Condensed Matter Physics 2006, Vol. 9, No 2(46), pp. 367–372

A recipe for an unpredictable random number generator

M.A.García-Ñustes, L.Trujillo, J.A.González

Centro de Física, Instituto Venezolano de Investigaciones Científicas (IVIC),
A.P 21827, Caracas 1020–A, Venezuela

Received February 27, 2006, in final form April 19, 2006

In this work we present a model for computing the random processes in digital computers which solves the problem of periodic sequences and hidden errors produced by correlations. We show that systems with non-invertible non-linearities can produce unpredictable sequences of independent random numbers. We illustrate our results with some numerical calculations related to random walks simulations.

Key words: *random number generator, independent random numbers*

PACS: *05.10.-a, 05.40.a, 05.40.Fb, 07.05.Tp*

Many challenging problems in computational physics are associated with reliable realizations of randomness (e.g. Monte Carlo simulations). In a typical 32-bit format a maximum of 2^{32} floating point numbers can be represented. Therefore, a recursive function $X_{n+1}=f(X_n, X_{n-1}, \ldots, X_{n-r+1})$ acting on these numbers generates a sequence $X_0, X_1, X_2, \ldots, X_{N-1}$ which should repeat itself. It is known that for any recursive function, a digital computer can only generate periodic sequences of numbers [1–7]. These generators are not unpredictable.

Definition of truly unpredictable process: The next values are not determined by the previous values. A process $X_n = P(\theta T Z^n)$ is said to be unpredictable if for any string of values $X_0, X_1, X_2, \ldots, X_m$ of length $m+1$, generated using some $\theta = \theta_1$, there are other values of θ for which function $X_n = P(\theta T Z^n)$ generates exactly the same string of numbers $X_0, X_1, X_2, \ldots, X_m$, but the next value X_{m+1} is different, where m is any integer. Note that this kind of process cannot be expressed as a map of type $X_{n+1} = f(X_n, X_{n-1}, \ldots, X_{n-r+1})$ [8–10].

All the known generators (in some specific physical calculations) give rise to incorrect results because they deviate from randomness [2,4,5,7]. It is trivial that any periodic process is not unpredictable. Suppose that m_T is the period of the sequence generated. Given any string of m_T values: $X_s, X_{s+1}, \ldots, X_{m_T-1}$; the next value X_{m_T} is always known because the process is periodic. On the other hand, for any generator of type $X_{n+1} = f(X_n, X_{n-1}, \ldots, X_{n-r+1})$, given any string of r values: $X_s, X_{s+1}, \ldots, X_{s+r-1}$; the next value X_{s+r} is always determined by the previous r values. Thus it is not unpredictable. So the subsequences must be correlated.

An example of this can be found in [5], where the authors have shown that using common pseudo random number generators, the produced random walks present symmetries, meaning that the generated numbers are not independent. On the other hand, the logarithmic plot of the mean distance $\langle d \rangle$ versus the number of steps N is not a straight line (as expected theoretically, $\langle d \rangle \sim N^{1/2}$) after $N > 10^5$ (in fact, it is a rapidly decaying function). Here d is defined as the end to end mean-square distance from the origin of the random walk as a function of the number of steps. Other papers on the effect of the pseudorandom number generator on random walk simulations are as follows [11–13].

In the following, we shall show that using non-invertible nonlinear functions, we can create an unpredictable random number generator which does not contain visible correlations while simulating a random walk with the length 10^9.

Let us investigate the following function[8–10]:

$$X_n = P(\theta\, T Z^n), \tag{1}$$

© M.A.García-Ñustes, L.Trujillo, J.A.González

where $P(t)$ is a periodic function, θ is a real number, T is the period of the function $P(t)$, and Z is a noninteger real number.

Let Z be a rational number expressed as $Z = p/q$, where p and q are relative prime numbers. Now let us define the following family of sequences

$$X_n^{(k,m,s)} = P\left[T\left(\theta_0 + q^m k\right)\left(\frac{q}{p}\right)^s\left(\frac{p}{q}\right)^n\right], \tag{2}$$

where k, m and s are non-negative integers. Parameter k distinguishes different sequences. For all sequences parametrized by k, the strings of $m+1$ values $X_s, X_{s+1}, X_{s+2}, \ldots, X_{s+m}$ are the same. This is because

$$X_n^{(k,m,s)} = P\left[T\theta_0\left(\frac{q}{p}\right)^s\left(\frac{p}{q}\right)^n + Tkp^{n-s}q^{(m-n+s)}\right] = P\left[T\theta_0\left(\frac{q}{p}\right)^s\left(\frac{p}{q}\right)^n\right]$$

for all $s \leqslant n \leqslant m+s$. Note that the number $k\, p^{(n-s)}q^{(m-n+s)}$ is an integer for $s \leqslant n \leqslant m+s$. So we can have an infinite number of sequences that share the same string of $m+1$ values.

