Artificial Intelligence

13. Bayesian Networks

Shashi Prabh

School of Engineering and Applied Science
Ahmedabad University

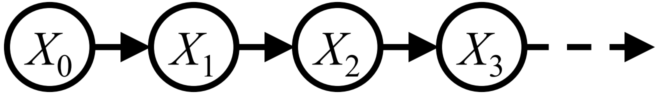
Uncertainty and Time

- Often, we want to reason about a sequence of observations where the state of the underlying system is changing
    - Speech recognition
    - Robot localization
    - User attention
    - Medical monitoring
    - Global climate

- Need to introduce time into our models

Markov Models (aka Markov chain/process)

- Value of X at a given time is called the state (usually discrete, finite)

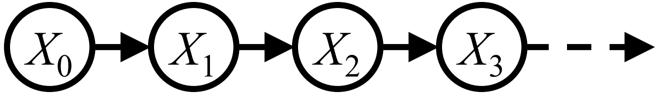


$P(X_0)$ $P(X_t \mid X_{t-1})$

- Transition model: $P(X_t \mid X_{t-1})$ how the state evolves over time
- Stationarity assumption: transition probabilities are the same at all times

Markov Models (aka Markov chain/process)

- Markov assumption: "future is independent of the past given the present"



$P(X_0)$ $P(X_t \mid X_{t-1})$

  - X_{t+1} is independent of $X_0, \ldots, X_{t-1}$ given X_t
  - This is a first-order Markov model (a kth-order model allows dependencies on k earlier steps)
- Joint distribution $P(X_0, \ldots, X_T) = P(X_0) \prod_t P(X_t \mid X_{t-1})$

Are Markov models a special case of Bayes nets?

- Yes and no!
- Yes:
  - Directed acyclic graph, joint = product of conditionals
- No:
  - Infinitely many variables (unless we truncate)
  - Repetition of transition model not part of standard Bayes net syntax

Reminder: elementary probability

- Basic laws: $0 \leq P(\omega) \leq 1$ $\sum_{\omega \in \Omega} P(\omega) = 1$
- Events: subsets of Ω: $P(A) = \sum_{\omega \in A} P(\omega)$
- Random variable $X(\omega)$ has a value in each ω
  - Distribution $P(X)$ gives probability for each possible value x
  - Joint distribution $P(X,Y)$ gives total probability for each combination x,y
- Summing out/marginalization: $P(X=x) = \sum_y P(X=x,Y=y)$
- Conditional probability: $P(X|Y) = P(X,Y)/P(Y)$
- Product rule: $P(X|Y)P(Y) = P(X,Y) = P(Y|X)P(X)$
  - Generalize to chain rule: $P(X_1,\ldots,X_n) = \prod_i P(X_i \mid X_1,\ldots,X_{i-1})$

Probabilistic Inference

- Probabilistic inference: compute a desired probability from a probability model
  - Typically for a query variable given evidence
  - E.g., P(airport on time | no accidents) = 0.90
  - These represent the agent's beliefs given the evidence

- Probabilities change with new evidence
  - P(airport on time | no accidents, 5 a.m.) = 0.95
  - P(airport on time | no accidents, 5 a.m., raining) = 0.80
  - Observing new evidence causes <u>beliefs to be updated</u>

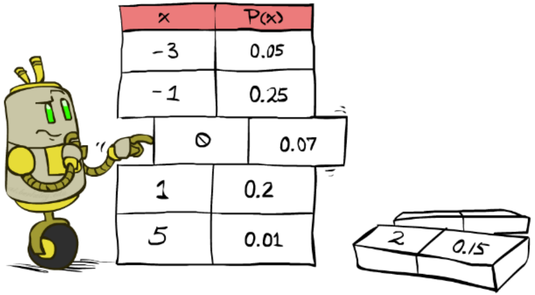
Inference by Enumeration

- Probability model $P(X_1, \ldots, X_n)$ is given
- Partition the variables $X_1, \ldots, X_n$ into sets as follows:
  - Evidence variables: $E = e$
  - Query variables: Q
  - Hidden variables: H

- We want:

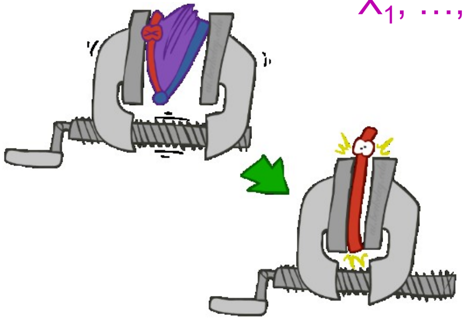
  $P(\mathbf{Q} \mid \mathbf{e})$

- Step 1: Select the entries consistent with the evidence



- Step 2: Sum out **H** from model to get joint of query and evidence

$$P(\mathbf{Q}, \mathbf{e}) = \sum_h \underbrace{P(\mathbf{Q}, \mathbf{h}, \mathbf{e})}_{X_1, \ldots, X_n}$$



- Step 3: Normalize

$$P(\boldsymbol{Q} \mid \boldsymbol{e}) = \alpha P(\boldsymbol{Q}, \boldsymbol{e})$$

Inference by Enumeration

- P(W)?

- P(W | winter)?

- P(W | winter, cold)?

Season	Temp	Weather	P
summer	hot	sun	0.35
summer	hot	rain	0.01
summer	hot	fog	0.01
summer	hot	meteor	0.00
summer	cold	sun	0.01
summer	cold	rain	0.05
summer	cold	fog	0.10
summer	cold	meteor	0.00
winter	hot	sun	0.10
winter	hot	rain	0.01
winter	hot	fog	0.01
winter	hot	meteor	0.00
winter	cold	sun	0.10
winter	cold	rain	0.10
winter	cold	fog	0.15
winter	cold	meteor	0.00

Issues with Inference by Enumeration

- Worst-case time complexity $O(d^n)$
  - exponential in the number of hidden variables
- Space complexity $O(d^n)$ to store the joint distribution
- $O(d^n)$ data points to estimate the entries in the joint distribution

Bayes' Rule

- Write the product rule both ways:

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

- Dividing left and right expressions, we get:

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

- Why is this at all helpful?
  - Lets us build one conditional from its reverse
  - Often one conditional is tricky but the other one is simple
  - Describes an "update" step from prior $P(a)$ to posterior $P(a \mid b)$
  - Foundation of many systems we'll see later (e.g. ASR, MT)
- In the running for most important AI equation!

Inference with Bayes' Rule

- Example: Diagnostic probability from causal probability:

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

- Example:
  - M: meningitis, S: stiff neck

$$\left.\begin{aligned} &P(s \mid m) = 0.8 \\ &P(s \mid \neg m) = 0.01 \\ &P(m) = 0.0001 \end{aligned}\right\} \text{Example givens}$$

$$P(m \mid s) = \frac{P(s \mid m)\, P(m)}{P(s)} \simeq \frac{0.8 \times 0.0001}{0.01}$$

  - Note: posterior probability of meningitis still very small: 0.008 (80x bigger – why?)
  - Note: you should still get stiff necks checked out! Why?

