SIAM J. APPL. MATH.
Vol. 62, No. 6, pp. 1980–2011

OCCUPANCY NUMBERS IN TESTING RANDOM NUMBER GENERATORS*

A. FIGOTIN†, A. GORDON‡, S. MOLCHANOV‡, J. QUINN‡, AND N. STAVRAKAS‡

Abstract. The classical occupancy problem where n balls are placed in N cells is used for testing of random number generators. We show that the statistics of appropriately chosen occupancy numbers are incompatible with the statistics of many pseudorandom number generators (PRNGs) even if they are truncated. More than that, the incompatibility shows up on relatively small samples long before the period of the PRNG is reached. We introduce generalized Fermi–Dirac models as idealized models for PRNGs. These models are used to make educated guesses of how large the sample sizes should be to detect the deficiencies of the PRNGs under study. We use the developed occupancy tests together with some other ones to examine the performance of several widely used random number generators, including *random()* of the UNIX C library. We found that *random()* failed two of the conducted tests rather badly. We also tested a true random number generator based on α-decay which passed all our relevant tests successfully.

Key words. occupancy numbers, testing, random number generators, pseudorandom number generators

AMS subject classifications. 65C10, 65C50, 65C60, 60F99

PII. S0036139900366869

1. Introduction.

The testing of long sequences of random numbers has grown in importance in pace with the number of applications of such sequences. Some important areas of application include all applications employing Monte Carlo methods, cryptography, and financial mathematics. Some indication of the range of applications of and interest in random sequences is evidenced by [5], [7], [17], [19], [20], [24], [25], [26], [27], and [34]. The development of parallel computing has significantly raised the requirements on the quality of randomizers (random number generators (RNGs)); see, for example, [1] and [2]. Requirements for quality and speed of RNGs in statistical mechanics applications [4] and in cryptographic applications (cf. [28]) can be extremely stringent.

For the most part, so-called pseudorandom number generators (PRNGs) are used to provide "random" numbers. Such generators are algorithmic and the numbers they produce are, of course, not truly random and can, at best, substitute for true random numbers with some level of success. Over the years, many researchers have found serious deficiencies in some well-known PRNGs; see, for example, [16, sections 3.3.2, 3.3.4 (*RANDU*), 3.6], [2], [11], [12], [21], [23], [29], [31], [35].

Our interest in testing RNGs was initially due to our work in developing a true random number generator (TRNG) based on radioactive decay. One problem with the

*Received by the editors January 13, 2000; accepted for publication (in revised form) January 24, 2002; published electronically July 3, 2002. The research of the first three authors was sponsored by the Air Force Office of Scientific Research, Air Force Materials Command, USAF, under grant number F49620-97-1-0229. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research or the U.S. Government.

http://www.siam.org/journals/siap/62-6/36686.html

†Department of Mathematics, University of California at Irvine, Irvine, CA 92697-3875 (afigotin@math.uci.edu).

‡Department of Mathematics, University of North Carolina at Charlotte, Charlotte, NC 28223 (aygordon@email.uncc.edu, smolchan@math.uncc.edu, jequinn@uncc.edu, nstvrks@email.uncc.edu).

testing of RNGs in general is in determining any particular set of tests that is sufficient to certify any given RNG. One reasonable approach suggested by Marsaglia in [23] is to combine tests that are derived for particular application areas with some battery of tests, like the ones he describes in [23]. The methods we employ in section 4 below shed at least some light on this question. They demonstrate that only certain statistics can be expected to detect certain flaws. Furthermore, even when one has chosen the "right" statistic, the sample size is of critical importance. Another problem with testing RNGs is that up to now essentially all tests have been based on asymptotics. One picks a statistic, shows it is asymptotically normal or Poisson or some other distribution, and then performs tests based on the asymptotic statistics. The difficulty here is that good control on errors is almost never available and hence the actual confidence levels of the tests are neither really known nor knowable. Again, our methods point to a solution to this problem in a couple of ways. In section 2, we establish the asymptotics of the first collision times (as well as the distribution of the intervals between subsequent collision times) but do a careful enough error analysis to provide bounds on confidence levels for tests involving the first collision times. These collision times are seen to be asymptotically Weibull. In section 4, we deal with a situation for which classical asymptotic approaches fall short either of providing satisfactory estimates on confidence levels or of even covering the full range of sample sizes of interest. In this case, we abandon asymptotics for a more straightforward approach based on moment methods that go back to Chebyshev in the 19th century. As a prelude to this, we use the same approach in section 3 to develop tests based on classical occupancy statistics.

Besides the analysis of the collision times in section 2 already mentioned, we review there the basic distributions and the occupancy statistics following von Mises' approach. The basic model is the Maxwell–Boltzmann (*MB*) model for distributing n balls to N cells with each ball having an equal probability of going into any of the cells. von Mises' approach is particularly well suited to the development of tests because of its precise expressions for the factorial moments of multiple occupancy statistics. For von Mises, the interesting cases involved asymptotics as the ratio of the number of balls to the number of cells goes to infinity. We extend his approach slightly to include another asymptotic regime in which this ratio goes to zero. This is more appropriate for the testing of RNGs when the word size is taken to be relatively large and is especially more appropriate for the testing of PRNGs since it aids in the development of tests with sample sizes that are only a fraction of their periods. Section 2 also includes a discussion of the RNGs considered in this paper as well as a discussion of the structure of hypothesis tests that are appropriate for randomizers.

In section 3, we discuss the issue of testing with confidence. The issue is whether one can calculate tail probabilities with enough accuracy to guarantee a particular significance level for a given test. We demonstrate that this can be done but it may require the adoption of nonasymptotic methods. An apparent "defect" of linear congruential generators is that they cannot repeat themselves within a period. This leads immediately to a "birthday" test that they cannot pass on both theoretical and practical grounds. One possible "fix" for this situation is to use a PRNG whose underlying dynamical system operates on a larger than required word space and then project onto some smaller factor. Detecting the difference between such a generator with hidden states and a true source of uniformly distributed binary integers is one of the things that the tests in section 4 are about.

In section 4, we introduce generalized Fermi–Dirac (*GFD*) models that serve as

idealized models for certain PRNGs and develop tests on them. In reality, the lessons illustrated by the methods of this development are more important than any particular test results. First, as alluded to above, to distinguish between a *GFD* model and the appropriate, corresponding *MB* model, a statistic is singled out by the Neyman–Pearson lemma. Essentially, only statistics that significantly correlate with this one have any chance of distinguishing these two distributions. Even more, the choice of sample size is seen to be critical—a sample can be too small, and too small can still be large. Second, the tests we develop in this section are nonasymptotic tests and provide excellent estimates for the confidence levels used. They even provide some indication of the power of the tests. Since we originally tried to develop asymptotic tests, we are in a position to compare the nonasymptotic approach with more standard approaches. We believe these "lessons" point the way to possibly fruitful directions for future investigations.

In section 5, all the tests we have applied to our generators are described. Finally, in section 6, the results of the testing are indicated.

2. Classical and generalized classical distributions. The basic model is associated with the classical problem of randomly distributing n particles or balls into N cells or boxes [9]. Every one of the n balls may occupy any of the N cells with probability $\frac{1}{N}$, and the occupations for different balls are completely independent. Therefore, there are N^n equally probable arrangements of the balls over the cells and the probability of each of these arrangements is $\frac{1}{N^n}$. We are interested in the statistics of the occupation numbers for the cells. These are defined by

$$x_i = \#\{\text{balls in the } i\text{th box}\}, \quad 1 \le i \le N.$$

Since the number of different ways to assign the n balls so that x_i go to cell $i, 1 \le i \le N$, is given by $\frac{n!}{x_1!x_2!\cdots x_N!}$, it follows that the probability of each event $x_1, x_2, \ldots, x_N$, $\sum_{i=1}^N x_i = n$, is given by $P_{MB}(x_1, x_2, \ldots, x_N) = \frac{n!}{x_1!x_2!\cdots x_N!}\frac{1}{N^n}$.

We refer to this model for given n, N and with the probability P_{MB} above as the *MB* model with parameters n, N $(MB(n, N))$.

If we have a source of uniformly distributed binary words of length k, each word $w \in W_k, w = (i_1, i_2, \ldots, i_k)$, $i_s \in \{0, 1\}, 1 \le s \le k$, corresponds to a box or cell among $N = 2^k$ boxes. The experiment of drawing an i.i.d. sample, $Z_1, Z_2, \ldots, Z_n$, is equivalent to an $MB(n, N)$ experiment. If we have a symmetric Bernoulli bit stream, $u_1, u_2, \ldots, u_i, \ldots$, then we can generate a uniform source of any size word, k, by taking $w_1 = (u_1, u_2, \ldots, u_k), w_2 = (u_{k+1}, \ldots, u_{2k}), \ldots$.

The Fermi–Dirac (*FD*) statistic is another classical model associated with distributing particles into cells. However, this time there is an exclusion principle that will not allow an already occupied cell to take an additional ball. In this case, the occupancy numbers satisfy $x_i \in \{0, 1\}, 1 \le i \le N$, and the probability of each event $x_1, x_2, \ldots, x_N$, $\sum_{i=1}^N x_i = n \le N$, is given by $P_{FD}(x_1, x_2, \ldots, x_N) = \frac{1}{\binom{N}{n}}$. We refer to this model for given n, N as the *FD* model with parameters n, N $(FD(n, N))$.

Finally, among the classical distributions, we mention the Bose–Einstein statistic, which is associated with the distribution of Bosons (indistinguishable particles) into cells. In this case, an already occupied cell is more likely to receive another particle than any particular unoccupied cell is. The n balls must now be distributed to the N cells in such a manner that each event $x_1, x_2, \ldots, x_N$, $\sum_{i=1}^N x_i = n$, is equally likely; here the x_i's are again the occupancy numbers. To assign probabilities, we need to count the number of ways to write n as a sum of N nonnegative integers, which is

equal to $\binom{N+n-1}{n}$. Thus,

$$P_{BE}(x_1, x_2, \ldots, x_N) = \frac{1}{\binom{N+n-1}{n}}. \tag{1}$$

Simulating Bose–Einstein statistics with a uniform source (as a ball in cell model) is a little more challenging, but it can be done as a variant on a Polya model as follows: Start with N boxes. The balls are placed in boxes one at a time. If the first ball goes into box i, place a new box over position i in a tower. The second ball can go into any of these now $N+1$ boxes with equal probability. So,

$$P(\text{second ball goes into box in tower } j \neq i) = \frac{1}{N+1}, \text{ while}$$
$$P(\text{second ball goes into box in tower } i) = \frac{2}{N+1}.$$

Continuing in this fashion, we see that each particular outcome (order counts) with occupation numbers $x_1, x_2, \ldots, x_N$, $\sum_{i=1}^{N} x_i = n$, is assigned probability

$$P_{order-counts}(x_1, x_2, \ldots, x_N) = \frac{x_1! x_2! \cdots x_N!}{N(N+1)\cdots(N+n-1)}.$$

Since there are $\frac{n!}{x_1! x_2! \cdots x_N!}$ different ways to distribute the n balls to get the same occupation numbers, it follows that the probability of a given event $x_1, x_2, \ldots, x_N$, $\sum_{i=1}^{N} x_i = n$, is

$$P(x_1, x_2, \ldots, x_N) = \frac{n!}{x_1! x_2! \cdots x_N!} \cdot \frac{x_1! x_2! \cdots x_N!}{N(N+1)\cdots(N+n-1)} = \frac{1}{\binom{N+n-1}{n}},$$

which is, of course, the same as (1).

For *GFD* models, we start with N_0 towers of boxes $T_i, 1 \leq i \leq N_0$, each with $N_1(i)$ boxes. Balls are placed in the $T = \sum_{i=1}^{N_0} N_1(i)$ boxes with an exclusion principle in force, i.e., if the first ball goes into a box in the ith tower, then no other ball may enter that box again, leaving the number of boxes in the ith tower that the next ball can go into diminished by one. We are interested in distributing $n \leq T$ balls and the statistics of the occupation numbers in the N_0 towers. Clearly the probability of a given event $x_1, x_2, \ldots, x_N$, $\sum_{i=1}^{N} x_i = n$ is given by

$$P_{GFD}(x_1, x_2, \ldots, x_N) = \frac{\binom{N_1(1)}{x_1}\binom{N_1(2)}{x_2}\cdots\binom{N_1(N_0)}{x_{N_0}}}{\binom{T}{n}}.$$

For a given n, N_0, N_1 (here $N_1 = N_1(i)$ is a function of the indices $1 \leq i \leq N_0$), we will refer to this model as the *GFD* model with parameters n, N_0, N_1 ($GFD(n, N_0, N_1)$).

In each of the models above, there is an underlying experiment for which the order counts. We will not have much use for these in this paper, but for the sake of the discussion we will give them some names at this time. In the *MB* case, we will denote the basic experiment of placing n balls in N cells where order counts as $MBO(n, N)$. Clearly $P_{MBO}(\text{each outcome}) = \frac{1}{N^n}$. For *FD*, the basic ordered experiment will be denoted by $FDO(n, N)$ with $P_{FDO}(\text{each outcome}) = \frac{1}{C(N,n)}$. We have already discussed the ordered version of the Bose–Einstein experiment; we denote it by $BEO(n, N)$. Finally, the ordered version of the *GFD* experiment is denoted by $GFDO(n, N_0, N_1)$ with $P_{GFDO}(x_1, \ldots, x_N) = \frac{C(N_1(1),x_1)\cdots C(N_1(N_0),x_{N_0})}{C(T,n)}$.

2.1. Random number generators. Every random number generator produces a sequence of numbers (or binary digits) $\{\xi_n\}$ which pretends to be random. Depending on how random are "random" number generators, they can be divided into three big categories:

(i) *True random number generators (TRNGs).* These RNGs are based on physical random processes such as radioactive decay, different kinds of thermal noises, phase fluctuations in oscillating processes, and more. An ideal RNG producing the Bernoulli sequence (see more below) we also count as a TRNG. It is useful as a theoretical tool and for purposes of comparison.

