Proceedings of the 3rd International Symposium on Uncertainty Quantification and Stochastic Modeling
D.A. Rade; A. M. G. Lima (Editors), ABCM, Maresias, SP, Brazil, February 15-19, 2016

Comparison of Sampling Schemes in Asymptotic Sampling

Magdalena Šmídová[1], Miroslav Vořechovský [2]

[1,2] Institute of Structural Mechanics, Faculty of Civil Engineering, Brno University of Technology, Veveří 331/95, 602 00 Brno, Czech republic, e-mail: {smidova.m@fce.vutbr.cz, vorechovsky.m@fce.vutbr.cz}

Abstract: This article deals with the possibility to use Asymptotic Sampling (AS) for estimation of the failure probability. The AS algorithm requires samples of multidimensional Gaussian random vector. There are many alternatives how to obtain such a sample and selection of the sampling strategy influences the performance of AS method. Three reliability problems (testing functions) are selected to test AS. First, the functions are analyzed using AS in combination with Monte Carlo designs and LHS designs optimized using Periodic Audze-Eglājs (PAE) Criterion. Afterwards, the same set of problems has been solved without AS procedure by direct estimation of failure probability. All the results are also compared with the exact value of the failure probability.

Keywords: Asymptotic Sampling, Failure Probability, Monte Carlo (MC), Latin Hypercube Sampling (LHS), Periodic Audze-Eglājs (PAE)

INTRODUCTION

Nowadays, European Design Standards for design of building structures allow for various ways to check the reliability of the designed structure. Apart from the commonly used approach using partial safety factors it is also possible to use the full probabilistic approach. In this case uncertain variables entering the problem (e.g. imposed load, dimensions of the structural elements or material characteristics) are defined as random variables with certain probability distribution. Usually, the probability of failure of the structure is required. Due to the complexity of the problem it is often impossible to solve the problem analytically (even with given probability distribution of inputs). Therefore, certain number of simulations are performed (Lemaire, 2009) and the subsequent statistical analysis of the simulations approximates the evaluation of the integrals corresponding to the quantities in question.

The simulations are generally performed according to a prearranged design plan. This plan defines values of all variables for each simulation of the experiment. One of the possibilities how to prepare such a design plan is to use Monte Carlo (MC) method (Lemaire, 2009; Montgomery, 2012). This method covers the space of input random variables (sampling space) randomly taking into account only the probability distributions of individual input variables. Considering very low desired probabilities of failure P_f (in case of structural problems around 10^{-5}) using MC method disproportionally many simulations would be necessary (approximately $10/P_f$). Higher number of simulations means greater computational demands which, in case of complex structures, might lead to insolvability of the problem in real time.

Various methods are being developed in order to decrease the necessary number of simulations. One of the commonly used Monte Carlo type methods is Latin Hypercube Sampling (LHS; Conover, 1975). Compared to the crude MC, this method improves the uniformity of representation of the marginal input variables with respect to probability. For certain type of problems, it allows to use lower number of simulations preserving high quality of the final estimates (variance reduction). Designs of experiments prepared by MC or LHS methods can be further optimized taking into account various criteria selected in view of required characteristics of the final design (Vořechovský and Novák, 2009; Fang and Ma, 2001; Fang, Ma and Winker, 2000).

Even more effective use of prepared designs is supposed to be achieved by the following special procedures. This article will describe and exploit the procedure of Asymptotic Sampling (AS; Bucher, 2009) combined with MC designs as well as LHS designs optimized using PAE criterion (Vořechovský and Eliáš, 2015). The resulting estimates of failure probability in pre-selected functions will be compared with the exact value (in the case when it is possible to obtain it analytically) or an estimate gained performing a very high number of simulations by crude MC.

Testing functions

The estimate of the probability of failure using AS has been performed on three testing functions that supposedly represent an engineering problem. The failure occurs if the function value is less or equal to zero. All the input variables have standard Gaussian distribution (their distribution function will be denoted as Φ throughout the paper).

Sum1D

The first function is summation of one random input variable with a constant. The function is defined as

$$g_1(x) = x + 1. \tag{1}$$

The failure occurs when $g_1(x) \le 0$, which is valid if the value of the input random variable $x \le 1$. Therefore, the safety index $\beta = 1$ and the probability of failure $P_f = \Phi(-\beta) = \Phi(-1) \doteq 0.1587$.

