How to beat randomness?

A Quasi-Monte Carlo methods overview

Pierre Marion

Sorbonne Université - Corps des Mines

pierre.marion@mines.org

Joint work with Pierre L'Ecuyer and Maxime Godin

November 15, 2019

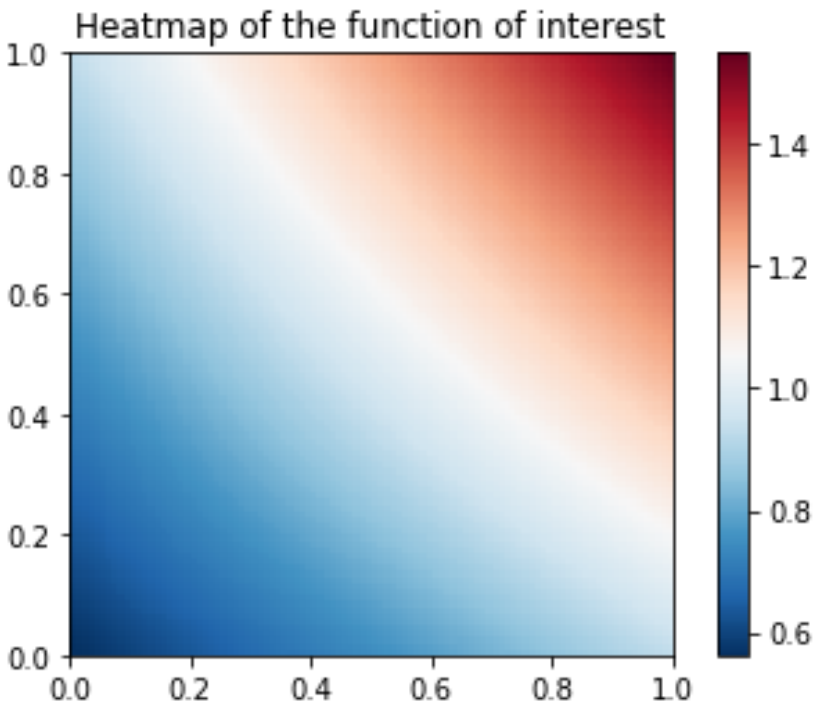
Overview

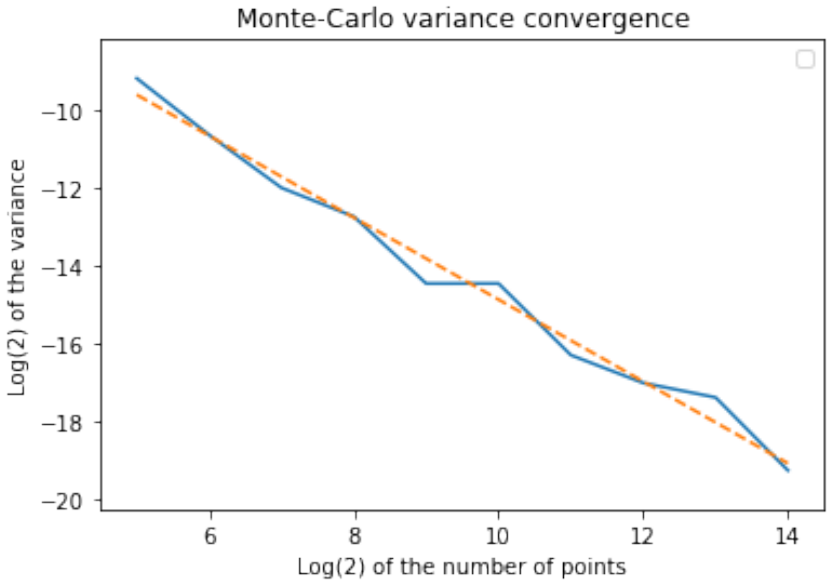
Monte-Carlo (MC) methods

- Invented in 1946 by Ulam, von Neumann, Metropolis (Los Alamos National Laboratory)
- Basic idea: solve problems using random sampling
- Applications for a huge number of purposes:
  - Simulation of random processes
  - Quantum Monte Carlo
  - Integration
  - Monte-Carlo Tree Search
  - Markov Chain Monte Carlo
  - etc., etc.

