Bayesian Inference

Chapter 13

Review

Making rational decisions when faced with uncertainty:

- *Probability*

  the precise representation of knowledge and uncertainty

- *Probability theory*

  how to optimally update your knowledge based on new information

- *Decision theory: probability theory + utility theory*

  how to use this information to achieve maximum expected utility

Marginalization

Weather =	sunny	rain	cloudy	snow
Cavity = T	0.144	0.02	0.016	0.02
Cavity = F	0.576	0.08	0.064	0.08

Weather =	sunny	rain	cloudy	snow
	0.72	0.1	0.08	0.1

$$P(W) = \sum_{i=1}^{2} P(C = c_i, W)$$

$$P(W = \text{sunny})$$
$$= P(\text{cavity=t,W=sunny}) + P(\text{cavity=f,W=sunny})$$

Marginalization

Weather =	sunny	rain	cloudy	snow
Cavity = T	0.144	0.02	0.016	0.02
Cavity = F	0.576	0.08	0.064	0.08

Cavity = T	0.2
Cavity = F	0.8

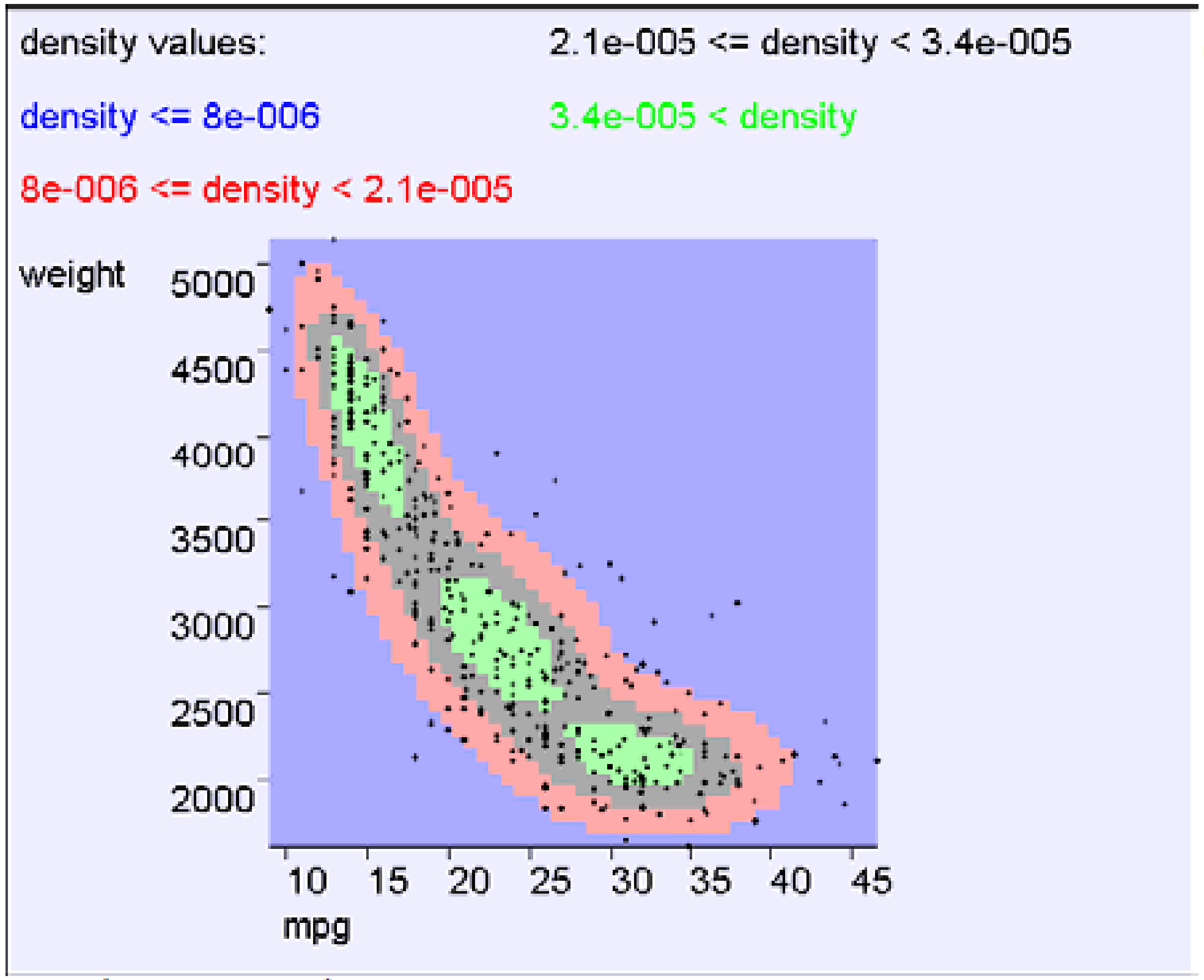
$$P(C) = \sum_{i=1}^{4} P(C, W = w_i)$$

Continuous Densities

In 2 dimensions

Let *X,Y* be a pair of continuous random variables, and let R be some region of (X,Y) space…

