Proceedings of the 1999 Winter Simulation Conference
P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, eds.

MONKEYS, GAMBLING, AND RETURN TIMES: ASSESSING PSEUDORANDOMNESS

Stefan Wegenkittl

Department of Mathematics
University of Salzburg
A 5026 Salzburg, AUSTRIA

ABSTRACT

We present a general construction kit for empirical tests of pseudorandom number generators which comprises a wide range of well-known standard tests. Within our setup we identify two important families of tests and check for connections between them. This leads us to quiery the existence of universal tests which claim to be able to detect *any* possible defect of a generator.

1 INTRODUCTION

Whereas the art of constructing pseudorandom number generators (PRNGs, see Knuth 1997, L'Ecuyer 1994, and Hellekalek 1998a for overviews) is that of carefully hiding the deterministic nature of the afterwards presumed random numbers, the art of empirical testing is to find the hidden correlations and to analyze their impact on simulation studies and Monte Carlo algorithms. Following Marsaglia and Zaman (1993), a good PRNG produces an output which does not differ significantly from that of a (memoryless and fair) monkey hitting keys on a numeric keyboard.

Theoretical tests (Hellekalek 1998b, Niederreiter 1992, 1995) - as they are often provided by the authors of a PRNG themselves - usually ensure the quality of the sample space which is the set of all possible realizations that can be obtained from the generator. We are left to empirical testing - in which the PRNG is treated as a black box - for gaining confidence in that the *samples* will suit the needs of our application. Here, we try to remodel important features of the target application in a test statistic and to seek for any "non-monkeyness" in the corresponding test results.

The huge amount of empirical tests presented in literature in various setups and styles (see Bratley, Fox, and Schrage 1983, Knuth 1997, and L'Ecuyer 1992 for surveys, and Altman 1988, Bernhofen et. al. 1996, DeMatteis and Pagnutti 1995, Dudewicz et. al. 1995, Eichenauer-Herrmann, Herrmann, and Wegenkittl 1997, Entacher, Uhl, and Wegenkittl 1998, Ferrenberg, Landau, and Wong 1992, Marsaglia 1985, and Vattulainen, Ala-Nissila, and Kankaala 1994, 1995 for examples) makes it difficult to rate on the differences and redundancies within these batteries. We present a construction kit which unifies many well-known approaches in a general setup in Section 2 and analyze two important families of tests in Sections 3 and 4. We will see that these families are strongly connected by the notion of entropy. We consider a scale ranging from highly specific to rather universal tests in Section 5 and examine the necessity of both types of tests.

2 CONSTRUCTION OF TESTS

The construction kit in Figure 1 exhibits the major building blocks of empirical testing procedures. From top downwards we have two input modules, the PRNG and the monkey, three feature extraction modules and a comparison module. The latter, **C**, is used to measure the extent of "non-monkeyness" of the PRNG with respect to the selected features. The test rejects the generator if the observed behavior is very unlikely to occur when replacing it by the monkey. As to the middle modules, we have

- a keyboard device **K** for turning the pseudorandom numbers (PRNs) into letters from a finite alphabet $\mathcal{A}$, $\#\mathcal{A} = \alpha$, such as bit cutting mechanisms, even-odd-testing, or coin-throw simulation. Under the monkey hypothesis, the sequence of letters is assumed to be an i.i.d. uniform random sequence on $\mathcal{A}^\infty$ with each letter $a \in \mathcal{A}$ having a fixed probability π_a.
- a finite state automaton **A** with state space $S = \{1, \ldots, m\}$ which makes transitions according to the input letters,
- and an observation unit **O** for reporting statistics on the state automata such as the number of visits in each state, see Section 3, or the return time to a certain state, see Section 4.

