

430.457

Introduction to Intelligent Systems

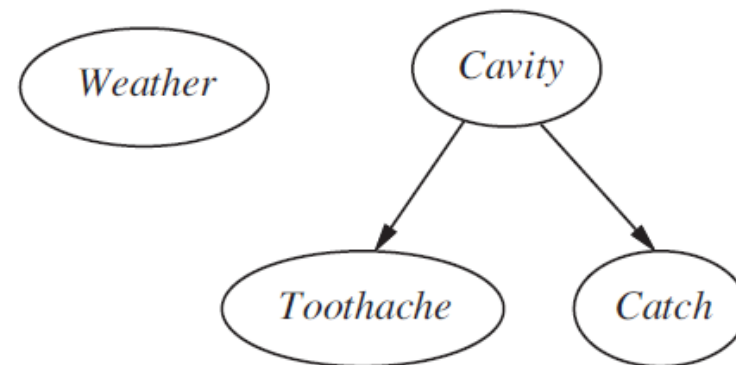
Prof. Songhwai Oh
ECE, SNU

Chapter 13. Probabilistic Reasoning

BAYESIAN NETWORKS

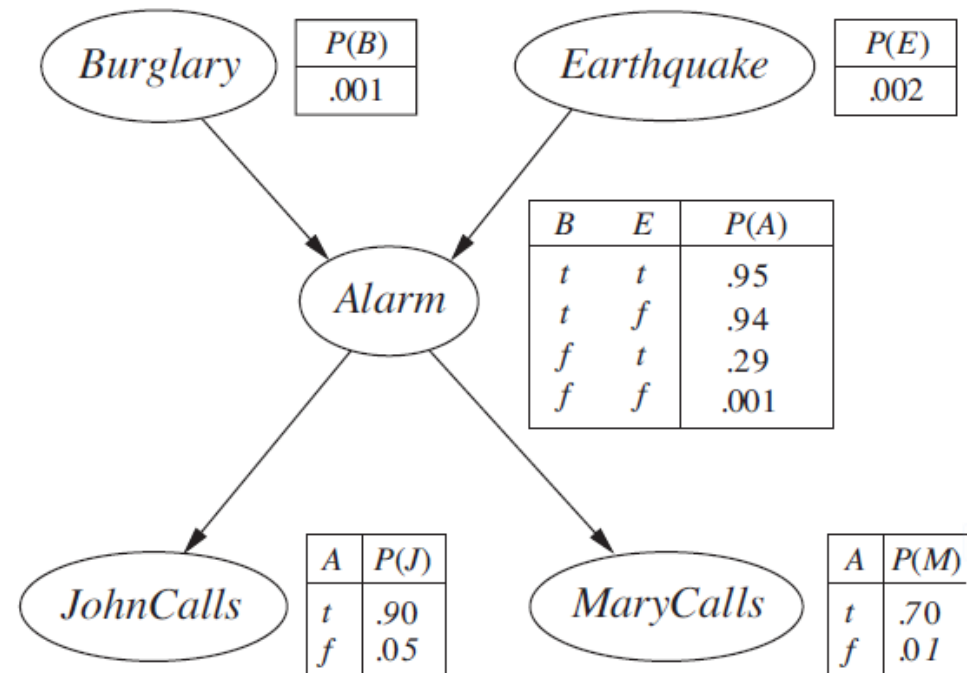
Bayesian Networks

- Probability theory + Graph theory
- Compact representation for a complex probability distribution (cf. full joint distribution)
 - By taking advantage of independence and conditional independence
- A Bayesian network is a DAG (directed acyclic graph)
 - Node = random variable
 - Node X_i has a conditional probability distribution $P(X_i | \text{Parents}(X_i))$



Example: Burglar Alarm

- A new burglar alarm at home
 - Fairly reliable at detecting a burglary, but also responds to minor earthquake
- Two neighbors, John and Mary, who calls you at work when they hear the alarm
 - John nearly always calls when he hears the alarm but sometimes confuses the telephone ringing with the alarm
 - Mary likes loud music and often misses the alarm
- Given the evidence of who has or has not called, what is the probability of a burglary?

