Bayesian Reasoning

Lecture 17

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

Statistics primer
for Bayesian thinking

BOOLEAN-VALUED RANDOM VARIABLES

Discrete Boolean-valued random variables

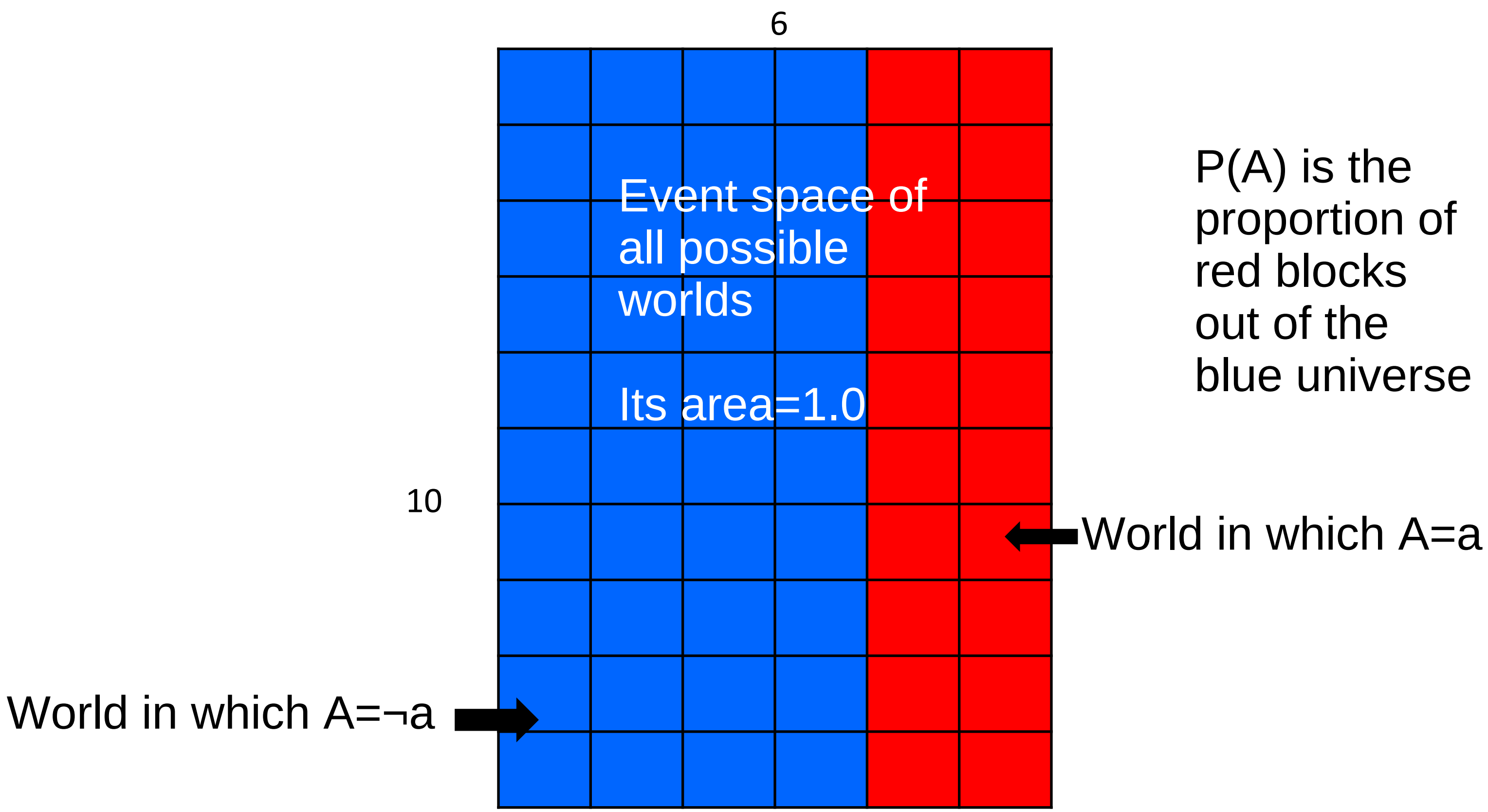
A is a Boolean-valued random variable if *A* denotes an event, and there is some degree of uncertainty as to whether *A* occurs or not.

Examples:

- P = True: The US president in 2024 will be male
- P=¬True: The US president will not be a male
- H = True: You wake up tomorrow with a headache
- H=¬True: No headache

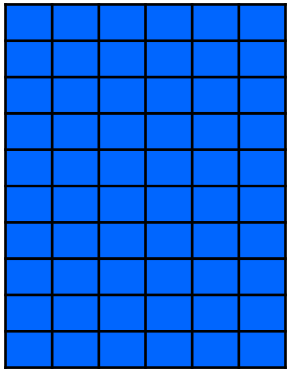
Probabilities

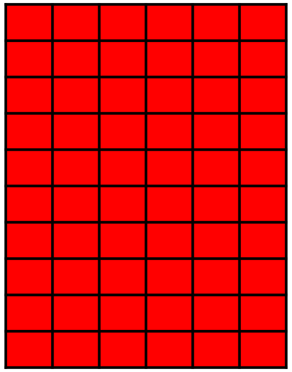
We write P(A=a), or P(A=true) or simple P(A) as "the fraction of possible worlds where A=a is true"



P(A) is the proportion of red blocks out of the blue universe

The Axioms of Probability

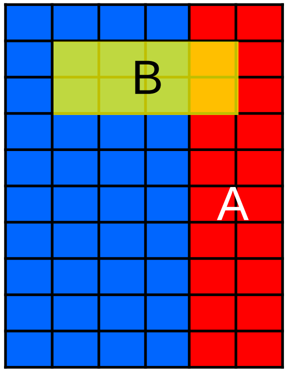




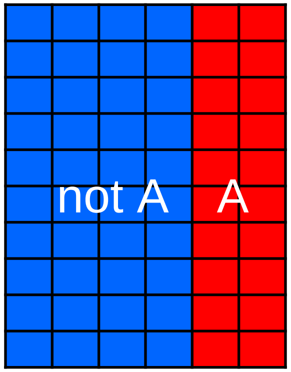
We do not need to prove that:

I. 0<= P(A=a)<=1

II. P(A or B)=P(A)+P(B)-P(A and B)

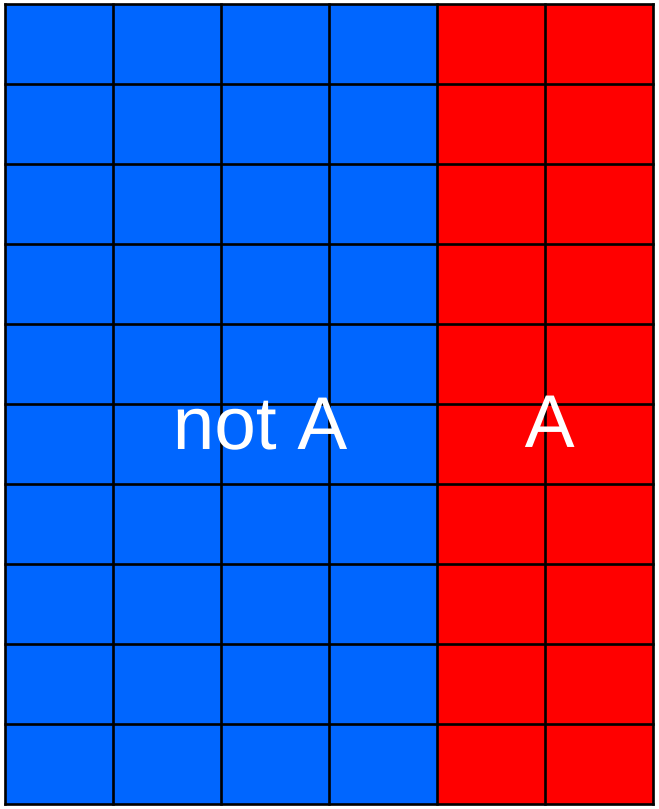


III. P(A)+P(¬A)=1



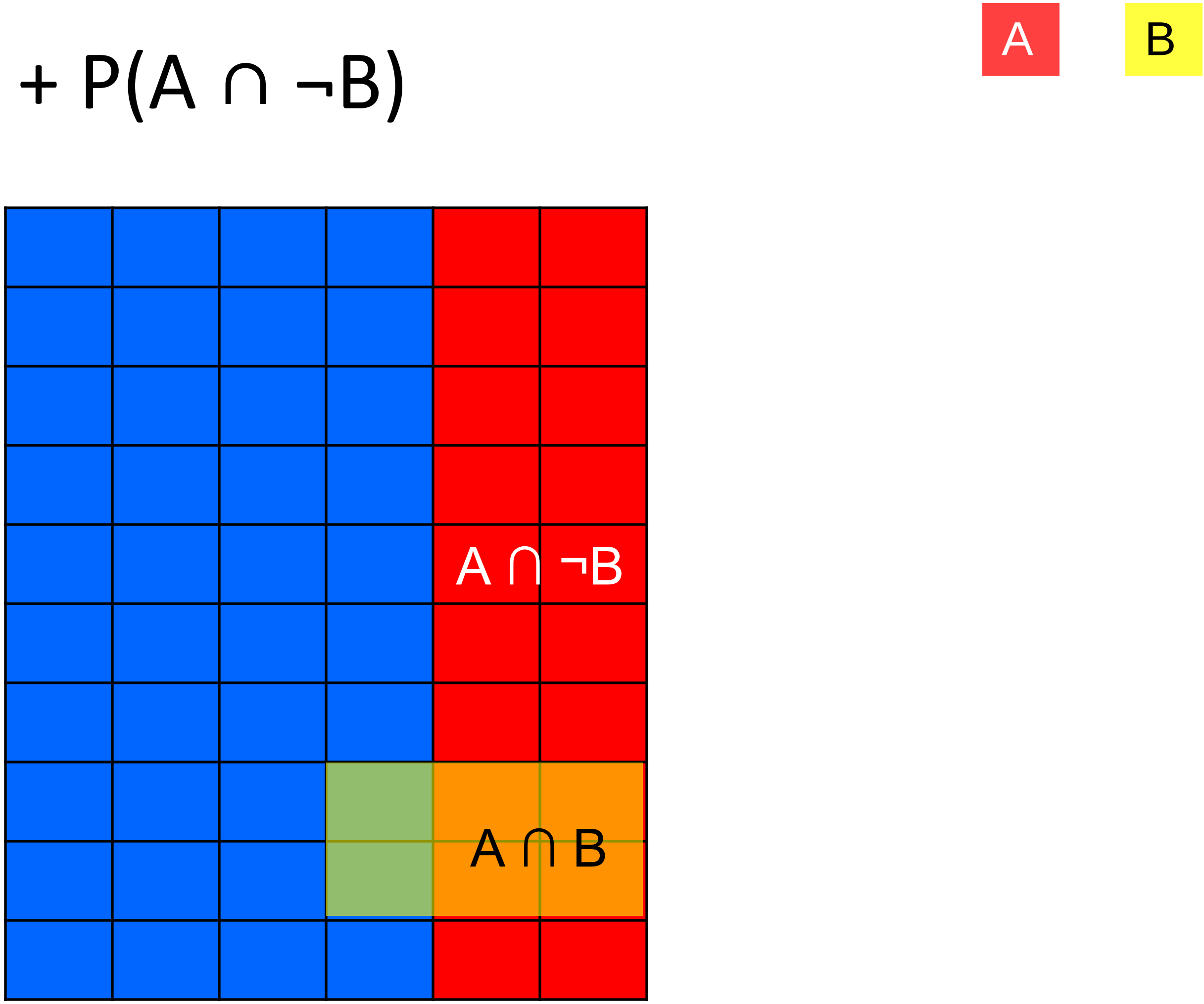
Theorems of Probability: Theorem 1

$P(\neg A) = 1 - P(A)$



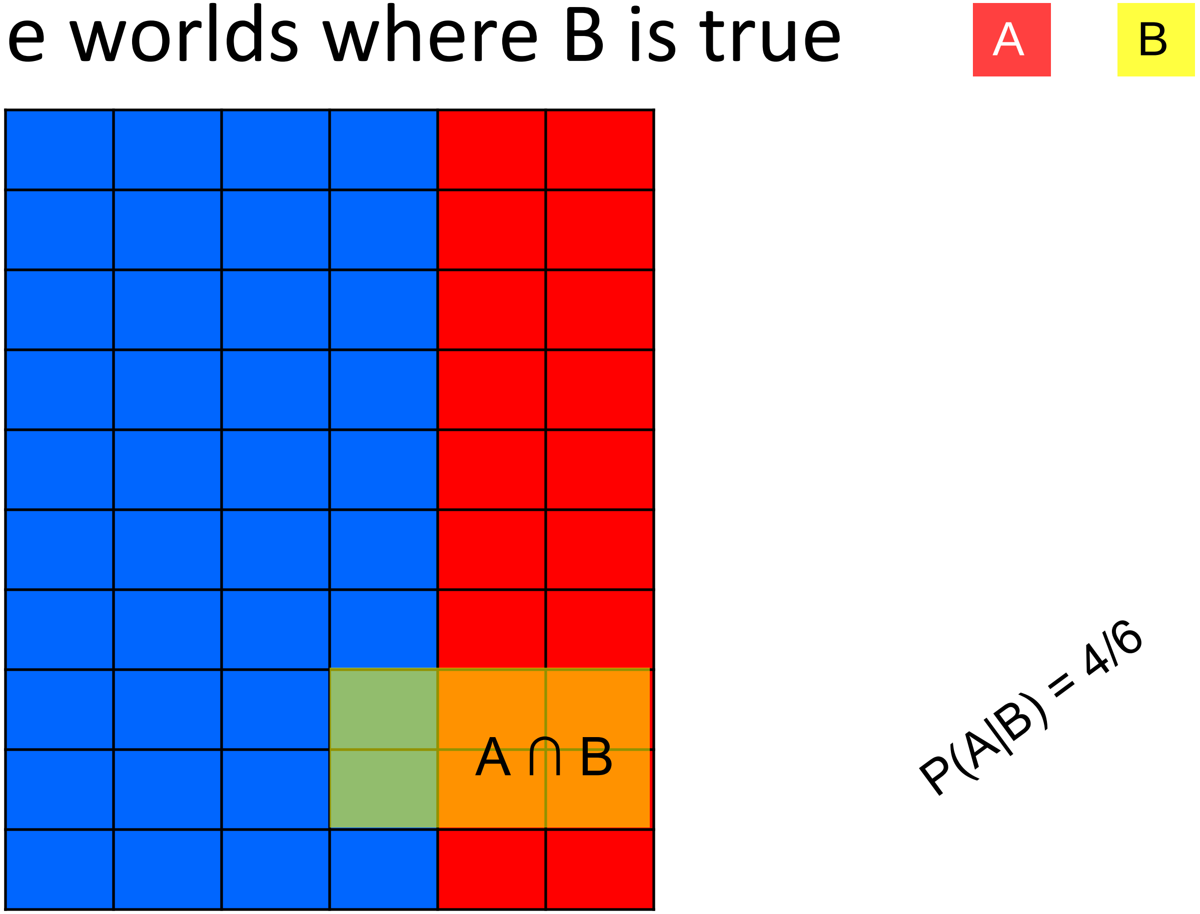
Theorems of Probability: Theorem 2

$P(A)=P(A \cap B) + P(A \cap \neg B)$



Conditional probability: *definition*

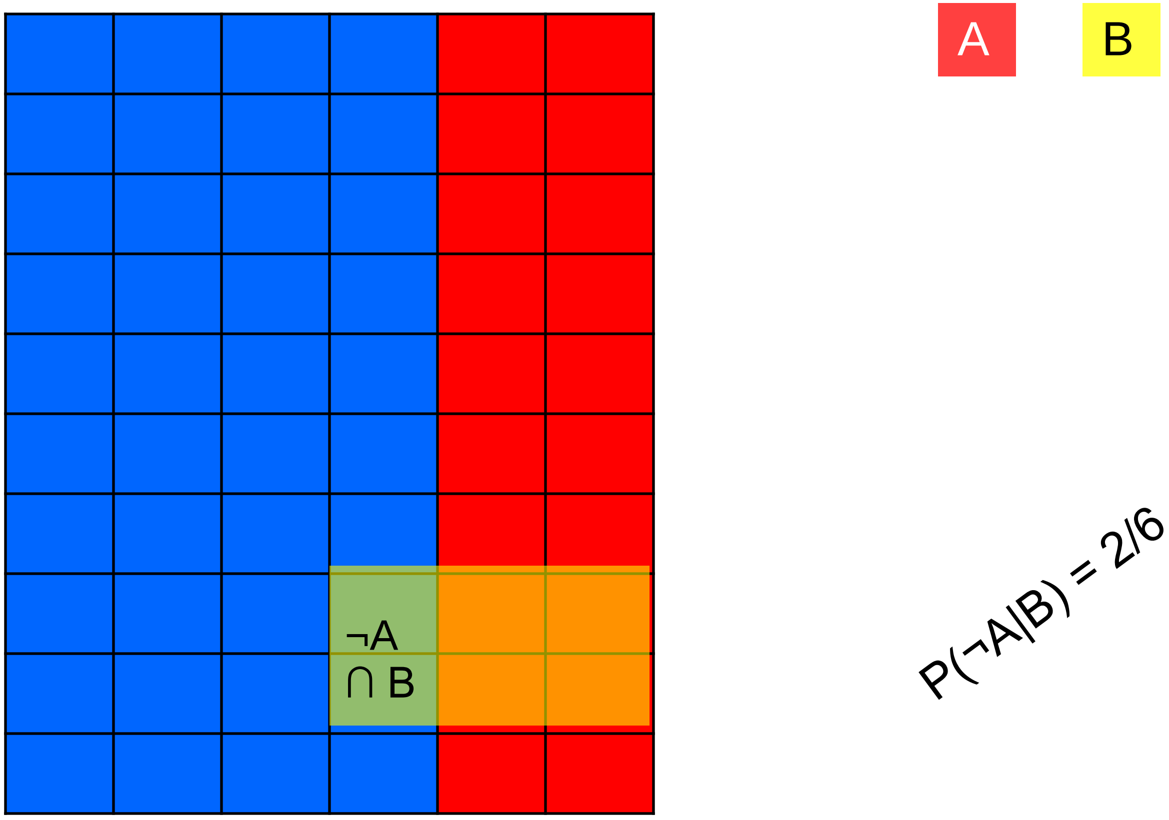
- P(A|B) = fraction of worlds in which A is true out of all the worlds where B is true



