Inference in Bayesian Networks Part 2

Gennaro De Luca, PhD
Arizona State University

The lecture is based on the slides developed by Prof. Yu Zhang from ASU School of Computing and Augmented Intelligence

ASU Ira A. Fulton Schools of Engineering
Arizona State University

Approximate Inference

Exact inference is NP-complete.

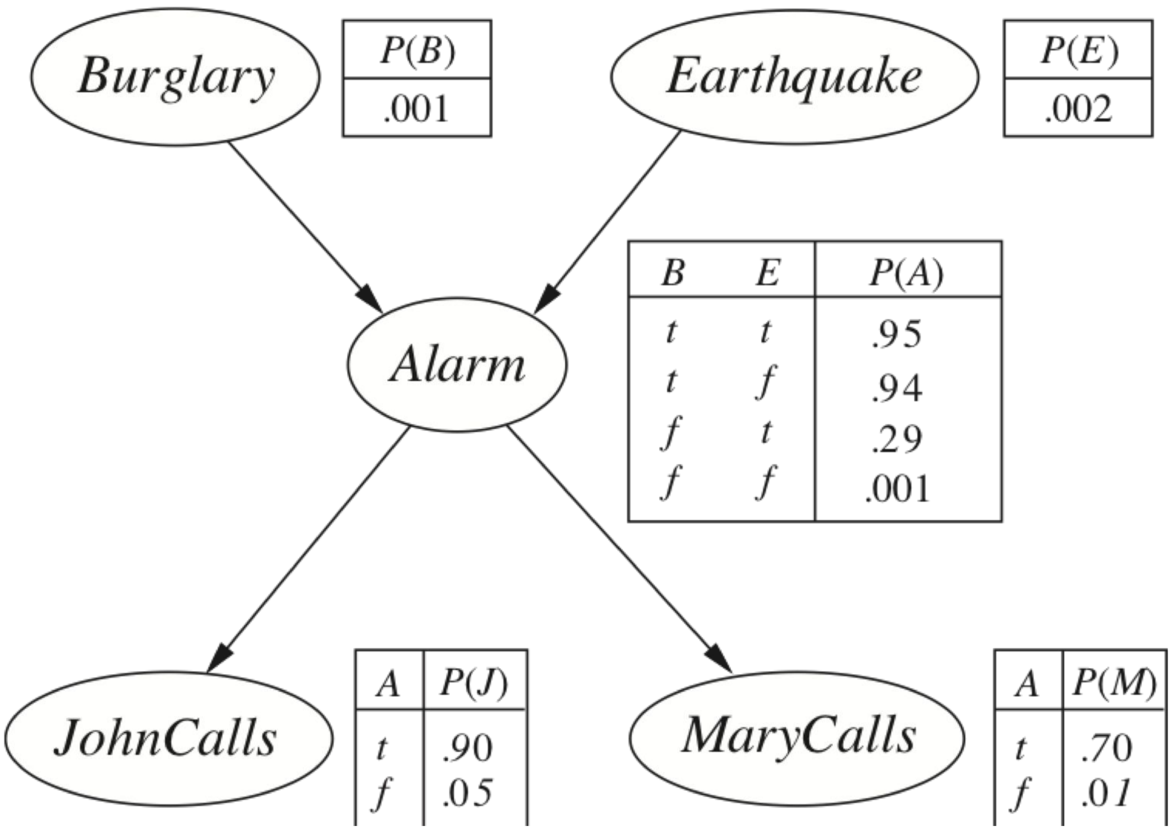
- Must enumerate all possible hidden variables.
- Variable Elimination can be used to improve performance.

Approximate Inference:

- Sampling methods:
  - Draw N samples from a sampling distribution S.
  - Compute an approximate posterior probability.
  - Show this converges to the true probability P.

