Ingo Rohlfing: Multiple non-exclusive hypotheses in Bayesianism

2020-06-24 21:57:53

In this post, I briefly illustrate for the discrete version of Bayes' theorem the implications of working with multiple non-exclusive hypotheses. I will explain what the specific issue with non-exclusive hypotheses is, how to take it into account when specifying the theorem and what the scope of the problem is.

The post is only a minor extension of what Sheryl Zaks has written about in two articles(2017, 2020). Based on an email exchange I just had with someone about this, I thought it might be useful to put up a short post on this.topic.

The issue with non-exclusive hypotheses

The toy example of Bayes theorem in the *process tracing* literature uses two exclusive hypotheses and one piece of evidence or a single body of evidence E. This is totally fine for introducing the theorem to someone who is not familiar with the theorem (it was for me). However, the toy example is as far as away from actual process tracing because we usually have more than two hypotheses; the hypotheses are not exclusive, at least not all of them; we have many more pieces of evidence Abell (2009) discusses this for more pieces of evidence.

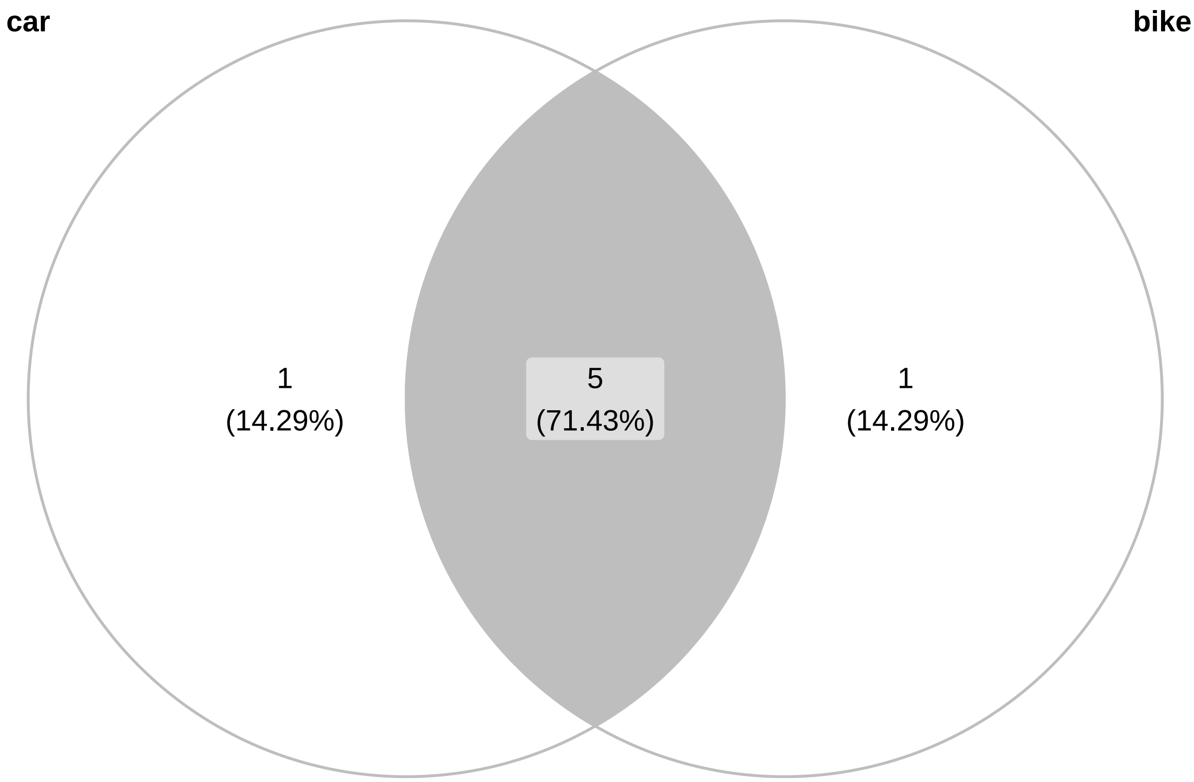
Concerning the hypotheses, the *priors* for the hypotheses need to add up to 1. With a substantive hypothesis H_A and a null hypothesis H_0, this is a piece of cake because $p(H_A) = 1 - p(H_0)$. (Of course, coming up with the priors in an empirical analysis is hard.) In practice, you can always only specify two hypotheses. Imagine the outcome is the position change of a party. H_A could be that the party responds to changes in the position of the median voter. A stands for 'median voter change' and H_A for the corresponding hypothesis. H_0 would be the negation and say that it is not the median voter change that makes a party move. This is a possible way of theorizing, but neither very informative nor elegant. H_0 includes the possibility that it is neither the median voter nor anything else that explains party behavior, or that it is not the median voter, but another cause B that makes parties move. This is way it has been recommended to break up H_0 and put forward multiple, substantively informative hypotheses. (In chapter 8 of my case study book and by Fairfield/Charman, for example.)

