Chapter 7 Loglinear Models for Contingency Tables

Loglinear Models For Contingency Tables

- Logistic regression and other models in Ch 3-6 distinguish between a response variable Y and explanatory vars x_1, x_2, etc.
- Loglinear models for contingency tables treat all variables as response variables, like multivariate analysis.

  Ex. Survey of high school seniors (see text, p.209):

  - Y_1: used alcohol? (yes, no)
  - Y_2: cigarettes? (yes, no)
  - Y_3: marijuana? (yes, no)

  Interested in patterns of dependence and independence among the variables:

  - Any variables (conditionally) independent?
  - Strength of associations?
  - Homogeneous associations?
  - Interactions?

Loglinear Models for Two-Way Tables

In a $I \times J$ table, X and Y are independent if

$$\mathrm{P}(X = i, Y = j) = \mathrm{P}(X = i)\mathrm{P}(Y = j) \quad \text{for all } i, j$$
$$\text{i.e.,} \quad \pi_{ij} = \pi_{i+}\pi_{+j}$$

For expected cell frequencies,

$$\begin{aligned} \mu ij &= n\pi_{ij} && \text{(in general)} \\ &= n\pi_{i+}\pi_{+j} && \text{(if } X, Y \text{ indep.)} \end{aligned}$$

Loglinear models treat cell counts n_{ij} as Poisson and use log link

$$\begin{aligned} \log(\mu_{ij}) &= \lambda + \lambda_i^X + \lambda_j^Y && \text{(if } X, Y \text{ indep.)} \\ \log(\mu_{ij}) &= \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY} && \text{(in general)} \end{aligned}$$

If X, Y indep., then

$$\mu_{ij} = e^{\lambda} \exp(\lambda_i^X) \exp(\lambda_j^Y),$$

where $\exp(\lambda_i^X) \propto \pi_{i+}$, $\exp(\lambda_j^Y) \propto \pi_{+j}$.

Poisson-Multinomial Connection

If $Y_1, \ldots, Y_J$ are independent Poisson random variables, and

$$Y_j \sim \text{Poisson}(\mu_j), \quad j = 1, \ldots, J$$

then **given** $Y_1 + \cdots + Y_J = n$,

$$(Y_1, Y_2, \ldots, Y_J) \sim \text{Multinom}(n; \pi_1, \pi_2, \ldots, \pi_J),$$

where

$$\pi_j = \frac{\mu_j}{\mu_1 + \ldots + \mu_J}.$$

Consider an $I \times J$ contingency table that cross-classifies n subjects.

X categories	Y categories $Y=1$	$Y=2$	$\cdots$	$Y=J$	X margin
$X=1$	n_{11}	n_{12}	$\cdots$	n_{1J}	n_{1+}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$X=I$	n_{I1}	n_{I2}	$\cdots$	n_{IJ}	n_{I+}
Y margin	n_{+1}	n_{+2}	$\cdots$	n_{+J}	n

Let $\{\pi_{ij}\}$ be the joint cell prob. $\pi_{ij} = \mathrm{P}(X=i, Y=j)$.

- In Ch 4-6, cell counts n_{ij} are modeled as binomial or multinomial.
- In Ch 7, n_{ij}'s are modeled as indep. **Poisson**

$$n_{ij} \sim \text{Poisson}(\mu_{ij}), \quad \text{where } \mu_{ij} = n\pi_{ij}.$$

  By the Poisson-Multinomial connection, given marginal total $n = n_{++}$ or n_{i+} or n_{+j}, the cell counts n_{ij} are still binomial or multinomial, consistent w/ the binomial or multinomial models in Ch 4-6.

Residual Degrees of Freedom

For a Poisson loglinear model,

$$\text{Residual df} = \# \text{ of Poisson counts} - \# \text{ of parameters}$$

Here $\#$ of Poisson counts $= \#$ cells in table.

Just like logistic models contingency tables, loglinear models like $\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y$ have redundant parameters.

Think of dummy variables for each variable.
Number of dummies is one less than number of levels of variable.
Products of dummy variables correspond to "interaction" terms.

