Les Cahiers du GERAD ISSN: 0711–2440

Entropy Tests for Random Number Generators

P. L'Ecuyer, A. Compagner
J.-F. Cordeau

G–96–41

September 1996

Entropy Tests for Random Number Generators

Pierre L'Ecuyer
Université de Montréal and GERAD

Aaldert Compagner
Technical University Delft

Jean-François Cordeau
Université de Montréal

September, 1996

Abstract

Uniformity tests based on a discrete form of entropy are introduced and studied in the context of empirical testing of uniform random number generators. Numerical results are provided. It turns out that some currently used generators fail the tests. The linear congruential and inversive generators with power-of-two modulus perform especially badly.

Résumé

Des tests d'uniformité basés sur l'entropie d'une loi de probabilité discrète sont introduits et étudiés pour tester empiriquement des générateurs de valeurs aléatoires uniformes. Des résultats numériques montrent que plusieurs générateurs couramment utilisés échouent les tests. Les générateurs à congruence linéaire et inversifs dont le module est une puissance de 2 sont particulièrement mauvais face à ces tests.

Introduction

Random number generators should generally be built based on proper *theoretical* analysis and understanding of their structural properties, and then tested *empirically* to further improve one's confidence in them. Different statistical tests are sensitive to different types of deficiencies in generators, so it is useful to apply a wide range of tests. For background on random number generators and statistical testing, see for example [17, 20, 21, 24, 27]. Since by necessity all statistical tests are applied to subsequences of finite length, they can never exclude the possibility that in extensive applications effects arise that escaped detection in testing. In fact, all sequences of numbers of a given length have the same total amount of correlation, as explained in [3, 4]. But what we ask for is that the generators pass a collection of tests that take a reasonable (or practically feasible) amount of computing time.

In this paper, we study uniformity and independence tests based on the concept of entropy for discrete uniform distributions, following the suggestion in [5] that entropy might provide a useful testing ground. In [5], the notation and terminology were taken from statistical mechanics, while here the point of view is that of probability theory and statistics.

Recall that for a discrete random variable X taking its values in a (discrete) set $\mathcal{S}$, with probability mass function $p_x = P[X = x]$ for all $x \in \mathcal{S}$, the *entropy* of p (or of X) is defined by

$$H_d = -\sum_{x \in \mathcal{S}} p_x \lg(p_x), \tag{1}$$

where lg denotes the logarithm in base 2. In particular, consider a string of L independent random bits, each bit being 0 with probability 1/2 and 1 with probability 1/2. There are $C = 2^L$ possible outcomes for the string, each having probability $1/C = 2^{-L}$. Identify each such outcome (or bit string) with the integer x that it represents in binary arithmetic. Then, the random variable X has the *discrete uniform* distribution over $\{0, 1, \ldots, C-1\}$ and its entropy is

$$H_d = -\sum_{x=0}^{C-1} (1/C)\lg(1/C) = L. \tag{2}$$

To test whether a random variable X distributed over $\mathcal{S} = \{0, 1, \ldots, C-1\}$ effectively has the uniform distribution, we might estimate its entropy and compare

the result with (2). Let $X_1, \ldots, X_n$ be a sample of n presumably independent copies of X and for each $x \in \mathcal{S}$, let N_x be the number of times the value x was obtained:

$$N_x = \sum_{i=1}^{n} I[X_i = x] \tag{3}$$

where I denotes the indicator function; that is, $I[X_i = x] = 1$ if $X_i = x$, 0 otherwise. A natural and standard *discrete entropy estimator* (or *empirical entropy*) is:

$$\hat{H}_d(C, n) = -\sum_{x=0}^{C-1} (N_x/n) \lg(N_x/n). \tag{4}$$

For a goodness-of-fit statistical test based on the statistic (4) to be practical, the distribution function of this statistic (or a good approximation of it) must be available. In Section 2, we provide explicit expressions for the exact mean and variance of $\hat{H}_d(C, n)$ under the uniformity assumption. Basharin [1] already showed its asymptotic normality as $n \to \infty$. For finite n, we shall use as an approximation the normal distribution whose mean and variance are the exact values for that n. We study the quality of this approximation by estimating the discrepancy between it and the exact distribution, as a function of n.

Statistical tests based on the entropy of a *continuous* distribution have already been proposed and applied [6, 13, 32]. Those tests are based on more complicated (continuous) entropy estimators and have very little in common with those proposed here. They are also discussed in [23].

In the next section, we explain how the bit strings can be constructed from the output values of a generator that produces real numbers between 0 and 1. We study the distribution of the empirical entropy and propose another test based on the linear correlation between successive values of the entropy. We also introduce a second type of entropy, still defined by (4), but based on all (overlapping) L-bit substrings within a circular n-bit string. From this, we define an average entropy test and a correlation test. In Section 2, we apply a selection of tests to a set of random number generators. This set is small and not necessarily representative of all the different methods proposed in the literature. However, our results certainly show that these entropy tests are powerful enough to detect certain defects in random number generators and are therefore justified.