Sum2D

The second function is the sum of two input random variables, that is, the summation in 2-dimensional space of standard Gaussian variables

$$g_2(x_1, x_2) = x_1 + x_2 + 4. \tag{2}$$

The failure domain in this problem is depicted in Fig. 1, left (see the shaded area below the line $g_2 = 0$).

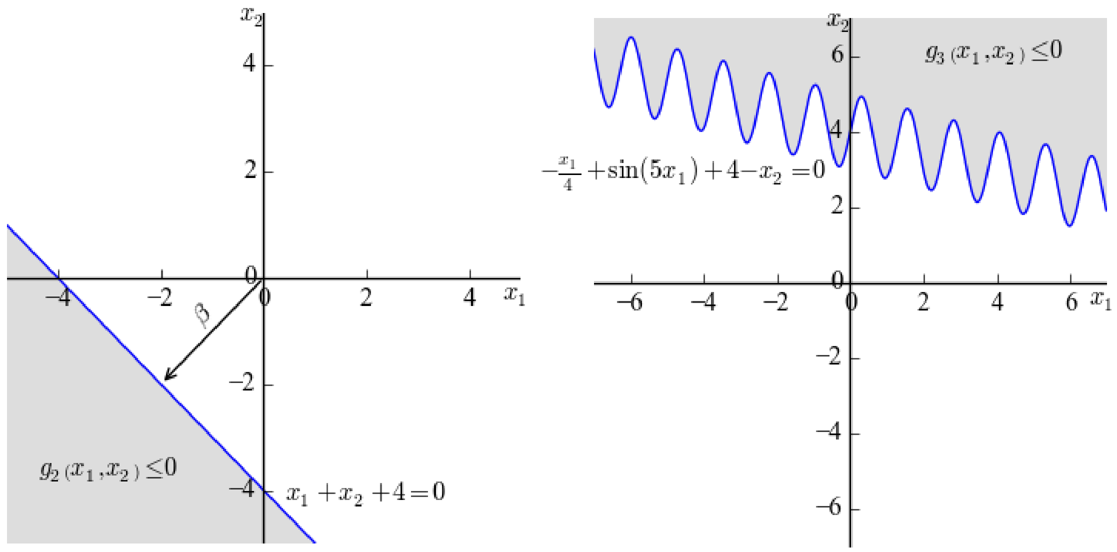

Figure 1 – Failure space for functions sum2D (left) and sin2D (right).

The exact value of safety index $\beta = 2\sqrt{2}$ follows from a simple geometry and it corresponds to the failure probability $P_f = 2.3339 \cdot 10^{-3}$.

Sin2D

The last function is again defined in the space of two standard Gaussian input variables

$$g_3(x_1, x_2) = -\frac{x_1}{4} + \sin(5x_1) + 4 - x_2. \tag{3}$$

Its failure space is bounded by a wavy boundary, see Fig. 1, right.

The accuracy of the estimate obtained using AS will be verified by comparison of the results with an estimate from crude MC experiment with 10^9 simulations. This experiment provides the failure probability estimate $P_f = 4.1508 \cdot 10^{-4}$ corresponding to the index $\beta = 3.3425$.

Asymptotic Sampling (AS)

AS, as presented by Bucher (2009; 2015; Sichani, Nielsen and Bucher, 2011), should serve to decrease the necessary number of simulations while solving real problems with low probability of failure and high number of random input variables. It exploits the relation between the standard deviation of input variables σ and the probability of failure P_f.

P_f is (for this procedure more appropriately) expressed by safety index β that can be evaluated

$$\beta = -\Phi^{-1}\left(P_f\right). \tag{4}$$

The actual distribution function of random vector featured in the problem can be mapped onto standard Gaussian random vector with independent marginal using e.g. the Nataf transformation (Nataf, 1962).

By artificial increase of the variance of individual underlying Gaussian variables, the total probability failure is increased, therefore it can be evaluated with lower number of simulations. This variance increase is performed repeatedly in several steps and in this way, few pairs of f and β are found, considering f as the inverse of standard deviation, i.e.

$$f = \frac{1}{\sigma}. \tag{5}$$

The increase of variance (and in the same time also its square root, which is the standard deviation) and concurrent increase if the probability of failure is shown in Fig. 2 for the sum1D function. The probability of failure is highlighted by hatching in this figure.