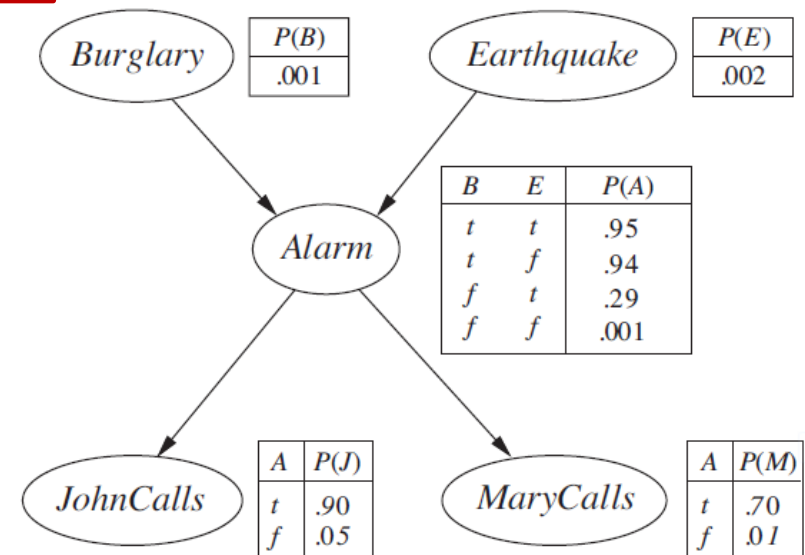


Semantics of Bayesian Networks

- Two views on the semantics of Bayesian networks
 - A representation of the joint distribution
 - An encoding of a collection of conditional independence statements
 - * Two views are equivalent
- Full joint distribution for a Bayesian network:

$$P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i | \text{parents}(X_i))$$

$$\begin{aligned} P(B, E, A, J, M) &= P(B)P(E)P(A|B, E) \\ &\times P(J|A)P(M|A) \end{aligned}$$



Constructing a Bayesian Network

- Chain rule:

$$\begin{aligned} P(x_1, \dots, x_n) &= P(x_n | x_{n-1}, \dots, x_1) P(x_{n-1} | x_{n-2}, \dots, x_1) \cdots P(x_2 | x_1) P(x_1) \\ &= \prod_{i=1}^n P(x_i | x_{i-1}, \dots, x_1) \end{aligned}$$

- Bayesian networks: $P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i | \text{parents}(X_i))$.
- Hence, we must have $P(X_i | X_{i-1}, \dots, X_1) = P(X_i | \text{Parents}(X_i))$, provided that $\text{Parents}(X_i) \subseteq \{X_{i-1}, \dots, X_1\}$.

Algorithm for constructing a Bayesian network

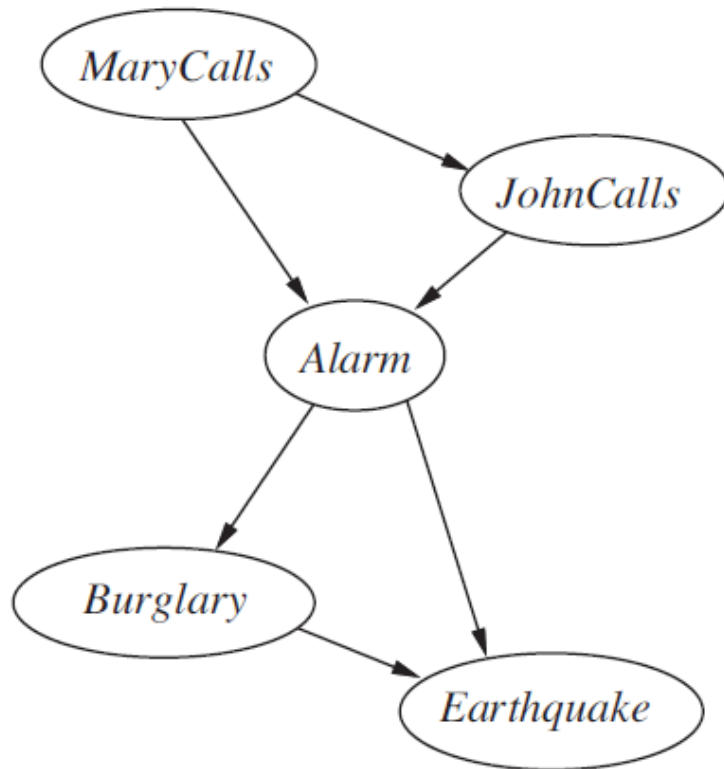
- Nodes: Determine the set of variables to model the domain and order them, $\{X_1, \dots, X_n\}$.
- Links: For $i = 1$ to n do:
 - Choose a minimal set of parents for X_i from $\{X_1, \dots, X_{i-1}\}$
 - For each parent insert a link from the parent to X_i
 - Associate the conditional probability tables, $P(X_i | \text{Parents}(X_i))$ to node X_i .

