Bayesian Belief Networks

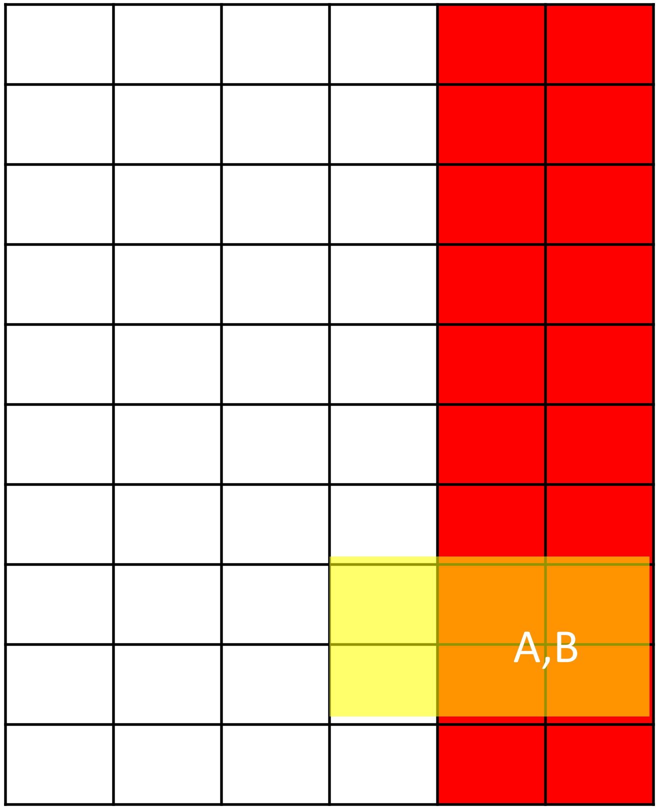
Lecture 02.03

Joint probability: A ∩ B

What is the general probability of both A and B being True

- General (global) probability of **A and B**:

$P(A,B) = P(B|A)*P(A)=4/20*20/60 = 4/60$



Joint probability: A ∩ B ∩ C

- General (global) probability of

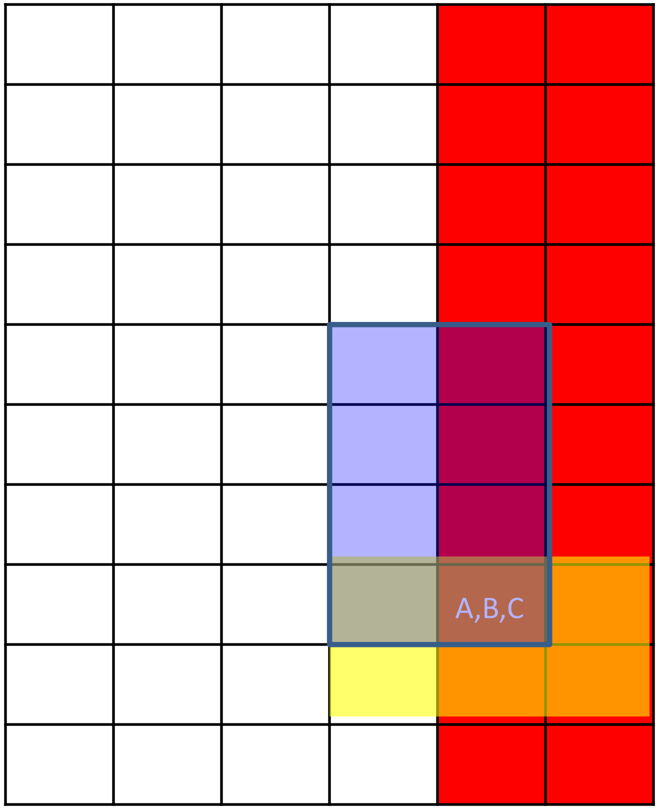
A and B and C:

P(A,B,C) =P(C|A,B)*P(A,B) =

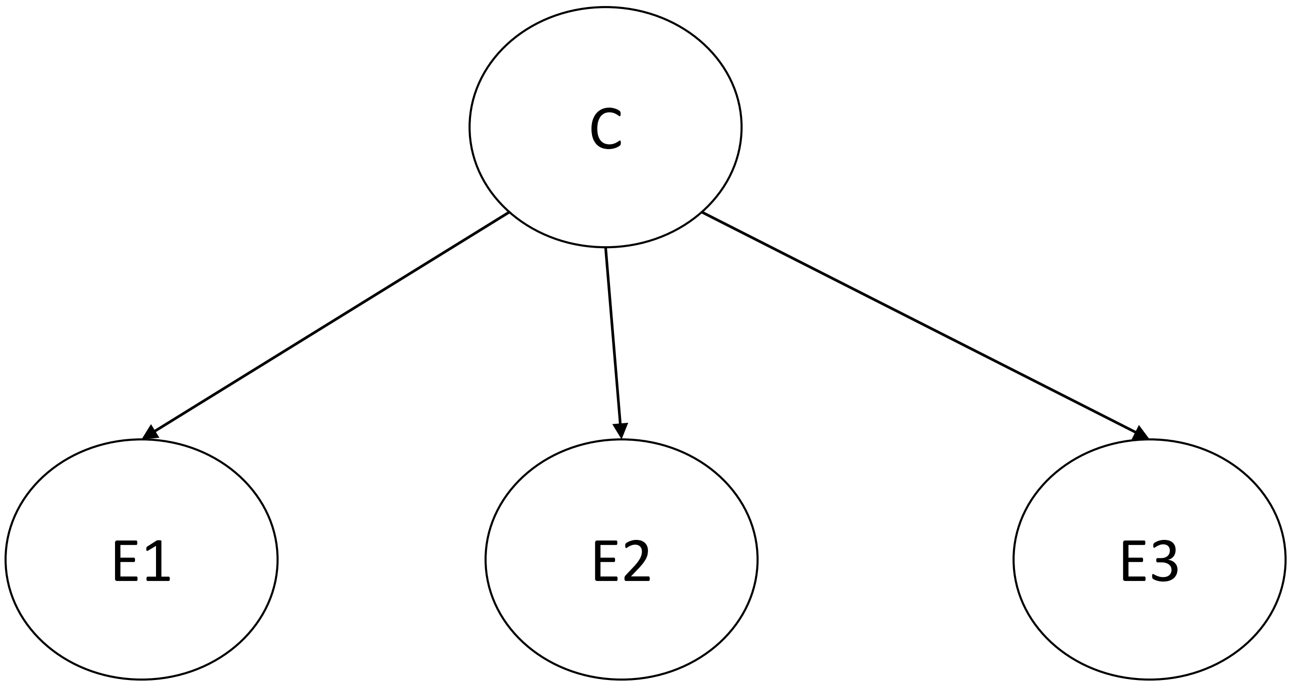
P(A,B) = P(B|A)*P(A)

P(C|A,B)*P(B|A)*P(A) = 1/4*4/60 =1/60

The Chain Rule

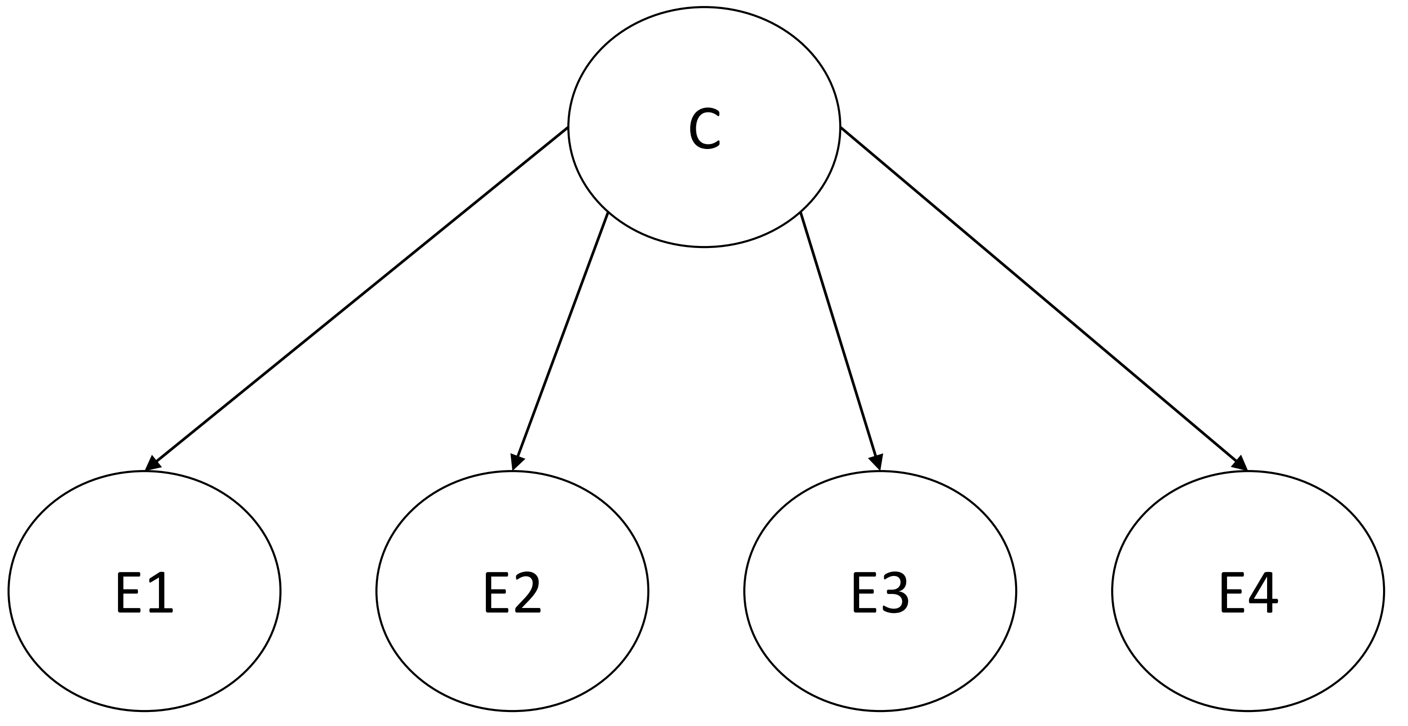


Naïve Bayes as a graph (network)



This graph states that there is a ***probabilistic dependence*** between C and each E_i. The probability of one of these variables (Class to predict) is influenced by the probabilities of the rest of the variables (set of evidences) and vice versa: $P(C|E) \neq P(C)$, and $P(E|C) \neq P(E)$

Bayesian networks model joint probability distribution for all variables



P(**c**|e1,e2,e3,e4)=P(**c**,e1,e2,e3,e4) /**P(e1,e2,e3,e4)**=α P(**c**,e1,e2,e3,e4)
P(**¬c**|e1,e2,e3,e4)=P(**¬c**,e1,e2,e3,e4)/**P(e1,e2,e3,e4)**= α P(**¬c**,e1,e2,e3,e4)

In fact, for prediction, it is enough to compute the joint probability of all known variables e1..e4 - for **c** and **¬c**, and to compare

Joint probability when e1-e3 are mutually **independent** events

- $P(c|e1,e2,e3) = P(c, e1,e2,e3) / P(e1,e2,e3)$
- $P(c|E)=P(c \cap E)/P(E)$

↑

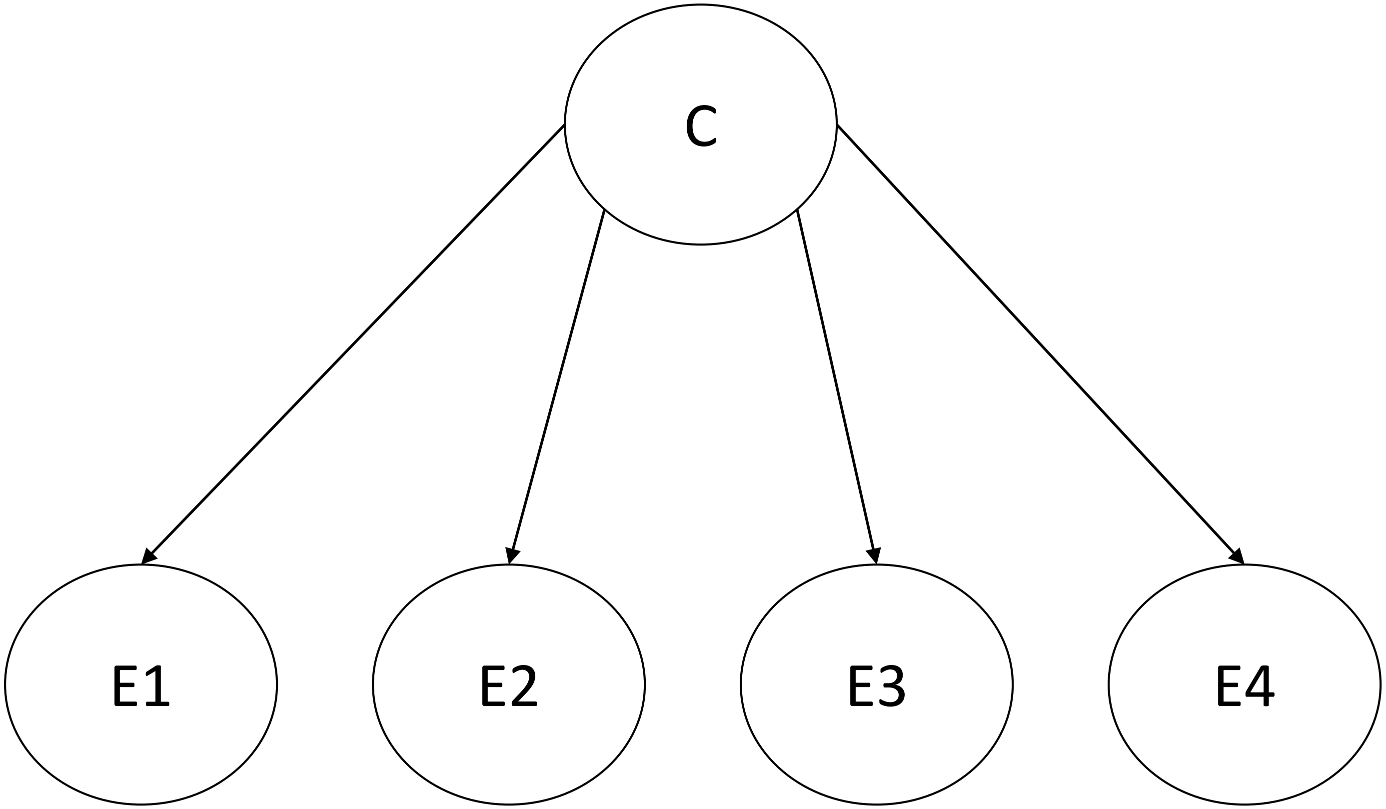
Joint probability of all variables in the network

We can compute the probability of all these events to happen together:

- $P(c \cap e1 \cap e2 \cap e3) = P(e1|c)\, P(e2|c)\, P(e3|c)\, {*}P(c)$

We multiply $P(e_i|c)$ because we assume: there is no probabilistic dependence between e_i and e_j, given the parent value C

Naïve Bayes

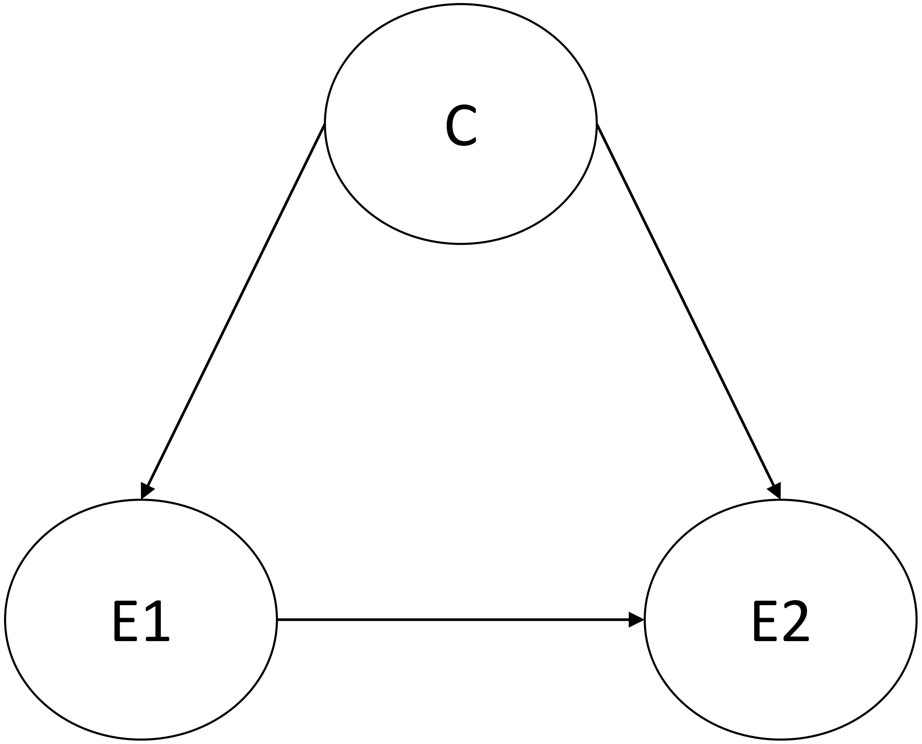


$P(\mathbf{c} \mid e1,e2,e3,e4) = \alpha\, P(c)P(e1 \mid c)P(e2 \mid c)P(e3 \mid c)$

$P(\neg\mathbf{c} \mid e1,e2,e3,e4) = \alpha\, P(\neg c)P(e1 \mid \neg c)P(e2 \mid \neg c)P(e3 \mid \neg c)$

Compare

More complex dependencies

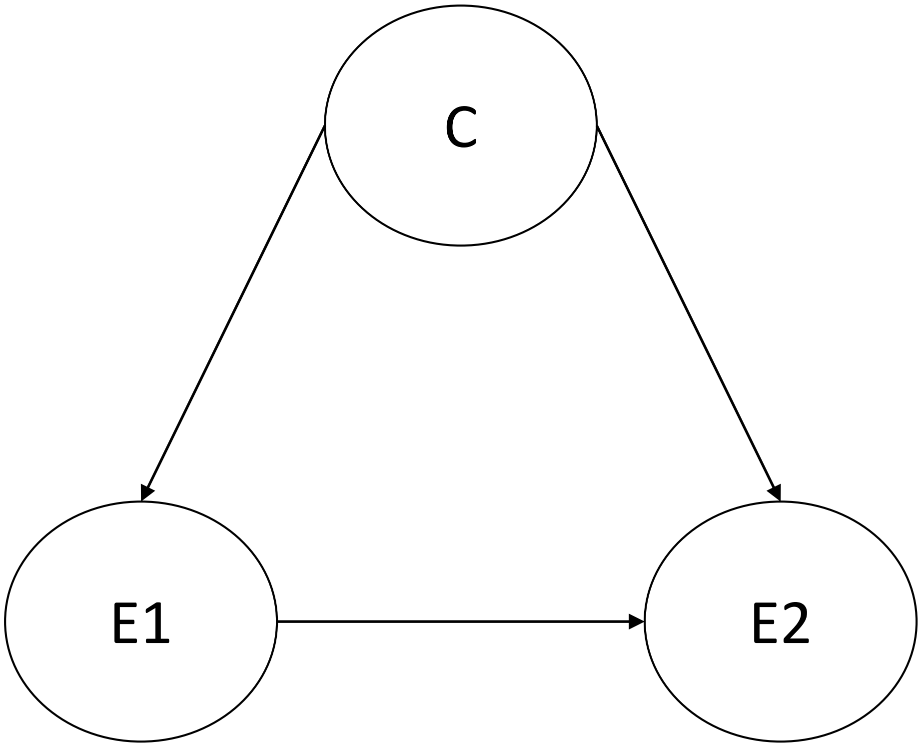


What if E1 and E2 are not independent?
For each node with more than 1 parent we need Conditional Probability Table (CPT) with probability distribution for all possible combinations of parent variables:

CPT for attribute E2

		E2	
C	E1	e2	¬e2
c	e1	P(e2|c,e1)	
c	¬e1		
¬c	e1		
¬c	¬e1		

Using the chain rule for complex dependencies



After all CPTs are computed for each node given all possible combinations of values of its parents, the joint probability is computed by the same chain rule.

