PHYSICAL REVIEW E VOLUME 52, NUMBER 5 NOVEMBER 1995

Operational conditions for random-number generation

A. Compagner
Laboratory of Applied Physics, P.O. Box 5046, 2600 GA Delft, The Netherlands

(Received 19 April 1995)

Ensemble theory is used to describe arbitrary sequences of integers, whether formed by the decimals of π or produced by a roulette or by any other means. Correlation coefficients of any range and order are defined as Fourier transforms of the ensemble weights. Competing definitions of random sequences are considered. Special attention is given to sequences of random numbers needed for Monte Carlo calculations. Different recipes for those sequences lead to correlations that vary in range and order, but the total amount of correlation is the same for all sequences of a given length (without internal periodicities). For maximum-length sequences produced by linear algorithms, most correlation coefficients are zero, but the remaining ones are of absolute value 1. In well-tempered sequences, these complete correlations are of high order or of very long range. General conditions to be obeyed by random-number generators are discussed and a qualitative method for comparing different recipes is given.

PACS number(s): 02.70.Lq, 02.50.Ng, 64.60.Cn, 75.40.Mg

I. INTRODUCTION

Random sequences are difficult to define. The mathematical problems involved were examined by Von Mises [1], Kolmogorov [2,3], Chaitin [4], Kac [5], Martin-Löf [6,7], Kirschenmann [8], Van Lambalgen [9,10], and many other authors. Random sequences are easy to generate in a computer, except when needed for reliable Monte Carlo simulations. This practical problem was also widely discussed; here, only the work of Golomb [11], Knuth [12], Marsaglia [13], Ripley [14], James [15], L'Ecuyer [16], and Niederreiter [17] is mentioned. In comparison, the connection between the two difficulties has received little attention.

In the case of binary sequences [18–20], operational definitions of randomness leading to reliable recipes for random-number sequences could be based on ensemble theory and on a hierarchy of correlation coefficients of arbitrary range and order. A generalization of these ideas is desirable, since many recipes for random-number generation produce sequences of integers modulo m. The present generalization finds its origin in statistical mechanics, and is also related to complexity theory and to the theory of multivariate distributions. Its first aim is to attach a precise meaning to notions like randomness and a truly random sequence. Several definitions are needed to cope with different situations.

The generalization turns out to be closely related to the Fourier analysis of linear-congruence sequences carried out, many years ago, by Coveyou and MacPherson [21]. Their unified theory, which led to the spectral test that is by now a standard technique for these sequences, foreshadows many of the present results, but details and context differ. The emphasis of their paper is on numerical criteria for special cases, whereas the main aim of the present paper is to find a general approach to the problem of random-number generation in terms of qualitative though operational conditions.

Such an approach is needed, since numerical tests are time consuming and yet inconclusive, the possible correlations being too many for a complete check. Monte Carlo calculations need efficient algorithms that *a priori* are reliable, but a compromise depending upon application and available machinery is often unavoidable. A combination of systematic and heuristic arguments may help to find reasonable solutions.

II. ENSEMBLES AND CORRELATION COEFFICIENTS

Consider a sequence $r_1, r_2, \dots, r_N$ of N integers modulo m. The elements of this object sequence are the components of a vector $\mathbf{r} \in \mathbf{Z}_m^N$. The end points of the m^N different vectors $\mathbf{r}$ form an N-dimensional cubic lattice, which becomes a probability space by assigning to each vertex $\mathbf{r}$ a probability $p(\mathbf{r})$. These probabilities obey

$$p(\mathbf{r}) \geq 0, \quad \sum_{\text{all } \mathbf{r}} p(\mathbf{r}) = 1 , \tag{1}$$

and are also called ensemble weights. They define a general ensemble of object sequences, which can be used to calculate ensemble averages of functions of $\mathbf{r}$. A natural measure for the amount of randomness present in the ensemble is the entropy

$$S(p(\mathbf{r})) \equiv - \sum_{\text{all } \mathbf{r}} p(\mathbf{r}) \log_2 p(\mathbf{r}) . \tag{2}$$

The number of degrees of freedom that an ensemble allows is defined by

$$n \equiv S(p(\mathbf{r})) / \log_2 m , \tag{3}$$

which in practice is taken to be an integer.

To discuss the stochastic properties of the sequence $\mathbf{r}$ induced by the ensemble, it is convenient to use the auxiliary quantities

$$s_i \equiv \exp \left[\frac{2\pi i}{m} r_i \right] , \tag{4}$$

which are generalizations of the parities ±1 that served as alternatives to the bits 0 and 1 in earlier papers on binary sequences [18–20]. They are similar to the spin variables that occur in the q-state Potts model in statistical mechanics. If the ensemble is such that the r_i are mutually independent, the s_i are too, and vice versa. Being located on the unit circle, the s_i are automatically normalized, and the expected mean value, even of powers and products of s_i, is zero.

An arbitrary test vector $\mathbf{k}\in\mathbf{Z}_m^N$ is used to define the correlation product

$$P(\mathbf{k},\mathbf{r})\equiv\prod_{i=1}^{N} s_i^{k_i}=\exp\left[\frac{2\pi i}{m}\mathbf{k}\cdot\mathbf{r}\right] . \tag{5}$$

The components $k_1, k_2, \ldots, k_N$ of $\mathbf{k}$ are another sequence of N integers modulo m, the test sequence. The hypercubic lattice of $\mathbf{Z}_m^N$ serves a dual purpose, being used as a correlation space also. The number of different vectors $\mathbf{k}$ in this space is again m^N. In analogy to the spin correlation functions in statistical mechanics, the complex correlation coefficients $C(\mathbf{k})$ are defined as ensemble averages of the correlation products,

$$C(\mathbf{k})\equiv\sum_{\text{all }\mathbf{r}} P(\mathbf{k},\mathbf{r})p(\mathbf{r})=\sum_{\text{all }\mathbf{r}} e^{(2\pi i/m)\mathbf{k}\cdot\mathbf{r}}p(\mathbf{r}) , \tag{6}$$

and turn out to be discrete Fourier transforms of the ensemble weights, with values located on or in the unit circle: $|C(\mathbf{k})|\le 1$. The structure of values $C(\mathbf{k})$ in correlation space is the dual of the structure formed by the ensemble weights $p(\mathbf{r})$ in probability space. In crystallographic terms, the ensemble weights are scattering strengths on a hypercubic lattice, and the correlation coefficients are the resulting diffraction pattern on the dual or reciprocal lattice.

Equation (6) is the discrete version of the characteristic function of a multivariate distribution; upon expansion, $C(\mathbf{k})$ and $\ln C(\mathbf{k})$ are found to be the generating functions of the moments and cumulants of that distribution; see, for instance, Van Kampen [22]. All information hidden in moments and cumulants is contained in $C(\mathbf{k})$, which is a more practical quantity for the present purposes. The complex correlation coefficients $C(\mathbf{k})$, based on general products of arbitrary powers of s_i, differ from the usual correlation coefficients, based on normalized products of only two elements r_i (from which mean values are subtracted).

The value of $C(\mathbf{k})$ in the origin, belonging to the test vector $\mathbf{k}=\mathbf{0}$ that does not measure a true correlation, is given by

$$C(0)=\sum_{\text{all }\mathbf{r}} p(\mathbf{r})=1 . \tag{7}$$

Inversion of the Fourier transform gives

$$p(\mathbf{r})=m^{-N}\sum_{\text{all }\mathbf{k}} e^{-(2\pi i/m)\mathbf{k}\cdot\mathbf{r}}C(k) , \tag{8}$$

which for $\mathbf{r}=\mathbf{0}$ leads to the mean value, averaged over all $\mathbf{k}$:

$$\langle C(\mathbf{k})\rangle\equiv m^{-N}\sum_{\text{all }\mathbf{k}} C(\mathbf{k})=p(0) . \tag{9}$$

The transformation is norm conserving (Parseval's theorem), the second moment being given by

$$\langle |C(\mathbf{k})|^2\rangle\equiv m^{-N}\sum_{\text{all }\mathbf{k}} |C(\mathbf{k})|^2=\sum_{\text{all }\mathbf{r}} [p(\mathbf{r})]^2 . \tag{10}$$

Equations (9) and (10) are conservation laws that characterize the ensemble. Equation (9) is a weak law: usually, the $C(\mathbf{k})$ cancel one another and $p(0)$ is zero in ensembles for random sequences. Equation (10) is a stronger law, giving

$$\Gamma\equiv m^N\langle |C(\mathbf{k})|^2\rangle=m^N\sum_{\text{all }\mathbf{r}} [p(\mathbf{r})]^2 \tag{11}$$

as measure for the total amount of correlation.

