Last update: November 29, 2012

Quantifying Uncertainty

CMSC 421: Chapter 13

Motivation

◊ Let action A_t = leave for airport t minutes before flight

- Will A_t get me there on time?

◊ Problems:

1) partial observability (road state, other drivers' plans, etc.)

2) noisy sensors (radio traffic reports)

3) uncertainty in action outcomes (flat tire, etc.)

4) immense complexity of modelling and predicting traffic

◊ Hence a purely logical approach either

1) risks falsehood: "A_{25} will get me there on time", or

2) leads to conclusions that are too weak for decision making:

◊ "A_{25} will get me there on time if there's no accident on the bridge,
and it doesn't rain,
and I don't get a flat tire,
and . . ."

Methods for handling uncertainty

◇ *Default* or *nonmonotonic* logic:

Assume my car does not have a flat tire

Assume A_{25} works unless contradicted by evidence

- What assumptions are reasonable? How to handle contradiction?

◇ *Rules with fudge factors*:

$A_{25} \mapsto_{0.3} AtAirportOnTime$

$Sprinkler \mapsto_{0.99} WetGrass$

$WetGrass \mapsto_{0.7} Rain$

- Problems with combination, e.g., $Sprinkler$ causes $Rain$?

◇ *Probability*

Given the available evidence, A_{25} will get me there on time with probability 0.04

- Mahaviracarya (9th C.), Cardamo (1565) theory of gambling
  - ◇ Note: *Fuzzy logic* handles **degree of truth**, **not** uncertainty
    e.g., $WetGrass$ is true to degree 0.2

Outline

- Probability
- Syntax and Semantics
- Inference
- Independence and Bayes' Rule

Probability

◇ Probabilistic assertions **summarize** effects of

- laziness: failure to enumerate exceptions, qualifications, etc.
- ignorance: lack of relevant facts, initial conditions, etc.

◇ *Subjective* or *Bayesian* probability:

- Probabilities relate propositions to one's own state of knowledge
  e.g., $P(A_{25}|\text{no reported accidents}) = 0.06$
- They are **not** claims of a "probabilistic tendency" in the current situation
- They might be learned from past experience of similar situations
- Probabilities of propositions change with new evidence:
  e.g., $P(A_{25}|\text{no reported accidents, 5 a.m.}) = 0.15$

Making decisions under uncertainty

◇ Suppose I believe the following:

$$\begin{aligned} P(A_{25} \text{ gets me there on time}|\ldots) &= 0.04 \\ P(A_{90} \text{ gets me there on time}|\ldots) &= 0.70 \\ P(A_{120} \text{ gets me there on time}|\ldots) &= 0.95 \\ P(A_{1440} \text{ gets me there on time}|\ldots) &= 0.9999 \end{aligned}$$

- Which action to choose?

◇ Depends on both the probabilities and my *preferences*

- e.g., P(missing flight) vs. getting to airport early and waiting

◇ *Utility theory* (Chapter 16) is used to represent and infer preferences

◇ *Decision theory* = utility theory + probability theory

Probability basics

◇ Begin with a set Ω called the *sample space*

◇ Each $\omega \in \Omega$ is called a *sample point*, *possible world*, or *atomic event*

◇ *Probability space* or *probability model*:

- given a sample space Ω, assign a number $P(\omega)$ (the *probability* of ω) to every atomic event $\omega \in \Omega$

◇ A probability space must satisfy the following properties:

  - ◇ $0 \leq P(\omega) \leq 1$ for every $\omega \in \Omega$
  - ◇ $\sum_{\omega \in \Omega} P(\omega) = 1$

- e.g., for rolling a die,
  - ◇ $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6}$
  - ◇ $P(1) + P(2) + P(3) + P(4) + P(5) + P(6) = 1$

◇ An *event* A is any subset of Ω

◇ $P(A) = \sum_{\omega \in A} P(\omega)$

- E.g., $P(\text{die roll} < 4) = P(1) + P(2) + P(3) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$

Random variables

◇ A *random variable* is a function from sample points to some range

◇ The book uses capitalized words for random variables

- e.g., rolling the die: $Odd(\omega) = \begin{cases} true \text{ if } \omega \text{ is even,} \\ false \text{ otherwise} \end{cases}$

◇ A *probability distribution* gives a probability for every possible value.

◇ If X is a random variable, then

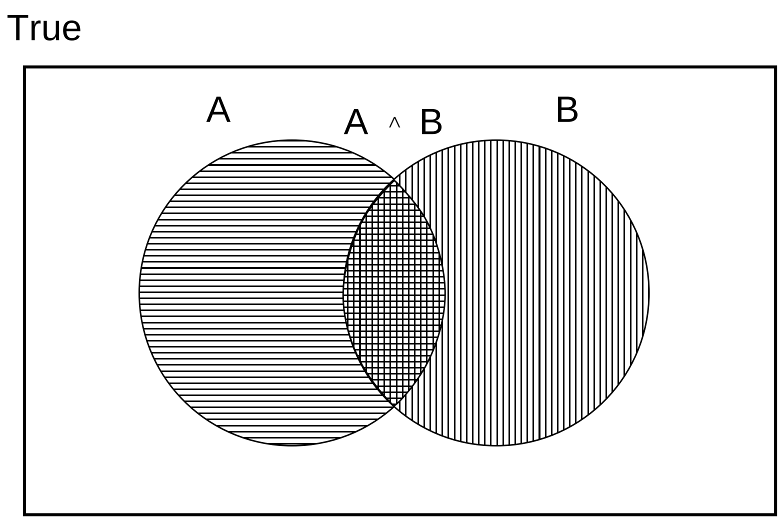
- $P(\underset{\substack{\uparrow \\ X \text{ here, not } X(\omega)}}{X} = x_i) = \sum \{P(\omega) : X(\omega) = x_i\}$

◇ E.g., $P(Odd = true) = P(1) + P(3) + P(5) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$

Why use probability?

