Lecture 11a: Inference in Bayesian Networks

CSCI 360

Introduction to Artificial Intelligence

USC

Here is where we are…

	3/1		Project 2 Out	
9	3/4 3/6	3/5 3/7	Quantifying Uncertainty Bayesian Networks	[Ch 13.1-13.6] [Ch 14.1-14.2]
10	3/11 3/13	3/12 3/14	*(spring break, no class)* *(spring break, no class)*	
11	3/18 3/20	3/19 3/21	Inference in Bayesian Networks Decision Theory	[Ch 14.3-14.4] [Ch 16.1-16.3 and 16.5]
	3/23		Project 2 Due	
12	3/25 3/27	3/26 3/28	***Advanced topics*** *(Chao traveling to National Science Foundation)* ***Advanced topics*** *(Chao traveling to National Science Foundation)*	
	3/29		Homework 2 Out	
13	4/1 4/3	4/2 4/4	Markov Decision Processes Decision Tree Learning	[Ch 17.1-17.2] [Ch 18.1-18.3]
	4/5 4/5		Homework 2 Due Project 3 Out	
14	4/8 4/10	4/9 4/11	Perceptron Learning Neural Network Learning	[Ch 18.7.1-18.7.2] [Ch 18.7.3-18.7.4]
15	4/15 4/17	4/16 4/18	Statistical Learning Reinforcement Learning	[Ch 20.2.1-20.2.2] [Ch 21.1-21.2]
16	4/22 4/24	4/23 4/25	Artificial Intelligence Ethics Wrap-Up and Final Review	
	4/26		Project 3 Due	
	5/3	5/2	**Final Exam** (2pm-4pm)	

Outline

- What is AI?
- Problem-solving agent (search)
- Knowledge-based agent (logical reasoning)
- **Probabilistic reasoning**
  - Quantifying Uncertainty
  - Bayesian Networks
  - **Inference in Bayesian Networks**
  - Decision Theory
  - Markov Decision Processes
- Machine learning

What we have learned so far…

- Early AI researchers largely rejected using probability in their systems
  - "People don't think that way…"

- However, neither **problem-solving** nor **logical reasoning** agents tolerate approximation well…
  - Need probabilistic modeling/reasoning

Recap: *Making decision*

Rational decision depends on

(1) The relative importance of various goals and

(2) likelihood that (and degree to which) they will be reached

Decision theory = Utility theory + Probability theory

Choose the action that yields the ***highest expected utility***, averaged over all the possible outcomes of the action

Recap: *Conditional (or posterior) probability*

- For any propositions *a* and *b*, we have

$$P(a \mid b) = \frac{P(a \wedge b)}{P(b)} \qquad \text{whenever } P(b) > 0.$$

- $\boxed{P(\underbrace{a \wedge b}_{\text{joint probability}}) = P(\underbrace{a \mid b}_{\text{conditional (or posterior) probability}})\, P(\underbrace{b}_{\text{unconditional (or prior) probability}})}$

Recap: *Probability distribution*

- Probabilities of all possible values of a random variable

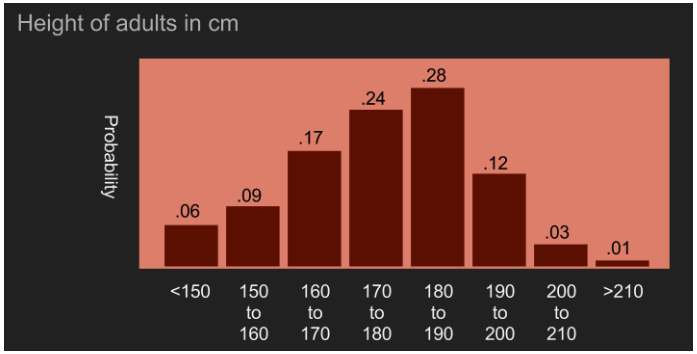
$$P(Weather = sunny) = 0.6$$
$$P(Weather = rain) = 0.1$$
$$P(Weather = cloudy) = 0.29$$
$$P(Weather = snow) = 0.01 \ ,$$

- In a vector format

$$\boxed{\mathbf{P}(Weather) = \langle 0.6, 0.1, 0.29, 0.01 \rangle}$$

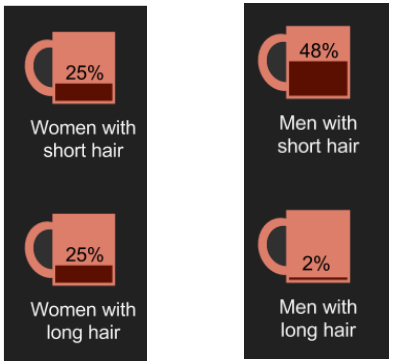
- Other examples

one-variable



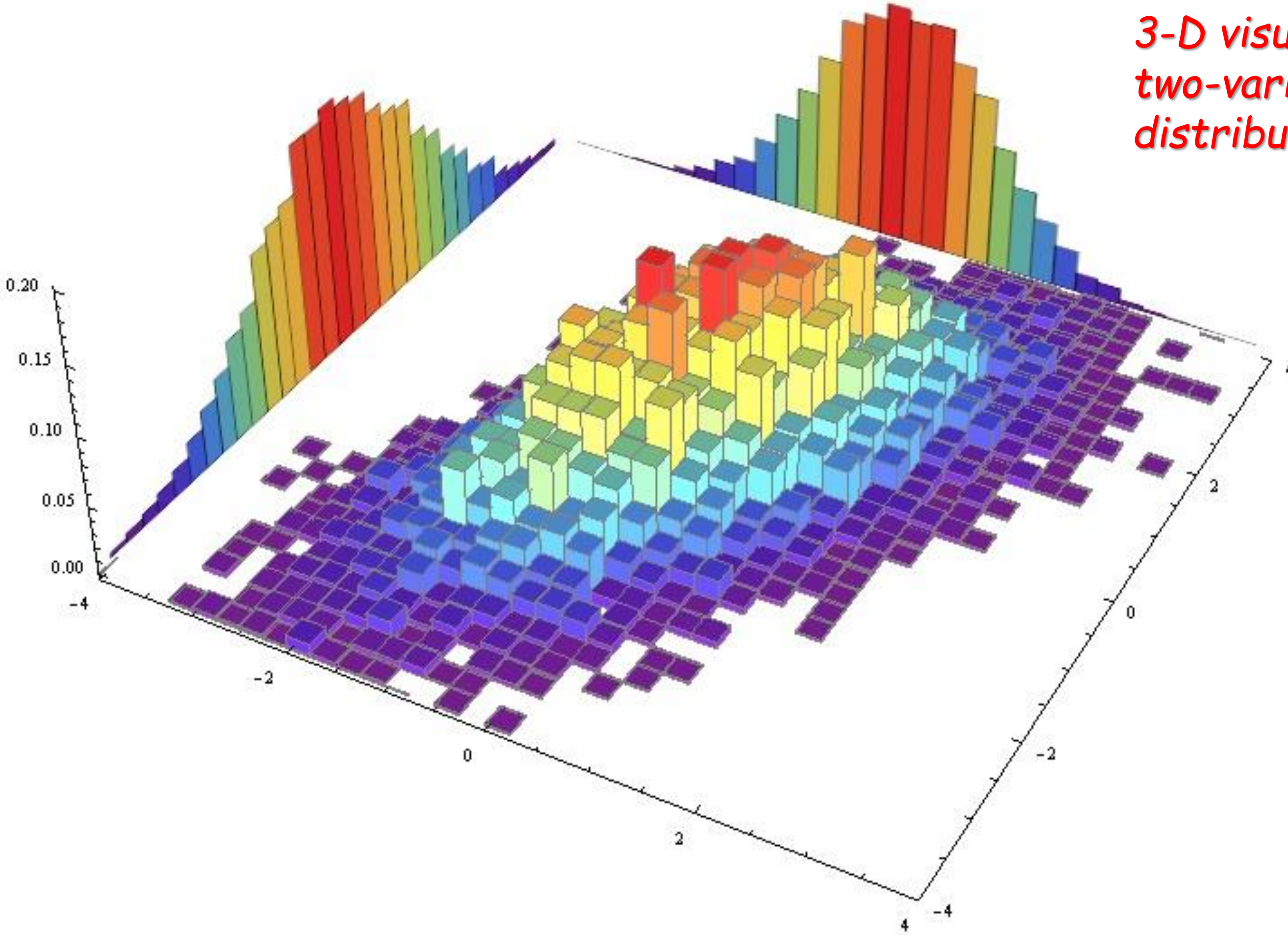
https://brohrer.github.io

multi-variable



Recap: *Joint probability distribution*

- Probabilities of all possible values of **multiple** variables



3-D visualization of two-variable joint distribution

Recap: *Marginal probability*

- Extracting the distribution **over a subset** of variables from the full joint distribution

