Quantile Estimation in Structural Reliability with Incomplete Dependence Structure

NAZIH BENOUMECHIARA
Université Pierre et Marie Curie – EDF R&D

Supervisor(s): Gérard Biau (UPMC), Bertrand Michel (ECN), Philippe Saint-Pierre (IMT), Roman Sueur (EDF), Nicolas Bousquet (Quantmetry, UPMC), Bertrand Iooss (EDF, IMT)

Ph.D. expected duration: July 2015 - June 2018

Adress: Laboratoire de Statistique Théorique et Appliquée, Université Pierre et Marie Curie, 4 place Jussieu, 75005 PARIS

Email: nazih.benoumechiara@upmc.fr

Abstract:

The study of some components presenting a high level of reliability requires the use of structural reliability methods. The approach is to evaluate an indicator of risk $y \in \mathbb{R}$ based on a numerical model $\eta : \mathbb{R}^d \to \mathbb{R}$ representing the behavior of the component according to input parameters $\mathbf{x} = (x_1, \dots, x_d)$. These parameters, considered as uncertain, are modeled by a vector of random variables $\mathbf{X} = (X_1, \dots, X_d)$, so that the risk indicator $Y = \eta(\mathbf{X})$ is itself a random variable. It is necessary to characterize the law of the vector $\mathbf{X}$ for any study concerning a criterion of interest $\mathscr{C}(Y)$ of the model output distribution, for example the quantile of which we are particularly interested. In many cases, the used probabilistic model can be incomplete when it is obtained from d separate sets of experimental data. In the absence of joint data for the vector $\mathbf{X}$, it is not possible to characterize its dependence structure, but only the marginal laws of $X_1, \dots, X_d$. The most common industrial practice is then to conduct the study assuming independence of these variables. Unfortunately, without theoretical justification or physical basis, this approach can lead to overly optimistic results. This was firstly shown by [2] which illustrated the influence of the correlation of random variables on a system reliability and concluded that correlation is an important information that must be taken into account in reliability calculations. Similarly, [5] indicates that the choice of the dependence structure can strongly influence $\mathscr{C}(Y)$ but confines itself to the case of two-dimensional dependencies.

The approaches we propose are intended to ensure the conservatism of the estimate of $\mathscr{C}(Y)$ by exploring a set of dependence structures. The value of $\mathscr{C}(Y)$ ultimately retained is the one considered as penalizing, i.e., the one leading to the most pessimistic estimate. The corresponding dependence structure, defined by the minimization of a stochastic quantity and data-dependent, is an extremum-estimator. For this exploration, different approaches are possible and are based on different characterizations of copulas describing the dependence structure of $\mathbf{X}$ [4]. Therefore, we chose a parametric framework based on multidimensional copulas constructed using *regular vines* (R-vines) [1]. We consider that the copula families of the R-vine are defined beforehand for each pair of variables and because some copulas have non-symmetric tail dependencies, it is important to correctly specify their orientations.

A first approach consists in minimizing the α-quantile of Y on a thin grid in the domain $\boldsymbol{\Theta}$ of the parameters $\boldsymbol{\theta}$ of the copula $C_{\boldsymbol{\theta}}$ of $\mathbf{X}$. We denote respectively $G_{\boldsymbol{\theta}}$ and $G_{\boldsymbol{\theta}}^{-1}$ as the cumulative distribution function and the quantile function of Y. The model η is considered as a black box. Therefore, the quantile $G_{\boldsymbol{\theta}}^{-1}(\alpha)$ cannot be known and is estimated given n realizations $(y^{(i)})_{i\in\{1,\dots,n\}}$ of the random variable Y. We chose the empirical quantile $\hat{G}_{\boldsymbol{\theta}}^{-1}(\alpha)$ as an estimator

of $G_{\boldsymbol{\theta}}^{-1}(\alpha)$. The minimization of this estimator is an extremum-estimator and is written as

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\operatorname{argmin}} \, \hat{G}_{\boldsymbol{\theta}}^{-1}(\alpha). \tag{1}$$

The consistency of the estimator $\hat{\boldsymbol{\theta}}$ is shown in Theorem 1.

Theorem 1 *Under regularity assumptions of the model η and the probability distribution $P_{\boldsymbol{\theta}}$, we have for a given $\alpha \in (0, 1)$ and for any $\varepsilon > 0$,*

$$P\left[\left|\hat{G}_{\hat{\boldsymbol{\theta}}}^{-1}(\alpha) - G_C^{-1\star}(\alpha)\right| > \varepsilon\right] \xrightarrow{n \to \infty} 0. \tag{2}$$

Moreover, if $\boldsymbol{\theta}_C^* \in \boldsymbol{\Theta}$ minimizes $\boldsymbol{\theta} \mapsto G_{\boldsymbol{\theta}}^{-1}(\alpha)$, then for any $h > 0$:

$$\mathbb{P}[|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_C^*| > h] \xrightarrow{n \to \infty} 0.$$

Because the set $\boldsymbol{\Theta}$ is not easy to explore continuously, we chose to approximate the set by a thin grid $\boldsymbol{\Theta}_K$ of cardinality K. Moreover, the consistency of $\hat{\boldsymbol{\theta}}_K$ is verified for a grid $\boldsymbol{\Theta}_K$ thin enough and under the same assumptions of Theorem 1.

The first approach considers a set of possible R-vine structures to determine $\hat{\boldsymbol{\theta}}$. However, the cost of this approach makes its use very limited due to the fast-growing number of R-vine structures with the dimension (e.g., 480 for $d = 5$ and $23,040$ for $d = 6$ [3]). Moreover, because the number of pairs increases with the dimension ($d \times (d-1)/2$), K should also increase to properly explore $\boldsymbol{\Theta}$. Therefore, when the number of pairs becomes too large, we propose a second approach to iteratively determine the most penalizing parameter $\boldsymbol{\theta}$. This method relies on the assumption that the influence of dependencies on the output quantile is only governed by a few pairs of variables. The aim of the method is to determine, at each iteration, the pair of variables that decreases the most the quantile and to consider the pair in the R-vine structure. The other pairs that are not considered influential are fixed to independence. Thus, this method can yield a more penalized dependence structure with a lower cost than the first approach by exploring low dimension subsets of $\boldsymbol{\Theta}$. Because the R-vine structure is composed of conditional and unconditional pairs, we chose to place iteratively the most influential pairs as unconditional. Moreover, the obtained minimum copula is greatly simplified and is only composed of some dependent pair-copulas.

References

[1] Tim Bedford and Roger M Cooke. Vines: A new graphical model for dependent random variables. *Annals of Statistics*, 30(4):1031–1068, 2002.

[2] Mircea Grigoriu and Carl Turkstra. Safety of structural systems with correlated resistances. *Applied Mathematical Modelling*, 3(2):130–136, 1979.

[3] Oswaldo Morales Nápoles. *Bayesian belief nets and vines in aviation safety and other applications*. PhD thesis, TU Delft, Delft University of Technology, 2010.

[4] Abe Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris*, 8:229–231, 1959.

[5] Xiao-Song Tang, Dian-Qing Li, Guan Rong, Kok-Kwang Phoon, and Chuang-Bing Zhou. Impact of copula selection on geotechnical reliability under incomplete probability information. *Computers and Geotechnics*, 49:264–278, 2013.

Short biography – Graduated from a master in modeling and simulation, I started a PhD in 2015 at UPMC under a CIFRE convention with EDF R&D, where I completed an internship the year prior. The present subject is motivated by industrial problems in structural reliability confronted by EDF. In my spare time, I am involved in different data science and machine learning projects.