

CS 1571 Introduction to AI

Lecture 20

Inferences in Bayesian belief networks

Milos Hauskrecht
milos@cs.pitt.edu
5329 Sennott Square

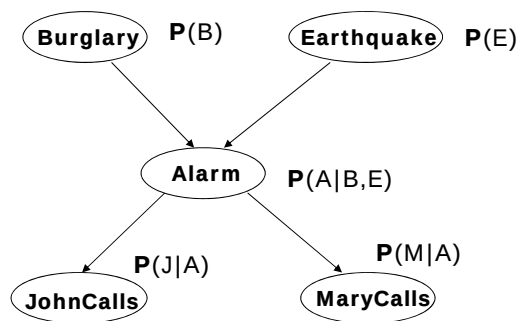
CS 1571 Intro to AI

M. Hauskrecht

Bayesian belief network.

1. Directed acyclic graph

- **Nodes** = random variables
Burglary, Earthquake, Alarm, Mary calls and John calls
- **Links** = direct (causal) dependencies between variables.
The chance of Alarm is influenced by Earthquake, The chance of John calling is affected by the Alarm



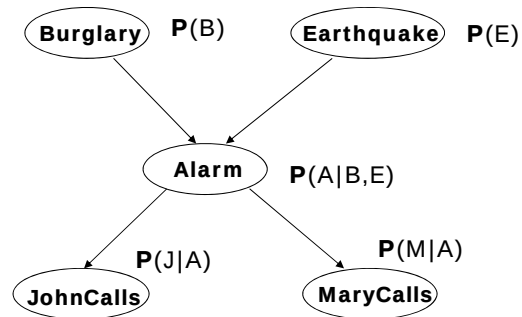
CS 1571 Intro to AI

M. Hauskrecht

Bayesian belief network.

2. Local conditional distributions

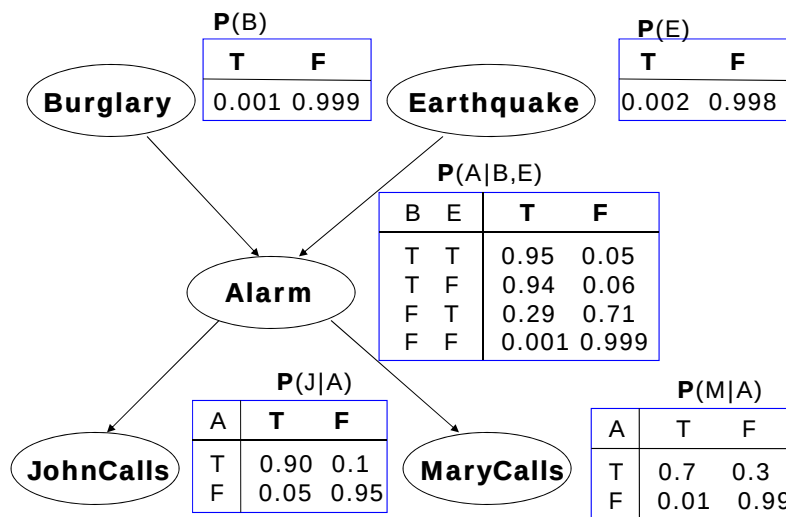
- relate variables and their parents



CS 1571 Intro to AI

M. Hauskrecht

Bayesian belief network.



CS 1571 Intro to AI

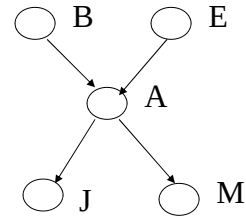
M. Hauskrecht

Bayesian belief networks (general)

Two components: $B = (S, \Theta_S)$

- **Directed acyclic graph**

- Nodes correspond to random variables
- (Missing) links encode independences



- **Parameters**

- Local conditional probability distributions for every variable-parent configuration

$$P(X_i | pa(X_i))$$

Where:

$pa(X_i)$ - stand for parents of X_i

$P(A|B,E)$

B	E	T	F
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

Full joint distribution in BBNs

Full joint distribution is defined in terms of local conditional distributions (obtained via the chain rule):

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1, \dots, n} P(X_i | pa(X_i))$$

Example:

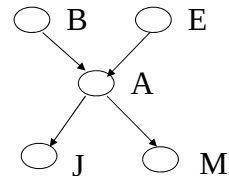
Assume the following assignment of values to random variables

$$B=T, E=T, A=T, J=T, M=F$$

Then its probability is:

$$P(B=T, E=T, A=T, J=T, M=F) =$$

$$P(B=T)P(E=T)P(A=T | B=T, E=T)P(J=T | A=T)P(M=F | A=T)$$



Parameter complexity problem

- In the BBN the **full joint distribution** is defined as:

$$\mathbf{P}(X_1, X_2, \dots, X_n) = \prod_{i=1, \dots, n} \mathbf{P}(X_i \mid pa(X_i))$$

- What did we save?

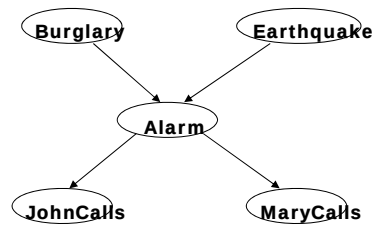
Alarm example: 5 binary (True, False) variables

of parameters of the full joint:

$$2^5 = 32$$

One parameter is for free:

$$2^5 - 1 = 31$$



Parameter complexity problem

- In the BBN the **full joint distribution** is defined as:

$$\mathbf{P}(X_1, X_2, \dots, X_n) = \prod_{i=1, \dots, n} \mathbf{P}(X_i \mid pa(X_i))$$

- What did we save?

Alarm example: 5 binary (True, False) variables

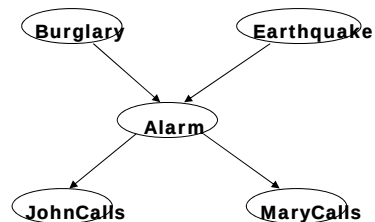
of parameters of the full joint:

$$2^5 = 32$$

One parameter is for free:

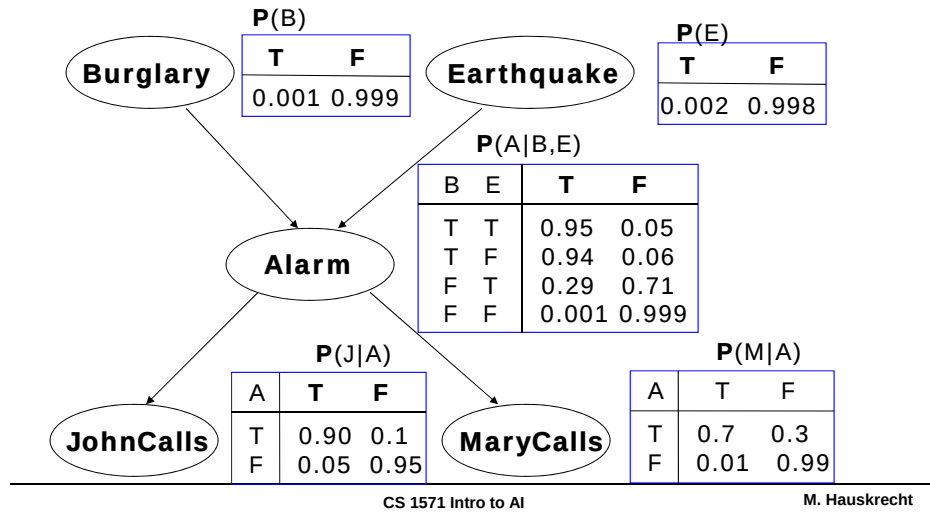
$$2^5 - 1 = 31$$

of parameters of the BBN: ?



