CS 2740 Knowledge representation
Lecture 20

Bayesian belief networks. Inference.

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

Project proposals

Due: Tuesday, 27, 2007

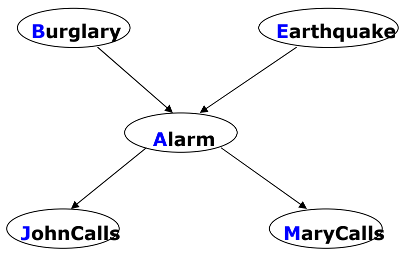
- **1-2 pages long**

Proposal

- **Written proposal:**
  1. Outline the problem you would like to tackle. Why is the problem important?
  2. Methods you plan to try and implement for the problem. References to previous work.
  3. How do you plan to test, run your solution.
  4. Schedule of work (approximate timeline of work)
- **A 3-slide PPT presentation summarizing points 1-4**

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

$$P(J = T) =$$

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

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

Computational cost:

Number of additions: 15

Number of products: 16*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)

$$P(J=T)=$$

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

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

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

Computational cost:

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

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

Variable elimination

- **Idea:** interleave sum and products one variable at the time during the inference
  - Typically relies on a special structure (called **joint tree**) that groups together multiple variables
  - E.g. Query $P(J=T)$ requires to eliminate A,B,E,M and this can be done in different order

$$P(J=T)=$$

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

Variable elimination

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

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

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

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

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

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

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

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

Inference in Bayesian network

- **Exact inference algorithms:**
  - **Variable elimination**
  - Recursive decomposition (Cooper, Darwiche)
  - Belief propagation algorithm (Pearl)
  - Arc reversal (Olmsted, Schachter)

- **Approximate inference algorithms:**
  - **Monte Carlo methods:**
    - Forward sampling, Likelihood sampling
  - Variational methods

Monte Carlo approaches

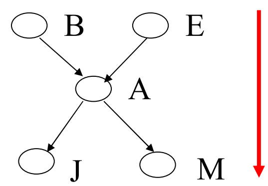
- **MC approximation**:
  - The probability is approximated using sample frequencies
  - **Example:**

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

$N_{B=T,J=T}$ ← $\# examples \;\; with \;\; B = T, J = T$

N ← $total \;\; \# \; examples$

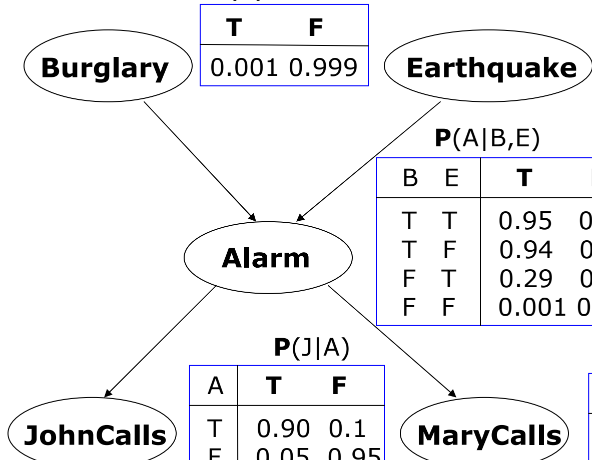
- **BBN sampling:**



Generate sample in the top down manner, following the links

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

BBN sampling example



P(B)

T	**F**
0.001	0.999

P(E)

T	**F**
0.002	0.998

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

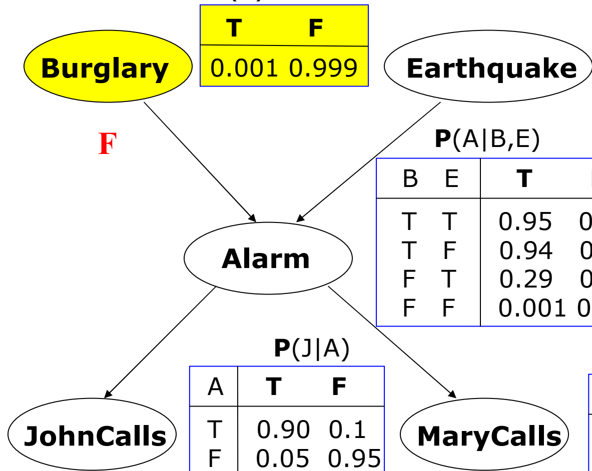
P(J|A)

A	**T**	**F**
T	0.90	0.1
F	0.05	0.95

P(M|A)

A	T	F
T	0.7	0.3
F	0.01	0.99

BBN sampling example



P(B)

T	F
0.001	0.999

P(E)

