Uncertainty

CHAPTER 13
HASSAN KHOSRAVI
SPRING2011

Outline

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

- In many cases, our knowledge of the world is incomplete (not enough information) or uncertain (sensors are unreliable).
- Often, rules about the domain are incomplete or even incorrect
- We have to act in spite of this!
- Drawing conclusions under uncertainty

Example

- Goal: The agent wants to drive someone to air port to catch a flight

  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 (traffic reports)
3. uncertainty in action outcomes (flat tire, etc.)
4. immense complexity of modeling 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.”

(A_{1440} might reasonably be said to get me there on time but I'd have to stay overnight in the airport …)

Uncertainty in logical rules

Example: Expert dental diagnosis system.

$$\forall p\ [Symptom(p, toothache) \Rightarrow Disease(p, cavity)]$$

$\rightarrow$ This rule is *incorrect*! Better:

$$\forall p\ [Symptom(p, toothache) \Rightarrow Disease(p, cavity) \lor Disease(p, gum\_disease) \lor \ldots]$$

… but we don’t know all the causes.

Perhaps a *causal* rule is better?

$$\forall p\ [Disease(p, cavity) \Rightarrow Symptom(p, toothache)]$$

$\rightarrow$ Does not allow to reason from symptons to causes & is still wrong!

Probability

- First order logic fails with medical diagnosis
  - laziness: failure to enumerate exceptions, qualifications, etc.
  - Theoretical ignorance: lack of relevant facts, initial conditions, etc.
  - Practical ignorance: Even if we know all the rules, a patience might not have done all the necessary tests.

- Probabilistic assertions summarize effects of
  - Laziness
  - Ignorance

- Degree of belief vs degree of truth
  - Probability of 0.8 does not mean 80% true.
  - A card is taken out of a deck of cards
    - The probability of it being Ace of clubs
    - The probability after seeing the card
  - Being 0.8 intelligence is not probabilistic. It means on a scale of 0 to 1 you are 0.8 intelligence

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?
-

Making decisions under uncertainty

Suppose I believe the following:

- $P(A_{25}$ gets me there on time | …) = 0.04
- $P(A_{90}$ gets me there on time | …) = 0.70
- $P(A_{120}$ gets me there on time | …) = 0.95
- $P(A_{1440}$ gets me there on time | …) = 0.9999
- Which action to choose? Which one is rational?
- Depends on my preferences for missing flight vs. time spent waiting, etc.
- Utility theory is used to represent and infer preferences
- Decision theory = probability theory + utility theory

The fundamental idea of decision theory is that an agent is rational if and only if it chooses the action that yields that highest expected utility, averaged over all the possible outcomes of the action.

Syntax

- Basic element: random variable
- Similar to propositional logic: possible worlds defined by assignment of values to random variables.
- Boolean random variables
  e.g., *Cavity* (do I have a cavity?)
- Discrete random variables
  e.g., *Weather* is one of <*sunny,rainy,cloudy,snow*>
- Elementary proposition constructed by assignment of a value to a random variable: e.g., *Weather* = *sunny*, *Cavity* = *false*
- Complex propositions formed from elementary propositions and standard logical connectives e.g., $Weather = sunny \vee Cavity = false$

Syntax

- Atomic event: A complete specification of the state of the world about which the agent is uncertain

  E.g., if the world consists of only two Boolean variables *Cavity* and *Toothache*, then there are 4 distinct atomic events:

  $Cavity = false \wedge Toothache = false$

  $Cavity = false \wedge Toothache = true$

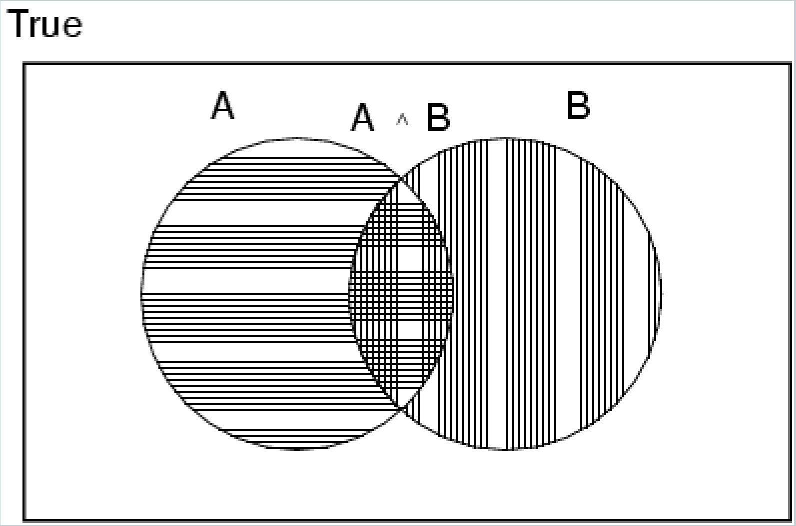
  $Cavity = true \wedge Toothache = false$

