

タイトル	超一様分布列のラングマイゼーションおよび一様性の評価
著者	串山，繁
引用	北海学園大学工学部研究報告，33：67-76
発行日	2006-02-20

超一様分布列のランダムイゼーションおよび一様性の評価

串 山 繁*

Randomized Low-Discrepancy Sequences and Assessment of Uniformity on Their Sequences

Shigeru KUSHIYAMA*

Abstract

Ordinary low-discrepancy sequences have some correlation between the points in higher dimensions. Because this leads to a large error on multi-dimensional integration, randomized (scrambled) techniques are used widely in QMC (Quasi-Monte Carlo methods). QMC using randomized sequences has improved drastically the calculation time for an expected value of payoff.

In this paper, some randomized techniques on Halton sequence and Faure sequence are introduced. Also, its efficiency is compared through L_2 star discrepancy.

1 序

ランダムイゼーションが施されていないオリジナルな超一様分布列においては、隣接次元の点列間に相関現象が生起する高次元問題が存在する¹⁾。その様な相関を有する数列を直接シミュレーションに用いることは、貧弱な積分近似値に繋がる恐れがある。しかし、ランダムイゼーションを施すことが点列間の相関を消失させ、その結果少ないシミュレーション回数で計算精度が大幅に改善されることが1990年代に実証された。

金融工学の分野では、既にデリバティブの価格決定に一般化Faure列 (Generalized Faure Sequence) などが使用されている。一方、MCMC (Markov Chain Monte Carlo Simulation) と組み合わせて用いる際には、擬似乱数を単純に超一様分布列に置き換えることは出来ないが、Lecot and Tuffin²⁾による成功例も報告されている。

本報告は、信頼性解析における超一様分布列の使用を前提としたシミュレーションの高速化

* 北海学園大学工学部建築学科

* Department of Architecture and Building Engineering, Faculty of Engineering, Hokkai-Gakuen University

を目指した基礎研究として、一般化Halton列、一般化Faure列のランダム化と L_2 スターディスクレパンシーによる一様性の評価について纏めたものである。

2 ディスクレパンシー

QMC (Quasi-Monte Carlo methods) は、擬似乱数を用いるMC (Monte Carlo methods) の代わりに確定的な超一様分布列 (準乱数) を用いるシミュレーション手法である。QMCが使用される際には、積分の真値を (1) 式、シミュレーションの近似値を (2) 式で以下のように表すものとする、

$$I = \int_{[0,1]^s} f(u) du \quad (1)$$

$$\hat{I} = \frac{1}{N} \sum_{n=1}^N f(\xi^{(n)}) \quad (2)$$

予測誤差の上界は、次式のKoksma-Hlawka不等式で表される。

$$\left| \frac{1}{N} \sum_{n=1}^N f(\xi^{(n)}) - \int_{[0,1]^s} f(u) du \right| \leq V(f) D_N^*(p) \quad (3)$$

ただし、 $V(f)$: Hardy and Krauseの意味における関数 f の分散

$D_N^*(p)$: 点列のスターディスクレパンシー

ディスクレパンシーは、 s 次元単位超立方体の点列の不規則性を表す測度であり、スターディスクレパンシー (star discrepancy) や極値ディスクレパンシー (extreme discrepancy) など数種のディスクレパンシーが存在する。図1に2次元のディスクレパンシー例を示す。

