

Numeric uncertainty

Sisukord

[peida]

1 A motivating example

2 Different types of uncertainties

2.1 Proper probabilities

2.1.1 Bayesians and frequentists

2.2 Subjective probability or confidence

2.3 Fuzzy measures

2.4 Well-knownness

2.5 Popularity

2.6 Context

3 Some example scenarios

4 Different ways to encode confidences in logic

4.1 Scenario introduction

4.2 Some notes about the scenario descriptions in FOL

4.3 Probabilities and confidences, object- and metalevel

4.3.1 Two-layer representation

4.3.2 Object level probabilities

4.3.3 Meta level probabilities and confidences

5 Rules with confidences

6 A scenario with different observations

6.1 Case 1: we estimate the error rate of the observer

6.2 Case 2: the observer estimates her error rate herself

6.3 Case 3: many observers, same information

6.4 Case 4: two observers, different confidences

- 7 Confidence calculation for the previous cases
- 8 Rules and confidences: next try
- 9 Box contents and multiple boxes
- 10 Calculating with both a statistical and meta-confidence attached
- 11 Information is lost when boxes are merged
- 12 Multiple boxes inside multiple boxes
- 13 Splitting and merging statistical and meta-confidences
- 14 Related literature

A motivating example

John observes something in the field which looks like a bird and estimates the probability of it being a bird as 80%. Mike observes the same object, but estimates the probability as 70%. I have an intuitive rule which says that birds can fly with the probability 90%. Let us formalize this as:

- $\text{bird}(\text{object1}) : 0.8 :: \text{John}$
- $\text{bird}(\text{object1}) : 0.7 :: \text{Mike}$
- $[\text{All } X. \text{bird}(X) \Rightarrow \text{canfly}(X)] : 0.9 :: \text{me}$

What can I derive from here? A simple idea is to combine the bird observation probabilities to a stronger one, using a standard probability calculation rule $P(a \vee b) = P(a) + P(b) - (P(a) * P(b))$, which holds in case a and b are independent observations. We get:

- $\text{bird}(\text{object1}) : 0.8 + 0.7 - (0.9 * 0.7) = 0.87 :: \text{John, Mike}$

Second, using the derivation rule, decreasing the probability according to a probability rule $P(a \& b) = P(a) * P(b)$, again in case when a and b are independent:

- $\text{canfly}(\text{object1}) : 0.87 * 0.9 = 0.783 :: \text{John, Mike, me}$

Now, what can go wrong with such calculations? A lot: there are nontrivial nuances. For example, does

- $\text{bird}(\text{object1}) : 0.1$

mean that object1 is unlikely to be a bird, or that it is a bit likely to be a bird? In other words, are small probability numbers really interpreted as positive or rather indications of a probability of a negation

- $\neg \text{bird}(\text{object1}) : 0.9$

For this particular question, we have to be careful to indicate what the probability number means: it can be interpreted in several ways.

What more could go wrong or require extra care?

- Are the calculation rules we used really true or applicable? In which context are they applicable? Can they lead to incorrect or unintuitive results in other cases?
- What about "independent": are these statements independent and to what degree?
- Are we sure our procedure does not use same input

several times to "over-strengthen" the probabilities?

- What is the "formalism" we used, ie how to encode that a rule/fact has a probability and sources? Can we use quantifiers also there, like quantifying over sources or something?

Different types of uncertainties

We intuitively understand and can use uncertain information: after all, almost all the information we have is uncertain, with some intuitive estimates we make.

Looking at the issue closer, we see that uncertainty stems from many reasons, has different contexts and usages and a lot of different ways to calculate.

The philosophical questions of what is probability at all and how it should be theorized about are complex. There are many different camps of people advocating different ways to handle uncertainty. Long time ago (before Kolmogorov) Bertrand Russell said in a 1929 lecture: "Probability is the most important concept in modern science, especially as nobody has the slightest notion what it means". The Stanford Encyclopedia of Philosophy article [probability interpretations](#) is a nice showcase of different ways to understand probability.

No common widely used method exists for using probabilities in knowledge representation and reasoning. Even if we choose one of the multitude of ways to handle

uncertainty, actually implementing the calculations for reasoning under uncertainty turns out to be much harder than the case with total certainty.

Hence most systems doing some type of nontrivial uncertainty reasoning are not too efficient and fall under the attack of other camps advocating different ways to do uncertainty reasoning. In other words, practical systems tend to perform only the most simple forms of uncertainty reasoning and often do not call their uncertainties "probabilities".

This does not mean that probabilities cannot be used in knowledge representation: just that one should be aware that this is a slowly developing complex area. In practice one should simply choose a way which seems to fit best for the expected use cases of the system.

For example, a machine learning guru Yann LeCun has said "he's ready to throw Probability Theory under the bus". Have a quick look at the [Medium article](#) explaining the probable reasons for LeCun's sentiment.

