Bayes’ Rule

Monty Hall Problem

"Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the other doors, opens another door, say No. 3, which has a goat. He then says to you, 'Do you want to pick door No. 2?' Is it to your advantage to take the switch?"

Marilyn vos Savant

Image Source

Image Source

Derivation

Facts

- P(A | B) = P(A and B) / P(B) implies P(A and B) = P(A | B) P(B)
- P(B | A) = P(A and B) / P(A) implies P(A and B) = P(B | A) P(A)
- P(A | B) P(B) = P(B | A) P(A) implies P(A | B) = P(B | A) P(A) / P(B)
- P(A | B) P(B) = P(B | A) P(A) implies P(B | A) = P(A | B) P(B) / P(A)

Rule

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)}$$

In words, if we know one conditional probability, then we can use Bayes' rule to find the other conditional probability.

Derivation

Facts

- P(A | B) = P(A and B) / P(B) implies P(A and B) = P(A | B) P(B)
- P(B | A) = P(A and B) / P(A) implies P(A and B) = P(B | A) P(A)
- P(A | B) P(B) = P(B | A) P(A) implies P(A | B) = P(B | A) P(A) / P(B)
- LOTP: P(B) = P(B and A) + P(B and A^c) = P(B | A) P(A) + P(B | A^c) P(A^c)

Rule

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

In words, if we know one conditional probability, then we can use Bayes' rule to find the other conditional probability.

Worked Example 1

There are 175 million adults in the United States. Suppose some rare disease affects 1 in 100,000 of us. Further suppose that we can screen for this disease. Given that someone tests positive, how likely is it that they have the disease?

Spoiler: Surprisingly low!

Worked Example 1

Assumptions in English

- The disease affects 1 in 100,000.
- If someone has the disease, the screen always flags them.
- If someone does not have the disease, the screen still flags them 1 in 10,000 times.

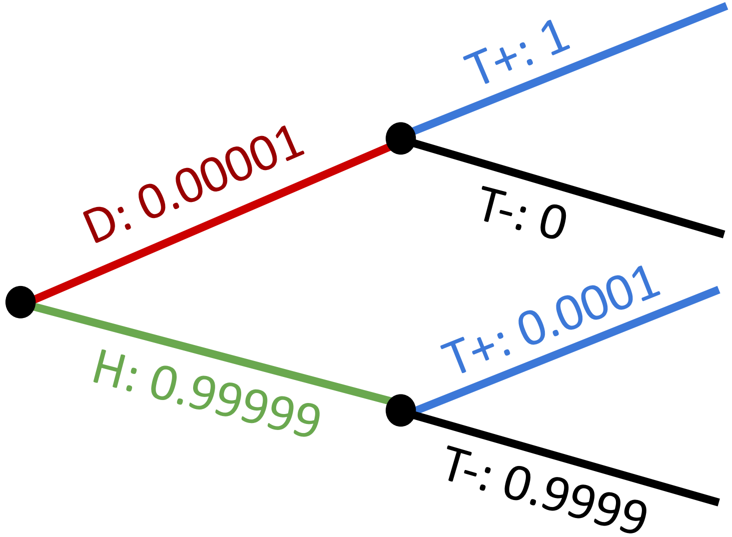
Assumptions as Probabilities

- P(Diseased) = 0.00001
- P(Test Positive|Diseased) = 1
- P(Test Positive|Healthy) = 0.0001

Question

- P(Diseased|Test Positive) = P(Test Positive|Diseased) P(Diseased) / P(Test Positive)