1 Tests based on the discrete empirical entropy

1.1 Constructing the bit strings and computing entropies

Let $u_1, u_2, u_3, \ldots$ be a sequence of successive output values of some random number generator, which are supposed to behave as independent $U(0,1)$ random variables. We want to test the uniformity and independence of, say, the first ℓ bits of the binary fractional expansion of those u_i's, where ℓ is a positive constant indicating the finite precision. If the binary expansion of u_i to its first ℓ bits is written as

$$u_i = \sum_{j=1}^{\ell} b_{i,j} 2^{-j},$$

then the null hypothesis H_0 to be tested can be formulated as: "the sequence

$$b_{1,1}, \ldots, b_{1,\ell}, b_{2,1}, \ldots, b_{2,\ell}, b_{3,1}, \ldots \tag{5}$$

is a sequence of independent random bits, each taking the value 1 with probability 1/2, independently of the others".

To test H_0, choose two positive integers n and L, extract n disjoint *blocks* (or substrings) of L bits each from this sequence, and compute the empirical entropy $\hat{H}_d(C, n)$ defined in (4). Suppose that this procedure is repeated N times, with disjoint parts of the sequence, and let $T_1, \ldots, T_N$ denote the N values of $\hat{H}_d(C, n)$ thus obtained. We examine the following two ways of testing H_0: (a) construct the empirical distribution of $T_1, \ldots, T_N$ and compare it with the theoretical distribution of $\hat{H}_d(C, n)$ under H_0 and (b) test if there is a significant correlation between the pairs (T_i, T_{i+1}) of successive values of the entropy.

The n bit strings of length L can be extracted from (5) in different ways, depending on the testing strategy that one has in mind. If one is interested in testing only the few most significant bits of each u_i, then one would take a small value of ℓ. For example, with $\ell = 1$, only the most significant bit is tested. On the other hand, to test the least significant bits, one may throw away (say) the r most significant bits of each u_i and extract only the $\ell - r$ bits that remain. The following setup covers these situations. Choose an integer r such that $0 \leq r < \ell$ and let $s = \ell - r$. (To

keep all the bits, just take $r = 0$.) Extract from each u_i the bits $b_{i,r+1}, \ldots, b_{i,r+s}$ and (conceptually) put them in a long string:

$$b_{1,r+1}, \ldots, b_{1,r+s}, b_{2,r+1}, \ldots, b_{2,r+s}, b_{3,r+1}, \ldots. \tag{6}$$

Partition this string into substrings (or blocks) of L consecutive bits, without overlap. To simplify the notation, assume that either s divides L, or L divides s. So, if $s \leq L$, the first substring is

$$b_{1,r+1}, \ldots, b_{1,r+s}, \ldots, b_{L/s,r+1}, \ldots, b_{L/s,r+s},$$

the second one is

$$b_{1+L/s,r+1}, \ldots, b_{1+L/s,r+s}, \ldots, b_{2L/s,r+1}, \ldots, b_{2L/s,r+s},$$

and so on. If $s > L$, then the first s/L blocks are taken from u_1, the next s/L ones from u_2, and so on. Later on, we will also consider taking the blocks with overlap.

1.2 The distribution of the sample entropy

To compare the empirical distribution of $T_1, \ldots, T_N$ with the theoretical one, one needs the distribution of $\hat{H}_d(C, n)$ under H_0, or at least a good approximation of it. Such an approximation is given by the next proposition. It says that for large n, $\hat{H}_d(C, n)$ is approximately normally distributed, and it also gives the exact mean and variance of $\hat{H}_d(C, n)$ (for any finite n).

PROPOSITION **1.** *Under the hypothesis H_0, one has:*

$$E[\hat{H}_d(C,n)] = -C \sum_{j=0}^{n} \frac{j}{n} \lg\left(\frac{j}{n}\right) \binom{n}{j} \frac{(C-1)^{n-j}}{C^n}, \tag{7}$$

$$\begin{aligned} \text{Var}\,[\hat{H}_d(C,n)] &= C \sum_{j=0}^{n} \left(\frac{j}{n} \lg\left(\frac{j}{n}\right)\right)^2 \binom{n}{j} \frac{(C-1)^{n-j}}{C^n} \\ &\quad + C(C-1) \sum_{j=0}^{n} \sum_{k=0}^{n} \frac{j}{n} \lg\left(\frac{j}{n}\right) \frac{k}{n} \lg\left(\frac{k}{n}\right) \binom{n}{j} \binom{n-j}{k} \frac{(C-2)^{n-j-k}}{C^n} \\ &\quad - E^2[\hat{H}_d(C,n)], \end{aligned} \tag{8}$$

and

$$S(C,n) \stackrel{\text{def}}{=} \frac{\hat{H}_d(C,n) - E[\hat{H}_d(C,n)]}{\sqrt{\text{Var }[\hat{H}_d(C,n)]}} \Rightarrow N(0,1) \tag{9}$$

as $n \to \infty$ *for fixed* C.

