CS 1675 Intro to Machine Learning
Lecture 20

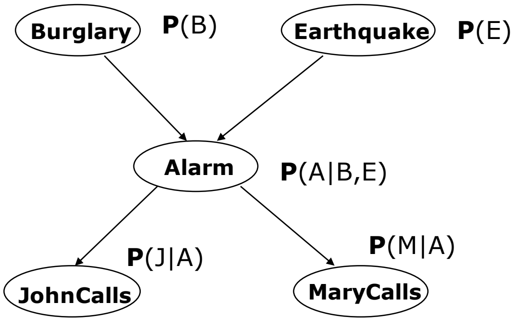
Bayesian belief networks IV
(Monte Carlo inference)

Milos Hauskrecht
milos@pitt.edu
5329 Sennott Square

Bayesian belief network

Belief network structure:

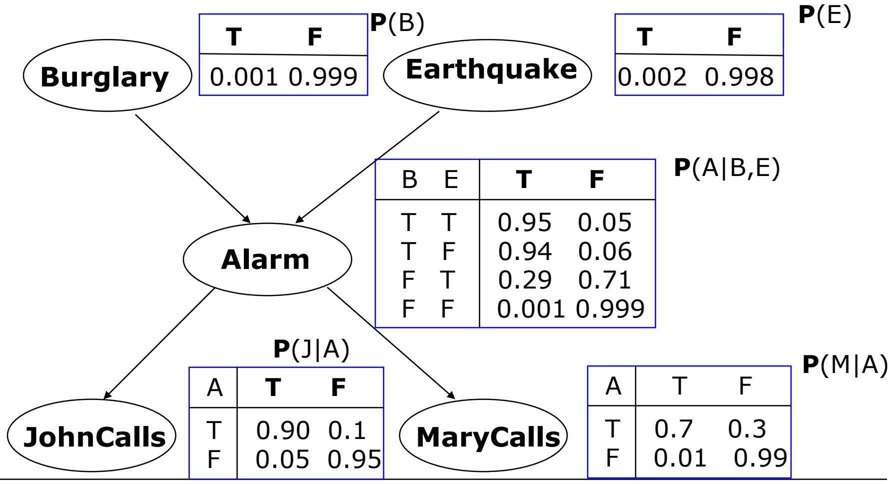
- **Nodes** = random variables
  Burglary, Earthquake, Alarm, Mary calls and John calls
- **Links** = direct (causal) dependencies between variables.
  The chance of Alarm being is influenced by Earthquake, The chance of John calling is affected by the Alarm



Bayesian belief network

2. Local conditional distributions

- relating variables and their parents



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

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

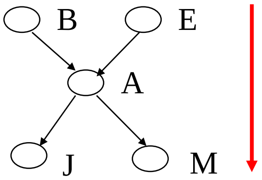
- **MC approximation**:
  - The probability is approximated using sample frequencies
  - **Example:**

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

 # samples with $B=T, J=T$

total # samples

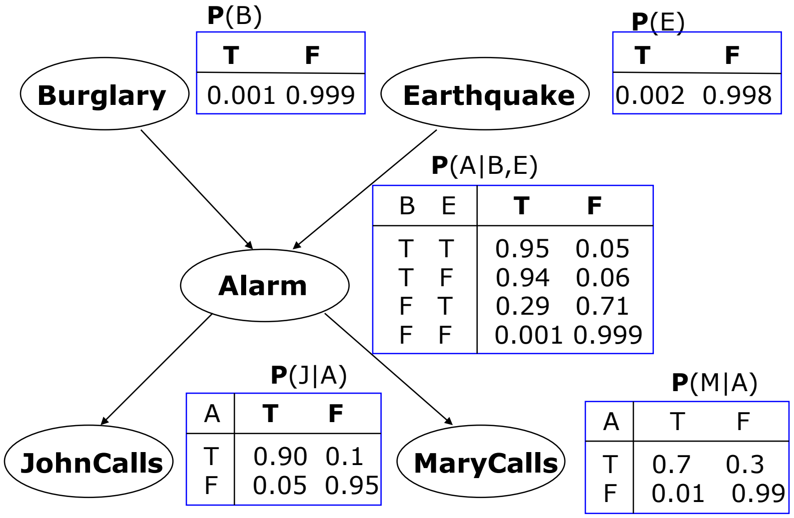
- **Sample generation: BBN sampling of the joint is easy**



Generate sample in a top down manner, following the links

- **One sample 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

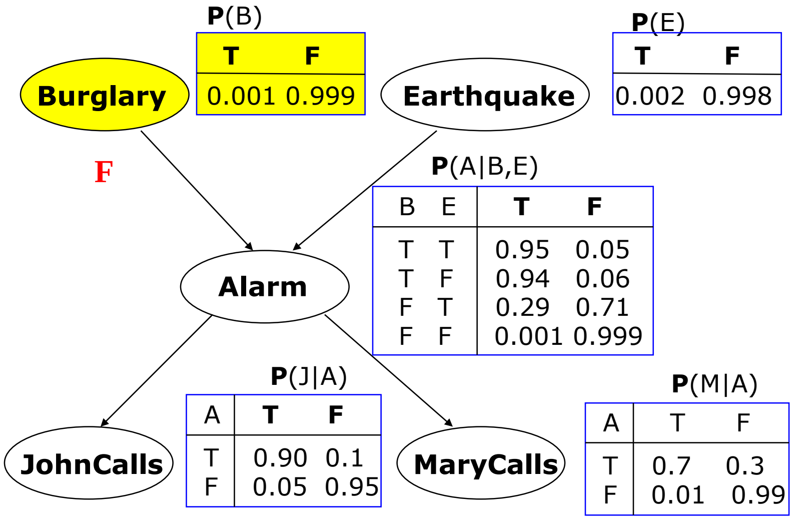
P(J|A)