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

- Example *-- getting rid of the other two variables*

$$P(cavity) =$$

Recap: *Marginal probability*

- Extracting the distribution **over a subset** of variables from the full joint distribution

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

- Example -- *getting rid of the other two variables*

$$P(cavity) = 0.108 + 0.012 + 0.072 + 0.008 = 0.2$$

Recap: *Normalization*

- The probability of **cavity**, or **no cavity**, given **toothache**

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

- Example

$$P(cavity \mid toothache) \; =$$

$$P(\neg cavity \mid toothache) \; =$$

Sum of the two is always 1.0

Recap: *Normalization*

- The probability of **cavity**, or **no cavity**, given **toothache**

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

- Example

No need to compute *P (toothache)* any more

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

$$= \frac{0.108 + 0.012}{\cancel{0.108 + 0.012 + 0.016 + 0.064}} = 0.6$$

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

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

Recap: *Normalization*

- The probability of **cavity**, or **no cavity**, given **toothache**

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

- Example

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

$$= \alpha\, \langle 0.12, 0.08 \rangle = \langle 0.6, 0.4 \rangle\ .$$

Assume that
α = 1 / (0.12+0.08)
= 1/0.2
= 5

Recap: *Independence to reduce table size*

- Consider ***P****(Toothache, Catch, Cavity, Weather)*, which has 32 entries in the **full joint distribution** table

	toothache		¬toothache		toothache		¬toothache		toothache		¬toothache		toothache		¬toothache	
	catch	¬catch	catch	¬catch	catch	¬catch	catch	¬catch	catch	¬catch	catch	¬catch	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008	0.108	0.012	0.072	0.008	0.108	0.012	0.072	0.008	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576	0.016	0.064	0.144	0.576	0.016	0.064	0.144	0.576	0.016	0.064	0.144	0.576

- Applying the product rule

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

- But weather is not influenced by dentistry!

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

$$\boxed{P(toothache, catch, cavity, cloudy) = P(cloudy) P(toothache, catch, cavity)}$$

Recap: *Independence to reduce table size*

- Consider ***P**(Toothache, Catch, Cavity, Weather)*, which has 32 entries in the **full joint distribution** table

	toothache		*¬toothache*		*toothache*		*¬toothache*		*toothache*		*¬toothache*		*toothache*		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*	*catch*	*¬catch*	*catch*	*¬catch*	*catch*	*¬catch*	*catch*	*¬catch*	*catch*	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008	0.108	0.012	0.072	0.008	0.108	0.012	0.072	0.008	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576	0.016	0.064	0.144	0.576	0.016	0.064	0.144	0.576	0.016	0.064	0.144	0.576

- The 32-element table can be reduced to a 8-element table and a 4-element table

	toothache		*¬toothache*	
	catch	*¬catch*	*catch*	*¬catch*
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

Recap: *Conditional independence to reduce table size*

- For *n* effects that are conditionally independent given the cause, the **full joint distribution** table size grows as *O(n)* instead of *O(2^n)*

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

Recap: *Bayes' rule*

- Derive **Bayes' rule** from the **product rule** of conditional probability

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

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

- Equating the right-hand sides and dividing by $P(a)$

$$\boxed{P(b \mid a) = \frac{P(a \mid b) P(b)}{P(a)}}$$

Recap: *Bayes' rule*

- Derive **Bayes' rule** from the **product rule** of conditional probability

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

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

- Equating the right-hand sides and dividing by $P(a)$

$$\boxed{P(b \mid a) = \frac{P(a \mid b)P(b)}{P(a)}}$$

This equation underlies most **modern AI systems** for **probabilistic inference…**

Recap: *Bayesian networks*

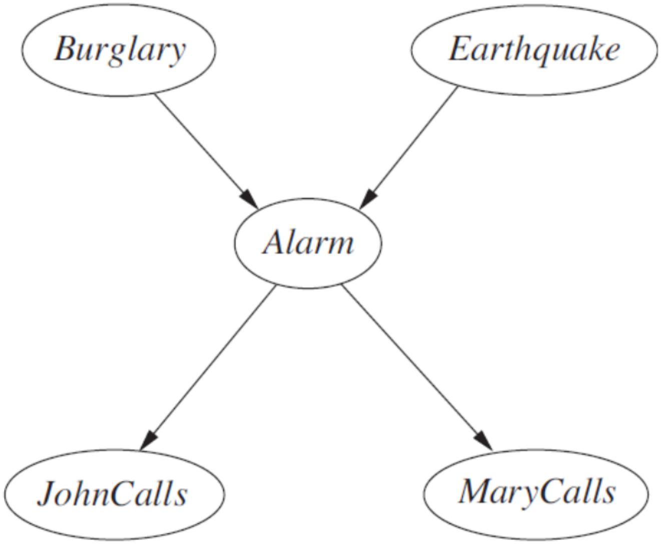
- **Full joint distribution** can be used to answer any query about the world
  - But table size is exponential in the number of variables
  - Independence relations are helpful, but can be unnatural and tedious to specify

- **Bayesian networks** is a data structure to represent both **joint distributions** and **dependencies** among variables

Recap: *Bayesian networks (definition)*

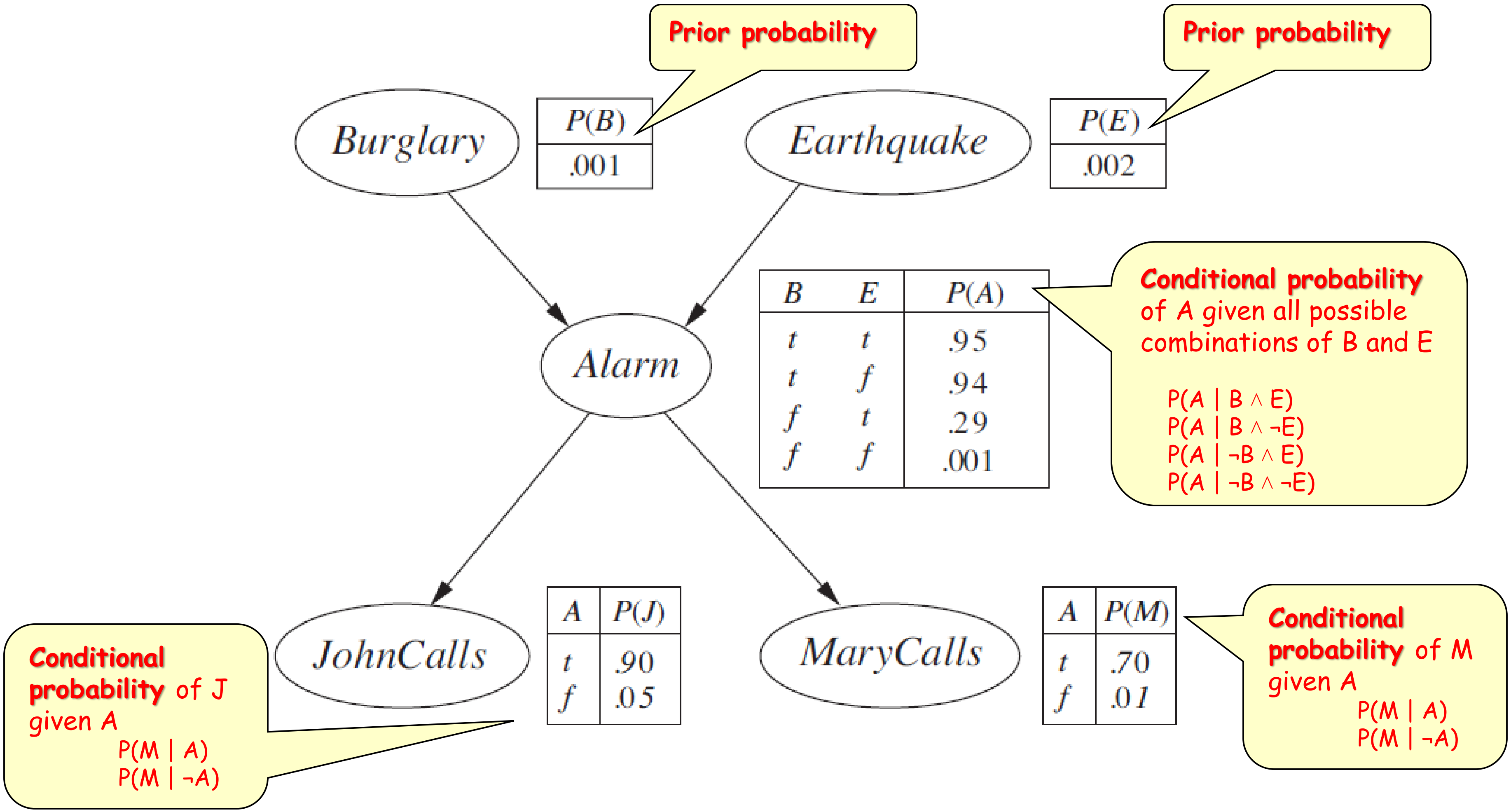
- A directed acyclic graph (DAG) where
  - Each node corresponds to a random variable,
  - Each edge from node X to node Y represents a direct influence of X on Y,
  - Each node X_i has a conditional probability distribution $P(X_i \mid Parents(X_i))$ that quantifies the effect of the parents on the node.

