Perfect maps (P-maps)

- I-maps are not unique and often not simple enough

- Define “simplest” $G$ that is an I-map for $P$
  - A BN structure $G$ is a perfect map for a distribution $P$ if $I(P) = I(G)$

- Our goal:
  - Find a perfect map!
  - Must address equivalent BNs
Inexistence of P-maps 1

- XOR (this is a hint for the homework)
  
  \[
  A = B \oplus C \\
  A \perp B \\
  B \perp C \\
  C \perp A \\
  \begin{align*}
  \mathcal{I} &:\quad A \\ 
  \text{Minimal} &:\quad A \perp B \\ 
  \not\exists &:\quad B \perp C \\
  \end{align*}
  
  \begin{align*}
  \text{P-MAP} &:\quad ? \\
  \text{Extra credit} &:\quad \\
  \end{align*}

Obtaining a P-map

- Given the independence assertions that are true for \( P \)

- Assume that there exists a perfect map \( G^* \)
  - Want to find \( G^* \)

- Many structures may encode same independencies as \( G^* \), when are we done?
  - Find all equivalent structures simultaneously!
I-Equivalence

- Two graphs $G_1$ and $G_2$ are **I-equivalent** if $I(G_1) = I(G_2)$

**Equivalence class** of BN structures
- Mutually-exclusive and exhaustive partition of graphs

- How do we characterize these equivalence classes?

Skeleton of a BN

- **Skeleton** of a BN structure $G$ is an *undirected graph* over the same variables that has an edge $X \rightarrow Y$ for every $X \rightarrow Y$ or $Y \rightarrow X$ in $G$

- (*Little*) Lemma: Two I-equivalent BN structures must have the same skeleton

Counter example
What about V-structures?

- **V-structures are key property of BN structure**

- **Theorem**: If $G_1$ and $G_2$ have the same skeleton and V-structures, then $G_1$ and $G_2$ are I-equivalent

  \[
  \text{not if and only if}
  \]

Same V-structures not necessary

- **Theorem**: If $G_1$ and $G_2$ have the same skeleton and V-structures, then $G_1$ and $G_2$ are I-equivalent

- Though sufficient, same V-structures not necessary
Immoralities & I-Equivalence

- Key concept not V-structures, but “immoralities” (unmarried parents 😊)
  - $X \rightarrow Z \leftarrow Y$, with no arrow between $X$ and $Y$
  - Important pattern: $X$ and $Y$ independent given their parents, but not given $Z$
  - (If edge exists between $X$ and $Y$, we have covered the V-structure)

- **Theorem:** $G_1$ and $G_2$ have the same skeleton and immoralities if and only if $G_1$ and $G_2$ are I-equivalent

Obtaining a P-map

- Given the independence assertions that are true for $P$
  - Obtain skeleton
  - Obtain immoralities

- From skeleton and immoralities, obtain every (and any) BN structure from the equivalence class
Measuring Independence

- Many ways to estimate from data: not always easy, but doable.
- A very simple approach (related to MLE): 
  \[(X \perp Y \mid Z) \iff I(X, Y \mid Z) = 0\]

\[I(X, Y \mid Z) = \sum_{x,y,z} P(x,y,z) \log \frac{P(x,y,z)}{P(x \mid z) P(y \mid z)} \]

Data: don't have \(P(x,y,z)\), estimate MLE \(\hat{P}(x,y,z) = \frac{\text{count}(x,y,z)}{m}\)

In practice: \(I(X, Y \mid Z) > 0\)

Independent enough when \(I(X, Y \mid Z) < \epsilon\)

Identifying the skeleton 1

- When is there an edge between X and Y?
  - Is it \(X \perp Y\)? Yes: \(X \rightarrow Z \rightarrow Y\) \(X \perp Y\)
  - Everything else? \(X \rightarrow Z \rightarrow Y\) but no edge \(X \rightarrow Y\)

- When is there no edge between X and Y?
  - X is not parent of Y & vice versa.
  - Local Markov assumption
    \[\exists U \subseteq X \setminus \{X,Y\}, \text{ such that } X \perp Y \mid U\]
    - Additional assumption: \# parents \(\leq d\)
    \[\exists U \subseteq X \setminus \{X,Y\}, \text{ such that } X \perp Y \mid U\]
Identifying the skeleton 2

- Assume $d$ is max number of parents (d could be $n$)

- For each $X_i$ and $X_j$
  - $E_{ij} \leftarrow \text{true}$
  - For each $U \subseteq X - \{X_i, X_j\}$, $|U| \leq d$
    - Is $(X_i \perp X_j \mid U)$?
      - $E_{ij} \leftarrow \text{false}$
    - If $E_{ij}$ is true
      - Add edge $X - Y$ to skeleton