III. GAMBLING ENSEMBLE AND SINGULAR ENSEMBLE

Consider the extreme case of the gambling ensemble, in which all sequences $\mathbf{r}$ are equally probable. It is defined by

$$p(\mathbf{r})=m^{-N} \tag{12}$$

for all $\mathbf{r}$, which factorizes into a product of probabilities $p(r_i)=m^{-1}$ since the elements r_i are independent. The gambling ensemble contains no information, and Eqs. (2) and (3) give the maximal values $S=N\ln_2 m$ (complete randomness) and $n=N$ (independence of elements) for the entropy and the number of degrees of freedom. Insertion into Eq. (6) leads to

$$C(\mathbf{k})=\delta_{\mathbf{k},0} . \tag{13}$$

While the ensemble weights are spread uniformly over probability space, all correlation is concentrated in the origin of correlation space. The quantities of Eqs. (9) and (10) are both equal to m^{-N}, and the total amount of correlation of Eq. (11) attains the minimal value $\Gamma=1$; only the term with $\mathbf{k}=\mathbf{0}$, which does not measure a true correlation, contributes. The property that all true correlation coefficients vanish is unique for the gambling ensemble: Eq. (12) follows from Eqs. (7) and (13). When randomness, in what will be called here the gambling definition, is identified with the absence of correlations, the use of the gambling ensemble is obligatory.

The opposite case of the gambling ensemble is the singular ensemble defined by

$$p(\mathbf{r})=\delta_{\mathbf{r},\mathbf{r}'} . \tag{14}$$

The Kronecker δ indicates that only the single sequence $\mathbf{r}'$ is present in the ensemble. Complete information is available, and the distribution in probability space is concentrated in one point. Insertion into Eqs. (3) and (4) gives minimal values $S=0$ and $n=0$ for the entropy and for the number of degrees of freedom: there is no randomness, and there are no arbitrary elements. Equation (6) leads immediately to

$$C(\mathbf{k})=e^{(2\pi i/m)\mathbf{k}\cdot\mathbf{r}'},\quad |C(\mathbf{k})|=1 . \tag{15}$$

For a given object sequence $\mathbf{r}'$ all test sequences $\mathbf{k}$ measure a complete correlation: in a fixed sequence, everything is correlated. Equation (9) reads

$$\langle C(\mathbf{k})\rangle \equiv m^{-N}\sum_{\text{all }\mathbf{k}} C(\mathbf{k})=\delta_{\mathbf{r}',0}\ , \tag{16}$$

since $\mathbf{r}'$ contributes only when all elements are zero. Equation (10) is easily verified for the singular ensemble. The total amount of correlation has the maximum value $\Gamma=m^N$.

The results obtained so far are both trivial and satisfactory. When randomness is identified with maximal entropy, at least some conceptual problems are solved. However, neither of the two ensembles offers a good description of the stochastic properties of the pseudorandom sequences that are used in practice.

IV. SCANNING ENSEMBLE

For that purpose the scanning ensemble is used, in which one object sequence $\mathbf{r}$ and its $N-1$ translated versions $T^j\mathbf{r}$ all have the same weight

$$p(T^j\mathbf{r})=N^{-1},\quad j=0,\ldots,N-1\ , \tag{17}$$

whereas all other sequences have weight zero. The translation matrix T (with 1's just above the diagonal and in the lower left-hand corner, and 0's everywhere else) shifts the elements of $\mathbf{r}$ one position to the left. In probability space, T is a rotation around the main body diagonal of the hypercube; T^N is the unit matrix. The scanning ensemble amounts to averaging over the sequence $\mathbf{r}$, which is taken to be periodic:

$$r_{i+N}=r_i \quad \text{for all } i\ . \tag{18}$$

The period N may be due to an intrinsic rule obeyed by the sequence, or the cyclic condition is just added to avoid boundary problems. To ensure that all sequences in the ensemble differ, it is assumed that $\mathbf{r}$ does not contain internal periodicities with a factor of N as period; otherwise, that factor should replace N (which amounts to having fewer sequences in the ensemble, each with a larger weight). Overcorrelated sequences with internal periods that are not a factor of N may be dismissed also.

Insertion of Eq. (17) into Eq. (2) gives $S=\log_2 N$ for the entropy, much less than the value $S=N\log_2 m$ in the gambling ensemble, but at least not zero as in the singular ensemble. Insertion into Eq. (6) gives

$$C(\mathbf{k})=\frac{1}{N}\sum_{j=0}^{N-1}\exp\left[\frac{2\pi i}{m}\mathbf{k}\cdot T^j\mathbf{r}\right]\ , \tag{19}$$

describing a circular autocorrelation of the sequence (in the sum over j one may replace $\mathbf{k}\cdot T^j\mathbf{r}$ by $T^{-j}\mathbf{k}\cdot\mathbf{r}$). Leaving differences in notation and derivation aside, one finds, after adding the limit $N\to\infty$, that $C(\mathbf{k})$ of Eq. (19) is identical with the Fourier transform $\phi(Q)$ used by Coveyou and MacPherson [21] to analyze linear-congruence sequences. The limit $N\to\infty$ tries to take immediate advantage of expected asymptotic properties, but is confusing. The finite and discrete form of Eq. (19) is preferred here.

Equations (9)–(11) give the following results for the scanning ensemble, valid for all sequences of length N without internal periodicities:

$$\langle C(\mathbf{k})\rangle=0,\quad \langle |C(\mathbf{k})|^2\rangle=\frac{1}{N},\quad \Gamma=\frac{m^N}{N}\ . \tag{20}$$

The mean over all $\mathbf{k}$ of $C(\mathbf{k})$ is zero, and its second moment is small, but the total amount of correlation Γ is huge, due to the large number m^N of different test sequences $\mathbf{k}$. This makes numerical tests inconclusive. By minimizing the correlations that a test happens to be sensitive for, one may even select sequences in which more harmful correlations are larger.

The number of degrees of freedom that agrees with $S=\log_2 N$ is

$$n=\log_2 N/\log_2 m\ . \tag{21}$$

A first step to avoid harmful correlations is to select equidistributed sequences, defined by requiring that the degrees of freedom are used to fit just all possible strings of n integers mod m into the period $N=m^n$ (overlaps allowed). Here this is called the equidistributing definition of a random sequence. Equidistributed sequences, also known as De Bruijn sequences, were called pseudorandom in earlier papers [18,19].

The scanning ensemble for a De Bruijn sequence of N integers mimics the gambling ensemble for strings of n integers: all true correlation coefficients $C(\mathbf{k})$ vanish for test sequences in which nonzero elements are at most n positions apart. Hence De Bruijn sequences are suitable candidates for random-number generation; many recipes for random numbers are based on them (or good imitations). However, when all correlations are considered, the permutation over the sequence of strings of size n becomes important: neighboring (nonoverlapping) strings should be as different as can be.

Table I is a summary of the properties of the different ensembles.

V. NORMAL BEHAVIOR

Henceforth only the scanning ensemble will be used. Consider first the case that the following relation holds, for a given combination of $\mathbf{r}$ and $\mathbf{k}$ but independent of j:

$$\mathbf{k}\cdot T^j\mathbf{r}=\mathbf{k}\cdot\mathbf{r}\ \mathrm{mod}m\equiv\gamma\ . \tag{22}$$

For this special case, Eq. (19) gives

$$C(\mathbf{k})=e^{(2\pi i/m)\gamma},\quad |C(\mathbf{k})|=1\ , \tag{23}$$

meaning that $\mathbf{k}$ indicates a complete correlation in $\mathbf{r}$. In general, $\mathbf{k}\cdot T^j\mathbf{r}$ varies with j, and contributions to $C(\mathbf{k})$ due to different j do appear at various places on the unit circle, with $|C(\mathbf{k})|<1$ as a result; the larger N is, the closer to the origin $C(\mathbf{k})$ tends to be. For arbitrary object sequences the distribution of $C(\mathbf{k})$ around the origin, which has to obey the mean and the second moment given by Eqs. (20), is expected to be a two-dimensional Gaussian.

Consider the object sequence formed by N decimals of π, to which the periodicity of Eq. (18) is added. For test sequences $\mathbf{k}=(1,0,\ldots,0)$ and $T^{-j}\mathbf{k}$, each digit contributes its own root of unity to $C(\mathbf{k})$; a value $C(\mathbf{k})\approx 0$ is

TABLE I. Survey of ensemble properties. The values m^{-N} in the gambling ensemble for $\langle C(\mathbf{k})\rangle$ and $\langle |C(\mathbf{k})|^2\rangle$ are due to $\mathbf{k}=0$ only. In the singular ensemble, $|C(\mathbf{k})|=1$ holds for all $\mathbf{k}$. The total amount of correlation Γ in the scanning ensemble is large for all sequences.

