Quasi-Monte Carlo

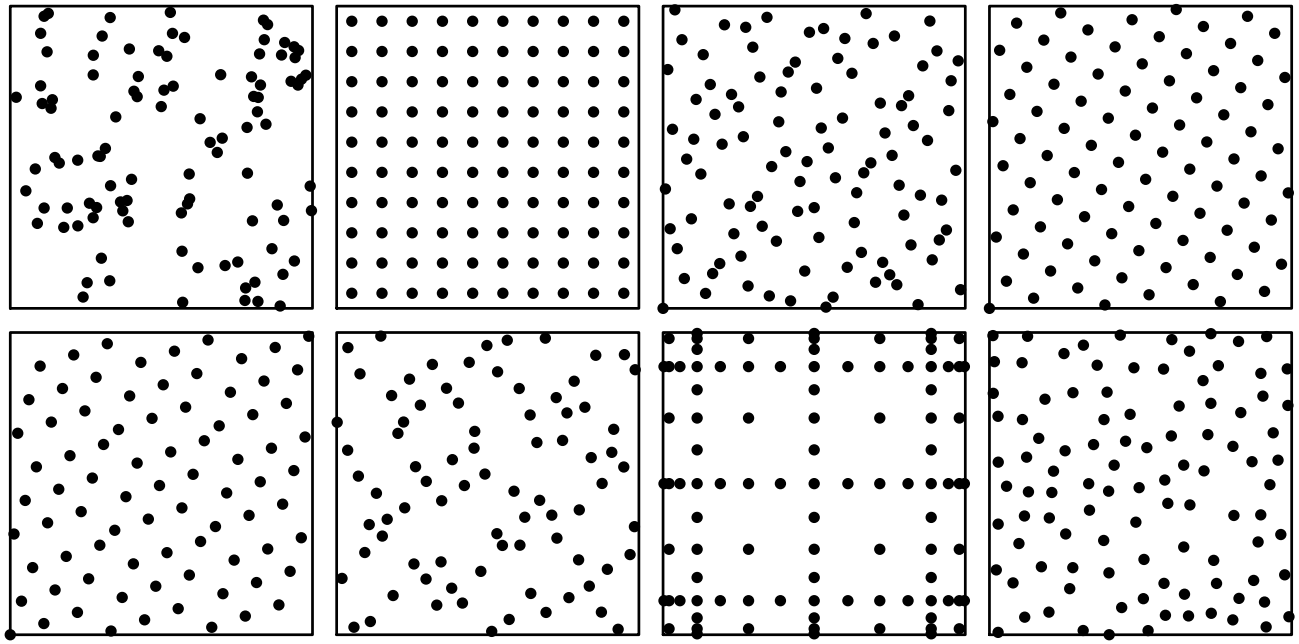
A tutorial introduction

Art B. Owen

Stanford University

MC and QMC and other points

Monte Carlo (MC) and quasi-Monte Carlo (QMC) methods come down to sampling the input space of a function.



Top row, left to right

MC, grid, and two QMC methods

Top row, left to right

Two more QMC, sparse grid, blue noise

Outline

1) What QMC is (utmost stratification)

2) Why it works (discrepancy, variation and Koksma-Hlawka)

3) How it works (constructions)

4) Randomized QMC

5) When it works best (effective dimension, tractability, weighted spaces)

   (room for Bayesian thinking here)

Landmark papers in MC

Some landmark papers where Monte Carlo was applied:

- Physics Metropolis et al. (1953)
- Chemistry (reaction equations) Gillespie (1977)
- Financial valuation Boyle (1977)
- Bootstrap resampling Efron (1979)
- OR (discrete event simulation) Tocher & Owen (1960)
- Bayes (maybe 5 landmarks in early days)
- 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 Alex Keller++

  (They got an Oscar!)
- Solving PDEs Frances Kuo, Christoph Schwab++, 2015
- Particle methods Chopin & Gerber (2015)

The next landmark methods

Some strong candidate areas:

- machine learning
- Bayes
- uncertainty quantification (UQ)

The next landmark

QMC methods dominate when

dimension is high, but

effective dimension is low

Best way to find out

try it and see

Low effective dimension

$f : [0, 1]^d \to \mathbb{R}$ is very nearly a sum of functions of just a few inputs

MC and QMC

We estimate

$$\mu = \int_{[0,1]^d} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \qquad \text{by} \qquad \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{x}_i), \quad \boldsymbol{x}_i \in [0,1]^d$$

In plain MC, the $\boldsymbol{x}_i$ are IID $\mathbf{U}[0,1]^d$. In QMC they're 'spread evenly'.

Non uniform

$$\mu = \int_{\Omega} f(\boldsymbol{x})p(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \qquad \text{and} \qquad \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} f(\psi(\boldsymbol{u}_i)), \quad \boldsymbol{u}_i \in [0,1]^s$$

$$\psi : [0,1]^s \to \mathbb{R}^d$$

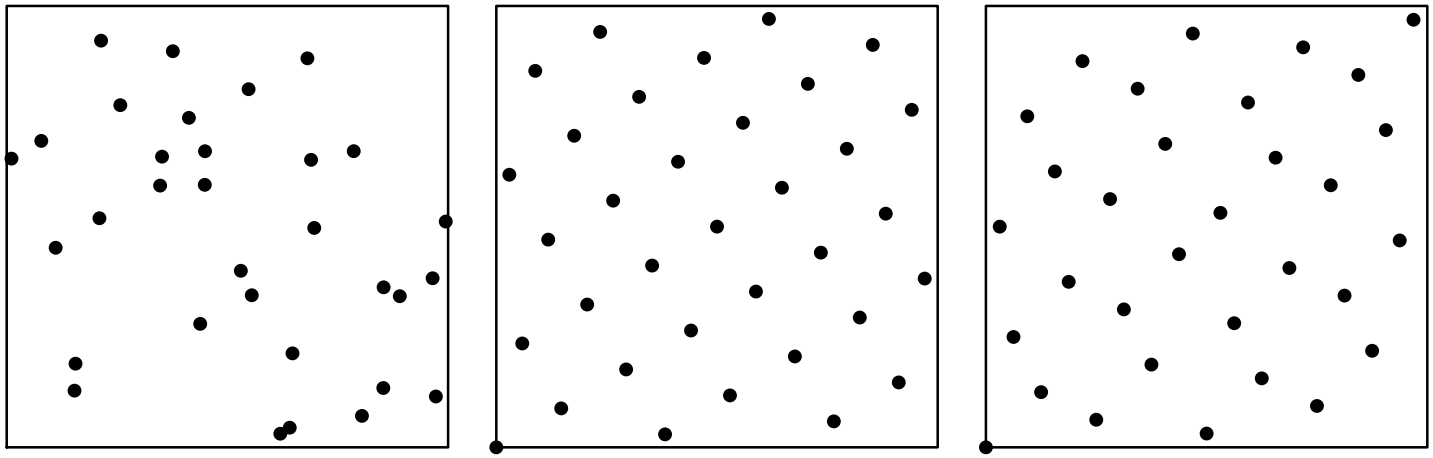
If $\boldsymbol{u} \sim \mathbf{U}[0,1]^s$ then $\boldsymbol{x} = \psi(\boldsymbol{u}) \sim p$

Many methods fit this framework. Devroye (1986)

Acceptance-rejection is a bit awkward.

