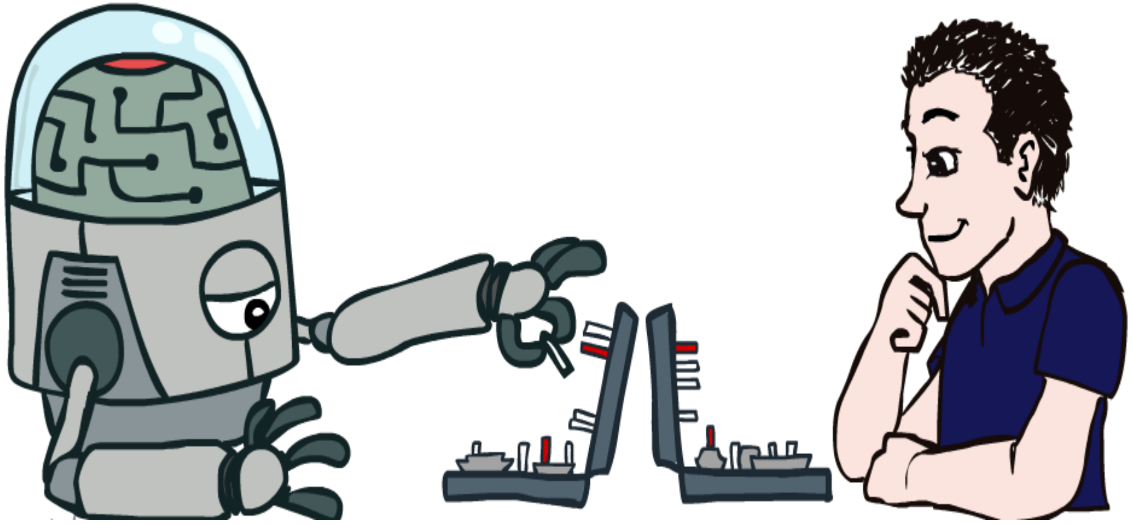
ARTIFICIAL INTELLIGENCE- CS411

Dr. Abdulelah Algosaibi



Credit: Dr. Alaa Sagheer

Acknowledgement

These slides were assembled by **Credit: Prof. Alaa Sagheer**, with grateful acknowledgement of the many other professors who made their course materials freely available online.

7. Uncertain Knowledge & Reasoning

Uncertainty

In which we see what agent should do when not all is crystal clear!

Lecture I

Introduction

- The world is not a well-defined place.

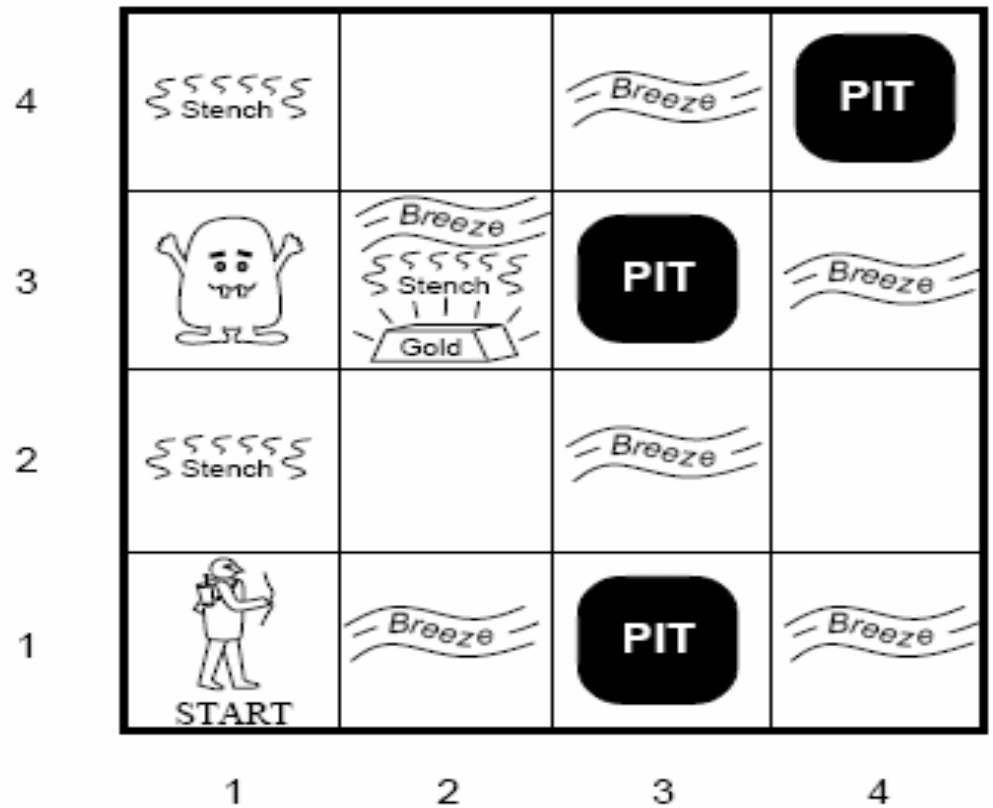
Subjection.

- There is uncertainty in the facts we know:
  - What’s the temperature? Imprecise measures
  - Is Trump a good president? Imprecise definitions
  - Where is the pit? Imprecise knowledge
- There is uncertainty in our inferences
  - If I have a blistery, itchy rash and was gardening all weekend I probably have poison ivy
- People make successful decisions all the time anyhow.

Uncertain Agent

- When an agent knows enough facts about its environment, the logical approach enables it to derive plans that are guaranteed to work.
- Unfortunately, agents almost never have access to the whole truth about their environment.
- Example: Wumpus!

Agents must, therefore, act under Uncertainty *



Sources of Uncertainty

what are the sources of uncertainty.

1 Uncertain data

- missing data, unreliable, ambiguous, imprecise representation, inconsistent, subjective, derived from defaults, noisy…

2 Uncertain knowledge

- Multiple causes lead to multiple effects
- Incomplete knowledge of causality in the domain
- Probabilistic/stochastic effects

3 Uncertain knowledge representation

- restricted model of the real system
- limited expressiveness of the representation mechanism

4 inference process

- Derived result is formally correct, but wrong in the real world
- New conclusions are not well-founded (eg, inductive reasoning)
- Incomplete, default reasoning methods

Where Uncertainty can be?

Knowlege

1. Uncertainty in prior knowledge

E.g., some causes of a disease are unknown and are not represented in the background knowledge of a medical-assistant agent.

Actions

2. Uncertainty in actions

E.g., actions are represented with relatively short lists of preconditions, while these lists are in fact arbitrary long

perception

3. Uncertainty in perception

E.g., sensors do not return exact or complete information about the world; a robot never knows exactly its position.

Example-1 of Uncertainty

For example, to drive my car in the morning:

- It must not have been stolen during the night
- It must not have flat tires
- There must be gas in the tank
- The battery must not be dead
- The ignition must work
- I must not have lost the car keys
- No truck should obstruct the driveway
- I must not have suddenly become blind or paralytic

etc…

Not only would it not be possible to list all of them, but would trying to do so be efficient? No.

Example-2 of Uncertainty

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, noisy sensors)
2. uncertainty in action outcomes (flat tire, etc.)
3. 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 …)

Definition of Uncertainty

What we call **uncertainty** is a summary of all that is not explicitly taken into account in the agent's KB

Basically, What is not defined in the KBs

Uncertain Agent

Can LOGIC or INFERENCE

solves UNCERTAINTY?

Uncertain Agent

Can Conditional PLANNING solves UNCERTAINTY?

Handling Uncertain Knowledge

Example

Diagnosis: (for medicine, automobile repair, etc..) is a task that involves uncertainty.

For *p* *Symptom (p, Toothache)* ➔ *Disease (p, Cavity) **

For *p* *Symptom (p, Toothache)* ➔ *Disease (p, Cavity) V Disease (p, GumDisease) V Disease (p, Abscess)*

For *p* *Disease (p, Cavity)* ➔ *Symptom (p, Toothache) ***

1. Theoretical/Practical Ignorance
2. Laziness (efficiency?)

Handling Uncertain Knowledge

Approaches: What are the approaches? to handle uncertain knowledge.

1. Default reasoning
2. Worst-case reasoning
3. Probabilistic reasoning

Default Reasoning

- The world is fairly normal. Abnormalities are rare,
- So, an agent assumes normality, until there is evidence of the contrary,
- E.g., if an agent sees a bird x, it assumes that x can fly, unless it has evidence that x is a penguin, an ostrich, a dead bird, a bird with broken wings, …

Worst Case Reasoning

That's me!

