

Today – Part 1: Uncertainty

- ▶ **Only a degree of belief**
- ▶ Use **probability theory**
 - ▶ Assign to each sentence a numerical degree of belief.
 - ▶ Not degree of truth!
 - ▶ Summarizing the uncertainty that comes from laziness, ignorance.
- ▶ Random variables
- ▶ Joint and marginal distributions
- ▶ Conditional distributions
- ▶ Product rule, chain rule, Bayes' rule
- ▶ Inference
- ▶ Independence, conditional independence

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

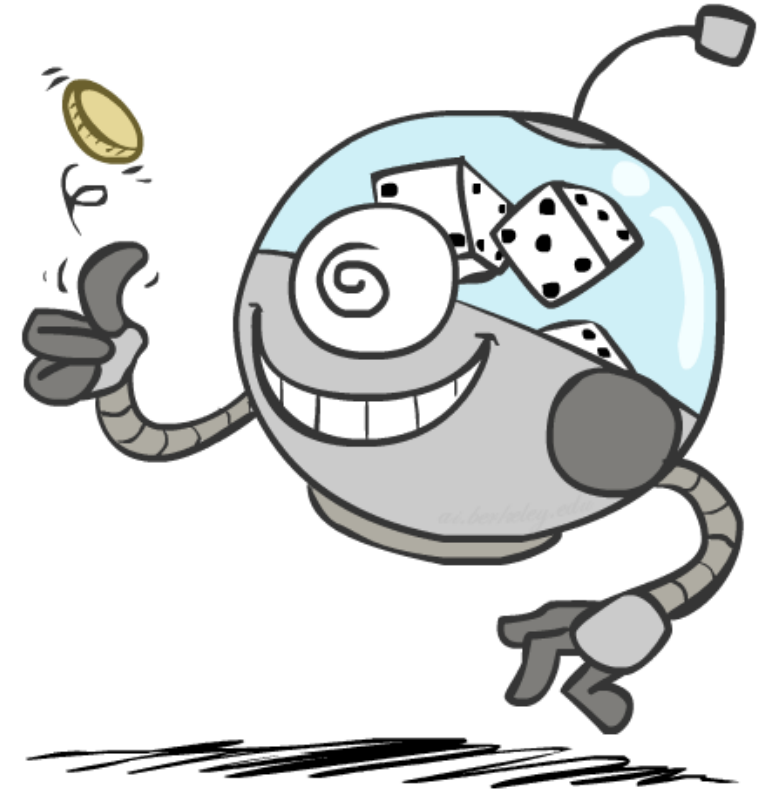
0.11	0.11	0.11
0.11	0.11	0.11
0.11	0.11	0.11

0.17	0.10	0.10
0.09	0.17	0.10
<0.01	0.09	0.17

<0.01	<0.01	0.03
<0.01	0.05	0.05
<0.01	0.05	0.81

Random Variables

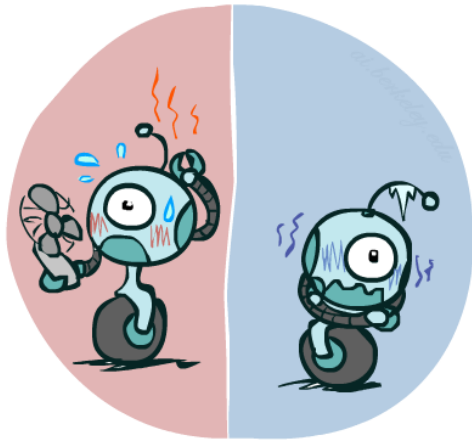
- A random variable is some aspect of the world about which we (may) have uncertainty
 - R = Is it raining?
 - T = Is it hot or cold?
 - D = How long will it take to drive to work?
 - L = Where is the ghost?
- We denote random variables with capital letters
- Like variables in a CSP, random variables have domains
 - R in $\{\text{true}, \text{false}\}$ (often write as $\{+r, -r\}$)
 - T in $\{\text{hot}, \text{cold}\}$
 - D in $[0, \infty)$
 - L in possible locations, maybe $\{(0,0), (0,1), \dots\}$



