Probabilistic Reasoning

(Reasoning under uncertainty with Bayesian Nets)

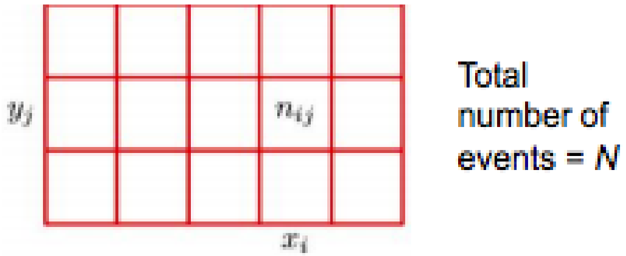
CS417

R. Hedjam

Recall

Joint Probability

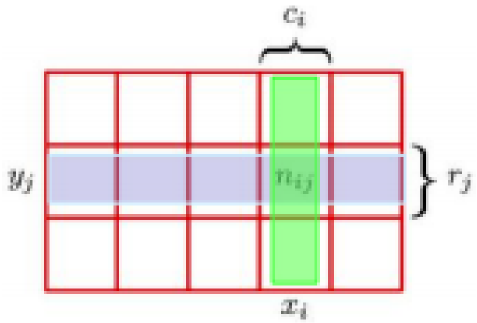
$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$



Marginal Probability

$$P(X = x_i) = \sum_j P(x_i, y_j) = \boxed{\frac{c_i}{N}}$$

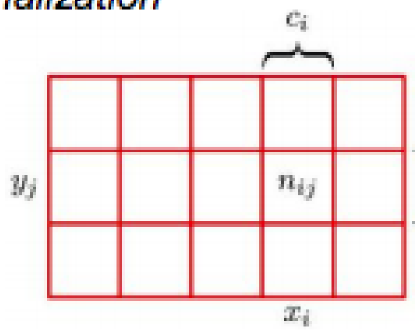
$$P(Y = y_j) = \sum_i P(x_i, y_i) = \boxed{\frac{r_j}{N}}$$



Summing out a variable is called *marginalization*

Conditional Probability

$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}$$



From: bishop

Recall: Bayes' Rule

$$P(x,y) = P(x \mid y)P(y) = P(y \mid x)P(x)$$

i.e.

$$\boxed{P(x \mid y) = \frac{P(y \mid x)\, P(x)}{P(y)}}$$

Normalization in Bayes' Rule

$$P(x \mid y) = \frac{P(y \mid x)\, P(x)}{P(y)} = \alpha\, P(y \mid x) P(x)$$

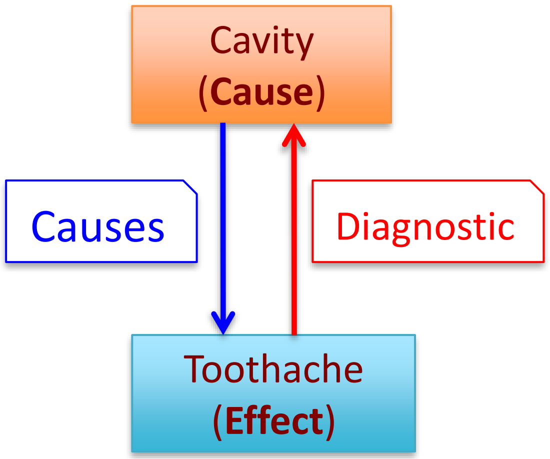
$$\alpha = \frac{1}{P(y)} = \frac{1}{\sum_x P(y,x)} = \frac{1}{\sum_x P(y \mid x) P(x)}$$

α is called the normalization constant
(can be calculated by summing over numerator values)

Why Bayes rule is useful?

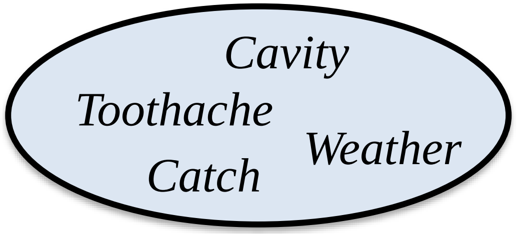
- Allows diagnostic reasoning from causal information

