STATISTICS, STATIONARITY AND RANDOM NUMBER GENERATION

Robert Bohrer and Peter B. Imrey
University of Illinois

ABSTRACT

This paper remarks upon some issues involved in evaluating the "randomness" of numerical sequences. The question of how much to test is addressed, particularly with respect to pseudorandom generators. Historical failures of seemingly random sequences are noted. The dependence of evaluation programme upon proposed use of the sequence is stressed. The meaning and importance of stationarity are considered, and results from statistical distribution theory useful in checking for it, and in further evaluation of a sequence, are described. An example illustrates differences in power of three tests directed against a particular class of stationary alternatives to Normal white noise.

1. INTRODUCTION

Recent developments in the foundations of probability theory (see, e.g., the review article by Chaitin (1)) have brought to the surface and spotlighted certain logical anomalies, involving the usage of algorithmically generated "pseudorandom" sequences which underlies the practice of simulation by a digital computer. In this paper we i) discuss, from a naive viewpoint, some implications of these anomalies for the simulator, and ii) survey some available procedures for detecting non-randomness, with particular attention to spectral tests and conditions under which they are most and least advisable.

Consider what might be called "standard practice." A stochastic model is developed which assumes that some portion of the phenomenon under study is describable as a random sample from a probability distribution. By a random sample we mean realizations of random variables that are a) each from the same probability distribution, and b) statistically independent. A search is then made for numbers which plausibly impersonate such realizations. Since thousands of numbers are frequently needed, the economics of computation suggest that methods of generating the numbers on-line are desirable. Economics further dictate that an on-line generating procedure must be short and sparing of memory to be usable; a simple deterministic algorithm is thus selected. A series so resulting, calculated deterministically in the hope that it may successfully masquerade as a probabilistically generated sequence, is called "pseudorandom," and the algorithm a pseudorandom number generator.

The generator is carefully chosen mathematically to avoid undesirable properties such as, say, periods shorter than the sequence length for planned Monte Carlo. In addition, number theoretic asymptotic approximations which indicate that certain aggregate properties of the pseudorandom sequence agree with predictions of the stochastic model can sometimes be derived. Finally, on the basis of a more or less vague feeling that there is no reason to suspect nonrandomness and that hypothesis testing is relevant to matters statistical, the stochastic model is elevated to the status of both a scientific and statistical null hypothesis for generator characteristics, and subjected to a battery of statistical tests using generated subsequences. If these all accept the null the generator is called acceptable and used, frequently by many researchers in a wide variety of situations.

What does a statistical programme for screening a random number generator accomplish? Essentially, it involves the rejection of generators on the basis of subsequences which are atypical of the hypothesized stochastic model. "Atypical" subsequences are defined by membership in the critical regions of some set of hypothesis tests. Ordinarily, a test statistic provides a measure of atypicality, from the point of view of a particular test, for each conceivable subsequence. The nature of a statistical screen therefore depends upon the choice of tests, which designates the types of sequences to be rejected, and the stringency of application in terms of the numbers of sequences in the critical regions of the various tests. All this seems eminently reasonable, until one considers the choices to be made in designing a statistical screen for a "general purpose" random number generator. Which test should one choose at what levels? How does one handle the multiple testing problem? Just as all readers of this paper are, in some combination of characterstics, unusual human beings, so sufficiently close examination of any pseudorandom sequence will reveal some features that are not shared in the aggregate by run-of-the-mill sequences from a hypothesized model. The question arises of when the passage of an intensive screening itself becomes a trait sufficiently unusual, on the basis of a stochastic model, to require rejection of a generator.

The severity of the paradox involved becomes clearest when viewed in the light of the algorithmic definition of randomness of Kolmogorov, Chaitin and Solomonoff (1). While the classical definition of randomness refers to the probabilistic origin of the sequence, the algorithmic formulation treats randomness as a trait of the sequence itself, independently of how it was arrived at. In particular, a sequence is called random if there exists no formal computer program for production of the sequence which is substantially smaller in information content (bits) than the sequence itself. If the complexity of a sequence is defined as the minimal amount of information necessary to tell a computer how to construct the sequence, then a random sequence is one whose complexity is approximately equal to the amount of information it carries. Moreover, the complexity of a sequence induces a partial ordering of randomness on the class of possible sequences of a fixed length; those of least complexity are the least random.

