ROLE OF RANDOM NUMBERS IN SIMULATIONS OF ECONOMIC PROCESSES

Dominika Crnjac Milić, Ph.D.
J.J. Strossmayer University in Osijek
Faculty of Electrical Engineering

Ljiljanka Kvesić, M.Sc.
University of Mostar, Bosnia and Hercegovina

Abstract

Various ways of selecting random numbers used in process simulations will be presented in this paper. Special attention will be given to complex phenomena not known enough to be precisely described. Modes of interaction are unknown; what is known are probabilities of interaction outcome. Such cases are found mostly in social and economic phenomena, such as population growth, economic predictions, decision-making risk analysis, etc.

Key words: random numbers, simulation, process, distribution, model.

1. Introduction

1.1 Types of simulation models

Classifications of simulation models resulted in four basic types of simulation models, which differ both in their approach to modeling and the type of the problem addressed, and in modeling and simulation techniques developed for such purposes. These are Monte Carlo simulation, continuous simulation, discrete event simulation and mixed continuous/discrete simulation (Law and Kelton, 1982; Kreutzer, 1986; Evans, 1988). All these types of simulation are dynamic simulations, excluding Monte Carlo simulation.

1.1.1. Monte Carlo simulation

Monte Carlo simulation[1] (statistical simulation), as the name implies, is linked to random phenomena. It is interesting that it is characterized as one of the first computer programming applications. The method was developed in Los Alamos during

[1] Vlatko Čerić, (1993), Simulacijsko modeliranje, Školska knjiga, Zagreb, p. 36

the Second World War II for the purpose of solving complex problems referring to creation of atomic bomb, such as calculation of dispersion of neutrons on the nucleus. However, the term itself is not used by full consent. Some authors call any type of software using random numbers Monte Carlo. As in the majority of references in relation with simulation modeling, in this text this term will be used only for static types of simulations by which problems are solved by creating samples from random variable distributions. In such cases, problems might be of either deterministic or stochastic character.

The following types of applications of Monte Carlo simulations are differentiated (Kleijnen, 1974):

(1) Deterministic problems whose solving is hard or expensive.

A typical example of this type is calculation of values of certain integrals that cannot be solved analytically, i.e. whose subintegral function is such that a solution in form of a mathematical expression cannot be found.

(2) Complex phenomena not known enough.

The second class of problems solved by Monte Carlo simulation refers to phenomena that are known insufficiently to be precisely described. Instead of knowing modes of elements interaction, only probabilities of interaction outcome are known, which are in Monte Carlo simulation used for execution of a series of experiments giving samples of possible states of dependent variables. Statistical analysis of such samples provides a distribution of probabilities of dependent variables of interest. Most frequently social or economic phenomena, such as population growth, economic predictions or risk analysis, are analyzed by this approach.

(3) Statistical problems with no analytical solution.

Statistical problems with no analytical solution are just one of broad classes of problems by solving of which Monte Carlo simulation is used. E.g. estimation of critical values or the power of testing new hypotheses belongs to this group. Generation of random numbers and variables is also used in problem solving.

In case of comparison of various regression methods, Monte Carlo simulation is used for generation of input data, which are then analyzed by means of various regression methods, providing estimations of regression parameters of these data. Since the input data are generated by some predetermined parameters, it is possible to compare the quality of various regression methods by accuracy of regression parameter estimations they give to known parameters by means of which the data are generated.

2. Random numbers

Existence of random values in a simulation model requires mechanisms which can generate values of variables from various probability distributions[2] during simulation experiments. A series of generated values of a random variable is a sample from the probability distribution describing that variable. We will describe random numbers which make the basis for random variable generation. We will present the way how to generate random numbers, describe a linear congruent random number generator and list fundamental methods of testing a random number generator.

2.1. Using random numbers in simulation experiments

Simulation process models containing components behaving randomly require corresponding methods of generating random numbers (Law and Kelton, 1982; Banks and Carson, 1984). During a simulation experiment, e.g. generation of a great number of servicing time values, demand size or interarrival times belonging to some probability distributions might be requested. Therefore it is necessary to have an effective and high-quality way of generating values of random numbers and variables. Unfortunately, the term "generation of random variables" is thereby not precise enough, i.e. this term implies generation of numerical values of random variables from corresponding probability distributions of the variable in question.

