Version: 20.7.1993 (to appear in ACM Transactions on Modelling and Computer Simulation, Vol 4, 1)

A Note on the Quality of Random Variates Generated by the Ratio of Uniforms Method

Wolfgang Hörmann
Inst. f. Statistik Wirtschaftuniv. Wien, Augasse 2-6, A-1090 Vienna, Austria.
e-mail: whoer@statrix2.wu-wien.ac.at

Abstract: The one-dimensional distribution of pseudo-random numbers generated by the ratio of uniforms methods using linear congruential generators (LCGs) as the source of uniform random numbers is investigated in this paper. Due to the two-dimensional lattice structure of LCGs there is always a comparable large gap without a point in the one-dimensional distribution of any ratio of uniforms method. Lower bounds for these probabilities only depending on the modulus and the Beyer quotient of the LCG are proved for the case that the Cauchy the normal or the exponential distribution are generated. These bounds justify the recommendation not to use the ratio of uniforms method combined with LCGs.

Categories and Subject Descriptors: G.3 [Probability and Statistics]: Random Number Generation
General Term: Algorithms
Additional Terms: Ratio of uniforms method, linear congruential generator, discrepancy

1. Introduction

The ratio of uniforms method suggested in [10] is a popular transformation method to generate non-uniform random variates out of uniform random numbers and can be applied to a variety of distributions. It is called exact in the sense that it transforms a sequence of independent uniform random numbers into a sequence of independent variates with the correct distribution. In practice we have no ideal uniform random numbers but only pseudo-random numbers generated by a certain form of recursion. Questions concerning the influence of this fact to the quality of non-uniform deviates were almost entirely neglected in literature as B. D. Ripley states in [19]. We only know some papers discussing the quality of the Box-Muller method (for example [16] and [1]) and the recent papers [3], [5] and [4]. In the current paper we investigate the one-dimensional distribution of the ratio of uniforms method for the case that a linear congruential generator (LCG) is used to generate the uniform random numbers.
The still very popular LCG with modulus m, multiplier a and increment c is based on the simple recursion; $x_{i+1} = (ax_i + c) \bmod m$; to get random numbers uniformly distributed on (0,1) this sequence is divided by m. It is well known that n-tuples produced by successive calls to a LCG form a lattice or grid. Further information on LCG's and their lattice structure can be found in [11].
As we want to examine the one-dimensional distribution of non-uniform random sequences we need the following definition of the one-dimensional discrepancy which is also given in [12]: For a sequence $x_1, \ldots, x_k$ of R the discrepancy with respect to the continuous distribution function $F(x)$ is defined by:

$$D_k(F) := sup_{\alpha \leq \beta} \mid \frac{\#\{x_i : \alpha < x_i \leq \beta\}}{k} - (F(\beta) - F(\alpha)) \mid$$

Other notions of discrepancy and the importance of discrepancy for Monte Carlo Integration and the assessment of uniform random number generators can be found in [17], [12] and [2]. In the sequel the discrepancy of the ratio of uniforms method combined with a LCG will be the discrepancy of all points (the period) that can be returned by that combination with respect to the corresponding distribution function. As only very short parts of that period should be used for a simulation experiment the whole period can be looked at as the parent population we are drawing a sample of. Therefore it is clear that the discrepancy of the method should be small as this implies that the parent population has the correct distribution. A low one-dimensional discrepancy is of course not a sufficient condition for a good method but it is a necessary one.

