10.2478/v10160-010-0021-4

Transactions of the VŠB – Technical University of Ostrava

No.2, 2010, Vol.X, Civil Engineering Series

paper #21

Petr KONEČNÝ[1]

EFFECT OF THE NUMBER OF RANDOM VARIABLES ON PRECISION OF PROBABILITY OF FAILURE ESTIMATION USING THE MONTE CARLO SIMULATION METHOD

Abstract

The paper deals with the relationship between the target probability, number of applied random variables and the number of Monte Carlo simulation steps needed to obtain satisfactory results. The precision of probability of failure estimation using crude Monte Carlo simulation is independent of the number of random variables in a studied case. Results obtained by Monte Carlo simulation are compared with Direct Determined Fully Probabilistic Method (DDFPM) that allows for fast computing solutions in the case of well-mapped tasks.

Keywords

Monte Carlo simulation. Probability, Necessary number of simulations, Reliability, DDFPM.

Abstrakt

Příspěvek se zabývá vztahem mezi cílovou pravděpodobností, počtem náhodně proměnných a počtem simulací Monte Carlo nutných k získání uspokojivého výsledku. Je zjištěno, že u přímé metody Monte Carlo ve studovaném případě nezávisí přesnost odhadu pravděpodobnosti poruchy na počtu náhodných veličin. Výsledky získané Monte Carlo simulací jsou porovnány s řešením pomocí přímého determinovaného pravděpodobnostního výpočtu (PDPV), přičemž užití PDPV může vézt u dobře zmapovaných úloh k rapidní úspoře výpočetního času.

Klíčová slova

Simulace Monte Carlo, Pravděpodobnost, Nutný počet simulací, Inženýrská spolehlivost, PDPV.

1 INTRODUCTION – PROBABILITY OF FAILURE AS RANDOM VARIABLE

If we analyze the engineering reliability with simulation tools such as Monte Carlo, used for example in SBRA (see [6], [5]), which use random number generators to estimate the probability of failure, it is necessary for the resulting probability estimation to be also regarded as a random variable (see e.g., [10], [11], [1] and [9]).

The precision of the probability of failure estimation by the Monte Carlo is dependent on the target probability of failure and the number of simulation steps, see e.g., [10], [11], [1]. While applying the direct Monte Carlo method, the number of random variables in the general form should have no effect on the error of estimation of the target probability (see [10], [11]). In [4] study, the verification for the target probabilities $P_t = 1\times10^{-2}$, $P_t = 1\times10^{-3}$, $P_t = 1\times10^{-4}$ a $P_t = 1\times10^{-5}$ was carried

[1] Ing. Petr Konečný, Ph.D., Department of Structural Mechanics, Faculty of Civil Engineering, VSB-Technical University Ostrava, Ludvíka Podéště 1875/17, 708 33 Ostrava - Poruba, Phone: (+420) 597 321 384, Email: petr.konecny@vsb.cz.

10.2478/v10160-010-0021-4

out in binary histograms. There was no relationship between the number of histograms and the precision of estimation of the probability of failure observed.

The paper aims to verify whether the number of random variables of the general form does not affect the precision of the target probability estimation using the direct Monte Carlo. The Monte Carlo computation is compared with the DDFPM method computation [2] [3].

2 METHODOLOGY

The principle of numerical experiments to verify the relationship between the number of random variables and the precision of estimation of the Monte Carlo simulation is as follows: If we create an example where the target probability of failure is known, then it is possible to verify the accuracy of the computation of this probability in Monte Carlo. Each Monte Carlo simulation of a given number of steps is viewed as one sample. These samples can be statistically evaluated if the sufficient quantity is obtained, and is done so for different numbers of simulation steps.

2.1 Product of histograms

As a model example the product of binary histograms used in study [4] is listed. The aim of this work was to verify the relationship between the number of random variables and the precision of estimation using simulation on the examples of the product of binary histograms with the known accurately computable probability. Example of the product of two histograms is shown in the Figure below. The probability of occurrence of 1 in the final product of *RF* is 1/100.

