Quasi-Monte Carlo

Art B. Owen

Stanford University

These slides are from a series of four lectures given at the Johann Radon Institute for Computational and Applied Mathematics (RICAM) held on March 24 and March 25 2021.

It was an honor to be asked to present on quasi-Monte Carlo (QMC) sampling in Austria, from where so much of QMC comes and has come. The talks were virtual; I would have otherwise made sure to get some Linzertorte. That will have to wait.

1. Quasi-Monte Carlo
2. Randomized Quasi-Monte Carlo
3. Quasi-Monte Carlo Outside the unit Cube
4. Variable Importance and Sobol' indices

A small number of corrections have been made since then.

More accurate title

$\frac{1}{2}$ of QMC

I will leave out most of lattice methods

focussing on digital nets and related constructions.

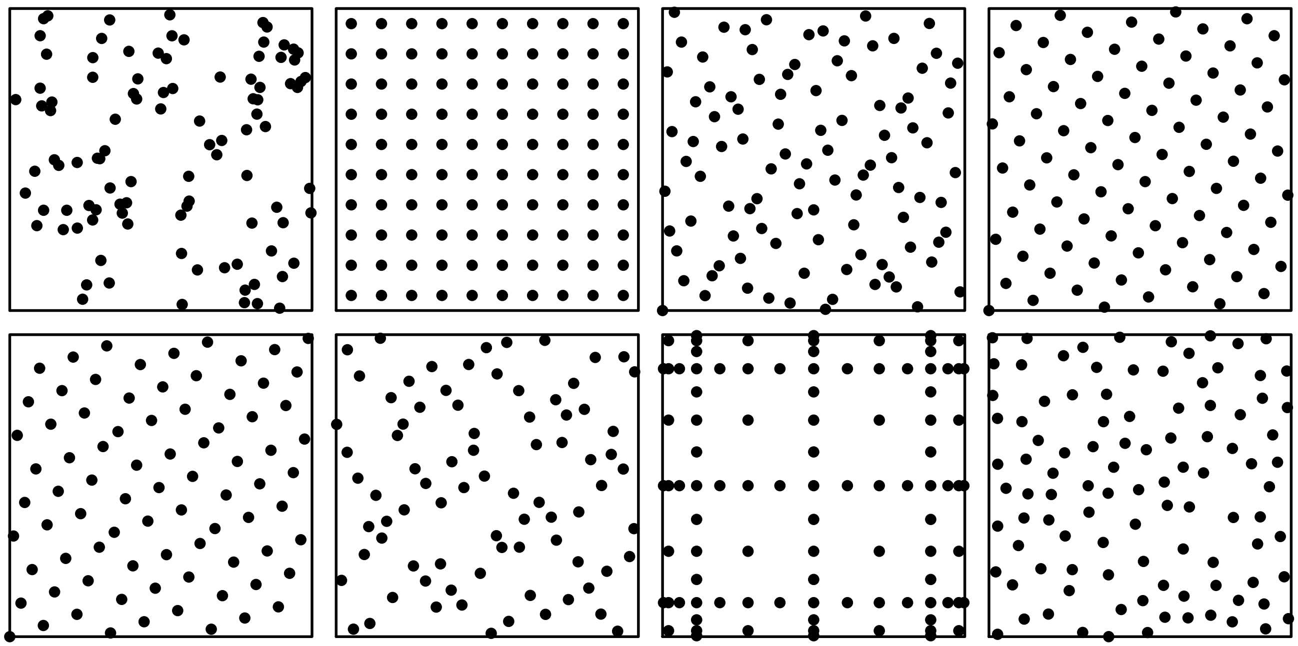
What is Quasi-Monte Carlo?

It is a method of sampling.

Mostly for integration over a continuum.

Designed to be better than random (Monte Carlo).

Points in boxes



We get a discrete approximation to a continuous problem.

Why QMC?

1) It solves real world problems to help people

2) The mathematics is interesting (curiosity) or elegant (aesthetics)

These are not separate

Example problems

1) Approximate $\mu = \int_{[0,1]^d} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$

2) Simulate some phenomenon many times to ‘see what happens’

These are not separate, either

How to integrate

Choose the first thing that works in this order

1) closed form expression
2) symbolic mathematics (e.g., Maple, Mathematica, Sage)
3) classical quadrature (e.g., Simpson's rule)
4) QMC or randomized QMC (RQMC)
5) Monte Carlo
6) Markov chain Monte Carlo (MCMC)
7) approximate MCMC

Sparse grids might be in position 3.5 or 4.

Maybe RQMC can absorb MC

Classic rules

$$I \equiv \int_a^b f(x)\,\mathrm{d}x \doteq (b-a)\sum_{i=0}^{n} w_i f(x_i)$$

for $a \leqslant x_0 < x_1 < x_2 < \cdots < x_n \leqslant b$ and $w_i \in \mathbb{R}$

Simpson's rule

$$\hat{I} = \frac{b-a}{3n}\Big[f(x_0)+4f(x_1)+2f(x_2)+4f(x_3)+\cdots+4f(x_{n-1})+f(x_n)\Big]$$

If $f^{(4)}$ continuous

$$|\hat{I} - I| \leqslant \frac{(b-a)^4}{180n^4}\max_{a\leqslant x\leqslant b}|f^{(4)}(x)| = O(n^{-4})$$

Dimension $d > 1$

Fubini:

$\int_0^1 \int_0^1 f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_0^1 I(y)\,\mathrm{d}y$ for $I(y) = \int_0^1 f(x,y)\,\mathrm{d}x$

Take $n \times n$ grid

Simpson's for x given y (inner)

then for y (outer)

Generally take $N = n \times n \times \cdots \times n$ point grid $\boldsymbol{x}_i \in [0,1]^d$

Product of Simpson's rules

At best $O(n^{-4}) = O(N^{-4/d})$

Imagine one dimensional integral of perfect $d-1$ dimensional integrals

Higher order

$O(n^{-r})$ over $[0,1] \implies$ at best $O(n^{-r/d})$ in dimension d.

Bakhvalov

For $r \geqslant 1$ and $M > 0$ define "nice" functions

$$\mathcal{F}_r^M = \left\{ f\colon [0,1]^d \to \mathbb{R} \;\middle|\; \left| \frac{\partial f(\boldsymbol{x})}{\partial x_1^{\alpha_1} \cdots x_d^{\alpha_d}} \right| \leqslant M, \quad \alpha_j \geqslant 0, \quad \sum_{j=1}^d \alpha_j = r \right\}$$

Curse of dimension

For any rule like $\hat{I} = \sum_{i=1}^n w_i f(\boldsymbol{x}_i)$ there is $k > 0$ and $f \in \mathcal{F}_r^M$ with

$$|\hat{I} - I| \geqslant k n^{-r/d}.$$

Consequence

Large $d \implies$ no good rule for **all** $f \in \mathcal{F}_r^M$

Fooling functions

$f(\boldsymbol{x}_i) = 0$ for $i = 1, \dots, n$

See monographs Novak, Wozniakowski

Monte Carlo

Integrals as expectations

$$\mu \equiv \mathbb{E}(f(\boldsymbol{x})) = \int_{\mathbb{R}^d} f(\boldsymbol{x})p(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

Take $\boldsymbol{x}_i \overset{\text{iid}}{\sim} p$ and

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

If $\mathbb{E}(f(\boldsymbol{x})^2) < \infty$

Then $\mathbb{E}\bigl((\hat{\mu}-\mu)^2\bigr) = O\Bigl(\frac{1}{n}\Bigr)$

Root mean squared error (RMSE) is $O(n^{-1/2})$ any d

Does not need $r > 0$ (smoothness)

Puzzle: does this break Bakhvalov's curse of dimension?

Quasi-Monte Carlo

We want to do better than RMSE $= O(n^{-1/2})$.

