Available online at http://scik.org
Eng. Math. Lett. 2014, 2014:20
ISSN: 2049-9337

ON THE EFFICIENCY OF PARTITION PSEUDO RANDOM NUMBER GENERATED IN INTEGRAL ESTIMATION

BEHROUZ FATHI-VAJARGAH AND ALI. A. L_ZADEH*

Department of Statistics, University of Guilan, Rasht, Iran

Abstract. We know the Monte Carlo method for high number of samples is an unbiased estimator, as well as converges with slowly speed. To address this problem, quasi- Monte Carlo methods were introduced. Quasi-Monte Carlo method has good efficiency and strong convergence, but as the dimension of the problem increases, the advantage of quasi-Monte Carlo method quickly decreases, and the error of quasi-Monte Carlo increases, so. To solve this problem, we use the hybrid Monte Carlo method. Hybrid Monte Carlo is composed of hybrid sequence. Hybrid sequence is a combination of Monte Carlo, quasi-Monte Carlo and randomize quasi-Monte Carlo. In this we use hybrid Monte Carlo method using Halton sequence based on partitioning the interval Monte Carlo method using Halton sequence based on partitioning the interval $(0,1)$ to k subintervals and control the number to follow must possible informity on $(0,1)$. Also, as an alternative, we use scramble on Halton sequence to increase the unifotmity on the all areas of desired sections whole parts.

Keywords: Monte Carlo, partition pseudo random number generated, hybrid method, quasi-Monte Carlo.

2010 AMS Subject Classification: 65C05.

Introduction

Quasi-Monte Carlo (QMC) methods are the deterministic version of Monte Carlo (MC) methods [11]. Halton sequence is a popular class of low-discrepancy sequences, and one of its important advantages is that the Halton sequence is easy to implement due to its definition via the radical inverse function [1]. This low-discrepancy sequence was introduced by Halton [6]. The Halton

*Corresponding author

Received April 26, 2014

sequence [1] is a general s-dimensional sequence in the unit hypercube $[0,1]^s$. The first dimension of the Halton sequence is the Van Der Corput sequence base 2 and the second dimension is the sequence using base 3. Dimension *s* of the Halton sequence is the sequence using the s^{th} prime number as the base. simplest hybrid sequence has been combined of random sequence and low-discrepancy sequence, such that some of researchers worked in this issue, for instance [2,9,10].

In this paper, in the first section we introduce Monte Carlo method and the new method. We call this method PMC or p-rand and based on this method, partition random numbers are generated. And we will express the corresponding algorithm for this method. In the second section we express low-discrepancy sequences and explain about Halton sequence. In the third section, we will explain about mixed sequence. And the last section we will show numerical results for estimate integral.

1. Monte Carlo Method (MC)

Monte Carlo method is an analytical technique for solving a problem by performing a large number of trial runs, called simulations, and inferring a solution from the collective results of the trial runs. Monte Carlo integration is to use random points for the numerical evaluation of an integral [15].

$$I = \int_a^b f(x)dx \approx \frac{b-a}{N}\sum\nolimits_{i=1}^{N} f(x_i)$$

in this method we approximate I by taking random variables X_i and arithmetic averaging by contribution $f(x_i)$. In the special case of the above equation is integration bounded on $[0,1)$,

$$I = \int_0^1 f(x)dx \approx \frac{1}{N}\sum\nolimits_{i=1}^{N} f(x_i).$$

The base of Monte Carlo method is to generate independent identically distributed random variables on $(0,1)$, which the interval number uniformly distributed on $[0,1)$ to generate these distributed random variables, we employ the rand function on Matlab programming software. Estimated results of the Monte Carlo integral are reliable, however we can significantly improve the results by managing and partitioning of the random numbers generated that is known as p-rand or PMC method.

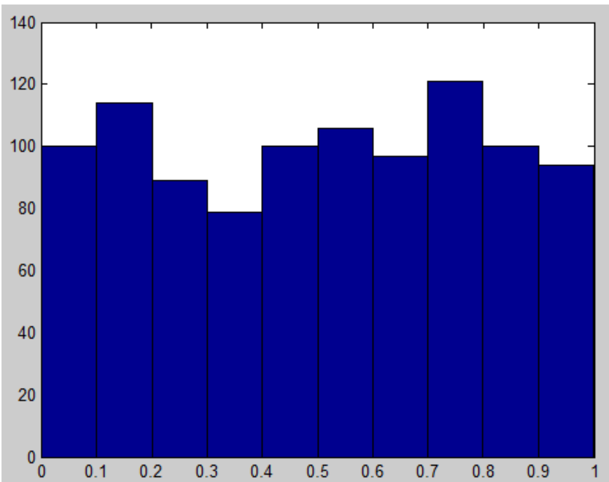
Algorithm of the method:

Step 1. Choose N as the number of random numbers.

Step 2. Choose number of sample points.

Step 3. Choose N/K number of sub-interval's j, with $j = 1,2, \dots, K$, where K is the a number of subintervals.

Results are shown separately in two histograms.