Proof. Note that each N_x is a binomial, with

$$P[N_x = j] = \binom{n}{j} \left(\frac{1}{C}\right)^j \left(1 - \frac{1}{C}\right)^{n-j}.$$

Replacing this in the definition of $E[\hat{H}_d(C,n)]$ yields (7). Similarly, for $1 \le x \le x' \le C$, one has

$$P[N_x = j, N_{x'} = k] = \binom{n}{j} \binom{n-j}{k} \left(\frac{1}{C}\right)^j \left(\frac{1}{C}\right)^k \left(1 - \frac{2}{C}\right)^{n-j-k}.$$

Equation (8) follows easily by using this in the definition of Var $[\hat{H}_d(C,n)] = E[\hat{H}_d^2(C,n)] - (E[\hat{H}_d(C,n)])^2$. Basharin [1] has shown the asymptotic normality of the estimator $\hat{H}_d(C,n)$, for fixed C, using its Taylor expansion in terms of $(\hat{p}_1 - p_1, \ldots, \hat{p}_C - p_C)$. His result is stated in the context of a general discrete distribution over a finite set, using asymptotic expressions with $O(n^{-2})$ error for the mean and variance. Here, we replace these approximations by the exact values. Note that (slightly more complicated) exact expressions for the mean and variance were also given for the case of a more general discrete distribution in [15], using a development based on the generating function of the multinomial distribution. □

This proposition provides the ingredients for a goodness-of-fit test based on the discrete empirical entropy: generate N independent values of $S(C,n)$, say $S_1, \ldots, S_N$, and compare their empirical distribution to the $N(0,1)$. These S_i's are in fact the normalized values of the T_i's introduced in the previous subsection: Under H_0, $S_i = (T_i - E[T_i])/(\text{Var }[T_i])^{1/2}$. Table 1 gives the exact mean and standard deviation of $\hat{H}_d(C,n)$ under H_0, computed from (7) and (8), for different values of L, and $n = C = 2^L$. As L increases, these values become rather costly to compute because of the double sum in (8).

Table 1: Mean and standard deviation of the sample entropy for $C = n = 2^L$.

L	$E[\hat{H}_d(C,n)]$	$(\text{Var }[\hat{H}_d(C,n)])^{1/2}$
1	0.50000	0.50000
2	1.32399	0.38950
3	2.24579	0.28677
4	3.20868	0.20647
5	4.19057	0.14725
6	5.18163	0.10455
7	6.17718	0.07408
8	7.17497	0.05244
9	8.17386	0.03710
10	9.17331	0.02624
11	10.17303	0.01856
12	11.17289	0.01312
13	12.17282	0.00928
14	13.17279	0.00656
15	14.17277	0.00464
16	15.17276	0.00328

1.3 Testing the goodness-of-fit via a KS statistic

A *Kolmogorov-Smirnov* (KS) test can be used to compare the empirical distribution of $S_1, \ldots, S_N$ to the standard normal distribution. It works as follows (see [17, 31] for more details). Let $S_{(1)} \leq \ldots \leq S_{(N)}$ be the N ordered values of $S_1, \ldots, S_N$ and Φ be the distribution function of the standard normal: $\Phi(z) = P[N(0,1) \leq z]$. Define

$$D_N^+ \;=\; \max_{1 \leq j \leq N} (j/N - \Phi(S_{(j)}))$$

and

$$D_N^- \;=\; \max_{1 \leq j \leq N} (\Phi(S_{(j)}) - (j-1)/N).$$

Approximations of the distribution of D_N^+ and D_N^- under H_0 are given in [17, 31] and other references there. Let d^+ and d^- be the values taken by D_N^+ and D_N^- in a given experiment and define the corresponding *significance levels* as $\delta^+ = P[D_N^+ > d^+]$ and $\delta^- = P[D_N^- > d^-]$, respectively. The hypothesis H_0 is *rejected* if one of the

significance levels, δ^+ or δ^-, is extremely close to zero or one. In case of doubt, one may replicate the entire test (independently) and reject H_0 if the significance levels are consistently too close to zero or one. We call this the *discrete entropy distribution test*.

1.4 How good is the normal approximation ?

For finite n, the $N(0,1)$ distribution is only an approximation of the exact distribution of $S(C,n)$. The quality of this approximation must be controlled, because even if the empirical distribution of $S_1,\ldots,S_N$ is very close to the true distribution of $S(C,n)$ under H_0, if the latter distribution is too far from the standard normal, then D_N^+ or D_N^- will take a large value. In other words, if the approximation is not good enough, the KS test may detect the approximation error and reject the generator because of this.

The distribution of $\hat{H}_d(C,n)$ is asymmetric about its mean, but closer to symmetry as $n \to \infty$. The normal approximation is poor for small n also because of larger jumps in the distribution. When n is small compared to C, only a few bit strings are observed more than once, so most values of N_x are 0 or 1. As a result, the distribution of $\hat{H}_d(C,n)$ is concentrated on just a few values and is thus far from normal.

To assess the quality of the approximation for moderately large n, we made the following empirical investigation. Our aim is to bound the maximum vertical distances between the true distribution function of $S(C,n)$ under H_0 and its normal approximation. More specifically, we would like to know the values of

$$\Delta^+(C,n) \stackrel{\text{def}}{=} \sup_{0\le u\le 1}(P[\Phi(S(C,n)) \le u] - u),$$
$$\Delta^-(C,n) \stackrel{\text{def}}{=} \sup_{0\le u\le 1}(u - P[\Phi(S(C,n)) \le u]).$$

Computing these error bounds exactly is too difficult in general, so we shall settle for estimators. It turns out that the most natural estimators of $\Delta^+(C,n)$ and $\Delta^-(C,n)$ are $\hat{\Delta}^+(C,n) = D_N^+$ and $\hat{\Delta}^-(C,n) = D_N^-$, respectively, for N as large as possible. These estimators can be computed by standard Monte Carlo simulation,

using "reliable" random number generators (for the results reported here, we checked our experiments with several random number generators of different types, and the results agreed). The percentage points of the KS distribution can be used to compute confidence intervals for $\Delta^+(C,n)$ and $\Delta^-(C,n)$, as explained in Section 4.5.4 of [30]. For a 95% confidence level, for example, the half-widths of the confidence intervals are approximately 0.043 for $N = 10^3$ and 0.004 for $N = 10^5$.