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

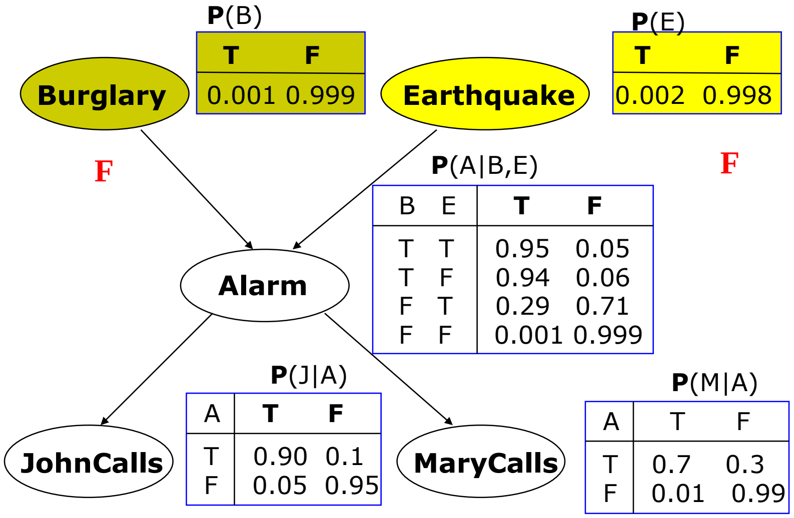
P(J|A)

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

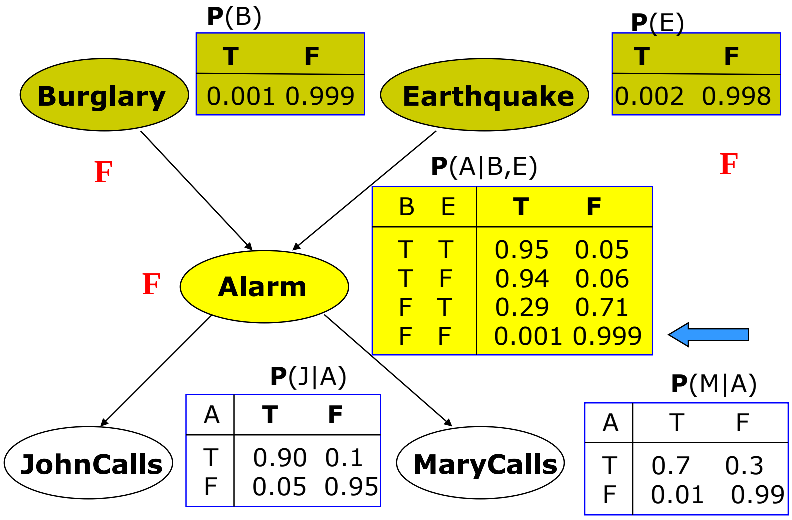
P(J|A)

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

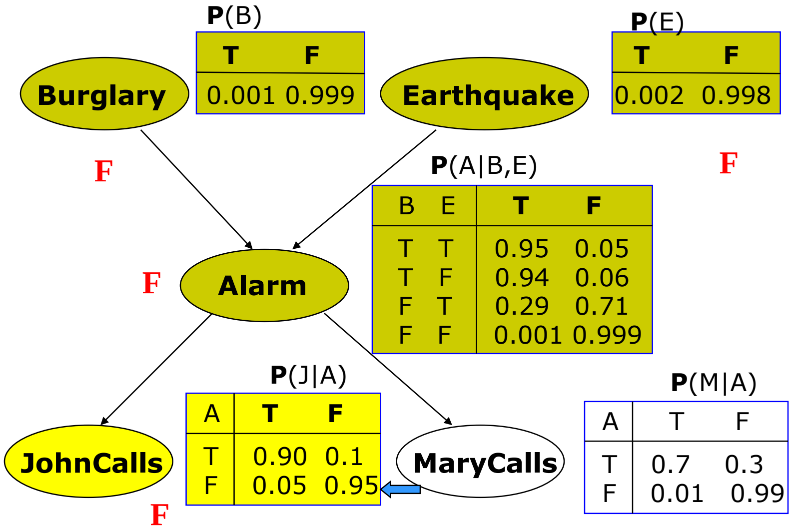
P(J|A)

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

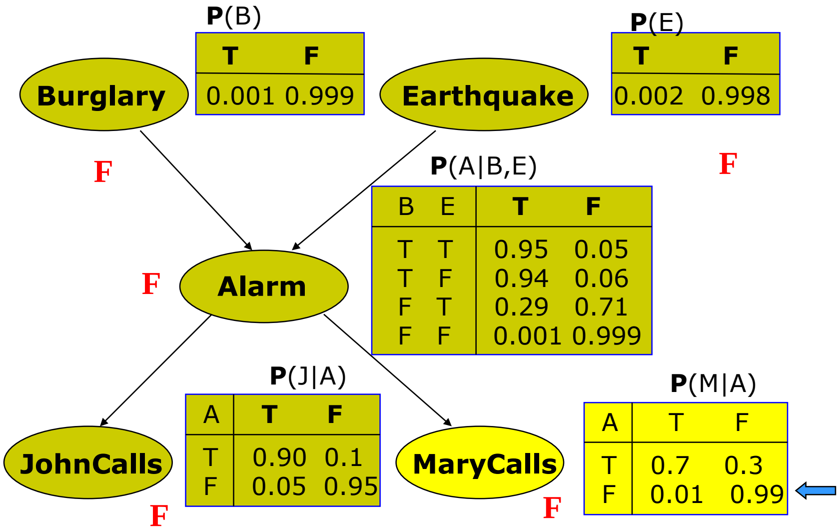
P(J|A)

