

Part II: Generalized Linear Models

Chapter II.5

Log-Linear Models

Log-Linear Models for Contingency Tables of Higher Dimension, Model Selection,
Connection between Log-Linear and Logit Models

Log-Linear Models for Contingency Tables of Higher Dimension

II.5.17 Log-linear models for k -way contingency tables

Consider the cross-classification of $X_i \in \{1, \dots, I_i\}$ to a k -way contingency table of observed counts $\mathbf{n} = (n_{i_1 \dots i_k})$, a realization of $\mathbf{N} = (N_{i_1 \dots i_k}) \sim \mathcal{M}(n, \boldsymbol{\pi})$. Log-linear models can be defined for k -way, analogously to two-way tables (s. Definitions II.5.6 and II.5.7), ranging from the model of complete independence to the saturated model. The saturated model is the model consisting of all main effects and all possible interactions among the classification variables, from two-way up to k -way interactions.

For example, for $k = 5$, $\ell = 1, \dots, 5$, and $i_\ell = 1, \dots, I_\ell$, possible non-saturated models are

$$\textcircled{1} \log(\mu_{i_1 \dots i_5}) = \lambda + \sum_{\ell=1}^5 \lambda_{i_\ell}^{X_\ell} + \lambda_{i_1 i_2}^{X_1 X_2} + \lambda_{i_3 i_5}^{X_3 X_5},$$

$$\textcircled{2} \log(\mu_{i_1 \dots i_5}) = \lambda + \sum_{\ell=1}^5 \lambda_{i_\ell}^{X_\ell} + \lambda_{i_1 i_2}^{X_1 X_2} + \lambda_{i_1 i_3}^{X_1 X_3} + \lambda_{i_2 i_3}^{X_2 X_3} + \lambda_{i_1 i_2 i_3}^{X_1 X_2 X_3},$$

$$\textcircled{3} \log(\mu_{i_1 \dots i_5}) = \lambda + \sum_{\ell=1}^5 \lambda_{i_\ell}^{X_\ell} + \lambda_{i_1 i_2}^{X_1 X_2} + \lambda_{i_1 i_2 i_3}^{X_1 X_2 X_3}.$$

► II.5.18 Interpretation of the log-linear model parameters

For higher dimensional contingency tables, the interpretation of the log-linear model parameters is analogue to the two-dimensional tables case and is based on the *conditional* odds ratios. For example under model (1) in Defintion II.5.17, the conditional OR comparing levels $i_3 \neq i'_3$ of X_3 for $X_5 = i_5$ vs $X_5 = i'_5$ ($i_5 \neq i'_5$), given that X_1, X_2 and X_4 are at fixed level i_1, i_2 and i_4 , respectively, is equal to

$$\log(\theta_{i_3 i_5 (i_1 i_2 i_4)}^{i'_3 i'_5}) = \log \left(\frac{\pi_{i_1 i_2 i_3 i_4 i_5} \pi_{i_1 i_2 i'_3 i_4 i'_5}}{\pi_{i_1 i_2 i'_3 i_4 i_5} \pi_{i_1 i_2 i_3 i_4 i'_5}} \right) = \lambda_{i_3 i_5}^{X_3 X_5} + \lambda_{i'_3 i'_5}^{X_3 X_5} - \lambda_{i'_3 i_5}^{X_3 X_5} - \lambda_{i_3 i'_5}^{X_3 X_5} .$$



See also Example II.5.22 below.

Log-Linear Models for Three-Way Tables

II.5.19 Non-Saturated Hierarchical Log-Linear Models for Three-Way Tables

Consider the $I \times J \times K$ table, produced by cross-classifying the categorical variables X , Y and Z , of levels I , J and K , respectively.

