Uncertainty

Larry Holder
School of EECS
Washington State University

Uncertainty

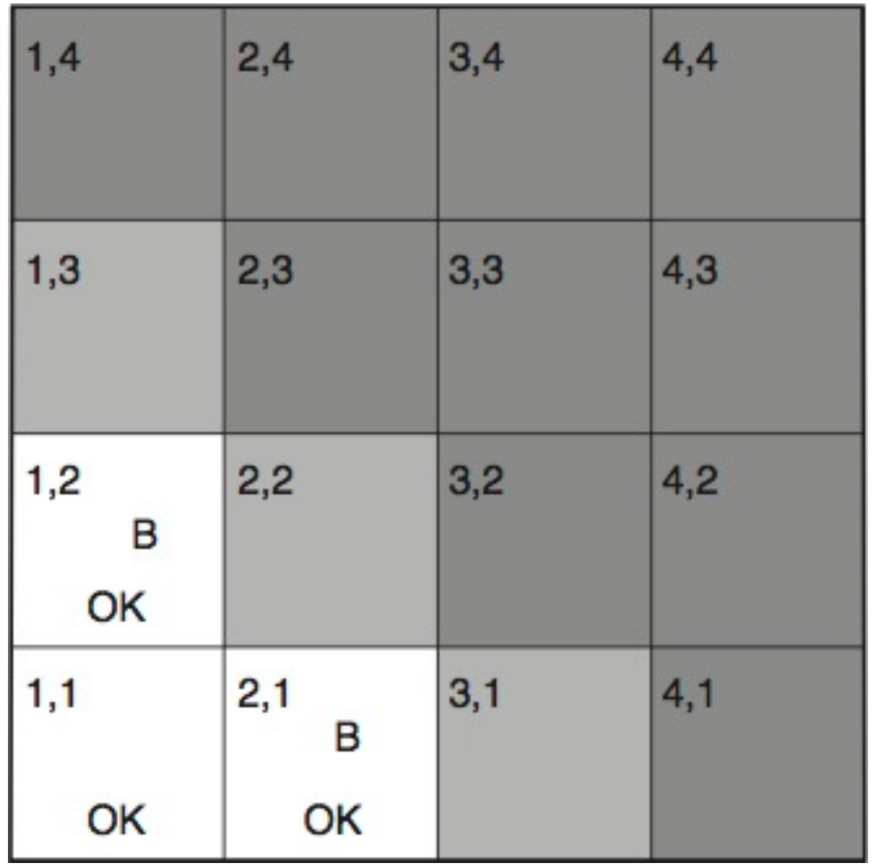
- Sometimes the truth or falsity of facts in the world is unknown
- Sources of agent’s uncertainty
  - Incompleteness of rules
  - Incorrectness of rules
  - Limited and ambiguous sensors
  - Imperfection and noise of actions
  - Unpredictable and dynamic nature of environment
  - Approximate nature of internal models and algorithms

Environment Properties

- Fully observable vs. partially observable
- Deterministic vs. stochastic
- Episodic vs. sequential
- Static vs. dynamic
- Discrete vs. continuous
- Single agent vs. multiagent

Wumpus World

- $P(Pit_{1,3}) = ?$
- $P(Pit_{2,2}) = ?$
- $P(Pit_{3,1}) = ?$



Rational Agent Approach

- Choose action A that maximizes expected utility
- I.e., maximizes Prob(A) * Utility(A)
- Prob(A) = probability A will succeed
- Utility(A) = value to agent of A's outcomes

$$\text{action} = \max_{a} P(a) * U(a)$$

Probability

- A probability model associates a numeric probability P(w) with each possible world w, where

$$0 \le P(w) \le 1 \quad and \quad \sum_{w} P(w) = 1$$

Probability

- The probability P(a) of a proposition 'a' is the sum of the probabilities of the worlds in which 'a' is true

