A formal model to compute uncertain continuous data

Jérôme Dantan[1,2], Yann Pollet[2], and Salima Taibi[1]

[1]Esitpa, Agri'terr, Mont-Saint-Aignan, France
{jdantan,staibi}@esitpa.fr
[2]CNAM, CEDRIC, Paris, France
yann.pollet@cnam.fr

Abstract. Current researches in the domain of Information and Communication Technologies describe and extend the existing formalisms to develop systems that compute uncertain data. Indeed, handling uncertain data is a great challenge for complex systems. In this article, we provide a formal model to compute such data rigorously. Such quantities may be interpreted as either possible or probable values, added to their interdependencies. For this, the algebraic structure we defined is a vector space. We then provide a particular way for mixing such continuous quantities.

Keywords: probability, possibility, discrete, continuous

1 Introduction

Current researches in the domain of ICT describe and extend existing formalisms to design systems that have to manage more and more data, because of the increasing number of sensors and means of storage. Many actors of scientific domains have to cope with data that may be uncertain, imprecise or incomplete, to assist humans in their decisions. For this, they have to merge data from many data sources. In the last forty years many well-known mathematical approaches to model imperfect data have been applied, such as probability based calculus. But increasing amounts of data have to be processed so that there is not enough time for data cleaning step. Decisions of experts from various fields are based on aggregations of data. We have therefore to take into account their imperfection by a rigorous approach. We propose an algebraic structure to model data, whose imperfections nature may be covered either by the classical probability theory, either possibility theory.

2 State of the art

For a long time, uncertainty modelling remained addressed by the probability theory, which is the mathematical study of phenomena characterized by randomness and uncertainty. However, this approach is little suitable for total ignorance representation, and objective interpretations of probabilities, assigned to such events remain difficult when handled knowledge are no longer linked to random and repeatable phenomena (Dubois and Prade, 1988). As against, it is possible to model uncertainty

adfa, p. 1, 2011.

thanks to the possibility theory.

The possibility theory (Dubois D., Prade H., 1988), (Zadeh L.A., 1978) removes the strong probability additive constraint and associates the events of Ω to a possibility measure denoted Π and a necessity measure denoted N, that are both applications from Ω to $[0;1]$, respectively satisfying: $\Pi(A \cup B) = \max(\Pi(A), \Pi(B))$ and $N(A \cap B) = \min(N(A), N(B))$. The relationship between the possibility of an event and its opposite is given by $\Pi(A) = 1 - N(\neg A)$ and total ignorance is then given by $\Pi(A) = \Pi(\neg A) = 1$, which implies: $N(A) = N(\neg A) = 0$. This approach also allows representing imprecision using notions of fuzzy sets and distributions of possibilities. Thus, a fuzzy set (Zadeh L.A., 1965) F of a set E is defined by a membership function μ_F from E to $[0;1]$, which associates each element x of E its membership degree $\mu_F(x)$, to the subset F (i.e.: x belongs "more or less" to F). When this membership function is normalized (i.e. a x value from E such as $\mu_F(x) = 1$ exists), $\mu_F(x)$ can then interpreted as the chance that F takes the value x ($\mu_F(x)$ is then a possibility distribution).

In (Dantan et al., 2015), we provided a formalism for both representing and manipulating quantities which may have a finite number of possible or probable values. Such quantities are values of the ℝ set. We provided an algebraic structure to operate the chained computations on such quantities with properties similar to ℝ, that does not allow the classical approaches based on fuzzy sets seen in the literature. In this paper, we provide an extension of this approach to continuous quantities that are on the one hand, probabilistic, and, on the other, possibilistic. The continuous mixed quantities are not in the scope of this article. In addition, we restrict ourselves to quantities belonging to the ℝ set.

3 Approach

The provided approach provides a formalism for both representing and manipulating rigorously quantities which may have a finite number of possible or probable values with their interdependencies. Then, we defined an algebraic structure to operate chained computations on such quantities with properties similar to ℝ.

