Communications in Statistics - Simulation and Computation

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lssp20

A New Test of Randomness for Lehmer Generators Based on the Manhattan Distance Between Pairs of Consecutive Random Numbers

Amin Khoshkenar [a] & Hashem Mahlooji [a]

[a] Department of Industrial Engineering , Sharif University of Technology , Tehran , Iran

Published online: 27 Sep 2012.

To cite this article: Amin Khoshkenar & Hashem Mahlooji (2013) A New Test of Randomness for Lehmer Generators Based on the Manhattan Distance Between Pairs of Consecutive Random Numbers, Communications in Statistics - Simulation and Computation, 42:1, 202-214, DOI: 10.1080/03610918.2011.633728

To link to this article: http://dx.doi.org/10.1080/03610918.2011.633728

Communications in Statistics—Simulation and Computation®, 42: 202–214, 2013

ISSN: 0361-0918 print/1532-4141 online
DOI: 10.1080/03610918.2011.633728

A New Test of Randomness for Lehmer Generators Based on the Manhattan Distance Between Pairs of Consecutive Random Numbers

AMIN KHOSHKENAR AND HASHEM MAHLOOJI

Department of Industrial Engineering, Sharif University of Technology, Tehran, Iran

This article considers the Marsaglia effect by proposing a new test of randomness for Lehmer random number generators. Our test is based on the Manhattan distance criterion between consecutive pairs of random numbers rather than the usually adopted Euclidian distance. We derive the theoretical distribution functions for the Manhattan distance for both overlapping (two dimensional) as well as non-overlapping cases. Extensive goodness-of-fit testing as well as empirical experimentation provides ample proof of the merits of the proposed criterion.

Keywords Lehmer random number generators; Manhattan distance; Marsaglia effect; Test of randomness.

Mathematics Subject Classification Primary 62C10; Secondary 11K45, 62G10.

1. Introduction

Random number generation is of interest in many application areas such as statistical sampling, randomized algorithms, cryptography, and particularly computer simulation (Knuth, 1997). The significance of random number generation in computer simulations is to such an extent that the validity of the generated results from highly popular Monte Carlo methods is completely dependent on the quality of the random number generator. In computational methods those generators are popular that employ algorithms with deterministic recursive formulas such as

$$\xi_i = f(\xi_{i-1}, \xi_{i-2}, \ldots, \xi_{i-q}); \quad i > q. \tag{1}$$

The deterministic recursive formula of a linear congruential random number generator is:

$$\xi_n = (a\xi_{n-1} + b) \bmod m, \tag{2}$$

Received August 15, 2011; Accepted October 14, 2011
Address correspondence to Amin Khoshkenar, Department of Industrial Engineering, Sharif University of Technology, Azadi Ave., POB: 11365-11155, Tehran, Iran; E-mail: Amin.khoshkenar@gmail.com

where ξ_n is the nth generated number (Knuth, 1997). In a multiplicative Lehmer generator, b is set equal to zero, m is known as the modulus, and a is known as the multiplier. The seed of the generator is ξ_0. The sequence of the random numbers generated by a Lehmer generator is periodic and the largest possible period length is controlled by m. If m is a prime number, ξ_0 is an arbitrary number smaller than m and $a < m$ is chosen in such a way that $\forall n < m-1 : a^n \neq 1 \bmod m$, then the period length will be maximum and equal to $m-1$ (Knuth, 1997). In order to normalize the generated numbers to the numbers with $U(0,1)$ distribution, one sets $X_i = \frac{\xi_i}{m}$; $i = 1, 2, \ldots, m-1$, where the new sequence is a permutation from the elements of the set $\left\{\frac{1}{m}, \frac{2}{m}, \ldots, \frac{m-1}{m}\right\}$.

The quality of random numbers generated by the Lehmer generator is completely dependent on the choices of m and a; a proper choice of these parameters can result in generating a stream of numbers with desirable qualities (L'Ecuyer, 1999). Until now, a rich literature has been created to identify the best choices of a and m. Fishman and Moore (1986) conducted a computer search with $m = 2^{31} - 1$ (Mersenne Prime, M_{31}) and arrived at 5 values for a that tend to maximize the generator's spectrum (Fishman and Moore, 1986). The sixth such value for a was proposed by Park and Miller (1988). They adopted the Mersenne prime in their algorithm and their generator is currently code named MINSTD in the Monte Carlo literature (Park and Miller, 1988).

