

Chapter 6



Quasi-Monte Carlo Methods (I)

Outline

- ◆ General principles of QMC;
- ◆ Measures of Uniformity;
- ◆ Error bounds;
- ◆ Low discrepancy sequences and their constructions
(Halton, Sobol, Faure and Niederreiter sequences);
- ◆ Good lattice rules; Korobov construction;
- ◆ Randomized quasi-Monte Carlo methods;
- ◆ Reproducing kernel Hilbert spaces,
weighted function spaces;
- ◆ ANOVA decomposition and effective dimensions;
- ◆ Advanced dimension reduction techniques.

1. General Principles of QMC

- Consider the evaluation of the integral:

$$I[f] = \int_{[0,1]^d} f(x) dx = E[f(X)].$$

Recall:

MC approximates the expectation/integral by

$$Q_N[f] = \frac{1}{N} \sum_{i=1}^N f(x_i), \quad x_i \sim U(0,1)^d,$$

where x_1, \dots, x_N are i.i.d. random samples from $U(0,1)^d$.

Recall: Key Features of MC

- **Unbiased:** $E[Q_N(f)] = I(f)$.
- **Convergence** (by strong law of large numbers):

$$\lim_{N \rightarrow \infty} Q_N[f] = I[f] \text{ (with probability 1).}$$

- **Error:** by Central Limit Theorem

$$\frac{\sqrt{N} (I[f] - Q_N[f])}{\sigma} \sim_d N(0,1),$$

$$\text{RMSE} = \sqrt{E[(I(f) - Q_N(f))^2]} = \sqrt{\sigma^2 / N}.$$

MC error is $O(N^{-1/2})$ independent of the dimension
--- breaks the curse of dimensionality.

Deficiencies of MC

- There are only probability error bounds (i.e., in any one instance one cannot be sure of integration error);
- The regularity of the integrand is not reflected (except that the function is square-integrable);
- The convergence order is very slow (and various variance reduction techniques can not increase the convergence order);
- Generating random samples (from complicated distributions) can be very difficult.

The Basic Idea of QMC

Replace random numbers in MC by deterministic numbers, which are cleverly constructed and are more uniformly distributed.

In QMC, to approximate the integral

$$I[f] = \int_{[0,1]^d} f(x)dx = E[f(X)]$$

we use the **same form** of approximation as in MC:

$$\mathcal{Q}_N[f] = \frac{1}{N} \sum_{i=1}^N f(x_i).$$

The points x_i are deterministic, which are cleverly and carefully chosen and are more uniformly distributed than random numbers.

1. General Principles of QMC

- We want a sequence x_1, x_2, \dots for which

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N f(x_i) = \int_{[0,1]^d} f(x) dx \quad (*)$$

holds for a reasonable class of integrands
(say, for all Riemann integrable function).

- A sequence $\{x_i\}$ is said to be **uniformly distributed** in $[0,1]^d$, if $(*)$ is satisfied, or equivalently,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N I_J(x_i) = m(J), \quad (**)$$

for all subintervals $J=[a, b]$ of $[0,1]^d$, where I_J is characteristic function and $m(J)$ is the measure of J .

- Equation $(**)$ says that each subinterval J should contain asymptotically the **right portion** of points.⁷

1. General Principles of QMC

- It is now clear that in order to obtain a QMC rule which converges to the actual value of the integral, the underlying nodes **must come from a uniformly distributed sequence**.

- **Questions:**

- How can we assess the quality of a point set?
- Can we obtain faster convergence than MC?
- How to construct points with better uniformity than random points?
- In which case or in which kind of applications, QMC can be better than MC?

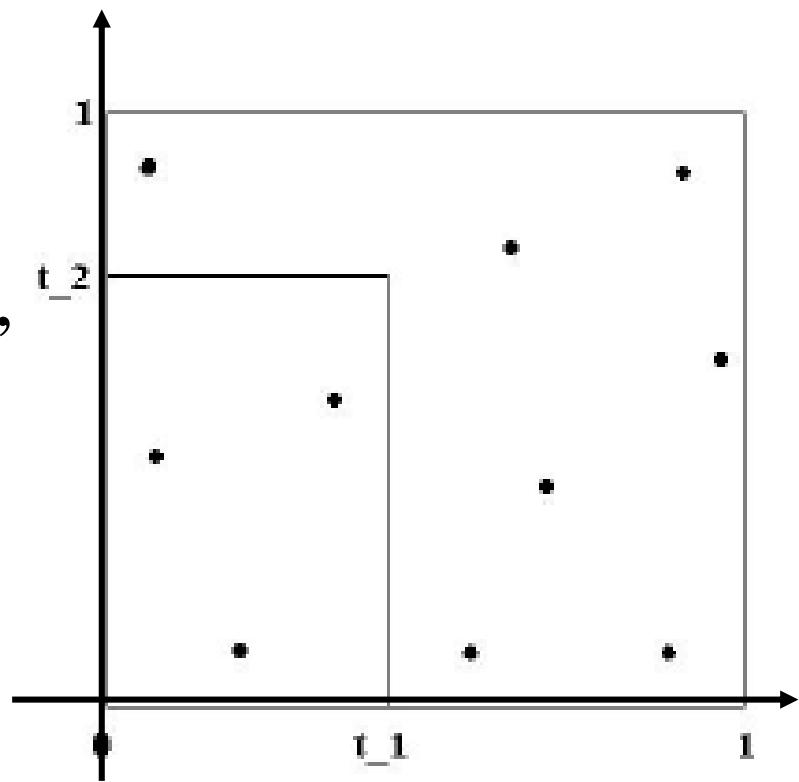
2. Measures of Uniformity: Discrepancy

Let $P = \{x_1, \dots, x_N\} \subset [0,1]^d$, $t = (t_1, \dots, t_d) \in [0,1]^d$.

Local discrepancy :

$$\Delta_P(t) = \frac{1}{N} \sum_{n=1}^N I_{[0,t]}(x_n) - \prod_{i=1}^d t_i,$$

$$\text{where } [0, t] = \prod_{i=1}^d [0, t_i].$$



Star-Discrepancy

Let P a point set consisting of $x_1, \dots, x_N \in [0,1]^d$

The **star-discrepancy** of the point set P is defined as

$$D_N^*(P) = \sup_{E \in J^*} \left| \frac{A(E; P)}{N} - m(E) \right|,$$

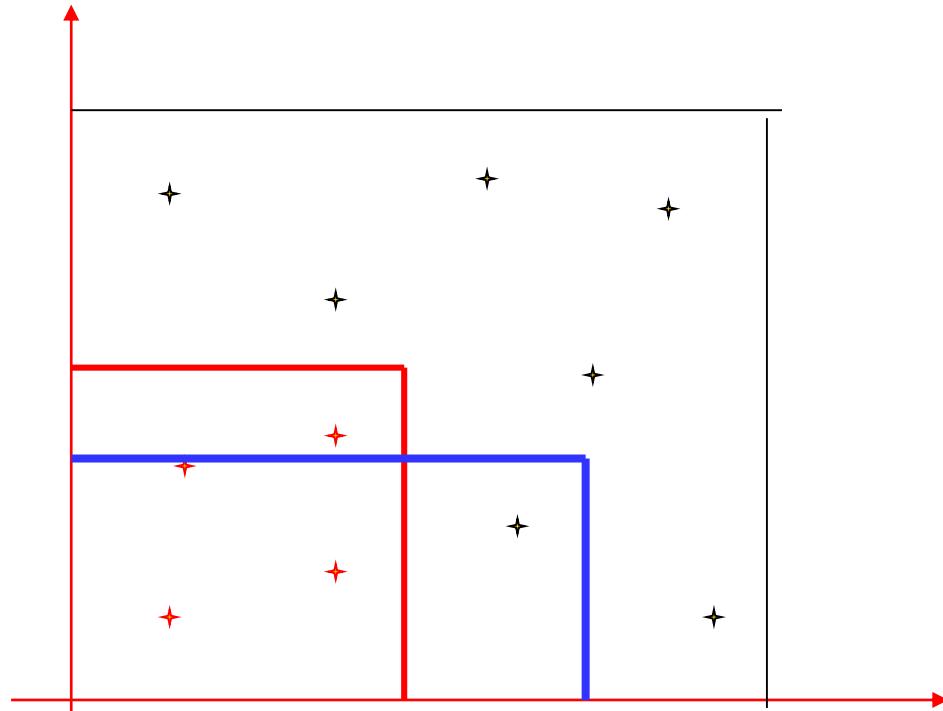
where

J^* is the family of all subintervals of $[0,1]^d$ of the form $\prod_{i=1}^d [0, u_i)$ (with a vertex at the origin),

$A(E, P)$ is the counting function that indicates the number of points of P falling in E ,

$m(E)$ is the measure of interval E : $m(E) = u_1 \dots u_d$.

Star-Discrepancy

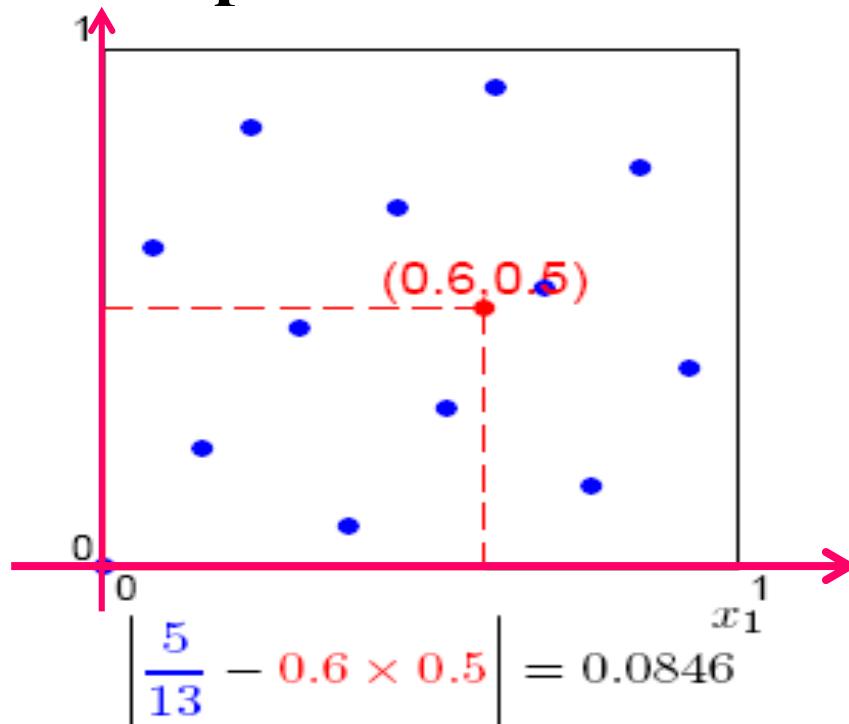


Star-discrepancy measures the deviation of the empirical distribution of a point set to the theoretical uniform distribution.

Star-Discrepancy

The star discrepancy is the **supremum**, over all boxes with a corner at the origin, of

$$\left| \frac{\text{number of points in the box}}{\text{total number of points}} - \text{volume of the box} \right|$$



Star-Discrepancy: an alternative interpretation

Let $F(x) = x^{(1)} \dots x^{(d)}$, $x = (x^{(1)} \dots x^{(d)})$;

$$\bar{F}(x) = \frac{1}{N} \sum_{n=1}^N I(x_n \leq x).$$

be the theoretical uniform distribution function and empirical distribution function of the points in P , respectively, then

$$D_N^*(P) = \sup_x |\bar{F}(x) - F(x)|.$$

Star-discrepancy is the Kolmogorov-Smirnov distance. It can be viewed as a quantitative measure for the deviation from the uniform distribution.

(Extreme)-Discrepancy

- The (extreme)-discrepancy of the point set P is defined as

$$D_N(P) = \sup_{E \in J} \left| \frac{A(E; P)}{N} - m(E) \right|,$$

where J is the family of all subintervals of $[0,1]^d$ of the form $\prod_{i=1}^d [u_i, v_i]$.

Discrepancies are very important notions in QMC.
They measure how well given point set is distributed.

Theorem: (relation between extreme and star discrepancies)

For any P consisting of N points in $[0,1]^d$, we have

$$D_N^*(P) \leq D_N(P) \leq 2^d D_N^*(P).$$

Proof. The left inequality is trivial.

For $d=1$, the right inequality is immediately obtained from

$$A([u,v]; P) = A([0,v]; P) - A([0,u]; P);$$

$$m([u,v]) = m([0,v]) - m([0,u]).$$

For $d > 1$ it is obtained from analogous identities.

Theorem: (relation between extreme and star discrepancies)

In dimension d , any subinterval of $[0,1]^d$ can be written as a composition of at most 2^d of subintervals of $[0,1]^d$ with one vertex at the origin.

For example, for $d=2$ and

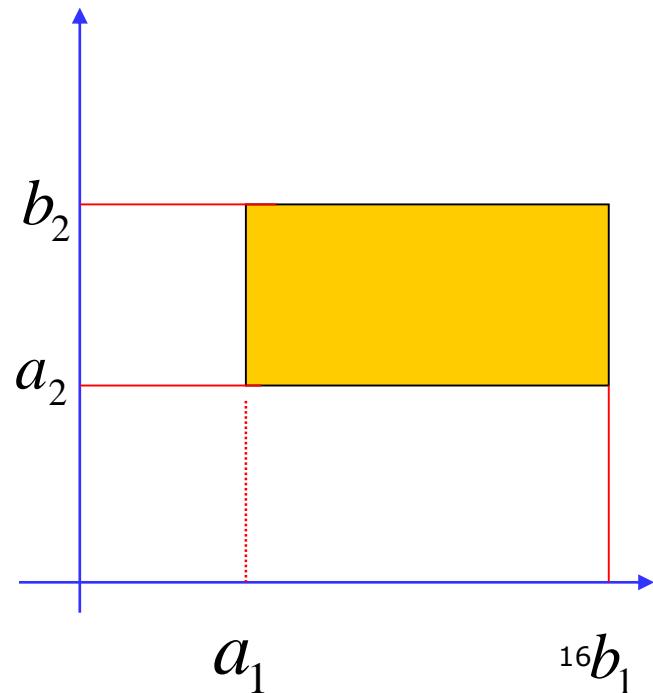
$a = (a_1, a_2), b = (b_1, b_2)$, then the

interval $[a,b) = [a_1, b_1) \times [a_2, b_2)$

is a combination of

$[0, b_1) \times [0, b_2), [0, a_1) \times [0, b_2),$

$[0, b_1) \times [0, a_2), [0, a_1) \times [0, a_2).$



2. Measures of Uniformity: Discrepancy

The smaller the discrepancy is, the better.

How small can be the extreme discrepancy?

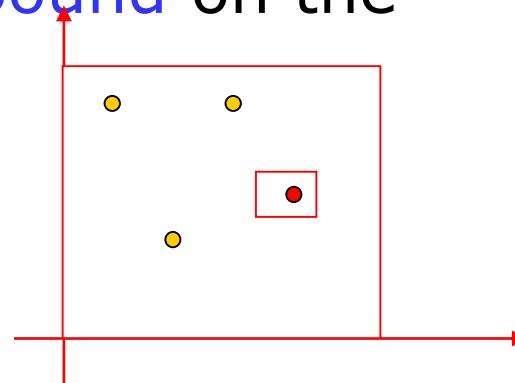
--- **It cannot be arbitrarily small !**

Let P be a point set of N points in $[0,1]^d$. Consider a subinterval J which contains exactly one point of P .

We can choose the subinterval with this property but with the volume $m(J)$ arbitrarily small.

Thus we have the **first lower bound** on the extreme discrepancy

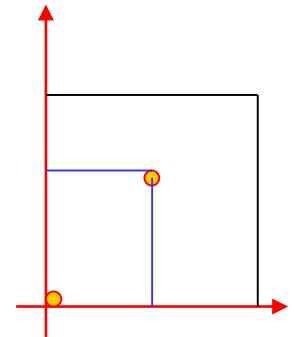
$$D_N(P) \geq 1/N.$$



L₂-Discrepancy

- Let $J(x)$ denote rectangle with one corner at 0 and another corner at x .
- L₂-Star-Discrepancy**

$$T_N^*(P) = \left(\int_{[0,1]^d} \left| \frac{A(J(x); P)}{N} - m(J(x)) \right|^2 dx \right)^{\frac{1}{2}},$$



- L₂-Discrepancy:**

$$T_N(P) = \left(\int_{[0,1]^{2d}} \left| \frac{A(J(x, y); P)}{N} - m(J(x, y)) \right|^2 dxdy \right)^{\frac{1}{2}},$$

where $J(x, y)$ denotes the rectangle with opposite corners at (x, y) .

2. Measures of Uniformity: Discrepancy

Clearly,

$$T_N^*(P) \leq D_N^*(P),$$

since L_∞ – norms are larger than L_2 – norms.

L_q -discrepancy is also possible, by replacing L_2 -norm by L_q -norm:

$$T_N^*(P) = \left(\int_{[0,1]^d} \left| \frac{A(J(x); P)}{N} - m(J(x)) \right|^q dx \right)^{\frac{1}{q}}, \quad 1 \leq q < \infty.$$

Formula to calculate L₂ discrepancy

- By direct integration, we have

$$\begin{aligned} \left(T_N^*(P)\right)^2 &= \int_{[0,1]^d} \left| \frac{A(J(x); P)}{N} - m(J(x)) \right|^2 dx \\ &= \int_{[0,1]^d} \left| \frac{1}{N} \sum_{i=1}^N I_{[0,x)}(t_i) - x_1 \dots x_d \right|^2 dx \\ &= \frac{1}{3^d} - \frac{1}{2^{d-1}} \frac{1}{N} \sum_{i=1}^N \prod_{k=1}^d (1 - t_{i,k}^2) + \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \prod_{k=1}^d [1 - \max(t_{i,k}, t_{j,k})], \end{aligned}$$

where $t_i = (t_{i,1}, \dots, t_{i,d})$, $P = \{t_1, \dots, t_n\}$.

- But there is no explicit formula for L_∞ -discrepancy.

