Unifying parameter learning and modelling complex systems with epistemic uncertainty using probability interval

C. Baudrit, S. Destercke, Pierre-Henri Wuillemin

Unifying parameter learning and modelling complex systems with epistemic uncertainty using probability interval

C. Baudrit[a,*], S. Destercke[b], P.H. Wuillemin[c]

[a]*INRA, I2M, USC 1368, F-33400 Talence, France.*
[b]*UTC, HEUDIASYC, UMR6599, F-60205 Compiègne cedex, France*
[c]*Sorbonne Universites, UPMC, Univ Paris 06, UMR 7606, LIP6, Paris, France*
CNRS, UMR 7606, CNRS, Paris, France

Abstract

Modeling complex dynamical systems from heterogeneous pieces of knowledge varying in precision and reliability is a challenging task. We propose the combination of dynamical Bayesian networks and of imprecise probabilities to solve it. In order to limit the computational burden and to make interpretation easier, we also propose to encode pieces of (numerical) knowledge as probability intervals, which are then used in an imprecise Dirichlet model to update our knowledge. The idea is to obtain a model flexible enough so that it can easily cope with different uncertainties (i.e., stochastic and epistemic), integrate new pieces of knowledge as they arrive and be of limited computational complexity.

Keywords: Dynamic credal networks, imprecise probability, Dirichlet model, knowledge integration, uncertainty, modelling.

1. Introduction

Firms and industrials of all sectors have to face up new challenging situations. On the one hand, citizens as well as public authorities have stronger demands in terms of quality, safety, . . . and on the other hand, they must adapt to the increase of population, global warming and the depletion of fossil resources. This means, among other things, that industrial projects have to integrate sustainability from local to world scale in their conception. Possessing adequate tools to model their systems is likely to make the task easier.

In order to provide relevant conclusions and recommendations, such tools should be able to integrate as much available knowledge as possible, however heterogeneous it is, both in terms of nature (e.g., qualitative expert knowledge vs statistical data) and quality (different precision or degrees of reliability). Such systems are also complex, meaning that the modeling tool

*Corresponding author
Email addresses: cedric.baudrit@u-bordeaux.fr (C. Baudrit), sebastien.destercke@hds.utc.fr (S. Destercke), pierre-henri.wuillemin@lip6.fr (P.H. Wuillemin)

must be able to cope with different scales (e.g., molecular to macroscopic) and with dynamic, time-varying processes. Current researches rely on the development of mathematical tools [53, 6] capable of helping decision-makers to deal with uncertainties, linked for instance to meteorological variations, to expert reliability, *etc.* To summarize, ideal modeling tools should be able to deal with:

- heterogeneous sources of knowledge (Web, data warehouse, experts, ...)
- mathematical formalisms used by different disciplines (differential equations, graphs, cognitive maps, ...)
- various manipulated scales (molecular, cellular, population, ...)
- different forms of uncertainty [32, 36, 40] (natural randomness, imprecision in expert opinions, data scarcity, vagueness, ...)

In this paper, we propose dynamic credal networks as a possible answer to these challenging tasks to describe complex dynamical systems tainted with stochastic and epistemic uncertainty. As an extension of dynamic Bayesian networks (DBNs) [51], their network structure provides an intuitively appealing interface for human experts to model highly-interacting sets of variables, resulting in a qualitative representation of knowledge. Stochastic and epistemic uncertainties pertaining to the system are then taken into account by quantifying dependence between variables by means of convex sets of conditional probability distributions. The concept of DCNs makes it possible to combine different sources of information, from qualitative expert knowledge to experimental data.

In this paper, we are specifically interested in the problem of parameter learning for a given network structure (assumed to be known), when faced with heterogeneous knowledge. Indeed, while DCN are very attractive modeling tools, they also come with a number of challenges, such as how to control their computational tractability, or how to combine efficiently and easily various pieces of information. For example, how to combine simulations coming from stochastic differential equations with an experimental database, both offering information for the same parameters? We propose to use an imprecise Dirichlet model [7] as a model of the conditional probabilities, and probability intervals as a common uncertainty model to treat different pieces of knowledge. Once transformed, these information pieces gradually increment the set of prior distributions according to the received knowledge, using the Generalized Bayes rule each time additional information arrives. Lower and upper expected *a posteriori* (EAP) are then used as probability bounds to draw inferences from the network. The combination of information is done through a weighted average, allowing us to weigh the importance of the different sources of knowledge.

Section 2 details the material regarding imprecise probabilities as well as the proposed updating scheme of a given parameter set. We then describe in Section 3 how various common sources of information can be transformed into probability intervals. Section 4 presents how we extend Dynamic Bayesian Networks to sets of conditional probabilities, while Section 5 illustrates the whole approach on a real-case scenario involving cheese ripening.

2. Imprecise probabilities and Dirichlet model

Let X be a variable [1] taking its values on the finite set $\mathcal{X} = \{x_1, \ldots, x_n\}$, and $p : \mathcal{X} \mapsto [0,1]$, $\sum_{x \in \mathcal{X}} p(x) = 1$ be a probability mass function over $\mathcal{X}$. $p(X)$ will denote the vector mass function, while $p(x)$ will denote the value taken by p for $X = x$. Such a mass function defines a measure $P_X(A) = \sum_{x \in A} p(x)$ for all $A \subseteq \mathcal{X}$.

2.1. Imprecise probability and credal sets

In general, identifying a single probability modelling our uncertainty about some variable X requires a lot of data and/or knowledge. When such knowledge is not available, a safer option is to model our uncertainty by convex sets of probabilities, often called *credal sets* [47, 61, 2]. A credal set associated with X, denoted $K(X)$, is a convex set of probability masses over $\mathcal{X}$. $K(X)$ represents the uncertainty about the unknown value of the variable X. From $K(X)$ are defined upper and lower probability measures of an event $A \subseteq \mathcal{X}$ as

$$\overline{P_X}(A) = \sup_{p \in K(X)} \sum_{x \in A} p(x), \; \underline{P_X}(A) = \inf_{p \in K(X)} \sum_{x \in A} p(x). \tag{1}$$

and, in particular, for any element $x \in \mathcal{X}$ we will have that the upper and lower probabilities are given by

$$\overline{p}(x) = \sup_{p \in K(X)} p(x), \tag{2}$$

$$\underline{p}(x) = \inf_{p \in K(X)} p(x) \tag{3}$$

In a subjectivist tradition, the lower probability $\underline{P_X}(A)$ can be interpreted as the maximal price one would be willing to pay for the gamble which pays 1 unit if event A occurs (and nothing otherwise) [61]. $\underline{P_X}(A)$ is therefore a measure of evidence in favour of event A, or in other words how much $K(X)$ supports event A, while $\overline{P_X}(A)$ measures the lack of evidence against A. $K(X)$ can also be given a robust interpretation, in which it models imperfect

[1] We adopt notations similar to those of [2, Ch.9] and [24].

knowledge of a precise, possibly frequentist, probability p. A credal set $K(X)$ contains a set $\mathcal{E}xt(K(X))$ of extreme probability masses, always finite in this paper, corresponding to the vertices of $K(X)$. Geometrically, $K(X)$ may be equivalently specified by the convex hull (denoted CH) of the set $\mathcal{E}xt(K(X))$, i.e.

$$K(X) = CH\{\mathcal{E}xt(K(X))\}. \quad (4)$$

The *vacuous* credal set

$$K_v(X) = \{p(X) : p(x) \geq 0,\ \forall x \in \mathcal{X},\ \sum_{x\in\mathcal{X}} p(x) = 1\} \quad (5)$$

that includes all probability masses over $\mathcal{X}$ plays an important role, as it models total ignorance, and should be the starting point of any model. We refer to Walley [61, Sec. 5.5.] for a discussion about uniform probability distribution not being a good model of ignorance.

In this paper, we will also be especially interested in particular credal sets $K(X)$ specified by means of *interval probability*

$$K(X) = \{p(X) :\ p(x) \in [l_x, u_x],\ 0 \leq l_x \leq u_x \leq 1,\ \sum_{x\in\mathcal{X}} p(x) = 1\}. \quad (6)$$

Indeed, such credal sets that focus over bounds of singletons have the advantage to be easier to manipulate, simulate and represent than general ones, while remaining expressive enough (they include both the vacuous and the precise models). We refer to De Campos *et al.* [11] for a detailed exposition, and will only limits ourselves to necessary elements in this paper.

Example 1. Consider an example with three possibilities $\mathcal{X} = \{x_1, x_2, x_3\}$ (e.g., the working states of a system such as "failing", "degraded functioning", "fully functioning"), and assume that previous experiments result in the following intervals

$$p(x_1) = [0; 0.2], \qquad p(x_2) = [0.3; 0.4], \qquad p(x_3) = [0.4; 0.6].$$

The credal set $K(X)$ is the set of all precise probabilities $P(X) = (p(x_1), p(x_2), (p(x_3))$ within these interval bounds. Here $K(X)$ is a polytope defined by the convex hull of its four vertices in a three dimensional space:

$$K(X) = CH\{(0, 0.4, 0.6); (0.2, 0.3, 0.5); (0.2, 0.4; 0.4); (0.1, 0.3, 0.6)\}.$$

Finding these vertices can be done by using classical tools of convex geometry [39], or by using algorithms proper to a given representation (an Algorithm is provided by De Campos *et al.* [11]). The set $K(X)$ is represented in Figure 1 in barycentric coordinates.

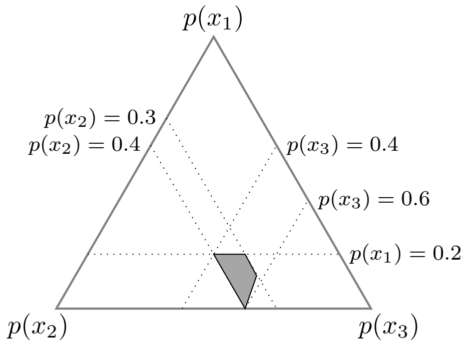

Figure 1: Example 1 credal set in Barycentric coordinates.

2.2. Robust Dirichlet model to learn $K(X)$

An important question is how the credal set $K(X)$ can be instantiated from actual evidence, or in other words how can we go from an initially vacuous knowledge towards a more precise state of knowledge. An instrumental tool to do that is to use a robustified version of the Dirichlet model, also commonly referred to as the Imprecise Dirichlet Model (IDM) [62, 7, 8, 60]. The basic model is based on two hyper-parameters: a positive real value s_0 associated to the strength of prior knowledge, and a vector $\epsilon_0 = (\epsilon_0(x_1), \ldots, \epsilon_0(x_n))$ associated to our initial beliefs about the probabilities of occurrence of elements x_i.

Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ be a vector of chances such that θ_i corresponds to the chance that $X = x_i$. The prior distribution of vectors $\boldsymbol{\theta}$ given by a Dirichlet model is then

$$\mathrm{Dir}(s_0; \xi_0)(\boldsymbol{\theta}) = \frac{\Gamma(s_0)}{\prod_{i=1}^{n} \Gamma(s_0 \xi_0(x_i))} \prod_{i=1}^{n} \theta_i^{s_0 \xi_0(x_i) - 1} \tag{7}$$

where Γ is the gamma function. A very easy way to make this model imprecise is to let the vector ϵ_0 become imprecise, and more precisely to consider the set of Dirichlet models

$$\mathcal{M}_{(s_0;\xi_0)} = \{\mathrm{Dir}(s_0; \xi_0)(\boldsymbol{\theta}) : \xi_0 \in \mathcal{T}\} \tag{8}$$

with

$$\mathcal{T} = \{\xi_0 : \ 0 < \xi_0(x_i) < 1 \ , \sum_{i=1}^{n} \xi_0(x_i) = 1\} \tag{9}$$

the open $(n-1)$-dimensional unit simplex. When ξ_0 is precise, the first moments of $\mathrm{Dir}(s_0; \xi_0)$ are given by $E(\theta_i|(s_0; \xi_0))= \xi_0(x_i)$, and they can be used as estimates of $p(x_i)$, *i.e.*

$$E(\theta_i|(s_0; \xi_0)) = \xi_0(x_i) = p(x_i). \tag{10}$$

When starting from a vacuous prior knowledge $\xi_0 \in \mathcal{T}$, the bounds over the first moments become

$$\underline{E}(\theta_i|(s_0;\xi_0)) = \min_{\xi_0 \in \mathcal{T}} \xi_0(x_i) = 0 \tag{11}$$

and

$$\overline{E}(\theta_i|(s_0;\xi_0)) = \max_{\xi_0 \in \mathcal{T}} \xi_0(x_i) = 1. \tag{12}$$

The credal set corresponding to these bounds is then the vacuous one (5).

We may then receive additional information from various m sources. A convenient way to encode this information is as a couple $s_k, \mathcal{P}_k$, $k = 1, \ldots, m$, with $\mathcal{P}_k \subseteq \mathcal{T}$ a convex polytope providing information about the possible chances θ_i, and $s_k \in \mathbb{R}^+$ modelling the strength of the information. We can then update the Dirichlet modelling our uncertainty about $\boldsymbol{\theta}|(s_k;\xi_k)_{k=0}^m$ into

$$\mathcal{M}_{(s_k;\xi_k)_{k=0}^m} = \left\{ \mathrm{Dir}\left((s_k;\xi_k)_{k=0}^m\right)(\boldsymbol{\theta}) : \xi_k \in \mathcal{P}_k \ \forall k \right\}. \tag{13}$$

We can then use the posterior first moments to make inferences on chances θ_i

$$\mathrm{E}(\theta_i|(s_k;\xi_k)_{k=0}^m) = p(x_i) = \frac{\sum_{k=0}^m s_k \xi_k(x_i)}{\sum_{k=0}^m s_k} \tag{14}$$

As information $\mathcal{P}_k$ are imprecise, we again obtain bounds in the form

$$\underline{E}(\theta_i|(s_k;\xi_k)_{k=0}^m) = \underline{p}(x_i) = \frac{\sum_{k=0}^m s_k \underline{\xi}_k(x_i)}{\sum_{k=0}^m s_k}, \tag{15}$$

$$\overline{E}(\theta_i|(s_k;\xi_k)_{k=0}^m) = \overline{p}(x_i) = \frac{\sum_{k=0}^m s_k \overline{\xi}_k(x_i)}{\sum_{k=0}^m s_k}. \tag{16}$$

where

$$\underline{\xi}_k(x_i) = \inf_{\xi_k \in \mathcal{P}_k} \xi_k(x_i) \tag{17}$$

$$\overline{\xi}_k(x_i) = \sup_{\xi_k \in \mathcal{P}_k} \xi_k(x_i). \tag{18}$$

These bounds then induce an updated credal set

$$K_{(s_k;\xi_k)_{k=0}^m}(X) = \left\{ p : p(x_i) \in \left[\frac{\sum_{k=0}^m s_k \underline{\xi}_k(x_i)}{\sum_{k=0}^m s_k}, \frac{\sum_{k=0}^m s_k \overline{\xi}_k(x_i)}{\sum_{k=0}^m s_k} \right] \right\} \tag{19}$$

that we can use as new knowledge. In practice, s_0 can be interpreted as the number of "unseen" data, and $s_k = s_0$ means that the kth information source has as much importance as our initial uncertainty.

Remark 1. The exact updated credal set

$$\tilde{K}_{(s_k;\xi_k)_{k=0}^m}(X) = \left\{ \frac{\sum_{k=0}^m s_k \xi_k}{\sum_{k=0}^m s_k} : \ \xi_k \in \mathcal{P}_k, \ \forall k = 1, \dots, m \right\} \tag{20}$$

is a subset of $K_{(s_k;\xi_k)_{k=0}^m}(X)$, i.e., $\tilde{K}_{(s_k;\xi_k)_{k=0}^m}(X) \subseteq K_{(s_k;\xi_k)_{k=0}^m}(X)$. The set (19) is thus an outer-approximation. Yet, the main advantages of using probability bounds as a basic representation are that

- their number of extreme points is bounded and relatively low, even when combining them through a weighted average. This is in general not the case if we consider averaging of heterogeneous simple representations: if we denote $|\mathcal{E}xt(\mathcal{P}_k)|$ the number of extreme points of the kth item of information, then their (Minkowsky) sum $\sum_{k=0}^m s_k \mathcal{P}_k$ may have as much as $\prod_{k=0}^m |\mathcal{E}xt(\mathcal{P}_k)|$ extreme points, an exponentially growing number;

- they are easy to explain and to represent graphically (e.g., as imprecise histograms), therefore offering a convenient way to communicate with domain experts or users not specialized in mathematics or computer science. This is not the case of more complex representations such as belief functions (Section 3.4);

- except for requiring a finite space, they do not require specific assumptions, such as the existence of an ordering between elements;

- they are expressive enough so that they can go from a fully precise probability to the complete ignorance model.

None of the other common practical models of information reviewed in Section 3 have all these advantages at once, making probability bounds a quite convenient model. Given this, using probability bounds seem a good general starting point in applications, not preventing one from investigating refined solutions if the results are unsatisfactory.

Of course, in some cases $K_{(s_k;\xi_k)_{k=0}^m}(X)$ may be a poor outer-approximation, however we shall see in Section 5 that it does not necessarily lead to completely void conclusions. Previous studies [1] also suggest that this kind of approximation may be in average reasonable.

Example 2. Let $\mathcal{X} = \{x_1, x_2, x_3\}$ and $\mathcal{P} = \{\xi : \xi(x_2) \geq \xi(x_1), \ 1/2 \geq \xi(x_3), \ \sum_{i=1}^3 \xi(x_i) = 1\}$ be an item of information. The credal set $\tilde{K}_\xi(X)$ over $\{x_1, x_2, x_3\}$ obtained by (20) has four extreme points $\{(0, 1, 0), (0, 0.5, 0.5), (0.5, 0.5, 0), (0.25, 0.25, 0.5)\}$ that are also extreme points of $K_\xi(X) = \{p : p(x_i) \leq \max_{\epsilon \in \mathcal{P}} \epsilon(x_i)\}$. However, the probability $(0.5, 0.25, 0.25)$ is an extreme point of $K_\xi(X)$ but not of $\tilde{K}_\xi(X)$.

3. Review of practical sources of information

Building $K_{(s_k;\xi_k)_{k=0}^m}(X)$ requires to obtain elements of information $\mathcal{P}_k$. In this section, we review different practical models and the bounds they induce over $\xi_k(x_i)$. We will also provide small examples illustrating what kind of information they can model. For the sake of brevity, we will denote $\xi(x_i)$ by ξ^i in this section. Note that our information is initially queried on observed values x_i, to be then transferred as knowledge on the parameters ξ^i. Hence we will consistently refer to knowledge about ξ^i, and to observation or information about x_i.

3.1. Precise evaluations

The most simple models is when the knowledge $\mathcal{P}$ is given by a precise vector, in which case $\xi^i = f_i$ is a precise number, and we have

$$\overline{\xi}^i = \underline{\xi}^i = f_i \tag{21}$$

A classical way to obtain such precise evaluations is when observing $\boldsymbol{m} = (m_1, \ldots, m_n)$ experiments, where m_i is the number of times x_i was observed. In such a case, a classical choice is to take as strength $s = m = \sum_{i=1}^n m_i$ and $\mathcal{P}$ is the vector $\boldsymbol{f} = (f_1, \ldots, f_n)$ where $f_i = {}^{m_i}/{}_{m}$.

Example 3. Assume we can observe three possibilities x_1, x_2, x_3 (e.g., severity of a disease, importance of a bacterial population), and we observed 3 times x_1, 6 times x_2 and one time x_3. We then have

$$f_1 = 0.3, f_2 = 0.6, f_3 = 0.1 \text{ and } s = m = 10 \tag{22}$$

Note that the more observations we accumulate, the stronger becomes this piece of knowledge. s can also be modulated to reflect the reliability of data. Note that this model is a degenerated case of probability intervals, and can therefore be exactly represented in our framework.

3.2. Numerical possibility distributions and fuzzy subsets

A possibility distribution π is simply a mapping from $\{\xi^1, \ldots, \xi^n\}$ to $[0, 1]$, with at least one element ξ^i such that $\pi(\xi^i) = 1$ [33]. In practice, we can see distribution π as an ordering $1 = \pi(\xi^{(1)}) \geq \ldots \geq \pi(\xi^{(n)})$ of the elements $x_1, \ldots, x_n$, from the most plausible to the least plausible one.