◇ The definitions imply that certain logically related events must have related probabilities

- E.g., $P(A = true \lor B = true)$
  $$= P(A = true) + P(B = true) - P(A = true \land B = true)$$



◇ de Finetti (1931): an agent who bets according to probabilities that violate the axioms of probability can be forced to bet so as to lose money, regardless of the outcome

- Related to rational preferences, utility theory

Propositions

◇ A random variable is *Boolean* or *propositional* if its range is $\{true, false\}$

◇ To represent the event that a propositional random variable is true, we'll use the corresponding lower-case word

◇ In the die example, $Odd(\omega) = \begin{cases} true \text{ if } \omega \text{ is even,} \\ false \text{ otherwise} \end{cases}$

- $P(odd) = P(Odd = true) = \frac{1}{6}$
- $P(\neg odd) = P(Odd = false) = \frac{5}{6}$

◇ Boolean formula = disjunction of the sample points in which it is true

- E.g., suppose that
  - ◇ a is the event $A = true$
  - ◇ b is the event $B = true$
- Then

$$(a \vee b) \;\equiv\; (\neg a \wedge b) \;\vee\; (a \wedge \neg b) \;\vee\; (a \wedge b)$$
$$P(a \vee b) = P(\neg a \wedge b) + P(a \wedge \neg b) + P(a \wedge b)$$

Syntax for propositions

- *Propositional* or *Boolean* random variables
  - e.g., $Cavity$ (do I have a cavity in one of my teeth?)
  - $Cavity = true$ is a proposition, also written $cavity$
- *Discrete* random variables (finite or infinite)
  - e.g., $Weather = sunny, rain, cloudy,$ or $snow$
    - Values must be exhaustive and mutually exclusive
  - $Weather = rain$ is a proposition
- *Continuous* random variables (bounded or unbounded)
  - e.g., $Temp = 21.6$
  - also allow propositions such as $Temp < 22.0$.
- Arbitrary Boolean combinations of basic propositions
  - e.g., $\neg cavity$ means $Cavity = false$
- *Probabilities* of propositions
  - e.g., $P(cavity) = 0.1$ and $P(Weather = sunny) = 0.72$

Syntax for probability distributions

- Represent a discrete probability distribution as a vector of probability values that sum to 1:
  - $\mathbf{P}(Weather) = \langle\ 0.72,\ \ 0.1,\ \ 0.08,\ \ 0.1\ \rangle$
  - probabilities of *sunny*, *rain*, *cloudy*, *snow*
- If B is a Boolean random variable, then $\mathbf{P}(B) = \langle P(b), P(\neg b)\rangle$
- A *joint probability distribution* for a set of n random variables gives the probability of every atomic event on those variables (i.e., every sample point)
  - Represent it as an n-dimensional matrix
- e.g., $\mathbf{P}(Weather, Cavity)$:

	$Weather =$			
	sunny	*rain*	*cloudy*	*snow*
$Cavity = true$	0.144	0.02	0.016	0.02
$Cavity = false$	0.576	0.08	0.064	0.08

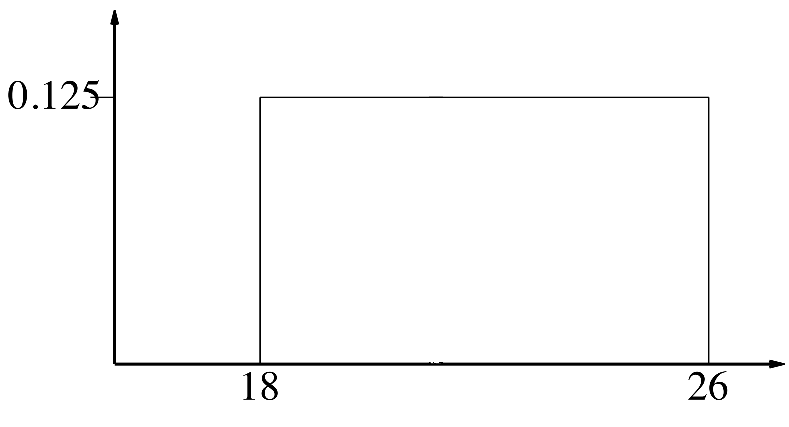
- Every event is a sum of sample points
  - its probability is determined by the joint distribution

Probability for continuous variables

◊ Express continuous probability distributions using parameterized *probability density* functions that integrate to 1

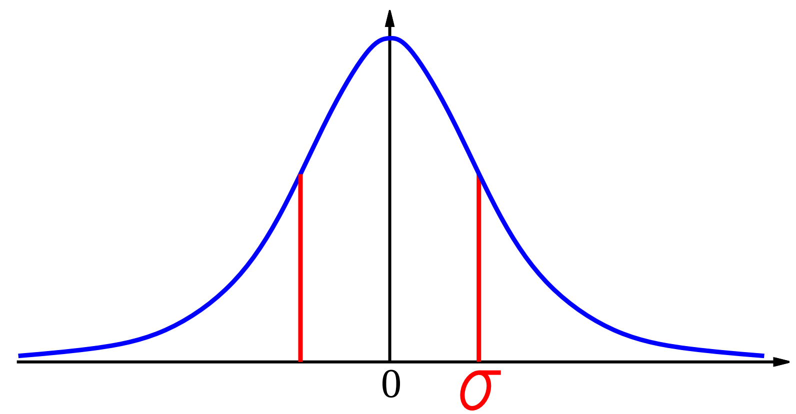
Uniform density between 18 and 26:

- $f(x) = U[18,26](x)$



Gaussian density:

- $P(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$



$$P(20 \leq X \leq 22) = \int_{20}^{22} 0.125 \, dx = 0.25$$

Conditional probability

◊ *Conditional* or *posterior probabilities*

- $P(cavity|toothache) = 0.8$
  - ◇ i.e., given that *toothache* is **all I know**
  - ◇ **not** "if *toothache* then 80% chance of *cavity*"

◊ Suppose we get more evidence, e.g., *cavity* is also given. Then

- $P(cavity|toothache, cavity) = 1$

◊ The less specific belief **remains valid**, but is not always **useful**

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

- $P(cavity|toothache, OriolesWin) = P(cavity|toothache) = 0.8$

◊ Notation for conditional distributions:

- $\mathbf{P}(Cavity \mid Toothache)$ represents a set of conditional probabilities:

$$\{P(cavity \mid toothache),\ \ P(\neg cavity \mid toothache),$$
$$P(cavity \mid \neg toothache), P(\neg cavity \mid \neg toothache)\}$$

Conditional probability

◇ Definition of conditional probability: $P(a \mid b) = P(a \wedge b)/P(b)$

- *Product rule* holds even if $P(b) = 0$: $P(a \wedge b) = P(a \mid b)\ P(b)$

◇ A general version holds for an entire probability distribution, e.g.,

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

◇ That **isn't** matrix multiplication, it's quantification. It means:

$$\begin{aligned} \forall\, w, c\ \ & P(Weather = w, Cavity = c) \\ & = P(Weather = w \mid Cavity = c)\ P(Cavity = c) \end{aligned}$$