Proper probabilities

[Probability theory](#) is a classical part of mathematics operating in situations where we have good statistical estimates of the situation at hand. Kolmogorov [Probability axioms](#) forms the crucial axiomatic base for the theory.

Say, if a symmetric dice is thrown, there is a $1/6$ chance

that the points will be 1. If two dice are thrown, there is a $1/12$ chance that the summed points will be 2, but $2/12$ chance that the summed points will be 3 (because there are two cases $1+2$ and $2+1$ both giving 3).

Bayesians and frequentists

Two main camps for handling more-classical kinds of probabilities are

- [Frequentists](#): E2 the core intuition is treating probabilities as limit values of statistical experiments, when more and more experiments are made. Dice-kinds of examples fit this camp very well.
- [Bayesianists](#) E2 are more subjective, interpreting probability as "reasonable expectation" (a probability is assigned to a hypothesis), yet basing their calculations on the famous [Bayes theorem](#) of [Conditional probability](#).

Subjective probability or confidence

In most cases in practical life there is no way to properly estimate the probabilities: not even enough for Bayesianists. We simply have too little information. Then we use "subjective probability" or our estimate of probability. We also understand that this subjective probability itself may have a higher or lower probability of being true. It may be better to talk about **confidence**, since there is no dice-like information available.

Importantly, we should be able to

- Compute "and" of two uncertain facts: this makes the confidence of any long chain of reasoning lower and lower. Thus long chains of reasoning are not feasible (and people are not built for long reasoning chains, for this same reason).
- Combine fact A from source 1 with confidence $c(1,A)$ with the same fact A from another source with confidence $c(2,A)$ to get higher confidence than $c(1,A)$ and $c(2,A)$. The more independent sources claiming the same thing, the higher the confidence.

It has been shown in many ways, that people estimate probabilities and use them in a fairly complex way to our advantage, which does not exactly correspond to the simple probability rules (if there is such a thing). For the most famous of these theories and findings, see

- The simpler [Expected utility hypothesis](#).
- A more advanced [Prospect theory](#) of Kahneman and Tverski.

There are many ways to do reasoning with subjective probability, none of them are really mainstream. There are almost no systems seriously attempting nontrivial reasoning with subjective probability.

Have a brief look at the following main important lines of reasoning/camps-of-thought, just to understand that

there are really many different approaches. Most of the following are somewhat similar to [Bayesian probability](#), sometimes using also some form of [Imprecise probability](#):
E2

- [Dempster-Shafer theory](#)
- [Probabilistic logic](#)
- [Markov logic network](#)
- [Subjective logic](#)
- [Bayesian networks](#)
- [Probabilistic logic network](#)
- [Principle of maximum entropy](#)

Observe that the multitude of such camps does not mean the problem is "solved". In contrary, it means none of these approaches work very well in most cases: they tend to work well in some specific scenarios, but not others.

Fuzzy measures

The words "red", "tall", "big", "warm" etc have no exact meaning in terms of color saturation, meters, diameter, temperature or such. People tend to call two meter tall guys "tall" and 1.5 meter not so. What happens in between is nontrivial.

The standard example picture is [like this](#).

[Fuzzy logic](#) E2 was built to handle such cases. The fuzzy logic approach works quite well in specific measurement

cases with rules, and is thus actually used in practice.

The main ideas of fuzzy logic are:

- Map words like "red", "tall", "warm" etc to concrete functions from color saturation, height etc to a function giving values between 0 and 1 (the fuzzy measure).
- Use [fuzzy operators](#) "min" for "and", "max" for "or" and "1-X" for "not X".

Well-knownness

Some facts are well-known (like Donald Trump is the U.S. president) and others less so (like Kersti Kaljulaid is the president of Estonia) and others even less (Arthur Dempster was born in 1929).

Clearly the well-knownness is an important measure. For example, it makes sense to look at the more well known facts and rules while searching for an answer to some question.

Popularity

Some things/people/places and some facts come more to people's minds than others. For example, a lot fewer people look at the wikipedia page for "Bread" or "Mother" than "Donald Trump" or "Eiffel tower", even though "Bread" and "Mother" and rules about these things are

more well-known than about Trump or Eiffel tower.

Context

Some objects are more popular in some contexts. Some facts are true or have a higher confidence in some contexts and not others. Say, "Hobbit has a small sword" is a well-known high-confidence fact in the Tolkien context, less so in the general fantasy context and even less so in everyday context (where it may actually be even wrong).

Clearly, we could (and probably should) attach context tags (like "Tolkien", "Fantasy", "Estonia", "General" etc) with different confidence and popularity numbers to objects and facts and rules.

Some example scenarios

Consider these scenarios. Each of these is somewhat different and may require a different way to calculate probability or confidence.

Dice scenario fits classical probability theory: we throw two dice at the same time. What is the probability that the sum of points on these two dice is 3?

