Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

CHAI
Humanities-Centered AI

Perception: Natural Language Processing and Computer Vision

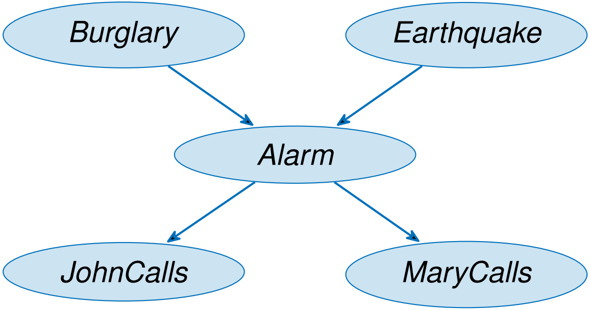
Malte Luttermann – Institute for Humanities-Centered Artificial Intelligence (CHAI)

April 24, 2026

Version from April 24, 2026

Topics for Today

- Bayesian networks semantics
- Exact inference in Bayesian networks
  - Variable elimination
  - Belief propagation
- Approximate inference in Bayesian networks
  - Prior sampling
  - Rejection sampling
  - Likelihood weighting
  - Markov Chain Monte Carlo



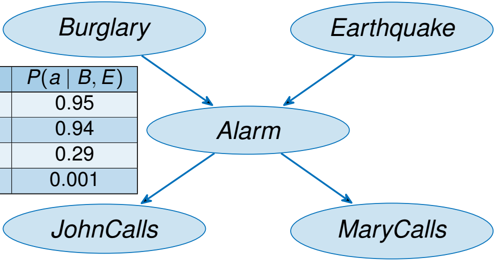
Recap – Bayesian Network Semantics

The full joint distribution is the product of local conditional distributions:

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

$P(b)$
0.001

$P(e)$
0.002



B	E	$P(a \mid B, E)$
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	$P(j \mid A)$
T	0.90
F	0.05

A	$P(m \mid A)$
T	0.70
F	0.01

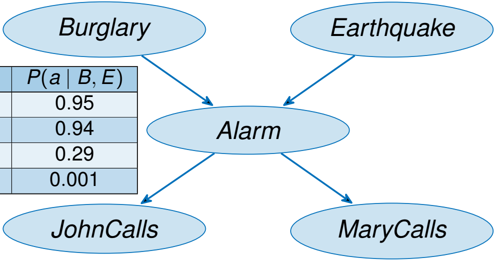
Each node X_i stores $P(X_i \mid Parents(X_i))$ as a *conditional probability table (CPT)*.

Recap – Bayesian Network Semantics

The full joint distribution is the product of local conditional distributions:

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

$P(b)$
0.001



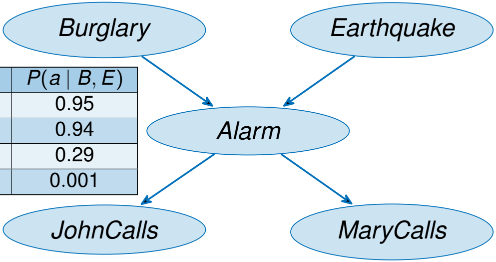
$P(e)$
0.002

B	E	$P(a \mid B, E)$
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	$P(j \mid A)$
T	0.90
F	0.05

A	$P(m \mid A)$
T	0.70
F	0.01

Recap – Bayesian Network Semantics



$P(b)$
0.001

$P(e)$
0.002

B	E	$P(a \mid B, E)$
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	$P(j \mid A)$
T	0.90
F	0.05

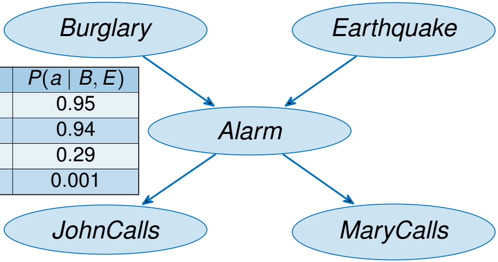
A	$P(m \mid A)$
T	0.70
F	0.01

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

Recap – Bayesian Network Semantics

$P(b)$
0.001

$P(e)$
0.002



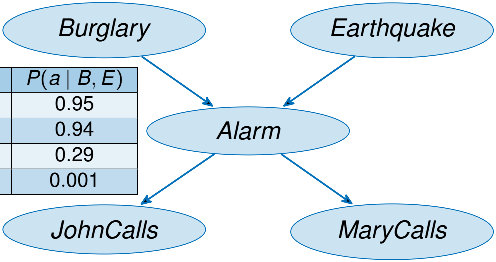
B	E	$P(a \mid B, E)$
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	$P(j \mid A)$
T	0.90
F	0.05

A	$P(m \mid A)$
T	0.70
F	0.01

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

Recap – Bayesian Network Semantics



$P(b)$
0.001

$P(e)$
0.002

B	E	$P(a \mid B, E)$
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	$P(j \mid A)$
T	0.90
F	0.05

