Part II: Generalized Linear Models

Chapter II.5

Log-Linear Models

Topics

To be discussed...

- Two-Way Contingency Tables / Sampling Schemes
- Log-Linear Models for Two-Way Tables
- Generalized Odds Ratios for $I \times J$ Contingency Tables
- Log-Linear Models for Contingency Tables of Higher Dimension
- Model Selection in Multi-Way Tables
- Connection between Log-Linear and Logit Models

5. Log-Linear Models

II.5.1 Notation (Two-way Contingency Tables)

Table of observed counts: $\mathbf{n} = (n_{ij}) \in \mathbb{N}^{I \times J}$
Realization of the random matrix of counts: $\mathbf{N} = (N_{ij}) \sim \mathcal{M}(n, \boldsymbol{\pi})$

Classification variables: $X \in \{1, \dots, I\}$ and $Y \in \{1, \dots, J\}$
☞ Y: response variable – X: explanatory or response variable

Y

X	1	2	$\cdots$	j	$\cdots$	J	marginal
1	$n_{11}(\pi_{11})$	$n_{12}(\pi_{12})$	$\cdots$	$n_{1j}(\pi_{1j})$	$\cdots$	$n_{1J}(\pi_{1J})$	$n_{1+}(\pi_{1+})$
2	$n_{21}(\pi_{21})$	$n_{22}(\pi_{22})$	$\cdots$	$n_{2j}(\pi_{2j})$	$\cdots$	$n_{2J}(\pi_{2J})$	$n_{2+}(\pi_{2+})$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
i	$n_{i1}(\pi_{i1})$	$n_{i2}(\pi_{i2})$	$\cdots$	$n_{ij}(\pi_{ij})$	$\cdots$	$n_{iJ}(\pi_{iJ})$	$n_{i+}(\pi_{i+})$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
I	$n_{I1}(\pi_{I1})$	$n_{I2}(\pi_{I2})$	$\cdots$	$n_{Ij}(\pi_{Ij})$	$\cdots$	$n_{IJ}(\pi_{IJ})$	$n_{I+}(\pi_{I+})$
marginal	$n_{+1}(\pi_{+1})$	$n_{+2}(\pi_{+2})$	$\cdots$	$n_{+j}(\pi_{+j})$	$\cdots$	$n_{+J}(\pi_{+J})$	$n \quad (1)$

II.5.2 Sampling Schemes

The frequencies in contingency tables usually arise from one of the following sampling schemes.

- Multnomial: The table of cell counts $\mathbf{n} = (n_{ij})$ consists an observation of a multinomial response $\mathcal{M}ult(n, \boldsymbol{\pi})$. The sample size n is *fixed* a priori.
- Product Multnomial: The row (or column) vectors of counts are realizations of independent multinomial responses. The sample sizes $n_{1+}, \ldots, n_{I+}$ (or $n_{+1}, \ldots, n_{+J}$) are *fixed*.
- Poisson: The cell counts (n_{ij}) are realizations of independent Poisson processes $\mathcal{P}(\mu_{ij})$. The total sample size n is a *random variable*.

II.5.3 Proposition

If $N_i \overset{ind}{\sim} \mathcal{P}(\mu_i)$, $i = 1, \ldots, k$, independent Poisson random variables, then their joint distribution, conditional on their sum $\sum_{j=1}^{k} N_j = n$, is multinomial

$$(N_1, N_2, \ldots, N_k) \mid \sum_{j=1}^{k} N_j = n \sim \mathcal{M}(n, (\pi_1, \pi_2, \ldots, \pi_k)), \quad \text{with } \pi_i = \mu_i / n \ .$$

II.5.4 Example

We are interested in investigating the relation between age (X, in four categories: “18–25”, “26–45”, “46–65”, “$>$65”) and attitude towards refugees (Y, on a 5-level scale from negative to positive) among adults in Aachen.

1. A random sample of 400 residents of Aachen is drawn, interviewed and cross-classified according to X and Y.
2. Four independent samples of residents of Aachen, one for each age group, are drawn, of sample sizes 120, 100, 80 and 100, respectively, and asked on Y.
3. The residents of Aachen visiting the ‘Citizen office Aachen’ in a pre-specified week are interviewed and cross-classified according to X and Y.

