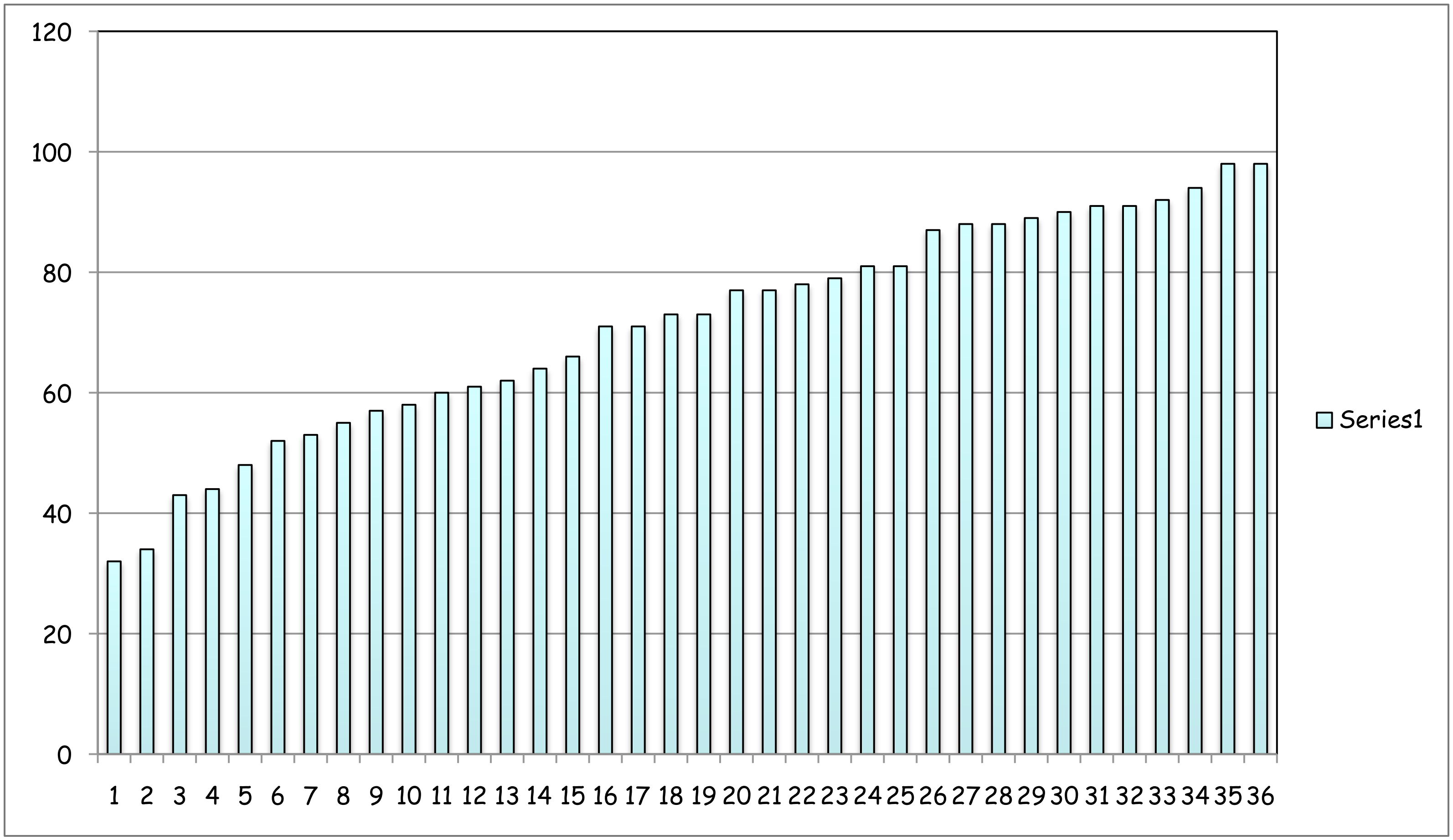
Dealing with Uncertainty

CMPSCI 383
October 25, 2011

Midterm grades



Average: 71

Outline

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

Uncertainty

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

Will A_t get me there on time?

- Problems
  - Partial observability (road state, other drivers' plans, etc.)
  - Noisy sensors (traffic reports)
  - Uncertainty in action outcomes (flat tire, etc.)
  - Immense complexity of modeling and predicting traffic
- A purely logical approach either...
  - Risks falsehood: "A_{25} will get me there on time", or
  - 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 etc." (qualification problem)
    - "A_{1440} might reasonably be said to get me there on time but I'd have to stay overnight in the airport …"

Options for handling uncertainty

- Belief states and contingency plans
- Default or nonmonotonic logic
  - Assume my car does not have a flat tire
  - Assume A_{25} works unless contradicted by evidence
  - However...
    - What assumptions are reasonable?
    - How to handle contradiction?
- Rules with fudge factors
  - $A_{25} \mapsto_{0.3}$ get there on time
  - $Sprinkler \mapsto_{0.99} WetGrass$
  - $WetGrass \mapsto_{0.7} Rain$
  - However...
    - 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

Probability

Probabilistic assertions summarize effects of

- Laziness — failure to enumerate exceptions, qualifications, etc.
- Ignorance —lack of relevant facts, initial conditions, etc.
- Fundamental stochastic nature of phenomena

Subjective probability

- Probabilities relate propositions to agent's own state of knowledge

  e.g., $P(A_{25} \mid \text{no reported accidents}) = 0.06$
- These are not assertions about the world, they are assertions about belief
- Probabilities of propositions change with new evidence:

  e.g., $P(A_{25} \mid \text{no reported accidents, 5 a.m.}) = 0.15$

Making 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
  - MEU principle: maximum expected utility

Basics of Probability Theory

Begin with a set Ω—the sample space

e.g., 6 possible rolls of a die.

