PROBABILITY

Today

- Reading
  - AIMA Chapter 13.2-13.5
- Goals
  - Normalization trick
  - Inference by Enumeration
  - (Independence)

Joint probability distribution

- A joint distribution over a set of r.v.s $\{X_1, X_2, \ldots, X_n\}$ assigns probabilities to each possible assignment:

$$p(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

$$p(x_1, x_2, \ldots, x_n)$$

P(W = w, T = t)

w	t	P
sunny	hot	0.4
rain	hot	0.1
sunny	cold	0.2
rain	cold	0.3

- Still subject to constraints:

$$0 \le p(x_1, x_2, \ldots, x_n) \le 1 \quad \text{and} \quad \sum_{(x_1,\ldots,x_n)} p(x_1, \ldots, x_n) = 1$$

If we have n variables with domain size d, what is the size of the probability distribution (the number of rows in the table)?

Marginal Distributions

- Sometimes we have the joint distribution but we're only interested in the distribution of a subset of the variables
  - Called the marginal distribution
  - We "marginalize out" the other variables by summing over them
  - Corresponds to a sub-table created by summing over rows

P(W = w, T = t)

w	t	P
sunny	hot	0.4
rain	hot	0.1
sunny	cold	0.2
rain	cold	0.3

$$p(X = x) = \sum_y P(X = x, Y = y)$$

Oftentimes, the events we're interested in are marginal distributions

P(W = w)

w	P
sunny	0.6
rain	0.4

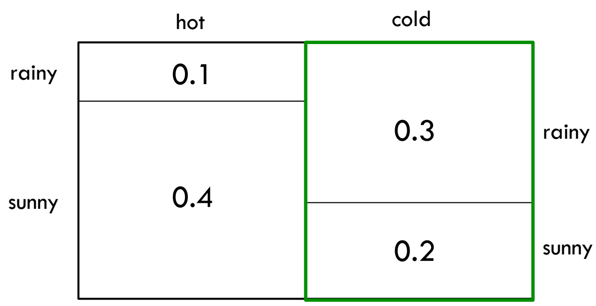
P(T = t)

t	P
hot	0.5
cold	0.5

Conditional (posterior) distribution

- The conditional distribution is given by the equation

$$p(X = x|Y = y) = \frac{p(X = x, Y = y)}{p(Y = y)}$$



p(W | T=cold)?

Whereas before the total probability mass was 1, the total probability mass is now p(T=cold). We compute everything in relation to this value.

Summary of distributions so far

$$p(x) = \sum_y p(x, y)$$

Marginal

Conditional

$$p(x|y) = \frac{p(x, y)}{p(y)}$$

Joint $p(x, y)$

Conditional and Joint are just a constant apart!

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

$$p(W = r|T = c) = \frac{p(W = r, T = c)}{p(T = c)} = \frac{0.3}{0.5} = 0.6$$

- Note that p(T=c) is constant no matter the value of W
- We call p(T=c) a normalization constant because:
  1. It is constant with respect to the distribution of interest p(W|T=c)
  2. It ensures that the distribution sums to 1 (i.e. it restores the distribution p(W|T=c) back to the "normal" condition of summing to 1)

Conditional and Joint are just a constant apart!

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

$$p(W = r|T = c) = \frac{p(W = r, T = c)}{p(T = c)} = \frac{0.3}{0.5} = 0.6$$

$$p(X, Y) \propto p(X|Y)$$

↑

"is proportional to"

Normalization Trick

- Step 1: Look up p(W, T=c) for all values of W
- Step 2: Sum these joint probabilities
- Step 3: Divide each joint probability by this sum

$$< \alpha \cdot 0.2,\ \alpha \cdot 0.3 >=< \frac{0.2}{0.2+0.3}, \frac{0.3}{0.2+0.3} >=< 0.4, 0.6 >$$

