

COMBINING PROBABILITIES

444 Lecture 6

Brian Weatherston

2024-02-02

DISCURSIVE DILEMMA

A PUZZLE

	p	q	p & q
A	✓	✓	✓
B	✓	✗	✗
C	✗	✓	✗
Group	✓	✓	✗

AN IMPORTANT EXAMPLE

- A, B and C are judges.
- In the case in front of them, the plaintiff wins if, and only if, both p and q are true.
- And the ticks and crosses in the columns under p and q are the individual judge's views on whether each of these parts of the case are satisfied.

AN IMPORTANT EXAMPLE

Here are two methods the judges could use to resolve the case:

1. Vote on p, vote on q, and if a majority says yes in each case, the plaintiff wins.
2. Vote on p&q, and if a majority says yes, the plaintiff wins.

Method 1 delivers a win to the plaintiff, method 2 delivers a win to the defendant.

PROBABILITY TO THE RESCUE

NOT A PUZZLE

	p	q	p & q
A	1	1	1
B	1	0	0
C	0	1	0
Group	2/3	2/3	1/3

GENERAL PRINCIPLE

The average of some probabilities is a probability.

If everyone in the group has consistent probabilities, the group will have consistent probabilities.

NATURAL RULE

The group probability is just the average of the probabilities of the members of the group.

By 'average' here, we're talking about the **arithmetic mean**: add the probabilities up and divide by how many judges there are.

ADVANTAGES

1. Simple
2. Intuitive
3. Consistent

ANOTHER ADVANTAGE - EQUAL BETWEEN VOTERS

One way to spell this out: no reference to people in the definition.

Another way - if two people switch views entirely, end up with same result.

ANOTHER ADVANTAGE - EQUAL BETWEEN OPTIONS

One way to spell this out: no reference to options in the definition. (Compare views that say don't make a change unless super-majority approves.)

Another way - if everyone switches their views between two options, the position of those two options in the group probability switches.

YET ANOTHER ADVANTAGE

- ATOMIC

To know the group view about the probability of p , you just need to know what each member thinks about p , and not about anything else.

COMPARISON TO PAPER

The last three slides have described the features they call

1. Anonymity
2. Neutrality
3. Irrelevant Alternatives

ADVANTAGES

I don't want to undersell these advantages.

Until not that long ago, I thought the problem of how to get a group probability out of individual probabilities was a simple problem with a simple solution ... this one.

But in this area we really aren't allowed to have nice things.

INDEPENDENCE

DEFINITION

According to probability function Pr , two propositions p and q are independent if any of the following, equivalent, conditions are met. (Caveat: weird things happen if probability of p or q is 0; set that case aside for now.)

$$\begin{aligned}\text{Pr}(p|q) &= \text{Pr}(p) \\ \frac{\text{Pr}(p \wedge q)}{\text{Pr}(q)} &= \text{Pr}(p) \\ \text{Pr}(p \wedge q) &= \text{Pr}(p)\text{Pr}(q)\end{aligned}$$

A PUZZLE

	p	q	p & q
A	0.9	0.9	0.81
B	0.1	0.1	0.01
Group	0.5	0.5	0.42

The two propositions are independent according to each member of the group, but not according to the group as a whole.

A PUZZLE ABOUT DEFERENCE

Imagine that you know nothing about p , q , but you know that A and B are experts, and think you should defer to them.

How do you defer to both of them?

Natural way - by setting your probability to the average of theirs.

Problem: You'll end up thinking something both of your experts reject, namely that p and q are not independent.

A PRINCIPLE ABOUT DEFERENCE

I'm going to use this heuristic a bit, and I just want to flag it because while I think it's useful, you might think it's where I go wrong.

- The group probability just is the probability you'll end up with if you defer to the members of the group collectively.

LEARNING

CONDITIONALISATION

When we are thinking about personal probabilities, we normally assume that learning goes by conditionalisation.