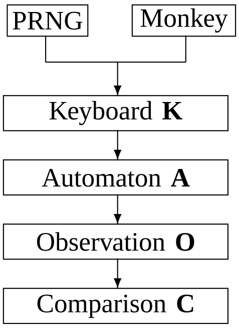

Figure 1: Construction Kit for Empirical Tests

The number of letters that is generated during a test is called sample size N. Depending on whether the ratio m/N is small or large, we distinguish the dense and sparse case, respectively. Well known examples are:

- the serial test (Good 1953, Knuth 1997) in dimension s (also called chi-square test or relative frequency test): this test compares the relative number of visits to the expected number of visits in each state of the automaton. It assumes independence of the successive states under the monkey hypothesis, see Section 3 for the dense case, and the tests in L'Ecuyer, Simard, and Wegenkittl (1998) for the sparse case.
- the overlapping serial test (also called overlapping M-tuple test in Marsaglia 1985), which is computed from overlapping s-tuples of successive letters, see again Section 3.
- more general serial tests (see Cressie and Read 1984 and the so-called monkey tests in Marsaglia and Zaman 1993) where **O** counts e.g. the number of states never visited (also called empty-cells test), or collisions (more than one visit in a state). Some of these tests are more powerful when the sparse case ($m >> N$) is considered, see L'Ecuyer, Simard, and Wegenkittl (1998) for the corresponding distributions and results.
- return time tests (see Choe and Kim 1999, and Maurer 1992), where the observation unit measures the time between two visits in a certain state, see Section 4.

Many other tests proposed in literature - such as the run and poker test and some of the tests based on random walks - can be reformulated in terms of the general scheme for serial tests considered in Section 3 or in terms of return times to certain states, see Section 4. In order to perform significance testing it is necessary to have a combination of units which allows the derivation of the asymptotic distribution or moments under the monkey hypothesis. Often, the test is performed several times and a second level test is used to rate on the results, see Knuth (1997), Chap. 3.3, and Leeb and Wegenkittl (1997).

3 SERIAL TESTS FOR MARKOV CHAINS

Consider a keyboard **K** for the finite alphabet $\mathcal{A} = \{1, \ldots, \alpha\}$ with cardinality α and an automaton unit **A** with state space $S = \{1, \ldots, m\}$. The transitions in **A** shall only depend on the incoming letters and on the current state, but not on the time when the letter is received. Under the monkey hypothesis, we may analyze the random sequence of states, $X = (X_0, X_1, \ldots, X_{N-1})$, $X_l \in S$, assumed by **A** as a Markov chain. Each entry p_{ij} in the transition matrix $\mathbb{P} = (p_{ij})_{(i,j)\in S^2}$, $p_{ij} = P[X_l = j | X_{l-1} = i]$, $1 \le l \le N-1$, denotes the probability of going from one state to another and equals the sum of all π_a of letters that yield the corresponding transition in **A**. We will restrict ourselves to the case of irreducible and aperiodic chains $(S, \mathbb{P})$ in the following so that we get the ergodicity of X and a unique stable distribution $\mathbf{P} = (p_1, \ldots, p_m)$, $\mathbf{P} \cdot \mathbb{P} = \mathbf{P}$, for free. We also assume that the initial state X_0 has distribution $\mathbf{P}$ so that X becomes a stationary sequence. In the dense case, the effect of an arbitrary initial distribution wears off exponentially so that we can neglect its influence in the following.

In a serial test, the observation unit **O** counts the number of visits in each state of the chain and reports the vector of relative frequencies, $\hat{\mathbf{P}}^{(N)} = (\hat{P}_1^{(N)}, \ldots, \hat{P}_m^{(N)})$, $\hat{P}_i^{(N)} = 1/N \cdot \#\{0 \le l \le N-1 \ : \ X_l = i\}$. Clearly, the expectation $E[\hat{\mathbf{P}}^{(N)}]$ equals the stable distribution $\mathbf{P}$. The matrix of variances and covariances $V[\hat{\mathbf{P}}^{(N)}]$ can, owing to the stationarity, be written $V[\hat{\mathbf{P}}^{(N)}] = \Sigma$ for a fixed matrix Σ. Both, $\mathbf{P}$ and Σ, can be computed from $\mathbb{P}$ numerically for modest m, see Wegenkittl (1998, 1999b). We will see below, that the numerical computations can be omitted for a fairly large class of automata **A** for which theoretical results are available. The Central Limit Theorem for Markov chains yields asymptotic normality, $\sqrt{N}(\hat{\mathbf{P}}^{(N)} - \mathbf{P}) \approx N(0, \Sigma)$, for large N.

