

LHS AS A TOOL FOR THE INPUT DATA

Jan Pruška¹, Miroslav Šedivý²

¹ CTU in Prague - FCE, Prague, Thakurova 7, 166 29, Czech Republic

² GeoTec – GS a.s., Prague, Chmelova 2920/6 , 106 00, Czech Republic

KEYWORDS

Latin Hypercube Sampling, stochastic estimation, statistical analysis.

INTRODUCTION

Natural materials (such as rock and soil) are invariably affected by structural defects, mineralogy, grain size, porosity, degree of weathering and anisotropy. Therefore the reliable estimates of their properties are very important. This can be easily achieved in the two steps. In the first one we obtained the probability distributions, which give us both the range of values that the variable could take and relative frequency of each value within the range and in the second one execute stochastic estimation. This paper is focused on the demonstrating the use of Latin Hypercube Sampling (LHS) to describe stochastic estimation of the rock mass parameters due the variability of their mechanical properties. The procedure of the founding the appropriate distribution for the measured properties of rock or soil specimens using QC-Expert software is described after short introduction. In the following part the basic simulation techniques are shortly mentioned while Latin Hypercubes Sampling method is described in detailed. The method of sampling is outlined in the first step. In the second step we described the implementation of the Log-normal distributions to the LHS sampling. At the end the application of the statistical analysis of the input parameters by LHS is mentioned.

ESTIMATING THE PARAMETERS OF A STATISTICAL MODEL

For the statistical processing for the measured properties of rock or soil specimens we used QC-Expert software with probability module (Trilobyte, 2012). This module provides the MLE method (Maximum Likelihood Estimate) for deriving estimations for given data. In general, MLE method selects parameters that produce distribution that gives the observed data the greatest probability (i.e., parameters for a given statistic that make the known likelihood function a maximum). This is a simple, compelling concept and it has a host of good statistical properties. On the Fig. 1 is an example of the graphical output of the QC-Expert software with MLE criterion.

DISTRIBUTION USED IN PROBABILITY ANALYSES

One of the essential questions in the probability analyses is determination of the proper distribution that can be used to fit a given data set. In addition to the commonly used normal distribution there are a number of alternative distributions that are used in probability analyses. Experts opinion on the choosing the distribution in accordance with the define properties or events is disunited and therefore we mentioned some of the most useful distribution only.

Normal or *Gaussian* distribution is the most common type of probability distribution function and the distributions of many random variables conform to this distribution. It is generally used for probabilistic studies in geotechnical engineering unless there are good reasons for selecting a different distribution. Typically, variables that arise as a sum of a number of random effects, none of which dominate the total, are normally distributed. The using of the normal distribution as a basis for a Monte Carlo analysis in certain investigations can cause numerical instability.

Beta distribution is very versatile distribution which can be used to replace almost any of the common distributions and which do not suffer from the extreme value problems because the range is bounded by specified values.

Exponential distribution is sometimes used to define events such as the length of discontinuities in a rock mass, rock bursts etc.

Lognormal distributions is useful when considering processes such as the crushing of aggregates in which the final particle size results from a number of collisions of particles of many sizes moving in different directions with different velocities.

Weibul distribution is used to the outcome of tests such as point load tests on rock core in which a few very high values may occur.

