Artificial Intelligence
Chapter 19
Reasoning with Uncertain Information

Biointelligence Lab
School of Computer Sci. & Eng.
Seoul National University

Outline

- Review of Probability Theory
- Probabilistic Inference
- Bayes Networks
- Patterns of Inference in Bayes Networks
- Uncertain Evidence
- D-Separation
- Probabilistic Inference in Polytrees

19.1 Review of Probability Theory (1/4)

- Random variables

$$V_1, V_2, \ldots, V_k$$

- Joint probability

$$p(V_1 = v_1, V_2 = v_2, \ldots, V_k = v_k)$$

Ex.

(B (BAT_OK), M (MOVES) , L (LIFTABLE), G (GUAGE))	Joint Probability
(True, True, True, True)	0.5686
(True, True, True, False)	0.0299
(True, True, False, True)	0.0135
(True, True, False, False)	0.0007
…	…

19.1 Review of Probability Theory (2/4)

- Marginal probability

Ex. $$p(B=b)=\sum_{B=b} p(B,M,L,G)$$

$$p(B=b,M=m)=\sum_{B=b,M=m} p(B,M,L,G)$$

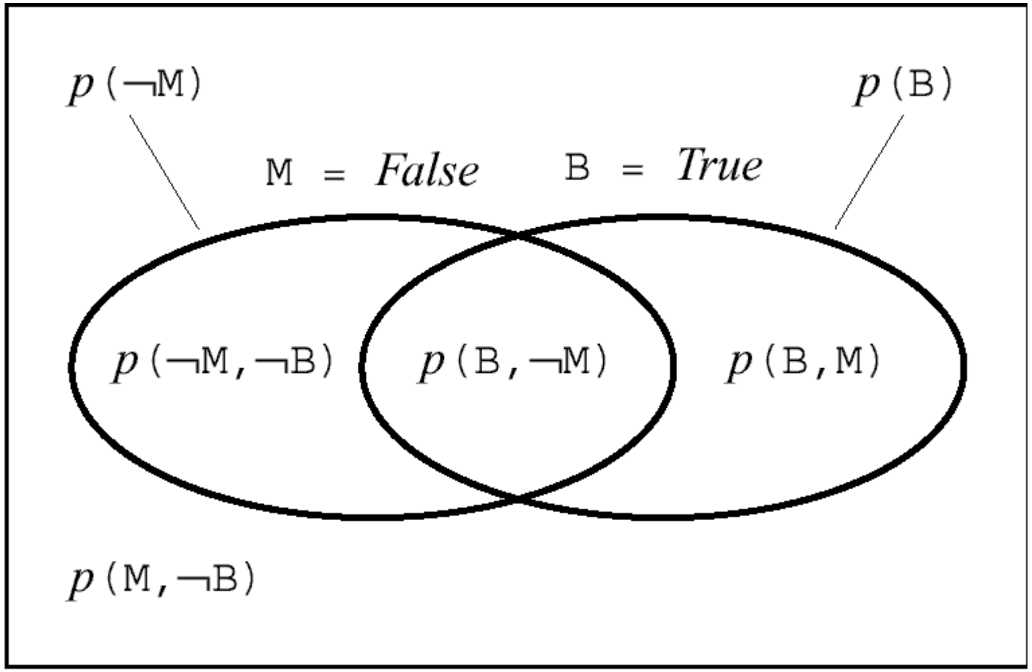
- Conditional probability

$$p\left(V_i \mid V_j\right)=\frac{p\left(V_i,V_j\right)}{p\left(V_j\right)}$$

  - ♦ Ex. The probability that the battery is charged given that the arm does not move

$$p\left(B=True \mid M=False\right)=\frac{p\left(B=True,M=False\right)}{p\left(M=False\right)}$$

19.1 Review of Probability Theory (3/4)



$p(B \mid \neg M) = p(B, \neg M) / p(\neg M)$

Figure 19.1 A Venn Diagram

19.1 Review of Probability Theory (4/4)

- Chain rule $p(B,L,G,M) = p(B \mid L,G,M)\,p(L \mid G,M)\,p(G \mid M)\,p(M)$

$$p(V_1, V_2, \ldots, V_k) = \prod_{i=1}^{k} p(V_i \mid V_{i-1}, \ldots, V_1)$$

- Bayes' rule

$$p(V_i \mid V_j) = \frac{p(V_j \mid V_i)\,p(V_i)}{p(V_j)}$$

- set notation $p(\mathcal{V})$
  - ♦ Abbreviation for $p(V_1, V_2, \ldots, V_k)$
    where $\mathcal{V} = \{V_1, V_2, \ldots, V_k\}$

19.2 Probabilistic Inference

- We desire to calculate the probability of some variable V_i has value v_i given the evidence $\mathcal{E}$ =**e**.