MC for integration

$$f(x) = \prod_{i=1}^{2} \left(1 + \frac{1}{2}\left(x_i - \frac{1}{2}\right)\right)$$





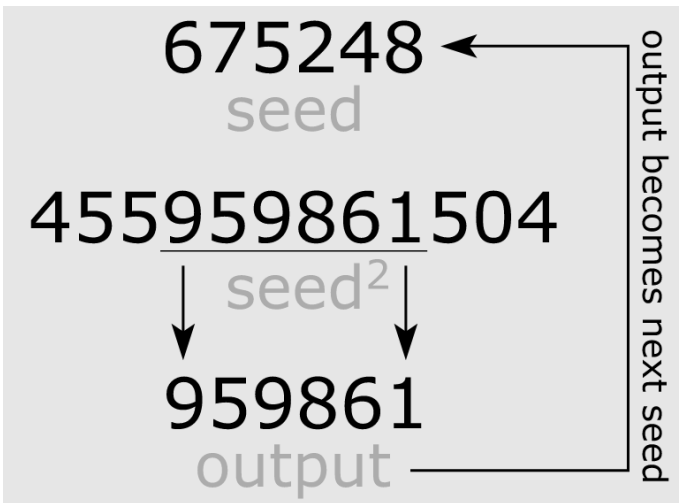
The integration error decreases as $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$.

Random number sampling

- True random number generators
  - Atmospheric noise produced in the ionosphere
  - `https://www.random.org/`
- List of random digits
  - First table (41,600 digits) by L.H.C Tippett in 1927
  - 'A Million Random Digits with 100,000 Normal Deviates' by the RAND Corporation in 1947

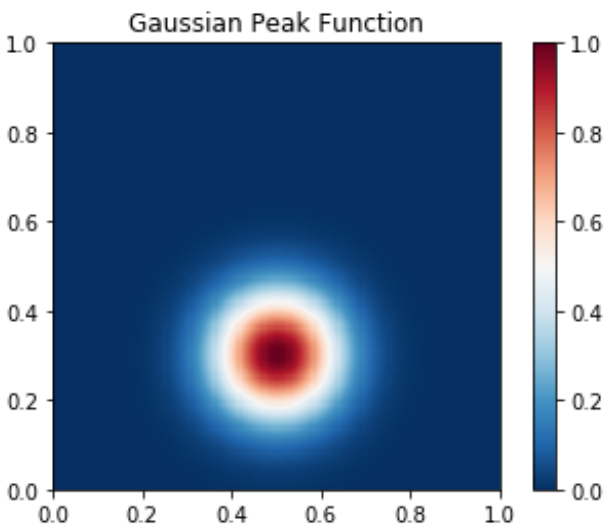
73735	45963	78134	63873
02965	58303	90708	20025
98859	23851	27965	62394
33666	62570	64775	78428
81666	26440	20422	05720

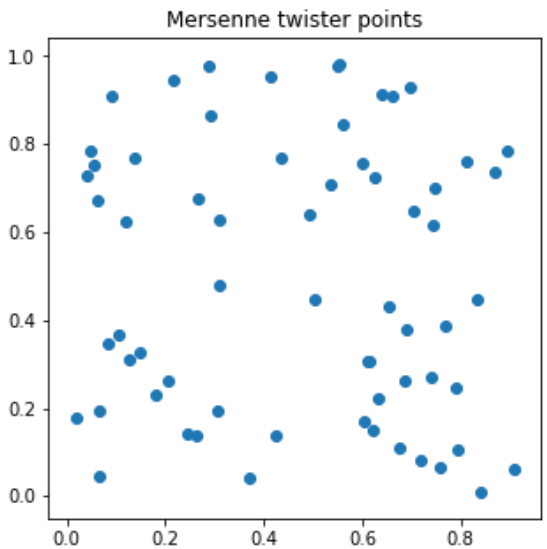
- Pseudorandom number generators
  - Middle-square method invented by John von Neumann (1949)

