SAND2016-7756R

Convergence Study in Global Sensitivity Analysis

Rebecca Harmon[1], Mohammad Khalil[2], Habib Najm[2] and Cosmin Safta[2]

[1]*Department of Chemical and Biological Engineering, Iowa State University, Ames, Iowa 50012 USA*
[2]*Sandia National Laboratories, Livermore, California 94550 USA*

August 11, 2016

Abstract

Monte Carlo (MC) sampling is a common method used to randomly sample a range of scenarios. The associated error follows a predictable rate of convergence of $1/\sqrt{N}$, such that quadrupling the sample size halves the error. This method is often employed in performing global sensitivity analysis which computes sensitivity indices, measuring fractional contributions of uncertain model inputs to the total output variance. In this study, several models are used to observe the rate of decay in the MC error in the estimation of the conditional variance, the total variance in the output, and the global sensitivity indices. The purpose is to examine the rate of convergence of the error in existing specialized, albeit MC-based, sampling methods for estimation of the sensitivity indices. It was found that the conditional variances and sensitivity indices all follow the $1/\sqrt{N}$ convergence rate. Future work will test the convergence of observables from more complex models such as ignition time in combustion.

I. INTRODUCTION

Modeling systems and physical phenomena has error abound from experimental error to the inability to completely model the system of interest. These issues are especially common in chemical kinetics and reaction modeling. In a seemingly straightforward reaction, there are tens, hundreds, even thousands of intermediate elementary reactions that make up the overall step. The reaction rate of each reaction depends on temperature, specific activation energy, and pre-exponential factor. That, in turn, leads to many more inputs into a model with more opportunities for introducing uncertainty into the output. There are two steps to combat the unfeasibly large models and the uncertain inputs that go with them: use Monte Carlo sampling with global sensitivity analysis to efficiently sample from the input space and determine which inputs significantly contribute to the uncertainty in the output. This process may then reduce the input space of a majority of the inputs, allowing the model dimensions to be reduced and more effective sampling methods such as sparse quadrature can be performed on the simplified model. In this paper, we discuss the convergence properties of Monte Carlo sampling to confirm the theoretical rates of convergence for global sensitivity analysis results. A simple and complex model are used to touch on the various sensitivity measures. It was expected that the convergence of the conditional variance and sensitivity indices would follow the $1/\sqrt{N}$ rate.

II. MONTE CARLO SAMPLING

Monte Carlo (MC) analysis is a method to randomly sample from the input space using each input's respective probability density function.[1] The Monte Carlo sampling methods known today stem from Los Alamos National Laboratory's efforts in the 1930's to estimate quantities related to controlled fission.[3] MC sampling has several key attributes that make it ideal for high-dimensional models such as combustion and other kinetic models. First, the error in estimation of an MC-based quantity of interest converges at a rate of $1/\sqrt{N}$, where N is the number of samples. As the number of samples quadruples, the error in estimation halves. Although this rate is slow, it is constant and independent of dimensions. A model may have 10 dimensions or 1,000 dimensions, and the error in estimation will still converge at the same rate. Second, MC sampling is robust and does not depend on the smoothness of model outputs. This is ideal for modeling chemical ignition, as there are sharp time gradients in many system outputs.

III. SENSITIVITY ANALYSIS

Sensitivity analysis is a tool used in a variety of fields to study the contribution and significance of input variation to variation in the output of a model.[1] An application of sensitivity analysis is to reduce the model of interest by identifying insignificant uncertain inputs and either eliminating the inputs from the model or setting their values. This is used often in chemical reaction models to reduce dimensionality from thousands of reactions to a handful that contribute to uncertainty.

Although there are various types of sensitivity analysis techniques, this research focused exclusively on global (variance-based) sensitivity analysis (GSA). The GSA method divides the uncertainty of the output among the uncertainty in the model input factors by varying all inputs simultaneously, unlike local sensitivity analysis, which varies one input at a time to determine its effects. The input uncertainty is based on the entire support and form of each input probability density function.[1]

The analysis results in sensitivity indices as discussed more thoroughly in the following section.