$$p(V_i = True \mid \mathcal{E} = e) = \frac{p(V_i = True, \mathcal{E} = e)}{p(\mathcal{E} = e)}$$

Example

p(P,Q,R)	0.3
p(P,Q,¬R)	0.2
p(P, ¬Q,R)	0.2
p(P, ¬Q,¬R)	0.1
p(¬P,Q,R)	0.05
p(¬P, Q, ¬R)	0.1
p(¬P, ¬Q,R)	0.05
p(¬P, ¬Q,¬R)	0.0

$$p(Q \mid \neg R) = \frac{p(Q, \neg R)}{p(\neg R)} = \frac{[p(P,Q,\neg R) + p(\neg P,Q,\neg R)]}{p(\neg R)}$$
$$= \frac{(0.2 + 0.1)}{p(\neg R)} = \frac{0.3}{p(\neg R)}$$

$$p(\neg Q \mid \neg R) = \frac{p(\neg Q, \neg R)}{p(\neg R)} = \frac{[p(P,\neg Q,\neg R) + p(\neg P,\neg Q,\neg R)]}{p(\neg R)}$$
$$= \frac{(0.1 + 0.0)}{p(\neg R)} = \frac{0.1}{p(\neg R)}$$

$$p(Q \mid \neg R) = 0.75$$
$$\because\ p(Q \mid \neg R) + p(\neg Q \mid \neg R) = 1$$

Statistical Independence

- Conditional independence

$$p(V_i, V_j \mid \mathcal{V}) = p(V_i \mid \mathcal{V})\, p(V_j \mid \mathcal{V})$$

$\mathcal{V}$: a set of variables

  - Intuition: V_i tells us nothing more about $\mathcal{V}$ than we already knew by knowing V_j

- Mutually conditional independence

$$p(V_1, V_2, \ldots, V_j \mid \mathcal{V}) = \prod_{i=1}^{k} p(V_i \mid V_{i-1}, V_{i-2}, \ldots, V_1, \mathcal{V}) = \prod_{i=1}^{k} p(V_i \mid \mathcal{V})$$

- Unconditional independence (When $\mathcal{V}$ is empty)

$$p(V_1, V_2, \ldots, V_j) = p(V_1)\, p(V_2) \ldots p(V_k)$$

19.3 Bayes Networks (1/2)

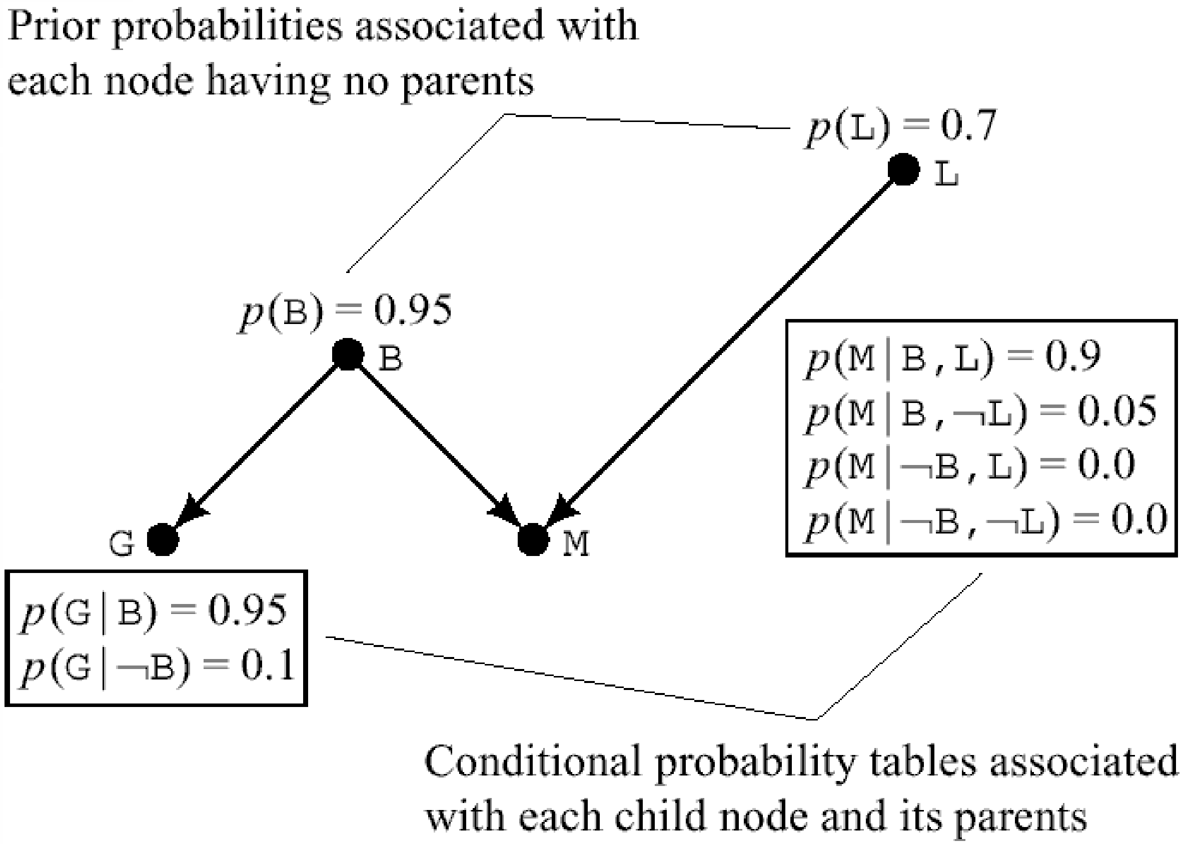
- Directed, acyclic graph (DAG) whose nodes are labeled by random variables
- Characteristics of Bayesian networks
  - Node V_i is conditionally independent of any subset of nodes that are not descendents of V_i given its parents