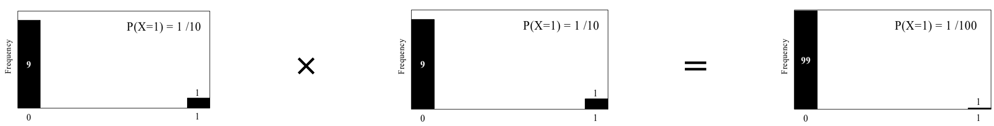

Fig.1: Product of two histograms $RF = X1 \times X2$

The probability is analyzed using the Monte Carlo simulation. It uses *Matlab* and simulation tool with its core formed by P. Praks, see [7] and [8].

The stochastic nature of the failure probability estimation is pointed out in Fig. 2, which shows the distribution of failure probability for the 300 calculations of the probability of failure P = 1/100 estimated ($N = 10^2$, 10^3.. 10^5). Variance of estimation decreases with an increase in the number of simulation steps.

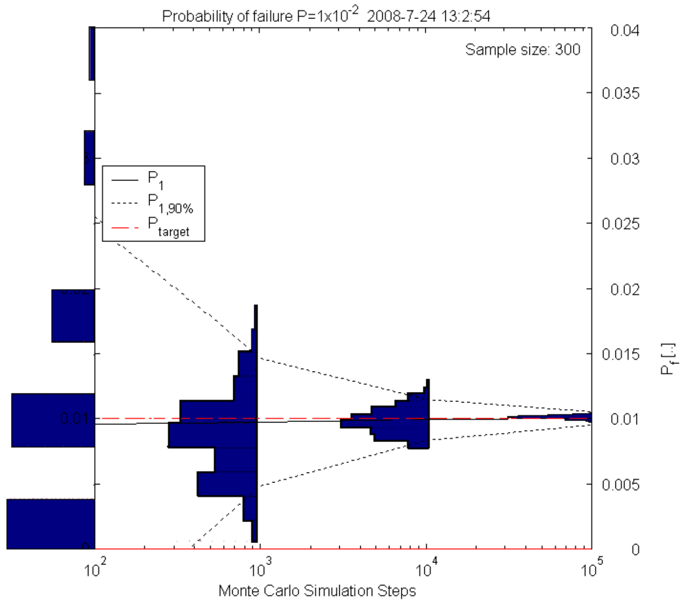

Fig. 2 Estimated probability of failure P_f (vertical axis) as a random variable estimated 300 times, depending on the number of Monte Carlo simulations (horizontal axis), the target probability P_t is 1/100

2.2 Sum of histograms

Using the Central Limit Theorem (CLT) it is possible to create a normal distribution (see e.g., [12]) from the general number of identical independent random variables, where the mean and standard deviation are known. A prerequisite is the I sum of the x distributions with the same distribution curve. The sum is the normal distribution of X:

$$X = \frac{1}{I}\sum_{i=1}^{I} x_i \,, \tag{1}$$

where:

I a number of the same, statistically independent distributions of x_i,

x_i distribution with mean $\mu_{x,i}$, and standard deviation $\sigma_{x,i}$,

X the resulting distribution with mean $\mu_{X,}$ and standard deviation $\sigma_{X,}$ which converges to a normal distribution with a sufficiently large amount of input distribution of x_i (and hence for the large I).

The mean μ_X of the resulting normal distribution X is equal to the mean μ_x of the random variable input. The standard deviation σ_X of the resulting file depends on the input standard deviation $\sigma_{x,i}$, and indirectly dependent on the square root of the number of random variables $\sqrt{I}$:

$$\sigma_X = \frac{\sigma_{x,i}}{\sqrt{I}} \tag{2}$$

If we obtain the parameters of normal distribution, we can construct its distribution function and determine its value for a selected probability of dropping below the mean. In the acquired normal distribution with the mean μ_X and standard deviation $\sigma_{X,}$ it is possible to derive from its inverse distribution function corresponding boundary values X_h for the selected probabilities of not exceding (e.g., $P_t = 1\times10^{-4}$, $P_t = 1\times10^{-3}$a $P_t = 1\times10^{-2}$).

$$X_h = \Phi^{-1}(P_t) \tag{3}$$

Acquired normal distribution can therefore be advantageously used to test dependence of the number of random variables on the precision estimation of the Monte Carlo method because the tested probability P_I is equaled to:

$$P_I = \mathrm{P}(X_{I,h} - X \le 0) \tag{4}$$

where:

P_I sought probability for the I of considered histograms,

X the resulting normal distribution with μ_X and standard deviation σ_X,

$X_{I,h}$ boundary value of the distribution X for I of random variables.

