Networks

Introduction to Bayesian Networks

Outline

- Motivation for Bayesian networks
  - A probabilistic model for the *lac* operon
- Definition of Bayesian networks
- Representation of conditional probability distributions in Bayesian networks
- Computational tasks for Bayesian networks

Probabilistic Model of *lac* Operon

- suppose we represent the system by the following discrete variables

L (lactose)	present, absent
G (glucose)	present, absent
I (lacI)	present, absent
C (CAP)	present, absent
lacI-unbound	true, false
CAP-bound	true, false
Z (lacZ)	high, low, absent

- suppose (realistically) the system is not completely deterministic
- the joint distribution of the variables could be specified by $2^6 \times 3 - 1 = 191$ parameters

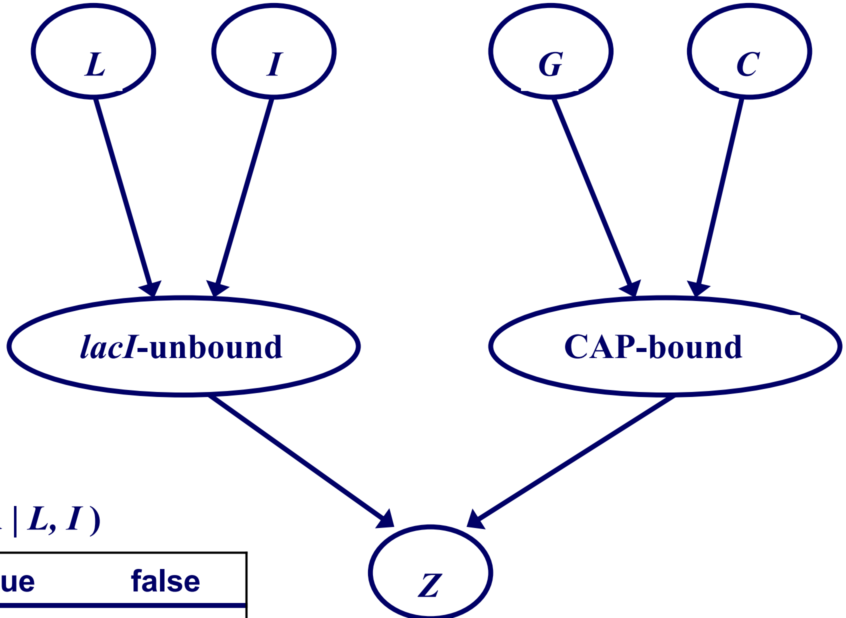
Motivation for Bayesian Networks

- Explicitly state (conditional) independencies between random variables
  - Provide a more compact model (fewer parameters)
- Use directed graphs to specify model
  - Take advantage of graph algorithms/theory
  - Provide intuitive visualizations of models

A Bayesian Network for the lac System

Pr (*L*)

absent	present
0.9	0.1



Pr (l*acI*-unbound | *L*, *I*)

L	*I*	true	false
absent	absent	0.9	0.1
absent	present	0.1	0.9
present	absent	0.9	0.1
present	present	0.9	0.1

Pr (*Z* | *lacI*-unbound, *CAP*-bound)

lacI-unbound	CAP-bound	absent	low	high
true	false	0.1	0.8	0.1
true	true	0.1	0.1	0.8
false	false	0.8	0.1	0.1
false	true	0.8	0.1	0.1

Bayesian Networks

- Also known as Directed Graphical Models
- a BN is a Directed Acyclic Graph (DAG) in which
  - the nodes denote random variables
  - each node X has a *conditional probability distribution* (CPD) representing $P(X \mid Parents(X))$