Another instrumental way is to encode the possibility distribution through the necessity measure N. This necessity measure N is such that

$$N(A_{(i)} = \{\xi^{(1)}, \ldots, \xi^{(i)}\}) = 1 - \pi(\xi^{(i+1)}) \tag{23}$$

with $\pi(\xi^{(n+1)}) = 0$. $N(A_{(i)})$ can be associated to a lower probability bound of event $A_{(i)}$. In particular, the sets $A_{(i)}$ can be interpreted as nested sets with an associated lower confidence, these nested sets being built by starting from the most plausible element $\xi^{(1)}$ and incrementally including the less plausible ones. Note that we may have $\pi(\xi^{(i)}) = \pi(\xi^{(i+1)})$, in which case elements $x_{(i-1)}$ and $x_{(i)}$ would be in the same confidence set.

In practice, an expert can provide a possibilistic information by giving confidence bounds over a collection of nested sets. Let $A_1 \subseteq \ldots \subseteq A_m$ be such sets with associated confidence levels $\alpha_1 \leq \ldots \leq \alpha_m$, then it encodes the knowledge $\mathcal{P}_\pi = \{\xi : \sum_{k=1}^{i} \xi^{(k)} \geq \alpha_i, \forall i\}$. From the knowledge on sets $A_1, \ldots, A_m$, one can always come back to an associated distribution π using

$$\pi(\xi^k) = \min_{i:\xi^k \in A_i} 1 - \alpha_{i-1} \tag{24}$$

with $\alpha_0 = 0$.

Another possibility is to use the formal equivalence between a possibility distribution π and a fuzzy set having π for membership function. This means that an expert conveying information in the form of linguistic assessment [64] can also be modelled by possibility distributions. Deriving bounds on ξ^i from $\mathcal{P}_\pi$ using the possibility distribution π is very easy, as

$$\underline{\xi}^i = 1 - \max_{\xi \neq \xi^i} \pi(\xi) \tag{25}$$

$$\overline{\xi}^i = \pi(\xi^i) \tag{26}$$

Example 4. Assume that an expert is interrogated about the temperature in a room that can be in three states x_1, x_2, x_3. Expert judges that x_2 is the most plausible state, then x_3 and x_1, meaning that $\xi^{(1)} = \xi^2, \xi^{(2)} = \xi^3, \xi^{(3)} = \xi^1$. The expert provides the following confidence values:

$$N(\{\xi^2\}) = 0.5$$

$$N(\{\xi^2, \xi^3\}) = 0.8$$

$$N(\{\xi^2, \xi^3, \xi^1\}) = 1$$

which means that the expert has a confidence 0.5 that x_2 will be the observed state, a confidence 0.8 that the observed state will be either x_2 or x_3, and finally is certain that the only observable states are x_1, x_2, x_3. From these values can be deduced the values of the corresponding possibility distribution $\pi(\xi^1) = 0.2$, $\pi(\xi^2) = 1$, $\pi(\xi^3) = 0.5$.

Alone, possibility distributions will often be simpler than probability intervals: they require less information (one value per element) and will have a maximal number of $2^{|\mathcal{X}|-1}$ extreme points [56]. Yet the average of multiple sets $\mathcal{P}_{\pi_1}, \ldots, \mathcal{P}_{\pi_m}$ would no longer be a possibility distribution, and the corresponding number of extreme points could explode. Also, possibility distributions cannot model precise probabilities, unless they are degenerate ones.

3.3. Probability boxes and clouds

A probability box [35] $\underline{F}, \overline{F}$ is an imprecise cumulative distribution. It can be modelled by two discrete non-decreasing functions $\underline{F}$ and $\overline{F}$ from $(\xi^1, \ldots, \xi^n)$ to $[0, 1]$ such that $\underline{F}(\xi^i) \leq \overline{F}(\xi^i)$ for all i in $\{1, \ldots, n\}$ and $\underline{F}(\xi^n) = \overline{F}(\xi^n) = 1$. The values $\underline{F}(\xi^i), \overline{F}(\xi^i)$ are interpreted as the following bounds

$$\underline{F}(\xi^i) \leq \sum_{i=1}^{n} \xi^i \leq \overline{F}(\xi^i)$$

and we can denote by $\mathcal{P}_{\underline{F} \leq \overline{F}}$ the knowledge modelled by a p-box. A p-box information provides us with estimates about the cumulated probabilities of events of the kind $\{x_1, \ldots, x_i\}$, hence assuming that the ordering induced by the indices do make sense.

In the case of p-boxes, the bounds over ξ^i are very easy to determine [59], and are equal to

$$\underline{\xi}^i = \max(0, \underline{F}(\xi^i) - \overline{F}(\xi^{i-1})) \tag{27}$$

$$\overline{\xi}^i = \overline{F}(\xi^i) - \underline{F}(\xi^{i-1}) \tag{28}$$

with the convention $\underline{F}(\xi^0) = \overline{F}(\xi^0) = 0$.

Example 5. Assume we have to assess the how likely it is that a bacterial population is below some threshold, or how likely it is that a component may function for a given period of time. The population sizes or time intervals may be discretized into x_1, x_2, x_3. Assume the following p-box has been given as information

$$\underline{F}(\xi^1) = 0.2, \underline{F}(\xi^2) = 0.7 \text{ and } \overline{F}(\xi^1) = 0.5, \overline{F}(\xi^2) = 0.9.$$

From it we can deduce the bounds

$$\underline{\xi}^1 = 0.2, \underline{\xi}^2 = 0.2, \underline{\xi}^3 = 0.1 \text{ and } \overline{\xi}^1 = 0.5, \overline{\xi}^2 = 0.7, \overline{\xi}^3 = 0.3.$$

P-boxes usually rely on the fact that the set $(\xi^1, \ldots, \xi^n)$ is naturally ordered, and provide confidence bounds over sets of the kind $\{\xi^1, \ldots, \xi^i\}$. However, one possibility is to extend this notion by considering that values ξ^i follows an arbitrary ordering $\xi^{(1)} \leq \ldots \leq \xi^{(n)}$ (for example, from the least to the most plausible element) and to ask to the expert to provide upper and lower confidence bounds about the fact that the truth lies in $\{x_{(1)}, \ldots, x_{(i)}\}$, thus obtaining $\underline{F}(\xi^{(i)})$ and $\overline{F}(\xi^{(i)})$. As in principle any ordering can be used, this is indeed a generalization of p-boxes, known as comonotonic clouds [31]. In particular, in the case where $\underline{F}(\xi^{(i)}) = 0$ for any i, we retrieve the notion of possibility distribution as a special case.

Up to now, what is the maximal number of extreme points of a p-box structure and how to efficiently enumerate them remains an open problem. However, as p-boxes are a special case of belief functions, one can use (potentially sub-optimal) algorithms and methods applicable to belief functions [17]. It is also clear that the maximal number of such points is bounded above by the maximal number of extreme point of a belief function ($n!$). Classical p-boxes suffer from the fact that a natural order must exist on $\mathcal{X}$, and when no such order exists, then the average of generalized p-boxes relying on different orders will not be a p-box.

3.4. Belief functions and random sets

Formally, a random set or belief function, initially introduced by Dempster [30] and Shafer [57], is defined as a positive mapping $\nu : 2^{\{\xi^1,\dots,\xi^n\}} \to [0,1]$ from the power set of $\{\xi^1,\dots,\xi^n\}$ to the unit interval, such that $\nu(\emptyset) = 0$ and $\sum_E \nu(E) = 1$. From this mapping can then be defined probability bounds $Bel(A), Pl(A)$ for any event that are equal to

$$Bel(A) = \sum_{E, E\subseteq A} \nu(E) \text{ and } Pl(A) = \sum_{E, E\cap A\neq\emptyset} \nu(E) = 1 - Bel(A^c) \quad (29)$$

that induce an information $\mathcal{P}_\nu$ such that

$$\mathcal{P}_\nu = \left\{ \xi : \sum_{E\subseteq A} \nu(E) \leq \sum_{\xi^i \in A} \xi^i \leq \sum_{E\cap A\neq\emptyset} \nu(E),\ \forall A \right\} \quad (30)$$

In particular, this means that given a function ν, the bounds over elementary events are given by

$$\underline{\xi}^i = Bel(\{\xi^i\}) = \nu(\{\xi^i\}) \quad (31)$$

$$\overline{\xi}^i = Pl(\{\xi^i\}) = \sum_{\xi^i \in E} \nu(E) \quad (32)$$

Belief functions are instrumental to model frequencies of imprecise observations, for example when multiple exclusive options can be chosen in surveys, or when some sensors sometimes send back imprecise observations. They also include p-boxes, comonotonic clouds and possibilities as special cases.

Example 6. Assume again that we can meet four different situations x_1, x_2, x_3 , x_4. Out of 20 observations, x_1, x_2, x_3, x_4 were each perfectly observed respectively $3, 2, 5, 6$ times, we observed 3 times the set $\{x_2, x_3, x_4\}$ (excluding x_1) and 2 times the set $\{x_1, x_2, x_3\}$. Such observations can be modelled on $\xi^1, \xi^2, \xi^3, \xi^4$ by the mass

$$\nu(\{\xi^1\}) = 3/20, \nu(\{\xi^2\}) = 2/20, \nu(\{\xi^3\}) = 5/20, \nu(\{\xi^4\}) = 6/20,$$

$$\nu(\{\xi^1, \xi^2, \xi^3\}) = 2/20, \nu(\{\xi^2, \xi^3, \xi^4\}) = 3/20.$$

From this, we can for example deduce $\underline{\xi}^3 = 0.25$ and $\overline{\xi}^3 = 0.5$.

Belief functions are general enough to deal with a lot of practical assessments, and share the properties of probability intervals that an average of belief functions is still a belief function. However, providing an intuitive graphical representation of a belief function is challenging, and their use may quickly lead to computational issues (e.g., their number of extreme points can be as high as $\mathcal{X}!$ [50])

3.5. Fuzzy random variables

Fuzzy random variables have been given different interpretations in the literature, depending on the nature of the fuzzy elements. For example, a fuzzy random variable can be seen as a random phenomenon with precise observations that are fuzzy in nature, or as a random phenomenon with imprecise observations. We refer to [19, 21, 22] for a detailed discussion. In this paper, Fuzzy random variables are interpreted as conditional possibility measures [4, 21], which consist in putting positive masses, not on subsets, but on possibility distributions. They can be modelled by a set $\pi_1, \ldots, \pi_k$ where each distribution receives probability mass $p(\pi_i)$. As each π_i can in turn be turned into a mass function ν_{π_i} defined this time over subsets, it is always possible to come back from a fuzzy random variable to a classical mass function, simply by computing for any subset E the value

$$\nu(E) = \sum_{i=1}^{k} p(\pi_i)\nu_{\pi_i}(E).$$

We obtain a weighted random sampling of subset E defining a belief function ν. Fuzzy random variables in this context may be cast into the framework of belief functions leading to the same formal advantages and disadvantages of them (see Section 3.4). Fuzzy random variables can result, for instance, from Monte-Carlo simulations of physical models mixing possibilistic and probabilistic uncertainty [4], or from the random observation of fuzzy sets (modelling an ill-calibrated scale, for instance [18]).