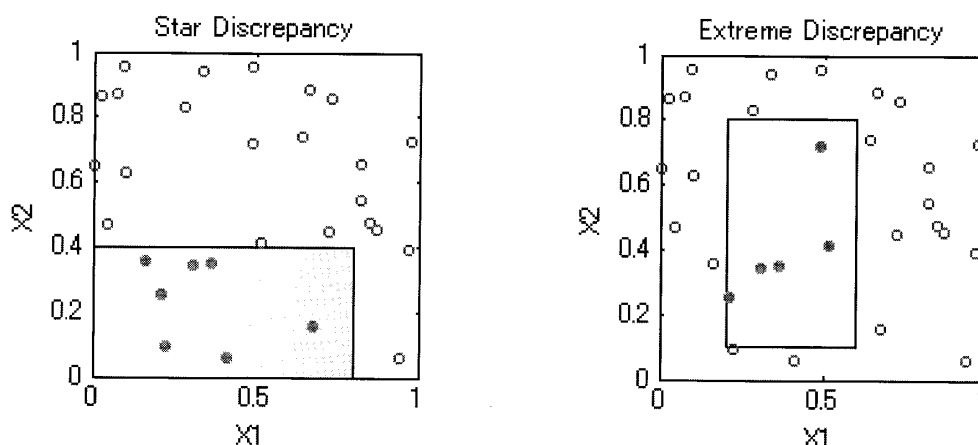


図1 スターディスクレパンシーと極値ディスクレパンシー

図1に示す様に、スターディスクレパンシーは四角形の頂点のひとつが原点に固定された条件で算定されるが、極値ディスクレパンシーにはその様な制約は無い。

L_∞ スターディスクレパンシーの定義は、次の様である．

$$L_\infty = D_N^*(B;P) = \sup_{B^* \in B} |R(B^*;P)| \quad (4)$$

ただし、 $R(B^*;P) = \frac{A(B^*;P)}{N} - \text{Vol}(B^*)$

$A(B^*;P)$; s 次元部分空間 B^* に存在する点列の数

$\text{Vol}(B^*)$; ルベーグ測度の意味での s 次元部分空間の体積 $\prod_{i=1}^s [0, u_i)$

B^* ; $[0,1)^s$ 内の全ての s 次元部分空間の族のひとつ

(4) 式は、スターディスクレパンシーがローカルディスクレパンシー ; $|R(B^*;P)|$ の最大値であることを示している．もし、点列の一様性が完全であるならば、 $|R(B^*;P)|$ の値は如何なる s 次元部分空間に対しても 0 に等しくなる．E. Thiernard³⁾ はスターディスクレパンシーの計算を試みているが多くの計算は困難であり、実際の取り扱いにおいては、(4) 式の L_∞ ノルムを次式の L_2 ノルムに置き換え、簡単化して一様性を判定している．

$$L_2 = T_N^*(B;P) = \left[\int_{[0,1)^s} R(B^*;P)^2 du \right]^{\frac{1}{2}} \quad (5)$$

L_2 スターディスクレパンシーは、 i 番目成分 $x_k^{(i)}$ によって次式で表される⁴⁾．

$$[T_N^*(P)]^2 = \frac{1}{N^2} \sum_{k=1}^N \sum_{m=1}^N \prod_{i=1}^s (1 - \max(x_k^{(i)}, x_m^{(i)})) - \frac{2^{1-s}}{N} \sum_{k=1}^N \prod_{i=1}^s (1 - x_k^{(i)}) + 3^{-s} \quad (6)$$

また、 $T_N^*(P)$ に対するrms (root-mean-square) は、次式で表される．

$$\langle T_N^*(P) \rangle = ((2^{-s} - 3^{-s})/N)^{1/2} \quad (7)$$

後述する一様性の判定比較においては、上記 (6), (7) 式を用いて評価する．

3 Halton列のランダムイゼーション

任意の整数 n を基底の素数 p に対して (8) 式のように p 進数に展開した後、ランダムイズされていないオリジナルのHalton列は、基底逆関数 (radical inverse function) を用いて (9) 式のように表される．

$$n = a_0(n) + a_1(n)p + a_2(n)p^2 + \cdots + a_i(n)p^i + \cdots + a_I(n)p^I = \sum_{i=0}^I a_i(n)p^i \quad (8)$$

ただし、 $I = \ell n(n)/\ell n(p)$

$$x_n = \Phi_p(n) = \frac{a_0(n)}{p} + \frac{a_1(n)}{p^2} + \cdots + \frac{a_I(n)}{p^{I+1}} = \sum_{i=0}^I \frac{a_i(n)}{p^{i+1}} \quad (9)$$

