Reasoning under Uncertainty: Marginalization, Conditional Prob., and Bayes

Computer Science cpsc322, Lecture 25

(Textbook Chpt 6.1.3.1-2)

June, 13, 2017

Lecture Overview

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

Recap: Possible World Semantics for Probabilities

Probability is a formal measure of subjective uncertainty.

- Random variable and probability distribution

X

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

$$\left.\begin{array}{l} x_1 \rightarrow P(x_1) \\ x_2 \rightarrow P(x_2) \\ x_3 \rightarrow P(x_3) \end{array}\right\} \sum = 1$$

- Model Environment with a set of random vars

$X \; Y \; Z$ binary $\quad 8$

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

formula

- Probability of a proposition f $\quad X = T \wedge Z = F$

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

Joint Distribution and Marginalization

$P(X,Y,Z)$

$P(cavity, toothache, catch)$

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

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

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

$P(cavity, toothache)$

	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

Joint Distribution and Marginalization

P(X,Y,Z)

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

cavity	*toothache*	*catch*	*µ(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

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

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

cavity	*catch*	A. *P(cavity , catch)*	B. *P(cavity , catch)*	C. *P(cavity , catch)*
T	T	.12	.18	.18
T	F	.08	.02	.72
F	T	…	….	….
F	F	…	….	….

Joint Distribution and Marginalization

$P(X,Y,Z)$

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

cavity	*toothache*	*catch*	*µ(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

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

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

iclicker.

cavity	*catch*	A. P(cavity , catch)	B. P(cavity , catch)	C. P(cavity , catch)
T	T	.12	.18	.18
T	F	.08	.02	.72
F	T	…		
F	F	…		

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

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

P(cavity)

	Toothache = T	Toothache = F
Cavity = T	.12	.08
Cavity = F	.08	.72

.2

.8

.2 .8

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**.
- Assume 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.

P(cavity = T | toothache = T) ?

Conditioning Example

- Prior probability of having a cavity

$P(cavity = T)$

- Should be revised if you know that there is toothache

$P(cavity = T \mid \underline{toothache = T})$

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

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

- What about ?

$P(cavity = T \mid \underline{sunny = T})$

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

$\Sigma = P(e) = .2$

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

$e = (cavity = T)$

$$\begin{cases} \mu_e(w) = \dfrac{\mu(w)}{P(e)} \\ \text{if } w \models e \end{cases}$$

$$\begin{cases} \mu_e(w) = 0 \\ \text{if } w \not\models e \end{cases}$$

How can we compute P(h|e)

$$P(h \mid e) = \sum_{w \models h} \mu_e(w)$$

$$P(toothache = F \mid cavity = T) = \sum_{w \models \, toothache = F} \mu_{cavity=T}(w)$$

cavity	*toothache*	*catch*	$\mu(w)$	$\mu_{cavity=T}(w)$
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

Semantics of Conditional Probability

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

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

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

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

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

$e = (cavity = T)$

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

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

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

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

Conditional Probability among Random Variables

TRY

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

$$\underline{P(X \mid Y)} = \underline{P(X , Y)} \,/\, \underline{P(Y)}$$

$$P(X \mid Y) = P(\underline{\text{toothache}} \mid \text{cavity})$$
$$= \boxed{P(\underline{\text{toothache}} \wedge \underline{\text{cavity}})} \,/\, \underline{P(\text{cavity})}$$

TWO PROB DISTRIBUTIONS

P(X, Y)

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

0.2

0.8

P(X | Y)

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

P(toothache | cavity = T)

P(toothache | cavity = F)

$\Sigma = 1$

$\Sigma = 1$

Product Rule

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

$t = n - 1$

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

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

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

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

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

Chain Rule: Example

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

$$P(cavity) * P(toothache \mid cavity) * P(catch \mid cavity, toothache)$$

P(*toothache, catch, cavity*) =

$$P(toothache) * P(catch \mid toothache) * P(cavity \mid toothache, catch)$$

In how many other ways can this joint be decomposed using the chain rule?

iclicker.

A. 4 B. 1 C. 8 D. 0

Chain Rule: Example

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

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

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

$$P(toothache) * P(catch \mid toothache) * P(cavity \mid toothache, catch)$$

these and the other four decompositions are OK

Lecture Overview

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

Using conditional probability

- Often you have causal knowledge (forward from cause to evidence):
  - For example
    - ✓ P(symptom | disease)
    - ✓ P(light is off | status of switches and switch positions)
    - ✓ P(alarm | fire)
  - In general: P(evidence e | hypothesis h)

- ... and you want to do evidential reasoning (backwards from evidence to cause):
  - For example
    - ✓ P(disease | symptom)
    - ✓ P(status of switches | light is off and switch positions)
    - ✓ P(fire | alarm)
  - In general: P(hypothesis h | evidence e)

Bayes Rule

- By definition, we know that :

$$P(h|e) = \frac{P(h \wedge e)}{P(e)} \qquad P(e|h) = \frac{P(e \wedge h)}{P(h)}$$

- We can rearrange terms to write

$$P(h \wedge e) = P(h|e) \times P(e) \qquad (1)$$

$$P(e \wedge h) = P(e|h) \times P(h) \qquad (2)$$

- But

$$P(h \wedge e) = P(e \wedge h) \qquad (3)$$

- From (1) (2) and (3) we can derive
- **Bayes Rule**

$$P(h|e) = \frac{P(e|h)P(h)}{P(e)} \qquad (3)$$

Example for Bayes rule

- On average, the alarm rings once a year
  - $P(alarm) = ?$

Prob. of ringing on a given day?

- If there is a fire, the alarm will almost always ring
- On average, we have a fire every 10 years
- The fire alarm rings. What is the probability there is a fire?

Example for Bayes rule

- On average, the alarm rings once a year
  - $P(alarm) = 1/365$
- If there is a fire, the alarm will almost always ring
  - $P(alarm|fire) = 0.999$
- On average, we have a fire every 10 years
  - $P(fire) = 1/3650$
- The fire alarm rings. What is the probability there is a fire?
  - Take a few minutes to do the math!

$$P(h \mid e) = \frac{P(e \mid h)P(h)}{P(e)}$$

iclicker

A. 0.999 | B. 0.9 | C. 0.0999 | D. 0.1

Example for Bayes rule

- On average, the alarm rings once a year
  - $P(alarm) = 1/365$
- If there is a fire, the alarm will almost always ring
  - $P(alarm|fire) = 0.999$
- On average, we have a fire every 10 years
  - $P(fire) = 1/3650$
- The fire alarm rings. What is the probability there is a fire?
- $P(fire|\mathrm{alarm}) = \frac{P(alarm|fire) \times P(fire)}{P(alarm)} = \frac{0.999 \times 1/3650}{1/365} = 0.0999$
  - Even though the alarm rings the chance for a fire is only about 10%!

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**

Next Class

- Marginal Independence
- Conditional Independence

Assignments

- Assignment 3 has been posted : due jone 20th

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