The Testing of Pseudorandom Generators Using Correlation Properties

Vladimír ŠEBESTA
Department of Radioelectronics
and
Pavel KILIÁN
Department of Biomedical Engineering
Faculty of Electrical Engineering
Technical University of Brno
Antonínská 1, 662 09 Brno
ČSFR

Abstract

The statistical properties of a pseudorandom sequence being characterized by the distribution of the values near to the uniform distribution may differ from the expected ones. On doing a quantitative evaluation of the differences, it is advantageous to start with the properties of an ideally random sequence. The results of the study of the absolute value of autocovariance sequence point show that the estimate is characterized by the unilateral normal distribution with the parameters dependent on the sequence length. Thanks to this it is possible to determine the criteria that enable the relatively simple testing of the behaviour of pseudorandom sequence generators. These criteria have been verified on the four generators.

It is often necessary to choose an appropriate pseudorandom generator for a given application. The authors of this article have shown how to quantify the correlation properties of a generator. As a means of achieving this, they used the absolute value of the deviation from zero of the estimated value of the covariance function. These properties are considered in this paper. The criteria devised here have been applied to determine the qualities of four generators.

1. Introduction

The occurence of new types of tasks for digital signal processing, as well as the advancing development of software, often set us the problem of testing and choosing a pseudorandom generator. The correlation properties of a pseudorandom generator, represent one of the features for the choice.

The autocovariance sequence of the stationary sequence of the random numbers which are perfectly independent of each other, is equal to zero under any circumstance, except for the initial value. However, the estimate of the covariance sequence, determined from the section of the final length of random numbers, involves various values not equal to zero thanks to the statistical indeterminateness of the estimates. The statistical characteristics of the estimate may serve as being relevant in evaluating the properties of the sequences of numbers obtained by means of the pseudorandom numbers generators.

Let us suppose the estimate $\hat{K}(m)$ of the covariance sequence $K(m)$ in the form

$$\hat{K}(m) = \hat{R}(m) - \hat{\mu}_x^2 \tag{1}$$

where $\hat{R}(m)$ is the estimate of the autocorrelation sequence and $\hat{\mu}_x$ is the estimate of the mean value of the sequence. The introduced variables are determined by the following relations

$$\hat{R}(m) = \frac{1}{N}\sum_{n=1}^{N} x(n)\,x(n+m) \tag{2}$$

where $x(n)$ are the values of random numbers and N is the length of the sequence of random numbers,

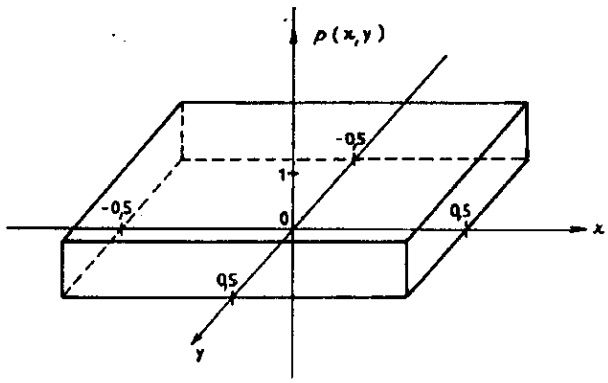

Fig.1, The probability density function of the variables x and y

$$\hat{\mu} = \frac{1}{N}\sum_{n=1}^{N} x(n) \tag{3}$$

2. An Ideal Generator

We will consider an ideal generator, for which the random variables $x(n)$ and $x(n+m)$ where $m \neq 0$ are independent, being distributed uniformly over the interval $<-0.5; 0.5>$. In the case considered $K(m) = 0$ for $m \neq 0$, but as a result of the finiteness of the number N, $\hat{K}(m)$ is not generally equal to zero. The object of our consideration will be to estimate the behaviour of the variable $|\hat{K}(m)|$.

By certain considerations and calculations it has been shown, that the variable $\hat{\mu}_x^2$ in relation (1) has a χ^2 distribution with dispersion $1/(6N^2)$ and with mean value $1/(12N)$. It can be shown, that for the case $N > 240$ it is

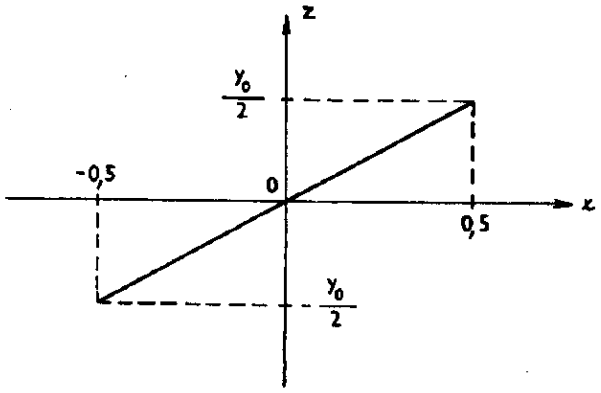

Fig. 2, The dependence of the variable z on x when y = y_0 and $y_0 > 0$.

possible to neglect $\hat{\mu}_x^2$ with respect to $\hat{K}(m)$. This permits the study of $\hat{R}(m)$ instead of $\hat{K}(m)$.