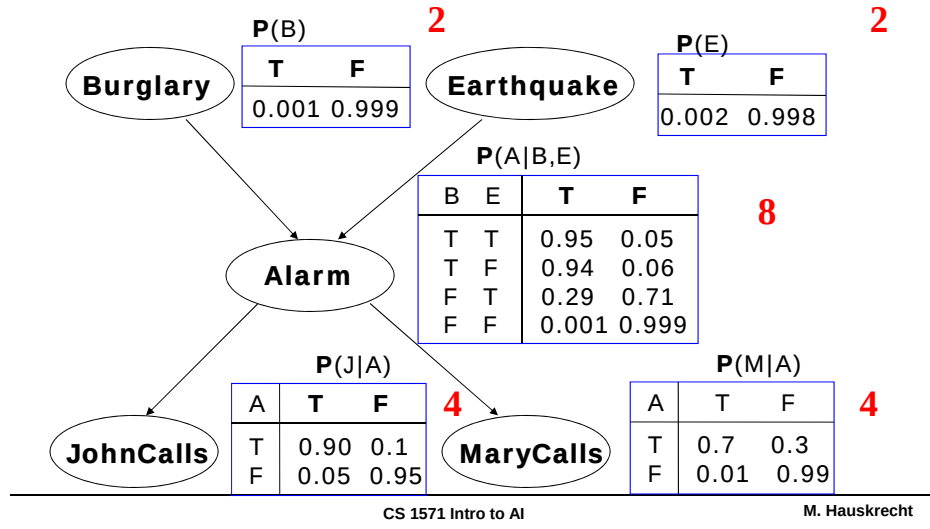
Bayesian belief network.

- In the BBN the **full joint distribution** is expressed using a set of local conditional distributions



Bayesian belief network.

- In the BBN the **full joint distribution** is expressed using a set of local conditional distributions



Parameter complexity problem

- In the BBN the **full joint distribution** is defined as:

$$\mathbf{P}(X_1, X_2, \dots, X_n) = \prod_{i=1, \dots, n} \mathbf{P}(X_i \mid \text{pa}(X_i))$$

- What did we save?

Alarm example: 5 binary (True, False) variables

of parameters of the full joint:

$$2^5 = 32$$

One parameter is for free:

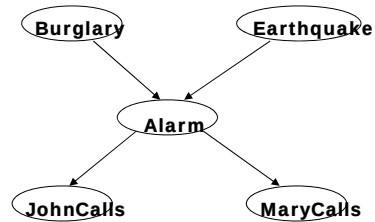
$$2^5 - 1 = 31$$

of parameters of the BBN:

$$2^3 + 2(2^2) + 2(2) = 20$$

One parameter in every conditional is for free:

?



Parameter complexity problem

- In the BBN the **full joint distribution** is defined as:

$$\mathbf{P}(X_1, X_2, \dots, X_n) = \prod_{i=1, \dots, n} \mathbf{P}(X_i \mid \text{pa}(X_i))$$

- What did we save?

Alarm example: 5 binary (True, False) variables

of parameters of the full joint:

$$2^5 = 32$$

One parameter is for free:

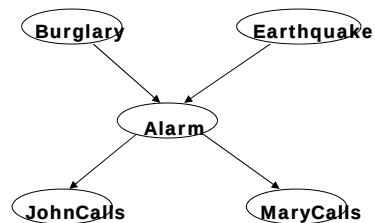
$$2^5 - 1 = 31$$

of parameters of the BBN:

$$2^3 + 2(2^2) + 2(2) = 20$$

One parameter in every conditional is for free:

$$2^2 + 2(2) + 2(1) = 10$$



BBNs built in practice

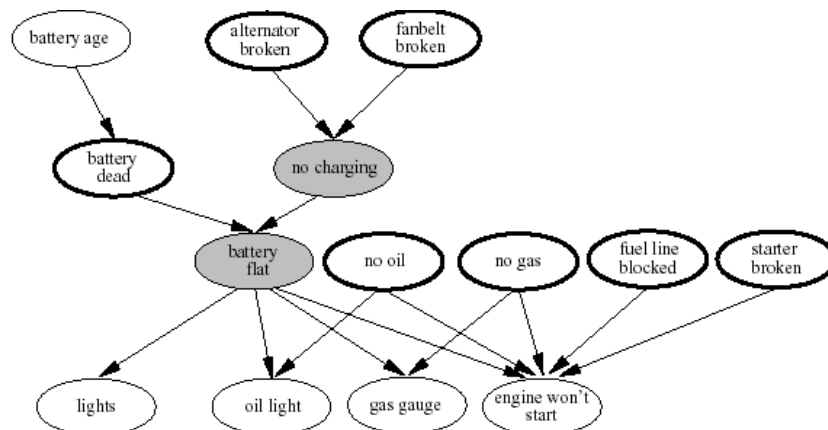
- In various areas:
 - Intelligent user interfaces (Microsoft)
 - Troubleshooting, diagnosis of a technical device
 - Medical diagnosis:
 - Pathfinder (Intellipath)
 - CPSC
 - Munin
 - QMR-DT
 - Collaborative filtering
 - Military applications
 - Business and finance
 - Insurance, credit applications

CS 1571 Intro to AI

M. Hauskrecht

Diagnosis of car engine

- Diagnose the engine start problem

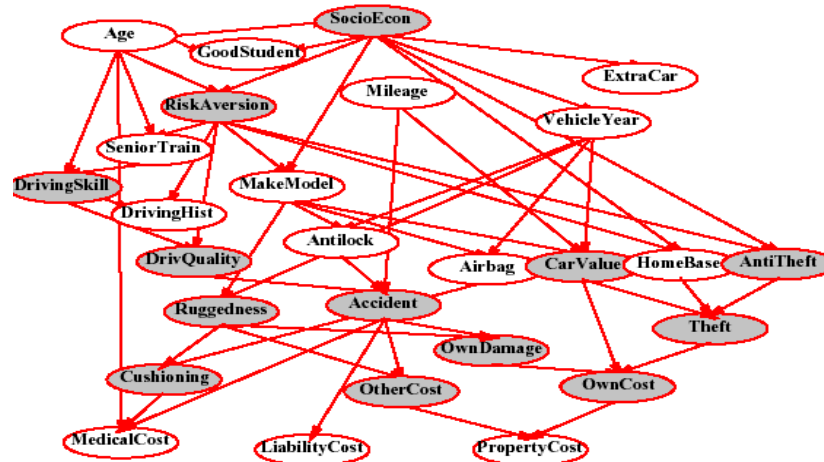


CS 1571 Intro to AI

M. Hauskrecht

Car insurance example

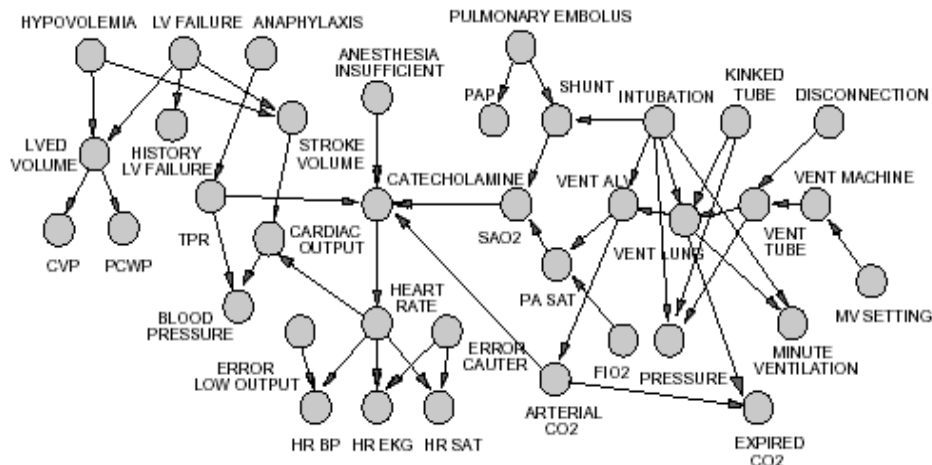
- Predict claim costs (medical, liability) based on application data



