Likelihood theory for logistic regression

$$Y_i = \begin{cases} 1 & \text{w prob } p_i \\ 0 & \text{w prob } 1-p_i \end{cases} \qquad p_i = p_i(\alpha, \Delta, \beta)$$

$$\underbrace{\prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{(1-y_i)}}_{} = \prod_{i=1}^{n} f(y_i) \qquad f(\cdot) \text{ density}$$

generic $\qquad$ fn for y_i

Prob(getting data $y_1, \dots, y_n$)

$= L(\alpha, \beta, \Delta; y_1, \dots, y_n)$ likelihood function Ch 4.1–4.4

$\hat\alpha, \hat\beta, \hat\Delta$ defined to maximize $L(\alpha, \beta, \Delta; y)$ makes our data "most probable"

Solution: $\quad \ell(\alpha, \beta, \Delta) = \log L(\alpha, \beta, \Delta)$

$$\frac{\partial \ell}{\partial \alpha}(\hat\alpha, \hat\beta, \hat\Delta) = 0 \; ; \; \frac{\partial \ell}{\partial \beta}(\hat\alpha, \hat\beta, \hat\Delta) \overset{\text{''}0}{\;} \; ; \; \frac{\partial \ell}{\partial \Delta}(\hat\alpha, \hat\beta, \hat\Delta) = 0$$

$$\boxed{\ell(\alpha,\beta,\Delta) = \log \prod_{i=1}^{n} \left(\frac{e^{\alpha+\Delta z_i+\beta(x_i-\bar{x})}}{1+e^{\alpha+\Delta z_i+\beta(x_i-\bar{x})}} \right)^{y_i} \left(\frac{1}{1+e^{\alpha+\Delta z_i+\beta(x_i-\bar{x})}} \right)^{1-y_i}}$$

ntbc

$$\boxed{= \sum_{i=1}^{n} \left\{ \alpha y_i + \Delta z_i y_i + \beta(x_i-\bar{x}) y_i - \log\left(1+e^{\alpha+\Delta z_i+\ldots}\right) \right\}}$$

$$\frac{\partial \ell}{\partial \alpha} \quad \frac{\partial \ell}{\partial \beta} \quad \frac{\partial \ell}{\partial \Delta} = 0$$

requires iteration (automatic in R) (Fisher scoring)

Back to general case

$$L(\theta; y) = \prod_{i=1}^{n} f(y_i; \theta)$$

y_i's independent

$$\log L = \ell(\theta; y) = \sum_{i=1}^{n} \log f(y_i; \theta) \qquad \ell'(\hat{\theta}) = 0$$

solve

$$\hat{\theta} = \hat{\theta}(y)$$

By considering the dist^n of $y_1, \ldots, y_n$, can get dist^n of $\hat{\theta}$

Lots of theory ... , conclusion

1) $$\underset{\sim}{\hat\theta} \;\dot\sim\; N_p\left(\underset{\sim}{\theta},\, \underline{j^{-1}(\hat\theta)}\right)$$

where $$j(\hat\theta) = -\,\partial^2 \ell(\theta; y)/\partial\theta\,\partial\theta^T \Big|_{\theta=\hat\theta}$$

2) $$2\{\ell(\underset{\sim}{\hat\theta}) - \ell(\underset{\sim}{\theta}_0)\} \;\dot\sim\; \chi^2_p \quad \text{if } \underset{\sim}{\theta} = \underset{\sim}{\theta}_0$$

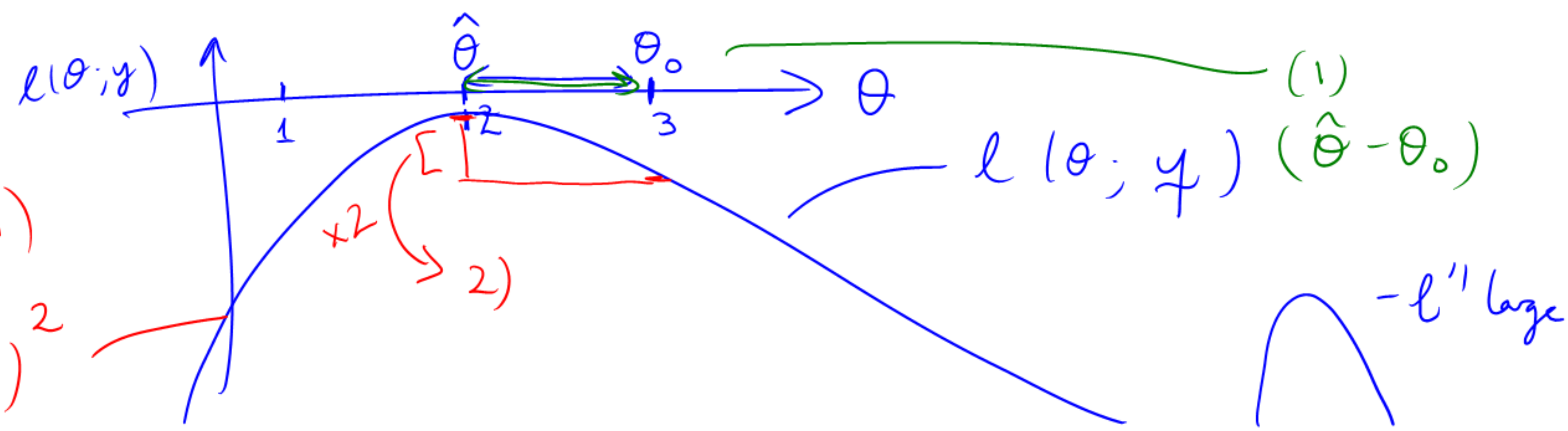
where $\underset{\sim}{\theta} = (\theta_1, \dots, \theta_p)$ $\quad \theta_0 = (\theta_{01}, \dots, \theta_{0p})$