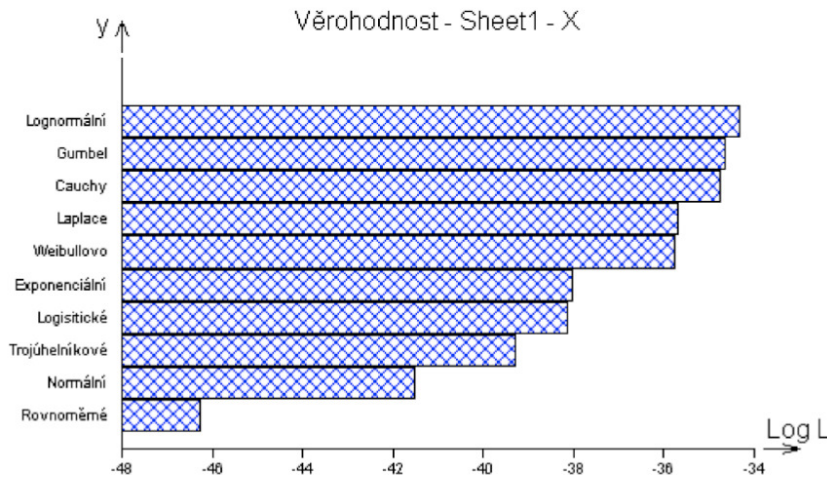


Fig. 1: Graphical output of the QC-Expert software

SAMPLING TECHNIQUES

Monte Carlo simulation is a well-known toll that is used to analyse random phenomena. In Monte Carlo simulation is the random problem transformed into several deterministic problems that are much easier to solve – sample inputs are used to generate sample outputs with statistical or probabilistic information about random output quantity. Monte Carlo simulation is simple to use and therefore has found much favour in geomechanics, particularly in stability analysis of rock slopes (McMahon, 1971). The simplest sampling scheme of a Monte Carlo simulation approach is to use a pseudorandom number generator to select random numbers between 0 and 1 and swap them with randomly generate values for each variable which will be input to the calculation. However, this simple (and best-known) random sampling scheme required for good accuracy and repeatability many samples – in practice for generating a probability distribution of the safety factor of rock slopes is the minimum number of selection between 200 up to 2000 (depending on the desired accuracy). The simulation output (random variable which depends upon random input variables, fields and processes) may be in several approaches. One way is to define the probability that safety factor F is less than a prescribed value F_p :

$$(1) P(F < F_0) = \frac{n}{N}$$

where: n = number of trials in which $F < F_n$,

N = number of selections.

Another approach (Mahtab 1992) is to plot a cumulative distribution of random F , which can be used to define the probability of F being less than a given value (Fig. 2).

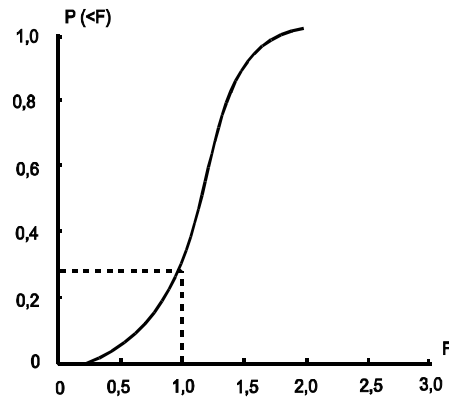


Fig. 2: Cumulative distribution of Monte Carlo results (Mahtab 1992)

The *Generalised Point Estimate Method*, developed by Rosenbleuth (Rosenbleuth, 1981) and discussed in detail by Harr (Harr, 1987) can be used for rapid calculation of the mean and standard deviation of a quantity such as a factor of safety that depends upon random behaviour of input variables. Hoek (Hoek, 1989), discussed the application of this technique to the analysis of surface crown pillar stability while Pine (Pine, 1992) has applied this technique to the analysis of slope stability and other mining problems.

The *Latin Hypercube* sampling technique Bažant (Bažant, 1985) is a relatively recent development which gives comparable results to the Monte Carlo technique Latin hypercube sampling preserves marginal probability distributions for each simulated variables. To fulfilled this aim, Latin hypercube sampling constructs a highly depend joint probability density function for the random variables in the problem, which allows good accuracy in the response parameter using only a small number of samples (At al. 2001).

PRINCIPLE OF LATIN HYPERCUBE SAMPLING

The principle of Latin Hypercube sampling will be outlined for the two layers area of the shale rock mass (weathered and strong shale). In the case of the statistical analysis of input parameters influence to the tunnel deformations modelling we can conveniently dealt the Mohr–Coulomb failure criterion with 5 basic parameters. The parameters do not have a single fixed value and hence these parameters are described as random variables q_1, \dots, q_5 :

- $q_1 = E_{def}$ deformation modulus,
- $q_2 = \nu$ Poisson number,
- $q_3 = \gamma$ unit weight,
- $q_4 = c$ cohesion,
- $q_5 = \phi$ friction angle.