このランダムイゼーションの措置を施していないオリジナルなHalton列は、図2に示す様に

低次では一様であるが、高次の隣接次元では点列間に相関がみられ一様ではない。

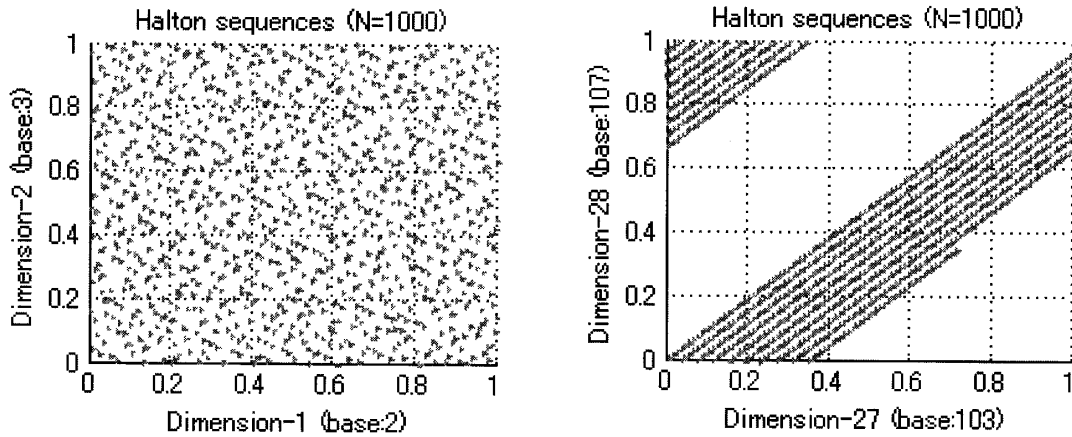


図2 オリジナルHalton列

これらの相関は、高次元で基底が大きくなることによるサイクル長の増大に由来する。しかし、この相関はディスキレパンシーの特性を保持したままで(9)式のデジット $a_i(n)$ をランダム化することにより小さくなり、一様性が改善されることが知られている。(9)式に対して、Braaten and Wellerはランダム化された基底逆関数を次のように定義した。

$$x_n = S_p(n) = \frac{\pi_p a_0(n)}{p} + \frac{\pi_p a_1(n)}{p^2} + \dots + \frac{\pi_p a_l(n)}{p^{l+1}} = \sum_{i=0}^l \frac{\pi_p a_i(n)}{p^{i+1}} \quad (10)$$

ただし、 π_p は $(0,1,2,\dots,p-1)$ の順列

具体的なランダム化の手法は、ディスキレパンシーを小さくするために $(0,1,2,\dots,p-1)$ の並び順を如何に構築するかで複数のアルゴリズムが存在している。以下では、参考文献⁵⁾に記載の3つのアルゴリズムとHongmei Chiの方法⁶⁾について簡潔に記す。

3.1 改良BW法

改良BW法は、Braaten and Wellerが³⁾提案したオリジナルのBW法（アルゴリズムは、1次元のディスキレパンシー； $T_N^*\left(\left\{\frac{\pi_p(1)}{p}, \dots, \frac{\pi_p(i)}{p}\right\}\right)$ を最小化する様に $\pi_p(i)$ を順列 $(0,1,2,\dots,p-1)$ から選択）についてVandewoestyne and Coolsが⁴⁾一部改良を加えたものである⁵⁾。基底19までを示すと表1の様である。

表1 改良BW法の順列（素数の値19まで）

$\pi_2 = (0 \ 1)$	$\pi_{11} = (0 \ 5 \ 8 \ 2 \ 10 \ 3 \ 6 \ 1 \ 9 \ 4 \ 7)$
$\pi_3 = (0 \ 2 \ 1)$	$\pi_{13} = (0 \ 6 \ 10 \ 2 \ 8 \ 4 \ 12 \ 1 \ 9 \ 5 \ 11 \ 3 \ 7)$
$\pi_5 = (0 \ 2 \ 4 \ 1 \ 3)$	$\pi_{17} = (0 \ 8 \ 13 \ 3 \ 11 \ 5 \ 16 \ 1 \ 10 \ 7 \ 14 \ 4 \ 12 \ 2 \ 15 \ 6 \ 9)$
$\pi_7 = (0 \ 3 \ 5 \ 1 \ 6 \ 2 \ 4)$	$\pi_{19} = (0 \ 9 \ 14 \ 3 \ 17 \ 6 \ 11 \ 1 \ 15 \ 7 \ 12 \ 4 \ 18 \ 8 \ 2 \ 16 \ 10 \ 5 \ 13)$

3.2 Faure法

これは, Faureにより提案された順列 π_p を生成する手法で, 順列生成アルゴリズムは表2に示す通りである. また表3は, このアルゴリズムに基づき得た順列であるが, アルゴリズムとの対応関係が分かり易い様に, 素数以外の p についても併記した.

