Cooperating Intelligent Systems

Uncertainty & probability

Chapter 13, AIMA

”When an agent knows enough facts about its environment, the logical approach enables it to derive plans that are guaranteed to work.”

”Unfortunately, agents almost never have access to the whole truth about the environment.”

From AIMA, chapter 13

Airport example

Let action A_t = leave for airport *t* minutes before the flight.

Will A_t get me there on time?

Problems:

•partial observability (road state, other drivers' plans, vehicle condition, ...)
•noisy sensors (KCBS traffic reports, news, ...)
•uncertainty in action outcomes (flat tire, missing an exit, ...)
•immense complexity in modelling and predicting state of the environment

Hence a purely logical approach either:

1)risks falsehood: *"A25 will get me there on time"*, or
2)leads to conclusions that are too weak for decision making: *"A25 will get me there on time if there's no accident on the bridge and it doesn't rain and my tires remain intact and ..."*

(A1440 might reasonably be said to get me to there on time but I'd have to stay overnight in the airport... it's not a good solution, even if it is "correct" in some sense).

Slide adapted from S. Russell

Why FOL fails?

Laziness: People who want to solve problems can’t be bothered to list all possible requirements

- FOL is often the very definition of „useless solution”

Ignorance: We have no theoretical models for the domain that would be complete

- and what we have isn't usually tested enough

Practical reasons: Even if we knew everything and could be persuaded to list everything, the rules would be totally impractical to use

- we can’t possibly account for / sense / test everything

Instead, use decision theory

Decision theory = probability theory + utility theory

Probability

- Assign probability to a proposition based on the percepts, i.e. the information the agent has.
- Proposition is either true or false. Probability means assigning a value that indicates how much do we believe in it being true/false.
- *Evidence* is all information that the agent receives. Probabilities can (will) change when more evidence is acquired.
- Prior/unconditional probability ⇔ no evidence at all.
- Posterior/conditional probability ⇔ after evidence is obtained.

Utility

- Plan does not need to be guaranteed to achieve the goal.
- To make choices, the agent must have preferences between the different possible outcomes of various plans.
- Utility represents the value of the outcomes, and utility theory is used to reason with the resulting preferences.
- An agent is rational if and only if it chooses the action that yields the highest expected utility, when averaged over all possible outcomes.

Inspired by Sun-Hwa Hahn

Probability

The dealer will pay $1 if you flip a coin and it lands head up...

How much will you pay to play this game?

Most people agree that they would pay up to $0.50

Probability

The dealer will pay $2 if you roll a die and it lands with a 6 up...

How much will you pay to play this game?

Most people agree that they would pay up to $0.33

Probability

The dealer will pay \$2 if the card you draw has a rank at least as high as the rank of the card he draws....

How much will you pay to play this game?

$$52*3 + (52*51 - 52*3) / 2 = 52 * 27$$

$$52 * 27/ (52 * 51) = 27 / 51$$

$$\$2 * 27 / 52 = \$1.03846$$

Symmetry

In principle, according to classical physics, we should be able to predict how the coin will land, if we know its initial position, the force exerted upon it by the flipper, the position of the surface on which it lands, the material properties of the coin and that surface, the air pressure, any winds that blow through the trajectory of the coin, etc.

We can apply Newton's laws of motion and the law of gravity, while accounting for the elasticity of the collisions between the coin and the surface, and for air friction, and for whatever other physical effects there are, to calculate the coin's motion until it stops.

In practice, of course, these calculations are too difficult for us to do exactly, at least in our heads, even if we knew all the relevant factors.

And furthermore, we do not know all of these factors.

Basic Principle of Counting

But there is a symmetry in the problem:

none of these factors differ significantly with the different sides of the coin

Thus the amount we are willing to pay to win \$1 if the coin lands *head up* is the same amount that we are willing to pay to win \$1 if the coin lands *tail up*

We believe that coin will land head up or tail up

$p + p = 1, \; p = 0.5$

We believe that a die will show a number 1..6

$6 * p = 1, \; p = 0.1666666666(6)$

Interpretation of Probability

Classical interpretation

The probability of an event is the ratio of the number of cases favorable to it, to the number of all cases possible when nothing leads us to expect that any one of these cases should occur more than any other

Frequency interpretation

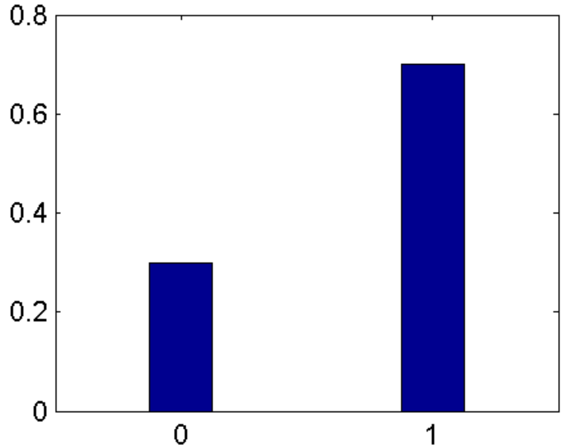
The probability of an event is the limit of its relative frequency in an infinitely large number of trials

Bayesian interpretation

The probability of an event depends on some prior probability, which is then updated in the light of new relevant data or observations

