CS 331: Artificial Intelligence
Fundamentals of Probability II

Thanks to Andrew Moore for some course material

Full Joint Probability Distributions

Coin	Card	Candy	P(Coin, Card, Candy)
tails	black	1	0.15
tails	black	2	0.06
tails	black	3	0.09
tails	red	1	0.02
tails	red	2	0.06
tails	red	3	0.12
heads	black	1	0.075
heads	black	2	0.03
heads	black	3	0.045
heads	red	1	0.035
heads	red	2	0.105
heads	red	3	0.21

The probabilities in the last column sum to 1

This cell means P(Coin=heads, Card=red, Candy=3) = 0.21

Joint Probability Distribution

From the full joint probability distribution, we can calculate any probability involving these three random variables.

e.g. P(*Coin* = *heads* OR *Card* = *red*)

Joint Probability Distribution

P(*Coin* = *heads* OR *Card* = *red*) =

P(*Coin=heads, Card=black, Candy=1*) +
P(*Coin=heads, Card=black, Candy=2*) +
P(*Coin=heads, Card=black, Candy=3*) +
P(*Coin=tails, Card=red, Candy=1*) +
P(*Coin=tails, Card=red, Candy=2*) +
P(*Coin=tails, Card=red, Candy=3*) +
P(*Coin=heads, Card=red, Candy=1*) +
P(*Coin=heads, Card=red, Candy=2*) +
P(*Coin=heads, Card=red, Candy=3*)

= 0.075 + 0.03 + 0.045 + 0.02 + 0.06 + 0.12 + 0.035 + 0.105 + 0.21 = 0.7

Marginalization

We can even calculate marginal probabilities (the probability distribution over a subset of the variables) e.g.:

P(*Coin=tails*, *Card=red*) =

P(*Coin=tails*, *Card=red*, *Candy=1*) +

P(*Coin=tails*, *Card=red*, *Candy=2*) +

P(*Coin=tails*, *Card=red*, *Candy=3*)

= 0.02 + 0.06 + 0.12 = 0.2

Marginalization

Or even:

P(*Card=black*) =
P(*Coin=heads, Card=black, Candy=1*) +
P(*Coin=heads, Card=black, Candy=2*) +
P(*Coin=heads, Card=black, Candy=3*) +
P(*Coin=tails, Card=black, Candy=1*) +
P(*Coin=tails, Card=black, Candy=2*) +
P(*Coin=tails, Card=black, Candy=3*)

= 0.075 + 0.03 + 0.045 + 0.015 + 0.06 + 0.09 = 0.315

Marginalization

The general marginalization rule for any **sets** of variables ***Y*** and ***Z***:

$$\boldsymbol{P}(\boldsymbol{Y}) = \sum_{\mathbf{z}} \boldsymbol{P}(\boldsymbol{Y}, \mathbf{z})$$

z is over all possible combinations of values of ***Z*** (remember ***Z*** is a set)

or

$$\boldsymbol{P}(\boldsymbol{Y}) = \sum_{\mathbf{z}} \boldsymbol{P}(\boldsymbol{Y} \mid \mathbf{z}) P(\mathbf{z})$$

Marginalization

For continuous variables, marginalization involves taking the integral:

$$\boldsymbol{P}(\boldsymbol{Y}) = \int \boldsymbol{P}(\boldsymbol{Y}, \mathbf{z}) d\mathbf{z}$$

CW: Practice

Compute $P(Candy = 2)$.

Coin	Card	Candy	P(Coin, Card, Candy)
tails	black	1	0.15
tails	black	2	0.06
tails	black	3	0.09
tails	red	1	0.02
tails	red	2	0.06
tails	red	3	0.12
heads	black	1	0.075
heads	black	2	0.03
heads	black	3	0.045
heads	red	1	0.035
heads	red	2	0.105
heads	red	3	0.21

Conditional Probabilities

We can also compute conditional probabilities from the joint. Recall:

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

Conditional Probabilities

$P(Coin = heads|Card = black)$

$$= \frac{P(Coin{=}heads, Card{=}black)}{P(Card{=}black)}$$

$$= \frac{0.075+0.03+0.045}{0.15+0.06+0.09+0.075+0.03+0.045} = 0.333$$

$P(Coin = tails|Card = black)$

$$= \frac{P(Coin{=}tails, Card{=}black)}{P(Card{=}black)}$$

$$= \frac{0.15+0.06+0.09}{0.15+0.06+0.09+0.075+0.03+0.045} = 0.667$$

Conditional Probabilities

$P(Coin = heads|Card = black)$

$$= \frac{P(Coin{=}heads, Card{=}black)}{P(Card{=}black)}$$

$$= \frac{0.075+0.03+0.045}{0.15+0.06+0.09+0.075+0.03+0.045} = 0.333$$

$P(Coin = tails|Card = black)$

Note that 1/P(*Card=black*) remains constant in the two equations.

$$= \frac{P(Coin{=}tails, Card{=}black)}{P(Card{=}black)}$$

$$= \frac{0.15+0.06+0.09}{0.15+0.06+0.09+0.075+0.03+0.045} = 0.667$$

Normalization

- In fact, $1/P(Card)$ can be viewed as a normalization constant for $\mathbf{P}(Coin|\ Card)$, ensuring it adds up to 1
- We will refer to normalization constants with the symbol α

$$\mathbf{P}(Coin|black) = \alpha\mathbf{P}(Coin, black)$$

13

CW: Practice

Compute $P(Candy = 1|Card = red)$.