T	F
0.002	0.998

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

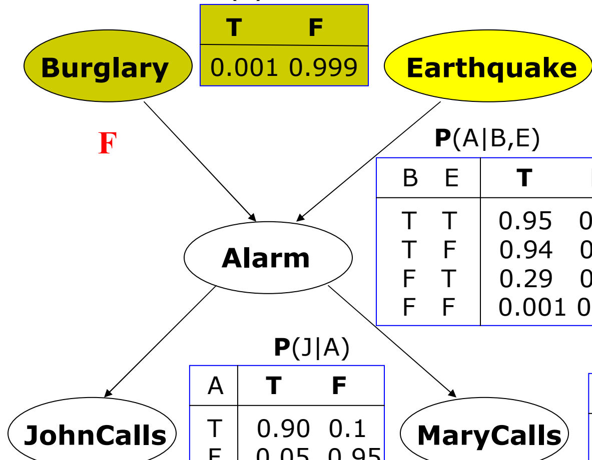
P(J|A)

A	T	F
T	0.90	0.1
F	0.05	0.95

P(M|A)

A	T	F
T	0.7	0.3
F	0.01	0.99

BBN sampling example



P(B)

T	F
0.001	0.999

P(E)

T	F
0.002	0.998

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

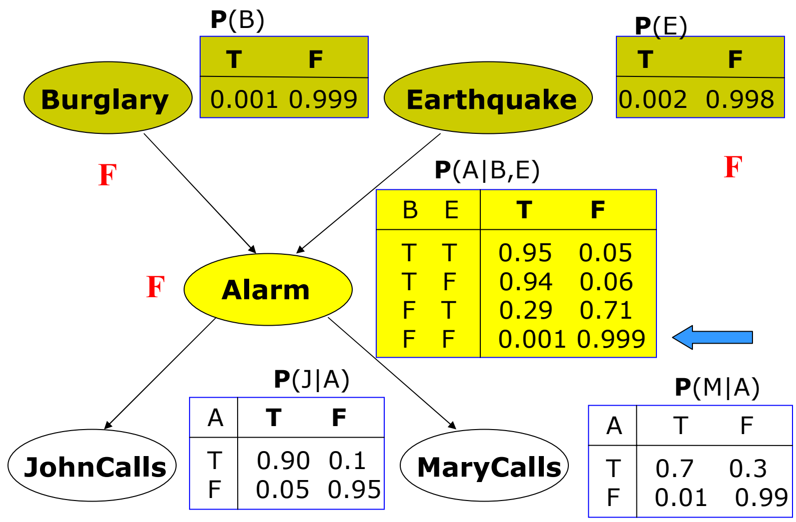
P(J|A)

A	T	F
T	0.90	0.1
F	0.05	0.95

P(M|A)

A	T	F
T	0.7	0.3
F	0.01	0.99

BBN sampling example



P(B)

T	F
0.001	0.999

P(E)

T	F
0.002	0.998

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

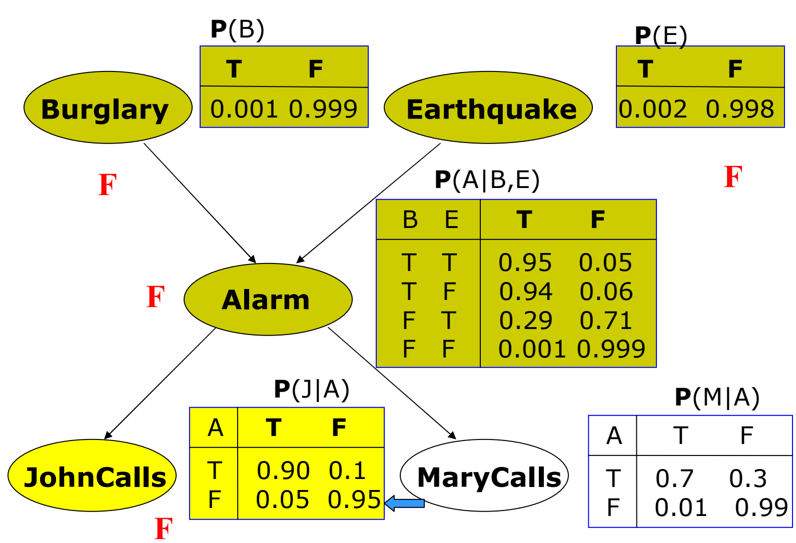
P(J|A)

A	T	F
T	0.90	0.1
F	0.05	0.95

P(M|A)

A	T	F
T	0.7	0.3
F	0.01	0.99

BBN sampling example



P(B)

T	F
0.001	0.999

P(E)

