3.2 Bayes' Rule

Monday, 10 October 2022 12:56

Summary	• Bayes' Rule

- Prior probability
- Posterior probability
- Where does Bayes' rule fit into this?

Baye's rule

- Often we have initial guesses about an event from which we can calculate prior probabilities, using the usual probability theory.
- Then, from sources such as data collection, sample, product field tests, we obtain more information about these events.
- Given, this new information, we can update our prior beliefs by calculating revised probabilities – this is called the posterior probability.
- Baye's rule is used to calculate the posterior probability if we have the initial belief (probability) and the additional sample information.

Baye's rule

- Suppose that a manufacturer receives same raw material from two different suppliers *S1* and *S2*.
- Currently 65% of the raw material comes from *S1* and remaining, 35%, comes from *S2*.
- Also, suppose that from the historical data available with the quality assurance department, we know that *S1* has 98% of the supplied raw material of good quality and *S2* has 95% of the raw material of good quality.
- That is, the probability of a "Good" quality raw material given that the supplier is *S1* is, *Pr(G|S1) = 0.98*. And for the second supplier, this probability is: *Pr(G|S2) = 0.95*.

65% of the raw material comes from S1	35% of the raw material comes from S2

	S1	S2	
Good	**98% of S1** P(G | S1)	**95% of S2** P(G | S2)	
Bad	**2% of S1** P(B | S1)	**5% of S2** P(B | S1)	
	P(s1) = **65%**	P(S2) = **35%**	

	S1	S2	
Good	0.98×0.65	0.95×0.35	
Bad	0.02×0.65	0.05×0.35	
	0.65	0.35	

Baye's rule

- What is the probability of the raw material being supplied by *S1* and it being good?
- Joint probability, of course!
- This can be calculated using the Baye's formula.

*Pr(S1, G) = Pr(S1 ∩ G) = Pr(S1)*Pr(G|S1) = 0.65*0.98 = 0.637*

*Pr(S2, G) = Pr(S2 ∩ G) = Pr(S2)*Pr(G|S2) = 0.35*0.95 = 0.3325*

- Now, knowing all this information so far, suppose the manufacturer inspects the incoming raw material on receipt and finds a bad quality material.
- He wants to know the supplier who needs to be contacted to complain!

	S1	S2	
Good	0.637	0.3325	0.9695
Bad	0.013	0.0175	0.0305
	0.65	0.35	**1**

- State Bayes' Rule

Baye's rule

- We are interested in the posterior probability that a particular supplier is guilty of supplying bad quality product *given that* we have bad quality raw material at our doorstep – *Pr(S1|B)* or *Pr(S2|B)*.
- This is an application of Baye's theorem – finding posterior probability given some initial facts and numbers.
- From Baye's formula we know that:

$$\Pr(S1 \mid B) = \frac{\Pr(S1 \cap B)}{\Pr(B)}$$

$$\Pr(S1) * \Pr(B \mid S1)$$

$$= \frac{\quad}{\Pr(B)}.$$

	S1	S2	
Good	0.98×0.65 = 0.637 **P(G | S1) × P(S1)**	0.95×0.35 = 0.3325 **P(G | S2) × P(S2)**	0.9695 **P(G)**
Bad	0.02×0.65 = 0.013 **P(B | S1) × P(S1)**	0.05×0.35 = 0.0175 **P(B | S2) × P(S2)**	0.0305 **P(B)**
	0.65 **P(S1)**	0.35 **P(S2)**	**1**

- We have this contingency table, which is prior info.
- Now we found a bad quality material - this is the additional sample info.
- We want to know given a bad quality material (additional info) what is the probability that it came from supplier S1 (or S2)? - Posterior probability
- $P(S1 \mid B) \times P(B) = P(B \mid S1) \times P(S1)$

$$P(S1 \mid B) = \frac{P(B \mid S1) \times P(S1)}{P(B)}$$

Bayes' Rule

$$\mathbf{P(X \mid Y) = \frac{P(Y \mid X) \times P(X)}{P(Y)}}$$

Baye's rule

- What is *Pr(B)*?
- That is the probability of receiving a bad quality raw material.
- Now bad quality raw material can from supplies of *S1* or *S2*.
- That is, the event *B* can occur with *S1* or with *S2*.

Pr(B) = Pr(S1 ∩ B) + Pr(S2 ∩ B)

- But *Pr(S1 ∩ B) = Pr(S1)*Pr(B|S1)*, and
- *Pr(S2 ∩ B) = Pr(S2)*Pr(B|S2)*

$$\Pr(S1 \mid B) = \frac{\Pr(S1)^* \Pr(B \mid S1)}{\Pr(S1)^* \Pr(B \mid S1) + \Pr(S2)^* \Pr(B \mid S2)}.$$

- $P(B) = P(S1 \cap B) + P(S2 \cap B)$

$P(B) = P(B \mid S1) \times P(S1) + P(B \mid S2) \times P(S2)$

- Observe how probabilities change with the prior and posterior.

Baye's rule

$$\Pr(S1 \mid B) = \frac{\Pr(S1)^* \Pr(B \mid S1)}{\Pr(S1)^* \Pr(B \mid S1) + \Pr(S2)^* \Pr(B \mid S2)}$$
$$= \frac{0.65 * 0.02}{0.65 * 0.02 + 0.35 * 0.05} = 0.426.$$
$$\Pr(S2 \mid B) = 0.574.$$

- Significance: Find posterior probabilities using prior information.
- Notice that we use *Pr(B|S1)* to find *Pr(S1|B)*.

	S1	S2	
Good	0.98×0.65 = 0.637 **P(G | S1) × P(S1)**	0.95×0.35 = 0.3325 **P(G | S2) × P(S2)**	0.9695 **P(G)**
Bad	0.02×0.65 = 0.013 **P(B | S1) × P(S1)**	0.05×0.35 = 0.0175 **P(B | S2) × P(S2)**	0.0305 **P(B)**
	0.65 **P(S1)**	0.35 **P(S2)**	**1**

- So, Prior was:
  - P (a randomly picked item was from S1) = 65%
  - P (a randomly picked item was from S2) = 35%
- Posterior:
  - P (a randomly picked item was from S1 | given that the item was bad) = 42.6%
  - P (a randomly picked item was from S2 | given that the item was bad) = 57.4%

	• So we have found association between two categorical variables: 1. Supplier (s1 & S2), and 2. Quality (good and bad)