2. Measures of Uniformity: Discrepancy

The **expected** L₂-star-discrepancy of random points:

$$E(T_N^*(P))^2 = \frac{1}{N} (2^{-d} - 3^{-d}).$$

This formula can also be obtained by using the theory of reproducing kernel Hilbert spaces (next chapter).

2. Measures of Uniformity: Discrepancy

□ Questions:

- For a given N , what is the **smallest discrepancy**?
That is, **how small** can be the discrepancy with a suitable choice of the points?
- **How to find** a point set (or sequence) with the smallest discrepancy?

One-dimensional case

Theorem:

If $0 \leq x_1 \leq x_2 \leq \dots \leq x_N \leq 1$, then

$$D_N^*(x_1, \dots, x_N) = \frac{1}{2N} + \max_{1 \leq n \leq N} \left| x_n - \frac{2n-1}{2N} \right|$$

Note: The theorem implies that we always have

$$D_N^*(x_1, \dots, x_N) \geq \frac{1}{2N}$$

and equality holds if $x_n = (2n-1)/(2N)$.

Proof. Put $x_0 = 0, x_{N+1} = 1, P = \{x_1, \dots, x_N\}$. We have

Next,

$$D_N^*(P) = \max_{0 \leq n \leq N} \sup_{x_n < u \leq x_{n+1}} \left| \frac{A([0, u); P)}{N} - u \right|$$

$$= \max_{0 \leq n \leq N} \sup_{x_n < u \leq x_{n+1}} \left| \frac{n}{N} - u \right|$$

$$= \max_{0 \leq n \leq N} \max \left(\left| \frac{n}{N} - x_n \right|, \left| \frac{n}{N} - x_{n+1} \right| \right)$$

$$= \max_{1 \leq n \leq N} \max \left(\left| \frac{n}{N} - x_n \right|, \left| \frac{n-1}{N} - x_n \right| \right) \quad (**)$$

$$= \max_{1 \leq n \leq N} \max \left(\left| x_n - \frac{2n-1}{2N} - \frac{1}{2N} \right|, \left| x_n - \frac{2n-1}{2N} + \frac{1}{2N} \right| \right)$$

$$= \frac{1}{2N} + \max_{1 \leq n \leq N} \left| x_n - \frac{2n-1}{2N} \right|.$$

Lower Bound for Discrepancy

- It has been **proved** that for **any N-element** point set P in dimension d, we have

$$D_N^*(P) \geq B_d \frac{(\log N)^{(d-1)/2}}{N}.$$

Theorem (Roth 1954). For every dimension d,
there exists a quantity $B_d > 0$ with the property :
For every N-element point set in $[0,1]^d$,

$$D_N(P) \geq D_N^*(P) \geq T_{2,N}(P) \geq B_d \frac{(\log N)^{(d-1)/2}}{N}.$$

Lower Bound for Discrepancy

- It is known that Roth's lower bound is **best** possible for the **L₂ discrepancy**:

For every d , $N \geq 2$, there exists an N -element point set P in $[0,1)^d$ such that

$$T_{2,N}(P) \leq B_d^* \frac{(\log N)^{(d-1)/2}}{N}.$$

2002

- On the other hand, Roth's lower bound is **not best** possible for the **star-discrepancy**.

Lower Bound for Discrepancy

□ Conjecture: (for star-discrepancy)

For any N -element point set P in dimension d , it is widely believed that

$$D_N^*(P) \geq C_d \frac{(\log N)^{(d-1)}}{N}.$$

This is true for $d=1, 2$, but it is still open for $d > 2$.

Theorem (Schmidt). *There exists a constant $c > 0$ such that for the star discrepancy of any N -element point set \mathcal{P} in $[0, 1]^2$ we have*

$$D_N^*(\mathcal{P}) \geq c \frac{\log N}{N}.$$

1972

The exact determination of the sharp lower bound on the star discrepancy seems to be a very difficult problem.

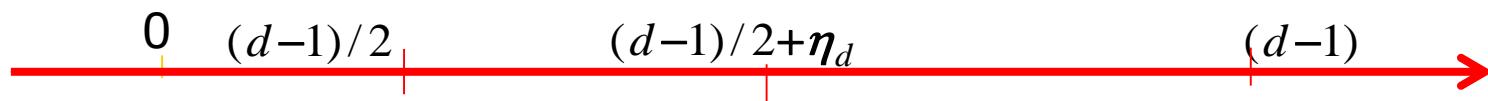
Lower Bound for Discrepancy

The present best lower estimate for star-discrepancy of finite point sets in dimension $d > 2$:

$$D_N^*(P) \geq C_d \frac{(\log N)^{(d-1)/2 + \eta_d}}{N}, \quad \eta_d \in (0, 1/2).$$

2008

Note:



Best for L_2

present best for star

Best for star ($d = 1, 2$)

$(d \geq 3)$

(conjecture : for all d)

Low discrepancy point set or sequence

- There are **point sets** satisfying

$$D_N^*(P) \leq C_d \frac{(\log N)^{d-1}}{N}.$$

- There are **sequences**, the first N terms of which satisfy

$$D_N^*(P) \leq C_d \frac{(\log N)^d}{N}.$$

Low discrepancy point set or sequence

- **Low discrepancy point set:**

If a point set has discrepancy satisfying

$$D_N^*(P) \leq C_d \frac{(\log N)^{d-1}}{N}.$$

then it is called a **low discrepancy point set**.

- **Low discrepancy sequence:**

If for **an infinite sequence** of points, the point set consisting of the first N terms satisfies

$$D_N^*(P) \leq C_d \frac{(\log N)^d}{N}.$$

then it is called a **low discrepancy sequence**.

Remark

- It is possible to get

$$D_N^*(P) \leq C_d \frac{(\log N)^{d-1}}{N}.$$

The gap subject to continued work.

$$D_N^*(P) \geq B_d \frac{(\log N)^{(d-1)/2}}{N}.$$

- **Theorem (Roth 1954).** For every dimension d , there exists a quantity $B_d > 0$ with the property :
For every N -element point set in $[0,1]^d$,

$$D_N(P) \geq D_N^*(P) \geq T_{2,N}(P) \geq B_d \frac{(\log N)^{(d-1)/2}}{N}.$$

2. Measures of Uniformity: Discrepancy

If P is a set of random points, then (with probability 1)

$$D_N^*(P) \leq C_d \sqrt{\frac{\log \log N}{N}}.$$

Let m be an integer and $N = m^d$. Define a set (regular centered lattice)

$$P = \left\{ \left(\frac{2l_1 - 1}{2m}, \dots, \frac{2l_d - 1}{2m} \right), 1 \leq l_i \leq m, i = 1, \dots, m \right\}.$$

We can prove that for this set

$$C_1(d)N^{-1/d} \leq D_N^*(P) \leq C_2(d)N^{-1/d}.$$

For $d > 1$, it turns out that this is not a good choice.

2. Measures of Uniformity: Discrepancy

Let m be an integer and $N = m^d$. Define

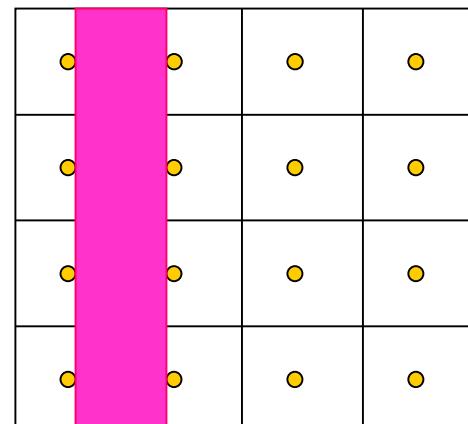
$$P = \left\{ \left(\frac{2l_1 - 1}{2m}, \dots, \frac{2l_d - 1}{2m} \right), 1 \leq l_i \leq m, i = 1, \dots, m \right\}.$$

Then

$$C_1(d)N^{-1/d} \leq D_N^*(P) \leq C_2(d)N^{-1/d}.$$

The volume of the coloured part is

$$\frac{1}{m} = N^{-1/d}.$$



3. The Error Bound

- MC has the CLT.
- QMC has the Koksma-Hlawka inequality.

3. The Error Bound

Theorem (one-dimensional case):

If f has bounded variation $V(f)$ on $[0,1]$, then for any $x_1, \dots, x_N \in [0,1]$, we have

$$\left| \frac{1}{N} \sum_{i=1}^N f(x_i) - \int_0^1 f(x) dx \right| \leq V(f) D_N^*(\{x_i\}).$$

Note: Variation $V(f) = \int_0^1 |f'(x)| dx = \int_0^1 |df|$.

Proof.

We can assume that $x_1 \leq x_2 \leq \dots \leq x_N$. Put $x_0 = 0, x_{N+1} = 1$. Using **integration by parts**, we obtain

$$\begin{aligned}\frac{1}{N} \sum_{i=1}^N f(x_i) - \int_0^1 f(x) dx &= -\sum_{i=1}^N \frac{i}{N} [f(x_{i+1}) - f(x_i)] + \int_0^1 x df(x) \\ &= \sum_{i=0}^N \int_{x_i}^{x_{i+1}} \left(x - \frac{i}{N} \right) df(x).\end{aligned}$$

By theorem 2.6 (Niederreiter, p.15) or (**) of p.26,

$$\left| x - \frac{i}{N} \right| \leq D_N^*(\{x_i\}).$$

The desired inequality follows immediately.

Alternative Proof:

Note that $f(x) = f(1) - \int_x^1 f'(y)dy$, we have

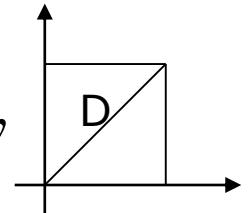
$$\int_0^1 f(x)dx - \frac{1}{N} \sum_{i=1}^N f(x_i)$$

$$= \frac{1}{N} \sum_{i=1}^N \int_{x_i}^1 f'(y)dy - \int_0^1 \int_x^1 f'(y)dy dx$$

$$= \int_0^1 \frac{1}{N} \sum_{i=1}^N J_{(x_i, 1]}(y) f'(y) dy - \int_0^1 \int_0^y f'(y) dx dy$$

$$= \int_0^1 \left[\frac{1}{N} \sum_{i=1}^N J_{(x_i, 1]}(y) - y \right] f'(y) dy$$

$$= \int_0^1 \left[\frac{1}{N} \sum_{i=1}^N J_{(0, y]}(x_i) - y \right] f'(y) dy$$



Local discrepancy

Alternative Proof (cont.)

$$\begin{aligned} \Rightarrow |\text{Error}| &\leq \int_0^1 \left| \frac{1}{N} \sum_{i=1}^N J_{(0,y]}(x_i) - y \right| |f'(y)| dy \\ &\leq \left(\int_0^1 \left| \frac{1}{N} \sum_{i=1}^N J_{(0,y]}(x_i) - y \right|^p dy \right)^{1/p} \left(\int_0^1 |f'(y)|^q dy \right)^{1/q}, \end{aligned}$$

for $p, q \geq 1$, $\frac{1}{p} + \frac{1}{q} = 1$.

The case $q = 1, p = \infty$ is the K - H inequality .

High-Dimensional Case

- We may develop a similar theory for **multivariate functions**. This can be done in a very elegant way by using the notion of **reproducing kernel Hilbert space** (next chapter).
- The **Koksma-Hlawka inequality** is the fundamental error estimate for QMC rules which **separates the influence of the integrand** and **of the underlying integration nodes** on the integration error.
- It was proved by Koksma in 1942 for dimension one and generalized by Hlawka to arbitrary dimension in 1961.

High-Dimensional Case

The variation in the sense of Hardy and Krause:

$$V(f) = \sum_{k=1}^d \sum_{1 \leq i_1 \leq \dots \leq i_k} V^{(k)}(f; i_1, \dots, i_k);$$

$$V^{(k)}(f; i_1, \dots, i_k) = \int_0^1 \dots \int_0^1 \left| \frac{\partial^k f}{\partial x_{i_1} \dots \partial x_{i_k}} \right|_{x_j=1, j \neq i_1, \dots, i_k} dx_{i_1} \dots dx_{i_k}$$

The restriction of differentiability is stronger than necessary and may be relaxed to the standard notion of bounded variation (we do not give its definition here).

The Koksma-Hlawka Inequality

Theorem:

If f has bounded variation in the sense of Hardy and Krause, then for any $x_1, \dots, x_N \in [0,1]^d$, we have

$$\left| \frac{1}{N} \sum_{i=1}^N f(x_i) - \int_{[0,1]^d} f(x) dx \right| \leq V(f) D_N^*(\{x_i\}).$$

The Koksma-Hlawka Inequality

$$\left| \frac{1}{N} \sum_{i=1}^N f(x_i) - \int_{[0,1]^d} f(x) dx \right| \leq V(f) D_N^*(\{x_i\}).$$

Remark:

The upper bound depends on **variation and discrepancy**.

- The variation $V(f)$ depends only on the integrand,
- The discrepancy $D_N^*(\{x_i\})$ depends on the point set.
- The inequality separates the effects of the function and the point set on the integration error.

Remark:

Assume that

$$\frac{\partial^d f}{\partial x_1 \dots \partial x_d}$$

exists and all partial derivatives $\frac{\partial^k f}{\partial x_{i_1} \dots \partial x_{i_k}}$ are bounded by L, then

$$V(f) \leq 2^d L$$

and

Depends exponentially on the dimension

$$\left| \frac{1}{N} \sum_{i=1}^N f(x_i) - \int_{[0,1]^d} f(x) dx \right| \leq 2^d L D_N^*(\{x_i\}).$$

3. The Error Bound

It is clear from [Koksma-Hlawka Inequality](#), to make the QMC integration error small, we should [choose point set with star discrepancy as small as possible](#).

If we use QMC algorithm based on a low discrepancy sequence, then QMC integration convergence order is

$$O(N^{-1}(\log N)^d).$$

Thus the [Koksma-Hlawka Inequality gives the](#) possible theoretical convergence of the QMC integration error.

3. The Error Bound

Koksma-Hlawka inequality is **sharp** in the following sense.

Theorem: (Niederreiter, 1992)

For any $x_1, \dots, x_N \in [0,1]^d$ and any $\varepsilon > 0$, there exists a function $f \in C^\infty([0,1]^d)$ with $V(f) = 1$ and

$$\left| \frac{1}{N} \sum_{i=1}^N f(x_i) - \int_{[0,1]^d} f(x) dx \right| > D_N^*(\{x_i\}) - \varepsilon.$$

So Koksma-Hlawka is tight.
Koksma-Hlawka is also very loose.

Remarks

- The Koksma-Hlawka error bound is very **loose** (**conservative**).
- Both the variation and the discrepancy are **very difficult to compute**.
- Thus the Koksma-Hlawka error bound **has little practical importance** on error estimation.
(It only has theoretical importance.)
- For QMC error estimation, we need to use **Randomized QMC** (to be discussed later).

Remarks

- For functions with **infinite variations**, $V(f)=\infty$, the Koksma-Hlawka error bound will not provide us with a useful bound of the integration error.
- This is a **common situation in finance** since in many instances we have problems cast in an unbounded domain (say, a European call option).
- Still, the numerical results support the claim that the method is **useful** for the considered problems.
- Theory is expected to be established for $V(f)=\infty$.

MC

vs

QMC

Use random points

Randomness, independence

Uniformity

Convergence $O(N^{-1/2})$

Probabilistic error bound

Does not depend on dimension

Error bound is useful in practice

Based on Law of Large Numbers

And on Central Limit Theory

Use low discrepancy points

Deterministic, correlated

Greater uniformity

$O(N^{-1}(\log N)^d)$

Deterministic error bound

Depend on dimension

Error bound is not useful in practice

Based on Number theory

MC vs QMC

- **On the one hand**, from the mathematical point of view, for any fixed dimension d , QMC converges **asymptotically** faster than MC:

$$O(N^{-1/2}) \quad \text{vs} \quad O(N^{-1} \log^d N)$$

In low dimensions QMC is clearly better than MC.

- **On the other hand**, for a large d , the factor $(\log N)^d$ is very essential. In order to have

$$N^{-1}(\log N)^d < N^{-1/2}, \quad (*)$$

N should be extremely large. For example, for $d = 8$, we need $N > 1.79 \times 10^{29}$ such that $(*)$ is satisfied.

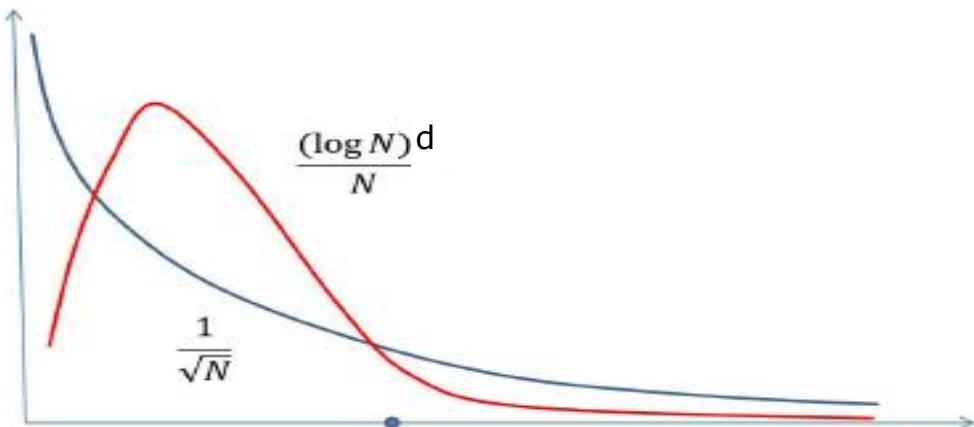
MC vs QMC: What about those logs?

The function $N \mapsto (\log N)^d / N$ does not start to decrease to zero until $N \geq \exp(d)$. For $N \leq \exp(d)$ this function is increasing, which means that for cardinality N in this range the discrepancy bounds are useless.