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

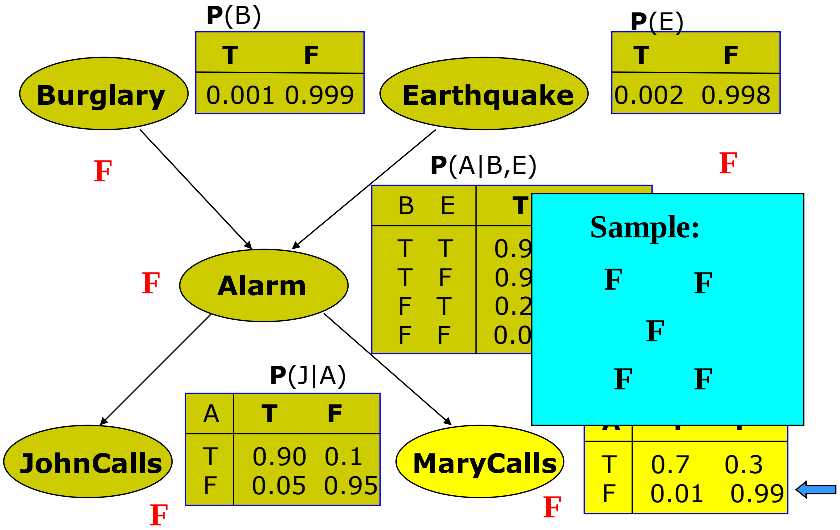
P(J|A)

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

Sample:

F **F**

F

F **F**

T	0.7	0.3
F	0.01	0.99

P(J|A)

A	T	F
T	0.90	0.1
F	0.05	0.95

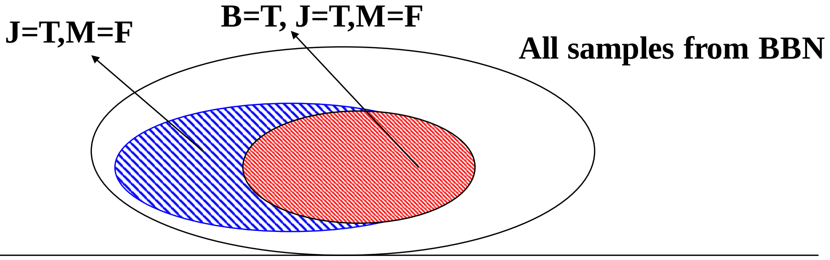
Monte Carlo approaches

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

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

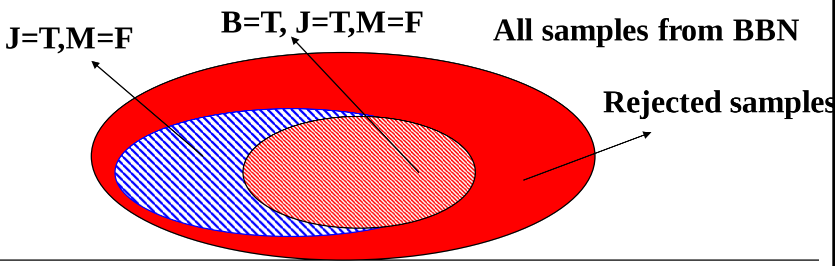
$$\tilde{P}(B=T \mid J=T, M=F) = \frac{N_{B=T,J=T,M=F}}{N_{J=T,M=F}}$$

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



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:

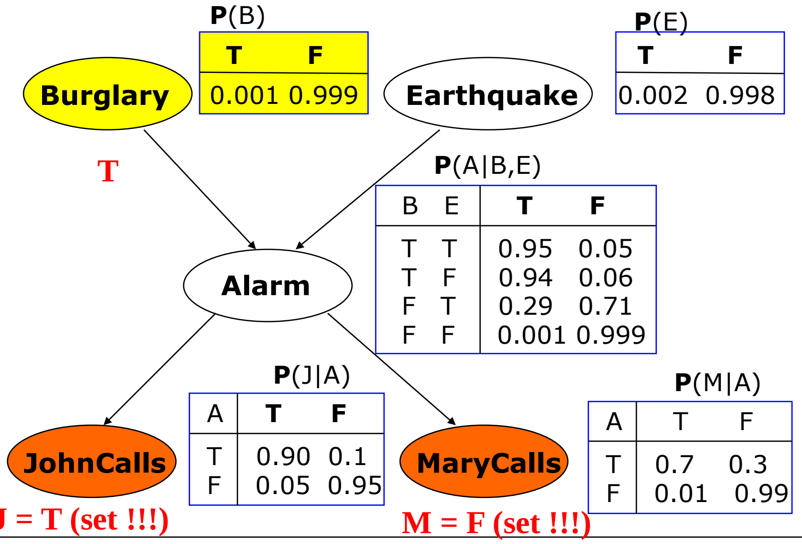
- **the distribution generated by enforcing the conditioning variables to set values is biased**
- simple counts are not sufficient to estimate the probabilities

Solution:

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

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

BBN likelihood weighting example



T

J = T (set !!!)

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**	**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

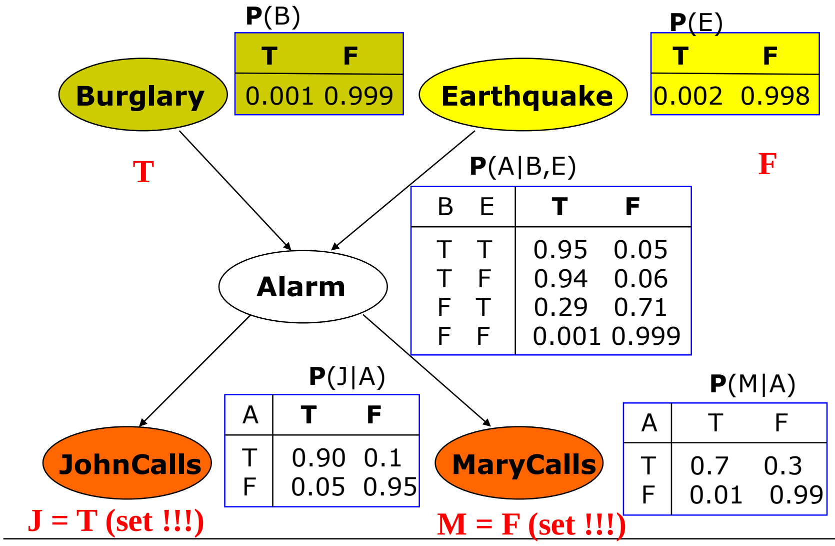
P(J|A)