(ii) *Pseudorandom number generators (PRNGs).* These RNGs are based entirely on algorithms and usually implemented as computer codes; see [16], [30], and references therein. PRNGs are deterministic and, of course, are not truly random. They just hope to simulate genuine randomness with some level of success, which depends on the application. Every PRNG involves some function F that operates on a space W_k of binary words of length k. Once a seed $w_0 \in W_k$ has been chosen, the generator produces words $w_j = F(w_{j-1}), 1 \le j$, until the first time T that $w_T = w_0$. The collection of words $\{w_0, w_1, \dots, w_{T-1}\}$ is then a cycle of the mapping F determined by the seed w_0. There may be many different cycles of varying lengths or there may be only one cycle. Within a given cycle $C = C(w_0)$, each word $w \in C$ occurs exactly once. However, many PRNGs follow F by some sort of projection $\Phi : W_k \to W_{k_0}, 1 \le k_0 \le k$. In these cases, the generator can produce the same output word more than once. We will refer to $T = T(w_0)$ as the period of the PRNG on the cycle determined by w_0. We will refer to k_0 as the number of visible states for the PRNG and we will say that a PRNG with $k_0 < k$ has hidden states. We do not consider any PRNGs for which k_0 itself depends on the choice of seed.

(iii) *Randomly seeded pseudorandom number generators.* These RNGs have been introduced to fix the fundamental problems associated with the periodicity and deterministic nature of regular PRNGs. The problem is partially solved by *random seeding* of a PRNG when we choose the initial value w_1 of the sequence $\{w_j\}$ at random and then apply the algorithm of a PRNG. To improve the quality of the RNG one can apply random seeding more often.

There are many different methods for generating pseudorandom numbers (see, for instance, [16], [30], and references therein), and with the advances in testing the better ones survive and improve. As to testing, we will follow Knuth [16, section 3.3.3] and call the tests exploring the fundamental limitations of PRNGs *theoretical tests.* Theorem P in [16, section 3.3.3] is an example of that kind of theoretical test for a class of linear congruential sequences, and in section 3.3.4 of [16] there is an analysis of the deficiencies of linear congruential RNGs. Knuth points out in [16, section 3.3.3] that the development of theoretical tests is quite difficult and the majority of results in this area are obtained for statistical tests made for the entire period of a PRNG. In practice the limitations of PRNGs associated with their periodicity are well known, and normally a PRNG is used to generate at most $T_{\max} \ll T$ numbers, where T is its period. Hence, it is justified to call a *theoretical test efficient if it detects a deficiency of a PRNG based only on the numbers* w_j, $1 \le j \le T_{\max}$.

The generators below are considered in this paper: We list them now with a brief description of their properties.

I. *RAND* is a linear congruential PRNG whose output is a sequence of 32-bit integers X_n such that $0 \leq X_n \leq 2^{32} - 1$, and X_n is defined by the recursive formula

$$X_n = 69069X_{n-1} + 1 \mathrm{mod} 2^{32}.$$

RAND has only one orbit consisting of the set of integers $\{0, 1, \ldots, 2^{32} - 1\}$. It has no hidden states. The best that *RAND* can simulate are *FD* statistics $FDO(n, 2^{32})$ and $FD(n, 2^{32}), 1 \leq n \leq 2^{32}$.

II. *RANDU* is a multiplicative linear congruential PRNG whose output is a sequence of (odd) 31-bit integers X_n such that $0 \leq X_n \leq 2^{31} - 1$ and X_n is defined by the recursive formula

$$X_n = 65539X_{n-1} \mathrm{mod} 2^{31}.$$

Here, X_0 is normally taken as odd. *RANDU* has three orbits: $C_0 = \{0\}$, $C_1 = C(1)$, $C_2 = C(2)$. The lengths of these orbits are 1, 2^{30}, and $2^{30} - 1$, respectively. *RANDU* has no hidden states. The best it can simulate are *FD* statistics $FDO(n, N_0)$ and $FD(n, N_0)$, where $N_0 \in \{2^{30} - 1, 2^{30}\}$ depending on whether the seed used is an even or odd nonzero integer.

III. *GGL* is a multiplicative linear congruential PRNG whose output is a sequence of 31-bit integers X_n such that $0 \leq X_n \leq 2^{31} - 1$ and X_n is defined by the recursive formula

$$X_n = 16807X_{n-1} \mathrm{mod} \left(2^{31} - 1\right).$$

This generator has two orbits: $C_0 = \{0\}$ and $C_1 = \{1, 2, \ldots, 2^{31} - 1\}$. The best it can do is simulate $FDO(n, 2^{31} - 1)$ and $FD(n, 2^{31} - 1), 1 \leq n \leq 2^{31} - 1$.

IV. Lagged Fibonacci generators use an initial set of seed values $x_1, x_2, \ldots, x_p$ and two lags p and q, $1 \leq q < p$, to generate successive elements x_i for $i > p$ by means of the recursion $x_i = x_{i-p} \circ x_{i-q}$. Here $\circ$ is some binary operation which might be $+, -, *$, or (xor) (exclusive or). Following Marsaglia [23], we designate such a generator by $F(p, q, \circ)$. Lagged Fibonacci generators of maximal period are obtained when an associated polynomial is primitive over approriate modular rings; see [6] and [22]. This is true, for example, of $F(17, 5, +)$ on integers $\mathrm{mod} 2^m$. (The period is $(2^{17} - 1)2^{m-1}$—every nonzero seed determines an orbit of this length.) Other good choices are $F(31, 13, +), F(55, 24, +)$, and $F(250, 103, +)$. The (xor) versions of these generators have maximal period $(2^p - 1)$ regardless of the word size. Some of these generators are known by other names: $F(250, 103, (xor))$ is known as R250, for example. The initial seeds of these generators can be represented by $m \times p$ binary matrices and thus they operate naturally on $W_{2^{mp}}$. Their visible output is obtained by projection onto the pth column vector and are thus in W_{2^m}. We have shown, but will not prove it here, that for $F(p, q, (xor))$ of maximal period, if the initial $m \times p$ seed matrix has rank $m, 1 \leq m \leq p$, then each nonzero m-bit binary integer will occur exactly 2^{p-m} times and 0 will occur exactly $2^{p-m} - 1$. Thus, the best that R250 can simulate, for example, are *GFD* models $GFDO(n, 2^m, N_1)$ and $GFD(n, 2^m, N_1)$, where $N_1(1) = 2^{250-m} - 1$ and $N_1(i) = 2^{250-m}, 2 \leq i \leq 2^{31}$. Here we have identified the indices i with the binary outputs $(i - 1)_2$. We will discuss what we can say in the " $\circ$ " = " $+$ " case below.

V. *random()* is a PRNG which is built into the UNIX operating system. Its output is a sequence of 31-bit integers X_n, $0 \le X_n \le 2^{31} - 1$. The algorithm on which *random()* is based is not disclosed; it is only said in the on-line manual that "*random()* uses a non-linear additive feedback random number generator employing a default table of size 31 long integers to return successive pseudorandom numbers in the range from 0 to $2^{31} - 1$." It is underscored that, unlike many RNGs whose several least significant digits are of poor quality, "all the bits generated by *random()* are usable." For example, the least significant bit of *random()* "will produce a random binary value." We expect that *random()* has a large number of hidden states but that all visible states appear with the same or nearly the same frequency. If this is true, then *random()* can at best simulate *GFD* statistics $GFDO(n, 2^{31}, N_1)$ and $GFD(n, 2^{31}, N_1)$ with constant or nearly constant tower heights.

VI. *TGGL* is obtained by truncating the last 10 digits of *GGL* producing a PRNG with 21 visible states and 10 hidden states. The best it can simulate are $GFDO(n, 2^{21}, N_1)$ and $GFD(n, 2^{21}, N_1)$, where $N_1(1) = 2^{10} - 1, N_1(i) = 2^{10}$, $2 \le i \le 2^{21}$. Here again we have identified the indices i with the binary outputs $(i-1)_2$.

VII. *TRNG* is output from a prototype true random number generator based on alpha decay.

A summary of the generators is given in Table 1. The pseudorandom generators were implemented on the Silicon Graphics Origin 200 system (SG 200). For a comparative study on PRNGs, see also [32].

TABLE 1
Summary of basic information on the tested random number generators.

	Name	Type	Output
1	*TRNG*	Physical (α-particles)	Binary digits
2	R250	Pseudorandom	1- to 31-bit words
3	*RAND*	Pseudorandom	Binary words of length 32
4	*RANDU*	Pseudorandom	Binary words of length 31
5	*GGL*	Pseudorandom	Binary words of length 31
6	*random()*	Pseudorandom	Binary words of length 31
7	*TGGL*	Pseudorandom	Binary words of length 21
8	$F(17, 5, (xor))$	Pseudorandom	1- to 17-bit words
9	$F(17, 5, +)$	Pseudorandom	1- to 17-bit words
10	$F(31, 13, (xor))$	Pseudorandom	1- to 31-bit words

TRNGs are normally bit generators and tests for them are derived from the assumption that their output is typical of a symmetric Bernoulli process. It is trivial to derive all of the *MB* models from such output. All of the tests that apply to this category of generator fall into the category of testing symmetric Bernoulli versus not symmetric Bernoulli (symmetric Bernoulli)^C. In many cases, the tests are most directly in the form of $MB(n, N)$ versus $(MB(n, N))^C$ or $MBO(n, N)$ versus $(MBO(n, N))^C$ for some suitable choice of n, N. All such tests can be applied to all generators.

In the case of PRNGs we have seen that they can, at best, simulate the ordered and/or unordered versions of appropriate *FD* or *GFD* models. This brings up two testing questions:

1. How well does the given PRNG simulate the appropriate ordered and/or unordered *FD* or *GFD* experiment(s)? This leads to testing *FD* versus $(FD)^C$, FDO versus $(FDO)^C$, *GFD* versus $(GFD)^C$, or $GFDO$ versus $(GFDO)^C$.

2. How well does the *FD* or *GFD* experiment simulate the corresponding *MB* models? This leads to tests of *FD* versus *MB*, *GFD* versus *MB*, as well as ordered versions of these.

The testing of all versions of *GFD* versus $(GFD)^C$ will be the subject of another paper. However, before leaving this subject, we would like to indicate some of its significance. The simplest cases involve the classical *FD* distributions, ordered and unordered. Let N be fixed and take $n = N$. Each run of $FDO(N, N)$ is a complete sampling that determines a random permutation of the integers $1, 2, \ldots, N$. We can associate with each such sampling a map $F : \{1, 2, \ldots, N\} \to \{1, 2, \ldots, N\}$, where $F(i_1) = i_2, F(i_2) = i_3, \ldots, F(i_N) = i_1$. This determines a PRNG associated with F. Furthermore, every PRNG that has an orbit of size N represents a possible complete sampling of the N integers $\{1, 2, \ldots, N\}$. If a PRNG provides a good simulation of $FDO(N, N)$, then it will provide good simulations of all $FDO(n, N)$ (and $FD(n, N)$) for $1 \leq n \leq N$. For each run $i_1, i_2, \ldots, i_n$ associate the sequence of indices $j(1), j(2), \ldots, j(n)$, where $i_{j(1)} < i_{j(2)} < \cdots < i_{j(n)}$. There are two independent statistics that can be associated with this run: (1) the random permutation of the integers $1, 2, \ldots, n$ given by $j(1), j(2), \ldots, j(n)$, and (2) the gaps $\Delta_0 = i_{j(1)}, \Delta_1 = i_{j(2)} - i_{j(1)}, \ldots, \Delta_n = N - i_{j(n)}$. Both of these statistics are uniformly distributed and together they are sufficient to determine $FDO(n, N)$. This last statement assumes that, for each n, repeated sampling is possible, which, for a PRNG is only true in a limited sense. It is significant to note, however, that Marsaglia considers his gap collision test [23], which is based on collisions in the sequence $\Delta_0, \Delta_1, \ldots, \Delta_n$, to be very powerful in detecting defective PRNGs.

2.2. Basic results on occupation statistics. For the distributions $MB(n, N)$, we are interested in the following set of random quantities:

$$\begin{aligned} \gamma_s &= \text{number of cells containing exactly } s \text{ balls} \\ &= \#\{i : x_i = s\}, \ s = 0, 1, \ldots, \end{aligned} \tag{2}$$

and their mean values,

$$A_s = \mathbb{E}\{\gamma_s\}, \quad \mathbb{E}\{\cdot\} \text{ is the expected value with respect to the } P_{MB}. \tag{3}$$

The Poisson distribution, with parameter, say, a, appears in the statistics of occupancy numbers very often and is denoted by $\mathcal{P}_a$, i.e.,

$$\mathcal{P}_a(m) = e^{-a}\frac{a^m}{m!}, \quad m = 0, 1, 2, \ldots, a \geq 0. \tag{4}$$

The distribution of a quantity γ taking on the values $0, 1, 2, \ldots$ is denoted by P_γ, i.e.,

$$P_\gamma(m) = P\{\gamma = m\}, \quad m = 0, 1, 2, \ldots. \tag{5}$$

The convergence of distributions is understood as the weak convergence in measure [10, section VIII.1].

