Analysis of Categorical Data
Chapter 10: Build Log linear model

Shaobo Jin

Department of Mathematics

Intended Learning Outcome

Through this chapter, you should be able to

1. Interpret the conditional independence graph,
2. Interpret the generator multigraph,
3. determine conditional independence and collapsibility from graphs,
4. determine decomposability,
5. obtain closed form expression of MLE.

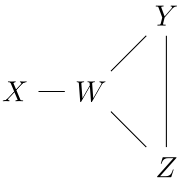
Hierarchical Loglinear Model

A loglinear model is a hierarchical loglinear model, if a higher-order term is in the model then all lower-order terms are also included in the model.

$$
\begin{aligned}
\log \mu_{ij} &= \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}, \\
\log \mu_{ijk} &= \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}, \\
\log \mu_{ijk} &= \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}, \\
\log \mu_{ijk} &= \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}, \\
\log \mu_{ijk} &= \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}.
\end{aligned}
$$

Graph

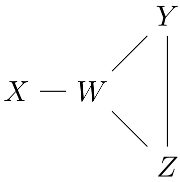
A graph consists of a set of vertices, and a set of edges connecting some vertices.



In a conditional independence graph for a hierarchical loglinear model, each vertex represents a variable.

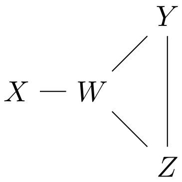
- The edges represent pairwise association, i.e., first-order interaction in the loglinear model.
- The absence of an edge connecting two variables represents conditional independence between them.

Reading Conditional Independence Graph



- The graph lacks XY and XZ.
- This can be a graph for (WX, WY, WZ, YZ).
- However, the loglinear model described by the graph is not unique. Two loglinear models with the same pairwise associations can have the same association graph. This can also be a graph for (WX, WYZ), since WYZ also includes WY, WZ, YZ.

Path



- Two vertices in a graph are adjacent if an edge joint hem.
- A path from vertex X to vertex Y is a sequence of edges leading from X to Y.
- A set C of vertices separates the two sets A and B of vertices, if all paths from any vertex in A to any vertex in B pass through at least one vertex in C.
  - Consider $A = \{X\}$ and $B = \{Y\}$. Then $C = \{W\}$ separates A and B. $C = \{W, Z\}$ also separates A and B.

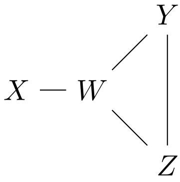
Independence

Unconditionally independent variables are said to lie in different connected components of the graph.

$$\begin{tikzcd} & & Y \\ X \arrow[r, dash] & W \arrow[dr, dash] & \\ & & Z \end{tikzcd}$$

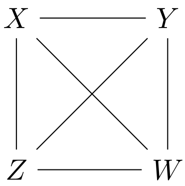
- (X, Z, W) is independent of Y.

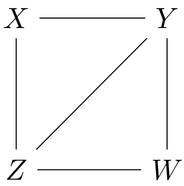
Global and Local Markov Properties



- Global Markov property: If a set C of variables separates two sets A and B of variables, then the variables in A are conditionally independent of the variables in B, given the variables in C.
  - $X \perp Y \mid W$, $X \perp Y \mid (W, Z)$.
- Local Markov property: a variable is conditionally independent of all other variables, given its adjacent neighbors to which it is connected with an edge.
  - $X \perp Y \mid W$, $X \perp (Y, Z) \mid W$.

Complete Graph

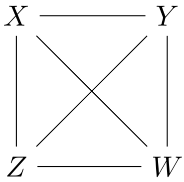


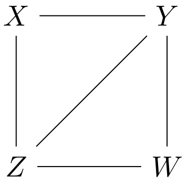


- A complete graph has an edge joining every pair of vertices.
  - LHS is a complete graph but RHS is not a complete graph.
- A maxclique in a graph is a set of vertices that form a complete graph that is not contained in a larger complete graph.
  - LHS has a maxclique: $(XYZW)$.
  - RHS has two maxcliques: (XYZ) and (YZW).

