# Bayesian Belief Networks

Lecture 6

Outline

- ▶ Extending the Naïve Bayes network to more complex networks. Joint probability

- Bayesian Belief networks (BBN). Definition. Types

- Query on BBN: what nodes to include. Markov blanket

- Query on BBN: how to compute. Simple example. Some variables may be hidden.

- Examples

- Finding network topology

- Applications of Bayesian networks

Naïve Bayes as a graph
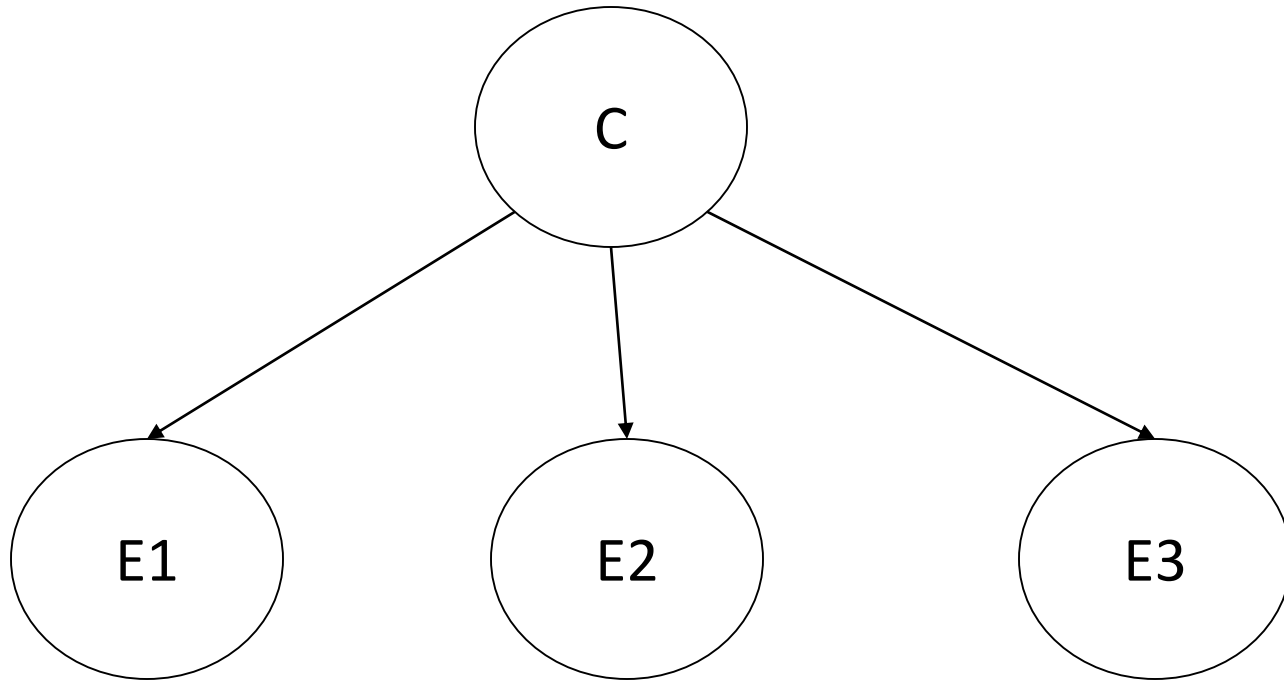
This graph states that there is a ***probabilistic dependence*** between C and each E_i. The probability of one of these variables (Class to predict) is influenced by the probabilities of the rest of the variables (set of evidences) and vice versa

Joint probability

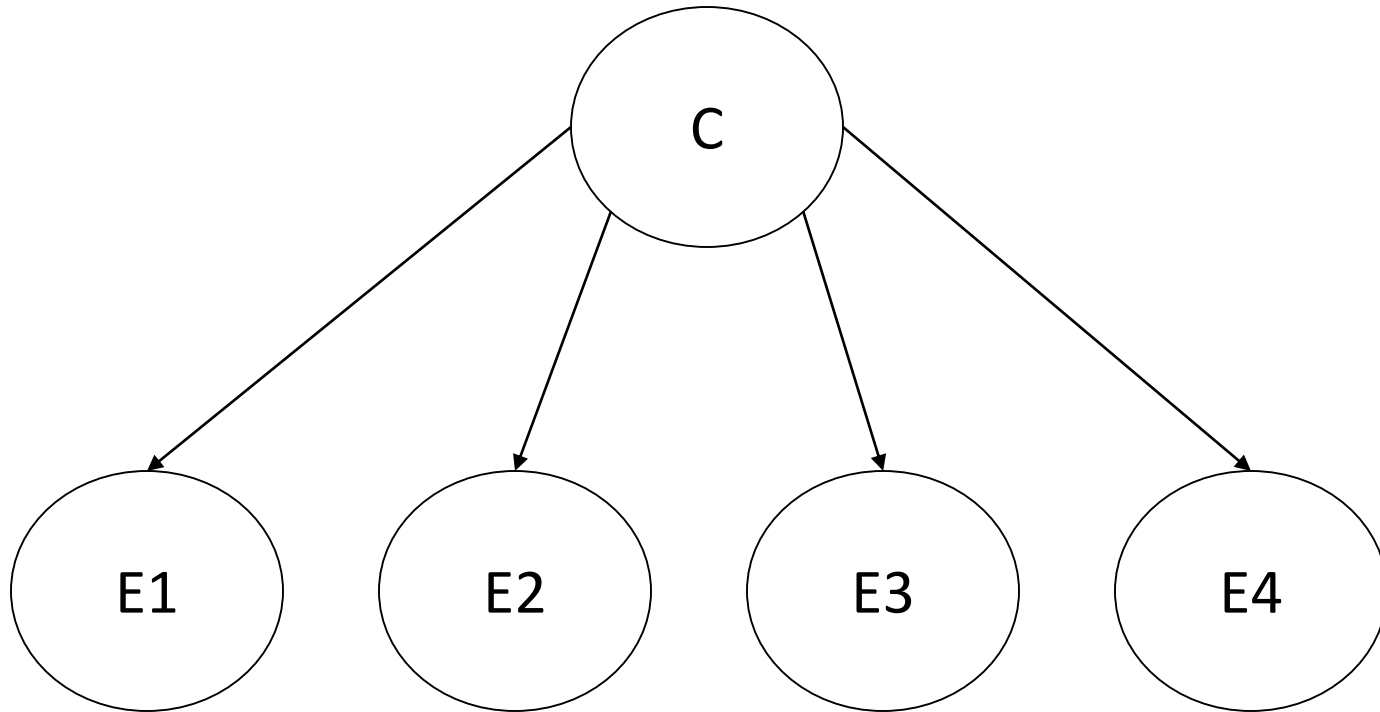- P(C|E1,E2,E3) = P(C, E1,E2,E3) / P(E1,E2,E3)
- P(C|E)=P(C ∩ E)/P(E)

Joint probability of all variables in the network

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

- P(C ∩ E1 ∩ E2 ∩ E3) = P(E1|C) P(E2|C) P(E3|C) P(E4|C) *P(C)

We can multiply because we assume: there is no probabilistic dependence between Ei and Ej, given the parent value C

Naïve Bayes models joint probability distribution for all variables in the network

$P(c|e1,e2,e3,e4)=P(c,e1,e2,e3,e4) /P(e1,e2,e3,e4)=\alpha\ P(c,e1,e2,e3,e4)$

$P(\neg c|e1,e2,e3,e4)=P(\neg c,e1,e2,e3,e4)/P(e1,e2,e3,e4)= \alpha\ P(\neg c,e1,e2,e3,e4)$

In fact, for prediction, it is enough to compute the joint probability of all participating variables for **c** and **¬c**, and to compare

Chain rule to find P(c,e1,e2,e3,e4)
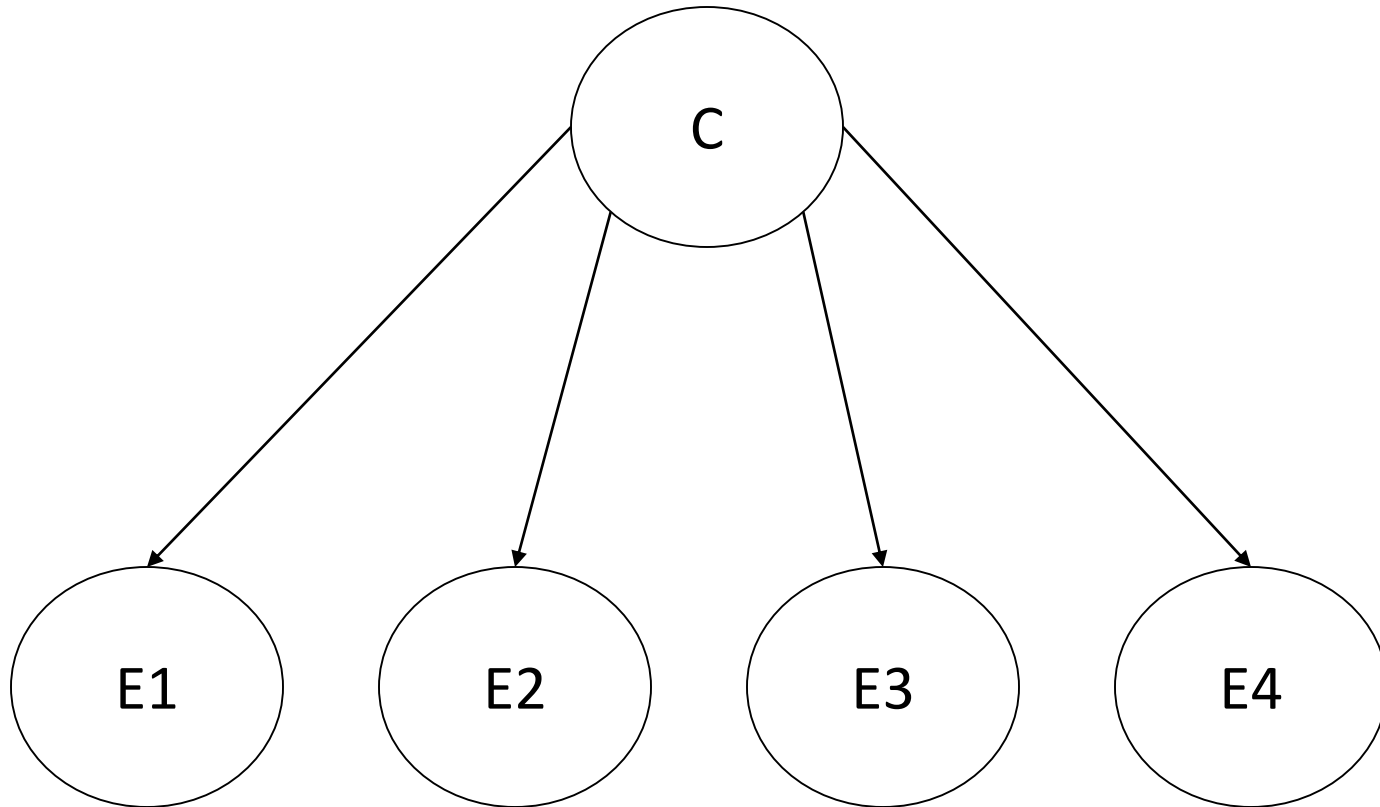
P($\textcolor{red}{c}$|e1,e2,e3,e4)=α P(c)P(e1|c)P(e2|c)P(e3|c)

P($\textcolor{red}{\neg c}$|e1,e2,e3,e4)= α P(¬c)P(e1|¬c)P(e2|¬c)P(e3|¬c)

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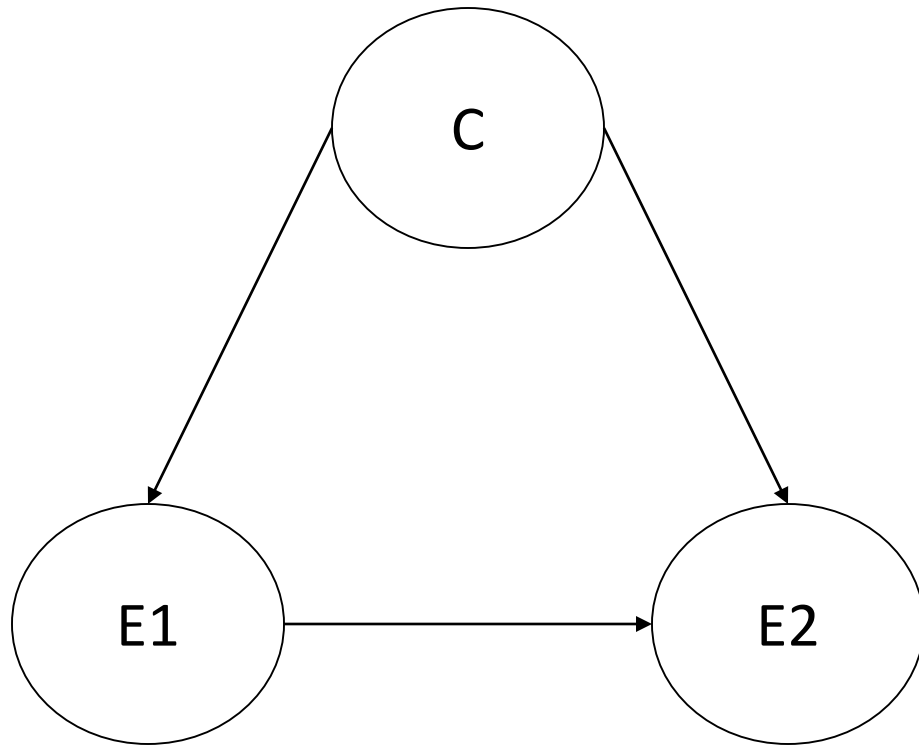
What if E1 and E2 are not independent?
For each node with more than 1 parent we need Conditional Probability Table (CPT) with probability distribution for all possible combinations of parent variables:

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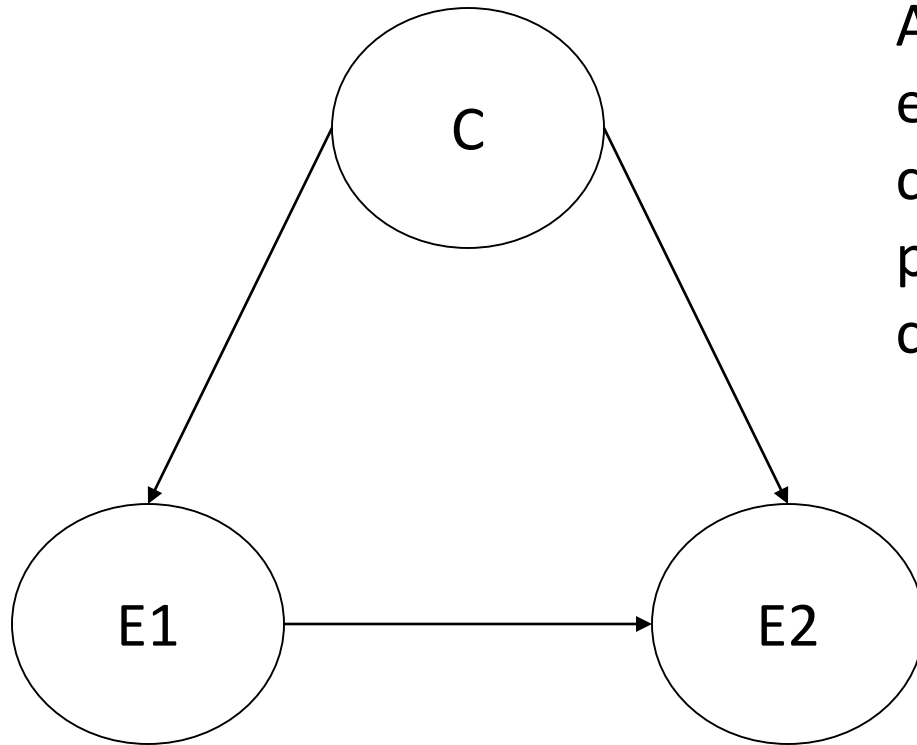
		E2	
C	E1	e2	¬e2
c	e1	P(e2|c,e1)	
c	¬e1		
¬c	e1		
¬c	¬e1		

More 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		

More 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.

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

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

E1 has 1 parent so its probability is conditioned on C

CPT for attribute E2
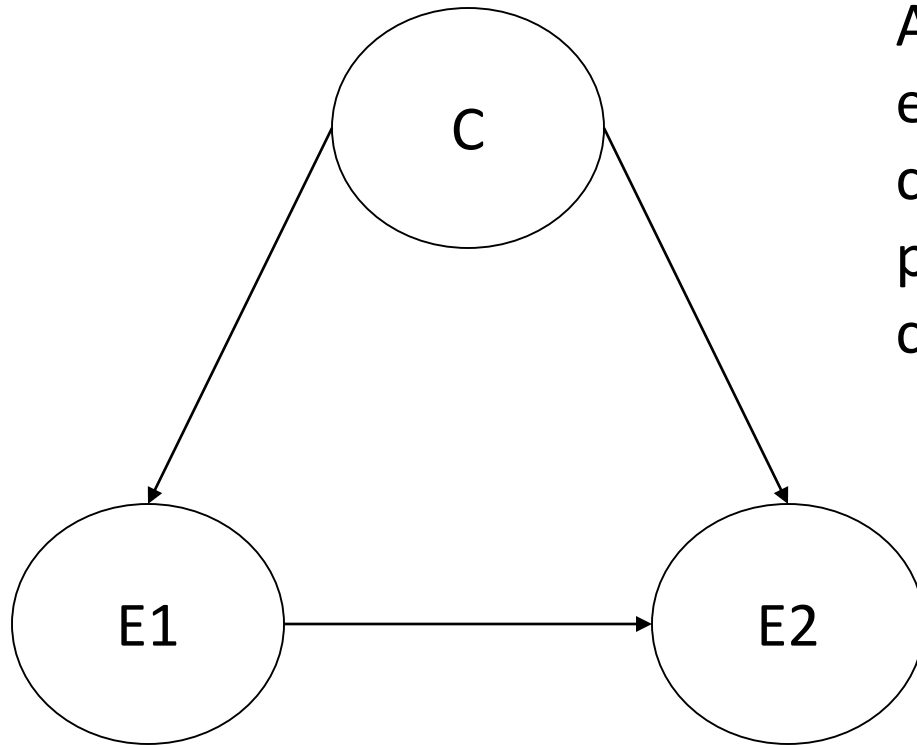
		E2	
C	E1	e2	¬e2
c	e1	$P(e2|c,e1)$	
c	¬e1		
¬c	e1		
¬c	¬e1		

More 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.

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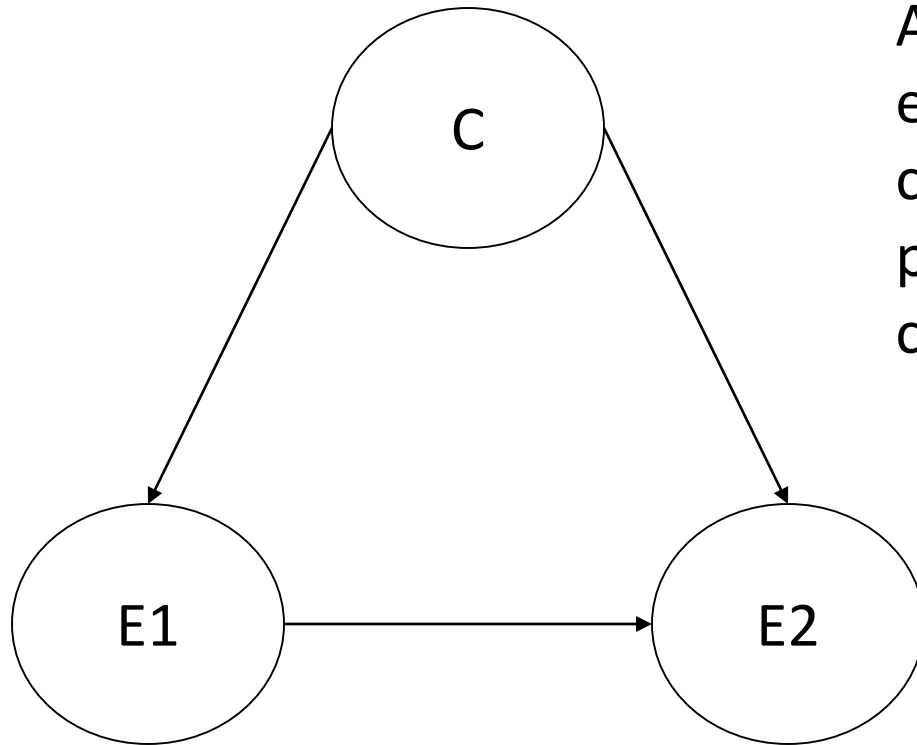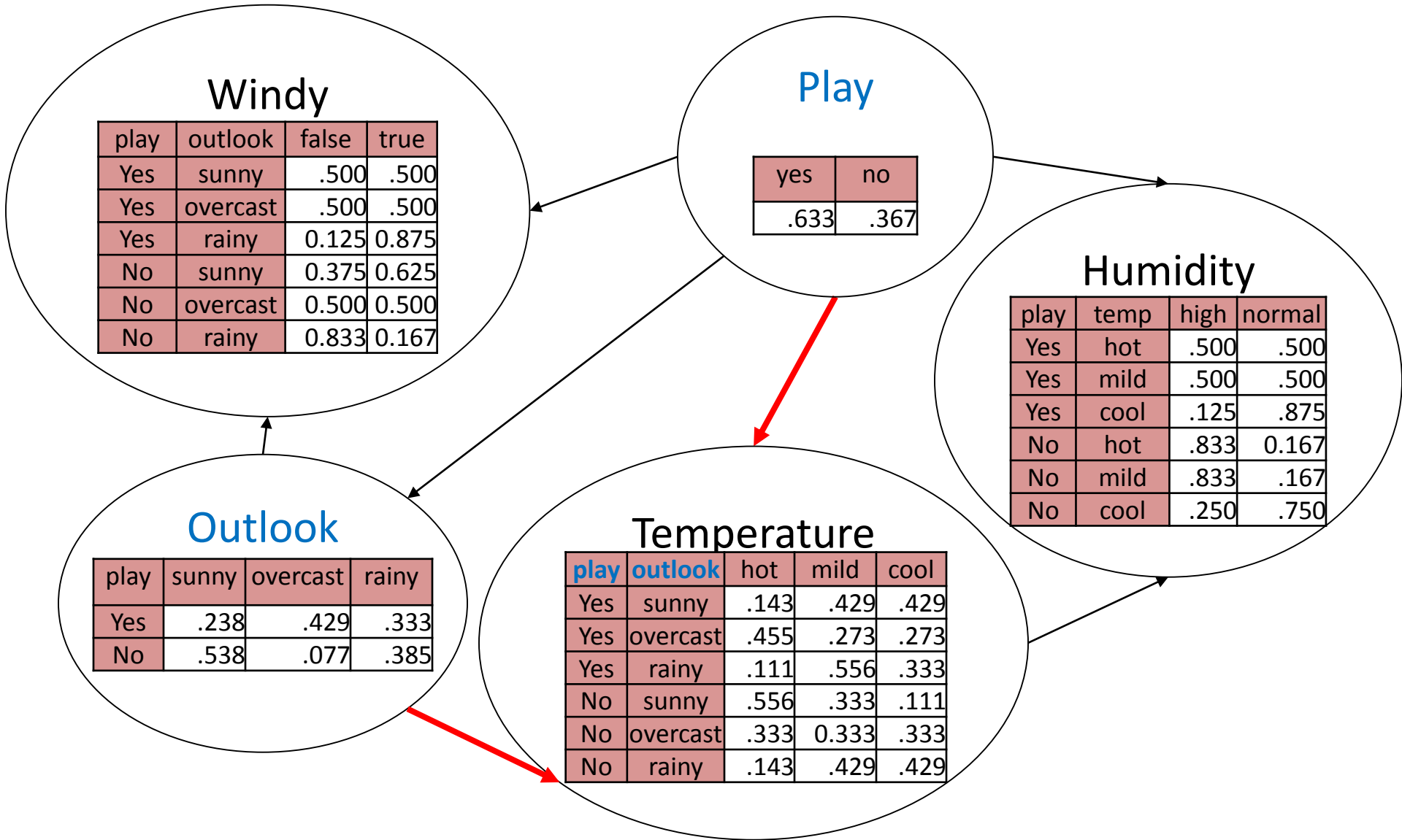
$P(\neg c|e1,e2)=\alpha\ P(\neg c)P(e1|\neg c)\ \textbf{P(e2|\neg c,e1)}$

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		

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	?

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

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	?

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

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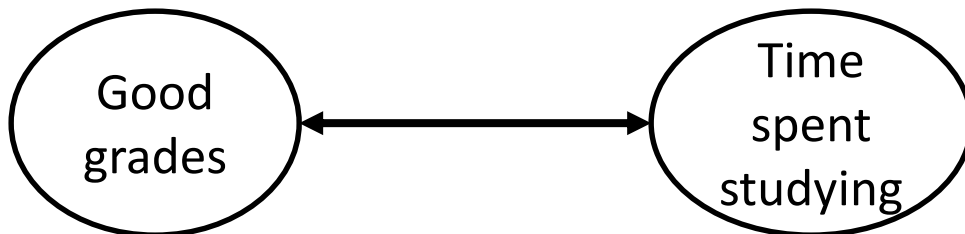All these probabilities are known – just plug them in and compute

Outline

- Extending the Naïve Bayes network to more complex networks. Joint probability
- ▶ Bayesian Belief networks (BBN). Definition. Types
- Query on BBN: what nodes to include. Markov blanket
- Query on BBN: how to compute. Simple example. Some variables may be hidden.
- Examples
- Finding network topology
- Applications of Bayesian networks

Bayesian Belief Networks

- 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

Good grades ◄──────► Time spent studying

The meaning of edges in BBNs: types of networks

Causal

Increased probability of A makes B more likely.

A causes B

Intercausal

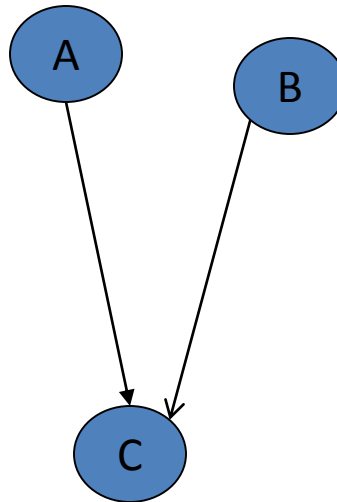A and B can each cause C. B explains C and so is evidence against A

Evidential

Increased probability of B makes A more likely.