It is clear that the underlying sampling scheme is in case (1) multinomial, in (2) product multinomial and in (3) independent Poisson in each cell.

II.5.5 Remark

Due to the connection between Poisson and Multinomial distributions, under either sampling scheme, the cell counts of a contingency table can be modeled using a Poisson GLM. In the multinomial case, the total count is the sample size and the cell probabilities are proportional to the Poisson means.

Log-Linear Models for Two-Way Tables

II.5.6 Independence Model ($\mathcal{M}_0$)

Let $\mathbf{N} = (N_{ij})$ be an $I \times J$ random table of cell counts, cross-classifying variables X (rows) and Y (columns), with $\mathbf{N} \sim \mathcal{M}ult(n, \boldsymbol{\pi})$. Then $\boldsymbol{\mu} = \mathsf{E}(\mathbf{N})$ is the table of expected cell counts with entries $\mu_{ij} = \mathsf{E}(N_{ij}) = n\pi_{ij}$. If X and Y are independent, it holds

$$\mu_{ij} = n\pi_{i+}\pi_{\cdot+}, \quad i = 1, \ldots, I; \; j = 1, \ldots, J.$$

The model of independence can equivalently be expressed in log-linear form

$$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y, \qquad i = 1, \ldots, I, \; j = 1, \ldots, J.$$

Identifiability constraints: $\lambda_1^X = \lambda_1^Y = 0$ [or $\sum_{i=1}^I \lambda_i^X = \sum_{j=1}^J \lambda_j^Y = 0$]

Degrees of Freedom: df($\mathcal{M}_0$)=$(I-1)(J-1)$

II.5.7 Saturated Model ($\mathcal{M}_1$)

Under the set–up of II.5.6, if the interaction between the variables X and Y is significant, then the saturated log-linear model is given by

$$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}, i = 1, \ldots, I, j = 1, \ldots, J.$$

Additional identifiability constraints: $\lambda_{1j}^{XY} = \lambda_{i1}^{XY} = 0, \forall\, i, j$ [or $\sum_{i=1}^{I} \lambda_{ij}^{XY} = \sum_{j=1}^{J} \lambda_{ij}^{XY} = 0$]

Degrees of Freedom: df($\mathcal{M}_1$)=0

Log-Linear Models as GLMs

II.5.8 Example

Consider the independence models for a 2×2 table with constraints $\lambda_1^X = \lambda_1^Y = 0$:

$$\begin{aligned}
\log(\mu_{11}) &= \lambda + \lambda_1^X + \lambda_1^Y = \lambda \\
\log(\mu_{12}) &= \lambda + \lambda_1^X + \lambda_2^Y = \lambda + \lambda_2^Y \\
\log(\mu_{21}) &= \lambda + \lambda_2^X + \lambda_1^Y = \lambda + \lambda_2^X \\
\log(\mu_{22}) &= \lambda + \lambda_2^X + \lambda_2^Y
\end{aligned}$$

In matrix form:

$$\log(\boldsymbol{\mu}) = \begin{pmatrix} \log(\mu_{11}) \\ \log(\mu_{12}) \\ \log(\mu_{21}) \\ \log(\mu_{22}) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} \lambda \\ \lambda_2^X \\ \lambda_2^Y \end{pmatrix} = \mathbf{X}\boldsymbol{\beta}$$

For the saturated model, the parameter vector becomes $\boldsymbol{\beta} = (\lambda, \lambda_2^X, \lambda_2^Y, \lambda_{22}^{XY})'$ and the model matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

II.5.9 Example

The data of the table below are from the General Social Survey basis for year 2008 (GSS2008). Responders are cross-classified by their opinion on the sufficiency of the amount of national spending for welfare and their educational level, measured by the highest degree they obtained.

☞ Is the opinion on national spending for welfare independent of the educational level of the responder?

