Loglinear Models

Stat 557
Heike Hofmann
Outline

• Relationship of loglinear models and logistic regression
• Association Graphs
• Association Models
Loglinear Model vs Logistic Regression

• Consider model GI, BI, LI, LBG from accident data

• Assume, I is response - i.e. what affects probability of injury:

\[
\log \frac{P(I = \text{yes} \mid G, B, L)}{P(I = \text{no} \mid G, B, L)} = \left( \lambda_I^{\text{yes}} - \lambda_I^{\text{no}} \right) + \left( \lambda_{GI}^{\text{yes}} - \lambda_{GI}^{\text{no}} \right) + \\
\left( \lambda_{LI}^{\text{yes}} - \lambda_{LI}^{\text{no}} \right) + \left( \lambda_{BI}^{\text{yes}} - \lambda_{BI}^{\text{no}} \right) + \\
\left( \lambda_{LGB}^{\text{yes}} - \lambda_{LGB}^{\text{no}} \right) = \alpha + \beta_G^I + \beta_I^L + \beta_B^I
\]
Loglinear Model vs Logistic Regression

- Not every loglinear model can be expressed as logistic regression
- But: every logistic regression has loglinear counterpart

**Conversion**: in loglinear model include highest interaction of explanatory variables and all interactions of response and explanatory variables
Association Graphs

• **Path** exists, if nodes A and B are connected

• A and B **separated** by C, if all paths intersect C (could be a set of variables)
  If such C exists, A and B are *separable*

X and V are separated by \{Z\}
U and W are not separable
U and V are separated by \{Z\} and by \{X,Z\}
and by \{X,Y,Z\}
Collapsibility Condition

- X and Y are conditionally independent, if they are separable
- Conditional associations are usually different from marginal associations (remember Simpson’s paradox)
- If collapsibility condition holds: marginal and conditional association are the same
Collapsibility Condition

• Three-way table:
  XY marginal and conditional association are the same, if either X,Z or Y,Z are conditionally independent

• Multiway table:
  for mutually exclusive subsets A,B,C , we can collapse over C w/o changing associations between A and B, if B separates A and C or A separates B and C
Marijuana - All

Alcohol

Cigarette

Marijuana

sex

race
Check 2-way relationships between response set and covariates.
Automatic Search

```r
mj.main <- glm(count~., family=poisson(link=log), data=mj)
summary(mj.main)

lower <- count~sex*race+(cigarette+alcohol+marijuana)^2

library(MASS)
mj.step <- stepAIC(mj.two, scope=list(upper=formula(mj.two), lower=lower), direction="both")
```

Step:  AIC=180.97
count ~ cigarette + alcohol + marijuana + sex + race + cigarette:alcohol +
cigarette:marijuana + alcohol:marijuana + alcohol:sex + alcohol:race +
marijuana:sex + marijuana:race + sex:race

<table>
<thead>
<tr>
<th>Df</th>
<th>Deviance</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;none&gt;</td>
<td>16.735</td>
<td>180.97</td>
</tr>
<tr>
<td>+ cigarette:sex</td>
<td>1</td>
<td>15.784</td>
</tr>
<tr>
<td>- marijuana:race</td>
<td>1</td>
<td>19.909</td>
</tr>
<tr>
<td>+ cigarette:race</td>
<td>1</td>
<td>16.317</td>
</tr>
<tr>
<td>- alcohol:race</td>
<td>1</td>
<td>21.290</td>
</tr>
<tr>
<td>- alcohol:sex</td>
<td>1</td>
<td>22.021</td>
</tr>
<tr>
<td>- marijuana:sex</td>
<td>1</td>
<td>25.811</td>
</tr>
</tbody>
</table>

Cigarette use is conditionally independent from sex and race given alcohol and Marijuana use
Closer Investigation

- interaction marijuana:race not significant (remove)
- would like to also remove marijuana:sex
- results in model with 28 deviance and 20 degrees of freedom - no apparent lack of fit
- then sex and race is linked to response variables only through alcohol use
Association Models
Example: Mental Health

- SES of parents and mental health of 1660 students recorded.
- SES is measured from A=high to F=low,
- Mental Health is rated in four levels from well to poor
Mental Health

prodplot(data=mh, count~status+ses, c("vspine","hspine")+aes(fill=status))
Mental Health

- Model of independence obviously does not fit
- But: there is some pattern discernible
- Idea: make use of ordinal structure in variables
Association Models

• For ordinal variable X we have scores $u_1 \leq u_2 \leq \ldots \leq u_p$ (most commonly $u_i = i$)
  similarly, for ordinal variable Y assume scores $v_j$

• describe interaction term using scores
Association Models

• Uniform association
  \[ \lambda_{ij}^{XY} = \beta u_i v_j \]

• Row/Column effects model
  \[ \lambda_{ij}^{XY} = \beta_i v_j \quad \lambda_{ij}^{XY} = \beta_j u_i \]

