Probability Models

Robert M. Haralick

Computer Science, Graduate Center
City University of New York
The Problem

- When there are many variables, the sample size is often too small.
- When the sample size is too small, the class conditional joint probability cannot be estimated directly.
- There must be some assumptions made to allow low order marginals to be combined in some manner to form class conditional joint probabilities to be used in the classification.
The Markov Assumption

\[ p(y_1 \mid y_2 \ldots y_N) = P(y_1 \mid y_2) \]
\[ p(y_2 \mid y_3 \ldots y_N) = P(y_2 \mid y_3) \]
\[ \vdots \]
\[ P(y_{N-2} \mid y_{N-1}, y_N) = P(y_{N-2} \mid y_{N-1}) \]

In general,

\[ P(y_n \mid y_{n+1} \ldots y_N) = P(y_n \mid y_{n+1}), \, n = 1, \ldots N - 1 \]
Now,
\[ P(x_1 \ldots x_N) = P(x_1 | x_2 \ldots x_N) P(x_2 \ldots x_N) \]
\[ = P(x_1 | x_2 \ldots x_N) P(x_2 | x_3 \ldots x_N) P(x_3 \ldots x_N) \]

Repeating the pattern,
\[ P(x_1 \ldots x_N) = \left[ \prod_{n=1}^{N-1} P(x_n | x_{n+1} \ldots x_N) \right] P(x_N) \]
Under the Markov Assumption

\[ P(x_n \mid x_{n+1} \ldots x_N) = P(x_n \mid x_{n+1}), \quad n = 1, \ldots N - 1 \]

Hence,

\[ P(x_1 \ldots x_N) = \left[ \prod_{n=1}^{N-1} P(x_n \mid x_{n+1} \ldots x_N) \right] P(x_N) \]

\[ = \left[ \prod_{n=1}^{N-1} P(x_n \mid x_{n+1}) \right] P(x_N) \]
The Markov Classifier

Assign \((x_1, \ldots x_N)\) to class \(c^*\) when

\[
P(x_1 \ldots x_N \mid c^*) > P(x_1 \ldots x_N \mid c), \ c \neq c^*
\]

\[
\left[ \prod_{n=1}^{N-1} P(x_n \mid x_{n+1}, c^*) \right] P(x_N \mid c^*) > \left[ \prod_{n=1}^{N-1} P(x_n \mid x_{n+1}, c) \right] P(x_N \mid c)
\]

for all other \(c\)
Let $i_1, \ldots, i_N$ be a permutation of $1, \ldots, N$. Assign $(x_1, \ldots x_N)$ to class $c^*$ when

$$P(x_1 \ldots x_N \mid c^*) > \prod_{n=1}^{N-1} P(x_{i_n} \mid x_{i_{n+1}}, c^*) P(x_{i_N} \mid c^*)$$

and

$$\prod_{n=1}^{N-1} P(x_{i_n} \mid x_{i_{n+1}}, c) P(x_{i_N} \mid c)$$

for all other $c$. 
How To Choose the Permutation

- Use the training data to estimate $P(x_i | x_j, c)$, $i \neq j$
- For permutation $i_1, \ldots, i_N$
- Use the first half of testing data to estimate the expected gain using $P(x_{i_n} | x_{i_{n+1}}, c)$
- Search for the permutation having the largest estimated expected gain
- For the best permutation, get an unbiased estimate of the estimated expected gain using the second half of the testing data
First Order Dependence Trees

\[ P(x_1, x_2, x_3, x_4, x_5) = p(x_1 \mid x_2)P(x_5 \mid x_2)P(x_3 \mid x_1)P(x_4 \mid x_1)P(x_2) \]
1 \quad = \quad \sum_{x_1} \sum_{x_2} \sum_{x_3} \sum_{x_4} \sum_{x_5} P(x_1, x_2, x_3, x_4, x_5) \\
\sum_{x_1} \sum_{x_2} \sum_{x_3} \sum_{x_4} \sum_{x_5} p(x_1 \mid x_2) P(x_5 \mid x_2) P(x_3 \mid x_1) P(x_4 \mid x_1) P(x_2) \\
\sum_{x_2} \sum_{x_3} \sum_{x_4} \sum_{x_5} P(x_2) \sum_{x_1} p(x_1 \mid x_2) \sum_{x_5} P(x_5 \mid x_2) \sum_{x_4} P(x_4 \mid x_1) \sum_{x_3} P(x_3 \mid x_1) \\
= \quad 1
First Order Dependence Trees

Precedence Function

<table>
<thead>
<tr>
<th>i</th>
<th>j(i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>
First Order Dependence Tree

\[
P(x_1, x_2, x_3, x_4, x_5) = p(x_1 | x_2)p(x_5 | x_2)p(x_3 | x_1)p(x_4 | x_1)p(x_2)
\]

\[
[N] = \{1, \ldots, N\}
\]

\[M \subset [N] \quad j : M \rightarrow N\]

\[G = ([N], E)\]

\[E = \{\{j(m), m\} \mid m \in M\}\]

\[
P(x_1, \ldots, x_N) = P(x_m : m \in [N] - M) \prod_{m \in M} P(x_m | x_{j(m)})
\]
Under the Markov assumption

\[ P(x_i, x_{i+1}, | x_{i+2} \ldots, x_N) = \frac{P(x_i, \ldots x_N)}{P(x_{i+2} \ldots x_N)} \]

\[ = \frac{P(x_i | x_{i+1} \ldots x_N)P(x_{i+1} \ldots x_N)}{P(x_{i+2} \ldots x_N)} \]

\[ = \frac{P(x_i | x_{i+1})P(x_{i+1} \ldots x_N)}{P(x_{i+2} \ldots x_N)} \]

\[ = \frac{P(x_i | x_{i+1})P(x_{i+1} | x_{i+2})P(x_{i+2} \ldots x_N)}{P(x_{i+2} \ldots x_N)} \]

\[ = P(x_i | x_{i+1})P(x_{i+1} | x_{i+2}) \]
Conditional Probability Products

\[ P(x_1 \mid x_2, x_3, x_4)P(x_4 \mid x_2, x_5, x_6)P(x_5 \mid x_6)P(x_2, x_6 \mid x_3)P(x_3) \]

- Does this product make a probability function?
- If it does, is the probability function an extension of these conditional probabilities?
Summing

Define,

\[ Q(x_1, \ldots, x_6) = P(x_1 | x_2, x_3, x_4)P(x_4 | x_2, x_5, x_6)P(x_5 | x_6) \]
\[ P(x_2, x_6 | x_3)P(x_3) \]

\[ Q(x_2, \ldots, x_6) = \sum_{x_1} Q(x_1, \ldots, x_6) \]

\[ = \sum_{x_1} P(x_1 | x_2, x_3, x_4)P(x_4 | x_2, x_5, x_6) \]
\[ P(x_5 | x_6)P(x_2, x_6 | x_3)P(x_3) \]

\[ = P(x_4 | x_2, x_5, x_6)P(x_5 | x_6)P(x_2, x_6 | x_3)P(x_3) \]
\[
Q(x_2, x_3, x_5, x_6) = \sum_{x_4} Q(x_2, \ldots, x_6)
\]
\[
= \sum_{x_4} P(x_4 | x_2, x_5, x_6) P(x_5 | x_6) P(x_2, x_6 | x_3) P(x_3)
\]
\[
= P(x_5 | x_6) P(x_2, x_6 | x_3) P(x_3)
\]
Summing

\[ Q(x_2, x_3, x_6) = \sum_{x_5} Q(x_2, x_3, x_5, x_6) \]

\[ = \sum_{x_5} P(x_5 \mid x_6) P(x_2, x_6 \mid x_3) P(x_3) \]

\[ = P(x_2, x_6 \mid x_3) P(x_3) \]

\[ = P(x_2, x_3, x_6) \]

\[ \sum_{x_2, x_3, x_6} Q(x_2, x_3, x_6) = \sum_{x_2, x_3, x_6} P(x_2, x_3, x_6) \]

\[ = 1 \]
Define,

\[ Q(x_1, \ldots, x_6) = P(x_1 | x_2, x_3, x_4)P(x_4 | x_2, x_5, x_6)P(x_5 | x_6) \]

\[ P(x_2, x_6 | x_3)P(x_3) \]

Is,

\[ Q(x_1 | x_2, x_3, x_4) = P(x_1 | x_2, x_3, x_4) \]
\[ Q(x_4 | x_2, x_5, x_6) = P(x_4 | x_2, x_5, x_6) \]
\[ Q(x_5 | x_6) = P(x_5 | x_6) \]
\[ Q(x_2, x_6 | x_3) = P(x_2, x_6 | x_3) \]
\[ Q(x_3) = P(x_3) \]
Define,

\[ Q(x_1, \ldots, x_6) = P(x_1 | x_2, x_3, x_4)P(x_4 | x_2, x_5, x_6)P(x_5 | x_6) \]
\[ P(x_2, x_6 | x_3)P(x_3) \]

Is,

\[ Q(x_1, x_2, x_3, x_4) = P(x_1, x_2, x_3, x_4) \]
\[ Q(x_4, x_2, x_5, x_6) = P(x_4, x_2, x_5, x_6) \]
\[ Q(x_5, x_6) = P(x_5, x_6) \]
\[ Q(x_2, x_3, x_6) = P(x_2, x_3, x_6) \]
\[ Q(x_3) = P(x_3) \]
Define,
\[ Q(x_1, \ldots, x_6) = P(x_1 | x_2, x_3, x_4)P(x_4 | x_2, x_5, x_6) \]
\[ \quad P(x_5 | x_6)P(x_2, x_6 | x_3)P(x_3) \]

Is
\[ Q(x_1 | x_2, x_3, x_4) = \frac{Q(x_1, x_2, x_3, x_4)}{Q(x_2, x_3, x_4)} \]

