CS 331: Artificial Intelligence Fundamentals of Probability II

Thanks to Andrew Moore for some course material

1

Coin	Card	Candy	P(Coin, Card, Candy)	
tails	black	1	0.15	
tails	black	2	0.06	The probabilities
tails	black	3	0.09	in the last column
tails	red	1	0.02	sum to 1
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	•

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) =

 $\begin{array}{l} \mathsf{P}(\ Coin=heads,\ Card=black,\ Candy=1\) +\\ \mathsf{P}(\ Coin=heads,\ Card=black,\ Candy=2\) +\\ \mathsf{P}(\ Coin=heads,\ Card=black,\ Candy=3\) +\\ \mathsf{P}(\ Coin=heads,\ Card=red,\ Candy=1\) +\\ \mathsf{P}(\ Coin=heads,\ Card=red,\ Candy=3\) +\\ \mathsf{P}(\ Coin=heads,\ Card=red,\ Candy=2\) +\\ \mathsf{P}(\ Coin=heads,\ Card=red,\ Candy=2\) +\\ \mathsf{P}(\ Coin=heads,\ Card=red,\ Candy=2\) +\\ \mathsf{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\end{array}$

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) +

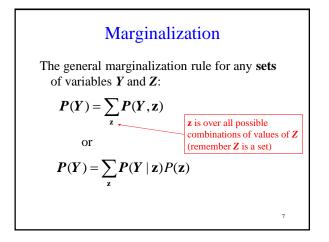
P(Coin=tails, Card=red, Candy=3)

$$= 0.02 + 0.06 + 0.12 = 0.2$$

Marginalization

Or even:

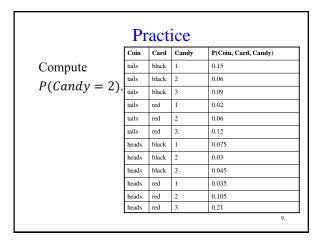
 $\begin{array}{l} \mathsf{P}(\ Card=black\)=\\ \mathsf{P}(\ Coin=heads,\ Card=black,\ Candy=1)+\\ \mathsf{P}(\ Coin=heads,\ Card=black,\ Candy=2\)+\\ \mathsf{P}(\ Coin=heads,\ Card=black,\ Candy=3\)+\\ \mathsf{P}(\ Coin=tails,\ Card=black,\ Candy=1)+\\ \mathsf{P}(\ Coin=tails,\ Card=black,\ Candy=2\)+\\ \mathsf{P}(\ Coin=tails,\ Card=black,\ Candy=3\)\\ =\ 0.075+0.03+0.045+0.015+0.06+0.09=0.315 \end{array}$



Marginalization

For continuous variables, marginalization involves taking the integral:

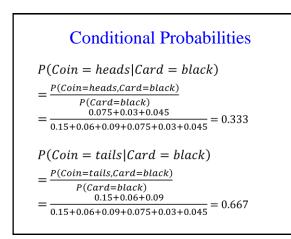
$$\boldsymbol{P}(\boldsymbol{Y}) = \int \boldsymbol{P}(\boldsymbol{Y}, \boldsymbol{z}) d\boldsymbol{z}$$

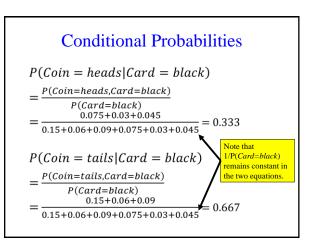


Conditional Probabilities

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

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





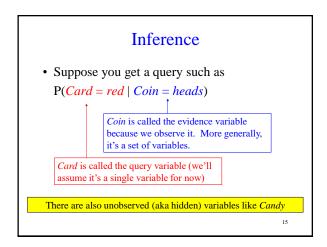
Normalization

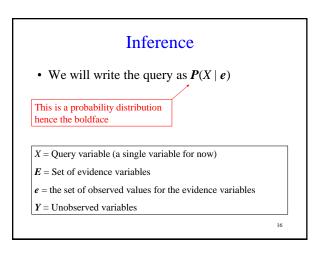
- In fact, 1/P(*Card*) can be viewed as a normalization constant for *P*(*Coin*| *Card*), ensuring it adds up to 1
- We will refer to normalization constants with the symbol α

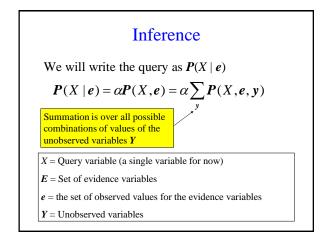
 $P(Coin|black) = \alpha P(Coin, black)$

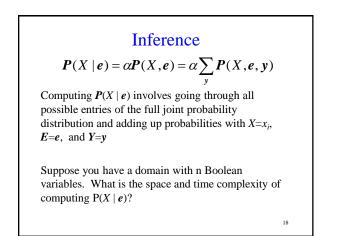
13

Р	racti	ce		
	Coin	Card	Candy	P(Coin, Card, Candy)
Compute	tails	black	1	0.15
P(Candy = 1 Card = red).	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









Independence

• How do you avoid the exponential space and time complexity of inference?

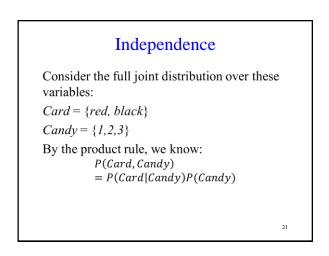
19

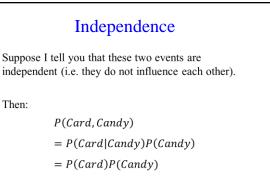
• Use independence (aka factoring)

Independence

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

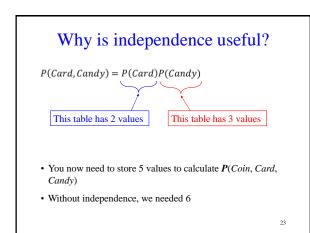
P(X | Y) = P(X) or P(Y | X) = P(Y) or P(X,Y) = P(X)P(Y)

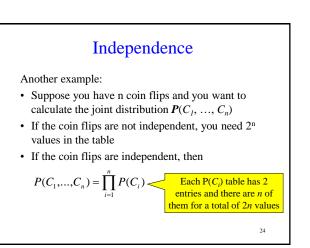




20

22





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

Practice Coin Card Candy P(Coin, Card, Candy) tails black 0.15 Are Coin and Card 1 tails black 0.06 independent in this 0.09 tails black distribution? tails red 0.02 0.06 tails red 2 tails red 0.12 3 Recall: heads black 1 0.075 $\boldsymbol{P}(X \mid Y) = \boldsymbol{P}(X)$ black 2 0.03 heads black 3 0.045 heads $\boldsymbol{P}(Y \mid X) = \boldsymbol{P}(Y)$ heads red 0.035 1 $\boldsymbol{P}(X,Y) = \boldsymbol{P}(X)\boldsymbol{P}(Y)^{\frac{\text{heads red}}{\text{heads red}}}$ 0.105 2 0.21 3 for independent X and Y 26