Last update: April 13, 2010

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 my tires remain intact etc etc."

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

Issues: What assumptions are reasonable? How to handle contradiction?

Rules with fudge factors:

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

$Sprinkler \mapsto_{0.99} WetGrass$

$WetGrass \mapsto_{0.7} Rain$

Issues: 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

(*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*
missing flight vs. getting to airport early and waiting, etc.

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 a *sample point*/*possible world*/*atomic event*
    e.g., 6 possible rolls of a die: $\{1, 2, 3, 4, 5, 6\}$

Probability space or *probability model*: take a sample space Ω, and 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 $\omega \in \Omega$

$\Sigma_{\omega \in \Omega} P(\omega) = 1$

e.g., for rolling the die, $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$.

An *event* A is any subset of Ω

$$P(A) = \Sigma_{\{\omega \in A\}} P(\omega)$$

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

Random variables

A *random variable* is a function from sample points to some range
We'll use 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 $P(X = x_i) = \Sigma\{P(\omega) : X(\omega) = x_i\}$

e.g., $P(Odd = true) = P(1) + P(3) + P(5) = 1/6 + 1/6 + 1/6 = 1/2$

Note that we don't write Odd's argument ω here.

Propositions

Odd is a *Boolean* or *propositional* random variable: its range is $\{true, false\}$

We'll use the corresponding lower-case word (in this case odd) for the event that a propositional random variable is true

e.g., $P(odd) = P(Odd = true) = 1/6$
$P(\neg odd) = P(Odd = false) = 5/6$

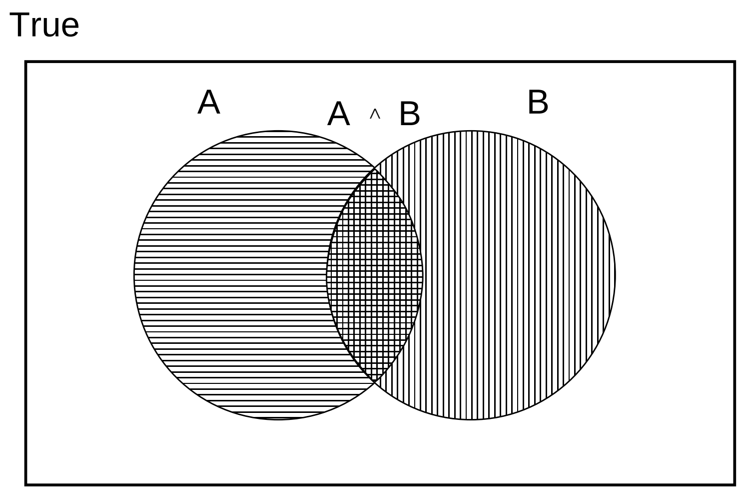
Boolean formula = disjunction of the sample points in which it is true

e.g., $(a \lor b) \equiv (\neg a \land b) \lor (a \land \neg b) \lor (a \land b)$
$\Rightarrow\ P(a \lor b) = P(\neg a \land b) + P(a \land \neg b) + P(a \land b)$

Why use probability?

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

E.g., $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$



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

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., $Weather$ is one of $\langle sunny, rain, cloudy, snow \rangle$

$Weather = rain$ is a proposition

Values must be exhaustive and mutually exclusive

Continuous random variables (bounded or unbounded)

e.g., $Temp = 21.6$; also allow, e.g., $Temp < 22.0$.

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:

$$\mathbf{P}(Weather) = \langle\ 0.72,\quad 0.1,\quad 0.08,\quad 0.1\ \rangle$$

probabilities of $sunny$, $rain$, $cloudy$, $snow$ (must sum to 1)

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., $\mathbf{P}(Weather, Cavity)$ is a 4×2 matrix:

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

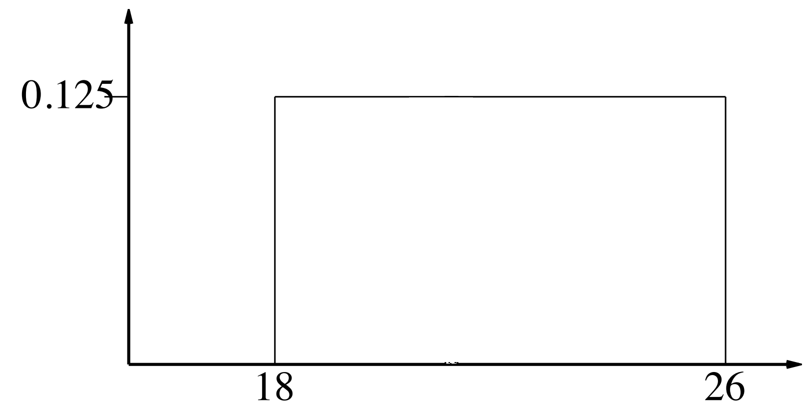
Every event is a sum of sample points, hence its probability is determined by the joint distribution