According to this definition, the very simplicity and brevity which make a pseudorandom number generator economically attractive insure that all sequences of at least moderate length produced by the generator will be markedly non-random. Consider for a moment the question of what to do in testing a sequence of unknown origin for randomness. Since we do not know if the sequence is pseudorandom, a statistical testing programme would still be appropriate according to the algorithmic formulation. However, the dependence of a statistical screen on the particular hypothesis test chosen is emphasized by consideration of what the most appropriate tests might look like in this situation. They would try to reject generators of low complexity. We might, for example, apply a test against a class of congruential or mixed congruential generators operable within current machine capacities. The test would reject whenever the new sequence satisfied one of the generator algorithms, and accept otherwise. If the generator was used to produce "continuous" pseudorandom numbers to a sufficient number of figures, and the class not too large, the test would have 100% power against the included congruential or mixed congruential generators, at a level close to zero under any continuous null distribution. Such a test is not less reasonable than more conventional tests of randomness.

The point we wish to draw from this admittedly labored discussion is that the concept of establishing a pseudorandom generator as a general purpose supplier of random numbers through a varied programme of statistical hypothesis testing is invalid. While statistical tests may provide useful ordering principles for the exclusion of potentially misleading sequences, their most appropriate application is in conjunction with a perturbation analysis of the proposed simulation model, to determine what sorts of violations of the stochastic model, and of what magnitude, would compromise the outcome. Once this is known, specific statistical tests of type, level and power such as are necessary to screen out generators which produce such sequences may be selected.

We do note a circularity inherent in this discussion. The perturbation analysis must be accomplished without simulating, as we have no baseline sequence for our simulation. And yet, the lack of a good mathematical perturbation analysis frequently is what justifies a simulation in the first place. The difficulty is hardly devastating, however. Frequently we know much but not all about a model to be simulated, and the knowledge that we have often implies that certain stochastic assumptions are more critical than others in the problem context. Certainly, we should test against critical violations extensively, but the emphasis we should put on supplementing such testing with examinations designed to delete violations without known deleterious effects is not clear. While undertesting may let very bad generators for the problem slip through, overtesting may ultimately suppress from the output of our simulations interesting and realistic variation, carrying the sort of information which makes simulation a valuable predictive tool. The recommendation of McLaren and Marsaglia (15), to test on a similar problem for which the answer is known, is certainly sound when possible. Ultimately, the user of pseudorandom sequences must accept the fact that he will occasionally take his lumps.

The remainder of this paper surveys somewhat more specifically what statistics can say about pseudorandom generators. In Section 2, real-life bad examples illustrate how care must be used in choosing a generator which gives those randomness properties of most concern to the user. Some tests from the literature are discussed in relation to the types of non-randomness and non-stationarity to which they are sensitive in Section 3 and finally, in Section 4, the characteristics of some tests for stationary series are reviewed in more detail.

2. HISTORICAL PERSPECTIVES

The history of random number generators, which purport to generate white noise, has been rather colorful. A few cases are presented here to emphasize by example how care must be used in adopting a generator.

Early in the century, Student and others used random physical measurements as random Normal numbers. These may be suspect because of abnormal human variation and they are inconvenient because the entire table of them must be stored in one's memory, computer or library. RAND Corporation (17), to overcome the first of these drawbacks, devised an electronic device to generate one million random numbers. Although the physical construction of the device would have seemed to guarantee randomness, RAND's statistical tests comfortably accepted the hypothesis at the beginning of the generation, but less comfortably so near the one-millionth number. Again, each of these numbers must be stored in memory for use.