As to the comparison unit **C**, we might compute a weak inverse $\overline{\Sigma}$ of Σ, i.e., a matrix satisfying $\Sigma\overline{\Sigma}\Sigma = \Sigma$, and use the quadratic form $I = (\hat{\mathbf{P}}^{(N)} - \mathbf{P}) \cdot \overline{\Sigma} \cdot (\hat{\mathbf{P}}^{(N)} - \mathbf{P})^*$, where the $*$ denotes the transposed vector. A more general statistic considered in Wegenkittl (1998, 1999a, 1999b) is the modified generalized ϕ-divergence $\tilde{I}_{\Sigma,\varphi,q}$, which we cite in a slightly simplified version:

Let the function φ, $[0, \infty) \to (-\infty, \infty]$ have continuous second derivatives on some nonempty interval $(1-\delta, 1+\delta) \subset [0, \infty)$ with $\varphi(1) = \varphi'(1) = 0$ and

$\varphi'' := \varphi''(1) \neq 0$. Further let $\overline{\Sigma}^{(l)} = (\overline{\sigma}_{ij}^{(l)})_{(i,j)\in S^2}$, $1 \leq l \leq q$, be a family of $m \times m$ real matrices. Put

$$\phi^{\varphi}(x, y) := 2\varphi\left(\frac{x+y}{2}\right) - \frac{\varphi(x) + \varphi(y)}{2},$$

and define the modified generalized ϕ-divergence

$$\tilde{I}_{\overline{\Sigma},\varphi,q}(\hat{\mathbf{P}}^{(N)}, \mathbf{P}) =$$

$$= 2n \sum_{l=1}^{q} \frac{1}{\varphi''} \sum_{i,j=1}^{m} \overline{\sigma}_{ij}^{(l)} P_i P_j \phi^{\varphi}\left(\frac{\hat{P}_i^{(N)}}{P_i}, \frac{\hat{P}_j^{(N)}}{P_j}\right).$$

On the conditions that $\sqrt{n}\left(\hat{\mathbf{P}}^{(N)} - \mathbf{P}\right) \overset{\mathrm{d}}{\to} N(\emptyset, \Sigma)$ for a covariance matrix Σ, and that $\overline{\Sigma} := \sum_{l=1}^{q} \overline{\Sigma}^{(l)}$ is a weak inverse of Σ, $\Sigma\overline{\Sigma}\Sigma = \Sigma$, this statistic is asymptotically distributed chi-square with $R(\Sigma)$ degrees of freedom,

$$\tilde{I}_{\overline{\Sigma},\varphi,q}(\hat{\mathbf{P}}^{(N)}, \mathbf{P}) \overset{\mathrm{d}}{\to} \chi^2_{R(\Sigma)} \text{ as } N \to \infty.$$

In the dense case $N >> m$, combining **K**, **A**, and **O** with an appropriately parameterized $\tilde{I}_{\overline{\Sigma},\varphi,q}$ statistic in the **C** unit, we get an empirical significance test with level of significance approximately equal to τ for the monkey hypothesis by rejecting values of $\tilde{I}_{\overline{\Sigma},\varphi,q}$ which differ from the expected $R(\Sigma)$ by more than ϵ, where $P[|\chi^2_{R(\Sigma)} - R(\Sigma)| \geq \epsilon] = \tau$. Note, that since ϵ is calculated from the asymptotic distribution of the test statistic, it gives reasonable approximation of τ only for large enough sample sizes. The following subsections discuss instances of the general scheme.

3.1 Counting S-Letter Words

We fix a dimension $s \geq 1$ and consider two automata, **A**-n ("n"onoverlapping) and **A**-o ("o"verlapping) with state space $S = \mathcal{A}^s$ of cardinality $m = \alpha^s$ each. Denote a state $\mathbf{a} \in S$ by $\mathbf{a} = (a_1, \ldots, a_s)$, $a_i \in \mathcal{A}$. **A**-n makes a transition from $\mathbf{a}$ to $\mathbf{b} = (b_1, \ldots, b_s)$ if it receives the s-letter word $\mathbf{b}$. **A**-o makes a transition from $\mathbf{a}$ to $\mathbf{b} = (a_2, \ldots, a_s, b)$ if it receives the (single) letter $b \in \mathcal{A}$. The current state thus reflects the last s incoming letters in both units. It is clear, how to write down the corresponding transition matrices $\mathbb{P}$-n and $\mathbb{P}$-o, say. The stable distribution $\mathbf{P}$ equals $\mathbf{P} = (P_{\mathbf{a}})_{\mathbf{a}\in S}$ with components $P_{\mathbf{a}} = \pi_{a_1} \cdot \pi_{a_2} \cdots \pi_{a_s}$ (arranged in lexicographic order) in both cases. We assume that the initial state $X_0 \in S$ is distributed according to this stable distribution.