Prior Sampling

Orderly sample local CPTs for each variable:



Sample:
-b, -e, -a, -j, -m

Prior Sampling (cont'd)

The sample distribution is consistent with the BN:

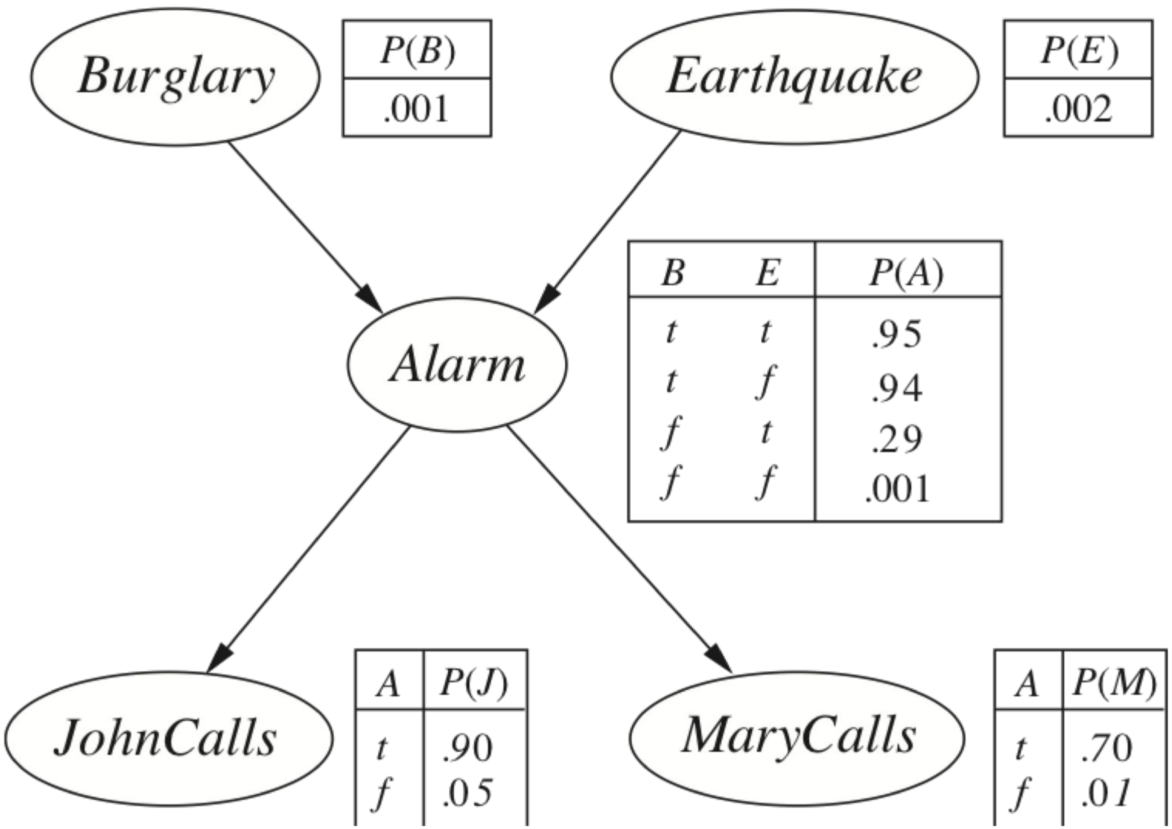
$$S(x_1, \ldots x_n) = \#(x_1, \ldots x_n) / N$$

$$\#(x_1, \ldots x_n)$$
$$\doteq N * \prod_{i=1\ldots n} P(x_i|\text{Parents}(X_i))$$
$$= N * P(x_1, \ldots x_n)$$

$$S(x_1, \ldots x_n) \doteq P(x_1, \ldots x_n)$$

Inference With Samples

Inference:



$P(B)$
.001

$P(E)$
.002

B	E	$P(A)$
t	t	.95
t	f	.94
f	t	.29
f	f	.001

A	$P(J)$
t	.90
f	.05

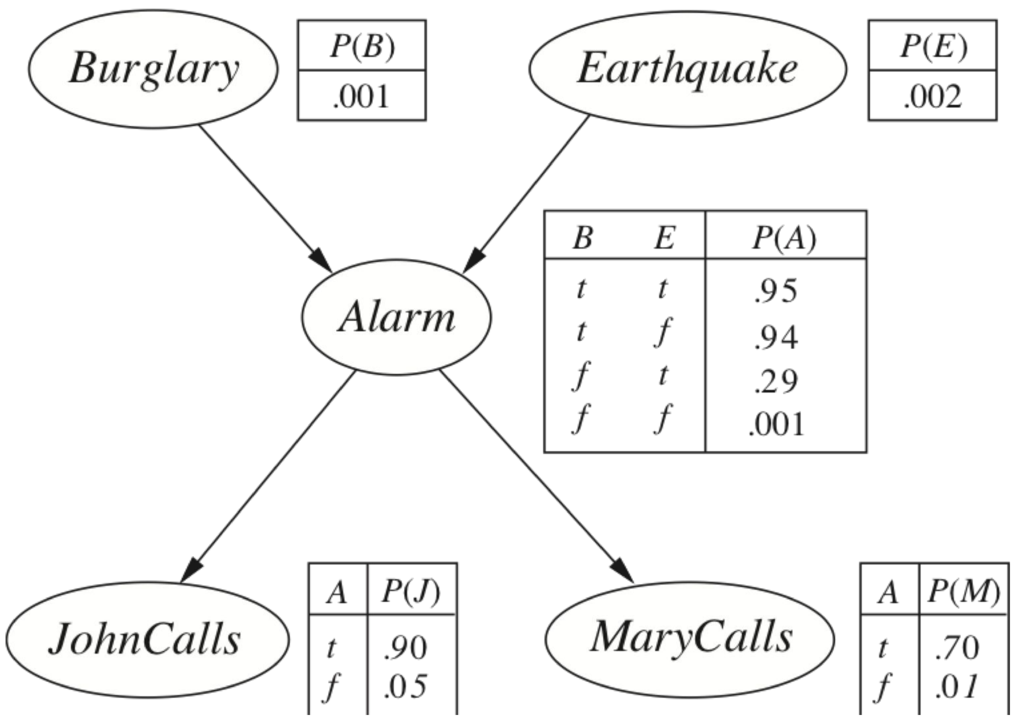
A	$P(M)$
t	.70
f	.01

Sample:
-b, -e, -a, -j, -m
-b, -e, -a, +j, -m
-b, -e, -a, -j, -m
-b, -e, -a, -j, -m
-b, -e, -a, -j, -m

P(X|y) = P(X, y)/P(y)
≐ #(X, y)/#(y)

Rejection Sampling

- To infer about *P(B | +j, +m)*, no need to keep all samples.
  - Reject samples not consistent with +j, +m.
  - Saves space and computation compared to prior sampling.