$P(c|e1,e2)=\alpha$ **P(c)** $P(e1|c)P(e2|c,e1)$

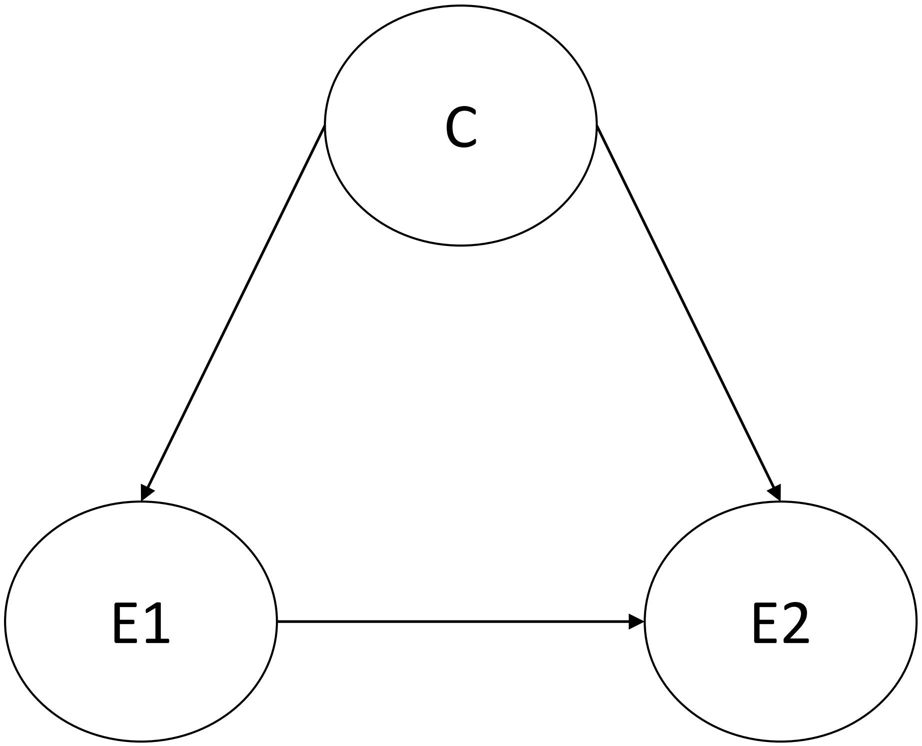
$P(\neg c|e1,e2)= \alpha$ **P(¬c)** $P(e1|\neg c)P(e2|\neg c,e1)$

↑ C does not have parents, so its probability is unconditional

CPT for attribute E2

		E2	
C	E1	e2	¬e2
c	e1	P(e2|c,e1)	
c	¬e1		
¬c	e1		
¬c	¬e1		

Using the chain rule for complex dependencies



$P(c|e1,e2)=\alpha\ P(c)\ \mathbf{P(e1|c)}\ P(e2|c,e1)$

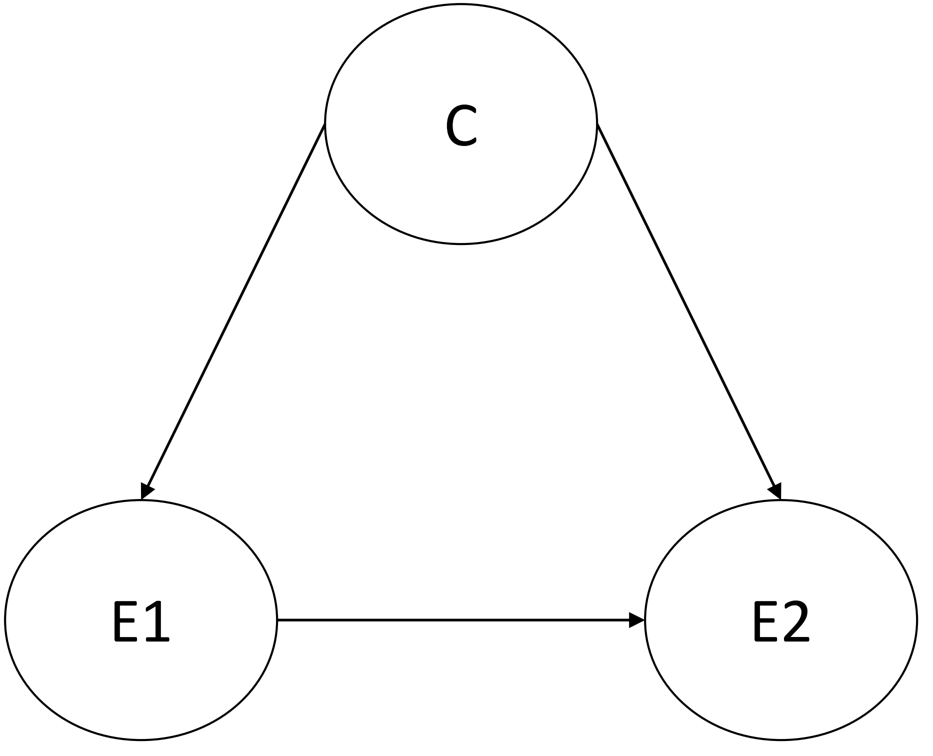
$P(\neg c|e1,e2)= \alpha\ P(\neg c)\ \mathbf{P(e1|\neg c)}\ P(e2|\neg c,e1)$

↑ E1 has 1 parent so its probability is conditioned on C

CPT for attribute E2

		E2	
C	E1	e2	¬e2
c	e1	P(e2|c,e1)	
c	¬e1		
¬c	e1		
¬c	¬e1		

Using the chain rule for complex dependencies



$P(c|e1,e2)=\alpha\ P(c)P(e1|c)\ \mathbf{P(e2|c,e1)}$

$P(\neg c|e1,e2)=\alpha\ P(\neg c)P(e1|\neg c)\ \mathbf{P(e2|\neg c,e1)}$

E2 has 2 parents so its probability is conditioned on both C and E1

CPT for attribute E2

		E2	
C	E1	e2	¬e2
c	e1	P(e2|c,e1)	
c	¬e1		
¬c	e1		
¬c	¬e1		

Estimating joint probabilities

- In a complex network of interrelated variables, it is easier to think in terms of joint probability of all known variables rather than a conditional probability of a class given evidence set
- This way we can predict not only a single attribute (Class) but also any other attribute, given that we know some evidences

- Instead of comparing:

P(c|e1,e2)=α **P(c)P(e1|c) P(e2|c,e1)**
P(¬c|e1,e2)= α **P(¬c)P(e1|¬c) P(e2|¬c,e1)**

- Compare just:

P(c)P(e1|c) P(e2|c,e1)
P(¬c)P(e1|¬c) P(e2|¬c,e1)

Probabilities of all 3 random events to happen together

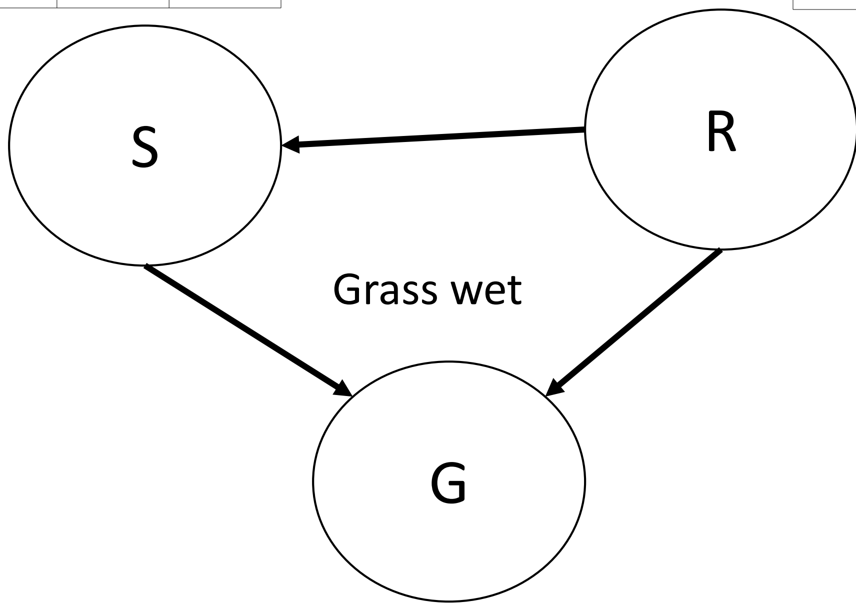
Explanation by example: predicting rain

	Sprinkler	
Rain	s	¬s
r	0.40	0.60
¬r	0.99	0.01

Rain	
r	¬r
0.20	0.80

We know that Sprinkler was off: S= ¬s
and grass is wet: G=g

Was it raining?
P(r|g, ¬s)=?
P(¬ r|g, ¬s)=?

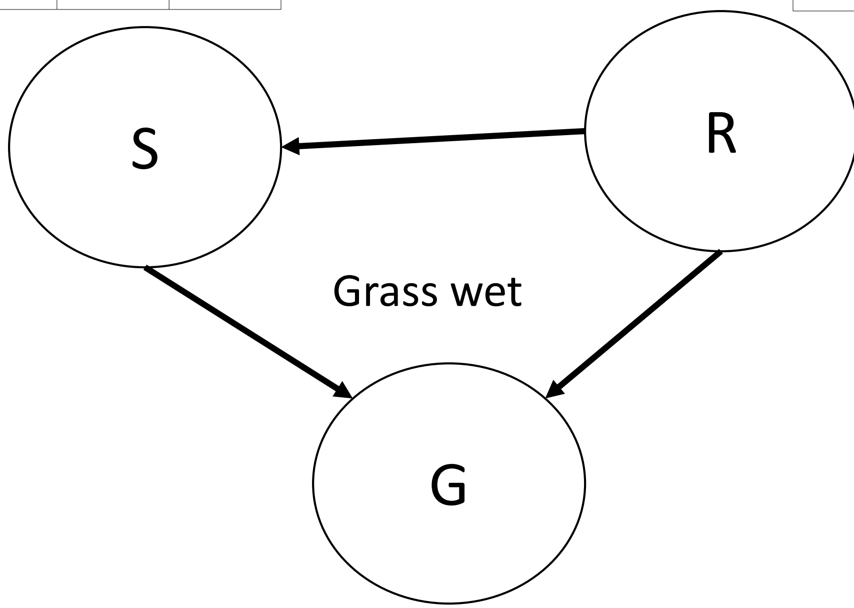


		Grass wet	
Rain	Sprinkler	g	¬g
r	s	0.99	0.01
r	¬s	0.8	0.2
¬r	s	0.90	0.10
¬r	¬s	0.01	0.99

Wet Grass example: predicting rain

	Sprinkler	
Rain	s	¬s
r	0.40	0.60
¬r	0.99	0.01

Rain	
r	¬r
0.20	0.80



		Grass wet	
Rain	Sprinkler	g	¬g
r	s	0.99	0.01
r	¬s	0.8	0.2
¬r	s	0.90	0.10
¬r	¬s	0.01	0.99

$S = \neg s$, $G = g$

$P(r|g, \neg s) = ?$

$P(\neg r|g, \neg s) = ?$

$P(r|g, \neg s) = \alpha P(r, g, \neg s)$

$= \alpha P(r) P(\neg s|r) P(g|r, \neg s)$

$P(\neg r|g, \neg s) = \alpha P(\neg r, g, \neg s)$

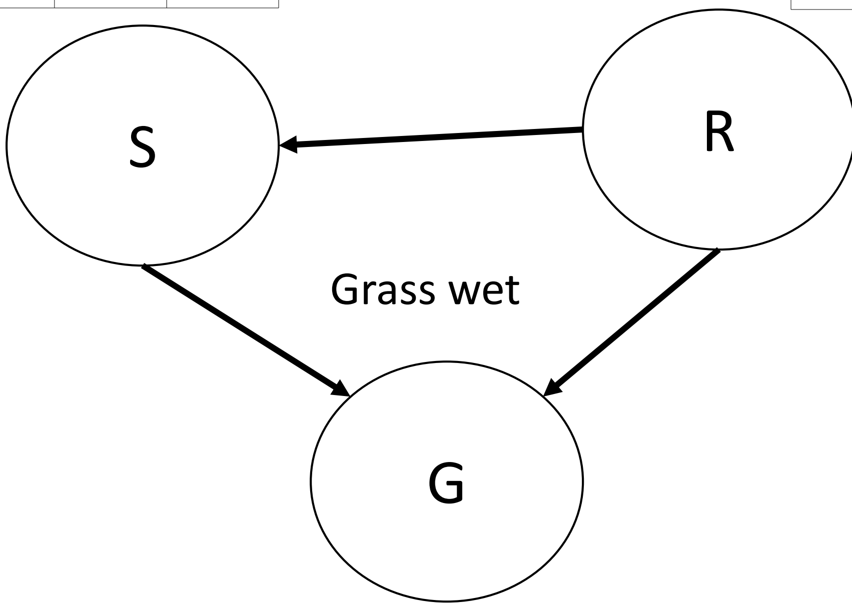
$= \alpha P(\neg r) P(\neg s|\neg r) P(g|\neg r, \neg s)$

All probabilities are given in CPTs, so we just plug in and compute

Wet Grass example: predicting rain

	Sprinkler	
Rain	s	¬s
r	0.40	0.60
¬r	0.99	0.01

Rain	
r	¬r
0.20	0.80



		Grass wet	
Rain	Sprinkler	g	¬g
r	s	0.99	0.01
r	¬s	0.8	0.2
¬r	s	0.90	0.10
¬r	¬s	0.01	0.99

$S = \neg s$, $G = g$

$P(r \mid g, \neg s) = ?$
$P(\neg r \mid g, \neg s) = ?$

$P(r \mid g, \neg s)$
$= \alpha\, P(r)\, P(\neg s \mid r)\, P(g \mid r, \neg s)$
$= \alpha\, 0.20 * 0.60 * 0.8$
$= \alpha\, 0.096$

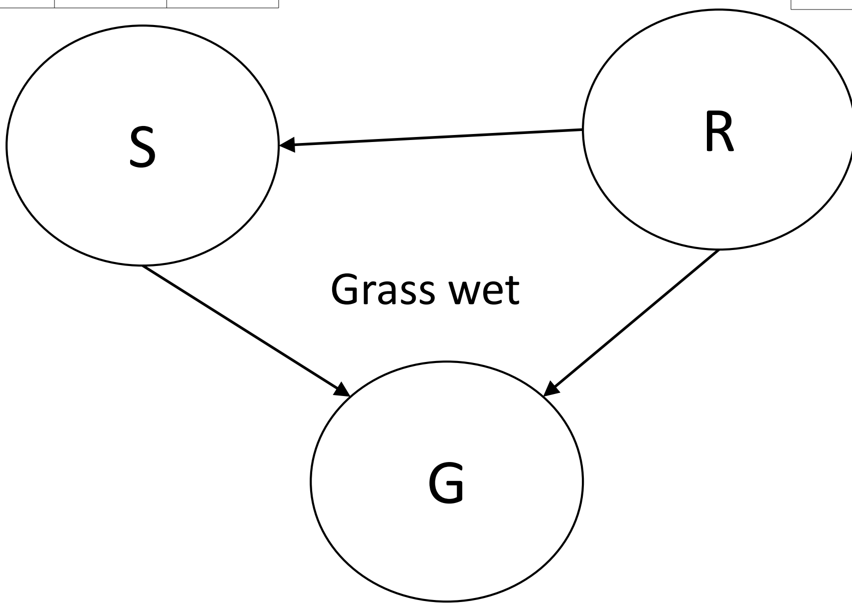
$P(\neg r \mid g, \neg s)$
$= \alpha\, P(\neg r)\, P(\neg s \mid \neg r)\, P(g \mid \neg r, \neg s)$
$= \alpha\, 0.80 * 0.01 * 0.01$
$= \alpha\, 0.00008$

Definitely, it was raining

Wet Grass example: hidden variables

	Sprinkler	
Rain	s	¬s
r	0.40	0.60
¬r	0.99	0.01

Rain	
r	¬r
0.20	0.80



		Grass wet	
Rain	Sprinkler	g	¬g
r	s	0.99	0.01
r	¬s	0.8	0.2
¬r	s	0.90	0.10
¬r	¬s	0.00	1.00

All we know that the grass is wet: G=g

P(r|g)=?

The value of S is unknown: S is a hidden variable which influences G and depends on R. We need to include it into the joint probability:

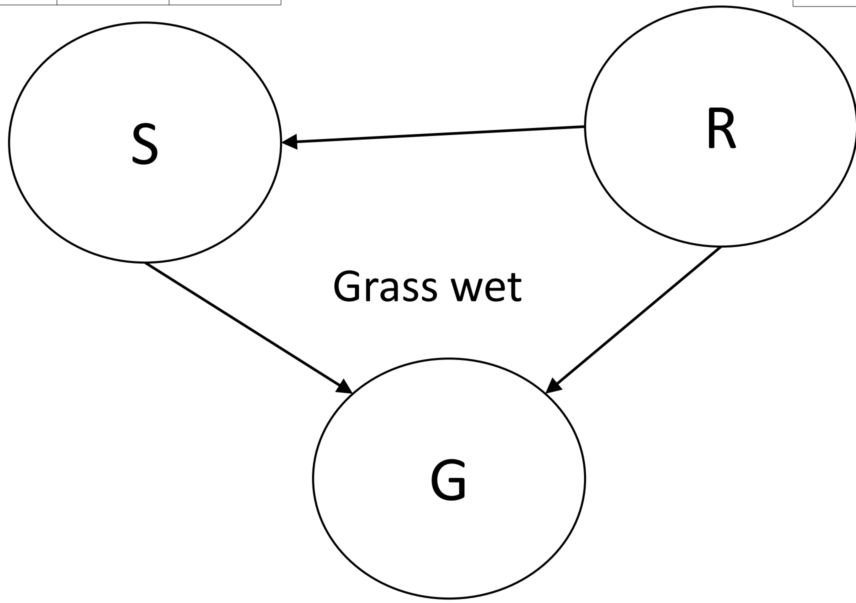
$P(r|g)=\alpha\, P(r, g, S)$
$= \alpha\, P(r)\, P(S|r)\, P(g|r, S)=$
$= \alpha\, P(r)\ *$
$[P(s|r)\, P(g|r, s)+ P(\neg s|r)\, P(g|r, \neg s)]$

Sprinkler was or on or off

Wet Grass example: hidden variables

	Sprinkler	
Rain	s	¬s
r	0.40	0.60
¬r	0.99	0.01

Rain	
r	¬r
0.20	0.80



		Grass wet	
Rain	Sprinkler	g	¬g
r	s	0.99	0.01
r	¬s	0.8	0.2
¬r	s	0.90	0.10
¬r	¬s	0.00	1.00

All we know that the grass is wet:
G=g

$P(r|g)=?$

$P(r|g)=\alpha\, P(r, g, S_{T \vee F})$
$= \alpha\, P(r)\, P(S_{T \vee F}|r)\, P(g|r, S_{T \vee F})=$
$= \alpha\, P(r)$ *
$[P(s|r)\, P(g|r, s) + P(\neg s|r)\, P(g|r, \neg s)]$

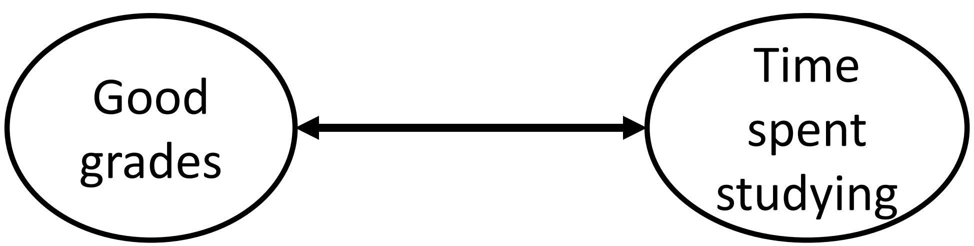
We add because we don’t know the value of S, and we consider it as being or false, or true.
We apply theorems V, VI from PROBABILITY slides

Hidden (missing) evidences

- For each hidden variable consider all possible values of this variable and perform summation by substituting this variable with all possible values in turn

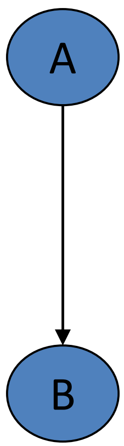
Bayesian Belief Networks (BBN)

- BBN is a graphical representation (Directed acyclic graph (DAG) – no cycles) of probabilistic dependencies between variables
- They combine reasoning with probabilities
- Nodes: random variables
- At each node: Conditional Probability Table (CPT) - the probabilities for all different values of the node variable given all possible value combinations of its parents
- The directed edges show probabilistic influence – dependence between variables. Edges can be drawn in any direction: the direction is application-dependent



BBN types: possible meaning of edges

Causal BBN

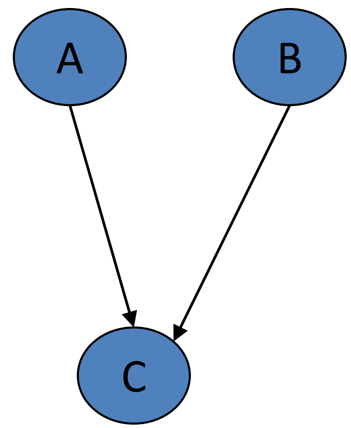


Increased probability of A makes B more likely.

A causes B

We know P(B|A) - diagnostics

Intercausal

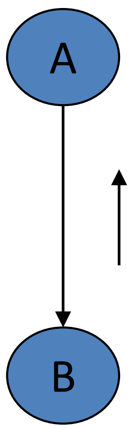


A and B can each cause C. B explains C and so is

evidence against A

We need to know P(C|A), P(C|B), and P(C|A,B)

Evidential



Increased probability of B makes A more likely.

B is evidence for A, A depends on B

We know P(B|A)

Using Bayesian Belief Networks for prediction

- Each query asks for a joint probability which is computed by applying the chain rule (multiplying corresponding conditional probabilities for each variable involved in the query and its dependants)

- This is because all conditional probabilities for each node given its parent are in CPTs, and each query for conditional probability of a parent given its children can be computed using Bayes theorem

Example: weather data

Windy

play	outlook	false	true
Yes	sunny	.500	.500
Yes	overcast	.500	.500
Yes	rainy	0.125	0.875
No	sunny	0.375	0.625
No	overcast	0.500	0.500
No	rainy	0.833	0.167

Play

yes	no
.633	.367

Humidity

play	temp	high	normal
Yes	hot	.500	.500
Yes	mild	.500	.500
Yes	cool	.125	.875
No	hot	.833	0.167
No	mild	.833	.167
No	cool	.250	.750

Outlook

play	sunny	overcast	rainy
Yes	.238	.429	.333
No	.538	.077	.385

Temperature

play	**outlook**	hot	mild	cool
Yes	sunny	.143	.429	.429
Yes	overcast	.455	.273	.273
Yes	rainy	.111	.556	.333
No	sunny	.556	.333	.111
No	overcast	.333	0.333	.333
No	rainy	.143	.429	.429

After all CPTs are filled in, we can perform any query on joint distribution

Joint probability: weather data

Outlook	Temp.	Humidity	Windy	Play
Sunny	Cool	High	True	?

Play

yes	no
.633	.367

Windy

play	outlook	false	true
Yes	sunny	.500	.500
Yes	overcast	.500	.500
Yes	rainy	0.125	0.875
No	sunny	0.375	0.625
No	overcast	0.500	0.500
No	rainy	0.833	0.167

Humidity

play	temp	high	normal
Yes	hot	.500	.500
Yes	mild	.500	.500
Yes	cool	.125	.875
No	hot	.833	0.167
No	mild	.833	.167
No	cool	.250	.750

Outlook

play	sunny	overcast	rainy
Yes	.238	.429	.333
No	.538	.077	.385

Temperature

play	outlook	hot	mild	cool
Yes	sunny	.143	.429	.429
Yes	overcast	.455	.273	.273
Yes	rainy	.111	.556	.333
No	sunny	.556	.333	.111
No	overcast	.333	0.333	.333
No	rainy	.143	.429	.429

P(Yes|Sunny,Cool,High,True)=αP(Yes,Sunny,Cool,High,True) = α **P(Yes) P(Sunny|Yes) P(Cool|Yes,Sunny) P(High|Yes, Cool) P (True| Yes, Sunny)**

Joint probability: weather data

Outlook	Temp.	Humidity	Windy	Play
Sunny	Cool	High	True	?

Play

yes	no
.633	.367

Windy

play	outlook	false	true
Yes	sunny	.500	.500
Yes	overcast	.500	.500
Yes	rainy	0.125	0.875
No	sunny	0.375	0.625
No	overcast	0.500	0.500
No	rainy	0.833	0.167

Humidity

play	temp	high	normal
Yes	hot	.500	.500
Yes	mild	.500	.500
Yes	cool	.125	.875
No	hot	.833	0.167
No	mild	.833	.167
No	cool	.250	.750

Outlook

play	sunny	overcast	rainy
Yes	.238	.429	.333
No	.538	.077	.385

Temperature

play	outlook	hot	mild	cool
Yes	sunny	.143	.429	.429
Yes	overcast	.455	.273	.273
Yes	rainy	.111	.556	.333
No	sunny	.556	.333	.111
No	overcast	.333	0.333	.333
No	rainy	.143	.429	.429

P(Yes|Sunny,Cool,High,True)=αP(Yes,Sunny,Cool,High,True) = α **P(Yes) P(Sunny|Yes) P(Cool|Yes,Sunny) P(High|Yes, Cool) P (True| Yes, Sunny)**

All these probabilities are known – just plug them in and compute

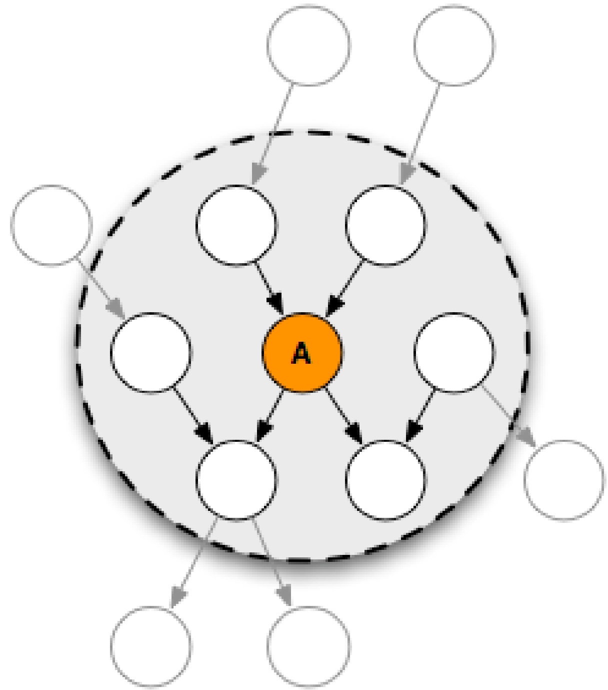
Markov Blanket Assumption

- All nodes in the network are connected in some way
- The key feature of Bayesian Networks, which allows us to use the chain rule, is the assumption that the probability of each node is influenced only by the nodes in the Markov blanket of this node:

- The *Markov blanket* of a node is its set of neighboring nodes: its ***parents***, its ***children***, and any other ***parents of its children***.

