

# Approximate Inference in Bayes Nets

## Sampling based methods

Mausam

(Based on slides by Jack Breese and  
Daphne Koller)

# Intuition

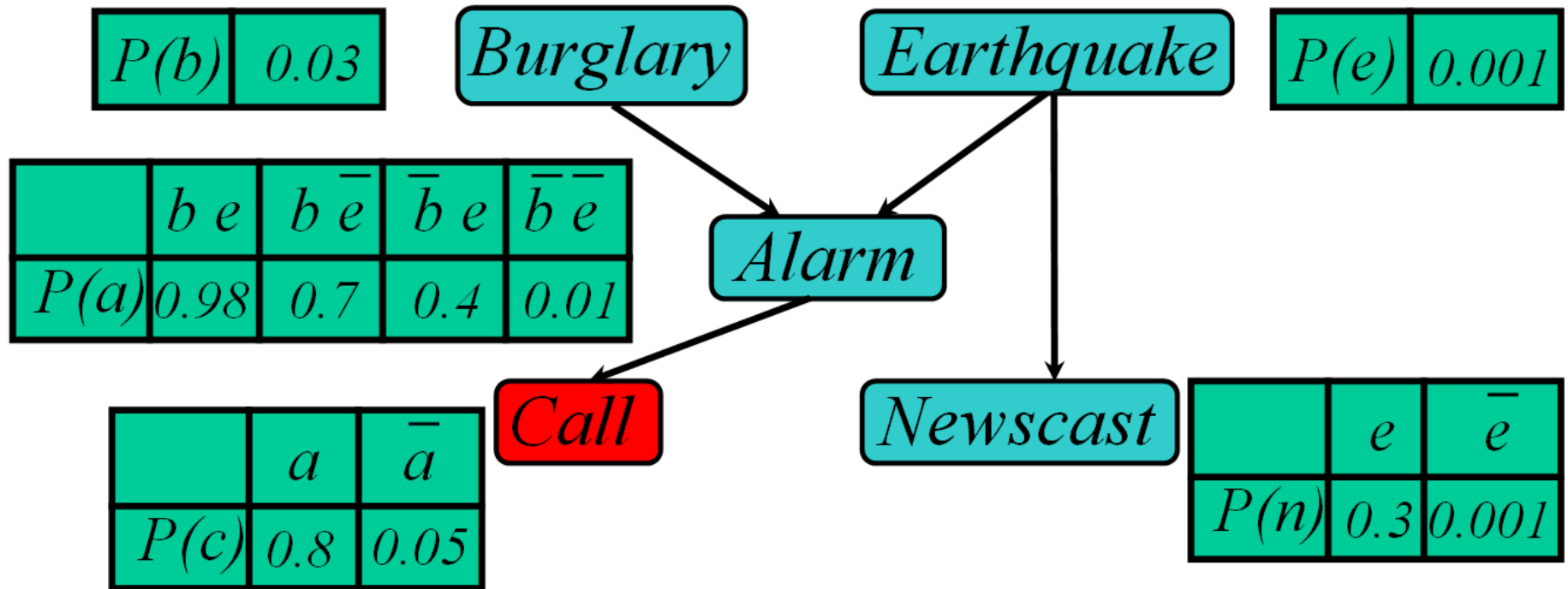
- Suppose I have a coin whose  $p(\text{heads})$  is unknown
- How could I estimate it?
- When will I get the correct probability?
- Bayes Net inference is not a learning problem
  - But similar intuitions apply
  - In particular, generate samples from a Bayes net
  - But the samples should be unbiased!

# Sampling

- Samples should be representative of the world
- Samples:  $P(\text{people} > 60 \text{ yrs age in Delhi})$ 
  - Computer Science class
  - Call on landline
  - Call on cellphone
  - Check facebook...
  - Count at election booth

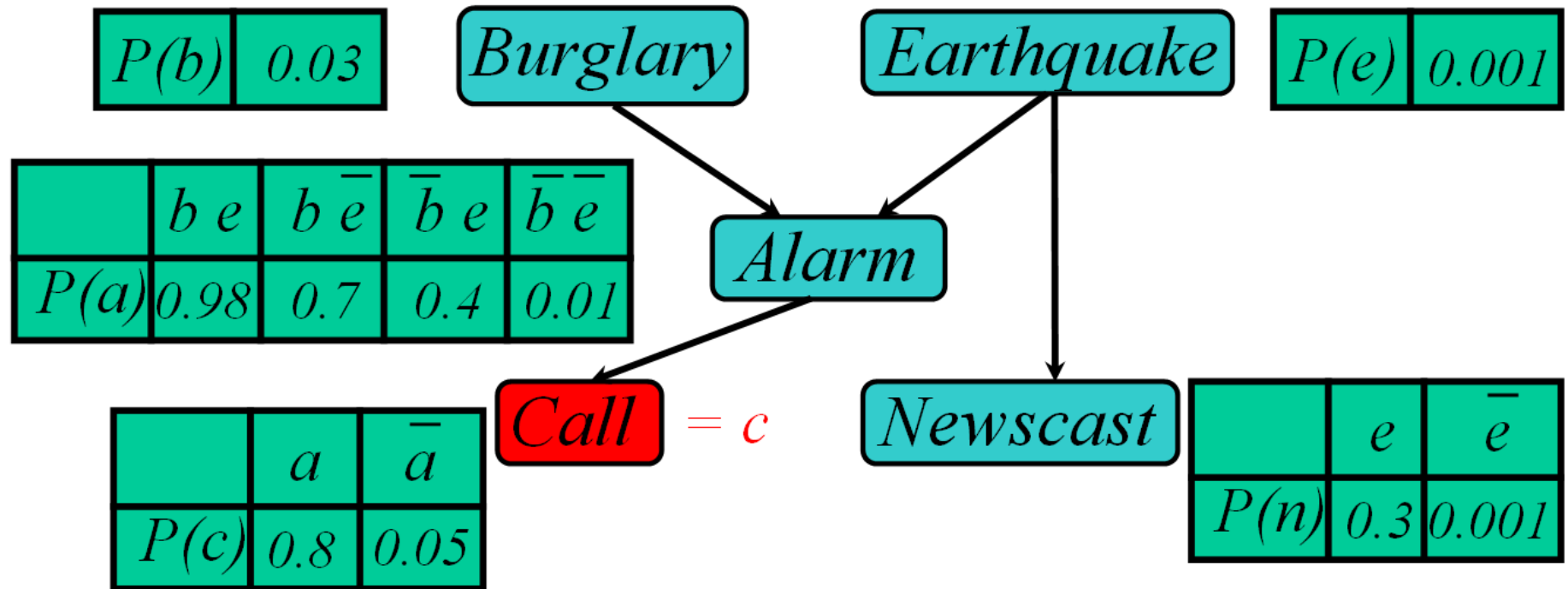
# Bayes Nets is a generative model

- We can easily generate samples from the distribution represented by the Bayes net
  - Generate one variable at a time in topological order