Marsaglia (1968) demonstrated that if a random number stream generated by a Lehmer generator is plotted in the form of ordered overlapping (OL) n-tuples in an n-dimensional space, the overlapping n-tuples like $(1, 2, \ldots, n)$, $(2, 3, \ldots, n+1), \ldots$ will lie on at most $(n!m)^{1/n}$ hyperplanes. Such behavior is known as the Marsaglia effect which is due to the correlation between consecutive points $(X_i, \ldots, X_{i+n-1}), (X_{i+1}, \ldots, X_{i+n}), \ldots$; $i = 1, 2, \ldots, n$ (Marsaglia, 1968). He showed that this proposition is also true for the ordered Non overlapping (NOL) n-tuples like $(1, 2, \ldots, n), (n+1, \ldots, 2n), \ldots$. Figure 1 shows such effect for a two-dimensional case with $m = 401$ and $a = 12$, for OL and NOL n-tuples where $n = 2$.

It is to be noted that in a good random number generator, the statistical properties like empirical and theoretical distributions of distance between consecutive n-tuples are very close to each-other. On the other hand, on the basis of the Marsaglia effect for Lehmer generators, the points lie on a number of hyperlanes in the n-dimensional space. Now, the fewer the number of the hyperplanes and the

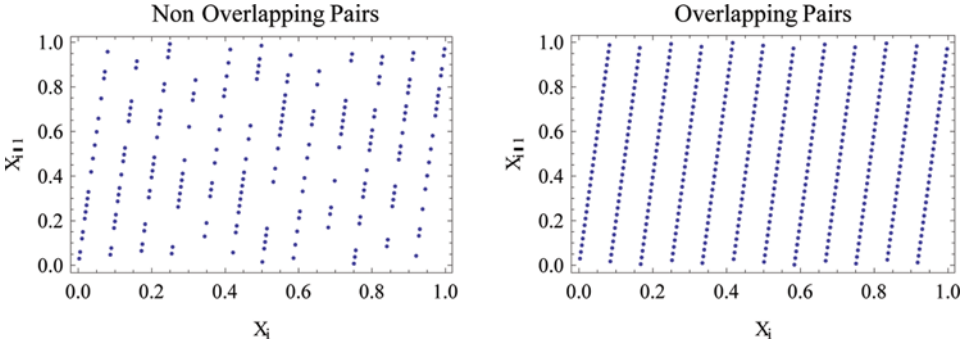

Figure 1. Marsaglia effect in two-dimensional case for NOL and OL n-tuples ($n = 2$). (color figure available online)

more closely the hyperplanes follow a simple pattern, the empirical distribution of distance between consecutive n-tuples would show a more pronounced difference with the theoretical distribution. Under such circumstances it can be concluded that the generator's quality would be more in doubt. Thus, a comparison of distance between consecutive pseudo random numbers with the distance between true random numbers can be used as an appropriate test for Lehmer generators, in the presence of the Marsaglia effect. Due to major drawbacks and disadvantages associated with the Euclidian distance in this context, we propose the Manhattan distance as a substitute measure in evaluation of the Lehmer generators. Two of the advantages of adopting the Manhattan distance are as follows.

1. Analytical derivation of theoretical distribution for Manhattan distance is considerably easier than for the Euclidean distance. While such derivation for Manhattan distance can be easily done for higher dimensions, the theoretical distribution in Euclidean distance has been carried out just up to two dimensions. In higher dimensions these calculations for Euclidean distance tend to get exponentially complicated (Dugan et al., 2005).
2. Computer programs for processing hypothesis tests on the basis of Manhattan distance distribution have lower computational complexity. In fact, this is mainly due to the simpler form of theoretical distribution function of Manhattan distance as well as faster data aggregation for construction of empirical distribution function (edf). It is simple to show that while the time complexity of such calculations for Euclidean distance is $O(n^2)$; for Manhattan distance it is just $O(n)$.

In order to compare two empirical and theoretical distributions, goodness-of-fit tests can be used (Lehmann, 1992). In 2005, Duggan et al. designed a test for Lehmer generators using the distribution function of Euclidean distance. Our proposed goodness-of-fit test used in this article is a modification of their work to cover the case of Manhattan distance.

Since the statistical structure of NOL n-tuples does not change by shifting the starting point for n-tuples, in obtaining a more precise empirical distribution we take all the consecutive n-tuples into account. For developing the proposed test, the statistical distributions of distances of NOL and OL n-tuples must be derived.

Figure 2(a) depicts the Euclidean distances of the points $(X_1, X_2), (X_3, X_4), \ldots$ for a Lehmer generator with $a = 14$ and $m = 401$. Euclidean distances for the same generator are displayed in Fig. 2(b) for the points $(X_2, X_3), (X_4, X_5), \ldots$. Figure 2(c), which is the set of all the distances in NOL form, is obtained by the superposition of Figs. 2(a) and (b). Figure 2(d) is the set of distances in OL form for the same generator.