Ensemble	General (all seq.)	Gambling (all seq.)	Singular ($\mathbf{r}'$ only)	Scanning ($\mathbf{r}$+transl.)
Number of sequences	m^N	m^N	1	N
Weight	$p(\mathbf{r})$	m^{-N}	$\delta_{\mathbf{r},\mathbf{r}'}$	N^{-1}
Entropy S	$-\sum p(\mathbf{r})\log_2 p(\mathbf{r})$	$N\log_2 m$	0	$\log_2 N$
Degrees of freedom n	$S/\log_2 m$	N	0	$\log_2 N/\log_2 m$
$\langle C(\mathbf{k})\rangle$	$p(0)$	m^{-N}	$\delta_{\mathbf{r}',0}$	0
$\langle |C(\mathbf{k})|^2\rangle$	$\sum [p(\mathbf{r})]^2$	m^{-N}	1	N^{-1}
Total amount of correl. Γ	$m^N\langle |C(\mathbf{k})|^2\rangle$	1	m^N	m^N/N

equivalent to an average value of about 4.5 for the digits. Less simple test sequences give similar results. The normal behavior of the digits of π discussed, for instance, by Pathria [23], Wagon [24], and Johnson and Leeming [25] corresponds to a two-dimensional Gaussian distribution in the unit circle for $C(\mathbf{k})$. The Gaussian shape that obeys Eqs. (20) was verified numerically by Heringa [26], who calculated $C(\mathbf{k})$ for the first $N=5$, 10, and 20 digits of π (using representative samples of $\mathbf{k}$). Although verification is more difficult for larger N, there is no room for doubt that the Gaussian is the correct asymptotic description.

This behavior of $C(\mathbf{k})$ corresponds closely to intuitive notions of randomness, and will be called the normal regime. The sequences and the definition of randomness implied are called normal too. As a version of the weak law of large numbers, the normal regime is expected to hold for the digits of most transcendental or irrational numbers, and even to some extent for the digits of most rational numbers (for large N, but not in an asymptotic sense, because in this case N has to be small in comparison with the periodicity of the decimals).

A similar situation was described by Kolmogorov [3] and Chaitin [4] in terms of the complexity of a sequence, which is the number of bits in the shortest algorithm generating the sequence. A sequence is then taken to be random when its complexity equals its length in bits, the shortest algorithm being the sequence itself (adding an asymptotically irrelevant copying instruction). This will be called the complexity definition of randomness. Asymptotically, almost all sequences are complexity random as well as normal, although most of them will never be identified.

Sequences of which the elements can be computed are not complexity random, and sequences produced by linear algorithms are not even normal (see Sec. VII). However, the combination of range and order as a measure for the amount of complication of a test sequence, used below to control nonzero correlation coefficients, is a modified form of complexity.

VI. RANGE AND ORDER OF CORRELATIONS

Monte Carlo results are misleading when correlations hidden in the random numbers and in the simulated system interfere constructively. The usual advice is to test random-number generators in the context of their application. When the nonzero correlation coefficients of the object sequence from which the random numbers are taken belong to very long or complicated test sequences, interferences are *a priori* unlikely.

To specify these test sequences, two parameters are needed, one for the length and one for the order of a test sequence. A length parameter is provided by the range $\Lambda(\mathbf{k})$ or the size $\lambda(\mathbf{k})$, defined by

$$\Lambda(\mathbf{k})\equiv\lambda(\mathbf{k})\log_2 m,\quad \lambda(\mathbf{k})\equiv j-i+1\geq 0\ , \tag{24}$$

where $\Lambda(\mathbf{k})$ serves to compare cases of different m, and where j and i are the indices of k_i and k_j, the first and last nonzero elements of $\mathbf{k}$. For the trivial case $\mathbf{k}=\mathbf{0}$, one has $\lambda(0)=0$. Test sequences with $k_1\neq 0$ are called basic; all nontrivial test sequences can be reduced to basic form by translations that leave $C(\mathbf{k})$ invariant. When only test sequences of maximum size λ are used, λ acts as a window through which the object sequence is seen; it may then be convenient to project the structures in $\mathbf{Z}_m^N$ formed by ensemble weights and correlation coefficients onto $\mathbf{Z}_m^\lambda$.

In the case of binary sequences [18–20], the order parameter chosen was the number of bits equal to 1 in a test sequence. An obvious choice for general m is to use the number of factors in the product of Eq. (5). This is the case $d=1$ in

$$q_d(\mathbf{k})\equiv\left[\sum_{i=1}^{N}k_i^d\right]^{1/d}\ , \tag{25}$$

where $d=2$ is another choice. To avoid that elements larger than $\frac{1}{2}m$ do contribute more than their modular complements, k_i could be replaced by the Brillouin-zone variable $k_i'\equiv\min\{k_i,m-k_i\}$. For $m=2$, all choices agree with the earlier definition. The so-called city-block distance q_1 is an integer order parameter, and the Euclidean measure q_2 is identical to the quantity $|Q|$ used by Coveyou and MacPherson [21] in the spectral test and by Ora and Jerry Percus [27] in an overall correlation measure.

To enable comparison between sequences with different moduli the order parameter chosen (in a certain analogy with complexity theory) is

$$q(\mathbf{k})\equiv\sum_{i=1}^{N} H(k_i)\ , \tag{26}$$

where $H(k_i)$ is the Hamming weight of k_i (the number of 1's in its binary form), a quantity taken from code theory. For $m=2$ also Eq. (26) agrees with the earlier definition. Again, the Brillouin-zone variable $k_1'\equiv\min\{k_i,m-k_i\}$ instead of k_i could lead to a more even-handed treatment of complementary values, but this would be less simple formally. The values of $q(\mathbf{k})$ obey $0\le q(\mathbf{k})\le\Lambda(\mathbf{k})$, and are concentrated around $\frac{1}{2}\Lambda(\mathbf{k})$. The number $f(\Lambda,q)$ of test sequences of range Λ or smaller and of order q is just the binomial coefficient:

$$f(\Lambda,q)=\begin{bmatrix}\Lambda\\ q\end{bmatrix}\ . \tag{27}$$

When joined together, the binary expansions of the elements k_i form the bit pattern of $\mathbf{k}$, which has a simple structure when $q(\mathbf{k})$ is small or close to $\Lambda(\mathbf{k})$, the complement of a small value. The combination of range and order can be used as a measure for the degree of complication of $\mathbf{k}$: when $q(\mathbf{k})$ is large and close to the optimal value $\frac{1}{2}\Lambda(\mathbf{k})$, the bit pattern of $\mathbf{k}$ is usually complicated. It is an incomplete measure, which does not exclude simple patterns like alternating bits 0 and 1, but it is a more operational quantity than the complexity of $\mathbf{k}$.

Nonzero correlation coefficients that belong to complicated test sequences, of long range and of optimal Hamming order, are assumed to be least harmful. A similar specification, but in terms of q_2 instead of q, was given by Coveyou and MacPherson [21].

VII. LINEAR PRODUCTION RULES

The attention is now limited to sequences produced by the efficient linear production rules used in many Monte Carlo calculations. Consider an object sequence $\mathbf{r}$ determined by the linear algorithm

$$r_{j+n}=\sum_{i=1}^{n} K_i r_{i+j-1}+c \bmod m\ , \tag{28}$$

where the constant c and the coefficients K_i (with at least $K_1\neq 0$) are integers mod$m$. From an initial condition or state in the form of a seed of n integers $r_1,\dots,r_n$ all other elements of the sequence are found by iteration, starting at $j=1$. A set of n elements $r_j,\dots,r_{j+n-1}$ is called a temporary state of the generator, which together with the state starting at $j+n$ forms a pair of successive states. When the period N is reached, the initial state is recovered.

For a modulus like $m=2^{32}$ the elements r_i can be used directly as random numbers in Monte Carlo simulations. For smaller moduli a number of elements (for instance, 32 for $m=2$) must be combined. When needed for such a combination or to compare sequences with different m, all elements are thought to be replaced by their binary representation. Problems that arise when m is not a power of 2 are ignored, as well as *ad hoc* attempts to enhance randomness by skipping or reshuffling parts of a sequence, by using carry bits, or by other operations that distract from a study of the direct consequences of Eq. (28).

Many usual recipes for random-number generation are special cases of Eq. (28). For $n=1$ the linear-congruence rule results. The case $c=0$ and $n>1$ with all K_i equal to zero except $K_1=K_{i'}=1$ (with $1<i'\le n$) is known as the lagged-Fibonacci method; if more coefficients K_i are 1 instead of 0, the Fibonacci rule with multiple lags results. The lagged-Fibonacci method with $m=2$ is the feedback-shiftregister rule for binary sequences, with two or more feedback positions.