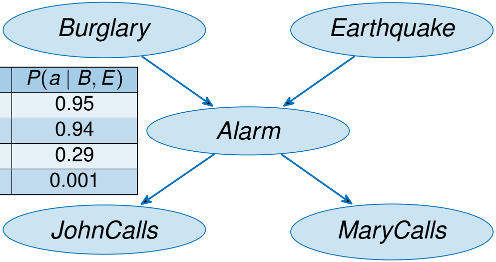
A	$P(m \mid A)$
T	0.70
F	0.01

$$
\begin{aligned}
P(j, m, a, \neg b, \neg e) &= P(j \mid a) \cdot P(m \mid a) \cdot P(a \mid \neg b, \neg e) \cdot P(\neg b) \cdot P(\neg e) \\
&= 0.90 \cdot 0.70 \cdot 0.001 \cdot 0.999 \cdot 0.998
\end{aligned}
$$

Recap – Bayesian Network Semantics

$P(b)$
0.001

$P(e)$
0.002



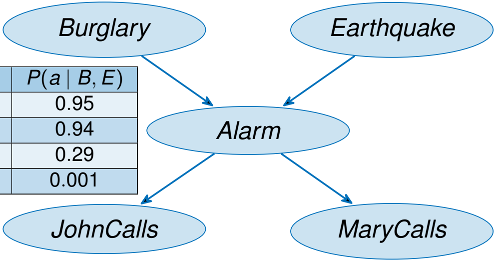
B	E	$P(a \mid B, E)$
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	$P(j \mid A)$
T	0.90
F	0.05

A	$P(m \mid A)$
T	0.70
F	0.01

$$
\begin{aligned}
P(j, m, a, \neg b, \neg e) &= P(j \mid a) \cdot P(m \mid a) \cdot P(a \mid \neg b, \neg e) \cdot P(\neg b) \cdot P(\neg e) \\
&= 0.90 \cdot 0.70 \cdot 0.001 \cdot 0.999 \cdot 0.998 \\
&\approx 0.00063
\end{aligned}
$$

Inference in Bayesian Networks



$P(b)$
0.001

$P(e)$
0.002

B	E	$P(a \mid B, E)$
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	$P(j \mid A)$
T	0.90
F	0.05

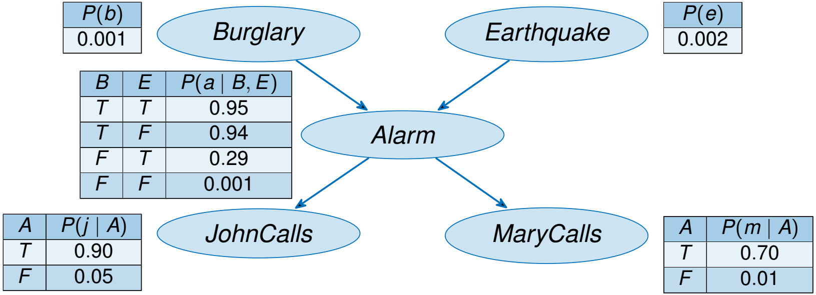
A	$P(m \mid A)$
T	0.70
F	0.01

Query answering in Bayesian networks:

$$P(X_i \mid \boldsymbol{E} = \boldsymbol{e}) \qquad \text{(simple query)}$$

$$P(X_i, X_j \mid \boldsymbol{E} = \boldsymbol{e}) \qquad \text{(conjunctive query)}$$

Inference in Bayesian Networks



$P(b)$
0.001

$P(e)$
0.002

B	E	$P(a \mid B, E)$
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	$P(j \mid A)$
T	0.90
F	0.05

A	$P(m \mid A)$
T	0.70
F	0.01

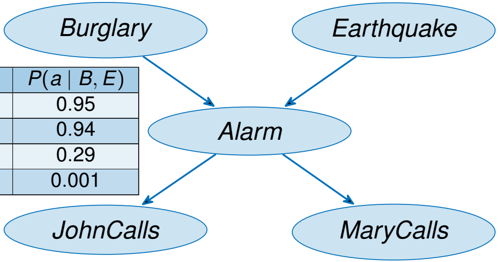
Example queries:

$$P(a) = ?$$

Inference in Bayesian Networks

$P(b)$
0.001

$P(e)$
0.002



B	E	$P(a \mid B, E)$
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	$P(j \mid A)$
T	0.90
F	0.05

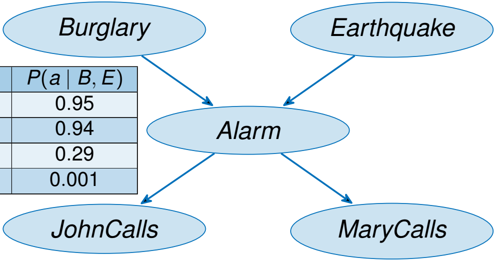
A	$P(m \mid A)$
T	0.70
F	0.01

Example queries:

$$P(a) = ?$$

$$P(j, m) = ?$$

Inference in Bayesian Networks



$P(b)$
0.001

$P(e)$
0.002

B	E	$P(a \mid B, E)$
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	$P(j \mid A)$
T	0.90
F	0.05

A	$P(m \mid A)$
T	0.70
F	0.01

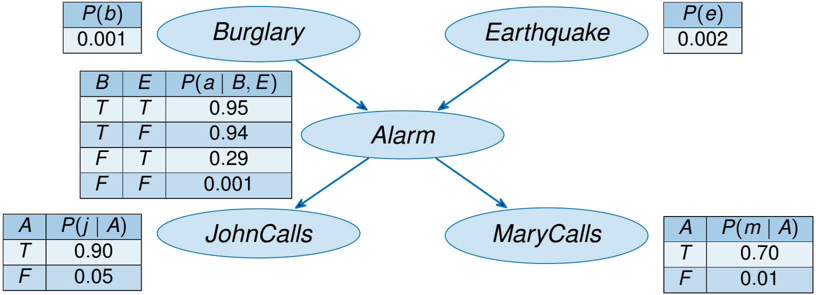
Example queries:

$$P(a) = ?$$

$$P(j \mid a) = ?$$

$$P(j, m) = ?$$

Inference in Bayesian Networks



$P(b)$
0.001

$P(e)$
0.002

B	E	$P(a \mid B, E)$
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	$P(j \mid A)$
T	0.90
F	0.05