The observation unit **O** reports the vector $\hat{\mathbf{P}}^{(N)}$ of relative frequencies as described above. The expectation $E[\hat{\mathbf{P}}^{(N)}]$ equals $\mathbf{P}$ for both, **A**-n and **A**-o. In the case of **A**-n, $V[\hat{\mathbf{P}}^{(N)}]$ is the covariance matrix of a multinomial distribution with parameters $\mathbf{P}$ and 1 since the corresponding chain yields an i.i.d. sequence of states. A weak inverse is given by $\overline{\Sigma} = diag(1/P_{(1,\ldots,1)}, \ldots, 1/P_{(\alpha,\ldots,\alpha)})$. For $q = 1$ we get

$$\tilde{I}_{\overline{\Sigma},\varphi,1}(\hat{\mathbf{P}}^{(N)}, \mathbf{P}) = \frac{2N}{\varphi''} \sum_{i=1}^{m} P_i \varphi(\frac{\hat{P}_i^{(N)}}{P_i}), \qquad (1)$$

which is Csiszár's (1963) famous φ-divergence. For the loss function $\varphi_2(u) := \frac{1}{2}(u-1)^2$, $\tilde{I}_{\overline{\Sigma},\varphi_2,1}$ equals the ordinary chi-square statistic, whereas $\varphi_1(u) := 1-u+u\ln(u)$ yields the I-Divergence of Kullback and Leibler (1951), $G^2(\hat{\mathbf{P}}^{(N)}, \mathbf{P}) = 2n \sum_{i=1}^{m} \hat{P}_i^{(N)} \ln(\frac{\hat{P}_i^{(N)}}{P_i})$, which is also called log-likelihood ratio statistic, see Cressie and Read (1984), and Wegenkittl (1998, 1999b). By the above, our setup includes standard serial tests for pseudorandomness.

Now we switch to the automaton **A**-o: Although $\Sigma = V[\hat{\mathbf{P}}^{(N)}]$ has an awfully complicated structure, a rather simple weak inverse $\overline{\Sigma}$ can be given, see Good (1953), Marsaglia (1985), Marsaglia and Zaman (1993), and Wegenkittl (1998). The nice thing about this weak inverse is that it can be written as the sum of two matrices such that for $q = 2$, $\tilde{I}_{\overline{\Sigma},\varphi,q}$ decomposes to the difference of two standard φ-divergences of type (1), one for dimension s and one for dimension $s - 1$. In other words, $\mathbf{P}$ and $\hat{\mathbf{P}}^{(N)}$ are computed for two **A**-n automata in dimensions s and $s-1$, respectively, for the same sequence of letters. $\tilde{I}_{\overline{\Sigma},\varphi,2}$ is asymptotically chi-square distributed with $R(\Sigma) = \alpha^s - \alpha^{s-1}$ degrees of freedom in this case.

By now, our setup also includes the overlapping serial tests in the dense case. We get a connection to entropy based testing if we assume uniform distribution on S, $\mathbf{P} = (\frac{1}{m}, \ldots, \frac{1}{m})$, and consider the linear transform of the statistic $\tilde{I}_{\overline{\Sigma},\varphi_1,2}$ with the loss function φ_1,

$$I_s := \log_2(m) - \frac{\tilde{I}_{\overline{\Sigma},\varphi_1,2}(\hat{\mathbf{P}}^{(N)}, (\frac{1}{m}, \ldots, \frac{1}{m}))}{2N\ln(2)}.$$

I_s is an estimator for the entropy of the sequence of letters under the condition that this sequence stems from a stationary ergodic source of order less than s, see Section 4 and the papers of Wegenkittl (1999a) and L'Ecuyer, Compagner, and Cordeau (1996) for details.

3.2 Gambling Tests

By the latter property of I_s we conclude that it might be important to consider serial tests in high dimensions in order to reveal defects of PRNGs. However, if the state space $S = \mathcal{A}^s$ is getting large, we run into difficulties in storing all the $m = \alpha^s$ components of $\hat{\mathbf{P}}^{(N)}$ in the main memory of the computer, and in performing *dense* tests with the required sample sizes $N >> m$.