- Example



Recap: *Bayesian networks (example)*

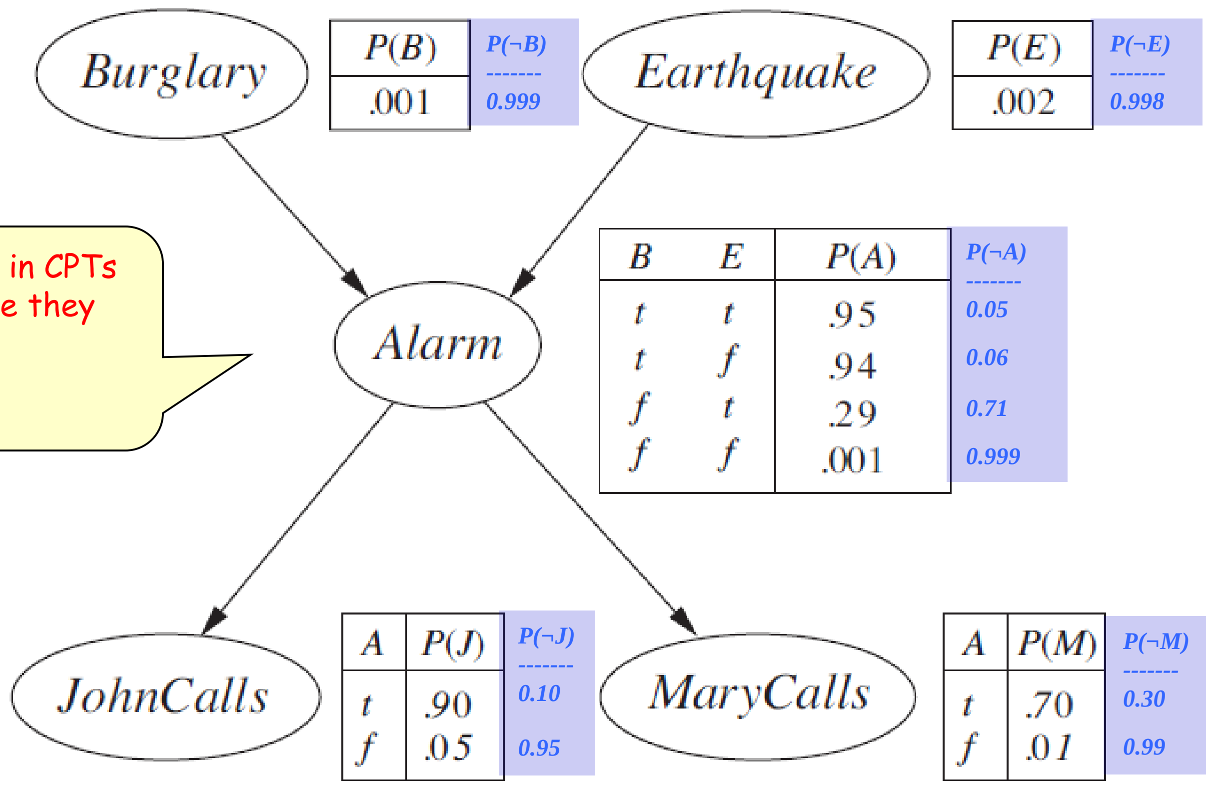
- Both the **topology** and the **conditional probability tables** (CPTs)



Recap: *Bayesian networks (semantics)*

- Both the **topology** and the **conditional probability tables** (CPTs)

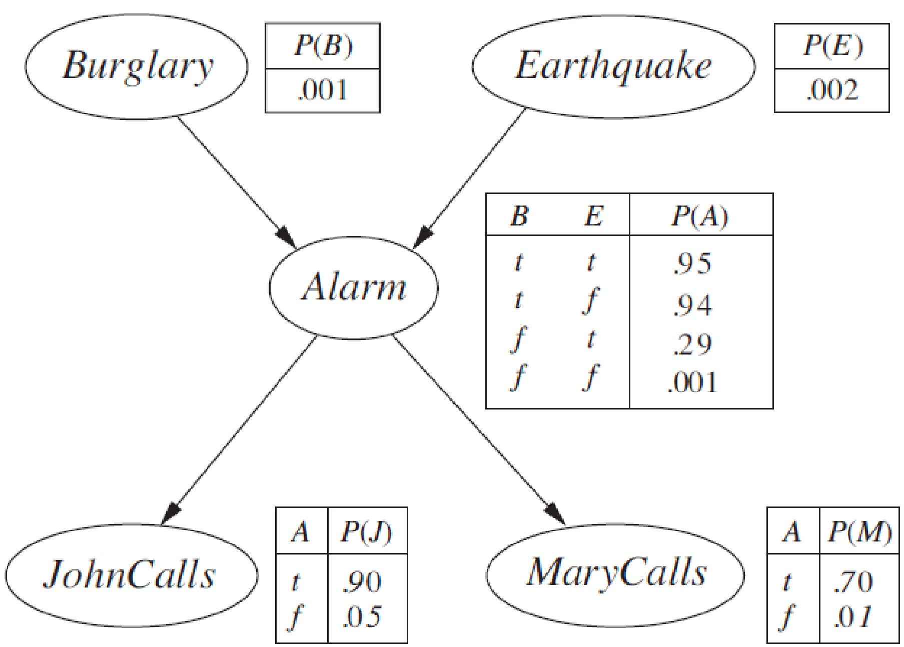
Half of the entries in CPTs are omitted because they can be inferred



Quiz 5: solution

- Each entry $P(x_1, \ldots, x_n)$ in the full joint distribution, which is the abbreviation of $P(X_1{=}x_1 \wedge \ldots \wedge X_n{=}x_n)$ is the product of the elements of the CPTs defined as follows:

$$\boxed{P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid parents(X_i))}$$



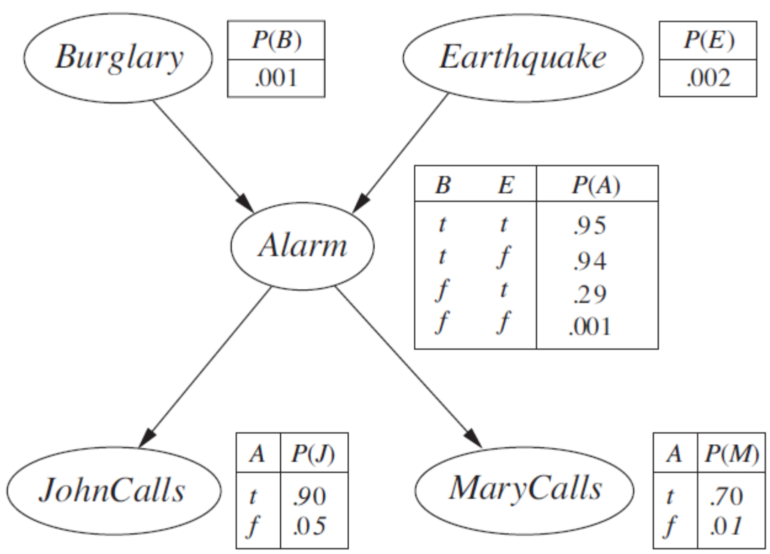
$P(j, m, a, b, e) = ?$

$P(\neg m, j, \neg a, \neg e, b) = ?$

Quiz 5: solution

- Each entry $P(x_1,\ldots,x_n)$ in the full joint distribution, which is the abbreviation of $P(X_1{=}x_1 \wedge \ldots \wedge X_n{=}x_n)$ is the product of the elements of the CPTs defined as follows:

$$\boxed{P(x_1,\ldots,x_n) = \prod_{i=1}^{n} P(x_i \,|\, parents(X_i))}$$



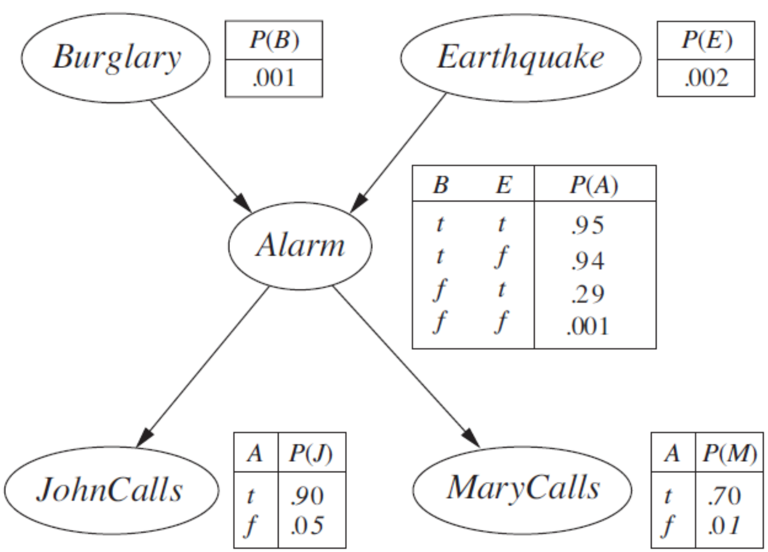
$P(j, m, a, b, e)$
$= P(j|a)\, P(m|a)\, P(a|b,e)\, \mathbf{P(b)\, P(e)}$
$= 0.90*0.70*0.95*\mathbf{0.001*0.002}$
$= 0.000001197$

$P(\neg m, j, \neg a, \neg e, b) = ?$

Quiz 5: solution

- Each entry $P(x_1,\ldots,x_n)$ in the full joint distribution, which is the abbreviation of $P(X_1=x_1 \wedge \ldots \wedge X_n=x_n)$ is the product of the elements of the CPTs defined as follows:

$$\boxed{P(x_1,\ldots,x_n) = \prod_{i=1}^{n} P(x_i \mid parents(X_i))}$$



$P(j, m, a, b, e) = 0.000001197$

$P(\neg m, j, \neg a, \neg e, b)$
$= P(\neg m|\neg a)P(j|\neg a)P(\neg a|b,\neg e)\, \mathbf{P(b)P(\neg e)}$
$=(1-0.01)*0.05*(1-0.94)*\mathbf{0.001*(1-0.002)}$
$=0.00000296406$

Outline

- Representing knowledge in Bayesian networks
- Semantics of Bayesian networks
- Efficient representation of conditional distributions
- Exact inference in Bayesian networks

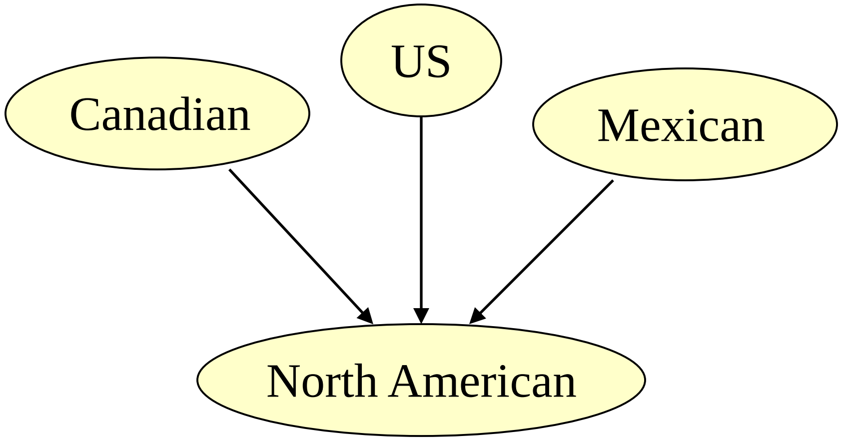
Efficient representation

- Conditional probability table (CPT) in Bayesian networks requires $O(2^k)$ numbers
  - Is there a more compact representation?
  - The answer is often “yes”

Efficient representation

- Conditional probability table (CPT) in Bayesian networks requires $O(2^k)$ numbers
  - Is there a more compact representation?
  - The answer is often “yes”

- Example #1:
  - **Deterministic node**: value determined by parents, with certainty



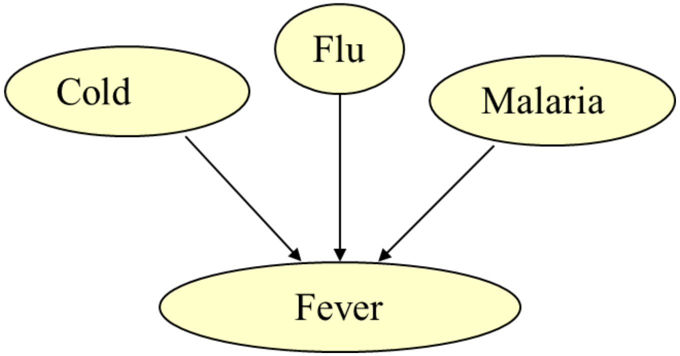
Canadian	*US*	*Mexican*	*N. A.*
…	…	…	…
…	…	T	T
…	…	…	…
…	T	…	T
…	…	…	…
T	…	…	T
…	…	…	…

NA = (Cananian ∨ US ∨ Mexican)

Efficient representation (2)

- What about uncertain relationships?
  – Noisy logical relationships (as opposed to deterministic logic)

- Example #2:
  – **Noisy-OR**: a generalization of the logical OR



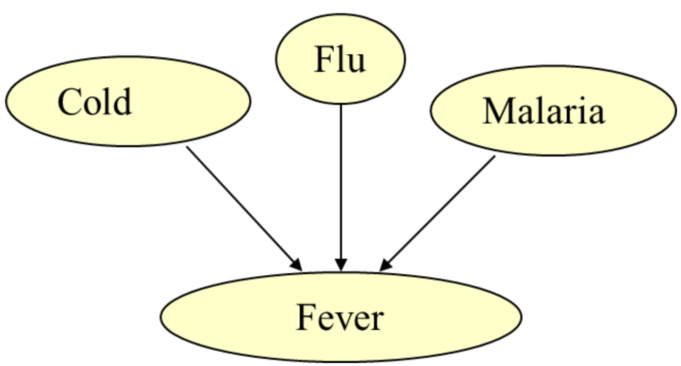
Cold	*Flu*	*Malaria*	*P(Fever)*
F	F	F	0.0
F	F	T	0.9
F	T	F	0.8
F	T	T	0.98
T	F	F	0.4
T	F	T	0.94
T	T	F	0.88
T	T	T	0.988

  – ***Assumption***: *inhibitions of the causal relationships are independent*

Efficient representation (2)

- What about uncertain relationships?
  - Noisy logical relationships (as opposed to deterministic logic)

- Example #2:
  - **Noisy-OR**: a generalization of the logical OR



Cold	*Flu*	*Malaria*	$P(Fever)$	$P(\neg Fever)$
F	F	F	0.0	
F	F	T	0.9	**0.1**
F	T	F	0.8	**0.2**
F	T	T	0.98	
T	F	F	0.4	**0.6**
T	F	T	0.94	
T	T	F	0.88	
T	T	T	0.988	

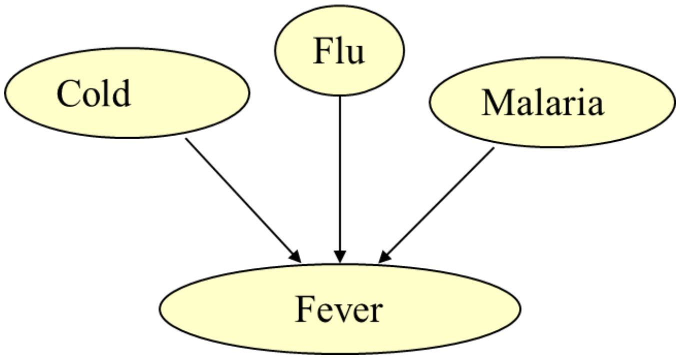
- *Assumption: inhibitions of the causal relationships are independent*

Efficient representation (2)

$$P(x_i \mid parents(X_i)) = 1 - \prod_{\{j: X_j = true\}} q_j$$

$$q_{\text{cold}} = P(\neg fever \mid cold, \neg flu, \neg malaria) = 0.6\ ,$$
$$q_{\text{flu}} = P(\neg fever \mid \neg cold, flu, \neg malaria) = 0.2\ ,$$
$$q_{\text{malaria}} = P(\neg fever \mid \neg cold, \neg flu, malaria) = 0.1\ .$$

