QMC Beyond the Cube

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. QMC Beyond the Cube
4. QMC and Variable Importance

A small number of corrections have been made since then.

QMC

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

Koksma-Hlawka

$$|\hat{\mu} - \mu| \leqslant D_n^*(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n) \times \|f\|_{\mathrm{HK}}$$

Discrepancy is with respect to **axis-oriented** boxes $[\mathbf{0},\boldsymbol{a}]$ or $[\boldsymbol{a},\boldsymbol{b}]$

Variation is based on **axis-oriented** differences of differences

Non-cubic domains

$$\mu = \int_\Omega f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

Triangle Simplex Cylinder

Disk Sphere Ball Spherical triangle

What axes?

For discrepancy and variation

Cartesian products

$$\Omega = \prod_{j=1}^{s} \Omega_j, \quad \Omega_j \subset \mathbb{R}^{d_j}$$

Disk $\times$ Sphere $\times$ Sphere $\times$ Interval $\times \cdots \times$ Spherical triangle

The cube

“You’ll never get out of the cube.”

The Cube Jim Henson & Jerry Jule (1969)

is a surreal and somewhat grim film about being stuck in a cube.

Other people can get in and out

German version

KUBUS by glassbooth (2008)

General measures

$D_n^*(\cdot;\mu) = D_n^*(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n;\mu)$ is star discrepancy wrt measure μ

Theorem from 'Gates of Hell' paper

Aistleitner, Bilyk & Nikolov (2016), For any normalized measure μ on $\mathbb{R}^d$ there exist points with $D_n^*(\cdot;\mu) \leqslant \log(n)^{d-1/2}/n$

References from GoH paper

- Aistleitner & Dick (2015)

  discrepancy and Koksma-Hlawka for general signed measures.

- Aistleitner & Dick (2014) For any normalized measure μ on $[0,1]^d$,

  $$D_n^*(\cdot;\mu) \leqslant 63\sqrt{d}\bigl(2+\log_2(n)^{(3d+1)/2}\bigr)/n.$$

- Beck (1984) had $\log(n)^{2d}$.

- Götz (2002) first Koksma-Hlawka for general measures.

QMC sampling

We emphasize **constructions**

1) Measure preserving maps from $[0, 1]^d$ onto Ω, and

2) Direct constructions, e.g., by recursively partitioning Ω.

Existence proofs

For users, they are frustrating.

- Constructions say how to do something.

  Yes! You can do this.

Existence proofs

For users, they are frustrating.

- Constructions say how to do something.

  Yes! You can do this.

- Non-existence results show that constructions don't exist.

  No! You can't do that.

Existence proofs

For users, they are frustrating.

- Constructions say how to do something.

  Yes! You can do this.

- Non-existence results show that constructions don't exist.

  No! You can't do that.

- Existence proofs show that non-existence proofs don't exist.

  Maybe! Keep looking.

However

They can be interesting, elegant or deep.

(And may hint at constructions.)

Non-cubic domains

We want

$$\mu = \int_{\Omega} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}, \quad \text{bounded } \Omega \subset \mathbb{R}^d, \quad \mathbf{vol}(\Omega) = 1$$

Transformations

For measure preserving $\tau : [0,1]^s \to \Omega$

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} (f \circ \tau)(\boldsymbol{x}_i), \quad \boldsymbol{x}_i \in [0,1]^s$$

But $f \circ \tau$ might not be well behaved. No problem for MC; challenge for QMC.

Choices for τ

Devroye (1986), Fang & Wang (1994), Pillards & Cools (2005)

The triangle

Brandolini, Colzani, Gigante & Travaglini (2013)

- define a 'trapezoid discrepancy' in the simplex and a variation
- prove a Koksma-Hlawka inequality

but gave no constructions of points with vanishing discrepancy.

Pillards & Cools (2005)

- lots of measure preserving mappings
- get variation & discrepancy & Koksma-Hlawka

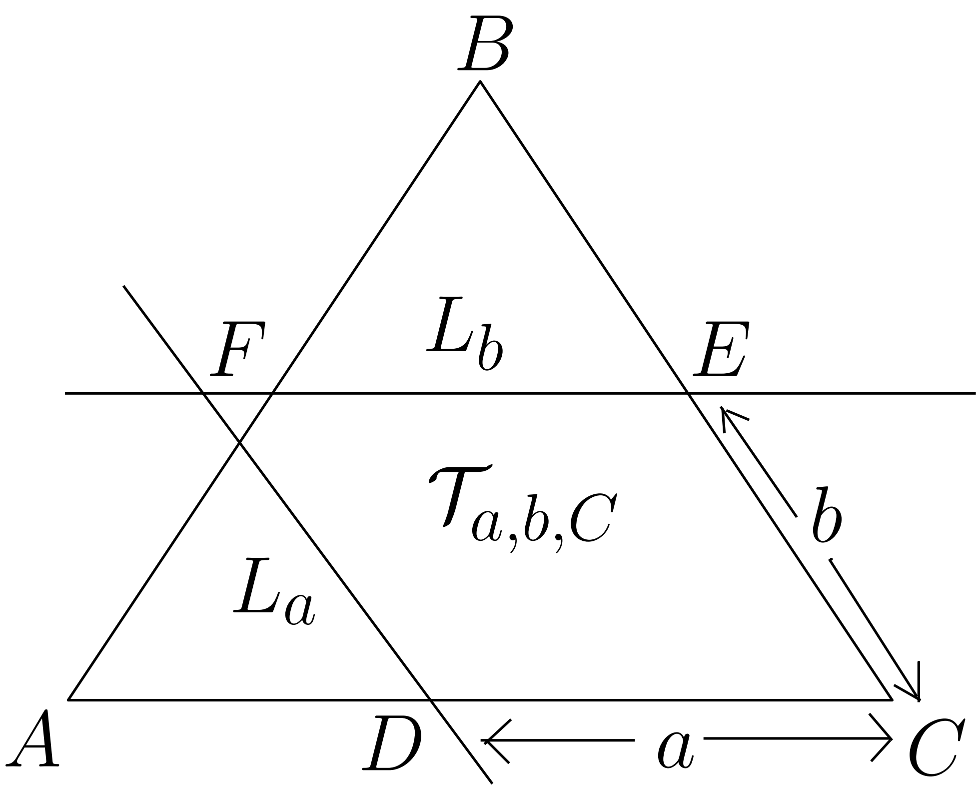
but gave no conditions for vanishing discrepancy of transformed points.