- Just the opposite! The world is ruled by Murphy's Law

(*Anything that can go wrong, will go wrong*)

- The agent assumes the worst case, and chooses the actions that maximizes a utility function in this case.
- Example:

Adversarial Search

Probabilistic Reasoning

- The world is not divided between "normal" and "abnormal", nor it is adversarial. Possible situations have various likelihoods (probabilities)
- Probability provides a way of **summarizing** the uncertainty that comes from our laziness and ignorance *.
- The agent has probabilistic beliefs – pieces of knowledge with associated probabilities (strengths) – and chooses its actions to maximize the expected value of some utility function

Probability Theory

- Assigning a probability of 0 to a given sentence corresponds to an unequivocal (no doubt) belief that the sentence is false,
- While assigning a probability of 1 corresponds to an unequivocal belief that the sentence is true. → degree of correctness
- Probabilities between 0 and **1** correspond to intermediate degrees of belief in the truth of the sentence *

A degree of belief is different from a degree of truth

- All probability statements must therefore indicate the evidence with respect to which the probability is being assessed.
- Before the evidence is obtained, we talk about prior or unconditional probability; after the evidence is obtained, we talk about posterior or conditional probability

Uncertainty and Rational Decisions

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 → closest to 1

- Which action to choose?

  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

Uncertainty and Rational Decisions

- The presence of uncertainty changes the way an agent makes decisions.
- To make decision, an agent must first have **preferences** between the different possible **outcomes** of the various plans*.

$$\underline{\textit{Decision theory}} = \underline{\textit{Probability theory} + \textit{Utility theory}}$$

- An agent is rational if and only if it chooses the action that yields the highest expected utility, averaged over all the possible outcomes of the action. **(Maximum Expected Utility MEU)**

Probability Notations

- **Proposition**: Degrees of belief are always applied to propositions (assertions that such-and-such is the case)
- **Random variable**: referring to a "part" of the world whose "status" is initially unknown
- **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:

$Cavity = false \wedge Toothache = false$
$Cavity = false \wedge Toothache = true$
$Cavity = true \wedge Toothache = false$
$Cavity = true \wedge Toothache = true$

Probability Notations*

- **Prior** (or **unconditional) probability** for a proposition "*a*" is the degree of belief accorded to it in the absence of any other information; written *P(a)*

- **Conditional** (or **posterior) probability** $P(a \mid b)$, *a* and *b* are propositions
  - A prior probability, such as *P(cavity),* can be thought of as a special case of the conditional probability *P(cavity | ??*), where the probability is conditioned on no evidence.
  - Conditional probabilities: in terms of unconditional probabilities

$$P(a \wedge b) = P(a \mid b)\, P(b)$$

Product rule: for *a* and *b* to be true, we need *b* to be true, and we also need *a* to be true given *b.*

Probability Notations

- Sometimes, we talk about the probabilities of all the possible values of a random variable. In that case, we use an expression such as **P(** *Weather),* which denotes a ***vector*** of values f for the probabilities of each individual state of the weather. Instead of writing the four equations

$P(Weather = sunny) = 0.7$ $\quad$ $P(Weather = rain) = 0.2$

$P(Weather = cloudy) = 0.08$ $\quad$ $P(Weather = snow) = 0.02$

We may simply write:

$$\mathbf{P}(Weather) = (0.7, 0.2, 0.08, 0.02)$$

This statement defines a prior **Probability Distribution.**

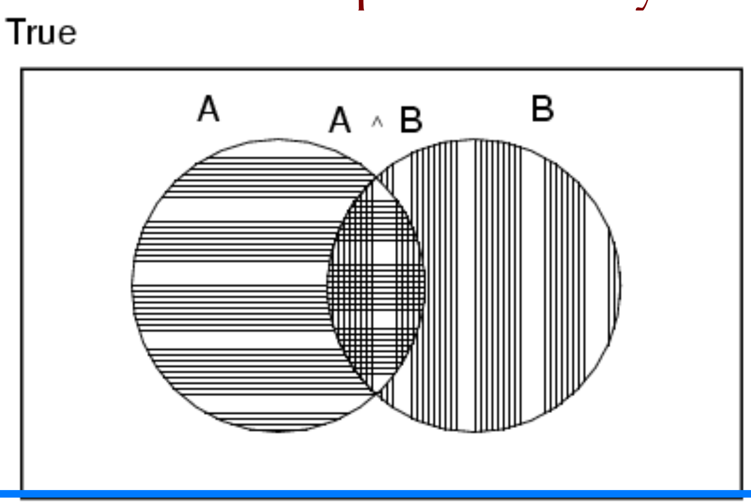
- Conditional distributions. $\mathbf{P}(X \mid Y)$ gives the values of $P(X = x_i \mid Y = y_j)$ for each possible i, j.

Also, we have $\quad$ $\mathbf{P}(X, Y) = \mathbf{P}(X \mid Y)\,\mathbf{P}(Y)$.

Axioms of Probability

- The probability of a proposition A is a real number P(A) between 0 and 1
- P(True) = 1 and P(False) = 0
- $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

Axioms of probability



$P(A \vee \neg A) = P(A)+P(\neg A)-P(A \wedge \neg A)$

$P(\text{True}) = P(A)+P(\neg A)-P(\text{False})$

$1 = P(A) + P(\neg A)$

So:

$P(A) = 1 - P(\neg A)$

Uncertainty

- Inference using Full Joint Distributions
- Bayes Rule and its use

Lecture II

Dentist's Terms

- *Toothache*: boolean variable indicating whether the patient has a toothache
- *Cavity*: boolean variable indicating whether the patient has a cavity
- *Catch*: whether the dentist's probe catches in the cavity

dentist's probe

Review

Prior Probability

Prior or unconditional probabilities of propositions

شدخل ؟

e.g., P(*Cavity* = true) = 0.1 and P(*Weather* = sunny) = 0.72 correspond to belief prior to arrival of any (new) evidence

- Probability distribution gives values for all possible assignments:

P(*Weather*) = <0.72,0.1,0.08,0.1> (normalized, i.e., sums to 1)

Review

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: **P**(*Cavity* | *Toothache*) = 2-element vector of 2-element vectors)
- If we know more, e.g., *catch or Gum disease* is also given, then we have

  P(*cavity* / *toothache, catch*) = 1

  P(cavity | toothache, Gum Disease) = 0.7

- Sometimes, evidence may be irrelevant, allowing simplification, e.g.,

  P(*cavity* / *toothache, sunny*) = P(*cavity* | *toothache*) = 0.8

  - This kind of inference, sanctioned or supported by domain knowledge, is crucial

Review

Conditional Probability

- Definition of conditional probability:

$$P(a \mid b) = P(a \wedge b) / P(b) \text{ if } P(b) > 0$$

- **Product rule** gives an alternative formulation:

$$P(a \wedge b) = P(a \mid b) P(b) = P(b \mid a) P(a)$$

- A general version holds for whole distributions, e.g.,

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

- **Chain rule** is derived by successive application of product rule:

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

Joint Probability Distribution

- **For distributions on multiple random variables**
  - We use commas between the variables: so **P**(Weather, Cavity) denotes the probabilities of all combinations of values of the two variables,
  - For discrete random variables we can use a tabular representation, in this case yielding a 4x2 table of probabilities,
    - This gives the ***joint probability distribution JPD*** of Weather & Cavity tabulates the probability of every atomic event on those random variables

P(*Weather, Cavity*) = a 4 × 2 matrix of values:

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

Every question about a domain can be answered by the joint distribution

- The distribution notation, **P**, allows compact expressions,

Joint Probability Distribution

- Sometimes it will be useful to think about the complete set of random variables used to describe the world.
- The sum of probabilities for the **ALL random variables** in which it holds,
- for the random variables Cavity, Toothache & Weather
  - there are **16 possible worlds**, this can be represented as a 2x2x4 table.
- Inference using full joint distributions can be called as **probabilistic inference**
  - Posterior probabilities can be computed for query propositions given observed evidence

Joint Probability Distribution

Adding up probabilities where it is true.

❑ Inference by enumeration

- Simply identify the possible worlds in which the proposition is true and add up their probabilities.

