International Journal of Approximate Reasoning 67 (2015) 1–36

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Conditioning, updating and lower probability zero

Jasper De Bock *, Gert de Cooman

Ghent University, SYSTeMS Research Group, Belgium

A R T I C L E   I N F O

Article history:
Received 10 December 2014
Received in revised form 18 August 2015
Accepted 24 August 2015
Available online 1 September 2015

Keywords:
Conditioning
Updating
Probability zero
Regular extension
Natural extension
Sets of desirable gambles

A B S T R A C T

We discuss the issue of conditioning on events with probability zero within an imprecise-probabilistic setting, where it may happen that the conditioning event has lower probability zero, but positive upper probability. In this situation, two different conditioning rules are commonly used: regular extension and natural extension. We explain the difference between them and discuss various technical and computational aspects. Both conditioning rules are often used to update an imprecise belief model after receiving the information that some event O has occurred, simply by conditioning on O, but often little argumentation is given as to why such an approach would make sense. We help to address this problem by providing a firm foundational justification for the use of natural and regular extension as updating rules. Our results are presented in three different, closely related frameworks: sets of desirable gambles, lower previsions, and sets of probabilities. What makes our justification especially powerful is that it avoids making some of the unnecessary strong assumptions that are traditionally adopted. For example, we do not assume that lower and upper probabilities provide bounds on some 'true' probability mass function, on which we can then simply apply Bayes's rule. Instead a subject's lower probability for an event O is taken to be the supremum betting rate at which he is willing to bet on O, and his upper probability is the infimum betting rate at which he is willing to take bets on O; we do not assume the existence of a fair betting rate that lies in between these bounds.

1. Introduction

Conditioning on events with probability zero is traditionally considered to be problematic because Bayes's rule cannot be applied. In those cases, depending on the approach that is taken, the conditional probability measure is either left undefined or chosen freely without enforcing any connection with the unconditional measure. The latter approach has the advantage of being more flexible, but it does not really solve the problem, because it provides no guidelines on how to come up with such a conditional model. In order to avoid all issues with countable additivity and measurability—which make the problem even more complicated—we restrict attention to finite state spaces, thereby allowing us to work with probability mass functions. In that case, the most common solution to this problem is to simply ignore it, because from a practical point of view, in a finite state space, events with probability zero are usually considered to be impossible anyway, which makes the task of conditioning on them rather irrelevant.

The situation becomes more complex when we consider a set of probability mass functions instead of a single one, because then, the lower and upper probability of an event need not coincide, and it may have lower probability zero, but

* Corresponding author.
E-mail addresses: Jasper.DeBock@UGent.be (J. De Bock), Gert.deCooman@UGent.be (G. de Cooman).

http://dx.doi.org/10.1016/j.ijar.2015.08.006

positive upper probability. This happens frequently in practice. For example, many common methods for deriving sets of probabilities from data will assign lower probability zero to an event for which there is no data—because it might well be impossible—but not upper probability zero—because the fact that you have not seen it yet does not imply that it is impossible. Clearly, such events cannot be ignored, and we need to be able to condition on them.

Fortunately, working with sets of probability mass functions is already part of the solution, because it allows us to condition on an event O that has probability zero in a trivial manner, simply by taking the conditional model to be the set of all probability mass functions on O; this is called the vacuous model on O. Since Bayes's rule does not impose any constraints, this vacuous model represents all we know about the conditional probability mass function on O.

If we start from a set $\mathscr{M}$ of mass functions, then three cases can be distinguished. If every mass function in $\mathscr{M}$ assigns probability zero to O, we use the vacuous model on O as our conditional model. If every element of $\mathscr{M}$ assigns positive probability to O, the conditional set of mass functions is obtained by applying Bayes's rule to every element of $\mathscr{M}$. The remaining possibility is the one that we have already mentioned: the case that O has probability zero according to some elements of $\mathscr{M}$—lower probability zero—but positive probability according to others—positive upper probability. Here, imprecise-probabilistic frameworks—such as, but as we shall see not limited to, sets of probability mass functions—usually distinguish between two main conditioning rules: natural and regular extension. Natural extension again uses the vacuous model on O, consisting of all probability mass functions on O. Regular extension ignores the probability mass functions in $\mathscr{M}$ that assign probability zero to O and simply applies Bayes's rule to the others.