Find expressions for \( Q(x_1, x_2, x_3, x_4) \) and \( Q(x_2, x_3, x_4) \)
\[ Q(x_1, x_2, x_3, x_4) = \sum_{x_5} \sum_{x_6} Q(x_1, \ldots, x_6) \]
\[ = \sum_{x_5} \sum_{x_6} P(x_1 | x_2, x_3, x_4)P(x_4 | x_2, x_5, x_6) \]
\[ \quad P(x_5 | x_6)P(x_2, x_6 | x_3)P(x_3) \]
\[ = P(x_1 | x_2, x_3, x_4) \sum_{x_5} \sum_{x_6} P(x_4 | x_2, x_5, x_6) \]
\[ Q(x_2, x_3, x_4) = \sum_{x_1} Q(x_1, x_2, x_3, x_4) \]
\[ = \sum_{x_1} P(x_1 | x_2, x_3, x_4) \sum_{x_5} \sum_{x_6} P(x_4 | x_2, x_5, x_6) \]
\[ P(x_5 | x_6)P(x_2, x_6 | x_3)P(x_3) \]
\[ = \sum_{x_5} \sum_{x_6} P(x_4 | x_2, x_5, x_6) \]
\[ P(x_5 | x_6)P(x_2, x_6 | x_3)P(x_3) \]
Weak Extension

\[ Q(x_1 \mid x_2, x_3, x_4) = \frac{Q(x_1, x_2, x_3, x_4)}{Q(x_2, x_3, x_4)} \]

\[
= \frac{P(x_1 \mid x_2, x_3, x_4) \sum_{x_5} \sum_{x_6} P(x_4 \mid x_2, x_5, x_6)P(x_5 \mid x_6)P(x_2, x_3, x_6)}{\sum_{x_5} \sum_{x_6} P(x_4 \mid x_2, x_5, x_6)P(x_5 \mid x_6)P(x_2, x_3, x_6)}
\]

\[ = \frac{P(x_1 \mid x_2, x_3, x_4)}{P(x_1 \mid x_2, x_3, x_4)} \]

So we have shown a weak extension for one conditional probability.
Conditional Independence

**Definition**

Random variables $X$ and $Y$ are **conditionally independent** given random variable $Z$ if and only if for all values $x, y, z$ in the domain of the respective variables $X, Y, Z$

$$P(X = x, Y = y|Z = z) = P(X = x|Z = z)P(Y = y|Z = z)$$

For the sake of compactness, we write

$P(x, y|z)$ for $P(X = x, Y = y | Z = z)$
Conditional Independence Notation

If random variables $X$ and $Y$ are conditionally independent of random variable $Z$ we write

$$X \perp Y \mid Z$$

Let $\{X_1, \ldots, X_N\}$ be a set of random variables. If $X_i$ is conditionally independent of $X_j$ given $X_k$ we write

$$i \perp j \mid k$$

Let $A, B, C \subset \{1, \ldots, N\}$ with

- $A \cap B = \emptyset$
- $A \cap C = \emptyset$
- $B \cap C = \emptyset$

If $\{X_i : i \in A\}$ is conditionally independent of $\{X_j : j \in B\}$ given $\{X_k : k \in C\}$, then we write

$$A \perp B \mid C$$
Theorem

\[ P(x, y|z) = P(x|z)P(y|z) \text{ if and only if } P(x|y, z) = P(x|z) \]

Proof.

Suppose \( P(x, y|z) = P(x|z)P(y|z) \). Consider \( P(x|y, z) \)

\[
P(x|y, z) = \frac{P(x, y, z)}{P(y, z)} = \frac{P(x, y|z)P(z)}{P(y, z)} = \frac{P(x|z)P(y|z)P(z)}{P(y, z)} = P(x|z)
\]

Suppose \( P(x|y, z) = P(x|z) \). Consider \( P(x, y|z) \).

\[
P(x, y|z) = \frac{P(x, y, z)}{P(z)} = \frac{P(x|y, z)P(y, z)}{P(z)} = \frac{P(x|z)P(y, z)}{P(z)} = P(x|z)P(y|z)
\]
Can we see if $x_5$ is conditionally independent of $x_2$ given $x_6$. Is $P(x_5 \mid x_2, x_6) = P(x_5 \mid x_6)$?

$$Q(x_2, x_4, x_5, x_6) = \sum_{x_1} \sum_{x_3} Q(x_1, \ldots, x_6)$$

$$= \sum_{x_1} \sum_{x_3} P(x_1 \mid x_2, x_3, x_4)P(x_4 \mid x_2, x_5, x_6)$$

$$P(x_5 \mid x_6)P(x_2, x_6 \mid x_3)P(x_3)$$

$$= P(x_4 \mid x_2, x_5, x_6)P(x_5 \mid x_6)P(x_2, x_6)$$
\[ Q(x_2, x_5, x_6) = \sum_{x_4} Q(x_2, x_4, x_5, x_6) \]

\[ = \sum_{x_4} P(x_4 \mid x_2, x_5, x_6) P(x_5 \mid x_6) P(x_2, x_6) \]

\[ = P(x_5 \mid x_6) P(x_2, x_6) \]
\[ Q(x_5, x_6) = \sum_{x_1} \sum_{x_2} \sum_{x_3} \sum_{x_4} Q(x_1, \ldots, x_6) \]

\[ = \sum_{x_1} \sum_{x_2} \sum_{x_3} \sum_{x_4} P(x_1 \mid x_2, x_3, x_4) P(x_4 \mid x_2, x_5, x_6) \]

\[ P(x_5 \mid x_6) P(x_2, x_6 \mid x_3) P(x_3) \]

\[ = \sum_{x_2} \sum_{x_3} \sum_{x_4} P(x_4 \mid x_2, x_5, x_6) P(x_5 \mid x_6) P(x_2, x_3, x_6) \]

\[ = \sum_{x_2} \sum_{x_3} P(x_5 \mid x_6) P(x_2, x_3, x_6) \]

\[ = P(x_5 \mid x_6) P(x_6) = P(x_5, x_6) \]
$Q(x_2, x_5, x_6) = P(x_5 \mid x_6)P(x_2, x_6)$

$Q(x_2, x_6) = \sum_{x_5} Q(x_2, x_5, x_6)$

$= \sum_{x_5} P(x_5 \mid x_6)P(x_2, x_6)$

$= P(x_2, x_6)$

$Q(x_5 \mid x_2, x_6) = \frac{Q(x_2, x_5, x_6)}{Q(x_2, x_6)}$

$= \frac{P(x_5 \mid x_6)P(x_2, x_6)}{P(x_2, x_6)} = P(x_5 \mid x_6)$
Now,
\[ Q(x_5 \mid x_2, x_6) = P(x_5 \mid x_6) \]

But,
\[ Q(x_5, x_6) = P(x_5, x_6) \]

Hence,
\[ Q(x_5 \mid x_6) = P(x_5 \mid x_6) \]

Therefore,
\[ Q(x_5 \mid x_2, x_6) = Q(x_5 \mid x_6) \]
\[ x_5 \perp x_2 \mid x_6 \]
Conditional Independence

Suppose, \( x_5 \perp x_2 \mid x_6 \)

\[
Q(x_5 \mid x_2, x_6) = Q(x_5 \mid x_6)
\]

Then,

\[
Q(x_5, x_2 \mid x_6) = Q(x_5 \mid x_6)Q(x_2 \mid x_6)
\]

\[
Q(x_5, x_2 \mid x_6) = \frac{Q(x_2, x_5, x_6)}{Q(x_6)}
\]

\[
= \frac{Q(x_5 \mid x_2, x_6)Q(x_2, x_6)}{Q(x_6)}
\]

\[
= \frac{Q(x_5 \mid x_6)Q(x_2, x_6)}{Q(x_6)}
\]

\[
= Q(x_5 \mid x_6)Q(x_2 \mid x_6)
\]
\[Q(x_4 \mid x_2, x_5, x_6) = \frac{Q(x_2, x_4, x_5, x_6)}{Q(x_2, x_5, x_6)}\]

\[= \frac{P(x_4 \mid x_2, x_5, x_6)P(x_5 \mid x_6)P(x_2, x_6)}{P(x_5 \mid x_6)P(x_2, x_6)}\]

\[= P(x_4 \mid x_2, x_5, x_6)\]
Additional Relationships You Work Out

\[ Q(x_2, x_3, x_6) = \sum \sum \sum Q(x_1, \ldots, x_6) \]
\[ = \sum \sum \sum P(x_1 \mid x_2, x_3, x_4)P(x_4 \mid x_2, x_5, x_6) \]
\[ P(x_5 \mid x_6)P(x_2, x_6 \mid x_3)P(x_3) \]
\[ = \sum \sum P(x_4 \mid x_2, x_5, x_6)P(x_5 \mid x_6)P(x_2, x_3, x_6) \]
\[ = \sum P(x_5 \mid x_6)P(x_2, x_3, x_6) \]
\[ = P(x_2, x_3, x_6) \]
Semi-Graphoid

Definition
Let $I$ be an index set containing the indexes of all the variables. Let $G$ be a collection of triples each of whose components are subsets of the index set $I$. $G$ is called a Semi-Graphoid if and only if

- Mutual Exclusivity: $(A, B, C) \in G$ implies
  - $A \cap B = \emptyset$, $A \cap C = \emptyset$, $B \cap C = \emptyset$
- Symmetry: $(A, B, C) \in G$ if and only if $(B, A, C) \in G$
- Decomposition: $(A, B \cup D, C) \in G$ implies $(A, B, C) \in G$
- Weak Union: $(A, B \cup C, D) \in G$ implies $(A, B, C \cup D) \in G$
- Contraction: $(A, B, C \cup D) \in G$ and $(A, C, D) \in G$ imply $(A, B \cup C, D) \in G$
Theorem

Let \( \{X_1, \ldots, X_N\} \) be a set of random variables. Let

\[
G = \{(A, B, C) \in [N]^3 \mid A \cap B = \emptyset, A \cap C = \emptyset, B \cap C = \emptyset, A \independent B \mid C\}\}
\]

Then \( G \) is a semi-graphoid.

Proof.

We need to prove Symmetry, Decomposition, Weak Union, and Contraction. We do so in the following propositions where \( X, Y \) and \( Z \) represent tuples of random variables.
Proposition

\( X \perp Y \mid Z \) implies \( Y \perp X \mid Z \)

Proof.

\( X \perp Y \mid Z \) implies \( P(xy \mid z) = P(x \mid z)P(y \mid z) \)

\( P(xy \mid z) = P(x \mid z)P(y \mid z) \) implies \( P(xy \mid z) = P(y \mid z)P(x \mid z) \)

\( P(xy \mid z) = P(y \mid z)P(x \mid z) \) implies \( Y \perp X \mid Z \)
Proposition

\( Y \perp Z_1, Z_2 \mid X \) implies \( Y \perp Z_1 \mid X \) and \( Y \perp Z_2 \mid X \).