$$P((X,Y)\in R)=\iint_{(x,y)\in R} p(x,y)\,dy\,dx$$

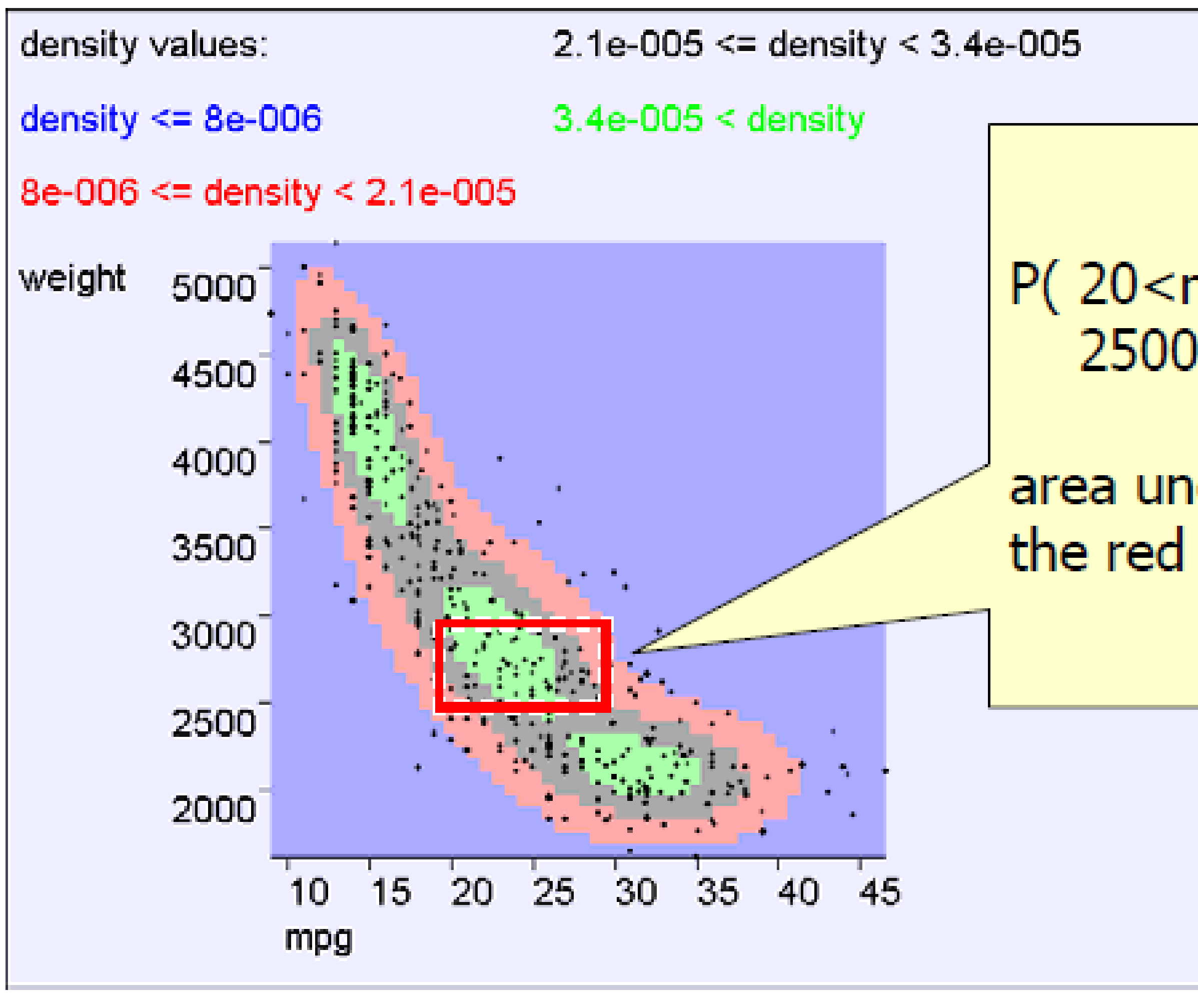


Probability of Joint Event

In 2 dimensions

Let X, Y be a pair of continuous random variables, and let R be some region of (X,Y) space…

