Monte Carlo’s Simulation in the area of project cost management: the importance to use an adequate number of interactions that can enlarge the sample number by obtaining reliable results

Vânia Veiga Rodrigues (UFF) vaniaveiga@uol.com.br

Carlos Alberto Pereira Soares (UFF) carlos.uff@globo.com

José Murilo Ferraz Saraiva (UFF) vaniaveiga@uol.com.br

Rogério Gomes Côrtes (UFF) rogerio@rgcortes.com.br

Abstract

The utilization of Monte Carlo’s Simulation in project cost management aim, above all, to make possible to the entrepreneur, during the process of decision, to appraise the risks that the enterprise may be subject to, and to take the most possible secure decision, resulting this way, in the optimization of the profits. However, the utilization of this method will only give reliable results, if adequate studies were conducted during the phase of all necessary number interaction determinations.

Initially, the actual work presents an explanation about Monte Carlo’s Simulation, that consists basically in an artificial generation of probability of occurrence of certain events, in other words, in the generation of random values that will belong to a pre-established probability density function, with defined characteristics by simulation.

Next, is described a methodology that enables a risk evaluation of not following cost/sq.m, previously proposed, making use of Monte Carlo’s Simulation.

Finally, the procedures adopted are proposed to choose all necessary number interaction on a practical application used to validate these procedures.

Key words: Monte Carlo’s simulation, Risk analysis, Civil construction.

1. Introduction

The theory and the methods of stochastic analysis have been developed in a significant way during the last years. The publication of several goods and researches, continuously, in the technician-scientific way prove the applicability of those methods in the development of solutions of countless complex problems in the most different Areas of Research.

Monte Carlo's Simulation has been used with success in the engineering problem solving that involves a high number of random variables and complex functional relationships among them. Interesting examples of the application of Monte Carlo's Simulation can be observed in the Studies of Structural Reliability, Reliability of Systems and Analysis of Risk and, more recently, according to Rafetery (1996), Valeriano (1998) and Kassai (2002) for Projects Cost Management.

One of the main difficulties in the application of Monte Carlo's method is directly related to the great number of simulations requested to evaluate the probability of occurrence of an event, particularly when this value is relatively small. However, the method is quite general and it can reach any level of approach. The disadvantage of the method is just of practical order, in that it is usually extremely computationally demanding due to the repeated analyses that have to take place. Due to the difficulties found by the traditional analytical methods, and being considered that the computational simulation with Monte Carlo's use is quite simple and of easy application, several techniques came to be developed with the intention of reducing the number of necessary interactions.

The present text will treat of all procedures and necessary cares to choose an adequate number of simulations, exemplified by an application of Monte Carlo's Simulation, in the evaluation of the risk in not accomplishing the cost/sq.m, initially established for a real estate enterprise, typical construction and incorporation of residential buildings.

2. The Monte Carlo's method

The Monte Carlo's method, according to Rodrigues (2002), is better compared to the example of the application of the concept related with the computation simulation of a random experiment. The base of the method is the Theorem of the Integral Transformation, whose demonstration is available in various publications. It enables that, being known the cumulative probability density function F of a random variable x, a random sample, acted for $(x_1, x_2, ..., x_n)$, will be created with the use of the expression:

$$x_i = F^{-1}(r_i) \tag{1}$$

Being, r_i = random number $\in [0,1]$

Therefore, the method of Monte Carlo's Simulation is used in the solution of problems that involve random variables with known probability density function. It is known that, if $f_X(x)$ represents the probability density function of the variable x, then:

$$F_X(x) = \int_{-\infty}^{+\infty} f_X(x)dx \tag{2}$$

Considering that the method allows the obtaining of probability of occurrence of certain events, the interest of research is in determining the probability of occurrence of smaller values than the estimated value for cost/sq.m in certain real estate enterprise. That probability can then be expressed in terms of frequency relative f_A, defined through the expression:

$$f_A = \frac{n_A}{n} \tag{3}$$

Where:

- n_A= number of occurrences of the variable studied inside of the limits of the event A
- n = represents the number of simulations

3. The beta distribution

The first step for application of this technique is the identification of the problabe distribution of the variable that is being the study object, in other words, of the variable cost/sq.m.