Worked Example 1



P(D and T+) = P(T+ | D) P(D) = 0.00001

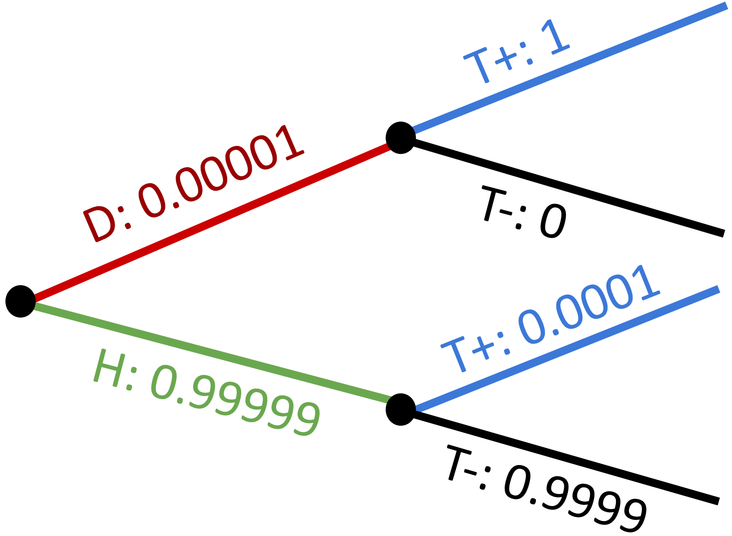
P(D and T-) = P(T- | D) P(D) = 0

P(H and T+) = P(T+ | H) P(H) ~ 0.0001

P(H and T-) = P(T- | H) P(H) ~ 0.99989

P(Test Positive) = P(D and T+) + P(H and T+) = 0.00001 + 0.0001 ~ 0.00011

Worked Example 1



P(D and T+) = 0.00001

P(D and T-) = 0

P(H and T+) ~ 0.0001

P(H and T-) ~ 0.99989

Bayes Rule

P(D | T+)
= P(T+ | D) P(D) / P(T+)
= (1) (0.00001) / 0.00011
~ 0.00001 / 0.00011
~ 0.09

If someone tests positive, they have about a 9% chance of actually having the disease.

Worked Example 2

There are 250 million adults in the United States. The government monitors them all (us!) to detect terrorism. Given that someone is suspected of terrorism, what is the probability that the person is a terrorist? Spoiler: Surprisingly low!

Worked Example 2

Assumptions in English

- There are 250 million adults in the United States, of whom 250 are terrorists.
- If someone is a terrorist, then the government's monitoring system will always flag them.
- If someone is not a terrorist, then the system will flag them anyway 1 in 10,000 times.