Nevertheless, the next value

$$X_{s+1}^{(k,m,s)} = P\left[T\ \theta_0\left(\frac{q}{p}\right)^s\left(\frac{p}{q}\right)^{(s+1)} + \frac{T\ kp^{(m+1)}}{q}\right]$$

is uncertain. In general $X_{s+1}^{(k,m,s)}$ can take q different values. In addition, the value $X_{s-1}^{(k,m,s)}$,

$$X_{s-1}^{(k,m,s)} = P\left[T\ \theta_0\left(\frac{q}{p}\right)^s\left(\frac{p}{q}\right)^{(s-1)} + \frac{T\ kq^{(m+1)}}{p}\right],$$

is also undetermined from the values of the string $X_s, X_{s+1}, X_{s+2}, \ldots, X_{s+m}$. There can be p different possible values. In the case of a generic irrational Z, there are infinite possibilities for the future and for the past. From the observation of the string $X_s, X_{s+1}, X_{s+2}, \ldots, X_{s+m}$, there is no method for determining the next and the previous values of the sequence.

But this is not the only feature of these functions. It can be shown that there are no statistical correlations between X_m and X_n if $m \neq n$, and that they are also independent in the sense that their probability densities satisfy the relationship $P(X_n, X_m) = P(X_n)P(X_m)$ [14,15].

Moreover, we shall show that, given the function (1), any string of sequences $X_s, X_{s+1}, \ldots, X_{s+r}$ constitutes a set of statistically independent random variables.

Without loss of generality, we assume that $P(t)$ has zero mean and can be expressed using the following Fourier representation $P(t) = \sum_{k=-\infty}^{\infty} a_k \mathrm{e}^{\mathrm{i}\pi k t}$.

We can calculate the r-order correlation functions [14,15]:

$$\begin{aligned}
E(X_{n_1}\cdots X_{n_r}) &= \int_X \mathrm{d}\theta P(T\theta Z^{n_1})\cdots P(T\theta Z^{n_r}) \\
&= \sum_{k_1=-\infty}^{\infty}\cdots\sum_{k_r=-\infty}^{\infty} a_{k_1}\cdots a_{k_r}\int_0^1 \mathrm{d}\theta \exp\{\mathrm{i}\pi(k_1 Z^{n_1} + \cdots + k_r Z^{n_r})T\theta\} \\
&= \sum_{k_1=-\infty}^{\infty}\cdots\sum_{k_r=-\infty}^{\infty} a_{k_1}\cdots a_{k_r}\delta(k_1 Z^{n_1} + \cdots + k_r Z^{n_r}, 0),
\end{aligned} \tag{3}$$

where the coefficients k_i can be different integers, and $\delta(n,m) = 1$ if $n = m$ or $\delta(n,m) = 0$ if $n \neq m$.

When all n_i are even, the following equation is satisfied

$$E(X_s^{n_1}X_{s+1}^{n_2}\cdots X_{s+r}^{n_r}) = E(X_s^{n_1})E(X_{s+1}^{n_2})\cdots E(X_{s+r}^{n_r}). \tag{4}$$

The main problem in this equation is when one of the numbers n_i is odd. In this case, the correlations $E(X_s^{n_1}X_{s+1}^{n_2}\cdots X_{s+r}^{n_r})$ should be zero. A nonzero correlation in equation (4) exists only for the sets $(n_1, n_2, \ldots, n_r)$ that satisfy the equation $k_1 Z^{n_1} + \cdots + k_r Z^{n_r} = 0$. For a typical real number Z, this equation is never satisfied.

If we use non-invertible nonlinear functions, type of (1), we can implement a Truly Random Number Generator (TRNG). In this case, we propose the following function

$$X_n = [\theta_s Z^n] \mod 1. \tag{5}$$

Function (5) is an example of the general case $X_n = P[\theta T Z^n]$ studied in this paper. We have shown that the subsequences $X_s, X_{s+1}, \ldots, X_{s+r}$ constitute a set of statistically independent random variables. The particular case of function (5) is well-known to produce uniformly distributed numbers [8–10].