CP definition: $P(A|B) = P(A \cap B) / P(B)$

Conditional probability: definition

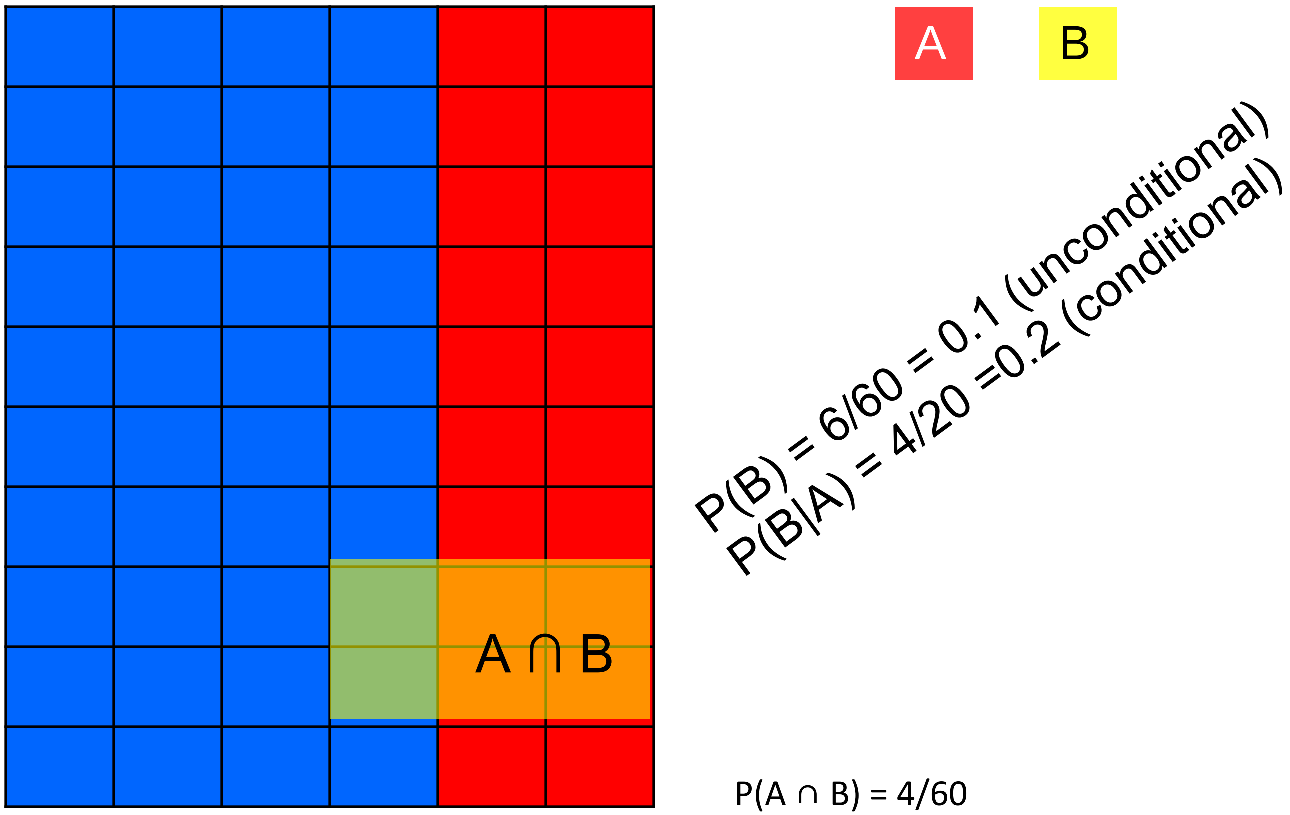
- P(A|B) = fraction of worlds in which A is true out of all the worlds where B is true



P (¬A|B) = P(¬A ∩ B) / P(B)

Conditional probability: definition

- P(B|A) = fraction of worlds in which B is true out of all the worlds where A is true



P(B|A)= P(A ∩ B) / P(A)

P(A ∩ B) = 4/60
P(A) = 20/60
P(B|A) = 4/60 : 20/60 = 0.2

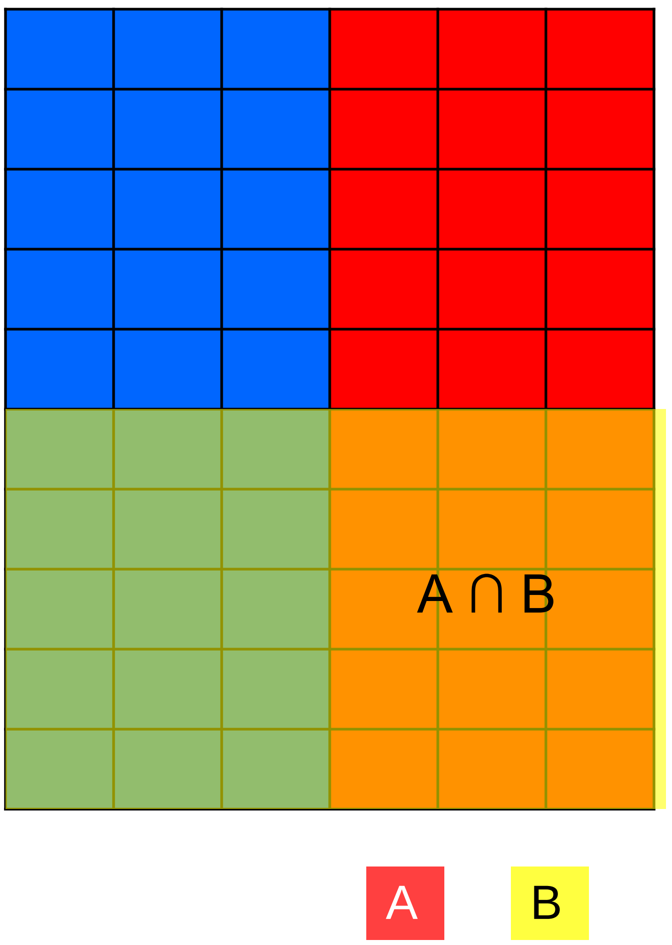
Probabilistic independence

Two random variables A and B are *mutually independent* if

$P(A|B) = P(A)$, which means that:

Knowing that B is true (or false)

does not change the probability of A

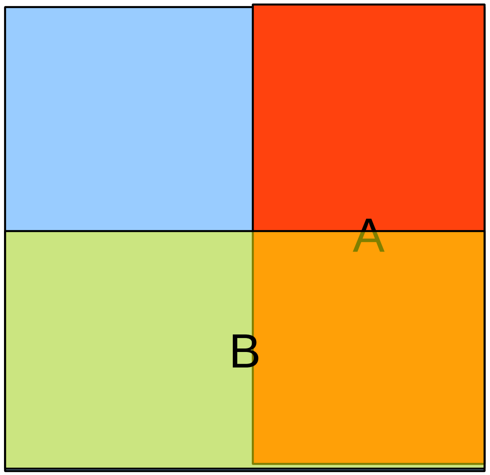


$P(A|B) = P(A)$	$15/30 = 30/60$
$P(\neg A|B) = P(\neg A)$	$15/30 = 30/60$
$P(A|\neg B) = P(A)$	$15/30 = 30/60$
$P(\neg A|\neg B) = P(A)$	$15/30 = 30/60$

Independent and mutually exclusive events

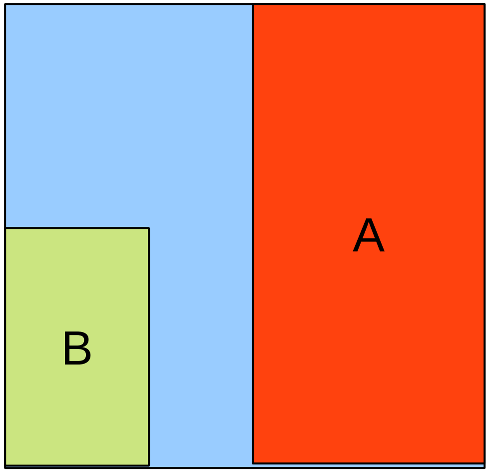
A is *independent* of B: knowing that B is true (or false) does not change the probability of A:

$$P(A|B) = P(A)$$



A and B are *mutually exclusive* – not independent variables: if A is true then B is false, if A is false then B is true with probability $P(B|\neg A)$

$$P(A \cap B)=0$$



Joint probability of two events

From the definition of conditional probabilities:

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

we can compute $P(A \cap B)$ – that both events happened together:

$$P(A \cap B) =P(A|B)P(B)$$

If A and B are *independent* that becomes:

$$P(A \cap B) =P(A)P(B)$$

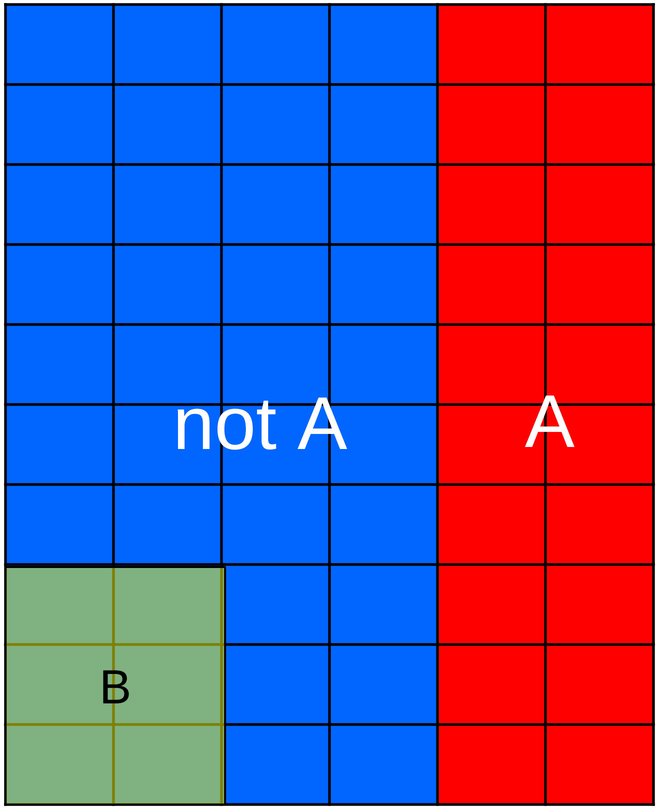
Probability of two mutually exclusive events

If A and B are mutually exclusive:

P(A or B)=P(A)+P(B)-~~P(A and B)~~

A and ¬A are mutually exclusive:

P(A or ¬A)=P(A) + P(¬A) = 1



Conditional independence

Conditional independence means that once you know the value of 1 random variable, other variables become independent

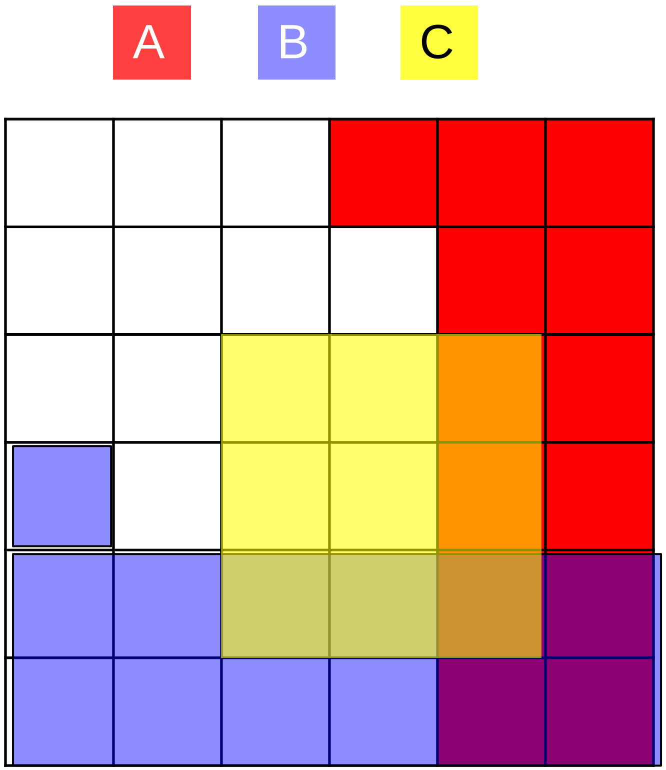
Example: (Height, Vocabulary) are not independent since very small people tend to be children, known for their more basic vocabularies

But given that two people are 19 years old (i.e., conditional on age) there is no reason to think that one person's vocabulary is larger if we are told that they are taller.

So given that we know Age, we can calculate joint conditional probability of (Height, Vocabulary) as a simple multiplication.

https://en.wikipedia.org/wiki/Conditional_independence

Conditional independence of 2 variables given the third



P(A ∩ B | C) = 1/9

P(A ∩ B) = 4/36 = 1/9

P(A ∩ B | C) = P(A ∩ B)

Definition of conditional independence

P(A|C) = 3/9

P(B|C) = 3/9

P(A ∩ B | C) = P(A|C)*P(B|C)

A and B are independent in the world where C is True (C is known, it occured)

However, in general A and B are not independent:

P(A) = 13/36 ≈ 0.36, and P(A|B) = 4/13 ≈ 0.31 → P(A) ≠ P(A|B)

Bayes theorem

$$P(A \cap B) = P(A|B)P(B)$$

From definition of Conditional probability

On the other hand:

$$P(B \cap A) = P(B|A)P(A)$$

⇓

$$P(A|B)P(B) = P(B|A)P(A)$$

and we can express conditional probability of A given B through conditional probability of B given A and unconditional probabilities of A and B:

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

Multiple Boolean random variables

All theorems for 2 Boolean-valued random variables can be extended to several random variables $C, E_1, E_2,\ldots,E_n$.

Let $C, E_1, E_2,\ldots,E_n$ be Boolean-valued random variables.

For convenience, we will let E denote the n-tuple of random variables $(E_1, E_2,\ldots,E_n)$

$$E_1, E_2,\ldots,E_n = E$$

Just a notation

$$P(C \cap E_1 \cap E_2 \cap \ldots \cap E_n)=P(C,E_1,E_2,\ldots E_n)=P(C,E)$$

Chain rule:

$$P(C,E)=P(C)P(E_1|C,E_2,\ldots E_n)P(E_2|C,E_1, E_3,\ldots,E_n)x\ldots xP(E_n|C,E_1,\ldots E_{n-1})$$

Multiple variables dependent on C

C – condition
E – evidence (event)

If $E_1,\ldots E_n$ are mutually *conditionally* independent given C:

$$P(C,E)=P(C)P(E_1|C)P(E_2|C)x\ldots xP(E_n|C)$$

And from Bayes theorem:

$$P(C|E)=P(C,E)/P(E)$$

That gives you a formula of the probability that the unknown condition C was true given a set of known evidences E

Main lecture

Bayesian Reasoning

Outline

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

Belief and evidence
Inductive reasoning

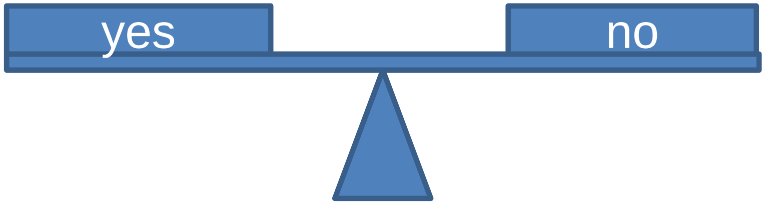
- Critical thinking: always have good reasons for your beliefs
- Some reasons are 100% true - some only probable
- Inductive reasoning with probabilities: you always have a chance of being wrong

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

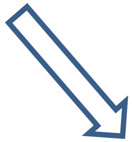


What are the odds?

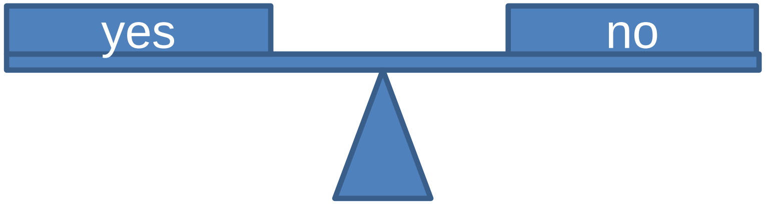
I believe that John will not be at the party