Illustration

MC and two QMC methods in the unit square



Monte Carlo Fibonacci lattice Hammersley sequence

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.

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

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

Discrepancy

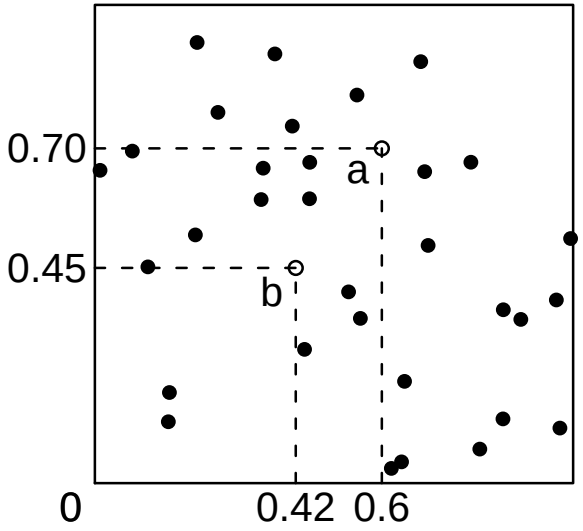
A discrepancy is a distance $\|F - \hat{F}_n\|$ between measures $F = \mathbf{U}[0,1]^d$ and $\hat{F}_n = \mathbf{U}\left\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\right\}$.

There are many discrepancies.

Local discrepancy

Did the box $[0, \boldsymbol{a})$ get it's fair share of points?

Local discrepancy at a, b



$$\delta(\boldsymbol{a}) = \hat{F}_n([0, \boldsymbol{a})) - F([0, \boldsymbol{a})) = \frac{13}{32} - 0.6 \times 0.7 = -0.01375$$

Star discrepancy

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

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

More 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

Also

Wrap-around discrepancies Hickernell

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

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

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

That's enough for good integration.

QMC's law of large numbers

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

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

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 CLT.

QMC has the Koksma-Hlawka inequality.

Koksma's inequality

For $d = 1$ $\quad |\hat{\mu} - \mu| \leqslant D_n^*(x_1, \ldots, x_n) \times \int_0^1 |f'(x)| \, \mathrm{d}x$

NB: $D_n^* = \|\delta\|_\infty$ and $\int_0^1 |f'(x)| \, \mathrm{d}x$ is the total variation $V(f)$.

Setup for the proof

$$\int_0^1 f(x) \, \mathrm{d}x = f(1) - \int_0^1 x f'(x) \, \mathrm{d}x$$ Integration by parts

$$\frac{1}{n} \sum_{i=1}^{n} f(x_i) = f(1) - \frac{1}{n} \sum_{i=0}^{n} i \underbrace{(f(x_{i+1}) - f(x_i))}_{= \int_{x_i}^{x_{i+1}} f'(x) \, \mathrm{d}x}$$ Summation by parts

A few more steps, via continuity of f'

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

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)

Multidimensional variation has a few surprises for us.

Puzzler

Is this a 100% confidence interval?

Rates of convergence

It is possible to get $D_n^* = O\Big(\dfrac{\log(n)^{d-1}}{n}\Big)$.

Then

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

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)

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

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

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

KH holds as an equality for a worst case function, e.g., $f' \doteq \pm\delta$.

It even holds if an adversary sees your $\boldsymbol{x}_i$ before picking f.

Koksma-Hlawka is also very loose

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

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

Just like Chebychev's inequality

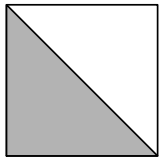
It is also tight and very loose. E.g., $\Pr\bigl(|\mathcal{N}(0,1)| \geqslant 10\bigr) \leqslant 0.01$ is loose.

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

Variation

Multidimensional 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

For general $\boldsymbol{a} \in \mathbb{R}^d$,

$$f = \max(\boldsymbol{a}^{\mathsf{T}}\boldsymbol{x}, 0) \implies V_{\mathrm{HK}}(f) = \infty \text{ for } d \geqslant 3$$

$$f = \max(\boldsymbol{a}^{\mathsf{T}}\boldsymbol{x}, 0)^2 \implies V_{\mathrm{HK}}(f) = \infty \text{ for } d \geqslant 4$$

QMC-friendly discontinuities

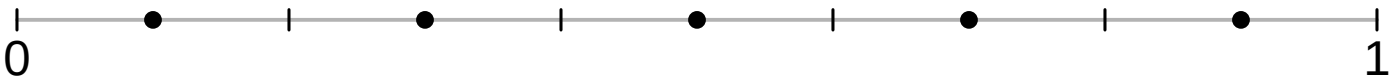
Axis parallel discontinuities may have $V_{\mathrm{HK}} < \infty$.

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

Extensibility

For $d = 1$, the equispaced 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 $D_n^* = O((\log n)^d/n)$.

No known extensible constructions get $O((\log n)^{d-1}/n)$.