$$P(Cause|Effect) = \frac{P(Effect|Cause)P(Cause)}{P(Effect)}$$



Independence

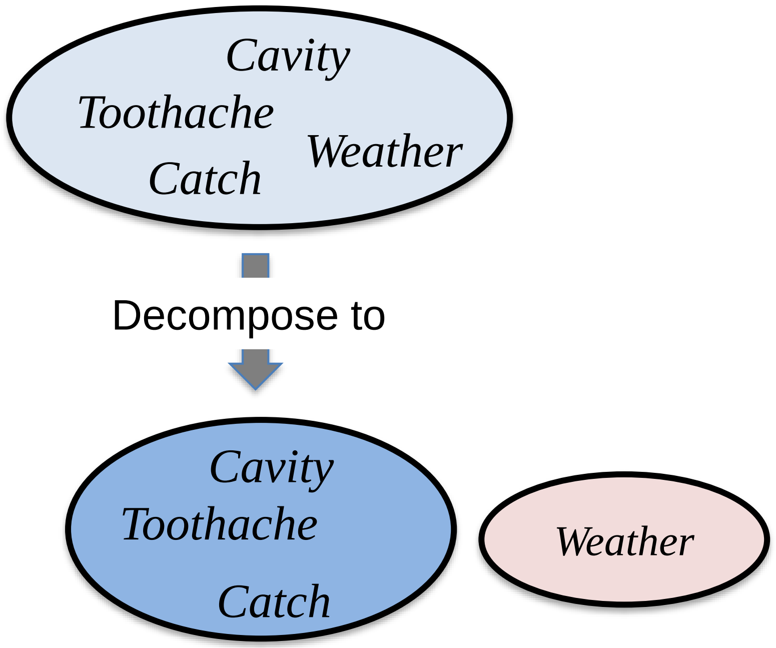
- **Given four (4) r.v.:**
  - **Toothache** = {**T**rue, **F**alse}
  - **Catch** = {**T**rue, **F**alse}
  - **Cavity** = {**T**rue, **F**alse}
  - **Weather** = {Sunny, Rainy, Cloudy, Snow}
- There are 2*2*2*4 = 32 possible values.
  - ➢ P(Toothache, Catch, Cavity, Weather)



Independence

- **Given four (4) r.v.:**
  - **Toothache** = {**T**rue, **F**alse}
  - **Catch** = {**T**rue, **F**alse}
  - **Cavity** = {**T**rue, **F**alse}
  - **Weather** = {Sunny, Rainy, Cloudy, Snow}
- There are 2*2*2*4 = 32 possible values.
  - P(Toothache, Catch, Cavity, Weather)
- Weather is ***independent*** from other r.v.
- There 2*2*2 + 4 possible values.
  - P(Toothache, Catch, Cavity, Weather) =

    P(Toothache, Catch, Cavity)P(Weather)



- *32 possible values reduced to 12.*
- *For n independent biased coins, $2^n \rightarrow n$*

Conditional Independence

Given:

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

Joint probability distribution:

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

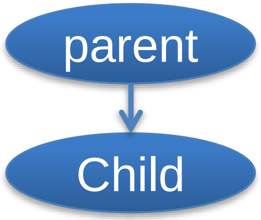
$= \mathbf{P}(Catch \mid Toothache, Cavity)\mathbf{P}(Toothache, Cavity)$

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

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

2 + 2 + 1

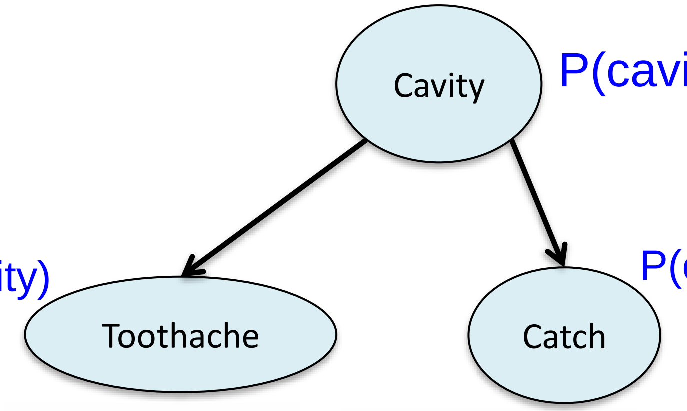
= 5 independent numbers



P(child | parent)

P(cavity=False)=1-
P(Cavity=True)

Cavity
True = 0.2



P(cavity)

P(Toothache | cavity)

P(catch | cavity)

CPT

Cavity	Toothache
true	0.8
false	0.1

Cavity	Catch
true	0.3
false	0.1

Catch=F
0.7
0.9

P(catch=T|Cavity)

Power of Conditional Independence

- Often, **CI** can reduce the storage complexity of the joint distribution from exponential to linear!!
- **CI** is the most basic and robust form of knowledge in uncertain environments.

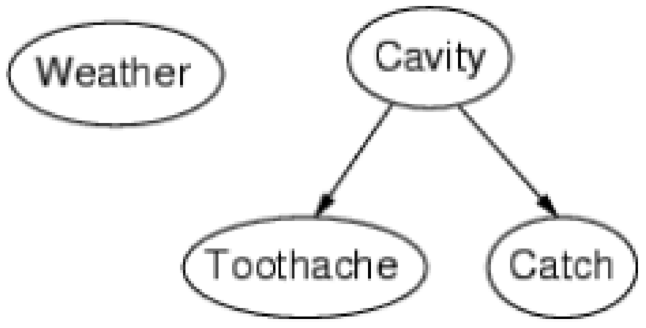
Bayesian Nets ?