$$P(a) = \sum_{w \in a} P(w)$$

Example

- What is the probability of the proposition that a two-dice roll totals 8?
- Consider the 36 possible outcomes of rolling two dice, each with probability 1/36
  - (1, 1) (1, 2) (1, 3) (1, 4) (1, 5) (1, 6)
  - (2, 1) (2, 2) (2, 3) (2, 4) (2, 5) (2, 6)
  - (3, 1) (3, 2) (3, 3) (3, 4) (3, 5) (3, 6)
  - (4, 1) (4, 2) (4, 3) (4, 4) (4, 5) (4, 6)
  - (5, 1) (5, 2) (5, 3) (5, 4) (5, 5) (5, 6)
  - (6, 1) (6, 2) (6, 3) (6, 4) (6, 5) (6, 6)
- $P(Total=8) = P(Die_1=2 \wedge Die_2=6) + P(Die_1=3 \wedge Die_2=5) + \ldots + P(Die_1=6 \wedge Die_2=2) = ?$

Probability

- An unconditional or prior probability is the degree of belief without any other information
  - E.g., P(Total=8)
- A conditional or posterior probability is the degree of belief given some evidence
  - E.g., given that the first die is a 2, what is the probability that the total will be 8?
  - $P(Total=8 \mid Die_1=2) = ?$
  - $P(Total=8 \mid Die_1=1) = ?$

Conditional Probability

$$P(a \mid b) = \frac{P(a \wedge b)}{P(b)} \quad assumes \quad P(b) > 0$$

- Fraction of worlds in which ‘a’ and ‘b’ are true out of the worlds in which ‘b’ is true
- $P(Total=8 \mid Die_1=2)$
  $= P(Total=8 \wedge Die_1=2) / P(Die_1=2) = ?$

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

Probability Theory

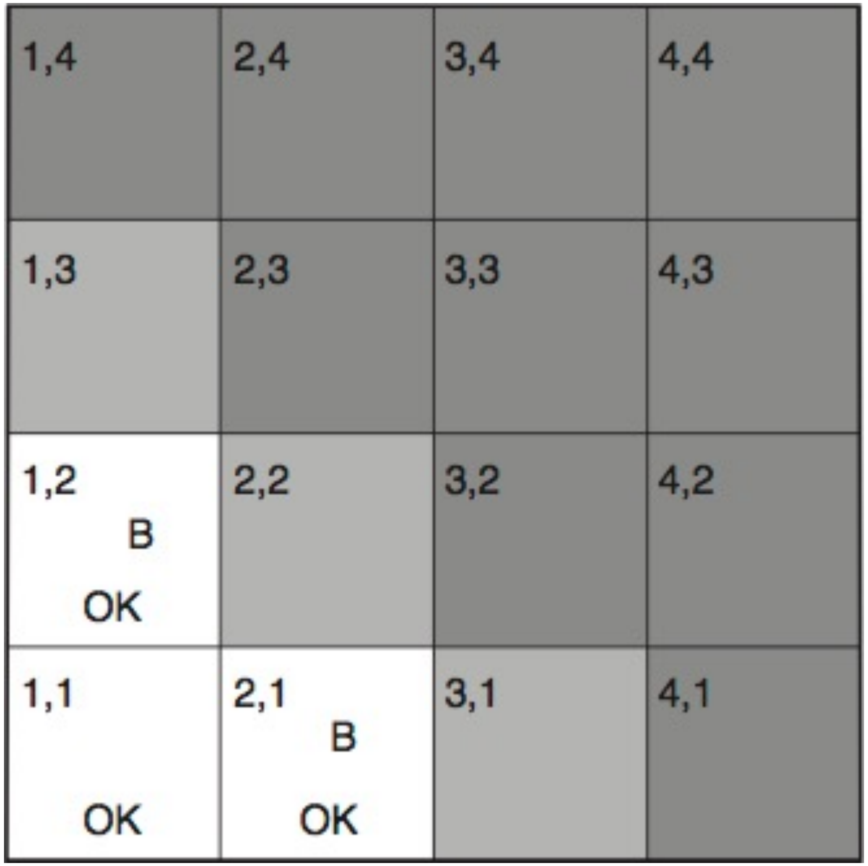
- Random variable: a variable in probability theory
  - E.g., Total, Die_1, Die_2
- Domain: set of possible values for a random variable
  - E.g., domain(Die_1) = {1,2,3,4,5,6}
- Probability distribution: probability of each value in a random variable's domain
  - E.g., $P(Die_1=1)=1/6$, …, $P(Die_1=6)=1/6$
  - Or, $\mathbf{P}(Die_1)=\langle 1/6, 1/6, 1/6, 1/6, 1/6, 1/6\rangle$