T	F
0.002	0.998

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

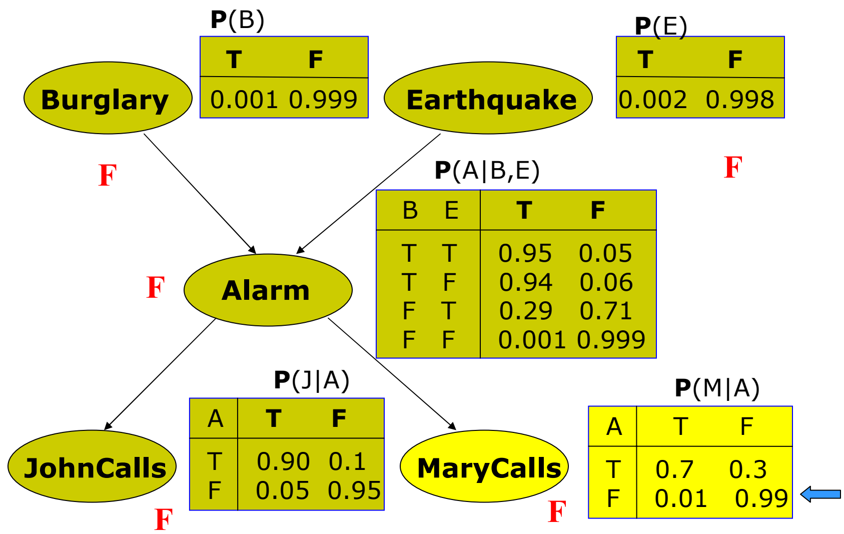
P(J|A)

A	T	F
T	0.90	0.1
F	0.05	0.95

P(M|A)

A	T	F
T	0.7	0.3
F	0.01	0.99

BBN sampling example



P(B)

T	F
0.001	0.999

P(E)

T	F
0.002	0.998

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

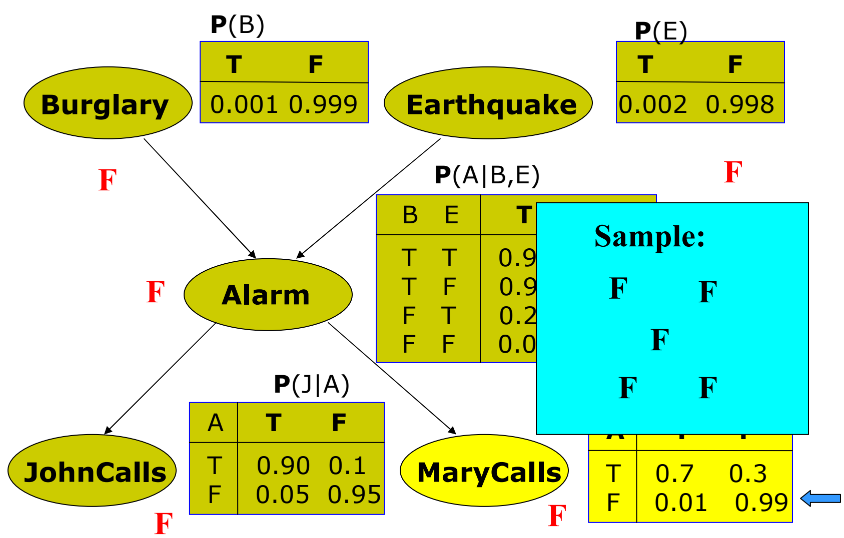
P(J|A)

A	T	F
T	0.90	0.1
F	0.05	0.95

P(M|A)

A	T	F
T	0.7	0.3
F	0.01	0.99

BBN sampling example



P(B)

T	F
0.001	0.999

P(E)

T	F
0.002	0.998

P(A|B,E)

B	E	T
T	T	0.9
T	F	0.9
F	T	0.2
F	F	0.0

P(J|A)

A	T	F
T	0.90	0.1
F	0.05	0.95

T	0.7	0.3
F	0.01	0.99

Monte Carlo approaches

- **MC approximation of conditional probabilities**:
  - The probability is approximated using sample frequencies
  - **Example:**