- Simple, graphical notation for conditional independence assertions
  - Allows compact specification on full joint distribution

Bayesian Nets ?

Example: Back at the Dentist's

- Topology of network encodes conditional independence assertions:



- *Weather* is independent of the other variables
- *Toothache* and *Catch* are conditionally independent of each other given *Cavity*

Node: variables (with domains)
- Can be assigned (observed) or
- Unassigned (unobserved)

Arcs: interactions
- indicate "direct influence" between variables
- Formally: encode conditional independence

Naïve Bayesian Model

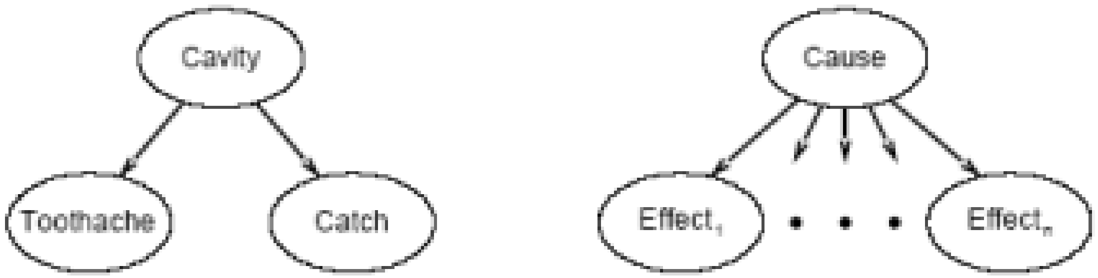
A special case of BN

Conditional Independence and the "Naïve Bayes Model"

$$
\begin{aligned}
&\mathbf{P}(Cavity|toothache \wedge catch) \\
&= \alpha\,\mathbf{P}(toothache \wedge catch|Cavity)\mathbf{P}(Cavity) \\
&= \alpha\,\mathbf{P}(toothache|Cavity)\mathbf{P}(catch|Cavity)\mathbf{P}(Cavity)
\end{aligned}
$$

This is an example of a *naive Bayes* model:

$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause)\Pi_i \mathbf{P}(Effect_i|Cause)$$



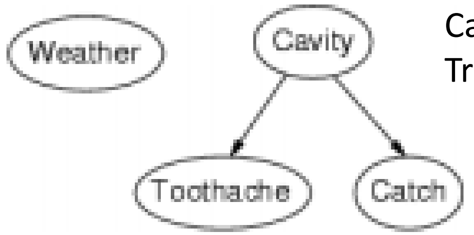
Total number of parameters is *linear* in n

P(Cavity, Tooth, Catch) = P(cavity)P(Tooth|cavity)P(catch|cavity)

Bayesian networks

- Syntax:
  - set of nodes, one per random variable
  - directed, acyclic graph (link ≈ "directly influences")
  - conditional distribution for each node given its parents:
    $P\ (X_i \mid Parents\ (X_i))$

- For discrete variables, conditional distribution = conditional probability table (CPT) = distribution over X_i for each combination of parent values



Cavity
True = 0.2

Cavity	Toothache
true	0.8
false	0.1

Cavity	Catch
true	0.3
false	0.1

CPT

A Bayes net = Topology (graph) + Local Conditional Probability

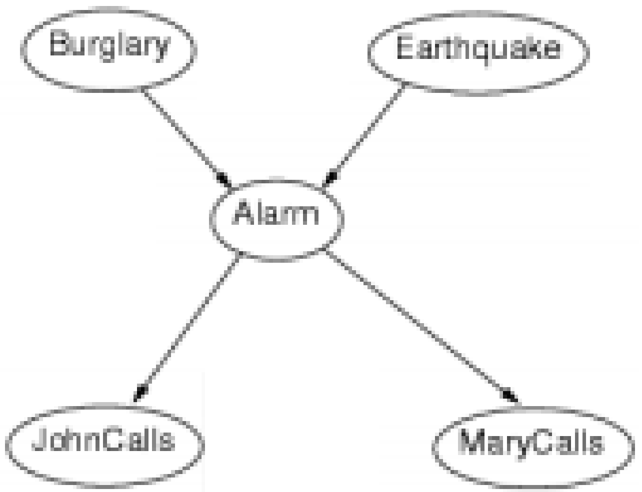
Example 2: Burglar and Earthquakes

- You have a new alarm at home that:
  - *Rings when there is a* ***burglary***;
  - *Sometimes rings when there is an* ***earthquake***.
- You have two neighbors who call you at the office if they hear the alarm.
  - ***John calls*** *all the time when he hears the alarm, but sometimes he confuses the phone with the alarm*.
  - ***Mary*** *likes to listen to loud music and sometimes she does not hear the alarm*.
- Knowing who called, what is the probability that there is a burglary?