B is evidence for A,

A depends on B

Outline

- Extending the Naïve Bayes network to more complex networks. Joint probability
- Bayesian Belief networks (BBN). Definition. Types
- Query on BBN: what nodes to include. Markov blanket
- Query on BBN: how to compute. Simple example. Some variables may be hidden.
- Examples
- Finding network topology
- Applications of Bayesian networks

Bayesian Belief Networks: query

- 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
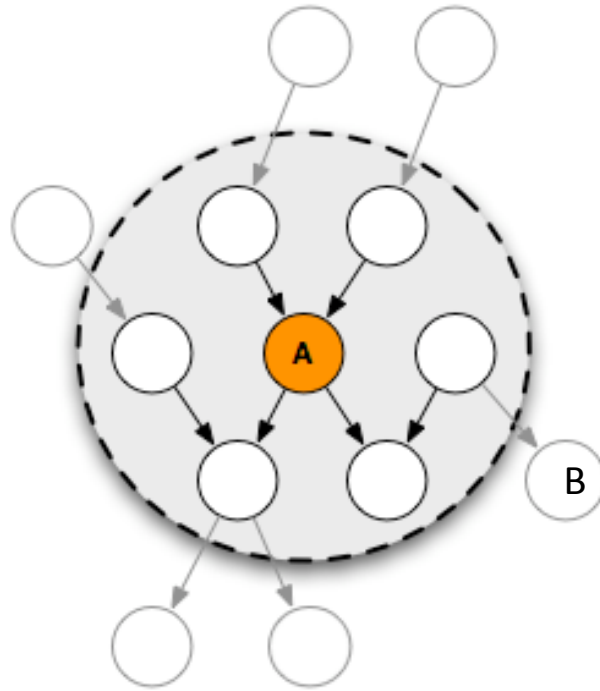
Markov Blanket Assumption

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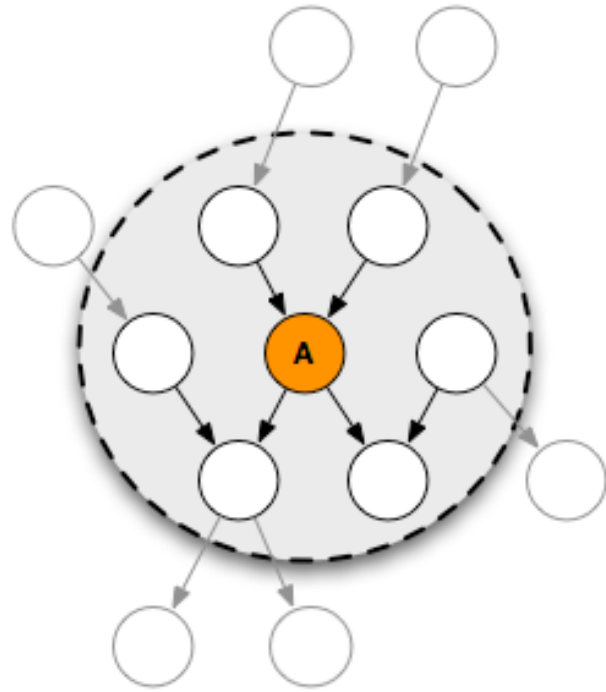
The Markov blanket of a node contains all the variables that shield the node from the rest of the network. This means that the Markov blanket of a node is the only knowledge needed to predict the behavior of that node.

The Markov blanket assumption

- 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

Markov blanket example 1

Cold

Sneezing

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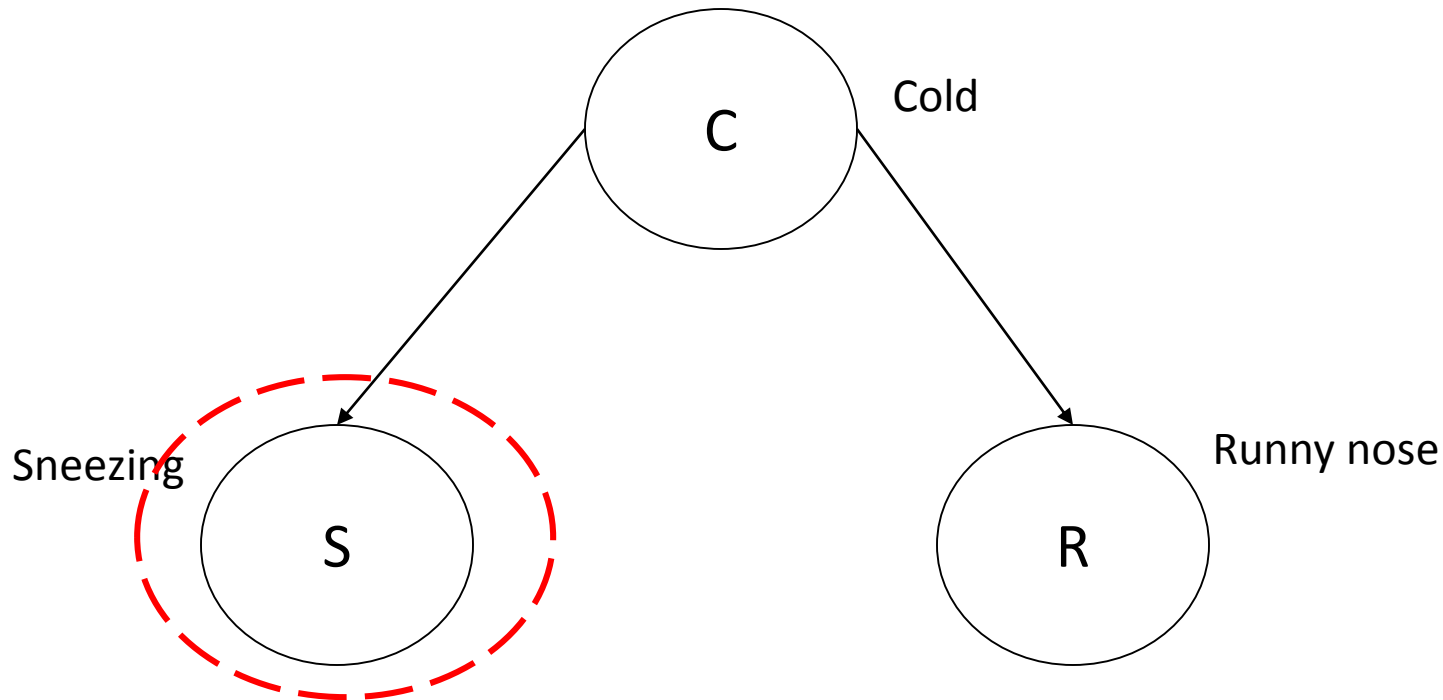
P(S|R)>P(S)
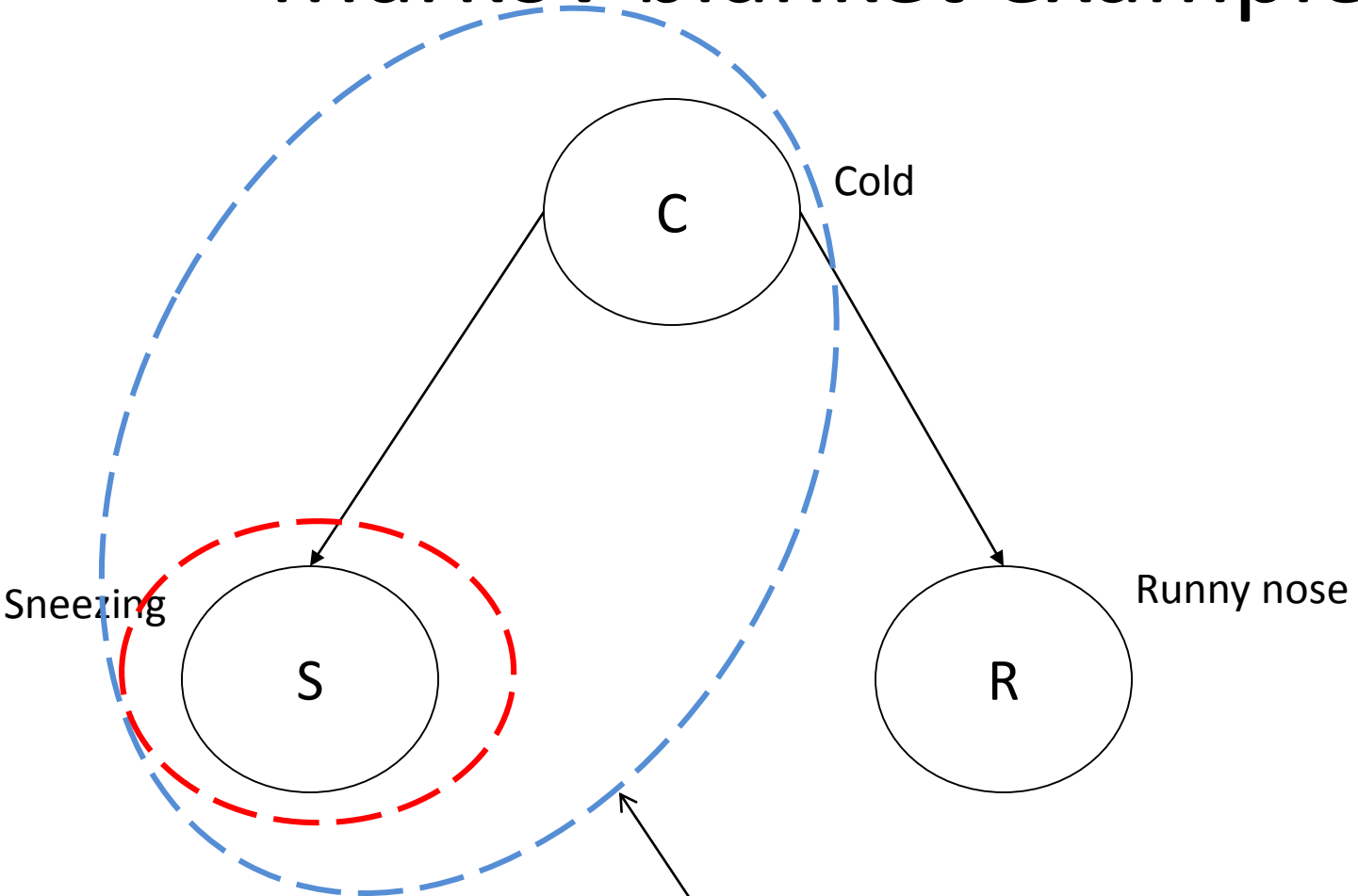Thus, S is not independent of R: R makes C more probable, which in turn influences the probability of S.
However, P(S|C) is independent of R: if we know the value of C (C is given), then R does not influence the probability of S:
P(S|C,R)=P(S|C)   - C 'shields' node S from the influence of R

Markov blanket example 1

Cold

C

Sneezing

S

Runny nose

R

Markov blanket of node S

Children of parents - No

Markov blanket example 2
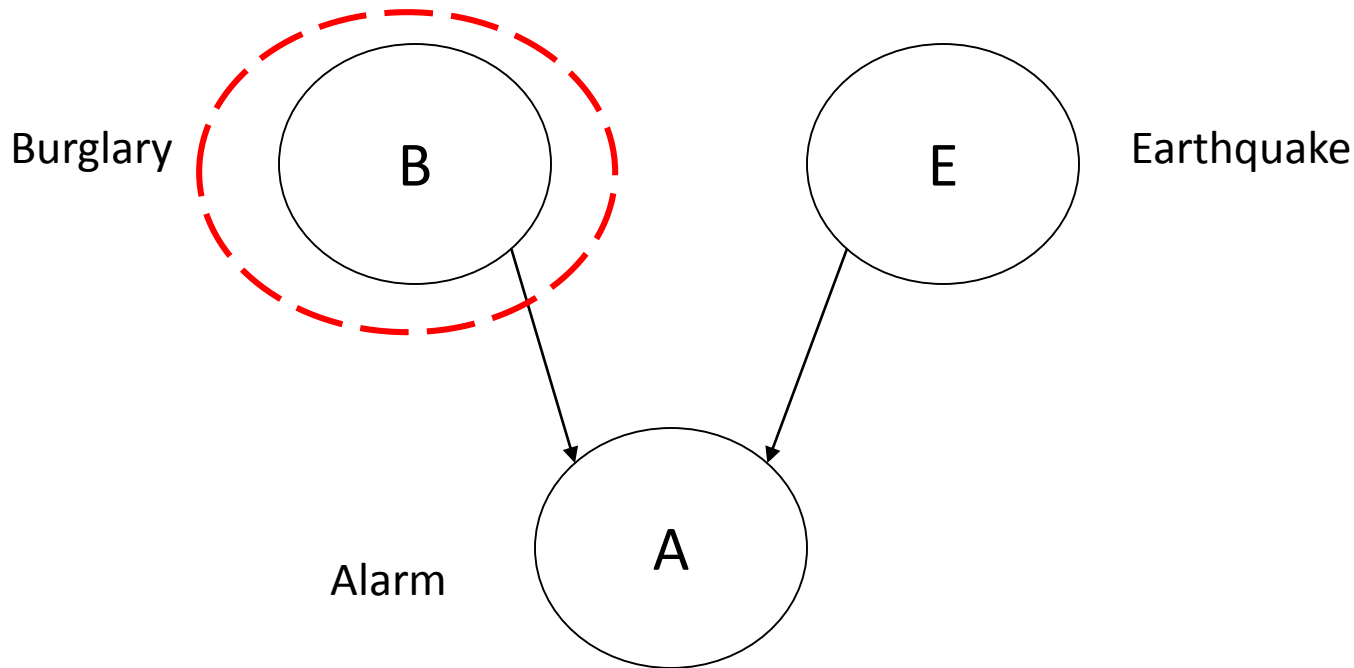
P(B|E)=P(B) (independent), but P(B|A,E)<P(B|A)
If you hear an alarm, you might evaluate the probability of B,
but if you know that it was an earthquake, this probability
decreases: E 'discounts' B, E is evidence against B, then it
should be included in its Markov blanket together with A

Markov blanket example 2

Burglary B → A ← E Earthquake

Alarm A

Markov blanket of node B

Parents of children - Yes

Markov blanket example 3

Cheating spouse — C

Dining with stranger — D

Seen dining with stranger — S
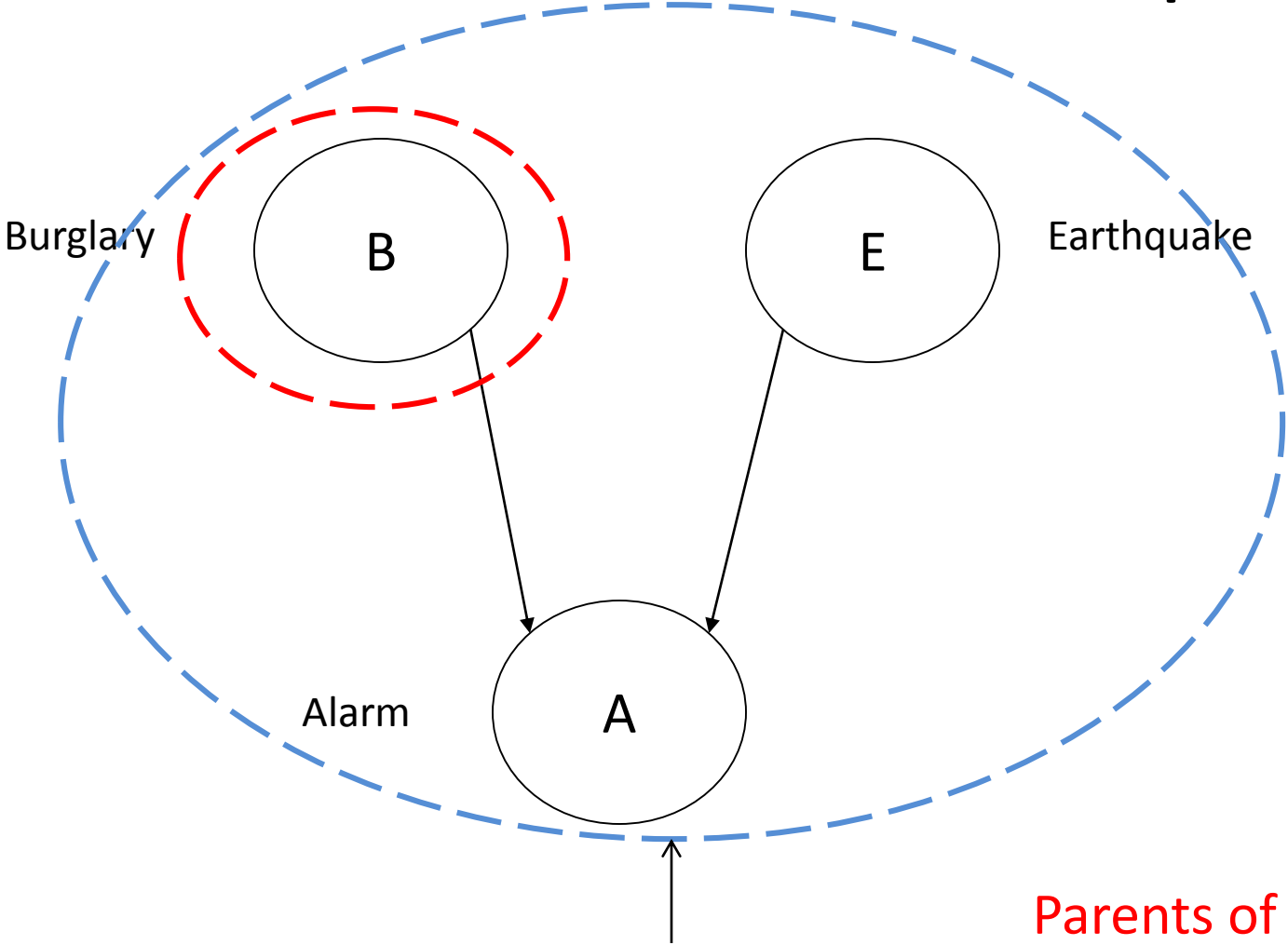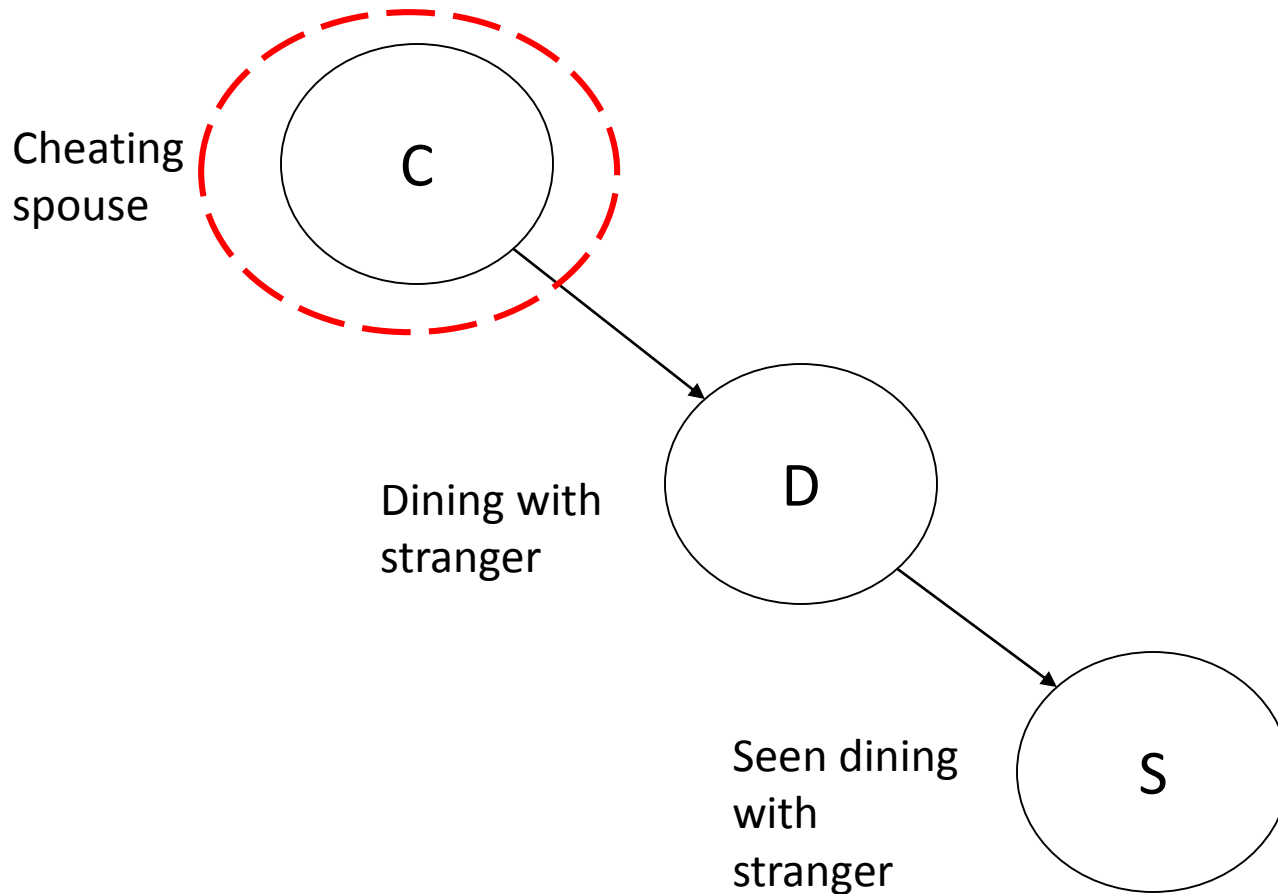
$P(C|S)>P(C)$
$P(C|D,S)=P(C|D)$

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

Markov blanket example 3

Cheating spouse

C

Dining with stranger

D

Seen dining with stranger

S
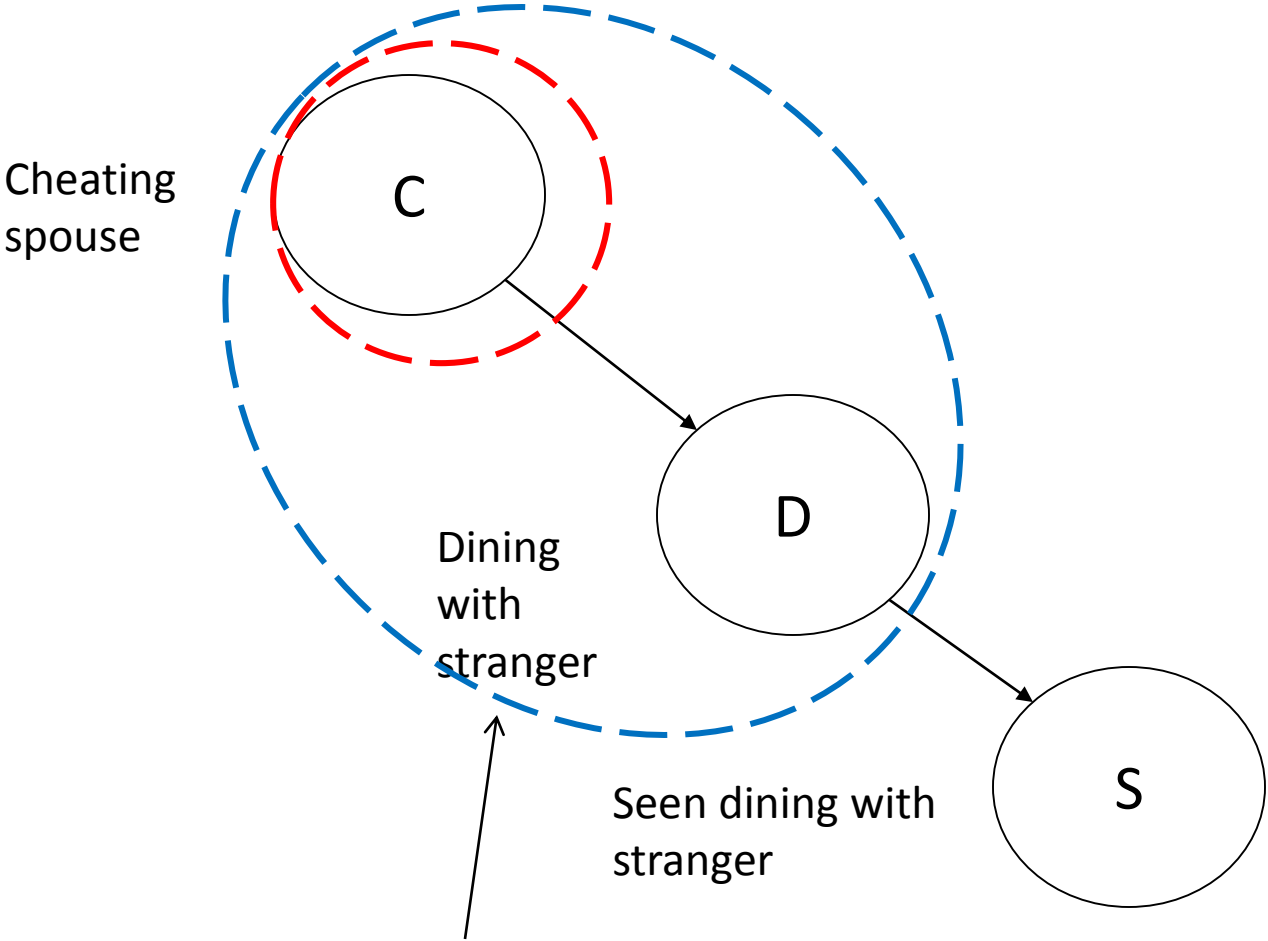
Markov blanket of node C

Children of children - No

Remember

- When performing a query on Bayesian network, include parents, children and the parents of the children of participating nodes into the calculation of joint probabilities

Market blanket of node C?
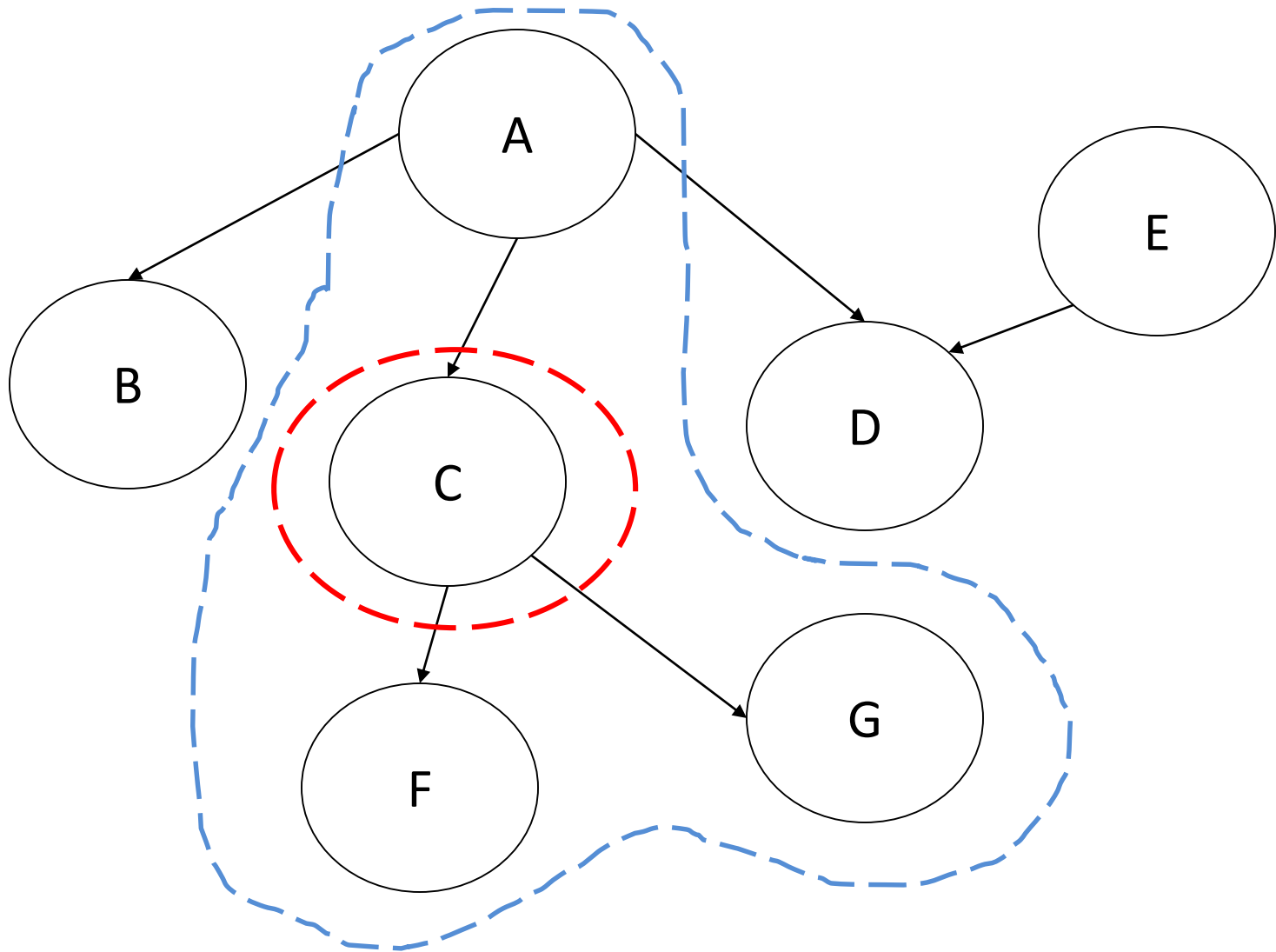
Market blanket of node C?