- Example #2:
  - **Noisy-OR**: a generalization of the logical OR



$Cold$	Flu	$Malaria$	$P(Fever)$	$P(\neg Fever)$
F	F	F		
F	F	T	0.9	**0.1**
F	T	F	0.8	**0.2**
F	T	T		
T	F	F	0.4	**0.6**
T	F	T		
T	T	F		
T	T	T		

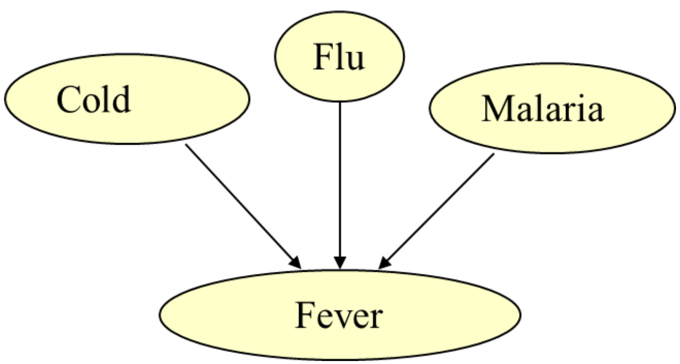
  - *Assumption: inhibitions of the causal relationships are independent*

Efficient representation (2)

$$P(x_i \mid parents(X_i)) = 1 - \prod_{\{j : X_j = true\}} q_j$$

$$q_{\text{cold}} = P(\neg fever \mid cold, \neg flu, \neg malaria) = 0.6\ ,$$
$$q_{\text{flu}} = P(\neg fever \mid \neg cold, flu, \neg malaria) = 0.2\ ,$$
$$q_{\text{malaria}} = P(\neg fever \mid \neg cold, \neg flu, malaria) = 0.1\ .$$

- Example #2:
  - **Noisy-OR**: a generalization of the logical OR

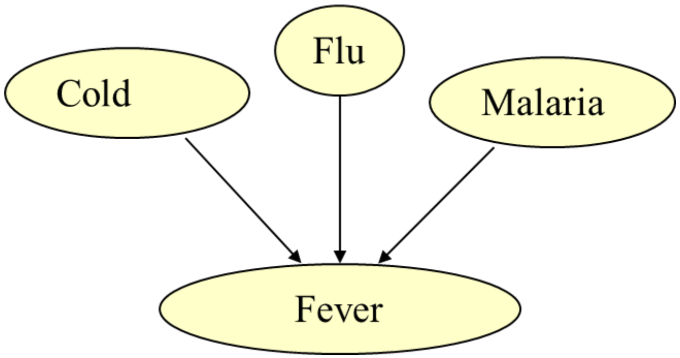


Cold	*Flu*	*Malaria*	$P(Fever)$	$P(\neg Fever)$
F	F	F	0.0	1.0
F	F	T	0.9	**0.1**
F	T	F	0.8	**0.2**
F	T	T	0.98	$0.02 = 0.2 \times 0.1$
T	F	F	0.4	**0.6**
T	F	T	0.94	$0.06 = 0.6 \times 0.1$
T	T	F	0.88	$0.12 = 0.6 \times 0.2$
T	T	T	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

  - *Assumption: inhibitions of the causal relationships are independent*

Efficient representation (2)

- Example #2:
  - **Noisy-OR**: a generalization of the logical OR



$$q_{\text{cold}} = P(\neg fever \mid cold, \neg flu, \neg malaria) = 0.6\ ,$$
$$q_{\text{flu}} = P(\neg fever \mid \neg cold, flu, \neg malaria) = 0.2\ ,$$
$$q_{\text{malaria}} = P(\neg fever \mid \neg cold, \neg flu, malaria) = 0.1\ .$$

$$\boxed{P(x_i \mid parents(X_i)) = 1 - \prod_{\{j: X_j = true\}} q_j}$$

  - *Assumption: inhibitions of the causal relationships are independent*

Outline

- Representing knowledge in Bayesian networks
- Semantics of Bayesian networks
- Efficient representation of conditional distributions
- Exact inference in Bayesian networks

Basic task

- Computing the **posterior probability distribution**, given some **observed event**
  - ***Query*** variables, ***Evidence*** variables, and ***Hidden*** variables

$Earthquake$, $Alarm$

$$\mathbf{P}(Burglary \mid JohnCalls = true, MaryCalls = true) = \langle 0.284, 0.716 \rangle$$

How to compute it?

Algorithms for probabilistic inference

- Two options
  - Inference by enumeration (basic)
  - Variable elimination (improved)

Inference by enumeration

- Computing **posterior probability distribution**, given **observed event**
  - ***Query*** variables, ***Evidence*** variables, and ***Hidden*** variables

$$\mathbf{P}(B \mid j, m) = \frac{\mathbf{P}(B, j, m)}{P(j, m)}$$

$$= \alpha \ \mathbf{P}(B, j, m)$$

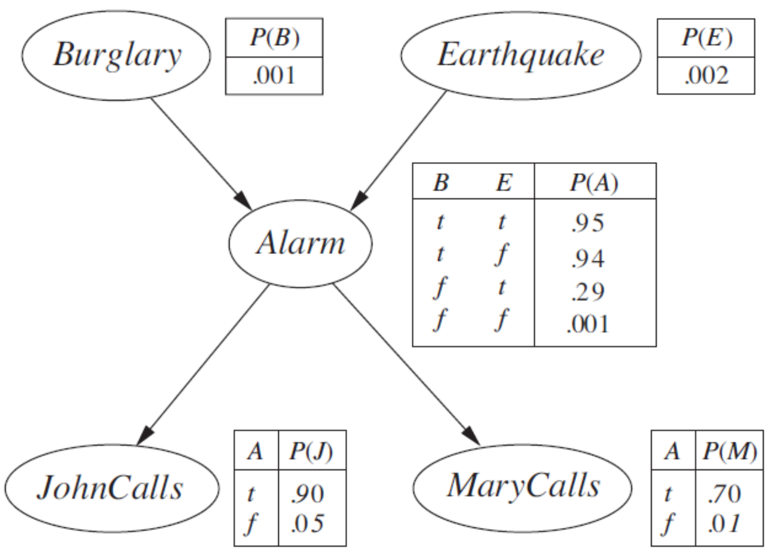
$$= \alpha \ \sum_e \sum_a \mathbf{P}(B, j, m, e, a)$$

$$= \langle \alpha \sum_e \sum_a P(b, j, m, e, a) \ , \ \alpha \sum_e \sum_a P(\neg b, j, m, e, a) \rangle$$

We knew how to compute joint distribution…

- Each entry $P(x_1,\ldots,x_n)$ in the full joint distribution, which is the abbreviation of $P(X_1=x_1 \wedge \ldots \wedge X_n=x_n)$ is the product of elements of CPTs defined as follows:

$$P(x_1,\ldots,x_n) = \prod_{i=1}^{n} P(x_i \mid parents(X_i))$$



$P(j, m, a, b, e)$
$= P(j|a)\, P(m|a)\, P(a|b,e)\, \mathbf{P(b)\, P(e)}$
$= 0.90*0.70*0.95*\mathbf{0.001*0.002}$
$= 0.000001197$

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

Inference by enumeration

- Computing **posterior probability distribution**, given **observed event**
  - ***Query*** variables, ***Evidence*** variables, and ***Hidden*** variables

$$P(b \mid j, m) = \alpha \sum_{e} \sum_{a} P(b) P(e) P(a \mid b, e) P(j \mid a) P(m \mid a)$$

5 -- the number of **all variables**

$\mathbf{2^2 = 4}$ -- the number of possible combinations of **hidden variables**

$\mathbf{2^2 * 5}$ -- in general, the computational cost can be $O(n\,2^n)$

Inference by enumeration *(improvement)*

- Computing **posterior probability distribution**, given **observed event**
  - ***Query*** variables, ***Evidence*** variables, and ***Hidden*** variables

