Probability Models

Robert M. Haralick

Computer Science, Graduate Center
City University of New York

Outline

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

$$\begin{aligned} p(y_1 \mid y_2 \ldots y_N) &= P(y_1 \mid y_2) \\ p(y_2 \mid y_3 \ldots y_N) &= P(y_2 \mid y_3) \\ &\vdots \\ P(y_{N-2} \mid y_{N-1}, y_N) &= P(y_{N-2} \mid y_{N-1}) \end{aligned}$$

In general,

$$P(y_n \mid y_{n+1} \ldots y_N) = P(y_n \mid y_{n+1}), n = 1, \ldots N-1$$

Conditional Probability

Now,

$$\begin{aligned} P(x_1 \ldots x_N) &= P(x_1 \,| x_2 \ldots x_N) P(x_2 \ldots x_N) \\ &= P(x_1 \,| x_2 \ldots x_N) P(x_2 \mid x_3 \ldots x_N) P(x_3 \ldots x_N) \end{aligned}$$

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

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

Hence,

$$\begin{aligned} P(x_1 \ldots x_N) &= \left[\prod_{n=1}^{N-1} P(x_n \,| x_{n+1} \ldots x_N)\right] P(x_N) \\ &= \left[\prod_{n=1}^{N-1} P(x_n \,| x_{n+1})\right] P(x_N) \end{aligned}$$

The Markov Classifier

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

$$\begin{aligned} P(x_1 \ldots x_N \mid c^*) &> P(x_1 \ldots x_N \mid c), c \neq c^* \\ \left[\prod_{n=1}^{N-1} P(x_n \mid x_{n+1}, c^*)\right] P(x_N \mid c^*) &> \left[\prod_{n=1}^{N-1} P(x_n \mid x_{n+1}, c)\right] P(x_N \mid c) \end{aligned}$$

for all other c

The General Markov Classifier

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

$$\begin{aligned} P(x_1 \ldots x_N \mid c^*) &> P(x_1 \ldots x_N \mid c), c \neq c^* \\ \left[\prod_{n=1}^{N-1} P(x_{i_n} | x_{i_{n+1}}, c^*)\right] P(x_{i_N} | c^*) &> \left[\prod_{n=1}^{N-1} P(x_{i_n} | x_{i_{n+1}}, c)\right] P(x_{i_N} | c) \end{aligned}$$

for all other c

Conditional Independence Assumption

Under the Markov assumption

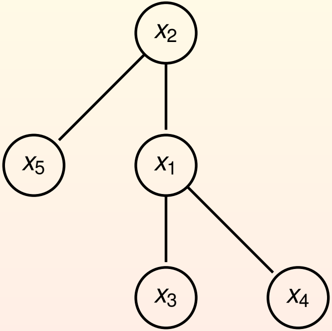
$$
\begin{aligned}
P(x_i, x_{i+1}, \mid x_{i+2} \ldots, x_N) &= \frac{P(x_i, \ldots x_N)}{P(x_{i+2} \ldots x_N)} \\
&= \frac{P(x_i \mid x_{i+1} \ldots x_N) P(x_{i+1} \ldots x_N)}{P(x_{i+2} \ldots x_N)} \\
&= \frac{P(x_i \mid x_{i+1}) P(x_{i+1} \ldots x_N)}{P(x_{i+2} \ldots x_N)} \\
&= \frac{P(x_i \mid x_{i+1}) P(x_{i+1} \mid x_{i+2}) P(x_{i+2} \ldots x_N)}{P(x_{i+2} \ldots x_N)} \\
&= P(x_i \mid x_{i+1}) P(x_{i+1} \mid x_{i+2})
\end{aligned}
$$

The General Markov Classifier

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

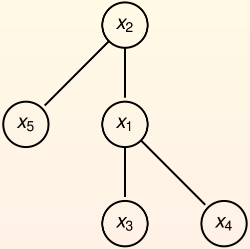


$$P(x_1, x_2, x_3, x_4, x_5) = p(x_1 \mid x_2)P(x_5 \mid x_2)P(x_3 \mid x_1)P(x_4 \mid x_1)P(x_2)$$

First Order Dependence Trees

$$
\begin{aligned}
1 &= \sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_4}\sum_{x_5} P(x_1, x_2, x_3, x_4, x_5) \\
& \sum_{x_1}\sum_{x_2}\sum_{x_3}\sum_{x_4}\sum_{x_5} p(x_1 \mid x_2) P(x_5 \mid x_2) P(x_3 \mid x_1) P(x_4 \mid x_1) P(x_2) \\
& \sum_{x_2} P(x_2) \sum_{x_1} P(x_1 \mid x_2) \sum_{x_5} P(x_5 \mid x_2) \sum_{x_4} P(x_4 \mid x_1) \sum_{x_3} P(x_3 \mid x_1) \\
&= 1
\end{aligned}
$$

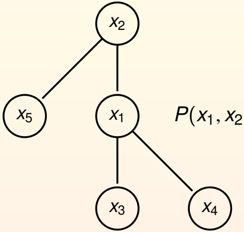
First Order Dependence Trees



Precedence Function

i	**j(i)**
1	2
5	2
3	1
4	1

First Order Dependence Tree



$$P(x_1, x_2, x_3, x_4, x_5) = p(x_1 \mid x_2)P(x_5 \mid x_2)P(x_3 \mid x_1)P(x_4 \mid x_1)P(x_2)$$

$$\begin{aligned}
[N] &= \{1, \ldots, N\} \\
M &\subset [N] \qquad j : M \to N \\
G &= ([N], E) \\
E &= \{\{j(m), m\} \mid m \in M\} \\
P(x_1, \ldots, x_N) &= P(x_m : m \in [N] - M) \prod_{m \in M} P(x_m \mid x_{j(m)})
\end{aligned}$$

The Optimal Dependence Tree

Of all possible dependence trees, is there an optimal one?

- The probabilities we are interested in are all conditional on class
- To save writing longer expression, we omit the class
- The probabilities in our dependence tree are order 2 $P(x_a, x_b)$
- The joint probability formed from the dependence tree product must be an extension of the marginal forming it
- The probability we want the product form to approximate is unknown
- The dependence tree product we seek must be as close to it as possible
- Given the complete set of the order 2 marginals,
  - The closest dependence tree product we can construct must use the marginals with largest entropy
  - But finding that out is a combinatorial problem

Dependence Tree Optimization

There are two papers In 1968 Chow and Liu published a major paper *Approximating Discrete Probability Distributions with Dependence Trees* in the IEEE Transactions on Information Theory.

The Lewis paper, published in Information And Control, appeared in 1959 and had the title *Approximating Probability Distributions to Reduce Storage Requirements*

proved that

- If we are just given low order marginals
- Whose product is a probability and extension of the marginals
- And we are approximating an unknown joint distribution
- Then the best we can do is a minimum information extension

Dependence Tree Optimization

In 1968 Chow and Liu published a major paper *Approximating Discrete Probability Distributions with Dependence Trees* in the IEEE Transactions on Information Theory. The paper gave the algorithm for forming the maximum mutual information extension.

- For every order two marginal, compute its Mutual Information
  - The Mutual Information between to variables x and y is defined by

$$I(x, y) = \sum_{x,y} P(x, y) \log \frac{P(x, y)}{P(x)P(y)}$$

- Make a graph whose nodes are labeled by the variable name
- Connect every pair of nodes, say node x with node y, with an edge
- Weight the edge by $I(x, y)$
- Use Kruskal's algorithm to find the maximum weighted

Kruskal's Maximal Spanning Tree Algorithm

- Sort the edges by decreasing weight
- Select the first edge as having the largest Mutual Information
- Select the next largest successive edge, that does not form a loop with the edges that have been previously chosen
- Stop when there are $N - 1$ edges where N is the number of nodes

Kruskal proved that when the algorithm stopped, the result was a spanning tree. Further the spanning tree was maximal.