Chen & Travaglini (2013)

prove existence of point sets with vanishing trapezoid discrepancy for the triangle

Trapezoid discrepancy

Brandolini et al. (2013,2014)



$\Omega = \triangle(A, B, C)$

Discrepancy for $\mathcal{T}_{a,b,C} \cap \Omega$

$\sup$ over trapezoids

Corresponding variation

Elegant argument $\cdots$

$\cdots$ extends to simplices

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 $x_1 = 0$.

Triangular van der Corput

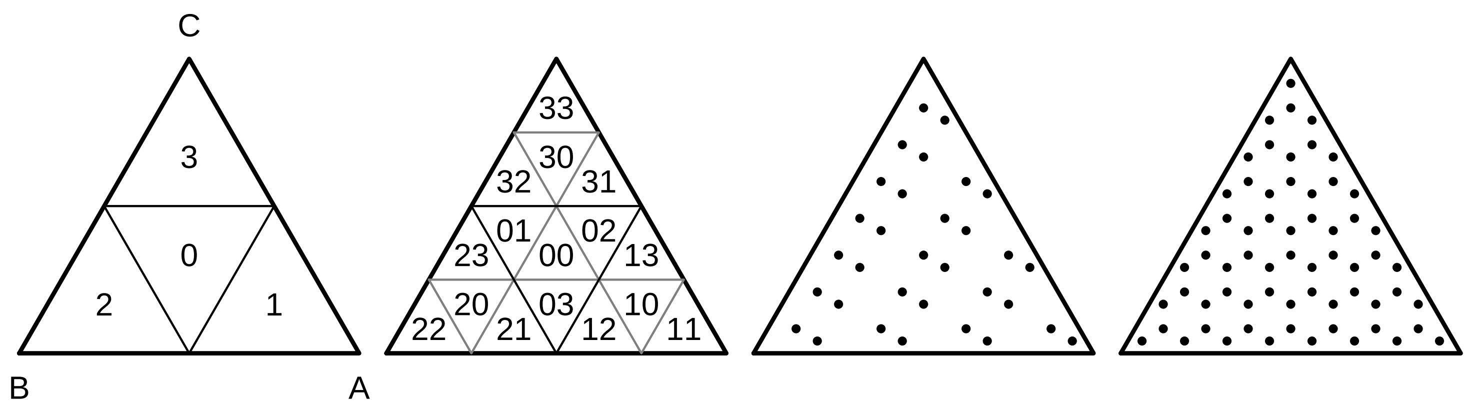
For i'th point in $T = \triangle(A, B, C)$, write

$$i = \sum_{k=1}^{K_i} d_{k,i} 4^{k-1}, \quad d_{k,i} \in \{0, 1, 2, 3\}$$

Split T into 4 congruent sub-triangles, $T(0), T(1), T(2), T(3)$

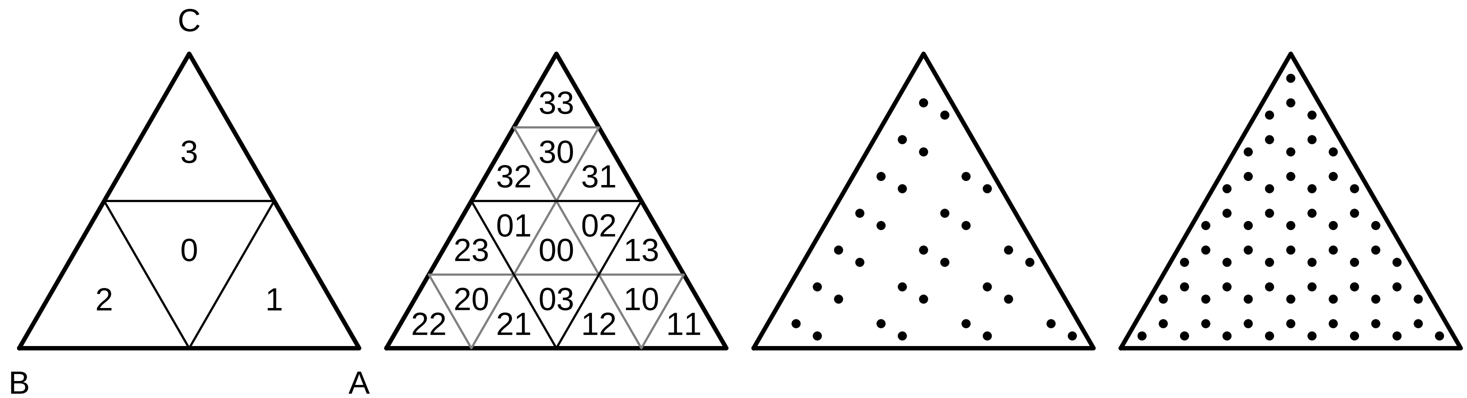
Place $\boldsymbol{x}_i$ in $T(d_{1,i})$