Using random numbers and variables in simulation models enables reproduction of irregular behavior of system elements without having the model with an excellent detailed description of that behavior. Random numbers and variables describe irregular behavior even in a compressed form.

The terms random number and random variable must be used very carefully, since it is not easy to say whether a certain series of numbers is random, although some series seems not to be random at all. This question cannot be answered correctly on the basis of knowing the way a series of numbers in question is formed. The only approach that might provide a satisfactory answer is to put that series of numbers under corresponding statistical tests which will prove whether it is a sample of a random number and to which extent. Quality of the random number generator is tested in the same way.

Precisely, the term "random number" implies a continuous random variable with the uniform distribution on the interval [0,1]. This distribution will be denoted by $U(0,1)$. Although this is the simplest continuous distribution, it is extremely important since random variables of all other probability distributions (normal, binomial,

[2] Željko Pauše, (1978), Vjerojatnost, informacija, stohastički proces, Školska Knjiga Zagreb, p. 248.

etc.) can be obtained by transforming3 the random variable $U(0,1)$ by means of independent identical distribution.

2.2. How can random numbers be generated?

In a long and interesting history of generating random numbers various methods of generation have been used. Let us mention a successful random number generator. Very successful generators of random numbers are nowadays linear congruent generators (abbr. LKG). LKG is represented by a choice of three positive integers:

a, multiplicator;

c, increment;

m, module; $m > a, m > c$

In order to obtain a desired series of random numbers $x_1, x_2, \ldots, x_N$, we generate a series of integers $X_1, X_2, \ldots, X_N$ starting from the initial value X_0 (called the seed). A series of numbers is generated by means of a recursive rule

$X_{n+1} = (aX_n + c) (\text{mod } m)$, $n=0,1,2,\ldots,N-1$.

After that, x_k is defined as X_k / m, $k = 1,2,\ldots,N$.

Every LKG produces a deterministic series, i.e. every number X_k, and therefore X_k is known, by which X_0, a, c and m.

LKG always produces a periodic series with $p<m$.

In order to obtain a high period p, module m is usually selected close to the greatest integer on the computer. It is usually $m=2^{31}-1$. In order to ensure a high quality of pseudorandom numbers, multiplicator a has to be selected carefully. One of the possibilities when $m=2^{31}-1$ is $a=16807$. For the purpose of generating a series $X_1, X_2, \ldots, X_n$, increment c is set to be 0. By selecting $c=0$, "randomicity" of a generated series (if $X_0 \neq 0$) is not reduced more significantly. The represented LKG

$$X_{n+1} = \left(16807 X_n\right) \left(\text{mod } 2^{31} - 1\right),$$

$$x_{n+1} = \frac{X_{n+1}}{2^{31} - 1},$$

is an example of a frequently used generator of pseudorandom numbers. Period p of this LKG is $2^{31}-2$.

---

[3] Vladimir Vranić, (1961), O graničnim teoremima teorije vjerojatnosti, Matematička biblioteka, Beograd , pp. 151-165.

Product a and X_n in LKG goes over the maximum value of the 32-bit integer if $X_n > 2^{31}/_a$. Thus, direct implementation of LKG with $m=2^{31}-1$ in an advanced language is impossible. Let us mention the following algorithm which gives us the opportunity to avoid that problem. Module is represented in the form $m=aq+r$, where $r<q$, and then $aX_n(\text{mod } m)$ is calculated by using the identity

$$aX_n(\text{mod}\, m) = \begin{cases} a(X_n(\text{mod}\, q)) - r[X_n/q], & \{a(X_n(\text{mod}\, q)) \geq r[X_n/q], \\ a(X_n(\text{mod}\, q)) - r[X_n/q] + m, & \text{otherwise}, \end{cases}$$

where $[X_n/q]$ is the integer of X_n/q. This expression does not include numbers greater than m.

There are special algorithms for random number generation from important distributions. For example, Box-Muller algorithm and polar algorithm are used for generation of random numbers from a standard normal distribution.

Numbers which are said to be strictly random can be generated. In 1955 RAND corporation published a table with a million random numbers obtained by such device. Input into the computer's memory, these obviously pure random numbers seem to be suitable for usage as well as pseudorandom numbers generated by means of LKG. However, the table might e too short. Scientific Monte-Carlo simulations require several billions of random numbers.