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

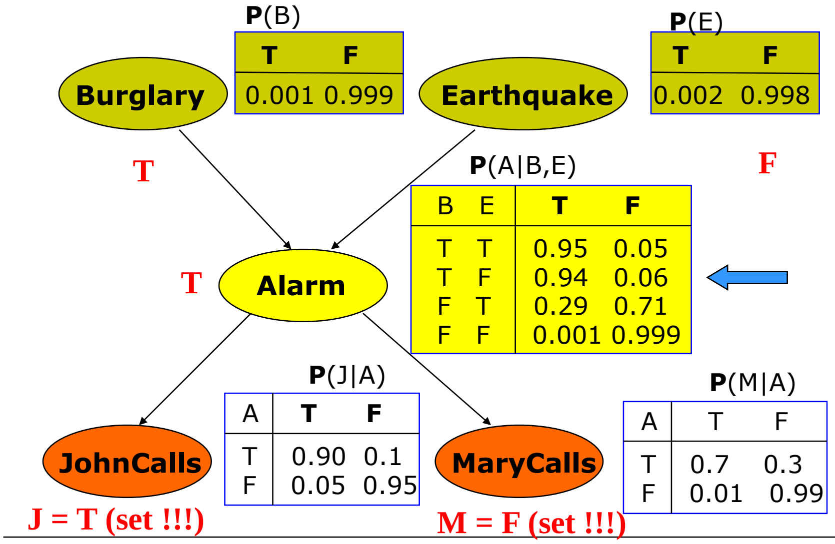
P(J|A)

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

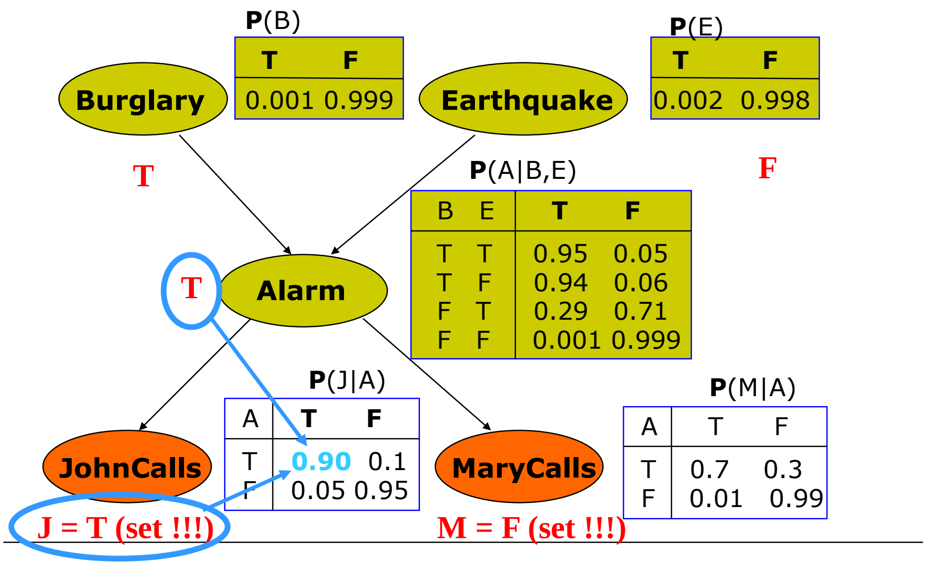
P(J|A)

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

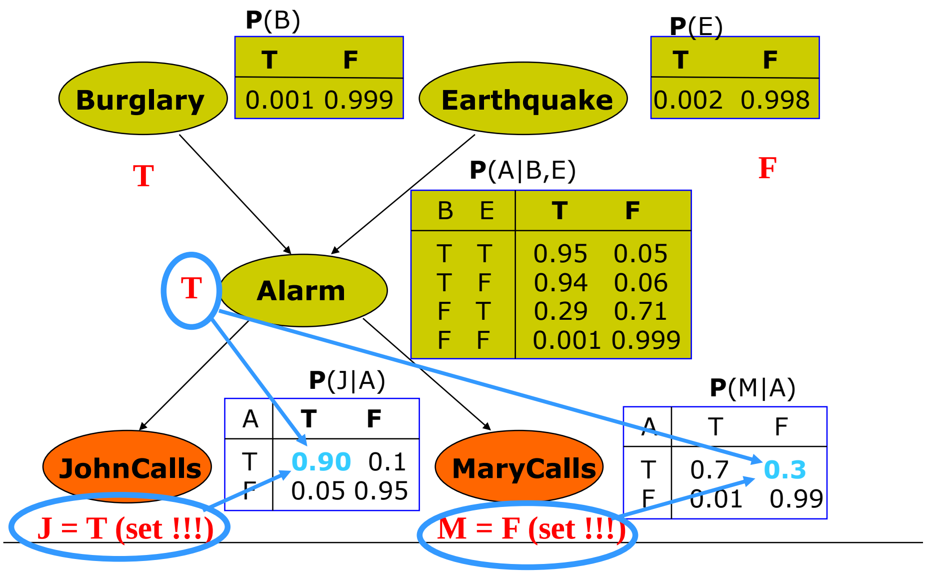
P(J|A)

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)