Let Ω be a universe, with both a possibility measure Π and a probability measure P, each having values belonging to ℝ. The considered values are respectively denoted: $a_1, a_2, \ldots, a_{n1}$, with possibilities $\alpha_1 = \Pi(a_1), \ldots, \alpha_{n1} = \Pi(a_{n1})$ ans $b_1, b_2, \ldots, b_{n2}$, with probabilities $\beta_1 = P(b_1), \ldots, \beta_{n2} = P(b_{n2})$. In the following of this article, we denote D1-Distributions on ℝ: D1D(ℝ) with possibilistic bases $B_I = \{X_{I,i}\ ; i = 1, \ldots\}$ (I fixed) and probabilistic bases $B^J = \{X^{J,j}\ ; j = 1, \ldots\}$ (J fixed). The types of structures here considered are aggregation (cartesian product) of elementary types taken among ℝ space. We define two types of bases, generating the D1D vector space of D1D on the infinite space but countable ℝ space.

We have defined an internal product on vectors: Z = X.Y that checks the following properties: (1) X.X = X (idempotency); (2) X.0 = 0_v, where 0_v is the null vector (absorbing element); (3) $i_1 \neq i_2$ implies $X_{I,i1}.X_{I,i2} = 0_v$; (4) $j_1 \neq j_2$ implies $X^{I,i1}.X^{I,i2} = 0_v$; (5) $X_{I1,i1}.X_{I2,i2} \neq 0_v$; (6) $X_{I1,i1}.X^{I2,i2} \neq 0_v$;(7) $X^{I1,i1}.X^{I2,i2} \neq 0_v$. With such two types of vectors bases, we are able distinguish the sources of uncertainty during combinations of values and then make rigorous computations.

4 Background: uncertainty on discrete quantities

We define a purely possibilistic D1D as a value "a" which may have a finite number n_1 of possible values: (a_1,…, a_{n1}) of K^{n_1} with their respective associated possibilities $\alpha_1 = \Pi(a_1)$, …, $\alpha_{n1} = \Pi(a_{n1})$ of $[0,1]^{n_1}$.

The canonical form of a purely possibilistic D1D "a" is the following expression: $= \sum_{i=1}^{n} a_i/\alpha_i . X_{I,i}$, with a_i are the possible values of a, α_i are the possibilities, associated to each value a_i (one of which at least is equal to 1) and $X_{I,\,i}$ (I fixed) correspond to the partition of the universe Ω corresponding to values of quantity a.

We define a purely probabilistic D1D as a value b which may have a finite number n_2 of probable values: (b_1, b_2,…, b_{n2}) of K^{n_2} with their respective associated probabilities $\beta_1 = P(b_1)$, …, $\beta_{n2} = P(b_{n2})$ of $[0,1]^{n_2}$. The n_2 values (b_1, b_2,…, b_{n2}) completely define the probability distribution b/β_i on ℝ, associated to the probabilistic variable b. β_i is equal to zero except on values b_1, b_2,…, $b_{n2.}$

The canonical form of a purely probabilistic D1D "b" is the following expression: $b = \sum_{i=1}^{n} b_i/\beta_i . X^{J,i}$, with b_j are the probable values of b, β_j are the probabilities, associated to each value b_j (the sum of β_j is equal to 1) and $X^{J,\,j}$ (J fixed) correspond to the partition of the universe Ω corresponding to values of quantity b.

4.1 Internal composition law (+)

The following expression: $a_1 + a_2 = \sum_{i=1}^{n_1} a_{1,i}/\alpha_{1,i}.X_{1,i} + \sum_{j=1}^{n_2} a_{2,j}/\alpha_{2,j}.X_{2,j}$ is evaluated with a composed sum of ($a_{1,i}$ + $a_{2,j}$) / ($\alpha_{1,i}$ min $\alpha_{2,j}$) . $X_{1,i}$. $X_{2,j}$, i.e. $a_1 + a_2 = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \quad (a_{1,i} + a_{2,j})/(\alpha_{1,i} \min \alpha_{2,j}).X_{1,i}\,.X_{2,j}$.

