Reasoning under Uncertainty
(Chapter 13)

Dr. Nasima Begum
Associate Professor
Dept. of CSE, UAP

Outline

- Uncertainty- *In which we see how an agent can tame uncertainty with degrees of belief*

- Probability
- Syntax and Semantics
- Inference
- Independence and Bayes' Rule

The Problem: Uncertainty

- We cannot always know everything relevant to the problem before we select an action:
  - Environments that are non-deterministic, partially observable
  - Noisy sensors
  - Some features may be too complex model

- **For Example: Trying to decide when to leave for the airport to make a flight**
  - Will I get me there on time?
  - **Uncertainties:**
    - Car failures (flat tire, engine failure) (non-deterministic)
    - Road state, accidents, natural disasters (partially observable)
    - Unreliable weather reports, traffic updates (noisy sensors)
    - Predicting traffic along route (complex modeling)

- A purely logical agent does not allow for strong decision making in the face of such uncertainty.
  - Purely logical agents are based on binary True/False statements, no may be
  - Forces us to make assumptions to find a solution --> weak solutions

Uncertainty

- Back to planning:
  - Let action A_t = denote leave for airport ***t*** minutes before the flight.
  - For a given value of t, will A_t get me there on time?

Uncertainties:

1. partial observability (road state, other drivers' plans, etc.)
2. noisy sensors (traffic reports)
3. uncertainty in action outcomes (flat tire, etc.)
4. immense complexity of modeling and predicting traffic

Hence a purely logical approach either

1. risks falsehood: "A_{25} will get me there on time", or
2. leads to conclusions that are too weak for decision making:

"A_{25} will get me there on time if there's no accident on the bridge and it doesn't rain and my tires remain intact etc."

(A_{1440} might reasonably be said to get me there on time but I'd have to stay overnight in the airport …)

How to deal with Uncertainty

- Implicit methods:
  - Ignore uncertainty as much as possible
  - Build procedures that are robust to uncertainty
  - This is the approach in the planning methods studied so far (e.g. monitoring and re-planning)

- *Explicit methods*
  - Build a *model* of the world that *describes the uncertainty* (about the system's state, dynamics, sensors, model)
  - Reason about the effect of actions given the model

Methods for handling uncertainty

- *Default (non-monotonic) logic*: Make assumptions unless contradicted by evidence.

–E.g. "Assume my car doesnt have a flat tire. Assume A_{25} works unless contradicted by evidence

- Issues: What assumptions are reasonable? How to handle contradictions?
- *Rules with fudge factor:*

–E.g. "Sprinkler $\rightarrow_{0.99}$ WetGrass", "WetGrass $\rightarrow_{0.7}$ Rain"

- Issues: Problems with combination (e.g. Sprinkler causes rain?)
- *Probability*:
  - Model agent's degree of belief
  - Given the available evidence, A_{25} will get me there on time with probability 0.04

–E.g. Given what I know, *A(25)* succeed with probability 0.2

- *Fuzzy logic*:

–E.g. WetGrass is true to degree 0.2

- Issues: Handles degree of truth, NOT uncertainty.

Probability

- A well-known and well-understood framework for dealing with uncertainty
- Has a clear semantics
- Provides principled answers for:
  - Combining evidence
  - Predictive and diagnostic reasoning
  - Incorporation of new evidence
- Can be learned from data
- Intuitive to human experts (arguably?)

Probability

- We use probability to describe uncertainty due to:
  - –Laziness: failure to enumerate exceptions, qualifications etc.
  - –Ignorance: lack of relevant facts, initial conditions etc.
  - –True randomness? Quantum effects? ...
- *Beliefs (Bayesian or subjective probabilities)* relate propositions to one's current state of knowledge
  - –E.g. *P(A(25)| no reported accident) = 0.1*
- These are *not assertions about the world / absolute truth*
- Beliefs change with new evidence:
  - –E.g. *P(A(25)| no reported accident, 5am) = 0.2*
- This is analogous to logical entailment: *KB given* the *KB*, but may not be true in general.

Making actions/decisions under uncertainty

Suppose I believe the following:

$P(A_{25}$ gets me there on time | …) = 0.04
$P(A_{90}$ gets me there on time | …) = 0.70
$P(A_{120}$ gets me there on time | …) = 0.95
$P(A_{1440}$ gets me there on time | …) = 0.9999

- Which action to choose?
  Depends on my preferences for missing flight vs. time spent waiting, etc.
  - Utility theory is used to represent and infer preferences
  - Decision theory = probability theory + utility theory

Syntax

- A *random variable* X describes an outcome that cannot be determined in advance
  - E..g. The roll of a die
  - E.g. Number of e-mails received in a day
- The *sample space (domain)* S of a random variable X is the set of all possible values of the variable
  - E.g. For a die, $S = \{1, 2, 3, 4, 5, 6\}$
  - E.g. For number of emails received in a day, S is the natural numbers
- An *event* is a subset of S.
  - E.g. $e = \{1\}$ corresponds to a die roll of 1
  - E.g. number of e-mails in a day more than 100

Syntax

- Basic element: random variable
- Similar to propositional logic: possible worlds defined by assignment of values to random variables.
- Boolean random variables

  e.g., *Cavity* (do I have a cavity?)
- Discrete random variables: e.g., *Weather* is one of <*sunny, rainy, cloudy, snow*>
- Domain values must be exhaustive and mutually exclusive
- Elementary proposition constructed by assignment of a value to a random variable: e.g., *Weather = sunny*, *Cavity = false* (abbreviated as $\neg$*cavity*)
- Complex propositions formed from elementary propositions and standard logical connectives e.g., *Weather = sunny* $\vee$ *Cavity = false*