We begin with

$$\mu = \int_{[0,1]^d} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \quad \text{i.e. } p = \mathbf{U}[0,1]^d$$

For general $\boldsymbol{x} \sim p$

Transformations $\psi(\cdot)$ from Devroye (1986)

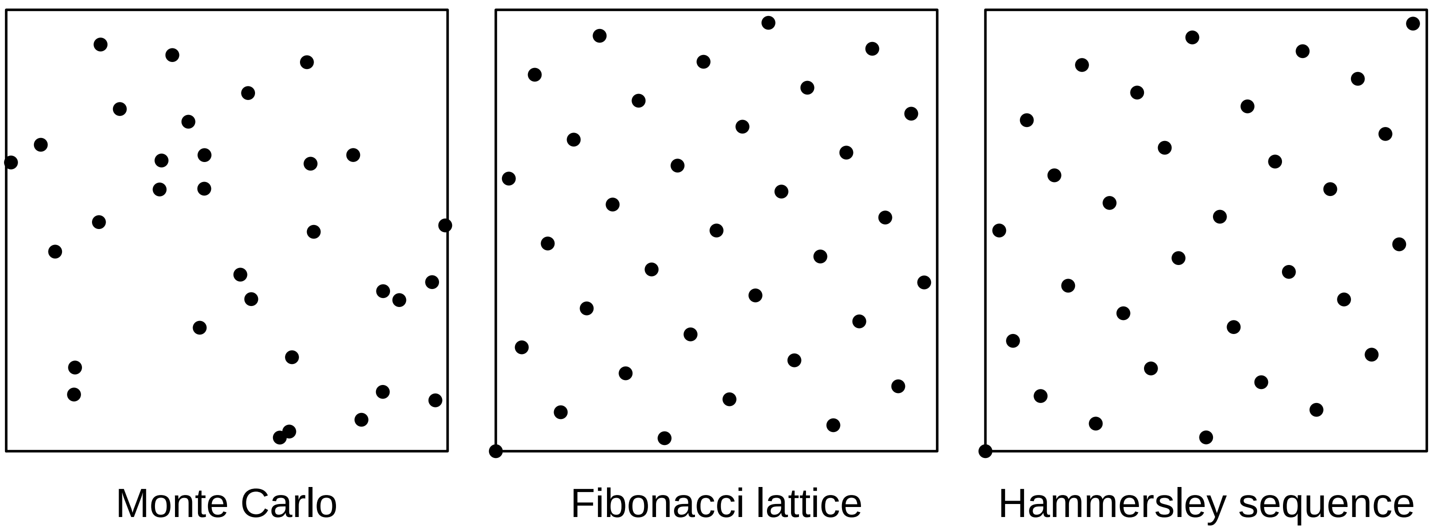
$$\boldsymbol{x} \sim \mathbf{U}[0,1]^d \implies \psi(\boldsymbol{x}) \sim p$$

$$\begin{aligned}\int_{\mathbb{R}^s} g(\boldsymbol{x})p(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} &= \int_{[0,1]^d} g(\psi(\boldsymbol{x}))\,\mathrm{d}\boldsymbol{x} \\ &= \int_{[0,1]^d} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \qquad f = g \circ \psi\end{aligned}$$

Caution: ψ can make f irregular

Illustration

MC and two QMC methods in the unit square



MC points always have clusters and gaps. What is random is where they appear.

QMC points avoid clusters and gaps to the extent that mathematics permits.

Idea let's pick points that are more equally spread out.

Measuring uniformity

We need a way to verify that the points $\boldsymbol{x}_i$ are ‘spread out’ in $[0,1]^d$.

The most fruitful way is to show that

$$\mathbf{U}\{\boldsymbol{x}_1,\boldsymbol{x}_2,\dots,\boldsymbol{x}_n\} \doteq \mathbf{U}[0,1]^d$$

Discrepancy

A discrepancy is a distance $\|F-\hat{F}_n\|$ between measures $F=\mathbf{U}[0,1]^d$ and $\hat{F}_n=\mathbf{U}\{\boldsymbol{x}_1,\boldsymbol{x}_2,\dots,\boldsymbol{x}_n\}$.

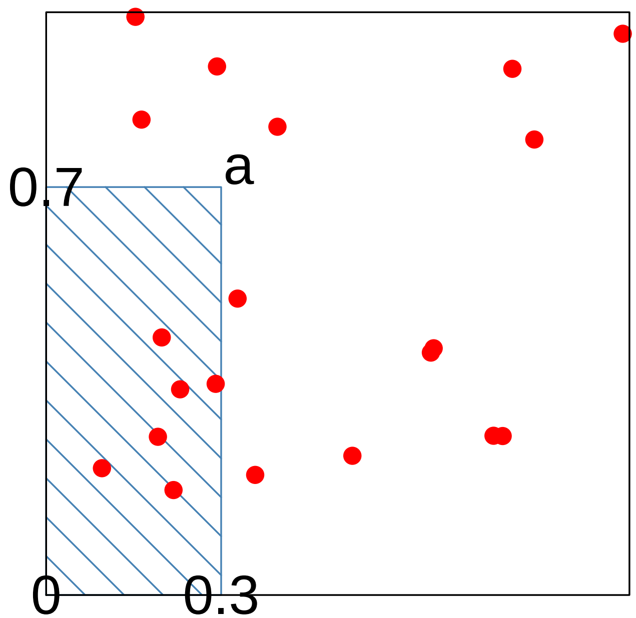
There are many discrepancies.

As integrals

$$\mu=\int f(\boldsymbol{x})\,\mathrm{d}F(\boldsymbol{x}) \qquad \hat{\mu}=\int f(\boldsymbol{x})\,\mathrm{d}\hat{F}_n(\boldsymbol{x})$$

so we want $\hat{F}_n \approx F$.

Local discrepancy



The box $[0, \boldsymbol{a})$ contains $6/20 =$ points and has $0.3 \times 0.7 = 0.21$ of the area.

$$\delta(\boldsymbol{a}) \equiv 0.30 - 0.21 = 0.09$$

Star discrepancy

$$D_n^* = D_n^*(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \sup_{\boldsymbol{a} \in [0,1)^d} |\delta(\boldsymbol{a})| = \|\delta\|_\infty$$

For $d = 1$ this is Kolmogorov-Smirnov distance

Discrepancies

$$D_n^* = \sup_{\boldsymbol{a}\in[0,1)^d} \left|\hat{F}_n([0,\boldsymbol{a})) - F([0,\boldsymbol{a}))\right|$$

$$D_n = \sup_{\boldsymbol{a},\boldsymbol{b}\in[0,1)^d} \left|\hat{F}_n([\boldsymbol{a},\boldsymbol{b})) - F([\boldsymbol{a},\boldsymbol{b}))\right|$$

$$D_n^* \leqslant D_n \leqslant 2^d D_n^*$$

L^p discrepancies

$$D_n^{*p} = \left(\int_{[0,1)^d} |\delta(\boldsymbol{a})|^p \,\mathrm{d}\boldsymbol{a}\right)^{1/p}$$

e.g., Warnock uses $p = 2$

Also

Discrepancies over (triangles, rotated rectangles, balls $\cdots$ convex sets $\cdots$).

Beck, Chen, Schmidt, Brandolini, Travaglini, Colzani, Gigante, Cools, Pillards

Wrap-around discrepancies Hickernell

Best results are **only** for axis-aligned hyper-rectangles.

That's enough for good integration.

Discrepancy

Classical field, dates from Weyl (1916)

Recent book:

A Panorama of Discrepancy Theory (2014)

Chen, Srivastav, Travaglini

Chapter 9: Dick & Pillichshammer

Discrepancy and Quasi-Monte Carlo Integration

Discrepancy has beautiful math.

It is fundamental

Computable finite approximations to mathematical problems:

in the continuum

or in combinatorially large settings

QMC's law of large numbers

1) If f is Riemann integrable on $[0,1]^d$, and

2) $D_n^*(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n)\to 0$

Then

$$\frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{x}_i)\to\int_{[0,1]^d} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

How fast?