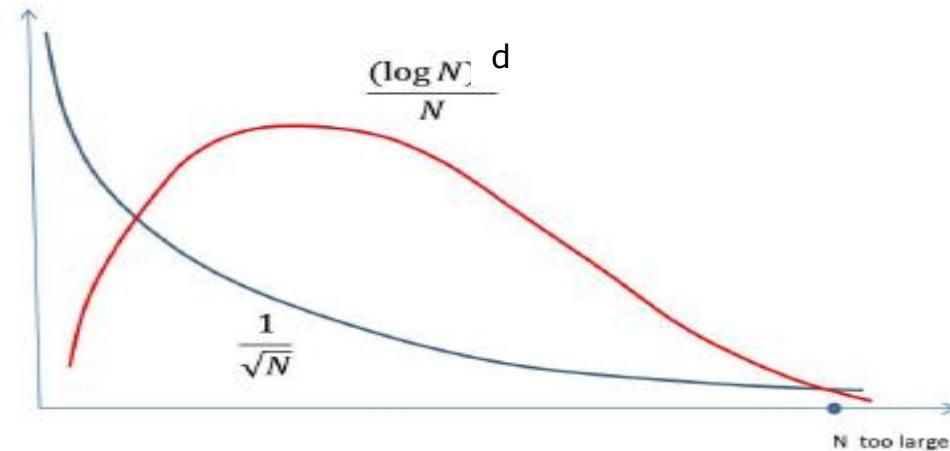
Point sets with cardinality $N \geq \exp(d)$ cannot be used for practical applications. The QMC error bound is useful only if N is large compared to the dimension.

MC vs QMC

For d given, not **too** large:



For d given, large:



MC vs QMC

- The classical theory (dependence of the discrepancy bounds on dimension) led to the **belief** that QMC can only be applied in very moderate dimensions, and could not be expected to work well for high dimension.
- Studies and comparisons (on some test problems and on discrepancy) **before 1994** showed that in dimension d larger than 12, QMC is not better than MC. Thus experts in QMC suggested then that QMC should not be used in high dimension.

MC vs QMC

- **Nevertheless, and surprisingly,** in 1995, people found that even in 360-dimensional financial problems (CMO problem), QMC is much more efficient than MC and the convergence of QMC is much more smoother than that of MC (Paskov & Traub) .
- Paskov & Traub (1995). Fast valuation of financial derivatives, Journal of Portfolio Management.

3. The Error Bound

- QMC rules often do very well and even work much better than we have any right to expect.
- The question of exactly ***why QMC rules should give good results*** is still not completely resolved.
- In 2010, at the MCQMC meeting in Warsaw, **Sloan** spoke in this context about
“The unreasonable effectiveness of QMC”.
- **It is challenging to explain the success of QMC for very high dimensional problems.**
This is nowadays an important stream of research.

3. The Error Bound

- Why is QMC more efficient than MC for many high-dim financial problems?

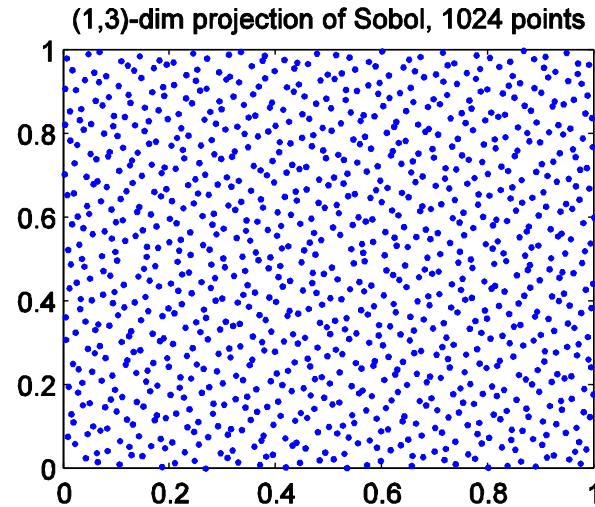
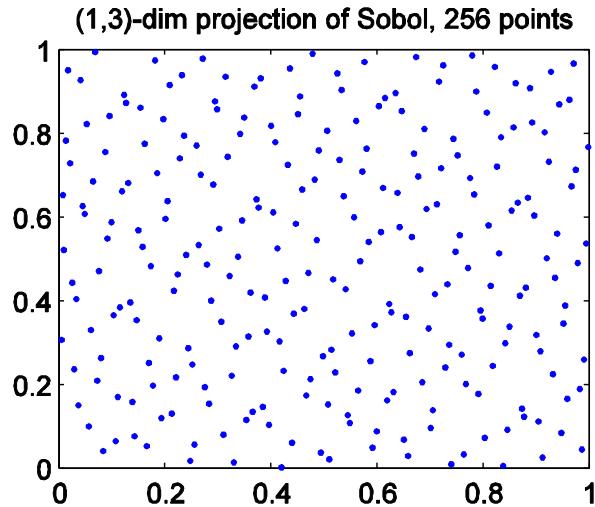
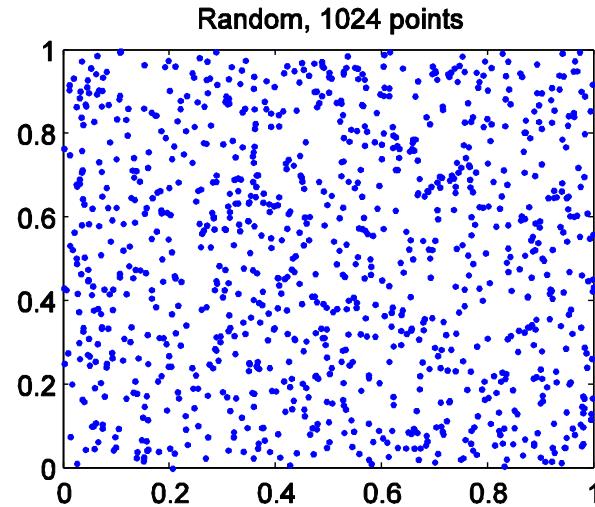
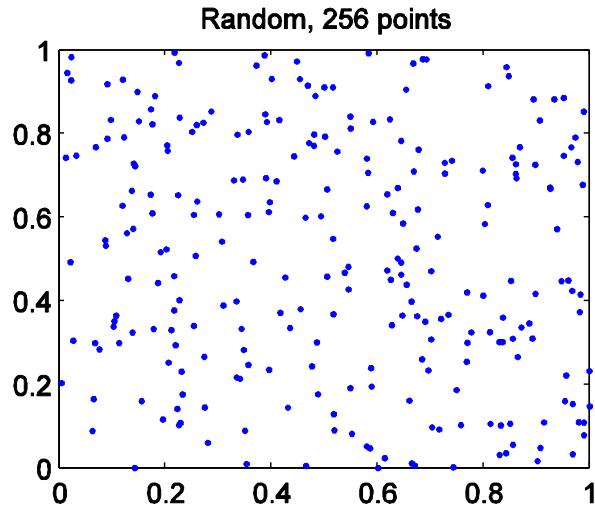
- Theory of weighted function spaces and tractability

Some coordinate directions are more important than others. The weights model the behavior of different coordinate directions and under suitable conditions one can break the curse of dimensionality.

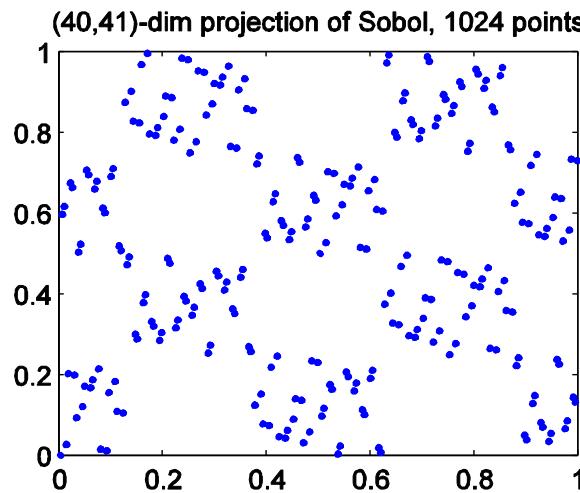
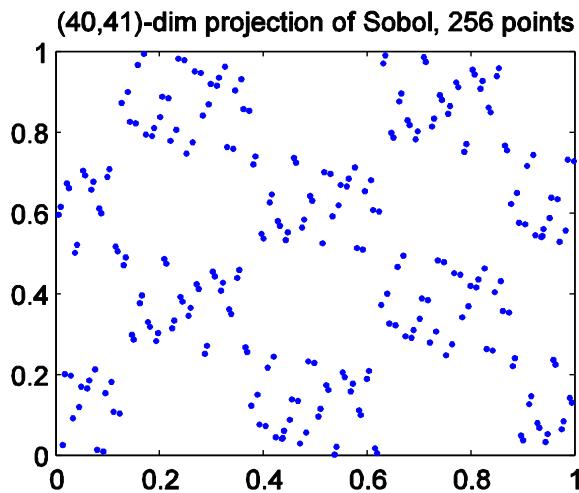
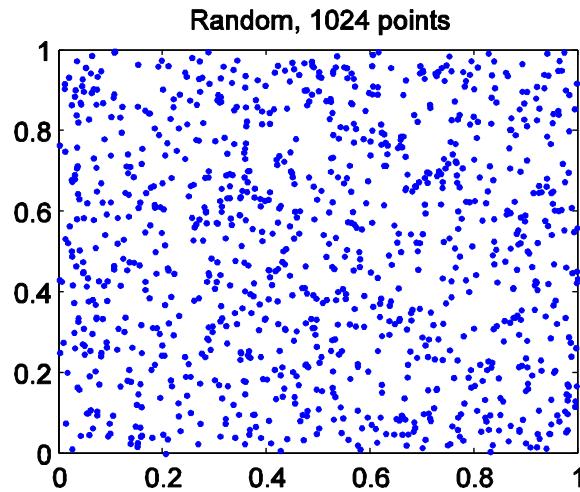
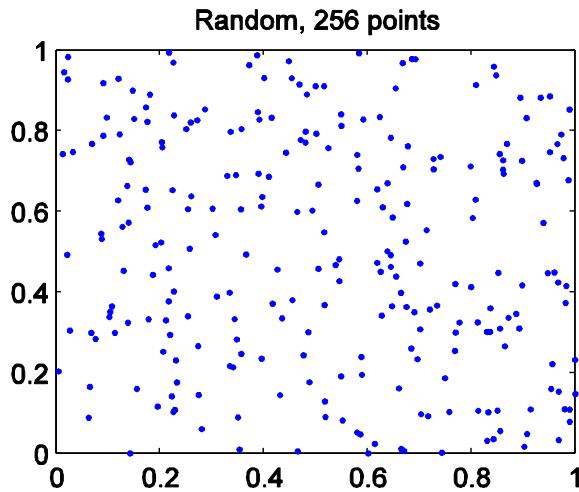
(Sloan & Wozniakowski 1998, Wang 2003).

- ANOVA and Effective dimension.
(Caflisch et al. 1997, Wang and Fang 2003)
- It is still a very active part of research.

Random vs QMC Points



QMC: High Dimensional Limitations



QMC in High Dimensions

- Typical low discrepancy sequences are better distributed in their **initial dimensions**.
- Their projections onto **later dimensions** and projections of high orders can exhibit **poor quality**.
- To improve QMC, there is strong motivation to reduce the **effective dimension** of the problem:
concentrating the variance on the first few dimensions, such that the better quality of the initial dimensions can be fully exploited
(-- next chapter).

4. (t,m,d) -net and (t,d) -sequence

□ Motivation

Goal: As many as possible subsets E contain the **correct portion** of points, i.e.,

$$\frac{A(E; P)}{N} = m(E), \quad (*)$$

If $(*)$ is satisfied, then we say that the point set P is fair for the subset.

From the lower bound for discrepancy, it is never possible that P is fair for all subintervals are fair, since there always exists an subinterval such that

$$\left| \frac{A(E, P)}{N} - m(E) \right| \geq B_d \frac{(\log N)^{(d-1)/2}}{N}. \quad \text{Lower bound}$$

4. (t,m,d) -net and (t,d) -sequence

- We want to construct point sets which are **fair** for as large as possible class of subintervals.
 - The class should not be too large (would lead to too restrictive demand).
 - The class should not be too small.

For example, consider a regular lattice:

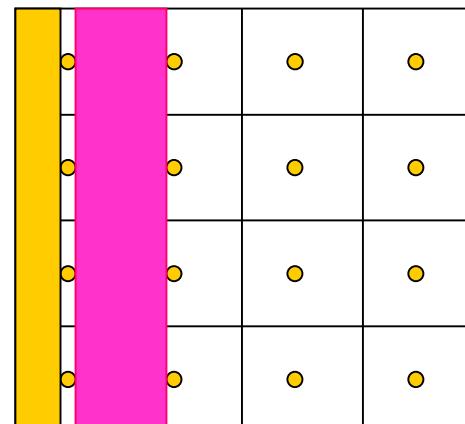
The 16 subintervals are fair.

But the interval

$[0,1/8] \times [0,1]$ or

$[1/8,3/8] \times [0,1]$

is not fair.



Elementary interval in base b: (b-ary box)

We fix dimension d and base b, a subinterval E of $[0,1)^d$ of the form

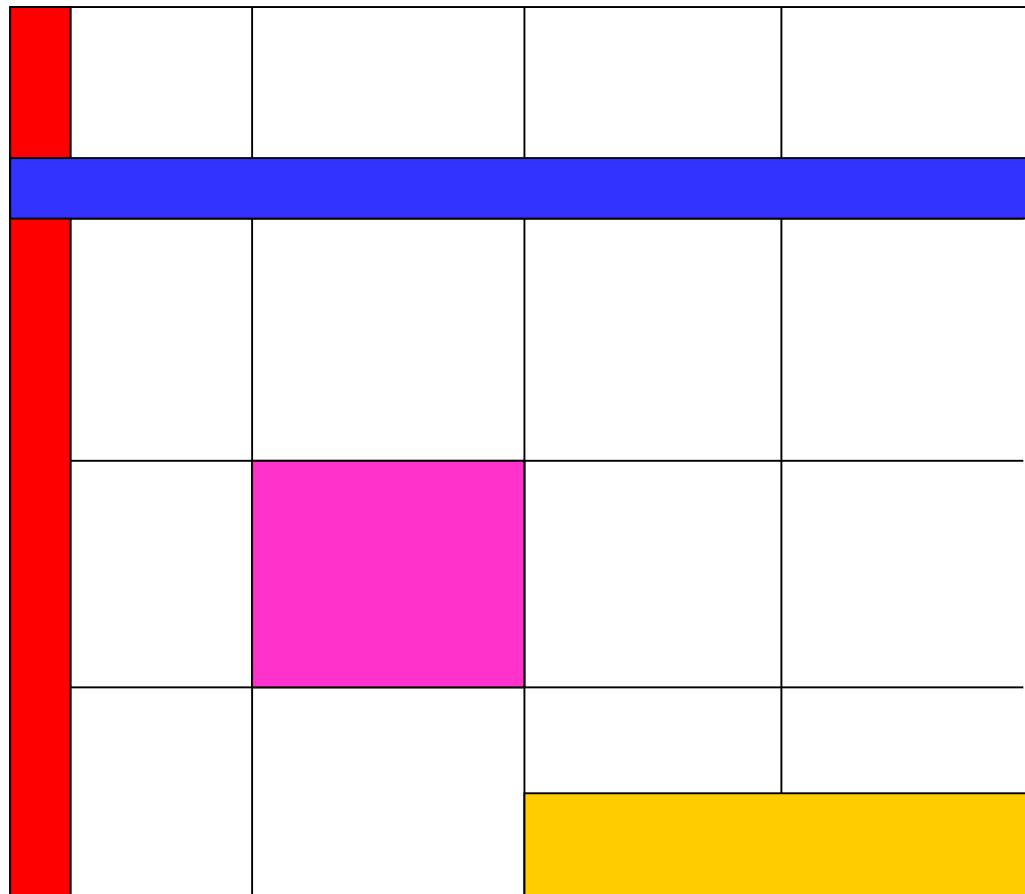
$$E = \prod_{i=1}^d \left[\frac{a_i}{b^{k_i}}, \frac{a_i + 1}{b^{k_i}} \right), \quad k_i \geq 0, \quad 0 \leq a_i \leq b^{k_i} - 1,$$

is called an **elementary interval in base b of order**
 $K = k_1 + \dots + k_d$.

The volume of **elementary interval E in base b is**

$$m(E) = 1/b^{k_1 + \dots + k_d} = b^{-K}.$$

Elementary intervals in base 2 of order 4 (area 1/16)



Definition: (t, m, d) -net

$(0, m, d)$ -net

For $t = 0$, a $(0, m, d)$ - net in base b is a point set P of b^m points in $[0,1)^d$ such that exactly one point falls in every elementary interval of volume $1/b^m$

(the point set P is fair ~~v...l...~~ ^{for} every elementary interval of volume $1/b^m$.)

Definition: (t, m, d) -net

Let $0 \leq t \leq m$ be integer.

A (t, m, d) - net in base b is a point set P of b^m points in $[0,1)^d$ such that exactly b^t points fall in every elementary interval of volume b^t / b^m .

(The point set P is fair for every elementary interval of order $m-t$).

The Net Property

- The fraction of points that lie in every **elementary interval** equals the volume of the interval.
- But this may not hold for **arbitrary geometric domain** of the same volume.
- A disadvantage of nets in base b is that the number of points is restricted to **a power of b** .
- The **parameter t** is called the **quality parameter** of the net.

The quality parameter t

Smaller values of t mean greater uniformity:

With smaller t, even small elementary interval contains the right number of points.

- With $t=0$, each elementary interval with volume contains exactly $b^0 = 1$ point
(Is this always possible? No! This depends on b and d).
- With $t=m$, the elementary interval with volume $b^t / b^m = 1$ (the interval $[0,1]^d$ itself) contains exactly b^m points. All point set in $[0,1]^d$ satisfies this.
All point set in $[0,1]^d$ is a (m, m, d) -net.

The quality parameter t

- Every (t, m, d) -net with $t < m$ is a $(t+1, m, d)$ -net.
- A point set of b^m points never can be fair for an interval with volume less than $1/b^m$.