Model of complete independence

$$\log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z, \quad i = 1, \dots, I, \quad j = 1, \dots, J, \quad k = 1, \dots, K,$$

Identifiability constraints: $\lambda_1^X = \lambda_1^Y = \lambda_1^Z = 0$

Model of joint independence of Y from X and Z

$$\log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ}, \quad \forall i, j, k.$$

Additional identifiability constraints: $\lambda_{1k}^{XZ} = \lambda_{i1}^{XZ} = 0$, for all possible values of i and k .

Model of conditional independence of X and Y , given Z

$$\log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}, \quad \forall i, j, k.$$

Additional identifiability constraints: $\lambda_{1k}^{YZ} = \lambda_{j1}^{YZ} = 0$, for all possible values of j and k .

Model of no three-factor interaction

$$\log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}, \quad \forall i, j, k.$$

Additional identifiability constraints: $\lambda_{1j}^{XY} = \lambda_{i1}^{XY} = 0$, for all possible values of i and j .

II.5.20 Hierarchical Models

The log-linear models considered in II.5.19 above share the following property. In all of them, whenever a higher order effect is in the model, then all possible lower order effects involving the variables of this higher order effect term are also in the model. Such models are called *hierarchical log-linear models* and are parsimoniously symbolized by the set of the highest order terms (with respect to all variables) that define them uniquely.

II.5.21 Example (hierarchical models)

- For instance, for two-way tables, the independence and the saturated model are symbolized by (X, Y) and (XY) , respectively. The model $\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_{ij}^{XY}$ is non-hierarchical, since it includes the term λ_{ij}^{XY} , without having the term λ_j^Y .
- Analogously, in three-way tables, the model of *joint independence* of Y from X and Z is denoted as (Y, XZ) and model $\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ}$ is non-hierarchical, due to the absence of the term λ_k^Z .
- In II.5.17 models (1) and (2) are hierarchical while (3) is not (it includes the term $\lambda_{i_1 i_2 i_3}^{X_1 X_2 X_3}$ without including $\lambda_{i_1 i_3}^{X_1 X_3}$ and $\lambda_{i_2 i_3}^{X_2 X_3}$).

Model Selection in Multi-Way Tables

As the dimension of the table increases, the number of possible models increases as well and a model selection procedure has to be adopted (s. Model Selection in GLMs).

Model selection will be demonstrated here in terms of an example.

▶ II.5.22 Example

Smoking habit vs major depressive disorder by gender for a sample collected in the St. Louis Epidemiologic Catchment Area Survey (Glassman et al., 1990).

Males			Females		
<i>Ever smoked</i>	<i>Major Depression</i>		<i>Ever smoked</i>	<i>Major Depression</i>	
	yes	no		yes	no
	yes	40 889	yes	104 840	
no	10 417		no	40 873	

→ Is the “Depression–Smoking” association significant? Is it the same for males and females?



The data are organized in a data frame as follows.

```
> freq <- c(40, 10, 889, 417, 104, 40, 840, 873)
> row <- rep(1:2, 4); col <- rep(1:2, each=2,2)
> lay <- rep(1:2, each=4); row.lb <- c(`yes`,`no`)
> col.lb <- c(`yes`,`no`); lay.lb <- c(`male`,`female`)
> S <- factor(row,labels=row.lb); D <- factor(col,labels=col.lb)
> G <- factor(lay, labels=lay.lb)
> depres.fr <- data.frame(freq,S,D,G)
```

The appropriate log-linear model is selected via the **backward stepwise** procedure based on *AIC*. Thus, we first save the saturated model under object `saturated` and then proceed with the backward model selection procedure, as follows

```
> saturated <- glm(freq ~ S*D*G, poisson, data = depres.fr)
> step(saturated, direction=`backward`)
```