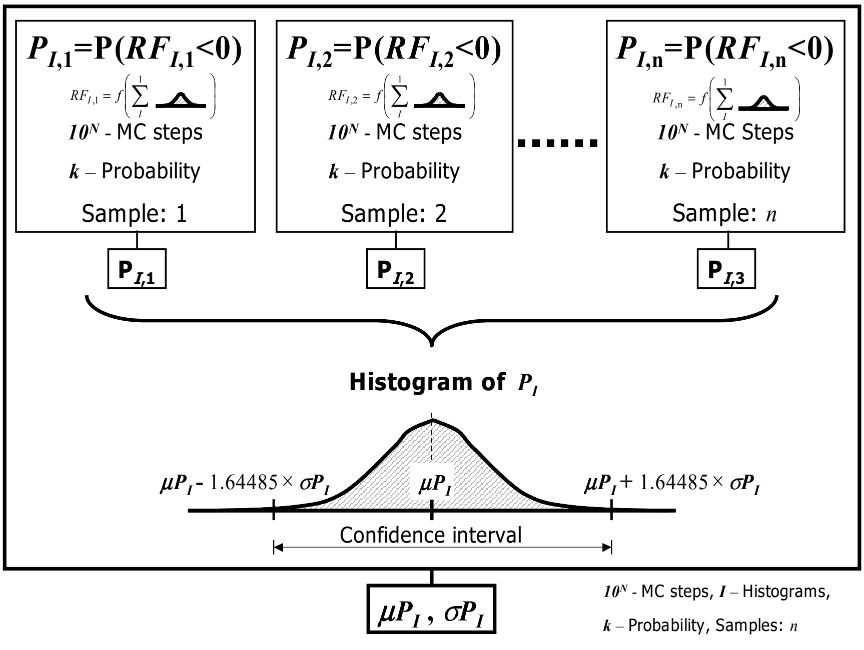

Fig. 3 Diagram of the calculation of one set of inputs for the parametric study. The mean μP_I as well as the standard deviation σP_I of the estimated probability correspond to a selected number of simulation steps 10^N, the number of histograms I, the target probability k, while the number of samples is marked n.

3 NUMERICAL EXAMPLES

3.1 Sum of random variables – Central Limit Theorem (CLT)

To verify the precision of Monte Carlo simulations using the Central Limit Theorem (CLT) a bounded normal distribution is used (see Fig. 4).

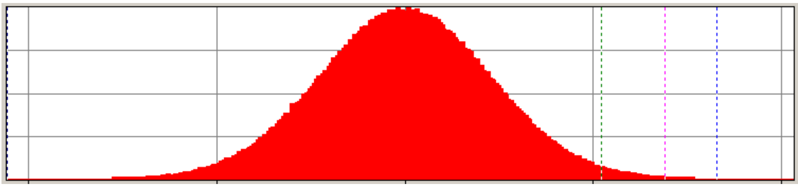

Fig. 4 Probability distribution of random input variables x_i - histogram normal3.dis

(See [5]); $\sigma_{x,i} = 0.9969$, $\mu_{x,i} = 6.224\times10^{-17} \approx 0$, <-3.5..3.5>

Summing up any number of independent implementations of the selected distribution according to equation (1), a normal distribution with a standard deviation as described in (2) can be obtained. Boundary values of the resulting normal distribution *X,* necessary to verify the precision of the Monte Carlo simulations are given in Table 1.

Table 1: Boundary values and statistical parameters of the resulting normal distribution X corresponding to the assumed probabilities of not exceeding P_t and a number of random variables *N*.