Recurse



Basu & O (2015)

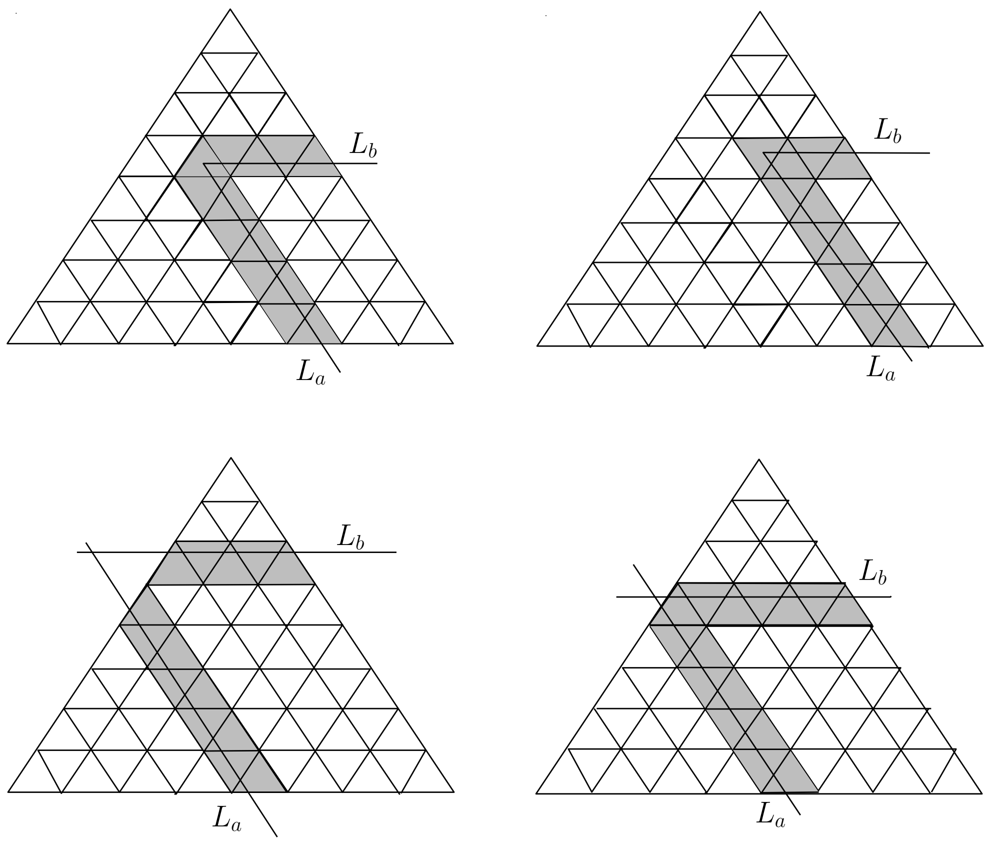
Construction continued



Corners of the subtriangle

$$T(d) = \begin{cases} \triangle\left(\frac{B+C}{2}, \frac{A+C}{2}, \frac{A+B}{2}\right), & d = 0, \\ \triangle\left(A, \frac{A+B}{2}, \frac{A+C}{2}\right), & d = 1, \\ \triangle\left(\frac{A+B}{2}, B, \frac{B+C}{2}\right), & d = 2, \\ \triangle\left(\frac{A+C}{2}, \frac{B+C}{2}, C\right), & d = 3. \end{cases}$$

For $n = 4^k$



- n subtriangles, 1 point each
- all discrepancy from within shaded triangles
- enumerate all possibilities
- upright vs inverted are different

Results

Let D_n^P be (anchored) parallelogram discrepancy.

First $n = 4^k$ points

$$D_n^P = \begin{cases} \frac{7}{9}, & n = 1 \\ \frac{2}{3\sqrt{n}} - \frac{1}{9n} & \text{else} \end{cases}$$