  $Cavity = true \wedge Toothache = true$

- Atomic events are
  - mutually exclusive: at most one is true
  - Exhaustive: at least one is true

Axioms of probability

- For any propositions A, B
- $0 \leq P(A) \leq 1$
  - $P(true) = 1$ and $P(false) = 0$
  - $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$
  -



Unconditional Probabilities (1)

$\boxed{P(A)}$ denotes the **unconditional** probability or **prior** probability that A will appear *in the absence of any other information*, for example:

$$\mathcal{P}(Cavity) = 0.1$$

Cavity is a proposition. We obtain prior probabilities from statistical analysis or general rules.

Unconditional Probabilities (2)

In general, a **random variable** can take on *true* and *false* values, as well as other values:

$$P(Weather{=}Sunny) = 0.7$$
$$P(Weather{=}Rain) = 0.2$$
$$P(Weather{=}Cloudy) = 0.08$$
$$P(Weather{=}Snow) = 0.02$$
$$P(Headache{=}TRUE) = 0.1$$

- Propositions can contain equations over random variables.
- Logical connectors can be used to build propositions, e.g. $P(Cavity \wedge \neg Insured) = 0.06$.

Unconditional Probabilities (3)

$\boxed{\mathbf{P}(x)}$ is the **vector of probabilities** for the (ordered) domain of the random variable X:

$$\mathbf{P}(\mathit{Headache}) = \langle 0.1, 0.9 \rangle$$
$$\mathbf{P}(\mathit{Weather}) = \langle 0.7, 0.2, 0.08, 0.02 \rangle$$

define the probability distribution for the random variables *Headache* and *Weather.*

P*(Headache, Weather)* is a 4x2 table of probabilities of all combinations of the values of a set of random variables.

	Headache = TRUE	Headache = FALSE
Weather = Sunny	P(W = Sunny $\wedge$ Headache)	P(W = Sunny $\wedge$ $\neg$Headache)
Weather = Rain		
Weather = Cloudy		
Weather = Snow		

Conditional Probabilities (1)

New information can change the probability.

Example: The probability of a cavity increases if we know the patient has a toothache.

If additional information is available, we can no longer use the **prior** probabilities!

$\boxed{P(A|B)}$ is the **conditional** or **posterior** probability of A given that *all we know* is B:

$$P(Cavity \mid Toothache) = 0.8$$

$\boxed{\mathbf{P}(X|Y)}$ is the table of all conditional probabilities over all values of X and Y.

Conditional Probabilities (2)

P$(Weather \mid Headache)$ is a 4x2 table of conditional probabilities of all combinations of the values of a set of random variables.

	Headache = TRUE	Headache = FALSE
Weather = Sunny	P(W = Sunny | Headache)	P(W = Sunny | ¬Headache)
Weather = Rain		
Weather = Cloudy		
Weather = Snow		

Conditional probabilities result from unconditional probabilities (if P(B)>0) (**per definition**)

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

Conditional Probabilities (4)

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

- Product rule: $P(A \wedge B) = P(A|B)\, P(B)$
- Analog: $P(A \wedge B) = P(B|A)\, P(A)$
- A and B are **independent** if $P(A|B) = P(A)$ (equiv. $P(B|A) = P(B)$).
  Then (and only then) it holds that $P(A \wedge B) = P(A)\, P(B)$.

Why are the Axioms Reasonable?

- If P represents an *objectively observable* probability, the axioms clearly make sense.
- But why should an agent respect these axioms when it models its own degree of belief?

- *Objective vs. subjective probabilities*
  - The axioms limit the set of beliefs that an agent can maintain.

- One of the most convincing arguments for why subjective beliefs should respect the axioms was put forward by de Finetti in 1931. It is based on the connection between actions and degree of belief.
  - If the beliefs are contradictory, then the agent will fail in its environment in the long run!