Graphical Models

A hierarchical loglinear model is called a graphical model when there is a one-to-one correspondence between the maxcliques in the conditional independence graph and the generating class of the loglinear model.

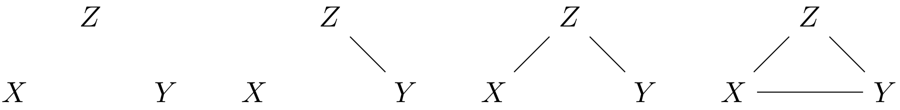




- Both $(XYZW)$ and (XYZ, WYZ, WX) have the same conditional independence graph on the left.
    - $(XYZW)$ is a graphical model. (XYZ, WYZ, WX) is not a graphical model since the generating class is not the same as the maxclique.
- The generating class (XYZ, YZW) yields the conditional independence graph on the right, and is graphical.

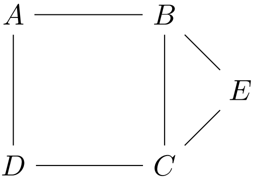
Maxclique and Minimal Sufficiency

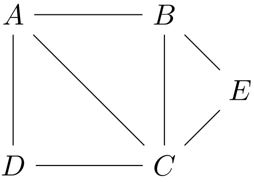
In a graphical model, the maxcliques are always related to the minimal sufficient statistics.



1. The maxclique is (X, Y, Z), same as generating class. The minimal sufficient statistic is $\{n_{i++}\}$, $\{n_{+j+}\}$, $\{n_{++k}\}$.
2. The maxclique is (X, YZ), same as generating class. The minimal sufficient statistic is $\{n_{i++}\}$ and $\{n_{+jk}\}$.
3. The maxclique is (XZ, YZ), same as generating class. The minimal sufficient statistic is $\{n_{i+k}\}$and $\{n_{+jk}\}$
4. The maxclique is (XYZ). If the generating class is (XY, XZ, YZ), then the model is not graphical.
   - The graphical model has minimal sufficient statistic $\{n_{ijk}\}$.

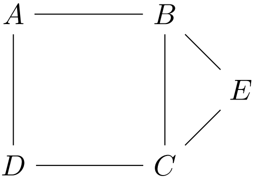
Chordal Graph

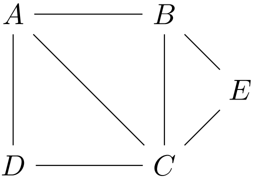




- A cycle is a sequence of edges that begins and ends at a given vertex, e.g., $B - C - E - B$ is a cycle.
- A chord is an edge between nonconsecutive vertices along a cycle.
  - Graph on the right: $A - B - C - D - A$ is a cycle, and the edge $A - C$ is a chord.
- A chordal graph is a graph where there exists a chord in every cycle of length four or more. If all cycles have length less than four, then the graph is also chordal.
  - Graph on the left is not chordal.
  - Graph on the right is chordal.

Decomposable Model





A decomposable model is defined to be a graphical model whose conditional independence graph is a chordal graph.

- The generating class (AB, BC, CD, AD, BCE) yields the graph on the left, and is a graphical model. The model is not decomposable because it is not chordal.
- The generating class (ABC, ACD, BCE) yields the graph on the right, and is a graphical model. The model is decomposable because it is also chordal.
- The generating class $(AB, BC, AC, AD, CD, BCE)$ yields the graph on the right, and is a chordal model, but not graphical.