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

Numerator: $\#\, samples\;\; with\;\; B=T, J=T, M=F$

Denominator: $\#\, samples\;\; with\;\; J=T, M=F$

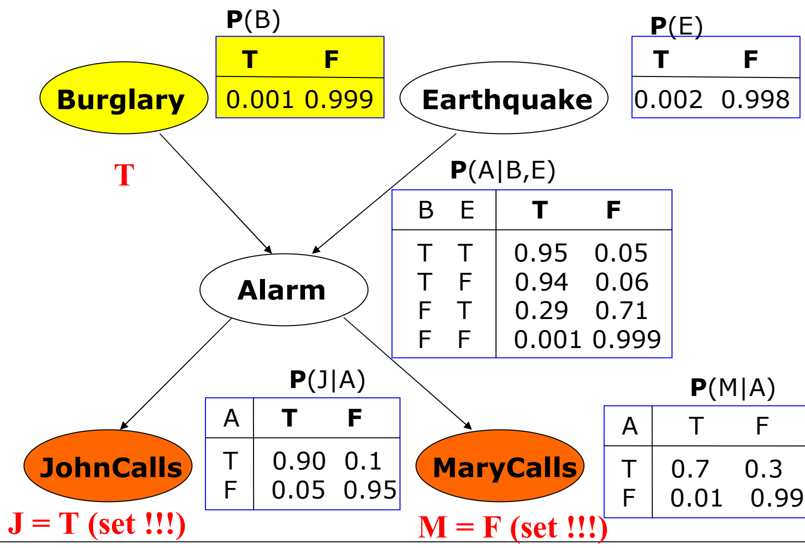
- **Rejection sampling:**
  - Generate samples from the full joint by sampling BBN
  - Use only samples that agree with the condition, the remaining samples are rejected
- **Problem:** many samples can be rejected

Likelihood weighting

- **Avoids inefficiencies of rejection sampling**
  - **Idea:** generate only samples **consistent with the evidence** (or conditioning event)
    - **the value of evidence nodes is not sampled**
- **Problem:** using simple counts is not enough since these may occur with different probabilities
- Likelihood weighting:
  - **With every sample keep a weight with which it should count towards the estimate**

$$\tilde{P}(B=T \mid J=T, M=F) = \frac{\sum_{samples\;\; with\;\; B=T,J=T,M=F} w_{B=T}}{\sum_{samples\;\; with\;\; J=T,M=F} w_{B=x}}$$

BBN likelihood weighting example



P(B)

T	F
0.001	0.999

P(E)

T	F
0.002	0.998

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

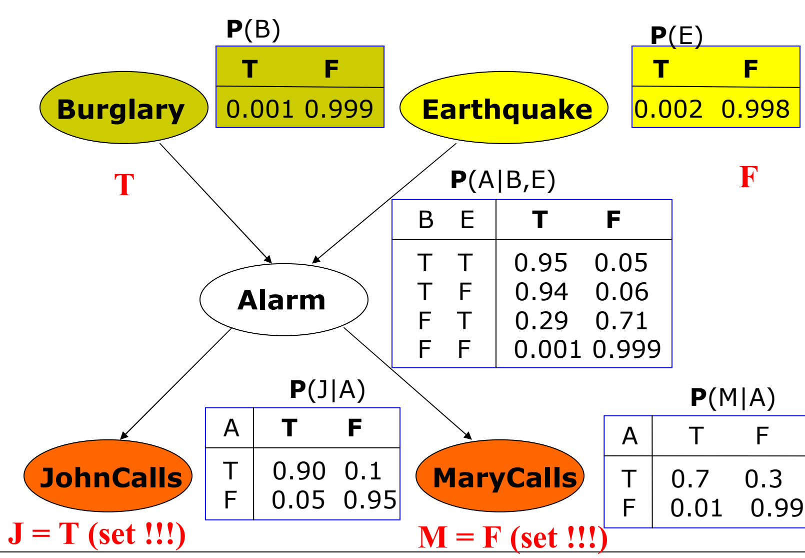
P(J|A)

A	T	F
T	0.90	0.1
F	0.05	0.95

P(M|A)

A	T	F
T	0.7	0.3
F	0.01	0.99

BBN likelihood weighting example



P(B)

T	F
0.001	0.999

P(E)

T	F
0.002	0.998

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

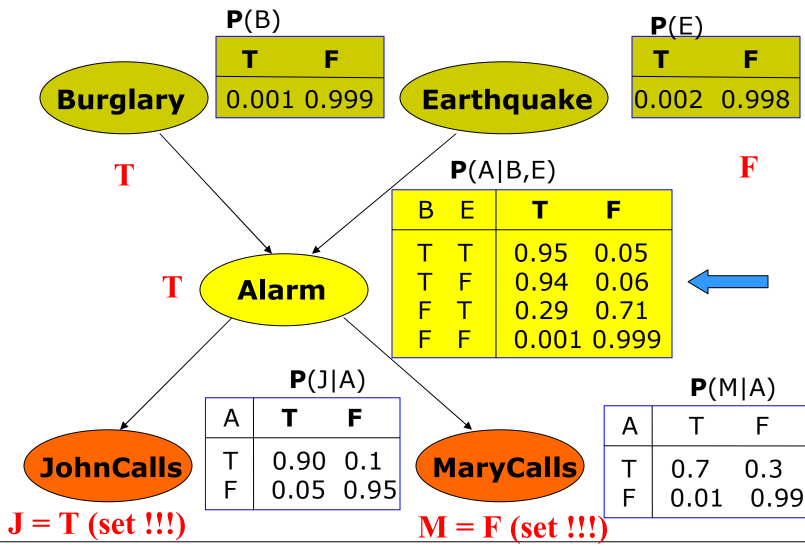
P(J|A)

