Inferences in Bayesian Networks

Yuqing Tang

Doctoral Program in Computer Science
The Graduate Center
City University of New York
ytang@cs.gc.cuny.edu

October 6, 2010

Outline

Inference tasks

- Simple queries: compute posterior marginal $P(X_i|\mathbf{E}=\mathbf{e})$
  e.g., $P(P = low|S = T, X = pos, D = T)$
- Conjunctive queries:
  $P(X_i, X_j|\mathbf{E}=\mathbf{e}) = P(X_i|\mathbf{E}=\mathbf{e}) \times P(X_j|X_i, \mathbf{E}=\mathbf{e})$
- Optimal decisions: decision networks include utility information; probabilistic inference required for $P(outcome|action, evidence)$
- Value of information: which evidence to seek next?
- Sensitivity analysis: which probability values are most critical?
- Explanation: why do I need a new starter motor?

Outline

Inference by enumeration

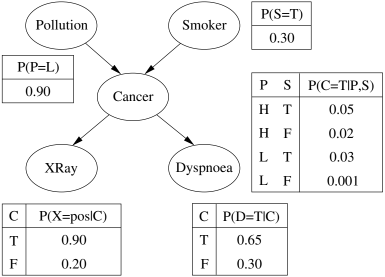
$$
\begin{aligned}
P(X|\mathbf{E}=\mathbf{e}) &= \Sigma_{\mathbf{y}\in Domain(\mathbf{Y})} P(X,\mathbf{Y}|\mathbf{E}=\mathbf{e}) \\
&= \alpha \Sigma_{\mathbf{y}\in Domain(\mathbf{Y})} P(X,\mathbf{Y},\mathbf{E}=\mathbf{e}) \\
P(X_1,X_2|\mathbf{E}=\mathbf{e}) &= \Sigma_{\mathbf{y}\in Domain(\mathbf{Y})} P(X_1,X_2,\mathbf{Y}|\mathbf{E}=\mathbf{e}) \\
&= \alpha \Sigma_{\mathbf{y}\in Domain(\mathbf{Y})} P(X_1,X_2,\mathbf{Y},\mathbf{E}=\mathbf{e})
\end{aligned}
$$

where $\alpha = \frac{1}{P(\mathbf{E}=\mathbf{e})}$.

- Recursive depth-first enumeration: $O(n)$ space, $O(|Domain(X_i)|^n)$ time

Inference by enumeration (example)

Simple query on the lung cancer example



$$
\begin{aligned}
& P(P|X=pos, D=T) \\
&\quad = \quad P(P, X=pos, D=T)/P(X=pos, D=T) \\
&\quad = \quad \alpha P(B, X=pos, D=T) \\
&\quad = \quad \alpha \, \Sigma_{v_c \in \{T,F\}} \, \Sigma_{v_s \in \{T,F\}} \, P(P, S=v_s, C=v_c, X=pos, D=T)
\end{aligned}
$$

where $\alpha = \frac{1}{P(X=pos, D=T)}$ is the normalizer.

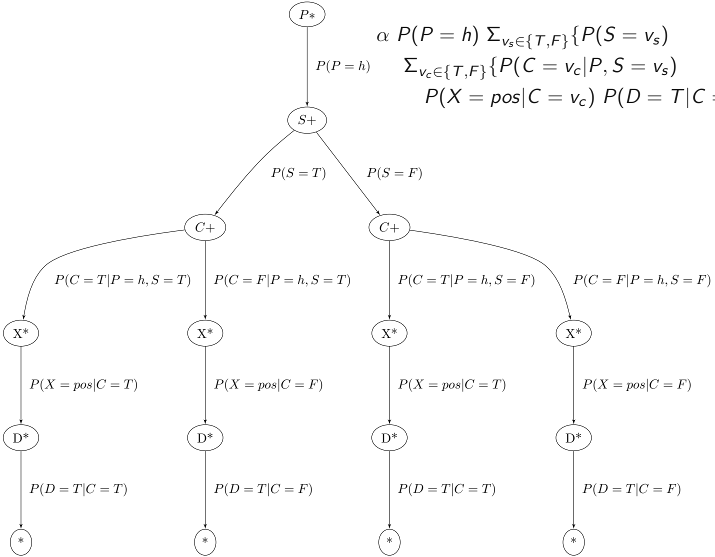
Inference by enumeration (example cont.)

Rewrite full joint entries using product of CPT entries:

$$
\begin{aligned}
& P(P|X = pos, D = T) \\
& \quad = \quad \alpha \; \Sigma_{v_c \in \{T,F\}} \; \Sigma_{v_s \in \{T,F\}} \{ \\
& \qquad\qquad P(P) \times P(S = v_s) \\
& \qquad\qquad \times P(C = v_c | P, S = v_s) \\
& \qquad\qquad \times P(X = pos | C = v_c) \times P(D = T | C = v_c) \} \\
& \quad = \quad \alpha \times P(P) \\
& \qquad\qquad \times \; \Sigma_{v_s \in \{T,F\}} \{ \\
& \qquad\qquad\quad P(S = v_s) \\
& \qquad\qquad\quad \times \; \Sigma_{v_c \in \{T,F\}} \{ P(C = v_c | P, S = v_s) \\
& \qquad\qquad\qquad \times P(X = pos | C = v_c) \\
& \qquad\qquad\qquad \times P(D = T | C = v_c) \} \\
& \qquad\qquad \}
\end{aligned}
$$

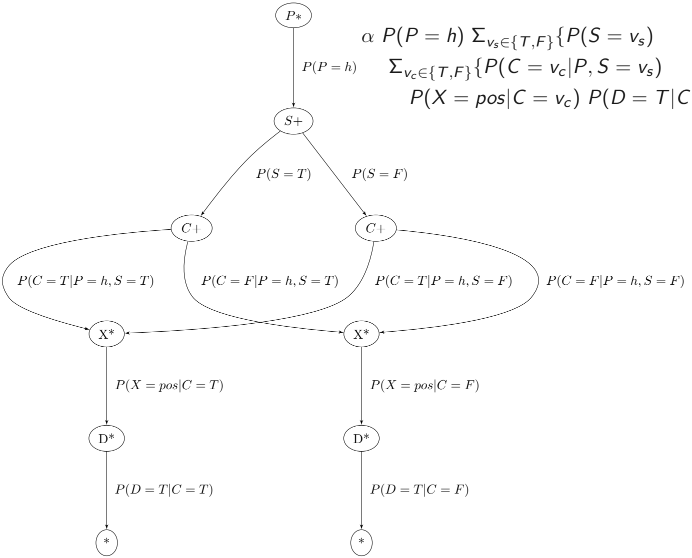
Evaluation tree

$$\alpha\ P(P=h)\ \Sigma_{v_s \in \{T,F\}} \{P(S=v_s)\ \Sigma_{v_c \in \{T,F\}} \{P(C=v_c|P,S=v_s)\ P(X=pos|C=v_c)\ P(D=T|C=v_c)\}\}$$