Here, the concept of *dimension reduction* by applying a linear transform to the counter vector in the observation unit can be helpful: instead of reporting $\hat{\mathbf{P}}^{(N)}$ itself, the observation unit computes a vector $\hat{\mathbf{P}}^{(N)} \cdot M$ for a real $s \times t$ matrix, $t < s$. This vector is asymptotically normal distributed with mean $\mathbf{P} \cdot M$ and covariance matrix $M^* \Sigma M$, so that we may still apply the modified generalized ϕ-divergence $\tilde{I}_{\overline{\Sigma},\varphi,q}$ in the comparison unit. M might for example be a matrix consisting of zeros and ones only in such a way, that the state space S is partitioned into a few classes which are considered to maintain the relevant information on the suspected defects of a PRNG, and that the relative frequencies of all states within each class are summed up. The resulting vectors have less components and, consequently, are more accessible to the comparison unit.

As an example consider the class of *gambling tests* from Wegenkittl (1998) and Wegenkittl and Matsumoto (1999): The keyboard **K** consists of the letters h and t (head, tail) and the monkey simulates fair coin tosses. In **A** we memorize the last $s-1$ coin throws and bet on h if it had occurred more often than $(s-1)/2$. In this case we can either win or loose depending on the next coin throw. Otherwise, we skip the round. In terms of the automaton unit we have a state space of cardinality 2^s which can be rather intractable in practice, see the examples below. **O** collapses the state space to the 3 dimensional vector denoting wins, losses, and skipped rounds. The test has shown the ability to detect theoretically foreseen deficiencies of two high period generators `T800` (Matsumoto and Kurita 1994) and `ranarray` (Knuth 1997) in dimensions $s = 53, 102$, as well as the improvement by tempering `T800` (yielding the excellent `TT800`) and discarding with high luxury parameter applied to `ranarray`, see again Wegenkittl and Matsumoto (1999).

4 RETURN TIME ANALYSIS

Instead of counting the number of visits of the Markov chain in each possible state, we might consider the return time to the initial state or a function thereof. According statistics have been studied in Choe and Kim (1999) and Maurer (1992). Let $T = \min\{l \geq 1 : X_l = X_0\}$ be the return time to the initial state, and let $E[T|X_0 = i]$ be the expected return time conditional on a start in state $i \in S$. The fundamental relationship to the stable distribution,

$$E[T|X_0 = i] = \frac{1}{P_i}, \quad i \in S, \tag{2}$$

holds in every stationary ergodic chain. By (2) we expect a connection between tests build on the estimation of the return time and the serial tests in the last section, in which $\hat{\mathbf{P}}^{(N)}$ estimates $\mathbf{P}$.

Indeed, the following setup for the logarithm of the return time also incorporates the concept of tests in dimension s: We choose an automaton **A** with state space S and define a return time **O**-r (“r”eturn) unit by letting

$$\hat{T} = \min\{l \geq 1 : (X_{l\cdot s}, X_{l\cdot s+1}, \ldots, X_{l\cdot s+s-1}) =$$

$$= (X_0, X_1, \ldots, X_{s-1})\},$$

note the non-overlapping treatment similar to **A**-n. Let $H(S, \mathbb{P}) = -\sum_{i\in S} P_i \sum_{j\in S} p_{ij} \log_2(p_{ij})$ be the entropy of the chain $(S, \mathbb{P})$, i.e. the entropy of the sequence of states assumed under the monkey hypothesis, and let $H(\mathbf{P}) = -\sum_{i\in S} P_i \log_2(P_i)$ be that of the stable distribution. It turns out (Wegenkittl 1999a) that

$$\lim_{s\to\infty} (E[\log_2(\hat{T})] - (s-1)H(S, \mathbb{P}) - H(\mathbf{P})) = C,$$

with $C = -\gamma/\ln(2) = -0.832746\ldots$, where γ stands for the Euler constant. By this, $I_r := E[\log_2(\hat{T})]/s$ gives an estimator of the entropy $H(S, \mathbb{P})$. To carry out the test, we compute I_r and compare it to the theoretical $H(S, \mathbb{P})$ under the monkey hypothesis in the comparison unit.