Example:

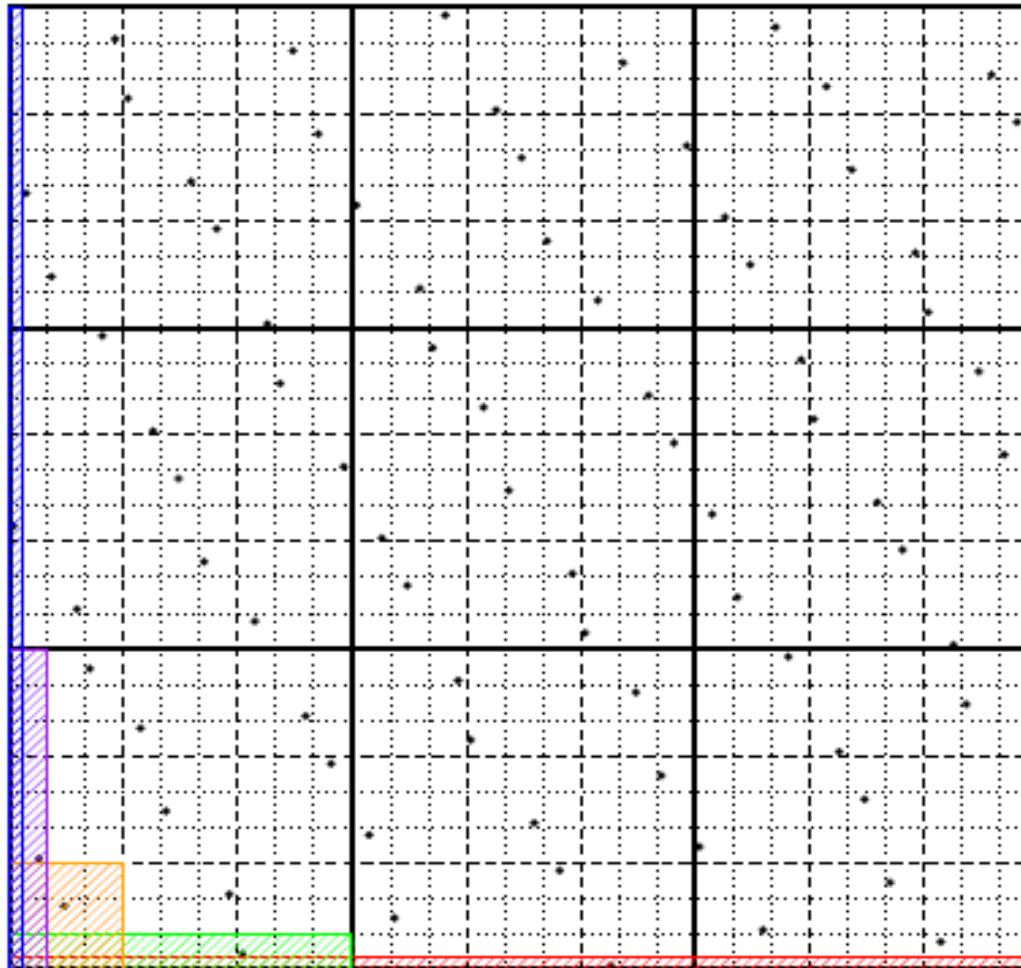
- ◆ **(0, 4, 2)-net in base 3: t=0, m=4, d=2, b=3**
- Total number of points: $b^m = 3^4 = 81$.
- Each elementary interval of volume $b^{(-m)} = 1/81$ contains exactly $b^0 = 1$ point (since $t=0$).
- The form of such elementary interval is

$$E = \prod_{i=1}^2 \left[\frac{a_i}{3^{d_i}}, \frac{a_i + 1}{3^{d_i}} \right),$$

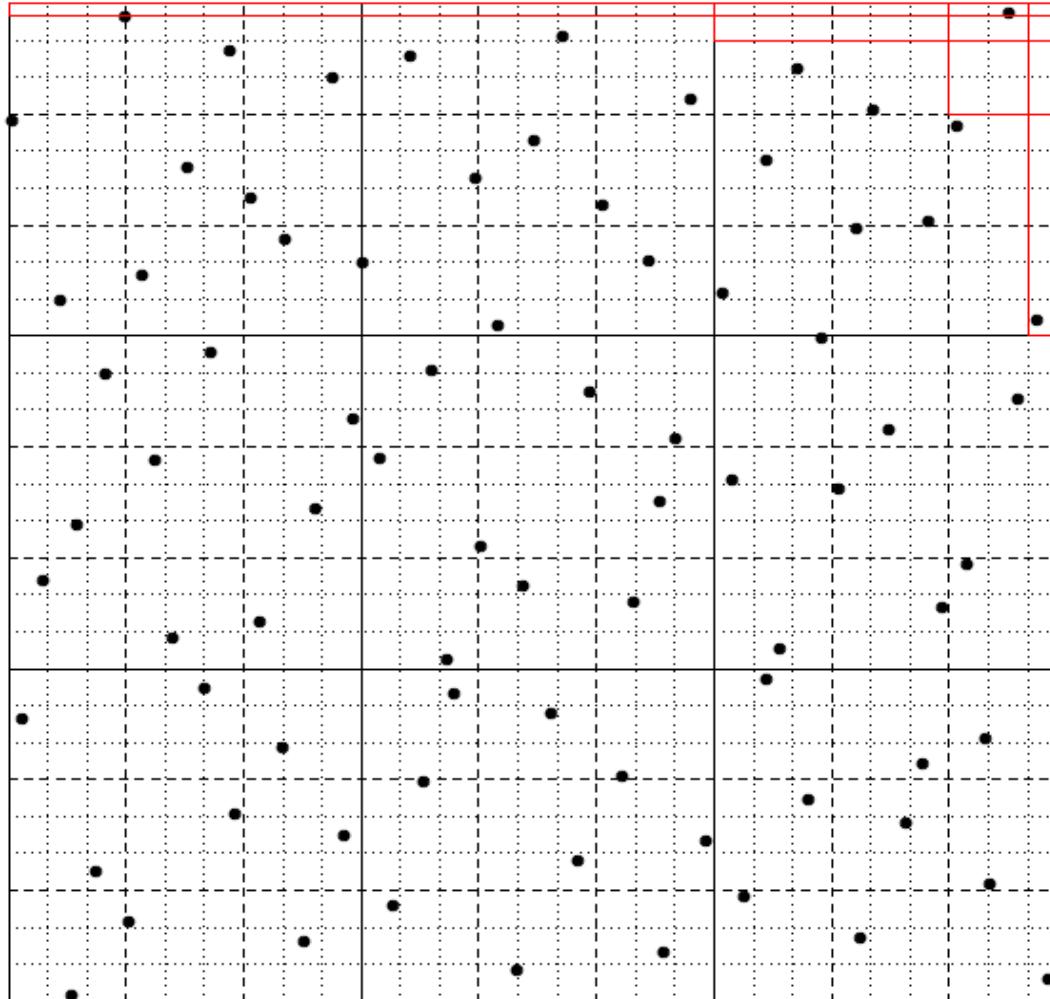
$$d_i \geq 0, d_1 + d_2 = 4, 0 \leq a_i \leq b^{d_i} - 1.$$

- Possible values of (d_1, d_2) with $d_1 + d_2 = 4$ are $(0,4), (1,3), (2,2), (3,1), (4,0)$

$(0,4,2)$ -net in base 3



(0,4,2)-net in base 3



Example: (0,4,2)-net in base 2 of 2^4 points

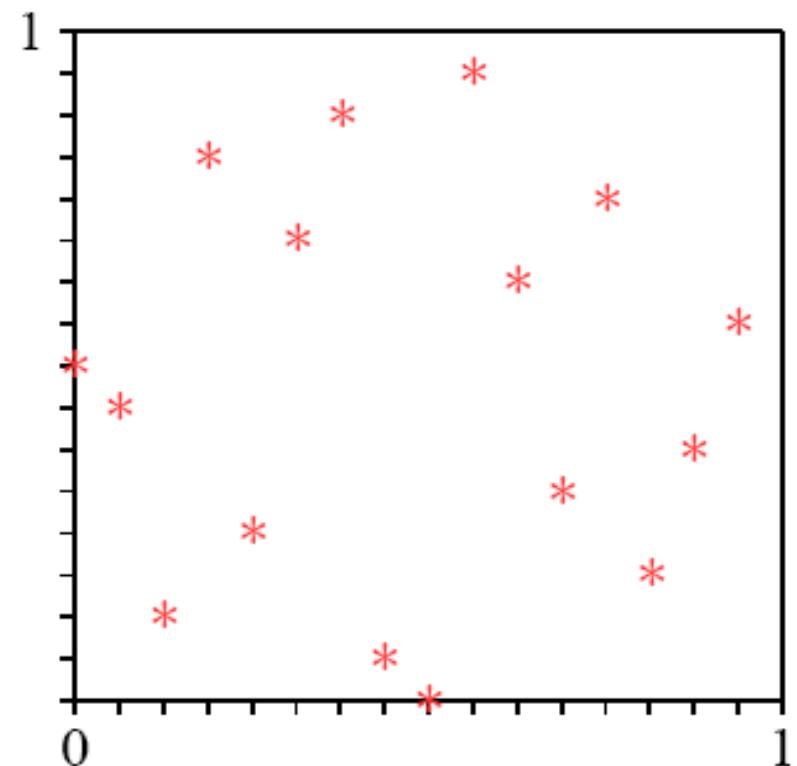
Each elementary interval of volume $1/16$ contains exactly one point (i.e., the point set is fair for each elementary interval of order 4).

$$E = \prod_{i=1}^2 \left[\frac{a_i}{2^{d_i}}, \frac{a_i + 1}{2^{d_i}} \right),$$

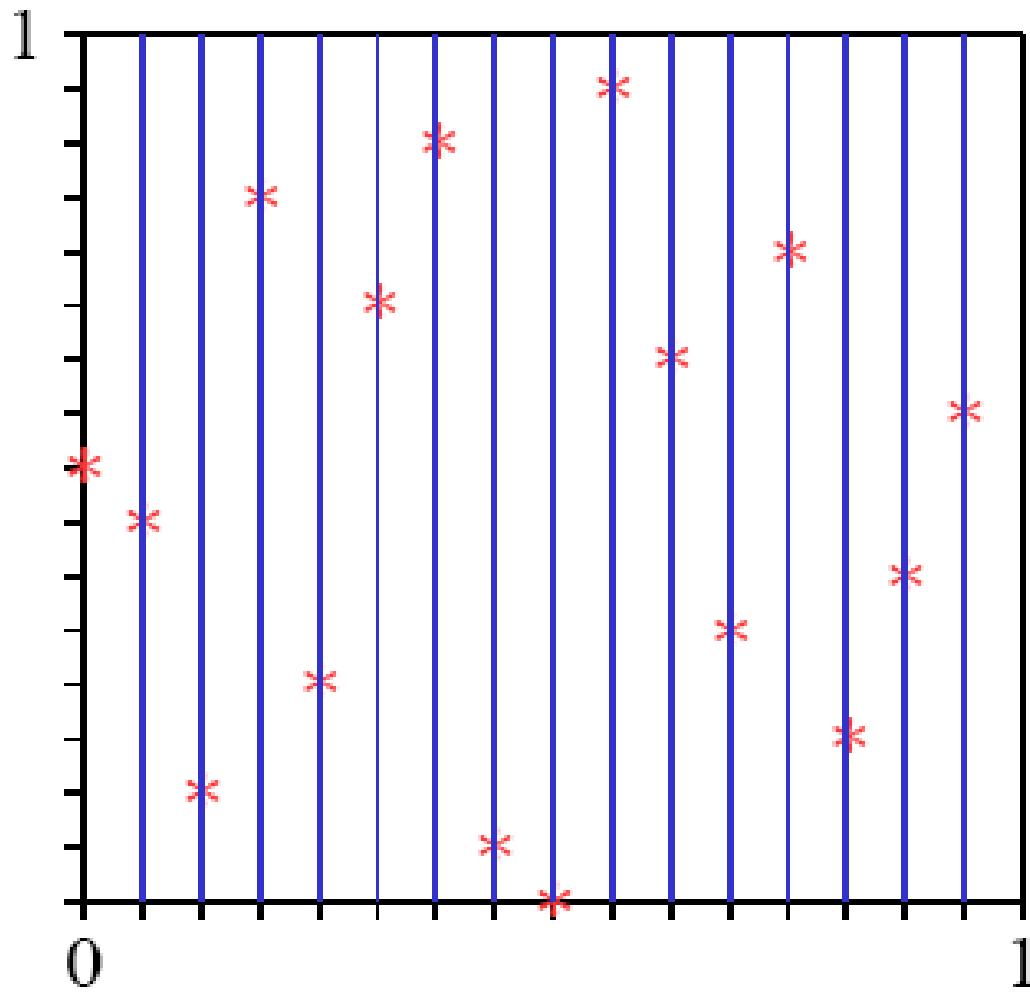
$$d_i \geq 0, d_1 + d_2 = 4,$$

$$0 \leq a_i \leq 2^{d_i} - 1.$$

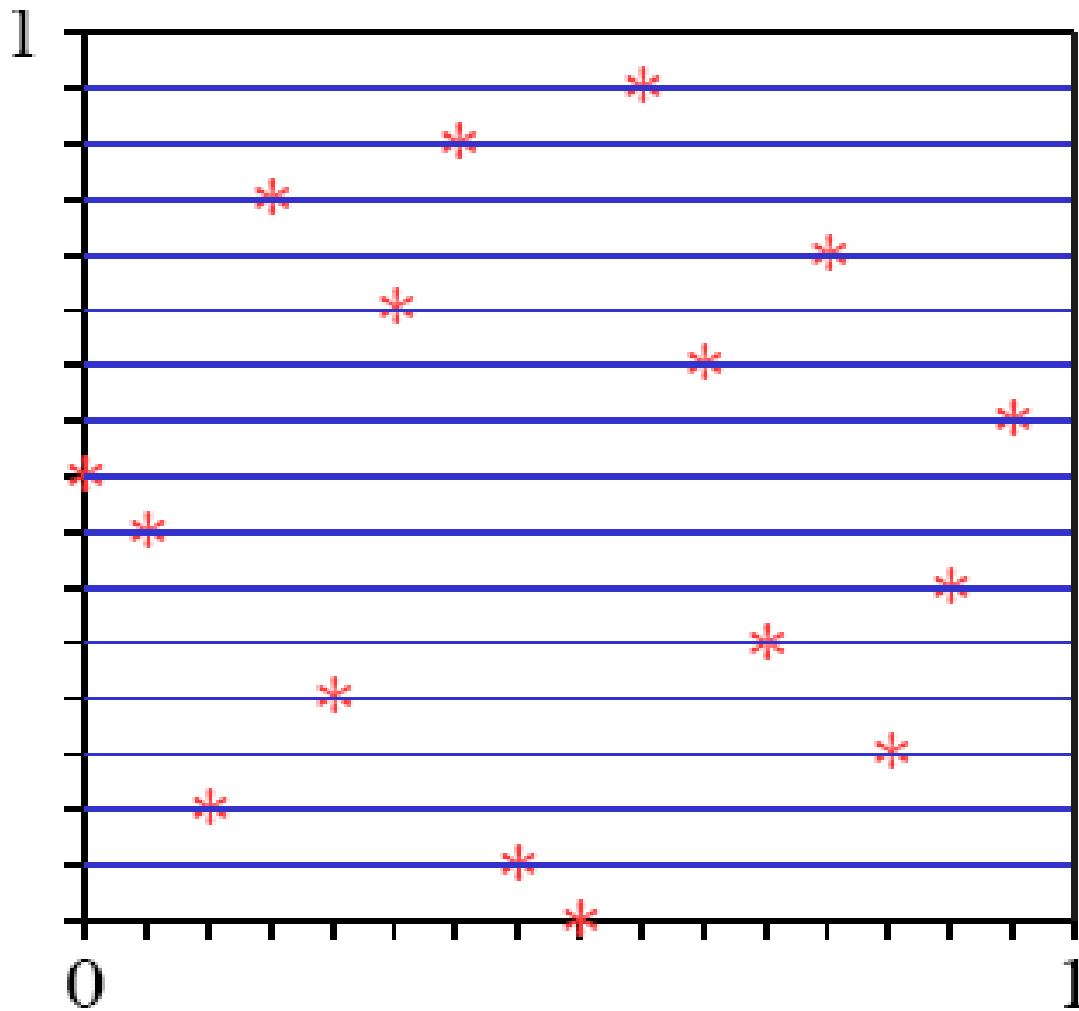
Possible values of (d_1, d_2) with are $(0,4)$, $(1,3)$, $(2,2)$, $(3,1)$, $(4,0)$.



