International Journal of Approximate Reasoning 74 (2016) 30–57

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Robust probability updating ☆

Thijs van Ommen [a,*], Wouter M. Koolen [b], Thijs E. Feenstra [c], Peter D. Grünwald [b,c]

[a] Universiteit van Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands
[b] Centrum Wiskunde & Informatica, Science Park 123, 1098 XG Amsterdam, The Netherlands
[c] Universiteit Leiden, Niels Bohrweg 1, 2333 CA Leiden, The Netherlands

A R T I C L E   I N F O

Article history:
Received 30 November 2015
Received in revised form 4 March 2016
Accepted 9 March 2016
Available online 15 March 2016

Keywords:
Probability updating
Maximum entropy
Loss functions
Minimax decision making

A B S T R A C T

This paper discusses an alternative to conditioning that may be used when the probability distribution is not fully specified. It does not require any assumptions (such as CAR: coarsening at random) on the unknown distribution. The well-known Monty Hall problem is the simplest scenario where neither naive conditioning nor the CAR assumption suffice to determine an updated probability distribution. This paper thus addresses a generalization of that problem to arbitrary distributions on finite outcome spaces, arbitrary sets of 'messages', and (almost) arbitrary loss functions, and provides existence and characterization theorems for robust probability updating strategies. We find that for logarithmic loss, optimality is characterized by an elegant condition, which we call *RCAR (reverse coarsening at random)*. Under certain conditions, the same condition also characterizes optimality for a much larger class of loss functions, and we obtain an objective and general answer to how one should update probabilities in the light of new information.

1. Introduction

There are many situations in which a decision maker receives incomplete data and still has to reach conclusions about these data. One type of incomplete data is *coarse* data: instead of the real outcome of a random event, the decision maker observes a subset of the possible outcomes, and knows only that the actual outcome is an element of this subset. An example frequently occurs in questionnaires, where people may be asked if their date of birth lies between 1950 and 1960 or between 1960 and 1970 et cetera. Their exact year of birth is unknown to us, but at least we now know for sure in which decade they are born. We introduce a simple and concrete motivating instance of coarse data with the following example.

Example A *(Fair die).* Suppose I throw a fair die. I get to see the result of the throw, but you do not. Now I tell you that the result lies in the set $\{1, 2, 3, 4\}$. This is an example of coarse data. You know that I used a fair die and that what I tell you is true. Now you are asked to give the probability that I rolled a 3. Likely, you would say that the probability of each of the remaining possible results is $1/4$. This is the knee-jerk reaction of someone who studied probability theory, since this is standard *conditioning*. But is this always correct?

☆ This work is adapted from dissertation [1, Chapters 6 and 7], which extends M.Sc. thesis [2].

* Corresponding author.
E-mail addresses: T.vanOmmen@uva.nl (T. van Ommen), wmkoolen@cwi.nl (W.M. Koolen), pdg@cwi.nl (P.D. Grünwald).

http://dx.doi.org/10.1016/j.ijar.2016.03.001
0888-613X/

Suppose that there is only one alternative set of results I could give you after rolling the die, namely the set $\{3, 4, 5, 6\}$. I can now follow a *coarsening mechanism*: a procedure that tells me which subset to reveal given a particular result of the die roll. If the outcome is 1, 2, 5, or 6, there is nothing for me to choose. Suppose that if the outcome is 3 or 4, the coarsening mechanism I use selects set $\{1, 2, 3, 4\}$ or set $\{3, 4, 5, 6\}$ at random, each with probability $1/2$. If I throw the die 6000 times, I expect to see the outcome 3 a thousand times. Therefore I expect to report the set $\{1, 2, 3, 4\}$ five hundred times after I see the outcome 3. It is clear that I expect to report the set $\{1, 2, 3, 4\}$ 3000 times in total. So for die rolls where I told you $\{1, 2, 3, 4\}$, the probability of the true outcome being 3 is actually $1/6$ with this coarsening mechanism. We see that the prediction of $1/4$ from the first paragraph was not correct, in the sense that the probabilities computed there do not correspond to the long-run relative frequencies. We conclude that the knee-jerk reaction is not always correct.

