Emulating the response PDF of stochastic simulators using sparse generalized lambda models

X. Zhu
ETH Zürich

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

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

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

Email: zhu@ibk.baug.ethz.ch

Abstract:

With increasing demands on the functionality of structures, more and more complex interdependent infrastructures and networks are developed in engineering. Design and maintenance of such systems require advanced computational models (a.k.a. simulators) to assess the reliability, control the risk and optimize the behaviour of the systems. Classically, numerical models are *deterministic*, meaning that repeated model evaluations with the same input parameters produce exactly the same output quantity of interest (QoI). In contrast, *stochastic simulators* provide different results when run twice with the same input values. In other words, the QoI of a stochastic simulator is a random variable for a given vector of input parameters. The case study that fosters this research work is the structural design of wind turbines: the input of the simulator is among others, a three-dimensional wind field, which is macroscopically defined only with a few parameters. A single realization of those parameters leads to different realizations of the wind field, and thus to different structural performance.

In the context of optimization or uncertainty quantification, surrogate models are often used to alleviate the computational burden. Deterministic surrogate methods have been successfully developed in the past two decades, yet they cannot be directly applied to emulate stochastic simulators due to the random nature of the latter. To build stochastic emulators, two categories of methods can be found in the literature. The first one focuses on estimating some statistical scalar quantities, *e.g.* mean and variance [2]. The second category aims to estimate the response distribution function but requires a large size of the data set, especially many replications of the runs of the simulator to capture the intrinsic stochasticity of the response [3]. In this work, we introduce a novel approach that does not require replications to build a sparse surrogate model that predicts the response probability density function (PDF) for any input parameter set.

For given input parameters $\boldsymbol{X} = \boldsymbol{x} \in \mathbb{R}^M$, we choose to approximate the PDF of the QoI $Y(\boldsymbol{x})$ using the four-parameter generalized lambda distribution (GLD) [1]. Following this setting, the distribution parameters $\boldsymbol{\lambda}$ are functions of the input parameters, *i.e.* $\boldsymbol{\lambda}(\boldsymbol{X})$. Under certain conditions, these functions can be represented by polynomial chaos (PC) expansions [4], and the coefficients associated with the PC basis functions are the model parameters to be estimated from data. In summary, the model is expressed as follows:

$$Y(\boldsymbol{X}) \sim GLD(\lambda_1(\boldsymbol{X}), \lambda_2(\boldsymbol{X}), \lambda_3(\boldsymbol{X}), \lambda_4(\boldsymbol{X})) \tag{1}$$

$$\lambda_s(\boldsymbol{X}) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} a_{s,\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{X}) \;\; s = 1, 2, 3, 4 \tag{2}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_M)$ denotes the multi-index defining the PC basis function $\psi_{\boldsymbol{\alpha}}(\boldsymbol{X})$ and $a_{s,\boldsymbol{\alpha}}$ is the associated coefficient with respect to $\lambda_s(\boldsymbol{X})$.

When fitting the GLD model to data, truncated series expansions using a finite set of multivariate orthogonal polynomials $\{\psi_{\boldsymbol{\alpha}}, \alpha \in \mathcal{A}_s\}$ is used for each $\lambda_s(\boldsymbol{X})$. If prior knowledge is available to fix $\mathcal{A}_s$, we propose the maximum likelihood estimation to estimate the model parameters $\boldsymbol{a}$ without the need for replications. However, due to the complexity of the GLD formulation, maximizing the likelihood can be time consuming with large data sets. Here, we derive analytically the gradient and Hessian matrix of the likelihood function with respect to $\boldsymbol{a}$ to efficiently apply derivative-based optimization algorithms. In the case of unknown $\mathcal{A}_s$, the classical "full" PC approximation cannot be applied due to the so-called *curse of dimensionality*, in the sense that the basis size increases exponentially with increasing input dimension or polynomial degree [4]. To overcome the difficulty, we propose a stepwise algorithm to adaptively construct a sparse PC approximation for each $\lambda_s(\boldsymbol{X})$. The method mainly consists of three steps: estimation of the conditional mean and variance, forward selection and backward elimination.

The performance of the proposed method is illustrated in various analytical examples. Further applications to wind turbine simulations is currently in progress.

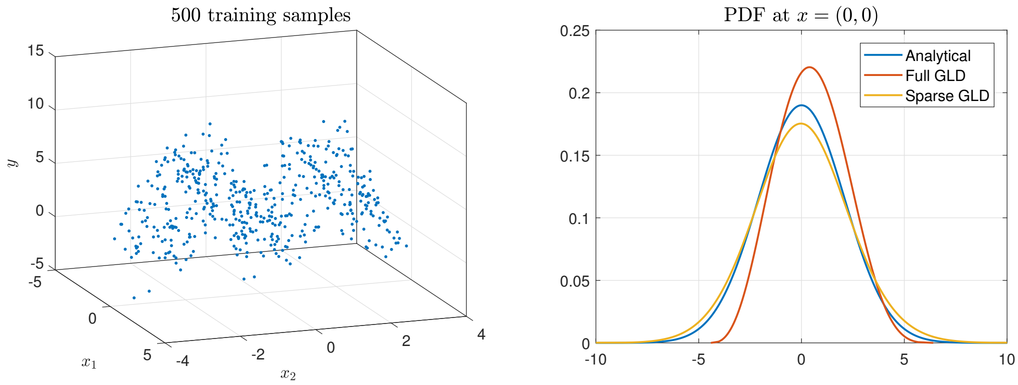

Figure 1: On the left: training data. On the right: comparison between the GLD model built with full PC approximations of $\boldsymbol{\lambda}(\boldsymbol{X})$ (denoted by full GLD) and the sparse GLD model. This toy example has normal distribution as the analytical solution with its mean and variance depending on the two-dimensional input parameters.

References

[1] M. Freimer, G. Kollia, G. S. Mudholkar, and C. T. Lin. A study of the generalized Tukey lambda family. *Comm. Stat. Theor. Meth.*, 17:3547–3567, 1988.

[2] A. Marrel, B. Iooss, S. Da Veiga, and M. Ribatet. Global sensitivity analysis of stochastic computer models with joint metamodels. *Stat. Comput.*, 22:833–847, 2012.

[3] V. Moutoussamy, S. Nanty, and B. Pauwels. Emulators for stochastic simulation codes. *ESAIM: Math. Model. Num. Anal.*, 48:116–155, 2015.

[4] B. Sudret. *Polynomial chaos expansions and stochastic finite element methods*, chapter 6, pages 265–300. Risk and Reliability in Geotechnical Engineering. Taylor and Francis, 2015.

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.