CS 1571 Introduction to AI
Lecture 21a

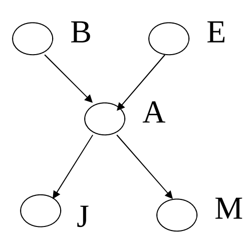
Bayesian belief networks: Inference

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

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

$$\mathbf{P}(X_i \mid pa(X_i))$$

Where:

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

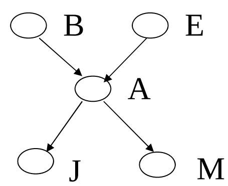
P(A|B,E)

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):

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



$$\mathbf{P}(B, E, A, J, M) = P(J \mid A)P(M \mid A)P(A \mid B, E)P(B)P(E)$$

Parameter complexity problem

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

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

- **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:

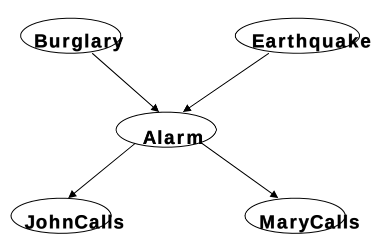
$$2^5 - 1 = 31$$

# of parameters of the BBN:

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

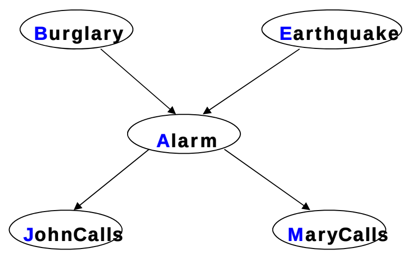
One parameter in every conditional is for free:

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



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)

$$P(J=T)=$$

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

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

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

Computational cost:

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

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

Inference in Bayesian network

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

Monte Carlo approaches

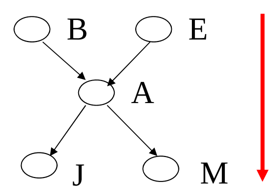
- **MC approximation**:
  - The probability is approximated using sample frequencies
  - **Example:**

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

$N_{B=T,J=T}$ ← *# samples with* $B=T, J=T$

N ← *total # samples*

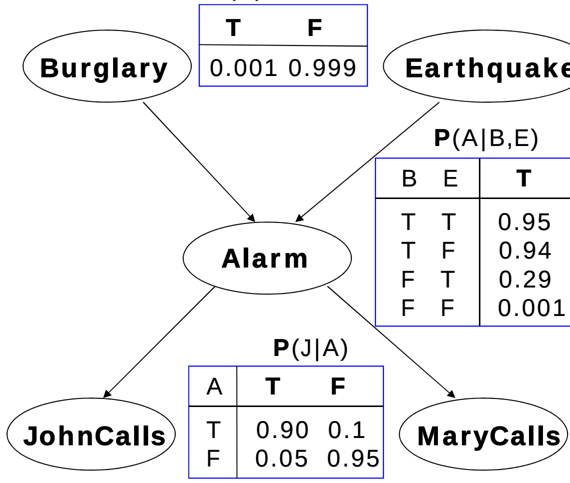
- **BBN sampling:**



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

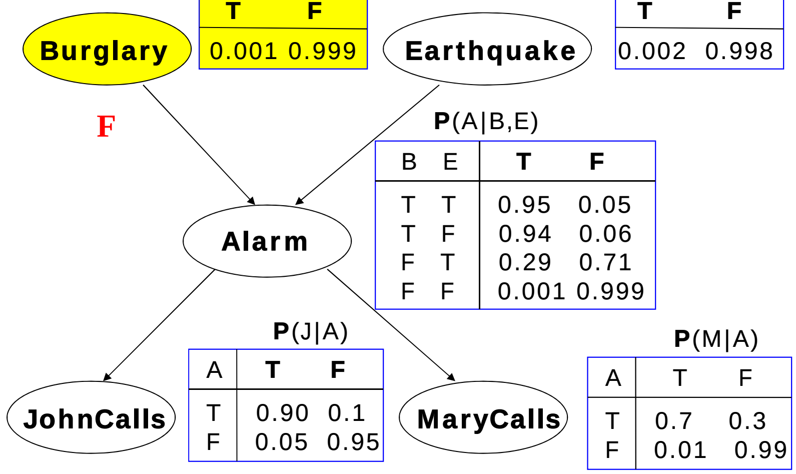
P(J|A)

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

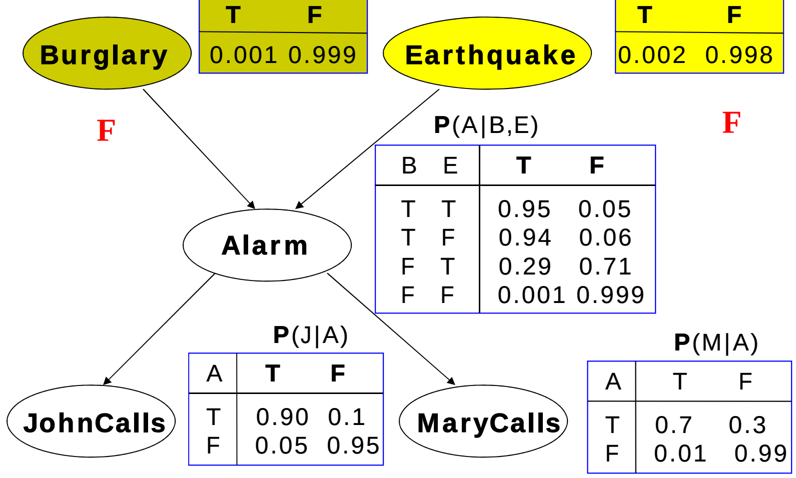
P(J|A)