Benefit of Decomposable Model

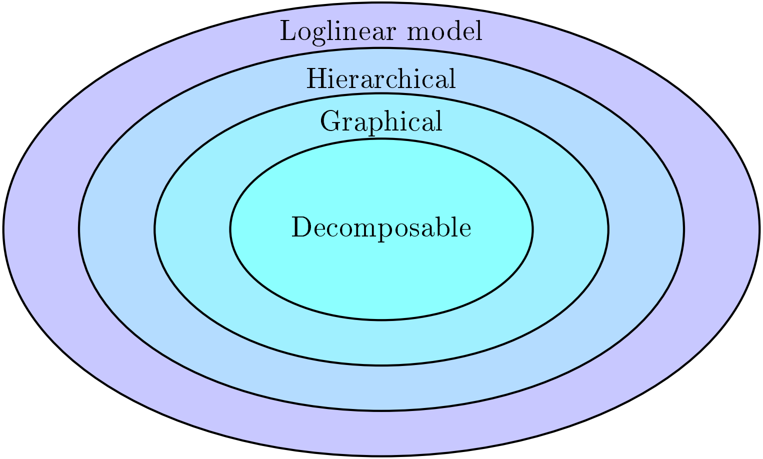
We have seen in Chapter 9 that a benefit of a decomposable model is that the joint probability of the contingency table can be factored in closed form with respect to the indices in the generating class, and the maximum likelihood estimators of μ are available in closed form only for decomposable models.

- The model (XZ, YZ) is decomposable. The MLEs are $\hat{\mu}_{i+k} = n_{i+k}$, $\hat{\mu}_{+jk} = n_{+jk}$. We also have

$$\pi_{ijk} = \frac{\pi_{i+k}\pi_{+jk}}{\pi_{++k}}.$$

  Hence, $\hat{\mu}_{ijk} = n_{i+k}n_{+jk}/n_{++k}$.
- The model (XY, XZ, YZ) is not decomposable because it is not graphical. Hence, it has no closed form probabilistic expression in terms of π_{ijk} and its marginals.

Summary of Loglinear Models



Collapsibility in Three-way Contingency Table

Collapsibility condition for three-way table: In a three-way table, a variable is collapsible with respect to the interaction between the other two variables if it is at least conditionally independent of one of the other two variables given the third variable.

The XY marginal odds ratio and the XY conditional odds ratio are identical if

1. either $X \perp Z \mid Y$,
2. or $Y \perp Z \mid X$,
3. or both.

For example, these conditions occur for loglinear models (XY, YZ), (XY, XZ), and (XY, Z). We say that Z is collapsible with respect to the XY association.

Collapsibility in the (XY, XZ) Model

Consider the (XY, XZ) model

$$\log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ}.$$

The XY marginal table satisfies

$$\begin{aligned}\log \mu_{ij+} &= \log\left[\sum_k \exp\left(\lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ}\right)\right] \\ &= \lambda + \underbrace{\lambda_i^X + \log\left[\sum_k \exp\left(\lambda_k^Z + \lambda_{ik}^{XZ}\right)\right]}_{\text{new main effect } \lambda_i^X} + \lambda_j^Y + \lambda_{ij}^{XY},\end{aligned}$$

where the interaction λ_{ij}^{XY} remains the same as in the partial table. Since the XY odds ratios are functions of λ_{ij}^{XY}, they are the same in both marginal and partial tables.

An Example

From the conditional independence graph, we can tell whether we can collapse a variable. Consider the conditional independence graph

$$A - M - C$$

The model corresponding model is (AM, CM), i.e., $A \perp C \mid M$. Hence, by the collapsibility condition,

- the AM conditional odds ratio is identical to the AM marginal odds ratio collapsed over C,
- the CM conditional odds ratio is identical to the CM marginal odds ratio collapsed over A,
- the AC conditional odds ratio is not necessarily identical to the AC marginal odds ratio collapsed over M.

Collapsibility in Multiway Contingency Tables

Collapsibility condition for multiway table: Suppose that a model for a multiway table partitions variables into three mutually exclusive subsets, A, B, C, such that B separates A and C. After collapsing the table over the variables in C,

1. all λ terms relating variables in A are unchanged,
2. all λ terms relating variables in A to variables in B are unchanged,
3. the λ terms relating variables in B are not invariant to collapsing.