表2 Faureの順列生成アルゴリズム

```

 $p = 2 ; \pi_p = (0,1)$ 
repeat
   $p = p + 2$ 
  if  $p$  is even then
     $\pi_p = (2\pi_{\frac{p}{2}}, 2\pi_{\frac{p}{2}} + 1)$ 
  else
     $\eta = \pi_{p-1} ; k = \frac{p-1}{2}$ 
    add 1 to each element of  $\eta$  which is  $\geq k$ 
     $\pi_p = \eta$  with  $k$  added in the middle
  end if
until all necessary permutation are found

```

表3 Faureによる順列

$\pi_2 = (0\ 1)$	$\pi_9 = (0\ 5\ 2\ 7\ 4\ 1\ 6\ 3\ 8)$
$\pi_3 = (0\ 1\ 2)$	$\pi_{10} = (0\ 6\ 4\ 2\ 8\ 1\ 7\ 5\ 3\ 9)$
$\pi_4 = (0\ 2\ 1\ 3)$	$\pi_{11} = (0\ 7\ 4\ 2\ 9\ 5\ 1\ 8\ 6\ 3\ 10)$
$\pi_5 = (0\ 3\ 2\ 1\ 4)$	$\pi_{12} = (0\ 4\ 8\ 2\ 6\ 10\ 1\ 5\ 9\ 3\ 7\ 11)$
$\pi_6 = (0\ 2\ 4\ 1\ 3\ 5)$	$\pi_{13} = (0\ 4\ 9\ 2\ 7\ 11\ 6\ 1\ 5\ 10\ 3\ 8\ 12)$
$\pi_7 = (0\ 2\ 5\ 3\ 1\ 4\ 6)$	$\pi_{14} = (0\ 4\ 10\ 6\ 2\ 8\ 12\ 1\ 5\ 11\ 7\ 3\ 9\ 13)$
$\pi_8 = (0\ 4\ 2\ 6\ 1\ 5\ 3\ 7)$	$\pi_{15} = (0\ 4\ 11\ 6\ 2\ 9\ 13\ 7\ 1\ 5\ 12\ 8\ 3\ 10\ 14)$

3.3 Reverse法およびHongmei Chiによる方法

Reverse法は, Vandewoestyne and Coolsが提案している方法⁵⁾で, π_p として順列 $(0,1,2,\dots,p-1)$ を文字通り逆順に $(0,p-1,\dots,2,1)$ としているシンプルな順列である.

一方, Hongmei Chiによる方法は, π_p を選択するために先ず線形の順列関数

$$f(x) = wx + c \pmod{p_i}$$

を仮定し, 具体的には $c = 0$ とした次式を用いて最適な重み w_i を求めるとしている.

$$\pi_{p_i}(a_j) = w_i a_j \pmod{p_i} \quad \text{ここに, } 1 \leq w_i \leq p_i - 1, \quad 0 \leq j \leq m$$

しかし, 最初の40次元に対する最適な重み w_i は参考文献⁶⁾表3.1に掲載されているが, w_i の具体的誘導法については明示されていない.

4 一般化Halton列の比較

図3は、スタート値 $n_{start} = 1$ としたランダム化なしのHalton列の一例である。それに対して図4は、上記で説明したランダム化を施した一般化Halton列である。一般化することにより、アルゴリズムの違いによって程度の差は見受けられるが、相関性が弱まり一様性が改善されていることが分かる。