MC has the RMSE and central limit theorem.

QMC has the Koksma-Hlawka inequality.

Koksma-Hlawka theorem

$$\left|\frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{x}_i) - \int_{[0,1)^d} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right| \leqslant D_n^* \times V_{\mathrm{HK}}(f)$$

V_{HK} is the **total variation** in the sense of Hardy (1905) and Krause (1903)

Koksma's inequality

For $d = 1$ and f' continuous

$$|\mu - \hat{\mu}| = \cdots = \left|\int_0^1 \delta(x) f'(x)\,\mathrm{d}x\right| \leqslant \|\delta\|_\infty \|f'\|_1 = D_n^* \times V(f)$$

$V(f) = \|f'\|_1 = \int_0^1 |f'(x)|\,\mathrm{d}x$ ordinary total variation

Rates of convergence

Can get $D_n^* = O\Bigl(\dfrac{\log(n)^{d-1}}{n}\Bigr) = o(n^{-1+\epsilon})$ any $\epsilon > 0$.

Then if $V_{\mathrm{HK}}(f) < \infty$

$$|\hat\mu - \mu| = o(n^{-1+\epsilon}) \quad \text{vs} \quad O_p(n^{-1/2}) \text{ for MC}$$

What about those logs?

Maybe $\log(n)^{d-1}/n \gg 1/\sqrt{n}$

Low effective dimension (later) counters them

As do some randomizations (later)

Roth (1954)

$$D_n^* = o\Bigl(\frac{\log(n)^{(d-1)/2}}{n}\Bigr) \quad \text{is unattainable}$$

Gap between $\log(n)^{(d-1)/2}$ and $\log(n)^{d-1}$ subject to continued work.

E.g., Lacey, Bilyk

Tight and loose bounds

They are not mutually exclusive.

Koksma-Hlawka is tight

$$|\hat{\mu}-\mu| \leqslant (1-\epsilon)D_n^*(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n)\times V_{\mathrm{HK}}(f) \quad \textbf{fails} \text{ for some } f$$

Equality for a worst case function, e.g., $f' \doteq \pm\delta$.

Koksma-Hlawka is also very loose

It can greatly over-estimate actual error. Usually δ and f' are dissimilar.

$$\hat{\mu}-\mu = -\langle \delta, f'\rangle$$

Like Chebychev's inequality

$$\Pr\Bigl(|x-\mathbb{E}(x)| \geqslant k\sqrt{\mathrm{Var}(x)}\Bigr) \leqslant \frac{1}{k^2}$$

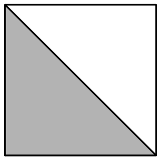
E.g., $\Pr\bigl(|\mathcal{N}(0,1)| \geqslant 10\bigr) \leqslant 0.01$ is loose.

Yes: $1.5\times 10^{-23} \leqslant 10^{-2}$

Variation

Hardy-Krause variation has surprises for us. O (2005)

$$f(x_1, x_2) = \begin{cases} 1, & x_1 + x_2 \leqslant 1/2 \\ 0, & \text{else} \end{cases}$$



$$V_{\mathrm{HK}}(f) = \infty \quad \text{on } [0,1]^2$$

$$V_{\mathrm{HK}}(f_\epsilon) < \infty, \quad \text{for some } f_\epsilon \text{ with } \|f - f_\epsilon\|_1 < \epsilon$$

Cusps, kinks, jumps

$f(\boldsymbol{x}) = \max(\theta_1^{\mathsf{T}} \boldsymbol{x}, \theta_2^{\mathsf{T}} \boldsymbol{x})$
generally has $V_{\mathrm{HK}}(f) = \infty$ with $d \geqslant 3$

QMC-friendly discontinuities

Axis parallel discontinuities may have $V_{\mathrm{HK}} < \infty$.

Used by e.g., X. Wang, I. Sloan, Z. He

Next

1) Famous first uses of MC
2) Famous first uses of QMC
3) QMC constructions

Landmark papers in MC

Some landmark papers where Monte Carlo was applied:

- Physics Metropolis et al. (1953)
- Discrete event simulation Tocher & Owen (1960)
- Chemistry (reaction equations) Gillespie (1977)
- Financial valuation Boyle (1977)
- Bootstrap resampling Efron (1979)
- Bayes (maybe 5 landmarks in early days of MCMC)
- Nonsmooth optimization Kirkpatrick et al. (1983)
- Computer graphics (path tracing) Kajiya (1988)

Landmark uses of QMC

- Particle transport methods in physics / medical imaging Jerome Spanier++
- Financial valuation, some early examples Paskov & Traub 1990s
- Graphical rendering Heinrich then Alex Keller++
- Solving PDEs in random environments

  Frances Kuo, Dirk Nuyens, Christoph Schwab++, 2015
- Particle methods Chopin & Gerber (2015)

The next landmark methods

Some strong candidate areas:

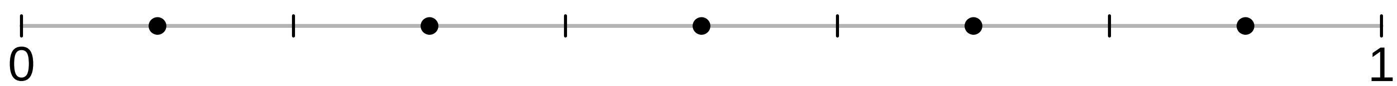
- machine learning
- Bayes
- uncertainty quantification (UQ)

These areas are already seeing QMC sampling.

Extensibility

For $d = 1$, the equisped points $x_i = (i - 1/2)/n$ have $D_n^* = \dfrac{1}{2n}$

Best possible.



But **where** do we put the $n\mathbf{+1}$**'st** point?

We cannot get $D_n^* = O(1/n)$ along a sequence $x_1, x_2, \ldots$.

Extensible sequences

Take first n points of $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \ldots, \boldsymbol{x}_n, \boldsymbol{x}_{n+1}, \boldsymbol{x}_{n+2}, \ldots$.

Then we can get $D_n^* = O((\log n)^{d}/n)$.

d not $d-1$

van der Corput

i				$\phi_2(i)$
1	1	0.1	$1/2$	0.5
2	10	0.01	$1/4$	0.25
3	11	0.11	$3/4$	0.75
4	100	0.001	$1/8$	0.125
5	101	0.101	$5/8$	0.625
6	110	0.011	$3/8$	0.375
7	111	0.111	$7/8$	0.875
8	1000	0.0001	$1/16$	0.0625
9	1001	0.1001	$9/16$	0.5625

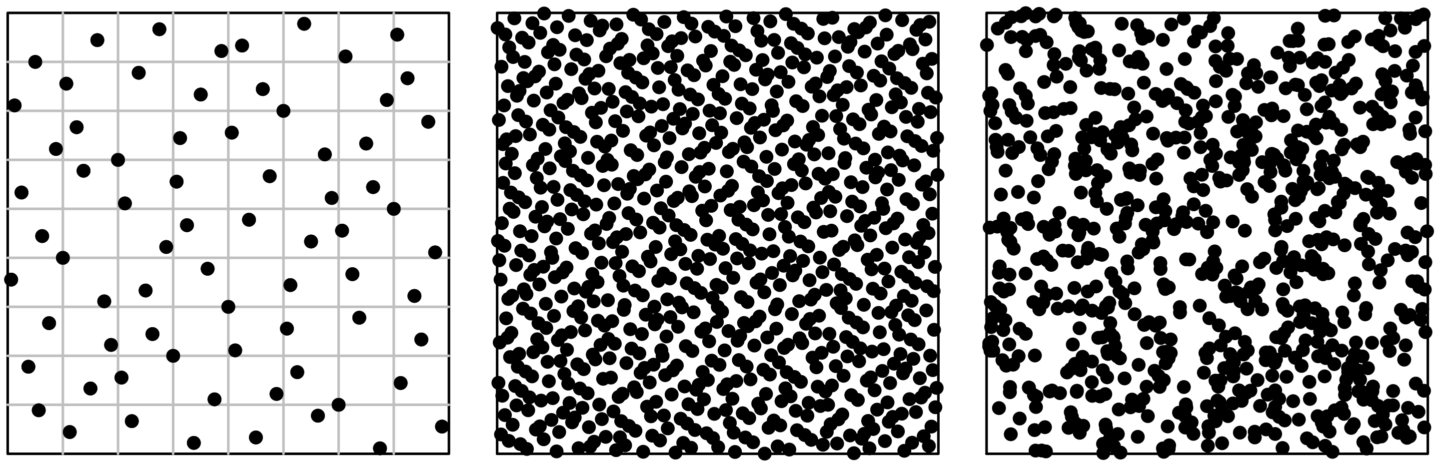
Take $x_i = \phi_2(i)$. Extensible with $D_n^* = O(\log(n)/n)$.