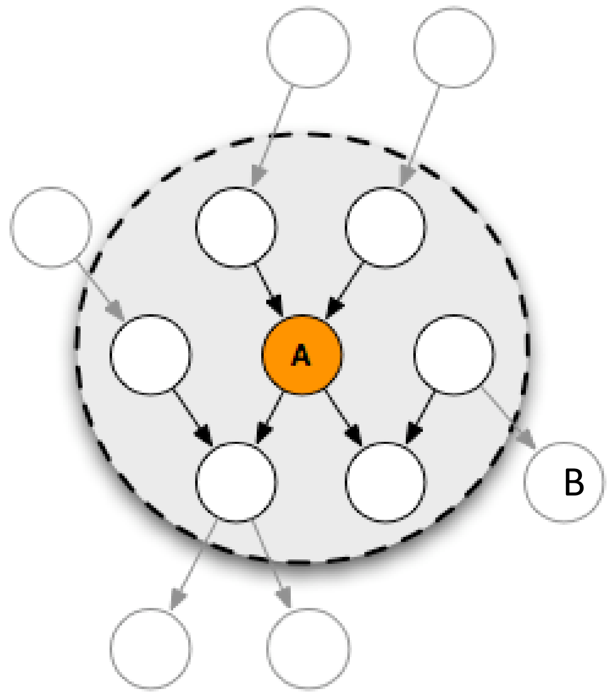
- **No grandparents, no grandchildren, no children of its parent.**

Markov Blanket of node A



- The Markov blanket of a node contains all the variables that shield the node from the rest of the network.
- This means that the Markov blanket of a node is the only knowledge needed to predict the behavior of that node.

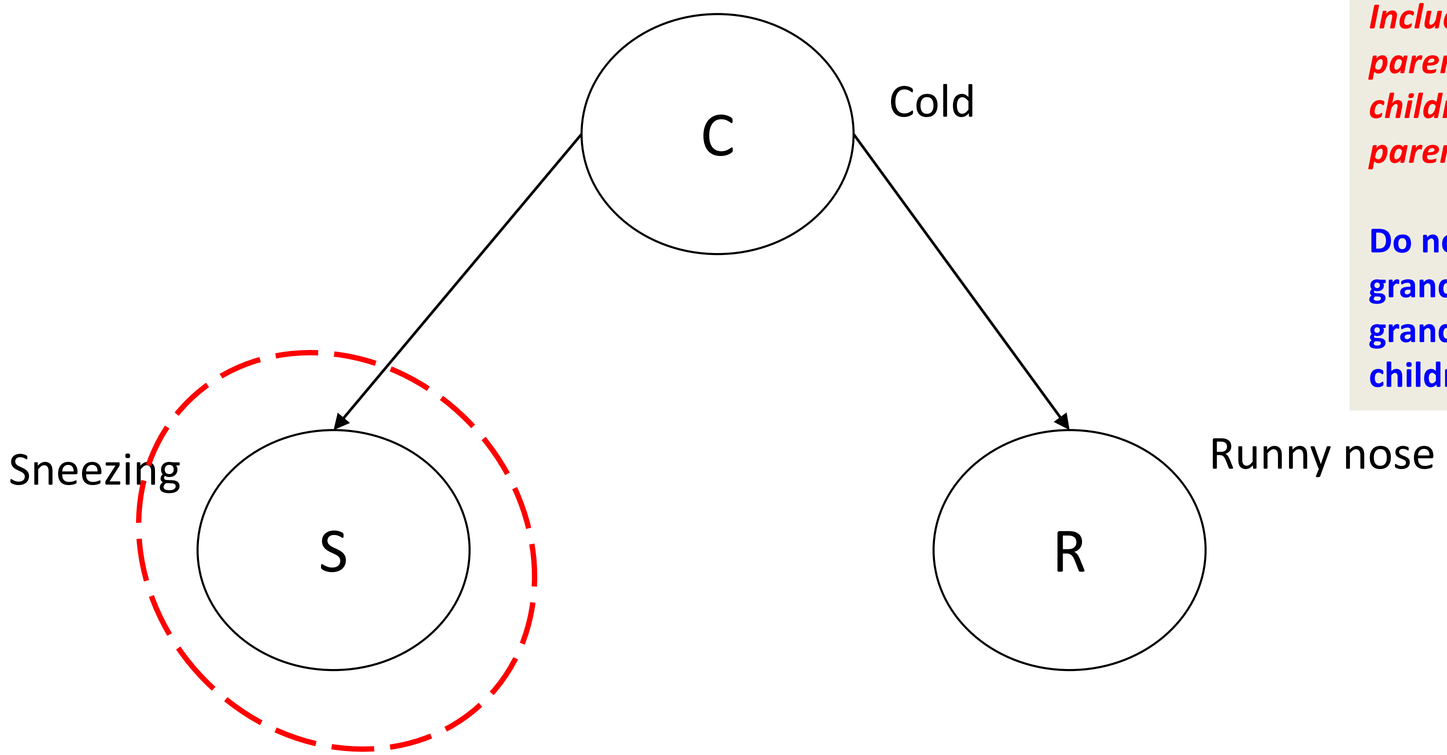
The Markov blanket assumption



- Markov blanket assumes that P(A|B)=P(A) – probability of A is not influenced by the value of B, if B is outside of the blanket

- This corresponds to our intuition about probabilistic influences

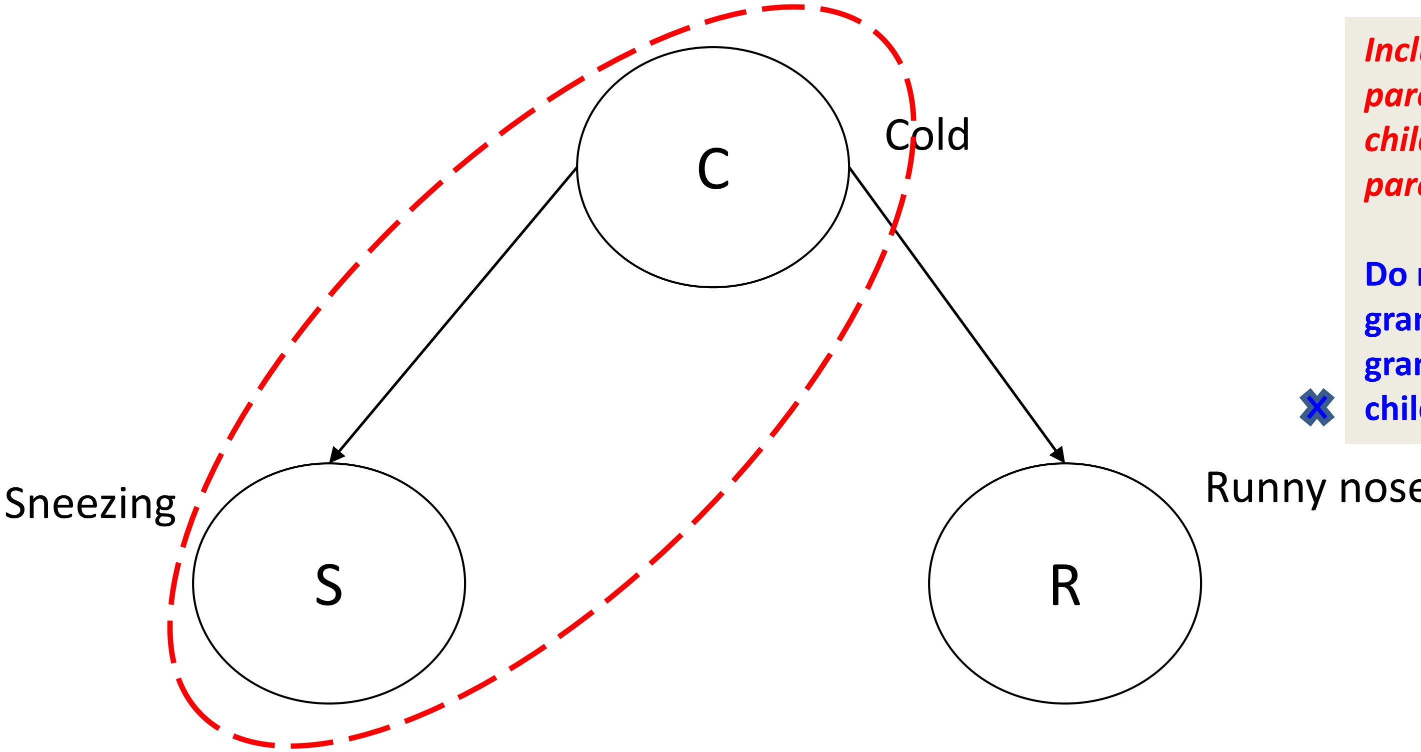
Example 1: Markov blanket of S



Include:
parents
children
parents of its children

Do not include:
grandparents
grandchildren
children of its parent

Example 1: Markov blanket of S

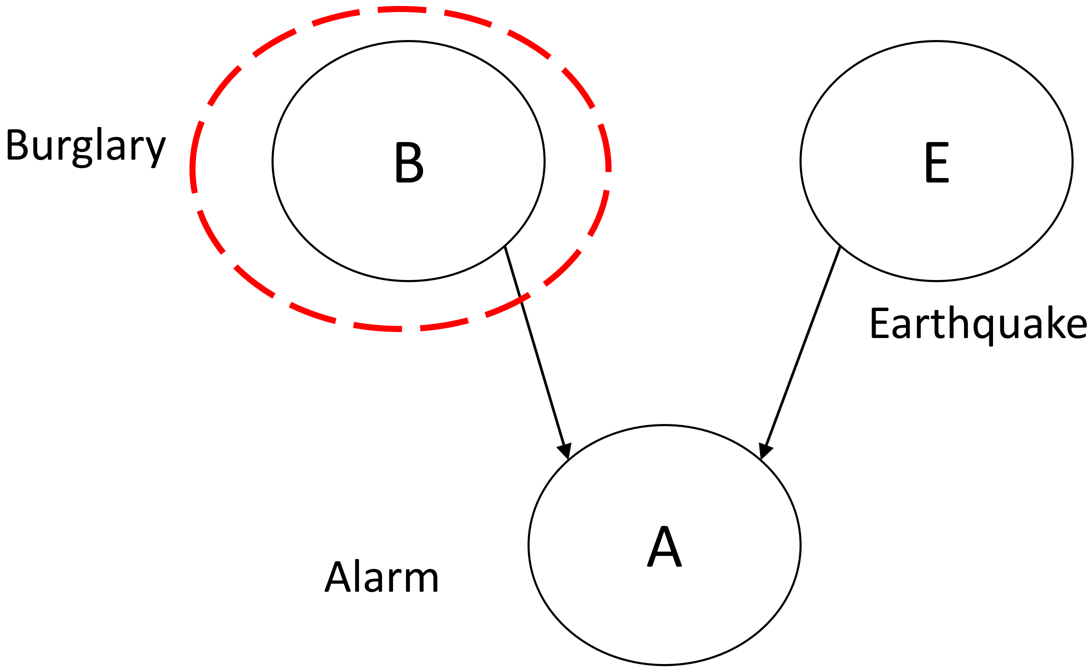


Include:
parents
children
parents of its children

Do not include:
grandparents
grandchildren
✖ **children of its parent**

P(S|R)>P(S)
Thus, S in general is not independent of R:
R makes C more probable, which in turn influences the probability of S.
However, P(S|C) is independent of R: if we know the value of C (C is given), then R does not influence the probability of S:
P(S|C,R)=P(S|C)   - C 'shields' node S from the influence of R

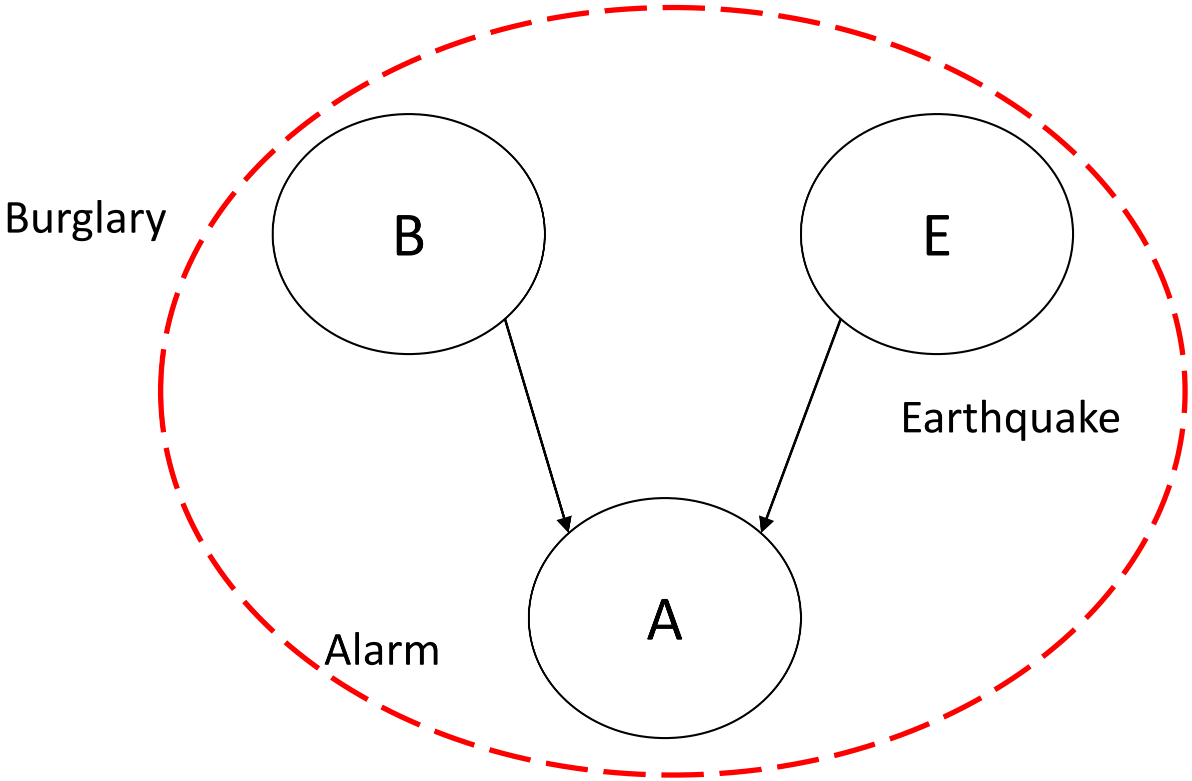
Example 2: Markov blanket of B



Include:
parents
children
parents of its children

Do not include:
grandparents
grandchildren
children of its parent

Example 2: Markov blanket of B

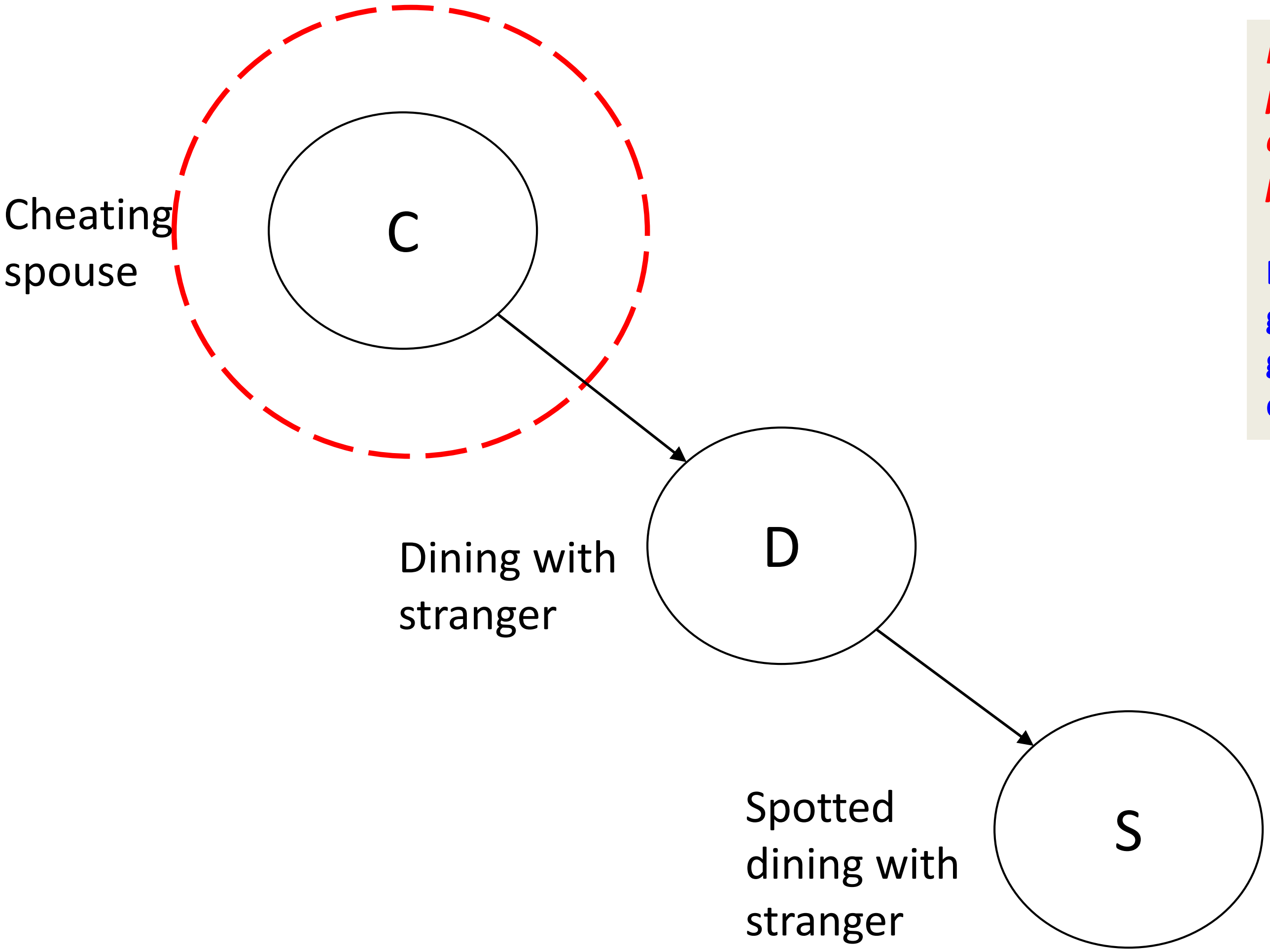


Include:
parents
children
parents of its children

Do not include:
grandparents
grandchildren
children of its parent

P(B|E)=P(B) (independent), but P(B|A,E)<P(B|A)
If you hear an alarm, you might evaluate the probability of B, but if you know that there was an earthquake, this probability decreases:
E 'discounts' B, E is evidence against B, and it should be included in its Markov blanket together with A

