Log-linear Models Part One[1]

STA 312: Fall 2012

[1] See last slide for copyright information.

Background: Re-parameterization

- Data are denoted $D \sim P_\theta, \theta \in \Theta$
- Likelihood function $\ell(\theta, D) = \ell(\theta)$
- Another, equivalent way of writing the parameter may be more convenient.
- Let $\beta = g(\theta), \beta \in \mathcal{B}$
- The function $g : \Theta \to \mathcal{B}$ is one-to-one, meaning $\theta = g^{-1}(\beta)$.
- Re-parameterize, writing the likelihood function in a different form.
- $\ell[\theta] = \ell[g^{-1}(g(\theta))] = \ell[g^{-1}(\beta)] = \ell_2[\beta]$.
- The largest value of $\ell[\theta]$ is the same as the largest value of $\ell_2[\beta]$.
- $\ell[\widehat{\theta}] = \ell_2[\widehat{\beta}]$

Invariance principle of maximum likelihood estimation

$$\widehat{\beta} = g(\widehat{\theta})$$

- Assume $\widehat{\theta}$ is unique, meaning $\ell(\widehat{\theta}) > \ell(\theta)$ for all $\theta \in \Theta$ with $\theta \neq \widehat{\theta}$.
- What if there were a $\beta \neq g(\widehat{\theta})$ in $\mathcal{B}$ with $\ell_2(\beta) \geq \ell_2(g(\widehat{\theta}))$.
- In that case we would have

$$\begin{aligned} & \ell[g^{-1}(\beta)] \geq \ell[g^{-1}(g(\widehat{\theta}))] \\ \Leftrightarrow \quad & \ell(\theta) \geq \ell(\widehat{\theta}) \end{aligned}$$

for some $\theta \neq \widehat{\theta}$. But that's impossible, so there can be no such β. ■

Main point about re-parameterization

- If you have a reasonable model, you can re-write the parameters in any way that's convenient, as long as it's one-to-one with (equivalent to) the original way.
- Maximum likelihood does not care how you express the parameters.
- Log-linear models depend heavily on re-parameterization.

Features of log-linear models

- Used to analyze multi-dimensional contingency tables.
- All variables are categorical.
- No distinction between explanatory and response variables.
- Build a picture of how all the variables are related to each other.
- ANOVA-like models for the logs of the expected frequencies.
- "Response variable" is a vector of log observed frequencies.
- Relationships between variables correspond to interactions in the ANOVA model.

ANOVA-like models

For the logs of the expected frequencies

- Relationships between variables are represented by two-factor interactions.
- Three-factor interactions mean the nature of the relationship *depends* ...
- Etc.

It's like the rotten potatoes example

Course

Passed	Catch-up	Mainstream	Elite
No	π_{11}	π_{12}	π_{13}
Yes	π_{21}	π_{22}	π_{23}

- No relationship means the conditional distribution of Course is the same, regardless of whether the student passed or not.
- Probabilities are proportional:

$$\frac{\pi_{11}}{\pi_{21}} = \frac{\pi_{12}}{\pi_{22}} = \frac{\pi_{13}}{\pi_{23}}$$

- Because $\mu_{ij} = n\pi_{ij}$, same applies to the expected frequencies.

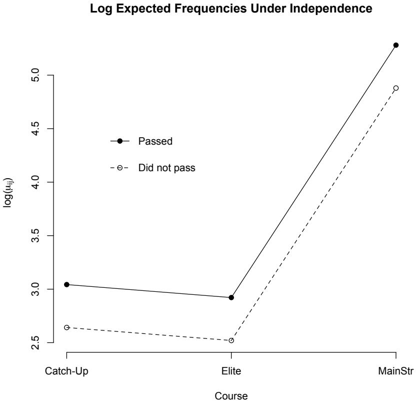
Expected frequencies are proportional

Under H_0 of independence

$$\frac{\mu_{11}}{\mu_{21}} = \frac{\mu_{12}}{\mu_{2}} = \frac{\mu_{13}}{\mu_{23}}$$

$$\Leftrightarrow \quad (\log \mu_{11} - \log \mu_{21}) = (\log \mu_{12} - \log \mu_{22}) = (\log \mu_{13} - \log \mu_{23})$$

So the profiles are parallel in the log scale — no interaction means no relationship.