Note: Boldfaced

Probability Theory

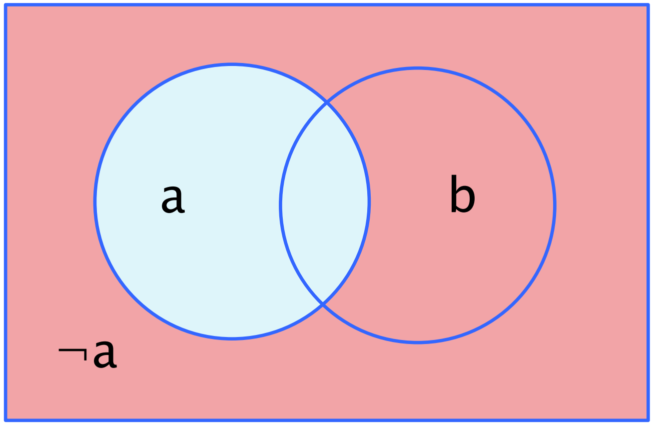
- Conditional probability distribution
  - **P**(X | Y) gives the values of $P(X=x_i \mid Y=y_j)$ for all possible i,j pairs
- Joint probability distribution
  - **P**(X,Y) denotes the probabilities of all possible combinations of X and Y
  - E.g., **P**(Die_1,Die_2) gives values for P(Die_1=1,Die_2=1), P(Die_1=1,Die_2=2), …, P(Die_1=6,Die_2=6)
  - E.g., P(Die_1=5,Die_2=1) abbreviated P(5,1)
- Full joint probability distribution
  - Joint probability distribution over all random variables in the world

Wumpus World



More Probability Axioms

- $P(\neg a) = 1 - P(a)$
- $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$



Probabilistic Inference

- Approach #1: Just ask Spock…

"Errand of Mercy" (1967)

Probabilistic Inference

- Given full joint probability distribution, we can answer any question about probabilities
- Example: Tooth World
  - Random variables: Toothache, Cavity, Catch
    - Domain: true, false

	Toothache = true		Toothache = false	
	Catch = true	Catch = false	Catch = true	Catch = false
Cavity = true	.108	.012	.072	.008
Cavity = false	.016	.064	.144	.576

Probabilistic Inference

- Answer questions (perform probabilistic inference) by summing probabilities
- Examples
  - P(toothache) = ?
  - P(cavity ∧ toothache) = ?
  - P(cavity | toothache) = ?

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

Note: For Boolean random variables, write X=true as x, and X=false as ¬x.

Probabilistic Inference

- Marginalization is the process of finding the probability distribution over a subset of variables Y:

$$P(Y) = \sum_{z \in Z} P(Y, z)$$

Z is all possible combinations of values

  - For example:

$$P(Cavity) = \sum_{z \in \{Catch, Toothache\}} P(Cavity, z)$$

z in {(catch, tooth), (catch, ¬tooth), (¬catch, tooth), (¬catch, ¬tooth)}

- Conditioning:

$$P(Y) = \sum_{z \in Z} P(Y \mid z) P(z)$$

Probabilistic Inference

- Conditional probabilities

$$
\begin{aligned}
P(cavity \mid toothache) &= \frac{P(cavity \wedge toothache)}{P(toothache)} \\
&= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6
\end{aligned}
$$

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

- Note that denominator is the same in both (normalization constant)