$$p(V_1, V_2, \ldots, V_k) = \prod_{i=1}^{k} p(V_i \mid Pa(V_i))$$

- Prior probability
- Conditional probability table (CPT)

19.3 Bayes Networks (2/2)

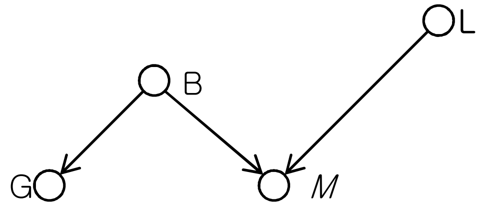
Bayes network about the block-lifting example



$$p(\mathtt{G},\mathtt{B},\mathtt{M},\mathtt{L}) = p(\mathtt{G} \mid \mathtt{B})\,p(\mathtt{M} \mid \mathtt{B},\mathtt{L})\,p(\mathtt{B})\,p(\mathtt{L})$$

19.4 Patterns of Inference in Bayes Networks (1/3)

- Causal or top-down inference
  - Ex. The probability that the arm moves given that the block is liftable



$$p(M \mid L) = p(M, B \mid L) + p(M, \neg B \mid L)$$

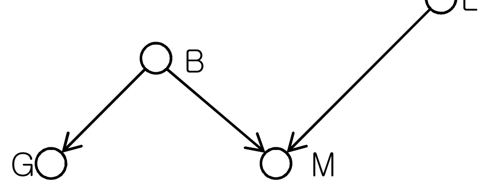
$$= p(M \mid B, L)p(B \mid L) + p(M \mid \neg B, L)p(\neg B \mid L)$$ (chain rule)

$$= p(M \mid B, L)p(B) + p(M \mid \neg B, L)p(\neg B)$$ (from the structure)

$$= 0.9 * 0.95 = 0.855 .$$

19.4 Patterns of Inference in Bayes Networks (2/3)

- Diagnostic or bottom-up inference
  - Using an effect (or symptom) to infer a cause
  - Ex. The probability that the block is not liftable given that the arm does not move.



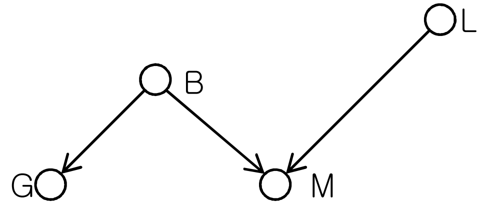
$$p(\neg M \mid \neg L) = 0.9525$$ (using causal reasoning)

$$p(\neg L \mid \neg M) = \frac{p(\neg M \mid \neg L)p(\neg L)}{p(\neg M)} = \frac{0.9525 \times 0.3}{p(\neg M)} = \frac{0.28575}{p(\neg M)}$$ (using Bayes' rule)

$$p(L \mid \neg M) = \frac{p(\neg M \mid L)p(L)}{p(\neg M)} = \frac{0.145 \times 0.7}{p(\neg M)} = \frac{0.1015}{p(\neg M)}$$ (using Bayes' rule)

$$p(\neg L \mid \neg M) = 0.7379$$

19.4 Patterns of Inference in Bayes Networks (3/3)



- Explaining away
  - One evidence: $\neg M$ (the arm does not move)
  - Additional evidence: $\neg B$ (the battery is not charged)