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

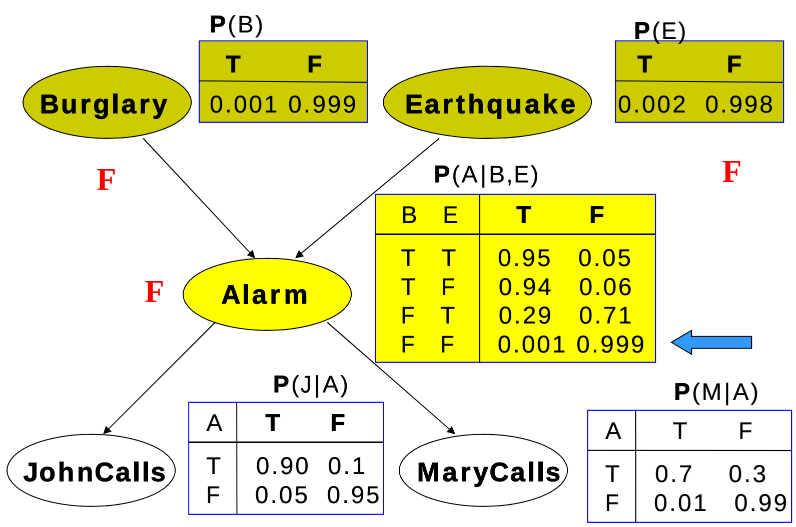
P(J|A)

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

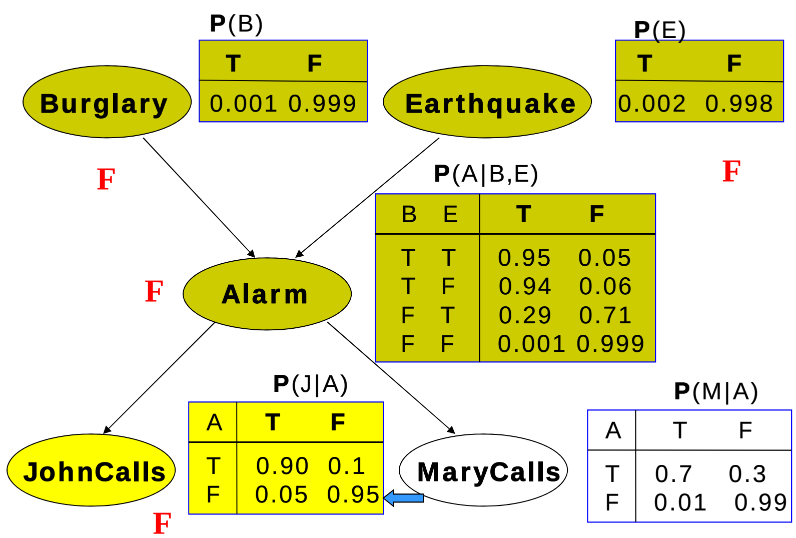
P(J|A)

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

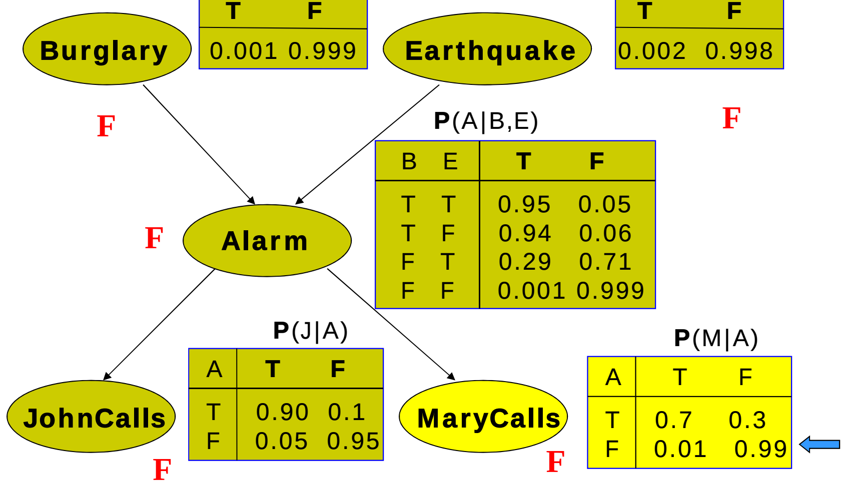
P(J|A)

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)