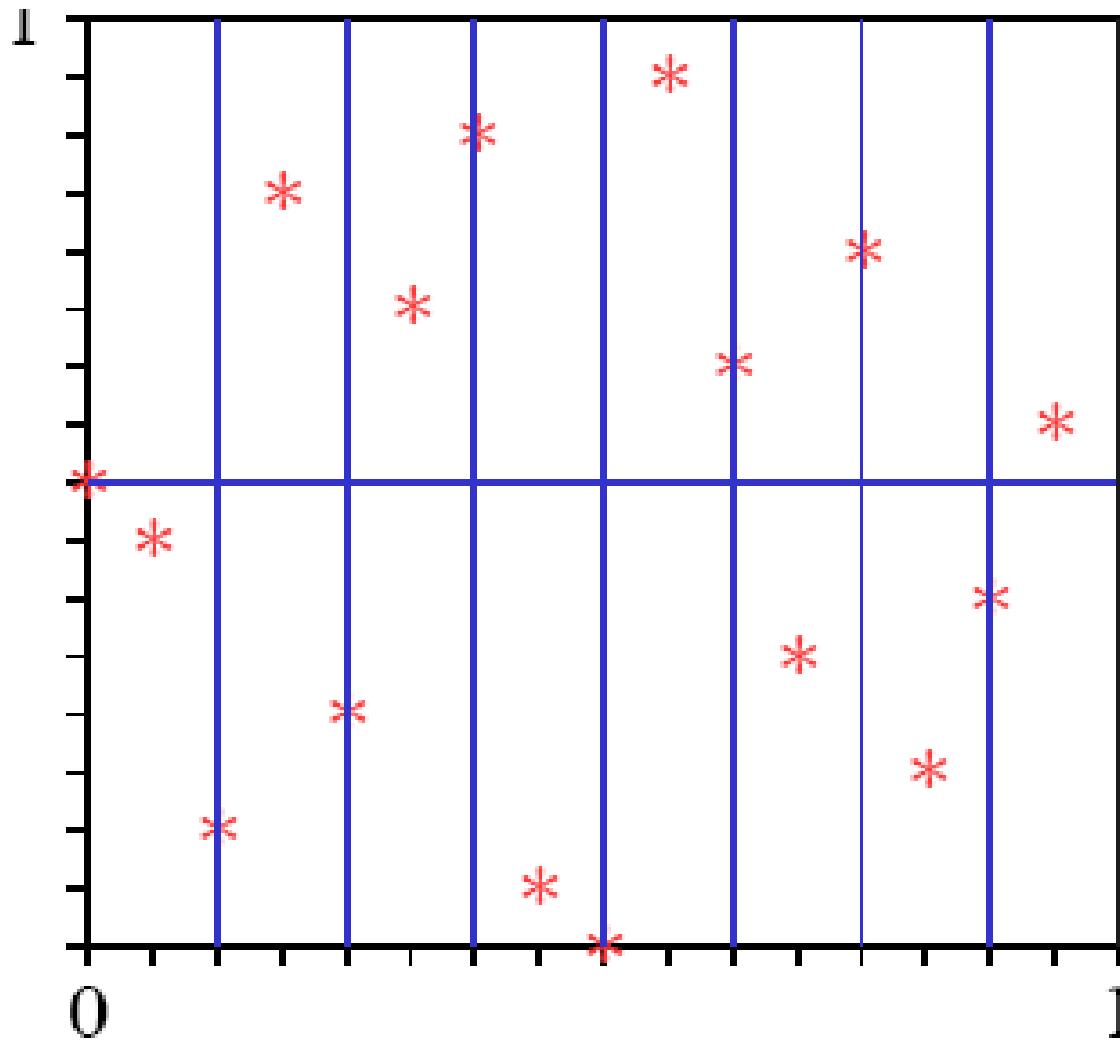
$(0,4,2)$ -net in base 2 of 2^4 points



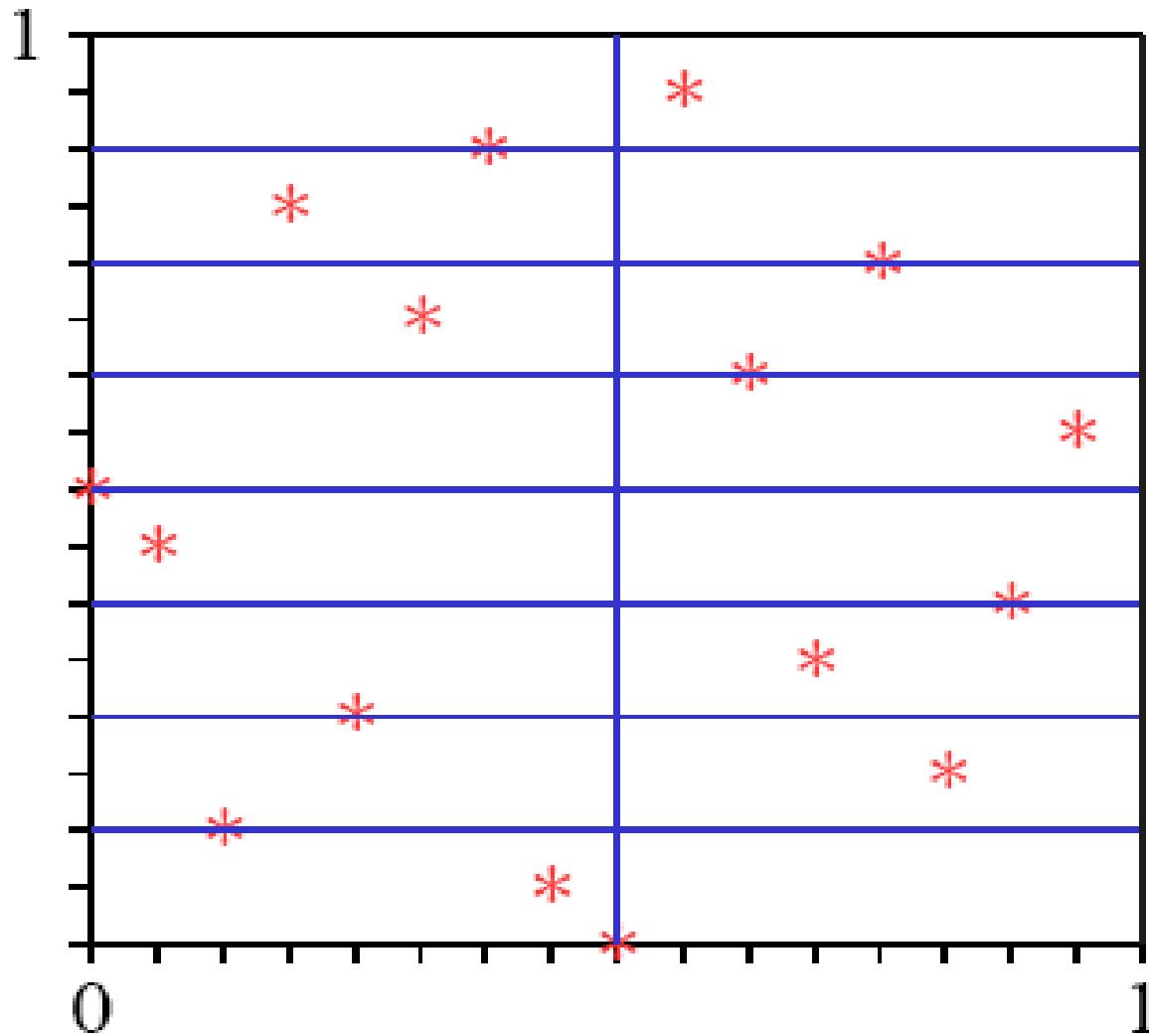
(0,4,2)-net in base 2 of 2^4 points



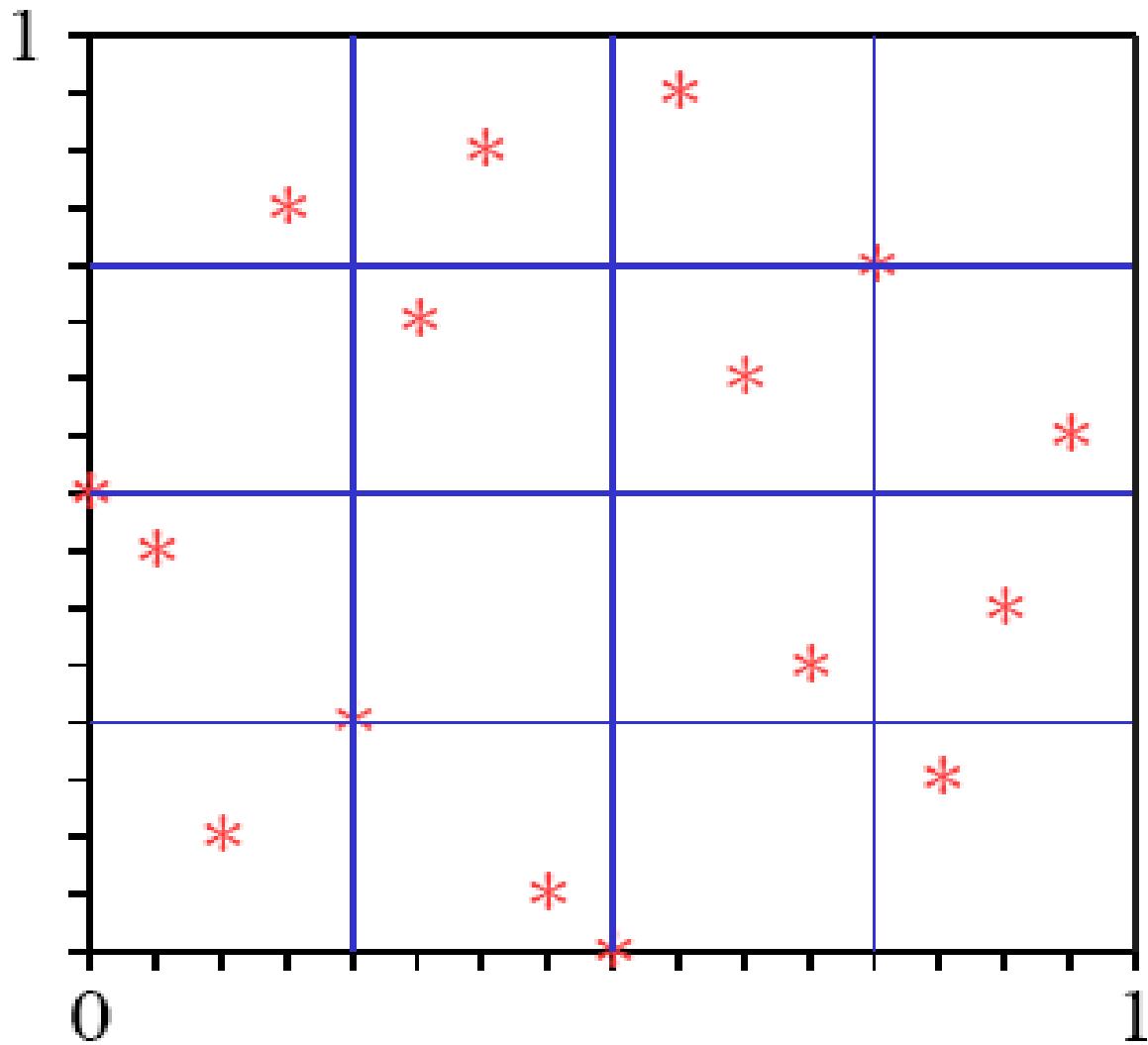
(0,4,2)-net in base 2 of 2^4 points



$(0,4,2)$ -net in base 2 of 2^4 points



$(0,4,2)$ -net in base 2 of 2^4 points



4. (t, m, d) -net and (t, d) -sequence

□ Definition: (t, d) -sequence

Let $t \geq 0$ be an integer.

An infinite sequence x_1, x_2, \dots of points in $[0,1]^d$ is a (t,d) -sequence in base b , if for all integers $k \geq 0$, and $m > t$, the point set

$$\{x_n : kb^m \leq n < (k+1)b^m\}$$

is a (t, m, d) -net.

Discrepancy bound for (t, m, d) -net

Theorem: (Niederreiter, Thm 4.10, p.56)

The star discrepancy of a (t, m, d) -net P in base b with $m > 0$ satisfies

$$ND_N^*(P) \leq B(d, b)b^t (\log N)^{d-1} + O(b^t (\log N)^{d-2}),$$

where $B(d, b) = \begin{cases} \left(\frac{b-1}{2 \log b}\right)^{d-1}, & \text{if either } s = 2 \text{ or } b = 2, s = 2, 3; \\ \frac{1}{(d-1)!} \left(\frac{[b/2]}{\log b}\right)^{d-1}, & \text{otherwise.} \end{cases}$

The convergence order of star discrepancy of (t, m, d) -net: ...

5. Constructions

- The (t, m, d) -nets and (t, d) -sequences in base b have the desired property of having **low star discrepancy** (at least in an asymptotic sense).
- Now the question arises **how to construct** concrete examples of (t, m, d) -nets and (t, d) -sequence in base b with good equidistribution properties.
- The concept of *digital nets* and *sequences* is a general framework for the construction of point sets and sequences, and it is the basis for virtually all concrete constructions of (t, m, d) -nets and (t, d) -sequences.

5. Constructions

□ **Digital Nets:**

Halton (1960)

Sobol(1967)

Faure (1982)

Niederreiter(1992)

Niederreiter and Xing (1995, 1996), ...

□ **Good Lattice Points:**

Korobov method (1960)

Hua-Wang method (1961)

Sloan method (recent years), ...

van der Corput sequence (one dimension)

1. Choose a prime b ., e.g. $b=2$;
2. Express each integer n in base b ,

$$n = \sum_{j \geq 0} a_j b^j = (\dots a_2 a_1 a_0)_b, \text{ e.g. } n = 6 = 1 \cdot 2^2 + 1 \cdot 2^1 + 0 \cdot 2^0 = (110)_2$$

3. Reflect the number about the decimal point,

$$X_n = \sum_{j \geq 0} a_j b^{-j-1} = (0.a_0 a_1 a_2 \dots)_b =: \phi_b(n),$$

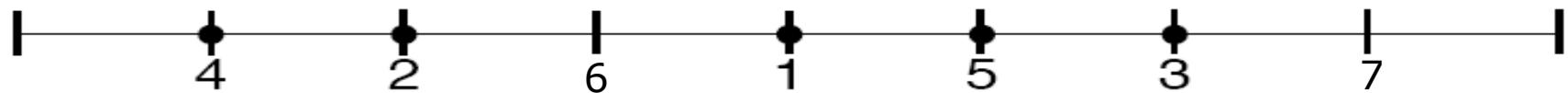
$$\text{e.g. } X_6 = \frac{0}{2} + \frac{1}{2^2} + \frac{1}{2^3} = (0.011)_2 = \frac{3}{8}$$

$\phi_b(n)$ is called **radical-inverse function**.

van der Corput sequence is: $\{\phi_b(n)\}_{n=1}^{\infty}$.

Example: (demonstration)

n	1	2	3	4	5	6	7
n in base 2	1	10	11	100	101	110	111
x_n in base 2	0.1	0.01	0.11	0.001	0.101	0.011	0.111
X_n	$1/2$	$1/4$	$3/4$	$1/8$	$5/8$	$3/8$	$7/8$



van der Corput sequence

Proposition

van der Corput sequence is a $(0, 1)$ -sequence in base b .

We need to prove that $\{x_n : kb^m \leq n < (k + 1)b^m\}$ is a $(0, m, 1)$ - net for integers $k \geq 0$ and $m \geq 1$, i.e., every elementary interval $[ab^{-m}, (a + 1)b^{-m})$ of length $1/b^m$, where $0 \leq a < b^m$, contains exactly one point of the set.

a_0, \dots, a_{m-1} 每一组可能（共
 b^m 种可能）刚好对应一个
interval of length b^{-m}

Proof

Assume $n = \underline{a_N \cdots a_m} \underline{a_{m-1} \cdots a_0}$, then $X_n = 0.\underline{a_0 \cdots a_{m-1}} \underline{a_m \cdots a_N}$.

For $kb^m \leq n < (k + 1)b^m$, the m least significant digits in the digital expansion of n can range freely, whereas the remaining leading digits are fixed.

For $x_n = \phi_b(n)$ this means that its first m leading digits after the "decimal point" can range freely, whereas the remaining digits are fixed. So each b -adic interval of length b^{-m} contains exactly one of these x_n .

Halton sequence in multiple dimension

Let b_1, b_2, \dots, b_d be relatively prime integers greater than 1. Often, we choose the d smallest primes b_1, b_2, \dots, b_d and define Halton sequence as:

$$x_n = (\phi_{b_1}(n), \phi_{b_2}(n), \dots, \phi_{b_d}(n)), n = 0, 1, 2, \dots$$

Halton sequence has discrepancy in the order $O(N^{-1}(\log N)^d)$:

$$D_N^*(P) \leq C_{\text{Halton}}(d) \frac{(\log N)^d}{N} + O\left(\frac{(\log N)^{d-1}}{N}\right).$$

So Halton sequence is a low discrepancy sequence in dimension d

Example:

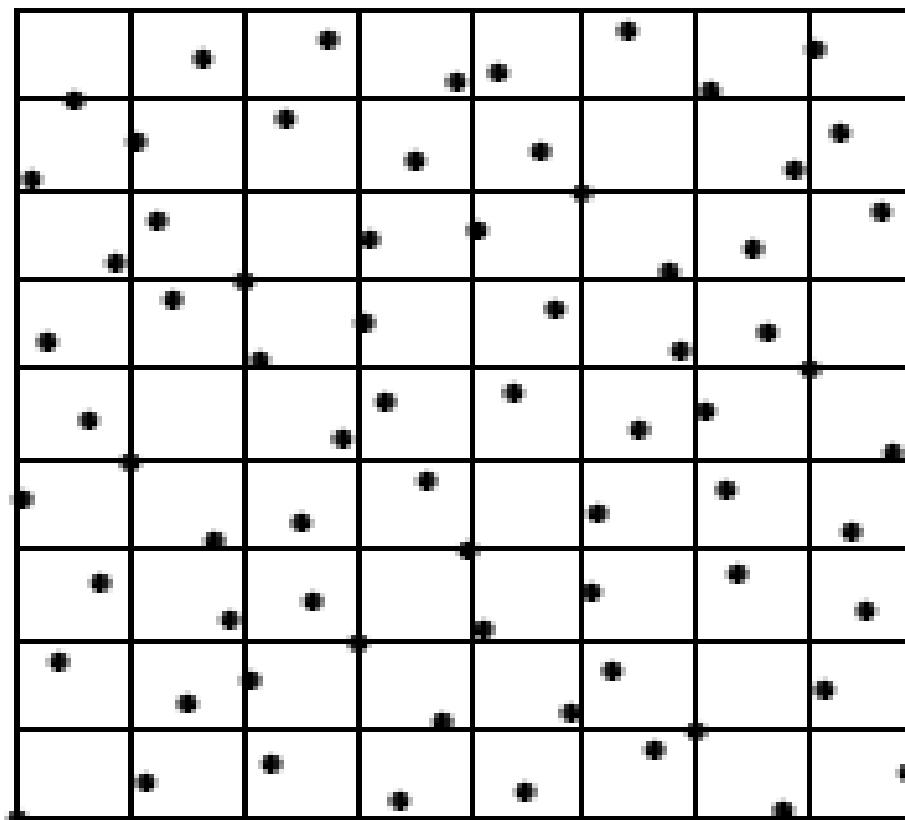
Two-dim Halton sequence, using bases 2 and 3:

$$x_n = (\phi_2(n), \phi_3(n)), n = 0, 1, 2, \dots$$

n	^	1	2	3	4	5	6	7	8	9	10	11
D1		1/2	1/4	3/4	1/8	5/8	3/8	7/8	1/16	9/16	5/16	13/16
D2		1/3	2/3	1/9	4/9	7/9	2/9	5/9	8/9	1/27	10/27	19/27

Halton sequence

The Halton sequence in bases 2 and 3 with $2^3 \times 3^2 = 72$ points:

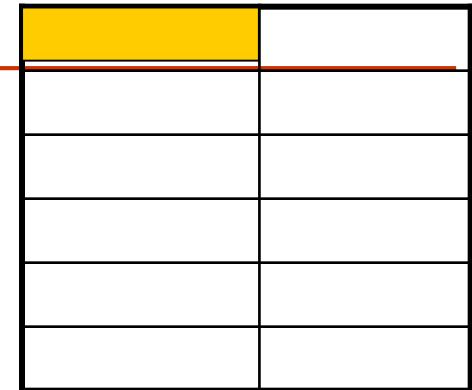


First few points of Halton sequence

Term	Dimension 1	Dimension 2	Dimension 3	Dimension 4
	base 2	base 3	base 5	base 7
1	1/2	1/3	1/5	1/7
2	1/4	2/3	2/5	2/7
3	3/4	1/9	3/5	3/7
4	1/8	4/9	4/5	4/7
5	5/8	7/9	1/25	5/7
6	3/8	2/9	6/25	6/7
7	7/8	5/9	11/25	1/49
8	1/16	8/9	16/25	8/49
9	3/16	1/27	21/25	15/49

Remark

The requirement that b_i are **relatively prime** is necessary for the sequence to fill the hypercube.



