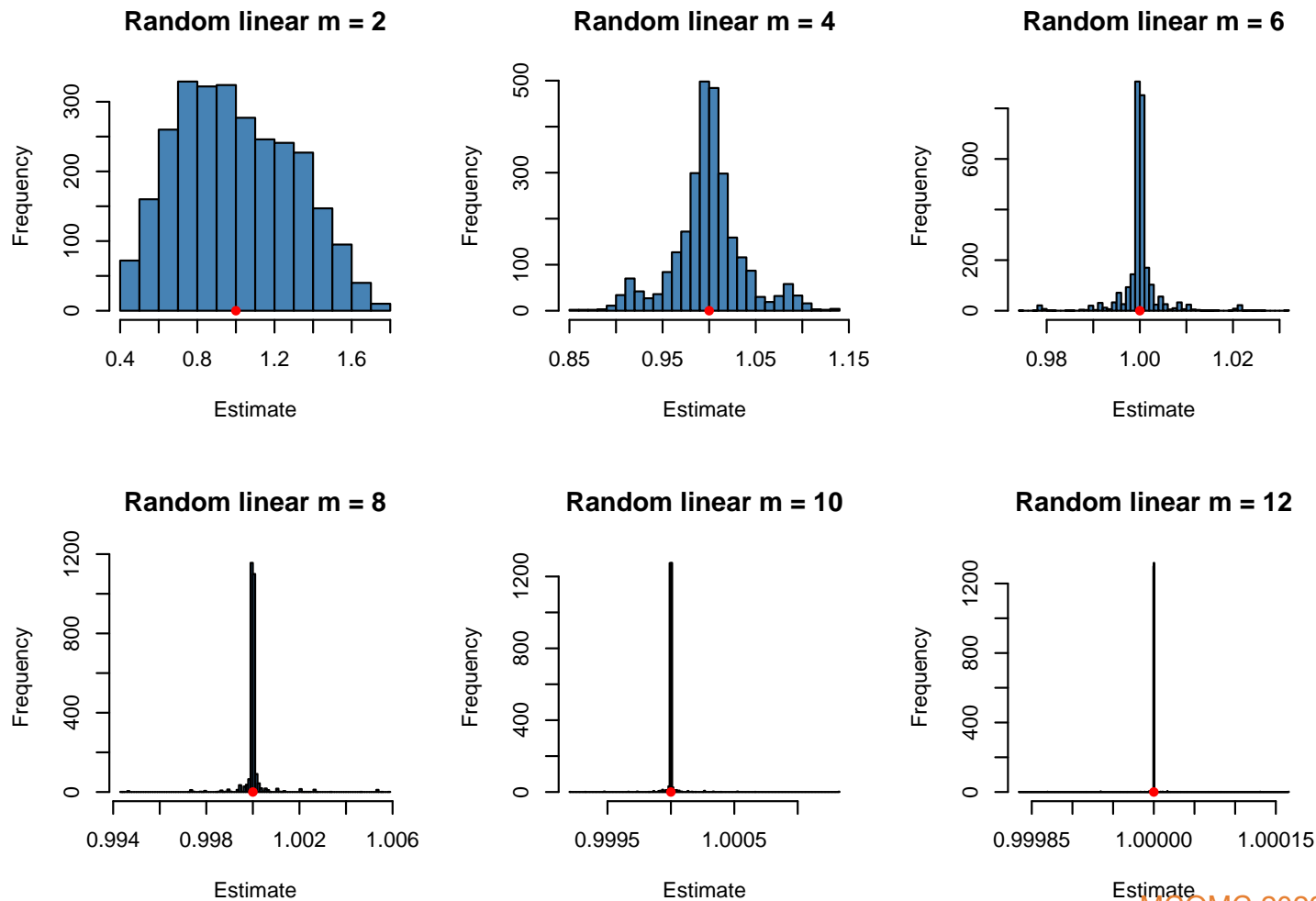
Apply digital shift

$$\vec{x}_i = MC\vec{i} + \vec{D} \text{ mod } 2 \quad D_k \stackrel{\text{iid}}{\sim} \mathbf{U}\{0, 1\}$$

