Sets of Independencies

Let $I(G)$ be the set of independence assertions according to graph $G$ (e.g., $A \perp D | B, C$).

Let $I(P)$ be the set of independence assertions that hold in a probability distribution $P$.

1. We say that $G$ is an I-map for $P$ iff:
   \[ I(G) \subseteq I(P) \]
2. If no edge can be removed from $G$ while remaining an I-map, then $G$ is a minimal I-map for $P$.
3. We say that $G$ is a P-map for $P$ iff:
   \[ I(G) = I(P) \]

I-Equivalence

- We say that two Bayesian network graphs are I-equivalent if they specify the same set of independencies:
  \[ I(G) = I(G') \]
- For example:

I-Equivalence Conditions

- **Sufficient conditions:**
  Two graphs are I-equivalent if they have the same undirected skeleton and v-structures.

- **Necessary conditions:**
  Two graphs are I-equivalent if they have the same undirected skeleton and immoralities.
  
  - An immorality is a v-structure with no edge connecting the parents.
  
    (That is, a child with unmarried parents.)

Exercise

- Are these two networks I-equivalent?

- Are these two networks I-equivalent?

Why does I-equivalence matter?

- Some probability distributions can be represented equally well by several different Bayesian networks.

- If $P$ factorizes over $G$ and $I(G) = I(G')$, then $P$ factorizes over $G'$. 
Sets of Independencies

Let \( I(G) \) be the set of independence assertions according to graph \( G \) (e.g., \( A \perp D \mid B, C \)).
Let \( I(P) \) be the set of independence assertions that hold in a probability distribution \( P \).
1. We say that \( G \) is an I-map for \( P \) iff:
   \[ I(G) \subseteq I(P) \]
2. If no edge can be removed from \( G \) while remaining an I-map, then \( G \) is a minimal I-map for \( P \).
3. We say that \( G \) is a P-map for \( P \) iff:
   \[ I(G) = I(P) \]

Example: Misconception

Alice studies with Bob and Debbie; Bob studies with Alice and Charles; Charles studies with Bob and Debbie; Debbie studies with Alice and Charles.
Communication between study partners may cause misconception to spread.
1. What are the independencies?
2. How can we capture them graphically?

Bayes Net 1

Conditional independencies:
1. \( A \perp C \mid B, D \)
2. \( B \perp D \mid A, C \)

Bayes Net 2

Conditional independencies:
1. \( A \perp C \mid B, D \)
2. \( B \perp D \mid A, C \)

Bayes Net 3

Conditional independencies:
1. \( A \perp C \mid B, D \)
2. \( B \perp D \mid A, C \)

Markov Network

Conditional independencies:
1. \( A \perp C \mid B, D \)
2. \( B \perp D \mid A, C \)
Warm-Up

Write down the factorization entailed by this Markov network.

Is A independent of H given E, G, and B?

Gibbs distributions / Markov networks

A distribution \( P_\Phi \) is a Gibbs distribution parameterized by a set of factors \( \Phi = \{ \phi_1(D_1), \ldots, \phi_k(D_k) \} \)

if it is defined as follows:

\[
P_\Phi(X_1, \ldots, X_n) = \frac{1}{Z} \tilde{P}_\Phi(X_1, \ldots, X_n)
\]

where \( \tilde{P}_\Phi = \phi_1(D_1)\phi_2(D_2)\cdots\phi_k(D_k) \)

and \( Z = \sum_{x_1, \ldots, x_n} \tilde{P}_\Phi(x_1, \ldots, x_n) \)

A Gibbs distribution factorizes over a Markov network \( H \) if there is a clique in \( H \) for each \( D_k \).

Hammersley-Clifford Theorem

If Distribution is strictly positive (\( P(x) > 0 \))

And Graph \( H \) encodes conditional independences

Then Distribution is product of potentials over cliques of graph (P “factorizes” over \( H \))

Inverse is also true.

Markov networks: Definition

- A Markov network (aka Markov Random Field, Gibbs distribution) consists of an undirected graph and a set of potential functions or factors defined over cliques in the graph.
- Conditional independence is determined by graph separation.
- Probability distribution is a product of factors, e.g.,

\[
P(A,B,C,D) = \frac{1}{Z} \phi(A,B)\phi(B,C)\phi(C,D)\phi(A,D)
\]

\[
P(X_1, \ldots, X_n) = \frac{1}{Z} \prod_{k=1}^n \phi(D_k)
\]

- \( Z \) is a normalization constant, called the partition function:

\[
Z = \sum_{x_1, \ldots, x_n} \phi(D_k)
\]
From BNs to MNs

How do we convert a BN to an MN?

Moralization

To convert a Bayesian network into a Markov network:
- For each variable:
  Add arcs between its parents ("marry" them)
- Remove arrows
- Reuse the original conditional probability distributions as potential functions

From MNs to BNs

• On a piece of paper, please write down (semi-anonymously):
  – What’s going well for you in this class
  – What’s not going so well for you in this class

• Then answer the following questions:

How many BN structures are I-equivalent to each of the following MN graphs:
Markov Nets vs. Bayes Nets

<table>
<thead>
<tr>
<th>Property</th>
<th>Markov Nets</th>
<th>Bayes Nets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form</td>
<td>Prod. potentials</td>
<td>Prod. potentials</td>
</tr>
<tr>
<td>Potentials</td>
<td>Arbitrary</td>
<td>Cond. probabilities</td>
</tr>
<tr>
<td>Cycles</td>
<td>Allowed</td>
<td>Forbidden</td>
</tr>
<tr>
<td>Partition func.</td>
<td>Z = ?</td>
<td>Z = 1</td>
</tr>
<tr>
<td>Indep. check</td>
<td>Graph separation</td>
<td>D-separation</td>
</tr>
<tr>
<td>Indep. props.</td>
<td>Some</td>
<td>Some</td>
</tr>
<tr>
<td>Inference</td>
<td>MCMC, BP, etc.</td>
<td>Convert to Markov</td>
</tr>
</tbody>
</table>

Social Network Analysis

- People who are connected in a social network are more likely to have similar preferences and behaviors.
- E.g., Smokers are likely to be friends with other smokers.