A	T	F
T	0.90	0.1
F	0.05	0.95

P(M|A)

A	T	F
T	0.7	0.3
F	0.01	0.99

BBN likelihood weighting example



P(B)

T	F
0.001	0.999

P(E)

T	F
0.002	0.998

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

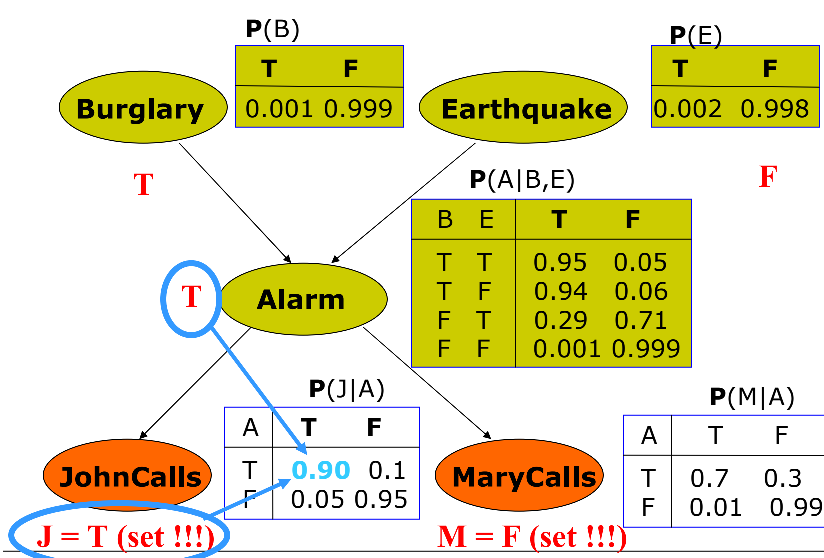
P(J|A)

A	T	F
T	0.90	0.1
F	0.05	0.95

P(M|A)

A	T	F
T	0.7	0.3
F	0.01	0.99

BBN likelihood weighting example



P(B)

T	F
0.001	0.999

P(E)

T	F
0.002	0.998

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

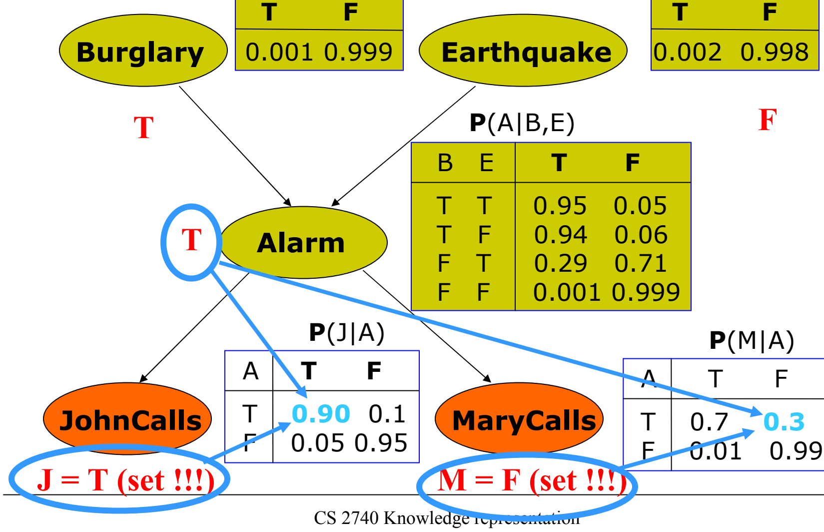
P(J|A)

A	T	F
T	0.90	0.1
F	0.05	0.95

P(M|A)

A	T	F
T	0.7	0.3
F	0.01	0.99

BBN likelihood weighting example



P(B)

T	F
0.001	0.999

P(E)

T	F
0.002	0.998

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

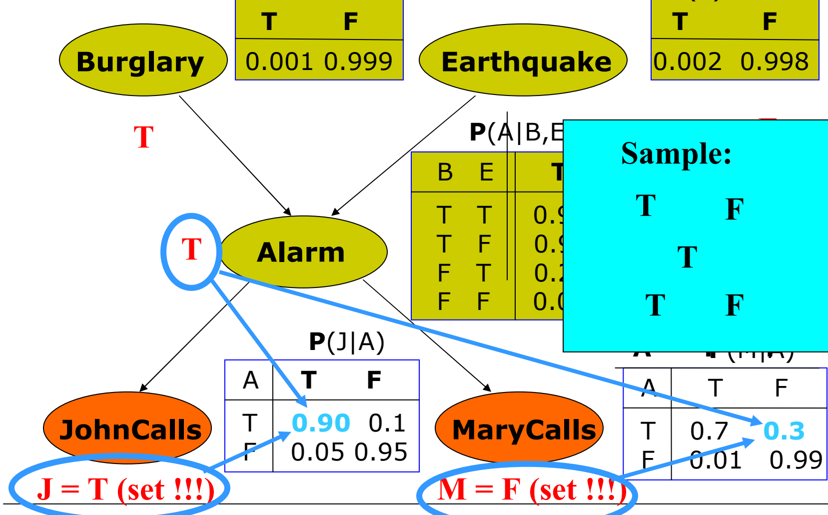
P(J|A)