A. Method of Sobol’

Sobol’ sensitivity indices represent an approach to estimate fractional contributions from individual parameters of a combination of paramters to the total variance in specific quantities of interest output from a model.[1] Calculating sensitivity indices using an MC-based algorithm requires $n(d+2)$ number of model evaluations where n is number of samples and d is number of independent random variables or the stochastic dimensionality of the model. As such, Monte Carlo sampling is best suited for this method because it can handle high-dimensional problems.

Based the work done by Sobol' and eloquently shared by T. Crestaux *et al.*,[4] the following highlights an understanding of the Sobol' method. First, the input range is described by the following equation:

$$\Omega^d = \underbrace{\Omega \times \cdots \times \Omega}_{\text{d times}} \tag{1}$$

where Ω is the space of an input. Each independent random variable has its own probability density function, $p(x_i)$, as well.

The fundamental aspect of the Sobol' method is the decomposition of the model into a collection of summands, as in

$$y = f(\mathbf{x}) = f(x_1, \ldots, x_d) = f_0 + \sum_{i=1}^{d} f_i(x_i) + \sum_{1 \le i < j \le d} f_{ij}(x_i, x_j) + \cdots + f_{1,2,\ldots,d}(x_1, \ldots, x_d) \tag{2}$$

where $f_0 = E[y]$ is the expected value of y. Each of the $2^d - 1$ functions within the individual summands can be used to evaluate the conditional variances, D_u or uncertainty from one or more of the random variables.

The conditional output variance for input u can be determined using the following equation:

$$D_u = \int_{\Omega_u} E[y|x_u]^2 p(x_u) dx_u - f_0^2 = Var(E[y|x_u]) \tag{3}$$

where Var and E are the probabilistic variance and expectation of a random variable.

The total variance of the output can be defined as

$$D = \int_{\Omega_d} (f(\mathbf{x}) - f_0)^2 p(\mathbf{x}) d(\mathbf{x}) = Var(y) \tag{4}$$

Once the total variance and conditional variances of the output are calculated for each random variable input, sensitivity indices can be found. Three sensitivity indices are discussed explicitly in this paper; however, $2^d - 1$ indices exist for a given model.

First, the first-order or main-effect sensitivity index is shown below. This index measures how a single input affects the output:

$$S_i^k = \frac{D_i}{D} = \frac{Var(E[y|x_i^k])}{Var(y)} \tag{5}$$

where k = size of the ensemble at N number of samples. Next is the second-order or joint sensitivity index, which measures the joint interactive effect between two input uncertainties:

$$S_{ij} = \frac{D_{ij}}{D} - S_i - S_j \tag{6}$$

The summation of all first-order, joint, and higher-order sensitivity indices for a model add up to 1 which makes it a useful indicator.

$$\sum_{i=1}^{k} S_i + \sum_{1 \leq i < j \leq d} S_{ij} + \cdots + S_{1,2,\ldots,d} = 1 \tag{7}$$

Finally, the total-order or total-effect sensitivity index, found in the following equation, describes the whole impact of a random variable's uncertainty on the output including joint interaction with other random variables.

$$TS_i = 1 - \frac{D_{\sim i}}{D} \tag{8}$$

where $\sim i$ denotes all inputs except i. The summation of all input total-order sensitivity indices do not necessarily add to one as is the case in Eq. (7). This is due to the incorporation of higher-order, joint interactions between the inputs.

IV. EXAMPLES

In order to work up to complex models with the theory, the convergence of the Sobol' indices and its constituent parts needs to be evaluated. To verify the expected convergence rate of Monte Carlo sampling, the following error in estimation equations were used for a simple and complex model, respectively.

$$E_N = \frac{1}{K}\sum_{k=1}^{K} |\phi_N^k - \phi_{exact}| \tag{9}$$

where ϕ_N^k = quantity of interest for k^{th} ensemble at N samples and ϕ_{exact} = exact value of quantity of interest.

$$E_N = \frac{\sum_{k=1}^{K} |\phi_N^k - \phi_{2N}^k|}{k} \tag{10}$$

Eq. (9) relies on knowing the analytical, exact solution to the quantity of interest. This becomes increasingly difficult the more complex a model, thus Eq. (10) relies on self-convergence. Each quantity of interest at various N is with respect to the same reference value— the exact solution. Finding the difference between two quantities of interest cancels out the unknown reference making it ideal for black-box or complex models.