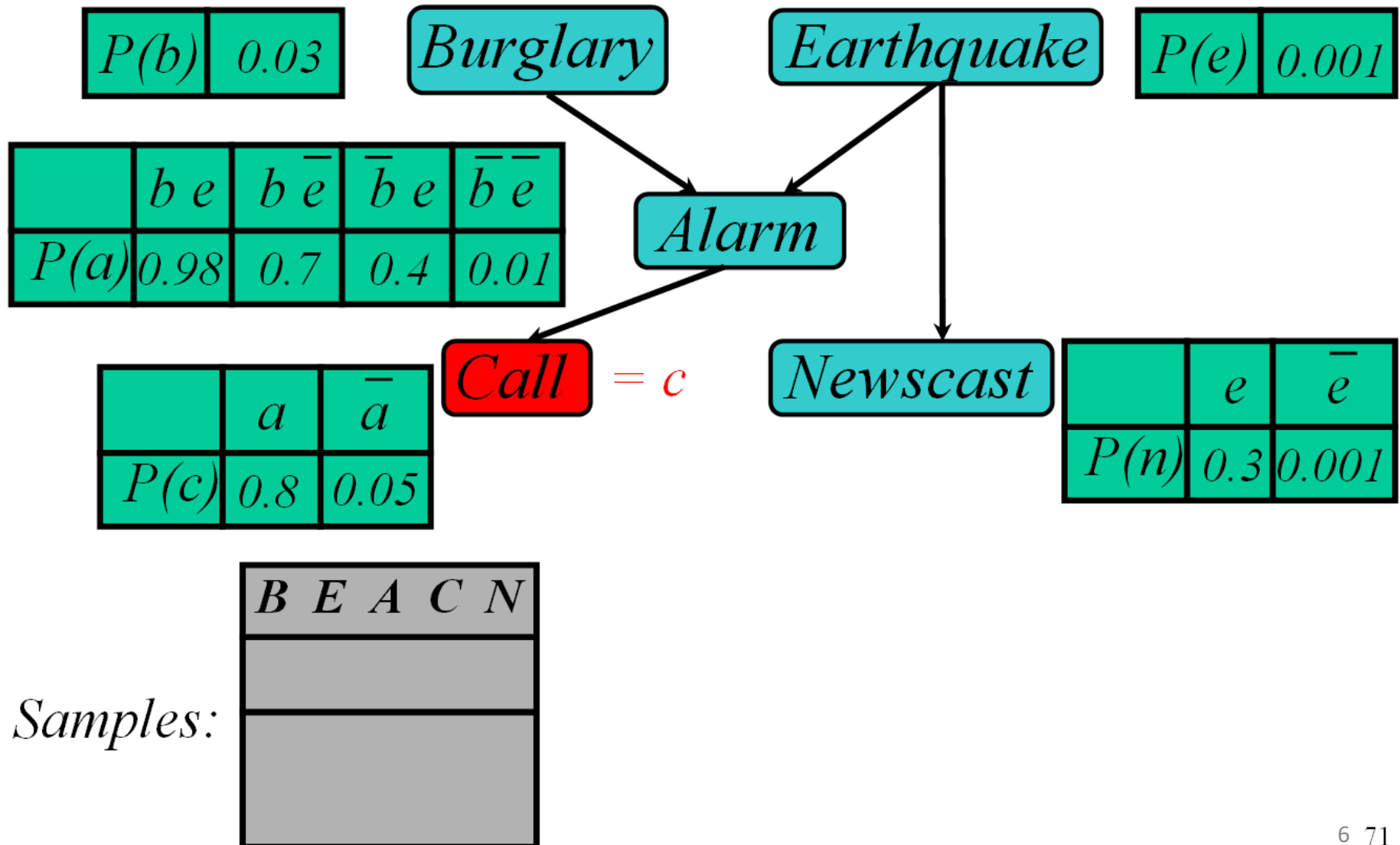


Use the samples to compute marginal probabilities, say  $P(c)$

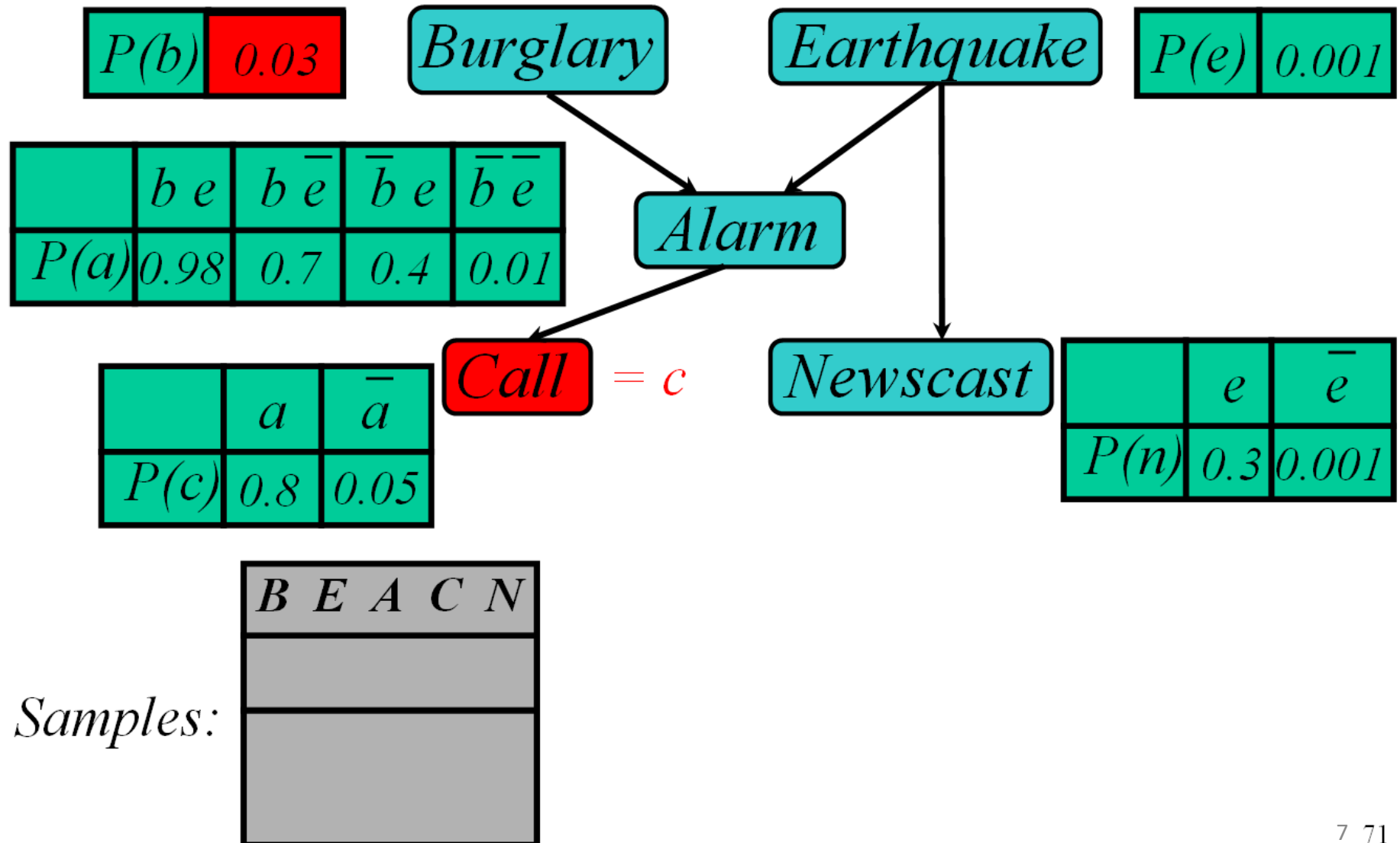
# Stochastic simulation $P(B|C)$



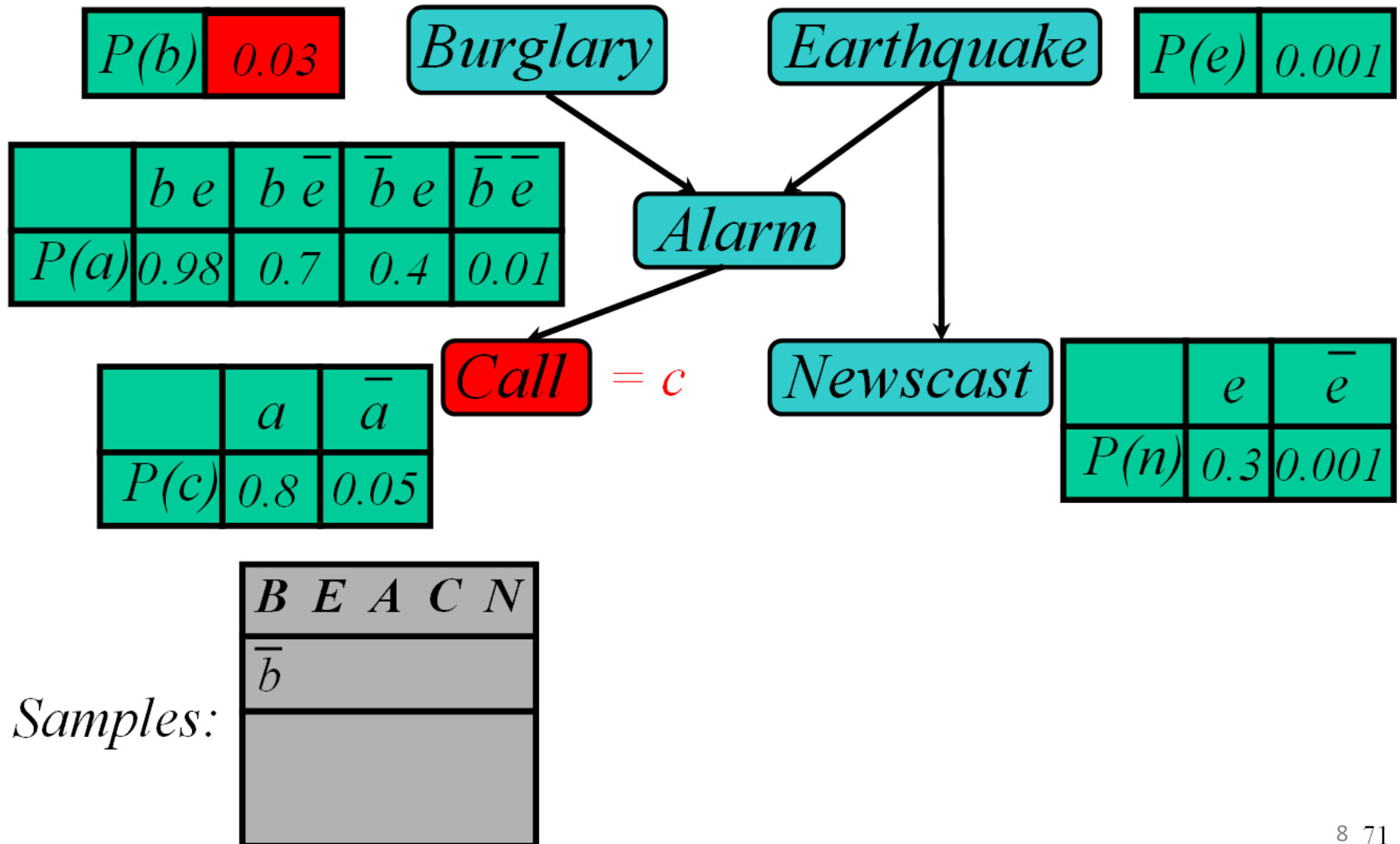
# Stochastic simulation $P(B|C)$



# Stochastic simulation $P(B|C)$

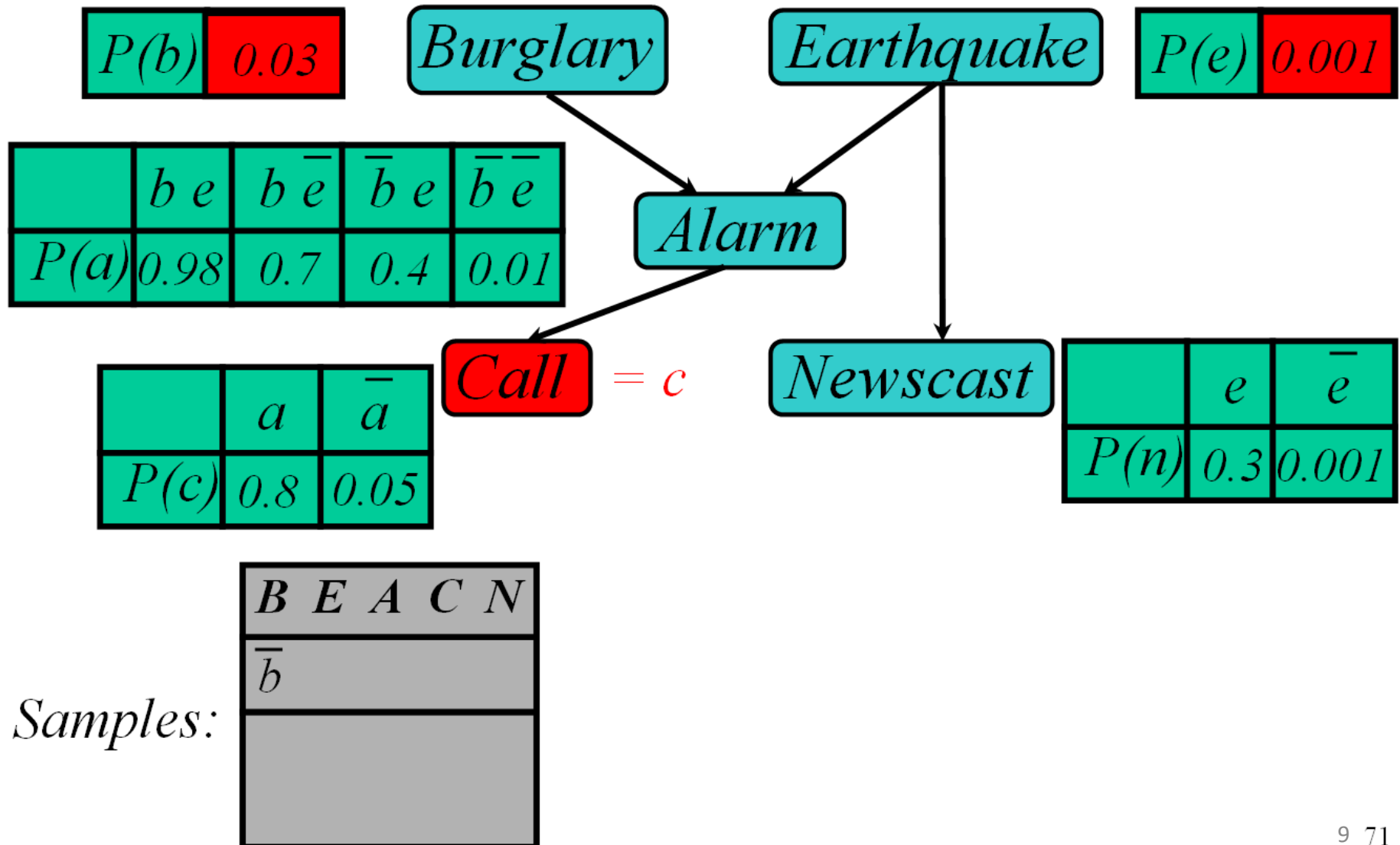


# Stochastic simulation $P(B|C)$

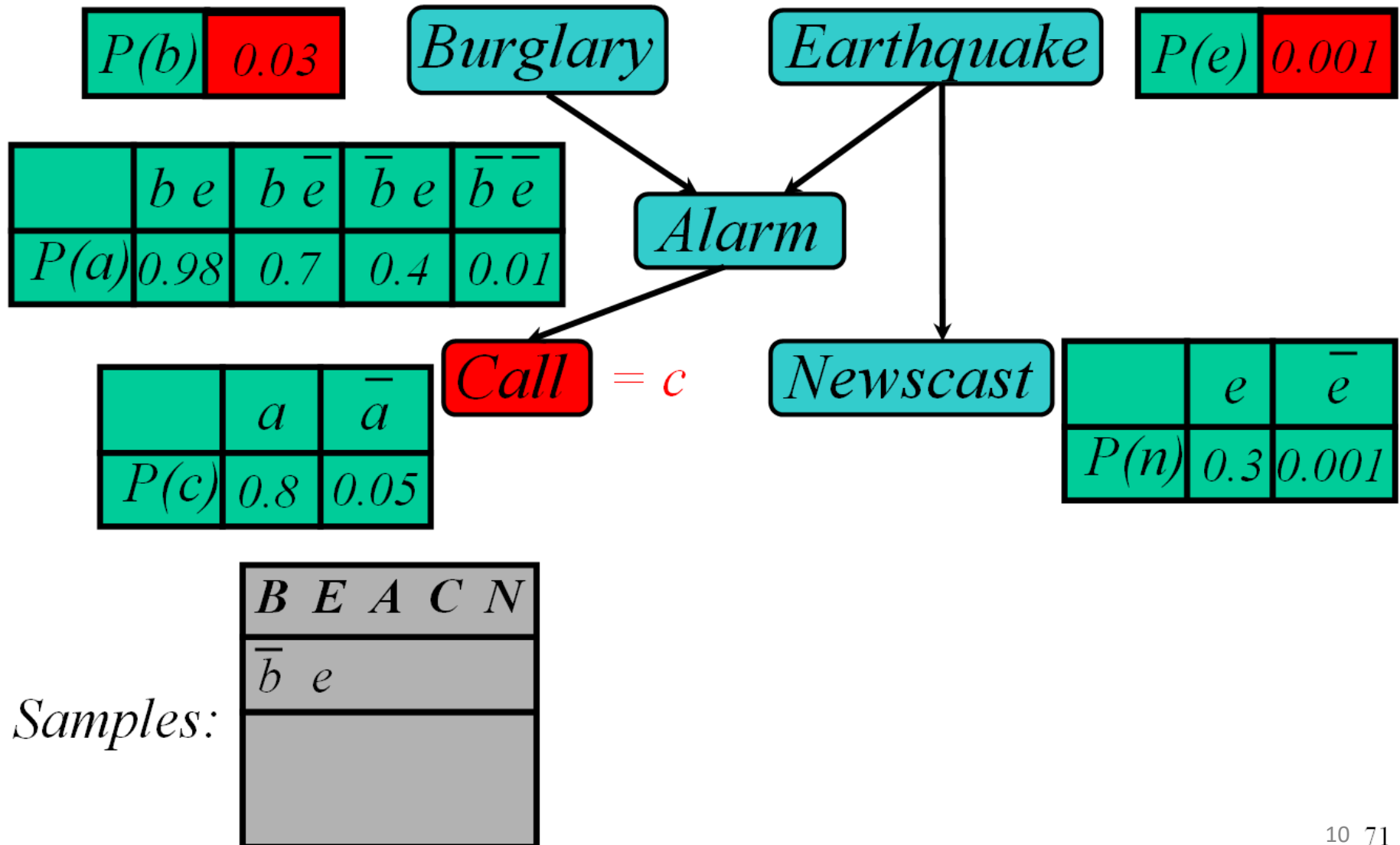




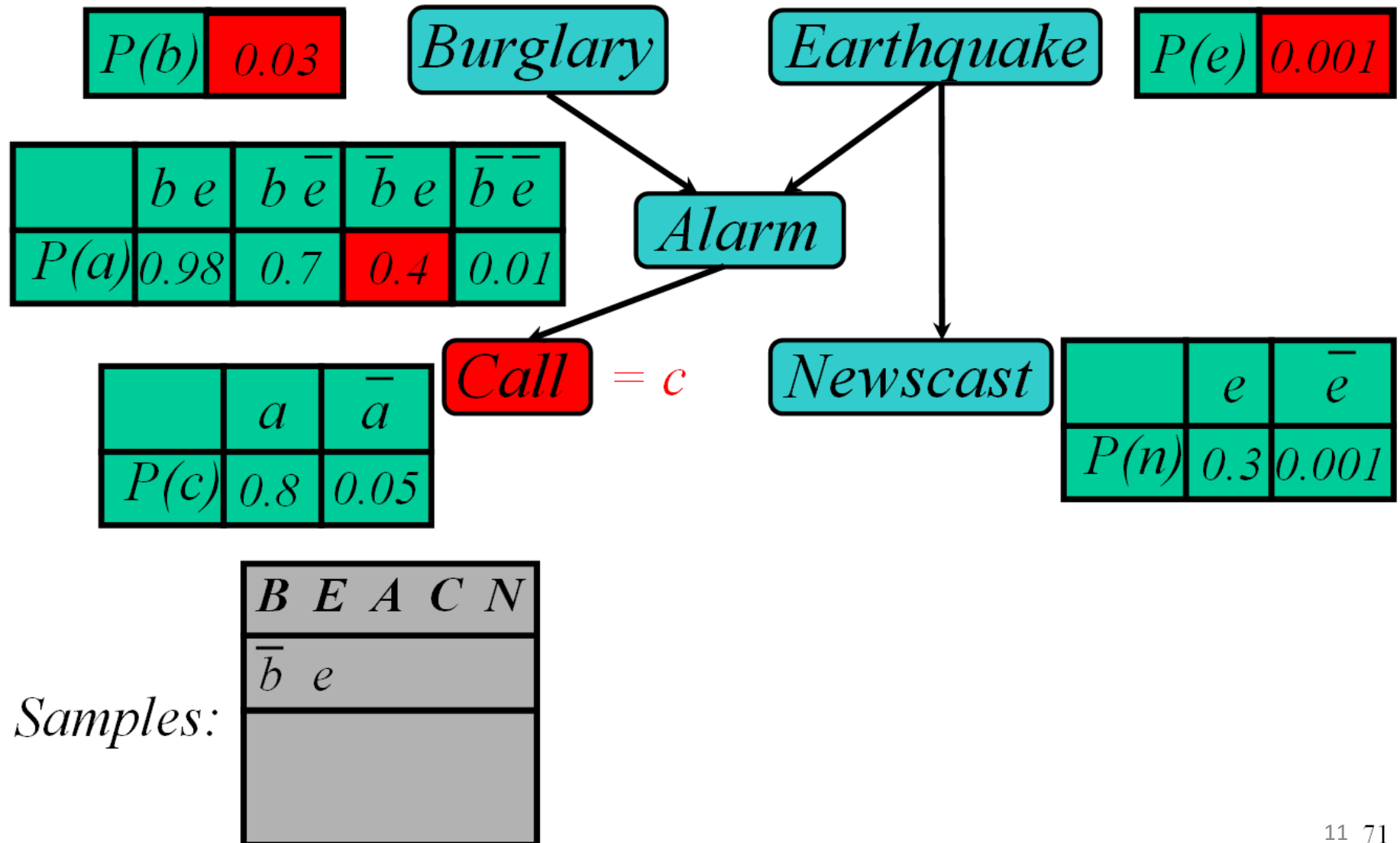
# Stochastic simulation $P(B|C)$



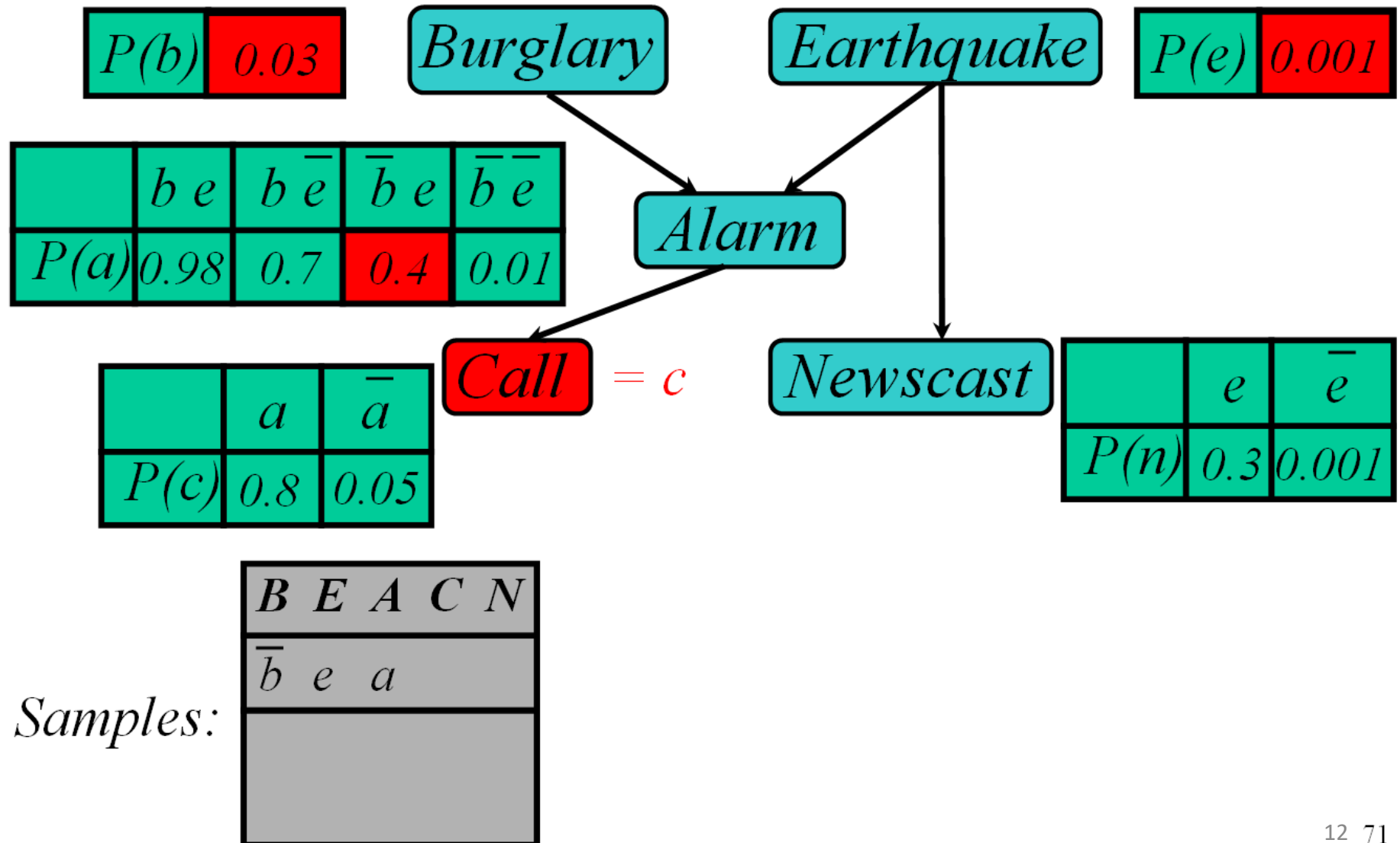
# Stochastic simulation $P(B|C)$



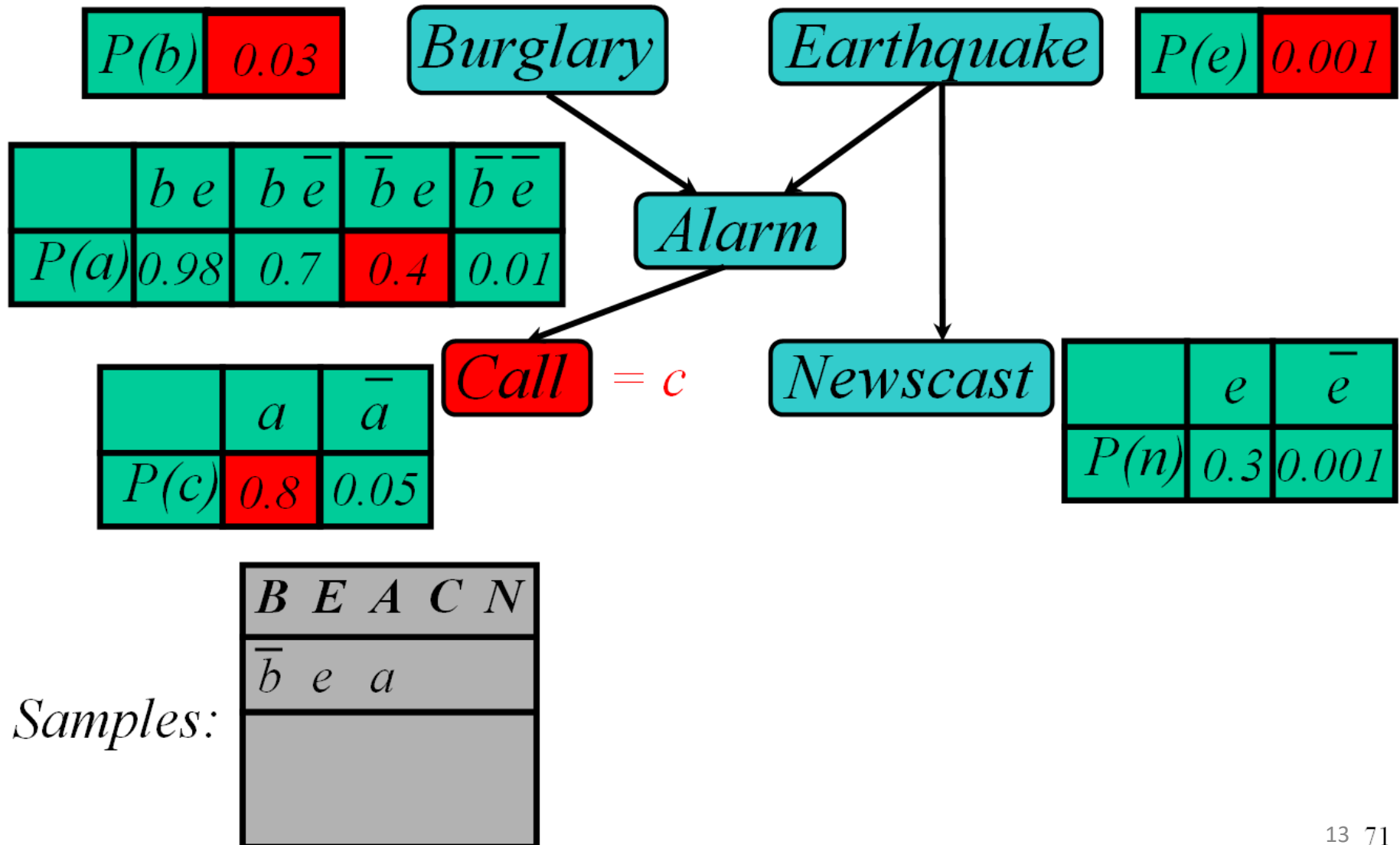