$$\implies \text{each } x_i \sim \mathbf{U}[0, 1] \quad \text{for } E = \infty$$

$$\int_0^1 x e^x dx = 1$$

$n = 2^m$ random linear **Same** shift D for all n points



Precision

$M \in \{0, 1\}^{E \times m}$ produces x_i to E bits

Infinite precision $E = \infty$ and $M \in \{0, 1\}^{\infty \times m}$

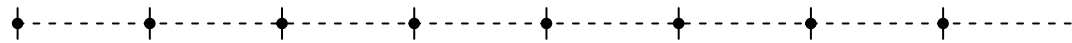
Produces $\hat{\mu}_\infty$

based on $n = 2^m$ points

Stratification

Without randomization we get

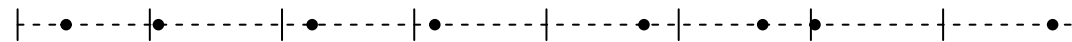
$$i/n \text{ for } 0 \leq i < n$$



in some order

With randomization we get one point in

$$[i/n, (i + 1)/n)$$



each $0 \leq i < n$

Bottleneck

Maybe


$$M(m+1, :) = (0, 0, \dots, 0) = \mathbf{0} \in \{0, 1\}^m$$


Then

$$x_{0,m+1} = x_{1,m+1} = \dots = x_{n-1,m+1} = D_{m+1}$$

This has probability $2^{-m} = 1/n$

Consequence

$D_{m+1} = 0 \implies$ all points in $[\frac{i}{n}, \frac{i+1/2}{n})$'s 

$D_{m+1} = 1 \implies$ all points in $[\frac{i+1/2}{n}, \frac{i+1}{n})$'s 

\implies Half of $[0, 1)$ is empty

$$\mathbb{E}(|\hat{\mu} - \mu| \mid M(m+1, :) = \mathbf{0}) = O(1/n) \quad (\text{Taylor})$$

Summary

This rare event contributes $O(1/n^3)$ to $\text{Var}(\hat{\mu}_\infty)$

Cannot reduce by more smoothness

More trouble

Problematic whenever

$$\sum_{\ell \in L} M(\ell, :) = \mathbf{0} \quad \text{for some } L \subseteq \mathbb{N}$$

Can't happen if $L \subseteq \{1, 2, \dots, m\}$ because M has full rank

Probability 2^{-m} otherwise: “last row” of $M(L, :)$ does it

That is

$$\Pr\left(\sum_{\ell \in L} M(\ell, :) = \mathbf{0}\right) = \begin{cases} 0, & L \subseteq \{1, 2, \dots, m\} \\ 2^{-m}, & L \not\subseteq \{1, 2, \dots, m\} \end{cases}$$

Magnitudes

$$|L| = \text{card}(L) \quad \|L\|_1 = \sum_{\ell \in L} \ell$$

Decomposition

$$\mathcal{L} := \{L \subset \mathbb{N} \mid 0 < |L| < \infty\}$$

Let f be analytic on $[0, 1]$ with

$$\left| f^{(k)}\left(\frac{1}{2}\right) \right| \leq A \alpha^k k!, \quad A < \infty, \quad \alpha < 2, \quad \text{all } k \in \mathbb{N}$$

$C \in \{0, 1\}^{m \times m}$ full rank

Theorem 3.1

$$\hat{\mu}_\infty - \mu = \sum_{L \in \mathcal{L}} \mathbf{1} \left\{ \sum_{\ell \in L} M(\ell, :) = \mathbf{0} \right\} S_L(D) 2^{-\|L\|_1} B_L$$

$$S_L(D) = \prod_{\ell \in L} (-1)^{D_\ell}$$

$$|B_L| \leq 6A(|L| - 1)! \left(\frac{\alpha/2}{1 - \alpha/2} \right)^{|L|}$$

The B_L involve Walsh coefficients of f

Significant use of [Dick & Pillichshammer \(2010\)](#)

