Handout 14

Log-Linear Models

In many cases, we have analyzed contingency tables whereby we have distinguished between the response and the independent variables. We have modeled counts as multinomial or independent multinomials or Poisson sampling schemes. We now look at a modeling scheme for contingency tables which incorporates all these different sampling schemes under one umbrella. In the analysis of contingence tables we have often focussed on probabilities. What we discuss now deals with expected frequencies, where the logarithm of expected frequencies are decomposed as in ANOVA with main effects and interactions. These are called log-linear models for contingency tables. This modeling scheme does not require declarations of independent or dependent variables.

Log-linear models can handle all three sampling schemes: multinomial, independent multinomial and Poisson. **However, the actual calculations are done assuming Poisson sampling scheme with counts as the dependent or the response variable.** It turns out that the estimated values of the expected frequencies remain the same irrespective of the sampling schemes, and as a result, LR or Pearson's chi-square statistics for comparing various models, and the corresponding degrees of freedom also remain the same. A short Appendix at the end of this handout presents the basic ideas.

Log-linear Models for Two-way Tables

Consider a two-way table with counts $\{n_{ij}\}$ with expected counts $\{\mu_{ij}\}$. For multinomial sampling, independence implies

$$\mu_{ij} = n\pi_{i+}\pi_{+j}, \text{ ie } \log(\mu_{ij}) = \log(n) + \log(\pi_{i+}) + \log(\pi_{+j}).$$

This is an additive model and can rewritten as

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

with constraints on $\{\lambda_i^X\}$ and $\{\lambda_j^Y\}$.

For the product multinomial model, where the rows are independent multinomial, independence implies

$$\begin{aligned} \lambda_{ij} &= n_{i+}\pi_{+j}, \text{ ie,} \\ \log(\mu_{ij}) &= \log(n_{i+}) + \log(\pi_{+j}) \\ &= \log(n) + \log(n_{i+}/\!/n) + \log(\pi_{+j}). \end{aligned}$$

This model can also be written as an additive model. However, the parameters $\{\lambda_i^X\}$ and $\{\lambda_j^Y\}$ have different interpretations. For this reason, it may not be always useful to try interpret the parameters. However, it is meaningful to compare models of independence to saturated models, or to compare model of conditional independence to the model of homogeneous association, etc.

For a two-way table with mean expected counts $\{\mu_{ij}\}$, the **saturated log-linear model is**

$$\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY},$$

with the constraints on $\{\lambda_i^X\}, \{\lambda_j^Y\}$ and $\{\lambda_{ij}^{XY}\}$. The main effects X, Y, and the interaction effects are $\{\lambda_i^X\}$, $\{\lambda_j^Y\}$ and $\{\lambda_{ij}^{XY}\}$ respectively. The model for independence is the additive model for $\log(\mu_{ij})$, ie, $\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y$.

For a 2×2 table, the odds ratio is given by

$$\begin{aligned}
\theta &= \frac{\mu_{11}\mu_{22}}{\mu_{12}\mu_{21}}, \text{ ie,} \\
\log(\theta) &= \log(\mu_{11}) + \log(\mu_{22}) - \log(\mu_{12}) - \log(\mu_{21}) \\
&= (\lambda + \lambda_1^X + \lambda_1^Y + \lambda_{11}^{XY}) + (\lambda + \lambda_2^X + \lambda_2^Y + \lambda_{22}^{XY}) \\
&\quad -(\lambda + \lambda_1^X + \lambda_2^Y + \lambda_{12}^{XY}) - (\lambda + \lambda_2^X + \lambda_1^Y + \lambda_{21}^{XY}) \\
&= \lambda_{11}^{XY} + \lambda_{22}^{XY} - \lambda_{12}^{XY} - \lambda_{21}^{XY}.
\end{aligned}$$

Under the model of independence, $\lambda_{ij}^{XY} = 0$ for all i and j, and $\log(\theta) = 0$.