Alarm example

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

P(B)
.001

P(E)
.002

Burglary

Earthquake

Alarm

JohnCalls

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

MaryCalls

A	P(M|A)
T	.70
F	.01
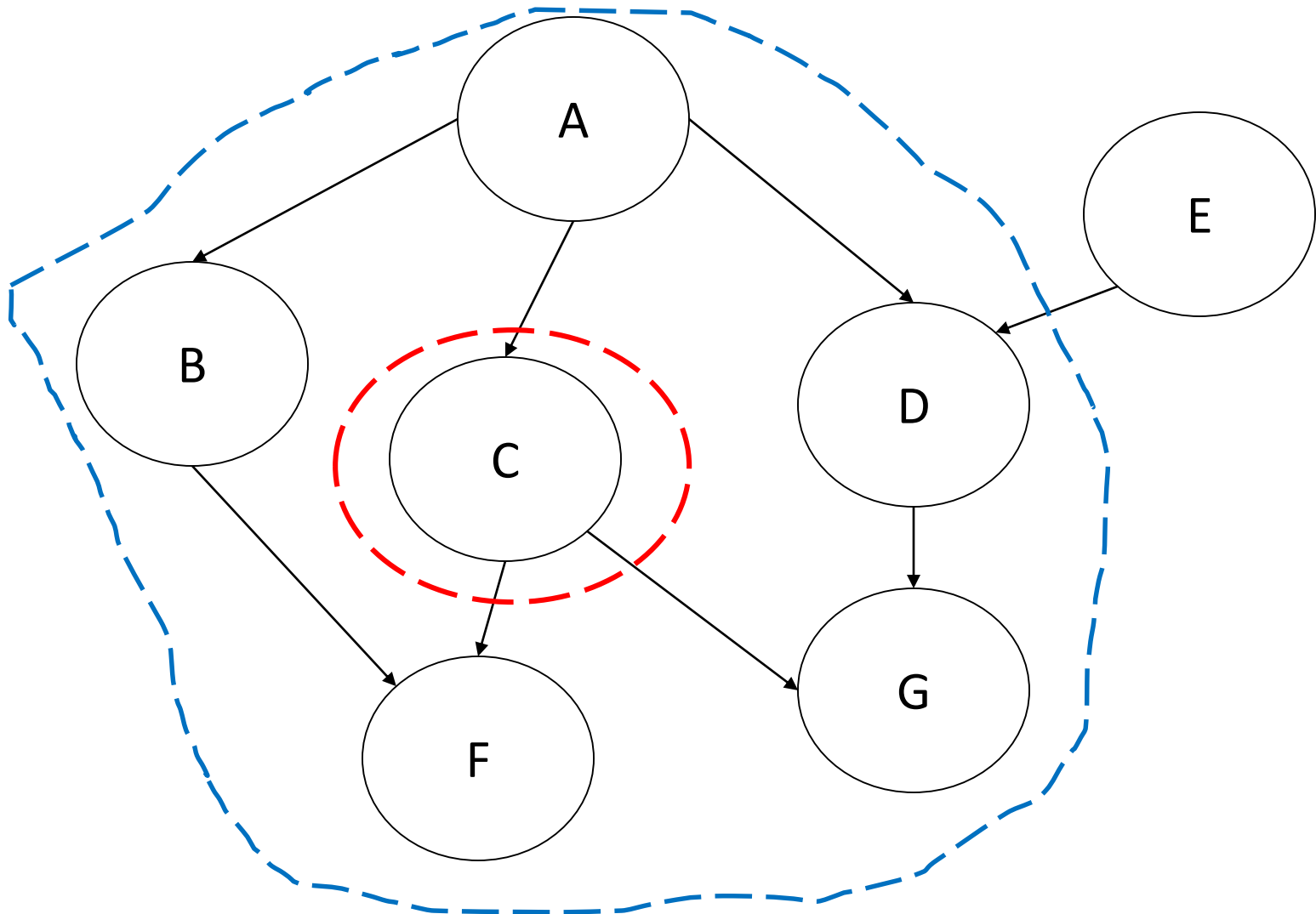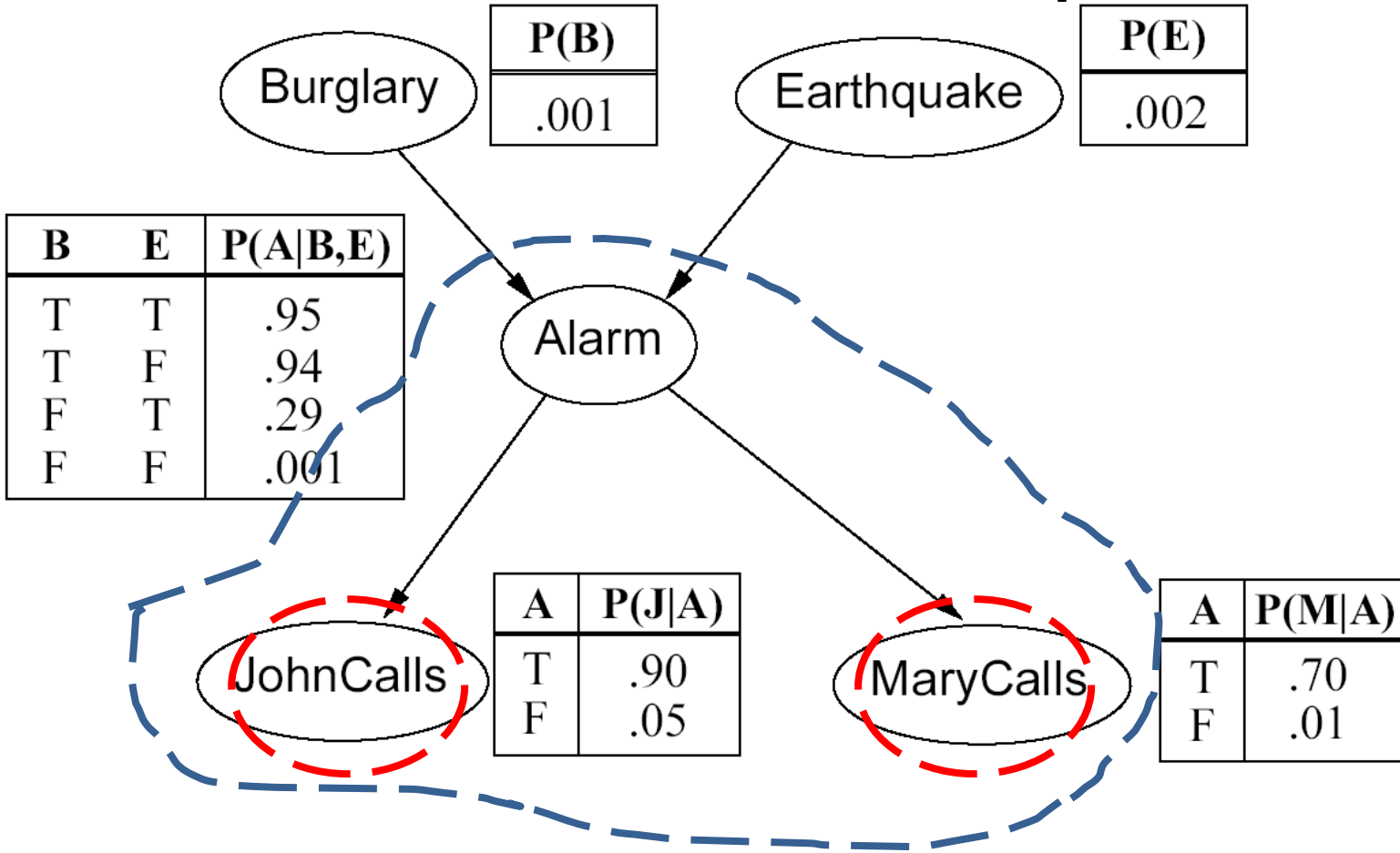
Query: what is the probability of John calling given that Mary called

Alarm example

	P(B)
	.001

Burglary

	P(E)
	.002

Earthquake

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

Alarm

JohnCalls

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

MaryCalls

A	P(M|A)
T	.70
F	.01
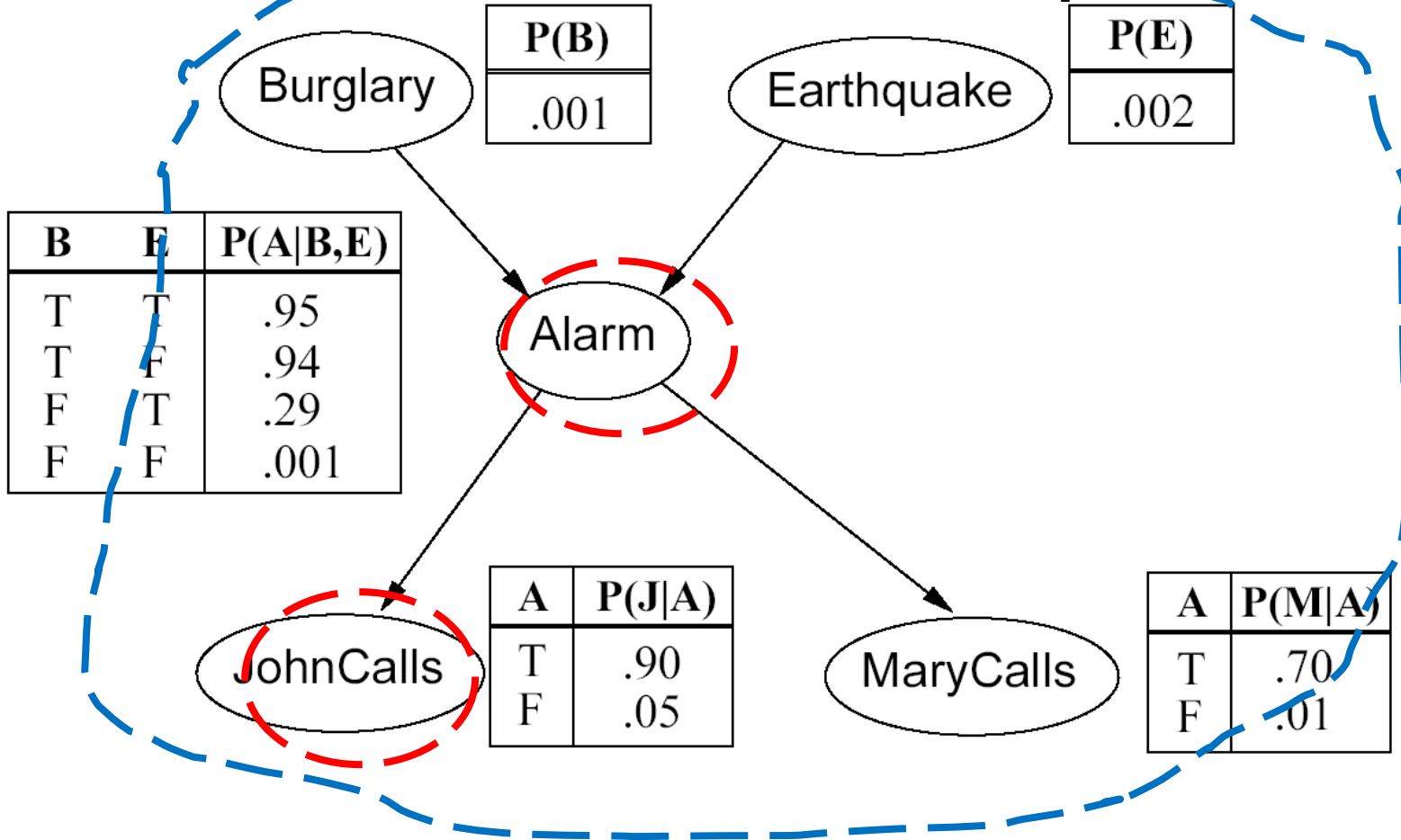
Query: what is the probability of Alarm given that John called

Alarm example

	P(B)
	.001

Burglary

Earthquake

	P(E)
	.002

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

Alarm

JohnCalls

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

MaryCalls

A	P(M|A)
T	.70
F	.01
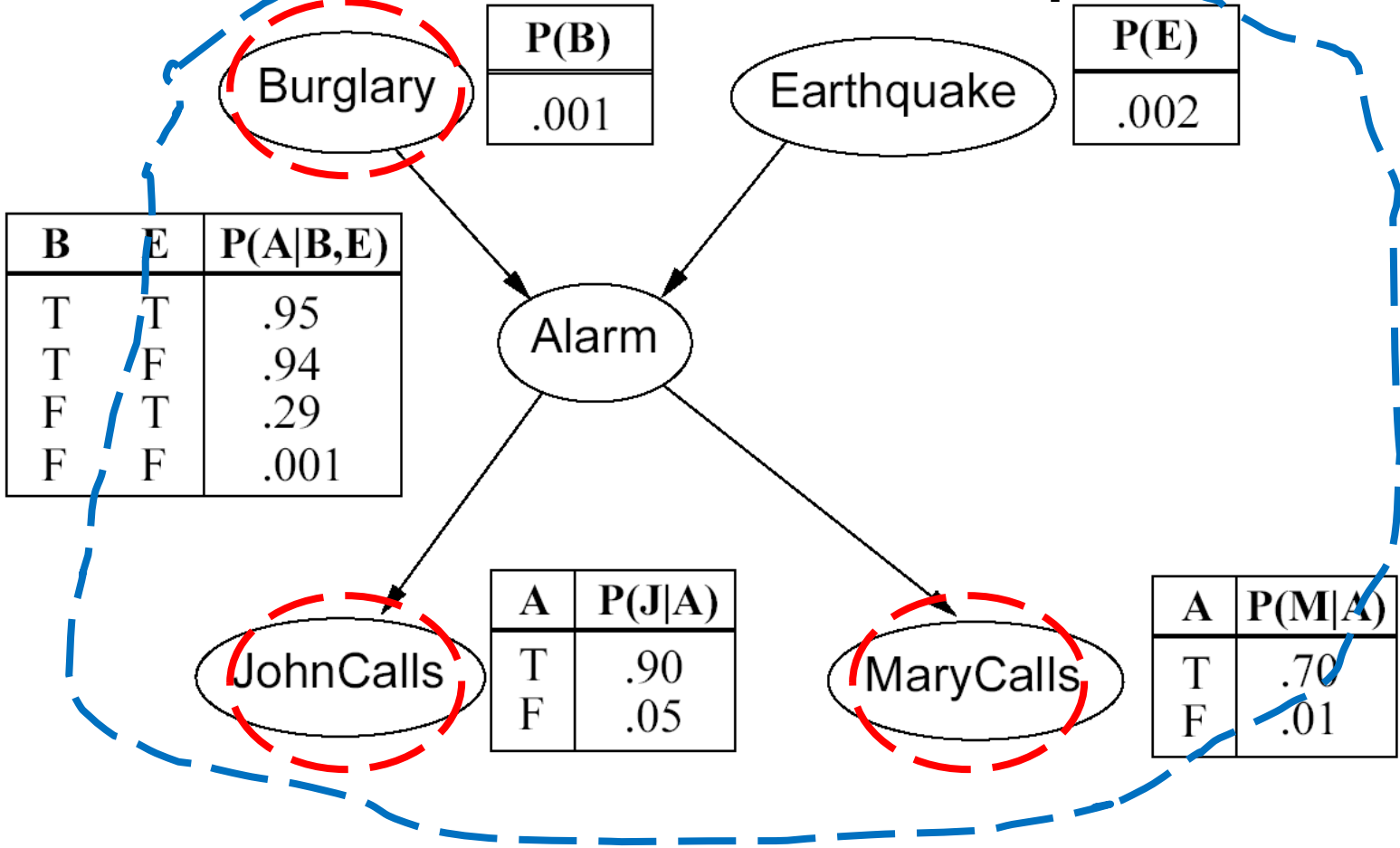
Query: what is the probability of Burglary given that John called and Mary called

Another example: High Blood Pressure

	H=Yes
D=Healthy	0.2
D=Unhealthy	0.35

	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

	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

	BP=High
HD=Yes	0.85
HD=No	0.2

E=Yes
0.7

D=Healthy
0.25
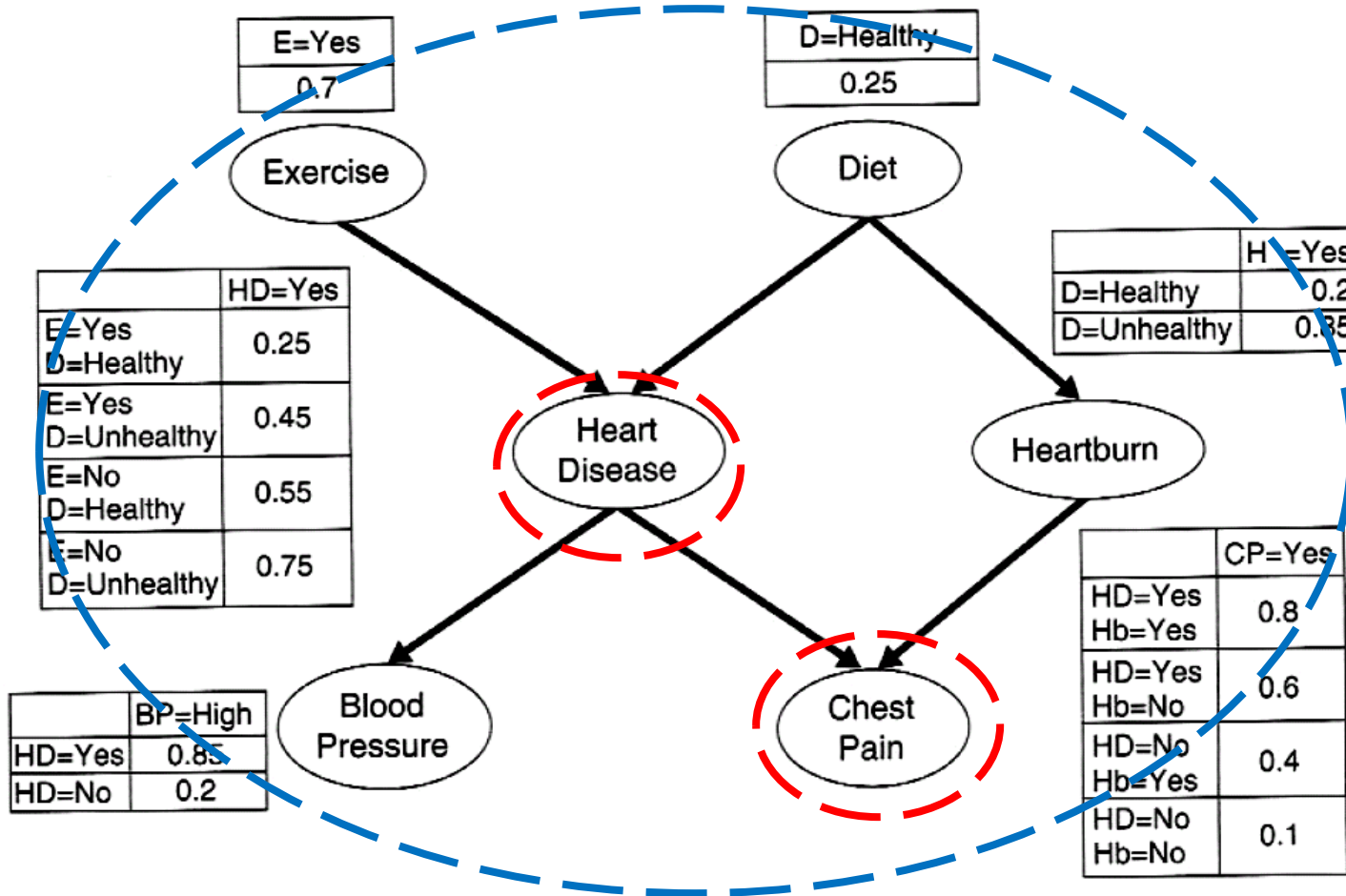
Query: what is the probability of Heart disease given chest pain

Outline

- Extending the Naïve Bayes network to more complex networks. Joint probability

- Bayesian Belief networks (BBN). Definition. Types

- Query on BBN: what nodes to include. Markov blanket

▶ - Query on BBN: how to compute. Simple example. Some variables may be hidden.

- Examples

- Finding network topology

- Applications of Bayesian networks

The recipe for computing any query

- In complex networks: select subset of nodes which are inside Market blankets of participating nodes

- Compute joint probabilities of these nodes by the chain rule, substituting random variables by the evidence values

Wet Grass example: predicting rain

	Sprinkler	
Rain	s	¬s
r	0.40	0.60
¬r	0.99	0.01
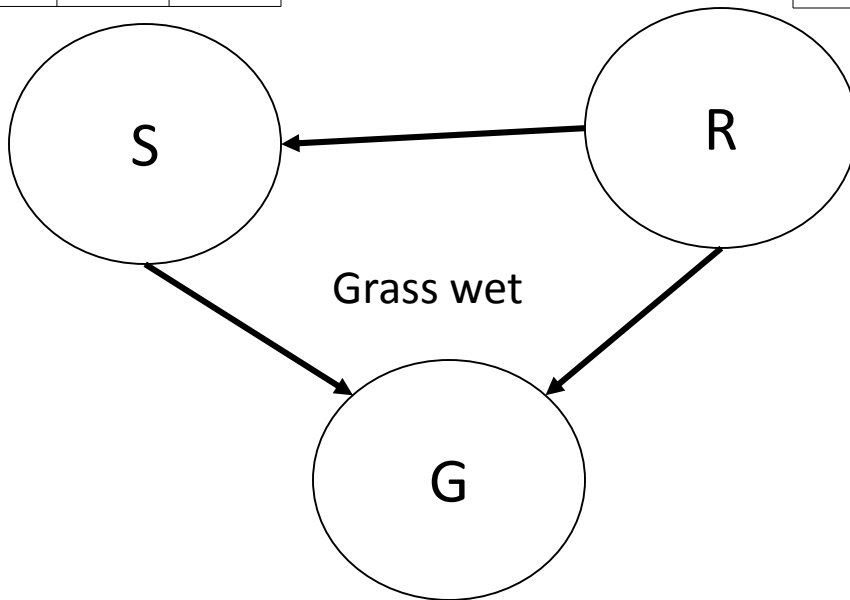
Rain	
r	¬r
0.20	0.80

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

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

S

R

Grass wet

G

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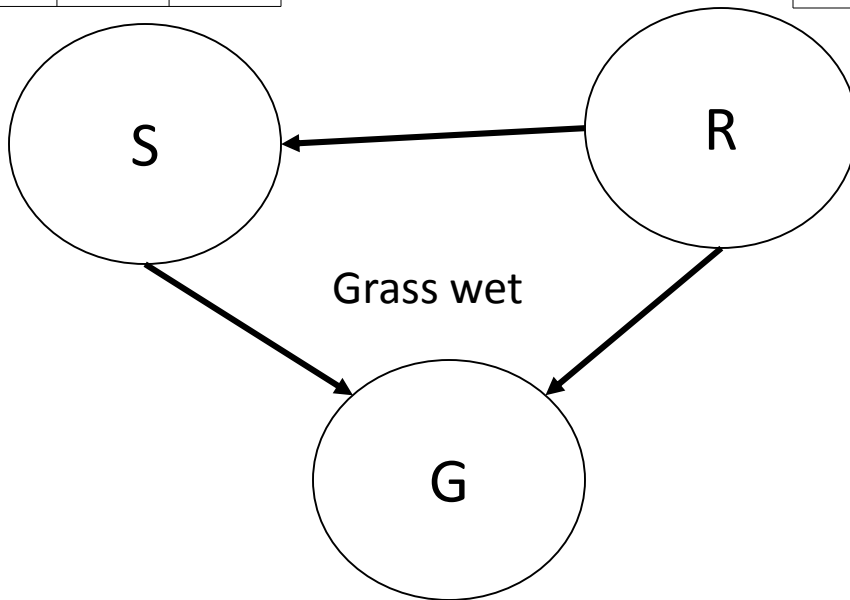
	Sprinkler	
Rain	s	¬s
r	0.40	0.60
¬r	0.99	0.01

Rain	
r	¬r
0.20	0.80

S = ¬s, G=g

$P(r|g, ¬s)=?$
$P(¬r|g, ¬s)=?$

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

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

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

		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

S= ¬s, G=g

P(r|g, ¬s)=?
P(¬ r|g, ¬s)=?