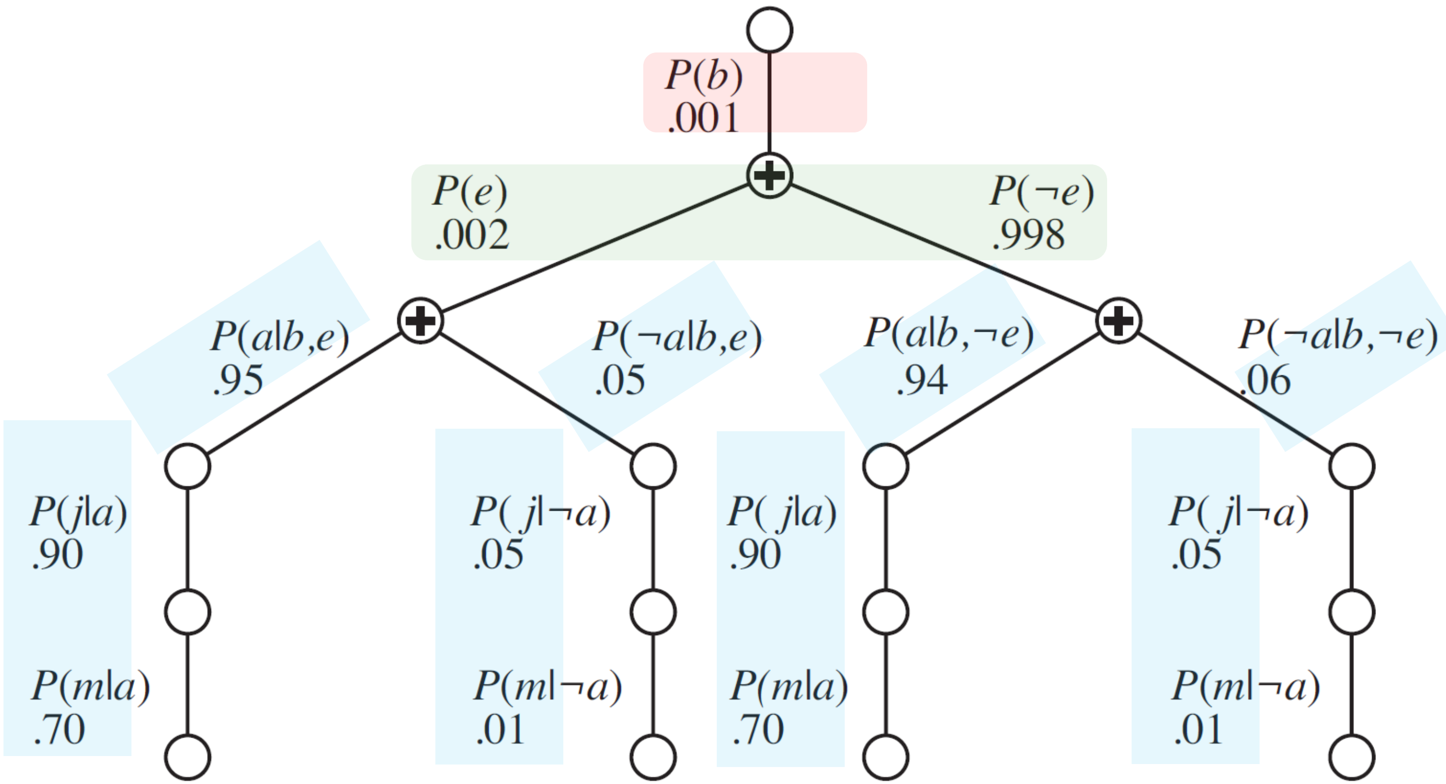
$$P(b \mid j, m) = \alpha \sum_e \sum_a P(b) P(e) P(a \mid b, e) P(j \mid a) P(m \mid a)$$

$$P(b \mid j, m) = \alpha P(b) \sum_e P(e) \sum_a P(a \mid b, e) P(j \mid a) P(m \mid a)$$

Inference by enumeration *(improvement)*

- Recursive evaluation of the expression

$$P(b \mid j, m) = \alpha\, P(b) \sum_e P(e) \sum_a P(a \mid b, e) P(j \mid a) P(m \mid a)$$

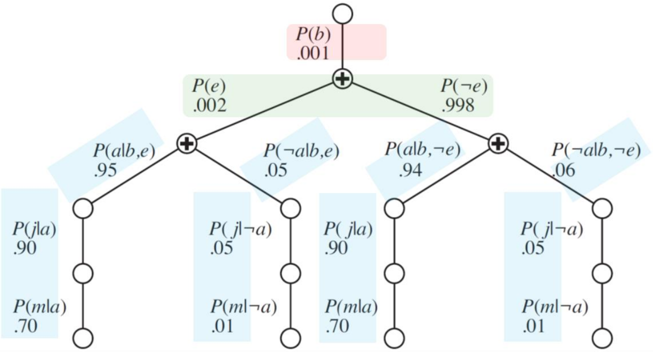


Enumeration Algorithm

function ENUMERATION-ASK(X, **e**, bn) **returns** a distribution over X
  inputs: X, the query variable
    e, observed values for variables **E**
    bn, a Bayes net with variables $\{X\} \cup \mathbf{E} \cup \mathbf{Y}$ /* **Y** = *hidden variables* */

  $\mathbf{Q}(X) \leftarrow$ a distribution over X, initially empty
  for each value x_i of X **do**
    $\mathbf{Q}(x_i) \leftarrow$ ENUMERATE-ALL(bn.VARS, $\mathbf{e}_{x_i}$)
      where $\mathbf{e}_{x_i}$ is **e** extended with $X = x_i$
  return NORMALIZE($\mathbf{Q}(X)$)

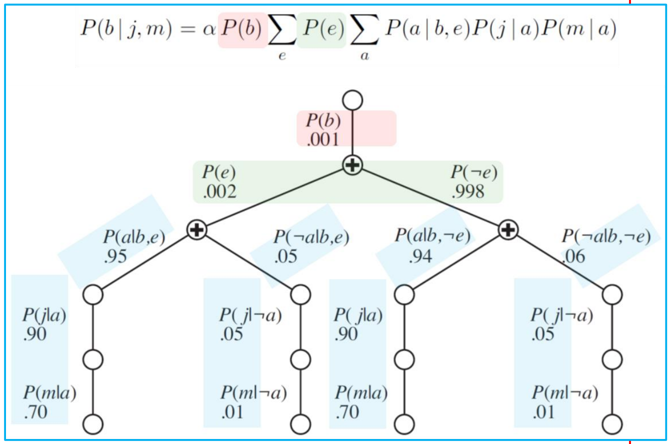
$$P(b \mid j, m) = \alpha\, P(b) \sum_e P(e) \sum_a P(a \mid b, e) P(j \mid a) P(m \mid a)$$



Enumeration Algorithm

function ENUMERATION-ASK(X, **e**, bn) **returns** a distribution over X
  inputs: X, the query variable
    e, observed values for variables **E**
    bn, a Bayes net with variables $\{X\} \cup \mathbf{E} \cup \mathbf{Y}$ /* **Y** = *hidden variables* */

  $\mathbf{Q}(X) \leftarrow$ a distribution over X, initially empty
  for each value x_i of X **do**
    $\mathbf{Q}(x_i) \leftarrow$ ENUMERATE-ALL(bn.VARS, $\mathbf{e}_{x_i}$)
      where $\mathbf{e}_{x_i}$ is **e** extended with $X = x_i$
  return NORMALIZE($\mathbf{Q}(X)$)

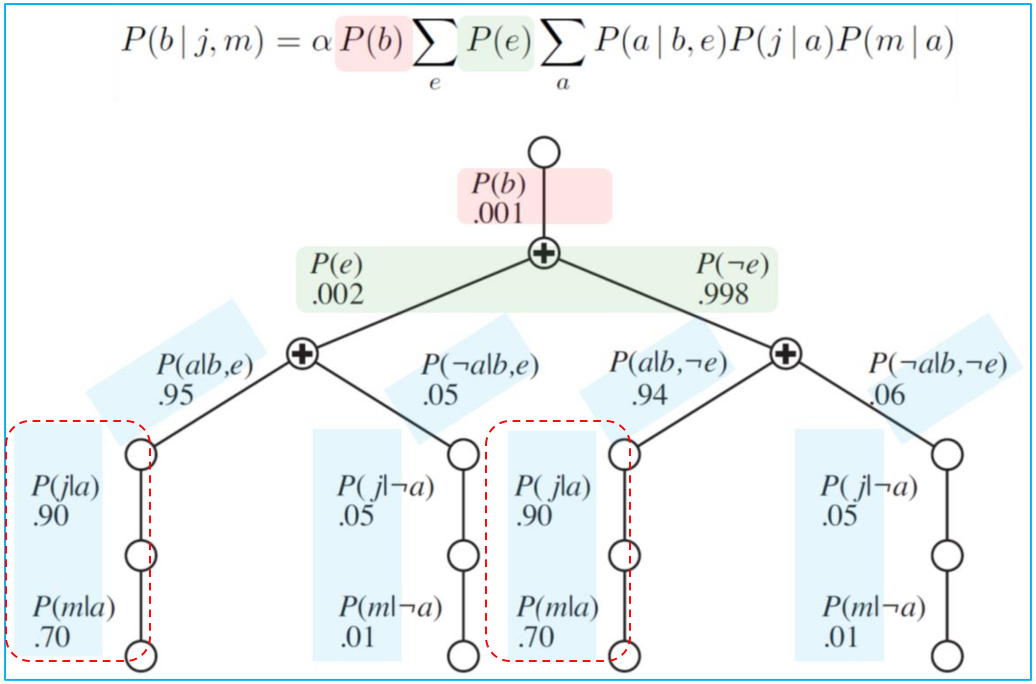


function ENUMERATE-ALL($vars$, **e**) **returns** a real number
  if EMPTY?($vars$) **then return** 1.0
  $Y \leftarrow$ FIRST($vars$)
  if Y has value y in **e**
    then return $P(y \mid parents(Y)) \times$ ENUMERATE-ALL(REST($vars$), **e**)
    else return $\sum_y P(y \mid parents(Y)) \times$ ENUMERATE-ALL(REST($vars$), $\mathbf{e}_y$)
      where $\mathbf{e}_y$ is **e** extended with $Y = y$