Pseudorandom generators, on the other hand, require only a few memory locations. These date back at least as far as von Neumann's (20) "mid squares" uniform pseudorandom generator, which obtains the next number as the middle n digits of the square of the previous number (n even). At first sight, there may seem little reason to doubt its randomness; however, trouble can ensue. For a simple example with n=2, consider the sequence beginning with the 2n digit number 4321, vis., 32, 02, 00, 00, 00, ... ad inf. Such an anomaly, if programmed into a computer and left there without periodic checking, in future use might generate ludicrous Monte Carlo results before discovery of the problem.

Asymptotic theorems from number theory show that the generator $X(N)=(Nk) \pmod 1$ has excellent frequency properties if $k=(-1-\sqrt{5})/2$; that is, the proportion of numbers in any subinterval of the unit interval is very close to the length of the interval, at least in long sequences. With this assurance, the first author used this generator with abandon in a class project with very poor Monte Carlo verification of theoretical results. Also, aleatoric (random) music was produced with this generator and was found to contain highly nonrandom melodies. Further statistical testing, based on consecutive pairs, verified that these numbers appeared not at all independent, although the frequency property was splendid. All properties of concern to the study should have been tested.

Multiplicative congruential generators were introduced by Lehmer (10), and number theoretic theorems can be used to suggest which starting points and multipliers to use with them. The generator RANDU by IBM is an example which was in somewhat general use. RANDU passed many statistical tests; however, McLaren and Marsaglia (15) and Coveyou and MacPherson (5) show that triples of such numbers appear very dependent. Richardson (18) reports tests of this and other such generators.

Marsaglia (14) has shown a general, geometrical non-randomness results from multiplicative congruential generators. If such non-randomness is an important property to avoid in one's simulation work, then one should avoid these generators. Recently one of our students used a local generator, which was designed to overcome this geometrical problem. Weird Monte Carlo results led him to statistical tests of the generator, which revealed very bad frequency properties, e.g., very significantly more than half of his uniform numbers were greater than one-half. Randomness properties not of vital concern should not be traded for those which are.

The standard Normal pseudorandom generator which approximates Normals by adding up 12 pseudouniforms, subtracting 6, and appealing to the Central Limit Theorem, provides another semi-fanciful example of why the user must test a generator for the properties of concern to him. If, for example, one wants to study the tail behavior of the normal distribution by considering only those simulation runs wherein the deviate exceeds 6.0, then he will wait forever if the generation is done with this method.

With recent widespread popularity of pocket calculators (and sale of digital watches which double as gaming houses), the performance of their random number generators is of some interest. It is noteworthy that Hewlett-Packard recalled its original generator for the HP 65. Miller (16) proposed a faster and purportedly better generator than the one supplied by the Texas Instruments SR 52.

3. SURVEY OF STATISTICAL TESTS

The range of utility of a statistical test for randomness applied to judge a pseudorandom sequence is determined by the interaction, through the statistic of interest, of quite separable characteristics of the sequence represented by adherence to or violation of four components of the random sample stochastic model: i) covariance, or weak sense, stationarity (Hannan (8)); ii) identical marginal distributions, given covariance stationarity; iii) independence, given covariance stationarity; iv) nature of distribution, given i)-ii). Every statistical test will be more sensitive to some alternatives than do others. For example, anyone with imagination can concoct non-stationary dependency structures against which any given test will be magnificant or powerless, depending on the point to be illustrated. With respect to i) we note that underlying the decision to test a pseudorandom generator by examining subsequences is a more stringent assumption of strong-stationarity of the generator; that is, that the statistical properties of the sequence tested are the same as those of all other sequences which may be obtained from the generator in the future. A generator that is not stationary is certainly unreliable. For every truly random sequence is stationary and, if a generator is highly non-stationary, no inference however precise about one subsequence can tell much about the potentially different properties of another subsequence. For example, the precise knowledge of the mean of one subsequence tells little about the mean of another subsequence unless the generator is stationary, insuring that the means are equal.

A list of tests of random numbers was compiled by the authors from a set of prominent papers in the field (6,9,11,12,15). No attempt was made to be exhaustive or to select a "best set" of papers.

We briefly comment on properties of the tests found.