The stepwise procedure concludes to the model of no three-factor interaction (SD , DG , SG), giving the following output.



```
Start: AIC=71.38
freq  S * D * G:
      Df      Deviance   AIC
- S:D:G      1      0.77135  70.155
<none>              0.00000  71.384

Step: AIC=70.16
freq ~ S + D + G + S:D + S:G + D:G
      Df      Deviance   AIC
<none>              0.771   70.155
- S:D      1      33.024  100.408
- D:G      1      34.386  101.769
- S:G      1     112.298  179.682

Call: glm(formula = freq~S+D+G+S:D+S:G+D:G, family=poisson,
data=depres.fr)
Coefficients:
Intercept      Sno          Dno    Gfemale
3.7393      -1.6684      3.0485    0.8850
Sno:Dno      Sno:Gfemale  Dno:Gfemale
0.9187      0.7834      -0.9369

Degrees of Freedom: 7 Total (i.e. Null); 1 Residual
Null Deviance: 3315
Residual Deviance: 0.7713  AIC: 70.16
```




The (SD, DG, SG) is the model of *homogeneous association*, since under this model the association in all two-way partial tables is homogeneous across the levels of the remaining third classification variable.

This model is fitted in R by

```
> hom.assoc <- glm(freq~S*D+S*G+D*G, poisson,data=depres.fr)
```

From the summary output (provided next), we read $\hat{\lambda}_{22}^{SD} = 0.91871$.

Under (SD, DG, SG) , λ_{22}^{SD} is equal to

$$\lambda_{22}^{SD} = \log \left(\frac{\mu_{11k}/\mu_{12k}}{\mu_{21k}/\mu_{22k}} \right), \quad k = 1, 2.$$

$\frac{\mu_{11k}/\mu_{12k}}{\mu_{21k}/\mu_{22k}}$ is the **conditional odds ratio**, denoted as θ_k^{SD} and is a measure of “Smoking–Depression” association, conditional on gender.

Under the homogeneous association model, $\theta_k^{SD} = \theta^{SD}$, for all k .

The MLE of the common θ^{SD} is

$$\hat{\theta}^{SD} = \exp \left(\hat{\lambda}_{22}^{SD} \right) = \exp(0.91871) = 2.506,$$

indicating that the odds of depression is 2.5 higher for smokers than for non-smokers, both for males and females.



```
> summary(hom.assoc)
```

```
Call:
glm(formula = freq~S*D+S*G+D*G, family=poisson, data=depres.fr)
Deviance Residuals:
    1      2      3      4      5      6      7      8
-0.32157  0.70555  0.06943 -0.10112  0.20418 -0.32157 -0.07131
0.07006
Coefficients:
              Estimate      Std. Error  z value    Pr(>|z|)
(Intercept)   3.73930      0.14417    25.936    < 2e-16 ***
SNo           -1.66844      0.17668    -9.443    < 2e-16 ***
DNo            3.04847      0.14705    20.731    < 2e-16 ***
Gfemale        0.88501      0.16620     5.325    1.01e-07 ***
SNo:DNo        0.91871      0.17059     5.385    7.23e-08 ***
SNo:Gfemale    0.78344      0.07529    10.405    < 2e-16 ***
DNo:Gfemale   -0.93691      0.17055    -5.493    3.94e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for poisson family taken to be 1)
Null deviance: 3315.40325 on 7 degrees of freedom
Residual deviance: 0.77135 on 1 degrees of freedom
AIC: 70.155
Number of Fisher Scoring iterations: 4
```

► II.5.23 Example

Asume that for $X = 1$, 50 observations are sampled for each Z level, while when $X = 2$ 60 observations are sampled for each Z level. For each sampled case we observe the response Y . This means that it is a priori known that $n_{1+k} = 50$ and $n_{2+k} = 60$ for $k = 1, 2$.

Z	X	Y		Total
		1	2	
1	1	35	15	50
	2	40	20	60
2	1	30	20	50
	2	45	15	60

Ignoring the sampling design, the simplest model that fits well the data is the $(Y) : G^2 = 4.79$ (df = 6, P -value = 0.57).