Commonly $x_i = \phi_2(i-1)$ starts at $x_1 = 0$.

Halton sequences

The van der Corput trick works for any base. Use bases $2, 3, 5, 7, \ldots$

Halton sequence in the unit square



72 Halton points — 864 Halton points — 864 random points

Via base b digital expansions

$$i = \sum_{k=0}^{K} b^k a_{ik} \qquad \to \qquad \phi_b(i) \equiv \sum_{k=0}^{K} b^{-1-k} a_{ik}$$

$$\boldsymbol{x}_i = (\phi_2(i), \phi_3(i), \ldots, \phi_p(i))$$

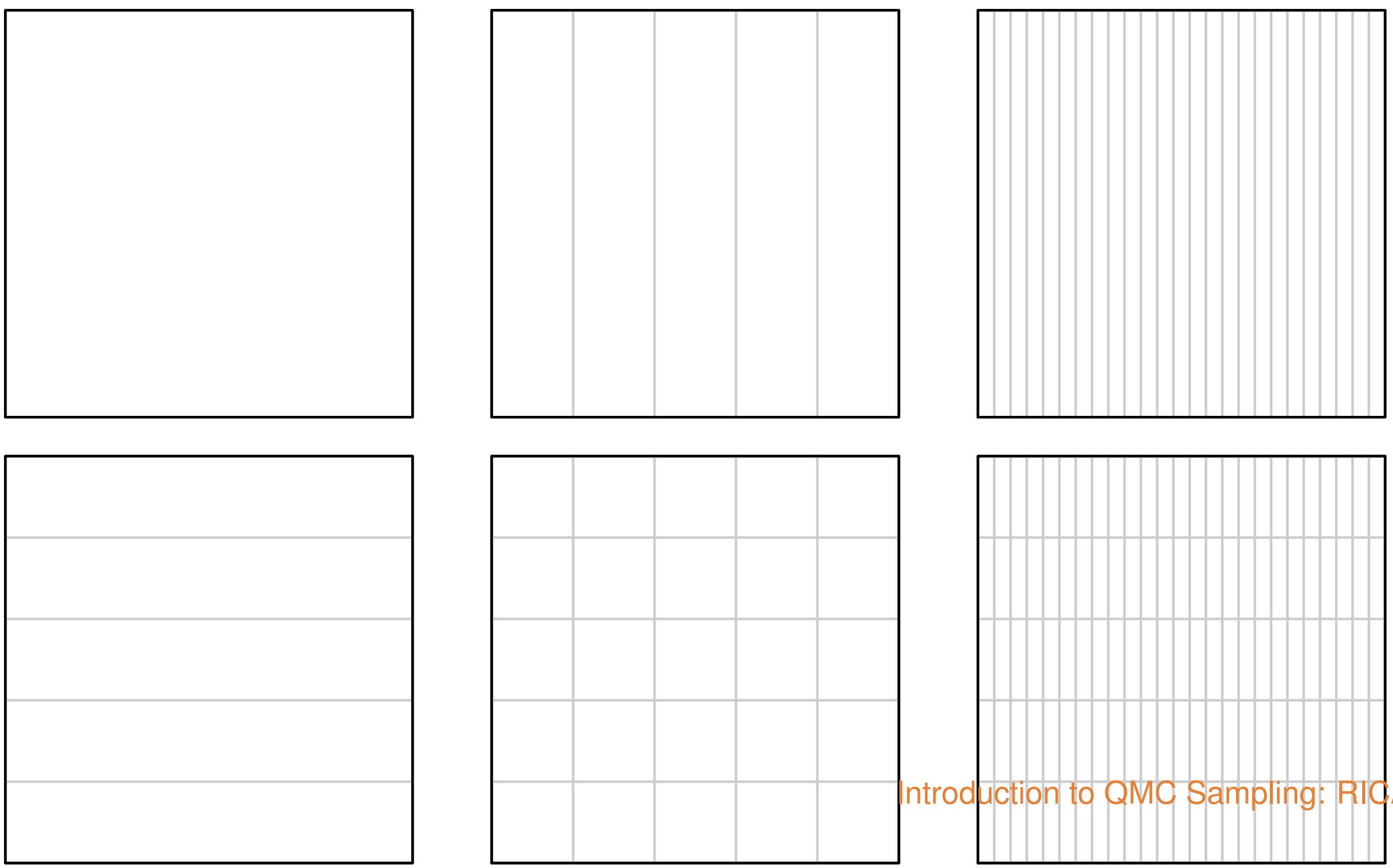
Digital nets

Halton sequences are balanced if n is a multiple of 2^a and 3^b and $5^c \ldots$

Digital nets use just one base $b \implies$ balance all margins equally.

Elementary intervals

Some elementary intervals in base 5



Digital nets

$$E = \prod_{j=1}^{s} \Bigl[\frac{a_j}{b^{k_j}}, \frac{a_j+1}{b^{k_j}}\Bigr), \quad 0 \leqslant a_j < b^{k_j}$$

$(0, m, s)$-net

$n = b^m$ points in $[0,1)^s$. If $\mathbf{vol}(E) = 1/n$ then E has one of the n points.

e.g. Faure (1982) points, prime base $b \geqslant s$

(t, m, s)-net

If E deserves b^t points it gets b^t points. Integer $t \geqslant 0$.

e.g. Sobol' (1967) and Niederreiter & Xing (1995) points $b = 2$

Smaller t is better (but a construction might not exist).

minT project

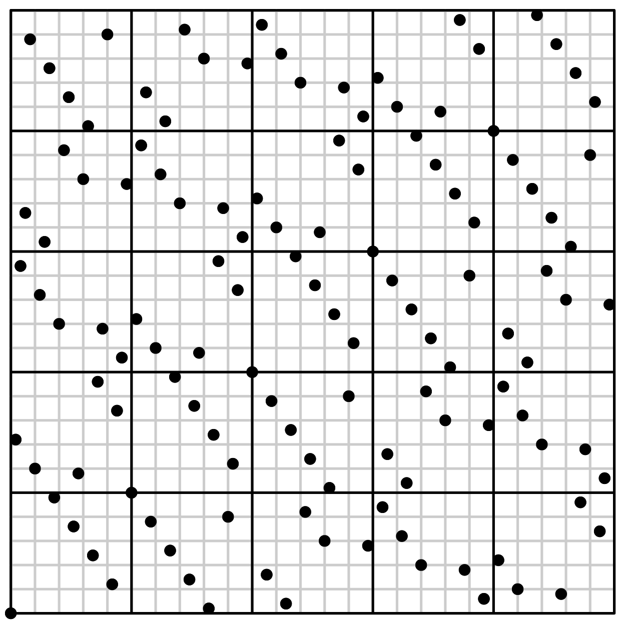
Schürer & Schmid (2006, 2009, 2010) give bounds on t given b, m and s