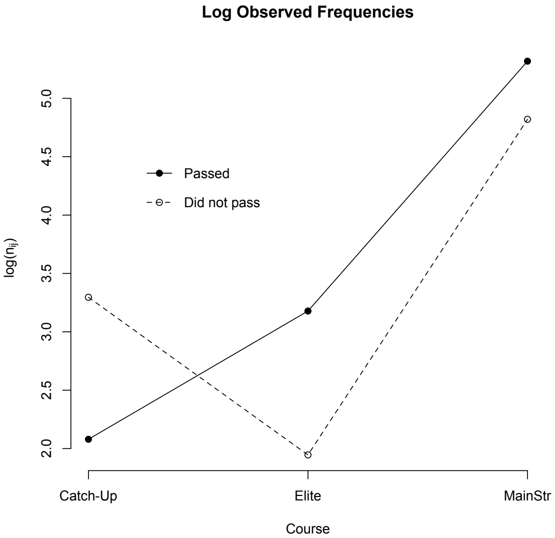
For the record: R code for the last plot

Log expected frequencies

```
# Using mathcat.data
# Get expected frequencies to plot logs
c1 = chisq.test(tab1)
tab0 = c1$expected; tab0
Course = c(1,2,3,1,2,3)
logexpect = log(c(tab0[1,],tab0[2,]))
# Plot
plot(Course,logexpect, pch=' ', frame.plot=F, axes=F,
     xlab="Course", ylab=expression(paste('log(',mu[ij],')') , xaxt='n') )
axis(side=1,labels=c("Catch-Up","Elite","MainStr"),at=1:3)
axis(side=2)
lines(1:3,logexpect[1:3],lty=2) # Did not pass
points(1:3,logexpect[1:3])
lines(1:3,logexpect[4:6],lty=1) # Yes Passed
points(1:3,logexpect[4:6],pch=19)
title("Log Expected Frequencies Under Independence")
legend(1.25,4.5,legend='Passed',lty=1,pch=19,bty='n')
legend(1.25,4.25,legend='Did not pass',lty=2,pch=1,bty='n')
```

Suggests plotting log *observed* frequencies

To see departure from independence



For the record: R code for the last plot

Log observed frequencies

```
# Using mathcat.data
Course = c(1,2,3,1,2,3)
logobs = log(c(tab1[1,],tab1[2,]))
# Plot
plot(Course,logobs, pch=' ', frame.plot=F, axes=F,
     xlab="Course", ylab=expression(paste('log(',n[ij],')')) , xaxt='n') )
axis(side=1,labels=c("Catch-Up","Elite","MainStr"),at=1:3)
axis(side=2)
lines(1:3,logobs[1:3],lty=2) # Did not pass
points(1:3,logobs[1:3])
lines(1:3,logobs[4:6],lty=1) # Yes Passed
points(1:3,logobs[4:6],pch=19)
title("Log Observed Frequencies")
legend(1.25,4.5,legend='Passed',lty=1,pch=19,bty='n')
legend(1.25,4.25,legend='Did not pass',lty=2,pch=1,bty='n')
```

It would be faster to do this in MS Excel.