Probability for continuous variables

Express continuous probability distributions using parameterized functions, e.g.,

Uniform density between 18 and 26

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

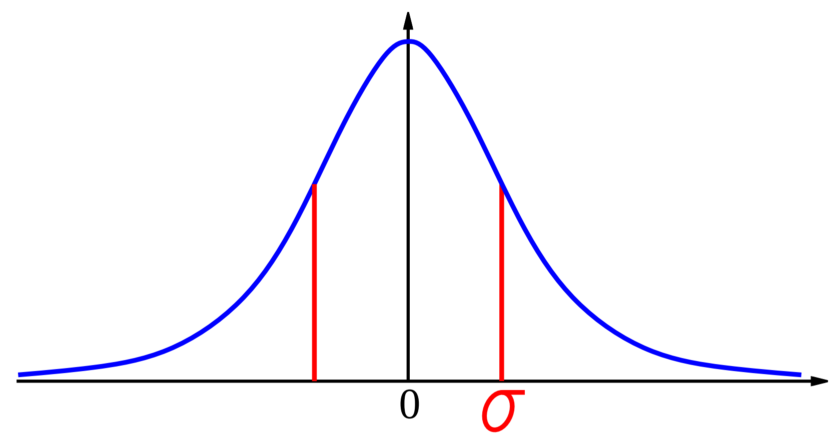


Here f is a *density*; integrates to 1.

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

Gaussian density

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



Conditional probability

Conditional or *posterior probabilities*

e.g., $P(cavity|toothache) = 0.8$

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

NOT "if $toothache$ then 80% chance of $cavity$"

(Notation for conditional distributions:

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

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

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

Note: the less specific belief **remains valid**, but is not always **useful**

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

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

Conditional probability

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

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

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

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

This is **not** matrix multiplication, it's a set of $4 \times 2 = 8$ equations:

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

Chain rule is derived by successive application of product rule:

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

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		$\neg$ *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:

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

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

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:

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

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

$$\begin{aligned} P(cavity \lor toothache) \\ &= 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 \\ &= 0.28 \end{aligned}$$

Inference by enumeration

Start with the joint distribution:

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

Can also compute conditional probabilities:

$$P(\neg cavity | toothache) = P(\neg cavity \land toothache)/P(toothache)$$
$$= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

General idea: compute distribution on *query variable* (e.g., $Cavity$)
by fixing *evidence variables* ($Toothache$)
and summing over all possible values of *hidden variables* ($Catch$)

Normalization

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

$$P(\neg cavity|toothache) = \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

$$P(cavity|toothache) = \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

The quantity $\alpha = 1/P(toothache) = 1/(0.108 + 0.012 + 0.016 + 0.064)$ can be viewed as a *normalization constant*. It's the multiplier that's needed to get $\mathbf{P}(Cavity|toothache)$ to sum to 1.

Thinking of α this way is useful because it enables us to compute α as a by-product of other computations

Normalization

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

Recall that *events* are $lower\ case$, *random variables* are $Capitalized$

For a set of n random variables, $\mathbf{P}$ is an n-dimensional table giving the probability of each possible combination of values

$$\mathbf{P}(Cavity|toothache) = \alpha\,\mathbf{P}(Cavity, toothache)$$
$$= \alpha\,[\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)]$$
$$= \alpha\,[\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle]$$
$$= \alpha\,\langle 0.12, 0.08\rangle$$
$$= \langle 0.6, 0.4\rangle \text{ since the entries must sum to 1}$$

Compute α directly from the last line, as $\alpha = 1/(0.12 + 0.08)$

Inference by enumeration, continued

Let $\mathbf{X} = \{\text{all the variables}\}$. Typically, we want the posterior (i.e., conditional) 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:

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

i.e., sum over every possible combination of values $\mathbf{h} = \langle h_1, \ldots, h_n \rangle$ of the hidden varables $\mathbf{H} = \langle H_1, \ldots, H_n \rangle$