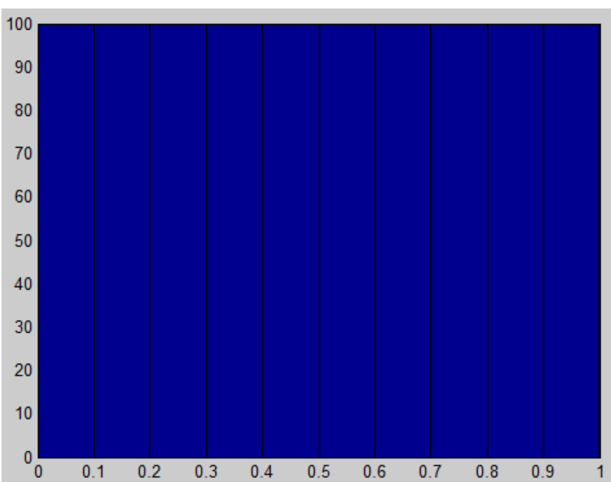


Fig 1. Histogram for N=1000, left histogram using the rand-function, and right histogram partitioning by use of the rand-function in Matlab programming software.

2. Quasi-Monte Carlo Method (QMC)

Monte Carlo method is used for integrals estimating, solving linear equations, finding eigenvalues of the matrix and etc, have revealed some demerits including lack of efficiency, low convergence speed, high radius of convergence. So, after so many achievements in this issue, eventually the Quasi-Monte Carlo methods has been proposed instead of the Monte Carlo method for covering all of the before mentioned inefficiencies.

Briefly, in QMC the following relation is hold;

$$I = \int_{I^s} f(\boldsymbol{U}) d\boldsymbol{U} \approx \frac{1}{N}\sum_{i=1}^{N} f(\boldsymbol{X}_i).$$

QMC methods can be viewed as deterministic versions of Monte Carlo methods.

2.1 Quasi-random sequences

The quasi-Monte Carlo methods are similar to the Monte Carlo method so that this method evaluates a multi-integral by substituting it with average values of the computed integrand at the generated deterministic points. Quasi- Monte Carlo methods are based on quasi-random sequences, and usually is known as low- discrepancy sequences. The best sequences and most commonly used of these sequences are: Halton, Sobol, Faure and Niederreiter sequences [1,3,6,11]. In this paper we use the following Halton sequence.

For an integer $b \geq 2$, let $Z_b = \{0, 1, \ldots, b-1\}$ denote the least residue system modulo b. Let

$$n = \sum_{j=1}^{\infty} a_j(n) b^{j-1}$$

with all $a_j(n) \in Z_b$ and $a_j(n) = 0$ for all sufficiently large j be the unique digit expansion of the integer $n \geq 0$ in base b , [5,15].

The radical-inverse function in base b_j for $j = 1,2, \ldots, s$ is defined by

$$\varphi_{b_j}(n) = \sum_{i} \sigma(a_i(j,n) b^{-i-1})$$

where σ is the permutations on $a_i(j,n)$'s.

For pairwise coprime integers $b_1, \ldots, b_s \geq 2$, the Halton sequence in the bases $b_1, \ldots, b_s$ is given by

$$X_n = (\varphi_{b_1}(n), \ldots, \varphi_{b_s}(n)) \in [0,1)^s$$

for $n = 0,1, \ldots,$ [6].

We remind that quasi-random sequence (low discrepancy sequence) for high dimension will lost their own effectiveness; in other word the high dimensions correlation of these points will be increased and concentration will be placed around the lines, therefore for making up this demerit these points will be scrambled. In this paper we will utilize the scrambling method for Halton sequence that introduced by Kocis and Whiten in [7], and this method called " RR2".

2.2 The Koksma-Hlawka Inequality

For any sequence $\{x_n\}_{1\leq n\leq N}$ and any function f with variation in the sense of Hardy-Krause, $V\ (f)$, bounded, I have

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

where D_N^* is the star-discrepancy of point set $\{x_1, \ldots, x_N\}$.

The star discrepancy of the first n Halton points in dimension s with relatively prime bases $b_1, \ldots, b_s$ in

$$D_N^*(X) \leq C_s \frac{1}{N}(logN)^s + O(\frac{1}{N}(logN)^{s-1})$$

that C_s is coefficient for the Halton [3,4,13],

$$C_s = \prod_{j=1}^{s} \frac{p_j - 1}{2logp_j}\ .$$

In order to minimize error, we should be reduce $D_N^*(X)$ or $V(f)$**.**

3. Hybrid Monte Carlo Method

In problem of moderate dimensions, the quasi-Monte Carlo method usually provides better estimate than the Monte Carlo method. However, as the dimension of the problem increase, the advantage the quasi-Monte Carlo methods diminish quickly. Some researchers have also proposed a method to solve this problem, and it called mixed method [14].

In this method we will disuse about methods that have been named mixed method. Mixed method is made up mixed sequence. Mixed sequence is combined MC sequence (or PMC sequence) by randomized (scrambled) QMC sequence (RQMC), [14]. Padding PMC sequence by randomized QMC (RQMC).

We are consider the problem of estimate the multidimensional integration

$$I = I(f) = \int_{I^s} f(x)dx = \int_0^1 \ldots \int_0^1 f(x_1, \ldots, x_s)dx_1, \ldots, dx_s\ .$$

According to the Monte Carlo method [15],

$$\hat{I} = \frac{1}{N}\sum_{k=1}^{N} f(x^{(k)})$$

where $x^{(k)}$ are s-dimensional vectors chosen adaptable, where $x^{(k)}$ for $k = 1,2,\ldots,N$, are sequences of variables s-dimensional. Let $(x_1,\ldots,x_{\mathrm{d}})$ be d-dimensional subset of variables $(x_1,\ldots,x_s)$, for $d \leq s$.