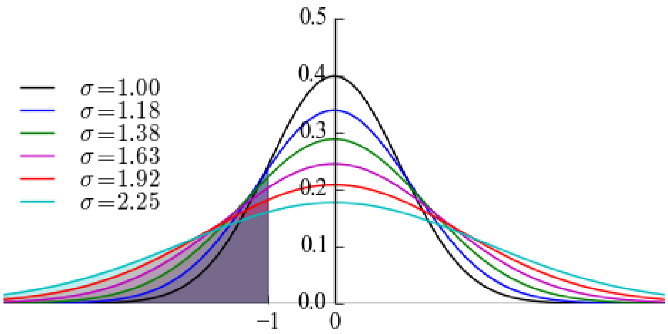

Figure 2 – Subsequent steps of the Asymptotic Sampling for the function sum1D (P_f is illustrated by hatching).

When enough pairs of f and β are obtained, it is possible to extrapolate (e.g. by linear regression) the value of the failure probability of the real problem with the initially defined standard deviation of marginal random input variables. According to Bucher (2015), a generally suitable regression model is a function:

$$\beta(f) = A \cdot f + \frac{B}{f^{p}}. \tag{6}$$

where p is an arbitrary positive number and its value may be adjusted to suit the problem in question. In this work p will be set to the value 1. A and B are regression parameters of the selected model.

AS requires to set certain parameters in advance. The first of them is the series of artificially increased standard deviation in subsequent steps of the procedure, then the number of simulations in each step in which the (f, β) pair is sought, the number of minimum failure simulations in a step to consider it as a valid step and the total number of pairs (f, β) that are to be used for extrapolation. Setting of this parameters for assessed functions is described in section Results.

Periodic Audze-Eglājs (PAE) criterion

Individual steps of AS can exploit simulations prepared by the crude MC method as well as for example optimized design plans. This article compares the results of AS performed on MC designs with those performed using LHS designs optimized with respect to PAE criterion (Šmídová *et al.*, 2015).

This criterion is derived from Audze-Eglājs criterion (Audze and Eglājs, 1977) and it is based on the analogy between experiment simulations represented by points in hypercube $[0,1]^{N_v}$ (N_v – number of random input variables) and points in space that repel each other with a repulsive force dependent on the distance between each pair of points in the design. The potential energy of the system is a sum of potential energies for all the pairs of points in the design plan

$$E^{\mathrm{PAE}} = \sum_{i=1}^{N_s} \sum_{j=i+1}^{N_s} \frac{1}{\overline{\overline{L}}_{ij}^{2}}. \tag{7}$$

In this formula, N_s stands for the number of simulations in the design and $\overline{L_{ij}}$ represents the Euclidean distance between point *i* and point *j* or its periodic image nearest to the point *i*, see Vořechovský and Eliáš (2015) for details. The optimization of the design plan with respect to PAE criterion aims at the decrease of the potential energy of the whole system which leads to equalization of the distances between points and uniform covering of the space. The design prepared in this way decreases the risk of neglecting significant part of the design space while repeatedly evaluating certain regions of the space during analyses due to clusters of simulation points in the design domain.

Results

AS method was used for estimation of failure probability of the three functions described above.

The value of parameter f was decreased by 5 % in each step of this method (standard deviation increased by approximately 5.3 %). Each step was performed using MC design or optimized LHS-PAE design containing 128 simulations (this number was chosen since the optimized designs with this number of simulations were available in previously prepared database). To perform the regression (in order to be able to extrapolate the probability of failure of the real problem, $\sigma = 1$) at least five pairs of f and β were required, i.e. five valid steps of the AS procedure had to be evaluated while the step is considered *valid* if at least $N_{f,step}$ failures occur. $N_{f,step}$ for each function was:

- Sum1D: $N_{f,step} = 30$
- Sum2D: $N_{f,step} = 10$
- Sin2D: $N_{f,step} = 7$

Generally, a higher value of $N_{f,step}$ means that the valid steps of AS procedure are more remote from the real problem which is the more pronounced the lower is the failure probability of the solved problem. Consequently, in the case of more complex functions with lower P_f the value of $N_{f,step}$ had to be decreased compared to the initial setting.

The following graphs show the resulting estimates for individual functions. To assess the variability associated with the random number generation, the experiments were performed repeatedly (ten times) for each setting. Red and green colors represent the results of AS in combination with MC or LHS-PAE designs. The results of the valid steps of the procedure are depicted as points in the plot. These were used to get (by least squares method) regression curves for extrapolation of P_f. The extrapolated failure probabilities were statistically analyzed and the graphs show the mean value and standard deviation of their estimate (see the plot legend).