Any consecutive $n = 4^k$ points

$$D_n^P \leqslant \frac{2}{\sqrt{n}} - \frac{1}{n}$$

First n points

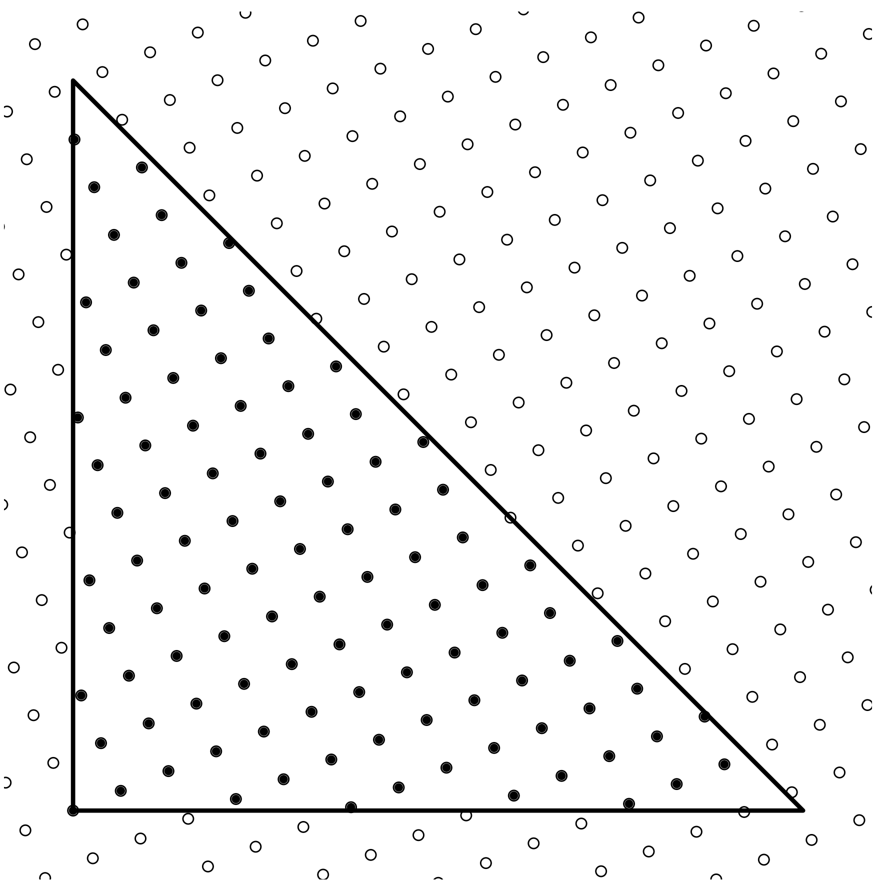
$$D_n^P \leqslant \frac{12}{\sqrt{n}}$$

Basu & O (2015)

This rate is not optimal

Kronecker lattice in the triangle

Basu & O (2015)



1) Place a square grid in $\mathbb{R}^2$
2) Rotate it α radians
3) Intersect with right triangle
4) Linear map to desired $\triangle$

Critical: choose good α

Kronecker continued

$\theta \in \mathbb{R}$ is **badly approximable** if there exists $c > 0$ with

$$\mathbf{dist}(n\theta, \mathbb{Z}) > c/n, \quad \forall n \in \mathbb{N}$$

Quadratic irrationals $\theta = (a + b\sqrt{c})/d$ are badly approximable.

Here $a, b, c, d \in \mathbb{Z}, \quad b, d \neq 0,$ square free $c > 1$

Chen & Travaglini (2007) There exist points with

Polygon discrepancy $= O(\log(n)/n)$

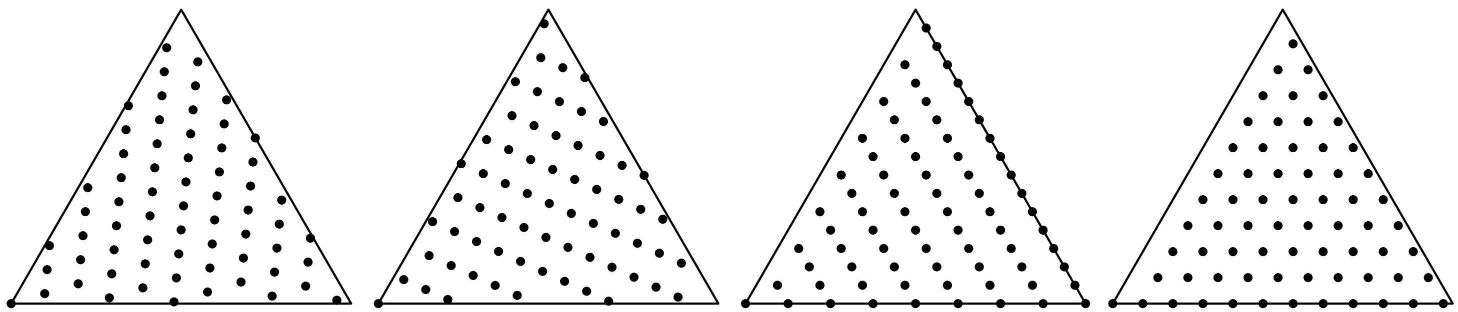
Basu & O (2015) For trapezoids:

rotate a grid by α radians where $\tan(\alpha)$ is a quadratic irrational.

E.g., for $\alpha = 3\pi/8, \quad \tan(\alpha) = 1 + \sqrt{2}$

Triangular Kronecker

Triangular lattice points



Angle 3pi/8	Angle 5pi/8	Angle pi/4	Angle pi/2
$\checkmark$	$\checkmark$	X	X

A grid with a ‘Kronecker rotation’ gets $D_n^P = O(\log(n)/n)$. Basu & O (2015)

This is the best possible rate. Chen & Travaglini (2013)

