CS6462

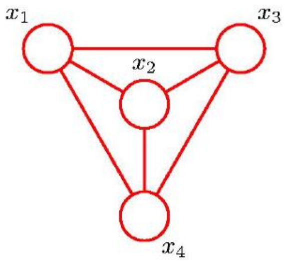
Probabilistic and Explainable AI

Lesson 10
Probabilistic Graph-Based Inference

by Emil Vassev

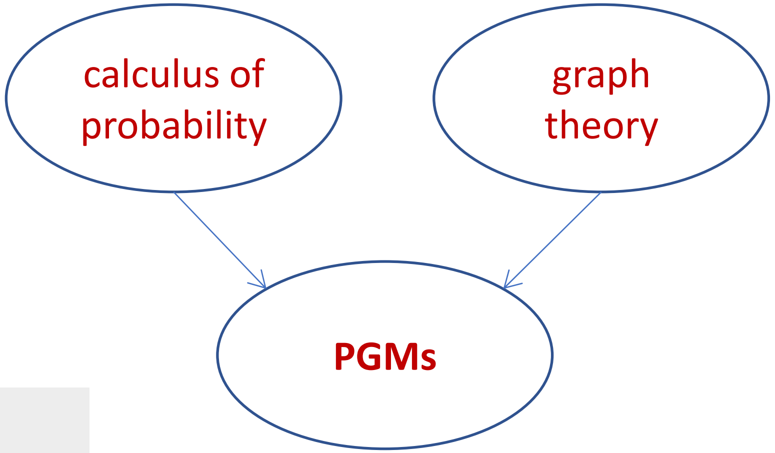
February 10, 2025

Probabilistic Graphical Models



Applications of graphical models:

- machine learning and AI
- computational biology
- statistical signal and image processing
- communication and information theory



Origin and subject:

Fundamental to the idea of a graphical model is the notion of modularity

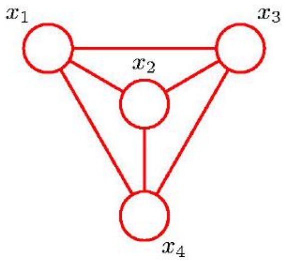
based on correspondences between graph theory and calculus of probability

deal with random variables: relationships and probability distributions

Definition:

PGMs provide a graphical representation of probability distributions addressing dependencies between random variables

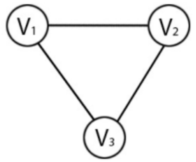
Probabilistic Graphical Models (cont.)



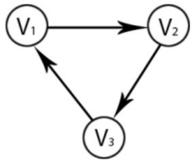
Structure, Elements & Roles:

- A graph contains a set of nodes (vertices) connected by links (edges or arcs)
- PGM - each node represents a random variable, and links represent probabilistic dependencies between random variables
- PGM - specifies the way in which the joint distribution over all random variables decomposes into a product of factors, where each factor depends on a subset of random variables

Undirected Graph



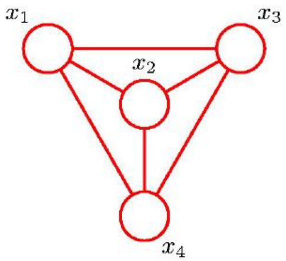
Directed Graph



Types of Graphical Models:

- Undirected Graphical Models (Markov Random Fields)
- Directed Graphical Models (Bayesian networks)
- Hybrid Graphical Models - combine directed and undirected graphical models

Independence Properties



Revision: Independent Events:

- the occurrence of ***F*** has no effect upon the specification of the probability of ***E***
- events ***E*** and ***F*** are independent if: ***P(E) = P(E|F)***

Independent Random Variables:

$X \perp Y$ – notation for independent variables

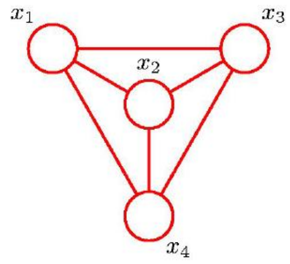
- $P(X=x) = P(X=x|Y=y)$, for $\forall$ $X = \{x_1, \ldots. x_n\}$ and $\forall$ $Y = \{y_1, \ldots. y_k\}$
- $P(X=x \cap Y=y) = P(X=x) * P(Y=y)$, for $\forall$ $X = \{x_1, \ldots. x_n\}$ and $\forall$ $Y = \{y_1, \ldots. y_k\}$

Conditional Independence:

- conditional distribution of ***X*** given ***Y*** and **Z**, is $P(X|Y \cap Z)$
- if $P(X=x|Y=y \cap Z=z)$ does not depend on any value of $Y = \{y_1, \ldots. y_k\}$:

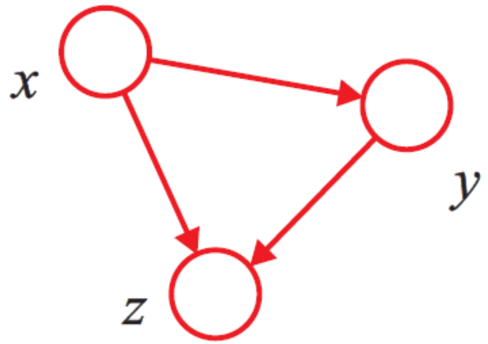
  $P(X=x|Y=y \cap Z=z) = P(X=x|Z=z)$, or ***X*** is conditionally independent of ***Y*** given ***Z***

Directed Graphical Models

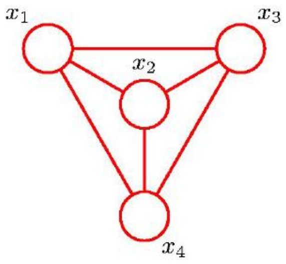


Features:

- *representation*: represented by a graph with its nodes being random variables and directed edges being statistical dependency relationships between them
- *direction of the edges*: determines the influence of one random variable on another
- *causal relationships*: can be used to express causal relationships between random variables
- *Directed Acyclic Graph (DAG)*: graph without cycles (closed chain)
- *Bayesian Networks:* generic name for *probabilistic directional graphs*
  - *DAGs* used to describe probability distributions
  - consider *joint distribution* of variables, e.g, ***X***, ***Y***, ***Z***
  - consider *conditional probability distribution* of variables
  - exploit *conditional independence* of variables



Bayesian Networks



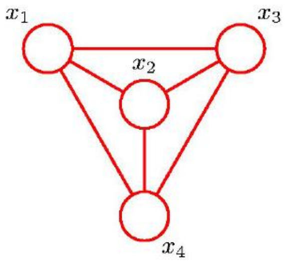
Definition:

- Bayesian Network is a DAG with random variables for vertices representing observable or latent random variables of a model
  - a variable can be continuous or discrete
  - directed edges represent conditional distributions

Distribution models:

- binary values of the vertices - conditional distributions may be Bernoulli
- continuous values of the vertices - conditional distributions may be Gaussian
- joint probability distribution - formulated as a product of conditional or total (marginal) probability distributions

Bayesian Networks – Joint Distribution



Recall:

joint probability distribution - a product of conditional or marginal probabilities

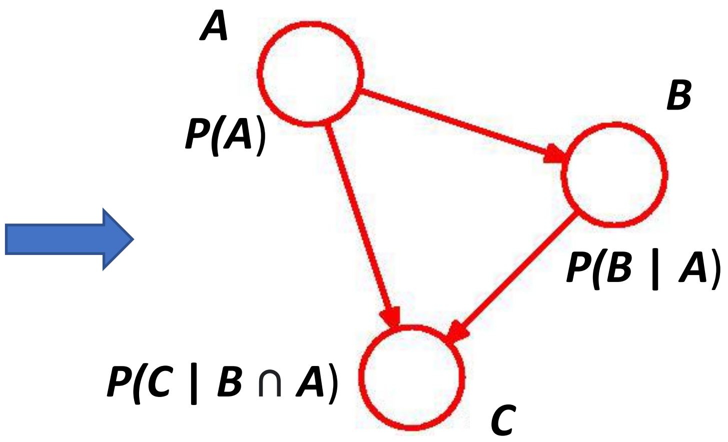
Example:

- an arbitrary joint distribution over three random variables ***A*** ***,B***, and ***C***
- product rule of probability:

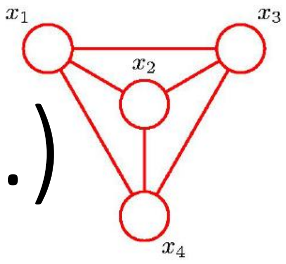
$P(A \cap B \cap C) = P(A) * P(B|A) * P(C \mid B \cap A)$, holds for any value of ***A***, ***B*** and ***C***

- representing the joint distribution in terms of a simple graphical model:

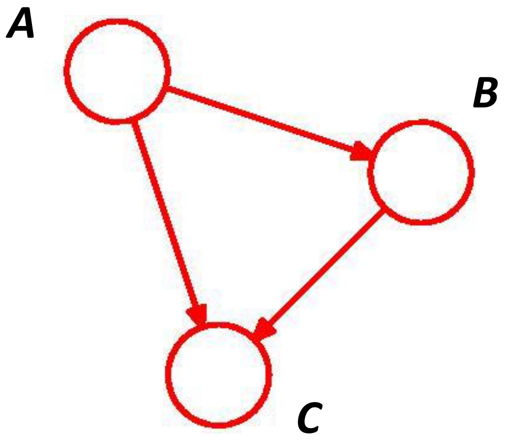
1) introduce a node for each of the random variables
2) associate each node with the corresponding conditional distribution factor from the joint distribution
3) for each conditional distribution we add directed links to the graph from the nodes corresponding to the variables on which the distribution is conditioned



Bayesian Networks – Joint Distribution (cont.)



Features:

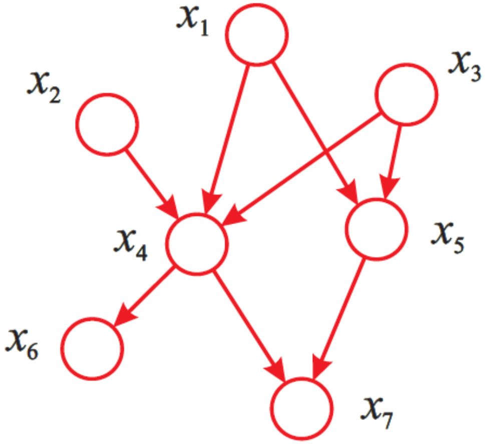


- if there is a link going from node ***A*** to node ***B***:
  - node ***A*** is a parent of node ***B***
  - node ***B*** is a child of node ***A***
- *fully connected graph*: there is a link between all pairs of nodes
- joint distribution over ***k*** random variables factorizes:

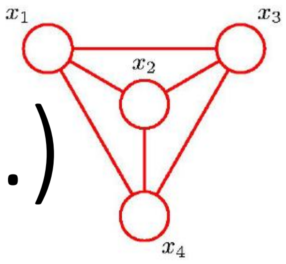
$$P(X_1 \cap X_2 \cap \ldots \cap X_k) = P(X_1) * P(X_1|X_2) * P(X_3|X_1 \cap X_2) * \ldots. * P(X_k|X_1 \cap X_2 \cap \ldots \cap X_{k-1})$$

- *absence of links*: graph is not fully connected

$$P(X_1 \cap X_2 \cap \ldots \cap X_7) = P(X_1)*P(X_2)*P(X_3)*P(X_4|X_1 \cap X_2 \cap X_3) * P(X_5|X_1 \cap X_3)*P(X_6|X_4)*P(X_7|X_4 \cap X_5)$$



Bayesian Networks – Joint Distribution (cont.)

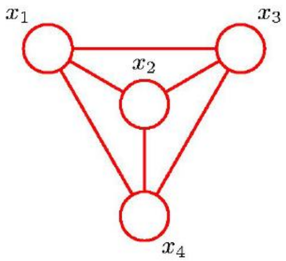


Factorization property:

- joint distribution defined by a graph is given by the product of conditional probability distributions for each node conditioned on its parents:

$$\boldsymbol{P(X_1 \cap X_2 \cap \cdots \cap X_k) = \prod_{i=1}^{k} P(X_i | Parents(X_i))}$$

- $\boldsymbol{Parents(X_i)}$ is a function that returns the joint probability distribution of the parents of node $\boldsymbol{X_i}$
- this equation expresses a key factorization property of the joint distribution for Bayesian Networks



Bayesian Networks – Marginal Distribution

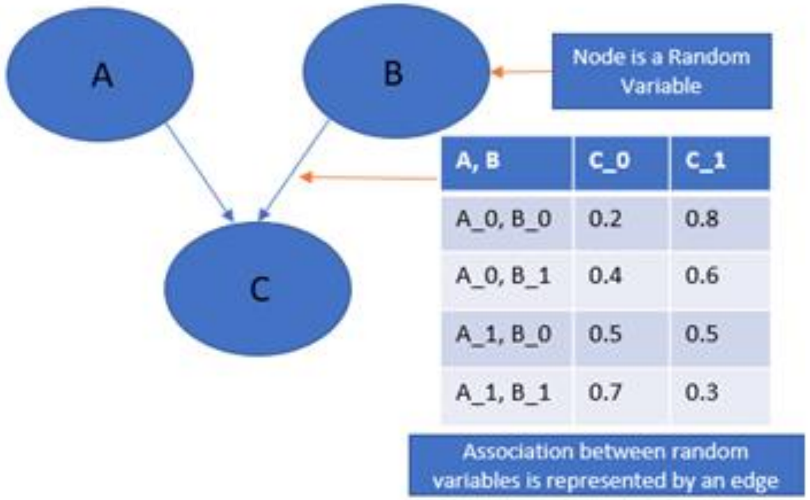
Definition & calculation:

- *definition*: distribution over a single random variable (or a subset of variables) from a larger set of variables, without any reference to an observed set of variables
- *calculation*: the marginal distribution for variable $\boldsymbol{X_i}$ is:

$$\boldsymbol{P(X_i) = \sum_{X_j \neq X_i,} P(X_1 \cap X_2 \cap \cdots \cap X_k)}, \text{ for } \forall\ \boldsymbol{X_j = \{x_1, \ldots. x_n\}}$$

where the sum is over the states of all variables $\boldsymbol{X_j \neq X_i}$ and can be computed by the sum of the products

Example: calculate the marginal distribution for ***C***



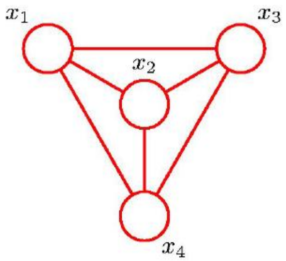
A, B	C_0	C_1
A_0, B_0	0.2	0.8
A_0, B_1	0.4	0.6
A_1, B_0	0.5	0.5
A_1, B_1	0.7	0.3

$$\boldsymbol{P(C) = 0.3 + 0.24 + 0.12 + 0.34 =}\ 1$$

A	B	C	Probability
A_0	B_0	C_0	0.15
A_0	B_0	C_1	0.15
A_0	B_1	C_0	0.08
A_0	B_1	C_1	0.16
A_1	B_0	C_0	0.05
A_1	B_0	C_1	0.07
A_1	B_1	C_0	0.02
A_1	B_1	C_1	0.32

A	B	Probability
A_0	B_0	0.3
A_0	B_1	0.24
A_1	B_0	0.12
A_1	B_1	0.34

Bayesian Network Inference

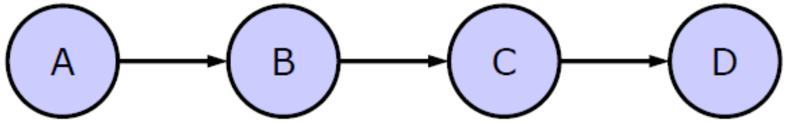


Given a Bayesian network, what queries might we want to ask?

- ***Parents(X) = evidence***
- general question: What's the marginal probability of each node?
- conditional probability query: ***P(X|Parents(X))*** - What is the probability for $\boldsymbol{X}$ for having a particular value $\boldsymbol{x}$ of ∈ $\boldsymbol{X}$**?**
- max posterior probability: How is maximizing the posterior probability ***P(X|evidence)*** equivalent to maximizing the likelihood ***P(evidence|X)*** ? What value of ***X*** maximizes the ***P(evidence|X)***?

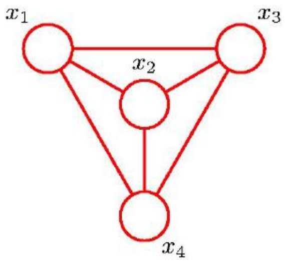
Example:

- Given the joint distribution over three variables, we can answer any question about the probability of a single value held by another variable by summing (or marginalizing) over the first three variables.



$$\boldsymbol{P(d_2)} = \sum_{\boldsymbol{a \in A}} \sum_{\boldsymbol{b \in B}} \sum_{\boldsymbol{c \in C}} \boldsymbol{P(A = a \cap B = b \cap C = c \cap D = d_2)}$$

Undirected Graphical Models

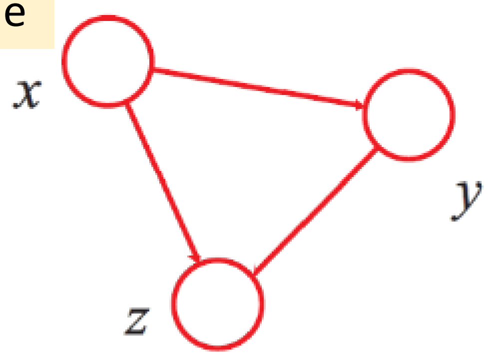


Definition & Features:

- *representation*: represented by a graph with its nodes being random variables and undirected edges being correlation relationships between them
- *Markov Random Field (MRF):* an undirected graph over a set of random variables **$\{X_1, \ldots X_n\}$** is called a Markov Random Field or Markov network
- *Conditional Independence*: independence can be established simply by graph separation: if every path from a node in ***X*** to a node in ***Z*** goes through a node ***Y***, we conclude that $\boldsymbol{X \perp Z | Y}$
- *Cliques:* sets of nodes that are fully connected
- *Joint distribution*: product of non-negative functions over the cliques of the graph

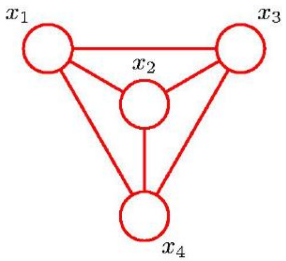
$$\boldsymbol{P(X_1 \cap X_2 \cap \cdots \cap X_k) = \frac{1}{Z} \prod_{S} \varphi_S(X_S)}$$

X, Y, Z are a clique



X and ***Z*** are not independent

Defining Graphical Models with FOPPL



Definition:*

1) graphical model ***G*** as a tuple ***G=(V, A, P, Y)***:

- ***V*** - a set of vertices that represent random variables
- ***A = V × V*** - a set of directed edges that represent conditional dependencies between variables
- ***P*** - a map mapping vertices to deterministic expressions that specify a probability density or mass function for each random variable
- ***Y*** - a partial map that for each observed random variable contains a deterministic expression ***E*** for the observed value

2) a set of translation rules used to compile any FOPPL program to a graphical model

```
(let [z (sample (bernoulli 0.5))
      mu (if (= z 0) -1.0 1.0)
      d (normal mu 1.0)
      y 0.5]
  (observe d y)
  z)
```

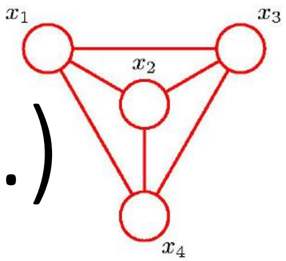
Defines a joint distribution $p(y = 0.5, z)$:

1) samples z from a Bernoulli distribution
2) sets a likelihood parameter μ to −1.0 or 1.0
3) observes a value y = 0.5 from a normal distribution with mean μ

inference problem: characterize the posterior distribution $p(z \mid y)$

** "An Introduction to Probabilistic Programming", book by J.W. van de Meent, B. Paige, H. Yang, F. Wood]*

Defining Graphical Models with FOPPL (cont.)



FOPPL Translation to Graphical Model:

Graphical Model ***G=(V, A, P, Y)*** :

- Vertices ***V*** contains two variables ***{z, y}***
- Arcs ***A*** contains a single pair ***(z, y)*** to mark the conditional dependence relationship between these two variables
- Map ***P****:*
  - the probability mass for ***z*** is defined as the target language expression (***Pbern z 0.5***)
  - the probability density function for *y* is defined using ***Pnorm***, which implements the probability density function for the normal distribution
- Map ***Y*** - holds the observed value for **y**

```
(let [z (sample (bernoulli 0.5))
      mu (if (= z 0) -1.0 1.0)
      d (normal mu 1.0)
      y 0.5]
  (observe d y)
  z)
```

$$V = \{z, y\},$$

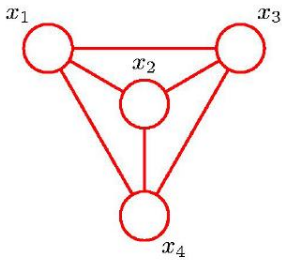
$$A = \{(z, y)\},$$

$$\mathcal{P} = [z \mapsto (p_{\text{bern}}\ z\ 0.5),$$

$$y \mapsto (p_{\text{norm}}\ y\ (\texttt{if}\ (=\ z\ 0)\ \texttt{-1.0 1.0})\ \texttt{1.0})],$$

$$\mathcal{Y} = [y \mapsto 0.5]$$

$$E = z$$



Summary

Probabilistic Graphical Models:

- A graph contains a set of nodes (vertices) connected by links (edges or arcs)
- PGM - each node represents a random variable, and links represent probabilistic dependencies between random variables

Directed Graphical Models – Bayesian Networks

Inference with Bayesian Networks

Undirected Graphical Models - Markov Random Fields

Defining Graphical Models with FOPPL

Next Lesson – FOPPL in Python

Thank You!

Questions?