Syntax

- Atomic event: A complete specification of the state of the world about which the agent is uncertain

E.g., if the world (problem) consists of only two Boolean variables *Cavity* and *Toothache*, then there are 4 distinct atomic events:

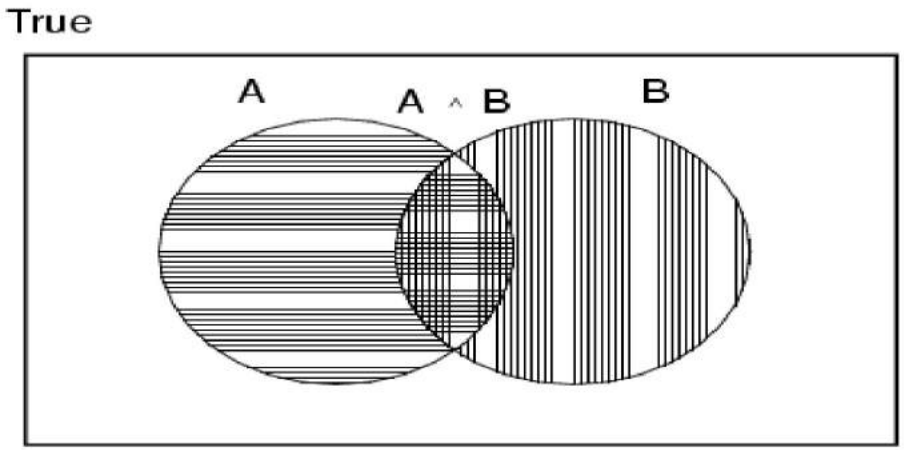
$Cavity = false \wedge Toothache = false$
$Cavity = false \wedge Toothache = true$
$Cavity = true \wedge Toothache = false$
$Cavity = true \wedge Toothache = true$

- Atomic events are mutually exclusive and exhaustive

Axioms of probability

- For any propositions *A, B*
  - $0 \leq P(A) \leq 1$
  - $P(true) = 1$ and $P(false) = 0$
  - $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$
- Alternatively, if A and B are mutually exclusive ($A \wedge B = F$) then:

  $P(A \vee B) = P(A) + P(B)$



Prior probability

- **Prior or unconditional probabilities** of propositions

  e.g., P(*Cavity* = true) = 0.1 and P(*Weather* = sunny) = 0.72 correspond to belief prior to arrival of any (new) evidence

- Probability distribution gives values for all possible assignments:
  P(*Weather*) = <0.72,0.1,0.08,0.1> (normalized, i.e., sums to 1)

- Joint probability distribution for a set of random variables gives the probability of every atomic event on those random variables

-

  P(*Weather, Cavity*) = a 4 × 2 matrix of values:

Weather =	sunny	rainy	cloudy	snow
Cavity = true	0.144	0.02	0.016	0.02
Cavity = false	0.576	0.08	0.064	0.08

- Every question about a domain can be answered by the joint distribution

Conditional probability

- Conditional or posterior probabilities
  e.g., P(*cavity* | *toothache*) = 0.8
  i.e., given that *toothache* is all I know

- (Notation for conditional distributions:
  P(*Cavity* | *Toothache*) = 2-element vector of 2-element vectors)

- If we know more, e.g., *cavity* is also given, then we have
  P(*cavity* / *toothache, cavity*) = 1

- New evidence may be irrelevant, allowing simplification, e.g.,
  P(*cavity* / *toothache, sunny*) = P(*cavity* | *toothache*) = 0.8

- This kind of inference, sanctioned by domain knowledge, is crucial

Conditional probability

- Definition of conditional probability:
  $P(a \mid b) = P(a \wedge b) / P(b)$ if $P(b) > 0$

- Product rule gives an alternative formulation:
  $P(a \wedge b) = P(a \mid b)\, P(b) = P(b \mid a)\, P(a)$

- A general version holds for whole distributions, e.g.,
  $\mathbf{P}(\mathit{Weather}, \mathit{Cavity}) = \mathbf{P}(\mathit{Weather} \mid \mathit{Cavity})\, \mathbf{P}(\mathit{Cavity})$
- (View as a set of 4 × 2 equations, not matrix mult.)

- Chain rule is derived by successive application of product rule:
-

$$
\begin{aligned}
\mathbf{P}(X_1, \ldots, X_n) &= \mathbf{P}(X_1,\ldots,X_{n-1})\, \mathbf{P}(X_n \mid X_1,\ldots,X_{n-1}) \\
&= \mathbf{P}(X_1,\ldots,X_{n-2})\, \mathbf{P}(X_{n-1} \mid X_1,\ldots,X_{n-2})\, \mathbf{P}(X_n \mid X_1,\ldots,X_{n-1}) \\
&= \ldots \\
&= \pi_{i=1}\hat{}n\, \mathbf{P}(X_i \mid X_1, \ldots, X_{i-1})
\end{aligned}
$$

Inference using Joint Distributions

- Suppose cavity, toothache and catch are three random variables.
- Start with the joint probability distribution:

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

- The ***unconditional probability*** of any proposition is computable as the sum of entries from the full joint distribution
- For any proposition ϕ, sum the atomic events where it is true: $P(\phi) = \Sigma_{\omega:\omega \models \phi} P(\omega)$

