CSE 473: Artificial Intelligence
Spring 2014

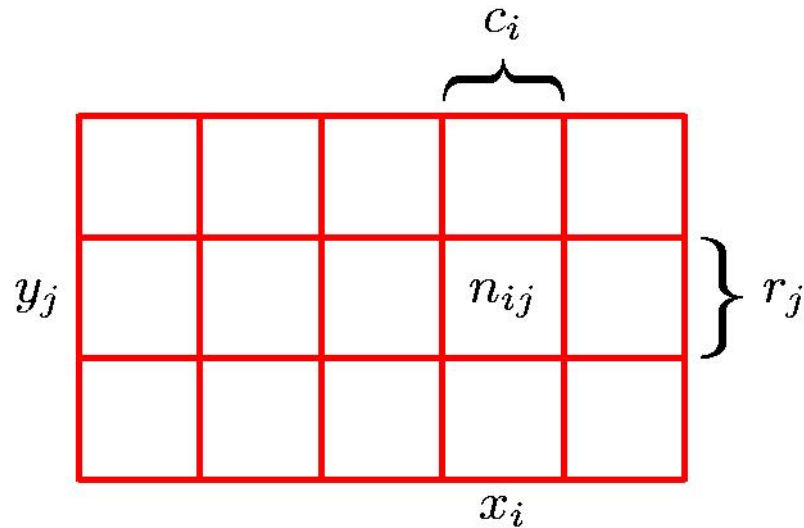## Uncertainty & Probabilistic Reasoning

Hanna Hajishirzi

Many slides adapted from Pieter Abbeel, Dan Klein, Dan Weld,Stuart Russell, Andrew Moore & Luke Zettlemoyer

Announcements

- Project 1 grades
- Resubmission policy

Terminology

Marginal Probability

$$p(X = x_i) = \frac{c_i}{N}.$$

Joint Probability

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$
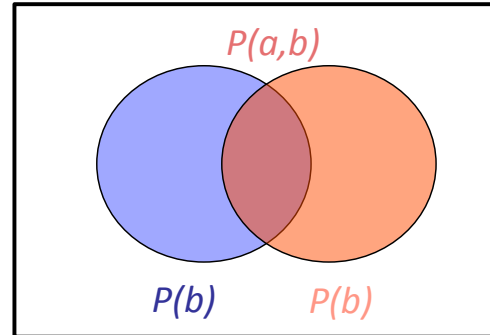
Conditional Probability

$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}$$

X value is given

Conditional Probability

- A simple relation between joint and conditional probabilities
  - In fact, this is taken as the *definition* of a conditional probability

$$P(a|b) = \frac{P(a,b)}{P(b)}$$

P(a,b)

P(b)   *P(b)*

$P(T,W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$$P(W = s|T = c) = \frac{P(W = s, T = c)}{P(T = c)} = \frac{0.2}{0.5} = 0.4$$

$$= P(W = s, T = c) + P(W = r, T = c)$$
$$= 0.2 + 0.3 = 0.5$$

Probabilistic Inference

- Diagnosis
- Speech recognition
- Tracking objects
- Robot mapping
- Genetics
- Error correcting codes
- ... lots more!

Probabilistic Inference

- **Probabilistic inference**: compute a desired probability from other known probabilities (e.g. conditional from joint)

- We generally compute conditional probabilities
  - P(on time | no reported accidents) = 0.90
  - These represent the agent's beliefs given the evidence

- Probabilities change with new evidence:
  - P(on time | no accidents, 5 a.m.) = 0.95
  - P(on time | no accidents, 5 a.m., raining) = 0.80
  - Observing new evidence causes beliefs to be updated

Inference by Enumeration

- P(sun)?

S	T	W	P
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

Inference by Enumeration

-

- P(sun | winter)?

S	T	W	P
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

Inference by Enumeration

-

-

- P(sun | winter, hot)?

S	T	W	P
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

Inference by Enumeration

- P(sun)?

S	T	W	P
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

Inference by Enumeration

-
- P(sun | winter)?

