Experience with a Hierarchical Stratified Sampling Method for Sensitivity Analysis

Charles H. Tong
Center for Applied Scientific Computing
Lawrence Livermore National Laboratory
Livermore, CA 94551-0808 *

A deterministic computer simulation of physical models takes an input vector $X \in \Re^n$ and generates the corresponding output vector $Y \in \Re^m$. Oftentimes users are interested in characterizing the output uncertainty given input and model uncertainties. Another question that often arises is how important are the individual inputs in affecting output uncertainty. Some of these questions on uncertainty quantification and sensitivity analysis can be answered by using stochastic finite element problems or solving the corresponding sensitivity equations. More often, the simulation code is treated as a black box and in this case sampling-based uncertainty and sensitivity analyses are more appropriate.

A conventional sampling-based method is Monte Carlo sampling. An alternative approach which gives better statistical estimates is stratified sampling methods such as Latin hypercube sampling. Since its introduction by McKay, Beckman, and Conover [4], the method has been widely and successfully used. In addition, the correlation ratios calculated from sample outputs generated by replicated Latin hypercube sampling are useful to extract information about important input parameters [3].

Compared to intrusive approaches such as stochastic finite element or polynomial chaos methods, sampling-based methods have the disadvantage that it requires many simulation runs. For large scale simulations that take days on massively parallel systems, these methods may not be feasible. Moreover, frequently the number of the simulations needed to give reasonable statistical estimates are not known a-priori, as demonstrated in [5].

To address the issue of overall computational costs and the sample size problem, we study the use of a hierarchical approach to generate sample points, integrated with sensitivity analysis at each level. The idea is to create an initial set of n Latin hypercube samples using n or n/r (if replicated LHS is used with r replications) symbols, analyze the outputs based on this set of computer experiments, and refine uniformly the set of sample points if further studies are needed. An obvious requirement of this approach is the adequacy of the tests for determining the adequacy of the current sample set. It can be shown that when the computer model exhibits certain properties such as monotonicity, a

*This work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48.

simple convergence test can be prescribed. An advantage of this approach is that the sample points generated in previous steps (or levels) can be reused subsequently which helps to lower computational costs. This integrated design and analysis approach offers promises to more efficiently utilize the computational resource as well as alleviate sample size effects on output analysis.

Hierarchical and adaptive approaches have been proposed before for design optimization. For example, the multilevel coordinate search [1] has been proposed for global optimization. Other such approaches are described in [7]. In our case, it is used for sensitivity analysis with uniform refinement.

We will discuss the application of this approach to performing global sensitivity analysis in three scenarios: (1) with replicated LHS sampling and correlation ratio analysis; (2) with LHS or grid samplings given local sensitivity information; and (3) with LHS or Monte Carlo sampling and the use of response surfaces. We will demonstrate the usefulness of this approach for some model problems.

References

[1] W. Huyer and A. Neumaier, *Global Optimization by Multilevel Coordinate Search,* Kluwer Academic Publishers, 1998.

[2] J. R. Koehler and A. B. Owen, *Computer Experiments,* Department of Statistics, Stanford University.

[3] M. McKay, *Evaluating Prediction Uncertainty,* LANL technical report NUREG/CR-6311, LA-12915-MS, 1995.

[4] M. McKay, R. Beckman, W. Conover, *A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code,* Technometrics, 21(2):239-245, 1979.

[5] M. McKay, M. A. Fitzgerald, and R. J.Beckman, *Sample Size Effects when Using R^2 to Measure Model Input Importance,* LANL technical report.

[6] A. B. Owen, *Orthogonal Arrays for Computer Experiments, integration, and visualization,* Statist. Sinica 2, pp. 439-452, 1992.

[7] G. G. Wang, *Adaptive Response Surface Method Using Inherited Latin Hypercube Design Points,* Transactions of the ASME, Journal of Mechanical Designs, 2002.