Naïve Bayes

Lecture 02.01

Statistics is a tool to aid and organize our reasoning and beliefs about the world

Today

- Belief and evidence
- Empirical reasoning: always probabilistic
- Inductive reasoning with probabilities
- Bayes method for updating beliefs
- Naïve Bayes classifier

Belief and evidence
Inductive reasoning

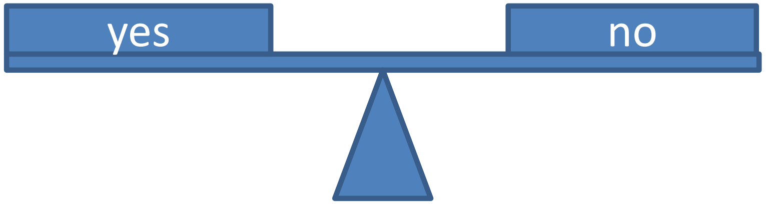
- Critical thinking: always have good reasons for your beliefs
- Some reasons are 100% true
- Some only probable
- Inductive reasoning with probabilities: you always have a chance of being wrong

http://www.starwars.com/video/never-tell-me-the-odds

I believe that John will not be at the party

In the absence of facts

John will not be at the party



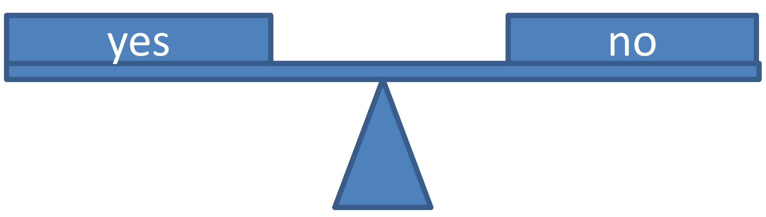
What are the odds?

I believe that John will not be at the party

Invalid reasoning

I do not like John

John will not be at the party



What are the odds?

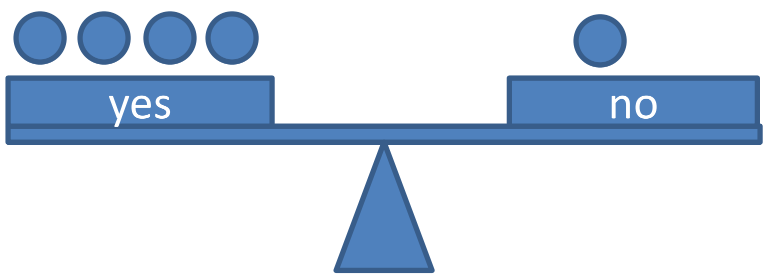
I believe that John will not be at the party

Probabilistic reasoning: valid fact (evidence)

I do not like John

John is very shy

John will not be at the party

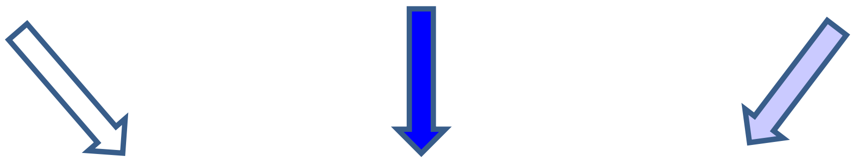


What are the odds given this fact?

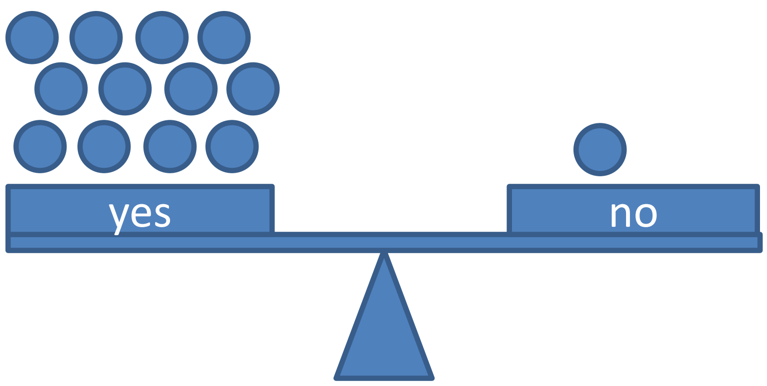
I believe that John will not be at the party

More facts – update your beliefs

I do not like John | John is in Beijing | John is very shy



John will not be at the party



What are the odds?