A basic result on the statistics of the occupancy numbers γ_s is that if

$$N \to \infty \text{ and } a_s = \frac{\exp\left(-\frac{n}{N}\right) n^s}{s! N^{s-1}} \to \bar{a}, \quad \bar{a} \geq 0, \tag{6}$$

then

$$\text{for } s \geq 2 \ \lim A_s = \bar{a} \text{ and } \lim P_{\gamma_s} = \mathcal{P}_{\bar{a}}. \tag{7}$$

This result is due to von Mises [33, section IV.9], and it also can be expressed as the following approximation formula:

$$P_{\gamma_s} \cong \mathcal{P}_{a_s}, \quad a_s = \frac{\exp\left(-\frac{n}{N}\right) n^s}{s! N^{s-1}}. \tag{8}$$

In developing tests based on the γ_s statistics, we will not actually use these asymptotic results directly. However, the asymptotics do provide insight and guidance in choosing sample sizes. For this reason, we want to be a bit more precise in our discussion of these results. von Mises' asymptotic results are based on his derivation of exact formulas for the factorial moments $\mu_r(s) = E(\gamma_s(\gamma_s - 1)\cdots(\gamma_s - (r-1)))$ of the random variables γ_s. He proves the following [33, section IV.9, formulas (59)–(62), (68)–(70)].

THEOREM 1. *The rth-factorial moment for the random variable γ_s associated with the MB model with parameters n, N is given by the expression*

$$\mu_r(s) = r! \binom{N}{r} \frac{n!}{(s!)^r (n - sr)!} \frac{(N-r)^{n-rs}}{N^n}. \tag{9}$$

Based on (9) we see that the rth-factorial moments of γ_s on $MB(n, N)$ satisfy the inequality

$$\begin{aligned} &\left[\frac{N}{s!}\left(\frac{n}{N}\right)^s e^{-\frac{n}{N}}\right]^r e^{-\frac{nr}{N(N-1)} - \frac{r(r-1)}{N-(r-1)} - \frac{sr(sr-1)}{n-(sr-1)} + \frac{r^2 s}{N-1}} \\ &\le \mu_r(s) \le \left[\frac{N}{s!}\left(\frac{n}{N}\right)^s e^{-\frac{n}{N}}\right]^r e^{-\frac{r(r-1)}{N} - \frac{sr(sr-1)}{n} + \frac{r^2 s}{N}}. \end{aligned} \tag{10}$$

The expression $a_s = \frac{N}{s!}\left(\frac{n}{N}\right)^s e^{-\frac{n}{N}}$ that appears on both extreme sides of these inequalities plays the crucial role in determining the asymptotics for γ_s when n and N are properly related. The natural range of the values for a_s is

$$0 \le a_s \le \max_{\alpha \ge 0} N \frac{e^{-\alpha}\alpha^s}{s!} = N \frac{\left(\frac{s}{e}\right)^s}{s!} = \frac{N}{\sqrt{2\pi s}\exp\left\{\frac{1}{12s + \theta_s}\right\}}, \quad 0 \le \theta_s \le 1. \tag{11}$$

Here, $\alpha = \frac{n}{N}$ and Stirling's formula was used in the last equality. Following von Mises, we impose the condition that

$$n \to \infty,\ N \to \infty, \quad a_s = N \frac{e^{-\alpha}\alpha^s}{s!} = \frac{\exp\left(-\frac{n}{N}\right) n^s}{s! N^{s-1}} \to \bar{a}, \quad \bar{a} \ge 0, \tag{12}$$

where the natural number $s \ge 2$ and $\bar{a}$ can be chosen arbitrarily. It turns out that there are two ways to keep a_s constant. To see this let us look at the related equation for $\alpha \ge 0$:

$$a_s = N \frac{e^{-\alpha}\alpha^s}{s!} = \bar{a}, \quad \bar{a} \ge 0, \tag{13}$$

where $s \ge 2$ is a fixed integer, $\bar{a}$ is a fixed nonnegative number, and $N \to \infty$. As follows from (11), for (13) to have a solution $\bar{a}$ must be in the following interval:

$$0 \le \bar{a} \le N \frac{\left(\frac{s}{e}\right)^s}{s!} = \frac{N}{\sqrt{2\pi s}\exp\left\{\frac{1}{12s + \theta_s}\right\}}, \quad 0 \le \theta_s \le 1. \tag{14}$$

An elementary analysis of (13) shows that if (14) is satisfied, then the equation has two positive solutions $\alpha_\pm = \alpha_\pm(s, \bar{a}, N)$ such that

$$0 < \alpha_-(s, \bar{a}, N) \le s \le \alpha_+(s, \bar{a}, N) < \infty, \tag{15}$$

$$\lim_{N\to\infty} \alpha_-(s, \bar{a}, N) = 0, \quad \lim_{N\to\infty} \alpha_+(s, \bar{a}, N) = \infty, \tag{16}$$

or, more precisely,

$$\alpha_-(s, \bar{a}, N) = (\bar{a}s!)^{1/s} N^{-1/s} (1 + o(1)) \text{ as } N \to \infty, \tag{17}$$

$$\alpha_+(s, \bar{a}, N) = \ln N + s \ln(\ln N)(1 + o(1)) \text{ as } N \to \infty. \tag{18}$$

For every fixed N there are, respectively, two values of n providing the same $\bar{a}$:

$$n = n_\mp(s, \bar{a}, N) = N\alpha_\mp(s, \bar{a}, N). \tag{19}$$

Evidently the relations (17), (18), and (19) describe two different scenarios of behavior for $\alpha = \alpha_\mp(s, \bar{a}, N)$ and $n_\mp(s, \bar{a}, N)$ as $N \to \infty$:

$$\text{low density, } \alpha \le s,\ n \ll N : n = n_- = (\bar{a}s!)^{1/s} N^{1-1/s} (1 + o(1)), \tag{20}$$

$$\text{high density, } \alpha \ge s,\ n \gg N : n = n_+ = N[\ln N + s \ln(\ln N)(1 + o(1))]. \tag{21}$$

The high density case (21) $\alpha = \alpha_+(s, \bar{a}, N) \ge s$ has been studied by von Mises [33, section IV.9]. In this case clearly the average number of balls per cell $\alpha = n/N$ approaches infinity, and, hence, the overwhelming number of cells is occupied by more than s balls, and, in addition to that, $A_s > A_{s-1} > \cdots > A_0$. More precisely, the statistics of the occupancy numbers are described by the following statement which is due to von Mises [33, section IV.9] (see also [9, section IV.2, formula (2.12)]).

PROPOSITION 2. *Under the von Mises condition* (12) *and* $\alpha = n/N \to \infty$ *(which, in fact, is the high density case* (21)*), the distribution of the random variable* γ_s *approaches the Poisson distribution, i.e.,*

$$P_{\gamma_s} \to \mathcal{P}_{\bar{a}}, \quad \lim P(\gamma_s = m) = e^{-\bar{a}} \frac{\bar{a}^m}{m!}, \quad m = 0, 1, 2, \ldots. \tag{22}$$

Moreover,

$$\lim A_s = \lim \mathbb{E}\{\gamma_s\} = \bar{a}. \tag{23}$$

In addition to that,

$$\lim P(\gamma_{s'} = 0) = 1, \quad 0 \le s' \le s - 1, \tag{24}$$

$$\lim P(\gamma_{s'} \ge K) = 1, \quad s' \ge s + 1, \text{ for all } K > 0. \tag{25}$$

In the low density case (20) the average number of balls per cell $\alpha = n/N$ approaches zero, and the overwhelming number of cells is empty, and, in addition to that, $A_s < A_{s-1} < \cdots < A_0$. More precisely, an examination shows that the method of von Mises [33, section IV.9] gives without any alterations the following result for the low density case (actually, even with less effort than in the high density case). In the case $\alpha \to 0$ the von Mises condition (12) turns into the following:

$$n \to \infty,\ N \to \infty, \quad a_s = N\frac{\alpha^s}{s!} = \frac{n^s}{s!N^{s-1}} \to \bar{a}, \quad \bar{a} \ge 0. \tag{26}$$

THEOREM 3. *Under asymptotic condition* (26) *and* $\alpha = n/N \to 0$ *(which, in fact, is the low density case* (20)*) the distribution of the random variable* γ_s *approaches the Poisson distribution, i.e.,*

$$P_{\gamma_s} \to \mathcal{P}_{\bar{a}}, \quad \lim P(\gamma_s = m) = e^{-\bar{a}} \frac{\bar{a}^m}{m!}, \quad m = 0, 1, 2, \ldots. \tag{27}$$

In particular,

$$\lim A_s = \lim \mathbb{E}\{\gamma_s\} = \bar{a}. \tag{28}$$

In addition to that,

$$\lim P(\gamma_{s'} \geq K) = 1, \quad 0 \leq s' \leq s-1, \text{ for all } K > 0, \tag{29}$$

$$\lim P(\gamma_{s'} = 0) = 1, \quad s' \geq s+1. \tag{30}$$

The relations (29) and (30) indicate that under the asymptotic condition (26) the statistics of $\gamma_{s'}$ for $s' \neq s$ collapse to trivial distributions. This suggests that after appropriate renormalization $\gamma^*_{s'}$ of the quantities $\gamma_{s'}$ their distributions may converge to one of the standard limit distributions. Indeed, it is shown in [13] (see also [33, section IV.9]) that under the von Mises condition (12) the following relations hold for $\gamma_s = \gamma_s(n, N)$, $s \geq 2$, and $N \to \infty$:

$$\lim \mathbb{E}\left\{\frac{\gamma_s}{N}\right\} = a_s, \text{ where } a_s = e^{-\bar{a}} \frac{\bar{a}^s}{s!}, \tag{31}$$

$$\lim \mathbb{E}\left\{\frac{[\gamma_s - \mathbb{E}\{\gamma_s\}]^2}{N}\right\} = a_s - a_s^2\left[1 + \frac{(s-\bar{a})^2}{\bar{a}}\right] = v_s, \tag{32}$$

and the random variables γ_s are asymptotically normal with mean value $\bar{a}_s N$ and variance $v_s N$. Thus the appropriate renormalization here will be

$$\gamma^*_s = \frac{\gamma_s - a_s N}{\sqrt{v_s N}} = \frac{\gamma_s - a_s N}{\sqrt{\left(a_s - a_s^2\left[1 + \frac{(s-\bar{a})^2}{\bar{a}}\right]\right) N}}, \text{ where } a_s = e^{-\bar{a}} \frac{\bar{a}^s}{s!}. \tag{33}$$

The normality of random variables γ_s under condition (12) for $N \to \infty$ has also been considered in [3], [14] (see also references therein).

REMARK 4. *Though the cases of the low and the high densities (respectively,* (20) *and* (21)*) are similar in many ways, there remain certain advantages to using the low density case for testing which are discussed in section* 3*. An additional advantage of the low density case is that in that case* $\alpha = \alpha_-(s, \bar{a}, N) \leq s$ *and, hence, fewer "balls" and, consequently, shorter sequences of random numbers are needed.*

2.3. Collision times. Another set of random quantities related to the classical occupancy problem are the collision times (waiting times) [15, section 3.3.2], [8, section II.7]. To introduce them we consider the process of placing n balls in N cells. We place the balls in the cells one after another until a ball is placed in a cell that is already occupied. In that case we say that a collision has occurred and assign to the collision the variable τ_1, where τ_1 is the number of the ball to enter the occupied cell. Having recorded τ_1, the time of the first collision, we go on until the next collision and record the time of this second collision as τ_2, which is clearly greater than τ_1. In this

fashion we generate the collision times $\tau_1, \tau_2, \tau_3, \ldots$. The collision times $\tau_1, \tau_2, \ldots$ have the following representations:

$$\begin{aligned}
\tau_1 &= \min\{t : t > 1 \text{ and there exists } s < t \text{ such that } w_t = w_s\}, \\
\tau_2 &= \min\{t : t > \tau_1 \text{ and there exists } s < t \text{ such that } w_t = w_s\}, \\
&\ldots \\
\tau_k &= \min\{t : t > \tau_{k-1} \text{ and there exists } s < t \text{ such that } w_t = w_s\}.
\end{aligned} \tag{34}$$

Here is a statement of the limiting form of the statistics of collision times.

THEOREM 5. *For fixed $s \geq 0$ and fixed $x_1, \ldots, x_s > 0$,*

$$P\left\{ \begin{array}{c} \frac{\tau_1(\tau_1-1)}{2N} > x_1, \frac{\tau_2(\tau_2-1)}{2N} - \frac{\tau_1(\tau_1-1)}{2N} > x_2, \\ \ldots, \frac{\tau_s(\tau_s-1)}{2N} - \frac{\tau_{s-1}(\tau_{s-1}-1)}{2N} > x_s \end{array} \right\} \to e^{(-x_1-x_2-\cdots-x_s)}.$$

That is, the joint distribution of the increments of the nonlinearly transformed collision times $\tau_1^*, \tau_2^*, \ldots, \tau_s^*$, where, for each $i = 1, 2, \ldots, s$,

$$\tau_i^* = \frac{\tau_i(\tau_i - 1)}{2N},$$

converges weakly to the product of s standard exponential distributions.

Another equivalent interpretation is that the collision times $\tau_1, \tau_2, \ldots, \tau_s$ after the nonlinear scaling transformation

$$\tau_i^* = \frac{\tau_i(\tau_i - 1)}{2N}, \quad i = 1, 2, \ldots,$$

form, asymptotically as $N \to \infty$, the Poissonian point flow with rate $\lambda_0 = 1$. We will prove this result without effective estimation of the remainder. We will, however, give bounds for the $s = 1$ case and some indications for the $s = 2$ case that can be used to estimate errors in these cases.