Algorithms for probabilistic inference

- Two options
  - Inference by enumeration (basic)
  - Variable elimination (improved)

There may be repeated computation



Variable elimination

$$\mathbf{P}(B \mid j,m) = \alpha \underbrace{\mathbf{P}(B)}_{\mathbf{f}_1(B)} \sum_e \underbrace{P(e)}_{\mathbf{f}_2(E)} \sum_a \underbrace{\mathbf{P}(a \mid B,e)}_{\mathbf{f}_3(A,B,E)} \underbrace{P(j \mid a)}_{\mathbf{f}_4(A)} \underbrace{P(m \mid a)}_{\mathbf{f}_5(A)}$$

Eliminating A …

$$\begin{aligned} \mathbf{f}_6(B,E) &= \sum_a \mathbf{f}_3(A,B,E) \times \mathbf{f}_4(A) \times \mathbf{f}_5(A) \\ &= (\mathbf{f}_3(a,B,E) \times \mathbf{f}_4(a) \times \mathbf{f}_5(a)) + (\mathbf{f}_3(\neg a,B,E) \times \mathbf{f}_4(\neg a) \times \mathbf{f}_5(\neg a)) \,. \end{aligned}$$

$$\mathbf{P}(B \mid j,m) = \alpha\, \mathbf{f}_1(B) \times \sum_e \mathbf{f}_2(E) \times \mathbf{f}_6(B,E)$$

Eliminating E…

$$\begin{aligned} \mathbf{f}_7(B) &= \sum_e \mathbf{f}_2(E) \times \mathbf{f}_6(B,E) \\ &= \mathbf{f}_2(e) \times \mathbf{f}_6(B,e) + \mathbf{f}_2(\neg e) \times \mathbf{f}_6(B,\neg e) \,. \end{aligned}$$

$$\mathbf{P}(B \mid j,m) = \alpha\, \mathbf{f}_1(B) \times \mathbf{f}_7(B)$$

Variable elimination *(with a different order)*

$$\mathbf{P}(B \mid j,m) = \alpha\, \underbrace{\mathbf{P}(B)}_{\mathbf{f}_1(B)} \sum_{e} \underbrace{P(e)}_{\mathbf{f}_2(E)} \sum_{a} \underbrace{\mathbf{P}(a \mid B,e)}_{\mathbf{f}_3(A,B,E)} \underbrace{P(j \mid a)}_{\mathbf{f}_4(A)} \underbrace{P(m \mid a)}_{\mathbf{f}_5(A)}$$

Equivalent transformation

$$\mathbf{P}(B \mid j,m) = \alpha\, \mathbf{f}_1(B) \times \sum_{a} \mathbf{f}_4(A) \times \mathbf{f}_5(A) \times \sum_{e} \mathbf{f}_2(E) \times \mathbf{f}_3(A,B,E)$$

Eliminating E first…

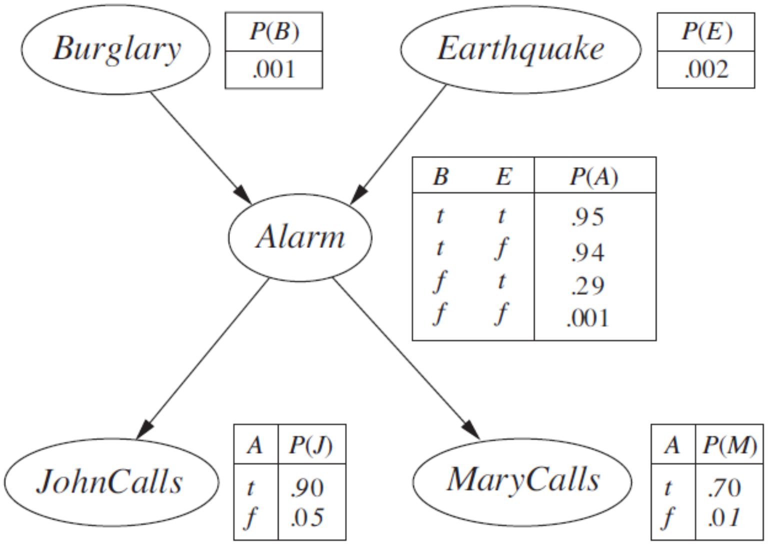
Eliminating A next…

Variable elimination *(another example)*

- Consider $\mathbf{P}(JohnCalls \mid Burglary=true)$

$$\boxed{\mathbf{P}(J \mid b) = \alpha\, P(b) \sum_e P(e) \sum_a P(a \mid b, e) \mathbf{P}(J \mid a) \sum_m P(m \mid a)}$$

This is always 100%

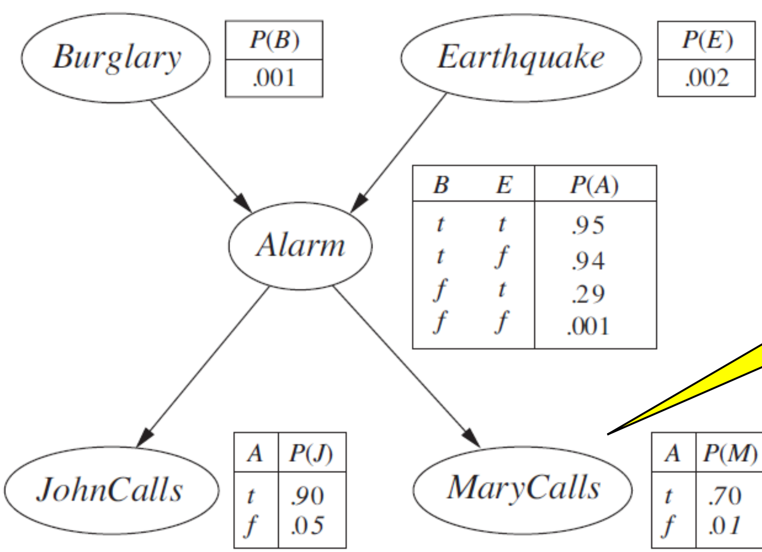


Variable elimination *(another example)*

- Consider $\mathbf{P}$ (*JohnCalls* | *Burglary=true*)

$$\mathbf{P}(J \mid b) = \alpha\, P(b) \sum_{e} P(e) \sum_{a} P(a \mid b, e) \mathbf{P}(J \mid a)$$

We can remove any variable that is not an ancestor of a "query" or "evidence" variable

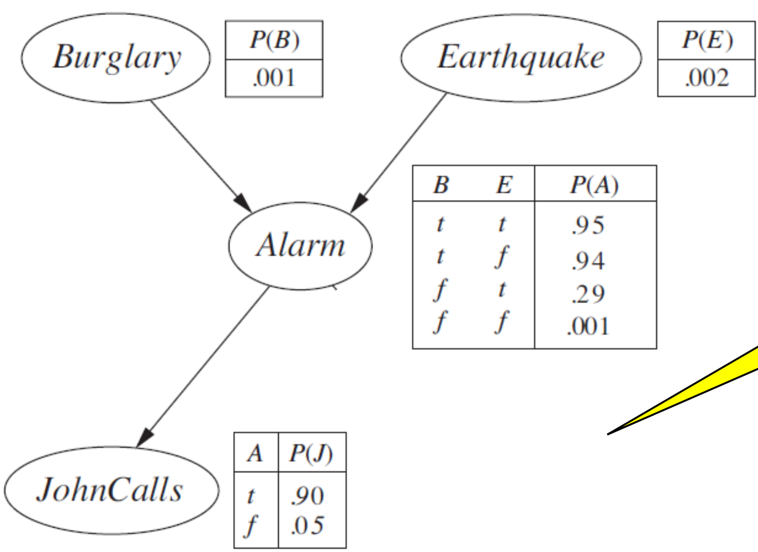


Variable elimination *(another example)*

- Consider $\mathbf{P}(\textit{JohnCalls} \mid \textit{Burglary=true})$

$$\mathbf{P}(J \mid b) = \alpha\, P(b) \sum_{e} P(e) \sum_{a} P(a \mid b, e)\mathbf{P}(J \mid a)$$