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?

Define,

$$
\begin{aligned}
Q(x_1, \ldots, x_6) &= P(x_1 \mid x_2, x_3, x_4) P(x_4 \mid x_2, x_5, x_6) P(x_5 \mid x_6) \\
&\quad P(x_2, x_6 \mid x_3) P(x_3)
\end{aligned}
$$

$$
\begin{aligned}
Q(x_2, \ldots, x_6) &= \sum_{x_1} Q(x_1, \ldots, x_6) \\
&= \sum_{x_1} P(x_1 \mid x_2, x_3, x_4) P(x_4 \mid x_2, x_5, x_6) \\
&\quad P(x_5 \mid x_6) P(x_2, x_6 \mid x_3) P(x_3) \\
&= P(x_4 \mid x_2, x_5, x_6) P(x_5 \mid x_6) P(x_2, x_6 \mid x_3) P(x_3)
\end{aligned}
$$

Summing

$$\begin{aligned} Q(x_2, x_3, x_5, x_6) &= \sum_{x_4} Q(x_2, \ldots, x_6) \\ &= \sum_{x_4} P(x_4 \mid x_2, x_5, x_6) P(x_5 \mid x_6) P(x_2, x_6 \mid x_3) P(x_3) \\ &= P(x_5 \mid x_6) P(x_2, x_6 \mid x_3) P(x_3) \end{aligned}$$

Summing

$$
\begin{aligned}
Q(x_2, x_3, x_6) &= \; = \sum_{x_5} Q(x_2, x_3, x_5, x_6) \\
&= \sum_{x_5} P(x_5 \mid x_6) P(x_2, x_6 \mid x_3) P(x_3) \\
&= P(x_2, x_6 \mid x_3) P(x_3) \\
&= P(x_2, x_3, x_6) \\
\sum_{x_2, x_3, x_6} Q(x_2, x_3, x_6) &= \sum_{x_2, x_3, x_6} P(x_2, x_3, x_6) \\
&= 1
\end{aligned}
$$

Weak Extension

Define,

$$\begin{aligned} Q(x_1, \ldots, x_6) &= P(x_1 \mid x_2, x_3, x_4) P(x_4 \mid x_2, x_5, x_6) P(x_5 \mid x_6) \\ &\quad P(x_2, x_6 \mid x_3) P(x_3) \end{aligned}$$

Is,

$$\begin{aligned} Q(x_1 \mid x_2, x_3, x_4) &= P(x_1 \mid x_2, x_3, x_4) \\ Q(x_4 \mid x_2, x_5, x_6) &= P(x_4 \mid x_2, x_5, x_6) \\ Q(x_5 \mid x_6) &= P(x_5 \mid x_6) \\ Q(x_2, x_6 \mid x_3) &= P(x_2, x_6 \mid x_3) \\ Q(x_3) &= P(x_3) \end{aligned}$$

Strong Extension

Define,

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

Is,

$$\begin{aligned}
Q(x_1, x_2, x_3, x_4) &= P(x_1, x_2, x_3, x_4) \\
Q(x_4, x_2, x_5, x_6) &= P(x_4, x_2, x_5, x_6) \\
Q(x_5, x_6) &= P(x_5, x_6) \\
Q(x_2, x_3, x_6) &= P(x_2, x_3, x_6) \\
Q(x_3) &= P(x_3)
\end{aligned}$$

Extension

Define,

$$\begin{aligned} Q(x_1, \ldots, x_6) &= P(x_1 \mid x_2, x_3, x_4) P(x_4 \mid x_2, x_5, x_6) \\ &\qquad P(x_5 \mid x_6) P(x_2, x_6 \mid x_3) P(x_3) \end{aligned}$$

Is

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

Find expressions for $Q(x_1, x_2, x_3, x_4)$ and $Q(x_2, x_3, x_4)$

$$\begin{aligned} Q(x_1, x_2, x_3, x_4) &= \sum_{x_5} \sum_{x_6} Q(x_1, \ldots, x_6) \\ &= \sum_{x_5} \sum_{x_6} P(x_1 \mid x_2, x_3, x_4) P(x_4 \mid x_2, x_5, x_6) \\ &\qquad P(x_5 \mid x_6) P(x_2, x_6 \mid x_3) P(x_3) \\ &= P(x_1 \mid x_2, x_3, x_4) \sum_{x_5} \sum_{x_6} P(x_4 \mid x_2, x_5, x_6) \\ &\qquad P(x_5 \mid x_6) P(x_2, x_6 \mid x_3) P(x_3) \end{aligned}$$

Extension

$$
\begin{aligned}
Q(x_2, x_3, x_4) &= \sum_{x_1} Q(x_1, x_2, x_3, x_4) \\
&= \sum_{x_1} P(x_1 \mid x_2, x_3, x_4) \sum_{x_5} \sum_{x_6} P(x_4 \mid x_2, x_5, x_6) \\
&\quad P(x_5 \mid x_6) P(x_2, x_6 \mid x_3) P(x_3) \\
&= \sum_{x_5} \sum_{x_6} P(x_4 \mid x_2, x_5, x_6) \\
&\quad P(x_5 \mid x_6) P(x_2, x_6 \mid x_3) P(x_3)
\end{aligned}
$$

Weak Extension

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

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

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

Conditional Independence Notation

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

- $X \perp\!\!\!\perp Y \mid Z$

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

- $i \perp\!\!\!\perp j \mid k$

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

- $A \perp\!\!\!\perp B \mid C$

Conditional Independence Characterization Theorem

Theorem

$P(x, y|z) = P(x|z)P(y|z)$ if and only if $P(x|y, z) = P(x|z)$

Proof.

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

$$\begin{aligned} P(x|y,z) &= \frac{P(x,y,z)}{P(y,z)} = \frac{P(x,y|z)P(z)}{P(y,z)} \\ &= \frac{P(x|z)P(y|z)P(z)}{P(y,z)} = P(x|z) \end{aligned}$$

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

$$\begin{aligned} P(x,y|z) &= \frac{P(x,y,z)}{P(z)} = \frac{P(x|y,z)P(y,z)}{P(z)} \\ &= \frac{P(x|z)P(y,z)}{P(z)} = P(x|z)P(y|z) \end{aligned}$$

□

Conditional Independence

Can we see if x_5 is conditionally independent of x_2 given x_6. Is $P(x_5 \mid x_2, x_6) = P(x_5 \mid x_6)$?

$$\begin{aligned}
Q(x_2, x_4, x_5, x_6) &= \sum_{x_1}\sum_{x_3} Q(x_1, \ldots, x_6) \\
&= \sum_{x_1}\sum_{x_3} P(x_1 \mid x_2, x_3, x_4) P(x_4 \mid x_2, x_5, x_6) \\
&\qquad P(x_5 \mid x_6) P(x_2, x_6 \mid x_3) P(x_3) \\
&= P(x_4 \mid x_2, x_5, x_6) P(x_5 \mid x_6) P(x_2, x_6)
\end{aligned}$$

Extension

$$
\begin{aligned}
Q(x_2, x_5, x_6) &= \sum_{x_4} Q(x_2, x_4, x_5, x_6) \\
&= \sum_{x_4} P(x_4 \mid x_2, x_5, x_6) P(x_5 \mid x_6) P(x_2, x_6) \\
&= P(x_5 \mid x_6) P(x_2, x_6)
\end{aligned}
$$