CS 1571 Intro to AI

M. Hauskrecht

(ICU) Alarm network

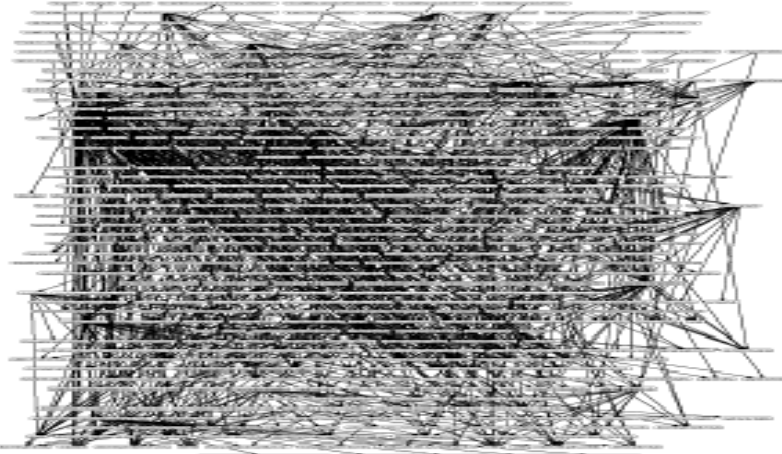


CS 1571 Intro to AI

M. Hauskrecht

CPCS

- Computer-based Patient Case Simulation system (CPCS-PM) developed by Parker and Miller (University of Pittsburgh)
- 422 nodes and 867 arcs



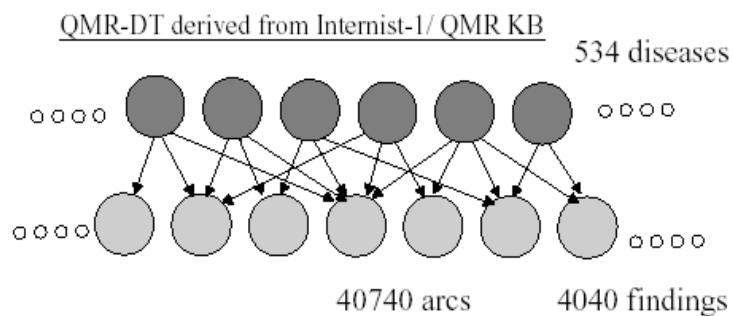
CS 1571 Intro to AI

M. Hauskrecht

QMR-DT

- Medical diagnosis in internal medicine

Bipartite network of disease/findings relations

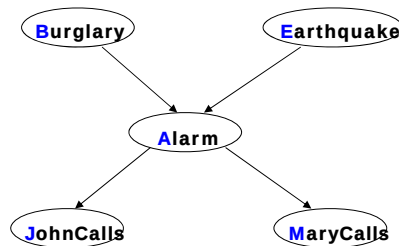


CS 1571 Intro to AI

M. Hauskrecht

Inference in Bayesian network

- **Bad news:**
 - Exact inference problem in BBNs is NP-hard (Cooper)
 - Approximate inference is NP-hard (Dagum, Luby)
- **But** very often we can achieve significant improvements
- Assume our Alarm network



- Assume we want to compute: $P(J = T)$

CS 1571 Intro to AI

M. Hauskrecht

Inference in Bayesian networks

Computing: $P(J = T)$

Approach 1. Blind approach.

- Sum out all un-instantiated variables from the full joint,
- express the joint distribution as a product of conditionals

$$\begin{aligned}
 P(J = T) &= \\
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} \sum_{m \in T, F} P(B = b, E = e, A = a, J = T, M = m) \\
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} \sum_{m \in T, F} P(J = T | A = a) P(M = m | A = a) P(A = a | B = b, E = e) P(B = b) P(E = e)
 \end{aligned}$$

Computational cost:

Number of additions: ?

Number of products: ?

CS 1571 Intro to AI

M. Hauskrecht

Inference in Bayesian networks

Computing: $P(J = T)$

Approach 1. Blind approach.

- Sum out all un-instantiated variables from the full joint,
- express the joint distribution as a product of conditionals

$$\begin{aligned}
 P(J = T) &= \\
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} \sum_{m \in T, F} P(B = b, E = e, A = a, J = T, M = m) \\
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} \sum_{m \in T, F} P(J = T | A = a) P(M = m | A = a) P(A = a | B = b, E = e) P(B = b) P(E = e)
 \end{aligned}$$

Computational cost:

Number of additions: 15

Number of products: $16 \cdot 4 = 64$

Inference in Bayesian networks

Approach 2. Interleave sums and products

- Combines sums and product in a smart way (multiplications by constants can be taken out of the sum)

$$\begin{aligned}
 P(J = T) &= \\
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} \sum_{m \in T, F} P(J = T | A = a) P(M = m | A = a) P(A = a | B = b, E = e) P(B = b) P(E = e) \\
 &= \sum_{b \in T, F} \sum_{a \in T, F} \sum_{m \in T, F} P(J = T | A = a) P(M = m | A = a) P(B = b) \left[\sum_{e \in T, F} P(A = a | B = b, E = e) P(E = e) \right] \\
 &= \sum_{a \in T, F} P(J = T | A = a) \left[\sum_{m \in T, F} P(M = m | A = a) \right] \left[\sum_{b \in T, F} P(B = b) \left[\sum_{e \in T, F} P(A = a | B = b, E = e) P(E = e) \right] \right]
 \end{aligned}$$

Computational cost:

Number of additions: $1 + 2 \cdot [1 + 1 + 2 \cdot 1] = ?$

Number of products: $2 \cdot [2 + 2 \cdot (1 + 2 \cdot 1)] = ?$

Inference in Bayesian networks

Approach 2. Interleave sums and products

- Combines sums and product in a smart way (multiplications by constants can be taken out of the sum)

$$\begin{aligned}
 P(J=T) &= \\
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} \sum_{m \in T, F} P(J=T | A=a) P(M=m | A=a) P(A=a | B=b, E=e) P(B=b) P(E=e) \\
 &= \sum_{b \in T, F} \sum_{a \in T, F} \sum_{m \in T, F} P(J=T | A=a) P(M=m | A=a) P(B=b) \left[\sum_{e \in T, F} P(A=a | B=b, E=e) P(E=e) \right] \\
 &= \sum_{a \in T, F} P(J=T | A=a) \left[\sum_{m \in T, F} P(M=m | A=a) \right] \left[\sum_{b \in T, F} P(B=b) \left[\sum_{e \in T, F} P(A=a | B=b, E=e) P(E=e) \right] \right]
 \end{aligned}$$

Computational cost:

Number of additions: $1+2*[1+1+2*1]=9$

Number of products: $2*[2+2*(1+2*1)]=?$

Inference in Bayesian networks

Approach 2. Interleave sums and products

- Combines sums and product in a smart way (multiplications by constants can be taken out of the sum)

$$\begin{aligned}
 P(J=T) &= \\
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} \sum_{m \in T, F} P(J=T | A=a) P(M=m | A=a) P(A=a | B=b, E=e) P(B=b) P(E=e) \\
 &= \sum_{b \in T, F} \sum_{a \in T, F} \sum_{m \in T, F} P(J=T | A=a) P(M=m | A=a) P(B=b) \left[\sum_{e \in T, F} P(A=a | B=b, E=e) P(E=e) \right] \\
 &= \sum_{a \in T, F} P(J=T | A=a) \left[\sum_{m \in T, F} P(M=m | A=a) \right] \left[\sum_{b \in T, F} P(B=b) \left[\sum_{e \in T, F} P(A=a | B=b, E=e) P(E=e) \right] \right]
 \end{aligned}$$

Computational cost:

Number of additions: $1+2*[1+1+2*1]=9$

Number of products: $2*[2+2*(1+2*1)]=16$

Variable elimination

- **Variable elimination:**

- Similar idea but interleave sum and products one variable at the time during inference
- E.g. Query $P(J=T)$ requires to eliminate A,B,E,M and this can be done in different order

$$P(J=T) =$$

$$= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} \sum_{m \in T, F} P(J=T | A=a) P(M=m | A=a) P(A=a | B=b, E=e) P(B=b) P(E=e)$$

