Loglinear Models for Contingency Tables

Jinhwan Suk

Department of Mathematical Science, KAIST

August 23, 2020

Contents

- Consider an $I \times J$ table that classifies n subjects.
- In Poisson form, independence is the loglinear model

$$\log \mu_{ij} = \lambda + \alpha_i + \beta_j \Leftrightarrow \mu_{ij} = \mu\alpha_i\beta_j$$

- Identifiability requires constraints such as

$$\alpha_I = \beta_J = 0$$

- If X and Y are statistically dependent...

$$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY} \tag{1}$$

- Test of independence

$$\lambda_{ij}^{XY} = 0 \to df = (I-1)(J-1)$$

- The number of parameters in model (1) $= IJ \overset{MLE}{\Longrightarrow} \hat{\mu}_{ij} = n_{ij}$
- For 2×2 tables,

$$\begin{aligned}\log \theta &= \log \frac{\mu_{11}\mu_{22}}{\mu_{12}\mu_{21}} \\ &= \lambda_{11}^{XY} + \lambda_{22}^{XY} - \lambda_{12}^{XY} - \lambda_{21}^{XY}\end{aligned}$$

- *Hierarchical model* includes all lower-order terms.
  (e.g) $\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}$
- An example of nonhierarchical model : $\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_{ij}^{XY}$
  $\rightarrow \log \mu_{Ij} = \lambda$ in every column

Contents

- **Mutual independence :** $\pi_{ijk} = \pi_{i++}\pi_{+j+}\pi_{++k}$

  The loglinear model is

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

- **Joint independence :** $\pi_{ijk} = \pi_{i+k}\pi_{+j+}$

  The loglinear model is

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

- **Conditional independence :** $\pi_{ij|k} = \pi_{i+|k}\pi_{+j|k}$

  The loglinear model is

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

- $\lambda_{ij}^{XY}, \lambda_{jk}^{YZ}$, and λ_{ik}^{XZ} pertain to conditionally dependent variables.
- Permitting all three pairs to be conditionally dependent,

$$\log \mu_{ijk} = \log \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ik}^{XZ} \tag{2}$$

Exponentiating both sides,

$$\pi_{ijk} = \psi_{ij}\phi_{jk}\omega_{ik}$$

- For the model (2), each pair has homogeneous association.
- The model (2) is called the loglinear model of *homogeneous association* or of *no three-factor interaction*.

- At a fixed level k of Z, the local odds ratios

$$\theta_{ij(k)} = \frac{\pi_{ijk}\pi_{i+1,j+1,k}}{\pi_{i,j+1,k}\pi_{i+1,j,k}}$$

  describe XY conditional association.

- Using model (2),

$$\log\theta_{ij(k)} = \log\frac{\mu_{ijk}\mu_{i+1,j+1,k}}{\mu_{i,j+1,k}\mu_{i+1,j,k}} = \lambda_{ij}^{XY} + \lambda_{i+1,j+1}^{XY} - \lambda_{i,j+1}^{XY} - \lambda_{i+1,j}^{XY}$$

- An absence of three-factor interaction is equivalent to

$$\theta_{ij(1)} = \theta_{ij(2)} = \cdots = \theta_{ij(k)}$$

- Any model not having three-factor interaction term has a homogeneous association for each pair of variables.

TABLE 8.3 Alcohol, Cigarette, and Marijuana Use for High School Seniors

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

Source: Data courtesy of Harry Khamis, Wright State University.

TABLE 8.4 Fitted Values for Loglinear Models Applied to Table 8.3

Alcohol Use	Cigarette Use	Marijuana Use	Loglinear Model[a] (A, C, M)	(AC, M)	(AM, CM)	(AC, AM, CM)	(ACM)
Yes	Yes	Yes	540.0	611.2	909.24	910.4	911
		No	740.2	837.8	438.84	538.6	538
	No	Yes	282.1	210.9	45.76	44.6	44
		No	386.7	289.1	555.16	455.4	456
No	Yes	Yes	90.6	19.4	4.76	3.6	3
		No	124.2	26.6	142.16	42.4	43
	No	Yes	47.3	118.5	0.24	1.4	2
		No	64.9	162.5	179.84	279.6	279

[a] A, alcohol use; C, cigarette use; M, marijuana use.

