Markov Equivalence

SS 2008 – Bayesian Networks

Multimedia Computing, Universität Augsburg
Rainer.Lienhart@informatik.uni-augsburg.de
www.multimedia-computing.{de,org}
Markov Equivalence

Different DAGs can have the same d-separations --> “equivalent”.

All and only d-separations: $I_G(\{Y\}, \{Z\} \mid \{X\})$

$I_G(\{X\}, \{W\} \mid \{Y,Z\})$
Markov Equivalence (1)

Definition 2.7: Let $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ be two DAGs containing the same set of nodes $V$. Then $G_1$ and $G_2$ are called Markov equivalent if for every three mutually disjoint subsets $A, B, C \subseteq V$, $A$ and $B$ are d-separated by $C$ in $G_1$ iff $A$ and $B$ are d-separated by $C$ in $G_2$. That is

$$I_{G_1}(A, B|C) \iff I_{G_2}(A, B|C)$$

Theorem 2.3: Two DAGs are Markov equivalent iff, based on the Markov condition, they entail the same conditional independencies.

Corollary 2.1: Let $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ be two DAGs containing the same set of random variables $V$. Then $G_1$ and $G_2$ are called Markov equivalent iff for every probability distribution $P$ of $V$, $(G_1, P)$ satisfies the Markov condition iff $(G_2, P)$ satisfies the Markov condition.
Markov Equivalence (2)

**Lemma 2.4**: Let $G = (V, E)$ be a DAG and $X, Y \in V$. Then $X$ and $Y$ are adjacent in $G$ iff they are not d-separated by some set in $G$.

**Corollary 2.2**: Let $G = (V, E)$ be a DAG and $X, Y \in V$. Then $X$ and $Y$ are d-separated by some set, they are d-separated either by the set consisting of the parents of $X$ or the set consisting of the parents of $Y$.

**Lemma 2.5**: Suppose we have a DAG $G = (V, E)$ and an uncoupled meeting $X \leftarrow Z \rightarrow Y$. Then the following are equivalent:
1. $X \leftarrow Z \rightarrow Y$ is a head-to-head meeting
2. There exists a set not containing $Z$ that d-separated $X$ and $Y$.
3. All the sets containing $Z$ do not d-separated $X$ and $Y$.

**Lemma 2.6**: If $G_1$ and $G_2$ are Markov equivalent, then $X$ and $Y$ are adjacent in $G_1$ iff they are adjacent in $G_2$. That is, Markov equivalent DAGs have the same links (edges without regard for direction).
Theorem 2.4: Two DAGs $G_1$ and $G_2$ are Markov equivalent iff they have the same links (edges without regard for direction) and the same set of uncoupled head-to-head meetings.

Markov Equivalent

Not Markov Equivalent

Search for uncoupled head-to-head meetings!
**Theorem 2.4**

**Theorem 2.4:** Two DAGs $G_1$ and $G_2$ are Markov equivalent iff they have the same links (edges without regard for direction) and the same set of uncoupled head-to-head meetings.

![Markov Equivalent DAGs](image)

- uncoupled head-to-head meetings!

**Markov Equivalent**

**Not Markov Equivalent**
A Markov equivalence class can be represented by a graph with
1. Same link and
2. Same uncoupled head-to-head meetings
Any assignment of directions to the undirected edges in this graph, that does not create a new uncoupled head-to-head meeting or a directed cycle, yields a member of the equivalence class.

**DAG pattern for a Markov equivalence class** = graph that has the same links as the DAGs in the equivalence class and has oriented all and only the edges common to all of the DAGs in the equivalence class. The directed links in a DAG pattern are called **compelled edges**.
Definition 2.8: Let $gp$ be a DAG pattern whose nodes are the elements of $V$, and $A$, $B$, and $C$ let be mutually disjoint subsets of $V$. We say $A$ and $B$ are d-separated by $C$ in $gp$ if $A$ and $B$ are d-separated by $C$ in any (and therefore every) DAG $G$ in the Markov equivalence class represented by $gp$.

Lemma 2.7: Let $gp$ be a DAG pattern and $X$ and $Y$ be nodes in $gp$. Then $X$ and $Y$ are adjacent in $gp$ iff they are not d-separated by some set in $gp$.