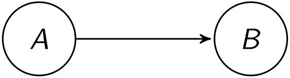
Reduced evaluation tree

$$\alpha \; P(P=h) \; \Sigma_{v_s \in \{T,F\}} \{ P(S=v_s) \\ \Sigma_{v_c \in \{T,F\}} \{ P(C=v_c | P, S=v_s) \\ P(X=pos | C=v_c) \; P(D=T | C=v_c) \} \}$$



Outline

Inference for two node network



- $P(B|A)$ can be read from the CPTs
- $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ by Bayes' rule
  - $P(A), P(B|A)$ can be read from the CPTs
  - $\alpha = \frac{1}{P(B)}$ can be computed by $\Sigma_{a \in Domain(A)} P(A = a|B) = 1$.

Inference for three node chain network I

$$A \longrightarrow B \longrightarrow C$$

- $P(B|A)$ and $P(C|B)$ can be read from the tables
- $P(A|B)$ and $P(C|B)$ can be computed with the Bayes' rule
- $P(C|A)$ can be computed by $\Sigma_{b \in Domain(B)} P(B|A)P(C|B,A) = \Sigma_{b \in Domain(B)} P(B|A)P(C|B)$
- $P(A|C)$ can be computed by $P(A|C) = \alpha P(C|A)\ P(A)$
- $P(C|A,B) = P(C|B)$ since C is independent of A given B
- $P(B|A,C) = \alpha P(C|A,B)P(B|A) = \alpha P(C|B)P(B|A)$ – first apply Bayes' rule conditional on A, and then simplify $P(C|A,B)$ by the condtional indepenence in the network
- $P(A|B,C) = \alpha P(C|A,B)P(A|B) = \alpha P(C|B)P(A|B)$ – first apply Bayes' rule conditional on B, and then simplify $P(C|A,B)$ into $P(C|B)$; $P(A|B)$ are computed above.

Inference for three node chain network II

$$A \longrightarrow B \longrightarrow C$$

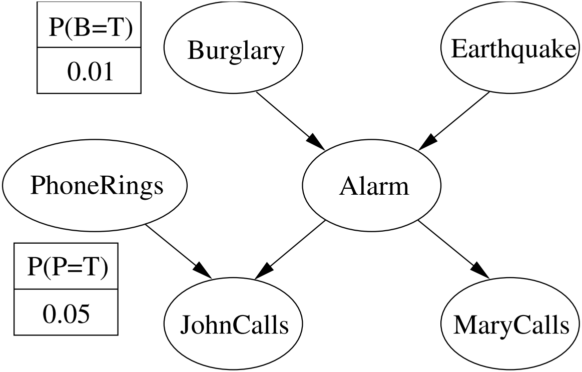
- $P(A, B|C) = \alpha P(C|A, B)P(A, B) = \alpha P(C|B)P(B|A)P(A)$ — first apply Bayes' rule by taking A, B and C as two events, and then simplify.
- $P(A, C|B) = P(A|B)P(C|B) = \alpha P(B|A)P(A)P(C|B)$ — first apply conditional independence of A and C given B, and apply Bayes' rule on $P(A|B)$
- $P(B, C|A) = P(B|A)P(C|A, B) = P(B|A)P(C|B)$ — apply chain rule contional on A, and the conditional independence

Outline

Polytrees (“forest”)

- Polytrees have at most one path between any pair of nodes
- Also called singly-connected networks

Earthquake Example



P(B=T)
0.01

P(E=T)
0.02

B	E	P(A=T|B,E)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

P(P=T)
0.05

P	A	P(J=T|P,A)
T	T	0.95
T	F	0.50
F	T	0.90
F	F	0.01

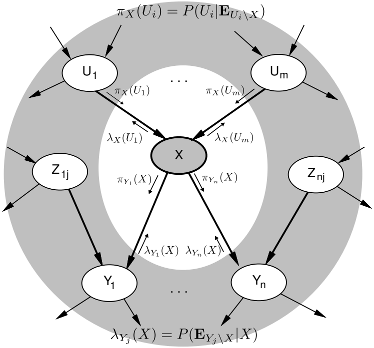
A	P(M=T|A)
T	0.70
F	0.01

The query

$$
\begin{aligned}
P(X|\mathbf{E}) &= \frac{P(X,\mathbf{E})}{P(\mathbf{E})} \\
&= \alpha\, P(X,\mathbf{E})
\end{aligned}
$$

Look at the neighbors, marginalize, and re-organize the variables

$$
\begin{aligned}
& P(X, \mathbf{E}) \\
&= \Sigma_{U_1,\ldots,U_m}\left[P(X, U_1, \mathbf{E}_{U_1\backslash X}, \ldots, U_m, \mathbf{E}_{U_m\backslash X}, \mathbf{E}_{Y_1\backslash X}, \ldots, \mathbf{E}_{Y_n\backslash X})\right] \\
&= \Sigma_{U_1,\ldots,U_m}\left[P(U_1, \mathbf{E}_{U_1\backslash X}, \ldots, U_m, \mathbf{E}_{U_m\backslash X}, X, \mathbf{E}_{Y_1\backslash X}, \ldots, \mathbf{E}_{Y_n\backslash X})\right]
\end{aligned}
$$