Proof. We begin with the following remark: if $x_1, \ldots, x_s$ are independent, standard exponential random variables, i.e., they have joint distribution density

$$p_s(x_1, \ldots, x_s) = \prod_{i=1}^{s} e^{-x_i} I_{\{x_i > 0\}}, \tag{35}$$

then the random variables $y_i = \frac{1}{2} + \sqrt{\frac{1}{4} + 2(x_1 + \cdots + x_i)}$ have joint distribution with density

$$q_s(y_1, \ldots, y_s) = \left(y_1 - \frac{1}{2}\right) \ldots \left(y_s - \frac{1}{2}\right) e^{-\frac{y_s(y_s-1)}{2}} I_{\{y_1 > 1, \ldots, y_s > 1\}}. \tag{36}$$

Here

$$x_1 = \frac{y_1(y_1 - 1)}{2}, \ldots, x_s = \frac{y_s(y_s - 1) - y_{s-1}(y_{s-1} - 1)}{2},$$

and substitution of these expressions into (35) together with calculation of the Jacobian $\frac{\partial(x_1, \ldots, x_s)}{\partial(y_1, \ldots, y_s)}$ gives (36). It is sufficient to check now that the joint distribution of the random variables

$$\hat{\tau}_1 = \frac{\tau_1}{\sqrt{N}}, \ldots, \hat{\tau}_s = \frac{\tau_N}{\sqrt{N}}$$

converges weakly in $C(R^d)$ to the limiting distribution with density (36). To simplify the calculations, let's discuss only the particular case of $s = 2$. We have for $1 \leq s_1 < s_2$,

$$\begin{aligned} & P\{\tau_1 = s_1, \tau_2 = s_2\} \\ &= \frac{N(N-1)\dots(N-(s_1-2))(s_1-1)\dots(N-(s_2-3))(s_2-1)}{N^{s_2}} \\ &= \frac{(s_1-1)(s_2-1)}{N^2} e^{-\frac{(s_2-3)(s_2-1)}{2N} + o(\frac{s_2^3}{N^2})}. \end{aligned}$$

If as

$$N \to \infty, \ \frac{s_1}{\sqrt{N}} \to y_1(1+o(1)), \text{ and } \frac{s_2}{\sqrt{N}} \to y_2(1+o(1)),$$

then on the set $1 < y_1 \leq y_2 < A$, for fixed A,

$$\begin{aligned} P\{\tau_1 = s_1, \tau_2 = s_2\} &= \frac{1}{\sqrt{N}} \frac{1}{\sqrt{N}} y_1 y_2 e^{-\frac{y_2(y_2-1)}{2}} \left(1 + o\left(\frac{1}{\sqrt{N}}\right)\right) \\ &= \frac{1}{\sqrt{N}} \frac{1}{\sqrt{N}} q\left(\frac{s_1}{\sqrt{N}}, \frac{s_2}{\sqrt{N}}\right) \left(1 + o\left(\frac{1}{\sqrt{N}}\right)\right). \end{aligned}$$

Thus, we have proved the theorem by, in fact, proving the local form of the theorem. □

Let's evaluate remainders for $s = 1$. We have for this case

$$P\{\tau_1 > n\} = \frac{N(N-1)\dots(N-n+1)}{N^n} = e^{\sum_{i=1}^{n-1} \ln(1-\frac{i}{N})}.$$

Assume that $0 < \epsilon = \frac{n-1}{N} < \frac{1}{2}$ and use inequalities

$$-\frac{\epsilon}{1-\epsilon} < \ln(1-\epsilon) < -\epsilon, \frac{1}{1-\epsilon} < 1 + 2\epsilon$$

to obtain

$$P\{\tau_1 > n\} \leq e^{-\sum_{i=1}^{n-1} \frac{i}{N}} = e^{-\frac{n(n-1)}{2N}} \tag{37}$$

and

$$P\{\tau_1 > n\} \geq e^{-\sum_{i=1}^{n-1} \frac{i}{N\left(1-\frac{n-1}{N}\right)}} \geq e^{-\frac{n(n-1)}{2N}\left(1+\frac{2(n-1)}{N}\right)}. \tag{38}$$

For any positive x let $k = k(x)$ be the unique integer such that

$$\frac{k(k-1)}{2N} \leq x \leq \frac{k(k+1)}{2N}.$$

Direct calculation shows that

$$-\frac{1}{2} + \sqrt{2Nx + \frac{1}{4}} \leq k \leq \frac{1}{2} + \sqrt{2Nx + \frac{1}{4}}.$$

Then, for $\tau_1^* = \frac{\tau_1(\tau_1-1)}{2N}$, we have

$$P\{\tau_1^* > x) \le e^{-\frac{k(k-1)}{2N}} = e^{-\frac{k(k+1)}{2N}+\frac{k}{N}} \le e^{-x+\frac{\sqrt{1+8Nx}+1}{2N}} \tag{39}$$

and

$$P\{\tau_1^* > x) \ge e^{-\frac{k(k+1)}{2N}\left(1+\frac{2k}{N}\right)} \ge e^{-x\left(1+\frac{1+\sqrt{1+8Nx}}{N}\right)}.$$

These estimations in (39) guarantee that

$$P\{\tau_1^* > x\} \sim e^{-x} \quad \text{for } x = O\left(N^{\frac{1}{3}}\right),\ N \to \infty, \tag{40}$$

and also provide bounds on the error of the approximation (40).

These estimates can also be used to calculate and provide bounds for the $s = 2$ case. For example, if we abbreviate $k(x_1) = k_1$ and $k(x_1 + x_2) = k_2$ and use the definition of $k\,(x)$, then

$$\begin{aligned} P\{\tau_1^* = x_1, \tau_2^* = x_1 + x_2\} &= P\{\tau_1 = k_1, \tau_2 = k_2\} = A_1 A_2 \\ &= \frac{k_1 - 1}{N}\frac{k_1 + k_2 - 2}{N} P\{\tau_1 > k_1 + k_2\}, \text{ where} \\ A_1 &= \frac{N(N-1)\dots(N-k_1+2)}{N^{k_1-1}}\frac{k_1-1}{N} \\ A_2 &= \frac{(N-k_1+1)\dots(N-(k_1+k_2)+2)}{N^{k_2-1}}\frac{k_1+k_2-2}{N}. \end{aligned}$$

The most important conclusion of this section is that

$$\tau_1 = O\left(\sqrt{N}\right).$$

For example, if $k = 32$, $N = 2^{32} \cong 4,295 \cdot 10^9$, then $\sqrt{N} = 2^{16} \cong 65 \cdot 10^3$ and we should expect the first collision to occur in a sample of this size.

3. Sampling with confidence. In previous sections we discussed the limit statistics for random variables γ_s as $N \to \infty$ and for collision times, again as $N \to \infty$. The limit statistics provide important insight and can be useful for practical purposes. However, most often, we have to deal with a given number N and it is always questionable if one can use the limit statistics in the place of a given one.

The testing problem that must be addressed is as follows: We fix $N = 2^k$ and consider a finite sequence of binary words $\{w_j\}_1^n$, where n is the number of words in the sequence. We would like to test whether $\{w_j\}_1^n$ is truly based on a uniform distribution on the sampling space W_k of binary words of length k as it would be if the words were generated by a truly random Bernoulli sequence. It is customary in sampling to choose a so-called significance level α, a number between 0 and 1, and use it to determine a critical region that will depend on the sample size n that has probability smaller than α of occurring by chance if the distribution is actually the assumed one. This assumes that tail probabilities for the actual distribution can be bounded from above in a meaningful way. However, as mentioned in the introduction, it can be challenging to calculate rigorous but useful bounds on tail probabilities. There are three approaches to doing so:

1. Provide direct estimates on tail probabilities as we were able to do in the case of the first collision times in the previous section.

2. Provide estimates on the error involved in using the limiting distribution instead of the actual distribution—Berry–Eseen estimates in the case of the normal distribution might be used for this sort of approach.
3. Use nonasymptotic moment techniques to estimate the distribution function of the actual distribution and to provide rigorous bounds on tail probabilities.

We illustrate the first approach below with a first collision time test utilizing the tail probability estimates (37) and (38) of the previous section. We illustrate the third approach both below in this section with tests based on the γ_s statistics and again in section 4, in a more complex situation, with tests designed to differentiate between $GFD(n, N_0, N_1)$ and $MB(n, N_0)$. The second approach is difficult to achieve. We will see in section 4 that Berry–Eseen estimates can be used to justify some tests that cover a part of our range of interest, but only a small part.

3.1. First collision test. The following table (Table 2) was computed using (37) and (38).

TABLE 2
Tail cutoffs for a range of significance levels for first collision times.

N	α	n such that $P(\tau_1 \geq n) \leq \alpha$	n such that $P(\tau_1 \leq n) \leq \alpha$
2^{31}	.05	113431	14844
2^{31}	.01	140638	6571
2^{31}	.005	150852	4641
2^{31}	.001	172247	2074
2^{31}	.0005	180688	1467

Based on this table, one can construct a number of hypothesis tests of the type MB versus $(MB)^C$. For example, at the .001 level of significance we have the following two-tailed test.

First collision test for MB versus $(MB)^C$. For $N = 2^{31}$, take a sample of size $n = 2^{18}$. Determine the time τ_1 of the first collision in this sample. Reject $\mathrm{H}_0 : MB$ and accept $\mathrm{H}_a : (MB)^C$ if either $\tau_1 \geq 180688$ or $\tau_1 \leq 1467$.

This two-tailed test is significant at the .001 level. The reader should note that $2^{18} > 180688$. Actually, this test is a theoretical test for PRNGs that have no hidden states. It does not have to be run for such generators since they are bound to never have any collisions within their period. In particular, *RANDU*, *RAND*, and *GGL* automatically flunk this test. Furthermore, the sample size needed to flunk them at the .001 significance level (180688 for a two-tailed test, 172247 for a one-tailed test) is the most efficient of the tests that we know of based on the inability of these generators to repeat over their period.

The first collision test can also detect problems for PRNGs with hidden states. For example, we form a PRNG with hidden states using the PRNG *GGL* by truncating bits 0–21 so only the most significant bits 22–30 (9 out of 31) are visible. The seed for *GGL* that we have chosen is 186739657. We use the resulting generator to generate random words of length 36 by combining four consecutive output words of length 9 bits each. We encountered no collisions among the first 1591139 36-bit words thus generated. The probability of such an event if these 36-bit words were truly uniformly distributed is less than 10^{-8}. Note that we have used only 4(1591139) (truncated) output words of *GGL*, which is less than 0.003 of *GGL*'s period of 2147483647.

3.2. Tests based on γ_s-statistics. It would be very difficult to estimate the errors in approximating the γ_s-distributions by their limiting Poisson distributions, and,

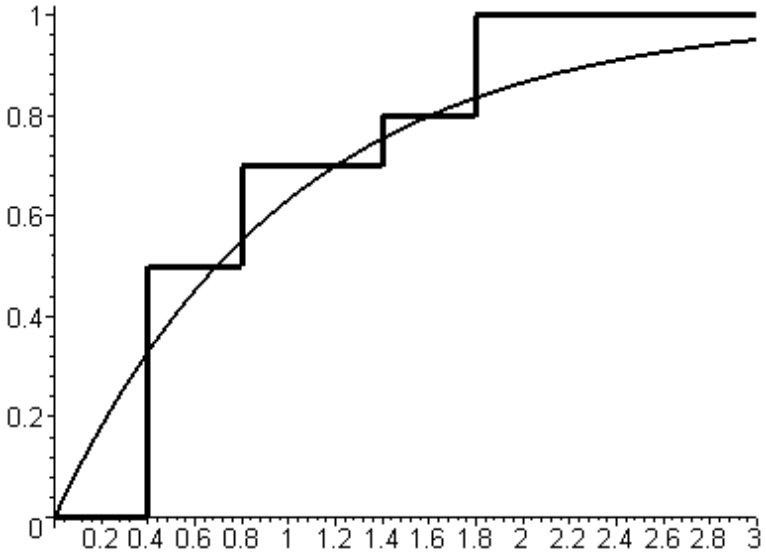

FIG. 1. *The smooth and the polygonal curves in this figure illustrate the way in which a distribution function (the smooth curve) must intersect its Chebyshev polygonal approximation: each of the segments of the polygonal plot is intersected exactly once by the continuous curve.*

therefore, we would not be able to rigorously bound the tail probabilities if we used the Poisson distributions to determine critical regions. Furthermore, for relatively small N and n the actual distributions are not even close to the limiting distributions. We avoid these difficulties by using MAPLE to calculate accurate approximations for the higher order moments of the γ_s random variables from the precise formulas for their factorial moments due to von Mises.

These highly accurate values now permit the use of classical moment theorems due to Chebyshev. In 1874, Chebyshev showed the following (see [33, section VIII, B]): Let V_1 be a distribution function that has moments of all orders, whose first $2m$ moments $M_0, M_1, \ldots, M_{2m-1}$ are known, and assume that V_1 increases at more than $m-1$ points. Then there exists a unique m-step function V whose first $2m$ moments agree with those of V_1. Furthermore, unless $V = V_1$, V and V_1 must intersect exactly $2m-1$ times, once on each of the m vertical segments of V and once on each of the $m-1$ interior flat (horizontal) segments of V (see Figure 1 for a typical situation).

It follows that if X_1 is a random variable with distribution V_1, and V has jumps $A_1, A_2, \ldots, A_m$ at the points $a_1, a_2, \ldots, a_m$, then $P(X_1 \le a_1) \le A_1$ and $P(X_1 \ge a_m) \le A_m$. Given the moments $M_0, M_1, \ldots, M_{2m-1}$, the following steps are required to determine the m-step function that has these same moments (see [33, section VIII, 4.3]):

1. Solve the system

$$\begin{aligned} M_0c_o + M_1c_1 + \cdots + M_{m-1}c_{m-1} &= -M_m, \\ \cdots &= \cdots \\ M_{m-1}c_0 + M_mc_1 + \cdots + M_{2m-2}c_{m-1} &= -M_{2m-1}. \end{aligned} \tag{41}$$

2. Find the roots $a_1, a_2, \ldots, a_m$ of the polynomial

$$x^m + c_{m-1}x^{m-1} + \cdots + c_1x + c_0. \tag{42}$$

There will be m distinct positive real roots of this polynomial.

