# MascotNum2016 Conference - Inverse Problem and Dependencies

Nazih Benoumechiara
*University Pierre et Marie Curie (LSTA) – EDF R&D (MRI)*

**Supervisors:** Gérard Biau (UPMC), Bertrand Michel (UPMC), Philippe Saint-Pierre (IMT), Roman Sueur (EDF), Nicolas Bousquet (EDF), Bertrand Iooss (EDF)

**Ph.D. expected duration:** 2015-2018

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

**Email**: nazih.benoumechiara@upmc.fr

**Abstract:**

A common problem seen in multiple applications is the lack of knowledge regarding dependencies. When the dependence structure between variables is unknown (or neglected for the commodity), the study is performed by considering margins exclusively. Somehow, the independence hypothesis can be too optimistic. In order to provide a robust and conservative uncertainty analysis of the problem, the assumption of independence should be rigorously justified.

We consider $\mathbf{X} = (X_1, \dots, X_d)^T$, a $d$-dimension vector of random variables and $Y$ the scalar random variable such as $Y = g(\mathbf{X})$. We take our interest on $\gamma(Y)$, a quantity of $Y$, such as the expectation $\mathbb{E}(Y)$, the variance $\mathrm{Var}(Y)$, an $\alpha$-quantile $Q_\alpha(Y)$ or any other scalar quantity. When no information is available on the dependence structure, we have no choice but to search for the one leading to the "most penalised case". This is defined as the configuration that minimises a predefined objective-function $U$, which only depends on the dependence structure of $\mathbf{X}$ and the quantity $\gamma(Y)$. Some questions follows: can we verify that the independence hypothesis is conservative? Is there a dependence structure (other than the independent one) leading to the "most penalised case"? Moreover, can we quantify the influence of the dependence on $\gamma(Y)$? These are the questions we will try to respond.

To describe the dependence between the margins we use a copula representation, which is adapted to our problem. Indeed, the univariate margins are assumed to be accurate and only the appropriate dependence structure is missing. Thus, the joint distribution of $\mathbf{X}$ can be described using copula [5]. With $F_{X_1} \dots F_{X_d}$ the cumulative distribution margins of $\mathbf{X}$, the joint cumulative distribution can be written as

$$F_{\mathbf{X}}(\mathbf{x}) = C_{\boldsymbol{\rho}}(F_{X_1}(x_1), \dots, F_{X_d}(x_d)),$$

with $C_{\boldsymbol{\rho}} : [0,1]^d \to [0,1]$ the copula function and $\boldsymbol{\rho} \in \mathcal{S}_{\boldsymbol{\rho}}$ its vector of parameters. The copula is chosen among parametric or non-parametric families and $\boldsymbol{\rho}$ is the only variable describing it. Thus, if the margins are supposed known, $\boldsymbol{\rho}$ remains the only unknown parameter of the problem.

The objective-function $U(\boldsymbol{\rho})$ is defined such as its minimum corresponds to the "most penalised case" and we aim to find the $\boldsymbol{\rho}^*$ leading to it. The problem can be expressed as an inverse problem such as,

$$\boldsymbol{\rho}^* = \arg\min_{\boldsymbol{\rho} \in \mathcal{S}_{\boldsymbol{\rho}}} U(\boldsymbol{\rho}).$$

Solving this inverse problem leads to a new dependence structure $C_{\boldsymbol{\rho}^*}$ which is more conservative, if the problem is well posed [3]. However, in practice $\boldsymbol{\rho}^*$ is costly to compute.

This work can be structured in two successive steps:

- Find the copula $C_{\boldsymbol{\rho}^*}$ leading to the most penalised case.
- Quantify how dependencies can impact the quantity of interest and which dependencies influence the most.

To begin, we choose the $\alpha$-quantile $Q_\alpha(Y)$ as the quantity of interest $\gamma(Y)$ and we aim to minimise it. We place ourselves in a structural reliability problem where, the lower the quantile is, the more penalised we are. Because $Y = g(\mathbf{X})$, we can easily state that the quantile is related to $\mathbf{X}$ and thus to $\boldsymbol{\rho}$. We can write that $U(\boldsymbol{\rho}) = Q_{\alpha,\boldsymbol{\rho}}(Y)$ and for convenience we define $q_\alpha(\boldsymbol{\rho}) = Q_{\alpha,\boldsymbol{\rho}}(Y)$.

Because $q_\alpha(\boldsymbol{\rho})$ can be very expensive to compute, a quantile regression method can be used. Among all these methods, the quantile regression forest [1, 4] is an interesting way to estimate $q_\alpha(\boldsymbol{\rho})$ at a low computational cost. Moreover, random forests can compute importance measures by permutation [2], which can be adapted quantify the impact of dependencies. A Bayesian approach is also studied to treat the problem from a different point of view.

## References


[1] Leo Breiman. Random forests. *Mach. Learn.*, 45(1):5–32, October 2001.

[2] B. Gregorutti, B. Michel, and P. Saint-Pierre. Correlation and variable importance in random forests. *ArXiv e-prints*, October 2013.

[3] Jacques Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton university bulletin*, 13(49-52):28, 1902.

[4] Nicolai Meinshausen. Quantile regression forests. *J. Mach. Learn. Res.*, 7:983–999, December 2006.

[5] M Sklar. *Fonctions de répartition à n dimensions et leurs marges.* Université Paris 8, 1959.


**Short biography** – After a first year master in physics and a second year in modelling and simulation, I started an internship at EDF (*Management des Risques Industriels*) on the study of different methods in reliability analysis. Since July 2015, I had the chance to start my PhD under CIFRE convention with EDF. The application of this PhD is the same as my previous internship, which is on the structural reliability of an irreplaceable material exposed to extreme constraints.