~~-b, -e, -a, -j, -m~~

~~-b, -e, -a, +j, -m~~

~~-b, -e, -a, -j, -m~~

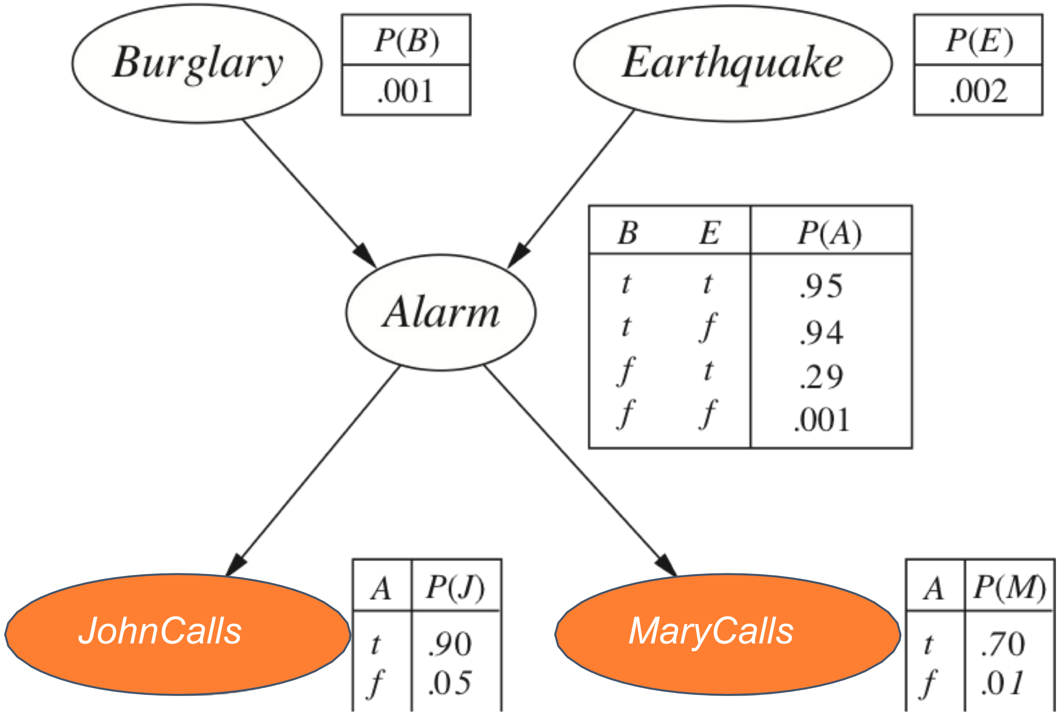
~~-b, -e, -a, -j, -m~~

~~-b, -e, -a, -j, -m~~

Likelihood Weighting

- If evidence is unlikely, e.g., for P(B | +j, +m), rejection sampling will reject many samples.

- Can we fix the evidence?



Sample:
-b, -e, -a, +j, +m
-b, -e, -a, +j, +m
-b, -e, -a, +j, +m
-b, -e, -a, +j, +m
-b, -e, -a, +j, +m

Likelihood Weighting (cont'd)

- **The sample distribution is not consistent with the BN.**

  $S(x_1, \ldots x_n) = \#(x_1, \ldots x_n) / N$

  $\#(x_1, \ldots x_n)$
  $\doteq N * \prod_{x_i \notin EP}(x_i|Parents(x_i))$

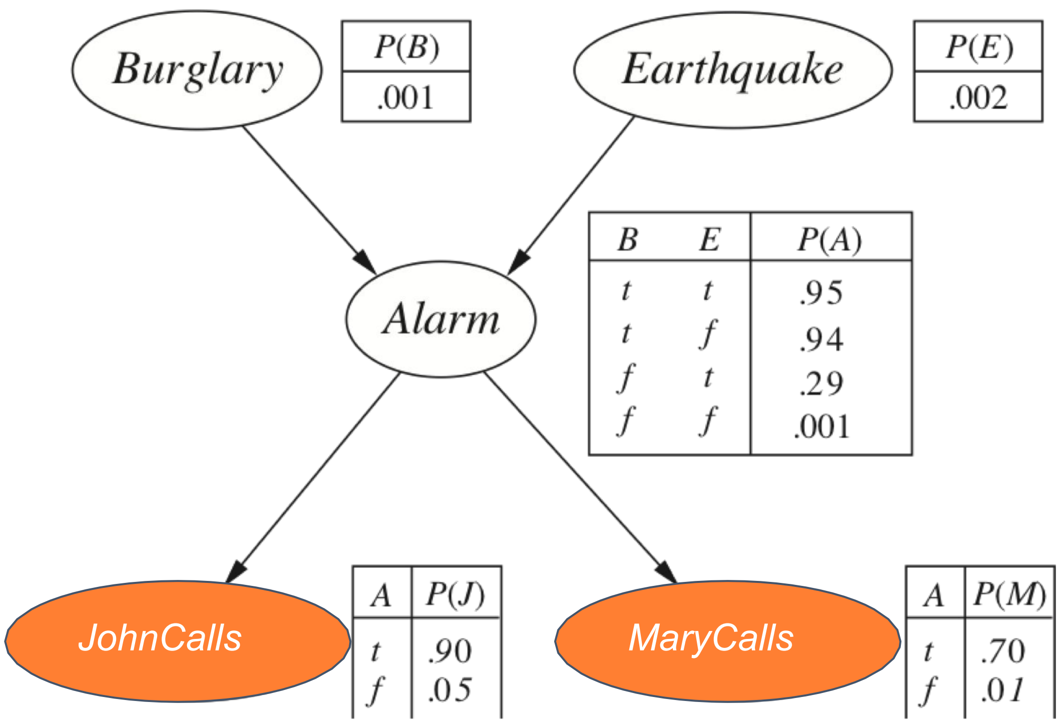
- **Weight each sample by $w = \prod_{x_i \in EP}(x_i|Parents(x_i))$:**

