

CS 1675 Introduction to Machine Learning
Lecture 16

**Bayesian belief networks:
Inference**

Milos Hauskrecht

milos@pitt.edu

5329 Sennott Square

Inference in Bayesian networks

- **BBN models compactly the full joint distribution by taking advantage of existing independences among variables**
- Simplifies the representation and learning of a model
- But we are interested in solving various **inference tasks**:

- **Diagnostic task. (from effect to cause)**

$$\mathbf{P}(\textit{Burglary} \mid \textit{JohnCalls} = T)$$

- **Prediction task. (from cause to effect)**

$$\mathbf{P}(\textit{JohnCalls} \mid \textit{Burglary} = T)$$

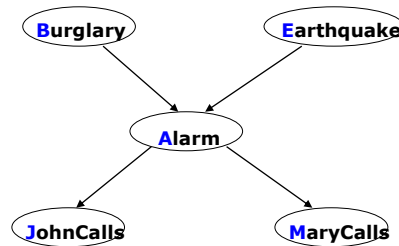
- **Other probabilistic queries** (queries on joint distributions).

$$\mathbf{P}(\textit{Alarm})$$

- **Main question:** Can we take advantage of independences to construct special algorithms and speeding up the inference?
-

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: $P(J = T)$

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

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

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
(constants can be taken out of the sum)

$$\sum_x a f(x) = a \sum_x f(x)$$

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

Computational cost:

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

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
(constants can be taken out of the sum)

$$\sum_x a f(x) = a \sum_x f(x)$$

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

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

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

Variable elimination

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

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

Variable elimination

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

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

Variable elimination

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

$$\begin{aligned}
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} \sum_{m \in T, F} P(J = T \mid A = a) P(M = m \mid A = a) P(A = a \mid B = b, E = e) P(B = b) P(E = e) \\
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} P(J = T \mid A = a) P(A = a \mid B = b, E = e) P(B = b) P(E = e) \left[\sum_{m \in T, F} P(M = m \mid A = a) \right] \\
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} P(J = T \mid A = a) P(A = a \mid B = b, E = e) P(B = b) P(E = e) \boxed{1}
 \end{aligned}$$


Variable elimination

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

$$\begin{aligned}
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} \sum_{m \in T, F} P(J = T \mid A = a) P(M = m \mid A = a) P(A = a \mid B = b, E = e) P(B = b) P(E = e) \\
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} P(J = T \mid A = a) P(A = a \mid B = b, E = e) P(B = b) P(E = e) \left[\sum_{m \in T, F} P(M = m \mid A = a) \right] \\
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} P(J = T \mid A = a) P(A = a \mid B = b, E = e) P(B = b) P(E = e) \quad 1 \\
 &= \sum_{a \in T, F} \sum_{b \in T, F} P(J = T \mid A = a) P(B = b) \left[\sum_{e \in T, F} P(A = a \mid B = b, E = e) P(E = e) \right]
 \end{aligned}$$

Variable elimination

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

$$\begin{aligned}
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} \sum_{m \in T, F} P(J = T \mid A = a) P(M = m \mid A = a) P(A = a \mid B = b, E = e) P(B = b) P(E = e) \\
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} P(J = T \mid A = a) P(A = a \mid B = b, E = e) P(B = b) P(E = e) \left[\sum_{m \in T, F} P(M = m \mid A = a) \right] \\
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} P(J = T \mid A = a) P(A = a \mid B = b, E = e) P(B = b) P(E = e) \quad 1 \\
 &= \sum_{a \in T, F} \sum_{b \in T, F} P(J = T \mid A = a) P(B = b) \left[\sum_{e \in T, F} P(A = a \mid B = b, E = e) P(E = e) \right]
 \end{aligned}$$

$$\tau_1(A = a, B = b)$$

$$\tau_1(A = a, B = b) =$$

	$A = T$	$A = F$	
$B = T$			$\sum_{e \in T, F} P(A = F \mid B = T, E = e) P(E = e)$
$B = F$			$\sum_{e \in T, F} P(A = F \mid B = F, E = e) P(E = e)$

Variable elimination

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

$$\begin{aligned}
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} \sum_{m \in T, F} P(J = T \mid A = a) P(M = m \mid A = a) P(A = a \mid B = b, E = e) P(B = b) P(E = e) \\
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} P(J = T \mid A = a) P(A = a \mid B = b, E = e) P(B = b) P(E = e) \left[\sum_{m \in T, F} P(M = m \mid A = a) \right] \\
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} P(J = T \mid A = a) P(A = a \mid B = b, E = e) P(B = b) P(E = e) \quad 1 \\
 &= \sum_{a \in T, F} \sum_{b \in T, F} P(J = T \mid A = a) P(B = b) \left[\sum_{e \in T, F} P(A = a \mid B = b, E = e) P(E = e) \right] \\
 &= \sum_{a \in T, F} \sum_{b \in T, F} P(J = T \mid A = a) P(B = b) \tau_1(A = a, B = b)
 \end{aligned}$$

Variable elimination

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

$$\begin{aligned}
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} \sum_{m \in T, F} P(J = T \mid A = a) P(M = m \mid A = a) P(A = a \mid B = b, E = e) P(B = b) P(E = e) \\
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} P(J = T \mid A = a) P(A = a \mid B = b, E = e) P(B = b) P(E = e) \left[\sum_{m \in T, F} P(M = m \mid A = a) \right] \\
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} P(J = T \mid A = a) P(A = a \mid B = b, E = e) P(B = b) P(E = e) \quad 1 \\
 &= \sum_{a \in T, F} \sum_{b \in T, F} P(J = T \mid A = a) P(B = b) \left[\sum_{e \in T, F} P(A = a \mid B = b, E = e) P(E = e) \right] \\
 &= \sum_{a \in T, F} \sum_{b \in T, F} P(J = T \mid A = a) P(B = b) \tau_1(A = a, B = b) \\
 &= \sum_{a \in T, F} P(J = T \mid A = a) \left[\sum_{b \in T, F} P(B = b) \tau_1(A = a, B = b) \right]
 \end{aligned}$$