Probabilistic Inference

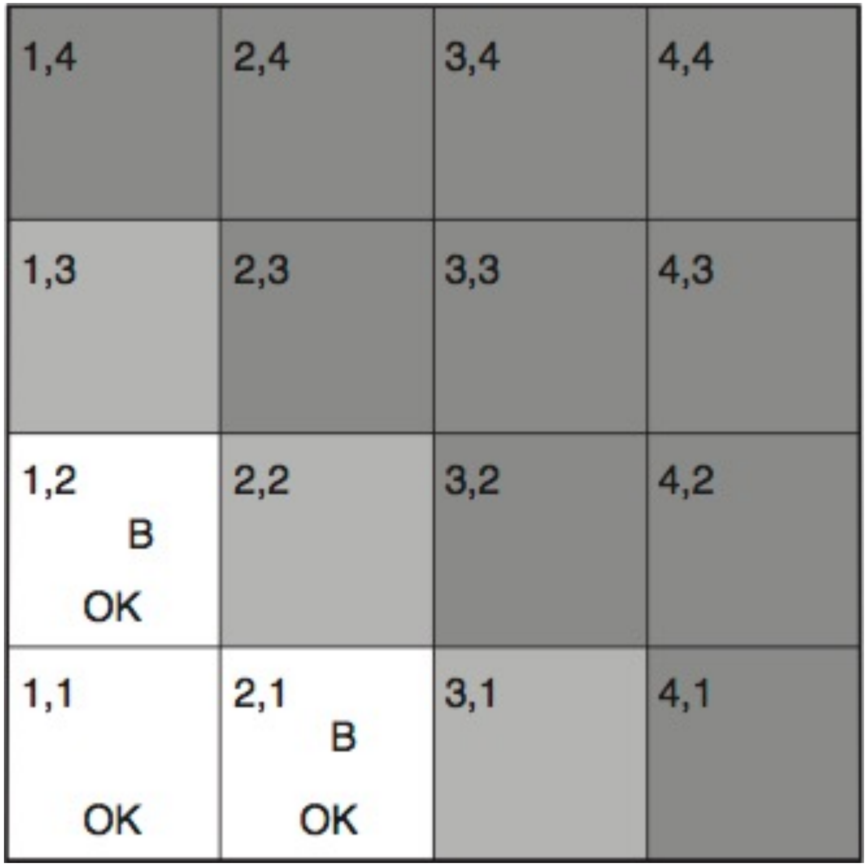
- Don't need normalization constant to compute probability distribution
  - For example
    - $P(cavity \mid toothache) = \alpha P(cavity \wedge toothache) = \alpha 0.12$
    - $P(\neg cavity \mid toothache) = \alpha P(\neg cavity \wedge toothache) = \alpha 0.08$
    - Since these have to sum to 1, $\alpha = 1/0.2 = 5$
    - Also means $P(toothache) = 1/5 = 0.2$, but we didn't need to know this
    - Also means we can determine which is more likely (cavity or ¬cavity) without knowing P(toothache)

Probabilistic Inference

- General rule
- Want to know the probability distribution over X given observed variables **e** (evidence) and unobserved variables **y**

$$\boldsymbol{P}(X \mid \boldsymbol{e}) = \alpha \boldsymbol{P}(X, \boldsymbol{e}) = \alpha \sum_{\boldsymbol{y}} \boldsymbol{P}(X, \boldsymbol{e}, \boldsymbol{y})$$

Wumpus World



Independence

- Full joint probability distributions are typically infeasible to write down
  - E.g., 2^n table entries for n Boolean variables
- If we know some variables are independent of others, then we can decompose the full table into smaller tables
- If two variables X and Y are independent, then
  - $P(X,Y) = P(X)P(Y)$
  - $P(X|Y) = P(X)$
  - $P(Y|X) = P(Y)$

Independence

- For example, add Weather variable to Tooth World
  - domain(Weather) = {sunny, cloudy, rain, snow}
  - 8 entry table now a 32 entry table
- Assuming your teeth don't affect the weather
  - I.e., Weather is independent of Toothache, Catch and Cavity
  - So, **P**(Weather, Toothache, Catch, Cavity) =
    P(Weather) * **P**(Toothache, Catch, Cavity)
  - E.g., P(Weather=cloudy, toothache, catch, cavity) =
    P(Weather=cloudy) * P(toothache, catch, cavity)
- Full joint distribution described by two tables of 8 and 4 entries

Bayes Rule

Thomas Bayes
(1701–1761)

$$P(b \mid a) = \frac{P(a \mid b)P(b)}{P(a)} \qquad \boldsymbol{P}(Y \mid X) = \frac{\boldsymbol{P}(X \mid Y)\boldsymbol{P}(Y)}{\boldsymbol{P}(X)}$$