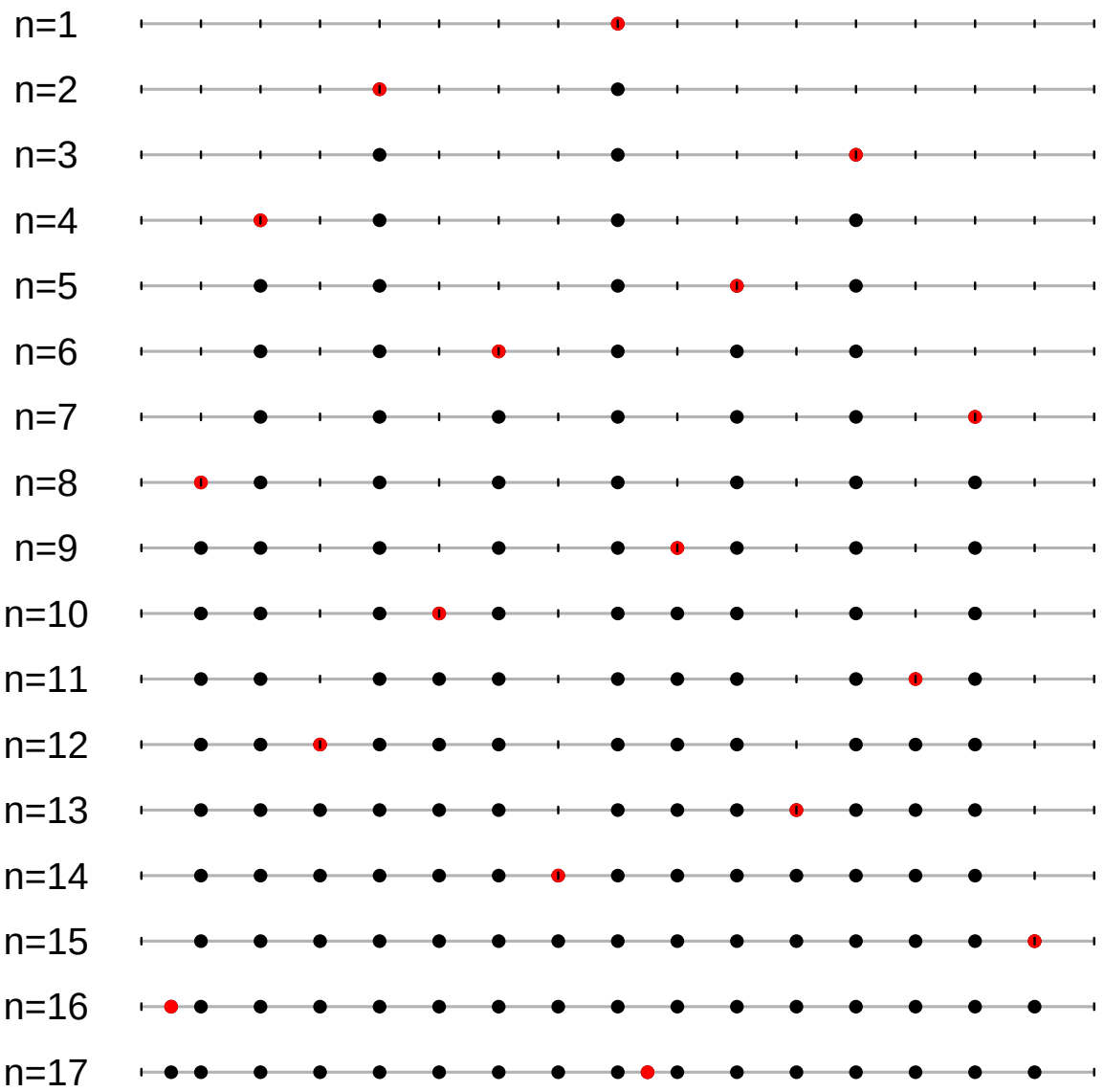
van der Corput

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 $x_1 = 0$.

