Propagation of Uncertainty involving Imprecision and Randomness

Cédric Baudrit **Didier Dubois** **Helène Fargier**
Institut de Recherche en Informatique de Toulouse
Université Paul Sabatier, 118 route de Narbonne 31062 Toulouse Cedex 4, France
e-mail: baudrit@irit.fr e-mail: dubois@irit.fr e-mail: fargier.irit.fr

Abstract

This paper presents a non exhaustive list of different methods of uncertainty propagation when the knowledge of some parameters of physical models is represented by probability measures, and others by possibility measures or belief functions.

Keywords: Fuzzy Number, Probability, Possibility, Belief function, dependence, correlation

1 Introduction

Currently, decisions pertaining to the management of potentially polluted sites very often rely on the evaluation of risks for man and the environment. This evaluation is carried out with the help of models which simulate the transfer of pollutants from a source to a vulnerable target, for different scenarii of exposure. It may happen, in practice, that some parameters of these models can be represented by probability distributions while others are better represented by possibility distributions, or by belief functions of Shafer (for lack of information). Many researchers are used to either one or the other of these modes of representation of uncertainty. But fewer researchers are interested in the question: how to combine these different modes of representation (probability, possibility, belief function).
Let T:$\mathbb{R}^n \rightarrow \mathbb{R}$ be a function (model) of n arguments x_i ($x = (x_1, ..., x_n)$). We want to know the extent to which the criterion $T(x) \leq e$ is satisfied. The knowledge on parameters x_i can be represented using a distribution of probability, of possibility or a mass function. The main issue is thus to carry the uncertainty attached to the variables over to $T(x)$ with the least possible loss of initial information. This is uncertainty propagation. Generally, in the evaluation of risks for man and the environment, one tries to estimate $P_{T(X)}(]-\infty, e])$ where e is an absorbed dose limit for exemple.
There are three difficulties: the first is to represent the available information faithfully (cf [1]), the second one is how to account for dependencies, correlations between the parameters in the propagation process (linear, non linear monotone dependency, interaction ...). For example the fact of assuming stochastic independence between parameters can generate too optimistic results. The last one is the choice of the propagation technique.

In Section 2 (resp Section 3), we consider propagation when arguments of T are purely represented by probabilities (resp possibilities). Section 4 deals with the case where some arguments of T are represented by probability, and others by possibility or mass functions. Examples of results provided by the different methods are presented in Section 5.

2 Random variables

When the knowledge on all arguments of T is represented using random variables (that is random variables X_i are associated to parameters x_i), then $T(X)$ is also a random variable where X is the random vector. We can estimate the cumulative distribution function F_T using Monte Carlo methods. Iman and Conover [10] take into consideration monotone dependencies between parameters (Spearman rank correlation) when they are known.

The situation is more complex when the form of the dependence is not known. Ferson [5], [6] discusses the consequences on the results if one makes dependence assumptions when one does not have this information. A few researchers [7] proposed to scan all correlation coefficients ranging between -1 and 1 when no information is available. The problem is that not all forms of dependence are covered, since this approach only deals with linear dependencies. Indeed, a zero correlation between two parameters does not imply independence.
We can also estimate the cumulative distribution function F_T approximating a probability distribution function using quantile/histograms. So Berleant and Goodman-Strauss [2] maximise and minimise F_T using histograms. They obtain an envelope of F_T without any knowledge of dependencies, for any continuous T. In other words, they obtain $F_T^- \leq F_T \leq F_T^+$, by scanning all possible joint probability distribution functions of random vector X.
Williamson and Downs [13] use copulas to achieve propagation without knowledge on dependence (binary operations). The use of copulas becomes very difficult as soon as the number of variables is higher than 3

3 Possibility

When the knowledge on all arguments of T is represented by possibility distributions π_i, the knowledge on T can also be represented by a possibility distribution π_T. To compute it, we can use the extension principe [4]: $\forall u \in \mathbb{R}^n$

$$\pi_T(u) = \sup_{x_1,\ldots,x_n, T(x_1,\ldots,x_n)=u} \pi(x_1,\ldots,x_n)$$

When there is no interaction between the parameters, we can write: $\forall u \in \mathbb{R}^n$

$$\pi_T(u) = \sup_{x_1,\ldots,x_n, T(x_1,\ldots,x_n)=u} \min(\pi_1(x_1),\ldots,\pi_n(x_n))$$