TABLE 8.5 Estimated Odds Ratios for Loglinear Models in Table 8.5

	Conditional Association			Marginal Association		
Model	AC	AM	CM	AC	AM	CM
(A, C, M)	1.0	1.0	1.0	1.0	1.0	1.0
(AC, M)	17.7	1.0	1.0	17.7	1.0	1.0
(AM, CM)	1.0	61.9	25.1	2.7	61.9	25.1
(AC, AM, CM)	7.8	19.8	17.3	17.7	61.9	25.1
(ACM) level 1	13.8	24.3	17.5	17.7	61.9	25.1
(ACM) level 2	7.7	13.5	9.7			

Contents

- X^2 and G^2 test
- df $=$ cell counts - the number of parameters

TABLE 8.6 Goodness-of-Fit Tests for Loglinear Models in Table 8.4

Model	G^2	X^2	df	P-value[a]
(A, C, M)	1286.0	1411.4	4	< 0.001
(A, CM)	534.2	505.6	3	< 0.001
(C, AM)	939.6	824.2	3	< 0.001
(M, AC)	843.8	704.9	3	< 0.001
(AC, AM)	497.4	443.8	2	< 0.001
(AC, CM)	92.0	80.8	2	< 0.001
(AM, CM)	187.8	177.6	2	< 0.001
(AC, AM, CM)	0.4	0.4	1	0.54
(ACM)	0.0	0.0	0	—

[a] P-value for G^2 statistic.

- Tests about *conditional associations* compare loglinear models.
- For model (XY, YZ, XZ), consider the hypothesis of XY conditional independence.
$$H_0 : \lambda_{ij}^{XY} = 0$$
- The test statistic is $G^2(XZ, YZ) - G^2(XY, YZ, XZ)$ with $df = (I-1)(J-1)$
- $G^2[(AM, CM)|(AM, CM, AC)] = 187.8 - 0.4 = 187.4$
$df = 2 - 1 = 1 \rightarrow P < 0.001$
- It provides strong evidence of AC conditional association.

- With large sample sizes, statistically significant effects can be weak and practically unimportant.
- **Confidence intervals** are more useful than tests for assessing associations.

TABLE 8.7 Output for Fitting Loglinear Model to Table 8.3

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value / DF
Deviance	1	0.3740	0.3740
Pearson Chi-Square	1	0.4011	0.4011

Parameter			Estimate	Standard Error	Wald Chi-Square	Pr>ChiSq
Intercept			5.6334	0.0597	8903.96	<.0001
a	1		0.4877	0.0758	41.44	<.0001
c	1		−1.8867	0.1627	134.47	<.0001
m	1		−5.3090	0.4752	124.82	<.0001
a*m	1	1	2.9860	0.4647	41.29	<.0001
a*c	1	1	2.0545	0.1741	139.32	<.0001
c*m	1	1	2.8479	0.1638	302.14	<.0001

LR Statistics

Source	DF	Chi-Square	Pr>ChiSq
a*m	1	91.64	<.0001
a*c	1	187.38	<.0001
c*m	1	497.00	<.0001

Contents

- 4 variables W, X, Y and Z
- Interpretations are simplest when the model has *no three-factor interaction terms*, so that each pairwise association is **homogeneous**.
- For model (WXY, WZ, XZ, YZ), each pair is conditionally dependent, but at each level of Z, the WX association, the WY association, the WY association vary.

TABLE 8.8 Loglinear Models for Injury, Seat-Belt Use, Gender, and Location[a]

Gender	Location	Seat Belt	Injury No	Injury Yes	(GI, GL, GS, IL, IS, LS) No	(GI, GL, GS, IL, IS, LS) Yes	(GLS, GI, IL, IS) No	(GLS, GI, IL, IS) Yes	Sample Proportion Yes
Female	Urban	No	7,287	996	7,166.4	993.0	7,273.2	1,009.8	0.12
		Yes	11,587	759	11,748.3	721.3	11,632.6	713.4	0.06
	Rural	No	3,246	973	3,353.8	988.8	3,254.7	964.3	0.23
		Yes	6,134	757	5,985.5	781.9	6,093.5	797.5	0.11
Male	Urban	No	10,381	812	10,471.5	845.1	10,358.9	834.1	0.07
		Yes	10,969	380	10,837.8	387.6	10,959.2	389.8	0.03
	Rural	No	6,123	1,084	6,045.3	1,038.1	6,150.2	1,056.8	0.15
		Yes	6,693	513	6,811.4	518.2	6,697.6	508.4	0.07

[a] G, gender; I, injury; L, location; S, seat-belt use.