The game

- Player1 gives a subjective probability “a” on the occurrence of an event “b”
- Player2 can then decide to bet either $“a” dollars against player1’s $“1-a” that b happens or $“1-a” dollars against player1 $“a” that b does happen

P(b) = 0.4

Player 2 bets $40 that “b”
player1 bets $60 that not b

Player 2 bets $60 that “b”
player1 Bets 40 dollar that not b

- Which decision is more rational?
  - To bet for b 0.4 *$60 - 0.6 *$40 = 0
  - To bet against b   0.6 *$40 - 0.4 *$60 = 0

Why are the Axioms Reasonable?

- P(a)= 0.4

  Player1 bets \$40 for a
  Player2 bets \$60 for $\neg$ a

  Player1 bets \$60 for a
  Player2 bet \$40 for $\neg$a

- P(b) = 0.3

  Player1 bets \$30 for b
  Player2 bets \$70 for $\neg$ b

  Player1 bets \$70 for $\neg$ b
  Player2 bets \$30 for b

- P($a \vee b$) = 0.8

  Player1 bets \$80 for ($a \vee b$)
  Player2 bets \$20 for $\neg(a \vee b)$

  Player1 bets \$20 for $\neg(a \vee b)$
  Player2 bets \$80 for ($a \vee b$)

Why are the Axioms Reasonable?

	a , b	Not a, b	a, not b	Not a, not b
a	6	-4	6	-4
b	7	7	-3	-3
$\neg(a \vee b)$	-2	-2	-2	8
	11	1	1	1

Inference by enumeration

- Start with the joint probability 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: $P(\varphi) = \Sigma_{\omega:\omega \models \varphi} P(\omega)$

Inference by enumeration

- Start with the joint probability 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: $P(\varphi) = \Sigma_{\omega:\omega \models \varphi} P(\omega)$

- P(*toothache*) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2

Inference by enumeration

- Start with the joint probability 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: $P(\varphi) = \Sigma_{\omega:\omega \models \varphi} P(\omega)$

- P(*toothache or cavity*) = 0.108 + 0.012 + 0.016 + 0.064 +0.072 + 0.008 = 0.28

Inference by enumeration

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

- Can also compute conditional probabilities:

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)}$$

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

Normalization

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)}$$

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)}$$

- Denominator can be viewed as a normalization constant α

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

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

Inference by enumeration, contd.

Typically, we are interested in
the posterior joint distribution of the query variables **Y**
given specific values **e** for the evidence variables **E**

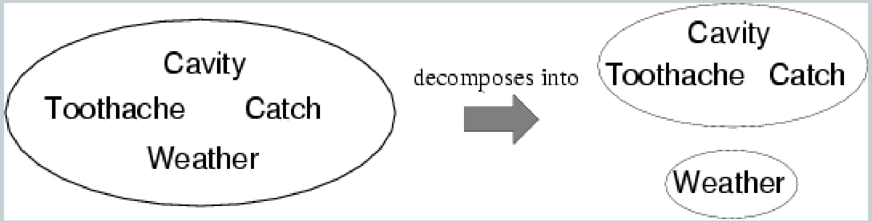
Let the hidden variables be **H** = **X** - **Y** – **E**
Then the required summation of joint entries is done by summing out the hidden variables:

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

- The terms in the summation are joint entries because **Y**, **E** and **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?

Independence

- A and B are independent iff
  $\mathbf{P}(A|B) = \mathbf{P}(A)$ or $\mathbf{P}(B|A) = \mathbf{P}(B)$ or $\mathbf{P}(\mathrm{A, B}) = \mathbf{P}(A)\, \mathbf{P}(B)$



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

- 32 entries reduced to 12; for n independent biased coins, $O(2^n) \rightarrow O(n)$
-
- Absolute independence powerful but rare
-
- Dentistry is a large field with hundreds of variables, none of which are independent. What to do?

Bayes' Rule

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

  $\Rightarrow$ Bayes' rule: $P(a \mid b) = P(b \mid a)\, P(a) / P(b)$

- or in distribution form
-

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

- Useful for assessing diagnostic probability from causal probability:
- $P(Cause|Effect) = P(Effect|Cause)\, P(Cause) / P(Effect)$
-

Bayes' Ruled

- Doesn't seem very useful
  - Requires three terms to compute one conditional
  - But useful in practice
- A doctor knows that the disease meningitis causes the patient to have a stiff neck 50% of the time. The prior that someone has meningitis is 1/50000 and the prior that someone has a stiff neck is 1/20, knowing that a person has a stiff neck what is the probability that they have meningitis?