A normal distribution has been assumed for the variables q_i . This distribution is described by Gauss – Laplace function (Fig. 2):

$$(2) f(t_0) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}}$$

where: t standard variable.

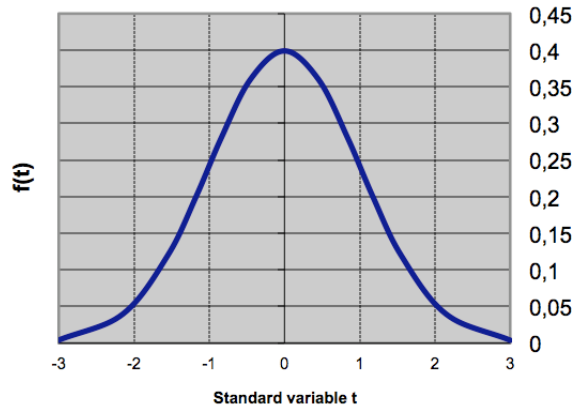


Fig. 3: Probability density function – Gaussian distribution

In the normal distribution the random variable is continuously distributed and the area under the probability density function (is always unity) is described by cumulative distribution function (the integral of the corresponding probability density function):

$$(3) F(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} \cdot dt = 1$$

where: t standard variable.

Cumulative distribution function gives the probability that the variable will have a value less than or equal to the selected value. For variables q_i we assumed that falling with probability 95,45% to the entered range:

$$(4) F(t) = \frac{1}{\sqrt{2\pi}} \int_{-2}^{+2} e^{-\frac{t^2}{2}} \cdot dt = 0,954$$

where: t standard variable.

The vertical axis of the graphical representations of a cumulative distribution function has been standardized so the values of the variables q_i are within range $<0,1>$ (Fig. 4). The correlation has been made using the following equation:

$$(5) q_i = \frac{a+b}{2} + q_i' \cdot \frac{b-a}{2}$$

where: a lower limit of the variable q_i ,
b upper limit of the variable q_i .

Diagram of the distribution function of the input parameters q_i has been uniformly divided into $N = 5$ non-overlapping intervals Δq_{ik} ($k=1,2,\dots,N$) with the same probability of accuracy. For the middle points of the non-overlapping intervals Δq_{ik} has been set the norm values q_i' (Table 1).

Table 1: Normed values q_i'

$F(q_i')$	0.1	0.3	0.5	0.7	0.9
q_i'	-0.64	-0.26	0	0.26	0.64
Interval No. used for q_i	1	2	3	4	5

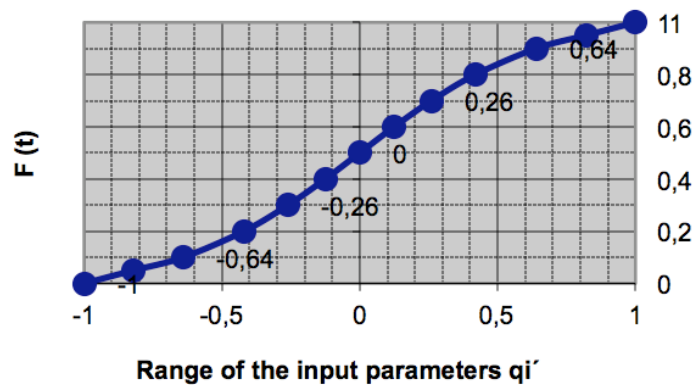


Fig. 4: Dividing the distribution function into 5 intervals

Afterwards the non normed input parameters q_i has been converted and then the table of the random permutation for individual parameters was generate (Fig. 5). Thus, on computer run number 1, the input data set 1 is formed by selecting the specific value of q_1 from the interval number 4 ($q_1=E$) and pairing this value with the specific value of q_2 selected from interval number 2 ($q_2=\square$). The vectors for the subsequent runs are constructed in a similar manner. Thus we can assemble matrix of input parameters q_i for each calculation (Table 2).