Generalization

Hexagon $=$ six triangles, et cetera

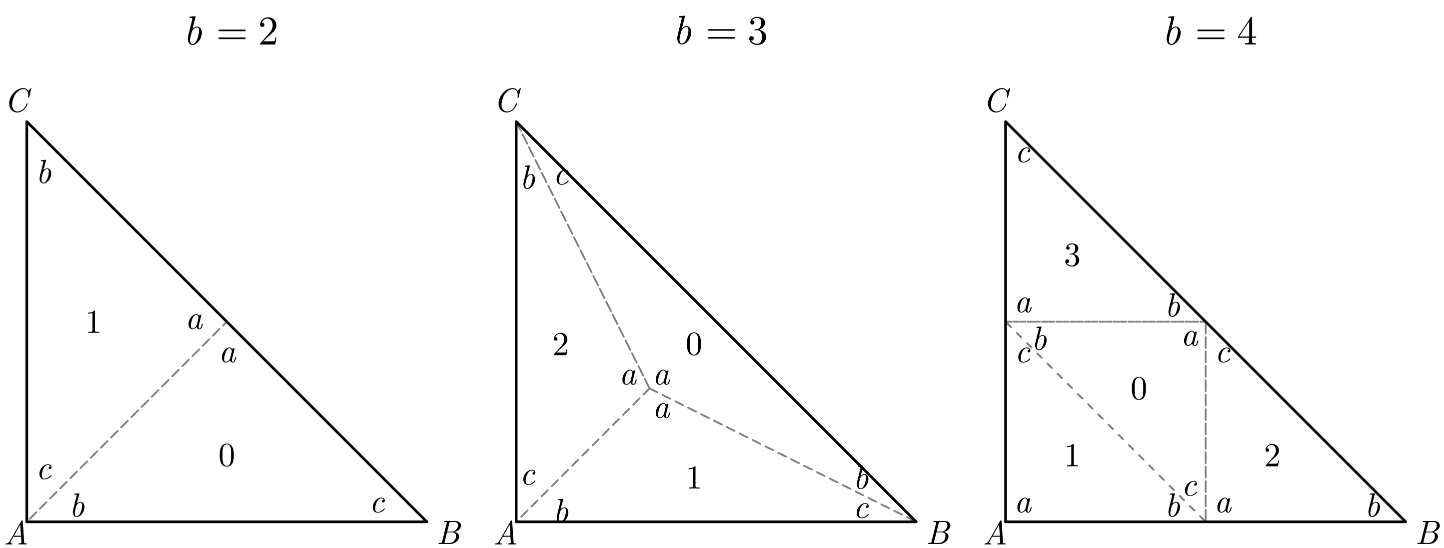
Very unlikely to generalize to higher dimensional simplices or Cartesian products of simplices. (D. Bilyk personal communication)

Geometric van der Corput

Map $i = 1, 2, 3, \dots$ into $\boldsymbol{x}_i \in \Omega$.

- replace triangle by more general set Ω
- split Ω into b equal volumes
- recursively

Splits of a triangle

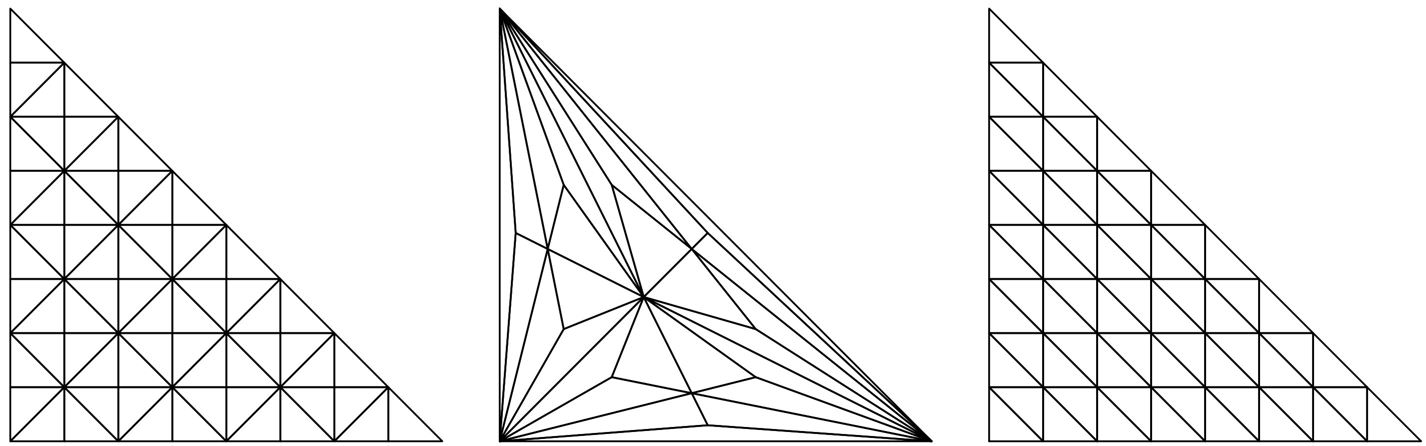


The triangle can be recursively split 2-fold, 3-fold or 4-fold.

This allows digital constructions in those bases.