Lemma 2.8: Suppose we have a DAG pattern $gp$ and an uncoupled meeting $X—Z—Y$. Then the following are equivalent:
1. $X—Z—Y$ is a head-to-head meeting
2. There exists a set not containing $Z$ that d-separated $X$ and $Y$.
3. All the sets containing $Z$ do not d-separated $X$ and $Y.$
Entailing Dependencies
Example 2.7 (1)

Assign 1/13 to each object $\Rightarrow I_P(\{V\},\{S\}\mid\{C\})$

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>Outcomes Mapped to this Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>V</td>
<td>$v1$</td>
<td>All objects containing a “1”</td>
</tr>
<tr>
<td></td>
<td>$v2$</td>
<td>All objects containing a “2”</td>
</tr>
<tr>
<td>S</td>
<td>$s1$</td>
<td>All square objects</td>
</tr>
<tr>
<td></td>
<td>$s2$</td>
<td>All round objects</td>
</tr>
<tr>
<td>C</td>
<td>$c1$</td>
<td>All black objects</td>
</tr>
<tr>
<td></td>
<td>$c2$</td>
<td>All white objects</td>
</tr>
</tbody>
</table>
Example 2.7 (2)

Assign 1/13 to each object $\Rightarrow I_{P}(\{V\}, \{S\} | \{C\})$

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>Outcomes Mapped to this Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V$</td>
<td>$v1$</td>
<td>All objects containing a “1”</td>
</tr>
<tr>
<td></td>
<td>$v2$</td>
<td>All objects containing a “2”</td>
</tr>
<tr>
<td>$S$</td>
<td>$s1$</td>
<td>All square objects</td>
</tr>
<tr>
<td></td>
<td>$s2$</td>
<td>All round objects</td>
</tr>
<tr>
<td>$C$</td>
<td>$c1$</td>
<td>All black objects</td>
</tr>
<tr>
<td></td>
<td>$c2$</td>
<td>All white objects</td>
</tr>
</tbody>
</table>

$P(c,v,s) = P(v,s|c) * P(c)$
$= P(s|v,c) * P(v|c) * P(c)$

$\Rightarrow$ satisfies Markov condition
$\Rightarrow$ Any distribution does so (no entailing independencies)
Example 2.7 (3)

Assign 1/13 to each object  $I_P(\{V\}, \{S\} \mid \{C\})$

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>Outcomes Mapped to this Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V$</td>
<td>$v_1$</td>
<td>All objects containing a “1”</td>
</tr>
<tr>
<td></td>
<td>$v_2$</td>
<td>All objects containing a “2”</td>
</tr>
<tr>
<td>$S$</td>
<td>$s_1$</td>
<td>All square objects</td>
</tr>
<tr>
<td></td>
<td>$s_2$</td>
<td>All round objects</td>
</tr>
<tr>
<td>$C$</td>
<td>$c_1$</td>
<td>All black objects</td>
</tr>
<tr>
<td></td>
<td>$c_2$</td>
<td>All white objects</td>
</tr>
</tbody>
</table>

$P(c,v,s) = P(v,s|c) \cdot P(c)$

$= P(v|s,c) \cdot P(s|c) \cdot P(c)$

$\Rightarrow$ satisfies Markov condition

$\Rightarrow$ Any distribution does so (no entailing independencies)
Definition 2.9

**Direct dependency** between $X$ and $Y$ in $P$ if $X$ and $Y$ are **not** conditionally independent given any subset of $V-\{X,Y\}$.

Markov condition implies “absence of edge between $X$ and $Y$ $\Rightarrow$ no direct dependency”. However, the existence of an edge between $X$ and $Y$ **doesn't** mean that there is a dependency.

But we want an edge to mean that there is a direct dependency!

**Definition 2.9:** Suppose we have a joint probability distribution $P$ of the random variables in some set $V$ and a DAG $G=(V,E)$. We say that $(G,P)$ satisfies the **faithfulness condition** if, based on the Markov condition, $G$ entails all and only conditional independencies in $P$. That is, the following two conditions hold:

1. $(G,P)$ satisfies the Markov condition.
2. All conditional independencies in $P$ are entailed by $G$, based on the Markov condition.
Example 2.8

For faithfulness show:

- Not $I_p(\{V\}, \{S\})$
- Not $I_p(\{V\}, \{C\})$
- Not $I_p(\{S\}, \{C\})$
- Not $I_p(\{V\}, \{C\} | \{S\})$
- Not $I_p(\{C\}, \{S\} | \{V\})$

implicitly from this it follows:

- Not $I_p(\{C\}, \{S,V\})$
- Not $I_p(\{S\}, \{C,V\})$
- Not $I_p(\{V\}, \{C,S\})$

$\mathcal{I}_p(\{V\}, \{S\} \parallel \{C\})$

DAG Pattern
Theorem 2.5

When \((G, P)\) satisfies the faithfulness condition, we say \(P\) and \(G\) are **faithful** to each other, and we say \(G\) is a **perfect map** of \(P\). When \((G, P)\) does not satisfy the faithfulness, we say they are **unfaithful** to each other.