Variable elimination

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

$$\begin{aligned}
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} \sum_{m \in T, F} P(J = T \mid A = a) P(M = m \mid A = a) P(A = a \mid B = b, E = e) P(B = b) P(E = e) \\
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} P(J = T \mid A = a) P(A = a \mid B = b, E = e) P(B = b) P(E = e) \left[\sum_{m \in T, F} P(M = m \mid A = a) \right] \\
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} P(J = T \mid A = a) P(A = a \mid B = b, E = e) P(B = b) P(E = e) \cdot 1 \\
 &= \sum_{a \in T, F} \sum_{b \in T, F} P(J = T \mid A = a) P(B = b) \left[\sum_{e \in T, F} P(A = a \mid B = b, E = e) P(E = e) \right] \\
 &= \sum_{a \in T, F} \sum_{b \in T, F} P(J = T \mid A = a) P(B = b) \tau_1(A = a, B = b) \\
 &= \sum_{a \in T, F} P(J = T \mid A = a) \left[\sum_{b \in T, F} P(B = b) \tau_1(A = a, B = b) \right] \\
 &\quad \quad \quad \tau_2(A = a)
 \end{aligned}$$

Variable elimination

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

$$\begin{aligned}
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} \sum_{m \in T, F} P(J = T \mid A = a) P(M = m \mid A = a) P(A = a \mid B = b, E = e) P(B = b) P(E = e) \\
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} P(J = T \mid A = a) P(A = a \mid B = b, E = e) P(B = b) P(E = e) \left[\sum_{m \in T, F} P(M = m \mid A = a) \right] \\
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} P(J = T \mid A = a) P(A = a \mid B = b, E = e) P(B = b) P(E = e) \cdot 1 \\
 &= \sum_{a \in T, F} \sum_{b \in T, F} P(J = T \mid A = a) P(B = b) \left[\sum_{e \in T, F} P(A = a \mid B = b, E = e) P(E = e) \right] \\
 &= \sum_{a \in T, F} \sum_{b \in T, F} P(J = T \mid A = a) P(B = b) \tau_1(A = a, B = b) \\
 &= \sum_{a \in T, F} P(J = T \mid A = a) \left[\sum_{b \in T, F} P(B = b) \tau_1(A = a, B = b) \right] \\
 &= \sum_{a \in T, F} P(J = T \mid A = a) \tau_2(A = a)
 \end{aligned}$$

Variable elimination

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

$$\begin{aligned}
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} \sum_{m \in T, F} P(J = T | A = a) P(M = m | A = a) P(A = a | B = b, E = e) P(B = b) P(E = e) \\
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} P(J = T | A = a) P(A = a | B = b, E = e) P(B = b) P(E = e) \left[\sum_{m \in T, F} P(M = m | A = a) \right] \\
 &= \sum_{b \in T, F} \sum_{e \in T, F} \sum_{a \in T, F} P(J = T | A = a) P(A = a | B = b, E = e) P(B = b) P(E = e) \quad 1 \\
 &= \sum_{a \in T, F} \sum_{b \in T, F} P(J = T | A = a) P(B = b) \left[\sum_{e \in T, F} P(A = a | B = b, E = e) P(E = e) \right] \\
 &= \sum_{a \in T, F} \sum_{b \in T, F} P(J = T | A = a) P(B = b) \tau_1(A = a, B = b) \\
 &= \sum_{a \in T, F} P(J = T | A = a) \left[\sum_{b \in T, F} P(B = b) \tau_1(A = a, B = b) \right] \\
 &= \sum_{a \in T, F} P(J = T | A = a) \tau_2(A = a) = P(J = T)
 \end{aligned}$$

Inference in Bayesian network

- **Exact inference algorithms:**
 - **Variable elimination**
 - Recursive decomposition (Cooper, Darwiche)
 - Symbolic inference (D'Ambrosio)
 - Belief propagation algorithm (Pearl)
 - Clustering and joint tree approach (Lauritzen, Spiegelhalter)
 - Arc reversal (Olmsted, Schachter)
- **Approximate inference algorithms:**
 - **Monte Carlo methods:**
 - Forward sampling, Likelihood sampling
 - Variational methods

Monte Carlo approaches

- **MC approximation:**

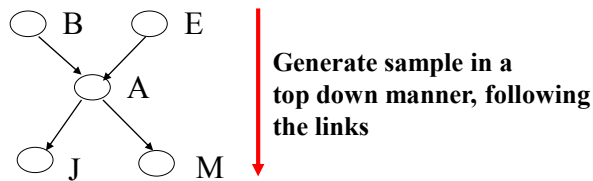
- The probability is approximated using **sample frequencies**

- **Example:**

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

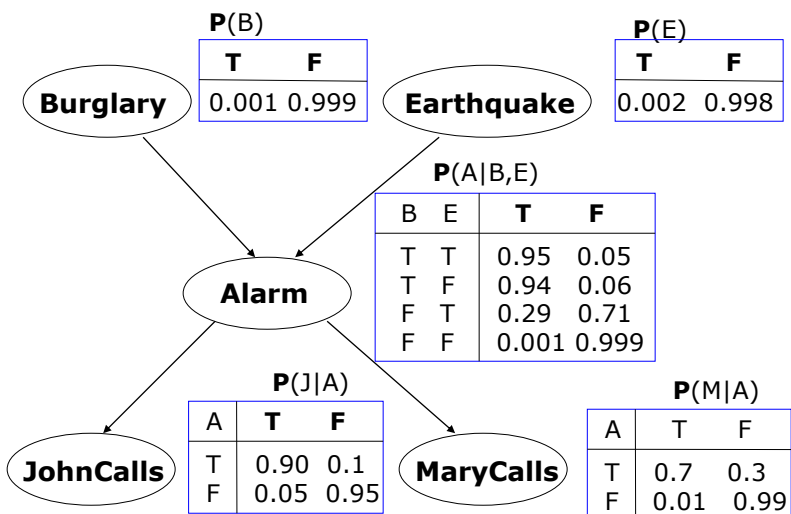
$\leftarrow$ #samples with $B=T, J=T$
 $\leftarrow$ total # samples

