Reviewing the Distinction between Uncertainty and Ignorance

Vedang Joshi
vedang.joshi.2018@bristol.ac.uk

EMAT30015 Uncertainty Modelling for Intelligent Systems 3

Department of Engineering Mathematics, University of Bristol

9th January 2020

1. Introduction

Uncertainty refers generally to situations with imperfect or lack of information. Uncertainty comes in many different types and we use many different interpretations of uncertainty for various situations. Uncertainty itself may be divided into two types, aleatoric and epistemic. Aleatoric uncertainty refers to uncertainty about a variable phenomenon while epistemic uncertainty refers to uncertainty stemming from a lack of knowledge [14]. Here the aleatoric interpretation of uncertainty may be considered to be the basis for probability theory whereas the epistemic interpretation of uncertainty is known as ignorance. According to Giang [5], these two 'extreme' interpretations form the basis on which all the other epistemic interpretations of uncertainty are 'spanned'. Ignorance exists when information is scarce, contradictory and unreliable. In risk assessments, ignorance can refer to situations with a high level of uncertainty on the likelihood of events and their consequences [1]. Ignorance may also refer to when an agent has difficulty in determining the set of alternate states [11].

Suppose a patient has a prescription of a particular drug, but the label on the bottle (with the information of when the dose should be taken, with what frequency is given) is missing. The patient goes to three experts for some advice on the dose and frequency required and they get different opinions from each expert (see Appendix A). Taking the response from expert A, the statement is very precise, with little uncertainty. The quantity of the drug to take is 100 mg. The patient must take it twice a day. The uncertainty here lies at the first dose i.e. when to take it, but after that, every twelve hours the patient should take another dose. The uncertainty here is part the execution, not the prescription. Examining the response from expert B, it is very imprecise/uncertain. It is unclear whether 100 mg is a minimal dose and the natural question is whether this dose will be effective. The optimum dose is unclear as well as the maximum dose that the patient takes. Examining the response from expert C, here the alternatives are inconsistent with each other. The patient may think that it is 100mg twice a day or 200mg once a day, and that the expert misspoke. The patient can make out exactly what the instructions say to do, yet they are contradictory. While expert A gives a precise opinion without (much) uncertainty, experts B and C give opinions fraught with uncertainty. Yet, the two opinions are nothing alike; expert B gives some indication as to the quantity of the drug and the frequency of ingesting it required yet drawing meaningful insights from this data is very difficult, as explained above. This is an example of aleatoric uncertainty. Expert C, gives contradictory alternatives on what quantity and when to take the drug. This is an excellent example of ignorance, as the patient has no idea which of the alternatives to pick, and inherently the choices are inconsistent.

In this paper, we will review the differences between uncertainty and ignorance. We will look at traditional Bayesianism and how it lacks in representing ignorance, while taking into account Laplace's principle of insufficient reasoning. We will finally go on to look at probability measures that are successful in distinguishing between ignorance and uncertainties.

2. Ignorance and Bayesianism

If we roll a fair, cubic die and you ask me what the probability of a 3 coming up is, the answer is '$\frac{1}{6}$'. Now assume we wish to assign a probability to the claim 'Roald Dahl's children's book *The BFG* has exactly 17122 words' (adapted from [2]). We would not be easily able to assign a probability to this claim, with our knowledge of Dahl's works allowing us to say that the value falls within some interval and that all values within this interval are equally plausible. But we would not be able to assign a probability to the claim of there being *exactly* 17122 words in 'The BFG'. The simplest version of Bayesianism cannot handle this scenario. In the case of the die, the principle of insufficient reasoning allows us to use a uniform probability distribution over some interval of values and assign the probability from that distribution; if there are n equally probable values in the plausible range of number of words for 'The BFG', each value will be assigned a probability of $\frac{1}{n}$.