In the lagged-Fibonacci method one may also consider the use of multiplication of two or more different elements of the sequence (instead of addition) in order to find additional elements. Although there is some empirical evidence of the proper behavior of the resulting sequences, one should be aware of the possible occurrence of regions of attraction in these nonlinear schemes (the amount of nonlinearity is of course still smaller than in the calculation of the decimals of π). Since they are inevitably slower than the linear recipes included in Eq. (28) and more difficult to analyze, they are not taken into account here.

For many different values of m and n, combinations of values of K_i and c are known (or can be found by standard techniques) for which the period of the resulting sequence $\mathbf{r}$ is close to or equal to the maximum period $N=m^n$. Then $\mathbf{r}$ is almost or exactly a De Bruijn sequence, and the length n of the seed is equal to the number of degrees of freedom. For instance, maximum-length sequences generated by shiftregisters with a primitive trinomial as a characteristic function have periods 2^n-1 (only a single 0 is missing); the difference from a true De Bruijn sequence is negligible. Usually, the sequences considered below are assumed to be of maximum length, at least approximately.

A De Bruijn sequence with n degrees of freedom, for which all $C(\mathbf{k})$ of size $\lambda(\mathbf{k})\le n$ vanish, must have test sequences with $C(\mathbf{k})\neq 0$ for all sizes larger than n, otherwise the period would be larger than m^n. For linear De Bruijn sequences, which are generated by a linear production rule, a stronger result holds. Rewrite Eq. (28) as

$$\mathbf{K}\cdot T^j\mathbf{r}\equiv\sum_{i=1}^{n} K_i r_{i+j-1}+(m-1)r_{j+n}\bmod m=m-c\ , \tag{29}$$

which is valid for any j, where $\mathbf{K}$ is the basic test sequence of range $\Lambda(\mathbf{K})=(n+1)\log_2 m$ with $|C(\mathbf{K})|=1$ given by

$$\mathbf{K}\equiv(K_1,\dots,K_n,m-1,0,\dots,0)\ . \tag{30}$$

Equation (29) is identical to Eq. (22) for $\gamma=m-c$ and

$\mathbf{k}=\mathbf{K}$. The constant c is often needed for maximum length, but its contribution to the correlation product is only a phase factor. Defining the first complete correlation, the test sequence $\mathbf{K}$ is equivalent to the production rule, which will also be indicated by $\mathbf{K}$. The range and order of $\mathbf{K}$ will always be denoted by $\Lambda(\mathbf{K})$ and $q(\mathbf{K})$ explicitly; they obey the relation

$$2 \leq q(\mathbf{K}) \leq \Lambda(\mathbf{K}) = (n+1)\log_2 m \ . \tag{31}$$

The period of a maximum-length sequence generated by a production rule $\mathbf{K}$ of range $\Lambda(\mathbf{K})$ is given by $N=m^n = 2^{\Lambda(\mathbf{K})}/m$.

The production rule generates all other correlations. Take a linear combination of $\mathbf{K}$ and its translated versions,

$$\mathbf{K}' \equiv \sum_{j=1}^{\lambda-n} b_j T^{1-j} \mathbf{K} \bmod m \ , \tag{32}$$

where T^{-1} is a shift to the right over one position, and where λ obeys $n+1 \leq \lambda \leq N$. The coefficients b_j are integers mod$m$. If at least one of them is not zero, $\mathbf{K}'$ can always be reduced to an equivalent basic test sequence of size $\lambda(\mathbf{K}') \geq n+1$, obeying Eq. (22) and indicating a complete correlation $|C(\mathbf{K}')|=1$. For $\lambda=n+1$ there are $m-1$ different choices for $b_1 \neq 0$, and for $\lambda=n+2$ there are $(m-1)^2$ nonzero choices for b_1 and b_2. Each time λ increases further, the number of sequences $\mathbf{K}'$ is multiplied by m. The number $\Delta(\lambda)$ of basic test sequences $\mathbf{K}'$ of size $\lambda=\lambda(\mathbf{K}')$ is

$$\Delta(\lambda) = \begin{cases} 0 \ \text{ for } 0 < \lambda \leq n \\ m-1 \ \text{ for } \lambda = n+1 \\ (m-1)^2 m^{\lambda-n-2} \ \text{ for } \lambda \geq n+2 \ . \end{cases} \tag{33}$$

Consider all completely correlated test sequences of size λ or smaller, with $\lambda \geq n$. Since a basic test sequence of size $\lambda' \leq \lambda$ corresponds to $\lambda-\lambda'+1$ test sequences of maximum size λ when its nonbasic translated versions are included, the total number $D(\lambda)$ of complete correlations of maximum size λ generated by the production rule is

$$\begin{aligned} D(\lambda) &\equiv \sum_{\lambda'=0}^{\lambda} \Delta(\lambda') \\ &= 1 + (\lambda-n)(m-1) \\ &\quad + \sum_{\lambda'=n+2}^{\lambda} (\lambda-\lambda'+1)(m-1)^2 m^{\lambda'-n-2} = m^{\lambda-n} \ , \end{aligned} \tag{34}$$

where $\Delta(0)=1$ is due to the trivial linear combination $\mathbf{K}'=\mathbf{0}$.

For $\lambda=N$ the number of complete correlations is equal to the total amount of correlation $\Gamma=m^N/N=m^{N-n}$ of Eqs. (20). Hence Eq. (34) counts all completely correlated test sequences $\mathbf{K}'$ with $|C(\mathbf{K}')|=1$, called the correlated test sequences for short; together, they form a complete set of conserved quantities for production rule $\mathbf{K}$. All other $m^N - m^{N-n}$ test sequences, the uncorrelated ones, give $C(\mathbf{k})=0$. For linear De Bruijn sequences, m^{N-n} correlation coefficients are located on the unit circle, while the remaining ones are located in the origin. Equations (20) are obeyed, but the contrast with normal behavior could not be greater.

VIII. WELL-TEMPERED SEQUENCES

Linear sequences may still be reliable sources of random numbers. Consider the cumulative distribution $d(\Lambda,q)$, defined as the number of correlated test sequences of range Λ or smaller and of order $q \leq \Lambda$ (the more detailed distribution for each range Λ separately does not change the argument, but leads to less simple formulas). When the uncorrelated test sequences are included, the distribution $f(\Lambda,q)$ of Eq. (27) results. For $\lambda \leq n$ one has $d(\Lambda,q)=0$, apart from $d(\Lambda,0)=1$ due to $C(0)=1$. The first $m-1$ nontrivial contributions to $d(\Lambda,q)$ occur at $\lambda=n+1$, due to the $m-1$ choices for $b_1 \neq 0$ in Eq. (32). For $\lambda \geq n+1$ the total number of (complete) correlations of range Λ or smaller is

$$\sum_{q=0}^{\Lambda} d(\Lambda,q) = D(\lambda) = m^{\lambda-n} = \frac{2^\Lambda}{N} = \frac{1}{N} \sum_{q=0}^{\Lambda} f(\Lambda,q) \ , \tag{35}$$

in varying disguises. It is equal to m for $\Lambda=\Lambda(K) = (n+1)\log_2 m$ and to m^N/N when at $\Lambda = N \log_2 m$ the full period is covered.

The order $q=q(\mathbf{K}')$ is determined by interference between the bits of K_i (including $K_{n+1}=m-1$) and b_j in Eqs. (30) and (32). When $\lambda-n$ increases, the production rule continues to feed additional correlations into $d(\Lambda,q)$, and q will increasingly behave as a stochastic variable. It is convenient to split $d(\Lambda,q)$ into a stochastic and a deterministic part,

$$d(\Lambda,q) \equiv d_s(\Lambda,q) + d_d(\Lambda,q) \ . \tag{36}$$

The stochastic part, defined by

$$d_s(\Lambda,q) \equiv m^{-n} f(\Lambda,q) = \frac{1}{N} \left[\begin{matrix} \Lambda \\ q \end{matrix}\right] \ , \tag{37}$$

obeys Eq. (35) all by itself and is identical for all De Bruijn sequences of the same period. For large Λ and $q \approx \frac{1}{2}\Lambda$, it is approximately equal to

$$d_s(\Lambda,q) \approx \frac{2^\Lambda}{N\sigma\sqrt{2\pi}} e^{-(q-\bar{q})^2/2\sigma^2} \ , \tag{38}$$

where the mean and the variance of the binomial distribution as well as its Gaussian approximation are given by $\bar{q}=\frac{1}{2}\Lambda$ and $\sigma^2 \equiv \overline{q^2} - \bar{q}^2 = \frac{1}{4}\Lambda$.