3. Solve the system

$$\begin{aligned} A_1 + A_2 + \cdots + A_m = M_0 &= 1, \\ a_1 A_1 + a_2 A_2 + \cdots + a_m A_m &= M_1, \\ a_1^2 A_1 + a_2^2 A_2 + \cdots + a_m^2 A_m &= M_2, \\ \cdots &= \cdots \\ a_1^{m-1} A_1 + a_2^{m-1} A_2 + \cdots + a_m^{m-1} A_m &= M_{m-1} \end{aligned} \tag{43}$$

for jump masses $A_1, A_2, \ldots, A_m$ (see the plot in Figure 3 of the next section for an illustration). To illustrate, tests involving γ_2 and γ_3 can be based on Table 3 below.

TABLE 3

This table can be used to determine a number of one- and two-tailed tests of MB versus $(MB)^C$. For example, using the first row, for a generator of 17-bit words, take a sample of size $n = 2^{12}$ and reject MB if $\gamma_2 \le 34$ or $\gamma_2 \ge 99$. This two-tailed test is valid at the .0005 significance level.

s	N	n	m	a_1	A_1	a_m	A_m
2	2^{17}	2^{12}	8	34.9	.0004	98.3	.00004
2	2^{55}	2^{30}	8	35.9	.0004	102.5	.00003
2	2^{31}	2^{18}	7	5.09	.0046	35.5	.00008
2	2^{31}	2^{19}	7	38.2	.0016	98.4	.0002
2	2^{31}	2^{18}	8	4.3	.0018	38	.00001
2	2^{31}	2^{19}	8	35.9	.0009	102.4	.00003
2	2^{32}	2^{19}	8	13.6	.0008	60.8	.00002
3	2^{31}	2^{23}	8	7.1	.0012	45.7	.00001
3	2^{31}	2^{24}	8	120.5	.0003	228.5	.00005

Once again, for *RAND*, *RANDU*, *GGL*, and the (xor) lagged Fibonacci generators (seeded with a full rank initial matrix), all of these tests, in particular those utilizing the γ_2-statistic, are theoretical tests. In particular, the sixth and seventh rows of the table show that a sample of size 2^{19} will be sufficient to fail *RANDU*, *GGL*, *RAND*, and $F(31, 13, (xor))$, provided it is seeded with a full rank initial matrix at the .001 significance level. The first row of the table shows that a sample of size 2^{12} will catch $F(17, 5, (xor))$ (provided it is seeded with a full rank matrix). Finally, the second row of the table shows that a sample of size 2^{30} will be sufficient to catch $F(55, 24, (xor))$ provided it is seeded with a full rank initial matrix. We estimate that a theoretical test for $F(250, 103, (xor))$ utilizing the γ_2-statistics would require a sample of size on the order of 2^{128}, which is not computationally feasible at this time.

Tests based on γ_s statistics can also be used on generators with hidden states, but it is not clear at this point what values of n and N should be looked at. The next section provides guidelines that help with this.

4. *GFD* versus *MB* and *MB* versus $(MB)^C$. We first develop tests for *GFD* models versus *MB* models. These are used to derive an empirical rule that is applied in developing *MB* versus $(MB)^C$ tests for PRNGs.

4.1. *GFD* model versus *MB* model. Let's consider as in section 2 a PRNG model based on a dynamical system $F : W_k \to W_k$ followed by a truncation operator $\Phi : W_k \to W_{k_0}$. Recall that k_0 is the number of visible digits. Let seed w_0 be chosen and let $T \le 2^k$ be the period of the PRNG on the cycle determined by w_0. This means that on the interval $[0, T-1]$ the system $X_{n+1} = F(X_n)$ has no self-intersection (NSI), i.e., the sequence $X(0), \ldots, X(T-1)$ is a permutation of the set

of all states $\{X(0), \ldots, X(T-1)\} \subset W_k$ when this set is endowed with the order inherited from W_k. This holds for each cycle and the cycles partition the phase space W_k. This is equivalent to the condition that the invariant measure of $X_n(\cdot)$ is the uniform distribution on W_k.

Put $N_0 = 2^{k_0} = |W_{k_0}|$, the number of visible states. For each $1 \leq i \leq N_0$, let $N_1(i)$ equal the number of times that the binary word $w = (i-1)(2)$ appears in the cycle $C(w_0)$; here $(i-1)(2)$ is the binary representation for the integer $i-1$, and note that $\sum_{i=1}^{N_0} N_1(i) = T(\omega_0)$. The "maximally random" experiments, incorporating the NSI property, the main feature of the dynamical systems, that we can associate with such a generator are the *GFD* experiments $GFDO(\circ, N_0, N_1)$ and $GFD(\circ, N_0, N_1)$ briefly described in section 2. Note that the functions $N_1 = N_1(i)$ in these models will usually depend on the seed ω_0.

An alternative description of a $GFDO(\circ, N_0, N_1)$ or a $GFD(\circ, N_0, N_1)$ experiment is as follows. A container holds T balls which are labeled uniquely by ordered pairs of integers (i, j), where $1 \leq i \leq N_0$, $1 \leq j \leq N_1(i)$, $T = \sum_{i=1}^{N_0} N_1(i)$. For $1 \leq n \leq T$, n of these balls are drawn randomly, without replacement, and assigned one at a time to N_0 cells, which have been labeled from 1 to N_0, according to the first coordinate of the ball's label, i.e., ball (i, j) is assigned to cell i. The *MB* experiment can be viewed as a *GFD* experiment with $N_1(i) = \infty$ for all i, or equivalently as one run for constant $N_1(i) = N_1 \geq 1$ where drawing is with replacement. Let $F : W_k \to W_k$ and $\Phi : W_k \to W_{k_0}$ be as above. There is a straightforward connection between the *GFD* scheme and PRNGs based on such dynamical systems. First, each F determines a partition of W_k into orbits $C(\omega_i), i = 1, \ldots, m$. On each orbit, the element $\omega \in C(\omega_i)$ can be relabeled in terms of pairs of indices (i, j), where $i = \Phi(\omega) \in W_{k_0}$ and $1 \leq j \leq N_1(i)$. Once this has been done, $F|_{C(\omega_i)}$ and each initial seed $\omega_0 \in C(\omega_i)$ determines a complete sampling $(i_1, j_1), (i_2, j_2), \ldots, (i_T, j_T)$ of the elements of the cycle and therefore could represent the possible outcomes of running an exhaustive *GFD* experiment on T elements. On the other hand, if we have a partition of W_k into disjoint sets $S_1, S_2, \ldots, S_m$ with $|S_i| = T_i, 1 \leq i \leq m$; and a projection $\Phi : W_k \to W_{k_0}$, we can label each $\omega \in S_i$ uniquely by indices $(i, \Phi(\omega), j)$. Then for each i an exhaustive sampling of $GFD(T_i, N_0, N_1)$, say $(i, r_1, j_1), (i, r_2, j_2), \ldots, (i, r_{T_i}, j_{T_i})$, will determine a map $F|_{S_i}$ by $F(i, r_1, j_1) = (i, r_2, j_2), \ldots, F(i, r_{T_i}, j_{T_i}) = (i, r_1, j_1)$. The map $F : W_k \to W_k$ determined by these restrictions combined with Φ will determine a PRNG associated with the given partition and the complete random-sampling-without-replacement of the partition sets. Of course any given PRNG could correspond to "nonrepresentative" results of the associated *GFD* experiments and for that reason a hypothesis test that distinguishes between $GFD(n, N_0, N_1)$ and $MB(n, N_0)$ will not necessarily distinguish between any particular PRNG and the *MB* model. However, all the PRNGs and the *GFD* models share the underlying NSI property and thus analyzing the *GFD* versus *MB* situation should provide guidance on the construction of hypothesis tests that have good potential for detecting differences caused by the NSI defect. This expectation is in fact confirmed in practice. In the remainder of this section, we first develop the most powerful (or nearly so) tests for detecting the difference between $GFD(\cdot, N_0, N_1)$ and $MB(\cdot, N_0)$, based on N_0, N_1 and a choice of significance level $\alpha > 0$. This analysis will produce estimates on the minimal range for n needed to detect the difference between these models for the given significance level. Again, we caution that for any given PRNG, the relevancy of the "test" must be established by actually running it.

Returning to the *GFD* scheme, for n, N_0, N_1, we are interested in the occupation

numbers $x_1, x_2, \ldots, x_{N_0}$ for the positions $i = 1, 2, \ldots, N_0$. Let $(x_1, \ldots, x_{N_0}) = \omega$ be a "typical" realization of the *GFD* experiment; the set of such realizations, as a rule, is smaller than the set of outcomes of *MB* statistics. The last set contains all sequences

$$\{x_i,\ i = 1, 2, \ldots, N_0\}, \quad \sum_{i=1}^{N_0} x_i = n, \quad 0 \le x_i \le n.$$

Let (Ω_{MB}, P_{MB}) and (Ω_{GFD}, P_{GFD}) be the probability spaces for the *MB* experiment and the *GFD* experiment, respectively. Clearly,

$$\Omega_{GFD} \subseteq \Omega_{MB}.$$

We have the following measures on Ω_{MB}:

$$P_{GFD}(\omega) = \begin{cases} \frac{\binom{N_1(1)}{x_1} \cdots \binom{N_1(N_0)}{x_{N_0}}}{\binom{T}{n}} & \text{if} \quad \omega \in \Omega_{GFD}, \\ 0 & \text{if} \quad \omega \in \Omega_{MB} - \Omega_{GFD}, \end{cases}$$

$$\begin{aligned} P_{MB}(\omega) &= \frac{n!}{x_1! \cdots x_{N_0}!} \left(\frac{1}{N_0}\right)^n \\ &= \frac{n!}{x_1! \cdots x_{N_0}!} \left(\frac{1}{N_0}\right)^{x_1} \cdots \left(\frac{1}{N_0}\right)^{x_{N_0}}. \end{aligned}$$

According to the Neyman–Pearson lemma, in order to distinguish the two hypotheses

$$\begin{aligned} &H_0 : \text{the underlying distribution is } MB, \\ &H_1 : \text{the underlying distribution is } GFD, \end{aligned}$$

it is optimal to study the ratio

$$\begin{aligned} \pi_0(\omega) = \frac{P_{GFD}(\omega)}{P_{MB}(\omega)} &= \frac{\left(\prod_{i=1}^{N_0} N_1(i)\,(N_1(i)-1)\cdots(N_1(i)-x_i+1)\right) N_0^n}{(T)\,(T-1)\cdots(T-n+1)} \\ &= \frac{\left(\prod_{i=1}^{N_0} N_1(i)\,(N_1(i)-1)\cdots(N_1(i)-x_i+1)\right)}{\overline{N_1}^n \left(1-\frac{1}{T}\right)\cdots\left(1-\frac{n+1}{T}\right)} \\ &= \prod_{i=1}^{N_0} \frac{N_1(i)}{\overline{N_1}} \frac{\prod_{i=1}^{N_0}\left(1-\frac{1}{N_1(i)}\right)\cdots\left(1-\frac{x_i-1}{N_1(i)}\right)}{\left(1-\frac{1}{N_0\overline{N_1}}\right)\cdots\left(1-\frac{n-1}{N_0\overline{N_1}}\right)}. \end{aligned}$$

The Neyman–Pearson lemma (cf. [18, section III, Theorem 1]) tells us that the distributions of the statistic π_0 with respect to the law P_{MB} (i.e., under the condition H_0) and with respect to the law P_{GFD} (i.e., under condition H_1) are maximally different. If the variance of N_1 is large, then using this statistic to test the null-hypothesis will require knowing a great deal about the generator and the particular cycle being used. The situation is greatly simplified if the $N_1(i)$ have little variance. Let's begin with

the situation when, for all $i, N_1(i) = \overline{N_1}$. In this case the expression for π_0 simplifies to

$$\pi_0(\omega) = \frac{\prod_{i=1}^{N_0}\left(1-\frac{1}{\overline{N_1}}\right)\cdots\left(1-\frac{x_i-1}{\overline{N_1}}\right)}{\left(1-\frac{1}{N_0\overline{N_1}}\right)\cdots\left(1-\frac{n-1}{N_0\overline{N_1}}\right)} = \frac{e^{\sum_{i=1}^{N_0}\sum_{s=1}^{x_i-1}\ln\left(1-\frac{s}{\overline{N_1}}\right)}}{e^{\sum_{j=1}^{n-1}\ln\left(1-\frac{j}{N_0\overline{N_1}}\right)}}.$$

In many cases this ratio $\pi_0(\omega)$, under the assumption that for all $i, N_1(i) = \overline{N_1}$, can be approximated by the simpler expression containing the quadratic forms, $x_i(x_i-1)$, of the occupation numbers x_i, $i = 1, 2, \ldots, N_0$, $\Sigma x_i = n$:

$$\pi_0(\omega) \approx e^{-\sum_{i=1}^{N_0}\frac{x_i(x_i-1)}{2\overline{N_1}}+\frac{n(n-1)}{2N_0\overline{N_1}}}.$$

Because of this, we choose to develop tests based on the statistic $Q_{n,N_1}(N_0) = \sum_{i=1}^{N_0} x_i(x_i-1)$. Since

$$\sum_{i=1}^{N_0} x_i(x_i-1) = \sum_{i=1}^{N_0} x_i^2 - \sum_{i=1}^{N_0} x_i = \sum_{i=1}^{N_0} x_i^2 - n,$$

we can deal instead with $X = \sum_{i=1}^{N_0} x_i^2$. This in turn is equivalent to the Pearson $\mathcal{X}^2$ statistic,

$$\mathcal{X}^2 = \sum \frac{\left(x_i - \frac{n}{N_0}\right)^2}{\sqrt{\frac{n}{N_0}}},$$

which (see [33, section IX, C]) for $n \gg N_0$ and $n, N_0 \to \infty$ has a $\chi^2_{N_0-1}$-distribution (which would certainly be approximately normal).