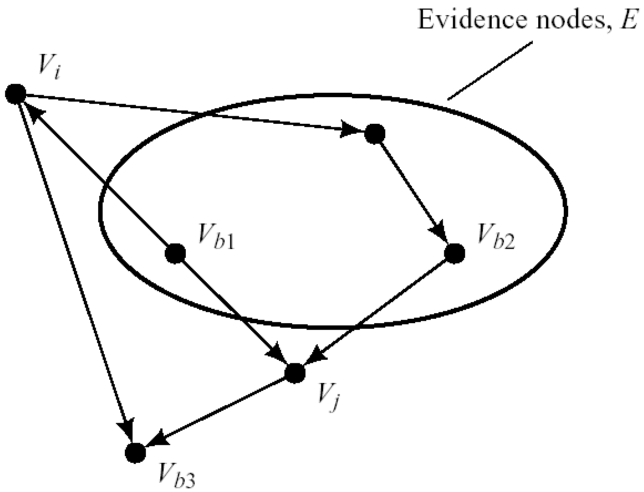
$$
\begin{aligned}
p(\neg L \mid \neg B, \neg M) &= \frac{p(\neg M, \neg B \mid \neg L)\, p(\neg L)}{p(\neg B, \neg M)} && \text{(Bayes' rule)} \\
&= \frac{p(\neg M \mid \neg B, \neg L)\, p(\neg B \mid \neg L)\, p(\neg L)}{p(\neg B, \neg M)} && \text{(def. of conditional prob.)} \\
&= \frac{p(\neg M \mid \neg B, \neg L)\, p(\neg B)\, p(\neg L)}{p(\neg B, \neg M)} && \text{(structure of the Bayes network)} \\
&= 0.30.
\end{aligned}
$$

  - $\neg B$ explains $\neg M$, making $\neg L$ less certain (0.30<0.7379)

19.5 Uncertain Evidence

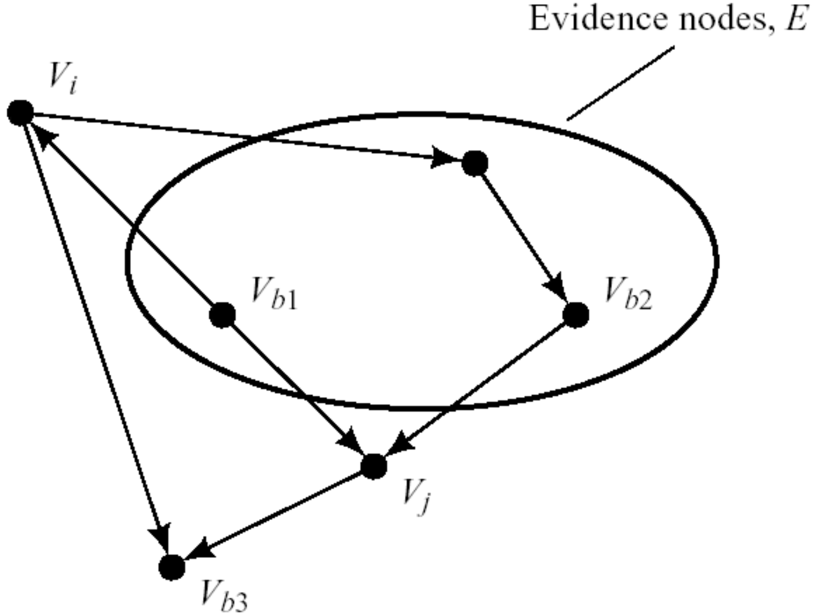
- We must be certain about the truth or falsity of the propositions they represent.
  - Each uncertain evidence node should have a child node, about which we can be certain.
  - Ex. Suppose the robot is not certain that its arm did not move.
    - Introducing M' : "The arm sensor says that the arm moved"
      - We can be certain that that proposition is either true or false.
    - $p(\neg L | \neg B, \neg M')$ instead of $p(\neg L | \neg B, \neg M)$
  - Ex. Suppose we are uncertain about whether or not the battery is charged.
    - Introducing G : "Battery guage"
    - $p(\neg L | \neg G, \neg M')$ instead of $p(\neg L | \neg B, \neg M')$

19.6 D-Separation (1/3)



- D-saparation: direction-dependent separation
- Two nodes V_i and V_j are conditionally independent given a set of nodes $\mathcal{E}$ if for every *undirected* path in the Bayes network between V_i and V_j, there is some node, V_b, on the path having one of the following three properties.
  - V_b is in $\mathcal{E}$, and both arcs on the path lead out of V_b
  - V_b is in $\mathcal{E}$, and one arc on the path leads in to V_b and one arc leads out.
  - Neither V_b nor any descendant of V_b is in $\mathcal{E}$, and both arcs on the path lead in to V_b.
- V_b blocks the path given $\mathcal{E}$ when any one of these conditions holds for a path.
- If all paths between V_i and V_j are blocked, we say that $\mathcal{E}$ *d-separates* V_i and V_j

19.6 D-Separation (2/3)



V_i is independent of V_j given the evidence nodes because all three paths between them are blocked. The blocking nodes are

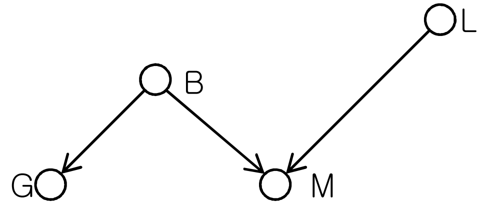
(a) V_{b1} is an evidence node, and both arcs lead out of V_{b1}.

