Emulating the response distribution of stochastic simulators

X. Zhu
ETH Zürich

Supervisor(s): Prof. Dr. B. Sudret (ETH Zürich)

Ph.D. expected duration: Oct. 2017 - Sep. 2021

Address: Chair of Risk, Safety and Uncertainty Quantification, Stefano-Franscini-Platz 5, 8093, Zürich, Switzerland

Email: zhu@ibk.baug.ethz.ch

Abstract: With increasing demand on the functionality of complex engineering systems, computational models, a.k.a. simulators are widely used to simulate various operational scenarios. This allows engineers to assess the reliability, control the risk and optimize the behavior of the system *in silico*. Classical simulators are usually deterministic: repeated model runs with the same input parameters give the exact same value of the output quantity of interest (QoI). In contrast, *stochastic simulators* produce different results when evaluated several times for the same inputs because of some inherent stochasticity associated wit latent (hidden) variables. In other words, the QoI corresponding to a given input is a random variable, and each model run provides a single realization of the latter. Due to this random nature, it is necessary to repeatedly evaluate the stochastic simulator with the same input parameters, (so-called *replications*), to fully characterize the probability distribution of the associated model response.

In the context of uncertainty quantification or optimization, many input parameters should be investigated to explore the input space. Together with the replications, the necessary number of model runs quickly becomes intractable for expensive simulators. To alleviate this prohibitive cost, surrogate models, which behaves similarly to the original but are much cheaper to evaluate, can be built based on a limited number of model runs. Because of the internal stochasticity, surrogate modeling approaches that have been successfully developed for deterministic simulators are not directly applicable.

In the past decade, large efforts have been dedicated to the development of methods that estimate some characteristic values of the response random variable, namely the mean, variance and quantiles. When it comes to emulations of the entire response distribution, mainly replication-based methods [1, 4] can be found in the literature. The idea is to use replications to characterize the response distribution through a few parameters which are then treated as outputs of a deterministic simulator. These parameters can be emulated as functions of the input variables by conventional surrogate models such as Gaussian processes and polynomial chaos expansions (PCE).

In this communication, we present two surrogate models that we recently developed for emulating the response distribution which do no require replications. The first approach is called *generalized lambda model* (GLaM) [2]. In this method, we propose using generalized lambda distributions (GLD) to represent the output probability distribution density function (PDF). The reason for this choice is that GLD is a highly flexible parametric family. With a suitable set of parameters, it can well approximate many conventional unimodal distributions, e.g., normal, lognormal, Student's t and extreme value distributions. Following the GLD approximation of the response PDF, the four distribution parameters are represented by PCE. Therefore, the model parameters of a GLaM are *in fine* the coefficients of the PCE, which can be calibrated from data by the maximum likelihood estimation [2].

The second approach is called *stochastic polynomial chaos expansions*, which is an extension of PCE that allows us to emulate stochastic simulators. Here, we introduce a latent variable and an additional Gaussian noise, on top of the well-defined input, to represent the randomness in the model output. As a result, for a given set of input parameters, the output of such a surrogate is a function of the latent variable with an additive Gaussian noise, thus a random variable. To construct such a surrogate model, we propose using maximum likelihood estimation to calibrate the coefficients associated with the polynomial chaos basis. The variance of the noise variable is a hyperparameter estimated by cross-validation.

Both the GLaM and the stochastic PCE are more flexible than restrictive parametric models like generalized linear models but less versatile than fully non-parametric models. Because of the use of GLD, GLaM should be mainly used for stochastic simulators with unimodel PDFs. In contrast, stochastic PCE remains applicable to simulators with multi-modal response distributions. We illustrate the performance of the two surrogate models by comparing with the state-of-the-art kernel estimator on various examples in epidemiology and mathematical finance. Finally, we present our recent developments on global sensitivity analysis for stochastic simulators using the proposed surrogates [3].

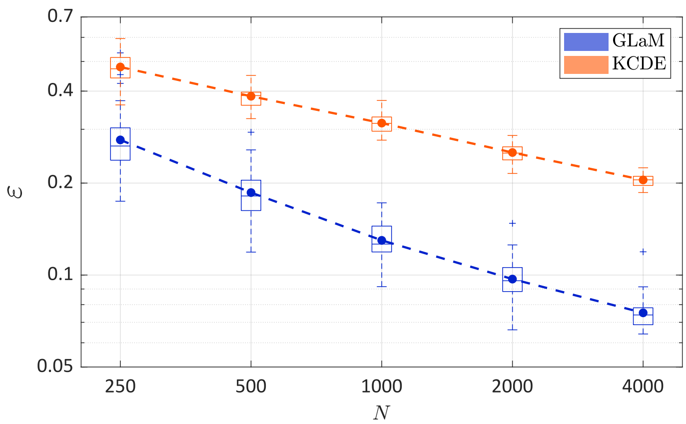

Figure 1: Comparison of the generalized lambda model (GLaM) and the kernel estimator (denoted by KCDE) on the SIR model in epidemiology

References

[1] V. Moutoussamy, S. Nanty, and B. Pauwels. Emulators for stochastic simulation codes. *ESAIM: Mathematical Modelling and Numerical Analysis*, 48:116–155, 2015.

[2] X. Zhu and B. Sudret. Emulation of stochastic simulators using generalized lambda models. *SIAM/ASA J. Unc. Quant.*, 2020. submitted.

[3] X. Zhu and B. Sudret. Global sensitivity analysis for stochastic simulators based on generalized lambda surrogate models. *Reliab. Eng. Sys. Safety*, 2020. Submitted.

[4] X. Zhu and B. Sudret. Replication-based emulation of the response distribution of stochastic simulators using generalized lambda distributions. *Int. J. Unc. Quant.*, 10(3):249–275, 2020.

Short biography – Xujia Zhu received his engineer degree from Ecole Polytechnique (France) in 2015. He also holds a MSc in computational mechanics from the Technical University of Munich. Since October 2017, he is a Ph.D. student at the Chair of Risk, Safety and Uncertainty Quantification with the thesis entitled "*Surrogate modelling for stochastic simulators using statistical approaches*" funded by the Swiss National Science Foundation.