In Example A we have seen that standard conditioning does not always give the correct answers. Heitjan and Rubin [3] answer the question under what circumstances standard conditioning of coarse data is correct. They discovered a necessary and sufficient condition of the coarsening mechanism, called *coarsening at random (CAR)*. A coarsening mechanism satisfies the CAR condition if, for each subset y of the outcomes, the probability of choosing to report y is the same no matter which outcome $x \in y$ is the true outcome. It depends on the arrangement of possible revealed subsets whether a coarsening mechanism exists that satisfies CAR. It holds automatically if the subsets that can be revealed partition the sample space. As noted by Grünwald and Halpern [4] however, as soon as events overlap, there exist distributions on the space for which CAR does not hold. In many such situations it even cannot hold; see Gill and Grünwald [5] for a complete characterization of the — quite restricted — set of situations in which CAR can hold. No coarsening mechanism satisfies the CAR condition for Example A.

We hasten to add that we neither question the validity of conditioning nor do we want to replace it by something else. The real problem lies not with conditioning, but with conditioning within the wrong sample space, in which the coarsening mechanism cannot be represented. If we had a distribution P on the correct, larger space, which allows for statements like 'the probability is α that I choose $\{1, 2, 3, 4\}$ to reveal if the outcome is 3', then conditioning would give the correct results. The problem with coarse data is though that we often do not have enough information to identify P — e.g. we do not know the value of α and do not want to assume that it is $1/2$. Henceforth, we shall refer to conditioning in the overly simple space as 'naive conditioning'. In this paper we propose update rules for situations in which naive conditioning gives the wrong answer, and conditioning in the right space is problematic because the underlying distribution is partially unknown. These are invariably situations in which two or more of the potentially observed events overlap.

We illustrate this further with a famously counter-intuitive example: the Monty Hall puzzle, posed by Selvin [6] and popularized years later in Ask Marilyn, a weekly column in Parade Magazine by Marilyn vos Savant [7].

Example B *(Monty Hall).* Suppose you are on a game show and you may choose one of three doors. Behind one of the doors a car can be found, but the other two only hide a goat. Initially the car is equally likely to be behind each of the doors. After you have picked one of the doors, the host Monty Hall, who knows the location of the prize, will open one of the other doors, revealing a goat. Now you are asked if you would like to switch from the door you chose to the other unopened door. Is it a good idea to switch?

At this moment we will not answer this question, but we show that the problem of choosing whether to switch doors is an example of the coarse data problem. The unknown random value we are interested in is the location of the car: one of the three doors. When the host opens a door different from the one you picked, revealing a goat, this is equivalent to reporting a subset. The subset he reports is the set of the two doors that are still closed. For example, if he opens door 2, this tells us that the true value, the location of the car, is in the subset $\{1, 3\}$. Note that if you have by chance picked the correct door, there are two possible doors Monty Hall can open, so also two subsets he can report. This implies that Monty has a choice in reporting a subset. How does Monty's coarsening mechanism influence your prediction of the true location of the car?

The CAR condition can only be satisfied for very particular distributions of where the prize is: the probability that the prize is hidden behind the initially chosen door must be either 0 or 1, otherwise no CAR coarsening mechanism exists [4, Example 3.3].[1] If the prize is hidden in any other way, for example uniformly at random as we assume, then CAR cannot hold, and naive conditioning will result in an incorrect conclusion for at least one of the two subsets.

Examples A and B are just two instances of a more general problem: the number of outcomes may be arbitrary; the initial distribution of the true outcome may be any distribution; and the subsets of outcomes that may be reported to the decision maker may be any family of sets. Our goal is to define general procedures that tell us how to update the probabilities of the outcomes after making a coarse observation, in such situations where naive conditioning is not adequate. We are aiming for modular methods that do not enforce a particular interpretation of probability. In Example A, we saw "objective" probabilities: the original distributions were known, and the updated probabilities we found could again be interpreted as frequencies over many repetitions of the same experiment. The original distribution of the outcomes could however also

[1] This uses the *weak* version of CAR in the terminology of Jaeger [8], in which outcomes with probability 0 are exempt from the equality constraint. A *strong* CAR coarsening mechanism does not exist regardless of the probabilities with which the prize is hidden.