Estimation (maximum likelihood estimator)

~~like~~ likelihood function (β)

① Ungrouped data $\quad y_i \sim \text{Bernoulli}(P_i) \quad \ln\left(\frac{P_i}{1-P_i}\right) = \underset{\sim}{x_i}^T \beta$

$$P_i^{y_i}(1-P_i)^{1-y_i}$$

$$L(\beta) = \prod_{i=1}^{n}\left(\frac{\exp(\underset{\sim}{x_i}^T\beta)}{1+\exp(\underset{\sim}{x_i}^T\beta)}\right)^{y_i}\left(\frac{1}{1+\exp(\underset{\sim}{x_i}^T\beta)}\right)^{1-y_i} \qquad y_i = 1, 0$$

$$= \prod_{i=1}^{n_1}\left(\frac{\exp(\underset{\sim}{x_i}^T\beta)}{1+\exp(\underset{\sim}{x_i}^T\beta)}\right)\prod_{i=n_1+1}^{n}\left(\frac{1}{\exp(\underset{\sim}{x_i}^T\beta)}\right) \qquad n_1 = \#\text{ of } (y_i = 1)$$

② grouped data $\quad r_i \sim \text{Binomial}(n_i, P_i)$

$$L(\beta) = \prod_{i=1}^{n}\binom{n_i}{r_i}\left(\frac{\exp(\underset{\sim}{x_i}^T\beta)}{1+\exp(\underset{\sim}{x_i}^T\beta)}\right)^{r_i}\left(\frac{1}{1+\exp(\underset{\sim}{x_i}^T\beta)}\right)^{n_i-r_i}$$

$\Rightarrow$ m.l.e. of β

Note that For linear regression, $\hat{\beta} = (\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T\underset{\sim}{Y}$

$$L(\beta) = \prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi}\sqrt{\sigma^2}}\right)\exp\left\{-\frac{\sum_{i=1}^{n}(y_i - \underset{\sim}{x_i}^T\beta)^2}{2\sigma^2}\right\}$$

$$= (2\pi)^{-\frac{n}{2}}(\sigma^2)^{-\frac{n}{2}}\exp\left\{-\boxed{\frac{\sum_{i=1}^{n}(y_i - \underset{\sim}{x_i}^T\beta)^2}{2\sigma^2}}\right\}$$

$\Downarrow$ min For fixed σ^2

l.s.e ~~min~~ ~~l.s.e~~.

mle = lse $\leftarrow$ linear regression

~~r.v.~~ $e \sim N$

logistic regression $\quad$ mle $\neq$ lse
of β $\quad$ of β

For grouped data, if n_i large $\forall i$

We can weighted least squares to handle the model of $\log\left(\frac{\hat{P}_i}{1-\hat{P}_i}\right)$ on x_1 —— , x_p with weight $W_i = n_i\hat{P}_i(1-\hat{P}_i)$

MATH 3423 ① $\hat{\theta}_{mle}$ – asymptotically ~~no~~ unbiased $\quad n \to \infty$

full efficient

sufficient

normally distributed

$$\hat{\underset{\sim}{\beta}}_{p' \times 1} \,\dot{\sim}\, MN(\underset{\sim}{\beta}, \underset{p' \times p'}{Var(\hat{\underset{\sim}{\beta}})})$$

$$\begin{pmatrix} Var(\hat{\beta}_i) & & cov(\hat{\beta}_i, \hat{\beta}_j) \\ & \ddots & \\ & & Var(\hat{\beta}_p) \end{pmatrix}$$

② CRLB

$$Var(\hat{\theta}_{unbiased}) \geq -\frac{1}{nE\left(\frac{\partial^2}{\partial\theta^2} \log f_{X_i}(x_i, \theta)\right)}$$

OR

$$\geq -\frac{1}{E\left(\frac{\partial^2}{\partial\theta^2} \log f_{\underset{\sim}{X}}(\underset{\sim}{x}, \theta)\right)}$$