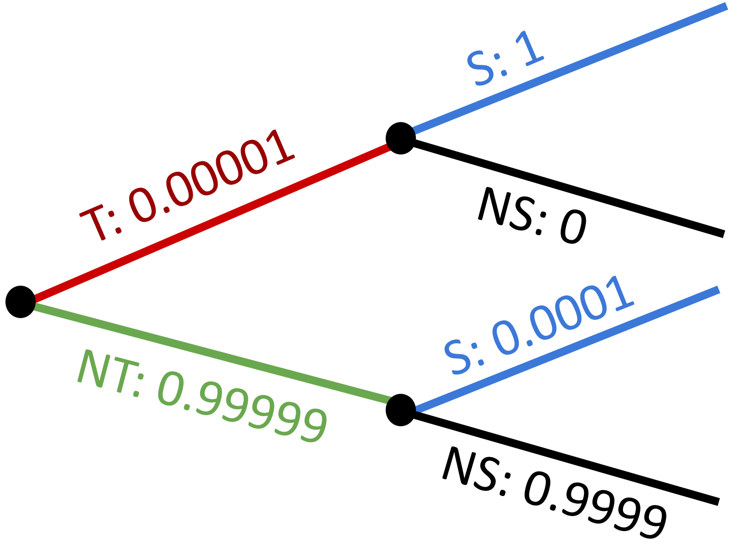
Assumptions as Probabilities

- $P(T) = 0.000001$
- $P(S|T) = 1$
- $P(S|NT) = 0.0001$

Question

- $P(T|S) = P(S|T)\ P(T)\ /\ P(S)$

Worked Example 2



$P(T \text{ and } S) = P(S \mid T)\, P(T) = 0.000001$

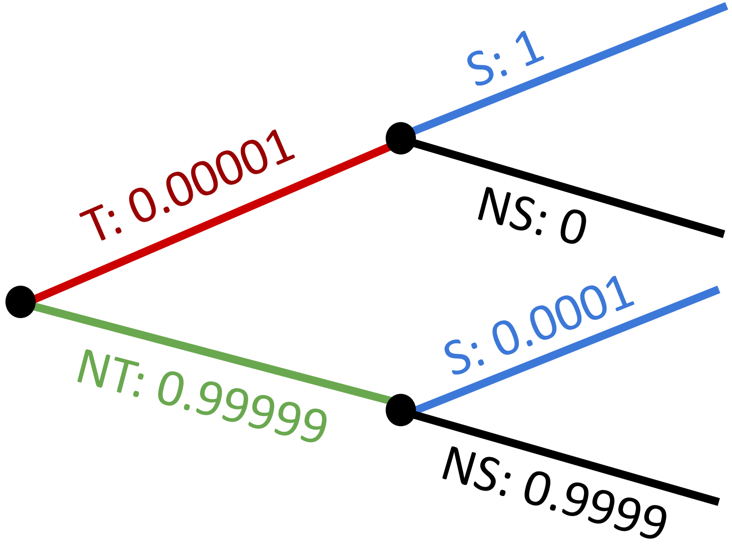
$P(T \text{ and } NS) = P(NS \mid T)\, P(T) = 0$

$P(NT \text{ and } S) = P(S \mid NT)\, P(NT) \sim 0.0001$

$P(NT \text{ and } NS) = P(NS \mid NT)\, P(NT) \sim 0.99989$

$P(S) = P(T \text{ and } S) + P(NT \text{ and } S) = 0.000001 + 0.0001 \sim 0.000101$

Worked Example 2



P(T and S) = 0.000001

P(T and NS) = 0

P(NT and S) = 0.0001

P(NT and NS) = 0.99989

Bayes Rule

P(T | S)
= P(S | T) P(T) / P(S)
= (1) (0.000001) /0.000101
~ 0.000001 / 0.000101
~ 0.01

If someone is suspected of terrorism, there is about a 1% chance they are actually a terrorist.

Worked Example 3

There are 250 million adults in the United States. The government monitors them all (us!) to detect terrorism. Given that someone is convicted of terrorism, what is the probability that the person is a terrorist? Spoiler: Higher, but arguably not high enough!

Worked Example 3

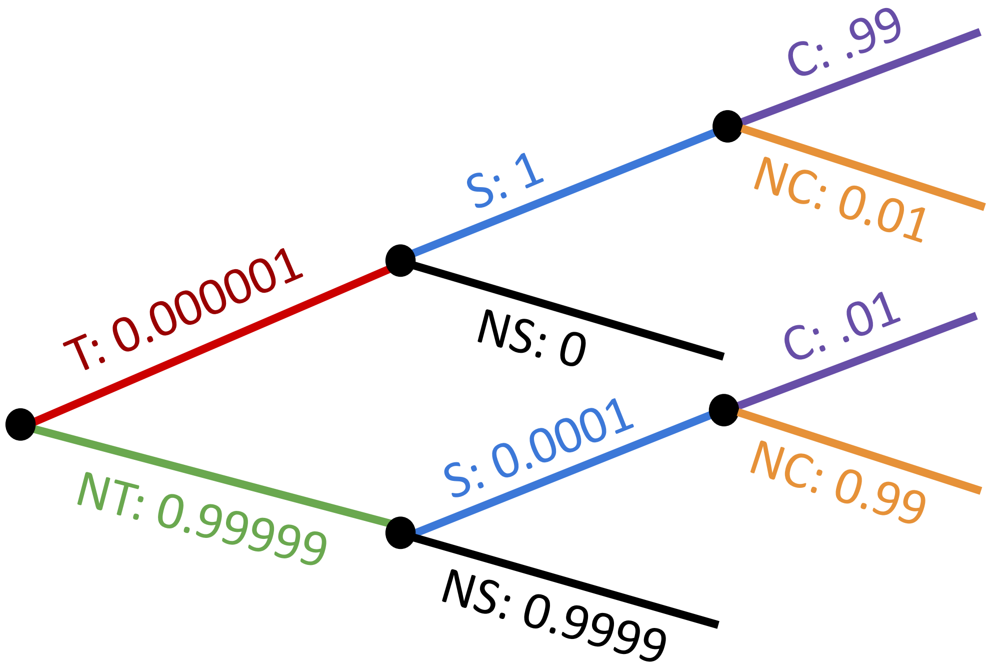
Assumptions in English:

- There are 250 million adults in the United States, of whom 250 are terrorists.
- If someone is a terrorist, then the government's monitoring system will always flag them.
- If someone is not a terrorist, then the system will flag them anyway 1 in 10,000 times.
- The criminal justice system's verdict is incorrect 1% of the time:
  - A terrorist who is flagged is deemed guilty 99% of the time.
  - A flagged civilian is judged to be innocent 99% of the time.

Assumptions as Probabilities:

- $P(T) = 0.000001$
- $P(S|T) = 1$
- $P(S|NT) = 0.0001$
- $P(C|S \text{ and } T) = .99$, $P(C|S \text{ and } NT) = .01$

Worked Example 3



P(C and S and T)
= P(C|S) P(S|T) P(T)
= (0.000001)(1)(0.99)
= 0.00000099

P(C)
= P(C and S and T) + P(C and S and NT)
= (0.000001)(1)(0.99) +
(0.999999)(0.0001)(0.01)
= 0.000001989999

P(T|C)
= P(C|T) P(T) / P(C)
= [P(C and S|T) + P(C and NS|T)] P(T) / P(C)
= [0.99 + 0] (0.000001) / 0.000001989999
= 0.00000099 / 0.000001989999
~ 0.497

If someone is convicted of terrorism, there is only about a 49.7% chance they are actually a terrorist!

Problem 1:
Morning after the GCB

Worked Example

I wake up exhausted after a night at the GCB (Grad Center Bar) and realize my alarm didn't go off. In classic fashion, I seem to have lost my phone, and I have no way to check what day it is.

What I do know is the sun is up, and it is time to fulfill my daily obligations. In particular, I have a very important meeting with a professor every **Wednesday** morning.

Help me find the likelihood that this morning is a Wednesday given that I went to the GCB the night before!

Worked Example

For simplicity, assume there are 52 weeks in the year, and 7 days per week (so 364 days in the year)

- I go to the GCB 3 Sundays of the year
- I go to the GCB 2 Mondays of the year
- I go to the GCB 1 Tuesday of the year
- I go to the GCB 12 Wednesdays of the year
- I go to the GCB 11 Thursdays of the year
- I go to the GCB 28 Fridays of the year
- I go to the GCB 34 Saturdays of the year

Worked Example

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)}$$

We want to find P (Wednesday| GCB). Recall Bayes' Rule. What probabilities do we need to find to get there?

How do we use the below information to compute these probabilities? Hint: one useful bit of information is that I go to the GCB a total of 91 days per year.

- 52 weeks per year, 364 days per year
- I go to the GCB 3 Sundays of the year
- I go to the GCB 2 Mondays of the year
- I go to the GCB 1 Tuesday of the year
- I go to the GCB 12 Wednesdays of the year
- I go to the GCB 11 Thursdays of the year
- I go to the GCB 28 Fridays of the year
- I go to the GCB 34 Saturdays of the year

Use Bayes' Rule

P(A|B) = P(B|A) P(A) / P(B)

P(Wednesday | GCB) = P(GCB| Wednesday) P(Wednesday) / P(GCB)

P(GCB) = 91/364 = 1/4 = 1/4

P(Wednesday) = 1/7

P(GCB | Wednesday) = 12/52 = 3/13

P(Wednesday | GCB) = (3/13) (1/7) / (1/4) = 12/91 ~ 13.19%

Problem 2: Monty Hall

Monty Hall Problem

Assumptions in English

- The car is equally likely to be behind 1 of *N* doors.
- We pick door 1.
- Monty Hall opens door 2.
- The argument is symmetric, so we could pick any door, and Monty could open either of the others.