- Foundational rule of probabilistic reasoning
- Turns a <u>diagnostic</u> question into a <u>causal</u> one

$$P(cause \mid effect) = \frac{P(effect \mid cause)P(cause)}{P(effect)}$$

- In general, given some evidence **e**:

$$\boldsymbol{P}(Y \mid X, \boldsymbol{e}) = \frac{\boldsymbol{P}(X \mid Y, \boldsymbol{e})\boldsymbol{P}(Y \mid \boldsymbol{e})}{\boldsymbol{P}(X \mid \boldsymbol{e})}$$

Bayes Rule

- Example: Diagnosing cancer
- Given
  - 1% of population has cancer
  - Test has 20% false positive rate
  - Test has 10% false negative rate
  - Patient tests positive for cancer
  - What is the probability that patient has cancer?

Bayes Rule

- Example: Diagnosing cancer
- Variables
  - Diagnosis $\in$ {cancer,healthy)
  - Test $\in$ {pos,neg}
- Given
  - P(cancer)=0.01, P(healthy)=0.99
  - P(neg|cancer)=0.1, P(pos|cancer)=0.9
  - P(pos|healthy)=0.2, P(neg|healthy)=0.8
- P(cancer|pos)=?
- P(cancer|pos) = P(cancer $\wedge$ pos) / P(pos) = ?

Bayes Rule

- Applying Bayes rule…

$$P(cancer \mid pos) = \frac{P(pos \mid cancer)P(cancer)}{P(pos)}$$

  - We know P(pos|cancer)=0.9 and P(cancer)=0.01
  - Can compute P(pos) by marginalization
  - P(pos) = ?

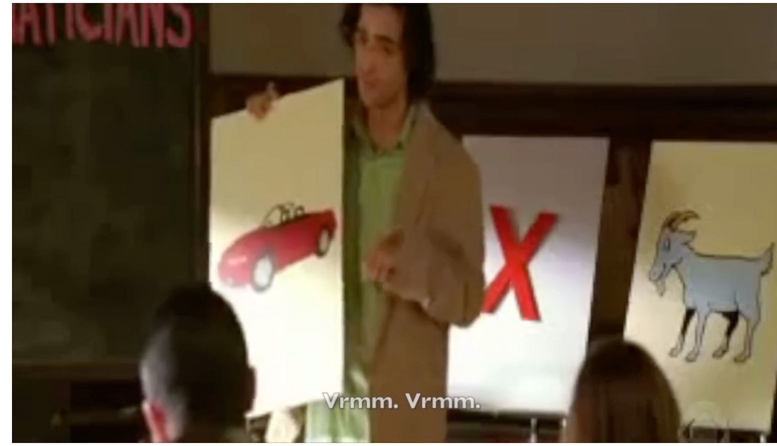
  - P(cancer | pos) = ?

Bayes Rule

- Could also compute P(pos) via normalization
- $P(cancer|pos) = \alpha P(pos|cancer)P(cancer) =$
- $P(healthy|pos) = \alpha P(pos|healthy)P(healthy) =$
- $\alpha =$
- $P(pos) = 1/\alpha =$
- Many times, P(effect|cause) is easier to determine than P(cause|effect)

Monty Hall Problem

- Ci = car behind Door i
- Oi = opens Door i
- Pick Door 2
- Opens Door 3



"Numb3rs" S1E13, 2005.

- Should you change your guess?
- $P(C1 \mid O3) = ?$

Combining Evidence in Bayes Rule

- How to compute $P(a \mid b \wedge c \wedge \ldots)$?
  - E.g., $\mathbf{P}(Cavity \mid toothache \wedge catch) = ?$
- Easy if we have full joint probability distribution
  - E.g., $\mathbf{P}(Cavity \mid toothache \wedge catch) = \alpha\langle P(cavity \wedge toothache \wedge catch), P(\neg cavity \wedge toothache \wedge catch)\rangle = ?$
- Or, using Bayes rule: $\alpha[P(b \wedge c \wedge \ldots \mid a)\, P(a)]$
  - E.g., $\alpha\ \mathbf{P}(toothache \wedge catch \mid Cavity)\ \mathbf{P}(Cavity)$
  - Still need to know many probabilities

