ADS2 Problem set: Bayesian Inference

ADS2 (based on MI Stefan)

Semester 2, 2023/24

We expect this problem set to take around one hour to complete. But professors are sometimes wrong![citation missing]. If this or future problem sets are too long, please let us know, so we can adjust and plan accordingly.

Please make use of the online discussion board to share your ideas and get help.

In vitro fertilization (IVF) and pregnancy tests

A girl underwent IVF, and she wants to know whether she is pregnant or not. From the previous studies, we know that the chance of conception after IVF is 0.222 (CDC, 2021).

She used an early pregnancy detection kit that gives the correct answer 95% time if the person is pregnant and 80% time if the person *is not* pregnant. That is: it detects the elevated level of hCG in 95% of cases and does not react below the threshold value in 80% of cases. Answer the following questions:

1. What is the chance that a girl with a positive test is really pregnant?
    - Is this result convincing?
2. What is the chance that a girl with a *negative* test is still pregnant?
3. Suppose, the girl got a positive result. Did her beliefs get updated?
4. She decided to be double sure. After some time, she used the same test once again, and she got a positive result *once again*. What would be the chance that she is pregnant given she got positive results *twice*?
5. What can you say about this test kit? Is it good? Is it specific? Is it sensitive? Support your claims with some calculations. **Note:** use odds and odds ratios to support your words.

A dicy problem - Part 2

How did you tackle the dice problem in the practical?

Here is some of the thinking I did before starting the coding:

We know that the friend threw 7 out of 20 sixes. That means that their die definitely does not have either 0 or 6 sixes. So we are left with five hypotheses:

- H_1: Their die has 1 six
- H_2: Their die has 2 sixes
- ...
- H_5: Their die has 5 sixes

Ultimately, we want to find the most probable hypothesis given the data we have, so we want to find the i that maximises $P_{(H_i|Data)}$.

We can easily (well, quite easily) compute $P_{(Data \mid H_i)}$ for each i, this is just a probability exercise.

We don't have $P_{(Data)}$, but does it matter? Not if all we want to do is find the most probable hypothesis. $P_{(Data)}$ does not depend on the hypothesis chosen, so it's just a constant factor. The maximum of $P_{(H_i|Data)}$ is also the maximum of $P_{(H_i|Data)} \times P_{(Data)}$. So if all we want to do is find the max, we can just go ahead and ignore $P_{(Data)}$.

What about priors though? This is what this exercise is about!

Here are three students with very different approaches to constructing priors.

Aditi: My experience in the world suggests that most playing dice have exactly one 6. In fact, probably all dice I have ever seen had exactly one 6. So $P(H_1)$ would be close to 1. So, I set $P(H_1)$ to 0.99 and $P(H_2) = P(H_3) = P(H_4) = P(H_5) = 0.0025$.

Bo: We know nothing about this die, and dealing with different priors is complicated, so we may as well assume $P(H_i) = 0.2$ for all i $(1 < i < 5)$.

Charlie: I know my friend to be a trickster and I don't trust them. I am 99 % sure the die is not fair, therefore $P(H_1) = 0.01$. I don't know how unfair exactly, so I assume $P(H_2) = P(H_3) = P(H_4) = P(H_5) = 0.2475$.

Does the choice of prior influence the end result?

Also, why do the priors for each person have to add up to 1?

And why does nobody choose priors of exactly 0 or exactly 1? What would happen if they did?

Updated by Dmytro Shytikov in 2024.