CS 1571 Intro to AI

M. Hauskrecht

Variable elimination

Assume order: M, E, B, A to calculate $P(J=T)$

$$= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} \sum_{m \in T, F} P(J=T | A=a) P(M=m | A=a) P(A=a | B=b, E=e) P(B=b) P(E=e)$$

$$= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} P(J=T | A=a) P(A=a | B=b, E=e) P(B=b) P(E=e) \left[\sum_{m \in T, F} P(M=m | A=a) \right]$$

$$= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} P(J=T | A=a) P(A=a | B=b, E=e) P(B=b) P(E=e) \quad 1$$

$$= \sum_{a \in T, F} \sum_{b \in T, F} P(J=T | A=a) P(B=b) \left[\sum_{e \in T, F} P(A=a | B=b, E=e) P(E=e) \right]$$

$$= \sum_{a \in T, F} \sum_{b \in T, F} P(J=T | A=a) P(B=b) \tau_1(A=a, B=b)$$

$$= \sum_{a \in T, F} P(J=T | A=a) \left[\sum_{b \in T, F} P(B=b) \tau_1(A=a, B=b) \right]$$

$$= \sum_{a \in T, F} P(J=T | A=a) \tau_2(A=a)$$

CS 1571 Intro to AI

M. Hauskrecht

Inference in Bayesian network

- **Exact inference algorithms:**

- ➡

Book

 - **Variable elimination**
 - **Recursive decomposition** (Cooper, Darwiche)
 - Symbolic inference (D'Ambrosio)
 - Belief propagation algorithm (Pearl)
- ➡

Book

 - **Clustering and joint tree approach** (Lauritzen, Spiegelhalter)
 - Arc reversal (Olmsted, Schachter)

- **Approximate inference algorithms:**

- ➡

Book

 - **Monte Carlo methods:**
 - Forward sampling, Likelihood sampling
 - Variational methods