The following section provides a concise review on the Manhattan distance. In Sec. 3, the theoretical distribution functions of distances for NOL forms in dimensions up to 5 and dimensions 1 and 2 for OL case are derived as reference distribution functions for goodness-of-fit tests in Sec. 4. This procedure can be applied in higher dimensions for finding high quality generators. In Sec. 5, the results for an exemplary modulus and three full-period multipliers are shown and the ability of the test is studied through Monte Carlo integration examples. Finally, concluding remarks are presented in the last section.

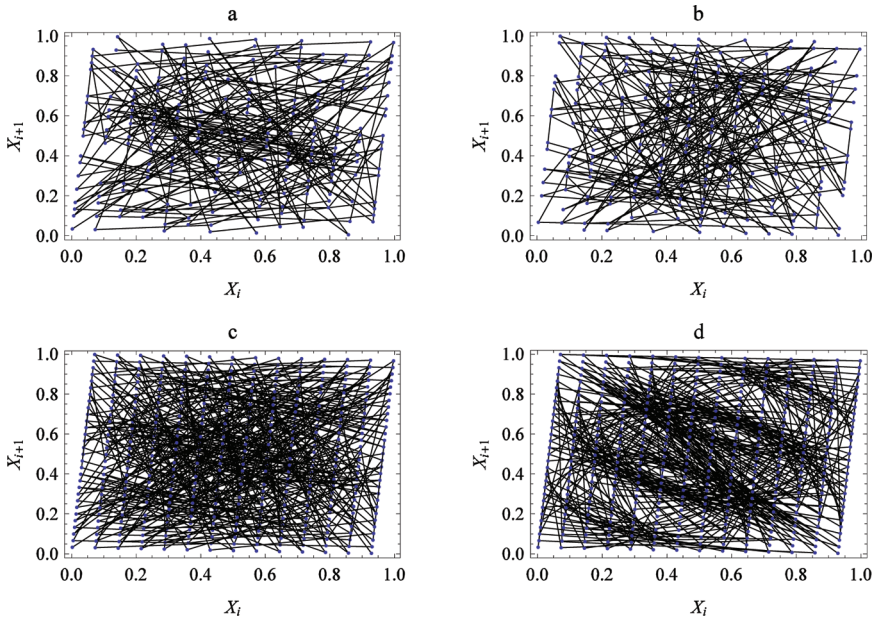

Figure 2. Differences in Euclidean distance patterns in NOL and OL pairs.

2. Manhattan Distance

Manhattan distance is also known as the *Rectangular distance* or *Taxicab metric*. Taxicab metric or taxicab geometry was first proposed as a means of creating a non Euclidean geometry by Herman Minkowski (1864–1909) early in the 20th century. The metric was one of a whole family of metrics Minkowski proposed to easily create non Euclidean geometries. From a computational perspective Manhattan distance is significantly less costly to calculate than Euclidean distance, as it does not require taking a square root. The formula for Manhattan distance between the points $p = (p_1, p_2, \ldots, p_n)$ and $q = (q_1, q_2, \ldots, q_n)$ is given by the following equation (Krause, 1986):

$$d(p, q) = \sum_{i=1}^{n} |p_i - q_i|; \quad i = 1, 2, \ldots, n. \tag{3}$$

Figure 3 shows the graphical Manhattan and Euclidean distances between two points p and q in two dimensions.

Today there is a whole spectrum of applications and implementations of taxicab geometry. Taxicab geometry is useful in a number of real-world situations such as the following (Krause, 1986).

- In chess, the distance between squares on the chessboard for rooks and knights is measured in taxicab distance.
- An extended version of taxicab geometry is used in fire-spread simulation with square-cell, grid-based maps.

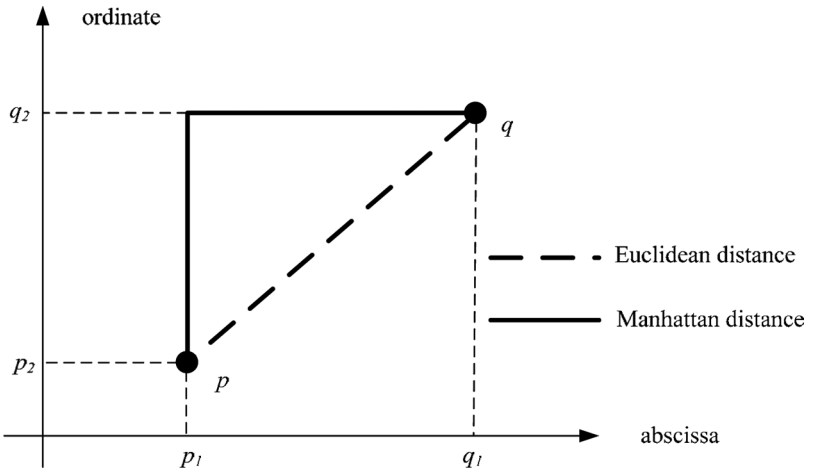

Figure 3. Manhattan distance and Euclidean distance in two dimensions.