Extension

$$
\begin{aligned}
Q(x_5, x_6) &= \sum_{x_1} \sum_{x_2} \sum_{x_3} \sum_{x_4} Q(x_1, \ldots, x_6) \\
&= \sum_{x_1} \sum_{x_2} \sum_{x_3} \sum_{x_4} P(x_1 \mid x_2, x_3, x_4) P(x_4 \mid x_2, x_5, x_6) \\
&\qquad P(x_5 \mid x_6) P(x_2, x_6 \mid x_3) P(x_3) \\
&= \sum_{x_2} \sum_{x_3} \sum_{x_4} P(x_4 \mid x_2, x_5, x_6) P(x_5 \mid x_6) P(x_2, x_3, x_6) \\
&= \sum_{x_2} \sum_{x_3} P(x_5 \mid x_6) P(x_2, x_3, x_6) \\
&= P(x_5 \mid x_6) P(x_6) = P(x_5, x_6)
\end{aligned}
$$

Conditional Independences

$$\begin{aligned}
Q(x_2, x_5, x_6) &= P(x_5 \mid x_6) P(x_2, x_6) \\
Q(x_2, x_6) &= \sum_{x_5} Q(x_2, x_5, x_6) \\
&= \sum_{x_5} P(x_5 \mid x_6) P(x_2, x_6) \\
&= P(x_2, x_6) \\
Q(x_5 \mid x_2, x_6) &= \frac{Q(x_2, x_5, x_6)}{Q(x_2, x_6)} \\
&= \frac{P(x_5 \mid x_6) P(x_2, x_6)}{P(x_2, x_6)} = P(x_5 \mid x_6)
\end{aligned}$$

Conditional Independence

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,

$$Q(x_5 \mid x_2, x_6) = Q(x_5 \mid x_6)$$

$$x_5 \perp\!\!\!\perp x_2 \mid x_6$$

Conditional Independence

Suppose, $x_5 \perp\!\!\!\perp x_2 \mid x_6$

$$Q(x_5 \mid x_2, x_6) = Q(x_5 \mid x_6)$$

Then,

$$Q(x_5, x_2 \mid x_6) = Q(x_5 \mid x_6) Q(x_2 \mid x_6)$$

$$\begin{aligned}
Q(x_5, x_2 \mid x_6) &= \frac{Q(x_2, x_5, x_6)}{Q(x_6)} \\
&= \frac{Q(x_5 \mid x_2, x_6) Q(x_2, x_6)}{Q(x_6)} \\
&= \frac{Q(x_5 \mid x_6) Q(x_2, x_6)}{Q(x_6)} \\
&= Q(x_5 \mid x_6) Q(x_2 \mid x_6)
\end{aligned}$$

Additional Relationships You Work Out

$$
\begin{aligned}
Q(x_4 \mid x_2, x_5, x_6) &= \frac{Q(x_2, x_4, x_5, x_6)}{Q(x_2, x_5, x_6)} \\
&= \frac{P(x_4 \mid x_2, x_5, x_6)P(x_5 \mid x_6)P(x_2, x_6)}{P(x_5 \mid x_6)P(x_2, x_6)} \\
&= P(x_4 \mid x_2, x_5, x_6)
\end{aligned}
$$

Additional Relationships You Work Out

$$
\begin{aligned}
Q(x_2, x_3, x_6) &= \sum_{x_1} \sum_{x_4} \sum_{x_5} Q(x_1, \ldots, x_6) \\
&= \sum_{x_1} \sum_{x_4} \sum_{x_5} P(x_1 \mid x_2, x_3, x_4) P(x_4 \mid x_2, x_5, x_6) \\
&\qquad P(x_5 \mid x_6) P(x_2, x_6 \mid x_3) P(x_3) \\
&= \sum_{x_4} \sum_{x_5} P(x_4 \mid x_2, x_5, x_6) P(x_5 \mid x_6) P(x_2, x_3, x_6) \\
&= \sum_{x_5} P(x_5 \mid x_6) P(x_2, x_3, x_6) \\
&= P(x_2, x_3, x_6)
\end{aligned}
$$

Graphical Models

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

Graphs

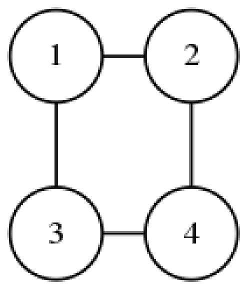
Definition

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.

Graphs

Here, $G = (N, E)$ where

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

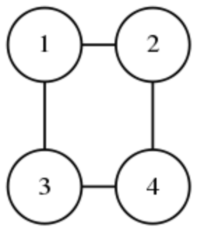


Boundary

Definition

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

$$bndry(i) = \{j \in N \mid \{i, j\} \in E\}$$



- $bndry(1) = \{2, 3\}$
- $bndry(2) = \{1, 4\}$
- $bndry(3) = \{1, 4\}$
- $bndry(4) = \{2, 3\}$

Conditional Independence Graph: Definition

Definition

A graph (N, E) is called a Conditional Independence Graph of a random variable set $\mathcal{X} = \{X_1, \ldots, X_M\}$ if and only if $N = \{1, \ldots, M\}$, the index set for the variables in $\mathcal{X}$, and

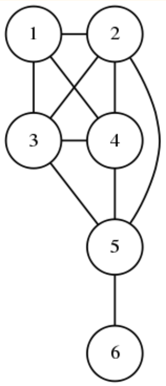
$$E^c = \{\{i, j\} \mid X_i \perp\!\!\!\perp X_j \mid \mathcal{X} - \{X_i, X_j\}\}$$

All graphs we discuss will be conditional independence graphs.

Conditional Independence Graph

Nodes correspond to indexes of variables in the variable set $\mathcal{X} = \{X_1, \ldots, X_6\}$

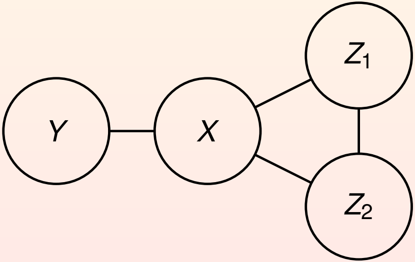
$\{i, j\}$ not in the edge set means $X_i \perp\!\!\!\perp X_j \mid \mathcal{X} - \{X_i, X_j\}$



Conditional Independence Graph

$\{Y, Z_1\}$ and $\{Y, Z_2\}$ not in edge set means

$$
\begin{aligned}
Y \perp\!\!\!\perp Z_1 &\quad | \quad \{X, Y, Z_1, Z_2\} - \{Y, Z_1\} \\
Y \perp\!\!\!\perp Z_2 &\quad | \quad \{X, Y, Z_1, Z_2\} - \{Y, Z_2\} \\
Y \perp\!\!\!\perp Z_1 &\quad | \quad \{X, Z_2\} \\
Y \perp\!\!\!\perp Z_2 &\quad | \quad \{X, Z_1\}
\end{aligned}
$$

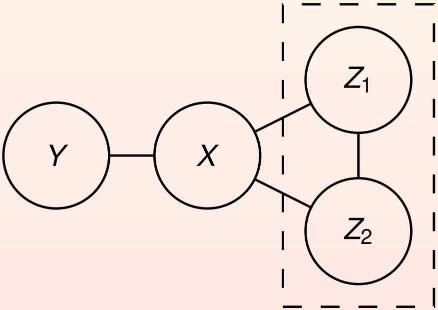


Block Independence Theorem

Y is conditionally independent of the block $\{Z_1, Z_2\}$ given X

Theorem

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

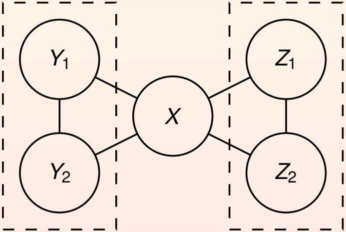


Reduction Theorem

Theorem

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

- $Y \perp\!\!\!\perp Z_1, Z_2 \mid X$ *if and only if* $Y \perp\!\!\!\perp Z_1 \mid X, Z_2$ *and* $Y \perp\!\!\!\perp Z_2 \mid X, Z_1$.
- $Y \perp\!\!\!\perp Z_1, Z_2 \mid X$ *implies* $Y \perp\!\!\!\perp Z_1 \mid X$ *and* $Y \perp\!\!\!\perp Z_2 \mid X$.