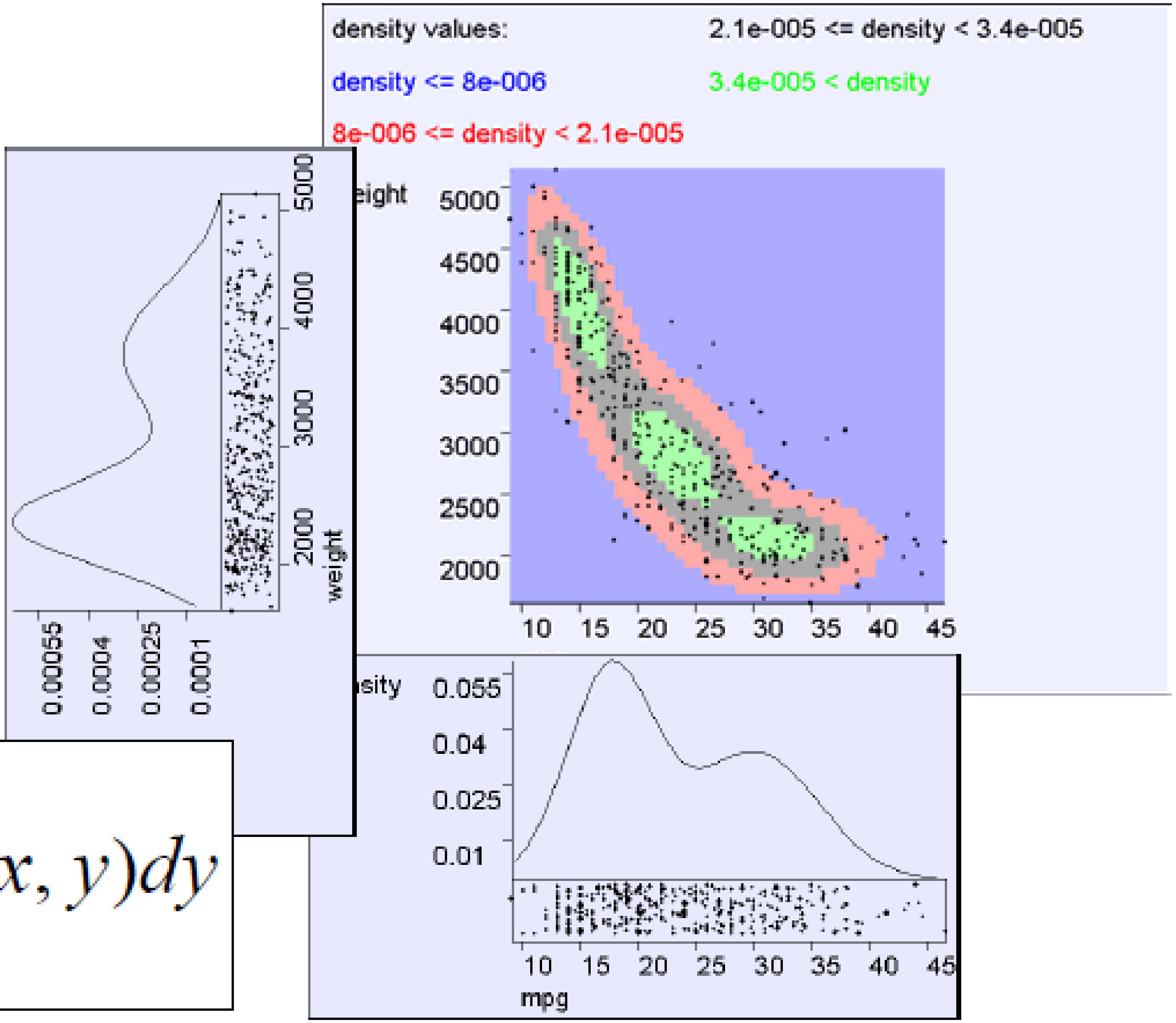
$$P((X,Y) \in R) = \iint_{(x,y)\in R} p(x,y)\,dy\,dx$$



P(20<mpg<30 and 2500<weight<3000) =

area under the 2-d surface within the red rectangle

Marginal Distributions



$$p(x) = \int_{y=-\infty}^{\infty} p(x,y)\,dy$$

Inference by Enumeration

Start with the joint distribution:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	.108	.012	.072	.008
$\neg$ *cavity*	.016	.064	.144	.576

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

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

Inference by Enumeration

Start with the joint distribution:

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	.108	.012	.072	.008
¬ *cavity*	.016	.064	.144	.576

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

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

$$P(toothache) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$$

Inference by Enumeration

Start with the joint distribution:

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ *cavity*	**.016**	**.064**	**.144**	**.576**

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

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

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

Inference by Enumeration

Start with the joint distribution:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	.108	.012	.072	.008
$\neg$ *cavity*	.016	.064	.144	.576