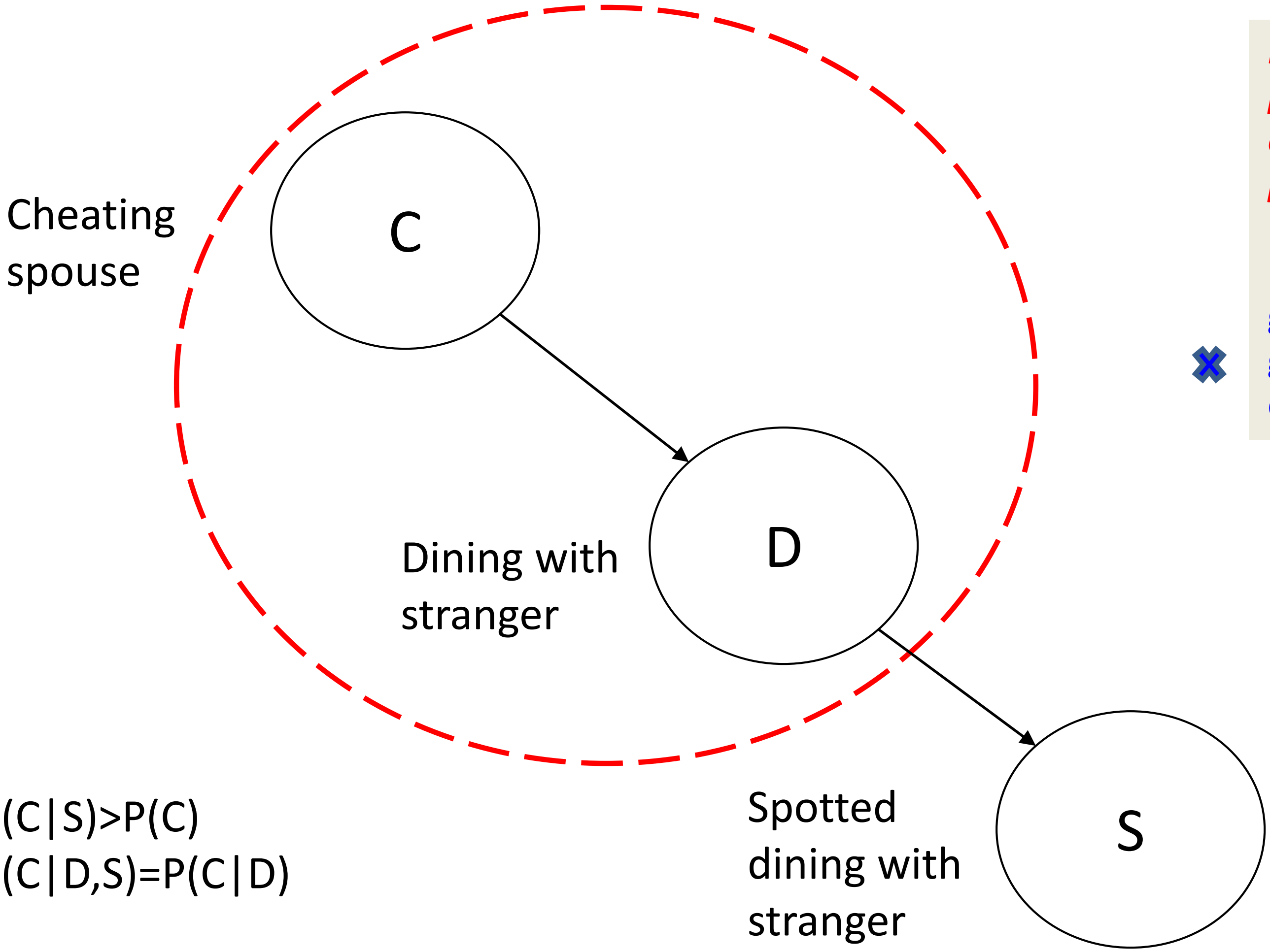
Example 3: Markov blanket of C



Include:
parents
children
parents of its children

Do not include:
grandparents
grandchildren
children of its parent

Example 3: Markov blanket of C



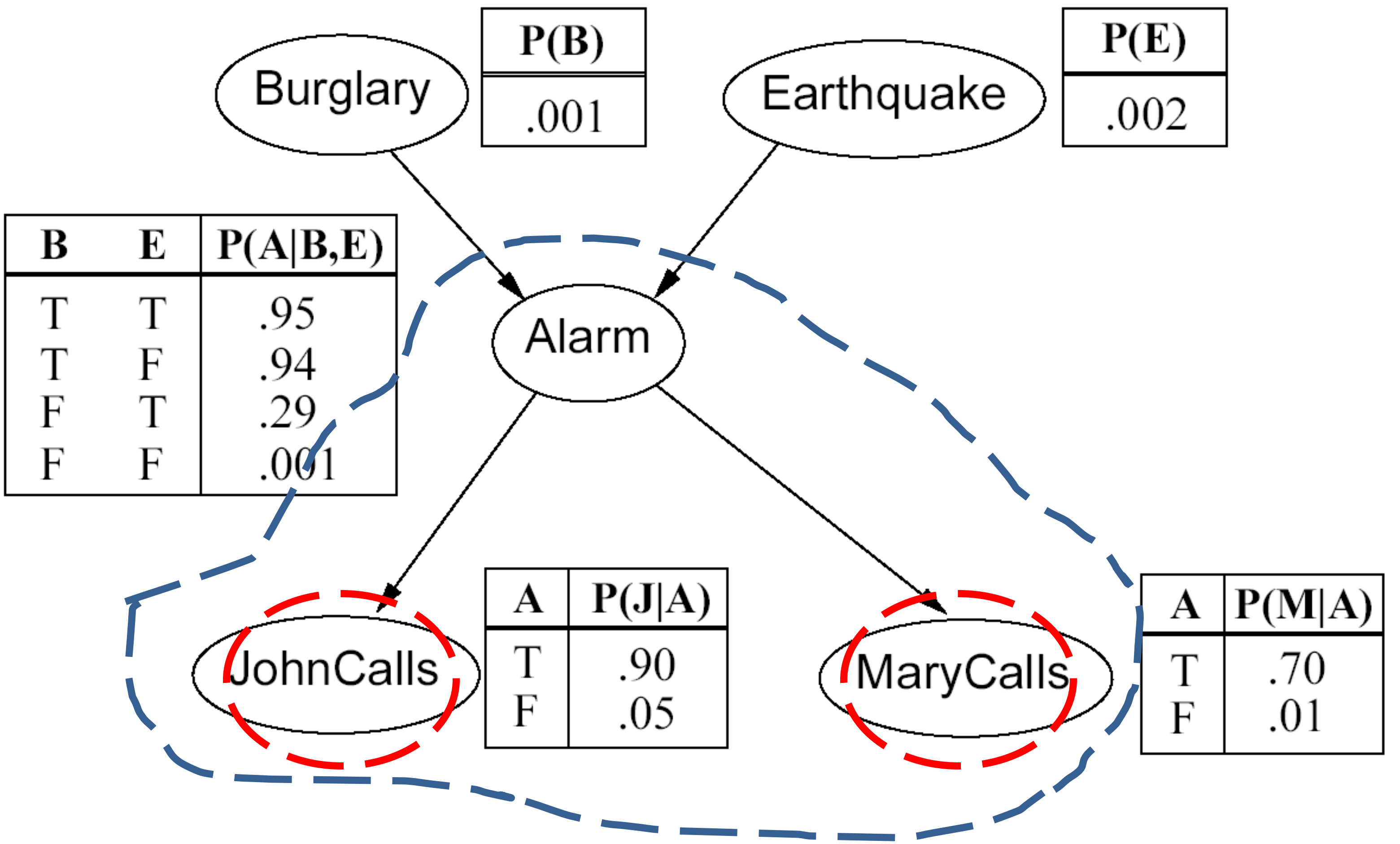
Include:
parents
children
parents of its children

Do not include:
grandparents
grandchildren
children of its parent

$P(C|S)>P(C)$
$P(C|D,S)=P(C|D)$

If D is known (given), then there is no influence of hearsay S on the probability of C

Alarm example



P(B)
.001

P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

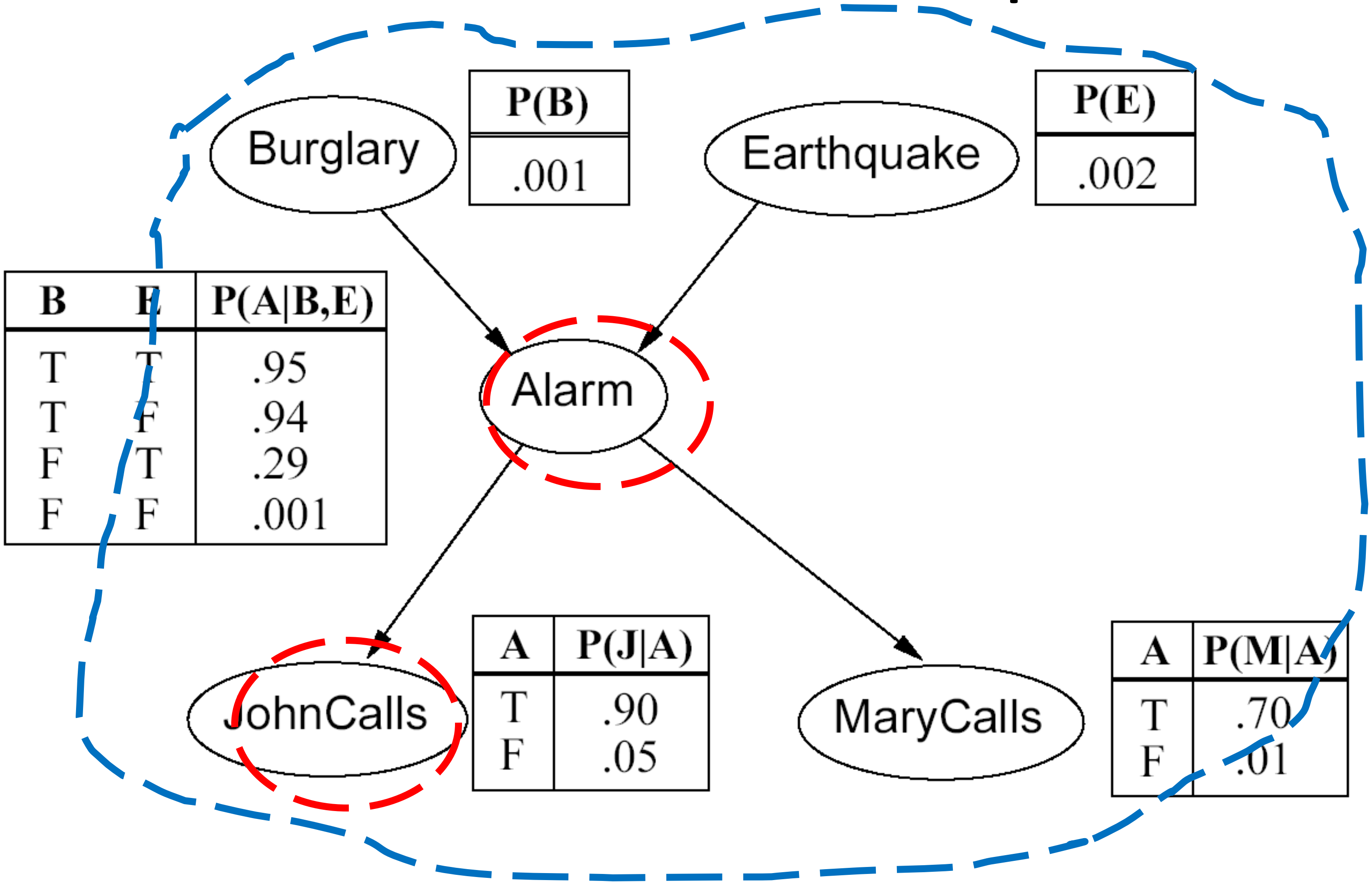
A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Query: what is the probability of John calling given that Mary called

parents
children
parents of its children

Alarm example



P(B)
.001

P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

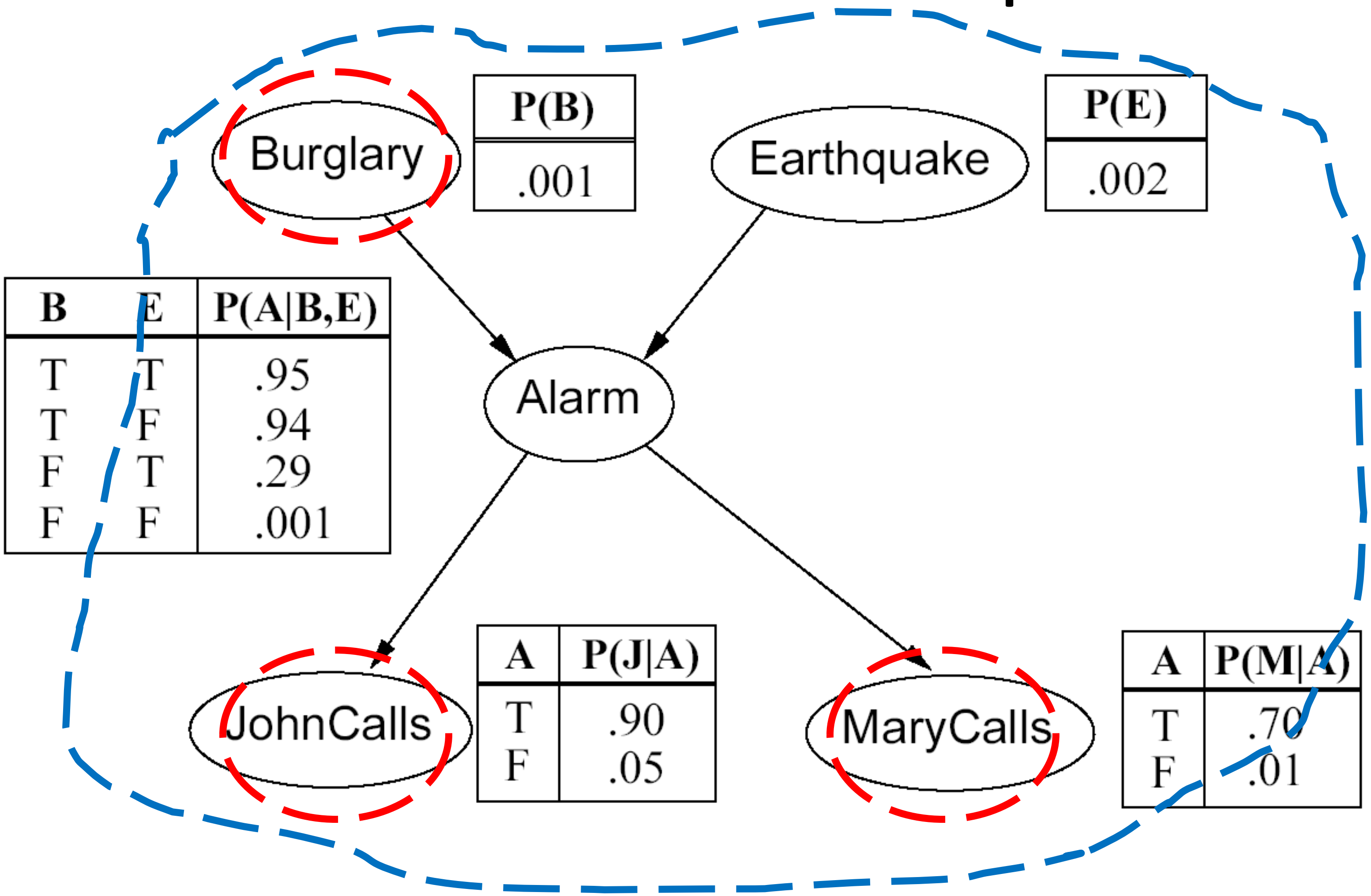
A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Query: what is the probability of Alarm given that John called

parents
children
parents of its children

Alarm example



P(B)
.001

P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

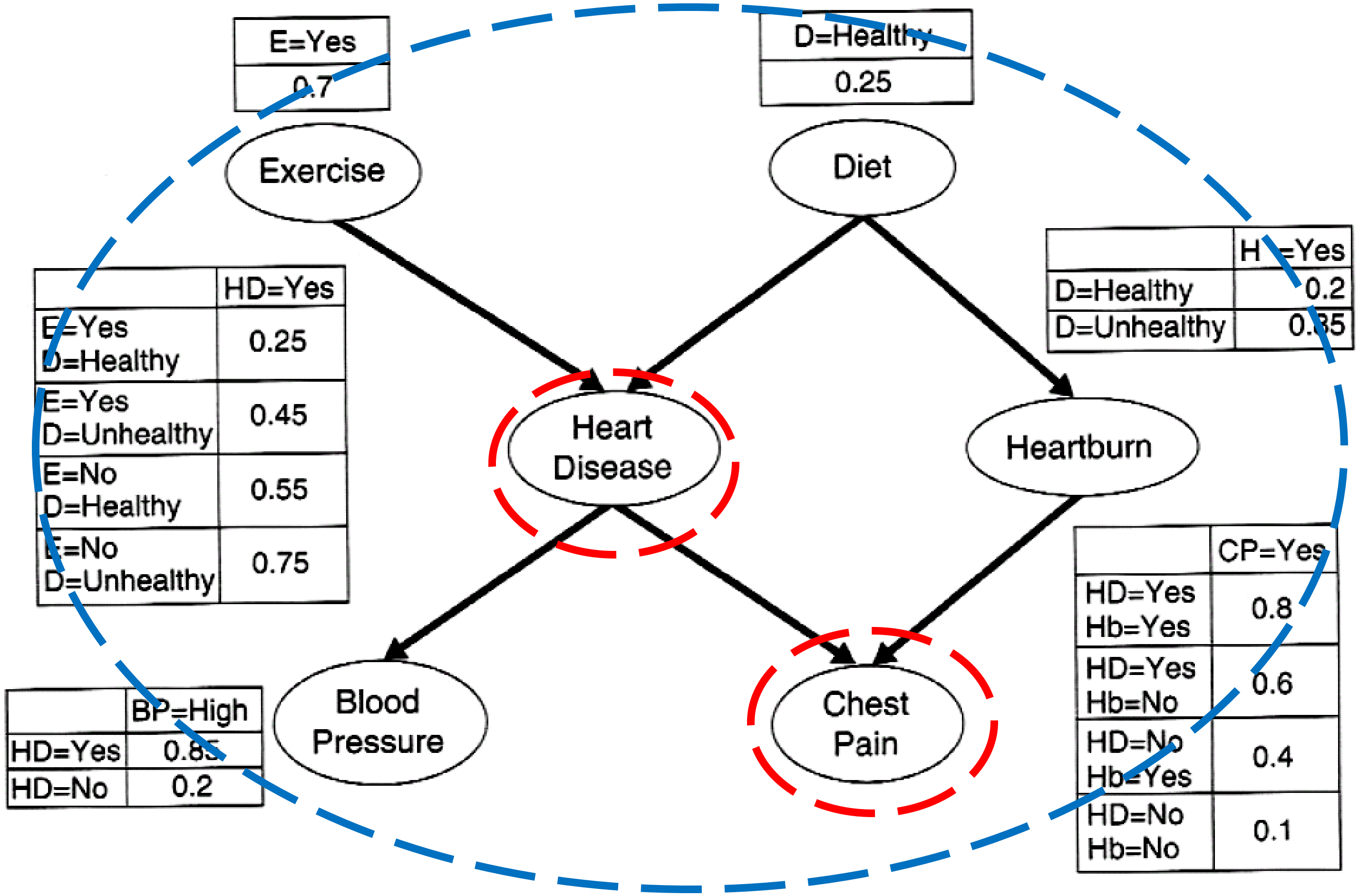
A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Query: what is the probability of Burglary given that John called and Mary called

parents
children
parents of its children

High Blood Pressure example



Query: what is the probability of Heart disease given chest pain

parents
children
parents of its children

Algorithm for classification using BBN

- In complex networks: select a subset of nodes which are inside Market blankets of nodes participating in the query
- Compute joint probabilities of all these nodes by the chain rule, substituting random variables by the evidence values
- If some of the values are unknown (hidden), sum up over all possible values

Bayesian Belief Networks: applications

- Very important technology in the Machine Learning / AI field
- A clean, clear, manageable language and methodology for expressing what you're certain and uncertain about
- Many practical applications in medicine, factories, helpdesks:

  P(this problem | these symptoms)

  anomalousness of this observation

  choosing next diagnostic test | these observations

Pathfinder system*

- Diagnostic system for lymph-node diseases.
- 60 diseases and 100 symptoms and test-results.
- 14,000 probabilities
- Experts consulted to make net. Apparently, the experts found it quite easy to invent the causal links and probabilities.
- 8 hours to determine variables.
- 35 hours for net topology.
- 40 hours for probability table values.
- Pathfinder is now outperforming the world experts in diagnosis.
- Being extended to several dozen other medical domains.

---

* Heckerman 1991, Probabilistic Similarity Networks, MIT Press, Cambridge MA

EXERCISES

I: Burglary

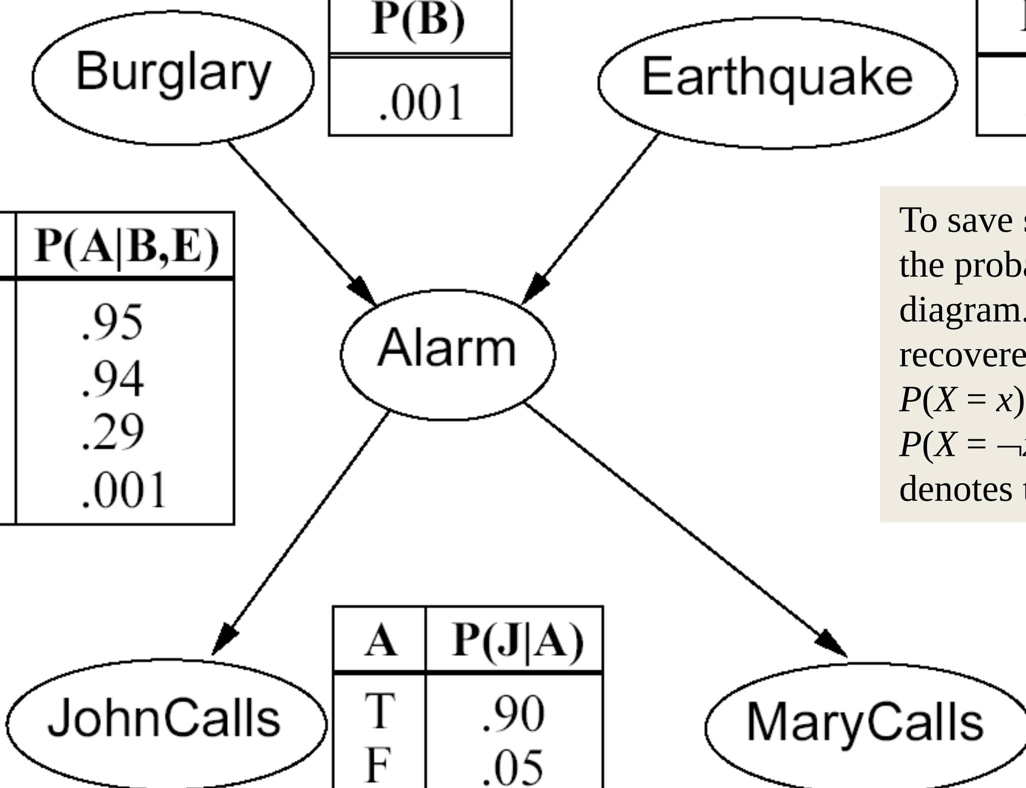
- I'm at work, neighbor John calls to say my alarm is ringing, and also my neighbor Mary calls. Sometimes the alarm is set off by minor earthquakes. Is there a burglar?

- John always calls when he hears the alarm, but sometimes confuses the telephone ringing with the alarm.
- Mary likes rather loud music and sometimes misses the alarm.

- 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

I: Burglary



P(B)
.001

P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

To save space, some of the probabilities have been omitted from the diagram. The omitted probabilities can be recovered by noting that
$P(X = x) = 1 - P(X = \neg x)$ and
$P(X = \neg x|Y) = 1 - P(X = \neg x|Y)$, where $\neg x$ denotes the opposite outcome of x.

The topology shows that burglary and earthquakes directly affect the probability of alarm, but whether Mary or John call depends only on the alarm.

Our assumptions are that they don't perceive any burglaries directly, and they don't confer before calling.

I: Prediction

- Suppose, we are given for the evidence variables $E_1,\dots,E_m$, their values $e_1,\dots,e_m$, and we want to predict whether the query variable X has the value x or not.
- For this we compute and compare the following:

$$P(x \mid e_1,\dots,e_m) = \frac{P(x, e_1,\dots,e_m)}{P(e_1,\dots,e_m)} = \alpha P(x, e_1,\dots,e_m)$$

$$P(\neg x \mid e_1,\dots,e_m) = \frac{P(\neg x, e_1,\dots,e_m)}{P(e_1,\dots,e_m)} = \alpha P(\neg x, e_1,\dots,e_m)$$

- How do we compute:

$$\alpha P(x, e_1,\dots,e_m)$$

and

$$\alpha P(\neg x, e_1,\dots,e_m)?$$

What about the hidden variables $Y_1,\dots,Y_k$?

