STAT 770 Feb. 16 Lecture

Extended Case Study with LRT Data Analysis

Reading for this lecture: Section 1.6 in Agresti, along with our LRT material mirroring section 16.3, plus R stuff. There is a little, but not much, on the Chapter 2 2-way table LRTs that we will go back to next week.

The R Script that will be followed in this Lecture is `Rscript5B.txt`.

Some mathematical formulas and associated reminders and comments are in these pdf slides.

Goodness of Fit Tests in One-Way Layouts

We know when count data with total n fall in cells $j = 1, \ldots, K$ and the cell probabilities π_j are unknown but modeled as $\pi_j(\beta)$, then we use the Wilks Theorem to test whether the model is 'adequate' through the Likelihood Ratio Statistic

$$\Lambda = G^2 = 2 \sum_{j=1}^{K} N_j \log \Big(N_j \Big/ (n\, \pi(\widehat{\beta}_r)) \Big)$$

or the X^2 statistic that approximates it,

$$X^2 = \sum_{k=1}^{K} (N_k - E_k)^2 / E_k \quad , \qquad E_k = n\, \pi(\widehat{\beta}_r)$$

All these statistics come from the multinomial likelihood proportional to $\prod_{j=1}^{K} \pi_j^{N_j}$ or $\prod_{j=1}^{K} \pi_j(\beta)^{N_j}$ and depend on large n to have approximate distribution $\chi^2_{K-1-dim(\beta)}$

More on Goodness of Fit

When total count is not restricted, e.g. the cell counts $N_j \sim$ Poisson(λ_j) are all random, and $n = \sum_{j=1}^{K} N_j$ is random, Wilks' Theorem applies but the details are a little different.

Likelihood: $\prod_{j=1}^{K} \left(e^{-\lambda_j} \lambda_j^{N_j} / N_j! \right)$ or $\prod_{j=1}^{K} \left(e^{-\lambda_j(\beta)} \lambda_j(\beta)^{N_j} / N_j! \right)$

So one of the less familiar problems that we have to consider in today's case-study is the restricted maximum likelihood calculation of the latter kind of product, with respect to β.

The next slide addresses why the useful maximization problems to do for LRT often involve the additional complication of **Grouped Data**.

Grouped Data LRTs

Suppose *iid* data W_a, $1 \le a \le n$ are categorized into intervals or regions $Y_a = j \;\Leftrightarrow W_a \in B_j$, $\{B_j\}_{j=1}^K$ partition the data-space

if $W_a \sim f(w, \beta)$, cell-prob $\pi_j = P(W_a \in B_j) = \pi_j(\beta) \equiv \int_{B_j} f(w, \beta)dw$

Grouping can be $\sum_{w \in B_j} p(w, \beta)$ **if** $W_j \sim p(w, \beta)$ **discrete**

Grouped-data Likelihood $L(\beta) = \prod_{j=1}^K \left(\sum_{w \in B_j} p(w, \beta) \right)^{N_j}$

Restricted MLE $=$ argmax$_\beta$ $L(\beta)$ used in LRT

Unrestricted MLEs again $\widehat{\pi}_j = N_j/n$, $N_j = \sum_{a=1}^n I_{[W_a \in B_j]}$