MLDS 401/IEMS 404-1 (Fall 2023): Lab 6 Solution – 11/07/2023

7.1 (2×2 **contingency table):**

(a) The logistic regression models for $x = 0$ and $x = 1$ are

$$\ln\left(\frac{p_0}{1-p_0}\right) = \beta_0 \quad \text{and} \quad \ln\left(\frac{p_1}{1-p_1}\right) = \beta_0 + \beta_1.$$

Hence

$$\beta_1 = \ln\left(\frac{p_1}{1-p_1}\right) - \ln\left(\frac{p_0}{1-p_0}\right) = \ln\left\{[p_1/(1-p_1)]/[p_0/(1-p_0)]\right\}.$$

The MLEs $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are found by substituting the MLEs $\widehat{p}_0 = s_0/n_0$ and $\widehat{p}_1 = s_1/n_1$ in the above equations and solving for $\widehat{\beta}_0$ and $\widehat{\beta}_1$. Hence

$$\widehat{\beta}_1 = \ln \widehat{\psi} = \ln\left\{[\widehat{p}_1/(1-\widehat{p}_1)]/[\widehat{p}_0/(1-\widehat{p}_0)]\right\}.$$

(b) The information matrix equals

$$\boldsymbol{\mathcal{I}} = \begin{bmatrix} n_0 p_0 q_0 + n_1 p_1 q_1 & n_1 p_1 q_1 \\ n_1 p_1 q_1 & n_1 p_1 q_1 \end{bmatrix},$$

where $q_i = 1 - p_i \ \ (i = 0, 1)$. The inverse of this matrix equals

$$\boldsymbol{\mathcal{I}}^{-1} = \frac{1}{n_0 p_0 q_0 n_1 p_1 q_1} \begin{bmatrix} n_1 p_1 q_1 & -n_1 p_1 q_1 \\ -n_1 p_1 q_1 & n_0 p_0 q_0 + n_1 p_1 q_1 \end{bmatrix}.$$

Hence

$$\text{Var}(\ln \widehat{\psi}) = \frac{n_0 p_0 q_0 + n_1 p_1 q_1}{n_0 p_0 q_0 n_1 p_1 q_1} = \frac{1}{n_0 p_0 q_0} + \frac{1}{n_1 p_1 q_1}.$$

The sample estimate of $\text{Var}(\ln \widehat{\psi})$ is obtained by substituting the sample estimates $\widehat{p}_i$ and $\widehat{q}_i = 1 - \widehat{p}_i \ (i = 0, 1)$ giving the desired result.

7.2 (Nonconvergence of MLEs in logistic regression):

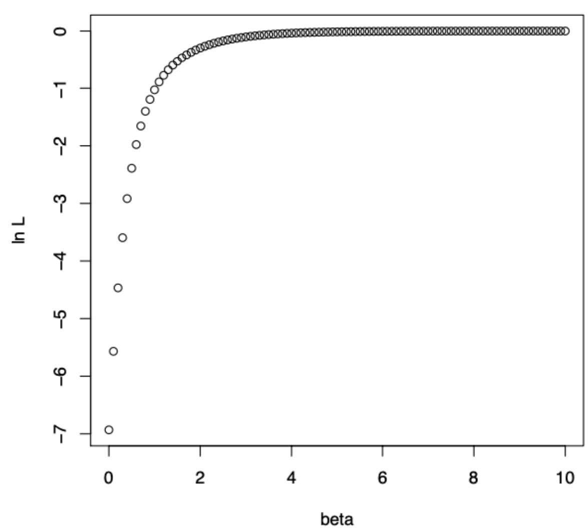
(a) Since $\beta_0 = 0$, denote β_1 simply by β. Then the likelihood function for the given data is

$$\begin{aligned} L &= \prod_{x=-5}^{-1}\left(\frac{1}{1+e^{\beta x}}\right)\prod_{x=1}^{5}\left(\frac{e^{\beta x}}{1+e^{\beta x}}\right) \\ &= \prod_{x=1}^{5}\left(\frac{1}{1+e^{-\beta x}}\right)\prod_{x=1}^{5}\left(\frac{e^{\beta x}}{1+e^{\beta x}}\right) \\ &= \prod_{x=1}^{5}\left(\frac{e^{\beta x}}{1+e^{\beta x}}\right)\prod_{x=1}^{5}\left(\frac{e^{\beta x}}{1+e^{\beta x}}\right) \\ &= \left[\prod_{x=1}^{5}\left(\frac{e^{\beta x}}{1+e^{\beta x}}\right)\right]^2. \end{aligned}$$