A	$P(m \mid A)$
T	0.70
F	0.01

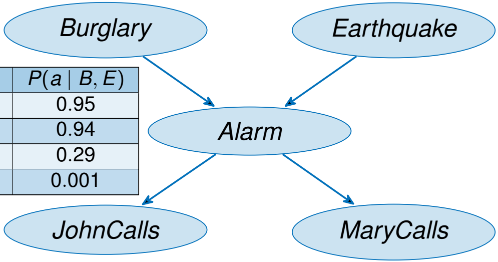
Example queries:

$$P(a) = ? \qquad P(j \mid a) = ?$$

$$P(j, m) = ? \qquad P(j, m \mid a) = ?$$

Inference in Bayesian Networks

$P(b)$
0.001



$P(e)$
0.002

B	E	$P(a \mid B, E)$
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	$P(j \mid A)$
T	0.90
F	0.05

A	$P(m \mid A)$
T	0.70
F	0.01

Example queries:

$$P(a) = ? \qquad P(j \mid a) = ? \qquad P(j, m \mid b) = ?$$

$$P(j, m) = ? \qquad P(j, m \mid a) = ? \qquad P(j, m \mid a, b) = ?$$

Variable Elimination

Goal: Compute $P(Q \mid E = e)$ without materialising the full joint.

Key objects: **Factors**

- Initial factors = CPTs from the network
- During elimination: New factors are created

Two basic operations:

(1) **Pointwise product** of factors
(2) **Summing out** a variable

Pointwise Product

$f_1(A, B) \cdot f_2(B, C) = f(A, B, C)$

Multiply entries that agree on shared variables.

Summing Out

$$\sum_x f_1 \cdot \ldots \cdot f_k = f_1 \cdot \ldots \cdot f_i \cdot \underbrace{\sum_x f_{i+1} \cdots f_k}_{f_X}$$

Move factors not depending on X outside the sum.

Pointwise Product

Pointwise Product by Hand

Given the following two factors over Boolean variables:

$f_1(A, B)$

A	B	$f_1(A, B)$
T	T	0.3
T	F	0.7
F	T	0.9
F	F	0.1

$f_2(B, C)$

B	C	$f_2(B, C)$
T	T	0.2
T	F	0.8
F	T	0.6
F	F	0.4

Compute the pointwise product $f_3(A, B, C) = f_1(A, B) \cdot f_2(B, C)$.

Pointwise Product

Pointwise Product by Hand

A	B	$f_1(A,B)$
T	T	0.3
T	F	0.7
F	T	0.9
F	F	0.1

$\cdot$

B	C	$f_2(B,C)$
T	T	0.2
T	F	0.8
F	T	0.6
F	F	0.4

$=$

A	B	C	$f_3(A,B,C)$
T	T	T	$0.3 \cdot 0.2 = 0.06$
T	T	F	$0.3 \cdot 0.8 = 0.24$
T	F	T	$0.7 \cdot 0.6 = 0.42$
T	F	F	$0.7 \cdot 0.4 = 0.28$
F	T	T	$0.9 \cdot 0.2 = 0.18$
F	T	F	$0.9 \cdot 0.8 = 0.72$
F	F	T	$0.1 \cdot 0.6 = 0.06$
F	F	F	$0.1 \cdot 0.4 = 0.04$

Summing Out

Summing Out by Hand

Given the previously computed factor $f_3(A, B, C)$:

$f_3(A, B, C)$

A	B	C	$f_3(A, B, C)$
T	T	T	$0.3 \cdot 0.2 = 0.06$
T	T	F	$0.3 \cdot 0.8 = 0.24$
T	F	T	$0.7 \cdot 0.6 = 0.42$
T	F	F	$0.7 \cdot 0.4 = 0.28$
F	T	T	$0.9 \cdot 0.2 = 0.18$
F	T	F	$0.9 \cdot 0.8 = 0.72$
F	F	T	$0.1 \cdot 0.6 = 0.06$
F	F	F	$0.1 \cdot 0.4 = 0.04$

Sum out B from $f_3(A, B, C)$ to obtain $f_4(A, C)$.