- $Y_1 \perp\!\!\!\perp Z_1 \mid X,\ Y_1 \perp\!\!\!\perp Z_2 \mid X,\ Y_2 \perp\!\!\!\perp Z_1 \mid X,\ Y_2 \perp\!\!\!\perp Z_2 \mid X$
- $Y_1, Y_2 \perp\!\!\!\perp Z_1 \mid X,\ Y_1, Y_2 \perp\!\!\!\perp Z_2 \mid X,\ Y_1, Y_2 \perp\!\!\!\perp Z_1, Z_2 \mid X$
- $Z_1, Z_2 \perp\!\!\!\perp Y_1 \mid X,\ Z_1, Z_2 \perp\!\!\!\perp Y_2 \mid X$

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 $n \in \{1, \ldots, N-1\}$.

Connectedness

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.

Separation

Definition

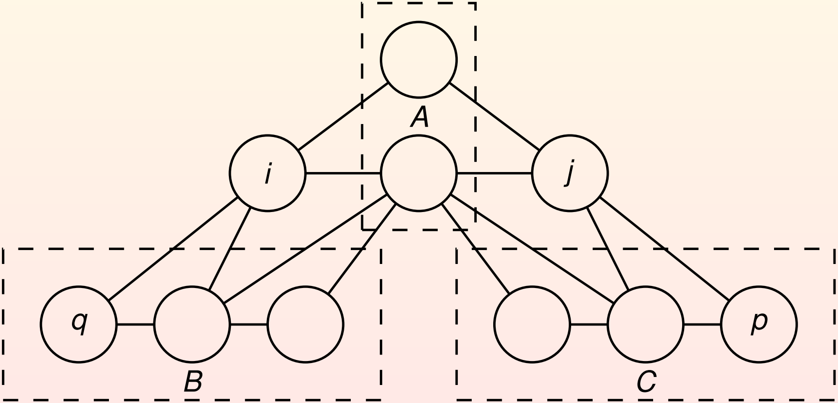
Let (G, E) be a graph and A, B, S be non-empty subsets of G. S separates A from B if and only if for every $a \in A$ and $b \in B$, every path in G that begins with a and ends with b has at least one node in S.

Separation Theorem

A separates $B \cup \{i\}$ from $C \cup \{j\}$

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

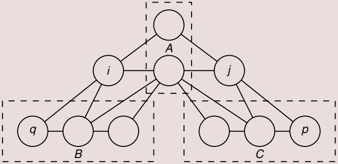
Then $i \perp\!\!\!\perp j \mid A$



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 $i \perp\!\!\!\perp j \mid A$.



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, $i \perp\!\!\!\perp j \mid N - \{i, j\}$ and $i \perp\!\!\!\perp p \mid N - \{i, p\}$. Application of the block independence theorem yields $i \perp\!\!\!\perp j, p \mid N - \{i, j, p\}$. Application of the reduction theorem yields $i \perp\!\!\!\perp j \mid N - \{i, j, p\}$. Repeated application using the remaining nodes of C yields $i \perp\!\!\!\perp j \mid N - \{i, j\} - C$. Similarly for using q. Repeated application yields $i \perp\!\!\!\perp j \mid N - \{i, j\} - B - C$. But $N - \{i, j\} - B - C = A$. Therefore $i \perp\!\!\!\perp j \mid A$. □

Local Markov Property

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 $B = N - \{n\} - bndry(n)$. Then $n \perp\!\!\!\perp B \mid bndry(n)$.

Proof.

The set $bndry(n)$ 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 $bndry(n)$.

Complete Graphs

Definition

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

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

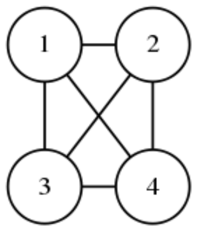


Figure: The Complete Graph on 4 Nodes

Graph Restriction

Definition

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

$$G\mid_A = (A, E\mid_A)$$

where

$$E\mid_A = \{\{i, j\} \in E \mid i, j \in A\}$$

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 $G \mid_A$ is a complete graph.

Maximally Complete

Definition

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.

Chordal Graphs

Definition

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

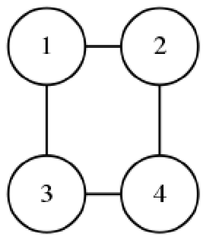


Figure: Non-chordal

Non-Chordal Graphs

Definition

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

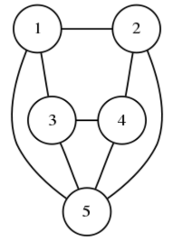


Figure: Non-chordal

Decomposable Graphs

Definition

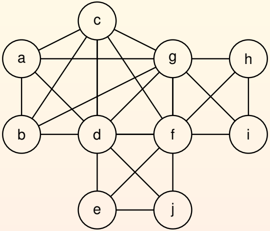
A Graph $G = (N, E)$ is Decomposable if and only if

- 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



C_1	$=$	$\{a, b, c, d, g\}$					
C_2	$=$	$\{c, d, f, g\}$	S_2	$=$	$C_2 \cap C_1$	$=$	$\{c, d, g\}$
C_3	$=$	$\{f, g, h, i\}$	S_3	$=$	$C_3 \cap (C_1 \cup C_2)$	$=$	$\{f, g\}$
C_4	$=$	$\{d, e, f, j\}$	S_4	$=$	$C_4 \cap (C_1 \cup C_2 \cup C_3)$	$=$	$\{d, f\}$

Notation

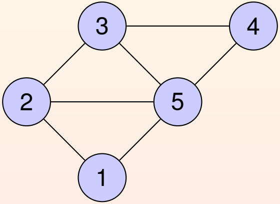
Let I be an index subset. If $I = \{1, 3, 7\}$, then

$$P(x_i : i \in I) = P(x_1, x_3, x_7)$$

Decomposable Graph

$$I = \{1, 2, 3, 4, 5\}$$

C_1	$=$	$\{1, 2, 5\}$	$1 \perp\!\!\!\perp 4$	$\mid$	$2, 5$
C_2	$=$	$\{2, 3, 5\}$	$1 \perp\!\!\!\perp 3$	$\mid$	$2, 5$
C_3	$=$	$\{3, 4, 5\}$	$2 \perp\!\!\!\perp 4$	$\mid$	$3, 5$
S_2	$=$	$\{2, 5\}$	$1 \perp\!\!\!\perp 4$	$\mid$	$3, 5$
S_3	$=$	$\{3, 5\}$	$1 \perp\!\!\!\perp 4$	$\mid$	$2, 3, 5$



$$\begin{aligned} P(x_i : i \in I) &= \frac{P(x_i : i \in C_1)P(x_i : i \in C_2)P(x_i : i \in C_3)}{P(x_i : i \in S_2)P(x_i : i \in S_3)} \\ &= P(x_i : i \in C_1)P(x_i : i \in C_2 - S_2 \mid S_2)P(x_i : i \in C_3 - S_3 \mid S_3) \end{aligned}$$

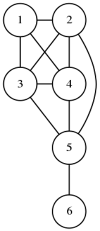
Decomposable Graphs

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

$$\begin{aligned} P(x_1, \ldots, x_N) &= \frac{\prod_{k=1}^{K} P(x_i : i \in C_k)}{\prod_{m=2}^{K} P(x_j : j \in S_m)} \\ &= P(x_i : i \in C_1) \prod_{k=2}^{K} P(x_i : i \in C_k - S_k \mid S_k) \end{aligned}$$