Normalization Trick

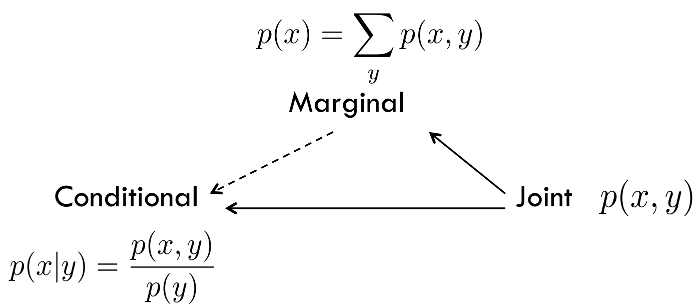
- Normalize the following distributions:

P(W = w, T = t)

w	t	P
sunny	hot	0.4
rain	hot	0.1
sunny	cold	0.2
rain	cold	0.3

p(W| T=hot)?
p(W| T=cold)?
p(T| W=sunny)?
p(T| W=rainy)?

Summary of distributions so far



Probabilistic Inference

- Probabilistic inference refers to the task of computing some desired probability given other known probabilities (evidence)
- Typically compute the conditional (posterior) probability of an event
  - p(on time | no accidents) = 0.80
- Probabilities change with new evidence
  - p(on time | no accidents, 5 a.m.) = 0.95
  - p(on time | no accidents, 5 a.m., raining) = 0.8

Inference by Enumeration

- Have a set of random variables $\{X_1, X_2, \ldots, X_n\}$
- Partition the set of random variables into:
  - Evidence variables: $E_1=e_1, E_2=e_2,\ldots E_k=e_k$
  - Query variables: Q
  - Hidden (misc.) variables: $H_1, H_2, \ldots, H_r$

We want to compute:
$p(Q \mid E_1=e_1, E_2=e_2, \ldots, E_k=e_k)$

Inference by Enumeration

Step One: select the entries in the table consistent with the evidence (this becomes our world)

Step Two: sum over the H variables to get the joint distribution of the query and evidence variables

Step Three: Normalize

$$p(Q, e_1, \ldots, e_k) = \sum_{(h_1, \ldots, h_r)} p(Q, e_1, \ldots, e_k, h_1, \ldots, h_r)$$

$$Z = \sum_q p(Q = q, e_1, \ldots, e_k)$$

$$p(Q|e_1, \ldots, e_k) = \frac{1}{Z} \cdot p(Q, e_1, \ldots, e_k)$$

Inference by Enumeration

Step One: select the entries in the table consistent with the evidence (this becomes our world)

Step Two: sum over the H variables to get the joint distribution of the query and evidence variables

Step Three: Normalize

p(W | S = winter)?

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

Step One

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

Inference by Enumeration

Step One: select the entries in the table consistent with the evidence (this becomes our world)

Step Two: sum over the H variables to get the joint distribution of the query and evidence variables

Step Three: Normalize

p(W | S = winter)?

S	W	T	P
winter	sunny	hot	0.10
winter	rain	hot	0.05
winter	sunny	cold	0.15
winter	rain	cold	0.20

Step Two

$$p(Q, e_1, \ldots, e_k) = \sum_{(h_1, \ldots, h_r)} p(Q, e_1, \ldots, e_k, h_1, \ldots, h_r)$$

Inference by Enumeration

Step One: select the entries in the table consistent with the evidence (this becomes our world)

Step Two: sum over the H variables to get the joint distribution of the query and evidence variables

Step Three: Normalize

p(W | S = winter)?

S	W	T	P
winter	sunny	hot	0.10
winter	rain	hot	0.05
winter	sunny	cold	0.15
winter	rain	cold	0.20

Step Three

$$Z = \sum_q p(Q = q, e_1, \dots, e_k)$$

$$p(Q|e_1, \dots, e_k) = \frac{1}{Z} \cdot p(Q, e_1, \dots, e_k)$$

Inference by Enumeration

Step One: select the entries in the table consistent with the evidence (this becomes our world)

Step Two: sum over the H variables to get the joint distribution of the query and evidence variables

Step Three: Normalize

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

Queries:

p(W | S=winter, T=hot)?

p(S, W)?

p(S, W | T=hot)?

Inference by Enumeration

- n random variables
- d domain size
- Worst-case time is $O(d^n)$
- Space is $O(d^n)$ to save entire table in memory

- Is there something better?

Notions of Independence

Lecture proceeds on the whiteboard!