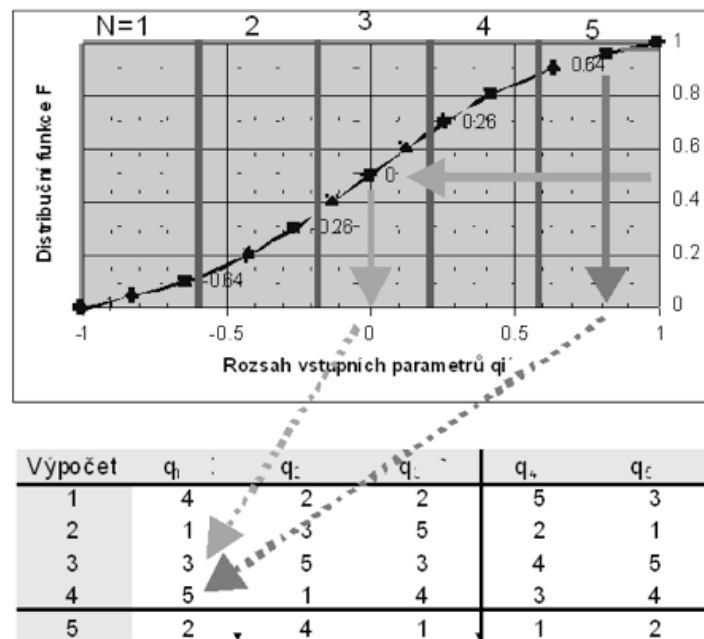


Fig. 5: Dividing the distribution function into 5 intervals

Table 2: Random permutations of the input parameters q_i

Data set	$q_1=E$	$q_2=\square$	$q_3=\square$	$q_4=c$	$q_5=\square$
1	4	2	2	5	3
2	1	3	5	2	1
3	3	5	3	4	5
4	5	1	4	3	4
5	2	4	1	1	2

LOG-NORMAL DISTRIBUTION IN LHS

This part of the paper describes precisely the implementation of the log-normal distribution in the LHS. The reason for this focus on the next distribution in the LHS is that for the describing distribution function of the log-normal distribution we cannot use the standard function (such as in the Gaussian distribution). Probability density function of the log-normal distribution is defined by function:

$$(6) f(x, \mu, \sigma) = \frac{1}{x \sigma \sqrt{\pi}} \cdot e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$$

where: x stochastic variable (measured value), $x > 0$,

- aritmetical average,
- standard deviation.

Aritmetical average and standard deviation can be expressed using the adequate values of the related normal distribution:

$$(7) \lambda = \ln(\mu) - \frac{\sigma^2}{2} \quad \zeta = \sqrt{\ln \left[1 + \left(\frac{\sigma}{\mu} \right)^2 \right]}$$

where: x variable,
 μ aritmetical average,
 σ standard deviation,
 λ aritmetical average $\ln(x)$,
 ζ standard deviation $\ln(x)$.

Cumulative distribution function (Fig. 6) can be expressed by error function:

$$(8) F(x, \mu, \sigma) = 0,5 + \frac{1}{2} \operatorname{erf} \left[\frac{\ln(x) - \mu}{\sigma \sqrt{2}} \right]$$

where: μ aritmetical average,
 σ standard deviation,
 erf Gaussian error function.

Function erf (Figure 7) is not the elementary function and can be described by Taylor row:

$$(9) \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{n!(2n+1)}$$

where: x variable,
 n serial number in the row.

We can determine the interval of the log-normal distribution (using margin q_{inf} , q_{sup}) that includes input parameters (measured values) with probability 95%:

$$(10) q_{inf}; q_{sup} = e^{\lambda m k_n \zeta} = e^{0,624 m 1,647 \times 0,357}$$

where: q_{inf} bottom limit of the interval (from left),
 q_{sup} upper limit of the interval (from right),
 λ aritmetical average $\ln(x)$,
 ζ standard deviation $\ln(x)$,
 k_n quantile.