The following expression: $b_1 + b_2 = \sum_{i=1}^{n_1} b_{1,i}/\beta_{1,i}.X^{1,i} + \sum_{j=1}^{n_2} b_{2,j}/\beta_{2,j}.X^{2,j}$ is likewise evaluated with a composed product of ($b_{1,i}$ * $b_{2,j}$) / ($\beta_{1,i}$. $\beta_{2,j}$) . $X^{1,i}$. $X^{2,j}$, $b_1 + b_2 = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \quad (b_{1,i} + b_{2,j})/(\beta_{1,i} * \beta_{2,j}).X^{1,i}.X^{2,j}$

As a special case, if both operands are expressed in the same basis (i.e. they are dependent, or in other words linked), then the formula is simplified. We check that it can take into account rigorously dependencies to not artificially explode the number of possible values.

4.2 Internal composition law (*)

Similarly to +, the product internal composition law is evaluated with composed products of ($a_{1,i}$ * $a_{2,j}$) / ($\alpha_{1,i}$ min $\alpha_{2,j}$) . $X_{1,i}$. $X_{2,j}$ (possibility) and $(b_{1,i} + b_{2,j})/(\beta_{1,i} * \beta\_(2,j))$ (probability). Finally, we showed that D1D (ℝ) has a vector space structure on ℝ .

5 Continuous quantities

5.1 Combinations of continuous possibilistic quantities (trapezoids)

In this case, we restrict ourselves to possibilistic coefficients and to trapezoids. So this is purely continuous possibilistic D1D. We then get the following canonical expression: $a = \sum_{i=1}^{n} a_i . T(\lambda_{1i} . a_i, \lambda_{2i} . a_i)/\alpha_i . X_{I,i}$, where $T(\lambda_{1i} . a_i, \lambda_{2i} . a_i)$ is a trapezoid centered on a_i, with kernel $a_i \pm \lambda_{1i}. a_i$ and support $a_i \pm \lambda_{2i}. a_i$

The resulting computations are degraded trapezoids. Considering that the symmetric trapezoid centered on a_i, with kernel $a_i \pm \lambda_{1i}. a_i$ and support $a_i \pm \lambda_{2i}. a_i$, $\hat{a}_i$ (i.e. $a_i . \mathcal{T}(\lambda_{1i}. a_i, \lambda_{2i}. a_i)$) is stable for the four main algebraic operations +, -, *, /. Here are the values for the resulting cores and supports four main algebraic operations:

- Addition: $a + b = \sum_i [^{\alpha i}{}_0]/ a_i . \mathcal{T}(\lambda_{1i}. a_i, \lambda_{2i}. a_i) + \sum_j [^{\eta j}{}_0]/ b_j . \mathcal{T}(\kappa_{1j}. b_j, \kappa_{2j}. b_j)$. We then get the following expression: $a + b = \sum_{i,j} [^{\alpha i \min \eta j}{}_0]/ (a_i + b_j) . \mathcal{T}(\lambda_{1i}. a_i + \kappa_{1j}. b_j, \lambda_{2i}. a_i + \kappa_{2j}. b_j)$.
- Subtraction: $a - b = \sum_i [^{\alpha i}{}_0]/ a_i . \mathcal{T}(\lambda_{1i}. a_i, \lambda_{2i}. a_i) - \sum_j [^{\eta j}{}_0]/ b_j . \mathcal{T}(\kappa_{1j}. b_j, \kappa_{2j}. b_j)$. We then get the following expression: $a - b = \sum_{i,j} [^{\alpha i \min \eta j}{}_0]/ (a_i - b_j) . \mathcal{T}(\lambda_{1i}. a_i + \kappa_{1j}.b_j, \lambda_{2i}. a_i + \kappa_{2j}. b_j)$.
- Product: $a * b = \sum_i [^{\alpha i}{}_0]/ a_i . \mathcal{T}(\lambda_{1i}. a_i, \lambda_{2i}. a_i) * \sum_j [^{\eta j}{}_0]/ b_j . \mathcal{T}(\kappa_{1j}. b_j, \kappa_{2j}. b_j)$. We then get the following expression: $a * b = \sum_{i,j} [^{\alpha i \min \eta j}{}_0]/ (a_i * b_j) . \mathcal{T}(a_i . b_j * (\lambda_{1i} + \kappa_{1j}), a_i . b_j (\lambda_{2i} + \kappa_{2j})$.
- Division: $a / b = \sum_i [^{\alpha i}{}_0]/ a_i . \mathcal{T}(\lambda_{1i}. a_i, \lambda_{2i}. a_i) / \sum_j [^{\eta j}{}_0]/ b_j . \mathcal{T}(\kappa_{1j}. b_j, \kappa_{2j}. b_j)$. We then get the following expression: $a / b = \sum_{i,j} [^{\alpha i \min \eta j}{}_0]/ (a_i / b_j) . \mathcal{T}(a_i . b_j * (\lambda_{1i} + \kappa_{1j}), a_i . b_j (\lambda_{2i} + \kappa_{2j})$.