Not all splits work well

2^6 Decomposition 3^3 Decomposition 4^3 Decomposition



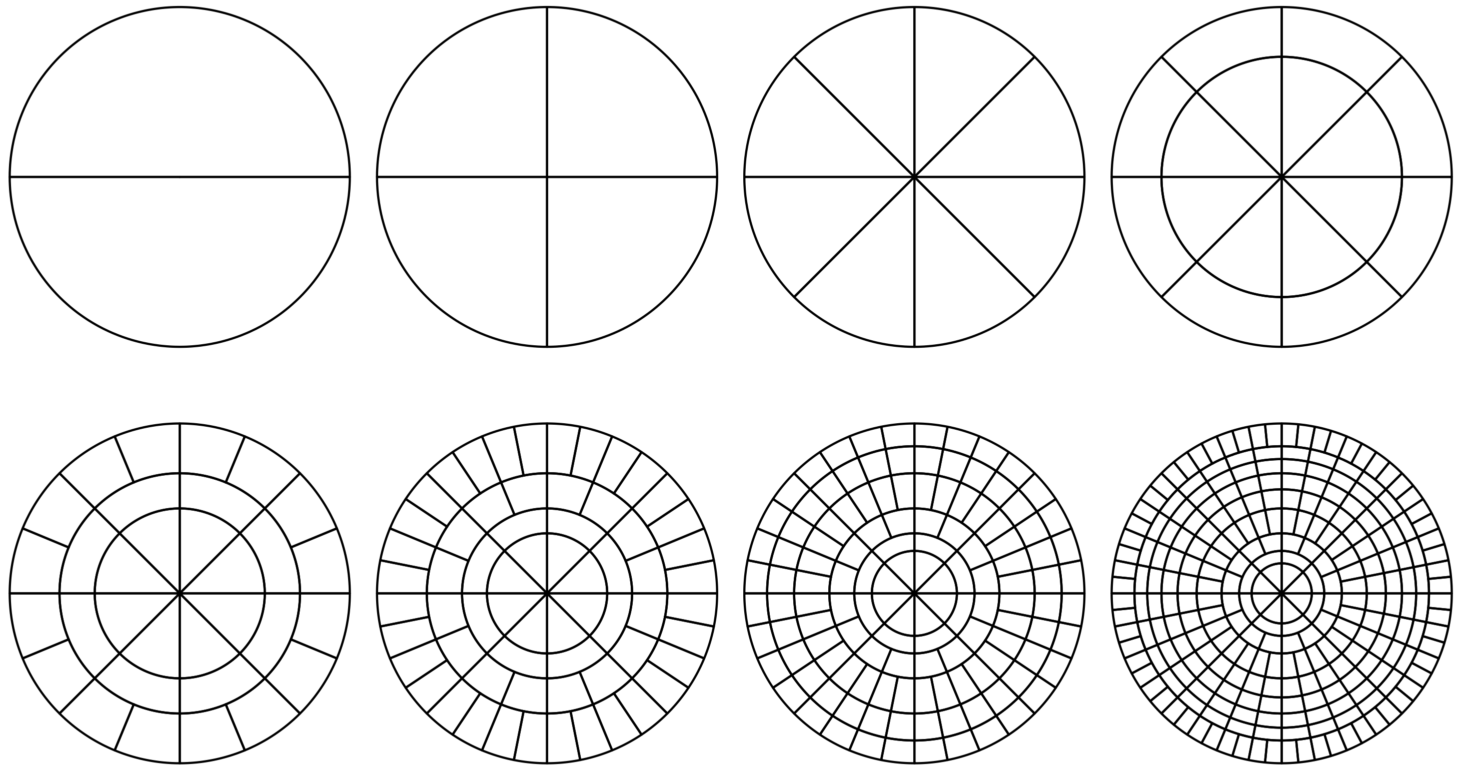
The base 3 split leads to very unfavorable aspect ratios.

The regions do not 'converge nicely' to a point.

E.g., Stromberg (1994) defines 'converge nicely'

(Bounded aspect ratios.)

Splits don't have to be congruent



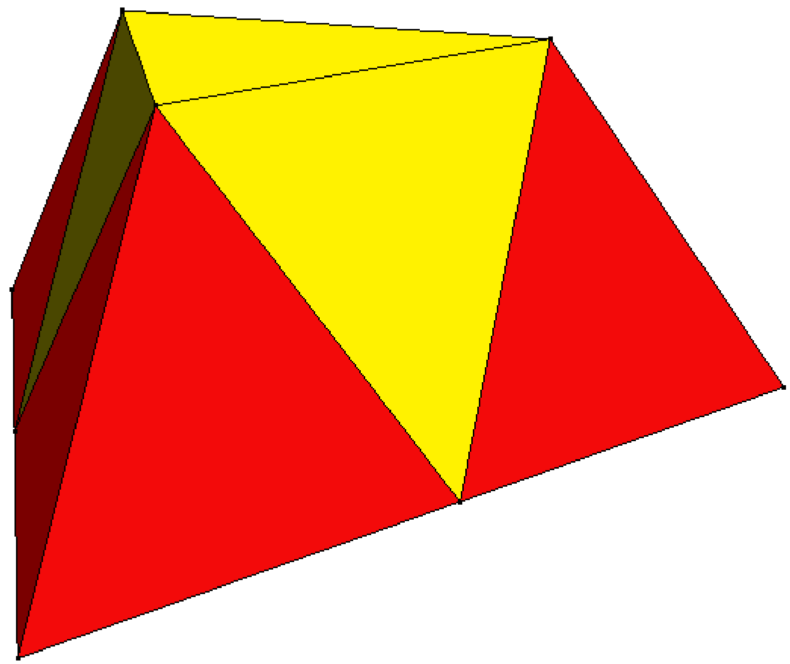
- Mix 'arc splits' and 'radial splits' to keep aspect ratio bounded
- Not a global alternation; different cells get different splits

Basu & O (2015)

See Beckers & Beckers (2012) for non-recursive splits

Tetrahedron

- chop off 4 tetrahedral corners
- remaining volume makes 4 more



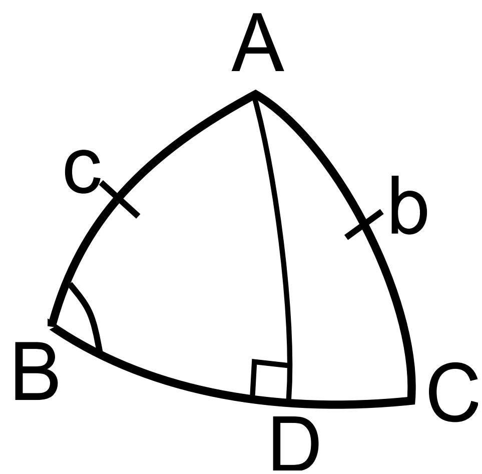
- but they're not congruent to first 4
- binary splits may be better (split a longest edge)