Laplace's principle of insufficient reasoning (PIR) stems from a lack of information about the initial degrees of beliefs about any states of a rational agent i.e. when the agent initially has no relevant evidence about the world, apart from the Bayesian fact that these degrees of belief should be probabilistic. Given the PIR exists, the agent now has inherent instructions on how to revise their degrees of beliefs when confronted with new evidence and also when no evidence is provided. PIR is an extremely intuitive concept independently justified in the literature, with considerations ranging from epistemic utility [10] to risk aversion [20] to evidential support [19]. Yet before the acceptance of any principle, it has become customary to besiege the principle with paradoxes, the most notable of which was put forth by Keynes and has since been reformulated in the literature. Keynes [7] asks after the country of a man whose origins are known to be one of France, Britain or Ireland. By PIR, we assign a probability of 1/3 to each state. We may disjunctively condense the same probability space by creating a disjunctive outcome 'British Isles', where British Isles = Britain $\cup$ Ireland. Again by PIR, we assign the probabilities 1/2 to each of France and British Isles. We have now assigned two incompatible probabilities 1/3 and 1/2 to the state 'France', and arrive at a contradiction. We may argue that PIR cannot be used to specify probabilities in cases of extensive ignorance, thereby providing constraints for when PIR may be used, and addressing such paradoxes rigorously.

Coming back to our BFG problem, according to Bayesianism, knowing the exact length of BFG is the same as assigning a low, yet non-zero probability to our claim. Suppose we were also discussing whether the Disney movie 'The BFG' is a good adaptation of the book and we claimed that Dahl himself authored the screenplay to the film, we would not use the PIR, but would assign the claim a very low probability. Bayesianism does not distinguish between these two claims; it views both as a confirmation of a low-probability claim. We may be surprised by the confirmation of a low-probability claim but not by the confirmation of a claim where we assign each possible state by the same degree of belief. In the absence of a neutral degree of belief, or ignorance, we must consider propositions which have a large number of possible states each with the same degree of belief, as equally improbable. Thus any new evidence and hypothesis which supports a particular state, more so than originally provided by the uniform probability distribution, will then be said to be confirmed.

John Norton [9] criticised these probabilistic notions of ignorance and objected to the claims that there can be a probability distribution over possible values of some variables when such values are cannot fully be determined by existing evidence. Norton also objected to the claims that the observed values have low probability rather than a neutral probability (ignorance). Norton believes that this inability to distinguish between a neutral degree of belief (ignorance) and low probability values is a crucial flaw in Bayesianism which itself stems from the Bayesian principle of additivity. Norton believes that a radically different framework can represent Bayesianism and Dempster-Shafer theory is one such framework that represents both ignorance and probabilistic uncertainty measures.

3. Dempster-Shafer Theory and Imprecise Probabilities

The Dempster-Shafer theory (DST) allows us not only to calculate probabilities for a certain event (as in Bayesianism) but also allows us to base degrees of belief for all the probability in an event [13]. An agent uses evidence that has bearing not only on a single state within the state of possible worlds but on a larger subset of states withing the set of all possible worlds. Dempster's rule assumes that the events for which we have probabilities are independent with respect to our prior probability judgements, but this independence vanishes when there is conflict between contradictory pieces of evidence. Because of the generality of DST, it avoids the Bayesian restriction that committing a belief measure to a possible state implies committing the remaining belief to its negation [6]. Fundamentally, this discerns DST from orthodox Bayesianism. DST quantifies uncertainty by using two measures of uncertainty, akin to the Bayesian theory of subjective probability, a belief measure and a plausibility measure. Let W be the set of all possible worlds. Then the power set, 2^{W}, is the set of all subsets of W. Assigning a belief mass, to all elements of the power set we obtain m: $2^{\mathrm{W}} \rightarrow [0,1]$ where $m(\phi) = 0$ and $\sum_{B \subseteq W} m(B) = 1$. From these mass assignments, DST introduces two bounds within which the probabilities can be calculated. The lower bound, *belief Bel*: $2^{\mathrm{W}} \rightarrow [0,1]$ quantifies the degree of belief assigned confirming the proposition and the upper bound *plausibility Pl*: $2^{\mathrm{W}} \rightarrow [0,1]$ quantifies the degree of belief assigned not dis-confirming the proposition. The two measures are directly related as shown in the example of a proposition, Θ. Here $\Theta \subseteq W$, $Bel(\Theta) \leq Pl(\Theta)$, $Pl(\Theta) = 1 - Bel(\Theta^c)$ and $Pl(\Theta) - Bel(\Theta)$ gives a measure of the ignorance in proposition Θ. Traditional probability measures are special cases of DST; when $Bel(\Theta) = Pl(\Theta) = \mathrm{P}(\Theta)$. DST also does not satisfy additivity which satisfies one of Norton's rules for a framework quantifying ignorance.