Probability Distributions

- Associate a probability with each value

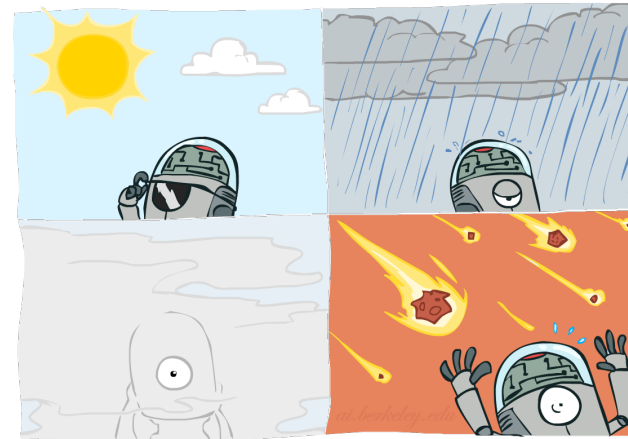
- Temperature:



$P(T)$

T	P
hot	0.5
cold	0.5

- Weather:



$P(W)$

W	P
sun	0.6
rain	0.1
fog	0.3
meteor	0.0

Probability Distributions

- Unobserved random variables have distributions

$P(T)$

T	P
hot	0.5
cold	0.5

$P(W)$

W	P
sun	0.6
rain	0.1
fog	0.3
meteor	0.0

Shorthand notation:

$$P(\text{hot}) = P(T = \text{hot}),$$

$$P(\text{cold}) = P(T = \text{cold}),$$

$$P(\text{rain}) = P(W = \text{rain}),$$

...

OK if all domain entries are unique

- A distribution is a TABLE of probabilities of values

- A probability (lower case value) is a single number

$$P(W = \text{rain}) = 0.1$$

- Must have: $\forall x \ P(X = x) \geq 0$

and

$$\sum_x P(X = x) = 1$$

Joint Distributions

- A *joint distribution* over a set of random variables: $X_1, X_2, \dots, X_n$ specifies a real number for each assignment (or *outcome*):

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

$$P(x_1, x_2, \dots, x_n)$$

$$P(x_1, x_2, \dots, x_n) \geq 0$$

$$\sum_{(x_1, x_2, \dots, x_n)} P(x_1, x_2, \dots, x_n) = 1$$

Size of distribution if n variables with domain sizes d ?

$$P(T, W)$$

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

- For all but the smallest distributions, impractical to write out!

Prior probability

Prior or unconditional probabilities of propositions

e.g., $P(\text{Cavity} = \text{true}) = 0.1$ and $P(\text{Weather} = \text{sunny}) = 0.72$
correspond to belief prior to arrival of any (new) evidence

Probability distribution gives values for all possible assignments:

$$\mathbf{P}(\text{Weather}) = \langle 0.72, 0.1, 0.08, 0.1 \rangle \text{ (normalized, i.e., sums to 1)}$$

Joint probability distribution for a set of r.v.s gives the probability of every atomic event on those r.v.s (i.e., every sample point)

$\mathbf{P}(\text{Weather}, \text{Cavity})$ = a 4×2 matrix of values:

<i>Weather</i> =	<i>sunny</i>	<i>rain</i>	<i>cloudy</i>	<i>snow</i>
<i>Cavity</i> = <i>true</i>	0.144	0.02	0.016	0.02
<i>Cavity</i> = <i>false</i>	0.576	0.08	0.064	0.08

Every question about a domain can be answered by the joint distribution because every event is a sum of sample points