![Social Network Example](image)

Fraud Detection

On an auction site, people committing fraud (red triangles) work together with each other and accomplish (yellow diamonds) to boost their reputation.

![Fraud Detection Example](image)

Another MN Example

**Goal:** Determine which IP addresses correspond to machines at a particular physical location.
- True = Located at a given address
- False = Not located at a given address

**Data:**
- Traceroutes:
- IP co-location
- Similar DNS names
- Some ground truth

![Another MN Example](image)

Markov networks: Applications

- Statistical physics
- Vision / Image processing
- Social networks
- Web page classification
- Binary code analysis
- Computational biology
- Etc.

Markov networks: Different factorizations, same graph

- $P(A,B,C) = \frac{1}{Z} \phi(A,B,C)$
- $P(A,B,C) = \frac{1}{Z} \phi_{AC}(AB) \phi_{BC}(AC)$
- $P(A,B,C) = \frac{1}{Z} \phi_{AC}(ABC) \phi_{BC}(AB) \phi_{AC}(C)$
Factor graphs:
Different factorizations, different graphs

\[ P(A, B, C) = \frac{1}{Z} \phi_{AB}(ABC) \]

\[ P(A, B, C) = \frac{1}{Z} \phi_{AB}(AB) \phi_{AC}(AC) \phi_{BC}(BC) \]

\[ P(A, B, C) = \frac{1}{Z} \phi_{ABC}(ABC) \phi_{AB}(AB) \phi_{C}(C) \]

Factor Graphs

A factor graph is a bipartite graph with a node for each random variable and each factor.
There is an edge between a factor and each variable that participates in that factor.

Log-linear Models

Log-factors can be represented as weighted functions, sometimes much more compactly:

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>( \phi(A, B) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>True</td>
<td>10</td>
</tr>
<tr>
<td>True</td>
<td>False</td>
<td>1</td>
</tr>
<tr>
<td>False</td>
<td>True</td>
<td>1</td>
</tr>
<tr>
<td>False</td>
<td>False</td>
<td>10</td>
</tr>
</tbody>
</table>

\[ \log \]

\[ (\log 10) \, f_{\text{AB}}(A, B) + (\log 1) \, f_{\text{A}}(A) + (\log 1) \, f_{\text{B}}(B) + (\log 10) \, f_{\text{D}}(D) \]

\[ (\log 10) \, f_{\text{AB}}(A, B) \]

Log-linear Models

Product of factors can be represented as an exponentiated sum of log-factors:

\[ P(X_1, \ldots, X_n) = \frac{1}{Z} \prod \phi(D_i) \]

\[ \log P(X_1, \ldots, X_n) = -\log Z + \sum \log \phi_i(D_i) \]

\[ P(X_1, \ldots, X_n) = \frac{1}{Z} \exp \left( \sum \log \phi_i(D_i) \right) \]

Log-linear Models

Putting it all together...

\[ P(X_1, \ldots, X_n) = \frac{1}{Z} \exp \left( \sum w_i f_i(D_i) \right) \]

\[ = \frac{1}{Z} \exp \left( \mathbf{w} \cdot \mathbf{\psi}(X) \right) \]
MN Independences

- **Global independence, \( I(G) \):**
  X and Y are conditionally independent given Z if all paths between X and Y go through Z.

- **Pairwise, \( I_P(G) \):**
  If there is no edge between X and Y, then they are conditionally independent given all other variables.

- **Markov blanket, \( I_l(H) \):**
  X is independent of all other variables given its neighbors in the graph.

Equivalent for positive distributions!
We can use any of these to construct MN graphs!

The Problem with Non-Positive Distributions

Consider \( P(A,B,C,D) \), which has probability 1/8 for the following configurations:

- \((0,0,0,0)\)
- \((1,0,0,0)\)
- \((1,1,0,0)\)
- \((1,1,1,0)\)
- \((0,0,0,1)\)
- \((0,0,1,1)\)
- \((0,1,1,1)\)
- \((1,1,1,1)\)

\( P \) satisfies the independencies of the MN graph:

However, \( P \) does not factorize over this graph.

Implementing Factor Operations

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>Φ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0.8</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>Φ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0.8</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.7</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Scope: [2, 3]
Stride: [1, 2]
Values: [0.5, 0.8, 0.1, 0.3, 0.9]

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>Φ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0.7</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Scope: [3, 2]
Stride: [1, 2]
Values: [0.5, 0.7, 0.1, 0.2]

Inference Queries

Let \( E \) be evidence variables, \( X \) be query variables, and \( H \) be other variables.

- **Probability of evidence:** \( P(e) = \Sigma_h P(e,x,h) \)
- **Conditional probability:** \( P(x|e) = P(x,e) / P(e) \)
- **Joint conditional probability:**
  \( P(X|e) = \Sigma_h P(X,e)/P(e) \)

- **Conditional marginals:**
  \( P(X_i|e) = \Sigma_{h,x_{-i}} P(X_i, e, h, x_{-i})/P(e) \)
- **Maximum a posteriori (MAP):**
  \( \text{argmax}_x P(x|e) = \text{argmax}_x \Sigma_h P(x,h,e) \)