Capability to fly. Here we have some initial probability-like estimates and then rules with some confidence. We detect a object on the ground. With a probability 90% we estimate it is a bird. What is the confidence that it can fly

away if we continue approaching? Maybe it is a very young bird, a sick bird, a dead bird, or a penguin? What if we asked the same question in antarcotics with lots of penguins or a zoo?

Apple scenario. This is a fuzzy logic case. Here we have a more-or-less measure, not really a probability or confidence. Red apples should be sorted to the left box, green apples to the right box and all others to the middle box. An apple under observation is a mixture of red and green, with a bit more red. Clearly it should not go to the "green" box, but should we put it to the left or the middle box? Looks like we need a **measure of redness** and **thresholds rules** for the boxes.

Traffic light scenario.: a Dempster-Shafer theory case. Here we need to combine two probabilities. We have two light sensors pointed at a traffic light far away. The traffic light can be red, yellow or green. The light sensors are not perfect and cannot see very well from distance: there may be errors, let us say, with the average probability that a light sensor makes an error being 10%. This probability can be obtained from say, 1000 measurements, simply counting the number of errors.

- Suppose both sensors say "red". What is the probability that the light is really red?
- Suppose one says "red" and another "yellow". What is the probability now?
- Suppose one says "yellow" and another "green".

What is the probability now?

Self-driving car scenario. The car is driving fast on the highway and there are other cars following. There is a human-sized object standing on the curb of the road ahead. We cannot yet identify with certainty what it is. Should the car brake? If it brakes, it wastes fuel and time and more importantly, there is a chance that a car behind will slam into our car.

- What is the confidence that if we brake (say, with a given brake pressure and computable slowdown) the car behind will notice too late?
- What is the confidence that the object will step on the road in front of us (and we should rather brake beforehand to avoid collision)?
- What is the confidence that the object is an animal? What is the probability that it is a human?
- What is the confidence that the human-sized figure is seriously drunk? Maybe the camera can find some wobbly behaviour and estimate that there is a chance $X\%$ that it is drunk. What is the confidence of the object stepping on the road now? Should be brake and risk the car behind slamming into us?

Observe that there is no way to get reasonable statistical information about wobbly human-like objects on the curb being really drunk or stepping in front of a car or even us getting slammed by a car behind us when we brake with a certain intensity. Yet humans somehow estimate these

things even without proper statistical information.

Different ways to encode confidences in logic

We will now have a look at what are the different ways to actually encode confidences or probabilities in a logical language.

The scenario we use is of the frequentist kind, where real probabilities do exist. This makes it possible to actually check whether and in which cases we can calculate actually correct probabilities. If we can do so, then we could use the same approach for subjective probabilities, which are our primary target.

- $\forall x \dots$ means "for all x ..."
- $\exists x \dots$ means "exists x so that ..."

Scenario introduction

We will consider a scenario with a box containing exactly three coloured balls r (red), g (green), y (yellow).

We will now start to describe this scenario in logic in various ways, assigning confidences to our statements. We will also consider small modifications to the scenario with more balls and optionally black dots added to some of the balls.

The end goal is to find the best ways to calculate

confidences of logically derived statements, comparing it to statistical probability where possible.

The text below contains conceptually two parts: the first part describes several ways of describing the probabilities and uncertainties in our scenario and the second part focuses on a concrete metalogic representation.

Some notes about the scenario descriptions in FOL

First, minimal information, where R means "red", G "green" and Y "yellow" and we use constants 1, 2, 3 as labels for things: observe that there is no intrinsic correspondence between the labels 1,2,3 and the real balls r,g,y. Unless we axiomatize against it, we could have all three labels attached to just one ball or some other object altogether. Intrinsically the three predicates R, G, Y could also mean exactly the same thing. Thus

- $R(1) \ \& \ G(2) \ \& \ Y(3)$

does not tell us much. Adding

- $(1 \neq 2) \ \& \ (2 \neq 3) \ \& \ (1 \neq 3)$

tells us there are at least three distinct objects.

Additionally

- $\exists x [x=1 \vee x=2 \vee x=3]$

tells us there are no more objects than these three.

Adding

- $\forall x. (R(x) \Rightarrow \text{not } G(x)) \ \& \ (R(x) \Rightarrow \text{not } Y(x)) \ \& \ (G(x) \Rightarrow \text{not } Y(x))$

tells us that all R,G,Y colored objects have a single color.

In the following examples we do not assume that any of these axioms are given.

Probabilities and confidences, object- and metalevel

Let us have a scenario with a box containing three coloured balls: red, green, yellow.

How to write down the probabilities in FOL?

There are several ways of describing probabilities in FOL, with different expressivity and different models. Let us have a brief look at several options, from a more general and powerful to weaker ones.