Compactness of Bayesian Networks

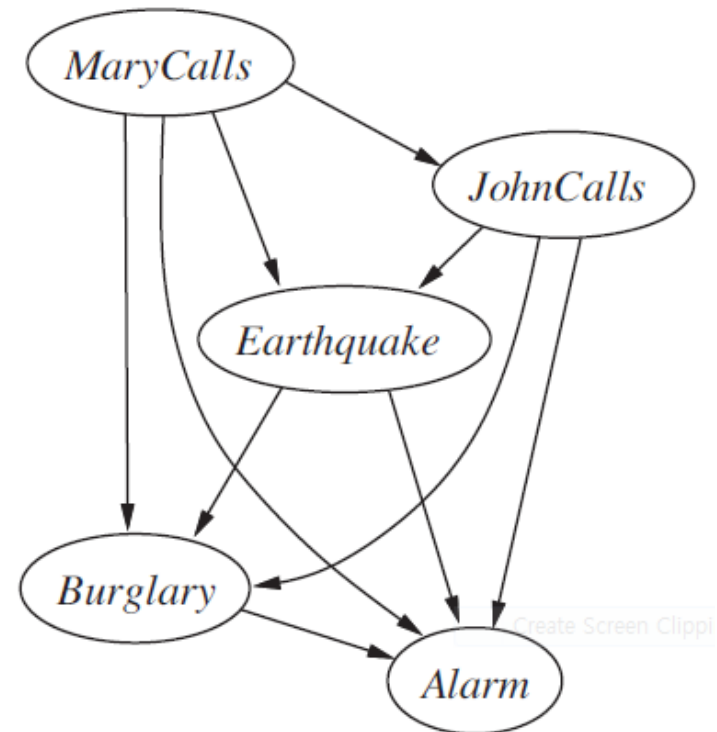
- Suppose we have n Boolean variables.
- Joint distribution requires 2^n numbers.
- A Bayesian network with at most degree k requires only $n2^k$ numbers.
- E.g., if $n = 30$ and each node has five parents ($k = 5$), a Bayesian network requires 960 numbers while the full joint distribution requires $2^{30} \approx 10^9 = 1,000,000,000$ numbers.
- The ordering of nodes at the construction of a Bayesian network determines the complexity of the resulting network.