- $(I-1)$ of $\{\lambda_i^X\}$ are non-redundant
- $(J-1)$ of $\{\lambda_j^Y\}$ are non-redundant
- $(I-1)(J-1)$ of $\{\lambda_{ij}^{XY}\}$ are non-redundant

For an $I \times J$ contingency table, number of cells $= IJ$:

- General model: $\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}$

Parameter	Nonredundant
λ	1
λ_i^X	$I-1$
λ_j^Y	$J-1$
λ_{ij}^{XY}	$(I-1)(J-1)$
Total	IJ

  Residual df $= \#$ of cells $- \#$ of parameters $= IJ - IJ = 0$
  So for 2-way table, the general model is the saturated model.

- Independence model: $\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y$

$$\begin{aligned} \#\text{ of parameters} &= 1 + (I-1) + (J-1) \\ &= I + J - 1 \\ \text{Residual df} &= IJ - (I+J-1) \\ &= (I-1)(J-1) \end{aligned}$$

Interpretation of Parameters

Under saturated model $\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}$,
log-odds-ratio comparing levels i and i' of X and j and j' of Y is

$$\begin{aligned} \log\left(\frac{\mu_{ij}\mu_{i'j'}}{\mu_{ij'}\mu_{i'j}}\right) &= \log\mu_{ij} + \log\mu_{i'j'} - \log\mu_{ij'} - \log\mu_{i'j} \\ &= \quad (\lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}) \\ &\quad + (\lambda + \lambda_{i'}^X + \lambda_{j'}^Y + \lambda_{i'j'}^{XY}) \\ &\quad - (\lambda + \lambda_i^X + \lambda_{j'}^Y + \lambda_{ij'}^{XY}) \\ &\quad - (\lambda + \lambda_{i'}^X + \lambda_j^Y + \lambda_{i'j}^{XY}) \\ &= \lambda_{ij}^{XY} + \lambda_{i'j'}^{XY} - \lambda_{ij'}^{XY} - \lambda_{i'j}^{XY}. \end{aligned}$$

For the independence model this is 0, and the odds-ratio is $e^0 = 1$.

- As the saturated model fits the data perfectly, the MLEs for the parameters of the loglinear model will make the model fitted odds ratio equal to the empirical odds ratio

$$\frac{\widehat{\mu}_{ij}\widehat{\mu}_{i'j'}}{\widehat{\mu}_{ij'}\widehat{\mu}_{i'j}} = \exp(\widehat{\lambda}_{ij}^{XY} + \widehat{\lambda}_{i'j'}^{XY} - \widehat{\lambda}_{ij'}^{XY} - \widehat{\lambda}_{i'j}^{XY}) = \frac{n_{ij}n_{i'j'}}{n_{ij'}n_{i'j}}$$

Remark

- Loglinear models (both independence and saturated one) treat X and Y symmetrically. Unlike, e.g., logistic models where Y = response, X = explanatory.
- To test the independence of X and Y, the LR test comparing the independence model and saturated model is equivalent to the G^2 test of independence in Section 2.5.

Loglinear Models for Three-Way Tables

In a $I \times J \times K$ table w/ cell counts $\{n_{ijk}\}$, the saturated model is the 3-way interaction model, denoted as (XYZ) is

$$\log(\mu_{ijk}) = \lambda + \underbrace{\lambda_i^X + \lambda_j^Y + \lambda_k^Z}_{\text{main effects}} + \underbrace{\lambda_{ij}^{XY} + \lambda_{jk}^{YZ} + \lambda_{ik}^{XZ}}_{\text{2-way interactions}} + \underbrace{\lambda_{ijk}^{XYZ}}_{\text{3-way interactions}}$$

- How many non-redundant parameters for each term?

