

MB: Binary regression (with the logit model)

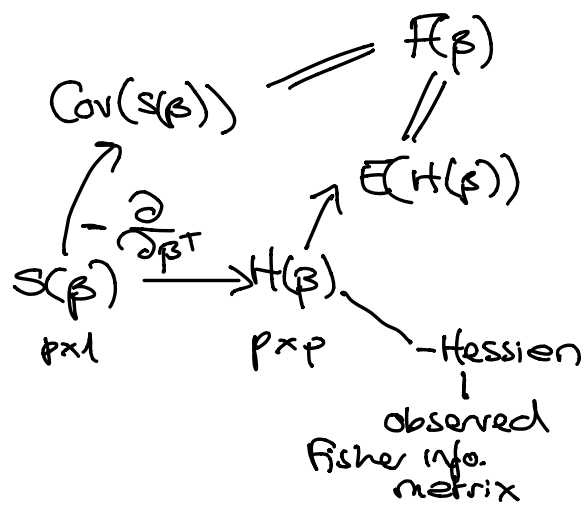
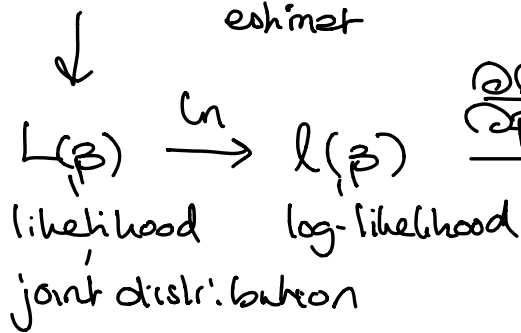
Week 3: 25.09.2017

Assumptions

1. $Y_j \sim \text{bin}(n_j, \pi_j)$
2. $\eta_j = x_j^T \beta$
3. $\eta_j = \ln\left(\frac{\pi_j}{1-\pi_j}\right) \leftarrow \text{logit link}$

Aim: estimate parameters β

maximize to find
maximum likelihood
estimate



$\hat{\beta}$: found from $s(\hat{\beta})=0$, by $\beta^{(t+1)} = \beta^{(t)} + F(\beta^{(t)})^{-1} s(\beta^{(t)})$

Fisher scoring / Newton
Raphson

$$\hat{\beta} \approx N_p(\beta, F^{-1}(\hat{\beta}))$$

$$\uparrow$$

$$(X^T W X)^{-1}$$

$$W = \text{diag}(\pi_j, \pi_j, (1-\pi_j))$$

$$\hat{\beta}_k \approx N(\beta_k, a_{kk}(\hat{\beta}))$$

$$\uparrow$$

$$A = F^{-1}, \text{ } a_{kk} \text{ is element}$$

$$\text{kk of } A$$

Q: why $a_{kk}(\hat{\beta})$ end not $a_{kk}(\beta_k)$?

Because $F(\hat{\beta})$ is a function of the full $\hat{\beta}$ vector, end
we take the inverse to get $\hat{C}_k(\hat{\beta}_k)$.

Wald test

$$H_0: C_{\beta} = d \quad \text{vs.} \quad H_1: C_{\beta} \neq d$$

$$r \times p$$

$$\hat{\beta} \approx N_p(\beta, F^{-1}(\hat{\beta}))$$

$$C_{\beta}^{\hat{\beta}} \approx N_r \left(\underset{d}{C_{\beta}}, C F^{-1}(\hat{\beta}) C^T \right)$$

$$W = (C_{\hat{\beta}}^{\hat{\beta}} - d)^T [C F^{-1}(\hat{\beta}) C^T]^{-1} (C_{\hat{\beta}}^{\hat{\beta}} - d) \approx \chi_r^2$$

$$\text{P-value: } P(\chi_r^2 > W)$$

||
when H_0
is true

Likelihood ratio test

A: larger model H_1 true $\Rightarrow \hat{\beta}_A$ with $L(\hat{\beta}_A)$

B: smaller model H_0 true $\Rightarrow \hat{\beta}_B$ $L(\hat{\beta}_B)$

Q: Can $L(\hat{\beta}_B) > L(\hat{\beta}_A)$? No, same as for SSE!
MLR

$$LRT = -2 (\ln L(\hat{\beta}_B) - \ln L(\hat{\beta}_A))$$

$$\approx \chi^2_{\# \text{par A} - \# \text{par B}}$$

\nearrow asymptotic \leftarrow Sler. Inf THAYZU (ch10)
when H_0 true

Deviance

G groups with n_j obs in group j

Saturated model: a perfect fit to each group

\nearrow
"imaginary model"

$$\hat{\pi}_j = \frac{y_j}{n_j}$$

$$\frac{\exp(x_j^T \hat{\beta})}{1 + \exp(x_j^T \hat{\beta})}$$

Candidate model: our model

$\hat{\beta}$ is our MLE

$$\hat{\pi}_j = h(\eta(\hat{\beta}))$$

Deviance = LRT with $A = \text{Saturated}$
 $B = \text{Candidate}$

$$D = -2 \left(\underset{\substack{\uparrow \\ \text{Cand}}}{\ell(\hat{\mu})} - \ell(\hat{\mu}) \right) \approx \chi^2_{G-p}$$

Used to evaluate if a model is "good".