DGOR/NSOR Tagung 1993 Amsterdam:

THE QUALITY OF NON-UNIFORM RANDOM NUMBERS

Wolfgang Hörmann, WU Wien

Summary: The quality of non-uniform random numbers is not only influenced by the quality of the uniform generator that is used but also by the transformation method applied to the uniform random numbers. This differences in quality between "exact" methods were almost entirely neglected in literature. So we compare the behaviour of four different transformation methods when combined with a linear congruential uniform generator (LCG). Heuristic considerations, the computation of two measures of approximation and a statistical test show that the inversion method performs best. Among the others rejection, when combined with a LCG with small multiplier, and ratio of uniforms perform worse. Their use could slightly change the results of some simulation studies.

Zusammenfassung: Die Qualität von nicht gleichverteilten Zufallszahlen wird nicht nur vom benützten Gleichverteilungsgenerator sondern auch von der verwendeten Transformationsmethode beeinflußt. Dieser Unterschied in der Qualität verschiedener "exakter" Transformationsmethoden wurde bisher in der Literatur fast völlig vernachlässigt. Darum vergleichen wir das Verhalten vier verschiedener Transformationsmethoden in Verbindung mit linearen Kongruenzgeneratoren (LKG). Heuristische Überlegungen, die Berechnung von zwei Maßen für die Approximation und ein statistischer Test zeigen, daß die Inversionsmethode am besten abschneidet. Unter den anderen Verfahren schneiden die Verwerfung – in Kombination mit einem LKG mit kleinem Multiplikator – und die Quotientenmethode am schlechtesten ab. Es ist nicht auszuschließen, daß ihre Verwendung die Resultate von Simulationsstudien verfälscht.

1. Introduction

The literature of random number generation falls into two parts. There are a lot of papers dealing with the quality of uniform pseudo-random number generators. On the other hand many papers discuss the generation of non-uniform variates by transforming an independent sequence of uniform random numbers and are mainly concerned with the speed or simplicity of the proposed algorithms. Concerning quality it is only stated that the method is exact which means that perfect uniform random numbers (which are not available) would be transformed into independent random numbers of the correct distribution. Investigations dealing with the effects that can occur when such an exact transformation method is combined with a pseudo-random sequence (this is done in every simulation which needs non-uniform random numbers) are very rare (see [1], [8], [7] and references there). But there are important differences in quality between exact transformation methods when combined with the same linear congruential pseudo-random number generator (LCG). This old but still very popular uniform generator is based on the simple recursion $x_i = (a \cdot x_{i-1} + c) \bmod m$; to get random numbers uniformly distributed on (0,1) this sequence is divided through m. It is well known that n-tuples produced by successive calls to a LCG form a lattice or grid and that the properties of this lattice depend on the choice of the multiplier a (see e.g. [9]).

In order to assess the quality of a uniform random number generator it is generally accepted that it is sensible (and more powerful than statistical tests) to investigate the distribution of all n-tuples that can be returned by that generator. For $n = 1$ this is necessary to guarantee the uniform distribution of the numbers, whereas the higher values are important for the order of the sequence and thus for the independence of the variates. A good uniform generator should have a regular distribution of all generated n-tuples for n between 1 and (at least) 20 as for a single simulation only a small portion of these points is taken as a sample. (In [10] it is demonstrated that not more than the 2/3 power of the period should be taken.)