- It is a better mathematical model than Euclidean geometry for solving problems in urban geography. In this application, taxicab geometry is a much simpler geometry, as well.

3. Theoretical Distribution Functions

3.1. *Non Overlapping Pairs*

If the sequence of $2n$ consecutive random numbers $\{X_i\}_{i=1}^{2n}$ is generated by a random number generator, the NOL distance for such consecutive numbers is given by:

$$R = d(p, q) = |X_1 - X_{n+1}| + |X_2 - X_{n+2}| + \cdots + |X_n - X_{2n}|, \tag{4}$$

which is the case of two $U(0, 1)^n$ points $p = (X_1, X_2, \ldots, X_n)$ and $q = (X_{n+1}, X_{n+2}, \ldots, X_{2n})$. Since the X_i's are i.i.d.$U(0, 1)$, the probability distribution function (pdf) of R is not dependent on the order of random variables X_i's in Eq. (4). For deriving the pdf of the random variable R, for $i = 1, 2, \ldots, n$ we define the random variables $Y_i = X_i - X_{n+i}$ and $Z_i = |Y_i|$ for $i = 1, 2, \ldots, n$ and write $R = \sum_{i=1}^{n} Z_i$. All this means that $-1 \le Y_i \le 1, 0 \le Z_i \le 1$ and hence $0 \le R \le n$.

In the first step, the pdf of random variable Y_i is derived. Since the X_i's are supposed to behave as i.i.d. $U(0, 1)$ random variables, their difference is triangular, e.g.,

$$f_{Y_i}(y_i) = \begin{cases} y_i + 1 & -1 < y_i \le 0 \\ 1 - y_i & 0 < y_i \le 1 \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Finding the pdf of random variable $Z_i = |Y_i|$ is the next step. By the transformation technique,

$$f_{Z_i}(z_i) = \begin{cases} 2(1 - z_i) & 0 \le z_i \le 1 \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Since the X_i's are independent, the Y_i's are independent as well and so are the Z_i's. Therefore, pdf of R would be attained from convolution of pdf's of random variables Z_i's:

$$f_R(r) = f_{Z_1}(z_1) * f_{Z_2}(z_2) * \cdots * f_{Z_n}(z_n). \tag{7}$$

pdfs and cumulative distribution functions (cdfs) of R for dimensions $n = 1, 2$ and 3 are:

$$f_{R_{1D}}(r) = \begin{cases} 2(1-r) & 0 \le r \le 1 \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

$$F_{R_{1D}}(r) = \begin{cases} 0 & r \le 0 \\ 2r - r^2 & 0 < r \le 1 \\ 1 & \text{otherwise,} \end{cases} \tag{9}$$

$$f_{R_{2D}}(r) = \begin{cases} 4r - 4r^2 + \dfrac{2}{3}r^3 & 0 < r \le 1 \\ \dfrac{16}{3} - 8r + 4r^2 - \dfrac{2}{3}r^3 & 1 < r \le 2 \\ 0 & \text{otherwise,} \end{cases} \tag{10}$$

$$F_{R_{2D}}(r) = \begin{cases} 0 & r \le 0 \\ 2r^2 - \dfrac{4}{3}r^3 + \dfrac{1}{6}r^4 & 0 < r \le 1 \\ 1 - \dfrac{(r-2)^4}{6} & 1 < r \le 2 \\ 1 & \text{otherwise,} \end{cases} \tag{11}$$

$$f_{R_{3D}}(r) = \begin{cases} 4r^2 - 4r^3 + r^4 - \dfrac{1}{15}r^5 & 0 < r \le 1 \\ -\dfrac{93}{15} + 21r - 22r^2 + 10r^3 - 2r^4 + \dfrac{2}{15}r^5 & 1 < r \le 2 \\ -\dfrac{(r-3)^5}{15} & 2 < r \le 3 \\ 0 & \text{otherwise,} \end{cases} \tag{12}$$

$$F_{R_{3D}}(r) = \begin{cases} 0 & r \le 0 \\ \dfrac{4}{3}r^3 - r^4 + \dfrac{1}{5}r^5 - \dfrac{1}{90}r^6 & 0 < r \le 1 \\ \dfrac{43}{30} - \dfrac{31}{5}r + \dfrac{21}{2}r^2 - \dfrac{22}{3}r^3 + \dfrac{5}{2}r^4 - \dfrac{2}{5}r^5 + \dfrac{1}{45}r^6 & 1 < r \le 2 \\ 1 - \dfrac{(r-3)^6}{90} & 2 < r \le 3 \\ 1 & \text{otherwise.} \end{cases} \tag{13}$$

pdfs and cdfs in higher dimensions are easily obtained by extending the convolution in Eq. (7).