Then one has the following points [14]:

1- Sample $(x_1,\ldots,x_d)$ using a d-dimensional MC sequence and for $(x_{d+1},\ldots,x_s)$ of the variables use (s-d)-dimensional partition low-discrepancy sequence (call mixed method, padding MC by PQMC);

2- Sample $(x_1,\ldots,x_d)$ using a d-dimensional partition Monte Carlo (p-rand) sequence and for $(x_{d+1},\ldots,x_s)$ of the variables use $(s-d)$-dimensional low-discrepancy sequence (call mixed method, padding PMC by QMC);

3- Sample $(x_1,\ldots,x_d)$ using a d-dimensional MC sequence and for $(x_{d+1},\ldots,x_s)$ of the variables use $(s-d)$-dimensional RQMC sequence (call mixed method, padding MC by RQMC);

4- Sample $(x_1,\ldots,x_d)$ using a d-dimensional Partition Monte Carlo (p-rand) sequence and for $(x_{\mathrm{d}+1},\ldots,x_{\mathrm{s}})$ of the variables use $(s-d)$-dimensional RQMC sequence (call mixed method, padding PMC by RQMC).

Here $x^{(k)} = (X^{(k)}, q^{(k)})$ be a s-dimensional sequence by concatenating by vectors $X^{(k)}$ and $q^{(k)}$. In the first strategy, $X^{(k)}$,$k \geq 1$, are independent random variables with the uniform distribution on $(0,1)^d$, and $(q^{(k)})_{k\geq 1}$ is a $(s-d)$-dimensional low-discrepancy sequence. In the second strategy, $X^{(k)}$,$k \geq 1$, are partitioned independent random variables with the uniform distribution on $(0,1)^d$, and $(q^{(k)})_{k\geq 1}$ is similar to the first strategy. In the third strategy, $X^{(k)}$,$k \geq 1$, are independent random variablres with the uniform distribution on $(0,1)^d$, and $(q^{(k)})_{k\geq 1}$ is a $(s-d)$-dimensional randomized Quasi-Monte Carlo (RQMC) sequence. Last one , $X^{(k)}$,$k \geq 1$, are similar to the second strategy, and $(q^{(k)})_{k\geq 1}$ is similar to the third strategy. All of the above defined sequences are called mixed sequences.

In this section we compute the star discrepancy for hybrid sequence and low-discrepancy sequence (Halton sequence). We note that, if the dimension of deterministic section ($(s-d)$-dimension of low-discrepancy sequence) be equal of s, then mixed sequence conversion to low-discrepancy sequence, also star discrepancy for hybrid Monte Carlo of the following equation is obtained [13].

Theorem Let $x^{(k)} = \left(X^{(k)}, q^{(k)}\right)$ be a s-dimensional mixed sequence, where $\left\{q^{(k)}\right\}_{k=1}^{\infty}$ be a $(s-d)$-dimensional low-discrepancy sequence with

$$D_N^*\left(q^{(k)}\right) \leq C_{(s-d)} \frac{1}{N} (logN)^{s-d} + O(\frac{1}{N} (logN)^{s-d-1})$$

and $X^{(k)}$ is a random variable with the uniform distribution on $(0,1)^d$. Then, for sufficiently large N, and for $0 < a < 1$ and the star discrepancy for mixed (s, d) sequence satisfies [13],

$$D_N^*\left(x^{(k)}\right) < \frac{1}{N^{a/2}} + C_{(s-d)} \frac{1}{N} (logN)^{s-d} + O(\frac{1}{N} (logN)^{s-d-1})$$

this star discrepancy with probability greater than or equal to $1 - 2e^{-2N^{1-a}}$.

In other words

$$P(D_N^*\left(x^{(k)}\right) < \frac{1}{N^{a/2}} + C_{(s-d)} \frac{1}{N} (logN)^{s-d} + O(\frac{1}{N} (logN)^{s-d-1})) \geq 1 - 2e^{-2N^{1-a}}$$

proof in $\ddot{O}kten$ et al [14].

Table 1

Bounds for the discrepancy

s	$d = s/2$	D_1^*	D_2^*
4	2	5.167×10^{-3}	3.655×10^{-4}
6	3	1.175×10^{-3}	6.855×10^{-4}
8	4	2.670×10^{2}	5.51×10^{-3}
10	5	6.070×10^{4}	7.825×10^{-2}

In the table, we compute D_1^* and D_2^*, that the D_1^* is lower bound the upper bound for the discrepancy of the s-dimensional Halton sequence, and D_2^* be the probabilistic upper bound for the corresponding mixed (s, d) sequence. The best values for C_d, $2 \leq d \leq 10$ for the Halton sequence, and corresponding mixed (s, d), sequence. In this table using three digits rounding, and constant values $N = 10^7$, $a = 0.99$, $d = s/2$, and $s = 4,6, \dots ,10$.

The numerical results are becoming clearer for them in drawing together the team apart from the figures.