Probabilistic Models

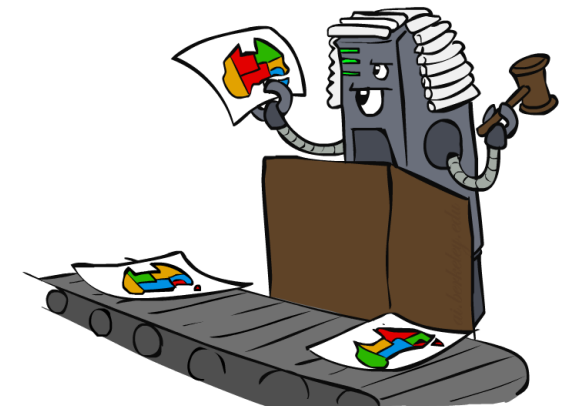
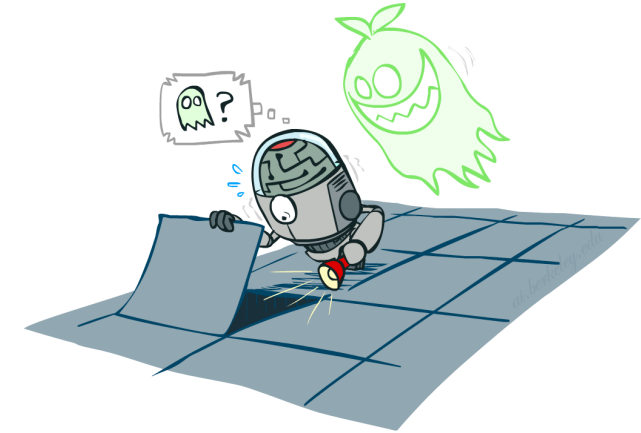
- A probabilistic model is a joint distribution over a set of random variables
- Probabilistic models:
 - (Random) variables with domains
 - Assignments are called *outcomes*
 - Joint distributions: say whether assignments (outcomes) are likely
 - *Normalized*: sum to 1.0
 - Ideally: only certain variables directly interact
- Constraint satisfaction problems:
 - Variables with domains
 - Constraints: state whether assignments are possible
 - Ideally: only certain variables directly interact

Distribution over T,W

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

Constraint over T,W

T	W	P
hot	sun	T
hot	rain	F
cold	sun	F
cold	rain	T



Events

- An *event* is a set E of outcomes

$$P(E) = \sum_{(x_1 \dots x_n) \in E} P(x_1 \dots x_n)$$

- From a joint distribution, we can calculate the probability of any event
 - Probability that it's hot AND sunny?
 - Probability that it's hot?
 - Probability that it's hot OR sunny?
- Typically, the events we care about are *partial assignments*, like $P(T=\text{hot})$

$P(T, W)$

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

Quiz: Events

■ $P(+x, +y)$?

■ $P(+x)$?

■ $P(-y \text{ OR } +x)$?

$P(X, Y)$

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

Marginal Distributions

- Marginal distributions are sub-tables which eliminate variables
- Marginalization (summing out): Combine collapsed rows by adding

$P(T, W)$

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



$$P(t) = \sum_s P(t, s)$$

$P(T)$

T	P
hot	0.5
cold	0.5

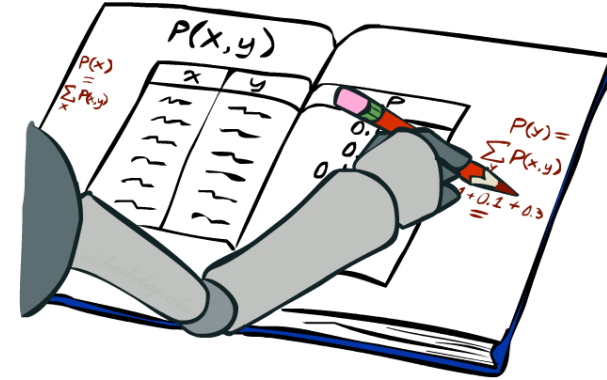


$$P(s) = \sum_t P(t, s)$$

$P(W)$

W	P
sun	0.6
rain	0.4

$$P(X_1 = x_1) = \sum_{x_2} P(X_1 = x_1, X_2 = x_2)$$



Quiz: Marginal Distributions

$P(X, Y)$

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1



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



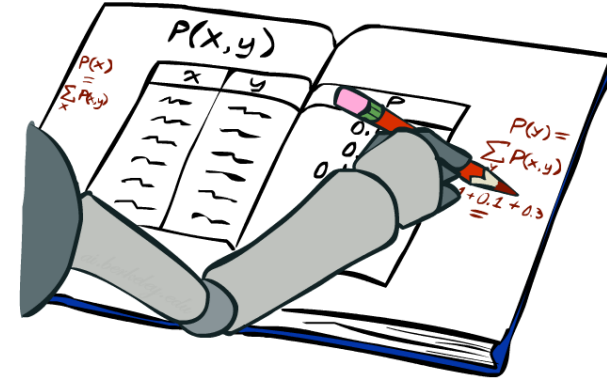
$$P(y) = \sum_x P(x, y)$$

$P(X)$

X	P
+x	
-x	

$P(Y)$

Y	P
+y	
-y	



Conditional probability

Conditional or posterior probabilities

e.g., $P(\text{cavity}|\text{toothache}) = 0.8$

i.e., **given that toothache is all I know**

NOT “if *toothache* then 80% chance of *cavity*”