A motivation is that, by the Markov property, $A \perp C \mid B$. In other words, conditional independence yields collapsibility.

- But such collapsibility condition is only a sufficient condition, not a necessary condition.
- Their numerical estimates may differ slightly.

Consequence of Collapsibility Theorem

Recall that the conditional odds ratio depend on λ. If B separates A and C, then one can collapse over the factors in C

- without distorting the the association between a factor in A and a factor in B,
- without distorting the the association between factors in A.

Consider the (XY, XZ) model, where X separates Y and Z. Hence,

$$\begin{aligned} \log \mu_{ijk} &= \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ}, \\ \log \mu_{ij+} &= \lambda + \beta_i^X + \lambda_j^Y + \lambda_{ij}^{XY}. \end{aligned}$$

That is, λ_j^Y and λ_{ij}^{XY} remain the same, but λ_i^X may not. At a fixed level $Z = k$, the local odds ratios satisfies

$$\log \theta_{ij(k)} = \log \left(\frac{\mu_{ijk}\mu_{i+1,j+1,k}}{\mu_{i+1,j,k}\mu_{i,j+1,k}} \right) = \lambda_{ij}^{XY} + \lambda_{i+1,j+1}^{XY} - \lambda_{i+1,j}^{XY} - \lambda_{i,j+1}^{XY}.$$

which remains unchanged.

Example: Collapsibility in Multiway Tables

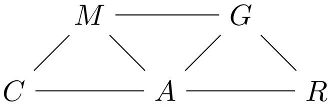
(WX, WY, WZ, YZ) model

1. Let $A = X$, $B = W$, and $C = \{Y, Z\}$. If we collapse over Y and Z, the WX association is unchanged.
2. Let $A = \{Y, Z\}$, $B = \{W\}$, and $C = \{X\}$. Associations among W, Y, Z remain the same after collapsing over X.
3. Let $A = \{X\}$, $B = \{W, Z\}$, and $C = \{Y\}$. Associations among W, Y, Z remain the same after collapsing over X. The association between W and Z is also unchanged. But if we collapse Y, the WZ association is changed.

(WX, WY, XY, Z) model

Let $A = \{X, Y\}$, $B = W$, and $C = \{Z\}$. If we collapse over Z, the associations among W, X, Y are unchanged.

Another Example



- The model is $(AC, AM, CM, AG, AR, GM, GR)$.
- $\{A, M\}$ separates C and $\{G, R\}$.
- $C \perp (G, R) \mid (A, M)$.
- The conditional associations between C and A, and between C and M are the same with the model (AC, AM, CM).

Generator Multigraph

The generator multigraph, or simply multigraph, is an alternative to the conditional independence graph for hierarchical loglinear models.

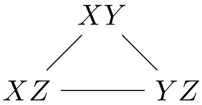
- Different from a conditional independence graph, a multigraph allows a vertex with more than one variable.

A multigraph consists of vertices and multiedges.

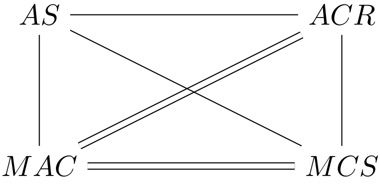
$$
\begin{aligned}
\text{vertex} &= \text{generators in the generating class} \\
\text{multiedge} &= \text{edges equal in number to the number of indices common} \\
&\quad\ \text{to the two generators being joined.}
\end{aligned}
$$

Examples of Multigraph

Consider the model with generating class (XY, XZ, YZ). The vertices are XY, XZ, and YZ. Each pair of vertices share a single index, so there is a single edge joining each pair. The multigraph is

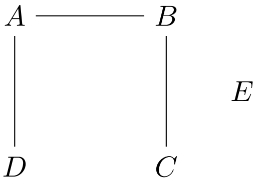


Consider the model with generating class (AS, ACR, MCS, MAC). The vertices are AS, ACR, MCS, and MAC. The pair (ACR, MAC) share two indices, so is the pair (MCS, MAC). The multigraph is

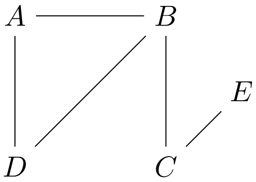


Tree

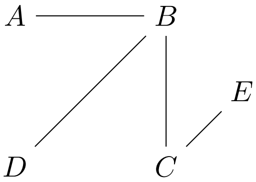
A connected graph is any graph for which there is at least one path from any vertex to any other vertex in the graph. A tree is a connected graph with no circuit (closed loop) that includes each vertex of the graph.