- let M be meningitis, S be stiff neck:
- P(m|s) = P(s|m) P(m) / P(s) = (0.5   1/50000)/ (1/20)= 0.0002
  - Note: posterior probability of meningitis still very small! One in every 5000
  - Stiff neck a strong indication
- Why not just store the number?
  - Based on facts if we had epidemic we know how to update facts

Bayes' Rule and conditional independence

$\mathbf{P}(Cavity \mid toothache \wedge catch)$

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

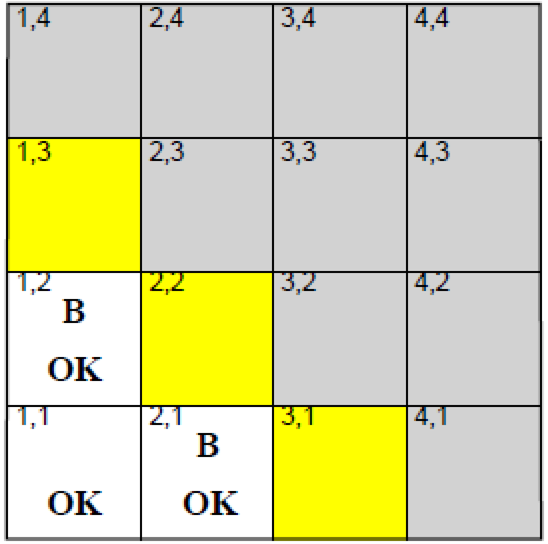
$= \alpha\mathbf{P}(toothache \wedge catch \mid Cavity)\, \mathbf{P}(Cavity)$

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

$\mathrm{P}(\mathrm{Cause} \mid \mathrm{Effect}_1, \ldots, \mathrm{Effect}_n) = \alpha\, \pi_i \mathbf{P}(\mathrm{Effect}_i \mid \mathrm{Cause})\, \mathbf{P}(\mathrm{Cause})$

- This is an example of a naïve Bayes model:
- $\mathbf{P}(\mathrm{Cause}, \mathrm{Effect}_1, \ldots, \mathrm{Effect}_n) = \mathbf{P}(\mathrm{Cause})\, \pi_i \mathbf{P}(\mathrm{Effect}_i \mid \mathrm{Cause})$
- Total number of parameters is linear in *n*

Wumpus World

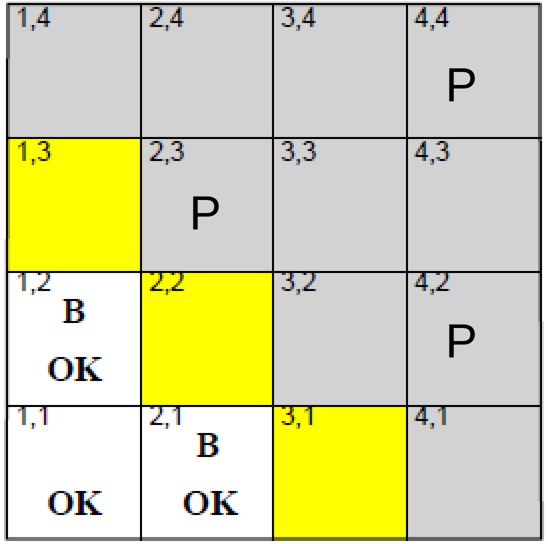


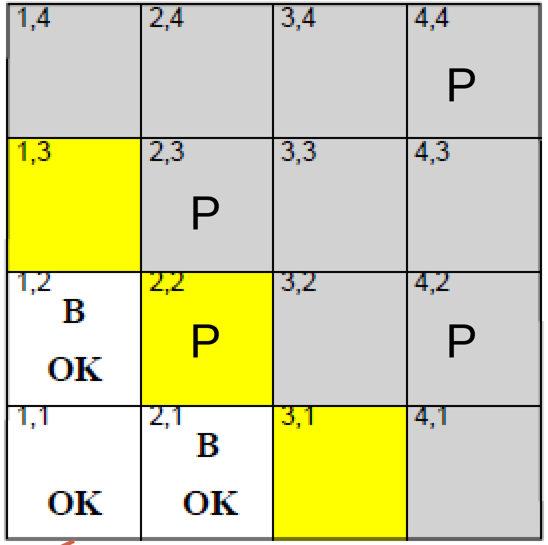
Let's define the random variables first

- P_{ij} = true if [i,j] contains a pit
- B_{ij} = true if [i,j] is breezy

- We want to be able to predict the probability of possbile boards

Specifying the 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: $\mathbf{P}(B_{1,1}, B_{1,2}, B_{2,1} \mid P_{1,1}, \ldots, P_{4,4})\mathbf{P}(P_{1,1}, \ldots, P_{4,4})$