Identifying immoralities

- Consider $X - Z - Y$ in skeleton, when should it be an immorality?
  - When $X \perp Y$ 
    - $X \rightarrow Z \leftarrow Y$ (immorality)
      - But no need to make $X \perp Y \mid U$ since $Z$ is a parent of $Y$
  - Must be $X \rightarrow Z \leftarrow Y$ (immorality):
    - When $X$ and $Y$ are never independent given $U$, if $Z \in U$, $\exists U \subseteq X - \{X, Y\}, Z \in U$, $X \perp Y \mid U$ (must not have $d$ parents or most $d$)
    - Must not be $X \rightarrow Z \leftarrow Y$ (not immorality):
      - When there exists $U$ with $Z \in U$, such that $X$ and $Y$ are independent given $U$
From immoralities and skeleton to BN structures

- Representing BN equivalence class as a partially-directed acyclic graph (PDAG)

- Immoralities force direction on some other BN edges

- Full (polynomial-time) procedure described in reading

What you need to know

- Minimal I-map
  - every $P$ has one, but usually many

- Perfect map
  - better choice for BN structure
  - not every $P$ has one
  - can find one (if it exists) by considering I-equivalence
  - Two structures are I-equivalent if they have same skeleton and immoralities
Announcements

- Recitation tomorrow
  - Don’t miss it!
- No class on Monday 😊

```
Everything so far Chapter 3
Now = Chapter 16
```
Thumbtack – Binomial Distribution

- P(Heads) = \theta, P(Tails) = 1-\theta

- Flips are i.i.d.:
  - Independent events
  - Identically distributed according to Binomial distribution

- Sequence $D$ of $\alpha_H$ Heads and $\alpha_T$ Tails
  $$P(D \mid \theta) = \theta^{\alpha_H} (1 - \theta)^{\alpha_T}$$

Maximum Likelihood Estimation

- **Data:** Observed set $D$ of $\alpha_H$ Heads and $\alpha_T$ Tails
- **Hypothesis:** Binomial distribution
- Learning $\theta$ is an optimization problem
  - What’s the objective function?
  - $\hat{\theta}_{MLE} = \arg \max_{\theta} P(D \mid \theta)$
- MLE: Choose $\theta$ that maximizes the probability of observed data:
  $$\hat{\theta}_{MLE} = \arg \max_{\theta} P(D \mid \theta) = \arg \max_{\theta} \ln P(D \mid \theta)$$
Your first learning algorithm

\[ \hat{\theta} = \arg \max_{\theta} \ln P(D | \theta) \]
\[ = \arg \max_{\theta} \ln \theta^{\alpha_H} (1 - \theta)^{\alpha_T} \]
\[ = \frac{\alpha_H}{\alpha_H + \alpha_T} \ln \theta + \frac{\alpha_T}{\alpha_H + \alpha_T} \ln (1 - \theta) \]

Set derivative to zero:

\[ \frac{d}{d\theta} \ln P(D | \theta) = 0 \]

One binary node in this BN

Learning Bayes nets

<table>
<thead>
<tr>
<th></th>
<th>Known structure</th>
<th>Unknown structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fully observable data</td>
<td>Easy</td>
<td>Hard structure learning</td>
</tr>
<tr>
<td>Missing data</td>
<td>Hard (EM)</td>
<td>Very hard later in the semester</td>
</tr>
</tbody>
</table>

Data

\[ X^{(1)} \]
\[ \ldots \]
\[ X^{(m)} \]

CPTs –

\[ P(X_i | \text{Pa}_{X_i}) \]

structure

parameters
Learning the CPTs

For each discrete variable $X_i$

$P(x_i) = \frac{\text{Count}(X_i = x_i)}{\text{Count}(x_i)}$

$P(x_i | P_{x_i}) = P(x_i | U)$

$\hat{P}_{\text{MLE}}(x_i | U) = \frac{\text{Count}(x_i, U = u)}{\text{Count}(U = u)}$

Why??

MLE: $P(X_i = x_i | X_j = x_j) = \frac{\text{Count}(X_i = x_i, X_j = x_j)}{\text{Count}(X_j = x_j)}$

Learning the CPTs

For each discrete variable $X_i$

MLE: $P(X_i = x_i | X_j = x_j) = \frac{\text{Count}(X_i = x_i, X_j = x_j)}{\text{Count}(X_j = x_j)}$

WHY????????????

if only one var
then take derivative, set to 0
all is good
Maximum likelihood estimation (MLE) of BN parameters – example

Given structure, log likelihood of data:

\[
\log P(D | \theta_G, G) = \log \prod_{i=1}^{n} P(x^{(i)} | \theta_{G_i}, G) = \sum_{i=1}^{n} \log P(x^{(i)} | \theta_{G_i}, G)
\]

\[
= \sum_{i=1}^{n} \left( \sum_{j=1}^{m} \log P(f^{(j)} | \theta_{G_i}, G) + \log P(d^{(j)} | \theta_{G_i}, G) + \log P(s^{(j)} | \theta_{G_i}, G) \right) + \sum_{i=1}^{n} \log P(o^{(i)} | \theta_{G_i}, G)
\]

Break up problem into independent subproblems: one for each CPT