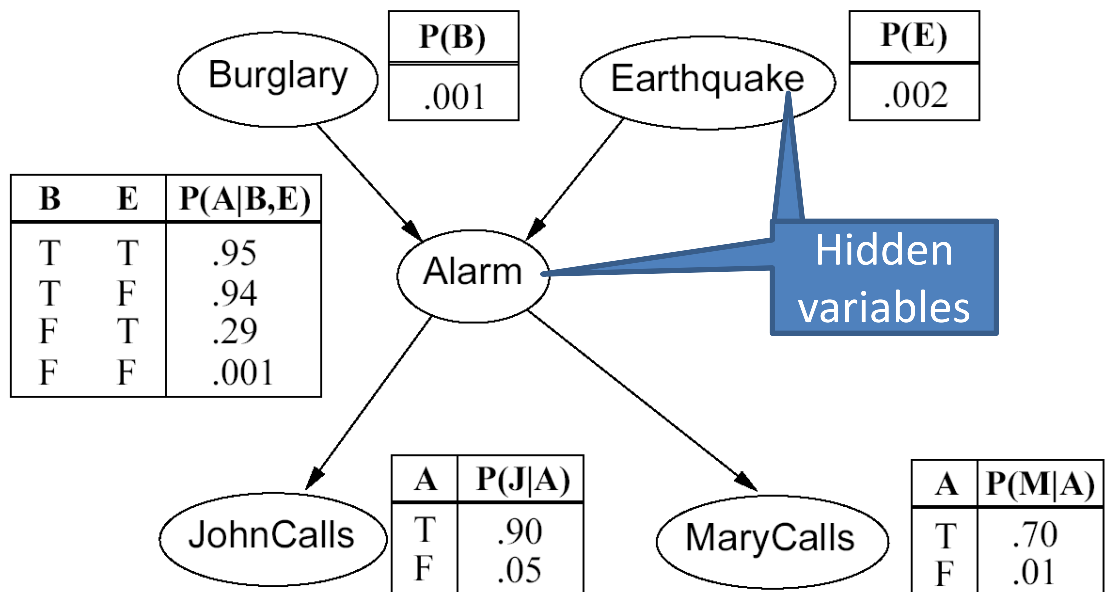
I: Classification of burglary and earthquake

- We are given for the evidence variables *J=j and M=m,* and we want to predict whether the query variable *B* has the value *b* or not *b*.
- However, to evaluate the probability of *B* we need to know: whether alarm really went off and whether it was an earthquake.
- *A* and *E* are hidden variables

$$\alpha P(x, e_1, \ldots, e_m)$$

and

$$\alpha P(\neg x, e_1, \ldots, e_m)?$$



I: Inference by enumeration

P(burglary | johhcalls, marycalls)? (Abbrev. $P(b|j,m)$)

$$P(b \mid j,m)$$
$$= \alpha P(b, j, m)$$
$$= \alpha \sum_a \sum_e P(b, j, m, A, E)$$
$$= \alpha \big(P(b, j, m, a, e) + P(b, j, m, \neg a, e) + P(b, j, m, a, \neg e) + P(b, j, m, \neg a, \neg e)\big)$$

Or

Alarm rings, earthquake

No alarm, earthquake

Alarm rings, no earthquake

No alarm, no earthquake

In general:

$$P(x \mid e_1,\ldots,e_m) = \alpha P(x, e_1,\ldots,e_m) = \sum_{y_1} \ldots \sum_{y_k} P(x, e_1,\ldots,e_m, y_1,\ldots,y_k)$$

and

$$P(\neg x \mid e_1,\ldots,e_m) = \alpha P(\neg x, e_1,\ldots,e_m) = \sum_{y_1} \ldots \sum_{y_k} P(\neg x, e_1,\ldots,e_m, y_1,\ldots,y_k)$$

where y1,…yk are hidden variables

I: $P(b \mid j,m)$

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

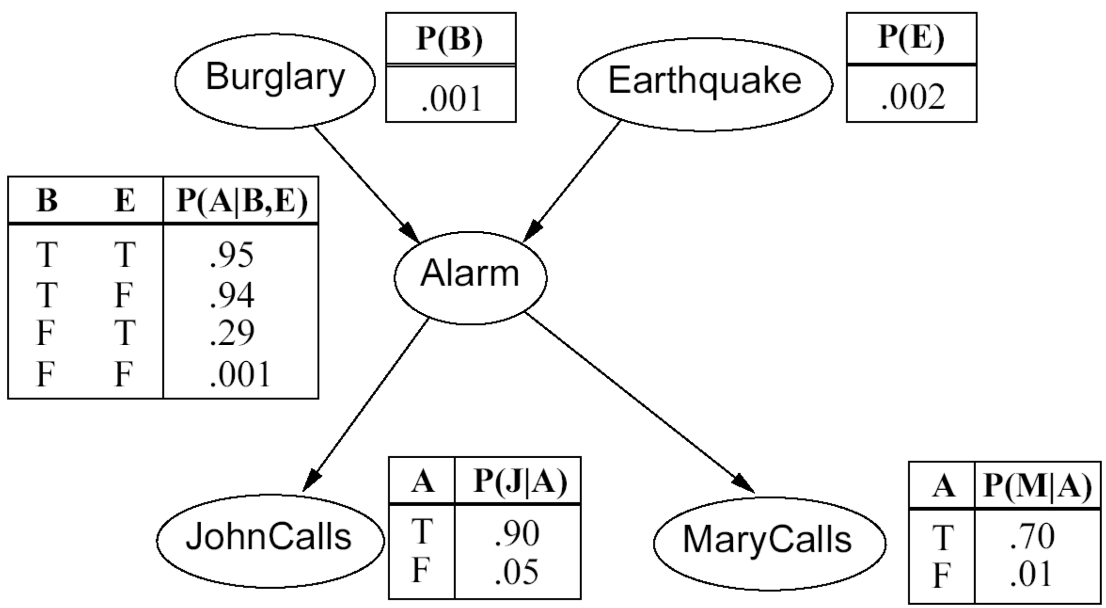
$= \alpha P(b) \sum_a P(j|a)P(m|a)(P(a|b,e)P(e) + P(a|b,\neg e)P(\neg e))$

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

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

$= \alpha * .001*(.9*.7*(.95*.002 + .94*.998) +.05*.01*(.05*.002 + .71*.998))$

$= \alpha * .00059$



I: $P(\neg b \mid j,m)$

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

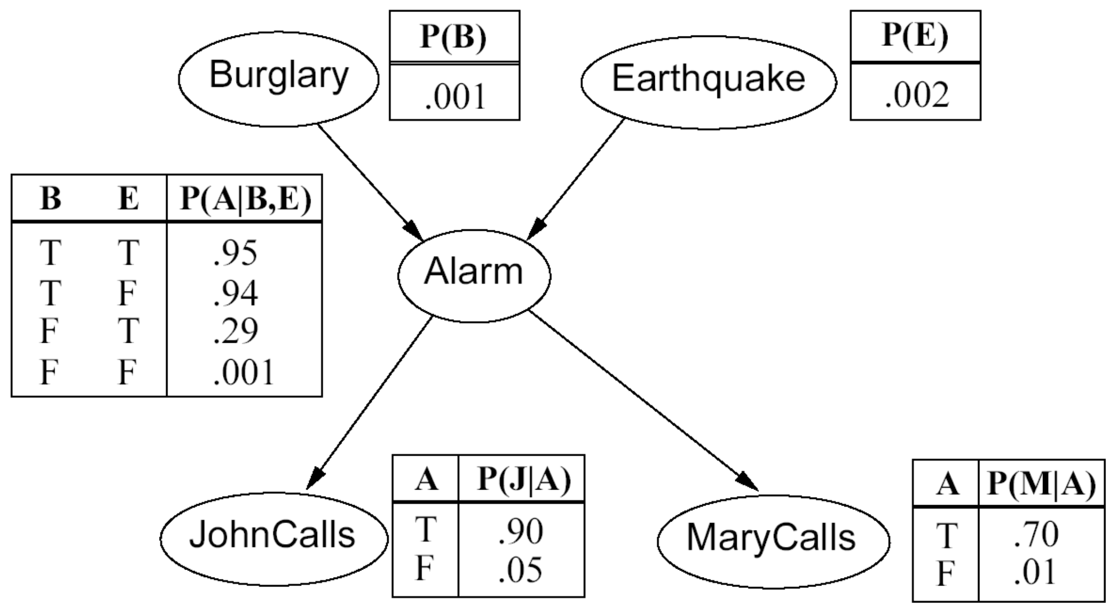
$= \alpha\, P(\neg b) \sum_a P(j|a)P(m|a)(P(a|\neg b,e)P(e) + P(a|\neg b,\neg e)P(\neg e))$

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

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

$= \alpha * .999*(.9*.7*(.29*.002 + .001*.998) +.05*.01*(.71*.002 + .999*.998)\ )$

$= \alpha * .0015$



P(B)
.001

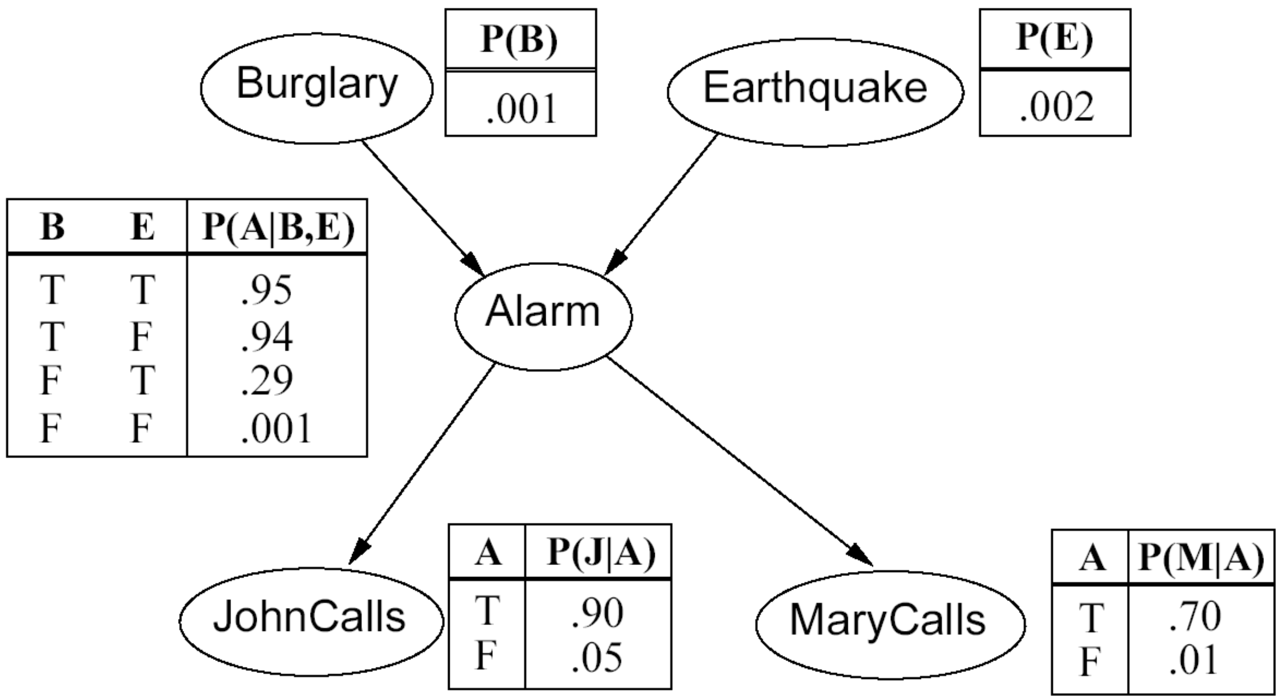
P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

I: Finally…



P(B)
.001

P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

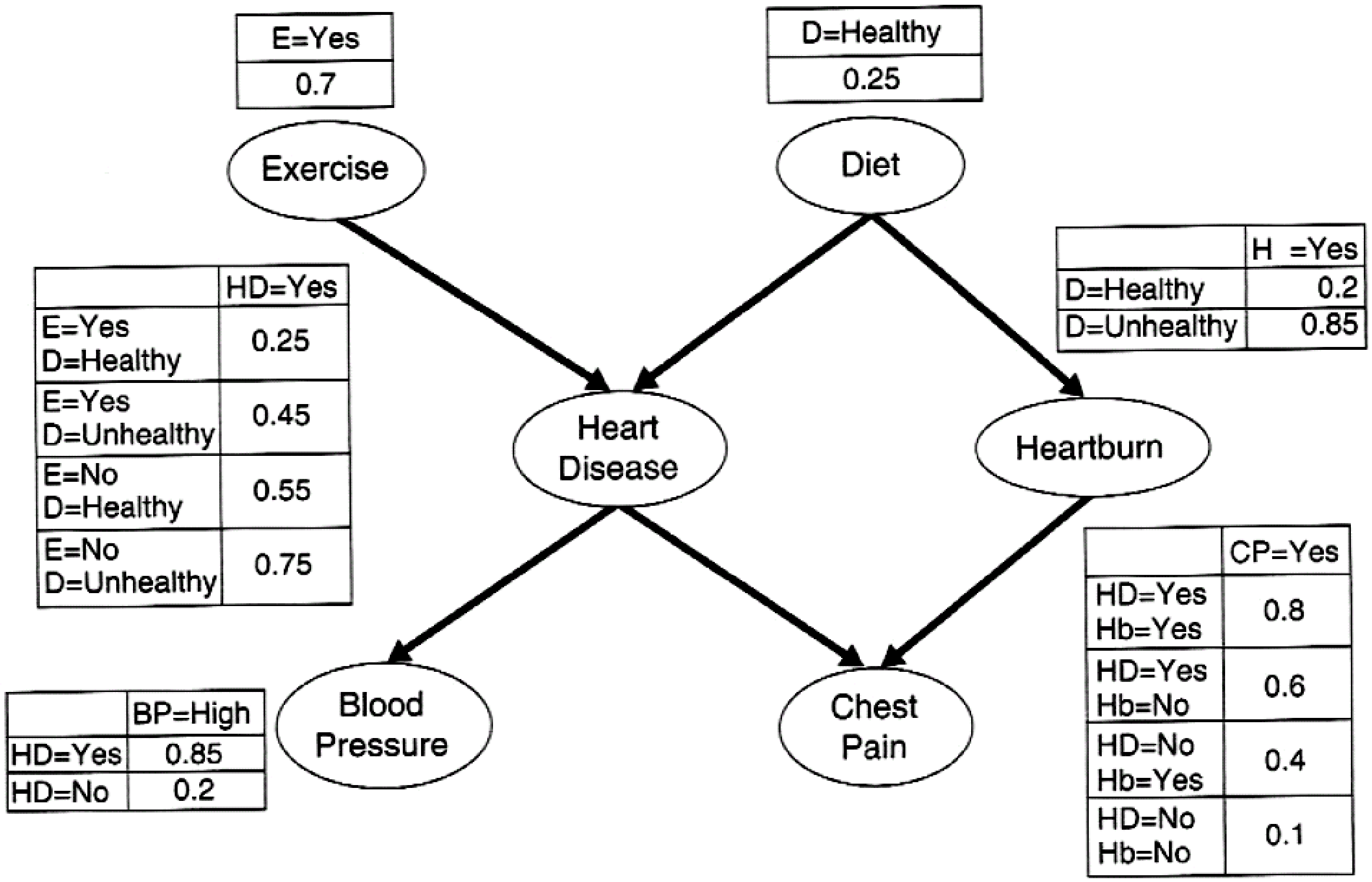
A	P(M|A)
T	.70
F	.01

$P(b \mid j,m) = \alpha\, P(b) \sum_a P(j|a)P(m|a)\sum_e P(a|b,e)P(e) = \ldots = \alpha * 0.00059$

$P(\neg b \mid j,m) = \alpha\, P(\neg b) \sum_a P(j|a)P(m|a)\sum_e P(a|\neg b,e)P(e) = \ldots = \alpha * 0.0015$

$\mathbf{P}(B \mid j,m) = \alpha \langle 0.00059, 0.0015\rangle =$ **<0.28, 0.72>**.

IIA: High Blood Pressure



E=Yes
0.7

D=Healthy
0.25

	HD=Yes
E=Yes D=Healthy	0.25
E=Yes D=Unhealthy	0.45
E=No D=Healthy	0.55
E=No D=Unhealthy	0.75

	H =Yes
D=Healthy	0.2
D=Unhealthy	0.85

	BP=High
HD=Yes	0.85
HD=No	0.2

	CP=Yes
HD=Yes Hb=Yes	0.8
HD=Yes Hb=No	0.6
HD=No Hb=Yes	0.4
HD=No Hb=No	0.1

Once the right topology has been found. the probability table associated with each node is determined from the data.

Estimating such probabilities similar to the approach used by naïve Bayes classifiers is done by counting rows where all the assignments of variables hold.

IIA: High Blood Pressure

- Suppose we get to know that the new patient has high blood pressure.
- What's the probability he has heart disease under this condition?

Heart disease: Yes

$$P(hd \mid bp) = \alpha \sum_e \sum_d \sum_h \sum_{cp} P(hd, E, D, H, CP, bp)$$

$$= \alpha \sum_e \sum_d \sum_h \sum_{cp} P(hd \mid E, D) P(H \mid D) P(CP \mid hd, H) P(bp \mid hd) P(E) P(D)$$

$$= \alpha P(bp \mid hd) \sum_e P(E) \sum_d P(hd \mid E, D) P(D) \sum_h P(H \mid D) \sum_{cp} P(CP \mid hd, H)$$

$$= \alpha P(bp \mid hd) \sum_e P(E) \sum_d P(hd \mid E, D) P(D)$$

$$= \alpha P(bp \mid hd) P(e) \left(P(hd \mid e, d) P(d) + P(hd \mid e, \neg d) P(\neg d) \right)$$

=1

$$+ \alpha P(bp \mid hd) P(\neg e) \left(P(hd \mid \neg e, d) P(d) + P(hd \mid \neg e, \neg d) P(\neg d) \right)$$

$$= \alpha * 0.85 * 0.7 * \left(0.25 * 0.25 + 0.45 * 0.75 \right)$$

$$+ \alpha * 0.85 * 0.3 * (0.55 * 0.25 + 0.75 * 0.75)$$

$$= \alpha * 0.4165$$

IIA: High Blood Pressure

Heart disease: No

$$P(\neg hd|bp) = \alpha \sum_{e}\sum_{d}\sum_{h}\sum_{cp} P(\neg hd, E, D, H, CP, bp)$$

$$= \alpha \sum_{e}\sum_{d}\sum_{h}\sum_{cp} P(\neg hd \mid e, d) P(h \mid d) P(CP \mid \neg hd, H) P(bp \mid \neg hd) P(E) P(D)$$

$$= \alpha P(bp|\neg hd) \sum_{e} P(E) \sum_{d} P(\neg hd|E, D) P(D) \sum_{h} P(H|D) \sum_{cp} P(CP|\neg hd, H)$$

$$= \alpha P(bp \mid \neg hd) \sum_{e} P(E) \sum_{d} P(\neg hd \mid E, D) P(D)$$

$$= \alpha P(bp \mid \neg hd) P(e) \left(P(\neg hd \mid e, d) P(d) + P(\neg hd \mid e, \neg d) P(\neg d) \right)$$
$$+ \alpha P(bp \mid \neg hd) P(\neg e) \left(P(\neg hd \mid \neg e, d) P(d) + P(\neg hd \mid \neg e, \neg d) P(\neg d) \right)$$

$$= \alpha * 0.2 * 0.7 * (0.75 * 0.25 + 0.55 * 0.75)$$
$$+ \alpha * 0.2 * 0.3 * (0.45 * 0.25 + 0.25 * 0.75)$$
$$= \alpha * 0.102$$

IIA: High Blood Pressure (α)

$P(\neg hd|bp) = \alpha * 0.102$

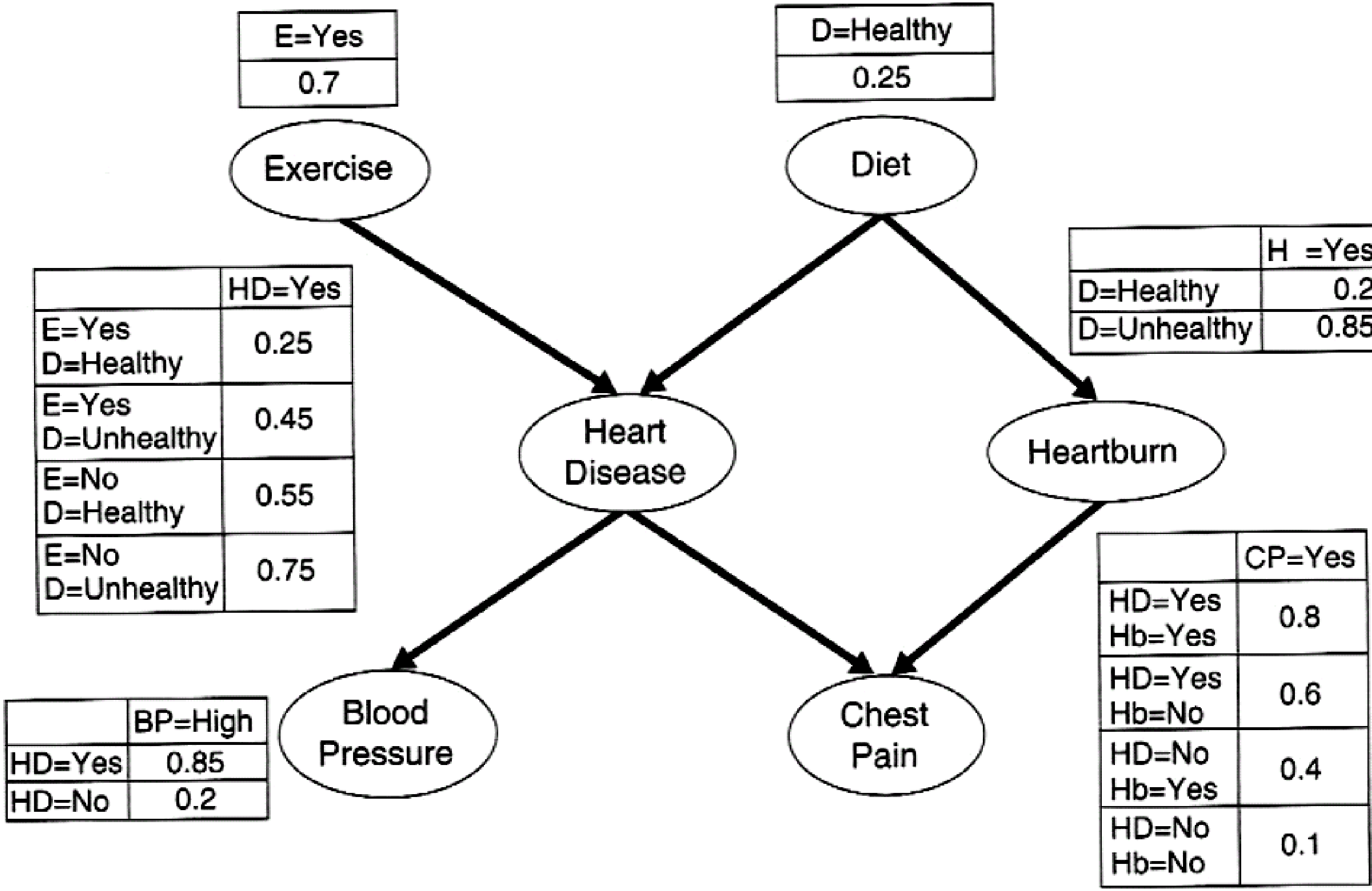
$P(hd|bp) = \alpha * 0.4165$

$$\alpha = \frac{1}{0.4165 + 0.1020} = \frac{1}{0.5185}$$

$$P(hd \mid bp) = \alpha * 0.4165 = 0.8033$$

$$P(\neg hd \mid bp) = \alpha * 0.102 = 0.1967$$

IIB: High Blood Pressure, Healthy Diet, Regular Exercise



E=Yes
0.7

D=Healthy
0.25

	HD=Yes
E=Yes D=Healthy	0.25
E=Yes D=Unhealthy	0.45
E=No D=Healthy	0.55
E=No D=Unhealthy	0.75