P(r|g, ¬s)
= α P(r) P(¬s|r) P(g|r, ¬s)
= α 0.20*0.60*0.8
= α 0.096

P(¬ r|g, ¬s)
= α P(¬ r) P(¬s| ¬ r) P(g| ¬ r, ¬s)
= α 0.80 * 0.01 *0.01
= α 0.00008

S → G

R → S

R → G

Grass wet

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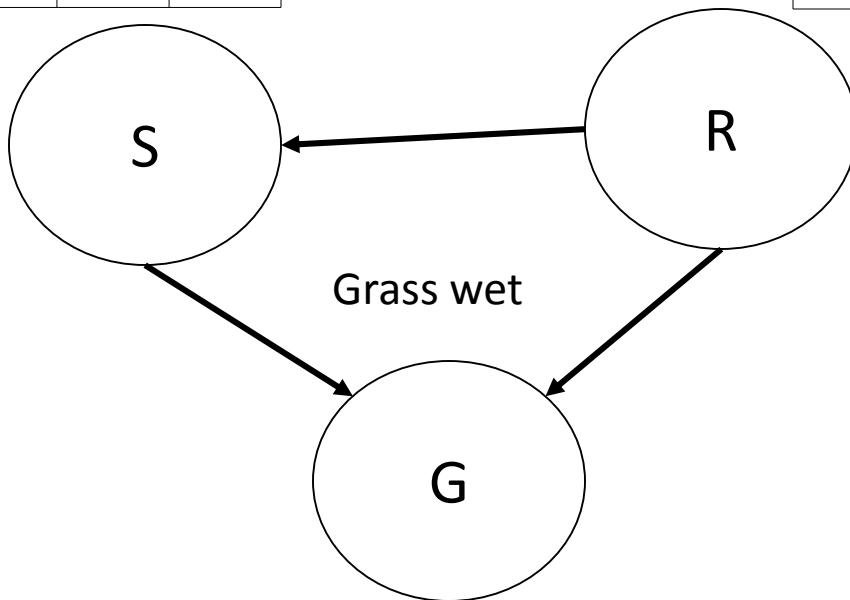
		Grass wet	
Rain	Sprinkler	g	¬g
r	s	0.99	0.01
r	¬s	0.8	0.2
¬r	s	0.90	0.10
¬r	¬s	0.01	0.99

Definitely, it was raining

Wet Grass example: hidden variables
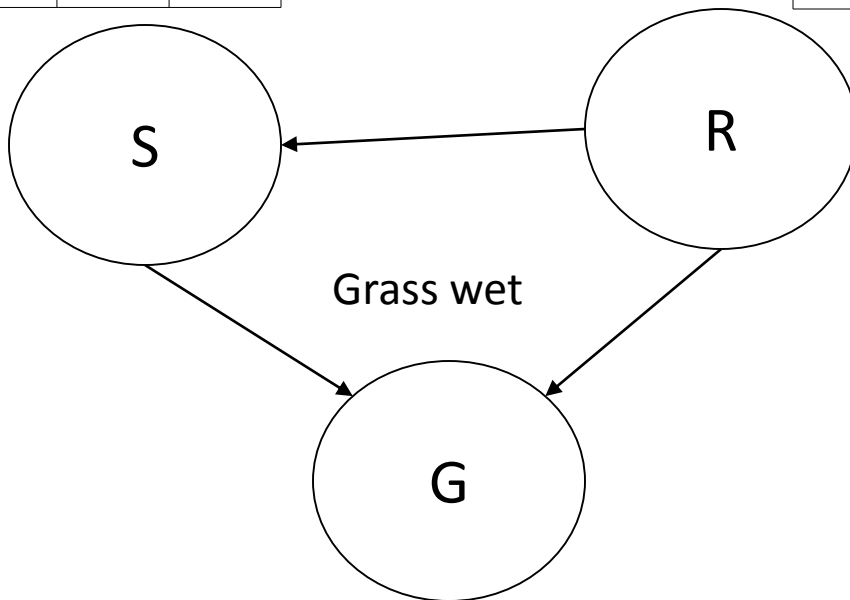
	Sprinkler	
Rain	s	¬s
r	0.40	0.60
¬r	0.99	0.01

Rain	
r	¬r
0.20	0.80

S

R

Grass wet

G

		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)=α P(r, g, S)
= α P(r) P(S|r) P(g|r, S)=
= α P(r) *
[P(s|r) P(g|r, s)+ P(¬s|r) P(g|r, ¬s)]

Sprinkler was or on or off

Wet Grass example: hidden variables
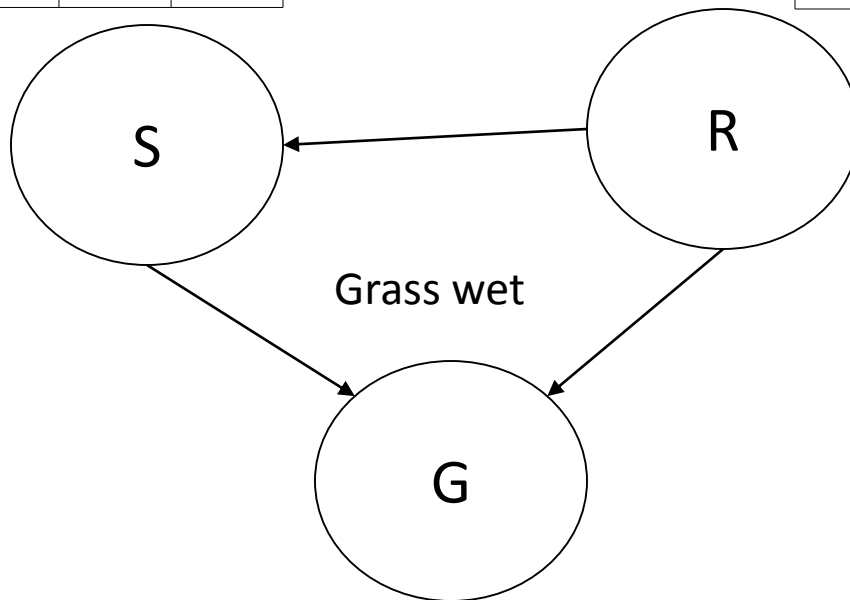
	Sprinkler	
Rain	s	¬s
r	0.40	0.60
¬r	0.99	0.01

Rain	
r	¬r
0.20	0.80

S

R

Grass wet

G

		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

Outline

- Extending the Naïve Bayes network to more complex networks. Joint probability
- Bayesian Belief networks (BBN). Definition. Types
- Query on BBN: what nodes to include. Markov blanket
- Query on BBN: how to compute. Simple example. Some variables may be hidden.
- ▶ Examples: I. High blood pressure
- Finding network topology
- Applications of Bayesian networks

Example IA: High Blood Pressure

E=Yes
0.7

D=Healthy
0.25

Exercise

Diet

	H =Yes
D=Healthy	0.2
D=Unhealthy	0.85

	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

Heart Disease

Heartburn

	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

	BP=High
HD=Yes	0.85
HD=No	0.2

Blood Pressure

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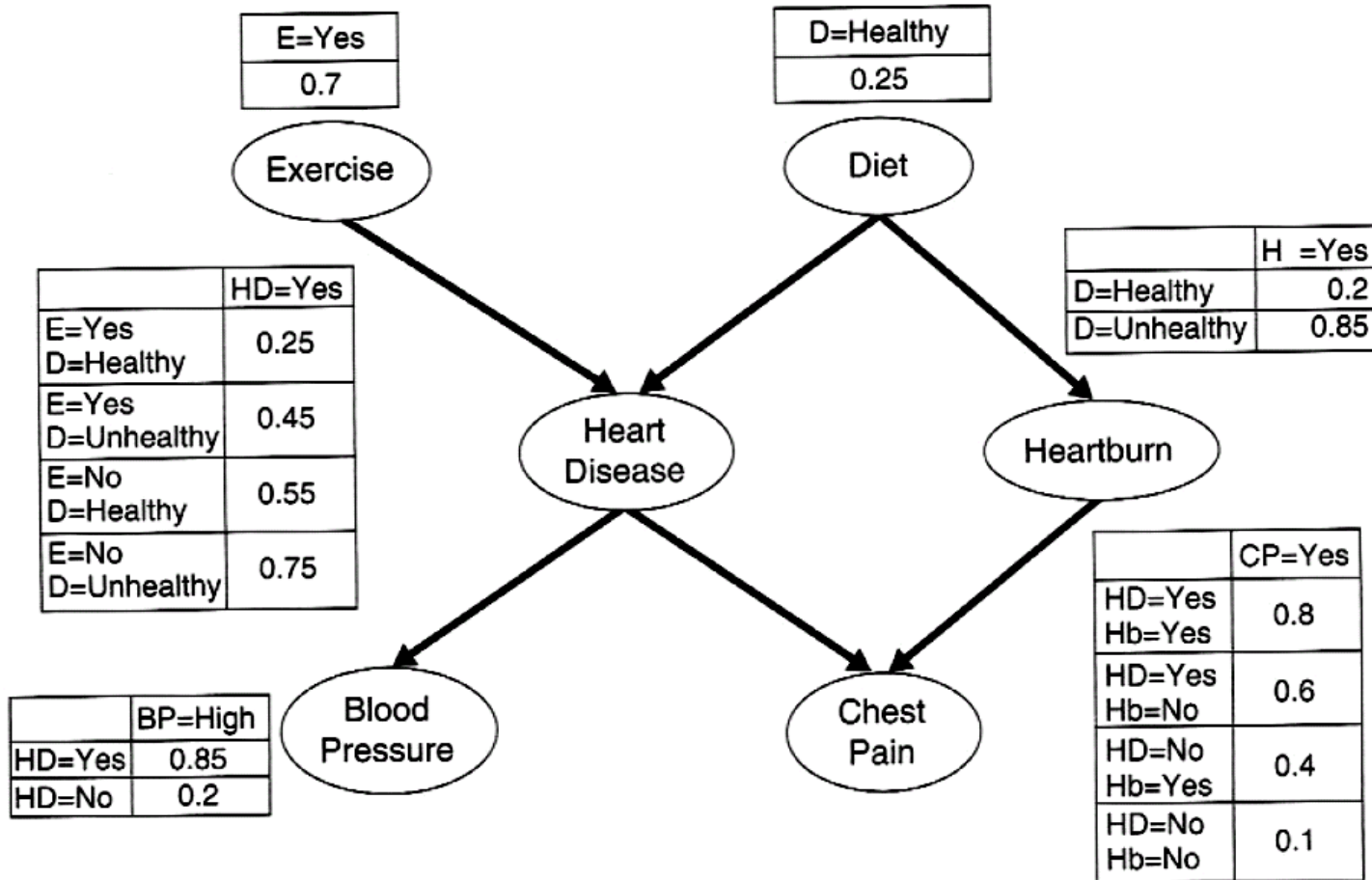
Once the right topology has been found. the probability table associated with each node is determined.

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

Example IA: 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?

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

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

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

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

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

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

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

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

$$= \alpha * 0.4165$$

=1

Example IA: High Blood Pressure

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

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

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

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

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

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

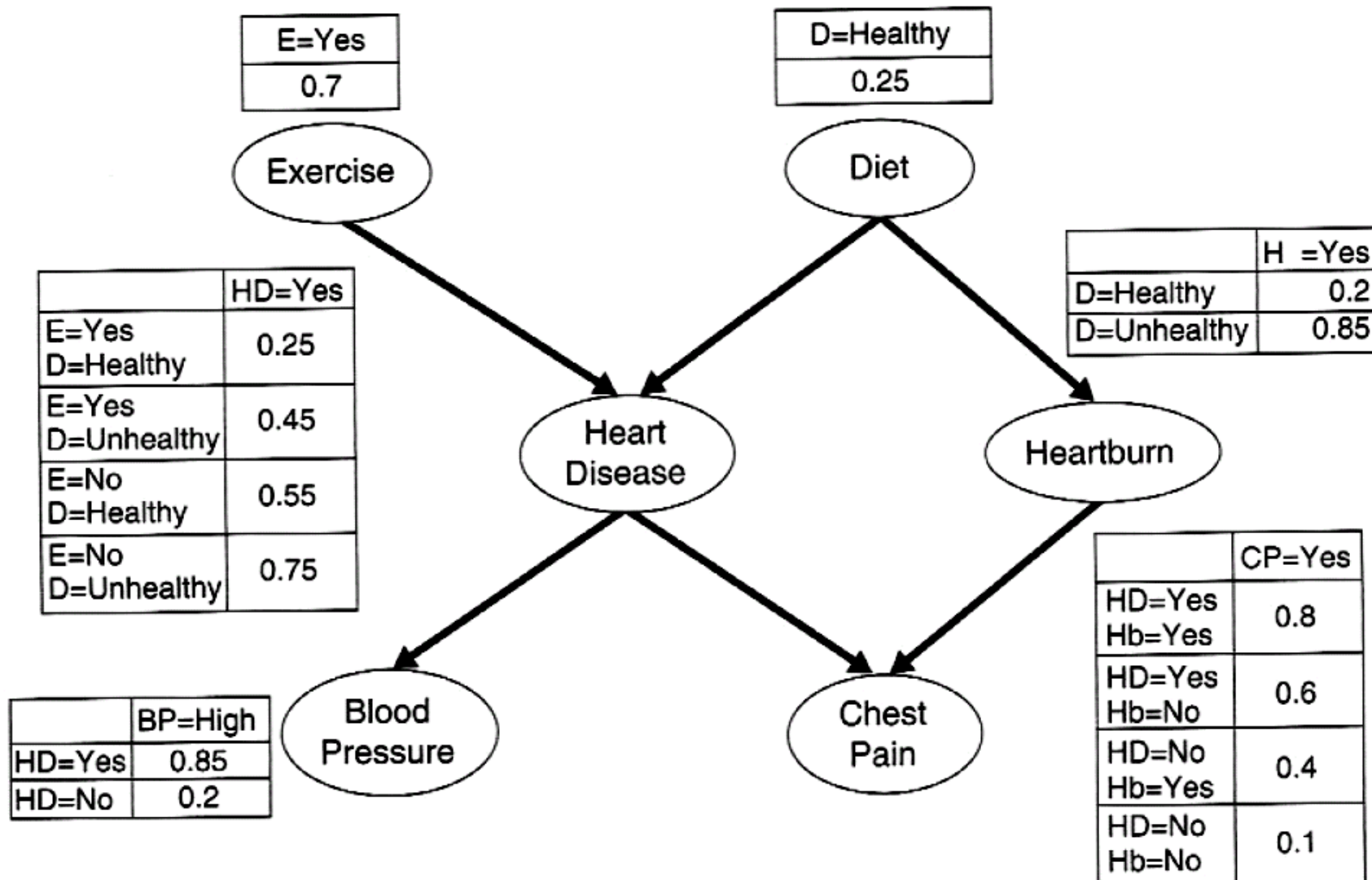# Example IA: High Blood Pressure (α)

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

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

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

Example IB:
High Blood Pressure, Healthy Diet, Regular Exercise

Example IB: Probability of heart disease

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

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

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

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

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

Example IB: Probability of not heart disease

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

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

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

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

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

Example I. High Blood Pressure, Healthy Diet, and Regular Exercise

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

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

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

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
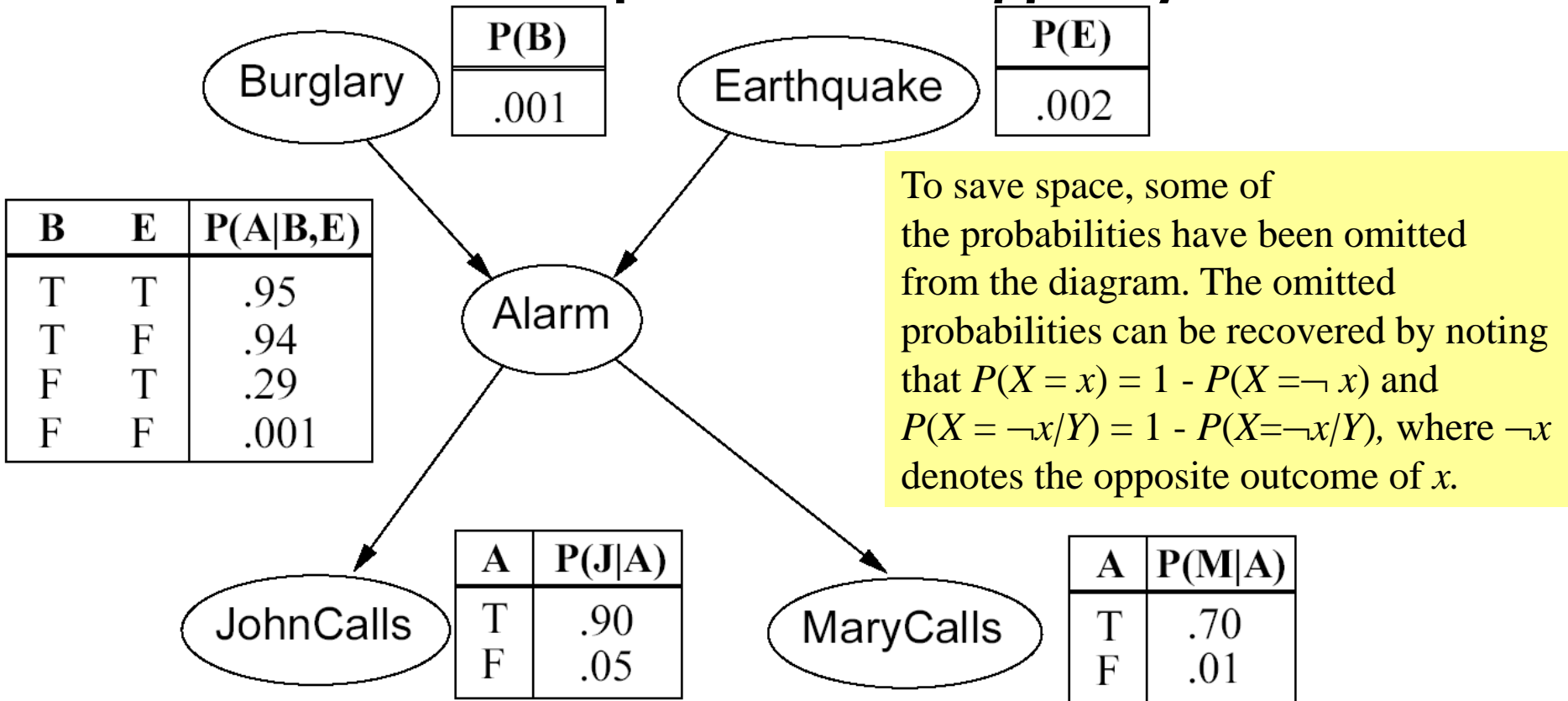
Outline

- Extending the Naïve Bayes network to more complex networks. Joint probability

- Bayesian Belief networks (BBN). Definition. Types

- Query on BBN: what nodes to include. Markov blanket

- Query on BBN: how to compute. Simple example. Some variables may be hidden.

▶ - Examples: II. Alarm and burglary

- Finding network topology

- Applications of Bayesian networks

Example II: Burglary

- I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's 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

Example II: Burglary

	P(B)
Burglary	.001

	P(E)
Earthquake	.002

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

Alarm

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.

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

JohnCalls

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

MaryCalls

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.

Example II: Classification

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

- For this we compute and compare the following:

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

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

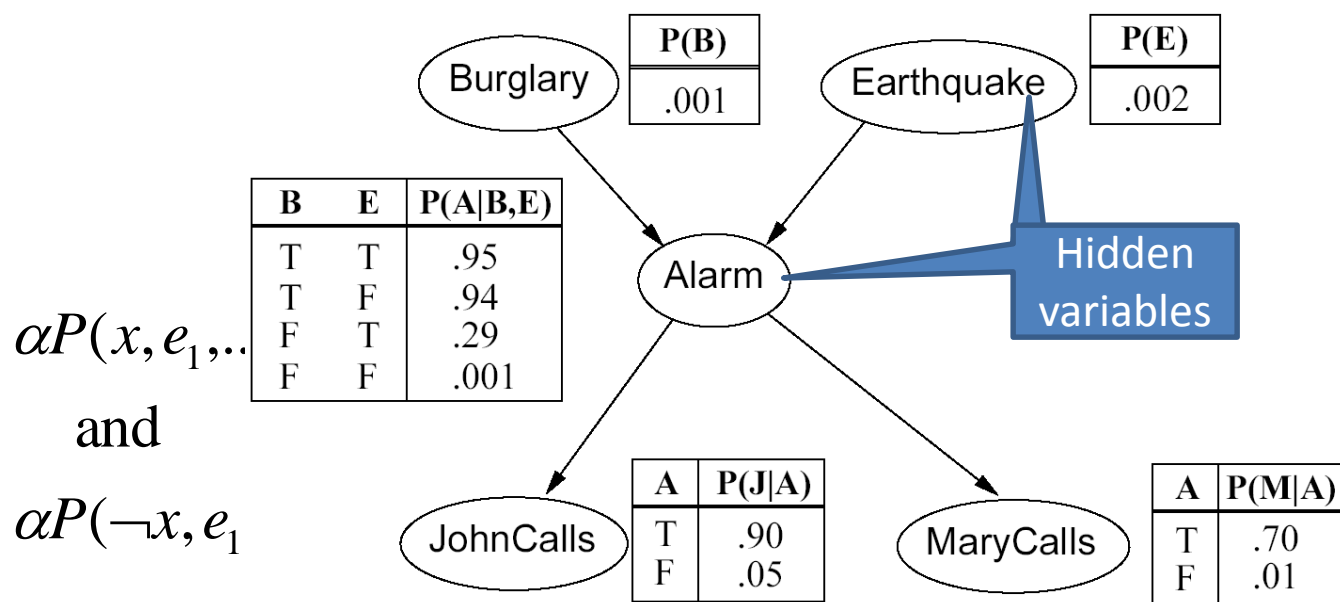- However, how do we compute:

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

and

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

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

Example II: Classification of burglary and earthquake

- Suppose, 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

	P(B)
Burglary	.001

	P(E)
Earthquake	.002

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

$\alpha P(x, e_1, ..$

and

$\alpha P(\neg x, e_1$

Alarm

Hidden variables

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

JohnCalls

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

MaryCalls

Example II: Inference by enumeration

Example: $P(burglary \mid johhcalls, marycalls)$? (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

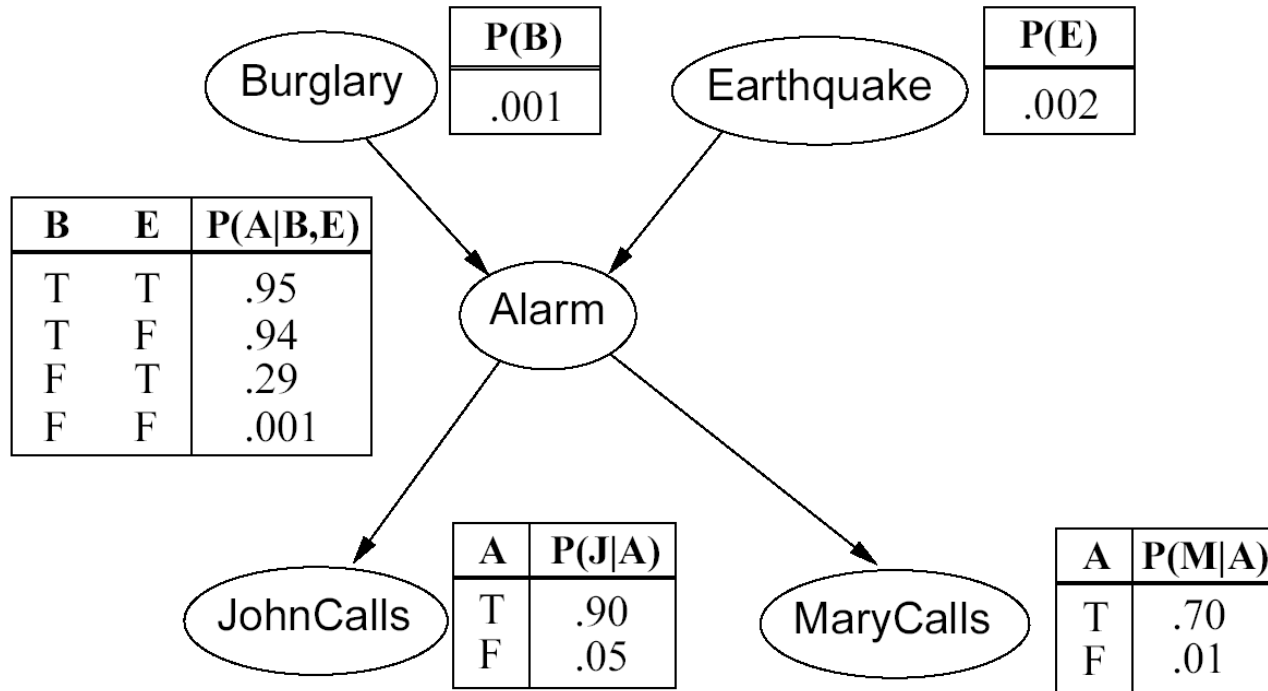Alarm rings, no earthquake

No alarm, no earthquake

In general:

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

and

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

where y1,...yk are hidden variables

Example II: Numerically…

P(B)
.001

P(E)
.002

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

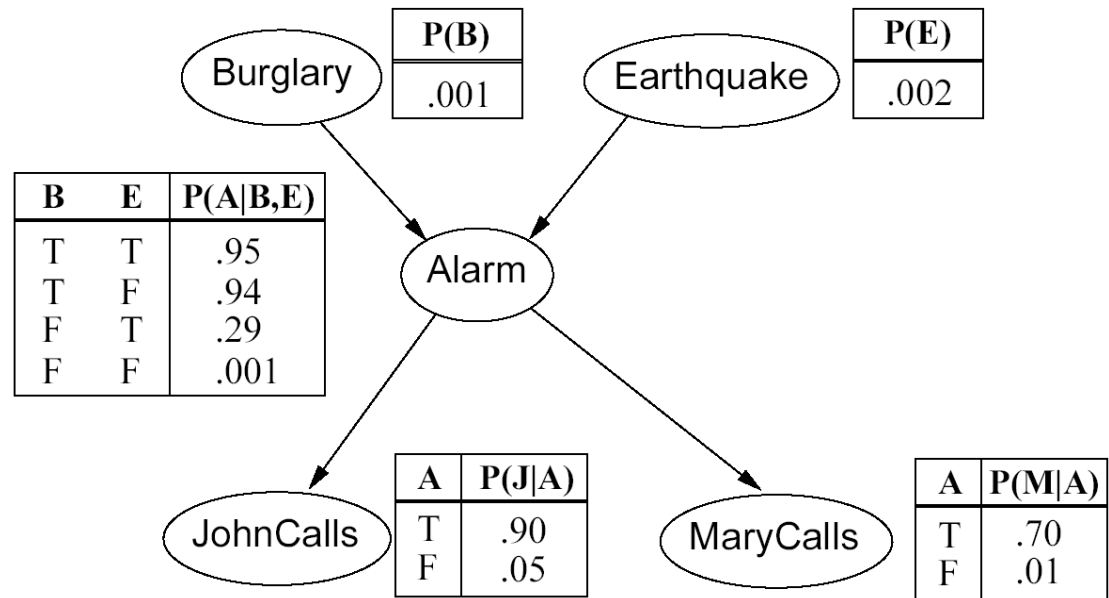| A	P(J|A)
T	.90
F	.05

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

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

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

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

Example II: $P(b \mid j,m)$

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

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

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

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

$= \alpha * .001*(.9*.7*(.95*.002 + .94*.998) +.05*.01*(.05*.002 + .71*.998) \,)$
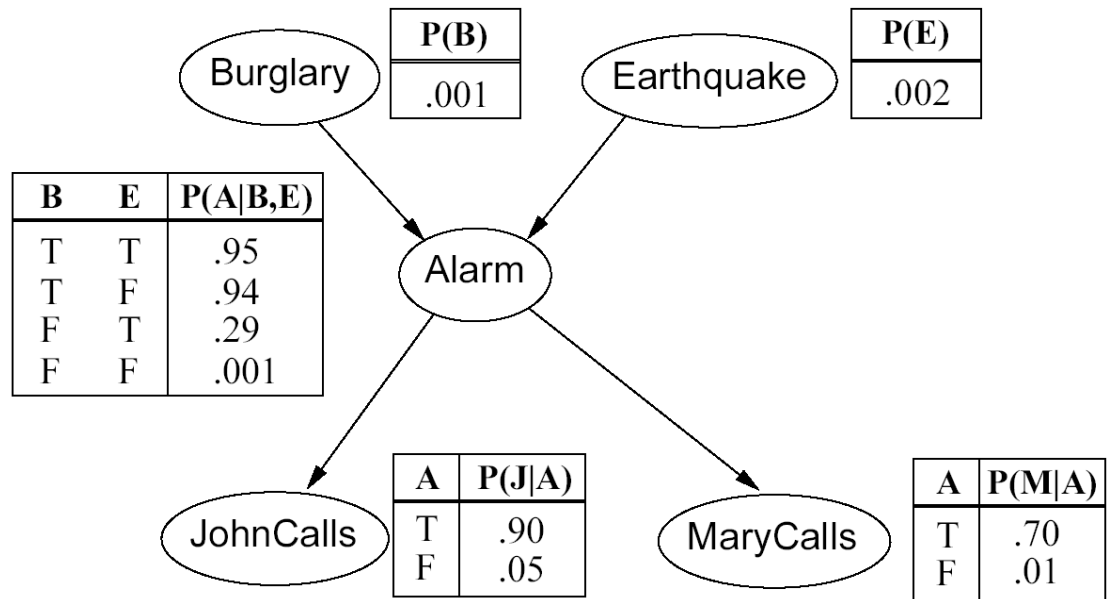
$= \alpha * .00059$

	P(B)
Burglary	.001

	P(E)
Earthquake	.002

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

Alarm

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

JohnCalls

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

MaryCalls

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

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

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

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

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

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

$= \alpha * .0015$

$\alpha\ = 1/(.00059 + .0015)$

$=\ 478.5$

$P(b \mid j,m) = 478.5 * .00059$

$=.28$

$P(\neg b \mid j,m) = 478.5 * .0015$

$=.72$

	P(B)
Burglary	.001

	P(E)
Earthquake	.002

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

Alarm

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

JohnCalls

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

MaryCalls

Outline

- Extending the Naïve Bayes network to more complex networks. Joint probability
- Bayesian Belief networks (BBN). Definition. Types
- Query on BBN: what nodes to include. Markov blanket
- Query on BBN: how to compute. Simple example. Some variables may be hidden.
- ▶ Examples: III Mystery
- Finding network topology
- Applications of Bayesian networks

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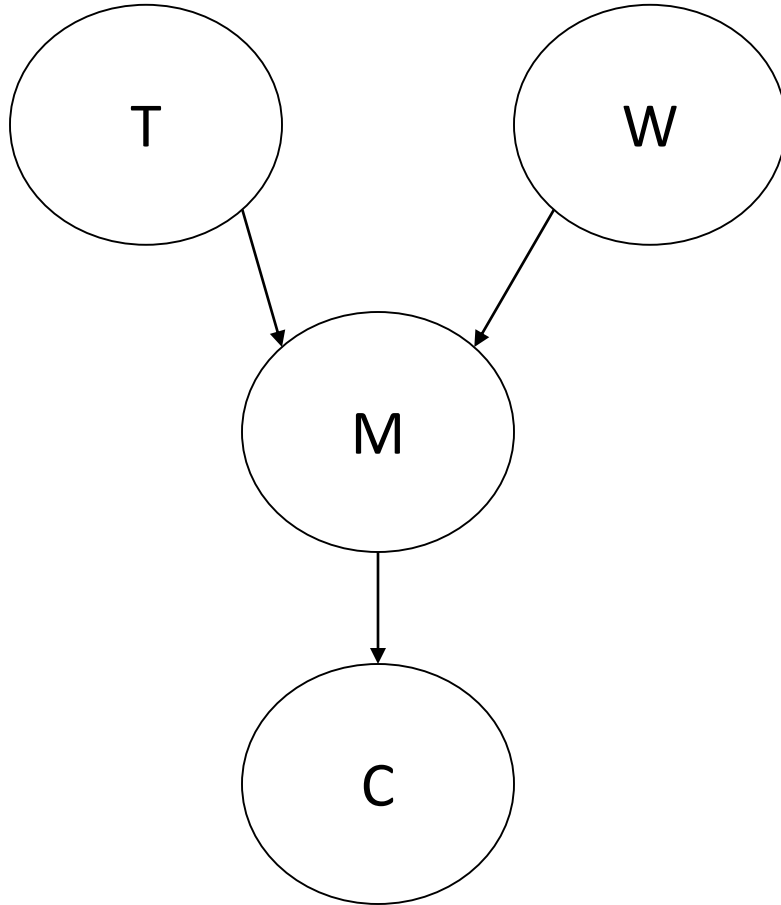
Example III: Network topology
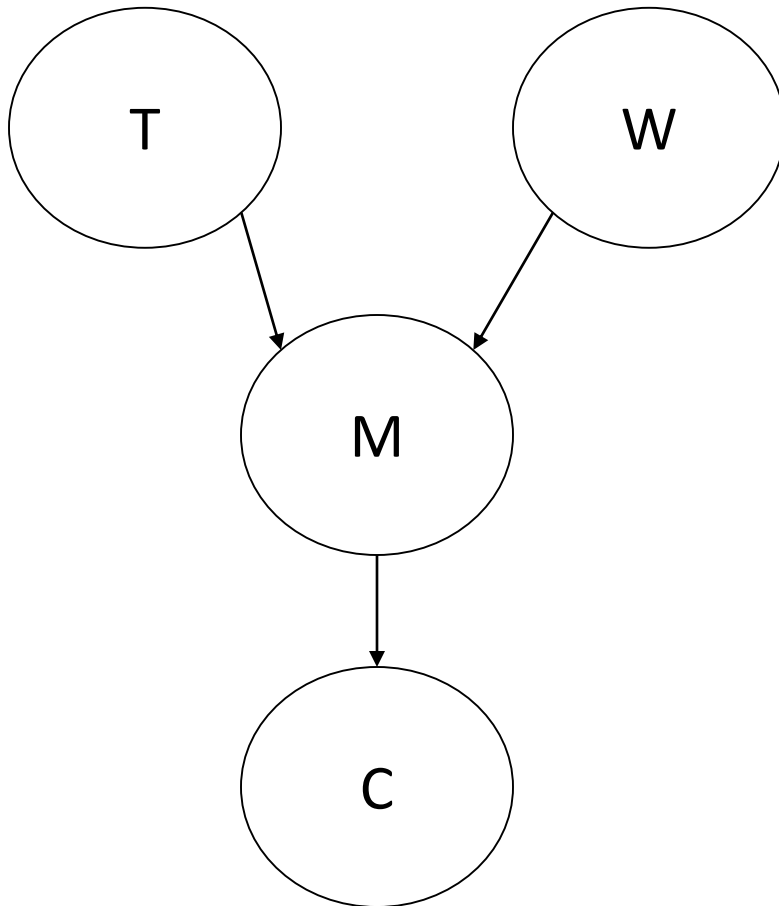
- T – time of day when the murder was committed: evening (e) or night (n)
- W – crime weapon: vacuum (v) or stick (s)
- M- murderer: maid (m) or butler (b)
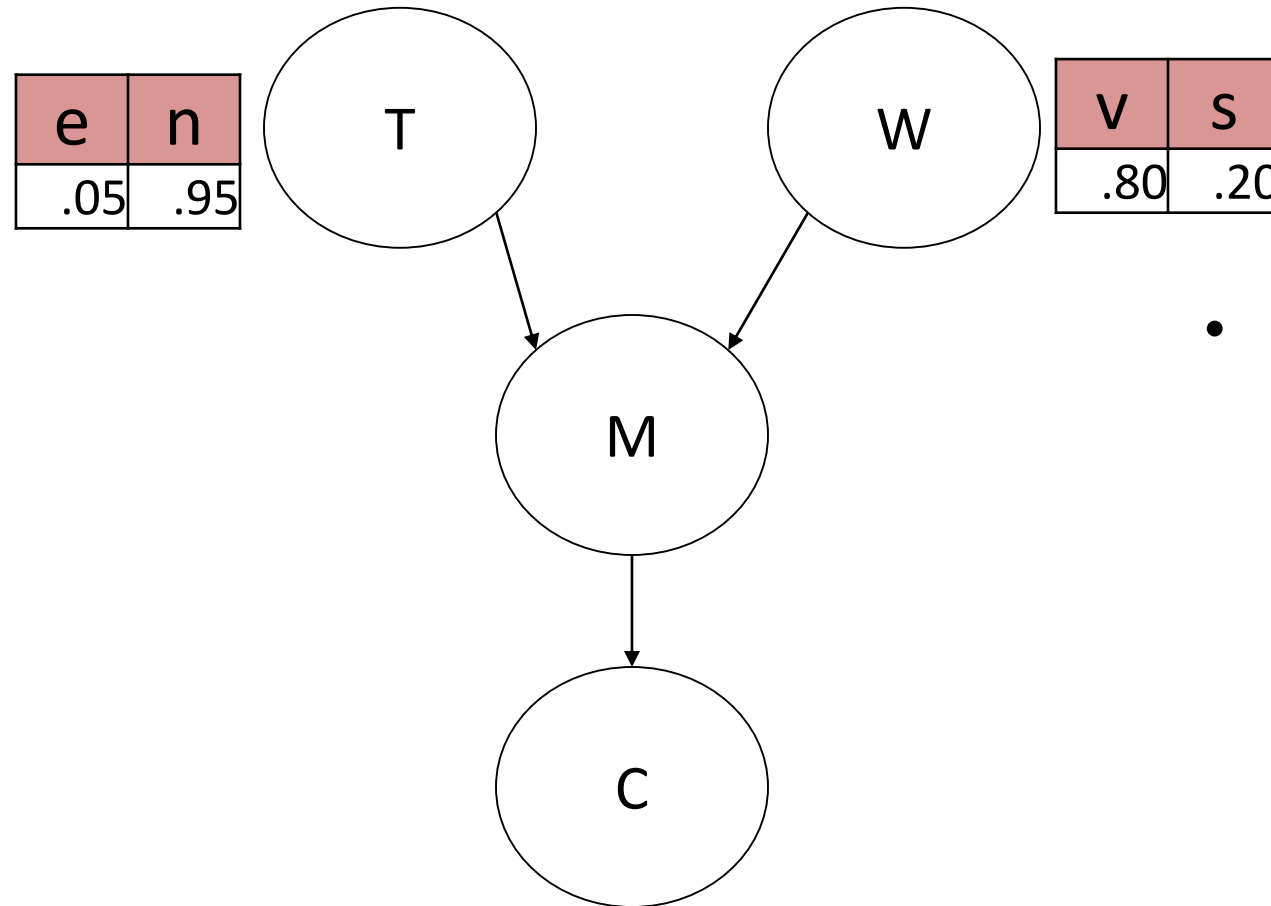- C – will confess: yes or no
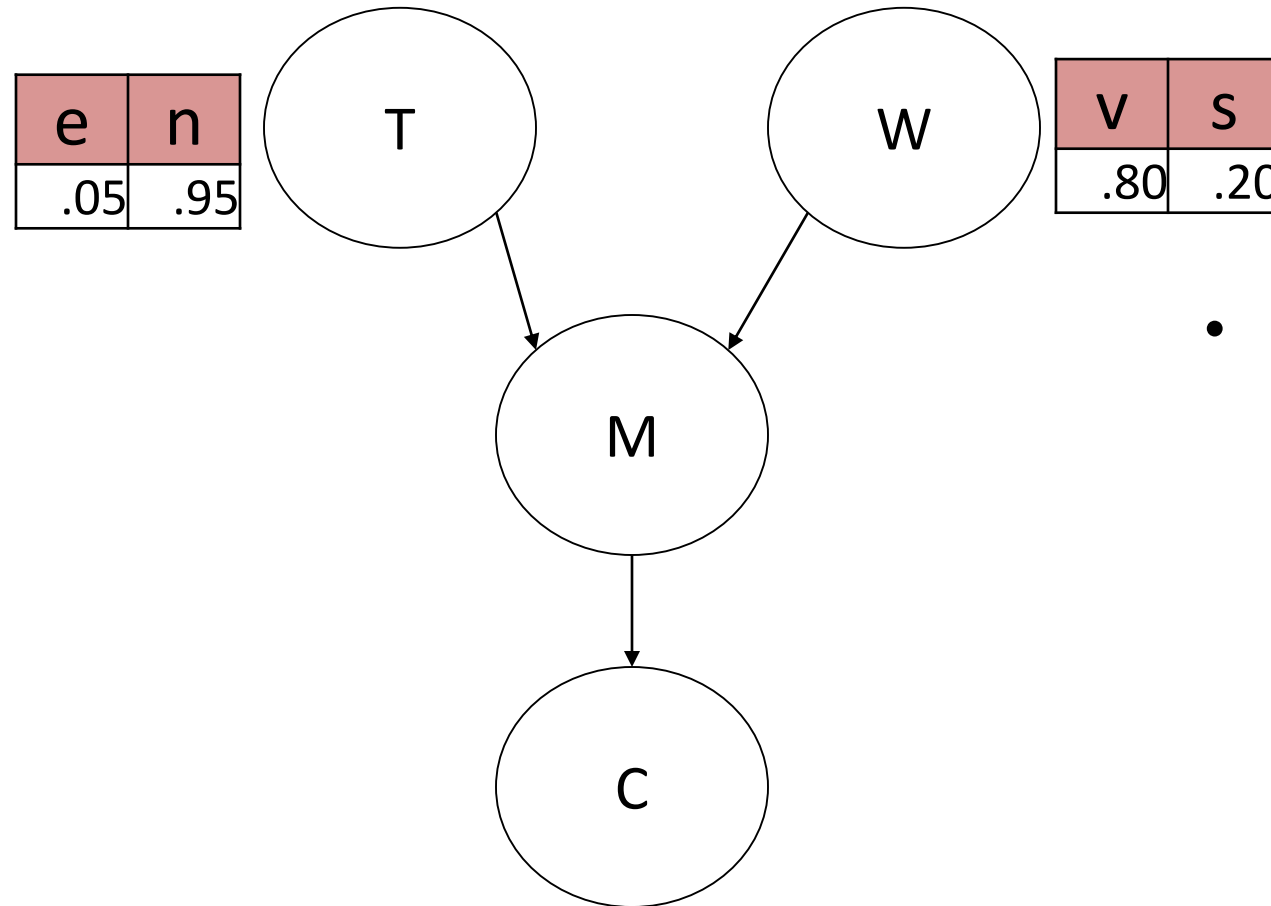
Example III: CPT for Time

e	n
.05	.95

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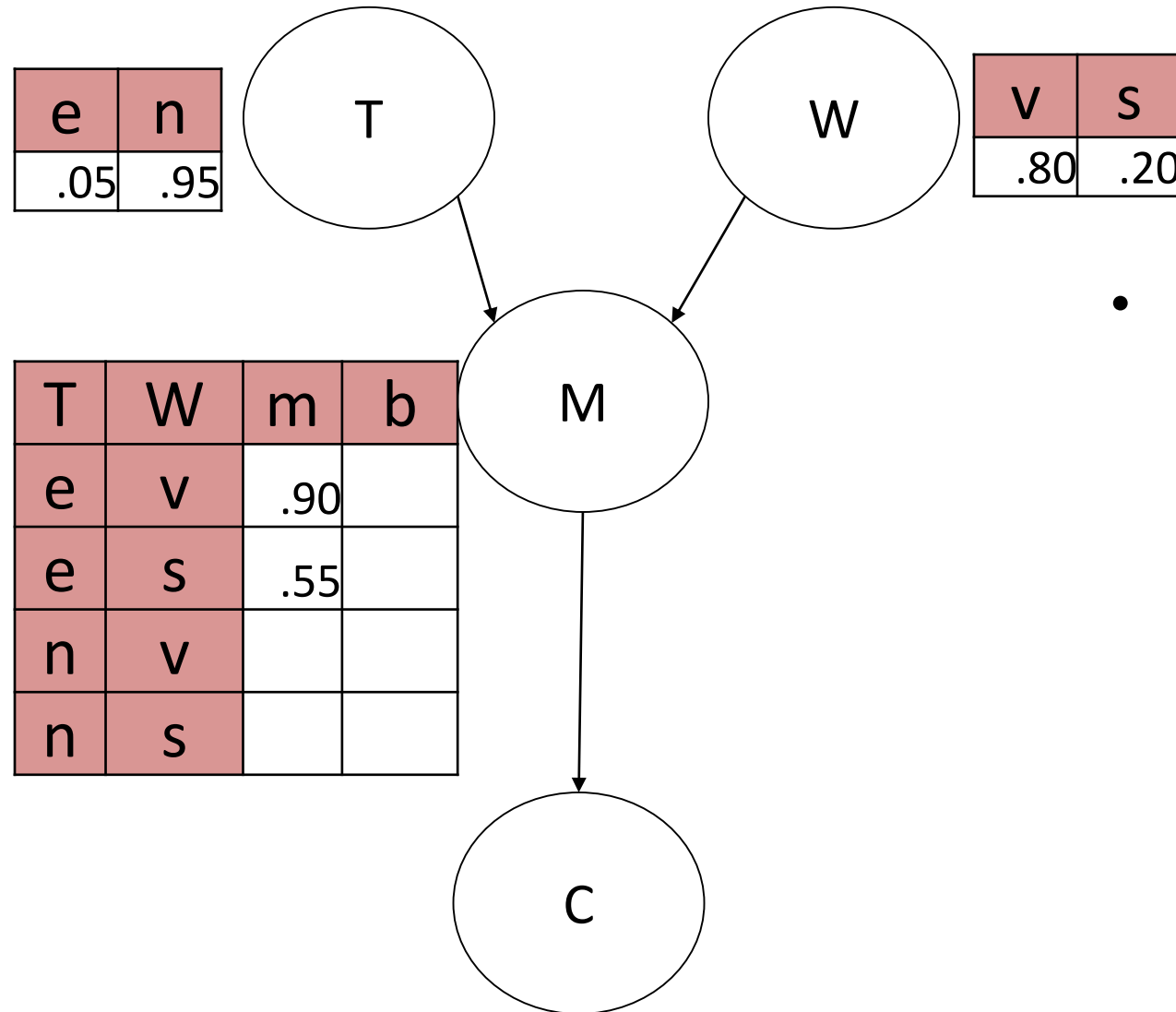
Example III: CPT for Weapon

e	n
.05	.95

T

W

v	s
.80	.20

M

C

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

Example III: CPT Murderer

e	n
.05	.95

v	s
.80	.20

T

W

M

C

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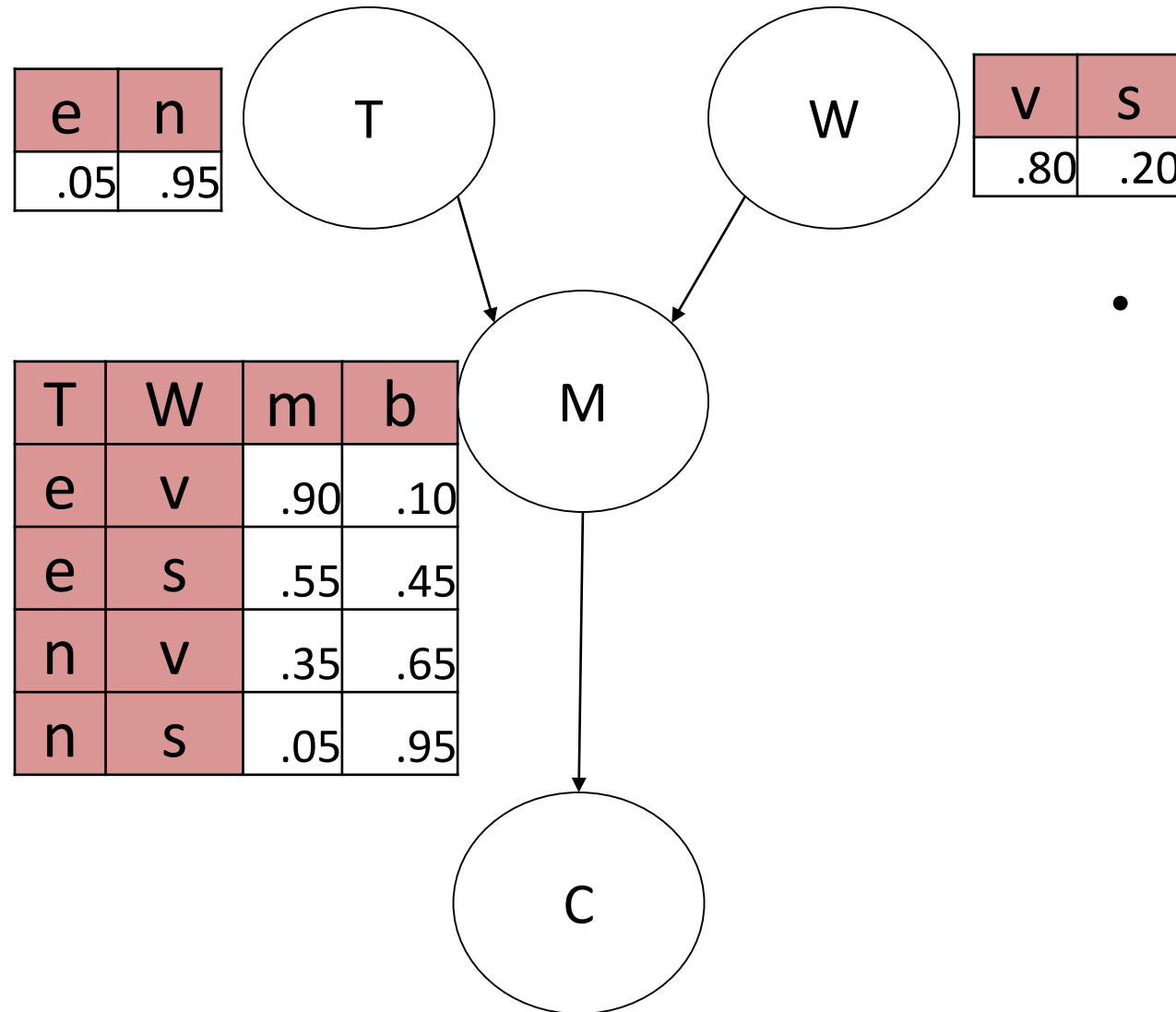
Example III: CPT Murderer

e	n
.05	.95

v	s
.80	.20

T

W

M

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

C

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

Example III: CPT Murderer

e	n
.05	.95

T

W

v	s
.80	.20

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

M

C

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

Example III: CPT Murderer

e	n
.05	.95

v	s
.80	.20

(T)

(W)

(M)

(c)