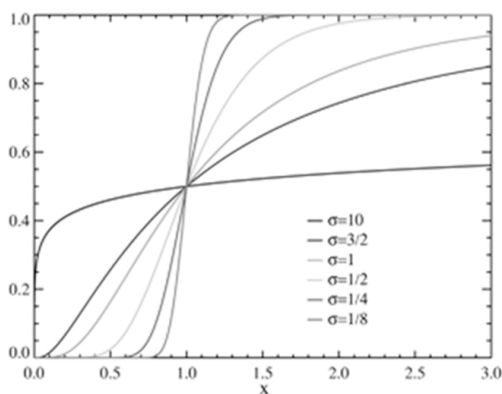


Fig. 6: Log-normal distribution – CDF

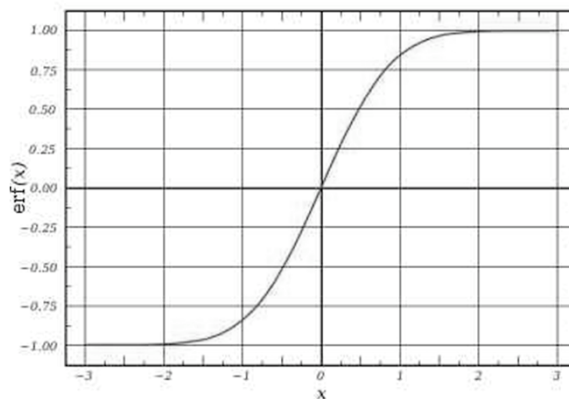


Fig. 7: Erf function

Mathematical relation between input mean, error function and parameters of the normal distribution (μ and σ) are describe using following equations:

$$(11) \sigma = \frac{\ln(\text{error factor})}{1,645}$$

$$\mu = \ln(\text{input mean}) - 0,5\sigma^2$$

where: σ standard deviation,
 μ aritmetical average,
input arithmetic mean > 0 ,
error factor > 1 .

LHS APPLICATION

Prague metro line A Extension is 6134 m long and comprises of four stations, initial works started in April 2010 and the construction will go into service in 2014. All underground stations (except Motol station) are mined using NATM. The running tunnel from Vypich to Motol (two tracks) is excavated using NATM, the other tunnels (single track) are driven by two tunneling machines of EPBS type. Above described LHS method was applied to verify a change in the Petřiny station construction concept. The Petřiny station is the single vault mined station with cross section area of 266 m² and the length of 217 m – Fig. 8. The station excavations sectioning consists of two side-wall drifts and one central core due to very difficult geological conditions. The side-wall drifts have cross section area of 70 m² and are sub-divided into top heading, bench and invert. The numerical analysis was carried out using 3D FEM (Fig. 9) by the MIDAS GTS program (Kožoušek 2010). The results of statistical study for station Petřiny are shown in Table 3. They show that final settlements of the tunnel lining will be with probability 95% between values 10,9 mm and 13,5 mm. The interval of final surface settlement above excavated station is from 5,5 mm 6,9 mm.

Table 3: Results for Petřiny station - settlement in mm

Calculation	Final surface	Final station	Left side-wall surface	Left side-wall tunnel
1	6,5	10,2	2,2	4,1
2	4,9	11,1	2,5	4,7
3	5,9	13	3,4	6,2
4	6,9	13	1,68	3,4
5	6,6	13,8	2,4	6,6
\bar{X} (average)	6,2	12,2	2,4	5,0
$\bar{X} + 2s, \bar{X} - 2s$	5,5 - 6,9	10,9 - 13,5	1,8 - 3,0	3,8 - 6,2