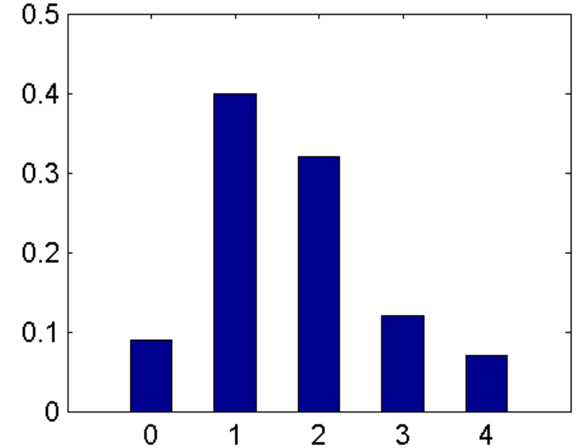
Basic probability notation

X : Random variable



Boolean: $P(X\text{=true}), P(X\text{=false})$

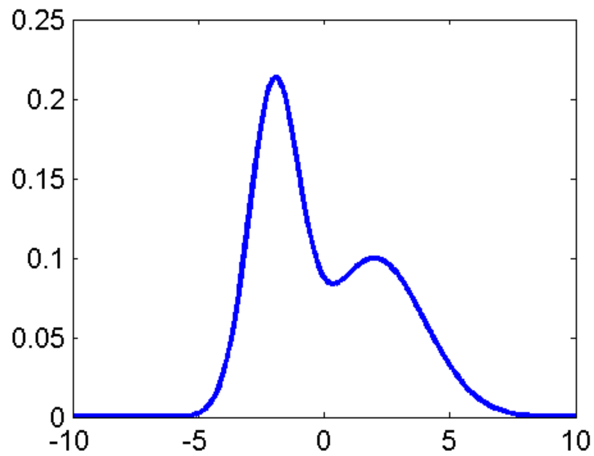
$P(X), P(\neg X)$



Discrete: $P(X{=}a), P(X{=}b), \ldots$

$P(a), P(b), \ldots$

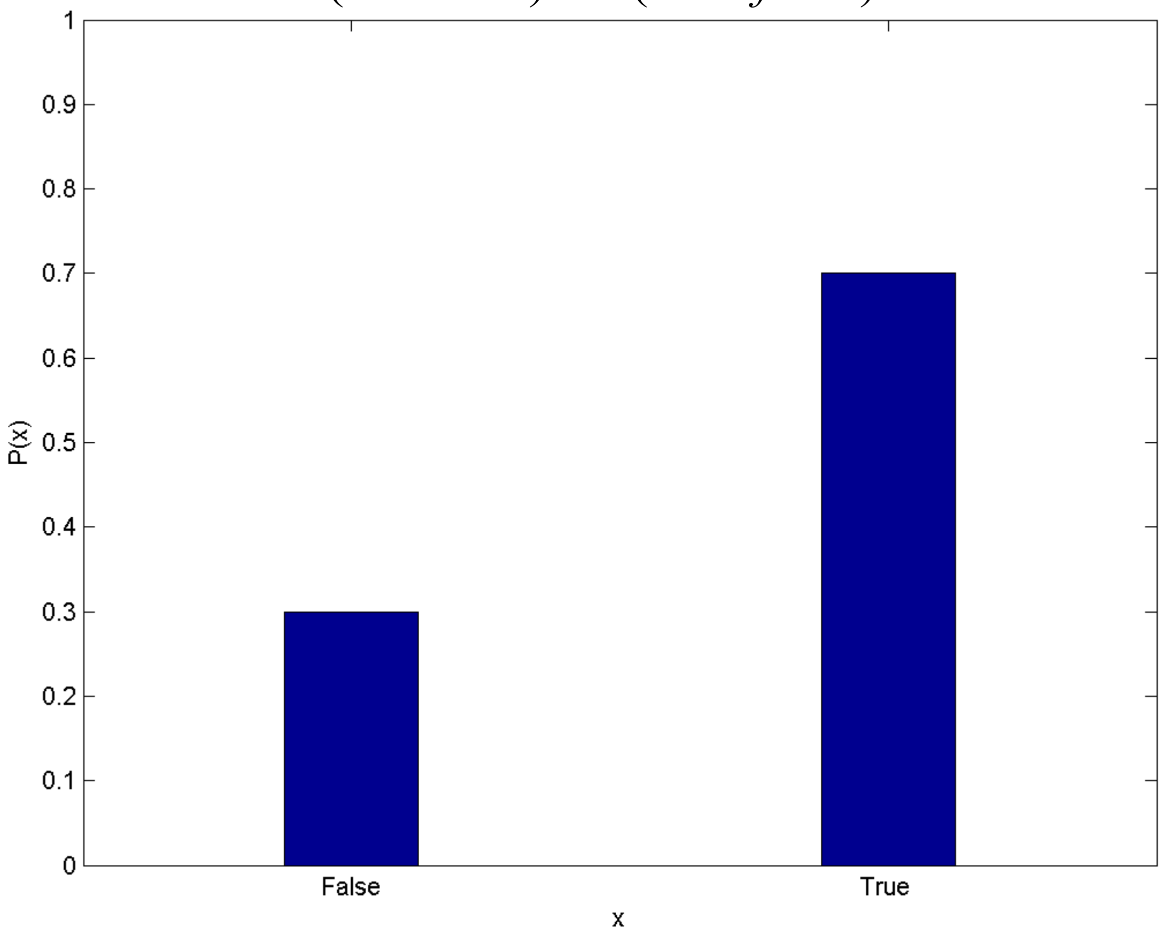
Continuous: $P(x)dx$



Boolean variable

$$X \in \{True, False\}$$

$$P(X = true) + P(X = false) = 1$$



Examples:

$P(W_{31})$

P(Survive)

P(CatchFlight)

P(Cancer)

P(Cavity)

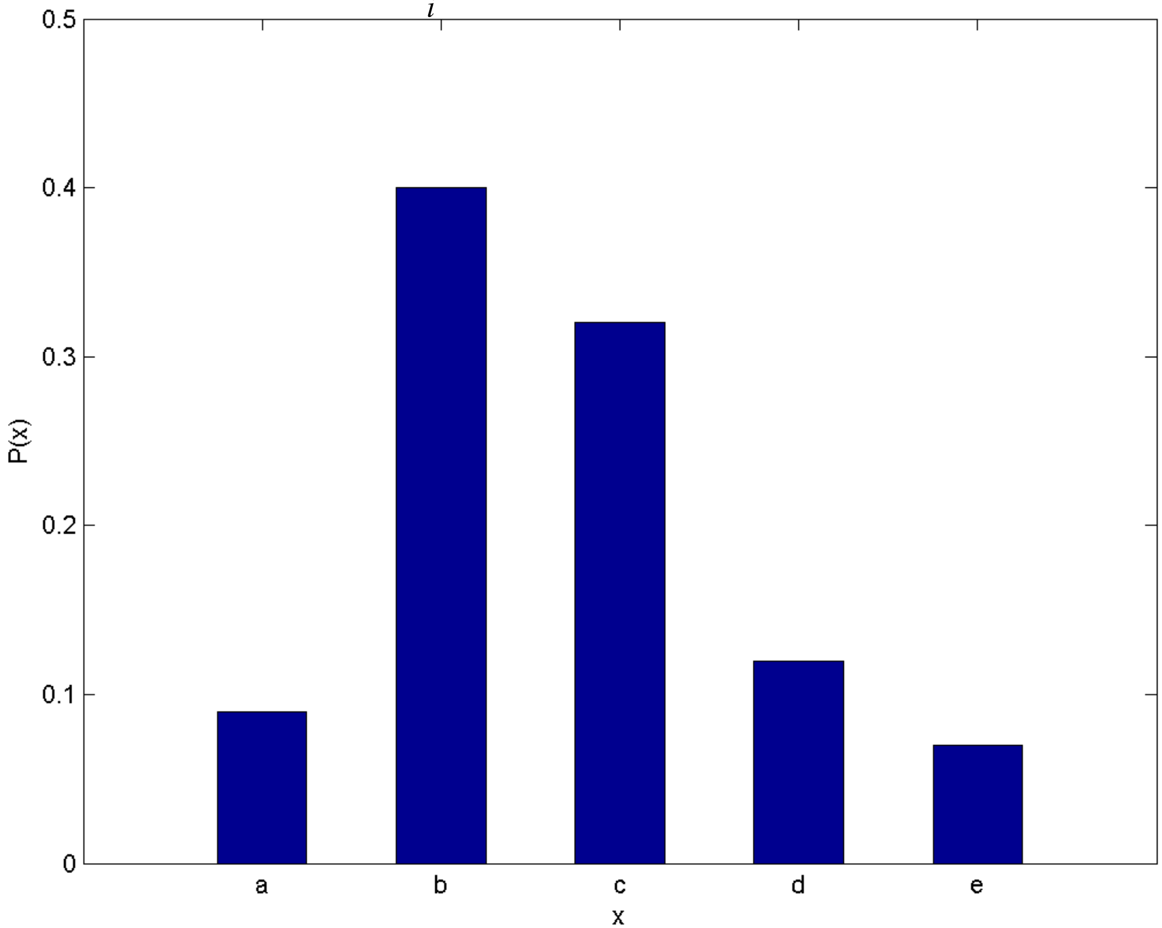
P(LectureToday)