	H =Yes
D=Healthy	0.2
D=Unhealthy	0.85

	BP=High
HD=Yes	0.85
HD=No	0.2

	CP=Yes
HD=Yes Hb=Yes	0.8
HD=Yes Hb=No	0.6
HD=No Hb=Yes	0.4
HD=No Hb=No	0.1

IIB: Probability of heart disease

$$P(hd \mid bp, d, e) = \alpha \sum_h \sum_{cp} P(hd, e, d, H, CP, bp)$$

$$= \alpha \sum_h \sum_{cp} P(hd \mid e, d) P(H \mid d) P(cp \mid hd, H) P(bp \mid hd) P(e) P(d)$$

$$= \alpha P(bp \mid hd) P(e) P(hd \mid e, d) P(d) \sum_h P(H \mid d) \sum_{cp} P(CP \mid hd, h)$$

$$= \alpha P(bp \mid hd) P(e) P(hd \mid e, d) P(d)$$

$$= \alpha * 0.85 * 0.7 * 0.25 * 0.25 = \alpha * 0.03719$$

IIB: Probability of not heart disease

$$P(\neg hd \mid bp, d, e) = \alpha \sum_h \sum_{cp} P(\neg hd, e, d, H, CP, bp)$$

$$= \alpha \sum_h \sum_{cp} P(\neg hd \mid e, d) P(h \mid d) P(CP \mid \neg hd, H) P(bp \mid \neg hd) P(e) P(d)$$

$$= \alpha P(bp \mid \neg hd) P(e) P(\neg hd \mid e, d) P(d) \sum_h P(H \mid d) \sum_{cp} P(CP \mid \neg hd, h)$$

$$= \alpha P(bp \mid \neg hd) P(e) P(\neg hd \mid e, d) P(d)$$

$$= \alpha * 0.2 * 0.7 * 0.75 * 0.25 = \alpha * 0.02625$$

II. High Blood Pressure, Healthy Diet, and Regular Exercise

$$P(hd \mid bp) = \alpha * 0.4165 = 0.8033$$
$$P(\neg hd \mid bp) = \alpha * 0.102 = 0.1967$$

$$\alpha = \frac{1}{0.03719 + 0.02625} = \frac{1}{0.06344}$$

$$P(hd \mid bp, d, e) = \alpha * 0.03719 = 0.5862$$

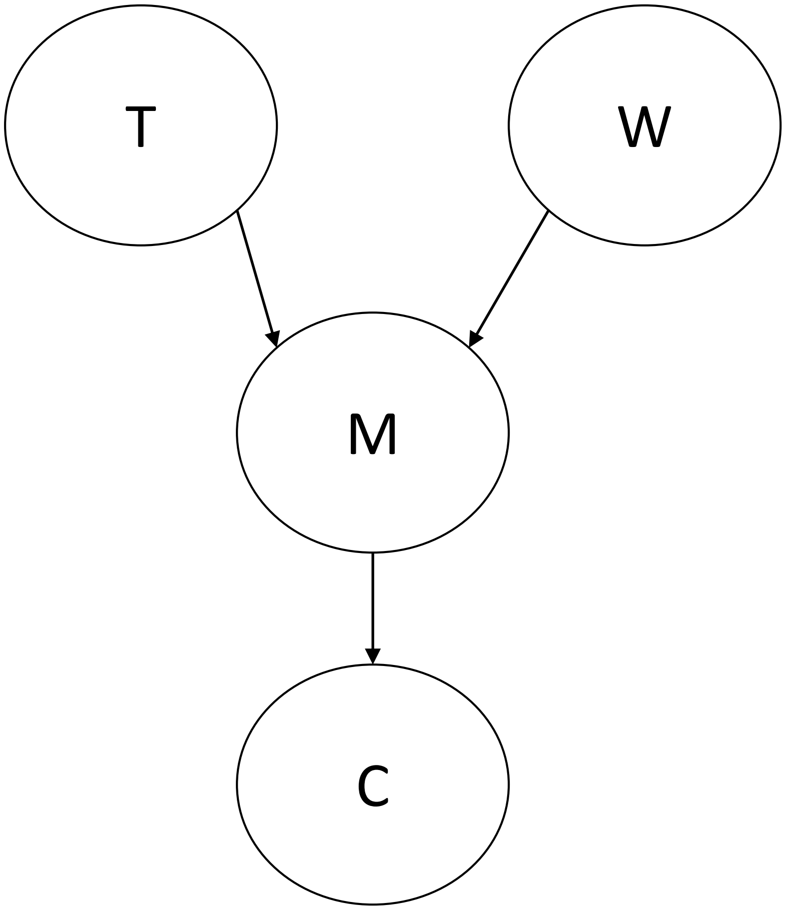
$$P(\neg hd \mid bp, d, e) = \alpha * 0.02625 = 0.4138$$

The model therefore suggests that eating healthily and exercising regularly may reduce a person's risk of getting heart disease, even if he has high blood pressure

III: Solving the mystery

- One early morning the maid was dusting the window when she saw something horrific. Right outside the window lay dead Mr. Boddy. She called the police and a detective was assigned to the case
- The detective, a former computer scientist, always tried to make his job as easy as possible.
- After a brief examination, he determined that Mr. Boddy has been hit over the head with a dull instrument, probably made of metal. The detective found two candidate weapons that matched the crime scene: an extension of Vacuum cleaner (V) used by the maid and a candle Stick (S) used by the butler.
- He took a brief statement from both the Made (M) and the Butler (B), the only two individuals who could have possibly committed the murder.
- Then he went to his office and decided to create BBN to determine whether the murderer is likely to confess

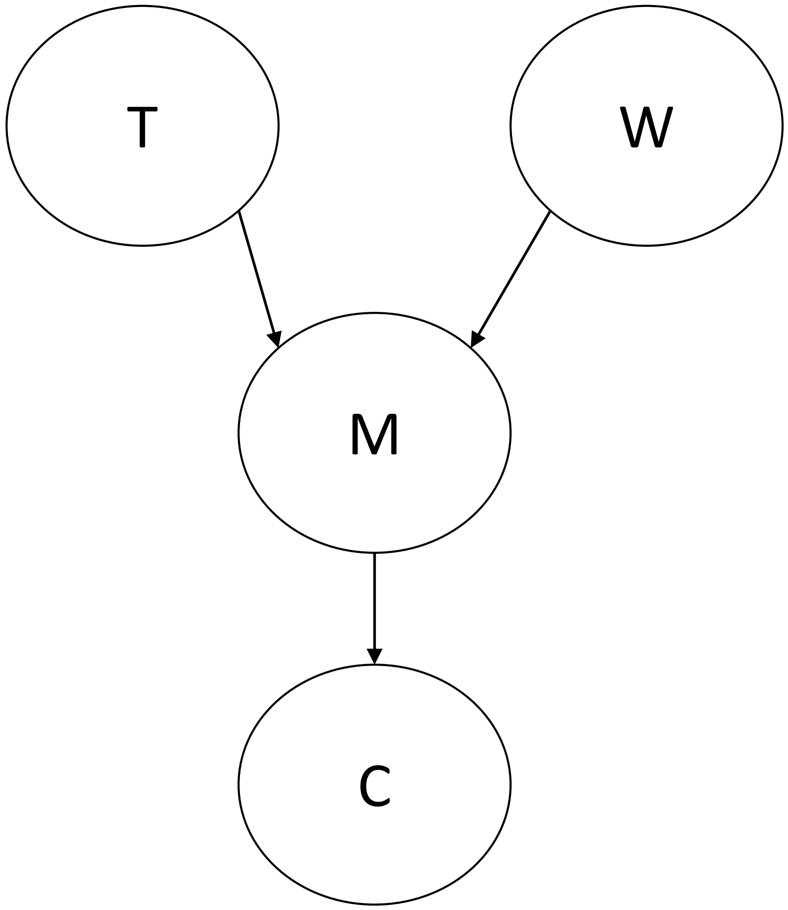
III: Network topology



- T – time of day when the murder was committed: evening (e) or night (n)
- W – crime weapon: vacuum (v) or stick (s)
- M- murderer: maid (m) or butler (b)
- C – will confess: yes or no

III: CPT for Time

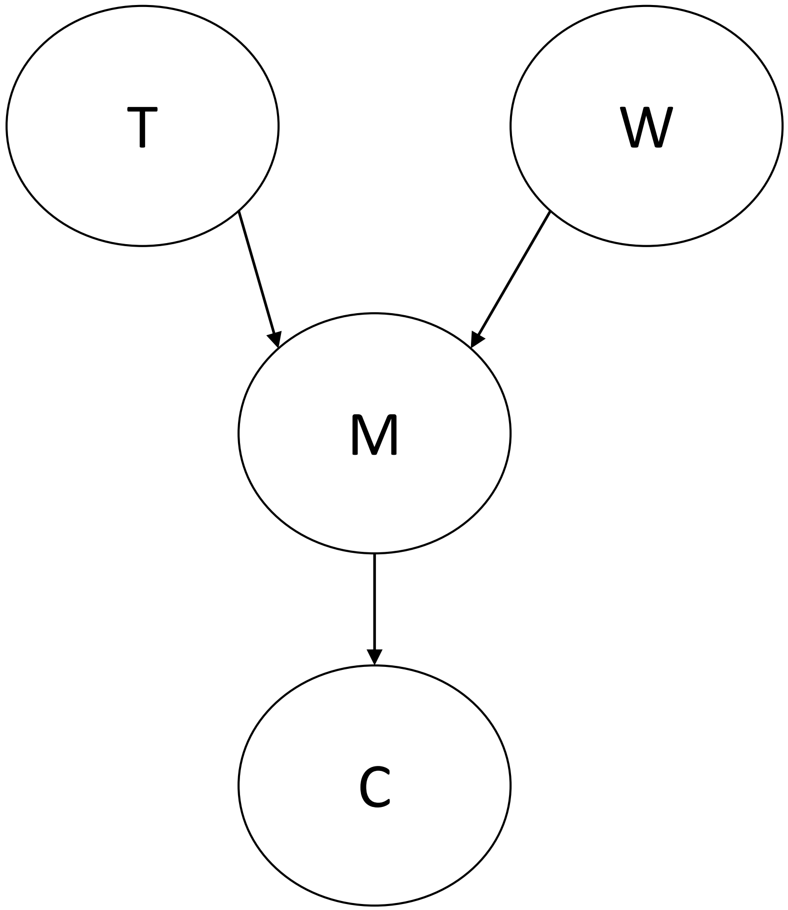
e	n
.05	.95



- T- time: the murder was committed in the evening (e) or at night (n), but much more likely at night

III: CPT for Weapon

e	n
.05	.95

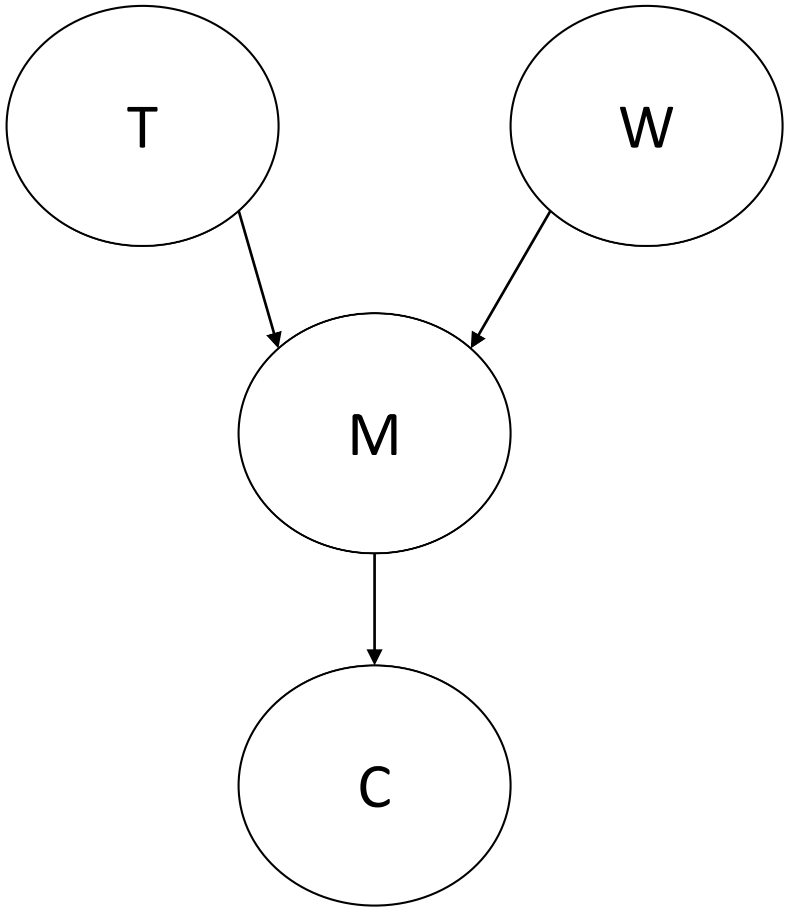


v	s
.80	.20

- W- weapon: the murder weapon was most likely an extension to vacuum cleaner than the candle stick

III: CPT Murderer

e	n
.05	.95

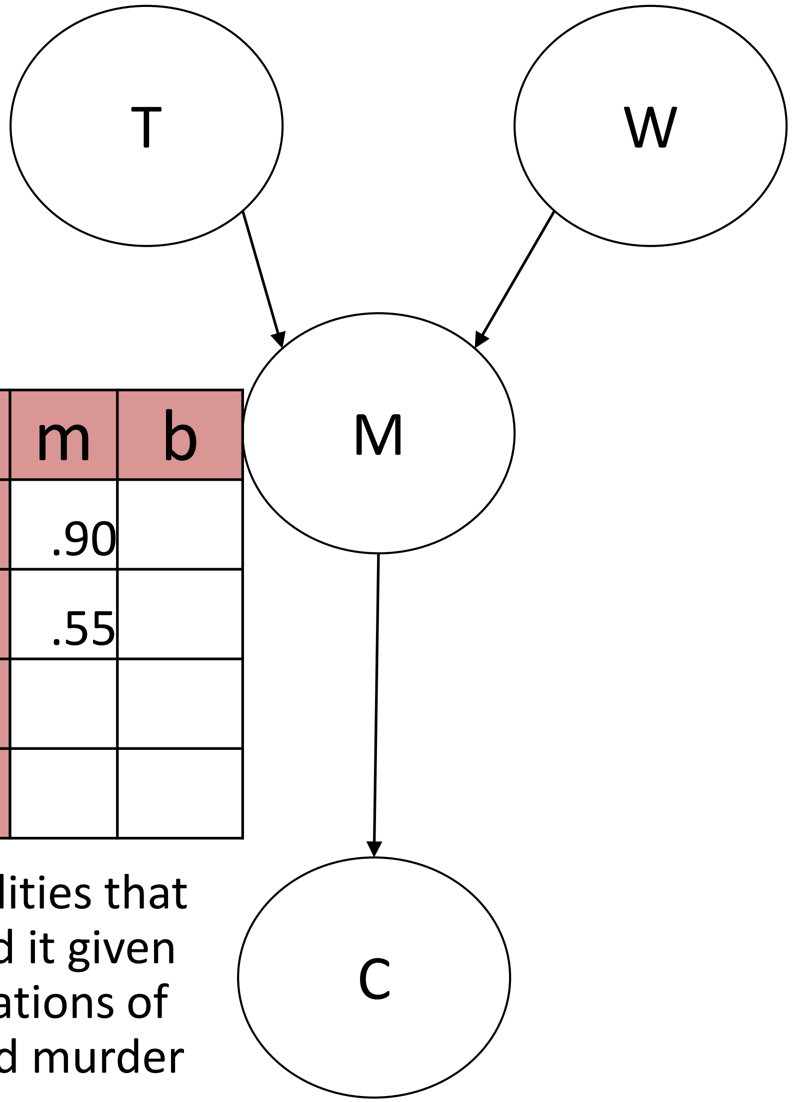


v	s
.80	.20

- M- murderer: Although it is possible for one employee to use the tool of another, it is very unlikely

III: CPT Murderer

e	n
.05	.95



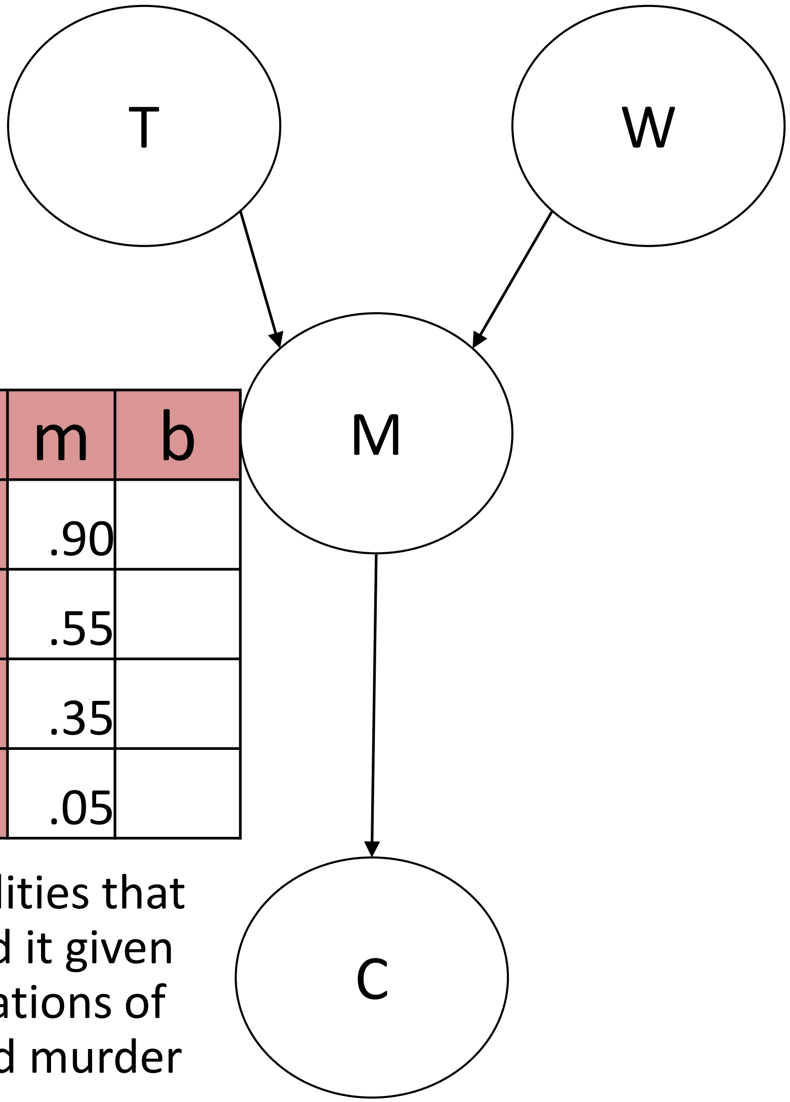
v	s
.80	.20

T	W	m	b
e	v	.90	
e	s	.55	
n	v		
n	s		

- Probabilities that maid did it given combinations of time and murder weapon

III: CPT Murderer

e	n
.05	.95



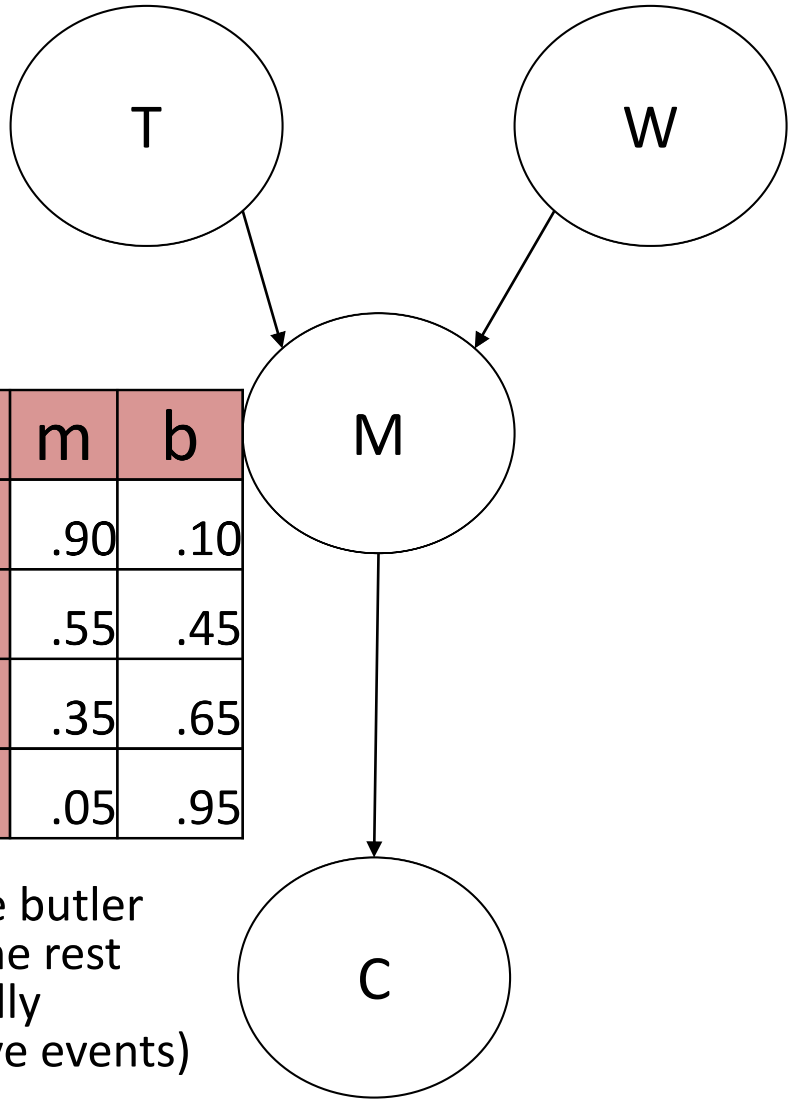
v	s
.80	.20

T	W	m	b
e	v	.90	
e	s	.55	
n	v	.35	
n	s	.05	

- Probabilities that maid did it given combinations of time and murder weapon

III: CPT Murderer

e	n
.05	.95



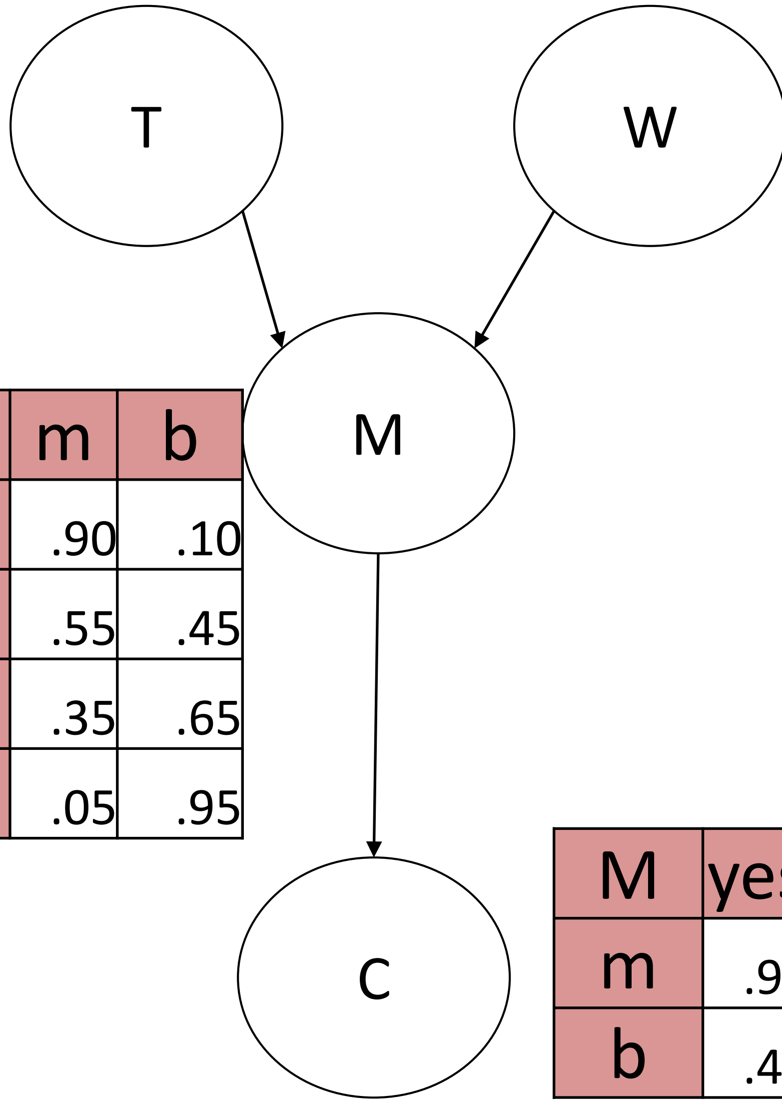
v	s
.80	.20

T	W	m	b
e	v	.90	.10
e	s	.55	.45
n	v	.35	.65
n	s	.05	.95

- And the butler takes the rest (mutually exclusive events)