Effects of Node Ordering

{ MaryCalls, JohnCalls,
Alarm, Burglary, Earthquake }



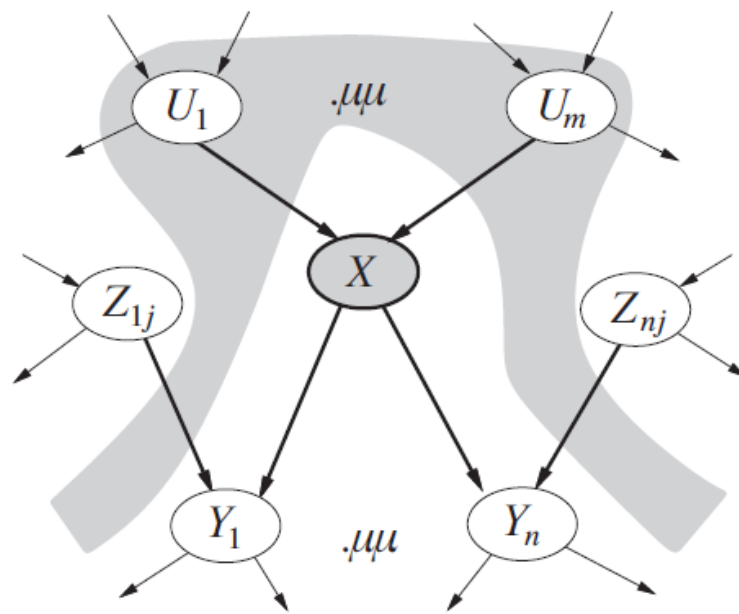
{ MaryCalls, JohnCalls,
Earthquake, Burglary, Alarm }



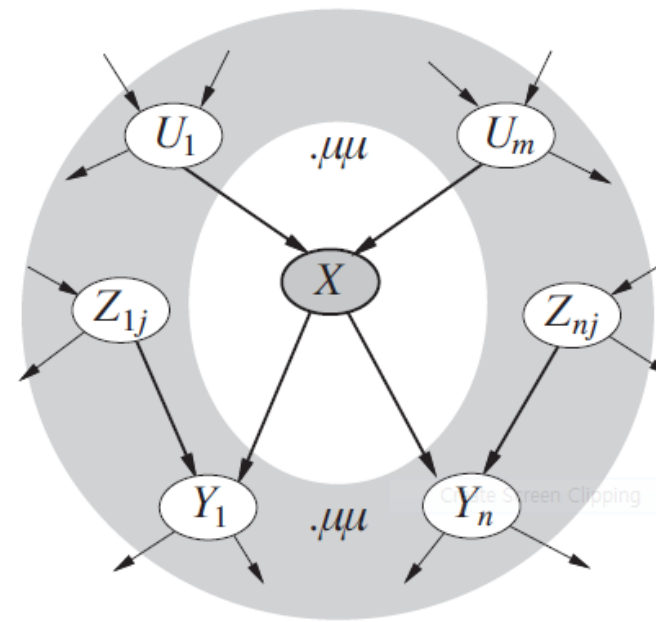
Compact representation if causes precede effects (causal model)

Conditional Independence

- a. Each variable is conditionally independent of its non-descendants given its parents (recall $P(X_i \mid \text{Parents}(X_i))$)
- b. A node is conditionally independent of all other nodes in the networks, given its parents, children, and co-parents (children's parents) - **Markov blanket**



(a)



(b)

EFFICIENT REPRESENTATION OF CONDITIONAL DISTRIBUTIONS

Noisy-OR Model

- Generalization of the logical OR. (E.g., Fever is true iff Cold, Flu, or Malaria is true.)
- Why noisy? There is an uncertainty about the ability of each parent to cause the child to be true.
- Assumptions: (1) all possible causes are listed, (2) inhibition of each parent is independent of inhibition of any other parents.
- Inhibition probabilities:

$$q_{cold} = P(\neg fever | cold, \neg flu, \neg malaria) = 0.6$$

$$q_{flu} = P(\neg fever | \neg cold, flu, \neg malaria) = 0.2$$

$$q_{malaria} = P(\neg fever | \neg cold, \neg flu, malaria) = 0.1$$

Fever is *false* iff all its *true* parents are inhibited.