2. The one-dimensional distribution

As we want to discuss the quality of non-uniform random numbers we will compare the empirical distribution function G_n of all possible points for several methods of generating non-uniform variates when a LCG is used as uniform generator. One of our measures for the quality of the approximation of the theoretical cumulative distribution function F by G_n will be the classical concept of discrepancy:

$$D_n(F) := \sup_x (G_n(x) - F(x)) + \sup_x (F(x) - G_n(x))$$

In [1] numerical comparisons of the discrepancy for various normal generators combined with LCG's are presented. But the main disadvantage of discrepancy is that it is sensitive only to the greatest difference. So we agree with Monahan in [12] that measuring the L_1-distance between G_n and F should give more information about the quality of the approximation.

$$E_n(F) := \int_R |F(x) - G_n(x)|\, dF(x)$$

In [12] accuracy of random variate generation is discussed under the assumption that truly random bits are available. In this paper we use distance criteria D and E to compare the quality of different transformation methods when combined with LCG's. For a LCG with maximal period m we have for the discrepancy $D = 1/m$. The value of E depends on how the integers x_i are transformed into floating point numbers u_i between 0 and 1. If $u_i = x_i/m$ we have $E = 1/(2m)$, for $u_i = (x_i + 0.5)/m$ (which was used for all computations of this paper) $E = 1/(4m)$. The only generation method for non-uniform deviates that preserves this good approximation is inversion. As long as we neglect numerical inaccuracies when computing the inverse distribution function, non-uniform random deviates generated by inversion have the same values for D and E as the LCG itself, all other methods are much worse. Before we present the empirical results we want to summarize the heuristic and theoretical considerations contained in [1], [8] and [7]. The accuracy of a LCG is better in low dimensions than in higher ones. Therefore a method that transforms one random number of the LCG into one non-uniform random number (like inversion) should be better than a method that transforms two or more uniform random numbers in one non-uniform deviate. But this simple rule does not explain all observations concerning the quality of non-uniform variate generation, mainly because most methods do not transform a fixed number of uniform deviates. The expected number of uniforms required by a certain transformation method and the rule, less uniforms imply better quality, give only a very crude rule of thumb for the quality of a transformation method when combined with a LCG. More insight is up to now only available for two of the most important transformation methods: For ratio of uniforms (see for example [3]) it is easy to see that all pairs (u, v) lying on one line through the origin are transformed into the same random number. Due to the lattice structure of the pairs generated by a LCG (see for example [9]) there must always be relatively large intervals without a random number when the ratio of uniforms method is combined with a LCG. (For details and bounds for the discrepancy see [8].) For the rejection method the simple observation that a LCG with a small multiplier (smaller than or close to $\sqrt{m}$) forms a two dimensional lattice with the shortest vector almost parallel to the y-axis has important consequences: Rejection algorithms combined with such LCG's (which were often recommended in literature because of their ease of implementation) have a bad approximation of the desired distribution. (For details see [7].)

In order to compare the quality of the approximation of the one-dimensional distribution for various transformation methods we computed criteria D and E for three different exponential generators: Transformed rejection (TR), a rejection technique that uses inversion to sample from the dominating distribution (see [6] and [5]), algorithm TRD that uses transformed rejection combined with decomposition to reduce the expected number of uniforms required (see [6] and [5]), and ratio of uniforms (RoU) (see eg. [3]). We restricted our attention to the exponential distribution only as it seems sensible to assume (and was checked by several numerical experiments) that the shape of a continuous distribution influences the quality of approximation of the different methods only slightly. To study the influence of different LCGs we combined each of the transformation methods with the following four LCGs. All of them have a good lattice structure up to 20 dimensions.

LCG 1: $m = 2^{32}$,	$a = 663608941$,	$c = 0$, Period=2^{30}	suggested in [2]
LCG 2: $m = 2^{32}$,	$a = 1099087573$,	$c = 0$, Period=2^{30}	suggested in [4]
LCG 3: $m = 2^{32}$,	$a = 69069$,	$c = 1$, Period=2^{32}	suggested in [11]
LCG 4: $m = 2^{32}$,	$a = 1589013525$,	$c = 1$, Period=2^{32}	suggested in [14]

To determine the exact value of D and E it is necessary to sort all random numbers that can be generated by a certain method. This was done for algorithm TR, for which it is easy to compute all generated random numbers in ascending order without storing them. For the remaining two algorithms we computed upper and lower bounds by counting the number of random variates, which fall into $2 \cdot 10^7$ intervals of equal probability. Table 1 contains the results of our computations for the distance criteria D and E divided through the value of D and E of the used LCG.