- i.e.,

$$\begin{aligned} P(sunny, cavity) &= P(sunny \mid cavity)\ P(cavity) \\ P(sunny, \neg cavity) &= P(sunny \mid \neg cavity)\ P(\neg cavity) \\ P(rain, cavity) &= P(rain \mid cavity)\ P(cavity) \\ P(rain, \neg cavity) &= P(rain \mid \neg cavity)\ P(\neg cavity) \\ P(cloudy, cavity) &= P(cloudy \mid cavity)\ P(cavity) \\ P(cloudy, \neg cavity) &= P(cloudy \mid \neg cavity)\ P(\neg cavity) \\ P(snow, cavity) &= P(snow \mid cavity)\ P(cavity) \\ P(snow, \neg cavity) &= P(snow \mid \neg cavity)\ P(\neg cavity) \end{aligned}$$

Chain rule

◇ The *chain rule* is derived by successive application of the product rule:

$$\begin{aligned}
&\mathbf{P}(X_1, \ldots, X_n) \\
&\quad = \mathbf{P}(X_1, \ldots, X_{n-1})\, \mathbf{P}(X_n \mid X_1, \ldots, X_{n-1}) \\
&\quad = \mathbf{P}(X_1, \ldots, X_{n-2})\, \mathbf{P}(X_{n-1} \mid X_1, \ldots, X_{n-2})\, \mathbf{P}(X_n \mid X_1, \ldots, X_{n-1}) \\
&\quad = \ldots \\
&\quad = \prod_{i=1}^{n} \mathbf{P}(X_i \mid X_1, \ldots, X_{i-1})
\end{aligned}$$

Inference by enumeration

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

- Start with the joint distribution
- For any proposition ϕ, sum the probabilities of the atomic events where ϕ is true:

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

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

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

Inference by enumeration (example)

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

◇ You have a toothache, and want the probability that you have a cavity:

$$P(cavity \mid toothache) = P(cavity, toothache)/P(toothache)$$
$$= (.108 + .012)/(.108 + .012 + .016 + .064) = .6$$

$$P(\neg cavity \mid toothache) = P(\neg cavity, toothache)/P(toothache)$$
$$= (.016 + .064)/(.108 + .012 + .016 + .064) = .4$$

◇ We computed the conditional distribution on a **query variable**, $Cavity$, from a known value of an **evidence variable**, $Toothache$, with a **hidden variable**, $Catch$

◇ Don't know $Catch$'s value, so sum over all possible values

Inference by enumeration (example)

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

$$\begin{aligned} P(cavity \mid toothache) &= P(cavity, toothache)/P(toothache) \\ &= (.108 + .012)/(.108 + .012 + .016 + .064) = .6 \end{aligned}$$

$$\begin{aligned} P(\neg cavity \mid toothache) &= P(\neg cavity, toothache)/P(toothache) \\ &= (.016 + .064)/(.108 + .012 + .016 + .064) = .4 \end{aligned}$$

◇ $\alpha = 1/(.108 + .012 + .016 + .064)$ is a **normalization coefficient**

- It's the multiplier that we need to use, to get $P(cavity \mid toothache)$ and $P(\neg cavity \mid toothache)$ to sum to 1

◇ Don't need to compute α explicitly

- Can get it as a by-product of other computations

Inference by enumeration (example)

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

$$\mathbf{P}(Cavity \mid toothache) = \alpha\, \mathbf{P}(Cavity, toothache)$$

$$= \alpha\, [\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)]$$

$$= \alpha\, [\langle 0.108, 0.016 \rangle + \langle 0.012, 0.064 \rangle]$$

$$= \alpha\, \langle 0.12, 0.08 \rangle$$

- ◊ The two entries must sum to 1, so $\alpha = 1/(0.12 + 0.08) = 5$
- ◊ Thus $\mathbf{P}(Cavity \mid toothache) = 5\langle 0.12, 0.08 \rangle = \langle 0.6, 0.4 \rangle$

Inference by enumeration (in general)

- Let $\mathbf{X} = \{\text{all the variables}\}$
  - *Evidence variables* $\mathbf{E} \subseteq \mathbf{X} = \{\text{the variables we know the values of}\}$, $\mathbf{e} = \{\text{the values of the evidence variables}\}$
  - *Query variables* $\mathbf{Y} \subseteq \mathbf{X} = \{\text{the variables we want find out about}\}$, i.e., we want $\mathbf{P}(\mathbf{Y} \mid \mathbf{E}=\mathbf{e})$
  - *Hidden variables* $\mathbf{H} = \mathbf{X} - \mathbf{Y} - \mathbf{E}$
- E.g., $\mathbf{E} = \{Toothache\}$, $\mathbf{e} = \{toothache\}$, $\mathbf{Y} = \{Cavity\}$, $\mathbf{H} = \{Catch\}$
- Get $\mathbf{P}(\mathbf{Y} \mid \mathbf{E}=\mathbf{e})$ by *summing out* the hidden variables
  - Sum over all possible combinations of values for $\mathbf{H}$

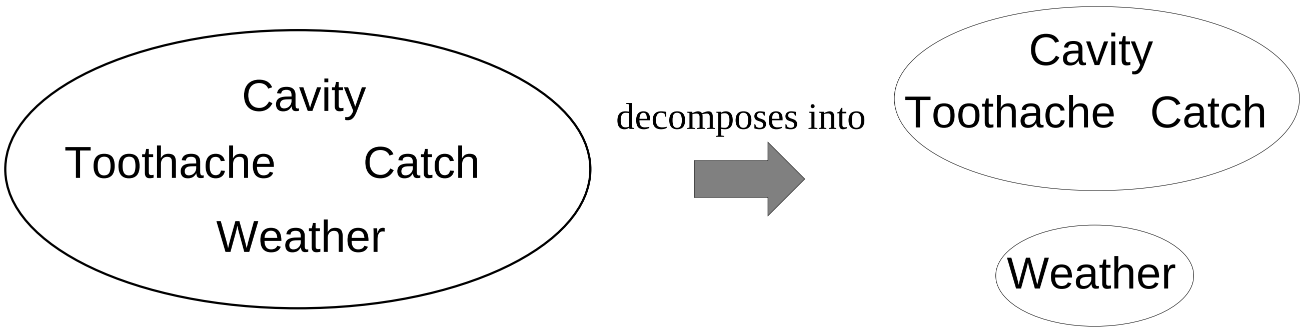
$$\mathbf{P}(\mathbf{Y} \mid \mathbf{E}=\mathbf{e}) = \alpha \mathbf{P}(\mathbf{Y}, \mathbf{E}=\mathbf{e}) = \alpha \sum_{\mathbf{h}} \mathbf{P}(\mathbf{Y}, \mathbf{E}=\mathbf{e}, \mathbf{H}=\mathbf{h})$$