E.g. P(toothache) = P(cavity, catch) +P(cavity, ¬ catch)+ P(¬ cavity, catch) +P(¬ cavity, ¬ catch)

= 0.108 + 0.012 + 0.016 + 0.064

= 0.2

Inference using Joint Distributions*

- **Conditional Probability**

- The basic statements in the Bayesian framework talk about *conditional probabilities*.
  - **$P(A \mid B)$ is the belief in event A given that event B is known with certainty**
- The ***product rule*** gives an alternative formulation:
  - $P(A \wedge B) = P(A \mid B) P(B) = P(B \mid A) P(A)$
  - Note: we often write $P(A, B)$ as a shorthand for $P(A \wedge B)$

Inference using Joint Distributions

- Start with the joint probability distribution:

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	.108	.012	.072	.008
¬ *cavity*	.016	.064	.144	.576

- Can also compute **conditional probabilities**:
-

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

$$= \frac{0.016+0.064}{0.108 + 0.012 + 0.016 + 0.064}$$

$$= 0.4$$

Normalization

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

- Denominator (divisor) can be viewed as a normalization constant α

P(*Cavity | toothache*) = α, **P**(*Cavity, toothache*)

= α, [**P**(*Cavity, toothache, catch*) + **P**(*Cavity, toothache, ¬ catch*)]

= α, [<0.108,0.016> + <0.012,0.064>]

= α, <0.12,0.08>

= <0.6, 0.4> ; [α = 0.2]

General idea: compute distribution on query variable by fixing evidence variables and summing over hidden variables

Chain Rule

Chain rule is derived by successive application of product rule:

$$
\begin{aligned}
& P(X_1, \ldots, X_n) = \\
= \quad & P(X_1, \ldots, X_{n-1}) P(X_n | X_1, \ldots, X_{n-1}) \\
= \quad & P(X_1, \ldots, X_{n-2}) P(X_{n-1} | X_1, \ldots, X_{n-2}) P(X_n | X_1, \ldots, X_{n-1}) \\
= \quad & \ldots \\
= \quad & \prod_{i=1}^{n} P(X_i | X_1, \ldots, X_{i-1})
\end{aligned}
$$

Bayes Rule

- *Bayes rule* is another alternative formulation of the product rule:

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

- The *complete probability formula* states that:

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

or more generally,

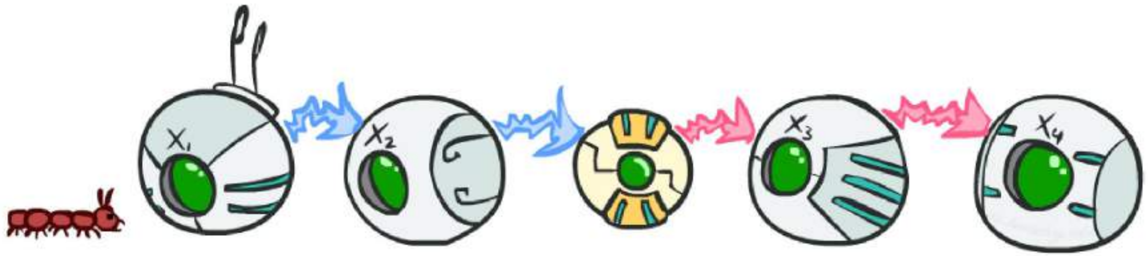
$$P(A) = \sum_i P(A|b_i)P(b_i),$$

where b_i form a set of exhaustive and mutually exclusive events.

Math on Naïve Bayes

Math on Naïve Bayes

Markov Models



[These slides were created by Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley.

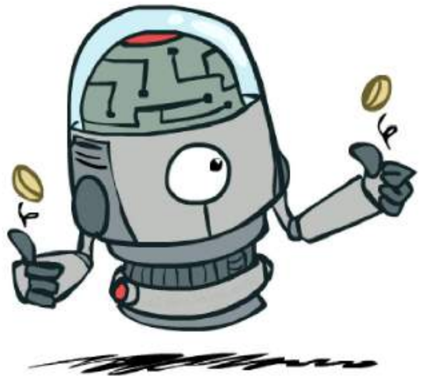
Independence

- Two variables are *independent* in a joint distribution if:

$$P(X, Y) = P(X)P(Y)$$

$$X \perp\!\!\!\perp Y$$

$$\forall x, y\, P(x, y) = P(x)P(y)$$



  - Says the joint distribution *factors* into a product of two simple ones
  - Usually variables aren't independent!

- Can use independence as a *modeling assumption*
  - Independence can be a simplifying assumption
  - *Empirical* joint distributions: at best "close" to independent
  - What could we assume for {Weather, Traffic, Cavity}?

- Independence is like something from CSPs (Constraint Satisfaction Problem)

Example: Independence?