Table 1

	Distance E			Distance D		
	TRD	TR	RoU	TRD	TR	RoU
LCG1	(231|432)	48	(213|424)	(2671|2719)	126	(50088|50196)
LCG2	(318|522)	77	(26|184)	(3150|3258)	68	(29366|29474)
LCG3	(43248|44962)	4545	(48|638)	(29159|29589)	18559	(117233|117663)
LCG4	(266|999)	143	(212|947)	(1013|1443)	131	(109322|109753)

The results of Table 1 show that there are large differences between the quality of the approximation of the desired distribution. Inversion has the best properties (the values for inversion in Table 1 would all be one), TR combined with LCGs 1, 2 and 4 ranks second. Both criteria show the very bad approximation of the rejection methods (TRD and TR) when combined with a LCG with small multiplier whereas the problems of RoU can only be seen in the very high values of D.

We want to emphasize again that E and D measure the quality of the approximation of the desired distribution by all numbers that can be returned by a generator. The question how these properties influence the behaviour of small samples can be tackled in two ways: The first and theoretical more sound possibility would be to study the properties of n-tuples of successive points for several values of n. With the exception of inversion, where the results for LCGs can be applied directly, it seems impossible to assess the quality of the approximation of the distribution in higher dimensions theoretically as there is no lattice any longer. The empirical investigation by computing bounds for the discrepancy or other measures of approximation is not possible as computing- and memory-requirements increase with the power of n. So we can only get some insight into the structure of the two-dimensional set of points generated by inversion, rejection and the ratio of uniforms method, which is done in Section 3. In Section 4 we present a statistical test that seems well suited to measure the degree of randomness of small samples.

3. The two-dimensional distribution

A fist and very helpful step to get some insight into the structure of the set of all pairs generated by a transformation method combined with a LCG is to plot pictures of these pairs for small uniform generators. We used the LCG with $m = 2^{11}$, $a = 605$ and $c = 1$, which has good lattice properties. Figure 1 shows all pairs generated with inversion together with that LCG and demonstrates that the lines of the lattice of the LCG are just transformed into curves. Strictly speaking a simple computation shows that by inversion the line $y = kx + d$ of the two-dimensional grid of the uniform generator is transformed into the curve $y = F^{-1}(k \cdot F(x) + d)$, which is $y = -log(1 - k(1 - exp(-x)) - d)$ for the exponential distribution. Thus we see that a good two-dimensional distribution of the LCG is preserved by inversion.