Source: Data courtesy of Cristanna Cook, Medical Care Development, Augusta, Maine.

TABLE 8.9 Goodness-of-Fit Tests for Loglinear Models in Table 8.8

Model	G^2	df	P-Value
(G, I, L, S)	2792.8	11	< 0.0001
(GI, GL, GS, IL, IS, LS)	23.4	5	< 0.001
(GIL, GIS, GLS, ILS)	1.3	1	0.25
(GIL, GS, IS, LS)	18.6	4	0.001
(GIS, GL, IL, LS)	22.8	4	< 0.001
(GLS, GI, IL, IS)	7.5	4	0.11
(ILS, GI, GL, GS)	20.6	4	< 0.001

- Mutual independence model (G, I, L, S) fits very poorly
- Model (GI, GL, GS, IL, IS, LS) fits much better but still has a lack of fit.
- Model (GIL, GIS, GLS, ILS) fits well but is complex

Loglinear Models for Higher Dimensions

Models for Four-Way Contingency Tables

TABLE 8.10 Estimated Conditional Odds Ratios for Models of Table 8.8

	Loglinear Model	
Odds Ratio	(GI, GL, GS, IL, IS, LS)	(GLS, GI, IL, IS)
GI	0.58	0.58
IL	2.13	2.13
IS	0.44	0.44
GL S = no	1.23	1.33
S = yes	1.23	1.17
GS L = urban	0.63	0.66
L = rural	0.63	0.58
LS G = female	1.09	1.17
G = male	1.09	1.03

- Model (GLS, GI, IL, IS) seemed to fit much better than (GI, GL, GS, IL, IS, LS).
- The difference in G^2 : $23.4 - 7.5 = 15.9$, $df = 5 - 4 = 1$(P=0.0001)
- However, the degree of three-factor interaction is weak
- With huge samples, it is better to focus on estimation rather than hypothesis testing.

- The *dissimilarity index*

$$\hat{\Delta} = \sum_i |n_i - \hat{\mu}_i|/2n = \sum_i |p_i - \hat{\pi}_i|/2$$

summarizes how far the model fit falls from the data.

- When $\hat{\Delta} < 0.02$ or 0.03, the sample data follow the model pattern quite closely.
- When Δ is near 0, $\hat{\Delta}$ tends to overestimate Δ, substantially so for small n.

Contents

- For three-way tables, the joint Poisson probability is

$$\prod_i \prod_j \prod_k \frac{e^{-\mu_{ijk}} \mu_{ijk}^{n_{ijk}}}{n_{ijk}!}$$

- The kernel of the log likelihood is

$$L(\boldsymbol{\mu}) = \sum_{i,j,k} n_{ijk} \log \mu_{ijk} - \sum_{i,j,k} \mu_{ijk}$$

- For saturated model, this simplifies

$$\begin{aligned} L(\boldsymbol{\mu}) =& n\lambda + \sum_i n_{i++}\lambda_i^X + \sum_j n_{+j+}\lambda_j^Y + \sum_k n_{++k}\lambda_k^Z \\ &+ \sum_{i,j} n_{ij+}\lambda_{ij}^{XY} + \sum_{i,k} n_{i+k}\lambda_{ik}^{XZ} + \sum_{j,k} n_{+jk}\lambda_{jk}^{YZ} \\ &+ \sum_{i,j,k} n_{ijk}\lambda_{ijk}^{XYZ} - \sum_{i,j,k} \exp(\lambda + \cdots + \lambda_{ijk}^{XYZ}) \end{aligned}$$