I	σ_X	μ_X	Boundary values of resulting distribution $X_{I,h}$		
			$P_t = (1\times10^{-2})$	$P_t = (1\times10^{-3})$	$P_t = (1\times10^{-4})$
5	0.445827	0.00E+00	-1.03715	-1.37772	-1.65824
10	0.315247	0.00E+00	-0.73337	-0.97419	-1.17255
20	0.222914	0.00E+00	-0.51857	-0.68886	-0.82912

The relationship (5) indicates an estimate of the probability P_5 using Monte Carlo for 5 histograms and a target probability $P_t = 1\times10^{-4}$.

$$P_5 = \mathrm{P}\left(X_{5,h} - X_5 \le 0\right) = \mathrm{P}\left(1.65824 - X \le 0\right) \approx 1\times10^{-4} \tag{5}$$

It should be noted that the resulting probabilities P_I for 5 and 10 random variables are slightly undervalued. This imprecision is acceptable with respect to the set target. Analogously, it is possible to derive the calculation for 10 and 20 used histograms, and further for $P_t = 1\times10^{-3}$ and $P_t = 1\times10^{-2}$.

3.1.1 Probability of failure $P_t = 1\times10^{-4}$

The following graph shows the relationship between the number of simulation steps applied and estimate variance of the probability of failure P_I. The estimated probability of failure is assessed using the Monte Carlo simulation tool. The probability $P_t = 1\times10^{-4}$ approximately corresponds to the value applied to the ultimate limit states ($P_\mathrm{d} = 0.7\times10^{-4}$), and is estimated using 5, 10 and 20 histograms. Outline of the exact solution shows the relationship (5).

Each probability is estimated for a selected number of simulation steps (N = 10^3, 10^4.. 10^6) 50-times and subsequently statistically evaluated. Intermediate values are connected by an axis. The graph in Fig. 5 contains the mean value as well as confidence interval limits. The estimate variance of the probability of failure ± 20 percent is considered, and it is done so with a 90 percent level of reliability. The required precision of the estimate when $P_t = 1{\times}10^{-4}$ can be achieved with approximately 676,000 simulation steps, shown in the following formula (see [11], [4]):

$$N_n = P_t(1-P_t)\left[\frac{t}{\varepsilon}\right]^2 = \frac{1}{10000}\left(1-\frac{1}{10000}\right)\left[\frac{1.64485}{\frac{0.2}{10000}}\right]^2 = 676315. \qquad (6)$$