- Problems:
  - Time complexity $O(d^n)$, where $d = \max_{h \in \mathbf{H}} |\{\text{possible values for } h\}|$
  - Space complexity $O(d^n)$ to store everything
  - How to find the numbers for $O(d^n)$ entries?

Independence

- Random variables A and B are *independent* iff

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



- $$\mathbf{P}(Toothache, Catch, Cavity, Weather) = \mathbf{P}(Toothache, Catch, Cavity)\,\mathbf{P}(Weather)$$
  - $2 \times 2 \times 2 \times 4 = 32$ entries reduced to $(2 \times 2 \times 2) + 4 = 12$ entries
- For n independent biased coins, 2^n entries reduced to n
- Absolute independence is powerful, but rare
  - E.g., dentistry is a large field with hundreds of variables, none of which are independent.
  - What to do?

Conditional independence

◇ Consider $\mathbf{P}(Toothache, Cavity, Catch)$

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

- $P(catch \mid toothache, cavity) = P(catch \mid cavity)$

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

- $P(catch \mid toothache, \neg cavity) = P(catch \mid \neg cavity)$

◇ Thus $Catch$ is *conditionally independent* of $Toothache$, given $Cavity$:

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

◇ Or equivalently:

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

Conditional independence, continued

◇ Use the chain rule on the full joint distribution:

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

◇ In many cases, conditional independence can reduce the size of the representation of the joint distribution
from exponential in n to linear in n.

- Example in next chapter: Bayes nets

Bayes' Rule

Product rule: $P(a \wedge b) = P(a \mid b)\ P(b) = P(b \mid a)\ P(a)$

$$\Rightarrow \textit{Bayes' rule } P(a \mid b) = \frac{P(b \mid a)\ P(a)}{P(b)}$$

or in probability distribution form,

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

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

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

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

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

Note: posterior probability of meningitis still very small!

Bayes' Rule and conditional independence

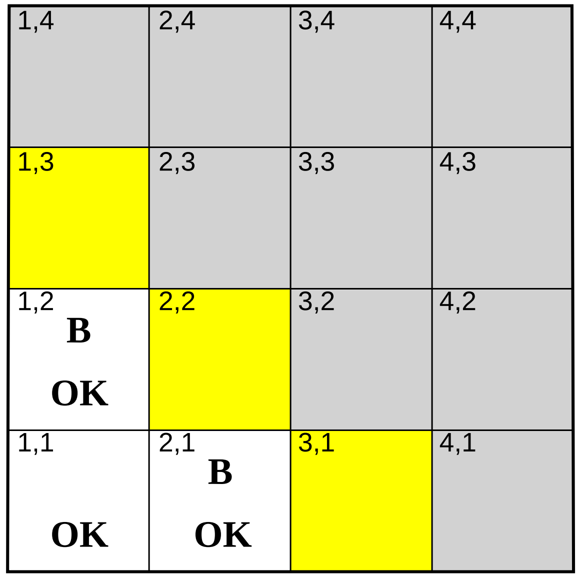
$$\mathbf{P}(Cavity \mid toothache \wedge catch)$$
$$= \mathbf{P}(toothache \wedge catch \mid Cavity)\, \mathbf{P}(Cavity) / P(toothache \wedge catch)$$
$$= \alpha\, \mathbf{P}(toothache \wedge catch \mid Cavity)\, \mathbf{P}(Cavity)$$
$$= \alpha\, \mathbf{P}(toothache \mid Cavity)\, \mathbf{P}(catch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

◊ *Naive Bayes* model: a mathematical model that assumes the effects are conditionally independent, given the cause

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

◊ Naive Bayes model ⇒ total number of parameters is **linear** in n

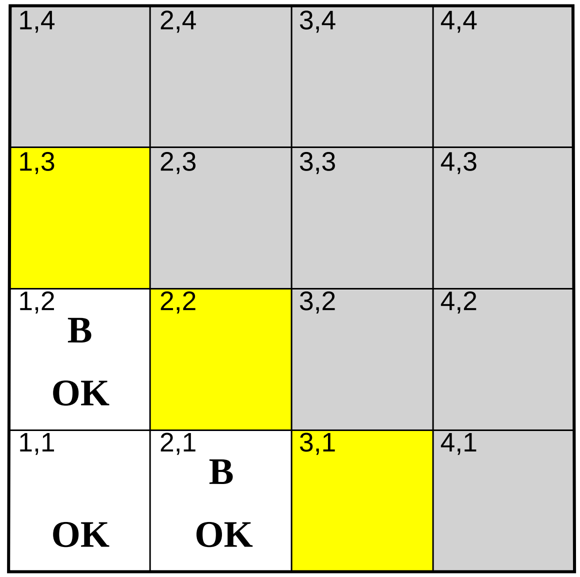
Wumpus World



- $P_{ij} = true$ iff $[i, j]$ contains a pit
- $B_{ij} = true$ iff $[i, j]$ is breezy
- The only breezes we care about are $B_{1,1}, B_{1,2}, B_{2,1}$; ignore all the others
- Then the joint distribution is

$$\mathbf{P}(P_{1,1}, \ldots, P_{4,4}, B_{1,1}, B_{1,2}, B_{2,1})$$

Specifying the probability model



◇ Apply the product rule to the joint distribution:

$$\mathbf{P}(P_{1,1}, \ldots, P_{4,4}, B_{1,1}, B_{1,2}, B_{2,1})$$
$$= \mathbf{P}(B_{1,1}, B_{1,2}, B_{2,1} \,|\, P_{1,1}, \ldots, P_{4,4})\ \ \mathbf{P}(P_{1,1}, \ldots, P_{4,4})$$

