STAT 770 Nov. 11 Lecture 21

Transforming Poisson GLM Parameters to Loglinear

Reading and Topics for this lecture: Chapter 9 Secs. 1-3, 5-7.

(1) Poisson Regression Model Matrix & Contrasts

(2) Transforming Parameters between Side Conditions

(3) Confidence Intervals for Parameters & Cell-Probs

(4) Multinomial Logistic as Loglinear / Poisson Model

(5) Iterative Proportional Fitting & Raking

Poisson Regression Model Matrix

Factors X, Z, W, Data $\{Y_a\}_{a=1}^{M}$, Combinations $a \leftrightarrow (i, j, k)$

Model (W XZ) matrix H, dummy columns $(IJK) \times (K + IJ - 1)$
– ordering of columns matters a lot!

Coefficients β maximize logLik$= \sum_a Y_a\,(H\beta)_a - \sum_a \exp((H\beta)_a)$

Model specification (in `glm` in `R`): $Y \sim W + X * Z$ either by directly supplying the dummy columns to reflect desired side-condition **or** using `contrasts()` function in `R`

These alternative ways of specifying the model fitting are covered in the current `R` script **Lec21Loglin.RLog**.
See especially Section (8) of that Script.

Mapping Between Coefficient-sets in the Example

The mapping from $\underline{\lambda} = (\lambda_0, \lambda_2^X, \lambda_2^Z, \lambda_2^W, \lambda_{22}^{XZ})$ to $\underline{\gamma} = (\gamma_0, \gamma_2^X, \gamma_2^Z, \gamma_2^W, \gamma_{22}^{XZ})$ is linear and invertible,

$$\underline{\gamma} = \begin{pmatrix} 1 & -1 & -1 & -1 & 1 \\ 0 & 2 & 0 & 0 & -2 \\ 0 & 0 & 2 & 0 & -2 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix} \underline{\lambda} \equiv C\, \underline{\lambda}$$

and the constraint determining λ_0 is unaffected by this mapping. Similarly:

$$\lambda_{xz}^{XZ} = \frac{(-1)^{x+z}}{4} \gamma_{22}^{XZ}, \quad \lambda_w^W = \frac{(-1)^w}{2} \gamma_2^W$$

$$\lambda_x^X = \frac{(-1)^x}{4} (2\gamma_2^X + \gamma_{22}^{XZ}), \quad \lambda_z^Z = \frac{(-1)^z}{4} (2\gamma_2^Z + \gamma_{22}^{XZ})$$

Using the Mapping for Estimates & Variances

If you know matrix C transforming between pararameter for different size-conditions, use it directly for estimates and variances.

Poisson GLM estmates give MLEs, variances $\widehat{\gamma}$, $\widehat{V}(\widehat{\gamma})$

Transformation property of MLE says:

$$\widehat{\lambda} = C^{-1}\widehat{\gamma}\,, \qquad \widehat{V}(\widehat{\lambda}) \;=\; C^{-1}\,\widehat{V}(\widehat{\gamma})\,(C^{-1})^{tr}$$

Implementation is shown in Sec. (6) of script `Lec21Loglin.RLog`

Confidence Intervals for Loglinear Coefficients

From $\hat{\lambda}$, $\hat{V}(\hat{\lambda})$ directly find Wald CIs $\hat{\lambda}_r \pm z_{\alpha/2} \left[\hat{V}(\hat{\lambda})_{rr} \right]^{1/2}$
or if estimates came directly from Poisson `glm`, use `confint`

For π_a CI, use definition as nonlinear function $\pi_a = g_a(\lambda)$ of λ plus Δ Method

Start from estimate $\hat{\lambda}$ (incl. intercept) and its model-matrix H^*

$$\hat{\pi}_a = \text{exp}((H^* \hat{\lambda})_a) / \sum_{b=1}^{M} \text{exp}((H^* \hat{\lambda})_b) \equiv g_a(\hat{\lambda})$$

for M-vector-valued function $\mathbf{g} = \{g_a\}_{a=1}^{M}$ with Jacobian $\mathcal{J}_{\mathbf{g}}$

$$\mathcal{J}_{\mathbf{g}}(\lambda) = (\nabla_\lambda \mathbf{g}^{tr}(\lambda))^{tr} = \left[\text{diag}(\underline{\pi}) - \underline{\pi}\, \underline{\pi}^{tr} \right] H^*$$