The choice of the distribution of probability is the most difficult part of the process that involves Monte Carlo`s Simulation and, according to Flanagan (1993), for the variable cost/sq.m, the more appropriate distribution is the beta distribution, once the distributions that form this group are the only ones capable to present the following characteristics simultaneously:

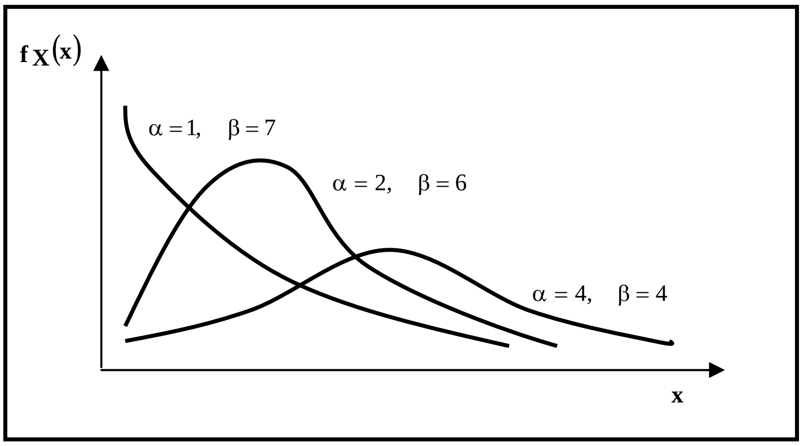
- To allow the updating with easiness, through the introduction of new historical data;
- To possess plenty of flexibility, in other words, it can take a great variety in ways in function of the form parameters α and β;
- To be easily identifiable for a certain sample.

It is known that the beta distribution is characterized by the following probability density function:

$$f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} \cdot x^{\alpha-1} \cdot (1-x)^{\beta-1}, \quad 0 \le x \le 1 \tag{4}$$

Where the parameters α and β just assume positive values. The coefficient $\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)}$ can be acted for $\frac{1}{B(\alpha,\beta)}$, where $B(\alpha,\beta) = \frac{\Gamma(\alpha)\cdot\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ is known as beta function.

The parameters α and β, form parameters are considered; different combinations of their values create a variety in ways of the density function, checking flexibility to the same. In figure 1 it has been plotted some of the possible ways.



Source: Rodrigues (2001)

Figure 1 - Beta distribution goes different values of the form parameters α and β

Chosen the distribution of probability, to proceed they should be generated n random numbers $r_i \in [\,0,1\,]$, for calculation of n random values through the equation (1), being X the variable cost/sq.m and F your cumulative probability density function. In the case of the beta function, the cumulative probability function is given for:

$$F(X;\alpha;\beta) = I_X(\alpha,\beta) \tag{5}$$

Being, $I_X(\alpha,\beta)$, the incomplete beta function.

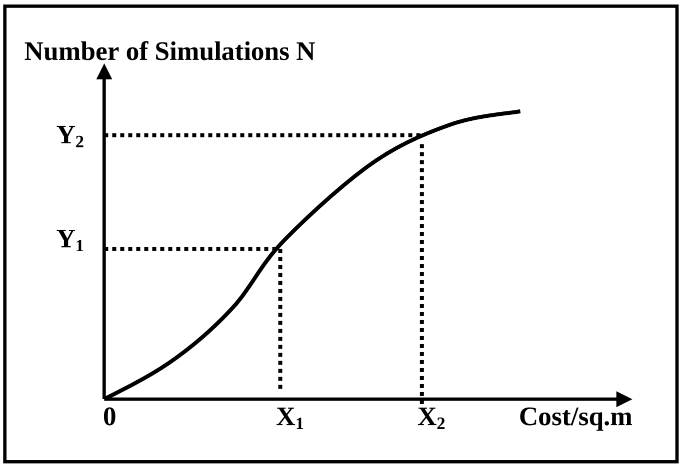
The random values of X can be generated in practice, with the use of mathematical computer programs. Rodrigues (2002) suggests the use of the applicative Mathcad 2000 Professional. This program possesses the function rbeta (m, s_1, s_2) that generates automatically m values of X, once supplied the parameters s_1 and s_2, that correspond respectively to the parameters α and β of the beta function, that are calculated through the resolution of the system formed by the related equations (6) and (7) to proceed:

$$m_x = \frac{\alpha}{\alpha+\beta} \tag{6}$$

$$\delta_X^2 = \frac{\alpha \cdot \beta}{(\alpha+\beta)^2 \cdot (\alpha+\beta+1)} \tag{7}$$

Where, m_x and δ_X^2 are, respectively, the values of the mean and of the variance of the sample.