TABLE 8.12 Minimal Sufficient Statistics for Fitting Loglinear Models

Model	Minimal Sufficient Statistics
(X, Y, Z)	$\{n_{i++}\}, \{n_{+j+}\}, \{n_{++k}\}$
(XY, Z)	$\{n_{ij+}\}, \{n_{++k}\}$
(XY, YZ)	$\{n_{ij+}\}, \{n_{+jk}\}$
(XY, XZ, YZ)	$\{n_{ij+}\}, \{n_{i+k}\}, \{n_{+jk}\}$

- Loglinear models have the form

$$\log \boldsymbol{\mu} = \boldsymbol{X}\beta$$

(e.g) consider the independence model, $\log \mu_{ij} = \lambda + \lambda_i^x + \lambda_j^Y$

$$\begin{bmatrix} \log \mu_{11} \\ \log \mu_{12} \\ \log \mu_{21} \\ \log \mu_{22} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ \lambda_1^X \\ \lambda_1^Y \end{bmatrix}$$

- We have $\log \mu_i = \sum_j x_{ij}\beta_j$ for all i.

- The log likelihood is

$$
\begin{aligned}
L(\boldsymbol{\mu}) &= \sum_i n_i \log \mu_i - \sum_i \mu_i \\
&= \sum_i n_i \left(\sum_j x_{ij} \beta_j \right) - \sum_i \exp \left(\sum_j x_{ij} \beta_j \right)
\end{aligned}
$$

- $\frac{\partial L(\boldsymbol{\mu})}{\partial \beta_j} = \sum_i n_i x_{ij} - \sum_i \mu_i x_{ij}$
- The likelihood equation is

$$
\boldsymbol{X}^t \boldsymbol{n} = \boldsymbol{X}^t \hat{\mu}
$$

Loglinear Model Fitting

Birch, Maximum Likelihood in Three-way Contingency Tables (1963)

- Equivalence of MLE's for
  1. Poisson Sampling
  2. Multinomial sampling
  3. Product-multinomial sampling
- The likelihood equations have a unique solution
- MLE's are the same under a wide variety of sampling conditions
- The estimates are sufficient statistics
- Generalization to multi-way contingency tables

Experiment 1. We have N independent Poisson distributions with expected frequencies $\mu_i(\boldsymbol{\theta})$, $i = 1, ..., N$ and corresponding observed frequencies n_i, $i = 1, ..., N$.

Experiment 2. We have $\rho - 1$ independent multinomial distributions. For the $\alpha-$th, we have taken a sample of fixed size

$$N_\alpha = \sum_{i=R(\alpha-1)+1}^{R(\alpha)} n_i$$

and we have observed numbers $n_{R(\alpha-1)+1}, \ldots, n_{R(\alpha)}$ respectively in the 1st, ... , $r(\alpha)$th categories, where

$$R(\alpha) = \begin{cases} \sum_{\beta=1}^{\alpha} r(\beta), & \text{for } \alpha \geq 1 \\ 0, & \text{for } \alpha = 0 \end{cases}$$

The corresponding expected frequencies are $\mu_{R(\alpha-1)+1}(\boldsymbol{\theta}), \ldots, \mu_{R(\alpha)}(\boldsymbol{\theta})$. We also have $r(\rho)$ Poisson distributions.

Theorem

Let us denote the MLE of $\boldsymbol{\theta}$ in experiment 1 by $\hat{\boldsymbol{\theta}}$ and the MLE of $\boldsymbol{\theta}$ in experiment 2 by $\hat{\boldsymbol{\theta}}^*$. Then, provided

$$\sum_{i=R(\alpha-1)+1}^{R(\alpha)} \hat{\mu}_i(\boldsymbol{\theta}) = N_\alpha,$$

$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}^*$ *and* $\hat{\mu}_i = \hat{\mu}_i{}^*$.

Theorem

Suppose now we have unrestricted sampling conditions.

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

Then the following statements hold.

