Available online at www.sciencedirect.com
SCIENCE DIRECT®
Physica A 356 (2005) 133–138
PHYSICA A
www.elsevier.com/locate/physa

Intensive statistical complexity measure of pseudorandom number generators

H.A. Larrondo[a,*], C.M. González[a], M.T. Martín[b], A. Plastino[b], O.A. Rosso[c]

[a]*Facultad de Ingeniería, Universidad Nacional de Mar del Plata, Juan B. Justo 4302, 7600 Mar del Plata, Argentina*
[b]*Instituto de Física, Universidad Nacional de La Plata, C.C. 727, 1900 La Plata, Argentina*
[c]*Instituto de Cálculo, FCEyN, Universidad de Buenos Aires, Pabellón II, Ciudad Universitaria. 1428 Ciudad de Buenos Aires, Argentina*

Received 26 February 2005
Available online 13 June 2005

Abstract

A Statistical Complexity measure has been recently proposed to quantify the performance of chaotic Pseudorandom number generators (PRNG) (Physica A 354 (2005) 281). Here we revisit this quantifier and introduce two important improvements: (i) consideration of an intensive statistical complexity (Physica A 334 (2004) 119), and (ii) following the prescription of Brand and Pompe (Phys. Rev. Lett. 88 (2002) 174102-1) in evaluating the probability distribution associated with the PRNG. The ensuing new measure is applied to a very well-tested PRNG advanced by Marsaglia.

Keywords: Random number generators; Statistical complexity

*Corresponding author.
E-mail addresses: larrondo@fi.mdp.edu.ar (H.A. Larrondo), cmgonzal@fi.mdp.edu.ar (C.M. González), mtmartin@venus.unlp.edu.ar (M.T. Martín), plastino@venus.unlp.edu.ar (A. Plastino), oarosso@fibertel.com.ar (O.A. Rosso).

0378-4371/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2005.05.025

1. Introduction

Random number generators (RNGs) are essential tools in several fields. They may be based either on physical noise sources or on mathematical algorithms, but in both cases truly random numbers are not obtained. Instead, any real implementation actually produces a Pseudorandom number generator (PRNG). Despite this restriction PRNGs have been developed to fulfil many statistical properties required in applications [1]. A widely employed test-suite, readily available to researchers in academia and industry who wish to analyze their newly developed PRNG, is Marsaglia's *Diehard* one [2]. Of course, a statistical test can never prove that a sequence generated by a PRNG is random (*because it is not!*), but it helps to detect certain kinds of weaknesses that a generator may have. Statistical complexity measures (SCMs) were recently proposed as quantifiers of the degree of physical structure in a signal [3–5]. They are null for total random processes. SCM for a given system's state are products of an entropic measure (H) times a distance to the equilibrium state (Q). In Ref. [6] Q was built up using Wootters' statistical distance and H is the normal Shannon entropy [5]. Regrettably enough, this SCM is neither an intensive nor an extensive quantity, although it yields useful results. A natural SCM improvement is to confer it with an intensive character, as proposed in Ref. [7]. This better SCM is the one be employed here to deal with PRNGs. Additional improvements can be expected if one modifies the manner in which the underlying probability distribution is extracted, by a better consideration of the system's dynamics. For this purpose we follow here the Brandt and Pompe prescription [8]. The bit string generated by a very well-tested PRNG proposed by Marsaglia [2] is tested and we show that the new SCM (C_J) is a stable quantifier when the size of the analyzed sequence increases. The improvement thus found is a factor $\sim$100 over the values obtained with previous procedures.

2. Intensive statistical complexity and its evaluation

The intensive statistical complexity measure is a functional $C_J[P]$ that can primarily be viewed as a quantity that characterizes the probability distribution P associated with the time series generated by the dynamical system under study. It quantifies not only randomness but also the presence of correlational structures [5]. The opposite extremes of perfect order and maximal randomness possess no structure to speak of. In between these two special instances, a wide range of possible degrees of physical structure exist, degrees that should be reflected in the features of the underlying probability distribution. Based on the ideas of Ref. [7] we cast the intensive SCM as

$$C_J[P] = Q_J[P, P_e] \cdot H_S[P]\,, \tag{1}$$

where with the probability distribution P we associate the entropic measure $H_S[P] = S[P]/S_{max}$, with $S_{max} = S[P_e]$ ($0 \leqslant H_S \leqslant 1$). P_e is the equilibrium distribution and S is the Shannon entropy. The disequilibrium Q_J is defined in terms of the

Jensen–Shannon divergence [7] and is given by

$$Q_J[P, P_e] = Q_0\{S[(P + P_e)/2] - S[P]/2 - S[P_e]/2\}. \quad (2)$$

with Q_0 being the normalization constant ($0 \leqslant Q_J \leqslant 1$). Thus, the disequilibrium Q_J is an intensive quantity.

