Module 3: Binary regression - week 2 18.09.2017
→ parameter estimation

Assumptions:

1. $Y_i \in \{0,1\}$, $Y_i \sim \text{bin}(n_i = 1, \pi_i)$ ($n_i > 1$ later)
   exponential family
   response
   $\mu_i = E(Y_i) = \pi_i$
   $\sigma_i^2 = \text{Var}(Y_i) = \pi_i(1-\pi_i)$

2. Linear predictor: $\eta_i = x_i^T \beta$
   $\underset{p\times 1}{x_i} = (1, x_{i1}, x_{i2}, \dots, x_{ik})$
   $\begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix} = \underset{n\times p}{X}$
   design matrix

3. Mean μ_i vs linear predictor η_i

$$\mu_i = h(\eta_i) \qquad \eta_i = g(\mu_i) \qquad (g^{-1} = h)$$

response function $(h^{-1} = g)$ link function

Logit (link) model: $\mu_i = \pi_i$

$$\pi_i = \frac{\exp(\eta_i)}{1+\exp(\eta_i)} \iff \eta_i = \underbrace{\ln\left(\frac{\pi_i}{1-\pi_i}\right)}_{\text{logit}}$$

What is unknown in our model?

Parameter estimation with maximum likelihood (ML)

$\beta_{(p \times 1)}$ $\qquad n_i = 1$

1) Likelihood $L(\beta) = \prod_{i=1}^{n} \pi_i^{y_i} (1-\pi_i)^{1-y_i}$

$\downarrow \ln$

2) Log-likelihood $\ell(\beta) = \sum_{i=1}^{n} \left(y_i \ln\left(\frac{\pi_i}{1-\pi_i}\right) + \ln(1-\pi_i) \right)$

$$= \sum_{i=1}^{n} \left(y_i \eta_i - \ln(1+\exp(\eta_i)) \right)$$

$\downarrow \frac{\partial \ell(\beta)}{\partial \beta} \ (p \times 1)$

3) Score function $s(\beta) = \sum_{i=1}^{n} x_i \ (y_i - \pi_i)$

$\pi_i = \left(\frac{e^{\eta_i}}{1+e^{\eta_i}}\right)$

$E(s(\beta)) = 0$ $\qquad$ 4) $\underset{p \times p}{\mathrm{Cov}}(s(\beta))$

$\qquad = F(\beta) =$ expected Fisher information matrix

$$F(\beta) = \sum_{i=1}^{n} \underbrace{x_i x_i^T}_{p \times p} \underbrace{\pi_i (1-\pi_i)}_{1 \times 1} = E\left(-\frac{\partial^2 \ell(\beta)}{\partial \beta \partial \beta^T}\right)$$

5) Observed Fisher information matrix $H(\beta)$

$$H(\beta) = -\frac{\partial^2 \ell}{\partial \beta \partial \beta^T} = \sum_{i=1}^{n} x_i x_i^T \pi_i (1-\pi_i)$$

1) Likelihood: (X_i, Y_i) independent pairs

$$L(\beta) = \prod_{i=1}^{n} f(y_i;\beta) = \prod_{i=1}^{n} \underbrace{\pi_i^{y_i}(1-\pi_i)^{1-y_i}}_{L_i(\beta)}$$

independent Y_i's

2) Log-likelihood

sum of individual contributions

$$\ell(\beta) = \ln L(\beta) = \sum_{i=1}^{n} \ln L_i(\beta) = \sum_{i=1}^{n} \ell_i(\beta)$$

$$\ell_i(\beta) = y_i \ln \pi_i + (1-y_i)\ln(1-\pi_i)$$

$$= y_i \underbrace{\ln\left(\frac{\pi_i}{1-\pi_i}\right)}_{\eta_i} + \ln(1-\pi_i)$$

$$\ln\left(\frac{1}{1+e^{\eta_i}}\right)$$

$$= y_i\eta_i - \ln(1+e^{\eta_i})$$

3) Score function:

sum of individual contributions

$$s(\beta) = \frac{\partial \ell(\beta)}{\partial \beta} = \sum_{i=1}^{n} \frac{\partial \ell_i(\beta)}{\partial \beta} = \sum_{i=1}^{n} s_i(\beta)$$