Example



Cliques in running intersection order: $\{1,2,3,4\}, \{2,3,4,5\}, \{5,6\}$
Separators: $\{2,3,4\}$, $\{5\}$

$$
\begin{aligned}
P(x_1,\ldots,x_6) &= \frac{P(x_1,x_2,x_3,x_4)P(x_2,x_3,x_4,x_5)P(x_5,x_6)}{P(x_2,x_3,x_4)P(x_5)} \\
&= P(x_1,x_2,x_3,x_4)P(x_5 \mid x_2,x_3,x_4)P(x_6 \mid x_5) \\
&= P(x_2,x_3,x_4,x_5)P(x_1 \mid x_2,x_3,x_4)P(x_6 \mid x_5)
\end{aligned}
$$

Product Form

The product form

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

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

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

$$\begin{aligned}
Q(x_1,x_2,x_3,x_4) &= \sum_{x_5}\sum_{x_6} Q(x_1,\ldots,x_6) \\
&= \sum_{x_5}\sum_{x_6} P(x_1,x_2,x_3,x_4)P(x_5 \mid x_2,x_3,x_4)P(x_6 \mid x_5) \\
&= P(x_1,x_2,x_3,x_4)\sum_{x_5} P(x_5 \mid x_2,x_3,x_4)\sum_{x_6} P(x_6 \mid x_5) \\
&= P(x_1,x_2,x_3,x_4)\sum_{x_5} P(x_5 \mid x_2,x_3,x_4) \\
&= P(x_1,x_2,x_3,x_4)
\end{aligned}$$

Product Form

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

$$\begin{aligned} Q(x_2, x_3, x_4, x_5) &= \sum_{x_1} \sum_{x_6} P(x_1, x_2, x_3, x_4) P(x_5 \mid x_2, x_3, x_4) P(x_6 \mid x_5) \\ &= P(x_5 \mid x_2, x_3, x_4) \sum_{x_1} P(x_1, x_2, x_3, x_4) \sum_{x_6} P(x_6 \mid x_5) \\ &= P(x_5 \mid x_2, x_3, x_4) P(x_2, x_3, x_4) = P(x_2, x_3, x_4, x_5) \end{aligned}$$

Product Form

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

$$\begin{aligned}
Q(x_2, x_3, x_4, x_5, x_6) &= \sum_{x_1} Q(x_1, \ldots, x_6) \\
&= \sum_{x_1} P(x_1, x_2, x_3, x_4) P(x_5 \mid x_2, x_3, x_4) P(x_6 \mid x_5) \\
&= P(x_2, x_3, x_4) P(x_5 \mid x_2, x_3, x_4) P(x_6 \mid x_5) \\
&= P(x_2, x_3, x_4, x_5) P(x_6 \mid x_5) \\
Q(x_5, x_6) &= \sum_{x_2} \sum_{x_3} \sum_{x_4} P(x_2, x_3, x_4, x_5) P(x_6 \mid x_5) \\
&= P(x_5) P(x_6 \mid x_5) = P(x_5, x_6)
\end{aligned}$$

Decomposable Graphs

$$
\begin{aligned}
S_k &= C_k \bigcap (\bigcup_{i=1}^{k-1} C_i), k = 2, \ldots, K \\
P(x_1, \ldots, x_N) &= P(x_i : i \in C_1) \prod_{k=2}^{K} P(x_i : i \in C_k - S_k \mid S_k)
\end{aligned}
$$

Proposition

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

Proof.

$$
\begin{aligned}
(C_k - S_k) \cap (\cup_{i=1}^{k-1} C_i) &= (C_k - (C_k \cap (\cup_{i=1}^{k-1} C_i)) \cap (\cup_{i=1}^{k-1} C_i) \\
&= (C_k - (\cup_{i=1}^{k-1} C_i)) \cap (\cup_{i=1}^{k-1} C_i) \\
&= \emptyset
\end{aligned}
$$

Decomposable Graphs: Summability

$$S_k = C_k \cap (\cup_{i=1}^{k-1} C_i), k = 2, \ldots, K$$

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

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

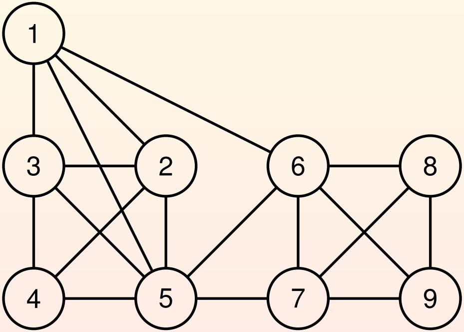
Proposition

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

Proof.

$$\begin{aligned} S &= \sum_{x_1} \sum_{x_2} \cdots \sum_{x_N} P(x_i : i \in C_1) \prod_{k=2}^{K} P(x_i : i \in C_k - S_k \mid S_k) \\ &= \sum_{C_1} \sum_{C_2 - S_2} \cdots \sum_{C_K - S_K} P(x_i : i \in C_1) \prod_{k=2}^{K} P(x_i : i \in C_k - S_k \mid S_k) \\ &= \sum_{C_1} P(x_i : i \in C_1) \sum_{C_2 - S_2} P(x_i : i \in C_2 - S_2 \mid S_2) \cdots \sum_{C_K - S_K} P(x_i : i \in C_K - S_K \mid S_K) \\ &= 1 \end{aligned}$$

Summability Example



$$
\begin{array}{llll}
C_1 = \{1,2,3,5\} & \\
C_2 = \{2,3,4,5\} & S_2 = \{2,3,5\} \\
C_3 = \{1,5,6\} & S_3 = \{1,5\} \\
C_4 = \{5,6,7\} & S_4 = \{5,6\} \\
C_5 = \{6,7,8,9\} & S_5 = \{6,7\}
\end{array}
$$

$$
\begin{aligned}
S &= \textstyle\sum_{x_1} \cdots \sum_{x_9} P(x_1 x_2 x_3 x_5) P(x_4|x_2 x_3 x_5) P(x_6|x_1 x_5) P(x_7|x_5 x_6) P(x_9|x_6 x_7) \\
&= \textstyle\sum_{x_1 x_2 x_3 x_5} P(x_1 x_2 x_3 x_5) \sum_{x_4} P(x_4|x_2 x_3 x_5) \sum_{x_6} P(x_6|x_1 x_5) \sum_{x_7} P(x_7|x_5 x_6) \sum_{x_8 x_9} P(x_8 x_9|x_6 x_7) \\
&= 1
\end{aligned}
$$

Separators

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.

Triangulated Graphs

Theorem

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.

Triangulated 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 $G_{A \cup S}$ and $G_{B \cup S}$ if and only if

- S separates A from B in G
- G_S is a complete graph
- $G_{A \cup S}$ and $G_{B \cup S}$ are each triangulated

Decomposable Graphs

Theorem

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.

Triangulated Graphs and Clique Finding

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.