Norm of L

The bound includes $2^{-\|L\|_1}$

Things are bad for **small** $\|L\|_1$

worst is $L = \{m + 1\}$ (bottleneck example)

Fortunately

Not too many $L \in \mathcal{L}$ with $\|L\|_1 = N$ small

$p(N) = \#$ ways to partition $N \in \mathbb{N}$ into a sum of natural numbers.

Analytic combinatorics

Hardy & Ramanujan (1918) show that

$$p(N) \sim \frac{1}{N 4\sqrt{3}} \exp\left(\pi \left(\frac{2N}{3}\right)^{1/2}\right) = o(2^N)$$

Bidar (2012) reduces this for **distinct** natural numbers

$$\frac{\pi}{2\sqrt{3N}} \exp\left(\pi \left(\frac{N}{3}\right)^{1/2}\right)$$

Trouble spots

Let $m = 3$

$\ L\ _1$	L
4	{4}
5	{5}, {1, 4}
6	{6}, {1, 5}, {2, 4}
7	{7}, {1, 6}, {2, 5}, {1, 2, 4}
8	{8}, {1, 7}, {2, 6}, {3, 5}, {1, 2, 5}, {1, 3, 4}
\vdots	\vdots

Weighted by $2^{-\|L\|_1}$

The rows grow in cardinality but they do not double in size.

Controlling bad $L \in \mathcal{L}$

For $\lambda = 3(\log(2)/\pi)^2 \approx 0.146$

$$|\{L \in \mathcal{L} \mid \|L\|_1 \leq \lambda m^2\}| \sim \frac{3^{1/4}}{2\pi\lambda^{1/4}} \frac{2^m}{\sqrt{m}}$$