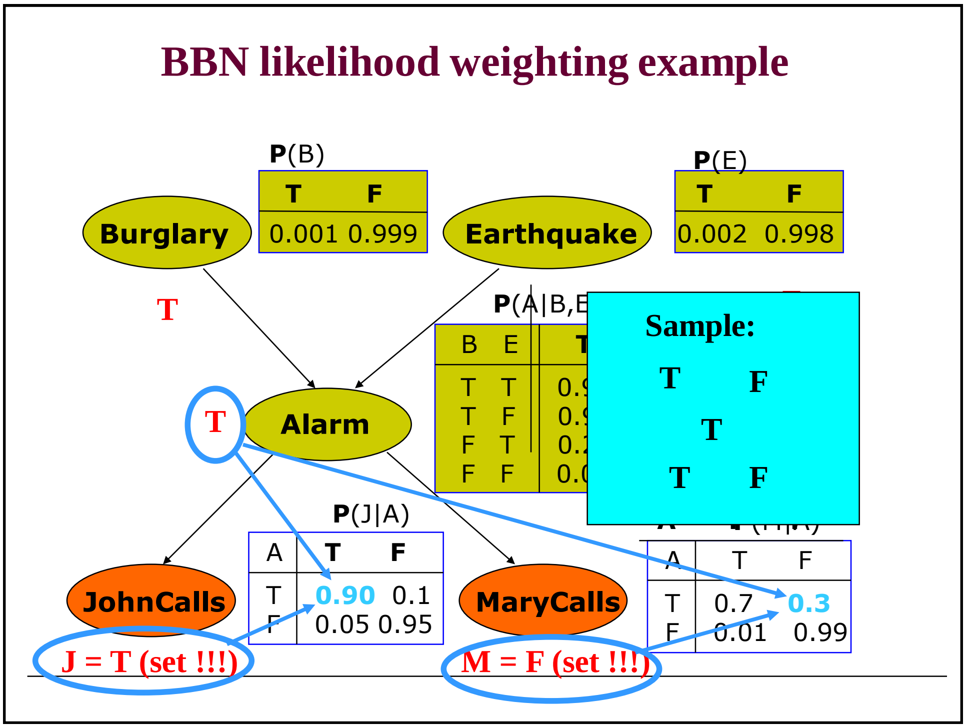
B	E	**T**	**F**
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

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

Burglary

Earthquake

T

Alarm

T

P(A|B,E

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

Sample:

T **F**

T

T **F**

P(J|A)

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

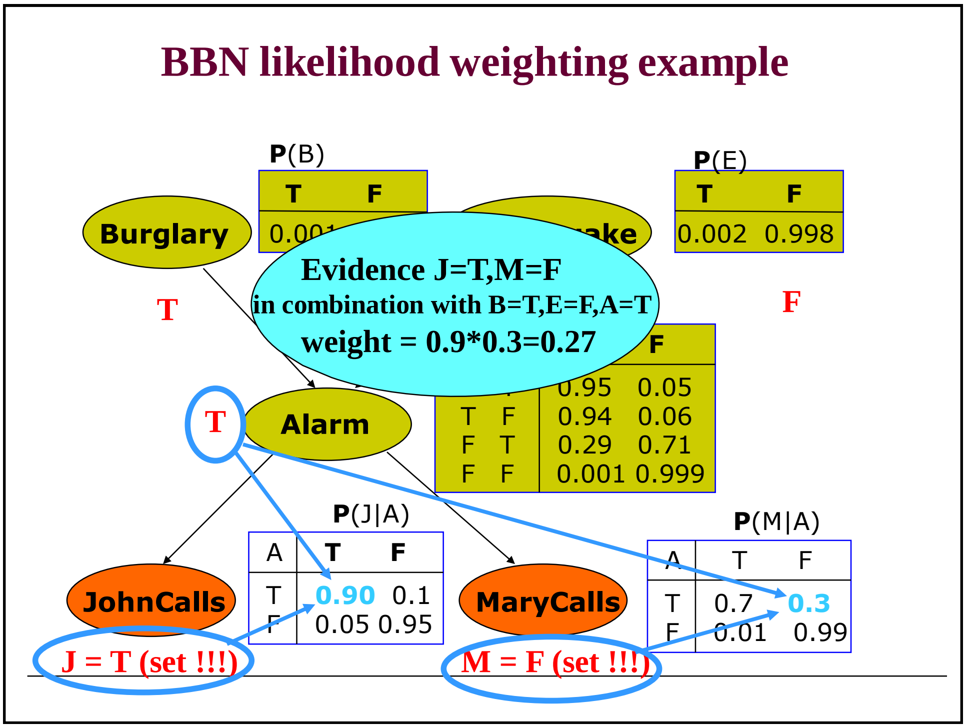
A	T	F
T	0.7	0.3
F	0.01	0.99

JohnCalls

MaryCalls

J = T (set !!!)

M = F (set !!!)



BBN likelihood weighting example

P(B)

T	**F**
0.001	

P(E)

T	**F**
0.002	0.998

Burglary

T

F

Evidence J=T,M=F

in combination with B=T,E=F,A=T

weight = 0.9*0.3=0.27

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

Alarm

T

P(J|A)

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

P(M|A)