2. The Table Mountain Distribution

The ratio of uniforms method is based on the following fact which is valid for any density function $f(x)$: If (U,V) is uniformly distributed over $C = \{(u,v) \mid 0 \leq u \leq \sqrt{c \cdot f(v/u)}\}$ $X = V/U$ has density $f(x)$. If f has sub-quadratic tails the area C is bounded. Thus it is possible to enclose C in a rectangle and to generate (U,V) uniformly distributed over C with the two-dimensional rejection method. For the Cauchy distribution the region of acceptance $C = \{(u,v) \mid 0 \leq u \leq 1, u^2 + v^2 \leq 1\}$ is the half circle with radius 1, for the normal distribution $C = \{(u,v) \mid 0 \leq u \leq 1, |v| \leq 2u\sqrt{-\log(u)}\}$ is enclosed in the rectangle $(0,1) \times (-\sqrt{2/e}, \sqrt{2/e})$, for the exponential distribution $C = \{(u,v) \mid 0 \leq u \leq 1, 0 \leq v \leq -2u\log(u)\}$ is enclosed in the rectangle $(0,1) \times (0, 2/e)$. For a detailed description of the ratio of uniforms method cf. [10] or [6]. If the region C is a rectangle of the form $(0,1) \times (v^-, v^+)$ it is easy to show ([7] or [20]) that the ratio of uniforms method samples from a "table mountain distribution" with density function $f(x) = \min(1/x^2, 1)/(2(v^+ - v^-))$. As the standard ratio of uniforms method for most distributions starts with a uniform distribution in a rectangle and uses rejection to get a pair uniformly distributed over the region C ratio of uniforms can be understood as rejection from the table mountain distribution ([7] or [20]).
It is obvious that the ratio of uniforms method transforms all points lying on one line through the origin into only one random number. As all points returned by a LCG form a lattice or grid this implies that there is a gap without a point in the direction of the shortest lattice vector (see [3] and [4]). The probability of this gap with respect to the desired distribution (prob(gap)) is of the order $1/\sqrt{m}$. Figure 1 shows this situation for the Cauchy distribution combined with the LCG with $m = 2^9$, $a = 117$ and $c = 1$. The left part of Figure 1 shows the region C together with the grid of the LCG, the right part the density functions of the corresponding table mountain distribution and of the Cauchy distribution and the points of the transformed grid. The segment of the unit circle on the left hand side is transformed into an interval without a point of the Cauchy distribution. To derive lower bounds for the probability of the gap we will first investigate the probability of the largest gap in the table mountain distribution. To keep the formulation short all theorems below are valid for LCGs with period m or $m-1$. For the case that the period is $m/4$ (for $m = 2^e$ and $c = 0$) the factor $1/\sqrt{m}$ in the bounds must be replaced by $2/\sqrt{m}$. For the symmetric case we arrived at the following theorem.

Theorem 1: For the ratio of uniforms method with $C = (0, u^+) \times (-v^+, v^+)$ combined with a LCG with modulus m there exists a gap (an interval without a point) with

$$\text{prob(gap)} \geq \frac{1}{\sqrt{m}}\sqrt{\frac{\sqrt{3}}{32}} \geq \frac{0.2326}{\sqrt{m}}$$

Proof: I) As a first step we consider the case $C = (0, m) \times (-m, m)$; The successive pairs returned by the LCG form a grid with basis vectors $(0, m)$ and $(1, a)$, in our case the y-values have to be multiplied by 2. For the Minkowski-reduced basis (cf. [2]) e_1 and e_2 of the lattice with $| e_1 | \leq | e_2 | \leq | e_1 \pm e_2 |$; the angle between e_1 and e_2 is $\alpha \geq \pi/3$; as the volume of a unit cell of the lattice is $2m$ we get: $|e_1||e_1| \leq |e_1||e_2| \leq 2m/\sin\alpha \leq 4m/\sqrt{3}$. Now we assume that at least two of the lines of the grid parallel to e_1 intersect the y-axis between $-m$ and m; the line intersecting closest to the origin below is called g_1 the one intersecting closest above is called g_2 . The (normal) distance between g_1 and g_2 denoted $dist = 2m/ | e_1 | \geq \sqrt{m\sqrt{3}}$.
The triangle (or quadrangle) A with the vertices origin, intersection of g_1 with the linear graph $((0, m), (m, m), (m, -m), (0, -m))$, intersection of g_2 contains no point and we have $\text{area}(A) \geq m\, dist/2$. Finally we arrive at: $\text{prob(gap)} = \text{area}(A)/\text{area}(C) \geq \sqrt{\frac{\sqrt{3}}{16m}}$
II) For the case that only one of the lattice lines parallel to e_1 intersects the y-axis between $-m$ and m we call this line g. (As the lattice has a basis vector $(0, 2m)$ one line intersects the y-axis in that part in any case.) If the intersection has the same sign as the ascent of g we get the same estimate as above, otherwise we have to consider two cases:
Case a) $| e_2 | \leq 2 | e_1 |$: We have: $| e_2 || e_2 | \leq 2\, | e_1 || e_2 | \leq 8m/\sqrt{3}$. Following the ideas of part I) with e_1 replaced by e_2 (in this case at least two lattice lines parallel to e_2 intersect the y-axis between $-m$ and m) we get: $\text{prob(gap)} \geq \sqrt{\frac{\sqrt{3}}{32m}}$
Case b): $| e_2 | \geq 2 | e_1 |$: We get $| e_1 | \leq \sqrt{2m/\sqrt{3}}$; as $(0, 2m)$ is a lattice vector the area of the triangle A with the vertices origin, $(0, m)$ or $(0, -m)$ and the intersection between g and the linear graph $((0, m), (m, m), (m, -m), (0, -m))$, is larger than $m\, dist/4$ which again includes $\text{prob(gap)} \geq \sqrt{\frac{\sqrt{3}}{32m}}$
As the mapping $(x, y) \to (\gamma\, x, \mu\, y)$ $\gamma, \mu > 0$ does not change the probability of the gap the proof is complete. $\square$