Table 2 gives the estimates $\hat{\Delta}^+(C,n)$ and $\hat{\Delta}^-(C,n)$ obtained with $N = 10^5$, for different values of L and n, with $n = C = 2^L$ and $n = C^2 = 2^{2L}$. The intersection of the interval $[\hat{\Delta}^+(C,n) - 0.004,\ \hat{\Delta}^+(C,n) + 0.004]$ with the interval $[0,1]$ yields a 95% confidence interval for $\Delta^+(C,n)$, and similarly for $\Delta^-(C,n)$. The approximation error clearly decreases with n as n and L increase simultaneously, and also seems to decrease with L for fixed n. For $n = C \geq 2^{10}$ and for $n = C^2 \geq 2^{16}$, the values observed in the table are mostly noise, in the sense that the confidence intervals on the error bounds contain zero.

Table 2: Values of $\hat{\Delta}^+(C,n)$ and $\hat{\Delta}^+(C,n)$

	$n = C = 2^L$			$n = C^2 = 2^{2L}$		
L	n	$\hat{\Delta}^+(C,n)$	$\hat{\Delta}^-(C,n)$	n	$\hat{\Delta}^+(C,n)$	$\hat{\Delta}^-(C,n)$
2	2^2	0.233	0.328	2^4	0.141	0.175
4	2^4	0.051	0.091	2^8	0.030	0.052
6	2^6	0.011	0.024	2^{12}	0.009	0.021
8	2^8	0.003	0.011	2^{16}	0.003	0.004
10	2^{10}	0.002	0.004			
12	2^{12}	0.001	0.003			
14	2^{14}	0.002	0.003			
16	2^{16}	0.001	0.003			

What are the practical effects of these approximation errors and how can one use the table to choose test parameters ? As an illustration, consider a test with parameters $N = 1000$ and $n = C > 1000$. With this N, for the significance level δ^+ to be smaller than 0.01 (say), D_N^+ must be larger than 0.05 (approximately). Since the approximation error $\Delta^+(C,n)$ in this case appears certainly smaller than 0.01, a value of D_N^+ larger than 0.05 should be caused by something else than this approximation

error. The bias on the significance level is acceptable. This is more so as n gets larger. The same also applies to D_N^-. Being conservative, for $N = 1000$ and $n = C \geq 2^{12}$, or for $N = 10000$ and $n = C \geq 2^{15}$, one can safely neglect the approximation error.

1.5 A correlation test

A second way to test H_0 is to check whether the pairs (T_i, T_{i+1}), for $1 \leq i \leq N-1$, are significantly correlated. If they are, it means that a low [high] entropy in one part of the sequence (5) tends to be followed by a low [high] entropy in the next part. Equivalently, one may test the correlation between the pairs (S_i, S_{i+1}), which is more convenient because the S_i have zero mean and unit variance. The sample correlation between the S_i's is simply

$$\hat{\rho}_N = \frac{1}{N-1} \sum_{i=1}^{N-1} S_i S_{i-1}. \tag{10}$$

Under H_0, as $N \to \infty$, $\hat{\rho}_N$ converges to zero with probability one and $\sqrt{N}\hat{\rho}_N$ converges in distribution to the $N(0,1)$. A statistical test readily follows from the latter property: compute $\sqrt{N}\hat{\rho}_N$ for a large value of N and reject H_0 if it is too far away from zero to be considered as a typical $N(0,1)$ variate. We call this the *discrete entropy correlation test*. Here, n can be small but N must be large, which is the opposite as for the discrete entropy distribution test.

1.6 Constructing the bit strings with overlap

In Section 1.1, the n bit strings of length L were constructed without overlap; that is, from disjoint parts of the sequence (6). We now consider a setup in which they are constructed with overlap. Take the first n bits of the sequence (6), relabel them as $b_1, \ldots, b_n$, and put them in a circle (i.e., define $b_0 = b_n$ and $b_j = b_{j \bmod n}$ for all integers j). For $i = 1, \ldots, n$, let X_i be the integer represented by the bit string of length L starting at position i:

$$X_i = \sum_{j=0}^{L-1} b_{i+L-j-1} 2^j. \tag{11}$$

Define N_x and the empirical entropy $\hat{H}_d(C,n)$ as in (3) and (4), and let T_1 be the value of $\hat{H}_d(C,n)$ thus obtained. Repeat the same procedure with the bits $n+1$ to $2n$ of the sequence (6), yielding an entropy value T_2, and so on. So, for $i \geq 1$, T_i is the value of $\hat{H}_d(C,n)$ obtained by putting the bits $(i-1)n+1,\ldots,in$ in a circle and looking at all n strings of L consecutive bits over that circle.