Proof.

\[
P(y, z_1, z_2 \mid x) = P(y \mid x)P(z_1, z_2 \mid x)
\]

\[
\sum_{z_2} P(y, z_1, z_2 \mid x) = \sum_{z_2} P(y \mid x)P(z_1, z_2 \mid x)
\]

\[
P(y, z_1 \mid x) = P(y \mid x)P(z_1 \mid x)
\]

Hence, \( Y \perp Z_1 \mid X \).

The proof for \( Y \perp Z_2 \mid X \) is similar with the roles of \( Z_1 \) and \( Z_2 \) interchanged.
Proposition

\[ Y \perp Z_1, Z_2 \mid X \text{ implies } Y \perp Z_1 \mid X, Z_2 \text{ and } Y \perp Z_2 \mid X, Z_1. \]

Proof.

Suppose \( Y \perp Z_1, Z_2 \mid X \). Consider \( P(Y, Z_1 \mid X, Z_2) \).

\[
P(y, z_1 \mid x, z_2) = \frac{P(x, y, z_1, z_2)}{P(x, z_2)} = \frac{P(y, z_1, z_2 \mid x)P(x)}{P(x, z_2)}
\]

\[
= \frac{P(y \mid x)P(z_1, z_2 \mid x)P(x)}{P(x, z_2)} = P(y \mid x)P(z_1 \mid x, z_2)
\]

But \( Y \perp Z_1, Z_2 \mid X \) implies \( Y \perp Z_2 \mid X \) so that \( P(y, z_2 \mid x) = P(y \mid x)P(z_2 \mid x) \).

Hence, \( P(y \mid x) = P(y, z_2 \mid x) / P(z_2 \mid x) \). Therefore,

\[
P(y, z_1 \mid x, z_2) = \frac{P(y, z_2 \mid x)}{P(z_2 \mid x)}P(z_1 \mid x, z_2)
\]

\[
= P(y \mid x, z_2)P(z_1 \mid x, z_2)
\]

Thus, \( Y \perp Z_1 \mid X, Z_2 \). Similarly, \( Y \perp Z_2 \mid X, Z_1 \). \( \square \)
Proposition

\[ X \perp Y \mid Z_1 \cup Z_2 \text{ and } X \perp Z_1 \mid Z_2 \text{ imply } X \perp Y \cup Z_1 \mid Z_2 \]

Proof.

\[ X \perp Y \mid Z_1 \cup Z_2 \text{ implies } P(xy \mid z_1, z_2) = P(x \mid z_1, z_2)P(y \mid z_1, z_2) \]

\[ X \perp Z_1 \mid Z_2 \text{ implies } P(xz_1 \mid z_2) = P(x \mid z_2)P(z_1 \mid z_2) \]

\[
P(xyz_1 \mid z_2) = \frac{P(xyz_1 z_2)}{P(z_2)} = \frac{P(xy \mid z_1 z_2)P(z_1 z_2)}{P(z_2)}
\]

\[ = P(x \mid z_1 z_2)P(y \mid z_1 z_2) \frac{P(z_1 z_2)}{P(z_2)}
\]

\[ = P(xz_1 z_2) \frac{P(y \mid z_1 z_2)}{P(z_2)} = P(xz_1 \mid z_2)P(y \mid z_1 z_2)
\]

\[ = P(x \mid z_2)P(z_1 \mid z_2) \frac{P(yz_1 z_2)}{P(z_1 z_2)}
\]

\[ = P(x \mid z_2) \frac{P(z_1 z_2)}{P(z_2)} \frac{P(yz_1 z_2)}{P(z_1 z_2)} = P(x \mid z_2)P(yz_1 \mid z_2)
\]
Theorem

\( G \) is a Semi-Graphoid if and only if

- \( A \perp B \mid C \) if and only if \( B \perp A \mid C \)
- \( A \perp B \cup C \mid D \) if and only if \( A \perp B \mid C \cup D \) and \( A \perp C \mid D \)
Graphoid

Definition

Let $I$ be an index set containing the indexes of all the variables. Let $G$ be a collection of triples each of whose components are subsets of the index set $I$. We write $A \perp B \mid C$ if and only if the triple $(A, B, C) \in G$.

$G$ is called a **Graphoid** if and only if

- **Mutual Exclusivity:** $(A, B, C) \in G$ implies $A \cap B = \emptyset$, $A \cap C = \emptyset$, $B \cap C = \emptyset$

- **Symmetry:** $(A, B, C) \in G$ if and only if $(B, A, C) \in G$

- **Decomposition:** $(A, B \cup D, C) \in G$ implies $(A, B, C) \in G$

- **Weak Union:** $(A, B \cup C, D) \in G$ implies $(A, B, C \cup D) \in G$

- **Contraction:** $(A, B, C \cup D) \in G$ and $(A, C, D) \in G$ imply $(A, B \cup C, D) \in G$

- **Intersection:** $(A, B, C \cup D) \in G$ and $(A, C, B \cup D) \in G$ imply $(A, B \cup C, D) \in G$
**Proposition**

$Y \independent Z_2 \mid X, Z_1$ and $Y \independent Z_1 \mid X, Z_2$ imply

$$P(y \mid x, z_2) = P(y \mid x, z_1) \text{ for all values } x, y, z_1, z_2.$$

**Proof.**

By the Conditional Independence Characterization Theorem, $Y \independent Z_2 \mid X, Z_1$ implies $P(y \mid x, z_1, z_2) = P(y \mid x, z_1)$ and with the roles of $Z_1$ and $Z_2$ interchanged, $Y \independent Z_1 \mid X, Z_2$ implies $P(y \mid x, z_1, z_2) = P(y \mid x, z_2)$. Now, $P(y \mid x, z_1, z_2) = P(y \mid x, z_1)$ and $P(y \mid x, z_1, z_2) = P(y \mid x, z_2)$ imply

$$P(y \mid x, z_1) = P(y \mid x, z_2).$$
Proposition

If \( P(y \mid x, z_1) = P(y \mid x, z_2) \) for all values \( x, y, z_1, z_2 \) of the random variables \( X, Y, Z_1, Z_2 \), and \( P(x, y, z_1) > 0 \), and \( P(x, y, z_2) > 0 \), then \( P(y \mid x) = P(y \mid x, z_1) = P(y \mid x, z_2) \).

Proof.

\[
\begin{align*}
P(y \mid x, z_1) &= P(y \mid x, z_2) = \frac{P(x, y, z_2)}{P(x, z_2)} \\
P(y \mid x, z_1)P(x, z_2) &= P(x, y, z_2) \\
\sum_{z_2} P(y \mid x, z_1)P(x, z_2) &= \sum_{z_2} P(x, y, z_2) = P(x, y) \\
P(y \mid x, z_1)P(x) &= P(x, y) \\
P(y \mid x, z_1) &= P(y \mid x)
\end{align*}
\]
Conditional Independence: Intersection

Proposition

Suppose that for any values for any group of joint variables, their probability is greater than zero. Then, \( Y \perp Z_1 | X, Z_2 \) and \( Y \perp Z_2 | X, Z_1 \) imply \( Y \perp Z_1, Z_2 | X \). (The Intersection Property holds.)

Proof.

\[
P(y, z_1, z_2 | x) = \frac{P(y, z_1, z_2, x)}{P(x)} = \frac{P(y, z_1 | x, z_2)P(x, z_2)}{P(x)}
\]

\[
= P(y | x, z_2)P(z_1 | x, z_2) \frac{P(x, z_2)}{P(x)}
\]

\[
= P(y | x, z_2)P(z_1, z_2 | x)
\]

But by the previous corollary, \( Y \perp Z_1 | X, Z_2 \) and \( Y \perp Z_2 | X, Z_1 \) implies \( P(y | x, z_2) = P(y | x) \). Hence,

\[
P(y, z_1, z_2 | x) = P(y | x)P(z_1, z_2 | x)
\]
Theorem

Suppose that for any values for any group of joint variables, the probability is greater than zero. Then, $Y \perp Z_1 \cup Z_2 \mid X$ if and only if $Y \perp Z_1 \mid X \cup Z_2$ and $Y \perp Z_2 \mid X \cup Z_1$.

Proof.

By weak union, $Y \perp Z_1 \cup Z_2 \mid X$ implies $Y \perp Z_1 \mid X \cup Z_2$ and $Y \perp Z_2 \mid Z_1 \cup X$. By intersection, $Y \perp Z_1 \mid X \cup Z_2$ and $Y \perp Z_2 \mid Z_1 \cup X$ implies $Y \perp Z_1 \cup Z_2 \mid X$. \qed
Graphical Models associates a graph, called the **conditional independence graph**, from which the all the conditional independencies can be easily seen.

When the conditional independence graph is triangulated, then the joint probability function can be expressed with a probability product form.

- The product form can be read off the graph
- The product form is a strong extension of the marginal terms of the product
A graph $G = (N, E)$ where $N$ is an index set and $E$, the edge set, is a collection of subsets of $N$ where each subset has exactly 2 elements of $N$. 
Here, $G = (N, E)$ where

$$N = \{1, 2, 3, 4\}$$

$$E = \{\{1, 2\}, \{2, 4\}, \{3, 4\}, \{3, 1\}\}$$
Boundary

Definition

Let \( G = (N, E) \) be a graph and \( i \in N \). The **boundary** of \( i \) is defined by

\[
\text{bndry}(i) = \{ j \in N \mid \{ i, j \} \in E \}
\]

- \( \text{bndry}(1) = \{2, 3\} \)
- \( \text{bndry}(2) = \{1, 4\} \)
- \( \text{bndry}(3) = \{1, 4\} \)
- \( \text{bndry}(4) = \{2, 3\} \)
A graph \((N, E)\) is called a **Conditional Independence Graph** of a random variable set \(X = \{X_1, \ldots, X_M\}\) if and only if \(N = \{1, \ldots, M\}\), the index set for the variables in \(X\), and

\[
E^c = \{\{i, j\} \mid X_i \perp \!
\!
\!
\perp X_j \mid X - \{X_i, X_j\} \}
\]
Nodes correspond to indexes of variables in the variable set $\mathcal{X} = \{ X_1, \ldots, X_6 \}$

\{i, j\} not in the edge set means $X_i \perp X_j \mid \mathcal{X} - \{X_i, X_j\}$
\{Y, Z_1\} and \{Y, Z_2\} not in edge set means

\[
\begin{align*}
Y \independent Z_1 & \mid \{X, Y, Z_1, Z_2\} - \{Y, Z_1\} \\
Y \independent Z_2 & \mid \{X, Y, Z_1, Z_2\} - \{Y, Z_2\} \\
Y \independent Z_1 & \mid \{X, Z_2\} \\
Y \independent Z_2 & \mid \{X, Z_1\}
\end{align*}
\]
Suppose that for any values for any group of joint variables, the joint probability is greater than zero. \( Y \perp Z_1, Z_2 \mid X \) if and only if \( Y \perp Z_1 \mid X, Z_2 \) and \( Y \perp Z_2 \mid X, Z_1 \).
**Theorem**

Suppose that for any values for any group of joint variables, the joint probability is greater than zero.