	Highest Degree Obtained					
Welfare Spending	LT High School	High School	Junior College	Bachelor	Graduate	Total
too little	45	116	19	48	23	251
about right	40	167	33	68	41	349
too much	47	185	34	63	26	355
Total	132	468	86	179	90	955

☞

In order to fit the model of independence by glm, the data-table needs to be given in a vector form (say, freq). Furthermore, the row and column factors, WELFARE and DEGREE, respectively, have to be constructed. They are combined in a data frame, named nt.frame, as shown below. The factors are defined for a frequency vector of length $IJ = 15$ that expands the cells of the table by rows.

```
> NI <- 3
> NJ <- 5
> freq <- c(45,116,19,48,23,40,167,33,68,41,47,185,34,63,26)
> row.lb <- c(``too little'',``about right'',``too much'')
> col.lb <- c(``LT HS'',``HS'', ``JColg'',``BA'', ``Grad'')
> WELFARE <- gl(NI,NJ,length=NI*NJ, labels=row.lb)
> DEGREE <- gl(NJ,1,length=NI*NJ, labels=col.lb)
> nt.frame <- data.frame(freq,WELFARE,DEGREE)
```

☞

Then, the model of independence is fitted as follows.

```
> I.glm <- glm(freq ~ WELFARE+DEGREE,family=poisson,data=nt.frame)
> summary(I.glm)
```

```
Call:
glm(formula = freq ~ WELFARE + DEGREE, family = poisson,data = nt.frame)

Deviance Residuals:
Min              1Q    Median          3Q    Max

-1.3419       -0.5377    -0.1352       0.3366    1.6724

Coefficients:
               Estimate   Std. Error   z value   Pr(>|z|)
(Intercept)    3.54654     0.10253     34.590     < 2e-16 ***
WELFARE2       0.32962     0.08276      3.983     6.81e-05 ***
WELFARE3       0.34666     0.08247      4.204     2.63e-05 ***
DEGREE2        1.26567     0.09855     12.843     < 2e-16 ***
DEGREE3       -0.42845     0.13858     -3.092      0.00199 **
DEGREE4        0.30458     0.11473      2.655      0.00793 **
DEGREE5       -0.38299     0.13670     -2.802      0.00508 **
---
Signif. codes:  0 ``***'' 0.001 ``**'' 0.01 ``*''0.05 ``.'' 0.1 `` '' 1
( Dispersion parameter for poisson family taken to be 1 )
   Null deviance: 478.046 on 14 degrees of freedom
Residual deviance: 10.363 on 8 degrees of freedom
AIC: 110.74
Number of Fisher Scoring iterations:  4
```

Goodness of Fit

The value of the G^2 statistic is reported under ‘Residual Deviance’ and is saved in `I.glm$deviance`, as can be verified by typing

```
> names(I.glm)
```

Its asymptotic p-value is not provided but can easily be calculated by

```
> p.value <- 1-pchisq(I.glm$deviance, I.glm$df.residual)
```

We find p-value=0.240, thus the independence model describes adequately this dataset.

Residuals

The residuals saved in object `I.glm` are the raw residuals. The Pearsonian residuals are calculated by

```
> residuals(I.glm, type = c(``pearson''))
```

while the deviance by changing the type option to ` ``deviance'' `. The standardized residuals are obtained by

```
> rstandard(I.glm).
```

☞

The items of the output object are all in vector form but can be transformed to the more friendly table form by `xtabs()`. For example, the MLEs of the expected cell frequencies under independence and the standardized residuals are derived in table form by

```
> MLEs <- xtabs(I.glm$fitted.values ~ WELFARE+DEGREE,data=nt.frame)
> stdres <- xtabs(rstandard(I.glm) ~ WELFARE+DEGREE,data=nt.frame)
```

Thus, the standardized residuals are

```
> stdres
```

```
             DEGREE
WELFARE           LT HS        HS      JColg         BA       Grad
  too little  2.0983151 -1.039894 -0.9517240  0.1790943 -0.1654438
  about right -1.6533505 -0.543633  0.3659428  0.4422752  1.7955727
  too much   -0.4040979  1.462390  0.4702723 -0.6127615 -1.7788921
```

The only standardized residual that exceeds in absolute value 1.96, corresponds to cell $(1,1)$. That is, responders with educational level lower than high school tend to believe that welfare spending is too little with higher probability than expected under the independence model.