What that means is that the probability of a hypothesis H after learning evidence E is $\Pr(H | E)$. Or in symbols, where \Pr_E is the probability after learning E :

$$\Pr_E(H) = \Pr(H|E) = \frac{\Pr(H \wedge E)}{\Pr(E)}$$

GROUP LEARNING

A version of the case we've already seen shows that the following two claims are inconsistent:

1. The group probability is the arithmetic mean of the individual probabilities.
2. Group learning goes by conditionalisation.

A PUZZLE ABOUT LEARNING

Imagine that we've just learned that q is true.

This doesn't affect A and B , because they think p and q are independent.

So the average of their probabilities is 0.5, and that's arguably the group probability.

But ...

A PUZZLE ABOUT LEARNING

Imagine that we'd:

1. Asked A and B their opinions yesterday.
2. Used the arithmetic mean to form a group probability.
3. Updated that group probability by conditionalisation.
So when (as it turns out) we learn q , the new $\Pr(p)$ is the old $\Pr(p|q)$.

That looks like it should give the same verdict, but it doesn't.

A BIG DISADVANTAGE OF THE VIEW

The principle that Russell et al call Conditionalisation basically says the following two things should give the same verdict, at least if A and B have learned the same things since yesterday.

1. Asking A and B their opinions today, and merging their opinions.
2. Asking A and B what their opinions were yesterday, merging their opinions, asking them what they've learned since, and updating.

GEOMETRIC MEAN

They note that there is one interesting rule that does satisfy this constraint.

It's moderately tricky to state in full generality, and I'm just going to state a rough version of it, and leave it to more math-centric classes to get the details fully right.

(I'm not sure the paper gets the details fully right, in cases where A and B have opinions over the distribution of a continuous variable. The math here is actually rather tricky.)

GEOMETRIC MEAN

If x and y are non-negative numbers, their geometric mean is:

$$\sqrt{xy}$$

In general the geometric mean of some non-negative numbers x_1, x_2, \dots, x_n is

$$\sqrt[n]{x_1 x_2 \dots x_n}$$

GEOMETRIC POOLING

1. Work out the geometric means of the probabilities of each 'world', i.e., maximally specific possibility.
2. Sum these to get a total t .
3. Multiply each of the sums by $1/t$ to get the probability of each world.
4. For each proposition that's true in multiple worlds, sum the probability of the worlds to get its probability.

A DIFFERENT PUZZLE

THREE DETECTIVES

I don't think trying to go through the math of that in detail is worth it; if you didn't get it straight away, nothing I say in 5-10 minutes will help. And the details don't matter for what we'll do next.

Instead I want to leave you with a puzzle.

THREE DETECTIVES

Three excellent detectives, A, B, and C, are investigating a crime

They've each seen all the evidence, but you know they have very different approaches to thinking about crimes and cases like this.

You also know they are much better at crime solving than you are.

They are also each incredibly self-confident; once they've formed a view, knowing what the others think won't change their view.

THREE DETECTIVES

You ask each of them what they think, and they each say it's 90% likely that the butler did it.

Question: How confident should you be that the butler did it?

OPTION 1

This is easy - 90%.

After all, they agree about this, and they're smarter/better informed/better at crime solving than you are.

OPTION 2

More than 90%. Here's why.

After hearing A say 90%, it would be natural to be 90% confident that the butler did it.

But then hearing B and C each say that is further evidence that the butler did it.

And when you get evidence like that, your probability should go up.

WHICH IS RIGHT?

I have no idea - I think it's a hard puzzle!

It's not strictly speaking a math puzzle. Given some mathematical models of combining probabilities we can do some algebra to work out which model agrees with which answer.

As it turns out, arithmetic mean agrees with the first, geometric mean with the second, and there are plenty of other views out there that take one or other side.

But the question here is what we want our models to do, not I think one that we should leave to be answered by the models.

FOR NEXT TIME

FOR NEXT TIME

We'll move from pooling probabilities to pooling preferences.

That is, we're going to start talking about voting systems.