International Journal of Approximate Reasoning 55 (2014) 1618–1622

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Discussion

Rejoinder on "Imprecise probability models for learning multinomial distributions from data. Applications to learning credal networks"

Andrés R. Masegosa, Serafín Moral *

Dpto. Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada, 18071 Granada, Spain

ARTICLE INFO

Article history:
Received 26 February 2014
Received in revised form 25 April 2014
Accepted 27 April 2014
Available online 10 May 2014

Keywords:
Imprecise probability
Learning
Near-ignorance
Imprecise prior models
Credal networks
Credal classification

ABSTRACT

In this paper we answer to the comments provided by Fabio Cozman, Marco Zaffalon, Giorgio Corani, and Didier Dubois on our paper 'Imprecise Probability Models for Learning Multinomial Distributions from Data. Applications to Learning Credal Networks'. The main topics we have considered are: regularity, the learning principle, the trade-off between prior imprecision and learning, strong symmetry, and the properties of ISSDM for learning graphical conditional independence models.

1. Fabio Cozman

In this section we consider the discussion by Fabio Cozman [1] on our paper [2]. We welcome the comments about *regularity* because this concept underlays some of our results about learning and it was not explicit in the paper. This is an important issue. Under the betting interpretation of upper and lower probabilities, we believe that probabilities should be positive, at least for feasible events in the finite case. Intuitively, there is always an $\epsilon > 0$ that we are ready to pay for any possible event A to gain 1 if A happens to be true. As it was pointed out by Cozman, this poses important technical difficulties in the infinite case. This does not mean that some weaker assumptions of regularity should be discarded. Here we concentrate on assuming it for measurable sets with Lebesgue measure greater than 0. But imprecise probabilities offer some other alternatives as for example to assume that if $\overline{P}(A) > 0$, then we must have $\underline{P}(A) > 0$, but allowing $\overline{P}(A) = \underline{P}(A) = 0$. In this form, it is avoided to assume regularity for any logically possible event. In our case, we have that $\overline{P}(A) > 0$, $\underline{P}(A) > 0$ for any measurable event with $|A| > 0$. If $|A| = 0$, we have that the upper and lower probabilities of A are equal to 0. With the IDM we have many measurable events for which $\overline{P}(A) > 0$ and $\underline{P}(A) = 0$. For example, in the case of intervals included in $[0, 1]$, $\underline{P}([a, b]) = 0$, for all the intervals except for the full $[0, 1]$ interval. However, there are some other measurable sets with positive lower probability (in other case we could not obtain meaningful inferences). For example, for any $\epsilon > 0$ the lower probability of $[0, \epsilon] \cup [1 - \epsilon, 1]$ is always positive. This is somewhat shocking. We cannot forget that upper and lower

DOI of original article: http://dx.doi.org/10.1016/j.ijar.2013.09.019.
DOIs of comments: http://dx.doi.org/10.1016/j.ijar.2014.04.016, http://dx.doi.org/10.1016/j.ijar.2014.05.001, http://dx.doi.org/10.1016/j.ijar.2014.04.002.
* Corresponding author.
E-mail addresses: andrew@decsai.ugr.es (A.R. Masegosa), smc@decsai.ugr.es (S. Moral).

http://dx.doi.org/10.1016/j.ijar.2014.04.017
0888-613X/

probabilities have a behavioral interpretation, and that we are ready to buy event $[0,\epsilon] \cup [1-\epsilon, 1]$ for a positive amount for any ϵ, but we are not ready to pay anything for the interval $[\epsilon, 1-\epsilon]$. The IDM is assuming that we are ready to bet some amount for the chances being in $[0, 0.0001] \cup [0.9999, 1]$, but nothing for the chances being in $[0.0001, 0.9999]$. In the ISSDM the last event has a greater lower probability than the former one (except in some extreme cases in which the lower value of the equivalent sample size s_1 is too small).