T	W	m	b
e	v	.90	.10
e	s	.55	.45
n	v	.35	.65
n	s	.05	.95
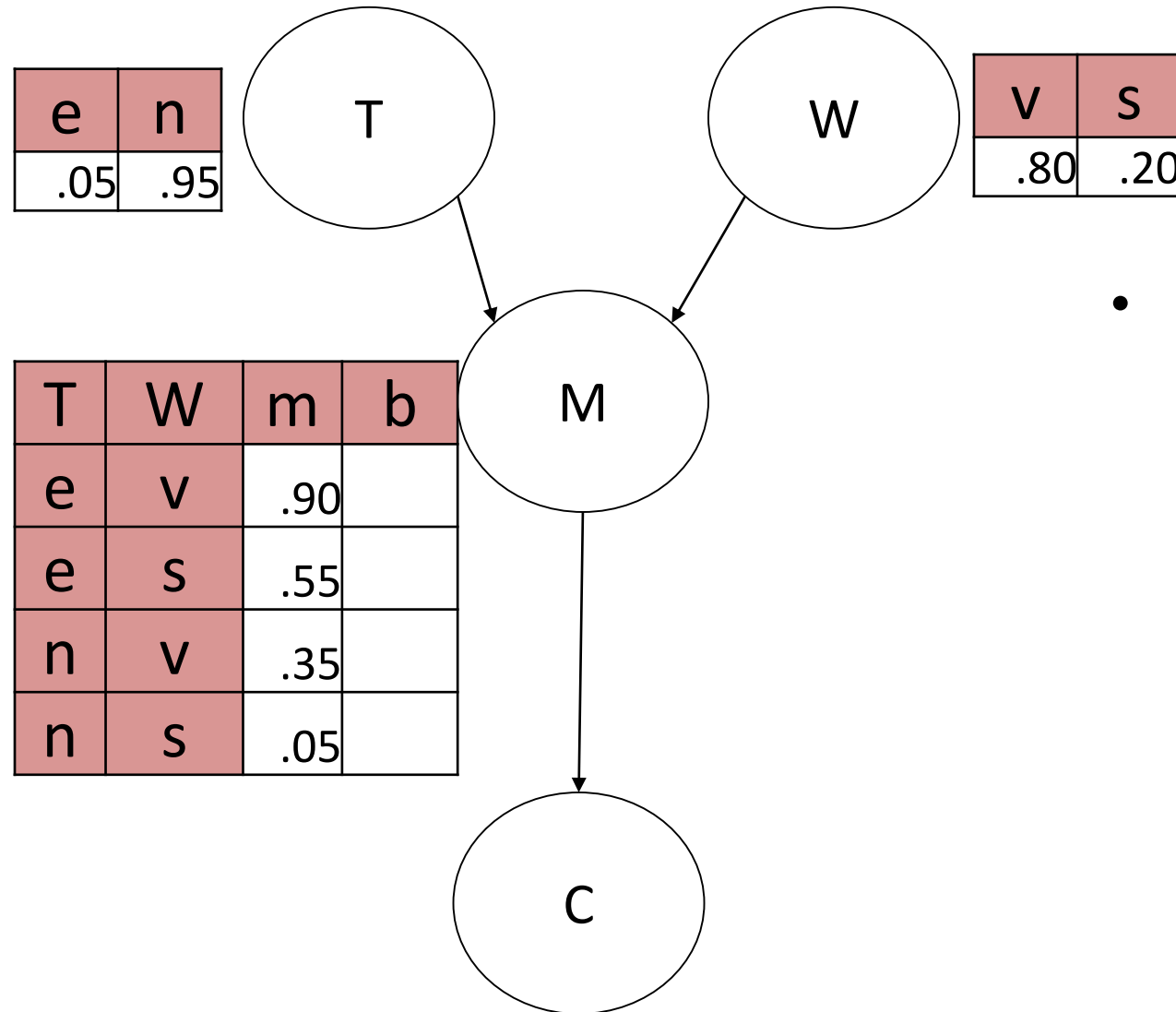
- And the butler takes the rest (mutually exclusive events)

Example III: CPT Confession

e	n
.05	.95

v	s
.80	.20

T

W

M

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

C

M	yes	no
m	.90	.10
b	.40	.60
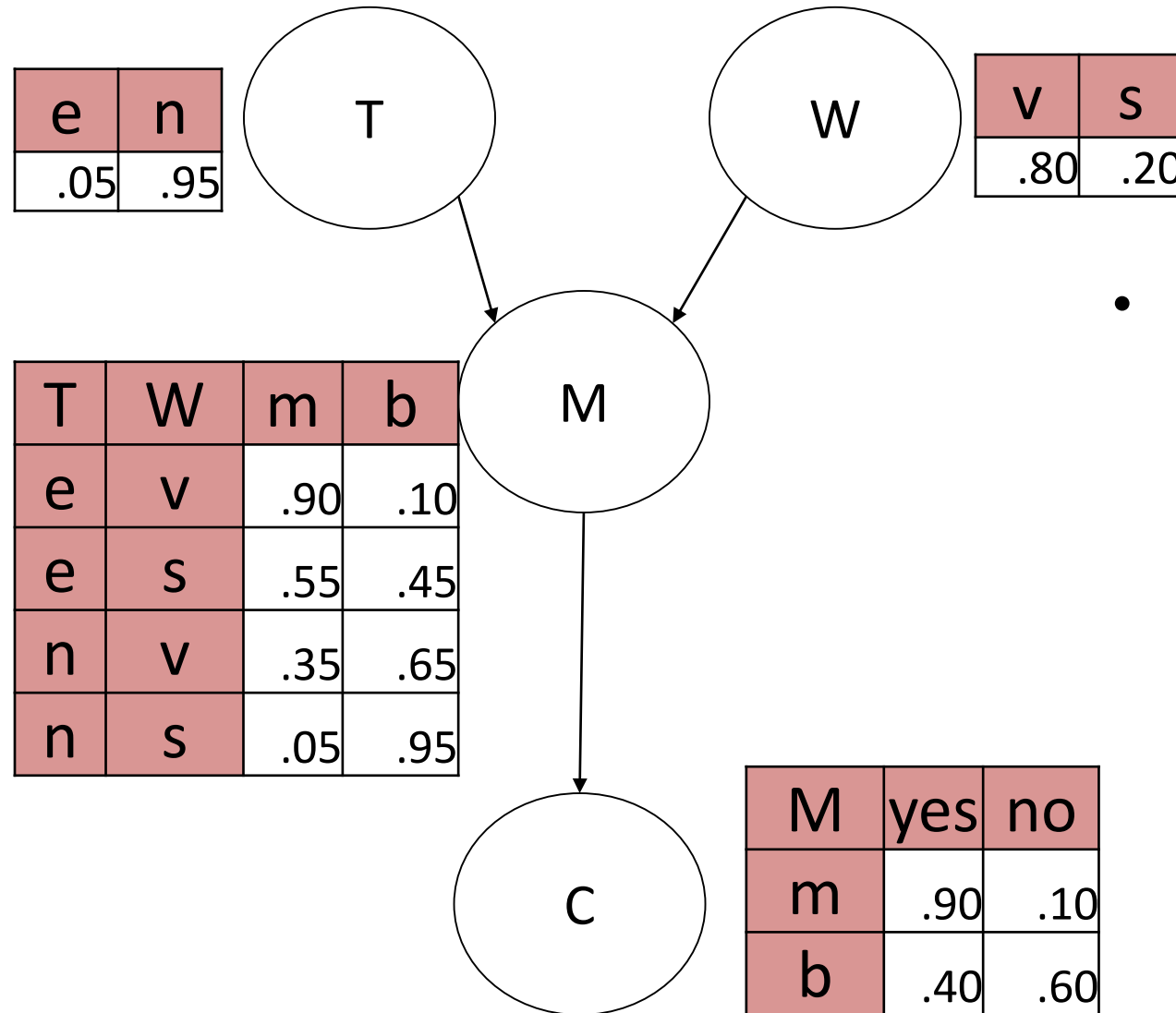
- The maid has a very strong conscience and she will eventually confess if she committed the murder. The butler is quite opposite
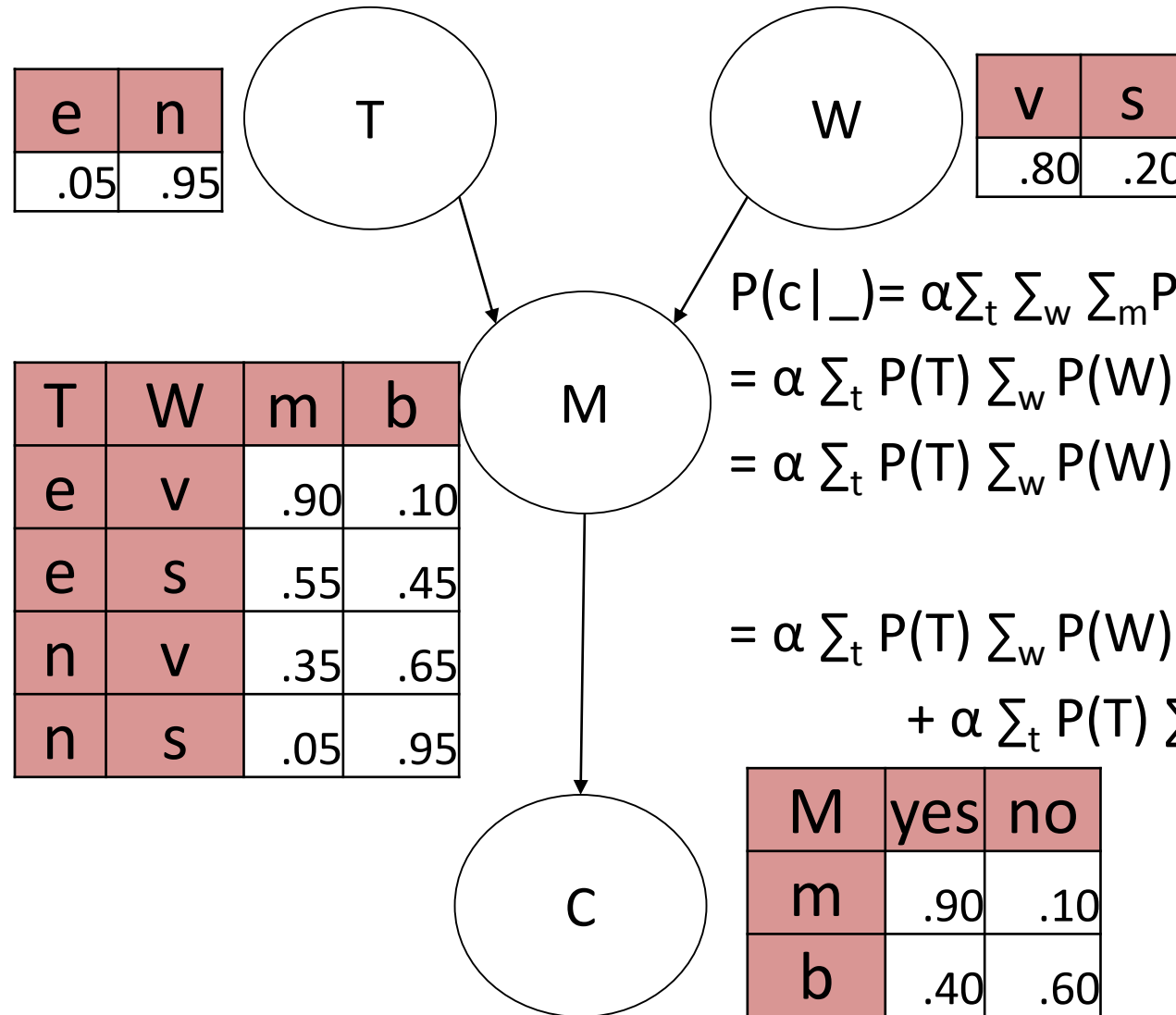
Example III: The probability of confession (nothing is given)

e	n
.05	.95

T

W

v	s
.80	.20

M

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

C

M	yes	no
m	.90	.10
b	.40	.60
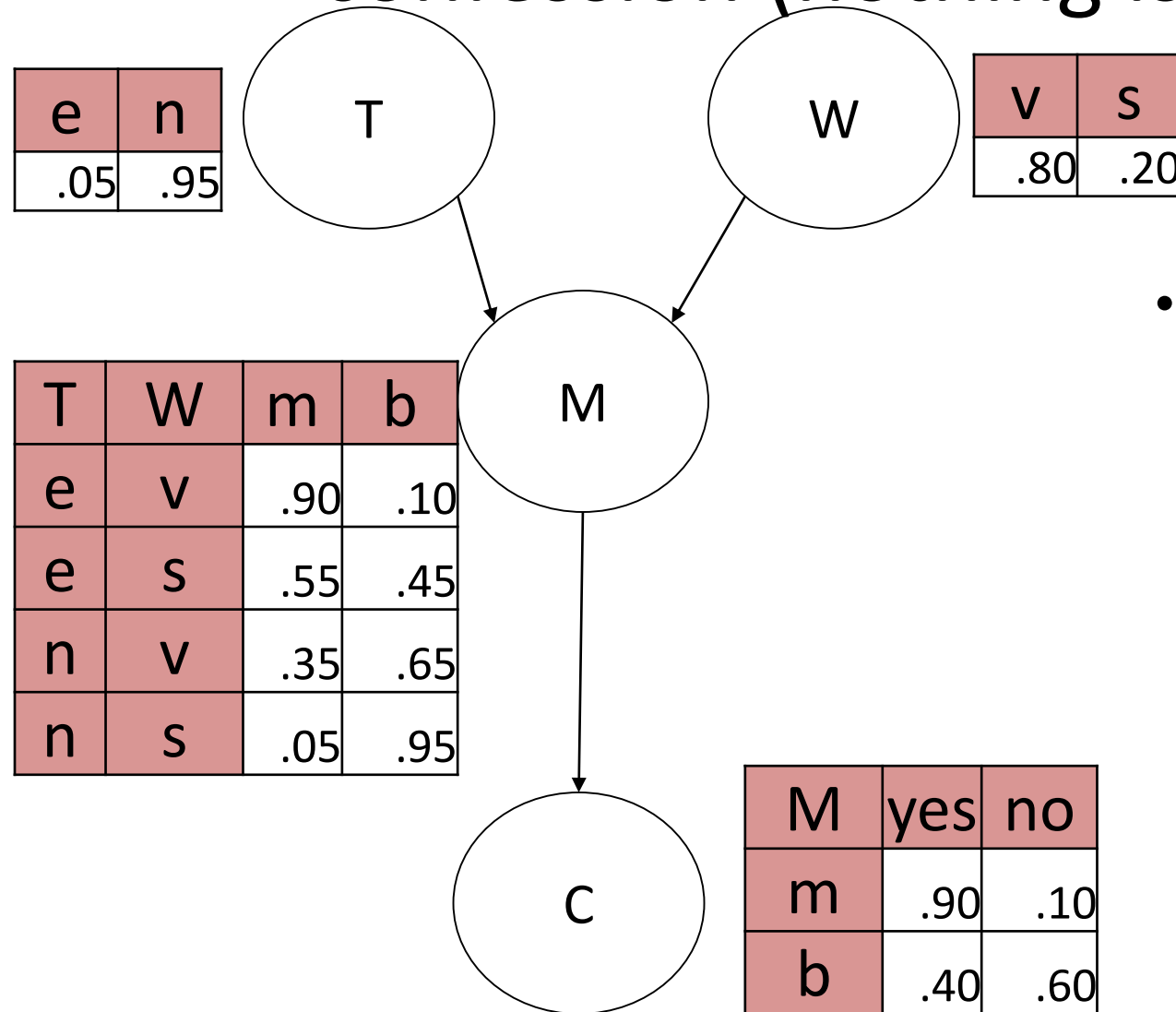
- The detective can evaluate the probability that the murderer will confess without having any real evidence
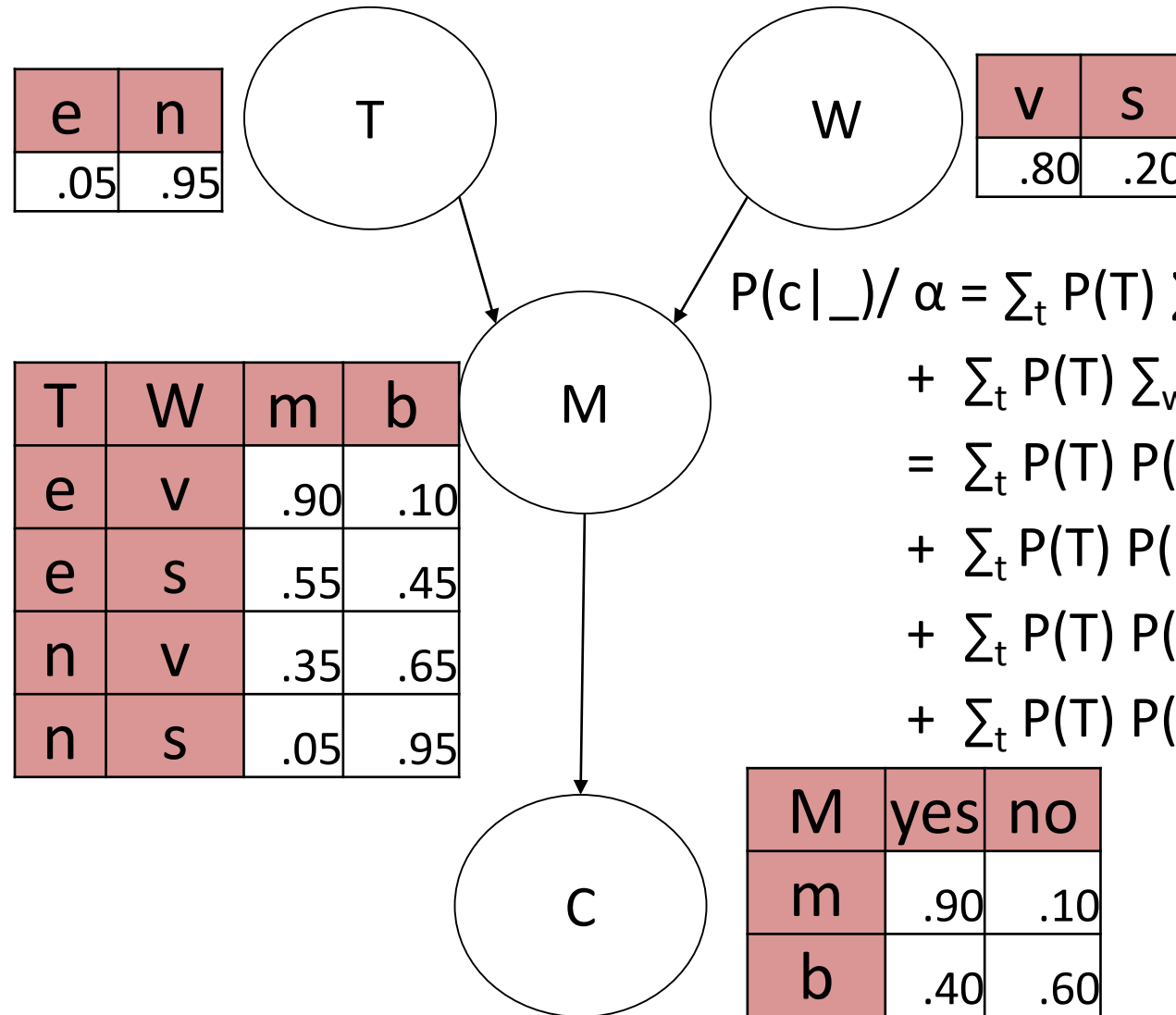
Example III:
The probability of confession

e	n
.05	.95

T

W

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

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

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

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

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

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

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

C

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

Example III:
The probability of confession

e	n
.05	.95

T

W

v	s
.80	.20

M

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

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

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

$= \sum_t P(T) P(v) P(m|T,v) *0.90$

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

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

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

C

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

Example III:
The probability of confession

e	n
.05	.95

T

W

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

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

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

$= \sum_t P(T) *0.80 * P(m|T,v) *0.90$

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

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

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

C

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

Example III:
The probability of confession

e	n
.05	.95

T

W

v	s
.80	.20

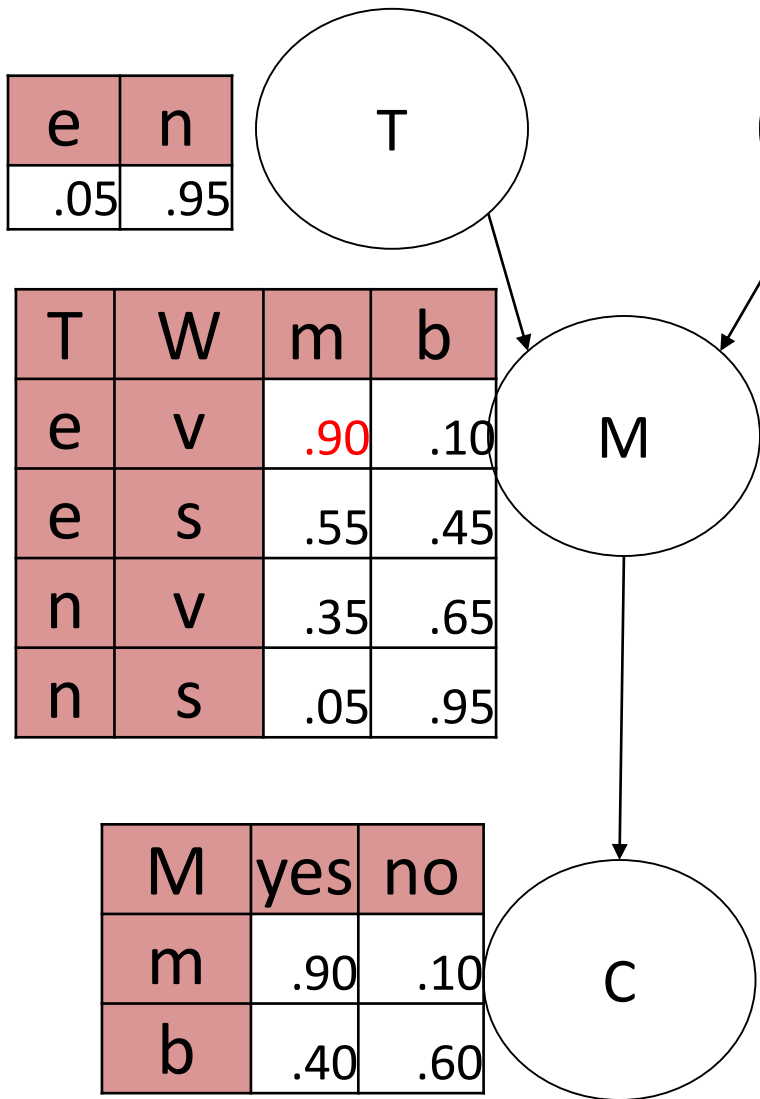| T	W	m	b
e	v	.90	.10
e	s	.55	.45
n	v	.35	.65
n	s	.05	.95

M

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

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

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

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

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

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

C

Example III:
The probability of confession

e	n
.05	.95

(T)　　　(W)

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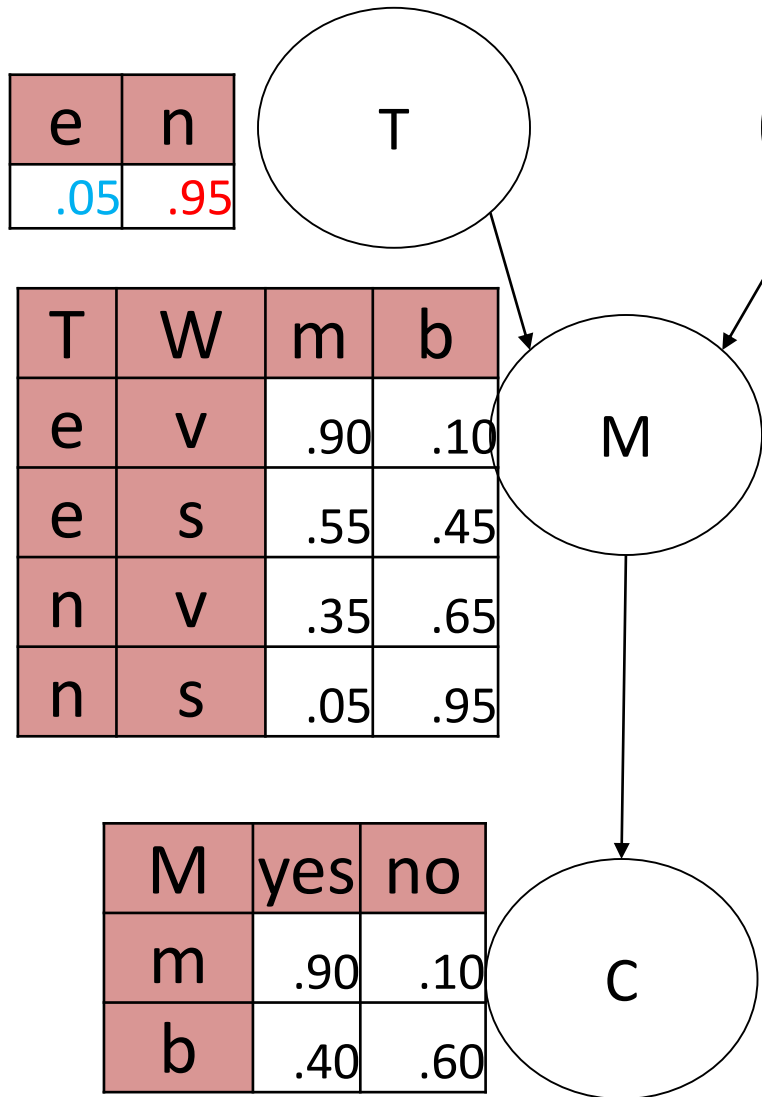
$P(c|\_)/\ \alpha = \sum_t P(T) *0.80 * P(m|T,v) *0.90$