- **Sample generation:** BBN sampling of the joint is easy

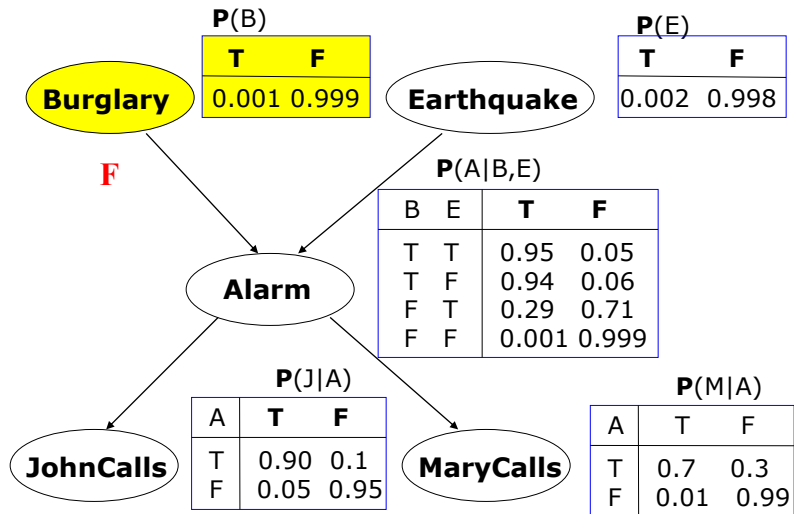


- One sample gives one assignment of values to all variables

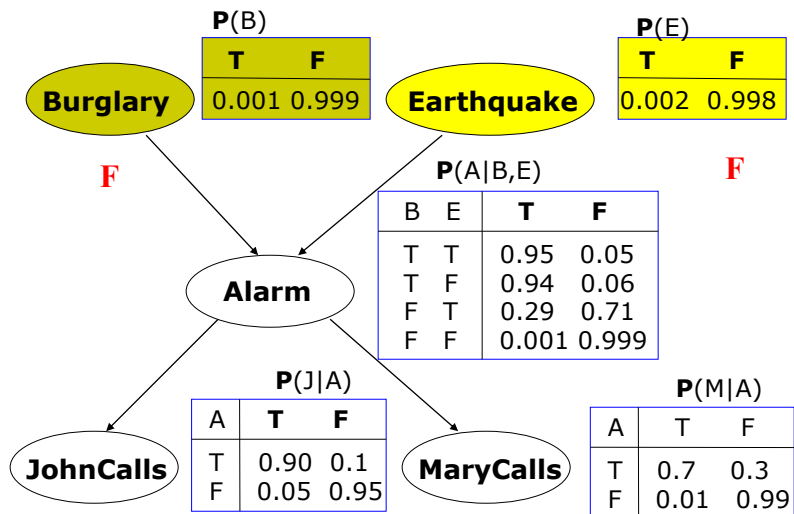
BBN sampling example



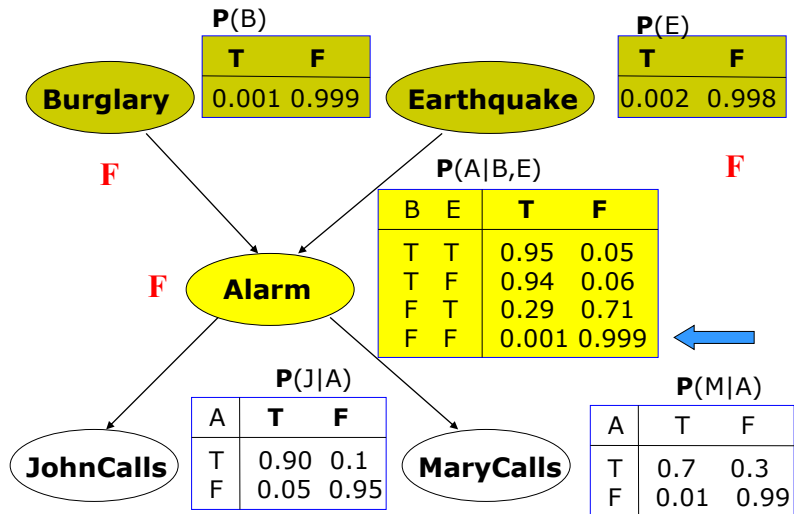
BBN sampling example



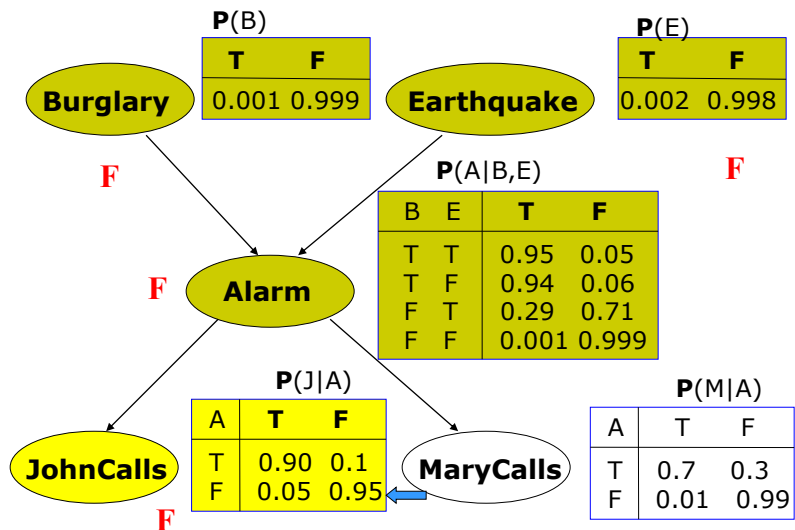
BBN sampling example



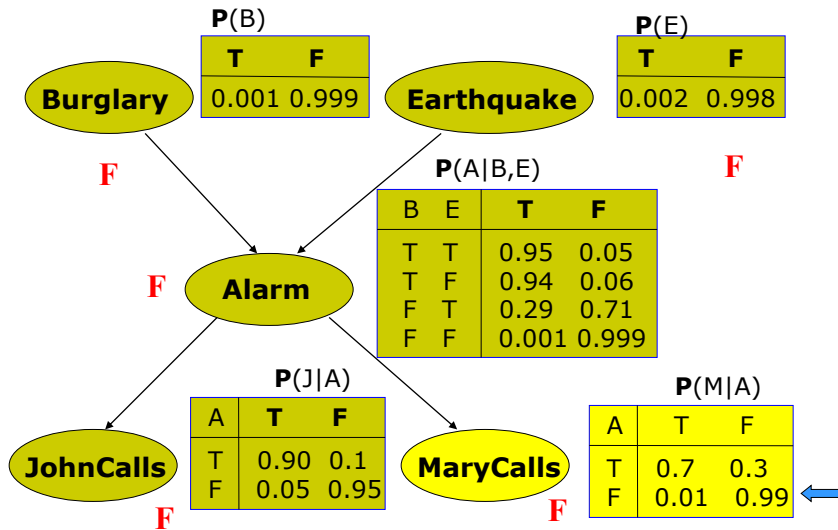
BBN sampling example



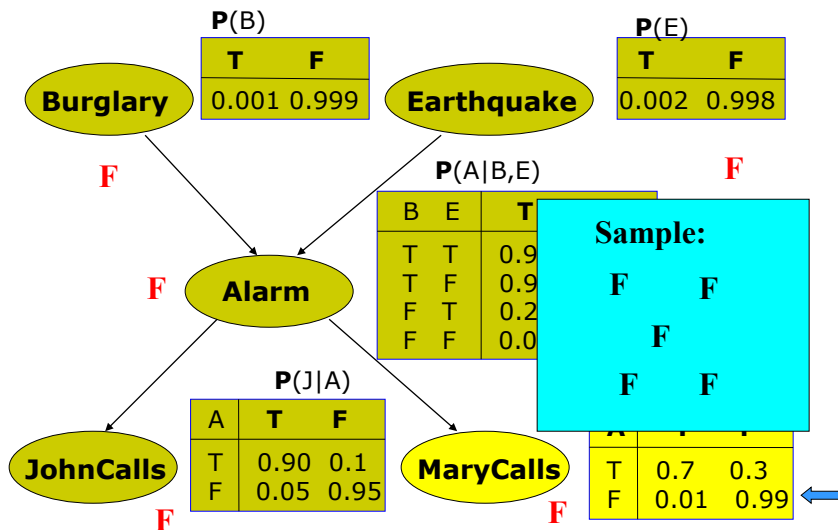
BBN sampling example



BBN sampling example



BBN sampling example



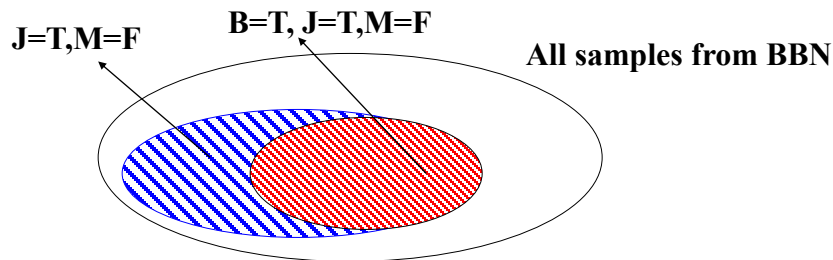