A	T	F
T	0.90	0.1
F	0.05	0.95

P(M|A)

A	T	F
T	0.7	0.3
F	0.01	0.99

BBN likelihood weighting example



P(B)

T	F
0.001	0.999

P(E)

T	F
0.002	0.998

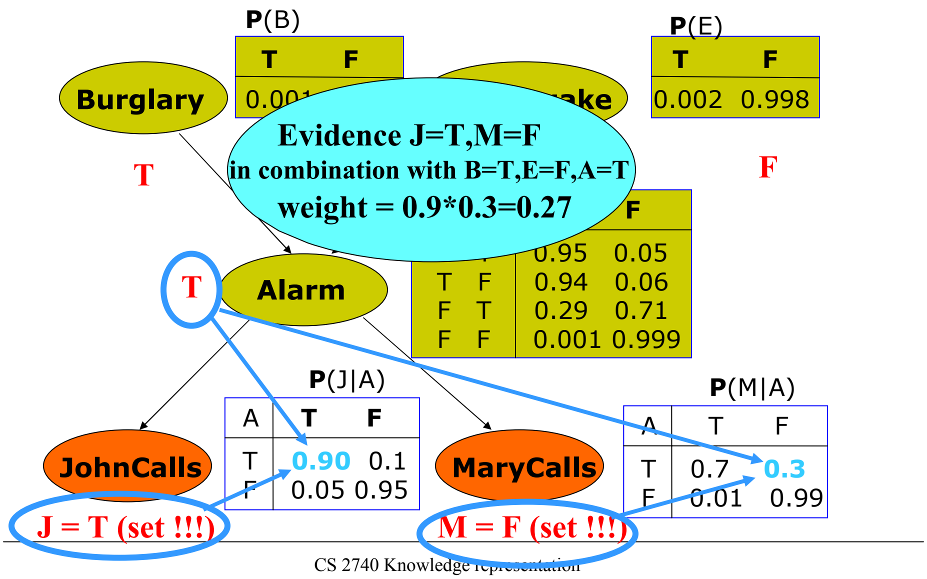
P(J|A)

A	T	F
T	0.90	0.1
F	0.05	0.95

P(M|A)

A	T	F
T	0.7	0.3
F	0.01	0.99

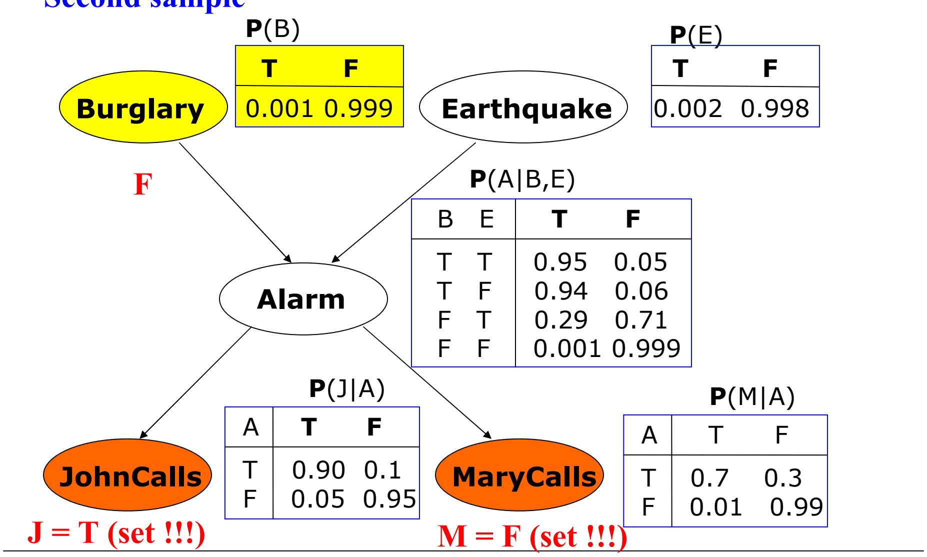
BBN likelihood weighting example



CS 2740 Knowledge representation

BBN likelihood weighting example

Second sample



P(B)