The use of min characterizes a lack of knowledge about the dependence between possibilistic variables, when the result is supposed to be a possibility distribution. Thus, we can compute the degrees of possibility $\Pi_{\pi_T}(]-\infty,e])$ and of necessity $N_{\pi_T}(]-\infty,e])$ of remaining under a threshold e. This calculus cannot serve as a conservative substitute to a probabilistic propagation under stochastic independence, that is, even if $\forall i \quad P_{X_i} \in \mathcal{F}_{X_i} = \{P_{X_i}, \forall C \subset \mathbb{R} \quad N_{\pi_i}(C) \leq P_{X_i}(C) \leq \Pi_{\pi_i}(C)\}$ with P_X joint probability of random vector X (assiociated vector parameters x of T) with projections P_{X_i}, it does not imply $P_{T(X)} \in \mathcal{F}_{T(X)} = \{P_{T(X)}, \forall C \subset \mathbb{R} \quad N_{\pi_T}(C) \leq P_{T(X)}(C) \leq \Pi_{\pi_T}(C)\}$. We shall observe this feature in example 1 (cf Section 5).

4 Probability and Possibility

Suppose there are $k < n$ random variables $(X_1,\ldots,X_k)$ associated to values $(x_1,\ldots,x_k)$ and $n-k$ possibilistic variables $(X_{k+1},\ldots,X_n)$ associated to values $(x_{k+1},\ldots,x_n)$.

4.1 Hybrid method [9]

If we fix the variables $(X_1 = x_1,\ldots,X_k = x_k)$, the knowledge on $T(X)$ value becomes a fuzzy subset. By randomizing this choice, $T(X)$ is a random fuzzy subset. In practice, we can combine a Monte Carlo technique with the extension principe. It generates a sample $(F_1,\ldots,F_m)$ of fuzzy subsets which estimate $T(X)$. Guyonnet et al. [9] propose to synthetise the result into a single fuzzy subset denoted F_d. For each α cut of the random fuzzy set obtained, Guyonnet et al. separately rearrange the left side and the right side of sets in increasing order. The set $[Finf_d^\alpha, Fsup_d^\alpha]$ is considered such that $P(leftside \leq Finf_d^\alpha) = 1-d\%$ and $P(rightside \leq Fsup_d^\alpha) = d\%$. Varying $\alpha \in [0,1]$, a fuzzy interval F_d is thus built. The standard value $d = 95$ is chosen.
It is worthwile noticing that within a Monte Carlo approach the rank correlation between the random variables [10] (if known) can be taken into consideration. This raises the following question for this method: how to take into consideration dependencies between the random and possibilistic variables if they exist?
However, we must be careful with this method. It is interesting if we want to estimate for exemple $P_{T(X)}(]-\infty,e])$. But, when checking whether $T(X)$ lies between two values e_1 and e_2, we get false estimates of $P_{T(X)}([e_1,e_2])$ with it. That is we obtain the same result whatever the imprecision on each F_i and the variability on the sample of all F_i (see figure 1). We get the same fuzzy number F_d whether we have large imprecision with a small variability as on the left part of fig.1 ,or we have little imprecsion with a great variability as on the right part. Indeed we

treat independently $leftside$ and $rightside$ whereas $rightside$ is entirely determined by $leftside$ and conversely since any α cut is generated as a whole.

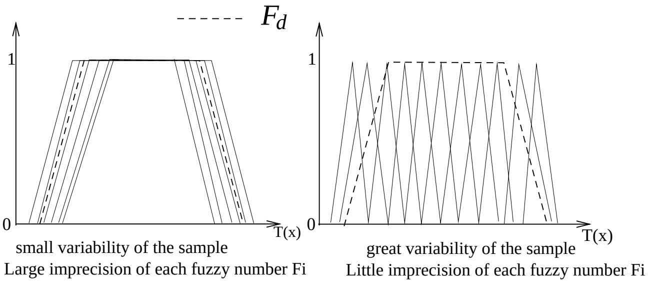

Figure 1: Two different possible results with hybrid method

To evaluate the imprecision of T, we can estimate an average fuzzy number F_d^{mean} with:

$$F_d^{mean} = \frac{1}{m}\sum_{i=1}^{m} F_i$$

and compute the area under F_d^{mean}.