3.6. Summary of types of knowledge

A final type of knowledge simply consists in directly providing bounds over the values of possible observations x_i. This means specifying, for each ξ^i, the bounds $\underline{f}_i = \underline{\xi}^i$ and $\overline{f}_i = \overline{\xi}^i$. Such bounds are formally equivalent to probability intervals [11].

There are multiple ways to derive such bounds: for instance by instantiating multinomial confidence intervals over observations, by requiring linguistic opinions of the type "probable", "very probable" from the experts and then translating them into numerical evaluations [54], by simply requiring numerical evaluations from the experts, when having imprecise histograms, ...

Table 1 summarises the most common type of practical information one can meet, to what type of information they correspond and how can be computed the lower/upper values $\underline{\xi}^i$ and $\overline{\xi}^i$. These are the values (used as a common mathematical tool) that are then combined and integrated into the learning process developed in Section 2.2.

Model	Usual type of information	$\underline{\xi}^i$	$\overline{\xi}^i$
Precise values	Sample/simulation	f_i	f_i
Possibility	Lower confidence on nested sets Linguistic assessments	$1 - \max_{\xi \neq \xi^i} \pi(\xi^i)$	$\overline{\xi}^i = \pi(\xi^i)$
P-boxes and clouds	Lower/upper confidence on nested sets	$\max(0, \underline{F}(\xi^{(i)}) - \overline{F}(\xi^{(i-1)}))$	$\overline{F}(\xi^{(i)}) - \underline{F}(\xi^{(i-1)})$
Belief functions	Imprecise sample	$\nu(\xi^i)$	$\sum_{\xi^i \in E} \nu(E)$
Fuzzy random variable	Fuzzy sample	$\sum_{i=1}^k p(\pi_i)\nu_{\pi_i}(\xi^i)$	$\sum_{\xi^i \in E} \sum_{i=1}^k p(\pi_i)\nu_{\pi_i}(E)$
Probability bounds	Linguistic assessments Multinomial confidence regions	$\underline{f}_i$	$\overline{f}_i$

Table 1: Summary of the different types of collectible information

4. Robust dynamic probabilistic graphical models

When modeling complex systems, we are not interested in a single variable, but in multiple variables interacting with each others and evolving over time. In theory, our knowledge about these variables, their interaction and evolution can be represented by a credal set defined over the Cartesian product of the corresponding spaces.

In practice, we need tool to represent these interactions, and to simplify the daunting task of specifying a full joint model. Credal networks are graphical (directed) models that aims at encoding our knowledge about variable interactions and at splitting the full joint into multiple, simple conditional models. This section introduces them, as well as their dynamical extension.

4.1. Credal networks

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a discrete random vector associated with the joint probability mass function $p(\mathbf{X})$ defined over $\prod_{i=1}^{n} \mathcal{X}_i$. Let $K(\mathbf{X})$ be the closed convex set of multivariate probability mass functions describing our knowledge of $\mathbf{X}$.

A credal network (CN) [24, 23] is an extension of Bayesian networks (BNs) where imprecision is introduced in probabilities by means of credal sets [47]. When working with probability sets rather than precise probabilities, the notion of stochastic independence can be extended in several ways [20]. Within graphical models, the most commonly used extension is *strong independence* (also called a *type 1 product* of the marginals in [61]), that induces the strong extension. It can be interpreted as a robust model of a precise yet ill-known BN. Under the *strong extension* [23] hypothesis, the joint credal set $K(\mathbf{X})$ over $\Omega_{\mathbf{X}}$ may be formulated as:

$$K(\mathbf{X}) = CH \left\{ p(\mathbf{X}) : p(\mathbf{X}) = \prod_{i=1}^{n} p_i, p_i \in K_i \right\} \tag{33}$$

where $p_i = p\left(X_i \mid \mathbf{U}_i\right)$, $\mathbf{U}_i$ denotes the set of parent nodes of the node X_i and $K_i = K\left(X_i \mid \mathbf{U}_i\right)$ is the closed convex set of probability mass function for the random variable X_i given $\mathbf{U}_i$. As mentioned in Section 2, it is sufficient to focus on $\mathcal{E}xt(K\left(X_i \mid \mathbf{U}_i\right))$ in Eq. (33).

In this work, we focus on the notion of strong independence and its extension to dynamical models, as this is the most widely used independence notion within graphical models and the one that fits the best with a robust interpretation of probability sets. Other independence notions that may even have asymmetrical versions such as epistemic irrelevance remain computationally intractable [25, 49], except for specific network structures [9] that are usually less complex than the one generally considered here.

4.2. Dynamic credal networks

Let $\mathbf{X}(\mathbf{t}) = (X_1(1), \ldots, X_n(1), \ldots, X_1(\tau), \ldots, X_n(\tau))$ be a discrete random vector process associated with the joint probability function $p(\mathbf{X}(\mathbf{t}))$ defined over $\prod_{t=1}^{\tau}\prod_{i=1}^{n}\mathcal{X}_i(t)$. Let $K(\mathbf{X}(\mathbf{t}))$ be the closed convex set of multivariate probability mass functions for $\mathbf{X}(\mathbf{t})$.

A dynamic credal network (CDN) [41] is a dynamic Bayesian network (DBNs) [51] where conditional probabilities $p\left(X_i\left(t\right) \mid \mathbf{U}_i\left(t\right)\right)$ (noted p_i^t) are replaced by credal sets $K\left(X_i(t) \mid \mathbf{U}_i(t)\right)$ (noted K_i^t). It is therefore a time-sliced model that can be used to describe a dynamic process or system[2].

We assume the same *first-order Markov property* as for DBN, meaning that parents only originate from the same or previous time slice, and also that conditional models remain the same at each times slice, that is

$$K\left(X_i(t) \mid \mathbf{U}_i(t)\right) = K\left(X_i(2) \mid \mathbf{U}_i(2)\right),\ \forall t \in [\![2, \tau]\!]. \tag{34}$$

Therefore, specifying the graphical structure of a DCN requires the same effort as the one of a DBN (that is, specifying only two consecutive time slices) but allows the user to provide conditional credal sets rather than probabilities if these latter cannot be reliably estimated (from data and/or experts).

4.2.1. Independence in DCN

Extending DBN to DCN requires to specify which kind of independence we consider within and also between each time-slice. We remind that we will only consider extensions relying on the strong independence (33). The most straightforward extension is to simply apply strong independence to the whole network, i.e.,

$$K(\mathbf{X}(\mathbf{t}))_{st} = CH\left\{p(\mathbf{X}(\mathbf{t})) : p(\mathbf{X}(\mathbf{t})) = \prod_{i=1}^{n}\prod_{t=1}^{\tau} p_i^t, p_i^t \in K_i^t\right\} \tag{35}$$

We call this extension the *dynamic strong extension* and it is worth noticing that we can have $p_i^t \neq p_i^{t'}$ for $t, t' \in [\![2, \tau]\!]$. That is, we do not assume probabilities within each time-slice to be identical. However, when stepping to dynamic models, Condition (34) allows us to use the notion of *repetitive independence* (also called a *type 2 product* of the marginals in [61]). This condition states that if two variables X, Y have the same set of possible outcomes, that is $\mathcal{X} = \mathcal{Y}$, and can be assumed to be governed by the same probability distribution belonging to $K\left(X\right)$, then the joint credal set $K\left(X, Y\right)$ is :

$$K\left(X, Y\right) = CH\{p(X)p(X) : p(X) \in K\left(X\right)\}. \tag{36}$$

[2]It should be noted that the network itself is static, but is used to represent a dynamic process.

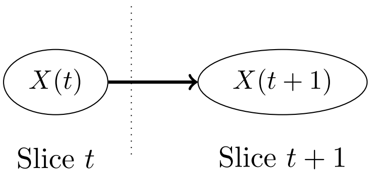

Figure 2: A simple dynamical graphical model

Adapting this notion of independence to DCN, so that probabilities of each time slice are assumed to be identical, leads to a second extension, i.e.,

$$K(\mathbf{X}(\mathbf{t}))_{rp} = CH \left\{ \begin{array}{l} p(\mathbf{X}(\mathbf{t})) : p(\mathbf{X}(\mathbf{t})) = \prod_{i=1}^{n} \prod_{t=1}^{\tau} p_i^t, \\ p_i^2 \in K_i^2 \text{ and } p_i^t = p_i^2 \ \forall t \in [\![2, \tau]\!] \end{array} \right\} \tag{37}$$

that we call the *dynamic repetitive extension*. We have $K(\mathbf{X}(\mathbf{t}))_{rp} \subseteq K(\mathbf{X}(\mathbf{t}))_{st}$, as $K(\mathbf{X}(\mathbf{t}))_{rp}$ is more constrained. In practice, the strong extension assumes that the dynamic network is ill-defined and that its behaviour can change between time slices, while the repetitive extension assumes that we seek a precise classical DBN who is partially known.

Example 7. Consider the very simple example where $\mathcal{X} = \{0, 1\}$ and the 2-slice network given in Figure 2, which is nothing else than a two-state imprecise Markov chain, and an observed value $X(1) = 1$. Assume furthermore that we have three time slices ($\tau = 3$), that $X(1) = 1$ is observed, and that

$$p(X(t) = 1 | X(t-1) = 0) = 0.8,$$
$$p(X(t) = 1 | X(t-1) = 1) \in [0.2, 0.5].$$

That is, the transition rates from state 0 are precisely known, but not the one from state 1 (although staying in state 1 is clearly less likely). The different extreme points over $(X(1), X(2), X(3))$ resulting from the strong and repetitive extension are summarized in Table 2, in which we adopt the notation $x(t)$ for $X(t) = 1$ for simplification purposes. Each cell of the table corresponds to a precise network obtained by a specific selection of extreme points. The non-specified transition probabilities can be retrieved by the formula $p(X(t) = 1 | X(t-1) = 1) = 1 - p(X(t) = 0 | X(t-1) = 1)$.