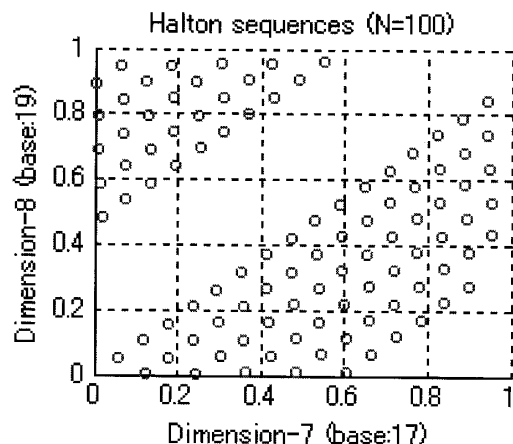


図3 unscrambling Halton sequence

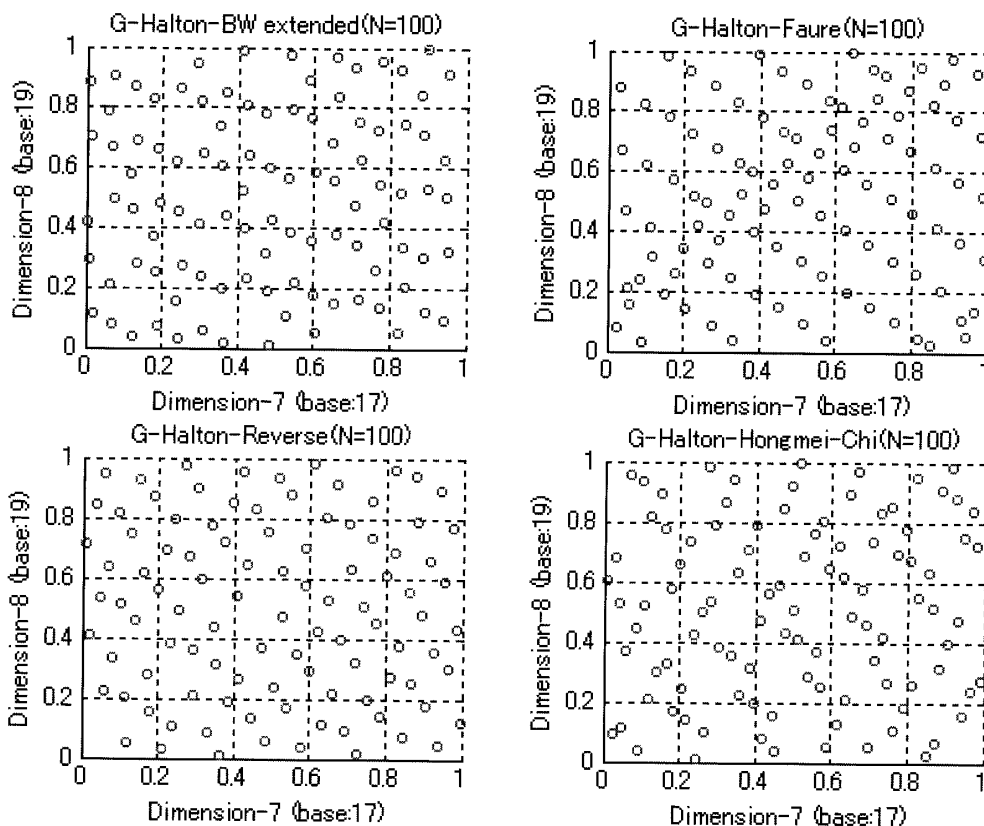


図4 scrambling Halton sequence

図5は、 L_2 スターディスクレパンシーを指標として一様性の改善を定量的に表したもので、図中のRMSディスクレパンシーの曲線は(7)式により描かれ、その他の折れ線グラフは(6)式の $T_N^*(p)$ を計算して描かれている。図から、一般化されていないunscrambling HaltonはRMSディスクレパンシーよりサンプル総数700以前ではディスクレパンシーが大きく、700以降ではやや小さくなっていることが分かる。これに対して一般化すると、サンプル総数が大きくなるにつれ、いずれのディスクレパンシーもRMSディスクレパンシーより小さくなり、明らかに一様性の改善効果が認められる。ただし、図5は8次元に対する比較である。

一方、図16は同様の計算を12次元に対して行った比較を示したものである。次元が増すと、8次元の場合よりも全てのケースのディスクレパンシーが大きくなる方向にシフトすること、サンプル総数が大きくなって漸くディスクレパンシーが小さくなることが分かる。