$P(T)$

T	P
hot	0.5
cold	0.5

$P_1(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$P_2(T, W) = P(T)P(W)$

T	W	P
hot	sun	0.3
hot	rain	0.2
cold	sun	0.3
cold	rain	0.2

$P(W)$

W	P
sun	0.6
rain	0.4

Example: Independence

- N fair, independent coin flips:

$P(X_1)$

H	0.5
T	0.5

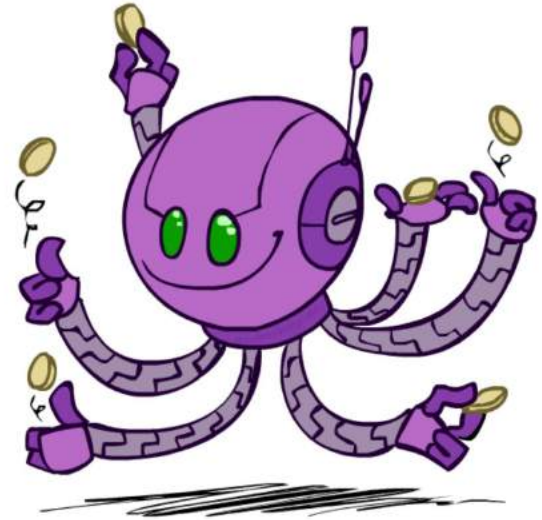
$P(X_2)$

H	0.5
T	0.5

$\cdots$

$P(X_n)$

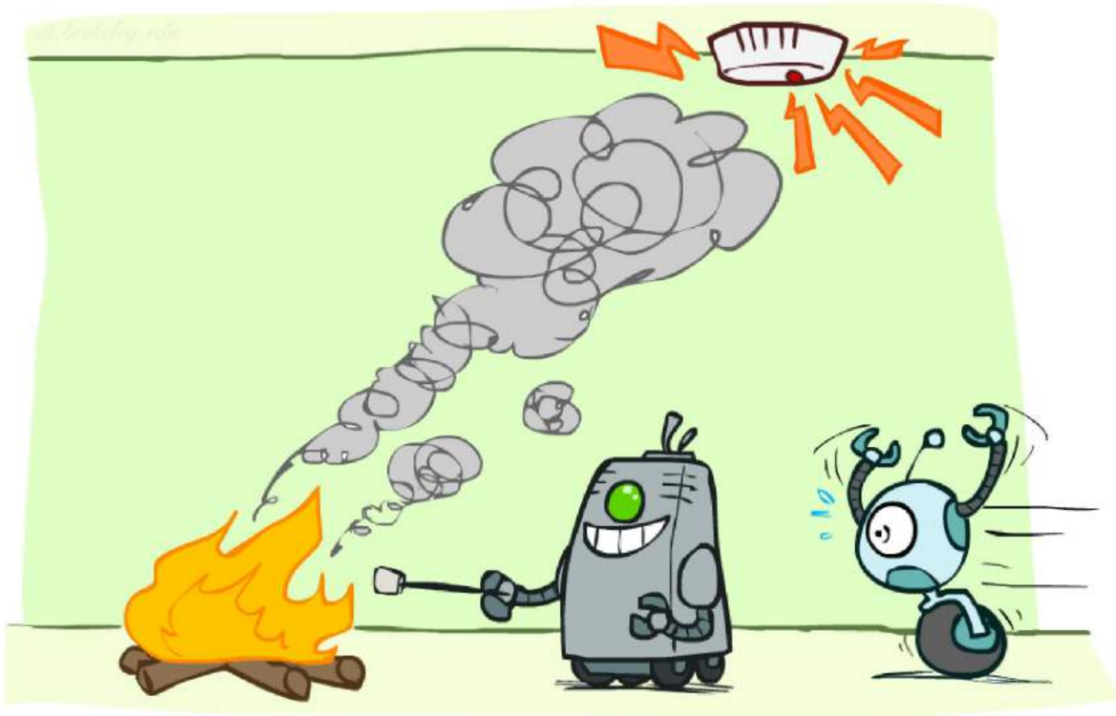
H	0.5
T	0.5



$P(X_1, X_2, \ldots X_n)$

2^n

Conditional Independence



Conditional Independence

In probability, we say two events are independent if knowing one event occurred doesn't change the probability of the other event.

- ❑ For example, the probability that a **fair coin** shows "heads" after being flipped is 1/2.

- What if we knew the day was Friday? Does this change the probability of getting "heads?"

- The probability of getting "heads," given that it's a Friday, is still 1/2.

- So the result of a coin flip and the day being Friday are independent events;

- Knowing it was a Friday didn't change the probability of getting "heads."

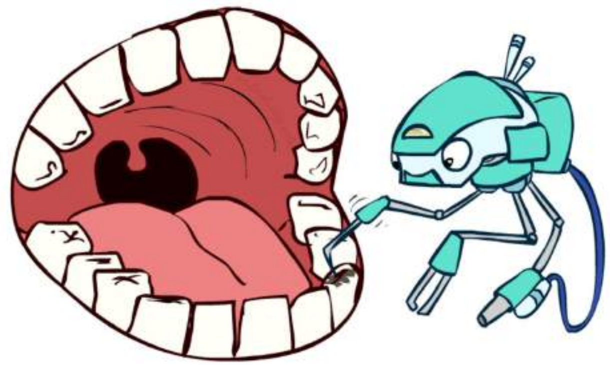
Conditional Independence

- P(Toothache, Cavity, Catch)

- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
  - P(+catch | +toothache, +cavity) = P(+catch | +cavity)

- The same independence holds if I don't have a cavity:
  - P(+catch | +toothache, -cavity) = P(+catch| -cavity)

- Catch is *conditionally independent* of Toothache given Cavity:
  - P(Catch | Toothache, Cavity) = P(Catch | Cavity)



- Equivalent statements:
  - P(Toothache | Catch , Cavity) = P(Toothache | Cavity)
  - P(Toothache, Catch | Cavity) = P(Toothache | Cavity) P(Catch | Cavity)
  - One can be derived from the other easily

Conditional Independence

- *Conditional independence* is the most basic and robust form of knowledge about uncertain environments.