In general, $d(\Lambda,q)$ itself is already expected to be almost similar to $f(\Lambda,q)$, because every correlated test sequence is surrounded by many uncorrelated ones of almost the same order; the reverse is not true. The deterministic part $d_d(\Lambda,q)$ defined by Eqs. (36) and (37) should be relatively small; its sum over q is zero. The main problem is that the region in q occupied by $d_d(\Lambda,q)$ may extend far beyond that of $d_s(\Lambda,q)$. When $q(\mathbf{K})$ is small and when the coefficients $b_j \neq 0$ chosen in Eq. (32) are few and of small Hamming weights, the resulting correlations occur at values of q where $d_s(\Lambda,q)$ is zero [at least initially, when $\Lambda \geq (n+1)$, $\log_2 m$ is still small] and must be accommodated in $d_d(\Lambda,q)$. Complementary circumstances when $q(\mathbf{K})$ is close to $\Lambda(\mathbf{K})$ have similar effects (formally, this could be taken into account by using Brillouin-zone variables).

In an earlier paper [28] an exact result for binary maximum-length sequences ($m=2$) was obtained for the case in which Λ is equal to the full period ($\Lambda=\lambda=N-1\equiv 2^n-1$). It reads

$$d(N-1,q) = \frac{1}{N}\left[\begin{matrix} N-1 \\ q \end{matrix}\right] + \frac{N-1}{N}(-1)^q(-1)^{q'/2}\left[\begin{matrix} N/2 \\ q'/2 \end{matrix}\right] , \quad (39)$$

with $q'=q$ for q even and $q'=q-1$ otherwise. For large N the second term is but a small deterministic ripple on the first one, which agrees with Eq. (37). However, $\lambda=N-1$ is far beyond the relevant region $n<\lambda \ll N$.

The precise shape of $d(\Lambda,q)$ depends on $\mathbf{K}$ and can only be determined numerically. Figure 1 shows the result for the very short binary sequence of maximum length generated by $\mathbf{K}=(1,1,0,0,1)$, with $n=4$, $q(\mathbf{K})=3$, and $\Lambda(\mathbf{K})=5$. The deviations between $d(\Lambda,q)$ and Eq. (37) are small. Extension to less trivial production rules is easy, but the steep increase with λ of the number $m^{\lambda-n}$ of completely correlated test sequences outruns any computer capacity long before rules used in practice come into sight. Exhaustive numerical checks of random-number generators are impossible.

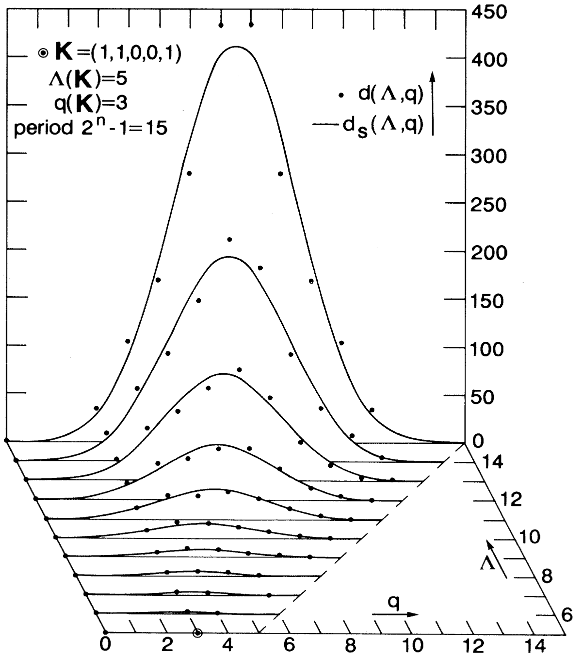

FIG. 1. The number $d(\Lambda,q)$ of correlations of range Λ and order q and its stochastic part $d_s(\Lambda,q)=(1/N)\binom{\Lambda}{q}$, for a binary maximum-length sequence of period $N=15$ generated by $\mathbf{K}=(1,1,0,0,1)$. When the production rule is of high order, $\Lambda(\mathbf{K})\gg 100$, the stochastic regime shifts to irrelevantly large values of Λ and q.

These results confirm that $d_d(\Lambda,q)$ is small compared with $d_s(\Lambda,q)$. In general, the distribution for the order of complete correlations in a linear De Bruijn sequence is expected to obey Eq. (37), apart from small deviations. For $\Lambda>\Lambda(\mathbf{K})$, first consider $d_s(\Lambda,q)$ separately, in which most correlations are located in a narrow region around the mean $\bar{q}=\frac{1}{2}\Lambda$. While Λ increases, the mean shifts to higher orders, but the total number of correlations $2^\Lambda/N$ grows so fast that the low-order tail of the distribution extends to even lower orders. As long as Λ remains smaller than a certain value Λ', all contributions to $d_s(\Lambda,q)$ are at least of order q'; when Λ' becomes smaller. When $\Lambda(\mathbf{K})$ is large, both q' and Λ' can be large enough to assume that any interference with correlations present in an application is excluded. See Sec. IX.

Most problems in linear algorithms are due to $d_d(\Lambda,q)$. When $q(\mathbf{K})$ is small, the tail of $d(\Lambda,q)$ at low order lies above $d_s(\Lambda,q)$, especially at small values of $\lambda-n$ and q where an adverse effect on the reliability of Monte Carlo calculations is largest. Similar effects arise when $q(\mathbf{K})$ is close to $\Lambda(\mathbf{K})$, due to excessive cancellations of bits 1 in Eq. (32). To minimize these effects, $\mathbf{K}$ should have a complicated bit pattern, of long range and of optimal order $q(\mathbf{K})\approx\frac{1}{2}\Lambda(\mathbf{K})$; then $\mathbf{K}$ feeds $d(\Lambda,q)$ at values of q near the maximum. When $\mathbf{K}$ has many elements of intermediate Hamming weight, bit mixing starts in Eq. (30) instead of having to wait for the coefficients b_j in Eq. (32), and a small number of coefficients $b_j\neq 0$ in Eq. (32) will not suffice to produce a correlated test sequence $\mathbf{K}'$ of low order $q(\mathbf{K}')$. Hence all correlations that occur outside the region where the stochastic part is active are then avoided.

Tentatively, a linear maximum-length sequence is a reliable source of random numbers when its production rule $\mathbf{K}$ obeys three conditions:

A. the range $\Lambda(\mathbf{K})$ is sufficiently large,

B. the order $q(\mathbf{K})$ is comparable to $\frac{1}{2}\Lambda(\mathbf{K})$,

C. the bit pattern of $\mathbf{K}$ is irregular and diffuse.

For condition A, the pragmatic estimate $\Lambda(\mathbf{K})\gg 100$ will be found below. Condition C implies that all parts of $\mathbf{K}$ contribute more or less equally to $q(\mathbf{K})$. Together, the conditions form the well-tempered definition of randomness. The well-tempered sequences that obey them should contain low-order correlations only at ranges $\Lambda>\Lambda'\gg\Lambda(K)$.

Table II summarizes the different definitions of randomness and the resulting properties of $C(\mathbf{k})$. Since condition C amounts to requiring the bit pattern of $\mathbf{K}$ to be random, the well-tempered definition is to some extent circular. A certain circularity is inherent to the subject, and can be traced back, as was argued in an earlier paper [18], to the seminal work of Von Mises [1]. Some of the circularity is removed by conditions A and B. Moreover, condition C refers to the production rule and is less stringent and easier to check than a condition of randomness applied to the sequence itself.

TABLE II. Definitions of randomness and corresponding properties of $C(\mathbf{k})$. Normal and complexity-random sequences are similar, but not equivalent. Well-tempered sequences are a special case of equidistributed sequences, for which all correlations are expected to be located in the stochastic region, at large values for range and order.

Definitions	Correlation coefficients
Gambling	No true correlations: $C(\mathbf{k})=0$ for $\mathbf{k}\neq 0$
Equidistributing	$C(\mathbf{k})=0$ for $0<\Lambda(\mathbf{k})\leq n\log_2 m$
Normal	Gaussian distribution of $C(\mathbf{k})$ in unit circle
Complexity	Unspecified (but often normal)
Well-tempered	$C(\mathbf{k})=0$ for $0<\Lambda(\mathbf{k})<\Lambda'$ and $0<q(\mathbf{k})<q'$

IX. BIT-MIXING PROCESS

The conditions are neither sharp nor independent, and a mixture of bits and arguments is needed to show their practical meaning. First consider condition A, the importance of which has of course been generally recognized; from an asymptotic point of view, all notions of randomness improve. In the present terminology, condition A together with maximum length is desirable because correlations of range $\Lambda<\Lambda(\mathbf{K})$ are absent, and because the number $2^{\Lambda}/N$ of complete correlations of maximum range Λ for $\Lambda>\Lambda(\mathbf{K})$ decreases when $N=2^{\Lambda(\mathbf{K})}/m$ increases, as Eq. (35) shows.