A possible advantage of this overlapping variant is that is squeezes more information from the bit string (6) compared to the non-overlapping case. However, in the overlapping case, N_x is no longer a binomial random variable and the mean and variance formulæ of Proposition 1 no longer apply. For small values of n, one can compute the exact mean and variance directly from their definitions (which involves a sum of 2^n terms corresponding to the 2^n possibilities for $\{b_1,\ldots,b_n\}$). The exact values for some pairs (L,n) are reported in Table 3. With these values in hand, one can compute $S_1,\ldots,S_N$, the N values of $S(C,n)$, as before. For large N,

$$\sqrt{N}\bar{S}_N = N^{-1/2}\sum_{i=1}^{N} S_i$$

is approximately $N(0,1)$ under H_0. The *overlapping average entropy test* computes this statistic to test the empirical mean of the entropy against its theoretical value. It rejects H_0 if $\sqrt{N}|\bar{S}_N|$ is too large.

Table 3: Mean and variance of the overlapping entropy for some pairs (L,n).

L	n	$E[\hat{H}_d(C,n)]$	Var $[\hat{H}_d(C,n)]$
2	4	1.375000	0.3593750
3	8	2.299772	0.1867293
4	16	3.238725	0.1007388
5	20	3.817000	0.0815392
5	25	4.014291	0.0694637
5	30	4.160005	0.0591489

The correlation test of the previous subsection can also be applied in the same way; we call this version the *overlapping entropy correlation test*. For large n (say, $n > 30$), the exact mean and variance of T_i take too much time to compute in reasonable time, but one can simply replace them by their sample counterparts. The sample correlation between $T_1,\ldots,T_N$ then becomes:

$$\hat{\rho}_N = \frac{(N-1)^{-1}\sum_{i=1}^{N-1} T_i T_{i+1} - (N-1)(\bar{T})^2}{\hat{\sigma}_T^2} \tag{12}$$

where $\bar{T}$ and $\hat{\sigma}_T^2$ are the sample mean and sample variance of the T_i's:

$$\bar{T} = \frac{1}{N}\sum_{i=1}^{N} T_i$$

and

$$\hat{\sigma}_T^2 = \frac{1}{N-1}\sum_{i=1}^{N}(T_i - \bar{T})^2.$$

Under H_0, since the T_i's are i.i.d., $\sqrt{N}\hat{\rho}_N$ again converges in distribution to the $N(0,1)$ as $N \to \infty$, so it can be used for a correlation test in the same way as (10) for large N.

2 Experimental results

2.1 A selection of random number generators

We selected a few popular or recently-proposed random number generators, listed in Table 4, and submitted them to entropy tests. Of course, this list is not exhaustive; there are several more good and poor generators that we could test. However, our aim here is not to test all known generators, nor to recommend any specific generator, but rather to investigate the power of entropy-based tests to detect deficiencies in certain types of generators.

First, we observe that the generators G1 to G12 are rather "baby" generators from our point of view: we think that their period lengths are much to short for current needs, so none of them can be recommended for general use. G13 to G16, on the other hand, have reasonable period lengths, althought G13 and G14 have well-documented statistical defects [4, 5, 26].

The generators G1 to G7 are well-known linear congruential generators (LCGs), based on a recurrence of the form $x_i = (ax_{i-1} + c) \bmod m$, with output $u_i = x_i/m$ at step i. G1 and G2 are recommended by Fishman and Moore [12], and Law and Kelton [18], respectively. G3 and G4 are in several software packages [2, 29], G5 is

Table 4: List of selected generators.

G1.	LCG with $m = 2^{31} - 1$ and $a = 742938285$.
G2.	LCG with $m = 2^{31} - 1$ and $a = 630360016$.
G3.	LCG with $m = 2^{31} - 1$ and $a = 16807$.
G4.	LCG with $m = 2^{32}$, $a = 69069$, and $c = 1$.
G5.	LCG with $m = 2^{31}$ and $a = 65539$.
G6.	LCG with $m = 2^{31}$ and $a = 452807053$.
G7.	LCG with $m = 2^{31}$, $a = 1103515245$, $c = 12345$.
G8.	Implicit inversive with $m = 2^{31} - 1$ and $a_1 = a_2 = 1$.
G9.	Explicit inversive with $m = 2^{31} - 1$ and $a = b = 1$.
G10.	Implicit inversive with $m = 2^{32}$, $a = b = 1$, and $z_0 = 5$.
G11.	Explicit inversive of [11] with $m = 2^{32}$, $a = 6$, and $b = 1$.
G12.	Modified explicit inversive of [9] with $m = 2^{32}$, $a = 6$, and $b = 1$.
G13.	GFSR-521 in the Appendix of [29].
G14.	GFSR proposed in [16].
G15.	Combined LCG in Fig. 3 of [19].
G16.	Combined MRG in Fig. 1 of [22].