Invalid (illogical) reasoning

I do not like John



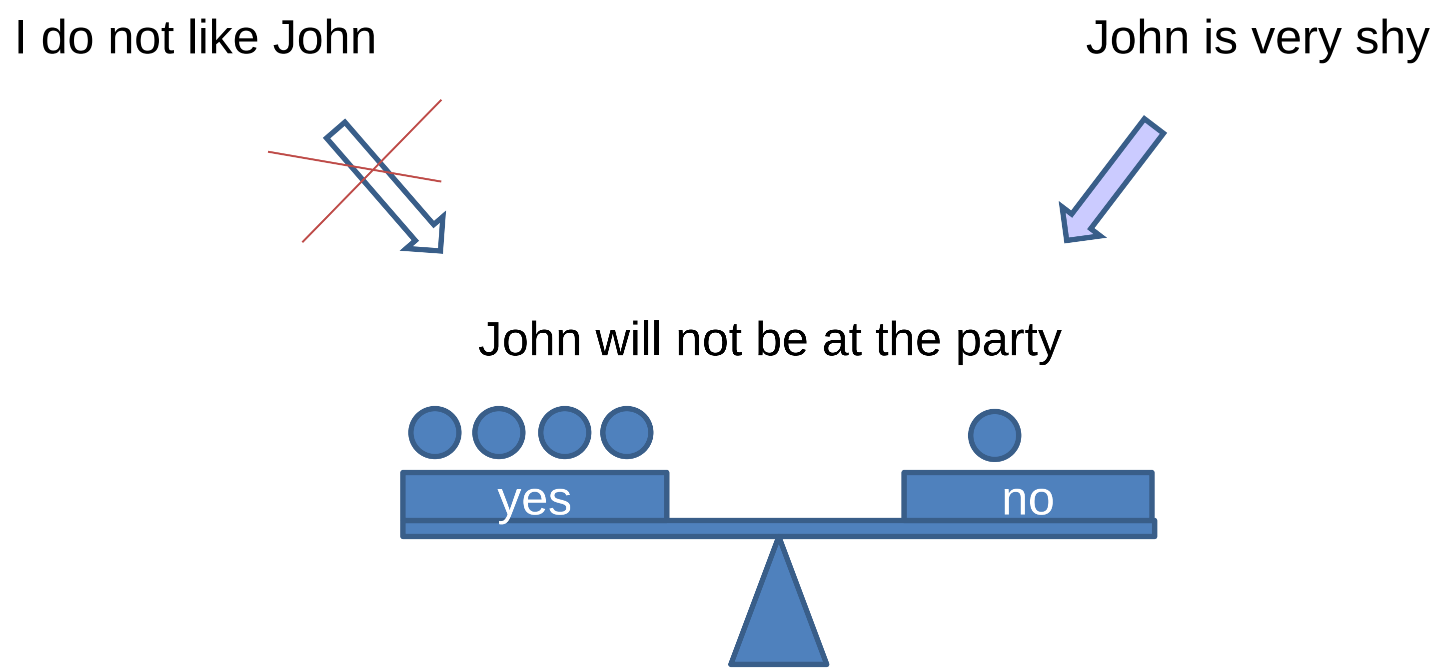
John will not be at the party



What are the odds?

I believe that John will not be at the party

Probabilistic reasoning: valid fact (evidence)



What are the odds given this fact?

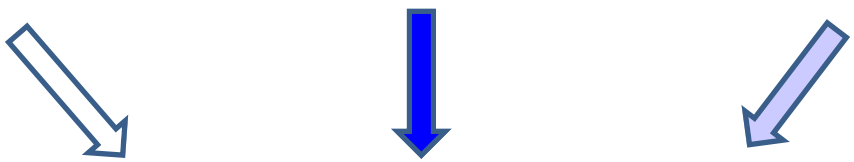
I believe that John will not be at the party

More facts – update your beliefs

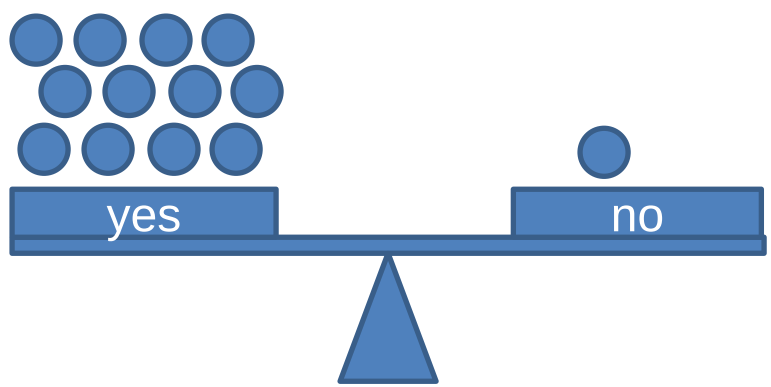
I do not like John

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

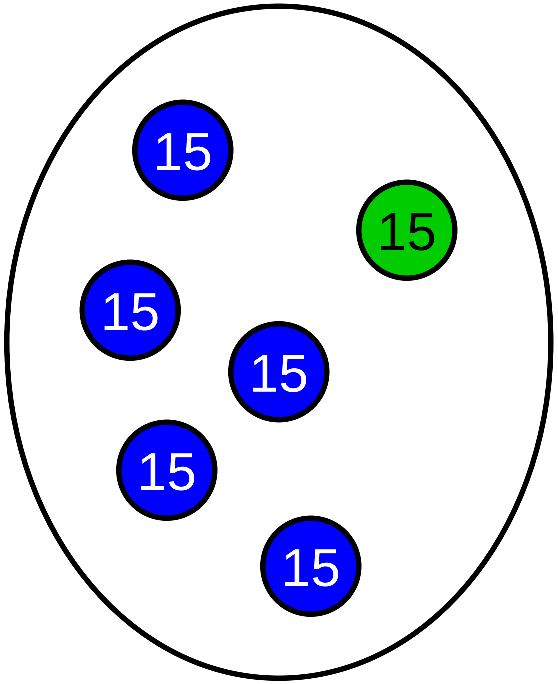
$$\frac{P(A|E)}{P(B|E)} = \frac{P(E|A)P(A)}{P(E|B)P(B)}$$

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:

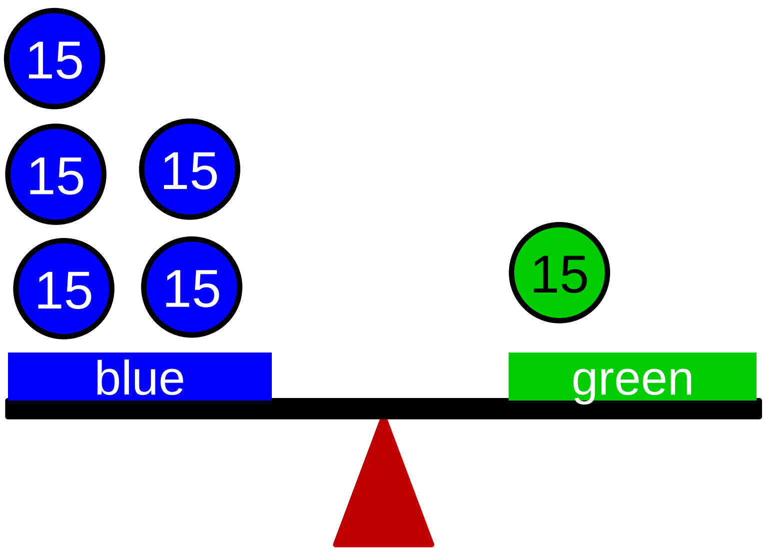
B or **G**

?



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: $P(G|E_G)$

vs.

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

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

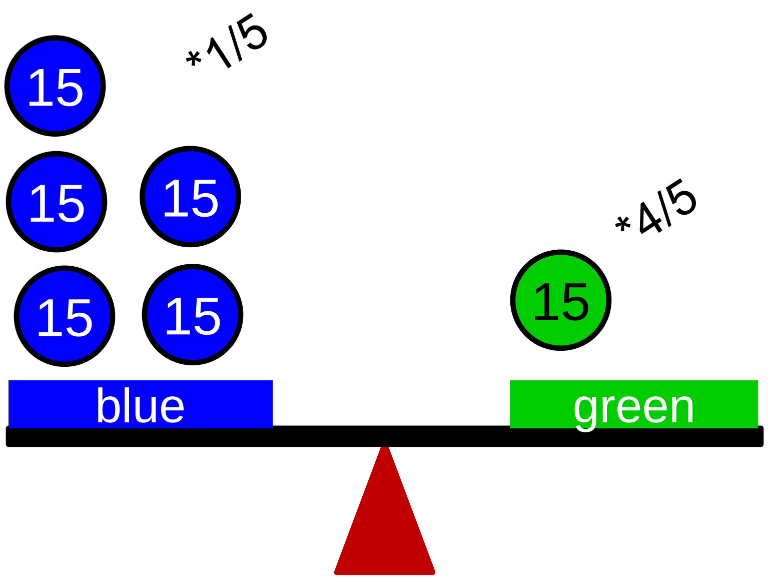
Back to hit-and-run

All cabs were on the streets:

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

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

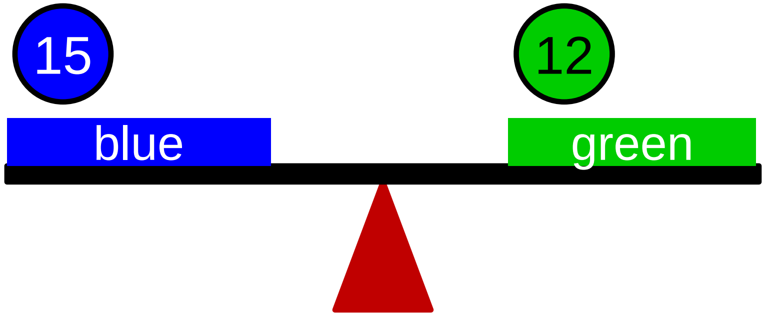
P(**E$_G$** | **G**)= 4/5 (correctly identified)
P(**E$_G$** | **B**)= 1/5 (incorrectly identified)



New odds

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

Still 5:4 odds that the car was **B**!



