An alternative estimator of a threshold-exceeding probability.

LUCIE BERNARD
University Pierre et Marie Curie

Supervisor(s): A. Guyader (UPMC-LSTA), F. Malrieu (University François Rabelais Tours-LMPT) and P. Leduc (STMicroelectronics Tours)

Ph.D. expected duration: 2015-2018

Adress: University UPMC 4 Place Jussieu, 75005 Paris

Email: lucie.bernard@live.fr

Abstract:

Despite all the preliminary precautions taken before the launch of a manufacturing process to make sure it runs smoothly, some operations are complicated to control and inevitably subjected to variability (typically, succession of heavy and complex machinery, environment challenging to maintain constant over time ...). Thus, we have to take into account the fact that some design parameter, also called *factors*, can take various numerical values from one manufactured product to another. Such variations are critical because, among all the possible numerical values that factors can then take, some configurations lead to manufactured products that do not fulfil the imposed specifications. We aim at measure the impact of manufacturing process variations on the products performances through the estimate of a threshold-exceeding probability.

We consider that the product under study is characterized by $d \geq 1$ factors. A particular set of numerical values taken by factors is the realization of a random variable denoted by $\mathbf{X}$, defined over the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in the measurable space $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$, where $\mathbb{X} \subseteq \mathbb{R}^d$. Its probability distribution is denoted by $\mathbb{P}_{\mathbf{X}}$. The performance of each manufactured product is measured from the output of a numerical simulation, which consists in an evaluation of a measurable and deterministic function $g : \mathbb{X} \to \mathbb{R}$ in a given realization $\mathbf{x}$ of the random variable $\mathbf{X}$. In particular, if the output $g(\mathbf{x})$ exceeds a prescribed threshold T, then the product characterized by $\mathbf{x}$ will be not satisfy the imposed specifications and considered as non-functional. Thus, we are interested in approximating the threshold-exceeding probability p defined by

$$p = \mathbb{P}_{\mathbf{X}}(\{\mathbf{x} \in \mathbb{X} : g(\mathbf{x}) \geq T\}) = \int_{\mathbb{X}} \mathbb{1}_{g(\mathbf{x}) \geq T} \mathrm{d}\mathbb{P}_{\mathbf{X}}(\mathbf{x}),$$

typically called *probability of failure*. The reference estimation method of p is the crude Monte Carlo method. However, it is excluded in our context to use it because we suppose that g is expensive to evaluate and has no available analytical expression (g is a black-box function). As a consequence, we only have access to a restraint number of data, corresponding to the evaluations of the function g at points of the design of experiments $\mathbf{D}^n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, $\mathbf{x}_i \in \mathbb{X}$.

We then make the assumption that g is a realization of the Gaussian process $\{\xi(\mathbf{x})\}_{\mathbf{x} \in \mathbb{X}}$. We condition the prior process ξ on the evaluations of g at $\mathbf{D}^n$ and obtain the posterior Gaussian process ξ_n. This approach refers to the Gaussian process regression or *Kriging* (see e.g [7] and [5]). We can then consider that, conditionally to the observations, the probability p is a realization of a random variable $P_n : (\Omega_0, \mathcal{F}_0, \mathrm{P}_0) \to [0, 1]$ defined for all $\omega_0 \in \Omega_0$ by

$$P_n(\omega_0) = \mathbb{P}_{\mathbf{X}}(\{\mathbf{x} \in \mathbb{X} : \xi_n(\mathbf{x}, \omega_0) \geq T\}) = \int_{\mathbb{X}} \mathbb{1}_{\xi_n(\mathbf{x}, \omega_0) \geq T} \, \mathrm{d}\mathbb{P}_{\mathbf{X}}(\mathbf{x}).$$

A natural estimator of p is the expected value of P_n, that is $\mathrm{E}_0[P_n]$ (see e.g [1], [2] and [4]). However, a realization of P_n requiring a realization of the process ξ_n and an integration w.r.t the law $\mathbb{P}_{\mathbf{X}}$, the estimates of moments, measures of dispersion and quantiles, require a long computation time.

We then propose an alternative random variable R, easy to simulate because it does not involve realizations of the process ξ_n. Indeed, a realization of R only requires a realization of a standard uniform random variable and an integration w.r.t the law $\mathbb{P}_{\mathbf{X}}$. Moreover, we show that P_n is smaller than R in the convex order, i.e for all convex function φ, we have $\mathrm{E}_0[\varphi(P_n)] \leq \mathrm{E}[\varphi(R)]$. More information about the convex order and its consequences are given e.g in [6], [3] and [8]. According to this result, R has the same expected value of P_n, which provides an alternative way to estimate p. Another consequence is, for example, that the variance of R is an upper bound of the variance of P_n. We explore the properties of the random variable R and use them to provide an efficient estimation of p.

References

[1] Y. Auffray, P. Barbillon, and J-M. Marin. Bounding rare event probabilities in computer experiments. *Comput. Statist. Data Anal.*, 80:153–166, 2014.

[2] J. Bect, D. Ginsbourger, L. Li, V. Picheny, and E. Vazquez. Sequential design of computer experiments for the estimation of a probability of failure. *Stat. Comput.*, 22(3):773–793, 2012.

[3] Buerle N. and Mller A. Stochastic orders and risk measures: Consistency and bounds. 2005.

[4] J. Oger, P. Leduc, and E Lesigne. A random field model and decision support in industrial production. *J. SFdS*, 156(3):1–26, 2015.

[5] C. E. Rasmussen and C. K. I. Williams. *Gaussian processes for machine learning.* Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, 2006.

[6] Ludger Rüschendorf. *Mathematical risk analysis.* Springer Series in Operations Research and Financial Engineering. Springer, Heidelberg, 2013. Dependence, risk bounds, optimal allocations and portfolios.

[7] J. Sacks, B. S. Schiller, and W.J. Welch. Designs for computer experiments. *Technometrics*, 31(1), 1989.

[8] M. Shaked and J.G. Shanthikumar. *Stochastic Orders.* Springer Series in Statistics. Springer, New York, 2007.

Short biography – My PhD thesis is hosted by the Laboratoire de Statistique Théorique et Appliquée at University Pierre et Marie Currie and the Laboratoire de Mathématiques et de Physique Théorique at University François Rabelais. My research are funded by the company STMicroelectronics which needs to developp computational products to estimate failure probabilities. Before starting my PhD in novembre 2015, I made a Master 2 in statistical research at University of Rennes 1.