We will first introduce the random variable

$$z(n) = x(n)\,x(n+m) = x(n)\,y(n) \tag{4}$$

where $y(n)$ denotes $x(n+m)$.

See Fig.3 for the conditional probability density function $p(z|y)$ for the case $y = y_0, y_0 > 0$. For the probability density function $p(z)$ we may then write the following:

$$p(z) = \int_{-\infty}^{\infty} p(y)\,p(z|y)\,dy =$$

$$= \int_{-0.5}^{-2z} -\frac{1}{y}dy + \int_{2z}^{0.5} \frac{1}{y}dy \tag{5}$$

After integration we get

$$p(z) = -2\,ln(4|z|); \quad \text{for } |z| \le 0.25$$

and

$$p(z) = 0; \qquad \text{for } |z| > 0.25 \tag{6}$$

Applying the relation which has just been derived, it is possible to demonstrate that the dispersion D_z of the random variable z equals $1/144$ and the mean value $\mu_z = 0$.

We know that $z(n)$ and $z(n+i)$ are statistically independent variables, except for the case $i = m$, when they are only uncorrelated. Thanks to this fact, it is possible, if $m > 0$, for the dispersion D of the variable $\hat{R}(m)$, to write the following:

$$D = \frac{1}{N^2} N D_z = \frac{1}{144N} \tag{7}$$

Similarly, the mean value μ of the variable $\hat{R}(m)$, is zero.

Taking into account that $\hat{R}(m)$ is, essentially, the sum of a large number of random variables, the distribution of the random variable $\hat{R}(m)$ will be practically gaussian. If $m > 0$ the mean value of the distribution will be zero and the value of the dispersion will be $1/144N$. If $N > 240$ it will be applicable even for $\hat{K}(m)$.

The variable $|\hat{K}(m)|$ will be characterized by the unilateral gaussian distribution. The relation of the unilateral distribution to the original gaussian distribution is shown e.g. in [5]. For the mean value we get

$$a = \sqrt{\frac{2D}{\pi}} = \sqrt{\frac{2}{144N\pi}} = \sqrt{\frac{1}{72\pi}} \tag{8}$$

This is then the expected value of the absolute value of $\hat{K}(m)$ when testing an ideal source of a sequence of random numbers being uniformly distributed over the interval $< -0.5; 0.5 >$.

The standard deviation of the variable $|\hat{K}(m)|$ is

$$\sigma = \sqrt{D\left(1 - \frac{2}{\pi}\right)} = \frac{1}{12}\sqrt{\frac{1}{N}\left(1 - \frac{2}{\pi}\right)} \tag{9}$$

For a 90% confidence interval we have derived the lower bound

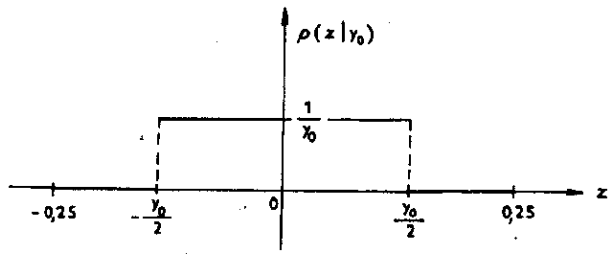

Fig.3, The conditional probability density function

$$K_L = \frac{5.22\ 10^{-3}}{\sqrt{N}} \tag{10}$$

and the upper

$$K_H = \frac{1.63\ 10^{-1}}{\sqrt{N}} \tag{11}$$

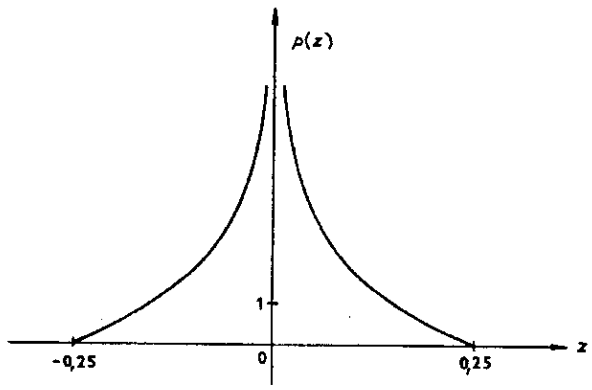

Fig 4, The probability density function of the variable z

3. A Real Generator

In pseudorandom generators the estimates of the mean value and those of the standard deviation of the absolute values of the covariance function will differ from the ex-

pected ones as determined by relations (8) and (9). This circumstance may become the basis of a number of criteria enabling one to judge whether the pseudorandom sequence in the section considered behaves according to our anticipations or not.

We are going to choose the following procedure. We will realize the pseudorandom sequence and estimate the covariance sequence values for $m > 0$. Then we will form the sequence of the absolute values of the covariance sequence and determine the mean value as well as the standard deviation of this sequence. We will compare these two parameters with the values derived for the ideal random sequence. We will also determine the number of cases in which the absolute values of the covariance sequence occur within the confidence interval and we will compare the determined number with the expected one.