Coin	Card	Candy	P(Coin, Card, Candy)
tails	black	1	0.15
tails	black	2	0.06
tails	black	3	0.09
tails	red	1	0.02
tails	red	2	0.06
tails	red	3	0.12
heads	black	1	0.075
heads	black	2	0.03
heads	black	3	0.045
heads	red	1	0.035
heads	red	2	0.105
heads	red	3	0.21

14

Inference

- Suppose you get a query such as $P(Card = red \mid Coin = heads)$

Coin is called the evidence variable because we observe it. More generally, it's a set of variables.

Card is called the query variable (we'll assume it's a single variable for now)

There are also unobserved (aka hidden) variables like *Candy*

15

Inference

- We will write the query as $\mathbf{P}(X \mid \mathbf{e})$

This is a probability distribution hence the boldface

X = Query variable (a single variable for now)

$\mathbf{E}$ = Set of evidence variables

$\mathbf{e}$ = the set of observed values for the evidence variables

$\mathbf{Y}$ = Unobserved variables

16

Inference

We will write the query as $\mathbf{P}(X \mid \mathbf{e})$

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\mathbf{P}(X, \mathbf{e}) = \alpha\sum_{\mathbf{y}}\mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

Summation is over all possible combinations of values of the unobserved variables $\mathbf{Y}$

X = Query variable (a single variable for now)

$\mathbf{E}$ = Set of evidence variables

$\mathbf{e}$ = the set of observed values for the evidence variables

$\mathbf{Y}$ = Unobserved variables

Inference

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\mathbf{P}(X, \mathbf{e}) = \alpha\sum_{\mathbf{y}}\mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

Computing $\mathbf{P}(X \mid \mathbf{e})$ involves going through all possible entries of the full joint probability distribution and adding up probabilities with $X=x_i$, $\mathbf{E}=\mathbf{e}$, and $\mathbf{Y}=\mathbf{y}$

Suppose you have a domain with n Boolean variables. What is the space and time complexity of computing $P(X \mid \mathbf{e})$?

18

Independence

- How do you avoid the exponential space and time complexity of inference?
- Use independence (aka factoring)

Independence

We say that variables X and Y are independent if any of the following hold: (note that they are all equivalent)

$\boldsymbol{P}(X \mid Y) = \boldsymbol{P}(X)$ or

$\boldsymbol{P}(Y \mid X) = \boldsymbol{P}(Y)$ or

$\boldsymbol{P}(X, Y) = \boldsymbol{P}(X)\boldsymbol{P}(Y)$

Independence

Consider the full joint distribution over these variables:

$Card = \{red, black\}$

$Candy = \{1,2,3\}$

By the product rule, we know:

$$P(Card, Candy)$$
$$= P(Card|Candy)P(Candy)$$

Independence

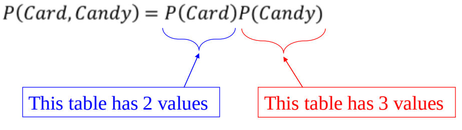
Suppose I tell you that these two events are independent (i.e. they do not influence each other).

Then:

$$P(Card, Candy)$$
$$= P(Card|Candy)P(Candy)$$
$$= P(Card)P(Candy)$$

Why is independence useful?

$P(Card, Candy) = P(Card)P(Candy)$



- You now need to store 5 values to calculate $\boldsymbol{P}(Coin, Card, Candy)$
- Without independence, we needed 6

Independence

Another example:

- Suppose you have n coin flips and you want to calculate the joint distribution $\boldsymbol{P}(C_1, \ldots, C_n)$
- If the coin flips are not independent, you need 2^n values in the table
- If the coin flips are independent, then

$$P(C_1, \ldots, C_n) = \prod_{i=1}^{n} P(C_i)$$

Each $P(C_i)$ table has 2 entries and there are *n* of them for a total of $2n$ values

Independence

- Independence is powerful!
- It required extra domain knowledge. A different kind of knowledge than numerical probabilities. It needed an understanding of relationships among the random variables.

25

CW: Practice

Are Coin and Card independent in this distribution?

Coin	Card	Candy	P(Coin, Card, Candy)
tails	black	1	0.15
tails	black	2	0.06
tails	black	3	0.09
tails	red	1	0.02
tails	red	2	0.06
tails	red	3	0.12
heads	black	1	0.075
heads	black	2	0.03
heads	black	3	0.045
heads	red	1	0.035
heads	red	2	0.105
heads	red	3	0.21

Recall:

$$\boldsymbol{P}(X \mid Y) = \boldsymbol{P}(X)$$

$$\boldsymbol{P}(Y \mid X) = \boldsymbol{P}(Y)$$

$$\boldsymbol{P}(X, Y) = \boldsymbol{P}(X)\boldsymbol{P}(Y)$$

for independent X and Y

26