1. *The marginal totals* $n_{ij+}, n_{i+k}, n_{+jk}$ *are sufficient statistics*
2. *The marginal totals are equal to the MLE of their expectations, i.e.* $n_{ij+} = \hat{m}_{ij+}$
3. *There is one set of* λ*'s for which the likelihood function is stationary.*
4. *The MLEs are determined uniquely by the marginal totals being equal to the MLE of their expectations.*

- From general expression $\log \mu_i = \sum_j x_{ij}\beta_j$, we obtained

$$\frac{\partial L(\boldsymbol{\mu})}{\partial \beta_j} = \sum_i n_i x_{ij} - \sum_i \mu_i x_{ij}$$

- The Hessian matrix has elements

$$\begin{aligned}
\frac{\partial^2 L(\boldsymbol{\mu})}{\partial \beta_j \partial \beta_k} &= -\sum_i x_{ij} \frac{\partial \mu_i}{\partial \beta_k} \\
&= -\sum_i x_{ij} \left[\frac{\partial}{\partial \beta_k} \left(\exp \left(\sum_h x_{ij}\beta_h \right) \right) \right] \\
&= -\sum_i x_{ij} x_{ik} \mu_i
\end{aligned}$$

- The information matrix is

$$\mathcal{J} = \boldsymbol{X}^{\boldsymbol{t}}\mathrm{Diag}(\boldsymbol{\mu})\boldsymbol{X}$$

- The asymptotic covariance matrix is

$$cov(\hat{\beta}) = \left[\boldsymbol{X}^{\boldsymbol{t}}\mathrm{Diag}(\boldsymbol{\mu})\boldsymbol{X}\right]^{-1}$$

- From model $\log \boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$,

$$L(\boldsymbol{\beta}) = \sum_i n_i \left(\sum_h x_{ij}\beta_h \right) - \sum_i \exp \left(\sum_h x_{ij}\beta_h \right)$$

Then

$$\nabla_j L(\boldsymbol{\beta}) = \sum_i n_i x_{ij} - \sum_i \mu_i x_{ij}$$
$$\nabla^2_{jk} L(\boldsymbol{\beta}) = -\sum_i \mu_i x_{ij} x_{ik}$$

- It generates the next value $\boldsymbol{\beta}^{(t+1)}$ by $\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} - (\boldsymbol{H}^{(t)})^{-1}\boldsymbol{u}^{(t)}$

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} + \left[\boldsymbol{X}^t \mathrm{Diag}(\boldsymbol{\mu})\boldsymbol{X}\right]^{-1} \boldsymbol{X}^t(\boldsymbol{n} - \boldsymbol{\mu}^{(t)})$$

- Alternatively, $\boldsymbol{\beta}^{(t+1)}$ can be expressed as

$$\boldsymbol{\beta}^{(t+1)} = -(\boldsymbol{H}^{(t)})^{-1}\boldsymbol{r}^{(t)}$$

  where $r_j^{(t)} = \sum_i \mu_i^{(t)} x_{ij}[\log \mu_i^{(t)} + (n_i - \mu_i^{(t)})/\mu_i^{(t)}]$

- The iterative process begins with $\mu_i^{(0)} = n_i$ or with an adjustment such as $\mu_i^{(0)} = n_i + \frac{1}{2}$

Steps :

1. Start with $\boldsymbol{\mu}^{(0)}$ satisfying a model no more complex than the one being fitted.

   (e.g) $\{\mu_i^{(t)} = 1\}$

2. By multiplying by appropriate factors, adjust $\mu_i^{(0)}$ successively to match each marginal table in the set of minimal sufficient statistics.

3. Continue until the maximum difference between the sufficient statistics and their fitted values is sufficiently close to zero.

Example :

- Consider the model (XY, XZ, YZ). Its minimal sufficient statistics are $\{n_{ij+}\}$, $\{n_{i+k}\}$, and $\{n_{+jk}\}$.
- IPF algorithm has three steps

$$\mu_{ijk}^{(t+1)} = \mu_{ijk}^{(t)} \frac{n_{ij+}}{\mu_{ij+}^{(t)}}, \quad \mu_{ijk}^{(t+2)} = \mu_{ijk}^{(t+1)} \frac{n_{i+k}}{\mu_{i+k}^{(t+1)}}, \quad \mu_{ijk}^{(t+3)} = \mu_{ijk}^{(t+2)} \frac{n_{+jk}}{\mu_{+jk}^{(t+2)}}$$

As the cycle progress, G^2 statistic is monotone decreasing.

The IPF algorithm produces ML estimates.