- First term: 1 if pits are adjacent to breezes, 0 otherwise
- Second term: pits are placed independently, probability 0.2 per square:

$$\mathbf{P}(P_{1,1}, \ldots, P_{4,4}) = \Pi_{i=1}^{4} \Pi_{j=1}^{4} \mathbf{P}(P_{i,j})$$

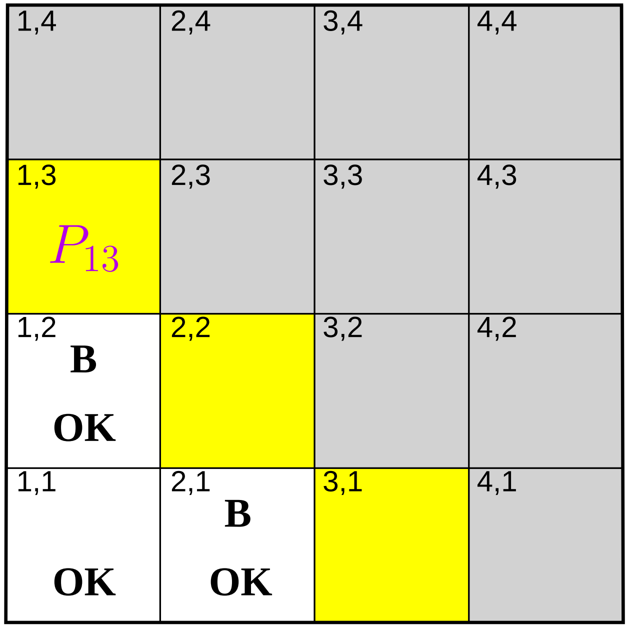
Inference by enumeration

◇ In general, $\mathbf{P}(\mathbf{Y} \mid \mathbf{E}=\mathbf{e}) = \alpha \mathbf{P}(\mathbf{Y}, \mathbf{E}=\mathbf{e}) = \alpha \sum_{\mathbf{h}} \mathbf{P}(\mathbf{Y}, \mathbf{E}=\mathbf{e}, \mathbf{H}=\mathbf{h})$

◇ Here, the evidence is from the 3 squares we've visited

◇ $\mathbf{e} = b^* \wedge p^*$, where

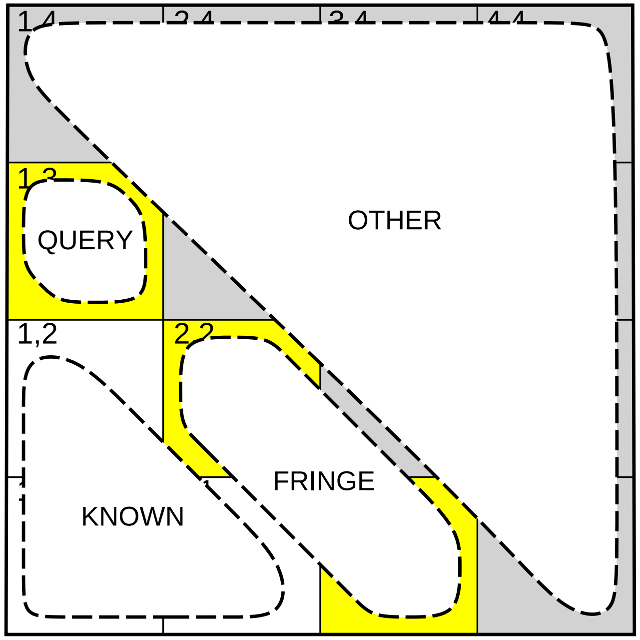
- $b^* = \neg b_{1,1} \wedge b_{1,2} \wedge b_{2,1}$
- $p^* = \neg p_{1,1} \wedge \neg p_{1,2} \wedge \neg p_{2,1}$



◇ So, $\mathbf{P}(P_{1,3} \mid p^*, b^*) = \alpha \sum_{unknown} \mathbf{P}(P_{1,3}, unknown, p^*, b^*)$

- $unknown$ = all P_{ij}s other than $P_{1,3}$ (the query variable) and $P_{1,1}, P_{1,2}, P_{2,1}$ (evidence variables)
- Two values for each P_{ij} $\Rightarrow$ grows exponentially with number of squares!

Using conditional independence



- Basic insight:
  - Given the *fringe* squares, b is conditionally independent of the *other* hidden squares

- The unknown variables are $Unknown = Fringe \cup Other$

$$\begin{aligned}\mathbf{P}(b^* \mid P_{1,3}, p^*, Unknown) &= \mathbf{P}(b^* \mid P_{1,3}, p^*, Fringe, Other) \\ &= \mathbf{P}(b^* \mid P_{1,3}, p^*, Fringe)\end{aligned}$$

  - Need to translate the query into a form where we can use this

Looks easy, doesn't it?

$$\mathbf{P}(P_{1,3}|p^*, b^*) = \mathbf{P}(P_{1,3}, p^*, b^*)/\mathbf{P}(p^*, b^*) = \alpha\mathbf{P}(P_{1,3}, p^*, b^*)$$

$$= \alpha\Sigma_{unknown}\mathbf{P}(P_{1,3}, unknown, p^*, b^*)$$

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

$$= \alpha\Sigma_{fringe}\Sigma_{other}\mathbf{P}(b^*|p^*, P_{1,3}, fringe, other)\mathbf{P}(P_{1,3}, p^*, fringe, other)$$

$$= \alpha\Sigma_{fringe}\Sigma_{other}\mathbf{P}(b^*|p^*, P_{1,3}, fringe)\mathbf{P}(P_{1,3}, p^*, fringe, other)$$

$$= \alpha\Sigma_{fringe}\mathbf{P}(b^*|p^*, P_{1,3}, fringe)\Sigma_{other}\mathbf{P}(P_{1,3}, p^*, fringe, other)$$

$$= \alpha\Sigma_{fringe}\mathbf{P}(b^*|p^*, P_{1,3}, fringe)\Sigma_{other}\mathbf{P}(P_{1,3})P(p^*)P(fringe)P(other)$$

$$= \alpha\, P(p^*)\mathbf{P}(P_{1,3})\Sigma_{fringe}\mathbf{P}(b^*|p^*, P_{1,3}, fringe)P(fringe)\Sigma_{other}P(other)$$

$$= \alpha'\,\mathbf{P}(P_{1,3})\Sigma_{fringe}\mathbf{P}(b^*|p^*, P_{1,3}, fringe)P(fringe)\Sigma_{other}P(other)$$

$$= \alpha'\,\mathbf{P}(P_{1,3})\Sigma_{fringe}\mathbf{P}(b^*|p^*, P_{1,3}, fringe)P(fringe)$$