With respect to the learning principle, we believe that it is important by itself. It is a basic requirement because situations such as the one in Example 4 in our paper [2] look counterintuitive and should be avoided. This principle is incompatible with representation invariance (RIP) but we find it more basic than RIP. Representation invariance looks appealing but refining and coarsening the categories is a non-trivial fact. In imprecise probability we very often neglect this and even we say that in an experiment the number of categories is unknown. But if we go to the bag of marbles example by Walley [3], the color of the balls is a continuum and to define the problem we have to determine the set of categories and the measurement procedure, i.e. how the observed frequencies of colors are going to be mapped in our set of categories. In any experimental setting there are prior assumptions and, in our case, the selection of categories is part of these assumptions. If according to a gamble, we are going to be paid an amount if a ball is red, we have to fix in advance the procedure which defines when a ball is classified as red but not brown, for example. It is possible that we have some set of categories that are not observed yet, but we should know which the categories are.

We appreciate the comment about connecting our proposal with existing literature on strict coherence. We believe that it is a good suggestion and, as it was said earlier, imprecise probability could throw new light on this issue, as it allows weaker formalizations of this idea.

With respect to the comment on dilation, we want to point out that being true that under the ISSDM the degree of imprecision can increase after receiving some information, it never happens that the interval of conditional probabilities strictly contains the interval of prior probabilities. For a value w_1 we can have a prior precise probability. After conditioning to observations, the conditional probabilities can become an imprecise interval $[a, b]$ but, this interval never contains the precise value. So, we do not have dilation in a strict sense [4]. When observations are uniform in the different categories we go back to precision, and imprecision only increases when data are unbalanced and relative frequencies are extreme (close to 0 or 1). It is in this case when imprecision appears in the ISSDM. We find this behavior reasonable as the extreme probabilities are the most risky ones: if our upper probability of an event is very close to 0, then we can take risky bets against this event.

In relation with the generalized version of the ISSDM and the necessity of specifying more parameters, we have shown that there are important problems in selecting the equivalent sample size in a Bayesian approach and in determining how this sample size should change in relation with the number of elements on W and the number of conditional distributions. There is not a clear way of doing it and we have shown examples, in which several options make sense. The generalized ISSDM tries to minimize the number of prior assumptions, by allowing different assignment functions to determine the equivalent sample sizes of prior probabilities for the different conditional distributions. In fact, the generalized ISSDM allows more possibilities than the restricted case. It usually happens that for being more imprecise we need more parameters. For example, we need two values to specify an interval probability against one value for a precise probability. But this is not necessarily a problem.

We consider the learning part as particularly important for the ISSDM. In some broad sense, we believe that this is a model for near-ignorance in relation with the dependence structure of a credal network. As it was pointed out in the paper, with low s values we are favoring independence and with high values of s dependence, so by allowing an interval for the s value, we are trying to be ignorant about this fact. It is similar to what IDM does for the parameters: α_i can be very low favoring low chances or high favoring high chances. Observed data overcome this ignorance and we can make more precise decisions when the number of observations increases.

There is a part in which we have to make many compromises in order to obtain a full procedure for learning generalized credal networks. This has been motivated by the idea of obtaining feasible computer implementations of the proposed methods and being able to show their behavior in some experiments. This part is not essential in itself. We have included it to show that this is not only a theory but something which can provide useful practical results. However, we agree with Fabio Cozman on the fact that moving from optimization to sampling for approximate computations is an idea that has some potential and deserves some attention for future work.

We thank Fabio Cozman for his comments in the generalized credal networks with ambiguous edges. We also believe that they are more intuitive when working with experts, and can also be a basis for interactive procedures for learning credal networks in which we can ask to experts about their beliefs about unsure edges in the line of [5].

Finally, we agree that evaluation of credal networks is not a simple issue, but here we have followed existing procedures in the literature as this is only an auxiliary topic in our paper.

2. Marco Zaffalon and Giorgio Corani

In this section we consider the comments by Marco Zaffalon and Giorgio Corani [6] on our paper [2]. We agree with Marco Zaffalon and with Giorgio Corani in that near-ignorance models have an important role to play in learning with no prior information. The only point is that there can be alternative models. As they say in their comments, there is a trade-off between learning and prior assumptions. If prior assumptions are weaker then the learning capabilities are lower. Models