Confidence Interval for Loglinear Cell-Prob

From last slide (π_a's do not actually depend on λ_0)

$$\widehat{\pi}_a \;=\; \exp((H^* \widehat{\lambda})_a) / \sum_{b=1}^{M} \exp((H^* \widehat{\lambda})_b) \;\equiv\; g_a(\widehat{\lambda})$$

for M-vector-valued function $\mathbf{g} = \{g_a\}_{a=1}^{M}$ with Jacobian $\mathcal{J}_{\mathbf{g}}$

$$\mathcal{J}_{\mathbf{g}}(\lambda) \;=\; (\nabla_\lambda \, \mathbf{g}^{tr}(\lambda))^{tr} \;=\; \left[\text{diag}(\underline{\pi}) \;-\; \underline{\pi}\, \underline{\pi}^{tr} \right] H^*$$

The Delta-Method variance-covariance matrix for $\underline{\widehat{\pi}}$ is

$$\mathcal{J}_{\mathbf{g}}(\widehat{\lambda}) \, \widehat{V}(\widehat{\lambda}) \, (\mathcal{J}_{\mathbf{g}}(\widehat{\lambda}))^{tr} = \left[\text{diag}(\underline{\pi}) - \underline{\pi}\, \underline{\pi}^{tr} \right] H^* \, \widehat{V}(\widehat{\lambda}) \, H^{*tr} \left[\text{diag}(\underline{\pi}) - \underline{\pi}\, \underline{\pi}^{tr} \right]$$

The diagonal-element square roots are the SE's of $\widehat{\pi}_a$

Multinomial Logistic Models are Conditional Loglinear

For discrete covariates $\underline{X}_a$, and $Y_a \in \{1, \ldots, K\}$ multinomial

multinomial-logistic says for $k = 2, \ldots, K$,

$$P(Y_a = k \,|\, \underline{X}_a) \;=\; \exp(\beta^{(k)tr}\underline{X}_a) \Big/ \Big[1 + \textstyle\sum_{r=2}^{K} \exp(\beta^{(r)tr}\underline{X}_a)\Big]$$

Therefore $(\underline{X}_a, Y_a)$ satisfy a loglinear model with $M \times K$ cells if $\underline{X}_a \in \{1, \ldots, M\}$ is loglinear on M cells (which may be obtained by multi-way cross-tabulation of factor levels).

This is true in particular if $K = 2$, the logistic case. And if $K > 2$, Agresti makes a similar observation about the model for outcomes conditioned to fall in a specified pair k_1, k_2 of outcome categories. That seems to be the main point made in Sec. 9.5.

Distinction between GLM and Loglinear Approaches

If a dataset (X_a, Y_a, Z_a, W_a) consists completely of discrete variables, loglinear models are GLMs.

GLMs can be specified more generally

- not requiring that all variables be modeled, and
- not requiring that all variables be discrete;
- and including cluster random effects (Ch. 13)

Loglinear models seem popular among (social-science) investigators focused on interactions as a way of understanding *causal pathways*

Iterative Proportional Fitting (Sec.9.7.2-9.7.4)

Start with counts (summed weights, in survey contexts) $\mu_a^{(0)}$

$a \leftrightarrow (i, j, k, \cdots)$ to be marginalized (some factor indices summed)

- restrict to example (i,j,k) in model (XZ, W)
- target counts $\mu_{xz+} = Y_{xz+}, \;\; \mu_{++w} = Y_{++w}$
- in iterative passes $m \geq 0$: hit targets exactly by multiplication using respective factors $f_{xz}, \, f_w$

$$\mu_{ijk}^{(2m+1)} = \mu_{ijk}^{(2m)} \frac{Y_{ij+}}{\mu_{ij+}^{(2m)}} \;, \quad \mu_{ijk}^{(2m+2)} = \mu_{ijk}^{(2m+1)} \frac{Y_{++k}}{\mu_{++k}^{(2m+1)}}$$

Convergence occurs under general conditions (usually quickly) when all $\mu_a^{(0)}$ and marginals are positive