Triangulated Graphs

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 $S_2, \ldots, S_K$,

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

then

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

Conditional Independence Graphs

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 $A \perp\!\!\!\perp B \mid S$

$$P(x_i : i \in A \cup B | x_j : j \in S) = P(x_i : i \in A | x_j : j \in S) P(x_i : i \in B | x_j : j \in S)$$

Generalized Products

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

$$\begin{aligned} f(x,y,z) &= f(x|z)f(y|z)f(z) \\ &= [f(x|z)][f(y|z)f(z)] \\ &= [f(x|z)][f(y,z)] \end{aligned}$$

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,...,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 $P(x_1, \ldots, x_N)$

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

Generalized Products

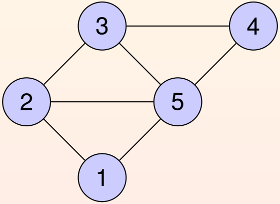
Let $B_1, \ldots, B_K$ be index subsets of {1,...,N}. Given marginal probability functions $P(x_i : i \in B_k), k = 1, \ldots, K$ find functions $a_k(x_i : i \in B_k)$ such that

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

Decomposable Graph

$$I = \{1, 2, 3, 4, 5\}$$

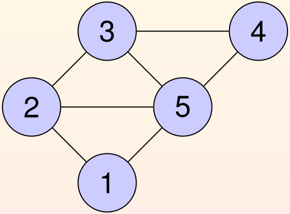
C_1	$=$	$\{1, 2, 5\}$	$1 \perp\!\!\!\perp 4$	$\mid$	$2, 5$
C_2	$=$	$\{2, 3, 5\}$	$1 \perp\!\!\!\perp 3$	$\mid$	$2, 5$
C_3	$=$	$\{3, 4, 5\}$	$2 \perp\!\!\!\perp 4$	$\mid$	$3, 5$
S_2	$=$	$\{2, 5\}$	$1 \perp\!\!\!\perp 4$	$\mid$	$3, 5$
S_3	$=$	$\{3, 5\}$	$1 \perp\!\!\!\perp 4$	$\mid$	$2, 3, 5$



$$\begin{aligned} P(x_i : i \in I) &= \frac{P(x_i : i \in C_1) P(x_i : i \in C_2) P(x_i : i \in C_3)}{P(x_i : i \in S_2) P(x_i : i \in S_3)} \\ &= P(x_i : i \in C_1) P(x_i : i \in C_2 - S_2 \mid S_2) P(x_i : i \in C_3 - S_3 \mid S_3) \end{aligned}$$

Decomposable Graph

In the conditional independence graph, there is no edge between node i and j if and only if X_i and X_j are conditionally independent given the rest of the variables.



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

$$= P_{15}(x_1, x_5) P_{2|15}(x_2 \mid x_1, x_5) P_{3|25}(x_3 \mid x_2, x_5) P_{4|35}(x_4 \mid x_3, x_5)$$

System Diagram 1

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

$$\begin{aligned} P_{12345}(x_1, x_2, x_3, x_4, x_5) &= \frac{P_{125}(x_1, x_2, x_5) P_{235}(x_2, x_3, x_5) P_{345}(x_3, x_4, x_5)}{P_{25}(x_2, x_5) P_{35}(x_3, x_5)} \\ &= P_{15}(x_1, x_5) P_{2|15}(x_2 \mid x_1, x_5) P_{3|25}(x_3 \mid x_2, x_5) P_{4|35}(x_4 \mid x_3, x_5) \end{aligned}$$

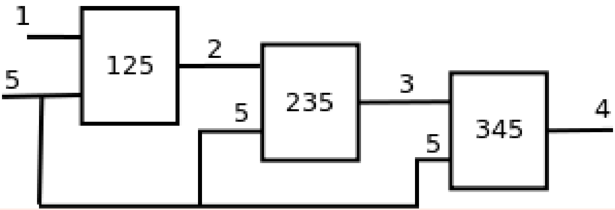


Figure: 1:System H

System Diagram 2

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

$$\begin{aligned} P_{12345}(x_1, x_2, x_3, x_4, x_5) &= \frac{P_{125}(x_1, x_2, x_5) P_{235}(x_2, x_3, x_5) P_{345}(x_3, x_4, x_5)}{P_{25}(x_2, x_5) P_{35}(x_3, x_5)} \\ &= P_{25}(x_2, x_5) P_{1|25}(x_1 \mid x_2, x_5) P_{3|25}(x_3 \mid x_2, x_5) P_{4|35}(x_4 \mid x_3, x_5) \end{aligned}$$

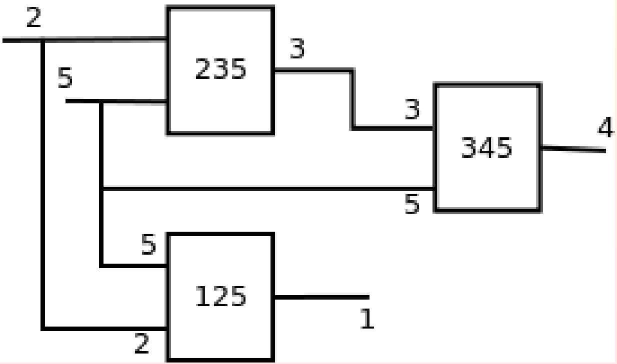


Figure: 1:System G

System Diagram 3

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

$$\begin{aligned} P_{12345}(x_1, x_2, x_3, x_4, x_5) &= \frac{P_{125}(x_1, x_2, x_5) P_{235}(x_2, x_3, x_5) P_{345}(x_3, x_4, x_5)}{P_{25}(x_2, x_5) P_{35}(x_3, x_5)} \\ &= P_{12}(x_1, x_2) P_{5|12}(x_5 \mid x_1, x_2) P_{3|25}(x_3 \mid x_2, x_5) P_{4|35}(x_4 \mid x_3, x_5) \end{aligned}$$

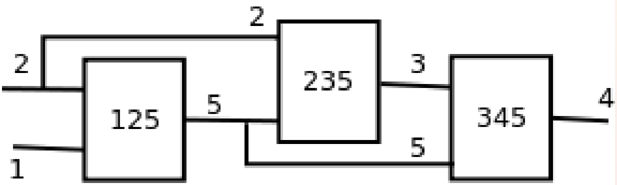


Figure: 1:System I

System Diagram 4

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

$$\begin{aligned} P_{12345}(x_1, x_2, x_3, x_4, x_5) &= \frac{P_{125}(x_1, x_2, x_5) P_{235}(x_2, x_3, x_5) P_{345}(x_3, x_4, x_5)}{P_{25}(x_2, x_5) P_{35}(x_3, x_5)} \\ &= P_{1|25}(x_1 \mid x_2, x_5) P_{2|35}(x_2 \mid x_3, x_5) P_{4|35}(x_4 \mid x_3, x_5) P_{35}(x_3, x_5) \end{aligned}$$

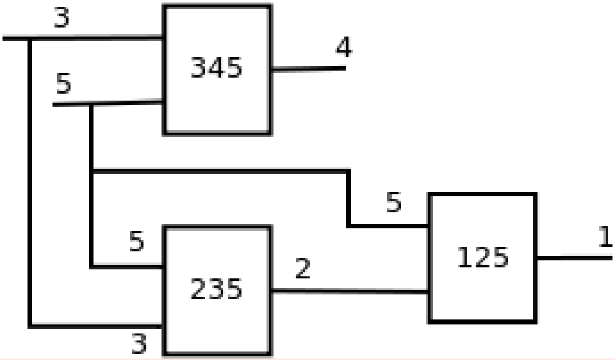


Figure: 2: System E

System Diagram 5

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

$$\begin{aligned} P_{12345}(x_1, x_2, x_3, x_4, x_5) &= \frac{P_{125}(x_1, x_2, x_5) P_{235}(x_2, x_3, x_5) P_{345}(x_3, x_4, x_5)}{P_{25}(x_2, x_5) P_{35}(x_3, x_5)} \\ &= P_{1|25}(x_1 \mid x_2, x_5) P_{2|35}(x_2 \mid x_3, x_5) P_{3|45}(x_3 \mid x_4, x_5) P_{45}(x_4, x_5) \end{aligned}$$