T	F
0.001	0.999

P(E)

T	F
0.002	0.998

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

P(J|A)

A	T	F
T	0.90	0.1
F	0.05	0.95

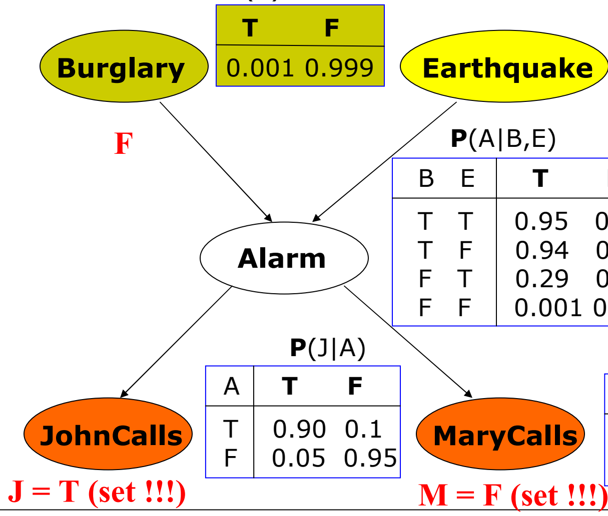
P(M|A)

A	T	F
T	0.7	0.3
F	0.01	0.99

CS 2740 Knowledge representation

BBN likelihood weighting example

Second sample



P(B)

T	F
0.001	0.999

P(E)

T	F
0.002	0.998

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

P(J|A)

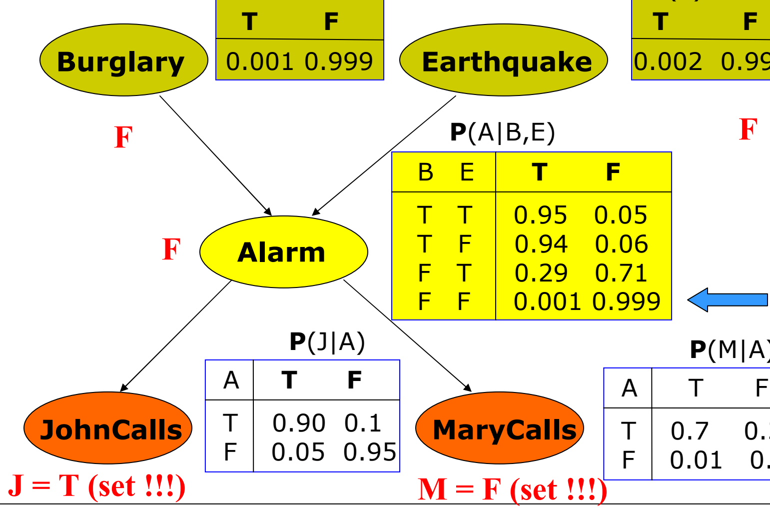
A	T	F
T	0.90	0.1
F	0.05	0.95

P(M|A)

A	T	F
T	0.7	0.3
F	0.01	0.99

BBN likelihood weighting example

Second sample



P(B)

T	F
0.001	0.999

P(E)

T	F
0.002	0.998

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

P(J|A)

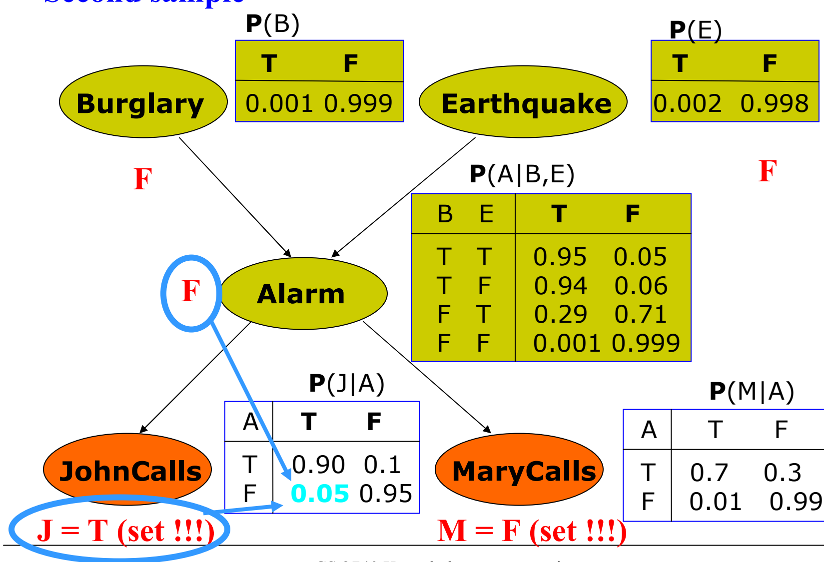
A	T	F
T	0.90	0.1
F	0.05	0.95

P(M|A)