Combining Evidence in Bayes Rule

- If b, c, … are "caused" by 'a', but not by each other, then
  - $P(b \wedge c \wedge \ldots \mid a) = P(b \mid a)\, P(c \mid a) \ldots$
  - $P(a \mid b \wedge c \wedge \ldots) = \alpha\, P(b \wedge c \wedge \ldots \mid a)\, P(a)$
  - $P(a \mid b \wedge c \wedge \ldots) = \alpha\, P(b \mid a)\, P(c \mid a) \ldots P(a)$
- I.e., b, c, … are independent given 'a'
- Two propositions (effects) are <u>conditionally independent</u> if they are independent given a third proposition (cause)
- Now we only need to know the probabilities of individual effects given the cause P(b | a) and the prior probability of the cause P(a)

Combining Evidence in Bayes Rule

- Example (assuming toothache and catch are conditionally independent given Cavity)
  - $\mathbf{P}(toothache \wedge catch \mid Cavity) = \mathbf{P}(toothache \mid Cavity)\, \mathbf{P}(catch \mid Cavity)$

- Using this in Bayes rule
  - $\mathbf{P}(Cavity \mid toothache \wedge catch) = \alpha\, \mathbf{P}(toothache \wedge catch \mid Cavity)\, \mathbf{P}(Cavity)$
  - $\mathbf{P}(Cavity \mid toothache \wedge catch) = \alpha\, \mathbf{P}(toothache \mid Cavity)\, \mathbf{P}(catch \mid Cavity)\, \mathbf{P}(Cavity)$

Combining Evidence in Bayes Rule

- In general, if random variables X and Y are conditionally independent given Z
  - $\mathbf{P}(X, Y \mid Z) = \mathbf{P}(X \mid Z)\, \mathbf{P}(Y \mid Z)$
  - $\mathbf{P}(X \mid Y,Z) = \mathbf{P}(X \mid Z)$
  - $\mathbf{P}(Y \mid X,Z) = \mathbf{P}(Y \mid Z)$

Naïve Bayes

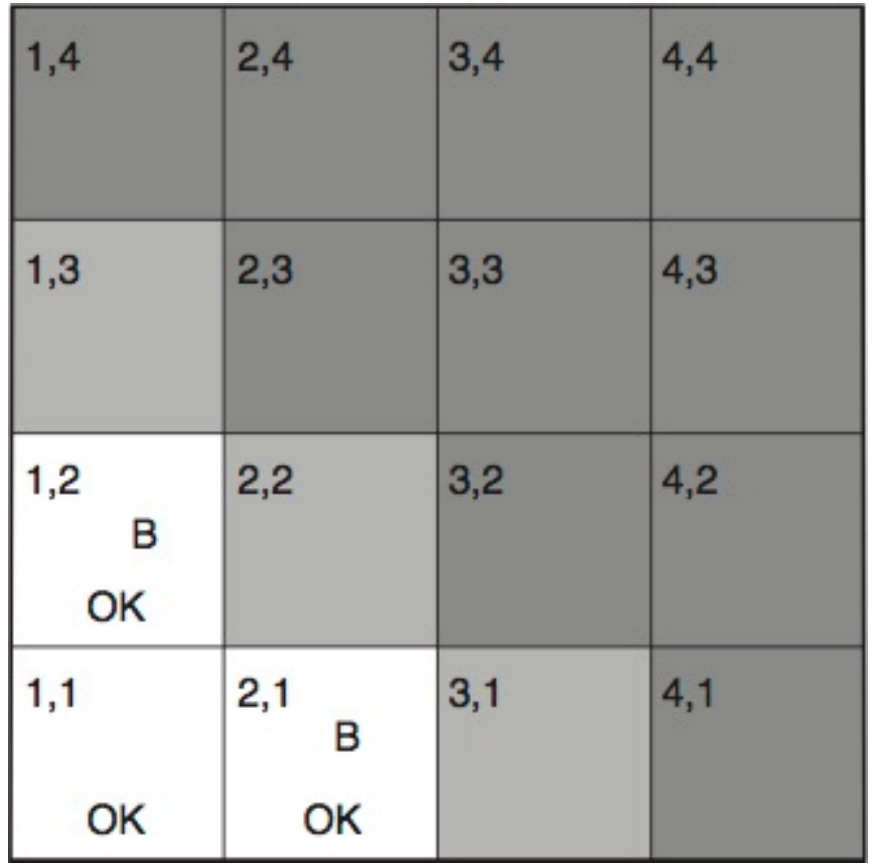
- Want **P**(Cause | $\text{Effect}_1,\text{Effect}_2,\ldots,\text{Effect}_n$)
- Assume $\text{Effect}_1,\text{Effect}_2,\ldots,\text{Effect}_n$ are conditionally independent given Cause
- Applying Bayes Rule
  - **P**(Cause | $\text{Effect}_1,\text{Effect}_2,\ldots,\text{Effect}_n$) = α**P**($\text{Effect}_1,\text{Effect}_2,\ldots,\text{Effect}_n$ | Cause) **P**(Cause) = α**P**(Cause) $\prod_i$ **P**(Effect_i | Cause)