$$\begin{aligned} & 1 + \underbrace{(I-1)}_{X \text{ main effects}} + \underbrace{(J-1)}_{Y \text{ main effects}} + \underbrace{(K-1)}_{Z \text{ main effects}} \\ & + \underbrace{(I-1)(J-1)}_{XY \text{ interactions}} + \underbrace{(J-1)(K-1)}_{YZ \text{ interactions}} + \underbrace{(I-1)(K-1)}_{XZ \text{ interactions}} \\ & + \underbrace{(I-1)(J-1)(K-1)}_{XYZ \text{ interactions}} \\ = & \, I \times J \times K = \# \text{ of cell counts} \end{aligned}$$

There are several simplifications of the saturated model for 3-way table. We denote them by their highest order interaction terms

- (XY, YZ, XZ) — Homogeneous Association Model
$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ} + \lambda_{ik}^{XZ}$$
- (YZ, XZ) — Conditional Independence Model
$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{jk}^{YZ} + \lambda_{ik}^{XZ}$$
- Independence Model (X, Y, Z)
$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z$$
    - X, Y, Z are independent.
    - XY, YZ, XZ odds ratios are all zero
- (X, YZ)
$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{jk}^{YZ}$$
    - X is independent of (Y, Z), though (Y, Z) might be dependent.

(XY, YZ, XZ) — Homogeneous Association Model

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

$X - Y$ odds ratios are the same at all levels of Z: if Z is fixed at k log-odds-ratio comparing levels i and i' of X and j and j' of Y is

$$\begin{aligned} \log\left(\frac{\mu_{ijk}\mu_{i'j'k}}{\mu_{ij'k}\mu_{i'jk}}\right) &= \log \mu_{ijk} + \log \mu_{i'j'k} - \log \mu_{ij'k} - \log \mu_{i'jk} \\ &= (\lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ} + \lambda_{ik}^{XZ}) \\ &\quad + (\lambda + \lambda_{i'}^X + \lambda_{j'}^Y + \lambda_k^Z + \lambda_{i'j'}^{XY} + \lambda_{j'k}^{YZ} + \lambda_{i'k}^{XZ}) \\ &\quad - (\lambda + \lambda_i^X + \lambda_{j'}^Y + \lambda_k^Z + \lambda_{ij'}^{XY} + \lambda_{j'k}^{YZ} + \lambda_{ik}^{XZ}) \\ &\quad - (\lambda + \lambda_{i'}^X + \lambda_j^Y + \lambda_k^Z + \lambda_{i'j}^{XY} + \lambda_{jk}^{YZ} + \lambda_{i'k}^{XZ}) \\ &= \underbrace{\lambda_{ij}^{XY} + \lambda_{i'j'}^{XY} - \lambda_{ij'}^{XY} - \lambda_{i'j}^{XY}}_{\text{does not depend on } k}. \end{aligned}$$

Similarly, $Y - Z$ odds ratio same at all levels of X, and $X - Z$ odds ratio same at all levels of Y, because model has no three-factor interaction.

(YZ, XZ) — Conditional Independence Model

$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{jk}^{YZ} + \lambda_{ik}^{XZ}$$

- X and Y are conditionally independent, given Z because

$$\log\left(\frac{\mu_{ijk}\mu_{i'j'k}}{\mu_{ij'k}\mu_{i'jk}}\right) = \lambda_{ij}^{XY} + \lambda_{i'j'}^{XY} - \lambda_{ij'}^{XY} - \lambda_{i'j}^{XY} = 0$$

- $X - Z$ odds ratio is the same at all levels of Y
  $Y - Z$ odds ratio same at all levels of X

Example (Alcohol, Cigarette, & Marijuana Use)