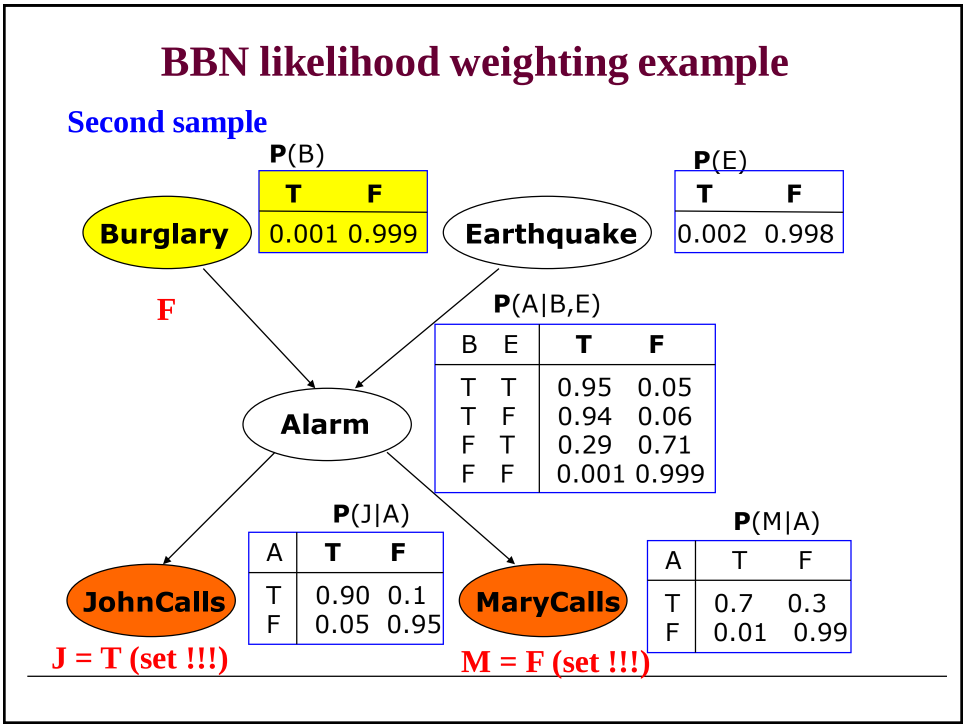
A	T	F
T	0.7	0.3
F	0.01	0.99

JohnCalls

MaryCalls

J = T (set !!!)

M = F (set !!!)



BBN likelihood weighting example

Second sample

P(B)

T	**F**
0.001	0.999

P(E)

T	**F**
0.002	0.998

F

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)

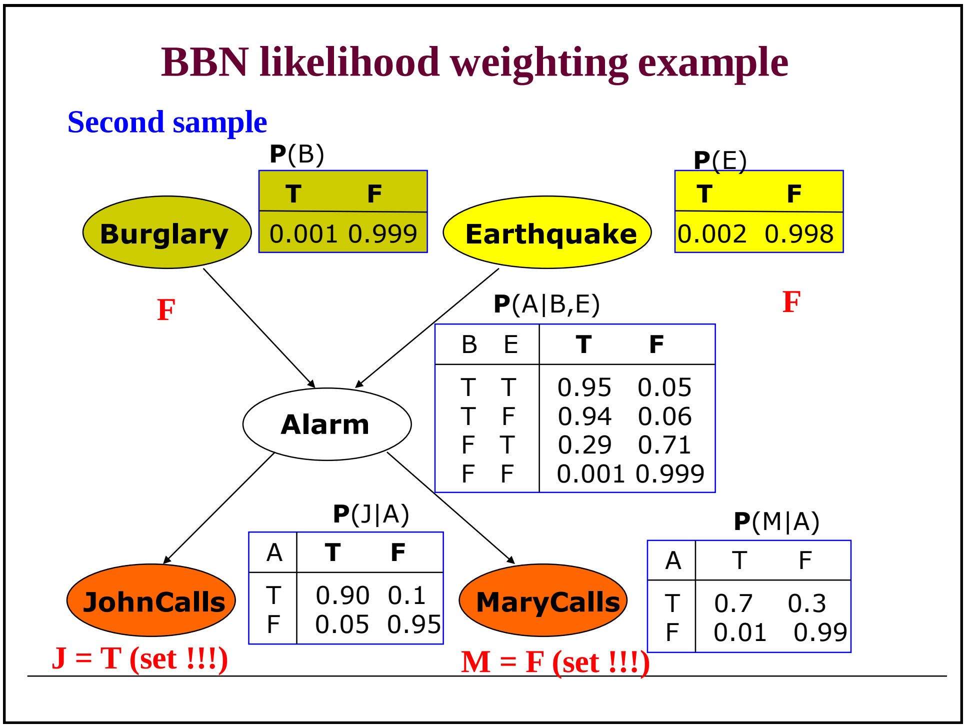
A	**T**	**F**
T	0.90	0.1
F	0.05	0.95

P(M|A)

A	T	F
T	0.7	0.3
F	0.01	0.99

J = T (set !!!)

M = F (set !!!)



BBN likelihood weighting example

Second sample

P(B)

T	**F**
0.001	0.999

P(E)

T	**F**
0.002	0.998

F

F

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)