the infamous RANDU, G6 corresponds to the URN12 generator of [7], and G7 is the LCG implemented in the `rand` function of the standard library of the C programming language [28]. The next five generators are inversive generators modulo m. Their output at step i is always $u_i = z_i/m$. G8 is an *implicit* inversive generator of the form $z_i = (a_1 + a_2 z_{n-1}^{-1}) \bmod m$, where $0^{-1} \bmod m$ is defined as 0 (see [8]). G9 is an *explicit* inversive generator of the form $x_i = (ai + b) \bmod m$, $z_i = x_i^{-1} \bmod m = x_i^{m-2} \bmod m$ [8, 14]. G10 is an implicit inversive generator with power-of-two modulus $m = 2^e$, based on the recurrence: $z_i = T(z_{i-1})$ where $T(2^\ell z) = (a_1 + 2^\ell a_2 z^{-1}) \bmod 2^e$ for odd z (see [10]). G11 and G12 are explicit inversive generators with power-of-two modulus; G11 is defined in [11] and G12 is defined as in [9], with the recurrence: $z_i = i(ai + c)^{-1} \bmod 2^e$. G13 is the GFSR generator based on the recurrence $x_i := x_{i-521} \oplus x_{i-32}$, where $\oplus$ denotes the bitwise exclusive-or, and with the initialization procedure given in the Appendix of Ripley [29]. G14 is another GFSR generator, given in Kirkpatrick and Stoll [16]. G15 and G16 are the combined LCG of L'Ecuyer [19] and the combined MRG given in Figure 1 of [22].

2.2 Results of discrete entropy distribution tests

We now report on a few experiments with the selected generators, using the entropy distribution test based on the statistic $S(C, n)$. We selected 9 parameter sets (or tests) with the generators of Table 4. Those parameter sets, called S1 to S9, are given in Table 5. The last column of the table gives the total number of calls to the generator for each test.

Table 5: Parameters for entropy distribution tests

Test	N	n	L	r	s	Nb. Calls
S1	1000	2^{12}	12	0	12	4 048 000
S2	1000	2^{12}	12	0	4	12 144 000
S3	1000	2^{12}	12	20	4	12 144 000
S4	1000	2^{16}	8	0	8	65 536 000
S5	1000	2^{16}	8	0	4	131 072 000
S6	1000	2^{16}	8	20	4	131 072 000
S7	1000	2^{16}	16	0	16	65 536 000
S8	1000	2^{16}	16	0	4	262 144 000
S9	1000	2^{16}	16	20	4	262 144 000

For each combination of generator and parameter set (or test), we computed the significance levels δ^+ and δ^- of the KS statistics for the entropy distribution tests. Table 6 reports the highly suspect significance levels; that is, those smaller than 0.01 or larger than 0.99. The other entries are left blank. The generators not mentioned in the table had no suspect significance levels (in this sense) for these tests. We also computed (in parallel) the sample correlation $\hat{\rho}_N$ in (10) for the same sets of parameters. The results of these entropy correlation tests were consistent with those of the distribution tests, in the sense that clear failures were observed for the same combinations of parameter set and generator.

All the LCGs with power-of-two modulus fail. Their failure for the tests based on the least significant bits (S3, S6, S9) was expected, because these bits have short period length for those generators. However, they also fail the tests based on the other bits. The inversive generators with power-of-two moduli also fail the tests based on the least significant bits, which can also be explained by the short period length of these bits. G5 (RANDU) is the only generator failing the test S2, which constructs 12-bit strings by taking 4 bits from each of 3 successive values. This is to be expected,

Table 6: Results of the entropy distribution tests

Test	Side	G1	G4	G5	G6	G7	G10	G11	G12
S1	δ^+								
	δ^-								
S2	δ^+			$< 10^{-10}$					
	δ^-			1.0000					
S3	δ^+		1.0000	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	1.0000
	δ^-		$< 10^{-10}$	1.0000	1.0000	1.0000	1.0000	1.0000	$< 10^{-10}$
S4	δ^+			7.0E-4	5.7E-5				
	δ^-				$< 10^{-10}$				
S5	δ^+				$< 10^{-10}$		0.0010		
	δ^-			$< 10^{-10}$	$< 10^{-10}$		0.9948		
S6	δ^+		$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$
	δ^-		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
S7	δ^+		0.9987	0.9926	0.9993				
	δ^-	0.0017		$< 10^{-10}$	$< 10^{-10}$				
S8	δ^+			$< 10^{-10}$	1.0000				
	δ^-			1.0000	$< 10^{-10}$				
S9	δ^+		$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$
	δ^-		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

because RANDU is known to be bad with respect to the equidistribution of its triples of successive values (they all lie in 16 equidistant hyperplanes [17, 18]). The test S2 successfully detects this. Besides the generators with power-of-two moduli, only G1 has a suspect significance level, for test S7.

2.3 Results of entropy tests with overlapping

Tables 8 and 9 report results of the average entropy test and entropy correlation test with overlapping. The test parameters are given in Table 7. The total number of calls to the generator for these tests is smaller than for the entropy distribution tests of Table 5, but we are doing more work with each number, so the computational times are roughly of the same order. They seem to detect as much (at least for these examples) as the tests of Table 5. The first four tests look at the 30 most significant bits of each number and compute the corresponding overlapping entropy. The last four take the bits 21 to 23 of each number, and 10 successive output values are used to construct each block of 30 bits. So, in this case, the testing concentrates on these 3 bits.