To obtain knowledge on the variability of T, we can work with a representative value v_i^r of each fuzzy number F_i. Then we can estimate a variance V with estimator:

$$V = \frac{1}{m}\sum_{i=1}^{m} {v_i^r}^2 - \frac{2}{m(m-1)}\sum_{j<i} v_j^r v_i^r$$

where v_i^r is a representative value of F_i. If V is small, the variability is small as on the left part of figure 1 for exemple. We can choose for the representative value v_i^r the middle point of the mean interval:

$$v_i^r = \int_0^1 \frac{(supF_i^{\alpha} + infF_i^{\alpha})}{2} d\alpha$$

As will be observed on some examples, the use of F_d may overweight extreme fuzzy values, in the case of precise results with high variability, even if some outliers can be deleted by the threshold d.

4.2 Homogeneous Approach with belief functions [3]

Both possibilistic and probabilistic information can be cast in the framework of Evidence theory. Evidence theory includes probability and necessity. It introduces a plausibility function (noted Pl) and a belief function (noted Bel) defined from a mass function $m : \mathcal{P}(\Omega) \rightarrow [0,1]$ such that $\sum_{E\in\mathcal{P}(\Omega)} m(E) = 1$. If $m(E) > 0$, E is a focal element, m(E) is the probability of knowing that $x \in E$ and nothing else. Moreover, $Bel(A) = \sum_{E,E\subseteq A} m(E)$ and $Pl(A) = \sum_{E,E\cap A\neq\emptyset} m(E)$. If the focal elements are nested, a belief function is a necessity measure and a plausibility function is a possibility measure. They indeed satisfy: $Bel(A\cap B) = \min(Bel(A), Bel(B))$ and $Pl(A\cup B) = \max(Pl(A), Pl(B))$.

If the focal elements are some disjoint intervals, belief and plausibility functions coincide with a probability measure for unions of such intervals. We can model each possibility distribution function and each probability distribution function by a mass function. We assign to the Cartesian product of focal sets attached to each variable X_i, and, assiming independence, the product of the corresponding masses (see [3]). The obtained joint mass is then assigned to the interval [min $T(x)$,max $T(x)$] obtained by interval analysis on each focal element of vector X (see [3]). This approach supposes independence between focal elements of each X_i.

Borrowing from [2], we propose here a "Dempster Shafer conservative approach": the idea is to compute extreme upper plausibility Pl_{max} and lower belief functions Bel_{min} without assuming knowledge on dependencies. It yields a linear optimization problem whose unknown is the joint mass function (see [2]). It ensures $Bel_{min}(T(X) \leq e) \leq P(T(X) \leq e) \leq Pl_{max}(T(X) \leq e)$. Consider an example involving two parameters x, y. The knowledge on x is represented by a probability measure approximated by a mass function m with focal elements $[0,1]$, $[1,2]$ and $[2,3]$; the knowledge on y is represented by a possibility measure with mass q and focal elements $[1,2]$ and $[0,6]$. Suppose we want to estimate $T(x,y) = x+y$. The table below shows the focal elements and the mass function obtained for $T(x,y) = x+y$ in the case of independent x and y.

$[1,3]$ $m_1q_1 = h_{11}$	$[0,7]$ $m_1q_2 = h_{12}$	$[0,1]$ m_1
$[2,4]$ $m_2q_1 = h_{21}$	$[1,8]$ $m_2q_2 = h_{22}$	$[1,2]$ m_2
$[3,5]$ $m_3q_1 = h_{31}$	$[2,9]$ $m_3q_2 = h_{32}$	$[2,3]$ m_3
$[1,2]$ q_1	$[0,6]$ q_2	x y

In the conservative approach, contrary to the standard Dempster-Shafer method, we must find h such that, with e fixed, $Pl(]-\infty,e])$ is maximal and $Bel(]-\infty,e])$ is minimal. For example, we obtain $Pl_{max}(x+y \leq 2)$ by solving the following problem:

$$\begin{array}{ll} Max & 1-h_{31} \\ & \sum_j h_{ij} = m_i \\ & \sum_i h_{ij} = q_j \\ & \sum_{i,j} h_{ij} = 1 \end{array}$$

Similary we obtain $Bel_{min}(x+y \leq 4)$ by solving the following problem:

$$\begin{array}{ll} Min & h_{11}+h_{21} \\ & \sum_j h_{ij} = m_i \\ & \sum_i h_{ij} = q_j \\ & \sum_{i,j} h_{ij} = 1 \end{array}$$