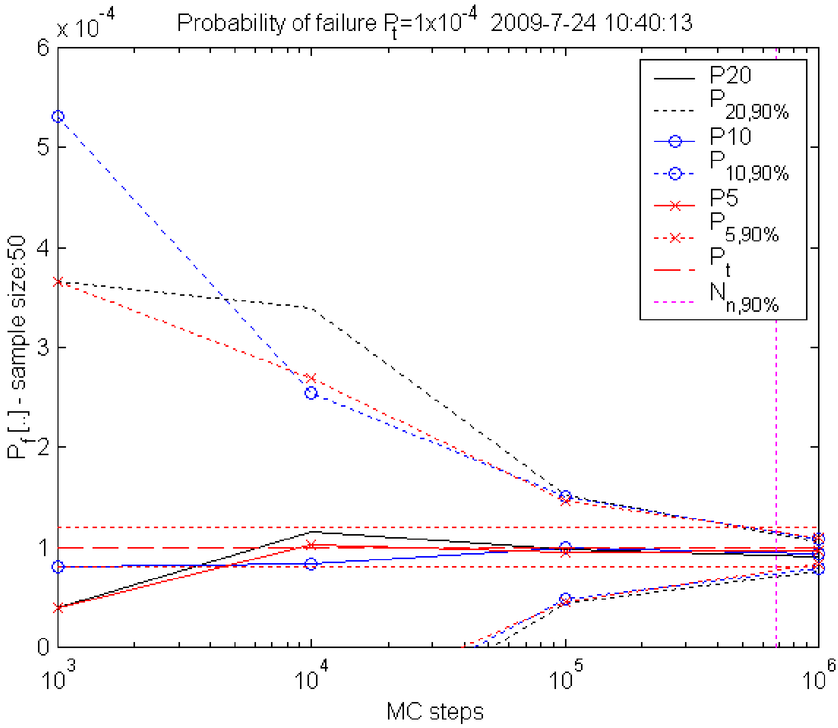

Fig.5 The probability of failure P_f as a random variable depending on the number of Monte Carlo simulations. The exact probability of failure $P_t = 10^{-4}$.
(P_{20} - 20 histograms, P_{10} - 10 histograms, P_5-5 histograms)

The result of numerical simulation in Fig. 5 shows that the number of histograms does not affect the precision of the estimate on the number of simulation steps. Means and variances of the resulting probabilities P_5, P_{10}, and P_{20} do not differ significantly for 10^5 and 10^6 simulation steps, which corresponds to the recommended 676,000 simulations.

3.1.2 Probabilities of failure $P_t = 1{\times}10^{-3}$ a $P_t = 1{\times}10^{-2}$

When estimating the probabilities $P_t = 1{\times}10^{-3}$ a $P_t = 1{\times}10^{-2}$, we are obtaining the same results as in the previous paragraph, see Fig. 6 and Fig. 7. The number of histograms does not affect, even in those cases, the precision of the estimated probability.

3.2 Results and Time of Calculation

Tab. 2 shows the calculated probabilities as well as the time needed to estimate the probability of failure. Time of calculation is shown for the Monte Carlo simulation and for the necessary number of simulation steps using *Anthill* programme.

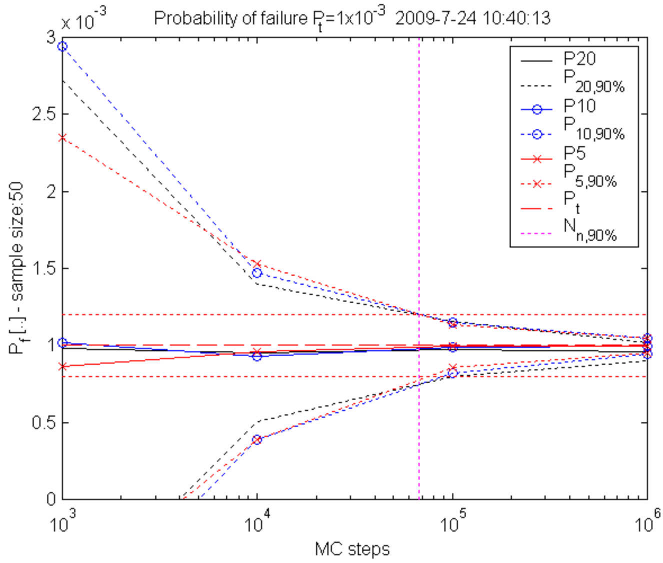

Fig. 6 The probability of failure P_f as a random variable depending on the number of Monte Carlo simulations. The exact probability of failure $P_t = 10^{-3}$.
(P_{20} - 20 histograms, P_{10} - 10 histograms, P_5 - 5 histograms)