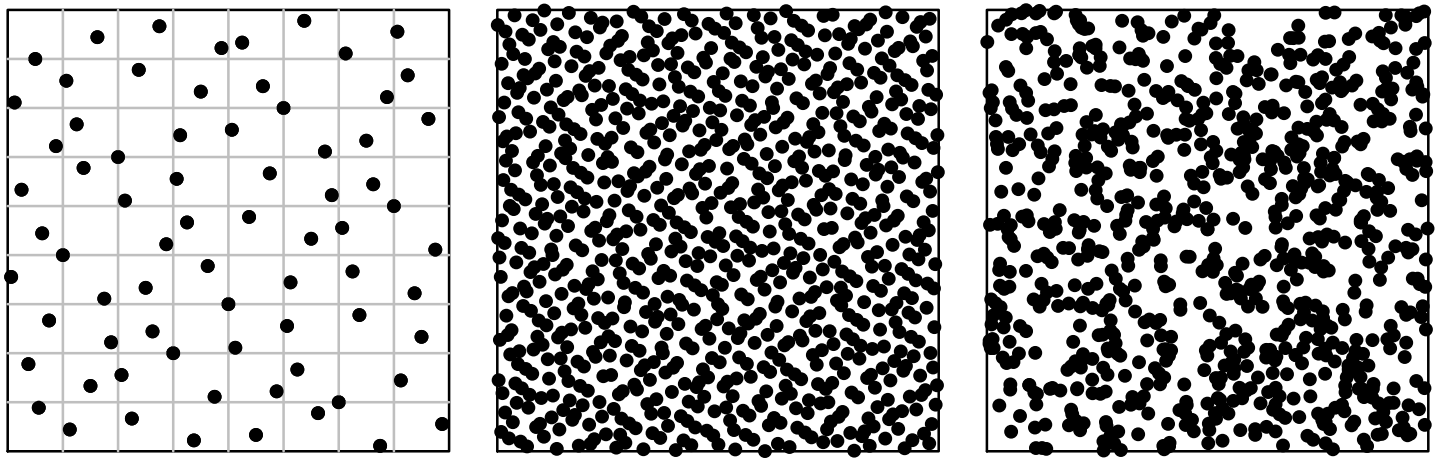
van der Corput



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

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

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

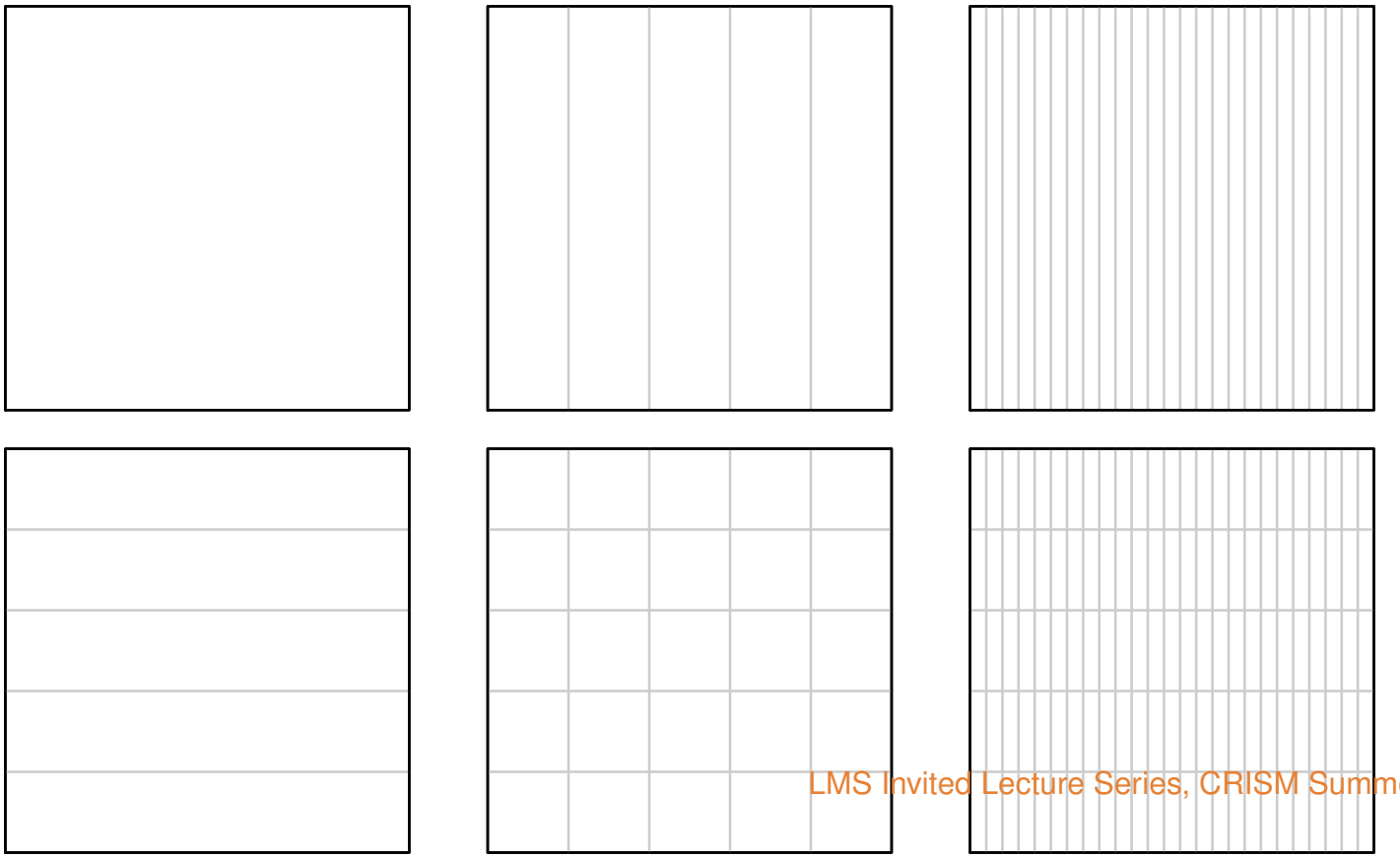
Digital nets

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

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

Elementary intervals

Some elementary intervals in base 5



Digital nets

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

$(0, m, s)$-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) points base 2

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

minT project

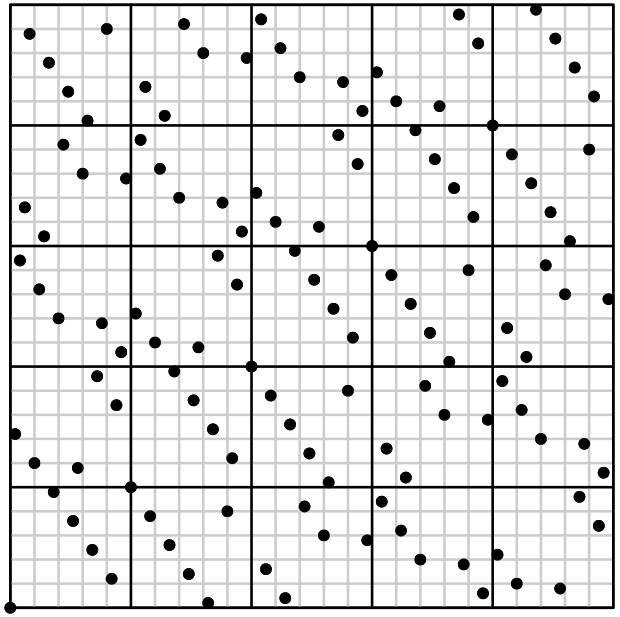
Schürer & Schmid give bounds on t given b, m and s

Monographs

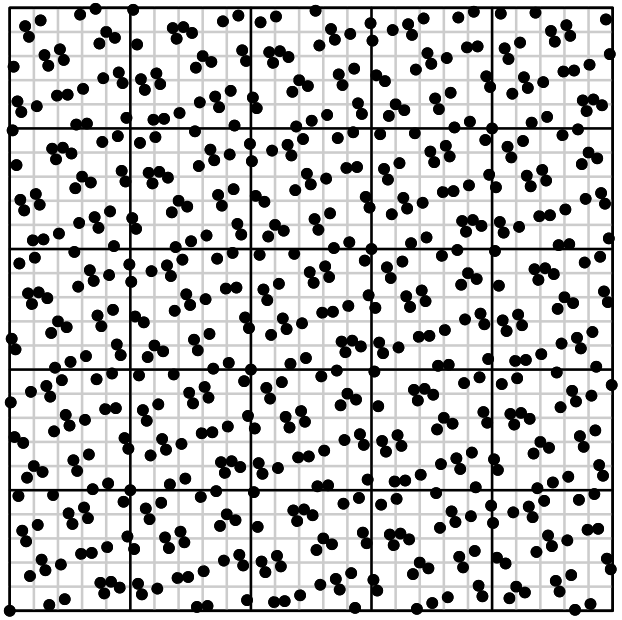
Niederreiter (1992) Dick & Pillichshammer (2010)

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.

Think of $43{,}750$ control variates for 625 obs.

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

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

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 $i = 0, \ldots, b^m - 1$ and $j = 1, \ldots, s$.

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

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

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.

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

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

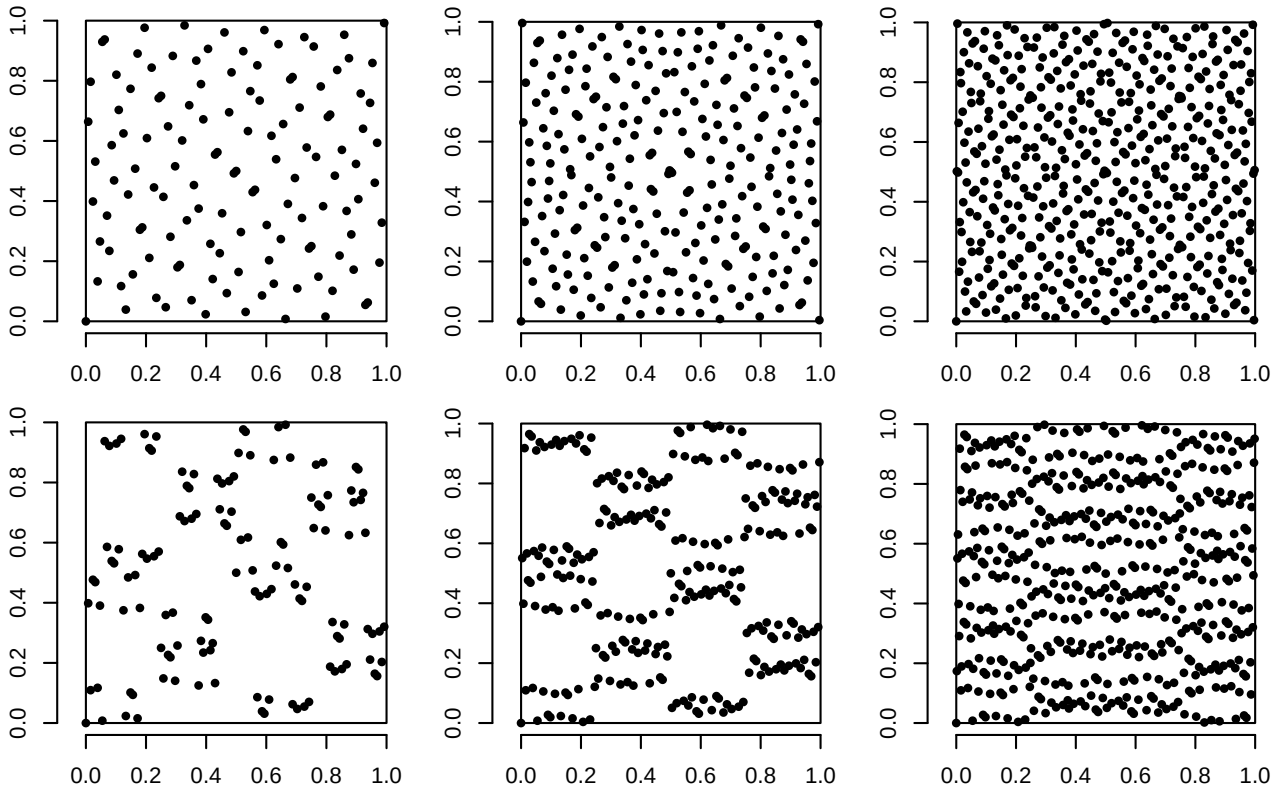
And recursively for all $m \geqslant t$.

