Logic, First Course, Winter 2020. Week 10, Lecture 1. Back to course website

Conditional Probability

Definition

The conditional probability of *h* given *e*, written $Pr(h \mid e)$, is defined as the fraction of $Pr(h \wedge e)$ over $Pr(e)$:
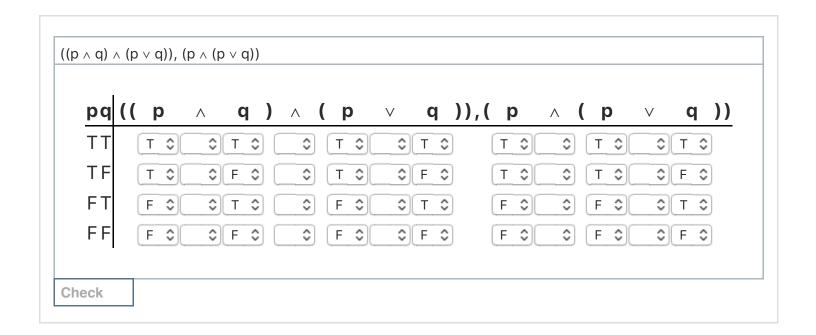
$$Pr(h \mid e) = \frac{Pr(h \wedge e)}{Pr(e)}$$

What is the underlying idea behind conditional probability? By definition, $Pr(h \mid e)$ is the probability of the *h* ∧ *e* rows over the probability of the *e* rows. In the case where the rows are all equally weighted, this comes down to asking what percentage of the *e* rows are also *h* ∧ *e* rows.

The choice of variables h, e here come from the canonical application, where h is a *hypothesis* and e is *evidence* which bears on the hypothesis. As we mentioned last time, one way to think about probability is as degrees of belief. To continue this, the idea is that $Pr(h \mid e)$ reflects one's degree of belief in the hypothesis h once one has come to learn evidence e. Hence, in the definition of conditional probability, when we form the conjunction of $h \wedge e$ and the evidence e, we are restricting attention to the e rows of the truth-table, since these are the rows where the evidence e is actually true.

Examples

Let us illustrate this with respect to the following, where we suppose that all rows have an equal chance of happening, namely a 1/4 chance of happening:

((p ∧ q) ∧ (p ∨ q)), (p ∧ (p ∨ q))

pq	((	p	∧	q	)	∧	(	p	∨	q	)),(	p	∧	(	p	∨	q	))
TT		T		T				T		T		T			T		T	
TF		T		F				T		F		T			T		F	
FT		F		T				F		T		F			F		T	
FF		F		F				F		F		F			F		F	

Check

Example 1. Let us calculate $Pr(p \wedge q \mid p \vee q)$.

$$Pr\left((p \wedge q) \mid (p \vee q)\right) = \frac{Pr\left((p \wedge q) \wedge (p \vee q)\right)}{Pr(p \vee q)}$$

$$= \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{\not{4}} \cdot \frac{\not{4}}{3} = \frac{1}{3}$$

Example 2. Let us calculate $Pr(p \vee q \mid p \wedge q)$.

$$Pr\left((p \vee q) \mid (p \wedge q)\right) = \frac{Pr\left((p \vee q) \wedge (p \wedge q)\right)}{Pr(p \wedge q)}$$

$$\overset{LCC}{=} \frac{Pr\left((p \wedge q) \wedge (p \vee q)\right)}{Pr(p \wedge q)} = \frac{\frac{1}{4}}{\frac{1}{4}} = 1$$

In this problem LCC is just our familiar abbreviation for the law of commutativity of conjunction. We can do this inside the probability operator due to the the connections between probability and logic.

Example 3. Let us calculate $Pr(p \mid p \vee q)$.

$$Pr(e \mid e \vee q) = \frac{Pr(e \wedge (e \vee q))}{Pr(e \vee q)}$$

$$= \frac{\frac{1}{2}}{\frac{3}{4}} = \frac{1}{2} \cdot \frac{4}{3} = \frac{2}{3}$$

Example 4. Let us calculate $Pr(p \vee q \mid p)$.

$$Pr((p \vee q) \mid p) = \frac{Pr((p \vee q) \wedge p)}{Pr(p)}$$

$$\overset{\text{def}}{=} \frac{Pr(p \wedge (p \vee q))}{Pr(p)} = \frac{\frac{1}{2}}{\frac{1}{2}} = 1$$

Here is a technique to slightly speed such calculations up. First, we know that $(p \wedge q) \wedge (p \vee q)$ is equivalent to $p \wedge q$:

Hence, we know from the connections between probability and logic that their $Pr((p \wedge q) \wedge (p \vee q))$ = $Pr(p \wedge q)$. This is useful because we could have just seen that $Pr(p \wedge q)$ has probability 1/4, without having had to do the truth-table up above (under the assumption at the rows are all equally weighted).