$$\Pr\left(\sum_{\ell \in L} M(\ell, :) = \mathbf{0}\right) = 2^{-m}$$

Expect $O\left(\frac{1}{\sqrt{m}}\right)$ problematic sets L

Quality of estimate

Bound $\Pr(|\hat{\mu}_\infty - \mu| \gg 2^{-\lambda m^2})$

Then median over infinite replicates has

$$|\text{median}(\hat{\mu}_\infty) - \mu| = o(2^{-(\lambda-\epsilon)m^2})$$

Sample median

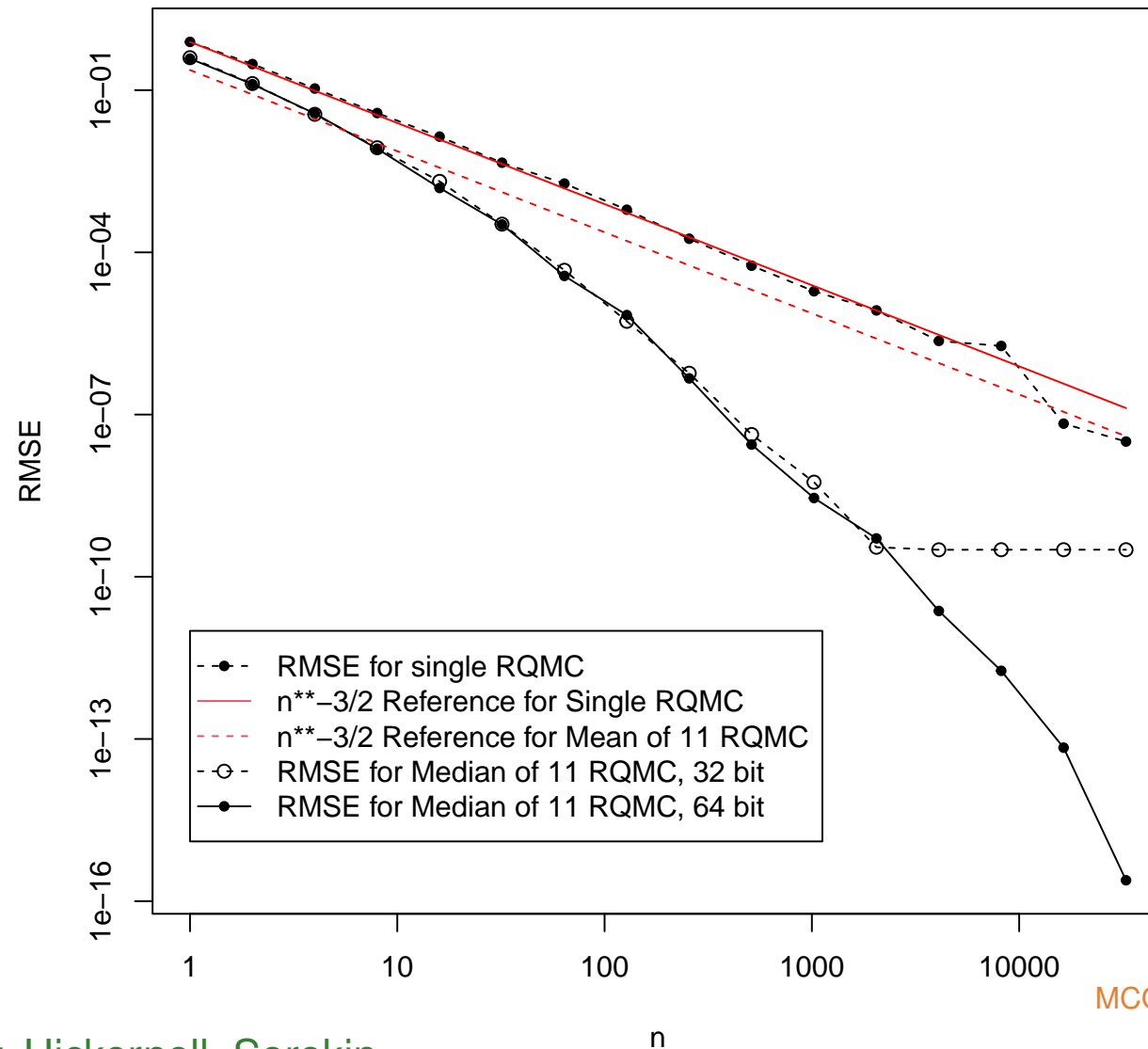
$$|\text{median}(\hat{\mu}_{\infty,1}, \hat{\mu}_{\infty,2}, \dots, \hat{\mu}_{\infty,2k-1}) - \text{median}(\hat{\mu}_\infty)|$$

letting $k = \Omega(m) = \Omega(\log_2(n))$.