**Theorem 2.5**: Suppose we have a joint probability distribution \(P\) of the random variables in some set \(V\) and a DAG \(G=(V,E)\). Then \((G, P)\) satisfies the faithfulness condition iff all and only conditional independencies in \(P\) are identified by d-separation in \(G\).

Example: not faithful
Theorem 2.6: If \((G, P)\) satisfies the faithfulness condition, then \(P\) satisfies this condition with all and only those DAGs that are Markov equivalent to \(G\). Furthermore, if we let \(gp\) be the DAG pattern corresponding to this Markov equivalence class, the d-separations in \(gp\) identify all and only conditional dependencies in \(P\). We say that \(gp\) and \(P\) are faithful to each other, and \(gp\) is a perfect map of \(P\).

A distribution \(P\) admits a faithful DAG representation if \(P\) is faithful to some DAG (and therefore some DAG pattern). --> not every \(P\) admits a faithful DAG representation.

Theorem 2.7: Suppose we have a joint probability distribution \(P\) of the random variables in some set \(V\) and a DAG \(G=(V,E)\). Then if \(P\) admits a faithful DAG representation, \(gp\) is the DAG pattern faithful to \(P\) if and only if the following hold:

1. \(X\) and \(Y\) are adjacent in \(gp\) iff there is no subsets \(S \subseteq V\) such that \(I_P(\{X\}, \{Y\} \mid S)\). That is, \(X\) and \(Y\) are adjacent iff there is a direct dependency between \(X\) and \(Y\).

2. \(X\overline{-->}Z\overline{-->}Y\) is a head-to-head meeting in \(gp\) iff \(Z \in S\) implies \(\neg I_P(\{X\}, \{Y\} \mid S)\).
**Embedded Faithfulness (1)**

**Definition 2.9:** Suppose we have a joint probability distribution $P$ of the random variables in some set $V$ and a DAG $G=(V,E)$. We say that $(G,P)$ satisfies the **faithfulness condition** if, based on the Markov condition, $G$ entails all and only conditional independencies in $P$. That is, the following two conditions hold:

1. $(G,P)$ satisfies the Markov condition.
2. All conditional independencies in $P$ are entailed by $G$, based on the Markov condition.

**Definition 2.10:** Let $P$ be a joint probability distribution of the variables in $V$ where $V \subseteq W$, and let $G=(W,E)$ be a DAG. We say $(G,P)$ satisfies the **embedded faithfulness condition** if the following two conditions hold:

1. Based on the Markov condition, $G$ entails only conditional independencies in $P$ for subset including only elements of $V$.
2. All conditional independencies in $P$ are entailed by $G$, based on the Markov condition.

We say $P$ is embedded faithfully in $G$. 
Let $P$ be a joint probability distribution of the variables in $W=\{L,V,C,S,F\}$, and $G=(W,E)$ be the following graph:

Let $(G,P)$ satisfies the faithfulness condition. Then, the only independencies involving only the variables $V, S, L$ and $F$ are:

\[
I_P(\{L\}, \{F, S\}) \quad I_P(\{L\}, \{S\}) \quad I_P(\{L\}, \{F\}) \\
I_P(\{F\}, \{L, V\}) \quad I_P(\{F\}, \{V\})
\]
Take the marginal distribution $P(v,s,l,f)$ of $P(v,s,c,l,f)$.

**Theorem 2.5** => a DAG is faithful to $P'=P(v,s,l,f)$ iff all and only conditional independencies in $P'$ are identified by $d$-separation in $G$.

$I_G([L],[F,S])$ $I_G([L],[S])$ $I_G([L],[F])$
$I_G([F],[L,V])$ $I_G([F],[V])$

**Lemma 2.4**

=> links in $G$ are $L-V$, $V-S$, and $S-F$

=> $L-V-S$ and $V-S-F$ are uncoupled meetings.

**Lemma 2.5**

$=> I_G([L],[S])$ implies $L-V-S$ is an uncoupled head-to-head meeting.

$=> I_G([V],[F])$ implies $V-S-F$ is an uncoupled head-to-head meeting.

$=>$ The result graph is not a DAG

$=>$ contradiction

$=>$ $P'$ does not have a faithful DAG representation.
**Definition 2.10**: Let $P'$ be a joint probability distribution of the variables in $V$ where $V \subseteq W$, and let $G=(W,E)$ be a DAG. We say $(G, P')$ satisfies the **embedded faithfulness condition** if the following two conditions hold:

1. Based on the Markov condition, $G$ entails only conditional independencies in $P'$ for subset including only elements of $V$.
2. All conditional independencies in $P'$ are entailed by $G$, based on the Markov condition.
**Theorem 2.8**: Let $P$ be a joint probability distribution of the variables in $W$ with $V \subseteq W$, and $G=(W,E)$. If $(G,P)$ satisfies the faithfulness condition, and $P'$ is the marginal distribution of $V$, then $(G,P')$ satisfies the embedded faithfulness condition.