Assumptions as Probabilities

- A_i: The car is behind door *i*.
- B_i: Monty opens door *i*.
- $P(A_i) = 1/3$
- $P(B_2 | A_1) = 1/2$
- $P(B_2 | A_2) = 0$
- $P(B_2 | A_3) = 1$

Monty Hall Problem (cont'd)

Assumptions in English

- The car is equally likely to be behind 1 of *N* doors.
- We pick door 1.
- Monty Hall opens door 2.
- The argument is symmetric, so we could pick any door, and Monty could open either of the others.

Assumptions as Probabilities

- A_i: The car is behind door *i*.
- B_i: Monty opens door *i*.
- $P(A_i) = 1/3$
- $P(B_2 | A_1) = 1/2$
- $P(B_2 | A_2) = 0$
- $P(B_2 | A_3) = 1$

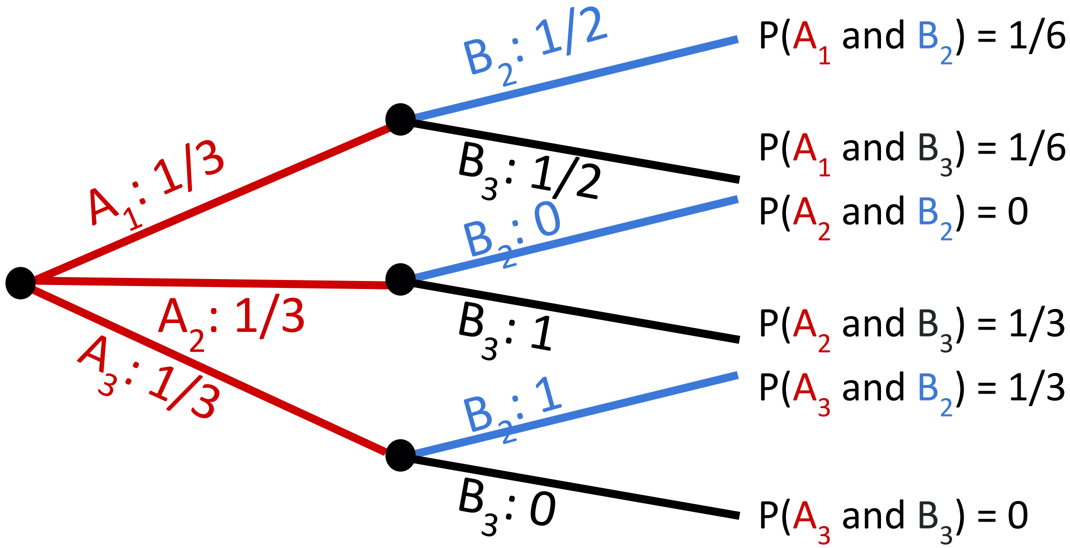
Should we switch?

Yes, if $P(A_1 | B_2) < ½$.

What is $P(A_1 | B_2)$?

$P(B_2 | A_1)\ P(A_1) / P(B_2)$

Monty Hall Problem (cont'd)



$$
\begin{aligned}
&P(B_2) \\
&= P(B_2 \mid A_1)\,P(A_1) + \\
&\quad P(B_2 \mid A_2)\,P(A_2) + \\
&\quad P(B_2 \mid A_3)\,P(A_3) \\
&= (\tfrac{1}{2})(\tfrac{1}{3}) + (0)(\tfrac{1}{3}) + (1)(\tfrac{1}{3}) \\
&= \tfrac{1}{6} + \tfrac{1}{3} = \tfrac{1}{2}
\end{aligned}
$$

$$
\begin{aligned}
&P(A_1 \mid B_2) \\
&= P(B_2 \mid A_1)\,P(A_1)\,/\,P(B_2) \\
&= (\tfrac{1}{2})(\tfrac{1}{3})\,/\,(\tfrac{1}{2}) \\
&= \tfrac{1}{3}
\end{aligned}
$$

If we pick door 1, and he shows us door 2, the probability we are correct is only ⅓, so we should switch to door 3!