…

Discrete variable

$$X \in \{a_1, a_2, a_3, a_4, \ldots\}$$

$$\sum_i P(X = a_i) = 1$$



Examples:

P(WumpusPos)

P(Weather)

P(Color)

P(Age)

P(LectureQual)

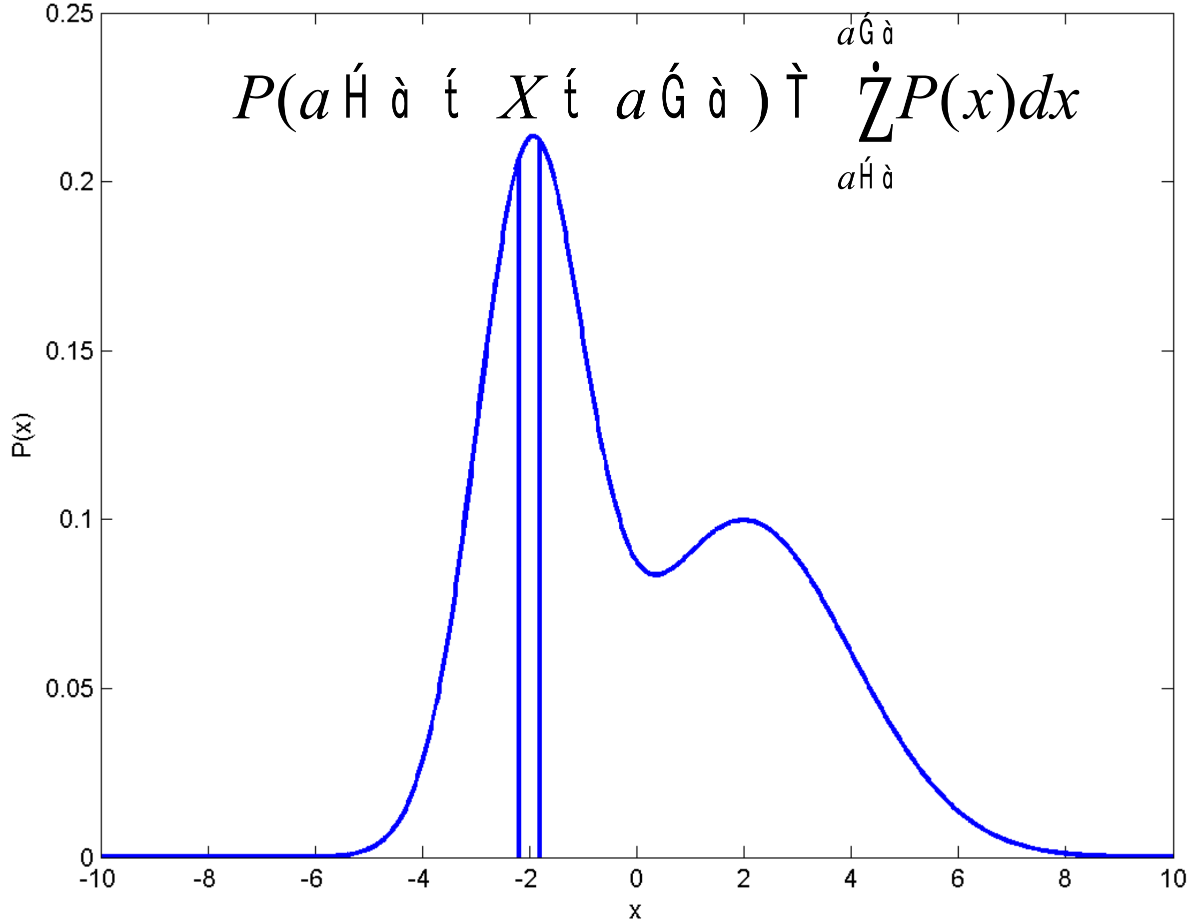
P(PlayingCard)

…

Continuous variable X

Probability density $P(x)$

$$\int_{-\infty}^{\infty} P(x)\,dx = 1$$



$$P(a-\delta < X < a+\delta) = \int_{a-\delta}^{a+\delta} P(x)\,dx$$

Examples:

P(Temperature)
P(WindSpeed)
P(OxygenLevel)
P(LengthLecture)
...