Imprecise probabilities are also generalisations of probability theory. Instead of a precise value p as the quantification of a probability measure of an event Θ, $P(\Theta) = p$, imprecise probability gives a set of probabilities bounded by the values $p1$ and $p2$ where $p1 \leq p \leq p2$. This succeeds in quantifying the aleatoric and epistemic uncertainties of the proposition in question. Imprecise probabilities may be considered an umbrella term for DST and other representations of probability theory incorporating the idea of distinguishing ignorance from low probability values. We give other such examples forthwith. Probability bound analysis evaluates uncertainty with pairs of lower and upper distribution functions, called p-boxes [4]. F-probability includes intervals into probability values which follow the Kolmogorov axioms [18]. The fuzzy probability takes into account probability distributions with fuzzy variables [8]. More recently, a new form of imprecise probability based on the generalised interval was proposed (A generalised interval being different to a traditional interval ([X] = $x \in R \mid a \leq x \leq b$) in the fact that the interval of [X] = [a,b] i.e a pair of values instead of a set of values) [17]. Possibility theory represents uncertainty and ignorance with necessity-possibility pairs analogous to belief-plausibility pairs in DST [3]. Let us formalise this notion of imprecise probabilities; following on from our earlier notation, let P(K) now be the set of probability distributions from an agent's prior knowledge of event, K. Let $P(K) \subseteq P$ be a set of probability measures on 2^{W}. The lower probability measure can then be defined as P_L: $2^{\mathrm{W}} \rightarrow [0,1]$. The upper probability measure can then be defined as P_U: $2^{\mathrm{W}} \rightarrow [0,1]$. Here, for some proposition Θ, $\mathrm{P_L}(\Theta) = \min\{\mathrm{P}(\Theta) : \mathrm{P} \in \mathrm{P(K)}\}$ and $\mathrm{P_U}(\Theta) = \max\{\mathrm{P}(\Theta) : \mathrm{P} \in \mathrm{P(K)}\}$. Similar to DST, we get $\mathrm{P_U}(\Theta) = 1 - \mathrm{P_L}(\Theta^c)$. We may consider the $Pl(\Theta)$-$\mathrm{P_U}(\Theta)$ pair and the $Bel(\Theta)$-$\mathrm{P_L}(\Theta)$ pair as analogous measures across the two theories. This formulation may also be generalised across possibility theory [21], another candidate for imprecise probabilities which uses possibility and necessity measures to quantify the uncertainty in events. It is clear however that necessity measures are very special cases of belief functions, however not all belief functions are necessity measures; examples being probability measures and epsilon-contamination models [16].

Walley [15] observes that there is no reason that the precise assessment $\mathrm{P_L}(\Theta) = \mathrm{P_U}(\Theta)$ may be made unlike the assessment for traditional probability measures when $Bel(\Theta) = Pl(\Theta) = \mathrm{P}(\Theta)$. Walley also observes that there is no justification for a Bayesian sensitivity analysis interpretation (SAI) of $\mathrm{P_L}(\Theta)$ and $\mathrm{P_U}(\Theta)$, which views them as lower and upper bounds for some unknown 'underlying linear prevision', $\mathrm{P}(\Theta)$. One of the most fundamental contributions of DST is that it emphasises that belief functions should not be given a SAI either. We must not confuse belief functions being represented by a lower envelope of a set of probability measures as the lower bound for some unknown probability measure [12]. In this respect, DST and other theories of imprecise probabilities are similar.