Both properties are, however, based on averaging correlation products over a full period. In practice, only relatively small subsequences are used; the larger $\Lambda(\mathbf{K})$ is, the more so. For subsequences of size $N'\ll N$, the relevant quantities are the subsequence coefficients, in which the correlation products of Eq. (5) are averaged over a part of the sequence only [in the scanning ensemble needed for the averaging, the periodicity of Eq. (18) is maintained, using N' instead of N as period]. The behavior of subsequence coefficients, even for $\Lambda<\Lambda(\mathbf{K})$, is somewhere between the normal regime of Sec. V and the opposite regime described at the end of Sec. VII, and depends on temporary states. Thus maximum-length sequences seem to lose their advantage, and condition A seems to be self-defeating.

Nevertheless, maximum length (apart from small deviations) as well as condition A remain recommendable. The decay of subsequence averages to their full-period values will not improve when the probability space from which successive temporary states are sampled is much smaller than allowed by the available degrees of freedom. The relative influence of quasistationary or imperfect resonant eigenstates of the production rule, states that are often characterized by containing many bits of some kinds of internal symmetries, is smaller when the probability space is larger. Whether a given generator is sampling this space effectively enough to ignore memory effects altogether, that problem is still to be discussed. First, condition A must be made more precise.

If the decay is fast enough, test sequences $\mathbf{K}'$ with the full-period value $|C(\mathbf{K}')|=1$ for $\Lambda(\mathbf{K}')\geq\Lambda(\mathbf{K})$ are the main cause of troubles. Most of these belong to $d_s(\Lambda,q)$, and can be made harmless by specifying what sufficiently large means in condition A. All correlations in $d_s(\Lambda,q)$ are of order $q(\mathbf{K}')\geq q'$, at least when $\Lambda(\mathbf{K}')$ is below the critical value Λ'. When q' is large enough, these stochastic correlations are irrelevant for most applications. The value q' is the order at which, for the given range $\Lambda'>\Lambda(\mathbf{K})$, the left tail of $d_s(\Lambda',q)$ in Eq. (37) is of magnitude 1; it is a crude estimate for the lowest order of the correlations present in the stochastic part. The resulting relation is

$$\frac{\Lambda'!}{q'!(\Lambda'-q')!}\approx m^n=N\ , \tag{40}$$

which leads to $\Lambda'\approx 2^6$, 2^9, or 2^{15}, respectively, for $N=2^{32}$, 2^{64}, or 2^{128} when the arbitrary value $q'=10$ is adopted; only the magnitude of the exponents matters. Since Λ' must be divided by the number of bits (say, 32) of one random number to see how many random numbers in a row can be assumed to be free of correlations of order below q', condition (a) boils down to $N\approx 2^{100}$ as a minimum, or $\Lambda(\mathbf{K})=\log_2(Nm)\gg 100$. For instance, linear-congruence sequences with $m=2^{32}$ appear to be unreliable sources of random numbers, at least for demanding tasks; as is easily verified, correlations of order $q'=10$ are likely already to be present in two successive elements. True, the lowest order of correlations tolerable in applications is a vague notion, and not all correlations are detrimental, but $q'=10$ is not an excessive value. For $\Lambda>\Lambda'$, more correlations of order q' and below will turn up soon. Since arguments based on the tail of a Gaussian are risky, condition A implies that extensive Monte Carlo calculations need values of $\Lambda(\mathbf{K})$ far beyond 100 to be reliable.

The remaining two questions, whether $d_d(\Lambda,q)$ is indeed harmless and whether the decay to full-period values of the correlation coefficients is fast enough, are directly related: low-order contributions to $d_d(\Lambda,q)$ that arrive early, at rather small values of $\Lambda-\Lambda(\mathbf{K})>0$, indicate that the bit-mixing process over the range Λ is too slow. Conditions B and C try to solve both problems by requiring $\mathbf{K}$ to be sufficiently complicated to avoid early low-order correlations; then successive temporary states are sufficiently independent to sample the large probability space that is effectively provided by condition A. Ideally, the balanced amounts of bits 0 and 1 required by condition B are scattered at random all over the bit pattern of $\mathbf{K}$.

Condition B is not very strict: if $q(\mathbf{K})$ is only rather close to the optimal value $\frac{1}{2}\Lambda(\mathbf{K})$, the back-to-normal process of bit mixing dictated by $\mathbf{K}$ will usually only lead to orders near the maximum of the stochastic regime. Interference effects that give rise to early correlations of low order, due to interference between different parts of the bit pattern of **K**, are not completely excluded, but they are unlikely. It is reasonable to suppose that the deterministic deviations are completely absorbed by $d_s(\Lambda,q)$ when $q(\mathbf{K})\approx\frac{1}{2}\Lambda(\mathbf{K})>100$ holds, together with condition C.

Obviously, condition C is not very strict either, in spite of being rather important. Sharp conditions do not exist; the quality of random numbers improves only gradually with that of the production rule, while the desirable quality depends upon applications and increases with time.

X. NUMERICAL TESTS

The above reasoning is very schematic. To check it, only numerical methods are available. Unfortunately, they require $\Lambda(\mathbf{K})$ to be so small that even condition A is not obeyed, while the whole argument depends on asymptotic considerations. A complete study of $d_d(\Lambda,q)$ is like looking for a needle in a haystack. Numerical checks of the stochastic behavior of the order of complete correlations are as unsatisfactory as for the normal behavior of the decimals of π.

This does not imply that they are not needed. Numerical studies of $d(\Lambda,q)$ for different production rules would be worthwhile, even when, at small values of $\Lambda(\mathbf{K})$, only a trend could become clear. The reduction of memory effects in temporary states due to conditions B and C is another interesting topic. For well-tempered sequences the decay of subsequence coefficients to full-period values (0 or 1) should be fast, independent of initial conditions. In particular, the decay from heavily biased or otherwise regular initial conditions would deserve attention.

The usual tests for random-number generators are designed to detect early correlations of low order in one way or another, often intuitively and without using an order parameter. At best, only indirect information about $d(\Lambda,q)$ is obtained. The spectral test for linear-congruence sequences is closest to the present approach and deserves special attention.

In this test, the projections onto $\mathbf{Z}_m^\lambda$ are studied for the structures in $\mathbf{Z}_m^N$ formed by $T^j\mathbf{r}$ (for all j) and by all linear combinations $\mathbf{K}'$ of **K**, where **r** is the sequence generated by **K**; see Eqs. (28)–(32) for details. The projections are dual hyper-rhombohedral sublattices of $\mathbf{Z}_m^\lambda$. The spectral test verifies whether these sublattices cover $\mathbf{Z}_m^\lambda$ uniformly; it requires that the shortest distance between nearest hyperplanes, as given by the inverse of the Euclidean order parameter q_2 of Eq. (25), is sufficiently large. Although stated in terms of q_2 rather than q, the spectral test is similar to the demand that the tail of $d(\Lambda,q)$ is zero for $q<q'$ and $\Lambda<\Lambda'$. Indeed, at the end of their paper, Coveyou and MacPherson [21] also asked attention for the binary code of the production rule.

In the past, the use of the spectral test has often been limited to small values of λ, but recently L'Ecuyer and Couture [29] developed efficient programming techniques to study very large values like $\lambda\approx30$. Perhaps these techniques can also be used to study a presumably Gaussian regime for q_2, or, after adaptations, to find the lowest values of q at which $d(\Lambda,q)$ differs from zero for a given value of $\Lambda>\Lambda(\mathbf{K})$.

Many other tests are found in the reviews cited above [11–17]; the information on correlations given by them is usually indirect. Recently, application specific tests were developed by Vattulainen, and co-workers [30–32]. The application involved is the Ising model of statistical mechanics, which, due to its known exact results and suitability for large-scale Monte Carlo simulations, was used as a testing ground in many other papers [33–37]. However, application tests move the unavoidable correlations only out of sight, to places where they are able to disturb the next application.

When random numbers are needed, the many recipes, tests, and rumors make it difficult to choose and to strike a balance between reliability and efficiency. Conditions A, B, and C offer a qualitative guideline. The examples in Sec. XI may indicate how they can be applied.

XI. COMPARISON OF LINEAR RULES

The familiar production rule for linear-congruence sequences,

$$r_{j+1}=K_1r_j+c \bmod m \ , \tag{41}$$

is equivalent to $\mathbf{K}=(K_1,K_2=m-1,0,\dots,0)$. The size and maximum-length period, respectively, are $\lambda(\mathbf{K})=n+1=2$ and $N=m$. A once popular choice, which together with similar ones has fallen into disrepute [15], is

$$m=2^{31}-1,\quad K_1=16\,807,\quad c=0\ , \tag{42}$$

with range $\Lambda(\mathbf{K})\approx62$ and order $q(\mathbf{K})\approx37$. Condition B is obeyed, though at the cost of the unbalanced bit pattern of **K** hiding behind the Hamming weights $H(K_1)=7$ and $H(K_2)=30$. Condition C is not obeyed. The value $H(K_{n+1})=H(m-1)$ is a handicap for rules with small n and with m close to a large power of 2; the Brillouin-zone variable K'_{n+1} would emphasize this effect. Condition A is also violated: $\Lambda(\mathbf{K})=62$ is not large enough to exclude early low-order correlations of the stochastic regime. These correlations are a plausible origin of the poor lattice structures that caused the decline of recipes based on moduli like $m=2^{31}$ or 2^{32}.