- X is conditionally independent of Y given Z

$$X \perp\!\!\!\perp Y | Z$$

  if and only if:

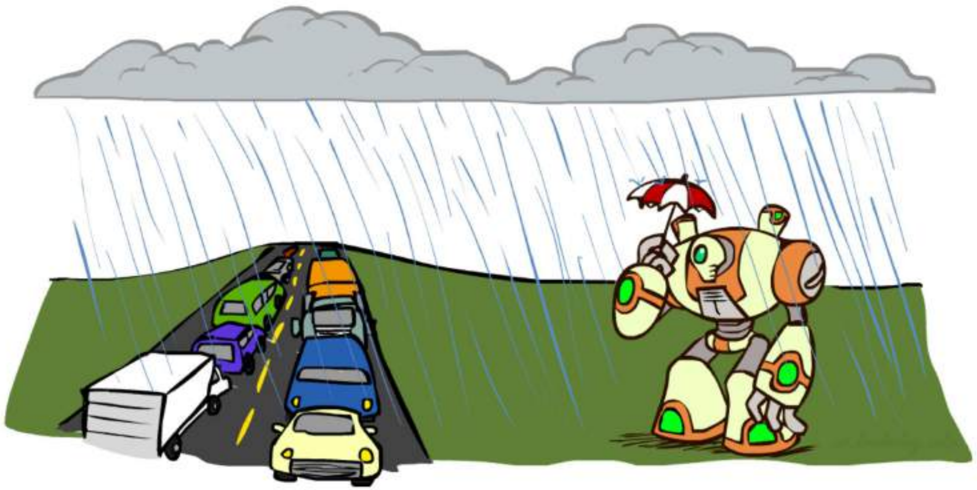
$$\forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$$

  or, equivalently, if and only if

$$\forall x, y, z : P(x|z, y) = P(x|z)$$

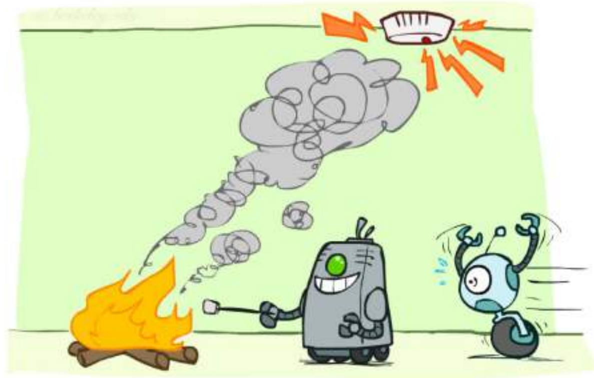
Conditional Independence

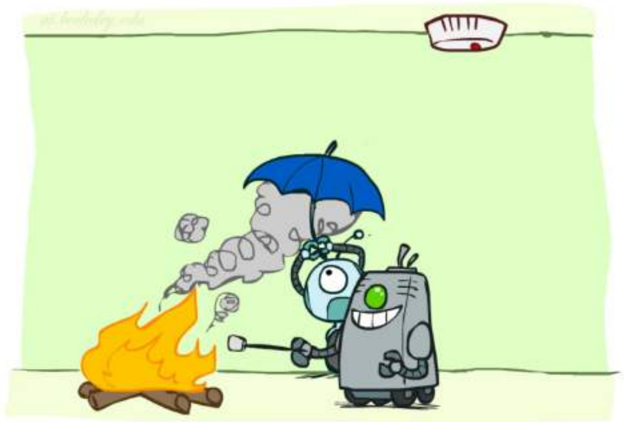
- What about this domain:
  - Traffic
  - Umbrella
  - Raining



Conditional Independence

- What about this domain:
  - Fire
  - Smoke
  - Alarm





Probability Recap

- Conditional probability

$$P(x|y) = \frac{P(x,y)}{P(y)}$$

- Product rule

$$P(x,y) = P(x|y)P(y)$$

- Chain rule

$$\begin{aligned} P(X_1, X_2, \ldots X_n) &= P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\ldots \\ &= \prod_{i=1}^{n} P(X_i|X_1, \ldots, X_{i-1}) \end{aligned}$$

- X, Y independent if and only if:

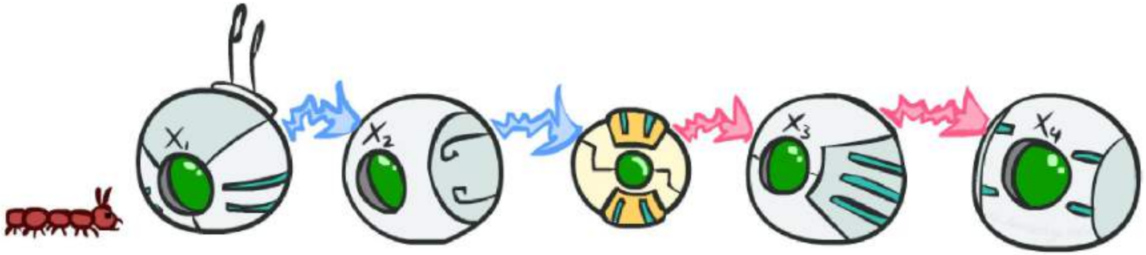
$$\forall x, y : P(x,y) = P(x)P(y)$$

- X and Y are conditionally independent given Z if and only if:

$$X \perp\!\!\!\perp Y | Z$$

$$\forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$$

Markov Models



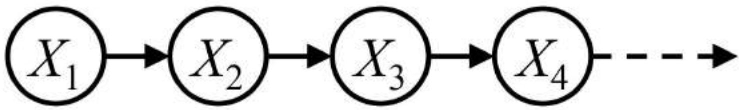
[These slides were created by Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley.