# Stochastic simulation $P(B|C)$



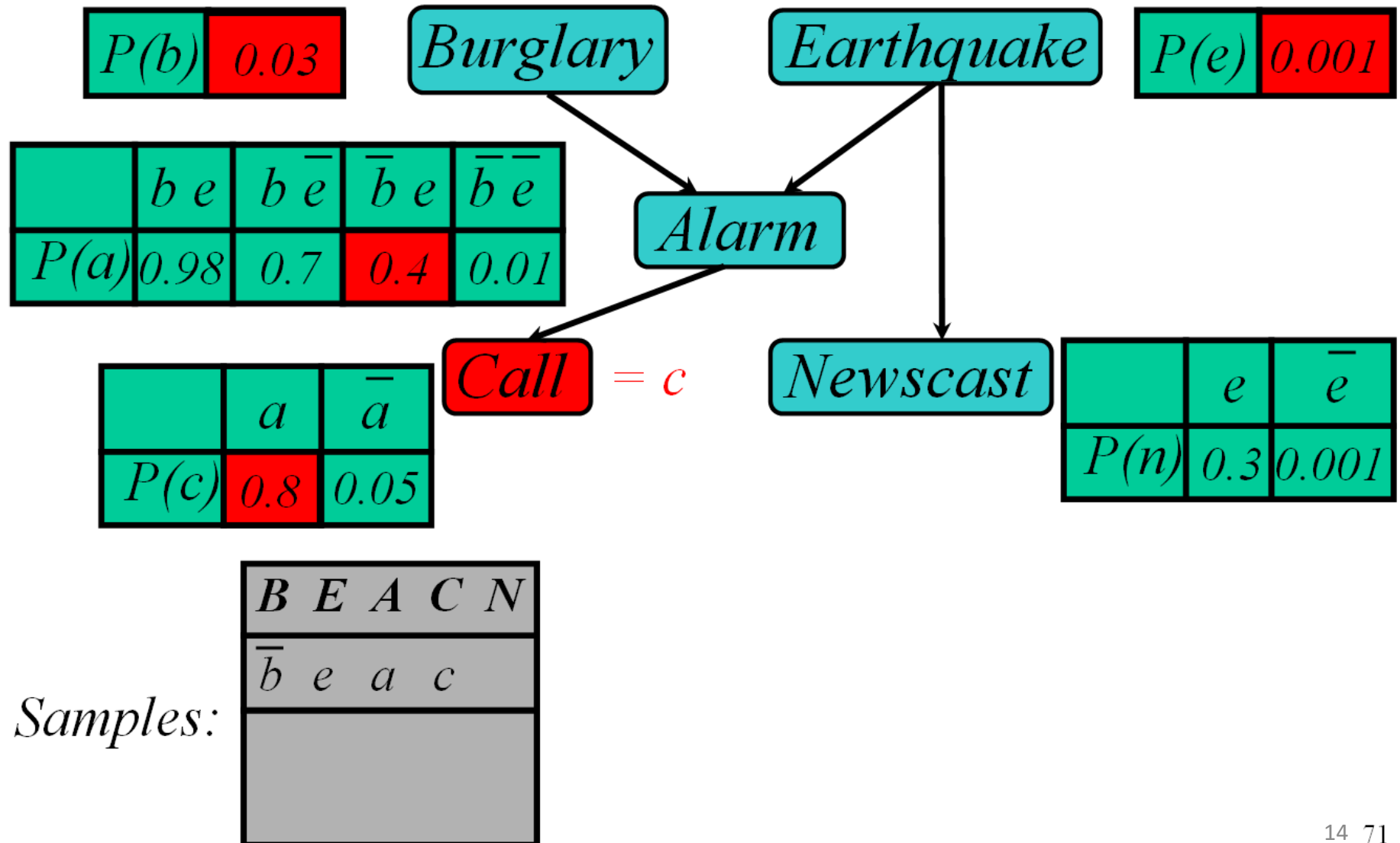
# Stochastic simulation $P(B|C)$



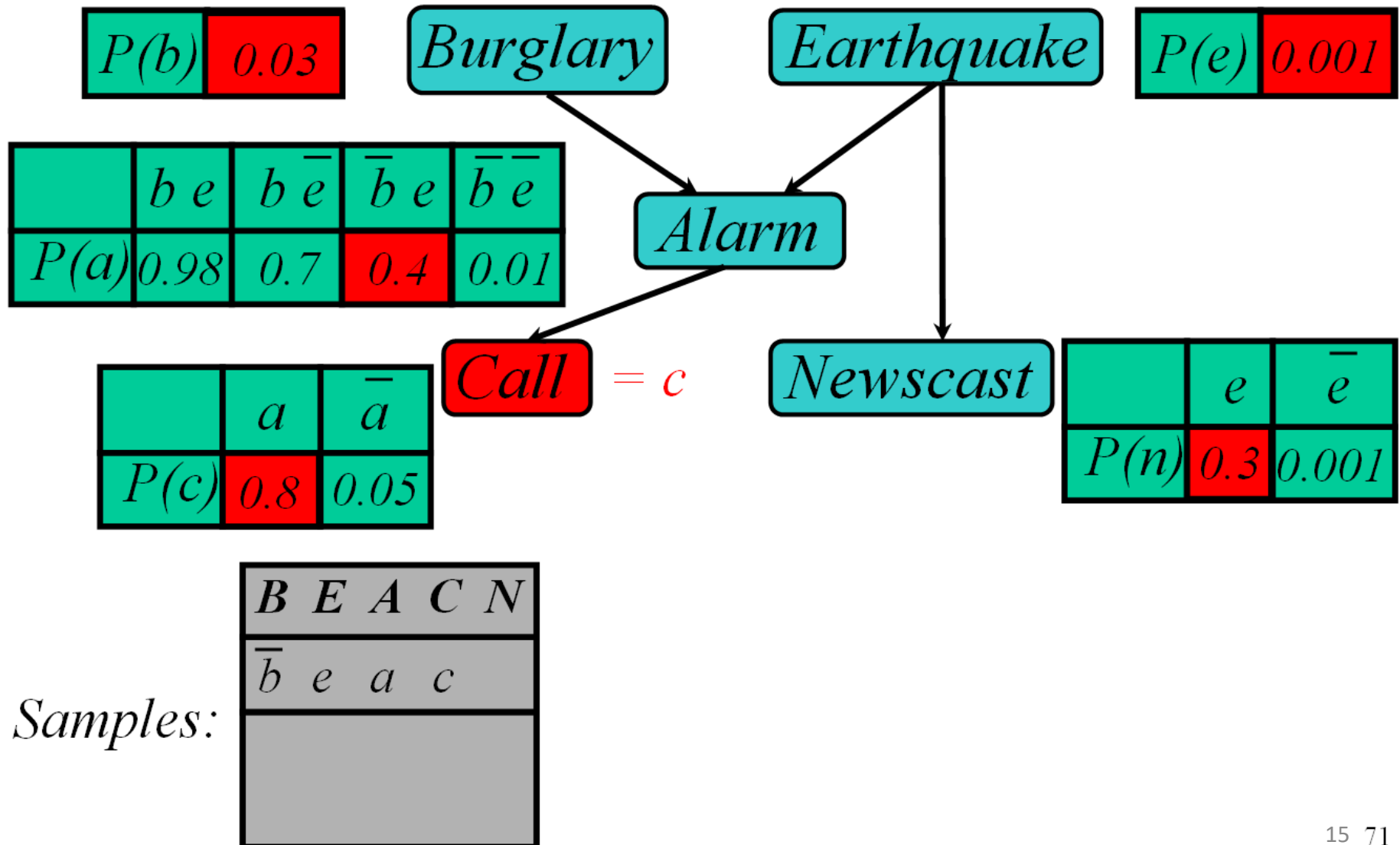
# Stochastic simulation $P(B|C)$



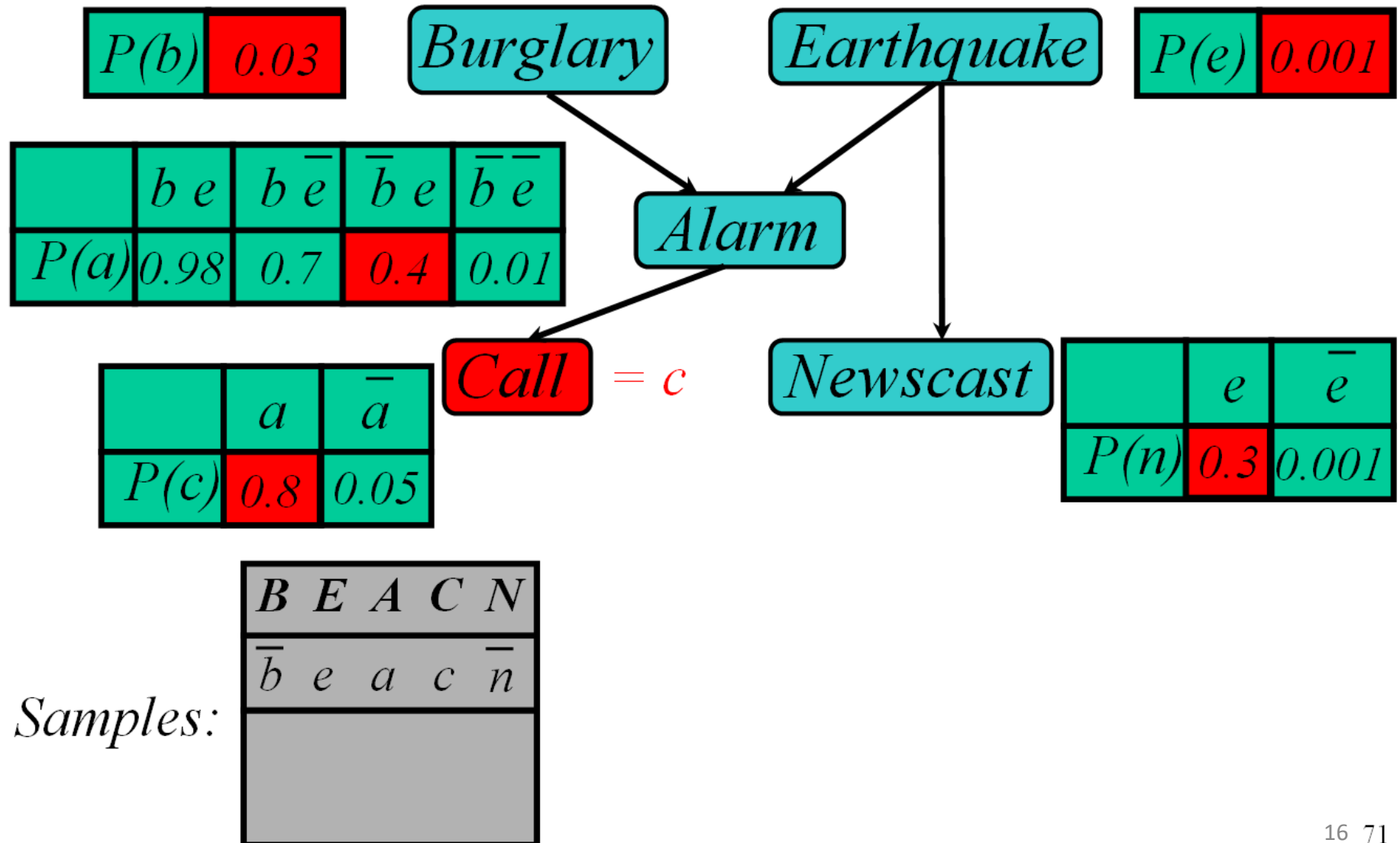
# Stochastic simulation $P(B|C)$



# Stochastic simulation $P(B|C)$

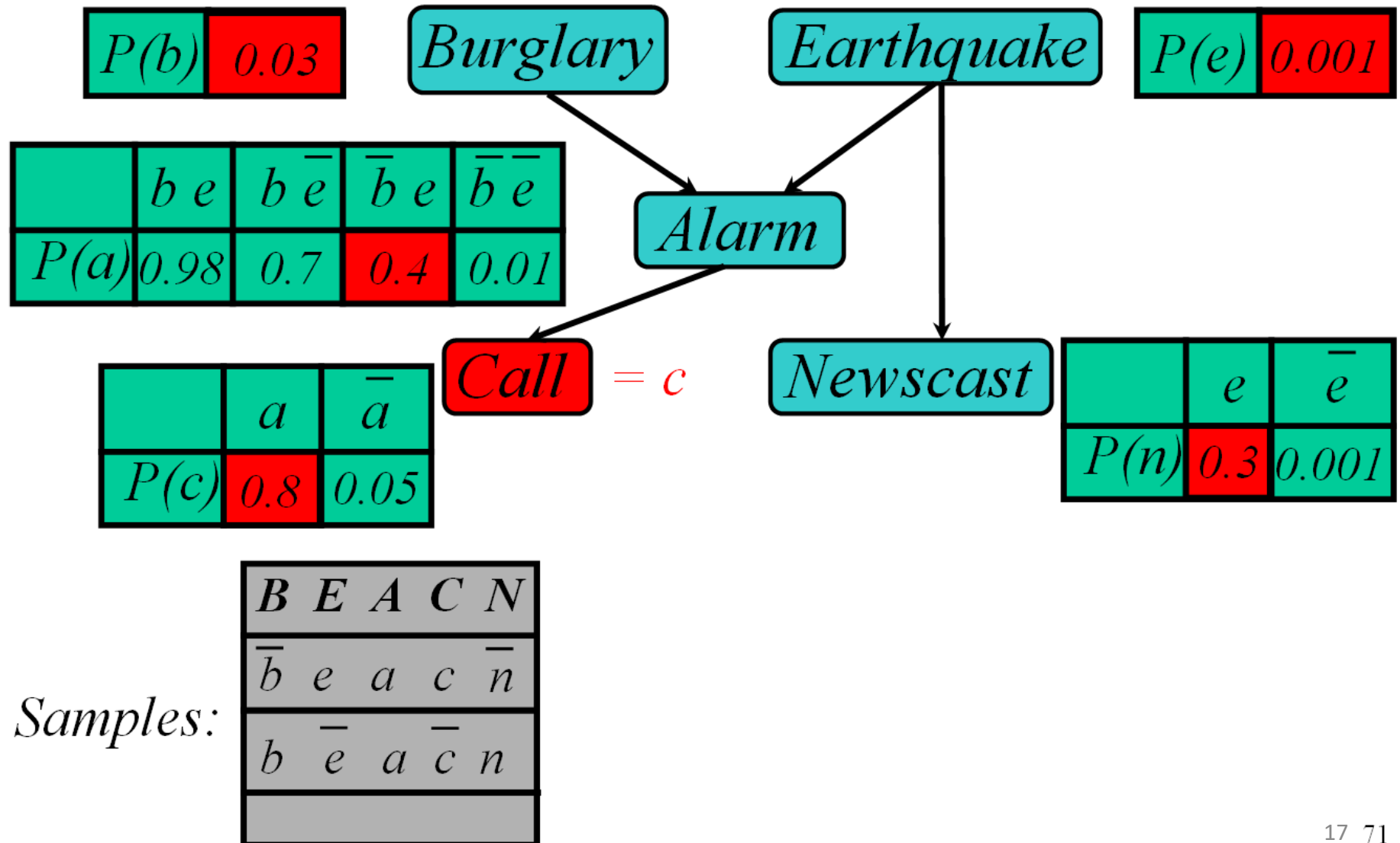


# Stochastic simulation $P(B|C)$

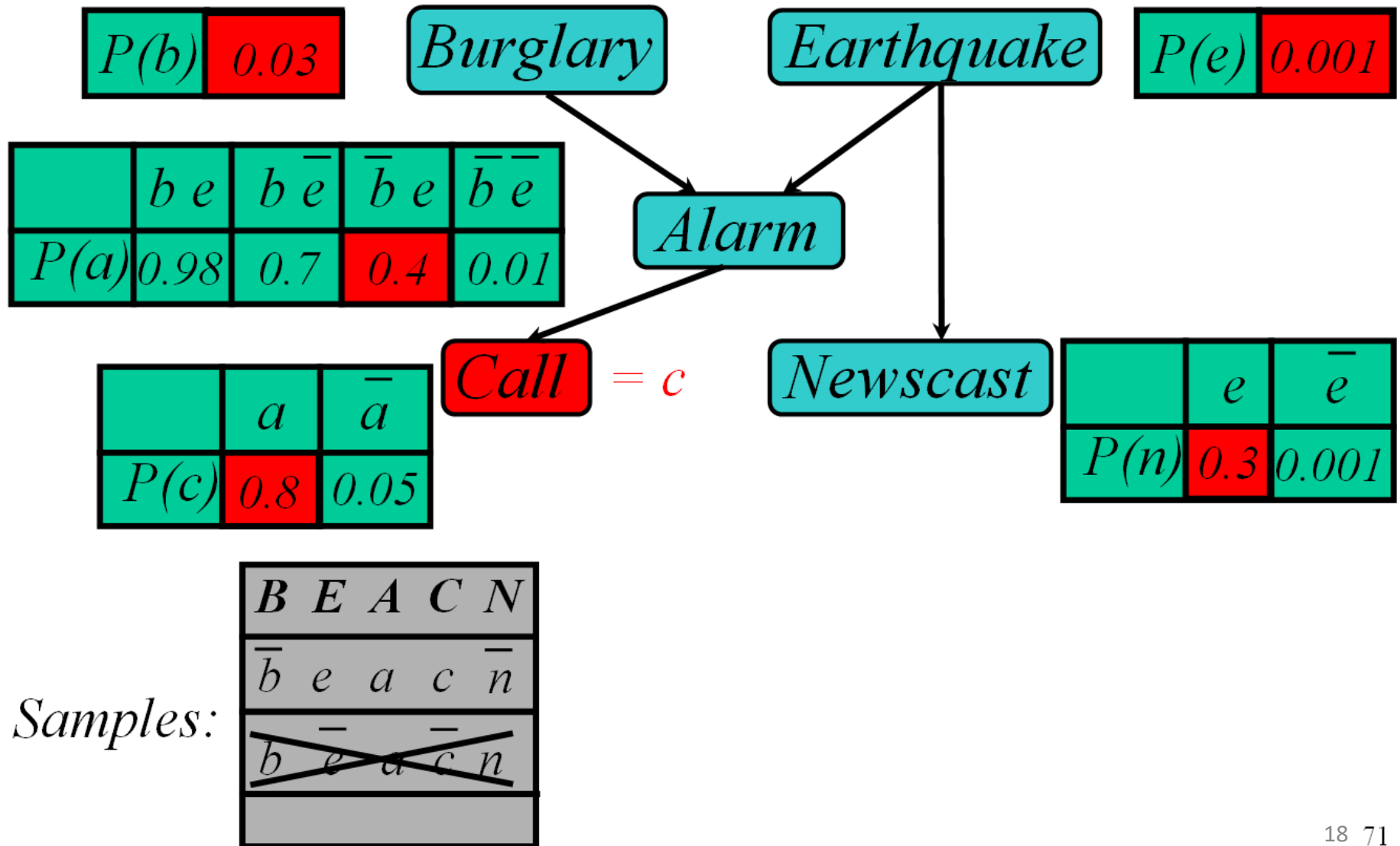




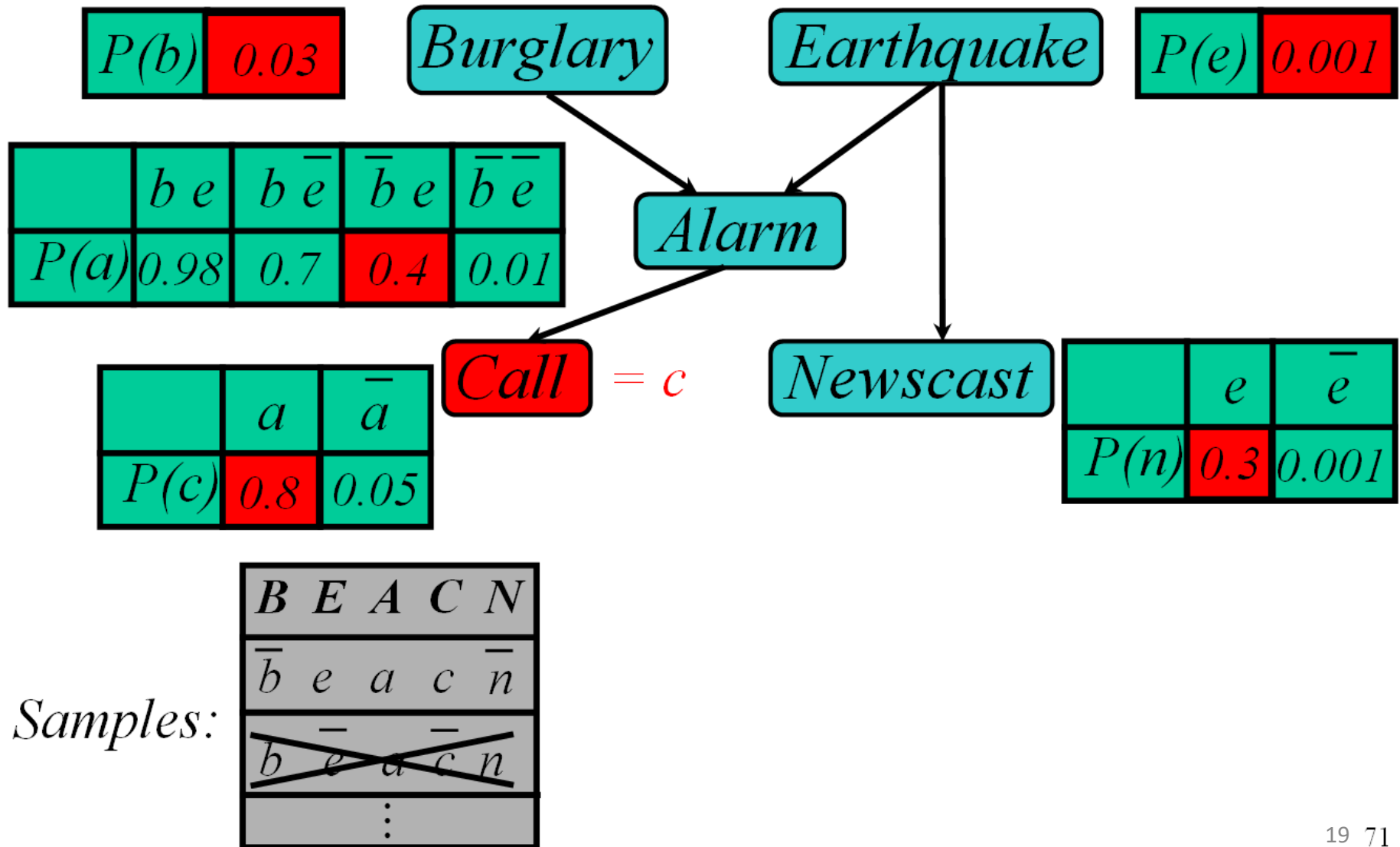
# Stochastic simulation $P(B|C)$



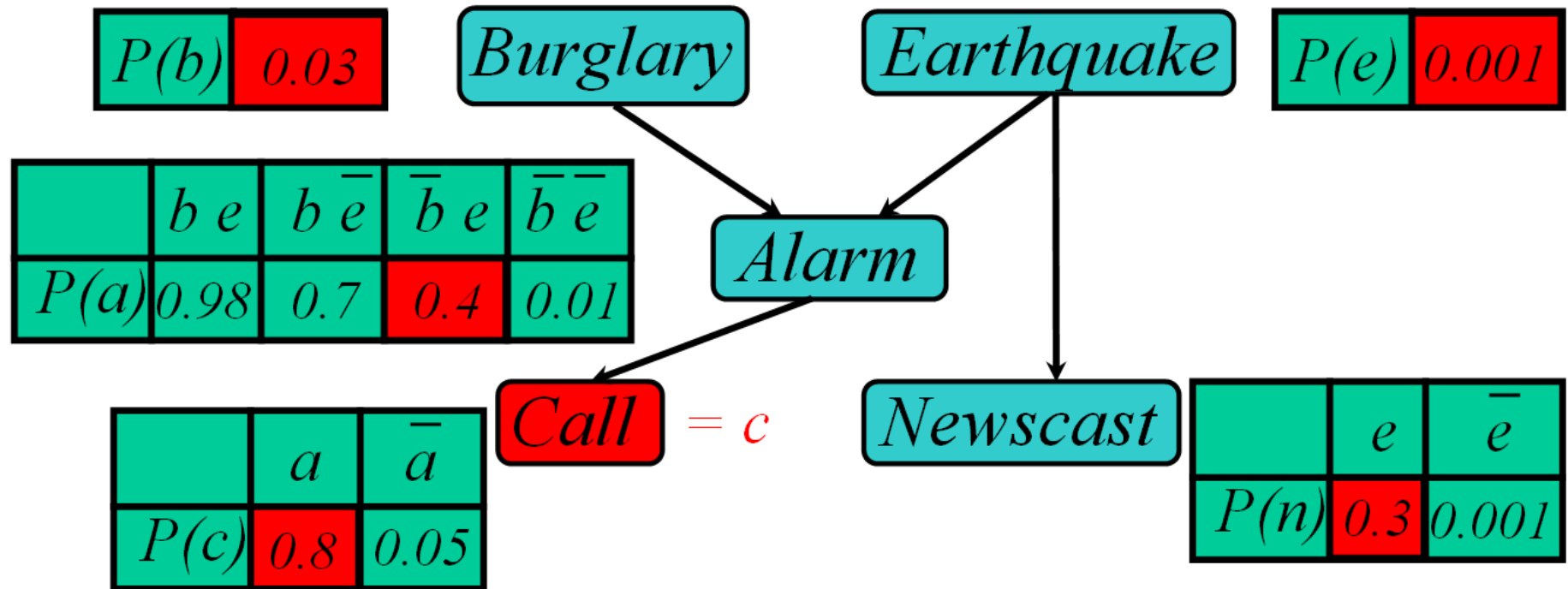
# Stochastic simulation $P(B|C)$



# Stochastic simulation $P(B|C)$



# Stochastic simulation $P(B|C)$



Samples:

$B$	$E$	$A$	$C$	$N$
$\bar{b}$	$e$	$a$	$c$	$\bar{n}$
<del><math>b</math></del>	<del><math>\bar{e}</math></del>	<del><math>\bar{a}</math></del>	<del><math>\bar{c}</math></del>	<del><math>n</math></del>
$\vdots$				

$$P(b|c) \sim \frac{\text{\# of live samples with } B=b}{\text{total \# of live samples}}$$

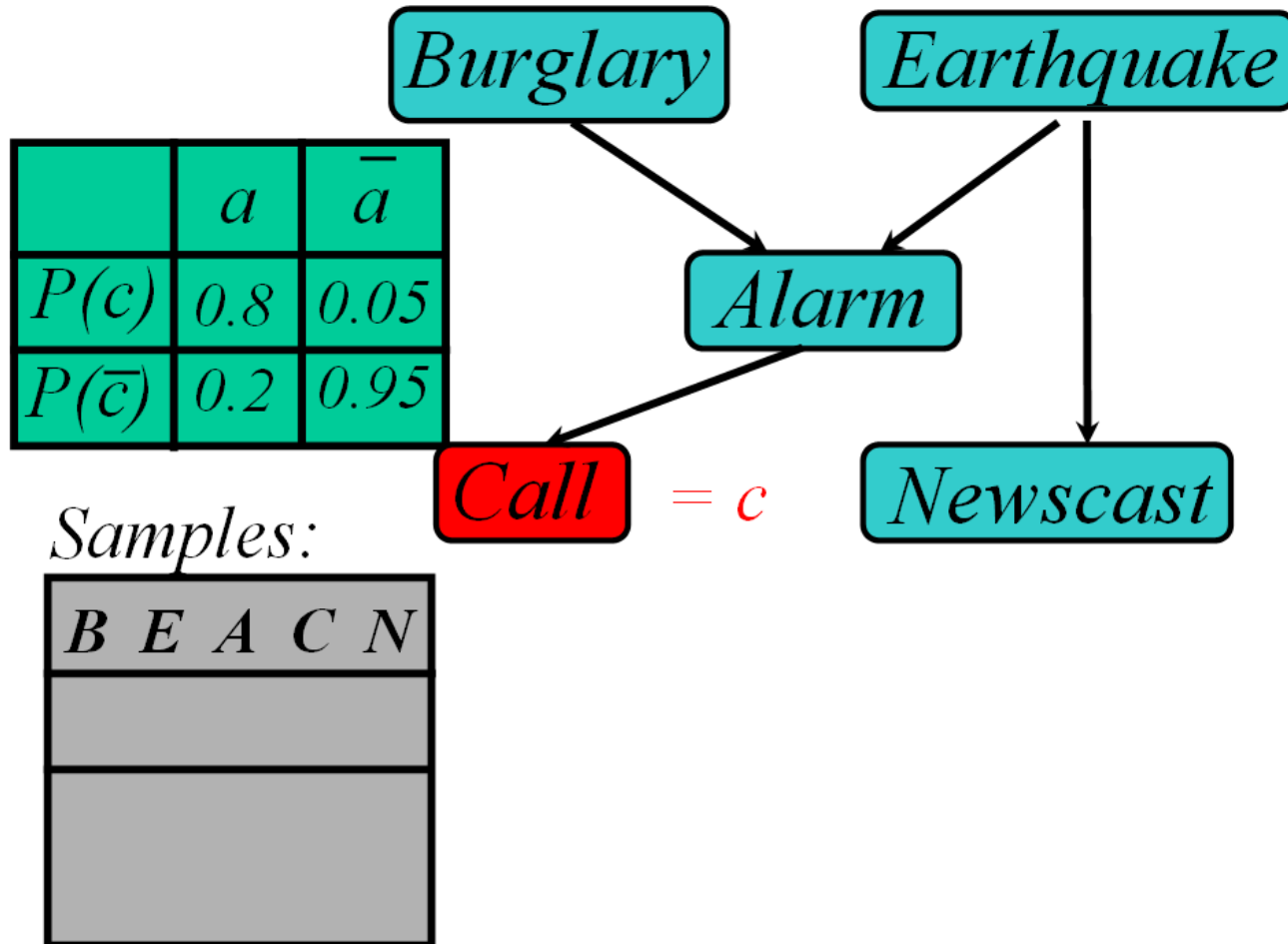
# Rejection Sampling

- Sample from the prior
  - reject if do not match the evidence
- Returns consistent posterior estimates
- Hopelessly expensive if  $P(e)$  is small
  - $P(e)$  drops off exponentially with no. of evidence vars