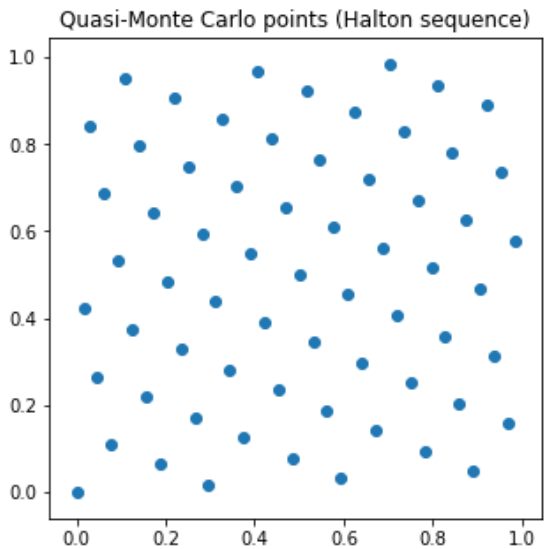


From pseudorandom to Quasi-Monte Carlo

- Modern pseudorandom number generators are based on linear algebra over $\mathbb{F}_2$ and are actually deterministic!
  - e.g. Mersenne Twister by Matsumoto, Nishimura (1997)
- Quasi-Monte Carlo methods as well!
  - Idea: obtain point sets with better equidistribution properties
  - Hope: better convergence rate than MC: $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$







Quasi-Monte Carlo roadmap

We need four ingredients:

1. a parametrized method to construct equidistributed point sets
   $\rightarrow$ Digital nets
2. a proper definition of equidistribution
   $\rightarrow$ t-value
3. a proof on integration error bounds
   $\rightarrow$ Koksma-Hlawka inequality
4. a method to find the best possible parameters
   $\rightarrow$ LatNet Builder software

The bible of digital nets: [J. Dick, F. Pillichshammer, 2010]

Digital nets: definition through example

$$\text{Let } M_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \text{ and } M_2 = \begin{bmatrix} 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \end{bmatrix}$$

With 16 points, how to generate the 7-th point?
$i = 7$ binary representation least-significant bits first: $[1, 1, 1, 0]$
multiply by one matrix per coordinate
Obtain

$$x = [1, 1, 1, 0] \text{ and } y = [0, 1, 1, 1]$$

then take the binary representation most-significant bits first, and divide by 2^4:

$$x = \frac{13}{16} \text{ and } y = \frac{7}{16}$$

Digital nets: a proper definition

Definition (Digital net)

Let $\mathbb{F}_2$ be the finite field with 2 elements (denoted 0 and 1). For a given dimension $s \geq 1$, positive integer k, let $M_1, \ldots, M_s$ be $k \times k$ matrices over $\mathbb{F}_2$.

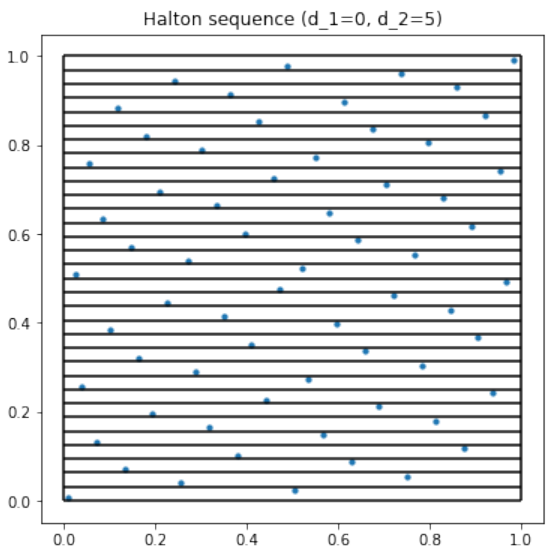
The *digital net* generated by these matrices is the s-dimensional point set $\{x_0, \ldots, x_{2^k-1}\}$ of cardinality 2^k defined as follows: for each $0 \leq i \leq 2^k - 1$, let $i = (a_0, \ldots, a_{k-1})^T$ be the digits (in $\mathbb{F}_2$) of the 2-adic expansion $i = \sum_{j=0}^{k-1} a_j 2^j$. For each coordinate l, let $y_l = M_l \times i = (y_{i,l,1}, \ldots, y_{i,l,r})^T$. Write $x_i = (x_{i,1}, \ldots, x_{i,s})^T$. Then:

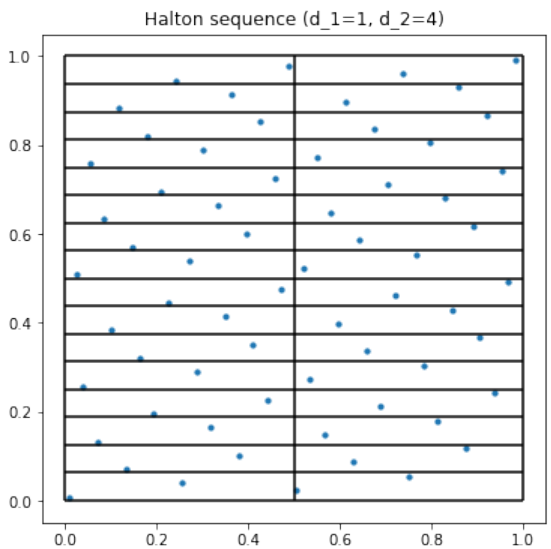
$$x_{i,l} = \sum_{j=1}^{k} y_{i,l,j} 2^{-j}.$$

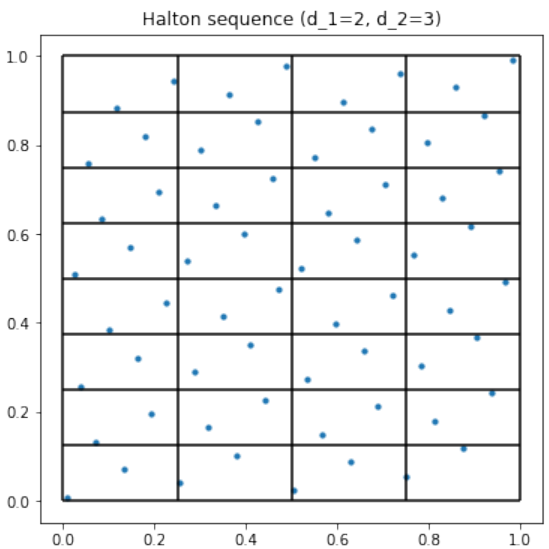
An equidistribution criterion: the t-value

$$J = \prod_{l=1}^{s} \left[\frac{A_l}{2^{d_l}}, \frac{A_l + 1}{2^{d_l}} \right) \text{ where } d_1 + \cdots + d_s = k - t \text{ and } A_l \in [0, 2^{d_l})$$

Example (The Halton sequence defines a $(1, 6, 2)$-net)







An equidistribution criterion: the t-value

Definition ((t, k, s)-net, t-value)

For a given dimension $s \geq 1$, a positive integer k and an integer t with $0 \leq t \leq k$, a point set $\mathcal{P}$ of 2^k points in $[0,1)^s$ is called a (t, k, s)-net if every interval of the form

$$J = \prod_{l=1}^{s} \left[\frac{A_l}{2^{d_l}}, \frac{A_l + 1}{2^{d_l}} \right) \text{ where } d_1 + \cdots + d_s = k - t \text{ and } A_l \in [0, 2^{d_l})$$

contains exactly 2^t points of $\mathcal{P}$. These intervals are of volume 2^{k-t}. The t-value of the net is the smallest t verifying this property, and we denote it $t(S)$ for a net generated by the set of matrices S.

An integration error bound

Theorem (Koksma-Hlawka inequality for (t, k, s)-nets)

W denotes the Sobolev space of functions on $[0,1)^s$ and is equipped with the norm

$$||f|| = \sum_{u \subseteq \{1,\dots,s\}} \int_{[0,1]^{|u|}} \left| \frac{\partial^{|u|}}{\partial z_u} f(z_u, 1) \right| dz_u.$$

For every $f \in W$, for every (t, k, s)-net $\mathcal{P} \subseteq [0,1)^s$,

$$\left| \int_{[0,1]^s} f(u)du - \frac{1}{n} \sum_{i=0}^{n-1} f(x_i) \right| \leq \frac{||f||\, 2^t}{n} \sum_{i=0}^{s-1} \binom{\log_2(n) - t}{i}.$$

The integration error decreases as $\mathcal{O}\left(\frac{(\log n)^s}{n}\right)$.