Under the saturated model, estimate of μ_{ij} is n_{ij} and the estimate of θ is

$$\hat{\theta} = \frac{n_{11}n_{22}}{n_{12}n_{21}}.$$

For a 2×2 table, the logit of $P(Y = 1|X = i)$ is

$$\begin{aligned}
&\log\left(\frac{P(Y = 1|X = i)}{P(Y = 2|X = i)}\right) \\
= \; &\log\left(\frac{\mu_{i1}}{\mu_{i2}}\right) = \log(\mu_{i1}) - \log(\mu_{i2}) \\
= \; &(\lambda + \lambda_i^X + \lambda_1^Y + \lambda_{i1}^{XY}) - (\lambda + \lambda_i^X + \lambda_1^Y + \lambda_{i2}^{XY}) \\
= \; &(\lambda_1^Y - \lambda_2^Y) + (\lambda_{i1}^{XY} - \lambda_{i2}^{XY}).
\end{aligned}$$

Under the model of independence, ie, $\lambda_{ij}^{XY} = 0$ for all i and j, the logit of $P(Y = 1|X = i)$ equals $\lambda_1^Y - \lambda_2^Y$, which is the same for all i.

Example 1 (Political ideology and opinion on death penalty)

This 2×2 contingency table was presented in Handout 1. The numbers in the brackets are the estimated expected counts ($n_{i+}n_{+j}/n$) under independence.

	Ideology		Total
Supports Death penalty	Liberal	Conservative	
Yes	37 (39.776)	51 (48.224)	88
No	76 (73.224)	86 (88.776)	162
Total	113	137	250

We recast this data in a table that is suitable for fitting a log-linear model. Also given in the brackets are the estimated expected counts under an additive model for $\log(\mu_{ij})$, ie, $\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y$. The R commands are

We fit a log-linear model after having declared Opinion and Ideology as factors. The R commands are

```
>ft=glm(Count ~Opinion + Ideology, family = "poisson")
>fittedCount=ft$fitted
```

The values of the fittedCount are given in the table below.

Note that the fitted values for the additive model coincide with those given in the table above where the estimated expected counts are calculated as $n_{i+}n_{+j}/n$.

Count	Death Penalty	Idelology
37 (39.776)	Yes	Liberal
51 (48.224)	Yes	Conservative
76 (73.224)	No	Liberal
86 (88.776)	No	Conservative

Here is the R output.

Call:

glm(formula = Count ~Opinion + Ideology, family = "poisson")

Deviance Residuals

1	2	3	4
-0.4454	0.3960	0.3224	-0.2962

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	4.48612	0.09732	46.094	$< 2e-16***$
OpinionYes	-0.61026	0.13243	-4.608	$4.06e-06***$
IdeologyLiberal	-0.19259	0.12708	-1.516	0.13

—

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 25.09006 on 3 degrees of freedom

Residual deviance: 0.54689 on 1 degrees of freedom

AIC: 30.238

Number of Fisher Scoring iterations: 3

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

The null and the alternative hypotheses are $H_0 : \log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y$ for all i, j, $H_1 : \log(\mu_{ij}) \neq \lambda + \lambda_i^X + \lambda_j^Y$ for some i, j.

Let L_M and L_S be the likelihoods under the independence model and the saturated models respectively. Then the residual deviance is $G^2 = -2[\log(L_M) - \log(L_S)] = 0.54689$ with df=1. The p-value is area to the right of 0.54689 under χ_1^2 and this area is about 0.460. Clearly, we cannot reject the null. Conclusion: we cannot reject the null hypothesis of independence of political ideology and opinion on death penalty.

Log-linear models for 3-way tables.

When there are three or more categorical variables, the modeling can be more complicated. For three way tables, there are many possibilities such as mutually independence, joint independence, conditional independence.

Mutual Independence

Log-linear model for three discrete variables X, Y and Z are mutually independent (all variables are independent of all others) is given by

$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z.$$

Joint Independence

When X and Y, and Y and Z are independent, but X and Z are dependent, the log-linear model is

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

Similarly, when X and Y, and X and Z are independent, but Y and Z are dependent, the model is

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

Finally, when X and Z, and Y and Z are independent, but X and Y are dependent, we have

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

So for given a data set, one can fit different models and use a criterion such as AIC to decide which one is the most appropriate.