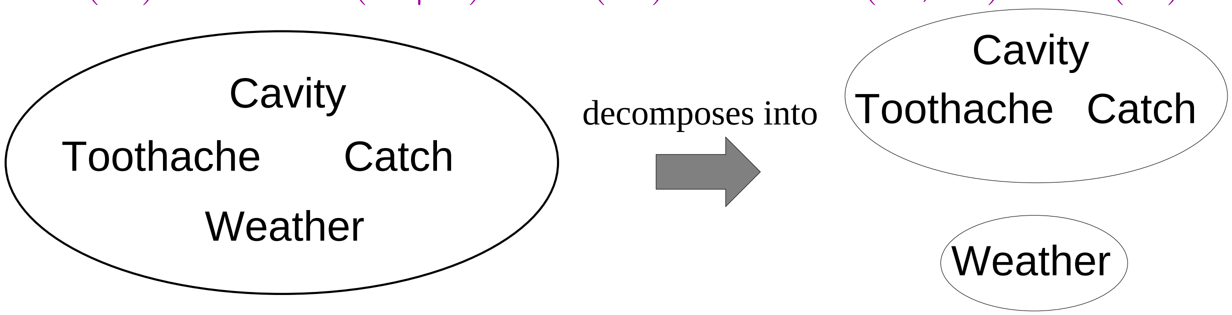
Obvious problems:

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

Independence

Random variables A and B are *independent* iff

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



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

$2 \times 2 \times 2 \times 4 = 32$ entries reduced to $(2 \times 2 \times 2) + 4 = 12$ entries

For n independent biased coins, 2^n entries reduced to n

Absolute independence 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:

$$P(catch|toothache, cavity) = P(catch|cavity)$$

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

$$P(catch|toothache, \neg cavity) = P(catch|\neg cavity)$$

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

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

Or equivalently:

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

Conditional independence, continued

Write out full joint distribution using chain rule:

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

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.

Bayes' Rule

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

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

or in probability 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!

Bayes' Rule and conditional independence

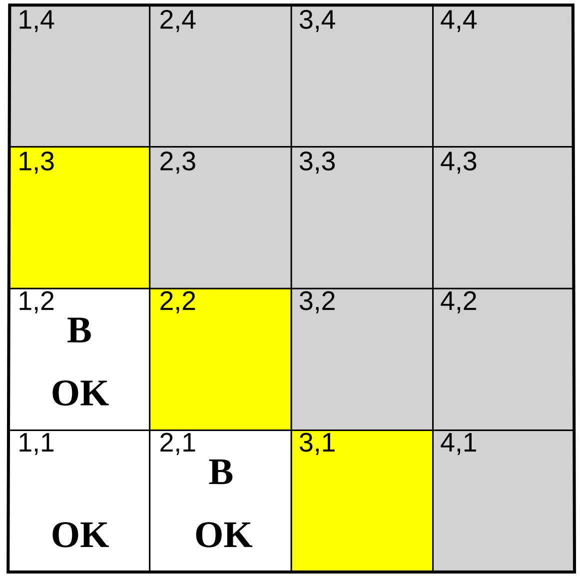
$$\mathbf{P}(Cavity|toothache \land catch)$$
$$= \mathbf{P}(toothache \land catch|Cavity)\mathbf{P}(Cavity)/P(toothache \land catch)$$
$$= \alpha\,\mathbf{P}(toothache \land catch|Cavity)\mathbf{P}(Cavity)$$
$$= \alpha\,\mathbf{P}(toothache|Cavity)\mathbf{P}(catch|Cavity)\mathbf{P}(Cavity)$$

A *naive Bayes* model is a mathematical model that assumes the effects are conditionally independent, given the cause

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

Naive Bayes model $\Rightarrow$ total number of parameters is **linear** in n

Wumpus World



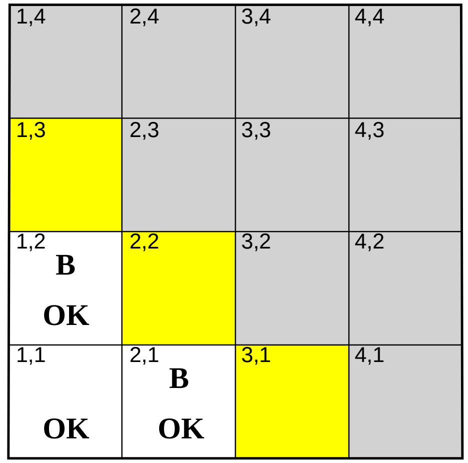
$P_{ij} = true$ iff $[i, j]$ contains a pit