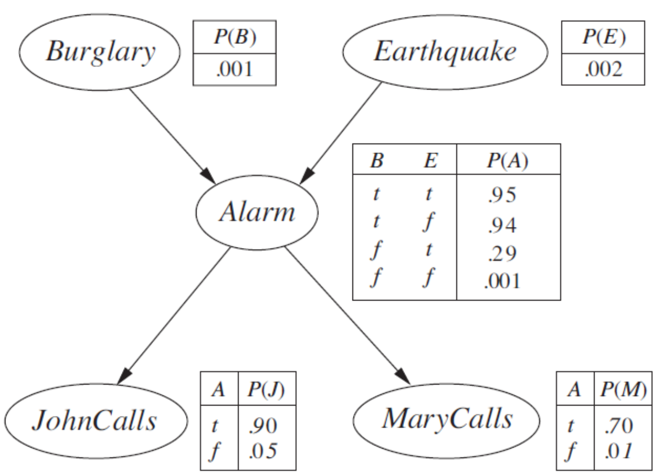
We can remove any variable that is not an ancestor of a "query" or "evidence" variable



Complexity of exact inference

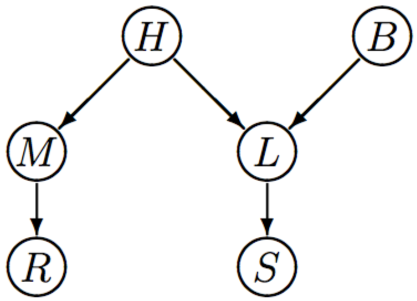
- In general it has **exponential time** and **space** complexity
  - NP-hard

- Special case: ***singly connected*** networks (or ***polytrees***)
  - Any two nodes are connected by at most one (undirected) path
  - Linear in the size of the network



Example for singly connected network

Yet another example



H = hardware problem
B = bug in lab
M = mailer is running
L = lab is running
R = received email
S = problem solved

Each node needs a probability table. Size of table depends on number of parents.

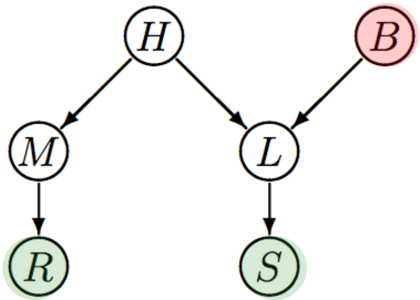
$\mathbf{P}(H)$	
$True$	$False$
0.01	0.99

	$\mathbf{P}(M \mid H)$	
H	$True$	$False$
$True$	0.1	0.9
$False$	0.99	0.01

		$\mathbf{P}(L \mid H, B)$	
H	B	$True$	$False$
$True$	$True$	0.01	0.99
$True$	$False$	0.1	0.9
$False$	$True$	0.02	0.98
$False$	$False$	1.0	0.0

$\ldots$, etc.

Yet another example



H = hardware problem
B = bug in lab
M = mailer is running
L = lab is running
R = received email
S = problem solved

- Compute $\mathbf{P}(B \mid \neg R, S)$

Each node needs a probability table. Size of table depends on number of parents.

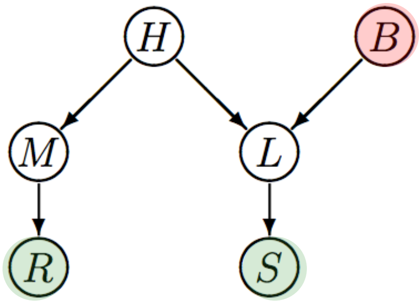
$\mathbf{P}(H)$	
$True$	$False$
0.01	0.99

	$\mathbf{P}(M \mid H)$	
H	$True$	$False$
$True$	0.1	0.9
$False$	0.99	0.01

		$\mathbf{P}(L \mid H, B)$	
H	B	$True$	$False$
$True$	$True$	0.01	0.99
$True$	$False$	0.1	0.9
$False$	$True$	0.02	0.98
$False$	$False$	1.0	0.0

…, etc.

Yet another example



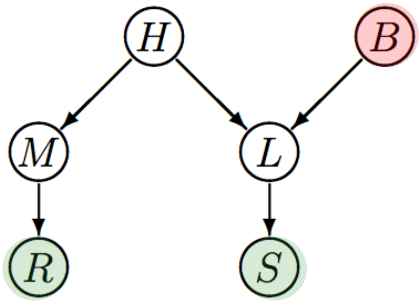
H = hardware problem
B = bug in lab
M = mailer is running
L = lab is running
R = received email
S = problem solved

- Compute **P** (B | ¬R,S)

1. Apply the conditional probability rule.

$$P(B \mid \neg R, S) = \frac{P(B, \neg R, S)}{P(\neg R, S)}$$

Yet another example



H = hardware problem
B = bug in lab
M = mailer is running
L = lab is running
R = received email
S = problem solved

- Compute **P** (B | $\neg R, S$)

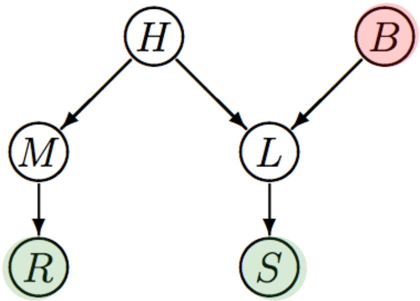
1. Apply the conditional probability rule.

$$P(B \mid \neg R, S) = \frac{P(B, \neg R, S)}{P(\neg R, S)}$$

2. Apply the marginal distribution rule to the unknown vertices. $P(B, \neg R, S)$ has 3 unknown vertices with $2^3 = 8$ possible value assignments.

$$\begin{aligned}
&P(B, \neg R, S) \\
&= P(B, \neg R, S, H, M, L) \\
&\quad + P(B, \neg R, S, H, M, \neg L) \\
&\quad + P(B, \neg R, S, H, \neg M, L) \\
&\quad + P(B, \neg R, S, H, \neg M, \neg L) \\
&\quad + P(B, \neg R, S, \neg H, M, L) \\
&\quad + P(B, \neg R, S, \neg H, M, \neg L) \\
&\quad + P(B, \neg R, S, \neg H, \neg M, L) \\
&\quad + P(B, \neg R, S, \neg H, \neg M, \neg L)
\end{aligned}$$

Yet another example



H = hardware problem
B = bug in lab
M = mailer is running
L = lab is running
R = received email
S = problem solved

- Compute **P** (B | $\neg R, S$)

1. Apply the conditional probability rule.

$$P(B \mid \neg R, S) = \frac{P(B, \neg R, S)}{P(\neg R, S)}$$

2. Apply the marginal distribution rule to the unknown vertices. $P(B, \neg R, S)$ has 3 unknown vertices with $2^3 = 8$ possible value assignments.

$$\begin{aligned} &P(B, \neg R, S) \\ &= P(B, \neg R, S, H, M, L) \\ &\quad + P(B, \neg R, S, H, M, \neg L) \\ &\quad + P(B, \neg R, S, H, \neg M, L) \\ &\quad + P(B, \neg R, S, H, \neg M, \neg L) \\ &\quad + P(B, \neg R, S, \neg H, M, L) \\ &\quad + P(B, \neg R, S, \neg H, M, \neg L) \\ &\quad + P(B, \neg R, S, \neg H, \neg M, L) \\ &\quad + P(B, \neg R, S, \neg H, \neg M, \neg L) \end{aligned}$$

3. Apply joint distribution rule for Bayesian networks.

$$\begin{aligned} &P(B, \neg R, S, H, M, L) \\ &= P(B)\,P(H) \\ &\quad P(M \mid H)\,P(\neg R \mid M) \\ &\quad P(L \mid H,B)\,P(S \mid L) \end{aligned}$$

Outline

- Representing knowledge in Bayesian networks
- Semantics of Bayesian networks
- Efficient representation of conditional distributions
- Exact inference in Bayesian networks

Quiz 6