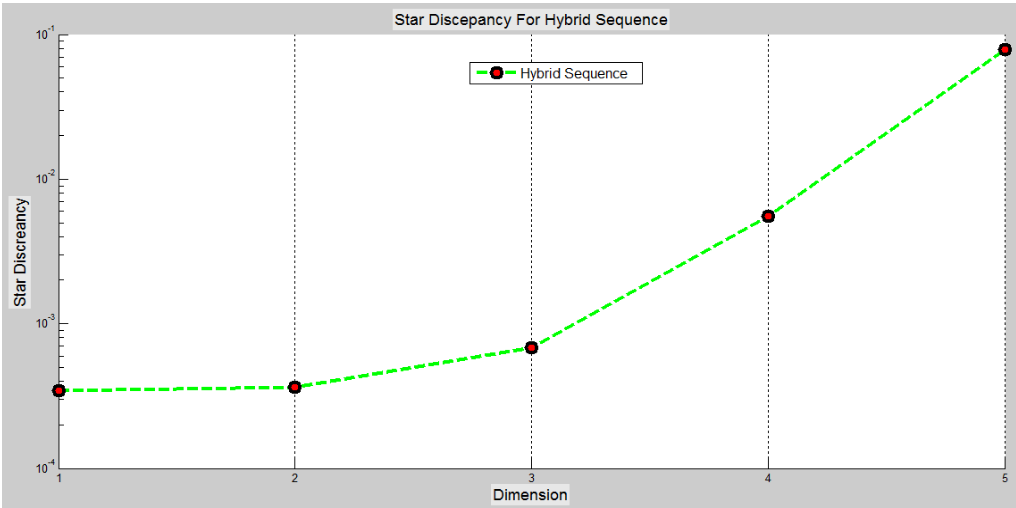


Fig 3. Star discrepancy for hybrid sequence padding MC by Halton

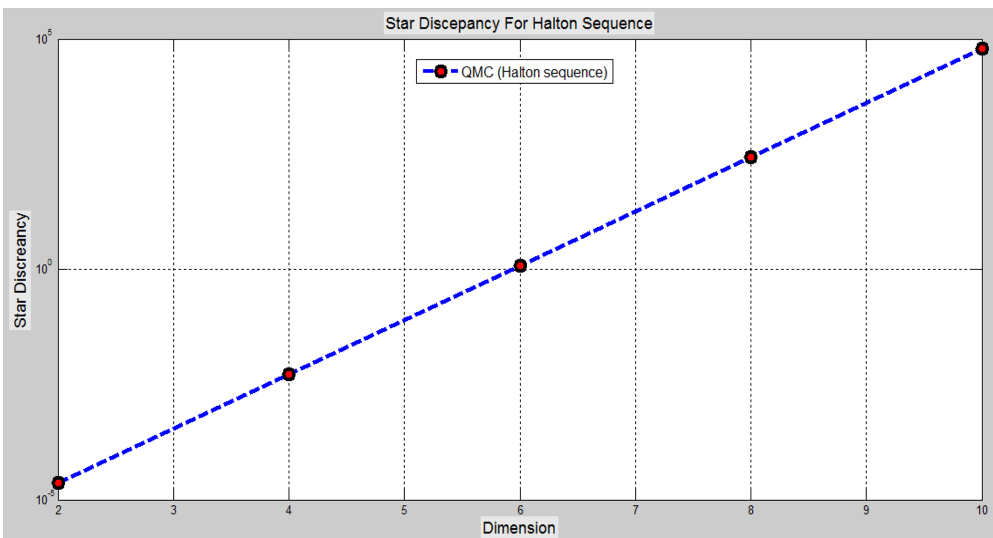


Fig 4. Star discrepancy for Halton sequence.

The efficiency of the hybrid sequence is illustrated by the following figures.