Examples

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

Sobol' points

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

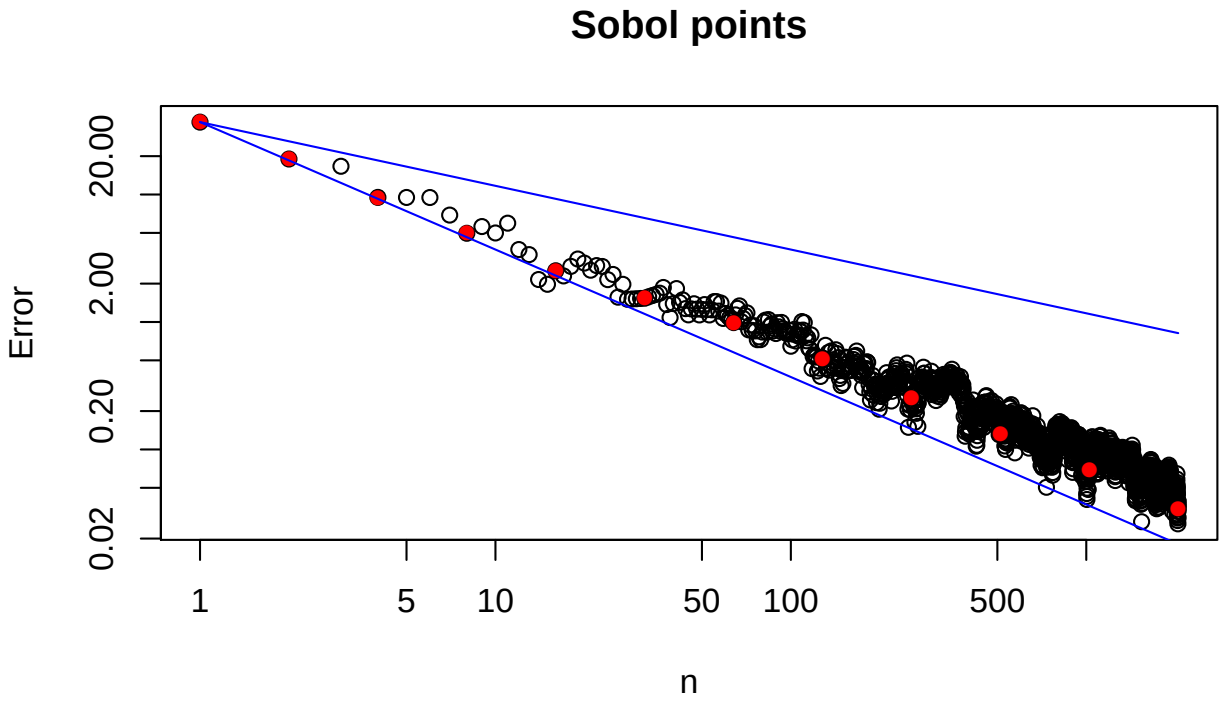


Using 'direction numbers' of Kuo and Joe

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 $n = 2^k$

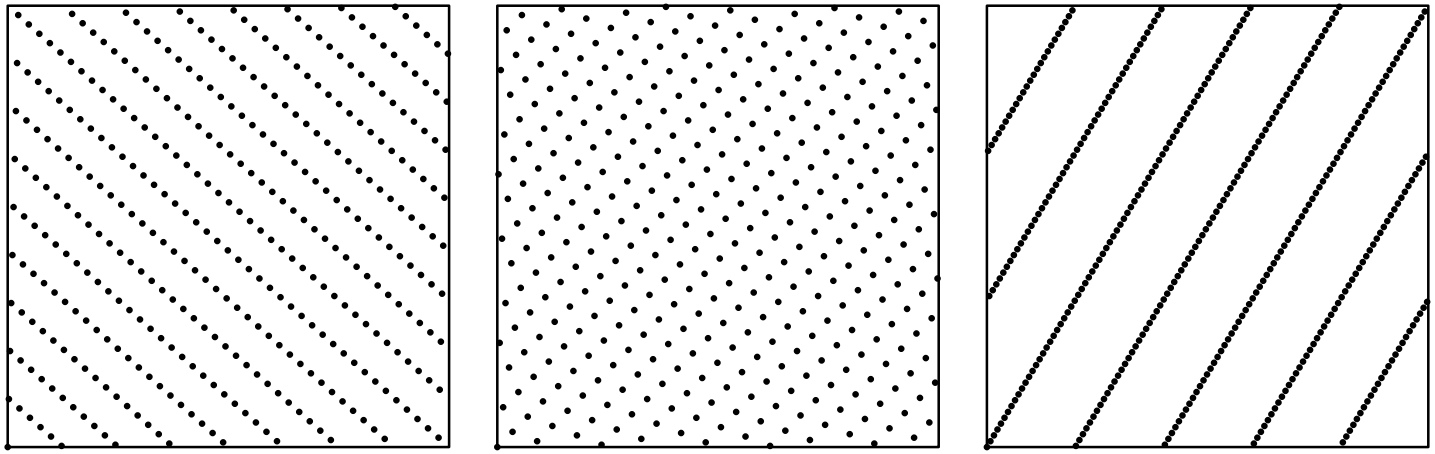


This integrand depends only on one or two inputs at a time.

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

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

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

QMC error estimation

$$|\hat{\mu} - \mu| \leqslant D_n^* \times V_{\mathrm{HK}}(f)$$

Not a 100% confidence interval

because not known to user

- D_n^* is hard to compute
- V_{HK} harder to get than μ

  For fixed n we get $|\hat{\mu} - \mu| < \infty$
- $V_{\mathrm{HK}} = \infty$ is common, e.g., $f(x_1, x_2) = 1_{x_1+x_2 \leqslant 1}$ 

  Then KH gives $|\hat{\mu} - \mu| \leqslant \infty$

Also

Koksma-Hlawka is worst case. It can be very conservative.

Recent work

GAIL project of Hickernell++ allows user specified error tolerance ϵ.

Randomized QMC

1) Make $\boldsymbol{x}_i \sim \mathbf{U}[0,1)^d$ individually,

2) keeping $D_n^*(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) = O(n^{-1+\epsilon})$ collectively.

R independent replicates

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

$$\widehat{\mathrm{Var}}(\hat{\mu}) = \frac{1}{R(R-1)}\sum_{r=1}^{R}(\hat{\mu}_r - \hat{\mu})^2$$

If $V_{\mathrm{HK}}(f) < \infty$ then

$$\mathbb{E}((\hat{\mu} - \mu)^2) = O(n^{-2+\epsilon})$$

Random shift Cranley & Patterson (1976)

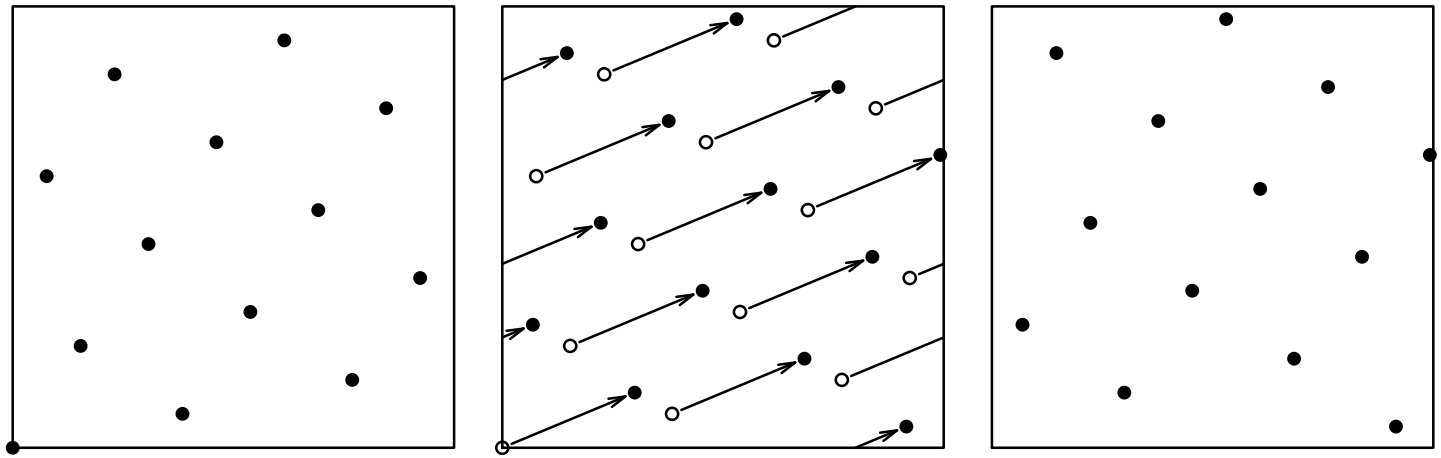
Scrambled nets O (1995,1997,1998)