Reasoning over Time or Space

- Often, we want to reason about a sequence of observations
  - Speech recognition
  - Robot localization
  - User attention
  - Medical monitoring
- Need to introduce time (or space) into our models

Markov Models

- Value of X at a given time is called the state



$$P(X_1) \qquad P(X_t|X_{t-1})$$

- Parameters: called transition probabilities or dynamics, specify how the state evolves over time (also, initial state probabilities)
- Stationary assumption: transition probabilities the same at all times
- Same as MDP transition model, but no choice of action
- A stochastic **model** which is used to **model** a randomly changing systems. It is assumed that the future states depend only on the current state, not on the events that occurred before it.

Joint Distribution of a Markov Model

$$X_1 \to X_2 \to X_3 \to X_4$$

$$P(X_1) \qquad P(X_t|X_{t-1})$$

- Joint distribution:

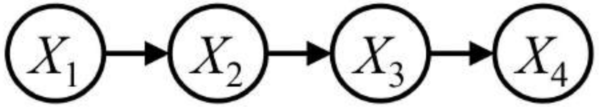
$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2|X_1)P(X_3|X_2)P(X_4|X_3)$$

- More generally:

$$\begin{aligned} P(X_1, X_2, \ldots, X_T) &= P(X_1)P(X_2|X_1)P(X_3|X_2)\ldots P(X_T|X_{T-1}) \\ &= P(X_1)\prod_{t=2}^{T} P(X_t|X_{t-1}) \end{aligned}$$

- Questions to be resolved:
  - Does this indeed define a joint distribution?
  - Can every joint distribution be factored this way, or are we making some assumptions about the joint distribution by using this factorization?

Chain Rule and Markov Models



- From the chain rule, every joint distribution over X_1, X_2, X_3, X_4 can be written as:

$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)P(X_4|X_1, X_2, X_3)$$

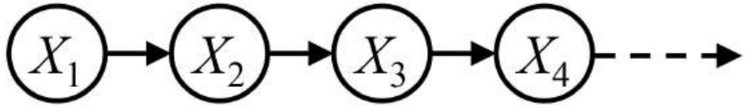
- Assuming that

$$X_3 \perp\!\!\!\perp X_1 \mid X_2 \quad \text{and} \quad X_4 \perp\!\!\!\perp X_1, X_2 \mid X_3$$

results in the expression posited on the previous slide:

$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2|X_1)P(X_3|X_2)P(X_4|X_3)$$

Chain Rule and Markov Models



- From the chain rule, every joint distribution over $X_1, X_2, \ldots, X_T$ can be written as:

$$P(X_1, X_2, \ldots, X_T) = P(X_1) \prod_{t=2}^{T} P(X_t | X_1, X_2, \ldots, X_{t-1})$$

- Assuming that for all *t*:

$$X_t \perp\!\!\!\perp X_1, \ldots, X_{t-2} \mid X_{t-1}$$

gives us the expression posited on the earlier slide:

$$P(X_1, X_2, \ldots, X_T) = P(X_1) \prod_{t=2}^{T} P(X_t | X_{t-1})$$

Implied Conditional Independencies

$$X_1 \to X_2 \to X_3 \to X_4$$

- We assumed: $X_3 \perp\!\!\!\perp X_1 \mid X_2$ and $X_4 \perp\!\!\!\perp X_1, X_2 \mid X_3$

- Do we also have $X_1 \perp\!\!\!\perp X_3, X_4 \mid X_2$?
  - Yes!
  - Proof:

$$\begin{aligned} P(X_1 \mid X_2, X_3, X_4) &= \frac{P(X_1, X_2, X_3, X_4)}{P(X_2, X_3, X_4)} \\ &= \frac{P(X_1)P(X_2 \mid X_1)P(X_3 \mid X_2)P(X_4 \mid X_3)}{\sum_{x_1} P(x_1)P(X_2 \mid x_1)P(X_3 \mid X_2)P(X_4 \mid X_3)} \\ &= \frac{P(X_1, X_2)}{P(X_2)} \\ &= P(X_1 \mid X_2) \end{aligned}$$

Markov Models Recap

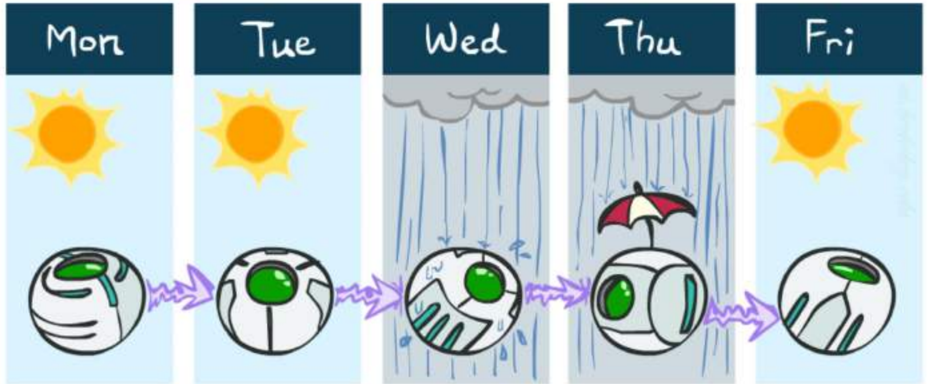
- Explicit assumption for all *t* : $X_t \perp\!\!\!\perp X_1, \ldots, X_{t-2} \mid X_{t-1}$
- Consequence, joint distribution can be written as:

$$P(X_1, X_2, \ldots, X_T) = P(X_1)P(X_2|X_1)P(X_3|X_2)\ldots P(X_T|X_{T-1})$$
$$= P(X_1)\prod_{t=2}^{T} P(X_t|X_{t-1})$$