On the other hand, a combination of LKGs might produce a series of pseudorandom numbers which is, in spite of having a finite period, for all practical needs as good as an infinite series.

A great advantage of LKGs is reproducibility of random numbers. If the same seed is specified at the beginning of two series, then it will produce the same series of pseudorandom numbers.

Monte Carlo simulation is mostly used for determining expectation $E(X)$ of a random variable X referring to a specific stochastic model. Model simulation results in output data x_1, realization of a random variable X. The second simulation gives a new output x_2. Simulation is continued until we accumulate a sufficiently great number n of outputs $x_1, x_2, \ldots, x_n$. Arithmetic mean of these output data $\mu = n^{-1}\sum_{k=1}^{n} x_k$ is used as an estimator for μ.

The second set of Monte Carlo simulations can be carried out with either the same or an independent set of random numbers. The same set of random numbers is e.g. used for:

1.Sensitivity analysis
2. Comparative simulations

1. Sensitivity analysis

If the distribution of a random variable X depends on the model parameter, say a, then the expectation of X is a parameter function a, i.e. $E(X)=A(a)$. One of the most important parts of simulation model analysis is sensitivity analysis.

Sensitivity analysis is used for determining whether simulation results, in our case expectation $A(a)$, significantly differ when the value of parameter a is changed. Two sets of simulations are carried out for two different values, say a_1 and a_2, of parameter a. It is important to use a joint set of random numbers for both simulations, since otherwise the effect of parameter change might be mixed up with the change of random numbers.

2. Comparative simulations

Two sets of simulations can be conducted in two different situations of the given model, or for two different model configurations. In such simulations a joint set of random numbers is to be used.

Let us explain the usage of the joint set of random numbers. Comparative simulation usually includes calculation of the difference

$$\Delta = A(a_1) - A(a_2).$$

Notice that by increasing the number of simulations n, variance of estimators $\hat{A}(a_1)$, $\hat{A}(a_2)$ and $\hat{\Delta}$ decreases. However, the number of simulations n is always finite, and thus the estimator

$$\hat{\Delta} = \hat{A}(a_1) - \hat{A}(a_2) = \frac{1}{n}\sum_{k=1}^{n} x_k(a_1) - \frac{1}{n}\sum_{k=1}^{n} x_k(a_2)$$

has a finite variance

$$Var\left(\hat{\Delta}\right) = n^{-1}\left(Var\left[X(a_1)\right] + Var\left[X(a_2)\right] - 2\,\mathrm{cov}\left[X(a_1), X(a_2)\right]\right).$$

Clearly, the greater the covariance of $X(a_1)$ and $X(a_2)$, the less the variance of estimator $\hat{\Delta}$. It is reasonable to assume that the joint set of random numbers produces a positive correlation for $|a_1-a_2|$ being small.

2.3. How many simulations does it take for a certain purpose?

Monte Carlo simulation is usually used when the expectation $A=E(X)$ of a random variable X referring to a certain stochastic model is to be determined. Simulation of the model results in the output data x_1, realization of a random variable X. The second independent simulation gives the second output data x_2. Simulations continue until we have the total of n output data $x_1,x_2,...,x_n$. Arithmetic mean of these output data $\hat{A}=n^{-1}\sum_{k=1}^{n}x_k$ is used as an estimator for A.

In typical numerical simulations the tolerance level ε is predetermined. Due to the stochastic nature of Monte Carlo simulations, prior to simulation execution, it is also necessary to specify the reliability level 1-a. The problem consists of the following: determine the value of n such that the difference between $\hat{A}$ and A is less than ε (maximum value of the allowed error) with the probability greater than or equal to 1-a. There are two common ways of measuring the difference: absolute error $\left|\hat{A}-A\right|$ and relative error $\left|\hat{A}-A\right|/\left|A\right|$.

It the absolute error is used as a measure, then the following argument can be used for the necessary number of simulations. Since simulations are independent, the Central Limit Theorem[4] says that for great values of n, the distribution of $n^{-1/2}\sum_{k=1}^{n}\left(x_k-A\right)$ is approximately normal with expectation zero and variance σ^2. That is,

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left(x_k-A\right)\approx N\left(0,\sigma^2\right),$$

where $N(a,b)$ is a normal random variable with expectation a and variance b.