3.2. *Overlapping Pairs (2 Dimensions)*

Here, the transformation method to obtain the pdf and cdf in two dimensions for OL pairs is presented. OL Manhattan distance for such consecutive numbers in two dimensions is given as:

$$R = |X_1 - X_2| + |X_2 - X_3|. \tag{14}$$

For calculating the cdf of the random variable R, one must calculate the following triple integral:

$$F_R(r) = P(R \le r) = \iiint_{[|X_1-X_2|+|X_2-X_3|\le r]\cap[0\le X_1,X_2,X_3\le 1]} dx_1 dx_2 dx_3. \tag{15}$$

To solve the integral in (15) we first determine all possible conditions that satisfy $[|X_1 - X_2| + |X_2 - X_3| \le r] \cap [0 \le X_1, X_2, X_3 \le 1]$. In order to simplify the transformation, we consider X_2 as a fixed term. Figure 4 shows the intended area in general in which X_3 is displayed as a linear function of X_1 where a, a', b, and b' are functions of r themselves.

After calculating all the transformation segments, the discontinuous pdf and cdf in two dimensions are found to be (the pdf and cdf in one dimension is not written

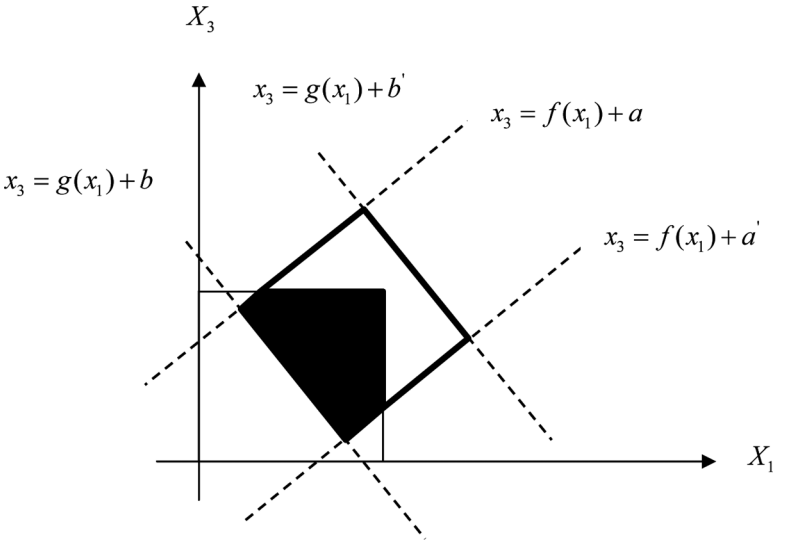

Figure 4. The integration area in general.

due to being identical in both NOL and OL cases):

$$f_{R_{2D}}(r) = \begin{cases} 0 & r \le 0 \\ 4r - \frac{7}{2}r^2 & 0 < r \le \frac{1}{3} \\ 4r - 4r^2 & \frac{1}{3} < r \le \frac{1}{2} \\ 5 - 12r + \frac{17}{2}r^2 & \frac{1}{2} < r \le 1 \\ 2 - 2r + \frac{1}{2}r^2 & 1 < r \le 2 \\ 0 & r > 2, \end{cases} \tag{16}$$

$$F_{R_{2D}} = \begin{cases} 0 & r \le 0 \\ -\frac{7}{6}r^3 + 2r^2 & 0 < r \le \frac{1}{3} \\ -\frac{4}{3}r^3 + 2r^2 & \frac{1}{3} < r \le \frac{1}{2} \\ \frac{17}{6}r^3 - 6r^2 + 5r - 1 & \frac{1}{2} < r \le 1 \\ \frac{1}{6}r^3 - r^2 + 2r - \frac{1}{3} & 1 < r \le 2 \\ 1 & r > 2. \end{cases} \tag{17}$$

4. Goodness-of-Fit Tests

Now the theoretical cdfs which were calculated in the previous section can be compared with edfs of distance between consecutive random numbers. The empirical distribution function of distance between points is: $F(x) = \frac{N(x)}{n}$, where n is the number of observations of the distance between consecutive random numbers and $N(x)$ is the number of observations which do not exceed x. By defining the empirical distribution function the hypothesis test can be stated as (Lehmann, 1992):

$$H_0 : F(x) = F_0(x)$$
$$H_1 : F(x) \neq F_0(x),$$

where $F_0(x)$ can be any of the theoretical cdfs derived in Sec. 3. Three goodness-of-fit tests are used in this article: Anderson-Darling (AD), Cramer-Von Mises (CVM), and Kolmogorov-Smirnov (KS) for hypothesis testing.