Monographs

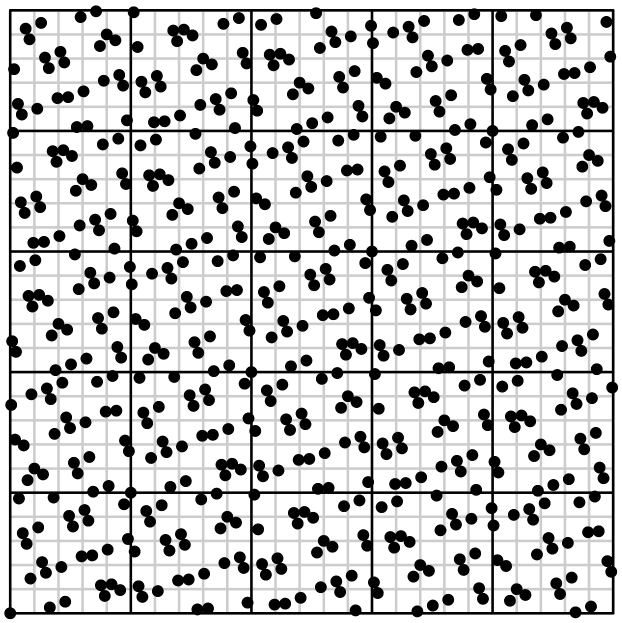
Niederreiter (1992), Dick & Pillichshammer (2010)

Example nets

Two digital nets in base 5



A (0,3,2) net



A (0,4,2) net

The $(0,4,2)$-net is a bivariate margin of a $(0,4,5)$-net.

The parent net has $5^4 = 625$ points in $[0,1)^5$.

It balances $43{,}750$ elementary intervals.

We should remove that diagonal striping artifact (later).

Digital net constructions

Write $i = \sum_{k=0}^{K} a_{ik} b^k$ (simplest for prime b) and let

$$\boldsymbol{x}_{i1} \equiv \begin{pmatrix} x_{i10} \\ x_{i11} \\ \vdots \\ x_{i1K} \end{pmatrix} = \begin{pmatrix} C_{11}^{(1)} & C_{12}^{(1)} & \dots & C_{1K}^{(1)} \\ C_{21}^{(1)} & C_{22}^{(1)} & \dots & C_{2K}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ C_{K1}^{(1)} & C_{K2}^{(1)} & \dots & C_{KK}^{(1)} \end{pmatrix} \begin{pmatrix} a_{i0} \\ a_{i1} \\ \vdots \\ a_{iK} \end{pmatrix} \mod b$$

Now put $x_{i1} \in [0,1]$ take $x_{i1} = \sum_{k=0}^{K} x_{i1k} b^{1-k}$.

Generally $\boldsymbol{x}_{ij} = C^{(j)} \boldsymbol{a}_i \mod b$ for $i = 0, \dots, b^m - 1$ and $j = 1, \dots, s$.

Good $C^{(j)}$ give small t.

See Dick & Pillichshammer (2010), Niederreiter (1991)

Finding matrices: L'Ecuyer, Nuyens, Kuo

Computational cost

About the same as a Tausworth random number generator.

Base $b = 2$ offers some advantages.

Extensible nets

Nets can be extended to larger sample sizes.

(t, s)-sequence in base b

Infinite sequence of (t, m, s)-nets.

$$\underbrace{\boldsymbol{x}_1, \dots, \boldsymbol{x}_{b^m}}\ \underbrace{\boldsymbol{x}_{b^m+1}, \dots, \boldsymbol{x}_{2b^m}}\ \cdots\ \underbrace{\boldsymbol{x}_{kb^m+1}, \dots, \boldsymbol{x}_{(k+1)b^m}}\ \cdots$$

$$\underbrace{\underbrace{(t,m,s)\mathrm{-net}}_{\text{1st}}\quad \underbrace{(t,m,s)\mathrm{-net}}_{\text{2nd}}\quad \cdots\quad \underbrace{(t,m,s)\mathrm{-net}}_{\text{b'th}}}_{(t,m+1,s)\mathrm{-net}}\ \cdots$$

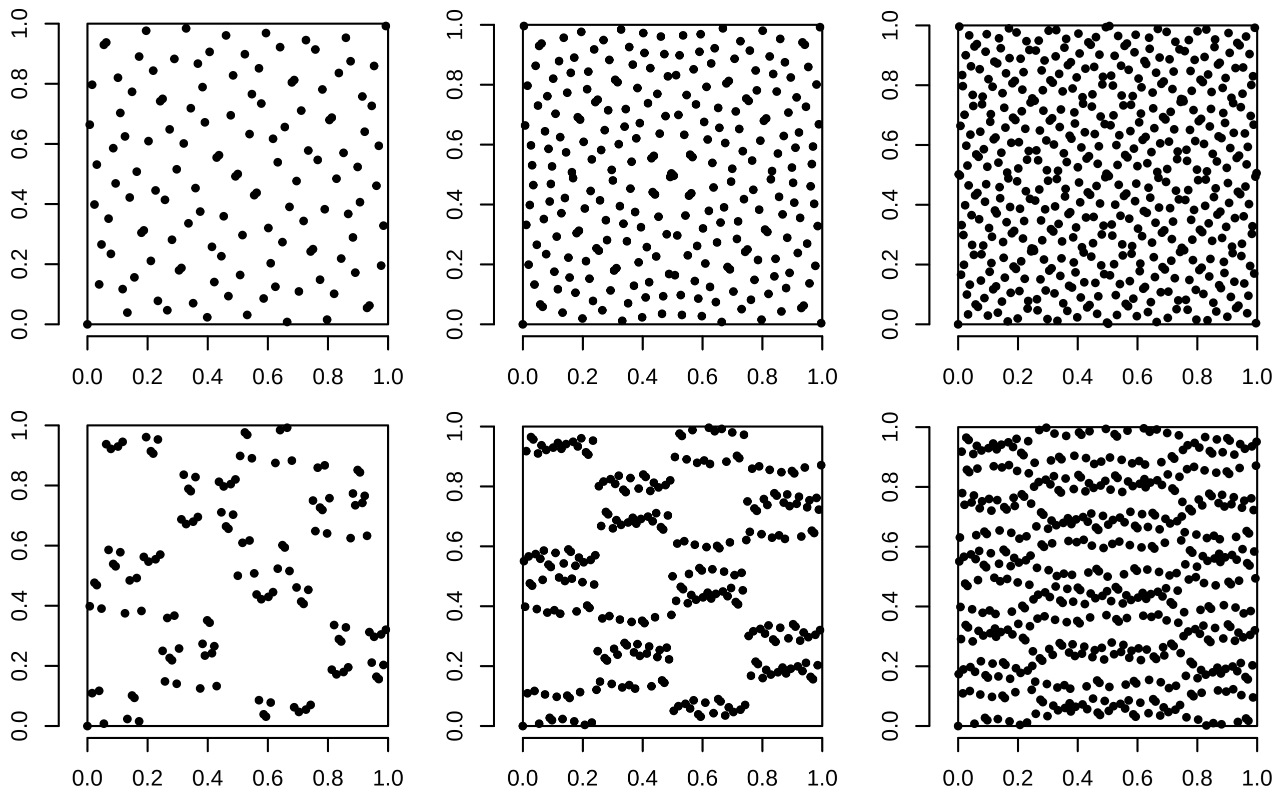
And recursively for all $m \geqslant t$.