For example, there are four possible worlds in which *toothache* holds. Adding these entries gives the unconditional/prior/marginal probability of *toothache.*

As given in the following example:

Joint Probability Distribution

Example

P(toothache)

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

- For any proposition a, sum the atomic events where it is true: $P(a) = \Sigma_{\omega:\omega \text{ s.t. } a=\text{true}} P(\omega)$
- P(*toothache*) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2

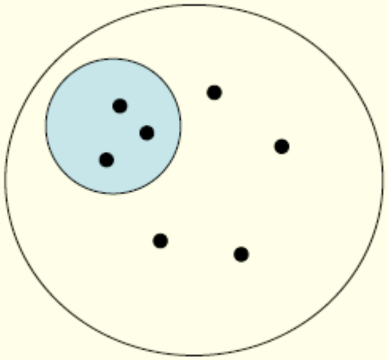
Adding the events where it is true

Joint Probability Distribution

Example

- For any proposition a, sum the atomic events where it is true: $P(a) = \Sigma_{\omega \text{ s.t. } a=true} P(\omega)$

$P(a)=1/7 + 1/7 + 1/7 = 3/7$



Joint Probability Distribution

❑ **Normalization** → making sure it Adds up to 1

- In computing the conditional probability, some constant probability terms can be considered as normalization constant, ensuring that it adds up to 1. (The normalization constant because it is the term whose value can be calculated by making sure that the probabilities of all outcomes sum to 1)

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

- Example
  - $\mathbf{P}(\mathbf{Y} \mid \mathbf{E} = \mathbf{e}) = \mathbf{P}(\mathbf{Y},\mathbf{E} = \mathbf{e})/\mathbf{P}(\mathbf{E} = \mathbf{e})$

  $P(a, b) = P(a \mid b)\, P(b)$

  Denominator can be viewed as a normalization constant α

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

  where H is the unobserved variable

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

The Probabilistic Wumpus World

- To use summation over the JPD:
  - Let *Unknown* be the set of $P_{i,j}$ variables for squares other than *Known* & [1,3]. So we have using ***probability enumeration***:

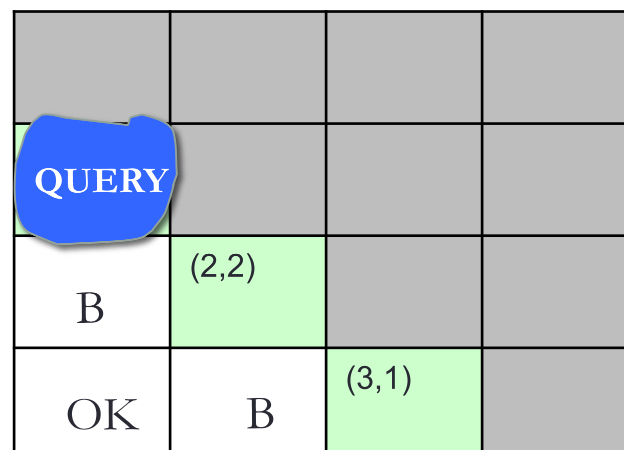
$$\mathbf{P}(P_{1,3} \mid known,\ b) = \mathbf{P}(P_{1,3},\ known,\ b) \,/\, \mathbf{P}(known,\ b)$$

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

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

Remember this form because we will use it later

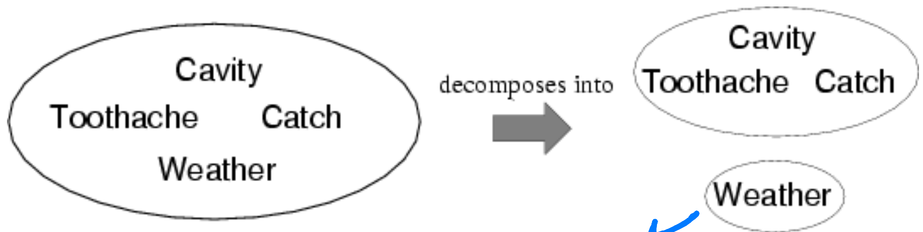
أبشر



Independence

- *A* and *B* are independent iff

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



Useless

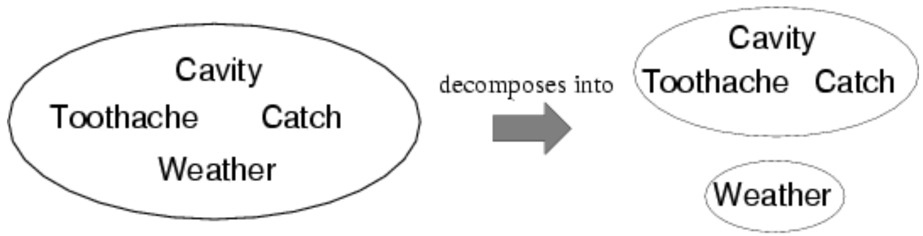
$$\mathbf{P}(\mathit{Toothache}, \mathit{Catch}, \mathit{Cavity}, \mathit{Weather}) = \mathbf{P}(\mathit{Toothache}, \mathit{Catch}, \mathit{Cavity})\,\mathbf{P}(\mathit{Weather})$$

- 32 entries reduced to 12; (2^3 +4)
- Then, for n independent biased coins, $O(2^n) \rightarrow O(n)$

Independence

- *A* and *B* are independent iff

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



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

Also, remember this rule because we will use it later

Example: In the Wumpus World, we have:

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

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

Conditional Independence: Example

- *Toothache*: boolean variable indicating whether the patient has a toothache
- *Cavity*: boolean variable indicating whether the patient has a cavity
- *Catch*: whether the dentist's probe catches in the cavity

dentist's probe

Conditional independence: Example

- If the patient has a cavity, the probability that the probe catches in it doesn't depend on whether he/she has a toothache

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

or

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

- Therefore, *Catch* is **conditionally independent** of *Toothache* given *Cavity*
- Likewise, *Toothache* is conditionally independent of *Catch* given *Cavity*

$$P(Toothache \mid Catch, Cavity) = P(Toothache \mid Cavity)$$

- Equivalent statement:

$$P(Toothache, Catch \mid Cavity) = P(Toothache \mid Cavity)\, P(Catch \mid Cavity)$$

Conditional independence contd.

- Write out full joint distribution using chain rule:

$P(a, b) = P(a \mid b)\, P(b)$

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

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

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

- 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*.
- Conditional independence is our most basic and robust form of knowledge about uncertain environments.

Bayes' Rule

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

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

$$\underbrace{P(a|b)}_{\text{Posterior}} = \frac{\overbrace{P(b|a)}^{\text{Likelihood}} * \overbrace{P(a)}^{\text{Prior}}}{\underbrace{P(b)}_{\text{Normalization}}}$$

$$\frac{P(2 \mid 1) * P(1)}{P(2)}$$

or in distribution form

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

Bayes' Rule - **Example**

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

- Useful for assessing diagnostic probability (**Bottom-up Inference**) from **causal probability** (Top-Down Inference)
  - P(Cause|Effect) = P(Effect|Cause) P(Cause) / P(Effect)
  - E.g., let *M* be meningitis, *S* be stiff neck:

    $$P(m|s) = P(s|m)\,P(m) / P(s) = 0.8 \times 0.0001 / 0.1 = 0.0008$$

  - Then: posterior probability of meningitis still very small!

Bayes Rule (overall)

- By definition, we know that $P(h|e) = \frac{P(h \wedge e)}{P(e)}$
- We can rearrange terms to show:
$$P(h \wedge e) = P(h|e) \times P(e)$$
- Similarly, we can show:
$$P(e \wedge h) = P(e|h) \times P(h)$$
- Since $e \wedge h$ and $h \wedge e$ are identical, we have:

Theorem (Bayes theorem, or Bayes rule)

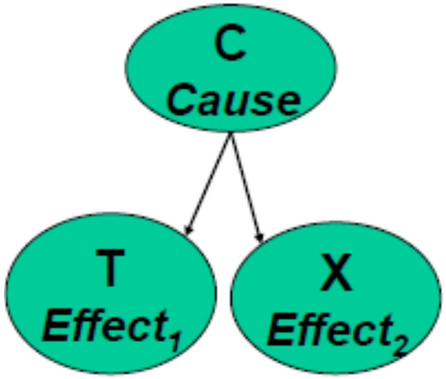
$$P(h|e) = \frac{P(e|h) \times P(h)}{P(e)}$$