An example of a lagged-Fibonacci rule is the subtract-with-borrow generator suggested by Marsaglia, Narasimhan, and Zaman [38],

$$r_{j+43}=-r_j+r_{j+21}-c_j \bmod(2^{32}-5)\ . \tag{43}$$

This amounts to $\mathbf{K}=(K_1=m-1,\ K_{22}=1,\ K_{44}=m-1)$, skipping zeros. The carry bit $c_j=0$ or 1 and the -5 in m are needed to ensure the maximum length $m^n\approx2^{32\times43}$. One finds $\Lambda(\mathbf{K})\approx1400$. Condition A seems to be obeyed, but this value is too flattering. The carry bit and the -5 contribute little to bit mixing, and the flow in probability space spirals too slowly away from the one for the case

$m=2^{32}$ without carry bit, which has maximum length $m2^n=2^{75}$. For $q(\mathbf{K})$ one finds 61, not optimal but rather large. However, also $q(\mathbf{K})=61$ is too flattering: the bit pattern of $\mathbf{K}$ (two large blocks of 1's separated by many 0's and a single 1 in the middle) is too simple for an efficient bit mixing; the use of Brillouin-zone variables would emphasize this. Conditions B and C are not met. These criticisms agree with the poor lattice structure found by Couture and L'Ecuyer [39] for Eq. (43). Together with subtract-with-borrow recipes, the related add-with-carry recipes are also unreliable.

Marsaglia, Narasimhan, and Zaman [38] suggest combining Eq. (43) with $r_j'=r_{j-1}'-c' \bmod 2^{32}$, where c' is a constant, into $r_j''=r_j-r_j' \bmod 2^{32}$. It is easily verified, however, that most of the added complication ends up, together with c', in the phase factor γ of Eqs. (22) and (23), where it does not change the correlation properties. Couture and L'Ecuyer [39] showed that the lattice structure of r_j'' is similar to the one for r_j.

Shiftregister generators are based on the production rule

$$r_{j+n}=1+\sum_{i=2}^{n} K_i r_{i+j-1} \bmod 2 \ , \tag{44}$$

which is equivalent to $\mathbf{K}=(K_1=1,K_2,\ldots,K_n,K_{n+1}=1)$, with $K_i=0$ or 1 for $2\le i\le n$. The same production rule can also be defined by means of the characteristic polynomial

$$R(\{i\},n)\equiv 1+\sum_{j\in\{i\}} x^j+x^n \ , \tag{45}$$

where $\{i\}$ is the set of indices $0<i<n$ for which $K_{i+1}=1$ holds. When the polynomial is primitive, the sequence has maximum length. See Golomb [11] for details.

A simple example is the two-bit feedback shiftregister rule defined by the primitive trinomial

$$R(103,250)=1+x^{103}+x^{250} \ , \tag{46}$$

on which the recipe of Kirkpatrick and Stoll [40] is based. This recipe, in which each bit of the random numbers is taken from a widely different subsequence, proved to be unreliable in applications [34–37]. The range and order are $\Lambda(\mathbf{K})=251$ and $q(\mathbf{K})=3$, which means that condition A is obeyed while conditions B and C are strongly violated. In addition, the third-order correlations are present in each bit of the random numbers. Because of the many early warnings against shiftregister sequences based on trinomials, the recipe should never have been used.

Shiftregister sequences are still attractive, due to their binary character. Primitive polynomials of degree $n\gg 100$ and of optimal order would be reliable recipes, but they are difficult to find (and difficult to implement). The alternative suggested in earlier papers [18,19] is to use reducible polynomials that are a product of M primitive trinomials,

$$R_M(\{i_j,n_j\})\equiv\prod_{j=1}^{M} R(i_j,n_j) \ . \tag{47}$$

The elements of the resulting sequence are the mod-2 sum of the elements of the sequences produced by the individual trinomials. When the periods of the individual sequences have no factor in common, their product is the period of the resulting sequence; when all n_j are large enough, the difference from maximum length is negligible. A further justification is found in an earlier paper [20]. The combined Tausworthe generators that were suggested by Tezuka and L'Ecuyer [41] are equivalent to Eq. (47).

Preliminary results of Monte Carlo simulations of the Ising model obtained by Shchur [42] and a series of tests carried out by Berdnikov and Turtia [43] show that considerable improvements above the use of a single trinomial occur already for $M=2$, for which the order of the resulting production rule (the number of terms in R_M) is at most 9. For the most demanding Monte Carlo calculations this is not sufficient, but $M=8$, with $q(\mathbf{K})=3^8$, is probably excessive. The following examples for the intermediate cases $M=3$ and 4 are given for further orientation.

In an extensive search for reliable and efficient recipes of this kind the following example for $M=3$ was given by L'Ecuyer [44]:

$$R_3(3,28;2,29;13,31)=(1+x^3+x^{28})(1+x^2+x^{29})\times(1+x^{13}+x^{31}) \ . \tag{48}$$

The degree of the polynomial is $n=\sum_j n_j=88$, and the range is $\Lambda(\mathbf{K})=89$. Since mod-2 cancellations due to double products are absent, the order, or the number of terms in the polynomial, is $q(\mathbf{K})=27$; this is slightly below optimal, but the bit pattern of $\mathbf{K}$ is found to be rather irregular. However, the range $\Lambda(\mathbf{K})=89$ is too small to conclude that condition A is obeyed convincingly.

For the case $M=4$ in Eq. (47), L'Ecuyer [44] found many different rules, satisfying additional conditions. All these rules (which are not explicitly given) have the same range, $\Lambda(\mathbf{K})=114$, which seems not yet large enough to obey condition A without hesitation. Also, at this value for the range, most polynomials with $M=4$ will not contain the maximum number $3^M=81$ of terms possible, due to accidental cancellations; the available degrees of freedom are not sufficient. When combined with a somewhat larger range, however, the value $M=4$ could serve to set a minimum standard. As such, the following polynomial could be used:

$$R_4=(1+x^6+x^{17})(1+x^{13}+x^{31})(1+x^{38}+x^{89})\times(1+x^{15}+x^{127}) \ . \tag{49}$$

All the trinomials are primitive and have a Mersenne exponent as degree. The range is $\Lambda(\mathbf{K})=1+\sum_j n_j=265$ and, since cancellations are absent, the order is $q(\mathbf{K})=81$. This is below optimal, but at the given value of the range there is no reason to worry: the stochastic regime, centered around $q\approx 132$, is far enough away to avoid early low-order correlations, and it will attract the deterministic contributions very effectively, since the bit pattern of $\mathbf{K}$ is highly irregular (as is found when the polynomial R_4 is calculated explicitly). The flow in probability space will never be parallel to one of the orthogonal subspaces defined by the trinomials.

Equation (49) is far from unique. Other combinations (of at least similar range and order) may be more efficient. For instance, due to its small degree the first factor in Eq. (49) may cause problems in implementation, but it can be replaced by any primitive trinomial taken from the lists cited below. In general, the range and order of the production rule $\mathbf{K}$ defined by the polynomial R_M of Eq. (47) are given by

$$\Lambda(\mathbf{K})\approx 1+\sum_{j=1}^{M} n_j,\quad q(\mathbf{K})\approx\min(3^M,\tfrac{1}{2}\Lambda(\mathbf{K}))\ , \qquad (50)$$

where occasional cancellations in R_M may prevent the order from reaching the maximum value 3^M, especially when this is larger than $\frac{1}{2}\Lambda(\mathbf{K})$. A list of primitive trinomials with a Mersenne exponent as degree was given by Zierler [45]; recently, it was extended by Kurita and Matsumoto [46] and by Heringa, Blöte, and Compagner [47] up to Mersenne exponent 132 049. When M trinomials of different degree are taken from these lists, the periods are Mersenne primes and have no factors in common, and the $\approx 3^M$ terms in R_M are uniformly scattered over the available degrees of freedom. Primitive trinomials of which the degree is not a Mersenne exponent were listed by Zierler and Brillhart [48], up to degree 1000; these can also be used as factors in R_M, as long as the maximum-length periods involved do not share factors.

Figure 2 is a plot of order versus range for the different examples. The reader is invited to put his own preferred recipe on the map.