$p \times 1$

$$\begin{bmatrix} \frac{\partial \ell_i}{\partial \beta[1]} \\ \frac{\partial \ell_i}{\partial \beta[2]} \\ \vdots \\ \frac{\partial \ell_i}{\partial \beta[p]} \end{bmatrix}$$

$$s_i(\beta) = \frac{\partial l_i(\beta)}{\partial \beta} = \frac{\partial l_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta}$$

$$\frac{\partial}{\partial \eta_i}\left[y_i \eta_i - \ln(1+\exp(\eta_i)) \right]$$

$$\frac{\partial}{\partial \beta}(x_i^T \beta) = x_i$$

$p \times 1$

$$= y_i - \frac{1}{1+e^{\eta_i}} \cdot e^{\eta_i}$$

$$= \Big(y_i - \underbrace{\frac{e^{\eta_i}}{1+e^{\eta_i}}}_{\pi_i} \Big)$$

$$s_i(\beta) = (y_i - \pi_i) \cdot x_i$$

$$\underset{p\times 1}{s(\beta)} = \sum_{i=1}^{n} (y_i - \pi_i) \cdot \underset{p\times 1}{x_i}$$

$$= \sum_{i=1}^{n} \left(y_i - \frac{\exp(x_i^T \beta)}{1+\exp(x_i^T \beta)} \right) x_i$$

Properties of $s(\beta)$

$S(\beta)$ is a function of Y_i and therefore a random vector.

NB $E(s_i(\beta)) = 0$

$$E(s(\beta)) = E\left(\sum_{i=1}^{n} (Y_i - \pi_i) \cdot x_i \right)$$

$$= \sum_{i=1}^{n} E\left((Y_i - \pi_i) \cdot x_i \right) = \sum_{i=1}^{n} (\underbrace{E(Y_i)}_{\pi_i} - \pi_i)\, x_i = \underline{\underline{0}}$$

4) $$\underset{F(\beta)}{\operatorname{Cov}}(s(\beta)) = \operatorname{Cov}\left(\sum_{i=1}^{n} s_i(\beta)\right)$$

$s_i(\beta) = (Y_i - \pi_i)\, x_i$

$$\underset{\substack{\uparrow \\ \text{Indep. pairs}}}{=} \sum_{i=1}^{n} \operatorname{Cov}(s_i(\beta)) = \sum_{i=1}^{n} E\left(\left(s_i(\beta) - \underbrace{E(s_i(\beta))}_{0}\right) \left(s_i(\beta) - \underbrace{E(s_i(\beta))}_{0}\right)^T \right)$$

$$= \sum_{i=1}^{n} \underbrace{E\left(s_i(\beta)\, s_i(\beta)^T \right)}_{F_i(\beta)} = \sum_{i=1}^{n} F_i(\beta)$$

$$F_i(\beta) = E\left(s_i(\beta)\, s_i(\beta)^T \right) = E\left(\underset{1\times 1}{(Y_i - \pi_i)}\, \underset{p\times 1}{x_i} \cdot \underset{1\times 1}{(Y_i - \pi_i)}\, \underset{1\times p}{x_i^T} \right)$$

$$= x_i x_i^T \underbrace{E\left((Y_i - \pi_i)^2 \right)}_{\operatorname{Var}(Y_i) = \pi_i(1-\pi_i)} = x_i x_i^T \pi_i (1 - \pi_i)$$

$$F(\beta) = \sum_{i=1}^{n} x_i x_i^T \pi_i (1 - \pi_i)$$

Useful relationship: Under mild regularity conditions (change $\int$ and $\frac{\partial}{\partial\beta}$):

$$F(\beta) = E\left(- \underbrace{\frac{\partial^2 \ell(\beta)}{\partial\beta\,\partial\beta^T}}_{\text{Hessian matrix of } \ell} \right)$$