The discrepancy is of course greater or equal to the probability of the largest gap. As the discrepancy is $1/m$ for all LCGs with maximal period m and $2/m$ for LCGs with m prime and period $m - 1$ and as non-uniform random numbers generated by inversion have the same discrepancy as the uniform generator the above theorem shows that the ratio of uniforms method with $C = (0, u^+) \times (-v^+, v^+)$ has a high one-dimensional discrepancy. For the case of the half table mountain ($C = (0, u^+) \times (0, v^+)$) the shortest lattice vector e_1 need not point in the direction of C. Therefore it is impossible to give a general bound only depending on m, but we can give a bound depending on m and on the Beyer quotient in two dimensions $q_2 = |e_1|/|e_2|$. As it is generally accepted that good LCGs must not have small Beyer quotients in several dimensions (cf. [2]) the theorems below give bounds for LCGs with a good two-dimensional lattice structure. Nevertheless it is possible to construct special examples of LCGs with a very low q_2 which lead to a low discrepancy when used together with the ratio of uniforms method with a certain C, but these special examples have no practical relevance.

Theorem 2: For the ratio of uniforms method with $C = (0, u^+) \times (0, v^+)$ combined with a LCG with modulus m and Beyer quotient q_2 there exists a gap with

$$\text{prob(gap)} \geq \frac{1}{\sqrt{m}}\sqrt{\frac{\sqrt{3}}{8}\frac{q_2}{q_2^2+1}}$$

Proof: Without loss of generality (cf. end of proof of Theorem 1) we take $C = (0,m) \times (0,m)$ and assume that e_1 and e_2 have positive x-coordinates. In the worst case both have negative y-coordinate. As e_1 and e_2 are a Minkowski basis easy geometric considerations show that either $e_1 - e_2$ or $e_2 - e_1$ point into C. For the angle α between e_1 and e_2 we have $\pi/2 \geq \alpha \geq \pi/3$; the volume of the unit cell of the grid is m; thus we get:

$$|e_1 - e_2|^2 \leq |e_1|^2 + |e_2|^2 = |e_2|^2(q_2^2+1) \leq \frac{m}{q_2 \sin\alpha}(q_2^2+1) \leq \frac{2m(q_2^2+1)}{\sqrt{3}q_2}$$

The grid line parallel to $e_2 - e_1$ intersecting the y-axis in the interval $[0,m)$ closest to the origin is called g_1, g_2 is the next parallel grid line below intersecting the x-axis in the interval $(0,m)$. As $(0,m)$ and $(m,0)$ are lattice vectors g_1 and g_2 are well defined. Therefore it is enough to proceed analogous to the proof of Theorem 1 part I) to complete the proof. □

Combining Theorem 1 and 2 it is not difficult to show a similar result for assymmetric table mountains. For $C = (0,u^+) \times (v^-,v^+)$ the probability of the gap is greater or equal $\frac{1}{\sqrt{m}}\sqrt{\frac{\sqrt{3}}{16}\frac{q_2}{q_2^2+1}}$. To avoid repetitions we omit the details.