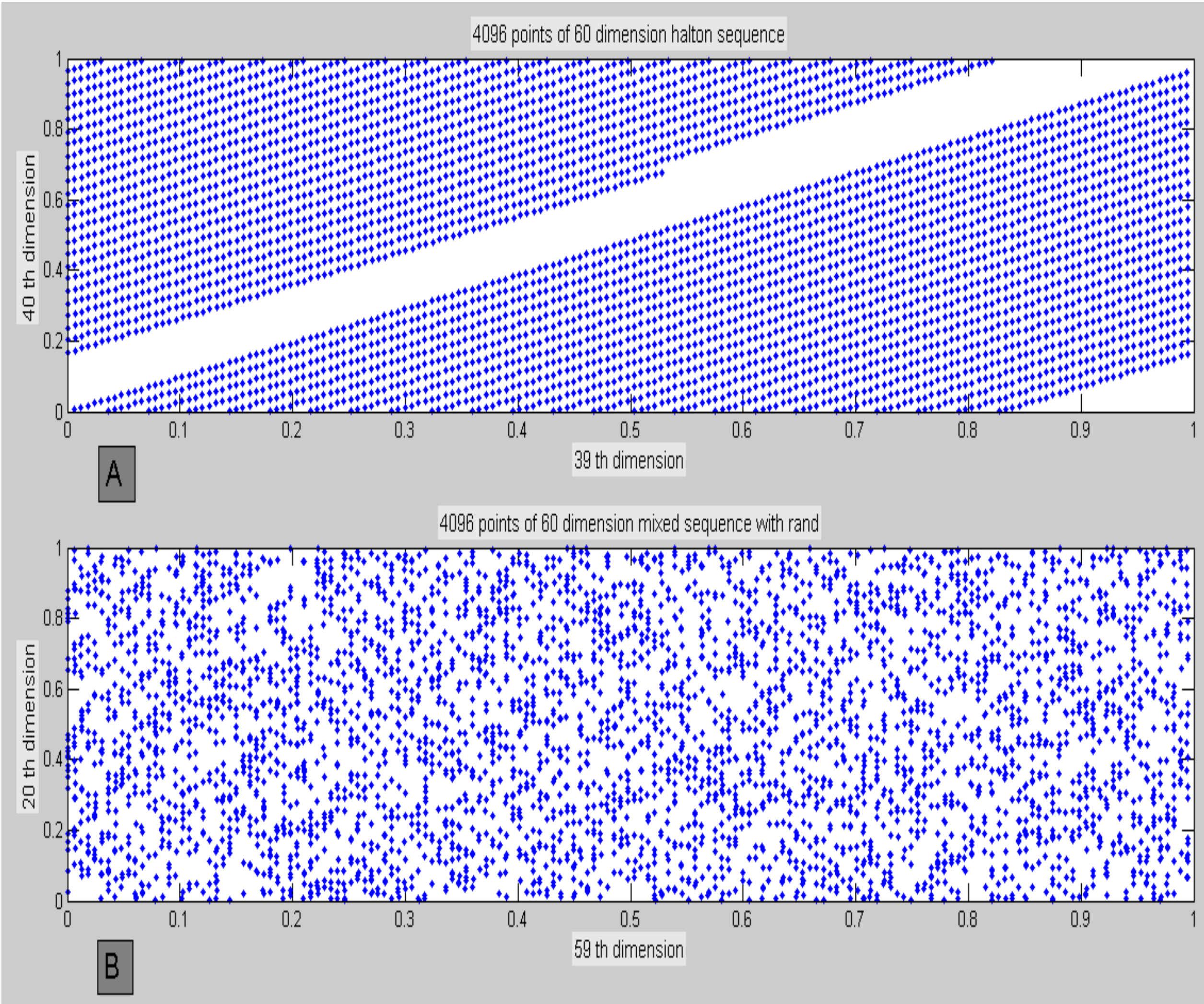


Fig 5. *Fig 5.A* ; represents the point sets for 60-dimensional Halton sequence which stand for bases of 167 and 173 with N = 4096. And *fig 5.B*; is allotted to the point sets for 60-dimensional mixed sequence, with X-axis Halton sequence for base of 167 and Y-axis random numbers generated in which N = 4096.

In *figure 5.A* the point sets of 60-dimensional Halton sequence for 39^{th} and 40^{th} dimension for the bases of 167 and 173 in which N = 4096 has been displayed. Also, *figure 5.B* demonstrated the point sets of 60-dimensional mixed sequence, where X-axis stand for Halton sequence for the base 167 and Y-axis stand for 20^{th} dimension random numbers generated where N = 4096. It is plainly visible that in *figure 5.A* correlation between points set of the Halton sequence for high dimensional is strong. *Figure 5.B* illustrates uniformity of the point sets of the mixed sequence utilized in hybrid Monte Carlo method in comparison with the points set of Halton sequence in *figure 5.A.* In other word, by using the mixed sequence correlation between the points set of Halton sequence for high dimension has been broken down [8].