Bayes' theorem

Formally, Bayes' theorem is just the following identity, whose parts we abbreviate as follows:

$$Pr(h \mid e) = \frac{Pr(e \mid h) \, Pr(h)}{Pr(e)}$$

$$= \frac{likelihood \times prior}{evidence}$$

The names for these parts are as follows, along with their intuitive meaning:

- *Pr(h | e)* is just the conditional probability of the hypothesis *h* given the evidence *e*. This is how likely the hypothesis *h* is made by observation of the evidence *e*.
- *Pr(e | h)* is the *likelihood*: this the conditional probability of the evidence given the hypothesis *h*. This is how likely the evidence is, given that we assume the hypothesis. More intuitively still, it is how likely the evidence is by the lights of the hypothesis.
- *Pr(h)* is the *prior* probability: it is how probable hypothesis *h* is full stop, and not conditional on any other proposition.
- *Pr(e)* is the probability of the evidence: it is how probable the evidence *e* is full stop, and not conditional on any other proposition.

The justification for Bayes' theorem just goes through the commutativity of conjunction and the definition of conditional probability:

$$\frac{Pr(e|h)\,Pr(h)}{Pr(e)} = \frac{\dfrac{Pr(e \wedge h)}{Pr(h)} \cdot Pr(h)}{Pr(e)}$$

$$= \frac{Pr(e \wedge h)}{Pr(e)} \overset{Lec}{=} \frac{Pr(h \wedge e)}{Pr(e)}$$

Bayes' theorem when the hypothesis implies the evidence

One traditional application of Bayes' theorem is when $h \vdash e$, or equivalently when $h \to e$ is a tautology (see the Deduction Theorem). This is the situation we are often in when we are trying to test a scientific hypothesis: we deduce that if the hypothesis h is true, then e is true. Further, there being a deduction of $h \to e$ implies that $e \wedge h$ is equivalent to e, as one can quickly check:

exercise

```
(h → e) ⊢ (h → (h ∧ e))

    1. h→e :assumption
```

⊤ ⊢ ((h ∧ e) → e)

1.

Hence if $h \to e$ is a tautology, then $Pr(e \wedge h) = Pr(h)$ by the connections between probability and logic, and thus

If $h \to e$ tautology

then $e \wedge h$ is equivalent to h

and so $Pr(h|e) = \dfrac{Pr(e \wedge h)}{Pr(h)} = \dfrac{Pr(h)}{Pr(h)} = 1$

This implies that if $h \to e$ is a tautology, then the conditional probability is just the fraction of the prior probability over the probability of the evidence:

$$\text{IF} \quad h \rightarrow e \quad \text{tautology}$$

$$\text{then} \quad Pr(h|e) = \frac{Pr(h)}{Pr(e)}$$

Further, recall from the connections between probability and logic that if $h \rightarrow e$ is a tautology, then $Pr(h) \leq Pr(e) \leq 1$.
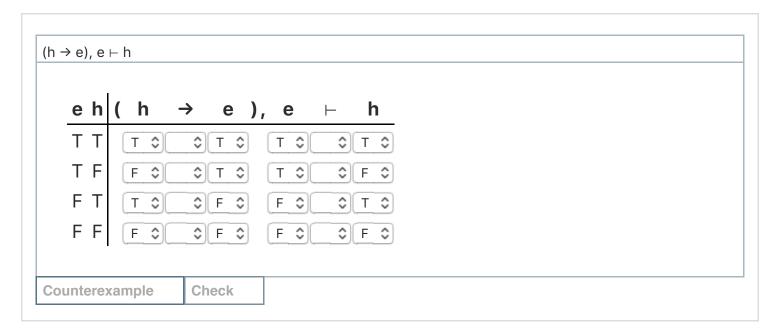
This has two immediate consequences:

- If $Pr(e)$ is as low as possible, namely close to $Pr(h)$, then $Pr(h|e)$ is close to one and so high as possible. Hence, updating on evidence e whose probability is as low as possible makes the conditional probability $Pr(h|e)$ as high as possible.
- If $Pr(e)$ is as high as possible, namely close to 1, then $Pr(h|e)$ is close to $Pr(h)$ and thus as low as possible. Hence, updating on evidence e whose probability is as high as possible makes the conditional probability $Pr(h|e)$ as low as possible.