(Notation for conditional distributions:

$\mathbf{P}(\text{Cavity}|\text{Toothache}) = 2\text{-element vector of 2-element vectors})$

If we know more, e.g., *cavity* is also given, then we have

$P(\text{cavity}|\text{toothache}, \text{cavity}) = 1$

Note: the less specific belief **remains valid** after more evidence arrives, but is not always **useful**

New evidence may be irrelevant, allowing simplification, e.g.,

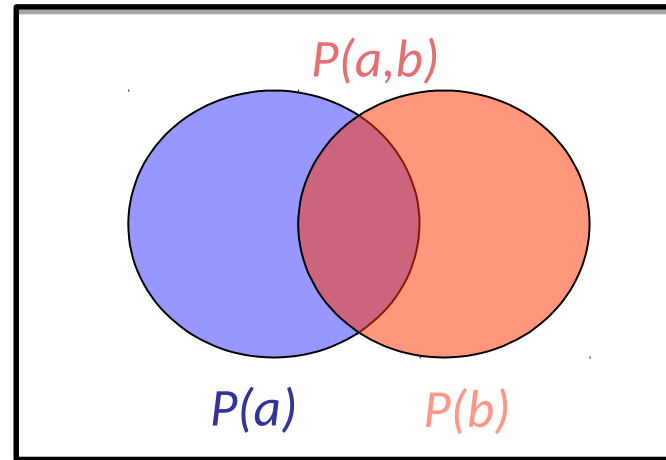
$P(\text{cavity}|\text{toothache}, \text{49ersWin}) = P(\text{cavity}|\text{toothache}) = 0.8$

This kind of inference, sanctioned by domain knowledge, is crucial

Conditional Probabilities

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

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



$P(T, W)$

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

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

$$\begin{aligned} &= P(W = s, T = c) + P(W = r, T = c) \\ &= 0.2 + 0.3 = 0.5 \end{aligned}$$

Quiz: Conditional Probabilities

■ $P(+x \mid +y) ?$

$P(X, Y)$

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

■ $P(-x \mid +y) ?$

■ $P(-y \mid +x) ?$

Conditional Distributions

- Conditional distributions are probability distributions over some variables given fixed values of others

Conditional Distributions

$P(W T)$	$P(W T = \text{hot})$	
	W	P
	sun	0.8
	rain	0.2
	$P(W T = \text{cold})$	
	W	P
	sun	0.4
	rain	0.6

Joint Distribution

$P(T, W)$		
T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

Normalization Trick

$$P(T, W)$$

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



$$P(W|T = c)$$

W	P
sun	
rain	

Normalization Trick

$P(T, W)$

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

$$\begin{aligned}P(W = s|T = c) &= \frac{P(W = s, T = c)}{P(T = c)} \\&= \frac{P(W = s, T = c)}{P(W = s, T = c) + P(W = r, T = c)} \\&= \frac{0.2}{0.2 + 0.3} = 0.4\end{aligned}$$



$P(W|T = c)$

W	P
sun	0.4
rain	0.6

$$\begin{aligned}P(W = r|T = c) &= \frac{P(W = r, T = c)}{P(T = c)} \\&= \frac{P(W = r, T = c)}{P(W = s, T = c) + P(W = r, T = c)} \\&= \frac{0.3}{0.2 + 0.3} = 0.6\end{aligned}$$

Normalization Trick

$$\begin{aligned}P(W = s|T = c) &= \frac{P(W = s, T = c)}{P(T = c)} \\&= \frac{P(W = s, T = c)}{P(W = s, T = c) + P(W = r, T = c)} \\&= \frac{0.2}{0.2 + 0.3} = 0.4\end{aligned}$$

$P(T, W)$

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

SELECT the joint probabilities matching the evidence



$P(c, W)$

T	W	P
cold	sun	0.2
cold	rain	0.3

NORMALIZE the selection (make it sum to one)



$P(W|T = c)$

W	P
sun	0.4
rain	0.6

$$\begin{aligned}P(W = r|T = c) &= \frac{P(W = r, T = c)}{P(T = c)} \\&= \frac{P(W = r, T = c)}{P(W = s, T = c) + P(W = r, T = c)} \\&= \frac{0.3}{0.2 + 0.3} = 0.6\end{aligned}$$

Quiz: Normalization Trick

■ $P(X \mid Y=-y)$?