Not connected.



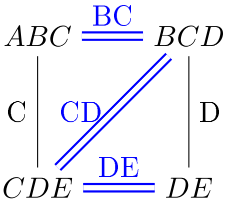
Connected graph, but not a tree because of the circuit including A, B, and D.



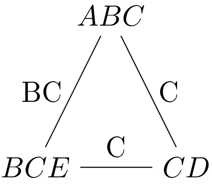
A tree.

Maximum Spanning Tree

For a multigraph, a maximum spanning tree is a tree where the sum of all the edges is maximum. The indices common to two vertices being joined in the maximum spanning tree are called the branches of the tree. The maximum spanning tree is not unique, neither is the branch set.



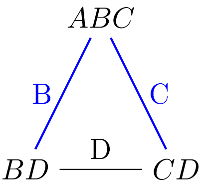
The maximum spanning tree is $[ABC]\,[BCD]\,[CDE]\,[DE]$. The branch set is $\{(BC)\,,(CD)\,,(DE)\}$



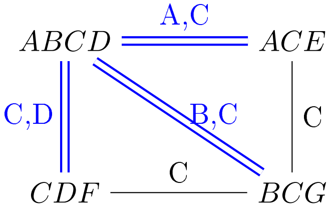
The maximum spanning tree can be $[ABC]\,[BCE]\,[CD]$ or $[CD]\,[ABC]\,[BCE]$. The branch set is $\{BC, C\}$ for both cases.

Sufficient and Necessary Condition

A hierarchical loglinear model is decomposable if and only if, in any maximum spanning tree of the multigraph, the number of factors equals to the number of indices added over the vertices minus the number of indices added over the branches of the maximum spanning tree.



Number of indices added over vertices is 7. Number of indices added over branches is 2. Number of factors is 4. Hence, not decomposable.



Number of indices added over vertices is 13. Number of indices added over branches is 6. Number of factors is 7. Hence, decomposable.

Joint Probability of Decomposable Model

Once we know that a loglinear model is decomposable, we want to factorize the joint probability of the contingency table in closed form with respect to the indices in the generating class.

- Let $P(\ell_1, \ell_2, ..., \ell_d)$ be the probability that a subject is in level ℓ_1 of the first factor, level ℓ_2 of the second factor, ..., and level ℓ_d of the dth factor.
- Let S be a subset of indices from the set $\{1, 2, ..., d\}$. We use p_S to denote the marginal probability having indices contained in S while all other indices are summed over.

Then, for a decomposable model,

$$P(\ell_1, \ell_2, ..., \ell_d) = \frac{\prod_{S \in V} p_S}{\prod_{S \in B} p_S},$$

where V is the set of indices in the vertex set of the maximum spanning tree and B represents the multiset of indices in the branch set.

Find Joint Probability: One Example

$$ABC \overset{\text{B,C}}{=\!=\!=} BCD$$

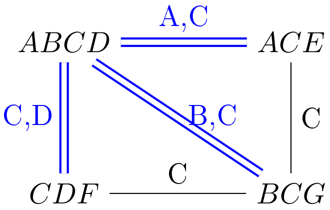
The maximum spanning tree is $[ABC]\,[BCD]$, and the branch set is $\{BC\}$.

- Number of indices added over vertices is 6.
- Number of indices added over branches is 2.
- Number of factors is 4. Hence, the loglinear model is decomposable.