Hit-and-run: full calculation

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

$P(E_G \mid G) = 4/5$ $P(E_G \mid B) = 1/5$

- 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:

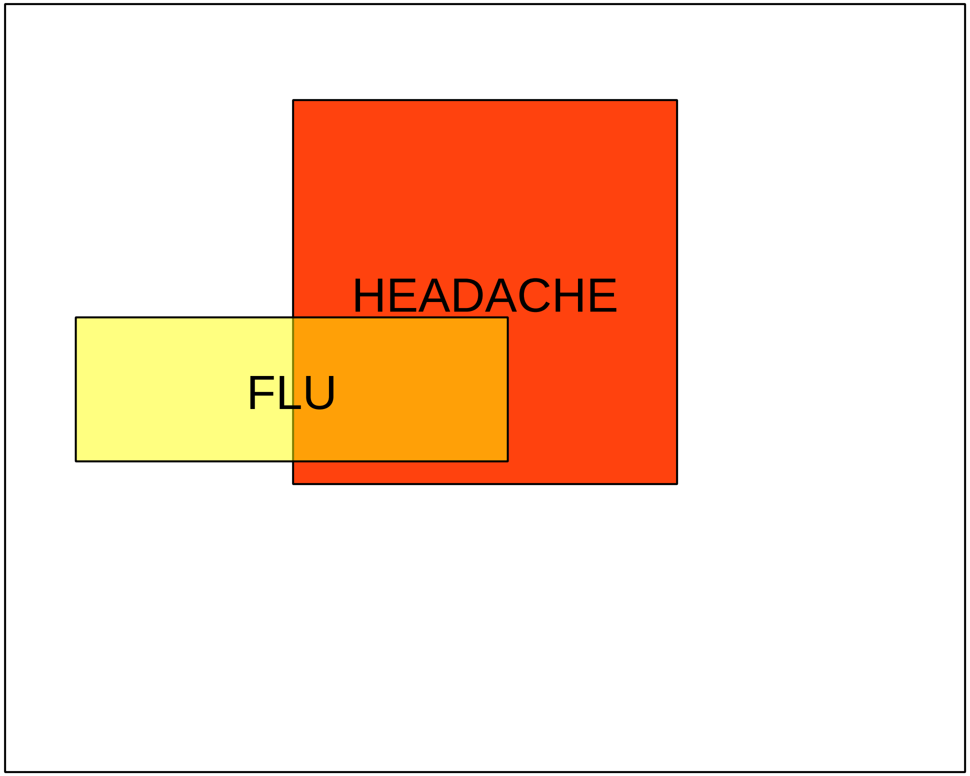
$P(B|E_G) = P(B) * P(E_G|B) / P(E_G) = [5/6 * 1/5] / P(E_G) = 5/30P(E_G)$

//- 5 parts of $30P(X_G)$

Bayes in 'real' life. Example 1

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

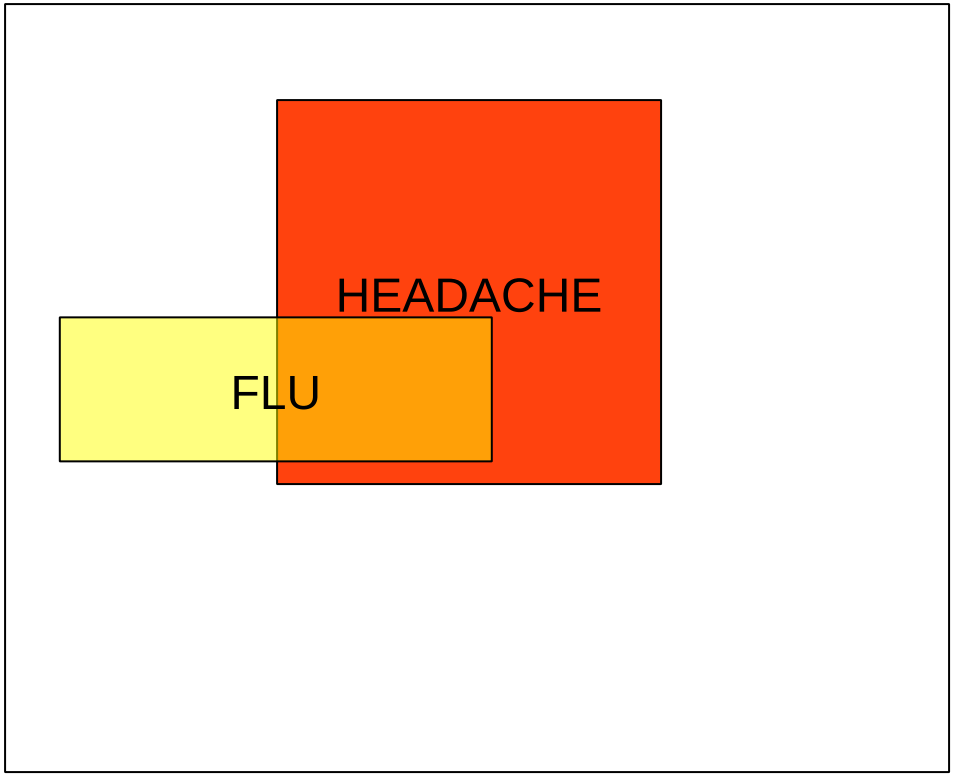
$P(F|H) =?$



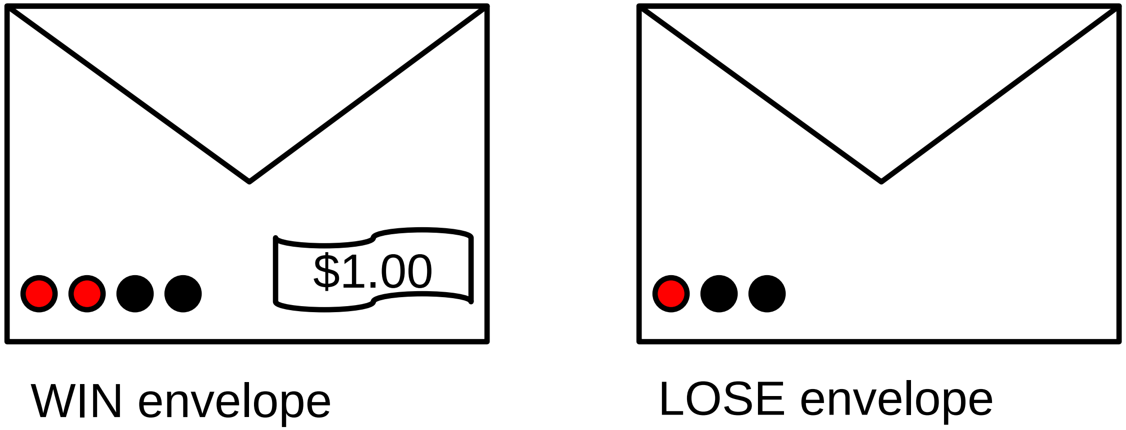
Bayes in 'real' life. Example 1

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

$P(F|H) = P(H|F)P(F)/P(H)$
$=1/2*1/40*10=1/8$

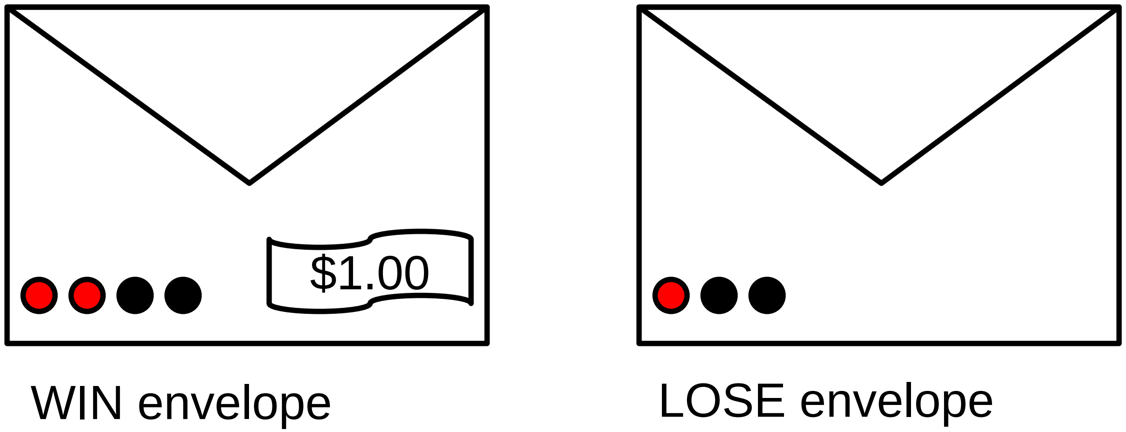


Bayes in ‘real’ life. Example 2



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

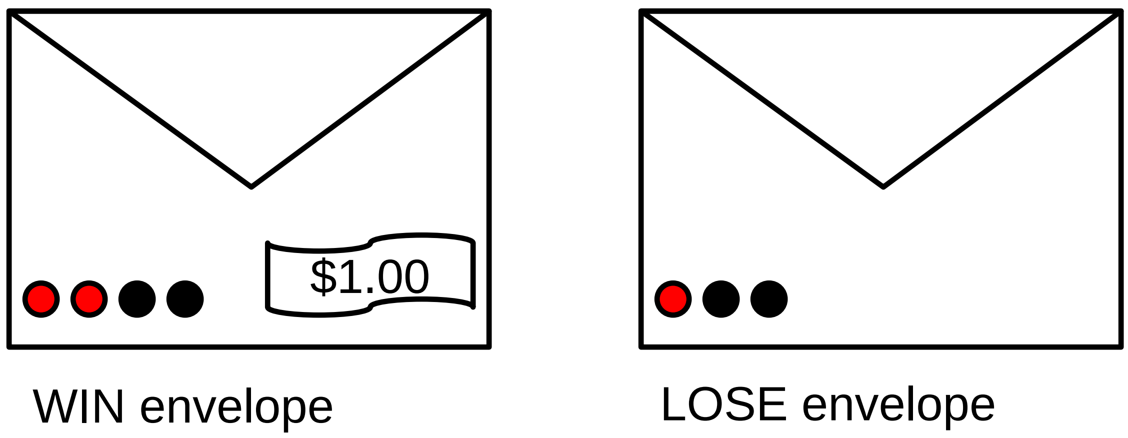


Variant: before deciding, you are allowed to see one bead drawn from the envelope.