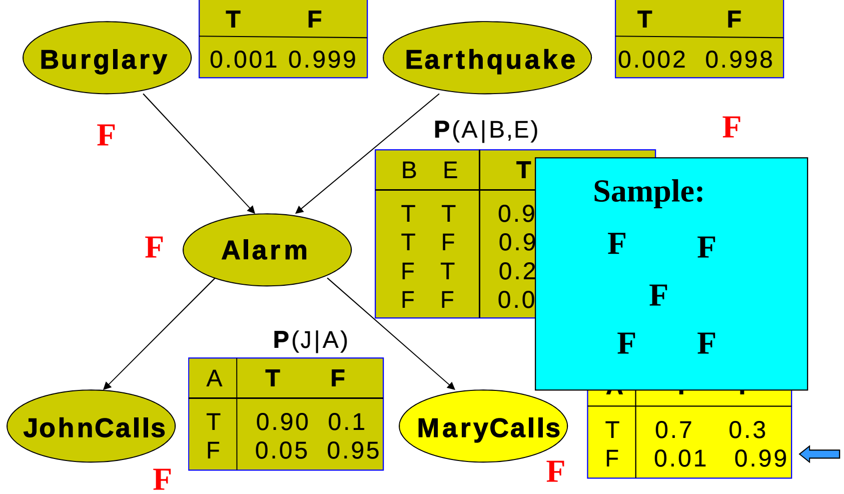
A	T	F
T	0.90	0.1
F	0.05	0.95

P(M|A)

A	T	F
T	0.7	0.3
F	0.01	0.99

CS 1571 Intro to AI M. Hauskrecht

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

CS 1571 Intro to AI M. Hauskrecht

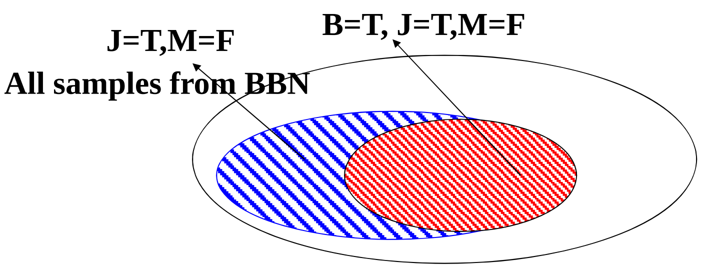
Monte Carlo approaches

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

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

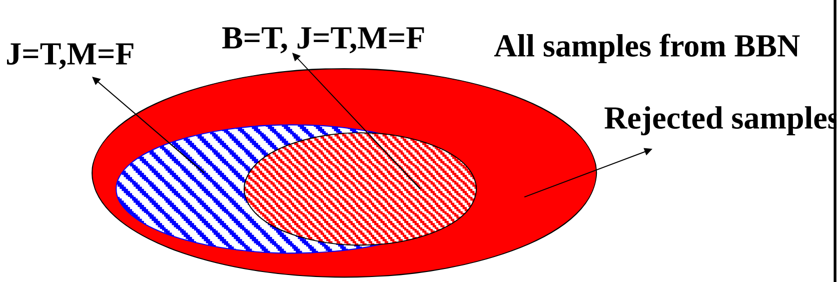
$N_{B=T,J=T,M=F}$: $\#\ samples\ \ with\ \ B=T, J=T, M=F$

$N_{J=T,M=F}$: $\#\ samples\ \ with\ \ J=T, M=F$



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

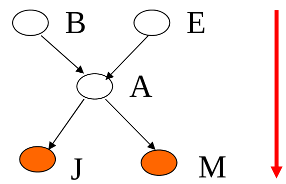
Idea: generate only samples consistent with an evidence (or the conditioning event)

– Benefit: Avoids inefficiencies of rejection sampling

Example:

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

- **Fix values of** J=T, M=F
- **Sample the rest to down**



Problem:

- **?**

Likelihood weighting

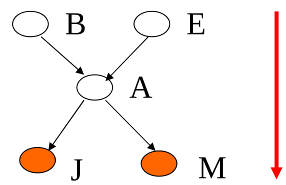
Idea: generate only samples consistent with an evidence (or the conditioning event)

– Benefit: Avoids inefficiencies of rejection sampling

Example:

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

- **Fix values of** J=T, M=F
- **Sample the rest to down**



Problem:

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

Likelihood weighting

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

– Benefit: Avoids inefficiencies of rejection sampling

Problem:

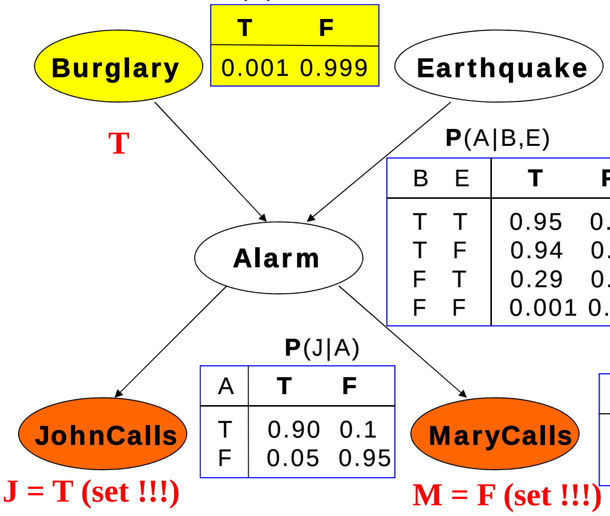
- **the distribution generated by enforcing the conditioning variables to set values is biased**