Chain rule

$$
\begin{aligned}
&P(U_1, \mathbf{E}_{U_1\setminus X}, \ldots, U_m, \mathbf{E}_{U_m\setminus X}, X, \mathbf{E}_{Y_1\setminus X}, \ldots, \mathbf{E}_{Y_n\setminus X}) = \\
&\quad P(U_1, \mathbf{E}_{U_1\setminus X}) \\
&\quad P(U_2, \mathbf{E}_{U_2\setminus X} | U_1, \mathbf{E}_{U_1\setminus X}) \\
&\quad P(U_3, \mathbf{E}_{U_3\setminus X} | U_1, \mathbf{E}_{U_1\setminus X}, U_2, \mathbf{E}_{U_2\setminus X}) \\
&\quad \ldots \\
&\quad P(U_m, \mathbf{E}_{U_m\setminus X} | U_1, \mathbf{E}_{U_1\setminus X}, \ldots, U_{m-1}, \mathbf{E}_{U_{m-1}\setminus X}) \\
&\quad P(X | U_1, \mathbf{E}_{U_1\setminus X}, \ldots, U_m, \mathbf{E}_{U_m\setminus X}) \\
&\quad P(\mathbf{E}_{Y_1\setminus X} | X, U_1, \mathbf{E}_{U_1\setminus X}, \ldots, U_m, \mathbf{E}_{U_m\setminus X}) \\
&\quad P(\mathbf{E}_{Y_2\setminus X} | X, U_1, \mathbf{E}_{U_1\setminus X}, \ldots, U_m, \mathbf{E}_{U_m\setminus X}, \mathbf{E}_{Y_1\setminus X}) \\
&\quad \ldots \\
&\quad P(\mathbf{E}_{Y_n\setminus X} | X, U_1, \mathbf{E}_{U_1\setminus X}, \ldots, U_m, \mathbf{E}_{U_m\setminus X}, \mathbf{E}_{Y_1\setminus X}, \ldots, \mathbf{E}_{Y_{n-1}\setminus X})
\end{aligned}
$$

Independence/conditional independence

$P(U_1, \mathbf{E}_{U_1\backslash X}, \ldots, U_m, \mathbf{E}_{U_m\backslash X}, X, \mathbf{E}_{Y_1\backslash X}, \ldots, \mathbf{E}_{Y_n\backslash X}) =$

$P(U_1, \mathbf{E}_{U_1\backslash X})$

$//U_i, \mathbf{E}_{U_i\backslash X}$ is independent of $U_1, \mathbf{E}_{U_1\backslash X}, \ldots, U_{i-1}, \mathbf{E}_{U_{i-1}\backslash X}$:

$P(U_2, \mathbf{E}_{U_2\backslash X})$

$\ldots$

$P(U_m, \mathbf{E}_{U_m\backslash X})$

// X is independent of $\mathbf{E}_{U_i\backslash X}$ given U_i ($i = 1, \ldots, m$):

$P(X|U_1, U_2, \ldots, U_m)$

// $\mathbf{E}_{Y_k\backslash X}$ is independent of $U_i s, \mathbf{E}_{U_i\backslash X} s$ and $\mathbf{E}_{Y'_k\backslash X}$ ($k' \neq k$) given X:

$P(\mathbf{E}_{Y_1\backslash X}|X)$

$P(\mathbf{E}_{Y_2\backslash X}|X)$

$\ldots$

$P(\mathbf{E}_{Y_n\backslash X}|X)$

One more re-write with the definition of conditionals

$$
\begin{aligned}
& P(U_1, \mathbf{E}_{U_1\setminus X}, \ldots, U_m, \mathbf{E}_{U_m\setminus X}, X, \mathbf{E}_{Y_1\setminus X}, \ldots, \mathbf{E}_{Y_n\setminus X}) \\
= & \\
& P(U_1|\mathbf{E}_{U_1\setminus X})\ P(\mathbf{E}_{U_1\setminus X}) \\
& P(U_2|\mathbf{E}_{U_2\setminus X})\ P(\mathbf{E}_{U_2\setminus X}) \\
& \ldots \\
& P(U_m|\mathbf{E}_{U_m\setminus X})\ P(\mathbf{E}_{U_m\setminus X}) \\
& P(X|U_1, U_2, \ldots, U_m) \\
& P(\mathbf{E}_{Y_1\setminus X}|X) \\
& P(\mathbf{E}_{Y_2\setminus X}|X) \\
& \ldots \\
& P(\mathbf{E}_{Y_n\setminus X}|X) \\
= & \ P(X|U_1, U_2, \ldots, U_m) \\
& \cdot \Pi_{i=1,\ldots,m} P(\mathbf{E}_{U_i\setminus X}) \cdot \Pi_{i=1,\ldots,m} P(U_i|\mathbf{E}_{U_i\setminus X}) \cdot \Pi_{k=1,\ldots,n} P(\mathbf{E}_{Y_k\setminus X}|X)
\end{aligned}
$$

Back to the query

$$
\begin{aligned}
& P(X|\mathbf{E}) \\
= \quad & \alpha \Sigma_{U_1,\ldots,U_m} \{ \\
& P(U_1, \mathbf{E}_{U_1 \backslash X}, \ldots, U_m, \mathbf{E}_{U_m \backslash X}, X, \mathbf{E}_{Y_1 \backslash X}, \ldots, \mathbf{E}_{Y_n \backslash X}) \} \\
= \quad & \alpha \Sigma_{U_1,\ldots,U_m} \{ P(X|U_1, U_2, \ldots, U_m) \\
& \Pi_{i=1,\ldots,m} P(\mathbf{E}_{U_i \backslash X}) \cdot \Pi_{i=1,\ldots,m} P(U_i|\mathbf{E}_{U_i \backslash X}) \cdot \Pi_{k=1,\ldots,n} P(\mathbf{E}_{Y_k \backslash X}|X) \} \\
= \quad & \alpha \cdot \Pi_{i=1,\ldots,m} P(\mathbf{E}_{U_i \backslash X}) \cdot \Sigma_{U_1,\ldots,U_m} \{ P(X|U_1, U_2, \ldots, U_m) \\
& \Pi_{i=1,\ldots,m} P(U_i|\mathbf{E}_{U_i \backslash X}) \cdot \Pi_{k=1,\ldots,n} P(\mathbf{E}_{Y_k \backslash X}|X) \}
\end{aligned}
$$

As $\Pi_{i=1,\ldots,m} P(\mathbf{E}_{U_i \backslash X})$ doesn't change when X takes different values, we can put it into the normalizing constant α:

$$
\begin{aligned}
P(X|\mathbf{E}) = \alpha \Sigma_{U_1,\ldots,U_m} \{ & P(X|U_1, U_2, \ldots, U_m) \\
& \Pi_{i=1,\ldots,m} P(U_i|\mathbf{E}_{U_i \backslash X}) \cdot \Pi_{k=1,\ldots,n} P(\mathbf{E}_{Y_k \backslash X}|X) \}
\end{aligned}
$$