In the evaluation of the probability distribution P associated with the time series (dynamical system) under study we follow the methodology proposed by Brandt and Pompe [8]. We consider partitions of the d-dimensional space revealing the ordinal-structure of a one-dimensional time series. Given the time series $\{x_t, t = 1, \ldots, T\}$ and an embedding dimension $d > 1$, we are interested in "ordinal patterns" of order d [8,9] generated by

$$(s) \mapsto (x_{s-(d-1)}, x_{s-(d-2)}, \ldots, x_{s-1}, x_s)\,, \quad (3)$$

which assign to each time s the d-dimensional vector of values at times $s, s-1, \ldots, s-(d-1)$. Clearly, the greater the d, the more the information on the past provided by our vectors. By the "ordinal pattern" related to the time (s) we mean the permutation $\pi = (r_0, r_1, \ldots, r_{d-1})$ of $(0, 1, \ldots, d-1)$ defined by

$$x_{s-r_{d-1}} \leqslant x_{s-r_{d-2}} \leqslant \cdots \leqslant x_{s-r_1} \leqslant x_{s-r_0}\,. \quad (4)$$

In order to obtain a unique result we set $r_i < r_{i-1}$ if $x_{s-r_i} = x_{s-r_{i-1}}$. Thus, for all the $d!$ possible permutations π of order d, the probability distribution $P = \{p(\pi)\}$ is defined by

$$p(\pi) = \frac{\sharp\{s | s \leqslant T - d + 1; (s), \texttt{has type } \pi\}}{T - d + 1}\,. \quad (5)$$

In this expression the symbol $\sharp$ stands for "number". The normalized entropy and the statistical complexity are then evaluated for this "permutation" probability distribution.

3. Applying C_J measure to a PRNG

The PRNG analyzed here is the one known as "bit.01" obtained from Marsaglia's web site [2]. This random $N_b = 80.000.000$ *bit* stream was obtained by combining two deterministic generators with a physical generator. The physical generator based on Johnson's Noise is used to prevent predictability and it also makes the periods infinite, but it does not pass the Diehard test suite (physical devices do not generate truly random numbers !). The deterministic generator is the sum of other two generators that pass all Diehard tests, *Mother* and *Kiss*, also from Marsaglia [2].

The *Mother* generator is as follows:

$$\begin{aligned} X_n &= Mod\{2111111111 X_{n-4} + 1492 X_{n-3} + 1776 X_{n-2} \\ &\quad + 5115 X_{n-1} + carry_{n-1};\ 2^{32}\}\,, \\ carry_n &= floor\left(\frac{X_n}{2^{32}}\right)\,. \end{aligned} \quad (6)$$

The initial conditions are random. X_n and $carry_n$ are 32-*bit* vectors. The period of *Mother* is about 2^{158}.

The *Kiss* generator is given by

$$X_n = Mod\{X_n^{(1)} + X_n^{(2)} + X_n^{(3)}; 2^{32}\}\,, \tag{7}$$

where

$$\begin{aligned} X_n^{(1)} &= Mod\{69069X_{n-1}^{(1)} + 1; 2^{32}\}\,, \\ X_n^{(2)} &= X_{n-1}^{(2)}(I + L^{13})(I + R^{17})(I + L^5)\,, \\ X_n^{(3)} &= Mod\{2X_{n-1}^{(3)} + X_{n-2}^{(3)} + carry_{(n-1)}; 2^{32}\}\,. \end{aligned} \tag{8}$$

I is the 32×32 identity matrix. L (R) produces a one-step shift to the left (right). L (R) are all 0's except for 1's on the principal subdiagonal (superdiagonal). The initial conditions are random. The period of *Kiss* is about 2^{127}.

Let $\{b_i, i = 1, \ldots, N_b\}$ be the bit string provided by the PRNG. The associated time series was obtained by grouping the bits in non-overlapping m-bit words $\{x_j, j = 1, \ldots, N_w = N_b/m\}$. Each word is an m-bit natural number $\in [0, 2^m]$. The $N_b = 80.000.000$ *bits* of file "bit.01" were processed as 8- and 16-bit words and the corresponding time series for each case was duly obtained. For each time series a *putative attractor* (embedding dimension $d = 6$) was reconstructed. Choosing the coordinates of the attractor as the starting point, the permutation probability distribution associated with the time series was obtained (see previous section) and then H_S and C_J were evaluated.