- General rule:

$$P(x_i | parents(X_i)) = 1 - \prod_{\{j: X_j = true\}} q_j$$

Noisy-OR Model

$$q_{cold} = P(\neg fever | cold, \neg flu, \neg malaria) = 0.6$$

$$q_{flu} = P(\neg fever | \neg cold, flu, \neg malaria) = 0.2$$

$$q_{malaria} = P(\neg fever | \neg cold, \neg flu, malaria) = 0.1$$

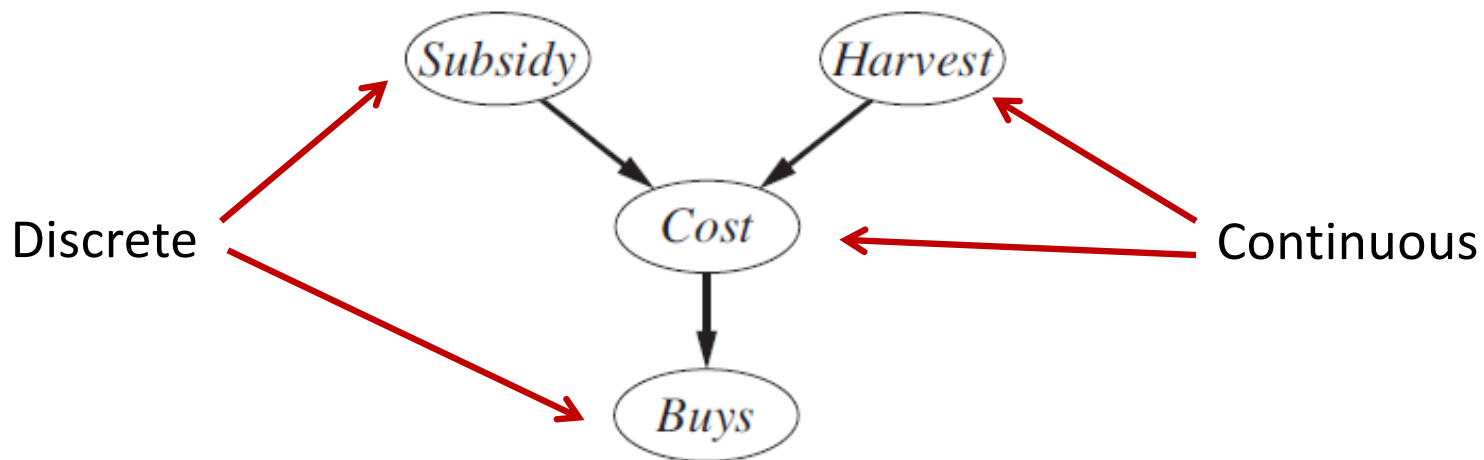
$$P(x_i | parents(X_i)) = 1 - \prod_{\{j: X_j = true\}} q_j$$

<i>Cold</i>	<i>Flu</i>	<i>Malaria</i>	$P(Fever)$	$P(\neg Fever)$
F	F	F	0.0	1.0
F	F	T	0.9	0.1
F	T	F	0.8	0.2
F	T	T	0.98	$0.02 = 0.2 \times 0.1$
T	F	F	0.4	0.6
T	F	T	0.94	$0.06 = 0.6 \times 0.1$
T	T	F	0.88	$0.12 = 0.6 \times 0.2$
T	T	T	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

- Requires only $O(k)$ parameters instead of $O(2^k)$ for the full conditional probability table, where k is the number of parents.

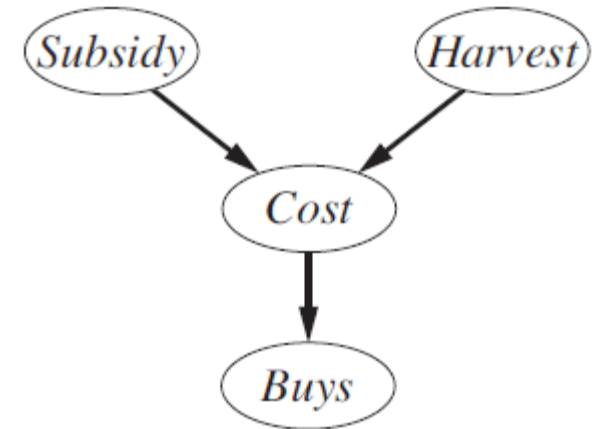
Hybrid Bayesian Networks

- Representing continuous variables in a Bayesian network
 - Discretization (may requires large CPTs)
 - Parametric model (Gaussian, Gamma, Beta, etc.)
 - Nonparametric model
- Hybrid Bayesian network
 - A network with both discrete and continuous variables