Solution:

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

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

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

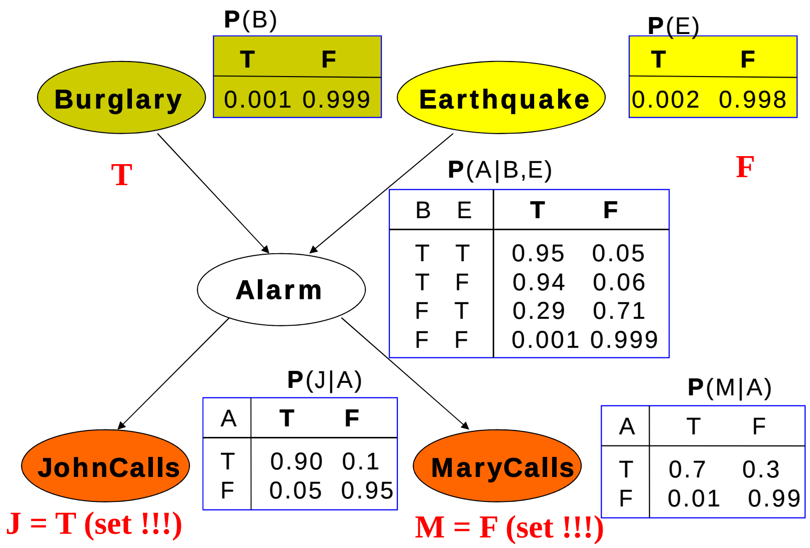
P(J|A)

A	T	F
T	0.90	0.1
F	0.05	0.95

P(M|A)