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

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

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

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

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

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

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

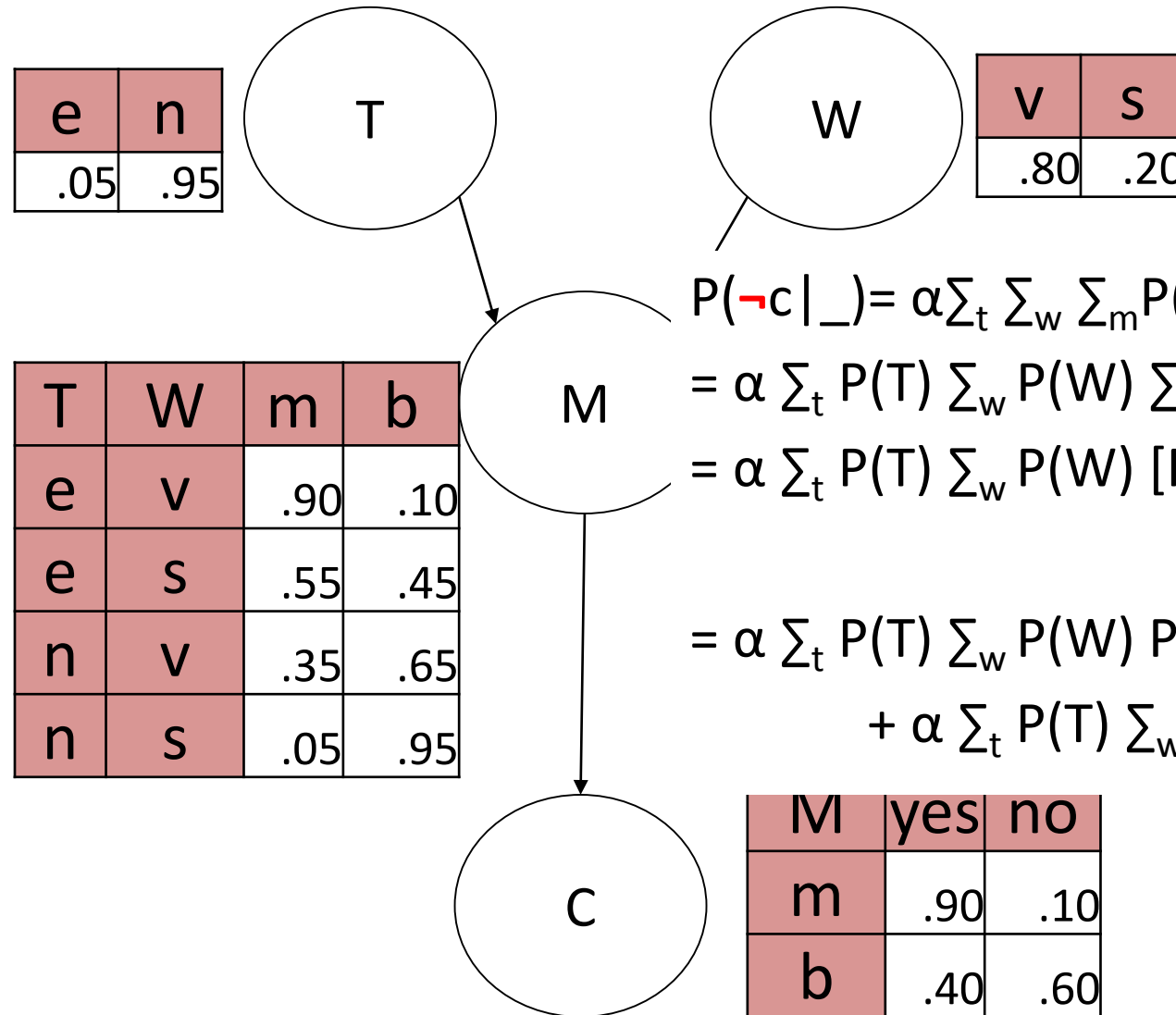| M	yes	no
m	.90	.10
b	.40	.60

(C)

Example III:
The probability of confession

e	n
.05	.95

T

W

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

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

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

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

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

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

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

C

Example III:
The probability of non-confession

e	n
.05	.95

T

W

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

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

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

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

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

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

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

c

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

Example 3:
The probability of non-confession

e	n
.05	.95

v	s
.80	.20

T   W   M   C

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

$$P(\neg c \mid \_)/\alpha = \sum_t P(T) \sum_w P(W) P(m \mid T,W) *0.90$$
$$+ \sum_t P(T) \sum_w P(W) P(b \mid T,W) *0.40$$
$$= \sum_t P(T) P(v) P(m \mid T,v) *0.10$$
$$+ \sum_t P(T) P(s) P(m \mid T,s) *0.10$$
$$+ \sum_t P(T) P(v) P(b \mid T,v) *0.60$$
$$+ \sum_t P(T) P(s) P(b \mid T,s) *0.60$$

Example 3:
The probability of non-confession

e	n
.05	.95

T

W

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

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

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

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

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

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

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

c

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

Example 3:
The probability of confession

e	n
.05	.95

 T

 W

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

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

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

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

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

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

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

 C

Example 3:
The probability of non-confession

e	n
.05	.95

T (node)

W (node)

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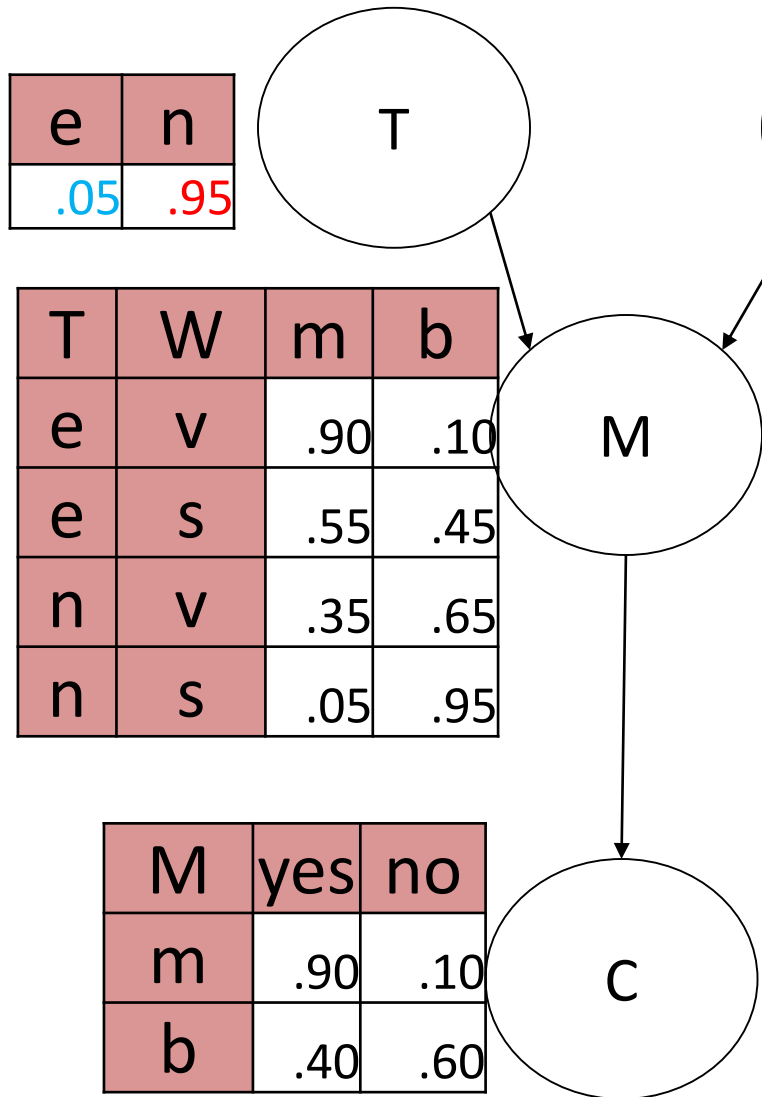
$P(\neg c | \_)/\alpha = \sum_t P(T) *0.80 * P(m|T,v) *0.90$

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

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

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

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

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

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

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

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

C (node)

Example 3:
The probability of non-confession

e	n
.05	.95

v	s
.80	.20

T   W   M   C (nodes)

T	W	m	b
e	v	.90	.10
e	s	.55	.45
n	v	.35	.65
n	s	.05	.95
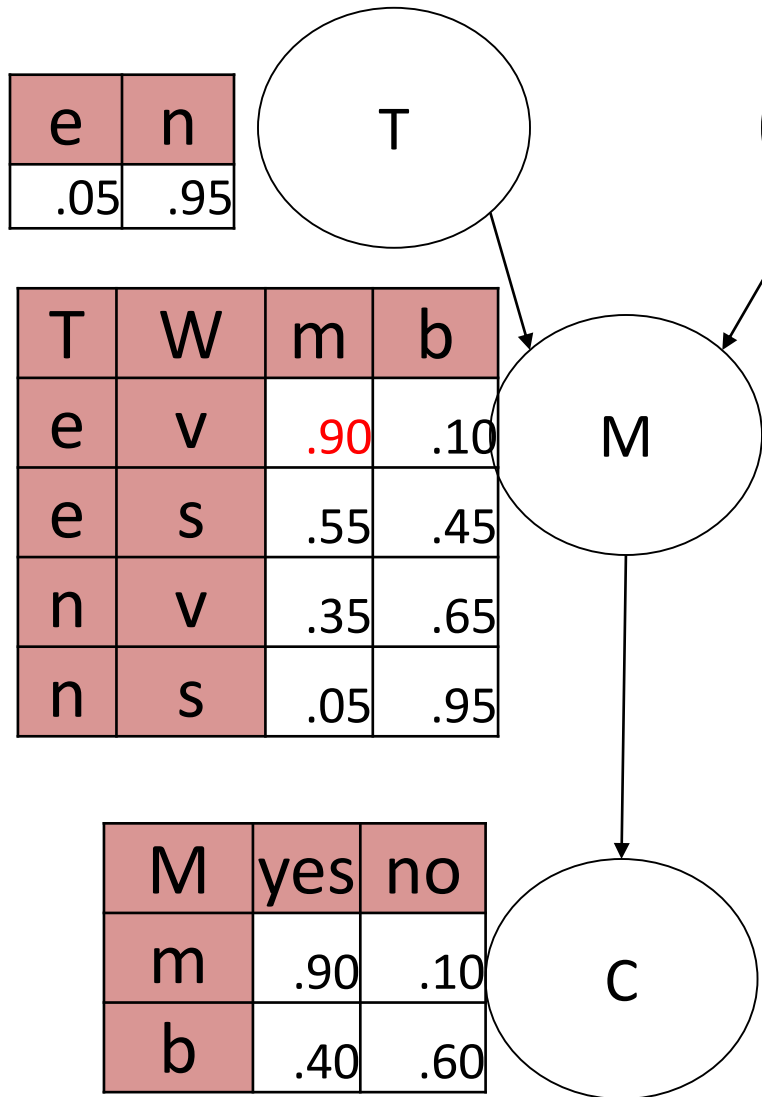
$P(c|\_)/ \alpha = \sum_t P(T) *0.80 * P(m|T,v) *0.90$

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

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

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

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

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

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

Outline

- Extending the Naïve Bayes network to more complex networks. Joint probability
- Bayesian Belief networks (BBN). Definition. Types
- Query on BBN: what nodes to include. Markov blanket
- Query on BBN: how to compute. Simple example. Some variables may be hidden.
- Examples
- Finding network topology
- Applications of Bayesian networks

Finding topology of Bayesian networks

1. Done by human expert
2. Automated learning of network from data

Constructing Bayesian networks by human experts

1. Choose an ordering of variables $X_1, \dots , X_n$
2. For $i = 1$ to n
   - add X_i to the network
   - select parents from $X_1, \dots , X_{i-1}$ such that
   $$\mathbf{P}(X_i \mid Parents(X_i)) = \mathbf{P}(X_i \mid X_1, \dots X_{i-1})$$

This choice of parents guarantees:
$$\mathbf{P}(X_1, \dots , X_n) = \prod_{i=1} \mathbf{P}(X_i \mid X_1, \dots , X_{i-1}) \quad \text{(chain rule)}$$
$$= \prod_{i=1} \mathbf{P}(X_i \mid Parents(X_i)) \quad \text{(by construction)}$$

- Choosing the parents from $X_1, \dots , X_{i-1}$ is done by domain human experts.

Expert construction: Example

- The ordering of variables is very important.
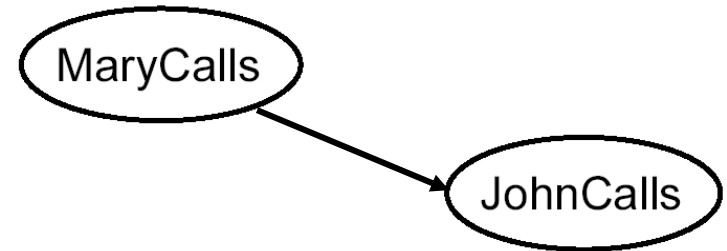- E.g. suppose we choose the ordering *M, J, A, B, E*

Adding MaryCalls: No parents

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

Is P(John calling) independent of P(Mary calling)?

Clearly not, since, on any given day, if Mary called, then the probability that John called is much better than the background probability that he called.

So, we add a link from MaryCalls to JohnCalls.

Expert construction: Example
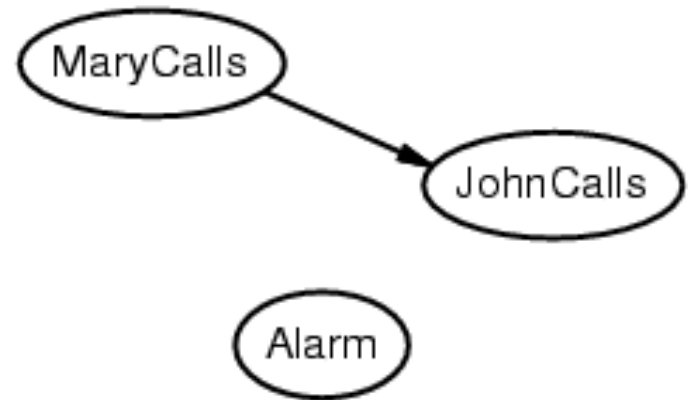
- Suppose we choose the ordering
  M, J, A, B, E

Adding the A (Alarm) node: Is

$P(A \mid J, M) = P(A \mid J)$?

$P(A \mid J, M) = P(A)$?

No.

Clearly, if both call, it's more likely that the alarm has gone off that if just one or neither call, so we need both MaryCalls and JohnCalls as parents.

Expert construction: Example

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

Adding B (Burglary) node: Is

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

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

Yes for the first. **No** for the second.

If we know the alarm state, then the call from John or Mary might give us information about the phone ringing or Mary's music, but not about burglary.

So, we need just Alarm as parent.
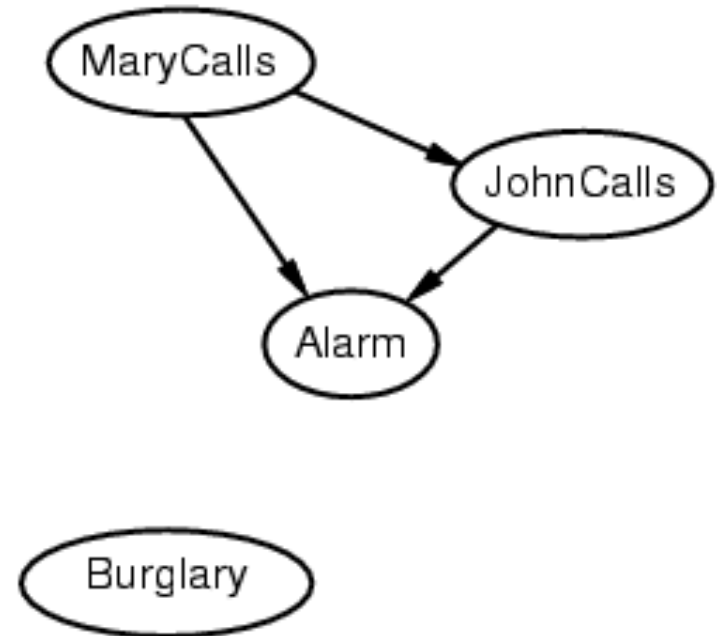
Expert construction: Example

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

- Adding E (Earthquake) node: Is

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

No for the first. **Yes** for the second.

If the alarm is on, it is more likely that there has been an earthquake.

But if we know there has been a burglary, then that explains the alarm, and the probability of an earthquake would be only slightly above normal.

Hence we need both Alarm and Burglary as parents.
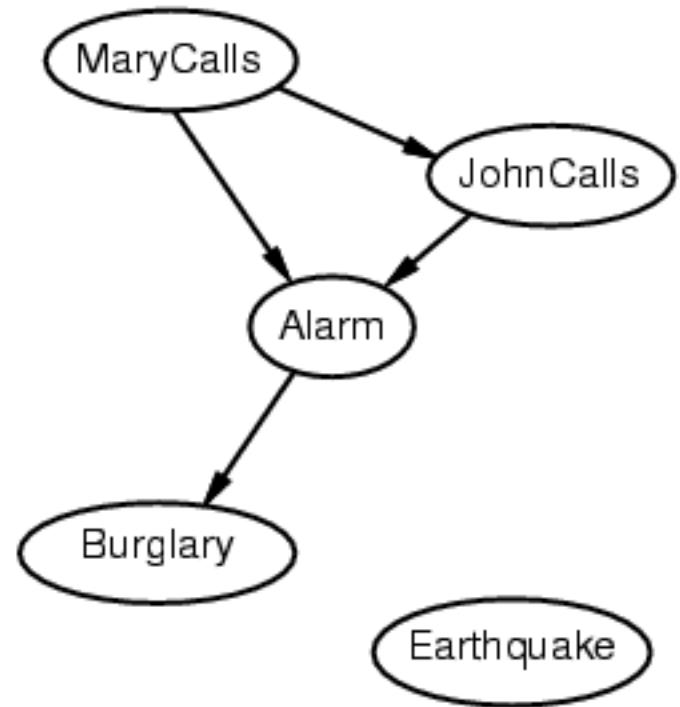
Expert construction: Example
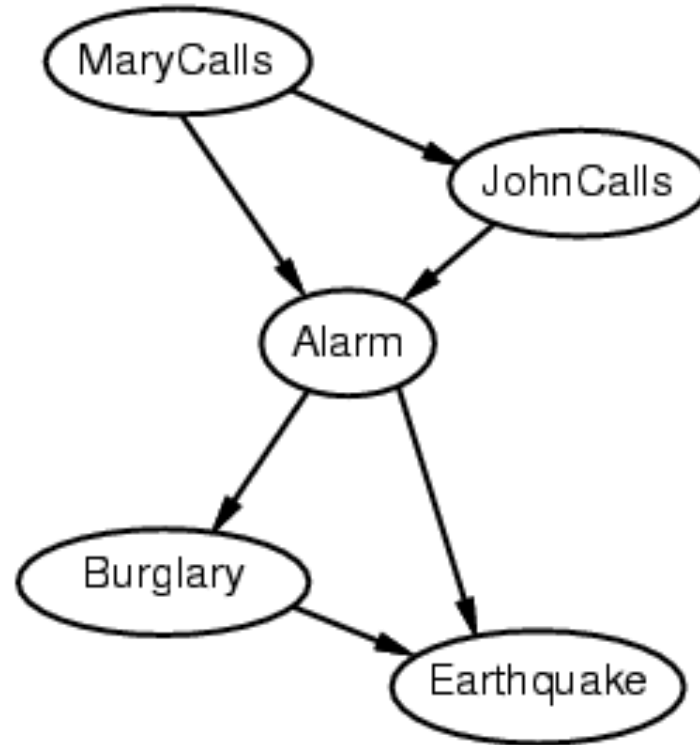
- So, the network is less compact if we go non-causal: 1 + 2 + 4 + 2 + 4 = 13 numbers needed instead of 10 if we go in causal direction.
- Deciding conditional independence is harder in noncausal directions
- Causal models and conditional independence seem hardwired for humans!

Right or Wrong Topology?

- If the topology is constructed manually, we (humans) tend to prefer the causal direction.
  - In domains such as medicine the graphs are *usually* less complex in the causal direction.

- But any topology is acceptable given there is probabilistic dependence between connected variables.

- The relative quality of a network is assessed by the performance of a corresponding BBN classifier.

Automatic construction

- Checking all combinations of variables for conditional independence is feasible only for very small datasets
- Uses local optimization techniques - heuristics

Weather data: Naïve Bayes network

Outlook	Temp.	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

Weka suggestion for BBN
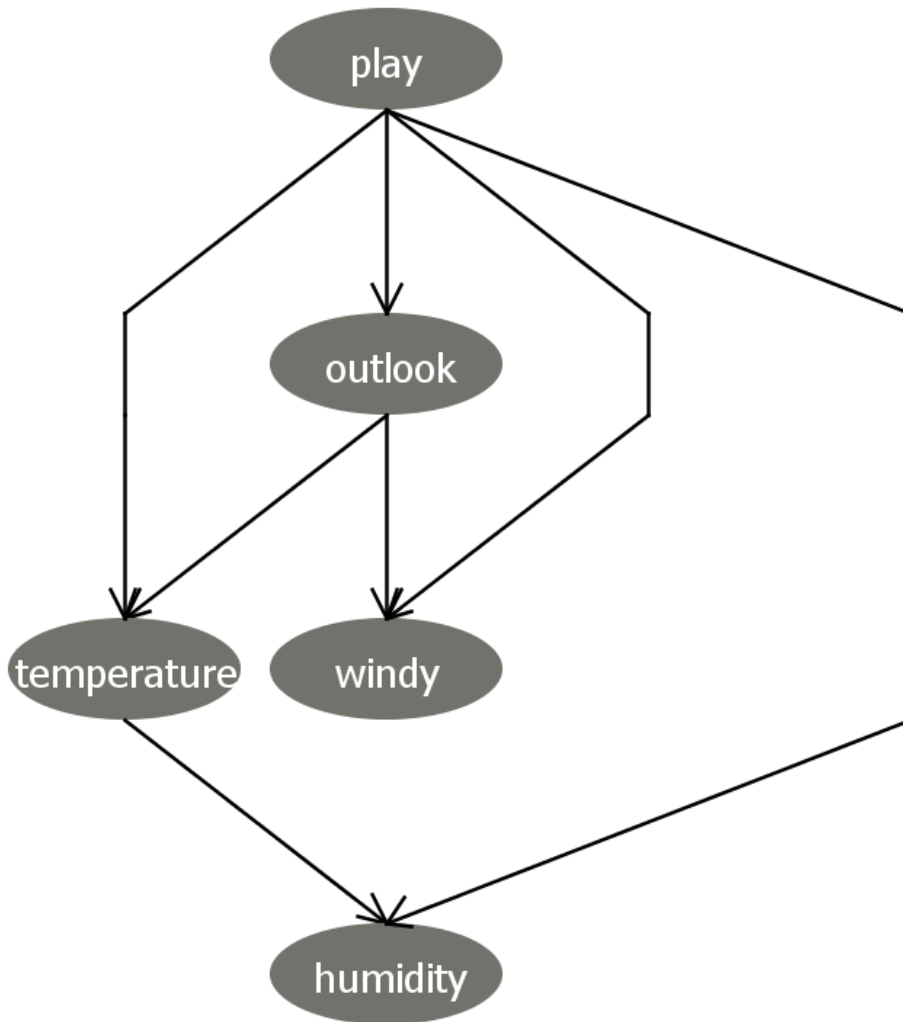
How Weka finds the shape of the graph?

Fixes an order of attributes (variables) and then adds and removes arcs until it gets the smallest estimated error (through cross-validation).

By default it starts with a Naïve Bayes network.