The frequency test, perhaps the most well known of all tests of randomness, is simply a chi-square goodness-to-fit test to the hypothesized distribution. It is useful against certain types of distributional shifts, such as produce very high frequencies in certain cells used in computing the test statistic, but is much more tolerant against less dramatic shifts. Since this is a general purpose non-parametric test, it is in particular situations inferior to its parametric alternatives. (This last comment applies to any test based on chi-square, Kolmorogov-Smirnov or other general goodness-of-fit statistics, no matter what the statistic is whose distribution is chosen to be tested.) The test is permutation invariant, so that it may function very poorly against ordered sequences of the sort that may result from dependencies, and behaves erratically in the face of distributional inhomogeneities, depending on the type of distributional differences encountered.

The n-tuples test is an n-dimensional generalization of the frequencies test. Cells used are identical subcubes of the n-dimensional hypercube, with n-tuples taken as points in n-space. The test statistic is invariant to permutations of n-tuples, or simultaneous identical permutations within n-tuples, but no others. Thus it is sensitive to some types of dependence of lag less than n, its advantage over the frequency test. Otherwise, comments on the frequency test apply. The minimum or maximum of n tests compare extrema of n numbers to their theoretical distributions using a chi-square test; thus, the test concerns tail behavior, and can be used to protect against distributional contamination by outliers, truncation alternatives, or non-stationarities such as trends. It is invariant to the same permutations as the n-tuples test. The analogous sum of n test is more useful against less dramatic slow shifts or slippages in underlying distribution, or against stationary distributional alternatives. These tests may perform very poorly when stationarity is absent.

The venerable poker hand test examines the distribution of the maximum number of matches of five digits. This test is sensitive to association of lag less than five, and to some distributional shifts. Different types of runs tests, such as runs above and below mean or median, or runs up and down, are useful against general alternatives to independence, and distributional shifts. Some are unaffected by stationary departures from assumed distribution; obviously, they utilize only information about the order of the sample, and are in no way permutation invariant. Gap tests and concatenated gap tests, which examine the distribution of gaps between similar digits in digit sequences, are excellent tests of general alternatives. It is possible to suggest sequences for which they fail entirely, however, since they are invariant to permutations of gaps. Thus, asymmetrical digit distributions which permute probabilities across digits at regular intervals produce sequences which generally will satisfy gap tests. Kendall and Smith, who proposed these tests (9), remark that a non-random series which escaped the gap test "would, it appears, have to have a very peculiar bias indeed, such as would hardly ever arise in practice." One would have to agree, but add that some simulations today are very peculiar, as is the concept of pseudorandom numbers which was unknown to those authors.

The hypersphere test and sums of squares test use the generator to approximate known properties of distributions of quadratic forms of generated numbers. Since each focuses on particular parameters of the associated distribution, it will be sensitive to departures from the distributional assumptions which change these parameters (in the case of the sums of squares test, these are particular percentiles of a chi-square distribution), and insensitive to departures not reflected in the parameters chosen. Thus, the sums of squares test will not be sensitive to distributional departures which preserve percentiles but not the basic shape of the relevant chi-square distribution. The choice of percentiles used may drastically alter the characteristics of the test. The concept of the test is sound for looking at heavily skewed distributional departures from Normality. Use of tail percentiles may be quite effective for detecting occasional dramatic distributional contamination.

We could not interpret the D^2 test of a random line, the binary tree test or the scattering experiment test, incompletely described (12), but suspect the properties of the first are generally similar to those of the n-tuple test.

The class of spectral tests, including the maximum autocorrelation to lag k test, median-spectrum test, the Grenander-Rosenblatt test (which is a Kolmogorov-Smirnov test on the integrated periodogram) or the modified Bartlett spectral test, are strongly dependent on covariance stationarity and may be totally ineffective against non-stationary alternatives. The tests differ essentially in which alternatives to a flat spectrum they select for exclusion. The maximum autocorrelation test is sensitive against strongly periodic spectra. The median-spectrum test rejects alternatives partial to high or low frequencies, the modified Bartlett test rejects clumping of the distribution into smaller frequency ranges, and the Grenander-Rosenblatt test is a broad but weak simultaneous screen against all patterns of autocorrelation. We remark further on spectral tests in Section 4.