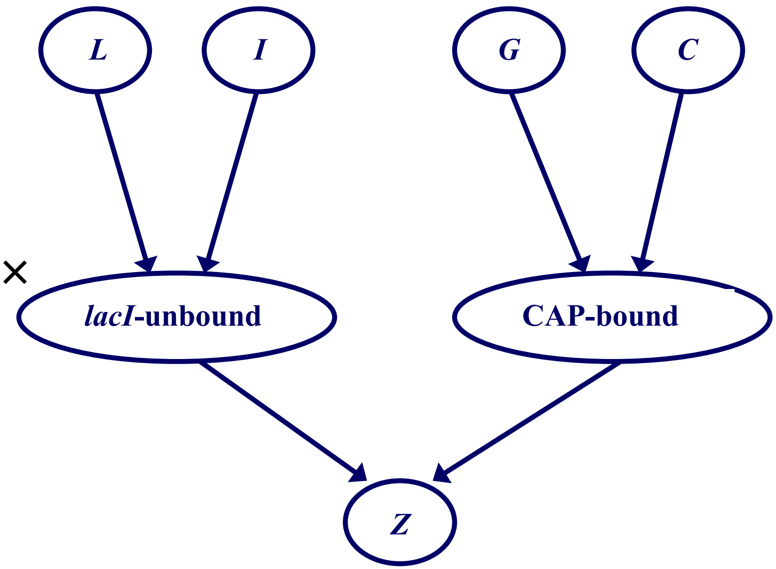
- the intuitive meaning of an edge from X to Y is that X *directly influences* Y

- formally: each variable X is independent of its non-descendants given its parents

Bayesian Networks

- a BN provides a *factored* representation of the joint probability distribution

$$\Pr(L,I,LU,G,C,CB,Z) =$$
$$\Pr(L) \times \Pr(I) \times$$
$$\Pr(LU \mid L,I) \times$$
$$\Pr(G) \times \Pr(C) \times$$
$$\Pr(CB \mid G,C) \times$$
$$\Pr(Z \mid LU,CB)$$



- this representation of the joint distribution can be specified with 20 parameters (vs. 191 for the unfactored representation)

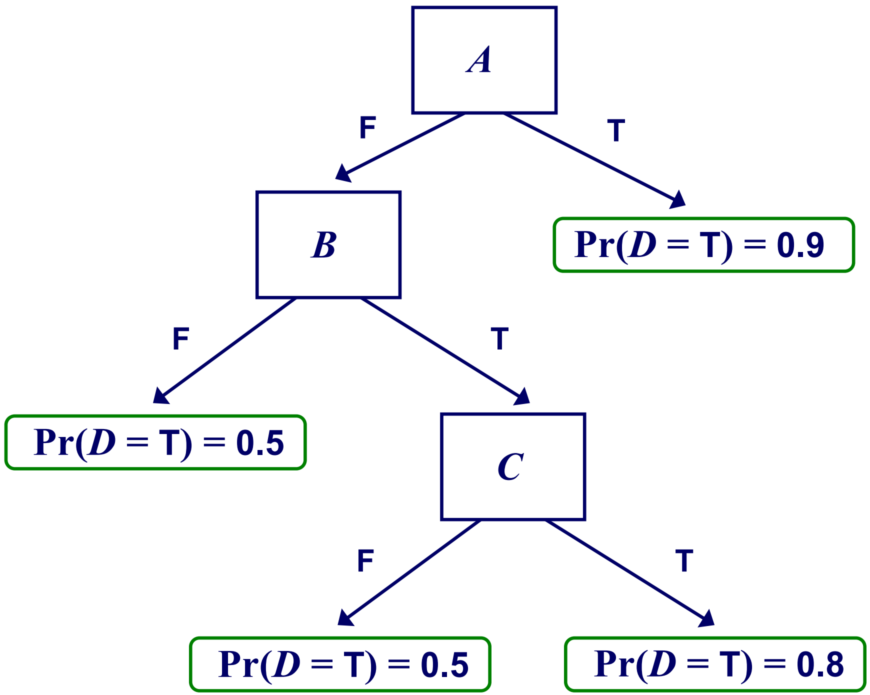
Representing CPDs for Discrete Variables

- CPDs can be represented using tables or trees
- consider the following case with Boolean variables *A*, *B*, *C*, *D*

Pr(*D* | *A*, *B*, *C*)

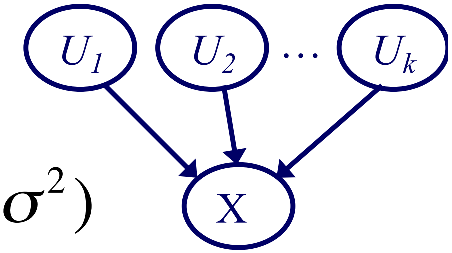
A	*B*	*C*	T	F
T	T	T	0.9	0.1
T	T	F	0.9	0.1
T	F	T	0.9	0.1
T	F	F	0.9	0.1
F	T	T	0.8	0.2
F	T	F	0.5	0.5
F	F	T	0.5	0.5
F	F	F	0.5	0.5