If the PRNG is an extremely good generator we can expect that "no attractor" will be reconstructed, that is, it will be quite reasonable to obtain a homogeneity cloud of points with a tendency to "fill" the d-dimensional space. Consequently, the associated permutation probability distribution will be $P \approx P_e$ and thus $H_S \approx 1$ and $C_J \approx 0$. In the case of periodic sequences one will have $H_S \approx 0$ and $C_J \approx 0$. Fig. 1 depicts the normalized entropy H_S and the intensive statistical complexity C_J as functions of the number of words. It is clear from this figure that C_J is stable and tends to a very small value when the number of bits tends to infinity. The normalized entropy tends to 1 as the number of words of the analyzed sequence increases. Consequently, the small value of C_J arises from the disequilibrium factor. The present results improve upon those presented in Ref. [6] by a factor of order 100.

4. Conclusions

We showed here that our intensive entropic statistical complexity measure is a stable quantifier of the PRNG randomness. Its value tends to zero for long sequences when good PRNGs are tested. The normalized entropy H_S is verified to be close to unity and the ensuing low C_J-values correspond to randomness. The preliminary

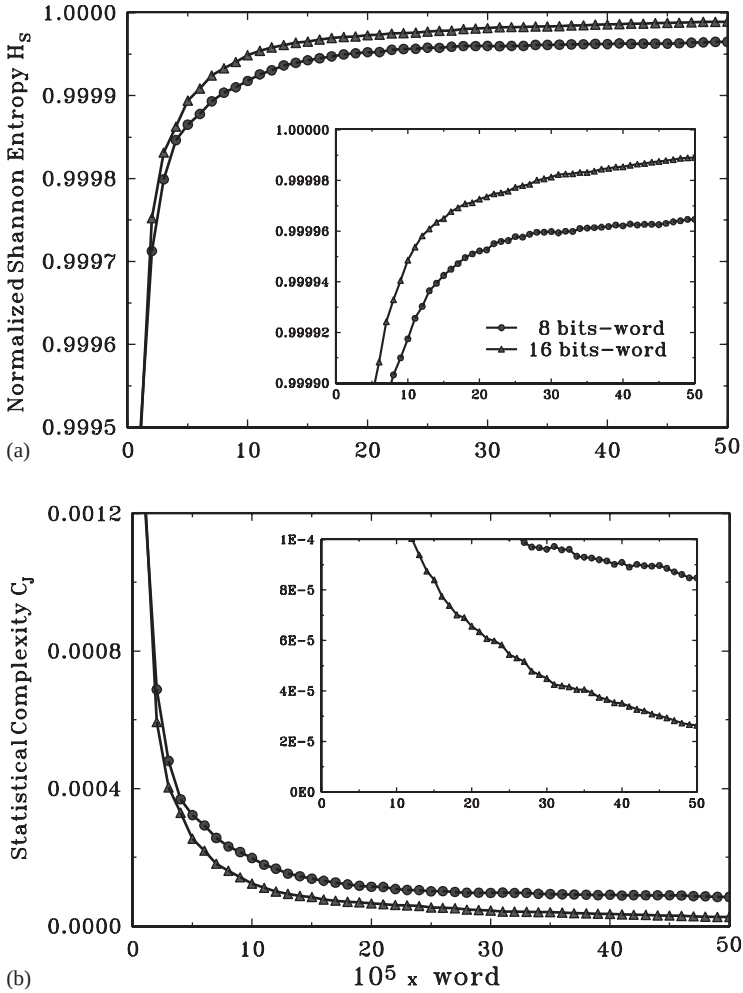

Fig. 1. (a) Normalized permutation entropy H_S and (b) intensive statistical complexity measure C_J for a Marsaglia generator, as functions of the number of words. In both cases words of 8 and 16 bits are considered.

results of new work in progress indicate that our complexity measure may be used to assess the quality of PRNGs different design steps.

Acknowledgements

This work was partially supported by Universidad Nacional de Mar del Plata, ANPCyT (PICTO 11-090076), CONICET(PIP 0029/98) (Argentina). CMG and HAL acknowledge fruitful discussions with Eduardo Boemo.

References

[1] See the excellent web page on random number generation at http://cgm.cs.mcgill.ca/~luc/rng.html.
[2] G. Marsaglia, Diehard Statistical Tests, available online at http://stat.fsu.edu/~geo/diehard.html.
[3] J.S. Shiner, M. Davison, P.T. Landsberg, Phys. Rev. E 59 (1999) 1459.
[4] R. López-Ruiz, H.L. Mancini, X. Calbet, Phys. Lett. A 209 (1995) 321.
[5] M.T. Martin, A. Plastino, O.A. Rosso, Phys. Lett. A 311 (2003) 126.
[6] C.M. González, H.A. Larrondo, O.A. Rosso, Phys. A 354 (2005) 281.
[7] P.W. Lamberti, M.T. Martin, A. Plastino, O.A. Rosso, Phys. A 334 (2004) 119.
[8] C. Brandt, B. Pompe, Phys. Rev. Lett. 88 (2002) 174102–1741021.
[9] K. Keller, H. Lauffer, Int. J. Bifurcation Chaos 13 (2003) 2657.