

## CSC384:Lecture9

- Lasttime
  - STRIPS and Regression planning
- Today
  - Reasoning under uncertainty (intro)
- Readings:
  - Today: 10.1– 10.3 (reasoning under uncertainty)
  - Next week: 10.3 (belief networks)

CSC 384 Lecture Slides (c) 2002, C. Boutilier

1

## Reasoning Under Uncertainty

- So far our planning problems have assumed:
  - start state is *known with certainty*
  - actions are *deterministic*
  - Both assumptions unrealistic (e.g., in robot domain)
- Knowledge:
  - What if Craig can be in office or lab?
  - Can robot know a priori if coffee made? mail waits?
- Actions:
  - Robbie grabs coffee: could fail (try again); could spill (make more)
  - Robbie does move(l,o): could end up elsewhere

CSC 384 Lecture Slides (c) 2002, C. Boutilier

2

## Classical Plans are Inadequate

- Classical plans (sequences of actions) fail to handle either of these
  - similarly for “deterministic domain” search procedures
- If actions have nondeterministic (uncertain) effects, suitable plans should *branch* on outcome
  - Pour coffee; **If** successful, go to office, **else** make more
- If uncertain knowledge, uncertain effects, there may be no plan guaranteed to achieve goal
  - Pour coffee; If spilled, there is no more and plan fails

CSC 384 Lecture Slides (c) 2002, C. Boutilier

3

## Some Decision Making Scenarios

- Suppose Craig wants coffee (as quickly as possible). Robbie, with coffee ready to deliver, knows Craig is in lab or office, but not which. Robbie is much closer to the lab than the office. What should it do?
  - What if Craig is *almost* always in his office?
  - What if peeking into the lab takes just a couple seconds? a couple minutes?
  - What if Craig is equally likely to be in lab or office?

CSC 384 Lecture Slides (c) 2002, C. Boutilier

4

## More Decision Making Scenarios

- Suppose Craig wants coffee or his mail (doesn't need both). Robbie only has enough battery power to do one or the other. What goals should it (attempt) to achieve?
  - What if Craig likes coffee much more than mail? mail more than coffee?
  - What if possibility of no coffee made?
  - What if Robbie usually spills coffee a couple of times before succeeding?
- Aside: consider the lottery. What would have to change for you to start (or stop) playing?

CSC 384 Lecture Slides (c) 2002, C. Boutilier

5

## Degrees of Belief and Preference

- The *right* decision/plan above requires that Robbie consider how important various objectives are, how likely he is to achieve them, and make tradeoffs among them.
- This generally requires that we *quantify* our preferences and our degrees of belief.
- We're going to start with *degrees of belief* and talk about preferences and decision making later.

CSC 384 Lecture Slides (c) 2002, C. Boutilier

6

## Reasoning under Uncertainty

- Logical representations (as seen so far) capture a weak form of uncertainty:
  - if query  $q$  not proved, could be true or false
- Too weak to allow (reasonable) tradeoffs
  - want to know *how likely*  $q$  is
- We'll *quantify* our beliefs using *probabilities*
  - $Pr(q)$  denotes probability that you believe  $q$  is true
  - We take subjectivist viewpoint (cf. *frequentist*)
- Note: statistics/data *influence* degrees of belief
- Let's formalize things...

CSC 384 Lecture Slides (c) 2002, C. Boutilier

7

## Random Variables

- Assume set  $V$  of *random variables*:  $X, Y$ , etc.
  - Each RV  $X$  has a *domain* of values  $Dom(X)$
  - $X$  can take on any value from  $Dom(X)$
  - Assume  $V$  and  $Dom(X)$  finite
- Examples
  - $Dom(X) = \{x_1, x_2, x_3\}$
  - $Dom(Weather) = \{sunny, cloudy, rainy\}$
  - $Dom(StudentInCraigsOffice) = \{pascal, georgios, veronica, tianhan, \dots\}$
  - $Dom(CraigHasCoffee) = \{T, F\}$  (boolean var)