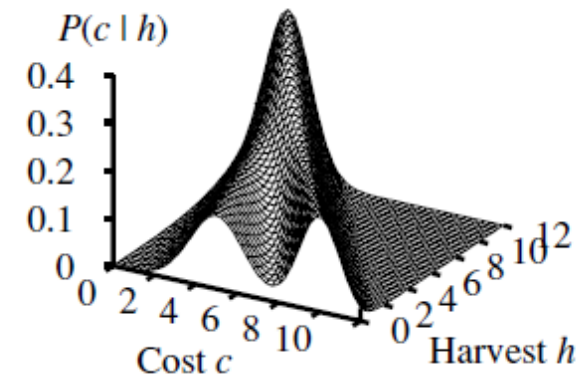
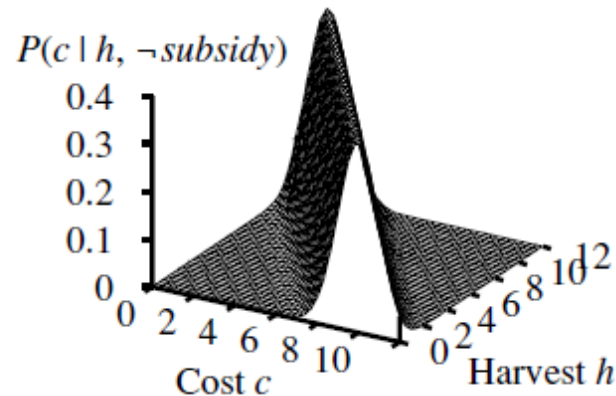
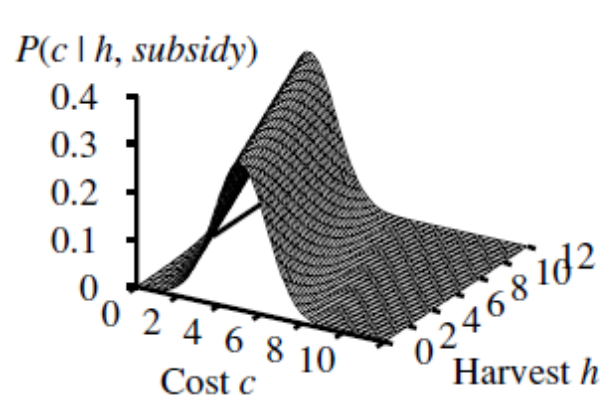
From Discrete to Continuous

- $P(\text{Cost}|\text{Harvest}, \text{Subsidy})$ can be specified using
 - $P(\text{Cost}|\text{Harvest}, \text{Subsidy})$ and
 - $P(\text{Cost}|\text{Harvest}, \neg \text{Subsidy})$.
- **Linear Gaussian model**



$$P(c|h, \text{subsidy}) = \mathcal{N}(c|a_t h + b_t, \sigma_t^2) = \frac{1}{\sqrt{2\pi}\sigma_t} \exp\left(-\frac{1}{2\sigma_t^2} (c - (a_t h + b_t))^2\right)$$

$$P(c|h, \neg \text{subsidy}) = \mathcal{N}(c|a_f h + b_f, \sigma_f^2) = \frac{1}{\sqrt{2\pi}\sigma_f} \exp\left(-\frac{1}{2\sigma_f^2} (c - (a_f h + b_f))^2\right)$$



From Continuous to Discrete

Specifying *Buys* given *Cost*:

- **Probit distribution:**

$$P(buys|Cost = c) = \Phi\left(\frac{-c + \mu}{\sigma}\right),$$

where $\Phi(x) = \int_{-\infty}^x \mathcal{N}(x|0, 1)dx$.

- **Logit distribution:**

$$P(buys|Cost = c) = \frac{1}{1 + \exp\left(-2\frac{-c + \mu}{\sigma}\right)},$$

where $1/(1 + e^{-x})$ is called a logistic function.

a.k.a. logistic function, sigmoid function
(related: softmax function)