Summing Out

Summing Out by Hand

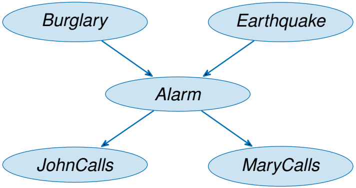
A	B	C	$f_3(A,B,C)$
T	T	T	$0.3 \cdot 0.2 = 0.06$
T	T	F	$0.3 \cdot 0.8 = 0.24$
T	F	T	$0.7 \cdot 0.6 = 0.42$
T	F	F	$0.7 \cdot 0.4 = 0.28$
F	T	T	$0.9 \cdot 0.2 = 0.18$
F	T	F	$0.9 \cdot 0.8 = 0.72$
F	F	T	$0.1 \cdot 0.6 = 0.06$
F	F	F	$0.1 \cdot 0.4 = 0.04$

$\Longrightarrow$

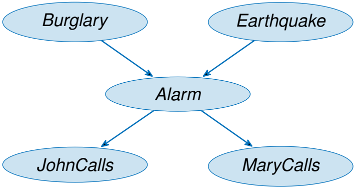
A	C	$f_4(A,C)$
T	T	$0.06 + 0.42 = 0.48$
T	F	$0.24 + 0.28 = 0.52$
F	T	$0.18 + 0.06 = 0.24$
F	F	$0.72 + 0.04 = 0.76$

Variable Elimination – Burglary Example

Query: $P(B \mid j, m)$ (evidence: $J = T, M = T$)



Variable Elimination – Burglary Example

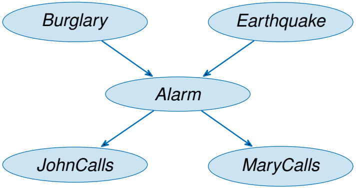


Query: $P(B \mid j, m)$ (evidence: $J = T$, $M = T$)

Step 1: Write the joint and push sums inward.

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

Variable Elimination – Burglary Example



Query: $P(B \mid j, m)$ (evidence: $J = T, M = T$)

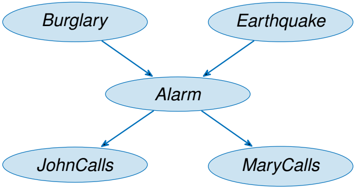
Step 1: Write the joint and push sums inward.

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

Factors:

$f_B(B) = P(B)$, $f_E(E) = P(E)$, $f_A(A, B, E) = P(A \mid B, E)$, $f_J(A) = P(j \mid A)$, $f_M(A) = P(m \mid A)$

Variable Elimination – Burglary Example



Query: $P(B \mid j, m)$ (evidence: $J = T$, $M = T$)

Step 1: Write the joint and push sums inward.

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

Factors:

$$f_B(B) = P(B), \quad f_E(E) = P(E), \quad f_A(A, B, E) = P(A \mid B, E), \quad f_J(A) = P(j \mid A), \quad f_M(A) = P(m \mid A)$$

Step 2: Choose elimination order: ***E*, then *A***.

$$P(B \mid j, m) = \alpha\, f_B(B) \cdot \sum_{a} f_J(a) \cdot f_M(a) \cdot \underbrace{\sum_{e} f_E(e) \cdot f_A(a, B, e)}_{f_1(A,B)}$$

Variable Elimination – Eliminate *E*

$$f_1(A, B) = \sum_e f_E(e) \cdot f_A(A, B, e)$$

E	$f_E(E)$
T	0.002
F	0.998

A	B	E	$f_A(A, B, E)$
T	T	T	0.95
T	T	F	0.94
T	F	T	0.29
T	F	F	0.001
F	T	T	0.05
F	T	F	0.06
F	F	T	0.71
F	F	F	0.999

Variable Elimination – Eliminate E

$$f_1(A, B) = \sum_e f_E(e) \cdot f_A(A, B, e)$$

E	$f_E(E)$
T	0.002
F	0.998

A	B	E	$f_A(A, B, E)$
T	T	T	0.95
T	T	F	0.94
T	F	T	0.29
T	F	F	0.001
F	T	T	0.05
F	T	F	0.06
F	F	T	0.71
F	F	F	0.999

$\Longrightarrow$

A	B	$f_1(A, B)$
T	T	$0.002 \cdot 0.95 + 0.998 \cdot 0.94$
T	F	$0.002 \cdot 0.29 + 0.998 \cdot 0.001$
F	T	$0.002 \cdot 0.05 + 0.998 \cdot 0.06$
F	F	$0.002 \cdot 0.71 + 0.998 \cdot 0.999$

Variable Elimination – Eliminate A

$$f_2(B) = \sum_a f_J(a) \cdot f_M(a) \cdot f_1(a, B)$$

A	$f_J(A)$
T	0.90
F	0.05

A	$f_M(A)$
T	0.70
F	0.01

A	B	$f_1(A, B)$
T	T	$0.002 \cdot 0.95 + 0.998 \cdot 0.94 = 0.94002$
T	F	$0.002 \cdot 0.29 + 0.998 \cdot 0.001 = 0.00158$
F	T	$0.002 \cdot 0.05 + 0.998 \cdot 0.06 = 0.05998$
F	F	$0.002 \cdot 0.71 + 0.998 \cdot 0.999 = 0.99842$

Variable Elimination – Eliminate A

$$f_2(B) = \sum_a f_J(a) \cdot f_M(a) \cdot f_1(a, B)$$

A	$f_J(A)$
T	0.90
F	0.05

A	$f_M(A)$
T	0.70
F	0.01

A	B	$f_1(A, B)$
T	T	$0.002 \cdot 0.95 + 0.998 \cdot 0.94 = 0.94002$
T	F	$0.002 \cdot 0.29 + 0.998 \cdot 0.001 = 0.00158$
F	T	$0.002 \cdot 0.05 + 0.998 \cdot 0.06 = 0.05998$
F	F	$0.002 \cdot 0.71 + 0.998 \cdot 0.999 = 0.99842$