From the variety of tests noted, it should be clear that researchers have approached the problem of checking randomness with considerable ingenuity and imagination. It should also be clear that the invention of tests for randomness is, however, an essentially easy problem, as the whole panoply of statistical methodology may be brought to bear.

Tests of virtually any hypothesis may be viewed as or converted to tests of randomness. Some very good tests against particular alternatives are immediate consequences of standard statistical methods, and are clearly more effective against particular classes of alternatives than any of those proposed. As an example, simple linear regression of generated number on place in sequence is an excellent test of slow linear slippage in mean under normality, homoscedasticity and independence, a test certainly superior under these conditions to any test previously mentioned. Covariance stationarity may be examined by application of analysis of variance proceudres to transformed sample autocorrelations derived from subsequences. Spectral tests which take into account that autocorrelations of lag nk, n = 2,3, . . . should appear in conjunction with true lag k autocorrelation can be developed. In general, likelihood ratio statistics may be constructed against specific alternatives to obtain asymptotically efficient tests of fixed sample size. Sequential methods may be used against some simple alternatives. The possibilities for development of tests explicitly tailored to alternatives are virtually unlimited, and should involve much fun for statisticians. But many such tests will overlap, and it is essential that they not be applied in shotgun fashion, since the use of redundant tests may serve to sanctify a bad generator and the use of too many tests of disjoint critical regions will result in the utilization of only numerical mush as input to our simulations. Pseudorandom number generators should be examined using a battery of tests of moderate proportions, members of the battery chosen specifically with reference to the nature of non-randomness which threatens the desired simulation. Overall, pseudorandom number generators may then serve us well, with a little bit of luck.

4. STATISTICS AND THE STATIONARY CASE

A generator is strongly stationary if its statistical properties do not vary over time. If they vary over time, then a) the generator cannot be producing identically and independently distributed random variables, as purported, and b) one cannot infer the properties of later sequences to be the same as those of sequences tested formerly. Thus, stationarity is an important property for a random number generator. Here we cite facts which, used in tandem, can provide powerful tests for stationarity and randomness, and we comment on the power of these tests.

As stressed previously, pseudorandom generators do not give truly random numbers, so it is essential to test one's generator for those, and only those, properties which are of concern to his study. Should this require several tests, then one must keep in mind the problem of multiple comparisons. That is, two (or more) tests are not as cheap as one, but rather have a higher chance of making at least one error when all hypotheses tested are true. When this is taken into account, the resulting inferences are necessarily less powerful for detecting any given departure from the hypotheses. One safe way to proceed is to do both individual and simultaneously valid inferences, rejecting any hypotheses rejected by the simultaneous procedure, while collecting fresh and independent samples with which to retest hypotheses which are accepted by the simultaneous test but rejected by the individual ones. For example, "reject if $|\overline{X}| \geq 1.64/\sqrt{n}$" is a test that the mean of unit variance Normal variables is zero, and this test has significant level .10. If two such n-samples are tested like this, simultaneously, and one requires .10 for the combined significance level, i.e., the chance that either test rejects when both means are zero, then each test must be of the form "reject when $|\overline{X}| \geq 1.95/\sqrt{n}$", which is somewhat less sensitive than the "1.64 tests" when a hypothesis is false. Should either $|\overline{X}|$ exceed 1.96, then its hypothesis should be rejected, while a value between 1.64 and 1.96 is cause for retesting with a fresh sample.

In testing for several properties of randomness in a generator, it should be kept in mind that tests for one such property may assume that other such properties are valid. If the tests are performed one after another, this will specify the logical order of testing.