Monte Carlo approaches

- **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 $J = T, M = F$

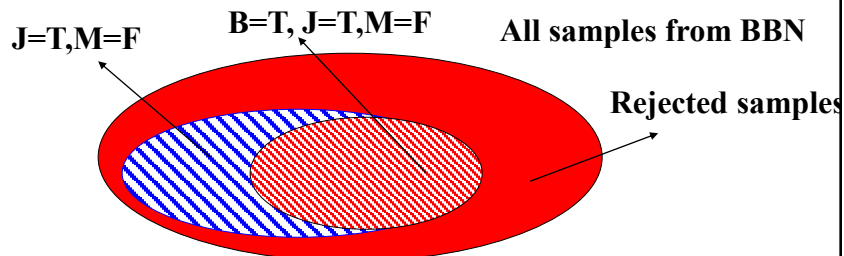


Monte Carlo approaches

- **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



Likelihood weighting

Idea: generate only samples consistent with an evidence (or conditioning event)

- Benefit: Avoids inefficiencies of rejection sampling

Problem:

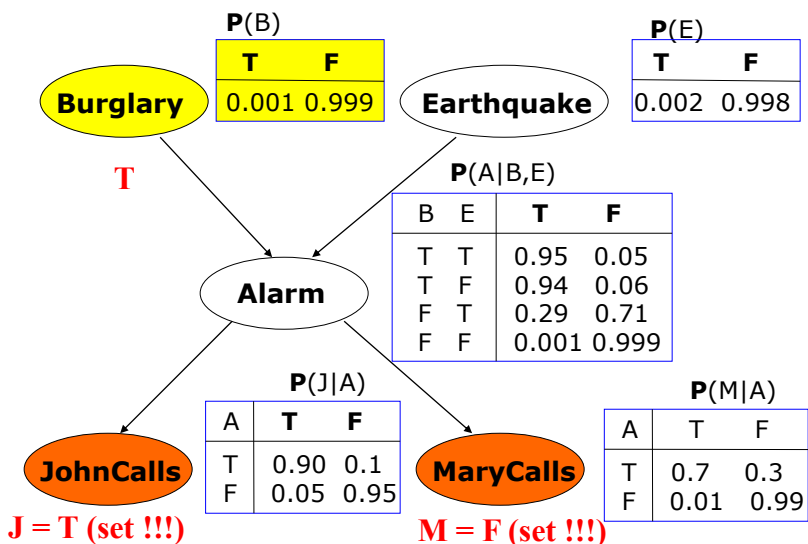
- the distribution generated by enforcing the conditioning variables to set values is biased
- simple counts are not sufficient to estimate the probabilities