What happens now when we add a second variable B that is position change of the *party supporters*? In this case, it makes most no theoretical sense to assume that an effect the median voter and of supporters are exclusive and that these are the only two possible explanations for party position change. It could be that both have an effect, or only one of the two, or neither of both, which means that the null hypothesis H_0 is also in the game here.

I want to illustrate the implications of two non-exclusive hypotheses plus the null hypothesis with a simple example that might be more intuitive than priors that we attach to hypotheses. Suppose you have data for 10 people and want to know how many own a car (A) or a bike (B). To make the example closer to the priors you need for Bayes' theorem, let's assume you want to know the probability that a randomly sampled person is a car owner or a bike owner.

The data is as follows: 5 people own a car and a bike; 1 person owns a car; 1 person owns a bike; 3 people own neither a car nor a bike. In a, say, naive analysis, you would once count the number of car owners (6) and the number of bike owners (6). This gives a sum of 12, meaning the probability of randomly choosing a car owner or a bike owner is 120%! The problem obviously is that you count the people who own a car and a bike twice. The Venn diagram below illustrates the issue and also points at ways how it can be addressed. The three people who neither have a car nor a bike are not explicitly displayed, but can be imagined as being located outside of the 'car' set and the 'bike' set.

```
library(ggVennDiagram) # available on CRAN
library(ggplot2)
carbike <- list(car = 1:6, bike = 2:7)
ggVennDiagram(carbike) +
  theme(legend.position = "none") + # remove legend
  scale_fill_gradient(low = "white", high = "gray") # reconfigure coloring
```



Specifying non-exclusive hypotheses

There are two ways to specify the priors when two hypotheses are non- exclusive. First, you can specify the marginal probability for each hypothesis and subtract the joint probability of the non-exclusive hypothesis (as discussed by Zaks): $p(H_A) + p(H_B) - p(H_A \cap H_B)$. If you add $p(H_0)$ to this, you would get 1 as a result. With regard to the car-bike example, this means your calculation would be 6+6-5 = 7 and a correct chance of 70% to randomly choose a car owner or a bike owner.

A second possibility is to "exclusivy" each prior (maybe there is a better term for this, I don't know.) The problem with the naive approach is to ignore that $p(H_A) = p(H_A \cap H_B) + p(H_A \cap \neg H_B)$ and that $p(H_B) = p(H_B \cap H_A) + p(H_B \cap \neg H_A)$ because of which $p(H_A \cap H_B)$ is counted twice. To "exclusify" the priors, you decompose the total probabilities $p(H_A)$ and $p(H_B)$ into the exclusive components:

- $p(H_A \cap \neg H_B)$ (corresponds to the 1 car owner out of 10 people)
- $p(H_B \cap \neg H_A)$ (corresponds to the 1 bike owner out of 10)
- $p(H_A \cap H_B)$ (corresponds to the 5 car and bike owners out of 10)
- $p(\neg H_A \cap \neg H_B)$ (corresponds to the 3 people out of 10 without a car and a bike)

I agree with Sheryl Zaks that the specification of priors is more challenging than one might believe because most examples (methodological and substantive) use the toy example with two exclusive hypotheses. (And I also wouldn't accept "Priors do wash out." as a counterargument.) For me, exclusifying the priors might be relatively easier because they represent the different theoretical arguments one can make better than the other first approach.