4.3 Homogeneous Approach with Monte Carlo [11]

We perform a random selection among focal elements, according to the probability provided by the mass function. For each variable X_i of T, we randomly select a focal element A_{ij} and we compute their image B_j by T. If we perform N samples, we assign mass 1/N to each B_j. We obtain:

$$Pl(T(X) \leq e) = \frac{1}{N} Card\{j, inf(B_j) \leq e\}$$

$$Bel(T(X) \leq e) = \frac{1}{N} Card\{j, sup(B_j) \leq e\}$$

Of course, this method has all limitations of Monte Carlo method as discussed by Ferson [7]. This sampling approach to estimating the *Bel* and the *Pl* function will almost always understimate *Pl* and overestimate *Bel*. This approach is adapted to the case when there are many focal elements due to a fine-grained discretization of distributions. In the discrete case, with few focal elements, the result can be computed directly. In the preceeding example, one would find $Pl(x+y \leq 2) = 1 - m_3 q_1$ and $Bel(x+y \leq 4) = q_1(m_1+m_2)$.

4.4 Casting possibilistic and probabilistic data in the setting of upper and lower probabilities

Let us finally suggest a last method. We can interpret possibilistic variables $(X_{k+1},...,X_n)$ of T, in terms of upper and a lower probability . That is we work with $F^*(x_j) = \Pi_{X_j}(]-\infty,x_j])$ and $F_*(x_j) = N_{X_j}(]-\infty,x_j])$ $\forall\, k+1 \leq j \leq n$. Ferson [8] uses upper and lower probabilities separately, to represent and to propagate imprecise knowledge. We showed in [1] that to work with F^*, F_* separately is less precise than to work with a possibility distribution $\pi = \min(F^*, 1-F_*)$ (when the latter is normalized) because the set $\mathcal{F}_X$ of probabilities induced by π is contained in the set $\{P, F_* \leq P \leq F^*\}$. However, we can use probabilistic methods on upper and lower cumulative distribution functions separately and take into consideration correlations between parameters if they exist. We obtain upper and lower cumulative distributions for T. As one works with upper and lower probabilities separately, this method makes sense only if T is monotonic.

5 Examples

We are going to show by means of examples how the different methods perform. In all figures, cumulative distributions are pictured. That is for all e, we represented $P_{T(X)}(]-\infty,e])$, $\Pi_T(]-\infty,e])$, $N_T(]-\infty,e])$, $Pl_T(]-\infty,e])$ or $Bel_T(]-\infty,e])$.

In Figure 2, we first model two variables A and B by means of a uniform probability distribution on [6,9]. The same variables A and B are modelled by a fuzzy number (support [6,9], core $\{6\}$) for possibilistic variables. We have $P_A \in \mathcal{F}_A$ and $P_B \in \mathcal{F}_B$ (see section 3). We plot A+B using Monte Carlo with different correlations. In the case of fuzzy numbers we used both fuzzy arithmetic and the Dempster-Shafer model. In this case, it is not necessary to represent upper distributions (possibility and plausibility) because they are the same step-function. The parallelogram [(0,12);(0,15);(1,18);(1,15)] is the tightest region that encloses all of the possible lower distributions for A+B that could arise under different dependencies between A and B. It is the best-possible bound on the set of all distributions resulting from addition of A and B under all possible dependency assumptions. This result comes from Fréchet bounds [13]. We can see the necessity function obtained by fuzzy arithmetic is inside the parallelogram. Moreover $N_{\pi_A} \leq P_A \leq \Pi_{\pi_A}$ and $N_{\pi_B} \leq P_B \leq \Pi_{\pi_B}$ do not imply $N_{\pi_{A+B}} \leq P_{A+B} \leq \Pi_{\pi_{A+B}}$. For instance, the figure shows $N_{\pi_{A+B}}(]12,14]) = 0.35 > P_{A+B}(]12,14]) = 0.2$ (the latter with independence assumption). However, by minimizing the resulting belief function we obtain the right side of the parallelogram. That means the maximal plausibility Pl_{max} and the minimal belief Bel_{min} are upper and lower probability bounds whatever knowledge of dependencies. Thus for all dependencies between A and B we can obtain $Bel_{min}(]-\infty,e]) \leq P_{A+B}(]-\infty,e]) \leq Pl_{max}(]-\infty,e])$.