Recursively compute $P(U_i|\mathbf{E}_{U_i\backslash X})$ and $P(\mathbf{E}_{Y_k}|X)$

$$P(X|\mathbf{E}) = \alpha\Sigma_{U_1,\ldots,U_m}\{ P(X|U_1,U_2,\ldots,U_m)\, \Pi_{i=1,\ldots,m}P(U_i|\mathbf{E}_{U_i\backslash X})\, \Pi_{k=1,\ldots,n}P(\mathbf{E}_{Y_k\backslash X}|X)\}$$

- $P(X|U_1,U_2,\ldots,U_m)$ can be looked from the CPTs
- Let $BN^{\pi}_{X}(U_i)$ be a sub-BN composed of U_i and all the nodes connecting to X through U_i.
  - Within $BN^{\pi}_{X}(U_i)$, $P(U_i|\mathbf{E}_{U_i\backslash X})$ can be computed recursively in the same manner
- Let $BN^{\lambda}_{Y_k}(X)$ be a sub-BN composed of Y_k and all the nodes connecting to X through Y_k.
  - Within $BN^{\lambda}_{Y_k}(X)$, $P(\mathbf{E}_{Y_k\backslash X}|X) = \Sigma_{Y_k}P(Y_k,\mathbf{E}_{Y_k\backslash X}|X)$ can be computed recursively in the same manner

Recursively compute $P(U_i|\mathbf{E}_{U_i\backslash X})$

Within $BN^{\pi}_X(U_i)$,

$$P(U_i|\mathbf{E}_{U_i\backslash X})) = \alpha\Sigma_{Parents^{BN^{\pi}_X(U_i)}(U_i)} \{ \\ P(U_i|Parents^{BN^{\pi}_X(U_i)}(U_i)) \cdot \\ \Pi_{Z_j \in Parents^{BN^{\pi}_X(U_i)}(U_i)} P(Z_j|\mathbf{E}_{Z_j\backslash U_i}) \cdot \\ \Pi_{Y_k \in Children^{BN^{\pi}_X(U_i)}(U_i)} P(\mathbf{E}_{Y_k\backslash X}|U_i) \Big\}$$

- $Parents^{BN_X(U_i)}(U_i)$ are the parents of U_i in $BN_X(U_i)$
- $Children^{BN_X(U_i)}(U_i)$ are the children of U_i in $BN_X(U_i)$
- When U_i is one of the evidence input
  - $P(U_i = u_{i,e}|\mathbf{E}_{U_i\backslash X}) = 1$ if $u_{i,e}$ is the evidence value entered
  - $P(U_i = u_{i,e}|\mathbf{E}_{U_i\backslash X}) = 0$ if $u_{i,e}$ is not the evidence value entered

Recursively compute $P(\mathbf{E}_{Y_k}|X)$ I

Take $P(\mathbf{E}_{Y_k}|X)$ as a joint probability on the evidences connecting to X through Y_k and all the probabilities in $BN^{\lambda}_{Y_k}(X)$ as conditional on X

$$P(\mathbf{E}_{Y_k \backslash X}|X) = \Sigma_{Y_k} P(Y_k, \mathbf{E}_{Y_k \backslash X}|X)$$

$$P(Y_k, \mathbf{E}_{Y_k \backslash X}|X) = \alpha \Sigma_{Parents^{BN^{\lambda}_{Y_k}(X)}(Y_k), Children^{BN^{\lambda}_{Y_k}(X)}(Y_k)} \Big\{ P(Y_k|Parents^{BN^{\lambda}_{Y_k}(X)}(Y_k), X) \cdot \prod_{Z_{k,j} \in Parents^{BN^{\lambda}_{Y_k}(X)}(Y_k)} P(Z_{k,j}|\mathbf{E}_{Z_{k,j} \backslash Y_k}, X) \cdot \prod_{W_{k,l} \in Children^{BN^{\lambda}_{Y_k}(X)}(Y_k)} P(\mathbf{E}_{W_{k,l} \backslash Y_k}|Y_k, X) \Big\}$$

- Observations
  - $P(Y_k|Parents^{BN^{\lambda}_{Y_k}(X)}(Y_k), X)$ is in the CPTs

Recursively compute $P(\mathbf{E}_{Y_k}|X)$ II

- $P(Z_{k,j}|\mathbf{E}_{Z_{k,j}\backslash Y_k}, X) = P(Z_{k,j}|\mathbf{E}_{Z_{k,j}\backslash Y_k})$ by conditional independence ($Z_{k,j}$ and X are marginal independence and there is no path between $Z_{k,j}$ and X through $\mathbf{E}_{Z_{k,j}}$)
- $P(\mathbf{E}_{W_{k,l}\backslash Y_k}|Y_k, X) = P(\mathbf{E}_{W_{k,l}\backslash Y_k}|Y_k)$ by conditional independence (Given Y_k, $\mathbf{E}_{W_{k,l}}$ and X are independence)

$$P(\mathbf{E}_{Y_k\backslash X}|X) = \alpha \Sigma_{Y_k} \Sigma_{Parents^{BN^{\lambda}_{Y_k}(X)}(Y_k), Children^{BN^{\lambda}_{Y_k}(X)}(Y_k)} \Big\{ P(Y_k|Parents^{BN^{\lambda}_{Y_k}(X)}(Y_k), X) \cdot \prod_{Z_{k,j} \in Parents^{BN^{\lambda}_{Y_k}(X)}(Y_k)} P(Z_{k,j}|\mathbf{E}_{Z_{k,j}\backslash Y_k}) \cdot \prod_{W_{k,l} \in Children^{BN^{\lambda}_{Y_k}(X)}(Y_k)} P(\mathbf{E}_{W_{k,l}\backslash Y_k}|Y_k) \Big\}$$

Recursively compute $P(\mathbf{E}_{Y_k}|X)$ III

- When Y_k is one of the evidence input, let $y_{k,e}$ be the evidence value entered, the marginalization over Y_k is replaced by setting $Y_k = y_{k,e}$ in the calculation.