Independence

- Two variables X and Y are (absolutely) independent if

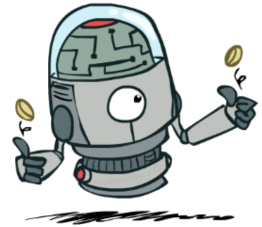
$$\forall x,y \qquad P(x, y) = P(x)\, P(y)$$

  - I.e., the joint distribution factors into a product of two simpler distributions

- Equivalently, via the product rule $P(x,y) = P(x|y)\, P(y)$,

$$P(x \mid y) = P(x) \quad \text{or} \quad P(y \mid x) = P(y)$$

- Example: two dice rolls $Roll_1$ and $Roll_2$
  - $P(Roll_1=5, Roll_2=3) = P(Roll_1=5)\, P(Roll_2=3) = 1/6 \times 1/6 = 1/36$
  - $P(Roll_2=3 \mid Roll_1=5) = P(Roll_2=3)$



Conditional Independence

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

- X is conditionally independent of Y given Z if and only if:

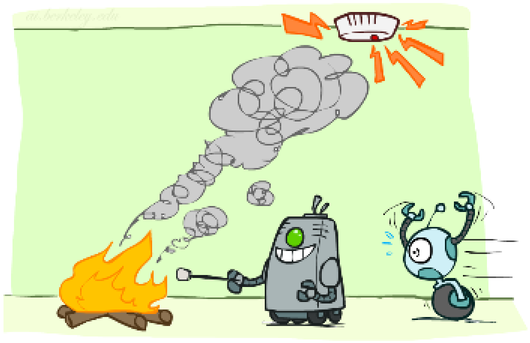
$$\forall x,y,z \qquad P(x \mid y, z) = P(x \mid z)$$

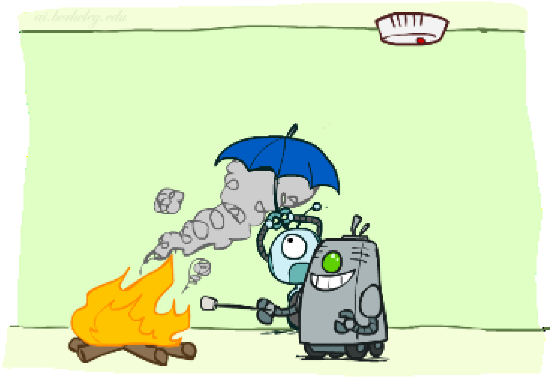
or, equivalently, if and only if

$$\forall x,y,z \qquad P(x, y \mid z) = P(x \mid z)\, P(y \mid z)$$

Conditional Independence

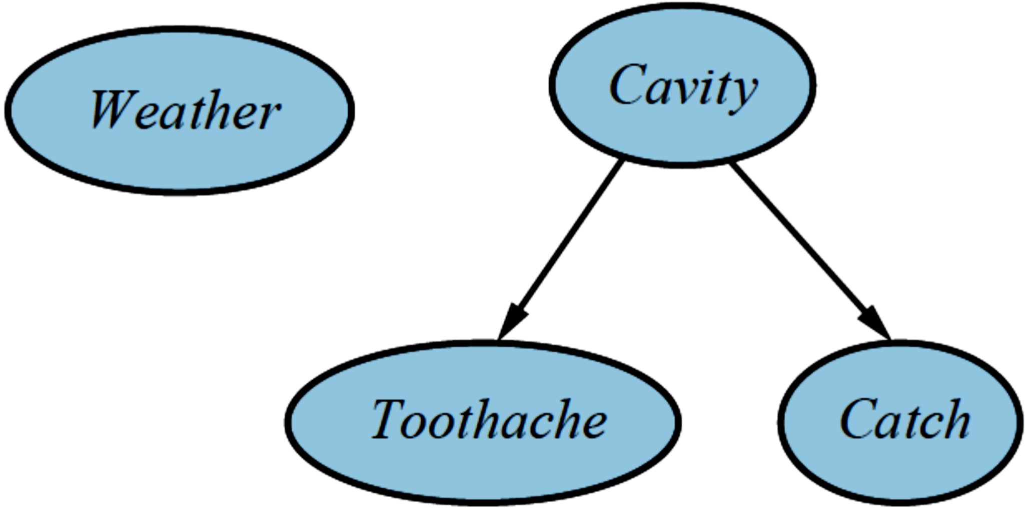
- What about this domain:
    - Fire
    - Smoke
    - Alarm





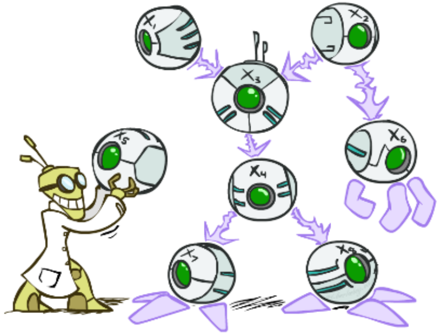
Conditional Independence

- What about this domain:
  - Cavity
  - Toothache
  - Catch
  - Weather



Bayes Nets: Big Picture

- Bayes nets: a technique for describing complex joint distributions (models) using simple, conditional distributions
  - A subset of the general class of graphical models
- Use local causality/conditional independence:
  - the world is composed of many variables,
  - each interacting locally with a few others

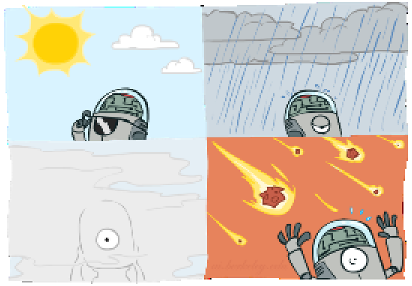


Bayes Nets

- Part I: Representation
- Part II: Exact inference
  - Enumeration (always exponential complexity)
  - Variable elimination (worst-case exponential complexity, often better)
  - Inference is NP-hard in general

Graphical Model Notation

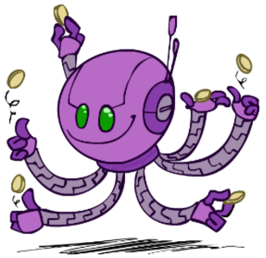
- Nodes: variables (with domains)
  - Can be assigned (observed) or unassigned (unobserved)



- Arcs: interactions
  - Indicate “direct influence” between variables
  - Formally: absence of arc encodes conditional independence (more later)