Can also compute conditional probabilities:

$$P(\neg cavity | toothache) = \frac{P(\neg cavity \land toothache)}{P(toothache)}$$

$$= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

Normalization

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ cavity	**.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

Independence

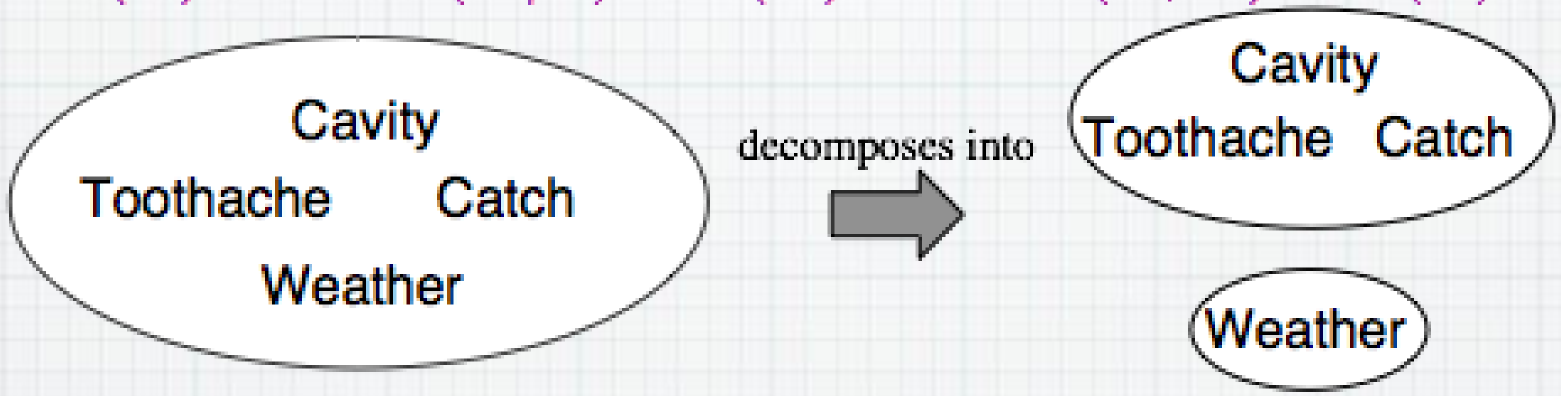
A and B are independent iff

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

Independence

A and B are independent iff

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

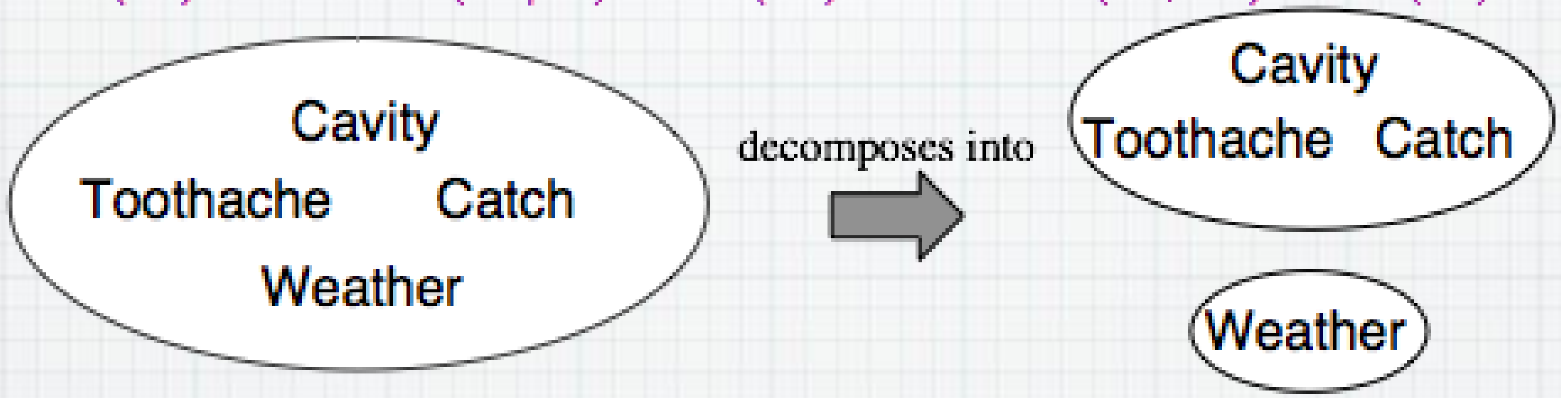


$$\mathbf{P}(Toothache, Catch, Cavity, Weather)$$
$$= \mathbf{P}(Toothache, Catch, Cavity)\mathbf{P}(Weather)$$

Independence

A and B are independent iff

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



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

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

Absolute independence powerful but rare

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

Inference Using Bayes Rule

Simple example: medical test results

- Test report for rare disease is positive, 90% accurate
- What's the probability that you have the disease?
- What if the test is repeated?

