Everyday Bayes

Bayes Theorem is vital for inference in all realms - not just scientific data analysis.

For example, the "What is the probability it will rain" problem in the HomeWork.

Bayes Theorem can also be applied to cases of determining the probability that a patient has a disease given an outcome of a medical test, or the probability of guilt or innocence of a suspect in jurisprudence.

Bayesian Spam Filtering

You have developed a new spam filter.

It is 100% accurate when it encounters spam!

But it is a little over-cautious and 5% of the time it results in a false positive, identifying a valid email as spam.

Your client is internet-savvy and is careful about distributing her email address. As such, she still gets spam, but only in 1 out of 1000 emails.

Given that your algorithm has identified an email as spam, what is the probability that it really is spam?

S - spam
$\sim S$ - not spam
Hypotheses

$+$ - filter result is positive
$-$ - filter result is negative

$P(S \mid I) = 0.001$

$P(+ \mid S) = 1.00$

$P(+ \mid \sim S) = 0.05$

$$P(S \mid +, I) = \frac{P(S \mid I) \cdot P(+ \mid S, I)}{P(S \mid I) \cdot P(+ \mid S, I) + P(\sim S \mid I) \cdot P(+ \mid \sim S, I)}$$

$$= \frac{0.001 \cdot 1.00}{0.001 \cdot 1.00 + 0.999 \cdot 0.05} = 0.0196$$

$P(S \mid +, I) \sim 2\%$ ONLY 2% of the time is an email that has been identified as spam actually spam!

Bayes in Medicine

You have just donated blood and the blood test has come back HIV positive.

One out of 1000 people have HIV.
If you have HIV, the blood test will detect it.
If you dont have HIV, the blood test will come back positive 5% of the time.

THIS IS A MADE-UP PROBLEM
DONT BASE A MEDICAL DECISION ON THESE RESULTS

Hypotheses	Data
HIV	+ - test results
~HIV	− - test result

$$P(HIV \mid I) = 0.001$$

$$P(+ \mid HIV, I) = 1.00$$

$$P(+ \mid \sim HIV, I) = 0.05$$

$$P(HIV \mid +, I) = \frac{P(HIV \mid I)\ P(+ \mid HIV, I)}{P(HIV \mid I) \cdot P(+ \mid HIV, I) + P(\sim HIV \mid I) \cdot P(+ \mid \sim HIV, I)}$$

$$= \frac{0.001}{0.001 + 0.999 \cdot 0.05} = 0.0196$$

$\sim 2\%$

WORKING WITH PROBABILITIES CAN BE NON-INTUITIVE

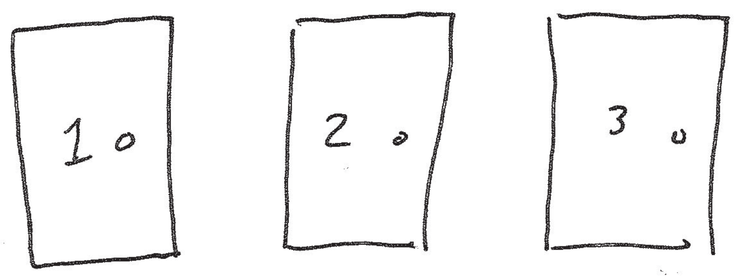
Three Doors

You are on a famous game show with Monty Hall (a 1970's game show host).

There are three doors.

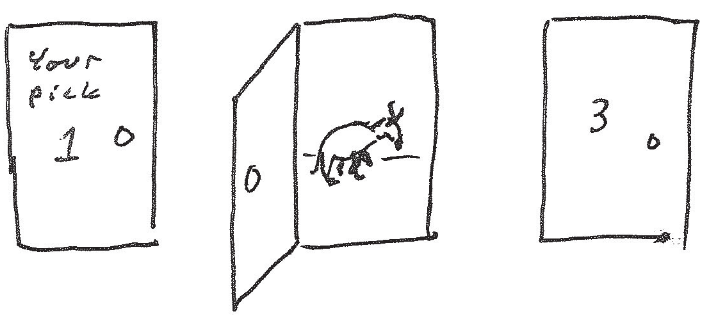
Behind one door is the prize... a brand new car!

Behind the other two doors are goats... not a prize.



You are asked to choose a door... call it door # 1.

Monty Hall then shows you what is behind another door... it is a goat.



Monty Hall now asks if you would like to change doors.

WHAT SHOULD YOU DO?

Three Doors ... The Solution

Prior Probabilities

$$P(1|I) = P(2|I) = P(3|I) = 1/3$$

where $P(1|I)$ is the prior probability that the prize is behind door 1.