Example 1. Consider a ternary state space $\Omega = \{a, b, c\}$ and let f be the real-valued map on Ω that is defined by $f(a) := 1$, $f(b) := 2$ and $f(c) := 3$. Let $\mathscr{M}$ be the set of all mass functions p on Ω for which the corresponding expectation $P_p(f) := \sum_{\omega\in\Omega} f(\omega)p(\omega)$ of f—also called the prevision of f—is higher than or equal to 2 or, equivalently, the largest set of mass functions on Ω such that the lower expectation

$$\underline{P}(f) := \inf_{p\in\mathscr{M}} P_p(f)$$

of f—also called the lower prevision of f—is equal to two. Since the unitary constraint for mass functions implies that $p(a) + 2p(b) + 3p(c) \geq 2 \Leftrightarrow p(c) \geq p(a)$, $\mathscr{M}$ is simply the set of all mass functions on Ω for which the probability of c is at least as high as that of a.

Now let $O = \{a, c\}$ be the event that b does not happen. According to $\mathscr{M}$, the lower and upper probability of O are equal to 0 and 1, respectively. The lower probability 0 is attained by a single degenerate probability mass function p_b that assigns all mass to b. If we use natural extension to condition the set $\mathscr{M}$ on the event O, then since O has lower probability zero, we obtain the set of all mass functions on O—the vacuous model on O. Regular extension, on the other hand, corresponds to applying Bayes's rule to every mass function in $\mathscr{M}$ except p_b, and results in a conditional model that is significantly smaller, as it only contains those conditional probability mass functions on O for which $p(c|O) \geq p(a|O)$. Hence, the conditional lower probability of c that is provided by regular extension is equal to $1/2$, whereas it is 0 according to natural extension. ◊

As this example shows, regular extension can be a lot more informative than natural extension, in the sense that the resulting conditional set of mass functions is smaller—twice as small in this particular case. One can even come up with examples where regular extension results in a single conditional mass function and natural extension still returns the vacuous model [38, Appendix J2].

For this reason, from a practical point of view, regular extension is more useful than natural extension. Conditioning is not merely a technical concept, but also a popular updating tool. After being informed that an event O has occurred, we wish to update our former model to obtain a new one, and the most popular approach for doing so is to condition the old model on the event O. For a single mass function, the use of Bayes's rule as an updating tool has been justified by many authors—see for example Refs. [29,28,30]—and this justification easily translates to sets of probabilities, by applying it elementwise. After updating, the updated set of mass functions takes the role of the original one, and is then used for inference, classification or decision making. Clearly, in order for this new model to be of practical use, it should be as informative as possible, and definitely not vacuous, which is why—as far as updating is concerned—regular extension is usually preferred over natural extension.

So far, so good. However, imprecision in probability theory can be represented by more models than just sets of probabilities. Consider for example the vacuous model, which is taken to model complete ignorance. In the framework of sets of probability mass functions, this corresponds to the use of the set of all possible mass functions. However, this is only a limited form of 'complete' ignorance, because it assumes the existence of such a mass function; we are only ignorant about which mass function it should be. Walley refers to this assumption as the dogma of ideal precision [38, Section 1.1.5]; we will simply call it ideal precision. It might seem like a trivial assumption, but it is in fact not. In order to understand why, we need to go back to the meaning of probability. Frequentists would argue that it is a limiting frequency. Indeed, in some cases, it is. But why should this be the case in general? And why should the frequencies converge? In any case, for a single experiment that is not repeated, this interpretation no longer works. People that favour the subjective approach will argue that probabilities are a subject's fair betting rates. However, from a subjective point of view, it seems hardly compelling that a subject should have such a fair betting rate for every event.