Becoming separated the interval understood between the smallest and the largest value generated in n intervals of class of same width; determining the number of values generated that belong to each class interval and considering the found numbers in an accumulated way, the curve is drawn cost/sq.m x number of simulations, represented in figure 2, that alone much can supply the probability of occurrence of the value goes smaller cost/sq.m than certain value estimate.



Source: Rodrigues (2001)

Figure 2. - Curve of cumulative distribution function of values generated through the Monte Carlo's simulation

This way, the curve of figure 2 indicates that, for a total of N simulations, the probability of happening a value of cost/sq.m smaller than X_1 is of Y_1/N and although a probability exists of Y_2/N happening for a smaller value than X_2.

4. Determination of the number of simulations

Another very important aspect in Monte Carlo's Simulation concerns the number of necessary interactions. In general as larger the number of simulations, better the curve of accumulated distribution will reflect the range of possible results. However, Flanagan (1993) indicates the use of the chi-square test to verify the amount of accomplished simulations it is enough.

Rodrigues (2002) suggests another simpler alternative, for the evaluation of the number of interactions would be supplied through the expression:

$$N \geq \frac{p(1-p)}{\varepsilon^2 \cdot \delta} \tag{8}$$

Where:

- p = probability to be estimated by Monte Carlo's Method;
- ε = maximum error for the estimate of p;
- δ = variation coefficient.

Being considered the percentile 95 as a reasonable value for the variable cost/sq.m and being taken the values of ε and δ, respectively 0,03 (3%) and 0,2, the application of the expression (8) will supply the N value of 264 interactions.

It is clear, that the value of N it should just be considered as an initial evaluation of the number of simulations. The application of the formula (8), simultaneously with the use of the chi-square test, will supply an appropriate strategy to guide the researcher in the sense of verifying the value of N is correct or not.

5. Practical application

It deals with the enterprise of builder R G Côrtes Engenharia S/A, consisting of a typical residential building, with total construction area of 7.259 sq.m and it is composed of 54 flats.

To estimate the cost/sq.m of the construction enterprise, the builder was based on costs/sq.m monetary corrected for the same date, of similar residential buildings that she had built previously, in other words, with same pattern of finish, same technical specifications and with total construction area as close as the one enterprise in analysis.

The considered enterprises possessed the following values in Brazilian Currency for construction cost/sq.m: 407,00; 413,00; 418,00; 426,00; 430,00; 435,00; 440,00; 446,00; 451,00; 455,00; 462,00; 467,00 and 470,00. With base in this sample, the builder adopted the value then of R$440,00 as being the construction cost/sq.m with more probability for the enterprise in analysis.

Using these described methodology in this present work, are found to the medium standard sample detour, respectively the values of R$440,00/sq.m and R$20,65/sq.m. Replacing the values found on equations (6) and (7) we have:

$$440{,}00 = \frac{\alpha}{\alpha+\beta}\,; \tag{9}$$

$$(20{,}65)^2 = \frac{\alpha \cdot \beta}{(\alpha+\beta)^2 \cdot (\alpha+\beta+1)}\,. \tag{10}$$

Being solved the system formed by the equations above, they are found the following values for α and β:

- $\alpha = 577{,}23$;
- $\beta = 734{,}64$

Informing the values of α and β in the Application Mathcad 2000 Professional, were accomplished 300 (three hundred) interactions, resulting in three hundred values for cost/sq.m.

Dividing the space comprehended between the lower and higher cost value of the activity by square meter of the construction sample, in seven class intervals, they were found the frequencies of values for the intervals considered related in the picture 1.