III: CPT Confession

e	n
.05	.95



v	s
.80	.20

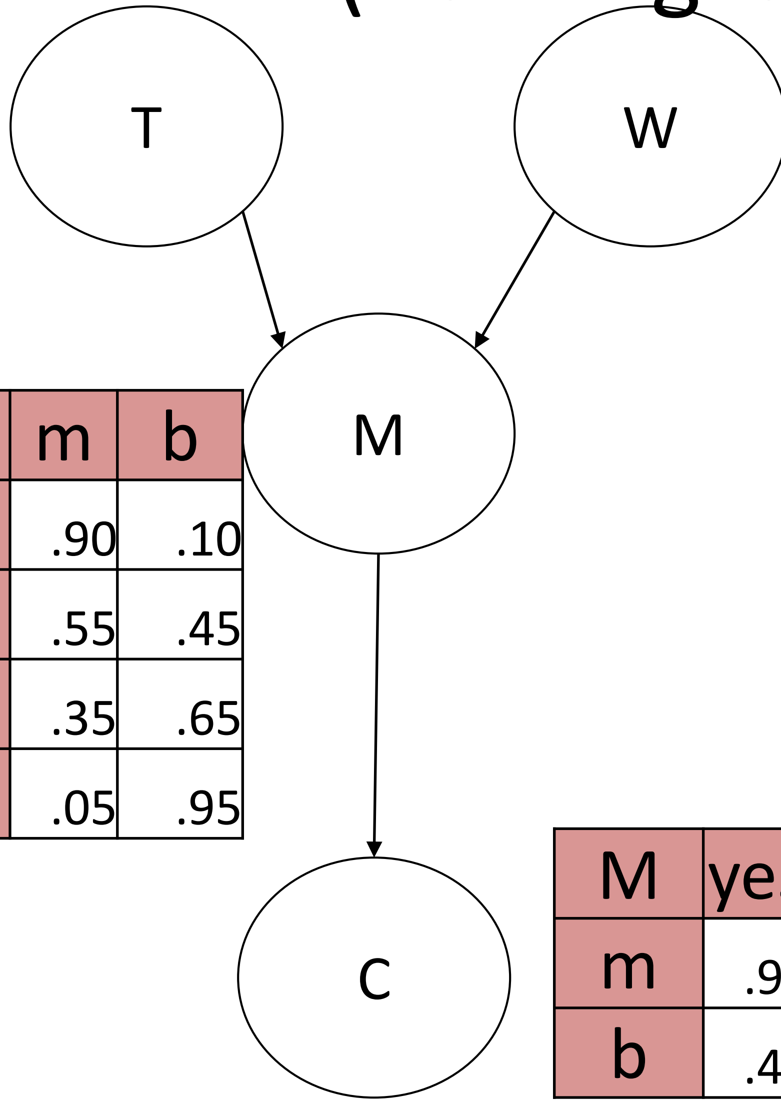
T	W	m	b
e	v	.90	.10
e	s	.55	.45
n	v	.35	.65
n	s	.05	.95

M	yes	no
m	.90	.10
b	.40	.60

- The maid has a very strong conscience and she will eventually confess if she committed the murder. The butler is quite opposite

III: The probability of confession (nothing is given)

e	n
.05	.95



v	s
.80	.20

T	W	m	b
e	v	.90	.10
e	s	.55	.45
n	v	.35	.65
n	s	.05	.95

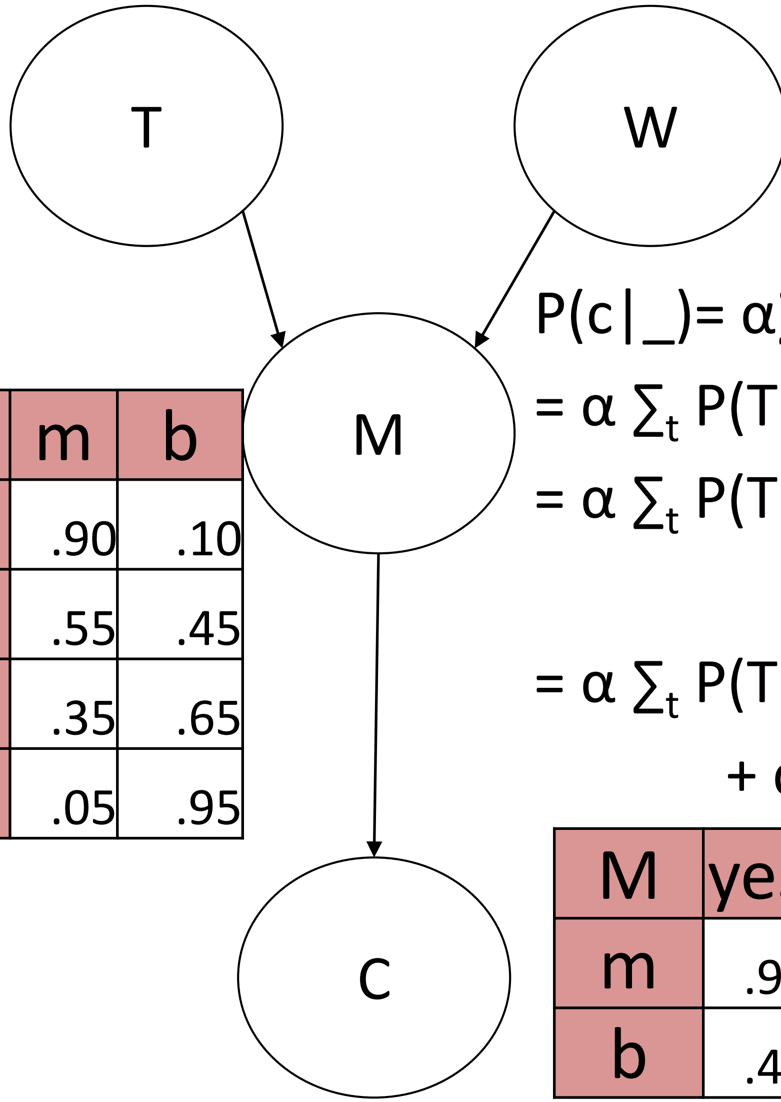
M	yes	no
m	.90	.10
b	.40	.60

- The detective can evaluate the probability that the murderer will confess without having any real evidence

III: The probability of confession

e	n
.05	.95

v	s
.80	.20



T	W	m	b
e	v	.90	.10
e	s	.55	.45
n	v	.35	.65
n	s	.05	.95

$P(c|\_) = \alpha \sum_t \sum_w \sum_m P(T)P(W)P(M|T,W)P(c|M)$

$= \alpha \sum_t P(T) \sum_w P(W) \sum_m P(M|T,W)P(c|M)$

$= \alpha \sum_t P(T) \sum_w P(W) [P(m|T,W)P(c|m)$

$+ P(b|T,W)P(c|b)]$

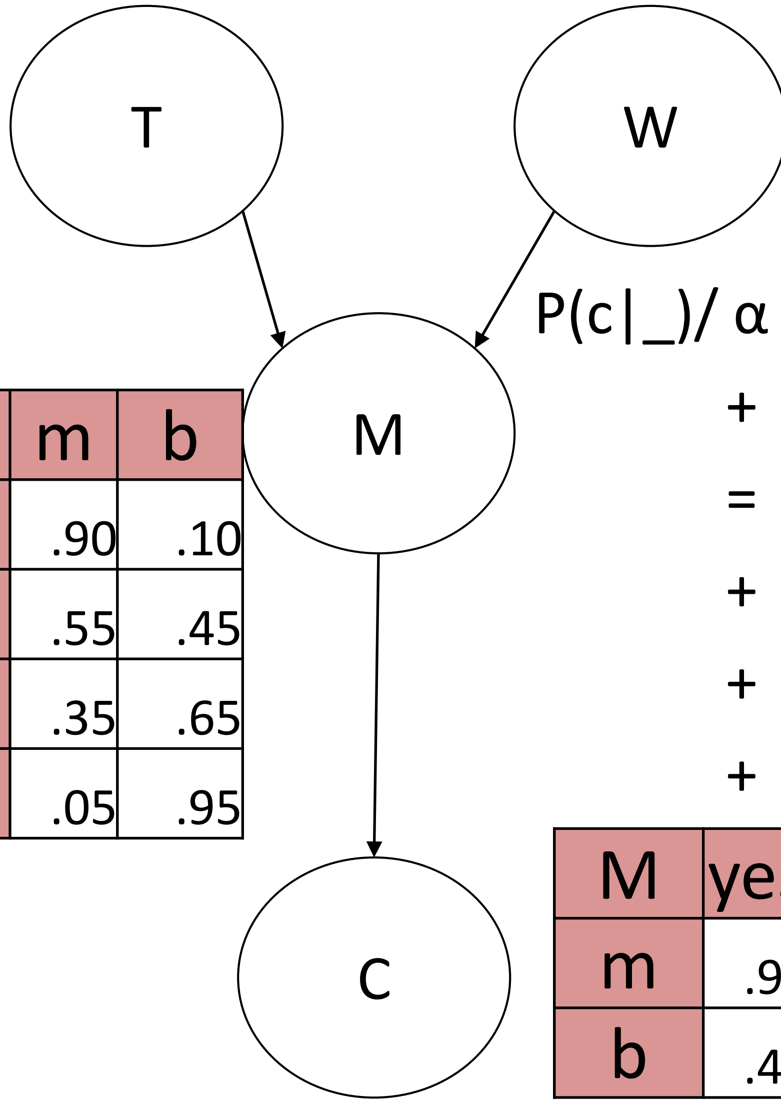
$= \alpha \sum_t P(T) \sum_w P(W) P(m|T,W) *0.90$

$+ \alpha \sum_t P(T) \sum_w P(W) P(b|T,W) *0.40$

M	yes	no
m	.90	.10
b	.40	.60

III: The probability of confession

e	n
.05	.95



v	s
.80	.20

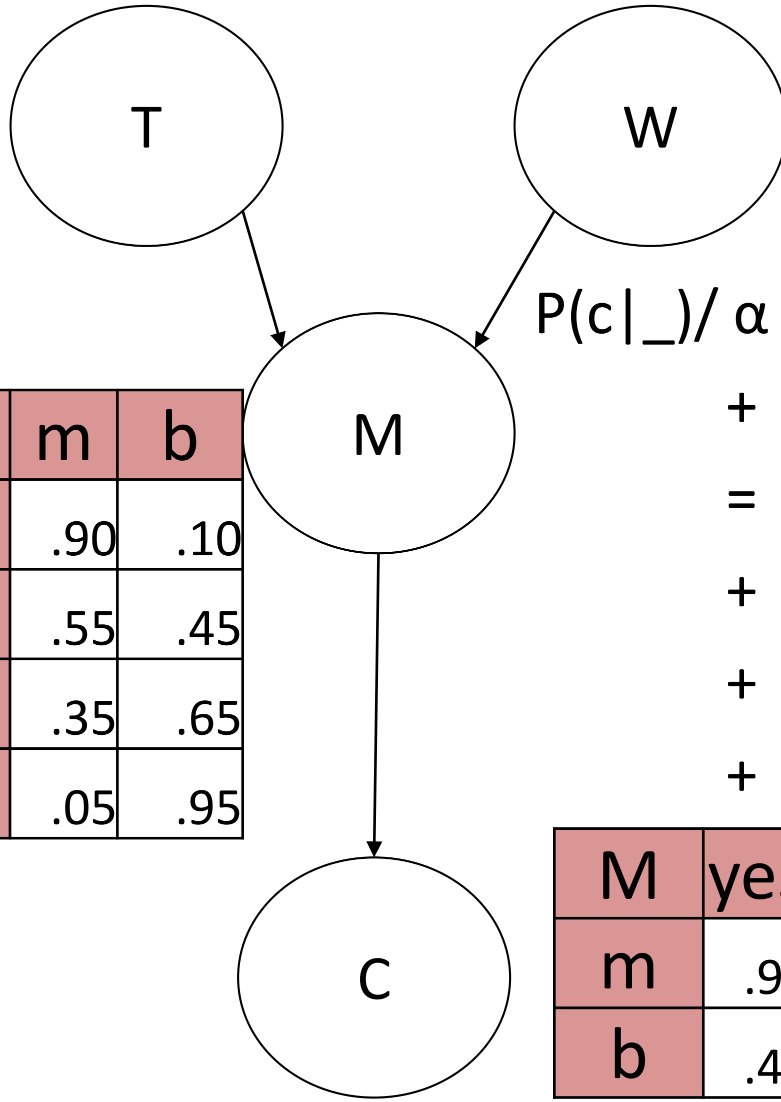
$$P(c|\_)/\alpha = \sum_t P(T) \sum_w P(W)\, P(m|T,W) *0.90$$
$$+ \sum_t P(T) \sum_w P(W)\, P(b|T,W) *0.40$$
$$= \sum_t P(T)\, P(v)\, P(m|T,v) *0.90$$
$$+ \sum_t P(T)\, P(s)\, P(m|T,s) *0.90$$
$$+ \sum_t P(T)\, P(v)\, P(b|T,v) *0.40$$
$$+ \sum_t P(T)\, P(s)\, P(b|T,s) *0.40$$

T	W	m	b
e	v	.90	.10
e	s	.55	.45
n	v	.35	.65
n	s	.05	.95

M	yes	no
m	.90	.10
b	.40	.60

III: The probability of confession

e	n
.05	.95



v	s
.80	.20

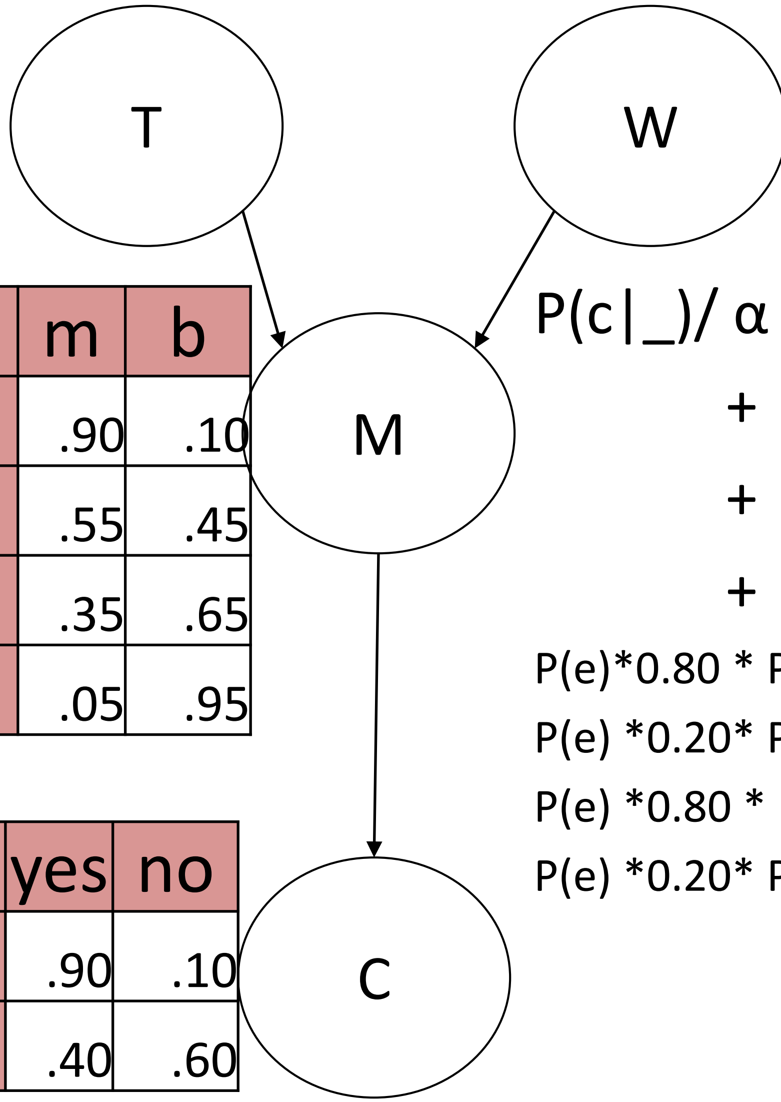
T	W	m	b
e	v	.90	.10
e	s	.55	.45
n	v	.35	.65
n	s	.05	.95

$$
\begin{aligned}
P(c|\_)/\alpha &= \textstyle\sum_t P(T) \sum_w P(W)\, P(m|T,W) *0.90 \\
&+ \textstyle\sum_t P(T) \sum_w P(W)\, P(b|T,W) *0.40 \\
&= \textstyle\sum_t P(T) *0.80* P(m|T,v) *0.90 \\
&+ \textstyle\sum_t P(T) *0.20* P(m|T,s) *0.90 \\
&+ \textstyle\sum_t P(T) *0.80* P(b|T,v) *0.40 \\
&+ \textstyle\sum_t P(T) *0.20* P(b|T,s) *0.40
\end{aligned}
$$

M	yes	no
m	.90	.10
b	.40	.60

III: The probability of confession

e	n
.05	.95



v	s
.80	.20

T	W	m	b
e	v	.90	.10
e	s	.55	.45
n	v	.35	.65
n	s	.05	.95

M	yes	no
m	.90	.10
b	.40	.60

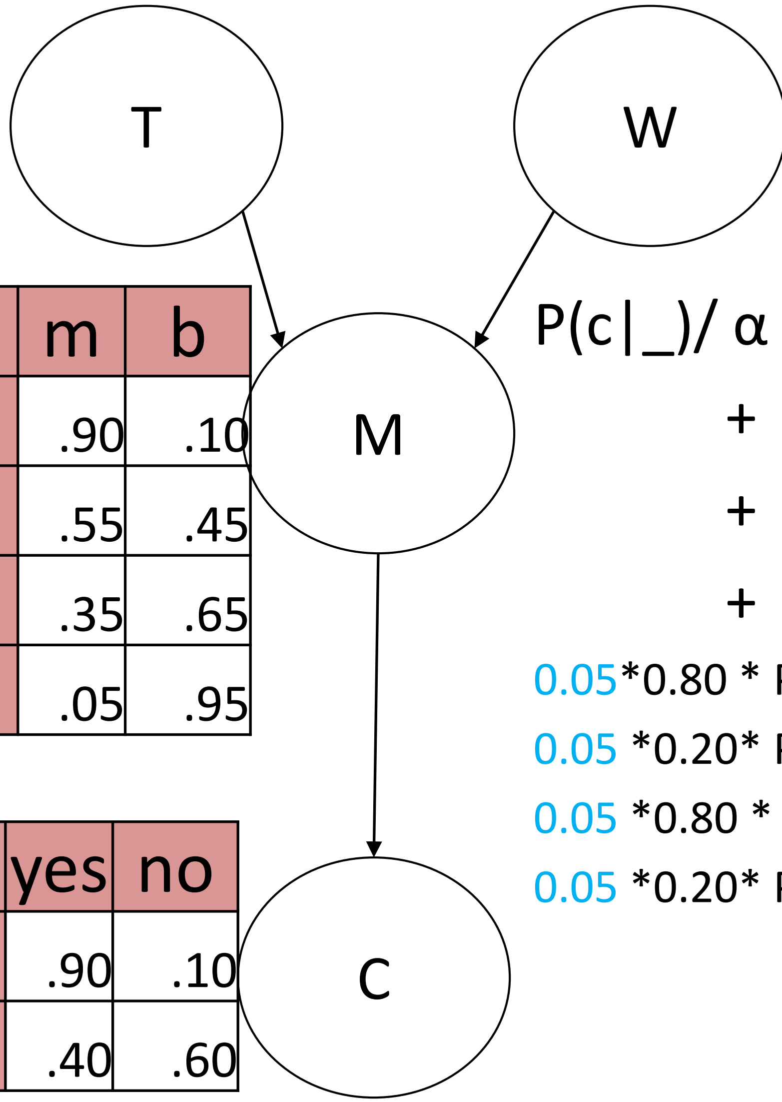
$$P(c|\_)/\alpha = \sum_t P(T) *0.80 * P(m|T,v) *0.90$$
$$+ \sum_t P(T) *0.20* P(m|T,s) *0.90$$
$$+ \sum_t P(T) *0.80 * P(b|T,v) *0.40$$
$$+ \sum_t P(T) *0.20* P(b|T,s) *0.40 =$$

$$P(e)*0.80 * P(m|e,v) * 0.90+P(n)*0.80 * P(m|n,v)*0.90 +$$
$$P(e) *0.20* P(m|e,s) *0.90+ P(n) *0.20* P(m|n,s) *0.90 +$$
$$P(e) *0.80 * P(b|e,v) *0.40+ P(n) *0.80 * P(b|n,v) *0.40 +$$
$$P(e) *0.20* P(b|e,s) *0.40+ P(n) *0.20* P(b|n,s) *0.40$$

III:The probability of confession

e	n
.05	.95

v	s
.80	.20



T	W	m	b
e	v	.90	.10
e	s	.55	.45
n	v	.35	.65
n	s	.05	.95

M	yes	no
m	.90	.10
b	.40	.60

$$P(c|\_)/\alpha = \sum_t P(T) *0.80 * P(m|T,v) *0.90$$
$$+ \sum_t P(T) *0.20* P(m|T,s) *0.90$$
$$+ \sum_t P(T) *0.80 * P(b|T,v) *0.40$$
$$+ \sum_t P(T) *0.20* P(b|T,s) *0.40 =$$