(b) V_{b2} is an evidence node, and one arc leads into V_{b2} and one arc leads out.

(c) V_{b3} is not an evidence node, nor are any of its descendants, and both arcs lead into V_{b3}.

Figure 19.3 Conditional Independence via Blocking Nodes

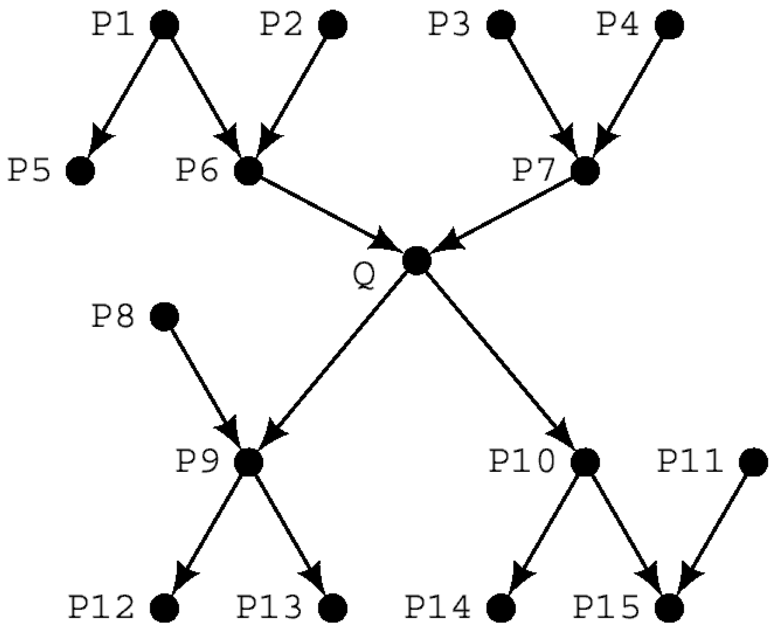
19.6 D-Separation (3/3)



- Ex.
  - $I(G, L|B)$ by rules 1 and 3
    - By rule 1, *B* blocks the (only) path between *G* and *L*, given *B*.
    - By rule 3, *M* also blocks this path given *B*.
  - $I(G, L)$
    - By rule 3, *M* blocks the path between G and *L*.
  - $I(B, L)$
    - By rule 3, *M* blocks the path between *B* and *L*.
- Even using *d*-separation, probabilistic inference in Bayes networks is, in general, NP-hard.

19.7 Probabilistic Inference in Polytrees (1/2)

- Polytree
  - A DAG for which there is just one path, along arcs in either direction, between any two nodes in the DAG.

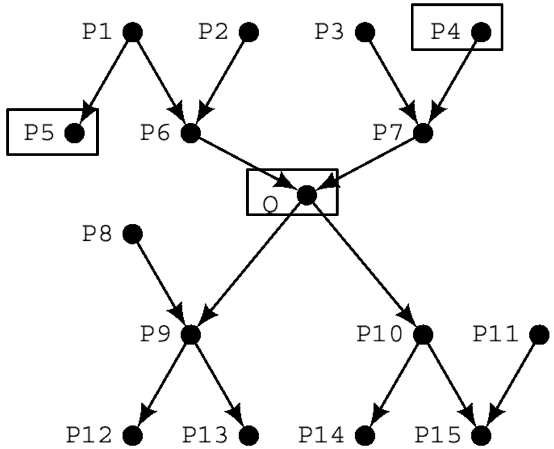


19.7 Probabilistic Inference in Polytrees (2/2)

- A node is above Q
  - The node is connected to Q only through Q's parents
- A node is below Q
  - The node is connected to Q only through Q's immediate successors.
- Three types of evidences
  - All evidence nodes are above Q.
  - All evidence nodes are below Q.
  - There are evidence nodes both above and below Q.