Bayes' Rule and Conditional Independence

By Bayes Rule $P(C|T,X) = \dfrac{P(T,X|C)P(C)}{P(T,X)}$

If T and X are *conditionally independent given C*:

$$P(C|T,X) = \frac{P(T|C)P(X|C)P(C)}{P(T,X)}$$

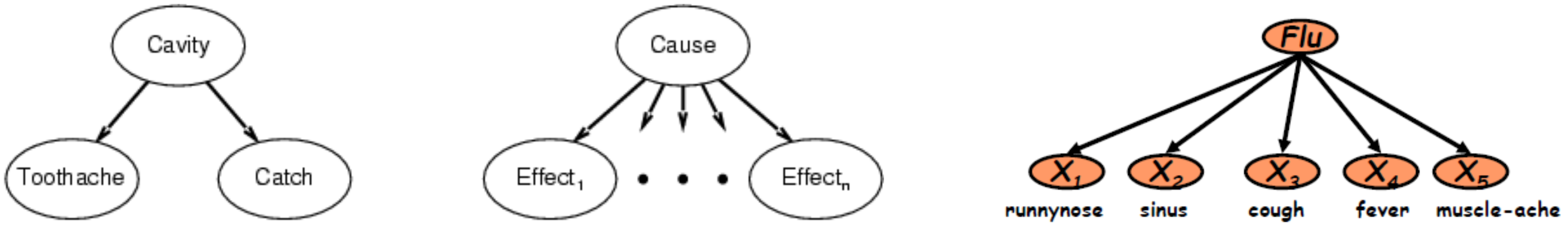


Bayes' Rule and Conditional Independence

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

- This is an example of a naïve Bayes model:

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



- ***i.e all effects assumed conditionally independent given Cause***

Exercise

100% بتجي بالفاينل !

Using the given table and the product rule:

Quiz

$$P(a \mid b) = P(a \wedge b) \,/\, P(b),$$

Calculate:

A: The probability of a cavity, given evidence of a toothache

B: The probability of *no* cavity, given evidence of a toothache

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

Answer

$$P(a \mid b) = P(a \wedge b) \,/\, P(b),$$

Composition .

A: The probability of a cavity, given evidence of a toothache

$$P(cavity \mid toothache) = P(cavity \wedge toothache) \,/\, P(toothache)$$

$$= (0.108 + 0.012) \,/\, (0.108 + 0.012 + 0.016 + 0.064) = \mathbf{0.6}$$

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

Answer

$$P(a \mid b) = P(a \wedge b) \,/\, P(b),$$

A: The probability of a cavity, given evidence of a toothache

$$P(cavity \mid toothache) = P(cavity \wedge toothache) \,/\, P(toothache)$$

$$= (0.108 + 0.012) \,/\, (0.108 + 0.012 + 0.016 + 0.064) = \mathbf{0.6}$$

B: The probability of *no* cavity, given evidence of a toothache

B. $P(\neg cavity \mid toothache) = P(\neg cavity \wedge toothache) \,/\, P(toothache)$

$$= (0.016 + 0.064) \,/\, (0.108 + 0.012 + 0.016 + 0.064 = \mathbf{0.4}$$

→ Normalized

 # Please note that 0.6 + 0.4 = 1

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

Exercise

Using the given table **Calculate:**

$$P(Cavity \vee Tooth)$$

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

Exercise

Using the given table **Calculate:**

P(Cavity ∨ Tooth) = 0.108 + 0.012 + 0.016 + 0.064 + 0.072 + 0.008

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

Exercise

Using the given table **Calculate:**

	Toothache	¬Toothache
Cavity	0.04	0.06
¬Cavity	0.01	0.89

$P(\neg Cavity \wedge Toothache)$

No Cavity and Toothache

$P(Cavity \wedge \neg Toothache)$

Cavity and no Tooth Ache

Exercise

Using the given table **Calculate:**

	Toothache	¬Toothache
Cavity	0.04	0.06
¬Cavity	0.01	0.89

0.04 0.01

- $P(Toothache) = P((Toothache \wedge Cavity) \vee (Toothache \wedge \neg Cavity))$

  $= P(Toothache \wedge Cavity) + P(Toothache \wedge \neg Cavity)$

  $= 0.04 + 0.01 = 0.05$

- $P(Toothache \vee Cavity)$

  $= P((Toothache \wedge Cavity) \vee (Toothache \wedge \neg Cavity) \vee (\neg Toothache \wedge Cavity))$

  $= 0.04 + 0.01 + 0.06 = 0.11$

Bayes' theorem is a fundamental concept in probability theory that provides a way to update our beliefs about the probability of an event based on new evidence or information. The theorem is named after Reverend Thomas Bayes, an 18th-century British mathematician and theologian who first formulated the idea.

Bayes' theorem states that the probability of an event A given some evidence B is equal to the probability of B given A multiplied by the prior probability of A, divided by the prior probability of B. Mathematically, it can be written as:

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

P(A | B) = P(B | A) * P(A) / P(B)

where P(A | B) is the probability of A given B, P(B | A) is the probability of B given A, P(A) is the prior probability of A, and P(B) is the prior probability of B.

In simpler terms, Bayes' theorem allows us to update our prior belief about the probability of an event, given new information. For example, suppose we want to know the probability of a person having a disease based on a positive test result. Bayes' theorem can help us calculate this probability by taking into account the prevalence of the disease in the population (prior probability) and the accuracy of the test (likelihood ratio).

In practice, Bayes' theorem has many applications in various fields such as statistics, machine learning, and artificial intelligence. It is a powerful tool for making predictions and informed decisions based on incomplete or uncertain information.

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

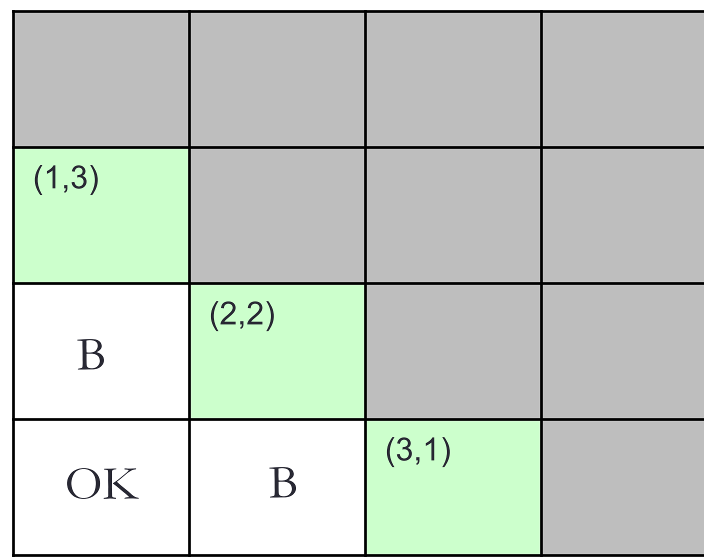
Uncertainty

The Wumpus World-Revisited

Lecture III

The Wumpus World **Revisited**

- We will combine the ideas of probabilities in order to solve the **probabilistic reasoning problems** in the wumpus world.
- Uncertainty arises in the wumpus world because the agent's sensors give only partial, local information about the world *.
- Pure logical inference can conclude nothing about which square is most likely to be safe, so a logical agent might be forced to choose the square randomly.
- Our new method: Choose and explore the square with the highest likelihood being safe.



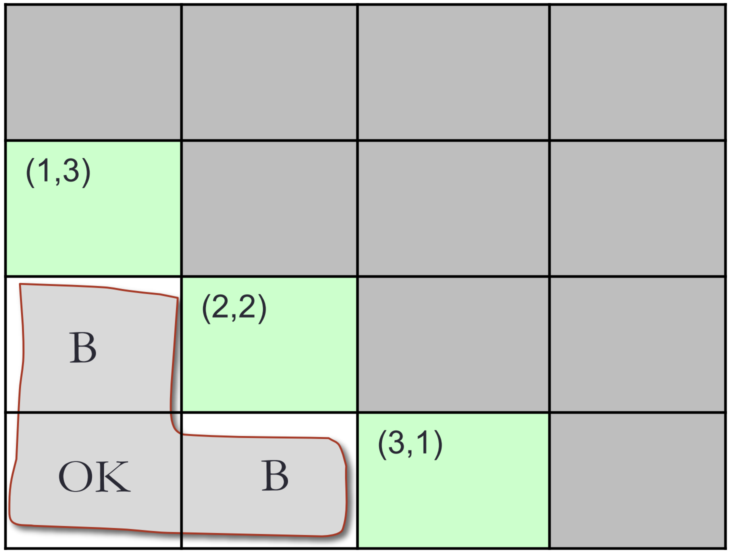
Probabilistic Agent vs. Logical Agent

The Probabilistic Wumpus World

- **Aim**: *calculate the probability that each of the three squares contains a Pit* *
  - Each square other than [1,1] contains a pit with probability 0.2, **WHY?**
  - Identify a set of random variables:

    1) A Boolean variable P_{ij} for each square iff it contains a Pit.