Searching for good digital nets

LatNet Builder: A General Software Tool For Constructing Highly Uniform Point Sets

- Developed at University of Montreal since 2012
- The most complete open-source software for QMC point set generation
- Written in modern C++ with interfaces in Python and Java
- Integrated with the Stochastic Simulation in Java software

`https://github.com/umontreal-simul/latnetbuilder`

Demo time!

Reminder: what is the t-value?

Definition ((t, k, s)-net, t-value)

For a given dimension $s \geq 1$, a positive integer k and an integer t with $0 \leq t \leq k$, a point set $\mathcal{P}$ of 2^k points in $[0,1)^s$ is called a (t, k, s)-net if every interval of the form

$$J = \prod_{l=1}^{s} \left[\frac{A_l}{2^{d_l}}, \frac{A_l + 1}{2^{d_l}} \right) \text{ where } d_1 + \cdots + d_s = k - t \text{ and } A_l \in [0, 2^{d_l})$$

contains exactly 2^t points of $\mathcal{P}$. These intervals are of volume 2^{k-t}. The t-value of the net is the smallest t verifying this property, and we denote it $t(S)$ for a net generated by the set of matrices S.

How to compute the t-value?

Definition (composition, linear independence parameter)

A *weak composition* of r in s parts is a tuple of non-negative integers $d_1, ..., d_s$ such that $\sum_{l=1}^{s} d_l = r$
For a set $S = \{M_1, ..., M_s\}$, the *linear independence parameter* $\rho(S)$ is the maximum integer r such that for every composition $(d_1, ..., d_s)$ of r in s parts, the d_1 first rows of M_1, the d_2 first rows of M_2, ..., up to the d_s first rows of M_s are all linearly independent.

$$C = \begin{bmatrix} M_1^{(d_1)} \\ M_2^{(d_2)} \\ \dots \\ M_s^{(d_s)} \end{bmatrix}.$$

Key property: $t(S) = k - \rho(S)$

Computation of the linear independence parameter

$$M_1 = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \quad M_2 = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix} \quad M_3 = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

To test if $\rho(M_1, M_2, M_3) \geq 3$, check if all these compositions matrices are full rank:

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \quad \ldots$$

Computation of the linear independence parameter (cont.)

- How to cycle through all the compositions?
  It is possible to cycle through the compositions changing one row at a time.
- How to update the reduction of a matrix when one row changes?
  By 'de-pivoting' the old row, replacing it with the new row, and pivoting the new row.
  It is possible with a *linear complexity*.

Wrap-up: complexity of the computation of the t-value

Theorem ([P.M., M. Godin, P. L'Ecuyer, 2019])

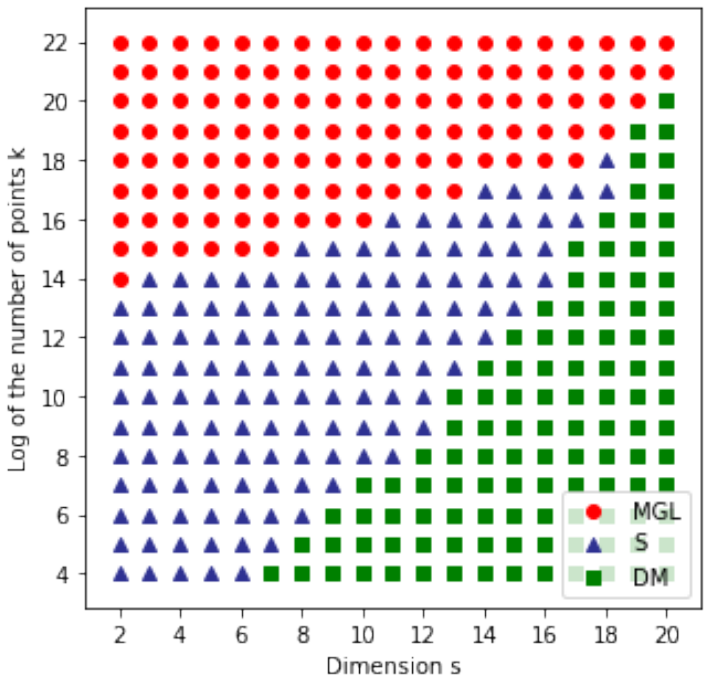
For every net $\mathcal{P} \subseteq [0,1)^s$ composed of 2^k points, its t-value can be computed with a time complexity of