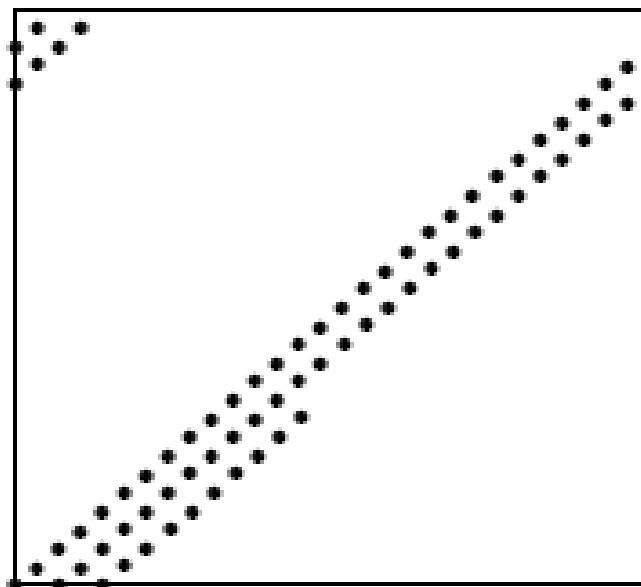
If $b_1 = 2$, $b_2 = 6$, then the corresponding sequence has no points in $[0,1/2) \times [5/6,1)$, since

- In order the first component x_1 to fall in $[0,1/2)$, i.e., $x_1 = (0.0***)_2$, **n must be even**;
- However, in order the second component x_2 to fall in $[5/6,1)$, i.e., $x_2 = (0.5***)_6$, **n must be odd**.

Please find other intervals that do not contain any points.

Remark:

Large bases in Halton sequence might cause problem:
Projections to lower dim faces may have very poor distribution properties. For practical applications such behavior is **not desirable**.



The Halton sequence in bases 29 and 31 with 72 points

Generalization

- The Halton sequence is easy to implement.
- It is a good tool for testing whether QMC will help on your problem.
- It is also well suited to homework problems.

Generalization

Randomized Halton sequence

(Wang & Hickernell,
Mathematical and Computer Modelling, 2000)

Using a form

$$x_n = g(x_{n-1}), n = 1, 2, \dots$$

Given $x_0 \sim U(0,1)$.

Very easy to implement.

N-element Hammersley point set

- We can achieve slightly better uniformity if we fix the number of points N :

$$x_n = \left(\frac{n}{N}, \phi_{b_1}(n), \dots, \phi_{b_{d-1}}(n) \right), n = 0, 1, 2, \dots, N-1.$$



Discrepancy: $O(N^{-1}(\log N)^{d-1})$.

$$D_N^*(P) \leq C_{\text{Halton}}(d-1) \frac{(\log N)^{d-1}}{N} + O\left(\frac{(\log N)^{d-2}}{N}\right).$$

5. Constructions

- General construction of **digital sequence**

Let b be a prime and $d > 1$. Assume we have d **generating matrices** C_1, C_2, \dots of dimension infinity with entries in \mathbb{Z}_b . Let

$$n = \sum_{k \geq 0} a_k(n) b^k,$$

Define

$$(d_{j,0}(n), d_{j,1}(n), \dots)^T = C_j (a_0(n), a_1(n), \dots)^T$$

The **jth coordinate of the nth point** of the digital sequence is given by

$$x_{nj} = \sum_{k \geq 0} d_{j,k}(n) b^{-k-1}.$$

Sobol' Sequence (1967)

□ One-dimensional case:

Generating one dimensional points in interval [0,1]:

(1) Find a set of **direction numbers**:

$$v_1, v_2, v_3, \dots$$

(2) Find the binary representation of the integer n:

$$n = (\dots b_3 b_2 b_1)_2.$$

(3) Set

$$x_n = b_1 v_1 \oplus b_2 v_2 \oplus b_3 v_3 \oplus \dots$$

where \oplus denotes the bit - by - bit exclusive - or operation.

How to obtain direction numbers: v_1, v_2, v_3, \dots ?

-
- (1) Choose a **primitive polynomial** of degree d in the field Z_2 :

$$P = x^d + a_1x^{d-1} + a_2x^{d-2} + \dots + a_{d-1}x + 1$$

- (2) Choose d initial values (integers):

m_1, \dots, m_d satisfying : m_i is odd and $m_i < 2^i$.

- (3) Use the recurrence to calculate (for $i=d+1, \dots$)

$$m_i = 2a_1m_{i-1} \oplus 2^2a_2m_{i-2} \oplus \dots \oplus 2^{d-1}a_{d-1}m_{i-d+1} \oplus 2^dm_{i-d} \oplus m_{i-d}.$$

- (4) Set $v_i = \frac{m_i}{2^i}, i = 1, 2, \dots$

Primitive polynomials of degree d

- 1. Irreducible** (cannot be factored into polynomials of smaller degree)
- 2. Does not divide the polynomial**

$$x^r + 1 \text{ for } r < 2^d - 1.$$

Example: $x^2 + x + 1$ is a primitive polynomial of degree 2, since it is irreducible and does not divide

$$x^r + 1 \text{ for } r < 2^2 - 1 = 3.$$

(but it does divide $x^3 + 1$)

$$m_i = 2a_1m_{i-1} \oplus 2^2a_2m_{i-2} \oplus \dots \oplus 2^{d-1}a_{d-1}m_{i-d+1} \oplus 2^da_{d-1}m_{i-d} \oplus m_{i-d}.$$

Demonstration:

- (1) Take a primitive polynomial $P = x^3 + x + 1$
- (2) Initialization: $m_1 = 1, m_2 = 3, m_3 = 7$.
- (3) Use recurrence $m_i = 4m_{i-2} \oplus 8m_{i-3} \oplus m_{i-3}$.
we obtain: $m_4 = 5, m_5 = 7, m_6 = 43, \dots$
- (4) Direction numbers are obtained by

$$\nu_i = \frac{m_i}{2^i}, i = 1, 2, \dots$$

Primitive polynomials

Degree	Primitive Polynomials
0	1
1	$3(x + 1)$
2	$7(x^2 + x + 1)$
3	$11(x^3 + x + 1), 13(x^3 + x^2 + 1)$
4	19, 25
5	37, 59, 47, 61, 55, 41
6	67, 97, 91, 109, 103, 115
7	131, 193, 137, 145, 143, 241, 157, 185, 167, 229, 171, 213, 191, 253, 203, 211, 239, 247
8	285, 369, 299, 425, 301, 361, 333, 357, 351, 501, 355, 397, 391, 451, 463, 487

Primitive polynomials of degree 8 or less. Each number in the right column, when represented in binary, gives the coefficients of a primitive polynomial.

Primitive polynomials

(an alternative representation)

Degree	Primitive Polynomials Modulo 2*
1	0 (i.e., $x + 1$)
2	1 (i.e., $x^2 + x + 1$)
3	1, 2 (i.e., $x^3 + x + 1$ and $x^3 + x^2 + 1$)
4	1, 4 (i.e., $x^4 + x + 1$ and $x^4 + x^3 + 1$)
5	2, 4, 7, 11, 13, 14
6	1, 13, 16, 19, 22, 25
7	1, 4, 7, 8, 14, 19, 21, 28, 31, 32, 37, 41, 42, 50, 55, 56, 59, 62
8	14, 21, 22, 38, 47, 49, 50, 52, 56, 67, 70, 84, 97, 103, 115, 122
9	8, 13, 16, 22, 25, 44, 47, 52, 55, 59, 62, 67, 74, 81, 82, 87, 91, 94, 103, 104, 109, 122, 124, 137, 138, 143, 145, 152, 157, 167, 173, 176, 181, 182, 185, 191, 194, 199, 218, 220, 227, 229, 230, 234, 236, 241, 244, 253
10	4, 13, 19, 22, 50, 55, 64, 69, 98, 107, 115, 121, 127, 134, 140, 145, 152, 158, 161, 171, 181, 194, 199, 203, 208, 227, 242, 251, 253, 265, 266, 274, 283, 289, 295, 301, 316, 319, 324, 346, 352, 361, 367, 382, 395, 398, 400, 412, 419, 422, 426, 428, 433, 446, 454, 457, 472, 493, 505, 508
*Expressed as a decimal integer representing the interior bits (that is, omitting the high-order bit and the unit bit).	

Implementation: Gray code

If $n = \dots g_3 g_2 g_1$

is the **Gray code** representation of n , then taking

$$x_n = g_1 v_1 \oplus g_2 v_2 \oplus g_3 v_3 \oplus \dots$$

does not affect the asymptotic discrepancy of Sobol' point set.

Gray code

- (1) The Gray code for n is obtained from the binary representation of $n = \dots b_3 b_2 b_1$ using

$$\dots g_3 g_2 g_1 = \dots b_3 b_2 b_1 \oplus \dots b_4 b_3 b_2.$$

- (2) **Important property:**

The Gray code for n and the Gray code for $n+1$ differ in only one position.

(If b_c is the rightmost zero-bit in the binary representation of n , then g_c is the bit whose value changes: see next page)

Important property

The Gray code for n and the Gray code for $n+1$ differ in only one position:

Set: $n = \dots b_{c+1} b_c 1 1 \dots 1$, b_c is the rightmost zero bit ($b_c = 0$)

Gray code of n :

$$\oplus \dots b_{c+2} b_{c+1} b_c 1 \dots 1 = \dots b_{c+1} 10 \dots 0$$

Since $n+1 = \dots b_{c+1} 100 \dots 0$

Gray code of $n+1$:

$$\oplus \dots b_{c+2} b_{c+1} 1 0 \dots 0 = \dots (b_{c+1} \oplus 1) 10 \dots 0$$

The position of b_c

Important property

The Gray code for n and the Gray code for $n+1$ differ in only one position:

Set: $n = \dots b_{c+1} b_c 1 1 \dots 1$,

b_c is the rightmost zero bit ($b_c = 0$)

Gray code of n :

$$\dots b_{c+1} b_c 1 1 \dots 1$$

$$\oplus \dots b_{c+2} b_{c+1} b_c 1 \dots 1 = \dots b_{c+1} 1 0 \dots 0$$

Since $n + 1 = \dots b_{c+1} 1 0 0 \dots 0$

The position of b_c

Gray code of $n + 1$:

$$\dots b_{c+1} 1 0 0 \dots 0$$

$$\oplus \dots b_{c+2} b_{c+1} 1 0 \dots 0 = \dots (b_{c+1} \oplus 1) 1 0 \dots 0$$

Sobol' Sequence using Gray code

Since $x_n = g_1v_1 \oplus g_2v_2 \oplus g_3v_3 \oplus \dots$

Using the properties above, we can calculate x_{n+1} in terms of x_n :

$$x_{n+1} = x_n \oplus \underline{v_c}$$

where c is the index of the rightmost zero-bit b_c in the binary representation of n (very efficient).

Note : Set : $n = \dots b_{c+1} b_c 1 1 \dots 1,$

b_c is the rightmost zero bit ($b_c = 0$)

Gray code of n : ... b_{c+2} b_{c+1} 10...0

Gray code of $n + 1$: ... b_{c+2} $(b_{c+1} \oplus 1)$ 10...0

High-dimensional Sobol' sequence

- Using different primitive polynomials for different components (using primitive poly.of lower degree). (large table of primitive polynomials in F_2 can be found in the Internet.)
- d-dim Sobol' sequence has discrepancy:

$$D_N^*(P) \leq C_{\text{Sobol}} \frac{(\log N)^d}{N} + O\left(\frac{(\log N)^{d-1}}{N}\right).$$

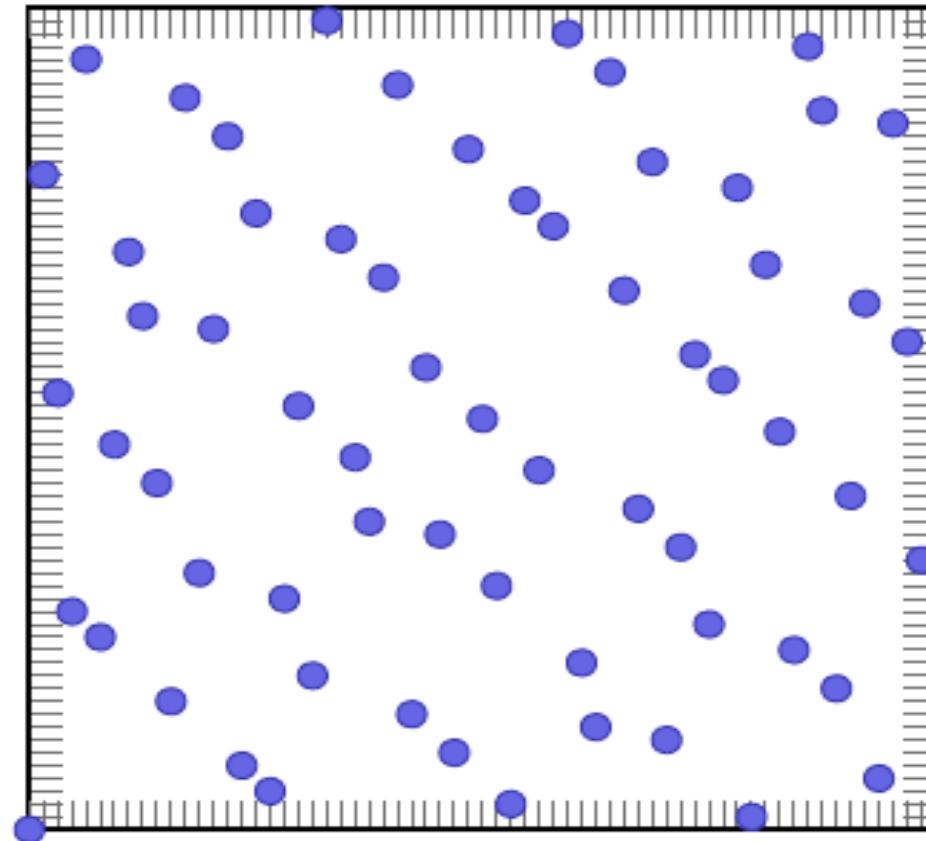
- d-dim Sobol's sequence is (t, d) -sequence in base 2, with

$$t = \sum_{i=1}^d (p_i - 1),$$

Wang, Math. Comput., 2003

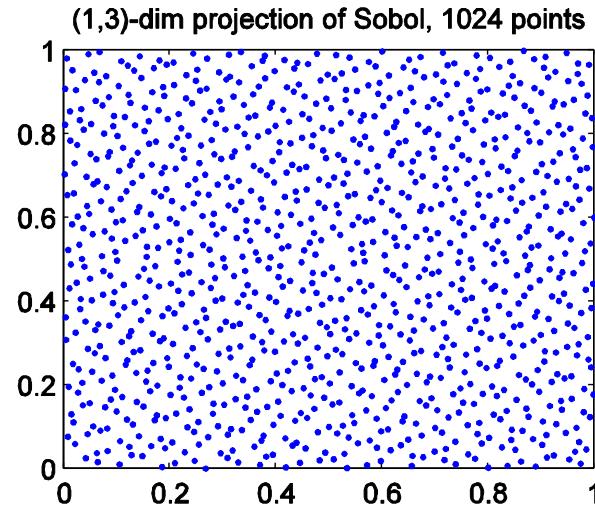
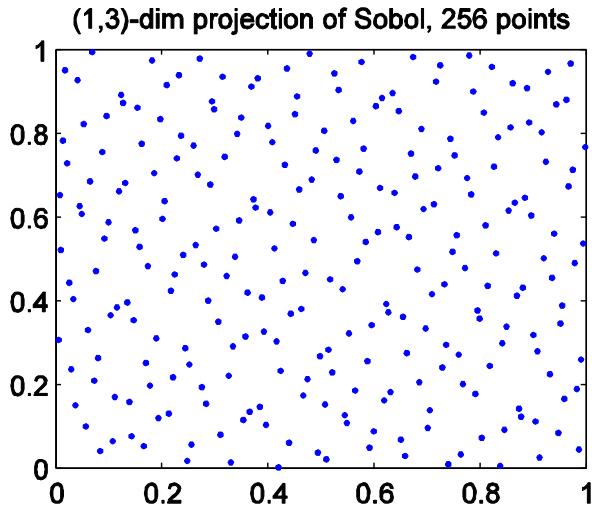
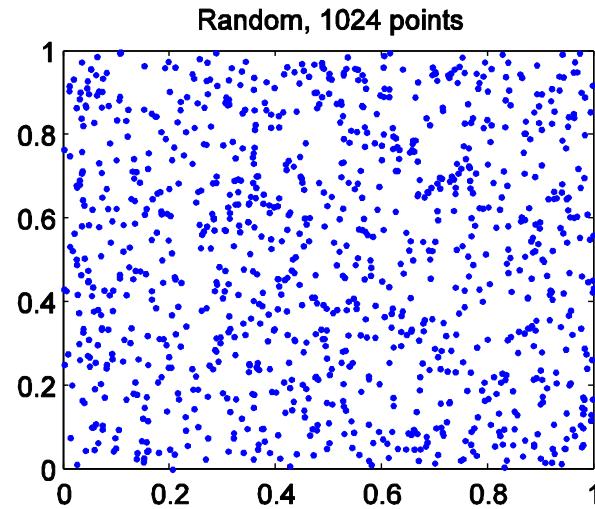
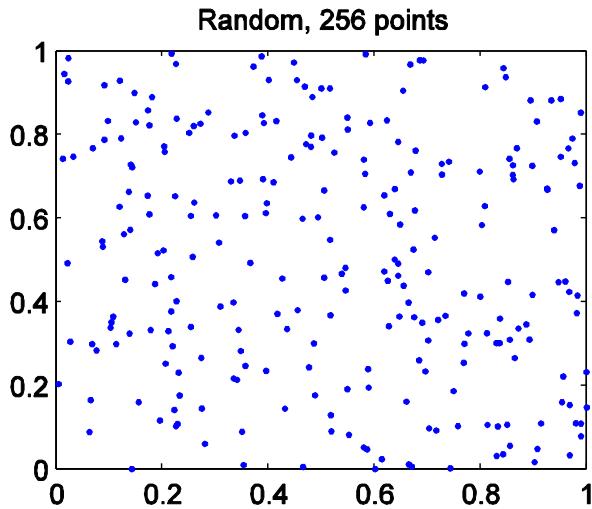
where p_i is the degree of the primitive poly used. ¹⁰⁶

Sobol' Sequence



The first 64 points of a Sobol sequence which form a $(0, 6, 2)$ -net in base 2.