Interval of Class	400 to 410	411 to 420	421 to 430	431 to 440	441 to 450	451 to 460	461 to 470
Frequency	7	16	44	78	83	43	29

Source: Rodrigues (2001)

Picture 1 - Picture of the frequencies obtained in 300 simulations

Through the cumulative distribution functions for the class intervals considered above, the curve of the figure 3 was drawn.

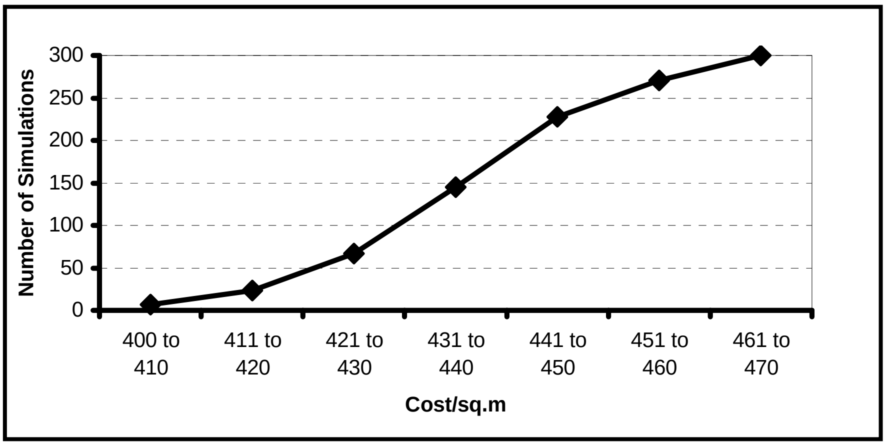


Figure 3 - Curve of cumulative distribution function of values generated through 300 simulations of Monte Carlo

6. Conclusions

However, as made calculations previously, through the expression (8), that it indicated the need of to be accomplished 264 interactions, it was done, according to the recommendations of Flanagan (1993), also already mentioned, the chi-square test (χ^2) for the data regarding 300 simulations, being verified a value for the χ^2_ϕ observed of (χ^2_ϕ) equal to 3,02. Being the critical χ_{ϕ}^{2} equal to 9,49, for a significance level of the 5%. This way, when the critical value (9,49) is bigger than the observed (3,02), it is had that the considered beta distribution is adjusted from a satisfactory way to the data of the problem. This way, the proposed procedures for this work, through equation (8), and of Flanagan (1993) are valid, respectively, by the initial evaluation to these same simulation numbers consist of an efficient method to determine N number of interactions necessary to the use of Monte Carlo's Simulation.

Through the cumulative distribution function curve acted in the figure 3, it is verified the approximate probability of only 48% to occur a value of cost/sq.m lower than the value of R$440,00 estimated by the builder, this way, was concluded that the adopted value was not the ideal one.

In case the builder had the intention of working with a reliable level of 95%, the accumulated frequency curve traced shows an approximate value of R$ 460,00 to be used in a variable construction cost/sq.m.

References

FLANAGAN, R. and NORMAN, G. *Risk Management and Construction*. Blackwell Science Ltd: London, 1993. 208 p.

KASSAI, J.R. et al. *Retorno de Investimento - Abordagem Matemática e Contábil do Lucro Empresarial*. Atlas:São Paulo, 2002. p. 98-141

RAFETERY, J. *Risk analysis in project management*. Chapman&Hall: London, 1996. 137 p.

RODRIGUES, V.V. *Modelo de Análise de Risco Aplicado a Estudos de Viabilidade para Construção e Incorporação de Prédios Residenciais*. 2001. 236 p. Dissertação (Mestrado em Engenharia Civil) - Centro Tecnológico da Universidade Federal Fluminense. Niterói. 2001.

RODRIGUES, V.V. et al. *The Use of Monte Carlo Simulation to Appraise the Risk of not Fulfilled Costs Previously Established for a Project*. In: National Conference With International Participation – In Situ Behavior of Constructions, 14 Th Edition, Galatzi, Bucuresti, România, Conference's Book, pp.227-235, 2002.

VALERIANO, D.L. *Gerência em Projetos - Pesquisa, Desenvolvimento e Engenharia*. Makron Books: São Paulo, 1998. 438 p.