Example 2: Burglars and Earthquakes

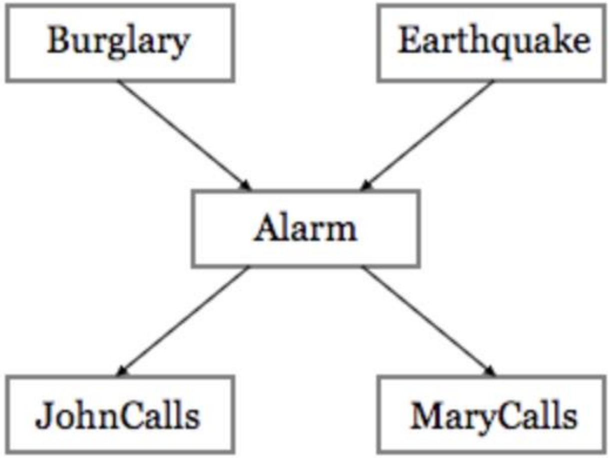
- Variables: *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, *MaryCalls*

- 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

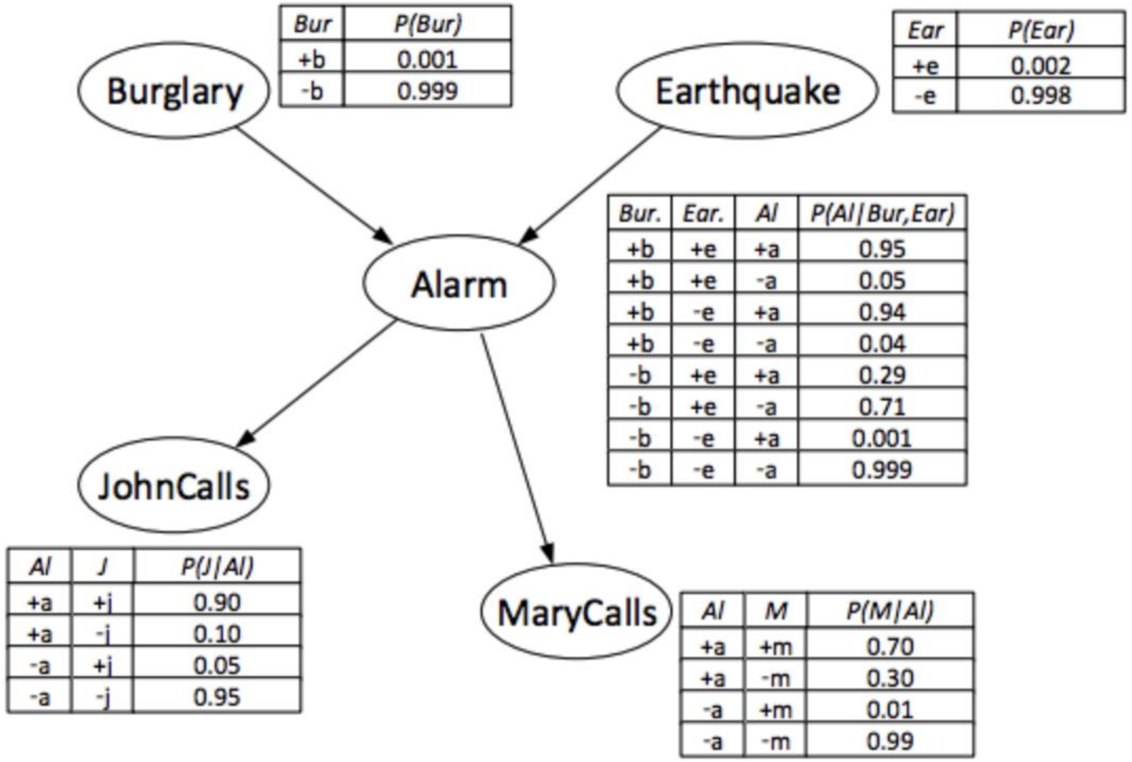


Example 2: Burglary and Earthquakes

- The topology of the network reflects the set of conditional independence relations:
  - Burglary and Earthquake affect directly the probability of triggering an alarm
  - The fact that John or Mary calls depends only on the alarm. John and Mary do not directly perceive burglary or minor earthquakes



Example 2: Burglar and earthquakes



Bur	P(Bur)
+b	0.001
-b	0.999

Ear	P(Ear)
+e	0.002
-e	0.998

Bur.	Ear.	Al	P(Al|Bur,Ear)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.04
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