$\mathbf{P}(B_{1,1}, B_{1,2}, B_{2,1} \mid P_{1,1}, \ldots, P_{4,4})$ =0

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

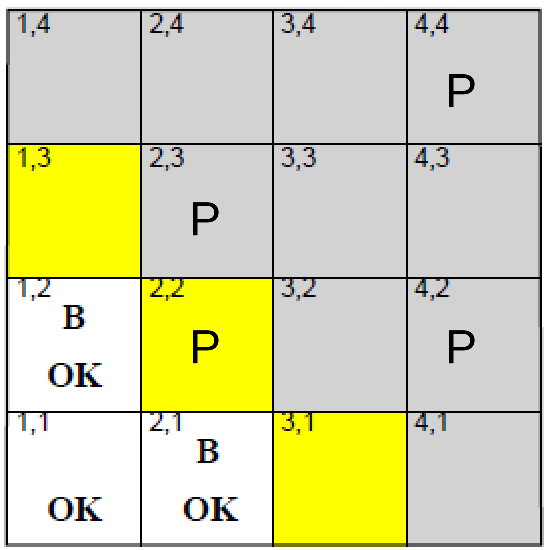
The full joint distribution is $\mathbf{P}(P_{1,1}, \ldots, P_{4,4}, B_{1,1}, B_{1,2}, B_{2,1})$

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

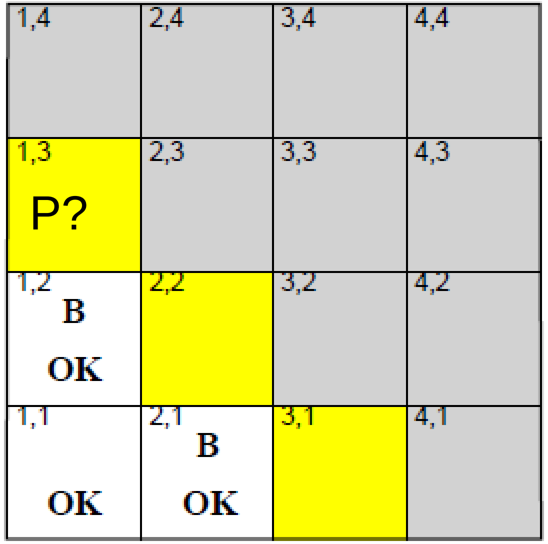
Second term: pits are placed randomly, probability 0.2 per square:

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

for n pits



$= 0.2^4 \times 0.8^{12}$



We know the following facts:

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

$$known = \neg p_{1,1} \wedge \neg p_{1,2} \wedge \neg p_{2,1}$$

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

For inference by enumeration, we have

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

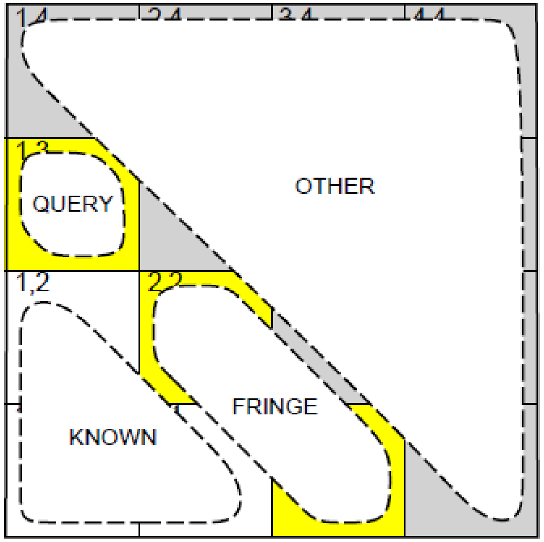
Done?

Unknown has 12 squares. $2^{12} = 4096$

Is P_{13} really related to P_{44}?