Cost is now $O(n \log(n))$

Example $d = 1$

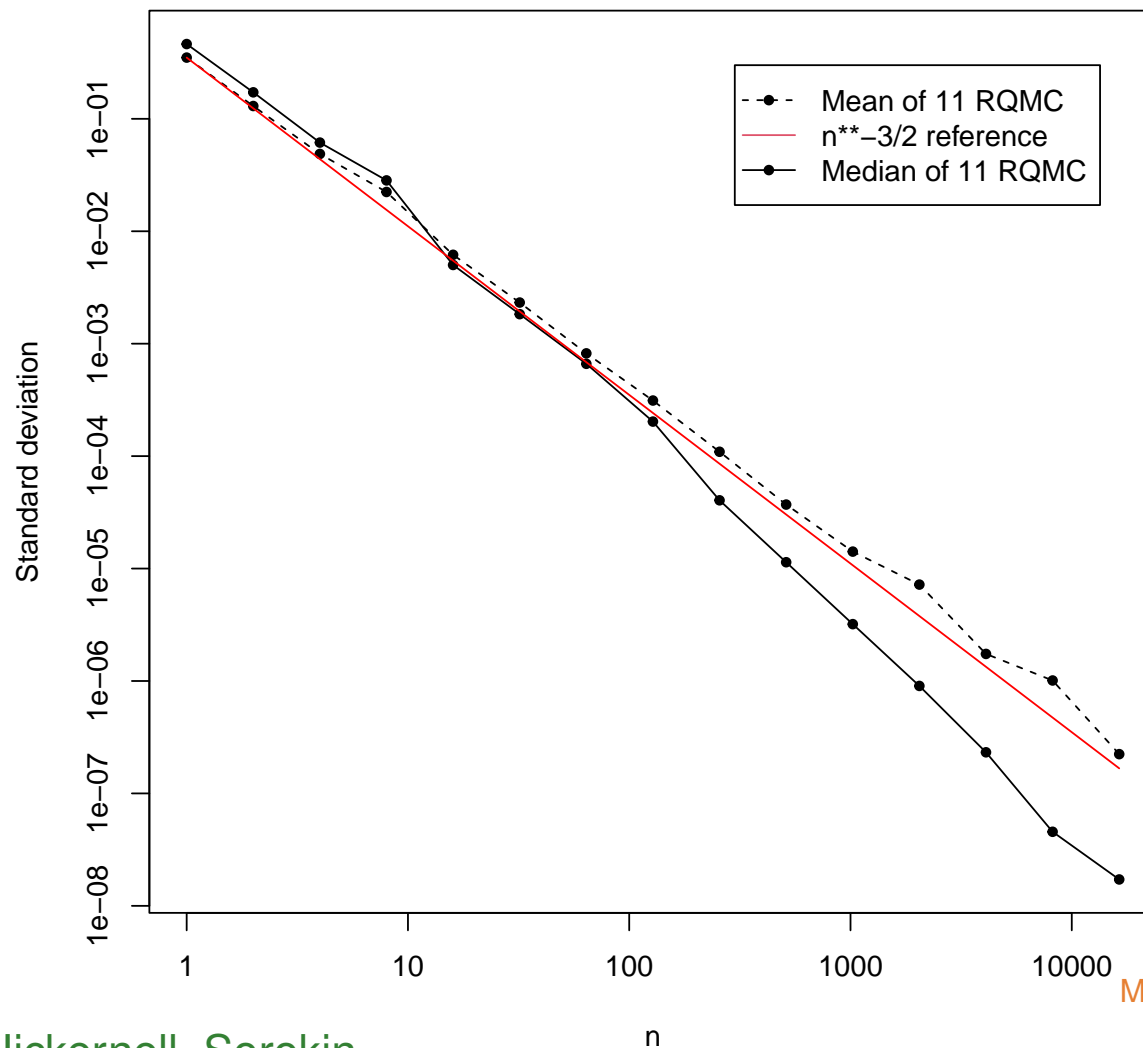
RMSE for $f(x) = x \cdot \exp(x)$



Example $d = 6$

Surjanovic & Bingham

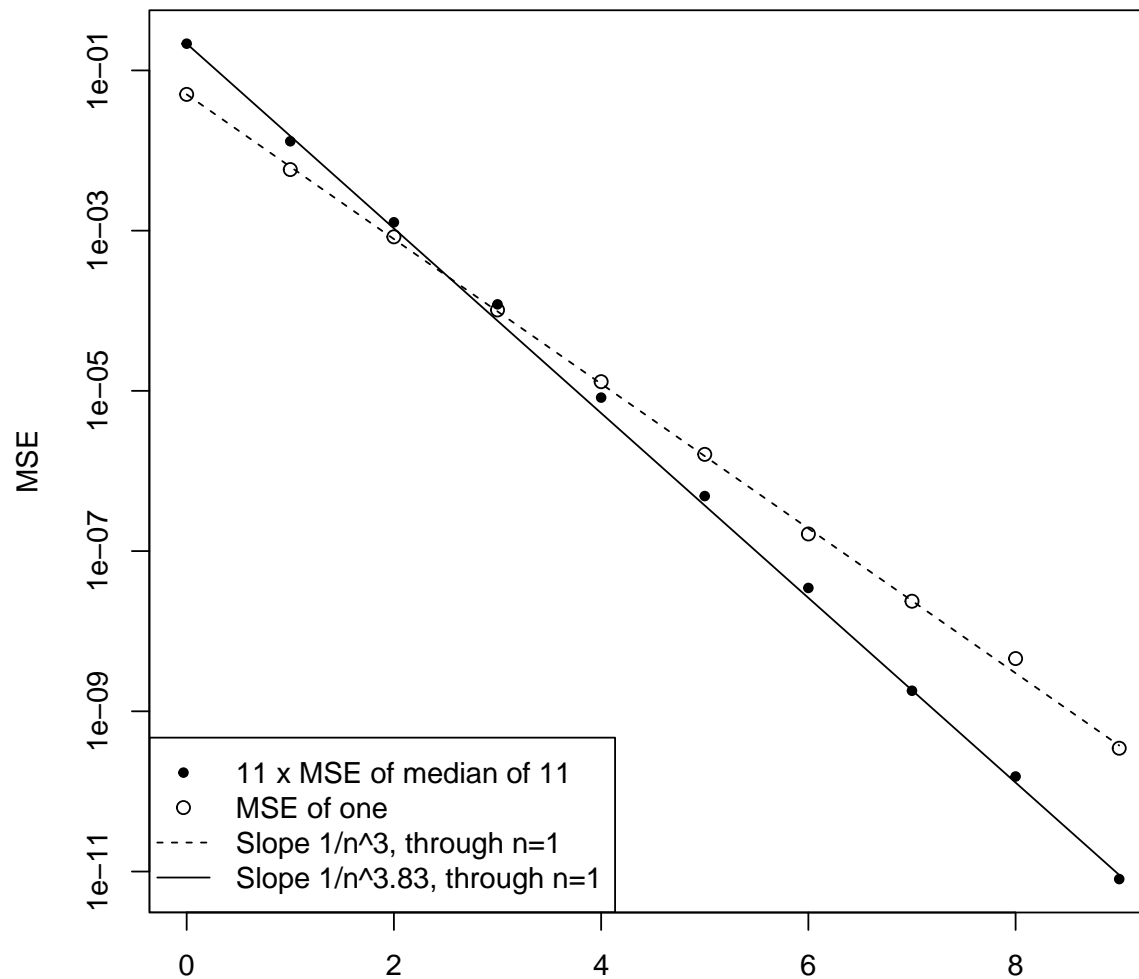
Standard Deviation for OTL Circuit Integrand



MCQMC 2022, Linz, July 2022

$$\int_0^1 (x - 1/3)_+ dx = 2/9$$

Random linear scramble with digital shift
 Kink: $f(x) = \max(x-1/3, 0) - 2/9$



Challenge

It is **hard** to estimate

$$\text{Var}(\text{median}(\hat{\mu}_1, \dots, \hat{\mu}_{2k-1}))$$

It is **easy** to get a confidence interval for

$$\text{median}(\text{Distribution of } \hat{\mu}_r)$$

But we **want** to know about

$$\int f(\mathbf{x}) \, d\mathbf{x} = \text{mean}(\text{Distribution of } \hat{\mu}_r)$$

Thanks

- Zexin Pan who could not be here today
- Fred Hickernell and Aleksei Sorokin for QMCPy help
- NSF IIS-1837931
- Aicke Hinrichs, Peter Kritzer, Friedrich Pillichshammer