4. Conclusion

Using a sequence of arguments based on examples, we have shown that there is indeed a significant difference between uncertainty and ignorance, the most stark of which is the total lack of knowledge or information regarding ignorance, which is not the case in uncertainties. We go on to show how the use of PIR may lead to paradoxes, but why it is still intuitive with lack of information, explaining its wide use in Bayesianism. We explain Bayesianism's inability to distinguish between low probability claims and ignorance, motivating the need for frameworks of uncertainty not based on additivity. We end by discussing DST and other imprecise probability measures measures showing the analogies between representing uncertainty in these theories ($Pl(\Theta)$-$\mathrm{P_U}(\Theta)$ pair and the $Bel(\Theta)$-$\mathrm{P_L}(\Theta)$ pair) thereby effectively having a generalised form for representing ignorance, as the difference between the upper and lower measures of probability.

References

[1] Aven, T. and Steen, R. (2010). The concept of ignorance in a risk assessment and risk management context. *Reliability Engineering & System Safety*, 95(11):1117–1122.

[2] Benétreau-Dupin, Y. (2015). The bayesian who knew too much. *Synthese*, 192(5):1527–1542.

[3] Dubois, D. and Prade, H. (2012). *Possibility theory: an approach to computerized processing of uncertainty*. Springer Science & Business Media.

[4] Ferson, S., Kreinovich, V., Ginzburg, L., Myers, D., and Sentz, K. (2002). Con-455 structing probability boxes and dempster-shafer structures. *Sandia Na-456 tional Laboratories Technical report SAND2002-4015, Albuquerque, NM*, 457:458.

[5] Giang, P. H. (2015). Decision making under uncertainty comprising complete ignorance and probability. *International Journal of Approximate Reasoning*, 62:27–45.

[6] Gordon, J. and Shortliffe, E. H. (1984). The dempster-shafer theory of evidence. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, 3:832–838.

[7] Keynes, J. M. (1921). *A treatise on probability*. Macmillan and Company, limited.

[8] Möller, B. and Beer, M. (2004). *Fuzzy randomness: uncertainty in civil engineering and computational mechanics*. Springer Science & Business Media.

[9] Norton, J. D. (2010). Cosmic confusions: Not supporting versus supporting not. *Philosophy of Science*, 77(4):501–523.

[10] Pettigrew, R. (2014). Accuracy, risk, and the principle of indifference. *Philosophy and Phenomenological Research*, 92(1):35–59.

[11] Pushkarskaya, H., Liu, X., Smithson, M., and Joseph, J. E. (2010). Beyond risk and ambiguity: Deciding under ignorance. *Cognitive, Affective, & Behavioral Neuroscience*, 10(3):382–391.

[12] Shafer, G. (1990). Perspectives on the theory and practice of belief functions. *International Journal of Approximate Reasoning*, 4(5-6):323–362.

[13] Shafer, G. (1992). Dempster-shafer theory. *Encyclopedia of artificial intelligence*, 1:330–331.

[14] Sullivan, T. J. (2015). *Introduction to uncertainty quantification*, volume 63. Springer.

[15] Walley, P. (1996). Measures of uncertainty in expert systems. *Artificial intelligence*, 83(1):1–58.

[16] Walley, P. (2000). Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning*, 24(2-3):125–148.

[17] Wang, Y. (2008). Imprecise probabilities with a generalized interval form. In *Proc. 3rd Int. Workshop on Reliability Engineering Computing (REC'08), Savannah, Georgia*, pages 45–59.

[18] Weichselberger, K. (2000). The theory of interval-probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning*, 24(2-3):149–170.

[19] White, R. (2010). Evidential symmetry and mushy credence. *Oxford studies in epistemology*, 3:161–186.

[20] Williamson, J. (2007). Motivating objective bayesianism: from empirical constraints to objective probabilities. *Probability and Inference: Essays in Honor of Henry E. Kyburg Jr*, pages 155–183.

[21] Zadeh, L. A. (1996). Possibility theory and soft data analysis. In *Fuzzy Sets, Fuzzy Logic, And Fuzzy Systems: Selected Papers by Lotfi A Zadeh*, pages 481–541. World Scientific.

A. Example 1: Advice from experts

Expert	Prescribed Treatment
A	100mg, once every 12 hours
B	At least 100mg, twice a day
C	200mg 4 times a day or 100mg once a day