CS 1571 Intro to AI

M. Hauskrecht

Monte Carlo approaches

- **MC approximation:**

- The probability is approximated using sample frequencies
- **Example:**

$$\tilde{P}(B = T, J = T) = \frac{N_{B=T, J=T}}{N}$$

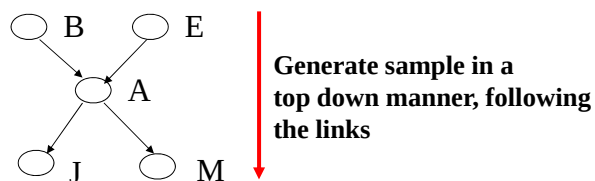
←

samples with $B = T, J = T$

←

total # samples

- **BBN sampling:**

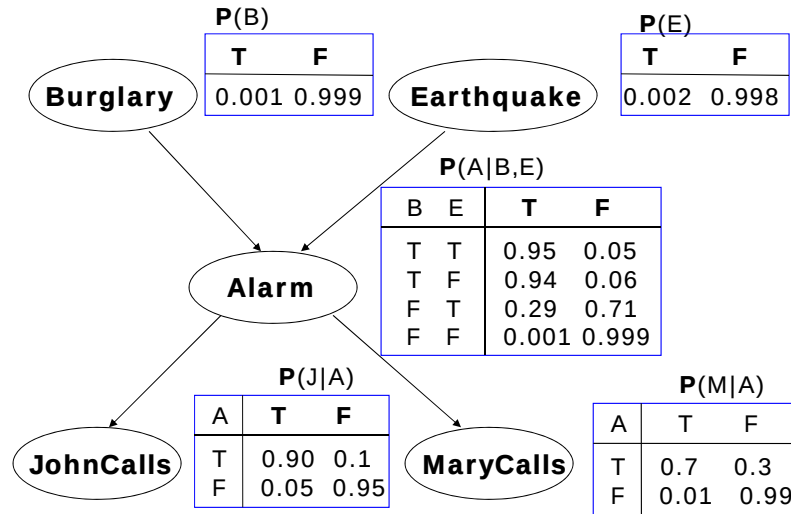


- **One sample gives one assignment of values to all variables**

CS 1571 Intro to AI

M. Hauskrecht

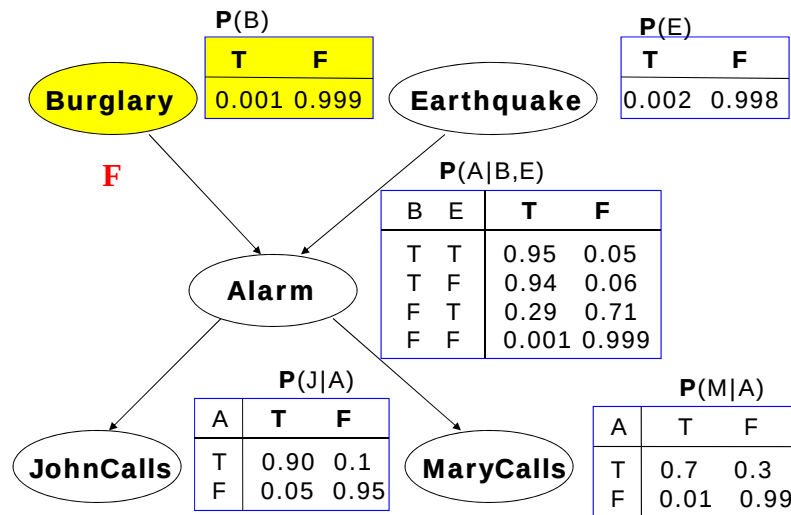
BBN sampling example



CS 1571 Intro to AI

M. Hauskrecht

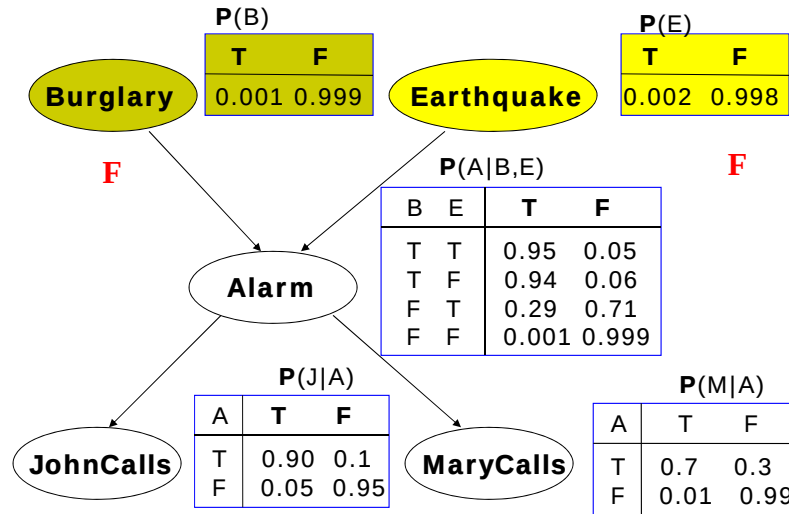
BBN sampling example



CS 1571 Intro to AI

M. Hauskrecht

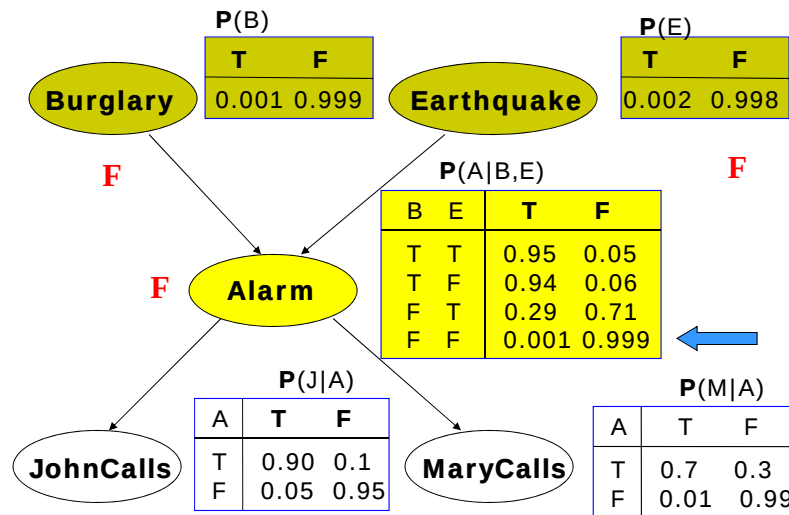
BBN sampling example



CS 1571 Intro to AI

M. Hauskrecht

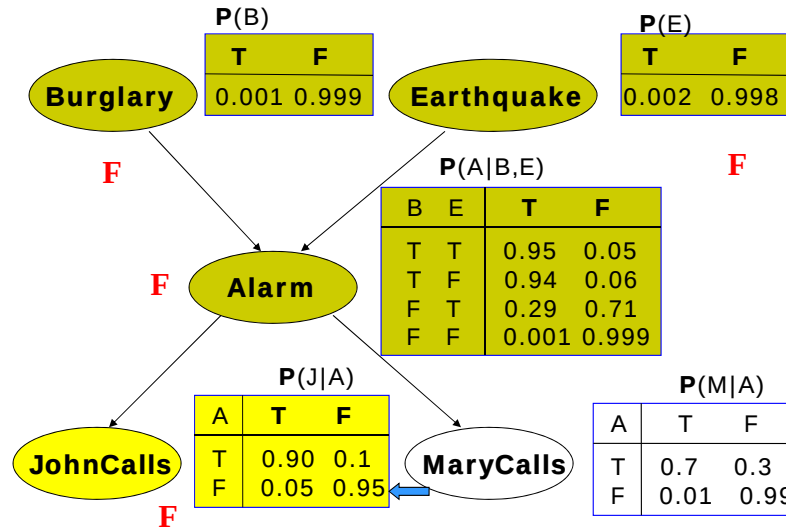
BBN sampling example



CS 1571 Intro to AI

M. Hauskrecht

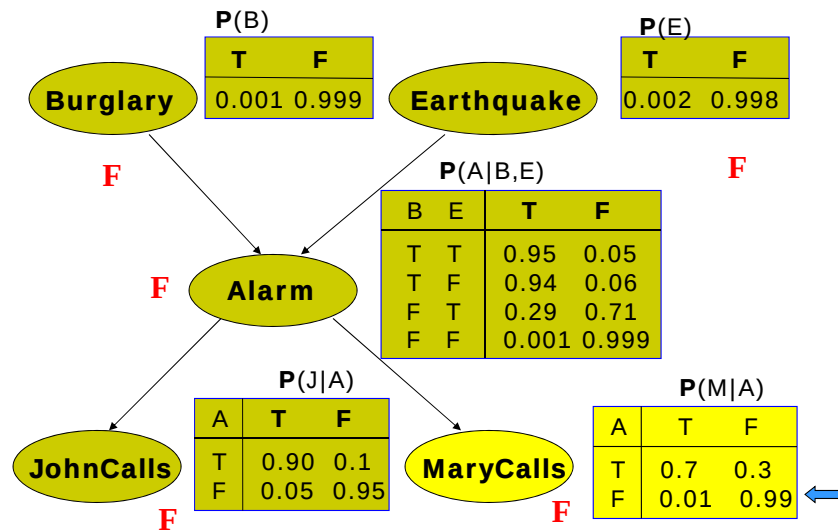
BBN sampling example



CS 1571 Intro to AI

M. Hauskrecht

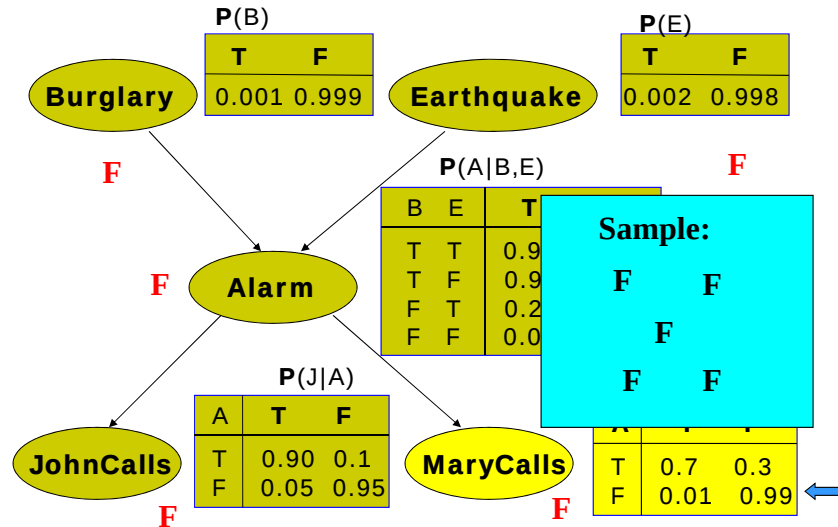
BBN sampling example



CS 1571 Intro to AI

M. Hauskrecht

BBN sampling example



CS 1571 Intro to AI

M. Hauskrecht

Monte Carlo approaches

- **MC approximation of conditional probabilities:**

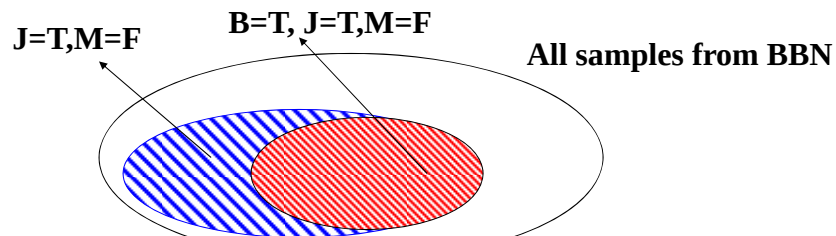
- The probability is approximated using sample frequencies

- **Example:**

$$\tilde{P}(B = T \mid J = T, M = F) = \frac{N_{B=T, J=T, M=F}}{N_{J=T, M=F}}$$

samples with $B = T, J = T, M = F$

samples with $J = T, M = F$

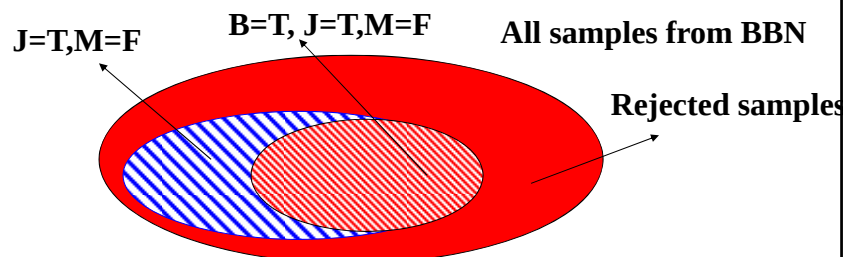


CS 1571 Intro to AI

M. Hauskrecht

Monte Carlo approaches

- **Rejection sampling**
 - Generate samples from the full joint by sampling BBN
 - Use only samples that agree with the condition, the remaining samples are rejected
- **Problem:** many samples can be rejected



CS 1571 Intro to AI

M. Hauskrecht

Likelihood weighting

- Idea:** generate only samples consistent with an evidence (or conditioning event)
- Benefit: Avoids inefficiencies of rejection sampling

Problem:

- the distribution generated by enforcing the conditioning variables to set values is biased
- simple counts are not sufficient to estimate the probabilities

Solution:

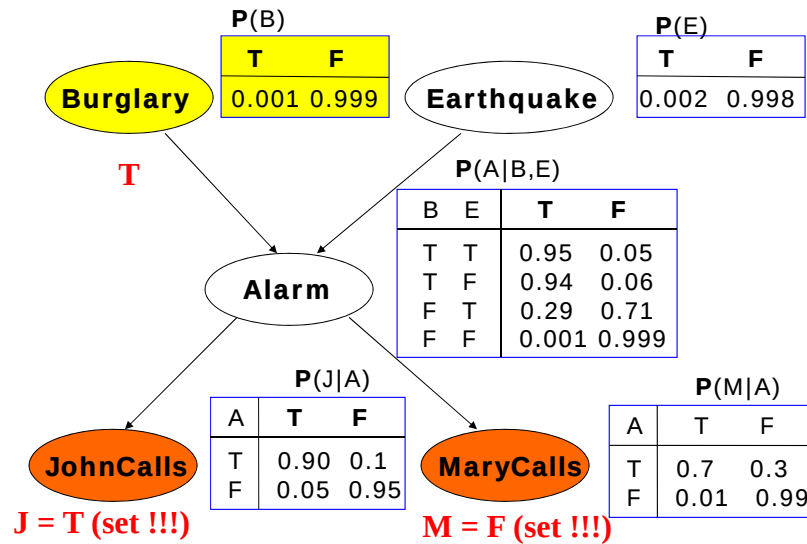
- With every sample keep a weight with which it should count towards the estimate

$$\tilde{P}(B = T \mid J = T, M = F) = \frac{\sum_{\text{samples with } B=T, M=F \text{ and } J=T} w_{B=T \mid J=T, M=F}}{\sum_{\text{samples with any value of } B \text{ and } J=T, M=F} w_{B=x \mid J=T, M=F}}$$