The average numbers of simulation, $N_{s,av}$, necessary for AS+MC and AS+LHS-PAE analysis were also detected. The results of MC and LHS-PAE analyses without the involvement of AS procedure are shown in the graphs. In the case of MC analysis designs with exactly $N_{s,av}$ simulation points were used. For the LHS-PAE analysis this was possible only for the function sum1D. For the functions with two input random variables the number of simulations was chosen with respect to the optimized designs available in the already existing database in such a way that the number of simulations of the optimized design is as close as possible to $N_{s,av}$.

Apart from that, the graphs also show the exact value of P_f (see the section Assessed Functions).

Sum1D

The resulting estimates obtained with sum1D function are depicted in Fig. 3.

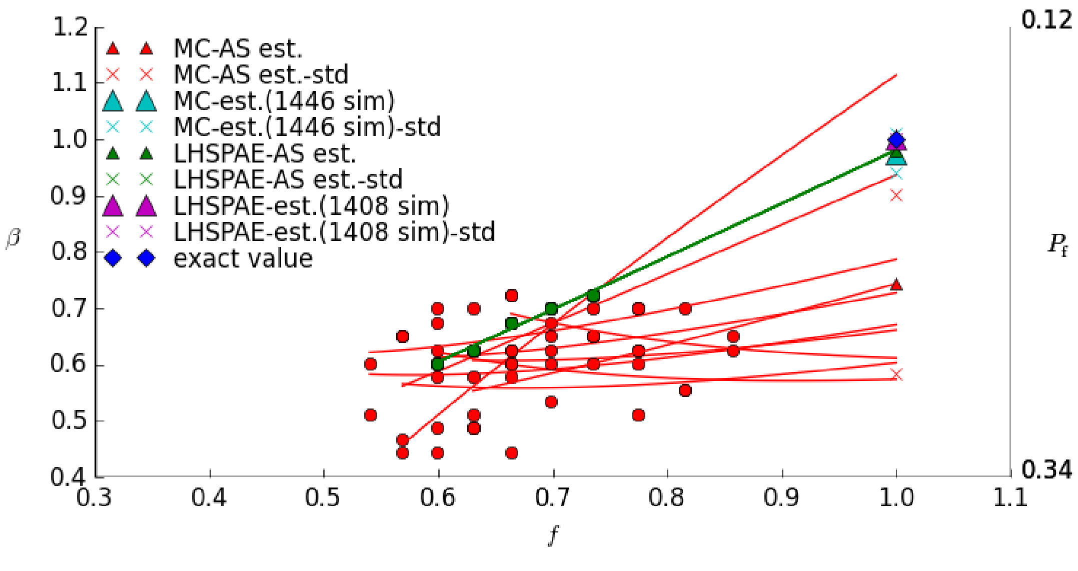

Figure 3 – Sum1D function.

Since this is a function of one random input variable, the estimates solved by LHS method are independent of the optimization criterion. This is because in the case of 1D problem the distribution of points in LHS median designs (Vořechovský and Novák, 2009) are deterministic and known in advance, as they depend only on the number of simulations in the particular design. Therefore LHS in combination with PAE criterion does not exhibit any variance and it provides relatively good estimate of P_f. Nevertheless, using LHS without AS with the same number of simulations that were needed for the whole procedure in combination with AS ($N_{s,av}$) the estimate of the result is even more accurate.

Using the crude MC method with AS leads to a large variance in the estimated β. The distance of the resulting estimate of the failure probability from the exact value is greater than one standard deviation of the set of resulting estimates. In this case it would be more advisable to perform the crude MC analysis without employing AS. The question is, whether this behavior holds also for different constant in Eq. 1.

Sum2D

The results of the function sum2D are shown in Fig. 4.