$\Longrightarrow$

B	$f_2(B)$
T	$0.90 \cdot 0.70 \cdot 0.94002 + 0.05 \cdot 0.01 \cdot 0.05998 = 0.59224$
F	$0.90 \cdot 0.70 \cdot 0.00158 + 0.05 \cdot 0.01 \cdot 0.99842 = 0.00149$

Variable Elimination – Final Result

$$P(B \mid j, m) = \alpha\, f_B(B) \cdot \sum_a f_J(a) \cdot f_M(a) \cdot \underbrace{\sum_e f_E(e) \cdot f_A(a, B, e)}_{f_1(A,B)}$$

$$= \alpha\, f_B(B) \cdot \sum_a f_J(a) \cdot f_M(a) \cdot f_1(a, B)$$

$$= \alpha\, f_B(B) \cdot \underbrace{\sum_a f_J(a) \cdot f_M(a) \cdot f_1(a, B)}_{f_2(B)}$$

$$= \alpha\, f_B(B) \cdot f_2(B)$$

Variable Elimination – Final Result

$$P(B \mid j, m) = \alpha\, f_B(B) \cdot f_2(B)$$

Step 3: Multiply with prior $f_B(B) = P(B)$ and normalise.

B	$f_B(B)$
T	0.001
F	0.999

$\cdot$

B	$f_2(B)$
T	0.59224
F	0.00149

$=$

B	$f_B(B) \cdot f_2(B)$
T	$0.001 \cdot 0.59224 = 0.000592$
F	$0.999 \cdot 0.00149 = 0.001489$

Normalise:

$$P(B = T \mid j, m) = \frac{0.000592}{0.000592 + 0.001489} \approx \mathbf{0.284}$$

$$P(B = F \mid j, m) = \frac{0.001489}{0.000592 + 0.001489} \approx \mathbf{0.716}$$

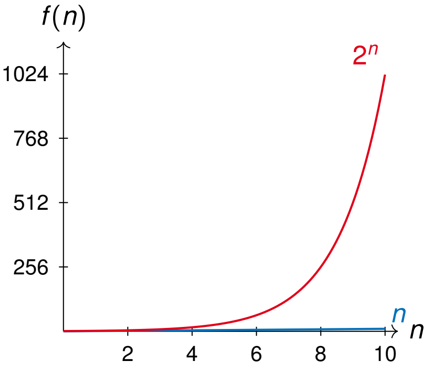
Polytree Networks

Definition: A BN is a **polytree** (singly connected network) iff there is at most one undirected path between any two nodes.

Key property:

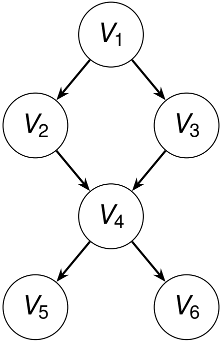
- Exact inference is NP-hard on general BNs
- But $O(n)$ on polytrees!

Idea: Do variable elimination *locally* at each node in parallel, sending/receiving messages.

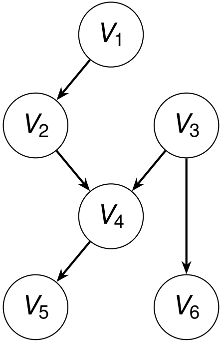


Polytree Networks

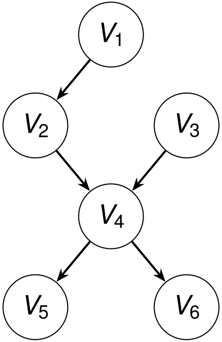
Definition: A BN is a **polytree** (singly connected network) iff there is at most one undirected path between any two nodes.



Polytree?



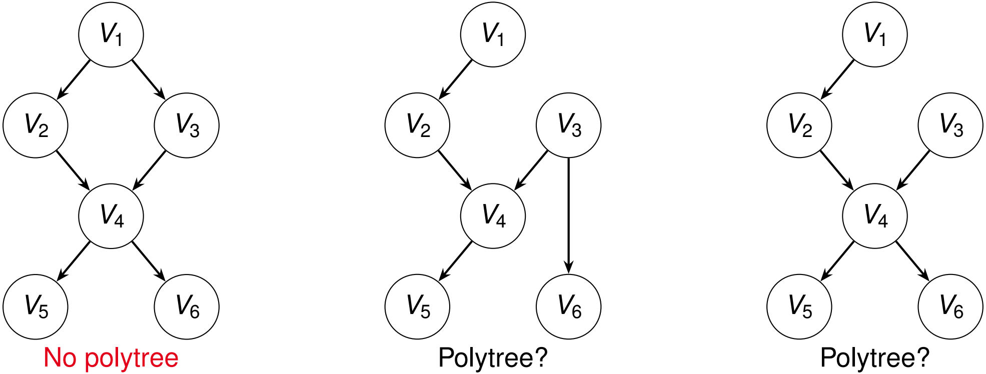
Polytree?



Polytree?

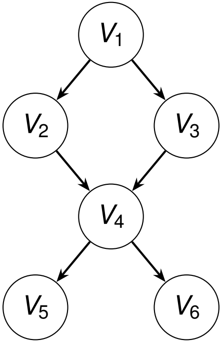
Polytree Networks

Definition: A BN is a **polytree** (singly connected network) iff there is at most one undirected path between any two nodes.