- Implied conditional independencies:
  - Past variables independent of future variables given the present

  i.e., if $t_1 < t_2 < t_3$ or $t_1 > t_2 > t_3$ then: $X_{t_1} \perp\!\!\!\perp X_{t_3} \mid X_{t_2}$
- Additional explicit assumption: $P(X_t \mid X_{t-1})$ is the same for all *t*

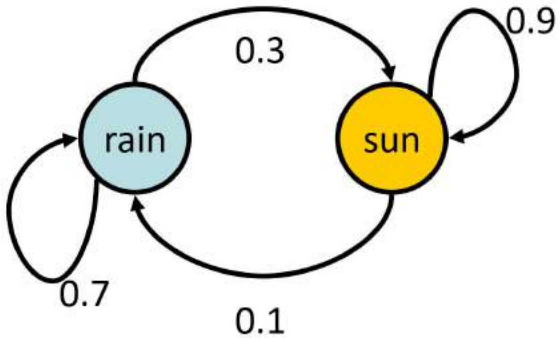
Example Markov Chain: Weather

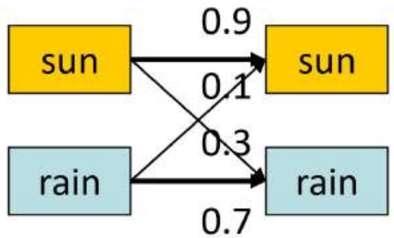


- States: X = {rain, sun}
- Initial distribution: 1.0 sun
- CPT $P(X_t \mid X_{t-1})$:

X_{t-1}	X_t	$P(X_t \mid X_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7

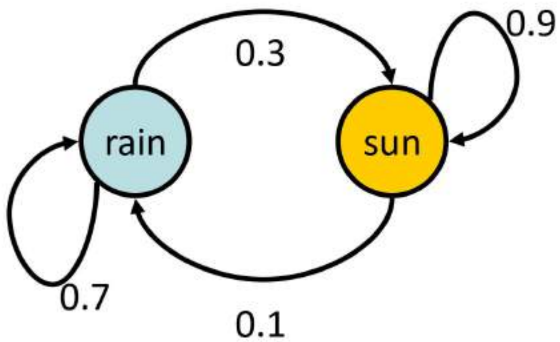
Two new ways of representing the same CPT





Example Markov Chain: Weather

- Initial distribution: 1.0 sun



- What is the probability distribution after one step?

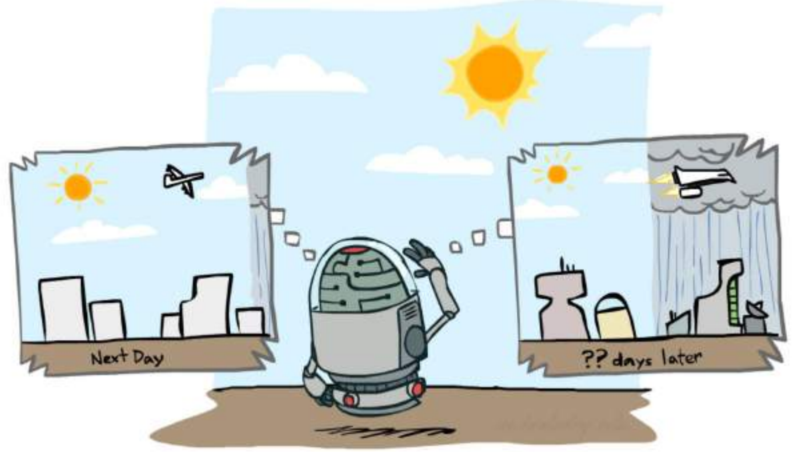
$$P(X_2 = \text{sun}) = \quad P(X_2 = \text{sun}|X_1 = \text{sun})P(X_1 = \text{sun}) + P(X_2 = \text{sun}|X_1 = \text{rain})P(X_1 = \text{rain})$$

$$0.9 \cdot 1.0 + 0.3 \cdot 0.0 = 0.9$$

Mini-Forward Algorithm

- Question: What's P(X) on some day t?

$$X_1 \to X_2 \to X_3 \to X_4 \dashrightarrow$$



$$P(x_1) = \text{known}$$

$$P(x_t) = \sum_{x_{t-1}} P(x_{t-1}, x_t)$$

$$= \sum_{x_{t-1}} P(x_t \mid x_{t-1}) P(x_{t-1})$$

Forward simulation

Example Run of Mini-Forward Algorithm

- From initial observation of sun

$$\left\langle \begin{matrix} 1.0 \\ 0.0 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.9 \\ 0.1 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.84 \\ 0.16 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.804 \\ 0.196 \end{matrix} \right\rangle \Rightarrow \left\langle \begin{matrix} 0.75 \\ 0.25 \end{matrix} \right\rangle$$