Figure 1

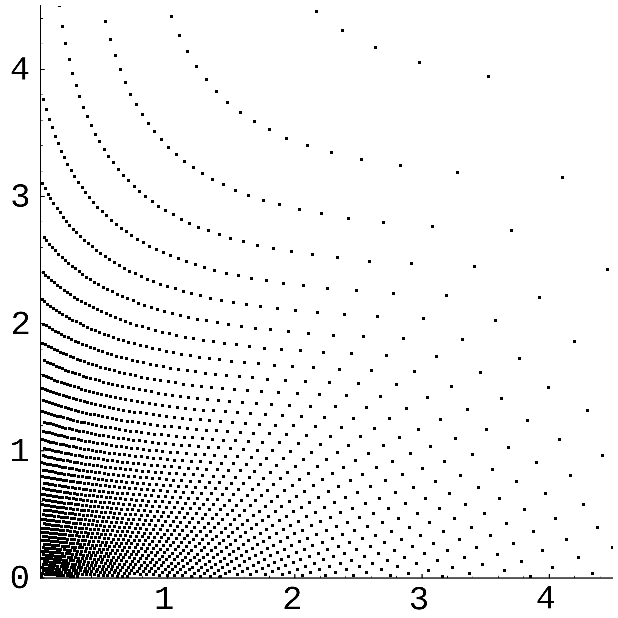

Figure 2

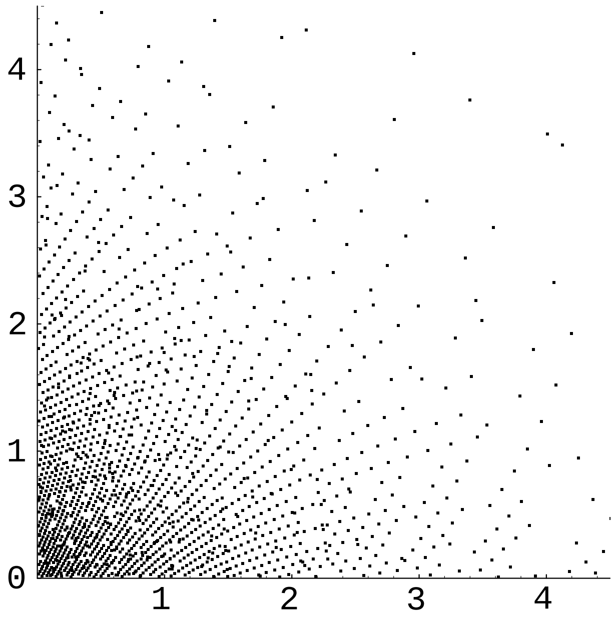

Figure 3

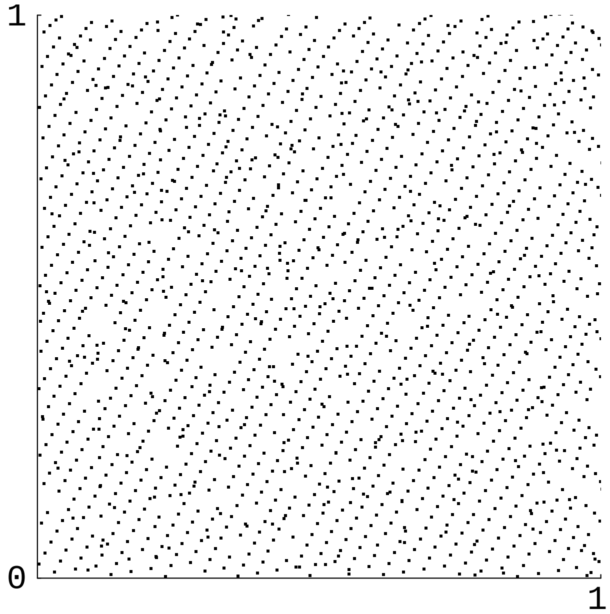

Figure 4

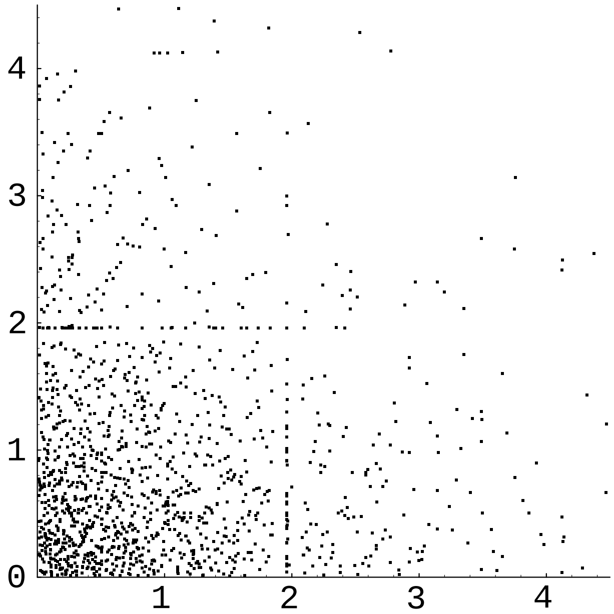

Figure 2 shows the set of all pairs generated with algorithm TR and the same LCG. At a first glance Figure 1 and Figure 2 look quite similar. This is not astonishing as the dominating distribution of TR is quite close to the exponential distribution and is generated by inversion. But it is not the 2-dimensional grid of all consecutive pairs of the LCG that is transformed. Due to the rejection algorithm a second uniform random number is necessary and so we see the transformed grid of all pairs of the LCG with one random number left out. Algorithm TR transforms at least 4 uniform random numbers into two exponential variates. So Figure 2 shows most points of the grid of all 4-tuples of the LCG transformed by the mapping $(x_1, x_2, x_3, x_4) \to (G(x_1), G(x_3))$ with G close to F^{-1} of the exponential distribution. Figure 3 shows the same points as Figure 2 but transformed by F of the exponential distribution. The grid is even more evident, but the lines are a bit curved due to the difference between the dominating distribution of TR and the exponential distribution. It is evident, especially when looking at Figure 3, that not all points lie in the regular grid and that some points of the regular grid are missing. This is due to the rejection logic that throws away some points of the main grid (about 16%) and replaces it by points of the transformed grid of all pairs of the LCG with three, five or more random numbers left out.

For TRD the two-dimensional distribution is quite similar. But for RoU (Figure 4) the set of all pairs looks entirely different. The grid structure of the LCG seems to have entirely disappeared. On the other hand RoU is quite similar to rejection as pairs of uniform random numbers are accepted or rejected. Thus the set of all points generated by RoU again consists mainly of points generated by transforming 4-tuples of the LCG. But this transformation $((x_1, x_2, x_3, x_4) \to (x_2/x_1, x_4/x_3)$, is not so simple as in the case of TR. One-dimensional affine subspaces of R^4 are transformed into hyperbolas of the form $y = (k_1 x + d_1)/(k_2 x + d_2)$ and any two-dimensional subspaces of R^4 containing the origin is transformed into one single point. Therefore the lattice structure of the LCG disappears entirely.

The above pictures and considerations seem to indicate that the quality of the approximation of the two-dimensional theoretical distribution in combination with LCGs is best for the inversion method, among the others ratio of uniforms seems to be worst. But as stated above it is very hard to compute measures for the quality of the approximation.