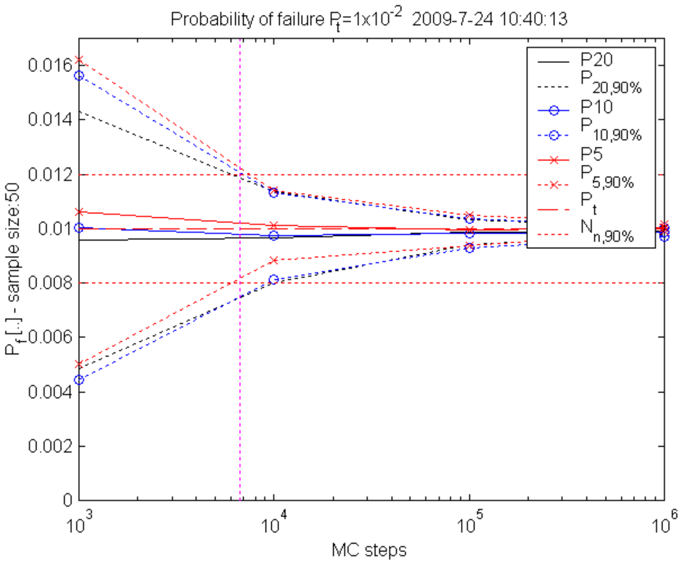

Fig. 7 The probability of failure P_f as a random variable depending on the number of Monte Carlo simulations. The exact probability of failure $P_t = 10^{-2}$.
(P_{20} - 20 histograms, P_{10} - 10 histograms, P_5 – 5 histograms)

This time is compared with the time of calculation obtained with the Direct Determined Fully Probabilistic Method (DDFPM, see [2] and [3]) using the *ProbCalc* programme. In all cases examined, the required probability was in the expected range of ± 20%, using both Monte Carlo method and DDFPM. It should be noted that for purposes of comparing computing time with *Anthill*, the definition of limit state using the DLL library is not used for the computation using *ProbCalc*. Time required for the computation using DDFPM method is dependent on the selected solution method. In the case of a simple combination of all histograms of 256 classes, $1{,}1\times10^{12}$ of simulations for 5 random variables would come into consideration, which would correspond to approx. 239 days of computation. To consider the simple sum of 20 variables is therefore unthinkable, because the computational complexity dramatically increases with the higher number of variables. For the computation itself using DDFPM method, there needs to be an adequate strategy. If individual random variables before the computation itself are added together, i.e. the *Combination* option in the ProbCalc programme [3], a drastic decrease in computational complexity occurs. Time of calculation is then in all considered cases equaled to less than 1 second.

Table 2 Comparison of target probabilities P_t with calculated probabilities P_I for the I histograms, including the simulation time using Monte Carlo direct method and DDFPM method. Monte Carlo shows a mean and confidence interval of 50 resulting probabilities.

P_t		Monte Carlo (Anthill)			DDFPM (ProbCalc)		
		$\varnothing P_I \pm P_{I,90}$	Number of steps	Time [sec]	P_I	Description	Time [sec]
$1\text{x}10^{-2}$	5	$(0.97\pm0.17)\times10^{-2}$	6.7 thous.	1	0.97×10^{-2}	The sum of random variables carried out using combination method. The calculation was carried out without optimization	<1
	10	$(0.97\pm0.16)\times10^{-2}$	6.7 thous.	1	0.99×10^{-2}		<1
	20	$(1.0\pm0.13)\times10^{-2}$	6.7 thous..	1	0.92×10^{-2}		<1
$1\text{x}10^{-3}$	5	$(0.98\pm0.18)\times10^{-3}$	67.5 thous.	11	0.93×10^{-3}		<1
	10	$(0.98\pm0.17)\times10^{-3}$	67.5 thous.	11	0.96×10^{-3}		<1
	20	$(0.99\pm0.14)\times10^{-3}$	67.5 thous.	11	0.94×10^{-3}		<1
$1\text{x}10^{-4}$	5	$(0.90\pm0.15)\times10^{-4}$	675.3 thous.	110	0.86×10^{-4}		<1
	10	$(0.94\pm0.15)\times10^{-4}$	675.3 thous.	110	0.93×10^{-4}		<1
	20	$(0.97\pm0.13)\times10^{-4}$	675.3 thous.	110	0.91×10^{-4}		<1

3 CONCLUSION

The paper suggests that the estimated probability of failure is also a random variable under the use of direct Monte Carlo. It has been found that under the direct Monte Carlo method in the example studied the precision of estimate of the probability of failure does not depend on a number of random variables. The paper verifies that the precision of estimate using Monte Carlo is affected by the size of the target probability P_t and the number of applied simulation steps, as indicated in, i.a., [11]. If the number of simulation steps is sufficiently large, it is possible to describe distribution of the probability of failure P_f by the normal distribution. Based on the normal distribution of the estimate of the probability of failure it is possible to estimate the required number of simulation steps to achieve the desired precision, and/or to estimate the precision of the result obtained (confidence interval).