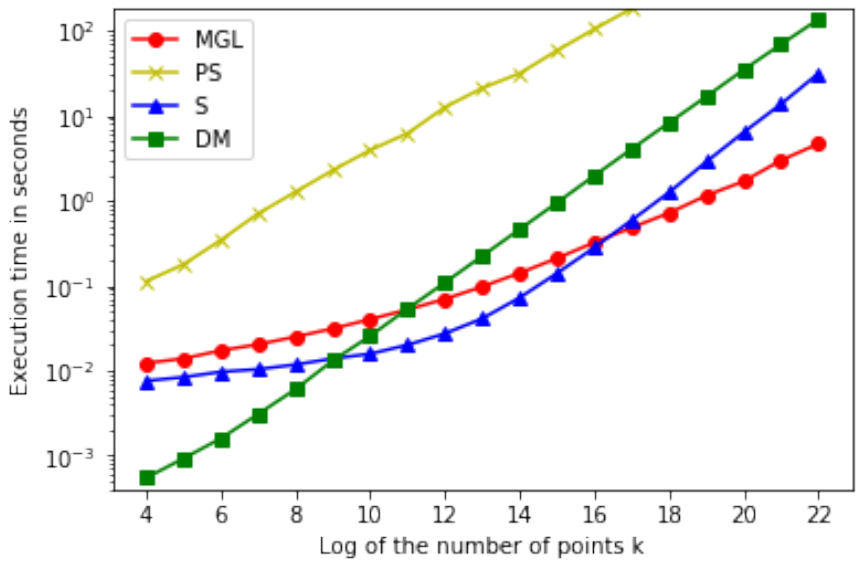
$$\mathcal{O}\left(k\binom{s+k}{s}\right) \sim \mathcal{O}\left(\frac{k^{s+1}}{s!}\right)$$

Baselines:

- [J. Dick, M. Matsumoto, 2013]: $\mathcal{O}(ks2^k)$
- [W.C. Schmid, 1999]: $\mathcal{O}(2^{k+s-1})$

Numerical experiments





Speed comparison for four methods to compute the t-value of a digital net with 2^k points, in s dimensions. In the left panel, the symbol indicates which method is fastest as a function of (s, k). In the right panel, we fix $s = 12$ and plot the execution time as a function of k, in log-log scale.

Take-home messages

QMC can lead to significant speed-up of your random algorithms

Whole theory with well-understood behavior (error bounds)

Software development efforts have been made to bring to you (relatively) easy ways to use QMC methods

U. of Montreal: LatNet Builder, Stochastic Simulation in Java
KU Leuven (D. Nuyens): `https://people.cs.kuleuven.be/~dirk.nuyens/`

Many extensions of the QMC framework

- Weighted projections
- Randomized Quasi-Monte Carlo (RQMC)
- Array-RQMC

References


W.C. Schmid (1999)
The exact quality parameter of nets derived from Sobol' and Niederreiter sequences
Recent Advances in Numerical Methods and Applications (1999) 287–295


J. Dick, F. Pillichshammer (2010)
Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration
Cambridge University Press, Cambridge, U.K., 2010


J. Dick, M. Matsumoto (2013)
On the fast computation of the weight enumerator polynomial and the t value of digital nets over finite Abelian groups
SIAM Journal on Discrete Mathematics 27 (2013) 1335–1359


P.M., M. Godin, P. L'Ecuyer (2019)
An algorithm to compute the t-value of a digital net and of its projection
eprint arXiv:1910.02277

Questions?