Solution:

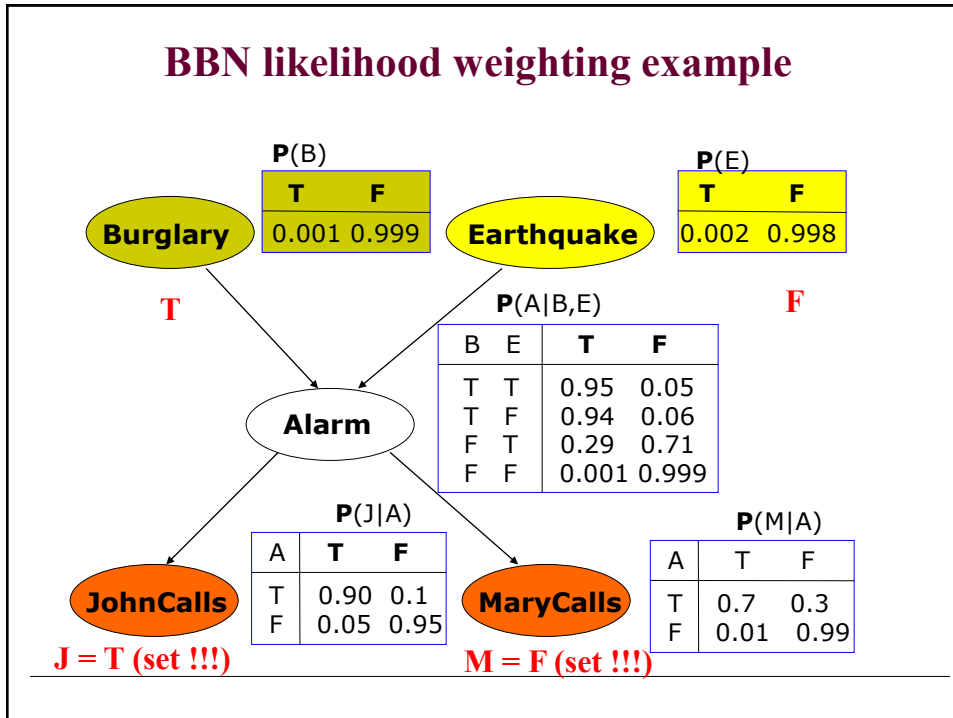
- With every sample keep a weight with which it should count towards the estimate