Bayesian beliefs

- How do we judge that something is true?
- Can mathematics help make judgments more accurate?
- Bayes: our believes should be **updated** as new evidence becomes available

T. Bayes.

1701 - 1761

Bayes' method for updating beliefs

- There are 2 events: **A** and not A (**B**) which you believe occur with probabilities P(**A**) and P(**B**). Estimation P(**A**):P(**B**) represents ***odds*** of A vs. B.
- Collect evidence data **E.**
- Re-estimate P(**A**|**E**):P(**B**|**E**) and update your beliefs.

Probabilities. Bayes theorem

Bayes theorem (formalized by Laplace)

$$P(A|E) = P(A \cap E)/ P(E)$$
$$P(E|A) = P(A \cap E)/ P(A)$$

↓

Probability of event A given evidence

Probability of evidence given event A

$$P(A|E) = P(E|A)P(A)/P(E)$$

Probability of event A without evidence (***prior probability***)

Inverse probabilities are typically easier to ascertain

Bayes' method with probabilities

- There are 2 events: **A** and not A (**B**) which you believe occur with probabilities P(**A**) and P(**B**). Estimation P(**A**):P(**B**) represents odds of A vs. B.
- Collect evidence data **E.**
- Re-estimate P(**A**|**E**):P(**B**|**E**) and update your beliefs.

The updated odds are computed as:

$$\frac{P(A|E)}{P(B|E)} = \frac{P(E|A)P(A)/P(E)}{P(E|B)P(B)/P(E)}$$

Bayes' method with probabilities

- There are 2 events: **A** and not A (**B**) which you believe occur with probabilities P(**A**) and P(**B**). Estimation P(**A**):P(**B**) represents odds of A vs. B.
- Collect evidence data **E.**
- Re-estimate P(**A**|**E**):P(**B**|**E**) and update your beliefs.

or simply

$$\frac{P(A|E)}{P(B|E)} = \frac{P(E|A)P(A)}{P(E|B)P(B)}$$

Explanation by example: hit-and-run (fictitious)

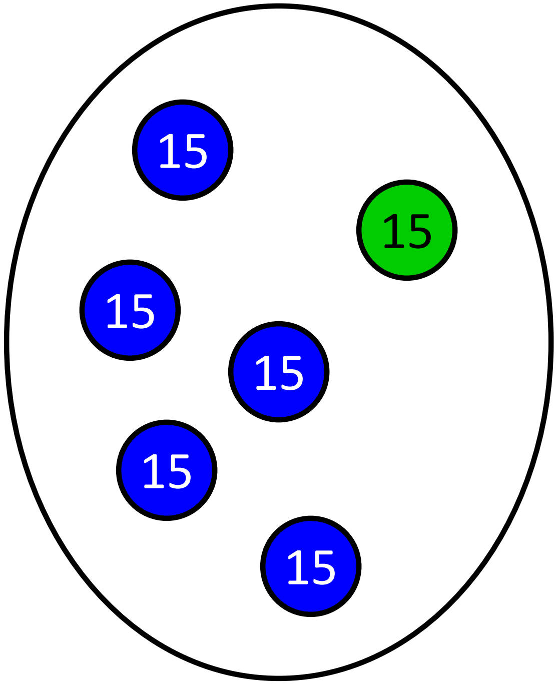
- Taxicab company has 75 blue cabs (**B**) and 15 green cabs (**G**)

- At night when there are no other cars on the street: hit-and-run episode

- Question: what is more probable:

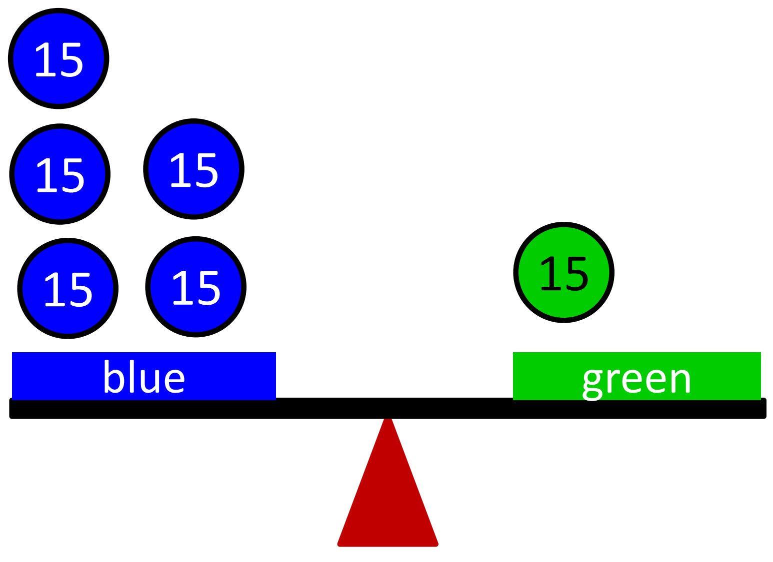
  B or **G**

  ?