- This is the simplest example of reasoning by combining sources of information.

How do we model the problem?

- Which is the correct description of “Test is 90% accurate” ?

$$
\begin{aligned}
P(T = \text{true}) &= 0.9 \\
\boxed{P(T = \text{true} | D = \text{true}) = 0.9} \\
P(D = \text{true} | T = \text{true}) &= 0.9
\end{aligned}
$$

- What do we want to know?

$$
\begin{gathered}
P(T = \text{true}) \\
P(T = \text{true} | D = \text{true}) \\
\boxed{P(D = \text{true} | T = \text{true})}
\end{gathered}
$$

- More compact notation:

$$
\begin{aligned}
P(T = \text{true} | D = \text{true}) &\rightarrow P(T|D) \\
P(T = \text{false} | D = \text{false}) &\rightarrow P(\bar{T}|\bar{D})
\end{aligned}
$$

Evaluating the posterior probability through Bayesian inference

- We want P(D|T) = "The probability of the having the disease given a positive test"
- Use Bayes rule to relate it to what we know: P(T|D)

$$\underbrace{P(D|T)}_{\textit{posterior}} = \frac{\overbrace{P(T|D)}^{\textit{likelihood}}\overbrace{P(D)}^{\textit{prior}}}{\underbrace{P(T)}_{\textit{normalizing constant}}}$$

- What's the prior P(D)?
- Disease is rare, so let's assume

$$P(D) = 0.001$$

- What about P(T)? $P(T) = P(T, D) + P(T, \bar{D})$
- What's the interpretation of that?

Evaluating the normalizing constant

$$\text{posterior} \quad P(D|T) = \frac{\overset{\text{likelihood}}{P(T|D)}\,\overset{\text{prior}}{P(D)}}{\underset{\text{normalizing constant}}{P(T)}}$$

- P(T) is the marginal probability of P(T,D) = P(T|D) P(D)
- So, compute with summation

$$P(T) = \sum_{\text{all values of D}} P(T|D)P(D)$$

- For true or false propositions:

$$P(T) = P(T|D)P(D) + P(T|\bar{D})P(\bar{D})$$

- Our complete expression is

$$P(D|T) = \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|\bar{D})P(\bar{D})}$$

- Plugging in the numbers we get:

$$P(D|T) = \frac{0.9 \times 0.001}{0.9 \times 0.001 + 0.1 \times 0.999} = 0.0089$$

$$P(D) = 0.001$$

- Does this make intuitive sense?

After positive test, 9 times more likely to have disease

Playing around with the numbers

$$P(D|T) \quad = \quad \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|\bar{D})P(\bar{D})}$$

- What if the test were 100% reliable?

$$P(D|T) = \frac{1.0 \times 0.001}{1.0 \times 0.001 + 0.0 \times 0.999} = 1.0$$

- What if the test was the same, but disease wasn't so rare?

$$P(D|T) = \frac{0.9 \times 0.1}{0.9 \times 0.1 + 0.1 \times 0.999} = 0.5$$

Repeating the test

- We can relax, P(D|T) = 0.0089, right?
- Just to be sure the doctor recommends repeating the test.
- How do we represent this?

$$P(D|T_1, T_2)$$

- Again, we apply Bayes' rule

$$P(D|T_1, T_2) = \frac{P(T_1, T_2|D)P(D)}{P(T_1, T_2)}$$

- How do we model $P(T_1,T_2|D)$?

Modeling repeated tests

$$P(D|T_1,T_2) = \frac{P(T_1,T_2|D)P(D)}{P(T_1,T_2)}$$

- Easiest is to assume the tests are *independent*.

$$P(T_1,T_2|D) = P(T1|D)P(T_2|D)$$

- This also implies:

$$P(T_1,T_2) = P(T1)P(T_2)$$

- Plugging these in, we have

$$P(D|T_1,T_2) = \frac{P(T_1|D)P(T_2|D)P(D)}{P(T_1)P(T_2)}$$

Evaluating the normalizing constant again

- Expanding as before we have

$$P(D|T_1, T_2) = \frac{P(T_1|D)P(T_2|D)P(D)}{\sum_{D=\{t,f\}} P(T_1|D)P(T_2|D)P(D)}$$

- Plugging in the numbers gives us

$$P(D|T) = \frac{0.9 \times 0.9 \times 0.001}{0.9 \times 0.9 \times 0.001 + 0.1 \times 0.1 \times 0.999} = 0.075$$

- Another way to think about this:
  - What's the chance of 1 false positive from the test?
  - What's the chance of 2 false positives?

