Representing Uncertainty

CSE 573

Many Techniques Developed

- Fuzzy Logic
- Certainty Factors
- Non-monotonic logic
- Probability

- Only one has stood the test of time!

Aspects of Uncertainty

- Suppose you have a flight at 12 noon
- When should you leave for SEATAC
  - What are traffic conditions?
  - How crowded is security?
- Leaving 18 hours early may get you there
  - But … ?

Decision Theory = Probability + Utility Theory

Min before noon	P(arrive-in-time)
20 min	0.05
30 min	0.25
45 min	0.50
60 min	0.75
120 min	0.98
1080 min	0.99

Depends on your ***preferences***

Utility theory: representing & reasoning about preferences

What Is Probability?

- **Probability:** Calculus for dealing with nondeterminism and uncertainty
- Cf. Logic
- **Probabilistic model:** Says how often we expect different things to occur
- Cf. Function

What Is Statistics?

- **Statistics 1:** Describing data
- **Statistics 2:** Inferring probabilistic models from data
  - Structure
  - Parameters

Why Should You Care?

- The world is full of uncertainty
  - Logic is not enough
  - Computers need to be able to handle uncertainty
- Probability: new foundation for AI (& CS!)
- Massive amounts of data around today
  - Statistics and CS are both about data
  - Statistics lets us summarize and understand it
  - Statistics is the basis for most learning
- Statistics lets data do our work for us

 7

Outline

- Basic notions
  - Atomic events, probabilities, joint distribution
  - Inference by enumeration
  - Independence & conditional independence
  - Bayes' rule
- Bayesian networks
- Statistical learning
- Dynamic Bayesian networks (DBNs)
- Markov decision processes (MDPs)

 8

Logic *vs.* Probability

Logic	Probability
Symbol: Q, R …	**Random variable: Q …**
Boolean values: T, F	**Domain: you specify e.g. {heads, tails} [1, 6]**
State of the world: Assignment to Q, R … Z	**Atomic event: complete specification of world: Q… Z** • **Mutually exclusive** • **Exhaustive**
	Prior probability (aka Unconditional prob: P(Q)
	Joint distribution: Prob. of every atomic event

 9

Syntax for Propositions

Propositional or Boolean random variables
  e.g., $Cavity$ (do I have a 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

 10

Propositions

- Assume Boolean variables
- Propositions:
  - A = true
  - B = false
  - $a \vee b$

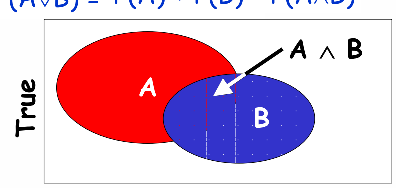
Proposition = disjunction of atomic events in which it is true
  e.g., $(a \vee b) \equiv (\neg a \wedge b) \vee (a \wedge \neg b) \vee (a \wedge b)$
  $\Rightarrow \ P(a \vee b) = P(\neg a \wedge b) + P(a \wedge \neg b) + P(a \wedge b)$

 11

Why Use Probability?

The definitions imply that certain logically related events must have related probabilities

E.g. $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$



de Finetti (1931): an agent who bets according to probabilities that violate these axioms can be forced to bet so as to lose money regardless of outcome.

 12

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:

$\mathbf{P}(Weather) = \langle 0.72, 0.1, 0.08, 0.1 \rangle$ (*normalized*, i.e., sums to 1)

Joint probability distribution for a set of r.v.s gives the probability of every atomic event on those r.v.s

$\mathbf{P}(Weather, Cavity) =$ a 4×2 matrix of values:

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

Any question 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

NOT "if $toothache$ then 80% chance of $cavity$"

(Notation for conditional distributions:

$\mathbf{P}(Cavity|Toothache) =$ 2-element vector of 2-element vectors)

If we know more, e.g., $cavity$ is also given, then we have

$P(cavity|toothache, cavity) = 1$

Note: the less specific belief *remains valid* after more evidence arrives, but is not always *useful*

New evidence may be irrelevant, allowing simplification, e.g.,

$P(cavity|toothache, 49ersWin) = P(cavity|toothache) = 0.8$

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

Conditional Probability

Def: $P(a|b) = \frac{P(a \wedge b)}{P(b)}$ if $P(b) \neq 0$

Product rule gives an alternative formulation:

$P(a \wedge b) = P(a|b)P(b) = p(b|a)P(a)$

A general version holds for whole distributions, e.g.,

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

(View as a 4×2 set of equations, *not* matrix mult.)

Chain rule is derived by successive application of product rule:

$$\begin{aligned}\mathbf{P}(X_1, \ldots, X_n) &= \mathbf{P}(X_1, \ldots, X_{n-1})\ \mathbf{P}(X_n|X_1, \ldots, X_{n-1}) \\ &= \mathbf{P}(X_1, \ldots, X_{n-2})\ \mathbf{P}(X_{n_1}|X_1, \ldots, X_{n-2})\ \mathbf{P}(X_n|X_1, \ldots, X_{n-1}) \\ &= \ldots \\ &= \Pi_{i=1}^{n} \mathbf{P}(X_i|X_1, \ldots, X_{i-1})\end{aligned}$$

Inference by Enumeration

Start with the joint distribution:

	$toothache$		$\neg\, toothache$	
	$catch$	$\neg\, catch$	$catch$	$\neg\, catch$
$cavity$	.108	.012	.072	.008
$\neg\, cavity$	.016	.064	.144	.576

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

$P(\phi) = \Sigma_{\omega:\omega \models \phi} P(\omega)$

P(toothache)=.108+.012+.016+.064

= .20 or 20%

This process is called "Marginalization"

Inference by Enumeration

Start with the joint distribution:

	$toothache$		$\neg\, toothache$	
	$catch$	$\neg\, catch$	$catch$	$\neg\, catch$
$cavity$	.108	.012	.072	.008
$\neg\, cavity$	.016	.064	.144	.576

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

$P(\phi) = \Sigma_{\omega:\omega \models \phi} P(\omega)$

P(toothache∨cavity = .20 + .072 + .008

.28

Inference by Enumeration

Start with the joint distribution:

	$toothache$		$\neg\, toothache$	
	$catch$	$\neg\, catch$	$catch$	$\neg\, catch$
$cavity$	.108	.012	.072	.008
$\neg\, cavity$	.016	.064	.144	.576

Can also compute conditional probabilities:

$$\begin{aligned}P(\neg cavity|toothache) &= \frac{P(\neg cavity \wedge toothache)}{P(toothache)} \\ &= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4\end{aligned}$$

Problems ??

- Worst case time: $O(n^d)$
  - Where d = max arity
  - And n = number of random variables
- Space complexity also $O(n^d)$
  - Size of joint distribution
- How get $O(n^d)$ entries for table??

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

Value of cavity & catch irrelevant - When computing P(toothache)