However, we are interested in keeping n as small as possible and in fact prefer to have $n \ll N_0$ if possible. Thus, in our region of interest, these asymptotic results are certainly questionable or downright invalid. There are ways around this: Chebyshev's inequality can be used, which of course is crude, or the *MB* experiment can be randomized, which reintroduces normality, but at a price. The experiment of randomly sampling ν from a Poisson distribution with parameter n and then running the $MB(\nu, N_0)$ experiment is equivalent to running N_0 independent Poisson processes each with parameter $\frac{n}{N_0}$. Berry–Esseen estimates for this randomized experiment can then be used to estimate bounds on significance levels in using the normal to approximate the sampling distribution. If $N_0 = 2^{31}$, the Berry–Esseen estimate is not useful for $n < 2^{26}$. This means that the randomized experiment is of no use below that number. But, for $N_0 = 2^{31}$, we are interested in $2^{17} \le n \le 2^{36}$, or so. For the statistic X, the variance under the randomized experiment is given by $Var(X) = 2N_0(\frac{n}{N_0})^2(1+2\frac{n}{N_0})$. The $2N_0(\frac{n}{N_0})^2$ factor in this expression is on the order of the variance of X under the nonrandomized experiment; however, the $(1+2\frac{n}{N_0})$ factor grows with $\frac{n}{N_0}$ and essentially renders the model useless beyond about $n = 2^{30}$,

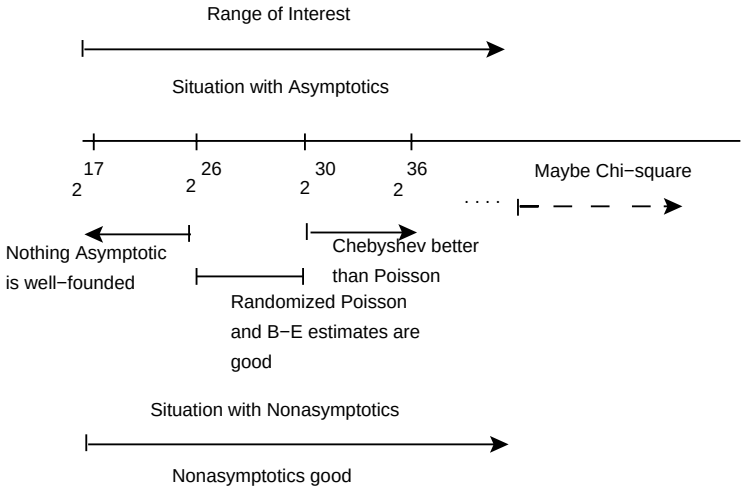

FIG. 2. *This diagram indicates regions of applicability of various asymptotic methods within the region of interest versus the much simpler situation for the nonasymptotic methods used in this paper.*

useless in the sense that it begins to perform worse than just using Chebyshev's inequality would. To go beyond 2^{30}, one has to appeal directly to Chebyshev's inequality until n is large enough to justify the Chi-square approximation to the multinomial. Even so, there are no rules for determining how large n needs to be in order to guarantee a given accuracy on the significance level. The inability to guarantee a bound on the significance level of tests based on asymptotic methods have plagued the testing of randomizers from the beginning. As far as we can tell, this problem has largely been ignored.

However, we can avoid all of these difficulties. We use MAPLE to calculate rational expressions in n, N_0 for higher order moments of the random variable X for the distribution $MB(n, N_0)$. These expressions are very complex and soon go beyond anyone's capacity to calculate by hand. From these expressions, we again use MAPLE to calculate exact rational values of the moments of X for appropriate ranges of values for n, N_0. These exact values now permit the use of the same classical moment theorems due to Chebyshev that were used in section 3 for the γ_s statistics. In fact, the situation now is better than it was then. The moment values in this case are really precise and no errors occur until the roots of the associated monic polynomial of step 2 (see (42)) are found. It is thus easier to control the errors in this case. In the end, for a given N_0, N_1, and significance level α, we are able to determine a critical value and a minimum sample size n needed to have a chance of distinguishing between X on $GFD(n, N_0, N_1)$ and X on $MB(n, N_0)$. Furthermore, all errors involved can be controlled precisely and the analyses are valid over the entire range of our interest without having to shift from one model to another. The diagram in Figure 2 provides a good visual summary of the situation.

Before proceeding with this program, let's consider the utility of X when N_1 is not assumed to be constant. Our test will compare $E_{GFD}(X)$ with $E_{MB}(X)$. This next result gives a very useful representation of $E_{GFD}(X)$.

THEOREM 6. *On* $GFD(n, N_0, N_1)$,

$$E_{GFD}(X) = n + \left[\frac{Var(N_1)}{\overline{N_1}} + \left(\overline{N_1} - 1\right)\right] \frac{n(n-1)}{N_0\overline{N_1} - 1}, \tag{44}$$

where

$$X = \sum_{i=1}^{N_0} x_i^2, \overline{N_1} = \frac{\sum_{i=1}^{N_0} N_1(i)}{N_0}, \text{ and } Var(N_1) = \frac{\sum_{i=1}^{N_0} (N_1(i) - \overline{N_1})^2}{N_0}.$$

Proof. Let $T = \sum_{i=1}^{N_0} N_1(i) = N_0\overline{N_1}$, and for each i, let $x_i = \epsilon_1(i) + \epsilon_2(i) + \cdots + \epsilon_{N_1(i)}(i)$, where $\epsilon_j(i)$ is the occupation number (either 0 or 1) of the jth box in ith tower of the *FD* model; then

$$E_{GFD}(\varepsilon_j(i)) = \frac{\binom{T-1}{n-1}}{\binom{T}{n}} = \frac{n}{T},$$

$$E_{GFD}(x_i) = \frac{nN_1(i)}{T},$$

$$E_{GFD}(\varepsilon_i\varepsilon_j) = \frac{\binom{T-2}{n-2}}{\binom{T}{n}} = \frac{n(n-1)}{(T)(T-1)},$$

$$\begin{aligned} E_{GFD}(x_i^2) &= N_1(i)E_{GFD}(\varepsilon_1(i)) + N_1(i)(N_1(i)-1)E_{GFD}(\varepsilon_1\varepsilon_2) \\ &= \frac{nN_1(i)}{T} + \frac{N_1(i)(N_1(i)-1)n(n-1)}{(T)(T-1)}, \end{aligned}$$

$$\begin{aligned} E_{GFD}(X) = E_{GFD}\left(\sum_{i-1}^{N_0} x_i^2\right) &= \frac{n\left(\sum_{i=1}^{N_0} N_1(i)\right)}{T} + \frac{n(n-1)\left(\sum_{i=1}^{N_0} N_1(i)(N_1(i)-1)\right)}{(T)(T-1)} \\ &= n + \frac{n(n-1)\left(\sum_{i=1}^{N_0} N_1(i)(N_1(i)-1)\right)}{(T)(T-1)} \\ &= n + \left[\frac{Var(N_1)}{\overline{N_1}} + \left(\overline{N_1} - 1\right)\right] \frac{n(n-1)}{N_0\overline{N_1} - 1}. \end{aligned}$$

□

If the $Var(N_1) = 0$, we get from (44) the expression for $E_{GFD}(X)$ under the assumption that for all $i, N_1(i) = \overline{N_1}$. However, it is clear from (44) that X will remain a valuable test statistic as long as the $\frac{Var(N_1)}{\overline{N_1}}$ term is small. If this term is large, then one might hope that $E_{GFD}(X)$ significantly overshoots $E_{MB}(X) = n + \frac{n(n-1)}{N_0}$, and we might still have a test that can discriminate between *GFD* and *MB*. As we will see, this in fact can happen in important cases.

We first deal with the case when $\frac{Var(N_1)}{\overline{N_1}}$ is negligible. Our strategy is the following: For a given $N_0, \overline{N_1}$ and a choice of significance level $\alpha > 0$, find an m such that the Chebyshev m-step distribution determined by the first $2m$ moments of the distribution function $P_{MB}(X \le y)$ has the property that $A_1 \le \alpha$ and $E_{GFD}(X) < a_1$. Then a_1 is a suitable critical value for rejection of the hypothesis H_0 that the distribution is $MB(n, N_0)$. To implement this strategy, we need to be able to calculate (a)

$E_{GFD}(X)$ and (b) $E_{MB}(X^k)$ for $k = 0, 1, 2, \ldots, 2m-1$ for a suitably chosen positive integer m. We have already calculated (a) above,

$$E_{GFD}(X) = n + \left(\overline{N_1} - 1\right) \frac{n(n-1)}{N_0 \overline{N_1} - 1}.$$

We are dropping the $\frac{Var(N_1)}{\overline{N_1}}$ term in (44) since we are presently assuming that it is negligible.

For *MB* we need to calculate $E_{MB}(X^k)$ for arbitrary k. We define $y_i = x_i^2$ so that

$$X^k = \left(\sum_{i=1}^{N_0} x_i^2\right)^k = \left(\sum_{i=1}^{N_0} y_i\right)^k.$$

The terms of the rightmost expression are of the form $y_{i_1}^{j_1} y_{i_2}^{j_2} \ldots y_{i_k}^{j_k}$, where

$$j_1 + j_2 + \cdots + j_k = k \text{ and some } j_i \text{ can be } 0.$$

We can assume these have been arranged in nonincreasing order so that $j_1 \geq j_2 \geq j_3 \geq \cdots \geq j_k$. Furthermore, for any such term,

$$E_{MB}\left(y_{i_1}^{j_1} y_{i_2}^{j_2} \ldots y_{i_k}^{j_k}\right) = E_{MB}\left(y_1^{j_1} y_2^{j_2} \ldots y_k^{j_k}\right).$$

Each term of the form $y_1^{j_1} y_2^{j_2} \ldots y_k^{j_k}$ can be uniquely associated with the nonincreasing finite sequence $(j_1, j_2, \ldots, j_k)$. Thus, there are three tasks involved in calculating a kth moment:

1. Generate all of the distinct, nonincreasing sequences $(j_1, j_2, \ldots, j_k)$. We call those sequences *admissible*.

2. Determine the number of terms in the expansion of $(\sum_{i=1}^{N_0} y_i)^k$ associated with each of these sequences.

3. Calculate $E_{MB}(y_1^{j_1} y_2^{j_2} \ldots y_k^{i_k})$ for each sequence.

We wrote a MAPLE procedure for generating the sequences $(j_1, j_2, \ldots, j_k)$. For example, the outcome of this procedure for $k = 4$ produces the following admissible sequences:

$$(1,1,1,1),\ (2,1,1,0),\ (2,2,0,0),\ (3,1,0,0),\ (4,0,0,0).$$

The number of terms associated with each sequence is given by the formula

$$\begin{aligned} &\#(j_1, j_2, \ldots, j_k) \\ &= \binom{k}{j_1, j_2, \ldots, j_k} \binom{N_0}{\beta_1} \binom{N_0 - \beta_1}{\beta_2} \cdots \binom{N_0 - \sum_{j=1}^{s-1} \beta_j}{\beta_s}, \end{aligned} \tag{45}$$

where the β_i are the multiplicities of the distinct nonzero j_r. More precisely, s is the number of distinct nonzero numbers in the list $\langle j_{s_1}, \ldots, j_{s_k} \rangle$ and $\beta_1, \ldots, \beta_m$ are the corresponding multiplicities of those nonzero numbers arranged in descending order. To establish the formula (45) we use the multinomial expansion for $(\sum_{i=1}^{N_0} y_i)^k$:

$$\begin{aligned} &\sum_{j_1 + j_2 + \cdots + j_{N_0} = k} \binom{k}{j_1, j_2, \ldots, j_{N_0}} y_1^{j_1} y_2^{j_2} \cdots y_{N_0}^{j_{N_0}} \\ &= \sum_{j_{s_1} + j_{s_2} + \cdots + j_{s_k} = k} \binom{k}{j_{s_1}, j_{s_2}, \ldots, j_{s_k}} y_{i_{s_1}}^{j_{s_1}} y_{i_{s_2}}^{j_{s_2}} \cdots y_{i_{s_k}}^{j_{s_k}}, \end{aligned}$$

TABLE 4

Based on appropriate Chebyshev 7-step functions we calculate the parameters needed for testing the relevant statistical hypotheses.

1	2	3	4	5	6	7	8	9
n	N_1	E_{GDF}	E_{MB}	a_1	$A_1{=}\alpha$	$\sum_1^5 A_i$	B_1	B_6
2^{17}	1	131072	131080	131072.45	0.04	.9978	.45	16
2^{19}	1	524288	524416	524364	.0016	.9838	76	44
2^{19}	2	524352	524416	524364	.0016	.9838	12	44
2^{22}	2^3	4201471	4202495	4202024	.00062	.9711	553	309
2^{22}	2^4	4201983	4202495	4202024	.00062	.9711	41	309
2^{24}	2^5	16904191	16908287	16906376	.00057	.9693	21822	1218
2^{24}	2^6	16906239	16908287	16906376	.00057	.9693	137	1218
2^{26}	2^7	69189631	69206015	69198344	.00055	.9689	8712	4853
2^{26}	2^8	69197823	69206015	69198344	.00055	.9689	520	4853
2^{28}	2^9	301924351	301959175	301924351	.00055	.9687	34823	19396
2^{28}	2^{10}	381957119	301959175	301924351	.00055	.9687	2055	19396
2^{30}	2^{11}	1610350592	1610692730	1610489858	.00055	.9687	139267	77566
2^{30}	2^{12}	1610481664	1610692730	1610489858	.00055	.9687	8195	77566
2^{32}	2^{13}	12883853310	12884901890	12884410350	.00055	.9687	557042	310245
2^{32}	2^{14}	12884377600	12884901890	12884410350	.00055	.9687	32754	310245
2^{34}	2^{15}	154614628300	154618822600	154616856500	.00055	.9687	2228141	1221096
2^{34}	2^{16}	159616725500	154618822600	154616856500	.00055	.9687	130989	1221096

assuming that

$$j_{s_1} \geq j_{s_2} \geq j_{s_3} \geq \cdots, \text{ where } j_{s_r} \geq 0 \text{ and}$$

$$j_{s_1} + j_{s_2} + \cdots + j_{s_k} = k.$$

The task of picking the $y_{i_{s_r}}$ corresponding to the nonzero j_{s_r} can be recognized as the task of picking the indices that will go with distinct values of the j_{s_r}. There will be $\binom{N_0}{\beta_1}$ ways of choosing the set $\{y_{i_{s_1}}, y_{i_{s_2}}, \ldots, y_{i_{s_{\beta_1}}}\}$, $\binom{N_0-\beta_1}{\beta_2}$ ways of choosing the set $\{y_{i_{s_{\beta_1}+1}}, \ldots, y_{i_{s_{\beta_1+\beta_2}}}\}$, and so on.