Linear scramble Matousek (1998)

Survey in L'Ecuyer & Lemieux (2005)

Rotation modulo 1

Cranley–Patterson rotation



Shift the points by $\boldsymbol{u} \sim \mathbf{U}[0,1)^s$ with wraparound:

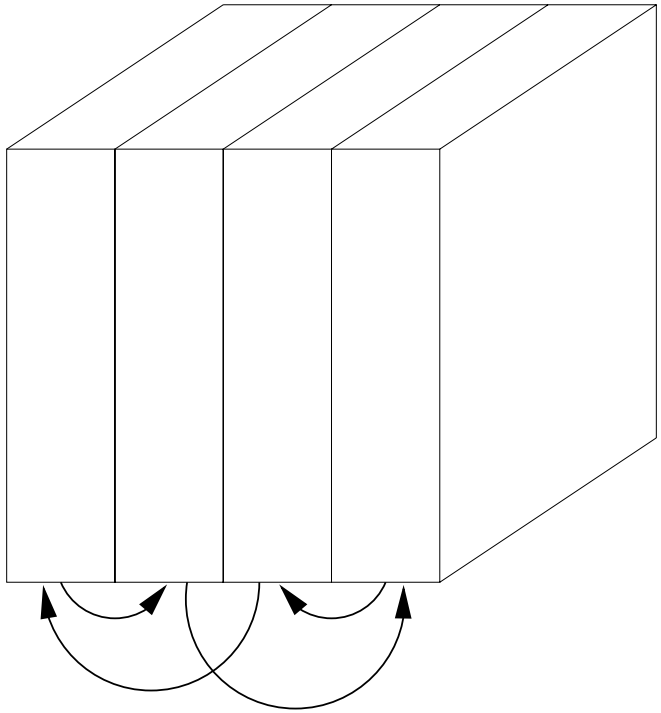
$$\boldsymbol{x}_i \to \boldsymbol{x}_i + \boldsymbol{u} \pmod 1.$$

Commonly used on lattice rules.

Can also be used with nets.

At least it removes $\boldsymbol{x}_1 = 0$.

Digit scrambling



1) Chop space into b slabs. Shuffle.
2) Repeat within each of b slabs.
3) Then within b^2 sub-slabs.
4) Ad infinitum b^3, b^4, . . .
5) And the same for all s coordinates.

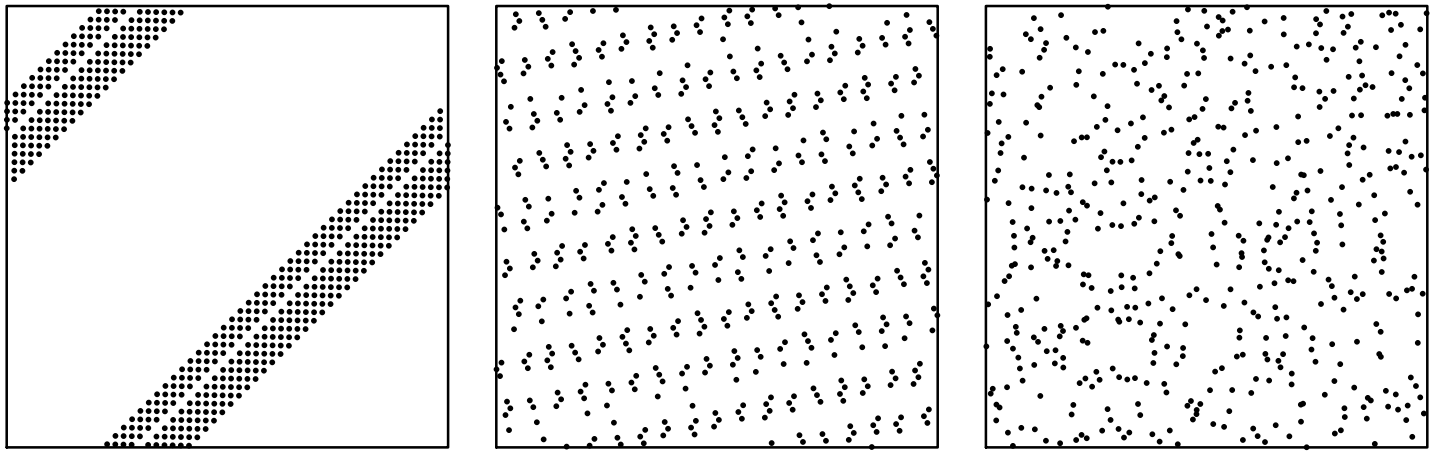
Each $\boldsymbol{x}_i \sim \mathbf{U}[0,1)^s$ and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ still a net (a.s.). O (1995)

Cheaper scrambles: digital shift and random linear.

Example scrambles

Two components of the first 530 points of a Faure $(0, 53)$-net in base 53.

Randomized Faure points



Digital shift | Random linear | Nested uniform

The digital shift is much like a Cranley-Patterson rotation.