A. Simple model

A simple model was used to examine the first-order sensitivity index and the conditional variance with respect to input x_1. This model is given by

$$y = f(x) = x_1 + x_2 \tag{11}$$

where $x_1 \sim$ N(0,2^2) and $x_2 \sim$ N(0,1^2)

An ensemble size, k, of 100 was used with N (number of samples) of {100, 200, 400, 800, 1600, 3200, 6400, 12800, 25600, 51200, 102400, 204800, 406900}. Eq. (9) was used after the analytical solutions to the quantities of interest were found.

Figures 1, 2, and 3 all relate to the simple model and plot the Eq. (9) at each number of samples compared to a $1/\sqrt{N}$ trend. Figure 1 shows a fairly smooth Monte Carlo trend for the conditional variance with respect to input x_1 with only 100 realizations in the ensemble.

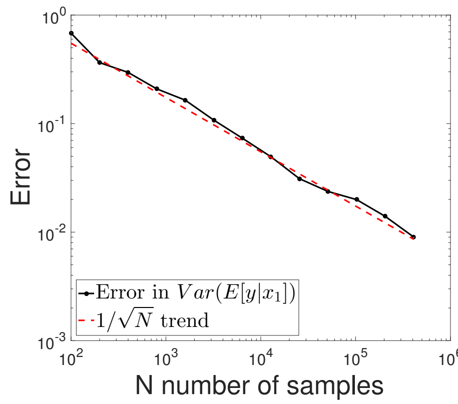

FIG. 1. Error in estimating the conditional variance of x_1 of a simple model as a function of N number of samples.

Similarly, the total variance in Figure 2 is mostly constant with a local dip at 25600 samples, due to random sampling error. The total variance was plotted as a check due to its simplicity in calculating the quantity of interest.

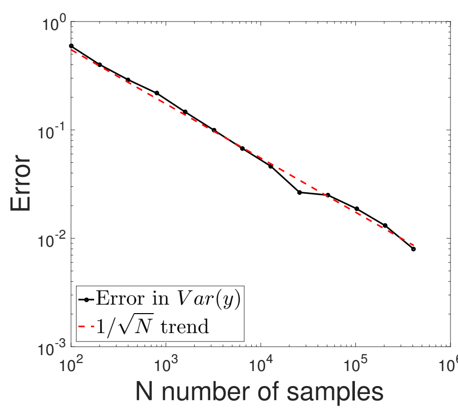

FIG. 2. Error in estimating the total variance of simple model as a function of N number of samples.

The error in estimating the first-order sensitivity index for x_1 also closely follows the MC convergence rate as the number of samples increases.

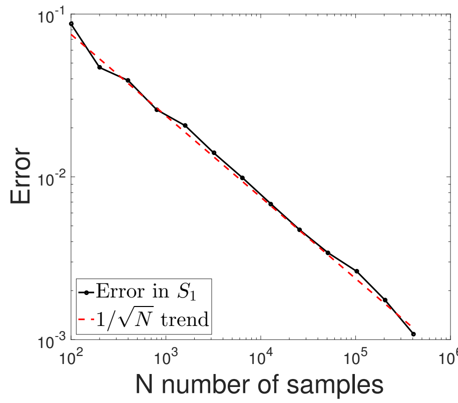

FIG. 3. Error in estimating the first-order sensitivity index of simple model input x_1 as a function of N number of samples.