CS 1571 Intro to AI

M. Hauskrecht

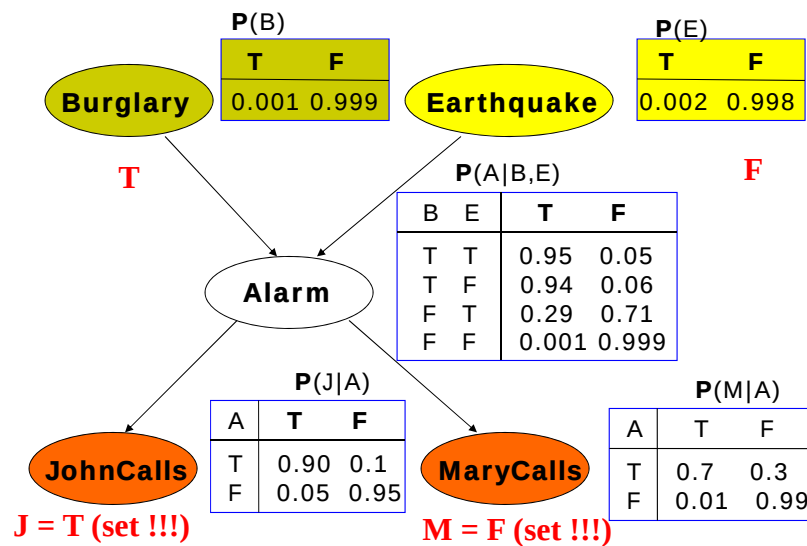
BBN likelihood weighting example



CS 1571 Intro to AI

M. Hauskrecht

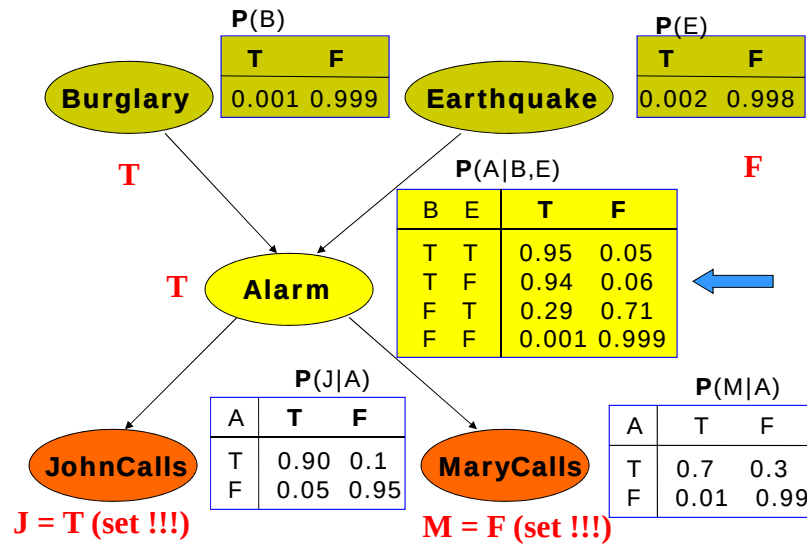
BBN likelihood weighting example



CS 1571 Intro to AI

M. Hauskrecht

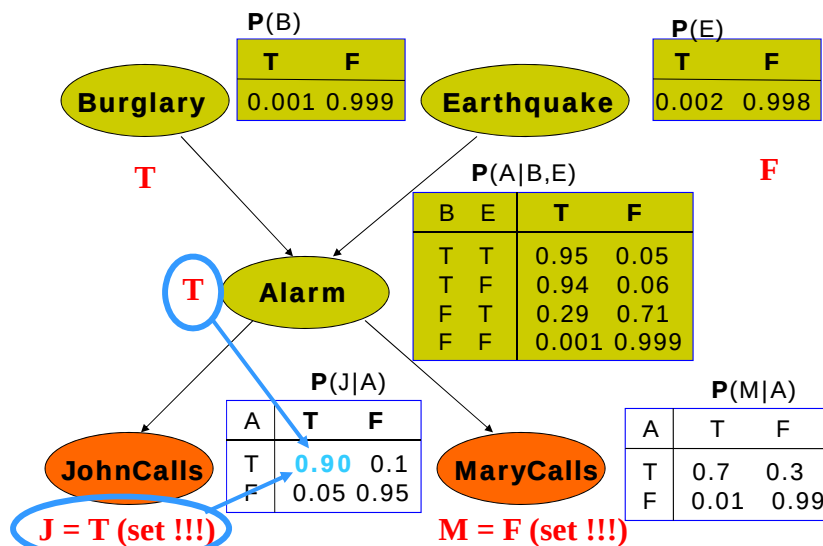
BBN likelihood weighting example



CS 1571 Intro to AI

M. Hauskrecht

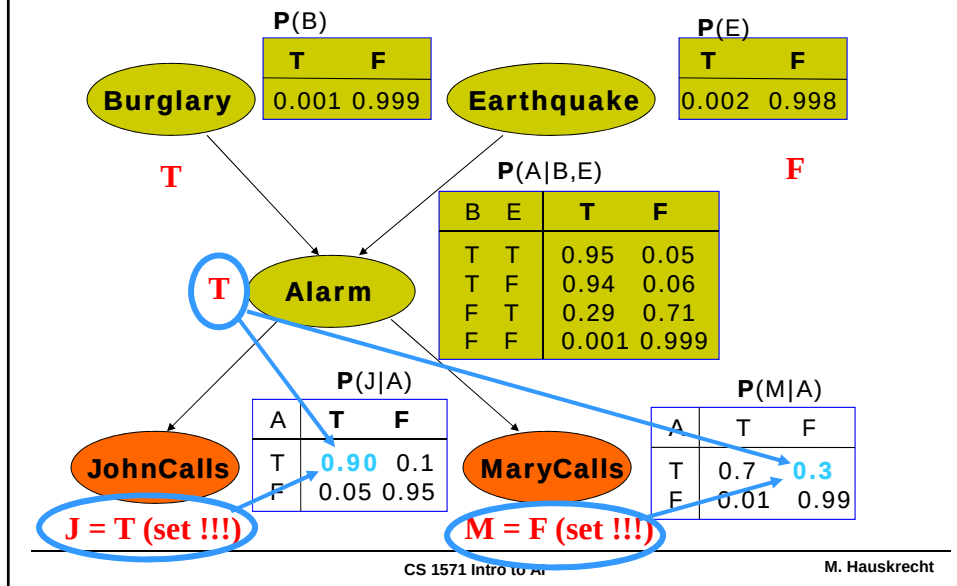
BBN likelihood weighting example



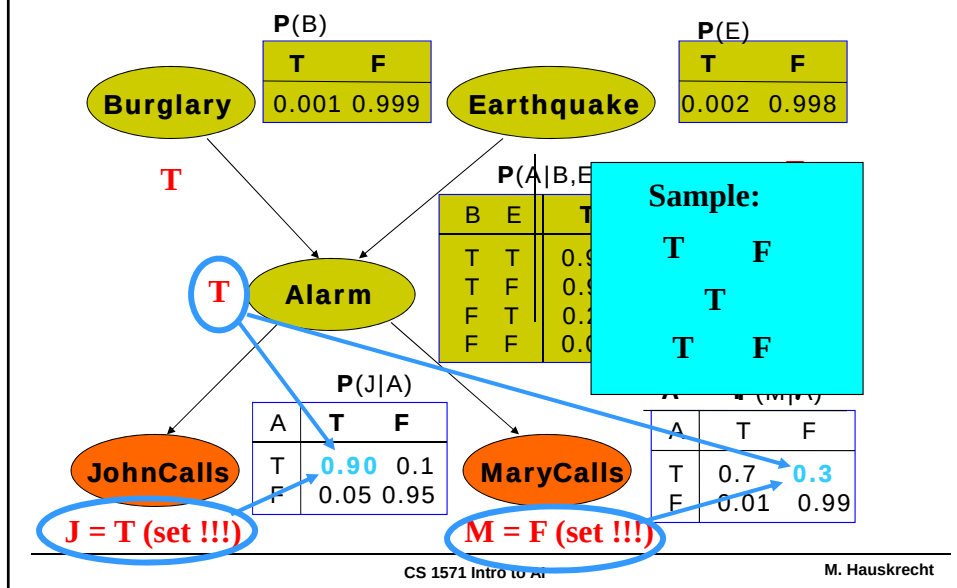
CS 1571 Intro to AI

M. Hauskrecht

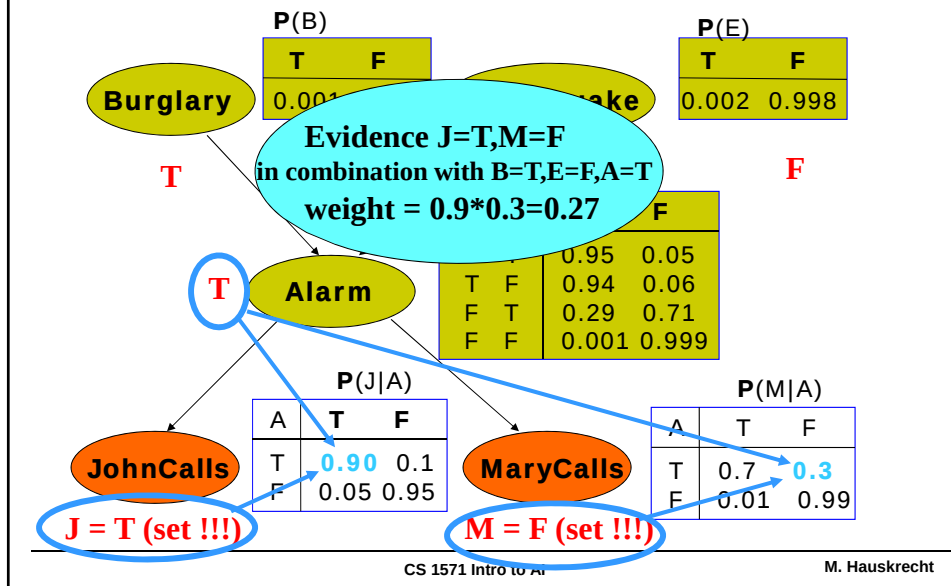
BBN likelihood weighting example



BBN likelihood weighting example

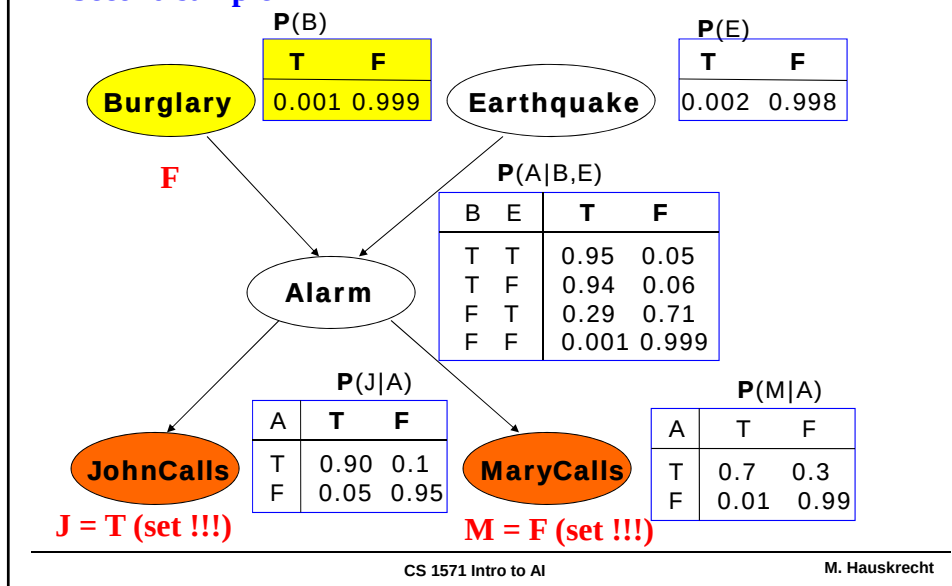


BBN likelihood weighting example



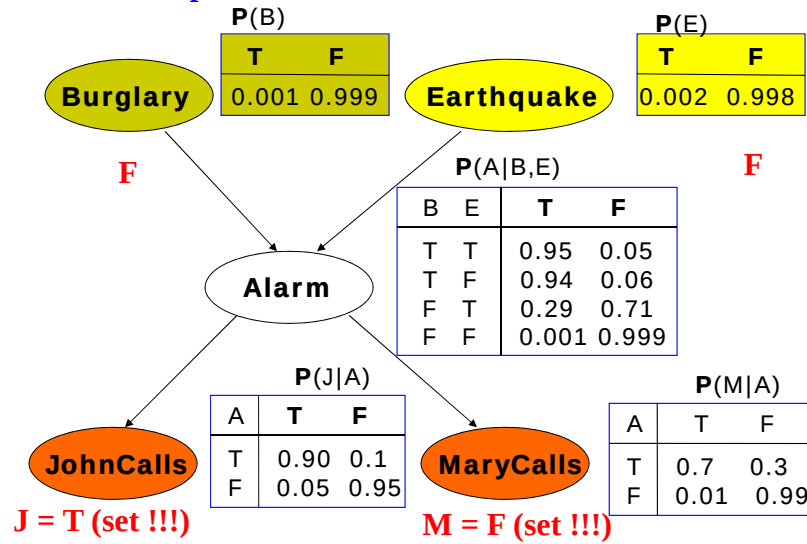
BBN likelihood weighting example

Second sample



BBN likelihood weighting example

Second sample

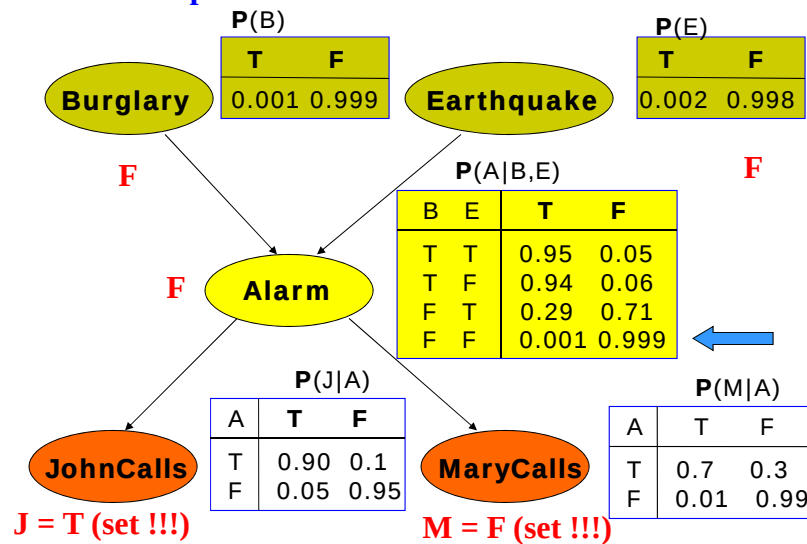


CS 1571 Intro to AI

M. Hauskrecht

BBN likelihood weighting example

Second sample