$$= 2\log\left\{\frac{L(\hat\theta)}{L(\theta_0)}\right\} = \log \text{ likelihood ratio statistic}$$

Wilk's statistic

$p = 1$

θ scalar



Normal dist $= (\theta, 1)$

$\ell = (y - \theta)^2$

If $j(\hat{\theta})$ is large, $\hat{\theta}$ well-determined

a.var$(\hat{\theta})$ small

If $j(\hat{\theta})$ is small, log-lik is flat near top

a.var$(\hat{\theta})$ large

In R "residual deviance" $= -2\log L(\text{maximum})$

So 2 different fits can be compared using χ^2 approximation to $2\{\ell(\hat{\theta}) - \ell(\tilde{\theta})\}$

$$L(\theta; y) = \prod_{i=1}^{n} f(y_i; \theta)$$

actual value of $L(\hat{\theta}; y)$ doesn't have an interpretation

not invariant to 1-1 transformations from y to $x = g(y)$

$$\frac{L(\theta_1; y)}{L(\theta_2; y)} \quad \text{is invariant}$$

?? $f(y) = f(g(x))\underbrace{|g'(x)|^{-1}}_{\text{Jacobian}}$??

This all holds, almost without change if $y_i \sim N(\mu_i, \sigma^2)$

with $\mu_i = \alpha + \Delta z_i + \beta(x_i - \bar{x})$

$L(\alpha, \Delta, \beta)$ is now $\displaystyle\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(y_i - \alpha - \Delta z_i - \beta(x_i - \bar{x}))^2}$

$$\ell(\alpha, \Delta, \beta) = -n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \alpha - \Delta z_i - \beta(x_i - \bar{x}))^2$$

$\hat{\alpha}, \hat{\Delta}, \hat{\beta}$ max. lik. est. = least squares estimators

In R, use glm instead of lm

$$\text{glm}(\overset{\text{fault}}{y} \sim I(\text{purity} - \text{mean}(\text{purity})) + \text{factor}(\text{process}),$$
$$\text{family} = \text{binomial})$$

?glm family = binomial, poisson, gamma, normal

default

$$y_i \sim \text{Bin}(m_i, p_i) \Rightarrow f(y_i) = \binom{m_i}{y_i} p_i^{y_i} (1-p_i)^{m_i - y_i} \quad y_i = 0, 1, \ldots, m_i$$

$$y_i \sim \text{Ber}(p_i) \qquad f(y_i) = p_i^{y_i} (1-p_i)^{1-y_i} \quad y_i = 0, 1$$

$$\text{Bin}(1, p_i)$$

glm(y ~ , family = binomial)

if y is numeric, assumed to be s_i/m_i # succ. / # trials if y = factor var. w/ 2 levels, or if $y = 0, 1$,

assumed to be binary

If y is a 2-column matrix, then the cols

are (# successes, # failures)

Example 10.18 data called "nodal" p. 491

m	r (# "successes")	age	stage	grade	xray	acid
6	5	0	1	1	1	1
6	1	0	0	0	0	1
⋮						

23 rows each Binomial(m_i, p_i) obs'd value r_i

But on HO, and in > library(boot)

> data(nodal)

e.g.

1st 6 rows

m	r	aged	stage	..		
1	1	0	1	1	1	1
1	1	0	1	1	1	1
1	1	0	1	1	1	1
1	1	0	1	1	1	1
1	1	0	1	1	1	1
1	0	0	1	1	1	1

53 rows – i.e. 53 patients, each with their own entry

covariates are: age, stage, grade, xray, acid

age	stage	grade	xray	acid	
0 <60	0 less serious	0	0	0	<0.6 serum level
1 >60	1 more serious	1	1	1	>0.6

Text p. 491 deviance ✗ 19.64 on 49 degrees of freedom
49.18 ✓

using binary regression.

Other examples: HW 2 Qu 3 $\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$

Table 1.3 & Fig 1.3 $\text{logit}\, p_i = \beta_0 + \beta_1 x_i$

temp

independent, Bernoulli or binomial,

$\prod$ $p_i^{y_i}(1-p_i)^{(1-y_i)}$ $\log\left(\frac{p_i}{1-p_i}\right) = x_i^T\beta$

or

$\binom{m_i}{y_i} p_i^{y_i}(1-p_i)^{(m_i-y_i)}$