STAT 151A Lecture 34

Henry Liev

14 November 2025

Remark 0.1 (Fisher Information for Logistic Regression)

Focus on simple logistic regression
$y_i \sim \text{Bern}(\pi_i)$
$\pi_i = \frac{1}{1+\exp(-\alpha-\beta x_i)}$

$$\mathcal{I}_n(\alpha,\beta) = -\sum \mathbb{E}\left[\frac{\partial^2}{\partial \begin{pmatrix}\alpha\\ \beta\end{pmatrix}^2} \log f(y_i|x_i,\alpha,\beta)\right] \in \mathbb{R}^{2\times 2}$$

$$-\sum \mathbb{E}\begin{bmatrix} \frac{\partial^2}{\partial\alpha^2}\log f(y_i|x_i,\alpha,\beta) & \frac{\partial^2}{\partial\alpha\partial\beta}\log f(y_i|x_i,\alpha,\beta) \\ \frac{\partial^2}{\partial\alpha\partial\beta}\log f(y_i|x_i,\alpha,\beta) & \frac{\partial^2}{\partial\beta^2}\log f(y_i|x_i,\alpha,\beta)\end{bmatrix}$$

$$-\mathbb{E}\begin{bmatrix} \frac{\partial^2}{\partial\alpha^2}\sum\log f(y_i|x_i,\alpha,\beta) & \frac{\partial^2}{\partial\alpha\partial\beta}\sum\log f(y_i|x_i,\alpha,\beta) \\ \frac{\partial^2}{\partial\alpha\partial\beta}\sum\log f(y_i|x_i,\alpha,\beta) & \frac{\partial^2}{\partial\beta^2}\sum\log f(y_i|x_i,\alpha,\beta)\end{bmatrix}$$

From Monday (Max likelihood problem):
$\frac{\partial l(y_1,\ldots,y_n)}{\partial\alpha} = \sum y_i - \sum \frac{1}{1+\exp(-\alpha-\beta x_i)}$
$\frac{\partial l}{\partial\beta} = \sum y_i x_i - \sum \frac{x_i}{1+\exp(-\alpha-\beta x_i)}$
$\frac{\partial^2 l}{\partial\alpha^2} = -\sum \frac{\exp(-\alpha-\beta x_i)}{[1+\exp(-\alpha-\beta x_i)]^2} = -\sum \pi_i(1-\pi_i)$
$-\sum \frac{\exp(-\alpha-\beta x_i)}{[1+\exp(-\alpha-\beta x_i)]^2}\cdot\frac{[\exp(\alpha+\beta x_i)]^2}{[\exp(\alpha+\beta x_i)]^2} = \sum \frac{\exp(\alpha+\beta x_i)}{[1+\exp(-\alpha-\beta x_i)]}$
$\frac{\partial^2 l}{\partial\alpha\partial\beta} = -\sum \pi_i(1-\pi_i)x_i$
$\frac{\partial^2}{\partial\beta^2} = \pi_i(1-\pi_i)x_i^2$

$$\boxed{\mathcal{I}_n(\alpha,\beta) = -\mathbb{E}\begin{bmatrix} -\sum\pi_i(1-\pi_i) & -\sum\pi_i(1-\pi_i)x_i \\ -\sum\pi_i(1-\pi_i)x_i & -\sum\pi_i(1-\pi_i)x_i^2\end{bmatrix}}$$

$$= \begin{pmatrix}\vec{1} & \vec{x}\end{pmatrix}^\intercal \operatorname{diag}(\pi_i(1-\pi_i))\begin{pmatrix}\vec{1} & \vec{x}\end{pmatrix}, \mathbf{V} = \operatorname{diag}(\pi_i(1-\pi_i))$$

$$\text{Var}\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\end{pmatrix} \overset{n\to\infty}{\approx} [\mathbf{X}^\intercal\mathbf{V}\mathbf{X}]^{-1}$$

Remark 0.2 (Tests and CIs)

$H_0 : \beta_j = 0$
From our theory, under H_0
$\frac{\hat{\beta}_j - 0}{\widehat{\mathrm{SE}}(\hat{\beta}_j)} \overset{n\to\infty}{\to} N(0,1), \widehat{\mathrm{SE}}(\hat{\beta}_j) = \sqrt{(\mathbf{X}^\intercal \mathbf{V} \mathbf{X})^{-1}_{jj}}$
In practice compute $z_j = \frac{\hat{\beta}_j}{\widehat{\mathrm{SE}}(\hat{\beta}_j)}$ Compare to $N(0,1)$ (Wald test)
Can get CIs: $\hat{\beta}_j \pm z_{1-\alpha/2} \cdot \widehat{\mathrm{SE}}(\hat{\beta}_j)$
What about F-style tests?
$H_0 : L\beta = c$
Under H_0: $\left[L\hat{\beta} - c\right]^\intercal \left[L\widehat{\mathrm{Var}}(\hat{\beta})L^\intercal\right]^{-1} (L\hat{\beta} - c) \overset{n\to\infty}{\sim} \chi^2_q, q = \mathrm{Rank}(L)$
Connect to NLM:
$H_0 : \beta_j = 0 \qquad \frac{\hat{\beta}_j - 0}{\widehat{\mathrm{SE}}(\hat{\beta}_j)} \sim t_{n-(p+1)} \overset{n\to\infty}{\to} N(0,1)$
$H_0 : L\beta = c \qquad \left[L\hat{\beta} - c\right]^\intercal \left[L\widehat{\mathrm{Var}}(\hat{\beta})L^\intercal\right]^{-1} (L\hat{\beta} - c) \sim F_{q,(n-(p+1))} \overset{n\to\infty}{\to} \chi^2_q$
These are all Wald tests on CI's
Based on Theorem B (Asymptotic distribution of $\hat{\beta}$)
Another way to do inference after MLE: based on asymptotic distribution of likelihood ratio
Let's say we fit 2 models full model: $\mathcal{L}_1$ and reduced model: $\mathcal{L}_0$
Compute $-2\log\left[\frac{\mathcal{L}_0}{\mathcal{L}_1}\right] \overset{n\to\infty}{\to} \chi^2_q, q = p_1 - p_0$
In practice, Likelihood ratio test and Wald test (of same hypothesis) differ slightly, likelihood ratio test is often preferred
In R anova(glm.full, glm.reduced, "wald" or "lrt")
More details: Fox appendix D.6.3
Likelihood Ratio gives use R^2 analog
A **saturated model** fits the data perfectly
For logistic regression, when $y_i = 1 \Rightarrow \pi_i = 1, y_i = 0 \Rightarrow \pi_i = 0$
$\mathcal{L}_{sat} = \prod y_i^{\pi_i}(1 - y_i)^{1-\pi_i} = \prod 1^1 \cdot 0^0 = 1$
TotSS = RegSS for a model where $\hat{y} = y$
Instead of $R^2 = \frac{RegSS}{TotSS}$
For logistic set residual deviance for model m: $ErrSS \leftrightarrow D_m = -2\log\left[\frac{\mathcal{L}_m}{\mathcal{L}_{sat}}\right] = -2\log\mathcal{L}_m$
R^2 analog: $1 - \frac{D_m}{D_0}$, where D_0 is the residual deviance for a model with just an intercept