As stated in Maurer (1992), such tests are not constructed with a certain defect of a PRNG in mind but with respect to the information theoretic meaning of the per-bit entropy. If a PRNG is used to produce keys for a cryptographic application, the effective key size can be estimated by using Maurer’s (1992) test, which is an optimized version of I_r. The connection to serial testing is established by choosing the trivial automaton **A** with state space $S = \mathcal{A}$ and transition to state $a \in S$ if letter a is received: in this case, I_r and I_s actually estimate the same quantity, namely the entropy of the stationary ergodic source which produces the sequence of letters and which equals $\log_2(\alpha)$ under the monkey hypothesis. In our empirical tests we have found that I_s slightly outperforms I_r because of the use of *overlapping* s-letter words and tends to reject bad PRNGs with smaller sample sizes N. Also, I_s gives an unbiased estimate if the dimension s is larger than the order of the (imaginative) letter source. Return time based tests are asymptotically (for large s) unbiased, but the results are difficult to interpret when s is small.

5 UNIVERSAL AND SPECIFIC TESTS

We have identified two important families of tests which are both built on the notion of turning a letter-driven finite state automaton into a Markov chain and analyzing different figures of merit, such as the relative frequency of each state and the expected logarithm of the return time to the initial state. Both families can be used to estimate the entropy of the letter generating source provided that the order of this source (i.e., the "memory" of the monkey) is smaller than the dimension s of the test. We may call the tests universal in the sense that any deviation of the fair, memoryless monkey is detected for large enough sample size and dimension s. Monkeys with higher memory will be able to hide the correlations in the letter sequence in a way that is inaccessible to the tests.

We identify the following test setups for empirical tests of PRNGs which we quote in increasing order of universality:

- special purpose test: the test is designed to show a presumed defect in the PRNs obtained from a selected family of PRNGs. If the design parameters like α, s, or M are well chosen, the test is highly efficient in finding the corresponding deviation from the monkey hypothesis, but even low-quality generators with other specific defects will pass the test. Consider for instance gambling tests which show the difference between `T800` and `TT800` but fail to reject standard linear congruential generators with bad lattice structure like RANDU in dimension 3. This type of test is used to put alert signs telling potential users that the deterministic origin of the PRNs may actually show up in very simple simulation problems. It also stresses the importance of cross-evaluation of simulations with PRNGs of different type.
- application related test: here, the design parameters are chosen with respect to the target application. The test becomes a toy-version of the simulation problem. Since we know the results in advance, we can rate on the generators performance. Although we consider this an important type of test, it is difficult to give families of tests which are general enough to be adjusted to the specific application. The general serial test presented in Section 3 for example still suffers from the complicated and often infeasible calculation of the covariance matrix Σ. We encourage further investigation of this problem.
- systematic testing: we fix a test and a family of generators and vary both, the design parameters like α, s, and N of the test, and the parameters of the generator such as the period length, modulus, or multiplier. Our aim is to systematically determine the region of rejection in the parameter space so that we shed light on the ability of the generator to hide its deterministic nature under different "pressure". As a result we get rules of thumb on the applicability of the generator depending on e.g. the number of bits and the sample sizes used in an application. Examples are the load-tests in Leeb and Wegenkittl (1997) which show the superiority of inversive– over linear congruential generators in overlapping serial tests, or the sparse case tests in L'Ecuyer, Simard, and Wegenkittl (1998).
- universal tests finally are the intended answer to the question "why do I have to use so many different tests?": although universal tests do exist in the theoretical sense, their practical value is limited by the computational effort that is necessary to do the test in high dimensions s. As we have seen, high s is necessary to reveal all potential correlations hidden in the PRNs.

Nowadays generators can be ordered among a scale of increasing state space. Usual congruential generators are on the lower end of the scale, since the period length is limited by the bit size of the CPU used for the modulus arithmetics. After all, the generation of the next PRN uses all bits in the state space in each step. This operation pretty mixes up the bits if the parameters of the generator are well selected. In higher dimensions, the small period length necessarily imposes limits on the quality of the PRNs. Huge period high-speed generators like `TT800` are on the opposite end. Only a small amount of bits in the state space is used to determine the next PRN. Such generators may suffer from long range correlations although the period length would admit better results in even high dimensions.