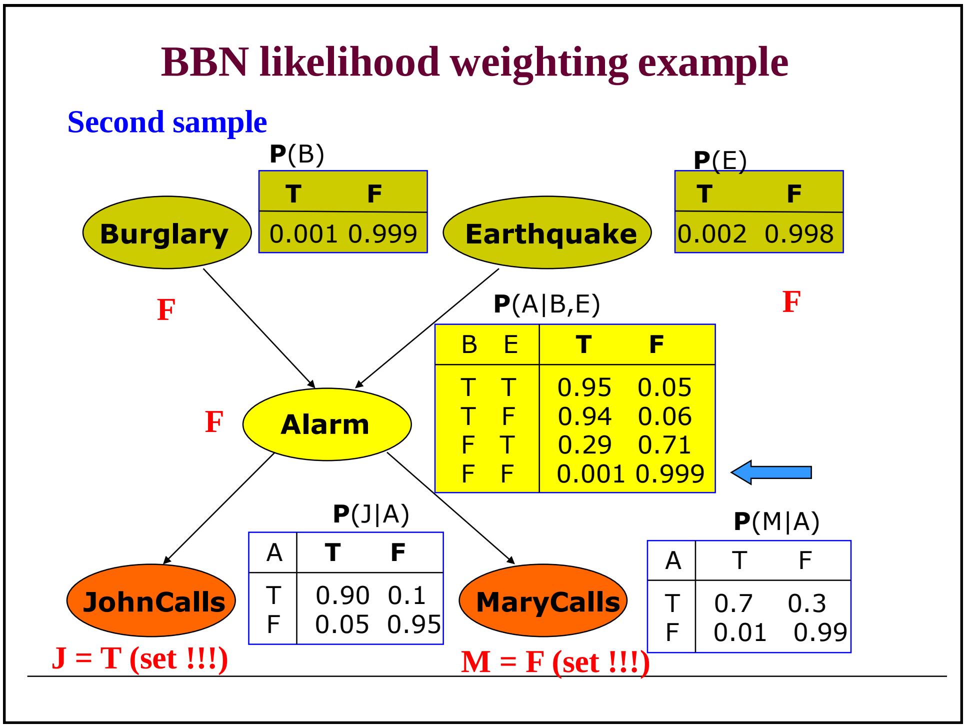
A	**T**	**F**
T	0.90	0.1
F	0.05	0.95

P(M|A)

A	T	F
T	0.7	0.3
F	0.01	0.99

J = T (set !!!)

M = F (set !!!)



BBN likelihood weighting example

Second sample

P(B)

T	**F**
0.001	0.999

Burglary

Earthquake

P(E)

T	**F**
0.002	0.998

F

F

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

F

Alarm

P(J|A)

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

P(M|A)

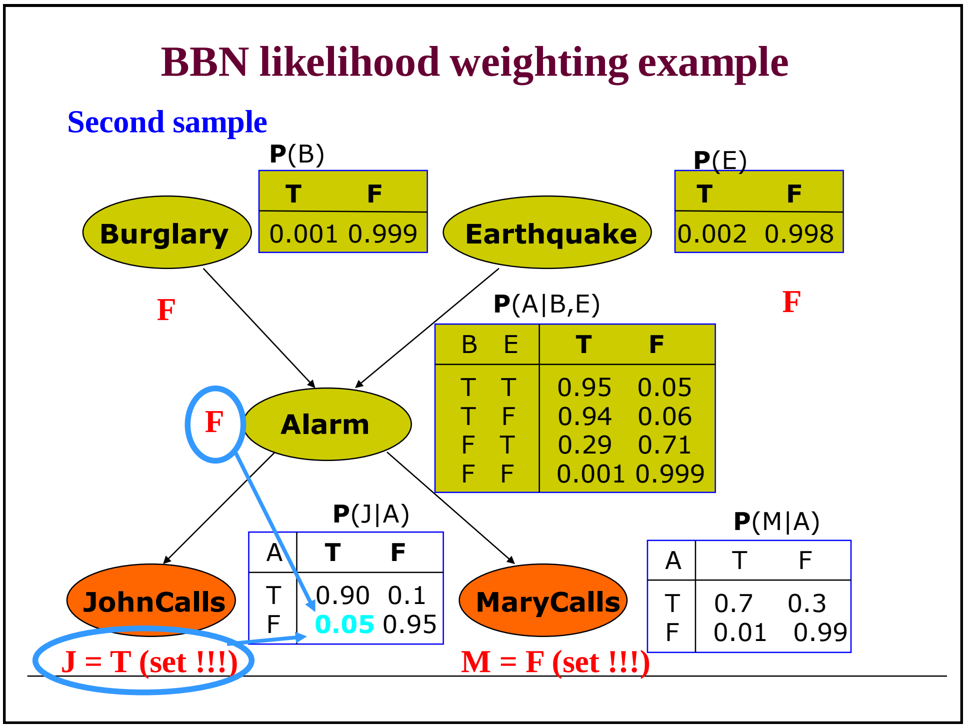
A	T	F
T	0.7	0.3
F	0.01	0.99

JohnCalls

MaryCalls

J = T (set !!!)

M = F (set !!!)



BBN likelihood weighting example

Second sample

P(B)

T	**F**
0.001	0.999

Burglary

Earthquake

P(E)

T	**F**
0.002	0.998

F

F

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

F

Alarm

P(J|A)

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

P(M|A)

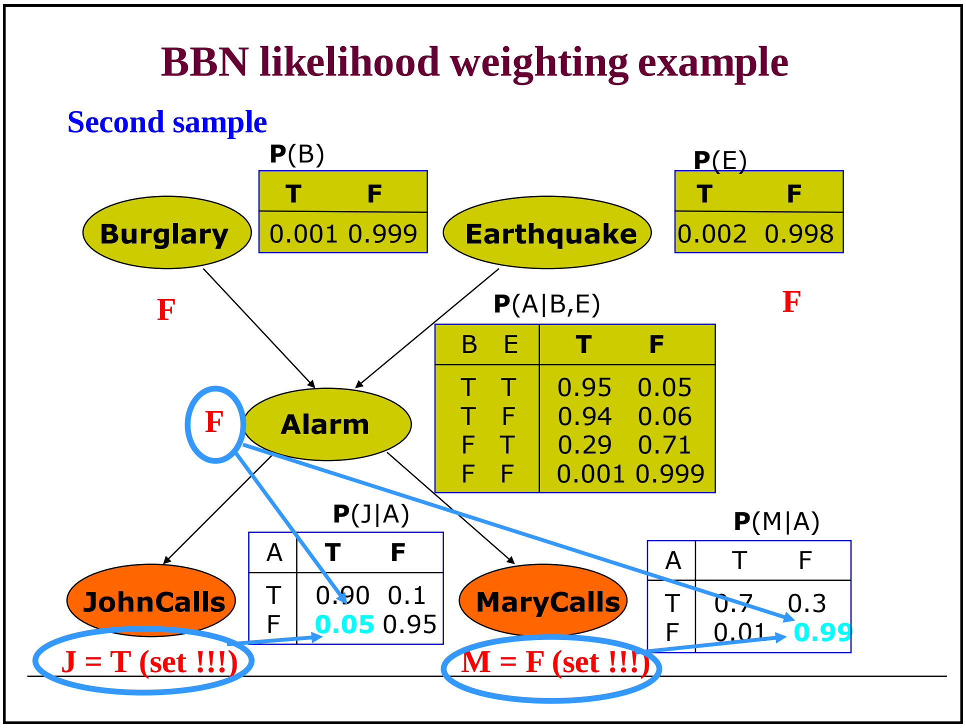
A	T	F
T	0.7	0.3
F	0.01	0.99

JohnCalls

MaryCalls

J = T (set !!!)