So the log-likelihood function equals

$$\ln L = 2\sum_{x=1}^{5}\left[\beta x - \ln(1+e^{\beta x})\right].$$

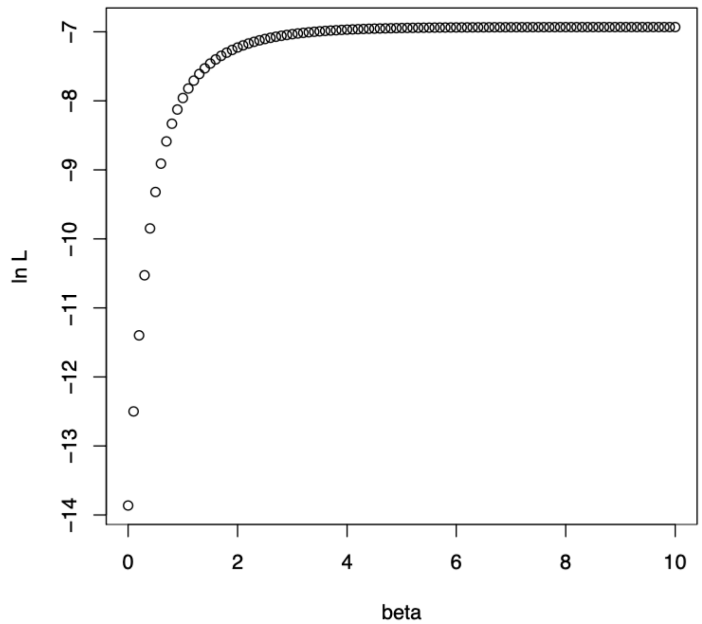
Note that $\ln(1+e^{\beta x}) \to \beta x$ and hence $\ln L \to 0$ or $L \to 1$ (the maximum value of L) as $\beta \to \infty$. So the maximum of L is not achieved for any finite β and so the MLE of β doesn't exist. See the graph below.



(b) If we add two observations $(x = 0, y = 0)$ and $(x = 0, y = 1)$ then each of these two observations adds a multiplying factor $1/2$ to L and hence

$$\ln L = 2\sum_{x=1}^{5}\left[\beta x - \ln(1+e^{\beta x}) - \ln 2\right].$$

As $\beta \to \infty$, $\beta x - \ln(1+e^{\beta x}) \to 0$ as shown above and hence $\ln L \to -10\ln 2 = -6.931$, so the MLE of β doesn't exist. See the graph below.



7.8 **(Radiation therapy)** Twenty four cancer patients were treated with radiation therapy for different number of days (x) and the presence ($y = 0$) or absence ($y = 1$) of tumor was observed.

Days (x)	Response (y)	Days (x)	Response (y)
21	1	51	1
24	1	55	1
25	1	25	0
26	1	29	0
28	1	43	0
31	1	44	0
33	1	46	0
34	1	46	0
35	1	51	0
37	1	55	0
43	1	56	0
49	1	58	0

Source: Tanner (1996), p. 28.

a) Fit a binary logistic regression model to the data.

b) Calculate a 95% confidence interval for the odds of absence of tumor vs. presence of tumor if the number of days of therapy is increased by 5 days.

c) Calculate the estimated success probabilities $\widehat{p}_i$ for the 24 patients in the sample. Find the optimum threshold p^* that maximizes the correct classification rate (CCR). Calculate sensitivity, specificity and the F_1-score for this p^*.