Listy Biometryczne - Biometrical Letters
Vol. 38(2001), No. 2, 81-87

A comparison of random number generators

Dorota Kozioł, Wojciech Zieliński

Department of Mathematical Statistics and Experimentation,
University of Agriculture,
ul. Rakowiecka 26/30, 02–568 Warszawa

SUMMARY

In investigation of stochastic properties of statistical procedures random number generators are applied. Obtained results depend on quality of a random number generator. In the paper the comparative analysis of five random number generators is provided.

KEY WORDS: random number generator

1. Introduction

Mathematical statistics creates new research tools such as estimators, statistical tests or decision procedures. It is very important to know stochastic properties of those procedures. In general, there are two methods of achieving this goal. One of them is analytical research of new statistical procedures. Unfortunately, in many situations this method could not be applied because of lack of appropriate mathematical tools, or it is very hard to use them. The second method is a computer simulation technique. This method is commonly used by statisticians.

Computer simulation technique known as the Monte Carlo method relies on multiple sampling and observation of the behavior of statistical procedures on simulated samples. For the sake of free access to computers and velocity of their work, a wide spectrum of potential possibilities may be tested in a short time. Consider a statistical model $(\mathcal{X}, \mathcal{B}, \{P_\theta, \theta \in \Theta\})$. We are interested in inference on θ. Let $\mathcal{T}$ be a statistical procedure. If $\mathcal{T}$ is an estimator, it is interesting to know, for example, its bias, variance or mean-square error. For $\mathcal{T}$ being a test statistic it would be interesting to know its size and/or power. Because theoretical approach may encounter technical troubles, therefore the Monte Carlo method is applied. For a given $\theta \in \Theta$, samples from the distribution P_θ are drawn. The procedure $\mathcal{T}$ is applied to the samples and on that basis the properties of the procedure are investigated.

A random number generator of uniform distribution $U(0,1)$ plays the main role in Monte Carlo studies. On the basis of samples from that generator, with the help of mathematical rules, random numbers from any distribution may be obtained. The theory of random number generators of uniform distribution and other probability distributions is well known (Wieczorkowski and Zieliński, 1997; Zieliński, 1972). In those books methods of constructing random numbers generators may be found. However, there are a lot of different programs with built-in generators on the computers market. Unfortunately, their numerical algorithms are not known to the users. Hence, to investigate their quality some statistical procedures must be applied. It is obvious, that $U(0,1)$ random number generator have to be a "good" one. In what follows we explain what does it mean.

Let $X_1,\ldots,X_n$ be a sequence of random numbers from a uniform generator. Theoretically, it should be a sample from the $U(0,1)$ distribution. So this sample should satisfy at least the following postulates:

1. $X_1,\ldots,X_n$ should be uniformly distributed on $(0,1)$ interval;
2. its characteristics should be compatible with characteristics of the $U(0,1)$ distribution;
3. it should be a random numbers sequence.

In the paper we test these characteristics of generator with the help of adequate statistical tests (Domański, 1979).

Five random number generators of distribution $U(0,1)$ were investigated: generator from Excel, generator from Statgraphics, generator from Statistica, Turbo Pascal generator and Ultra generator (Marsaglia and Zaman, 1991).

2. Statistical tests

We use the following theorem in testing generators.

THEOREM. *If X is a random variable with distribution function F, then the random variable $F(X)$ have the uniform distribution on interval* $(0,1)$.

Proof. $P\{F(X)\leq x\} = P\{X\leq F^{-1}(x)\} = F(F^{-1}(x)) = x$ $\square$

Let $X_1,\ldots,X_n$ be numbers from generator of the uniform distribution on interval $(0,1)$. These numbers will be treated as a realization of a random variable X. We want to check if X is distributed as $U(0,1)$. So, the hypothesis

$$H_0 : X \sim U(0,1) \qquad (*)$$

is verified. Let T be test statistic of a test of hypothesis $(*)$. Let F be the distribution function of null distribution of the statistic (i.e., its distribution when the null hypothesis is true). Due to the theorem statistic $F(T)$ has $U(0,1)$ distribution. Note that the value $1 - F(T)$ is known as a *p-value*.

Suppose that the generation of n numbers is repeated m times. For each simulated n numbers the appropriate *p-value* is calculated. Let $F(T_i)$ be the *p-value* obtained for i-th simulated n numbers. If hypothesis $(*)$ is true then $F(T_1), \ldots, F(T_m)$ should be a sample from uniform distribution. Hence checking the quality of a random number generator may be based on checking if $F(T_1), \ldots, F(T_m)$ is a sample from $U(0,1)$ distribution. To gain this aim some statistical test will be applied. These tests are described below. Described non–parametric tests based on series should check randomness of a sequence of generated numbers.