4. Properties of small samples

In [10] Maclaren suggests a test to detect excess of uniformity for random number generators. It is based on a result of Moran (see [13]), who shows that the sum of squares of the spacings of a true random U(0,1) sample of size N is asymptotically normal with mean $2/N$ and variance $4/N^3$. Maclaren uses this test to show that any uniform pseudorandom number generator that is based on sampling without replacement (like LCG's for example) returns too uniform samples (the sum of squares of the spacings is too small) if the sample size exceeds the 2/3 power of the period of the generator. Of course the test can also be used to detect too little uniformity if the sum of squares of the spacings is too large. By transforming variates with the cumulative distribution function the test can also be applied to non-uniform random numbers. We tested all of the transformation methods and all LCG's contained in Table 1 for samples of the size $N = 10^5$, 10^6 and $3 \cdot 10^6$. Table 2 gives the number of cases the sum of squares of the spaces was larger than the median (the sample was less uniform than an average random sampling) out of the hundred replications we made for each combination. For a real random sample the value lies with probability 0.95 between 40 and 60.

Table 2

	$N = 10^5$				$N = 10^6$				$N = 3 \cdot 10^6$			
	INV	TRD	TR	RoU	INV	TRD	TR	RoU	INV	TRD	TR	RoU
LCG1	48	43	44	46	17	9	0	96	0	0	0	100
LCG2	44	42	42	49	13	10	1	71	0	0	0	100
LCG3	52	48	66	55	38	100	100	57	23	100	100	93
LCG4	48	40	51	52	35	46	46	52	26	11	13	92

The results of Table 2 support the consideration in [10]: The uniformity of the samples of INV (i.e the results of the pure LCG) grows with the sample size. For $N = 3 \cdot 10^6$ they are definitely too uniform. Apparent are also the differences between LCGs 1 and 2 with period 2^{30} and LCG's 3 and 4 with period 2^{32}. For the other transformation methods it is evident that the whole period considerations of Section 2 and 3 have an influence on the properties of shorter samples. RoU and rejection (TR and partly TRD) combined with a LCG with a small multiplier (LCG 3) are less uniform than the used LCG and also less uniform than a truly random sample. For RoU the deviation is smaller but it occurs for all LCGs. The results for TRD and TR with LCGs 1, 2 and 4 are even more uniform than the results of INV. An explanation could be the fact that the period of TRD and TR is shorter than the period of the LCG itself.