Example: Coin Flips

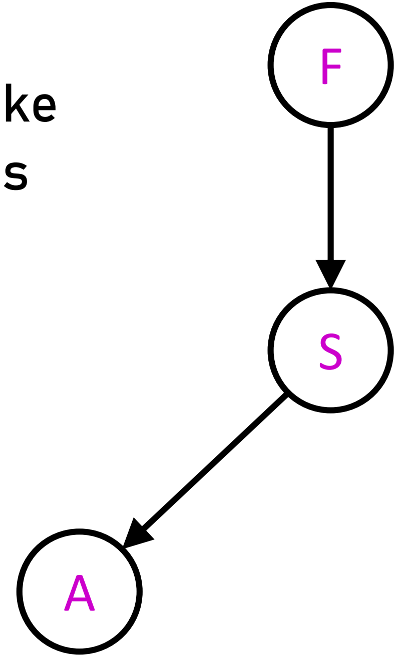
- N independent coin flips

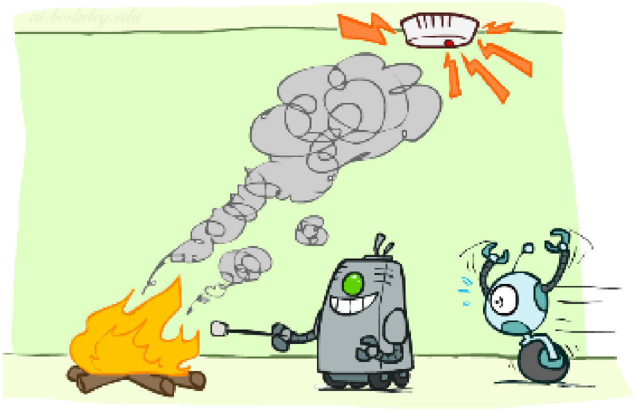


- No interactions between variables: absolute independence

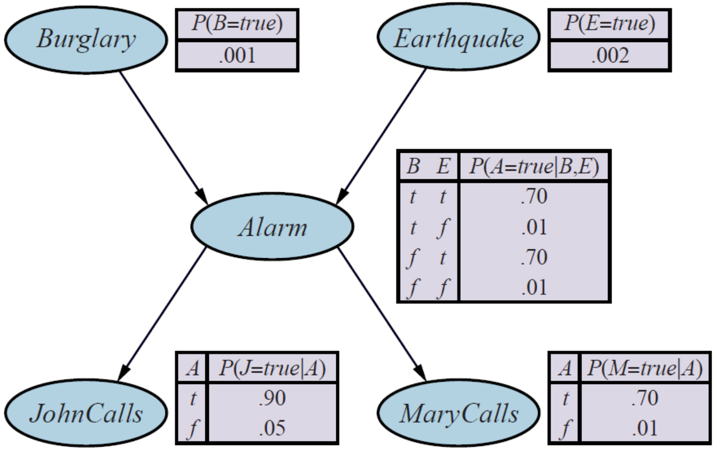
Example: Smoke alarm

- Variables:
  - F: There is fire
  - S: There is smoke
  - A: Alarm sounds





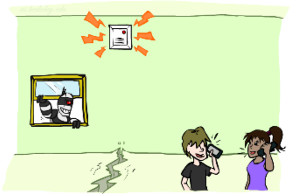
Bayes Net Syntax



- A set of nodes, one per variable X_i
- A directed, acyclic graph
- A conditional distribution for each node given its parent variables in the graph
  - CPT (conditional probability table); each row is a distribution for child given values of its parents

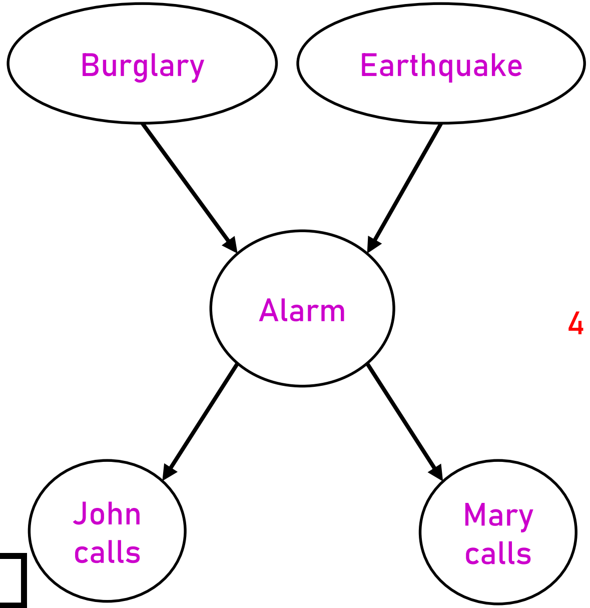
Bayes net = Topology (graph) + Local Conditional Probabilities

Example: Alarm Network



P(B)	
true	false
0.001	0.999

1



1

P(E)	
true	false
0.002	0.998

4

B	E	P(A|B,E)	
		true	false
true	true	0.95	0.05
true	false	0.94	0.06
false	true	0.29	0.71
false	false	0.001	0.999

A	P(J|A)	
	true	false
true	0.9	0.1
false	0.05	0.95

2

2

A	P(M|A)	
	true	false
true	0.7	0.3
false	0.01	0.99

Number of free parameters in each CPT:

1. Parent range sizes $d_1,\ldots,d_k$
2. Child range size d
3. Each row must sum to 1

$$(d-1)\ \Pi_i\ d_i$$

General formula for sparse BNs

- Suppose
  - n variables
  - Maximum range size is d
  - Maximum number of parents is k
- Full joint distribution has size $O(d^n)$
- Bayes net has size $O(n \cdot d^k)$
  - Linear scaling with n as long as causal structure is local

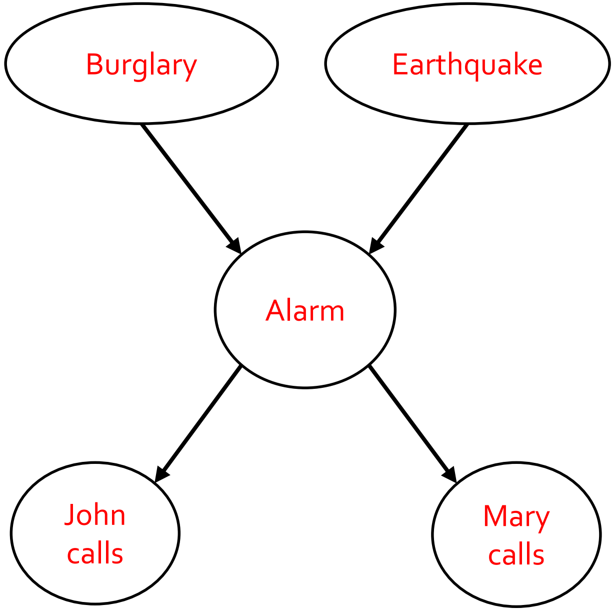
Bayes net global semantics

- Bayes nets encode joint distributions as product of cond distributions on each variable:
  - $P(X_1,..,X_n) \;=\; \prod_i P(X_i \mid Parents(X_i))$