$\omega \in \Omega$ is a sample point/possible world/atomic event

A probability space or probability model is a sample space with an assignment $P(\omega)$ for every $\omega \in \Omega$ s.t.

$0 \leq P(\omega) \leq 1$

$\Sigma_\omega P(\omega) = 1$

e.g., $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$.

An event A is any subset of Ω

$$P(A) = \Sigma_{\{\omega \in A\}} P(\omega)$$

E.g., $P(\text{die roll} < 4) = P(1) + P(2) + P(3) = 1/6 + 1/6 + 1/6 = 1/2$

Sample point/Atomic Event/Possible world

- A *complete* specification of the state of the world about which the agent is uncertain

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

  $Cavity = false \wedge Toothache = false$

  $Cavity = false \wedge Toothache = true$

  $Cavity = true \wedge Toothache = false$

  $Cavity = true \wedge Toothache = true$

- Sample points are mutually exclusive and collectively exhaustive

Random Variables

A random variable is a function from sample points to some range, e.g., the reals or Booleans

e.g., $Odd(1) = true$.

P induces a probability distribution for any r.v. X:

$$P(X = x_i) = \Sigma_{\{\omega : X(\omega) = x_i\}} P(\omega)$$

e.g., $P(Odd = true) = P(1) + P(3) + P(5) = 1/6 + 1/6 + 1/6 = 1/2$

Some Syntax

Propositional or Boolean random variables

e.g., $Cavity$ (do I have a cavity?)

$Cavity = true$ is a proposition, also written $cavity$

Discrete random variables (finite or infinite)

e.g., $Weather$ is one of $\langle sunny, rain, cloudy, snow \rangle$

$Weather = rain$ is a proposition

Values must be exhaustive and mutually exclusive

Continuous random variables (bounded or unbounded)

e.g., $Temp = 21.6$; also allow, e.g., $Temp < 22.0$.

Arbitrary Boolean combinations of basic propositions

Propositions and Probabilities

Think of a proposition as the event (set of sample points) where the proposition is true

Given Boolean random variables A and B:

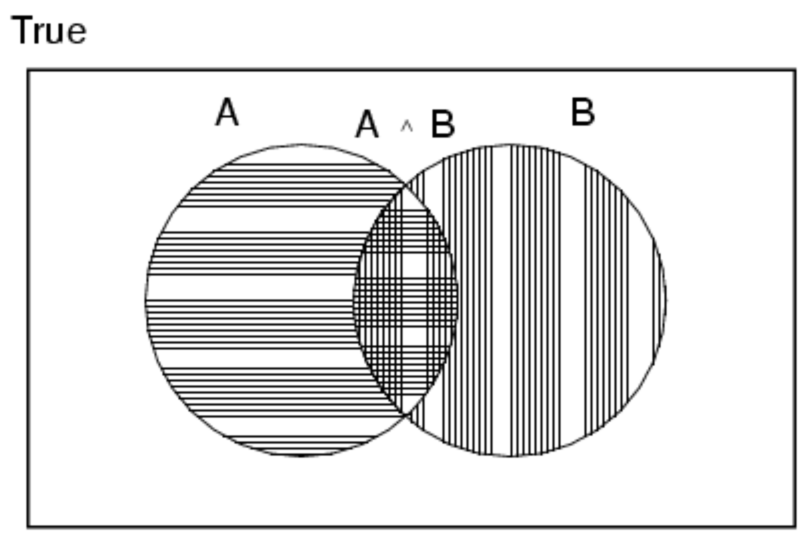
event a = set of sample points where $A(\omega) = true$
event $\neg a$ = set of sample points where $A(\omega) = false$
event $a \wedge b$ = points where $A(\omega) = true$ and $B(\omega) = true$

Often in AI applications, the sample points are defined by the values of a set of random variables, i.e., the sample space is the Cartesian product of the ranges of the variables

With Boolean variables, sample point = propositional logic model
e.g., $A = true$, $B = false$, or $a \wedge \neg b$.
Proposition = disjunction of atomic events in which it is true
e.g., $(a \vee b) \equiv (\neg a \wedge b) \vee (a \wedge \neg b) \vee (a \wedge b)$
$\Rightarrow P(a \vee b) = P(\neg a \wedge b) + P(a \wedge \neg b) + P(a \wedge b)$

Axioms of probability

- For any propositions *A, B*
  - $0 \leq P(A) \leq 1$
  - $P(true) = 1$ and $P(false) = 0$
  - $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$



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 probabilistic question about a domain can be answered by its joint distribution*

Conditional probability

- Conditional or posterior probabilities

  e.g., P(*cavity* | *toothache*) = 0.8

  i.e., given that *toothache* is all I know
  NOT "if *toothache* then 80% chance of *cavity*"

- Notation for complete conditional distribution:

  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

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:

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

Probabilistic Inference

- Unconditional or prior probability: belief in absence of additional information
- Evidence: additional information
- Conditional or posterior probability: belief given the evidence in addition to the prior

- Probabilistic Inference: computing posterior probabilities for query propositions given prior probabilities and evidence.

