A Hierarchical Approach to Learning with Imprecise Probabilities*

Serafín Moral SMC@DECSAI.UGR.ES
Dpto. Ciencias de la Computación e Inteligencia Artificial, University of Granada, Spain

The results of the paper "Learning with Imprecise Probabilities as Model Selection and Averaging" [5] will be discussed. That paper proposes a hierarchical approach to learning. It considers that learning is equivalent to selecting the parameters for a given model, understanding that parameters can also represent the structure of a model. So, that learning implies model selection.

The procedure follows the basic scheme of Gärdenfors and Shalin [3]. The parameters are classified in two sets: Θ and B. Θ is the top set and conditioned to each $\theta \in \Theta$ we have a precise Bayesian model for the parameters in B and the variables of interest, in order to follow a model averaging approach (by averaging with respect to the posterior probability in B given the observations). At the same time, a set of observations defines a likelihood in the top set Θ. Different methods of using this likelihood have been proposed in the literature [1, 3], but in general they do not provide a sound and well-founded integration model. The idea in this paper is essentially similar to the α-cut conditioning by Cattaneo [1], but it gives a justification based on decision making with imprecise probabilities. The final procedure proposed here is quite flexible and can accommodate very different available procedures, such as maximum likelihood, likelihood intervals, Bayesian high-density regions, imprecise probability methods such as the imprecise Dirichlet model, and others. It can also serve as a basis for proposing new imprecise probabilities methods for learning generalized credal networks.

An important fact about the procedure is that the prior information on Θ is given by a coherent set of desirable gambles [2, 8, 6]. This setting is what allows us to present our uniform model, which is essentially the same than the classical uniform model for finite Θ, but quite different from the uniform density in the infinite case. The use of desirable gambles is essential, since there is no equivalent representation as a set of probability measures or credal set. The final model is based on the discounting of this uniform information on Θ. The proposed discounting is a generalization of the concept of discounting a belief function [7] or the ε-contaminated robust models [4], but here a behavioural interpretation based on sets of desirable gambles is provided.

References

[1] Marco EGV Cattaneo. A continuous updating rule for imprecise probabilities. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 426–435. Springer, 2014.

[2] Inés Couso and Serafín Moral. Sets of desirable gambles: conditioning, representation, and precise probabilities. *International Journal of Approximate Reasoning*, 52(7):1034–1055, 2011.

[3] Peter Gärdenfors and Nils-Eric Sahlin. Unreliable probabilities, risk taking, and decision making. *Synthese*, 53(3): 361–386, 1982.

[4] Peter J. Huber. *Robust Statistics*. Wiley, New York, 1981.

[5] Serafín Moral. Learning with imprecise probabilities as model selection and averaging. *International Journal of Approximate Reasoning*, 109:111–124, 2019.

[6] Erik Quaeghebeur. Desirability. In *Introduction to Imprecise Probabilities*, chapter 1, pages 1–27. John Wiley & Sons, 2014. ISBN 9781118763117. doi:10.1002/9781118763117.ch1.

[7] Glenn Shafer. *A Mathematical Theory of Evidence*, volume 42. Princeton University Press, 1976.

[8] Peter Walley. Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning*, 24: 125–148, 2000.

* *Acknowledgments* This research was supported by the Spanish Ministry of Economy and Competitiveness under project TIN2016-77902-C3-2-P, and the European Regional Development Fund (FEDER).