Conditional Independence.

When X and Y are independent given Z, the log-linear model is

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

Similarly, when X and Z are independent given Y, the model is

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

If Y and Z are independent given X, we have

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

Model Without 3-factor Interactions.

Consider the following model without the three factor interactions

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

In a $2 \times 2 \times 2$ table, for this model, odds for (X, Y) given $Z = k$ does not depend on k, ie, this is the case of homogeneous association of (X, Y) given Z. It also turns out that the odds for (X, Z) given $Y = j$ does not depend on j, and the odds for (Y, Z) given $X = i$ does not depend on i.

The Saturated Model.

The most general model is the saturated model and it is given by

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

Notations.

Instead of writing down each model explicitly, we often adopt some simple notations.

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

Example 2 (Berkeley Admissions Data).

We have seen the Berkeley Admissions Data in Handout 13. The data are given below and it is put in yet another table below which makes it easier to use log-linear model.

Accept	Reject	Gender	Department
353	207	M	A
17	8	F	A
120	205	M	B
202	391	F	B
22	351	M	C
24	317	F	C

Count	Admission	Gender	Dept
353	Accept	M	A
207	Reject	M	A
17	Accept	F	A
8	Reject	F	A
120	Accept	M	B
205	Reject	M	B
202	Accept	F	B
391	Reject	F	B
22	Accept	M	C
351	Reject	M	C
24	Accept	F	C
317	Reject	F	C

Call:
glm(formula = Count ~Admit * Gender + Admit * Dept + Gender *
Dept, family = "poisson")

Deviance Residuals:

1	2	3	4	5	6	7	8
−0.07603	0.09965	0.35716	−0.47838	0.36239	−0.27248	−0.27447	0.19922
9	10	11	12				
−0.51321	0.13353	0.52826	−0.13981				

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.74532	0.20754	13.228	$< 2e-16 ***$
AdmitReject	-0.50138	0.14693	-3.412	$0.000644 ***$
GenderM	3.12519	0.20975	14.900	$< 2e-16 ***$
DeptB	2.58220	0.21219	12.169	$< 2e-16 ***$
DeptC	0.32293	0.25430	1.270	0.204137
AdmitReject:GenderM	-0.04334	0.12469	-0.348	0.728136
AdmitReject:DeptB	1.13248	0.13316	8.504	$< 2e-16 ***$
AdmitReject:DeptC	3.19988	0.18275	17.509	$< 2e-16 ***$
GenderM:DeptB	-3.69848	0.21836	-16.938	$< 2e-16 ***$
GenderM:DeptC	-2.99494	0.22860	-13.101	$< 2e-16 ***$

—

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 1535.4802 on 11 degrees of freedom

Residual deviance: 1.2725 on 2 degrees of freedom

AIC: 98.766

Number of Fisher Scoring iterations: 4

Appendix

Two-way Tables

Before we proceed further, let us define a few concepts from two-way ANOVA. For the table of numbers $\{\log(\mu_{ij}) = \lambda_{ij} : 1 \leq i \leq I, 1 \leq j \leq J\}$, we can define

$$\begin{aligned}\lambda &= \sum_i \sum_j \lambda_{ij}/IJ, \bar{\lambda}_{i+} = \sum_j \lambda_{ij}/J, \bar{\lambda}_{+j} = \sum_i \lambda_{ij}/I,\\ \lambda_i^X &= \bar{\lambda}_{i+} - \lambda, \lambda_j^Y = \bar{\lambda}_{+j} - \lambda,\ \lambda_{ij}^{XY} = \lambda_{ij} - \lambda - \lambda_i^X - \lambda_j^Y.\end{aligned}$$

Thus we can write

$$\begin{aligned}\lambda_{ij} &= \lambda + (\bar{\lambda}_{i+} - \lambda) + (\bar{\lambda}_{+j} - \lambda) + (\lambda_{ij} - \bar{\lambda}_{i+} - \bar{\lambda}_{+j} + \lambda)\\ &= \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}. \qquad (1)\end{aligned}$$

If $\lambda_{ij}^{XY} = 0$ for all i and j, then we have

$$\lambda_{ij} = \lambda + \lambda_i^X + \lambda_j^Y,$$

which is called an additive model.