CSC 384 Lecture Slides (c) 2002, C. Boutilier

8

## Random Variables/Possible Worlds

- A *formula* is a logical combination of variable assignments:
  - $X = x_1; (X = x_2 \vee X = x_3) \wedge Y = y_2; (x_2 \vee x_3) \wedge y_2$
  - $chc \wedge \sim cm$ , etc...
  - let  $\mathcal{L}$  denote the set of formulae (our language)
- A *possible world* is an assignment of values to each RV
  - these are analogous to truth assts (interpretations)
  - Let  $W$  be the set of worlds

CSC 384 Lecture Slides (c) 2002, C. Boutilier

9

## Probability Distributions

- A probability distribution  $Pr: \mathcal{L} \rightarrow [0, 1]$  s.t.
  - $0 \leq Pr(\alpha) \leq 1$
  - $Pr(\alpha) = Pr(\beta)$  if  $\alpha$  is logically equivalent to  $\beta$
  - $Pr(\alpha) = 1$  if  $\alpha$  is a tautology
  - $Pr(\alpha \vee \beta) = Pr(\alpha) + Pr(\beta) - Pr(\alpha \wedge \beta)$
- $Pr(\alpha)$  denotes our *degree of belief* in  $\alpha$ ; e.g.
  - $Pr(X = x_1) = Pr(x_1) = 0.9$
  - $Pr((x_2 \vee x_3) \wedge y_2) = 0.9$
  - $Pr(loc(craig) = off) = 0.6$
  - $Pr(loc(craig) = off \vee loc(craig) = lab) = 1.0$
  - $Pr(loc(craig) = lounge) = 0.0$

CSC 384 Lecture Slides (c) 2002, C. Boutilier

10

## Semantics of Prob. Distributions

- A probability measure  $\mu: W \rightarrow [0, 1]$  s.t.
  - $\sum_{w \in W} \mu(w) = 1$
- Intuitively,  $\mu(w)$  measures the probability that the actual world is  $w$  (your *belief* in  $w$ ). Thus, the relative likelihood of any world you consider possible is specified. If  $w$  has measure 0, you consider it to be impossible!

CSC 384 Lecture Slides (c) 2002, C. Boutilier

11

## Semantics of Distributions

- Given measure  $\mu$ , we determine degree of belief in formula  $Pr(\alpha)$ 
  - simply sum the measures of all worlds satisfying the formula of interest

$$Pr(\alpha) = \sum_{w \in W} \{\mu(w) : w \models \alpha\}$$

CSC 384 Lecture Slides (c) 2002, C. Boutilier

12

## Example Distribution

T - mail truck outside  
M - mail waiting  
C - craig wants coffee  
A - craig is angry

t c m a	0.162	$\bar{t}$ c m a	0.0
t c $\bar{m}$ a	0.018	$\bar{t}$ c $\bar{m}$ a	0.0
t c m $\bar{a}$	0.016	$\bar{t}$ c m $\bar{a}$	0.0
t c $\bar{m}$ $\bar{a}$	0.004	$\bar{t}$ c $\bar{m}$ $\bar{a}$	0.0
$\bar{t}$ c m a	0.432	$\bar{t}$ c m $\bar{a}$	0.0
$\bar{t}$ c $\bar{m}$ a	0.288	$\bar{t}$ c $\bar{m}$ $\bar{a}$	0.0
$\bar{t}$ c m $\bar{a}$	0.008	$\bar{t}$ c m $\bar{a}$	0.0
$\bar{t}$ c $\bar{m}$ $\bar{a}$	0.072	$\bar{t}$ c $\bar{m}$ $\bar{a}$	0.0

$\Pr(t) = 1$   
 $\Pr(\neg t) = 0$   
 $\Pr(c) = .2$   
 $\Pr(\neg c) = .8$   
 $\Pr(m) = .9$   
 $\Pr(a) = .618$   
 $\Pr(c \& m) = .18$   
 $\Pr(c \vee m) = .92$   
 $\Pr(a \rightarrow m)$   
 $= \Pr(\neg a \vee m)$   
 $= 1 - \Pr(a \& \neg m)$   
 $= .976$