    2) A Boolean variables B_{ij} for each square iff the observed squares is breezy;



Assume that there was NO wunpus at (1,3) how will the agent knows if there is a pit or not in each of (3,1), (2,2) and (1,3).

The Probabilistic Wumpus World

- We begin with the joint probability distribution JPD:

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

applying the product rule yields:

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

- **1st term**: the conditional probability of a breeze configuration given a pit configuration (**P**(Effect | Cause))
  - $B_{i,j}$ values in the first term are 1 if the breeze adjacent to a Pit, 0 otherwise
- **2nd term**: the prior probability of a Pit configuration
  - Pits are placed randomly, independent of each other, with probability 0.2 for any square, so:

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

  - which, for a particular configuration that has n pits is $\mathbf{P}(P_{1,1}, \ldots, P_{4,4}) = 0.2^{n} \times 0.8^{16-n}$

The Probabilistic Wumpus World

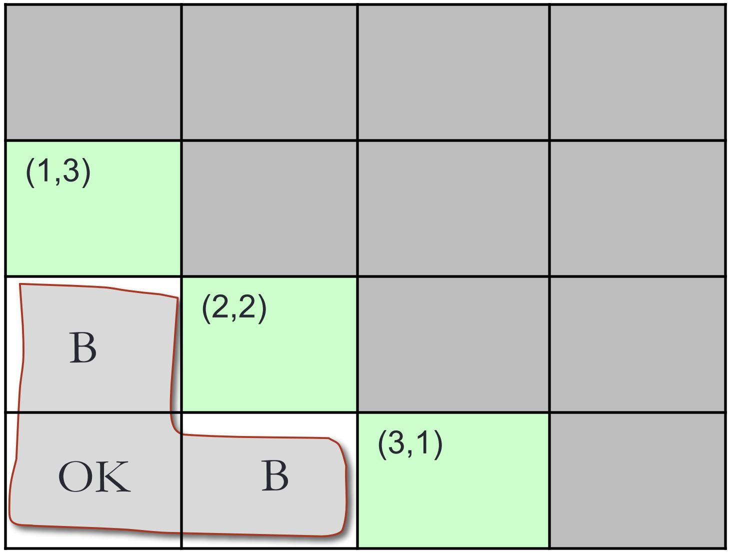
- In the example, we have as observed evidence
  - A breeze or not in each visited square + no pit in any visited square, abbreviated as b & known:

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

  - An example **query** concerns the safety of other squares: what's the probability of a pit at [1,3], given the evidence so far?

$$\mathbf{P}(P_{1,3} \mid known, b)$$

  - We could answer by summing over cells in the JPD.



The Probabilistic Wumpus World

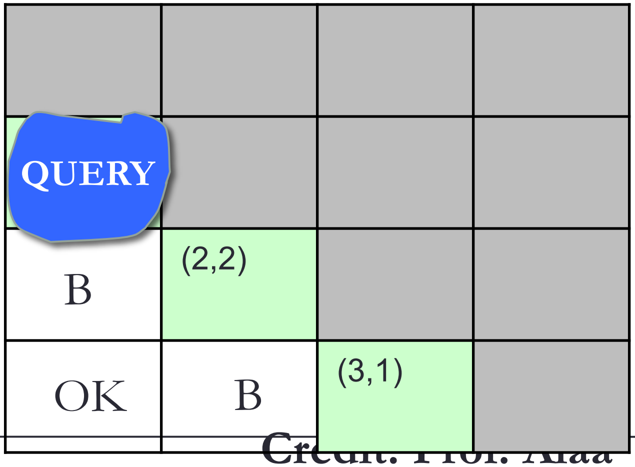
- To use summation over the JPD:
  - Let *Unknown* be the set of $P_{i,j}$ variables for squares other than *Known* & [1,3]. So we have using ***probability enumeration***:

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

(*back to slide 34*)

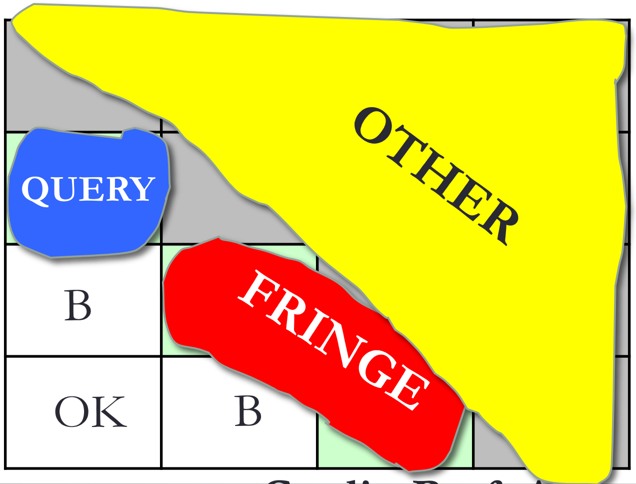
$$= \alpha\ \Sigma_{\text{unknown}}\ \mathbf{P}(P_{1,3},\ unknown,\ known,\ b)$$

  - That is, we can just sum over the entries in the JPD but with 12 unknown squares we have 2^{12} = 4096 terms in the summation, so the calculation is <u>exponential</u> in the number of squares.
  - So we'll need to simplify from insight about independence!
    - We note: *<u>not all unknown squares are equally relevant to the query</u>* *



The Probabilistic Wumpus World

- To use summation over the JPD with independence:
  - Let ***Fringe*** be the variables (other than the query variable) that are adjacent to visited squares, in this case just [2,2] and [3,1].
  - Also, let ***Other*** be the variables for the other unknown squares; in this case, there are 10 other squares,
  - With this description, we see that the observed breezes are *conditionally independent* of the **other** variables, given the ***known***, ***fringe*** & ***query*** variables.
  - Then, we manipulate the query formula into a form in which the breezes are conditioned on all the other variables, as follows:



The Probabilistic Wumpus World

- Using the conditional independence simplification

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

then using the product rule: $\mathbf{P(X, Y) = P(X \mid Y) P(Y)}$.

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

then partitioning unknown into fringe & other

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

Then using the <u>conditional independence</u> of *observed breezes* from *other* variables, given the *known, $P_{1,3}$ & fringe*, dropping *other* from the first term:

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

since the 1st term now does not depend on other, move the summation inward

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

The Probabilistic Wumpus World

- Use independence to factor the prior term:

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

- Then reorder the terms:

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

- Fold P(known) into the normalizing constant:
  - & use $\Sigma_{other} \, P(other) = 1$

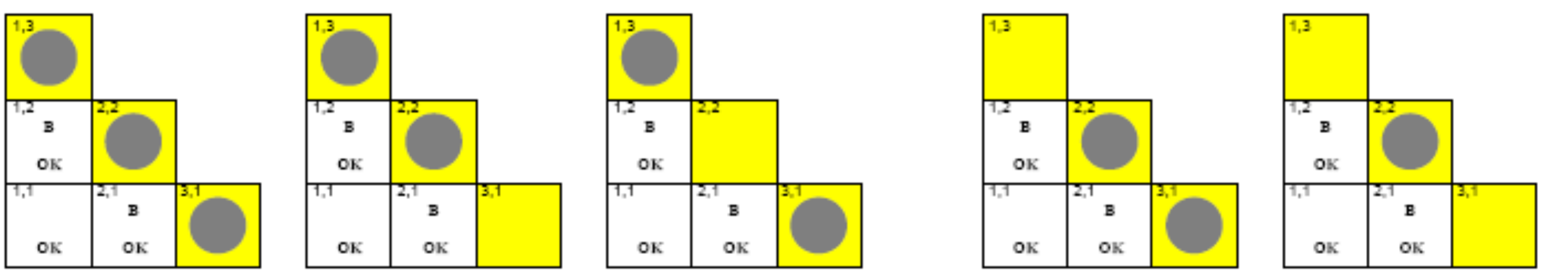
$$= \alpha' \, \mathbf{P}(P_{1,3}) \, \Sigma_{fringe} \mathbf{P}(b \mid known, P_{1,3}, fringe) \, P(fringe)$$

The Probabilistic Wumpus World

- Using conditional independence & independence
  - has yielded an expression with just 4 terms in the summation over the fringe variables $P_{2,2}$ & $P_{3,1}$ eliminating other squares

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

  - The expression $\mathbf{P}(b \mid known, P_{1,3}, fringe)$ is 1 when the fringe is consistent with the breeze observations, 0 otherwise. *
  - This figure shows the models & the associated priors P(fringe)
  - So to get each value of $P_{1,3}$ we sum over the logical models for fringe variables that are consistent with known facts