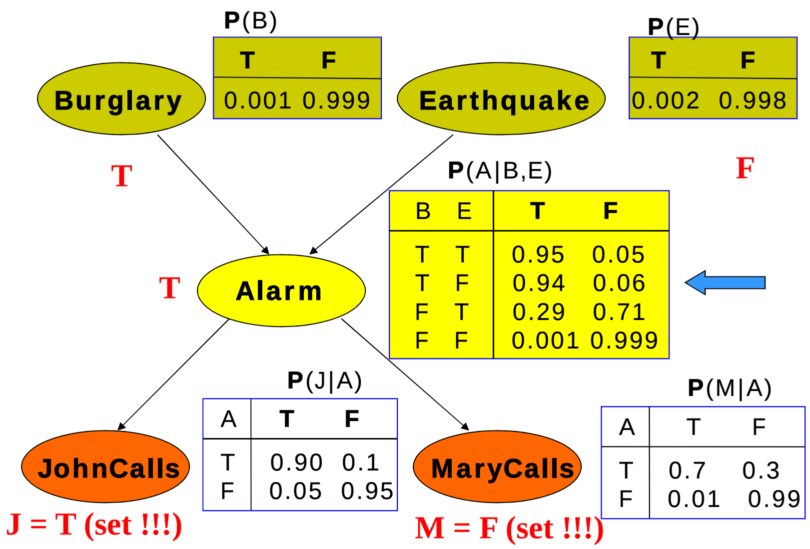
A	T	F
T	0.7	0.3
F	0.01	0.99

BBN likelihood weighting example



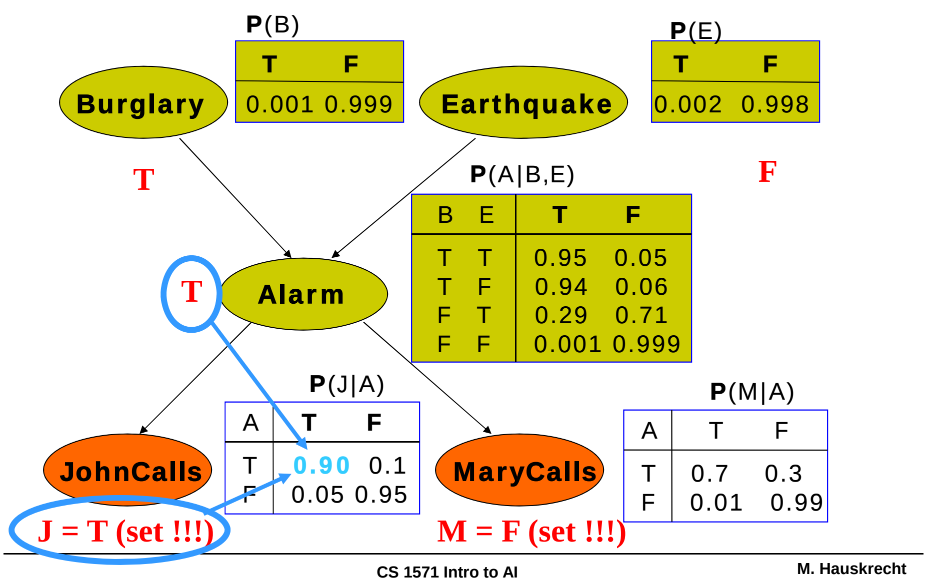
CS 1571 Intro to AI M. Hauskrecht

BBN likelihood weighting example



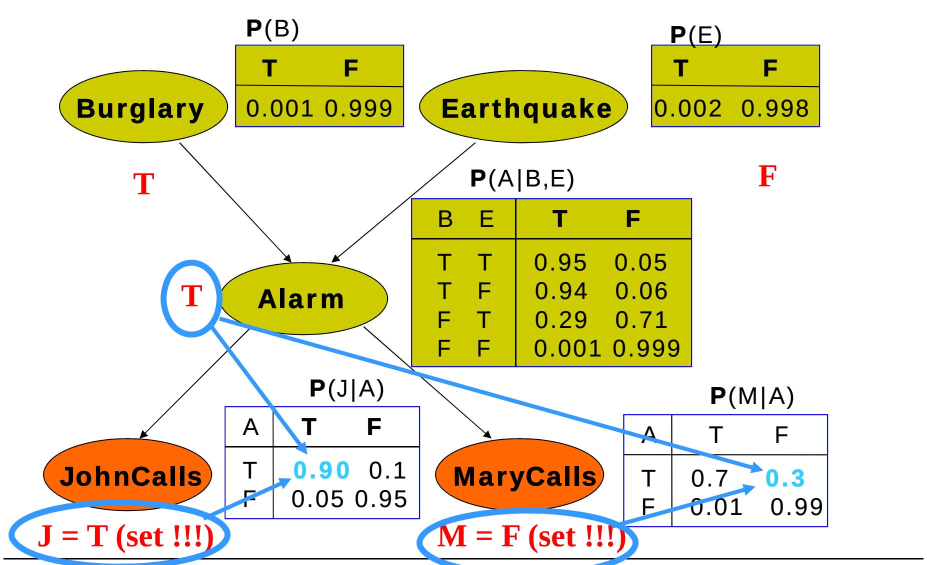
CS 1571 Intro to AI M. Hauskrecht

BBN likelihood weighting example



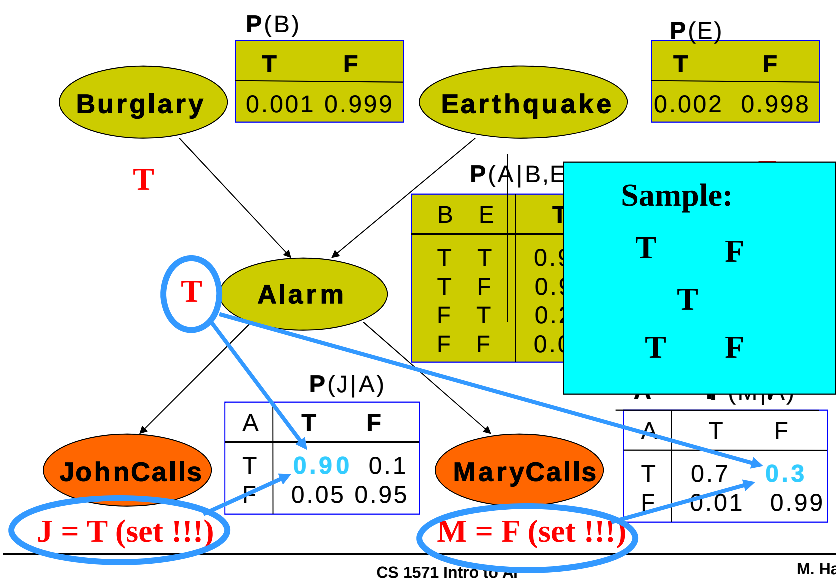
CS 1571 Intro to AI M. Hauskrecht

BBN likelihood weighting example

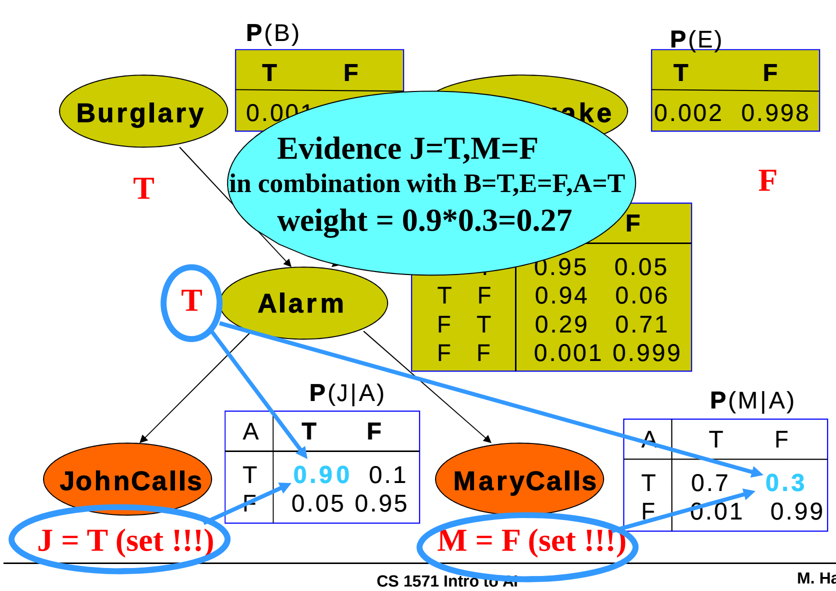


