Modelling dependence under constraint

R. Fischer
Université Paris-Est

Affiliation: *Department of Industrial Risk Management*, EDF R&D, 1 Avenue du Général de Gaulle, 92140 Clamart, France

Email: richard.fischer@edf.fr – **URL**: http://www.cermics.enpc.fr/~fischerr/home.html

Master: Université Paris-Dauphine

Ph.D. (2013-2016): Université Paris-Est

Supervisor(s): Prof. Jean-François Delmas (ENPC) and Prof. Cristina Butucea (UPE-MLV) and Anne Dutfoy (EDF R&D)

Abstract:

In various cases of modelling the propagation of uncertainity in complex numerical simulation schemes, the engineer needs to create a probabilistic model that takes into consideration the available statistical information (or expert judgement). In most of the cases this means that the marginal distribution of each uncertain quantity is determined, but it can also be a physical constraint between these quantities, for example an order relationship. The construction of such a model is quintessential if we want to evaluate a probabilistic reliability criterion by Monte Carlo methods. In order to estimate the expectation of a random variable derived from the uncertain quantities, or the probability that it exceeds a certain threshold, we need to generate i.i.d. samples, which requires the complete knowledge of the joint distribution of these quantities.

My thesis focuses on the modelling of multivariate random vectors where the marginal distributions of the components are given and the components verify certain deterministic constraints. As a first problem we consider ordering constraits, that is the components of the random vector are ordered almost surely. Among the possible joint distributions (if there exists any), our aim is to find the most "random" possible. To measure the uncertanity of our joint model, we use the Shannon entropy defined for a d-dimensional random variable $X = (X_1, \ldots, X_d)$ as:

$$H(X) = -\int f_X(x) \log\left(f_X(x)\right)\, dx,$$

where f_X denotes the density of X. We are looking for the joint distribution of X that maximizes this measure. From an information theory point of view, this is the distribution that adds the least amount of information in addition to the initial constraints, leading to a more realistic modelling of the random vector.

Since the marginal distributions are fixed, the theory of copula functions seems appropriate to model the dependence structure between the uncertain quantities. Copulas are multivariate distribution functions whose marginals are uniformly distributed on $[0, 1]$. Sklar's theorem states that the distribution function F of a d-dimensional random variable X with continuous marginal distribution functions F_i, $i = 1 \ldots d$ can be written as:

$$F(x_1, \ldots, x_d) = C\left(F_1(x_1), \ldots, F_d(x_d)\right)$$

where C is a uniquely determined copula. This allows us to decompose the information on the marginals and the information on the dependence, which is incorporated in the copula. Furthermore, we have the following decomposition of the entropy of X:

$$H(X) = \sum_{i=1}^{d} H(X_i) + H(U),$$

where $H(X_i)$ is the entropy of the univariate random variable X_i, and $H(U)$ is the entropy of a d-dimensional random variable U whose distribution function is C, the copula associated to X. Since the

marginals X_i are given, we can observe that $H(X)$ is maximal if and only if $H(U)$ is maximal. Therefore we turned our attention to finding the copula that maximizes the entropy under the constraints imposed by the marginals and the ordering. We have found that in the case of $d = 2$, this copula can be expressed with the help of the copula which maximizes the entropy under the constraint that its diagonal section, defined by $\delta(t) = C(t, t)$, $t \in [0, 1]$, is given. The function δ only depends on the marginals distribution functions of X. We present the explicit density of the unique optimal copula with given diagonal section δ. We give an explicit criterion on the diagonal section for the existence of the optimal copula as well as the closed formula for its entropy. We also general Using this copula we determine explicitly the distribution function of the random variable X which maximizes the entropy under the ordering and marginal constraints. For $d > 2$, we have shown that the maximum entropy copula of ordered random vectors with given marginals is related to the maximum entropy symmetric copula whose order statistics have fixed distributions. That is, if U is a random variable with distribution function C, then the distribution function of the i-th largest element of U is fixed and given by $\delta_{(i)}$ for all $1 \leq i \leq d$. As in the case of $d = 2$, we give the explicit density of the optimal copula with a closed formula for its entropy for a special class of marginal distributions. We also present the density of the random variable X which maximizes the entropy under the ordering and marginal constraints for general $d \geq 2$ for this special class of marginals.

The results will be illustrated on various choices for the marginal distributions, and an example arising from an industrial problem will also be presented.

Short biography – I'm a second-year Ph.D. student in Applied Probability from Hungary at Université Paris-Est, affiliated to the laboratories CERMICS of ENPC and LAMA of UPE-MLV. My work is a collaboration between the Department of Industrial Risk Management of EDF R&D and the aforementioned laboratories under a CIFRE contract. My studies are jointly funded by the organization ANRT and EDF R&D.