STAT 151A Lecture 37

Henry Liev

21 November 2025

Remark 0.1 (Multinomial Logistic Regression)

Suppose $y_i \in \begin{Bmatrix} 1 \\ 2 \\ \vdots \\ m \end{Bmatrix}$ categories

In math, $y_i \sim \text{Mult}(1, [\pi_1, \dots, \pi_m])$

We also have p regressors $\vec{x}_1, \dots, \vec{x}_p$ and an intercept $\vec{1}$

Our model: Define $\pi_{ij} = \mathbb{P}(y_i = j | x_i)$

$\pi_{ij} = \frac{\exp(\gamma_{0j} + \gamma_{1j}x_{1i} + \dots + \gamma_{pj}x_{pi})}{1 + \sum_{l=1}^{m-1} \exp(\gamma_{0l} + \gamma_{1l}x_{1i} + \dots + \gamma_{pl}x_{pi})}, j \in \{1, \dots, m-1\}$

$\pi_{im} = 1 - \sum_{l=1}^{m-1} \pi_{il}$

Total number of parameters: $(m-1)(p+1)$

If $m = 3$

$\begin{bmatrix} \gamma_{0voc} & \gamma_{1voc} & \cdots & \gamma_{pvoc} \\ \gamma_{0acad} & \gamma_{1acad} & \cdots & \gamma_{pacad} \end{bmatrix}$

Always: $0 < \pi_{ij} < 1$, $\sum_{j=1}^{m} \pi_{ij} = 1$

Fit via MLE just like logistic regression

Interpretation:

$\log\left(\frac{\pi_{ij}}{\pi_{im}}\right) = \gamma_0 + \gamma_{1j}x_{1i} + \dots + \gamma_{pj}x_{pi}$

So regression coefficient γ_{lj} (Coefficient of varaible l describing category j)

Tells us how different conditional by odds of category j versus category m are for individuals 1-unit different in covariate l

Odds of $j = \frac{\mathbb{P}(y_i = j)}{\mathbb{P}(y_i \neq j)}$

Conditional odds of j given j or $m = \frac{\mathbb{P}(y_i = j)|\text{j or m}}{P(y_i \neq j|\text{j or m})}$

Basically, m is the natural reference category

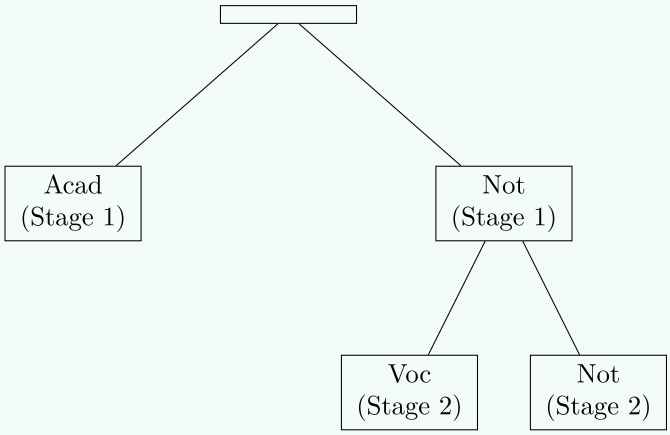
All γ_{lj} are defined relative to m

Also

$$\log\left(\frac{\pi_{ij}}{\pi_{ik}}\right) = \log\left(\frac{\pi_{ij}/\pi_{im}}{\pi_{ik}/\pi_{im}}\right) = \log\left(\frac{\pi_{ij}}{\pi_{im}}\right) - \log\left(\frac{\pi_{ik}}{\pi_{im}}\right)$$
$$= (\gamma_{0j} - \gamma_{0k}) + (\gamma_{1j} - \gamma_{1k})x_{1i} + \dots + (\gamma_{pj - \gamma_{pk}})x_{pi}$$

Remark 0.2 (Inference)

Also just like regression, using MLE
$H_0 : \gamma_{j1} = \gamma_{j2} = \cdots = \gamma_{j(m-1)} = 0$
Usually do likelihood ratio tests



$\widehat{\mathbb{P}}(voc_i) = \left[1 - \widehat{\mathbb{P}}(acad_i)\right] \widehat{\mathbb{P}}(voc_i)$
Cool fact: do inference for whether writing score matters in both models at once
Get LR statistic for dropping writing in model 1
Get LR statistic for dropping writing in model 2
Add them, compare to χ^2_{df}