These might initially seem counterintuitive. But consider that a hypothesis can deductively imply many things: if it deductively implies e then it deductively implies $e \lor f$. Hence, deducing a proposition of high probability should not be a mark of distinction, since high probability are kind of like near-tautologies and almost as easy to come by as tautologies. By contrast, if a hypothesis implied an observation which was startling because it is unlikely, we would be more interested in the hypothesis. It is of course a further step to argue that we ought thereby come to believe such a hypothesis to a stronger degree. This kind of application of probability to claims about what we ought to believe gets studied in the setting of *epistemology,* as well as *confirmation theory* within the *philosophy of science.* (The name comes from the idea that the evidence is confirming the hypothesis by making it more likely).[1]

Revisiting affirming the consequent

Recall the fallacy of affirming the consequent:

$(h \rightarrow e), e \vdash h$

e	h	(	h		→		e	),	e		⊢		h
T	T		T			T		T			T		T
T	F		F			T		T			F		F
F	T		T			F		F			T		T
F	F		F			F		F			F		F

Counterexample   Check

This is a paradigmatic example of an invalid argument. But we noted that invalid arguments can be good in other ways. For instance, when take

> h = it is raining

> e = the sidewalks are wet

then the argument actually seems very reasonable: since you know that rain makes the sidewalks wet, if you observed the sidewalks being wet, you might think it likely to be true that it is raining. There is no guarantee of truth here since the sprinklers may have been on and since you may not have been able to see the sky (imagine you are looking out a far away window and your view of the sky is obstructed). The assessment that the truth of e makes h more likely is predicted by the apparatus of conditional probability. For, it is something like a tautology that rain makes the sidewalks wet-- this has something to do with the meaning of "rain" and "sidewalks" and knowledge of the environment we live in. Then one has that $P(h|e)$ is $P(h)$ divided by $P(e)$. And if e.g. $Pr(h) = 1/3$ and $Pr(e) = 2/3$ then $Pr(h|e) = 1/2$, so that the probability of h conditional on evidence e is higher than the prior probability of the hypothesis. Hence, probability -- a slight extension of truth-tables-- allows us to see why the paradigmatic example of an invalid argument might help raise our estimation in the truth of its conclusion.

Revisiting the paradoxes of the material conditional

Stated in terms of proofs, one of the [paradoxes of the material conditional](#) which C.I. Lewis drew attention to were:

```
exercise

 q ⊢ (p → q)

   1. q :assumption
```

C.I. Lewis thought this was implausible because of

> q = New Year's is coming

> p = Rome is still burning

Maybe [Grice's notion of conversational implicature](#) has some role to play here. For instance, maybe the idea would be that when the speaker says $p \rightarrow q$, they conversationally implicate that they do not believe the deductively stronger q. But that cannot be the whole story. For we all believe that New Year's is coming but few of us believe "if Rome is still burning then New Year's is coming."

One explanation of this is to note that since $q \vdash p \rightarrow q$, then by the [the connections between probability and logic](#) we have:

- $Pr(q) \leq Pr(p \rightarrow q)$

Again, this seems implausible if probability is degrees of belief. For, again, we all like that it is likely that New Year's is coming but few think it likely that "if Rome is still burning then New Year's is coming."

However, perhaps what has gone wrong is that we have simply misidentified the meaning of "if . . . then . . . " Perhaps when we are asserting "if p then q" we are not asserting the truth (or high probability) of p->q, but rather we are asserting that the probability of p conditional on q is high.

And then one can note that it is not in general true that:

- $Pr(q) \leq Pr(q \mid p)$

For instance, if $Pr(p) = 3/4$ and $Pr(q) = 3/4$ and $Pr(p \wedge q) = 1/4$, then $Pr(q) = 3/4$ which is greater than $Pr(q \mid p) = (1/4) \times (4/3) = 1/3$. Hence if assertions of "if . . . then . . . " statements were expressions of conditional probability, then we would not have that q makes "if p then q" likely. Exploring this proposal is an active area within contemporary philosophy and logic.[2]

These are lecture notes written for this course.[3]

---

1. See Earman, John. Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory. Cambridge, MA: MIT Press, 1992, or Pettigrew, Richard & Weisberg, Jonathan (eds.) (2019). The Open Handbook of Formal Epistemology.↩

2. See Edgington, Dorothy. 2006. "The Pragmatics of the Logical Constants." In The Oxford Handbook to the Philosophy of Language, edited by Ernest Lepore and Barry Smith, 768–93. Oxford University Press.↩

3. It is run on the Carnap software, which is↩