Table 7: Parameters for the entropy tests with overlapping

Test	N	n	L	r	s	Nb. Calls
C1	10^4	30	5	0	30	10 000
C2	10^5	30	5	0	30	100 000
C3	10^6	30	5	0	30	1 000 000
C4	10^7	30	5	0	30	10 000 000
C5	10^4	30	5	20	3	100 000
C6	10^5	30	5	20	3	1 000 000
C7	10^6	30	5	20	3	10 000 000
C8	10^7	30	5	20	3	100 000 000

Table 8: Results of the overlapping average entropy test

Test	G4	G5	G6	G7	G9	G10	G11	G12
C1		6.26E-8			1.0000			
C2	0.0106	$< 10^{-10}$	2.50E-7		0.9999			
C3		$< 10^{-10}$	$< 10^{-10}$					
C4		$< 10^{-10}$	$< 10^{-10}$					
C5		1.0000	$< 10^{-10}$	0.0001			1.0000	1.0000
C6		1.0000	$< 10^{-10}$	$< 10^{-10}$		1.0000	1.0000	1.0000
C7	0.9993	1.0000	$< 10^{-10}$	$< 10^{-10}$		1.0000	1.0000	1.0000
C8	1.0000	1.0000	$< 10^{-10}$	$< 10^{-10}$		1.0000	1.0000	1.0000

Table 9: Results of the overlapping entropy correlation test

Test	G4	G5	G6	G7	G9	G10	G11	G12	G13
C1		$< 10^{-10}$			$< 10^{-10}$	$< 10^{-10}$		1.6E-4	
C2	0.0056	$< 10^{-10}$	0.0015		4.1E-10	$< 10^{-10}$			0.9997
C3	0.0009	$< 10^{-10}$	$< 10^{-10}$	0.9993	1.83E-5	$< 10^{-10}$	6.7E-5		
C4	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	1.0000		$< 10^{-10}$	$< 10^{-10}$	1.1E-6	
C5			1.0000	0.0007		0.9999	$< 10^{-10}$		
C6			1.0000	$< 10^{-10}$		1.0000	$< 10^{-10}$		
C7	1.0000		1.0000	$< 10^{-10}$		1.0000	$< 10^{-10}$		
C8	1.0000		1.0000	$< 10^{-10}$		1.0000	$< 10^{-10}$		1.2E-7

Again, all the generators with power-of-two moduli clearly fail the tests. Most fail even with rather small sample sizes. For the average entropy tests, the inversive generators with power-of-two moduli only fail the tests based on low order bits (the average entropy tends to be too low), but for the correlation tests, they also fail for the high order bits. The explicit inversive generator G9 also fails, but only with small sample sizes for the high order bits. This may appear curious at first sight, but can be explained as follows: the first n values produced by this generator are the inverses (modulo $2^{31} - 1$) of the first n positive integers, divided by $2^{31} - 1$, and it turns out that the inverses of smaller integers tend to have lower entropy for their high order bits. We applied the same tests to the explicit inverse generator with parameter values $m = 2^{31} - 1$, $a = 1$ and $b = 993652$ (the last value was chosen randomly), and it passed. The GFSR recommended by Ripley (G15) fails the last entropy correlation test.

We actually made more experiments with these tests than what is reported here. For example, we tried the same tests as C5–C8, but with $r = 0$ instead of $r = 20$ (i.e., testing the 3 most significant bits), and replicated three times the entire set of tests. The GFSR generators G13 and G14 failed some of the tests (at significance levels less than 10^{-5}) in some of the replications, but passed in others. So, these tests often detect that there is something wrong with those generators, but not always (depending on the initial seed of the generator). However, for most of the results reported here, the spectacular failures (significance levels of 1.0000 or $< 10^{-10}$) observed for one seed are typically observed for almost any random seed. One exception is the generator G9, for which changing the seed changes the behavior as explained previously.

3 Conclusion

The entropy tests reported here are powerful to detect defects in linear and nonlinear generators with power-of-two moduli. The tests reject those generators after looking at only a small fraction of the period length. They also show problems with the explicit inversive and GFSR generators, whose behavior is sensitive to their initial state. The GFSR generators are know to have important weaknesses, such as poor bit-mixing properties [5, 25, 26], but for most initial seeds of the generators, those defects do not show up in the entropy tests. To our knowledge, this paperis the first

to report problems with inversive generators based on empirical testing.The linear congruential generators with prime moduli considered here fail other tests [20], but not these entropy tests.

Of course, if one increases the sample sizes, all generators will eventually fail, because of their finite period length and because of the conservation law for the total amount of correlation valid for all finite sequences [5]. But if the generator is well-designed and has long-enough period, a test may require a (practically) infeasible amount of computing time before failure occurs.

Ideally, meaningful statistical tests should be sensitive to the weaknesses that are regarded most harmful in arbitrary applications. However, without restricting the class of admissible applications, this is an elusive requirement. General purpose random number generators should pass a rich battery of statistical tests of different types. Since entropy is one of the most fundamental measures of randomness, entropy tests are certainly a useful addition to the existing collection of tests for random number generators.

This work has been supported by NSERC-Canada grant # ODGP0110050, FCAR-Québec grant # 93ER1654, and by the Dutch grants NWO B62-424 and STW-DTI66.4085.

References

[1] G. P. Basharin. On a statistical estimate for the entropy of a sequence of independent random variables. *Theory of Probability and its Applications*, 4:333–336, 1959. Translated from Russian.