Comparative solution using the DDPFM method resulted in obtaining satisfactory probabilities of failure, which similarly to the results obtained by the Monte Carlo, were within an expected tolerance. While using the DDPFM method, it is important to select the right solution strategy, in order to that time of calculation can be significantly reduced. In case of the well-mapped tasks, using the DDPFM can significantly save calculation time, even in comparison with direct Monte Carlo.

Further works should be focused on the direct comparison of the direct Monte Carlo method and the more advanced Monte Carlo methods of the Importance Sampling and Latin Hypercube Sampling type, with regard to a number of random input variables and a required number of simulation steps.

ACKNOWLEDGEMENTS

The project was implemented with financial support through the Grant Agency of the Czech Republic. The registration number of the project GA CR 105/07/1265.

REFERENCES

[1] FEGAN, G. Chapter: „Precision Of Simulation Results.“ in [5], 2003.

[2] JANAS, P., KREJSA, M. Chapter 24.5 Using a Direct Determined Probabilistic Solution in the Framework of SBRA Method. In CD-ROM of [5], 2003.

[3] JANAS, P., KREJSA, M. Numerický výpočet pravděpodobnosti užitím useknutých histogramů při posuzování spolehlivosti konstrukcí (Numericla computation of probability usány bounded histograms applicable in the structural reliability assessment). In *Sborník vědeckých prací Vysoké školy báňské - Technické univerzity Ostrava*, 2002, vol. II., (č. 1), s. 47-58. ISSN 1213-1962 (in Czech).

[4] KONEČNÝ, P. Přesnost odhadu pravděpodobnosti poruchy (Precison of the probability of failure estimation), In *Sborník vědeckých prací Vysoké školy báňské - Technické univerzity Ostrava*. Číslo 1, rok 2008, ročník VIII, řada stavební, článek č. 33, pp. 333-344, 2008, ISBN 978-80-248-1883-2, ISSN 1213-1962 (in Czech).

[5] MAREK P., BROZZETTI J., GUŠTAR M., TIKALSKY P., Editors. *Probabilistic Assessment of Structures using Monte Carlo Simulation. Basics, Exercises, Software, (Second extended edition).* Publisher: ITAM Academy of Sciences of Czech Republic, Prosecká 76, 190 00 Prague 9, Czech Republic, 2003. ISBN 80-86246-19-1.

[6] MAREK, P., GUŠTAR, M., BATHON, L. *Simulation-Based Reliability Assesment for Structural Engineers*. Boca Taton, Florida, CRC Press, 1995, ISBN 0-8493-8286-6.

[7] PRAKS, P. Numerical aspects of Simulation Based Reliability Assessment of Systems. In *International Colloquium Euro-SiBRAM'2002*. Volume II. ITAM, Academy of Sciences of the Czech Republic, Prague, 2002. ISBN 80-86246-17-5.

[8] PRAKS, P. *Analýza spolehlivosti s iteračními řešiči*. Doctoral dissertation thesis, VŠB – Technical University of Ostrava, Faculty of Electrical Engineering and Computer Science, Department of Applied mathematics, December, 2005.

[9] PRAKS, P., KONEČNÝ, P. Chapter „Direct Monte Carlo Method vs. Improved Methods Considering Applications in Designers Every Day Work“ in CD-ROM of [5], 2003.

10.2478/v10160-010-0021-4

[10] SHOOMAN, M.L. *Probabilistic Reliability: An Engineering Approach*. MCGRAW-HILL, New York, 1968.

[11] SCHUËLLER, G. Past, present & Future of Simulation-based Structural Analysis In *International Colloquium Euro-SiBRAM'2002*. Volume II. Institute of Theoretical and Applied Mechanics, Academy of Sciences of the Czech Republic, Prague, June 2002. ISBN 80-86246-17-5.

[12] MATH WORLD – Central Limit Theorem - http://mathworld.wolfram.com/CentralLimitTheorem.html

Reviewer:

Ing. Miroslav Sýkora, Ph.D., CTU in Prague - Klokner Institute