Finally, $E_{MB}(y_1^{j_1} y_2^{j_2} \ldots y_k^{i_k})$ may be calculated using the generating function

$$g_k = \left(1 - \frac{k}{N_0} + \frac{e^{\alpha_1}}{N_0} + \frac{e^{\alpha_2}}{N_0} + \cdots + \frac{e^{\alpha_k}}{N_0}\right)^n.$$

Another MAPLE procedure we wrote calculates $\#(j_1, j_2, \ldots, j_k) \cdot E_{MB}(y_1^{j_1} y_2^{j_2} \ldots y_k^{i_k})$ for each admissible $(j_1, j_2, \ldots, j_k)$.

The two MAPLE procedures mentioned above can be used to calculate the kth moments of X for $k = 1, 2, \ldots$.

Since the moments M_j in the system (41) are rational, MAPLE finds exact values for the unknowns c_j. In step 2 using MAPLE we can find the roots of the polynomial (42) with any desired accuracy but not exactly. By choosing appropriate tolerances in step 2 we controlled the accuracy of A_j in system (43). In particular, the numbers reported in Table 4 have the indicated accuracy.

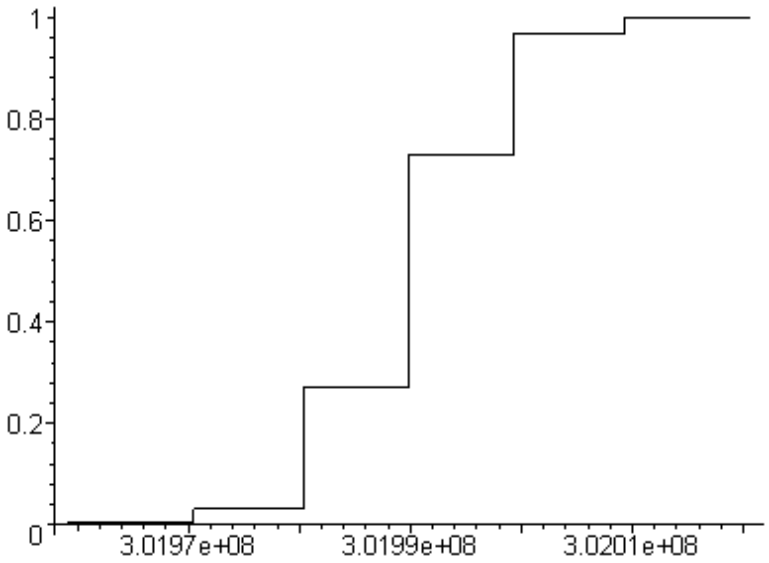

FIG. 3. *This is the plot of the Chebyshev 7-step function for the random variable X on $MB(n, N_0)$, where $N_0 = 2^{31}$, $n = 2^{28}$. The MAPLE procedure produces $a_1 = 3.0192 \cdot 10^8$ and $A_1 = .00054$.*

In Table 4, we use the Chebyshev 7-step function for X on $MB\left(n, 2^{31}\right)$ to determine the minimum sample sizes n needed to distinguish between $MB\left(n, 2^{31}\right)$ and $GDF\left(n, 2^{31}, N_1\right)$ at the calculated significance level, where $1 \leq N_1 \leq 2^{16}$. Since when $N_1 = 1$ $E_{GFD}(X) = n$, it follows that every PRNG with no hidden bits will fail the test in line 1 of Table 4. In Figure 3, we plot the Chebyshev 7-step function for X on $MB(2^{28}, 2^{31})$.

The only thing questionable about this test is the fact that $\alpha = .04$. However, by running the test two or three times, one can achieve satisfactory significance levels and still reject the hypothesis that the model is *MB*. Thus, tests for which $\alpha \leq .001$ will reject the hypothesis *MB* based on a sample of no more than 2^{19} points, if these points are being generated by a PRNG with the NSI property. The last column in the table is being used to give some indication of the power of the tests. The upward variability of the *GFD* model is less than that of the *MB* model for the statistic X. The column headed by $\sum_1^5 A_i$ gives a lower bound on $P_{MB}(X \leq a_6)$; thus we expect, but have not proved, that

$$P_{GFD}(X \leq E_{GFD}(X) + a_6 - E_{MB}(X)) \geq P_{MB}(X \leq a_6 - E_{MB}(X)).$$

Our expectation, therefore, is that the tests for which the $B_1 = a_1 - E_{GFD}$ entry is greater than the $B_6 = a_6 - E_{MB}$ entry will have power at least that of the entry in column 7. Thus, if $N_1 = 2^5$ and we run a test with $n = 2^{24}$ and the real underlying model is the *GFD* model, then the statistic X produced should fail the test at least 97% of the time. For the other rows where column 8 is less than column 9, we don't presume anything about the power of the test, but the test is still valid and can be tried.

We add to Table 4 this next table (Table 5) which provides some idea concerning what happens when N_0 is also allowed to vary.

TABLE 5

n	$N = N_0$	$N_1 = \overline{N_1}$	$E_{GFD}(X)$	a_1	A_1
2^{15}	2^{10}	2^7	1073128	1076416	.0011
2^{24}	2^{10}	2^{16}	27489047600	274892157900	.0011
2^{38}	2^{10}	2^{16}	$7378697650 \cdot 10^{10}$	$7378697653 \cdot 10^{10}$	.0011
2^{26}	2^{15}	2^{16}	137503963200	137504129300	.0006
2^{23}	2^{21}	2^{10}	41910268	41912390	.0006
2^{24}	2^{21}	2^{10}	150863864	150933636	.0006

The values of n in this table are the smallest powers of 2 for which $E_{GFD}(X) \leq a_1$ for the given N_0 and N_1.

Now, if $\frac{Var(N_1)}{\overline{N_1}}$ is not negligible, one could not expect the above approach to lead to a test that would detect the difference between the *GFD* and *MB* models at hand. The difference between $E_{MB}(X) = n + \frac{n(n-1)}{N_0}$ and the term $n + (\overline{N_1} - 1)\frac{n(n-1)}{N_0\overline{N_1}-1}$ is $\frac{n(n-1)}{N_0\overline{N_1}-1}(1 - \frac{1}{N_0})$. If N_0 is large, then this difference is on the order of $\frac{n(n-1)}{N_0\overline{N_1}-1}$. But then, $E_{GFD}(X) - E_{MB}(X) = n + (\overline{N_1} - 1)\frac{n(n-1)}{N_0\overline{N_1}-1} + \frac{Var(N_1)}{\overline{N_1}}(\frac{n(n-1)}{N_0\overline{N_1}-1}) - E_{MB}(X) \approx (\frac{Var(N_1)}{\overline{N_1}} - 1)(\frac{n(n-1)}{N_0\overline{N_1}-1})$. Since the X statistic on $MB(n, N_0)$ for much of the ranges of n, N_0 of interest is nearly symmetric, if $n + (\overline{N_1} - 1)\frac{n(n-1)}{N_0\overline{N_1}-1} < a_1$ and $\frac{Var(N_1)}{\overline{N_1}} \geq 2$, there is a good chance that $E_{GFD}(X) > a_7$. Based on these observations as well as the calculations in Tables 4 and 5, we offer the following empirical rule for the testing of *GFD* versus *MB*.

REMARK 7 (empirical rule for testing *GFD* versus *MB*). *Let N_0 and N_1 be known.*

- *If $\frac{Var(N_1)}{\overline{N_1}}$ is negligible, then a test for $GFD(n, N_0, N_1)$ versus $MB(n, N_0)$ is feasible for $n \geq n^*$, where $n^* \approx 2^{\frac{\log_2(N_0)}{2} + \log_2(\overline{N_1}) + 3}$. (Reject $MB(n, N_0)$ if $X \leq a_1$.)*
- *If $\frac{Var(N_1)}{\overline{N_1}}$ is large (≥ 2), then a test for $GFD(n, N_0, N_1)$ versus $MB(n, N_0)$ is feasible for $n \geq n^*$, where $n^* \approx 2^{\frac{\log_2(N_0)}{2} + \log_2(\overline{N_1}) + 3}$. (Reject $MB(n, N_0)$ if $X \geq a_7$).*
- *If $\frac{Var(N_1)}{\overline{N_1}}$ is moderate, we don't expect X to be able to distinguish GFD and MB.*

Note that if $\frac{Var(N_1)}{\overline{N_1}}$ is significantly larger than 2, then somewhat smaller values for n^* might be tried than the one indicated above.

4.2. Tests of MB versus $(MB)^C$ guided by the empirical rule. Based on insights from the above, Table 6 was generated to support tests for various PRNGs with hidden states.

We ran experiments to test $F(17, 5, (xor))$, $F(31, 13, (xor))$, *TGGL*, and $F(17, 5, +)$ with the results shown in Table 7.

Based on these results

- we reject *MB* for $F(17, 5, (xor))$ at the .0011 significance level using the first row of Table 6 ($X = 1073122 < 1076416 = a_1$);
- we reject *MB* for $F(31, 13, (xor))$ at the .0006 significance level using the 5th row of Table 6;
- we reject *MB* for *TGGL* at the .0006 significance level using the last row of Table 6;

TABLE 6

n	$N{=}N_0$	a_1	A_1	a_7	A_7
2^{15}	2^{10}	1076416	.0011	1087338	.0003
2^{22}	2^{10}	17183431930	.0011	17184829250	.0003
2^{23}	2^{10}	68726602210	.0011	68729396850	.0003
2^{36}	2^{10}	4611686076799232761	.0011	46116860996928967299	.0003
2^{26}	2^{15}	137504129300	.0006	137508062500	.0005
2^{23}	2^{21}	41912391	.0006	41973838	.0005
2^{24}	2^{21}	150933636	.0006	151056531	.0005
$8 \cdot 10^6$	2^{21}	38488348	.0006	38546949	.0005

TABLE 7

Generator	N_0	N_1	n	X = test statistic
$F(17,5,(xor))$	2^{10}	2^3	2^{15}	1073122
$F(31,13,(xor))$	2^{15}	2^{16}	2^{26}	137503783146
$TGGL$	2^{21}	2^{10}	$8{\cdot}10^6$	38482128
$F(17,5,+)$	2^{10}	$\overline{N_1} = 2^{16}$	2^{22}	17185098488
$F(17,5,+)$	2^{10}	$\overline{N_1} = 2^{16}$	2^{23}	68730103632

- we reject *MB* for $F(17,5,+)$ at the .0003 significance level using either row 4 of Table 7 with row 2 of Table 6 or row 5 of Table 7 with row 3 of Table 6. In this case, the rejection is based on the fact that $X > a_7$.

Note that in the first three cases, the sample size n corresponds to that predicted by the empirical rule. For example, the empirical rule indicates the X statistic should distinguish between $GFD(n, 2^{10}, 2^7)$ and $MB(n, 2^{10})$ for $n \geq 2^{5+3+7} = 2^{15}$. In the case of $F(17,5,+)$ though, the empirical rule would call for a sample size $n = 2^{5+3+16} = 2^{24}$. We tried the smaller sizes because several full period runs of the generator indicated that $\frac{Var(N_1)}{\overline{N_1}}$ was ranging between 3 and 6. The run for $n = 2^{23}$ is actually a fairly typical run and the generator will be rejected at the .0003 level of significance most of the time. In the case of $n = 2^{22}$, we had to try a few runs, but not many, to find a seed that produced $\frac{Var(N_1)}{\overline{N_1}} \approx 5.3$. The $X = 17185098488$ reported in row 4 of Table 7 resulted from the experiment run with that seed.

Row 4 of Table 6 we believe would support the rejection of *MB* for the generator $F(31,13,+)$, but we have not run the test because of our limited computer resources. It is feasible, however, to run this test with current hardware. The empirical rule would indicate a need for a sample of size 2^{38}; however, if this generator acts similarly to $F(17,5,+)$, we expect the discrepancy to show up a little sooner.

If a statistic correlates significantly with X, one would expect its distribution under *MB* and *GFD* to be different. Thus, such a statistic also should be able to detect the difference between these two distributions and hence might serve as a basis for testing PRNGs. Until we did the above analysis, we had been unable, when a generator had a significant number of hidden states, to demonstrate any significant difference using the occupation statistics γ_s. But on $MB(n,N)$, $X = \sum_{s=1}^{n} \gamma_s s^2$. Furthermore, any actual sum will run over a relatively short number of indices, say $K << n$. So, it should be possible to use some suitably chosen γ_s as a test statistic. This in fact turns out to be the case, but we need to use everything we have learned in order to help aim the test at the right place. To illustrate, consider $F(17,5,(xor))$. For reasons we will explain in a little bit, we consider $N_0 = 2^{14}$. Then $N_1 = 2^3$

TABLE 8

s	$n = 2^{13}$		$n = 2^{14}$	
	a_1	A_1	a_1	A_1
2	1131	.0001	2834	.0001
3	156	.0002	897	.0001
4	9	.0008	194	.0002
5	0	.08	26	.0005
6	NA	NA	1	.005

(essentially) and the empirical rule predicts that X will detect the difference around $n = 2^{13}$ or 2^{14}. Table 8 gives the cutoffs required to reject *MB* at the indicated significance levels.