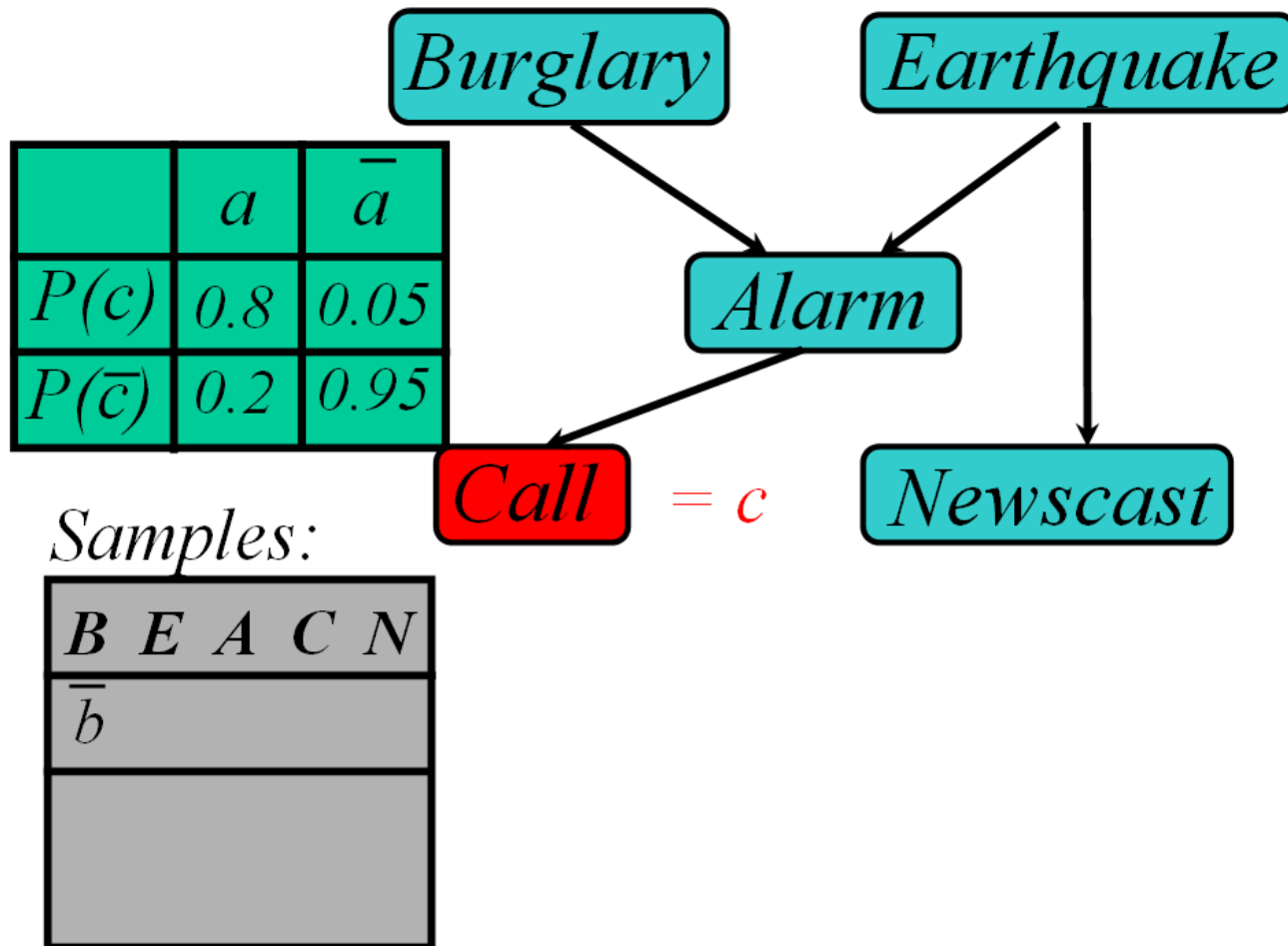
# Likelihood Weighting

- Idea
  - each sample agrees with evidence
  - pays some price for the agreement (weight)
- Algorithm
  - fix evidence variables
  - sample only non-evidence variables
  - weight each sample by the likelihood of evidence