In the hypothesis tests we choose $\alpha = 0.05$ as the level of significance. For the AD test, when $n \ge 5$ if $A_n^2 \ge 2.492$, the null hypothesis should be rejected where (Anderson and Darling, 1952):

$$A_n^2 = -\sum_{i=1}^{n} \frac{2i-1}{n} \{\log[F_0(t_{(i)})] + \log[1 - F_0(t_{(n+1-i)})]\}, \tag{18}$$

where $t_{(i)}$ is the ith smallest data observation. For CVM test, if $\left(W_n^2 - \frac{0.4}{n} + \frac{0.6}{n^2}\right)\left(1 + \frac{1}{n}\right) \geq 0.461$, the null hypothesis should be rejected for any value of $n \geq 1$ where (Lehmann, 1992):

$$W_n^2 = \sum_{i=1}^{n} \left(F_0(t_{(i)}) - \frac{i - \frac{1}{2}}{n} + \frac{1}{12n} \right). \tag{19}$$

For the KS test, also if $D_n\left(\sqrt{n} + 0.12 + \frac{0.11}{\sqrt{n}}\right) \geq 1.358$ the null hypothesis should be rejected for any value of $n \geq 5$ where (Lehmann, 1992):

$$D_n^+ = \max_{i=1,2,\ldots,n} \left(\frac{i}{n} - F_0(t_{(i)}) \right), \tag{20}$$

$$D_n^- = \max_{i=1,2,\ldots,n} \left(F_0(t_{(i)}) - \frac{i-1}{n} \right), \tag{21}$$

$$D_n = \max(D_n^+, D_n^-). \tag{22}$$

On the basis of the results of the tests of hypothesis, each random number generator can be classified in one of three categories: Good: the null hypothesis is not rejected in any test; Suspect: the null hypothesis is rejected in one or two of the tests; and Bad: the null hypothesis is rejected in all three tests.

Since tests are performed for NOL pairs in dimensions up to 5, altogether 15 tests are performed for each generator. A good generator is a generator that is labeled as "Good" in all five dimensions; a bad generator is a generator that is decided to be "Bad" in all five dimensions; otherwise, the generator is labeled as a "Suspect" generator. This approach is also implemented for the case of OL pairs up to dimension 2.

5. Numerical Results and Discussion

In order to evaluate the merits of our proposed method, a set of random numbers are generated via a Lehmer generator with $m = 401$. There are 160 full-period multipliers for this generator (Leemis and Park, 2006). The purpose of this section is to identify those full period multipliers for which the edf of a reasonably large sample is close enough to the theoretical cdf of the Manhattan distance for various dimensions in the presence of the Marsaglia effect. To achieve this purpose, the edf's of the Manhattan distance between the generated points are compared with the theoretical distributions calculated in Sec. 3. Then, goodness-of-fit tests are used in dimensions up to 5 for NOL pairs and dimensions 1 and 2 for OL pairs. Finally, the best and the worst multipliers are determined. Table 1 shows the number of multipliers that lead to the rejection of the null hypothesis for dimensions up to 5 for NOL. This table also shows the number of the "Good," "Bad," and "Suspect" multipliers for each dimension. Table 2 presents the number of the "Good," "Bad," and "Suspect" multipliers across all 3 tests and all 5 dimensions. Table 3 presents the results corresponding to the OL case.

While each of the "Good" multipliers in Table 1 has successfully passed only 3 goodness-of-fit tests, it should be noticed that each of the "Good" multipliers in Table 2 has been successful in 15 tests. As such, the "Good" multipliers in Table 2 enjoy a better quality than a typical multiplier which is classified as "Good" in

Table 1
Results up to five dimensions (NOL pairs)

Dimensions	Failed AD	Failed CVM	Failed KS	Good	Suspect	Bad
1	10	4	6	150	6	4
2	36	32	32	124	4	32
3	32	22	38	118	20	22
4	56	40	56	96	28	36
5	36	24	36	114	24	22

Table 2
Overall results (NOL pairs)

Good	Suspect	Bad
42	118	0

Table 3
Results for one and two dimensions (OL pairs)

Dimensions	Failed AD	Failed CVM	Failed KS	Good	Suspect	Bad
1	10	4	6	150	6	4
2	86	98	136	20	58	82

Table 4
Overall results (OL pairs)

Good	Suspect	Bad
18	138	4

Table 5
Results up to five dimensions (NOL pairs)