[2] P. Bratley, B. L. Fox, and L. E. Schrage. *A Guide to Simulation*. Springer-Verlag, New York, second edition, 1987.

[3] A. Compagner. Definitions of randomness. *American Journal of Physics*, 59:700–705, 1991.

[4] A. Compagner. The hierarchy of correlations in random binary sequences. *Journal of Statistical Physics*, 63:883–896, 1991.

[5] A. Compagner. Operational conditions for random number generation. *Physical Review E*, 52(5-B):5634–5645, 1995.

[6] E. J. Dudewicz and E. C. van der Meulen. Entropy-based tests of uniformity. *Journal of the American Statistical Association*, 76(376):967–974, 1981.

[7] E. J. Dudewicz, E. C. van der Meulen, M. G. SriRam, and N. K. W. Teoh. Entropy-based random number evaluation. *Americal Journal of Mathematical and Management Sciences*, 15:115–153, 1995.

[8] J. Eichenauer-Herrmann. Inversive congruential pseudorandom numbers: A tutorial. *International Statistical Reviews*, 60:167–176, 1992.

[9] J. Eichenauer-Herrmann. Modified explicit inversive congruential pseudorandom numbers with power-of-two modulus. *Statistics and Computing*, 6:31–36, 1996.

[10] J. Eichenauer-Herrmann and H. Grothe. A new inversive congruential pseudorandom number generator with power of two modulus. *ACM Transactions on Modeling and Computer Simulation*, 2(1):1–11, 1992.

[11] J. Eichenauer-Herrmann and K. Ickstadt. Explicit inversive congruential pseudorandom numbers with power of two modulus. *Mathematics of Computation*, 62(206):787–797, 1994.

[12] G. S. Fishman and L. S. Moore III. An exhaustive analysis of multiplicative congruential random number generators with modulus $2^{31} - 1$. *SIAM Journal on Scientific and Statistical Computing*, 7(1):24–45, 1986.

[13] D. V. Gokhale. On entropy-based goodness-of-fit tests. *Computational Statistics and Data Analysis*, 1:157–165, 1983.

[14] P. Hellekalek. Inversive pseudorandom number generators: Concepts, results, and links. In C. Alexopoulos, K. Kang, W. R. Lilegdon, and D. Goldsman, editors, *Proceedings of the 1995 Winter Simulation Conference*, pages 255–262. IEEE Press, 1995.

[15] K. Hutcheson and L. R. Shenton. Some moments of an estimate of Shannon's measure of information. *Communications in Statistics*, 3(1):89–94, 1974.

[16] S. Kirkpatrick and E. Stoll. A very fast shift-register sequence random number generator. *Journal of Computational Physics*, 40:517–526, 1981.

[17] D. E. Knuth. *The Art of Computer Programming, Volume 2: Seminumerical Algorithms*. Addison-Wesley, Reading, Mass., second edition, 1981.

[18] A. M. Law and W. D. Kelton. *Simulation Modeling and Analysis*. McGraw-Hill, New York, second edition, 1991.

[19] P. L'Ecuyer. Efficient and portable combined random number generators. *Communications of the ACM*, 31(6):742–749 and 774, 1988. See also the correspondance in the same journal, 32, 8 (1989) 1019–1024.

[20] P. L'Ecuyer. Testing random number generators. In *Proceedings of the 1992 Winter Simulation Conference*, pages 305–313. IEEE Press, Dec 1992.

[21] P. L'Ecuyer. Uniform random number generation. *Annals of Operations Research*, 53:77–120, 1994.

[22] P. L'Ecuyer. Combined multiple recursive generators. *Operations Research*, 1996. To appear.

[23] P. L'Ecuyer. Tests based on sum-functions of spacings for uniform random numbers. In preparation, 1996.

[24] G. Marsaglia. A current view of random number generators. In *in Computer Science and Statistics, Sixteenth Symposium on the Interface*, pages 3–10, North-Holland, Amsterdam, 1985. Elsevier Science Publishers.

[25] M. Matsumoto and Y. Kurita. Twisted GFSR generators II. *ACM Transactions on Modeling and Computer Simulation*, 4(3):254–266, 1994.

[26] M. Matsumoto and Y. Kurita. Strong deviations from randomness in m-sequences based on trinomials. *ACM Transactions on Modeling and Computer Simulation*, 6(2), 1996. To appear.

[27] H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*, volume 63 of *SIAM CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, Philadelphia, 1992.

[28] P. J. Plauger. *The Standard C Library*. Prentice Hall, Englewood Cliffs, New Jersey, 1992.

[29] B. D. Ripley. Thoughts on pseudorandom number generators. *Journal of Computational and Applied Mathematics*, 31:153–163, 1990.

[30] M. S. Stephens. Tests based on EDF statistics. In R. B. D'Agostino and M. S. Stephens, editors, *Goodness-of-Fit Techniques*. Marcel Dekker, New York and Basel, 1986.

[31] M. S. Stephens. Tests for the uniform distribution. In R. B. D'Agostino and M. S. Stephens, editors, *Goodness-of-Fit Techniques*, pages 331–366. Marcel Dekker, New York and Basel, 1986.

[32] O. Vasicek. A test for normality based on sample entropy. *Journal of the Royal Statistical Society: Series B*, 38:54–59, 1976.