3. The Cauchy, Normal and Exponential Distributions

As already stated the ratio of uniforms method can be interpreted as rejection from a table mountain distribution. In the previous section we have given estimates for the probability of the gap of the table mountain. In this section we examine the influence of this gap when generating distributions of practical importance. Is it possible that the large gap is entirely rejected? We will restrict ourselves to the three most prominent examples: the Cauchy, the normal and the exponential distribution, as these examples are enough to see that a gap with comparable large probability will occur for any of the continuous distributions (including the t- and the gamma distribution) the ratio of uniforms method is often recommended for. For the Cauchy distribution we obtain as a simple corollary of Theorem 1.

Theorem 3: For the ratio of uniforms method with $C = \{(u,v) \mid 0 \leq u \leq 1, u^2+v^2 \leq 1\}$ (Cauchy distribution) combined with a LCG with modulus m there exists a gap with

$$\text{prob(gap)} \geq \frac{1}{\sqrt{m}}\sqrt{\frac{\sqrt{3}}{2}\frac{1}{\pi}} \geq \frac{0.2962}{\sqrt{m}}$$

Proof: As for Theorem 1 replacing the triangle A with the corresponding sector of the circle with radius m; the estimations of area(A) for the three different cases are valid again and area$(C) = m^2\pi/2$. □

For the normal and exponential distributions it is not possible to give a bound only depending on m. A LCG with a very low q_2 and e_1 almost parallel to the y-axis could serve as a counter example as the gap is not in region C. But it is not difficult to prove the following.

Theorem 4: For the ratio of uniforms method with $C = \{(u,v) \mid 0 \leq u \leq 1, \mid v \mid \leq 2u\sqrt{-\log(u)}\}$ (normal distribution) combined with a LCG with modulus m and Beyer quotient q_2 there exists a gap with

$$\mathrm{prob(gap)} > \min\left(\frac{0.4315}{\sqrt{m}}\sqrt{\frac{q_2}{q_2^2+1}}\,,\ 0.04312\right)$$

Proof: Without changing the probability of the largest gap we use $C = \{(u,v) \mid 0 \leq u \leq m, \mid v \mid \leq k(u) = \sqrt{e/2}u\sqrt{-\log(u/m)} \leq m/2\}$ (cf. Figure 2). Between the lines $v = u$ and $v = -u$ the minimal distance from the origin to $k(u)$ is $dmin > 0.6776m$. Following the proof of Theorem 2 at least the vector $e_1 - e_2$ points into an interesting direction. The distance between the grid lines g_1 intersecting the y-axis closest above the origin and g_2 intersecting closest below, both parallel to $e_1 - e_2$ is $dist \geq \sqrt{m\sqrt{3}/2\frac{q_2}{q_2^2+1}}$. If g_1 or g_2 is not intersecting $k(u)$ the probability of the largest gap is larger than the area between $k(u)$ and the line $v = u$ divided through the area of C (≥ 0.04312).
Otherwise we look at the domain A bordered by the curve $k(u)$ and the lines from the origin to $k(u) \cap g_1$ and $k(u) \cap g_2$ and get:

$$\mathrm{prob(gap)} = \frac{\mathrm{area}(A)}{\mathrm{area}(C)} \geq \frac{dist\; dmin/2}{m^2\sqrt{e\pi}/4} > \frac{0.4315}{\sqrt{m}}\sqrt{\frac{q_2}{q_2^2+1}}$$

□

Theorem 5: For the ratio of uniforms method with $C = \{(u,v) \mid 0 \leq u \leq 1, 0 \leq v \leq -2u\log(u)\}$ (exponential distribution) combined with a LCG with modulus m and Beyer quotient q_2 there exists a gap with

$$\mathrm{prob(gap)} > \min\left(\sqrt{\frac{q_2}{m}}\min\left(0.4420\,,\ \frac{0.6362}{\sqrt{q_2^2+0.5177q_2+1}}\right)\,,\ 0.06419\right)$$