Example

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

P(B)	
true	false
0.001	0.999



P(E)	
true	false
0.002	0.998

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

$=.001x.998x.94x.1x.3=.000028$

B	E	P(A|B,E)	
		true	false
true	true	0.95	0.05
true	false	0.94	0.06
false	true	0.29	0.71
false	false	0.001	0.999

A	P(J|A)	
	true	false
true	0.9	0.1
fals e	0.05	0.95

A	P(M|A)	
	true	false
true	0.7	0.3
false	0.01	0.99

Conditional independence in BNs

- Compare the Bayes net global semantics

$$P(X_1,..,X_n) = \prod_i P(X_i \mid Parents(X_i))$$

with the chain rule identity

$$P(X_1,..,X_n) = \prod_i P(X_i \mid X_1,\ldots,X_{i-1})$$

Conditional independence in BNs

- Let $X_1,\ldots,X_n$ be sorted in topological order according graph, i.e., parents before children, so

$$\text{Parents}(X_i) \subseteq X_1,\ldots,X_{i-1}$$

- So the Bayes net asserts conditional independences

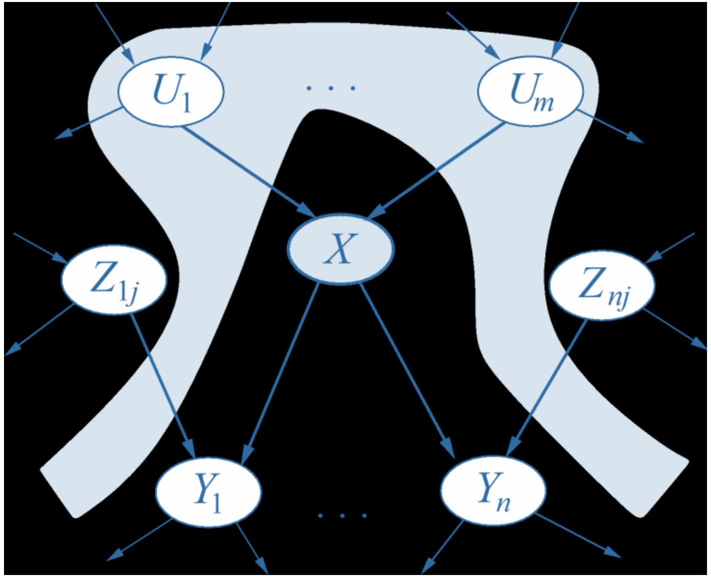
$$P(X_i \mid X_1,\ldots,X_{i-1}) = P(X_i \mid \text{Parents}(X_i))$$

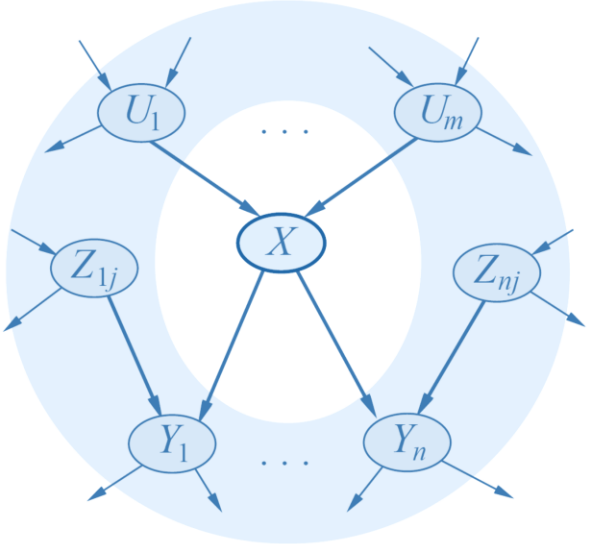
  - To ensure these are valid, choose parents for node X_i that "shield" it from other predecessors

- $P(M \mid J, A, E, B) = P(M \mid A)$

Conditional independence semantics

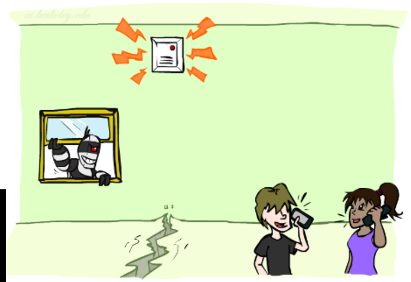
- Every variable is conditionally independent of its non-descendants given its parents
  - Markov blanket: parents, children and children's parents
- Conditional independence semantics <=> global semantics





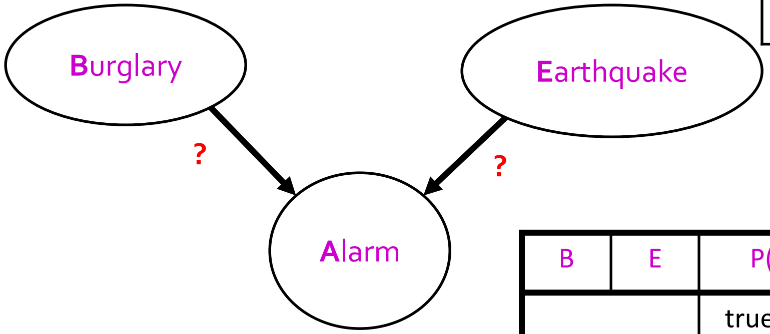
Example: Burglary

- Burglary
- Earthquake
- Alarm



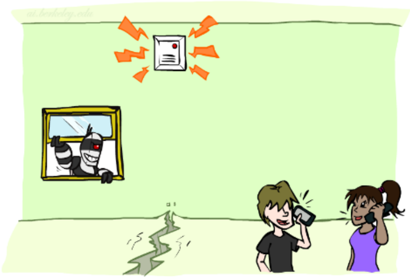
P(B)	
true	false
0.001	0.999

P(E)	
true	false
0.002	0.998



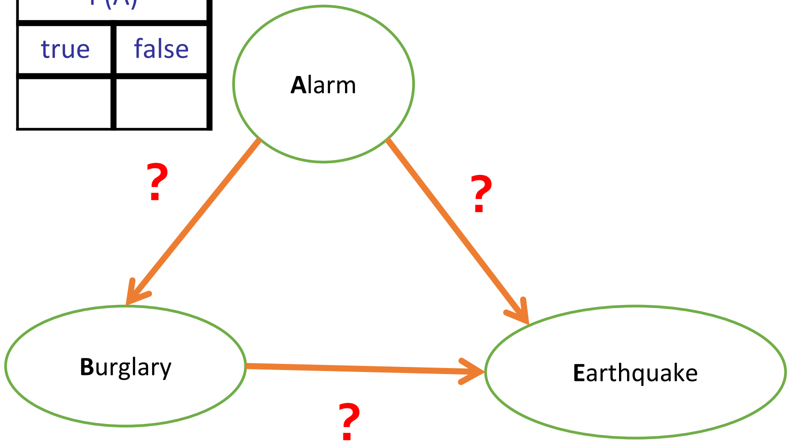
B	E	P(A|B,E)	
		true	false
true	true	0.95	0.05
true	false	0.94	0.06
false	true	0.29	0.71
false	false	0.001	0.999

