Logistic Regression Model Fitting

Maximum likelihood estimation of model parameters

- Data: $(y_i, m_i, \boldsymbol{x}_i)$, $i = 1, \cdots, n$
- Model: $y_i \sim Bin(m_i, \pi_i)$ where $g(\pi_i) = \boldsymbol{x}_i^T \boldsymbol{\beta}$
- Log likelihood:
  $l(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[y_i \log(\frac{\pi_i}{1-\pi_i}) + m_i \log(1-\pi_i) + \log \begin{pmatrix} m_i \\ y_i \end{pmatrix} \right]$
- Fisher scoring algorithm (IRLS)
- Convergence: not always obtainable

Goodness of Fit

Overall goodness-of-fit (for grouped data)

- Generalized Pearson χ^2 statistic

$$G(\hat{\boldsymbol{\pi}}, \boldsymbol{y}) = \sum_{i=1}^{n} \frac{(y_i - m_i\hat{\pi}_i)^2}{m_i\hat{\pi}_i(1-\hat{\pi}_i)}$$

- Deviance

$$2\sum_{i=1}^{n}\left[y_i \log \frac{y_i}{m_i\hat{\pi}_i} + (m_i - y_i)\log \frac{m_i - y_i}{m_i(1-\hat{\pi}_i)}\right]$$

Note: *if $y_i = 0$ or $y_i = m_i$, the zero-valued linear term overrides the log term, so the deviance is still well defined.*

- Both approximately follow $\chi^2(n-p)$ distribution when m_i's are large
- What about ungrouped data?

For sparse data (m_i small)

- Deviance and Generalized Pearson χ^2 statistics cannot be used for goodness-of-fit measure.
- Hosmer-Lemeshow statistic: χ^2_{HL}
  - Group observations into g (≈ 10) categories based on covariate patterns

$$\chi^2_{HL} = \sum_{i=1}^{g} \frac{(O_i - n_i\bar{\pi}_i)^2}{n_i\bar{\pi}_i(1-\bar{\pi}_i)}$$

  - O_i is the number of observed 1s in the ith group
  - $\bar{\pi}_i$ is the average of $\hat{\pi}_{ij}$ in the ith group
  - Approximately $\chi^2(g-2)$
  - Large value means lack of fit

Confidence Interval

For $\boldsymbol{\beta}$, we know that

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \overset{asy}{\sim} N(0, \mathcal{I}(\boldsymbol{\beta})^{-1}),$$

Thus the CI for β_j is $\hat{\beta}_j \pm Z_{1-\alpha/2} se(\hat{\beta}_j)$, where $Z_{\alpha/2}$ is the $\alpha/2$ quantile of the standard normal distribution.

What about $f(\boldsymbol{\beta})$?

- The asymptotic distribution of $f(\hat{\boldsymbol{\beta}})$ is

$$N(f(\boldsymbol{\beta}), \frac{\partial f}{\partial \boldsymbol{\beta}^T} \mathcal{I}(\boldsymbol{\beta})^{-1} \frac{\partial f}{\partial \boldsymbol{\beta}})$$

- Thus the CI for $f(\boldsymbol{\beta})$ is

$$f(\hat{\boldsymbol{\beta}}) \pm Z_{1-\alpha/2} se(f(\hat{\boldsymbol{\beta}}))$$

- For example, the asymptotic distribution of $\hat{\eta}_* = \boldsymbol{x}_*^T \hat{\boldsymbol{\beta}}$ is

$$N(\boldsymbol{x}_*^T \boldsymbol{\beta}, \boldsymbol{x}_*^T \mathcal{I}(\hat{\boldsymbol{\beta}})^{-1} \boldsymbol{x}_*)$$

- Subsequently, the asymptotic CI for $\eta_* = \boldsymbol{x}_*^T \boldsymbol{\beta}$ is

$$\boldsymbol{x}_*^T \hat{\boldsymbol{\beta}} \pm Z_{\alpha/2} \sqrt{\boldsymbol{x}_*^T \mathcal{I}(\hat{\boldsymbol{\beta}})^{-1} \boldsymbol{x}_*},$$

- Ultimately, we are interested in obtaining the CI for π_*.

- $\pi_* = g^{-1}(\eta_*)$

- Point estimate: $\hat{\pi}_* = g^{-1}(\boldsymbol{x}_*^T \hat{\boldsymbol{\beta}})$

- $(1-\alpha)100\%$ CI for $\eta_* = \boldsymbol{x}_*^T \boldsymbol{\beta}$ is $(\hat{\eta}_L, \hat{\eta}_R)$, where

$$\hat{\eta}_L = \boldsymbol{x}_*^T \hat{\boldsymbol{\beta}} - z_{\alpha/2}\sqrt{\boldsymbol{x}_*^T \mathcal{I}(\hat{\boldsymbol{\beta}})^{-1} \boldsymbol{x}_*}$$

$$\hat{\eta}_R = \boldsymbol{x}_*^T \hat{\boldsymbol{\beta}} + z_{\alpha/2}\sqrt{\boldsymbol{x}_*^T \mathcal{I}(\hat{\boldsymbol{\beta}})^{-1} \boldsymbol{x}_*}$$

- CI for π_*:

$$\left[g^{-1}(\hat{\eta}_L), g^{-1}(\hat{\eta}_R)\right]$$

Example

- **Show/no-show:**
  Investigating the relation between show/no-show and appointment lag.
- **Peer reviewed publication:**
  Comparing urology fellows with and without time off in terms of their proportions of urological publications.