Proof: Without changing the probability of the largest gap we use $C = \{(u,v) \mid 0 \leq u \leq m, v \leq k(u) = -e\,u\log(u/m) \leq m\}$ (cf. Figure 3). Between the line $v = u\tan(5\pi/12)$ and the x-axis the minimal distance from origin to $k(u)$ is $dmin > 0.929247m$, which occurs for the line through the origin with ascent 0.43867. The proof is analogous to the proof of Theorem 4. Only the fact that the "interesting" angle is only $5\pi/12$ leads to some additional work to estimate the shortest lattice vector in an "interesting" direction. In the worst case (e_1 and e_2 not in the correct angle) we have to distinguish two cases for the angle α between e_1 and e_2:
Case A: $\pi/2 \leq \alpha \leq 7\pi/12$; there is an integer $l \geq 1$ with $e_2 - le_1$ in the right direction and $|e_2|\tan(\pi/12) \leq l|e_1| \leq 2\tan(\pi/12)|e_2|$; if $q_2 > \tan(\pi/12)$ it is enough to take $l = 1$; we get the estimate $|e_2 - e_1|^2 \leq |e_2|^2(q_2^2 + 0.5177q_2 + 1)$; otherwise we get $|e_2 - le_1|^2 \leq |e_2|^2(1 + (2\tan(\pi/12))^2 - 2\cos(7\pi/12)2\tan(\pi/12)) < 1.56459|e_2|^2$
Case B: $\pi/3 \leq \alpha < \pi/2$; again there is an integer l as above; geometric considerations (e_1 must not intersect the mid-perpendicular of e_2) and easy computations show that $l = 1$

is enough for $q_2 \geq \sqrt{2-\sqrt{3}}$ and that $l|e_1| \leq |e_2| 2\sqrt{2-\sqrt{3}}$; putting together the parts as in the proof of Theorem 4 (area$(C) = m^2 e/4$) completes the proof. □

As the above theorems give only lower bounds we computed the probability of the largest gaps for the ratio of uniforms method combined with some LCGs recommended in literature. These results and the values of the bounds for the Theorems 3 to 5 are contained in Table 1. The results show that for LCGs with $q_2 > 0.7$ the probability of the largest gap is about $0.5/\sqrt{m}$ for any of the three examined distributions.

Table 1: prob(gap) $\sqrt{m}$

m	a	c	cf.	q_2	Cauchy		Normal		Exponential	
					exact	Th. 3	exact	Th. 4	exact	Th. 5
2147483399	40692	0	[13]	0.769	0.553	0.296	0.599	0.300	0.500	0.388
$2^{31}-1$	16807	0	[14],[18]	0.132	0.878	0.296	0.248	0.155	0.256	0.161
$2^{31}-1$	742938285	0	[8]	0.849	0.569	0.296	0.621	0.303	0.647	0.399
2^{32}	69069	1	[15]	0.943	0.569	0.296	0.671	0.305	0.691	0.401
2^{32}	1589013525	1	[17]	0.965	0.479	0.296	0.618	0.305	0.743	0.401

4. Conclusions

The above results can be interpreted in different ways. Depending on the point of view one can conclude that either the ratio of uniforms method or LCGs are bad. That the discrepancy of the ratio of uniforms method can be reduced dramatically by using a uniform generator with a better two-dimensional distribution indicates the below theorem in comparison with Theorem 1 and 2. It is valid for any uniform generator which returns all integer pairs between 0 and m (e.g. multiple recursive congruential generators) and has therefore a one-dimensional discrepancy of $1/m$ itself.

Theorem 6: The ratio of uniforms method with $C = (0, u^+) \times (v^-, v^+)$ combined with a uniform generator that returns all grid points in $[0, m) \times [0, m)$ with the same frequency has a one-dimensional discrepancy D with: $D \leq 2/m$
Proof: Without loss of generality we assume $C = [0, m) \times [v^-, v^+)$ and $v^- + v^+ = m$. We define $A(k) = C \cap \{v | v \leq k \cdot u\}$; $\#A(k)$ = number of grid points in $A(k)$ (the interesting grid has the basis (1,0) and (0,1) and contains the point $(0, v^-)$); $\Psi(k) = (\text{area}(A(k)) - \#A(k))/m^2$; it is easy to give bounds for Ψ as the difference between area and # can only occur in the squares of the grid hit by the line $k \cdot u$. Easy considerations show that we get (when restricting our attention to lines not touching grid points) $-(m + \lfloor v^- \rfloor)/m^2 \leq \Psi(k) \leq 0$ for $k \leq 0$ and $-1/m \leq \Psi(k) \leq \lfloor v^+ \rfloor / m^2$ for $k \geq 0$ which completes the proof if we assume that points on the y-axis are transformed into $+\infty$ or $-\infty$. □