Same thing, step by step

$$\mathbf{P}(P_{1,3}|\underline{p^*, b^*}) = \mathbf{P}(P_{1,3}, p^*, b^*)/\underline{\mathbf{P}(p^*, b^*)}$$

- Use the definition of conditional probability

Same thing, step by step

$\mathbf{P}(P_{1,3}|p^*, b^*) = \mathbf{P}(P_{1,3}, p^*, b^*)/\underline{\mathbf{P}(p^*, b^*)} = \underline{\alpha}\mathbf{P}(P_{1,3}, p^*, b^*)$

- $\mathbf{P}(p^*, b^*) = P(p^*, b^*)$ is a scalar constant; use as a normalization constant

Same thing, step by step

$$\mathbf{P}(P_{1,3}|p^*, b^*) = \mathbf{P}(P_{1,3}, p^*, b^*)/\mathbf{P}(p^*, b^*) = \alpha\mathbf{P}(P_{1,3}, p^*, b^*)$$

$$= \alpha\Sigma_{\underline{unknown}}\mathbf{P}(P_{1,3}, \underline{unknown}, p^*, b^*)$$

- Sum over the unknowns

Same thing, step by step

$$\mathbf{P}(P_{1,3}|p^*, b^*) = \mathbf{P}(P_{1,3}, p^*, b^*)/\mathbf{P}(p^*, b^*) = \alpha\mathbf{P}(P_{1,3}, p^*, b^*)$$

$$= \alpha\Sigma_{unknown}\mathbf{P}(P_{1,3}, unknown, p^*, \underline{b^*})$$

$$= \alpha\Sigma_{unknown}\mathbf{P}(\underline{b^*}|P_{1,3}, p^*, unknown)\mathbf{P}(P_{1,3}, p^*, unknown)$$

- Use the product rule

Same thing, step by step

$\mathbf{P}(P_{1,3}|p^*, b^*) = \mathbf{P}(P_{1,3}, p^*, b^*)/\mathbf{P}(p^*, b^*) = \alpha\mathbf{P}(P_{1,3}, p^*, b^*)$

$= \alpha \Sigma_{unknown} \mathbf{P}(P_{1,3}, unknown, p^*, b^*)$

$= \alpha \Sigma_{unknown} \mathbf{P}(b^*|P_{1,3}, p^*, \underline{unknown})\mathbf{P}(P_{1,3}, p^*, \underline{unknown})$

$= \alpha \Sigma_{fringe} \Sigma_{other} \mathbf{P}(b^*|p^*, P_{1,3}, \underline{fringe, other})\mathbf{P}(P_{1,3}, p^*, \underline{fringe, other})$

- Separate *unknown* into *fringe* and *other*

Same thing, step by step

$$\mathbf{P}(P_{1,3}|p^*, b^*) = \mathbf{P}(P_{1,3}, p^*, b^*)/\mathbf{P}(p^*, b^*) = \alpha\mathbf{P}(P_{1,3}, p^*, b^*)$$

$$= \alpha\Sigma_{unknown}\mathbf{P}(P_{1,3}, unknown, p^*, b^*)$$

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

$$= \alpha\Sigma_{fringe}\Sigma_{other}\mathbf{P}(b^*|p^*, P_{1,3}, \underline{fringe, other})\mathbf{P}(P_{1,3}, p^*, fringe, other)$$

$$= \alpha\Sigma_{fringe}\Sigma_{other}\mathbf{P}(b^*|p^*, P_{1,3}, \underline{fringe})\mathbf{P}(P_{1,3}, p^*, fringe, other)$$

- b^* is conditionally independent of $other$ given $fringe$

Same thing, step by step

$$\mathbf{P}(P_{1,3}|p^*, b^*) = \mathbf{P}(P_{1,3}, p^*, b^*)/\mathbf{P}(p^*, b^*) = \alpha\mathbf{P}(P_{1,3}, p^*, b^*)$$

$$= \alpha\Sigma_{unknown}\mathbf{P}(P_{1,3}, unknown, p^*, b^*)$$

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

$$= \alpha\Sigma_{fringe}\Sigma_{other}\mathbf{P}(b^*|p^*, P_{1,3}, fringe, other)\mathbf{P}(P_{1,3}, p^*, fringe, other)$$

$$= \alpha\Sigma_{fringe}\Sigma_{\underline{other}}\underline{\mathbf{P}(b^*|p^*, P_{1,3}, fringe)}\mathbf{P}(P_{1,3}, p^*, fringe, other)$$

$$= \alpha\Sigma_{fringe}\underline{\mathbf{P}(b^*|p^*, P_{1,3}, fringe)}\Sigma_{\underline{other}}\mathbf{P}(P_{1,3}, p^*, fringe, other)$$

- Move $\mathbf{P}(b^*|p^*, P_{1,3}, fringe)$ outward

Same thing, step by step

$$\mathbf{P}(P_{1,3}|p^*, b^*) = \mathbf{P}(P_{1,3}, p^*, b^*)/\mathbf{P}(p^*, b^*) = \alpha\mathbf{P}(P_{1,3}, p^*, b^*)$$

$$= \alpha\Sigma_{unknown}\mathbf{P}(P_{1,3}, unknown, p^*, b^*)$$

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

$$= \alpha\Sigma_{fringe}\Sigma_{other}\mathbf{P}(b^*|p^*, P_{1,3}, fringe, other)\mathbf{P}(P_{1,3}, p^*, fringe, other)$$

$$= \alpha\Sigma_{fringe}\Sigma_{other}\mathbf{P}(b^*|p^*, P_{1,3}, fringe)\mathbf{P}(P_{1,3}, p^*, fringe, other)$$

$$= \alpha\Sigma_{fringe}\mathbf{P}(b^*|p^*, P_{1,3}, fringe)\Sigma_{other}\underline{\mathbf{P}(P_{1,3}, p^*, fringe, other)}$$

$$= \alpha\Sigma_{fringe}\mathbf{P}(b^*|p^*, P_{1,3}, fringe)\Sigma_{other}\underline{\mathbf{P}(P_{1,3})P(p^*)P(fringe)P(other)}$$