$$
\begin{aligned}
P(\mathbf{E}_{Y_k \setminus X}|X) = \alpha \Sigma_{Parents^{BN^{\lambda}_{Y_k}(X)}(Y_k), Children^{BN^{\lambda}_{Y_k}(X)}(Y_k)} \{ \\
P(Y_k = y_{k,e}|Parents^{BN^{\lambda}_{Y_k}(X)}(Y_k), X) \cdot \\
\Pi_{Z_{k,j} \in Parents^{BN^{\lambda}_{Y_k}(X)}(Y_k)} P(Z_{k,j}|\mathbf{E}_{Z_{k,j} \setminus Y_k}) \cdot \\
\Pi_{W_{k,l} \in Children^{BN^{\lambda}_{Y_k}(X)}(Y_k)} P(\mathbf{E}_{W_{k,l} \setminus Y_k}|Y_k = y_{k,e}) \}
\end{aligned}
$$

Shorten the notation for $\pi_X(U_i)$

Let

$$\begin{aligned}\pi_X(U_i) &= P(U_i|\mathbf{E}_{U_i\backslash X}) \\ \lambda_{Y_k}(X) &= P(\mathbf{E}_{Y_k\backslash X}|X)\end{aligned}$$

$$\begin{aligned}&\pi_X(U_i) \\ &= P(U_i|\mathbf{E}_{U_i\backslash X}) \\ &= \alpha\Sigma_{Parents^{BN_X^{\pi}(U_i)}(U_i)}\{ \\ &\quad P(U_i|Parents^{BN_X^{\pi}(U_i)}(U_i))\cdot \\ &\quad \Pi_{Z_j\in Parents^{BN_X^{\pi}(U_i)}(U_i)}\pi_{U_i}(Z_j)\cdot \\ &\quad \Pi_{Y_k\in Children^{BN_X^{\pi}(U_i)}(U_i)}\lambda_{Y_k}(U_i)\}\end{aligned}$$

When U_i is one of the evidence input

- $\pi_X(U_i = u_{i,e}) = 1$ if $u_{i,e}$ is the evidence value entered
- $\pi_X(U_i = u_{i,e}) = 0$ if $u_{i,e}$ is not the evidence value entered

Shorten the notation for $\pi_X(U_i)$ (cont.)

Let

$$\begin{aligned} \pi_X(U_i) &= P(U_i|\mathbf{E}_{U_i \setminus X})) \\ \lambda_{Y_k}(X) &= P(\mathbf{E}_{Y_k \setminus X}|X) \end{aligned}$$

$$\begin{aligned} &\pi_X(U_i) \\ &\quad = P(U_i|\mathbf{E}_{U_i \setminus X})) \\ &\quad = \alpha \Sigma_{Parents(U_i)} \{ \\ &\qquad P(U_i|Parents(U_i)) \cdot \\ &\qquad \Pi_{Z_j \in Parents(U_i)} \pi_{U_i}(Z_j) \cdot \\ &\qquad \Pi_{Y_k \in Children(U_i) \setminus \{X\}} \lambda_{Y_k}(U_i) \} \end{aligned}$$

When U_i is one of the evidence input

- $\pi_X(U_i = u_{i,e}) = 1$ if $u_{i,e}$ is the evidence value entered
- $\pi_X(U_i = u_{i,e}) = 0$ if $u_{i,e}$ is not the evidence value entered

Shorten the notation for $\lambda_{Y_k}(X)$

Let

$$\begin{aligned}\pi_X(U_i) &= P(U_i|\mathbf{E}_{U_i\backslash X}))\\ \lambda_{Y_k}(X) &= P(\mathbf{E}_{Y_k\backslash X}|X)\end{aligned}$$

$$\begin{aligned}&\lambda_{Y_k}(X)\\ &= \Sigma_{Y_k} P(Y_k, \mathbf{E}_{Y_k\backslash X}|X)\\ &= \Sigma_{Y_k} \Sigma_{Parents^{BN^{\lambda}_{Y_k}(X)}(Y_k)} \{\\ &\quad P(Y_k|Parents^{BN^{\lambda}_{Y_k}(X)}(Y_k), X)\cdot\\ &\quad \prod_{Z_{k,j}\in Parents^{BN^{\lambda}_{Y_k}(X)}(Y_k)} \pi_{Y_k}(Z_{k,j})\cdot\\ &\quad \prod_{W_{k,l}\in Children^{BN^{\lambda}_{Y_k}(X)}(Y_k)} \lambda_{W_{k,l}}(Y_k)\Big\}\end{aligned}$$

When Y_k is one of the evidence input, let $y_{k,e}$ be the evidence value entered, the marginalization over Y_k is replaced by setting $Y_k = y_{k,e}$ in the calculation.

Shorten the notation for $\lambda_{Y_k}(X)$ (cont.) I

Let

$$\begin{aligned}\pi_X(U_i) &= P(U_i|\mathbf{E}_{U_i\setminus X}) \\ \lambda_{Y_k}(X) &= P(\mathbf{E}_{Y_k\setminus X}|X)\end{aligned}$$

$$\begin{aligned}&\lambda_{Y_k}(X) \\ &= \Sigma_{Y_k} P(Y_k, \mathbf{E}_{Y_k\setminus X}|X) \\ &= \Sigma_{Y_k}\Sigma_{Parents(Y_k)\setminus\{X\}} \{ \\ &\quad P(Y_k|Parent(Y_k)) \cdot \\ &\quad \Pi_{Z_{k,j}\in Parents(Y_k)\setminus\{X\}} \pi_{Y_k}(Z_{k,j}) \cdot \\ &\quad \Pi_{W_{k,l}\in Children(Y_k)} \lambda_{W_{k,l}}(Y_k) \}\end{aligned}$$

Shorten the notation for $\lambda_{Y_k}(X)$ (cont.) II

When Y_k is one of the evidence input, let $y_{k,e}$ be the evidence value entered, the marginalization over Y_k is replaced by setting $Y_k = y_{k,e}$ in the calculation.

$$
\begin{aligned}
&\lambda_{Y_k}(X) \\
&\quad = \quad P(Y_k = y_{k,e}, \mathbf{E}_{Y_k \setminus X} | X) \\
&\quad = \quad \Sigma_{Parents(Y_k) \setminus \{X\}} \{ \\
&\qquad\quad P(Y_k = y_{k,e} | Parent(Y_k)) \cdot \\
&\qquad\quad \Pi_{Z_{k,j} \in Parents(Y_k) \setminus \{X\}} \pi_{Y_k}(Z_{k,j}) \cdot \\
&\qquad\quad \Pi_{W_{k,l} \in Children(Y_k)} \lambda_{W_{k,l}}(Y_k) \}
\end{aligned}
$$