Alcohol Use	Cigarette Use	Marijuana Use Yes	Marijuana Use No
Yes	Yes	911	538
	No	44	456
No	Yes	3	43
	No	2	279

```
> A = gl(2, 4, length = 8, labels = c("Y","N"))
> C = gl(2, 2, length = 8, labels = c("Y","N"))
> M = gl(2, 1, length = 8, labels = c("Y","N"))
> Freq = c(911,538,44,456,3,43,2,279)
> teens.df = data.frame(A,C,M, Freq)
> teens.df
  A C M Freq
1 Y Y Y  911
2 Y Y N  538
3 Y N Y   44
4 Y N N  456
5 N Y Y    3
6 N Y N   43
7 N N Y    2
8 N N N  279
```



```
> teens.AC.AM.CM = glm(Freq ~ A*C + C*M + A*M,
                       family=poisson, data=teens.df)
> summary(teens.AC.AM.CM)
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  6.81387    0.03313 205.699  < 2e-16 ***
AN          -5.52827    0.45221 -12.225  < 2e-16 ***
CN          -3.01575    0.15162 -19.891  < 2e-16 ***
MN          -0.52486    0.05428  -9.669  < 2e-16 ***
AN:CN        2.05453    0.17406  11.803  < 2e-16 ***
CN:MN        2.84789    0.16384  17.382  < 2e-16 ***
AN:MN        2.98601    0.46468   6.426 1.31e-10 ***
---
(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 2851.46098  on 7  degrees of freedom
Residual deviance:    0.37399  on 1  degrees of freedom
AIC: 63.417
```

The (AC, AM, CM) model fits well: Deviance $= 0.37$ on 1 df.

Note: As a LRT, goodness-of-fit on previous slide is comparing to saturated model.

```
> teens.ACM = update(teens.AC.AM.CM, . ~ A*C*M)
> anova(teens.AC.AM.CM, teens.ACM, test="Chisq")
Analysis of Deviance Table

Model 1: Freq ~ A * C + C * M + A * M
Model 2: Freq ~ A + C + M + A:C + A:M + C:M + A:C:M
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1         1    0.37399
2         0    0.00000  1  0.37399   0.5408
```

And none of the interaction terms can be dropped:

```
> drop1(teens.AC.AM.CM,test="Chisq")
Single term deletions
Model:
Freq ~ A * C + C * M + A * M
       Df Deviance    AIC    LRT  Pr(>Chi)
<none>        0.37  63.42
A:C     1   187.75 248.80 187.38 < 2.2e-16 ***
C:M     1   497.37 558.41 497.00 < 2.2e-16 ***
A:M     1    92.02 153.06  91.64 < 2.2e-16 ***
```

Just like all GLMs, one can use likelihood ratio test to compare between models.

E.g., to test for conditional independence of A and C given M:

```
> teens.AM.CM = update(teens.AC.AM.CM, . ~ A*M + C*M)
> anova(teens.AM.CM, teens.AC.AM.CM, test="Chisq")
Analysis of Deviance Table

Model 1: Freq ~ A + M + C + A:M + M:C
Model 2: Freq ~ A * C + C * M + A * M
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1         2    187.754
2         1      0.374  1   187.38 < 2.2e-16 ***
```

Strong evidence that A, C are not conditionally indep. given M.

Example (Automobile Accidents)

68,694 passengers of autos and light trucks accidents in Maine

Gender	Location	Seat Belt	Injury No	Injury Yes
Female	Urban	No	7,287	996
		Yes	11,587	759
	Rural	No	3,246	973
		Yes	6,134	757
Male	Urban	No	10,381	812
		Yes	10,969	380
	Rural	No	6,123	1,084
		Yes	6,693	513

```
> G = gl(2, 8, 16, labels = c("F","M"))          # Gender
> L = gl(2, 4, 16, labels = c("Urban","Rural"))  # Location
> S = gl(2, 2, 16, labels = c("N","Y"))          # Seat-belt
> I = gl(2, 1, 16, labels = c("N","Y"))          # Injured
> G = relevel(G, ref="M")
> S = relevel(S, ref="Y")
> Freq = c(7287, 996, 11587, 759, 3246, 973, 6134, 757,
        10381, 812, 10969, 380, 6123, 1084, 6693, 513)
```



```
> G.L.S.I = glm(Freq ~ G+L+S+I, family="poisson")
> GL.GS.GI.LS.LI.SI = glm(Freq ~ (G+L+S+I)^2, family="poisson")
> GIL.GIS.GLS.ILS = glm(Freq ~ (G+L+S+I)^3, family="poisson")
> deviance(G.L.S.I)
[1] 2792.771
> deviance(GL.GS.GI.LS.LI.SI)
[1] 23.35099
> deviance(GIL.GIS.GLS.ILS)
[1] 1.325317
```

Goodness of Fit:

Model	Deviance	d.f.	P-value
(G, I, L, S)	2792.78	11	$< 2.2 \times 10^{16}$
(GI, GL, GS, IL, IS, LS)	23.35	5	0.00029
(GIL, GIS, GLS, ILS)	1.325	1	0.25

- Why are the df. for the 3 models 11, 5, and 1?
- need a model more complex than (GI, GL, GS, IL, IS, LS) but simpler than (GIL, GIS, GLS, ILS).