Define $\underset{p' \times p'}{\underset{\sim}{C}}$ = observed Fisher Information matrix

$$C_{ii} = -\frac{\partial^2}{\partial\hat{\beta}_i^2} \log L(\hat{\underset{\sim}{\beta}}) \qquad i = 0, \dots, p$$

$$C_{ij} = -\frac{\partial^2}{\partial\hat{\beta}_i \partial\hat{\beta}_j} \log L(\hat{\underset{\sim}{\beta}}) \qquad j = 0, \dots, p$$

$$Var(\hat{\underset{\sim}{\beta}}) = \underset{p' \times p'}{\underset{\sim}{C}^{-1}} = \begin{pmatrix} c^{00} & & c^{ij} \\ & \ddots & \\ & & c^{pp} \end{pmatrix}$$

Inverse of Fisher Information matrix

$$\hat{\beta}_i \,\dot{\sim}\, N(\beta_i, c^{ii})$$

$\Rightarrow$ ① C.I. of $\beta_i = \hat{\beta}_i \pm z_{\alpha/2} \sqrt{c^{ii}}$ $\quad$ Wald C.I.

② Hypothesis testing $\quad H_0 : \beta_i = 0$ vs $H_a : \beta_i \neq 0$

$$z_{obs} = \frac{\hat{\beta}_i}{\sqrt{c^{ii}}} \quad \text{OR} \quad \left(\frac{\hat{\beta}_i}{\sqrt{c^{ii}}}\right)^2 \dot{\sim}\, \chi^2_{(1)}$$

Reject H_0 if $|z_{obs}| > z_{\alpha/2}$

Wald test

Example 1

Wald test

$$\left(\frac{\hat{\beta}_i}{\sqrt{c^{ii}}}\right)^2 \dot{\sim} \chi^2_{(1)}$$

$i=0, \rightarrow p$

Parameter estimates with confidence interval

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-5.5784	0.3682	229.4877	<.0001
logload	1	1.1400	0.0893	163.0932	<.0001

← p-value

← $H_0: \beta_0 = 0$

← $H_0: \beta_1 = 0$

(a) Fitted line

(*) If p-value is small (<0.05) Reject H_0

Covariance matrix

$\underset{\sim}{C}^{-1}$

Estimated Covariance Matrix

Parameter	Intercept	logload
Intercept	0.1356	-0.03253
logload	-0.03253	0.007968

$Var(\hat{\beta}_0)$

$cov(\hat{\beta}_0, \hat{\beta}_1)$

$Var(\hat{\beta}_1)$

$cov(\hat{\beta}_1, \hat{\beta}_0)$

(a) Fitted line

$$\log_e\left(\frac{\hat{p}}{1-\hat{p}}\right) = -5.5784 + 1.14 * \log load$$

Or

$$\hat{p} = \frac{\exp(-5.5784 + 1.14 * \log load)}{1+\exp(-5.5784 + 1.14 * \log load)}$$

(b) Find the 95% C.I. of unknown parameters

95% C.I. for β_0

$$\Rightarrow \hat{\beta}_0 \pm \underbrace{\boxed{z_{0.025}}}_{=1.96} \sqrt{0.1356}$$

$$\Rightarrow -5.5784 \pm 1.96\sqrt{0.1356}$$

$$\Rightarrow (-6.3001, -4.8567)$$

95% C.I. for β_1

$$\Rightarrow \hat{\beta}_1 \pm 1.96\sqrt{c^{11}}$$

$$\Rightarrow 1.14 \pm 1.96\sqrt{0.007968}$$

$$\Rightarrow (0.965, 1.3149)$$

(c) Odds ratio ratio of two odds when logload increases one unit

$P/Q = P/(1-P)$ $\ln\left(\frac{p_i}{1-p_i}\right) = \underset{\sim}{x}_i^T \underset{\sim}{\beta}$

$$\text{Odds ratio} = \frac{\frac{P_{a+1}}{1-P_{a+1}} \Big|\ \text{odds at } x=a+1}{\frac{P_a}{1-P_a} \Big|\ \text{odds at } x=a}$$

