Reasoning under Uncertainty: Conditional Prob., Bayes and Independence

Computer Science cpsc322, Lecture 25

(Textbook Chpt 6.1-2)

March, 11, 2008

Lecture Overview

- **Recap Semantics of Probability**
- **Marginalization**
- Conditional Probability
- Chain Rule
- Bayes' Rule
- Independence

Recap: Possible World Semantics for Probabilities

Probability is a formal measure of subjective uncertainty.

X

- Random variable and probability distribution

$dom(X) = \{x_1, x_2, x_3\}$

$x_1 \rightarrow P(x_1)$

$x_2 \rightarrow P(x_2)$

$x_3 \rightarrow \ldots$

$\sum P(x_i)$

- Model Environment with a set of random vars

$X_1 X_2 X_3$

$$\sum_{\omega \in W} \mu(\omega) = 1$$

- Probability of a proposition *f*

$$P(f) = \sum_{\omega \models f} \mu(\omega)$$

Joint Distribution and Marginalization

cavity	toothache	catch	$\mu(w)$
T	T	T	.108
T	T	F	.012
T	F	T	.072
T	F	F	.008
F	T	T	.016
F	T	F	.064
F	F	T	.144
F	F	F	.576

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

Given a joint distribution, e.g. *P(X,Y, Z)* we can compute distributions over any smaller sets of variables

$$P(X,Y) = \sum_{z \in dom(Z)} P(X,Y,Z=z)$$

catch

P(cavity=T, toothache=T)

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

cavity	toothache	P(cavity , toothache)
T	T	.12
T	F	.08
F	T	.08
F	F	.72

Why is it called Marginalization?

P(X, Y)

cavity	*toothache*	*P(cavity , toothache)*
T	T	.12
T	F	.08
F	T	.08
F	F	.72

A

$$P(X) = \sum_{y \in dom(Y)} P(X, Y = y)$$

B

	Toothache = T	*Toothache = F*	
Cavity = T	.12	.08	.2
Cavity = F	.08	.72	.8
	.2	.8	

P(cavity)

P(toothache)

Lecture Overview

- Recap Semantics of Probability
- Marginalization
- **Conditional Probability**
- **Chain Rule**
- Bayes' Rule
- Independence

Conditioning
(Conditional Probability)

- We model our environment with a set of random variables.
- We have the joint, we can compute the probability of any formula.
- Are we done with reasoning under uncertainty?
- What can happen?
- Think of a patient showing up at the dentist office. Does she have a cavity?

Conditioning (Conditional Probability)

- Probabilistic conditioning specifies **how to revise beliefs based on new information.**
- You build a probabilistic model (for now the joint) taking all background information into account. This gives the prior probability.
- All other information must be conditioned on.
- If evidence *e* is all of the information obtained subsequently, the conditional probability $P(h|e)$ of *h* given *e* is the posterior probability of h.

Conditioning Example

- Prior probability of having a cavity

$P(cavity = T)$

- Should be revised if you know that there is toothache

$P(cavity = T \mid toothache = T)$

- It should be revised again if you were informed that the probe did not catch anything

$P(cavity = T \mid toothache = T, catch = F)$

- What about ?

$P(cavity = T \mid sunny = T)$

How can we compute P(h|e)

- What happens in term of possible worlds if we know the value of a random var (or a set of random vars)?
- Some worlds are ruled out. The other become ….

more likely

$P(e) = .2$

cavity	*toothache*	*catch*	$\mu(w)$	$\mu_e(w)$
T	T	T	.108	.54
T	T	F	.012	.06
T	F	T	.072	.36
T	F	F	.008	.04
~~F~~	~~T~~	~~T~~	~~.016~~	0
~~F~~	~~T~~	~~F~~	~~.064~~	0
~~F~~	~~F~~	~~T~~	~~.144~~	0
~~F~~	F	F	.576	0

$e = (cavity = T)$

$$\mu_e(w) = \frac{\mu(w)}{P(e)}$$

Semantics of Conditional Probability

$$\mu_e(w) = \begin{cases} \boxed{\dfrac{1}{P(e)} \times \mu(w)} & if \quad w \models e \\ 0 & if \quad w \not\models e \end{cases}$$