# Likelihood weighting $P(B|C)$

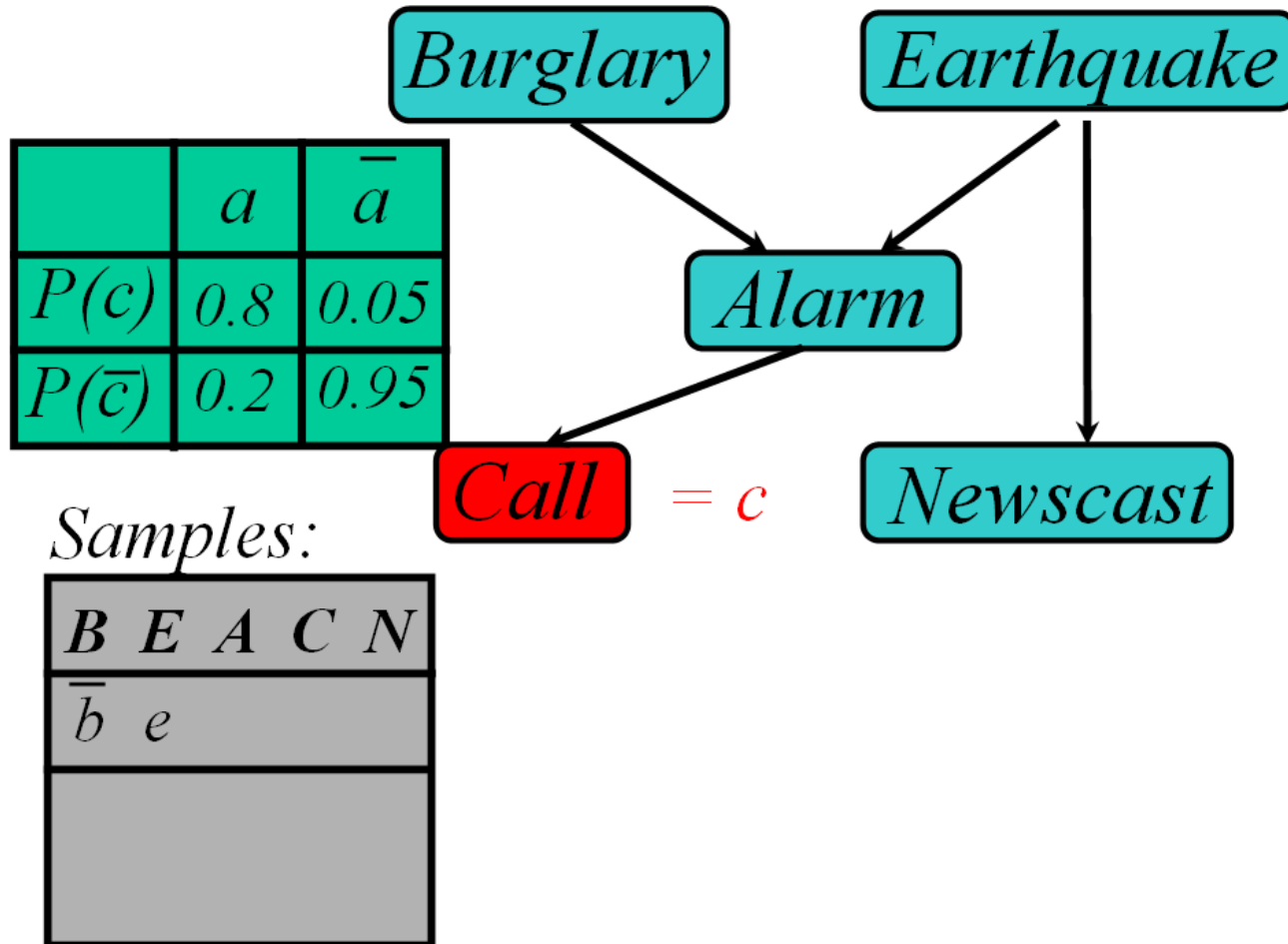


# Likelihood weighting $P(B|C)$

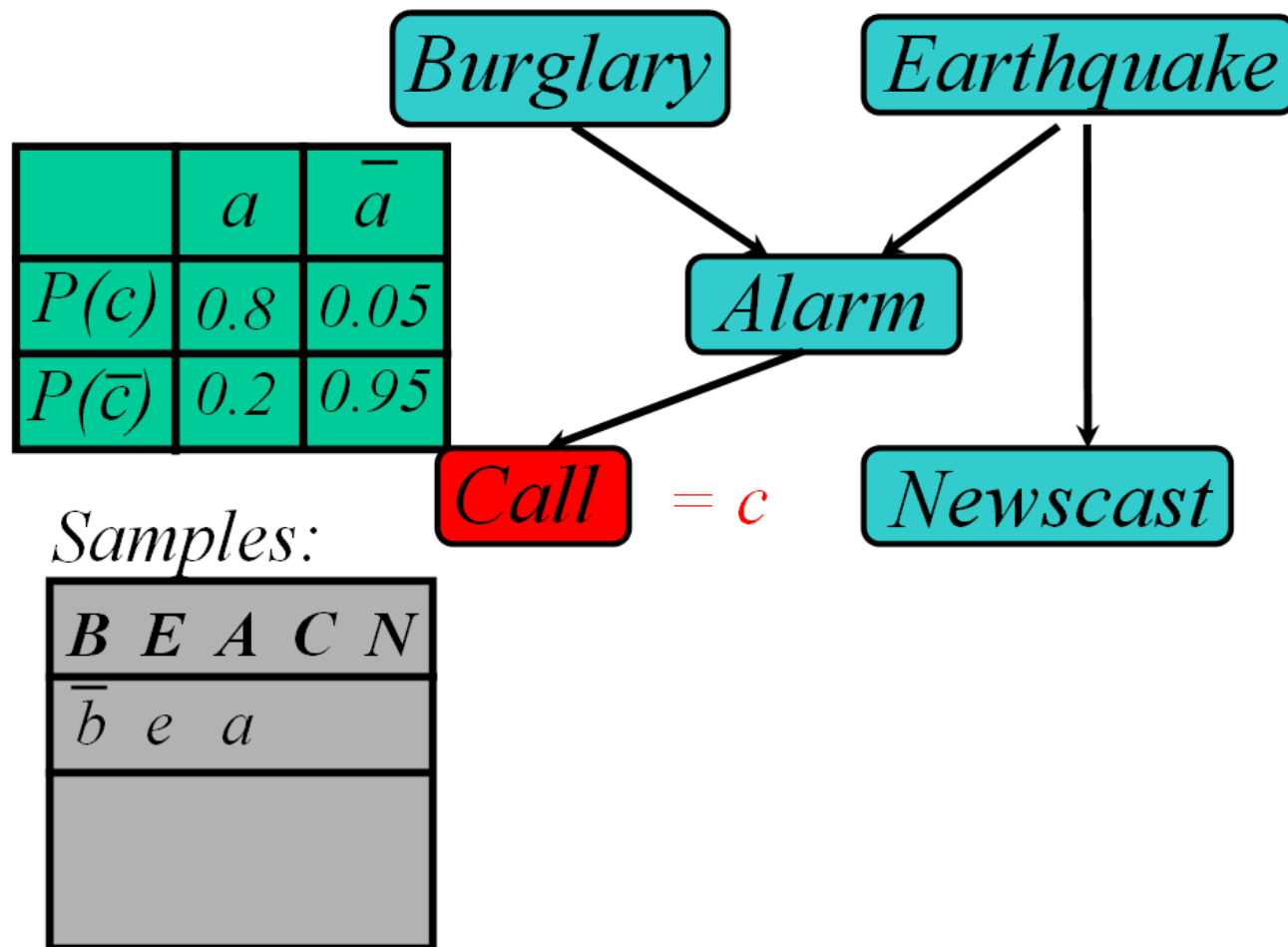




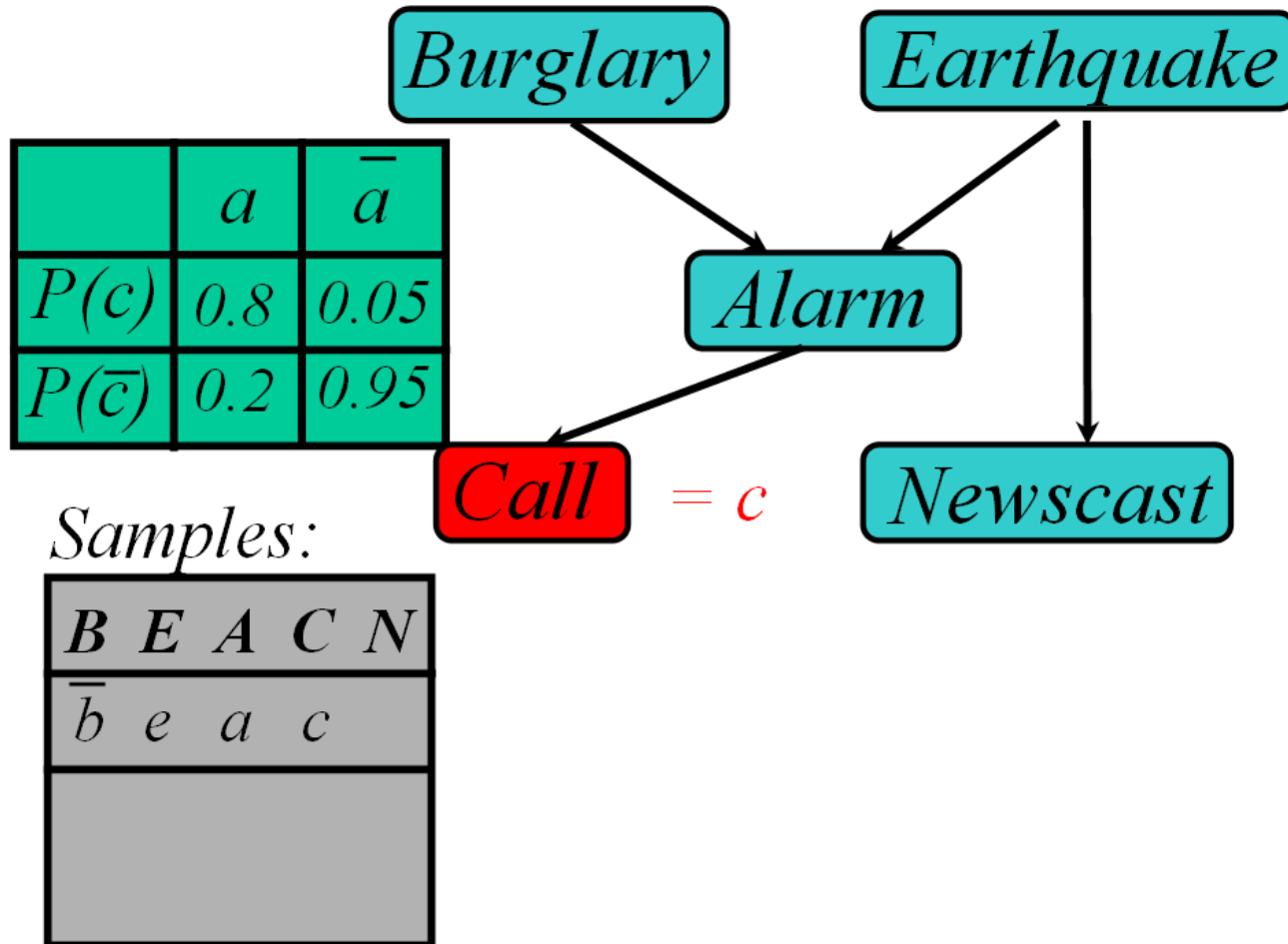
# Likelihood weighting $P(B|C)$



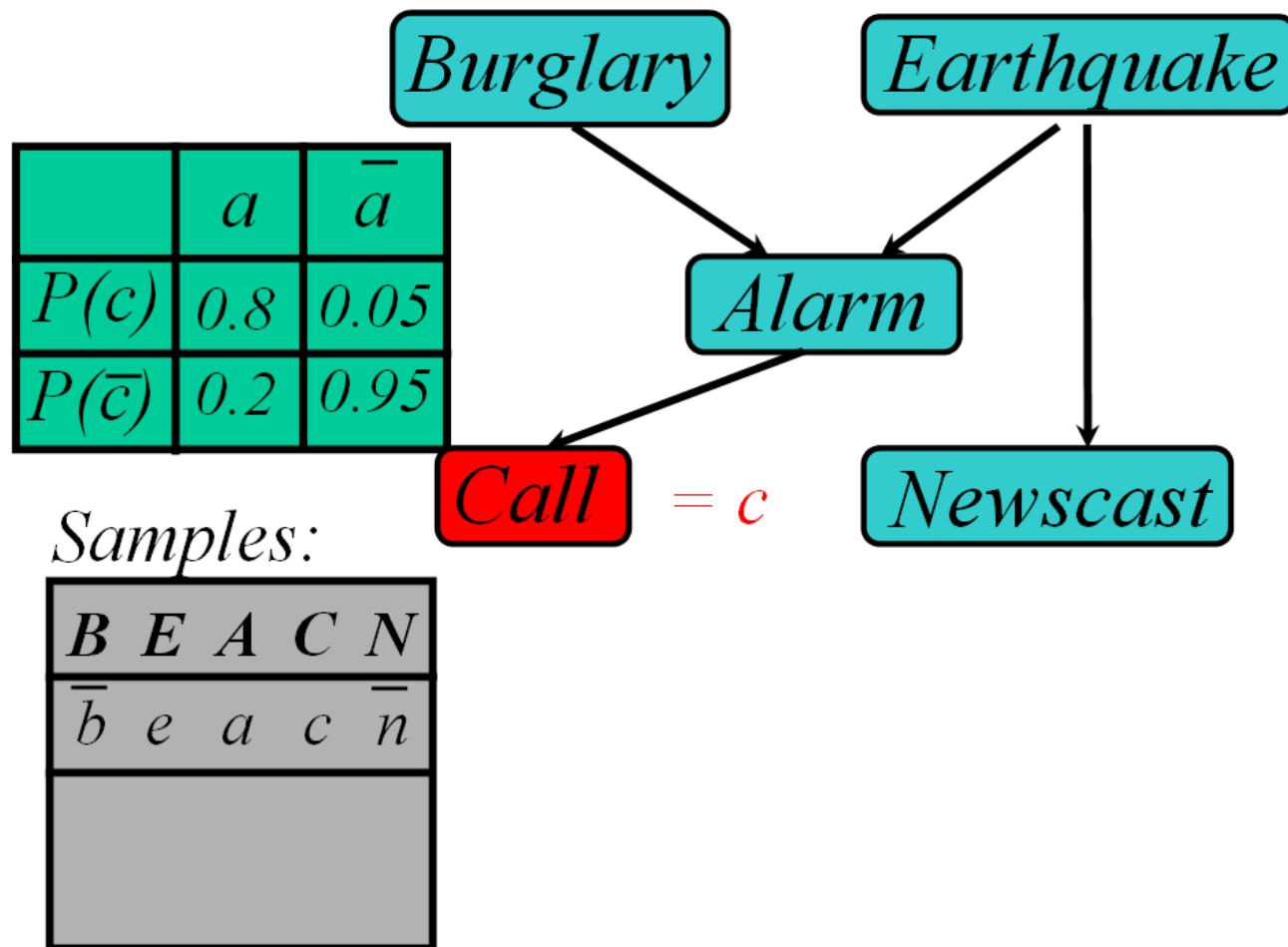
# Likelihood weighting $P(B|C)$



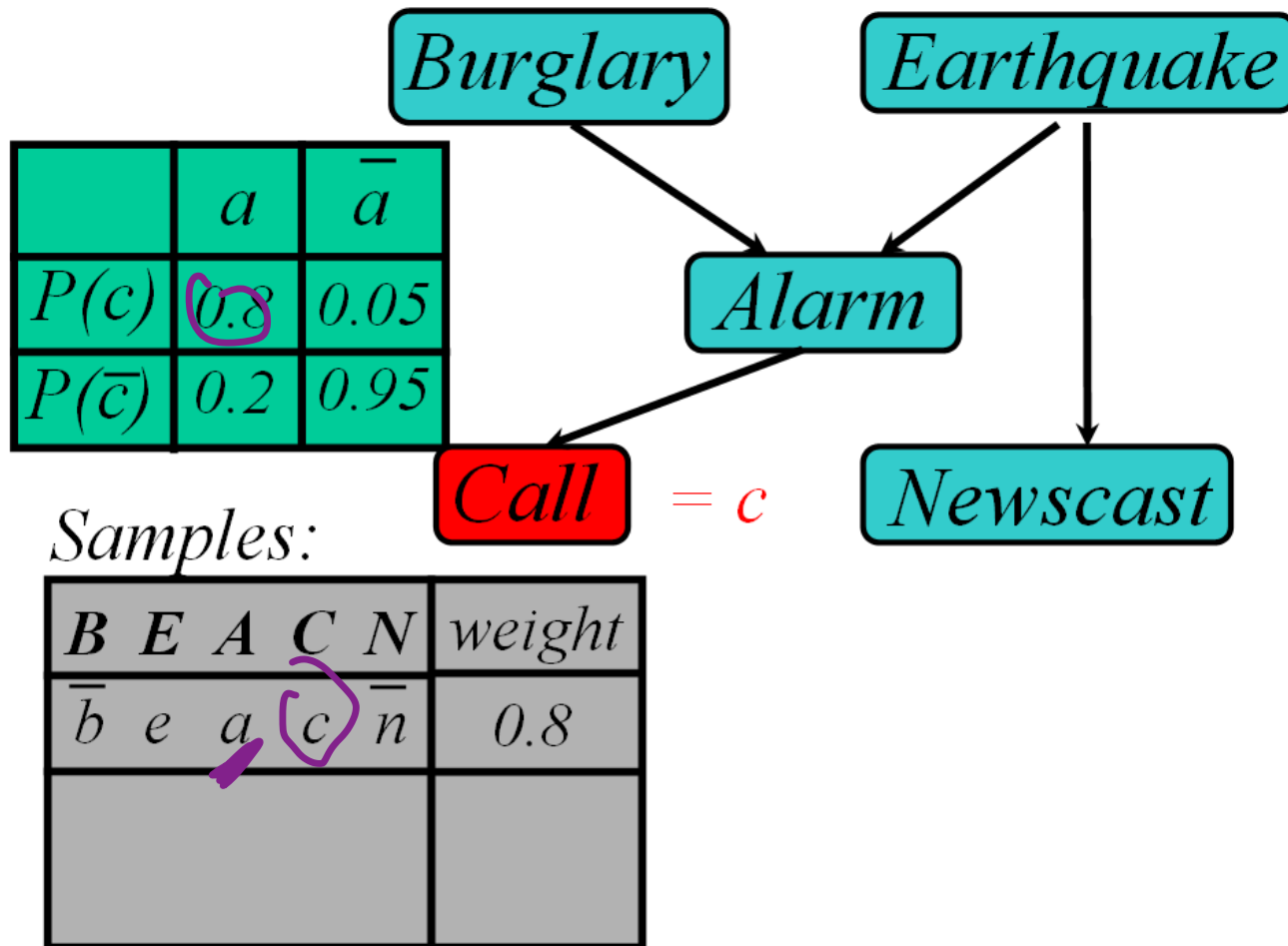
# Likelihood weighting $P(B|C)$



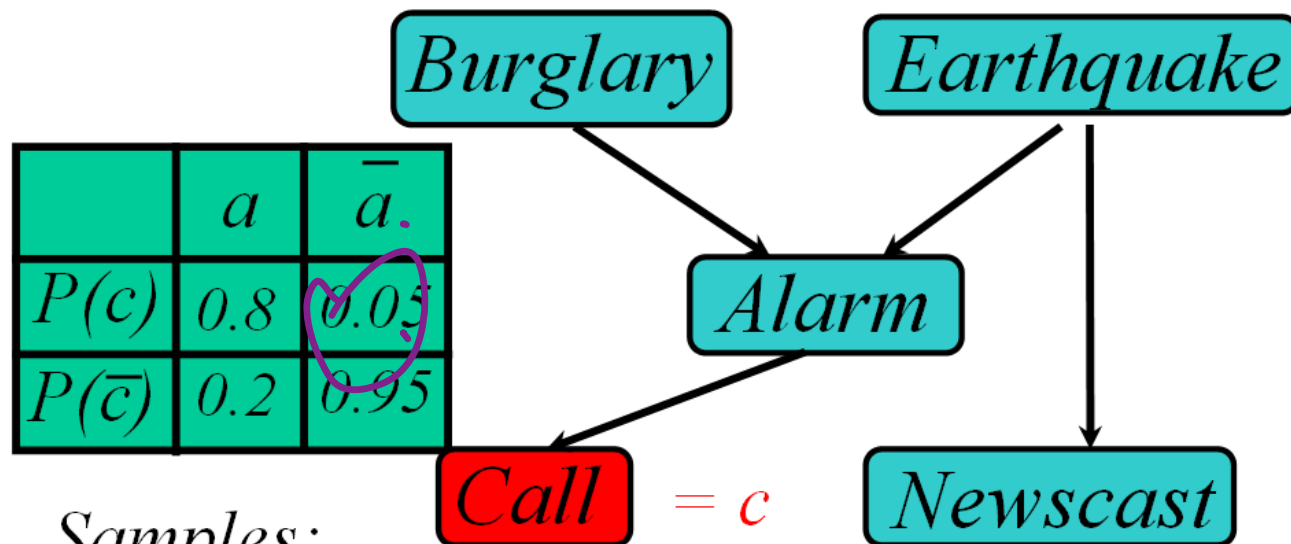
# Likelihood weighting $P(B|C)$



# Likelihood weighting $P(B|C)$



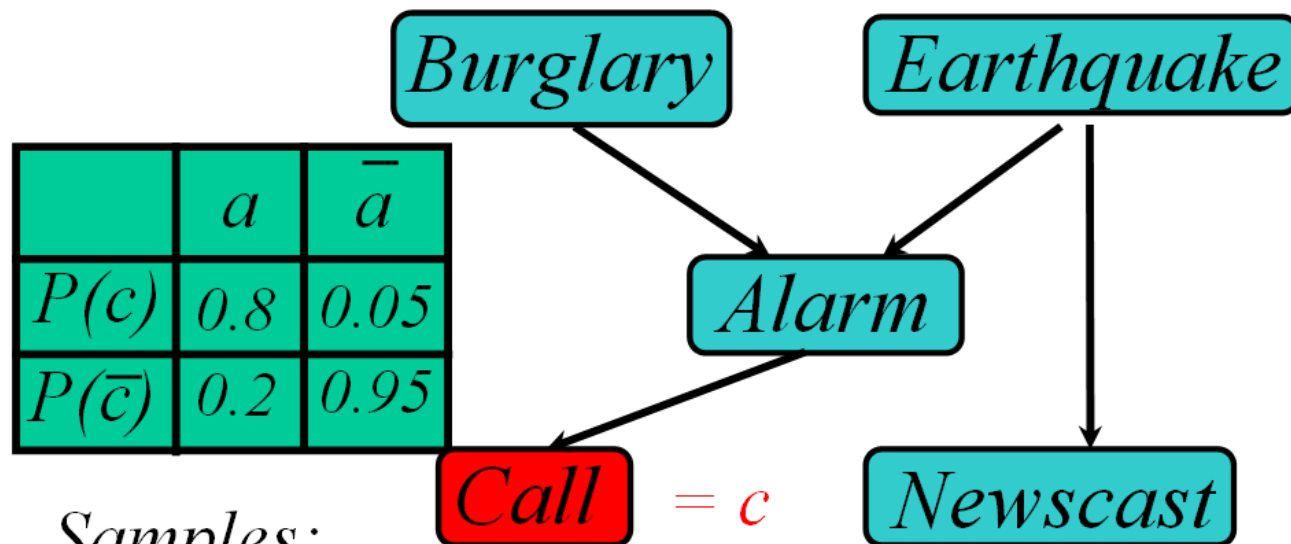
# Likelihood weighting $P(B|C)$



Samples:

$B$	$E$	$A$	$C$	$N$	weight
$\bar{b}$	$e$	$a$	$c$	$\bar{n}$	0.8
$b$	$\bar{e}$	$\bar{a}$	$c$	$n$	0.05

# Likelihood weighting $P(B|C)$

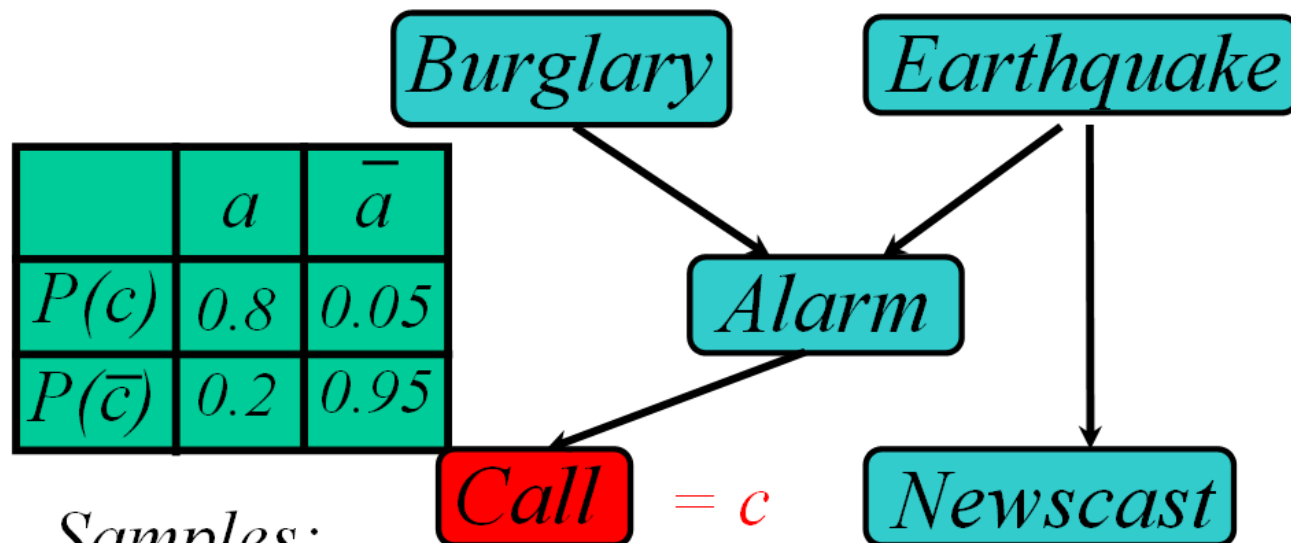


	$a$	$\bar{a}$
$P(c)$	0.8	0.05
$P(\bar{c})$	0.2	0.95

Samples:

$B$	$E$	$A$	$C$	$N$	weight
$\bar{b}$	$e$	$a$	$c$	$\bar{n}$	0.8
$b$	$\bar{e}$	$\bar{a}$	$c$	$n$	0.05
$\vdots$					

# Likelihood weighting $P(B|C)$



Samples:

$B$	$E$	$A$	$C$	$N$	weight
$\bar{b}$	$e$	$a$	$c$	$\bar{n}$	0.8
$b$	$\bar{e}$	$\bar{a}$	$c$	$n$	0.05
$\vdots$					

$$P(b|c) = \frac{\text{weight of samples with } B=b}{\text{total weight of samples}}$$



# Likelihood Weighting

- **Sampling probability:**  $S(z, e) = \prod_i P(z_i \mid \text{Parents}(Z_i))$ 
  - Neither prior nor posterior
- **Wt for a sample  $\langle z, e \rangle$ :**  $w(z, e) = \prod_i P(e_i \mid \text{Parents}(E_i))$
- **Weighted Sampling probability**  $S(z, e)w(z, e)$ 
$$= \prod_i P(z_i \mid \text{Parents}(Z_i)) \prod_i P(e_i \mid \text{Parents}(E_i))$$
$$= P^i(z, e)$$
- ➔ returns consistent estimates
- performance degrades w/ many evidence vars
  - but a few samples have nearly all the total weight
  - late occurring evidence vars do not guide sample generation

# MCMC with Gibbs Sampling

- Fix the values of observed variables
- Set the values of all non-observed variables randomly
- Perform a random walk through the space of complete variable assignments. On each move:
  1. Pick a variable  $X$
  2. Calculate  $\Pr(X=\text{true} \mid \text{all other variables})$
  3. Set  $X$  to true with that probability
- Repeat many times. Frequency with which any variable  $X$  is true is its posterior probability.
- Converges to true posterior when frequencies stop changing significantly
  - stationary distribution, mixing

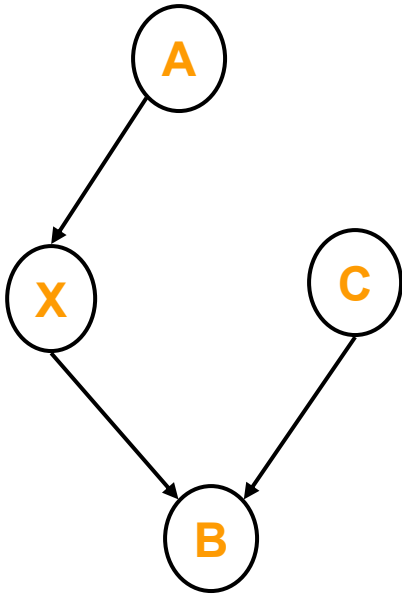
# Markov Blanket Sampling

- How to calculate  $\Pr(X=\text{true} \mid \text{all other variables})$  ?
- Recall: a variable is independent of all others given it's Markov Blanket
  - parents
  - children
  - other parents of children
- So problem becomes calculating  $\Pr(X=\text{true} \mid \text{MB}(X))$ 
  - We solve this sub-problem exactly
  - Fortunately, it is easy to solve

$$P(X) = \alpha P(X \mid \text{Parents}(X)) \prod_{Y \in \text{Children}(X)} P(Y \mid \text{Parents}(Y))$$

