ECE521:
Lecture 16

13 March 2017:

Bayesian networks continued,
conditional independence

With thanks to Brendan Frey and others

This week

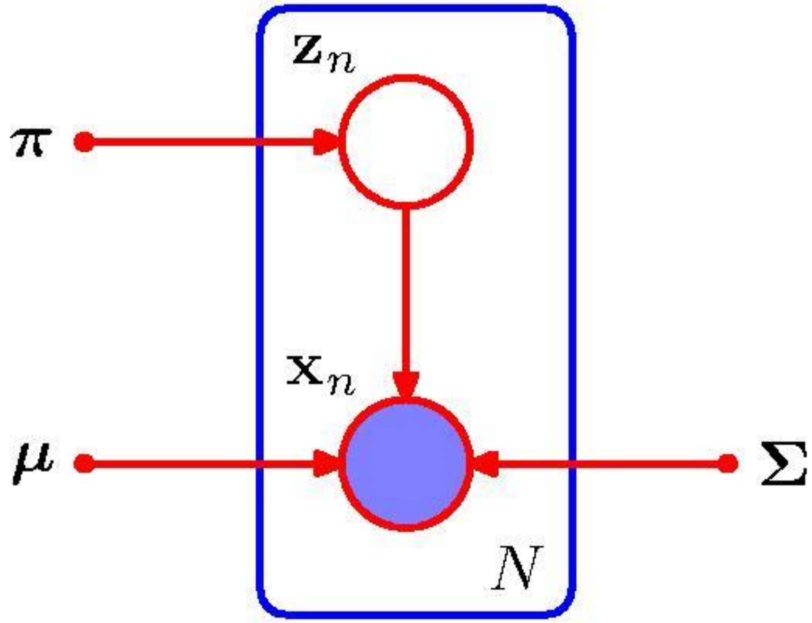
- Exploring both types of graphical model (directed and undirected)

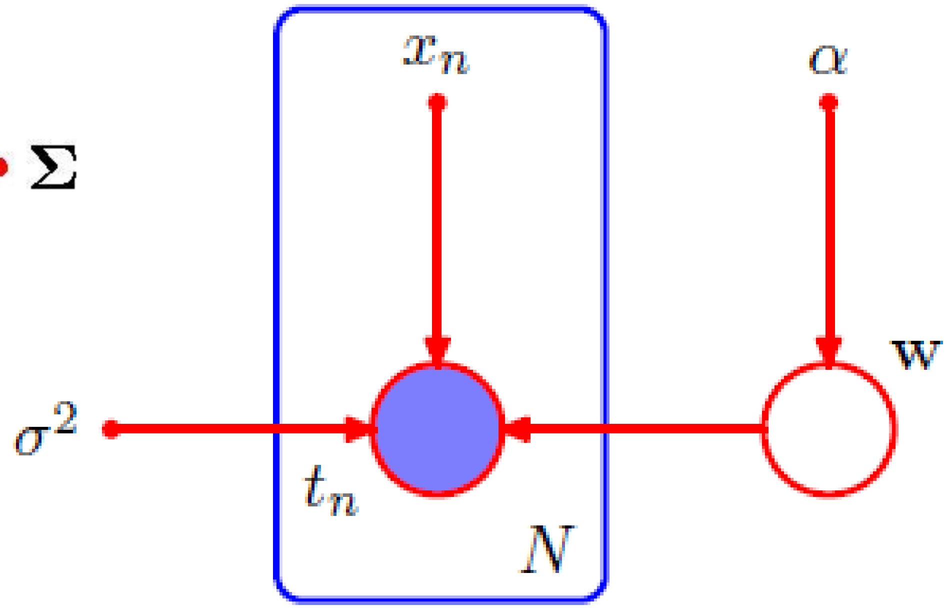
- Examples of additional perspectives:
  - Bishop 2006: parts of chap. 8
  - Murphy 2012: parts of chap. 10
  - Russell & Norvig, 2009: parts of chap. 14 *(Artificial Intelligence: A Modern Approach*)