$P(X, Y)$

X	Y	P
+x	+y	0.2
+x	-y	0.3
-x	+y	0.4
-x	-y	0.1

SELECT the joint
probabilities
matching the
evidence



NORMALIZE the
selection
(make it sum to one)



Probabilistic Inference

- Probabilistic inference: compute a desired probability from other known probabilities (e.g. conditional from joint)
- We generally compute conditional probabilities
 - $P(\text{on time} \mid \text{no reported accidents}) = 0.90$
 - These represent the agent's *beliefs* given the evidence
- Probabilities change with new evidence:
 - $P(\text{on time} \mid \text{no accidents, 5 a.m.}) = 0.95$
 - $P(\text{on time} \mid \text{no accidents, 5 a.m., raining}) = 0.80$
 - Observing new evidence causes *beliefs to be updated*



Inference by enumeration

Start with the joint distribution:

	<i>toothache</i>		$\neg$ <i>toothache</i>	
	<i>catch</i>	$\neg$ <i>catch</i>	<i>catch</i>	$\neg$ <i>catch</i>
<i>cavity</i>	.108	.012	.072	.008
$\neg$ <i>cavity</i>	.016	.064	.144	.576

For any proposition ϕ , sum the atomic events where it is true:

$$P(\phi) = \sum_{\omega: \omega \models \phi} P(\omega)$$

Inference by enumeration

Start with the joint distribution:

	<i>toothache</i>		$\neg$ <i>toothache</i>	
	<i>catch</i>	$\neg$ <i>catch</i>	<i>catch</i>	$\neg$ <i>catch</i>
<i>cavity</i>	.108	.012	.072	.008
$\neg$ <i>cavity</i>	.016	.064	.144	.576

For any proposition ϕ , sum the atomic events where it is true:

$$P(\phi) = \sum_{\omega: \omega \models \phi} P(\omega)$$

$$P(\text{toothache}) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$$

Inference by enumeration

Start with the joint distribution:

	<i>toothache</i>		$\neg$ <i>toothache</i>	
	<i>catch</i>	$\neg$ <i>catch</i>	<i>catch</i>	$\neg$ <i>catch</i>
<i>cavity</i>	.108	.012	.072	.008
$\neg$ <i>cavity</i>	.016	.064	.144	.576

For any proposition ϕ , sum the atomic events where it is true:

$$P(\phi) = \sum_{\omega: \omega \models \phi} P(\omega)$$

$$P(\text{cavity} \vee \text{toothache}) = 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28$$

Inference by enumeration

Start with the joint distribution:

	<i>toothache</i>		$\neg$ <i>toothache</i>	
	<i>catch</i>	$\neg$ <i>catch</i>	<i>catch</i>	$\neg$ <i>catch</i>
<i>cavity</i>	.108	.012	.072	.008
$\neg$ <i>cavity</i>	.016	.064	.144	.576

Can also compute conditional probabilities:

$$\begin{aligned}
 P(\neg \text{cavity} | \text{toothache}) &= \frac{P(\neg \text{cavity} \wedge \text{toothache})}{P(\text{toothache})} \\
 &= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4
 \end{aligned}$$

Normalization

	<i>toothache</i>		$\neg$ <i>toothache</i>	
	<i>catch</i>	$\neg$ <i>catch</i>	<i>catch</i>	$\neg$ <i>catch</i>
<i>cavity</i>	.108	.012	.072	.008
$\neg$ <i>cavity</i>	.016	.064	.144	.576

Denominator can be viewed as a **normalization constant** α

$$\begin{aligned}
 \mathbf{P}(Cavity|toothache) &= \alpha \mathbf{P}(Cavity, toothache) \\
 &= \alpha [\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)] \\
 &= \alpha [\langle 0.108, 0.016 \rangle + \langle 0.012, 0.064 \rangle] \\
 &= \alpha \langle 0.12, 0.08 \rangle = \langle 0.6, 0.4 \rangle
 \end{aligned}$$

General idea: compute distribution on query variable
by fixing **evidence variables** and summing over **hidden variables**

Inference by enumeration, contd.