- $p(H_A \cap \neg H_B)$ means that only hypothesis A is expected to be correct;
- $p(H_B \cap \neg H_A)$ that only hypothesis B is correct;
- $p(H_A \cap H_B)$ that both hypotheses are correct and both variables (or conditions or mechanisms) play a role;

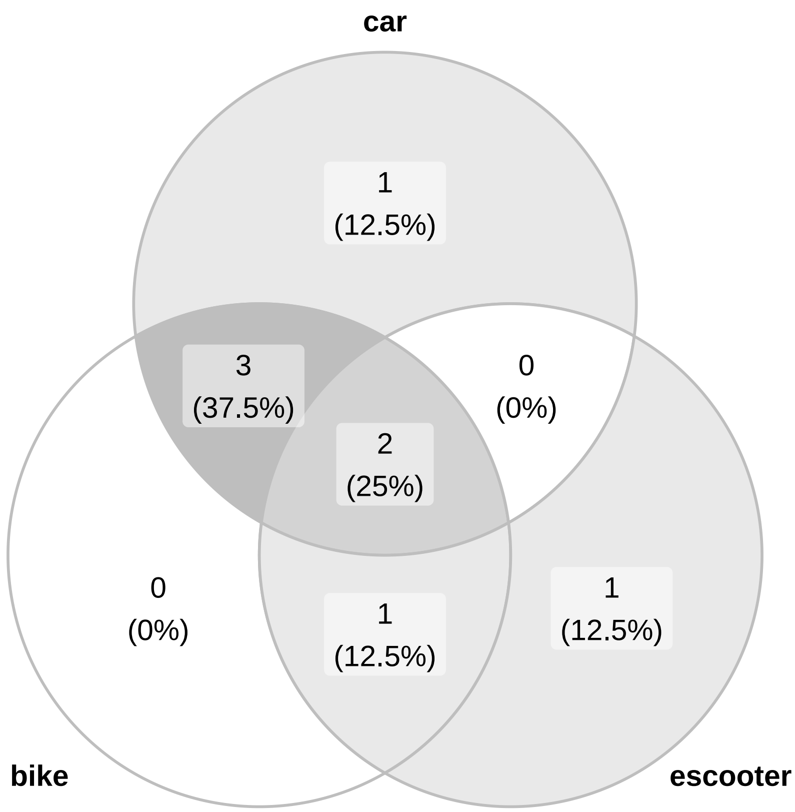
- $p(\neg H_A \cap \neg H_B)$ captures that neither might be correct and H_0 is true.

Now one might say: I am not interested in the estimation part of a Bayesian analysis (the posteriors for theoretical arguments), but in the likelihoods and the testing part where one assesses the strength of evidence (which has been the main focus in the process tracing literature). I don't need priors for this. True, one does not need priors, but the issue is the same here: When you want to make arguments about the evidential strength of an observation for the hypothesis $H_A \cap \neg H_B$, you need to compare the likelihood for $H_A \cap \neg H_B$ against $H_A \cap H_B$, and separately against $H_B \cap \neg H_A$ and against $\neg H_A \cap \neg H_B$. The hypothesis $H_A \cap \neg H_B$ might receive stronger support in one comparison, but less support in a second comparison, which you can only find out when calculating three likelihood ratios. (For theoretical reasons, one might reduce the number of comparisons, but right now I think it is better to take a more comprehensive perspective.)

The scale of the issue

The simple car-bike example indicates that the challenges of prior specification can escalate quickly. Moving from the toy example to two non-exclusive hypotheses means that we have to specify four priors instead of two; we have eight priors when we add a third non-exclusive hypothesis. You might see where this is going: The number of priors increases *exponentially* with 2^H, where H is the number of substantive hypotheses (meaning H_0 is excluded). If some, but not all hypotheses are exclusive, the number is somewhere between the number of hypotheses, including H_0, and H.
Let us add one type of ownership to the example to see how the Venn diagram gets larger and the number of combinations increases exponentially. We add a third type 'e-scooter' ownership and are now interested in the probability of randomly sampling a person who either owns a car, or a bike or an e-scooter.

```r
carbikeesc <- list(car = 1:6, bike = 2:7, escooter = c(2:3, 7:8))
ggVennDiagram(carbikeesc) +
  theme(legend.position = "none") +
  scale_fill_gradient(low = "white", high = "gray")
```



We now have seven types of ownership constellations plus the possibility that someone owns neither of the three devices. A naive analysis would again add the total number of owners per device, which is 16. A proper analysis would subtract the number of owners that we counted multiple times from 16: the car-bike owners (5), the car-bike-escooter owners (2) and the bike-escooter owners (3). This would tell us that 8 people own some device and that the sampling probability is 0.8.

All this certainly does *not* invalidate Bayesian process tracing as an approach and its use for empirical research. I only mean to show for one selected element what it requires to do process tracing that complies with Bayes' theorem. It quickly becomes demanding when we work with non-exclusive hypotheses that are, in my view, the rule in the social sciences.