Al	J	P(J|Al)
+a	+j	0.90
+a	-j	0.10
-a	+j	0.05
-a	-j	0.95

Al	M	P(M|Al)
+a	+m	0.70
+a	-m	0.30
-a	+m	0.01
-a	-m	0.99

$$P(+b, -e, +a, -j, +m) =$$
$$P(+b)P(-e)P(+a|+b, -e)P(-j|+a)P(+m|+a) =$$
$$0.001 \times 0.998 \times 0.94 \times 0.1 \times 0.7$$

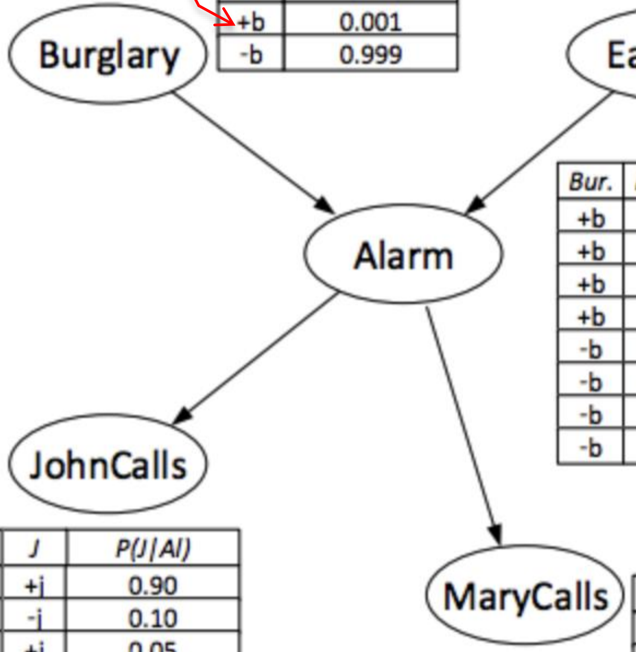
P(+e|+m) = a P(+e, +m) = a [P(+e,+m, +b, +j, +a) + P(+e,+m, -b, +j, +a) + P(+e, +m, -b, -j, +a)…]

Example 2: Alarm Network

P(-e,+b,+a,-j,+m) = P(-e) P(+b)P(+a|+b,-e) P(-j|+a)P(+m|+a)

R.V

Its instance



Bur	P(Bur)
+b	0.001
-b	0.999

Ear	P(Ear)
+e	0.002
-e	0.998

P(-e) = 1-P(+e)

Bur.	Ear.	Al	P(Al|Bur,Ear)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.04
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

P(A ,B,E,J,M) = P(B)P(E)P(A|B,E)P(M|A)P(J|A)
P(+b)P(-e)P(+a|+b,-e)P(+m|+a)P(-j|+a)

Al	J	P(J|Al)
+a	+j	0.90
+a	-j	0.10
-a	+j	0.05
-a	-j	0.95

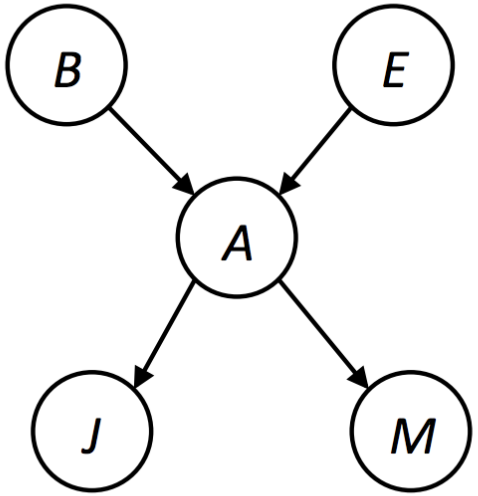
Al	M	P(M|Al)
+a	+m	0.70
+a	-m	0.30
-a	+m	0.01
-a	-m	0.99

+b: there is burglary
-e: there is no earthquake
+a: the alarm goes off
-j: John did not call
+m: Marry calls

$$P(+b,-e,+a,-j,+m) =$$
$$P(+b)P(-e)P(+a|+b,-e)P(-j|+a)P(+m|+a) =$$
$$0.001 \times 0.998 \times 0.94 \times 0.1 \times 0.7$$

Example 2: Burglar and Earthquakes

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998

A	J	P(J|A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

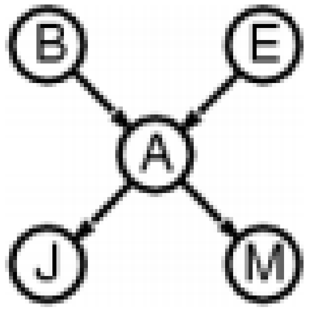
B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

$$P(+b, -e, +a, -j, +m) =$$
$$P(+b)P(-e)P(+a| + b, -e)P(-j| + a)P(+m| + a) =$$
$$0.001 \times 0.998 \times 0.94 \times 0.1 \times 0.7$$