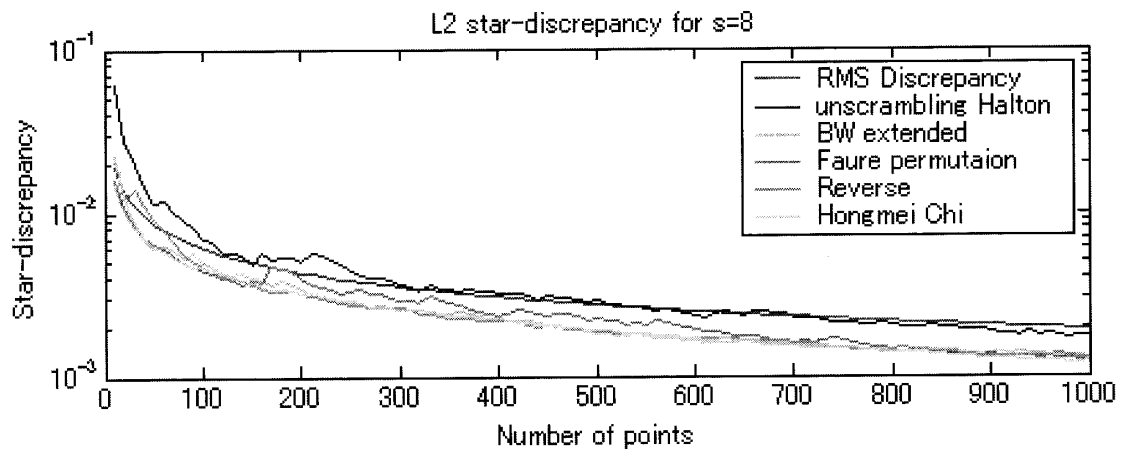


図5 L_2 -star discrepancy for 8-dimensional sequence

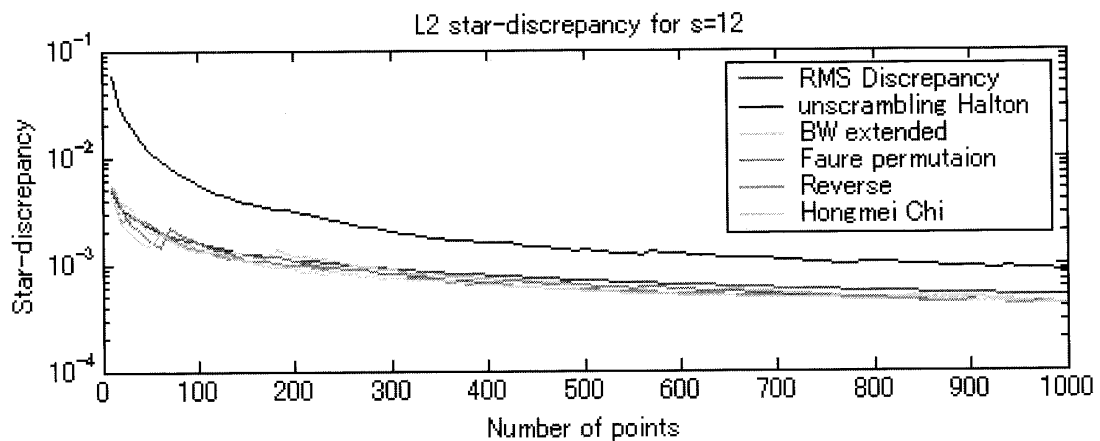


図6 L_2 -star discrepancy for 12-dimensional sequence

5 Faure列のランダムイゼーション

以下に比較の対象として取り挙げた，Faure列のランダムイゼーション構築法について提示する．

5.1 原始根を用いる方法（白川法）

白川は，(10) 式の π_p のランダムイズ順列として，原始根 q を利用することを提案し， $p_i = q^i \pmod{p}$ ，ただし $i = 1, 2, 3, \dots$ を対角要素とするランダムイズ項を生成マトリックス(generator matrix)に左から乗じて，一般化Faure列を次式の様に置いた⁷⁾．

$$\begin{pmatrix} a_0^{(1)} \\ a_1^{(1)} \\ \vdots \\ a_I^{(1)} \end{pmatrix} = \begin{bmatrix} p_i & 0 & 0 & 0 \\ & p_i & 0 & 0 \\ & & \ddots & 0 \\ sym. & & & p_i \end{bmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_I \end{pmatrix} \pmod{p} \quad \text{for } i = 1 \quad (11)$$

$$\begin{pmatrix} a_0^{(1)} \\ a_1^{(1)} \\ \vdots \\ a_I^{(1)} \end{pmatrix} = \begin{bmatrix} p_i & 0 & 0 & 0 \\ & p_i & 0 & 0 \\ & & \ddots & 0 \\ sym. & & & p_i \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & \dots \\ 0 & 1 & 2 & \dots \\ 0 & 0 & 1 & \dots \\ 0 & 0 & 0 & \ddots \end{bmatrix}^{i-1} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_I \end{pmatrix} \pmod{p} \quad \text{for } i \geq 2 \quad (12)$$