Examples

Sobol' $b = 2$ Faure $t = 0$ Niederreiter & Xing $b = 2$ (mostly)

Sobol' points

Top row: $(x_{i,1}, x_{i,2})$ Bottom row: $(x_{i,10}, x_{i,11})$

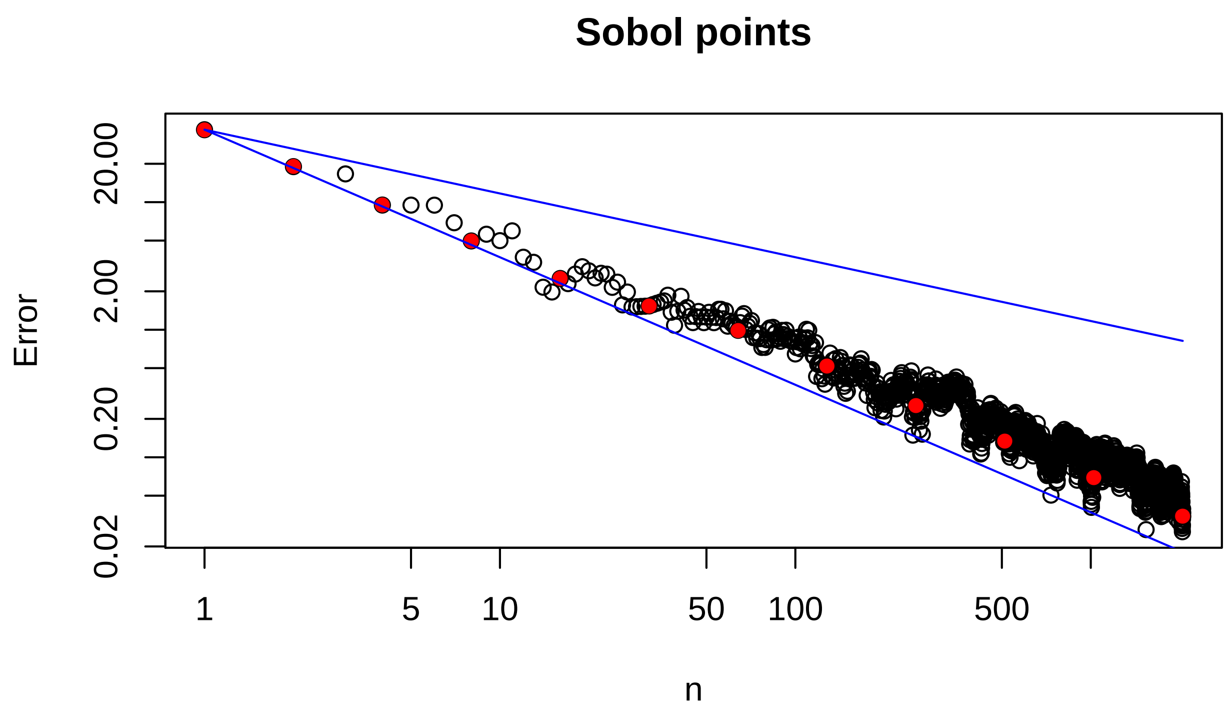


Using 'direction numbers' of Kuo and Joe

Very simple example

$$f(\boldsymbol{x}) = \left(\sum_{j=1}^{d} x_j\right)^2 \qquad \mathbb{E}(f(\boldsymbol{x})) = \frac{d^2}{4} + \frac{d}{12} \qquad d = 12$$

Reference lines $\propto n^{-1/2}$ and n^{-1}, ● for $n = 2^k$



This integrand depends only on one or two inputs at a time.

A finance problem

Paskov & Traub considered a financial valuation problem with $d = 360$

360 monthly interest rate fluctuations

$f(\boldsymbol{x})$ was value of a tranche of a portfolio of mortgages

No assurance that it would work

Recall Bakhvalov

Also $\log(n)^{360}/n$

is increasing in n until $n = e^{360} \approx 2.2 \times 10^{156}$

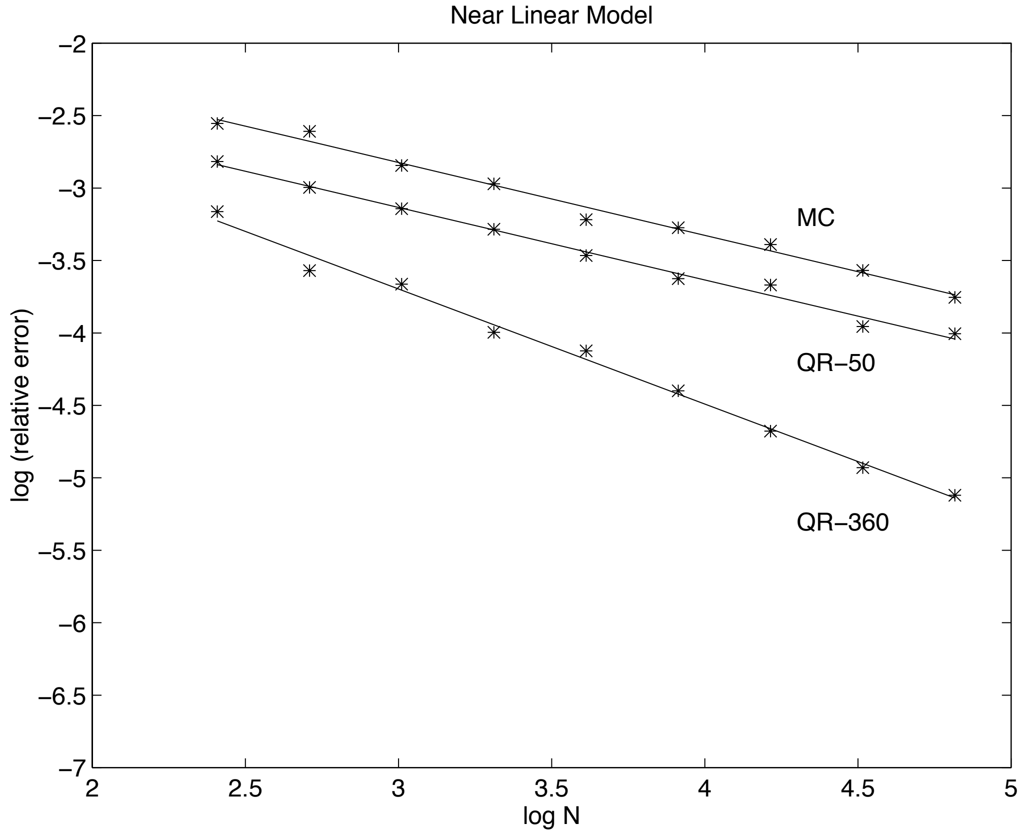
Surprise!

QMC worked very well

Why?

Finance integrand

Figure 4 from Caflisch, Morokoff & O (1997)



QMC worked ok for $d = 360$. Better results for other strategies in that paper

Hybrid: 50 QMC and 310 MC

For more hybrids in finance see Del Chicca & Larcher (2014)

Hybrids generally Spanier, Hofer, Niederreiter, Kritzer, Puchhammer ++

What if we succeed for large d?

Sometimes we get high accuracy for large d.

We didn't beat the curse of dimensionality.

We may have just had an easy, non-worst case function.

Bakhvalov never promised universal failure:

only the existence of hard cases.

Two kinds of easy

- Truncation: only the first $s \ll d$ components of $\boldsymbol{x}$ matter
- Superposition: the components only matter "s at a time"

Either way

f might not be "fully d-dimensional".

Studying the good cases

Two main tools to describe it

- Weighted spaces and tractability
- ANOVA and effective dimension

Decompositions

$$f(\boldsymbol{x}) = f_{\text{easy}}(\boldsymbol{x}) + f_{\text{hard}}(\boldsymbol{x})$$

$$\frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{x}_i) = \frac{1}{n}\sum_{i=1}^{n} f_{\text{easy}}(\boldsymbol{x}_i) + \frac{1}{n}\sum_{i=1}^{n} f_{\text{hard}}(\boldsymbol{x}_i)$$

If we're lucky then QMC works well on f_{easy} while f_{hard} is tiny.

