Available online at www.sciencedirect.com

SCIENCE DIRECT®

Physics Letters A 352 (2006) 421–425

PHYSICS LETTERS A

www.elsevier.com/locate/pla

Random number generators and causality

H.A. Larrondo [a], M.T. Martín [b], C.M. González [a], 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 (IFLP), Facultad de Ciencias Exactas, Universidad Nacional de La Plata and Argentina's National Council (CONICET), C.C. 727, 1900 La Plata, Argentina*

[c] *Chaos & Biology Group, Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Pabellón II, Ciudad Universitaria, 1428 Ciudad de Buenos Aires, Argentina*

Received 8 September 2005; received in revised form 11 November 2005; accepted 3 December 2005

Available online 12 December 2005

Communicated by F. Porcelli

Abstract

We advance a prescription to randomize physical or algorithmic Random Number Generators (RNG's) *that do not pass* Marsaglia's DIEHARD test suite and discuss a special physical quantifier, based on an intensive statistical complexity measure, that is able to adequately assess the improvements produced thereby. Eight RNG's are evaluated and the associated results are compared to those obtained by recourse to Marsaglia's DIEHARD test suite. Our quantifier, which is evaluated using causality arguments, can forecast whether a given RNG will pass the above mentioned test.

PACS: 03.67.-a; 89.70.+c; 03.65.Bz

Keywords: Random number generators; Statistical complexity

1. Introduction

Random Number Generators (RNG) are extensively used in physics and engineering. In point of fact, most computer systems possess RNG's, although not all of them are of good enough quality for statistical purposes. Recourse to special physical devices is also made so as to produce RNG's. It is then appropriate to classify RNG's as either "physical" or "algorithmic" ones. A very important RNG-issue is its adequate testing for specific applications such as encryption schemes and/or sophisticated Monte Carlo simulations. An easily available and very stringent statistical test-suite is that provided by Marsaglia [1]. In his website one finds two impressive contributions to the problem: (a) an RNG obtained by *mixing* algorithmic and physical generators, and (b) a complete RNG-test suite (called "DIEHARD"). The above referred to mixing is required because, quoting Marsaglia, "no physical devices I have considered pass the stringent randomness requirements of my DIEHARD battery of tests. But the deterministic methods do". In fact, he used the algorithmic RNG's to get good statistical properties and the physical devices to make the RNG's unpredictable.

In physics one assumes that a model can be manufactured for any system of interest and, in a sense, unpredictability may be understood as meaning that the model's features are not already known in sufficient detail. Furthermore, a complex dynamics does not imply a complex model if nonlinearity is present; low-dimensional chaotic dynamical systems are nice examples of such an assertion. Consequently, chaotic systems are good candidates to model devices used to produce physical RNG's.

The latest available version of Marsaglia's DIEHARD produces a matrix with about three hundred values as a *Test Summary*. These values are expected to be distributed in a rather close fashion to the uniform distribution in the interval [0, 1).

[*] Corresponding author. Tel./fax: +54 11 4576 3375.
E-mail addresses: larrondo@fi.mdp.edu.ar (H.A. Larrondo), mtmartin@venus.unlp.edu.ar (M.T. Martín), cmgonzal@fi.mdp.edu.ar (C.M. González), plastino@venus.unlp.edu.ar (A. Plastino), oarosso@fibertel.com.ar, rosso@ba.net (O.A. Rosso).

0375-9601/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.physleta.2005.12.009

A global quantifier is also provided by a KStest (Kolmogorov–Smirnov test) for the same interval.

The present effort purports to effect two main contributions to the above, ongoing discussion: (1) a prescription to randomize those physical or algorithmic RNG's *that do fail to pass* the DIEHARD test suite; and (2) advancing *the use of a special physical quantifier*, based on an intensive statistical complexity measure (MPR statistical complexity [2–5]), to assess the improvements achieved by recourse to the prescription of (1) above. This quantifier takes causality effects into account. Eight RNG's are evaluated and the concomitant results are compared to those obtained by recourse to Marsaglia's DIEHARD test suite.

Our work should be of relevance for Monte Carlo simulations [6], cryptography [7], communications theory [8] and some aspects of nanotechnology [9]. From a theoretical point of view it is of interest to point out that we are here linking the concept of Kolmogorov–Chaitin's algorithmic complexity [10] with that of *statistical* complexity. The former is adequately treated according to the so-called Pompe [11] procedure, adopted in this Letter to assign a probability distribution P to a given time series.