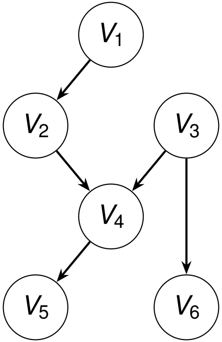


Polytree Networks

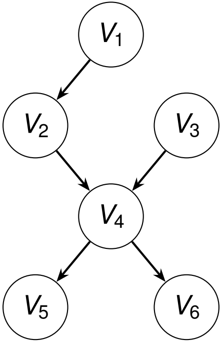
Definition: A BN is a **polytree** (singly connected network) iff there is at most one undirected path between any two nodes.



No polytree



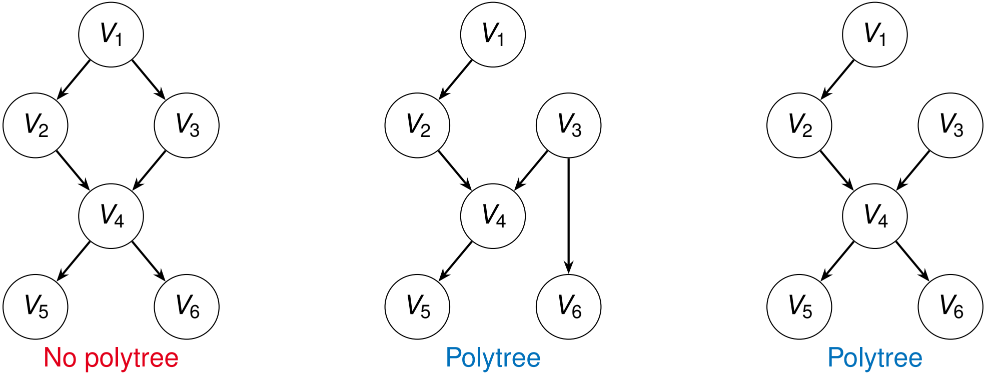
Polytree



Polytree?

Polytree Networks

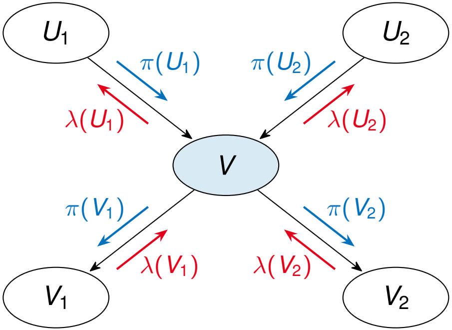
Definition: A BN is a **polytree** (singly connected network) iff there is at most one undirected path between any two nodes.



Belief Propagation

Node V splits the network into two disjoint evidence parts:

- E_V^+: evidence accessible through **parents** (causal/predictive)
- E_V^-: evidence accessible through **children** (diagnostic)



Two message types:

- π**-messages**: parents to children
  - $\pi(V) = P(V \mid E_V^+)$
- λ**-messages**: children to parents
  - $\lambda(V) = P(E_V^- \mid V)$

Belief: $\mathrm{BEL}(V) = \alpha\,\lambda(V)\,\pi(V)$

Belief Propagation – Algorithm

Initialisation:

- Evidence nodes $V_i = e_i$:
  $\lambda(x_i) = \pi(x_i) = \begin{cases} 1 & x_i = e_i \\ 0 & \text{otherwise} \end{cases}$
- Nodes without parents:
  $\pi(x_i) = P(x_i)$ (prior)
- Nodes without children:
  $\lambda(x_i) = 1$ (uniform)

Iteration (until convergence; Y_j are children, U_j are parents):

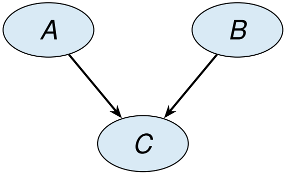
(1) Compute $\pi(x) = \sum_{u_1,\ldots,u_p} P(x \mid u_1, \ldots, u_p) \prod_j \pi_{U_j X}(u_j)$
(2) Compute $\lambda(x) = \prod_j \lambda_{Y_j X}(x)$
(3) Send $\pi_{XY_j}(x) = \alpha\, \pi(x) \prod_{k \neq j} \lambda_{Y_k X}(x)$
(4) Send $\lambda_{Y_j X}(x) = \sum_j \lambda(y_j) \sum_{u_1,\ldots,u_q} P(y_j \mid u_1, \ldots, u_q) \prod_{k, U_k \neq X} \pi_{U_k Y_j}(u_k)$

Result: $\mathrm{BEL}(X) = \alpha\, \lambda(X)\, \pi(X)$

Belief Propagation – Example

Consider the following polytree with evidence $E = \{C = \mathit{true}\}$:

$P(a)$
0.4



$P(b)$
0.3

A	B	$P(c \mid A, B)$
T	T	0.9
T	F	0.7
F	T	0.8
F	F	0.1

Compute $P(A \mid C = \mathit{true})$ using belief propagation.

Belief Propagation – Example: Initialisation

Step 1: Initialise using the BP algorithm rules.

- C is an **evidence node** ($C = T$), so: $\lambda(c) = 1$, $\lambda(\neg c) = 0$.
- A and B are **root nodes** (no parents), so their π-values are the priors:

$$\pi(a) = P(a) = 0.4, \quad \pi(\neg a) = 0.6, \quad \pi(b) = P(b) = 0.3, \quad \pi(\neg b) = 0.7$$

Belief Propagation – Example: λ-Message from C to A

Step 2: Compute the λ-message that C sends to its parent A.