Naïve Bayes

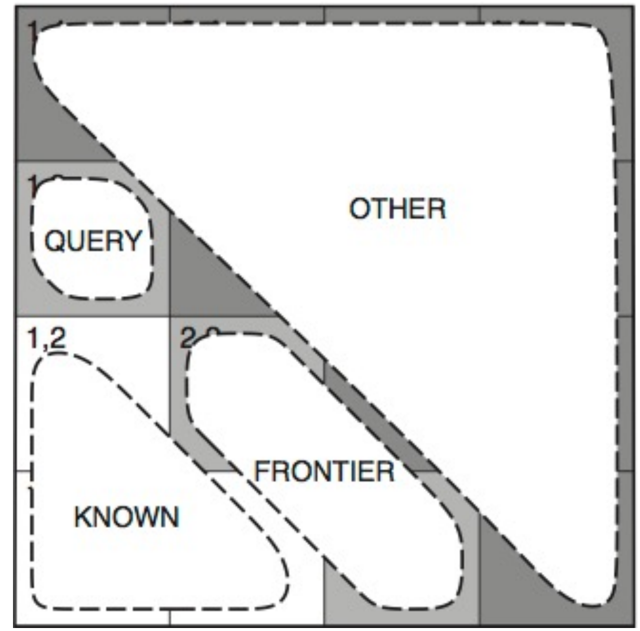
- Be careful
  - Conditional independence rarely absolutely true
    - E.g., Cold weather makes my teeth hurt
  - Probabilities estimated from data

Wumpus World Revisited

- $P(Pit_{1,3}) = ?$
- $P(Pit_{2,2}) = ?$
- $P(Pit_{3,1}) = ?$

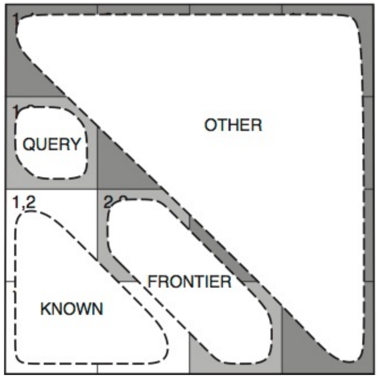


Wumpus World Revisited

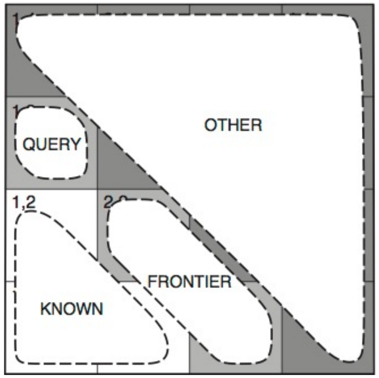


- "query" = $\text{Pit}_{1,3}$
- "frontier" = $\{\text{Pit}_{2,2}\ ,\ \text{Pit}_{3,1}\}$
- "other" = other 10 pit variables
- "known" = $\neg\text{pit}_{1,1} \wedge \neg\text{pit}_{1,2} \wedge \neg\text{pit}_{2,1}$
- "breeze" = $\neg\text{breeze}_{1,1} \wedge \text{breeze}_{1,2} \wedge \text{breeze}_{2,1}$
- Note: "breeze" is conditionally independent of "other" given "known", "frontier" and "query"