Message passing (bottom-up instead of recursion) I

- Query decomposition

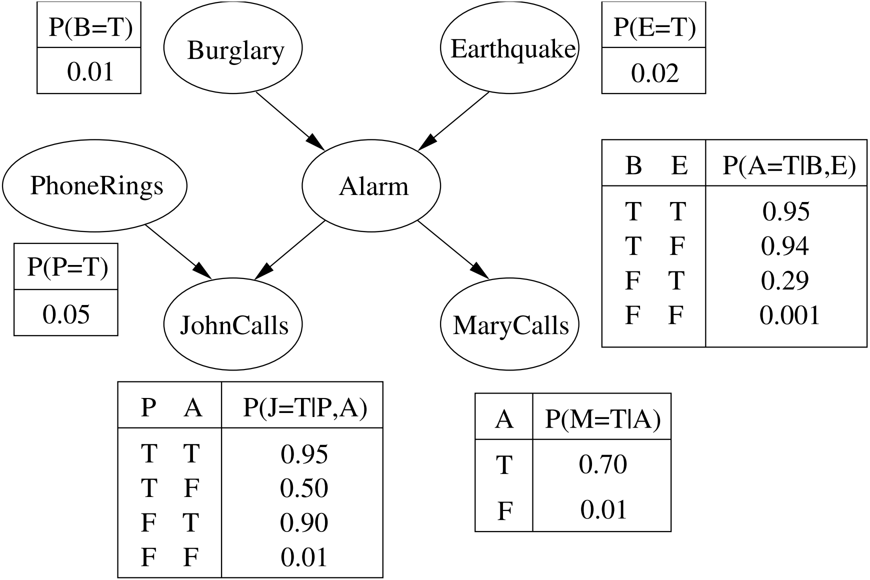
$$
\begin{aligned}
P(X|\mathbf{E}) = \alpha\Sigma_{U_1,\ldots,U_m}\{ & \\
& P(X|U_1, U_2, \ldots, U_m) \cdot \\
& \Pi_{i=1,\ldots,m}\pi_{U_i}(X) \cdot \\
& \Pi_{k=1,\ldots,n}\lambda_{Y_k}(X)\}
\end{aligned}
$$

- Message passing:
  - Start with the nodes that don't need any messages to compute π
  - Start with the nodes that don't need any messages to compute λ
  - Once a node gets the π and λ messages required to compute its own messages sent to its children or parents, compute the messages and send them out. Notice that
    - ★ U_i send to its child X: $\pi_X(U_i)$ — computing $\pi_X(U_i)$ only requires messages from nodes in $BN_X^{\pi}(U_i)$
    - ★ Y_k send to its parent X: $\lambda_{Y_k}(X)$ — computing $\lambda_{Y_k}(X)$ only requires messages from nodes in $BN_{Y_k}^{\lambda}(X)$

Message passing (bottom-up instead of recursion) II

- X computes query $P(X|\mathbf{E})$ using its conditional table $P(X|Parents(X))$, and the message $\pi_{U_i}(X)$ received from its parents, the message $\lambda_{Y_k}(X)$ received from its children
- For queries other than $P(X|\mathbf{E})$, X send messages
  - X send message $\pi_{Y_k}(X)$ to its child Y_k
  - X send message $\lambda_X(U_i)$ to its parent U_i

Earthquake Example



P(B=T)
0.01

P(E=T)
0.02

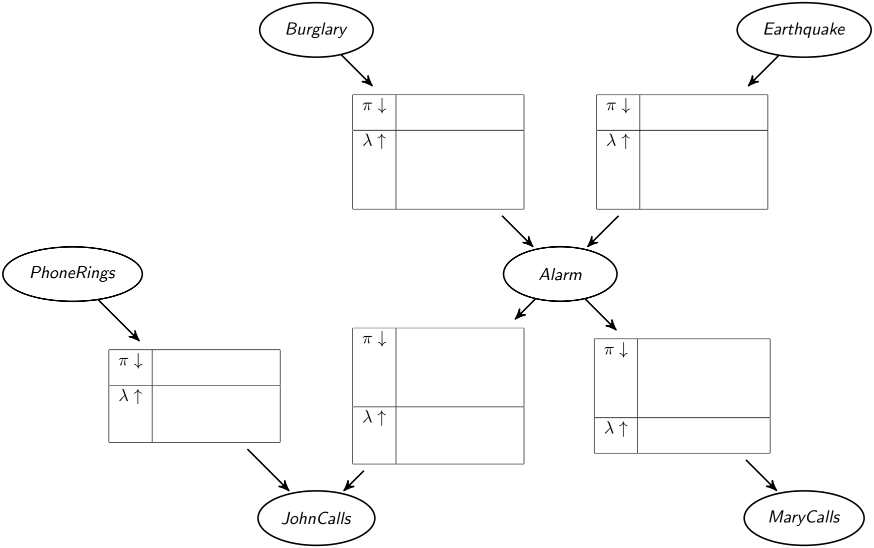
P(P=T)
0.05

B	E	P(A=T|B,E)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

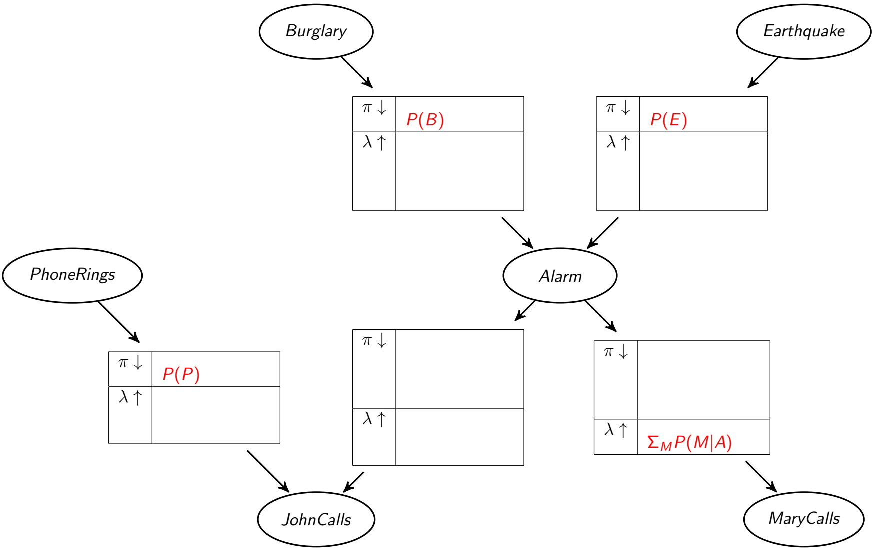
P	A	P(J=T|P,A)
T	T	0.95
T	F	0.50
F	T	0.90
F	F	0.01