A good compromise seems to be given by combined generators (see e.g. L'Ecuyer 1996, 1999a, and 1999b) which use a rather large fraction of the state space in each step, but limit the single operations to small modulus. Such generators can often be viewed as tricky implementations of large period congruential generators.

Empirical testing strives to detect the finite memory (i.e. state space) of the generator by allowing the test to have a finite memory S itself. From the universality argument, we know that large s and $S = \mathcal{A}^s$ will enable certain tests to distinguish between monkey and PRNG. Some generators require additional tricks like carefully selected dimension reduction in order to implement the test for the necessary size of s. The more tricks we need to reveal the "non-monkeyness" the higher our confidence is that an arbitrary application will not implement similar feature extraction and that using the PRNG yields reasonable simulation results.

In summary, rejection by an empirical test gives valuable information on the generator. It is not a lack of quality in the first place since every generator fails in the same number of tests from a theoretical point of view. The important question is, whether the application and the test depend on similar features. Instead of using lots of different tests, we stress the importance of systematic testing by varying the design parameters. Doing such tests gives knowledge on a specific generator, doing cross evaluation with different generators gives knowledge on a specific problem, simulation study, or Monte Carlo algorithm.

ACKNOWLEDGMENT

Research supported by the Austrian Science Foundation (FWF), grant no. P13480-MAT, e-mail : Stefan.Wegenkittl @sbg.ac.at, WWW: http://random.mat.sbg.ac.at.

REFERENCES

Altman, N. S., 1988. Bit–wise behavior of random number generators. *SIAM J. Sci. Stat. Comput.*, **9**(5):941–949.

Bernhofen, L.T., Dudewicz, E.J., Levendovszky, J., and van der Meulen, E.C. 1996. Ranking of the best random number generators via entropy-uniformity theory. *American Journal of Mathematical and Management Sciences*, **16**(1,2):49–88.

Bratley, P., Fox, B., and Schrage, L.E., 1983. *A Guide to Simulation*. Springer-Verlag, New York.

Choe, G.H., and Kim, D.H., 1999. The first return time test of pseudorandom number generators. Submitted for publication.

Cressie, N., and Read, T.. 1984. Multinomial Goodness-of-fit Tests. *J. R. Statist. Soc. B*, **46**(3):440–464.

Csiszár, I., 1963. Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Magyar Tud. Akad. Mat. Kutató Int. Közl*, **8**:85–108.

DeMatteis, A., and Pagnutti, S., 1995. Controlling correlations in parallel Monte Carlo. *Parallel Computing*, **21**:73–84.

Dudewicz, E.J., van der Meulen, E.C., SriRam, M.G., and Teoh, N.K.W., 1995. Entropy-based random number evaluation. *American Journal of Mathematical and Management Sciences*, **15**(1,2):115–153.

Eichenauer-Herrmann, J., Herrmann, E., and Wegenkittl, S., 1997. A survey of quadratic and inversive congruential pseudorandom numbers. In H. Niederreiter, P. Hellekalek, G. Larcher, and P. Zinterhof, editors, *Monte Carlo and Quasi-Monte Carlo Methods 1996*, number 127 in Lecture Notes in Statistics, pages 66–97. Springer, New York.

Entacher, K., Uhl, A., and Wegenkittl, S., 1998. Linear and Inversive Pseudorandom Numbers for Parallel and Distributed Simulation. In *Twelfth Workshop on Parallel and Distributed Simultation* `PADS'98`, May 26th - 29th, 1998, pages 90–97, Banff, Alberta, Canada. IEEE Computer Society, Los Alamitos, California.

Ferrenberg, A. M., Landau, D.P., and Wong, Y.J., 1992. Monte Carlo simulations: hidden errors from “good” random number generators. *Phys. Rev. Lett.*, **69**(23):3382–3384.

Good, I. J., 1953. The serial test for sampling numbers and other tests for randomness. *Proc. Cambridge Philosophical Society*, **49**:276–284.

Hellekalek, P., 1998a. Good random number generators are (not so) easy to find. *Mathematics and Computers in Simulation*, **46**:485–505.

Hellekalek, P., 1998b. On Correlation Analysis of Pseudorandom Numbers. In H. Niederreiter, P. Hellekalek, G. Larcher, and P. Zinterhof, editors, *Monte Carlo and Quasi-Monte Carlo Methods 1996*, volume 127 of *Lecture Notes in Statistics*, pages 251–265. Springer.