Regression-like model of independence for the log expected frequencies: No interaction

Use effect coding

$$\log \mu = \beta_0 + \beta_1 p_1 + \beta_2 c_1 + \beta_3 c_2$$

Passed	**Course**	p_1	c_1	c_2	$\log \mu$
No	Catch-up	1	1	0	$\beta_0 + \beta_1 + \beta_2$
No	Elite	1	0	1	$\beta_0 + \beta_1 + \beta_3$
No	Mainstream	1	-1	-1	$\beta_0 + \beta_1 - \beta_2 - \beta_3$
Yes	Catch-up	-1	1	0	$\beta_0 - \beta_1 + \beta_2$
Yes	Elite	-1	0	1	$\beta_0 - \beta_1 + \beta_3$
Yes	Mainstream	-1	-1	-1	$\beta_0 - \beta_1 - \beta_2 - \beta_3$

Notice how this assumes there are no zero probabilities.

Model of independence has main effects only

No interaction terms

$$\log \mu = \beta_0 + \beta_1 p_1 + \beta_2 c_1 + \beta_3 c_2$$

Course

Passed	Catch-up	Elite	Mainstream	Mean
No	$\beta_0 + \beta_1 + \beta_2$	$\beta_0 + \beta_1 + \beta_3$	$\beta_0 + \beta_1 - \beta_2 - \beta_3$	$\beta_0 + \beta_1$
Yes	$\beta_0 - \beta_1 + \beta_2$	$\beta_0 - \beta_1 + \beta_3$	$\beta_0 - \beta_1 - \beta_2 - \beta_3$	$\beta_0 - \beta_1$
Mean	$\beta_0 + \beta_2$	$\beta_0 + \beta_3$	$\beta_0 - \beta_2 - \beta_3$	β_0

- Grand mean is β_0.
- Main effects for Passed are β_1 and $-\beta_1$.
- Main effects for Course are β_2, β_3 and $-\beta_2 - \beta_3$.
- Effects always add up to zero.
- This is an additive model.

$\log \mu_{ij}$ = Grand Mean + Main effect for factor A + Main effect for factor B

Textbook's notation for the additive model

$\log \mu_{ij}$ = Grand Mean + Main effect for factor A + Main effect for factor B

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

Course

Passed	Catch-up	Elite	Mainstream	Mean
No	$\lambda + \lambda_1^X + \lambda_1^Y$	$\lambda + \lambda_1^X + \lambda_2^Y$	$\lambda + \lambda_1^X + \lambda_3^Y$	$\beta_0 + \beta_1$
Yes	$\lambda + \lambda_2^X + \lambda_1^Y$	$\lambda + \lambda_2^X + \lambda_2^Y$	$\lambda + \lambda_2^X + \lambda_3^Y$	$\beta_0 - \beta_1$
Mean	$\beta_0 + \beta_2$	$\beta_0 + \beta_3$	$\beta_0 - \beta_2 - \beta_3$	β_0

There is more than one parameterization. I like this one:

$\lambda = \beta_0$		The grand mean
$\lambda_1^X = \beta_1$		The main effect for $X = 1$
$\lambda_2^X = -\beta_1$		The main effect for $X = 1$
$\lambda_1^Y = \beta_2$		The main effect for $Y = 1$
$\lambda_2^Y = \beta_3$		The main effect for $Y = 2$
$\lambda_3^Y = -\beta_2 - \beta_3$		The main effect for $Y = 3$

Some effects are redundant
Just like in classical ANOVA models

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

where

$$\sum_{i=1}^{I} \lambda_i^X = 0 \text{ and } \sum_{j=1}^{J} \lambda_j^Y = 0$$

Explore the meaning of the parameters

- This is a multinomial model (of independence).
- Set of unique main effects must correspond somehow to the set of unique marginal probabilities.
- But how?
- First, how many parameters are there?

Count the parameters

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

- There are $(I-1)+(J-1)$ unique marginal probabilities.
- There are $(I-1)+(J-1)$ unique main effects.
- Plus the grand mean λ.
- Parameterizations cannot be one-to-one unless number of parameters is the same.
- It turns out that the grand mean is redundant, but not in the way you might think.