B. Complex model

After a successful trial with the simple model and the first-order sensitivity indices, a slightly more complex model is used, as seen in the following equation:

$$y = f(x) = x_1 + x_2 + (x_1 * x_2) \tag{12}$$

where $x_1 \sim$ N(0,2) and $x_2 \sim$ N(0,1)

The final term adds a joint interaction between the two inputs, x_1 and x_2, allowing for a study of the second-order and total-order sensitivity index error convergence. The same ensemble size, k, of 100 and sample size range was used for the complex model.

Using Eq. (10), Figure 4 clearly shows a $1/\sqrt{N}$ convergence rate for the joint sensitivity index between x_1 and x_2. Figure 5 plots the error in estimating the total sensitivity index for x_1 and also converges at a steady $1/\sqrt{N}$ rate.

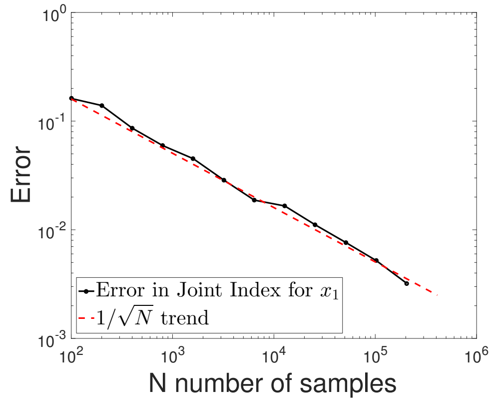

FIG. 4. Error in estimating the joint sensitivity index of complex model inputs x_1 and x_2 as a function of N number of samples.

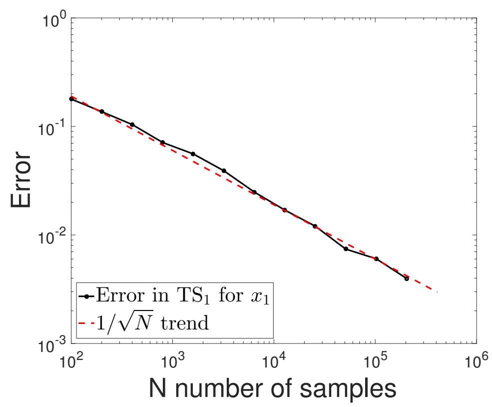

FIG. 5. Error in estimating the total-order sensitivity index of complex model input x_1 as a function of N number of samples.

After studying the convergence of error in estimating Monte Carlo-based sensitivity indices and the conditional variance, the expectation that each quantity of interest would converge at a Monte Carlo sampling rate of $1/\sqrt{N}$ holds.

V. CONCLUSION

Global sensitivity analysis using Monte Carlo sampling is an invaluable tool in model evaluation and simplification. This study of convergence verifies that the methods of sampling and sensitivity analysis follow what is expected from general Monte Carlo sampling. Future work will continue to push more complex chemical kinetic models and investigate the sensitivity of various observables from the models such as peak concentrations and temperature. There is more to explore as models become more elaborate.

ACKNOWLEDGEMENTS

R.H. wishes to thank Jonathan Wang for his assistance and input throughout the project. Support for this work was provided by the U.S. Department of Energy Office of Science Student Undergraduate Laboratory Internship (SULI) program. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94-AL85000.

References

[1] A. Saltelli, K. Chan, and E. M. Scott, Hitchhiker's guide to sensitivity analysis, in *Sensitivity Analysis*, (West Sussex, England: Wiley: 2000), Sect. 4, Chap. 2, p. 21.

[2] A. Saltelli, K. Chan, and E. M. Scott, Method of Sobol', in *Sensitivity Analysis*, (West Sussex, England: Wiley: 2000), Sect. 3, Chap. 8, pp. 174-5.

[3] R. L. Harrison, in *Introduction To Monte Carlo Simulation: AIP conference proceedings PMC 1 Jan 2011*, (National Institutes of Health 5 Jan 2010), 1204 pp. 17?21.

[4] T. Crestaux, O. Le Maître, and J.-M. Martinez, "Polynomial chaos expansion for sensitivity analysis" Reliability Engr. and Sys. Safety. **94**, 1161-1172 (2009).