Goodness of Fit:

Model	Deviance	d.f.	P-value
(GIL, GS, IS, LS)	18.5693	4	0.00095
(GIS, GL, IL, LS)	22.8468	4	0.00014
(GLS, GI, IL, IS)	7.4645	4	0.1133
(ILS, GI, GL, GS)	20.6334	4	0.00037
(GLS, ILS, GI)	3.5914	3	0.3091
(GLS, GIL, IS)	4.4909	3	0.21310
$\vdots$	$\vdots$	$\vdots$	$\vdots$
(GIL, GLS, ILS)	1.3670	2	0.5048
(GIL, GIS, ILS)	16.1391	2	0.00031
(GIS, GLS, ILS)	3.5624	2	0.16843
(GIL, GIS, GLS)	4.3720	2	0.11237

- (GLS, GI, IL, IS) is the simplest model that looks acceptable in Goodness of Fit

```
> add1(GL.GS.GI.LS.LI.SI, scope= ~.+G*I*L+G*I*S+G*L*S+I*L*S)
Single term additions

Model:
Freq ~ (G + L + S + I)^2
       Df Deviance    AIC
<none>     23.3510 198.81
G:L:I   1  18.5693 196.03
G:S:I   1  22.8468 200.31
G:L:S   1   7.4645 184.92
L:S:I   1  20.6334 198.09

> drop1(GIL.GIS.GLS.ILS)
Single term deletions

Model:
Freq ~ (G + L + S + I)^3
       Df Deviance    AIC
<none>      1.3253 184.78
G:L:S   1  16.1391 197.60
G:L:I   1   3.5624 185.02
G:S:I   1   1.3670 182.83
L:S:I   1   4.3720 185.83
```


Model (GI, GL, GS, IL, IS, LS)

```
> summary(GL.GS.GI.LS.LI.SI)
Coefficients:
(... estimates for main effects are omitted ...)
GF:LRural   -0.209922   0.016124  -13.019  < 2e-16 ***
GF:SN       -0.459925   0.015682  -29.328  < 2e-16 ***
GF:IY        0.540528   0.027219   19.859  < 2e-16 ***
LRural:SN   -0.084926   0.016194   -5.244 1.57e-07 ***
LRural:IY    0.755025   0.026949   28.017  < 2e-16 ***
SN:IY        0.813995   0.027618   29.473  < 2e-16 ***
```

- SI conditional odds ratio: $e^{0.814} \approx 2.257$
  Odds of injury when not wearing seal-belt are 2.257 times of the odds when wearing, constant across levels of G and L
- 95% Wald CI for SI conditional odds ratio:
  $$e^{0.814 \pm 1.96 \times 0.0276} = (e^{0.760}, e^{0.868}) = (2.138, 2.382)$$
- GS conditional odds ratio: $e^{-0.460} \approx 0.63$
  Odds of not wearing seat belt for women are 0.63 times the odds for men, constant across levels of I and L

Model (GLS, GI, IL, IS)

```
> GLS.GI.IL.IS = glm(Freq ~ G*L*S+(G+L+S)*I, family="poisson")
> summary(GLS.GI.IL.IS)
Coefficients:
(... estimates for main effects are omitted ...)
GF:LRural    -0.154164   0.021346   -7.222 5.12e-13 ***
GF:SN        -0.413282   0.019555  -21.134  < 2e-16 ***
LRural:SN    -0.028939   0.021442   -1.350    0.177
GF:IY         0.544829   0.027266   19.982  < 2e-16 ***
LRural:IY     0.758058   0.026972   28.105  < 2e-16 ***
SN:IY         0.817097   0.027651   29.551  < 2e-16 ***
GF:LRural:SN -0.128580   0.032277   -3.984 6.78e-05 ***
```