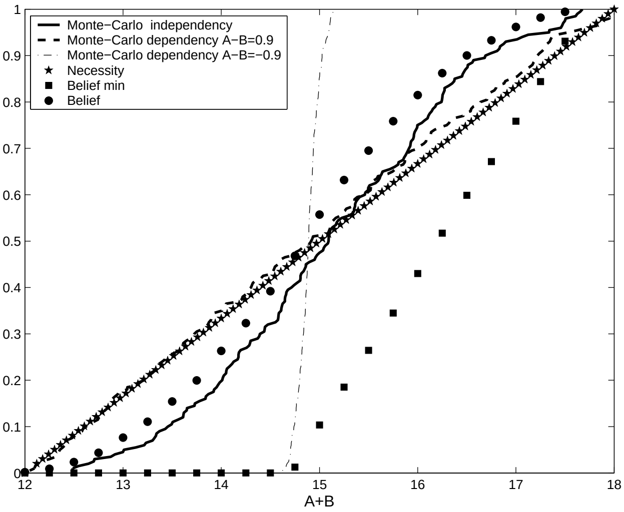

Figure 2: Various computations of A+B.

In the second example (cf Figure 3), we do the same kind of calculation as previously. Consider three variables A, B, C. Suppose A=Unif([1,5]), B=Unif([3,8]) and C=Norm(7.5,1) in the pure random case. We use fuzzy numbers (supp(A)=[1,5], core(A)=[2,3]; supp(B)=[3,8], core(B)=[4,6]; supp(C)=[4,11], core(C)=[7,8]) in the possibilistic case. We compute (A+B)/C in each pure case. Note that the minimal belief function is greater than the necessity function for x=2.5. It is due to a coarse discretisation of possibility distributions in order to build focals elements.

The last example (in Figure 4) is the most interesting. We have two possibilistic variables A, B (fuzzy numbers supp(A)=[1,5], core(A)=[2,3]; supp(B)=[3,8], core(B)=[4,6]). We have one random variable C=Norm(7.5,1). We try to estimate (A+B)/C. We use the Dempster-Shafer calculation (also the case of ignored dependencies; see Section 4.2), the hybrid method, the Dempster-Shafer calculation combined with Monte Carlo and the Monte Carlo with upper and lower probability for A and B. We can see that methods presented in Section 4.2, 4.3 et 4.4 return almost the same results. Indeed, we suppose

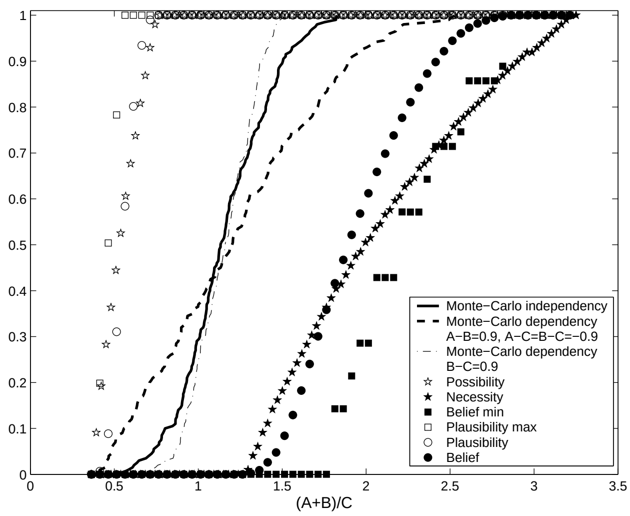

Figure 3: Various computations of (A+B)/C.

the independency of the variables with these methods. Dempster-Shafer conservative approach (see Section 4.2) produces an envelope for all results except for hybrid necessity. This due to the construction of F_d (see Section 4.1). Indeed, when Guyonnet et al [9] compute F_d, they transform the variability of T into imprecision. In fact, the hybrid method does not account for variability and may put excessive weights on randomly generated fuzzy numbers located on the sides of F_d. It seems that too conservative a result is obtained in this way. Bel_{min} and Pl_{max} remain the most credible conservative bounds on the cumulative distribution function of T. For example we have: $0 \leq P(\frac{A+B}{C} \leq 0.5) \leq 0.28$ or $0.85 \leq P(\frac{A+B}{C} \leq 2) \leq 1$. We have not information on $P(\frac{A+B}{C} \leq 1)$ because we only know: $0 \leq P(\frac{A+B}{C} \leq 1) \leq 1$.