Image: By Tomruen - Own work, CC BY-SA 3.0, `wikipedia`

Spherical triangles

- 4 way split at arc midpoints $\cdots$ not equal area
- 4 way equal area split of Song, Kimerling, Sahr (2002) uses 'small circle' boundaries, not great circles
- binary splits may be better $\cdots$ use first step in Arvo (1995)
  Arvo's work underlies much of modern movies and computer games

More about Arvo



Arvo shows how to pick D so

$$\frac{\mathbf{vol}(ABD)}{\mathbf{vol}(ABC)} = u$$

We can use $u = 1/2$.

Image by Peter Mercator - Own work, CC BY-SA 3.0, Wikipedia

Geometric nets

We want points in Ω^s for $\Omega \subset \mathbb{R}^d$

E.g., light path

$$\text{camera} \to \triangle \to \triangle \to \triangle \to \cdots \to \triangle \to \text{light source}$$

Use digital nets

A (t, m, s)-net, $b = 4$ or $b = 2$, puts $\boldsymbol{x}_i \in \triangle^s$ (componentwise)

Use other partitions

Other b-fold equal area recursive partitions can be used for $\Omega \neq \triangle$

Scramble the nets

Unbiasedness and error cancellation benefits under smoothness.

More generally

$$\Omega = \prod_{j=1}^{s} \Omega_j, \quad \Omega_j \subset \mathbb{R}^{d_j}$$

$\tau_j : [0,1] \to \Omega_j$ digital map, base b

Take $\boldsymbol{u}_i = (u_{i1}, \ldots, u_{is}) \in [0,1]^s$,
(t,m,s)-net or (t,s)-sequence in base b.

Componentwise map: $\boldsymbol{x}_i = \tau(\boldsymbol{u}_i)$

$$\boldsymbol{x}_i = (x_{i1}, \ldots, x_{is})$$

$$x_{ij} = \tau_j(u_{ij})$$

Scrambled geometric nets

Take $\mathbf{vol}(\Omega_j) = 1$ and $\Omega = \prod_{j=1}^s \Omega_j$ and let

$$\mu = \int_\Omega f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}, \qquad \hat\mu = \frac{1}{n}\sum_{i=1}^n f(\boldsymbol{x}_i)$$

where $\boldsymbol{x}_i$ are scrambled geometric nets.

For $f \in L^2(\Omega)$

$$\mathbb{E}(\hat\mu) = \mu \qquad \mathrm{Var}(\hat\mu) = o\Bigl(\frac{1}{n}\Bigr) \qquad \mathrm{Var}(\hat\mu) \leqslant \Gamma \times \frac{\sigma^2}{n}$$

where $\sigma^2 = \int_\Omega (f(\boldsymbol{x}) - \mu)^2\,\mathrm{d}\boldsymbol{x}$, and
Γ is the largest gain coefficient of the (t, m, s)-net

E.g., $t = 0$ implies $\Gamma \leqslant \exp(1) \doteq 2.718$

Convergence rates

$\mu = \int_\Omega f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}, \qquad \Omega = \mathcal{X}^s, \quad \mathcal{X} \subset \mathbb{R}^d$

For smooth f, nested uniform scrambled nets and nice partitions

$$\mathrm{Var}(\hat\mu) = O\Bigl(\frac{(\log n)^{s-1}}{n^{1+2/d}}\Bigr)$$

Basu & O (2017)

Using smoothness of f and 'hole free' Sobol' extensibility condition on $\mathcal{X}$'s

Note

Effect of d more critical than s

Additionally

Central limit when $t = 0$ Basu & Mukherjee (2017)

Tractability over products of simplices Basu (2015)

Strong law of large numbers O & Rudolf (2020)

Transformations

Let τ transform $\mathbf{U}[0,1]^m$ into $\mathbf{U}(\Omega)$.

$$\int_\Omega f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \int_{[0,1]^m} (f\circ\tau)(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}$$

We want $f\circ\tau \in$ BVHK for QMC and mixed partials in L^2 for RQMC

BVHK compositions

For $f\circ\tau : \mathbb{R}\to\mathbb{R}\to\mathbb{R}$:

$f\in$ Lipschitz, $\tau\in$ BV $\implies$ $f\circ\tau\in$ BV. Josephy (1981)

No such simple rule in higher dimensions.

Variation is bounded via integrated absolute mixed partials.

So we study derivatives of $f(\tau(\boldsymbol{u}))$.

Faà di Bruno

Derivatives of composite functions, $\mathbb{R} \to \mathbb{R} \to \mathbb{R}$

Faà di Bruno (1855,1857), Arbogast (1800)

$$\begin{aligned} h(x) &= f(g(x)) \\ h'(x) &= f'(g(x))g'(x) \\ h''(x) &= f''(g(x))g'(x)^2 + f'(g(x))g''(x) \\ h'''(x) &= f'''(g(x))g'(x)^3 + 3f''(g(x))g'(x)g''(x) + f'(g(x))g'''(x) \end{aligned}$$

Our map is

$$\mathbb{R}^D \to \mathbb{R}^d \to \mathbb{R}$$

which has many more terms

Constantine & Savits (1996) give a general Faà di Bruno theorem

Basu & O (2016) simplify it for

$$\partial^u (f \circ \tau), u \subseteq \{1, \dots, D\}$$

i.e., differentiate at most once wrt each x_j

Allows tests of BVHK.

