STOR 556: ADV METH DATA ANAL
Instructor: Richard L. Smith

Class Notes #14:

February 26, 2019

THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

Homework 5

- Chapter 5, Problems 2 and 5 (pages 99–101)
- Due date: Tuesday March 5
- I have posted the TA's solutions to HW1–4. Please treat these as *for personal use only* and do not pass on to anyone outside the class!

Poisson Model for 2×2 Table

- Counts y_{ij}: assume for convenience i, j are each 0 or 1
- $\mathrm{E}(y_{ij}) = \lambda_{ij}$, write in form

$$\log \lambda_{ij} = \mu + \alpha i + \beta j + \gamma ij.$$

- Poisson log likelihood

$$\ell = \sum_i \sum_j \left(y_{ij} \log \lambda_{ij} - \lambda_{ij} - \log y_{ij}! \right). \qquad (1)$$

- Saturated model: assume all of μ, α, β, γ, number of parameters equals number of observations, solve exactly, $\hat{\lambda}_{ij} = y_{ij}$ and

$$\ell = \sum_i \sum_j \left(y_{ij} \log y_{ij} - y_{ij} - \log y_{ij}! \right). \qquad (2)$$

Poisson Model With No Interaction

- $\log \lambda_{ij} \; = \; \mu + \alpha i + \beta j.$
- $\ell \; = \; \sum_i \sum_j \left(y_{ij}(\mu + \alpha i + \beta j) - e^{\mu + \alpha i + \beta j} - \log y_{ij}! \right).$
- **Claim:** The MLE is achieved when $\lambda_{ij} = \frac{y_{i+} y_{+j}}{y_{++}}$ where the $+$ signs in the suffices denote summation over the relevant indices.
- Comparing formulas (1) and (2), we have

$$\text{Deviance} \; = \; 2 \sum_i \sum_j y_{ij} \log \frac{y_{ij}}{\hat{\lambda}_{ij}}.$$

- We also have the Pearson Chi-Square statistic

$$X^2 \; = \; \sum_i \sum_j \frac{(y_{ij} - \hat{\lambda}_{ij})^2}{\hat{\lambda}_{ij}}.$$

- Both are asymptotically χ_1^2.

Multinomial Model

- Assume total sample size n is fixed, divided into four boxes with probabilities $\{p_{ij}\}$. The probability of a particular allocation $\{y_{ij}\}$ is then

$$\frac{n!}{\prod_i \prod_j y_{ij}!} \prod_i \prod_j p_{ij}^{y_{ij}}.$$

- Ignoring constants, the log likelihood is of the form

$$\ell = \sum_i \sum_j y_{ij} \log p_{ij}$$

essentially same as Poisson model

Binomial Model

- Assume *either* the row totals or the column totals are fixed

- Then, e.g. given n_0, n_1 totals in rows 0,1, assume

$$y_{ij} = \text{Bin}(n_i, p_{ij})$$

  with formulas for p_{ij} similar to the multinomial model

- The interpretations of the p_{ij}'s are different however — they are conditional probabilities (given the row totals), not marginal probabilities

- A similar model is possible fixing column totals

- All these models lead to very similar math

Hypergeometric Distribution

- Population of N items, k of which are "success"
- Sample of size n, drawn randomly without replacement
- Probability exactly x elements in the sample are successes:

$$\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

  valid for any x for which the numerator is > 0.

- This may also be written as

$$\frac{k!(N-k)!n!(N-n)!}{x!(k-x)!(n-x)!(N-k-n+x)!N!}$$

  which (when applied to a 2×2 table) is equivalent to the formula on page 108 of the text.

- Leads to *Fisher's Exact Test* (`fisher.test` in R.)