Outline for today

- Review from last week
- Example of a famous Bayesian network
- Inference in Bayesian networks:
  - Exact
  - Approximate
- Conditional independence in BNs

Graphical models





Example Uses of Conditional Independence

- Naive Bayes assumption: all dimensions are independent given the label *z*

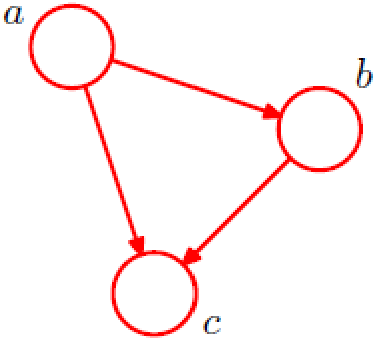
$$p(\mathbf{x}|z=k) = p(x_1, \ldots, x_D|z=k)$$
$$= \prod_{d=1}^{D} p(x_d|z=k)$$

- Markov assumption: the future is independent of the past given the present

$$p(x_d|x_1, \ldots, x_{d-1}) = p(x_d|x_{d-1})$$

Bayesian Networks

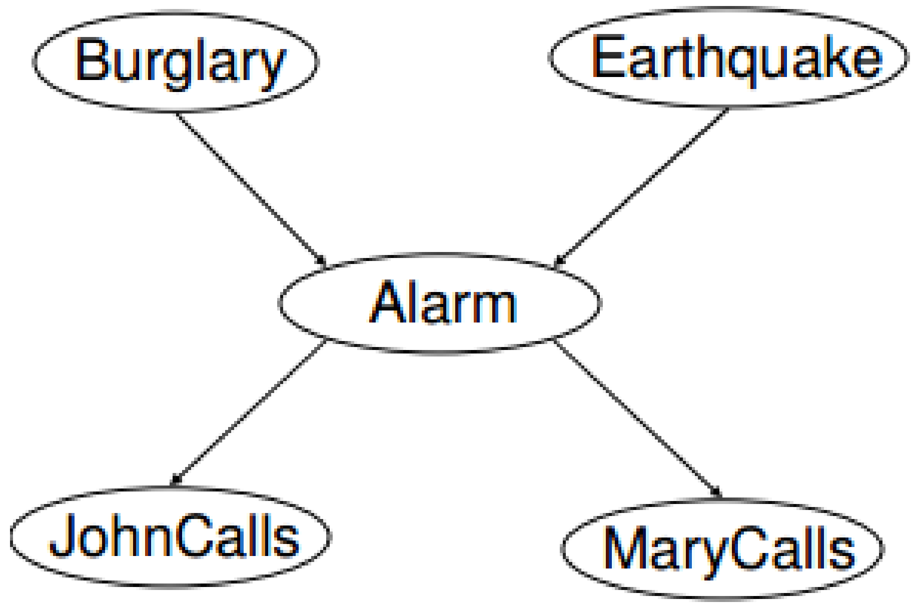
- Simple and visual: you can put conditional probability tables next to nodes
- Can be a compact representation of the full joint distribution, for locally structured (sparse) cases