5. Conclusions

We also did some simulation of an M/M/1-queue to find out, if the differences of the transformation method would influence the simulation results. As long as we used LCGs with good lattice properties all of the above transformation methods lead to correct results when compared with the theoretical solution. On the other hand a test with a LCG with bad lattice structure in three dimensions showed that the results for RoU were much worse than for the other transformation methods. Therefore even the simulation of this simple queue supports the claim that not only the choice of the uniform generator but also the choice of the transformation method can influence the results of a simulation. We are convinced that the theoretical and empirical considerations of this paper justify the recommendation: In combination with LCGs use inversion or, for the case that this is very slow, rejection with a large multiplier for the LCG. The use of ratio of uniforms combined with any LCG or rejection combined with a LCG with small multiplier could possibly influence the results of simulation studies.

References

[1] Afflerbach, L. and Hörmann, W.: Nonuniform random numbers: a sensitivity analysis for transformation methods. in: International Workshop on Computationally Intensive Methods in Simulation and Optimization, U. Dieter and G. Ch. Pflug, eds., Lecture Notes in Econom. Math. Systems 374, Springer-Verlag, New York (1992)

[2] Ahrens, J. H., Dieter, U. and Grube, A.: Pseudo-random numbers a new proposal for the choice of multiplicators. Computing 6, 121-138 (1970)

[3] Devroye, L.: Non-Uniform Random Variate Generation. Springer-Verlag, New York (1986)

[4] Fishmann G. S.: Multiplicative congruential random number generators with modulus 2^β: An exhaustive analysis for $\beta = 32$ and a partial analysis for $\beta = 48$. Math. Comp. 54, 331-344 (1990)

[5] Hörmann, W. and Derflinger, G.: The transformed rejection method for generating random variables, an alternative to the ratio of uniforms method. Manuskript, Institut f. Statistik, Wirtschaftsuniversität Wien, (1991)

[6] Hörmann, W.: New generators of normal and poisson deviates based on the transformed rejection method. in: Operations Research Proceedings 1992, ed. K.-W. Hansmann et al., Springer Verlag, Heidelberg (1993)

[7] Hörmann, W. and Derflinger, G.: A portable uniform random number generator well suited for the rejection method. ACM Transact. on Math. Softw., (to appear)

[8] Hörmann, W.: A note on the quality of random variates generated by the ratio of uniforms method. ACM Transactions on Modeling and Computer Simulation, (to appear)

[9] Knuth, D. E.: The Art of Computer Programming. Vol. II. Addison-Wesley, Menlo Park, London, Amsterdam, Don Mills, Sydney (1981)

[10] Maclaren, N.M.: A limit on the usable length of a pseudorandom sequence. J. Statist. Comput. Simul. 42, 47-54 (1992)

[11] Marsaglia G.: The structure of linear congruential sequences. in: Applications of Number Theory to Numerical Analysis, S. K. Zaremba, ed., Academic Press, New York (1972)

[12] Monahan, F.: Accuracy in random number generation. Math. Comp. 45, 559-568 (1985)

[13] Moran, P.A.P.: The random division of an interval. J. Royal Statist. Soc. (Supplement) 9, 92-98 (1947)

[14] Niederreiter, H.: Quasi-monte carlo methods and pseudo-random numbers. Bulletin of the American Math. Soc. 84, 6, 957-1041 (1978)