Function decompositions

$$\begin{aligned} f(\boldsymbol{x}) = \mu &+ f_1(x_1) + f_2(x_2) + \cdots + f_d(x_d) \\ &+ f_{1,2}(x_1, x_2) + \cdots + f_{d-1,d}(x_{d-1}, x_d) \\ &+ \cdots + f_{1,2,\ldots,d}(x_1, x_2, \ldots, x_d) \end{aligned}$$

There are many decompositions

See Kuo, Sloan, Wasilkowski (2010)

We will look at the “Analysis of variance” (ANOVA)

Hoeffding (1948), Sobol’ (1969), Efron & Stein (1981)

More tersely

$$f(\boldsymbol{x}) = \mu + \sum_{r=1}^{d} \sum_{1 \leqslant j_1 < j_2 < \cdots < j_r \leqslant d} f_{j_1, j_2, \ldots, j_r}(x_{j_1}, \ldots, x_{j_r})$$

Decompositions continued

$$f(\boldsymbol{x}) = \mu + \sum_{r=1}^{d} \sum_{1\leqslant j_1<j_2<\cdots<j_r\leqslant d} f_{j_1,j_2,\dots,j_r}(x_{j_1},\dots,x_{j_r})$$

Write $\boldsymbol{x}_u = (x_j)_{j\in u}$ E.g. $x_{\{1,3,4\}} = (x_1, x_3, x_4)$

Even more tersely

$$f(\boldsymbol{x}) \equiv \sum_{u\subseteq 1:d} f_u(\boldsymbol{x})$$

$f_u(\boldsymbol{x}) = f(\boldsymbol{x}_u)$

I.e., $f_u(\boldsymbol{x})$ only depends on x_j for $j \in u$

$f_\varnothing$ is constant μ

Nearly additive

Caflisch, Morokoff & O (1997)

Numerical inspection $\Longrightarrow$

99.96% of the variance in the finance function from additive approximation

$$f_{\text{add}}(\boldsymbol{x}) = \mu + \sum_{j=1}^{360} f_j(x_j)$$

$$\sigma^2 \equiv \mathrm{Var}(f(\boldsymbol{x})), \qquad \boldsymbol{x} \sim \mathbf{U}[0,1]^{360}$$

$$\mathrm{Var}(f_{\text{add}}(\boldsymbol{x})) \doteq 0.9996\sigma^2$$

$$\mathrm{Var}(f(\boldsymbol{x}) - f_{\text{add}}(\boldsymbol{x})) \doteq 0.0004\sigma^2$$

Additive functions are easy to integrate

They're essentially one dimensional

K-H on a decomposition

$$f(\boldsymbol{x}) = \sum_{u \subseteq 1:d} f_u(\boldsymbol{x}_u)$$

Then

$$\hat{\mu} - \mu = \sum_{u \subseteq 1:d} \left(\frac{1}{n} \sum_{i=1}^{n} f_u(\boldsymbol{x}_{i,u}) - \int f_u(\boldsymbol{x}_u) \, \mathrm{d}\boldsymbol{x}_u \right)$$

$$|\hat{\mu} - \mu| \leqslant \sum_{u \subseteq 1:d} D_n^*(\boldsymbol{x}_{1,u}, \dots, \boldsymbol{x}_{n,u}) \times V_{\mathrm{HK}}(f_u)$$

Often $D_n^*(\boldsymbol{x}_{i,u}) \ll D_n^*(\boldsymbol{x}_i)$ for small $|u|$. E.g. $\log(n)^{|u|-1}/n$

If also $V_{\mathrm{HK}}(f_u)$ is small for large u,

then all the terms are small.

Effective dimension

A function might not be “fully d-dimensional”

$$|u| := \text{cardinality of } u$$

$$\lceil u \rceil := \max\{j \in 1{:}d \mid j \in u\}, \quad \lceil \varnothing \rceil = 0$$

Effective dimension s

Caflisch, Morokoff & O (1997)

$\sigma_u^2 = \mathrm{Var}(f_u(\boldsymbol{x}))$ under $\boldsymbol{x} \sim \mathbf{U}[0,1]^d$

$$\sum_{\lceil u \rceil \leqslant s} \sigma_u^2 \geqslant 0.99 \sum_{u \subseteq 1:d} \sigma_u^2 \qquad \text{truncation sense}$$

$$\sum_{|u| \leqslant s} \sigma_u^2 \geqslant 0.99 \sum_{u \subseteq 1:d} \sigma_u^2 \qquad \text{superposition sense}$$

Using ANOVA decomp

First QMC paper

Richtmyer (1951) has a truncation version

Effective dimension

- Explains in hindsight why QMC worked

- Motivates methods that reduce effective dimension

  Find $\tilde{f}$ with $\mathbb{E}(\tilde{f}(\boldsymbol{x})) = \mathbb{E}(f(\boldsymbol{x}))$ but $\tilde{f}$ has lower effective dimension

  Brownian bridge Moskowitz & Caflisch (1996)

  Principal components Ackworth, Broadie & Glasserman (1998)

  Adaptive Imai & Tan (2014)

- But when can we expect QMC to work in the future?

Weighted spaces

Function classes $\mathcal{F}$ where

$f \in \mathcal{F} \implies$ QMC likely to work

Weighted spaces

Hickernell (1996), Sloan & Wozniakowski (1998)

Then Dick, Kuo, Novak, Wasilkowski many more

Most recently: Kritzer, Pillichshammer & Wasilkowski (2021)

$$\partial^u \equiv \prod_{j\in u} \frac{\partial}{\partial x_j} \qquad \text{assume } \partial^{1:d} f \text{ exists}$$

Inner product, weights $\gamma_u > 0$

$$\mathcal{F}_{\boldsymbol{\gamma},C} = \{f \mid \|f\|_{\boldsymbol{\gamma}} \leqslant C\}$$

$$\|f\|_{\boldsymbol{\gamma}}^2 = \sum_{u\subseteq 1:d} \frac{1}{\gamma_u} \int_{[0,1]^u} \left| \int_{[0,1]^{-u}} \partial^u f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}_{-u} \right|^2 \mathrm{d}\boldsymbol{x}_u$$

How it works

Small $\gamma_u \implies$ only small $\|\partial^u f\|$ in ball

$\implies$ sampling $\boldsymbol{x}_u$ well cannot be “important”

So make γ_u large for small $|u|$ and $\lceil u \rceil$

Product weights

$\gamma_u = \prod_{j \in u} \gamma_j$ where $1 \geqslant \gamma_j$ decrease rapidly with j.

Larger $|u| \implies$ smaller γ_u

Increasing $\lceil u \rceil$ reduces γ_u

$$\gamma_{\{1,3,6\}} \leqslant \gamma_{\{1,3,5\}}$$

Now $f \in \mathcal{F}_{\boldsymbol{\gamma},C}$ implies $\partial^u f$ small when $|u|$ large.

Common choices

$$\gamma_j = j^{-a} \qquad a \in \{1,2\}$$

Non product weights

Many more weight choices: Dick, Kuo, Sloan (2013)

Tractability

Sloan & Wozniakowski (1996) and many more

For $f \in \mathcal{F}_{\boldsymbol{\gamma},C}$ and $n = 0$ data $\cdots$ guess $\hat{\mu}_0 = 0$ Initial error

$$\mathrm{Err}(d, n = 0) = \sup_{f \in \mathcal{F}_{\boldsymbol{\gamma},C}} |\mu(f)|$$

For method $\hat{\mu}$

$$\mathrm{Err}(d, n) \equiv \sup_{f \in \mathcal{F}} |\hat{\mu} - \mu|$$

$n_*(d, \epsilon) \equiv$ First n with $\mathrm{Err}(d, n) \leqslant \epsilon \times \mathrm{Err}(d, 0)$

Weak tractability

$$n_*(d, \epsilon) = \mathrm{poly}(d, 1/\epsilon)$$

Strong tractability

$n_*(d, \epsilon) = \mathrm{poly}(1/\epsilon)$ $\quad \sum_{j=1}^{\infty} \gamma_j < \infty$ suffices

See Dick, Kuo, Sloan (2013)

Open or partly open problems

- **Which** weighted space to use?