- The chance of 2 false positives is still 10x more likely than the a prior probability of having the disease.

Simpler: Combining information the Bayesian way

- Let's look at the equation again:

$$P(D|T_1, T_2) = \frac{P(T_1|D)P(T_2|D)P(D)}{P(T_1)P(T_2)}$$

- If we rearrange slightly:

$$P(D|T_1, T_2) = \frac{P(T_2|D)P(T_1|D)P(D)}{P(T_2)P(T_1)}$$

We've seen this before!

- It's the posterior for the first test, which we just computed

$$P(D|T_1) = \frac{P(T_1|D)P(D)}{P(T_1)}$$

The old posterior is the new prior

- We can just plugin the value of the old posterior
- It plays exactly the same role as our old prior

$$P(D|T_1, T_2) = \frac{P(T_2|D)P(T_1|D)P(D)}{P(T_2)P(T_1)}$$

$$P(D|T_1, T2) = \frac{P(T_2|D) \times 0.0089}{P(T_2)}$$

This is how Bayesian reasoning combines old information with new information to update our belief states.

- Plugging in the numbers gives the same answer:

$$P(D|T) \quad = \quad \frac{P(T|D)P'(D)}{P(T|D)P'(D) + P(T|\bar{D})P'(\bar{D})}$$

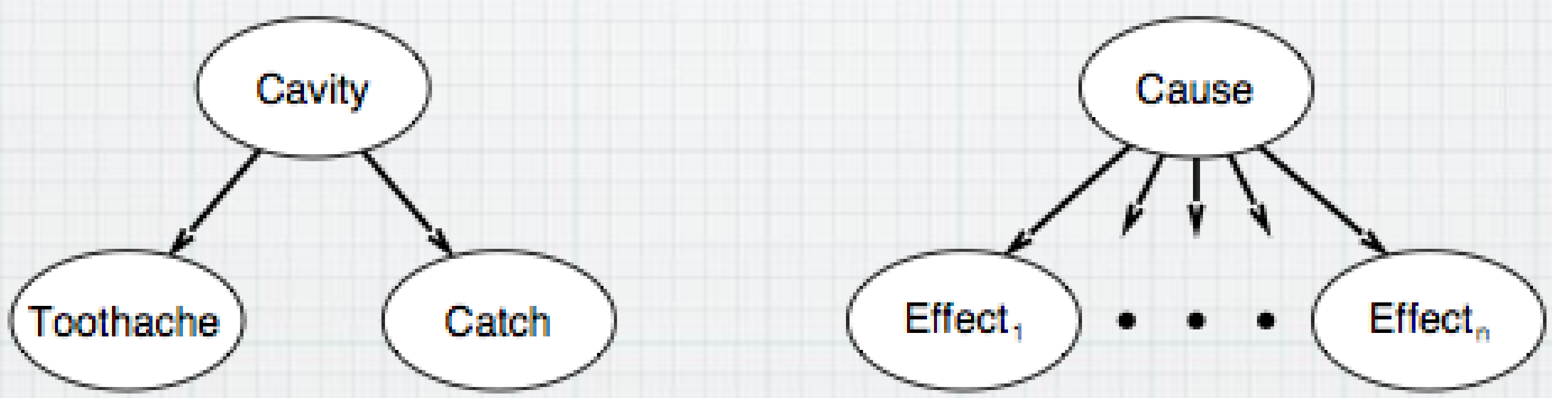
$$P(D|T) = \frac{0.9 \times 0.0089}{0.9 \times 0.0089 + 0.1 \times 0.9911} = 0.075$$

Bayes' Rule and Conditional Independence

$$\mathbf{P}(Cavity|toothache \wedge catch)$$
$$= \alpha\, \mathbf{P}(toothache \wedge catch|Cavity)\mathbf{P}(Cavity)$$
$$= \alpha\, \mathbf{P}(toothache|Cavity)\mathbf{P}(catch|Cavity)\mathbf{P}(Cavity)$$