Adopted from: **The numbers behind NUMB3RS:** solving crime with mathematics by Devlin and Lorden.

What is more probable:
B or G



P(B):P(G)=5:1

New evidence

- Witness: “I saw a green cab”: E_G
- What is the probability that the witness really saw a green car?
- Witness is tested at night conditions: identifies correct color 4 times out of 5

- The eyewitness test shows:

$P(E_G \mid G) = 4/5$ (correctly identified)

$P(E_G \mid B) = 1/5$ (incorrectly identified)

Updating the odds

- In our case we want to compare:

the car was **G** given a witness testimony E_G: $P(\mathbf{G}|\mathbf{E_G})$

vs.

the car was **B** given a witness testimony E_G: $P(\mathbf{B}|\mathbf{E_G})$

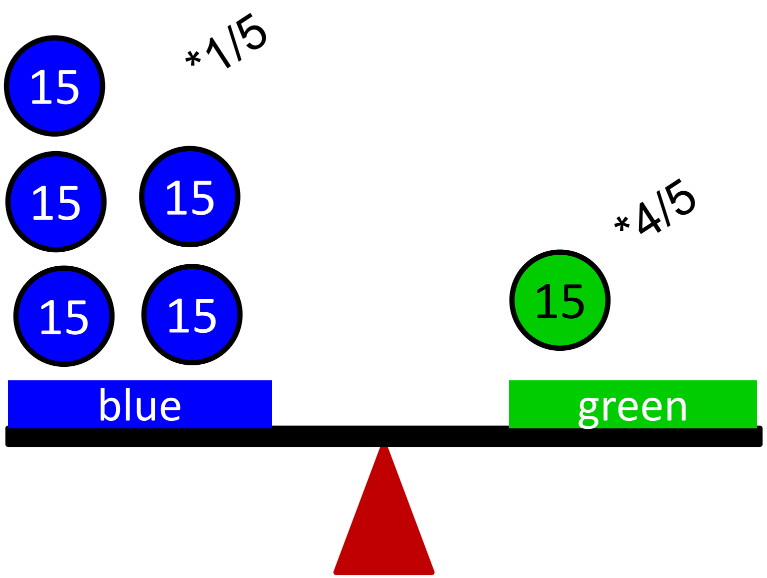
Note: There is no way to know which of 2 was true, we just estimate

Back to hit-and-run

All cabs were on the streets:

Prior odds ratio: P(**B**) : P(**G**) = 5/1

Updated odds ratio: $\frac{P(B|E_G)}{P(G|E_G)} = \frac{P(B)*P(E_G|B)}{P(G)*P(E_G|G)}$



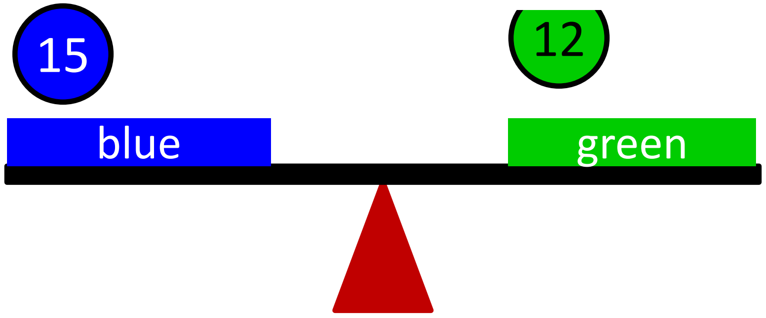
$P(E_G|G)$= 4/5 (correctly identified)

$P(E_G|B)$= 1/5 (incorrectly identified)

New odds

$$\frac{P(\mathbf{B}|E_G)}{P(\mathbf{G}|E_G)} = \frac{P(\mathbf{B})*P(E_G|\mathbf{B})}{P(\mathbf{G})*P(E_G|\mathbf{G})}$$

<u>Still 5:4 odds that the car was **B**!</u>



Hit-and-run: full calculation

P(**B**) =5/6, P(**G**) = 1/6

P(E_G | **G**)= 4/5 P(E_G | **B**)= 1/5

- Probability that car was **green** given the evidence E_G:

P(**G**|E_G)= P(**G**)* P(E_G|**G**) /P(E_G) = [1/6 * 4/5] / P(E_G) =4/30P(E_G)

//- 4 parts of 30P(X_G)

- Probability that car was **blue** given the evidence X_G:

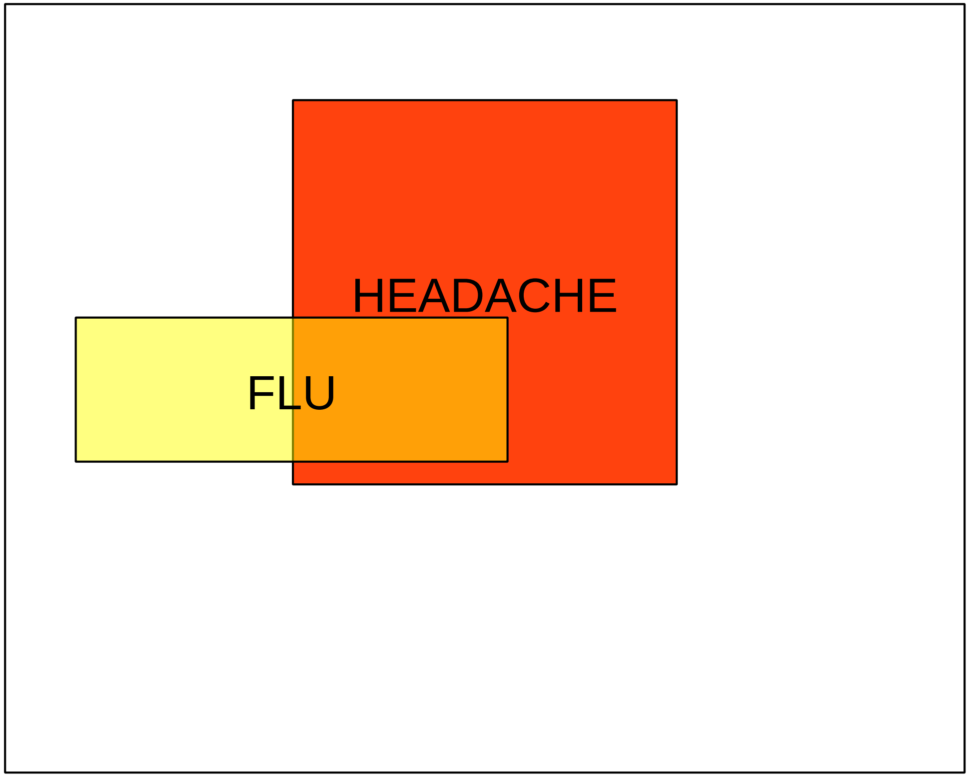
P(**B**|E_G) = P(**B**)* P(E_G|**B**) /P(E_G) = [5/6 * 1/5] /P(E_G) =5/30P(E_G)

//- 5 parts of 30P(X_G)