  I.e., how to choose $\boldsymbol{\gamma}$?

- By **which factor** ϵ should we reduce the initial error?

  Does it depend on d?

- ANOVA captures L^2 magnitude of f_u for small $|u|$ but not their smoothness.

  Griebel, Kuo & Sloan (2010) give conditions where low order ANOVA components are smooth.

Context

We design an algorithm for all $f \in \mathcal{F}$

The user may have only one f

or a few or a finite space of them

Choosing γ

Each γ corresponds to a reproducing kernel Hilbert space (RKHS) $\mathcal{H} = \mathcal{F}_{\boldsymbol{\gamma}}$

The question

Which RKHS should we use in a given problem?

$\mathcal{H}_1$ or $\mathcal{H}_2$ or $\cdots$ or $\mathcal{H}_J$ $\cdots$

1) sometimes $f \in \mathcal{H}_j$ all $j = 1, \dots, J$
   and $f \in \mathcal{H}_1$ vs $\mathcal{H}_2$ have very different implications

2) sometimes f belongs to none of them.
   while $|f - \widetilde{f}| \leqslant \epsilon$ where $\widetilde{f} \in \mathcal{H}$

Starts

Wang & Sloan have suggestions for finance

Kuo, Schwab +co-authors have suggestions when solving PDEs

Novak told me about 'fat F problem' in Wozniakowski (1986) IBC

Uncertainty

Important for $\hat{\mu}$ to be accurate

$\Longrightarrow$ important to **measure** accuracy

Koksma-Hlawka

Tells us $|\hat{\mu} - \mu| = o(n^{-1+\epsilon})$ but misses the lead constant

$$|\hat{\mu} - \mu| \leqslant D_n^* \times V_{\mathrm{HK}}(f) = ? \times ??? = ????$$
$$|\hat{\mu} - \mu| \leqslant \epsilon \times \mathrm{Err}(d, 0) = \epsilon \times ???$$

What we get

$$V_{\mathrm{HK}}(f) < \infty \implies |\hat{\mu} - \mu| < \infty$$
$$V_{\mathrm{HK}}(f) = \infty \implies |\hat{\mu} - \mu| \leqslant \infty$$

Next lecture

Randomized QMC for error estimation

Another approach

GAIL by Hickernell and others

Thanks

- Johannes Kepler Universität Linz
- RICAM: Johann·Radon·Institute for Computational and Applied Mathematics
- MCQMC series & Harald Niederreiter
- U.S. NSF: grants up to and including IIS-1837931
- Invitation: Peter Kritzer, Gerhard Larcher, Lucia Del Chicca
- Introductions: Peter Kritzer, Gerhard larcher, Gunther Leobacher
- Organization: Melanie Traxler

Post presentation

The next three slides were not shown. The first has an open problem about whether one can construct a persistent fooling function f with $\|f\|_{\mathrm{HK}} = 1$ for which $|\hat{\mu} - \mu| > c\log(n)^{d-1}/n$ for all n and some $c > 0$. Taking $\boldsymbol{x}_i$ to be a given QMC sequence like that of Halton or Sobol' would be interesting. Finding that any sequence must have such a fooling function would be more interesting. I have had email discussions about this but do not yet know the answer. For $d = 1$ and $f(x) = x$ the van der Corput sequence has error $1/(2n) = O(\log(n)^{d-1}/n)$ at $n = 2^m$. Bad choices of n can make it like $\log(n)/n$.

After that come slides on lattices and higher order nets.

Open problem

What is the worst case for

$$\limsup_{n\to\infty} |\hat\mu - \mu| \times \frac{n}{\log(n)^{?}}$$

for $V_{\mathrm{HK}}(f) = 1$

Base case

E.g., for $\boldsymbol{x}_i$ a Sobol' sequence

More general

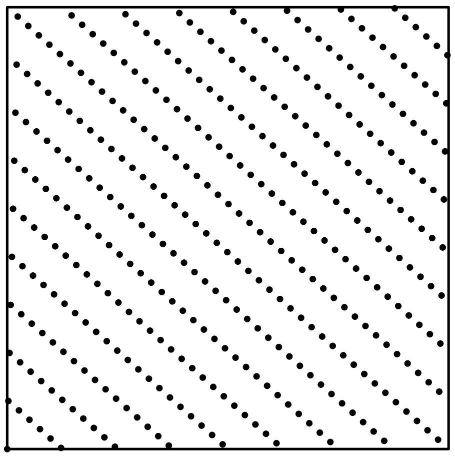
For minimax $\boldsymbol{x}_i$

Discussions with Novak, Hickernell, Nuyens

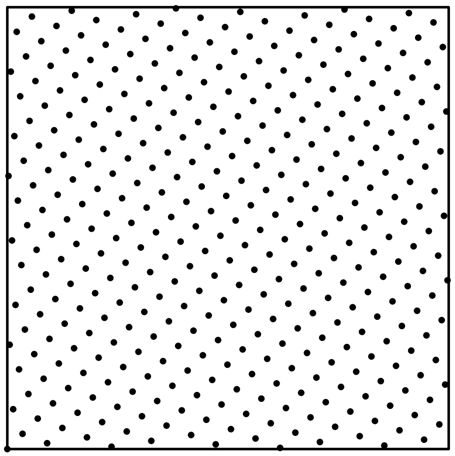
Lattices

The other main family of QMC points. An extensive literature, e.g., Sloan & Joe, Kuo, Nuyens, Dick, Cools, Hickernell, Lemieux, L'Ecuyer· · ·

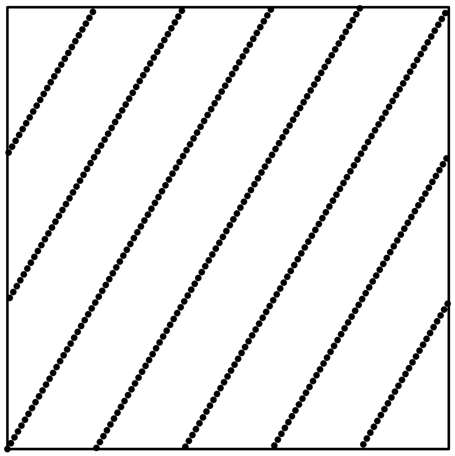
Some lattice rules for n=377



z = (1,41)



z = (1,233)



z = (1,253)

Computation like congruential generators

$$\boldsymbol{x}_i = \left(\frac{i}{n}, \frac{Z_2 i}{n}, \frac{Z_3 i}{n}, \dots, \frac{Z_d i}{n}\right) \pmod 1 \quad Z_j \in \mathbb{N}$$

choose $\boldsymbol{Z} = (1, Z_2, Z_3, \dots, Z_d)$ wisely

Higher order nets

Results from Dick, Baldeaux

Start with a net $\boldsymbol{z}_i \in [0,1)^{2s}$ dimensions.

'Interleave' digits of two variables to make a new one:

$$\begin{aligned} z_{i,2j} &= 0.g_1 g_2 g_3 \cdots \\ z_{i,2j+1} &= 0.h_1 h_2 h_3 \cdots \end{aligned} \quad \longrightarrow \quad x_{i,j} = 0.g_1 h_1 g_2 h_2 g_3 h_3 \cdots .$$

Error is $O(n^{-2+\epsilon})$ under increased smoothness: $\dfrac{\partial^{2s}}{\partial x_1^2 \dots \partial x_d^2} f$

Scrambling gets RMSE $O(n^{-2-1/2+\epsilon})$

Even higher

Start with ks dimensions interleave down to s.

Get $O(n^{-k+\epsilon})$ and $O(n^{-k-1/2+\epsilon})$ (under still higher smoothness)

Very promising

Cost: many inputs and much smoothness.

Starting to be used in PDEs. Kuo, Nuyens, Scwhab