$\Rightarrow P_i = \dfrac{\exp(\underset{\sim}{x}_i^T \underset{\sim}{\beta})}{1+\exp(\underset{\sim}{x}_i^T \underset{\sim}{\beta})}$ prob. at $\underset{\sim}{x}_i$

$$= \frac{\exp(\beta_0 + \beta_1 * (a+1))}{\exp(\beta_0 + \beta_1 * a)}$$

odds $= \Rightarrow \dfrac{P_i}{1-P_i} = \dfrac{\frac{\exp(\underset{\sim}{x}_i^T \underset{\sim}{\beta})}{1+\exp(\underset{\sim}{x}_i^T \underset{\sim}{\beta})}}{\frac{1}{1+\exp(\underset{\sim}{x}_i^T \underset{\sim}{\beta})}} = \exp(\underset{\sim}{x}_i^T \underset{\sim}{\beta})$

$$= \exp(\beta_1)$$

OR, $\Rightarrow \beta_1 = \ln(\text{odds ratio})$ when x increases one unit

$\hat{\beta}_1 = 1.14$ $\exp(\hat{\beta}_1) = \exp(1.14) = 3.127$

mle By invariance prop. $\exp(\hat{\beta}_1)$ – m.l.e. of $\exp(\beta_1)$

OR

From C.I. of β_1

95% C.I. of odds ratio

$\approx (\exp(0.965), \exp(1.3149))$

$= (2.625, 3.724)$

C.I. $\widehat{\text{odds ratio}} \pm 1.96 * \text{s.e.}$

Delta method

$\text{Var}(\widehat{\text{odds ratio}}) \approx (\exp(\hat{\beta}_1))^2 * \text{Var}(\hat{\beta}_1)$

(d) Prob at logload = 4

$$= \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 * 4)}{1+\exp(\hat{\beta}_0 + \hat{\beta}_1 * 4)} = \frac{\exp(-5.5784 + 1.14 * 4)}{1+\exp(-5.5784 + 1.14 * 4)}$$

$= 0.2653$

$g(\theta) = \dfrac{\exp(\theta)}{1+\exp(\theta)}$

$\hat{\theta}$ $g(\hat{\beta}_0 + \hat{\beta}_1 * 4)$

C.I. $\widehat{\text{Prob}} \pm 1.96 * \text{s.e.}$

$\text{Var}(\boxed{\hat{\beta}_0 + \hat{\beta}_1 * 4})$ $\hat{\theta}$ $\text{Var}(g(\boxed{\hat{\beta}_0 + \hat{\beta}_1 * 4}))$

$= \text{Var}(\hat{\beta}_0) + 16\,\text{Var}(\hat{\beta}_1) + 8\,\text{cov}(\hat{\beta}_0, \hat{\beta}_1)$ $\approx (g'(\hat{\theta}))^2\, \text{Var}(\hat{\theta})$

$= 0.1356 + 16 * 0.007968 - 8 * 0.03253 = 0.002848$

OR (a) To find ~~C.I.~~ C.I. of $\beta_0 + 4\beta_1$

$$\Rightarrow (\hat{\beta}_0 + 4\hat{\beta}_1) \pm 1.96\sqrt{Var(\hat{\beta}_0 + 4\hat{\beta}_1)}$$

$$\Rightarrow (-1.123, -0.9138)$$

(b) 95% C.I. of Prob. at logload = 4

$$\frac{\exp(\beta_0 + 4\beta_1)}{1+\exp(\beta_0 + 4\beta_1)}$$

$$\approx \left(\frac{\exp(-1.1233)}{1+\exp(-1.1233)}, \frac{\exp(-0.9138)}{1+\exp(-0.9138)} \right)$$

$$= (0.2455, 0.2862)$$

(e) Estimate odds at logload = 4 & find the 95% C.I.