4.2.2. Inference algorithms in DCN

(D)CNs can be queried as in (D)BNs to get information about the state of a variable given evidence about other variables, with respect to the chosen network *extension*. However, the use of credal sets makes the updating problem much harder, as it becomes an optimization problem. As such, the computation of the lower bound on $p(\mathbf{X}_Q \mid \mathbf{X}_E)$ requires to minimize a fraction containing polynomials :

Strong extension				Repetitive extension			
100	101	110	111	100	101	110	111
$P(x(t)\|x(t-1)) = 0.5$				$P(x(2)\|x(1)) = 0.5, P(x(3)\|x(2)) = 0.2$			
0.1	0.4	0.25	0.25	0.1	0.4	0.4	0.1
$P(x(t)\|x(t-1)) = 0.2$				$P(x(2)\|x(1)) = 0.2, P(x(3)\|x(2)) = 0.5$			
0.16	0.64	0.16	0.04	0.4	0.4	0.1	0.1

Table 2: Simple DCN extreme probabilities

$$\underline{p}(\mathbf{X}_Q(t) \mid \mathbf{X}_E(t)) = \min_{p(\mathbf{X}(\mathbf{t}))\in K(\mathbf{X}(\mathbf{t}))_\omega} \frac{\sum_{X_i(t)\in\mathbf{X}(\mathbf{t})\setminus\mathbf{X}_Q(t)\cup\mathbf{X}_E(t)} \prod_{i=1}^{n}\prod_{t=1}^{\tau} p_i^t}{\sum_{X_i(t)\in\mathbf{X}(\mathbf{t})\setminus\mathbf{X}_E(t)} \prod_{i=1}^{n}\prod_{t=1}^{\tau} p_i^t} \tag{38}$$

with $p(\mathbf{X}(\mathbf{t})) \in K(\mathbf{X})_\omega$ belonging to the *dynamic strong extension* ($\omega = st$) or *dynamic repetitive extension* ($\omega = rp$) of the network. An upper bound can be obtained by maximizing (38). It is known that such a minimum (or maximum) is obtained at a vertex of the *dynamic strong/repetitive extension*. Depending on (1) the structure of network, (2) the number of modality of variables and (3) the chosen extension (strong/repetitive), the updating problem will be more or less complex to solve. Because inferences are already hard in static credal networks, little work has been done on DCNs [41]. By unrolling a two-time slice network over T time steps, the number of possible vertex combinations goes from $\prod_{i,t=0} \#\mathcal{E}xt(K_i^t) \prod_{i,t=1} \#\mathcal{E}xt(K_i^t)$ in the case of repetitive independence, to $\prod_{i,t=0} \#\mathcal{E}xt(K_i^t) \prod_{i,t=1} \#\mathcal{E}xt(K_i^t)^{\tau-1}$ in the case of strong independence. Given the potential number of vertices, approximate algorithms seem more appropriate regarding DCNs.

Many algorithms, exact and approximate, have been proposed to deal with CN. Some are generalizations of well known (D)BNs algorithms. Among the approximate algorithms, there are those that compute inner bounds, i.e. bounds that are enclosed by the exact ones, outer bounds, which enclose the exact ones, and those that perform randomly. The 2U algorithm [34] performs an exact rapid inference in the case of binary tree-shaped (D)CNs with the assumption of *strong independence*. The CCM transformation [15] turns a (D)CN into a (D)BN by adding transparent nodes before performing an Maximum A Posteriori (MAP) estimation over the latter to find the best combination of vertices. It has the same complexity as credal network inference, that is $NP^{PP}Complete$, and performs poorly with separately specified credal networks such as the one we used during our trials (because of the sheer number of vertices). Optimization techniques such as branch and bound over local vertices of credal sets [27, 13] are also well suited to medium-sized networks and can be stopped at any time to give an approximate answer. Other algorithms are based on a variable elimination scheme from (D)BNs, such as Separable Variable Evaluation [26, 55] which keeps the separately specified credal sets as separated as possible during propagation, and can be mapped to an integer or a multi-linear program [29, 28]. Regarding binary and DAG-shaped (DAG : Directed Acyclic Graph) credal networks, algorithm L2U (Loopy 2U) [44] (similar to LBP (Loopy Belief Propagation) [63]) produces either inner or outer approximations. Its efficiency is due both to the bounded cardinality of variables and to ignoring loops. Another way to handle credal sets complexity is to represent them by simpler means. Variational methods [43, 42] choose a family of functions to approximate the exact combination of credal sets to decrease computational costs. Those functions are optimized according to some criteria until convergence and the inference is then realized in the network with the original credal sets replaced by the new found functions. The A\R(+) algorithm [27] uses interval probability arithmetic to approximate credal sets in a propagation scheme in tree-shaped networks (with the use of some additional constraints limiting the information loss in its enhanced version). The intervals produced are outer bounds of the real ones. Although those algorithms are fast in medium-sized network, they either produce too many approximations or are too complex to work with DCNs. Another popular family of approximate algorithms producing inner bounds is based on Monte-Carlo sampling [38]. Several methods have been proposed to better guide the search (simulated annealing [12], genetic algorithms [14]) among the vertices of the (conditional) local credal sets, but they require some tuning for more accurate results, otherwise they can lead to poor approximations.

Although there exist several inference algorithms, none allows to do inference, in a realistic and practical way, on networks capable of representing global complex system of Life Sciences. In further inferences, we used a simple Monte-Carlo sampling algorithm [38] which has the advantage to be a good starting point, as it applies with the same easiness to *dynamic repetitive* and *strong extensions* (with a faster convergence for *dynamic repetitive extension*).

4.2.3. Robust parameter learning

Let p_{ijk}^t be the probability that $X_i(t) = x_k$, given that its parents have instantiation[3] x_j (corresponding itself to a vector where j represents the

[3]Possible values of variables according to its discretization.

vector of parents of i), *i.e.*

$$p_{ijk}^t = p(X_i(t) = x_k | \mathbf{U}_i(t) = x_j) \qquad \begin{array}{l} i = 1, \ldots, n \\ j = 1, \ldots, c_i \\ k = 1, \ldots, r_i \end{array} \tag{39}$$

where r_i is the number of values that node i can take and c_i is the number of distinct configurations of $\mathbf{U}_i(t)$. Parameter learning consists in estimating p_{ij}^t faced with available information [45, 10]. For the sake of clarity, parameters p_{ij}^t will be denoted p_{ij} since parameters p_{ij}^t are time-invariant in the case of *repetitive extension* assumption and it is sufficient to only consider information limited to each time slice in the case of *strong extension* assumption. According to section 2.2, for all $i \in \{1, \ldots, n\}$; $j \in \{1, \ldots, c_i\}$ the credal set $\tilde{K}_{(s_l;\xi_l)_{l=0}^m}(X_i| \ \mathbf{U}_i = x_j)$ may be approximated by using the outer credal set $K_{(s_l;\xi_l)_{l=0}^m}(X_i| \ \mathbf{U}_i = x_j)$ defined by

$$K_{(s_l;\xi_l)_{l=0}^m}(X_i|\mathbf{U}_i = x_j) = \left\{ p_{ij} : \ p_{ijk} \in [\underline{p}_{ijk}, \overline{p}_{ijk}], \sum_k p_{ijk} = 1 \right\} \tag{40}$$

where $[\underline{p}_{ijk}, \overline{p}_{ijk}]$ is estimated and updated from Eq. (19) according to the available sources of knowledge $(S_0, \ldots, S_m)$.

4.3. Practical robust parameter learning example

Wood is essentially composed of cellulose (denoted C) that is a polymer whose quantity characterizes the nature of wood (denoted T) namely hardwood or softwood. Imagine that we want to determine the kind of wood according to its chemical composition tainted with uncertainties, that is we are interested in $P(T|C)$. For the sake of clarity, we choose $C = \{x_1 = 20\%, x_2 = 40\%, x_3 = 60\%\}$ meaning that there is 20%, 40% or 60% of cellulose inside wood, $T = \{x_1 = \text{Soft}, x_2 = \text{Hard}\}$ and all sources s_i have the same confidence level, *i.e.* $s_i = 1$ for all i. We thus need to estimate the following parameters:

$$p_{jk} = p(T = x_k | C = x_j) \tag{41}$$

according to the available knowledge described in the following. The credal sets $K(T|C = x_j)$ are initialized by

$$K_{s_0}(T|C = x_j) = \{p_{j.} : p_{jk} \geq 0, \sum_k p_{jk} = 1\}, \ \forall j = 1, \ldots, 3 \tag{42}$$

1. Precise measures are provided $\{(20, \text{Soft}), (20, \text{Hard}), (40, \text{Hard}), (60, \text{Soft})\}$ leading to update by Eq. (19)

   - $K_{(s_0,s_1)}(T|C = 20) = \{p_{1.} : \frac{1}{4} \leq p_{11} \leq \frac{3}{4}, p_{12} = 1 - p_{11}\}$,

- $K_{(s_0,s_1)}(T|C=40) = \{p_{2.} : 0 \leq p_{21} \leq \frac{1}{2}, p_{22} = 1 - p_{21}\}$,
- $K_{(s_0,s_1)}(T|C=60) = \{p_{3.} : \frac{1}{2} \leq p_{31} \leq 1, p_{32} = 1 - p_{31}\}$.

2. A first expert says that the more cellulose there is, the harder the wood. This information may be formalized by means of the following fuzzy numbers or possibility distribution (see Section 3.2):

   - $\pi(T = \text{Hard}|C=20) = 0.5, \quad \pi(T = \text{Soft}|C=20) = 1$
   - $\pi(T = \text{Hard}|C=40) = \pi(T = \text{Soft}|C=40) = 1$
   - $\pi(T = \text{Hard}|C=20) = 1, \quad \pi(T = \text{Soft}|C=20) = 0.5$

   meaning for instance that $P(T = \text{Hard}|C = 20) \leq 0.5$ leading to update

   - $K_{(s_0,s_1,s_2)}(T|C=20) = \{p_{1.} : \frac{1}{3} \leq p_{11} \leq \frac{5}{6}, p_{12} = 1 - p_{11}\}$,
   - $K_{(s_0,s_1,,s_2)}(T|C=40) = \{p_{2.} : 0 \leq p_{21} \leq \frac{2}{3}, p_{22} = 1 - p_{21}\}$,
   - $K_{(s_0,s_1,,s_2)}(T|C=60) = \{p_{3.} : \frac{1}{3} \leq p_{31} \leq \frac{5}{6}, p_{32} = 1 - p_{31}\}$.