CS 1571 Intro to AI M. Hauskrecht

BBN likelihood weighting example

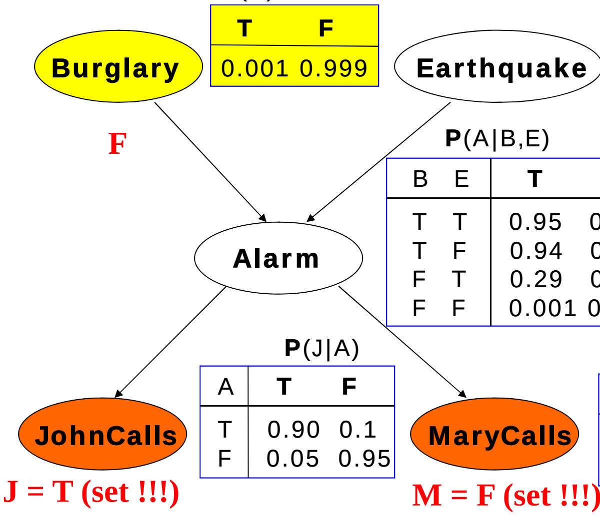


BBN likelihood weighting example



BBN likelihood weighting example

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)

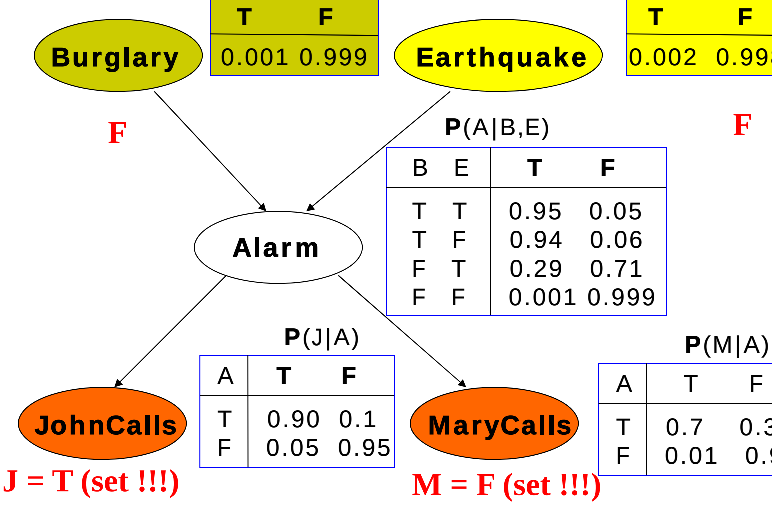
A	T	F
T	0.90	0.1
F	0.05	0.95

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)

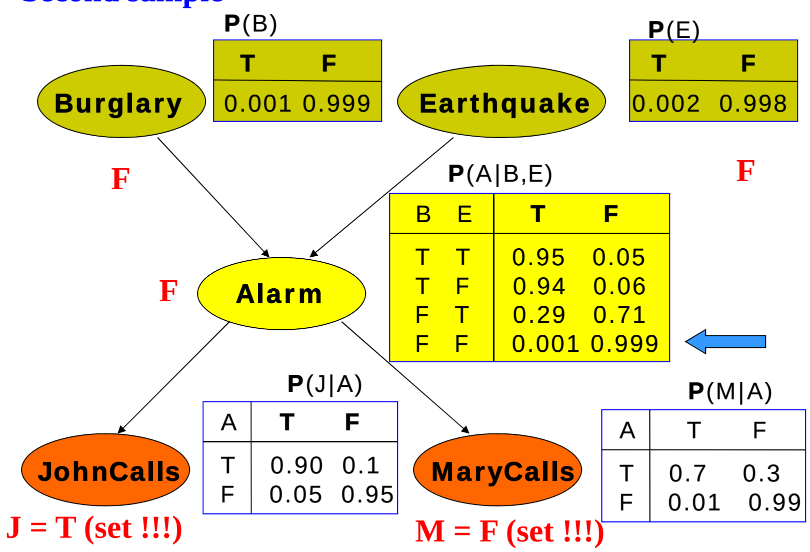
A	T	F
T	0.90	0.1
F	0.05	0.95

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)