A	T	F
T	0.7	0.3
F	0.01	0.99

BBN likelihood weighting example

Second sample



P(B)

T	F
0.001	0.999

P(E)

T	F
0.002	0.998

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

P(J|A)

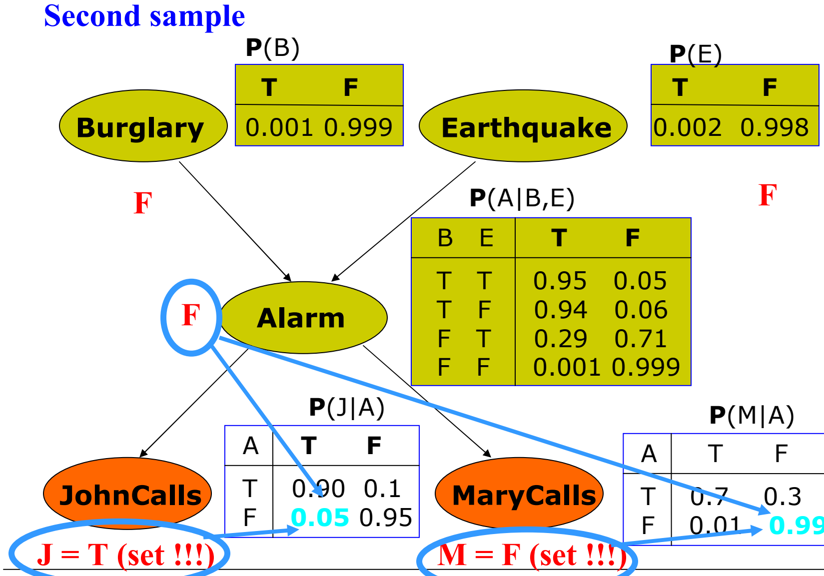
A	T	F
T	0.90	0.1
F	0.05	0.95

P(M|A)

A	T	F
T	0.7	0.3
F	0.01	0.99

BBN likelihood weighting example

Second sample



P(B)

T	F
0.001	0.999

P(E)

T	F
0.002	0.998

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

P(J|A)

A	T	F
T	0.90	0.1
F	0.05	0.95

P(M|A)

A	T	F
T	0.7	0.3
F	0.01	0.99

BBN likelihood weighting example

Second sample

Burglary — F

Earthquake

Alarm — F

JohnCalls — J = T (set !!!)

MaryCalls — M = F (set !!!)

P(B)

T	F
0.001	0.999

P(E)

T	F
0.002	0.998

P(A|B,E

B	E	T
T	T	0.9
T	F	0.9
F	T	0.2
F	F	0.0

Sample:

F F
F
T F

P(J|A)

A	T	F
T	0.90	0.1
F	0.05	0.95

A	T	F
T	0.7	0.3
F	0.01	0.99

BBN likelihood weighting example

Second sample

Burglary — F

Alarm — F

F

JohnCalls — J = T (set !!!)

MaryCalls — M = F (set !!!)

P(B)

T	F
0.001	

P(E)

T	F
0.002	0.998

Evidence J=T,M=F
in combination with B=F,E=F,A=F
weight = 0.05*0.99=0.0495

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

P(J|A)

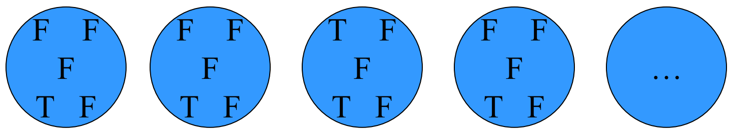
A	T	F
T	0.90	0.1
F	0.05	0.95

P(M|A)

A	T	F
T	0.7	0.3
F	0.01	0.99

Likelihood weighting

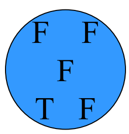
- Assume we have generated the following M samples:



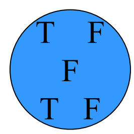
M

How to make examples consistent with the original distribution?

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



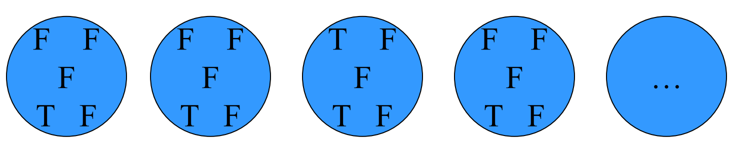
← **Weight 0.0495**



← **Weight 0.27**

Likelihood weighting

- **Likelihood weighting:**
  - **With every sample keep a weight with which it should count towards the estimate**



M

w $^{(1)}$ **w** $^{(2)}$ **w** $^{(3)}$ **w** $^{(4)}$

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