3. A second expert provides more accurate estimation in terms of confidence

   - $P(T = \text{Soft}|C=20) \geq 95\%$,
   - $P(T = \text{Hard}|C=40) \geq 60\%$,
   - $P(T = \text{Hard}|C=60) \geq 95\%$.

   which can be modeled again by a possibility distribution. This leads to update

   - $K_{(s_0,\ldots,s_3)}(T|C=20) = \{p_{1.} : 0.49 \leq p_{11} \leq 0.875, p_{12} = 1 - p_{11}\}$,
   - $K_{(s_0,\ldots,s_3)}(T|C=40) = \{p_{2.} : 0 \leq p_{21} \leq 0.6, p_{22} = 1 - p_{21}\}$,
   - $K_{(s_0,\ldots,s_3)}(T|C=60) = \{p_{3.} : 0.25 \leq p_{31} \leq 0.64, p_{32} = 1 - p_{31}\}$.

4. Defective sensors and measurements provide joint imprecise observations, summarized in Table 3 and producing a joint belief function (Section 3.4).

   From this information lower and upper probability bounds over parameters are given by

$$Bel(T=t|C=c) = \frac{Bel(T=t, C=c)}{Bel(T=t, C=c) + \sum_{t' \neq t} Pl(T=t', C=c)}$$

$$Pl(T=t|C=c) = \frac{Pl(T=t, C=c)}{Pl(T=t, C=c) + \sum_{t' \neq t} Bel(T=t', C=c)}$$

		Type		
	Focal sets	Soft	Hard	{Soft,Hard}
Cellulose	20	1	2	0
	40	4	6	1
	60	0	10	1
	$\{20, 40\}$	10	5	1
	$\{20, 60\}$	0	0	5
	$\{40, 60\}$	1	3	10
	$\{20, 40, 60\}$	0	0	8

Table 3: Focal sets occurrences

For example

$$Bel(T = \text{Soft}|C = 20) = \frac{1/68}{1/68 + 21/68} = 0.045$$

$$Pl(T = \text{Soft}|C = 20) = \frac{24/48}{24/48 + 2/48} = 0.926$$

Credal set $K_{(s_0,\ldots,s_4)}$ is then updated by

- $K_{(s_0,\ldots,s_4)}(T|C = 20) = \{p_{1.} : 0.4 \leq p_{11} \leq 0.89, p_{12} = 1 - p_{11}\}$,
- $K_{(s_0,\ldots,s_4)}(T|C = 40) = \{p_{2.} : 0.021 \leq p_{21} \leq 0.65, p_{22} = 1 - p_{21}\}$,
- $K_{(s_0,\ldots,s_4)}(T|C = 60) = \{p_{3.} : 0.2 \leq p_{31} \leq 0.67, p_{32} = 1 - p_{31}\}$.

5. Real-life case study

To illustrate the feasibility and practical use of our approach in a real case, we have focused on the ripening process of the Camembert type soft mould cheese that represents an ecosystem and a bioreactor difficult to apprehend from a global point of view [37, 52]. Based on recent works carried out by Baudrit *et al.* [5]; Sicard *et al.* [48], a simplified sub-structure of dynamic Bayesian networks has been extracted (see Figure 3) providing a qualitative representation of the coupled dynamics of yeast behaviour *Kluyveromyces marxianus* (Km, colony forming unit/g of Fresh Cheese in decimal logarithmic scale) with its lactose substrate (lo, g/Kg of Fresh Cheese) influenced by temperature (T, ^{o}C) inside the ripening chamber and involving odour changes ($Od = \{$Fresh,Mushroom,Camembert$\}$).

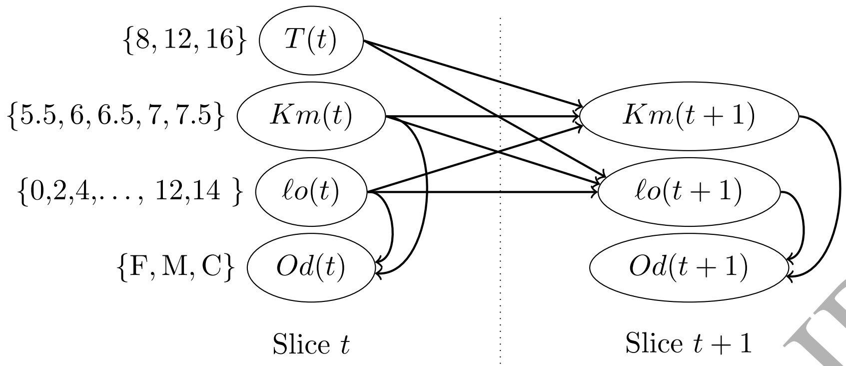

Figure 3: Structure of the dynamic credal network and the values of each variables representing the coupled dynamics Km growth versus lo consumptions influenced by temperature involving odour changes during the cheese ripening process (F=Fresh,M=Mushroom,C=Camembert).

5.1. Parameter learning

Assuming *repetitive extension* for computational reason mentioned in Section 4.2.2, we present, in the following, how parameters

$$\begin{aligned}
&\boldsymbol{p_1} = p(Km(1)),\\
&\boldsymbol{p_2} = p(lo(1)),\\
&\boldsymbol{p_3} = p(T(1)),\\
&\boldsymbol{p_4} = p(Km(2)|(Km(1), lo(1), T(1))),\\
&\boldsymbol{p_5} = p(lo(2)|(Km(1), lo(1), T(1))),\\
&\boldsymbol{p_6} = p(Od(1)|(Km(1), lo(1)))
\end{aligned}$$

may be estimated by using the robust hybrid parameter learning when we have several sources of knowledge (denoted S_i) tainted with stochastic and epistemic uncertainty.

1. Initialization (S_1).
   All DCN parameters are initialized by:
   - An experimental database $S_{experiments}$ of six cheese ripening trials carried out for temperatures varying from T = 8 to 16 ^{o}C is available.
   - the vacuous credal sets leading to bracket parameters by [0,1] when no information is available.

   With s_1 corresponding to the confidence level about experimental trials $S_{experiments}$, according to (19) we have:

$$p_{ijk} \in \left[\frac{s_1 f_{ijk}}{s_0 + s_1}, \frac{s_0 + s_1 f_{ijk}}{s_0 + s_1}\right] \tag{43}$$

where f_{ijk} represents the observed frequency corresponding to sample information in Table 1 and linked to Section 3.1.

2. Integration of partial mechanistic model tainted with uncertainties (S_2). The yeast Km is one of the dominant species in the yeast flora of Camembert cheeses and its principal activity is the consumption of lactose (lo) [46]. Models to determine the growth of microorganisms have been studied in the fermentation industry [58], and the description of the growth of Km is obtained by performing material balances on biomass Km and lactose lo [3]:

$$(S)\left\{\begin{array}{lcl} \dfrac{dKm}{dt} & = & \mu\dfrac{lo}{K_{lo}+lo}Km - b\cdot Km \\ \dfrac{dlo}{dt} & = & -\dfrac{\mu}{\beta}\dfrac{lo}{K_{lo}+lo}Km \end{array}\right. \tag{44}$$

where μ (the maximum specific growth rate of Km), $K_{lo}(T)$ (the half saturation constant for growth), b (the decay coefficient) and β (the yield coefficient for Km on lactose), depending on temperature, are tainted with stochastic and epistemic uncertainties, due to the natural variability of yeast population and the imperfection of the model. The background knowledge about parameters $\boldsymbol{p_1}$, $\boldsymbol{p_2}$, $\boldsymbol{p_4}$ and $\boldsymbol{p_5}$ are then updated regardless of the rest of network by using a simulated database $S_{simulated}$ resulting from Monte Carlo simulation coupled to interval analysis [4] leading to manage a joint random set $([\underline{Km}(t),\overline{Km}(t)],[\underline{lo}(t),\overline{lo}(t)],T(t))_l$ associated with mass $\nu_l = 1/\#S_{simulated}$ such that for instance

$$p_{4jk} \in \left[\frac{s_1 f_{4jk} + s_2\underline{\xi}_{4jk}}{s_0+s_1+s_2}, \frac{s_0 + s_1 f_{4jk} + s_2\overline{\xi}_{4jk}}{s_0+s_1+s_2}\right] \tag{45}$$

where

$$\underline{\xi}_{4jk} = \frac{bel(j,k)}{bel(j,k)+\sum_{l\neq k} pl(l,j)} \text{ and } \overline{\xi}_{4jk} = \frac{pl(j,k)}{pl(j,k)+\sum_{l\neq k} bel(l,j)} \tag{46}$$

and

$$pl(j,k) = \sum_{\substack{l,\,[\underline{km}(t+1),\overline{km}(t+1)]_l \cap \{km_k\} \neq \emptyset \\ \left[\underline{km}(t),\overline{km}(t)\right]_l \cap \{km_j\} \neq \emptyset \\ \left[\underline{lo}(t),\overline{lo}(t)\right]_l \cap \{lo_j\} \neq \emptyset \\ T_l(t) = T_j}} \nu_l \tag{47}$$

and

$$bel(j,k) = \sum_{\substack{l,\{km_k\} \subseteq [\underline{km}(t+1), \overline{km}(t+1)]_l \\ \{km_j\} \subseteq \left[\underline{km}(t), \overline{km}(t)\right]_l \\ \{lo_j\} \subseteq \left[\underline{lo}(t), \overline{lo}(t)\right]_l \\ T_l(t) = T_j}} \nu_l \qquad (48)$$

This kind information is linked to Sections 3.4, 4.3 and corresponds to Belief functions in Table 1.

3. Integration of expert knowledge, (S_3).

In cheese ripening, as in every complex food process, most of the control measures are performed on the basis of the expert's sensory perceptions. Indeed, experts have in mind the ripening process that they oversee and they are able to explain part of the complex reactions through their perception of quality changes [16]. Expert elicitation [48] informs us that during the exponential growing of the yeast Km, a characteristic fresh or lactic odour is released. Mushroom odour appears when the concentration of the yeast Km begins to stabilize and typical Camembert odour appears when the population of Km begins to decay. From this qualitative information, general rules may be deduced such as "it is impossible to have a Camembert odour with a weak (*resp.* high) concentrations of Km (*resp.* lo)". That means for several combinations of Km and lo concentrations, likely values about variable *Odour* may be formalized by means of possibility distributions $\pi_{Odour}(.|Km, lo)$. That is:

- When there is a high (*resp.* weak) concentration of lactose (*resp.* the yeast Km), having a fresh odour is the most plausible state, followed by Mushroom and Camembert odours, which can be formalized as the following possibility distribution:

$$\pi_{Odour}(\text{Fresh}|j) = 1$$
$$\pi_{Odour}(\text{Mushroom}|j) = 0.8$$
$$\pi_{Odour}(\text{Camembert}|j) = 0.2$$

  where $j = (Km \leq 6.5, lo \geq 8)$.