5.2 Combinations of continuous probabilistic quantities (gaussian distribution)

In this case, we restrict ourselves to probabilistic coefficients and to Gaussian distributions. We then get the following expression: $b = \sum_{i=1}^{n} b_i . N(\lambda_i . b_i)/\beta_i . X^{J,i}$. The probability density on $\mathbb{R}$ is then the weighted sum of n gaussian densities. In the internal composition law + (sum) case , to combine e.g. (a+b), with $b = \sum_{i=1}^{n} b_i . N(\lambda_i . b_i)/\beta_i . X^{J,i}$, we have to make a convolution of two density functions of the following formula: $\sum_i p_i . \mathcal{N}(\sigma_i (x))$. We reduce such computations to gaussian densities $N(\sigma_1) * N(\sigma_2)$ convolutions, that result to Gaussian densities of the form $N\left(\sqrt{\sigma_1^2 + \sigma_2^2}\right)$.

In the internal composition law * (product) case, a residue term appears when calculating the product of two gaussian distributions. This term might be ignored considering that random variations are small compared to the main values. However, in other cases, this residue does not give a gaussian distribution but a Bessel function, which means that the law continuous probabilistic D1D are not stable by the law *.

6 Conclusion

The provided approach provides a formalism for both representing and manipulating rigorously quantities which may have possible or probable values with their interdependencies. Then, we define an algebraic structure to operate chained computations on such quantities with properties similar to $\mathbb{R}$.

We have extended our formalism on continuous quantities. In special cases such as trapezoids for possibilities and normal distributions for probabilities), some algebraic properties of D1D have been maintained. However, combinations of continuous quantities, including probabilistic ones require additional assumptions that make the computations not mathematically rigorous.

The next steps are to compute mixed continuous quantities, by considering either the trapezoidal possibility distributions as intervals of cumulative distribution functions (Destercke S., Dubois D., 2009) or probability-possibility transformations (Dubois et al, 2004).

7 References

Dantan, J., Pollet, Y. and Taïbi, S. (2015), *Combination of Imperfect Data in Fuzzy and Probabilistic Extension Classes*. Journal of Environmental Accounting and Management. pp. 123-150 | DOI: 10.5890/JEAM.2015.06.004.

Dempster, A. (1967), Upper and lower probabilities induced by multivalued mapping, *Annals of Mathematical Statistics* **AMS-38**: 325-339.

Destercke S., Dubois D. and Chojnacki E. (2007), *On the relationships between random sets, possibility distributions, p-boxes and clouds.* 28th Linz Seminar on fuzzy set theory, 2007, Linz, Austria.

Dubois, D. and Prade, H. (1988), *Théorie des possibilités, Application à la représentation des connaissances en informatique*. Masson 1988. (In French).

Dubois D., Foulloy L., Mauris G., Prade H., 2004. Probability-Possibility Transformations, Triangular Fuzzy Sets, and Probabilistic Inequalities. Reliable Computing. August 2004, Volume 10, Issue 4, pp 273-297.

Shafer, G. (1976), *A mathematical theory of evidence*. Princeton University Press.

Shapiro, A.F. (2012), *Implementing Fuzzy Random Variables*. In proceedings of ARCH 2013.1. Society of Actuaries. University of Manitoba, Winnipeg, MB, Canada, August 1-4, 2012.

Smets, Ph. and Kennes, R. (1994), The transferable belief model, *Artificial Intelligence* **66**: 191-234.

Zadeh, L.A. (1965), Fuzzy Sets, *Information and Control* **8**: Academic Press.

Zadeh, L.A. (1978), Fuzzy Sets as a basis for a Theory of Possibility, *Fuzzy Sets and Systems*, 1.