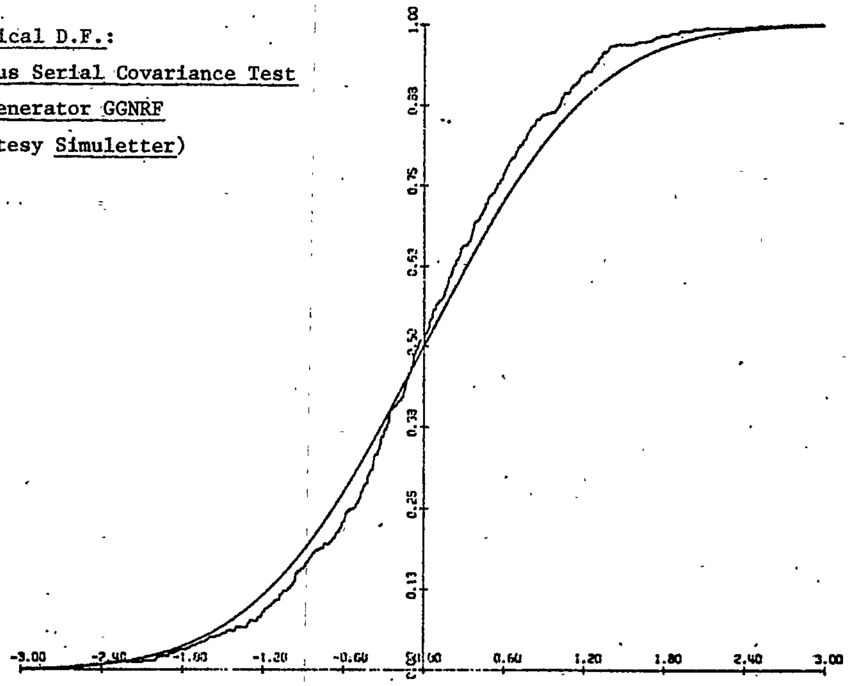
The basic building blocks for testing generators come from statistical distribution theory. Next, a comprehensive but not exhaustive collection of these results and their application, particularly useful for testing for or under stationarity, is given along with the authors' favorite references to these results. Some of these results are given in more detail by Bohrer and Putnam (2).

Assuming stationarity, unit variance and Normality (or, for a conservative test, fourth moment less than 3), independence can be tested using estimators C(K) for the autocovariance at lag K. The null and alternative distributions for these are asymptotically normal, and under the null hypothesis, the C(K) are independent for different values of K. This provides a "serial covariance" test at any lag K, vis, reject if $C(K) \geq (N-K)^{-\frac{1}{2}}$. A test which is valid simultaneously for several lags K, and which provides a better omnibus picture of the randomness situation, is based on the null distribution of $\sqrt{(N-K)}\, C(K)$ being that of a random sample from the standard normal distribution. Thus, the empirical distribution function for the several covariance estimators, so normed, can be compared to the standard Normal, e.g., by picture, Kolmogorov-Smirnov statistic, or Cramer-von Mises statistic; see Wilks (21).

Figure 1 shows the appropriate picture for a sequence from the IMSL generator GGNRF (ref. Bohrer and Putnam (2)). Alternatively, assuming stationarity within the sequence tested, independence can be tested using the normal distribution theory for the periodogram (see Brillinger (3), Chapter 5) or its integral (see Grenander and Rosenblatt (7), Chapter 6). This procedure seems first to have been used for randomness testing by Lewis et al. (11). Like the omnibus serial covariance test it assumes stationarity, but guards against dependence at all lags simultaneously. The Grenander-Rosenblatt test compares the integrated periodogram with the

ILLUSTRATION 1

Empirical D.F.:
Omnibus Serial Covariance Test
For Generator GGNRF
(Courtesy Simuletter)



right triangle of height .5, which would be expected under independence; the critical points for the maximum absolute difference between the hypothesized and empirical distributions are given on page 196 of Grenander-Rosenblatt (7). Figures 2, A and B, show how the non-aleatoric generator of Section 2 and the PDP-11 generator, respectively, fail and pass this test. Figures 2A and 2B note that the distribution theory applies asymptotically and exactly for Normal random sequences, and conversatively for uniform sequences. To test from a specified distribution, empirical distribution functions can be compared visually with the hypothesized distribution; if independence can be assumed, than Kolmogorov-Smirnov distribution theory can be used for hypothesis testing.

