STAT 770 Nov. 4 Lecture 19B

Loglinear Models versus GLMs

Reading and Topics for this lecture: Chapter 9 Sections 1-3.

(1) Definition, side-conditions

(2) Interpretation of Parameters (Low-dim examples)

(3) Interpretation as GLM

Loglinear Model with 2 Categorical Factors

Data: X_a, Z_a categorical factors, $a = 1, \ldots, n$ indexes obs

Multinomial counts N_{xz}, cell probabilities π_{xz}

indexed by levels $x = 1, \ldots, I$ for X, $z = 1, \ldots, J$ for Z

Model: $\log \pi_{xz} = \lambda + \lambda_x^X + \lambda_z^Z + \lambda_{xz}^{XZ}$ **saturated**

Side-conditions: $\sum_{xz} p_{xz} = 1$, and for all x, z :

$$\sum_{i=1}^{I} \lambda_i^X = \sum_{j=1}^{J} \lambda_j^Z = \sum_{i=1}^{I} \lambda_{iz}^{XZ} = \sum_{j=1}^{J} \lambda_{xj}^{XZ} = 0$$

Parameter dimension: $1+(I-1)+(J-1)+(I-1)\cdot(J-1) = I\cdot J$

Could exclude: λ^{XZ}, *or* λ^X, λ^{XZ} *or* λ^Z, λ^{XZ}

Interpretation of Model Coefficients

Consider the loglinear model from previous slide:

$$\log(p_x/p_{x'}) \;=\; \log\left(\frac{\sum_z p_{xz}}{\sum_z p_{x'z}}\right) \;=\; \lambda_x^X - \lambda_{x'}^X \;+\; \log\left(\frac{\sum_z \exp(\lambda_z^Z + \lambda_{zx}^{XZ})}{\sum_z \exp(\lambda_z^Z + \lambda_{zx'}^{XZ})}\right)$$

cancellation occurs in the last term only if λ^{XZ} is absent

the same issue mentioned in the last slide of Lec. 19A

Also:

$$\theta_{xz} \equiv \frac{p_{xz}\, p_{x+1,z+1}}{p_{x,z+1}\, p_{x+1,z}} \;=\; \exp\left(\lambda_{xz}^{XZ} + \lambda_{x+1,z+1}^{XZ} - \lambda_{x,z+1}^{XZ} - \lambda_{x+1,z}^{XZ}\right)$$

Loglinear Model with 3 Categorical Factors

Suppose the model is:

$$\log \pi_{xzw} = \lambda + \lambda_x^X + \lambda_z^Z + \lambda_w^W + \lambda_{xz}^{XZ} + \lambda_{xw}^{XW}$$

with side-conditions $0 = \lambda_+^X = \lambda_+^Z = \lambda_+^W = \lambda_{+z}^{XZ} = \lambda_{x+}^{XZ} \ldots$ Now

$$\theta_{xz(w)} = \frac{p_{xzw}\, p_{x+1,z+1,w}}{p_{x,z+1,w}\, p_{x+1,zw}} = \exp\left(\lambda_{xz}^{XZ} + \lambda_{x+1,z+1}^{XZ} - \lambda_{x,z+1}^{XZ} - \lambda_{x+1,z}^{XZ}\right)$$

is the same for all w*: the effect of having no 3-way interaction*

R Software for Log-Linear Models

(1). Direct specification using indices for factors and interactions – `loglin`

(2). Reconstructing model as GLM – `glm`

would require re-interpreting GLM coeff's !

(3). Fitting loglinear model but specifying in GLM syntax – `loglm` in `MASS`

The difficulty in moving between model-specifications is the lack of side-conditions in logistic-regression interaction factors.

See `Lec19BLogLin` Script for further discussion!!

GLM Interpretation of Loglinear Model

logLik for loglinear 3-factor model is: $\sum_{x,z,w} N_{xzw} \log \pi_{xzw}$

log probabilities in (X, Z, W, XZ) canonical Poisson model of the same form without same side-conditions, uses dummy columns

$\left(I_{[X_i=x]},\ 1 < x \le I\right), \quad \left(I_{[Z_i=z]},\ 1 < z \le J\right), \quad \ldots\ ,$

$\left(I_{[X_i=x,Z_i=z]},\ 1 < x \le I,\ 1 < z \le J\right)$ with coeff's $\beta_x^X,\ \beta_z^Z,\ \beta_{xz}^{XZ}, \ldots$

Poisson regression logLik has the **same** regression terms as loglinear, but coefficients become linearly (invertibly) transformed. Will describe the parameter transformation in detail next time.