Dimensions	Multiplier	AD	CVM	KS	Classification
1	$a = 309$	pass	pass	pass	Good
	$a = 129$	pass	pass	pass	Good
2	$a = 309$	pass	pass	pass	Good
	$a = 129$	fail	fail	fail	Bad
3	$a = 309$	pass	pass	pass	Good
	$a = 129$	pass	pass	pass	Good
4	$a = 309$	pass	pass	pass	Good
	$a = 129$	fail	fail	fail	Bad
5	$a = 309$	pass	pass	pass	Good
	$a = 129$	fail	fail	pass	Suspect

Table 6
Overall results (NOL pairs)

Multiplier	1D	2D	3D	4D	5D	Ultimate classification
$a = 309$	Good	Good	Good	Good	Good	Good
$a = 129$	Good	Bad	Good	Bad	Suspect	Suspect

Table 1. A similar argument holds for the OL case where results are displayed in Tables 3 and 4. For an in-depth evaluation we perform three tests of AD, KS, and CVM on the set of points generated by 2 multipliers: $a = 129$ (a Suspect multiplier) and 309 (a Good multiplier) for dimensions up to 5 for NOL pairs. Table 5 shows the results.

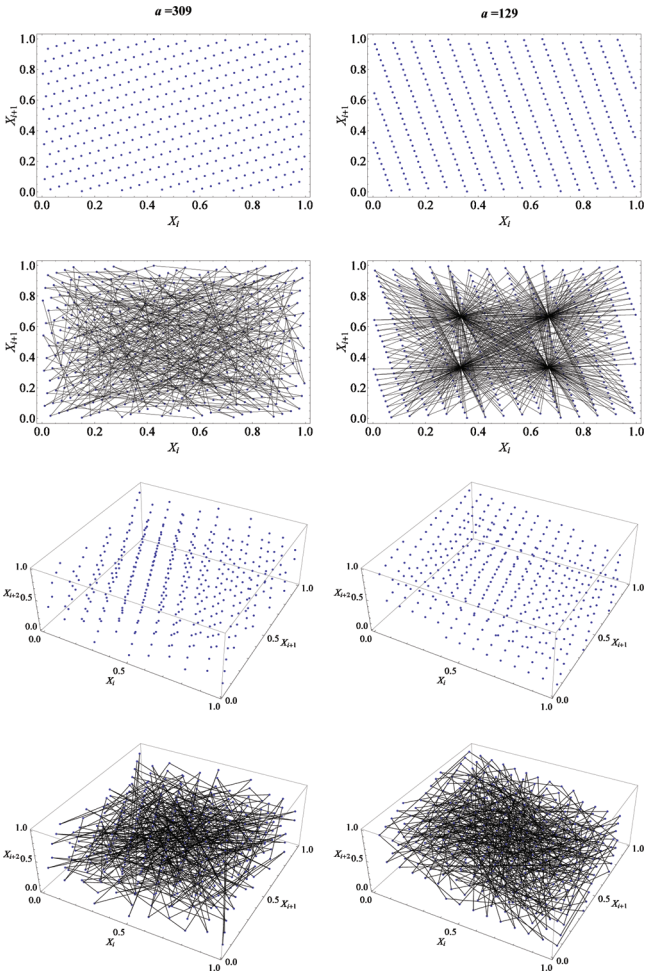

Figure 5. Plotted point and line diagrams for $a = 309$ and $a = 129$ in two and three dimensions (NOL pairs).

Table 7
Monte Carlo integration results (NOL pairs)

Multiplier	Classification	Computational error (%)
363	Good	0.6034
209	Suspect	3.4201
134	Bad	10.0193

Table 6 summarizes the results contained in Table 1. Line and point diagrams can be seen for these 2 multipliers in 2 and 3 dimensions in Fig. 5. This figure shows that the plotted diagrams agree with the test results. As a good multiplier, $a = 309$ has more hyperplanes (19) without showing little pattern in two dimensions. On the other hand, $a = 129$ as a suspect multiplier has fewer hyperplanes in 2 dimensions (16) than $a = 309$ and shows a pronounced pattern.

Monte Carlo integration is one of the major application areas for random numbers. As a further test, it was decided to numerically evaluate the following integral by sampling from the Lehmer generator when a "Good," "Bad," and "Suspect" multiplier is used.

$$\int_{y=-1}^{1} \int_{x=-\sqrt{1-y^2}}^{\sqrt{1-y^2}} x^2 + y^2 dx\, dy$$

Table 7 and Fig. 6 show the results of integration for NOL pairs. The good random number generator has the minimum computational error while the bad random number generator has the largest computational error. The suspect random number generator has an error between these two values. The maximum number of