- SI conditional odds ratio $e^{0.817} \approx 2.264$, still homogeneous across levels of G and L
- GS conditional odds ratio is $e^{-0.413} \approx 0.66$ in urban area, and $e^{-0.413-0.1286} \approx 0.58$ in rural area
  Men are less likely to wear seat belt then women, and even more so in rural area
- still have homogeneous GI, IL, IS association, but not GL, GS, and LS.

Odds Ratio	(GI, GL, GS, IL, IS, LS)	(GLS, GI, IL, IS)
GI (f,m) v.s. (yes,no)	1.72	1.72
LI (rural, urban) v.s. (yes,no)	2.13	2.13
SI (no,yes) v.s. (yes, no)	2.26	2.26
GL (f,m) v.s. (rural, urban) (S=yes)	0.81	0.86
GL (f,m) v.s. (rural, urban) (S=no)	0.81	0.75
GS (f,m) v.s. (no, yes) (L=urban)	0.63	0.66
GS (f,m) v.s. (no, yes) (L=rural)	0.63	0.58
LS (rural, urban) v.s. (no,yes) (G=m)	0.92	0.97
LS (rural, urban) v.s. (no,yes) (G=f)	0.92	0.85

Large Samples and Statistical Versus Practical Significance

For large sample sizes, statistically significant effects can be weak and unimportant.

- Though model (GLS, GI, IL, IS) seems to fit better than (GI, GL, GS, IL, IS, LS). However, the three-factor interaction is weak as shown in the Table on the previous slide.

Loglinear Cell Residuals

```
> std.res1 = round(rstandard(GL.GS.GI.LS.LI.SI,type="pearson"),2)
> std.res1 = xtabs(std.res1 ~ G+L+S+I)
> ftable(std.res1, col.vars=c("S","I"))
        S     Y           N
        I     N     Y     N     Y
G L
M Urban    3.84 -0.49 -2.66 -1.72
  Rural   -3.58 -0.31  2.37  2.29
F Urban   -4.70  2.04  3.64  0.15
  Rural    4.53 -1.32 -3.45 -0.79

> std.res2 = round(rstandard(GLS.GI.IL.IS,type="pearson"),2)
> std.res2 = xtabs(std.res2 ~ G+L+S+I)
> ftable(std.res2, col.vars=c("S","I"))
        S     Y           N
        I     N     Y     N     Y
G L
M Urban    0.63 -0.63  1.16 -1.16
  Rural   -0.28  0.28 -1.40  1.40
F Urban   -2.48  2.48  0.71 -0.71
  Rural    2.20 -2.20 -0.46  0.46
```


Loglinear-Logit Connection

Loglinear models

- all variables are response variables
- examine relationships between all variables
- model joint probabilities
  e.g., for 3-way tables, model $\pi_{ijk} = \mathrm{P}(X = i, Y = j, Z = k)$

Logistic models

- One (binary) response variable Y and the rest are explanatory $X, Z, W \ldots$
- examine relationship between the response Y and explanatory variables $(X, Z, W \ldots)$
  but ignore relationships between explanatory variables $(X, Z, W \ldots)$
- model conditional probabilities
  e.g., for 3-way tables, model $\mathrm{P}(Y = j | X = i, Z = k)$

E.g., loglinear model (XYZ) for 3-way tables:

$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ} + \lambda_{ik}^{XZ} + \lambda_{ijk}^{XYZ}$$