Let $\mathbf{X}$ be all the variables. Typically, we want the posterior joint distribution of the query variables $\mathbf{Y}$ given specific values $\mathbf{e}$ for the evidence variables $\mathbf{E}$

Let the hidden variables be $\mathbf{H} = \mathbf{X} - \mathbf{Y} - \mathbf{E}$

Then the required summation of joint entries is done by summing out the hidden variables:

$$\mathbf{P}(\mathbf{Y}|\mathbf{E} = \mathbf{e}) = \alpha \mathbf{P}(\mathbf{Y}, \mathbf{E} = \mathbf{e}) = \alpha \sum_{\mathbf{h}} \mathbf{P}(\mathbf{Y}, \mathbf{E} = \mathbf{e}, \mathbf{H} = \mathbf{h})$$

The terms in the summation are joint entries because $\mathbf{Y}$, $\mathbf{E}$, and $\mathbf{H}$ together exhaust the set of random variables

Obvious problems:

- 1) Worst-case time complexity $O(d^n)$ where d is the largest arity
- 2) Space complexity $O(d^n)$ to store the joint distribution
- 3) How to find the numbers for $O(d^n)$ entries???

Inference by Enumeration

- $P(W)$?
- $P(W \mid \text{winter})$?
- $P(W \mid \text{winter, hot})$?

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

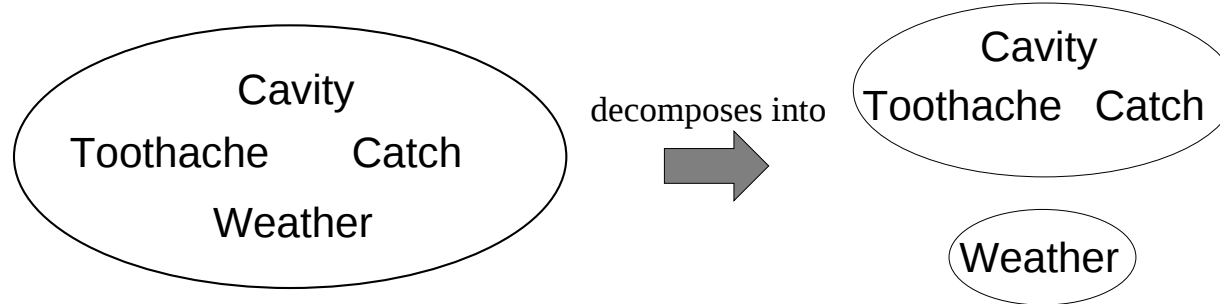
Inference by Enumeration

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

Independence

A and B are independent iff

$$\mathbf{P}(A|B) = \mathbf{P}(A) \quad \text{or} \quad \mathbf{P}(B|A) = \mathbf{P}(B) \quad \text{or} \quad \mathbf{P}(A, B) = \mathbf{P}(A)\mathbf{P}(B)$$



$$\begin{aligned} &\mathbf{P}(\textit{Toothache}, \textit{Catch}, \textit{Cavity}, \textit{Weather}) \\ &= \mathbf{P}(\textit{Toothache}, \textit{Catch}, \textit{Cavity})\mathbf{P}(\textit{Weather}) \end{aligned}$$

32 entries reduced to 12; for n independent biased coins, $2^n \rightarrow n$

Absolute independence powerful but rare

Dentistry is a large field with hundreds of variables,
none of which are independent. What to do?

