Esssentials of Applied Data Analysis IPSA-USP Summer School 2017

Handout - The Basics of Set Theory

Leonardo Sangali Barone
leonardo.barone@usp.br

jan/17

Bayes Rule

Suppose there are A_1, A_2,..., A_n events and $A_1 \cup A_2 \cup ... \cup A_n$ = S. A_i events are mutually exclusive. Suppose there is another event B. For each A_i, B has a different joint conditional probability $P(A_i|B)$. B can be described as:

$$P(B) = P(A_1) * P(B|A_1) + P(A_2) * P(B|A_2) + ... + P(A_n) * P(B|A_n)$$

The probability that an specific A_i will occur given B is defined as:

$$P(A_i|B) = \frac{P(A_i) * P(B|A_i)}{P(A_1) * P(B|A_1) + P(A_2) * P(B|A_2) + ... + P(A_n) * P(B|A_n)}$$

Bayes' rule - a first example

There are two political parties, A_1, with 60 members and A_2, with 40 members. 80% of party A_1 members are against the country engaging in a new war. In party A_2, one third of the members are against the new war, but the others are in favor of it. nW is the event of being **against** enganging in war.

You turn on the radio and there is a politician speaching against war. You don't know her and you want to figure out her party. What is the probability that she belongs to party A_1 given she is against war? Or party A_2?

Let's think about the problem for a second. Whithout any calculations, is she more likely to belong to which party? Make your guess.

What is the probability that a a politician at random belongs to party A_1? And party A_2?

$$P(A_1) = \frac{40}{100} = \frac{2}{5}$$
$$P(A_2) = \frac{60}{100} = \frac{3}{5}$$

Or

$$P(A_2) = 1 - P(A_1) = 1 - \frac{2}{5} = \frac{3}{5}$$

So without knowing that she is against the war, it is more likely that she belgons to party A_1, right? $2/5$ for A_1 and $3/5$ for A_2 would be a good first guess. But we do know she is against war and we can **update** our "guess" based on this information.

What is the probability that a politician from A_1 is against war?

$$P(nW|A_1) = \frac{4}{5} = 0.8$$

And from A_2?

$$P(nW|A_2) = \frac{1}{3}$$

Since we know that being or not in favor of war is conditioned on the party, we can update our guess using Baye's rule.

$$P(A_1|nW) = \frac{P(A_1) * P(nW|A_1)}{P(A_1) * P(nW|A_1) + P(A_2) * P(nW|A_2)} = \frac{\frac{3}{5} * \frac{4}{5}}{\frac{3}{5} * \frac{4}{5} + \frac{2}{5} * \frac{1}{3}} =$$

$$= \frac{\frac{12}{25}}{\frac{12}{25} + \frac{2}{15}} = \frac{\frac{12}{25}}{\frac{36+10}{75}} = \frac{\frac{12}{25}}{\frac{46}{75}} = \frac{12}{25} * \frac{75}{46} = \frac{36}{46} = \frac{18}{23} \simeq 0.7826 > P(A_1) = 0.6$$

So $P(A_1|nW) > P(A_1)$, our initial guess (which makes sense, since most of party A_1 members are against war). Let's look at what happens with A_2:

$$P(A_2|nW) = \frac{P(A_2) * P(nW|A_2)}{P(A_1) * P(nW|A_1) + P(A_2) * P(nW|A_2)} = \frac{\frac{2}{5} * \frac{1}{3}}{\frac{3}{5} * \frac{4}{5} + \frac{2}{5} * \frac{1}{3}} =$$

$$= \frac{\frac{2}{15}}{\frac{12}{25} + \frac{2}{15}} = \frac{\frac{2}{15}}{\frac{36+10}{75}} = \frac{\frac{2}{15}}{\frac{46}{75}} = \frac{2}{15} * \frac{75}{46} = \frac{10}{46} = \frac{5}{23} \simeq 0.2173 < P(A_2) = 0.4$$

As expected, $P(A_2|nW) < P(A_2)$, since the minority of party A_2 is against war and the politian was speaching against war.

Before we move on, think or a second about what we just did. We wanted to guess the political party of a politician and we had an expectation about the distribution of politicians within political parties ($A_1 = 0.6$ and $A_2 = 0.4$). We collected new information (data!!!!) about this specific politician and we updated our believes about what political parties she *is more likely* to belong to.

Bayes' rule - a textbook-like example

There are 3 boxes, each containing 10 colored balls. In the first box (A_1) there are 8 blue balls and 2 red balls. In the second box (A_2) there are 3 blue balls and 7 red balls. There are only red balls in the third box (A_3). You have randomly chosen a box and pick up randomly one ball from it. The ball was blue ($Blue$). What is the probability that you got the first box (A_1)?

$$P(A_1|B) = \frac{P(A_1) * P(Blue|A_1)}{P(A_1) * P(Blue|A_1) + P(A_2) * P(Blue|A_2) + P(A_3) * P(Blue|A_3)}$$

$$= \frac{1/3 * 8/10}{1/3 * 8/10 + 1/3 * 3/10 + 1/3 * 0} = \frac{8/30}{8/30 + 3/30 + 0} = \frac{8}{11} = 0.73$$

What is the probability that I got the second box (A_2)?

$$P(A_2|Blue) = \frac{P(A_2) * P(Blue|A_1)}{P(A_1) * P(Blue|A_1) + P(A_2) * P(Blue|A_2) + P(A_3) * P(Blue|A_3)}$$

$$= \frac{1/3 * 3/10}{1/3 * 8/10 + 1/3 * 3/10 + 1/3 * 0} = \frac{3/30}{8/30 + 3/30 + 0} = \frac{3}{11} = 0.27$$

And the third box (A_3)?

$$P(A_3|Blue) = \frac{P(A_2) * P(Blue|A_1)}{P(A_1) * P(Blue|A_1) + P(A_2) * P(Blue|A_2) + P(A_3) * P(Blue|A_3)}$$

$$= \frac{1/3 * 0}{1/3 * 8/10 + 1/3 * 3/10 + 1/3 * 0} = 0$$

Bayes' rule - another way to think

Another to think about Bayes rule is as a theorem that explains how we ought to change our's initial/hypothetical/subjective probability in response to empirical evidence. Let's rearrenge our notation:

T = theory or hypothesis

E = represents a new piece of evidence that seems to confirm or disconfirm the theory.

P(T) = Prior Belief = Probability of (T). We can think of it as the expectation that our theory is correct **BEFORE** empirical investigation.

P(T—E) = Posterior Probability. We can think of it as the expectation that our theory is correct **AFTER** empirical investigation.

Then, we can use Bayes rule to compute the Posterior Probability of T being correct:

$$P(T|E) = \frac{P(T) * P(E|T)}{P(T) * P(E|T) + P(T^c) * P(E|T^c)} =$$