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

The Probabilistic Wumpus World

- Sum over the logical models for frontier variables that are consistent with known facts

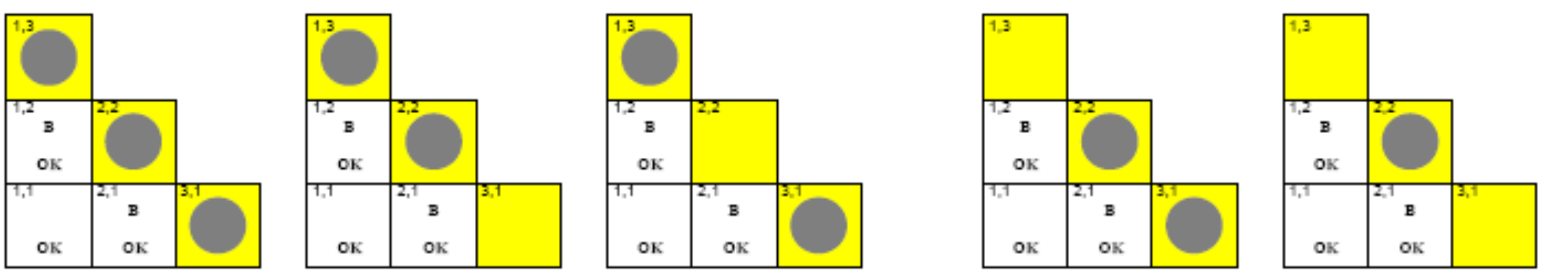
$$\mathbf{P}(P_{1,3} \mid known, b) = \alpha' \, \mathbf{P}(P_{1,3}) \, \Sigma_{fringe} \, P(fringe)$$

$$\mathbf{P}(P_{1,3} \mid known, b) = \alpha' \langle 0.2(0.04 + 0.16 + 0.16), 0.8(0.04 + 0.16) \rangle$$

with $\alpha' = 4.31$;

$$\approx \langle 0.31, 0.69 \rangle$$

- Consistent models for fringe variables $P_{2,2}$ & $P_{3,1}$ with P(fringe) for each model, for $P_{1,3}$ = true & $P_{1,3}$ = false



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

The Probabilistic Wumpus World

- Note that $P_{1,3}$, $P_{3,1}$ are symmetric
  - So by symmetry, [3,1] would contain a pit about 31% of the time:

    $$\mathbf{P}(P_{3,1} | known, b) = \langle 0.31, 0.69 \rangle$$

  - It is clear to the probabilistic agent where **not** to go next.

  - By a similar calculation, [2,2] can be shown to contain a pit with about

    ?? probability: $\mathbf{P}(P_{2,2}|known, b) \approx \langle ??, ?? \rangle$

IT IS HOMEWORK

The Probabilistic Wumpus World

- The logical agent & the probabilistic agent
  - Strictly logical inferencing can only yield known safe/known unsafe/unknown
  - The probabilistic agent knows which move is relatively safer, relatively more dangerous
  - For efficient probabilistic solutions we can use independence & conditional independence among variables to simplify the summations involved
    - fortunately, these often match our natural understanding of how the problem should be decomposed

Summary

- **Joint Probability Distribution JPD**
  - specifies the probability for every assignment of values to random variables, & when available allows summation over possible worlds to answer queries, but has complexity exponential in the number of variables
- **Absolute independence**
  - allows decomposition of a problem's random variables into smaller joint distributions, reducing complexity, but is rare
- **Conditional independence**
  - derives from shared causal relationships in the domain & may allow factoring of the JPD into smaller conditional distributions
- In Wumpus World
  - by simplifying calculations via conditional independence the agent may calculate probabilities for unobserved variables and so do better than a purely logical agent

Example of Uncertainty

Nowadays Applications

Korean Competition Shows Weather Still a Challenge for Autonomous Cars

http://spectrum.ieee.org/cars-that-think/transportation/advanced-cars/japan-competition-shows-weather-still-a-challenge-for-autonomous-cars/?utm_source=carsthatthink&utm_medium=email&utm_campaign=1112

7. Uncertain Knowledge & Reasoning

Uncertainty

Representing knowledge in an uncertain domain
Using Bayesian Networks

Lecture IV

Bayesian networks

- A simple, graphical notation for **conditional independence** assertions
- A compact way to represent the Joint Distribution of a set of Random Variables. The nodes represent Random Variables. Random variables are variables that provides a mapping from values to probabilities.

- Syntax:
  - a set of nodes, one per variable
  - a directed, acyclic graph (link ≈ "directly influences")
  - a conditional distribution for each node given its parents:

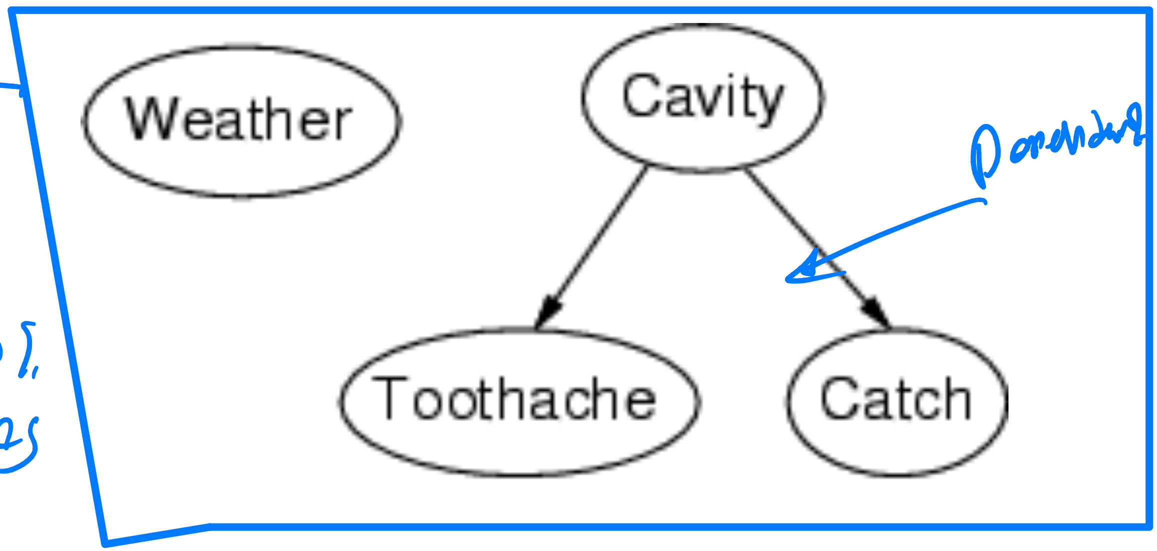
$$\mathbf{P}\,(X_i \mid \text{Parents}\,(X_i))$$

- In the simplest case, conditional distribution represented as a conditional probability table (CPT) giving the distribution over X_i for each combination of parent values.

Example -1

- Topology of network encodes conditional independence assertions:

Bayesian Network of Independencies, Dependencies



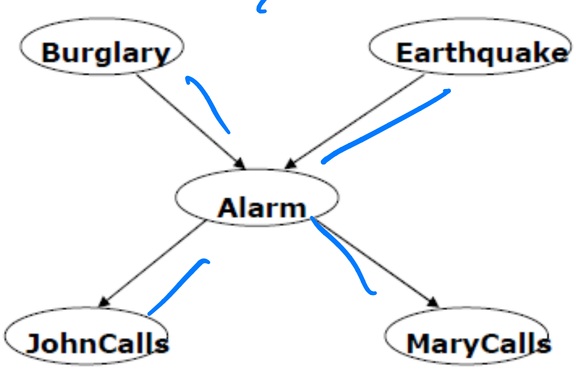
- Weather is **<u>independent</u>** of the other variables
- Toothache and Catch are **<u>conditionally independent</u>** given Cavity

Example -1

- Assume your house has an **alarm system** against **burglary**. You live in the seismically active area and the alarm system can get occasionally set off by an **earthquake**. You have two neighbors, **Mary** and **John**, who do not know each other. If they hear the alarm they call you, but this is not guaranteed.
- We want to represent the probability distribution of events:
  - Burglary, Earthquake, Alarm, Mary calls and John calls