Inference by enumeration

- Start with the joint probability distribution:

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

- For any proposition φ, sum the atomic events where it is true

Inference by enumeration

- Start with the joint probability distribution:

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

- For any proposition φ, sum the atomic events where it is true
- P(*toothache*) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2

Inference by enumeration

- 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 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>

General idea: compute distribution on query variable by fixing observed variables and summing over unobserved variables

Inference by enumeration, contd.

Typically, we are interested in
the posterior joint distribution of the query variables **Y**
given specific values **e** for the evidence (observed) variables **E**

Let the hidden (unobserved) variables be **H** = **X** – **Y** – **E**

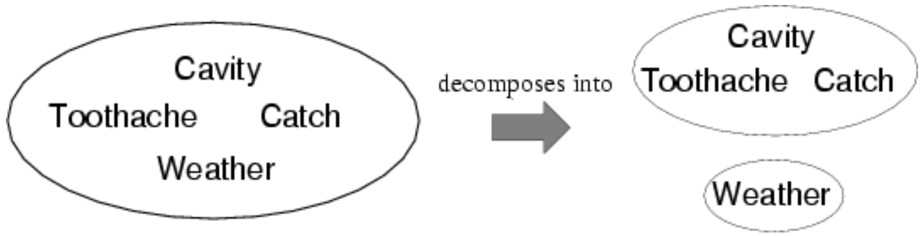
Then the required summation of joint entries is done by summing out the hidden variables:

$$\mathbf{P}(\mathbf{Y} \mid \mathbf{E} = \mathbf{e}) = \alpha \mathbf{P}(\mathbf{Y}, \mathbf{E} = \mathbf{e}) = \alpha \Sigma_{\mathbf{h}} \mathbf{P}(\mathbf{Y}, \mathbf{E} = \mathbf{e}, \mathbf{H} = \mathbf{h})$$

- The terms in the summation are joint entries because **Y**, **E** and **H** together exhaust the set of random variables
- Obvious problems:
  1. Worst-case time complexity $O(d^n)$ where d is the largest arity
  2. Space complexity $O(d^n)$ to store the joint distribution
  3. How to find the numbers for $O(d^n)$ entries?

Independence

- A and B are independent iff
  $\mathbf{P}(A|B) = \mathbf{P}(A)$ or $\mathbf{P}(B|A) = \mathbf{P}(B)$ or $\mathbf{P}(A, B) = \mathbf{P}(A)\,\mathbf{P}(B)$



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

- 32 entries reduced to 12; for n independent biased coins, $2^n \rightarrow n$
- Absolute independence powerful but rare
- Dentistry is a large field with hundreds of variables, none of which are independent. What to do?

Conditional independence

- $\mathbf{P}(Toothache, Cavity, Catch)$ has $2^3 - 1 = 7$ independent entries
- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:

  (1) $\mathbf{P}(catch \mid toothache, cavity) = \mathbf{P}(catch \mid cavity)$

- The same independence holds if I haven't got a cavity:

  (2) $\mathbf{P}(catch \mid toothache, \neg cavity) = \mathbf{P}(catch \mid \neg cavity)$

- *Catch* is conditionally independent of *Toothache* given *Cavity*:

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

- Equivalent statements:

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

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

Conditional independence contd.

- Write out full joint distribution using chain rule:

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

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

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

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

  i.e., 2 + 2 + 1 = 5 independent numbers

- In most cases, the use of conditional independence reduces the size of the representation of the joint distribution from exponential in *n* to linear in *n*.

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

Bayes' Rule

- Product rule $P(a \wedge b) = P(a \mid b) P(b) = P(b \mid a) P(a)$

  $\Rightarrow$ Bayes' rule: $P(a \mid b) = P(b \mid a) P(a) / P(b)$

- or in distribution form

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

- Useful for assessing diagnostic probability from causal probability:
  - $P(Cause|Effect) = P(Effect|Cause)\, P(Cause) / P(Effect)$
  - E.g., let *M* be meningitis, *S* be stiff neck:

    $P(m|s) = P(s|m)\, P(m) / P(s) = 0.8 \times 0.0001 / 0.1 = 0.0008$
  - Note: posterior probability of meningitis still very small!

Summary

- Probability is a rigorous formalism for uncertain knowledge
- Joint probability distribution specifies probability of every event
- Queries can be answered by summing over events
- For nontrivial domains, we must find a way to reduce the size of the joint distribution
- Independence and conditional independence provide the tools

Next Class

- Bayesian networks
- Sec 4.1 – 4.3