Two-layer representation

We introduce a predicate P taking a full FOL formula as a first argument and the probability that it is true as the second argument and freely intermingle this predicate with variables and quantifiers out of the scope of P . we could write

- $P(\text{"}\forall x. R(x) \vee G(x) \vee Y(x)\text{"}, 1)$

Suppose we have a box where half of the red and green balls have a dot and balls of other colors do not have a dot. Say, there are four red balls, one with a dot and four green balls, three with a dot. There are also many yellow balls, all without a dot. Let $D(X)$ mean "X has a dot". Then we could write:

- $\forall X, XP. P(\text{"}R(X) \vee G(X)\text{"}, XP) \Rightarrow P(\text{"}D(X)\text{"}, XP/2).$

Clearly this representation, while highly powerful, creates immense problems in both model-theoretic interpretation and proof search. We will not look at it any further.

Object level probabilities

We could introduce two-argument predicates RP , GP and YP , respectively indicating that the probability of the first argument being red, green or yellow, respectively, is given by the second argument:

- $(\forall x. RP(x, 1/3))$
- $(\forall x. GP(x, 1/3))$
- $(\forall x. YP(x, 1/3))$

We cannot write the formula for describing the situation in the section above: there is no way to write down a probability of a disjunction $R(X) \vee G(X)$.

Although simpler than the representation in the previous

section, this representation also creates significant problems.

Meta level probabilities and confidences

In the following we will write the statements about the probability and confidence of FOL statements as metalogic statements like this:

- $S : c$
- $S : p$

Where S is a FOL statement, c is the confidence we assign to the statement, p is the probability we assign to the statement.

Observe that the metalogical statements

- $(\forall x. R(x)) : p \ 1/3$
- $(\forall x. G(x)) : p \ 1/3$
- $(\forall x. Y(x)) : p \ 1/3$

have a different meaning than the statements containing RP , GP , YP above: they mean that ALL the objects in the box are red (resp green or yellow), with the probability $1/3$.

Such statements could be correct if we considered the possibility of several boxes, where some of these boxes contain only red, some contain only green balls, etc.

In literature this approach is often treated similarly to modal logics, where different boxes are considered as

different worlds, and then the distinction is made between two kinds of probabilities:

- Statistical "object probability" in a given box,
- Probability that we are given a certain box or a certain kind of a box.

Now, let us have a single box with three balls of color red, green and blue. Suppose we pick one ball from the box, labelled 1. We do not yet look at the ball and hence do not know its color. The following statements with probabilities 0.33 are very close to the statements with the number $1/3$, but not really the same:

(e)

- $R(1) : p \ 0.33$
- $G(1) : p \ 0.33$
- $Y(1) : p \ 0.33$

The 0.33 above is not actually the correct probability: $1/3$ is. In any but the totally abstract situation the actual, exact probabilities are unknowable.

We will rewrite the same statements with confidences like this:

(f)

- $R(1) : 0.33$
- $G(1) : 0.33$

- $Y(1) : 0.33$

For cases where a true probability p exists for some statement S (although we may not know its value) then we state the confidence correctness criteria: a confidence statement $S : c$ is true if and only if $c \leq p$ (less or equal than).

For cases where no true probability can be given for S , such a condition is obviously impossible to apply.

One of our goals is to find formulas for calculating the confidences of derived FOL formulas so that

- they obey the confidence correctness criteria
- we strive towards maximising the confidences

In other words, if both are given

- $S : c$
- $S : c'$

where $c' > c$ and c' obeys the confidence correctness criteria, we should prefer c' as a confidence. However, a statement may still be given with arbitrarily many different confidences: we only prefer the highest, but we do not say that the lower confidences are incorrect.

Again, consider a scenario where half of the red balls in a box have dots, and balls of other color do not have dots.

We could axiomatize that non-red balls are not dotted

using metalogic, since we do not need the probability variable for calculation:

- $(A \ X \ . \ -R(X) \Rightarrow -D(X)) : 1$

However, what about half of the red balls being dotted, for which we wrote in object logic

- $A \ X. \ RP(X,1) \Rightarrow DP(X,0.5)$ or
- $A \ X,XP. \ RP(X,XP) \Rightarrow DP(X,XP/2)$

What would it mean in our metalogic if we wrote the following formula?

- $(A \ X. \ R(X) \Rightarrow D(X)) : 0.5$

Let us approach this question by first looking at simpler cases.

Rules with confidences

Let us consider a scenario where we have six balls in the box, two of each color (red, green, yellow) and that one of the red balls has a black dot and no other balls have a black dot. We pick a ball labelled 1 from the box, but do not yet look at it, thus we do not know its color or other properties.

Suppose we write rules for this specific scenario as

(g)

- $R(1) : 0.33$
- $[A \ X. R(X) \Rightarrow D(X)] : 0.5$

Now, using modus ponens we derive

- $D(1)$

and suppose we will use multiplication of input confidences for calculating the result confidence $0.33 \cdot 0.5 = 0.165$, giving

- $D(1) : 0.165$

which is a correct confidence: the real probability of $D(1)$ exists and is very slightly higher than 0.165.