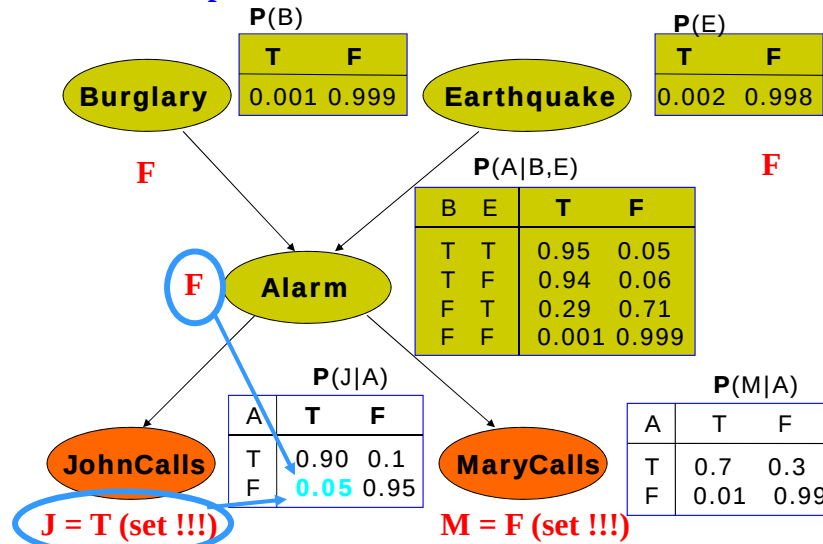


CS 1571 Intro to AI

M. Hauskrecht

BBN likelihood weighting example

Second sample

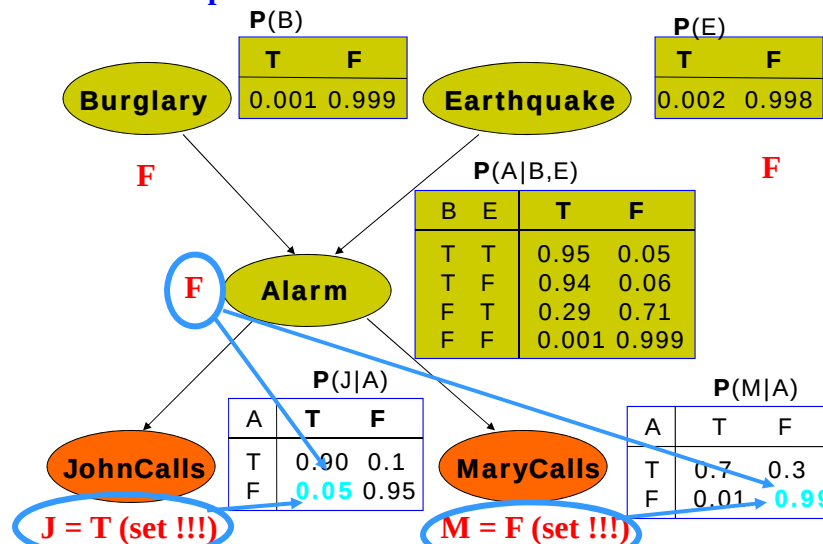


CS 1571 Intro to AI

M. Hauskrecht

BBN likelihood weighting example

Second sample

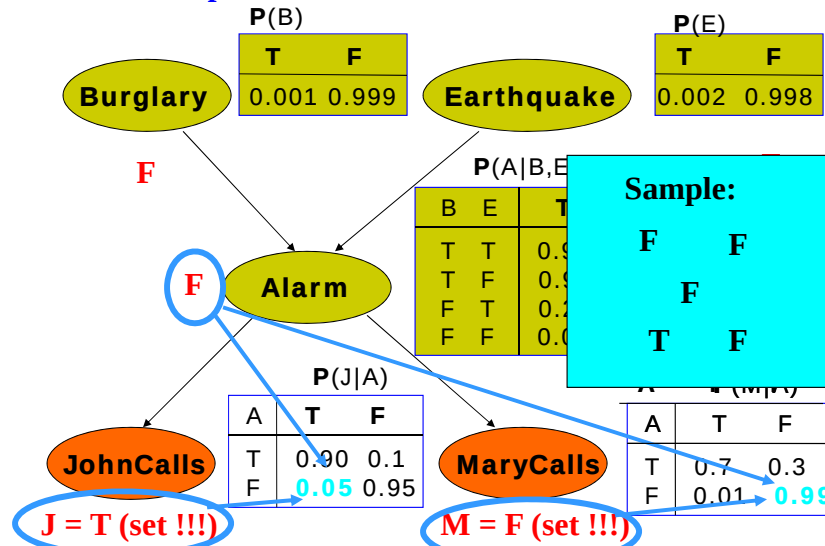


CS 1571 Intro to AI

M. Hauskrecht

BBN likelihood weighting example

Second sample

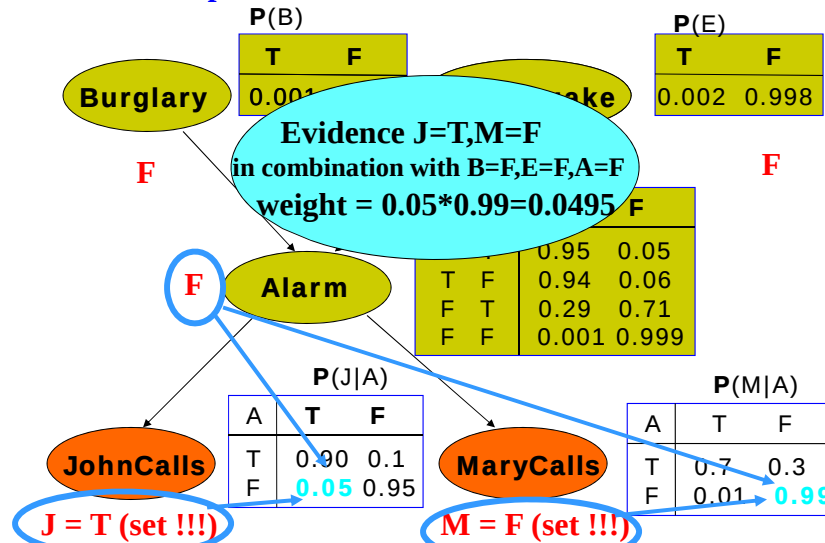


CS 1571 Intro to AI

M. Hauskrecht

BBN likelihood weighting example

Second sample

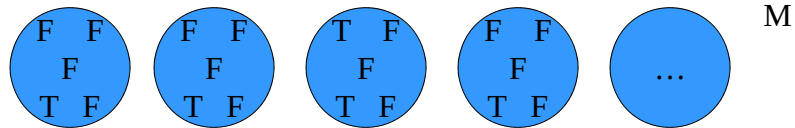


CS 1571 Intro to AI

M. Hauskrecht

Likelihood weighting

- Assume we have generated the following M samples:



How to make the samples consistent? Weight each sample by probability with which it agrees with the conditioning evidence $P(e)$.



$$\tilde{P}(B = T \mid J = T, M = F) = \frac{\sum_{\text{samples with } B=T, M=F \text{ and } J=T} w_{B=T \mid J=T, M=F}}{\sum_{\text{samples with any value of } B \text{ and } J=T, M=F} w_{B=x \mid J=T, M=F}}$$