$P(X_1)$ $P(X_2)$ $P(X_3)$ $P(X_4)$ $P(X_\infty)$

- From initial observation of rain

$$\left\langle \begin{matrix} 0.0 \\ 1.0 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.3 \\ 0.7 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.48 \\ 0.52 \end{matrix} \right\rangle \quad \left\langle \begin{matrix} 0.588 \\ 0.412 \end{matrix} \right\rangle \Rightarrow \left\langle \begin{matrix} 0.75 \\ 0.25 \end{matrix} \right\rangle$$

$P(X_1)$ $P(X_2)$ $P(X_3)$ $P(X_4)$ $P(X_\infty)$

- From yet another initial distribution $P(X_1)$:

$$\left\langle \begin{matrix} p \\ 1-p \end{matrix} \right\rangle \quad \cdots \quad \Rightarrow \left\langle \begin{matrix} 0.75 \\ 0.25 \end{matrix} \right\rangle$$

$P(X_1)$ $P(X_\infty)$

Stationary Distributions

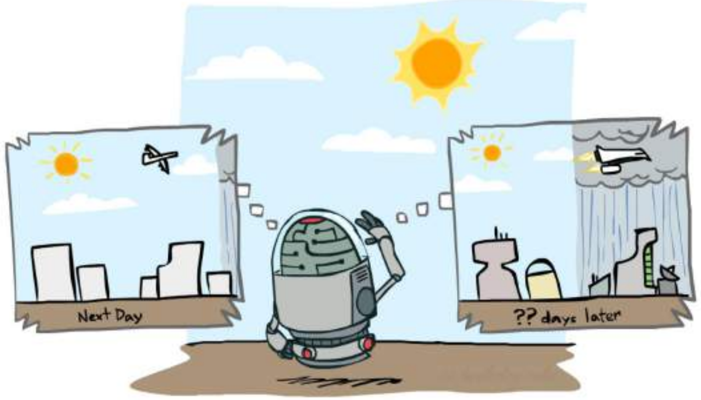
- For most chains:
  - Influence of the initial distribution gets less and less over time.
  - The distribution we end up in is independent of the initial distribution
- Stationary distribution:
  - The distribution we end up with is called the stationary distribution P_∞ of the chain
  - It satisfies

$$P_\infty(X) = P_{\infty+1}(X) = \sum_x P(X|x)P_\infty(x)$$

Example: Stationary Distributions

- Question: What's P(X) at time t = infinity?

$$X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow X_4 \dashrightarrow$$



$$P_\infty(sun) = P(sun|sun)P_\infty(sun) + P(sun|rain)P_\infty(rain)$$
$$P_\infty(rain) = P(rain|sun)P_\infty(sun) + P(rain|rain)P_\infty(rain)$$

$$P_\infty(sun) = 0.9P_\infty(sun) + 0.3P_\infty(rain)$$
$$P_\infty(rain) = 0.1P_\infty(sun) + 0.7P_\infty(rain)$$

$$P_\infty(sun) = 3P_\infty(rain)$$
$$P_\infty(rain) = 1/3P_\infty(sun)$$

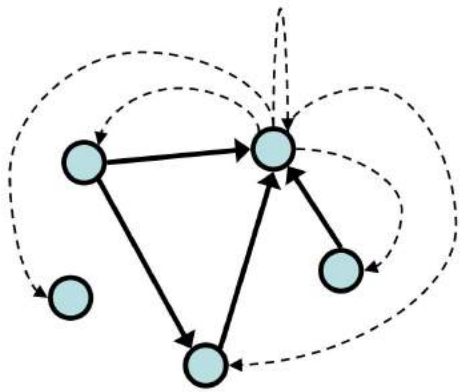
X_{t-1}	X_t	$P(X_t \mid X_{t-1})$
sun	sun	0.9
sun	rain	0.1
rain	sun	0.3
rain	rain	0.7

Also: $P_\infty(sun) + P_\infty(rain) = 1$

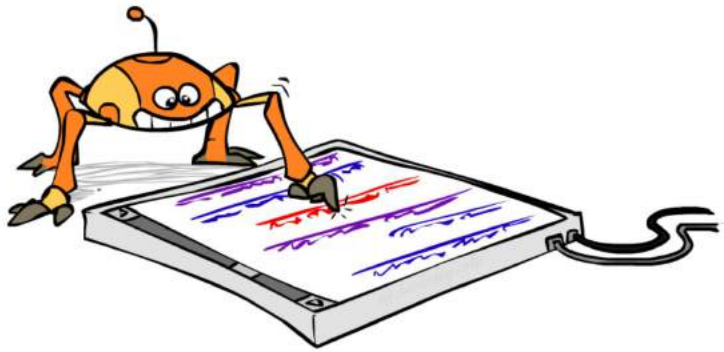
$$\Rightarrow \quad P_\infty(sun) = 3/4 \qquad P_\infty(rain) = 1/4$$

Application of Stationary Distribution: Web Link Analysis

- PageRank over a web graph
  - Each web page is a state
  - Initial distribution: uniform over pages
  - Transitions:
    - With prob. c, uniform jump to a random page (dotted lines, not all shown)
    - With prob. 1-c, follow a random outlink (solid lines)



- Stationary distribution
  - Will spend more time on highly reachable pages
  - E.g. many ways to get to the Acrobat Reader download page
  - Somewhat robust to link spam
  - Google 1.0 returned the set of pages containing all your keywords in decreasing rank, now all search engines use link analysis along with many other factors (rank actually getting less important over time)



Acknowledgement

- AIMA = Artificial Intelligence: A Modern Approach by Stuart Russell and Peter Norving (3rd edition)
- UC Berkeley (Some slides were created by Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley)
- U of toronto
- Other online resources

Thank You