The formula (from the algorithm) is:

$$\lambda_{CA}(a) = \sum_b \lambda(c) \cdot P(c \mid a, b) \cdot \pi_{BC}(b)$$

Since $\lambda(c) = 1$ (evidence) and $\pi_{BC}(b) = \pi(b) = P(b)$ (root node with only child C):

$$\begin{aligned}\lambda_{CA}(a) &= P(c \mid a, b) \cdot \pi(b) + P(c \mid a, \neg b) \cdot \pi(\neg b) \\ &= 0.9 \cdot 0.3 + 0.7 \cdot 0.7 = \mathbf{0.76}\end{aligned}$$

$$\begin{aligned}\lambda_{CA}(\neg a) &= P(c \mid \neg a, b) \cdot \pi(b) + P(c \mid \neg a, \neg b) \cdot \pi(\neg b) \\ &= 0.8 \cdot 0.3 + 0.1 \cdot 0.7 = \mathbf{0.31}\end{aligned}$$

Belief Propagation – Example: Compute Belief

Step 3: Combine π and λ to get the belief at *A*.

The belief formula is:

$$\mathrm{BEL}(A) = \alpha\, \lambda_{CA}(A) \cdot \pi(A)$$

Substituting the values:

$$\mathrm{BEL}(A) = \alpha\, \langle 0.76 \cdot 0.4,\ 0.31 \cdot 0.6 \rangle = \alpha\, \langle 0.304,\ 0.186 \rangle$$

Step 4: Normalise ($\alpha = \frac{1}{0.304+0.186} = \frac{1}{0.49}$):

$$P(a \mid c) = \frac{0.304}{0.49} \approx \mathbf{0.62}, \qquad P(\neg a \mid c) = \frac{0.186}{0.49} \approx \mathbf{0.38}$$

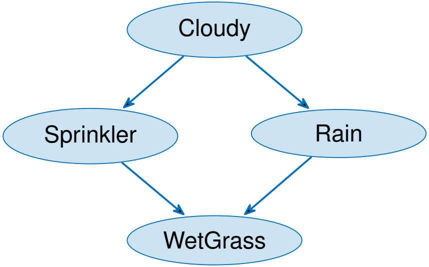
Approximate Inference

- Exact inference in general BNs is NP-hard
- Even approximate inference is NP-hard in the worst case!
- Most real-world BNs are *not* polytrees

- Randomised sampling algorithms
- Accuracy depends on number of samples

Example Bayesian Network

$P(c)$
0.5



C	$P(s \mid C)$
T	0.10
F	0.50

C	$P(r \mid C)$
T	0.80
F	0.20

S	R	$P(w \mid S, R)$
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.00

Prior Sampling

Idea: Sample each variable in **topological order** using the CPTs.

Example: Generate one sample from the weather network.

Variable	Distribution	Sampled value
C	$P(C) = 0.5$	T
S	$P(S \mid C = T) = 0.10$	F
R	$P(R \mid C = T) = 0.80$	T
W	$P(W \mid S = F, R = T) = 0.90$	T

$\Rightarrow$ Sample: $[C = T,\ S = F,\ R = T,\ W = T]$

Property: Samples are drawn from the **true joint distribution**.
After $N = 1000$ samples: $N(\text{Rain} = T) = 511 \ \Rightarrow\ \hat{P}(\text{Rain} = T) = 0.511$.

Rejection Sampling

Goal: Estimate a *conditional* probability, e.g. $P(R \mid S = T)$.

Procedure: Generate prior samples, **reject** those inconsistent with evidence.

Example: $P(\text{Rain} \mid \text{Sprinkler} = T)$ with 100 prior samples:

	Count	**Action**
$S = F$ (evidence mismatch)	73	**rejected**
$S = T, R = T$	8	kept
$S = T, R = F$	19	kept

$$\hat{P}(R = T \mid S = T) = \frac{8}{8+19} \approx 0.296$$

Problem: 73% of samples wasted! Gets worse with more evidence variables.

Likelihood Weighting

Idea: Fix evidence variables, **weight** each sample by their likelihood.

Example: $P(R \mid S = T, W = T)$. Generate one weighted sample:

Variable	Action	Value	Weight w
C	sample from $P(C)$	T	1.0
S	**evidence**: $w \leftarrow w \cdot P(S = T \mid C = T)$	T	$1.0 \cdot 0.1 = 0.1$
R	sample from $P(R \mid C = T)$	T	0.1
W	**evidence**: $w \leftarrow w \cdot P(W = T \mid S = T, R = T)$	T	$0.1 \cdot 0.99 = 0.099$

$\Rightarrow$ Sample $[T, T, T, T]$ with weight 0.099.
Estimate $P(R = r \mid S = T, W = T)$ by summing the weights of samples with $R = r$ and dividing by the total weight of all samples.

Advantage: No samples wasted!

Limitation: Only considers evidence from *parents*, not children.

Likelihood Weighting – Example

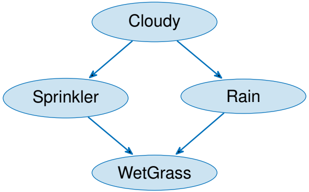
Query: $P(R \mid C = T, W = T)$

(1) Compute the weight for the sample $[C = T, S = F, R = T, W = T]$.

(2) Compute the weight for the sample $[C = T, S = F, R = F, W = T]$.