Tests for stationarity of a generator, between sequences of observations, can be made if the hypothesis of randomness within each sequence is acceptable. For then, the random variables can be transformed to normal variables, and stationarity is tested by a standard, one-way analysis of variance; see Scheffe (19), Chapter 3.

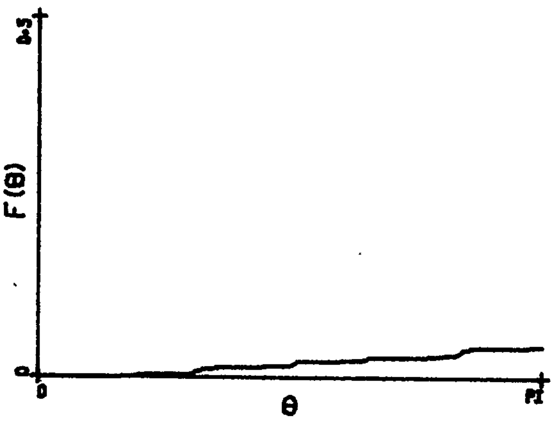
Omnibus tests of several hypotheses tested independently, or about the same hypothesis tested on independent sequences of data, may be combined by using the fact that the null distribution of the negative of twice the summed logged P-levels of the individual tests has a chi-square distribution with degrees of freedom twice the number of tests; ref. Anderson and Bancroft (1), Chapter 12.

Power

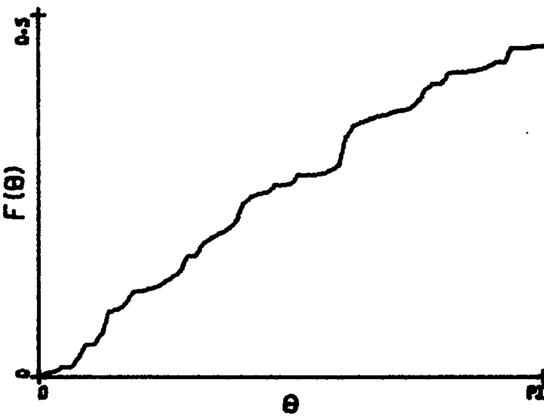
Assume now we are sampling from a large pseudo-random sequence. Consider a single example to compare the sensitivity of poker hand, lag-5 serial covariance, and omnibus spectral test, to departures from randomness. The non-random case considered is one with fifth order dependence, vis. X(N)= (W(N)+pW(N-5)), where W(N) is a sequence of independent standard Normals. Since the poker hand test considers only 5-tuples, and since no 5-tuple contains both W(N) and W(N-5), it has power equal to the significance level, near zero. The power of the serial covariance test is calculated using Hannan's (8, page 40) formula for the variance and is tabulated in Table 1, as a function of p, and sample size. By an argument using Normal distribution theory and Stein's proof for termination of the sequential probability ratio test, a lower bound on the power of the spectral test can be obtained. This lower bound increases to one as sample size increases. The

ILLUSTRATION 2

Integrated Periodograms



A. Non-Aleatoric Generator



B. PDP-11 Generator

(Courtesy Porges-Cheung-Drasgo)

bound is lower than the power of the lag 5 serial correlation test, because it is a bound, but probably more because it protects against more alternatives, and hence must pay the price in power. See Table 2.

Powerlessness

Of course, tests which assume stationarity, may be very bad for series' which are not stationary. For example, the spectral test with significance level α has power α for protecting against the non-stationary alternative $X(N)=(W(N)+(-1)^N * pW(N-1))$, for any $p\geq 0$, where the W(N) are identically and independently distributed.