  $S(x_1, \ldots x_n) = \#(x_1, \ldots x_n) * w / M$

  $S(x_1, \ldots x_n) \doteq P(x_1, \ldots x_n) * (N/M) \propto P(x_1, \ldots x_n)$

Likelihood Weighting (cont'd)

- Fix the evidence, sample and weight each sample:



P(B)
.001

P(E)
.002

B	E	P(A)
t	t	.95
t	f	.94
f	t	.29
f	f	.001

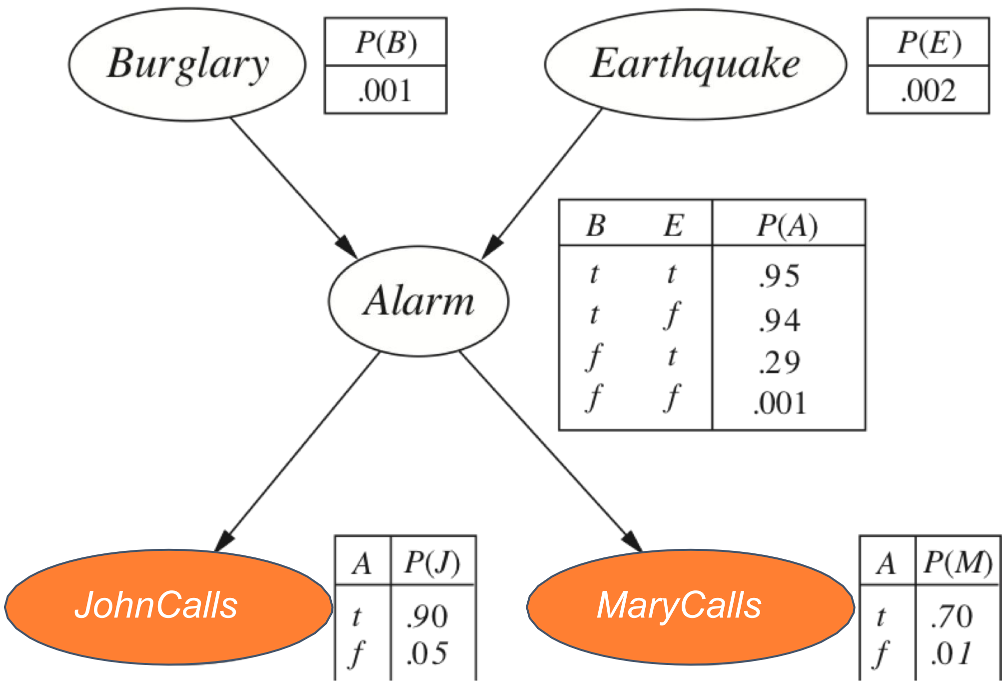
A	P(J)
t	.90
f	.05

A	P(M)
t	.70
f	.01

Sample:
-b, -e, -a, +j, +m
w = 0.05 * 0.01

Markov Chain Monte Carlo

- Start with an arbitrary $x_1, \ldots, x_n$ that is consistent with $e_1, \ldots, e_k$.
- Repeat: sample $P(X_i \mid x_1,\ldots,x_{i-1}, x_{i+1},\ldots x_n)$ $X_i \notin E$.



Sample:

-b, -e, -a, +j, +m →

sample P(B | -e, -a, +j, +m)

+b, -e, -a, +j, +m →

sample P(E | +b, -a, +j, +m)

+b, -e, -a, +j, +m

… (repeat many times for each sample)

Markov Chain Monte Carlo (cont'd)

Sampling from the conditional distribution:

$$P(X_i \mid x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$$

$$= P(x_1, \ldots x_n) / \sum x_i\, P(x_1, \ldots, x_n)$$

$$= \prod f(x_i) / \sum x_i \prod f(x_i)$$

f are all CPTs with x_i