- All of the pit locations are independent

Same thing, step by step

$$\mathbf{P}(P_{1,3}|p^*, b^*) = \mathbf{P}(P_{1,3}, p^*, b^*)/\mathbf{P}(p^*, b^*) = \alpha\mathbf{P}(P_{1,3}, p^*, b^*)$$

$$= \alpha\Sigma_{unknown}\mathbf{P}(P_{1,3}, unknown, p^*, b^*)$$

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

$$= \alpha\Sigma_{fringe}\Sigma_{other}\mathbf{P}(b^*|p^*, P_{1,3}, fringe, other)\mathbf{P}(P_{1,3}, p^*, fringe, other)$$

$$= \alpha\Sigma_{fringe}\Sigma_{other}\mathbf{P}(b^*|p^*, P_{1,3}, fringe)\mathbf{P}(P_{1,3}, p^*, fringe, other)$$

$$= \alpha\Sigma_{fringe}\mathbf{P}(b^*|p^*, P_{1,3}, fringe)\Sigma_{other}\mathbf{P}(P_{1,3}, p^*, fringe, other)$$

$$= \alpha\Sigma_{fringe}\mathbf{P}(b^*|p^*, P_{1,3}, fringe)\Sigma_{other}\underline{\mathbf{P}(P_{1,3})P(p^*)P(fringe)}P(other)$$

$$= \alpha\,\underline{P(p^*)\mathbf{P}(P_{1,3})}\Sigma_{fringe}\mathbf{P}(b^*|p^*, P_{1,3}, fringe)\underline{P(fringe)}\Sigma_{other}P(other)$$

- Move $P(p^*)$, $\mathbf{P}(P_{1,3})$, and $P(fringe)$ outward

Same thing, step by step

$$\mathbf{P}(P_{1,3}|p^*, b^*) = \mathbf{P}(P_{1,3}, p^*, b^*)/\mathbf{P}(p^*, b^*) = \alpha\mathbf{P}(P_{1,3}, p^*, b^*)$$

$$= \alpha\Sigma_{unknown}\mathbf{P}(P_{1,3}, unknown, p^*, b^*)$$

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

$$= \alpha\Sigma_{fringe}\Sigma_{other}\mathbf{P}(b^*|p^*, P_{1,3}, fringe, other)\mathbf{P}(P_{1,3}, p^*, fringe, other)$$

$$= \alpha\Sigma_{fringe}\Sigma_{other}\mathbf{P}(b^*|p^*, P_{1,3}, fringe)\mathbf{P}(P_{1,3}, p^*, fringe, other)$$

$$= \alpha\Sigma_{fringe}\mathbf{P}(b^*|p^*, P_{1,3}, fringe)\Sigma_{other}\mathbf{P}(P_{1,3}, p^*, fringe, other)$$

$$= \alpha\Sigma_{fringe}\mathbf{P}(b^*|p^*, P_{1,3}, fringe)\Sigma_{other}\mathbf{P}(P_{1,3})P(p^*)P(fringe)P(other)$$

$$= \alpha\, P(p^*)\mathbf{P}(P_{1,3})\Sigma_{fringe}\mathbf{P}(b^*|p^*, P_{1,3}, fringe)P(fringe)\,\underline{\Sigma_{other}P(other)}$$

$$= \alpha\, P(p^*)\mathbf{P}(P_{1,3})\Sigma_{fringe}\mathbf{P}(b^*|p^*, P_{1,3}, fringe)P(fringe)$$

- Remove $\Sigma_{other}P(other)$ because it equals 1

Same thing, step by step

$$\mathbf{P}(P_{1,3}|p^*, b^*) = \mathbf{P}(P_{1,3}, p^*, b^*)/\mathbf{P}(p^*, b^*) = \alpha\mathbf{P}(P_{1,3}, p^*, b^*)$$

$$= \alpha\Sigma_{unknown}\mathbf{P}(P_{1,3}, unknown, p^*, b^*)$$

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

$$= \alpha\Sigma_{fringe}\Sigma_{other}\mathbf{P}(b^*|p^*, P_{1,3}, fringe, other)\mathbf{P}(P_{1,3}, p^*, fringe, other)$$

$$= \alpha\Sigma_{fringe}\Sigma_{other}\mathbf{P}(b^*|p^*, P_{1,3}, fringe)\mathbf{P}(P_{1,3}, p^*, fringe, other)$$

$$= \alpha\Sigma_{fringe}\mathbf{P}(b^*|p^*, P_{1,3}, fringe)\Sigma_{other}\mathbf{P}(P_{1,3}, p^*, fringe, other)$$

$$= \alpha\Sigma_{fringe}\mathbf{P}(b^*|p^*, P_{1,3}, fringe)\Sigma_{other}\mathbf{P}(P_{1,3})P(p^*)P(fringe)P(other)$$

$$= \alpha\, P(p^*)\mathbf{P}(P_{1,3})\Sigma_{fringe}\mathbf{P}(b^*|p^*, P_{1,3}, fringe)P(fringe)\Sigma_{other}P(other)$$

$$= \underline{\alpha\, P(p^*)}\mathbf{P}(P_{1,3})\Sigma_{fringe}\mathbf{P}(b^*|p^*, P_{1,3}, fringe)P(fringe)$$

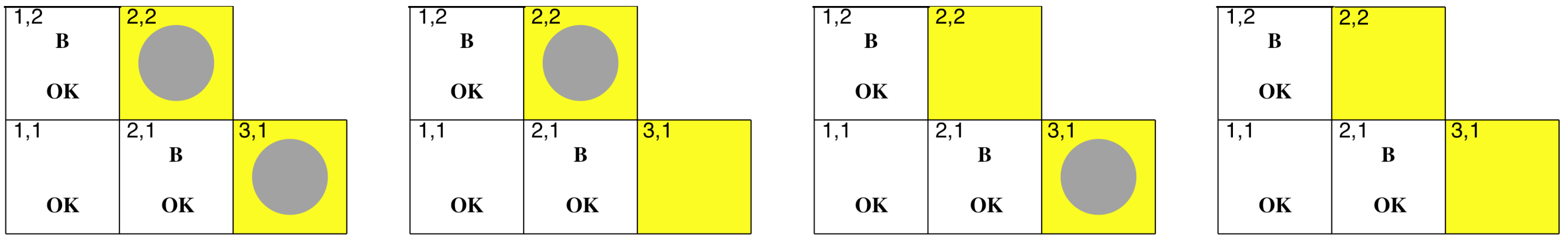
$$= \underline{\alpha'}\,\mathbf{P}(P_{1,3})\Sigma_{fringe}\mathbf{P}(b^*|p^*, P_{1,3}, fringe)P(fringe)$$