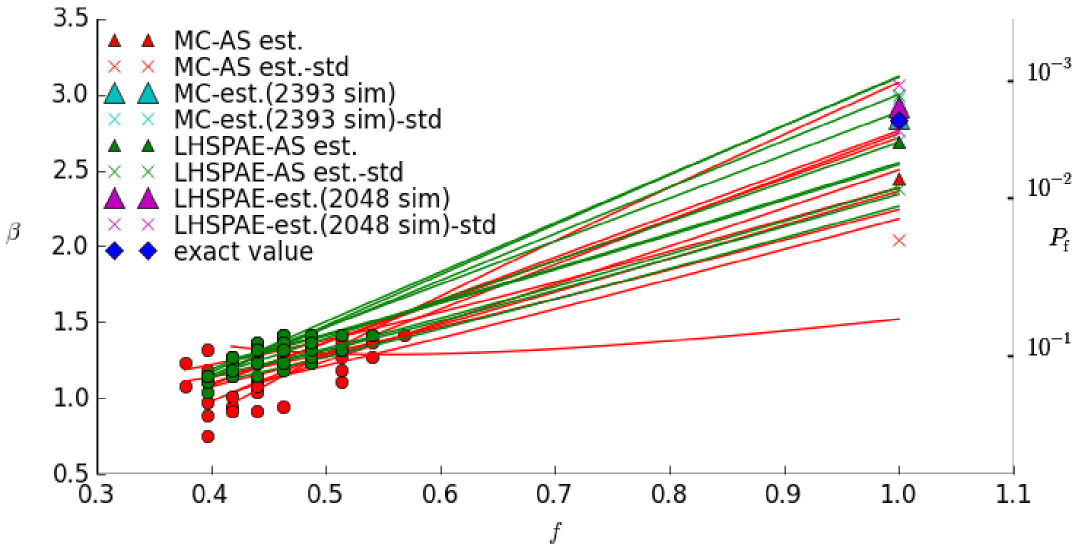

Figure 4 – Sum2D function.

As this is a function of two input random variables, the distribution of points in the design plane for optimized LHS-PAE designs is no more unique, as it depends on the heuristic optimization (Vořechovský and Novák, 2009). Thus, a certain variance of the resulting estimates is detectable for LHS designs with and also without AS procedure. Both cases provide fairly good estimates of the probability of failure.

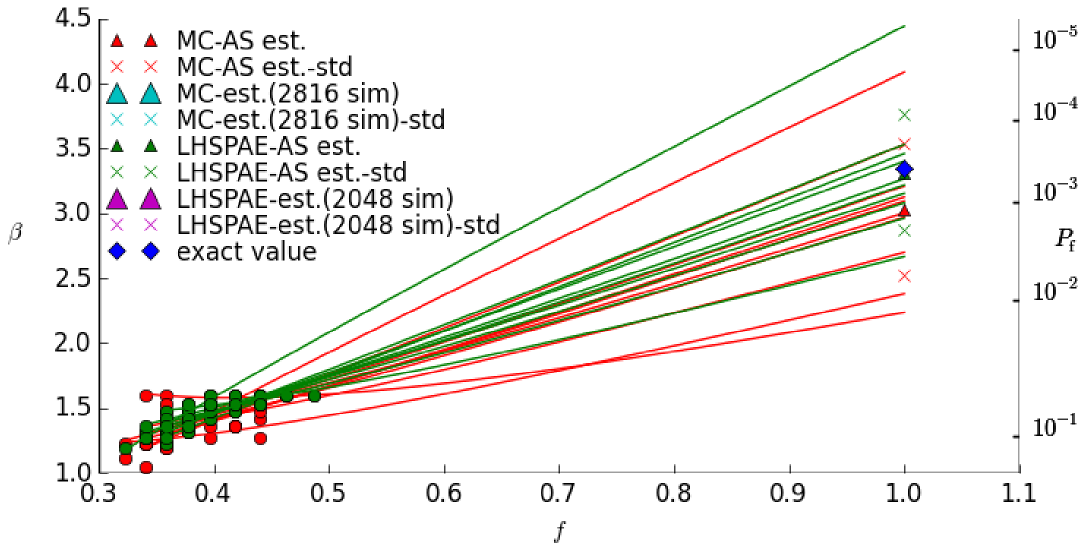

Figure 5 – Sin2D function.

As expected, AS+MC shows greater variance than AS+LHS-PAE, on the other side MC without the involvement of AS has (surprisingly) approximately the same variance as optimized LHS. This may be caused by higher number of simulations used for MC designs (optimized LHS designs with this number of simulations were not available in the database). MC without AS, similarly as for the previous function, provides the result that is closer to the exact value (using an identical number of evaluations of the function g_2) than the estimate of AS+MC and even the variance is distinctively smaller.