- \( Y \perp Z_1, Z_2 \mid X \) if and only if \( Y \perp Z_1 \mid X, Z_2 \) and \( Y \perp Z_2 \mid X, Z_1 \).
- \( Y \perp Z_1, Z_2 \mid X \) implies \( Y \perp Z_1 \mid X \) and \( Y \perp Z_2 \mid X \).

\[
\begin{align*}
Y_1 & \perp Z_1 \mid X, \ Y_2 \perp Z_2 \mid X, \ Y_1 \perp Z_1 \mid X, \ Y_2 \perp Z_2 \mid X \\
Y_1, Y_2 & \perp Z_1 \mid X, \ Y_1, Y_2 \perp Z_2 \mid X, \ Y_1, Y_2 \perp Z_1, Z_2 \mid X \\
Z_1, Z_2 & \perp Y_1 \mid X, \ Z_1, Z_2 \perp Y_2 \mid X
\end{align*}
\]
Paths

Definition

Let \((G, E)\) be a graph and \(g_1, \ldots, g_N \in G\). \(< g_1, \ldots, g_N >\) is a path in \((G, E)\) if and only if \(\{g_n, g_{n+1}\} \in E\) for every \(n \in \{1, \ldots, N - 1\}\).
Definition

Let \((G, E)\) be a graph and \(A, B\) be subsets of \(G\). \(A\) and \(B\) are said to be connected if and only if for some \(a \in A\) and \(b \in B\), there is a path \(< a, g_1, \ldots, g_N, b >\) in \(G\).
Definition

Let \((G, E)\) be a graph and \(A, B, S\) be non-empty subsets of \(G\). \(S\) separates \(A\) from \(B\) if and only if for every \(a \in A\) and \(b \in B\), every path in \(G\) that begins with \(a\) and ends with \(b\) has at least one node in \(S\).
A separates $B \cup \{i\}$ from $C \cup \{j\}$

$N = A \cup B \cup C \cup \{i, j\}$

Then $i \perp j \mid A$
Separation Theorem

**Theorem**

Let $G = (N, E)$ be a connected conditional independence graph for a set of random variables whose joint probability is positive. If $A \subset N$ is any node set that separates two nodes $i$ and $j$, then $i \perp j \mid A$.

**Proof.**

Let $B$ be the set of nodes that either connect to $i$ directly or through $A$. Let $C$ be the set of nodes that either connect to $j$ directly or through $A$. Hence, $\{A, B, C, \{i, j\}\}$ form a partition of $N$. By construction of the conditional independence graph, $i \perp j \mid N - \{i, j\}$ and $i \perp p \mid N - \{i, p\}$. Application of the block independence theorem yields $i \perp j, p \mid N - \{i, j, p\}$. Application of the reduction theorem yields $i \perp j \mid N - \{i, j, p\}$. Repeated application using the remaining nodes of $C$ yields $i \perp j \mid N - \{i, j\} - C$. Similarly for using $q$. Repeated application yields $i \perp j \mid N - \{i, j\} - B - C$. But $N - \{i, j\} - B - C = A$. Therefore $i \perp j \mid A$. □
All conditional independences can be read off the Conditional Independence Graph.

**Corollary**

Let \( G = (N, E) \) be a conditional independence graph and \( n \in N \). Define \( B = N - \{n\} - \text{bndry}(n) \). Then \( n \perp B \mid \text{bndry}(n) \).

**Proof.**

The set \( \text{bndry}(n) \) separates \( n \) from \( B \).

**Definition**

Let \( G = (N, E) \) be a conditional independence graph and \( n \in N \). The **Markov Blanket** of node \( n \) is \( \text{bndry}(n) \).
Complete Graphs

**Definition**

A graph $G = (N, E)$ is **complete** if and only if

$$E = \{\{i, j\} \mid i, j \in N, i \neq j\}$$

**Figure:** The Complete Graph on 4 Nodes
Graph Restriction

Definition

Let $G = (N, E)$ be a graph and $A \subseteq N$. The graph of $G$ restricted to $A$, $G|_A$, is defined by

$$G|_A = (A, E|_A)$$

where

$$E|_A = \{\{i, j\} \in E \mid i, j \in A\}$$
Completeness

Definition

Let $G = (N, E)$ be a graph. Let a subset of nodes $A \subset N$ be given. We say $A$ is complete if and only if $G|_A$ is a complete graph.
A subset of nodes $A \subset N$ is **maximally complete** if and only if

- $G|_{A}$ is complete
- $B \supset A$ and $G|_{B}$ complete implies $B = A$
Clique

**Definition**

Let $G = (N, E)$ be a graph. A maximally complete subset $A \subset N$ is called a **clique** of $G$. 
A graph is chordal (triangulated, decomposable) if and only if every cycle of length 4 or more has a chord.

Figure: Non-chordal
A graph is **chordal** (triangulated, decomposable) if and only if every cycle of length 4 or more has a chord.

**Figure**: Non-chordal
A Graph $G = (N, E)$ is \textbf{Decomposable} if and only if

- $G$ is chordal
- The cliques of $G$ can be put in running intersection order $C_1, \ldots, C_K$ with separators $S_2, \ldots, S_K$ where

$$S_k = C_k \bigcap \left( \bigcup_{i=1}^{k-1} C_i \right), \ k = 2, \ldots, K - 1$$

such that $S_k$ is complete.
Example

\[
\begin{align*}
C_1 &= \{a, b, c, d, g\} \\
C_2 &= \{c, d, f, g\} \\
C_3 &= \{f, g, h, i\} \\
C_4 &= \{d, e, f, j\}
\end{align*}
\]

\[
\begin{align*}
S_2 &= C_2 \cap C_1 = \{c, d, g\} \\
S_3 &= C_3 \cap (C_1 \cup C_2) = \{f, g\} \\
S_4 &= C_4 \cap (C_1 \cup C_2 \cup C_3) = \{d, f\}
\end{align*}
\]
Decomposable Graph

\[ I = \{1, 2, 3, 4, 5\} \]

\[
\begin{array}{c|c|c}
C_1 &=& \{1, 2, 5\} \quad 1 \parallel 4 \quad | \quad 2, 5 \\
C_2 &=& \{2, 3, 5\} \quad 1 \parallel 3 \quad | \quad 2, 5 \\
C_3 &=& \{3, 4, 5\} \quad 2 \parallel 4 \quad | \quad 3, 5 \\
S_2 &=& \{2, 5\} \quad 1 \parallel 4 \quad | \quad 3, 5 \\
S_3 &=& \{3, 5\} \quad 1 \parallel 4 \quad | \quad 2, 3, 5 \\
\end{array}
\]

\[
P(x_i : i \in I) = \frac{P(x_i : i \in C_1)P(x_i : i \in C_2)P(x_i : i \in C_3)}{P(x_i : i \in S_2)P(x_i : i \in S_3)}
\]

\[
= P(x_i : i \in C_1)P(x_i : i \in C_2 - S_2 | S_2)P(x_i : i \in C_3 - S_3 | S_3)
\]
Let \( I \) be an index subset. If \( I = \{1, 3, 7\} \), then

\[
P(x_i : i \in I) = P(x_1, x_3, x_7)
\]
Decomposable Graphs

**Theorem**

If $G$ is a decomposable graph with cliques in running intersection order $C_1, \ldots, C_K$ and separators $S_2, \ldots, S_K$ then

$$P(x_1, \ldots, x_N) = \frac{\prod_{k=1}^K P(x_i : i \in C_k)}{\prod_{m=2}^K P(x_j : j \in S_m)}$$

$$= P(x_i : i \in C_1) \prod_{k=2}^K P(x_i : i \in C_k - S_k \mid S_k)$$
Cliques in running intersection order: \{1, 2, 3, 4\}, \{2, 3, 4, 5\}, \{5, 6\}
Separators: \{2, 3, 4\}, \{5\}

\[ P(x_1, \ldots, x_6) = P(x_1, x_2, x_3, x_4)P(x_5 | x_2, x_3, x_4)P(x_6 | x_5) \]
The product form

\[ Q(x_1, \ldots, x_6) = P(x_1, x_2, x_3, x_4)P(x_5 | x_2, x_3, x_4)P(x_6 | x_5) \]

is an extension of the marginals

- \( P(x_1, x_2, x_3, x_4) \)
- \( P(x_2, x_3, x_4, x_5) \)
- \( P(x_5, x_6) \)
Product Form

\[ Q(x_1, \ldots, x_6) = P(x_1, x_2, x_3, x_4) P(x_5 \mid x_2, x_3, x_4) P(x_6 \mid x_5) \]

\[ Q(x_1, x_2, x_3, x_4) = \sum_{x_5} \sum_{x_6} Q(x_1, \ldots, x_6) \]
\[ = \sum_{x_5} \sum_{x_6} P(x_1, x_2, x_3, x_4) P(x_5 \mid x_2, x_3, x_4) P(x_6 \mid x_5) \]
\[ = P(x_1, x_2, x_3, x_4) \sum_{x_5} P(x_5 \mid x_2, x_3, x_4) \sum_{x_6} P(x_6 \mid x_5) \]
\[ = P(x_1, x_2, x_3, x_4) \sum_{x_5} P(x_5 \mid x_2, x_3, x_4) \]
\[ = P(x_1, x_2, x_3, x_4) \]
Product Form

\[ Q(x_1, \ldots, x_6) = P(x_1, x_2, x_3, x_4)P(x_5 \mid x_2, x_3, x_4)P(x_6 \mid x_5) \]

\[ Q(x_2, x_3, x_4, x_5) = \sum_{x_1} \sum_{x_6} P(x_1, x_2, x_3, x_4)P(x_5 \mid x_2, x_3, x_4)P(x_6 \mid x_5) \]