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" or "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*

Outline

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

Need for probabilistic learners

- Given the evidence (data),

can we certainly derive

the diagnostic classification 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 association 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

Classifier based on Bayes rule

- Given data – evidence - we can build a classifier which will classify a new record as class *C* (yes or no) by comparing probabilities of yes and of no

- In this case all the attributes except *C* are evidences *E*

- The machine learning task is to evaluate $P(E|C)$ from historical data and based on $P(E|C)$ and prior probabilities $P(C=\text{Yes})$ and $P(C=\text{No})$ compare $P(C=\text{Yes}|E)$ and $P(C=\text{No}|E)$ using Bayes rule.

Single-evidence classifier: priors

event (class)

Humidity	Play
High	No
High	No
High	Yes
High	Yes
Normal	Yes
Normal	No
Normal	Yes
High	No
Normal	Yes
Normal	Yes
Normal	Yes
High	Yes
Normal	Yes
High	No

- Prior probabilities:
  P(Play=yes)=9/14, P(play=no)=5/14

- From recording only 'play'/'not play' we have 5:9 odds for play to be canceled today

Single-evidence classifier: humidity

evidence Humidity	event (class) Play
High	No
High	No
High	Yes
High	Yes
Normal	Yes
Normal	No
Normal	Yes
High	No
Normal	Yes
Normal	Yes
Normal	Yes
High	Yes
Normal	Yes
High	No

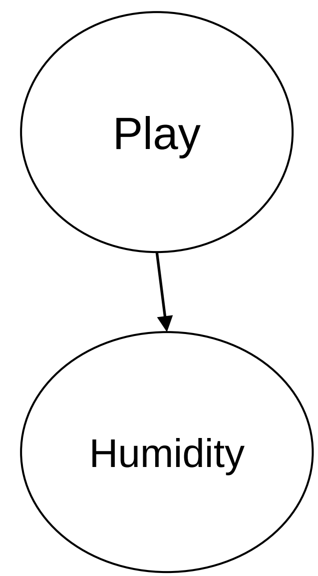
- Priors: P(Play=yes)=9/14, P(play=no)=5/14
- After adding evidence about Humidity we have:

How many times Humidity=normal out of all 9 Yes's:

P(normal|yes)=6/9

How many times Humidity=normal out of all 5 No's:

P(normal|no)=1/5



- Similarly:

P(high|yes)=3/9

P(high|no)=4/5

Single-evidence classifier: prediction

evidence Humidity	event (class) Play
High	No
High	No
High	Yes
High	Yes
Normal	Yes
Normal	No
Normal	Yes
High	No
Normal	Yes
Normal	Yes
Normal	Yes
High	Yes
Normal	Yes
High	No

- P(yes)=9/14, P(no)=5/14
- P(high|yes)=3/9
- P(high|no)=4/5

Today is a high humidity day, what is the probability to play?

- P(yes|high)=P(yes)*P(high|yes)/P(high)
- P(no|high)=P(no)*P(high|no)/P(high)

Single-evidence classifier: prediction

evidence **Humidity**	event (class) **Play**
High	No
High	No
High	Yes
High	Yes
Normal	Yes
Normal	No
Normal	Yes
High	No
Normal	Yes
Normal	Yes
Normal	Yes
High	Yes
Normal	Yes
High	No

P(yes)=9/14, P(no)=5/14

P(high|yes)=3/9

P(high|no)=4/5

Today is a high humidity day, what is the probability to play?

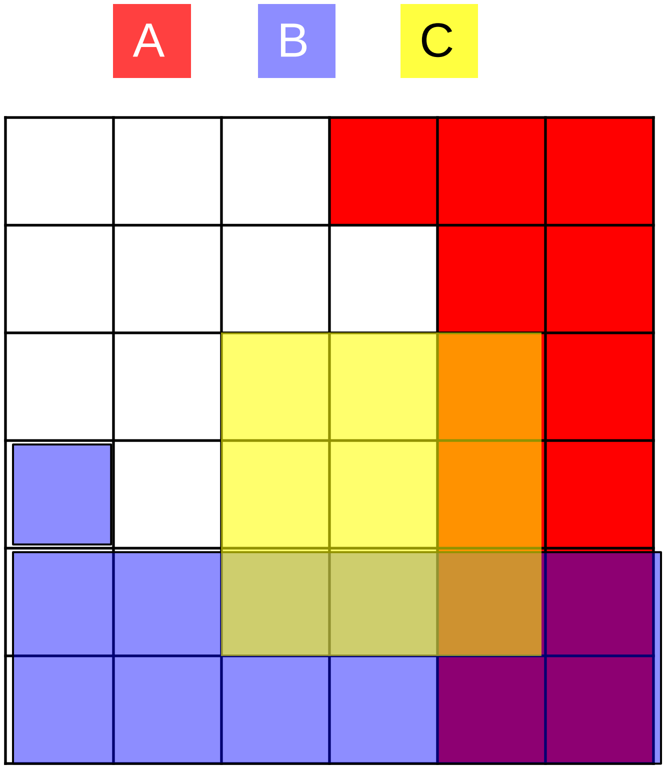
P(yes|high)=P(yes)*P(high|yes)/P(high) = [9/14*3/9] * 1/P(high) = 3/14 α

P(no|high)=P(no)*P(high|no)/P(high) = [5/14*4/5] * 1/P(high) = 4/14 α

3:4 odds to play given high humidity

(vs. 9:5 before the evidence)

Conditional independence of 2 variables given the third



P(A ∩ B | C) = 1/9

P(A ∩ B) = 4/36 = 1/9

P(A ∩ B | C) = P(A ∩ B)

Definition of conditional independence

P(A|C) = 3/9

P(B|C) = 3/9

P(A ∩ B | C) = P(A|C)*P(B|C)

A and B are independent in the world where C is True (C is known, it occured)

However, in general A and B are not independent:

P(A) = 13/36 ≈ 0.36, and P(A|B) = 4/13 ≈ 0.31 → P(A) ≠ P(A|B)

Bayes' rule – two evidences

$$P(class = B|evidence1,evidence2) = \frac{P(evidence1,evidence2|class=B) * P(class=B)}{P(evidence1,evidence2)}$$

If *evidence1* is ***conditionally** independent* of *evidence2* given class value:

$$P(class = B|evidence1,evidence2) = \frac{P(evidence1|class=B) * P(evidence2|class=B)P(class=B)}{P(evidence1,evidence2)}$$

Comparing probability of 2 classes

$$P(\text{class} = A|\text{evidence1}, \text{evidence2})$$

$$= \frac{P(\text{evidence1}|\text{class=A}) * P(\text{evidence2}|\text{class=A}) * P(\text{class=A})}{P(\text{evidence1}) * P(\text{evidence2})}$$

$$= \propto P(\text{evidence1}|\text{class} = A) * P(\text{evidence2}|\text{class} = A) * P(\text{class} = A)$$

The same – let's call it 1/α

$$P(\text{class} = B|\text{evidence1}, \text{evidence2})$$

$$= \frac{P(\text{evidence1}|\text{class=B}) * P(\text{evidence2}|\text{class=B}) * P(\text{class=B})}{P(\text{evidence1}) * P(\text{evidence2})}$$

$$= \propto P(\text{evidence1}|\text{class} = B) * P(\text{evidence2}|\text{class} = B) * P(\text{class} = B)$$

This approach only holds if **we assume conditional independence** between evidence1, evidence2

Generalized

for *N* evidences

$$P(\text{class} = A|\text{evidence1}, \text{evidence2}, \ldots, \text{evidenceN})$$

$$= \frac{P(\text{evidence1}|\text{class=A}) * \cdots * P(\text{evidence}N|\text{class=A}) * P(\text{class=A})}{P(\text{evidence1}) * \cdots * P(\text{evidence}N)}$$

$$= \propto P(\text{evidence1}|\text{class} = A) * \cdots * (\text{evidenceN}|\text{class} = A) * P(\text{class} = A)$$

- Two assumptions:

  Attributes (evidences) are:

  – equally important

  – conditionally independent (given the class value)

- This means that knowledge about the value of a particular attribute doesn’t tell us anything about the value of another attribute **given the class value** (inside the same class)

Naïve Bayes classifier

To predict class value for a set of attribute values (evidences) - for each class value A_i compute and compare:

$$P(\text{class} = A|\text{evidence1}, \text{evidence2}, \ldots, \text{evidenceN})$$

$$= \frac{P(\text{evidence1}|\text{class=A}) * \cdots * P(\text{evidence}N|\text{class=A}) * P(\text{class=A})}{P(\text{evidence1}) * \cdots * P(\text{evidence}N)}$$

$$= \propto P(\text{evidence1}|\text{class} = A) * \cdots * (\text{evidenceN}|\text{class} = A) * P(\text{class} = A)$$

- **Naïve – because it assumes *conditional* independence of variables**

- Although based on assumptions that are almost never correct, this scheme works well in practice!