Note that we still do not have a satisfactory meaning for the statement

- $[A \ X. R(X) \Rightarrow D(X)] : 0.5$

but we decided to treat it similarly to the object logic rule

- $[A \ X, XP . RP(X, XP) \Rightarrow D(X, XP/2)]$

Next, suppose we know two balls in the box have a black dot: one is red and the other is green or yellow.

The previous axioms (g) are still valid.

We will now add a new axiom

- $D(1) : 0.33$

and consider the previously derived result

- $D(1) : 0.165$

We have thus two different confidences for the same statement. Both are lower than the the real probability of $D(1)$ which is slightly higher than 0.33, hence both are correct. We should prefer the higher confidence, that is, 0.33.

Now, let us consider the idea of calculating the cumulative confidence as $0.33 + 0.165 - (0.33 * 0.165) = 0.44055$ This would be an incorrect confidence: it is higher than the actual probability.

The last formula assumes that the statements for which we had confidences 0.33 and 0.165 are independent. To disallow the formula application we should show that the confidence-giving statements are dependent.

This is a hard task which we cannot satisfactorily solve yet. We can say though that no observations of a random variable have been made for this calculation. Since there are no observations, there cannot be independent statements?

A scenario with different observations

We have a large unknown number of balls in the box: we know there are many different colors, but not how many colors or how many balls of each there are. One ball with

the label 1 is taken from the box and shown to an observer, who cannot see very clearly: the room is dark. The observer has to name the color.

Case 1: we estimate the error rate of the observer

Observer o1 says: it is red.

How to estimate the likelihood that the ball is really red?

Let us make an experiment with the same observer in the same room with different balls and register all cases where the observer says "it is red"

Suppose the observer is correct in 90% of the cases.

Knowing the error rate of the observer as 90% we write the confidence that the ball with the label 1 is red as:

- $R(1) : 0.9$

Case 2: the observer estimates her error rate herself

Observer o1 says: I think 90% it is red, but I am not 100% sure.

We make no experiments about the error rate of the observer, but trust her own estimate (maybe she has made such experiments herself and knows her error rate).

Again, we write:

- $R(1) : 0.9$

Case 3: many observers, same information

There are 10 observers, each saying: I think 90% it is red, but I am not 100% sure. Again, we trust the estimates of the observers.

We write:

- $R(1): 0.9 :: O1$
- $R(1): 0.9 :: O2$
- ...
- $R(1): 0.9 :: O10$

denoting also the observer id-s in addition to the confidences.

How should we estimate the cumulative confidence that the ball is red?

The probability of all the observers independently making an error is 0.1 to the power of 10.

Hence we calculate the cumulative confidence as

- $R(1) : 1-(0.1 ** 10) :: O1,...,O10$

Case 4: two observers, different confidences

There are 2 observers, giving different confidences, which we write as

- $R(1) : 0.8 : O_1$
- $R(1) : 0.5 : O_2$

Again, we trust the error estimates of the observers.

The probability of both observers making an error is $0.2 \cdot 0.5 = 0.1$, thus we calculate the cumulative confidence as

- $R(1) : 0.9 :: O_1, O_2$

Confidence calculation for the previous cases

Generally for all the cases when we have independent confidences c_1, \dots, c_n , we calculate the cumulative confidence with the standard rule as

- $1 - ((1-c_1) \cdot (1-c_2) \cdot \dots \cdot (1-c_n))$

which for two observers is the standard rule

- $1 - ((1-c_1) \cdot (1-c_2)) = 1 - (1-c_2-c_1+c_1 \cdot c_2) = c_1 + c_2 - c_1 \cdot c_2$

Rules and confidences: next try

Let us have a box with colored balls, some of which may have dots. We do not know how many colors or balls there

are, but observers will make educated guesses.

This scenario is similar to the idea of having multiple boxes (worlds) with different content. The observer does not know which box is chosen.

Let us consider the previously given rules for red balls and dotted balls as:

- (g)
- $R(1) : 0.33$
- $[A\ x. R(x) \Rightarrow D(x)] : 0.5$

But this time let us interpret the rules as educated guesses by different observers about what could be in the box. Maybe they have done some experiments by picking some balls from the box, but they do not know what is really in the box.

- $R(1) : 0.33 :: o1$

meaning that the observer $o1$ guesses that the statement

- $R(1)$

holds, but her confidence that it is correct is only 0.33, This guess could be motivated by earlier experiments by $o1$ taking three balls from the box and finding that only one is red.

- $[A\ x. R(x) \Rightarrow D(x)] : 0.5 :: o2$

meaning that the observer o2 guesses that the rule

- $[A \ x. R(x) \Rightarrow D(x)] : 0.5$

holds, but her confidence that the rule is correct is only 0.5.