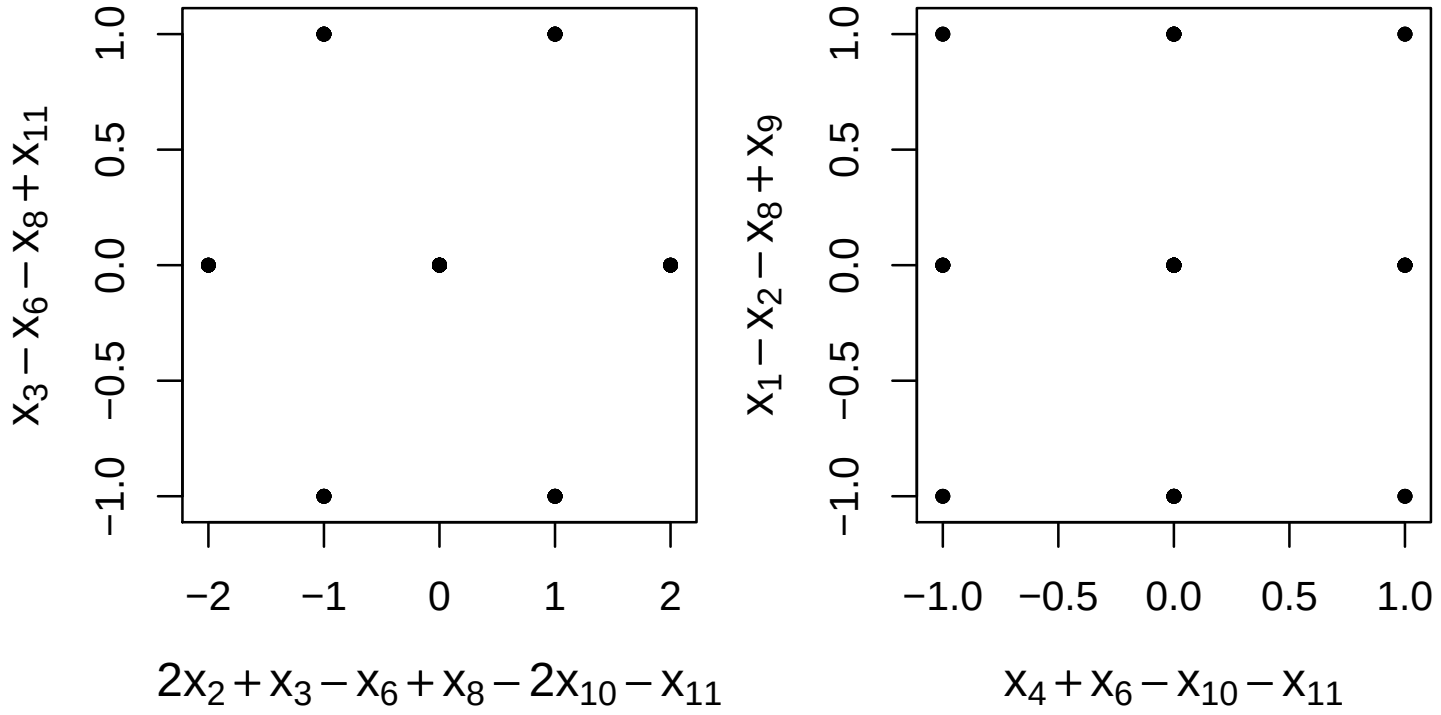
It uses just one random $\boldsymbol{u}$ for all points: $\widetilde{\boldsymbol{x}}_i = \boldsymbol{x}_i \oplus \boldsymbol{u}$.

Random linear Matousek (1998) and nested uniform O (1995) have the same $\mathrm{Var}(\hat{\mu})$.

Unscrambled Faure

First $n = 11^2 = 121$ points of Faure $(0, 11)$-net in $[0, 1]^{11}$.

Two projections of 121 Faure points



Unscrambled points are very structured.

Scrambling breaks it up.

Scrambled net properties

Using $\sigma^2 = \int (f(\boldsymbol{x}) - \mu)^2 \, \mathrm{d}\boldsymbol{x}$

If	Then	N.B.
$f \in L^2$	$\mathrm{Var}(\hat{\mu}) = o(1/n)$	even if $V_{\mathrm{HK}}(f) = \infty$
$f \in L^2$	$\mathrm{Var}(\hat{\mu}) \leqslant \Gamma_{t,b,s} \, \sigma^2/n$	if $t = 0$, $\Gamma \leqslant e \doteq 2.718$
$\partial^{1,2,\ldots,s} f \in L^2$(etc)	$\mathrm{Var}(\hat{\mu}) = O(\log(n)^{s-1}/n^3)$	O (1997,2008)

$\Gamma < \infty$ rules out $(\log n)^{s-1}$ catastrophe at finite n.

Loh (2003) has a CLT for $t = 0$ (and **fully** scrambled points).

Geometrically

Scrambling Faure breaks up the diagonal striping of the nets.

Scrambling Sobol' points moves the full / empty blocks around.

Improved rate

RMSE is $O(n^{-1/2})$ better than QMC rate (cancellation).

Holds for nested uniform and nested linear scrambles.

Scrambling vs shifting

Consider $n = 2^m$ points in $[0, 1)$.

QMC

van der Corput points $(i-1)/n$ for $i = 1, \ldots, n$.

Shift

Shift all points by $U \sim \mathbf{U}(0, 1)$ with wraparound.

Get one point in each $[(i-1)/n, i/n)$

Scramble

Get a stratified sample, **independent** $x_i \sim \mathbf{U}[(i-1)/n, i/n)$

Random errors cancel yielding an $O(n^{-1/2})$ improvement.

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

The curse of dimension

Curse of dimension: larger d makes integration harder.

$$C_M^r = \left\{ f : [0,1]^d \to \mathbb{R} \;\middle|\; \left| \prod_j \frac{\partial^{\alpha_j}}{\partial x_j^{\alpha_j}} f \right| \leqslant M,\ \sum_j \alpha_j = r,\ \alpha_j \geqslant 0 \right\}$$

Bahkvalov I:

For any $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in [0,1]^d$ there is $f \in C_M^r$ with $|\hat{\mu}_n - \mu| \geqslant k n^{-r/d}$

Ordinary QMC like $r = d$

Bahkvalov II:

Random points can't beat RMSE $O(n^{-r/d-1/2})$

Ordinary MC like $r = 0$

What if we beat those rates?

Sometimes we get high accuracy for large d.

It does not mean we beat the curse of dimensionality.

Bahkvalov never promised universal failure.

Only the existence of hard cases.

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

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".

Dimensional decomposition

For $u = \{j_1, j_2, \ldots, j_r\} \subset 1{:}d \equiv \{1, 2, \ldots, d\}$ let

$$\boldsymbol{x}_u = (x_{j_1}, \ldots, x_{j_r})$$
$$\boldsymbol{x}_{i,u} = (x_{ij_1}, \ldots, x_{ij_r})$$

Via ANOVA or other method, write

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

Then

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

Often $D_n^*(\boldsymbol{x}_{i,u}) \ll D_n^*(\boldsymbol{x}_i)$ for small $|u|$. If also $\|f_u\|$ is small for large u, then all the terms are small.

Studying the good cases

Two main tools to describe it

- Weighted spaces and tractability
- ANOVA and effective dimension

Implications

Neither causes the curse to be lifted. They describe the happy circumstance where the curse did not apply.

Both leave important gaps described below. I'll raise as an open problem later.

Weighted spaces

Hickernell (1996), Sloan & Wozniakowski (1998),

Dick, Kuo, Novak, Wasilkowski, many more

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

Inner product, weights $\gamma_u > 0$

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

Function ball $B_{\gamma,C} = \{f \mid \|f\|_\gamma \leqslant C\}$

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

Product weights

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

Now $f \in B_{\gamma,C}$ implies $\partial^u f$ small when $|u|$ large.

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

ANOVA and effective dimension

Caflisch, Morokoff & O (1997)

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

Often f is dominated by its low order interactions.

Then RQMC may make a huge improvement.

Let $\sigma_u^2 = \mathrm{Var}(f_u)$ variance component

Truncation dim. $s \leqslant d$

$$\sum_{u \subseteq 1:s} \sigma_u^2 \geqslant 0.99 \sum_{u \subseteq 1:d} \sigma_u^2$$

Superposition dim. $s \leqslant d$

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

Mean dimension

$$\frac{\sum_u |u| \sigma_u^2}{\sum_u \sigma_u^2}$$

Liu & O (2006) Easier to estimate via Sobol' indices.