Outline for today

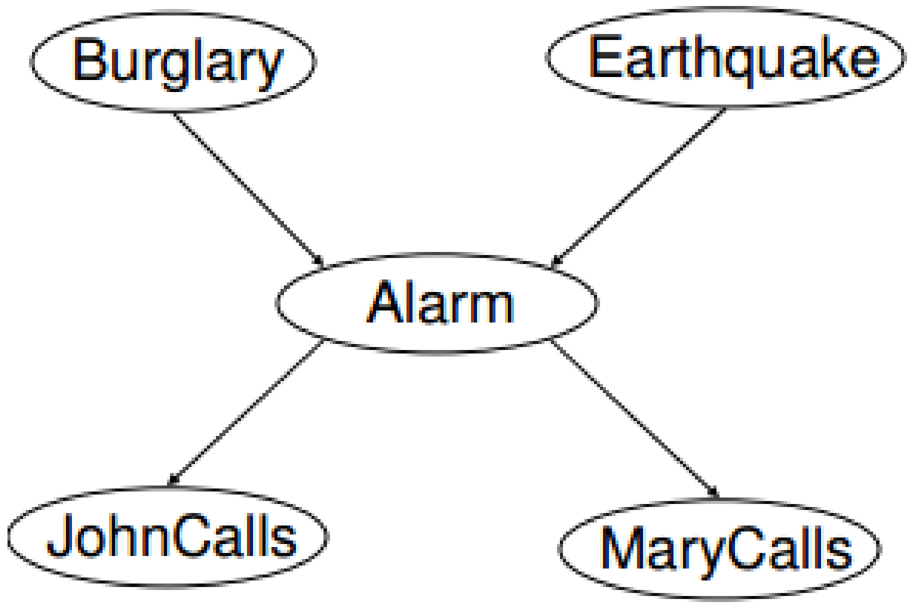
- Review from last week
- **Example of a famous Bayesian network**
- Inference in Bayesian networks:
  - Exact
  - Approximate
- Conditional independence in BNs

Bayesian Network Example



Bayesian Network Example

P(B) = 0.001



P(E) = 0.002

B	E	P(A = True|B=b,E=e)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(J = True|A=a)
T	0.90
F	0.05

A	P(M = True|A=a)
T	0.70
F	0.01

Calculating on a Bayesian Network

Recall that $$p(x_1, \ldots x_D) = \prod_{d=1}^{D} p(x_d | X_{\mathcal{A}_d})$$

So, P(B,E,A,J,M) = ?

Answer: ~1.2 x 10^{-6}

P(B|J,M) = ?

Answer: Tricky!

Outline for today

- Review from last week
- Example of a famous Bayesian network
- **Inference in Bayesian networks**
  - Exact
  - Approximate
- Conditional independence in BNs

Exact Inference in BNs

- $\mathrm{P}(B|J, M) = \frac{P(B,J,M)}{P(B,J,M)+P(\bar{B},J,M)}$

where $P(B, J, M) = \sum_E \sum_A P(B, E, A, J, M)$

$$= P(B) \sum_E P(E) \sum_A P(A|B, E) P(J|A) P(M|A)$$

$\mathrm{P}(B|J, M) \approx 0.284$

Direct Sampling: simulating a graphical model

- Put the nodes in ancestral order (parents coming before children)
- Sample each variable given its parents
- The probability of an event can be estimated as the fraction of all complete events generated that match the partially specified event. e.g. if 6 out of 2000 samples have A=true, $P(A) \approx 0.003$

Rejection Sampling:

- Helps us to estimate

  $$P(X|E) = P(X,E) \,/\, P(E)$$

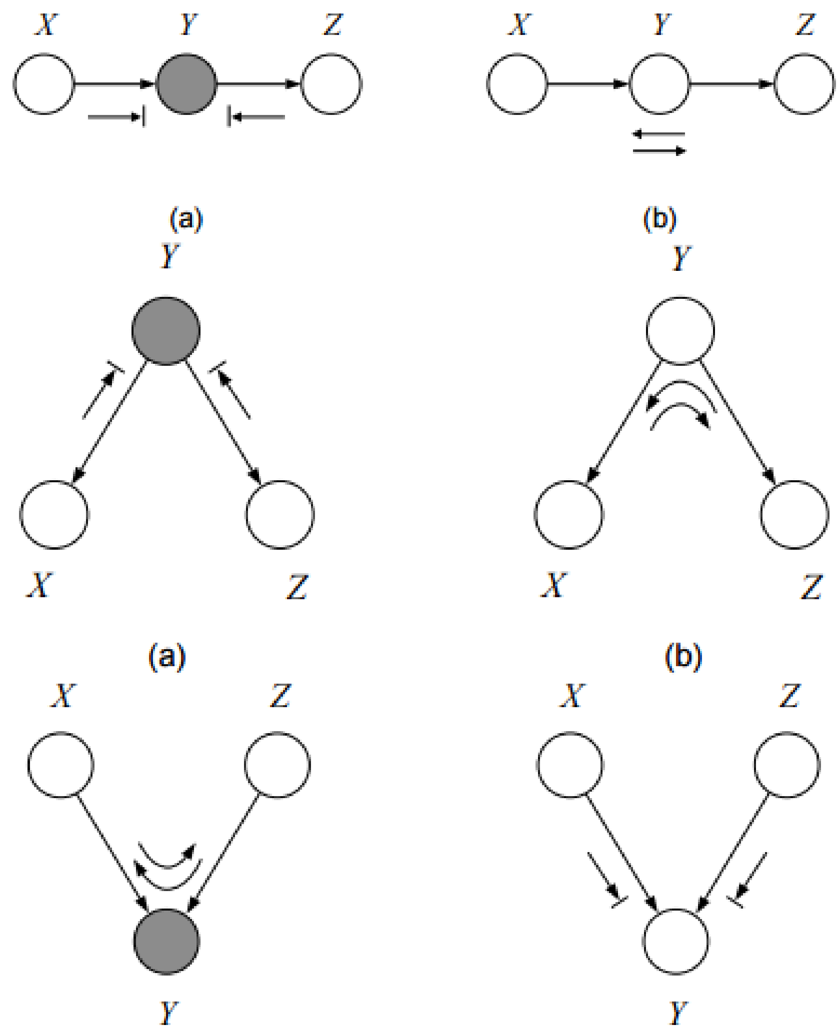
  for a query variable X and evidence E
- e.g. we can estimate $P(J|M)$
- Sample 1000 times and reject all samples in which M=false. From the remaining N samples (M=true), estimate: $P(J|M) \approx N_{J=\text{true}} \,/\, N$

Outline for today

- Review from last week
- Example of a famous Bayesian network
- Inference in Bayesian networks:
  - Exact
  - Approximate
- **Conditional independence in BNs**

Who is conditionally independent of whom?

- *Bayes ball algorithm*
- Quickly determines whether $X \perp\!\!\!\perp Z \mid Y$

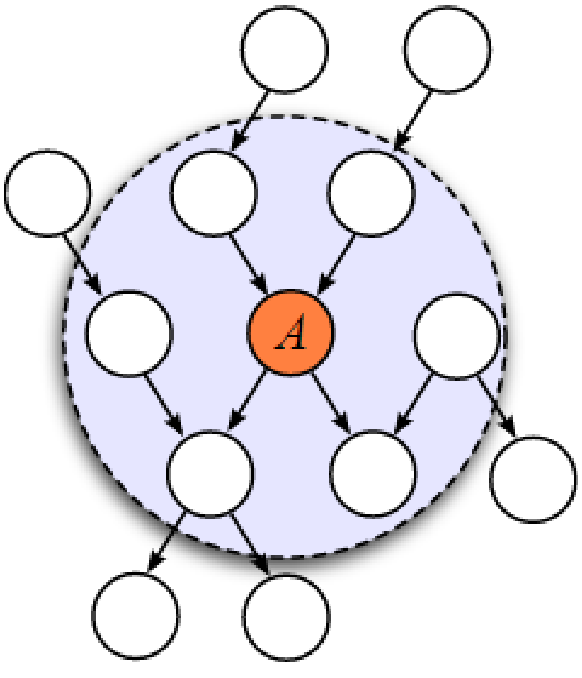


Comparing one node vs another

Conditional independence relations in Bayesian Networks

There are two, equivalent specifications:

- A node is C.I. of its non-descendants given its parents
- A node is C.I. of all other nodes, given its Markov blanket (parents, children, and co-parents)



Comparing one node vs rest of network

How does this relate?

- For a given node, A, the Markov blanket of A is the minimal set of nodes which *Bayes-ball*-separates (renders C.I.) node A from all other nodes in the network

- True/false:
  - $M \perp\!\!\!\perp J \mid A$
  - $B \perp\!\!\!\perp E \mid A$
  - $M \perp\!\!\!\perp E \mid A$