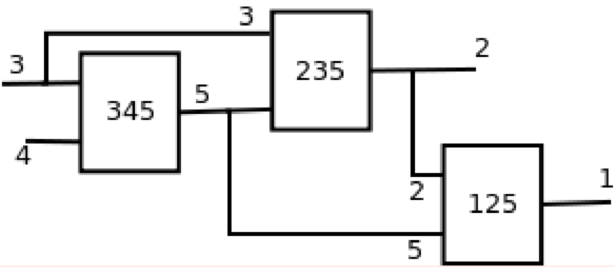
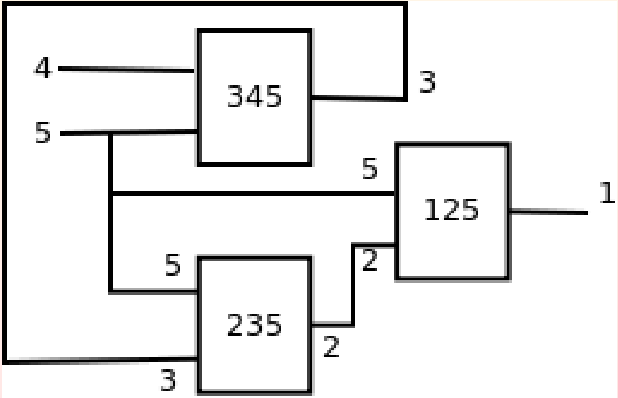


Figure: 2:System L

System Diagram 6

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

$$\begin{aligned} P_{12345}(x_1,x_2,x_3,x_4,x_5) &= \frac{P_{125}(x_1,x_2,x_5)P_{235}(x_2,x_3,x_5)P_{345}(x_3,x_4,x_5)}{P_{25}(x_2,x_5)P_{35}(x_3,x_5)} \\ &= P_{1|25}(x_1 \mid x_2,x_5)P_{2|35}(x_2 \mid x_3,x_5)P_{5|34}(x_5 \mid x_3,x_4)P_{34}(x_3,x_4) \end{aligned}$$



System Diagram 7

$$\{125 : 25\}, \{345 : 35\}$$

$$\begin{aligned} P_{12345}(x_1, x_2, x_3, x_4, x_5) &= \frac{P_{125}(x_1, x_2, x_5) P_{235}(x_2, x_3, x_5) P_{345}(x_3, x_4, x_5)}{P_{25}(x_2, x_5) P_{35}(x_3, x_5)} \\ &= P_{1|25}(x_1 \mid x_2, x_5) P_{4|35}(x_4 \mid x_3, x_5) P_{2|35}(x_2 \mid x_3, x_5) P_{35}(x_3, x_5) \end{aligned}$$

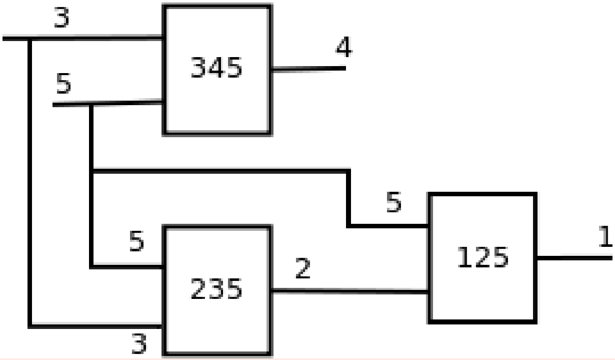


Figure: 3:System E

System Diagram 8

$$\{125 : 25\}, \{345 : 35\}$$

$$\begin{aligned} P_{12345}(x_1, x_2, x_3, x_4, x_5) &= \frac{P_{125}(x_1, x_2, x_5) P_{235}(x_2, x_3, x_5) P_{345}(x_3, x_4, x_5)}{P_{25}(x_2, x_5) P_{35}(x_3, x_5)} \\ &= P_{1|25}(x_1 \mid x_2, x_5) P_{4|35}(x_4 \mid x_3, x_5) P_{3|25}(x_3 \mid x_2, x_5) P_{25}(x_2, x_5) \end{aligned}$$

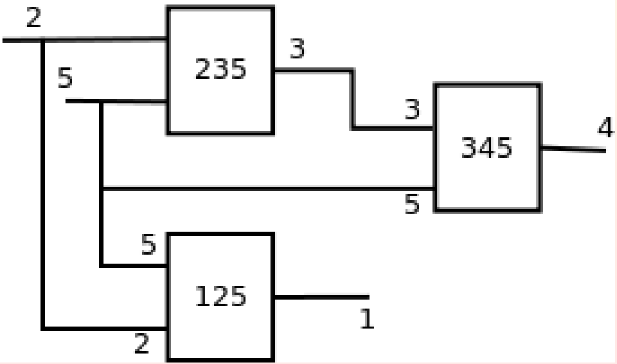


Figure: 3:System G

System Diagram 9

$$\{125:25\},\{345:35\}$$

$$\begin{aligned} P_{12345}(x_1,x_2,x_3,x_4,x_5) &= \frac{P_{125}(x_1,x_2,x_5)P_{235}(x_2,x_3,x_5)P_{345}(x_3,x_4,x_5)}{P_{25}(x_2,x_5)P_{35}(x_3,x_5)} \\ &= P_{1|25}(x_1\mid x_2,x_5)P_{4|35}(x_4\mid x_3,x_5)P_{5|23}(x_5\mid x_2,x_3)P_{23}(x_2,x_3) \end{aligned}$$

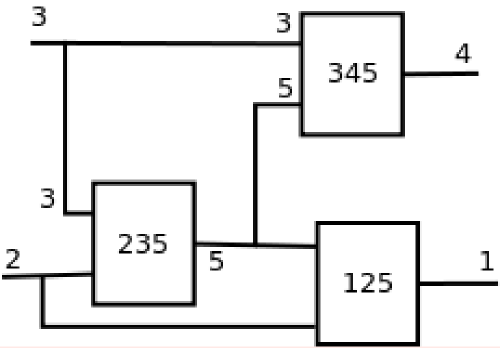


Figure: 3:System J

Feed Forward System Conditional Independences

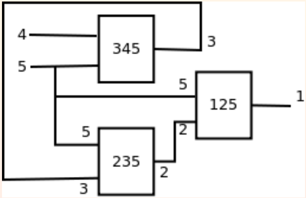
$$
\begin{aligned}
P^A_{12345}(x_1,x_2,x_3,x_4,x_5) &= P_{45}(x_4,x_5)P_{3|45}(x_3|x_4,x_5)P_{1|25}(x_1|x_2,x_5)P_{2|35}(x_2|x_3,x_5)\\
P^E_{12345}(x_1,x_2,x_3,x_4,x_5) &= P_{35}(x_3,x_5)P_{4|35}(x_4|x_3,x_5)P_{1|25}(x_1|x_2,x_5)P_{2|35}(x_2|x_3,x_5)\\
P^G_{12345}(x_1,x_2,x_3,x_4,x_5) &= P_{25}(x_2,x_5)P_{3|25}(x_3|x_2,x_5)P_{1|25}(x_1|x_2,x_5)P_{4|35}(x_4|x_3,x_5)\\
P^H_{12345}(x_1,x_2,x_3,x_4,x_5) &= P_{15}(x_1,x_5)P_{2|15}(x_2|x_1,x_5)P_{3|25}(x_3|x_2,x_5)P_{4|35}(x_4|x_3,x_5)\\
P^I_{12345}(x_1,x_2,x_3,x_4,x_5) &= P_{12}(x_1,x_2)P_{5|12}(x_5|x_1,x_2)P_{3|25}(x_3|x_2,x_5)P_{4|35}(x_4|x_3,x_5)\\
P^J_{12345}(x_1,x_2,x_3,x_4,x_5) &= P_{23}(x_2,x_3)P_{1|25}(x_1|x_2,x_5)P_{5|23}(x_5|x_2,x_3)P_{4|35}(x_4|x_3,x_5)\\
P^L_{12345}(x_1,x_2,x_3,x_4,x_5) &= P_{34}(x_3,x_4)P_{1|25}(x_1|x_2,x_5)P_{2|35}(x_2|x_3,x_5)P_{5|34}(x_5|x_3,x_4)
\end{aligned}
$$