If Y is binary and is treated as response,

$$\begin{aligned}
\text{logit}[\mathrm{P}(Y=1)] &= \log\left(\frac{\mathrm{P}(Y=1)}{1-\mathrm{P}(Y=1)}\right) = \log\left(\frac{\mathrm{P}(Y=1|X,Z)}{\mathrm{P}(Y=2|X,Z)}\right)\\
&= \log\left(\frac{\mu_{i1k}}{\mu_{i2k}}\right) = \log(\mu_{i1k}) - \log(\mu_{i2k})\\
&= \ (\lambda + \lambda_i^X + \lambda_1^Y + \lambda_k^Z + \lambda_{i1}^{XY} + \lambda_{1k}^{YZ} + \lambda_{ik}^{XZ} + \lambda_{i1k}^{XYZ})\\
&\quad - (\lambda + \lambda_i^X + \lambda_2^Y + \lambda_k^Z + \lambda_{i2}^{XY} + \lambda_{2k}^{YZ} + \lambda_{ik}^{XZ} + \lambda_{i2k}^{XYZ})\\
&= \underbrace{(\lambda_1^Y - \lambda_2^Y)}_{\alpha} + \underbrace{(\lambda_{i1}^{XY} - \lambda_{i2}^{XY})}_{\beta_i^X} + \underbrace{(\lambda_{1k}^{YZ} - \lambda_{2k}^{YZ})}_{\beta_k^Z}\\
&\quad + \underbrace{(\lambda_{i1k}^{XYZ} - \lambda_{i2k}^{XYZ})}_{\beta_{ik}^{XZ}}\\
&= \alpha + \beta_i^X + \beta_k^Z + \beta_{ik}^{XZ}
\end{aligned}$$

Example (Alcohol, Cigarette, & Marijuana Use)

If treat M (Marijuana Use) as the binary response,

```
> teens.df
  A C M Freq
1 Y Y Y  911
2 Y Y N  538
3 Y N Y   44
4 Y N N  456
5 N Y Y    3
6 N Y N   43
7 N N Y    2
8 N N N  279
> M.yes = Freq[c(1,3,5,7)]
> M.no =  Freq[c(2,4,6,8)]
> teensM.df = data.frame(teens.df[c(1,3,5,7),1:2],M.yes,M.no)
> teensM.df
  A C M.yes M.no
1 Y Y   911  538
3 Y N    44  456
5 N Y     3   43
7 N N     2  279
```


Observe the correspondence between coefficients of the loglinear models and the logistic model.

```
> ACM = glm(Freq ~ A*C*M, family="poisson",data=teens.df)
> summary(ACM)
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  6.81454    0.03313 205.682  < 2e-16 ***
AN          -5.71593    0.57830  -9.884  < 2e-16 ***
CN          -3.03035    0.15435 -19.633  < 2e-16 ***
MN          -0.52668    0.05437  -9.686  < 2e-16 ***  <-- (a)
AN:CN        2.62489    0.92583   2.835  0.00458 **
AN:MN        3.18927    0.59962   5.319 1.04e-07 ***  <-- (b)
CN:MN        2.86499    0.16696  17.159  < 2e-16 ***  <-- (c)
AN:CN:MN    -0.58951    0.94236  -0.626  0.53160      <-- (d)

> Mlogit.AC = glm(cbind(M.yes,M.no)~A*C,
                 family="binomial", data=teensM.df)
> summary(Mlogit.AC)
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.52668    0.05437   9.686  < 2e-16 *** <-- (a)
AN          -3.18927    0.59962  -5.319 1.04e-07 *** <-- (b)
CN          -2.86499    0.16696 -17.159  < 2e-16 *** <-- (c)
AN:CN        0.58951    0.94236   0.626    0.532     <-- (d)
```

We said the loglinear model (XYZ) and the logistic model $(X * Z)$ are equivalent

Likewise, for 3-way table if Y is the (binary) response induced logistic model for loglinear model are as follows

Loglinear Model	Logistic Model	Logistic Model Symbol	Equivalent?
(X, Y, Z)	α	(1)	
(Y, XZ)	α	(1)	Yes
(X, YZ)	$\alpha + \beta_k^Z$	(Z)	
(Z, XY)	$\alpha + \beta_i^X$	(X)	
(XY, XZ)	$\alpha + \beta_i^X$	(X)	Yes
(YZ, XZ)	$\alpha + \beta_k^Z$	(Z)	Yes
(XY, YZ)	$\alpha + \beta_i^X + \beta_k^Z$	$(X + Z)$	
(XY, YZ, XZ)	$\alpha + \beta_i^X + \beta_k^Z$	$(X + Z)$	Yes
(XYZ)	$\alpha + \beta_i^X + \beta_k^Z + \beta_{ik}^{XZ}$	$(X * Z)$	Yes