Also, it maintains a score of graph complexity, trying to keep the complexity low.
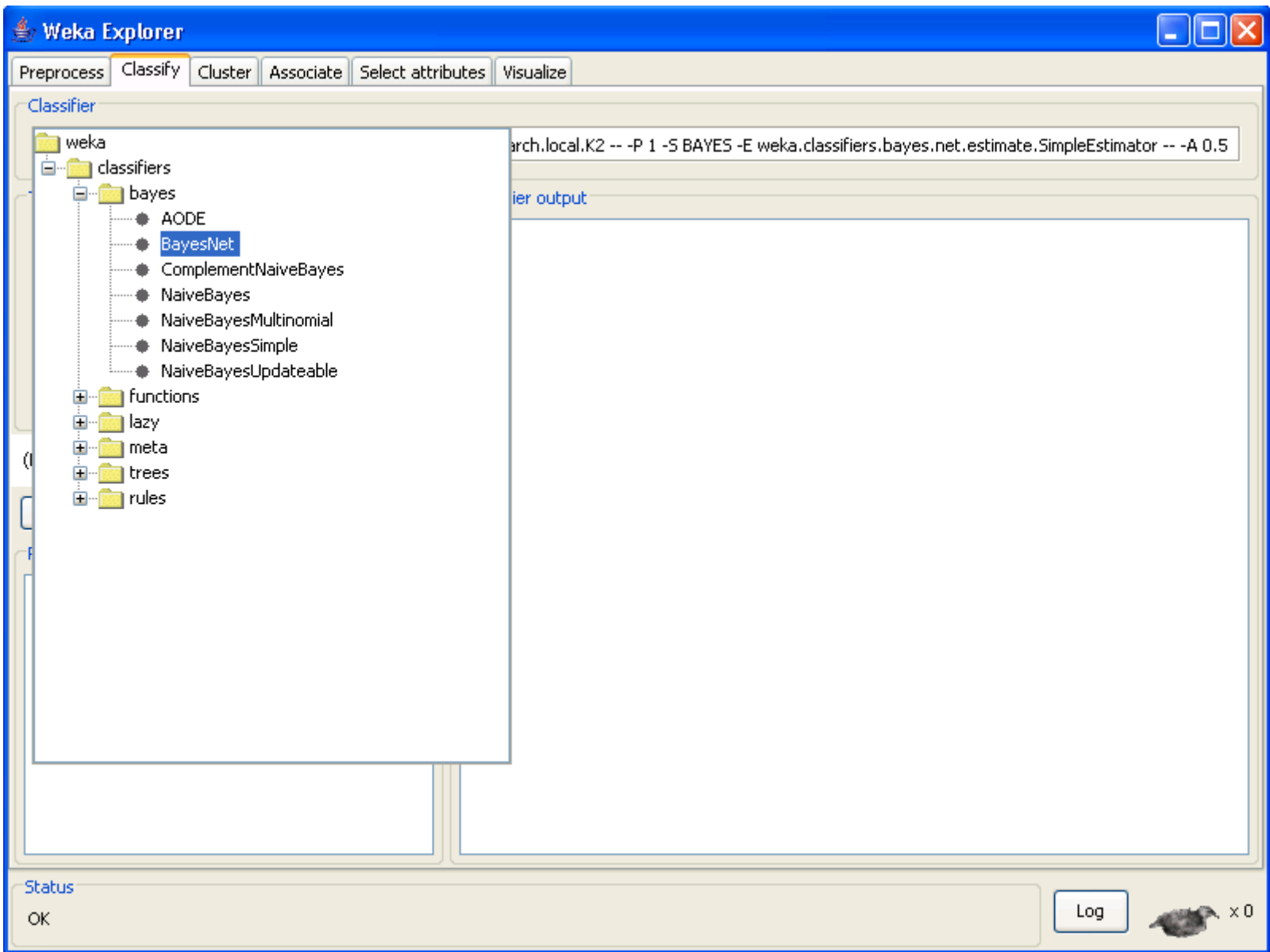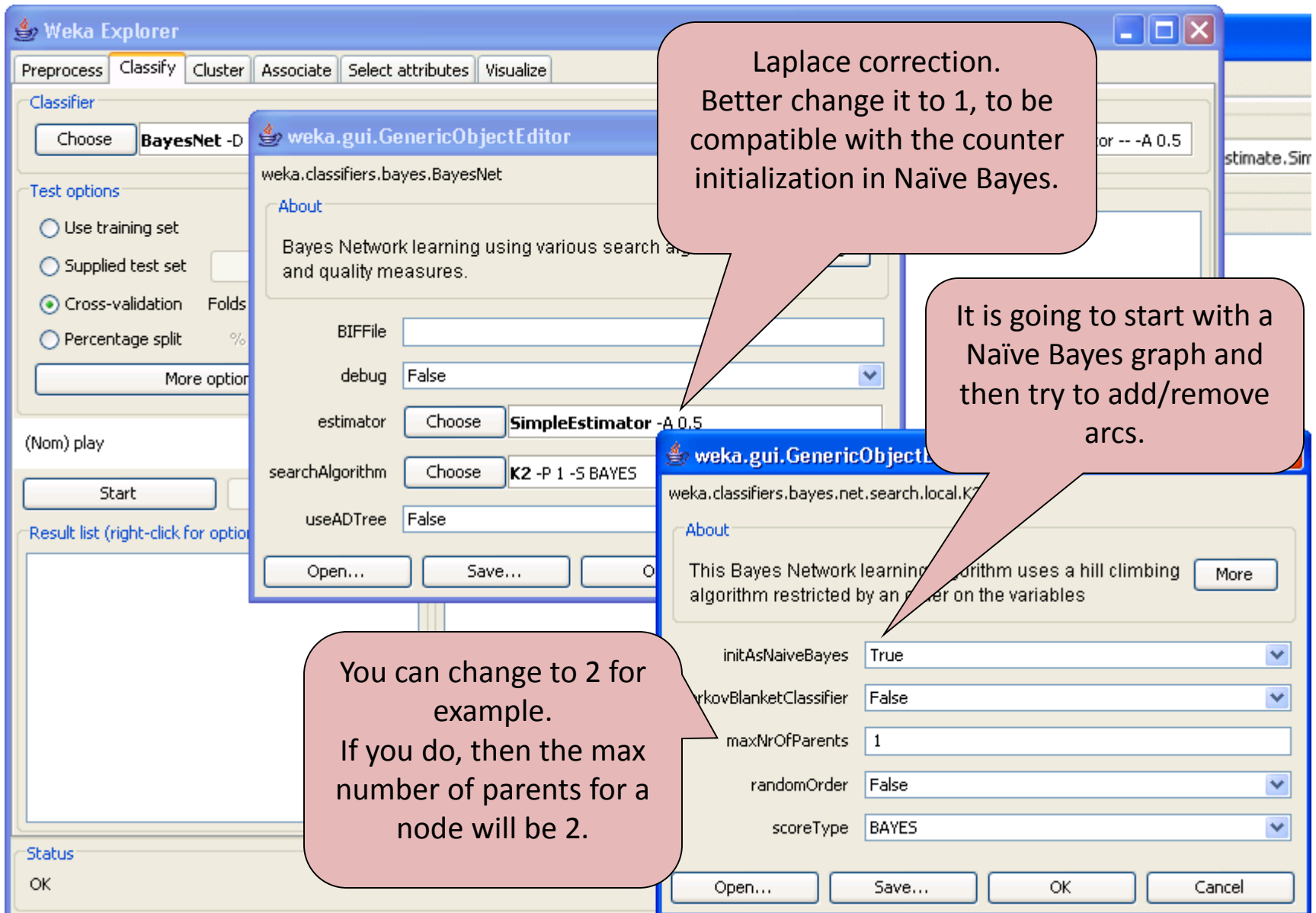
Weka Explorer

| Preprocess | Classify | Cluster | Associate | Select attributes | Visualize |

Classifier

arch.local.K2 -- -P 1 -S BAYES -E weka.classifiers.bayes.net.estimate.SimpleEstimator -- -A 0.5

ier output

```
weka
  classifiers
    bayes
        AODE
        BayesNet
        ComplementNaiveBayes
        NaiveBayes
        NaiveBayesMultinomial
        NaiveBayesSimple
        NaiveBayesUpdateable
    functions
    lazy
    meta
    trees
    rules
```

Status

OK

Log

x 0

Play probability table

yes	no
0.625	0.375
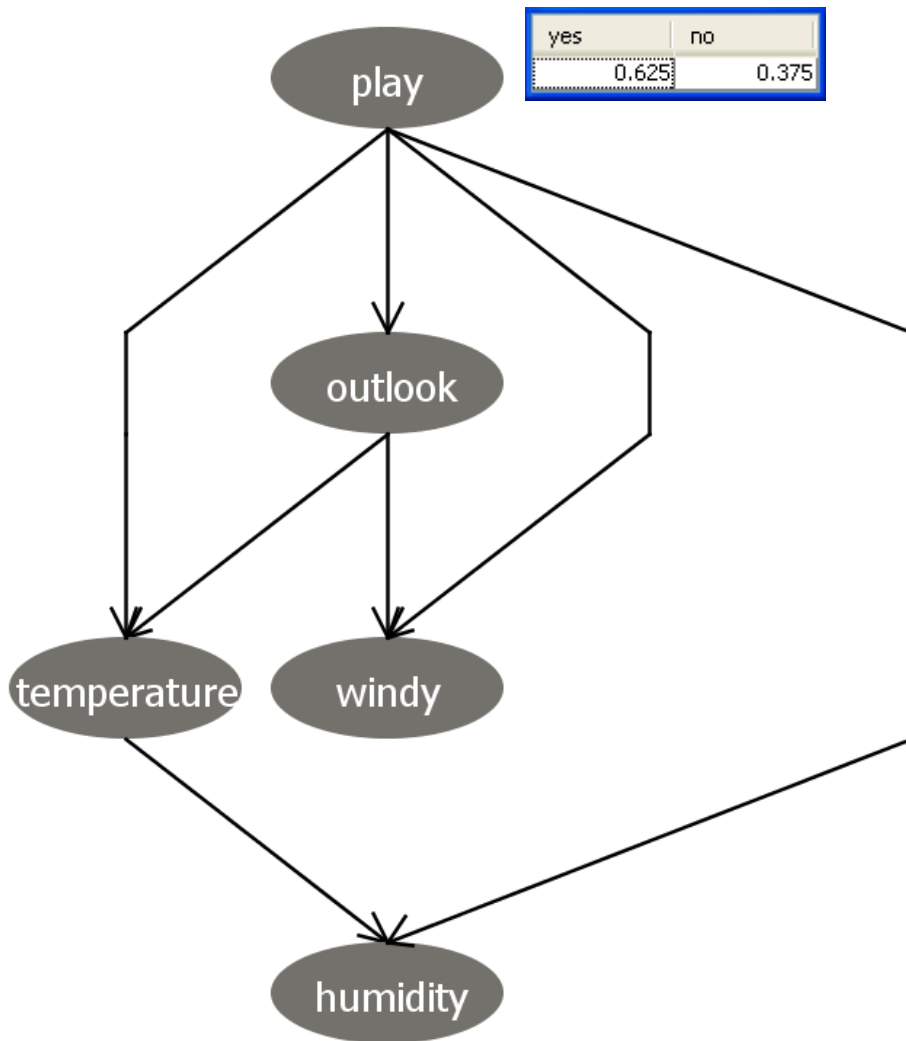
Based on the data…

P(play=yes) = 9/14
P(play=no) = 5/14

Let's correct with Laplace …

P(play=yes) = (9+1)/(14+2) = .625
P(play=yes) = (5+1)/(14+2) = .375

Outlook	Temp.	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

Outlook probability table

yes	no
0.625	0.375

play	sunny	overcast	rainy
yes	0.25	0.417	0.333
no	0.5	0.125	0.375
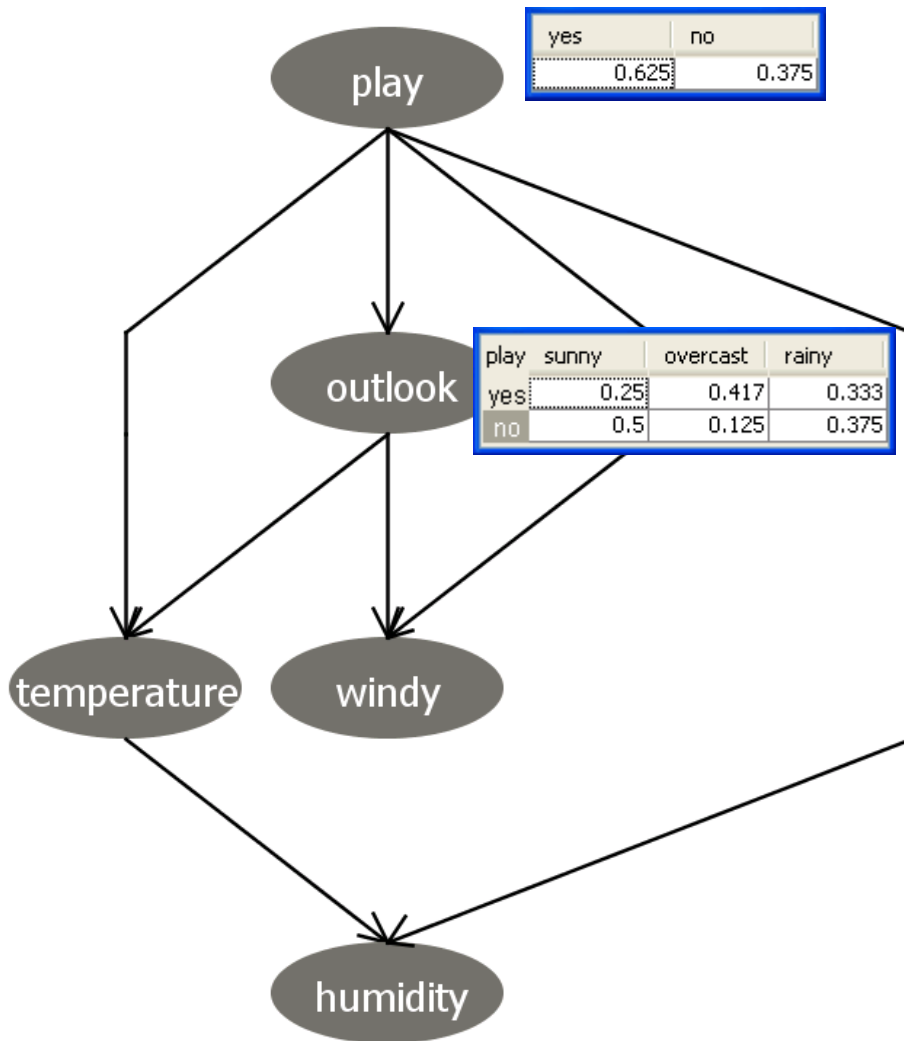
Based on the data…

P(outlook=sunny|play=yes) = (2+1)/(9+3) = .25
P(outlook=overcast|play=yes) = (4+1)/(9+3) = .417
P(outlook=rainy|play=yes) = (3+1)/(9+3) = .333

P(outlook=sunny|play=no) = (3+1)/(5+3) = .5
P(outlook=overcast|play=no) = (0+1)/(5+3) = .125
P(outlook=rainy|play=no) = (2+1)/(5+3) = .375

Windy probability table

Based on the data…let's find the conditional probabilities for "windy"

P(windy=true|play=yes,outlook=sunny) = (1+1)/(2+2) = .5

play	yes	no
	0.625	0.375

play	sunny	overcast	rainy
yes	0.25	0.417	0.333
no	0.5	0.125	0.375

play	outlook	TRUE	FALSE
yes	sunny	0.5	0.5
yes	overcast	0.5	0.5
yes	rainy	0.2	0.8
no	sunny	0.4	0.6
no	overcast	0.5	0.5
no	rainy	0.75	0.25

Outlook	Temp.	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

Windy probability table

play	yes	no
	0.625	0.375

play	sunny	overcast	rainy
yes	0.25	0.417	0.333
no	0.5	0.125	0.375

play	outlook	TRUE	FALSE
yes	sunny	0.5	0.5
yes	overcast	0.5	0.5
yes	rainy	0.2	0.8
no	sunny	0.4	0.6
no	overcast	0.5	0.5
no	rainy	0.75	0.25

P(windy=true|play=yes,outlook=sunny)
        = (1+1)/(2+2) = .5

P(windy=true|play=yes,outlook=overcast)
        = 0.5

P(windy=true|play=yes,outlook=rainy) = 0.2

P(windy=true|play=no,outlook=sunny) = 0.4

P(windy=true|play=no,outlook=overcast)
        = 0.5

P(windy=true|play=no,outlook=rainy)   = 0.75

Final figure

Classification I

play	yes	no
	0.625	0.375

Outlook	Temp.	Humidity	Windy	Play
Sunny	Cool	High	True	?

play	sunny	overcast	rainy
yes	0.25	0.417	0.333
no	0.5	0.125	0.375

play	outlook	TRUE	FALSE
yes	sunny	0.5	0.5
yes	overcast	0.5	0.5
yes	rainy	0.2	0.8
no	sunny	0.4	0.6
no	overcast	0.5	0.5
no	rainy	0.75	0.25

play	outlook	hot	mild	cool
yes	sunny	0.2	0.4	0.4
yes	overcast	0.429	0.286	0.286
yes	rainy	0.167	0.5	0.333
no	sunny	0.5	0.333	0.167
no	overcast	0.333	0.333	0.333
no	rainy	0.2	0.4	0.4

play	temperature	high	normal
yes	hot	0.5	0.5
yes	mild	0.5	0.5
yes	cool	0.2	0.8
no	hot	0.75	0.25
no	mild	0.75	0.25
no	cool	0.333	0.667

P(play=yes|outlook=sunny, temp=cool,humidity=high, windy=true) =

α*P(play=yes)
*P(outlook=sunny|play=yes)
*P(temp=cool|play=yes, outlook=sunny)
*P(humidity=high|play=yes, temp=cool)
*P(windy=true|play=yes, outlook=sunny)

= α*0.625*0.25*0.4*0.2*0.5
= α*0.00625

Classification II

	yes	no
play	0.625	0.375

outlook

play	sunny	overcast	rainy
yes	0.25	0.417	0.333
no	0.5	0.125	0.375

temperature

play	outlook	hot	mild	cool
yes	sunny	0.2	0.4	0.4
yes	overcast	0.429	0.286	0.286
yes	rainy	0.167	0.5	0.333
no	sunny	0.5	0.333	0.167
no	overcast	0.333	0.333	0.333
no	rainy	0.2	0.4	0.4

windy

play	outlook	TRUE	FALSE
yes	sunny	0.5	0.5
yes	overcast	0.5	0.5
yes	rainy	0.2	0.8
no	sunny	0.4	0.6
no	overcast	0.5	0.5
no	rainy	0.75	0.25

humidity

play	temperature	high	normal
yes	hot	0.5	0.5
yes	mild	0.5	0.5
yes	cool	0.2	0.8
no	hot	0.75	0.25
no	mild	0.75	0.25
no	cool	0.333	0.667

Outlook	Temp.	Humidity	Windy	Play
Sunny	Cool	High	True	?

P(play=no|outlook=sunny, temp=cool,humidity=high, windy=true) =

α*P(play=no)
*P(outlook=sunny|play=no)
*P(temp=cool|play=no, outlook=sunny)
*P(humidity=high|play= no, temp=cool)
*P(windy=true|play=no, outlook=sunny)

= α*0.375*0.5*0.167*0.333*0.4
= α*0.00417

Classification III

Classify it

	yes	no
play	0.625	0.375

outlook

play	sunny	overcast	rainy
yes	0.25	0.417	0.333
no	0.5	0.125	0.375

temperature

play	outlook	hot	mild	cool
yes	sunny	0.2	0.4	0.4
yes	overcast	0.429	0.286	0.286
yes	rainy	0.167	0.5	0.333
no	sunny	0.5	0.333	0.167
no	overcast	0.333	0.333	0.333
no	rainy	0.2	0.4	0.4

windy

play	outlook	TRUE	FALSE
yes	sunny	0.5	0.5
yes	overcast	0.5	0.5
yes	rainy	0.2	0.8
no	sunny	0.4	0.6
no	overcast	0.5	0.5
no	rainy	0.75	0.25

humidity

play	temperature	high	normal
yes	hot	0.5	0.5
yes	mild	0.5	0.5
yes	cool	0.2	0.8
no	hot	0.75	0.25
no	mild	0.75	0.25
no	cool	0.333	0.667

Outlook	Temp.	Humidity	Windy	Play
Sunny	Cool	High	True	?

P(play=yes|outlook=sunny, temp=cool,humidity=high, windy=true)
= α*0.00625

P(play=no|outlook=sunny, temp=cool,humidity=high, windy=true)
= α*.00417

α = 1/(0.00625+0.00417)
=95.969

P(play=yes|outlook=sunny, temp=cool,humidity=high, windy=true)
= 95.969*0.00625 = **0.60**

Classification IV (missing values or *hidden variables*)
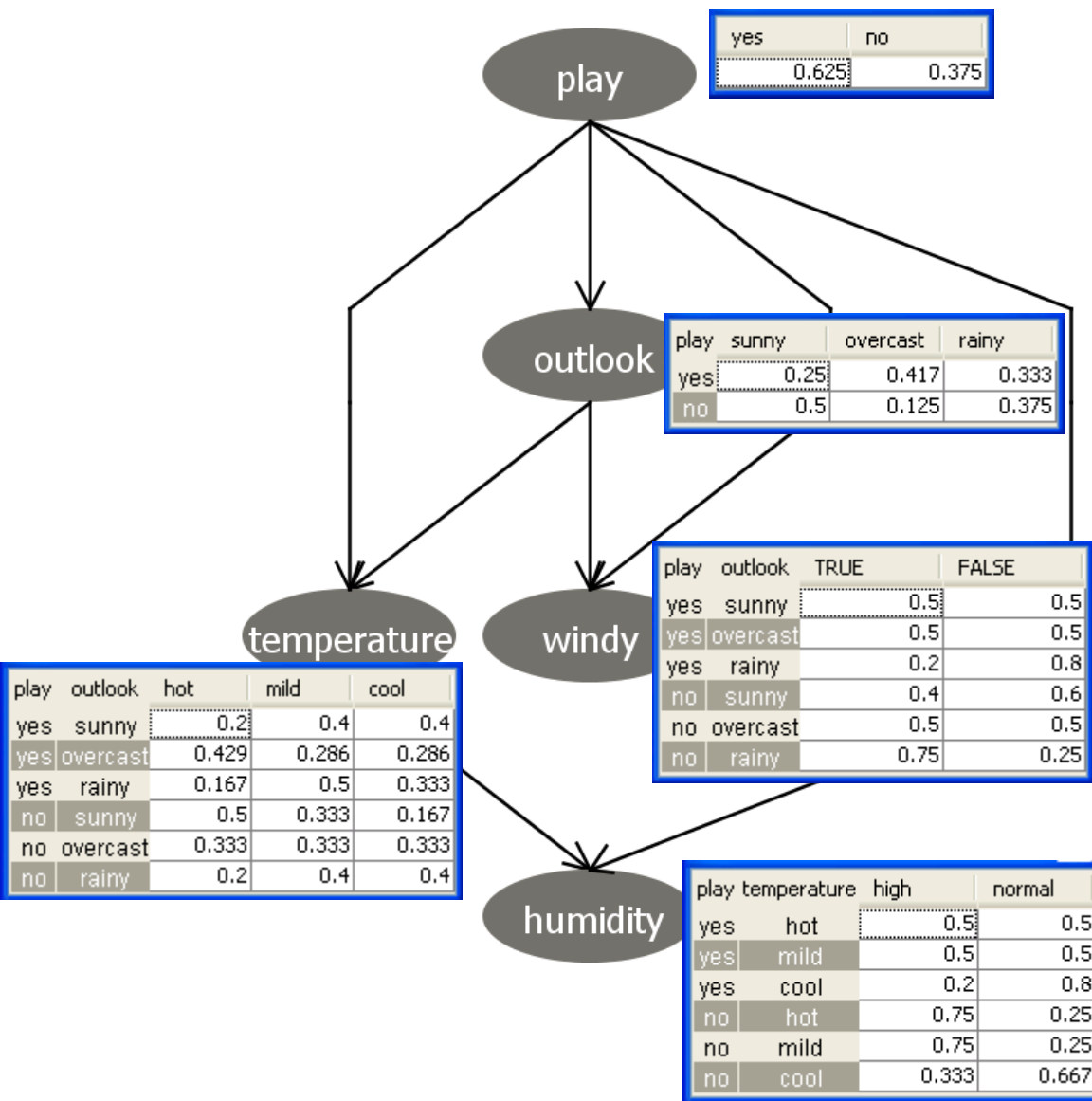
Outlook	Temp.	Humidity	Windy	Play
?	Cool	High	True	?

P(play=yes | temp=cool, humidity=high, windy=true)

$= \alpha * \Sigma_{outlook} P(\text{play=yes})$
$* P(\text{outlook} | \text{play=yes})$
$* P(\text{temp=cool} | \text{play=yes, outlook})$
$* P(\text{humidity=high} | \text{play=yes, temp=cool})$
$* P(\text{windy=true} | \text{play=yes, outlook})$

=…(*next slide*)

play

yes	no
0.625	0.375

outlook

play	sunny	overcast	rainy
yes	0.25	0.417	0.333
no	0.5	0.125	0.375

windy

play	outlook	TRUE	FALSE
yes	sunny	0.5	0.5
yes	overcast	0.5	0.5
yes	rainy	0.2	0.8
no	sunny	0.4	0.6
no	overcast	0.5	0.5
no	rainy	0.75	0.25

temperature

play	outlook	hot	mild	cool
yes	sunny	0.2	0.4	0.4
yes	overcast	0.429	0.286	0.286
yes	rainy	0.167	0.5	0.333
no	sunny	0.5	0.333	0.167
no	overcast	0.333	0.333	0.333
no	rainy	0.2	0.4	0.4

humidity

play	temperature	high	normal
yes	hot	0.5	0.5
yes	mild	0.5	0.5
yes	cool	0.2	0.8
no	hot	0.75	0.25
no	mild	0.75	0.25
no	cool	0.333	0.667

Classification V (missing values or *hidden variables*)

P(play=yes|temp=cool, humidity=high, windy=true)

= $\alpha * \Sigma_{outlook}$P(play=yes)*P(outlook|play=yes)*P(temp=cool|play=yes,outlook)
    *P(humidity=high|play=yes,temp=cool)*P(windy=true|play=yes,outlook)

= α*[
P(play=yes)*P(outlook= sunny|play=yes)*P(temp=cool|play=yes,outlook=sunny)
*P(humidity=high|play=yes,temp=cool)*P(windy=true|play=yes,outlook=sunny)

+P(play=yes)*P(outlook= overcast|play=yes)*P(temp=cool|play=yes,outlook=overcast)
*P(humidity=high|play=yes,temp=cool)*P(windy=true|play=yes,outlook=overcast)

+P(play=yes)*P(outlook= rainy|play=yes)*P(temp=cool|play=yes,outlook=rainy)
*P(humidity=high|play=yes,temp=cool)*P(windy=true|play=yes,outlook=rainy)
]

= α*[
  0.625*0.25*0.4*0.2*0.5 + 0.625*0.417*0.286*0.2*0.5 + 0.625*0.33*0.333*0.2*0.2]
=α***0.01645**

Classification VI (missing values or *hidden variables*)

P(play=no|temp=cool, humidity=high, windy=true)

= α*Σ~outlook~P(play=no)*P(outlook|play=no)*P(temp=cool|play=no,outlook)
    *P(humidity=high|play=no,temp=cool)*P(windy=true|play=no,outlook)

= α*[
P(play=no)*P(outlook=sunny|play=no)*P(temp=cool|play=no,outlook=sunny)
*P(humidity=high|play=no,temp=cool)*P(windy=true|play=no,outlook=sunny)

+P(play=no)*P(outlook= overcast|play=no)*P(temp=cool|play=no,outlook=overcast)
*P(humidity=high|play=no,temp=cool)*P(windy=true|play=no,outlook=overcast)

+P(play=no)*P(outlook= rainy|play=no)*P(temp=cool|play=no,outlook=rainy)
*P(humidity=high|play=no,temp=cool)*P(windy=true|play=no,outlook=rainy)
]

= α*[
 0.375*0.5*0.167*0.333*0.4 + 0.375*0.125*0.333*0.333*0.5 + 0.375*0.375*0.4*0.333*0.75
]
=α***0.0208**

Classification VII (missing values or *hidden variables*)

P(play=yes|temp=cool, humidity=high, windy=true) =α*0.01645
P(play=no|temp=cool, humidity=high, windy=true) =α*0.0208

α=1/(0.01645 + 0.0208)= 26.846

P(play=yes|temp=cool, humidity=high, windy=true) = 26.846 * 0.01645 = 0.44
P(play=no|temp=cool, humidity=high, windy=true) = 26.846 * 0.0208 = 0.56

I.e. P(play=yes|temp=cool, humidity=high, windy=true) is 44% and
    P(play=no|temp=cool, humidity=high, windy=true) is 56%

So, we predict 'play=no.'

Outline

- Extending the Naïve Bayes network to more complex networks. Joint probability
- Bayesian Belief networks (BBN). Definition. Types
- Query on BBN: what nodes to include. Markov blanket
- Query on BBN: how to compute. Simple example. Some variables may be hidden.
- Examples
- Finding network topology
- ▶ Applications of Bayesian networks

Applications of Bayesian Belief Networks

- Probably, the most important technology in the Machine Learning / AI field to have emerged in the last 10 years.

- 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