$B_{ij} = true$ 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



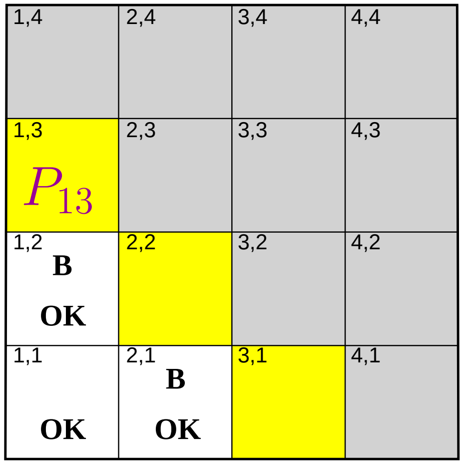
Apply the product rule to the joint distribution:

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

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



General form of query: $\mathbf{P}(\mathbf{Y}|\mathbf{E}=\mathbf{e}) = \alpha \mathbf{P}(\mathbf{Y}, \mathbf{E}=\mathbf{e}) = \alpha \Sigma_{\mathbf{h}} \mathbf{P}(\mathbf{Y}, \mathbf{E}=\mathbf{e}, \mathbf{H}=\mathbf{h})$

In our case, query is $\mathbf{P}(P_{1,3}|p^*, b^*)$, where the evidence is

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

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

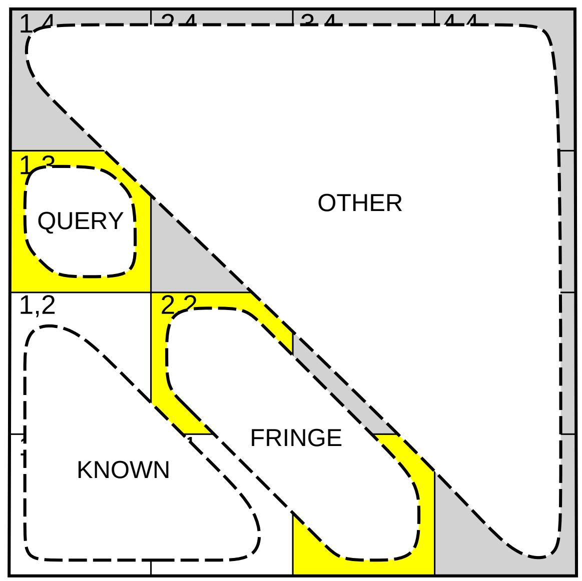
Sum over hidden variables: $\mathbf{P}(P_{1,3}|p^*, b^*) = \alpha \Sigma_{unknown} \mathbf{P}(P_{1,3}, unknown, p^*, b^*)$

$unknown =$ all P_{ij}s other than $P_{1,3}$ and the known squares $(P_{1,1}, P_{1,2}, P_{2,1})$

Two values for each P_{ij} $\Rightarrow$ grows exponentially with number of squares!

Using conditional independence

Basic insight: Given the *fringe* squares (see below), b is conditionally independent of the *other* hidden squares



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

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

Next: translate the query into a form where we can use this

Using conditional independence, continued

Looks easy, doesn't it?