4. The Results of the Experiments

The results of the tests on four pseudorandom generators are shown in tables 1 and 2. On individual lines one can find (a) the averages of the estimates of the absolute values, of the 99 or 511 autocovariance sequence elements, (b) the

Table 1
N = 132000, m = 1 ... 99

	gen. A	gen. B	gen. C	gen. D	expect ed
a	1.78E-4	3.95E-4	1.74E-4	1.65E-4	1.83E-4
σ	1.40E-4	6.24E-4	1.47E-4	1.29E-4	1.38E-4
a/σ	1.27	0.63	1.18	1.28	1.32
α	90	74	86	94	89.1

standard deviation of these estimates, (c) the ratio of the average and the standard deviation and, finally, (d) the number of cases, in which the estimate of the absolute value of the autocovariance sequence element coincided with the confidence interval.

The results from four different generators of pseudorandom numbers are shown in the four columns of the tables 1 and 2. The theoretically expected values of an ideal generator can be found in the last column.

Table 2
N = 512, m = 1 ... 511

	gen. A	gen. B	gen. C	gen. D	expect ed
a	3.44E-3	6.07E-3	3.35E-3	3.68E-3	2.94E-3
σ	2.47E-3	4.36E-3	2.52E-3	2.68E-3	2.22E-3
a/σ	1.39	1.39	1.33	1.37	1.32
α	443	327	446	432	459.5

Generator A is the generator described in [3], page 187. The starting value was 1. Generator B is the generator described in [6], page 266. The starting value was 1. Generator C is the generator built-in the Turbo Pascal v.5.5 of the firm Borland, the starting value was 0. Generator D is the generator described in [2], appendix A. The starting value was 12357.

5. Conclusions

The authors of this paper have tried to provide workers in the field a relatively simple instrument for testing the correlation properties of the pseudorandom sequences sections.

The distribution of the variable $x(n)\,x(n+m)$ has been determined for an ideal random sequence, characterized by the uniform distribution over the interval $< -0.5; 0.5 >$. Next the distribution and the parameters of the estimates of the autocorrelation sequence elements $R(m)$ were derived. It was shown that the estimates of the autocovariance sequence did not differ much from the estimates of the autocorrelation sequence when $N > 240$. It has been shown that the variable $|\hat{K}(m)|$ is characterized by the unilateral normal distribution with the parameters given in equations (8) and (9).

The results of the tests shown in tables 1 and 2 prove that the simple generator B behaved atypically in the sequence sections tested. In the sequence of length 132000 the value of the ratio $a/\sigma = 0.63$ differs greatly from the expected value of 1.32. This proves that the distribution of the values $|\hat{K}(m)|$ for the generator being tested is different from the expected one i. e. from the unilateral normal distribution.

References

[1] KORN, G. A. - KORN, T. M. : Mathematical Handbook for Scientist and Engineers. 2nd edition. New York, McGraw - Hill Book Company 1968.

[2] WIDROW, B. - STEARNS, S. D. : Adaptive Signal Processing. New York, Prentice - Hall Inc. 1985.

[3] KOMO, J. J. : Random Signal Analysis in Engineering Systems. Orlando, Academic Press, Inc. 1987.

[4] GOLENKO, D. I. : Computer modeling and statistical analysis of pseudorandom numbers. Moskva, Nauka 1965.

[5] ZAJEZDNYJ, A. M. : Foundations of calculations in statistical radioengeneering. Moskva, Svjaz 1969.

[6] OLEHLA, M. - VĚCHET, V. - OLEHLA, J. : Řešení úloh matematické statistiky ve fortranu. 1. vydání. Praha, Nadas 1982.

[7] BERSHAD, N. J. : Analysis of a First Order Complex Recursive LMS Adaptive Predictor. IEEE Proceedings - F, vol 138, No. 4, August 1991, p. 321 - 330.

[8] ŠEBESTA, V. - KILIÁN, P. : Modelování náhodných procesů. Sborník semináře "Radioelektronika '91", Brno, duben 1991.

[9] ŠEBESTA, V. - KILIÁN, P. : Posuzování kvality generátoru z hlediska rovnoměrnosti rozdělení. Knižnice odborných a vědeckých spisů VUT v Brně, ročník 1991, svazek B.

About author, ...

Vladimír Šebesta was born in Předín, Czechoslovakia, in1939. He received the M. E. degree in electrical engineering from the ČVUT Praha, in 1961, and the CSc. (Ph.D.) degree in radioelectronics from the VUT Brno, in 1973. He is currently the Associate professor at the Radioelectronic department of the VUT Brno. Mr. Šebesta is a member of the Czechoslovak Society for Signal Analysis and Processing.

Pavel Kilián was born in Brno, Czechoslovakia, in 1968. He received the M.E. degree in electrical engineering from the VUT Brno, in 1991. He is currently a postgraduate student at the Department of Biomedical Engineering of the VUT Brno.