For $n = 2^{13}$, on three out of seven runs $\gamma_5 = 0$, and on three out of seven runs $\gamma_4 \leq 9$. The strongest evidence though is that on all seven runs, $\gamma_3 \leq 172 = a_2$ and $A_2 + A_1 = .014$ and this permits rejection of *MB* at the 10^{-12} significance level.

For $n = 2^{14}$, we only made two runs. Both times, $\gamma_6 \leq 1, \gamma_5 \leq 26$, and $\gamma_4 \leq 194$.

The reason we chose $N_0 = 2^{14}$ for the above illustration is that the empirical rule applied to the given generator indicates a sample size for the X statistic that is on the same order or a little smaller than N_0. In most cases when one has $N_1 > 1$, it is necessary to take n somewhat larger than N_0. If n is larger than N_0, the matrix of coefficients in the first step (41) of the Chebyshev moment method becomes ill-conditioned, especially for small s. It therefore becomes challenging to control the errors in applying this method. Even in the case considered, we had to take great care to control errors. This, by the way, is one advantage of using the X statistic. We were able to calculate the relevant moments for it without any errors at all and therefore the condition of the matrix of coefficients in the first step of the process was of no concern. Because of certain terms in the factorial moment formulas for the γ_s-statistics, it is often necessary to approximate the moments in (41).

5. Description of other conducted tests. The development of tests based on hard estimates of tail probabilities and employing nonasymptotic methods constitute the major thrust of this paper and almost all that is new in it. However, part of our original motivation was to apply appropriate tests for possibly true RNGs. Because of this we developed a number of more standard tests that are based on traditional asymptotic methods. We describe these tests here and then, in the final section, summarize the results of all our tests on the various generators considered in this paper.

Homogeneity test. The binary digits are divided into ix groups of equal length jx each; we find the quantity

$$s = \sum_{i=0}^{ix-1} \sum_{b=0}^{1} \frac{\left(y(i,b) - \frac{jx}{2}\right)^2}{\frac{jx}{2}},$$

where $y(i,b)$ is the number of elements equal to b in the ith group. The random variable s should have approximately a χ^2-distribution with $ix-1$ degrees of freedom. We find $F(s)$, where $F(\cdot)$ is the corresponding distribution function. This quantity $F(s)$ is called "the left tail" in the program output; a generator passes the test if $F(s)$ is between $1 - \frac{\alpha}{2}$ and $\frac{\alpha}{2}$, where α is the level of significance; for instance, $0.005 < F(s) < 0.995$, if $\alpha = 0.99$.

Arcsine test. We use the (presumably independent and symmetrically distributed) random bits to simulate a random walk. We generate a large number of

independent trajectories of such a walk, all of them having the same (large) time length. For each trajectory we find the fraction of time that the particle spends on the positive half-axis. This random variable should have the so-called arcsine distribution. We compare graphically two curves: the sample distribution of this random variable and its theoretical (arcsine) distribution.

Correlation function test. We find the sample correlation function for a long sample of consecutive random bits. Its values at different integer points should be approximately normally distributed and independent (if we consider not too many integer points). We find the sample distribution of these values, normalize it so that its theoretical distribution becomes approximately uniform in $[0, 1]$, and then compare it with this theoretical distribution.

Collision times test. Note this is not the same as the "first collision test" discussed in section 2.3. We combine consecutive random bits into words of some fixed length. We generate these random words and fix the moment of the first coincidence of the new word with some previously generated one. Then we find the moment of the second repetition of words (this new repeating word may be the same or different from the word that repeated previously). Similarly we find the moment of the third repetition. Theory predicts that if the length of the words is large enough, then certain combinations of these moments should be approximately uniformly distributed in $[0, 1]$ (and statistically independent). Repeating our experiment many times, we obtain sample distributions of these random variables and then compare them with the theoretical (uniform) distribution.

Short range correlation test. We fix the length of a binary word, generate a large number of words, and find the frequency of appearance of each possible word. A particular χ^2 sum constructed from these frequencies should have (approximately) a prescribed χ^2-distribution. We find the probability that a random variable with this χ^2-distribution does not exceed the value that we found. This probability depends on the sample and hence is a random variable. It should be distributed (approximately) uniformly in $[0, 1]$. We repeat our experiment many times, find the sample distribution of this random variable, and check how far it is from the theoretical (uniform) distribution. This deviation, denoted by D_{KS}, is measured by the Kolmogorov–Smirnov statistic. Then we find the probability, p_{KS}, for a Kolmogorov–Smirnov random variable not to exceed the value D_{KS}. The generator fails the test if $p_{KS} < 1 - \frac{\alpha}{2}$ or $p_{KS} > \frac{\alpha}{2}$, where α is the level of significance.

Maximal triple correlation test. We choose two integers, n and m, and for each pair i, j such that $0 < i \le j \le m$ we find the sum

$$s(i,j) = \sum_{k=0}^{n-1} (2\beta_k - 1)\left(2\beta_{k+i} - 1\right)\left(2\beta_{k+j} - 1\right),$$

where β_k is the kth integer (0 or 1) provided by the RNG. Each of the quantities

$$t(i,j) = \frac{s(i,j)}{\sqrt{n}}$$

has (approximately) a standard normal distribution. We evaluate

$$T(n,k) = \max_{0<i\le j\le m} |t(i,j)|,$$

which is therefore the maximum of (absolute values of) $m \times m$ normalized sample triple correlations of the sequence β_k. At the same time it is (approximately) the

TABLE 9
Summary of "other" tests results.

		TRNG	R250	RAND	GGL	random
H	Homogeneity	P/2M	P/2M	F/2M	P/2M	P/2M
A	Arcsine	P/.5M	F/.5M	F/.5M	P/.5M	P/.5M
C	Correlation function	P/.26M	P/.26M	F/.26M	P/.26M	F/.26M
B	Collision times birthdays	P/17M	P/17M	F/17M	P/17M	P/17M
S	Short correlations	P/4.6M	P/4.6M	F/4.6M	P/4.6M	P/4.6M
T	Max. triple correlations	P/4K	F/4K	P/4K	P/4K	F/4K

TABLE 10
Summary of the data sizes used in the tests of Table 9.

	The number of bits used
Homogeneity	2,097,152
Arcsine	538,624
Correlation function	262,444
Collision times birthdays	About 3405 bits for the first collision to occur
Short correlations	4,608000
Max. triple correlations	4000

maximum of the absolute values of the standard normal random variables. These random variables are strongly dependent, which makes it difficult to find the distribution of that maximum; hence we find an upper estimate, p_T, of the probability of the event that the above maximum is greater than $T(n,k)$. This probability is

$$p_T = n^2\sqrt{\frac{2}{\pi}}\int_{T(n,k)}^{\infty}\exp\left(-\frac{x^2}{2}\right)dx.$$

We reject the hypothesis that the sequence β_k is a symmetric Bernoulli sequence, if p_T is too small, say, less than 10^{-4} (which means that the maximal triple correlation is too far from 0). In fact, instead of the sum $s(i,j)$, we compute a slightly different sum $s^*(i,j)$ that consists of independent terms; this change enhances the accuracy of the normal approximation.

6. Testing results for some generators. The test results are presented in Table 9 where **P** means "passing," **F** means "failing," M is 10^6, K is 10^3, B means "beyond available computing resources," and F/T means "fails on theoretical grounds." The data sizes used in the testing are collected in Table 10.

In Table 11, we summarize the results of the tests based on first collisions, FCT, occupation statistics, γ_s, and the X statistic. These are developed and fully discussed in sections 3 and 4. Here $FF^*(xor)$ stands for an xor lagged Fibonacci seeded by a full-rank matrix. The tests of *RAND*, *RANDU*, *GGL*, and $FF^*(xor)$ are theoretical with the sample sizes required for rejection determined by the empirical rule of Remark 7.

TABLE 11

	FCT	γ_s	X
RAND	F	F	F
RANDU	F	F	F
GGL	F	F	F
$FF^*(xor)$	F	F	F
$F(17,5,+)$	NR	F	F
TGGL	NR	NR	F
$F(17,5,xor)$	NR	F	F
$F(31,13,xor)$	NR	NR	F

REFERENCES

[1] S. Aluru, *Lagged Fibonacci random number generators for distributed memory parallel computers*, J. Parallel Distrib. Comput., 45 (1997), pp. 1–12.

[2] S. L. Anderson, *Random number generators on vector supercomputers and other advanced architectures*, SIAM Rev., 32 (1990), pp. 221–251.

[3] E. A. Bender, *Asymptotic methods in enumeration*, SIAM Rev., 16 (1974), pp. 485–515.

[4] K. Binder, *Applications of Monte Carlo methods to statistical physics*, Rep. Progr. Phys., 60 (1997), pp. 487–559.

[5] M. N. Barber, R. B. Pearson, D. Toussaint, and J. L. Richardson, *Finite-size scaling in the three-dimensional Ising model*, Phys. Rev. Lett. B, 32 (1985), pp. 1720–1730.

[6] R. P. Brent, *On the periods of generalized Fibonacci recurrences*, Math. Comp., 63 (1994), pp. 389–401.

[7] J. Debieerre and R. Bradley, *Fragmentation of percolation cluster perimeters*, J. Phys. A, 29 (1996), pp. 2337–2348.

[8] L. Devroge, *Non-Uniform Random Number Generation*, Springer, Berlin, 1986.

[9] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1, 3rd ed., John Wiley and Sons, New York-London-Sydney, 1968.

[10] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 2, 2nd ed., John Wiley and Sons, New York-London-Syndney, 1971.

[11] A. M. Ferrenberg, D. P. Landau, and Y. J. Wong, *Monte Carlo simulations: Hidden errors from "good" random number generators*, Phys. Rev. Lett., 69 (1992), pp. 3382–3384.

[12] B. Gammel, *Hurst's rescaled range statistical analysis for pseudorandom number generators used in physical simulations*, Phys. Rev. E (3), 58 (1998), pp. 2586–2597.

[13] H. Geiringer, *Sur les variables aléatoires arbitrairement liées*, Rev. Math. Union Interbalkan., 2 (1938), pp. 1–26.

[14] B. Harris and C. Park, *The limiting distribution of the sample occupancy numbers from the multimonomial distribution with equal cell probabilities*, Ann. Inst. Statist. Math., 23 (1971), pp. 125–133.

[15] M. Isichenko, *Percolation, statistical topology, and transport in random media*, Rev. Modern Phys., 64 (1992), pp. 961–1043.

[16] D. E. Knuth, *The Art of Computer Programming, Vol.* 2*: Seminumerical Algorithms*, 2nd ed., Addison-Wesley, Reading, MA, 1981.

[17] R. Kubo, *Statistical Mechanics*, North-Holland, Amsterdam, 1965.

[18] E. L. Lehmann, *Testing Statistical Hypotheses*, 2nd ed., Springer Texts in Statist., Springer-Verlag, New York, 1997.

[19] A. Law and W. Kelton, *Simulation Modeling and Analysis*, 2nd ed., McGraw-Hill, New York, 1991.

[20] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, CBMS-NSF Regional Conf. Ser. in Appl. Math. 63, SIAM, Philadelphia, 1992.

[21] G. Marsaglia, *Random numbers fall mainly in the planes*, Proc. Natl. Acad. Sci. USA, 61 (1968), pp. 25–28.

[22] G. Marsaglia, *The structure of linear congruential sequences*, in Applications of Number Theory to Numerical Analysis, S. K. Zaremba, ed., Academic Press, New York, 1972, pp. 249–285.

[23] G. Marsaglia, *A current view of random number generators*, in Computer Science and Statistics, L. Billard, ed., North-Holland, Amsterdam, New York, 61 (1985), pp. 3–10.

[24] A. Menezes, P. Oorschot, and S. Vanstone, *Handbook of Applied Cryptography*, CRC Press, Boca Raton, FL, 1996.

[25] S. Pincus and B. H. Singer, *Randomness and degrees of irregularity*, Proc. Natl. Acad. Sci. USA, 93 (1996), pp. 2083–2088.

[26] S. Pincus and R. E. Kalman, *Not all (possibly) "random" sequences are created equal*, Proc. Natl. Acad. Sci. USA, 94 (1997), pp. 3513–3518.

[27] R. Y. Rubinstein and B. Melamed, *Modern Simulation and Modeling*, Wiley Ser. Probab. Stat.: Appl. Probab. Stat., John Wiley and Sons, New York, 1998.

[28] B. Schneider, *Applied Cryptography*, 2nd ed., John Wiley and Sons, New York-Toronto, 1996.

[29] L. Shchur and H. Blöte, *Cluster Monte Carlo: Scaling of systematic errors in the two-dimensional Ising model*, Phys. Rev. E (3), 55 (1997), pp. 4905–4908.

[30] S. Tezuka, *Uniform Random Numbers: Theory and Practice*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1995.

[31] Yu. N. Tyurin and V. E. Figurnov, *On the testing of random number generators*, Theory Probab. Appl., 35 (1990), pp. 180–183.

[32] I. Vattulainen, K. Kankaala, J. Saarinen, and T. Ala-Nissila, *A comparative study of some pseudorandom number generators*, Comput. Phys. Comm., 86 (1995), pp. 209–226.
[33] R. von Mises, *Mathematical Theory of Probability and Statistics*, Academic Press, New York-London, 1964.
[34] Z. Yin, *New methods for simulation of fractional Brownian motion*, J. Comput. Phys., 127 (1996), pp. 66–72.
[35] B. Zhang, M. Gyulassy, and Y. Pang, *Equation of state and collision rate tests of parton cascade models*, Phys. Rev. C (3), 58 (1998), pp. 1175–1182.