5) Observed Fisher information matrix $H(\beta)$

$$H(\beta) = -\frac{\partial^2 \ell(\beta)}{\partial\beta\,\partial\beta^T} = -\frac{\partial\, s(\beta)}{\partial\beta^T} = \frac{\partial}{\partial\beta^T}\Big[\sum_{i=1}^n (\underset{\uparrow\,\beta}{\pi_i} - y_i)\, x_i\Big]$$

$$-s(\beta) = \sum_{i=1}^n (\pi_i - y_i)\, x_i$$

$$H(\beta) = \sum_{i=1}^n \frac{\partial}{\partial\beta^T}\left[x_i \pi_i - x_i y_i\right] = \sum_{i=1}^n \frac{\partial}{\partial\beta^T} x_i \pi_i$$

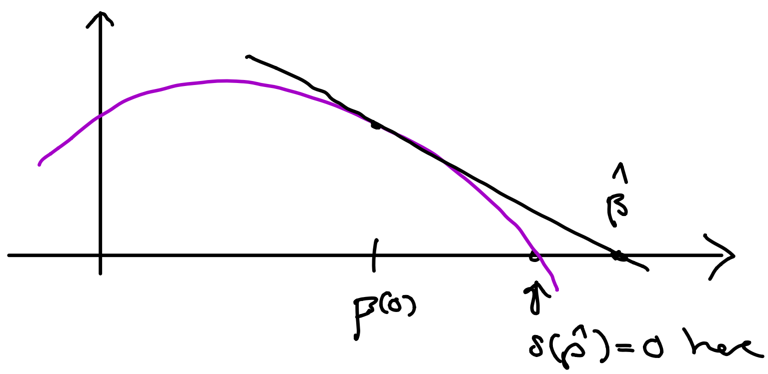
$$= \sum_{i=1}^n x_i \cdot \frac{\partial\pi_i}{\partial\eta_i}\,\frac{\partial\eta_i}{\partial\beta} \qquad \frac{\partial\, x_i^T\beta}{\partial\beta^T} = \left(\frac{\partial\, x_i^T\beta}{\partial\beta}\right)^T = x_i^T$$

$$\frac{\partial\left(\frac{e^{\eta_i}}{1+e^{\eta_i}}\right)}{\partial\eta_i} \overset{\text{book}}{=} \ldots = \pi_i(1-\pi_i)$$

$$H(\beta) = \sum_{i=1}^{n} x_i \underbrace{\pi_i(1-\pi_i)}_{1\times 1} x_i^T = \underline{\underline{\sum_{i=1}^{n} x_i x_i^T \pi_i(1-\pi_i)}}$$

Parameter estimation in practice

$$S(\hat\beta) \approx \underbrace{S(\beta^{(0)}) + \overbrace{\boxed{\left.\frac{\partial S(\beta)}{\partial \beta^T}\right|_{\beta=\beta^{(0)}}}}^{-H(\beta^{(0)})} (\hat\beta - \beta^{(0)})}_{\text{approx } S(\hat\beta)}$$



$S(\hat\beta) = 0 \iff$ solve for $\hat\beta$ ...

$$S(\beta^{(0)}) - H(\beta^{(0)})(\hat\beta - \beta^{(0)}) = 0$$

$$H(\beta^{(0)})\hat\beta = S(\beta^{(0)}) + H(\beta^{(0)})\beta^{(0)}$$

$$\underline{\hat\beta = \beta_0 + H(\beta^{(0)})^{-1} S(\beta^{(0)})}$$

$$\beta^{(t+1)} = \beta^{(t)} + H(\beta^{(t)})^{-1} S(\beta^{(t)})$$

Newton – Raphson

Substitute H with F → Fisher-scoring

Now: convergence & properties of the estimator on Module page

Reminder: Newton's method for solving $g(z)=0$

$$z_{t+1} = z_t - \frac{g(z_t)}{g'(z_t)}$$