These decompositions correspond to the same Decomposable Graphical Model

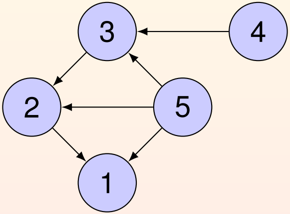
$$
P_{12345}(x_1,x_2,x_4,x_4,x_5) = \frac{P_{345}(x_3,x_4,x_5)P_{125}(x_1,x_2,x_5)P_{235}(x_2,x_3,x_5)}{P_{25}(x_2,x_5)P_{35}(x_3,x_5)}
$$

Feedforward Systems: Bayesian Networks

System A

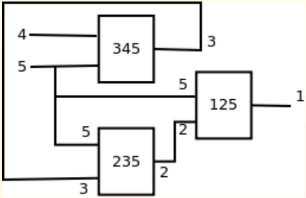


Associated Bayesian Network



System A	$P(x_1, x_2, x_3, x_4, x_5)$	$=$	$P_{45}(x_4, x_5) P_{3\mid 45}(x_3 \mid x_4, x_5) P_{2\mid 35}(x_2 \mid x_3, x_5) P_{1\mid 25}(x_1 \mid x_2, x_5)$
Bayesian Network	$P(x_1, x_2, x_3, x_4, x_5)$	$=$	$P_4(x_4) P_5(x_5) P_{3\mid 45}(x_3 \mid x_4, x_5) P_{2\mid 35}(x_2 \mid x_3, x_5) P_{1\mid 25}(x_1 \mid x_2, x_5)$

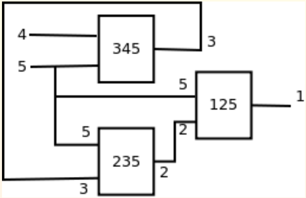
Causal Structure



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

$$
\begin{aligned}
J_1 &= \{3,4,5\} \\
I_1 &= \{4,5\} \\
O_1 &= \{3\} \\
J_2 &= \{1,2,5\} \\
I_2 &= \{2,5\} \\
O_2 &= \{1\} \\
J_3 &= \{2,3,5\} \\
I_3 &= \{3,5\} \\
O_3 &= \{2\}
\end{aligned}
$$

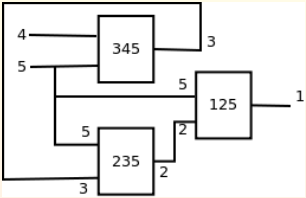
Causal Structure



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: $X_1 \nrightarrow X_3$
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: $X_3 \nrightarrow X_1 \mid X_2, X_5$
Given X_2, X_5, X_3 is conditionally independent of X_1: $X_3 \perp\!\!\!\perp X_1 \mid X_2, X_5$

Conditional Independence 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
Given its parents, each variable is conditionally independent
of its non-descendants
Given X_3 and X_5, X_2 is conditionally independent X_4: $X_2, \perp\!\!\!\perp X_4 \mid X_3, X_5$

Conditional Independence Structure

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

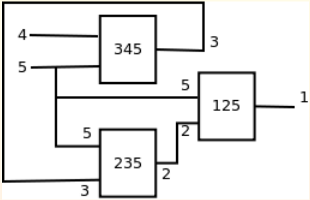
$$\begin{aligned}
P_{24|35}(x_2, x_4 \mid x_3, x_5) &= \sum_{x_1} \frac{P_{125}(x_1, x_2, x_5)P_{235}(x_2, x_3, x_5)P_{345}(x_3, x_4, x_5)}{P_{25}(x_2, x_5)P_{35}(x_3, x_5)P_{35}(x_3, x_5)} \\
&= \frac{P_{235}(x_2, x_3, x_5)P_{345}(x_3, x_4, x_5)}{P_{25}(x_2, x_5)P_{35}(x_3, x_5)P_{35}(x_3, x_5)} P_{25}(x_2, x_5) \\
&= \frac{P_{235}(x_2, x_3, x_5)P_{345}(x_3, x_4, x_5)}{P_{35}(x_3, x_5)P_{35}(x_3, x_5)} \\
&= P_{2|35}(x_2 \mid x_3, x_5)P_{4|35}(x_4 \mid x_3, x_5)
\end{aligned}$$

Possible Causal System Structure

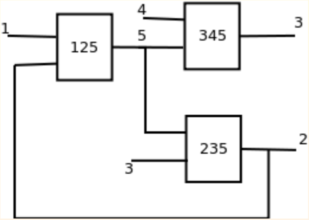
Let us consider all the possibilities where each subsystem has exactly one output variable and no two different subsystems produce the same output variables.

System	subsystem	output	subsystem	output	subsystem	output
A	345	3	235	2	125	1
B	345	3	235	2	125	5
C	345	3	235	5	125	1
D	345	3	235	5	125	2
E	345	4	235	2	125	1
F	345	4	235	2	125	5
G	345	4	235	3	125	1
H	345	4	235	3	125	2
I	345	4	235	3	125	5
J	345	4	235	5	125	1
K	345	4	235	5	125	2
L	345	5	235	2	125	1
M	345	5	235	3	125	1
N	345	5	235	3	125	2

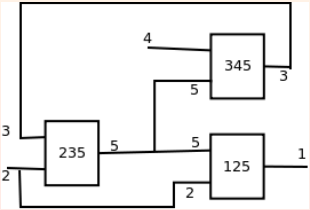
System Diagrams



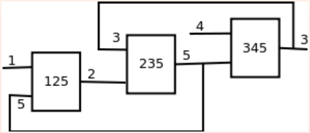
(a) System A: Feedfoward



(b) System B: Feedback

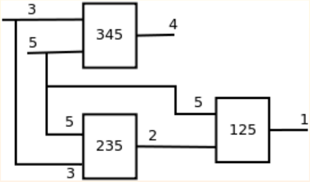


(c) System C: Feedback

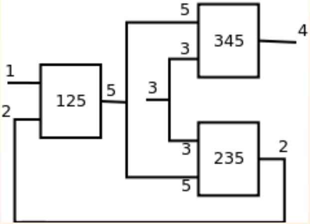


(d) System D: Feedback

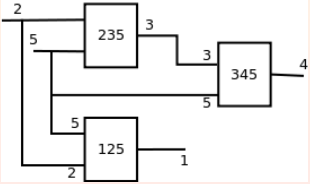
System Diagrams



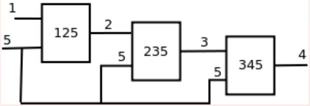
(e) System E: Feedfoward



(f) System F: Feedback

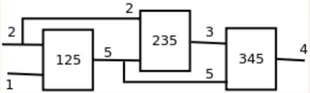


(g) System G: Feedfoward

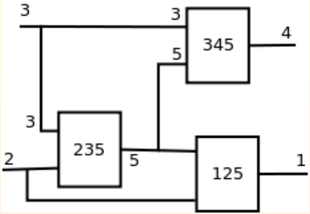


(h) System H: Feedfoward

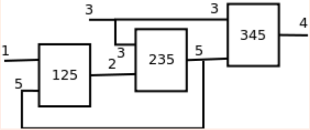
System Diagrams



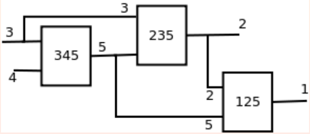
(i) System I: Feedfoward



(j) System J: Feedfoward



(k) System K: Feedback



(l) System L: Feedfoward