Evidence Above (1/2)

- Bottom-up recursive algorithm
- Ex. *p(Q|P5, P4)*



$$
\begin{aligned}
p(Q \mid P5, P4) &= \sum_{P6,P7} p(Q, P6, P7 \mid P5, P4) \\
&= \sum_{P6,P7} p(Q \mid P6, P7, P5, P4)\, p(P6, P7 \mid P5, P4) \\
&= \sum_{P6,P7} p(Q \mid P6, P7)\, p(P6, P7 \mid P5, P4) \\
&= \sum_{P6,P7} p(Q \mid P6, P7)\, p(P6 \mid P5, P4)\, p(P7 \mid P5, P4) \\
&= \sum_{P6,P7} p(Q \mid P6, P7)\, p(P6 \mid P5)\, p(P7 \mid P4)
\end{aligned}
$$

(Structure of The Bayes network)

(d-separation)

(d-separation)

Evidence Above (2/2)

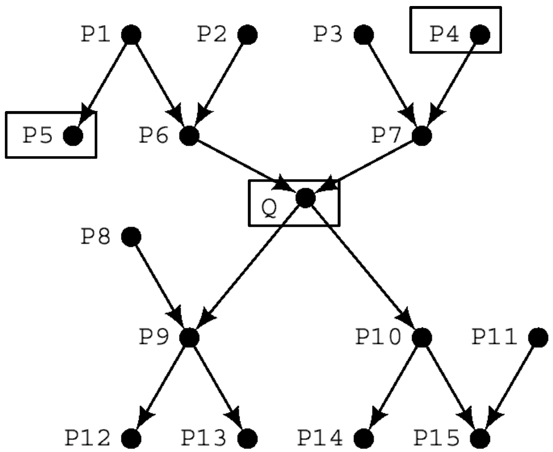
- Calculating $p(P7|P4)$ and $p(P6|P5)$

$$p\left(P7 \mid P4\right) = \sum_{P3} p\left(P7 \mid P3, P4\right) p\left(P3 \mid P4\right) = \sum_{P3} p\left(P7 \mid P3, P4\right) p\left(P3\right)$$

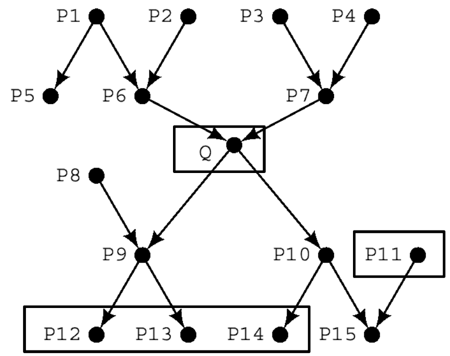
$$p\left(P6 \mid P5\right) = \sum_{P1,P2} p\left(P6 \mid P1, P2\right) p\left(P1 \mid P5\right) p\left(P2\right)$$

- Calculating $p(P1|P5)$
  - Evidence is "below"
  - Here, we use Bayes' rule

$$p(P1 \mid P5) = \frac{p(P5 \mid P1)p(P1)}{p(P5)}$$



Evidence Below (1/2)



$$p\left(Q \mid P12, P13, P14, P11\right) = \frac{p\left(P12, P13, P14, P11 \mid Q\right) p\left(Q\right)}{p\left(P12, P13, P14, P11\right)}$$
$$= kp\left(P12, P13, P14, P11 \mid Q\right) p\left(Q\right)$$
$$= kp\left(P12, P13 \mid Q\right) p\left(P14, P11 \mid Q\right) p\left(Q\right)$$

- Using a top-down recursive algorithm

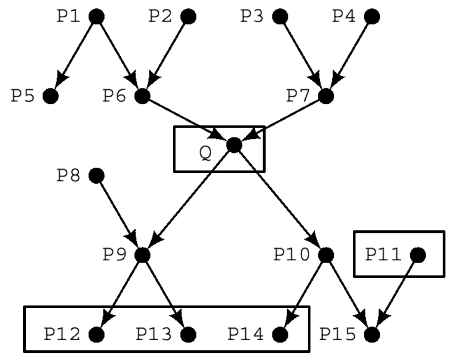
$$p(P12, P13 \mid Q) = \sum_{P9} p(P12, P13 \mid P9, Q) p(p9 \mid Q)$$
$$= \sum_{P9} p(P12, P13 \mid P9) p(P9 \mid Q)$$

(d-separation)

$$p(P9 \mid Q) = \sum_{P8} p(P9 \mid P8, Q) p(P8)$$

$$p(P12, P13 \mid P9) = p(P12 \mid P9) p(P13 \mid P9)$$

Evidence Below (2/2)