Example: Burglary



- Alarm
- Burglary
- Earthquake

P(A)	
true	false



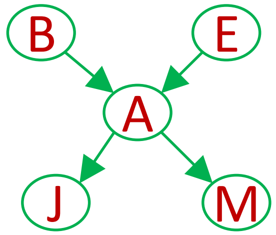
A	P(B|A)	
	true	false
true	?	
false		

A	B	P(E|A,B)	
		true	false
true	true	?	
true	false		
false	true		
false	false		

Inference by Enumeration in Bayes Net

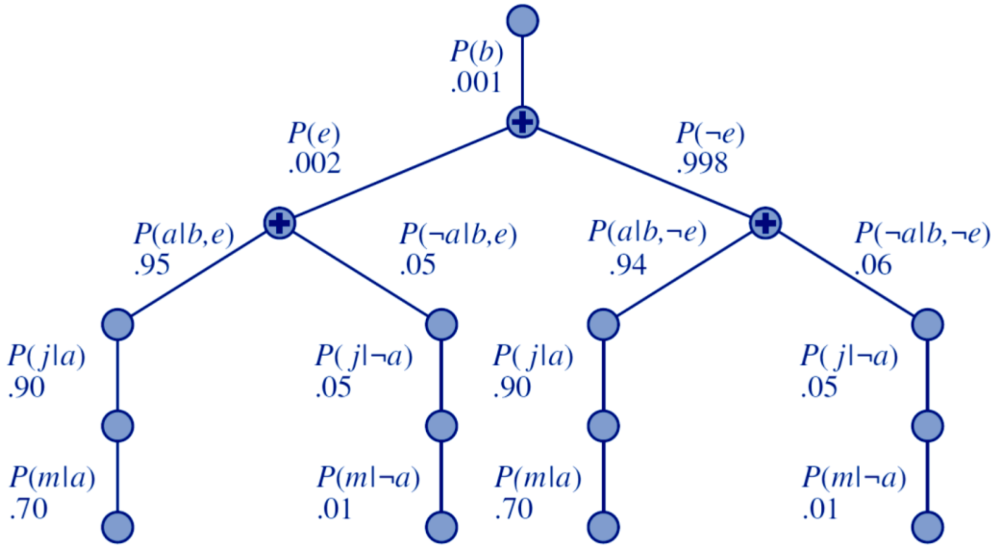
- Reminder of inference by enumeration
  - Any probability of interest can be computed by summing entries from the joint distribution: $P(Q \mid e) = \alpha \sum_h P(Q, h, e)$
  - Entries from the joint distribution can be obtained from a BN by multiplying the corresponding conditional probabilities
- $P(B \mid j, m) = \alpha \sum_e \sum_a P(B, e, a, j, m)$

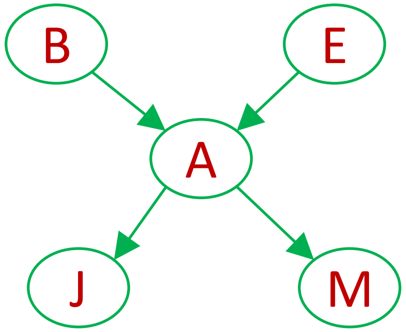
  $= \alpha \sum_e \sum_a P(B)\, P(e)\, P(a \mid B,e)\, P(j \mid a)\, P(m \mid a)$
- So inference in Bayes nets means computing sums of products of numbers: sounds easy!!
- Problem: sums of exponentially many products!



Inference by Enumeration in Bayes Net

- $P(B \mid j, m) = \alpha \sum_e \sum_a P(B, e, a, j, m)$
  $= \alpha \sum_{e,a} P(B)\, P(e)\, P(a \mid B, e)\, P(j \mid a)\, P(m \mid a)$
- $P(b \mid j, m) = \alpha * 0.00059224$, $P(\neg b \mid j, m) = \alpha * 0.0014919$
  - $P(B \mid j, m) = \langle 0.284, 0.716 \rangle$





Can we do better?

$\sum_{e,a} P(B)\ P(e)\ P(a|B,e)\ P(j|a)\ P(m|a)$

$= P(B)P(e)P(a|B,e)P(j|a)P(m|a) + P(B)P(\neg e)P(a|B,\neg e)P(j|a)P(m|a)$

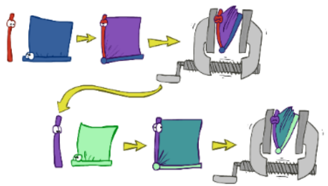
$+ P(B)P(e)P(\neg a|B,e)P(j|\neg a)P(m|\neg a) + P(B)P(\neg e)P(\neg a|B,\neg e)P(j|\neg a)P(m|\neg a)$

Lots of repeated subexpressions!

Can we do better?

- Consider $uwy + uwz + uxy + uxz + vwy + vwz + vxy + vxz$
  - 16 multiplies, 7 adds
  - Lots of repeated subexpressions!
- Rewrite as $(u+v)(w+x)(y+z)$
  - 2 multiplies, 3 adds

Variable elimination: The basic ideas

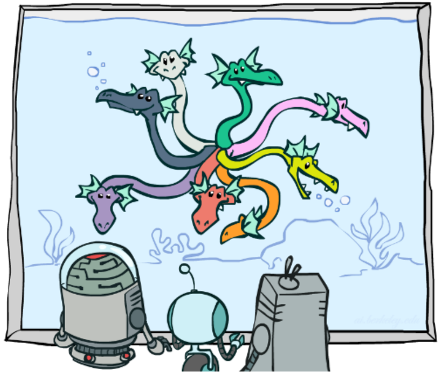


- Eliminate repeated calculations
- Move summations inwards as far as possible

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

- Do the calculation from the inside out
  - i.e., sum over a first, then sum over e
  - Problem: P(a | B, e) isn't a single number, it's a bunch of different numbers depending on the values of B and e
  - Solution: use arrays of numbers (of various dimensions) with appropriate operations on them; these are called factors

Factors



- Joint distribution: P(X,Y)
  - Entries P(x,y) for all x, y
  - |X| x |Y| matrix
  - Sums to 1

P(A,J)

A \ J	true	false
true	0.09	0.01
false	0.045	0.855

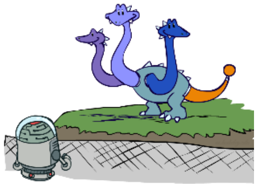
- Projected joint: P(x,Y)
  - A slice of the joint distribution
  - Entries P(x,y) for one x, all y
  - |Y|-element vector
  - Sums to P(x)