☞

The mosaic plot of the standardized residuals (stdres) is constructed by

```
> natfare <- xtabs(freq ~ WELFARE+DEGREE)
> mosaic(natfare, gp=shading_Friendly, residuals=stdres,
+    residuals_type="Std\nresiduals",labeling = labeling_residuals)
```

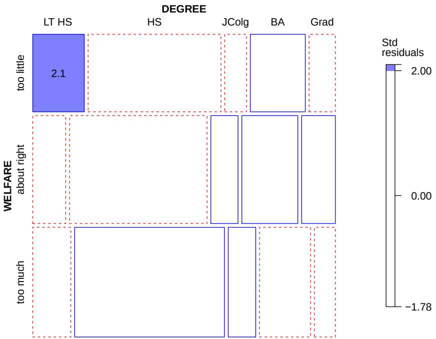

Figure: Mosaic plots of standardized residuals for the independence model
Requires: library(vcdExtra).

Odds Ratios and Log-Linear Models

☞ Log-Linear Model Parameters' Interpretation through the Odds Ratios

II.5.10 Odds Ratio

The *odds ratio* is a fundamental measure of association for a 2×2 contingency table. Consider

	Outcome	
Group	success	failure
A	π_{11}	π_{12}
B	π_{21}	π_{22}

Within group success probability $(j = 1)$:

$$\pi_{1|\mathsf{A}} = \pi_{1|i=1} = \pi_{11}/(\pi_{11} + \pi_{12})$$
$$\pi_{1|\mathsf{B}} = \pi_{1|i=2} = \pi_{21}/(\pi_{21} + \pi_{22})$$

Then, the *odds ratio* is defined by

$$\theta = \frac{\pi_{1|\mathsf{A}}/(1 - \pi_{1|\mathsf{A}})}{\pi_{1|\mathsf{B}}/(1 - \pi_{1|\mathsf{B}})} = \frac{\pi_{11}/\pi_{12}}{\pi_{21}/\pi_{22}} = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} \; .$$

II.5.11 Properties

- $\theta = 1 \Leftrightarrow \log(\theta) = 0$: independence.
- $\log(\theta) > 0$: positive dependence ($< 0 \Rightarrow$ negative dependence)
  Dependence becomes stronger as $|\log\theta|$ increases.
- The MLE of θ is $\hat{\theta} = \hat{\theta}(\mathbf{N}) = \frac{N_{11}N_{22}}{N_{12}N_{21}}$ with estimate $\hat{\theta}(\mathbf{n}) = \frac{n_{11}n_{22}}{n_{12}n_{21}}$, based on an observed table $\mathbf{n}$. Asymptotically it holds:

$$\log\hat{\theta} \sim \mathcal{N}(\log\theta,\ \frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}})\ .$$

Generalized Odds Ratios for $I \times J$ Contingency Tables

II.5.12 Generalzied Odds Ratios

For an $I \times J$ probability table $\boldsymbol{\pi}$, a *minimal set* of $(I-1)(J-1)$ *generalized odds ratios* (GORs) can be defined. Some commonly used GORs are