\[ = P(x_5 \mid x_2, x_3, x_4) \sum_{x_1} P(x_1, x_2, x_3, x_4) \sum_{x_6} P(x_6 \mid x_5) \]

\[ = P(x_5 \mid x_2, x_3, x_4)P(x_2, x_3, x_4) = P(x_2, x_3, x_4, x_5) \]
Product Form

\[ Q(x_1, \ldots, x_6) = P(x_1, x_2, x_3, x_4)P(x_5 | x_2, x_3, x_4)P(x_6 | x_5) \]

\[ Q(x_2, x_3, x_4, x_5, x_6) = \sum_{x_1} Q(x_1, \ldots, x_6) \]
\[ = \sum_{x_1} P(x_1, x_2, x_3, x_4)P(x_5 | x_2, x_3, x_4)P(x_6 | x_5) \]
\[ = P(x_2, x_3, x_4)P(x_5 | x_2, x_3, x_4)P(x_6 | x_5) \]
\[ = P(x_2, x_3, x_4, x_5)P(x_6 | x_5) \]

\[ Q(x_5, x_6) = \sum_{x_2} \sum_{x_3} \sum_{x_4} P(x_2, x_3, x_4, x_5)P(x_6 | x_5) \]
\[ = P(x_5)P(x_6 | x_5) = P(x_5, x_6) \]
Decomposable Graphs

$$S_k = C_k \bigcap \left( \bigcup_{i=1}^{k-1} C_i \right), \, k = 2, \ldots, K$$

$$P(x_1, \ldots, x_N) = P(x_i : i \in C_1) \prod_{k=2}^{K} P(x_i : i \in C_k - S_k | S_k)$$

**Proposition**

$$(C_k - S_k) \cap (\bigcup_{i=1}^{k-1} C_i) = \emptyset$$

**Proof.**

$$(C_k - S_k) \cap (\bigcup_{i=1}^{k-1} C_i) = (C_k - (C_k \cap (\bigcup_{i=1}^{k-1} C_i)) \cap (\bigcup_{i=1}^{k-1} C_i)$$

$$= (C_k - (\bigcup_{i=1}^{k-1} C_i)) \cap (\bigcup_{i=1}^{k-1} C_i)$$

$$= \emptyset$$
Decomposable Graphs: Summability

\[ S_k = C_k \cap \left( \bigcup_{i=1}^{k-1} C_i \right), \quad k = 2, \ldots, K \]

\[ P(x_1, \ldots, x_N) = P(x_i : i \in C_1) \prod_{k=2}^{K} P(x_i : i \in C_k - S_k | S_k) \]

\[ (C_k - S_k) \cap \left( \bigcup_{i=1}^{k-1} C_i \right) = \emptyset \]

**Proposition**

\[ \sum_{x_1} \sum_{x_2} \cdots \sum_{x_N} P(x_i : i \in C_1) \prod_{k=2}^{K} P(x_i : i \in C_k - S_k | S_k) = 1 \]

**Proof.**

\[ S = \sum \sum_{x_2} \cdots \sum_{x_N} P(x_i : i \in C_1) \prod_{k=2}^{K} P(x_i : i \in C_k - S_k | S_k) \]

\[ = \sum_{C_1} \sum_{C_2-S_2} \cdots \sum_{C_K-S_K} P(x_i : i \in C_1) \prod_{k=2}^{K} P(x_i : i \in C_k - S_k | S_k) \]

\[ = \sum_{C_1} P(x_i : i \in C_1) \sum_{C_2-S_2} P(x_i : i \in C_2 - S_2 | S_2) \cdots \sum_{C_K-S_K} P(x_i : i \in C_K - S_K | S_K) \]

\[ = 1 \]
Summability Example

\[ S = \sum_{x_1} \cdots \sum_{x_9} P(x_1 x_2 x_3 x_5) P(x_4 | x_2 x_3 x_5) P(x_6 | x_1 x_5) P(x_7 | x_5 x_6) P(x_9 | x_6 x_7) \]
\[ = \sum_{x_1 x_2 x_3 x_5} P(x_1 x_2 x_3 x_5) \sum_{x_4} P(x_4 | x_2 x_3 x_5) \sum_{x_6} P(x_6 | x_1 x_5) \sum_{x_7} P(x_7 | x_5 x_6) \sum_{x_8 x_9} P(x_8 x_9 | x_6 x_7) \]
\[ = 1 \]

\[ C_1 = \{1, 2, 3, 5\} \]
\[ C_2 = \{2, 3, 4, 5\} \quad S_2 = \{2, 3, 5\} \]
\[ C_3 = \{1, 5, 6\} \quad S_3 = \{1, 5\} \]
\[ C_4 = \{5, 6, 7\} \quad S_4 = \{5, 6\} \]
\[ C_5 = \{6, 7, 8, 9\} \quad S_5 = \{6, 7\} \]
Definition

Let \( G = (V, E) \) be a connected graph. A non-empty subset \( S \subset V \) is called a **Separator** of \( G \) if and only if \( G(V - S, E|_{V-S}) \) is not connected. Let \( A, B, \) and \( S \) be disjoint non-empty subsets of \( V \). \( S \) is a **Separator of \( A \) from \( B \)** in graph \( G \) if and only if in the restricted graph \( G|_{V-S} \), there exists no \( a \in A \) and \( b \in B \) such that \( \{a, b\} \in E|_{V-S} \).

A separator \( S \) is called a **Minimal Separator** if and only if \( T \) a separator with \( T \subset S \) implies \( T = S \).

Theorem

A graph is triangulated if and only if each minimal separator is maximally complete.
Theorem

A graph $G$ is a triangulated graph if and only if the vertices of $G$ can be partitioned into three nonempty subsets $A$, $S$, and $B$, such that

- $G|_{A \cup S}$ and $G|_{B \cup S}$ are triangulated subgraphs of $G$
- $S$ separates $A$ from $B$

This is one of the reasons that triangulated graphs are called decomposable graphs.
Definition

Let $G(V, E)$ be a graph and $\{A, B, S\}$ be a non-trivial partition of $V$. $(A, B, S)$ is called a Decomposition of $G$ into $G_{A∪S}$ and $G_{B∪S}$ if and only if

- $S$ separates $A$ from $B$ in $G$
- $G_S$ is a complete graph
- $G_{A∪S}$ and $G_{B∪S}$ are each triangulated
A graph is decomposable if and only if either $G$ is complete or there exists a decomposition $(A, B, S)$ of $G$ into $G_A \cup S$ and $G_B \cup S$. 
Triangulated Graphs

Definition

A Perfect Elimination Ordering in a graph is an ordering of the vertices of the graph such that, for each vertex \( v \), \( v \) and the neighbors of \( v \) that occur after \( v \) in the ordering form a maximally complete graph.

Theorem

A graph is triangulated if and only if it has a perfect elimination ordering.

Theorem

A graph is triangulated if and only if its cliques can be put in running intersection order.
A triangulated graph can have only linearly many cliques, while non-chordal graphs may have exponentially many. Therefore clique finding in triangulated graphs can be done in polynomial time.
Theorem

If a graph \( G \) is triangulated graph and \( C_1, \ldots, C_K \) are the cliques of \( G \) put in running intersection order with separators \( S_2, \ldots, S_K \),

\[
S_k = C_k \cap \left( \bigcup_{i=1}^{k-1} C_i \right), \quad k = 2, \ldots, K
\]

then

\[
P(x_1, \ldots, x_N) = \prod_{k=1}^{K} \frac{P(x_i : i \in C_k)}{\prod_{k=2}^{K} P(x_i : i \in S_k)}
\]
Theorem

Let $P(x_1, \ldots, x_N) > 0$ and $G$ be the conditional independence graph of $P$. If $\{A, B, S\}$ is a non-trivial partition of $\{1, \ldots, N\}$ and $S$ is a separator of $A$ from $B$ in $G$, then $A \perp B \mid S$

$$P(x_i : i \in A \cup B \mid x_j : j \in S) = P(x_i : i \in A \mid x_j : j \in S)P(x_i : i \in B \mid x_j : j \in S)$$
What happens if the conditional independence graph is not triangulated? Can the joint probability distribution be written in a product form?
Generalized Products

Theorem

Let $f$ be a probability distribution. Then $X$ is Conditionally Independent of $Y$ given $Z$ if and only if

$$f(x, y, z) = g(x, z)h(y, z)$$

Proof.

By definition of conditional independence, $X$ is conditionally independent of $Y$ given $Z$ if and only if

$$f(x, y|z) = f(x|z)f(y|z)$$

Hence $X$ is conditionally independent of $Y$ given $Z$ if and only if

$$f(x, y, z) = f(x|z)f(y|z)f(z)$$

$$= [f(x|z)][f(y|z)f(z)]$$

$$= [f(x|z)][f(y, z)]$$

Take $g(x, z) = f(x|z)$ and $h(y, z) = f(y, z)$ □
Generalized Products

Definition

Let $B_1, \ldots, B_K$ be index subsets of $\{1, \ldots, N\}$. The product form
\[
\prod_{k=1}^{K} a_k(x_i : i \in B_k)
\]
is called a generalized product form if and only if for some probability function $P(x_1, \ldots, x_N)$
\begin{itemize}
  \item $P(x_1, \ldots, x_N) = \prod_{k=1}^{K} a_k(x_i : i \in B_k)$
  \item $P(x_1, \ldots, x_N)$ is an extension of $P(x_i : i \in B_k), k = 1, \ldots, K$
\end{itemize}
Generalized Products

Let $B_1, \ldots, B_K$ be index subsets of $\{1, \ldots, N\}$. Given marginal probability functions $P(x_i : i \in B_k), k = 1, \ldots, K$ find functions $a_k(x_i : i \in B_k)$ such that

- $P(x_1, \ldots, x_N) = \prod_{k=1}^{K} a_k(x_i : i \in B_k)$
- $P(x_1, \ldots, x_N)$ is an extension of $P(x_i : i \in B_k), k = 1, \ldots, K$
**Definition**

Let $S = \{s_1, \ldots, s_M\}$ be an index subset of $\{1, \ldots, N\}$. $\pi_S(x_1, \ldots, x_N)$ is called the *projection* of $(x_1, \ldots, x_N)$ onto the index set $S$. $\pi_S(x) = (x_{s_1}, \ldots, x_{s_M}) = (x_i : i \in S)$.

