Loglinear models

Independence and Collapsility Graphs

Demetris Athienitis

Independence Graph

A graphical representation for conditional independence. They are non-directed and there are multiple models that correspond to the same independence graph.

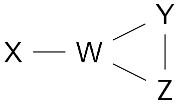
Graphical models are a subclass of loglinear models.

- Within this class there is a unique model for each independence graph
- For any group of variables having no missing edges, graphical model contains the highest order interaction term for those variables

The graphs consist of:

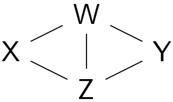
- Vertices (or nodes) represent variables
- Connected by edges: a missing edge between two variables represents a conditional independence between the variables

4-way examples



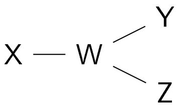
- (WX, WY, WZ, YZ) loglinear model leads to this independence graph
- (WX, WYZ) also leads to this graph and is the graphical model
- X and Y are conditionally independent
  - Conditional on subset $\{W, Z\}$ or simply W
  - Not necessarily marginally independent

4-way examples



- Many possible models for this graph
  - (WX, XZ, WY, WZ, YZ)
  - (WX, XZ, WYZ)
  - (WXZ, WY, YZ)
  - (WXZ, WYZ) is the graphical model
- X and Y are conditionally independent given $\{W, Z\}$

4-way examples



- (WX, WY, WZ) is the only model that fits this graph
- All pairs of X, Y, Z are conditionally independent given W

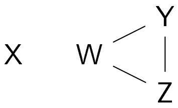
4-way examples

W — X — Y — Z

What is the model that fits this graph?

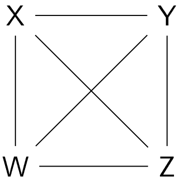
(WX, XY, YZ) is the only model that fits this graph

4-way examples



- (X, WY, WZ, YZ) and (X, WYZ) are both possible
- X is both marginally and conditionally independent of the other variables

4-way examples



There are many models that are possible here

- (WX, WY, WZ, XYZ)
- (WX, WYZ, XYZ)
- $(WXYZ)$
- Many others

Collapsibility

To simplify higher-order contingency tables we can always collapse them into lower-order tables.

For instance, if we have X, Y and Z, we can collapse to just have X and Y by summing over partial $X - Y$ tables for each level of Z

- This can lead to misleading results depending on what you are interested in
- X and Y can be marginally associated but conditionally independent

Collapsibility

For a three-way table, the XY marginal and conditional odds ratios are identical if either Z and X are conditionally independent or if Z and Y are conditionally independent.

- Conditions say control variable Z is either:
  - conditionally independent of X given Y, as in model (XY, YZ)
  - or conditionally independent of Y given X, as in (XY, XZ)
- I.e., XY association is identical in the partial tables and the marginal table for models with independence graphs

X — Y — Z   Y — X — Z

Example (Teen substance usage)

```
> ftable(teens, row.vars=c("alc","cigs"))
         mj yes  no
alc cigs
yes yes     911 538
    no       44 456
no  yes       3  43
    no        2 279
```

- A = alcohol use
- C = cigarette use
- M = marijuana use

Example (continued)

AC conditional independence model, (AM, CM), has graph

$$\text{A} — \text{M} — \text{C}$$

Consider *AM* association, treating *C* as control variable. Since *C* is conditionally independent of *A*, the AM conditional odds ratios are the same as the *AM* marginal odds ratio collapsed over *C*

$$\frac{(909.24)(142.16)}{(438.84)(4.76)} = \frac{(45.76)(179.84)}{(555.16)(0.24)} = \frac{(955)(322)}{(994)(5)} = 61.9$$

with the expected values derived by software in next slide.

Example (continued)

```
> AM.CM.fitted = teens
> AM.CM.fitted[,,] = predict(teens.AM.CM, type="response")
> AM.CM.fitted[,"yes",]
     alc
mj           yes         no
  yes 909.239583   4.760417
  no  438.840426 142.159574
> AM.CM.fitted[,"no",]
     alc
mj            yes          no
  yes  45.7604167   0.2395833
  no  555.1595745 179.8404255
  > AM.CM.fitted[,"yes",] + AM.CM.fitted[,"no",]
     alc
mj    yes  no
  yes 955   5
  no  994 322
```

Example (continued)

Or by using the loglinear model

```
> exp(coef(teens.AM.CM)[5])
alcyes:mjyes
    61.87324
```

Similarly, CM association is collapsible over A

Example (continued)

$$A \text{ --- } M \text{ --- } C$$

The AC association is not collapsible, because M is conditionally dependent with both A and C in model (AM, CM). Thus, A and C may be marginally dependent, even though conditionally independent.

$$\frac{(909.24)(0.24)}{(45.76)(4.76)} = \frac{(438.84)(179.84)}{(555.16)(142.16)} = 1$$

$$\frac{(1348.08)(180.08)}{(600.92)(146.92)} = 2.75 \neq 1$$

Example (continued)

```
> AM.CM.fitted["yes",,]
     alc
cigs          yes         no
  yes 909.2395833  4.7604167
  no   45.7604167  0.2395833
> AM.CM.fitted["no",,]
     alc
cigs       yes       no
  yes 438.8404 142.1596
  no  555.1596 179.8404
> AM.CM.fitted["yes",,] + AM.CM.fitted["no",,]
     alc
cigs      yes      no
  yes 1348.08  146.92
  no   600.92  180.08
```

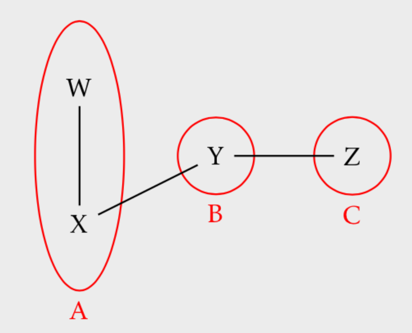
Collapsibility Conditions for Multiway Tables

$$A — B — C$$

If the variables in a model for a multiway table partition into three mutually exclusive subsets, A, B, C, such that B separates A and C (that is, if the model does not contain parameters linking variables from A directly to variables from C), then when the table is collapsed over the variables in C, model parameters relating variables in A and model parameters relating variables in A with variables in B are unchanged.

Example

Consider the (WX, XY, YZ) model (drawn slightly differently)

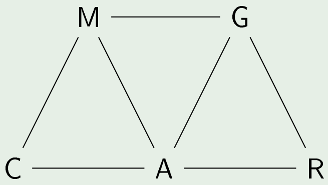


Then collapsing over Z:

- WX and XY associations are unchanged
- W and Y are still conditionally independent given X

Example (Teen substance usage continued)

In addition to the variables seen so far data exists on the race and gender of each teen. Text suggests loglinear model $(AC, AM, CM, AG, AR, GM, GR)$



See class notes

We learned

Independence graphs are a graphical representation of the relationship among variables than help us determine whether collapsing over a certain variable (or set of), will change the relationship among the other variables.