STA4030: Categorical Data Analysis
Loglinear Model Fitting: Likelihood Equations

Instructor: Bojun Lu

School of Data Science
CUHK(SZ)

December 8, 2020

Agenda

10.6 Loglinear Model Fitting: Likelihood Equations

- We next discuss loglinear model fitting. Results with models for three-way tables are presented.
- Fir simplicity, derivations use the Poisson sampling model, compared with multinomial sampling model, Poisson sampling model does not require constraints on $\{\mu_{ijk}\}$.
- For three-way tables, the joint Poisson distribution that cell counts $\{Y_{ijk} = n_{ijk}\}$ is

$$L(\boldsymbol{\mu}) = \prod_i \prod_j \prod_k \frac{e^{-\mu_{ijk}} \mu_{ijk}^{n_{ijk}}}{n_{ijk}!}, \tag{1}$$

  with product taken over all cells of the table.

10.6 Loglinear Model Fitting: Likelihood Equations

- The kernel of the log likelihood is

$$l(\boldsymbol{\mu}) = \sum_i \sum_j \sum_k n_{ijk} \log \mu_{ijk} - \sum_i \sum_j \sum_k \mu_{ijk}. \tag{2}$$

- Recall the general loglinear model (which is also the saturated model),

$$\log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}. \tag{3}$$

- For the saturated model (3), equation (2) is reduced as,

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

10.6 Loglinear Model Fitting: Likelihood Equations

- For simpler loglinear models, certain parameters are zero and the relative log likelihood $l(\boldsymbol{\mu})$ in terms of $\lambda, \ldots, \lambda_{ijk}^{XYZ}$ simplifies as well.
- The fitted values for a model are solutions to the likelihood equations.
- Next we try to derive likelihood equations in terms of a general formula for a loglinear model.
- For the N cells of a contingency table, we introduce notations,

$$\boldsymbol{n} = (n_1, \ldots, n_N)^T, \quad n = \sum_{i=1}^{N} n_i,$$

and

$$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_N)^T,$$

which denotes the column vectors of observed and expected counts for the N cells.

10.6 Loglinear Model Fitting: Likelihood Equations

- We can also represent loglinear models in the following matrix form,

$$\log \boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}, \tag{4}$$

  where $\boldsymbol{X}$ denotes a model matrix and $\boldsymbol{\beta}$ denotes model parameters.

- For example, for a 2×2 table, consider the independence model,

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

  with constraints $\lambda_2^X = \lambda_2^Y = 0$. We have that,

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

10.6 Loglinear Model Fitting: Likelihood Equations

- For the model $\log \boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$, we have that,

$$\log \mu_i = \sum_j x_{ij}\beta_j, \quad \forall i = 1, 2, \ldots, N.$$

- Then equation (2) can be rewritten as,

$$\begin{aligned} l(\boldsymbol{\mu}) &= \sum_i n_i \log \mu_i - \sum_i \mu_i \\ &= \sum_i n_i \left(\sum_j x_{ij}\beta_j \right) - \sum_i \exp\left(\sum_j x_{ij}\beta_j \right). \end{aligned} \tag{5}$$

- Next we derive the likelihood equations,

$$\frac{\partial}{\partial \beta_j} l(\boldsymbol{\mu}) = 0, \quad j = 1, 2, \ldots, p.$$

10.6 Loglinear Model Fitting: Likelihood Equations

- Calculate the following partial derivatives,

$$\frac{\partial}{\partial \beta_j}\left[\exp\left(\sum_j x_{ij}\beta_j\right)\right] = x_{ij}\exp\left(\sum_j x_{ij}\beta_j\right) = x_{ij}\mu_i,$$

and

$$\frac{\partial}{\partial \beta_j} l(\boldsymbol{\mu}) = \sum_i n_i x_{ij} - \sum_i x_{ij}\mu_i, \quad j = 1, 2, \ldots, p.$$

- Thus can derive the likelihood equations in the following form,

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

- Note that based on GLM theory, the sufficient statistic for β_j is its coefficient $\sum_i n_i x_{ij}$, thus the likelihood equations also equate the sufficient statistics to their expected values.

10.7 Direct Calculation of Fitted Values

- To illustrate how to solve likelihood equations, we continue the analysis of model (XZ, YZ),

$$\log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}. \tag{6}$$

- Actually, Model (6) assumes that categorical variables X and Y are conditionally independent, given Z. That is independence holds for each partial table within which Z is fixed,

$$\pi_{ij(k)} = \pi_{i+(k)}\pi_{+j(k)}, \quad \forall i, j, k.$$

or for joint probabilities, equivalently,

$$\pi_{ijk} = \pi_{i+k}\pi_{+jk}/\pi_{++k}, \quad \forall i, j, k.$$

10.7 Direct Calculation of Fitted Values

- Recall the saturated model (3) and the relative log-likelihood,

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

- For Model (6) with $\lambda^{XY} = \lambda^{XYZ} = 0$, the log-likelihood is simplified.