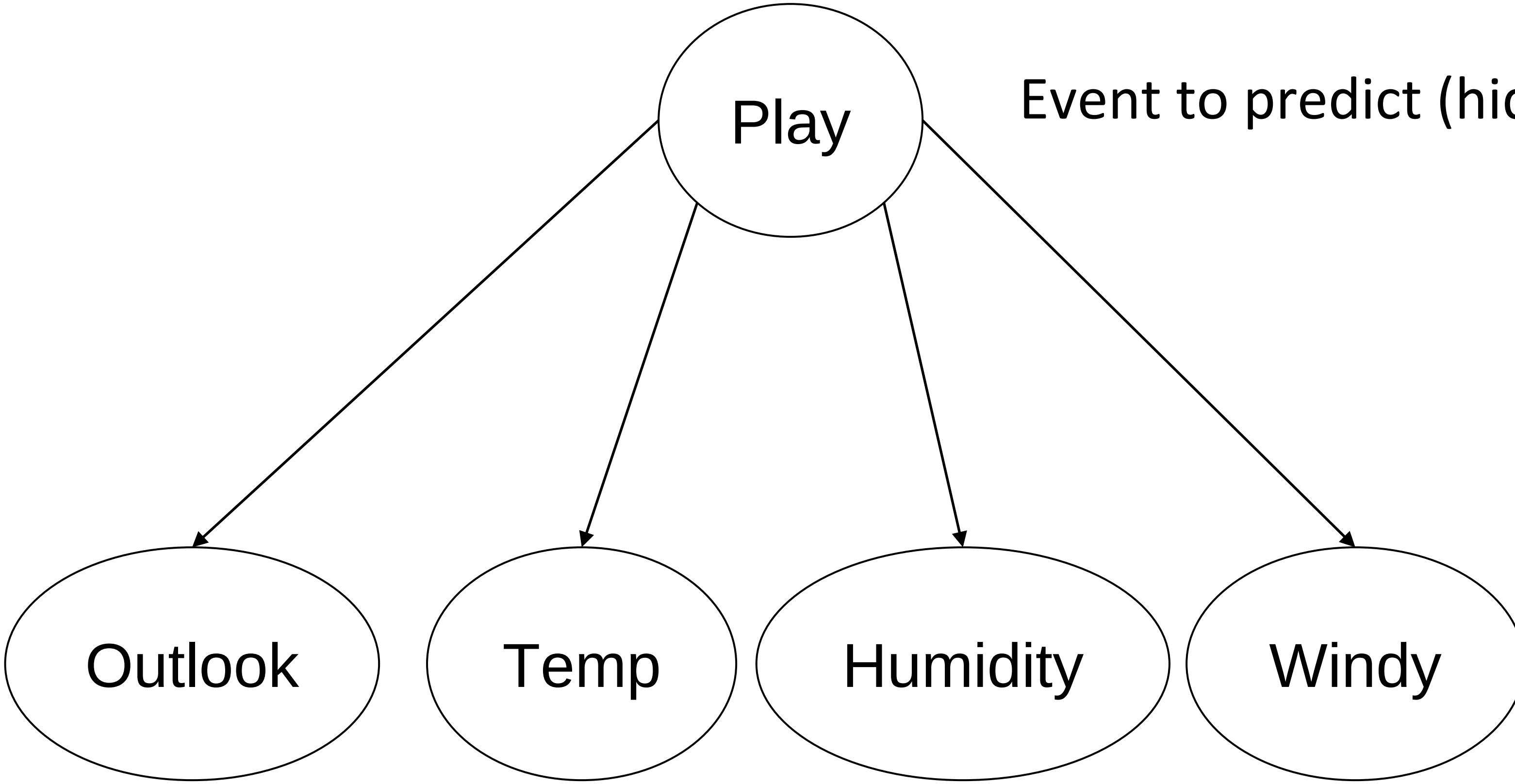
The weather data example

Outlook	Temp.	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

- A new day:

Outlook	Temp.	Humidity	Windy	Play
Sunny	Cool	High	True	?

Multi-evidence classifier



Event to predict (hidden)

Set of evidences (demonstrate themselves)

The weather data example: probabilities

Outlook	Temp.	Humidity	Windy	Play
Sunny	Cool	High	True	?

← *Evidence E*

Play	Sunny	Cool	High humidity	Windy= true
Yes: 9	**2**/9	**3**/9	**3**/9	**3**/9
No: 5	**3**/5	**1**/5	**4**/5	**3**/5
Total	**5**	**4**	**7**	**6**

Outlook	Temp.	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

The weather data example: yes

Outlook	Temp.	Humidity	Windy	Play
Sunny	Cool	High	True	?

← *Evidence E*

P(yes | E) =
P(Sunny | yes) *
P(Cool | yes) *
P(Humidity=High | yes) *
P(Windy=True | yes) *
P(yes) / P(E) =
= (2/9) *
(3/9) *
(3/9) *
(3/9) *
(9/14) / P(E) = 0.0053 / P(E)

Play	Sunny	Cool	High humidity	Windy= true
Yes: 9	2/9	3/9	3/9	3/9
No: 5	3/5	1/5	4/5	3/5
Total	5	4	7	6

Don’t worry about the 1/P(E):
It’s alpha - the normalization constant.

The weather data example: no

Outlook	Temp.	Humidity	Windy	Play
Sunny	Cool	High	True	?

← *Evidence E*

P(no | E) =

P(Sunny | no) *

P(Cool | no) *

P(Humidity=High | no) *

P(Windy=True | no) *

P(no) / P(E) =

= (3/5) *

(1/5) *

(4/5) *

(3/5) *

(5/14) / P(E) = 0.0206 / P(E)

Play	Sunny	Cool	High humidity	Windy= true
Yes: 9	2/9	3/9	3/9	3/9
No: 5	3/5	1/5	4/5	3/5
Total	5	4	7	6

The weather data example: decision

Outlook	Temp.	Humidity	Windy	Play
Sunny	Cool	High	True	?

← *Evidence E*

P(yes | E) = 0.0053 / P(E)

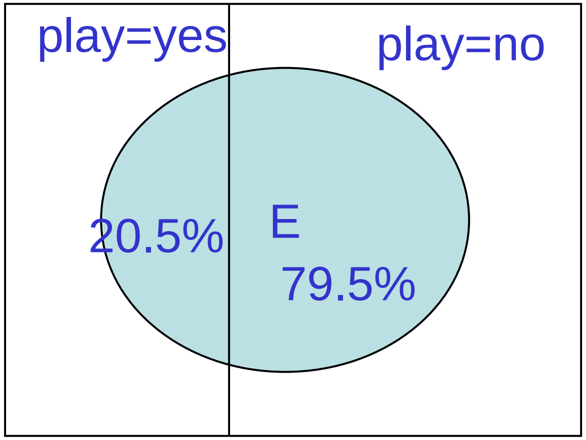
P(no | E) = 0.0206 / P(E)

More probable: no.

It would be nice to give the actual probability estimates

Outlook	Temp.	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

Normalization constant 1/P(E)



P(play=yes | E) + P(play=no | E) = 1   i.e.

0.0053 / P(E) + 0.0206 / P(E) = 1   i.e.

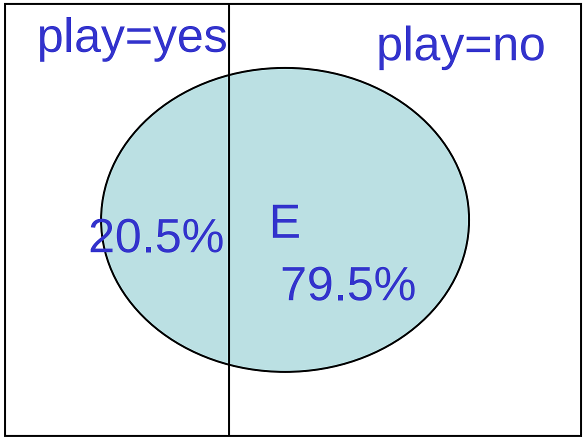
P(E) = 0.0053 + 0.0206

So,

P(play=yes | E) = 0.0053 / (0.0053 + 0.0206) = **20.5%**

P(play=no | E) = 0.0206 / (0.0053 + 0.0206) = **79.5%**

In other words:



P(play=yes | E) + P(play=no | E) = 1

P(play=yes |E) / P (play=no | E) = 0.0053 : 0.0206 = 0.26

0.26 * P (play=no | E) + P (play=no | E) = 1

P (play=no | E) = 1/1.26 = 79%

The remaining goes to yes: P(play=yes |E) = 21%