I-Equivalence and Markov Networks

Daniel Lowd
CIS 473/573

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:
   \[ I(G) = I(P) \]
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 ⊥ C | B, D
2. B ⊥ D | A, C

Markov Network

Conditional independencies:
1. A ⊥ C | B, D
2. B ⊥ D | A, C
Warm-Up

Write down the factorization entailed by this Markov network.

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

Misconception Example

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>$\phi_1(A,B)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a^0$</td>
<td>$b^0$</td>
<td>30</td>
</tr>
<tr>
<td>$a^0$</td>
<td>$b^1$</td>
<td>5</td>
</tr>
<tr>
<td>$a^1$</td>
<td>$b^0$</td>
<td>1</td>
</tr>
<tr>
<td>$a^1$</td>
<td>$b^1$</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A</th>
<th>D</th>
<th>$\phi_3(A,D)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a^0$</td>
<td>$d^0$</td>
<td>100</td>
</tr>
<tr>
<td>$a^0$</td>
<td>$d^1$</td>
<td>1</td>
</tr>
<tr>
<td>$a^1$</td>
<td>$d^0$</td>
<td>1</td>
</tr>
<tr>
<td>$a^1$</td>
<td>$d^1$</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B</th>
<th>C</th>
<th>$\phi_2(B,C)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b^0$</td>
<td>$c^0$</td>
<td>100</td>
</tr>
<tr>
<td>$b^0$</td>
<td>$c^1$</td>
<td>1</td>
</tr>
<tr>
<td>$b^1$</td>
<td>$c^0$</td>
<td>1</td>
</tr>
<tr>
<td>$b^1$</td>
<td>$c^1$</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C</th>
<th>D</th>
<th>$\phi_4(C,D)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c^0$</td>
<td>$d^0$</td>
<td>1</td>
</tr>
<tr>
<td>$c^0$</td>
<td>$d^1$</td>
<td>100</td>
</tr>
<tr>
<td>$c^1$</td>
<td>$d^0$</td>
<td>100</td>
</tr>
<tr>
<td>$c^1$</td>
<td>$d^1$</td>
<td>1</td>
</tr>
</tbody>
</table>
### 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.,

\[
P(A,B,C,D) = \frac{1}{Z} \phi_1(A,B) \phi_2(B,C) \phi_3(C,D) \phi_4(A,D)
\]

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

- **Z** is a normalization constant, called the **partition function**:

\[
Z = \sum_X \prod_i \phi_i(D_i)
\]
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.
From BNs to MNs

How do we convert a BN to an MN?

E  B

A

J  M

From BNs to MNs

How do we convert a BN to an MN?

E  B

A

J  M

E  B

A

J  M
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
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:

- A → B
- D → C

- A → B
- D → C

- A → B
- D → C
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>

Markov networks: Applications

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

Fraud Detection

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

Markov networks:
Different factorizations, same graph

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

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

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

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

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

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

\[
P(X_1, \ldots, X_n) = \frac{1}{Z} \exp \left( \sum_i \log \phi_i(D_i) \right)
\]
Log-linear Models

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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>( \psi_i(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) \( f_1(A, B) \) 
+ (log 1) \( f_2(A, B) \) 
+ (log 1) \( f_3(A, B) \) 
+ (log 10) \( f_4(A, B) \) 

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

Log-linear Models

Putting it all together...

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

\[ = \frac{1}{Z} \exp \left( w \cdot \psi(X) \right) \]
**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:

![MN Graph]

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

<table>
<thead>
<tr>
<th>X₁</th>
<th>X₂</th>
<th>Φ₁</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>X₂</th>
<th>X₃</th>
<th>Φ₂</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: [2,1]
Stride: [1,2]
Values: [0.5, 0.8, 0.1, 0, 0.3, 0.9]

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

Scope: list of variables included in this factor
Stride: index increment for each variable
Values: 1D array of all table entries
(Page 358, Box 10.A)

Inference Queries

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

• Probability of evidence: \( P(e) = \sum_{h,x} P(e,x,h) \)
• Conditional probability: \( P(x|e) = \frac{P(x,e)}{P(e)} \)
• Joint conditional probability:
  \( P(X|e) = \sum_h P(X,e)/P(e) \)
• Conditional marginals:
  \( P(X_i|e) = \sum_{h,x-i} P(X_i, e, h, x_i)/P(e) \)
• Maximum a posteriori (MAP):
  \( \arg\max_x P(x|e) = \arg\max_x \sum_h P(x,h,e) \)