A	P(M=T|A)
T	0.70
F	0.01

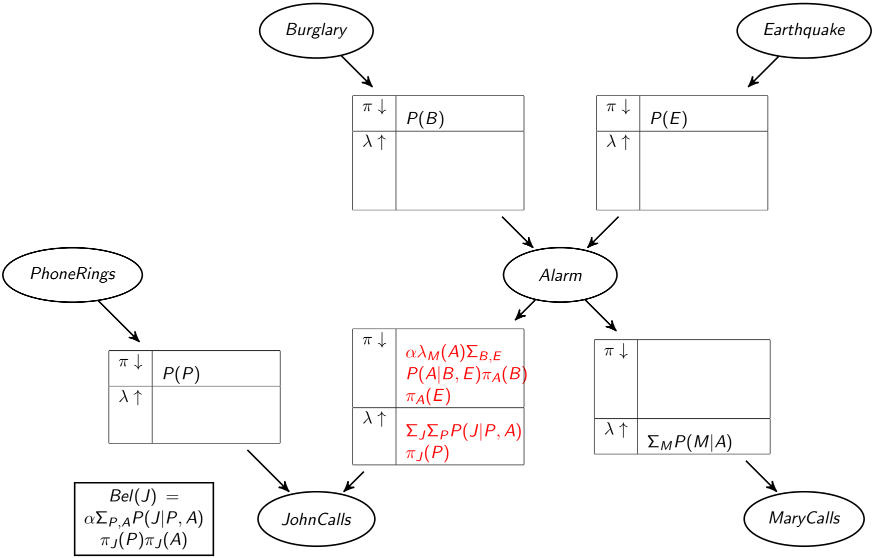
Message passing – No Evidence



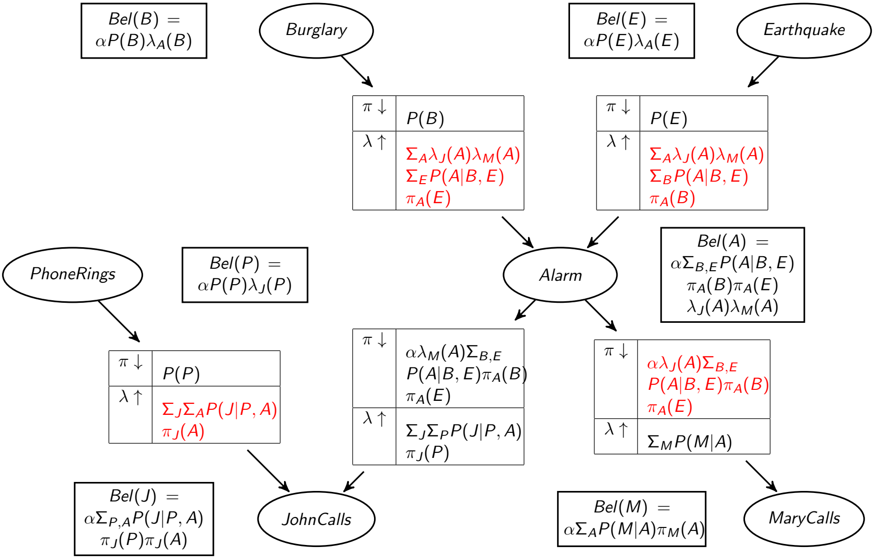
Message passing – No Evidence



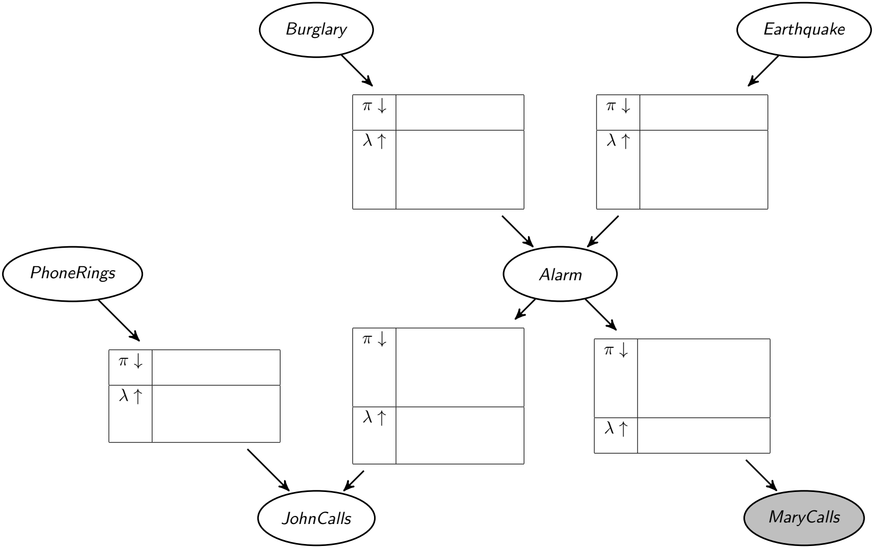
Message passing – No Evidence



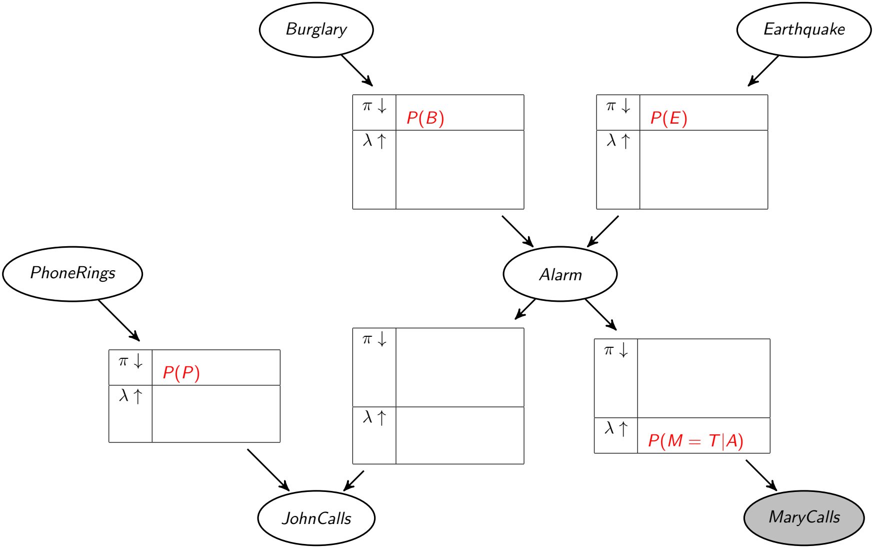
Message passing – No Evidence



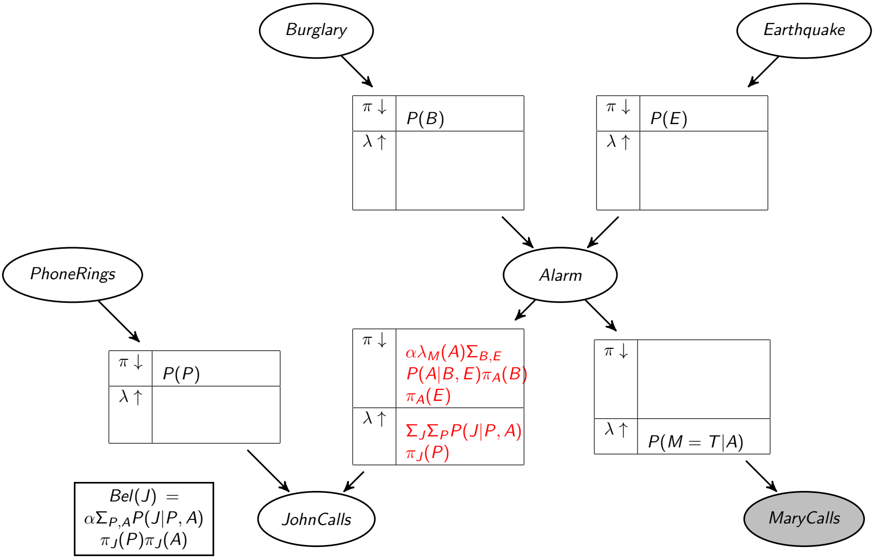
Message passing – Evidence Example



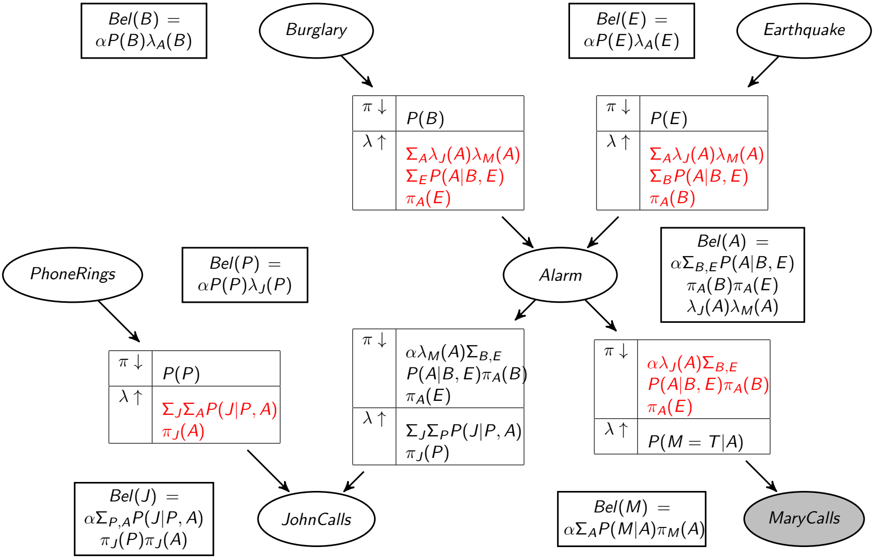
Message passing – Evidence Example



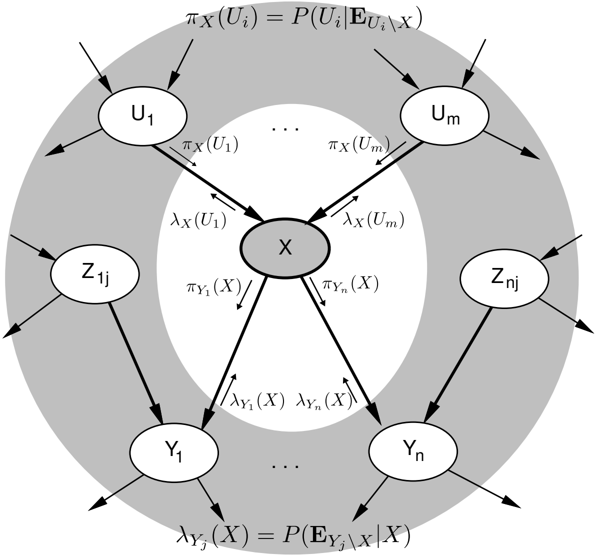
Message passing – Evidence Example



Message passing – Evidence Example



Overview of message passing



Message passing complexity

- Rounds of message passing: The maxium length of paths from X to roots (bounded by the number variables in the BN)
- Number of summations: It can be exponentional in the number of parents and children of each node

Outline

Summary

- Probabilistic inference: compute the probability distribution for query variables, given evidence variables

$$P(X|\mathbf{E})$$

- BN Inference is very flexible: can enter evidence about any node and update beliefs in any other nodes.
- The speed of inference in practice depends on the structure of the network: how many loops; numbers of parents; location of evidence and query nodes.
- Inference methods
  - Exact inference by enumeration
  - Polytree message passing
  - Junction tree algorithms for general BNs (next lecture)
  - Approximation inference with stochastic simulation (next lecture)

Acknowledgments

Lecture 5 is composed the instructor's own understanding of the subject, and materials from [Korb and Nicholson, 2003, Chapter 3] and [Russell and Norvig, 2009, Chapter 16] with the instructor's own interpretations. The instructor takes full responsibility of any mistakes in the slides.

References I

- K. Korb and A. E. Nicholson.
  Bayesian Artificial Intelligence.
  Chapman & Hall /CRC, 2003.

- Stuart J. Russell and Peter Norvig.
  Artificial Intelligence: A Modern Approach (3nd Edition).
  Prentice Hall, December 2009.