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

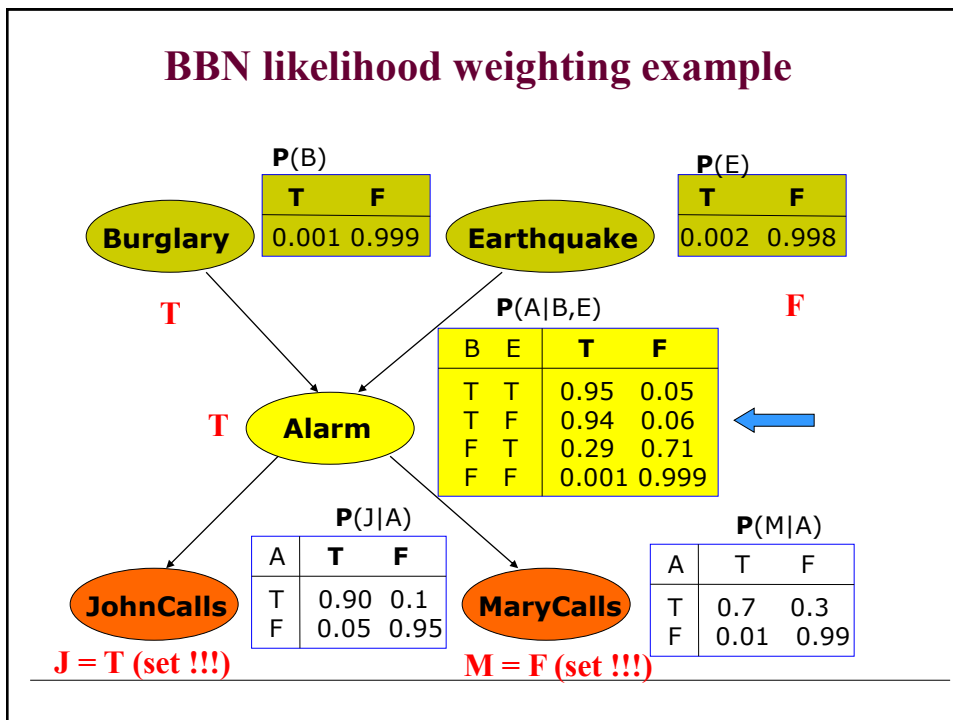
BBN likelihood weighting example



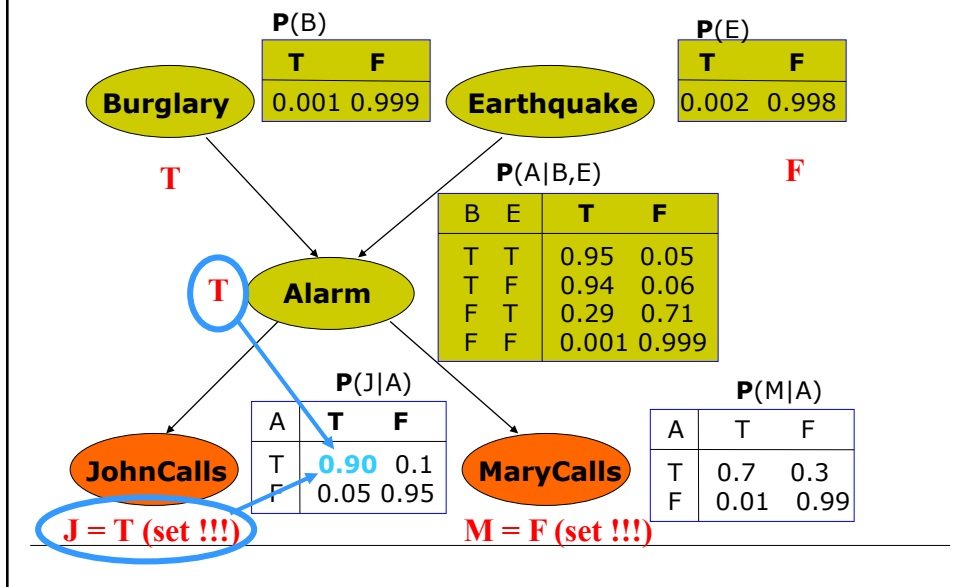
BBN likelihood weighting example



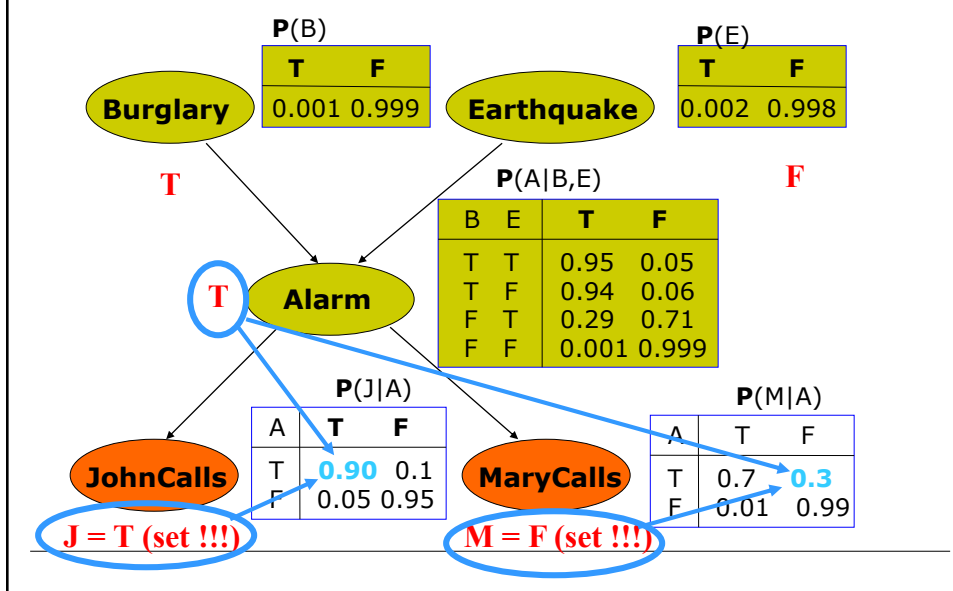
BBN likelihood weighting example



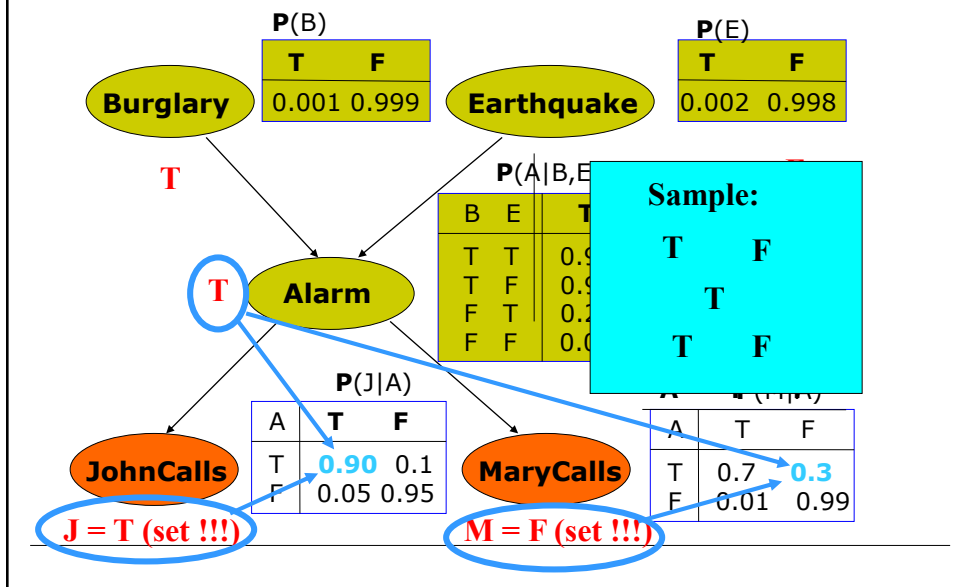
BBN likelihood weighting example



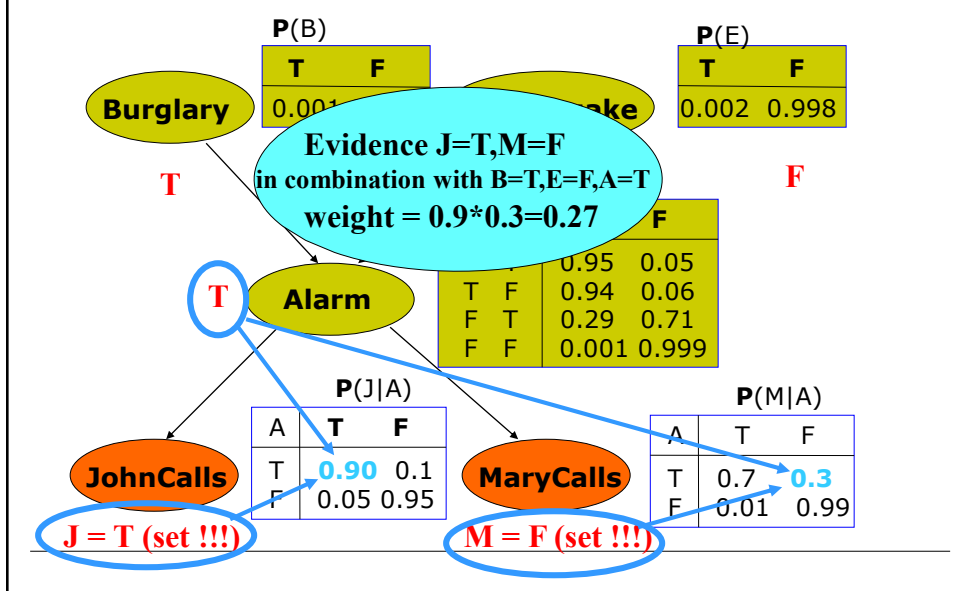
BBN likelihood weighting example



BBN likelihood weighting example

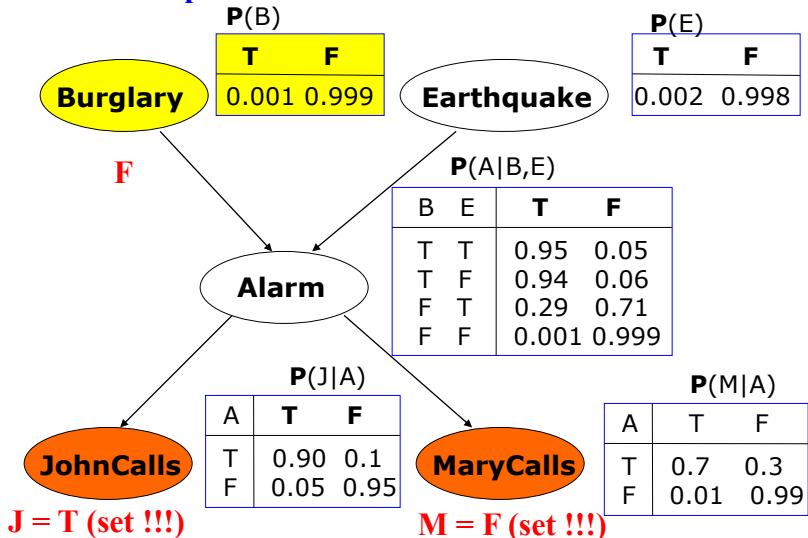