Probability Axioms

The probability of an event is a non-negative real number:

$$P(E) \in \mathbb{R} \wedge P(E) \geq 0 \qquad \forall E \in F$$

The probability that *some* elementary event in the entire sample space will occur is 1:

$$P(\Omega) = 1$$

Any countable sequence of pairwise disjoint (i.e. mutually exclusive) events $E_1, E_2, \ldots$ satisfies:

$$P(E_1 \cup E_2 \cup \cdots) = \sum_{i=1}^{\infty} P(E_i).$$

From Wikipedia

Properties of Probability

Monotonicity:

$$P(A) \leq P(B) \quad \text{if} \quad A \subseteq B.$$

The probability of the empty set:

$$P(\emptyset) = 0.$$

The numeric bound:

$$0 \leq P(E) \leq 1 \qquad \text{for all } E \in F.$$

The addition law of probability:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

The inclusion-exclusion principle:

$$P(\Omega \setminus E) = 1 - P(E)$$

From Wikipedia

Expected Value

$$\mathrm{E}[X] = x_1 p_1 + x_2 p_2 + \ldots + x_k p_k \ .$$

$$\mathrm{E}[X] = \sum_{i=1}^{\infty} x_i\, p_i,$$

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x)\, \mathrm{d}x.$$

From Wikipedia

Propositions

Elementary:

Cavity = True, W_{31} = True, Cancer = False,

Card = A♠, Card = 4♥, Weather = Sunny,

Age = 40, (20ĶC < Temperature < 21ĶC),

(2 hrs < LengthLecture < 3 hrs)

Complex: (Elementary + connective)

¬Cancer ∧ Cavity

Card_1 = A♠ ∧ Card_2 = A♣

Sunny ∧ (30ĶC < Temperature) ∧ ¬Beer

Propositions

Cavity = True, W_{31} = True, Cancer = False

2 hrs < LengthLecture < 3 hrs

$Card_1$ = A♠ ∧ $Card_2$ = A♣

Sunny ∧ (30ĶC < Temperature) ∧ ¬Beer

In the physical world, every of those propositions is either true or false

– we are not talking about degrees of truth here (that's called *fuzzy logic)*

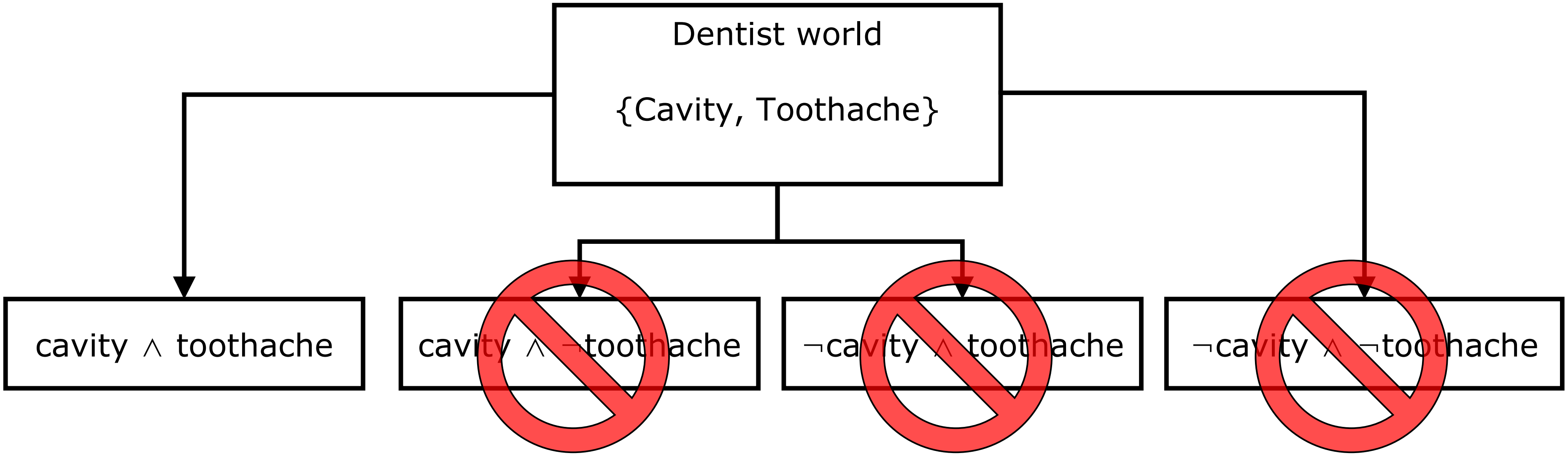
But we can discuss different degrees to which we believe that various propositions are true or not

Extension of propositional calculus...

Atomic event

is a complete specification of one of the possible states of the world/environment

- Mutually exclusive
- Exhaustive



Prior & posterior

- **Prior probability** P(X=a) ≡ P(a)
  Our belief in X=a being true before any information is collected

- **Posterior probability** P(X=a|Y=b) ≡ P(a|b)
  Our belief that X=a is true when we know that Y=b is true
  (and we don't know anything else)

- **Joint probability** P(X=a, Y=b) ≡ P(a ∧ b) ≡ P(a,b)
  Our belief that (X=a ∧ Y=b) is true.
  P(a,b) = P(a|b)P(b)