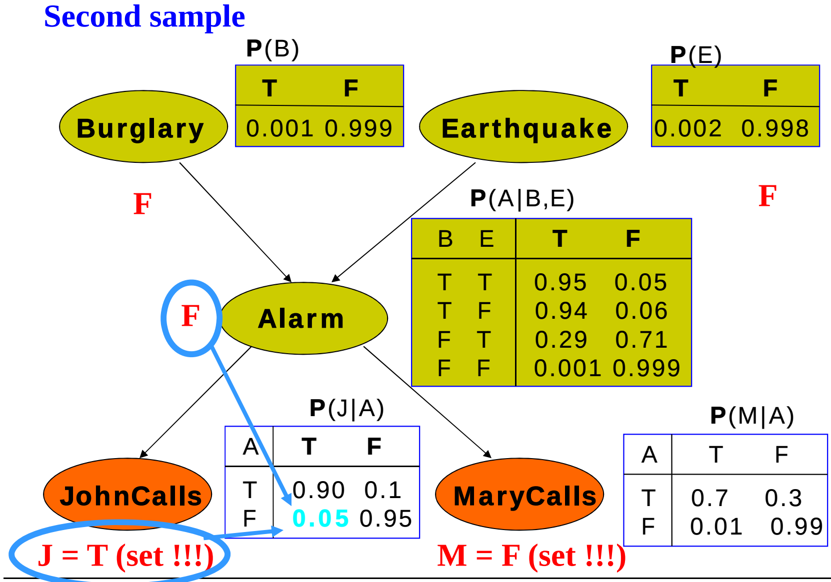
A	T	F
T	0.90	0.1
F	0.05	0.95

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)

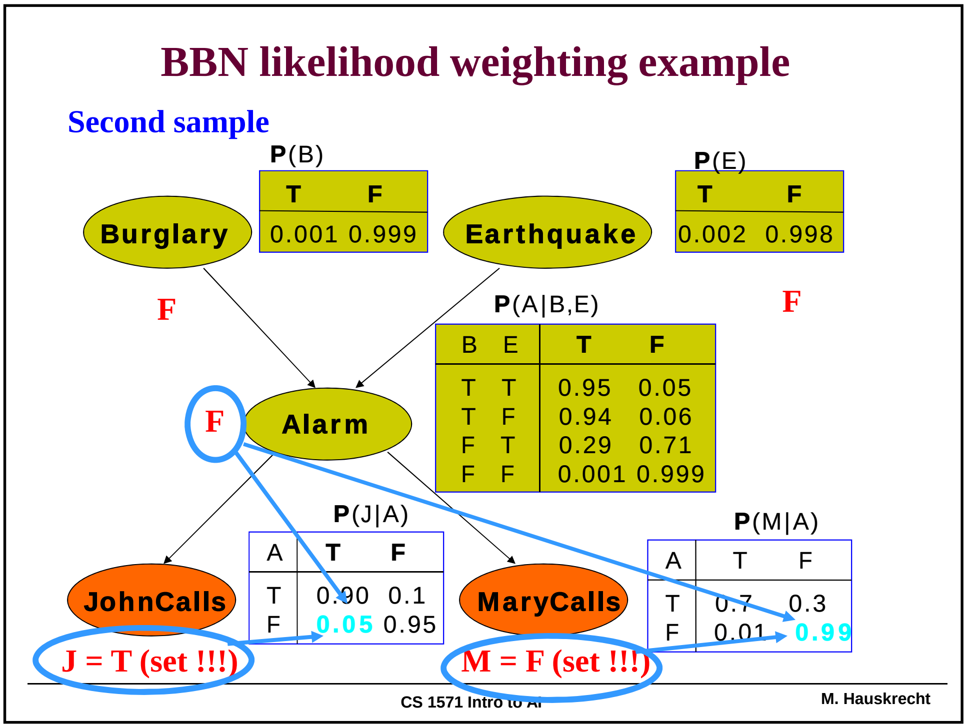
B	E	T	F
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

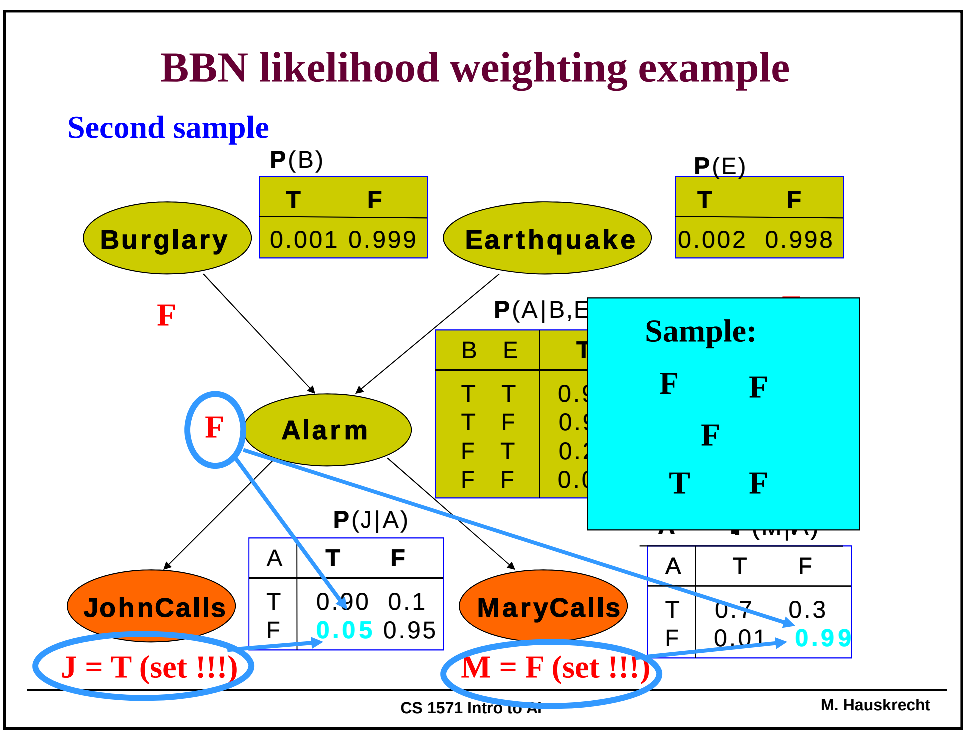
P(J|A)