Bayesian Network Semantics

- Full joint distribution is defined as product of local conditional distributions:

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



- e.g., Joint probability of all variables being true = ?

$$P(j \wedge m \wedge a \wedge b \wedge e)$$
$$= P\,(j \mid a)\; P\,(m \mid a)\; P\,(a \mid b, e)\; P\,(b)\; P\,(e)$$

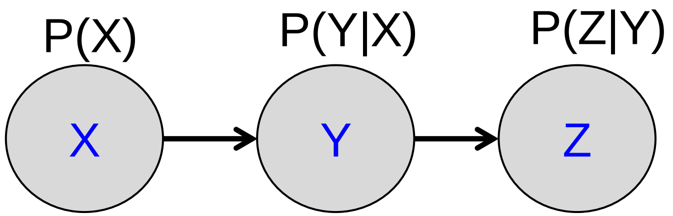
- Similarly, $P(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

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

Causal Chains

- Assume 3 r.v.

X: Low pressure
Y: Rain
Z: Traffic



$$P(x,y,z) = P(x)P(y|x)P(z|y)$$

Are X and Z independent?

Example:

Low pressure causes rain causes traffic, high pressure causes no rain causes no traffic.

$$\boxed{\begin{aligned} P(z|x,y) = \frac{P(x,y,z)}{P(x,y)} &= \frac{\cancel{P(x)}\cancel{P(y|x)}P(z|y)}{\cancel{P(x)}\cancel{P(y|x)}} \\ &= P(z|y) \end{aligned}}$$

So, X and Z are independent
Evidence along the chain, *<u>blocks</u>* the influence.

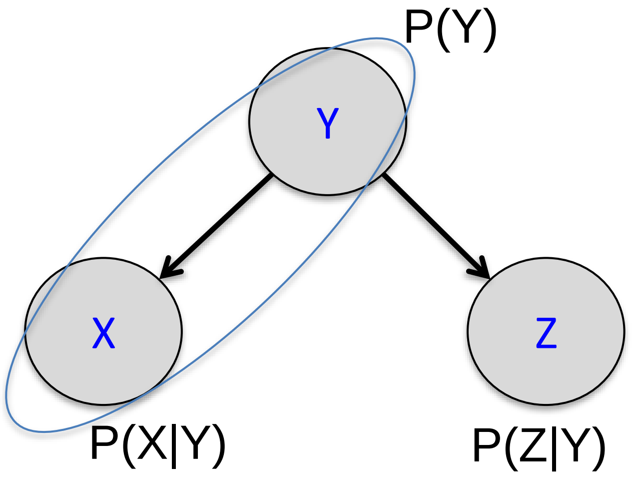
Comment Parent

Two effects of the same parent

- X: News group busy
- Y: Project due
- Z: Lab full

Are X and Z independent?

Are X and Z independent given Y?

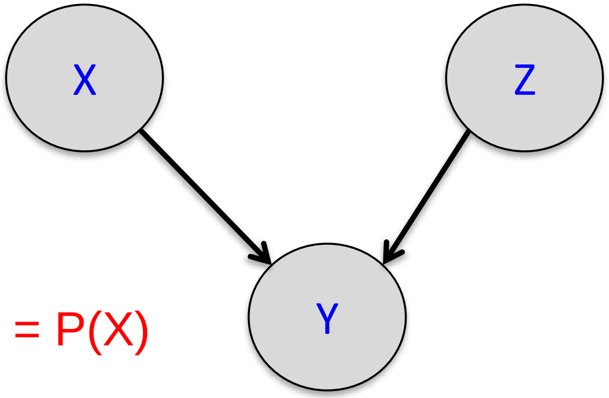


$$P(z|x,y) = \frac{P(x,y,z)}{P(x,y)} = \frac{\cancel{P(x)}\cancel{P(y|x)}P(z|y)}{\cancel{P(x)}\cancel{P(y|x)}}$$
$$= P(z|y)$$

So, X and Z are independent

Observing the cause, *<u>blocks</u>* influence between effects.

Common Effect



Two causes of one effect

- X: Raining
- Y: Ballgame
- Z: Traffic
- Are X and Z independent?
  - Yes, the ballgame and the rain cause traffic, but they are not correlated.
  - Still need to prove they must be.