Law of Great Numbers[5] says that the sample variance

$$\hat{\sigma}^2\equiv\frac{1}{n-1}\sum_{k=1}^{n}\left(x_k-\hat{A}\right)^2$$

is close to the unknown variance σ^2 for sufficiently great values of n. From the previous

$$\frac{n^{-1}\sum_{k=1}^{n}\left(x_k-A\right)}{\sqrt{\left(n-1\right)^{-1}\sum_{k=1}^{n}\left(x_k-\hat{A}\right)^2}}\approx N\left(0,1\right).$$

[4] Željko Pauše, (1978), Vjerojatnost, informacija, stohastički proces, Školska knjiga, Zagreb, p. 147
[5] Nikola Sarapa, (1987), Teorija vjerojatnosti, Školska knjiga, Zagreb, p. 647

In tables indicating normal distributions we can find the value z_a such that

$$P\left[-z_a \le N(0,1) \le z_a\right] = 1-a\,.$$

Then

$$P\left[-z_a \sigma \le \sum_{k=1}^{n}\left(x_k - A\right) \le z_a \sigma\right] \approx 1-a\,,$$

or

$$P\left[\left|\mu - A\right| \le z_a \sqrt{\frac{1}{n(n-1)}\sum_{k=1}^{n}\left(x_k - \mu\right)^2}\right] \approx 1-a\,.$$

The following procedure might be used if we want to see whether the desired accuracy is reached or not.

1. Generate n_0=30 output data $x_1, x_2, \ldots, x_{30}$.

2. If

$$z_a \sqrt{\frac{1}{n_0(n_0-1)}\sum_{k=1}^{n}\left(x_k - \mu\right)^2} \le \varepsilon\,,$$

,

then the desired accuracy is reached.

Otherwise, generate output data x_k until

$$z_a \sqrt{\frac{1}{n(n-1)}\sum_{k=1}^{n}\left(x_k - \mu\right)^2} \le \varepsilon\,,$$

knowing that the Central Limit Theorem and the Law of Great Numbers give correct results only for great values of n. As mentioned previously, estimator σ^2 is close to the unknown variance σ^2, therefore it is not altered significantly with n for $n \ge n_0$. Hence, the number of simulations necessary for the desired accuracy should be close to

$$n_* = \frac{z_a^2}{\varepsilon^2(n_0-1)}\sum_{k=1}^{n_0}\left(x_k - \mu\right)^2\,.$$

A similar argument might be applied if the relative error $\left|\mu - A\right| / \left|A\right|$ is used as an accuracy measure.

1. Generate n_0=30 output data $x_1, x_2, \ldots, x_{30}$..
2. If

$$\frac{z_a}{|\bar{x}|}\sqrt{\frac{1}{n_0(n_0-1)}\sum_{k=1}^{n}\left(x_k-\bar{x}\right)^2} \le \frac{\varepsilon}{1+\varepsilon},$$

then the desired accuracy is reached.

Otherwise, generate output data x_k until

$$\frac{z_a}{|\bar{x}|}\sqrt{\frac{1}{n(n-1)}\sum_{k=1}^{n}\left(x_k-\bar{x}\right)^2} \le \frac{\varepsilon}{1+\varepsilon}.$$

Conclusion

In this paper we analyze ways of selecting random numbers and their usage in simulations, especially in Monte Carlo simulation. Application of independent sets of random numbers in sensitivity analysis, numerical calculations and comparative simulation is shown. The answer to the question how a simulation should be conducted for certain purposes is given.

References:

Vlatko Čerić (1993) Simulacijsko modeliranje, Školska knjiga, Zagreb.

A. M. Law, W. D. Kelton (1982) Simulation Modeling and Analysis, Me Grew-Hill, New York.

Željko Pauše (1978) Vjerojatnost, informacija, stohastički proces, Školska knjiga, Zagreb.

Nikola Sarapa (1987), Teorija vjerojatnosti, Školska knjiga, Zagreb.

Vladimir Vranić (1961) O graničnim teoremima teorije vjerojatnost, Matematička biblioteka, Beograd, pp.151-165

Vladimir Vranić (1970) Vjerojatnost i statistika, Tehnička knjiga, Zagreb.

B. P. Zeigler (1976) Theory of Modelling and Simulation, Wiley, New York.