- $P(p^*)$ is constant, so make it part of the normalization constant

How to get the answer?

$$\mathbf{P}(P_{1,3}|p^*, b^*) = \alpha' \mathbf{P}(P_{1,3}) \Sigma_{fringe} \mathbf{P}(b^*|p^*, P_{1,3}, fringe) P(fringe)$$

◇ Not hard to compute, because there are only four possible fringes:



0.2 x 0.2 = 0.04   0.2 x 0.8 = 0.16   0.8 x 0.2 = 0.16   0.8 x 0.8 = 0.64

◇ Start by rewriting as two separate equations:

$$P(p_{1,3}|p^*, b^*) = \alpha' P(p_{1,3}) \quad \Sigma_{fringe} P(b^*|p^*, p_{1,3}, fringe) P(fringe)$$

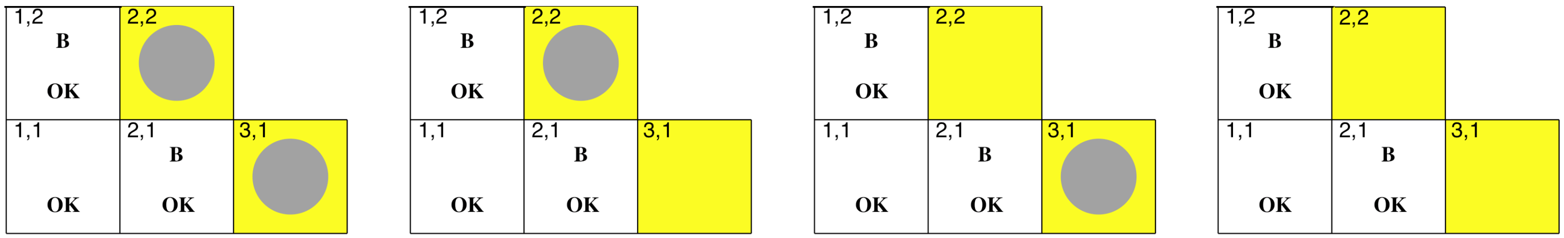
$$P(\neg p_{1,3}|p^*, b^*) = \alpha' P(\neg p_{1,3}) \Sigma_{fringe} P(b^*|p^*, \neg p_{1,3}, fringe) P(fringe)$$

◇ For each fringe, $P(b^*|p^*, p_{1,3}, fringe)$ is 1 if the breezes occur, 0 otherwise

- 1, 1, 1, and 0 for the four fringes above

◇ Similarly, for $P(b^*|p^*, \neg p_{1,3}, fringe)$, we get 1, 1, 0, and 0

Getting the answer



0.2 x 0.2 = 0.04 0.2 x 0.8 = 0.16 0.8 x 0.2 = 0.16 0.8 x 0.8 = 0.64

$$\sum_{fringe} P(b^*|p^*, p_{1,3}, fringe)P(fringe) = 1(0.04) + 1(0.16) + 1(0.16) + 0 = 0.36$$

$$\sum_{fringe} P(b^*|p^*, \neg p_{1,3}, fringe)P(fringe) = 1(0.04) + 1(0.16) \ + 0 \ + 0 = 0.2$$

so $\mathbf{P}(P_{1,3}|p^*, b^*) \ = \ \alpha' \, \mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b^*|p^*, P_{1,3}, fringe)P(fringe)$

$$= \alpha' \, \langle 0.2, \ 0.8 \rangle \, \langle 0.36, 0.2 \rangle = \alpha' \, \langle 0.072, 0.16 \rangle$$

so $\alpha' \ = \ 1/(0.072 + 0.16) = 1/0.232 \approx 4.31$

so $\mathbf{P}(P_{1,3}|p^*, b^*) \ = \ \langle 0.072 \, \alpha', 0.16 \, \alpha' \rangle \approx \langle 0.31, 0.69 \rangle$

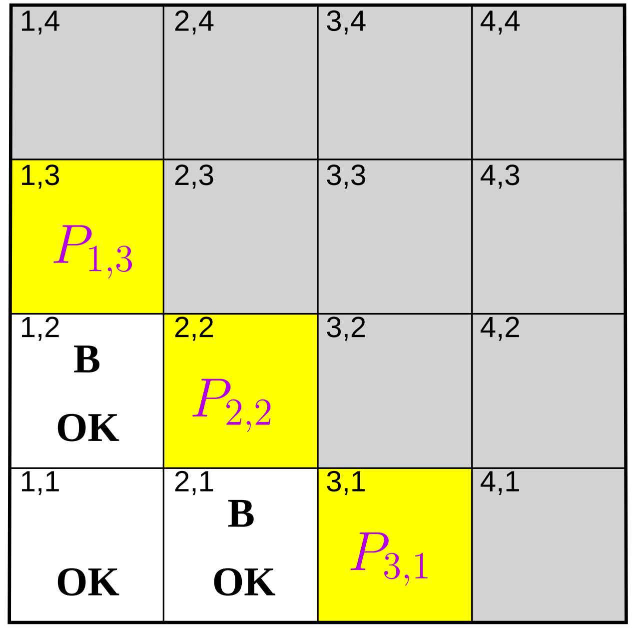
Additional answers

◇ We have

- $\mathbf{P}(P_{1,3}|p^*, b^*) \approx \langle 0.31, 0.69 \rangle$

◇ Similarly,

- $\mathbf{P}(P_{2,2}|p^*, b^*) \approx \langle 0.86, 0.14 \rangle$
- $\mathbf{P}(P_{3,1}|p^*, b^*) \approx \langle 0.31, 0.69 \rangle$



◇ Questions:

- Why don't these add up to 1?
- Which square should we move to?

Summary

- ♢ Probability is a rigorous formalism for uncertain knowledge
- ♢ *Joint probability distribution* specifies probability of every *atomic event*
- ♢ Queries can be answered by *inference by enumeration* (summing over atomic events)
- ♢ Can reduce combinatorial explosion using *independence* and *conditional independence*

Homework assignment

I'll post it on Piazza