If $(x_1, x_2, x_3, x_4, x_5) = (1, 5, 4, 3, 0)$ and $S = \{1, 4, 5\}$, then $\pi_S(1, 5, 4, 3, 0) = (x_i : i \in S) = (1, 3, 0)$. 
Inverse Projection

**Definition**

Let $h$ be a tuple whose components are indexed in index set $S$. Let $I$ be the index set for all the variables. The *inverse projection* $\pi_I^{-1} h$ of $h$ with respect to $I$ is defined by

$$
\pi_I^{-1}(h) = \{(x_1, \ldots, x_N) \mid \pi_S(x_1, \ldots, x_N) = h\}
$$
Let $P$ be a probability function on $N$ variables $(x_1, \ldots, x_N)$. Let $S_0, \ldots, S_{K-1}$ be $K$ index sets of $\{1, \ldots, N\}$ covering $\{1, \ldots, N\}$. Fix $k$. Let $h$ be a tuple whose components are indexed in index set $S_k$: $h = (x_i : i \in S_k)$.

$$
P(h) = P(x_i : i \in S_k) = \sum_{(x_1, \ldots, x_N) \in \pi_i^{-1}(h)} P(x_1, \ldots, x_N)
$$
Let $P$ be a probability function on $N$ variables $(x_1, \ldots, x_N)$. Let $S_0, \ldots, S_{K-1}$ be $K$ index sets of $I = \{1, \ldots, N\}$. Let $f_k$ be marginal probability functions defined on tuples $h_k = (x_i : i \in S_k)$, $k = 0, \ldots, K - 1$. $P$ is an extension of marginals $f_1, \ldots, f_K$ if and only if

$$f_k(h_k) = \sum_{(x_1, \ldots, x_N) \in \pi_I^{-1}(h_k)} P(x_1, \ldots, x_N)$$
Iterative Proportional Fitting

Let $(j) = j \mod K$. Let $S_0, \ldots, S_{K-1}$ be $K$ index sets of $I = \{1, \ldots, N\}$. Let the range sets for the variables be $L_1, \ldots, L_N$. Let $f_k$ be marginal probability functions defined on tuples $h_k = (x_i : i \in S_k) \in \times_{i \in S_k} L_i$, $k = 0, \ldots, K - 1$. Let $a_k$ be defined on the variables indexed by $S_k$, $a_k : \times_{i \in S_k} L_i \rightarrow [0, 1]$ satisfy

$$\sum_{(x_1, \ldots, x_N)} \prod_{k=0}^{K-1} a_k(\pi_{S_k}(x_1, \ldots, x_N)) = 1$$

For $j \geq K - 1$, iterative proportional fitting defines $a_K, a_{K+1}, \ldots, a_m : \times_{i \in S(m)} L_i \rightarrow [0, 1]$, $m = K, K + 1, \ldots$, by

$$a_{j+1}(h) = \frac{f_{(j+1)}(h)}{\sum_{(x_1, \ldots, x_N) \in \pi_{l}^{-1}(h)} \prod_{m=j+2-K}^{j} a_m(\pi_{S(m)}(x_1, \ldots, x_N))}$$
Iterative Proportional Fitting

For $j \geq K - 1$, iterative proportional fitting defines $a_K, a_{K+1}, \ldots, a_m : \prod_{i \in S(m)} L_i \to [0, 1]$, $m = K, K + 1, \ldots$, by

$$a_{j+1}(h) = \frac{f_{j+1}(h)}{\sum_{(x_1, \ldots, x_N) \in \pi^{-1}_j(h)} \prod_{m=j+2-K}^{j} a_m(\pi_{S(m)}(x_1, \ldots, x_N))}$$

$$a_0 \quad a_1 \quad \ldots \quad a_{K-1} \quad a_K \quad a_{K+1} \quad \ldots \quad a_{2K-1} \quad a_{2K} \quad a_{2K+1} \quad \ldots$$

$$f_0 \quad f_1 \quad \ldots \quad f_{K-1} \quad f_0 \quad f_1 \quad \ldots \quad f_{K-1} \quad f_0 \quad f_1 \quad \ldots$$
For $j \geq K - 1$, iterative proportional fitting defines $a_K, a_{K+1}, \ldots$, $a_m : \times_{i \in S(m)} L_i \rightarrow [0, 1]$, $m = K, K + 1, \ldots$, by

$$a_{j+1}(h) = \frac{f_{(j+1)}(h)}{\sum_{(x_1, \ldots, x_N) \in \pi_j^{-1}(h)} \prod_{m=j+2-K}^{j+1} a_m(\pi_{S(m)}(x_1, \ldots, x_N))}$$

$j + 2 - K \ldots, j$ indexes the last $K - 1$ $a$ functions not including the $a$ function associated with $f_{(j+1)}$

$$\hat{f}^{(j+1)}_{(j+1)}(h) = \sum_{(x_1, \ldots, x_N) \in \pi_j^{-1}(h)} \prod_{m=j+2-K}^{j+1} a_m(\pi_{S(m)}(x_1, \ldots, x_N))$$
Let $x_1, x_2, x_3$ be three binary $\{0, 1\}$ valued variables. Let marginals $f_0(x_1, x_2), f_1(x_1, x_3), f_2(x_2, x_3)$ be given. The $a$ functions are defined on the same domains as the marginals.

$$a_0(x_1, x_2), a_1(x_1, x_3), a_2(x_2, x_3)$$

$$(x_1, x_2) = (0, 0): \quad a_3(0, 0) = \frac{f_0(0, 0)}{a_1(0, 0)a_2(0, 0) + a_1(0, 1)a_2(0, 1)}$$

$$(x_1, x_2) = (0, 1): \quad a_3(0, 1) = \frac{f_0(0, 1)}{a_1(0, 0)a_2(1, 0) + a_1(0, 1)a_2(1, 1)}$$
Iterative Proportional Fitting

\[ a_{j+1}(h) = \frac{f(j+1)(h)}{\sum_{(x_1,\ldots,x_N) \in \pi^{-1}_{S(j+1)}(h)} \prod_{m=j+2-K}^{j} a_m(x_i : i \in S(m))} \]

- \( P^{j+1}(x_1, \ldots, x_N) = \prod_{m=j+2-K}^{j+1} a_m(x_i : i \in S(m)) \) is a probability function and extension of \( f_{j+1} \)
- The iterative process converges
- In the limit, \( P^j \) is an extension of all the marginals \( f_0, \ldots f_{K-1} \)
- It is the unique minimal information extension
Decomposable Graph

\[ l = \{1, 2, 3, 4, 5\} \]

\[
\begin{array}{|c|c|}
\hline
C_1 & \{1, 2, 5\} \quad 1 \perp 4 \quad | \quad 2, 5 \\
\hline
C_2 & \{2, 3, 5\} \quad 1 \perp 3 \quad | \quad 2, 5 \\
\hline
C_3 & \{3, 4, 5\} \quad 2 \perp 4 \quad | \quad 3, 5 \\
\hline
S_2 & \{2, 5\} \quad 1 \perp 4 \quad | \quad 3, 5 \\
\hline
S_3 & \{3, 5\} \quad 1 \perp 4 \quad | \quad 2, 3, 5 \\
\hline
\end{array}
\]

\[
P(x_i : i \in I) = P(x_i : i \in C_1)P(x_i : i \in C_2)P(x_i : i \in C_3) \]
\[
P(x_i : i \in S_2)P(x_i : i \in S_3) \]
\[
= P(x_i : i \in C_1)P(x_i : i \in C_2 - S_2 \mid S_2)P(x_i : i \in C_3 - S_3 \mid S_3)
\]
Decomposable Graph

In the conditional independence graph, there is an edge between node $i$ and $j$ if and only if $X_i$ and $X_j$ are conditionally independent given the rest of the variables.

\[
P_{12345}(x_1, x_2, x_3, x_4, x_5) = \frac{P_{125}(x_1, x_2, x_5)P_{235}(x_2, x_3, x_5)P_{345}(x_3, x_4, x_5)}{P_{25}(x_2, x_5)P_{35}(x_3, x_5)}
= P_{15}(x_1, x_5)P_{2|15}(x_2 \mid x_1, x_5)P_{3|25}(x_3 \mid x_2, x_5)P_{4|35}(x_4 \mid x_3, x_5)
\]
\[
P_{12345}(x_1, x_2, x_3, x_4, x_5) = \frac{P_{125}(x_1, x_2, x_5)P_{235}(x_2, x_3, x_5)P_{345}(x_3, x_4, x_5)}{P_{25}(x_2, x_5)P_{35}(x_3, x_5)}
\]
\[
= P_{15}(x_1, x_5)P_{2|15}(x_2 \mid x_1, x_5)P_{3|25}(x_3 \mid x_2, x_5)P_{4|35}(x_4 \mid x_3, x_5)
\]

Figure: 1: System H
\[
P_{12345}(x_1, x_2, x_3, x_4, x_5) = \frac{P_{125}(x_1, x_2, x_5)P_{235}(x_2, x_3, x_5)P_{345}(x_3, x_4, x_5)}{P_{25}(x_2, x_5)P_{35}(x_3, x_5)}
\]

\[
= P_{25}(x_2, x_5)P_{1|25}(x_1 \mid x_2, x_5)P_{3|25}(x_3 \mid x_2, x_5)P_{4|35}(x_4 \mid x_3, x_5)
\]
\[ P_{12345}(x_1, x_2, x_3, x_4, x_5) = \frac{P_{125}(x_1, x_2, x_5)P_{235}(x_2, x_3, x_5)P_{345}(x_3, x_4, x_5)}{P_{25}(x_2, x_5)P_{35}(x_3, x_5)} \]

\[ = P_{12}(x_1, x_2)P_{5|12}(x_5 | x_1, x_2)P_{3|25}(x_3 | x_2, x_5)P_{4|35}(x_4 | x_3, x_5) \]

\{235 : 25\}, \{345 : 35\}

Figure: 1: System I
\[
P_{12345}(x_1, x_2, x_3, x_4, x_5) = \frac{P_{125}(x_1, x_2, x_5)P_{235}(x_2, x_3, x_5)P_{345}(x_3, x_4, x_5)}{P_{25}(x_2, x_5)P_{35}(x_3, x_5)}
\]

\[
= P_{1\mid 25}(x_1 \mid x_2, x_5)P_{2\mid 35}(x_2 \mid x_3, x_5)P_{4\mid 35}(x_4 \mid x_3, x_5)P_{35}(x_3, x_5)
\]