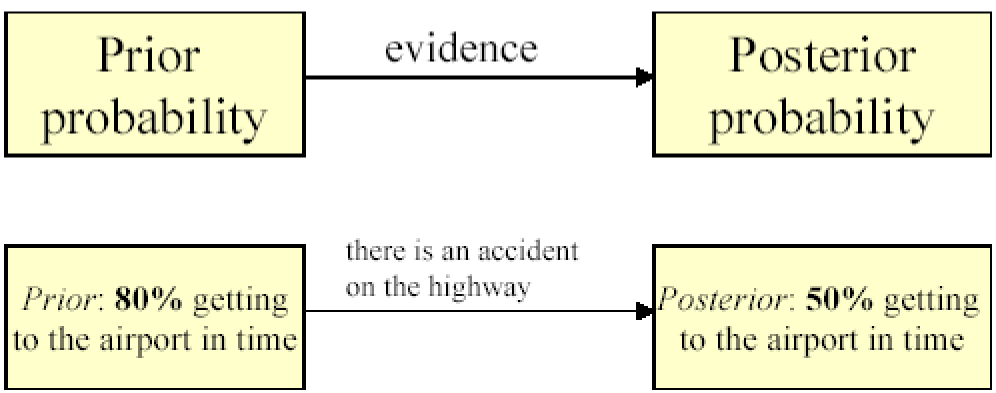
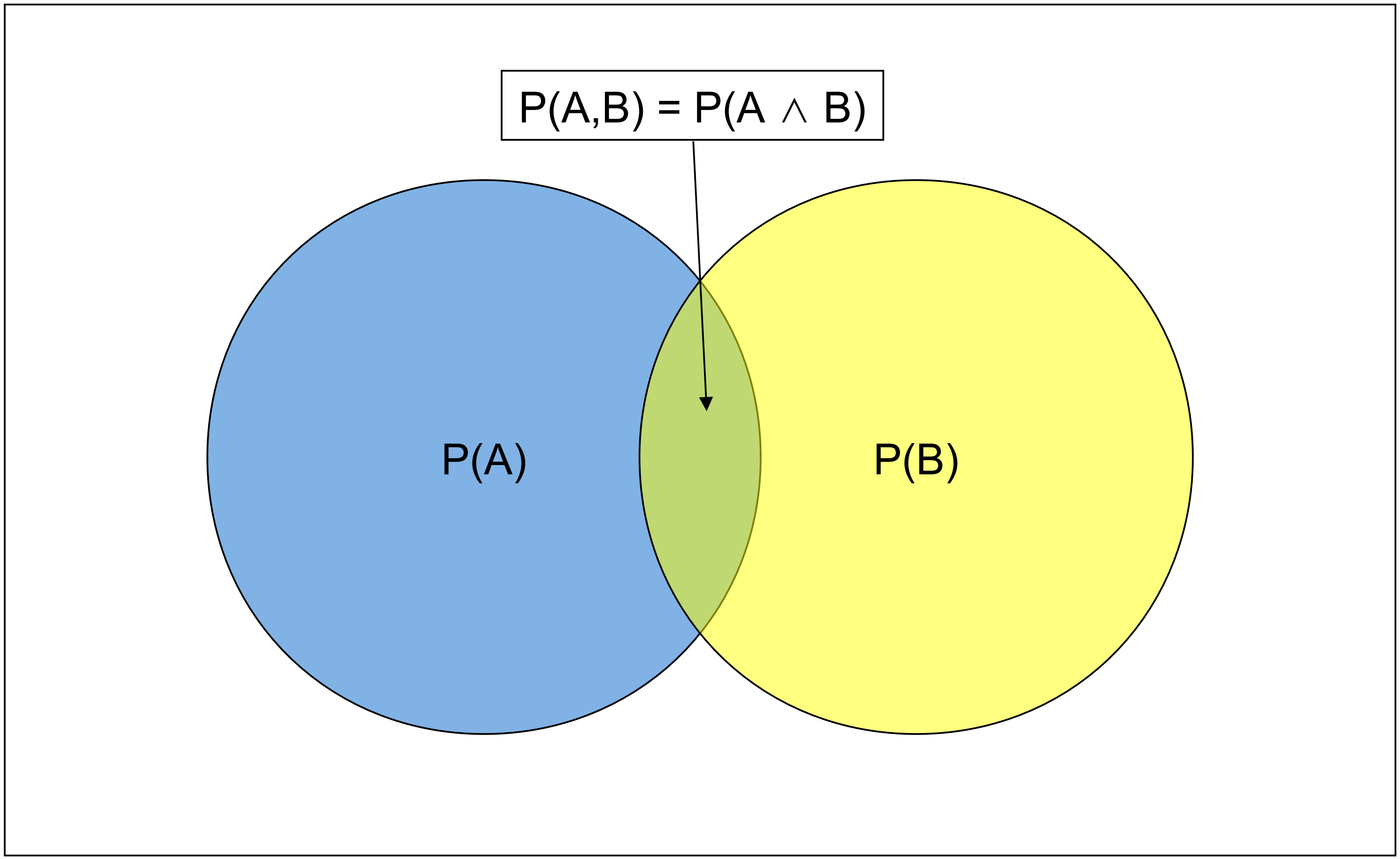


Image borrowed from Lazaric & Sceffer

Conditional probability
(Venn diagram)



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

Conditional probability examples

1. You draw two cards randomly from a deck of 52 playing cards. What is the conditional probability that the second card is an ace if the first card is an ace?

2. In a course I'm giving with oral examination the examination statistics over the period 2002-2005 have been: 23 have passed the oral examination in their first attempt, 25 have passed it their second attempt, 7 have passed it in their third attempt, and 8 have not passed it at all (after having failed the oral exam at least three times). What is the conditional probability (risk) for failing the course if the student fails the first two attempts at the oral exam?

3. (2005) 84% of the Swedish households have computers at home, 81% of the Swedish households have both a computer and internet. What is the probability that a Swedish household has internet if we know that it has a computer?

Conditional probability examples

1. You draw two cards randomly from a deck of 52 playing cards. What is the conditional probability that the second card is an ace if the first card is an ace?

$$P(2\text{nd ace} \,|\, 1\text{st ace}) = \frac{3}{51}$$

What's the probability that both are aces?

2. In a course I'm giving with oral examination the examination statistics over the period 2002-2005 have been: 23 have passed the oral examination in their first attempt, 25 have passed it their second attempt, 7 have passed it in their third attempt, and 8 have not passed it at all (after having failed the oral exam at least three times). What is the conditional probability (risk) for failing the course if the student fails the first two attempts at the oral exam?

3. (2005) 84% of the Swedish households have computers at home, 81% of the Swedish households have both a computer and internet. What is the probability that a Swedish household has internet if we know that it has a computer?

Conditional probability examples

1. You draw two cards randomly from a deck of 52 playing cards. What is the conditional probability that the second card is an ace if the first card is an ace?

$$P(2 = ace|1 = ace) = \frac{3}{51} \qquad P(2 = \text{ace}, 1 = \text{ace}) = P(2 = \text{ace} \mid 1 = \text{ace})P(1 = \text{ace}) = \frac{3}{51} \cdot \frac{4}{52} = 0.5\%$$

2. In a course I'm giving with oral examination the examination statistics over the period 2002-2005 have been: 23 have passed the oral examination in their first attempt, 25 have passed it their second attempt, 7 have passed it in their third attempt, and 8 have not passed it at all (after having failed the oral exam at least three times). What is the conditional probability (risk) for failing the course if the student fails the first two attempts at the oral exam?

3. (2005) 84% of the Swedish households have computers at home, 81% of the Swedish households have both a computer and internet. What is the probability that a Swedish household has internet if we know that it has a computer?

Conditional probability examples

1. You draw two cards randomly from a deck of 52 playing cards. What is the conditional probability that the second card is an ace if the first card is an ace?

$$P(2 = \text{ace} \,|\, 1 = \text{ace}) = \frac{3}{51} \qquad P(2 = \text{ace}, 1 = \text{ace}) = P(2 = \text{ace} \,|\, 1 = \text{ace})P(1 = \text{ace}) = \frac{3}{51} \cdot \frac{4}{52} = 0.5\%$$

2. In a course I'm giving with oral examination the examination statistics over the period 2002-2005 have been: 23 have passed the oral examination in their first attempt, 25 have passed it their second attempt, 7 have passed it in their third attempt, and 8 have not passed it at all (after having failed the oral exam at least three times). What is the conditional probability (risk) for failing the course if the student fails the first two attempts at the oral exam?

$$P(\text{Fail course} \,|\, \text{Fail OE 1 \& 2}) = \frac{P(\text{Fail course, Fail OE1 \& 2})}{P(\text{Fail OE1 \& 2})} = \frac{8/63}{15/63} = 53\%$$

$$63 = 23 + 25 + 7 + 8$$

$$15 = 63 - 23 - 25$$

3. (2005) 84% of the Swedish households have computers at home, 81% of the Swedish households have both a computer and internet. What is the probability that a Swedish household has internet if we know that it has a computer?

