Surrogating the response PDF of stochastic simulators using semi-parametric representations

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:

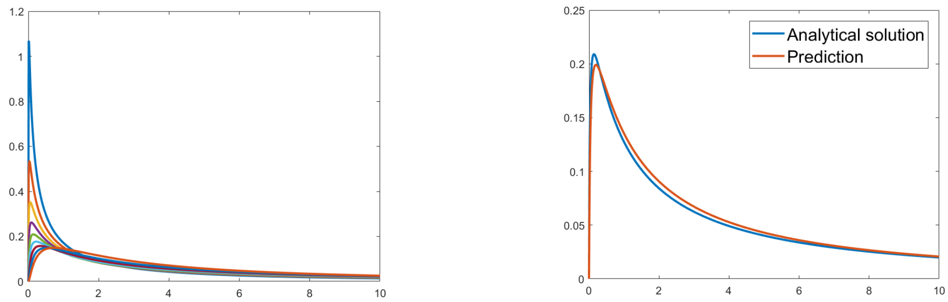
Different from the *deterministic* simulator where the same input parameters $\boldsymbol{x}_0 \in \mathcal{D}_X \subset \mathbb{R}^M$ have a unique corresponding output quantities of interest (QoI) $\boldsymbol{y} = \mathcal{M}(\boldsymbol{x}_0) \in \mathbb{R}^Q$, the *stochastic simulator* provides different results when running the same input $\boldsymbol{x}_0$ several times. The reason for the presence of randomness is that inside the model, some source of stochasticity is not taken into account in the input parameters. More precisely, the simulator is a deterministic function of two vectors $(\boldsymbol{x}, \boldsymbol{z})$ *i.e.* $\mathcal{M}_D : (\boldsymbol{x}, \boldsymbol{z}) \mapsto \boldsymbol{y} = \mathcal{M}_D(\boldsymbol{x}, \boldsymbol{z})$, but $\boldsymbol{z}$ are difficult to be considered due to lack of knowledge or to the model complexity. Consequently, with the latent parameters $\boldsymbol{z}$ varying in their range of definition, the output w.r.t each set of input parameters $\boldsymbol{x}$ is a random variable Y (resp. a random vector $\boldsymbol{Y}$ if $Q > 1$). Repeated runs for the same $\boldsymbol{x}$, named as *replications* in the sequel, are required to characterize the associated probability distribution.

To overcome the difficulty issued from the high computational cost in the case of deterministic codes, surrogate models (a.k.a emulators) which produce similar results to simulators w.r.t a certain measure but require much less computational time have been developed over the last decade, *e.g.* polynomial chaos expansions [5], low-rank tensor approximations [2], Gaussian processes [4]. To get rid of the computation burden encountered in the case of *stochastic simulators*, V. Moutoussany et al. [3] treated the probability density function (PDF) as a functional-valued output: the proposed method projects each PDF onto a subspace of L^2 with a finite dimension and then applies regression techniques to emulate the associated coefficients as a function of $\boldsymbol{x}$. However, due to the specific requirements for PDFs (positivity and integral equal to one) and to the algorithm employed for basis selection, this approach has certain shortcomings. T. Browne et al. [1] applied the same technique to quantile functions, which turns out to perform well in a specific application. In this project, we propose to build a framework based on the state-of-the-art surrogate modelling techniques that have been successfully applied to deterministic simulators. The main goal is to predict the probability density function (PDF) with a new set of parameters $\boldsymbol{x}$ by means of analysing the *replications* for $\{\boldsymbol{x}_i\}_{i=1\ldots N_{\boldsymbol{x}}}$ (*design points*).

Under certain regularity condition, the map $f : \boldsymbol{x} \in \Omega_{\boldsymbol{x}} \mapsto f(\cdot; \boldsymbol{x})$ is continuous w.r.t some norm $\|\cdot\|$, *i.e.* $\forall \boldsymbol{x}_0 \in \Omega_{\boldsymbol{x}} \lim\limits_{\boldsymbol{x} \to \boldsymbol{x}_0} \|f(\cdot; \boldsymbol{x}) - f(\cdot; \boldsymbol{x}_0)\| = 0$, where $f(\cdot; \boldsymbol{x})$ denotes the PDF with input parameters $\boldsymbol{x}$. We assume that several *replications* ($R \sim \mathcal{O}(100)$) exist at each design point. Depending on the nature of the data, two types of approach have been proposed:

- Y can be considered as a random field with $\boldsymbol{x}$ as index. Under the assumption that for the same replication index r the output Y with different $\boldsymbol{x}$ are a realization of the random field, we seek therefore to approximate the random field based on R realizations.
- At each design point $\boldsymbol{x}_i$, the PDF is estimated from *replications*. Some techniques can then be employed to predict the PDF with a new set of parameters.

Since the assumption for the first category is quite severe, the second approach has been firstly investigated. In the current contribution, it is assumed that the model response distributions belong to a parametric family (e.g. Gaussian, Weibull, lognormal, *etc.*). Therefore, the distribution at design points can be estimated via parametric estimation (*e.g.* maximum likelihood). Then some regression techniques, notably Gaussian processes, help emulate the variability of the distribution parameters as a function of the input parameters $\boldsymbol{x}$. In the end, the PDF with new parameters $\boldsymbol{x}$ is addressed through the estimated distribution parameters. At the second step, the parametric assumption will be relaxed by seeking to decompose the output into a set of random variables generated by some independent random variables $\boldsymbol{\Xi}$, *e.g.* in the case of scalar-valued output, $Y = \sum_{i=1}^{N_P} a_i h_i(\Xi)$ where h_i are orthogonal w.r.t the instrumental random variable Ξ *i.e.* $\int h_i(\xi) h_j(\xi) \cdot f_\Xi(\xi) d\xi = \delta_{i,j}$. Like the parametric approach, the coefficient a_i would then be extended to a new set of parameters resorting to regression techniques.



(a) PDF at design points

(b) Comparaison of the predictive PDF to the analytical solution at a new point

Figure 1: Toy example

References

[1] T. Browne, B. Iooss, L. Le Gratiet, J. Lonchampt, and E. Rémy. Stochastic simulators based optimization by Gaussian process metamodels – Application to maintenance investments planning issues. *Quality and Reliability Engineering International*, 32(6):2067–2080, 2016.

[2] M. Chevreuil, R. Lebrun, A. Nouy, and P. Rai. A least-squares method for sparse low rank approximation of multivariate functions. *SIAM/ASA J. Uncertainty Quantification*, 3(1):897–921, 2015.

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

[4] C.E. Rasmussen and C.K.I Williams. *Gaussian processes for machine learning.* Adaptive computation and machine learning. MIT Press, Cambridge, Massachusetts, Internet edition, 2006.

[5] B. Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliab. Eng. Sys. Safety*, 93:964–979, 2008.

Short biography – Xujia Zhu received his engineer degree from the Ecole Polytechnique (France) in 2015. He also holds a MSc in computational mechanics from the Technical University of Munich. In 2013, he joined PSA Peugeot Citroën for a 2-month internship and in 2015 he worked as a research intern at the research center of EDF (Electricity of France). Xujia Zhu prepared his master thesis in 2017 at the Chair for Computation in Engineering (TU Munich) and the research center of ESI group. Since October 2017, he is a PhD student at the Chair of Risk, Safety and Uncertainty Quantification with the thesis entitled *surrogate modelling for stochastic simulators.*