$P(a,J) = P_a(J)$

A \ J	true	false
true	0.09	0.01

Number of random variables (Capitals) = table's dimensionality

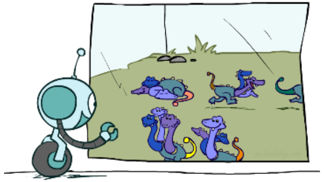
Factors



- Single conditional: $P(Y \mid x)$
  - Entries $P(y \mid x)$ for fixed x, all y
  - Sums to 1

$P(J|a)$

A \ J	true	false
true	0.9	0.1



- Family of conditionals:
- $P(X \mid Y)$
  - Multiple conditionals
  - Entries $P(x \mid y)$ for all x, y
  - Sums to $|Y|$

$P(J|A)$

A \ J	true	false	
true	0.9	0.1	} - $P(J|a)$
false	0.05	0.95	} - $P(J|\neg a)$

Operation 1: Pointwise product

- First basic operation: **pointwise product of factors** (similar to a database join, not matrix multiply!)
  - New factor has union of variables of the two original factors
  - Each entry is the product of the corresponding entries from the original factors

$P(J|A) \times P(A) = P(A,J)$

- Example:

P(A)

true	0.1
false	0.9

x

P(J|A)

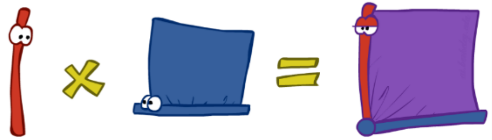
A \ J	true	false
true	0.9	0.1
false	0.05	0.95

=

P(A,J)

A \ J	true	false
true	0.09	0.01
false	0.045	0.855

Operation 1: Pointwise product



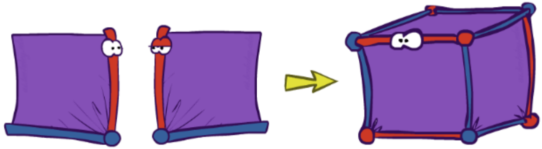
- $f(X_1, \ldots, X_j, Y_1, \ldots, Y_k) * g(Y_1, \ldots, Y_k, Z_1, \ldots, Z_l) = h(X_1, \ldots, X_j, Y_1, \ldots, Y_k, Z_1, \ldots, Z_l)$
  - If all variables are binary, the pointwise product g has 2^{j+k+l} entries

X	Y	$\mathbf{f}(X,Y)$	Y	Z	$\mathbf{g}(Y,Z)$	X	Y	Z	$\mathbf{h}(X,Y,Z)$
t	t	.3	t	t	.2	t	t	t	$.3 \times .2 = .06$
t	f	.7	t	f	.8	t	t	f	$.3 \times .8 = .24$
f	t	.9	f	t	.6	t	f	t	$.7 \times .6 = .42$
f	f	.1	f	f	.4	t	f	f	$.7 \times .4 = .28$
						f	t	t	$.9 \times .2 = .18$
						f	t	f	$.9 \times .8 = .72$
						f	f	t	$.1 \times .6 = .06$
						f	f	f	$.1 \times .4 = .04$

Figure 13.12 Illustrating pointwise multiplication: $\mathbf{f}(X,Y) \times \mathbf{g}(Y,Z) = \mathbf{h}(X,Y,Z)$.

Example: Making larger factors

- Example: P(A,J) x P(A,M) = P(A,J,M)



P(A,J)

A \ J	true	false
true	0.09	0.01
false	0.045	0.855

x

P(A,M)

A \ M	true	false
true	0.07	0.03
false	0.009	0.891

=

P(A,J,M)

A=false

J \ M	true	false
true		.8
false		

A=true

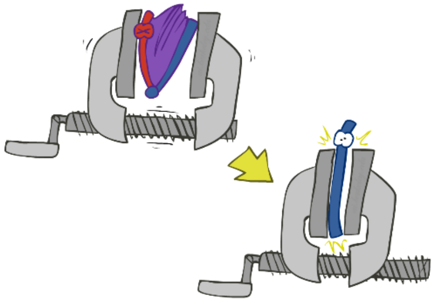
J \ M	true	false
true		
false		.0003

Making larger factors

- Example: P(U,V) x P(V,W) x P(W,X) = P(U,V,W,X)
- Sizes: [10,10] x [10,10] x [10,10] = [10,10,10,10]
  - i.e., 300 numbers blows up to 10,000 numbers!
  - Factor blowup can make VE very expensive

Operation 2: Summing out a variable

- Second basic operation: summing out (or eliminating) a variable from a factor
  - Shrinks a factor to a smaller one
- Example: $\sum_j P(A,J) = P(A,j) + P(A,\neg j) = P(A)$



P(A,J)

A \ J	true	false
true	0.09	0.01
false	0.045	0.855

Sum out J $\rightarrow$

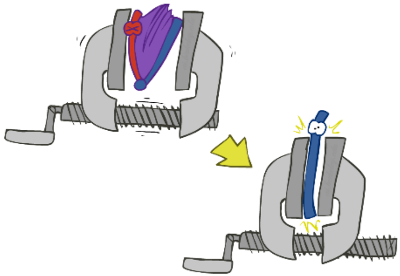
P(A)

true	0.1
false	0.9

Summing out from a product of factors

- Project the factors each way first, then sum the products

Example:

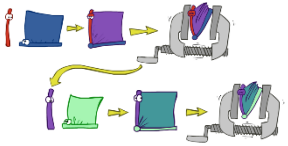


$\sum_a$ P(a|B,e) x P(j|a) x P(m|a)
= P(a|B,e) x P(j|a) x P(m|a)
+ P(¬a|B,e) x P(j|¬a) x P(m|¬a)

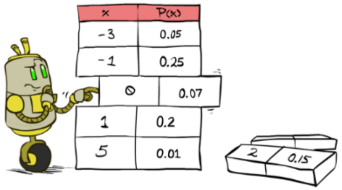
$$\mathbf{h}_2(Y,Z) = \sum_x \mathbf{h}(X,Y,Z) = \mathbf{h}(x,Y,Z) + \mathbf{h}(\neg x,Y,Z)$$

$$= \begin{pmatrix} .06 & .24 \\ .42 & .28 \end{pmatrix} + \begin{pmatrix} .18 & .72 \\ .06 & .04 \end{pmatrix} = \begin{pmatrix} .24 & .96 \\ .48 & .32 \end{pmatrix}$$

Variable Elimination



- Query: $P(Q \mid E_1=e_1, \ldots, E_k=e_k)$
- Start with initial factors:
  - Local CPTs (but instantiated by evidence)



- While there are still hidden variables (not Q or evidence):
  - Pick a hidden variable H_j
  - Eliminate (sum out) H_j from the product of all factors mentioning H_j