Knuth, D. E., 1997. *The Art of Computer Programming*, volume 2: Seminumerical Algorithms. Addison-Wesley, Reading, MA, third edition.

Kullback, S., and Leibler, R.. 1951. On information and sufficiency. *Ann. Math. Statist.*, 22:79–86.

L’Ecuyer, P., 1992. Testing random number generators. In J.J. Swain, D. Goldsman, R.C. Crain, and J.R. Wilson, editors, *Proceedings of the 1992 Winter Simulation Conference*, pages 305–313. IEEE Press, Piscataway, N.J.

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

L’Ecuyer, P., 1996. Maximally equidistributed combined Tausworthe generators. *Mathematics of Computation*, **65**(213):203–213.

L’Ecuyer, P., 1999a. Good parameters and implementations for combined multiple recursive random number generators. *Operations Research*, **47**(1):159–164.

L’Ecuyer, P., 1999b. Tables of Maximally Equidistributed Combined LFSR Generators. *Mathematics of Computation*, **68**(225):261–269.

L’Ecuyer, P., Compagner, A., and Cordeau, J.F., 1996. Entropy tests for random number generators. published electronically at `www.iro.umontreal.ca/~lecuyer`.

L’Ecuyer, P., Simard, R., and Wegenkittl, S., 1998. Sparse serial tests of uniformity for random number generators. Submitted for publication.

Leeb, H., and Wegenkittl, S., 1997. Inversive and linear congruential pseudorandom number generators in empirical tests. *ACM Trans. Modeling and Computer Simulation*, **7**(2):272–286.

Marsaglia, G., 1985. A current view of random number generators. In L. Billard, editor, *Computer Science and Statistics: The Interface*, pages 3–10. Elsevier Science Publishers B.V.

Marsaglia, G., and Zaman, A., 1993. Monkey Tests for Random Number Generators. *Computers Math. Applic.*, 26(9):1–10.

Matsumoto, M., and Kurita, Y., 1994. Twisted GFSR generators II. *ACM Trans. Model. Comput. Simul.*, **4**:254–266.

Maurer, U.M., 1992. A universal statistical test for random bit generators. *J. Cryptology*, **5**:89–105.

Niederreiter, H., 1992. *Random Number Generation and Quasi-Monte Carlo Methods*. SIAM, Philadelphia, USA.

Niederreiter, H., 1995. New developments in uniform pseudorandom number and vector generation. In H. Niederreiter and P. Jau-Shyong Shiue, editors, *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, volume 106 of *Lecture Notes in Statistics*. Springer.

Vattulainen, I., Ala-Nissila, T., and Kankaala, K., 1994. Physical tests for random numbers in simulations. *Physical Review Letters*, 73(19):2513–2516, 11.

Vattulainen, I., Ala-Nissila, T., and Kankaala, K., 1995. Physical models as tests of randomness. *Physical Review E*, 52(3):3205–3213.

Wegenkittl, S., 1998. *Generalized ϕ-Divergence and Frequency Analysis in Markov Chains*. PhD thesis, Universität Salzburg, Österreich. HTML version: `random.mat.sbg.ac.at/˜ste/diss`.

Wegenkittl, S., 1999a. Entropy estimators and serial tests for ergodic chains. Submitted for publication in *IEEE Transactions on Information Theory*.

Wegenkittl, S., 1999b. A generalized ϕ-divergence for asymptotically multivariate normal models. Submitted for publication in *Journal of Multivariate Analysis*.

Wegenkittl, S., and Matsumoto, M., 1999. Getting rid of correlations among pseudorandom numbers: Discarding versus tempering. Submitted for publication in *ACM TOMACS*.

AUTHOR BIOGRAPHY

STEFAN WEGENKITTL is doing research in mathematical statistics (φ-divergence) and metric number theory within the pLAB research group of Prof. Hellekalek at the Department of Mathematics of the University Salzburg, Austria. His work is suported by the Austrian Science Foundation (FWF), grant no. P13480-MAT. Special areas of interest are stochastic processes (dynamic systems and chaos), distance measures for probability measures, stochastic simulation, and empirical tests for pseudorandom number generators, see `http://random.mat.sbg.ac.at/team/`.