- The conditional probability of formula ***h*** given evidence **e** is

$$P(h \mid e) = \underbrace{\sum_{w \models h} \mu_e(w)}_{A} = \sum_{w \models h \wedge e} \boxed{\frac{1}{P(e)} \times \mu(w)} = \frac{1}{P(e)} \boxed{\sum_{w \models h \wedge e} \mu(w)} =$$

$$= \underbrace{\frac{P(h \wedge e)}{P(e)}}_{B}$$

Semantics of Conditional Prob.: Example

cavity	*toothache*	*catch*	$\mu(w)$	$\mu_e(w)$
T	T	T	.108	.54
T	T	F	.012	.06
T	F	T	.072	.36
T	F	F	.008	.04
F	T	T	.016	0
F	T	F	.064	0
F	F	T	.144	0
F	F	F	.576	0

$e = (cavity = T)$

$$\frac{P(h \wedge e)}{P(e)} \quad B$$

$$\frac{.12}{.2} = .6$$

$$P(h \mid e) = P(\underbrace{toothache = T}_{h} \mid \underbrace{cavity = T}_{e}) =$$

A $\sum_{w \models h} \mu_e(w) = .6$

Conditional Probability among Random Variables

TRY

P(cavity | toothache)

$$P(X \mid Y) = P(X, Y) / P(Y)$$

$$P(X \mid Y) = P(toothache \mid cavity)$$
$$= P(toothache \wedge cavity) / P(cavity)$$

	Toothache = T	Toothache = F
Cavity = T	.12 /.2	.08 /.2
Cavity = F	.08 /.8	.72 /.8
	.2	.8

P(X, Y)

0.2

0.8

	Toothache = T	Toothache = F
Cavity = T	.6	.4
Cavity = F	.1	.9

P(X|Y)

P(toothache | cavity = T)

P(toothache | cavity = F)

$\Sigma = 1$

$\Sigma = 1$

TWO PROB. DISTRIBUTIONS

Product Rule

- Definition of conditional probability:
  - $P(X_1 \mid X_2) = P(X_1 \wedge X_2) / P(X_2)$
- Product rule gives an alternative, more intuitive formulation:
  - $P(X_1 \wedge X_2) = P(X_2)\, P(X_1 \mid X_2) = \boxed{P(X_1)\, P(X_2 \mid X_1)}$
- Product rule general form:

$$\mathbf{P}(X_1, \ldots, X_n) = P(X_1 \ldots X_t, X_{t+1} \ldots X_n)$$

$$= \mathbf{P}(X_1, \ldots, X_t)\, \mathbf{P}(X_{t+1} \ldots X_n \mid X_1, \ldots, X_t)$$

Chain Rule

- Product rule general form:

$$\mathbf{P}(X_1, \ldots, X_n) =$$

$$= \mathbf{P}(X_1,\ldots,X_t)\, \mathbf{P}(X_{t+1} \ldots X_n \mid X_1,\ldots,X_t)$$

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

$t = n-1$

$$\mathbf{P}(X_1, \ldots X_{n-1}, X_n) =$$

$$= \mathbf{P}(X_1,\ldots,X_{n-1})\, \mathbf{P}(X_n \mid X_1,\ldots,X_{n-1})$$

$$= \mathbf{P}(X_1,\ldots,X_{n-2})\, \mathbf{P}(X_{n-1} \mid X_1,\ldots,X_{n-2})\, \mathbf{P}(X_n \mid X_1,\ldots,X_{n-1}) = \ldots$$

$$= \mathbf{P}(X_1)\, \mathbf{P}(X_2 \mid X_1) \ldots \mathbf{P}(X_{n-1} \mid X_1,\ldots,X_{n-2})\, \mathbf{P}(X_n \mid X_1,\ldots,X_{n-1})$$

$$= \prod_{i=1}^{n} \mathbf{P}(X_i \mid X_1, \ldots, X_{i-1})$$

Chain Rule: Example

P(*cavity* , *toothache*, *catch*) =

$P(\text{cavity}) * P(\text{toothache} \mid \text{cavity}) *$

$* \; P(\text{catch} \mid \text{cavity}, \text{toothache})$

P(*toothache*, *catch*, *cavity*) =

$P(\text{toothache}) * P(\text{catch} \mid \text{toothache}) * P(\text{cavity} \mid \text{toothache}, \text{catch})$

these and the other four decompositions are OK

Lecture Overview

- Recap Semantics of Probability
- Marginalization
- Conditional Probability
- Chain Rule
- **Bayes' Rule**
- **Independence**

Bayes' Rule

- From Product rule :
  - $P(X , Y) = P(Y)\, P(X \mid Y) = P(X)\, P(Y \mid X)$

STOP HERE

Do you always need to revise your beliefs?