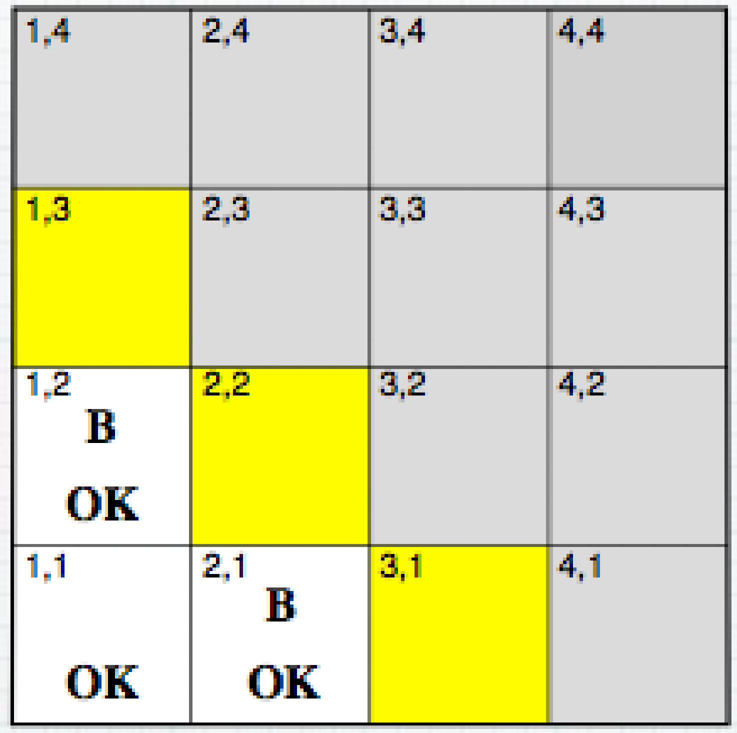
This is an example of a naive Bayes model:

$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause)\Pi_i \mathbf{P}(Effect_i|Cause)$$



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

Wumpus Example



$P_{ij} = true$ iff $[i, j]$ contains a pit

$B_{ij} = true$ iff $[i, j]$ is breezy
Include only $B_{1,1}, B_{1,2}, B_{2,1}$ in the probability model

Wumpus Probability Model

The full joint distribution is $\mathbf{P}(P_{1,1}, \ldots, P_{4,4}, B_{1,1}, B_{1,2}, B_{2,1})$

Apply product rule: $\mathbf{P}(B_{1,1}, B_{1,2}, B_{2,1} \,|\, P_{1,1}, \ldots, P_{4,4})\mathbf{P}(P_{1,1}, \ldots, P_{4,4})$

(Do it this way to get $P(Effect|Cause)$.)

First term: 1 if pits are adjacent to breezes, 0 otherwise

Second term: pits are placed randomly, probability 0.2 per square:

$$\mathbf{P}(P_{1,1}, \ldots, P_{4,4}) = \Pi_{i,j=1,1}^{4,4}\mathbf{P}(P_{i,j}) = 0.2^n \times 0.8^{16-n}$$

for n pits.

Observations and Query

We know the following facts:

$b = \neg b_{1,1} \land b_{1,2} \land b_{2,1}$
$known = \neg p_{1,1} \land \neg p_{1,2} \land \neg p_{2,1}$

Query is $\mathbf{P}(P_{1,3}|known, b)$

Define $Unknown = P_{ij}$s other than $P_{1,3}$ and $Known$

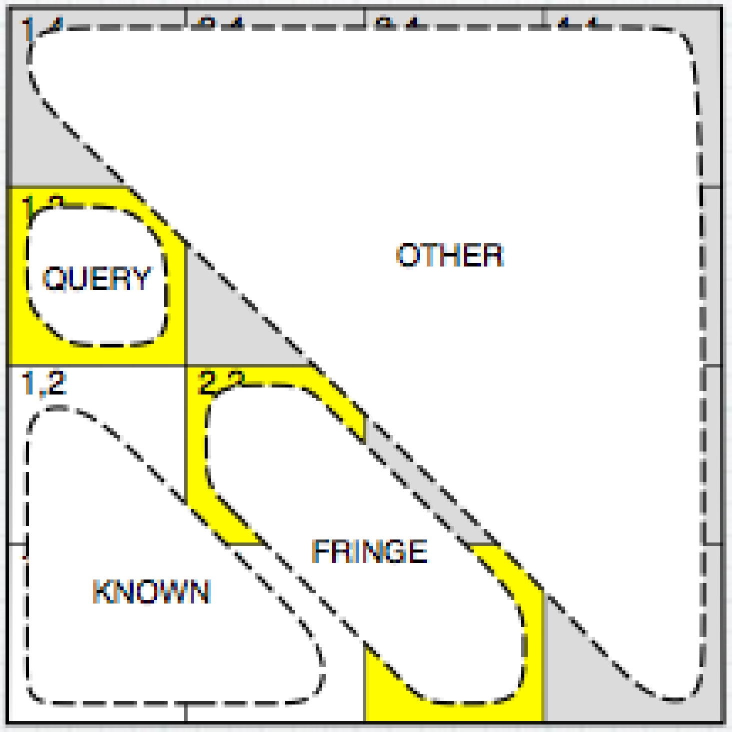
For inference by enumeration, we have

$$\mathbf{P}(P_{1,3}|known, b) = \alpha \Sigma_{unknown} \mathbf{P}(P_{1,3}, unknown, known, b)$$

Grows exponentially with number of squares!