Pr(*D* | *A*, *B*, *C*)



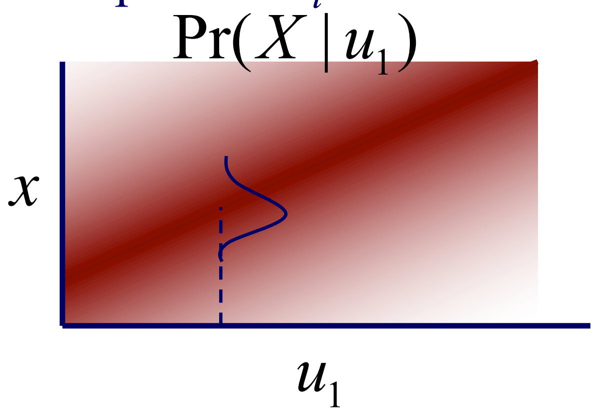
Representing CPDs for Continuous Variables

- we can also model the distribution of continuous variables in Bayesian networks
- one approach: *linear Gaussian models*



$$\Pr(X \mid u_1, \ldots, u_k) \sim N(a_0 + \sum_i a_i \times u_i,\ \ \sigma^2)$$

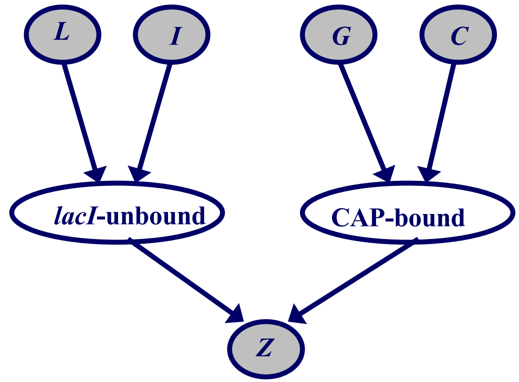
- X normally distributed around a mean that depends linearly on values of its parents u_i



The Inference Task in Bayesian Networks

Given: values for some variables in the network (*evidence*), and a set of *query* variables

L	G	I	C	lacI-unbound	CAP-bound	Z
present	present	present	present	?	?	low



Do: compute the posterior distribution over the query variables

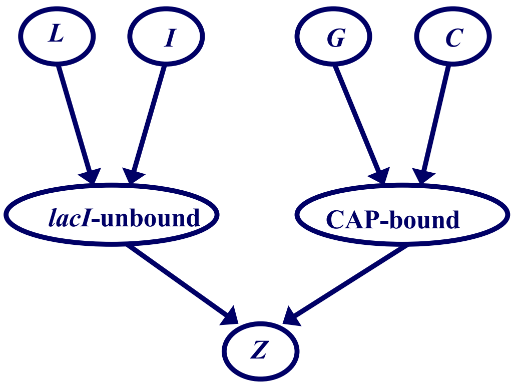
$$P(CAP-bound = true \mid L = present, G = present, I = present, C = present, Z = low)$$

- variables that are neither evidence variables nor query variables are *hidden* variables
- the BN representation is flexible enough that any set can be the evidence variables and any set can be the query variables

The Parameter Estimation Task

- **Given**: a set of training instances, the graph structure of a BN

L	G	I	C	lacI-unbound	CAP-bound	Z
present	present	present	present	true	false	low
present	present	present	present	true	false	absent
absent	present	present	present	false	false	high
			...			



- **Do**: infer the parameters of the CPDs

- this is straightforward when there aren't missing values, hidden variables

The Structure Learning Task

- **Given**: a set of training instances

L	G	I	C	lacI-unbound	CAP-bound	Z
present	present	present	present	true	false	low
present	present	present	present	true	false	absent
absent	present	present	present	false	false	high
			…			

- **Do**: infer the graph structure (and perhaps the parameters of the CPDs too)

Summary

- Bayesian networks provide a compact and visually intuitive representation of joint probability distributions
- Independence statements -> compactness
- Variety of representations for CPDs
- Key computational tasks
  - Inference, parameter estimation, structure learning