Bayes in 'real' life. Example 1

$P(H)=1/10$
$P(F)=1/40$
$P(H|F)=1/2$

$P(F|H) = ?$



Bayes in 'real' life. Example 1

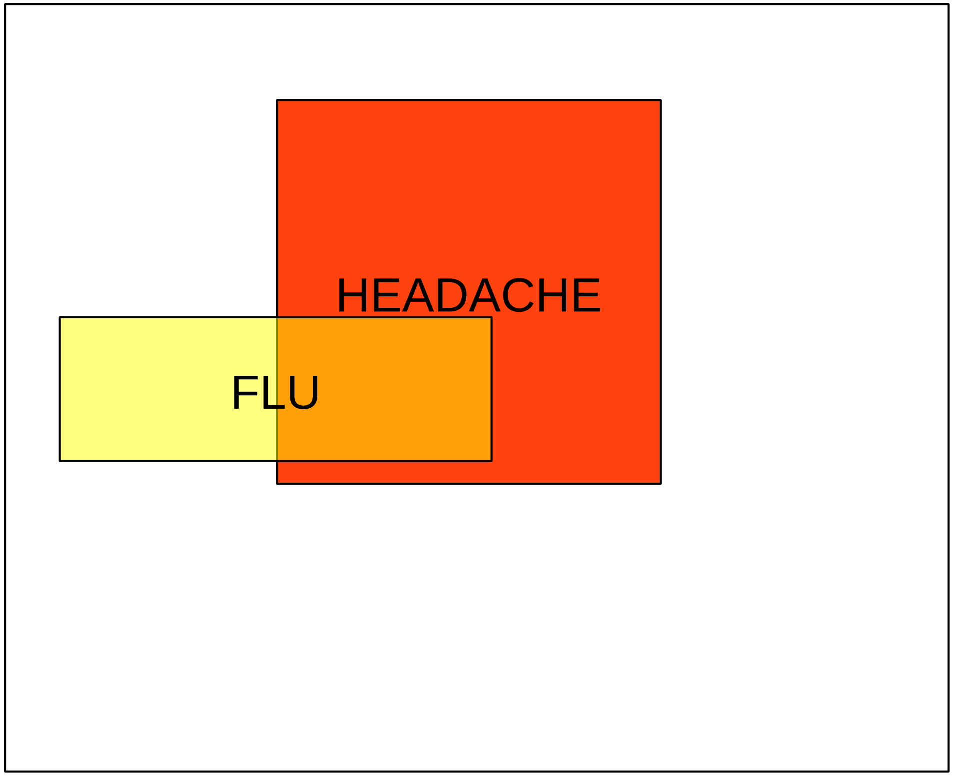
P(H)=1/10

P(F)=1/40

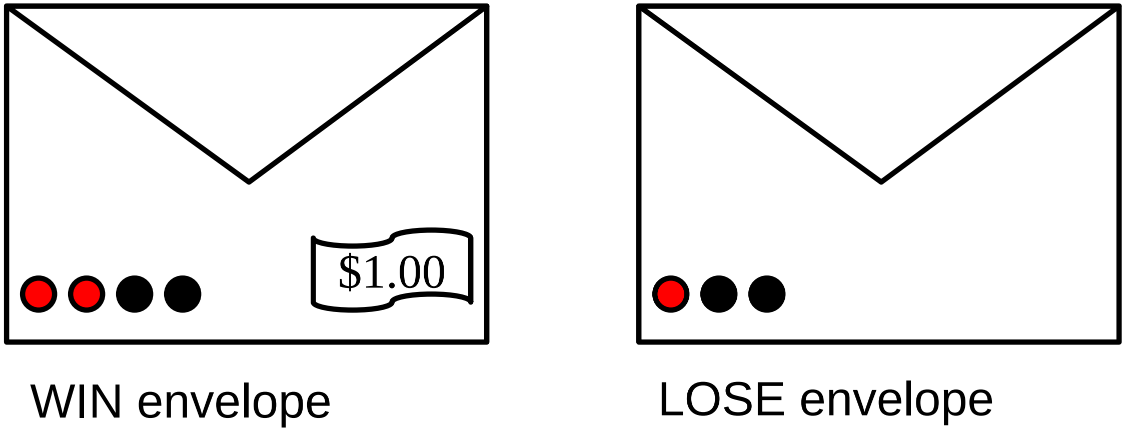
P(H|F)=1/2

P(F|H) =P(H|F)P(F)/P(H)