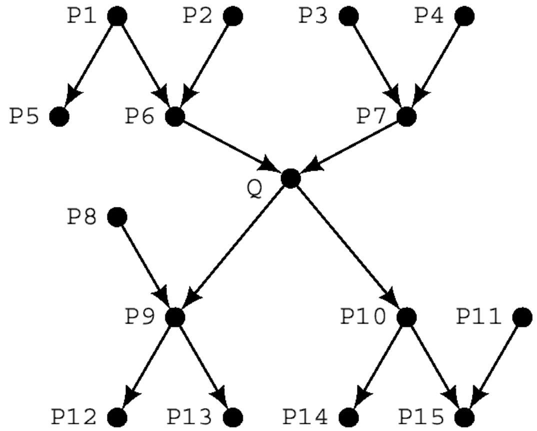
$$p(P14, P11 \mid Q) = \sum_{P10} p(P14, P11 \mid P10) p(P10 \mid Q)$$
$$= \sum_{P10} p(P14 \mid P10) p(P11 \mid P10) p(P10 \mid Q)$$

$$p(P11 \mid P10) = \sum_{P15} p(P11 \mid P15, P10) p(P15 \mid P10)$$
$$p(P15 \mid P10) = \sum_{P11} p(P15 \mid P10, P11) p(P11)$$
$$p(P11 \mid P15, P10) = \frac{p(P15, P10 \mid P11) p(P11)}{p(P15, P10)} = k_1 p(P15, P10 \mid P11) p(P11)$$
$$p(P15, P10 \mid P11) = p(P15 \mid P10, P11) p(P10 \mid P11) = p(P15 \mid P10, P11) p(P10)$$

Evidence Above and Below

$$p(Q \mid \{P5, P4\}, \{P12, P13, P14, P11\})$$

$\mathcal{E}^{+}$ = {P5, P4} $\mathcal{E}^{-}$ = {P12, P13, P14, P11}



$$p\left(Q \mid \mathcal{E}^{+}, \mathcal{E}^{-}\right) = \frac{p\left(\mathcal{E}^{-} \mid Q, \mathcal{E}^{+}\right) p\left(Q \mid \mathcal{E}^{+}\right)}{p\left(\mathcal{E}^{-} \mid \mathcal{E}^{+}\right)}$$

$$= k_2 p\left(\mathcal{E}^{-} \mid Q, \mathcal{E}^{+}\right) p\left(Q \mid \mathcal{E}^{+}\right)$$

$$= k_2 p\left(\mathcal{E}^{-} \mid Q\right) p\left(Q \mid \mathcal{E}^{+}\right)$$

(d-separation)

(We have calculated two probabilities already)

A Numerical Example (1/2)

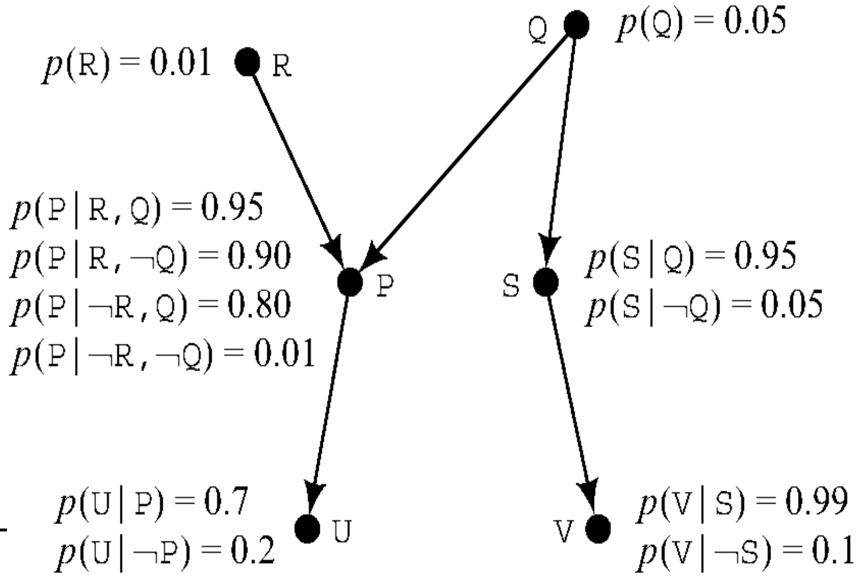
- We want to calculate $p(Q|U)$

$$p(Q|U) = kp(U|Q)p(Q)$$ (Bayes' rule)

$$p(U|Q) = \sum_P p(U|P)p(P|Q)$$

$$p(P|Q) = \sum_R p(P|R,Q)p(R)$$
$$= p(P|R,Q)p(R) + p(P|\neg R,Q)p(\neg R)$$
$$= 0.95 \times 0.01 + 0.8 \times 0.99 = 0.80$$