素数基底 p に関する原始根 q を適当に選択し，その i 乗の値について p を法とする剰余は， $1 \sim p-1$ 乗する間に $1 \sim p-1$ の数が重複なく一見ランダムな順番で丁度1回現れる．

なお，原始根は素数に対して最低1個以上，通常複数あり，一様性の評価は原始根の選択に依存するが，後述する比較においては最小の原始根を使用した．

5.2 Reverse法およびHongmei Chiによる方法

Vandewoestyne and Coolsは，Faure列に関してReverse法の適用を試みていないが，Halton列と同様の考え方で簡単に適用できる．

一方，Hongmei Chiはランダムイズ項として，Tezuka⁴⁾のI-binomial scramblingに準じた次式の行列を提案した⁶⁾．

$$A^{(j)} = \begin{bmatrix} h_1 & 0 & 0 & \dots \\ g_2 & h_1 & 0 & \dots \\ 0 & g_2 & h_1 & \dots \\ 0 & 0 & \ddots & \ddots \end{bmatrix} \quad (13)$$

ただし，Halton列同様，Faure列でも具体的な上記行列要素の誘導法を明示しておらず，基底101に適用する場合についてのみ， $h_1 = 28$ ， $g_2 = 83$ であるとしている．

6 一般化Faure列の比較

上記 (13) 式のただし書きの制約により、以下では素数基底101とした100次元迄のFaure列の比較に限定する。

図7左上図に示す様に、ランダムイゼーションされていない場合には、明らかに隣接次元の点列間に著しい相関が生じている。それに対して、いずれの一般化手法も一様性の改善効果が認められる。なお、下図の表示は、次元97-98の場合であるが、表示次元を違えた場合にはいずれの一般化手法も改善効果が認められるものの見掛けが劣る場合があった。3者の中では、概ねHongmei Chiの方法による結果が良好であったが、白川の原始根を利用する方法とReverse法とでは差が認められない場合もあった。この比較では原始根の恣意的な選択を避けるために最小の原始根2を用いているが、異なる原始根を用いる場合には上記の傾向と異なることも予想される。

図8は、 L_2 スターディスクレパンシーの相互比較である。同図から、ランダムイゼーションなしのFaure列は次元が大きいこともあり、極端にRMSディスクレパンシーより大きい。それに対して、一般化されたFaure列のディスクレパンシーはいずれもRMSのそれより小さい結果が得られた。ただし、図7で見られた様な3者間の差はディスクレパンシーには殆んど反映されておらず、僅かにHongmei Chiの方法によるディスクレパンシーが小さいだけである。