2. Methodology

In a recent contribution, López-Ruiz, Mancini and Calbet (LMC) have proposed a statistical complexity measure (SCM) based on the notion of "disequilibrium" as a quantifier of the degree of physical structure in a time series [12]. Given a probability distribution associated with a system's state, the LMC measure is the product of an entropy H times a distance to the uniform-equilibrium state Q. It vanishes for a totally random process. Martín, Plastino and Rosso (MPR) [2] improved on this measure by modifying the distance-component (in the concomitant probability space). In Ref. [2], Q is built-up using Wootters' statistical distance while H is a normalized Shannon-entropy. Regrettably enough, the ensuing statistical complexity measure is neither an intensive nor an extensive quantity, although it yields useful results [3]. Also, a reasonable complexity measure should be able to distinguish among different degrees of periodicity and it should vanish only for periodicity unity. In order to attain such goals any natural improvement should give this statistical measure an intensive character. Lamberti, Martín, Plastino and Rosso [4] obtained a statistical complexity measure (SCM) that is (i) able to grasp essential details of the dynamics, (ii) an intensive quantity, and (iii) capable of discerning among different degrees of periodicity and chaos. Such complexity measure is the one to be employed here to deal with RNG's. It has been shown in Refs. [3,4] that the MPR intensive statistical complexity measure provides one with more detailed information than the one obtained using just Shannon's entropy, which may confuse high degree of chaoticity with randomness.

Evaluation of the probability distribution P associated to a dynamical system or time series under study is a physical problem. Additional improvements can be expected if the underlying probability distribution is "extracted" by more appropriate consideration regarding causal effects in the system's dynamics.

The essence of symbolic dynamics is to associate a symbol sequence with each trajectory of a continuous or discrete dynamical system, by means of a suitable partition of the state-space. This process is described in the context of a *delay-embedding* of the time series into a d-dimensional space in Ref. [13]. Special generating partitions yield in the limit for a fine resolution the Kolmogorov–Sinai entropy. But these partitions are very difficult to ascertain even in the case of two-dimensional systems. Bandt and Pompe [11] advanced a method that "naturally" determines the adequate symbol sequence from the time series' values, without further model assumptions. They determine partitions of the state-space given by comparison of neighboring series' values. For any given series they look for certain *ordinal patterns* of order d. From the symbol occurrence frequency, they deduce a *permutation* probability distribution [11,14,15]. The advantages of Bandt and Pompe's method reside in (i) its simplicity, (ii) extremely fast calculation-process, (iii) robustness, and (iv) invariance with respect to nonlinear monotonous transformations. Using Kolmogorov–Chaitin's algorithmic complexity is another recourse that could be taken advantage of to overcome these problems, although this poses is much more difficult task.

All the RNG's assessed in this Letter are deterministic but *some* of them come from "discretised" chaotic differential equations that may be thought of as models for real physical processes (physical RNG's) while *others* come from recurrence rules (algorithmic RNG's). In order to convert any of them into an electronically realizable RNG, the following scheme is to be applied. (Step 1): a discretising process followed by a biased, scaling transformation that transforms our RNG-"signal" into natural numbers belonging to the interval $[0, 2^n - 1]$; after this step, each random number can be regarded as an n-bit word. (Step 2): a bit stream is assigned to each word. The length of this bit stream can be selected in different ways, the simplest one being to use all the bits of each word (ALL version). By generating five-million 16-bits-words we obtain an 80 million bit stream. We demonstrate below that this procedure yields poor results. It is much better to store just a portion of each word. In this Letter we follow, two strategies: (a) we use, for each n-bit-word's, only the *most significative* bit[1] (MSB version) to generate the bit stream. This is equivalent to the standard symbolic dynamic procedure of assigning a "1" if the number belongs to the range $[2^{n-1}, 2^n - 1]$ and a "0" if it lies within $[0, 2^{n-1} - 1]$. (b) Pick up, for each n-bit-word's, only the *least significative* bit (see footnote 1) (LSB version) to generate the bit stream. This bit represents a small perturbation and our results show that option LSB is the best one because it eliminates low frequency components of the Fourier spectrum. The bit streams obtained with the above described procedures (ALL, MSB, and LSB) are grouped again into m-bit-words and the MPR intensive statistical complexity measure [4] is now eval-

[1] Most (least) significative in the sense of most (least) important. Not to be confuse with the statistical significance.