CSC 384 Lecture Slides (c) 2002, C. Boutilier

13

## Relationship

- For any measure  $\mu$  the induced mapping  $\Pr$  is a distribution.
- For any distribution  $\Pr$  there is a corresponding  $\mu$  measure that induces  $\Pr$ .
- Thus, the syntactic and semantic restrictions correspond (soundness and completeness)

CSC 384 Lecture Slides (c) 2002, C. Boutilier

14

## Some Important Properties

- $\Pr(a) = 1 - \Pr(\neg a)$ , where  $a$  can be "generalized"
- $\sum \{\Pr(x) : x \in \text{Dom}(X)\} = 1$ 
  - e.g.,  $\Pr(\text{sunny}) + \Pr(\text{cloudy}) + \Pr(\text{rainy}) = 1$
- $\Pr(a \vee b) = 1$  if  $a \vdash \neg b$
- $\Pr(x) = \sum_{y \in \text{Dom}(Y)} \Pr(x \wedge y)$ 
  - this is called the *summing out property*
  - e.g.,  $\Pr(a) = \Pr(a \& m) + \Pr(a \& \neg m)$

CSC 384 Lecture Slides (c) 2002, C. Boutilier

15

## Conditional Probability

- Conditional probability critical in inference

$$\Pr(b | a) = \frac{\Pr(b \wedge a)}{\Pr(a)}$$

- if  $\Pr(a) = 0$ , we often treat  $\Pr(b|a)=1$  by convention

CSC 384 Lecture Slides (c) 2002, C. Boutilier

16

## Semantics of Conditional Prob.

- Semantics of  $\Pr(b|a)$ :
  - denotes relative weight of b-worlds among a-worlds
  - $\neg a$ -worlds play no role

$$\Pr(b | a) = \frac{\sum \{\mu(w) : w \models a \wedge b\}}{\sum \{\mu(w) : w \models a\}}$$

CSC 384 Lecture Slides (c) 2002, C. Boutilier

17

## Intuitive Meaning of Cond. Prob.

- Intuitively, if you learned  $a$ , you would change your degree of belief in  $b$  from  $\Pr(b)$  to  $\Pr(b|a)$
- In our example:
  - $\Pr(m|c) = 0.9$
  - $\Pr(m|\neg c) = 0.9$
  - $\Pr(a) = 0.618$
  - $\Pr(a|\neg m) = 0.27$
  - $\Pr(a|\neg m \& c) = 0.8$
- Notice the *nonmonotonicity* in the last three cases when additional evidence is added
  - contrast this with logical inference

CSC 384 Lecture Slides (c) 2002, C. Boutilier

18

## Some Important Properties

▪ **Product Rule:**  $\Pr(ab) = \Pr(a|b)\Pr(b)$

▪ **Summing Out Rule:**

$$\Pr(a) = \sum_{b \in \text{Dom}(B)} \Pr(a|b)\Pr(b)$$

▪ **Chain Rule:**

$$\Pr(abcd) = \Pr(a|bcd)\Pr(b|cd)\Pr(c|d)\Pr(d)$$

- holds for any number of variables

CSC 384 Lecture Slides (c) 2002, C. Boutilier

19

## Bayes Rule

▪ **Bayes Rule:**

$$\Pr(a|b) = \frac{\Pr(b|a)\Pr(a)}{\Pr(b)}$$

- Bayes rule follows by simple algebraic manipulation of the defn of condition probability
  - why is it so important? why significant?
  - usually, one "direction" easier to assess than other

CSC 384 Lecture Slides (c) 2002, C. Boutilier

20

## Example of Use of Bayes Rule

- Disease  $\in \{\text{malaria, cold, flu}\}$ ; Symptom = fever
  - Must compute  $\Pr(D|\text{fever})$  to prescribe treatment
- Why not assess this quantity directly?
  - $\Pr(\text{mal} | \text{fever})$  is not natural to assess;  $\Pr(\text{fever} | \text{mal})$  reflects the underlying "causal" mechanism
  - $\Pr(\text{mal} | \text{fever})$  is not "stable": a malaria epidemic changes this quantity (for example)
- So we use Bayes rule:
  - $\Pr(\text{mal} | \text{fever}) = \Pr(\text{fever} | \text{mal}) \Pr(\text{mal}) / \Pr(\text{fever})$
  - note that  $\Pr(\text{fever}) = \Pr(\text{m} | \text{fever}) + \Pr(\text{c} | \text{fever}) + \Pr(\text{fl} | \text{fever})$
  - so if we compute  $\Pr$  of each disease given fever using Bayes rule, normalizing constant is "free"