**Theorem 2.9**: Let $P$ be a joint probability distribution of the variables in $V$ where $V \subseteq W$, and $G=(W,E)$. Then $(G,P)$ satisfies the embedded faithfulness condition iff all and only conditional independencies in $P$ are identified by d-separation in $G$ restricted to elements of $V$. 
Minimality
Markov Blankets & Boundaries
Minimality (1)

**Definition 2.11**: Suppose we have a joint probability distribution $P$ of the random variables in some set $V$ and a DAG $G=(V,E)$. We say that $(G,P)$ satisfies the **minimality condition** if the following two conditions hold:

1. $(G,P)$ satisfies the Markov condition.
2. If we remove any edges from $G$, the resultant DAG no longer satisfies the Markov condition with $P$.

**Theorem 2.11**: Suppose we have a joint probability distribution $P$ of the random variables in some set $V$ and a DAG $G=(V,E)$. If $(G,P)$ satisfies the faithfulness condition, then $(G,P)$ satisfies the minimality condition. However, $(G,P)$ can satisfies the minimality condition without satisfying the faithfulness condition.
Minimality (2)

Every probability distribution $P$ satisfies the minimality condition with some DAG:

**Theorem 2.12**: Suppose we have a joint probability distribution $P$ of the random variables in some set $V$. Create an arbitrary ordering of the nodes in $V$. For each $X \in V$, let $B_X$ be the set of all nodes that come before $X$ in the ordering, and let $PA_X$ be a minimal subset of $B_X$ such that $I_P(\{X\}, B_X | PA_X)$. Create a DAG $G$ by placing an edge from each node in $PA_X$ to $X$. Then $(G,P)$ satisfies the minimality condition. Furthermore, if $P$ is strictly positive (i.e., there are no probability values equal 0), then $PA_X$ is unique relative to the ordering.

**NOTE**: A DAG satisfying the minimality condition with a distribution is not necessarily minimal with respect to have a minimal number of edges! However, a faithful DAG is minimal in this sense, too.
Markov Blankets & Boundaries
**Markov Blanket (1)**

**Definition 2.12:** Let $V$ be a set of random variables, $P$ be their joint probability distribution, and $X \in V$. Then a Markov blanket $M_X$ of $X$ is any set of variables such that $X$ is conditionally independent of all the other variables given $M_X$. That is,

$$I_P(\{X\}, V - (M_X \cup \{X\})|M_X)$$

**Theorem 2.13:** Suppose $(G, P)$ satisfies the Markov condition. Then, for each variable $X$, the set of all parents of $X$, children of $X$, and parents of children of $X$ is a Markov blanket of $X$.

Examples:

$M_X = \{T,Y,Z\}$

---

© 2005-2008 Prof. Dr. Rainer Lienhart, Head of Multimedia Computing, Universität Augsburg
Eichleitnerstr. 30, D-86135 Augsburg, Germany; email: Rainer.Lienhart@informatik.uni-augsburg.de
**Definition 2.12**: Let $V$ be a set of random variables, $P$ be their joint probability distribution, and $X \in V$. Then a Markov blanket $M_X$ of $X$ is any set of variables such that $X$ is conditionally independent of all the other variables given $M_X$. That is,

$$I_P(\{X\}, V - (M_X \cup \{X\})|M_X)$$

**Theorem 2.13**: Suppose $(G, P)$ satisfies the Markov condition. Then, for each variable $X$, the set of all parents of $X$, children of $X$, and parents of children of $X$ is a Markov blanket of $X$.

Examples:

$M_X = \{T, Y, Z\}$

--> not minimal any more

$M_X = \{Y, Z\}$
**Markov Boundary**

**Definition 2.13**: Let $V$ be a set of random variables, $P$ be their joint probability distribution, and $X \in V$. Then a **Markov boundary** of $X$ is any Markov blanket such that none of its proper subsets is a Markov blanket of $X$.

**Theorem 2.14**: Suppose $(G, P)$ satisfies the faithfulness condition. Then for each variable $X$, the set of all parents of $X$, children of $X$, and parents of children of $X$ is the unique Markov boundary of $X$.

**Theorem 2.15**: Suppose $P$ is a strictly positive probability distribution of the variables in $V$. Then for each $X \in V$ there is a unique Markov boundary of $X$. 

© 2005-2008 Prof. Dr. Rainer Lienhart, Head of Multimedia Computing, Universität Augsburg
Eichleitnerstr. 30, D-86135 Augsburg, Germany; email: Rainer.Lienhart@informatik.uni-augsburg.de