Random vs Sobol' points



Remarks

The most important ingredient is the choice of the **direction numbers**.

◆ **Simple method:**

For a primitive polynomial, choose d initial values (integers) **randomly** and

m_1, \dots, m_d satisfying: m_i is odd and $m_i < 2^i$.

◆ **More advanced method:**

Choose the initial direction numbers such that some properties are satisfied, say, 2-D projections are more uniformly distributed (further reading ...)

5. Constructions

□ Faure sequence (1982)

We generate d-dimensional Faure sequence:

- 0) Choose b as **smallest prime not less than d** ($b \geq d$)
- 1) Representing the integer n in base b as

$$n = \sum_{j \geq 0} a_j^1(n) b^j, \quad n = 0, 1, 2, \dots$$

- 2) The **first component** of nth Faure point is
(the same as in **van der Corput sequence**)

$$\phi_b^1(n) = \sum_{j \geq 0} a_j^1(n) b^{-j-1},$$

↑

Faure sequence (1982)

3) For $k = 2, \dots, d$, define (for $j=0, 1, \dots$)

$$a_j^k(n) = \sum_{i \geq j} \binom{i}{j} a_i^{k-1}(n) \bmod b,$$

$$\phi_b^k(n) = \sum_{j \geq 0} a_j^k(n) b^{-j-1}.$$

k-th component

4) The n-th point of Faure sequence is

$$\phi_b(n) = (\phi_b^1(n), \phi_b^2(n), \dots, \phi_b^d(n)), \quad n = 1, 2, \dots$$

Remark:

Define a function $C(x)$ on the b -adic rational in $[0,1]$:
if x is a rational and if

$$x = \sum_{j \geq 0} x_j b^{-j-1},$$

then

$$C(x) = \sum_{j \geq 0} y_j b^{-j-1} \text{ with } y_j = \sum_{i \geq j} \binom{i}{j} x_i \bmod b.$$

The n -th point of Faure sequence can be written as

$$\phi_b(n) = \left(\phi_b^1(n), C(\phi_b^1(n)), \dots, C^{d-1}(\phi_b^1(n)) \right).$$

Faure sequence (1982)

Assume

$$n = \sum_{j=0}^m a_j^1(n) b^j,$$

$$\begin{pmatrix} (0) & (1) & (2) & (3) & \dots \\ 0 & (1) & (2) & (3) & \dots \\ 0 & 0 & (2) & (3) & \dots \\ 0 & 0 & 0 & (3) & \dots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix}.$$

Then

$$a^k(n) := \begin{pmatrix} a_0^k(n) \\ a_1^k(n) \\ \dots \\ a_m^k(n) \end{pmatrix} = C \begin{pmatrix} a_0^{k-1}(n) \\ a_1^{k-1}(n) \\ \dots \\ a_m^{k-1}(n) \end{pmatrix} \text{ mod } b$$

where C is a $(m+1) \times (m+1)$ matrix with entries

$$C(p, q) = \binom{q-1}{p-1} \text{ for } q \geq p \text{ and zero otherwise}.$$

Faure Sequence

- Faure sequence has discrepancy

$$D_N^*(P) \leq C_{\text{Faure}} \frac{(\log N)^d}{N} + O\left(\frac{(\log N)^{d-1}}{N}\right).$$

- d-dim Faure sequence is **(0, d)-sequence in base b**, where b is the smallest prime not less than d.
- For a detailed description of the construction of Faure sequence, see

Joy, C., P. Boyle, K. S. Tan.

Quasi-Monte Carlo methods in numerical finance.

Management Science, 42, 926-938, 1996.

Ex: 3-D Faure sequence

- Choose $b=3$
- Consider the first 9 points: $n=0, 1, 2, \dots, 8$.
- The coefficients of these n in base 3 are

$$a^1(n) : \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \end{pmatrix},$$

个位 ←

- The matrix C is

$$C = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

- Note:

$$C(p, q) = \binom{q-1}{p-1} \text{ for } q \geq p \text{ and zero otherwise.}$$

Faure sequence

- The vector

$$a^2(n) = Ca^1(n) : \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix},$$

- The vector

$$a^3(n) = Ca^2(n) : \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix},$$

- The 9 Faure points are

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix}, \begin{pmatrix} 2/3 \\ 2/3 \\ 2/3 \end{pmatrix}, \begin{pmatrix} 1/9 \\ 4/9 \\ 7/9 \end{pmatrix}, \begin{pmatrix} 4/9 \\ 7/9 \\ 1/9 \end{pmatrix}, \begin{pmatrix} 7/9 \\ 1/9 \\ 8/9 \end{pmatrix}, \begin{pmatrix} 2/9 \\ 8/9 \\ 2/9 \end{pmatrix}, \begin{pmatrix} 5/9 \\ 2/9 \\ 5/9 \end{pmatrix}, \begin{pmatrix} 8/9 \\ 5/9 \\ 2/9 \end{pmatrix}.$$

5. Constructions

Kronecker sequence

Let $\alpha_1, \dots, \alpha_s \in \mathbb{R}$ be linearly independent over \mathbb{Q} .

Let

$$x_n = (\{n\alpha_1\}, \dots, \{n\alpha_s\}) \quad \text{for } n = 0, 1, 2, \dots$$

For instance, one can choose $\alpha_i = \sqrt{p_i}$ where p_i is the i th prime number.

Remarks

Extensibility

- Halton sequence and Sobol' sequence are **extensible** both in dimension and in the number of points.
- Faure sequence is **extensible in the number of points**, but **not** in the dimension.

$$x_n = (x_{n,1}, x_{n,2}, \dots, x_{n,d}), n = 0, 1, 2, \dots$$

Extensible
in number
of points



Extensible in dimension

Remarks

Uniformity in high dimension:

- For Halton sequence, different components use different bases, with **latter** components corresponding to **large bases**, thus may have bad uniformity.
- For Sobol' sequence, different components use different primitives polynomials, with **the latter** components corresponding to primitives polynomials of **higher degrees**, thus may have bad uniformity (**larger t value**).
- For Faure sequence, for large dimension, we need to use a base **no smaller** than the dimension. Large base may lead to bad uniformity.

6. Good lattice Rules

- A **rank-1 lattice rule** of N points in dimension d is a set of points:

$$\left\{ \left\{ \frac{k}{N} z \right\} = \frac{k}{N} z \bmod 1, k = 0, 1, \dots, N-1 \right\}$$

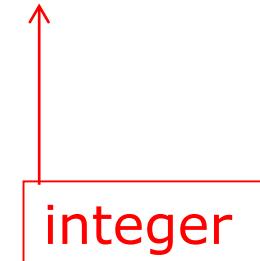
where $z = (z_1, \dots, z_d)^T$ is a d -dimensional vector of integers (called **generating vector**).

The symbol $\{x\}$ means the fractional part of x , i.e., $\{x\} = x - \lfloor x \rfloor$.

6. Good lattice Rules

The other main family of QMC points

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



integer

6. Good lattice Rules

To ensure that this set does indeed contain N distinct points, we require $\gcd(N, z_i) = 1$, for $i=1, \dots, d$.

For N prime, there are $(N-1)^d$ possible choices of the generating vectors.

For large d , this is a huge number.

Question:

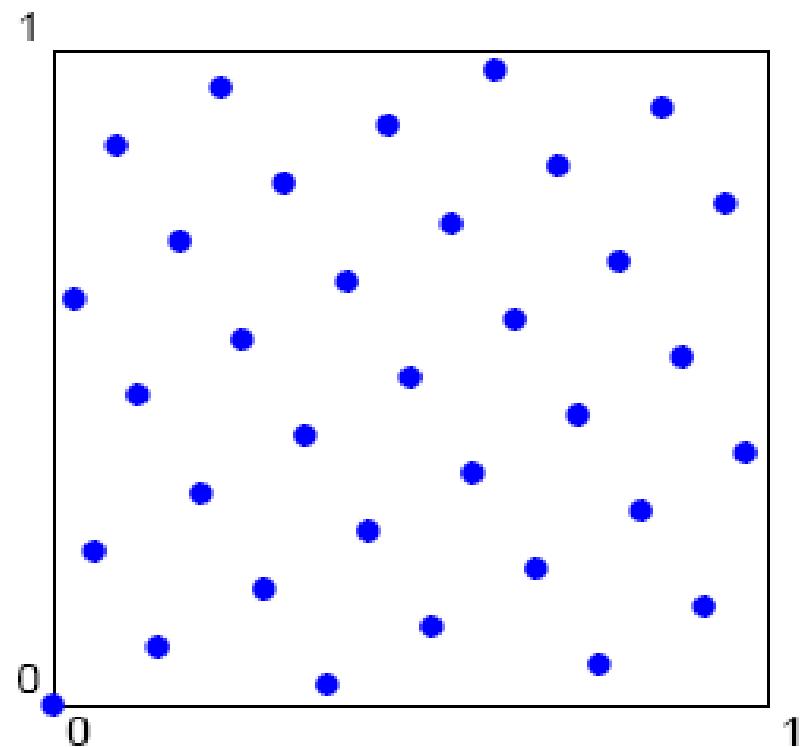
- ✓ How to choose a “good” generator?
- ✓ What is the criterion for good lattice?

6. Good lattice Rules

Example of good lattice points:

$$N=34, z=(1,21)$$

$$\left\{ \begin{array}{l} \left\{ \frac{k}{N} z \right\} = \frac{k}{N} z \bmod 1, \\ k = 0, 1, \dots, N-1 \end{array} \right\}$$



6. Good lattice Rules

□ Korobov lattice rule:

A rank-1 lattice rule with the **generating vector** of the form

$$z = (1, a, a^2, \dots, a^{d-1}) \bmod N$$

for some ~~well-chosen integer~~ $a \in \{1, 2, \dots, N-1\}$.

The Korobov form is a good choice on theoretical as well as practical grounds, only **one parameter** a must be chosen (reducing search from $(N-1)^d$ to $N-1$).

Reduced search space. Minor performance penalty.

Remark

Suppose we find a integer $a \in \{1, 2, \dots, N-1\}$,

How do we find the Korobov generator

$z = (1, a, a^2, \dots, a^{d-1}) \bmod N$ or find $a^k \bmod N$?

$$a^2 = I_1 \cdot \mathbf{N} + R_1 \Rightarrow a^2 \bmod N = R_1.$$

$$a^3 = a \cdot a^2 = a \cdot (I_1 \cdot \mathbf{N} + R_1) = aI_1 \cdot \mathbf{N} + aR_1$$

$$\Rightarrow a^3 \bmod N = aR_1 \xleftarrow{\text{mod } N} R_2$$

$$a^4 = a \cdot a^3 = a(I_2 \mathbf{N} + R_2) = aI_2 \mathbf{N} + aR_2,$$

$$\Rightarrow a^4 \bmod N = aR_2 \xleftarrow{\text{mod } N}$$

.....

6. Good lattice Rules

- A **rank- r lattice rule** of N points in dimension d is a set of points:

$$\left\{ \sum_{i=1}^r \frac{k_i}{N_i} h_i \bmod 1, k_i = 0, 1, \dots, N_i - 1; i = 1, \dots, r \right\}$$

where h_i are independent d -dim vectors of integers.
The integers

$N_1, \dots, N_r \geq 2$ and each N_i divides $N_{i+1}, i = 1, \dots, r-1$,
and $N_1 \dots N_r = N$.

We require that $\gcd(N_i, h_i) = 1$.

Remarks

- Nowadays, there are various procedures (computer search) to find appropriate generating vectors.
- Most of them are based on **minimizing the integration error** over some function classes, say, certain **weighted function spaces** (next chapter).
- No explicit construction is available in $\dim d > 2$.
- An explicit construction in dimension $d=2$ is presented latter.

6. Good lattice Rules

- Classically, the **function space** is defined as

$$E_\alpha(K) = \left\{ f : |F(h)| \leq \frac{K}{(\bar{h}_1 \dots \bar{h}_d)^\alpha} \right\}, \text{ where } \bar{h}_j = \max(|h_j|, 1),$$

$$f(x) = \sum_{h \in \mathbb{Z}^d} F(h) e^{2\pi i h \cdot x}, \quad F(h) = \int_{[0,1]^d} f(x) e^{-2\pi i h \cdot x} dx.$$

The function class is the class of functions f whose Fourier coefficients satisfy the condition above.

This is essentially a class of functions of a certain smoothness, given by α .

Theorem: (Sloan & Joe 1994)

- The integration error of QMC algorithm based on a **rank-1 lattice rule** with generating vector z satisfy

$$|I(f) - Q_N(f)| \leq K P_\alpha(z),$$

$$P_\alpha(z) = \sum_{h \cdot z \equiv 0 \pmod{N}, h \neq 0} \frac{1}{(\bar{h}_1 \dots \bar{h}_d)^\alpha}$$

Equality holds for integrands that are constant multiples of

$$\Phi_\alpha = \sum_{h \in \mathbb{Z}^d} \frac{e^{2\pi i h \bullet x}}{(\bar{h}_1 \dots \bar{h}_d)^\alpha} = \prod_{k=1}^d F_\alpha(x_k),$$

where $F_\alpha(x) = \left(1 + \sum_{h \neq 0} \frac{e^{2\pi i h x}}{|h|} \right).$

6. Good lattice Rules

- Usually, the $P_\alpha(z)$ is taken as the quality measure.
- For even α , the quantity $P_\alpha(z)$ is easy to compute, since

$$F_\alpha(x) = 1 - (-1)^{\alpha/2} \frac{(2\pi)^\alpha B_\alpha(x)}{\alpha!}, (\alpha \text{ even})$$

$B_\alpha(x)$ – Bernoulli polynomial of degree α .

Ex: $F_2(x) = 1 + 2\pi^2 B_2(x) = 1 + \underline{2\pi^2}(x^2 - x + 1/6).$

- The first few Bernoulli polynomials:

$$B_0(x) = 1,$$

$$B_1(x) = x - 1/2,$$

$$B_2(x) = x^2 - x + 1/6,$$

$$B_3(x) = x^3 - 3x^2/2 + x/2,$$

$$B_4(x) = x^4 - 2x^3 + x^2 - 1/30.$$

Theorem:

If $N > 1$ is any integer and $d > 1$, **there exists** a generator $z \in \mathbb{Z}^d$ such that

$$P_\alpha(z) \leq \frac{(2 \log N)^{\alpha d}}{N^\alpha} + O\left(\frac{(\log N)^{\alpha d - 1}}{N^\alpha}\right).$$

Remarks

- There are many tables of generating vectors in low dimensions (Hua & Wang 1981, Sloan & Joe 1994).
- It was pointed out in Wang & Sloan (SIAM J. Numer. Anal., 2006) that it is dangerous to use the lattice rules obtained based on the classical quality measure for high-dim financial problems (**why?**).

□ **Question:**

How to modify the constructions such that the obtained lattice rules are suitable for high-dim financial problems?

(Wang, SIAM J. Sci. Comput., 2007, based on the theory of weighted function spaces, next chapter)

Example: 2D Fibonacci lattice rule

- Fibonacci number:

$$F_1 = F_2 = 1, F_k = F_{k-1} + F_{k-2}, \text{ for } k \geq 3,$$

i.e., 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, ...

- For each Fibonacci number F_k with $k > 2$, there is a two-dim Fibonacci lattice rule of order F_k (with the number of points F_k)

$$x_j = \left(\frac{j}{N}, \left\{ \frac{jF_{k-1}}{N} \right\} \right), \quad N = F_k, \quad j = 0, \dots, N-1.$$

- Corresponding generator: $Z = (1, F_{k-1})$.

6. Good lattice Rules

Recent research on good lattice Rules:

- ✓ Extensible in the number of points
or/and in dimension

- ✓ Polynomial lattice points

7. Randomized Quasi-Monte Carlo

- **Goal:** Error estimation for QMC

Consider

$$I(f) = \int_{[0,1]^d} f(x) dx, \quad Q(f) = \frac{1}{N} \sum_{i=1}^N f(x_i).$$

- Assume that

$$P = \{x_1, x_2, \dots, x_N\}, x_i = (x_{i,1}, \dots, x_{i,d}).$$

is a low discrepancy point set.