Conditional probability examples

1. You draw two cards randomly from a deck of 52 playing cards. What is the conditional probability that the second card is an ace if the first card is an ace?

$$P(2 = \text{ace} \mid 1 = \text{ace}) = \frac{3}{51} \qquad P(2 = \text{ace}, 1 = \text{ace}) = P(2 = \text{ace} \mid 1 = \text{ace})P(1 = \text{ace}) = \frac{3}{51} \times \frac{4}{52} = 0.5\%$$

2. In a course I'm giving with oral examination the examination statistics over the period 2002-2005 have been: 23 have passed the oral examination in their first attempt, 25 have passed it their second attempt, 7 have passed it in their third attempt, and 8 have not passed it at all (a ... es). What is the conditi ... tudent fails the first two attempts at the oral exam?

What's the prior probability for not passing the course?

$$P(\text{Fail course} \mid \text{Fail OE 1 \& 2}) = \frac{P(\text{Fail course, Fail OE1 \& 2})}{P(\text{Fail OE1 \& 2})} = \frac{\boxed{8/63 \; =13\%}}{15/63} = 53\%$$

3. (2005) 84% of the Swedish households have computers at home, 81% of the Swedish households have both a computer and internet. What is the probability that a Swedish household has internet if we know that it has a computer?

Conditional probability examples

1. You draw two cards randomly from a deck of 52 playing cards. What is the conditional probability that the second card is an ace if the first card is an ace?

$$P(2 = \text{ace} \mid 1 = \text{ace}) = \frac{3}{51} \qquad P(2 = \text{ace}, 1 = \text{ace}) = P(2 = \text{ace} \mid 1 = \text{ace})P(1 = \text{ace}) = \frac{3}{51} \cdot \frac{4}{52} = 0.5\%$$

2. In a course I'm giving with oral examination the examination statistics over the period 2002-2005 have been: 23 have passed the oral examination in their first attempt, 25 have passed it their second attempt, 7 have passed it in their third attempt, and 8 have not passed it at all (after having failed the oral exam at least three times). What is the conditional probability (risk) for failing the course if the student fails the first two attempts at the oral exam?

$$P(\text{Fail course} \mid \text{Fail OE } 1 \,\&\, 2) = \frac{P(\text{Fail course, Fail OE1} \,\&\, 2)}{P(\text{Fail OE1} \,\&\, 2)} = \frac{8/63}{15/63} = 53\%$$

3. (2005) 84% of the Swedish households have computers at home, 81% of the Swedish households have both a computer and internet. What is the probability that a Swedish household has internet if we know that it has a computer?

$$P(\text{Internet} \mid \text{Computer}) = \frac{P(\text{Internet, Computer})}{P(\text{Computer})} = \frac{0.81}{0.84} = 96\%$$

Inference

- Inference means computing

  $\mathbf{P}$(State of the world | Observed evidence)

  $\mathbf{P}(Y \mid \mathbf{e})$

For example: The probability for having a cavity if I have a toothache

$$P(cavity \mid toothache)$$

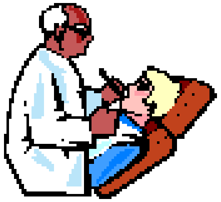
Or

The probability for having a cavity if I have a toothache and the dentist finds a catch in my tooth during inspection

$$P(cavity \mid toothache, catch)$$

Inference w/ full joint distribution

- The **full joint probability distribution** is the probability distribution for all random variables used to describe the world.

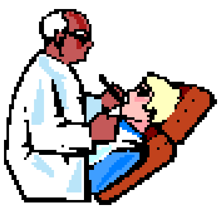


Dentist example {Toothache, Cavity, Catch}

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Inference w/ full joint distribution

- The **full joint probability distribution** is the probability distribution for all random variables used to describe the world.



Dentist example {Toothache, Cavity, Catch}

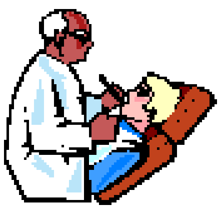
	toothache		¬toothache		
	catch	¬catch	catch	¬catch	
cavity	0.108	0.012	0.072	0.008	0.200
¬cavity	0.016	0.064	0.144	0.576	

P(cavity) = 0.2

Marginal probability (marginalization)

Inference w/ full joint distribution

- The **full joint probability distribution** is the probability distribution for all random variables used to describe the world.



Dentist example {Toothache, Cavity, Catch}

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576
	0.200			

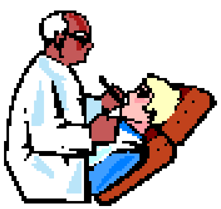
P(toothache) = 0.2

P(cavity) = 0.2

Marginal probability (marginalization)

Inference w/ full joint distribution

- The **full joint probability distribution** is the probability distribution for all random variables used to describe the world.



Dentist example {Toothache, Cavity, Catch}

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z}} \mathbf{P}(\mathbf{Y},\mathbf{z})$$

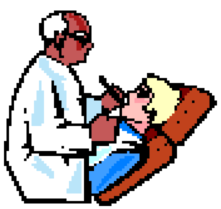
$$\mathbf{P}(\mathbf{Y}) = \sum_{\mathbf{z}} \mathbf{P}(\mathbf{Y} \mid \mathbf{z})\mathbf{P}(\mathbf{z})$$

Marginal probability
(marginalization)

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)} = \frac{\boxed{0.12}}{\boxed{0.2}} = 0.6$$

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)} = \frac{0.08}{0.2} = 0.4$$

$$P(cavity \mid \neg toothache) = \frac{P(cavity \wedge \neg toothache)}{P(\neg toothache)} = \frac{0.08}{0.8} = 0.1$$



Dentist example {Toothache, Cavity, Catch}

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Inference w/ full joint distribution

The general inference procedure

$$\mathbf{P}(Y \mid \mathbf{e}) = \alpha \mathbf{P}(Y, \mathbf{e}) = \alpha \sum_{z} \mathbf{P}(Y, \mathbf{e}, z)$$

where α is a normalization constant. This can always be computed if the full joint probability distribution is available.

$$\mathbf{P}(Cavity \mid toothache) = \alpha \left[\mathbf{P}(Cavity \mid toothache, catch) + \mathbf{P}(Cavity \mid toothache, \neg catch) \right]$$

Completely impossible to do in practice: $O(2^n)$ complexity

Independence

Independence between variables can dramatically reduce the amount of computation.