Using Conditional Independence

Basic insight: observations are conditionally independent of other hidden squares given neighbouring hidden squares



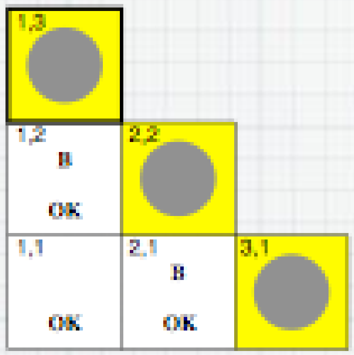
Define $Unknown = Fringe \cup Other$
$\mathbf{P}(b|P_{1,3}, Known, Unknown) = \mathbf{P}(b|P_{1,3}, Known, Fringe)$

Manipulate query into a form where we can use this!

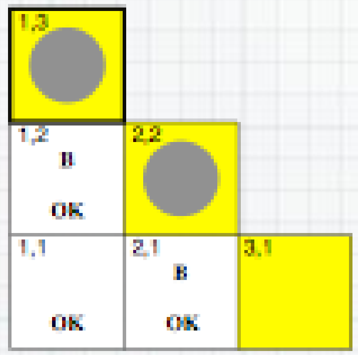
Using Conditional Independence

$$
\begin{aligned}
\mathbf{P}(P_{1,3}|known, b) &= \alpha \sum_{unknown} \mathbf{P}(P_{1,3}, unknown, known, b) \\
&= \alpha \sum_{unknown} \mathbf{P}(b|P_{1,3}, known, unknown)\mathbf{P}(P_{1,3}, known, unknown) \\
&= \alpha \sum_{fringe} \sum_{other} \mathbf{P}(b|known, P_{1,3}, fringe, other)\mathbf{P}(P_{1,3}, known, fringe, other) \\
&= \alpha \sum_{fringe} \sum_{other} \mathbf{P}(b|known, P_{1,3}, fringe)\mathbf{P}(P_{1,3}, known, fringe, other) \\
&= \alpha \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe) \sum_{other} \mathbf{P}(P_{1,3}, known, fringe, other) \\
&= \alpha \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe) \sum_{other} \mathbf{P}(P_{1,3})P(known)P(fringe)P(other) \\
&= \alpha\, P(known)\mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe)P(fringe) \sum_{other} P(other) \\
&= \alpha'\, \mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe)P(fringe)
\end{aligned}
$$

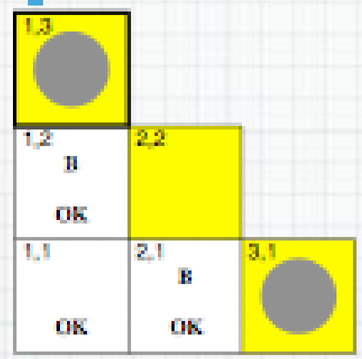
Using Conditional Independence



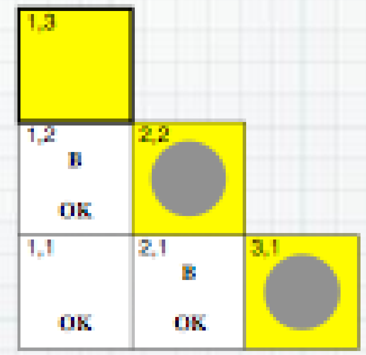
0.2 x 0.2 = 0.04



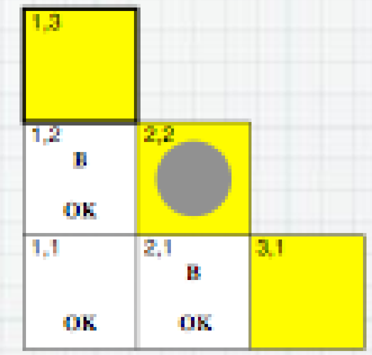
0.2 x 0.8 = 0.16



0.8 x 0.2 = 0.16



0.2 x 0.2 = 0.04



0.2 x 0.8 = 0.16

$$\begin{aligned}\mathbf{P}(P_{1,3}|known, b) &= \alpha' \langle 0.2(0.04+0.16+0.16),\ 0.8(0.04+0.16)\rangle \\ &\approx \langle 0.31, 0.69\rangle\end{aligned}$$

$$\mathbf{P}(P_{2,2}|known, b) \approx \langle 0.86, 0.14\rangle$$

Summary

Probability is a rigorous formalism for uncertain knowledge

Joint probability distribution specifies probability of every atomic event

Queries can be answered by summing over atomic events

For nontrivial domains, we must find a way to reduce the joint size

Independence and conditional independence provide the tools