• Column & Row effects model (not linear)
  \[ \lambda_{ij}^{XY} = \beta_i \beta_j \]
```r
glm(formula = count ~ ses + status + sesn:statusn, family = poisson(link = log),
     data = mh)

Deviance Residuals:
     Min       1Q   Median       3Q      Max
-1.2663  -0.3285  0.2025   0.3912  1.0820

Coefficients:
                         Estimate Std. Error    z value  Pr(>|z|)
(Intercept)          3.8316927  0.0954496  40.143180  < 2e-16 ***
  sesB                  0.1768209  0.0988097   1.790748  0.07353 .
  sesC                  0.5703522  0.1195459   4.770975 1.84e-06 ***
  sesD                  1.0880165  0.1452784   7.488930 6.94e-14 ***
  sesE                  0.9347058  0.1785394   5.234586 1.65e-07 ***
  sesF                  0.9435705  0.2088280   4.517503 6.23e-06 ***
  statusmoderate       0.2646119  0.0939875   2.815275  0.00487 **
  statusmild         1.0888208  0.1291145   8.432427  < 2e-16 ***
  statuswell          0.7099136  0.1746037   4.065739 4.79e-05 ***
sesn:statusn         -0.0906915  0.0150058  -6.043254 1.51e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 217.3998  on 23  degrees of freedom
Residual deviance:  9.8951  on 14  degrees of freedom
AIC: 174.07

Number of Fisher Scoring iterations: 4
```
Raw vs Fitted

Combine ses A and B?
## Party Affiliation (in 04)

<table>
<thead>
<tr>
<th>Gender</th>
<th>Race</th>
<th>Party</th>
<th>Number</th>
<th>Number stats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>Black</td>
<td>Democrat</td>
<td>4</td>
<td>Min.: 4.00</td>
</tr>
<tr>
<td>Male</td>
<td>White</td>
<td>Independent</td>
<td>4</td>
<td>1st Qu.: 14.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Republican</td>
<td>4</td>
<td>Median: 91.50</td>
</tr>
</tbody>
</table>

- Mean: 83.42
- 3rd Qu.: 130.50
- Max.: 176.00

**Charts:**
- **Gender** by Party affiliation:
  - Female:
    - Democrat: 6
    - Independent: 6
    - Republican: 4
  - Male:
    - White: 6
    - Independent: 4
    - Republican: 4

- **Race** by Party affiliation:
  - Black:
    - Democrat: 6
    - Independent: 4
    - Republican: 4
  - White:
    - Democrat: 6
    - Independent: 4
    - Republican: 4
Party Affiliation

```
glm(formula = Number ~ Party + Race + Gender + Racen:Partyn +
     Gendern:Partyn, family = poisson(link = log), data = party)

Deviance Residuals:
   1       2       3       4       5       6       7
 0.148681 -0.014165 -0.117567 -0.008349  0.107635  0.076275 -0.111428
   8      9     10     11     12
-0.165061 -0.163533 -0.380321  0.135880  0.449677

Coefficients:
               Estimate Std. Error   z value     Pr(>|z|)
(Intercept)    2.600823  0.145496     17.876 < 2e-16 ***
PartyIndependent -2.862458  0.304212    -9.409 < 2e-16 ***
PartyRepublican -5.410051  0.618489    -8.747 < 2e-16 ***
RaceWhite       -0.005983  0.228950    -0.026    0.979151
GenderMale     -0.576295  0.158528    -3.635     0.000278 ***
Racen:Partyn    1.135962  0.147876     7.682     1.57e-14 ***
Partyn:Gendern  0.289683  0.075876     3.818     0.000135 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 711.35267  on 11  degrees of freedom
Residual deviance:  0.48532  on  5  degrees of freedom
AIC: 82.555

Number of Fisher Scoring iterations: 3
```
Party & Ideology

Self-proclaimed ideology and Party affiliation

![Graph showing the distribution of political affiliation by ideology.]
Models

• Main effect obviously not good (not independent)
• Uniform fits better, but is not a good fit yet
• Row effects for Affiliation almost perfect fit
glm(formula = count ~ Affil + Ideology + Affil:Ideon, family = poisson(link = log),
data = table.9.5)

Deviance Residuals:

          1         2         3         4         5         6         7         8
 0.5405  -0.4337  -0.3924  -0.9929   0.6721   0.3759   0.6506  -0.3993

         9
-0.1389

Coefficients: (1 not defined because of singularities)

            Estimate Std. Error z value Pr(>|z|)
(Intercept)  6.06984   0.18677 32.500   < 2e-16 ***
AffilIndependent -0.81690   0.21899  3.730   0.000191 ***
AffilDemocrat  -1.53045   0.23518 -6.508 7.64e-11 ***
IdeologyModerate  0.58891   0.09619  6.122  9.22e-10 ***
IdeologyLiberal  0.37791   0.13355  2.830  0.004658 **
AffilRepublican:Ieon  -1.21336   0.13042 -9.304  < 2e-16 ***
AffilIndependent:Ieon -0.27074   0.09061 -2.988  0.002809 **
AffilDemocrat:Ieon         NA         NA         NA         NA

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 246.5190  on 8 degrees of freedom
Residual deviance:  2.8149  on 2 degrees of freedom
AIC: 74.9

Number of Fisher Scoring iterations: 3
Party & Ideology

Raw vs Fit