10.7 Direct Calculation of Fitted Values

- The derivatives become,

$$\frac{\partial}{\partial \lambda_{ik}^{XZ}} l(\boldsymbol{\mu}) = n_{i+k} - \mu_{i+k},$$

  and

$$\frac{\partial}{\partial \lambda_{jk}^{YZ}} l(\boldsymbol{\mu}) = n_{+jk} - \mu_{+jk}.$$

- Then,

$$\hat{\mu}_{i+k} = n_{i+k}, \quad \forall i, k,$$

  and

$$\hat{\mu}_{+jk} = n_{+jk}, \quad \forall j, k,$$

$$\hat{\mu}_{++k} = n_{++k}, \quad \forall k.$$

10.7 Direct Calculation of Fitted Values

- Setting $\pi_{ijk} = \mu_{ijk}/n$, then

$$\pi_{ijk} = \pi_{i+k}\pi_{+jk}/\pi_{++k}, \quad \forall i, j, k,$$

can be reduced to

$$\mu_{ijk} = \mu_{i+k}\mu_{+jk}/\mu_{++k}, \quad \forall i, j, k.$$

- Then we have that,

$$\begin{aligned} \hat{\mu}_{ijk} &= \frac{\hat{\mu}_{i+k}\hat{\mu}_{+jk}}{\hat{\mu}_{++k}} \\ &= \frac{n_{i+k}n_{+jk}}{n_{++k}}, \quad \forall i, j, k. \end{aligned}$$

- Note that ML estimates of functions of parameters are the same functions of the ML estimates of those parameters.

10.7 Direct Calculation of Fitted Values

- For models having explicit formulas for $\hat{\mu}_{ijk}$, the estimates are said to be direct.
- Many loglinear models do not have direct estimates. ML estimation then requires iterative methods.

Table 1: Fitted Values for Loglinear Models in Three-Way Tables

Model	Probabilistic Form	Fitted Value
(X, Y, Z)	$\pi_{ijk} = \pi_{i++}\pi_{+j+}\pi_{++k}$	$\hat{\mu}_{ijk} = \frac{n_{i++}n_{+j+}n_{++k}}{n^2}$
(XY, Z)	$\pi_{ijk} = \pi_{ij+}\pi_{++k}$	$\hat{\mu}_{ijk} = \frac{n_{ij+}n_{++k}}{n}$
(XY, XZ)	$\pi_{ijk} = \frac{\pi_{ij+}\pi_{i+k}}{\pi_{i++}}$	$\hat{\mu}_{ijk} = \frac{n_{ij+}n_{i+k}}{n_{i++}}$
(XY, XZ, YZ)	$\pi_{ijk} = \psi_{ij}\phi_{jk}\omega_{ik}$	Iterative methods
(XYZ)	No restriction	$\hat{\mu}_{ijk} = n_{ijk}$

10.8 Loglinear Model Fitting: Iterative Methods

- When a loglinear model does not have direct estimates, iterative algorithm such as Newton-Raphson can solve the likelihood equations.
- For the Newton-Raphson method, we identify $l(\boldsymbol{\beta})$ as the log-likelihood for Poisson loglinear models.
- Recall Equation (5), and let,

$$l(\boldsymbol{\beta}) = \sum_i n_i \left(\sum_j x_{ij}\beta_j \right) - \sum_i \exp\left(\sum_j x_{ij}\beta_j \right). \tag{7}$$

- Then calculate the first order and second order derivatives,

$$u_j := \frac{\partial}{\partial \beta_j} l(\boldsymbol{\beta}) = \sum_i n_i x_{ij} - \sum_i \mu_i x_{ij}.$$

$$h_{jk} := \frac{\partial^2}{\partial \beta_j \partial \beta_k} l(\boldsymbol{\beta}) = -\sum_i \mu_i x_{ij} x_{ik}.$$

10.8 Loglinear Model Fitting: Iterative Methods

- Therefore, the tth approximation yields,

$$u_j^{(t)} = \sum_i x_{ij}(n_i - \mu_i^{(t)}),$$

  and

$$h_{jk}^{(t)} = -\sum_i \mu_i^{(t)} x_{ij} x_{ik}.$$

- Note that the tth approximation $\boldsymbol{\mu}^{(t)}$ for $\hat{\boldsymbol{\mu}}$ derives from $\boldsymbol{\beta}^{(t)}$ through

$$\boldsymbol{\mu}^{(t)} = \exp(\boldsymbol{X}\boldsymbol{\beta}^{(t)}).$$

- The iterative relationship is as follows,

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

  This in turn produces $\boldsymbol{\mu}^{(t+1)}$, and so on.

10.8 Loglinear Model Fitting: Iterative Methods

- The iterative process begins with all $\mu_i^{(0)} = n_i$, or with an adjustment such as $\mu_i^{(0)} = n_i + 0.5$ if any $n_i = 0$.
- For loglinear models with a concave $l(\boldsymbol{\beta})$, $\boldsymbol{\mu}^{(t)}$ and $\boldsymbol{\beta}^{(t)}$ usually converge rapidly to the ML estimates $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\beta}}$ as t increases.

(End of the extended discussion of ML estimation for Loglinear models.)