$$\exp(\boxed{\beta_0 + \beta_1 * 4})$$

$$\text{est. of odds at logload} = 4 = \exp(\hat{\beta}_0 + \hat{\beta}_1 * 4)$$

$$= \exp(-5.5784 + 1.14 * 4)$$

$$= 0.3611$$

95% C.I. of $\beta_0 + 4\beta_1$)

$$= (-1.123, -0.9138)$$

95% C.I. of ~~ode~~ odds at logload = 4

$$\approx (\exp(-1.123), \exp(-0.9138))$$

$$= (0.3253, 0.4009)$$

C.I

$\widehat{odds} \pm 1.96 * SE$

$Var(\widehat{odds})$

$\approx (\exp(\hat{\theta}))^2 *$

$Var(\hat{\theta})$

MATH3423 asymptotic likelihood ratio test $n \to \infty$

$$\text{likelihood ratio} = \frac{\sup_{\theta \in \Theta_0} \{L(\underset{\sim}{\theta})\}}{\sup_{\theta \in \Theta} \{L(\underset{\sim}{\theta})\}}$$

asymptotic dist. of likelihood ratio

$$\Rightarrow -2 \log [\text{likelihood ratio}] \;\dot{\sim}\; \chi^2_r$$

$r = \#$ of free para. in Θ − # of free para in Θ_0

$\Rightarrow$ Reject H_0 if

$$-2 \log [\text{likelihood ratio}] \geq \chi^2_{\alpha, r}$$

Hypothesis testing Assume: $y_i \sim \text{Bernoulli}(P_i)$

OR $Y_i \sim \text{Binomial}(n_i, P_i)$

where $P_i = \dfrac{\exp(\underset{\sim}{x}_i^T \underset{\sim}{\beta})}{1+\exp(\underset{\sim}{x}_i^T \underset{\sim}{\beta})}$

$$H_0 : P_i = \frac{\exp(\underset{\sim}{x}_i^T \underset{\sim}{\beta})}{1+\exp(\underset{\sim}{x}_i^T \underset{\sim}{\beta})}$$

OR $H_0 : \ln\left(\dfrac{P_i}{1-P_i}\right) = \underset{\sim}{x}_i^T \underset{\sim}{\beta}$

If reject H_0 $\Rightarrow$ make transformation on x

$\Rightarrow$ ~~Overf~~ Overdispersion

① Ungrouped data $y_i \sim \text{Bernoulli}(P_i)$

$$\cancel{L(\underset{\sim}{P})} = L(\underset{\sim}{P}) = \prod_{i=1}^{n_1} P_i \prod_{i=n_1+1}^{n} (1-P_i) \qquad n_1 = \#\text{ of }(y_i = 1)$$

Numerator $\sup_{\underset{\sim}{\theta} \in \Theta_0} \{L(\underset{\sim}{\theta})\} = \prod_{i=1}^{n_1} \dfrac{\exp(\underset{\sim}{x}_i^T \hat{\underset{\sim}{\beta}})}{1+\exp(\underset{\sim}{x}_i^T \hat{\underset{\sim}{\beta}})} \prod_{i=n_1+1}^{n} \dfrac{1}{1+\exp(\underset{\sim}{x}_i^T \hat{\underset{\sim}{\beta}})}$

<u>Denominator</u> $\sup_{\underset{\sim}{\theta} \in \textcircled{H}} \{L(\underset{\sim}{\theta})\}$ $\quad \underset{\sim}{\theta} = (P_1, \dots, P_n)$

$$= \prod_{i=1}^{n_1} y_i^{y_i} \prod_{i=n_1+1}^{n} (1-y_i)^{1-y_i} \qquad y_i = 1 \quad \hat{P}_i = 1$$

$$y_i = 0 \quad \hat{P}_i = 0$$

$= 1$ if all x_i are distinct

$-2 \log\{\text{Likelihood ratio})$

$$= -2 \log \left\{ \frac{L(\hat{\underset{\sim}{\beta}})}{L(\hat{\underset{\sim}{\beta}})} \right\} \overset{\text{def}}{=} \lambda(\underset{\sim}{\beta}) \text{ — deviance}$$

$$\dot{\sim}\ \chi^2_r$$

$$r = \underbrace{\# \text{ of free para in } \textcircled{H}}_{n} - \underbrace{\# \text{ of free para in } \textcircled{H}_0}_{p'}$$

n if ~~xi~~ all x_i are distinct

e.g.