A	T	F
T	0.90	0.1
F	0.05	0.95

P(M|A)

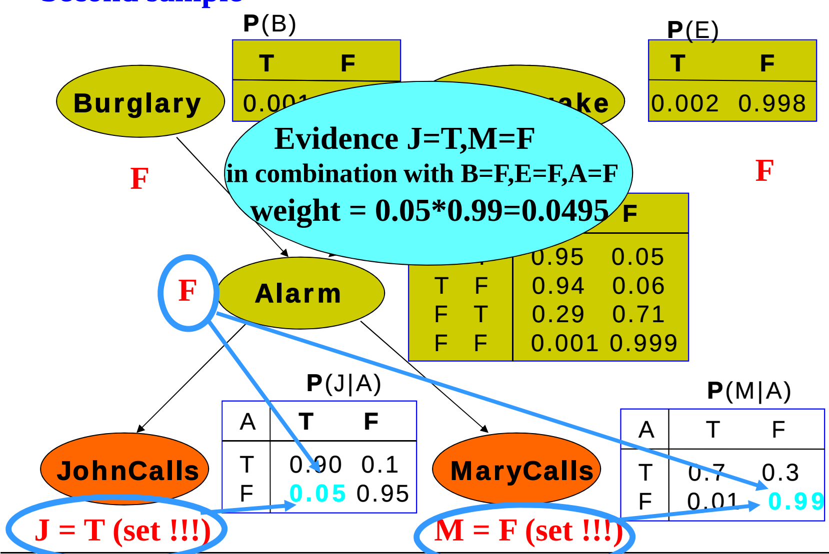
A	T	F
T	0.7	0.3
F	0.01	0.99


BBN likelihood weighting example
Second sample
P(B)
T F
0.001 0.999
P(E)
T F
0.002 0.998
Burglary
Earthquake
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 !!!)
CS 1571 Intro to AI
M. Hauskrecht


BBN likelihood weighting example
Second sample
P(B)
T F
0.001 0.999
P(E)
T F
0.002 0.998
Burglary
Earthquake
F
P(A|B,E
B E
T T
T F
F T
F F
Sample:
F F
F
T F
F
Alarm
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 !!!)
CS 1571 Intro to AI
M. Hauskrecht

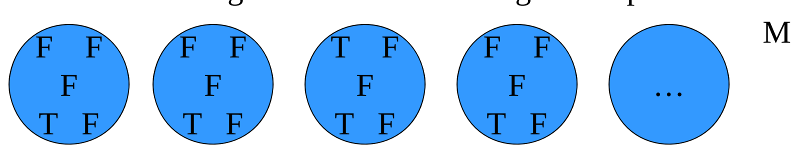
BBN likelihood weighting example

Second sample

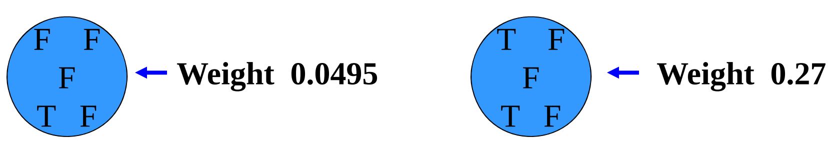


Likelihood weighting

- Assume we have generated the following M samples:



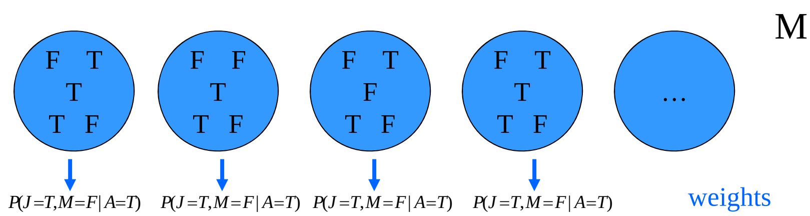
How to make the samples consistent? Weight each sample by probability with which it agrees with the conditioning evidence P(e).



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

Likelihood weighting

- Assume M samples where evidence is enforced:



- We can use P(e) to weight each sample and correct the bias.

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

M. Hauskrecht