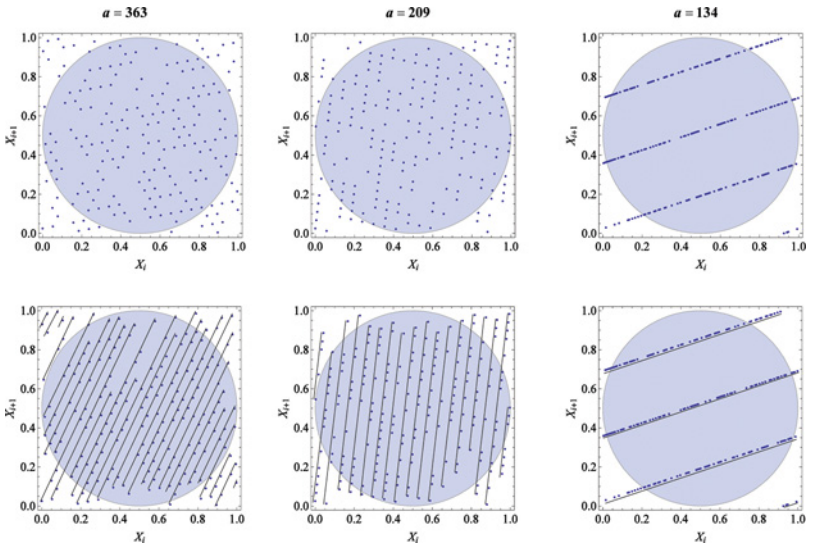

Figure 6. Graphical results of integration in NOL pairs. The second row shows more details about the hyperplanes generated by each multiplier. (color figure available online)

hyperplanes in two dimensions is given by

$$(mn!)^{\frac{1}{n}} = (401 \times 2!)^{\frac{1}{2}} \approx 28.$$

Regarding Fig. 6, the number of hyperplanes that cover all of the generated points from good to bad came out to be 28, 18, and 4, respectively. Thus, using $a = 363$ as a "Good" multiplier leads to the smallest computational error while attaining the maximum number of hyperplanes.

6. Conclusion

According to Marsaglia effect, the pseudo random numbers generated by a Lehmer generator, fall on a finite number of hyperplanes. In this article, we proposed another distance criterion based on the Manhattan distance. Having introduced this type of distance, we derived the theoretical distribution functions of the distances between points for NOL and OL pairs. Such derivation can lead to the exact pdf and cdf of the distance for NOL for any arbitrary dimension and dimensions 1 and 2 for the OL case. To shed light on the merits of the proposed criterion, it was decided to resort to three goodness-of-fit tests. Implementation of these tests was carried out by comparison of edf of the distance against its corresponding theoretical distribution function. Such comparison was performed for dimensions 1–5 in NOL case as well as dimensions 1 and 2 for the OL case. In this way we established the fact that using this criterion, leads to very good full-period multipliers in terms of attaining the maximum number of hyperplanes as well as having the minimum computational errors.

References

Anderson, T. W., Darling, D. A. (1952). Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes. *Ann. Mathemat. Statist.* 23:193–212.

Dugan, M. J., Drew, J. H., Leemis, L. M. (2005). A test of randomness based on the distance between consecutive random number pairs. In: Kuhl, M. E., Steiger, N. M., Armstrong, F. B., Joines, J. A., eds. *Proceedings of the 2005 Winter Simulation Conference*. Piscataway, NJ: Institute of Electrical and Electronics Engineers, Inc., pp. 741–748.

Fishman, G. S., Moore, L. R. (1986). An exhaustive analysis of multiplicative congruential random number generators with modulus 2^{31}-1. *SIAM J. Scientif. Statist. Comput.* 7:24–45.

Knuth, D. E. (1997). *The Art of Computer Programming, Seminumerical Algorithms*. 3rd ed. Vol. 2. Boston: Addison-Wesley.

Krause, E. F. (1986). *Taxicab Geometry: An Adventure in Non-Euclidean Geometry*. Mineola, NY: Dover.

L'Ecuyer, P. (1999). Tables of linear congruential generators of different sizes and good lattice structure. *Math. Computat.* 68(225):249–260.

Leemis, L. M., Park, S. K. (2006). *Discrete-Event Simulation: A First Course*. Upper Saddle River, NJ: Prentice–Hall.

Lehmann, E. L. (1992). *Introduction to Neyman and Pearson (1933) on the Problem of the Most Efficient Tests of Statistical Hypotheses, in Break-throughs in Statistics*. Vol. 1. Heidelberg: Springer-Verlag.

Marsaglia, G. (1968). Random numbers fall mainly in the planes. *Proc. Nat. Acad. Sci.* 61:25–28.

Park, S. K., Miller, K. W. (1988). Random number generators: Good ones are hard to find. *Commun. ACM* 31(10):1192–1201.