Figure: 2: System E
\[
P_{12345}(x_1, x_2, x_3, x_4, x_5) = \frac{P_{125}(x_1, x_2, x_5)P_{235}(x_2, x_3, x_5)P_{345}(x_3, x_4, x_5)}{P_{25}(x_2, x_5)P_{35}(x_3, x_5)}
= P_{1|25}(x_1 | x_2, x_5)P_{2|35}(x_2 | x_3, x_5)P_{3|45}(x_3 | x_4, x_5)P_{45}(x_4, x_5)
\]

\{125 : 25\}, \{235 : 35\}

\textbf{Figure: 2: System L}
$P_{12345}(x_1, x_2, x_3, x_4, x_5) = \frac{P_{125}(x_1, x_2, x_5)P_{235}(x_2, x_3, x_5)P_{345}(x_3, x_4, x_5)}{P_{25}(x_2, x_5)P_{35}(x_3, x_5)} = P_{1|25}(x_1 | x_2, x_5)P_{2|35}(x_2 | x_3, x_5)P_{5|34}(x_5 | x_3, x_4)P_{34}(x_3, x_4)$
\[ P_{12345}(x_1, x_2, x_3, x_4, x_5) = \frac{P_{125}(x_1, x_2, x_5)P_{235}(x_2, x_3, x_5)P_{345}(x_3, x_4, x_5)}{P_{25}(x_2, x_5)P_{35}(x_3, x_5)} \]

\[ = P_{1|25}(x_1 | x_2, x_5)P_{4|35}(x_4 | x_3, x_5)P_{2|35}(x_2 | x_3, x_5)P_{35}(x_3, x_5) \]
\[
P_{12345}(x_1, x_2, x_3, x_4, x_5) = \frac{P_{125}(x_1, x_2, x_5)P_{235}(x_2, x_3, x_5)P_{345}(x_3, x_4, x_5)}{P_{25}(x_2, x_5)P_{35}(x_3, x_5)}
= P_{1|25}(x_1 \mid x_2, x_5)P_{4|35}(x_4 \mid x_3, x_5)P_{3|25}(x_3 \mid x_2, x_5)P_{25}(x_2, x_5)
\]

Figure: 3: System G
\{125 : 25\}, \{345 : 35\}

\[
P_{12345}(x_1, x_2, x_3, x_4, x_5) = \frac{P_{125}(x_1, x_2, x_5)P_{235}(x_2, x_3, x_5)P_{345}(x_3, x_4, x_5)}{P_{25}(x_2, x_5)P_{35}(x_3, x_5)}
= P_{1\cdot|25}(x_1 \mid x_2, x_5)P_{4\cdot|35}(x_4 \mid x_3, x_5)P_{5\cdot|23}(x_5 \mid x_2, x_3)P_{23}(x_2, x_3)
\]
Feed Forward System Conditional Independences

\[ P_{12345}^A(x_1, x_2, x_3, x_4, x_5) = P_{45}(x_4, x_5) P_{345}(x_3 | x_4, x_5) P_{125}(x_1 | x_2, x_5) P_{235}(x_2 | x_3, x_5) \]
\[ P_{12345}^E(x_1, x_2, x_3, x_4, x_5) = P_{35}(x_3, x_5) P_{435}(x_4 | x_3, x_5) P_{125}(x_1 | x_2, x_5) P_{235}(x_2 | x_3, x_5) \]
\[ P_{12345}^G(x_1, x_2, x_3, x_4, x_5) = P_{25}(x_2, x_5) P_{325}(x_3 | x_2, x_5) P_{125}(x_1 | x_2, x_5) P_{435}(x_4 | x_3, x_5) \]
\[ P_{12345}^H(x_1, x_2, x_3, x_4, x_5) = P_{15}(x_1, x_5) P_{215}(x_2 | x_1, x_5) P_{325}(x_3 | x_2, x_5) P_{435}(x_4 | x_3, x_5) \]
\[ P_{12345}^I(x_1, x_2, x_3, x_4, x_5) = P_{12}(x_1, x_2) P_{512}(x_5 | x_1, x_2) P_{325}(x_3 | x_2, x_5) P_{435}(x_4 | x_3, x_5) \]
\[ P_{12345}^J(x_1, x_2, x_3, x_4, x_5) = P_{23}(x_2, x_3) P_{125}(x_1 | x_2, x_5) P_{523}(x_5 | x_2, x_3) P_{435}(x_4 | x_3, x_5) \]
\[ P_{12345}^L(x_1, x_2, x_3, x_4, x_5) = P_{34}(x_3, x_4) P_{125}(x_1 | x_2, x_5) P_{235}(x_2 | x_3, x_5) P_{534}(x_5 | x_3, x_4) \]

These decompositions correspond to the same Decomposable Graphical Model

\[ P_{12345}(x_1, x_2, x_4, x_4, x_5) = \frac{P_{345}(x_3, x_4, x_5) P_{125}(x_1, x_2, x_5) P_{235}(x_2, x_3, x_5)}{P_{25}(x_2, x_5) P_{35}(x_3, x_5)} \]
Feedforward Systems: Bayesian Networks

System A

Associated Bayesian Network

System A

\[ P(x_1, x_2, x_3, x_4, x_5) = \]
\[ P_4(x_4)P_5(x_5)P_{3|45}(x_3 | x_4, x_5)P_{2|35}(x_2 | x_3, x_5)P_{1|25}(x_1 | x_2, x_5) \]

Bayesian Network

\[ P(x_1, x_2, x_3, x_4, x_5) = \]
\[ P_4(x_4)P_5(x_5)P_{3|45}(x_3 | x_4, x_5)P_{2|35}(x_2 | x_3, x_5)P_{1|25}(x_1 | x_2, x_5) \]
$J = \{1, \ldots, N\}$
- Input set of subsystem $k$ is $I_k$
- Output set of subsystem $k$ is $O_k$
- $I_k \cup O_k = J_k$
- $I_k \cap O_k = \emptyset$
- $O_m \cap O_n = \emptyset, \ m \neq n$

The system structure is defined by $\{(I_k, O_k, P_k)\}_{k=1}^{K}$
- Input Set $I_k$
- Output Set $O_k$
- Behavior $P_k$

$$P(x_j : j \in J) = P(x_m : m \in J - \bigcup_{k=1}^{K} O_k) \prod_{k=1}^{K} P_k(x_o : o \in O_k | x_i : i \in I_k)$$

The System Structure is Causal Structure
Causal Structure

System A:
4,5 are the direct cause of 3
2,5 are the direct cause of 1
3,5 are the direct cause of 2

\[ J_1 = \{3, 4, 5\} \]
\[ I_1 = \{4, 5\} \]
\[ O_1 = \{3\} \]
\[ J_2 = \{1, 2, 5\} \]
\[ I_2 = \{2, 5\} \]
\[ O_2 = \{1\} \]
\[ J_3 = \{2, 3, 5\} \]
\[ I_3 = \{3, 5\} \]
\[ O_3 = \{2\} \]
Causal Structure

**System A:**
4,5 are the direct cause of 3
2,5 are the direct cause of 1
3,5 are the direct cause of 2

\[
\begin{align*}
J_1 &= \{3, 4, 5\} \\
I_1 &= \{4, 5\} \\
O_1 &= \{3\} \\
J_2 &= \{1, 2, 5\} \\
I_2 &= \{2, 5\} \\
O_2 &= \{1\} \\
J_3 &= \{2, 3, 5\} \\
I_3 &= \{3, 5\} \\
O_3 &= \{2\}
\end{align*}
\]
Causal Structure

System A

$X_4, X_5$ is the direct cause of $X_3$
$X_2, X_5$ is the direct cause of $X_1$
$X_3, X_5$ is the direct cause of $X_2$
$X_4$ is an indirect cause of $X_1$
$X_1$ has no causal influence on $X_3$: $X_1 \not\rightarrow X_3$
$X_3$ has causal influence on $X_1$: $X_3 \rightarrow X_1$
Given $X_2, X_5$, $X_3$ has no causal influence on $X_1$: $X_3 \not\rightarrow X_1 \mid X_2, X_5$
Given $X_2, X_5$, $X_3$ is conditionally independent of $X_1$: $X_3 \perp X_1 \mid X_2, X_5$
$X_4, X_5$ is the direct cause of $X_3$

$X_2, X_5$ is the direct cause of $X_1$

$X_3, X_5$ is the direct cause of $X_2$

$X_4$ is an indirect cause of $X_1$

Given its parents, each variable is conditionally independent of its non-descendants

Given $X_3$ and $X_5$, $X_2$ is conditionally independent of $X_4$: $X_2 \perp X_4 \mid X_3, X_5$
Conditional Independence Structure

\[ P_{12345}(x_1, x_2, x_4, x_4, x_5) = \frac{P_{345}(x_3, x_4, x_5)P_{125}(x_1, x_2, x_5)P_{235}(x_2, x_3, x_5)}{P_{25}(x_2, x_5)P_{35}(x_3, x_5)} \]

\[ P_{24|35}(x_2, x_4 \mid x_3, x_5) = \sum_{x_1} \frac{P_{125}(x_1, x_2, x_5)P_{235}(x_2, x_3, x_5)P_{345}(x_3, x_4, x_5)}{P_{25}(x_2, x_5)P_{35}(x_3, x_5)P_{35}(x_3, x_5)} \]

\[ = \frac{P_{235}(x_2, x_3, x_5)P_{345}(x_3, x_4, x_5)}{P_{25}(x_2, x_5)P_{35}(x_3, x_5)P_{35}(x_3, x_5)}P_{25}(x_2, x_5) \]

\[ = \frac{P_{235}(x_2, x_3, x_5)P_{345}(x_3, x_4, x_5)}{P_{35}(x_3, x_5)P_{35}(x_3, x_5)} \]

\[ = P_{2|35}(x_2 \mid x_3, x_5)P_{4|35}(x_4 \mid x_3, x_5) \]
Let \( \{ (I_k, O_k, R_k) \}_{k=1}^K \) be a system.

- Input Set \( I_k \)
- Output Set \( O_k \)
- Behavior \( P_k \)

Define the associated system digraph \((J, E)\) by

\[
\begin{align*}
J &= \bigcup_{k=1}^K I_k \cup O_k \\
E &= \bigcup_{k=1}^K I_k \times O_k
\end{align*}
\]

**Definition**

A system \( \{ (I_k, O_k, R_k) \} \) is called a **feedforward** system if and only if the digraph \((J, E)\) is acyclic. A system that is not feedforward is called a **feedback** system.
Possible Causal System Structure

Let us consider all the possibilities where each subsystem has exactly one output variable and no two different subsystems produce the same output variables.