CSC 384 Lecture Slides (c) 2002, C. Boutilier

21

## Probabilistic Inference

- By probabilistic inference, we mean
  - given a *prior* distribution  $\Pr$  over variables of interest, representing degrees of belief
  - and given new evidence  $E=e$  for some var  $E$
  - Revise your degrees of belief: *posterior*  $\Pr_e$
- How do your degrees of belief change as a result of learning  $E=e$  (or more generally  $E=\mathbf{e}$ , for set  $\mathbf{E}$ )
- Contrast with "logical" reasoning over probabilistic assertions

CSC 384 Lecture Slides (c) 2002, C. Boutilier

22

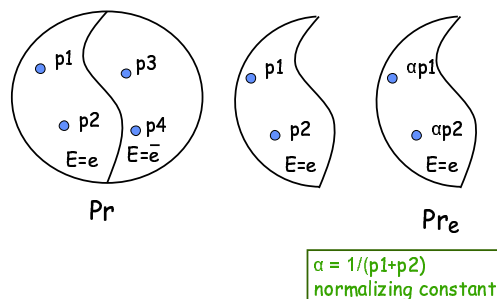
## Conditioning

- We define  $\Pr_e(a) = \Pr(a|e)$
- That is, we produce  $\Pr_e$  by *conditioning* the prior distribution on the observed evidence  $e$
- Semantically, we take original measure  $\mu$ 
  - we set  $\mu(w) = 0$  for any world falsifying  $e$
  - we set  $\mu(w) = \mu(w) / \Pr(e)$  for any  $e$ -world
  - last step known as normalization (ensures that the new measure sums to 1)

CSC 384 Lecture Slides (c) 2002, C. Boutilier

23

## Semantics of Conditioning



CSC 384 Lecture Slides (c) 2002, C. Boutilier

24

## Inference: Computational Bottleneck

- Semantically/conceptually, picture is clear; but several issues must be addressed
- Issue 1: How do we specify the full joint distribution over  $X_1, X_2, \dots, X_n$ ?
  - exponential number of possible worlds
  - e.g., if the  $X_i$  are boolean, then  $2^n$  numbers (or  $2^n - 1$  parameters/degrees of freedom, since they sum to 1)
  - these numbers are **not robust/stable**
  - these numbers are **not natural** to assess (what is probability that "Craig wants coffee; it's raining in Coquitlam; robot charge level is low; ..."?)

CSC 384 Lecture Slides (c) 2002, C. Boutilier

25

## Inference: Computational Bottleneck

- Issue 2: Inference in this rep'n frightfully slow
  - Must sum over exponential number of worlds to answer query  $Pr(a)$  or to condition on evidence  $e$  to determine  $Pr_e(a)$
- How do we avoid these two problems?
  - no solution in general
  - but in practice there is structure we can exploit
- We'll use conditional independence

CSC 384 Lecture Slides (c) 2002, C. Boutilier

26

## Independence

- Recall that  $x$  and  $y$  are **independent** iff:
  - $Pr(x) = Pr(x|y)$  iff  $Pr(y) = Pr(y|x)$  iff  $Pr(xy) = Pr(x)Pr(y)$
  - intuitively, learning  $y$  doesn't influence beliefs about  $x$
- $x$  and  $y$  are **conditionally independent given  $z$**  iff:
  - $Pr(x|z) = Pr(x|yz)$  iff  $Pr(y|z) = Pr(y|xz)$  iff  $Pr(xy|z) = Pr(x|z)Pr(y|z)$  iff ...
  - intuitively, learning  $y$  doesn't influence your beliefs about  $x$  **if you already know  $z$**
  - e.g., learning someone's mark on 384 exam can influence the probability you assign to a specific GPA; but if you already knew **final** 384 grade, learning the exam mark would **not** influence your GPA assessment