Rules:

- Drop the "Y" in terms that involve Y, e.g., $Y \to 1$, $XY \to X$, $YZ \to Z$, $XYZ \to XZ$
- Drop all terms not involving Y

However, not all induced logistic models are equivalent to the loglinear model they are induced from. Why?

Equivalent Loglinear and Logistic Models

For 3-way table if Y is the (binary) response, induced logistic model for loglinear model (XY, YZ, XZ) and (XY, YZ) are both the additive logistic model $(X + Z)$

$$\text{logit}[\mathrm{P}(Y = 1)] = \alpha + \beta_i^X + \beta_k^Z$$

However, only the loglinear model (XY, YZ, XZ) is equivalent to the logistic model model $(X + Z)$ but (XY, YZ) is not because...

Observe the correspondence between coefficients of the loglinear models (AC, AM, CM) and the logistic model $(A + C)$.

```
> AC.AM.CM = glm(Freq ~ A*C + C*M + A*M,
                      family=poisson, data=teens.df)
> summary(AC.AM.CM)
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  6.81387    0.03313 205.699  < 2e-16 ***
AN          -5.52827    0.45221 -12.225  < 2e-16 ***
CN          -3.01575    0.15162 -19.891  < 2e-16 ***
MN          -0.52486    0.05428  -9.669  < 2e-16 *** <--- (a)
AN:CN        2.05453    0.17406  11.803  < 2e-16 ***
CN:MN        2.84789    0.16384  17.382  < 2e-16 *** <--- (b)
AN:MN        2.98601    0.46468   6.426 1.31e-10 *** <--- (c)
---
Residual deviance:    0.37399  on 1  degrees of freedom
> Mlogit.A.C = glm(cbind(M.yes,M.no)~A+C, family="binomial",
                  data=teensM.df)
> summary(Mlogit.A.C)
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.52486    0.05428   9.669  < 2e-16 *** <--- (a)
AN          -2.98601    0.46468  -6.426 1.31e-10 *** <--- (c)
CN          -2.84789    0.16384 -17.382  < 2e-16 *** <--- (b)
---
Residual deviance:   0.37399  on 1  degrees of freedom
```


No correspondence between fitted coefficients between log-linear model (AM, CM) and logistic model $(A + C)$.

```
> AM.CM = glm(Freq ~ C*M + A*M,
               family=poisson, data=teens.df)
> summary(AM.CM)
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  6.81261    0.03316 205.450   <2e-16 ***
CN          -2.98919    0.15111 -19.782   <2e-16 ***
MN          -0.72847    0.05538 -13.154   <2e-16 ***
AN          -5.25227    0.44837 -11.714   <2e-16 ***
CN:MN        3.22431    0.16098  20.029   <2e-16 ***
MN:AN        4.12509    0.45294   9.107   <2e-16 ***
---
Residual deviance:  187.75  on 2  degrees of freedom

> Mlogit.A.C = glm(cbind(M.yes,M.no)~A+C, family="binomial",
                  data=teensM.df)
> summary(Mlogit.A.C)
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.52486    0.05428   9.669  < 2e-16 ***
AN          -2.98601    0.46468  -6.426 1.31e-10 ***
CN          -2.84789    0.16384 -17.382  < 2e-16 ***
```


Summary of Equivalent Loglinear and Logistic Models

A loglinear model has an equivalent logistic model must contain the highest order interaction term between ALL explanatory variables.

— A logistic model ignores relationships among explanatory variables, so it assumes nothing about their association structure

Equivalent loglinear and logistic models

- have identical fitted counts for all cell
- have identical deviance and hence the same goodness of fit.
- coefficients of logistic models can be derived from the equivalent loglinear model