Data				
<i>Z</i>	<i>X</i>	<i>Y</i>		Total
		1	2	
1	1	35	15	50
	2	40	20	60
2	1	30	20	50
	2	45	15	60

Estimates under (<i>Y</i>)				
<i>Z</i>	<i>X</i>	<i>Y</i>		Total
		1	2	
1	1	37.5	17.5	55
	2	37.5	17.5	55
2	1	37.5	17.5	55
	2	37.5	17.5	55

The estimated expected under this model marginals are different than the observed ones. Since this is not possible to hold, the terms λ^X , λ^Z and λ^{XZ} must be included in the model.



Data				
Z	X	Y		Total
		1	2	
1	1	35	15	50
	2	40	20	60
2	1	30	20	50
	2	45	15	60

Estimates under (Y, XZ)				
Z	X	Y		Total
		1	2	
1	1	34.1	15.9	50
	2	40.9	19.1	60
2	1	34.1	15.9	50
	2	40.9	19.1	60

This is the appropriate treatment for this case.

➤ II.5.24 Remark

If the marginal frequencies of some partial tables are a priori fixed *by the design* then the corresponding interaction terms should remain in the model, even if non-significant.

For example, in case of a $2 \times 2 \times 2$ contingency table with classification variables X , Y , Z , if the study is designed so that a certain number of observations is sampled for every combination of the levels of X and Z , then the log-linear model fitted should definitely include the λ^{XZ} term.

Connection between Log-Linear and Logit Models

Consider a GLM model with a **binomial** random response Y , a categorical explanatory variable X and the **logit** link (s. Definition II.4.11 and Remark II.3.4):

$$\text{logit}(\pi_{1|i}) = \log \frac{\pi_{1|i}}{1 - \pi_{1|i}} = \beta_0 + \beta_i, \quad i = 1, \dots, I,$$

with $\beta_1 = 0$ for identifiability reasons, which is equivalently expressed in terms of success probabilities (s. II.3.7) as

$$\pi_{1|i} = \frac{\exp(\beta_0 + \beta_i)}{1 + \exp(\beta_0 + \beta_i)}, \quad i = 1, \dots, I.$$

- The hypothesis of independence is equivalent to

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_I = 0.$$

➤ II.5.25 Connection between log-linear and binary logit models

Considering further the saturated log-linear model for the $I \times 2$ contingency table, we derive the standard expression for a logit model:

$$\text{logit}(\pi_{1|i}) = \log \pi_{i1} - \log \pi_{i2} = \underbrace{(\lambda_1^Y - \lambda_2^Y)}_{\beta_0} + \underbrace{(\lambda_{i1}^{XY} - \lambda_{i2}^{XY})}_{\beta_i}, \quad i = 1, \dots, I.$$


➤ 11.5.26 Log-linear vs. logit models for higher dimensional contingency tables

For two-way tables there exists a one-to-one correspondence between the logit models and the class of hierarchical log-linear models. This is not the case for contingency tables of higher order!

For example, for an $I_1 \times I_2 \times 2$ table with two explanatory variables, X_1 and X_2 , and a binary response Y , the logit model is

$$\text{logit}(\pi_{1|i_1 i_2}) = \beta_0 + \beta_{i_1}^{X_1} + \beta_{i_2}^{X_2}, \quad i_\ell = 1, \dots, I_\ell, \quad \ell = 1, 2,$$

which corresponds to the log-linear model $(X_1 X_2, X_1 Y, X_2 Y)$, as can be easily verified (analogous to 11.5.25, with $\text{logit}(\pi_{1|i_1 i_2}) = \log(\pi_{i_1 i_2 1}) - \log(\pi_{i_1 i_2 2})$).

 In logit framework, focus lies on the dependence of the response on the explanatory variable while *the association structure among the explanatory variables is not of interest and their interactions are considered in the greatest possible level.*

Thus, for the three-way table considered above, model $(X_1 Y, X_2 Y)$ has not a logit analogue.