CSC 384 Lecture Slides (c) 2002, C. Boutilier

27

## Variable Independence

- Two **variables**  $X$  and  $Y$  are conditionally independent given variable  $Z$  iff  $x, y$  are conditionally independent given  $z$  for all  $x \in \text{Dom}(X), y \in \text{Dom}(Y), z \in \text{Dom}(Z)$ 
  - Also applies to sets of variables  $X, Y, Z$
  - Also to unconditional case ( $X, Y$  independent)
- If you know the value of  $Z$  (**whatever** it is), nothing you learn about  $Y$  will influence your beliefs about  $X$ 
  - these defs differ from earlier ones (which talk about events, not variables)

CSC 384 Lecture Slides (c) 2002, C. Boutilier

28

## What does independence buys us?

- Suppose (say, boolean) variables  $X_1, X_2, \dots, X_n$  are mutually independent
  - we can specify full joint distribution using only  $n$  parameters (linear) instead of  $2^n - 1$  (exponential)
- How? Simply specify  $Pr(x_1), \dots, Pr(x_n)$ 
  - from this I can recover probability of any world or any (conjunctive) query easily
  - e.g.  $Pr(x_1 \sim x_2 x_3 x_4) = Pr(x_1) (1 - Pr(x_2)) Pr(x_3) Pr(x_4)$
  - we can condition on observed value  $X_k = x_k$  trivially by changing  $Pr(x_k)$  to 1, leaving  $Pr(x_i)$  untouched for  $i \neq k$

CSC 384 Lecture Slides (c) 2002, C. Boutilier

29

## The Value of Independence

- Complete independence reduces both **representation of joint** and **inference** from  $O(2^n)$  to  $O(n)$ : pretty significant!
- Unfortunately, such complete mutual independence is very rare. Most realistic domains do not exhibit this property.
- Fortunately, most domains do exhibit a fair amount of conditional independence. And we can exploit conditional independence for representation and inference as well.
- **Bayesian networks** do just this

CSC 384 Lecture Slides (c) 2002, C. Boutilier

30

## Exploiting Conditional Independence

- Let's see what conditional independence buys us
- Consider a story:

- If Craig woke up too early E, Craig probably needs coffee C; if C, Craig needs coffee, he's likely angry A. If A, there is an increased chance of an aneurysm (burst blood vessel) B. If B, Craig is quite likely to be hospitalized H.



E - Craig woke too early    A - Craig is angry    H - Craig hospitalized  
C - Craig needs coffee    B - Craig burst a blood vessel

CSC 384 Lecture Slides (c) 2002, C. Boutilier

31

## An Aside on Notation

- $\Pr(X)$  for variable X (or set of variables) refers to the *(marginal) distribution* over X.  $\Pr(X|Y)$  refers to family of conditional distributions over X, one for each  $y \in \text{Dom}(Y)$ .
- Distinguish between  $\Pr(X)$  -- which is a distribution -- and  $\Pr(x)$  or  $\Pr(\sim x)$  (or  $\Pr(x_i)$  for nonboolean vars) -- which are numbers. Think of  $\Pr(X)$  as a function that accepts any  $x_i \in \text{Dom}(X)$  as an argument and returns  $\Pr(x_i)$ .
- Similarly, think of  $\Pr(X|Y)$  as a function that accepts any  $x_i$  and  $y_k$  and returns  $\Pr(x_i | y_k)$ . Note that  $\Pr(X|Y)$  is not a single distribution; rather it denotes the family of distributions (over X) induced by the different  $y_k \in \text{Dom}(Y)$

CSC 384 Lecture Slides (c) 2002, C. Boutilier

32

## Cond'l Independence in our Story



- If you learned any of E, C, A, or B, your assessment of  $\Pr(H)$  would change.
  - E.g., if any of these are seen to be true, you would increase  $\Pr(h)$  and decrease  $\Pr(\sim h)$ .
  - So H is *not independent* of E, or C, or A, or B.
- But if you knew value of B (true or false), learning value of E, C, or A, would not influence  $\Pr(H)$ . Influence these factors have on H is mediated by their influence on B.
  - Craig doesn't get sent to the hospital because he's angry, he gets sent because he's had an aneurysm.
  - So H is *independent* of E, and C, and A, *given* B