Sin2D

The results obtained with sin2D function are presented in Fig. 5.

The estimates of the failure probability of LHS-PAE designs combined with AS procedure are significantly better than the estimates obtained by MC+AS; moreover, the variance is smaller.

In the case of sin2D function, a certain advantage of AS approach becomes evident. Evaluating this function without AS, all the simulations in the experiment ($N_{s,av}$ = 2816) were in the safe region and the failure did not occur ($\beta = \infty$). This statement is valid for both LHS+AS and MC+AS combinations, therefore, these estimates are not drawn in the figure.

Concluding remarks

The paper presents a pilot study focused on selection of sampling scheme in Asymptotic Sampling for evaluation of failure probability. The obtained results suggest that the application of the AS procedure onto problems with high P_f (analyzable without AS with reasonably low number of simulations) does not bring any advantages. This approach, however, enables to get at least general idea about the P_f value for functions that have the failure probability so low, that it is not possible to determine P_f with the same number of simulations by usual technique without AS method.

The complete procedure would deserve further attention with regard to appropriate selection of parameters of the AS (especially the number of simulations in individual steps of the procedure, minimum number of failures in each step – $N_{f,step}$, definition of regression function for extrapolation of the real failure probability) and their influence on the quality and variance of the resulting estimates.

ACKNOWLEDGMENTS

The authors acknowledge financial support provided by the Ministry of Education, Youth and Sports under the „National Sustainability Programme I" project No. LO1408 "AdMaS UP - Advanced Materials, Structures and Technologies", and also by the project No. 16-22230S supported by the Czech Science Foundation (GAČR).

REFERENCES

Audze, P.P. and Eglājs, V.O., 1977, "New approach for planning out of experiments", Problems of Dynamics and Strengths, Zinatne Publishing House, Vol. 35, pp. 104-107.

Bucher, C., 2009, "Asymptotic sampling for high-dimensional reliability analysis", Probabilistic Engineering Mechanics, Vol. 24, pp. 504-510, ISSN 0266-8920.

Bucher, C., 2015, "Asymptotic sampling – a tool for efficient reliability computation in high dimensions", Proceedings in Applied Mathematics and Mechanics (PAMM), Vol. 15, Issue 1, pp. 549-550, ISSN 1617-7061.

Conover, W.J., 1975, "On a Better Method for Selecting Input Variables", unpublished manuscript Los Alamos National Laboratories, printed as appendix A of "Latin Hypercube Sampling and the Propagation of Uncertainty in Analyses of Complex Systems" of the authors Helton, J.C. and Davis, F.J., Sandia National Laboratories, report SAND2001-0417.

Fang, K.T. and Ma, C.X., 2001, "Wrap-Around L_2-Discrepancy of Random Sampling, Latin Hypercube and Uniform Designs", Journal of Complexity, Vol. 17, Issue 4, pp. 608-624.

Fang, K.T., Ma, C.X. and Winker, P., 2000, "Centered L2-Discrepancy of Random Sampling and Latin Hypercube Design, and Construction of Uniform Designs", Mathematics of Computation, pp. 275-296.

Lemaire, M., 2009, "Structural Reliability", in collaboration with Chateauneuf, A. and Mitteau, J.-C., Ed. Mazars, Wiley-ISTE, London, UK, ISBN 978-1-84821-082-0.

Montgomery, D.C., 2012, "Design and Analysis of Experiments", Wiley, 8th Edition, ISBN 978-1118146927.

Nataf, A., 1962, "Détermination des distributions de probabilités don't les marges sont donées", Comptes Rendus de l'Académie des Sciences.

Sichani, M.T., Nielsen, S.R.K. and Bucher, C., 2011, "Applications of asymptotic sampling on high dimensional structural dynamic problems", Structural Safety, Vol. 33, pp. 305-316, ISSN 0167-4730.

Vořechovský, M. and Novák, D., 2009, "Correlation control in small-sample Monte Carlo type simulations I: A simulated annealing approach", Probabilistic Engineering Mechanics, Vol. 23, Issue 3, pp. 452-462, ISSN 0266-8920.

Vořechovský, M. and Eliáš, J., 2015, “Improved formulation of Audze-Eglājs criterion for space-filling designs, Proc. of 12th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP 12, Vancouver, Canada, doi: 10.14288/1.0076173.

RESPONSIBILITY NOTICE

The authors are the only responsible for the material included in this paper.