x	y
x_a	1
x_a	0

$\Rightarrow \hat{P}_i = \frac{1}{2}$

$\Rightarrow$ # of ~~free~~ free para in $\textcircled{H} = n-1$

② <u>Grouped data</u> $\quad r_i \sim \text{Binomial}(n_i, P_i) \quad i = 1, \dots, \textcircled{S}$ # of groups

$$\text{Likelihood function} = \prod_{i=1}^{S} \binom{n_i}{r_i} P_i^{r_i} (1-P_i)^{n_i - r_i}$$

<u>Numerator</u> $\sup_{\underset{\sim}{\theta} \in \textcircled{H}_0} \{L(\underset{\sim}{\theta})\}$

$$L(\hat{\underset{\sim}{\beta}}) = \prod_{i=1}^{S} \left(\frac{\exp(\underset{\sim}{x}_i^T \hat{\underset{\sim}{\beta}})}{1+\exp(\underset{\sim}{x}_i^T \hat{\underset{\sim}{\beta}})} \right)^{r_i} \left(\frac{1}{1+\exp(\underset{\sim}{x}_i^T \hat{\underset{\sim}{\beta}})} \right)^{n_i - r_i} \binom{n_i}{r_i}$$

<u>Denominator</u> $\sup_{\underset{\sim}{\theta} \in \Theta} \{ L(\underset{\sim}{\theta}) \}$

$\underset{\sim}{\theta} = (P_1, \text{---}, P_s)$

$\hat{P}_i = \frac{r_i}{n_i} \quad i = 1, \text{---}, s$

$$L(\hat{\underset{\sim}{P}}) = \prod_{i=1}^{s} \left(\frac{r_i}{n_i}\right)^{r_i} \left(1 - \frac{r_i}{n_i}\right)^{n_i - r_i} \binom{n_i}{r_i}$$

$$-2 \log \left\{ \frac{L(\hat{\underset{\sim}{\beta}})}{L(\hat{\underset{\sim}{P}})} \right\} \overset{\circ}{\sim} \chi^2_r$$

r = # of free param in Θ − # of free param in Θ_0

$= s - p'$

$$\text{Pearson Chi-square} = \sum_{i=1}^{s} \frac{(r_i - n_i \hat{P}_i)^2}{n_i \hat{P}_i (1 - \hat{P}_i)}$$

$$\sum_{i=1}^{s} \left[\frac{(r_i - n_i \hat{P}_i)^2}{n_i \hat{P}_i} + \frac{(n_i - r_i - n_i(1 - \hat{P}_i))^2}{n_i (1 - \hat{P}_i)} \right]$$

x =load

Deviance and Pearson Goodness-of-Fit Statistics				
Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	36.2181	3	12.0727	<.0001
Pearson	34.3607	3	11.4536	<.0001

$H_0: p_i = \frac{\exp(x_i^T \beta)}{1+\exp(x_i^T \beta)}$

Reject H_0

$\Rightarrow$ transformation of x

naive if $\frac{\max(x_i)}{\min(x_i)} > 10$

then make log-transformation

x =logload

Deviance and Pearson Goodness-of-Fit Statistics				
Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	5.3883	3	1.7961	0.1455
Pearson	5.3792	3	1.7931	0.1460

Fit a logistic regression ~~of~~ on logload

$$\Rightarrow \log_e\left(\frac{\hat{p}}{1-\hat{p}}\right) = -5.5784 + 1.14 \times \text{logload}$$

$$\frac{L(\hat{\beta})}{L(\hat{p})} = \frac{\prod_{i=1}^{s}\left(\frac{\exp(x_i^T\hat{\beta})}{1+\exp(x_i^T\hat{\beta})}\right)^{r_i}\left(\frac{1}{1+\exp(x_i^T\hat{\beta})}\right)^{n_i-r_i}}{\prod_{i=1}^{s}\left(\frac{r_i}{n_i}\right)^{r_i}\left(\frac{n_i-r_i}{n_i}\right)^{n_i-r_i}}$$