- Join all remaining factors and normalize

Example

Query P(B | j,m)

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

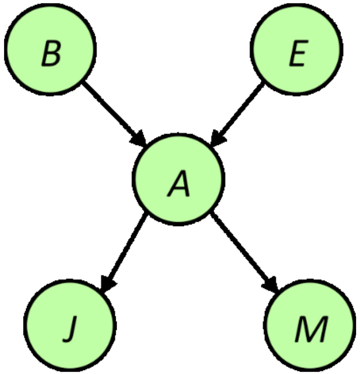
P(B)	P(E)	P(A|B,E)	P(j|A)	P(m|A)

Choose A

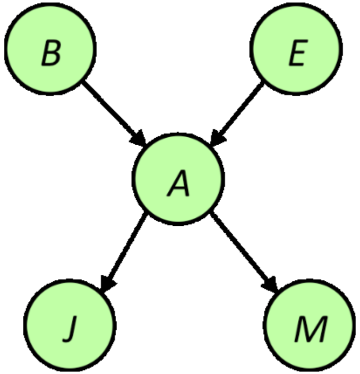
P(A|B,E)
P(j|A)
P(m|A)

$\xrightarrow{\blacksquare}$ $\xrightarrow{\sum}$ P(j,m|B,E)

P(B)	P(E)	P(j,m|B,E)



Example



Query P(B | j,m)

P(B)	P(E)	P(j,m|B,E)

Choose E

P(E)
P(j,m|B,E) $\xrightarrow{\blacksquare}$ $\xrightarrow{\sum}$ P(j,m|B)

P(B)	P(j,m|B)

Finish with B

P(B)
P(j,m|B) $\xrightarrow{\blacksquare}$ P(j,m,B) $\xrightarrow{\text{Normalize}}$ P(B | j,m)

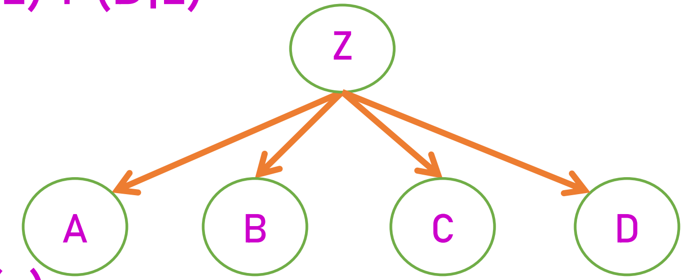
Order matters

- Order the terms Z, A, B C, D

  $P(D) = \alpha \sum_{z,a,b,c} P(z)\, P(a|z)\, P(b|z)\, P(c|z)\, P(D|z)$

  $\quad = \alpha \sum_{z} P(z) \sum_{a} P(a|z) \sum_{b} P(b|z) \sum_{c} P(c|z)\, P(D|z)$

  - Largest factor has 2 variables (D,Z)



- Order the terms A, B C, D, Z

  $P(D) = \alpha \sum_{a,b,c,z} P(a|z)\, P(b|z)\, P(c|z)\, P(D|z)\, P(z)$

  $\quad = \alpha \sum_{a} \sum_{b} \sum_{c} \sum_{z} P(a|z)\, P(b|z)\, P(c|z)\, P(D|z)\, P(z)$

  - Largest factor has 4 variables (A,B,C,D)
  - In general, with n leaves, factor of size 2^n
- Finding optimal ordering is intractable!

Order matters

- Exercise: $P(J \mid b) = ?$

Order matters

- Exercise: $P(J \mid b) = \alpha\, P(b) \sum_e P(e) \sum_a P(a \mid b, e)\, P(J \mid a) \sum_m P(m|a)$

- Every variable that is not an ancestor of query or evidence does not matter
  - M in this example

VE: Computational and Space Complexity

- The computational and space complexity of variable elimination is determined by the largest factor (and it's space that kills you)

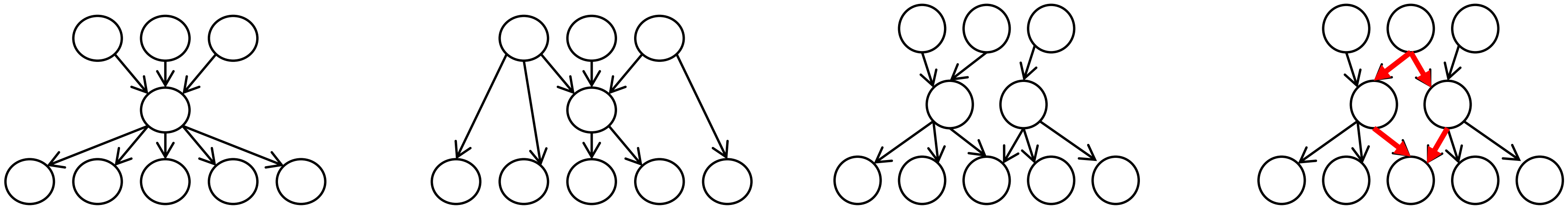
- The elimination ordering can greatly affect the size of the largest factor.
  - E.g., previous slide's example 2^n vs. 2

- Does there always exist an ordering that only results in small factors?
  - No!

Polytrees

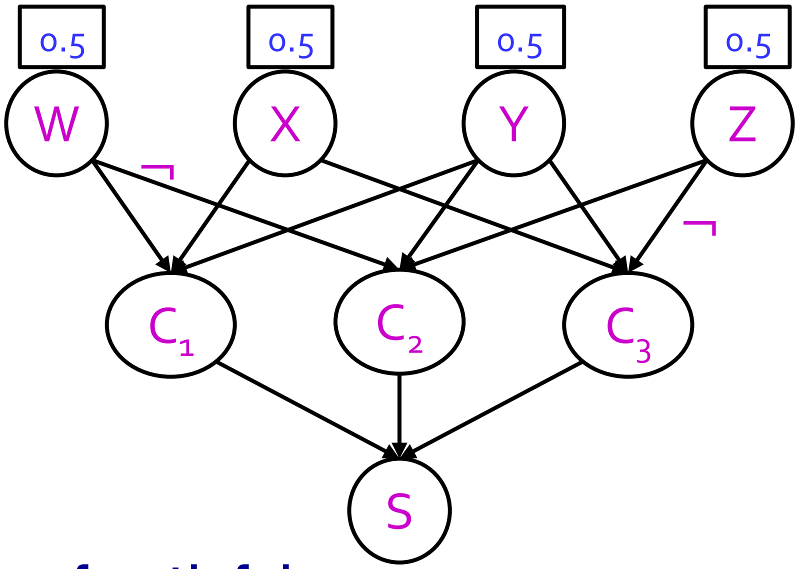
- A polytree is a directed graph with no undirected cycles

- For polytrees the complexity of variable elimination is linear in the network size (number of CPT entries) if you eliminate from the leave towards the roots