All dependant on the Alarm

Causal relations



Example -1

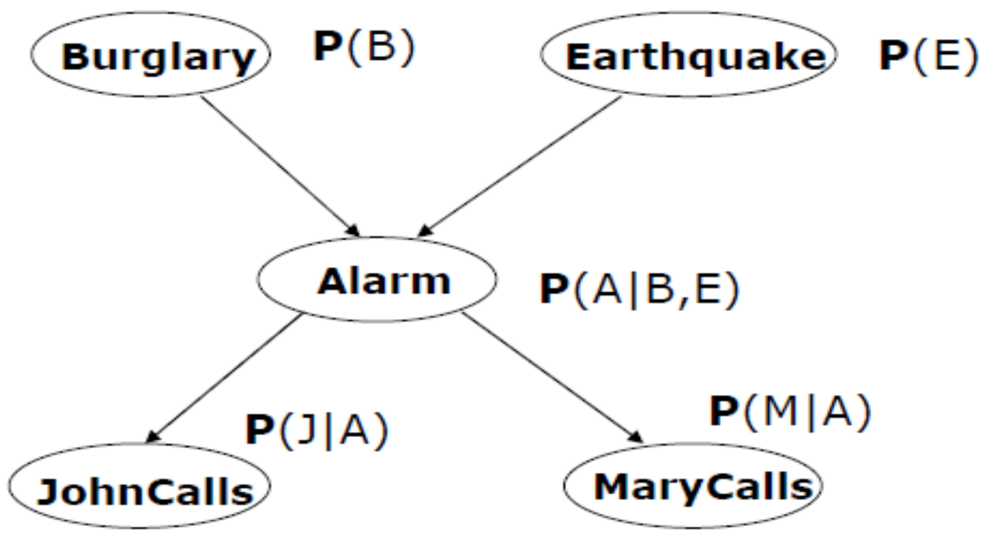
- **Variables:** *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, *MaryCalls*

- **Network topology reflects "causal" knowledge**:
  - A burglar can set the alarm off
  - An earthquake can set the alarm off
  - The alarm can cause Mary to call
  - The alarm can cause John to call

Example -1

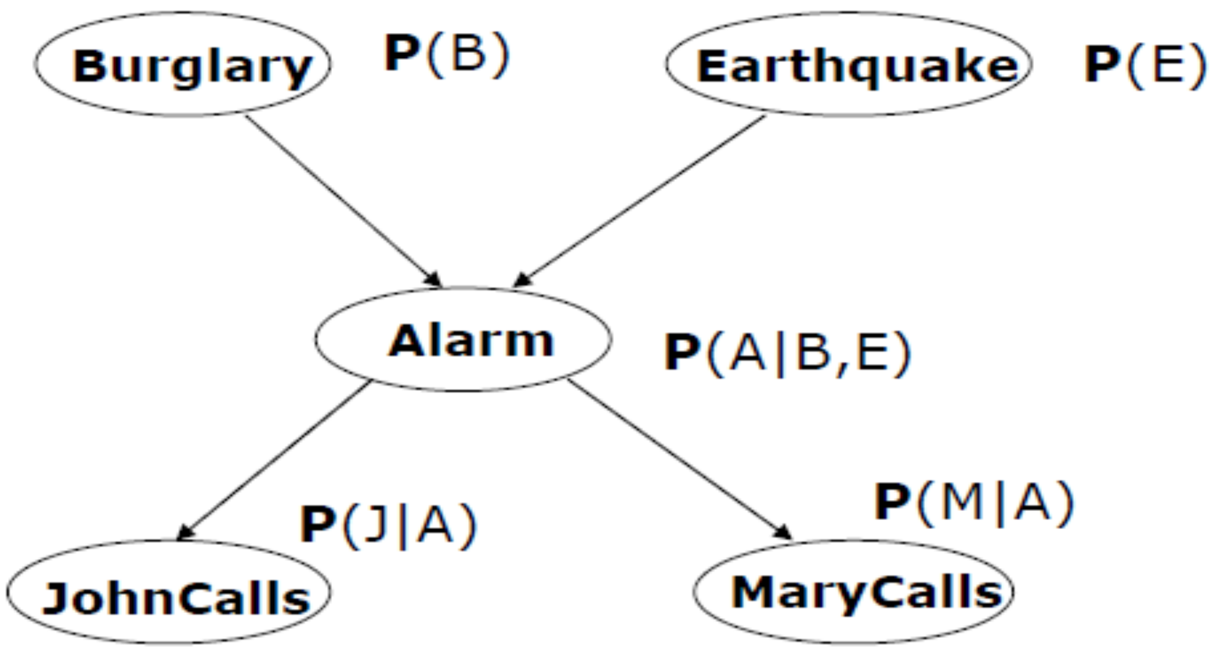
1. **Directed acyclic graph**
   - **Nodes** = random variables
     Burglary, Earthquake, Alarm, Mary calls and John calls
   - **Links** = direct (causal) dependencies between variables.

     The chance of Alarm is influenced by Earthquake, The chance of John calling is affected by the Alarm

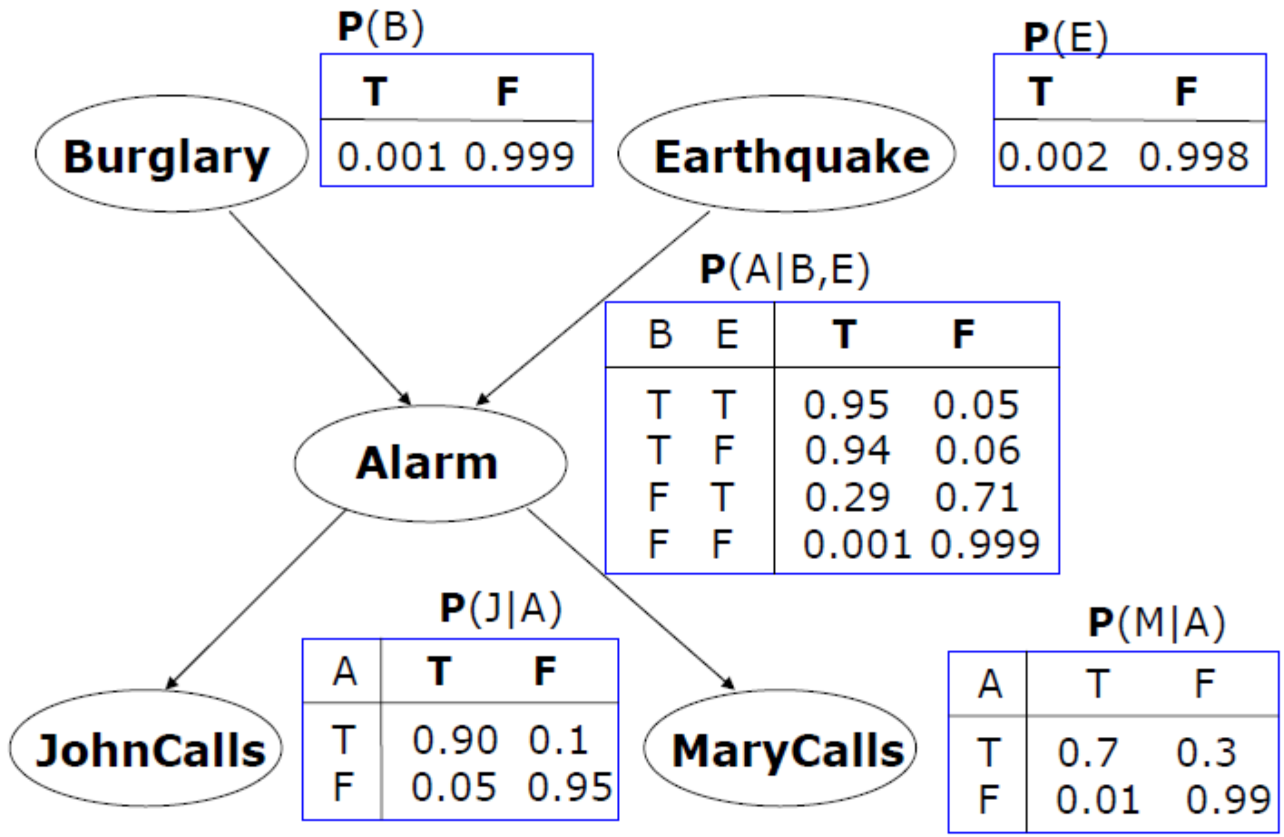


Example -1

2. **Local conditional distributions**
   - relate variables and their parents



Example -1



P(B)

T	**F**
0.001	0.999

P(E)

T	**F**
0.002	0.998

P(A|B,E)

B	E	**T**	**F**
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

P(J|A)

A	**T**	**F**
T	0.90	0.1
F	0.05	0.95

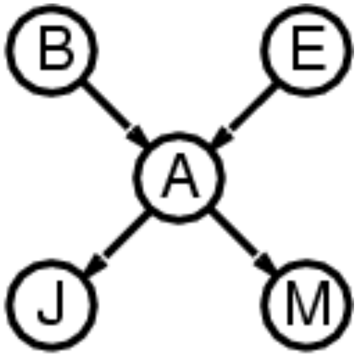
P(M|A)

A	T	F
T	0.7	0.3
F	0.01	0.99

Next to each node, you have conditional probability tables (**CPT**), these represent the conditional probability of the underlying Random Variable conditioned on its parents.