<table>
<thead>
<tr>
<th>System</th>
<th>subsystem</th>
<th>output</th>
<th>subsystem</th>
<th>output</th>
<th>subsystem</th>
<th>output</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>345</td>
<td>3</td>
<td>235</td>
<td>2</td>
<td>125</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>345</td>
<td>3</td>
<td>235</td>
<td>2</td>
<td>125</td>
<td>5</td>
</tr>
<tr>
<td>C</td>
<td>345</td>
<td>3</td>
<td>235</td>
<td>5</td>
<td>125</td>
<td>1</td>
</tr>
<tr>
<td>D</td>
<td>345</td>
<td>3</td>
<td>235</td>
<td>5</td>
<td>125</td>
<td>2</td>
</tr>
<tr>
<td>E</td>
<td>345</td>
<td>4</td>
<td>235</td>
<td>2</td>
<td>125</td>
<td>1</td>
</tr>
<tr>
<td>F</td>
<td>345</td>
<td>4</td>
<td>235</td>
<td>2</td>
<td>125</td>
<td>5</td>
</tr>
<tr>
<td>G</td>
<td>345</td>
<td>4</td>
<td>235</td>
<td>3</td>
<td>125</td>
<td>1</td>
</tr>
<tr>
<td>H</td>
<td>345</td>
<td>4</td>
<td>235</td>
<td>3</td>
<td>125</td>
<td>2</td>
</tr>
<tr>
<td>I</td>
<td>345</td>
<td>4</td>
<td>235</td>
<td>3</td>
<td>125</td>
<td>5</td>
</tr>
<tr>
<td>J</td>
<td>345</td>
<td>4</td>
<td>235</td>
<td>5</td>
<td>125</td>
<td>1</td>
</tr>
<tr>
<td>K</td>
<td>345</td>
<td>4</td>
<td>235</td>
<td>5</td>
<td>125</td>
<td>2</td>
</tr>
<tr>
<td>L</td>
<td>345</td>
<td>5</td>
<td>235</td>
<td>2</td>
<td>125</td>
<td>1</td>
</tr>
<tr>
<td>M</td>
<td>345</td>
<td>5</td>
<td>235</td>
<td>3</td>
<td>125</td>
<td>1</td>
</tr>
<tr>
<td>N</td>
<td>345</td>
<td>5</td>
<td>235</td>
<td>3</td>
<td>125</td>
<td>2</td>
</tr>
</tbody>
</table>
System Diagrams

(a) System A: Feedfoward

(b) System B: Feedback

(c) System C: Feedback

(d) System D: Feedback
System Diagrams

(e) System E: Feedfoward

(f) System F: Feedback

(g) System G: Feedfoward

(h) System H: Feedfoward
System Diagrams

(i) System I: Feedforward

(j) System J: Feedforward

(k) System K: Feedback

(l) System L: Feedforward
System Diagrams

(m) System M: Feedback

(n) System N: Feedforward
Remove any subsystem not part of the feedback loop

Break the feedback loop
  - This prevents the output variable $y$ of the feedback loop to connect to a prior subsystem input variable $x$.
  - This makes the system a feedforward system

Calculate the feedforward system behavior

Add the equation $x = y$

Calculate the new results
System C

System C with subsystem 125 removed and feedback loop broken. Output variable 3 renamed to 6.

- Variable $x_k$ has $N_k$ possible values
- Fix variables $x_2 = a_2$ and $x_4 = a_4$
- Use a matrix notation
<table>
<thead>
<tr>
<th>Matrix Notation Conventions</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{6</td>
</tr>
<tr>
<td>$P_{5</td>
</tr>
<tr>
<td>$P_{6</td>
</tr>
<tr>
<td>$P_{3</td>
</tr>
</tbody>
</table>
The feedforward matrix equation relating the output variable $x_6$ to the input variable $x_3$ when input variable $x_2$ is fixed to value $a_2$ and input variable $x_4$ is fixed to value $a_4$ is then

$$P_6|_{N_6 \times 1}^{N_6 \times 1} x_2 = a_2 \quad x_4 = a_4 \quad P_6|_{N_6 \times N_5}^{N_6 \times N_5} x_4 = a_4 \quad P_5|_{N_5 \times 3}^{N_5 \times 3} x_2 = a_2 \quad x_4 = a_4 \quad P_3|_{N_3 \times 1}^{N_3 \times 1} x_2 = a_2 \quad x_4 = a_4$$
Set variable $x_6 = x_3$, noting that $N_6 = N_3$ and that variable $x_6$ and $x_3$ have the same range sets. The resulting matrix equation is

$$P_3|_{N_3 \times 1} \begin{cases} x_2 = a_2 \\ x_4 = a_4 \end{cases} = P_3|_{N_3 \times N_5} \begin{cases} x_4 = a_4 \end{cases} P_5|_{N_5 \times N_3} \begin{cases} x_2 = a_2 \\ x_4 = a_4 \end{cases}$$

This equation can be easily solved for $P_3$ since it is the eigenvector corresponding to eigenvalue of 1 of the matrix

$$P_3|_{N_3 \times N_5} \begin{cases} x_4 = a_4 \end{cases} P_5|_{N_5 \times N_3} \begin{cases} x_2 = a_2 \end{cases}$$
Thus for each different value of the externally set input variables \( x_2 \) and \( x_4 \), there will be different distribution for \( x_3 \). Once, the distribution of \( x_3 \) is known, the joint distribution of all variables, can be calculated by means of the corresponding conditional probabilities.

\[
P_3 \mid_{x_2 = a_2, x_4 = a_4}^{N_3 \times 1} \text{ is really the conditional probability } P_{3\mid24}(x_3 \mid a_2, a_4).
\]

\[
P_{12345}(x_1, x_2, x_3, x_4, x_5) = P_{1\mid25}(x_1 \mid x_2, x_5) P_{3\mid24}(x_3, x_2, x_4) P_{5\mid23}(x_5 \mid x_2, x_3) P_{24}(x_2, x_4)
\]
Fix the external variables \( x_1 = a_1 \) and \( x_4 = a_4 \).

Take the combined variable \((x_2, x_3)\) as the feedback variable.

The conditional probability matrix for \((x_2, x_3)\) given \( x_5 \) is \( N_2 N_3 \times N_5 \).

\[
P_{23|5}^{N_2 N_3 \times N_5}_{x_1 = a_1, x_4 = a_4} = P_{2|15}^{N_2 \times N_5}_{x_1 = a_1} \otimes P_{3|45}^{N_3 \times N_5}_{x_4 = a_4}
\]

where \( \otimes \) is the kronecker matrix product and simply allows us to denote a conditional probability matrix where one of the variables is the joint variable \((x_2, x_3)\).
First we break the feedback loops and rename the output variable $x_5$ to $x_6$. Now we can write

$$P_6|_{N_6 \times 1} \begin{bmatrix} x_1 = a_1 \\ x_4 = a_4 \end{bmatrix} = P_6|_{23} \begin{bmatrix} x_1 = a_1 \\ x_4 = a_4 \end{bmatrix} N_6 \times N_2 N_3$$

$$= P_{23}|_{5} \begin{bmatrix} x_1 = a_1 \\ x_4 = a_4 \end{bmatrix} N_2 N_3 \times N_5 P_5|_{N_5 \times 1} \begin{bmatrix} x_1 = a_1 \\ x_4 = a_4 \end{bmatrix}$$

Now we connect the feedback loop. We set variable $x_6 = x_5$, noting that $N_6 = N_5$ and that variable $x_6$ and $x_5$ have the same range sets. The resulting matrix equation is

$$P_5|_{N_5 \times 1} \begin{bmatrix} x_1 = a_1 \\ x_4 = a_4 \end{bmatrix} = P_5|_{23} \begin{bmatrix} x_1 = a_1 \\ x_4 = a_4 \end{bmatrix} N_5 \times N_2 N_3$$

$$= P_{23}|_{5} \begin{bmatrix} x_1 = a_1 \\ x_4 = a_4 \end{bmatrix} N_2 N_3 \times N_5 P_5|_{N_5 \times 1} \begin{bmatrix} x_1 = a_1 \\ x_4 = a_4 \end{bmatrix}$$
As before, this equation is easily solved as $P_5|_{N_5 \times 1}^{N_5 \times 1}$ is just the eigenvector having eigenvalue 1 of the matrix

\[
P_5|_{N_5 \times 1}^{N_5 \times 1} = P_5|_{N_5 \times 1}^{N_5 \times 1} \quad x_1 = a_1 \\
\quad x_4 = a_4 \\
N_5 \times N_2 N_3 \\
P_23|_{N_5 \times 1}^{N_5 \times 1} \quad x_1 = a_1 \\
\quad x_4 = a_4 \\
P_5|_{N_5 \times 1}^{N_5 \times 1} 
\]

\[
P_5|_{N_5 \times 1}^{N_5 \times 1} = P_5|_{N_5 \times 1}^{N_5 \times 1} \quad x_1 = a_1 \\
\quad x_4 = a_4 \\
N_5 \times N_2 N_3 \\
P_23|_{N_5 \times 1}^{N_5 \times 1} \quad x_1 = a_1 \\
\quad x_4 = a_4 \\
P_5|_{N_5 \times 1}^{N_5 \times 1} 
\]

\[
P_5|_{N_5 \times 1}^{N_5 \times 1} = P_5|_{N_5 \times 1}^{N_5 \times 1} \quad x_1 = a_1 \\
\quad x_4 = a_4 \\
N_5 \times N_2 N_3 \\
P_23|_{N_5 \times 1}^{N_5 \times 1} \quad x_1 = a_1 \\
\quad x_4 = a_4 \\
P_5|_{N_5 \times 1}^{N_5 \times 1} 
\]
Multiple Connected Feedback Loops: Joint Probability

\[ P_{5|N_5}^{x_1, x_4} \] is the conditional probability \( P_{5|14}(x_5|a_1, a_4) \)

\[ P_{12345}(x_1, x_2, x_3, x_4, x_5) = P_{5|14}(x_5|a_1, x_4)P_{14}(x_1, x_4)P_{2|15}(x_2|a_1, x_5)P_{3|45}(x_3|a_4, x_5) \]