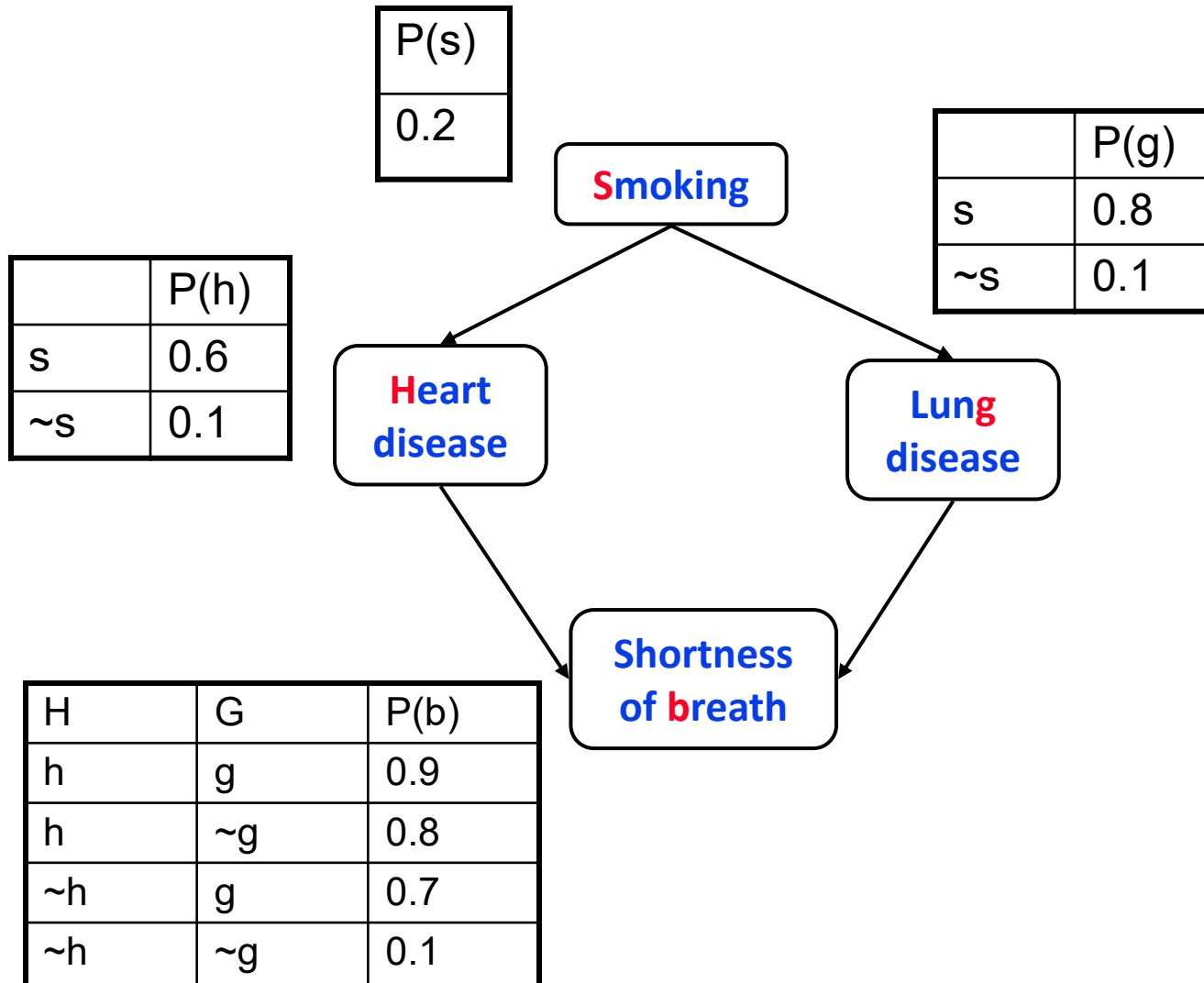
# Example

$$P(X) = \alpha P(X \mid \text{Parents}(X)) \prod_{Y \in \text{Children}(X)} P(Y \mid \text{Parents}(Y))$$

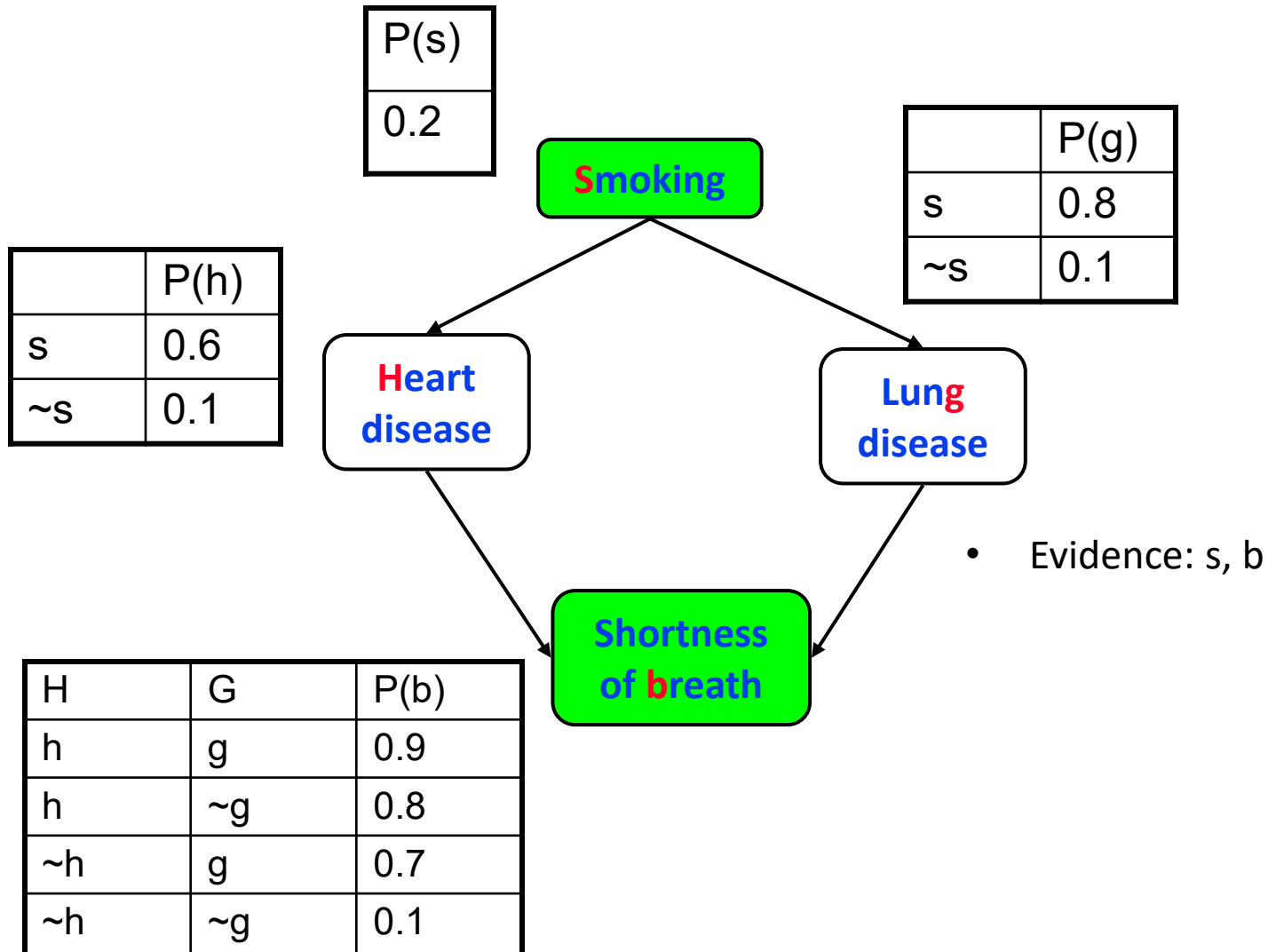


$$\begin{aligned} P(X \mid A, B, C) &= \frac{P(X, A, B, C)}{P(A, B, C)} \\ &= \frac{P(A)P(X \mid A)P(C)P(B \mid X, C)}{P(A, B, C)} \\ &= \left[ \frac{P(A)P(C)}{P(A, B, C)} \right] P(X \mid A)P(B \mid X, C) \\ &= \alpha P(X \mid A)P(B \mid X, C) \end{aligned}$$