Hence,

$$\pi_{ijkl} = \frac{\pi_{ijk+}\pi_{+jkl}}{\pi_{+jk+}}.$$

Find Joint Probability: Another Example



The maximum spanning tree is $[ABCD]\,[CDF]\,[ACE]\,[BCG]$, and the branch set is $\{(AC)\,,(BC)\,,(CD)\}$.

- Number of indices added over vertices is 13.
- Number of indices added over branches is 6.
- Number of factors is 7. Hence, the loglinear model is decomposable.

Hence,

$$\pi_{abcdefg} = \frac{\pi_{abcd+++}\pi_{++cd+f+}\pi_{a+c+e++}\pi_{+bc+++g}}{\pi_{a+c++++}\pi_{+bc++++}\pi_{++cd+++}}.$$

Fundamental Conditional Independence Set

Given a multigraph, we would like to find a fundamental conditional independence set.

- Suppose that the fundamental conditional independence set is $[C_1 \otimes C_2 \otimes \cdots \otimes C_k \mid S]$, where d factors in a contingency table have been partitioned by the $k+1$ sets of factors C_1,C_2, ..., C_k, and S.
- We can replace S with S' such that $S \subseteq S'$, and replace each C_i with C_i' such that $C_i' \subseteq C_i$, subject to

$$\left(C_1' \cup C_2' \cup \cdots \cup C_k'\right) \cap S' \quad = \quad \emptyset.$$

- Then, $[C_1' \otimes C_2' \otimes \cdots \otimes C_k' \mid S']$.

For example, if the fundamental conditional independence set is $[A, B \otimes D \mid E]$, then $[A \otimes D \mid B, E]$ and $[B \otimes D \mid A, E]$.

Decomposable Model

Denote a multigraph by M. Suppose that the resulting loglinear model is decomposable. We can obtain the fundamental conditional independence set using the following steps.

- Suppose that there are d factors in the contingency table. Let S be a subset of these factors.
- We construct a new multigraph M/S by removing each factor of S from each vertex in the multigraph, and removing each edge corresponding to that factor.
- Then, we have mutual independence of the sets of factors in the disconnected components of M/S, conditional on S.

We can simply let S be the branches of the maximum spanning tree. The resulting conditional independence sets

Conditional Independence: One Example

$$ABC \overset{\text{B,C}}{=\!=\!=} BCD$$

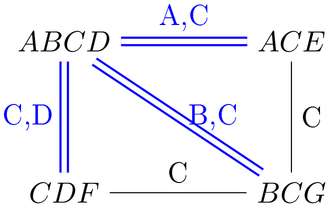
We have shown that the model is decomposable. The branch set is $\{BC\}$

- Let $S = \{B, C\}$, the branch set.
- To construct M/S, we remove all indices in S from each vertex and the edge corresponding to them.

$$A \qquad\qquad D$$

- M/S results in two disconnected components, i.e., $\{A\}$ and $\{D\}$.
- Hence, $A \perp D \mid (B, C)$. We often write $[A \otimes D \mid B, C]$ as well, in case there are more than two disconnected components.

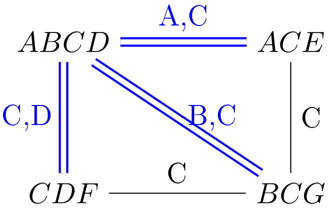
Conditional Independence: Another Example

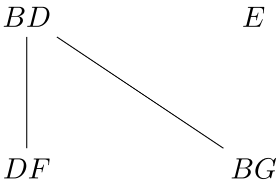


We have shown that the model is decomposable. The branch set is $\{(AC), (BC), (CD)\}$.

- We can choose $S = \{A, C\}$, or $S = \{B, C\}$, or $S = \{C, D\}$.