As a second possibility the discrepancy can be reduced by replacing the ratio of uniforms method. Not only the inversion method itself but any rejection method that uses inversion to sample from the dominating density has a much lower one-dimensional discrepancy when combined with LCGs with multiplier a not to small (cf. [4] and [9] for empirical results). As it is pointed out in [4] it is also possible to convert ratio of uniforms into ordinary rejection by generating the table-mountain distribution by inversion. This adaption has low one-dimensional discrepancy but we prefer the use of the transformed rejection algorithms suggested in [9], which have much higher acceptance probabilities and are as simple as the ratio of uniforms method.

The most important conclusion that should be drawn from the results of this paper is: Though LCGs and the ratio of uniforms method are frequently used avoid their combination, especially for applications that need a good one-dimensional resolution of the non-uniform random numbers generated.

References

[1] Afflerbach, L. and Wenzel, K. Normal random numbers lying on spirals and clubs, Statistical Papers 29 (1988), 237-244.

[2] Afflerbach, L. Criteria for the assessment of random number generators, J. Comput. and Applied Math. 31 (1990), 3-10.

[3] Afflerbach, L. Die Gütebewertung von Pseudo-Zufallszahlen-Generatoren aufgrund theoretischer Analysen und algorithmischer Berechnungen, Grazer Mathematische Berichte Nr. 309, Graz, 1990.

[4] 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.

[5] Deng, L. Y. Robustness study of some random variate generators, in: Computer Science and Statistics: 20^{th} Symposium on the Interface, eds. E. J. Wegman et. al. (1988), 624-626.

[6] Devroye, L. Non-Uniform Random Variate Generation, Springer-Verlag, New York, 1986.

[7] Dieter, U. Mathematical aspects of various methods for sampling from classical distributions, in: Proceedings of the 1989 Winter Simulation Conference, eds. E.A. MacNair et al. (1989), 477-483.

[8] Fishman G. S. and Moore L. R. An exhaustive analysis of multiplicative congruential random number generators with modulus $2^{31}-1$, SIAM Journal Sci. Stat. Comput. (1986), 24-45.

[9] 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).

[10] Kinderman, A. J. and Monahan, F. J. Computer generation of random variables using the ratio of uniform deviates, ACM Trans. Math. Soft. 3 (1977), 257-260.

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

[12] Kuipers, L. and Niederreiter, H. Uniform Distribution of Sequences, Wiley, New York, 1974.

[13] P. L'Ecuyer Efficient and portable combined random number generators, Communications of the ACM 31 (1988), 742-749.

[14] Lewis P. A., Goodman A. S. and Miller J. M. A pseudo-random number generator for the system/360, IBM Syst. J. 8, 2 (1969), 136-146.

[15] G. Marsaglia The structure of linear congruential sequences, in: Applications of Number Theory to Numerical Analysis, S. K. Zaremba, ed. , 249-286 Academic Press, New York, 1972.

[16] Neave, R. On using the Box-Muller transformation with multiplicative congruential pseudo-random number generators, Appl. Statist. 22 (1973), 92-97.

[17] Niederreiter, H. Quasi-Monte Carlo methods and pseudo-random numbers, Bulletin of the American Math. Soc. 84, 6 (1978), 957-1041.

[18] Park S. K. and Miller K. W. Random number generators: good ones are hard to find, Communications of the ACM 31 (1988), 1192-1201.

[19] Ripley, B. D. Uses and abuses of statistical simulation, Mathematical Programming 42 (1988), 53-68.

[20] Stadlober, E. Sampling from Poisson, binomial and hypergeometric distributions: Ratio of uniforms as a simple and fast alternative, Berichte der Math.-Stat. Sektion in der Forschungsgesellschaft Joanneum, Nr. 303 , Graz, 1989.