S	T	W	P
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

Uncertainty

- **General situation:**

  - **Observed variables (evidence)**: Agent knows certain things about the state of the world (e.g., sensor readings or symptoms)

  - **Unobserved variables**: Agent needs to reason about other aspects (e.g. where an object is or what disease is present)

  - **Model**: Agent knows something about how the known variables relate to the unknown variables

- Probabilistic reasoning gives us a framework for managing our beliefs and knowledge

Inference by Enumeration

- General case:
    - Evidence variables: $E_1 \ldots E_k = e_1 \ldots e_k$
    - Query* variable: Q
    - Hidden variables: $H_1 \ldots H_r$

$X_1, X_2, \ldots X_n$

All variables

- We want: $P(Q|e_1 \ldots e_k)$

- First, select the entries consistent with the evidence

- Second, sum out H to get joint of Query and evidence:

$$P(Q, e_1 \ldots e_k) = \sum_{h_1 \ldots h_r} \underbrace{P(Q, h_1 \ldots h_r, e_1 \ldots e_k)}_{X_1, X_2, \ldots X_n}$$

- Finally, normalize the remaining entries to conditionalize

Supremacy of the Joint Distribution

- P(sun)?

- P(sun | winter)?

- P(sun | winter, hot)?

S	T	W	P
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

Problems with Enumeration

- **Obvious problems:**
  - Worst-case time complexity $O(d^n)$
  - Space complexity $O(d^n)$ to store the joint distribution

- **Solutions**
  - Better techniques
  - Better representation
  - Simplifying assumptions

The Product Rule

- Sometimes have conditional distributions but want the joint
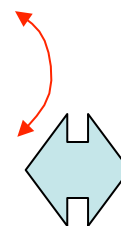
$$P(x|y) = \frac{P(x,y)}{P(y)} \quad \Longleftrightarrow \quad P(x,y) = P(x|y)P(y)$$

- Example:

$P(D|W)$

$P(D,W)$

$P(W)$

W	P
sun	0.8
rain	0.2

D	W	P
wet	sun	0.1
dry	sun	0.9
wet	rain	0.7
dry	rain	0.3

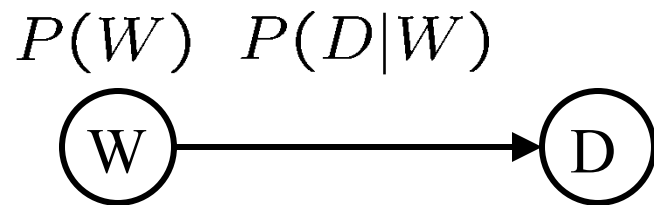| D	W	P
wet	sun	0.08
dry	sun	0.72
wet	rain	0.14
dry	rain	0.06

The Product Rule

- Sometimes have conditional distributions but want the joint

$$P(x|y) = \frac{P(x,y)}{P(y)} \qquad \Longleftrightarrow \qquad P(x,y) = P(x|y)P(y)$$

- Example:

$$P(D,W)$$

$$P(W) \quad P(D|W)$$

$$\text{W} \longrightarrow \text{D}$$

The Chain Rule

- More generally, can always write any joint distribution as an incremental product of conditional distributions?

$$P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)$$

$$P(x_1, x_2, \ldots x_n) = \prod_i P(x_i|x_1 \ldots x_{i-1})$$

- Why is this always true?

Bayes' Rule

- Two ways to factor a joint distribution over two variables:

$$P(x, y) = P(x|y)P(y) = P(y|x)P(x)$$

That's my rule!

- Dividing, we get:

$$P(x|y) = \frac{P(y|x)}{P(y)} P(x)$$

- Why is this at all helpful?
    - Lets us build a conditional from its reverse
    - Often one conditional is tricky but the other one is simple
    - Foundation of many systems we'll see later

- In the running for most important AI equation!

Inference with Bayes' Rule

- Example: Diagnostic probability from causal probability:

$$P(\text{Cause}|\text{Effect}) = \frac{P(\text{Effect}|\text{Cause})P(\text{Cause})}{P(\text{Effect})}$$

- Example:

  - m is meningitis, s is stiff neck

  $$\left.\begin{array}{l} P(s|m) = 0.8 \\ P(m) = 0.0001 \\ P(s) = 0.1 \end{array}\right\}$$ Example givens

$$P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{0.8 \times 0.0001}{0.1} = 0.0008$$

  - Note: posterior probability of meningitis still very small
  - Note: you should still get stiff necks checked out! Why?