Some mappings

The following mappings work well for MC, but not QMC

Triangle $\mathbb{T}^2 \subset \mathbb{R}^3$

$$\boldsymbol{u} \in [0,1]^3, \quad x_j = \tau_j(\boldsymbol{u}) = \frac{\log(u_j)}{\sum_{i=1}^3 \log(u_i)} \quad \boldsymbol{x} \sim \mathbf{U}(\mathbb{T}^2)$$

Even $x_j(\boldsymbol{u}) \notin \mathrm{BVHK}([0,1]^3)$.

Sphere $\mathbb{S}^{d-1} \subset \mathbb{R}^d$

$$x_j = \tau_j(\boldsymbol{u}) = \frac{\Phi^{-1}(u_j)}{\sqrt{\sum_{i=1}^d \Phi^{-1}(u_i)^2}}, \quad \boldsymbol{x} \sim \mathbf{U}(\mathbb{S}^{d-1})$$

Again, $x_j(\boldsymbol{u}) \notin \mathrm{BVHK}([0,1]^d)$.

BVHK compositions

For $\boldsymbol{u} \in [0,1]^D$ and

$$f(\tau_1(\boldsymbol{u}), \dots, \tau_d(\boldsymbol{u}))$$

If these hold

1) $\partial^v \tau_j(\boldsymbol{u}_v{:}\mathbf{1}_{-v}) \in L^{p_j}([0,1]^{|v|}), \quad p_j \in [1,\infty] \quad v \subseteq \{1,2,\dots,D\}$
2) $\sum_{j=1}^d 1/p_j \leqslant 1$
3) $f \in C^{(d)}(\mathbb{R}^d)$

Then

$$f \circ \tau \in \mathrm{BVHK}$$

RQMC smooth

1) $\partial^v \tau_j \in L^{p_j}([0,1]^D)$, $p_j \in [2, \infty]$, and
2) $\sum_{j=1}^d 1/p_j \leqslant 1/2$
3) $f \in C^{(d)}(\mathbb{R}^d)$

make $f \circ \tau$ smooth enough for RMSE$= O(n^{-3/2+\epsilon})$ under RQMC.

$f \in C^{(d)}$ can be weakened if p_j are increased

Fang & Wang (1993)

Three mappings to a simplex, one to the sphere, and one to a ball.

Example

$$A_d = \{(x_1, \ldots, x_d) \mid 0 \leqslant x_1 \leqslant x_2 \leqslant \cdots \leqslant x_d \leqslant 1\}$$

Transformation

$$\begin{aligned}
x_1 &= \tau_1(\boldsymbol{u}) = u_1 \\
x_2 &= \tau_2(\boldsymbol{u}) = u_1 \times u_2^{1/2} \\
x_3 &= \tau_3(\boldsymbol{u}) = u_1 \times u_2^{1/2} \times u_3^{1/3} \\
&\vdots \\
x_d &= \tau_d(\boldsymbol{u}) = u_1 \times u_2^{1/2} \times u_3^{1/3} \times \cdots \times u_d^{1/d}
\end{aligned}$$

Results

All five Fang & Wang mappings τ are in BVHK.

So composing with f has a chance.

None of them yield τ with mixed partials in L^2.

So no $\mathrm{RMSE} = O(n^{-3+\epsilon})$.

Still have

$\mathrm{RMSE} = o(1/n)$ and $\mathrm{RMSE} \leqslant \Gamma\sigma^2/n$

Markov chains

Sometimes no known transformation yields

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

Markov chain Monte Carlo

$$\boldsymbol{x}_i = \phi(\boldsymbol{x}_{i-1}, \boldsymbol{u}_i), \quad \boldsymbol{u}_i \sim \mathbf{U}[0,1]^d$$

$$\boldsymbol{x}_i \xrightarrow{\mathrm{d}} p \quad \text{as } i \to \infty$$

Estimate

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

$$\hat{\mu} = \frac{1}{n}\sum_{i=b+1}^{b+n} f(\boldsymbol{x}_i) \qquad \text{burn-in}$$

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{x}_{k\times i}) \qquad \text{thinning}$$

What about QMC for MCMC?

We need a stream $\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \cdots \in [0,1]^d$

$$\underbrace{v_1, v_2, \dots, v_d}_{\boldsymbol{u}_1}\ \underbrace{v_{d+1}, v_{d+2}, \dots, v_{2d}}_{\boldsymbol{u}_2} \cdots \underbrace{v_{(n-1)d+1}, v_{(n-1)d+2}, \dots, v_{nd}}_{\boldsymbol{u}_n}$$

Now we need a “driving sequence” $v_1, v_2, \cdots \in [0,1]$

Completely uniformly distributed

$$D_n^*\Bigl((v_1,\dots,v_s),(v_2,\dots,v_{s+1}),\cdots(v_N,\dots,v_{N+s})\Bigr) \to 0 \qquad \text{all } s \geqslant 1$$

Not just $s = d$

References

Empirical better rates in thesis of Seth Tribble

Consistency results O & Tribble (2005), Chen, Dick & O (2011)

Proved better rates in thesis of Su Chen under strong conditions

Some constructions Chen, Matsumoto, Nishimura, O (2012), Harase (2021)

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

Especially Kinjal Basu