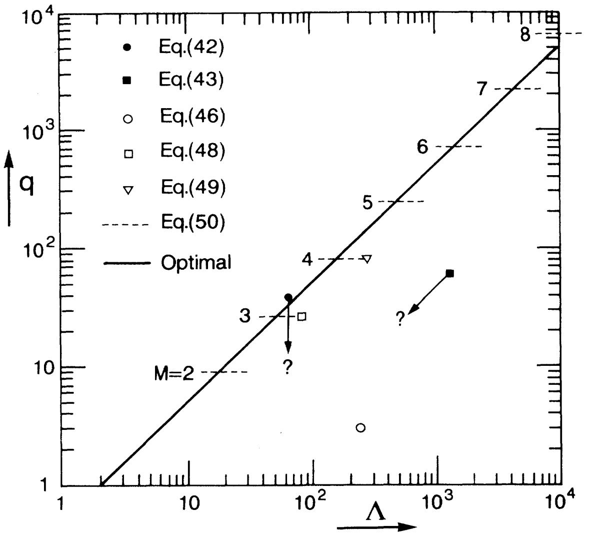

FIG. 2. Values of $q(\mathbf{K})$ for some typical production rules $\mathbf{K}$ are compared with the optimal value $q(\mathbf{K})\approx\frac{1}{2}\Lambda(\mathbf{K})$. Question marks refer to cases with misleading values of $\Lambda(\mathbf{K})$ or $q(\mathbf{K})$; see text. A lower bound for reliable random-number generation by means of maximum-length production rules $\mathbf{K}$ is $q(\mathbf{K})\approx\frac{1}{2}\Lambda(\mathbf{K})>100$. The open triangle indicates a minimum standard.

XII. CONCLUSIONS

Ensemble theory and general correlation coefficients are suitable instruments for discussing random-number generation. Any quest for a truly random sequence is bound to fail, since the total amount of correlation is conserved. Nonlinear recipes, leading to a normal behavior of the correlation coefficients in the unit circle, are closest to intuitive notions of randomness.

For the more efficient linear recipes all correlations are complete and consist of linear combinations of the production rule. The order of these correlations is approximately Gaussian distributed. A few general conditions for the production rule suffice to ensure that the sequence is well tempered: correlations of short range and low order are probably absent. In particular, the order of the production rule should obey the lower bound $q(\mathbf{K})\approx\frac{1}{2}\Lambda(\mathbf{K})>100$.

A promising alternative for the order parameter q, which is not a complete measure for the amount of complication and which also lacks the symmetry of a Brillouin-zone variable, is a local entropy, based on the frequencies with which different strings of a given number of elements occur in a subsequence. A frequency-based entropy of the bit pattern of a production rule leads to a measure of irregularity that would remove any circularity from the conclusion that random sequences are generated by irregular production rules.

In some cases, adding a single element 1 to a binary maximum-length sequence of period N of which the first element is also 1, one finds the bit pattern of a production rule for a maximum-length sequence of period 2^N-1. Since the frequency-based entropy of maximum-length sequences (for strings of size $n=\log_2 N$) is maximal, one may wonder whether iteration of this process would lead to ever more random recipes and sequences, of exponentially increasing ranges and periods.

In any case, entropy should be an essential element of a general theory for random sequences.

ACKNOWLEDGMENTS

Special thanks are due to Jouke Heringa in Delft and Dakai Wang in Xian, who contributed to the early stages of this research. The author is also indebted to Ora and Jerry Percus in New York, to Sacha Berdnikov and Sergey Turtia in St. Petersburg, to Lev Shchur and Andrei Talapov in Chernogolovka, and to Pierre L'Ecuyer in Montreal for many friendly and helpful discussions. This research was supported in part by a NATO grant (CRG 900661), and by a grant from the Dutch Organisation for Scientific Research (NWO 30-007).

[1] R. von Mises, *Probability, Statistics and Truth* (MacMilan, New York, 1957).

[2] A. N. Kolmogorov, *Foundations of the Theory of Probability* (Chelsea, New York, 1950).

[3] A. N. Kolmogorov, Russ. Math. Surv. **38** (4), 29 (1983).

[4] G. J. Chaitin, *Information, Randomness and Incompleteness* (World Scientific, Singapore, 1987).

[5] M. Kac, *Statistical Independence in Probability, Analysis and Number Theory* (Wiley, New York, 1959).

[6] P. Martin-Löf, Inf. Control **8**, 602 (1966).

[7] P. Martin-Löf, in *Intuitionism and Proof Theory,* edited by A. Kino, J. Myhill, and R. E. Vesley (North-Holland, Amsterdam, 1970), p. 73.

[8] P. Kirschenmann, J. Philos. Logic **1**, 395 (1972).

[9] M. van Lambalgen, J. Symb. Logic **52**, 725 (1987).

[10] M. van Lambalgen, Ph.D. thesis, University of Amsterdam, 1987.

[11] S. W. Golomb, *Shift Register Sequences* (Holden-Day, San Francisco, 1967).

[12] D. E. Knuth, *The Art of Computer Programming* (Addison-Wesley, Reading, 1981), Vol. 2.

[13] G. Marsaglia, in *Computer Science and Statistics,* edited by L. Billard (Elsevier, Amsterdam, 1985), p. 3.

[14] B. D. Ripley, J. Comput. Appl. Math. **31**, 153 (1990).

[15] F. James, Comput. Phys. Commun. **60**, 329 (1990).

[16] P. L'Ecuyer, Commun. ACM **33** (10), 87 (1990).

[17] H. Niederreiter, Ann. Op. Res. **31**, 323 (1991).

[18] A. Compagner, J. Stat. Phys. **63**, 883 (1991).

[19] A. Compagner, Am. J. Phys. **59**, 700 (1991).

[20] D. Wang and A. Compagner, Math. Comput. **60**, 363 (1993).

[21] R. R. Coveyou and R. D. MacPherson, J. Assoc. Comput. Mach. **14**, 100 (1967).

[22] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).

[23] R. K. Pathria, Math. Comput. **16**, 188 (1962).

[24] S. Wagon, Math. Intelligencer **7** (3), 65 (1985).

[25] B. R. Johnson and D. J. Leeming, Sankhyā Ind. J. Stat. **52B**, 183 (1990).

[26] J. R. Heringa (private communication).

[27] O. E. Percus and J. K. Percus, Combin. Prob. Comput. **1**, 161 (1992).

[28] A. Compagner and A. Hoogland, J. Comput. Phys. **71**, 391 (1987).

[29] P. L'Ecuyer and R. Couture, ORSA J. Comput. (to be published).

[30] I. Vattulainen, Ph.D. thesis, University of Helsinki, 1994; see also I. Vattulainen *et al.*, Phys. Rev. Lett. **73**, 2513 (1994).

[31] I. Vattulainen, K. Kankaala, J. Saarinen, and T. Ala-Nissilä, Comput. Phys. Commun. **86**, 209 (1995).

[32] K. Kankaala, T. Ala-Nissilä, and I. Vattulainen, Phys. Rev. E **48**, R4211 (1993).

[33] A. Hoogland, J. Spaa, B. Selman, and A. Compagner, J. Comput. Phys. **51**, 250 (1983).

[34] A. M. Ferrenberg, D. P. Landau, and Y. J. Wong, Phys. Rev. Lett. **69**, 3382 (1992).

[35] P. Grassberger, J. Phys. A **26**, 2769 (1993).

[36] W. Selke, A. L. Talapov, and L. N. Shchur, Pis'ma Zh. Eksp. Teor. Fiz. **58**, 684 (1993) [JETP Lett. **58**, 665 (1993)].

[37] W. Selke, L. N. Shchur, and A. L. Talapov, in *Annual Reviews of Computer Physics,* edited by D. Stauffer (World Scientific, Singapore, 1994), p. 17.

[38] G. Marsaglia, B. Narasimhan, and A. Zaman, Comput. Phys. Commun. **60**, 345 (1990).

[39] R. Couture and P. L'Ecuyer, Math. Comput. **62**, 799 (1994).

[40] S. Kirkpatrick and E. P. Stoll, J. Comput. Phys. **40**, 517 (1981).

[41] S. Tezuka and P. L'Ecuyer, ACM Trans. Mod. Comput. Sim. **1**, 99 (1991).

[42] L. Shchur (private communication).

[43] A. S. Berdnikov and S. B. Turtia, in *Computational Physics, Proceedings of the CP90 Europhysics Conference* (World Scientific, Singapore, 1991); and (unpublished).

[44] P. L'Ecuyer, Math. Comput. (to be published).

[45] N. Zierler, Inf. Control **15**, 67 (1969).

[46] Y. Kurita and M. Matsumoto, Math. Comput. **56**, 817 (1991).

[47] J. R. Heringa, H. W. J. Blöte, and A. Compagner, Int. J. Mod. Phys. C **3**, 561 (1992).

[48] N. Zierler and J. Brillhart, Inf. Control **13**, 541 (1968); **14**, 566 (1969).