(3) Which sample gets a higher weight? Why does this make sense intuitively?

$P(c)$
0.5



C	$P(s \mid C)$
T	0.10
F	0.50

C	$P(r \mid C)$
T	0.80
F	0.20

S	R	$P(w \mid S, R)$
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.00

Likelihood Weighting – Example

Evidence: $C = \text{true}$ and $W = \text{true}$.

(1) Sample $[T, F, T, T]$:

- $C = T$ is evidence: $w = 1.0 \cdot P(C = T) = 0.5$
- $S = F$: no evidence, no weight change
- $R = T$: no evidence, no weight change
- $W = T$ is evidence: $w = 0.5 \cdot P(W = T \mid S = F, R = T) = 0.5 \cdot 0.90 = \mathbf{0.45}$

Likelihood Weighting – Example

Evidence: $C =$ true and $W =$ true.

(1) Sample $[T, F, T, T]$:

- $C = T$ is evidence: $w = 1.0 \cdot P(C = T) = 0.5$
- $S = F$: no evidence, no weight change
- $R = T$: no evidence, no weight change
- $W = T$ is evidence: $w = 0.5 \cdot P(W = T \mid S = F, R = T) = 0.5 \cdot 0.90 = \mathbf{0.45}$

(2) Sample $[T, F, F, T]$:

- $C = T$ is evidence: $w = 1.0 \cdot P(C = T) = 0.5$
- $W = T$ is evidence: $w = 0.5 \cdot P(W = T \mid S = F, R = F) = 0.5 \cdot 0.00 = \mathbf{0.0}$

Likelihood Weighting – Example

Evidence: $C =$ true and $W =$ true.

(1) Sample $[T, F, T, T]$:

- $C = T$ is evidence: $w = 1.0 \cdot P(C = T) = 0.5$
- $S = F$: no evidence, no weight change
- $R = T$: no evidence, no weight change
- $W = T$ is evidence: $w = 0.5 \cdot P(W = T \mid S = F, R = T) = 0.5 \cdot 0.90 = \mathbf{0.45}$

(2) Sample $[T, F, F, T]$:

- $C = T$ is evidence: $w = 1.0 \cdot P(C = T) = 0.5$
- $W = T$ is evidence: $w = 0.5 \cdot P(W = T \mid S = F, R = F) = 0.5 \cdot 0.00 = \mathbf{0.0}$

(3) Intuition: Without rain and sprinkler, the grass is not wet.

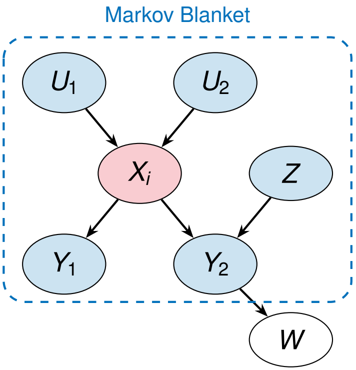
Markov Chain Monte Carlo (MCMC)

Idea: Generate each new state by making a **random change** to the current state.

Gibbs Sampling Procedure:

(1) Start with an arbitrary state (evidence variables fixed)
(2) Pick a non-evidence variable X_i
(3) Resample X_i conditioned on its **Markov blanket**
(4) Record the new state
(5) Repeat

Markov blanket of X_i:
Parents + children + children's other parents



Markov Chain Monte Carlo – Example

Weather network with evidence $S = \text{true}$, $W = \text{true}$.

Query: $P(R \mid S = T, W = T)$

Initial state: $[C = T,\ S = T,\ R = F,\ W = T]$

(1) What is the Markov blanket of C?

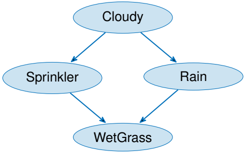
(2) What is the Markov blanket of R?

(3) We sample C from $P(C \mid S = T, R = F)$. Suppose we get $C = F$. What is the new state?

(4) Next we sample R from $P(R \mid C = F, S = T, W = T)$. Why does this condition on *both* C (parent) and W (child)?

(5) Assume that after many iterations, we visited 20 states with $R = T$ and 60 with $R = F$. What is the MCMC estimate of $P(R \mid S = T, W = T)$?

$P(c)$
0.5



C	$P(s \mid C)$
T	0.10
F	0.50

C	$P(r \mid C)$
T	0.80
F	0.20

S	R	$P(w \mid S, R)$
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.00

Markov Chain Monte Carlo – Example

(1) Markov blanket of C is $\{S, R\}$
(2) Markov blanket of R is $\{C, W, S\}$
(3) New state: $[C = F,\ S = T,\ R = F,\ W = T]$
(4) C and W are both in R's Markov blanket and the Markov blanket makes R conditionally independent of *all other nodes*
(5) $\hat{P}(R \mid S = T, W = T) = \text{NORMALISE}(\langle 20, 60 \rangle) = \langle \mathbf{0.25},\ \mathbf{0.75} \rangle$

Overview of Inference Methods

Overview of Inference Methods

Method	**Exact?**	**Handles evidence**
Variable Elimination (VE)	Yes	Any
Belief Propagation (BP)	Yes*	Any
Prior sampling	No	None
Rejection sampling	No	Any
Likelihood weighting (LW)	No	Any†
Markov Chain Monte Carlo (MCMC)	No	Any

*Only on polytrees.

†Degrades when evidence is on leaf nodes.