M = F (set !!!)



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)

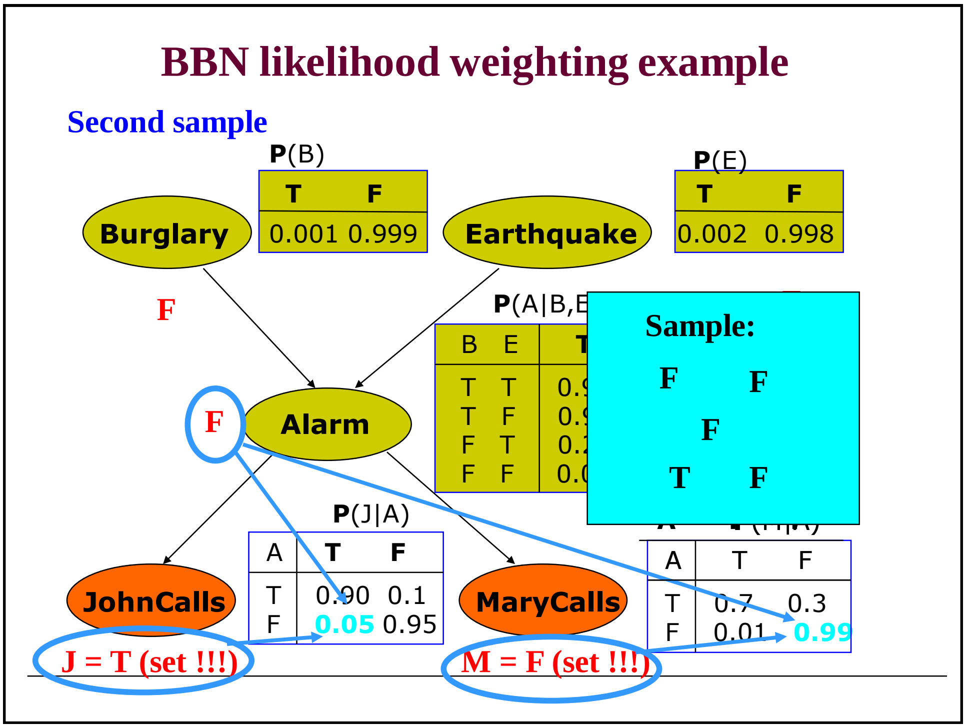
B	E	**T**	**F**
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

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
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

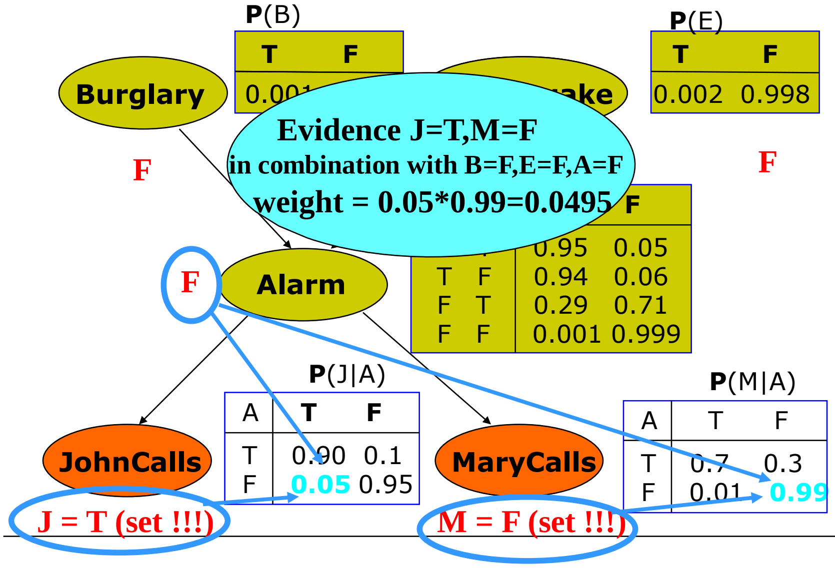
A	T	F
T	0.7	0.3
F	0.01	0.99

Sample:

F F
F
T F

BBN likelihood weighting example

Second sample



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

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)

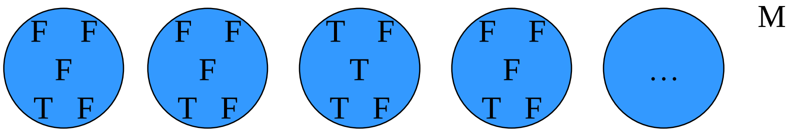
A	T	F
T	0.7	0.3
F	0.01	0.99

J = T (set !!!)

M = F (set !!!)

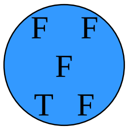
Likelihood weighting

- Assume we have generated the following M samples:

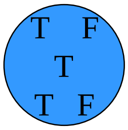


How to make the samples consistent?

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



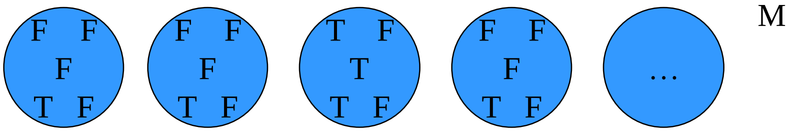
← **Weight 0.0495**



← **Weight 0.27**

Likelihood weighting

- Assume we have generated the following M samples:



0.0495 **0.0495** **0.27** **0.0495**

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