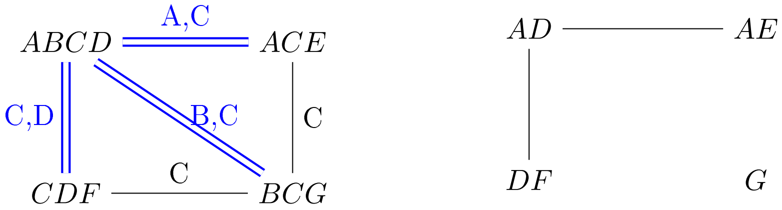
Conditional Independence: Another Example (2)





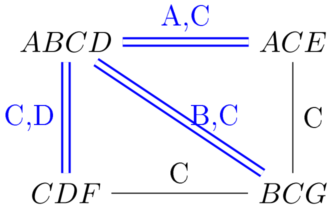
- If we choose $S = \{A, C\}$ and removing the factors in S from each vertex, then M/S is the graph on the right.
- M/S results in two disconnected components, i.e., $\{B, D, F, G\}$ and $\{E\}$.
- Hence, $[B, D, F, G \otimes E \mid A, C]$, or $(B, D, F, G) \perp E \mid (A, C)$.

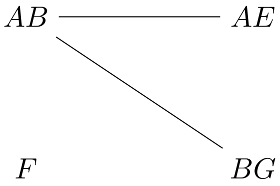
Conditional Independence: Another Example (3)



- If we choose $S = \{B, C\}$ and removing the factors in S from each vertex, then M/S is the graph on the right.
- M/S results in two disconnected components, i.e., $\{A, D, E, F\}$ and $\{G\}$.
- Hence, $[A, D, E, F \otimes G \mid B, C]$, or $(A, D, E, F) \perp G \mid (B, C)$.

Conditional Independence: Another Example (4)





- If we choose $S = \{C, D\}$ and removing the factors in S from each vertex, then M/S is the graph on the right.
- M/S results in two disconnected components, i.e., $\{A, B, E, G\}$ and $\{F\}$.
- Hence, $[A, B, E, G \otimes F \mid C, D]$, or $(A, B, E, G) \perp F \mid (C, D)$.

Nondecomposable Model

To find the fundamental conditional independence set for a nondecomposable model, we need to use a new concept called edge cutsets.

- An edge cutset of a multigraph is an inclusion-minimal set of multiedges whose removal disconnects the multigraph.

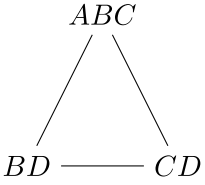
After removing Z from the graph, we obtain disconnected components. Hence, the edge cutset is $\{Z\}$.

$$XZ \text{ ——— } YZ \qquad\qquad\qquad X \qquad\qquad Y$$

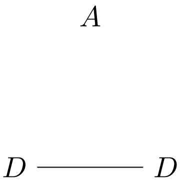
- For decomposable models, the edge cutsets are the branches of the maximum spanning tree.

Edge Cutset: Example

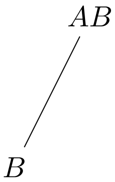
Consider the nondecomposable model on the right.



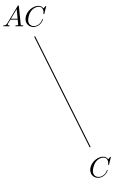
Edge cutset
$S_1 = \{B, C\}$



Edge cutset
$S_2 = \{C, D\}$



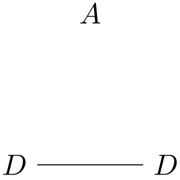
Edge cutset
$S_3 = \{B, D\}$



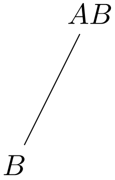
Identify Fundamental Conditional Independence Set

To identify the fundamental conditional independence set for nondecomposable models, we follow the same steps as in decomposable models, except that the indices in S come from the edge cutsets instead of the branches of a maximum spanning tree.

Edge cutset $S_1 = \{B, C\}$. Hence, $[A \otimes D \mid B, C]$.



Edge cutset $S_2 = \{C, D\}$. Only one component left, so no conditional independence set.



Edge cutset $S_3 = \{B, D\}$. Only one component left, so no conditional independence set.