TABLE 1

Power of the Lag-5 Serial Covariance Test

Level	P	Sample Size	Power
0.05	0.05	100	.08
0.05	0.10	100	.18
0.05	0.20	100	.52
0.05	0.05	250	.13
0.05	0.10	250	.36
0.05	0.20	250	.86
0.05	0.05	500	.20
0.05	0.10	500	.61
0.05	0.20	500	.99
0.10	0.05	100	.15
0.10	0.10	100	.27
0.10	0.20	100	.63
0.10	0.05	250	.21
0.10	0.10	250	.48
0.10	0.20	250	.92
0.10	0.05	500	.31
0.10	0.10	500	.72
0.10	0.20	500	1.00

TABLE 2

LOWER BOUNDS ON POWER OF THE OMNIBUS SPECTRAL TEST

LEVEL	P	SAMPLE	POWER
0.05	0.3	500	0.00073
0.05	0.5	500	0.12732
0.05	0.7	500	0.84418
0.05	0.9	500	0.99804
0.05	0.3	1000	0.00608
0.05	0.5	1000	0.64754
0.05	0.7	1000	0.99954
0.05	0.9	1000	1.00000
0.01	0.3	500	0.00001
0.01	0.5	500	0.01105
0.01	0.7	500	0.39317
0.01	0.9	500	0.95699
0.01	0.3	1000	0.00009
0.01	0.5	1000	0.13475
0.01	0.7	1000	0.96762
0.01	0.9	1000	1.00000

BIBLIOGRAPHY

1. Anderson, R. L. and Bancroft, T. A. (1952). Statistical Theory in Research. McGraw-Hill, New York.
2. Bohrer, R. and Putnam, P. (1977). "Statistical tests of random number generators." Simuletter VIII, 85-91.
3. Brillinger, D. R. (1975), Time Series: Data Analysis and Theory. Holt, Rinehart and Winston, New York.
4. Chaitin, G. J. "Randomness and mathematical proof," Scientific American, 47-52.
5. Coveyou, R. R. and MacPherson, R. D. (1967). "Fourier analysis of uniform random number generation," The Statistician, 20:55-66.
6. Craddock, J. M. and Farmer, S. A. (197), "Two robust methods of random number generation," The Statistician, 20:55-66.
7. Grenander, U. and Rosenblatt, M. (1957), Statistical Analysis of Stationary Time Series, John Wiley and Sons, New York.
8. Hannan, E. J. (1969), Time Series Analysis, Methuen and Co. Ltd., London.
9. Kendall, M. G. and Babington-Smith, B. (1938), "Randomness and random sampling numbers," J. Roy. Stat. Soc., 101:147-66.
10. Lehmer, D. H. (1949), "Mathematical methods in large scale computing units," Proceedings of the Second Symposium on Large-Scale Digital Calculating Machinery, 141-6.
11. Lewis, P. A. W., Goodman, A. S. and Miller, J. M. (1969). "A pseudorandom number generator for the System/360," IBM Systems Journal, 8:136-146.
12. Lewis, T. G. and Payne, W. H. (1973). "Generalized feedback shift register pseudorandom number algorithm," J. Assoc. Comp. Mach., 20: 456-68.
13. Marsaglia, G. (1968). "Random numbers fall mainly in the planes," Proc. N.A.S., 61:25-28.
14. Marsaglia, G. (1970). "Regularities in congruential random number generators," Numer. Math, 16:8-10.
15. McLaren, M. D. and Marsaglia, G. (1965). "Uniform random number generators," J. Assoc. Comp. Mach., 12:83-88.
16. Miller, A. (1977). Note in CSIRO Newsletter, No. 28.
17. RAND Corporation (1955). A Million Random Digits with 100,000 Normal Deviates, Free Press, Glencoe.
18. Richardson, B. (1969). "Random number generation on the IBM 360," University of Illinois DCS Report No. 329.
19. Scheffe, H. (1959). The Analysis of Variance. John Wiley and Sons, New York.
20. von Neumann, J. (1959). "Various tests used in connection with random digits." Monte Carlo methods, National Bureau of Standards. Applied Math, Series 12, 36-38.
21. Wilks, S. S. (1962). Mathematical Statistics, John Wiley and Sons, New York.

ACKNOWLEDGEMENTS

This research was supported in part by NSF Grant MPC 77-07978. The authors are grateful to Connie Parsley and Gerry Swift for extraordinary assistance in manuscript preparation.