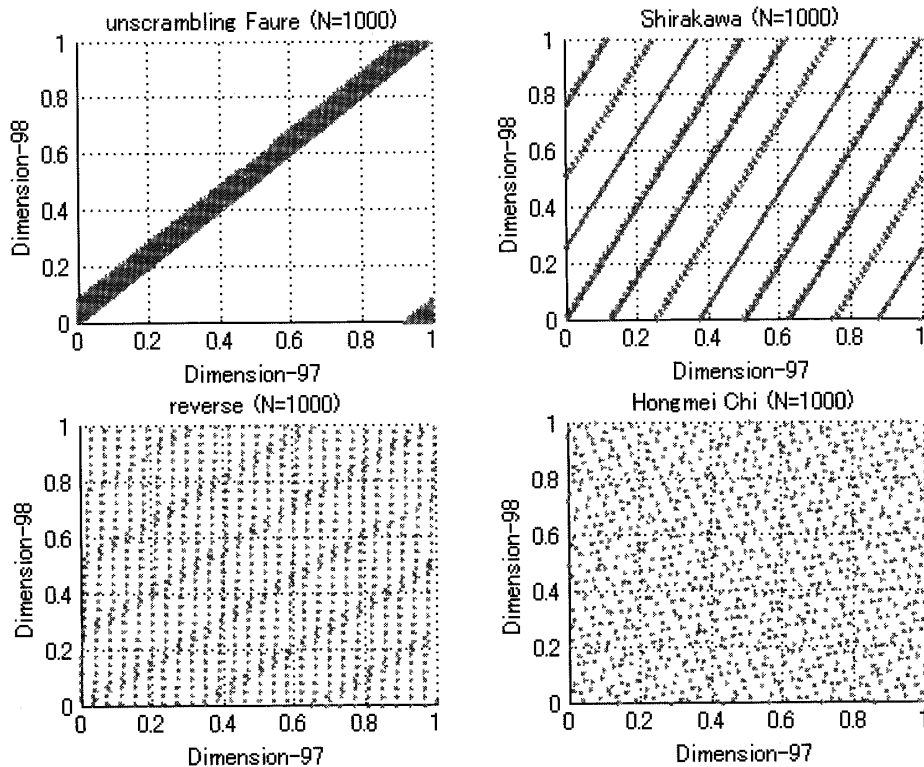


図7 scrambling Faure sequence

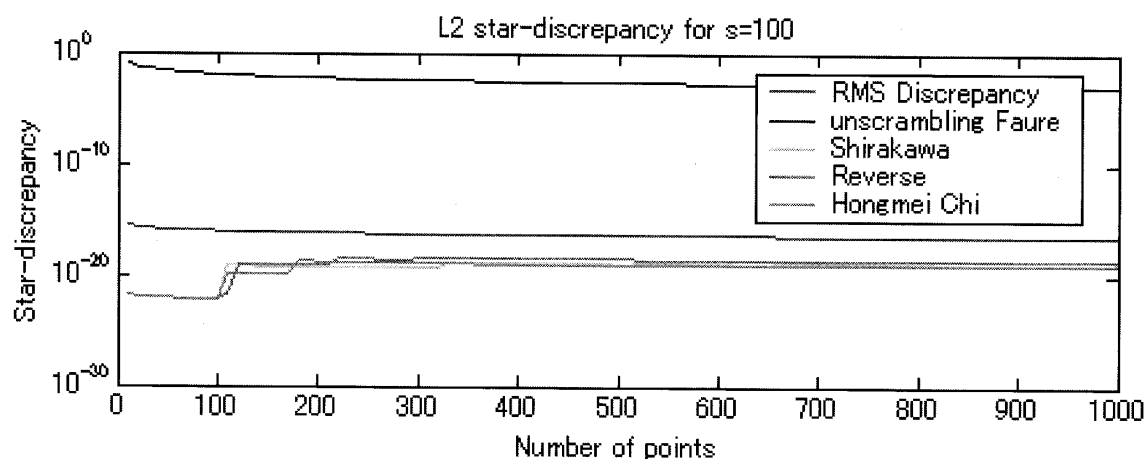


図8 L_2 -star discrepancy for 100-dimensional sequence

7 結語

超一様数列のランダム化について、経験的に得られた幾つかの手法を紹介した。現在、一般的で合理的な数学的意味も踏まえた超一様分布列、ランダム化の取り扱いが盛んに研究されている。(t,m,s) ネットおよびOwenのランダム化をはじめとするよりスマートな取り扱いについては、今後の課題としたい。

参考文献

- 1) 串山 繁：超一様分布列の生成方法，北海学園大学大学院工学研究科紀要 工学研究第5号，pp.93–103，2005年10月。
- 2) C. Lecot and B. Tuffin：Comparison of Quasi-Monte Carlo-Based Methods for the Simulation of Markov Chains, Monte Carlo Methods and Applications, 10(3–4), pp.377–384, 2004.
- 3) E. Thiemard：An Algorithm to Compute Bounds for the Star Discrepancy, Journal of Complexity 17, pp.850–880, 2001.
- 4) S. Tezuka, Uniform Random Numbers：Theory and Practice, Kluwer Academic Publishers, 1995.
- 5) B. Vandewoestyne and R. Cools：Good permutation for scrambled Halton sequences in terms of L_2 -discrepancy, Dept. of Computer Science, K. U. Leuven, Report TW406, September 2004.
- 6) Hongmei Chi：Scrambled Quasirandom Sequences and Their Applications, Phd. Thesis, the Florida State University, 2004.
- 7) 田村勉，白川浩：一般化Faure列による準乱数とそのオプション評価への応用，
http://www.craft.titech.ac.jp/s_lab/download/paper.html.