=1/2*1/40 *10=1/8

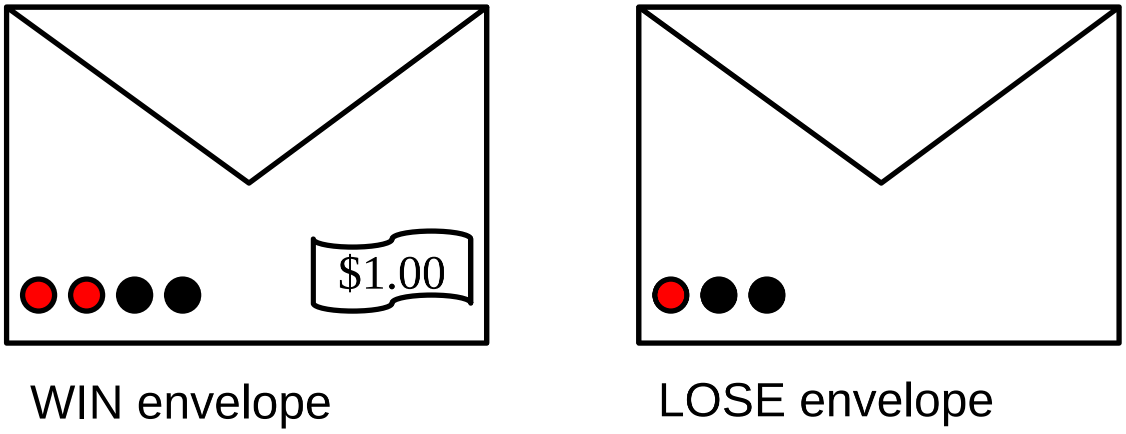


Bayes in 'real' life. Example 2



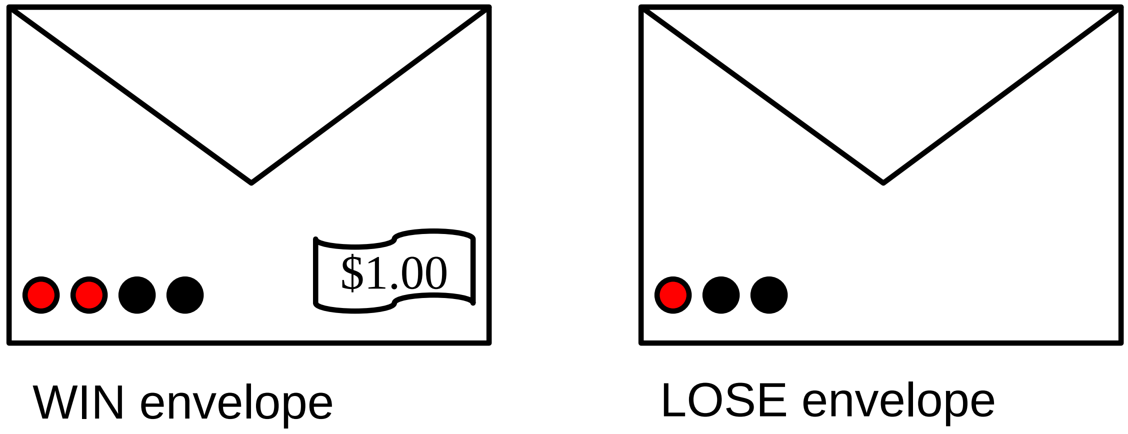
Someone draws an envelope at random and offers to sell it to you.
How much should you pay?
The probability to win is 1:1. Pay no more than 50c.

Bayes in ‘real’ life. Example 2



Variant: before deciding, you are allowed to see one bead drawn from the envelope.
Suppose it’s black: How much should you pay?
Suppose it’s red: How much should you pay?

Bayes in ‘real’ life. Example 2



Variant: before deciding, you are allowed to see one bead drawn from the envelope.

Suppose it’s black: How much should you pay?

P(W|b)=P(b|W)P(W)/P(b) =(1/2*1/2)/P(b)=1/4 *1/P(b)

P(L|b)=P(b|L)P(L)/P(b)=(2/3*1/2)/P(b) = 1/3 * 1/P(b)

Probability to win is now 3:4 – pay not more than $(3/7)

Suppose it’s red: How much should you pay? – the same logic

When you want to

- Determine the probability of having a medical condition after positive test results
- Find out a probable outcome of political elections
- Improve machine-learning performance
- Even to "prove" and "disprove" the existence of God

Use Bayesian Reasoning

Mathematical *predictions*

- We can 'predict' where the spacecraft will be at noon in 2 months from now
- We cannot predict where you will be tomorrow at noon
- But, based on numerous observations (evidence), we can *estimate the probability*

Need for probabilistic learners

- Given the evidence (data),

can we certainly derive

the diagnostic rule:

if Toothache=true then Cavity=true ?

Name	**Toothache**	**…**	**Cavity**
Smith	true	…	true
Mike	true	…	true
Mary	false	…	true
Quincy	true	…	false
…	…	…	…

Historical data

- This rule isn't right always.
  - Not all patients with toothache have cavities - some of them have gum disease, an abscess, etc.

- We could try an inverted rule:

if Cavity=true then Toothache=true

- But this rule isn't necessarily right either; not all cavities cause pain.

Certainty and Probability

- The connection between toothaches and cavities is not a certain logical consequence in either direction.

- However, we can provide a **probability** that given an evidence (toothache) the patient has cavity.

- For this we need to know:
  - Prior probability of having cavity: how many times dentist patients had cavities: $P(cavity)$
  - The number of times that the evidence (toothache) was observed among all cavity patients: $P(toothache|cavity)$

Bayes' Rule
for diagnostic probability

Bayes' rule:

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

- Useful for assessing **diagnostic** probability from **symptomatic** probability as:

P(Cause|Symptom) = P(Symptom|Cause) P(Cause) / P(Symptom)

- Bayes's rule is useful in practice because there are many cases where we do have good probability estimates for these three numbers and need to compute the fourth