Now we shall formulate a central limit theorem. Using theorems proved in previous studies [14–19] and the results obtained from this paper, we obtain the following formula: If Z is a generic real number and $X_n = 2(Y_n - 1/2)$, $Y_n = [\theta Z^n] \mod 1$, then

$$\lim_{r\to\infty} P\left\{\alpha < \frac{X_1 + X_2 + \cdots X_r}{\sqrt{r}} < \beta\right\} = \frac{1}{\sqrt{\pi}} \int_\alpha^\beta \mathrm{e}^{-\xi^2} \mathrm{d}\xi. \tag{6}$$

The Gaussian distribution of the sums is correct even for other functions $X_n = P[\theta Z^n]$, where $P(t)$ is periodic. This has been shown in numerical simulations [14].

The numbers $X_n = [\theta Z^n] \mod 1$ are uniformly distributed[8–10]. We can simulate different stochastic processes (with different distributions) using different functions $X_n = P[\theta T Z^n]$. As $\rho(X_n) = 1$, $\rho(X_{n+1}) = 1$, $\rho(X_n, X_{n+1}) = 1$, it is trivial that they are independent.

It is interesting to check the theoretical predictions using numerical simulations of the behavior of different stochastic processes.

For instance, let us study the function

$$U_n = \cos[2\pi\theta Z^n]. \tag{7}$$

All the moments and higher-order correlations can be exactly calculated [14,15]:

For odd m:

$$E(U_n^m) = 0. \tag{8}$$

If any n_i is odd, then

$$E(U_s^{n_0} U_{s+1}^{n_1} \cdots U_{s+r}^{n_r}) = 0. \tag{9}$$

Suppose now that all n_i are even:

$$\begin{aligned}
E(U_s^{n_0} U_{s+1}^{n_1} \cdots U_{s+r}^{n_r}) &= 2^{-(n_0+n_1+\cdots+n_r)} \begin{pmatrix} n_0 \\ \frac{n_0}{2} \end{pmatrix} \begin{pmatrix} n_1 \\ \frac{n_1}{2} \end{pmatrix} \cdots \begin{pmatrix} n_r \\ \frac{n_r}{2} \end{pmatrix}, && (10)\\
E(U_s^{n_0}) &= 2^{-n_0} \begin{pmatrix} n_0 \\ \frac{n_0}{2} \end{pmatrix}, && (11)\\
E(U_{s+1}^{n_1}) &= 2^{-n_1} \begin{pmatrix} n_1 \\ \frac{n_1}{2} \end{pmatrix}, \ldots, && (12)\\
E(U_{s+r}^{n_r}) &= 2^{-n_r} \begin{pmatrix} n_r \\ \frac{n_r}{2} \end{pmatrix}. && (13)
\end{aligned}$$

Note that the condition for independence is satisfied

$$E(U_s^{n_0} U_{s+1}^{n_1} \cdots U_{s+r}^{n_r}) = E(U_s^{n_0}) E(U_{s+1}^{n_1}) \cdots E(U_{s+r}^{n_r}), \tag{14}$$

for all integers $n_0, n_1, \ldots n_r$.

We have performed extensive numerical simulations that confirm the values of these moments and the independent conditions.

An additional checking is as follows.

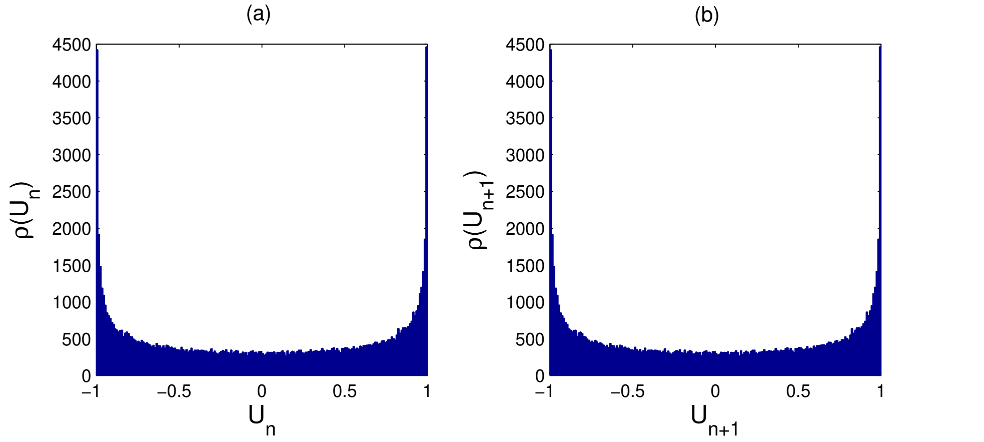

Figure 1. Probability densities for random variables $U_n = \cos[2\pi\theta Z^n]$ and $V_n = U_{n+1}$. (a) $\rho(U) = (\pi\sqrt{1-U^2})^{-1}$; (b) $\rho(V) = (\pi\sqrt{1-V^2})^{-1}$.