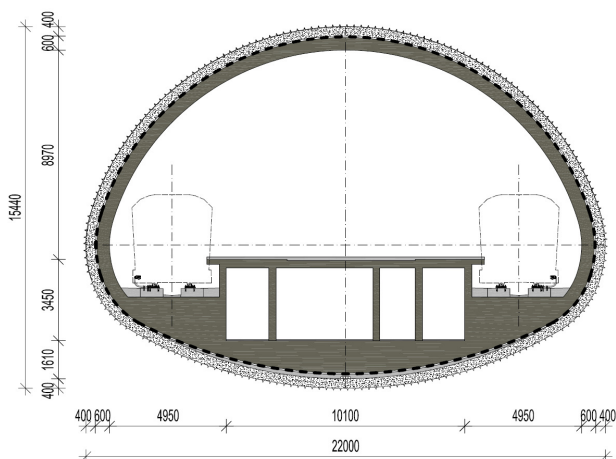


Fig. 8: Cross-section through Petřiny station

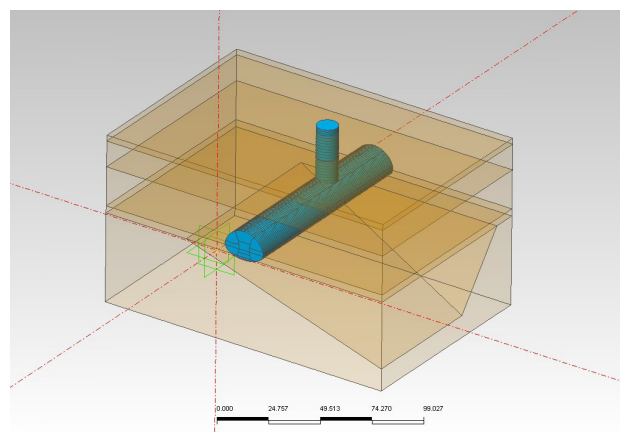


Fig. 9: 3D model of the Petřiny station in Midas - GTS

CONCLUSIONS

Latin Hypercubes method is advantageous procedure for qualified statistical evaluation of rock or soil parameters. This method makes possible radical time saving in regards to common statistical methods (Monte Carlo, estimations of probability moments etc.). Also Latin Hypercubes method is not bonded with concrete rock mass classification. LHS ensures that the ensemble of random numbers is representative of the real parameters variability. The methods of the LHS application described in this paper are crucial to the use of numerical models (common FEM) in geotechnical analyses.

ACKNOWLEDGEMENTS

The authors wish to express their gratitude to Technology Agency of the Czech Republic for research project TA01011816 “Probabilistic Evaluation of the Rock Mass Input Parameters and Underground Structures Reliability Analysis with Numerical Method Utilization” for allowing to develop this article.

REFERENCES

- Bažant, Z.P., Kwang, L.L. (1985), *Random Creep and Shrinkage in Structures*: Sampling Journal of Structural Engineering, ASCE, Vol. 111, No. 5, pp. 1113-1134
- Et al. (2001) *Spolehlivost konstrukcí : Téma Cesta k pravděpodobnostnímu posudku bezpečnosti, provozuschopnosti a trvanlivosti konstrukcí*. Ostrava. Dům techniky, ISBN80-02-01410-3 (in Czech)
- Harr, M.E. (1987), *Reliability-based design in civil engineering*. New York: McGraw-Hill.
- Hoek, E. (1989), *A limit equilibrium analysis of surface crown pillar stability*. In Surface crown pillar evaluation for active and abandoned metal mines, (ed. M.C. Betourney), 3-13. Ottawa: Dept. Energy, Mines & Resources Canada.
- Kožoušek, A. (2004) *Modelling of the underground station using 3D FEM*. CTU in Prague, 2010 – diploma thesis (in Czech)
- Mahtab, M.A., Grasso, O. (1992), *Geomechanics Principles in the Design of Tunnels and Caverns in Rock*. Elsevier, ISBN 0-4440883088
- McMahon, B. K. (1971), *A Statistical method for the design of rock slopes*. Proceedings, 1st Australia - New Zealand Conference on Geomechanics, Melbourne, Australia, Vol. 1, pp. 314-321.
- Pine. R.J. (1992), *Risk analysis design applications in mining geomechanics*. Trans. Inst. Min. Metall. (Sect.A) 101, 149-158
- Rosenbleuth, E. (1981), *Two-point estimates in probabilities*. J. Appl. Math. Modelling 5,
- Trilobyte. (2012), Statistical software QC – Expert, www.trilobyte.cz (5/2012)