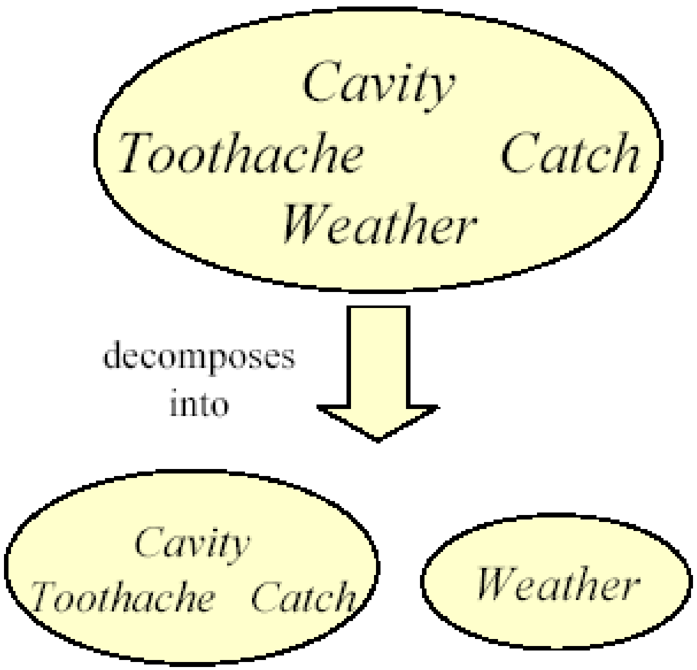
$$\mathbf{P}(X, Y) = \mathbf{P}(X)\mathbf{P}(Y)$$

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

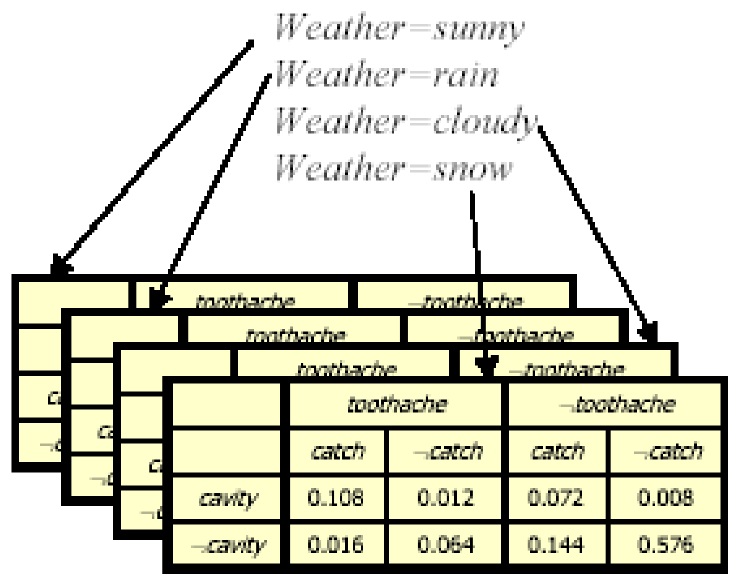
$$\mathbf{P}(Y \mid X) = \mathbf{P}(Y)$$

We don’t need to mix independent variables in our computations

Independence for dentist example



Oral health is independent of weather



decomposes into

	Weather
sunny	0.4
rain	0.15
cloudy	0.4
snow	0.05

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Image borrowed from Lazaric & Sceffer

Bayes' theorem

$$P(A,B) = P(A \mid B)P(B) = P(B \mid A)P(A)$$

$$\boxed{P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}}$$

Bayes theorem example

Joe is a randomly chosen member of a large population in which 3% are heroin users. Joe tests positive for heroin in a drug test that correctly identifies users 95% of the time and correctly identifies nonusers 90% of the time.

Is Joe a heroin addict?

Example from http://plato.stanford.edu/entries/bayes-theorem/supplement.html

Bayes theorem example

Joe is a randomly chosen member of a large population in which 3% are heroin users. Joe tests positive for heroin in a drug test that correctly identifies users 95% of the time and correctly identifies nonusers 90% of the time.

Is Joe a heroin addict?

$$P(H \mid pos) = \frac{P(pos \mid H)P(H)}{P(pos)}$$

$$P(H) = 3\% = 0.03,\ P(\neg H) = 1 - P(H) = 0.97$$

$$P(pos \mid H) = 95\% = 0.95,\ P(pos \mid \neg H) = 10\% = 1 - 0.90$$

$$P(pos) = P(pos \mid H)P(H) + P(pos \mid \neg H)P(\neg H) = 0.1255$$

$$P(H \mid pos) = 0.227 \approx 23\%$$

Example from http://plato.stanford.edu/entries/bayes-theorem/supplement.html

Bayes theorem:
The Monty Hall Game show

In a TV Game show, a contestant selects one of three doors; behind one of the doors there is a prize, and behind the other two there are no prizes.

After the contestant select a door, the game-show host opens one of the remaining doors, and reveals that there is no prize behind it. The host then asks the contestant whether he/she wants to SWITCH to the other unopened door, or STICK to the original choice.

What should the contestant do?