You might think that since under independence

$$\begin{aligned}\mu_{ij} &= n\pi_{ij} \\ &= n\pi_{i+}\pi_{+j} \\ \Leftrightarrow \log\mu_{ij} &= \log n + \log\pi_{i+} + \log\pi_{+j} \\ &= \lambda + \lambda_i^X + \lambda_j^Y\end{aligned}$$

- We should have $\lambda = \log n$,
- And $\lambda_i^X = \log\pi_{i+}$
- And $\lambda_j^Y = \log\pi_{+j}$
- But it's not so simple.

Expressing λ in terms of the other parameters

$$
\begin{aligned}
n &= \sum_{i=1}^{I}\sum_{j=1}^{J}\mu_{ij} \\
&= \sum_{i=1}^{I}\sum_{j=1}^{J} e^{\lambda+\lambda_i^X+\lambda_j^Y} \\
&= e^{\lambda}\sum_{i=1}^{I}\sum_{j=1}^{J} e^{\lambda_i^X+\lambda_j^Y} \\
\Leftrightarrow e^{\lambda} &= \frac{n}{\sum_{i=1}^{I}\sum_{j=1}^{J} e^{\lambda_i^X+\lambda_j^Y}} \\
\Leftrightarrow \lambda &= \log\frac{n}{\sum_{i=1}^{I}\sum_{j=1}^{J} e^{\lambda_i^X+\lambda_j^Y}} \neq \log n
\end{aligned}
$$

Connection of main effects to marginal probabilities

- Consider 2×2 case
- Simplify the notation

		Y: 1	Y: 2	
X	1	$\frac{1}{n}e^{\beta_0+\beta_1+\beta_2}$	$\frac{1}{n}e^{\beta_0+\beta_1-\beta_2}$	a
	2	$\frac{1}{n}e^{\beta_0-\beta_1+\beta_2}$	$\frac{1}{n}e^{\beta_0-\beta_1-\beta_2}$	$1-a$
		b	$1-b$	1

$=$

		Y: 1	Y: 2	
X	1	$\frac{e^{\beta_1+\beta_2}}{s}$	$\frac{e^{\beta_1-\beta_2}}{s}$	a
	2	$\frac{e^{-\beta_1+\beta_2}}{s}$	$\frac{e^{-\beta_1-\beta_2}}{s}$	$1-a$
		b	$1-b$	1

where $s = e^{\beta_1+\beta_2} + e^{\beta_1-\beta_2} + e^{-\beta_1+\beta_2} + e^{-\beta_1-\beta_2}$

Four equations in two unknowns

Solve for β_1 and β_2

		Y = 1	Y = 2	
X	1	$\frac{e^{\beta_1+\beta_2}}{s} = ab$	$\frac{e^{\beta_1-\beta_2}}{s} = a(1-b)$	a
	2	$\frac{e^{-\beta_1+\beta_2}}{s} = (1-a)b$	$\frac{e^{-\beta_1-\beta_2}}{s} = (1-a)(1-b)$	$1-a$
		b	$1-b$	1

$$\begin{aligned} \text{Odds}(Y=1|X=1) &= e^{2\beta_2} = \frac{ab}{a(1-b)} = \frac{b}{1-b} \\ \text{Odds}(X=1|Y=1) &= e^{2\beta_1} = \frac{ab}{(1-a)b} = \frac{a}{1-a} \end{aligned}$$

So

$$\begin{aligned} \beta_1 &= \frac{1}{2}\log\frac{a}{1-a} \\ \beta_2 &= \frac{1}{2}\log\frac{b}{1-b} \end{aligned}$$

Regression coefficients (Main Effects)

$$\beta_1 = \frac{1}{2}\log\frac{a}{1-a}$$

$$\beta_2 = \frac{1}{2}\log\frac{b}{1-b}$$

- Are functions of the marginal log odds.
- More generally, they are functions of log odds ratios.
- Notice $\beta_1 = 0 \Leftrightarrow a = 1/2$.
- Zero main effects correspond to equal probabilities, if there are no interactions involving that factor.