6 Conclusion

This paper has proposed a preliminary investigation of possible methods for handling uncertain parameters in environmental risk analysis. It is based on the idea of jointly exploiting incomplete and probabilistic information. First experiments suggest that the possibilistic modeling of ill-known data, interpreted in terms of upper and lower probabilities and belief functions can be useful for the propagation step. It is recalled that contrary to what is sometimes claimed, fuzzy interval analysis based on the extension principle does not necessarily produce conservative probabilistic bounds. It could even be locally less conservative than a standard Monte-Carlo approach. How-

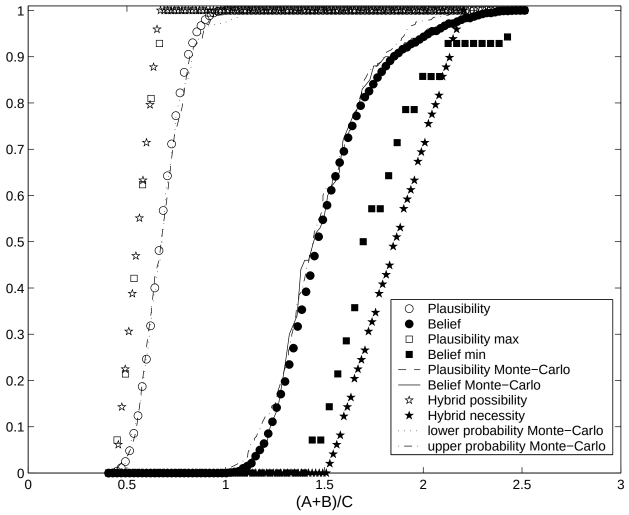

Figure 4: Hybrid computation of (A+B)/C.

ever the setting of belief functions can provide such conservative bounds, using discrete random intervals and no dependence assumption. In the case of mixed probabilistic/possibilistic data, the exploitation of random fuzzy outputs produced by Monte-Carlo simulation (as in [9]) may also produce results sometimes locally less conservative than a standard Dempster-Shafer approach, sometimes more conservative than the Dempster-Shafer approach without dependence assumption. More investigations are needed for an in-depth comparison of uncertainty propagation methods so as to properly explain some of the obtained results.

Acknowledgements

This work is as part of a collaboration with the French Institutes B.R.G.M, I.R.S.N and I.N.E.R.I.S.

References

[1] C. Baudrit, D. Dubois, H. Fargier. Reprsentation de la connaissance probabiliste incomplte. Soumis à LFA 03, Tours, France, 2003.

[2] D. Berleant & C. Goodman-Strauss. Bounding results of Arithmetic Operations on Random Variables of Unknown Dependencies using Interval Arithmetic. *Reliable Computation*, 4, 147-165, 1998.

[3] D. Dubois & H. Prade. Random sets and Fuzzy Interval Analysis. *Fuzzy Sets and Systems*, 42, 87-101, 1991.

[4] D. Dubois, E. Kerre, R. Mesiar, H. Prade . Fuzzy interval analysis. *Fundamentals of Fuzzy Sets*, Dubois,D. Prade,H., Eds: Kluwer , Boston, Mass, 483-581, 2000.

[5] S. Ferson & M.A. Burman. Correlation, dependency bounds and extinction risks. *Biological Conservation*, 73, 101-105, 1995.

[6] S. Ferson & T.F. Long. Conservative uncertainty propagation in environmental risk assessments. *Environmental Toxicology and Risk Assessment*. ASTM STP 1218, Ed J.S Hughes et al. Am. Soc. for Testing and Materials, Philadelphia, 97-110, 1994

[7] S. Ferson. What Monte Carlo methods cannot do. *Human and Ecology Risk Assessment*, 2, 990-1007, 1996.

[8] S.Ferson, L.Ginzburg, R.Akcakaya. Whereof one cannot speak: when input distributions are unknown. A paraître dans*Risk Analysis*, 2003.

[9] D. Guyonnet, B. Bourgine, D. Dubois, H. Fargier, B. Cme, J.P. Chils. Hybrid approach for addressing uncertainty in risk assessments. *Journal of Environmental Engineering*, 126, 68-78, 2003.

[10] R.L. Iman & W.J. Connover. A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables. *Technometric*, 3, 311-334, 1982.

[11] S. Moral & N. Wilson. Importance sampling Monte-Carlo algorithms for the calculation of Dempster-Shafer belief. *Tech. Rep. UTAI University of Granada*, 1996.

[12] G. Shafer. A Mathematical Theory of Evidence. *Princeton University Press*, 1976.

[13] R.C. Williamson & T. Downs. Problematic Arithmetic.I: Numerical Methods for Calculating Convolutions and Dependency Bounds. *International Journal of Approximate Reasonning*, 4, 89-158, 1990.