$P(X/Z)=P(X,Z)/P(Z) = P(X)P(Z)/P(Z) = P(X)$

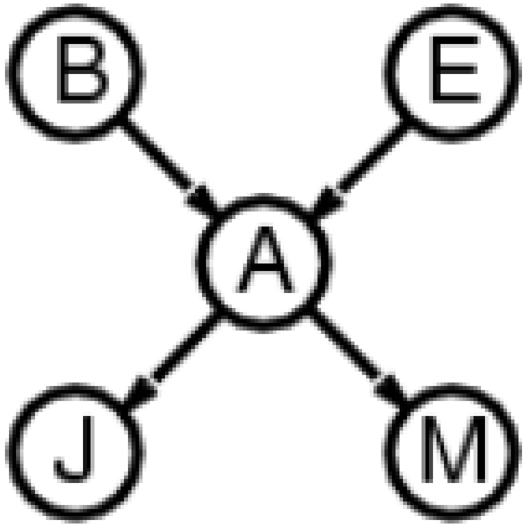
Probabilistic Inference in BNs

- The graphical independence representation yields efficient inference schemes
- We generally want to compute
  - *P(X|E)* where *E* is evidence from sensory measurements etc. (known values for variables)
  - Sometimes, may want to compute just *P(X)*
- One simple inference algorithm:
  - *variable elimination (VE)*

What is the probability of burglary given that John and Mary called?

Compute P(B=true | J=true, M=true)

Inference by enumeration



$$P(b|j,m) = \alpha\, P(b,j,m) = \alpha\, \Sigma_{e,a}\, P(b,j,m,e,a)$$

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

Inference by Enumeration

- General case:
  - Evidence variables: $E_1 \ldots E_k = e_1 \ldots e_k$
  - Query* variable: Q
  - Hidden variables: $H_1 \ldots H_r$

$X_1, X_2, \ldots X_n$

All variables

- We want: $P(Q|e_1 \ldots e_k)$
- First, select the entries consistent with the evidence
- Second, sum out H to get joint of Query and evidence:

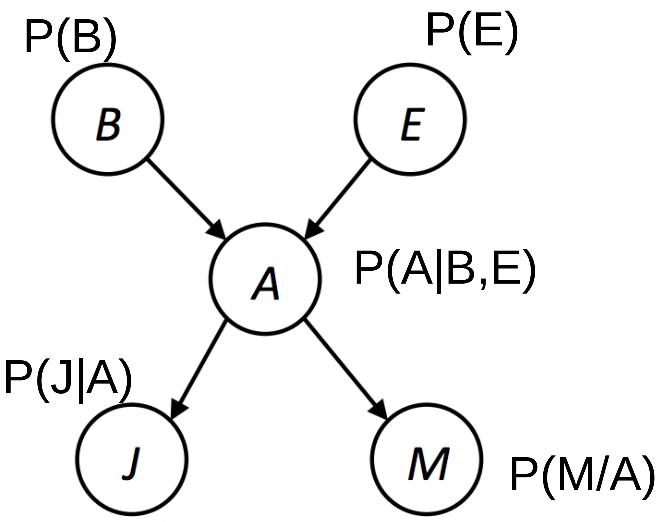
$$P(Q, e_1 \ldots e_k) = \sum_{h_1 \ldots h_r} \underbrace{P(Q, h_1 \ldots h_r, e_1 \ldots e_k)}_{X_1, X_2, \ldots X_n}$$

- Finally, normalize the remaining entries to conditionalize
- Obvious problems:
  - Worst-case time complexity $O(d^n)$
  - Space complexity $O(d^n)$ to store the joint distribution

From: H. Hajishirzi

Inference by Enumeration

- Given unlimited time, inference in BNs is easy
- Reminder of inference by enumeration by example:



$$P(B \mid +j, +m) = \alpha \; P(B, +j, +m)$$

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

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

$$= \alpha \; P(B)P(+e)P(+a|B,+e)\underline{P(+j|+a)P(+m|+a)}$$

$$+ P(B)P(+e)P(-a|B,+e)\underline{P(+j|-a)P(+m|-a)}$$

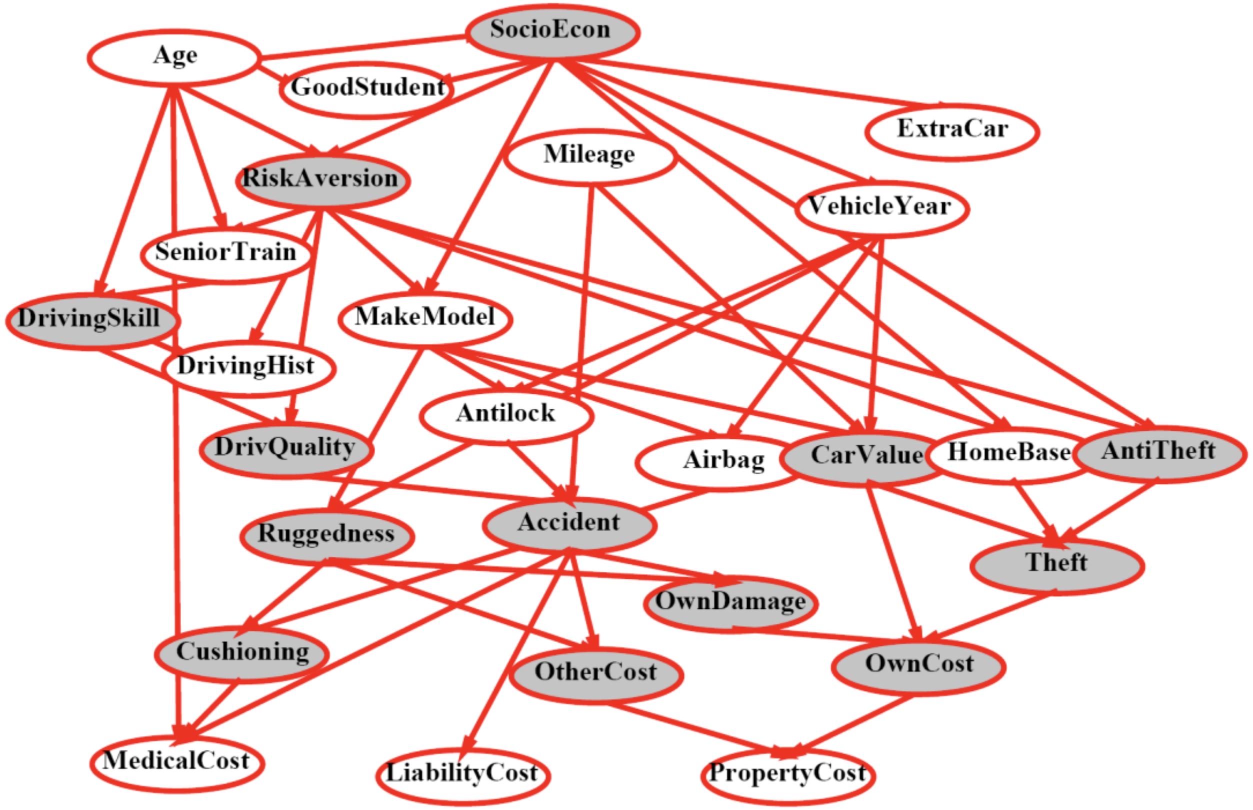
$$+ P(B)P(-e)P(+a|B,-e)\underline{P(+j|+a)P(+m|+a)}$$

$$+ P(B)P(-e)P(-a|B,-e)\underline{P(+j|-a)P(+m|-a)}$$

Repetition

From: H. Hajishirzi

Inference by Enumeration



P(Antilock | given some observation) = ?????

From: H. Hajishirzi

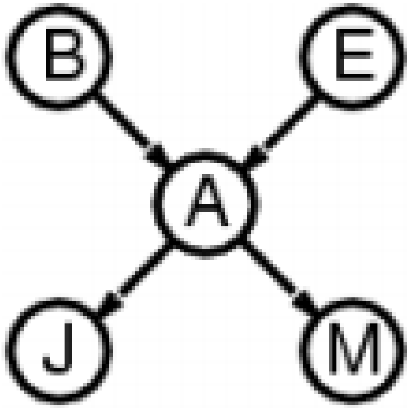
Variable Elimination (VE) Algorithm

- Eliminate variables one-by-one until there is a factor with only the query variables:
    1. *join* all factors containing that variable, multiplying probabilities
    2. *sum out* the influence of the variable

$$\underset{\downarrow}{P(b|j,m)} = \alpha\, \overbrace{P(b)\, \overbrace{\Sigma_e P(e)\, \overbrace{\Sigma_a P(a|b,e)P(j|a)P(m|a)}^{\text{Eliminate a}}}^{\text{Eliminate e}}}^{\text{Remaining factor is a function of b, j, m}}$$

Function of b,j,m

Example of VE: P(J)



$P(J)$

$= \Sigma_{M,A,B,E}\, P(J,M,A,B,E)$

$= \Sigma_{M,A,B,E}\, P(J|A)P(M|A)\, P(A|B,E)\, P(B)\, P(E)$

$= \Sigma_A P(J|A)\, \Sigma_M P(M|A)\, \Sigma_B P(B)\, \Sigma_E P(A|B,E)P(E)$

$= \Sigma_A P(J|A)\, \Sigma_M P(M|A)\, \Sigma_B P(B)\, f1(A,B)$

$= \Sigma_A P(J|A)\, \Sigma_M P(M|A)\, f2(A)$

$= \Sigma_A P(J|A)\, f3(A)$

$= f4(J)$