What if there are interactions?

$$\log \mu = \beta_0 + \beta_1 p_1 + \beta_2 c_1 + \beta_3 c_2 + \beta_4 p_1 c_1 + \beta_5 p_1 c_2$$

- Five parameters correspond to five probabilities
- A saturated model

Passed	**Course**	p_1	c_1	c_2	$p_1 c_1$	$p_1 c_2$	Interactions only
No	Catch-up	1	1	0	1	0	β_4
No	Elite	1	0	1	0	1	β_5
No	Mainstream	1	-1	-1	-1	-1	$-\beta_4 - \beta_5$
Yes	Catch-up	-1	1	0	-1	0	$-\beta_4$
Yes	Elite	-1	0	1	0	-1	$-\beta_5$
Yes	Mainstream	-1	-1	-1	1	1	$\beta_4 + \beta_5$

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

Interactions are departures from an additive model

	Course			
Passed	Catch-up	Elite	Mainstream	Sum
No	β_4	β_5	$-\beta_4 - \beta_5$	0
Yes	$-\beta_4$	$-\beta_5$	$\beta_4 + \beta_5$	0
Sum	0	0	0	0

- Add to zero down each row and across each column.
- Unique interaction effects are easy to count.
- They correspond to products of dummy variables.
- If non-zero, they make the profiles non-parallel.

Why probabilities and effects (β values) are one-to-one in general

- Since we know n, π_{ij} and μ_{ij} are one-to-one.
- μ_{ij} and $\log \mu_{ij}$ are one-to-one.
- So if we have all the β values, we can solve for the π_{ij}.

Suppose we have all the π_{ij} values. Can we solve for the βs?

- We can get the $\log \mu_{ij}$ values.
- β_0 is the mean of all the $\log \mu_{ij}$.
- Look how easy it is to solve for the main effects.

Course

Passed	Catch-up	Elite	Mainstream	Mean
No	$\log \mu_{11}$	$\log \mu_{12}$	$\log \mu_{13}$	$\beta_0 + \beta_1$
Yes	$\log \mu_{21}$	$\log \mu_{22}$	$\log \mu_{23}$	$\beta_0 - \beta_1$
Mean	$\beta_0 + \beta_2$	$\beta_0 + \beta_3$	$\beta_0 - \beta_2 - \beta_3$	β_0

- Interaction terms are just differences between differences (the difference depends).
- So we can get all the βs.

Extension to higher dimensional tables

- Relationships between variables are represented by two-factor interactions.
- Three-factor interactions mean the nature of the relationship *depends* ...etc.
- This holds provided all lower-order interactions involving the factors are in the model.
- Stick to *hierarchical* models, meaning if an interaction is in the model, then all main effects and lower-order interactions involving those factors are also in the model.

Bracket notation for hierarchical models

- Enclosing two or more factors (variables) in brackets means they interact.
- And all lower-order effects are automatically in the model.
- Suppose there are 4 variables, A, B, C, D
- (AB) (CD) means A is related to B and C is related to D, but A is independent of C and D, and B is independent of C and D.
- The log-linear model includes 4 main effects and 2 interactions.

More examples

- $(A)(B)(C)(D)$ means mutual independence.
- $(AB)(AC)(AD)(BC)(BD)(CD)$ means all two-way relationships are present, but the form of those relationships do not depend on the values of the other variables.
- Sometimes called "homogeneous association."

Given bracket notation, write the model in λ notation

- $(XY)(Z)$

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

- (XYZ)

$$\begin{aligned}\log \mu_{ijk} \quad = \quad & \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z \\ & +\lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} \\ & +\lambda_{ijk}^{XYZ}\end{aligned}$$

Parameter estimation: Iterative proportional model fitting

- Indirect maximum likelihood: Goes straight to estimated expected frequencies, and then estimates all the parameters (unique or not) from there.
- Just specify a list of vectors: Bracket notation.
- Each vector contains a set of indices corresponding to variables
- 1=rows, 2=cols, etc.

Copyright Information