Kolmogorow test. The numbers $X_1, \ldots, X_n$ are ordered in nondecreasing order: $X_{(1)} \leq \cdots \leq X_{(n)}$. Let $F_n(x)$ denote an empirical distribution function. Kolmogorow test statistic of hypothesis $(*)$ is given by formula

$$D_n = \sup_{-\infty < x < \infty} |F(x) - F_n(x)| = \max_{1 \leq i \leq n} \left\{ \left| X_{(i)} - \frac{i}{n} \right|, \left| \frac{i-1}{n} - X_{(i)} \right| \right\}.$$

For a large n the random variable

$$\mathcal{D}_n = \sqrt{n} D_n + \frac{1}{\sqrt{6n}}$$

has a Kolmogorow distribution with distribution function K. It means that the random variable $K(\mathcal{D}_n)$ has $U(0,1)$ distribution. The distribution function K is given for example in Zieliński and Zieliński (1990).

Mean test. For a sample $X_1, \ldots, X_n$ the arithmetic mean $\bar{X}$ is calculated. If hypothesis $(*)$ is true, then for large n the random variable $\bar{X}$ has approximately the normal distribution:

$$P\{\bar{X} \leq x\} \approx \Phi\left(\left(x - \frac{1}{2}\right)\sqrt{12n}\right),$$

where Φ denotes the distribution function of the standard normal distribution. That is, the random variable

$$\Phi\left(\left(\bar{X} - \frac{1}{2}\right)\sqrt{12n}\right)$$

has the $U(0,1)$ distribution.

Second moment test. For a sample $X_1, \ldots, X_n$ the statistic $\overline{X^2} = \sum X_i^2/n$ is calculated. Under null hypothesis $(*)$ for large n the random variable $\overline{X^2}$ has approximately the following distribution

$$P\{\overline{X^2} \leq x\} \approx \Phi\left(\left(x - \frac{1}{3}\right)\sqrt{\frac{45n}{4}}\right) - \frac{0.1065}{\sqrt{N}}\varphi^{(2)}\left(\left(x - \frac{1}{3}\right)\sqrt{\frac{45n}{4}}\right).$$

Here $\varphi^{(2)}$ denotes the second derivative of Φ. That is, the random variable

$$\Phi\left(\left(\overline{X^2}-\frac{1}{3}\right)\sqrt{\frac{45n}{4}}\right)-\frac{0.1065}{\sqrt{N}}\varphi^{(2)}\left(\left(\overline{X^2}-\frac{1}{3}\right)\sqrt{\frac{45n}{4}}\right)$$

has the $U(0,1)$ distribution.

Maximum test. The test is based on

$$X_{\max}=\max\{X_1,\ldots,X_n\}$$

for a sample $X_1,\ldots,X_n$. If hypothesis $(*)$ is true then $P\{X_{\max}\leq x\}=x^n$, that is, the random variable $X_{\max}^n$ has the $U(0,1)$ distribution.

Minimum test. The test is based on

$$X_{\min}=\min\{X_1,\ldots,X_n\}$$

for a sample $X_1,\ldots,X_n$. If hypothesis $(*)$ is true then $P\{X_{\min}\leq x\}=1-(1-x)^n$, that is, the random variable $1-(1-X_{\min})^n$ has the $U(0,1)$ distribution.

Range test. For a sample $X_1,\ldots,X_n$ the statistic

$$V=X_{\max}-X_{\min}$$

is calculated. If hypothesis $(*)$ is true, then $P\{V\leq x\}=nx^{n-1}-(n-1)x^n$, that is, the random variable $nV^{n-1}-(n-1)V^n$ has the $U(0,1)$ distribution.

Series tests. Let p be a number from interval $(0,1)$. For sample $X_1,\ldots,X_n$ create a sequence of letters A and B in such a manner that if $X_i<p$ than A is put, elsewhere B is put. The test statistic is based on the number R_p of series of letters. If the hypothesis $(*)$ is true then, with $n_2=n-n_1$, the statistic R_p has the following distribution:

$$P\{R_p=r\}$$
$$=\begin{cases} p^n+(1-p)^n, & r=1\\ 2\sum_{n_1=k}^{n-k}\binom{n_1-1}{k-1}\binom{n_2-1}{k-1}p^{n_1}(1-p)^{n_2}, & r=2k\\ \sum_{n_1=k}^{n-k}\left(\binom{n_1-1}{k-1}\binom{n_2-1}{k}+\binom{n_1-1}{k}\binom{n_2-1}{k-1}\right)p^{n_1}(1-p)^{n_2}, & r=2k+1.\end{cases}$$

3. Results

In the paper five random number generators of uniform distribution on interval $(0,1)$ were analyzed:

- generator from Excel;
- generator from Statgraphics;
- generator from Statistica;
- generator from Turbo Pascal;
- generator Ultra.

Samples of sizes $n = 50, 100, 150, 200$ were simulated by each of those generators. For each sample p-values of above described tests were calculated. Such simulations were repeated 10000 times. Hence for each sample size and for each generator 10000 p-values were obtained. Agreement of those values with the $U(0,1)$ distribution was checked by a classical chi-square test. Results are presented in Tables 1 and 2.