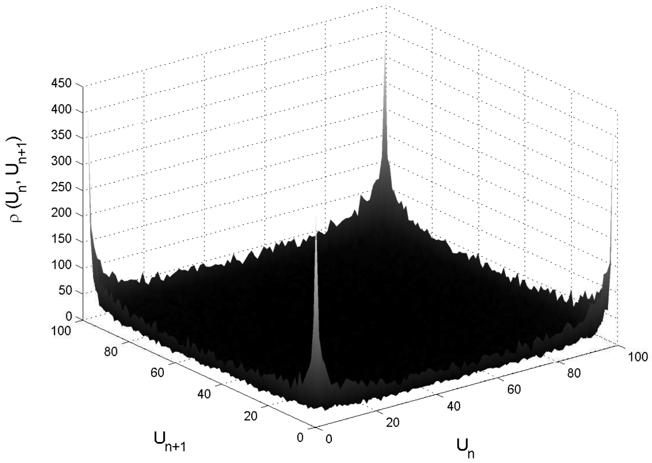

Figure 2. Probability density $\rho(U_n, U_{n+1})$, when $U_n = \cos[2\pi\theta Z^n]$. Here $\rho(U,V) = (\pi^2\sqrt{(1-U^2)(1-V^2)})^{-1}$.

The probability density of U_n is $\rho(U) = (\pi\sqrt{1-U^2})^{-1}$. Define $V_n = U_{n+1}$. The probability density of V_n is $\rho(V) = (\pi\sqrt{1-V^2})^{-1}$. We have checked both theoretically and numerically that $\rho(U,V) = (\pi^2\sqrt{(1-U^2)(1-V^2)})^{-1}$, that is $\rho(U,V) = \rho(U)\rho(V)$. This can be observed in figure 1 and figure 2. In order to avoid computation problems, we have used the following procedure. We change parameter θ after each set of M values of X_n, where M is the maximum number for which there are no overflow problems, such that the next value of X_{n+1} is obtained with the new θ. Let us define

$$\theta_s = A(C_s + X_s) + 0.1, \tag{15}$$

where C_s is a sequence obtained using the digits of the Champernowne's number [20] (i.e., $0.1234567891011\ldots$): $C_0 = 0.123456$, $C_1 = 0.234567$, $C_2 = 0.345678$, $C_3 = 0.456789$, $C_4 = 0.567891$, and so on. This sequence is nonperiodic. Index s is the order number of θ, such that $s = 0$ corresponds to the θ used for the first set of M sequence values $X_1, X_2, \ldots, X_M$; $s = 1$ for the second set $X_{M+1}, X_{M+2}, \ldots, X_{2M}$, and so on. X_0 represents the TRNG's seed.

Using this method we have generated a very long sequence of random numbers without computational problems.

To test function (5) as a truly random number generator, we have implemented a random walk simulation program in C++. We have made a sampling test of a random walk with $N = 10^9$ steps with 100 realizations with different initial seeds. The mean distance $\langle d \rangle$ was being calculated every 1000 steps of the random walk.

The Champernowne sequence of numbers used in the generator was produced previously by a short C++ program, who created a sequence of a maximum of 40000 Champernowne's numbers. If a larger amount of values to C_s is necessary, it can be obtained using a segment code that has 40 thousand values already stored in C_s and mixing them, e.g. the algorithm takes the first value of the series C_1, the third C_3 and so on, and adds them at the end of the series, obtaining $C_{s+1} = C_1$, $C_{s+2} = C_3, \ldots$; if more values are necessary, this procedure or cycle is repeated but now skipping two values C_1, C_3, $C_5, \ldots$ three values C_1, C_4, C_7 and so on. In this way, we can make the C_s sequence as large as we wish.

We present a logarithmic plot of the mean distance $\langle d \rangle$ versus the number of steps N with $N = 10^9$ steps with $A = 6.9109366$ and $Z = \pi/2$ (See figure 3). We can verify that there is no deviation from the theoretical straight line, even for $N \gg 10^5$ steps, which is a very good test of the reliability of the Random Number Generator used in the random walk simulations.