When Monty Hall shows you Door 2 ... that is data (information)

He could either have opened door 2 or door 3, but only if it had a goat. Let O2 = opened 2 and O3 = opened 3

$P(O2|2,I) = 0$ He would not open door 2 if the prize was behind door 2

$P(O2|3,I) = 1$ If the prize was behind door 3, he could have only opened door 2.

$P(O2|1,I) = 0.5$ If the prize was behind door 1, he could have opened either door 2 or 3.

$$P(1|O2,I) = \frac{P(1|I)\,P(O2|1,I)}{P(1|I)\,P(O2|1,I) + P(3|I)\,P(O2|3,I)}$$

$$= \frac{1/3 \cdot 1/2}{1/3 \cdot 1/2 + 1/3 \cdot 1} = \frac{1/6}{1/2} = 1/3$$

$$(3|O2,I) = \frac{P(3|I)\,P(O2|1,I)}{P(1|I)\,P(O2|1,I) + P(3|I)\,P(O2|3,I)} = \frac{1/3}{1/2} = 2/3$$

SWITCH DOORS! He didn't open #3 => INFORMATION

Two Children I

I have two children.

One is a boy.

What is the probability that both are boys?

Notation

B1 - child #1 is a boy

B2 - child #2 is a boy

S1 - child #1

S2 - child #2

$B1 \wedge B2$ - Both children are boys

$(B1 \wedge S2) \vee (S1 \wedge B2)$ - One child is a boy

Goal: Find $P(B1 \wedge B2 \mid (B1 \wedge S2) \vee (S1 \wedge B2), I)$

Bayes Theorem flips things around

$$P(B1 \wedge B2 \mid (B1 \wedge S2) \vee (S1 \wedge B2), I)$$

$$= \frac{P((B1 \wedge S2) \vee (S1 \wedge B2) \mid B1 \wedge B2, I)\, P(B1 \wedge B2 \mid I)}{P((B1 \wedge S2) \vee (S1 \wedge B2) \mid I)}$$

Two Children I

First consider the denominator

$$P((B1 \wedge S2) \vee (S1 \wedge B2) \mid I) =$$

$$P(B1 \wedge S2 \mid I) + P(S1 \wedge B2 \mid I) - P(B1 \wedge B2 \mid I)$$

$$P(B1 \wedge S2 \mid I) = P(B1 \mid S2, I)\, P(S2 \mid I)$$

$$= P(B1 \mid I)\, P(S2 \mid I)$$

$P(B1 \mid I) = P(B1 \mid S2, I)$ Since B1 cannot depend on S2

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

$$= \frac{1}{2}$$

$$P(S1 \wedge B2 \mid I) = \frac{1}{2}$$

$$P(B1 \wedge B2 \mid I) = P(B1 \mid B2, I)\, P(B2 \mid I)$$

$$= P(B1 \mid I)\, P(B2 \mid I)$$

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

$$= \frac{1}{4}$$

$$P((B1 \wedge S2) \vee (S1 \wedge B2) \mid I) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}$$

NOTE: THE $\vee$ IS NOT DISJOINT

This is what makes the problem difficult.

Consider the likelihood

$$P((B1 \wedge S2) \vee (S1 \wedge B2) \mid B1 \wedge B2, I) =$$

$$P(B1 \wedge S2 \mid B1 \wedge B2, I) + P(S1 \wedge B2 \mid B1 \wedge B2, I)$$

$$- P(B1 \wedge B2 \mid B1 \wedge B2, I)$$

$$P(B1 \wedge S2 \mid B1 \wedge B2, I) = P(B1 \mid B1 \wedge B2 \wedge S2, I)\, P(S2 \mid B1 \wedge B$$

$$= P(B1 \mid B1, I)\, P(S2 \mid B2, I)$$

$$= 1$$

$$P(S1 \wedge B2 \mid B1 \wedge B2, I) = 1$$

$$P(B1 \wedge B2 \mid B1 \wedge B2, I) = 1$$

Therefore

$$P((B1 \wedge S2) \vee (S1 \wedge B2) \mid B1 \wedge B2, I) = 1$$

Last

$$P(B1 \wedge B2 \mid I) = P(B1 \mid B2, I)\, P(B2 \mid I)$$

$$= P(B1 \mid I)\, P(B2 \mid I)$$

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

Two Children I

$$P(B1 \wedge B2 \mid (B1 \wedge S2) \vee (S1 \wedge B2), I) = \frac{1 \cdot \frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

The probability that both children are Boys is $\frac{1}{3}$