Wumpus World Revisited

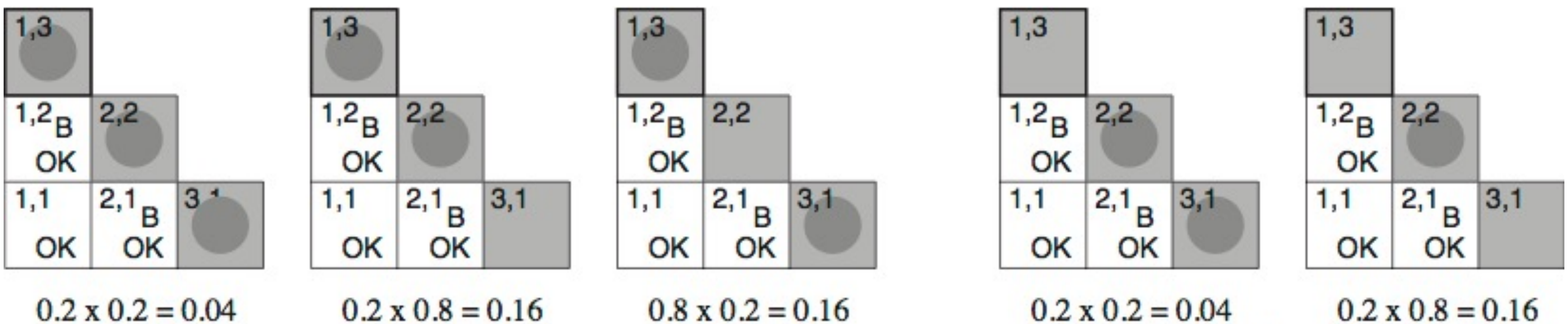


Wumpus World Revisited



Wumpus World Revisited

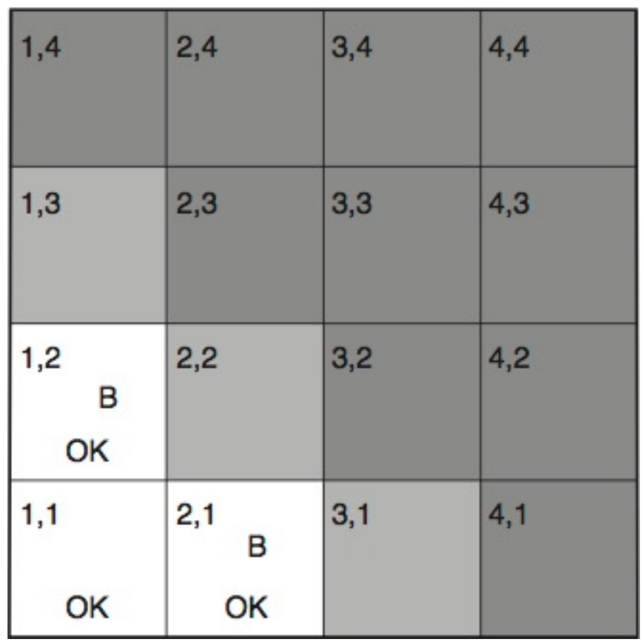
- And given independence of pits $P_{i,j}$
- $\mathbf{P}(P_{1,3} \mid known, breeze) =$
  $\alpha\, \mathbf{P}(P_{1,3})\, \Sigma_{frontier}\, \mathbf{P}(breeze \mid known, P_{1,3}, frontier)\ P(frontier) =$
  $\alpha\langle 0.2(0.04+0.16+0.16),\ 0.8(0.04+0.16)\rangle = \langle 0.31, 0.69\rangle$



$P_{1,3}$=true $P_{1,3}$=false

Wumpus World Revisited

- So, $P(P_{1,3}=true) = P(P_{3,1}=true) = 0.31$
- $P(P_{2,2}=true) = 0.86$
- Probabilistic agent "knows more" than logical agent



Summary: Uncertainty

- Probability theory allows us to reason about uncertainty
- Bayes rule allows us to change diagnostic questions to causal questions
- Conditional independence allows us to simplify complexity
- Probabilistic agent can outperform logical agent in partially observable and stochastic environments