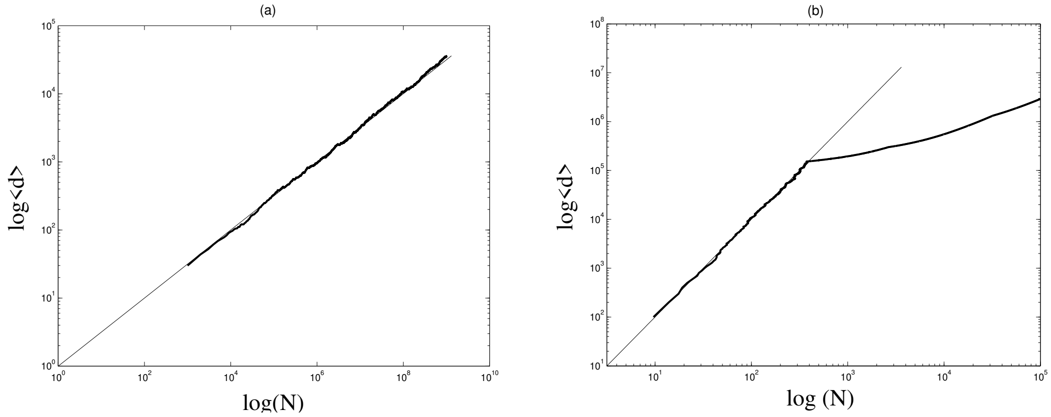

Figure 3. Logarithmic plot of the mean distance $\langle d \rangle$ versus the number of steps $N = 10^9$ steps. (a) for generator (5); (b) the same simulation for a generator of type $X_{n+1} = aX_n \mod T$.

We have presented a random number generator based on the properties of non-invertible transformations of truncated exponential functions. The obtained random process is unpredictable in the sense that the next values are not determined by the previous values. We have applied this generator to the numerical simulation of statistically independent random variables. In the simulation of a random walk with the length 10^9, the random process does not contain visible correlations.

References

1. Ferrenberg A.M., Landau D.P., Wong Y.J., Phys. Rev. Lett., 1992, **69**, 3382.
2. Grassberger P., Phys. Lett. A, 1993, **181**, 43.
3. Vattulainen I., Ala-Nissila T., Kankaala K., Phys. Rev. Lett., 1994, **73**, 2513.
4. D'Souza R.M., Bar-Yam Y., Kardar M., Phys. Rev. E, 1998, **57**, 5044.
5. Nogués J., Costa-Krämer J.L., Rao K.V., Physica A, 1998, **250**, 327.
6. L'Ecuyer P., Oper. Res., 1999, **47**, 159.
7. Bauke H., Mertens S., J. Stat. Phys., 2004, **114**, 1149.
8. González J.A., Reyes L.I., Suárez J.J., Guerrero L.E., Gutiérrez G., Phys. Lett. A, 2002, **295**, 25.
9. González J.A., Reyes L.I., Suárez J.J., Guerrero L.E., Gutiérrez G., Physica D, 2003, **178**, 26.
10. Trujillo L., Suárez J.J., González J.A., Europhys. Lett., 2004, **66**, 638.
11. Grassberger P., J. Phys. A: Math. Gen., 1993, **26**, 2769.
12. Shchur L.N., Heringa J.R., Blöte H.W.J., Physica A, 1997, **241**, 579.
13. Shchur L.N., Comput. Phys. Comm., 1999, **121**, 83.

14. González J.A., Trujillo L., Acta Physica Pol. B, 2005, **37**, 2394.
15. González J.A., Trujillo L., J. Phys. Soc. Japan, 2006, **75**, 023002.
16. Kac M., Stud. Mathematica, 1936, **6** , 46.
17. Kac M., Steinhaus H., Stud. Mathematica, 1936, **6**, 59.
18. Kac M., Steinhaus H., Stud. Mathematica, 1936, **6**, 89.
19. Kac M., Steinhaus H., Stud. Mathematica, 1937, **7**, 1.
20. Champernowne D.G., J. London Math. Soc., 1933, **8**, 254.

Припис для непередбачного генератора випадкових чисел

М.А.Гарсіа-Нюстес, Л.Трухійо, Х.А.Гонсалес

Центр фізики, Венесуельський Інститут наукових досліджень,
пошт. скринька 21827, Каракас 1020–А, Венесуела

Отримано 27 лютого 2006 р., в остаточному вигляді – 19 квітня 2006 р.

В цій роботі ми пропонуємо модель випадкових процесів в числових комп'ютерах, яка вирішує проблему спричинених кореляціями періодичних послідовностей і прихованих помилок. Ми показуємо, що системи з незворотними нелінійностями можуть породжувати непередбачувані послідовності незалежних випадкових чисел. Наші результати ілюструються обчисленнями, пов'язаними з симуляцією випадкових блукань.

Ключові слова: *генератор випадкових чисел, незалежні випадкові числа*

PACS: *05.10.-a, 05.40.a, 05.40.Fb, 07.05.Tp*