$0.05*0.80 * P(m|e,v) * 0.90+0.95*0.80 * P(m|n,v)*0.90 +$

$0.05 *0.20* P(m|e,s) *0.90+ 0.95 *0.20* P(m|n,s) *0.90 +$

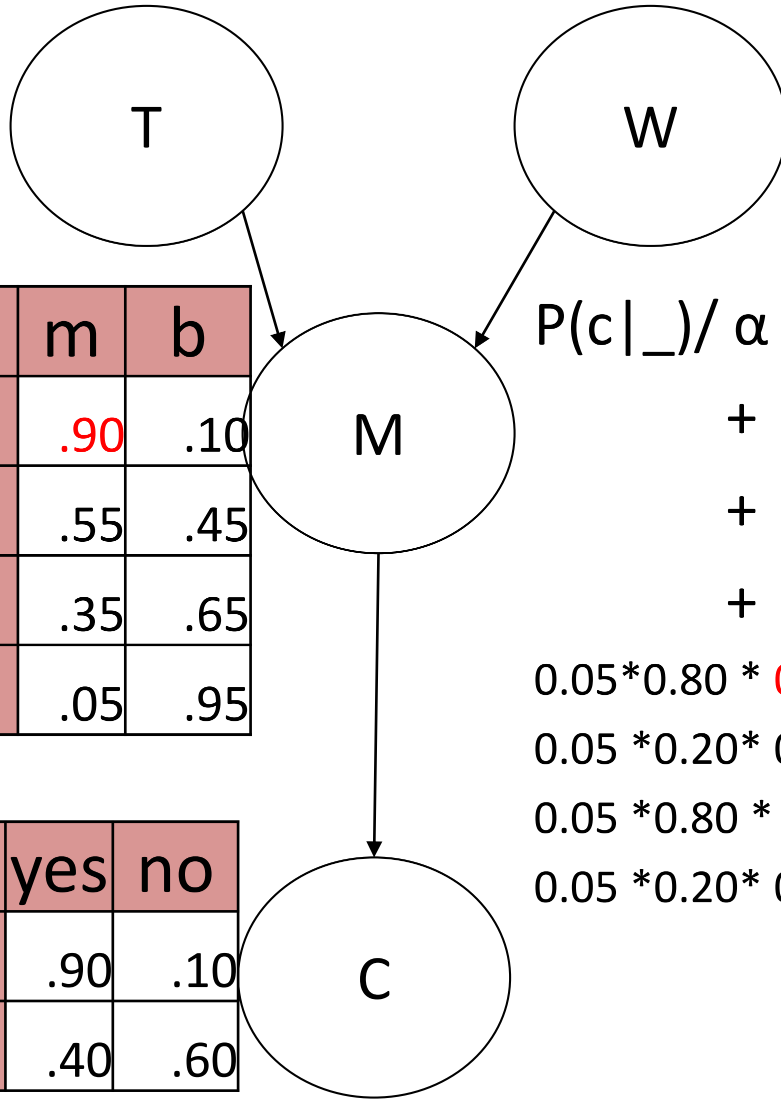
$0.05 *0.80 * P(b|e,v) *0.40+ 0.95*0.80 * P(b|n,v) *0.40 +$

$0.05 *0.20* P(b|e,s) *0.40+ 0.95 *0.20* P(b|n,s) *0.40$

III: The probability of confession

e	n
.05	.95

v	s
.80	.20



T	W	m	b
e	v	.90	.10
e	s	.55	.45
n	v	.35	.65
n	s	.05	.95

M	yes	no
m	.90	.10
b	.40	.60

$P(c|\_)/\alpha = \sum_t P(T) *0.80 * P(m|T,v) *0.90$

$+ \sum_t P(T) *0.20* P(m|T,s) *0.90$

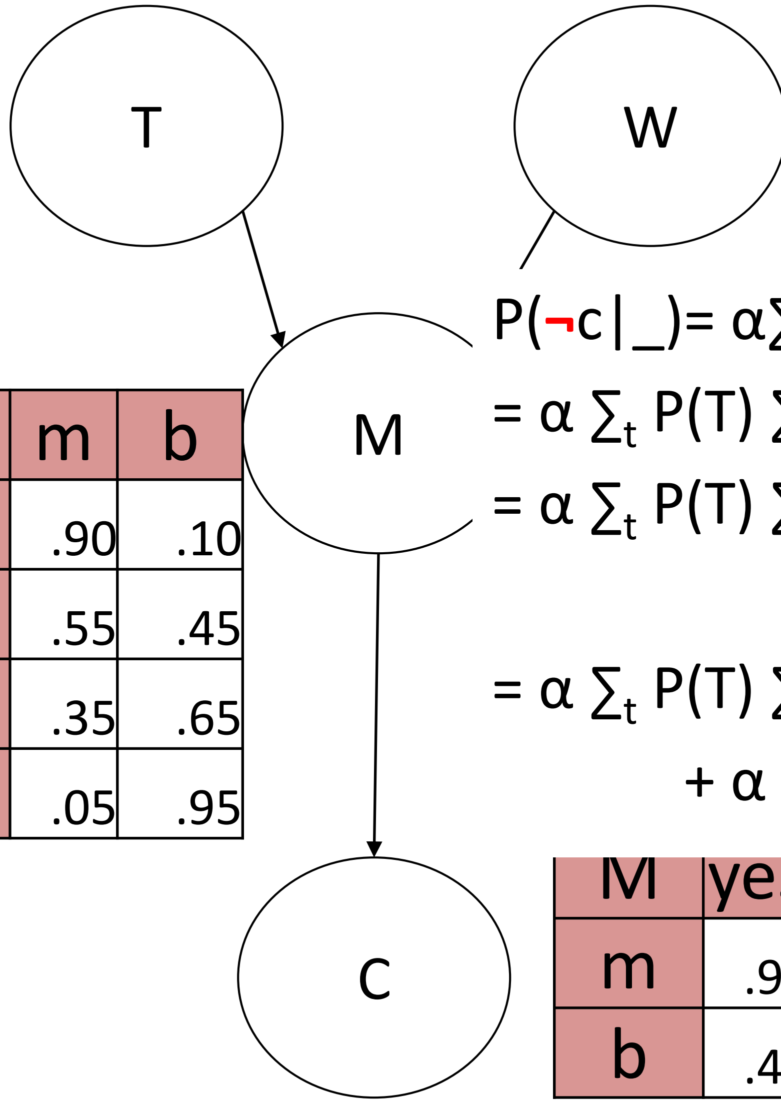
$+ \sum_t P(T) *0.80 * P(b|T,v) *0.40$

$+ \sum_t P(T) *0.20* P(b|T,s) *0.40 =$

0.05*0.80 * 0.90 * 0.90+0.95*0.80 * 0.35*0.90 +
0.05 *0.20* 0.55 *0.90+ 0.95 *0.20* 0.05 *0.90 +
0.05 *0.80 * 0.10 *0.40+ 0.95*0.80 *0.65 *0.40 +
0.05 *0.20* 0.45 *0.40+ 0.95 *0.20* 0.95 *0.40 = 0.56

III: The probability of non-confession

e	n
.05	.95



v	s
.80	.20

T	W	m	b
e	v	.90	.10
e	s	.55	.45
n	v	.35	.65
n	s	.05	.95

$P(\neg c|\_) = \alpha \sum_t \sum_w \sum_m P(T)P(W)P(M|T,W)P(\neg c|M)$

$= \alpha \sum_t P(T) \sum_w P(W) \sum_m P(M|T,W)P(\neg c|M)$

$= \alpha \sum_t P(T) \sum_w P(W) [P(m|T,W)P(\neg c|m)$

$+ \sum_w P(b|T,W)P(\neg c|b)]$

$= \alpha \sum_t P(T) \sum_w P(W)\ P(m|T,W) * 0.10$

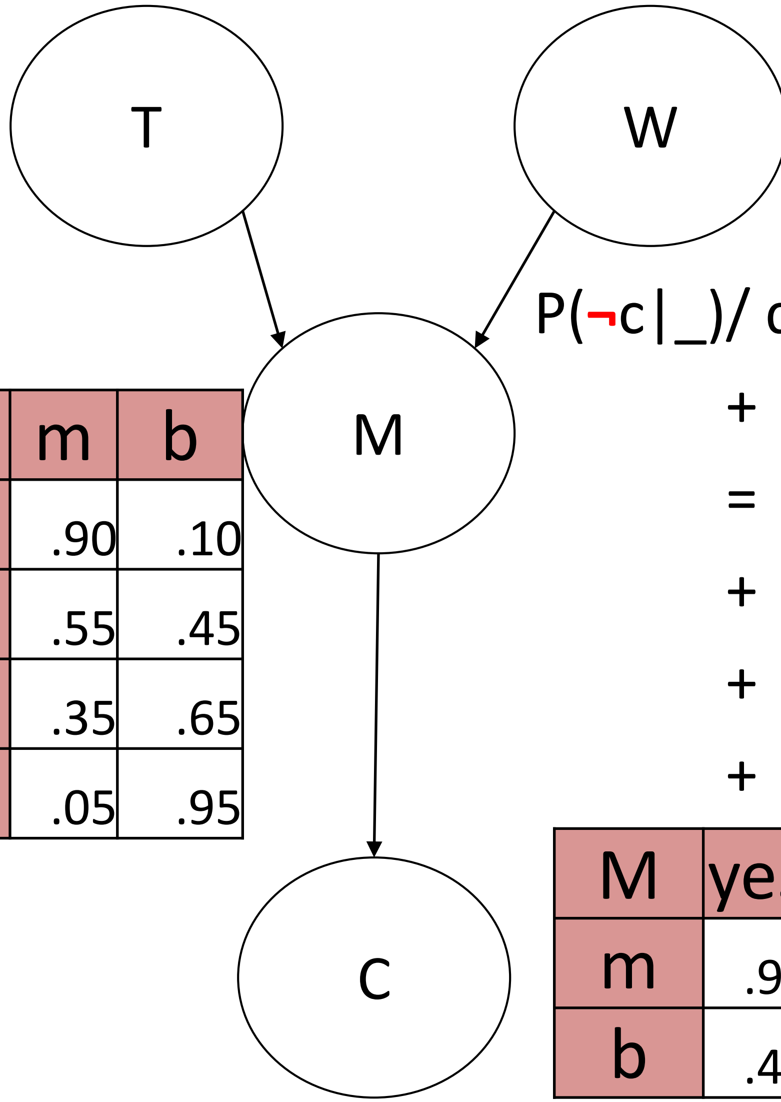
$+ \alpha \sum_t P(T) \sum_w P(W)\ P(b|T,W) * 0.60$

M	yes	no
m	.90	.10
b	.40	.60

III: The probability of non-confession

e	n
.05	.95

v	s
.80	.20



T	W	m	b
e	v	.90	.10
e	s	.55	.45
n	v	.35	.65
n	s	.05	.95

$P(\neg c|\_)/\alpha = \sum_t P(T) \sum_w P(W)\ P(m|T,W) * 0.90$

$+ \sum_t P(T) \sum_w P(W)\ P(b|T,W) * 0.40$

$= \sum_t P(T)\ P(v)\ P(m|T,v) * 0.10$

$+ \sum_t P(T)\ P(s)\ P(m|T,s) * 0.10$

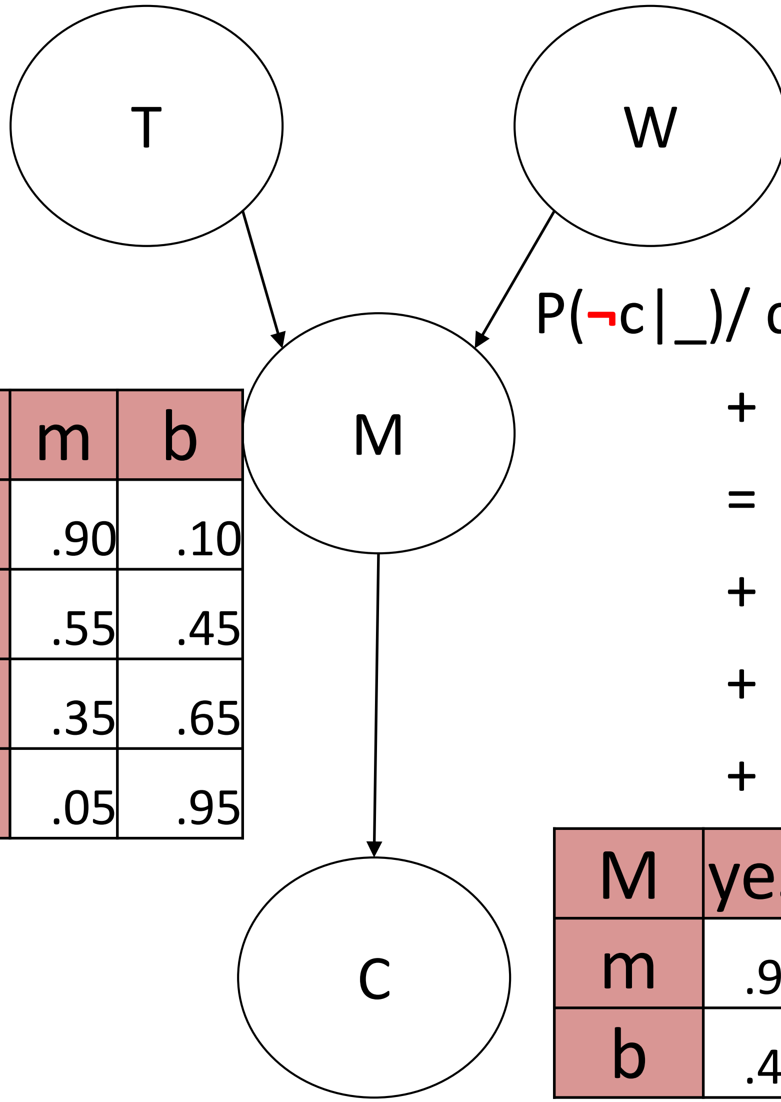
$+ \sum_t P(T)\ P(v)\ P(b|T,v) * 0.60$

$+ \sum_t P(T)\ P(s)\ P(b|T,s) * 0.60$

M	yes	no
m	.90	.10
b	.40	.60

III: The probability of non-confession

e	n
.05	.95



v	s
.80	.20

T	W	m	b
e	v	.90	.10
e	s	.55	.45
n	v	.35	.65
n	s	.05	.95

$$P(\neg c|\_)/\ \alpha = \sum_t P(T) \sum_w P(W)\ P(m|T,W) *0.90$$

$$+ \sum_t P(T) \sum_w P(W)\ P(b|T,W) *0.40$$

$$= \sum_t P(T) *0.80 * P(m|T,v) *0.10$$

$$+ \sum_t P(T) *0.20* P(m|T,s) *0.10$$

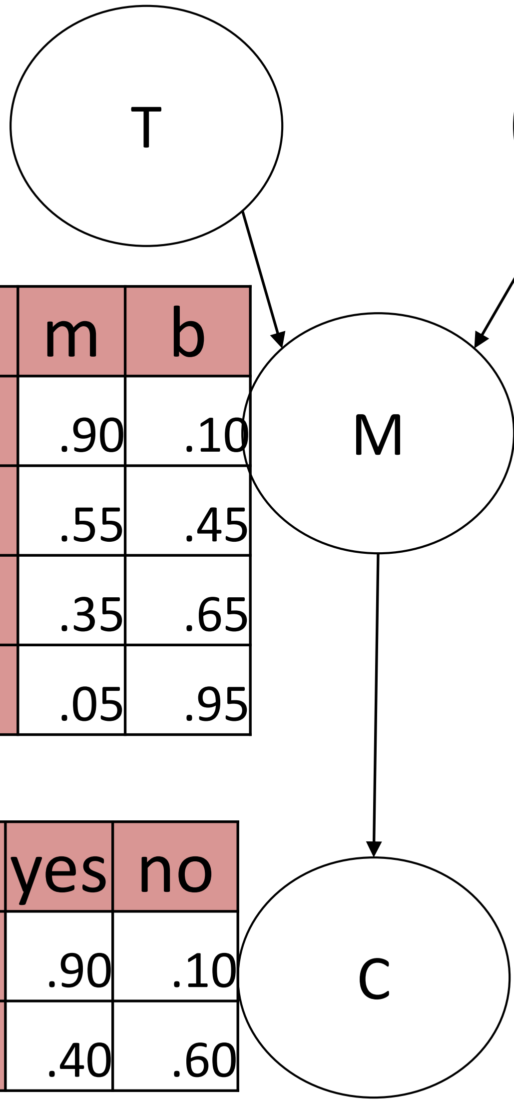
$$+ \sum_t P(T) *0.80 * P(b|T,v) *0.60$$

$$+ \sum_t P(T) *0.20* P(b|T,s) *0.60$$

M	yes	no
m	.90	.10
b	.40	.60

III: The probability of non-confession

e	n
.05	.95



v	s
.80	.20

T	W	m	b
e	v	.90	.10
e	s	.55	.45
n	v	.35	.65
n	s	.05	.95

M	yes	no
m	.90	.10
b	.40	.60

$$P(\neg c|\_)/\alpha = \sum_t P(T) * 0.80 * P(m|T,v) * 0.10$$
$$+ \sum_t P(T) * 0.20 * P(m|T,s) * 0.10$$
$$+ \sum_t P(T) * 0.80 * P(b|T,v) * 0.60$$
$$+ \sum_t P(T) * 0.20 * P(b|T,s) * 0.60 =$$

$P(e)*0.80 * P(m|e,v) * 0.10+P(n)*0.80 * P(m|n,v)*0.10 +$

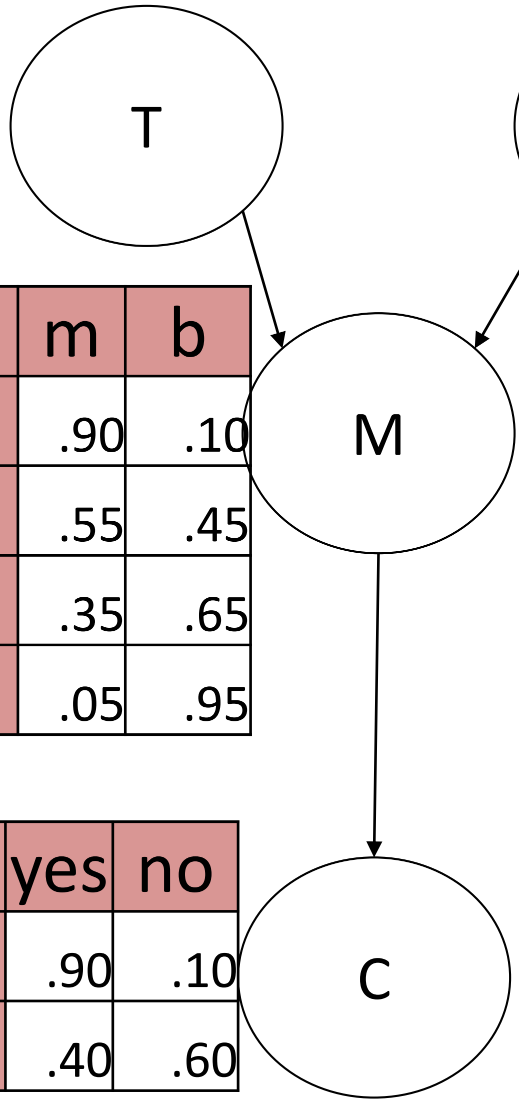
$P(e) *0.20* P(m|e,s) *0.10+ P(n) *0.20* P(m|n,s) *0.10 +$

$P(e) *0.80 * P(b|e,v) *0.60+ P(n) *0.80 * P(b|n,v) *0.60 +$

$P(e) *0.20* P(b|e,s) *0.60+ P(n) *0.20* P(b|n,s) *0.60$

III: The probability of non-confession

e	n
.05	.95



v	s
.80	.20

T	W	m	b
e	v	.90	.10
e	s	.55	.45
n	v	.35	.65
n	s	.05	.95

M	yes	no
m	.90	.10
b	.40	.60

$$P(\neg c|\_)/\alpha = \sum_t P(T) *0.80 * P(m|T,v) *0.90$$

$$+ \sum_t P(T) *0.20* P(m|T,s) *0.90$$

$$+ \sum_t P(T) *0.80 * P(b|T,v) *0.40$$

$$+ \sum_t P(T) *0.20* P(b|T,s) *0.40 =$$

$0.05*0.80 * P(m|e,v) * 0.10+0.95*0.80 * P(m|n,v)*0.10 +$

$0.05 *0.20* P(m|e,s) *0.10+ 0.95 *0.20* P(m|n,s) *0.10 +$

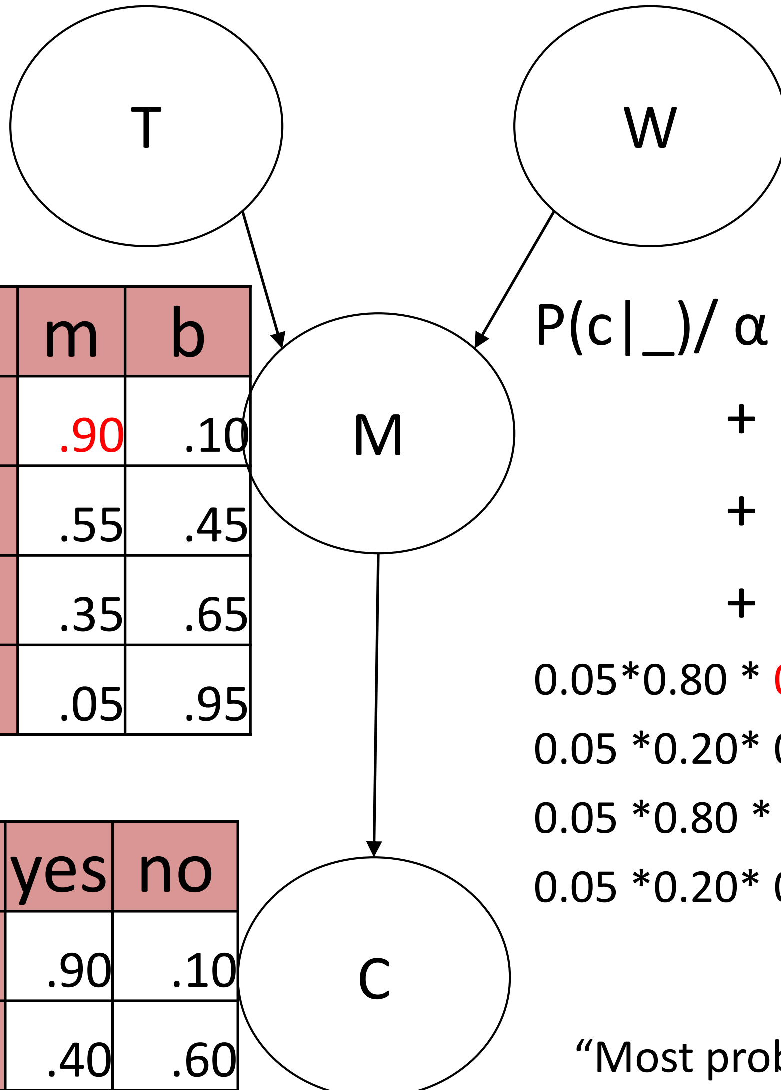
$0.05 *0.80 * P(b|e,v) *0.60+ 0.95*0.80 * P(b|n,v) *0.60 +$

$0.05 *0.20* P(b|e,s) *0.60+ 0.95 *0.20* P(b|n,s) *0.60$

Example 3: The probability of non-confession

e	n
.05	.95

v	s
.80	.20



T	W	m	b
e	v	.90	.10
e	s	.55	.45
n	v	.35	.65
n	s	.05	.95

M	yes	no
m	.90	.10
b	.40	.60

$$P(c|\_)/\alpha = \sum_t P(T) *0.80 * P(m|T,v) *0.90$$
$$+ \sum_t P(T) *0.20* P(m|T,s) *0.90$$
$$+ \sum_t P(T) *0.80 * P(b|T,v) *0.40$$
$$+ \sum_t P(T) *0.20* P(b|T,s) *0.40 =$$

0.05*0.80 * 0.90 * 0.10+0.95*0.80 * 0.35*0.10 +
0.05 *0.20* 0.55 *0.10+ 0.95 *0.20* 0.05 *0.10 +
0.05 *0.80 * 0.10 *0.60+ 0.95*0.80 *0.65 *0.60 +
0.05 *0.20* 0.45 *0.60+ 0.95 *0.20* 0.95 *0.60 = 0.44

"Most probably they will confess" - concluded the detective, and went home