CSC 384 Lecture Slides (c) 2002, C. Boutilier

33

## Cond'l Independence in our Story



- So H is *independent* of E, and C, and A, *given* B
- Similarly:
  - B is *independent* of E, and C, *given* A
  - A is *independent* of E, *given* C
- This means that:
  - $\Pr(H | B, \{A, C, E\}) = \Pr(H | B)$ 
    - i.e., for any subset of  $\{A, C, E\}$ , this relation holds
  - $\Pr(B | A, \{C, E\}) = \Pr(B | A)$
  - $\Pr(A | C, \{E\}) = \Pr(A | C)$
  - $\Pr(C | E)$  and  $\Pr(E)$  don't "simplify"

CSC 384 Lecture Slides (c) 2002, C. Boutilier

34

## Cond'l Independence in our Story

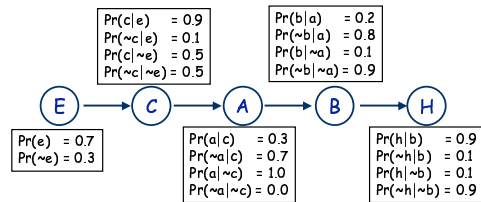


- By the chain rule (for any instantiation of H...E):
  - $\Pr(H, B, A, C, E) = \Pr(H | B, A, C, E) \Pr(B | A, C, E) \Pr(A | C, E) \Pr(C | E) \Pr(E)$
- By our independence assumptions:
  - $\Pr(H, B, A, C, E) = \Pr(H | B) \Pr(B | A) \Pr(A | C) \Pr(C | E) \Pr(E)$
- We can specify the full joint by specifying five *local conditional distributions*:  $\Pr(H | B)$ ;  $\Pr(B | A)$ ;  $\Pr(A | C)$ ;  $\Pr(C | E)$ ; and  $\Pr(E)$

CSC 384 Lecture Slides (c) 2002, C. Boutilier

35

## Example Quantification



- Specifying the joint requires only 9 parameters (if we note that half of these are "1 minus" the others), instead of 31 for explicit representation
  - linear in number of vars instead of exponential!
  - linear generally if dependence has a chain structure

CSC 384 Lecture Slides (c) 2002, C. Boutilier

36

## Inference is Easy



- Want to know  $P(a)$ ? Use summing out rule:

$$\begin{aligned}
 P(a) &= \sum_{c_i \in \text{Dom}(C)} \Pr(a | c_i) \Pr(c_i) \\
 &= \sum_{c_i \in \text{Dom}(C)} \Pr(a | c_i) \sum_{e_i \in \text{Dom}(E)} \Pr(c_i | e_i) \Pr(e_i)
 \end{aligned}$$

These are all terms specified in our local distributions!

CSC 384 Lecture Slides (c) 2002, C. Boutilier

37

## Inference is Easy



- Computing  $P(a)$  in more concrete terms:

- $P(c) = P(c|e)P(e) + P(c|\neg e)P(\neg e)$   
 $= 0.8 * 0.7 + 0.5 * 0.3 = 0.78$
- $P(\neg c) = P(\neg c|e)P(e) + P(\neg c|\neg e)P(\neg e) = 0.22$   
  - $P(\neg c) = 1 - P(c)$ , as well
- $P(a) = P(a|c)P(c) + P(a|\neg c)P(\neg c)$   
 $= 0.7 * 0.78 + 0.0 * 0.22 = 0.546$
- $P(\neg a) = 1 - P(a) = 0.454$

CSC 384 Lecture Slides (c) 2002, C. Boutilier

38

## Bayesian Networks

- The structure above is a *Bayesian network*. A BN is a *graphical representation* of the direct dependencies over a set of variables, together with a set of *conditional probability tables* quantifying the strength of those influences.
- Bayes nets generalize the above ideas in very interesting ways, leading to effective means of representation and inference under uncertainty.
- We'll get into the details and formal definitions next time!

CSC 384 Lecture Slides (c) 2002, C. Boutilier

39