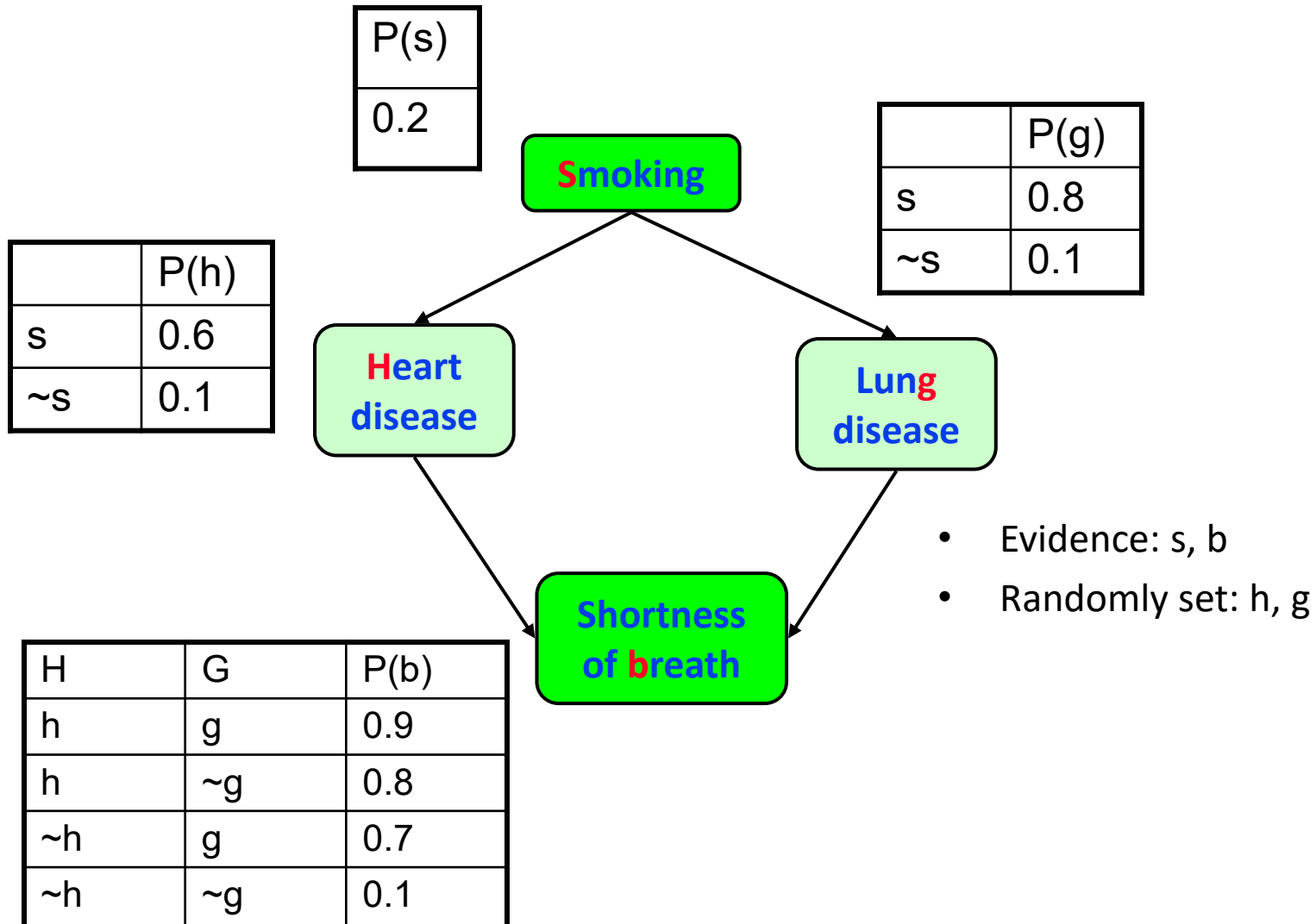
# Example



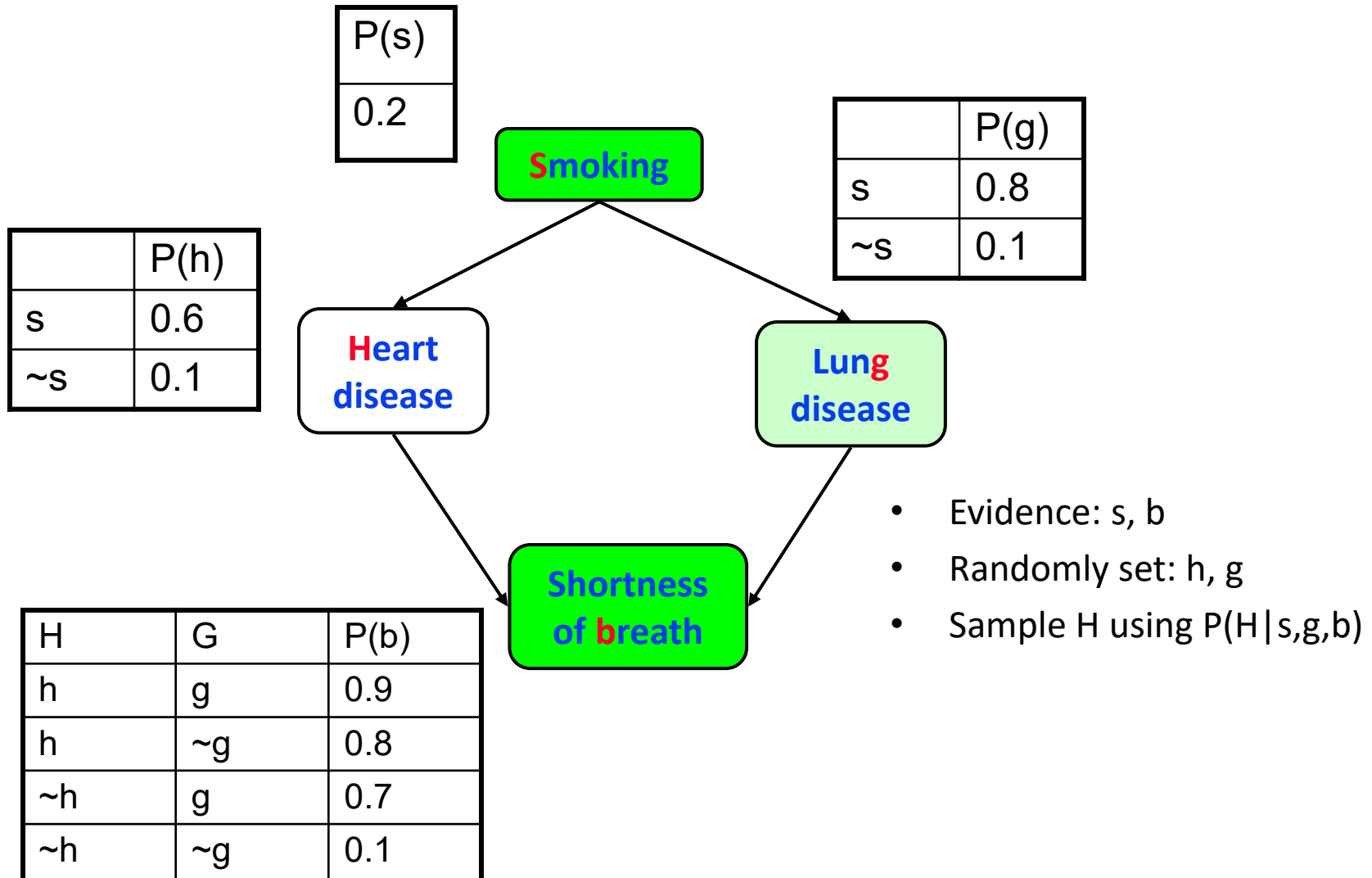
# Example



# Example

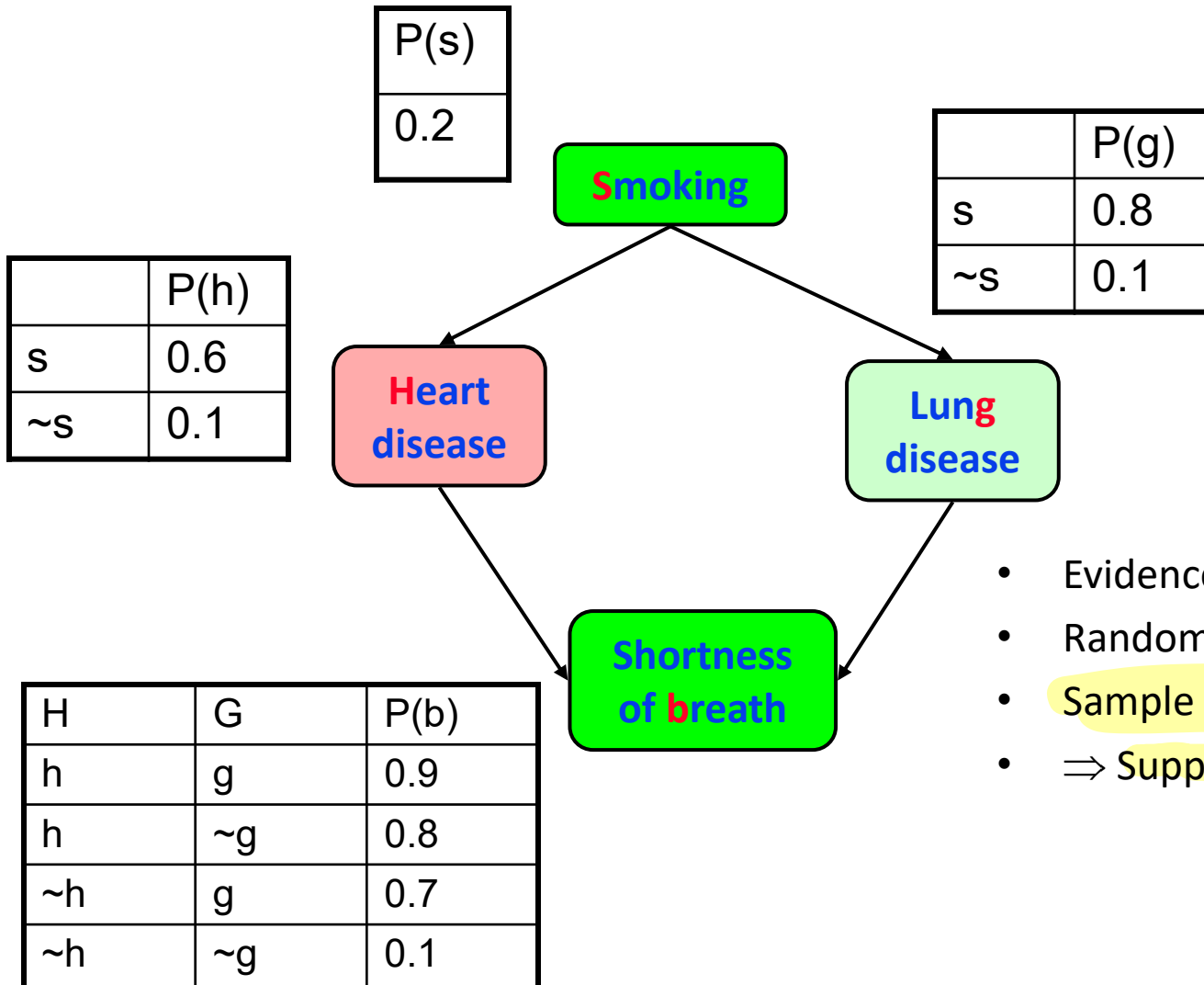


# Example



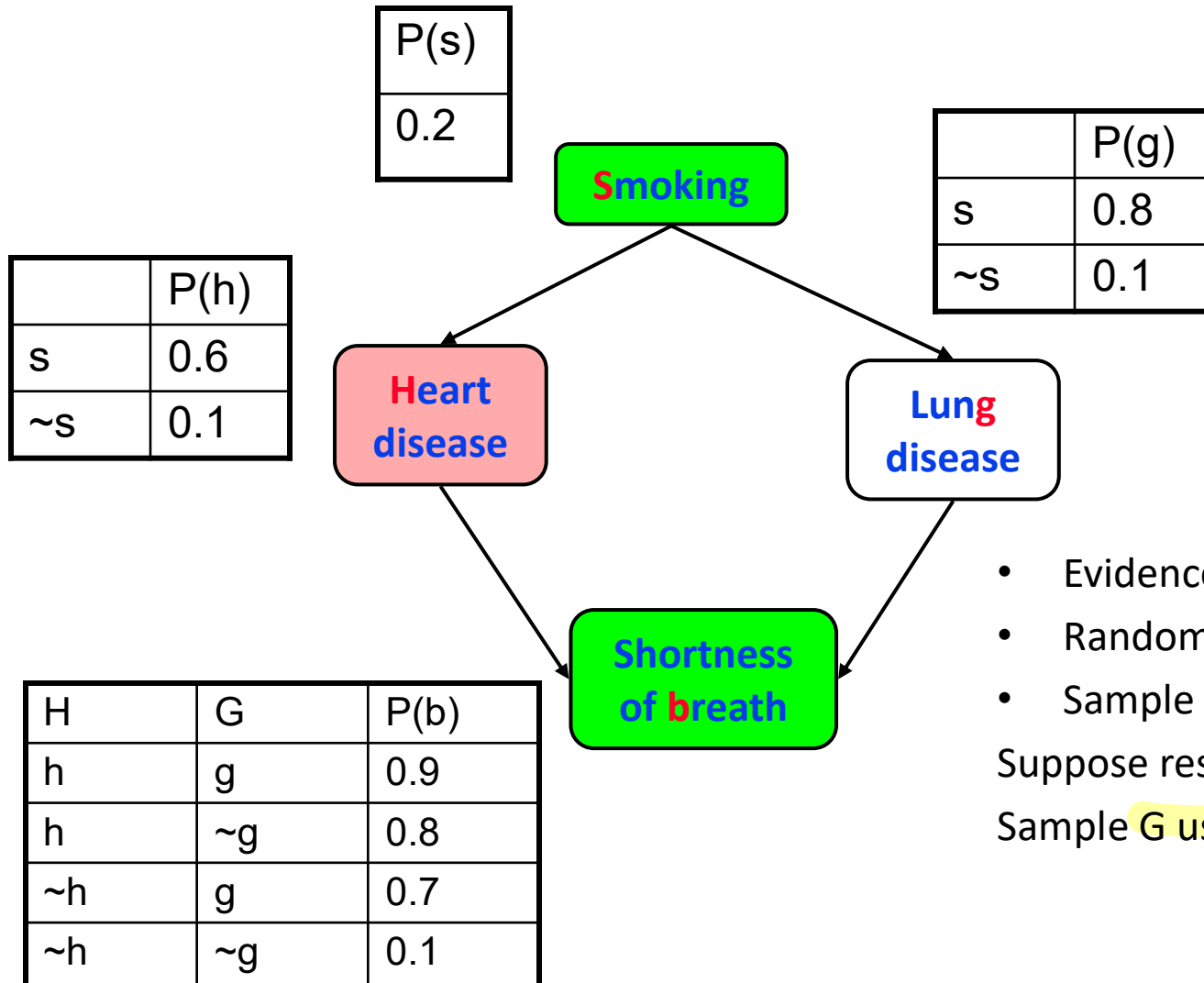


# Example



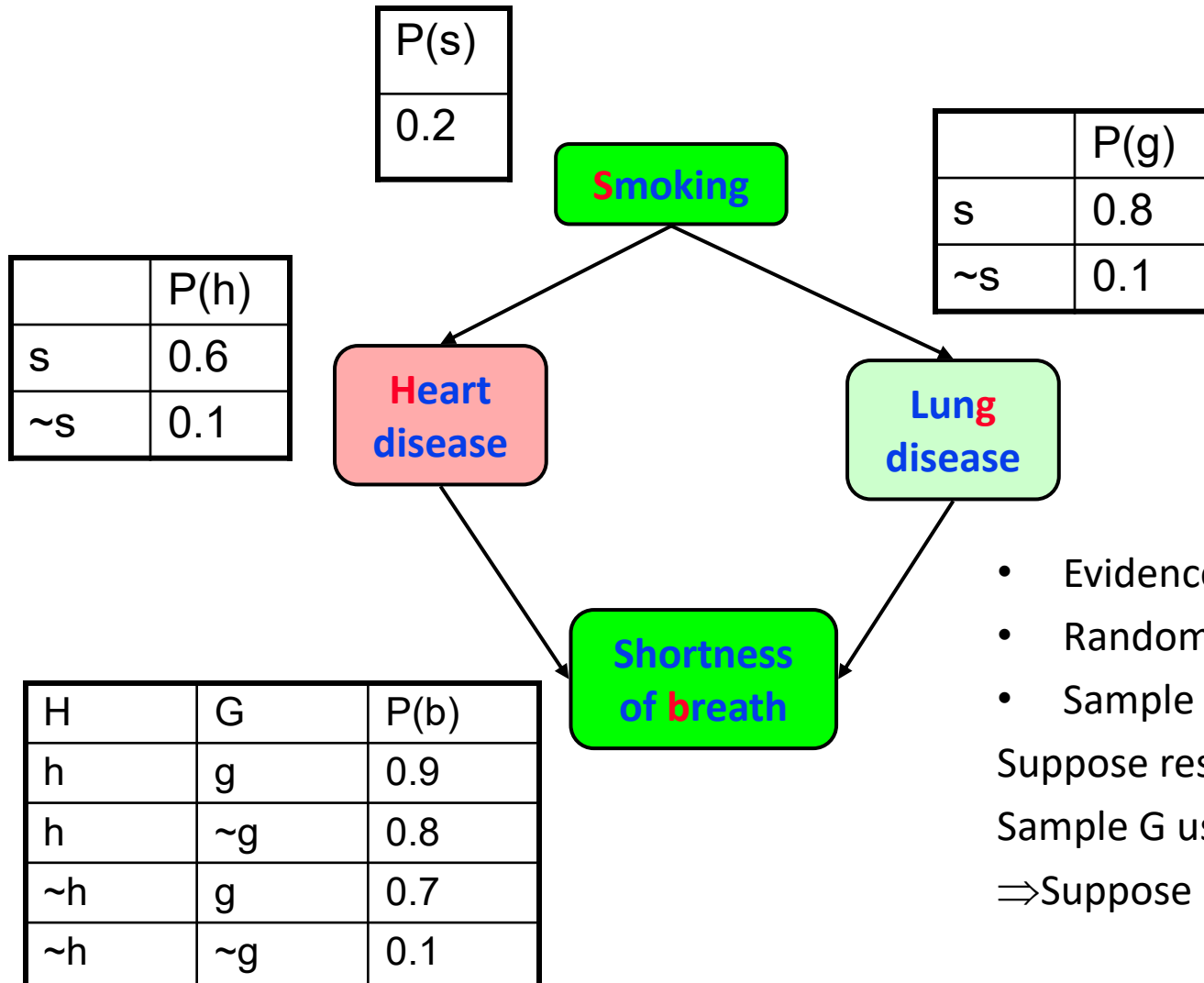
- Evidence: s, b
- Randomly set: h, g
- Sample H using  $P(H|s,g,b)$
- $\Rightarrow$  Suppose result is  $\sim h$

# Example



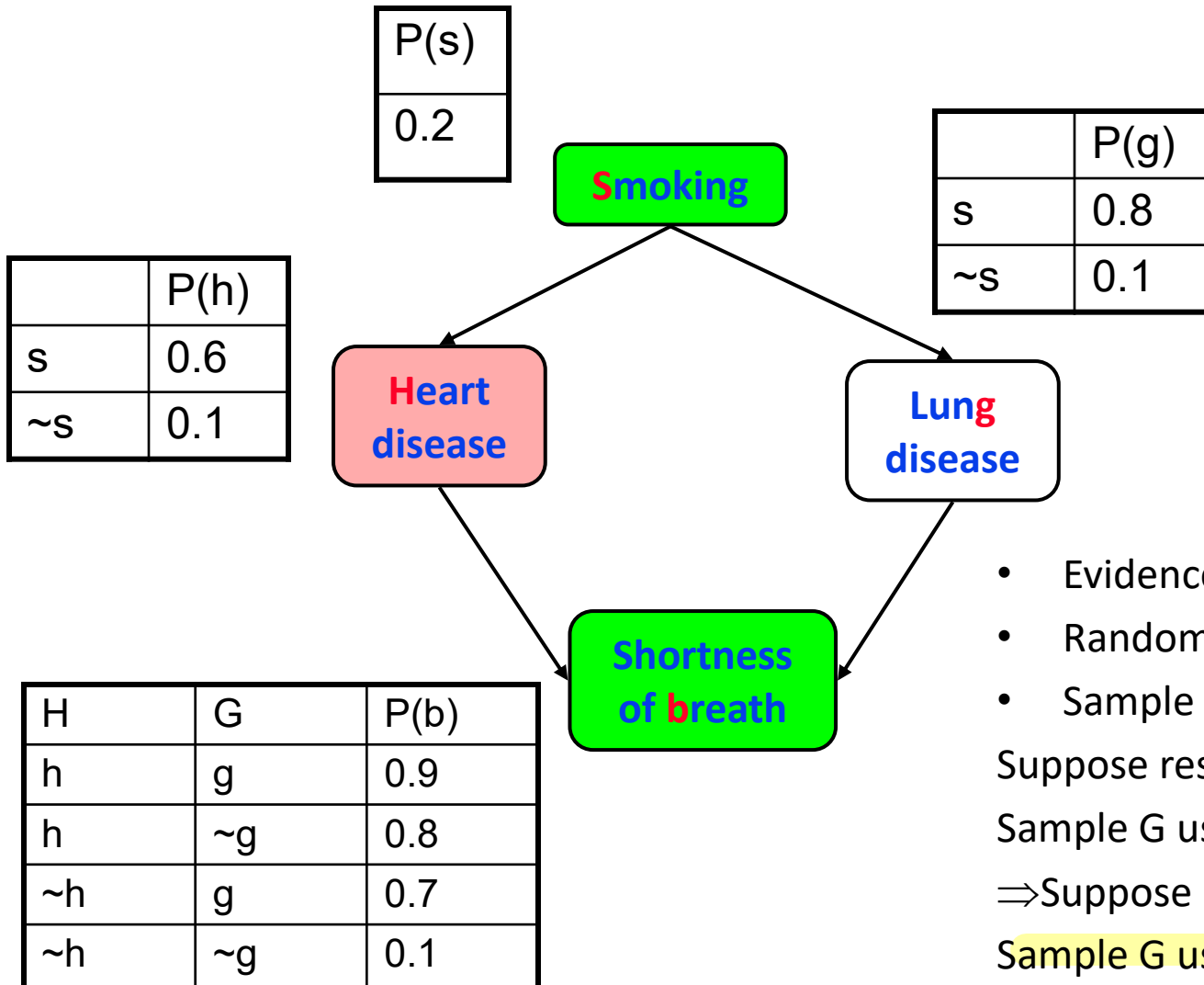
- Evidence: s, b
  - Randomly set: h, g
  - Sample H using  $P(H|s,g,b)$
- Suppose result is ~h
- Sample G using  $P(G|s,\sim h,b)$

# Example



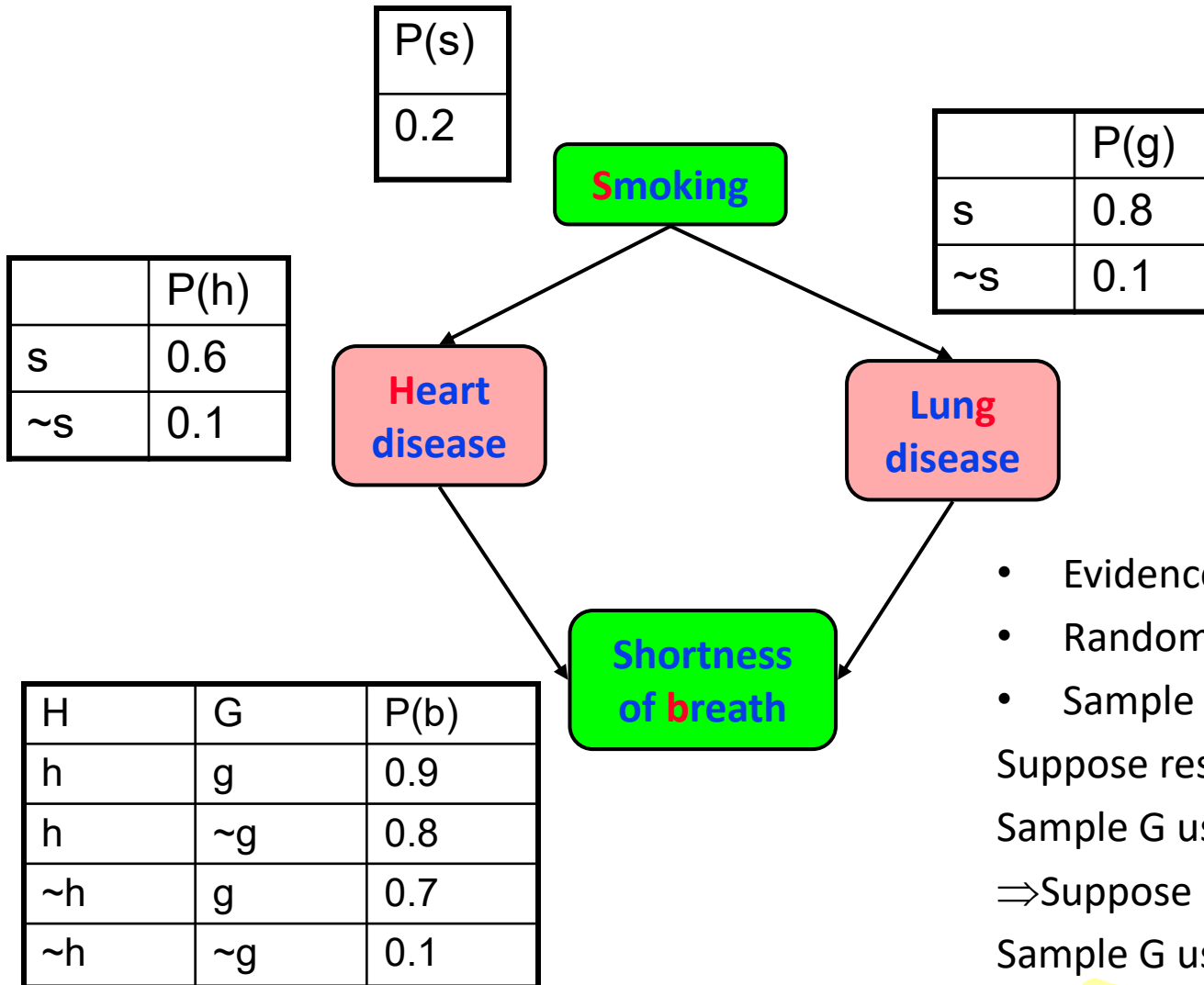
- Evidence: s, b
- Randomly set: h, g
- Sample H using  $P(H|s,g,b)$   
Suppose result is  $\sim h$   
Sample G using  $P(G|s,\sim h,b)$   
 $\Rightarrow$  Suppose result is g

# Example



- Evidence: s, b
  - Randomly set: h, g
  - Sample H using  $P(H|s,g,b)$
- Suppose result is ~h
- Sample G using  $P(G|s,\sim h,b)$
- ⇒ Suppose result is g
- Sample G using  $P(G|s,\sim h,b)$

# Example



- Evidence: s, b
- Randomly set: h, g
- Sample H using  $P(H|s,g,b)$

Suppose result is ~h

Sample G using  $P(G|s,\sim h,b)$

⇒ Suppose result is g

Sample G using  $P(G|s,\sim h,b)$

⇒ Suppose result is ~g

# Gibbs MCMC Summary

$$P(X|E) = \frac{\text{number of samples with } X=x}{\text{total number of samples}}$$

- Advantages:

- No samples are discarded
- No problem with samples of low weight
- Can be implemented very efficiently
  - 10K samples @ second

- Disadvantages:

- Can get stuck if relationship between two variables is *deterministic*
- Many variations have been devised to make MCMC more robust

# Other inference methods

- Exact inference
  - Junction tree
- Approximate inference
  - Belief Propagation
  - Variational Methods
  - Metropolis-Hastings