See http://www.io.com/~kmellis/monty.html

The Monty Hall Game Show

prize behind door $\in \{1,2,3\}$, $\text{open}_i \in$ Host opens door i

© Let's make a deal (A Joint Venture)

The Monty Hall Game Show

prize behind door Ì $\{1,2,3\}$, open_i Ì Host opens door i

Contestant selects door 1

Host opens door 2 Ŭ open_2

$$P(1 \mid \text{open}_2) \text{ Ì } \frac{P(\text{open}_2 \mid 1)P(1)}{P(\text{open}_2)} \text{ ì}$$

$$P(3 \mid \text{open}_2) \text{ Ì } \frac{P(\text{open}_2 \mid 3)P(3)}{P(\text{open}_2)} \text{ ì}$$

© Let's make a deal (A Joint Venture)

The Monty Hall Game Show

$\text{prize behind door} = \{1,2,3\},\ \text{open}_i = \text{Host opens door } i$

Contestant selects door 1

$\text{Host opens door 2} \Rightarrow \text{open}_2$

© Let's make a deal (A Joint Venture)

$$P(1 \mid \text{open}_2) = \frac{P(\text{open}_2 \mid 1)P(1)}{P(\text{open}_2)} = 1/3$$

$$P(3 \mid \text{open}_2) = \frac{P(\text{open}_2 \mid 3)P(3)}{P(\text{open}_2)} = 2/3$$

$$P(\text{open}_2) = \sum_{i=1}^{3} P(\text{open}_2 \mid i)P(i) = 1/2$$

$$P(\text{open}_2 \mid 1) = 1/2,\ P(\text{open}_2 \mid 2) = 0,\ P(\text{open}_2 \mid 3) = 1$$

Bayes theorem:
The Monty Hall Game show

In a TV Game show, a contestant selects one of three doors; behind one of the doors there is a prize, and behind the other two there are no prizes.

After the contestant select a door, the game-show host opens one of the remaining doors, and reveals that there is no prize behind it. The host then asks the contestant whether he/she wants to SWITCH to the other unopened door, or STICK to the original choice.

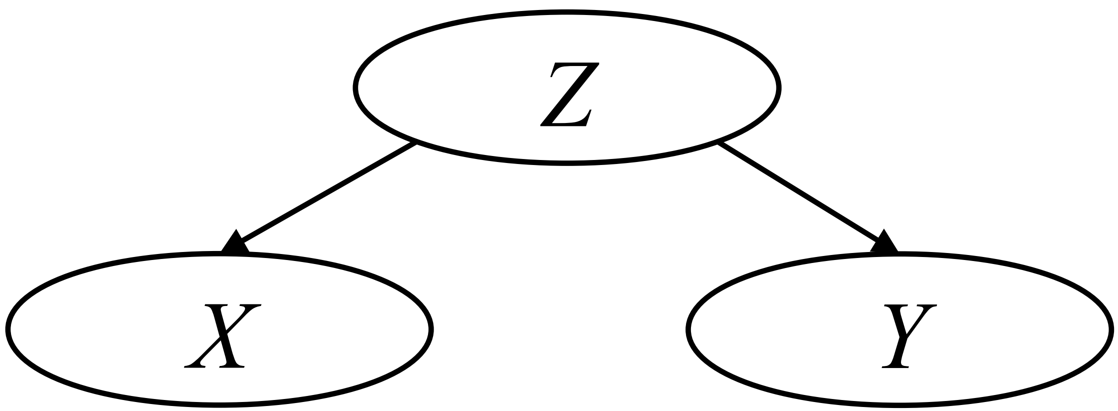
What should the contestant do?

The host is actually asking the contestant whether he/she wants to SWITCH the choice to both other doors, or STICK to the original choice. Phrased this way, it is obvious what the optimal thing to do is.

Conditional independence

We say that X and Y are conditionally independent if

$$\mathbf{P}(X, Y \mid Z) = \mathbf{P}(X \mid Z)\mathbf{P}(Y \mid Z)$$



What's the relation between independence and conditional independence?

Naive Bayes: Combining evidence

Assume full conditional independence and express the full joint probability distribution as:

$$\mathbf{P}(Effect_1, Effect_2, \ldots, Effect_n, Cause) =$$

$$\mathbf{P}(Effect_1, Effect_2, \ldots, Effect_n \mid Cause)\mathbf{P}(Cause) =$$

$$\mathbf{P}(Effect_1 \mid Cause) \cdots \mathbf{P}(Effect_n \mid Cause)\mathbf{P}(cause) =$$

$$\left[\prod_{i=1}^{n} \mathbf{P}(Effect_i \mid Cause)\right]\mathbf{P}(Cause)$$

Naive Bayes: Dentist example

$$\mathbf{P}(Toothache, Catch, Cavity) =$$

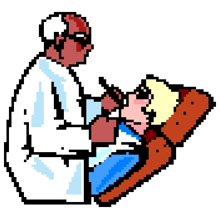
$$\mathbf{P}(Toothache, Catch \mid Cavity)\mathbf{P}(Cavity) =$$

$$\mathbf{P}(Toothache \mid Cavity)\mathbf{P}(Catch \mid Cavity)\mathbf{P}(Cavity)$$

$$\Rightarrow \quad \mathbf{P}(toothache, catch, cavity) =$$

$$\frac{(0.108 + 0.012)}{0.2} \times \frac{(0.108 + 0.072)}{0.2} \times 0.2 = 0.108$$

$$\text{True value}: \mathbf{P}(toothache, catch, cavity) = 0.108$$



	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Naive Bayes: Dentist example

	catch	¬catch
cavity	0.180	0.020
¬cavity	0.160	0.640

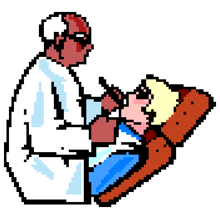
2 independent numbers

	toothache	¬toothache
cavity	0.120	0.080
¬cavity	0.080	0.720

2 independent numbers

cavity	0.200
¬cavity	0.800

1 independent number



	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Full table has 2^3-1=7 independent numbers [$O(2^n)$]

Naive Bayes application: Learning to classify text

- Use a dictionary with words (not too frequent and not to infrequent), e.g. w_1 = airplane, w_2 = algorithm, ...
- Estimate conditional probabilities $P(w_i$ | interesting) and $P(w_i$ | uninteresting)
- Compute P(text | interesting) and P(text | uninteresting) using Naive Bayes (and assuming that word position in text is unimportant)

$$P(\text{text} \mid \text{interesting}) \propto \prod_i P(w_i \mid \text{interesting})$$

Where w_i are the words occuring in this particular text.

Naive Bayes application: Learning to classify text

- Then compute the probability that the text is interesting (or uninteresting) using Bayes’ theorem

$$P(\text{interesting} \mid \text{text}) \propto \frac{P(\text{interesting})P(\text{text} \mid \text{interesting})}{P(\text{text})}$$

P(text) is just a normalization factor; it is not necessary to compute it since we are only interested in knowing whether

P(interesting | text) > P(uninteresting | text)

Conclusions

- Uncertainty in knowledge arises because of laziness and ignorance
  - this is unavoidable in the real world
- Probability expresses the degree to which the agent believes in possibility of different outcomes
- Decision theory helps the agent to act in a rational manner in the face of uncertainty
  - rational = maximize it's own expected utility
- In order to get efficient algorithms to deal with probability we need to explore the concept of conditional independence among variables