Conditional independence

$\mathbf{P}(\textit{Toothache}, \textit{Cavity}, \textit{Catch})$ has $2^3 - 1 = 7$ independent entries

If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:

$$(1) \ P(\textit{catch}|\textit{toothache}, \textit{cavity}) = P(\textit{catch}|\textit{cavity})$$

The same independence holds if I haven't got a cavity:

$$(2) \ P(\textit{catch}|\textit{toothache}, \neg\textit{cavity}) = P(\textit{catch}|\neg\textit{cavity})$$

Catch is **conditionally independent** of *Toothache* given *Cavity*:

$$\mathbf{P}(\textit{Catch}|\textit{Toothache}, \textit{Cavity}) = \mathbf{P}(\textit{Catch}|\textit{Cavity})$$

Equivalent statements:

$$\mathbf{P}(\textit{Toothache}|\textit{Catch}, \textit{Cavity}) = \mathbf{P}(\textit{Toothache}|\textit{Cavity})$$

$$\mathbf{P}(\textit{Toothache}, \textit{Catch}|\textit{Cavity}) = \mathbf{P}(\textit{Toothache}|\textit{Cavity})\mathbf{P}(\textit{Catch}|\textit{Cavity})$$

Conditional independence contd.

Write out full joint distribution using chain rule:

$$\begin{aligned} & \mathbf{P}(Toothache, Catch, Cavity) \\ &= \mathbf{P}(Toothache|Catch, Cavity)\mathbf{P}(Catch, Cavity) \\ &= \mathbf{P}(Toothache|Catch, Cavity)\mathbf{P}(Catch|Cavity)\mathbf{P}(Cavity) \\ &= \mathbf{P}(Toothache|Cavity)\mathbf{P}(Catch|Cavity)\mathbf{P}(Cavity) \end{aligned}$$

I.e., $2 + 2 + 1 = 5$ independent numbers (equations 1 and 2 remove 2)

In most cases, the use of conditional independence reduces the size of the representation of the joint distribution from exponential in n to linear in n .

Conditional independence is our most basic and robust form of knowledge about uncertain environments.

Bayes' Rule

Product rule $P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$

$$\Rightarrow \text{Bayes' rule } P(a|b) = \frac{P(b|a)P(a)}{P(b)}$$

or in distribution form

$$\mathbf{P}(Y|X) = \frac{\mathbf{P}(X|Y)\mathbf{P}(Y)}{\mathbf{P}(X)} = \alpha \mathbf{P}(X|Y)\mathbf{P}(Y)$$

Useful for assessing **diagnostic** probability from **causal** probability:

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

E.g., let M be meningitis, S be stiff neck:

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

Note: posterior probability of meningitis still very small!

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

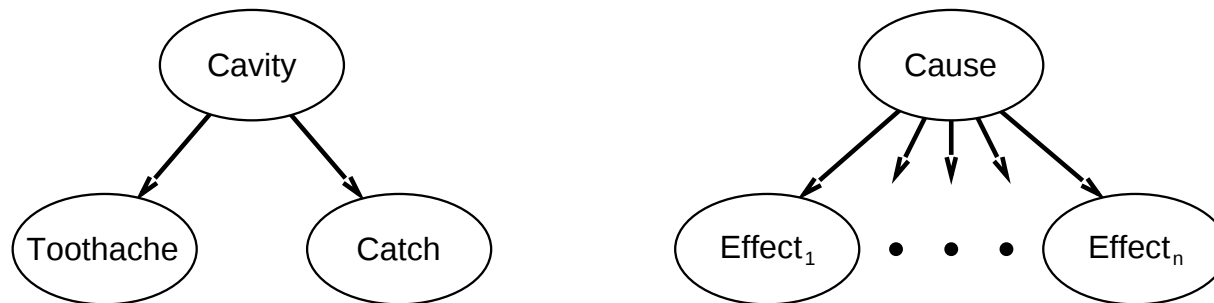
- What is $P(W \mid \text{dry})$?

Bayes' Rule and conditional independence

$$\begin{aligned} & \mathbf{P}(Cavity|toothache \wedge catch) \\ &= \alpha \mathbf{P}(toothache \wedge catch|Cavity)\mathbf{P}(Cavity) \\ &= \alpha \mathbf{P}(toothache|Cavity)\mathbf{P}(catch|Cavity)\mathbf{P}(Cavity) \end{aligned}$$

This is an example of a **naive Bayes** model:

$$\mathbf{P}(Cause, Effect_1, \dots, Effect_n) = \mathbf{P}(Cause) \prod_i \mathbf{P}(Effect_i|Cause)$$



Total number of parameters is **linear** in n