Compactness

- A CPT for Boolean X_i with k Boolean parents has 2^k rows for the combinations of parent values.



- Each row requires one number p for $X_i = true$ (the number for $X_i = false$ is just $1\text{-}p$)
- If each variable has no more than k parents, the complete network requires $O(n \cdot 2^k)$ numbers for each node.
- i.e., grows linearly with n, vs. **$O(2^n)$** for the full joint distribution
- For burglary net, 1 + 1 + 4 + 2 + 2 = **10** numbers (vs. 2^5-1 = 31)

Research Assignment # 1

Try to study the Bayesian Nets (BN) and its variant BBN, in the scope of the following elements:

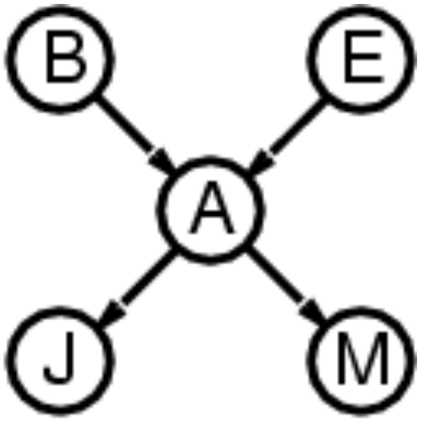
- Definition of BN and BBN
- What is the relation with FJD
- What is semantic of BN
- How can you achieve Inference using BN/BBN
- Applications

- Grading will be out of 2.5 Marks
- What you will study will be included in the final exams
- **NO cheat/No C+P. In such case (Penalty is -2 Marks)..TAKE CARE ☹**

Semantics

The full joint distribution is defined as the product of the local conditional distributions:

$$\mathbf{P}(X_1, \ldots, X_n) = \prod_{i=1}^{n} \mathbf{P}(X_i \mid Parents(X_i))$$



e.g., $\mathbf{P}(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

$= \mathbf{P}(j \mid a)\, \mathbf{P}(m \mid a)\, \mathbf{P}(a \mid \neg b, \neg e)\, \mathbf{P}(\neg b)\, \mathbf{P}(\neg e)$

Constructing Bayesian networks

- 1. Choose an ordering of variables $X_1, \ldots, X_n$
- 2. For i = 1 to n

  add X_i to the network

  select parents from $X_1, \ldots, X_{i-1}$ such that

  $$\mathbf{P}(X_i \mid Parents(X_i)) = \mathbf{P}(X_i \mid X_1, \ldots X_{i-1})$$

This choice of parents guarantees:

$$\mathbf{P}(X_1, \ldots, X_n) = \prod_{i=1}^{n} \mathbf{P}(X_i \mid X_1, \ldots, X_{i-1}) \text{ (chain rule)}$$

…Back to slide 28

$$= \prod_{i=1}^{n} \mathbf{P}(X_i \mid Parents(X_i)) \text{ (by construction)}$$

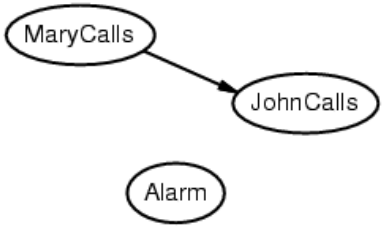
Example

- Suppose we choose the ordering M, J, A, B, E

$\mathbf{P}(J \mid M) = \mathbf{P}(J)$?

Example

- Suppose we choose the ordering M, J, A, B, E

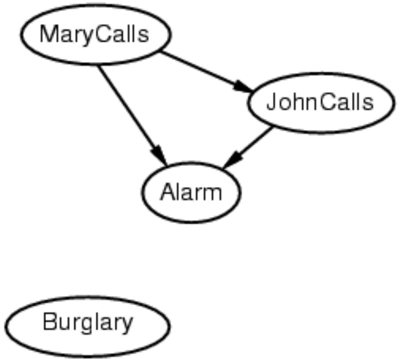


P(J | M) = **P**(J)? **No**

P(A | J, M) = **P**(A | J)? **P**(A | J, M) = **P**(A)?

Example

- Suppose we choose the ordering *M, J, A, B, E*



$\mathbf{P}(J \mid M) = \mathbf{P}(J)$? **No**

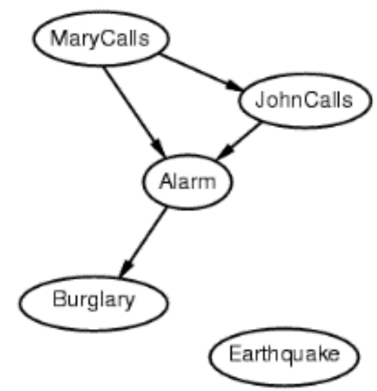
$\mathbf{P}(A \mid J, M) = \mathbf{P}(A \mid J)$? $\mathbf{P}(A \mid J, M) = \mathbf{P}(A)$? **No**

$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B \mid A)$?

$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B)$?

Example

- Suppose we choose the ordering M, J, A, B, E



$\mathbf{P}(J \mid M) = \mathbf{P}(J)$? **No**

$\mathbf{P}(A \mid J, M) = \mathbf{P}(A \mid J)$? $\mathbf{P}(A \mid J, M) = \mathbf{P}(A)$? **No**

$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B \mid A)$? **Yes**

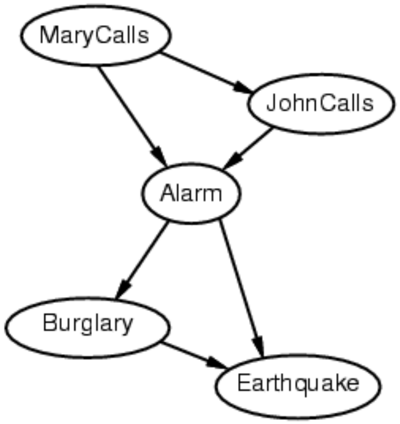
$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B)$? **No**

$\mathbf{P}(E \mid B, A, J, M) = \mathbf{P}(E \mid A)$?

$\mathbf{P}(E \mid B, A, J, M) = \mathbf{P}(E \mid A, B)$?

Example

- Suppose we choose the ordering M, J, A, B, E



$\mathbf{P}(J \mid M) = \mathbf{P}(J)$? **No**

$\mathbf{P}(A \mid J, M) = \mathbf{P}(A \mid J)$? $\mathbf{P}(A \mid J, M) = \mathbf{P}(A)$? **No**

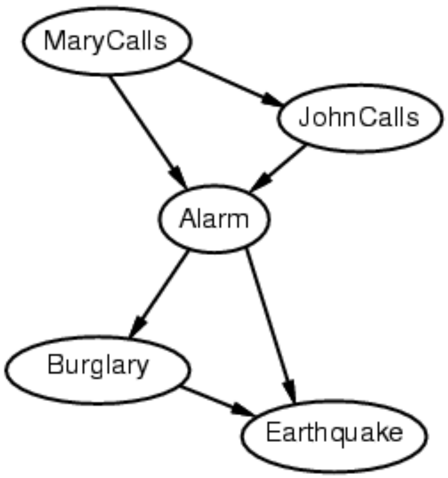
$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B \mid A)$? **Yes**

$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B)$? **No**

$\mathbf{P}(E \mid B, A, J, M) = \mathbf{P}(E \mid A)$? **No**

$\mathbf{P}(E \mid B, A, J, M) = \mathbf{P}(E \mid A, B)$? **Yes**

Example



- Deciding conditional independence is hard in noncausal directions
- (Causal models and conditional independence seem hardwired for humans!)
- Network is less compact: 1 + 2 + 4 + 2 + 4 = 13 numbers needed

Summary

- Bayesian networks provide a natural representation for (causally induced) conditional independence
- Topology + CPTs = compact representation of joint distribution
- Generally easy for domain experts to construct


Any questions?