- When there is a medium concentration of lactose and Km, the Mushroom odour is the most plausible state but we cannot exclude having Fresh or Camembert odours, formalized by:

$$\pi_{Odour}(\text{Fresh}|j) = 0.2$$
$$\pi_{Odour}(\text{Mushroom}|j) = 1$$
$$\pi_{Odour}(\text{Camembert}|j) = 0.2$$

where $j = (Km = 7, 2 < lo < 8)$.

- When having very weak (*resp.* high) concentration of lactose (*resp.* the yeast Km), the Camembert odour is the most plausible state, followed by Mushroom and Fresh odours, formalized by:

$$\pi_{Odour}(\text{Fresh}|j) = 0.2$$
$$\pi_{Odour}(\text{Mushroom}|j) = 0.8$$
$$\pi_{Odour}(\text{Camembert}|j) = 1$$

where $j = (Km > 7, lo < 2)$.

Parameter $\boldsymbol{p}_6$ is then updated by using

$$p_{6jk} \in \left[\frac{s_1 f_{6jk} + s_3 \underline{\xi}_{6jk}}{s_0 + s_1 + s_3}, \frac{s_0 + s_1 f_{6jk} + s_3 \overline{\xi}_{6jk}}{s_0 + s_1 + s_3}\right] \tag{49}$$

where

$$\overline{\xi}_{6jk} = \pi_{Odour}(k|j) \text{ and } \underline{\xi}_{6jk} = 1 - \max_{l \neq k} \pi_{Odour}(l|j) \tag{50}$$

This kind of information is linked to the Section 3.2 and corresponds to possibilistic model in Table 1.

5.2. Inference results and discussion

We attempt to estimate the lower and upper mean time evolution of Km, lo and $Odour$ for a temperature control according to the previous parameter learning. That is

$$\underline{E}(X(t)|\mathbf{U}(t)) = \sum_k x_k \underline{p}(X(t) = x_k|\mathbf{U}(t)) \tag{51}$$

for the lower bounds where X may be Km, lo, $Odour$; $\mathbf{U}(t) = (lo(0), Km(0), T(0), \ldots, T(t))$ and

$$\underline{p}(X(t) = x_k|\mathbf{U}(t)) = \inf_{p \in K(X(t)|\mathbf{U}(t))} p(X(t) = x_k|\mathbf{U}(t)) \tag{52}$$

by assuming $s_0 = s_1 = s_2 = s_3 = 1$ and the *repetitive independence*, since we assume that transition probabilities remain the same along the process (there is no reason to assume a change in the bacteria population behaviour), but are ill-known due to insufficient experiments and information. Figure 4 displays the lower and upper simulated mean evolution of Km, lo, $Odour$ versus experimental data over the cheese ripening carried out at 12^oC each time a source of information is added. Supported by Table 4, we may observe that the imprecision of simulated results well decreases (characterized by the surface in gray).

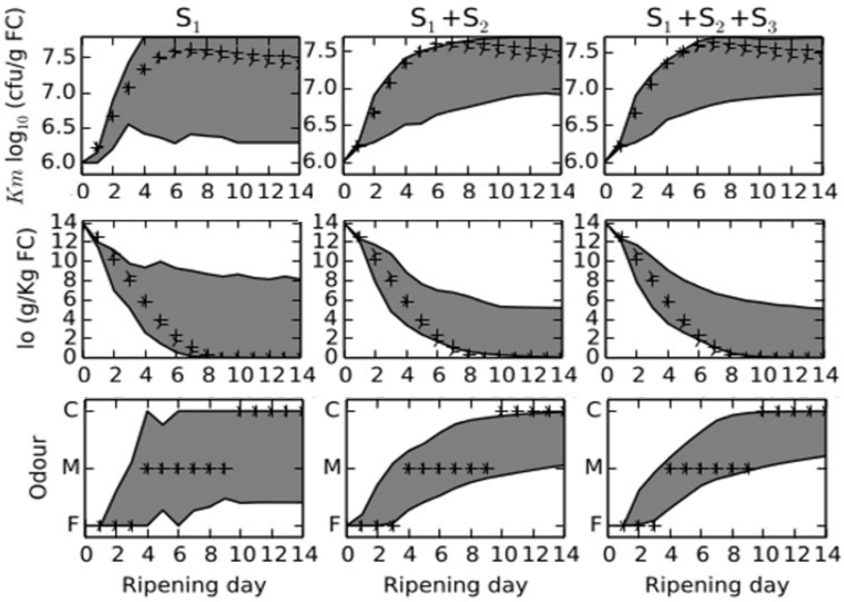

Figure 4: Incremental DCN average simulation versus raw data (dotted) of Km, lactose (lo) and Odour for a ripening carried out at and T=12^o C each time a new source of information is integrated.

	Km	lactose	Odour
Source 1	18.62	105.38	20.91
Source 1 & 2	11.37	71.06	13.77
Source 1& 2& 3	11.03	71.32	10.99

Table 4: Area between the lower and upper bounds of the simulated mean time evolution

6. Conclusion

There are complex dynamical processes for which no deterministic model describing the complete process exists. In such cases, dynamic credal networks are convenient models that allow to include expert knowledge, data and variable interaction in a single framework. They allow a faithful representation of incomplete knowledge or scarce data, that are inherent to the complexity of bio-physicochemical phenomena occurring in Life Sciences. In this paper, we attempted to implement a practical methodology coupling interval analysis and Dirichlet model in the framework of dynamical credal networks for building mathematical model capable of representing complex systems. Moreover, the concepts of dynamic repetitive and strong extensions have been proposed. While the latter can be seen as a straightforward extension of classical credal networks, the former considers repetitive independence to allow the model to preserve a temporal regularity. Methodology has been applied to a simplified real-case study concerning microbial population growth involving sensory evolution during cheese ripening. These experiments have shown that including information reduces imprecision about result simulations. Next tools should consider to manipulate, to combine convex sets in order to not lose information during incremental parameter learning. In further works, DCNs should enable us to determine the contribution of imprecision and/or incompleteness on the outcomes of a model in order to know if an ambiguous answer is due to a lack of information or due to a random phenomenon. That is, we plan to develop refined sensitivity analysis techniques based on their use. They should thus determine key variables and/or key phenomena for which it will be necessary to acquire more information. Finally, we also plan to investigate their usefulness in determining optimal commands.

References

[1] A. Antonucci and F. Cuzzolin. Credal sets approximation by lower probabilities: application to credal networks. In *Computational Intelligence for Knowledge-Based Systems Design*, pages 716–725. Springer, 2010.

[2] T. Augustin, F.P.A Coolen, G. de Cooman, and M.C.M. Troffaes. *Introduction to imprecise probabilities.* John Wiley & Sons, 2014.

[3] D. Barba, F. Beolchini, G. Del Re, G. Di Giacomo, and F. Veglió. Kinetic analysis of kluyveromyces lactis fermentation on whey: batch and fed-batch operations. *Process Biochemistry*, 36(6):531 – 536, 2001.

[4] C. Baudrit, I. Couso, and D. Dubois. Joint propagation of probability and possibility in risk analysis: Towards a formal framework. *International Journal of Approximate Reasoning*, 45(1):82–105, 2007.

[5] C. Baudrit, M. Sicard, P.H. Wuillemin, and N. Perrot. Towards a global modelling of the camembert-type cheese ripening process by coupling heterogeneous knowledge with dynamic bayesian networks. *Journal of Food Engineering*, 98(3):283–29, 2010.

[6] C. Baudrit, P.H. Wuillemin, and N. Perrot. Parameter elicitation in probabilistic graphical models for modelling multi-scale food complex systems. *Journal of Food Engineering*, 115(1):1 – 10, 2013.

[7] J.M. Bernard. An introduction to the imprecise Dirichlet model for multinomial data. *International Journal of Approximate Reasoning*, 39(2-3):123–150, June 2005.

[8] J.M. Bernard. The imprecise Dirichlet model. *Int. J. Approx. Reasoning*, 50(2):201–203, 2009.

[9] J. De Bock and G. De Cooman. State sequence prediction in imprecise hidden markov models. In *Proceedings of the seventh International Symposium on Imprecise Probabilities: Theory and Applications*, pages 159–168, 2011.

[10] W. Buntine. A guide to the literature on learning probabilistic networks from data. *Knowledge and Data Engineering, IEEE Transactions on*, 8(2):195–210, 1996.

[11] L.M. De Campos, J.F. Huete, and S. Moral. Probability intervals : a tool for uncertain reasoning. *International Journal of Uncertainty*, 2(2):167 – 196, 1994. Fuzziness and Knowledge-Based Systems.

[12] A. Cano, J. Cano, and S. Moral. Convex sets of probabilities propagation by simulated annealing on a tree of cliques. In *In: Proceedings of Fifth International Conference on Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 1994*, pages 4–8, 1994.

[13] A. Cano, M. Gómez, S. Moral, and J. Abellán. Hill-climbing and branch-and-bound algorithms for exact and approximate inference in credal networks. *Int. J. Approx. Reasoning*, 44(3):261–280, 2007.

[14] A. Cano and S. Moral. A genetic algorithm to approximate convex sets of probabilities. In *Proc. of the Int. Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 859–864, 1996.

[15] A. Cano and S. Moral. Using probability trees to compute marginals with imprecise probabilities. *International Journal of Approximate Reasoning*, 29(1):1 – 46, 2002.

[16] W. G. Chase and H. A. Simon. Perception in chess. *Cognitive Psychology*, 4(1):55 – 81, 1973.

[17] A. Chateauneuf and J.-Y. Jaffray. Some characterizations of lower probabilities and other monotone capacities through the use of möbius inversion. *Mathematical social sciences*, 17(3):263–283, 1989.

[18] I. Couso and D. Dubois. On the Variability of the Concept of Variance for Fuzzy Random Variables. *IEEE Transactions on Fuzzy Systems*, 17(5):1070–1080, 2009.