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

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

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

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

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

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

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

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

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

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

Same thing, step by step

Use the definition of conditional probability

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

Same thing, step by step

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

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

Same thing, step by step

Sum over the unknowns

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

Same thing, step by step

Use the product rule

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

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

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

Same thing, step by step

Separate $unknown$ into $fringe$ and $other$

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

Same thing, step by step

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

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

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

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

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

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

Same thing, step by step

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

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

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

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

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

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

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

Same thing, step by step

All of the pit locations are independent

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

Same thing, step by step

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

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

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

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

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

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

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

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

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

Same thing, step by step

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

$$
\begin{aligned}
&\mathbf{P}(P_{1,3}|p^*, b^*) = \mathbf{P}(P_{1,3}, p^*, b^*)/\mathbf{P}(p^*, b^*) = \alpha \mathbf{P}(P_{1,3}, p^*, b^*)\\
&\quad = \alpha \sum_{unknown} \mathbf{P}(P_{1,3}, unknown, p^*, b^*)\\
&\quad = \alpha \sum_{unknown} \mathbf{P}(b^*|P_{1,3}, p^*, unknown)\mathbf{P}(P_{1,3}, p^*, unknown)\\
&\quad = \alpha \sum_{fringe} \sum_{other} \mathbf{P}(b^*|p^*, P_{1,3}, fringe, other)\mathbf{P}(P_{1,3}, p^*, fringe, other)\\
&\quad = \alpha \sum_{fringe} \sum_{other} \mathbf{P}(b^*|p^*, P_{1,3}, fringe)\mathbf{P}(P_{1,3}, p^*, fringe, other)\\
&\quad = \alpha \sum_{fringe} \mathbf{P}(b^*|p^*, P_{1,3}, fringe) \sum_{other} \mathbf{P}(P_{1,3}, p^*, fringe, other)\\
&\quad = \alpha \sum_{fringe} \mathbf{P}(b^*|p^*, P_{1,3}, fringe) \sum_{other} \mathbf{P}(P_{1,3})P(p^*)P(fringe)P(other)\\
&\quad = \alpha\, P(p^*)\mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b^*|p^*, P_{1,3}, fringe)P(fringe) \underline{\sum_{other} P(other)}\\
&\quad = \alpha\, P(p^*)\mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b^*|p^*, P_{1,3}, fringe)P(fringe)
\end{aligned}
$$

Same thing, step by step

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

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

How to get the answer?

We have

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

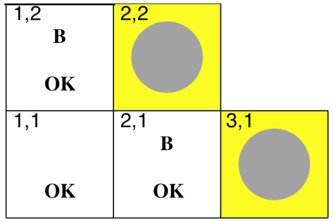
- It won't be hard to compute $\sum_{fringe} \mathbf{P}(b^*|p^*, P_{1,3}, fringe) P(fringe)$, because there are only four possible fringes (see next slide)
- We know that $\mathbf{P}(P_{1,3}) = \langle 0.2, 0.8 \rangle$.
- We can compute the normalization coefficient α' afterwards; it's whatever number will make the probabilities sum to 1.

Start by rewriting as two separate equations:

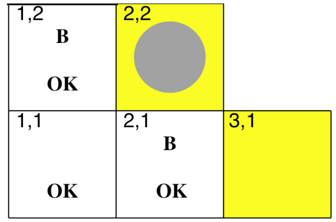
$$\begin{aligned} P(p_{1,3}|p^*, b^*) &= \alpha' \, P(p_{1,3}) \sum_{fringe} P(b^*|p^*, p_{1,3}, fringe) P(fringe) \\ P(\neg p_{1,3}|p^*, b^*) &= \alpha' \, P(\neg p_{1,3}) \sum_{fringe} P(b^*|p^*, \neg p_{1,3}, fringe) P(fringe) \end{aligned}$$

Getting the answer

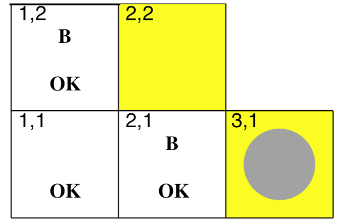
Four possible fringes:



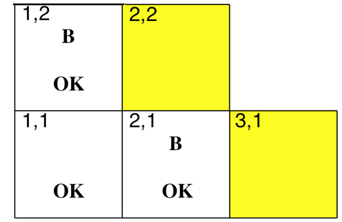
0.2 x 0.2 = 0.04



0.2 x 0.8 = 0.16



0.8 x 0.2 = 0.16



0.8 x 0.8 = 0.64

For each of them, $P(b^*|\ldots)$ is 1 if the breezes occur, 0 otherwise

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

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

so $$\begin{aligned}\mathbf{P}(P_{1,3}|p^*, b^*) &= \alpha' \, \mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b^*|p^*, P_{1,3}, fringe)P(fringe) \\ &= \alpha' \langle 0.2,\; 0.8\rangle \, \langle 0.36, 0.2\rangle \\ &= \alpha' \langle 0.072, 0.16\rangle\end{aligned}$$

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

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

Similarly, $\mathbf{P}(P_{2,2}|p^*, b^*) \approx \langle 0.86, 0.14\rangle$

Summary

Probability is a rigorous formalism for uncertain knowledge

Joint probability distribution specifies probability of every *atomic event*

Queries can be answered by *inference by enumeration* (summing over atomic events)

Can reduce combinatorial explosion using *independence* and *conditional independence*

Homework assignment

I'll post it to the discussion forum