…… when your knowledge of **Y**'s value doesn't affect your belief in the value of **X**

DEF. Random variable **X** is marginal independent of random variable **Y** if, for all $x_i \in dom(X)$, $y_k \in dom(Y)$,

$$P(X= x_i \mid Y= y_k) = P(X= x_i)$$

Consequence:

$$P(X= x_i , Y= y_k) = P(X= x_i \mid Y= y_k) P(Y= y_k) =$$

$$= P(X= x_i) P(Y= y_k)$$

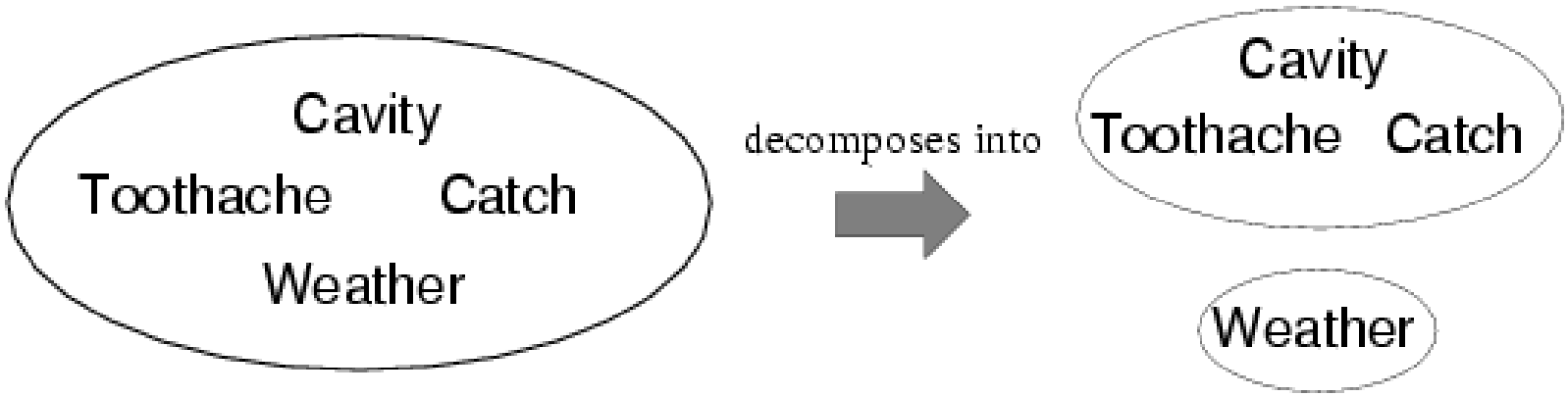
Marginal Independence: Example

- *A* and *B* are independent iff:

  $\mathbf{P}(A|B) = \mathbf{P}(A)$ or $\mathbf{P}(B|A) = \mathbf{P}(B)$ or $\mathbf{P}(A, B) = \mathbf{P}(A)\,\mathbf{P}(B)$

- That is new evidence B (or A) does not affect current belief in A (or B)
- Ex: $\mathbf{P}(Toothache, Catch, Cavity, Weather)$
  $= \mathbf{P}(Toothache, Catch, Cavity)\,\mathbf{P}(Weather)$

- JPD requiring entries is reduced to two smaller ones (and)



Learning Goals for today's class

- **You can:**
- Given a joint, compute distributions over any subset of the variables
- Prove the formula to compute $P(h|e)$
- Derive the **Chain Rule** and **the Bayes Rule**
- Define **Marginal Independence**

Next Class

- Conditional Independence
- Belief Networks…….

Assignments

- I will post Assignment 3 this evening
- Assignment2
  - Will post solutions for first two questions
  - Generic feedback on programming (see WebCT)
  - If you have more programming questions, office hours next M W (Jacek)

Plan for this week

- Probability is a rigorous formalism for uncertain knowledge
- Joint probability distribution specifies probability of every possible world
- Probabilistic queries can be answered by summing over possible worlds
- For nontrivial domains, we must find a way to reduce the joint distribution size
- **Independence** *(rare)* and **conditional independence** *(frequent)* provide the tools

Conditional probability (irrelevant evidence)

- New evidence may be irrelevant, allowing simplification, e.g.,
  - P(*cavity* | *toothache, sunny*) = P(*cavity* | *toothache*)
  - We say that Cavity is conditionally independent from Weather (more on this next class)
- This kind of inference, sanctioned by domain knowledge, is crucial in probabilistic inference