[19] I. Couso, D. Dubois, and L. Sánchez. *Random Sets and Random Fuzzy Sets as Ill-Perceived Random Variables: An Introduction for Ph. D. Students and Practitioners*. Springer, 2014.

[20] I. Couso, S. Moral, and P. Walley. A survey of concepts of independence for imprecise probabilities. *Risk Decision and Policy*, 5:165–181, 2000.

[21] I. Couso and L. Sanchez. Higher order models for fuzzy random variables. *Fuzzy Sets and Systems*, 159:237–258, 2008.

[22] I. Couso and L. Sánchez. Upper and lower probabilities induced by a fuzzy random variable. *Fuzzy Sets and Systems*, 165(1):1 – 23, 2011. Theme: Fuzzy intervals and applications.

[23] F. Cozman. Credal networks. *Artificial Intelligence*, 120(2):199–233, 2000.

[24] F. Cozman. Graphical models for imprecise probabilities. *International Journal of Approximate Reasoning*, 39(2-3):167–184, June 2005.

[25] F. Cozman. Separation properties of sets of probability measures. *CoRR*, abs/1301.3845, 2013.

[26] J.C.F. da Rocha and F.G. Cozman. Inference with separately specified sets of probabilities in credal networks. In *Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence*, UAI'02, pages 430–437, San Francisco, CA, USA, 2002. Morgan Kaufmann Publishers Inc.

[27] J.C.F. da Rocha, F.G. Cozman, and C.P. de Campos. Inference in polytrees with sets of probabilities. In *Proceedings of the Nineteenth conference on Uncertainty in Artificial Intelligence*, UAI'03, pages 217–224, San Francisco, CA, USA, 2003. Morgan Kaufmann Publishers Inc.

[28] C.P. de Campos and F.G. Cozman. Inference in credal networks using multilinear programming. In *Proceedings of the 2nd Starting AI Researchers Symposium*, pages 50–61, 2004.

[29] C.P. de Campos and F.G. Cozman. Inference in credal networks through integer programming. In *Proceedings of the Fifth International Symposium on Imprecise Probability: Theories and Applications*, pages 145–154, 2007.

[30] A.P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.

[31] S. Destercke, D. Dubois, and E. Chojnacki. Unifying practical uncertainty representations. ii: Clouds. *International Journal of Approximate Reasoning*, 49(3):664–677, 2008.

[32] D. Dubois. Uncertainty theories: a unified view. In *Cybernetic Systems, Dublin (Ireland)*, pages 4–94, http://www.ieee.org/, 2007. IEEE.

[33] D. Dubois, H.T. Nguyen, and H. Prade. Possibility theory, probability and fuzzy sets: misunderstandings, bridges and gaps. . In D. Dubois and H. Prade, editors, *Fundamentals of Fuzzy Sets* , The Handbooks of Fuzzy Sets Series, pages 343–438. Kluwer, Boston, Mass., 2000.

[34] E. Fagiuoli and M. Zaffalon. 2U: an exact interval propagation algorithm for polytrees with binary variables. *Artificial Intelligence*, 106(1):77 – 107, 1998.

[35] S. Ferson, L. Ginzburg, V. Kreinovich, D.M. Myers, and K. Sentz. Constructing probability boxes and dempster-shafer structures. Technical report, Sandia National Laboratories, 2003.

[36] S. Ferson and L. R. Ginzburg. Different methods are needed to propagate ignorance and variability. *Reliability Engineering & System Safety*, 54(2-3):133 – 144, 1996. Treatment of Aleatory and Epistemic Uncertainty.

[37] P.F. Fox. *Cheese - Chemistry, Physics and Microbiology; 3rd ed.* Elsevier, San Diego, CA, 2004.

[38] J.E. Gentle. Monte carlo methods. *Encyclopedia of Statistical Sciences*, 2006.

[39] B. Grünbaum. Convex polytopes. 1967. *Interscience, New York*, 1967.

[40] J.C. Helton and W.L. Oberkampf. Alternative representations of epistemic uncertainty. *Reliability Engineering & System Safety*, 85(1-3):1 – 10, 2004. Alternative Representations of Epistemic Uncertainty.

[41] M. Hourbracq, P.H. Wuillemin C. Baudrit, and S. Destercke. Dynamic credal networks: introduction and use in robustness analysis. In F. Cozman, T. Denœux, S. Destercke, and T. Seidenfeld, editors, *ISIPTA'13:*

Proceedings of the Eighth International Symposium on Imprecise Probability: Theories and Applications, pages 159–168, Compiègne, 2013. SIPTA.

[42] J.S. Ide and F.G. Cozman. Approximate Inference in Credal Networks by Variational Mean Field Methods. In *International Symposium on Imprecise Probabilities and Their Applications*, pages 203–212, 2005.

[43] J.S. Ide and F.G. Cozman. Approximate algorithms for credal networks with binary variables. *International Journal of Approximate Reasoning*, 48(1):275 – 296, 2008. Special Section: Perception Based Data Mining and Decision Support Systems.

[44] J.S. Ide and Cozman F.G. Ipe and l2u: Approximate algorithms for credal networks. In *Proceedings of the second starting AI Researcher Symposium*, pages 118–127. IOS Press, 2004.

[45] S.L. Lauritzen. The em algorithm for graphical association models with missing data. *Computational Statistics & Data Analysis*, 19(2):191–201, 1995.

[46] M.N. Leclercq-Perlat, F. Buono, D. Lambert, E. Latrille, E. Spinnler, and G. Corrieu. Controlled production of camembert-type cheeses. part i: Microbiological and physicochemical evolutions. *Journal of Dairy Research*, 71:346–354, 8 2004.

[47] I. Levi. *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance.* MIT press, 1983.

[48] S. Mariette, C. Baudrit, M.N. Leclerc-Perlat, P.H. Wuillemin, and N. Perrot. Expert knowledge integration to model complex food processes. application on the camembert cheese ripening process. *Expert Syst. Appl.*, 38(9):11804–11812, 2011.

[49] D.D. Mauá, C.P. de Campos, A. Benavoli, and A. Antonucci. Probabilistic inference in credal networks: new complexity results. *Journal of Artificial Intelligence Research*, 50(1):603–637, 2014.

[50] E. Miranda, I. Couso, and P. Gil. Extreme points of credal sets generated by 2-alternating capacities. *International Journal of Approximate Reasoning*, 33(1):95–115, 2003.

[51] K. Murphy. *Dynamic Bayesian Networks: Representation, Inference and Learning.* PhD thesis, UC Berkeley, Computer Science Division, July 2002.

[52] N. Perrot, L. Agioux, I. Ioannou, G. Mauris, G. Corrieu, and G. Trystram. Decision support system design using the operator skill to control

cheese ripening application of the fuzzy symbolic approach. *Journal of Food Engineering*, 64(3):321 – 333, 2004.

[53] N. Perrot, I.C. Trelea, C. Baudrit, G. Trystram, and P. Bourgine. Modelling and analysis of complex food systems: State of the art and new trends. *Trends in Food Science & Technology*, 22(6):304 – 314, 2011.

[54] A. Piatti, A. Antonucci, and M. Zaffalon. Building knowledge-based systems by credal networks: a tutorial. *Advances in Mathematics Research*, 11, 2010.

[55] J.C.F. Rocha and F.G. Cozman. Evidence propagation in credal networks: An exact algorithm based on separately specified sets of probability. In G. Bittencourt and G.L. Ramalho, editors, *Advances in Artificial Intelligence*, volume 2507, pages 376–385. 2002.

[56] G. Schollmeyer. On the number and characterization of the extreme points of the core of necessity measures on finite spaces. In *ISIPTA'15: Proceedings of the Ninth International Symposium on Imprecise Probability: Theories and Applications*, pages 277–286, Pescara, 2015. SIPTA.

[57] G. Shafer. *A mathematical Theory of Evidence.* Princeton University Press, New Jersey, 1976.

[58] L.M.M. Tijskens, L.A.T.M. Hertog, and B.M. Nicolaï. *Food Process Modelling.* Woodhead Publishing Series in Food Science, Technology and Nutrition. Woodhead, 2001.

[59] M. C. M. Troffaes and S. Destercke. Probability boxes on totally preordered spaces for multivariate modelling. *International Journal of Approximate Reasoning*, 52(6):767–791, 2011.

[60] L.V. Utkin. Probabilities of judgments provided by unknown experts by using the imprecise dirichlet model. In *Risk, Decision and Policy, 9(4):391 - 400*, page 400, 2004.

[61] P. Walley. *Statistical reasoning with imprecise probabilities.* Chapman and Hall London, 1991.

[62] P. Walley. Inferences from multinomial data: learning about a bag of marbles. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):3–57, 1996.

[63] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Generalized belief propagation. In *NIPS*, volume 13, pages 689–695, 2000.

[64] L.A. Zadeh. Fuzzy sets. *Information Control*, 8:338–353, 1965.

879 **Vitae**

Cédric Baudrit is a researcher for the French Nationale Institute for Agricultural Research (INRA), in knowledge representation and reasoning. He is interested in the development of mathematical tools capable of (1) integrating fragmented heterogeneous knowledge stemming from different sources; (2) taking into account stochastic and epistemic uncertainty in order to model global complex system. In 2005, he received the Ph.D. degree in computer science from Université Paul Sabatier, in Toulouse (France). In 2002, he holds a master in Applied Mathematics and Computer Science from the University of Orléans, France.

Destercke Sebastien graduated in 2004 as an engineer from the Faculté Polytechnique de Mons in Belgium. In 2008, he earned a Ph.D. degree in computer science from Université Paul Sabatier, in Toulouse (France). He now works as a researcher for the French National Research Center (CNRS), in the Heuristic and Diagnostic of Complex Systems (Heudiasyc) Laboratory. His main research interests are in the field of uncertainty reasoning with imprecision-tolerant models (DS theory, imprecise probabilities, possibility theory, ...), with a focus on issues related to reliability and risk analysis, decision making and machine learning..

Pierre-Henri Wuillemin is Associate Professor in the DESIR (Decision, Intelligent System and operational Research) department of LIP6 (UPMC, CNRS). His research field concerns Graphical Models for uncertainty reasoning and decision making. He is interested both in theoretical aspects - e.g. improving inference algorithms and PGM formulation, structure and parameters learning methods, credal networks - and in applications - e.g. object oriented framework for modeling PGM, sensitivity analysis, troubleshooting, (PO)MDPs.