1. nominal (with reference category (1,1)):
$$\theta_{ij}^{11} = \frac{\pi_{j|i}/\pi_{1|i}}{\pi_{j|1}/\pi_{1|1}} = \frac{\pi_{ij}\pi_{11}}{\pi_{i1}\pi_{1j}} \quad , \quad i = 2,\ldots,I,\ j = 2,\ldots,J,$$
2. local:
$$\theta_{ij}^{L} = \frac{\pi_{j|i}/\pi_{j+1|i}}{\pi_{j|i+1}/\pi_{j+1|i+1}} = \frac{\pi_{ij}\pi_{i+1,j+1}}{\pi_{i,j+1}\pi_{i+1,j}} \quad , \quad i = 1,\ldots,I-1,\ j = 1,\ldots,J-1,$$
3. continuation:
$$\theta_{ij}^{CO} = \frac{\pi_{j|i}/(\sum_{k>j}\pi_{k|i})}{\pi_{j|i+1}/(\sum_{k>j}\pi_{k|i+1})} \quad , \quad i = 1,\ldots,I-1,\ j = 1,\ldots,J-1,$$
4. cumulative:
$$\theta_{ij}^{C} = \frac{(\sum_{k\le j}\pi_{k|i})/(\sum_{k>j}\pi_{k|i})}{(\sum_{k\le j}\pi_{k|i+1})/(\sum_{k>j}\pi_{k|i+1})} \quad , \quad i = 1,\ldots,I-1,\ j = 1,\ldots,J-1,$$
5. global:
$$\theta_{ij}^{G} = \frac{P(Y\le j|X\le i)/P(Y>j|X\le i)}{P(Y\le j|X>i)/P(Y>j|X>i)} \quad , \quad i = 1,\ldots,I-1,\ j = 1,\ldots,J-1.$$

II.5.13 Remark

- Beyond the nominar ORs, also the local ORs can be applied for nominal classification variables.
- The continuation and cumulative ORs require at least one of the classification variables to be ordinal (here Y).
- The global ORs require both classification variables to be ordinal.
- ☞ The interpretation of interaction parameters of log-linear models is based on nominal ORs (and consequently for ORs formed by any two rows $i \neq i'$ and columns $j \neq j'$ of the table).

II.5.14 Parameters' interpretation for the independence model

The parameters' interpretation for model $\mathcal{M}_0$: $\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y$ is simple.

☞ Given we are at the i-th X category, the *logit* of the j category of Y vs the first category 1 is

$$\log\left(\frac{\pi_{ij}}{\pi_{i1}}\right) = \log\left(\frac{\mu_{ij}}{\mu_{i1}}\right) = \log \mu_{ij} - \log \mu_{i1}$$
$$= (\lambda + \lambda_i^X + \lambda_j^Y) - (\lambda + \lambda_i^X - \lambda_1^Y) = \lambda_j^Y - \lambda_1^Y = \lambda_j^Y \ .$$

More general, at the i-th X category and for $j \neq j'$ we have

$$\log(\pi_{ij}/\pi_{ij'}) = \log(\mu_{ij}/\mu_{ij'}) = \lambda_j^Y - \lambda_{j'}^Y$$

☞ Hence, under the independence model, the difference between two λ^Y parameters, corresponding to two categories of Y, equals the logit of the 1st vs the 2nd of them, *independently* from the category i of X.

☞ The interpretation ot the λ^X parameters is similar.

II.5.15 Parameters' interpretation for the saturated model

The *saturated* model $\mathcal{M}_1$ has another term, the interaction term: $\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}$

☞ Only the $(I-1)(J-1)$ of the λ^{XY} parameters are 'free' and are equal to the corresponding nominal ORs:

$$\lambda_{ij}^{XY} = \log(\theta_{ij}^{11}), \; i = 2, \ldots, I, \; j = 2, \ldots, J.$$

The interpretation of the interaction parameters follows by the *log odds ratio* formed by two rows, i, i' and two columns j, j':

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

More specific for a 2×2 table: $\log\left(\frac{\mu_{11}\mu_{22}}{\mu_{12}\mu_{21}}\right) = \lambda_{11}^{XY} + \lambda_{22}^{XY} - \lambda_{12}^{XY} - \lambda_{21}^{XY} = \lambda_{22}^{XY}$.

II.5.16 Remark

Overall, for the interpretation of the parameters of a log-linear model holds:

- Saturated model:
  Parameters λ^{XY} are related to *odds ratios*.
  ☞ Parameters λ^{X} and λ^{Y} are not interpreted.

- Independence model:
  Parameters λ^{X} (λ^{Y}) are related to *odds* for the categories of X (Y).
  ☞ Due to independence, all possible *odds ratios* are equal to 1 (all *log odds ratios* to 0).

- Intercept:
  Parameter λ is related to the sample size n. If the sampling scheme is Poisson, it is related to the parameter of the corresponding Poisson distribution. In any case, it is not informative about the association of X and Y.