This guess could be motivated by an experiment by o2 with several boxes, taking several balls from each box and for half the boxes observing that any time a ball was red, it had a dot, while for other boxes a red ball never had a dot.

We could argue that the same considerations apply for such general observations about the balls in the box as they do apply for observations about the color of the concrete balls taken from the box as we examined earlier.

How should we estimate the confidence that the following logical deduction from these two statements holds:

- $D(1)$

The chance of both of these statements

- $R(1)$
- $[A \ x. R(x) \Rightarrow D(x)]$

being true is just $0.33 * 0.5 = 0.165$

Now, if both statements do indeed hold, then so do their logical conclusions, hence we can write

- $D(1) : 0.165 :: o1, o2$

Suppose now we have a third observer $o3$ who simply states

- $D(1) : 0.5 :: o3$

We now have two independent observations about the box where the first one is the combination of $o1$ and $o2$ and the second is made by $o3$:

- $D(1) : 0.165 :: o1, o2$
- $D(1) : 0.5 :: o3$

and we can apply the cumulative confidence rule giving $0.165 + 0.5 - (0.165 * 0.5) = 0.5825$ as confidence:

- $D(1) : 0.5825 :: o1, o2, o3$

Now suppose the second observer has also made an additional claim

- $D(1) : 0.6 :: o2$

in addition to her previous sole claim

- $[A \ x. R(x) \Rightarrow D(x)] : 0.5 :: o2$

Notice that these two claims by $o2$ are not inconsistent.

In order to apply the cumulative rule to these two

- $D(1) : 0.5825 :: o1, o2, o3$

- $D(1) : 0.6 :: o2$

we'd need to know that they are independent. But, since the first of these is dependent on all three observers, we should not apply the rule. However, if the error rates in these statements are correct, we should simply prefer the better result:

- $D(1) : 0.6 :: o2$

Box contents and multiple boxes

The scenario with multiple boxes having different content means that we could separately give

- object confidences of properties of a ball taken from one concrete box,
- meta-confidences of a selection of boxes.

Clearly these two are similar. The edge cases would be

- contents of all the boxes moved to a single box,
- or many boxes with only one ball in each.

Notice that in case some boxes contain many more balls than other boxes, then moving all balls from all boxes to a single box would mean that the distribution of balls in larger original boxes will dominate the distribution in smaller boxes.

Let us again have a look at the possibilities of describing

object confidences, i.e. contents of a single box. The (meta)confidences in the previous section do not allow to describe that, say, one third of the balls in the box are red or that half of the red balls have a dot.

We will introduce a weak notation of object confidences to describe these properties as "statistical confidence" written

- $G :s c$

where G is in the prefix form, will mean the following: c is less or equal than the proportion of combinations of the value tuples of all the leading universally quantified variables such that G holds for exactly these value tuples. Example:

- $(\forall X. R(X)) :s 1/3$

means that R holds only for $1/3$ of the values for X , similarly

- $(\forall X. R(X) \Rightarrow D(X)) :s 1/3$

means that $R(X) \Rightarrow D(X)$ holds for $1/3$ of the values for X .

Such a statement makes clear sense only if the domain of the selection of the variable X is finite. Saying

- $(\forall X. \text{integer}(X) \Rightarrow \text{even}(X)) :s 0.5$

would seem to be incorrect: the sets of integers and even

integers have the same cardinality, even though the statement is true for any finite slice of integers.

For finite domains of the variable value selection the multiplication

As a more general example,

- $(\forall X, Y. M(X, Y) \Rightarrow M(Y, X)) :s 1/3$

means that $M(X, Y) \Rightarrow M(Y, X)$ holds for 1/3 of the tuples of X, Y .

The multi-box confidence treated in the previous section will be applicable to the whole statistical confidence statement, written like this

- $(\forall X. R(X) \Rightarrow D(X)) :s 1/3 : 1/2$

meaning that at least half of the boxes have a property

- $(\forall X. R(X) \Rightarrow D(X)) :s 1/3$

Let us again consider an example from the previous section:

- $R(1) : 0.33 :: o1$
- $[A x. R(x) \Rightarrow D(x)] : 0.5 :: o2$

but this time using the statistical confidence for the second rule:

- $[A x. R(x) \Rightarrow D(x)] :s 0.5 :: o2$

Does the applying the statistical rule lead us to the same seemingly obvious logical conclusion?

- $D(1)$

Not necessarily: for a box with two red balls 1 and 2 the statement

- $[A \ x. R(x) \Rightarrow D(x)] :s \ 0.5$

would hold both in case the dot is on ball 1

- $D(1) \ \& \ -D(2)$ where
- $(R(1) \Rightarrow D(1)) \ \& \ -(R(2) \Rightarrow D(2))$

and when the dot is on ball 2

- $-D(1) \ \& \ D(2)$ where
- $-(R(1) \Rightarrow D(1)) \ \& \ (R(2) \Rightarrow D(2))$

Hence we cannot really say that $D(1)$ is logically derivable, although it is derivable in some situations (worlds in modal logic).