Quiz: Bayes Rule

■ Given:

$$P(W)$$

R	P
sun	0.8
rain	0.2

$$P(D|W)$$

D	W	P
wet	sun	0.1
dry	sun	0.9
wet	rain	0.7
dry	rain	0.3

■ What is P(W | dry) ?

Ghostbusters, Revisited

- **Let's say we have two distributions:**
  - Prior distribution over ghost location: P(G)
    - Let's say this is uniform
  - Sensor reading model: P(R | G)
    - Given: we know what our sensors do
    - R = reading color measured at (1,1)
    - E.g. P(R = yellow | G=(1,1)) = 0.1

- **We can calculate the posterior distribution P(G|r) over ghost locations given a reading using Bayes' rule:**

$$P(g|r) \propto P(r|g)P(g)$$

Independence

- **Two variables are *independent* if:**

$$\forall x, y : P(x, y) = P(x)P(y)$$

  - This says that their joint distribution *factors* into a product two simpler distributions
  - Another form:

$$\forall x, y : P(x|y) = P(x)$$

  - We write: $X \perp\!\!\!\perp Y$

- **Independence is a simplifying *modeling assumption***
  - *Empirical* joint distributions: at best "close" to independent
  - What could we assume for {Weather, Traffic, Cavity, Toothache}?

Example: Independence?

$P(T)$

T	P
hot	0.5
cold	0.5

$P_1(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$P(W)$

W	P
sun	0.6
rain	0.4

Example: Independence

- N fair, independent coin flips:

$$P(X_1, X_2, \ldots X_n)$$

2^n
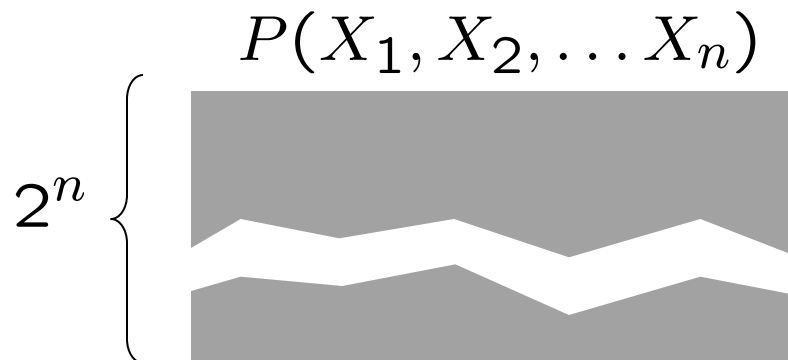
Conditional Independence

- P(Toothache, Cavity, Catch)
- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
  - P(+catch | +toothache, +cavity) = P(+catch | +cavity)
- The same independence holds if I don't have a cavity:
  - P(+catch | +toothache, -cavity) = P(+catch| −cavity)
- Catch is *conditionally independent* of Toothache given Cavity:
  - P(Catch | Toothache, Cavity) = P(Catch | Cavity)
- Equivalent statements:
  - P(Toothache | Catch , Cavity) = P(Toothache | Cavity)
  - P(Toothache, Catch | Cavity) = P(Toothache | Cavity) P(Catch | Cavity)
  - One can be derived from the other easily

Conditional Independence

- Unconditional (absolute) independence very rare (why?)

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

$$\forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$$

$$\forall x, y, z : P(x|z, y) = P(x|z)$$

$$X \perp\!\!\!\perp Y \,|\, Z$$

- What about this domain:
  - Traffic
  - Umbrella
  - Raining

Probability Summary

- **Conditional probability** $P(x|y) = \dfrac{P(x,y)}{P(y)}$

- **Product rule** $P(x,y) = P(x|y)P(y)$

- **Chain rule**
$$P(X_1, X_2, \ldots X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\ldots$$
$$= \prod_{i=1}^{n} P(X_i|X_1, \ldots, X_{i-1})$$

- **X, Y independent if and only if:** $\forall x, y : P(x,y) = P(x)P(y)$

- **X and Y are conditionally independent given Z if and only if:** $X \perp\!\!\!\perp Y | Z$
$$\forall x, y, z : P(x,y|z) = P(x|z)P(y|z)$$