Things you wanted to know about the Latin hypercube design and were afraid to ask

Felipe A. C. Viana
GE Global Research
Niskayuna, NY, 12309

Abstract

Computer models are often used in sensitivity analysis, reliability assessment, design optimization and a number of other studies which tend to require a large number of function evaluations. Very often, there is limited previous knowledge (particularly in situations like conceptual design) and engineers, designers, and analysts tend explore a large number of input variables (defined over relatively large domains). The growing power of computers enabled techniques coined for design and analysis of simulations to be applied to a large spectrum of problems and reach high level of acceptance among practitioners. Statistical modeling of computer experiments embraces the set of methodologies for generating a surrogate model (also known as metamodel or response surface approximation) used to replace an expensive simulation code [1]. The goal is constructing an approximation of the response of interest based on a limited number of expensive simulations.

With that said, careful planning of the inputs for the computer codes is one of the most crucial steps for successful statistical modeling of the simulations. This is clearly elucidated in the vast literature about experimental designs for computer experiments. Often, few statistical assumptions are made about the input/output relationship of computer models. That might be of the reasons why the initial sample is planned to cover most of the considered domain (leading to space-filling experimental designs). Among strategies coined for computer experiments, Latin hypercube designs [2], [3] are particularly popular (other strategies include maximin and minimax distance designs [4] and orthogonal arrays [5]).

This paper aims at providing a short overview of the research in Latin hypercube design of experiments highlighting the reasons of its widespread use. First, a brief discussion on the differences between physical and computer experiments is presented offering one possible explanation for the interest in strategies coined specifically for computer experiments. Next, given that Latin hypercube designs can create samples that poorly cover the input domain, optimization of the Latin hypercube is discussed. Then, a quick comparison of Latin hypercube and alternative designs for computer experiments is presented. After going through these two topics, the new practitioners should have a good understanding of why peers recommend them to use Latin hypercube designs. Then, the pitfalls of using always Latin hypercube designs for selecting experimental designs are highlighted. Finally, the research in Latin hypercube designs is situated in the current state of the art and opportunities for future work are also presented.

Keywords: Design and analysis of computer experiments, Latin hypercube sampling, space-filling designs, sequential sampling.

References

[1] TW Simpson, V Toropov, V Balabanov, and FAC Viana, "Design and analysis of computer experiments in multidisciplinary design optimization: a review of how far we have come – or not," *12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Victoria, Canada, Sep 10-12, 2008. AIAA 2008-5802.

[2] MD Mckay, RJ Beckman, and WJ Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, Vol. 21 (2), pp. 239-245, 1979.

[3] RL Iman, WJ Conover, "Small sample sensitivity analysis techniques for computer models, with an application to risk assessment," *Communications in Statistics, Part A. Theory and Methods*, Vol. 17, pp. 1749-1842, 1980.

[4] M Johnson, L Moore, and D. Ylvisaker, "Minimax and maximin distance designs," *Journal of Statistical Planning and Inference*, Vol. 26, pp. 131-148, 1990.

[5] AB Owen, "Orthogonal arrays for computer experiments, integration and visualization," *Statistics Sinica*, Vol. 2, pp. 439-452, 1992.