From the definitions of $\{\lambda_i^X\}$, $\{\lambda_j^Y\}$ and $\{\lambda_{ij}^{XY}\}$, it follows that there are a number of constraints

$$\begin{aligned}\sum_i \lambda_i^X &= 0,\ \sum_j \lambda_j^Y = 0,\\ \sum_i \lambda_{ij}^{XY} &= 0 \text{ for each } j, \text{ and } \sum_j \lambda_{ij}^{XY} = 0 \text{ for each } i.\end{aligned}$$

So if we know the values of $\lambda_1^X, \ldots, \lambda_{I-1}^X$, then we automatically know the value of λ_I^X as $-(\lambda_1^X + \cdots + \lambda_{I-1}^X)$. Thus we can say that if we know any $I-1$ of the values of $\{\lambda_i^X\}$, we automatically know the value of the rest. If $\{\lambda_i^X\}$ are unknown parameters, we need to estimate only $I-1$ of them. Similarly, if $\{\lambda_j^Y\}$ are unknown parameters, we need to estimate only $J-1$ of them. Since each row sum and each column sum of $\{\lambda_{ij}^{XY}\}$ equal zero, we need to estimate only $(I-1)(J-1)$ of them.

Alternate representations are also possible. In many packages such as R, the constraints are given as follows:
(a) $\lambda_1^X = 0$ and $\lambda_2^X, \ldots, \lambda_I^X$ are considered parameters to be estimated.
(b) $\lambda_1^Y = 0$, and $\lambda_2^Y, \ldots, \lambda_J^Y$ are considered parameters to be estimated.
(c) $\lambda_{i1}^{XY} = 0$ for all i, $\lambda_{1j}^{XY} = 0$ for all j, and $\lambda_{ij}^{XY}, i = 2, \ldots, I, j = 2, \ldots, J$, are considered parameters to be estimated.

Three-way tables.

For an $I \times J \times K$ table, the saturated model is three-way table is

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

with the same types of constraints on the main effects and two-factor interactions as in the two-way table. For the three factor interactions, the constraints are

$$\begin{aligned} \sum_k \lambda_{ijk}^{XYZ} &= 0 \text{ for each } i,\ j, \\ \sum_j \lambda_{ijk}^{XYZ} &= 0 \text{ for each } i,\ k, \\ \sum_i \lambda_{ijk}^{XYZ} &= 0 \text{ for each } j,\ k. \end{aligned}$$

Thus there are $(I-1)(J-1)(K-1)$ free $\{\lambda_{ijk}^{XYZ}\}$ parameters.

As in two-way tables, alternative representations are used in many computer packages such as R:
(a) $\lambda_1^X = 0$ and $\lambda_2^X, \ldots, \lambda_I^X$ are considered parameters to be estimated.
(b) $\lambda_1^Y = 0$, and $\lambda_2^Y, \ldots, \lambda_J^Y$ are considered parameters to be estimated.
(c) $\lambda_1^Z = 0$, and $\lambda_2^Z, \ldots, \lambda_K^Z$ are considered parameters to be estimated.
(d) $\lambda_{ij}^{XY}, 2 \le i \le I, 2 \le j \le J$, are considered parameters to be estimated, and the rest are set to zero.
(e) $\lambda_{ik}^{XZ}, 2 \le i \le I, 2 \le k \le K$, are considered parameters to be estimated, and the rest are set to zero.
(f) $\lambda_{jk}^{YZ}, 2 \le j \le J, 2 \le k \le K$, are considered parameters to be estimated, and the rest are set to zero.,
(g) $\lambda_{ijk}^{XYZ}, 2 \le i \le I, 2 \le j \le J, 2 \le k \le K$, are considered parameters to be estimated, and the rest are set to zero..