Table 1. p-values of chi-square goodness-of-fit test for Kolmogorow, mean, second moment, minimum, maximum and range tests

Generator	Sample size	D	$\overline{X}$	$\overline{X^2}$	$X_{\max}$	$X_{\min}$	V
Excel	50	0.30356	0.56050	0.57700	0.88961	0.15550	0.49619
	100	0.07550	0.35266	0.36571	0.36571	0.62717	0.83678
	150	0.08357	0.87724	0.83678	0.54414	0.59363	0.23922
	200	0.09235	0.87724	0.42093	0.03968	0.61036	0.12939
Statgraphics	50	0.11225	0.37907	0.27000	0.46532	0.51197	0.90134
	100	0.14196	0.66089	0.79201	0.80732	0.54414	0.94969
	150	0.64402	0.61036	0.64402	0.99452	0.69453	0.42093
	200	0.54414	0.94969	0.94147	0.87724	0.92282	0.94147
Statistica	50	0.74422	0.00756	0.00864	0.33992	0.46532	0.88961
	100	0.02222	0.62717	0.46532	0.57700	0.14861	0.67774
	150	0.59363	0.69453	0.23922	0.04949	0.14861	0.09703
	200	0.22021	0.42093	0.20236	0.03348	0.05513	0.11774
Pascal	50	0.05815	0.14196	0.06463	0.01969	0.82225	0.74422
	100	0.25943	0.40668	0.62717	0.19387	0.72781	0.93250
	150	0.12939	0.64402	0.42093	0.74422	0.13556	0.52796
	200	0.07550	0.45026	0.31536	0.49619	0.56050	0.79201
Ultra	50	0.52796	0.00707	0.03751	0.05815	0.43545	0.14861
	100	0.10697	0.25943	0.37907	0.32748	0.28087	0.51197
	150	0.79201	0.71124	0.77636	0.91242	0.36571	0.29206
	200	0.17005	0.61036	0.66089	0.54414	0.14861	0.57700

Table 2. p-values of chi-square goodness-of-fit for series test

Generator	Sample size	$R_{0.1}$	$R_{0.3}$	$R_{0.5}$	$R_{0.7}$	$R_{0.9}$
Excel	50	0.50102	0.13331	0.86143	0.48435	0.17518
	100	0.67072	0.61259	0.06106	0.08642	0.06787
	150	0.84951	0.06120	0.04444	0.88321	0.38171
	200	0.14720	0.02908	0.66922	0.00000	0.27295
Statgraphics	50	0.25141	0.33150	0.00718	0.70867	0.36969
	100	0.10726	0.00290	0.51783	0.49961	0.73959
	150	0.32600	0.11855	0.43354	0.02529	0.00000
	200	0.02082	0.00018	0.27714	0.02492	0.01264
Statistica	50	0.90883	0.09471	0.96249	0.10354	0.04100
	100	0.98389	0.08074	0.36920	0.06148	0.27582
	150	0.36339	0.47912	0.12796	0.00013	0.00149
	200	0.37636	0.15114	0.03672	0.55696	0.23849
Pascal	50	0.18766	0.00000	0.56421	0.23165	0.27180
	100	0.56353	0.59889	0.26566	0.05826	0.54584
	150	0.72479	0.28396	0.57080	0.59295	0.00000
	200	0.19196	0.02560	0.00000	0.63014	0.32792
Ultra	50	0.06729	0.75015	0.22626	0.40378	0.90243
	100	0.00000	0.23275	0.68639	0.37751	0.07233
	150	0.06942	0.17206	0.15137	0.19116	0.30639
	200	0.02091	0.24782	0.59883	0.60958	0.45369

4. Conclusions

Table 1 shows results of investigation of compatibility of generated numbers with the uniform distribution $U(0,1)$. Generally speaking, adequate hypothesis are not rejected on standard significance level 0.05. Due to the probability theory some hypotheses are rejected. Similar "regularity" may be observed in Table 2.

Therefore we can admit that each of five tested generators give a simple sample (Table 2) from uniform distribution $U(0,1)$ (Table 1). We may consider investigated generators as "good" ones.

The generators of normal distribution were not considered here. Random samples from normal distribution are obtained through different arithmetical operations on samples of uniform distribution. "Normal" and "random" generated samples depend on statistical properties of the uniform generator and on arithmetical transformations. Because the method of transformation of uniform numbers to normal ones is in the know-how, so the quality of normal generators can be investigated only by statistical methods. Of course above mentioned tests may be applied in such a research.

REFERENCES

Domański Cz. (1979). *Statystyczne testy nieparametryczne*, PWE, Warszawa.

Marsaglia G., Zaman A. (1991). A new class of random number generators. *Ann. Appl. Prob.* **1–3**, 462–480.

Wieczorkowski R., Zieliński R. (1997). *Komputerowe generatory liczb losowych.* WNT, Warszawa.

Zieliński R. (1972). *Generatory liczb losowych.* WNT, Warszawa (Wyd. 2).

Zieliński R., Zieliński W. (1990). *Tablice statystyczne*, PWN, Warszawa.

Received 15 November 2001

Porównanie generatorów liczb losowych

STRESZCZENIE

W badaniach własności procedur statystycznych wykorzystywane są coraz częściej generatory liczb losowych. Uzyskiwane rezultaty uzależnione są od jakości stosowanego generatora. W pracy dokonana jest analiza porównawcza pięciu generatorów liczb losowych.

SŁOWA KLUCZOWE: generatory liczb losowych