In case the probability of such different situations is equal, the probability of worlds where $D(1)$ holds is 0.5.

If so, then the chance that $D(1)$ indeed holds is at least 0.33, thus the following statement should hold:

- $D(1) : 0.33 \cdot 0.5 = 0.165 : o1, o2$

Calculating with both a statistical and meta-confidence attached

Next, consider the situation where we have a set of similar statements, but the second rule has a meta-confidence 0.1 attached:

- $R(1) : 0.33 :: o1$
- $[A \ x. R(x) \Rightarrow D(x)] :s\ 0.5 : 0.1 :: o2$

meaning that there is a 0.1 chance that the statistical two-to-one relation holds in the box. For all the other cases we know nothing. Clearly we have no certain grounds for calculating the confidence of $D(1)$ holding.

In order to calculate anything sensible, we have to make assumptions. An arbitrary, but simple assumption is that the same rule

- $[A \ x. R(x) \Rightarrow D(x)] :s\ 0.5$

holds also for all the other cases, about which we know nothing.

If we make this assumption, the result of a derivation is the same as in the previous section:

- $D(1) : 0.33 * 0.5 = 0.165 :: o1, o2$

Now consider two boxes with different statistical and meta-confidences:

- $(\text{All } X. R(X) \Rightarrow D(X)) :s 0.5 : 0.9 :: o2$
- $(\text{All } X. R(X) \Rightarrow D(X)) :s 0.25 : 0.1 :: o3$

Again, what is the confidence that $D(1)$ holds?

In case the box picked was the first one, the confidence is as calculated before:

- $D(1) : 0.33*0.5=0.165 : 0.9 :: o1,o2$

In case the box picked was the second one, the confidence is analogously

- $D(1) : 0.33*0.25=0.0825 : 0.1 :: o1,o3$

The chance of picking the first box was 10 times higher than that of the second. If these were independent choices, we could calculate the cumulative confidence as

- $D(1) : (0.165*0.9)+(0.0825*0.1)-((0.165*0.9)*(0.0825*0.1)) = 0.15552487 : o1,o2,o3$

The sources for the last derivation share a common element $o1$, while $o2$ and $o3$ are different:

- $D(1) : 0.33*0.5=0.165 : 0.9 :: o1,o2$
- $D(1) : 0.33*0.25=0.0825 : 0.1 :: o1,o3$

Question: is it certain that these statements are really independent? If yes, what are the suitable independence criteria regarding depending-upon statement lists like $o1,o2$ and $o1,o3$

Information is lost when boxes are merged

Let us now consider the idea of taking several boxes and moving all their balls into a single box. If some boxes contain many more balls than other boxes, then moving all balls from all boxes to a single box would mean that the distribution of balls in larger original boxes will dominate the distribution in smaller boxes.

Say in box 1 we have 100 balls, 8 of them red, of these four dotted: it holds that

- $(\text{All } X. R(X) \Rightarrow D(X)) :s 0.5$

and in box 2 we have 10 balls, 4 of them red, of these one dotted: it holds that

- $(\text{All } X. R(X) \Rightarrow D(X)) :s 0.25$

The end result of merging the boxes gives us 110 balls, 12 red, five of these dotted. It holds that

- $(\text{All } X. R(X) \Rightarrow D(X)) :s 5/12$

Now consider the combined object and meta-confidence statements for the original boxes, where we say that the confidence of picking the first box vs the second box is proportional to the number of balls in them:

- $(\text{All } X. R(X) \Rightarrow D(X)) :s 0.5 : 0.9$

- $(\forall X. R(X) \Rightarrow D(X)) :s 0.25 : 0.1$

From this information we cannot obtain the statistical confidence for the merged boxes:

- $(\forall X. R(X) \Rightarrow D(X)) :s 5/12$

To get the latter number, we'd need the information about the number of red balls in each, which is not given in the statements.

Multiple boxes inside multiple boxes

As a commonsense scenario we could imagine that in a normal day in a city in Europe, whenever I see a bird, this bird can fly with a very high statistical confidence:

- $(\forall X. \text{bird}(X) \Rightarrow \text{canfly}(X)) :s 0.99$

whereas on the antarctic coast a large percentage of birds are penguins, thus the percentage of flying birds like seagulls is much lower:

- $(\forall X. \text{bird}(X) \Rightarrow \text{canfly}(X)) :s 0.5$

A city in europe would thus be seen as one box and antarctic coastline as another box. The confidence that we are in one or another situation (box) could be given, for example, as

- $(\forall X. \text{bird}(X) \Rightarrow \text{canfly}(X)) :s 0.99 : 0.9$
- $(\forall X. \text{bird}(X) \Rightarrow \text{canfly}(X)) :s 0.5 : 0.01$

covering just a part of all the available boxes, like zoos, australian deserts, jungles, etc.