BBN likelihood weighting example



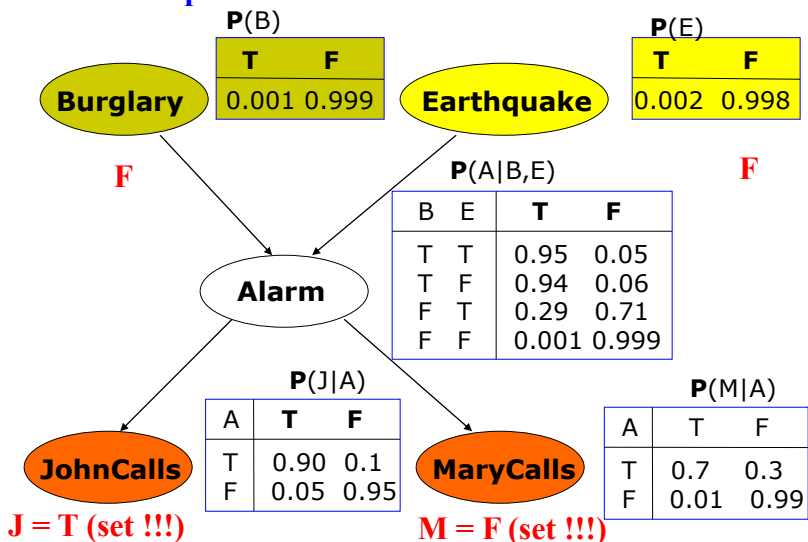
BBN likelihood weighting example

Second sample



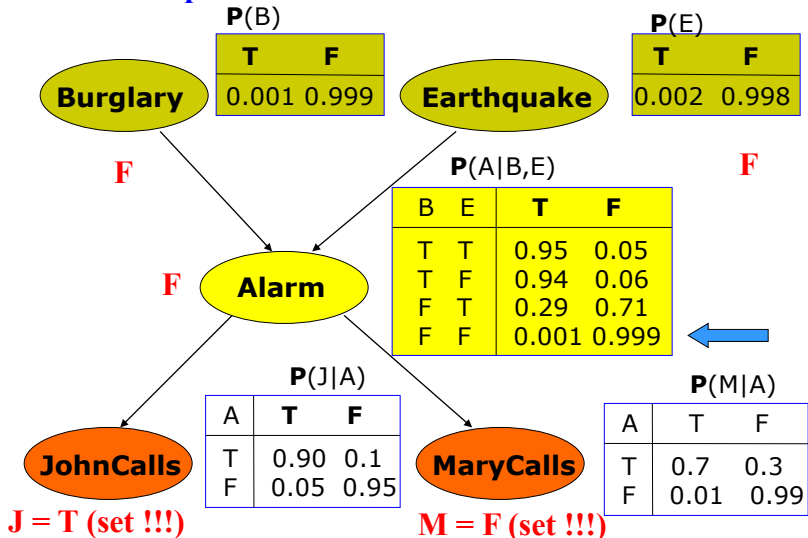
BBN likelihood weighting example

Second sample



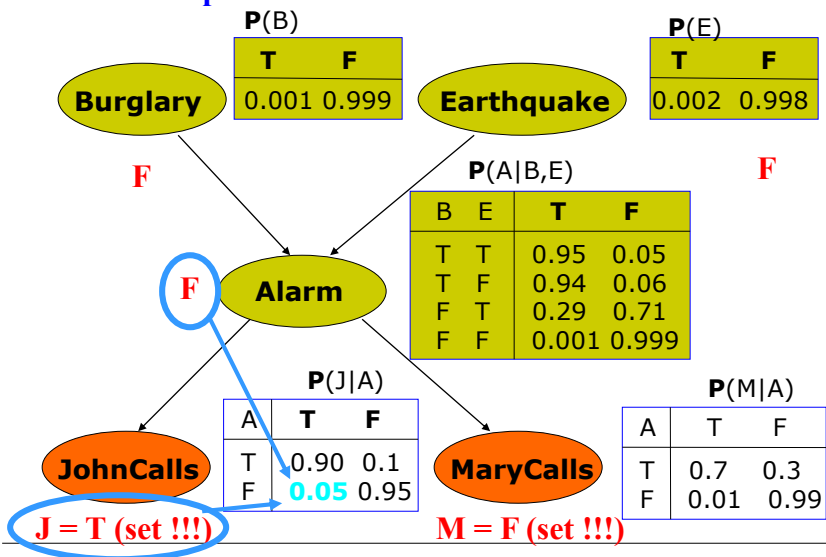
BBN likelihood weighting example

Second sample



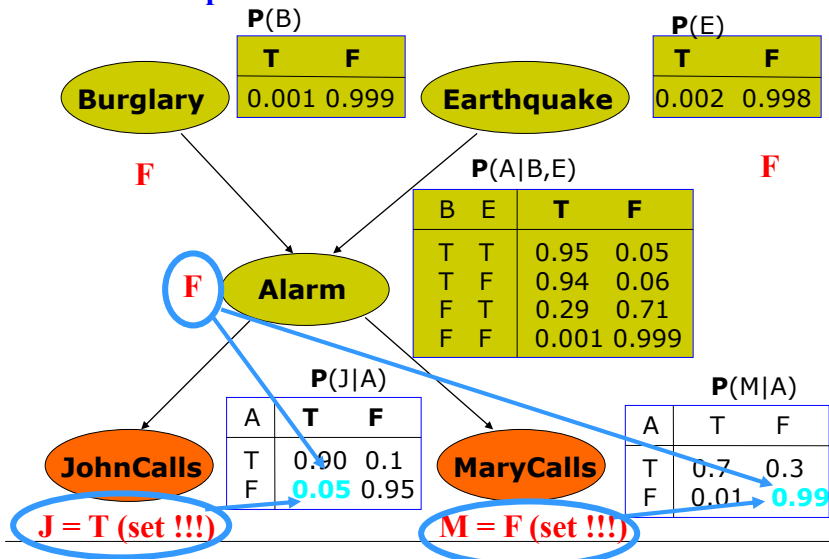
BBN likelihood weighting example

Second sample



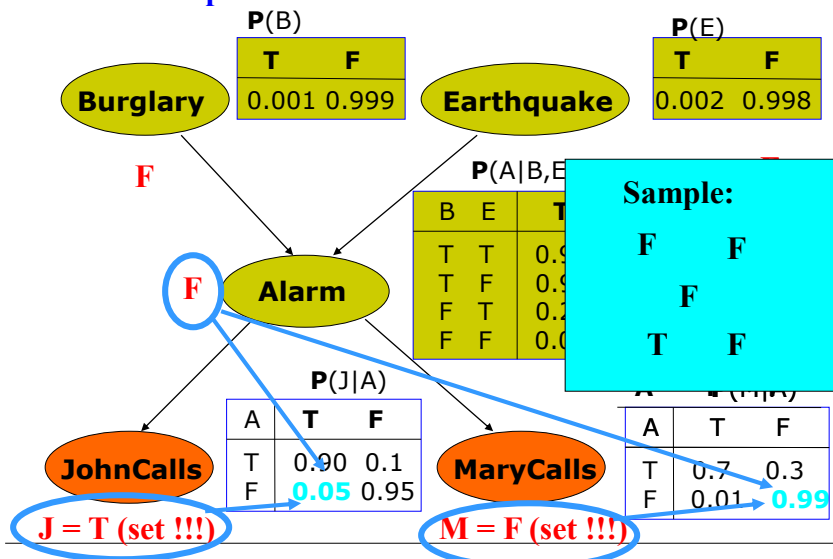
BBN likelihood weighting example

Second sample



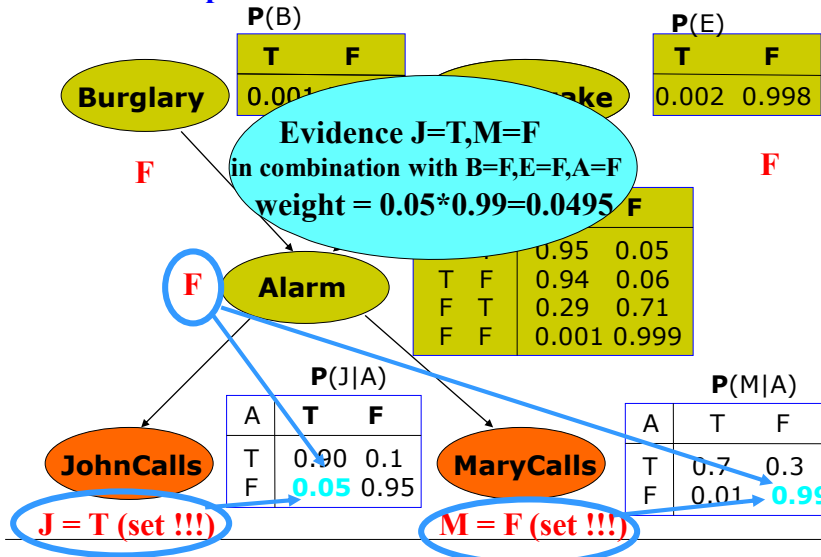
BBN likelihood weighting example

Second sample



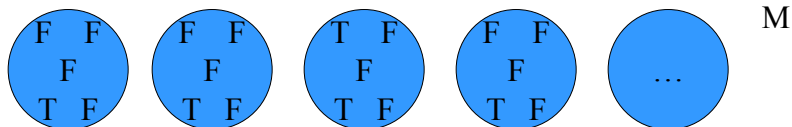
BBN likelihood weighting example

Second sample



Likelihood weighting

- Assume we have generated the following M samples:



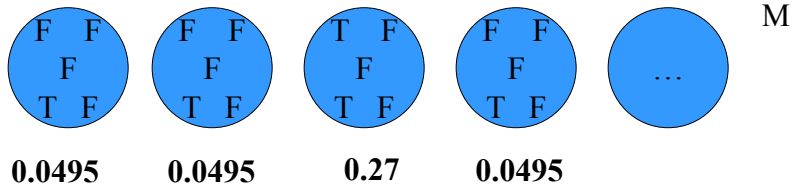
How to make the samples consistent?

Weight each sample by probability with which it agrees with the conditioning evidence $P(e)$.



Likelihood weighting

- Assume we have generated the following M samples:



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