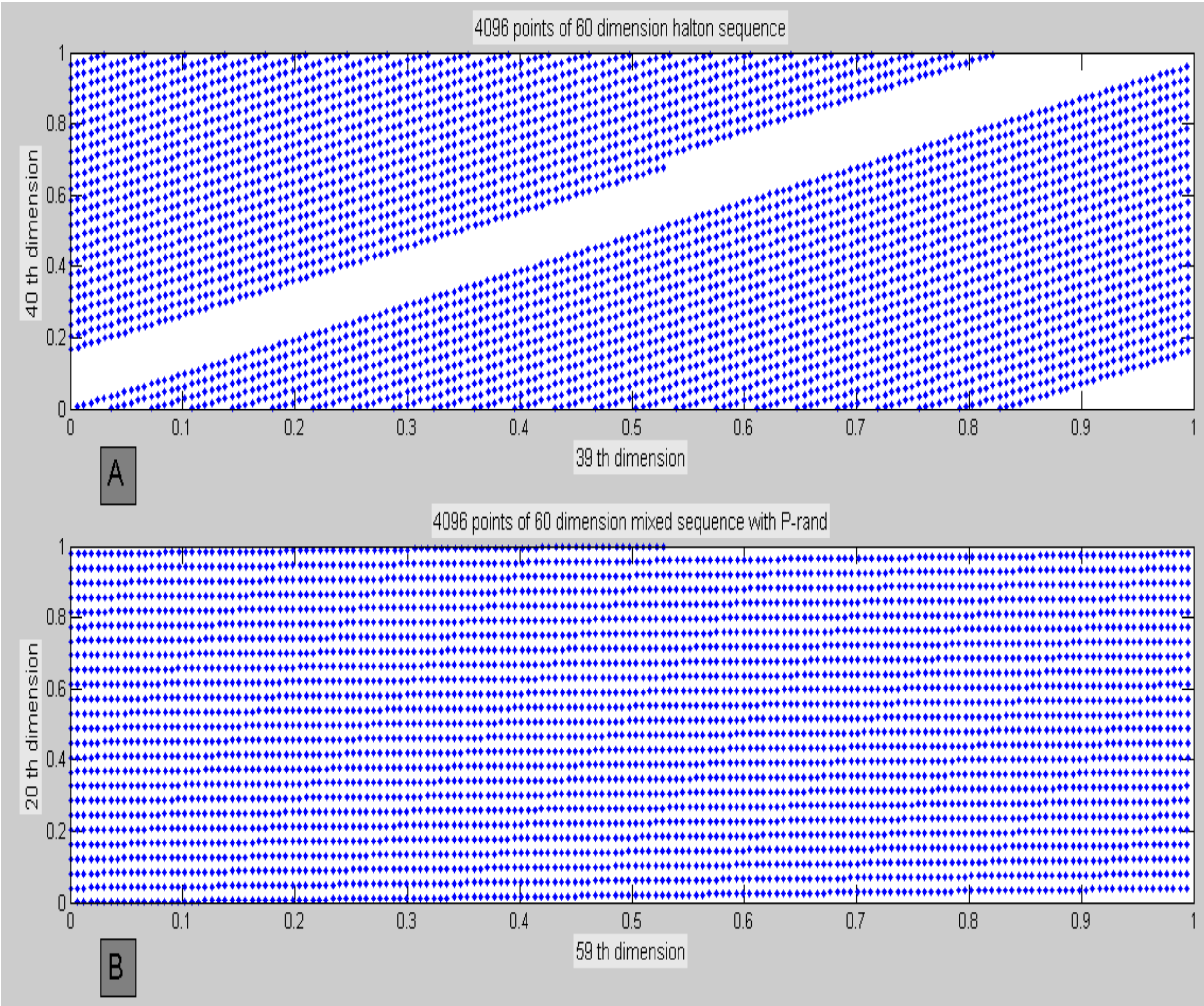


Fig 6. *Fig 6. A* ; represents the point sets for 60-dimensional Holton sequence which stand for bases of 167 and 173 with N = 4096. And *Fig 6. B*; is allotted to the point sets for 60-dimensional mixed sequence, with X-axis Halton sequence for base of 167 and Y-axis partitioning pseudo-random numbers generated in which N = 4096.

In *figure 6.A* the point sets of 60-dimensional Halton sequence for 39th and 40th dimension for the bases of 167 and 173 in which $N = 4096$ has been displayed. Also, *figure 6.B* demonstrated the point sets of 60-dimensional mixed sequence, where X-axis stand for Halton sequence for the base 167 and Y-axis stand for 20th dimension partitioning random numbers generated where N =

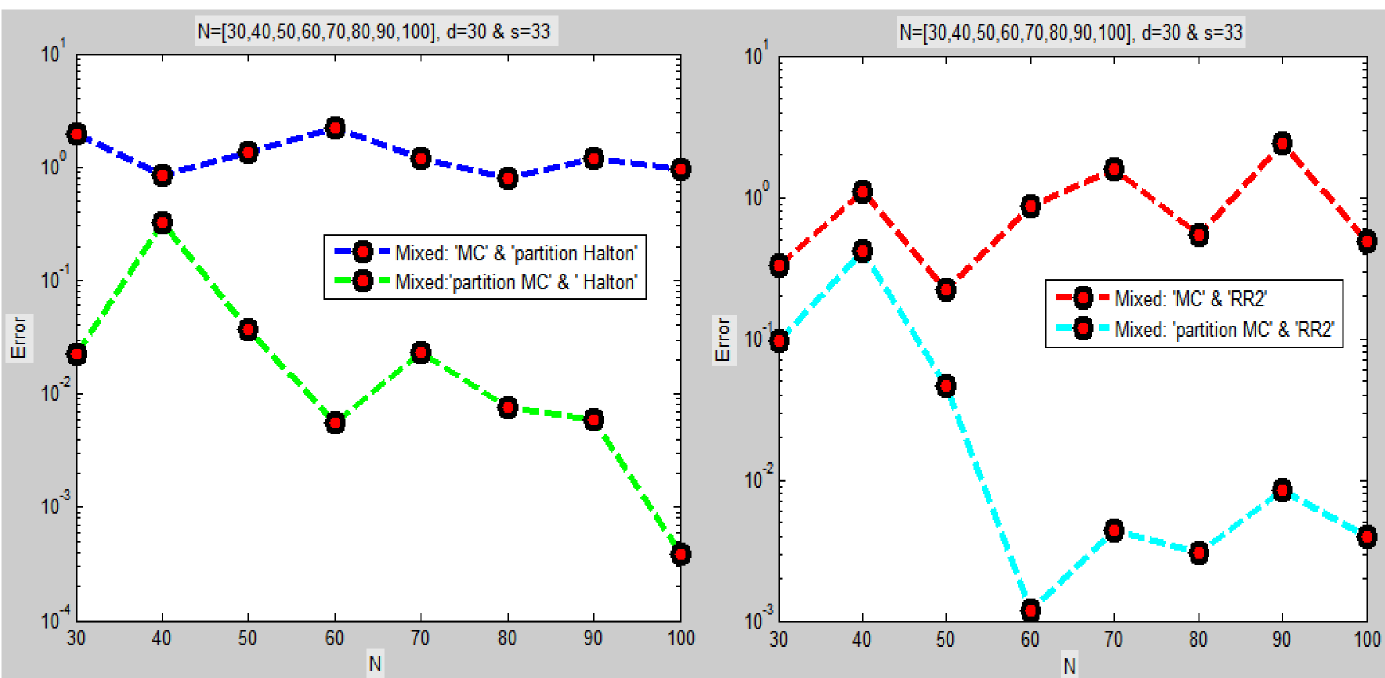
4096. It is plainly visible that in *figure 6.A* correlation between points set of the Halton sequence for high dimensional is strong. *Figure 6.B* illustrates uniformity of the point sets of the mixed sequence utilized in hybrid Monte Carlo method in comparison with the points set of Halton sequence in *figure 6.A*. In *Figure 6.B* mixed sequence is padding PMC(p-rand) by Halton. Also this figure shows that by partitioning random numbers generated lows discrepancy between these points and distributed uniformity.