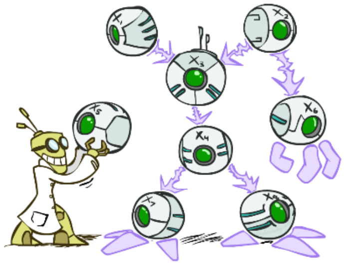
Worst Case Complexity – Reduction from SAT

- Variables: W, X, Y, Z
- CNF clauses:
  1. $C_1 = W \vee X \vee Y$
  2. $C_2 = Y \vee Z \vee \neg W$
  3. $C_3 = X \vee Y \vee \neg Z$
- Sentence $S = C_1 \wedge C_2 \wedge C_3$
- $P(S) > 0$ iff S is satisfiable
  - => NP-hard
- $P(S) = K \times 0.5^n$ where K is the number of satisfying assignments for clauses
  - => #P-hard



Summary

- Independence and conditional independence are important forms of probabilistic knowledge
- Bayes net encode joint distributions efficiently by taking advantage of conditional independence
  - Global joint probability = product of local conditionals
- Exact inference = sums of products of conditional probabilities from the network



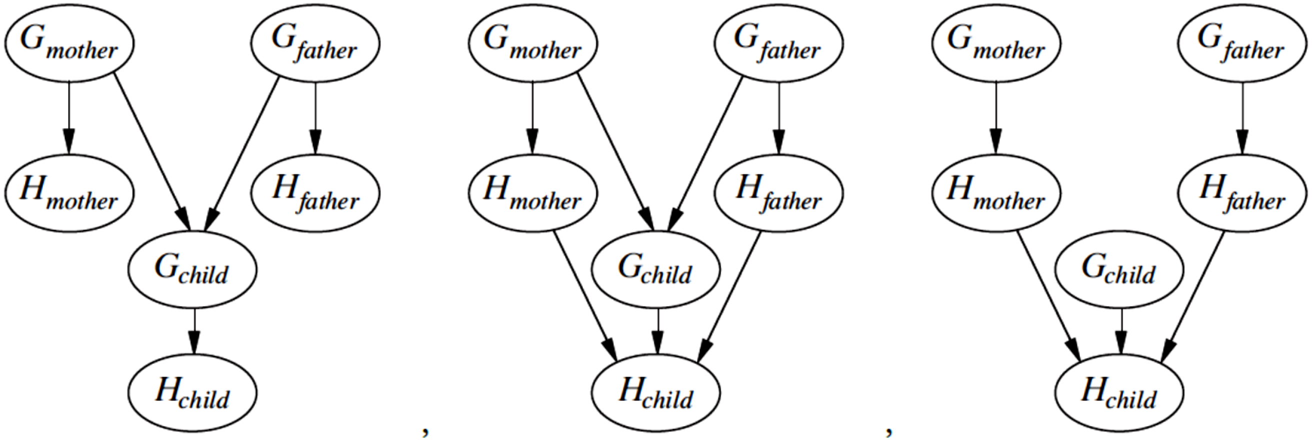
13.1 We have a bag of three biased coins a, b, and c with probabilities of coming up heads of 20%, 60%, and 80%, respectively. One coin is drawn randomly from the bag (with equal likelihood of drawing each of the three coins), and then the coin is flipped three times to generate the outcomes X_1, X_2, and X_3.

a. Draw the Bayesian network corresponding to this setup and define the necessary CPTs.

b. Calculate which coin was most likely to have been drawn from the bag if the observed flips come out heads twice and tails once.

13.5 Let H_x be a random variable denoting the handedness of an individual x, with possible values l or r. A common hypothesis is that left- or right-handedness is inherited by a simple mechanism; that is, perhaps there is a gene G_x, also with values l or r, and perhaps actual handedness turns out mostly the same (with some probability s) as the gene an individual possesses. Furthermore, perhaps the gene itself is equally likely to be inherited from either of an individual's parents, with a small nonzero probability m of a random mutation flipping the handedness.

a. Which of the three networks in Figure 2 claim that
$\mathbf{P}(G_{father}, G_{mother}, G_{child}) = \mathbf{P}(G_{father})\mathbf{P}(G_{mother})\mathbf{P}(G_{child})$?

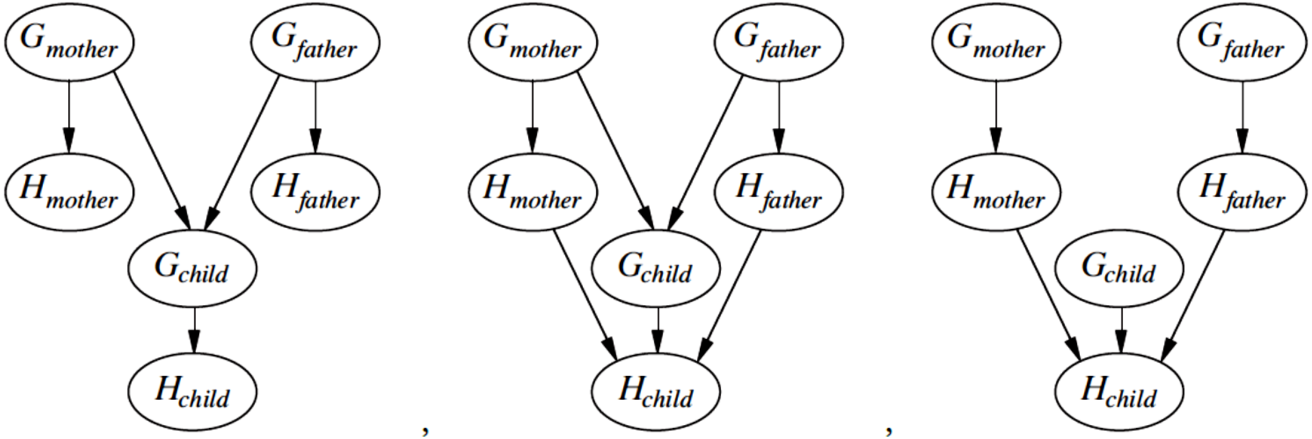


13.5 Let H_x be a random variable denoting the handedness of an individual x, with possible values l or r. A common hypothesis is that left- or right-handedness is inherited by a simple mechanism; that is, perhaps there is a gene G_x, also with values l or r, and perhaps actual handedness turns out mostly the same (with some probability s) as the gene an individual possesses. Furthermore, perhaps the gene itself is equally likely to be inherited from either of an individual's parents, with a small nonzero probability m of a random mutation flipping the handedness.

b. Which of the three networks make independence claims that are consistent with the hypothesis about the inheritance of handedness?

c. Which of the three networks is the best description of the hypothesis?

d. Write down the CPT for the G_{child} node in network (a), in terms of s and m.



Next time

- Markov models
- Hidden Markov models