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!

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

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

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

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

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

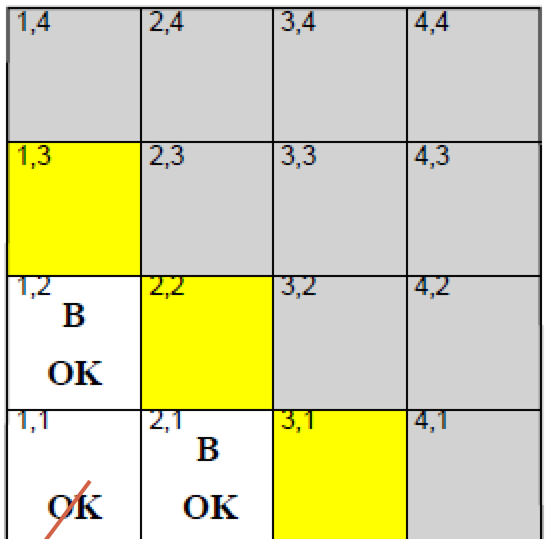
$$= \alpha \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe) \sum_{other} \mathbf{P}(P_{1,3}, known, fringe, other)$$

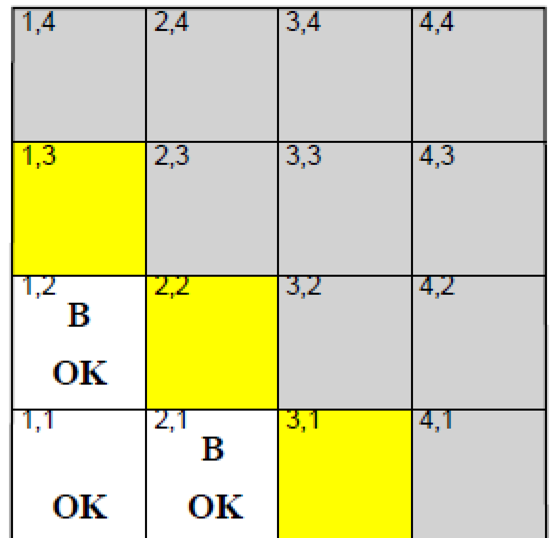
$$= \alpha \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe) \sum_{other} \mathbf{P}(P_{1,3})P(known)P(fringe)P(other)$$

$$= \alpha \, P(known)\mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe)P(fringe) \cancelto{1}{\sum_{other} P(other)}$$

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

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

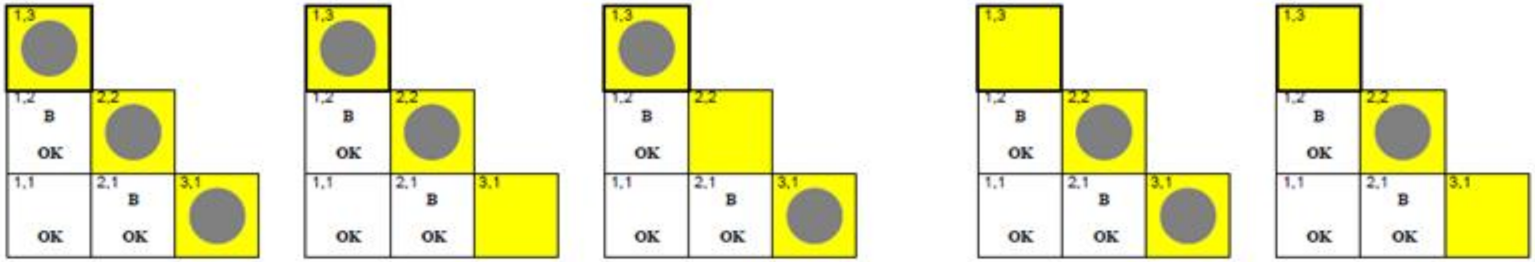




$\mathbf{P}(B_{1,1}, B_{1,2}, B_{2,1} \mid P_{1,1}, \ldots, P_{4,4})$ =0

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

Using conditional independence contd.



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

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

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

$$\begin{aligned}\mathbf{P}(P_{2,2}|known, b) &= \alpha' \langle 0.2(0.04 + 0.16 + 0.16 + 0.64),\ 0.8(0.04)\rangle \\ \mathbf{P}(P_{2,2}|known, b) &\approx \langle 0.86, 0.14\rangle\end{aligned}$$

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
-