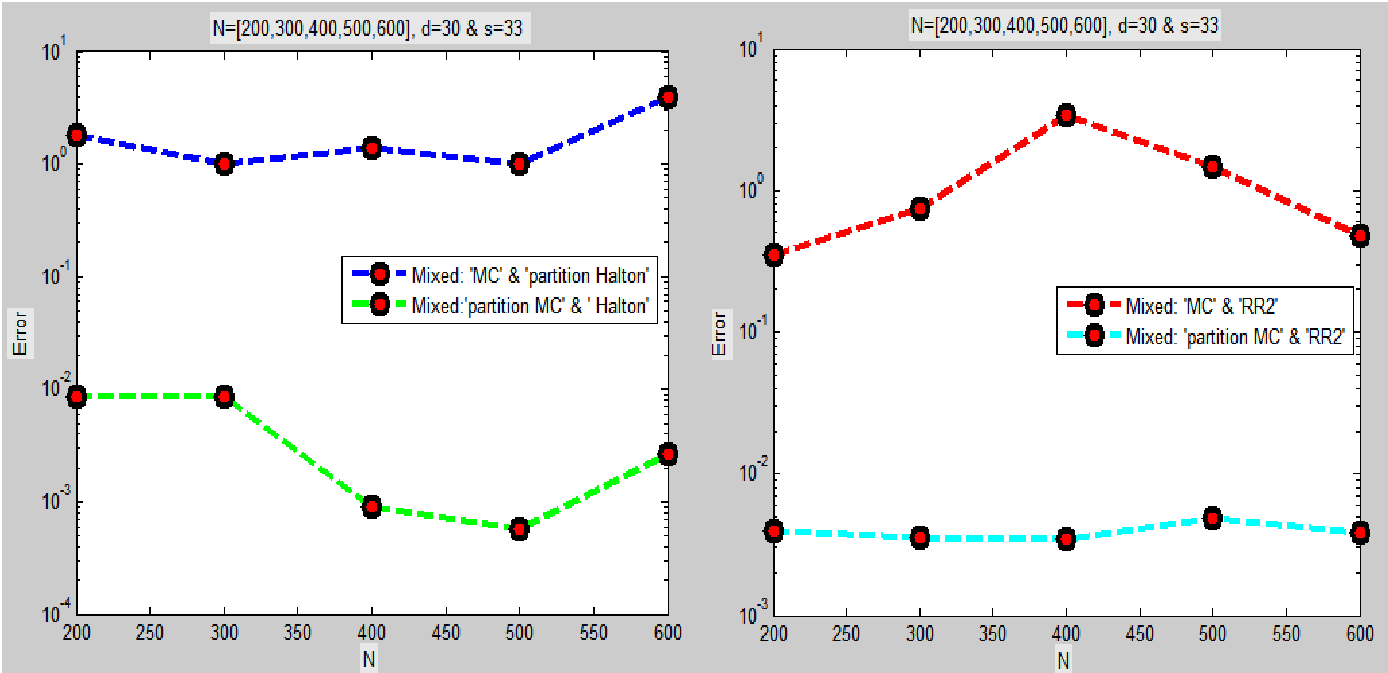
In other word, by using the mixed sequence (padding PMC (p-rand) by Halton) correlation between the points set of Halton sequence for high dimension has been broken down and these points distributed uniformity.