$$p(\neg P|Q) = 0.20$$

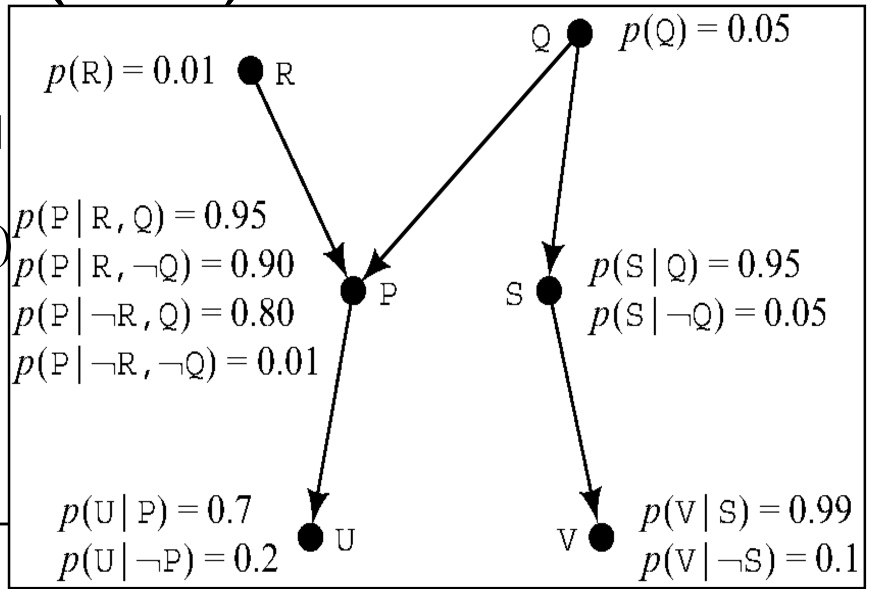


$$p(U|Q) = p(U|P) \times 0.8 + p(U|\neg P) \times 0.2$$
$$= 0.7 \times 0.8 + 0.2 \times 0.2 = 0.60$$

$$p(Q|U) = k \times 0.6 \times 0.05 = k \times 0.03$$

To determine *k*, we need to calculate $p(\neg Q|U)$

A Numerical Example (2/2)



$$p\left(\neg Q \mid U\right) = kp\left(U \mid \neg Q\right) p\left(\neg Q\right)$$ (Bayes' rule)

$$p\left(U \mid \neg Q\right) = \sum_P p(U \mid P)p(P \mid \neg Q)$$

$$p(P \mid \neg Q) = \sum_R p(P \mid R, \neg Q)p(R)$$
$$= p(P \mid R, \neg Q)p(R) + p(P \mid \neg R, \neg Q)p(\neg R)$$
$$= 0.90 \times 0.01 + 0.01 \times 0.99 = 0.019$$

$$p(\neg P \mid Q) = 0.98$$

$$p(U \mid \neg Q) = p(U \mid P) \times 0.019 + p(U \mid \neg P) \times 0.98$$
$$= 0.7 \times 0.19 + 0.2 \times 0.98 = 0.21$$

$$p(\neg Q \mid U) = k \times 0.21 \times 0.95 = k \times 0.20$$

Finally

$$p(Q \mid U) = k \times 0.6 \times 0.05 = k \times 0.03$$
$$p(\neg Q \mid U) = k \times 0.21 \times 0.95 = k \times 0.20$$
$$\therefore k = 4.35, \quad p(Q \mid U) = 4.35 \times 0.03 = 0.13$$

Other methods for Probabilistic inference in Bayes Networks

- Bucket elimination
- Monte Carlo methods (when the network is not a polytree)
- Clustering

Additional Readings (1/5)

- [Feller 1968]
  - Probability Theory
- [Goldszmidt, Morris & Pearl 1990]
  - Non-monotonic inference through probabilistic method
- [Pearl 1982a, Kim & Pearl 1983]
  - Message-passing algorithm
- [Russell & Norvig 1995, pp.447ff]
  - Polytree methods

Additional Readings (2/5)

- [Shachter & Kenley 1989]
  - Bayesian network for continuous random variables
- [Wellman 1990]
  - Qualitative networks
- [Neapolitan 1990]
  - Probabilistic methods in expert systems
- [Henrion 1990]
  - Probability inference in Bayesian networks

Additional Readings (3/5)

- [Jensen 1996]
  - Bayesian networks: HUGIN system
- [Neal 1991]
  - Relationships between Bayesian networks and neural networks
- [Hecherman 1991, Heckerman & Nathwani 1992]
  - PATHFINDER
- [Pradhan, et al. 1994]
  - CPCSBN

Additional Readings (4/5)

- [Shortliffe 1976, Buchanan & Shortliffe 1984]
  - MYCIN: uses certainty factor
- [Duda, Hart & Nilsson 1987]
  - PROSPECTOR: uses sufficiency index and necessity index
- [Zadeh 1975, Zadeh 1978, Elkan 1993]
  - Fuzzy logic and possibility theory
- [Dempster 1968, Shafer 1979]
  - Dempster-Shafer's combination rules

Additional Readings (5/5)

- [Nilsson 1986]
  - Probabilistic logic
- [Tversky & Kahneman 1982]
  - Human generally loses consistency facing uncertainty
- [Shafer & Pearl 1990]
  - Papers for uncertain inference
- Proceedings & Journals
  - Uncertainty in Artificial Intelligence (UAI)
  - International Journal of Approximate Reasoning