As a side note, observe that if we have two statements

- $F :s\ p1 : c1 :: s1$
- $G :s\ p2 : c2 :: s2$

where G is a generalization (or identical to) F and the set of depending-upon formulas $s2$ is a subset of $s1$ and $p2 \geq p1$ and $c2 \geq c1$ then using the second statement is always preferable to the first statement: anything derivable from the first can be derived from the second, just potentially with higher confidences.

Clearly it would be useful to know which box is given or which boxes are more likely to be taken. One way to express this would be to assign each box a name like "europe" or "antarctica" or "zoo" and attach this name to the formula like

- $(\lambda X. \text{bird}(X) \Rightarrow \text{canfly}(X)) :s\ 0.99 : \text{europe}\ 0.9$
- $(\lambda X. \text{bird}(X) \Rightarrow \text{canfly}(X)) :s\ 0.5 : \text{antarctica}\ 0.01$

to use just this statement if a situation is known.

Now, suppose we are in a context where the situation may switch between europe and antarctica with equal probability; then the given meta-confidences 0.9 and 0.01 would be misleading: instead we could choose

- $(A \text{ X. bird}(X) \Rightarrow \text{canfly}(X)) :s \text{ 0.99 :europe 0.5}$
- $(A \text{ X. bird}(X) \Rightarrow \text{canfly}(X)) :s \text{ 0.5 :antarctica 0.5}$

The contexts might also not be mutually exclusive: a zoo context like

- $(A \text{ X. bird}(X) \Rightarrow \text{canfly}(X)) :s \text{ 0.9 :zoo 0.1}$

could hold at the same time as

- $(A \text{ X. bird}(X) \Rightarrow \text{canfly}(X)) :s \text{ 0.99 :europe 0.5}$

and in such cases it would make sense to pick the more specific statement, if possible: here "zoo" instead of "europe".

Apparently the selection of situations may, in a general case, be hierarchical, with different branches having different confidences.

Splitting and merging statistical and meta-confidences

We assume that in the applications of commonsense reasoning we cannot always obtain both statistical and meta-confidences. Instead, we are likely to obtain confidence numbers which approximate both at the same time. In other words, given a statement

- $S :? n$

we should estimate the values for

- $S : s \ c1 : c2$

Question: what would be the sensible assumptions?

Perhaps there is no need to split them into two parts, but to operate with a single confidence number in all cases?

Related literature

An Analysis of First - Order Logics of Probability Joseph Y. Halpern <https://www.ijcai.org/Proceedings/89-2/Papers/084.pdf>

A Logic for Reasoning about Evidence Joseph Y. Halpern
Riccardo Pucella <https://arxiv.org/pdf/1407.7185.pdf>

A First-Order Logic of Probability and Only Knowing in
Unbounded Domains Vaishak Belle Gerhard Lakemeyer
Hector J. Levesque
<https://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/download/12271/11680>

Probabilistic Reasoning Thomas Lukasiewicz
<http://www.cs.ox.ac.uk/people/thomas.lukasiewicz/sli-kr16a.pdf>

A Logic for Default Reasoning About Probabilities Manfred
Jaeger <https://arxiv.org/pdf/1302.6822.pdf>

Unifying Logic and Probability STUART RUSSELL
<https://people.eecs.berkeley.edu/~russell/papers/cacm15-oupm.pdf>

General-Purpose MCMC Inference over Relational Structures Brian Milch Stuart Russell

<https://arxiv.org/pdf/1206.6849.pdf>

First-Order Probabilistic Languages: Into the Unknown
Brian Milch 1 and Stuart Russell 2

<http://people.csail.mit.edu/milch/papers/ilp06-fopl.pdf>

Representing and Reasoning With Probabilistic Knowledge: A Bayesian Approach Marie desJardins

<https://arxiv.org/pdf/1303.1481.pdf>

Probabilistic Description Logics: Reasoning and Learning

Riccardo Zese <http://2017.ruleml-rr.org/wp-content/uploads/2017/07/Riccardo-Zese-RuleMR-RR-tutorial.pdf>

Probabilistic Datalog+/- under the Distribution Semantics
Fabrizio Riguzzi, Elena Bellodi, and Evelina Lamma

http://ceur-ws.org/Vol-846/paper_25.pdf

Coherence Branden Fitelson <http://fitelson.org/coherence/>

http://fitelson.org/coherence/coherence_duke.pdf

Chapters 5 & 6 of Foundations of Measurement: Volume I
Krantz, Luce, Tversky, and Suppes,

http://fitelson.org/coherence/fom_chs_5_6.pdf