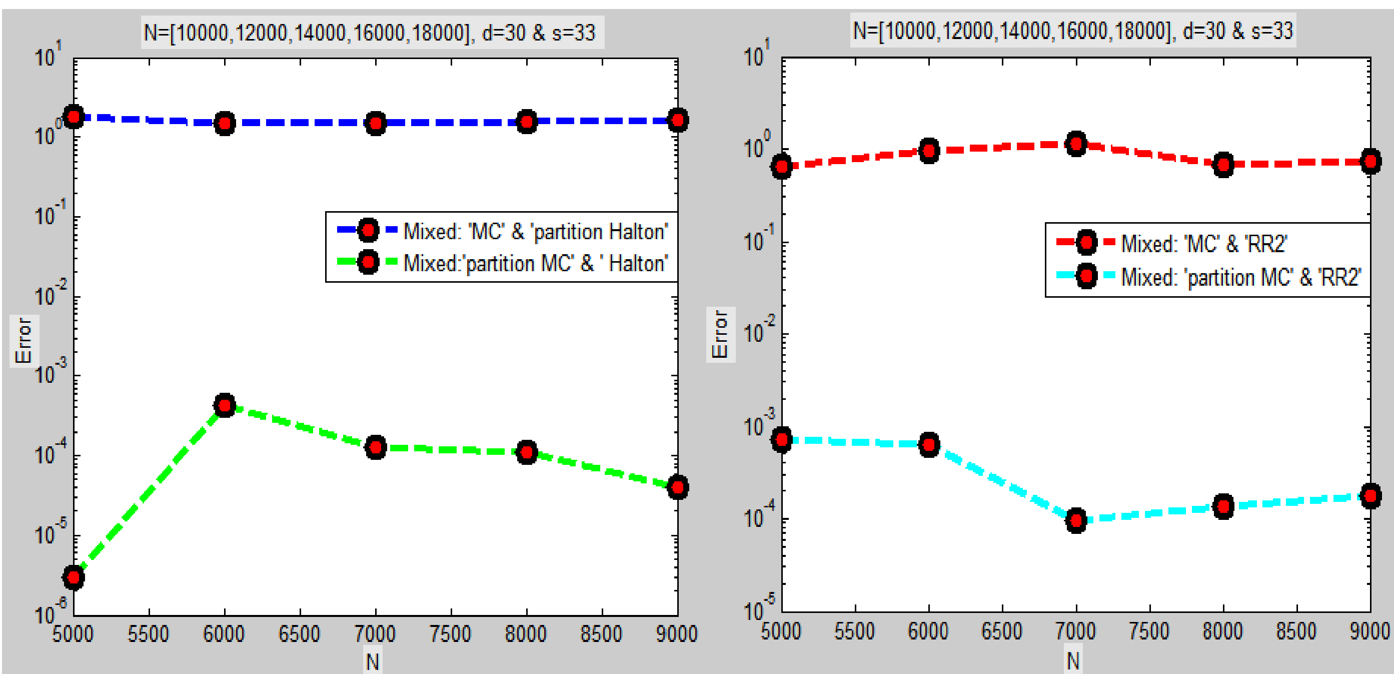
Numerical results

Also we can partition low discrepancy sequences similar to PMC. In this section we compare hybrid sequences together and testify the efficiency of the PMC sequence the following integrals will be examined: $I = \int_0^1 \dots \int_0^1 \prod_{i=1}^{s} \frac{\pi}{2} sin(\pi x_i) dx_1 \dots dx_s = 1$.

We apply the PMC for the integral I. According to the study it is concluded that if the dimension of scramble or deterministic section be equal 3, then the integral estimation error would be the lowest one.







Conflict of Interests

The author declares that there is no conflict of interests.

REFERENCES

[1] Acworth, Peter A., Mark Broadie, and Paul Glasserman. A comparison of some Monte Carlo and quasi Monte Carlo techniques for option pricing. Monte Carlo and Quasi-Monte Carlo Methods 1996. Springer New York, 1998. 1-18.

[2] Caflisch, R., Moskowitz, B., Smoothness and dimension reduction in quasi-Monte Carlo methods, Math. Comput.sModelling 23 (1996), 37–54.

[3] Chi, H., Scrambled quasirandom sequences and their application. The florida state university, college of arts and science, 2004.

[4] Domingo, G., Roswitha, H. and Niederreiter, H. A general discrepancy bound for hybrid sequences involve Halton sequence. Uniform Distribution Theory. 8 (2013), 31–45.

[5] Fox, B. Implementation and relative efficiency of quasirandom sequence generators. ACM Trans. on Mathematical Software, 12(1986), 362–376.

[6] Halton, J. On the efficiency of Certain Quasirandom Sequences of Points in Evaluating Multi Integrals, Numerisce Mathematic, 2 (1960), 84–90.

[7] Kocis, L., and W. J. Whiten. Computational Investigations of Low-Discrepancy Sequences, ACM Transactions on Mathematical Software. 23(1997), 266–294.

[8] Krykova, I. Thesis, Evaluating of path-dependent securities with low discrepancy methods. Faculty of the Worcester polytechnic institute, 2003.

[9] Matousek, J., On the L2-discrepancy for anchored boxes, J. Complex. 14 (1998) 527–556.

[10] Morokoff, William J., and Russel E. Caflisch. A quasi-Monte Carlo approach to particle simulation of the heat equation. SIAM Journal on Numerical Analysis 30 (1993), 1558-1573.

[11] Niederreiter, Harald. Random number generation and quasi-Monte Carlo methods. Vol. 63. Philadelphia: Society for Industrial and Applied mathematics, 1992.

[12] Ökten, G., A probabilistic result on the discrepancy of a hybrid-Monte Carlo sequence and applications, Monte Carlo Methods Appl. 2(1996), 255–270.

[13] Ökten, G., High dimensional simulation, Mathematics and Computers in Simulation 55 (2001), 215–222.

[14] Ökten, G., Tuffin,B., Burago,V. A central limit theorem and improved error bounds for a hybrid-Monte Carlo sequence with applications in computational finance, Journal of Complexity 22 (2006), 435 – 458.

[15] Kroese, Dirk P., Thomas Taimre, and Zdravko I. Botev. Handbook of Monte Carlo Methods. Vol. 706. John Wiley & Sons, 2011.