Open problem

- ANOVA captures magnitude of the low dimensional parts but not their smoothness. Even when it verifies that f is dominated by low dimensional parts it does not assure small $|\hat{\mu} - \mu|$.
- Weighted space models assure accurate estimation at least asymptotically. However, it is not easy to decide **which** weighted space to use.
- Given a weighted space, there are algorithms to tune QMC points for it.
- ANOVA approaches may support a strategy for choosing good weights for a given problem, building on Sobol' indices Sobol'++, Saltelli++, Prieur++, Kucherenko++ or active subspaces Constantine++

The problems are about how to combine the approaches and when / whether a resulting adaptive algorithm will be effective.

Example

Kuo, Schwab, Sloan (2012) consider quadrature for

$$f(\boldsymbol{x}) = \frac{1}{1 + \sum_{j=1}^{d} x_j^{\alpha}/j!}, \quad 0 < \alpha \leqslant 1.$$

For $\alpha = 1$ and $d = 500$

$R = 50$ replicated estimates of $\sum_v |v| \sigma_v^2/\sigma^2$ using $n = 10{,}000$

had mean 1.0052 and standard deviation 0.0058.

Upshot

$f(\boldsymbol{x})$ is nearly additive

mean dimension between 1.00356 and 1.00684

(± 2 standard errors)

Lowering effective dimension

Sometimes we can make f more suited to QMC.

E.g., cram importance into first few components of $\boldsymbol{x}$.

MC vs QMC

MC places lots of effort on variance reduction.

For QMC we gain by reducing effective dimension.

Or Hardy-Krause variation (but there asymptotics are slow).

E.g., when turning $\boldsymbol{u} \sim \mathbf{U}[0,1]^d$ into $\boldsymbol{z} \sim \mathcal{N}(0, \Sigma)$

The choice of $\Sigma^{1/2}$ affects QMC performance.

- Caflisch, Morokoff & O (1997)
- Acworth, Broadie & Glasserman (1998)
- Imai & Tan (2014)

Best Σ can depend strongly on f.

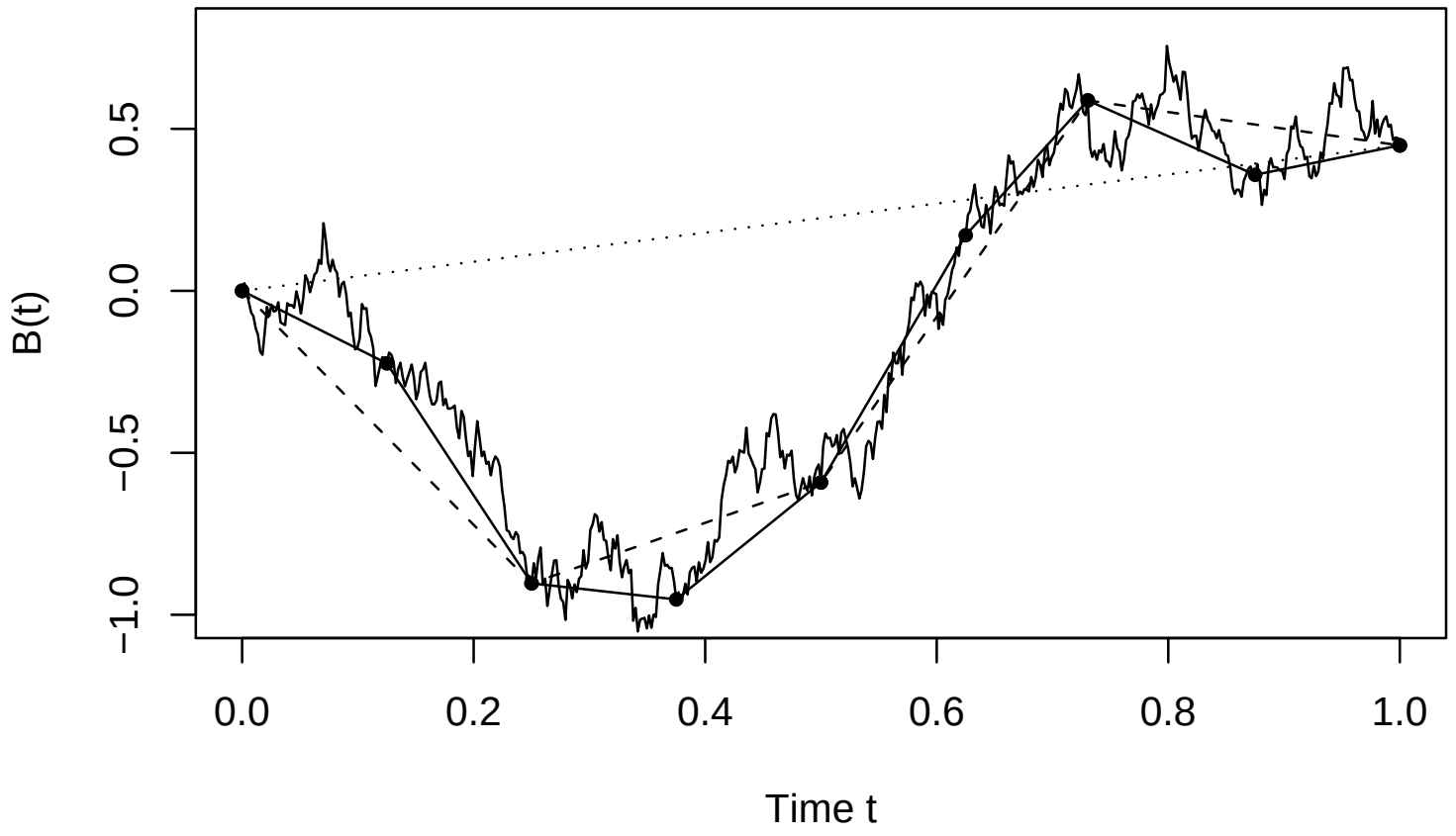
- Papageorgiou (2002)

Sampling Brownian motion

Feynman-Kac/Brownian bridge

First few variables define a ‘skeleton’. The rest fill in.

Brownian bridge construction of Brownian motion



See also Mike Giles++ on multi-level MC.

Choosing γ

Each γ corresponds to a reproducing kernel Hilbert space (RKHS)

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 $j = 1, \ldots, J$
   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 $\widetilde{f} \in \mathcal{H}$

Bayes and empirical Bayes ideas might help choose $\mathcal{H}$

Maybe we want an $\mathcal{H}$ where f is 'typical'.

A natural γ has

$$\gamma_u \propto \int_{[0,1]^d} \left(\partial^u f(\boldsymbol{x})\right)^2 \mathrm{d}\boldsymbol{x}$$

NB: the constant of proportionality is also important

Thanks

- Lecturers: Nicolas Chopin, Mark Huber, Jeffrey Rosenthal
- Guest speakers: Michael Giles, Gareth Roberts
- The London Mathematical Society: Elizabeth Fisher, Iain Stewart
- CRISM & The University of Warwick, Statistics
- Sponsors: Amazon, Google
- Partners: ISBA, MCQMC, BAYSM
- Poster: Talissa Gasser, Hidamari Design
- NSF: DMS-1407397 & DMS-1521145
- Planners: Murray Pollock, Christian Robert, Gareth Roberts
- Support: Paula Matthews, Murray Pollock, Shahin Tavakoli