Three methods of randomization:

- ✓ **Random shift,**
- ✓ **Digital b-ary shift,**
- ✓ **Scrambling.**

Introduce random element
to deterministic points.

1. Random Shift

(0) Generate a low discrepancy point set

$$P = \{x_1, x_2, \dots, x_N\}, x_i = (x_{i,1}, \dots, x_{i,d})$$

(1) Generate M random vectors

$$\Delta_j \sim U([0,1]^d), j=1, \dots, M.$$

(2) Form a “**random shift version**” of P and calculate

$$P + \Delta_j := \{y_i \mid y_i = (x_i + \Delta_j) \bmod 1, i=1, \dots, N\},$$

↑ ↑ $j = 1, \dots, M.$

Deterministic random

Note: The same shift for all points in the set P .¹³⁶

1. Random Shift

(3) Calculate:

$$Q_j(f, \Delta_j) = \frac{1}{N} \sum_{i=1}^N f((x_i + \Delta_j) \bmod 1), \quad j = 1, \dots, M.$$

(4) Calculate the overall average and the corresponding variance

$$\bar{Q}(f) = \frac{1}{M} \sum_{j=1}^M Q_j(f, \Delta_j),$$

$$\text{Var}(\bar{Q}(f)) \approx \frac{1}{M(M-1)} \sum_{j=1}^M [Q_j(f, \Delta_j) - \bar{Q}(f)]^2.$$

Note: $\bar{Q}(f)$ is an unbiased estimate of $I(f)$.

1. Random Shift

(5) Based on the estimated variance, one can compare the efficiency of the estimate with MC

$$\text{Eff}(\bar{Q}(f)) = \frac{\text{Var}(\text{MC based on } M \text{ samples})}{\text{Var}(\bar{Q}(f))}.$$

One also can calculate the 95% (approximate) confidence interval:

$$(\bar{Q}(f) - 1.96\sqrt{\text{Var}(\bar{Q}(f))}, \bar{Q}(f) + 1.96\sqrt{\text{Var}(\bar{Q}(f))})$$

Note: $\text{Var}(\bar{Q}(f)) = \frac{1}{M} \text{Var}(Q_j)$, $\text{Var}(Q_j) \approx \frac{1}{(M-1)} \sum_{j=1}^M [Q_j(f, \Delta_j) - \bar{Q}(f)]^2$.

7. Randomized Quasi-Monte Carlo

Randomized QMC combines the best of the MC and QMC:

- (1) Fast convergence**
- (2) Unbiasedness**
- (3) Simple error estimation**

$$Q_j(f, \Delta_j) = \frac{1}{N} \sum_{i=1}^N f((x_i + \Delta_j) \bmod 1), \quad j = 1, \dots, M.$$

Remarks

$$\bar{Q}(f) = \frac{1}{M} \sum_{j=1}^M Q_j(f, \Delta_j),$$

For a given number of function evaluation **NM**,

- **Large value of N, small value of M:** better accuracy in $Q_j(f, \Delta_j)$ (but poorer variance estimate).
- **Small value of N, large value of M:** better accuracy in the final variance estimation (but poorer error).
- In practice, I usually take $N = 10M$.
- Random shift method is especially suitable for good lattice rules. But it is **not the first choice for digital nets**.
- By using randomized QMC, we can estimate error in the usual way and get a confidence interval.

Digital b-ary shift

(0) Let P be a digital net in base b :

$$P = \{x_1, \dots, x_N\}, \quad x_i = (x_{i,1}, \dots, x_{i,d}), \quad x_{i,j} = \sum_{k=1}^{\infty} a_{i,j,k} b^{-k}.$$

(1) Generate M random vector

$$\Delta = (\Delta_1, \dots, \Delta_d) \sim U([0,1]^d), \quad \Delta_j = \sum_{k=1}^{\infty} \delta_{j,k} b^{-k}.$$

(2) Form a “**Digital shift version**” of P

$$P + \Delta := \{x_i \oplus \Delta, i = 1, \dots, N\},$$

Using Operations
over \mathbb{F}_b

$$x_i \oplus \Delta = (y_{i,1}, \dots, y_{i,d}), \quad y_{i,j} = \sum_{k=1}^{\infty} (a_{i,j,k} \oplus \delta_{j,k}) b^{-k}$$

The remaining steps are the same as random shift.⁴¹

Digital b-ary shift

- For example, for Sobol' sequences, digital shift is achieved by bitwise exclusive-or operation so that

$$\begin{aligned} & 0.1010011 \\ \oplus & 0.0110110 \\ = & 0.1100101 \end{aligned}$$

- The benefit of the digital shift is that it maintains the special properties of the Sobol' sequence.

Note: The same shift for all points in the set P .

Scrambling

(1) Let P be a digital net in base b :

$$P = \{x_1, \dots, x_N\}, \quad x_i = (x_{i,1}, \dots, x_{i,d}), \quad x_{i,j} = \sum_{k=1}^{\infty} a_{i,j,k} b^{-k}.$$

(2) Form a “**Scrambled version**” of P

$$P_{\text{scrambling}} = \{y_1, \dots, y_N\},$$

$$y_i = (y_{i,1}, \dots, y_{i,d}), \quad y_{i,j} = \sum_{k=1}^{\infty} d_{i,j,k} b^{-k},$$

$$d_{i,j,k} = \pi_{j, a_{i,j,1}, \dots, a_{i,j,k-1}}(a_{i,j,k}).$$



All the permutations are independent, and the permutations applied to the k th digit depends on j (but not on i) and on the preceding $k-1$ digits.

Alternatively:

- Let $\mathbb{Z}_b = \{0, 1, \dots, b - 1\}$ and let

$$x = \frac{x_1}{b} + \frac{x_2}{b^2} + \dots, \quad x_1, x_2, \dots \in \mathbb{Z}_b.$$

- Randomly choose permutations $\sigma, \sigma_{x_1}, \sigma_{x_1, x_2}, \dots : \mathbb{Z}_b \rightarrow \mathbb{Z}_b$.
- Let

$$y_1 = \sigma(x_1)$$

$$y_2 = \sigma_{x_1}(x_2)$$

$$y_3 = \sigma_{x_1, x_2}(x_3)$$

.....

- Set

$$y = \frac{y_1}{b} + \frac{y_2}{b^2} + \dots.$$

The same scrambling for
all points in the set P.

Remarks:

- For practical computation we must truncate the b -ary expansion of the points at some finite K_{\max} . Take K_{\max} large enough that $b^{-K_{\max}}$ is small compared to error committed in approximating real values in computer.
- An alternative choice of K_{\max} is take $K_{\max}=M$, if there are at most b^M points that will ever be used.
- Some simplifications of scrambling are possible.

Remarks about RQMC

- Random shift: for good lattice rules;
- Digital shift and scrambling: for digital nets.
- The primary purpose of randomized QMC is for error estimation.
- However, randomization can be shown to enhance the accuracy of the integration rules and improve the equidistribution of point sets.

For smooth f ,

$$V_{SNET}(\hat{I}) = O(n^{-3}(\log n)^{d-1})$$

RQMC

- The idea of RQMC is to introduce some randomness in the QMC method in order to produce a confidence interval, or from a dual viewpoint to use the QMC approach as a variance reducer in the MC method.
- Finally, randomized QMC is a specific (and not so easy to handle) variance reduction method, not a QMC speeding up method, which suffers from one drawback of regular QMC: the difficulty to identify the class of functions with finite variations.

8. Using QMC in Finance

- ❑ Many QMC points have the property that the points are more uniformly distributed through the lowest dimensions.
- ❑ Consequently, it is important to think about **how the dimensions are allocated to the problem**.
- ❑ Ideally, we'd like to use a change of variables, so the function we're integrating **depends mainly on the first coordinates**.
- ❑ We have studied the construction of BM motion. We have generated correlated Normals through $Y = A X$ with X i.i.d. $N(0, 1)$. These methods are changes of variables (may reduce the effective dimension).

Remark (very important!)

In simulating paths of the underlying by QMC,
use **ONE QMC POINT** to simulate **ONE PATH**.

A single asset at multiple dates : (S_{t1}, \dots, S_{td}) :

$$(U_1, \dots, U_d) \Rightarrow (Z_1, \dots, Z_d) \Rightarrow (B_{t1}, \dots, B_{td}) \Rightarrow (S_{t1}, \dots, S_{td})$$

Multiple assets at one date : (S_T^1, \dots, S_T^m) :

$$(U_1, \dots, U_m) \Rightarrow (Z_1, \dots, Z_m) \Rightarrow (B_T^1, \dots, B_T^m) \Rightarrow (S_T^1, \dots, S_T^m)$$

Multiple assets at multiple dates : $(S_{t1}^1, \dots, S_{t1}^m, \dots, S_{td}^1, \dots, S_{td}^m)$:

$$(U_1, \dots, U_{md}) \Rightarrow (Z_1, \dots, Z_{md}) \Rightarrow \begin{pmatrix} B_{t1}^1, \dots, B_{td}^1 \\ \dots \\ B_{t1}^m, \dots, B_{td}^m \end{pmatrix}$$

$$\Rightarrow (S_{t1}^1, \dots, S_{t1}^m, \dots, S_{td}^1, \dots, S_{td}^m)$$

Have different ways to
assign the coordinates

Remark:

Multiple assets at multiple dates :

$$(U_1, \dots, U_{md}) \Rightarrow (Z_1, \dots, Z_{md})$$

Have different ways to assign the coordinates

Standard BM	correlated BM
$\begin{pmatrix} W_{t1}^1, \dots, W_{td}^1 \\ \dots \\ W_{t1}^m, \dots, W_{td}^m \end{pmatrix}$	$\begin{pmatrix} B_{t1}^1, \dots, B_{td}^1 \\ \dots \\ B_{t1}^m, \dots, B_{td}^m \end{pmatrix}$
$\Rightarrow (S_{t1}^1, \dots, S_{t1}^m, \dots, S_{td}^1, \dots, S_{td}^m)$	

Using either PCA or BB to construct uncorrelated BM using

- $Z_1, Z_1+m, Z_1+2m, Z_1+3m, \dots$ for first dim
- $Z_2, Z_2+m, Z_2+2m, Z_2+3m, \dots$ for second dim
- $Z_3, Z_3+m, Z_3+2m, Z_3+3m, \dots$ for third dim
- etc.
- **This uses the “best” dimensions of Z for the overall behavior of the paths.**

8. Using QMC in Finance

❑ New ingredient:

How best to use QMC inputs to generate Brownian increments?

- Methods that **DO NOT** take into account the underlying problem (random walk, BB, PCA).
- Methods that **do** take into account the underlying problem (linear transformation, orthogonal transformation, QR decomposition, CQR, GPCA, ...):
 - ✓ Deal with high dimensionality
 - ✓ Deal with discontinuity in the payoff functions
 - ✓ Deal with both high dimensionality and discontinuity

8. Using QMC in Finance

- ❑ **Path Generation Methods in QMC**

- ❑ **Group I:**

- Standard method (Cholesky decomposition)

- Brownian bridge (BB)

- Principal Component Analysis (PCA)

- ❑ **Group II:**

- Linear transformation (LT)

- Orthogonal transformation (OT)

- QR decomposition

- CQR method

- GPCA method

- (latter)

8. Using QMC in Finance

- **Why does dimension reduction help QMC?**
 - The one dimensional projections are extremely uniform;
 - The initial dimension projections are better than others.
- **Use QMC in the first key dimensions.**

Remarks

- QMC offers large computational savings over the standard MC approach;
- QMC can be used in combination with various variance reduction techniques (AV, CV, IS, ...);
- Best to use randomized QMC to regain confidence intervals, at the cost of slightly poorer accuracy;
- Very important to use PCA, BB construction, or some more advanced constructions to create discrete Brownian increments --- can be much better than “standard” approach, which is equivalent to the Cholesky factorization of the covariance matrix.

Assignment 4 (Dead line: May19)

1. Prove that for L₂ star discrepancy, we have

$$\begin{aligned} \left(T_N^*(P)\right)^2 &= \int_{[0,1]^d} \left| \frac{A(J(x); P)}{N} - m(J(x)) \right|^2 dx \\ &= \frac{1}{3^d} - \frac{1}{2^{d-1}} \frac{1}{N} \sum_{i=1}^N \prod_{k=1}^d (1 - t_{i,k}^2) + \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \prod_{k=1}^d [1 - \max(t_{i,k}, t_{j,k})], \end{aligned}$$

where $t_i = (t_{i,1}, \dots, t_{i,d})$, $P = \{t_i\}$.

Prove also that the square expected L₂ star discrepancy of random points is

$$E\left(T_N^*(P)\right)^2 = \frac{1}{N} \left(2^{-d} - 3^{-d} \right).$$

Assignment 4 (Dead line: May 19)

2. Determine the L_1 -discrepancy, L_2 -discrepancy and the star-discrepancy of the following point sets

$$(1) P = \{n/N : n = 0, 1, \dots, N-1\};$$

$$(2) P = \left\{ \frac{2n-1}{2N} : n = 0, 1, \dots, N-1 \right\};$$

3. Construct "by hand" a $(0, 2, 2)$ -net in base 3.
(Refer to Faure points).

Assignment 4 (Dead line: May 19)

- 4. Write programs to generate Halton, Sobol, or Faure sequences in dimensions up to 50.**
- 5. Empirically compare the L₂ discrepancies of Halton, Sobol, or Faure sequences, and compare with the expected L₂ discrepancy.**
- 6. Compare the efficiency of Halton, Sobol, or Faure sequences for high-dimensional financial problems (say, option pricing) and compare their efficiency with MC.**

Assignment 4:

- 7. Using Halton, Sobol, or Faure sequences, in conjunction with random work construction, Brownian bridge construction and PCA construction of Brownian motion, for option pricing.**
- 8. Using good lattice rules (say, Korobov lattice rules), in conjunction with BB or PCA, for option pricing.**
- 9*. Using randomized QMC for error estimation and variance estimation when doing option pricing (and when comparing the efficiency of different methods).**

Projects:

- ◆ Using QMC methods (digital nets or good lattice point) to price Asian option under the framework of Black-Scholes model. The payoff is

$$f_A = \max\left(0, \sum_{j=1}^d w_j S(t_j) - K\right), \quad \sum_{j=1}^d w_j = 1.$$

The weights :

- (A). $w_j = 1/d, j = 1, \dots, d;$
- (B). $w_j = b 2^j, j = 1, \dots, d, \sum w_j = 1;$
- (C). $w_j = c 2^{-j}, j = 1, \dots, d, \sum w_j = 1;$

Use Path generation method (standard, BB or PCA).

Projects:

- ◆ Using QMC methods (Sobol or korobov points) to price options or other derivatives, combining with various variance reduction techniques:
 - (1) Antithetic variate;
 - (2) Control variables;
 - (3) Importance sampling.

When using importance sampling, you may also investigate the influence of path generation methods (standard, BB, PCA).

Projects:

Consider the mortgage-basked securities problem (see the paper below). Implement the BB and PCA techniques for this problem. Compare the variance of the randomly digital shifted Sobol points with or without BB or PCA for both the “nearly linear” and “nonlinear” parameter sets in that paper.

Reference:

Caflisch, R. E., W. Morokoff and A. Owen. 1997.
Valuation of mortgage backed securities using
Brownian bridges to reduce effective dimension.
J. Comput. Finance, 1, 27-46.

Projects:

- ◆ Use QMC (digital nets or good lattice point) in combination with BB or PCA to price American put option or American Asian option based on **Least-Square Monte Carlo methods**.
- About Least-Square Monte Carlo methods, please read:

Longstaff, F. A. and E. S. Schwartz. 2001.
Valuing American options by simulation:
a least-squares approach.
Rev. Financial Stud., 14, 113-148.

Projects:

- ◆ In Black-Scholes model,
- What is an optimal method of path generation for a given option (i.e., payoff function)?
- What is a “good” option for a given path generation?
- ◆ Why BB or PCA can lose power (i.e., performs worse than standard construction) in QMC (say, for digital option or binary option)?
- ◆ Use QMC in models beyond Black-Scholes (say, Levy process). Develop path generation methods (BB-like or PCA-like) for such models.
- ◆ Construct better digital sequence or good lattice points for finance applications.

Projects:

- ❑ Use QMC (digital nets or good lattice points) for option pricing in Heston's stochastic volatility model:

$$dS(t) = rS(t) + \sigma(t)S(t) \left[\rho dW_1(t) + \sqrt{1 - \rho^2} dW_2(t) \right],$$

$$d\sigma^2(t) = \kappa[\theta - \sigma^2(t)]dt + \sigma_v \sigma(t)dW_1(t),$$

where $W_1(t)$ and $W_2(t)$ are two independent standard Brownian motion, κ is the speed of mean reversion, $\theta > 0$ is the long - run mean variance, $\sigma_v > 0$ is the volatility of the volatility process, ρ is the correlation between BMs driving $S(t)$ and $\sigma^2(t)$.

-
- Note: about Heston's model, see
C. Lemieux. Monte Carlo and Quasi-Monte
Carlo Sampling. 2009, Springer.
(Pages 257-258).

Projects:

- **Use good lattice points (given in the following reference, say) for finance applications:**
 - (1) In conjunction with different variance reduction techniques;
 - (2) In conjunction with different path generation methods (BB, PCA, or others)

Reference:

X. Wang. 2007. Constructing robust good lattice rules for computational finance.
SIAM J. Sci. Comput. Vol. 29, 598-621.

Projects:

Summary: We have

- (1) Different models (BS, stochastic volatility, jump-diffusion ...)
- (2) Different payoff functions (exotic options, ...)
- (3) Different point sets (random, Sobol', Korobov, ...)
- (4) Different variance reduction techniques (AV, CV, IS, ...)
- (5) Different path simulation methods (STD, BB, PCA, ...)

We may use MC or QMC for

- American option pricing,
- Risk measures estimation (VaR, CVaR),
- Greeks (sensitivities) estimation (latter), ...

End of Chapter 6